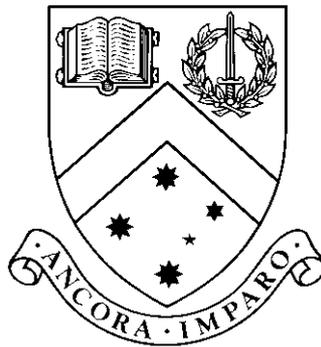


Sustaining Future Digital Forensics Through Intelligent Automation

by

Janis Toms Dalins, Master of Information Technology(Honours),
Bachelor of Computing



Thesis

Submitted by Janis Toms Dalins
for fulfillment of the Requirements for the Degree of
Doctor of Philosophy (0190)

Supervisor: Dr. Campbell Wilson

Associate Supervisor: Dr. Mark J. Carman

Caulfield School of Information Technology
Monash University

February, 2019

© Copyright

by

Janis Toms Dalins

2019

To MT, J and E.

Sustaining Future Digital Forensics Through Intelligent Automation

Janis Toms Dalins, Master of Information Technology(Honours), Bachelor of Computing
`janis.dalins@monash.edu.au`
Monash University, 2019

Supervisor: Dr. Campbell Wilson
`campbell.wilson@monash.edu.au`
Associate Supervisor: Dr. Mark J. Carman
`mark.carman@monash.edu.au`

Abstract

The field of Digital Forensics (DF) has become integral to law enforcement, but demand in terms of quantities of data and items seized is growing at a pace unmatched by available resources, both technical and human. Despite five-fold personnel increases, infrastructure improvements and the introduction of workload reduction techniques such as triage and self-service for routine, low-risk acquisitions, the demand for DF assistance within organisations such as the Australian Federal Police (AFP) remains insatiable. Research into efficiency improvements tends to focus upon data reduction at time of collection or improvements in presentation during analysis - a noble intention, but one incompatible with some jurisdictions' requirement for complete examination in all prosecutions, including uncontested matters. In either case, both approaches do nothing to support the review of the data itself for evidentiary value. Numerous studies report high levels of stress amongst practitioners due to excessive workloads, exacerbated by exposure to offensive materials such as Child Exploitation Material (CEM). The impact of requiring an investigator, analyst or even judicial officer to review 500,000+ CEM files renders any efficiency gains made through data reduction and presentation improvements immaterial, and places personnel and even the judiciary at an unacceptable health and safety risk.

Without adequate research and development of automation both for efficiency and health & safety purposes, the degeneration of DF support levels within law enforcement will continue, if not accelerate.

The efficacy of DF related research lags behind similar and related fields such as image retrieval. Commercial considerations no doubt contribute, but academic and even inter-jurisdictional collaboration are discouraged by legal & ethical considerations regarding the sharing of evidentiary data and even the portability of findings - the definition of seemingly consistent terms such as 'child' varying according to location and context.

This thesis lays the foundations for the sustainable research, development and implementation of automated tools for use throughout the investigation process, regardless of jurisdiction. We ontologise criminal behaviour, introducing the Tor-use Motivation Model

(TMM) as a simple yet robust method for recording not only online behaviour but motivation, consistent with the common law criminal elements of *actus reus* and *mens rea*. We design and develop a deep learning based CEM classifier, in the process identifying and documenting shortcomings of currently used schemas when applied to machine learning. We counter these deficiencies by introducing the *Majura Schema*, an objective, backward compatible, age agnostic pornography ontology focused solely upon tangible attributes rather than abstract concepts such as severity. Finally, we demonstrate the viability of these emergent technologies through the introduction of Monte Carlo Filesystem Search (MCFS), a lightweight, unsupervised and adaptive crawl strategy capable of exploiting their outputs as guidance to significantly accelerate searches for evidentiary materials.

DECLARATION

I hereby declare that this thesis contains no material which has been accepted for the award of any other degree or diploma at any university or equivalent institution and that, to the best of my knowledge and belief, this thesis contains no material previously published or written by another person, except where due reference is made in the text of the thesis.

This thesis includes three original papers published in peer reviewed journals. The core theme of the thesis is sustaining Digital Forensics through automation, principally the automated recognition of evidentiary or otherwise 'of interest' data from a law enforcement perspective. The ideas, development and writing up of all the papers in the thesis were the principal responsibility of myself, the student, working within the Faculty of Information Technology under the supervision of Drs Campbell Wilson and Mark J. Carman.

The inclusion of co-authors reflects the fact that the work came from active collaboration between researchers and acknowledges input into team-based research.

In the case of Chapters 3,4 & 5 my contribution to the work involved the following:

Thesis Chapter	Publication Title	Status <i>(published, in press, accepted or returned for revision)</i>	Nature and % of student contribution	Co-author name(s) Nature and % of Co-author's contribution*	Co-author(s), Monash student Y/N*
3	Criminal motivation on the dark web: A categorisation model for law enforcement	Published	80% Concept, collection of data (including legal clearances), conducting of experiments, writing first draft	Campbell Wilson – Ethical clearance, input into manuscript 10% Mark Carman – Input into manuscript: 10%	No
4	Laying foundations for effective machine learning in law enforcement Majura–A labelling schema for child exploitation materials	Published	65% Concept, collection of data, data analysis, safety framework, conducting of experiments, writing first draft.	Yuriy Tyshetskiy – Data analysis (classifier design) 10% Campbell Wilson – Input into manuscript 10% Mark Carman – Input into manuscript 5% Douglas Boudry – Input into manuscript – 10%	No
5	Monte-Carlo Filesystem Search–A crawl strategy for digital forensics	Published	65% collection of data (including legal clearances), conducting experiments, writing manuscript	Mark Carman – Concept and input into manuscript: 20% Campbell Wilson – Guidance/supervision, input into manuscript 15%	No

I have renumbered sections of submitted or published papers in order to generate a consistent presentation within the thesis.

Student signature:



Date: 31 JUL 18

The undersigned hereby certifies that the above declaration correctly reflects the nature and extent of the student's and co-authors' contributions to this work. In instances where I am not the responsible author I have consulted with the responsible author to agree on the respective contributions of the authors.

Main Supervisor signature:



Date: 31/7/2018

Acknowledgments

I wish to personally thank the Collier Charitable Fund for financial assistance in purchasing computing infrastructure used within this research.

Permission to conduct the crawl of Tor detailed within Chapter 3 was granted by The Honourable Michael Keenan, then (Commonwealth) Minister for Justice, with the endorsement of the AFP and the (Australian) Attorney-General's Department, Cyber-crime Division. The process was managed within the AFP's Specialist Operations, with particular credit and thanks due to Ms Slazana Ristevska.

The CEM classifier detailed within Chapter 4 was developed in conjunction with Dr Yuriy Tyshetskiy of Data61, as part of a secondment granted under an APS 'Data Fellowship', administered by the Department of Prime Minister & Cabinet. To Dr Simon Walsh, sincere thanks for your support and sponsorship of my application for this initiative.

Labelling and annotation experiments in support of Chapter 4 were conducted by AFP volunteers. I won't name the participants given the personal nature these exercises, but to these persons, my personal thanks.

Data for these experiments was made available by AFP DF personnel, on a scale and speed most researchers can only dream of. To the members responsible for the compilation and delivery of what could only be described as one of the most disturbing corpora in existence, thank you.

The vast bulk (if not totality) of this thesis would not have been possible without the support of Supt Doug Boudry, whose coordination and provisioning of resources including the aforementioned assistance were critical to Chapters 3 and 4.

My personal thanks to my supervisors, Drs Campbell Wilson & Mark J. Carman, whose patience in dealing with my mathematical ignorance, non-academic speech and "conversational" writing was (and remains) Herculean.

This research commenced as a rush of blood to the head almost a decade ago. J & E, neither of you was born when this commenced. You have shown a patience and maturity with your wayward parent far beyond your years. Now this work is finished, dad will be around more to be, well, dad!

Above all, my personal love and gratitude to MT. You have sacrificed and persevered through good times and bad. Here's to enjoying more of the former and dealing with the latter, together. Here comes the sun...

Janis Toms Dalins

Monash University
February 2019

Vita

Publications arising from this thesis include:

Dalins, J. , Wilson, C. and Carman, M. (June 2015), Monte-Carlo Filesystem Search
- A crawl strategy for digital forensics. In *Digital Investigation*

Dalins, J. , Wilson, C. and Carman, M. (February 2018), Criminal Motivation on
the Dark Web: A Categorisation Model for Law Enforcement. in *Digital Investiga-
tion*

Dalins, J. , Tyshetskiy, Y. , Wilson, C. Carman, M. and Boudry, D. (September 2018)
Laying Foundations for Effective Machine Learning in Law Enforcement. in *Digital
Investigation*

Permanent Address: Caulfield School of Information Technology
Monash University
Australia

This thesis was typeset with L^AT_EX 2_ε¹ by the author.

¹L^AT_EX 2_ε is an extension of L^AT_EX. L^AT_EX is a collection of macros for T_EX. T_EX is a trademark of the American Mathematical Society. The macros used in formatting this thesis were written by Glenn Maughan and modified by Dean Thompson and David Squire of Monash University.

Contents

Abstract	vii
Acknowledgments	xi
Vita	xii
List of Tables	xvii
List of Figures	xx
Preface	1
1 Introduction	7
1.1 Motivation	8
1.2 Research Questions	9
1.3 Contributions	10
1.4 Chapter Summaries	11
2 Digital Forensics Research & Practice	13
2.1 Digital Forensic Frameworks	16
2.1.1 Process Based Frameworks	16
2.1.2 Technically Focused Frameworks	19
2.1.3 Frameworks Summary	20
2.2 Preservation and Collection	20
2.2.1 Legal Authority	20
2.2.2 Digital Forensic Triage	21
2.2.3 Data Collection	23
2.2.4 Preservation and Collection in Practice	24
2.2.5 Data Reduction	24
2.3 Analysis	27
2.3.1 Visualisation	28
2.3.2 Clustering	29
2.3.3 Crawl Strategies	32
2.3.4 Crawling a Dark Web	35
2.4 Summarising Existing Research	41

2.5	Challenges in Digital Forensics	42
2.5.1	Dangers of CEM Exposure	44
2.6	Towards Automation of Digital Forensics	45
2.6.1	Metadata Based CEM Detection	46
2.6.2	Content Based CEM Detection	49
2.6.3	Cryptographic Digests	50
2.6.4	Fuzzy Hashing	50
2.6.5	Skin Tone Analysis	52
2.6.6	Introducing Machine Learning	54
2.6.7	Defining NSFW, Pornography, and CEM	59
2.6.8	Applying Machine Learning to CEM Detection	63
2.6.9	An Unfortunate Dearth of Data	63
2.7	Conclusions	65
3	Understanding and Categorising Online Criminal Activity	67
3.1	Methodology & Scope	68
3.2	Conducting the Crawl	69
3.2.1	Legal and Ethical Considerations	71
3.2.2	Implementation and Security	71
3.2.3	Seeding and Steering the Crawl	72
3.2.4	Labelling	72
3.3	Results - Applying the TMM	73
3.3.1	Seed Sites	74
3.3.2	All Sites	75
3.4	Conclusions	81
4	Recognising & Classifying CEM	83
4.1	Ensuring Safety	84
4.1.1	CEM Corpora	85
4.1.2	External/'Simulated' Corpora	86
4.1.3	Designing a Classifier Workflow	87
4.1.4	Occlusion maps	93
4.2	Experiments	95
4.2.1	Tor Imagery	96
4.2.2	ImageNet	96
4.2.3	Test Corpus	97
4.3	Classifier Results	98
4.4	Limitations of Existing Schemas	101
4.5	Towards a 'Base' for Cooperation	104
4.5.1	Defining Child Exploitation Imagery	105
4.5.2	Building a <i>Concrete</i> Taxonomy	105
4.5.3	Testing the Schema	107
4.6	Skin Tone Analysis - A Final Critique	111

4.7	Conclusions	113
5	Accelerating Search	115
5.1	Introduction	115
5.2	Developing a File System Crawl Strategy	116
5.3	Monte-Carlo Tree Search	117
5.4	Monte-Carlo Filesystem Search	119
5.4.1	File System Structures & Removing Redundant Visits	120
5.4.2	Treating Content Directories Differently with Virtual Branches	120
5.5	Integrating Domain Knowledge and Heuristics	121
5.5.1	Scoring Nodes	123
5.6	Experiments	125
5.6.1	Dataset	125
5.6.2	File Scoring	127
5.6.3	Node Scoring	128
5.6.4	Test Approaches	129
5.6.5	Evaluation	129
5.7	Calibrating Performance on the Training Set	129
5.7.1	Balancing Exploration vs Exploitation	129
5.7.2	Known Ignorables	130
5.7.3	Optimising File Scorer Weights for Unknown Files	131
5.7.4	Selecting Prioritisers for Informed Search	134
5.8	Findings and Discussion	134
5.8.1	File Scorers	134
5.8.2	Informed Search vs MCFS	135
5.8.3	Internal vs External Devices	135
5.9	Implementing MCFS	139
5.9.1	Demo Crawler	140
5.9.2	Classifier Server	144
5.10	Conclusion	145
6	Conclusions	147
6.1	Future Work	149
6.2	Practical Impacts	150
	Appendix A Chapter 3 Appendices	153
A.1	Initial 500 Page Tag/Label Combinations	154
	Appendix B Chapter 4 Appendices	161
B.1	CETS & COPINE Scales	162
B.2	Test Corpus CAT1 Skin Tone Results	165
B.3	Test Corpus CAT1 Classifier Results	173
B.4	Test Corpus CAT2 Skin Tone Results	181
B.5	Test Corpus CAT2 Classifier Results	189

B.6	Test Corpus CAT3 Skin Tone Results	197
B.7	Test Corpus CAT3 Classifier Results	205
B.8	Test Corpus CAT4 Skin Tone Results	213
B.9	Test Corpus CAT4 Classifier Results	221
B.10	Test Corpus CAT5 Skin Tone Results	229
B.11	Test Corpus CAT5 Classifier Results	237
B.12	Test Corpus CAT7 Skin Tone Results	245
B.13	Test Corpus CAT7 Classifier Results	253
B.14	ImageNet corpus Skin Tone Results	261
B.15	ImageNet Corpus Classifier Results	267
B.16	TorCrawl corpus Skin Tone Results	275
B.17	TorCrawl Corpus Classifier Results	282
B.18	Adult Pornography corpus Skin Tone Results	289
B.19	Adult Pornography Corpus Classifier Results	297
Appendix C Annotation Schema Test v1 Results		305
Appendix D Chapter 5 Appendices		317
D.1	Training Corpus C_p Tuning Results	318
D.2	Training Corpus MD5 Scorer Parameter Tuning Results	320
D.3	PhotoDNA Scorer Parameter Tuning Results	334
D.4	Skin Tone Scorer Parameter Tuning Results	341
D.5	Best First Results	348
D.6	Test Corpus C_p Tuning Results	349
D.7	Test Corpus MD5 Scorer Parameter Tuning Results	352
References		359
Glossary		375

List of Tables

2.1	Suitability Guidelines for Digital Forensic Research (Palmer, 2001)	15
2.2	Investigative Process for Digital Forensic Science	16
2.3	Performance characteristics - search crawl strategies	33
2.4	A comparison of identified Tor hidden site topics or uses	38
2.5	Dark Web crawler seeding methods	41
2.6	Cryptographic Hashes for Figure 2.6	50
2.7	CETS, abbreviated.	62
2.8	COPINE, abbreviated.	62
3.1	TMM Topics	70
3.2	TMM Motivations	70
3.3	Tor Bootstrap sources.	72
3.4	Unique English language seed sites - by category	76
3.5	(Virtual) domains by category	77
3.6	(Virtual) domains by motivation	77
4.1	External corpora unique file counts and descriptions	86
4.2	Tor Top 10 images	97
4.3	Top 10 ImageNet results	98
4.4	Test Corpus Classifier Results	99
4.5	CETS category limitations and advantages - a summary.	104
4.6	Majura Schema - Pornography	107
4.7	Majura Schema - Nudity	107
4.8	Majura Schema - Penetration	108
4.9	Majura Schema - BDSM	108
4.10	Majura Schema - Props	108
4.11	Majura Schema - Virtual	108
4.12	Majura Schema - Bodily Fluids	109
4.13	Majura Schema: Participants	109
4.14	Skin Tone percentages (with median)	112
5.1	Exploration vs Exploitation	118
5.2	Device summary - Training Corpus	126
5.3	Device summary - Test Corpus	127

5.4	Node Scorer algorithms explained	128
5.5	Training set best performing file type weight combinations	138
A.1	Raw labelling data - initial 500 pages	160
B.1	CETS	163
B.2	COPINE	164
B.3	Skin Tone - Test corpus CAT1 top 20 results	168
B.4	Skin Tone - Test corpus CAT1 bottom 20 results	172
B.5	Classifier - Test corpus CAT1 top 20 results	176
B.6	Classifier - Test corpus CAT1 bottom 20 results	180
B.7	Skin Tone - Test corpus CAT2 top 20 results	184
B.8	Skin Tone - Test corpus CAT2 bottom 20 results	188
B.9	Classifier - Test corpus CAT2 top 20 results	192
B.10	Classifier - Test corpus CAT2 bottom 20 results	196
B.11	Skin Tone - Test corpus CAT3 top 20 results	200
B.12	Skin Tone - Test corpus CAT3 bottom 20 results	204
B.13	Classifier - Test corpus CAT3 top 20 results	208
B.14	Classifier - Test corpus CAT3 bottom 20 results	212
B.15	Skin Tone - Test corpus CAT4 top 20 results	216
B.16	Skin Tone - Test corpus CAT4 bottom 20 results	220
B.17	Classifier - Test corpus CAT4 top 20 results	224
B.18	Classifier - Test corpus CAT4 bottom 20 results	228
B.19	Skin Tone - Test corpus CAT5 top 20 results	232
B.20	Skin Tone - Test corpus CAT5 bottom 20 results	236
B.21	Classifier - Test corpus CAT5 top 20 results	240
B.22	Classifier - Test corpus CAT5 bottom 20 results	244
B.23	Skin Tone - Test corpus CAT7 top 20 results	248
B.24	Skin Tone - Test corpus CAT7 bottom 20 results	252
B.25	Classifier - Test corpus CAT7 top 20 results	256
B.26	Classifier - Test corpus CAT7 bottom 20 results	260
B.27	Skin Tone - ImageNet top 20 results	263
B.28	Skin Tone - ImageNet bottom 20 results	266
B.29	Classifier - ImageNet top 20 results	270
B.30	Classifier - ImageNet bottom 20 results	274
B.31	Skin Tone - TorCrawl top 20 results	278
B.32	Skin Tone - TorCrawl bottom 20 results	281
B.33	Classifier - TorCrawl top 20 results	285
B.34	Classifier - TorCrawl bottom 20 results	288
B.35	Skin Tone - Test corpus CAT8 top 20 results	292
B.36	Skin Tone - Test corpus CAT8 bottom 20 results	296
B.37	Classifier - Test corpus CAT8 top 20 results	300
B.38	Classifier - Test corpus CAT8 bottom 20 results	304

D.1	C_p tuning - MD5 Scorer, First FOI found	318
D.2	C_p tuning - MD5 Scorer, All FOI found	319
D.3	Training Corpus - MD5 Scorer (with ignorable file hashset) Parameter Tuning Results - First Hit ($C_p = 0.1$)	320
D.4	Training Corpus - MD5 Scorer (without ignorable file hashset) Parameter Tuning Results - First Hit($C_p = 0.1$)	323
D.5	Training Corpus - MD5 Scorer (with ignorable file hashset) Parameter Tuning Results - All FOI Found ($C_p = 0.2$)	326
D.6	Training Corpus - MD5 Scorer (without ignorable file hashset) Parameter Tuning Results - All FOI Found ($C_p = 0.2$)	330
D.7	Training Corpus - PhotoDNA Scorer (with ignorable file hashset) Parameter Tuning Results - First Hit	334
D.8	Training Corpus - PhotoDNA Scorer (with ignorable file hashset) Parameter Tuning Results - All FOI Found	337
D.9	Training Corpus - Skin Tone Scorer (with ignorable file hashset) Parameter Tuning Results - First Hit	341
D.10	Training Corpus - Skin Tone Scorer (with ignorable file hashset) Parameter Tuning Results - All Found	344
D.11	Best First Test Results - First File Of Interest Found	348
D.12	Best First Test Results - All Files Of Interest Found	348
D.13	Test Corpus C_p tuning - MD5 Scorer, First FOI found	350
D.14	Test Corpus C_p tuning - MD5 Scorer, All FOI found	351
D.15	Test Corpus - MD5 Scorer (with ignorable file hashset) Parameter Tuning Results - First Hit	352
D.16	Test Corpus - MD5 Scorer (without ignorable file hashset) Parameter Tuning Results - First Hit	352
D.17	Test Corpus - MD5 Scorer (with ignorable file hashset) Parameter Tuning Results - All FOI Found	352
D.18	Test Corpus - MD5 Scorer (without ignorable file hashset) Parameter Tuning Results - All FOI Found	355

List of Figures

2.1	Digital Forensic Analysis Cycle (Quick and Choo, 2013a)	18
2.2	DFRWS framework subset	20
2.3	Breadth-first and depth-first crawl examples	33
2.4	Sample screen shot of Silk Road marketplace (Farivar, 2013).	36
2.5	Histogram - Top 100 index terms	48
2.6	Original and altered images	50
2.7	Skin Tone Examples	53
2.8	Skin Tone Mask - <i>Demolition Man</i>	54
2.9	Example Neural Network	56
2.10	Example activation functions	56
2.11	Sample Feature Maps	57
2.12	Microsoft Azure Computer Vision API demo	60
3.1	Tor-use Motivation Model	69
3.2	TMM Labelling App	73
3.3	Top 10 media/content types - The Onion Router (Tor) vs World-Wide Web (WWW)	74
3.4	Site sizes and actual unique content - Tor vs WWW	74
3.5	Labelled Categories - Unique English language sites	75
3.6	(Virtual) domains by category	78
3.7	(Virtual) domains by motivation	79
3.8	(Virtual) domains categorised as Drugs/Narcotics related - motivations	79
3.9	(Virtual) domains categorised as Child Exploitation related - motivations	80
3.10	(Virtual) domains categorised as Illicit/Illegal Pornography related - motivations	80
3.11	(Virtual) domains categorised as adult (i.e. legal) pornography related - motivations	81
4.1	Training and Test Corpora unique image count	87
4.2	OpenNSFW pornography confidences - Test Corpus	88
4.3	Original VGG-16 architecture & fine-tuning	89
4.4	Validation set Receiver Operating Curves (ROC)	92
4.5	Sample Occlusion Maps - 'isChild' classifier	94
4.6	CAT5 movie example classification	95

4.7	Figure 4.6 as stacked area plot.	96
4.8	Imagenet ‘Top 10’ False Positives	97
4.9	Classifier performance on test corpus.	98
4.10	Module Two confidences for images passed from module one (isPorn confidence ≥ 0.8)	100
4.11	Test corpus receiver operating characteristic plots (ROCs). Left: Module 1 (‘isPorn’/OpenNSFW). Right: Module 2 (isChild)	100
4.12	Module Three Classification Confidences - Per Test Corpus Category	102
4.13	Module 3 (CETS) sample occlusion maps.	103
4.14	‘Deepfake’ Katy Perry image	105
4.15	Male, nude (Choi, n.d.). Is it pornography?	111
5.1	MCFS inputs and outputs - a simplified view.	119
5.2	Selection criteria, Prioritisers and Tokenizers	125
5.3	Training Corpus - Testing known ignorables and tuning exploration vs exploitation - First File of Interest Found.	130
5.4	Training Corpus - Testing known ignorables and tuning exploration vs exploitation - All Files of Interest Found.	131
5.5	Training Corpus - File type weight tuning - First File of Interest found.	132
5.6	Training Corpus - File type weight tuning - All Files of Interest found.	133
5.7	Training Corpus - File type weight tuning - First File of Interest found	134
5.8	Training Corpus - File type weight tuning - All Files of Interest found	135
5.9	Training Corpus - Granular scorer file type weight tuning.	136
5.10	Training Corpus - MCFS vs Uninformed and Informed Search	136
5.11	Test Corpus - MCFS vs Uninformed and Informed Search	137
5.12	Example internal device - Full crawl progress.	137
5.13	Example external device - full crawl progress	139
5.14	Crawler Demo Interface	141
5.15	Stonefish Crawler $score(f)$ heatmap	142
5.16	Crawler Demo Report	143
5.17	Classifier Server Sample	144
5.18	Classifier Server Sample - Image Used	145

Preface

Caveat

The research conducted and this resulting dissertation is focused upon improving the efficacy of Digital Forensics (DF) within law enforcement - specifically, within the Commonwealth of Australia. Whilst we don't anticipate any inconsistencies with DF activities within fields such as military and commerce (and indeed have conducted some work with such organisations in confidence), the research, experiments and findings detailed within this work haven't been specifically tested outside our specific context.

By way of background, I am a sworn member of the Australian Federal Police (AFP), bearing the rank of Leading Senior Constable and designation of Federal Agent. I joined the AFP after working in private industry as a software developer and natural language speech recognition subject matter expert, graduating from recruit training in March 2003. Since then I have served in areas including:

- General Duties (i.e. uniform) policing in the Australian Capital Territory (one year);
- Counter Terrorism (four years); and
- Digital Forensics (nine years).

During this time I have had the opportunity to work closely with domestic and international partners, conducting training in Indonesia (Jakarta Center for Law Enforcement Cooperation), contributing to software development projects in the United States of America, and, of course, collaborating on investigations across Australia and in countries such as Vanuatu, the United Kingdom, and the United States of America. I have also worked in areas responsible for public policy and governance, including direct briefing of senior management and government.

Throughout this thesis we focus upon the application of the research within investigations into online child abuse and Child Exploitation Material (CEM) production/distribution. This is due to both personal and pragmatic reasons. Firstly, the creation and distribution of CEM is a particularly abhorrent crime. Secondly, due to the nature of both offending and prosecution, we have observed such investigations' tendency to produce large quantities of high quality, manually annotated data. These techniques are easily ported to differing crime types such as Counter Terrorism, but such a move will require the assembly of datasets of a scale beyond the feasibility of this research.

In terms of workflow, this thesis places a heavy emphasis upon the use of tools and techniques during *triage* - in particular, the filtering/whittling down of items identified during search warrant execution as possibly containing data of interest, ideally to those *definitely* doing so (though this can also include items requiring further examination due to technical limitations of on-site facilities and equipment). The tools and techniques examined and proposed within this thesis are not designed to be used at any specific point in the investigative cycle, nor in any particular jurisdiction. The tools proposed have been designed to be lightweight, portable, and fast in order to facilitate their use in situations of minimal infrastructure, personnel and time - i.e. during the execution of most search warrants. If one succeeds in providing tools for the search warrant use case, then optimisation in the lab is simple.

Ethical Guidance

This research has been conducted in compliance with the inherent values of policing within Australia, particularly in terms of ethical use of data and the avoidance of any ‘hacking’ type methodologies. This is in addition to the standard ethical considerations and clearance processes conducted by Monash University as an academic institution.

Policing in countries such as the United Kingdom and Australia is described as *policing by consent*, a series of considerations and policies commonly referred to as “Peelian Principles” or “Robert Peel’s 9 Principles of Policing”. Whilst their exact provenance is unclear (Robert Peel is not thought to have written them himself), they appear to have evolved from a set of guidelines for Constables during the early days of what is now the Metropolitan Police. Overall, they are perhaps best summarised as an acknowledgment that Police are part of the community (and vice-versa), and that unlike the military, their legitimacy, powers and very existence are only granted by virtue of public support, consent and respect (Home Office (United Kingdom), 2012).

Whereas the provenance, application and even relevance of the principles have been subjected to debate (Loader, 2016), they remain a strong influence on, and also reflection of, the values of ethical policing.

This work’s focus upon data mining and web crawls can be viewed as relating to electronic surveillance, a topic having gained particular prominence in public debate since revelations regarding the activities of the National Security Agency (NSA) made by Edward Snowden. These relate to government intelligence operations, but understandably, the vast bulk of public opinion simply regards such matters as ‘government’ activity, placing little distinction on law enforcement.

There is a place for data/telephone intercepts and mining of data obtained through such means within policing, but these activities *must* be justifiable on a *per instance* case. Within the Commonwealth of Australia all such activities must be judicially authorised on a targeted (per-individual or service) basis, supported by sufficient grounds to justify (in the eyes of the issuing officer) what amounts to an invasion of privacy. No avenue exists to lawfully target groups or the general public.

Peelian Principles

Named after Sir Robert Peel, who as Home Secretary of the United Kingdom founded the organisation now known as the Metropolitan Police, a move often regarded as the foundation of 'modern' policing. The nine principles were used as part of general instructions for recruits from 1829 (Home Office (United Kingdom), 2012), and have since provided guidance for ethical policing worldwide.

Whereas not all below listed principles are relevant to this work (1–3 being the most applicable), all nine are provided for context.

1. To prevent crime and disorder, as an alternative to their repression by military force and severity of legal punishment.
2. To recognise always that the power of the police to fulfil their functions and duties is dependent on public approval of their existence, actions and behaviour and on their ability to secure and maintain public respect.
3. To recognise always that to secure and maintain the respect and approval of the public means also the securing of the willing co-operation of the public in the task of securing observance of laws.
4. To recognise always that the extent to which the co-operation of the public can be secured diminishes proportionately the necessity of the use of physical force and compulsion for achieving police objectives.
5. To seek and preserve public favour, not by pandering to public opinion; but by constantly demonstrating absolutely impartial service to law, in complete independence of policy, and without regard to the justice or injustice of the substance of individual laws, by ready offering of individual service and friendship to all members of the public without regard to their wealth or social standing, by ready exercise of courtesy and friendly good humour; and by ready offering of individual sacrifice in protecting and preserving life.
6. To use physical force only when the exercise of persuasion, advice and warning is found to be insufficient to obtain public co-operation to an extent necessary to secure observance of law or to restore order, and to use only the minimum degree of physical force which is necessary on any particular occasion for achieving a police objective.
7. To maintain at all times a relationship with the public that gives reality to the historic tradition that the police are the public and that the public are the police, the police being only members of the public who are paid to give full time attention to duties which are incumbent on every citizen in the interests of community welfare and existence.
8. To recognise always the need for strict adherence to police-executive functions, and to refrain from even seeming to usurp the powers of the judiciary of avenging individuals or the State, and of authoritatively judging guilt and punishing the guilty.
9. To recognise always that the test of police efficiency is the absence of crime and disorder, and not the visible evidence of police action in dealing with them.

All activities undertaken and outputs sought during this research were considered specifically to remain consistent with the mindset represented by these laws and principles. We therefore self-imposed the following restrictions:

Public data were only obtained where a complete lack of restrictions existed. For example, during our crawl of The Onion Router (Tor) (chapter 3) we only accessed sites listed within public indices or linked to from other such pages. `Robots.txt` files were respected, and logins were never made. Similarly, ‘captchas’ were regarded as a withdrawal of consent for such examination, and sites hosting such infrastructure were dropped from further consideration.

Private data were only made available to this research when already lawfully obtained via search warrant, and even then *only* when established to be illegal via manual review. *No* such private data was shared further, with direct ‘view’ access restricted to authorised Australian Federal Police (AFP) staff. In total, two persons were granted such access throughout this work.

Some data (still images) taken from the aforementioned sources is included within appendices to this dissertation. However, *all* lawful materials are blurred, and illegal materials fully redacted. A ‘law enforcement only’ copy may be produced in future and stored within the AFP library for authorised research purposes, but only for work where the granting of such access is obviously in the public interest.

Technical vulnerabilities are seen as out of scope for our work, given the absence of any individual justification for undermining third parties’ data security. In chapter 3 we mention past research into Tor undertaken using technical exploits to undermine anonymity. Whilst the researchers didn’t de-anonymise any users or attempt to access any normally inaccessible data, we regard the indiscriminate use of such methods as against the *policing by consent* model. Therefore, no attempts were made to identify any contemporary equivalents to past exploits.

Ontologisation of networks and data was undertaken specifically to aid the avoidance of inefficient search - not only for economic reasons, but also to ensure the accessing of innocent third parties’ public data is kept to an absolute minimum. Given the established legal and ethical restrictions around accessing of private data, the accessing of such information is regarded as completely out of scope for our research.

This research has not been conducted to enable increased (in terms of targets) surveillance, nor does it seek or introduce any methods for undermining existing privacy safeguards - technical, physical or legal.

A Final Caveat

Unless otherwise cited, any opinions and anecdotal observations described within this thesis are directly drawn from first hand experience. They should not necessarily be regarded as those of the AFP, Monash University, or any other parties.

Chapter 1

Introduction

Forensic, *adj. and n.*: Pertaining to, connected with, or used in courts of law (Oxford English Dictionary, 2018)

The term ‘forensics’ tends to invoke visions based on popular television programs such as “CSI” and its many offshoots, whereby fearless forensic analysts rapidly examine crime scenes, cadavers and computers, invariably identifying methodologies, motivations and offenders. Arrests almost invariably follow within a one hour period, including commercials. Setting aside the sheer investigative speeds implied, the only accurate element within this perception is the examination of possible crime scenes and things (including persons) of interest. The remainder can largely be summarised as misconceptions made for entertainment purposes alone.

To paraphrase a broad topic, we define Digital Forensics (DF) as the **identification, evaluation and presentation of data from electronic devices for the primary purpose of presentation in court**. As a law enforcement field, it has evolved from individuals and ad-hoc teams working on specialist matters to dedicated laboratories embedded within relevant agencies worldwide, largely in response to the vast quantities of electronic devices and data emerging as a result of pervasive computing. Furthermore, the significance of *data* as evidence has evolved in sync with that of data in general throughout business and social life worldwide. Combined, these have resulted in the field encountering near immeasurable growth in workload and importance within investigations and prosecutions globally.

The reality of DF is somewhat less glamorous than that alluded to in the aforementioned television programs, and certainly not as efficient. DF is undeniably more important within criminal investigations now than it was ten years ago, and is more often critical to the success of complex investigations, particularly where *mens rea*¹ and conspiracy are required proofs. For example, a senior Counter Terrorism investigator from ██████████² explained that during a recent investigation into a (foiled) terrorist attack involving home made explosives, they’d rather have had more DF analysts

¹‘Guilty mind’ - summarisable as intent to commit a crime

²Given under ‘Chatham House rules’ of non-attribution

than crime scene examiners, as establishing the presence of explosives and their precursors in a location is relatively simple³. Proving the mindset and intended *actions* of a bomb manufacturer is more difficult, often relying upon past communications and interactions as a means for examining motivation, likely paths of radicalisation, possible targets, and even co-conspirators. In the case of so-called *lone wolf*⁴ attackers, the absence of known, identifiable physical interactions makes electronic records particularly important (if not essential).

1.1 Motivation

Whilst DF is typically separated from investigators organisationally, the rise of pervasive computing has resulted in something of a parallel evolution in law enforcement. Even the ‘lowest tech’ offences such as affray (typified by drunken brawling) will result in the examination of Closed Circuit Television (CCTV) systems and mobile phones for recorded footage, rendering DF itself ubiquitous, if not integral, to criminal investigations.

This “ubiquitous DF” means law enforcement has encountered something of a double hit - not only are more investigators requesting DF assistance, they are requesting it for more aspects of their cases. This has placed an unmeetable load on resources, leading to requests for assistance being postponed, reduced or rejected outright. Like many such organisations, the AFP has secured increased funding in the area (Borys, 2017), but the growth of data and devices requiring examination (including on-site triage) is too large to deal with through additional personnel and infrastructure alone.

A further complicating factor contributing to the challenges facing DF is the *psychological* harm being encountered by practitioners. *Burn out*⁵ relating to increased workloads is an obvious symptom, but the damage caused by long term exposure to offensive materials such as Child Exploitation Material (CEM) and violence/gore (associated with online radicalisation) is now being recognised as a major workplace health & safety concern. The risks associated with CEM exposure have been known for some time, with issues such as secondary traumatic stress (Seigfried-Spellar, 2017) and secondary victimhood (Brown et al., 1999) well known for investigators. Contrastingly, the dangers associated with *analysts* appear to have been underestimated or overlooked, with recent work starting to identify stressors encountered even by those without direct victim/offender interactions (Seigfried-Spellar, 2017; Powell et al., 2015).

DF research lags behind other information retrieval related fields. Beyond commercial considerations (DF largely being a government and specialist field), access to quality data corpora in and of itself is difficult, largely due to legal and ethical restrictions. CEM related research is particularly constrained by understandably strict laws on the storage and transmission of such materials - made all the more onerous when shared between jurisdictions with differing laws.

³We emphasise not ‘easy’

⁴Individuals seemingly radicalised, preparing and conducting terrorist acts in isolation

⁵Mental exhaustion, typically associated with long periods of stress

In summary, DF within law enforcement can't keep pace with demand without a significant effort at reducing analyst workloads and stressors. This thesis does not aim to provide complete solutions for the challenges facing DF today - this will only be achieved through cooperation between industry, government, the legal fraternity and judiciary within each affected jurisdiction. An ambitious goal, to say the least. As detailed within the Preface, this research is primarily aimed at reducing stresses and harms directly resulting from large workloads of often monotonous, repetitive and psychologically harmful tasks. Improved automation is an obvious first step towards this goal, but can only be achieved if DF is lifted beyond mere ubiquity to integration throughout the investigations lifecycle, with tools, inputs and outputs easily transported between teams, organisations and jurisdictions.

1.2 Research Questions

In this work we specifically note that the *identification* of a crime having occurred is outside the traditional role of 'forensics' - a view resulting in the "DF as a service" model, with knowledge siloed within investigations and forensic teams. We believe these silos directly contributed to the underestimation (if not complete ignorance) of practitioner welfare issues. Whilst not responsible for detecting criminal behaviour, DF as an organisational unit and science needs to be informed by investigators in order to carry out its pivotal role - the preservation of evidence. If we wish to automate the recognition and classification of evidentiary materials, we must be able to algorithmically explain what it is we are seeking. Such 'mapping' of specialist concepts is not in and of itself novel, but as we will go on to discuss, the absence of a global standard directly applicable to law enforcement has directly limited the ability to share tools and techniques across jurisdictions. We therefore ask:

Can online criminality be robustly ontologised? Can online criminal behaviour be classified according to a jurisdictionally independent ontology? If so, can that ontology be sufficiently flexible without becoming confusing and ambiguous?

An established ontology directly informs automated classification, providing a class structure for the target materials - in this case, imagery. Image classification is a mature topic, particularly in commercially valuable fields such as search. The *safe* annotation and supervised training of a classifier using materials known to cause psychological harm (refer Section 2.5.1) is a challenge still being felt by industry and government, yet the efficiency and safety dividends presented by such automation has the potential to provide the richest rewards to law enforcement. We therefore ask:

Can offensive materials be automatically recognised and classified reliably, with minimal labelling? Can broad topics such as child exploitation be automatically identified against existing labelling schemas currently in use by law enforcement?

The first two research questions aid in preparation for preservation and collection of evidence. However, the vast bulk of evidentiary data is typically encountered, preserved and collected on site during the execution of search warrants - situations where even basic facilities such as electricity and a safe working environment can't be assured. Whilst the automated classification of materials introduced in the previous research questions is undoubtedly of value, there is the potential for improving the performance of the search itself. In this work we therefore also discuss crawl strategies, a family of techniques used to traverse networks in a highly efficient manner. Directing a crawler to intelligently focus upon materials of interest is not a novel concept, and indeed we could implement a manually directed crawler quite easily. However, the adaptation of such a methodology, using the results output by an automated classifier, can lead not only to faster search (both in terms of time and files examined), but also *unattended* search. We therefore ask the question:

Can automated classifiers prioritise search for evidentiary electronic materials?

Can classifiers efficiently inform automated search within digital forensics, particularly in time-critical situations where additional computational infrastructure is scarce or unavailable?

1.3 Contributions

Our research makes the following research contributions:

TMM The Tor-use Motivation Model (TMM) is a two dimensional taxonomy for online behaviour on *dark webs*, with a particular interest in criminal & 'of interest'⁶ behaviour. Whereas existing schemas tend to categorise content according to underlying subject matter ('pornography', 'narcotics' etc), the TMM incorporates *motivation* as a means for improving granularity without introducing ambiguity. *The TMM has been peer reviewed and published in the Journal of Digital Investigation (Dalins, Wilson and Carman, 2018).*

Automated CEM Classifier We construct and demonstrate a deep learning based classifier for the automated classification of CEM. Unlike previous research in the field, and with close and rigorously regulated co-operation with the Australian Federal Police (AFP), the classifier was trained and validated on data taken from numerous real world cases. Furthermore, it was tested on thousands of images taken from an unrelated criminal investigation, ensuring the quality of results. Using deep learning, we show the classifier can be effectively trained with annotations already generated by law enforcement personnel as part of investigations, rather than requiring manual identification of features.

Majura Schema During the development of the classifier, we observed that existing CEM schemas are too abstract for the purpose of adequately training automated classifiers. In response, we introduce the Majura Schema, an ontology focused upon visible content

⁶to law enforcement

rather than abstract concepts such as ‘sadism’. *The Majura Schema and aforementioned CEM classifier have been peer reviewed and published together in the Journal of Digital Investigation (Dalins, Tyshetskiy, Wilson, Carman and Boudry, 2018).*

MCFS Monte Carlo Filesystem Search (MCFS) is a Monte-Carlo based tree search algorithm, specifically adapted and tested for optimising searches across file systems within electronic media - a key DF activity, particularly during search warrant execution. The algorithm is lightweight and modular, capable of being directed by existing methodologies such as cryptographic hashes. MCFS does not require domain specific knowledge, enabling portability between media types, source applications and even languages used in naming files & directories. However, it is capable of exploiting knowledge such as metadata when prioritising steps, further improving performance. *MCFS has been peer reviewed and published in the Journal of Digital Investigation (Dalins et al., 2015).*

1.4 Chapter Summaries

Chapter 2: This chapter provides an overview of DF, existing research and the current state of practice. It demonstrates how the field has evolved from a subordinate of information security to become a specialised ‘forensic’ role, focusing specifically on identifying what *has* occurred, rather than securing against what *could* occur. A level of disagreement remains between practitioners as to the role of DF in protection, though as the chapter demonstrates, it is perhaps wise to view DF’s outputs as *informing* security efforts rather than being an integrated part of such matters. The chapter details the challenges facing DF as a field, particularly due to the growth of data and device volumes at a rate unmatched by any feasible allocation of infrastructure and human resources. The chapter provides a case study of these challenges, showing how exposure to offensive media such as CEM adds further psychological burden on police and DF practitioners, at levels previously underestimated by law enforcement and academia alike. Finally, the chapter details the ‘dearth’ of available data, and its impact on research into technological responses for addressing these urgent challenges.

Chapter 3: A perhaps underestimated challenge in developing automated tools for identifying and classifying content is understanding precisely what is (and is not) being sought. One need only attempt to define pornography⁷ to understand how context dependent, constantly evolving and vague such a ‘simple’ concept can actually be. Introducing elements required for identifying illegality and then translating such a definition into machine readable form adds further complication, with such ontologies needing to find balance between comprehensiveness, flexibility and clarity. Whilst not designed or implemented for illegal activities, the anonymity afforded by *dark webs* such as The Onion Router (Tor) (Dingleline et al., 2004) has attracted an illicit user base, making their content an ideal starting point for defining what constitutes criminal behaviour online. Whilst numerous schemas

⁷In this case, lawful pornography

and ontologies exist for dark webs, in Chapter 2 we show these to be inadequate for law enforcement use, with ambiguity or vagueness contributing to a lack of accuracy (CEM being categorised as ‘adult content’ in one instance). Chapter 3 details our extensive and unrestricted crawl of Tor, achieved through unprecedented levels of support from the Australian Government and law enforcement agencies⁸. We use all data obtained as a means for developing an ontology specifically suited for police use. We demonstrate how an uncontrolled vocabulary approach provides simplicity, but rapidly leads to ambiguity and confusion - online criminality being a far richer and more diverse ‘community’ than we’d anticipated. We therefore introduce the TMM as a means for dealing with this diversity, demonstrating how the introduction of *motivation* as a second dimension provides a far greater level of granularity. This is critical when one considers the ‘edge’/hard to anticipate cases caused by the constant evolution and variability of criminality itself.

Chapter 4: This chapter demonstrates the introduction of machine learning to illegal content recognition, with the development and testing of a deep learning based classifier to detect and classify materials against a CEM schema currently in use across Australian jurisdictions. We show that the classifier performs particularly well on the one class actually focused upon an unambiguous feature, with the remaining classes’ reliance upon abstract or vague concepts directly contributing to poor performance. We therefore develop and introduce *Majura*, a labelling schema focused primarily upon unambiguous attributes. Whilst more complex (and therefore possibly more onerous) than existing labelling schemas such as the Child Exploitation Tracking System (CETS) and the Combating Paedophile Information Networks in Europe (COPINE) index, its flexibility also allows for the creation and use of datasets of lawful materials as proxies for CEM, opening the field to research and collaboration beyond law enforcement and associated entities.

Chapter 5: One of the challenges identified by DF practitioners is a difficulty in applying current research to their everyday work. The aforementioned classifier is presentable via a RESTful API, making it accessible to the vast majority of tools and applications available today. Chapter 5 takes this one step further, demonstrating how focused crawls can exploit existing knowledge such as CEM hashsets as a means for not only identifying known items of interest, but also for doing so more quickly and without the need for operator supervision. We develop and demonstrate the efficacy of MCFS across numerous ‘real world’ media seized as part of AFP investigations, showing it to robustly outperform traditional search methodologies, particularly during device triage.

Chapter 6: This chapter concludes the dissertation, summarising our work, detailing future directions (including some currently in progress) and practical implications of our findings.

⁸In keeping with Australian law, written permission was obtained from the Commonwealth Minister for Justice in order to allow the accessing, transmission and storage of CEM. The Attorney General’s Department and AFP also granted specific permission, with State authorities informed via joint policing arrangements.

Chapter 2

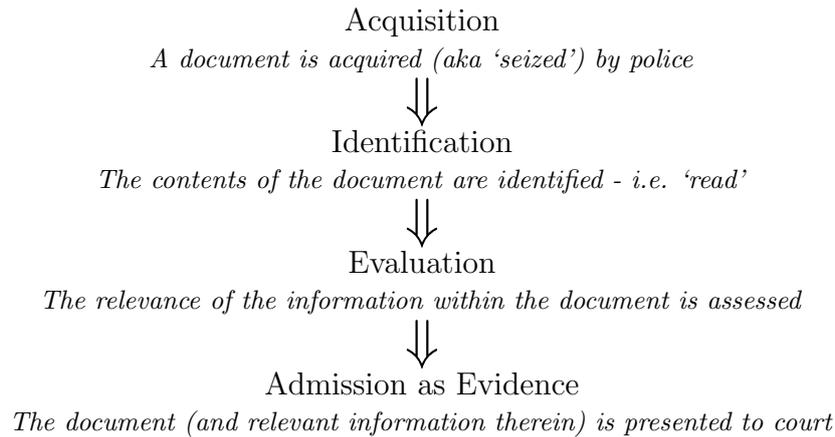
Digital Forensics Research & Practice

Digital Forensic Science is not in the
business of protection (Palmer, 2001)

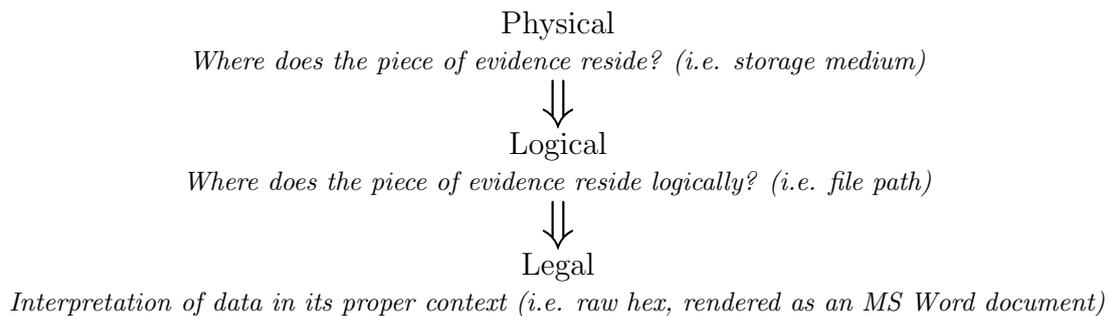
The emergence, evolution and growth of Digital Forensics (DF) as a profession and field of research appears to have followed that of mobile communications and pervasive computing, where at some point, electronic data started to *regularly* become relevant within criminal investigations and intelligence matters. In the specific case of the Australian Federal Police (AFP), this occurred around the early 1990s. In the absence of any guidelines, standards, accepted projections or even definitions, practitioners in this nascent field largely worked in relative isolation.

The earliest peer reviewed paper regarding computer forensics¹ we could identify is by Pollitt (1995), who provides a somewhat philosophical description of the field as “... *the application of science and engineering to the legal problem of digital evidence. It is a synthesis of science and law. At one extreme is the pure science of ones and zeros. At this level, the laws of physics and mathematics rule. At the other extreme, is the courtroom.*”. The author, an FBI Special Agent, specifically contrasts electronic evidence with the “Paper Paradigm”, whereby a paper-based item of evidence goes through a four part process within an investigation, being:

¹A decreasingly used synonym of digital forensics



Under this process, specialist knowledge (provided via an ‘expert² witness’) is only required in the last two stages (if at all), and only to help establish relevance. Simply put, all stages beyond acquisition are plainly visible, and judges & juries are perfectly capable of examining physical items and reading printed text. Contrastingly, electronic evidence is seen as going through a separate process, with three “contexts”:



The entire process requires specialist knowledge, with only the final context’s output properly aligning with the ‘Paper Paradigm’, but only after the data has been rendered in a readily human interpretable form. Without the specialist knowledge offered by Digital Forensics (DF), electronic data isn’t presentable as evidence, and as the author states, if information is not admitted as evidence, “it doesn’t exist”.

The Digital Forensic Research Workshop (DFRWS) appears to be the first forum dedicated to peer reviewed research and practice in DF. Before this time, relevant research tended to find an audience within information security and legal conferences and journals. Held annually since 2001 with a stated audience of “*military, civilian and law enforcement professionals...*”, the workshop has influenced the establishment and definition of DF as an independent field, particularly beyond its use in government. A decisive, inaugural item of business was the definition of DF as a science - not a simple task when one considers participants’ differing environments and priorities, as shown in Table 2.1.

²defined in the State of Victoria (Australia) as a person who has specialised knowledge based on the person’s training, study or experience (Supreme Court of Victoria (Australia), 2015)

Digital Forensics (DF) in the Australian Federal Police (AFP)

The AFP's DF capability has evolved from ad-hoc networks of technically minded staff (\lesssim mid 1990s), to small investigative teams (\approx 2000), to a specialist capability offering, for want of a better term, DF 'as a service' (\gtrsim early 2002), whereby investigators request support and analysis on a per-case basis not dissimilar to an outsourcing arrangement. To give an idea of the scale of growth, the AFP DF function has grown from six people in 2004 to >50 members in 2018, despite the introduction of 'self service' functionality for routine, simple items and aggressive data triage (discussed later in the chapter). In spite of this growth, a material proportion of requests for DF assistance are rejected, deferred, or severely restricted due to lack of resources.

Area	Primary Objective	Secondary Objective	Environment
Law Enforcement	Prosecution		After the fact
Military IW ³ Operations	Continuity of Operations	Prosecution	Real Time
Business & Industry	Availability of Service	Prosecution	Real Time

Table 2.1: Suitability Guidelines for Digital Forensic Research (Palmer, 2001)

Table 2.1 shows the main difference in differing practitioners' priorities lies with law enforcement, whose focus upon prosecution and post incident investigations contrasts with remaining participants' focus upon availability and real-time response. The evolution of law enforcement itself means these guidelines are somewhat dated, though by no means obsolete. Policing no longer has a sole focus on prosecution as a primary objective, with organisations such as the AFP moving to also support disruption of criminal activity and networks (Australian Federal Police, 2017), not only due to public safety considerations (e.g. counter terrorism) but also as a means of overcoming jurisdictional limitations caused by the international nature of contemporary crime.

The differing priorities prompted the introduction of a holistic definition of DF as a science, included here as *Digital Forensic Science*. Doubtlessly wordy, it nonetheless provides a good snapshot of the activities performed within the wider field:

Digital Forensic Science

The use of scientifically derived and proven methods toward the preservation, collection, validation, identification, analysis, interpretation, documentation and presentation of digital evidence derived from digital sources for the purpose of facilitating or furthering the reconstruction of events found to be criminal, or helping to anticipate unauthorized actions shown to be disruptive to planned operations. (Palmer, 2001)

An important caveat provided by the participants reflects the focus on evidentiary proofs rather than security faults, cementing the role of DF within incident response: "*Digital Forensic Science is not in the business of protection*".

³Information Warfare

-	Name	Description
1	<i>Identification</i>	<i>The establishment of the crime/incident having occurred.</i>
2	Preservation	The preservation of evidence in a forensically sound manner (e.g. imaging of data, chain of custody).
3	<i>Collection</i>	<i>The collection of data/items relevant to the case.</i>
4	Examination	Identification of deleted/hidden items, basic pattern matching/text search.
5	Analysis	Exploitation and further interpretation of data, including activities such as data mining.
6	<i>Presentation</i>	<i>Documentation and testimony associated with analyst findings from the previous steps.</i>
7	<i>Decision</i> (Out of scope)	<i>Implementation of changes to processes/infrastructure in response to incident.</i>

Table 2.2: “Investigative Process for Digital Forensic Science”(Palmer, 2001). *Italics* denote debate regarding status as ‘forensic’ categories.

2.1 Digital Forensic Frameworks

Now that we have established what DF is and *isn't*, we need to understand what the actual process entails. Numerous digital investigation frameworks have been defined, usually process based and focusing either upon the entire investigative life cycle or specific components (often associated with the authors’ specific skill sets).

The inaugural DFRWS (Palmer, 2001) issued a technical report including a seven category “*Investigative Process for Digital Forensic Science*”, detailed in Figure 2.2. Items listed in *italics* are acknowledged as being most open to debate as ‘forensic’ categories. The paper also acknowledges some discussion as to whether Preservation is a subset of Collection - a reasonable view, as the primary purpose of collecting evidence is to preserve its existence for later analysis and ultimate production in court. The inclusion of the (admittedly debatable) ‘Decision’ category reflects some degree of disagreement between participants, being contradictory to the previously quoted and unequivocal statement regarding non-involvement of DF in “protection”. Nonetheless, the document (plus the underlying investigative process) appears to have influenced government guidelines, with the National Institute of Standards and Technology (NIST) guide to introducing DF into incident response (Kent et al., 2006) loosely tracking the categories.

By being exclusively focused upon post-incident identification of events, our research takes the opinion of DF mirroring the traditional role of law enforcement - supporting or refuting allegations/charges through the establishment of proofs. Item 7 is therefore regarded as out of scope and is not further discussed within this dissertation.

2.1.1 Process Based Frameworks

The DFRWS focus upon process has heavily influenced subsequent framework proposals, with most at least reflecting elements therein. Reith et al. (2002) identify a lack of standardisation within DF frameworks, largely due to a focus on specific technologies rather than the underlying process. The authors provide an example of a methodology (Prosis

et al., 2003)⁴, detailing DFRWS process-esque steps such as incident detection and response. Acknowledging the methodology as “well thought out”, the authors nonetheless criticise the specification of “Windows NT/2000, UNIX and Cisco Routers”, pointing out that the focus on “computer crime” means devices such as PDAs, mobile phones and peripherals are not considered. This remains a prescient critique, as the evolution of mobile computing in the form of tablets, smart phones, etc. has greatly increased the potential significance of such items within investigations, regardless of crime type. The authors therefore move away from such pitfalls by expressly proposing “*An Abstract Digital Forensics Model*”, differing heavily from the DFRWS model. For example, the authors introduce preparation and strategic steps between Identification and Preservation, and also add a “Returning Evidence” stage to close out the process.

Stephenson (2003) introduces End-to-End Digital Investigation (EEDI), a “collection of steps to be taken in conjunction with the DFRWS framework”. A key element of the paper is the introduction of Digital Investigation Process Language (DIPL), a formal language for documenting the investigation process. This focus on language results in the provision of definitions for key elements within the DFRWS process. According to the author, a by-product of heavier process regulation is the ability to work with more sophisticated tools such as link analysers, but the primary benefit of this approach appears to be more focused toward providing a framework for ensuring the quality and reproducibility of the investigative process - not unlike ISO accreditation.

Based upon crime scene theory such as Locard’s Exchange Principle⁵, the “Integrated Digital Investigation Process” (IDIP) model (Carrier and Spafford, 2003) defines the ‘digital crime scene’. Whilst building on work such as the EEDI, the IDIP differs by moving closer to traditional forensic models, with seventeen phases across five groups. The groups themselves more closely resemble traditional crime scene sciences, with physical and digital crime scene investigation phases.

Ieong (2006) proposes the FORensics ZAchman (FORZA) framework⁶ as a means to integrate legal considerations within digital investigations. Heavy on defining individuals’ roles within investigations, it is light on detail in terms of *technical* aspects, instead defining desired outcomes for abstract layers such as the *data analysis layer*.

Beebe and Clark (2005), identifying the limitations of single tier process models, instead design a hierarchical framework, based upon investigative objectives. The authors contrast forensic and non-forensic investigations, using two hypothetical use scenarios (including an investigation into Child Exploitation Material (CEM) trading) focused upon the Analysis phase as a means to demonstrate the advantages of focusing upon objectives rather than tasks. This paper represents a strong attempt at designing a framework capable of representing overarching holistic views right through to detailed, device-specific taskings. The authors themselves argue the framework is incomplete, targeted more at

⁴We have been unable to source a copy of the original text

⁵Summarisable as “*Every contact leaves a trace*” (Horswell and Fowler, 2004)

⁶Based upon an unpublished work relating to enterprise architecture

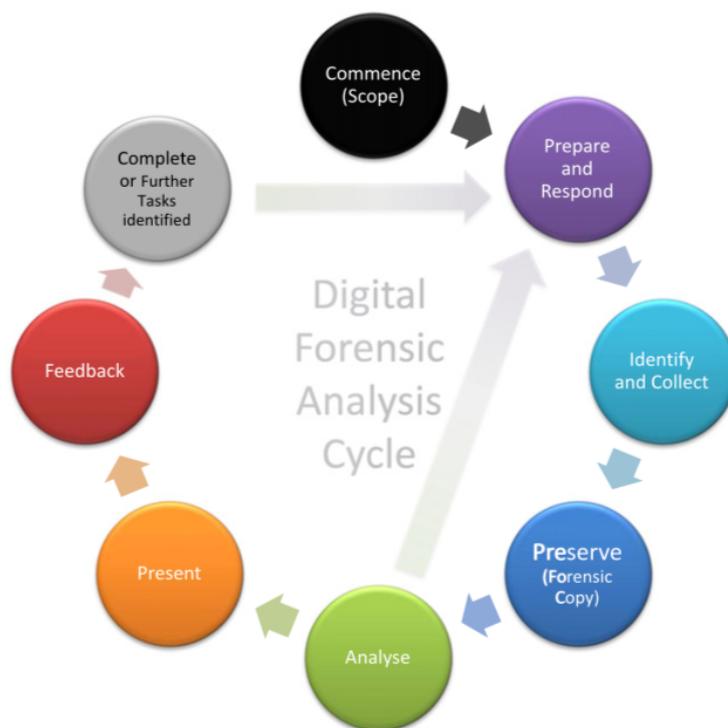


Figure 2.1: Digital Forensic Analysis Cycle (Quick and Choo, 2013a)

generating discussion within the DF community than representing a finalised, ‘as is’ solution. The authors openly admit that platform-specific tasking frameworks are conceivable, and in fact use this to support their objective based approach.

The emergence of cloud computing as a storage and processing medium is being addressed by research such as that by Martini and Choo (2012). A conceptual work closely aligned with the aforementioned NIST guidelines (Kent et al., 2006), it identifies the iterative nature of identifying and working with remote data at time of evidence collection *and* during examination & analysis, where artefacts may indicate the use of previously unanticipated providers. An evolution of this work by Quick and Choo (2013a) constructs a more granular process for working with cloud hosted data, as displayed in Figure 2.1. Of particular note, every step is bidirectional, acknowledging the very real possibility that an unknown remote storage account (as an example) could be identified at any time, necessitating a return to a previous step - in the case of our example, most likely to either “Prepare and Respond” or “Identify and Collect”. Whilst other cloud frameworks have been proposed, these two are of particular relevance to our research, as firstly, the research is conducted from the perspective of Australian practitioners⁷, but secondly, the framework has been tested on scenarios using MS SkyDrive, DropBox (Quick and Choo, 2013b) and Google Drive (Quick and Choo, 2014), representing a large proportion (if not clear majority) of the domestic cloud storage market.

⁷The authors are based in South Australia, with Quick being an Electronic Evidence analyst with the SA Police

2.1.2 Technically Focused Frameworks

A less holistic but far more nuanced alternative to focusing upon process is to generate frameworks around specific tasks and tools. Carrier (2003) moves beyond process design, instead providing an abstraction layer based framework for digital analysis tools. The author provides the example of File Allocation Table (FAT) file system analysis, separating the task of listing directory contents into seven layers, extending from the raw file system (used for boot sector values) up to extrapolating file details from allocated clusters.

By introducing the Computer History Model, Carrier and Spafford (2006) move to define digital investigation process categories based upon device history, rather than how existing investigations and tools operate. As the authors state, *“the unique contribution of this paper is that the categories are based on how digital evidence is created...”*

Petroni et al. (2006) provide another example of a purely technically focused framework, with FATKit (Forensic Analysis ToolKit) proposed as a modular method for improving the efficiency of low-level data extraction/analysis as it relates to volatile memory. The main value presented by FATKit is the abstraction of analysis, with the software providing means for automatically identifying objects within memory. Primarily targeting the C programming language, the authors’ focus upon modularity provides the means to allow other languages and processes to be added to FATKit’s functionality.

The PyFlag (Cohen, 2008) framework is similar to FATKit, with a target of network communications. It provides analysts with abstraction tools for rendering otherwise complex, lower level data (for example, by rendering HTML pages from raw packet captures). Whilst impossible to objectively quantify, the emergence of tools such as PyFlag was very significant, coming at a time when law enforcement was encountering extremely large increases in telecommunications intercept (‘wire tap’) data volumes in line with a boom in mobile communications⁸.

Attempts at technical frameworks for cloud computing also exist, but due to the architecture’s very nature, direct DF practitioner access to infrastructure and media is limited, typically by geography but also legal and practical considerations. Search warrants give law enforcement the right to enter premises and seize data, but disruptions to businesses (particularly innocent third parties) must be kept to a minimum, and only when reasonable and necessary. Unless a hosting service is seen as a co-offender or otherwise hostile to law enforcement, a mass outage of innocent third parties’ services is not justifiable. Furthermore, the proprietary, distributed nature of storage within major providers would make a physical examination technically unfeasible. Alex and Kishore (2017) propose the use of external computing infrastructure and a “forensic monitoring plane” as a means for bypassing these restrictions, but the approach appears to rely upon telecommunications intercepts, a methodology strongly restricted within Australia and well outside the ‘response’ model typically associated with DF. Realistically, DF responses to cloud hosted data and services at a technical level will be dictated by the providers themselves, and will be limited to data dumps rather than physical level examinations.

⁸Personal observation by author.

2.1.3 Frameworks Summary

Whilst broad in their application, the frameworks listed above share common origins from the “Investigative Process for Digital Forensic Science” (Palmer, 2001) discussed earlier within this section. Figure 2.2 displays a subset of this framework, focusing upon the categories directly relevant to the research reported within this work. Preservation and Collection have been merged, consistent with the authors’ reference to debate regarding the discrete nature of these stages.

Preservation & Collection	Examination	Analysis
Legal Authority	Data Preservation	Data Preservation
Data Acquisition	Keyword Searches	Data Mining
Data Reduction	Data carving	Link Analysis
Sampling	Hidden Data	<i>Clustering</i>
<i>Crawl Strategy</i>	Discovery/Extraction	

Figure 2.2: Subset of DFRWS framework - *Italics* denote proposed activities/enhancements

2.2 Preservation and Collection

As shown within Figure 2.2, the initial framework category relates to the *preservation* and *collection* of data, both from technical and legal perspectives. Legislative powers and restrictions are particularly prevalent within criminal investigations, whilst issues such as data acquisition and data reduction are concerns across all facets of DF.

2.2.1 Legal Authority

The direct acquisition of evidence during investigations typically occurs via the execution of search warrants, granting investigators access to premises and data for assessment and seizure (as required).

The scope of examinations undertaken during search warrants is set by relevant legislation. Search warrants in Australian Commonwealth criminal investigations are typically issued by virtue of Section 3E of the *Crimes Act 1914* (Cth), giving investigators permission to search specified premises, conveyances, and/or person(s) and seize evidential material⁹ where reasonable grounds for suspicion that such evidential materials are present or will be present within 72 hours.

The concept of ‘seizing’ logical data (as opposed to physical items) is taken into account. To paraphrase the relevant legislation, electronic items can be seized if the investigator believes on reasonable grounds that an item **or data accessed by operating the**

⁹a “thing” relevant to an indictable or summary Commonwealth offence or State offence with a federal aspect

item is evidential material¹⁰ - pending any debate regarding the process, the presentation of data *copied* from an item is equivalent to presenting the item itself. Helpful in avoiding unreasonable disruptions to businesses and innocent third parties, this provision is critical for accessing remotely stored data.

Section 3L (“Use of electronic equipment at premises”) of the *Crimes Act 1914* (Cth) regulates access to and copying of data, including data “not held at the premises”. This distinction is quite important, as online/‘cloud’ data storage services such as Microsoft SkyDrive™ and Google Drive™ provide free, convenient location transparent storage services worldwide. Such data can be obtained via mutual assistance request (MAR), but such requests can be slow and rely completely upon the service provider and data storage location being known and the relevant jurisdiction(s) being cooperative. Furthermore, services such as SpiderOak(Spideroak, 2015) now offer encrypted storage, whereby the service provider itself has no ability to access and interpret user data. Therefore, the ability to at least cursorily examine any relevant online storage services becomes critical during search warrant execution.

An investigator executing a search warrant under Australian Commonwealth legislation is therefore faced with the challenge of examining all potentially relevant electronic devices within or accessible from a target premises before being able to seize items and/or copy data. Typically, this analysis will involve manual browsing of data (perhaps with the targeting of specific features), or the calculation of file hashes with subsequent comparisons against pre-established hash sets of known files. Both processes are resource intensive from a computational, bandwidth, and/or human perspective.

The quantity and nature of items encountered during search warrant execution is largely unpredictable, making the process of allocating resources extremely difficult. The process of thoroughly examining individual items *prior* to seizure is largely a luxury nowadays, with analysts employing triage methods not unlike hospital emergency wards for prioritising and allocating taskings.

2.2.2 Digital Forensic Triage

Digital forensic triage, defined by Roussev et al. (2013) as “*a partial forensic examination conducted under (significant) time and resource constraints*”, is the process of examining items for the purposes of finding a subset of relevant data - typically, ‘enough’ to achieve the task at hand. The subset needn’t be specifically pre-identified - in our scenario of search warrant execution, the desired outcome is typically finding evidence sufficient for the investigator(s) to establish *belief* that the item or data constitutes evidential material, and is therefore seizable.

Using an example of a CEM investigation, upon commencing a search of premises, an investigator should prioritise examination of electronic items and media according to their capability and practicability in being used to commission¹¹ the offence. For example, a games console such as the Nintendo Wii™ has web browsing capability, but the

¹⁰Section 3K *Crimes Act 1914* (Cth) allows for the temporary moving of items offsite for examination, but this is subject to time restrictions and other considerations

¹¹Perform/enact/commit the act of

impracticality of using a games controller for navigation makes a network connected PC far more likely to be used for such activity. Therefore, examination of these devices would be prioritised accordingly.

Device examination can occur either *live*/'as is' (typically when the device is located powered on and operational), or 'dead', via either a specialised boot device or through the removal of storage media and read-only connection to examiner devices via a write blocker. The latter is the preferred option, due to the inherent safeguards against inadvertently corrupting or otherwise altering data. Depending upon examiner preference and individual circumstances, devices found powered on and operational at time of examination may be subjected to more thorough examination. Given the resources required for such examination, this tends to be limited to situations where a danger of data loss exists, typically due to encryption or other obfuscation methods.

A search for CEM can be approached from several angles. A triage approach will typically commence at the search for 'low hanging fruit'¹², continuing through to slower, more thorough examinations. For example, the process may include:

1. **'Live' search:** In cases where the device is running, the examiner may look for relevant user behaviour (recent files, internet history etc), or potential obstructions such as encryption.
2. **Locations of interest file search:** A quick check of relevant, known locations of interest for CEM related files - for example, the *Downloads* directory in a case of browser-based access and downloads. Such an approach is reliant upon the suspect user(s) maintaining such default behaviours, and not moving/deleting data of interest.
3. **Relevant application history search:** An examination of logs, libraries, history files etc associated with the application(s) suspected to have been used during the offence. For example, *Limewire*(Internet Archive, 2015), a now discontinued P2P file sharing application, saved attributes of files encountered on the network and download logs by default, providing analysts with a ready source of information regarding user behaviour.
4. **Filename search:** Fast, lightweight search of filenames for terms known to be associated with the specific matter, and CEM in general. In matters where specific files are sought (e.g. where the suspect is thought to have been party to specific file transfers), a search of filenames akin to the Unix/POSIX *find*(The Open Group, 2017) utility can be extremely fast and effective.
5. **Hash Digest search:** A comparison of all accessible files against a known file of interest (FOI) hashset. Whilst slow, this has the advantage of low user interaction due to the negligible (if not non-existent) level of false positives encountered using cryptographic hashes such as Message Digest 5 (MD5) or Secure Hash Algorithm 1

¹²Easily identifiable items of interest

(SHA-1). Similarity digests (described below) provide a more flexible but more computationally expensive approach, exploiting similarities as a means for identifying data of interest without being affected by minor text/image changes.

6. **Data carving (file recovery):** Search of the device’s raw contents (bypassing the file system), recovering files based upon known signatures such as file headers. Computationally expensive, plus completely reliant upon (1) the crawler recognising such signatures, and (2) in the case of deleted files, being intact enough for detection.
7. **Full text search:** As with data carving the device’s raw data is searched, this time for text strings of interest. This step can be carried out in isolation, but will not detect strings stored in other formats (e.g. deleted docx files) unless the crawler is capable of recognising and parsing such formats.

Roussev and Quates (2012) identify the slow performance of ‘deep forensic’ examinations, instead choosing to extend the use of hash-based searches to similarity digests. The move away from searching for *identical* data provides a means for identifying relationships across sources and establishing “an initial framework of understanding” - for example, by finding similar text files across devices. This approach could serve as an effective means for identifying related or altered data (e.g. chat logs from other parties’ perspectives, cropped/altered photographs, etc.). Interestingly, the authors use sequential access of the target physical storage devices as a means for accelerating search by removing seek time latencies. This effectively ignores the logical file system layout, and therefore is completely reliant upon the host operating system’s implementation of physical storage. A great deal of early sectors within devices will typically be used during Operating System (OS) installation, and therefore the crawler could risk becoming bogged down in irrelevant OS files rather than user generated data. The authors identify metadata based prioritisation as an option for improving performance, and this indeed would most likely be a suitable approach for lower latency devices such as SSDs.

An interesting analysis of the performance impacts caused by triage methods is presented by Roussev et al. (2013), who perform typical investigative tasks on a reference target using ‘workstation’ and ‘server’ configurations, reflecting on-site and lab-based triage. Whereas metadata extraction and cryptographic hashing perform well on the workstation, more intensive methods such as indexing and similarity hashing “*become somewhat feasible on the server*”.

2.2.3 Data Collection

Complete examination (as opposed to triage) follows the decision for seizure. Under optimal conditions, an item’s storage devices are *imaged* - a copying process where raw (i.e. binary level) data is acquired, rather than logical copies of files presented by the file system(s), prior to further examinations being undertaken. This is a slow and resource intensive process, but the additional information obtained from unused sectors, file system metadata and even the *physical* distribution of data can provide vital context and background to otherwise plainly visible files. Of course, the detection/‘carving’ of deleted or

otherwise obfuscated data is heavily reliant on having a complete copy of all data present on devices.

2.2.4 Preservation and Collection in Practice

Preservation and collection of data is typically conducted on a physical, per-device basis. If the triage and preview process identifies seizable material on (for example) a computer or HDD, the device is then seized as per the authorising legislation or body (e.g. search warrant, subpoena etc). The device itself is effectively ‘preserved’, in so far as physical access is now limited to authorised persons working under the direct authority and control of the investigating body. In cases of online services such as Google Drive, the level of service available to the end user is collected as per the investigator’s authority - for example, by downloading an account holder’s data using a service such as Google Takeout (Google, n.d.), or via the use of legal authorities such as subpoenas in the service provider’s jurisdiction.

The main exemption to the ‘per device’ focus comes from practicality, common sense, and/or legislative considerations. For example, data seized from innocent third parties (e.g. a suspect’s bank or university) is typically conducted on a targeted basis - the seizing of a university’s email server in order to obtain a copy of a single student’s mailbox would be deemed excessive in all but the most extreme cases - not just from the processing of large quantities of irrelevant data, but also from the disruption to third parties. One would have extreme difficulty justifying the shutdown and seizure of a bank’s servers in order to obtain a suspect’s financial records.

To date, this differentiation is largely carried out on a common sense basis, informed by investigators’ experiences in dealing with subject organisations and devices. Targeted data seizure has also been a focus of academic research, principally around data reduction.

2.2.5 Data Reduction

A strict, unrelenting requirement for the complete preservation and collection of relevant storage devices and services is a noble desire, providing investigators with a maximum quantity of data - both incriminating and possibly exonerating. It is also terribly inefficient, as evidentiary value needn’t correlate with logical size, and data of relevance needn’t constitute a material proportion of a storage device or service’s capacity. In fact, the increasing size of storage devices and use of cloud processing and storage infrastructure have been identified as issues for DF practitioners, even being described as part of a “coming digital forensic crisis” (Garfinkel, 2010). Unlike encryption, neither issue directly prevents the preservation/collection and analysis of data within individual cases - the effect is far more widespread, slowing down the machinery of law enforcement through the need for increased infrastructure, resources and time to store and process data.

Spafford (Palmer, 2001) identified the challenge inherent in the standard approach of “collecting everything”, leading “to examination and scrutiny of volumes of data heretofore unheard of”. He cited a research focus involving identifying what data is required to ensure the highest levels of accurate analysis.

Ferraro and Russell (2004) provide a strong discussion on the challenges encountered within digital investigations, particularly from an organisational perspective. One such challenge is the perceived tension between *expert examination* and *investigative review*, resulting in issues such as the the “*usually unnecessary*” use of full-blown forensic examination (an issue discussed later within this chapter). The authors cite the example of an investigation into child pornography possession, stating that “*an examination looking for questionable images should be sufficient to obtain the necessary evidence*”.

A response proposed within academia and industry is the process of data reduction: limiting analysis to data known (or highly likely) to be of interest to the investigator, either at time of collection/preservation or analysis.

Reduction At Collection

Data reduction at time of collection relates to the triage of *data*, as opposed to *devices*, - changing from a “shotgun” approach to that of a “sniper” (Pogue, 2011). As with the sufficiency argument presented within section 2.2.2, known irrelevant data can be discarded from the investigative process, though at the risk of introducing false negatives - incorrectly disregarded pertinent information. This is a natural response to the ever increasing storage media sizes encountered, even within the domestic use market. (Culley, 2003) identified the questionable efficiency of imaging all data¹³.

Post Preservation

A more cautious approach to data reduction comes after preservation/collection, with the most typical scenario being through the use of selective acquisition and analysis of seized items.

An example of a readily implementable post preservation analytical framework is provided by Quick and Choo (2016), who introduce Data Reduction by Selective Imaging (DRbSI). To paraphrase the paper, DRbSI provides a means for reducing processing time and infrastructure requirements by acquiring “information dense” files, most likely to be of interest to investigators, including internet history, email containers, images and multimedia. Acquired data is compressed as part of the process, with images and movies converted to thumbnails instead of being copied ‘as is’. The authors quote results of near 99% reduction in the quantity of data copied for analysis, obtaining “74% of the information in 11% of the processing time”.

Such efficiencies are seductive from a workflow perspective, but problematic when one considers the criminal prosecution landscape rather than simple process acceleration. Firstly, the requirement to prove guilt “beyond reasonable doubt” requires assertions to be supported by evidence capable of surviving strong scrutiny within court. Context is therefore key. The paper’s authors describe unallocated clusters and system files such as `pagefile` and `hiberfile` as “information light”, requiring a large amount of space whilst providing minimal information. From a purely volumetric perspective, this is a reasonable

¹³As an aside, the article identifies a 500GB external HDD as costing \$1000 at the time of writing, perhaps inadvertently giving us an indication as to the drop in storage media costs worldwide

assumption. As mentioned previously, the evidentiary significance of data needn't correlate with logical size - a web page or password cached within a 4GB+ sized pagefile can be critical evidence, but only constitute an immaterial proportion of data by size.

Discussion

The design and implementation of workable, peer-reviewed frameworks, procedures and guidelines in fields such as DF can be of great value, particularly in adversarial applications such as law enforcement. Work carried out by relatively junior, inexperienced investigators and analysts can be endorsed and defended by leaders in the field if it can be shown that such frameworks were followed.

The more prescriptive a framework, the more easily it can be consistently implemented and defended in court. On the other hand, unrelenting rigidity in a field as variable and unpredictable as law enforcement can also be a burden. A simple approach such as copying and preserving a target user's home directory (e.g. `c:\users\JanisDalins`) will quite possibly gather sufficient evidence to support an investigator's allegations, particularly in cases involving unsophisticated offenders (assuming single drive, self contained systems). Such an approach implicitly labels the remainder of the storage device as 'not of interest', raising the very real risk of false negatives¹⁴ in the form of overlooked evidence.

An issue with triage and data reduction previously identified by Pollitt (2013) is that of inadvertently limiting one's scope. Investigators will tend to look for evidence obviously supporting the underlying allegation(s), possibly overlooking evidence of a different crime, unexpected aspects (e.g. conspiracy), or worse yet, exculpatory evidence¹⁵. Such a risk is always present, but is arguably increased when time and resources are at a premium, as is the case during search warrants and other such field work. The collocation of differing, seemingly unrelated materials of interest such as CEM and violent imagery has been noted by Edelman (2010); Powell et al. (2015). Anecdotal reports from within the AFP and associated agencies indicate an over-representation of CEM being identified on electronic devices seized during Counter Terrorism (CT) investigations. Numerous theories have been posited by investigators, but we are unaware of any quantified research into the reasons for this particular correlation.

On this basis, the use of data reduction methodologies during the preservation/collection phases is potentially problematic, as attempts at obtaining subsequent access to subject devices will most likely be restricted by (a) reluctance by issuing authorities to authorise repeated search warrants, lest they constitute harassment; and (b) the high likelihood of evidentiary material being lost, hidden or destroyed once law enforcement interest is revealed.

It is impossible to objectively measure the risk of such false negatives on a case-by-case basis, and we are yet to observe any larger scale research into such an approach based upon real-world datasets.

¹⁴The labelling of 'positive' (of interest) data as ignorable/'negative'.

¹⁵Evidence *refuting* an allegation - for example, an alibi.

False negatives are difficult to detect without corroborating information, and often impossible to overcome. Obfuscation and concealment are common in criminal behaviour, with encryption and steganography examples of readily available technologies commonly encountered by DF practitioners. In a search warrant situation involving data reduction prior to preservation/collection, an analyst is required to select ‘relevant’ files and data from storage devices containing potentially thousands of candidate documents, creating a risk of critical information being missed. The risk of such a false negative is difficult (if not impossible) to quantify without a suspect’s (unlikely) cooperation, denying practitioners the opportunity to undertake informed risk acceptance, mitigation or rejection.

For these reasons, prescriptive frameworks are unsuitable for data reduction, particularly at time of collection. This is an issue identified by Quick and Choo (2016), who rightly acknowledge the need for the DF practitioner and/or investigator to make their own decisions when selecting data for extraction, based upon personal expertise and experience. The procedures proposed by the authors are, in our opinion, the most complete and readily applicable of this type for law enforcement, at least from the Australian perspective. This is largely due to the authors’ own insistence that much of the selection of data of interest is largely at the investigator/analyst’s discretion - no ‘one size fits all’ claims are made, including during which phase data reduction should be undertaken.

We are sceptical of data reduction at the time of collection/preservation, and cautious as to the value of data reduction post collection, for practical rather than academic reasons. The gathering of evidence is specifically required due to an absence of certainty regarding an investigation - in other words, an absence of reasonable doubt means a conviction is already achievable and any further investigative activity is redundant. The overlooking of evidence due to a ‘blinkered’ search for specific data is dangerous, and courts typically require an expert witness to ensure *all* relevant examinations are undertaken (Supreme Court of Victoria (Australia), 2015). For example, it would be exceedingly difficult (if not impossible) to disprove a defence involving malware/viruses if a complete copy of all executable files (and related libraries) was not made.

Unanticipated requests for additional analysis can occur even in situations of full confession and undisputed facts. For example, courts within the State of Victoria (Australia) commonly request full analysis of suspects’ devices (including manual annotation of CEM) in uncontested matters as a means for informing sentencing. It is also not unheard of for a court to request analysis be undertaken to refute *potential* (as opposed to actual) defences in cases, though this could be due to concerns regarding a defendant’s motivations for pleading guilty. For this reason we posit that data reduction may be used to accelerate processing and analysis, but only in situations where either a full master copy (or the source device itself) is preserved and remains available *throughout* the investigation and subsequent prosecution/appeals processes.

2.3 Analysis

The analysis phase of investigations refers largely to the *exploitation* of data acquired during the preceding stages, by identifying wider patterns and/or characteristics between files,

devices or overall sources. Typically focused upon text search and known *of interest/not of interest* search, pattern matching is the fundamental element of investigation - string matching being perhaps the base level method for identifying documents of interest. The efficiency of such searches varies widely and can be particularly affected by the original terms sought. Searches for long, less frequent character combinations will provide high levels of precision, but will suffer from misspellings and regional variations (e.g. ‘organise’ vs. ‘organize’). On the other hand, searches for common abbreviations (often seen in child exploitation investigations) suffer from large numbers of false positives (100,000+ false positives *per device* is not uncommon, due to some abbreviations being common to CEM and internal Windows operating system processes), presenting the risk of valuable data being overlooked due to resource limitations.

2.3.1 Visualisation

The first (and surprisingly complex) step in improving search result usability within DF is presentation - interpreting hundreds (if not thousands) of keyword search results is tedious work, leading to increased probabilities of incorrectly labelled and overlooked data. Beebe et al. (2011) identified that the presentation of search hits in two leading DF products actually underperformed a random walk¹⁶, largely due to the inability to filter or prioritise results. A more effective means of presenting data to possibly non-expert users needs to be found.

The field of DF is under-served in terms of data visualisation, particularly when compared with the broader field of data analytics. Forensic analysis software such as EnCase[®] (Guidance Software, 2018) and X-Ways[®] (X-Ways Software AG, n.d.) both feature text-heavy GUIs and processes, supporting in-depth technical analysis on a *per-device* basis. Data such as identified entities are presented in a largely tabular fashion, with more specialised components usually introduced as add-on modules using proprietary scripting languages. Nuix[®] (Nuix, n.d.), an e-discovery product gaining support in law enforcement, features a ‘context’ module, plotting entities and relationships as edge/node graphs. We believe this constitutes the first commercially available entity visualisation tool *within* a forensic analysis product¹⁷.

Teelink and Erbacher (2006) designed and implemented two visualisation tools specifically for DF - one hierarchical, and one non-hierarchical. The hierarchical tool represents the file system structure using a tree map, with a colouring scheme for distinguishing branches, and node sizes determined by logical size. The non-hierarchical representation simply shows a ‘flattened’ view of a directory’s contents, including subdirectories. Both representations utilise colour as a means for reporting differentiation (e.g. logical size). The non-hierarchical plot effectively provides a means for reporting file clustering to the user, where unusual/outlier files’ colour (and potentially other properties) can be used to identify anomalies and hidden data. The use of colour and size allows even untrained users

¹⁶A strategy whereby one traverses a network using completely random steps

¹⁷Timeline visualisations are available in some products, but these tend to be focused upon readily available metadata rather than identified and extracted entities

to rapidly observe patterns of files within a device: The authors report experiments comparing their tool with text-based (Linux shell) tools, observing a 35% reduction in time identifying files of interest, and a 57% reduction in identifying the first file of interest. These are impressive results, but a comparison with an industrially accepted tool such as EnCase[®] or Autopsy/TSK (Carrier, n.d.) would provide a more like-for-like comparison. Furthermore, the test focused upon the identification of altered or hidden files - a valid scenario, but perhaps not truly representative of the wider spread of scenarios typically encountered by DF practitioners.

2.3.2 Clustering

Clustering is the process of grouping individual documents into coherent topics/categories, identified via statistical analysis.

Early research into clustering within DF introduced the concept of adding processing steps during post-retrieval text string search, with data being classified prior to being searched, ranked and/or clustered, finally being presented to the examiner for analysis (Beebe and Dietrich, 2007). Kohonen Self-Organizing Maps were proposed as the means for clustering, largely due to their unsupervised nature and linear scaling against data set size (Beebe and Clark, 2007).

Separate research into the efficacy of clustering within digital forensics saw the use of partitional (K-means, K-medoids), hierarchical (Single Link, Complete Link, Average Link) and cluster ensemble (CSPA) algorithms on real-world datasets from Brazilian Federal Police investigations (da Cruz Nassif and Hruschka, 2013). Average Link and Complete Link algorithms performed best on the datasets, with “suitably initialized” K-means and K-medoids also performing well. The paper’s findings include:

- **File names:** The authors found that despite file names being insufficient for computing dissimilarities between files in isolation, they did assist the clustering process;
- **Relevant/irrelevant:** Documents tended to cluster around relevant and irrelevant materials, justifying the clustering process.
- **Outliers:** The removal of outliers during clustering didn’t result in improved performance. This is an interesting result (particularly in terms of simplified processing), but the authors note this could be data dependent and may not generalize to other examinations.

Beebe et al. (2011) build upon the proposals listed within Beebe and Dietrich (2007); Beebe and Clark (2007), clustering documents containing keyword search hits as specified by a volunteer examiner. The authors implement and measure precision and recall using a real-world dataset (a divorce case). This paper presents a large jump over the authors’ earlier papers, with a full test system implemented, including a GUI for user interaction. Key points of note from this paper:

- **Documents:** Whereas da Cruz Nassif and Hruschka (2013) focused their examinations on clustering documents, the authors here actually include full (byte level) search of the device, treating slack space as individual documents. Documents containing examiner-defined keyword terms are clustered prior to presentation to the user.
- **SSOM:** Scalable Self-Organizing Maps are used for clustering, due to the authors' desire for this solution to readily scale.
- **Performance metrics:** Precision¹⁸ and recall¹⁹ are measured as a function of search term priority²⁰, measured at cut-off points in the user's activity. In other words, instead of measuring precision in terms of raw hits/matches, *relevance* is included as a metric: $P = \frac{n \text{ relevant hits}}{N \text{ hits}}$.
- **Search hit analysis:** The authors use a GUI to present keyword search hits to the user, displaying hits by document clusters. The user is free to leave a cluster once he/she suspects all relevant data has been found.
- **Reproducibility:** The time taken by the user to evaluate hits (relevant/irrelevant) is recorded, and an average used to measure clock time in further simulations.
- **Real-world comparison:** The evaluation is extended to cover the same behaviour using traditional text search on commonly used forensic software, namely EnCase[®] and FTK[®].

An interesting finding (discussed in the previous section) by the authors is the poor performance of EnCase[®] and FTK[®] in terms of precision and recall. The authors theorise this is due to investigators selecting search terms purely on relevance to the investigation at hand, with no consideration made to potential noise²¹. This appears a reasonable assumption, though it remains surprising and concerning that a random walk of search hits can outperform these tools. In all likelihood, this is due to a traditional focus upon identifying and returning *all* potential evidence, at the expense of potentially swamping analysts with extraneous data.

Metadata Based Clustering

Fei et al. (2005) provide a broad overview of the potential of self organizing maps (SOMs) as a means for improving efficiency for examiners undertaking manual examination of data for CEM imagery. The authors provide the example of temporary internet files,

¹⁸The fraction of correct results within a search result. If a search returns 5 results and 2 are valid, then precision = 0.4

¹⁹The fraction of correct results *found* by a search - e.g. if there are 10 correct records and the search returns 8, recall = 0.8

²⁰Search terms are assumed to be of varying interest/priority to the analyst, with some terms generating large degrees of valid (but irrelevant) hits - i.e. 'noise'.

²¹A common issue when short (less than four character) strings are sought across large, compressed datasets, where even random distribution will result in thousands of matches

images typically generated and stored by the user's web browser as a by-product of general browsing. In this instance, an SOM was used to generate cluster maps based upon files' extension, creation date and creation time²². This experiment showed that images' clustering could be used to identify browsing habits, potentially allowing an examiner to quickly focus upon suspect areas based upon user behaviour.

The practicality of the scenario presented by this paper is somewhat limited, for the following reasons:

- **File extension:** The authors' claimed use of file extensions is erroneous, as they state elsewhere that their forensic software (FTK) will identify graphics with false extensions. This is a move beyond simple metadata analysis.
- **Files analysed:** The use case is based upon temporary internet files, which are programmatically created by users' web browser(s). The focus upon only one source of 'interesting' files undermines the analytical process, as another approach would be to simply focus all analysis upon directories known to be used by web browsers.
- **Low detail of dataset:** The authors' dataset is a selection of files extracted using forensic software, with no detail given as to the distribution of relevant/irrelevant files, nor logical structure.

The aforementioned items should not be taken to be criticisms of the paper, which appears to be more an introduction to the possibilities of such an approach than an in-depth analysis. The limited nature of available results (temporal analysis of online user activity being the primary outcome) leads to the belief that at least some analysis of file *content* is required for more effective search.

Content Based Clustering

Research on topic modelling within DF focuses largely on text clustering, with a view to accelerating manual investigator reviews of seized data by increasing precision with an acceptable trade-off of recall. Common observations of the DF landscape are that typical datasets are heterogeneous, unstructured (Beebe et al., 2011), and there often is little or no prior knowledge of the data (da Cruz Nassif and Hruschka, 2013).

Arguably, some degree of *a priori* knowledge may be available in circumstances where specific domains of digital investigations are being searched. Unlike search of an electronic storage device, an investigator/analyst searching a company's email archives will have some degree of anticipation as to what will be encountered, at least in terms of format and the general nature of emails²³. For this reason, research into general text search on devices and domain-specific search (in this case, emails) are separated within this section.

Clustering Email Search

Decherchi et al. (2009) investigate whether individual users' roles within a company (Enron) could be identified via clustering of their individual email mailbox contents. In this

²²All fields were converted into numeric values for processing

²³It is arguably unlikely to encounter multiple GB multimedia files as attachments, for example.

case, the cluster count was set to 10, due to the authors' desire to "obtain a limited number of informative groups". Some commentary indicates a level of success in identifying randomly selected staff members' activities, but the paper provides little (if any) critical analysis of the algorithm's performance, with a brief discussion sufficing as a summary. At best, the research and findings of this paper could be used to partially guide investigators undertaking a holistic overview of activities, but it does not seem to provide any true means for increasing precision or at least effectively prioritising accounts of interest.

Estimating Cluster Counts

A key challenge identified within da Cruz Nassif and Hruschka (2013) is that of accurately estimating the number of clusters present within a dataset, especially considering the limited (if not non-existent) knowledge *a priori*. This issue was also identified by Stoffel et al. (2010), though admittedly in a related but not identical field. DF related research has involved the use of empirically selected counts (Beebe and Liu, 2014; Decherchi et al., 2009). da Cruz Nassif and Hruschka (2013) use *silhouette*, a relative validity index, as a means for conducting topic count estimation. Such an approach comes with a non-trivial processing overhead, often requiring repeated passes over the data. One imagines some degree of *a priori* knowledge of a storage device's contents could assist in at least estimating topic count, but as previously mentioned, this would be rarely forthcoming.

The problem of estimating the number of clusters in a data set is difficult, underlined by the fact that there is no clear definition of a 'cluster' (Tibshirani et al., 2001). Any proposed topic count solution requires testing and validation prior to use, presumably against a known good clustering dataset. The establishment of such a dataset isn't as simple as it may appear, with the seemingly simple task of validating the cluster heavily influenced by the selection of a relevant null model - the specification of which is "*not a straightforward matter*" (Gordon, 1996).

We are unaware of any discussion of null model selection for cluster validation within the DF focused papers cited within this chapter, as most measures of efficacy within these works are focused upon comparison with more traditional search practices. This perhaps comes as a result of the complexities associated with measuring clustering performance - beyond challenges associated with formulating null models, Tibshirani et al. (2001) point out that the difficulty of estimating cluster numbers is "*underlined by the fact that there is no clear definition of a 'cluster'*".

2.3.3 Crawl Strategies

A performance improvement, rather than a forensic process in and of itself, the efficacy of crawl strategies²⁴ has already been proven in unrelated fields such as World-Wide Web (WWW) search and indexing. Currently, a typical search within a digital investigation will be undertaken using arbitrary rules - at the logical/file system level, breadth or depth first (refer Figure 2.3) are simple approaches, with best first²⁵ a more complicated but

²⁴The methods used to optimise paths taken by crawlers across networks such as the WWW.

²⁵Selection of nodes according to pre-defined rules for establishing priority

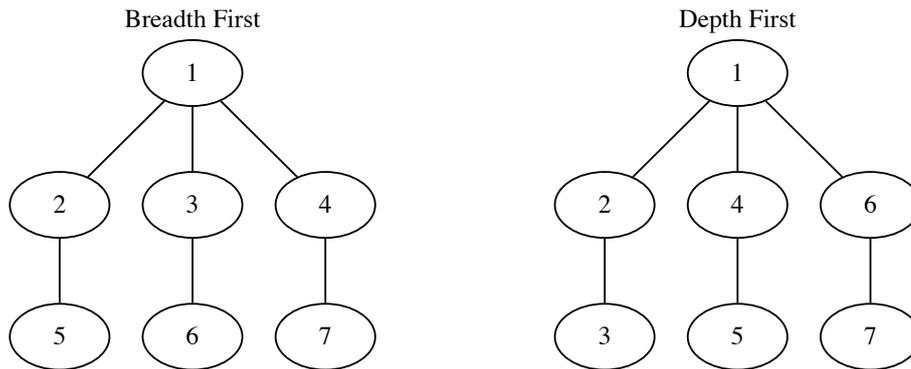


Figure 2.3: Node selection orders taken by breadth-first (left) and depth-first (right) crawls.

potentially more rewarding approach. Variations of these can be used at the physical level, based upon the *physical* properties of a device. These approaches have particular merits, as displayed within table 2.3.

	Memory Overhead	Speed
Breadth/Depth First	Minor	Device dependent (seek time latencies)
Best First (Metadata based)	Classifier Dependent	
Physical (stream)	Negligible	Fastest

Table 2.3: Performance characteristics - search crawl strategies

Focused Crawling

The bedrock principle behind our methods for accelerating forensic analysis is making relevant data available for examination as quickly as possible, particularly during time critical activities such as search warrant execution. An obvious starting point is to assess the application of a *focused crawler* for traversing file systems resident within target media.

Chakrabarti et al. (1999) first proposed focused crawlers as a resource whose goal “*is to selectively seek out pages that are relevant to a pre-defined set of topics*”. Designed around the challenge of indexing hypertext documents within a rapidly growing WWW, the authors found that combining a classifier evaluating document relevance with a *distiller* recognising documents linking to relevant information was an effective method for improving indexing efficiency.

The authors provide a very good justification for intelligent crawl strategies within large-scale search, by contrasting the then-dominant *Altavista* and *Inktomi* search engines’ use of (contemporarily) high-end hardware clusters, in contrast to the authors’ use of a Pentium II PC. They argue the aforementioned search engines attempted indexing as much data as possible in order to answer every possible user query, and therefore traversed and indexed a great deal of irrelevant and low value data during the process. Focused crawlers’

ability to classify document relevance allows them to identify high value locations (in terms of specialised sites and/or high proportions of links to relevant documents), achieving knowledge of relevant data without large-scale crawls. Hence the use of comparatively low-end hardware is not due to the budgetary restrictions commonly associated with research, but rather because the crawler achieves its goals quickly, and has “*relatively little to do*”.

Diligenti et al. (2000) experiment with *context graphs* for focused crawlers, introducing the Context Focused Crawler (CFC). Identifying the optimal focused crawler as retrieving “...the maximal set of relevant pages while simultaneously traversing the minimal number of irrelevant documents...”, the authors identify performance risks such as pursuing immediate links of middling value at the expense of less obvious (but more valuable) links. Introducing context knowledge therefore allows the crawler to predict links’ value more effectively (particularly in terms of multiple hops), and indeed the authors report up to 50-60% performance improvement, in terms of relevant documents obtained during a set period.

Chakrabarti et al. (2002) take a different approach to automating focused crawler training, introducing an *apprentice* to the focused crawler model. The apprentice serves to prioritise unvisited pages, whilst the original classifier becomes a trainer for the apprentice. This approach is shown to be an effective means for improving the process of predicting unvisited pages’ value (reducing false positives by up to 90%), but the authors rely upon Document Object Model (DOM) features within referencing pages - an information source unique to the WWW.

Contrast with File System Search

It is most likely apparent to the reader that the logical structure of the WWW is quite different from the file systems present within typical electronic media. Beyond the structure changing from network to tree, links between WWW pages are established *within* documents, giving a crawler a great deal more context. The tree structures inherent within file systems are defined by metadata outside a document, and whilst influenced by user behaviour, are largely dictated by OS and installed applications.

The CFC approach of Diligenti et al. (2000) has tradeoffs. Context graphs need to be generated for a “*reasonable fraction*” of seed documents, which in turn requires reverse links to be known. The authors acknowledge this limitation, utilising Google for this task. It is unlikely that such data would be readily available for seeding in many criminal investigations - particularly CEM matters! The context graph approach represents an enhancement of Chakrabarti et al. (1999)’s *distiller*, rather than a generational leap in approach.

Similarly, the *apprentice* approach taken by Chakrabarti et al. (2002) is designed solely for the WWW landscape, heavily utilising DOM features as a source of information. Such information is absent within our targeted landscape, but the authors do argue that other approaches could be valid within the apprentice model.

An alternate approach to the distiller enhancement (Chakrabarti et al., 1999) may be feasible, but it still needs to be remembered that the number of links emanating from

indexable documents within the WWW far outweighs the relative number of equivalent links within file system search. *One to many* and *many to many* relationships are common between nodes within the WWW, whilst file systems typically only see *one to many* - the relationship between parent (typically a directory) and child.

2.3.4 Crawling a Dark Web

Confusingly, terms such as ‘dark web’, ‘deep web’, ‘invisible web’ and ‘hidden web’ are often used interchangeably to denote a broad spectrum of mostly exclusive concepts, with changing definitions including:

- websites and services not indexed or made available via search engines (Guitton, 2013);
- dynamically generated materials inaccessible via search engines due to the need for user input, rather than any desire for covertness or privacy (Florescu et al., 1998; Schadd et al., 2012);
- unseemly or nefarious content such as that created by extremist/hate groups (Abbasi and Chen, 2007; Yang et al., 2009; Li et al., 2013); and
- networks providing anonymity for content users and content providers (Iliou et al., 2016).

Such definitions are overly broad, potentially encompassing a significant proportion of the WWW. The proposal (Fielding, 1994) and later widespread adoption of a voluntary standard for restricting crawlers’ access to websites (often referred to as `robots.txt`) effectively made large amounts of the WWW a ‘dark web’ according to the definition used by Guitton (2013). Similarly, the use of ‘captchas’ to detect and block automated crawlers and bots renders a large degree of the WWW ‘dark’ according to the definition posited by Florescu et al. (1998); Schadd et al. (2012). Defining a ‘dark’ web according to the nature of content as per Abbasi and Chen (2007); Yang et al. (2009); Li et al. (2013) ignores accessibility and accountability, with the added disadvantage of introducing subjectivity.

We refer to ‘dark web’ as referring to networks providing anonymity, though with further clarification. Following Guitton (2013), we regard anonymity to be the *non-coordinatability of traits* (Wallace, 1999). Non-coordinatability does not exist on the WWW, where any party involved in or capable of observing a particular communication can immediately glean each party’s IP address, or at least that of an upstream provider²⁶. As alluded to by Guitton (2013), any sufficiently legally or technically empowered party could exploit such information to further efforts at identifying the actual user(s). We regard dark webs to be networks which *at a technical level* do not rely upon elements capable of supporting coordinatability of traits - users can *choose* to make themselves identifiable, even inadvertently, but the network itself does not require them to do so as part of normal operation.

²⁶For example, a Virtual Private Network (VPN) host

The scope of research into the ‘dark’ or ‘deep’ webs within this thesis is restricted to The Onion Router (Tor), an anonymising network readily accessible to users via an open-source software download²⁷. Tor utilises a protocol employing circuits of relays for the purposes of ensuring user anonymity, and can be used as a ‘one way’ anonymiser, concealing client access to a known internet node (such as a WWW site). Tor also supports ‘hidden sites’, being websites and/or associated services hosted within its router network, thereby providing two-way anonymity - the ultimate ‘anonymous’ network.

The basic, ‘client’ software package includes an integrated web browser, providing a user experience not dissimilar to established software such as Firefox and Google Chrome, effectively making the network’s use near-identical to everyday web browsing. This simplicity has arguably contributed to its rise in usage, with well-publicised services such as the *Silk Road* marketplace effectively providing a pull factor for users and commerce perhaps not traditionally associated with the internet.

The Silk Road

The “Silk Road” was an online marketplace for the sale of any goods and services *excepting* those involving violence, accessible exclusively via Tor. It gained infamy largely around the sale of narcotics and false identification documents, achieving scale enabling vendor and buyer feedback akin to legitimate sites, effectively providing reputation assurance and earning it monikers such as the “Ebay for drugs” (Barratt, 2012). The enterprise was shut down by the FBI in 2013, with the founder, Ross Ulbricht (aka “Dread Pirate Roberts”), arrested and subsequently convicted on numerous charges including narcotics trafficking and money laundering. An indication of the operation’s scale is given by the USD\$183,961,921 forfeiture order issued against Ulbricht during his trial, the judge finding the amount “no more significant than the revenue that was generated” by the site’s operations (*United States of America vs Ross William Ulbricht*, 2015).

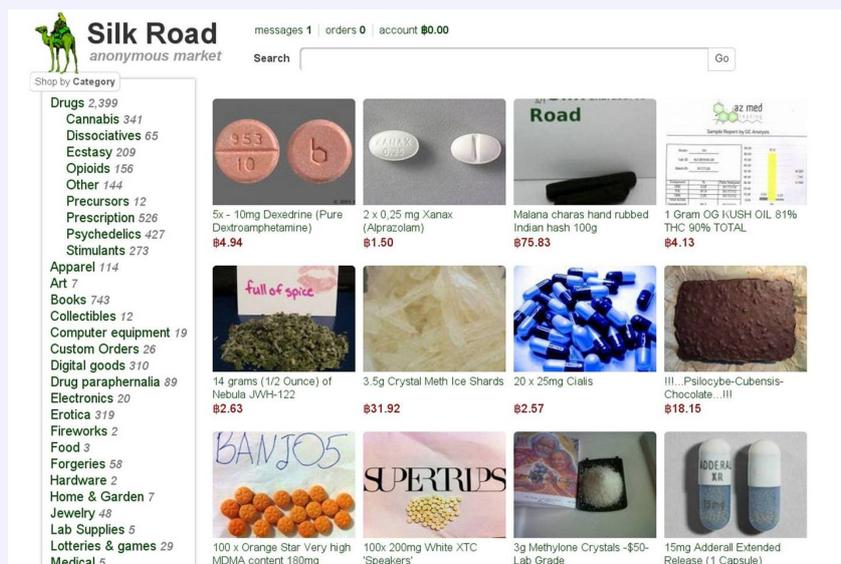


Figure 2.4: Sample screen shot of Silk Road marketplace (Farivar, 2013).

²⁷Refer <https://www.torproject.org/>

The surveying of dark webs and subsequent development of taxonomies has occurred, from different perspectives. Guitton (2013) conducted a crawl of 1171 Tor hidden services, building a 23 category schema split broadly into *ethical* and *unethical* services - finding unethical services to be so pervasive that “...*further development of Tor hidden services should hence stop*”.

Moore and Rid (2016) also conducted a crawl of Tor, creating a twelve category taxonomy. As with the other crawlers, the authors’ aim was to provide a ‘snapshot’ of Tor and the services offered therein, without any specific interests or use cases. Of note, the authors specifically restricted their crawler to textual content, due to the high risk of inadvertently accessing illegal materials such as child pornography and terrorist publications - a common issue within this field.

With respect to law enforcement applications, research into the dark web tends to focus upon specific content types. Chen (2012) establish a schema for categories of use for the web (also applied to ‘dark webs’), focused upon terrorist organisations. Westlake et al. (2017) create a web crawler specifically focused upon locating CEM. Guided by seed websites plus a combination of keywords for identifying CEM plus ‘safe sites’ for domains previously identified as not being of interest, the crawler was successful in following the topic, though limited to three ‘categories’ based upon Canadian legal definitions of CEM. Whereas both studies ultimately relied upon manual labelling (with some automated features for steering and/or identifying content of interest), their focus upon particular content limits their use in identifying wider taxonomies.

Table 2.4 summarises and contrasts the various taxonomies developed in the aforementioned papers, ranging from particularly holistic (Biryukov et al., 2014) to targeted (Chen, 2012).

Automated classification of materials of interest is obviously of value in identifying illicit materials online, though as with previously detailed work, such work tends to be focused on narrow topics. For example, Fu et al. (2010) develop a crawler focused upon extremist discussion fora on the ‘dark web’²⁸, with a need to access and analyse multimedia resulting in a mixed approach of URL tokens used to identify links associated with discussion fora and file hosting, plus page *levels* used to ensure collection of multimedia files potentially housed behind several links (typically encountered with third party file hosting services). Once identified, URLs are prioritised according to either best first or depth first search, depending upon the nature of the relevant forum. Sabbah et al. (2016) utilised the data generated by Fu et al. (2010) as a test-bed for further statistical analysis as a means for document classification, though with the rather basic categorisation of ‘dark’ for terrorist activities such as weapons/explosives manufacture. It is unclear where less clearly ‘illegal’ discussions such as recruitment and incitement to extremism lie within this definition. Contrastingly, Scanlon and Gerber (2014) focus specifically on detecting online recruitment by violent extremists, with a correspondingly simple annotation schema - ‘recruitment’/‘not recruitment’.

²⁸In this case, content requiring input in order to be accessible

Moore and Rid (2016)	Guitton (2013)	
Arms Drugs Extremism Finance Hacking Illegitimate Pornography Nexus Other illicit Social Violence Other None	<i>Unethical Services</i> Hacking Black Market Pornography (excl. child) Drugs Gen. forum with unethical topics Hit man Weapons Racial Discrimination	<i>Ethical Services</i> Personal File sharing Informatics Bitcoin Everything Search Engine Subversion of state power Surveillance Politics Anarchism Energy politics Communism Ethical & specific topic (other)
Biryukov et al. (2014)	Chen (2012) (Terrorist use)	
Adult Drugs Politics Counterfeit Weapons FAQs, Tutorials Security Anonymity Hacking Software/Hacking Art Services Games Science Digital libs Sports Technology Other	Communications Fundraising Sharing Ideology Propaganda (insiders) Propaganda (outsiders) Virtual Community	

Table 2.4: A comparison of identified Tor hidden site topics or uses

It is readily apparent that existing schemas and ontologies reflect the motivations surrounding their creation. Chen (2012) shows an extremely efficient approach to recording a specific area of concern, reducing terrorist use of communications fora to six distinct categories. Unfortunately, such a tight focus would probably best be described as ‘myopic’ if used within law enforcement, due to the complete absence of other crime types or areas of concern - in fact, the absence of *any* ‘other’ type category.

The demarcation of ‘ethical’ and ‘unethical’ topics by Guitton (2013) takes an approach more relevant to law enforcement, but realistically, the collection, labelling and classification of ‘ethical’ materials by law enforcement is unlikely, if not for efficiency purposes,

then at least due to their own ethical limitations - large scale monitoring of ‘not of interest’/‘legal’ materials running dangerously close to mass surveillance. A more immediate issue, however, is the relatively fluid definition of ‘ethical’. In this instance, ‘Anarchism’ and ‘Subversion of state power’ are listed as ‘ethical’ activities. It is difficult to imagine either topic in and of itself to be regarded as ‘unethical’ by a reasonable person - realistically, the *severity* of the writings and proposed actions/incitements (if applicable) should most heavily influence such opinion.

Research into user networks within the dark web continues a focus upon extremist activities. Xu et al. (2006) examined the online topologies of terrorist groups. Identification of groups and their interactions is a vital part of intelligence gathering, but the authors’ heavy reliance upon manual identification and cleaning of URLs throughout the crawl process limits its application within our work.

Where to Begin?

A crawl of a network is effectively an attempt to survey, observe, and record a virtual topography - relatively simple in terms of the WWW whereby a hierarchy of name servers takes responsibility for managing a repository of registered domain names and processing queries. Updates and changes of address are administered through this process, ensuring timely (though not immediate) updates across the WWW.

Extending such a crawl down the stack to IP addresses is more complicated with large quantities (sometimes entire countries) of devices effectively concealed behind routers utilising Network Address Translation (NAT). Admittedly, NAT isn’t a technology best suited for use in service provision to external parties, but new connections *could* be forwarded via pre-agreed ports for disambiguation, in a large-scale version of a service provided in consumer-level domestic ADSL routers. This is not exactly practical, but definitely feasible.

Surveying a ‘dark’ web is far more complicated - whilst the term itself has myriad definitions, most (if not all) share one attribute - unlike the WWW’s Domain Name System (DNS) service, they don’t advertise entry points. Hidden services on Tor issue *hidden service descriptors*. A requesting user queries the Tor network with a hidden service’s URL²⁹, receiving addressing information if the hidden service exists. Being decentralised and intentionally obfuscated, seeking and accessing hidden services is by its very nature slower than DNS. Ignoring the wider performance implications of a ‘brute force’ approach to identifying active services, such an approach would also be incredibly slow - there are 32^{16} possible top level domains for Tor hidden services, meaning a single threaded crawler experiencing extremely optimistic query times of six seconds per lookup would require 2.3×10^{17} years to traverse each valid address.

We have therefore established that whilst Tor currently supports a finite address range for hidden services, simply trying valid addresses isn’t feasible. Another option is to undermine Tor’s security as a means to removing obfuscation. Biryukov et al. (2014)

²⁹Tor hidden service addresses consist of a sixteen character, base32 (a-z2-7) string, appended with **.onion** - refer <https://www.torproject.org/docs/hidden-services.html.en>

utilised an exploit available in February 2013 (since patched) to detect 39,824 hidden service addresses, their technical services offered, and their relative popularity. Wang et al. (2011) and Ling et al. (2013) freely detail other vulnerabilities, subsequently fixed. Research utilising technical exploits in Tor is relatively rare, with the reliability of any exploit or vulnerability extremely tentative at best (particularly post publication).

Undermining a service commonly used for lawful (and ethically sound) activities for the purposes of research and/or indiscriminate surveillance is unethical and quite possibly illegal in many jurisdictions. The mere publication of the existence of any weaknesses in an otherwise ‘secure’ network can lead to further attacks. We are uncomfortable with surveys, crawls and port scans undertaken through bulk or otherwise indiscriminate identification of hidden services, as utilised in previously mentioned research. We acknowledge no ‘secret’ data was necessarily accessed (let alone published), but no permission was received to access what is effectively a secret address, plus the sheer volume of network traffic undertaken as part of the process would impact the Tor network’s performance (however slightly). For these reasons, we regard the use of technical exploits as ‘off limits’ for our research.

The open-source nature of the Tor project, together with its high profile, means many such weaknesses are quickly identified, publicised, and patched. The pace of such work is perhaps best exemplified by the software itself - the project’s client software release notes (Tor Project, 2017) list 150 versions, with over 600 minor and major bugfix summaries from underlying projects. Disregarding the ethical issues previously detailed, the rapid repair of any publicised security shortcomings also makes the reproducibility of any results unlikely.

The best (if not ‘least bad’) option is therefore bootstrapping - exploiting lists of known ‘live’ sites, either from external providers or ‘rolling your own’. Table 2.5 lists the methods used in previous research in this area.

Research	Dataset Focus	Bootstrap/Seeding method
Fu et al. (2010)	Terrorist & Extremist Groups	Dark Web Portal (Chen, 2012)
Guitton (2013)	Tor Hidden Service Sites	Aggregate list of services taken from Tor search engine/databases: Hidden Wiki, Snapp BBS, and Ahmia.fi
Yang et al. (2010)	Externally sourced dataset	Ansar1 English language based forum.
Al-Rowaily et al. (2015)	Cyber Security & extremism	Dark Web Portal (Chen, 2012)
Biryukov et al. (2013)	Tor Hidden Service Sites	Port scanning and service identification, via an identified flaw/vulnerability in the Tor protocol
Chen (2012)	Terrorist & Extremist Groups	Manual identification of URLs associated with previously identified, extremist groups and associated entities. Includes use of search engines, government reports and research centers.
Anwar and Abulaish (2012)	Extremist Group	Forums hosted by manually identified Neo-Nazi website (http://www.stormfront.org)
Xu et al. (2006)	Terrorist & Extremist groups	URLs from US State Department and FBI reports relating to manually identified organisations of interest.
L’huillier et al. (2010)	Terrorist & Extremist fora (English language)	Subset of Dark Web Portal (Chen, 2012)
Chen et al. (2008)	Terrorist & ”Jihadist” groups	Seeded URL list, backlink searches via search engines

Table 2.5: Dark Web crawler seeding methods

For reasons including the above, this thesis’ interest in anonymous networks is limited to the efficient mapping of online behaviour from a law enforcement perspective, with a view to better informing future data analytics within policing.

2.4 Summarising Existing Research

Garfinkel (2010) characterises 1999-2007 as a digital forensics ‘golden age’, largely due to consistency in the infrastructure and behaviours encountered (for example Windows XP as OS, ‘relatively few’ file formats of interest). However, at time of writing he predicted a looming crisis in investigative matters, due to (amongst others):

- **Growing storage device capacities** leading to the inability to completely image a device or completely analyse data if acquired;
- **Multiple storage devices** being involved in matters, requiring increased data correlation and examination;

- **Cloud storage and processing** effectively splitting single devices into multiple structures;
- **OS and file format proliferation**, leading to additional complexities in search exploitation.

Garfinkel's paper was published eight years ago, but the identified challenges largely remain contemporary. Growth in both the number and size of electronic storage media encountered during investigations simply reflects pervasive computing, and is most strongly felt during digital forensic triage, when infrastructure and time both tend to be in short supply. Approaches such as similarity digests can be effective in establishing links, but simple pattern matching still requires extensive reading and interrogation of data stored within a medium.

Research into data reduction typically espouses the value of limiting data *analysis* to files/areas known not to be of interest to an investigation, but does extend to reducing *acquisition*. This approach ostensibly makes sense when preservation is not in question. Indeed, the 'sniper forensics' proposal mentioned earlier uses the example of an investigation into unauthorised system access, presumably on behalf of the system's owners. A central tenet of science and the judicial system alike is the requirement to look for evidence contradicting the asserted theory or allegation. A risk associated with data reduction is the overlooking of *exculpatory* data. Unlike the example of an investigation into unauthorised system access, the assertion that child pornography possession investigations need only focus upon the search for relevant imagery (Ferraro and Russell, 2004) may be true in terms of establishing relevant proofs, but perhaps doesn't take sufficient account of external influences. As discussed in this chapter, courts in the State of Victoria (Australia) often require full forensic reports even in uncontested matters. Comprehensive forensic reports can't be generated if analysis is reduced to a simple investigative review. Debate over the need for full analysis across all digital investigations is valid, but external considerations such as those imposed by the relevant judiciary are outside the scope of this work.

Prior research is focused largely on the technical and practical challenges faced by DF. An issue largely under-appreciated until the last few years is that of exposure to offensive materials - the aggravating factors of which lead to significant challenges within the field.

2.5 Challenges in Digital Forensics

The move towards pervasiveness of computers across most (if not all) crime types is not a recent event - Ferraro and Russell (2004) specifically predicted this pattern fourteen years ago, warning that if forensic examination is expected for every crime, "Every law enforcement agency will either have a laboratory of its own or rely upon a computer forensics laboratory to process its evidence". This prediction has largely come to pass, with every law enforcement agency (and many government agencies with investigative functions) now maintaining in-house DF capabilities or contracted third party analytics.

Powell et al. (2014) conducted a survey of 32 law enforcement personnel across all Australian jurisdictions, querying their opinions on issues faced during investigations into online child exploitation. DF challenges identified included:

- Limited access to “image scanning” software - most likely a reference to CETS (refer Table B.1) or another cryptographic digest based content recognition system (refer Section 2.6);
- Inadequate staffing, including a lack of relevant digital forensics experience; and
- The need for “complete” examination - courts requiring every relevant item (image/video) to be reviewed and categorised, rather than accepting a representative sample. A respondent quotes a staff member “going through 500,000 images”.

More recently, Franqueira et al. (2017) conducted a targeted survey of DF practitioners worldwide, seeking their comments on challenges in the field of online child exploitation. The survey returned similar results in regard to the stresses and impacts of exposure to such imagery, but the authors’ stronger focus on technical specialists³⁰ resulted in a differing set of reported challenges:

- Emerging technologies such as automatic age estimation are not ‘translating’ into workable tools for improving practices;
- Stressful working conditions associated with viewing CEM, with recommendations for improving automation to “minimize exposure in the first place”; and
- A need to standardise operations, procedures and legal frameworks globally, necessitating an *“internationally recognised scale of indecency levels and a taxonomy of terms to bridge language and cultural differences”*

The absence of standardisation as a challenge is notable in Powell et al. (2014). Nine jurisdictions are included (6 States, 2 Territories, plus Federal), each with some degree of individual case law and procedures, but de-facto standardisation has occurred - both through the establishment of Joint Anti Child Exploitation Teams (JACETs) in each State/Territory, and the alignment of State legislation and availability of Federal legislation to State Police. Whilst not in blanket use across all prosecutions, the Child Exploitation Tracking System (CETS) scale (discussed further in Section 2.6) is used across Australia as a standardised measure of offending, greatly simplifying joint investigations and cross-jurisdictional prosecutions.

³⁰The authors use ‘DF’ in a broad sense, encompassing first responders, consultants and other roles regularly exposed to such materials

A Loose Quantification of Sentencing

$sentence = (\sum_{c \in categories} n_c \times \beta_c) \times ((1 + \alpha) - \omega) \times \tau$, where n denotes the quantity of materials identified for a category and β the weighting for the category. α and ω denote weightings with values $[0, 1]$ for aggravating and mitigating circumstances. All weightings and circumstances are exclusively the preserve of sentencing officer (Magistrate, Judge etc), but are informed by precedents set by equivalent matters in like jurisdictions (denoted by τ).

Whereas standardisation is undoubtedly a benefit for knowledge sharing within law enforcement, the existence of a means to compare offending has had the unintended consequence of encouraging the requesting by courts of ‘complete’ examinations within online child exploitation prosecutions, particularly as an input for sentencing (detailed in *A Loose Quantification of Sentencing*). This in turn has increased workload related stressors, further aggravated through the dangers of repeat exposure to CEM.

2.5.1 Dangers of CEM Exposure

First-hand exposure to traumatic and offensive events is long documented as psychologically harmful. Surveys of police officers in provincial England (Brown et al., 1999) and New York State (USA) (Violanti and Aron, 1995) indicated comparatively high levels of stress associated with exposure to traumatic events involving children. Both studies pre-date the mainstream emergence of online child sex abuse, but a key point of note appears to be stress associated with dealing with *victims* of crimes such as rape and child abuse being quite high, with police officers seen as potentially “becoming secondary victims” (Brown et al., 1999) in such cases.

The absence of studies into the effects of exposure to child exploitation on forensic analysts and other persons involved in the investigation/prosecution process is noted by Edelman (2010), who observe that employers such as the Metropolitan Police provide mandatory counselling to staff routinely exposed to such imagery.

In Powell et al. (2015), the authors of the aforementioned Powell et al. (2014) also specifically recorded the surveyed participants’ reported impacts of exposure to CEM³¹ within internet child exploitation investigations. Critically, the survey included not only sworn police, but also “computer analysts” - a role arguably requiring even more regular and in-depth exposure to materials during the course of normal duties. Interestingly, some respondents indicated an experience akin to the previously mentioned ‘secondary victimhood’, though contrastingly, some perceived exposure to CEM as less harmful than direct ‘interaction with victims of assault’³².

Specific factors were listed by survey respondents as increasing a risk of long-term effects from exposure:

- Perceived resemblances between victim(s) and child(ren) known to the reviewer (particularly the reviewer’s own children);

³¹Referred to as “internet child exploitation” materials within the paper

³²It is unclear if this refers to *sexual* or physical assault, given the context

- ‘Unexpected’ viewing of child exploitation materials;
- Repeat exposure to specific images or offenders; and
- Viewing the progression of an offender from viewer to contact offender³³.

An anonymous survey of US law enforcement personnel by Seigfried-Spellar (2017) identified differences in psychological distress between investigators and forensic analysts, with persons conducting both duties in CEM related cases reporting higher levels of traumatic stress than those working single roles. The author hypothesizes this is due to their requirement to both review CEM and interact with victims and offenders, a theory consistent with the “secondary victimhood” identified by Brown et al. (1999). Furthermore, whilst respondents *generally* used healthy coping strategies, those working dual roles “may be more likely to use sedatives . . . as a coping mechanism”.

Given the nature of the survey, it is difficult to quantify the level of ‘overlap’ encountered by respondents, as this often will be set by their respective organisations. For example, whilst DF practitioners within the AFP attend search warrants and interact to a limited extent with offenders, victims etc, the vast bulk of CEM review is conducted by investigators. As investigators are also responsible for offender and victim interactions, they effectively perform the more stressful ‘dual’ role.

Powell et al. (2015) note that due to the large number of variables involved, individual investigators’ reactions to CEM exposure are impossible to predict - interestingly, they also note reports of increased distress due to *text* content such as filenames, an aspect traditionally viewed as low/no risk by law enforcement.

Given the general reluctance by police to seek assistance, combined with a low (16%) level of mandatory counselling offered by the respondents’ agencies (Seigfried-Spellar, 2017), it appears quite feasible that the extents of exposure **and** related stress & harm are both underreported across law enforcement.

As stated by Powell et al. (2015), “purchase of technological strategies for global reduction in exposure to images is therefore warranted”.

2.6 Towards Automation of Digital Forensics - A CEM Case Study

Automated CEM detection is best divided into two challenges - discovering known items of interest (i.e. images previously observed), and discovering previously unknown items of interest.

In terms of known materials, previously observed and annotated files tend to be ‘recorded’ in the form of cryptographic digests, due largely to their inherent confidentiality, broad acceptance across industry and government, and their relative efficiency in terms of storage and processing. Fuzzy hashing, whilst computationally more expensive, can identify slightly altered or otherwise highly similar data. Perceptual hashing, a multimedia-focused extension of fuzzy hashing, has the most relevance and displays the

³³The *abuser*, as opposed to viewer of abuse.

most promise for offensive multimedia. The exploitation of file metadata is lightweight and fast, but can only provide hints as to the likely nature of content. Unfortunately, both fuzzy hashing and file metadata exploitation can only provide automated detection of materials highly similar to known items - a sort of ‘semi unknown’ detection at best.

In regards to unknown materials, the only accepted, readily available method for automatically recognising pornography and CEM is skin tone detection, of use only in pornography related investigations. This approach has numerous limitations (listed below), and for this reason, appears largely used as a triage tool of last resort within criminal investigations.

2.6.1 Metadata Based CEM Detection

Metadata, or ‘data about data’, is a lightweight means for identifying *likely* materials of interest, but with restrictions due to the fact that metadata provides descriptors rather than content itself. A rough but simple approach during preview/triage is to filter candidate files by type, according to the nature of data being sought. For example, device examinations during multimedia focused matters such as CEM investigations can often start with a simple filter for files with extensions such as *jpg, jpeg, gif or png*. Non user-generated files such as thumbnails and/or caches can be ignored or downgraded through the use of a further filter or descending sort based on file size.

Such an approach is effective for separating disparate files, but not when filtering like data formats. Offenders don’t necessarily *only* collect illegal materials - for example, adult pornography can be colocated with CEM (Powell et al., 2015; Franqueira et al., 2017). Intricately organised and managed libraries of electronic paedophilia are not unheard of, but the more usual experience is a highly unstructured dump of disparate media, often within the default download directory for the P2P/torrent application *du jour*.

A dominant academic focus for automated CEM recognition is filename/textual analysis of likely content, particularly in the context of P2P networks - architectures such as LimeWire and BitTorrent (BitTorrent inc, n.d.) allow the collection of metadata *without* downloading actual content, enabling researchers to stop short of breaching local laws and ethical boundaries. Steel (2009) provides a snapshot of the Gnutella network, using tokenised query responses to categorise files as likely child pornography. Whilst most terms are sanitised for ethical reasons, the author provides some indications of common ages and advertised features/actions. Panchenko et al. (2012) identified textual features from filenames of *known* CEM files (provided by law enforcement), providing a level of confidence unachievable from query-based studies. Latapy et al. (2013) also observed specialised vocabulary exclusive to online paedophile activity, a finding supported by Peersman et al. (2016).

Based on first-hand experience, we can confirm the presence of such ‘red flag’ terms, though their presence seems to be highly correlated with distribution via P2P networks - we hypothesise this is due to uploaders ‘advertising’ the files to make them more attractive for download, possibly in order to maintain required upload/download quotas in such systems.

Filenames are an obvious next step for further prioritising or filtering results, particularly the long, descriptive file names dominant within sharing services such as BitTorrent. CEMs often display domain specific vocabulary, with terms such as *Bibcam*, *PTSC* and *PTHC* giving a strong indication of content³⁴ - a hypothesis supported by Latapy et al. (2013); Peersman et al. (2016); Steel (2009). The identification of textual features from filenames of *known* CEM files (provided by law enforcement) provides a further level of confidence (Panchenko et al., 2012) from what amounts to an understandably limited access to source data (none of the aforementioned studies actually accessed or downloaded any CEM, limiting their involvement to searching for files of likely interest).

File extensions are another option, whereby a ranking algorithm identifies likely multimedia files for examination (e.g. MP4, AVI). File extensions remain a convention rather than rule within computing, with file headers (typically an n -long series of characters at the start of a file's content) actually used by software for parsing. Therefore, one can't assume the presence of a specific extension actually reflects a file's content.

File size can be a reliable indicator of multimedia files, particularly high definition movies. However, it doesn't reflect the nature of content, leading to extremely poor results where a device contains large quantities of files from disparate genres - a common scenerio identified by Franqueira et al. (2017).

The creation of a reliable classifier utilising the aforementioned features is extremely challenging, and presuming one could be made to work reliably, a prioritiser based on such is limited to identifying likely *files* of interest (and perhaps directories with suspect names). This is perfectly acceptable in 'first past the post' situations, where only the *presence* of such data is being sought, but one needs to consider the bigger picture - identifying a number of disparate files of interest may imply guilt, but doesn't provide investigators with an overall impression of the nature or scale of offending. Presuming a suspect has been (or will be) arrested as a result of finding the data of interest, the investigation may be subject to time limits prior to charges being laid and the suspect being delivered to a court or released on summons. A taped record of interview (or equivalent) will typically be conducted during this time, during which questions will be asked of the suspect, and the suspect given an opportunity to put forward their version of events. Typically, this is the *only* opportunity for such direct interaction, making the timely provision of accurate data (particularly relating to proofs and potential defences) critical for investigators.

Exploitation of metadata is lightweight and fast, but shouldn't be entirely relied upon for forensic tasks, for multiple reasons:

1. Metadata is machine readable, and therefore also easily editable. A simple obfuscation technique such as changing file extensions to unrelated type (for example, *jpg* to *zip*) will overcome the filter/prioritisation approach previously described. In the absence of an obviously incongruous extension (e.g. `txt` for a 1+GB file), examination of at least part of the candidate file's content is required to establish actual file type. Most file types contain a unique header (aka *magic number*) as a means to

³⁴All CEM related filenames listed here have been published prior to our research. A number of other such terms exist and are common knowledge within the field, but have not been listed for ethical reasons

Trialling Filename Based Detection

As a means of testing the approach and findings of Panchenko et al. (2012), we constructed a corpus of filenames associated with (a) general files, (b) adult pornography (sourced from The Pirate Bay download lists) and (c) CEM. These filenames were then processed using the Apache Solr^a default English language indexing analyser, which in turn was used to train a multinomial naive Bayes classifier. Encouraging signs were observed in the obviously distinct term distributions amongst the three classes, as shown by the histogram within figure 2.5.

The classifier initially performed encouragingly, with 0.9+ precision and recall levels observed with data limited to individual cases. Contrastingly, such performance dropped off rapidly once the classifier was trained and tested across unrelated cases. Panchenko et al. (2012) anticipate drops in performance when disambiguating adult pornography and CEM due to overlapping vocabulary, but we suspect poor performance is due more to CEM being a diverse field rather than a distinct genre. From personal investigative experience, we can state that seemingly specialised fields such as online child exploitation can actually be quite broad, with individual users' tastes and predilections heavily influencing the volume, nature and format of materials being traded, generated and stored. This is compounded by domain-specific influences during the sharing phase - for example, bitTorrent traded files' names tend to be quite descriptive, given the propensity for the names themselves to be the only advertising 'feature' seen by end users when searching for material.

Filenames therefore tend to have specialised vocabularies. Common terms and phrases are present across classes and genres (for example *fucks*), but these tend to also be seen within lawful and otherwise uninteresting data, raising the real possibility of high rates of false positive (inconvenient) and false negative (disastrous) rates within investigations. We have no doubt that a designer with access to sufficient data will at some point in the future be able to build a more successful classifier, once sufficient volumes of data *from disparate sources* are aggregated.

^a<https://lucene.apache.org/solr/>

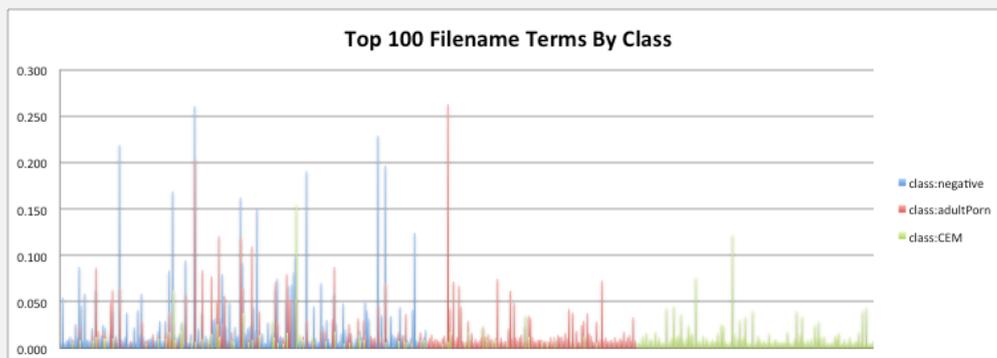


Figure 2.5: Histogram - Top 100 index terms for (a) General Files, (b) Adult Pornography, and (c) CEM

inform rendering applications of their format, so a prioritiser may only need examine data of the order of dozens of bytes per file.

2. File size is indicative by nature, but does not constitute a guarantee about a file's provenance or nature. Compression and ever-increasing definition ("Megapixel") resolutions mean the size range of user-generated content is expanding.
3. Descriptive filenames tend to be added to files as they are shared, particularly when distributed via torrents. Therefore, this method tends to detect materials the further they are from the original source, potentially missing detection of contact offenders' generated materials (i.e. CEM still bearing the camera's default naming convention of `IMG_XXXX.jpg`);
4. Files can be incorrectly named and/or labelled. We have observed this to be a common issue in file sharing services and protocols such as eMule(eMule Project, n.d.), BitTorrent and the now discontinued Limewire. Whilst this could be due to user error, one suspects that elements of advertising and inticement to download (both true and false) play a greater role in this behaviour, particularly in networks maintaining minimum user sharing rates; and
5. Individual offenders tend to have specific tastes. Of the sample cases used for Figure 2.5, one offender was particularly prolific with several hundred thousand descriptively named images. One prototype filename-based classifier we trained using this sample data showed an extremely strong bias for pubescent boys, completely ignoring file names denoting female involvement. This is a classic symptom of a biased training corpus, but the fact remains that the remaining non-specific textual features had more in common with innocent/general files than other CEM files.

Ultimately, a file's metadata can only represent an author's claims, and should be regarded as an indicative rather than objective or complete representation of rendered content.

Beyond legal considerations, it is likely at least some of the aforementioned research was able to be conducted because researchers and their host organisations viewed the avoidance of direct exposure to materials as akin to avoidance of risk, a fallacy already discussed in Section 2.5.1. Further work even at low exposure levels should only be conducted with the mandatory involvement of psychological screening, monitoring and counselling.

2.6.2 Content Based CEM Detection

Content analytics is a more intensive, but arguably also more robust approach. Currently, the dominant method for automatically detecting known materials is via cryptographic hash (e.g. MD5, SHA-1) comparison, recognising identical materials at the *binary* level. Specialist algorithms such as PhotoDNA(Ith, 2015) can measure similarity, allowing the automated recognition of resized, skewed or otherwise slightly altered still images.

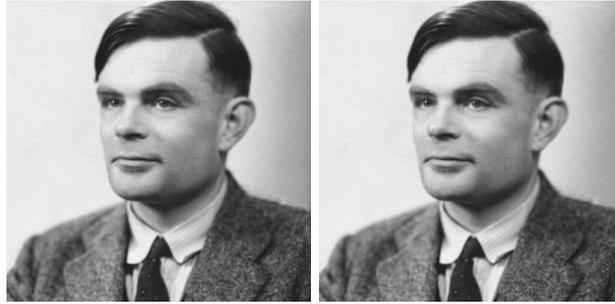


Figure 2.6: Original and altered images

Image	MD5	SHA-1
Left	5ca29e467c30aa2ade4195d53b011163	dde22a84c894ed68a6dd9312b82050e65205c3f5
Right	92cd72ae43ba10358a2aa4053e1fd4fe	96e55bc2498d48bb0610b07201a9b7fd35a070a8
Hamming Distance	25 (maximum distance: 32)	34 (maximum distance: 40)

Table 2.6: Cryptographic Hashes for Figure 2.6

2.6.3 Cryptographic Digests

Currently, law enforcement capture and distribution of known materials (both of interest and ignorable) is largely reliant upon cryptographic digests (commonly referred to as ‘hashes’) – traditionally MD5, with the newer SHA-1 algorithm used to a lesser (but growing) extent. Both are widely accepted methods for reliably recognising identical data, being designed as means for easily detecting unauthorised alterations, and work at the *binary* level. In simple terms, cryptographic digests look at binary data, and do not make any attempt to measure or understand the perceived/rendered output. To take the example of photographic images, Figure 2.6 shows an original and an (arguably imperceptibly) altered image, with a slight change in exposure levels applied. As anticipated, their respective cryptographic hashes (refer Table 2.6) detect their difference, giving distinctly different outputs. Whilst both algorithms’ outputs see Hamming Distances exceeding $\frac{2}{3}$ of their lengths, it must be remembered that *every* Hamming Distance greater than zero is equal in this context, simply denoting the data as ‘different’.

Cryptographic digests are designed specifically to *not* allow similarity measurement, both in order to preserve the confidentiality of materials being ‘hashed’, but also to minimise the value of brute-force efforts to force ‘collisions’ - different data returning identical digests.

2.6.4 Fuzzy Hashing

‘Fuzzy’ hashing is something of a similar technology (one-way compression of data to a reproducible ‘fingerprint’) to cryptographic digests, but for what amounts to a different application - in this case, measuring *similarity* between data. Several algorithms exist for measuring similarities at a binary (or flexible) level, for example ssdeep (Kornblum et al., 2018), Trend Micro Locality Sensitive Hash (Micro, 2018) and SDHash (Roussev, 2010). Effective for detecting lightly altered files such as binaries and text documents,

fuzzy hashing is susceptible to technologies opaque and/or of little relevance to the end user, such as compression. For example, the similarity of the images in Figure 2.6 at a binary level would differ widely if they were saved using differing file formats (e.g. jpg vs png).

Perceptual Hashing

Perceptual hashing applies fuzzy hashing to *rendered* content - in other words, measuring similarities between materials as they appear to the viewer. By way of example, pHash (Klinger, 2010), an open source perceptual hash algorithm, measures the similarity of the images in Figure 2.6 as 0.999992 (Using Peak of Cross Correlation), with 0.85 seen as the threshold for ‘identical’ (in terms of perception) images (Klinger, 2018).

PhotoDNA (Microsoft, 2015) has received strong support in law enforcement, having been acquired and distributed by Microsoft primarily as a CEM detection technology. Whilst ‘free’ for law enforcement use in the financial sense, PhotoDNA’s usage remains subject to licensing restrictions by Microsoft, limiting its use outside law enforcement. A cloud-based service is advertised for use in detecting CEM by the general public, but access is subject to user vetting³⁵, and requires users to upload images rather than hashes - a model reportedly adopted by Facebook as a means for detecting *revenge porn*³⁶ (Gribbin, 2017). Both services require user trust that the uploaded data is not going to be misused, but the transmission of CEM, even in good faith, potentially places users at risk of breaching laws in some jurisdictions.

A freely available algorithm nicknamed dHash (Krawetz, 2013) was utilised in our early forensic crawl acceleration experiments, and observed to be surprisingly fast and effective, particularly in light of the algorithm’s simplicity and reproducibility. Its use was dropped in favour of PhotoDNA for compatibility with existing datasets. It is difficult to quantify performance against other algorithms, as no experiments were conducted measuring the impact of ‘skewing’, rotating or otherwise altering imagery in a manner beyond cropping or changing resolution. Furthermore, no research was conducted into the rate of false positives.

Limitations of Fuzzy Hashing

It is unquestionable that fuzzy hashing, and in particular perceptual hashing, can be of great value in automatically identifying previously observed and annotated imagery of interest. Beyond licensing restrictions and efficiency of the underlying algorithms themselves, the principal challenge around fuzzy hashing in a law enforcement context is technical - in particular, dimensionality. To simplify, cryptographic hashes are one dimensional, in that if any character differs between two hashes, they are as distant as if all characters differ. A Hamming Distance value of 1 (one character difference) is therefore as dissimilar as a

³⁵Presumably to avoid counter-intelligence type efforts by offenders to measure law enforcement knowledge of certain materials

³⁶The release of explicit media without a depicted person’s permission, popularly associated with jilted/scorned ex-partners seeking comeuppance after termination of a relationship

value of the entire hash length. Comparison of cryptographic hashes can therefore cease at the point of encountering the first differing character.

To take the example of an MD5 hash, of the 32^{16} available combinations, any particular file has $\frac{1}{16}$ probability of matching the first character of a previously known hash/digest. Therefore when searching through a dataset, about $(1 - \frac{1}{16}) \cong 94\%$ of known hashes can be disregarded at the first character, and $(1 - \frac{1}{16}^2) \cong 99.7\%$ at the second. 99.9999% of candidate hashes can be disregarded at the fifth character.

DHash (the fuzzy hash previously mentioned) can be stored as 64 bits (coincidentally half the length of an MD5 hash), represented on paper as a sixteen character hexadecimal hash. Comparisons are made on a per-bit level, with users reporting a Hamming distance less than 10 typically denoting a ‘match’ between images. Assuming a pseudorandom distribution of bits³⁷, each bit has a 50% probability of increasing the Hamming Distance, meaning a hash comparison has a 50% chance of reaching the 20th bit before crossing the threshold distance of 10 and being disregarded. This is far more computationally complicated than a straight test for completely identical values, and also requires memory for maintaining Hamming Distance (minimal, but with potential impacts on parallel processing). Furthermore, an image-based similarity measure will also by its nature require rendering by the chosen algorithm, further increasing cost of computation.

It should be noted that methods for accelerating queries around fuzzy hashes exist, typically involving indexing subsets of hashes. Such an approach trades memory for speed, requiring such indices to be available within memory at query time and therefore again requiring additional computational resources.

Whilst fuzzy hashing is more robust to changed data than cryptographic digests, both methods are restricted to recognising previously observed and annotated materials, restricting their value to ‘downstream’ in the sharing process - older imagery/movies, most likely shared numerous times between production and detection. An obvious help by accelerating analysis, this does little to aid law enforcement in targeting producers and victims.

2.6.5 Skin Tone Analysis

A simple method for automatically detecting previously unseen pornographic content is the measurement of skin tone occurrence as a proportion of the image. Strictly speaking this is a nudity detector (we will discuss definitions of pornography in section 2.6.7), with a small number of ‘value added’ options such as disambiguation of adults and children (Islam et al., 2011) also being proposed. We posit that analysis based upon skin tone is inherently unreliable, due to a combination of the broad spectrum and non-exclusivity of possible skin colours, combined with the wide range of possible lighting situations available within legible imagery.

Figure 2.7 shows three images, together with corresponding skin tone masks calculated using the ‘Uniform Daylight Illumination’ algorithm by Kovac et al. (2003). The first

³⁷Problematic, given the pattern actually reflects what is seen in the image, meaning ‘common’ image types such as passport photographs will tend to have some correlation

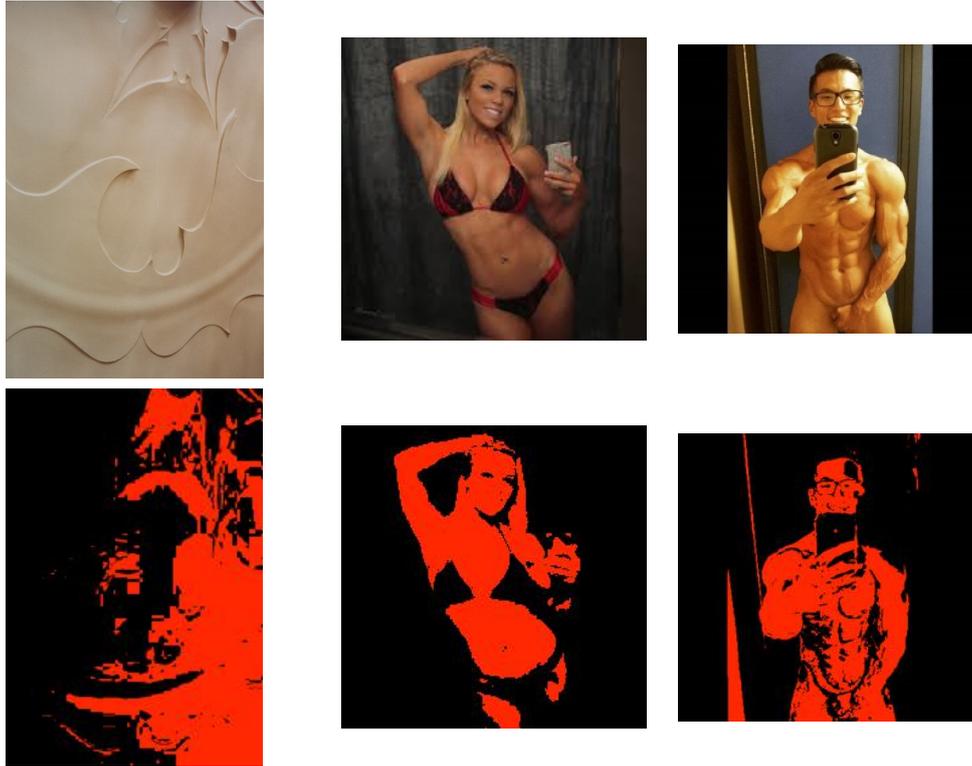


Figure 2.7: Ceiling skin tone >Bikini skin tone >Nude male skin tone

Top row: Ceiling(Shapiro, n.d.) (34% skin tone), Female bikini ‘selfie’(Drain, 2015) (24%), Male nude ‘selfie’(Choi, n.d.)(23% as received, 34% after cropping).

Bottom row: ‘Uniform Daylight Illumination’ (Kovac et al., 2003) Skin tone ‘masks’ of top row images. **Red** denotes pixels displaying skin tone, **black** denotes all other colours

image, a picture of a ceiling, reports skin tone appearing in 34% of pixels, an obvious result of the cream coloured paint. The second image is of a female bikini ‘selfie’, taken in front of a dark background - this only displays 24% skin tone, almost a third less than image one. As received, the third image of a Korean male bodybuilder displays 23% skin tone, increasing to 34% after the borders are removed. An admittedly selective sample, these examples shows how irrelevant features such as background, clothing colour and lighting can dominate rather than merely affect results. Interestingly, a material proportion of exposed skin in image three appears to have been sufficiently shaded or brightened to not be recorded as such, most likely due to contours arising from the subject’s muscle tone. Skin tone analysis is unable to disambiguate an image of a nude male from that of ceiling.

Beyond theoretical false positives, reliance upon skin tone in isolation can result in a great deal of false negatives when the imagery depicts persons of non-Caucasian heritage, a prime example of ‘whitewashing’. Figure 2.8 shows an image from the set of *Demolition Man*. The filter captures almost all of Sylvester Stallone’s exposed skin (excluding his heavily shaded hands), whilst Wesley Snipes is almost completely missed - his bleached hair and a small section of his ear being all that remains.

Moreover, measuring skin tone in isolation simply reflects the *percentage* of pixels within an image showing such colours - not the nature of the content. A passport photo



Figure 2.8: Scene from the 1993 movie *Demolition Man*(IMDB.com, n.d.), plus skin tone mask. **Red** denotes skin tone pixels, **black** denotes all other colours

arguably contains a high proportion of skin tone, given the dominance of a ‘nude’ face across most of the image.

An excellent, in-depth review of colour based approaches for adult pornography detection can be found in Ries and Lienhart (2014). Vitorino et al. (2018) include a comprehensive overview on numerous pornography detection methods and products.

2.6.6 Introducing Machine Learning

Thus far, we have shown how the vast bulk (if not all) of existing automated CEM detection doesn’t seek out or otherwise identify CEM. Technologies such as cryptographic and fuzzy hashes merely recognise identical or similar materials, respectively. The *nature* of the content is not in any way relevant to search. The closest existing technology commonly applied to this task is skin tone detection, and this merely detects and measures the presence of pre-defined colours within an image.

Machine Learning (ML) in its many guises has certainly increased in profile within academia and industry over the last few years. Whilst the terms ‘artificial intelligence’ and ‘machine learning’ are popularly correlated with killer cyborgs a la James Cameron’s 1984 movie *The Terminator* and HAL in Stanley Kubrick’s *2001: A Space Odyssey*, such methods are comparatively mundane. ML is perhaps best summarised as an approach whereby computers ‘learn’ their requisite tasks through the use of data, rather than their every action being explicitly programmed by human operators. It is better to think of ML as a means towards artificial intelligence, driving the development of improved tools, particularly for well-defined, repetitive, and possibly mundane tasks. The sheer breadth of tasks, approaches and even applications of ML make it something more of a *class* of computing rather than a field of endeavour, though it should be noted that development of a truly ‘artificial mind’ remains elusive within the foreseeable future.

ML algorithms fall into two broad approaches: *supervised* and *unsupervised* learning. Supervised learning involves an algorithm being provided with labelled training data, and therefore allows tuning (typically via the manual design, implementation and weighting of feature extractors) towards a known desired outcome. Unsupervised learning is not

provided such information, making it suitable for extremely large datasets and application in knowledge discovery tasks such as document clustering (discussed in section 2.3.2).

We have already detailed the dangers of working with offensive materials such as CEM. The manual identification and extraction of features relevant to such materials inevitably requires *trainer* exposure during development and tuning, effectively protecting investigators and analysts, though at the unfortunate expense of data scientists. Beyond this serious ethical limitation, such an approach would be highly inefficient and likely doomed to failure. Manually identifying and extracting features is cumbersome, particularly within complex tasks such as image recognition. For example, the volume of features relevant to pornographic imagery, either in isolation or combined, is probably impossible to quantify. Human trainers are effectively required to wholly define pornography *through examples*, a task more difficult than the already difficult challenge of defining the concept through words alone.

Neural Networks

As the name implies, *artificial neural networks* (aka ‘neural networks’) are a machine learning approach inspired by the biological networks within the brain. They consist of nodes (emulating neurons), connected via links/edges (emulating synapses) to each node on neighbouring layers³⁸. The only observable elements of the network are the *input* and *output* layers, with remaining layers’ behaviour and performance *hidden* from view³⁹. Figure 2.9 displays a simple, one layer neural network.

Nodes (‘neurons’) are responsible for processing inputs received from preceding nodes (or in the case of input nodes, from outside the network - e.g. pixel RGB values), performing *activations* (in short, processing these inputs according to pre-defined functions) and outputting results via outbound connections. Activation functions are typically hard thresholds or logistic functions, though a more recent approach seen as providing good performance is the Rectified Linear Unit (ReLU) (LeCun et al., 2015) - a function returning 0 for any negative numbers, and the original input value for positive values. Figure 2.10 displays an example for each activation function.

Connections are responsible for transmission of activations between nodes. They are typically weighted for tuning purposes, and can operate according to two structures: *Feed-forward*, where information travels unidirectionally towards the output, with the network forming a stateless Directed Acyclic Graph (DAG) (Russell and Norvig, 2009). *Recurrent* networks can send outputs to its own inputs, giving it some degree of statefulness/‘memory’ at the cost of increased complexity. Feed-forward networks are well suited for static data such as still images, whilst recurrent networks have gained interest in video processing, due specifically to their ability to transfer knowledge between iterations - useful for retaining context when examining individual frames.

³⁸The number of layers forms a network’s ‘depth’

³⁹Methods exist (*back-propagation*) for inferring layers’ performance, but these are out of scope for this chapter

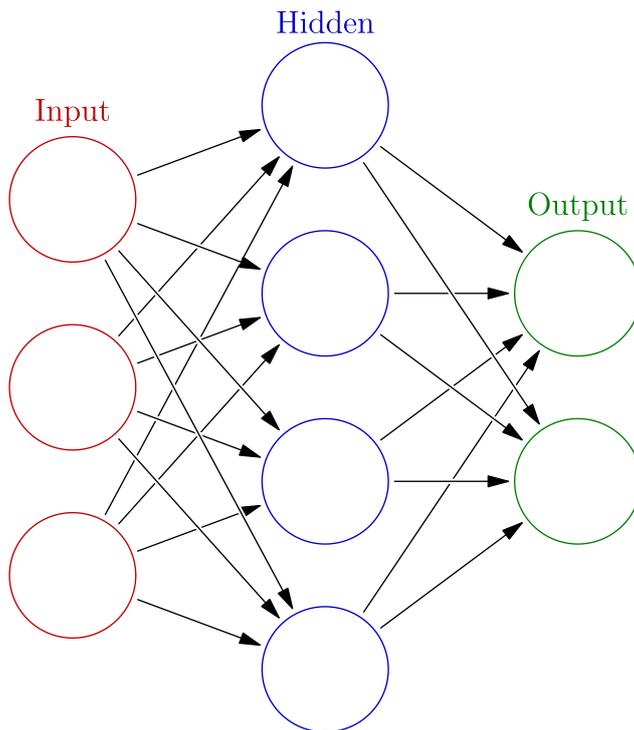


Figure 2.9: Example neural network. Note *feed forward* architecture, single hidden layer. (CBurnett, n.d.)

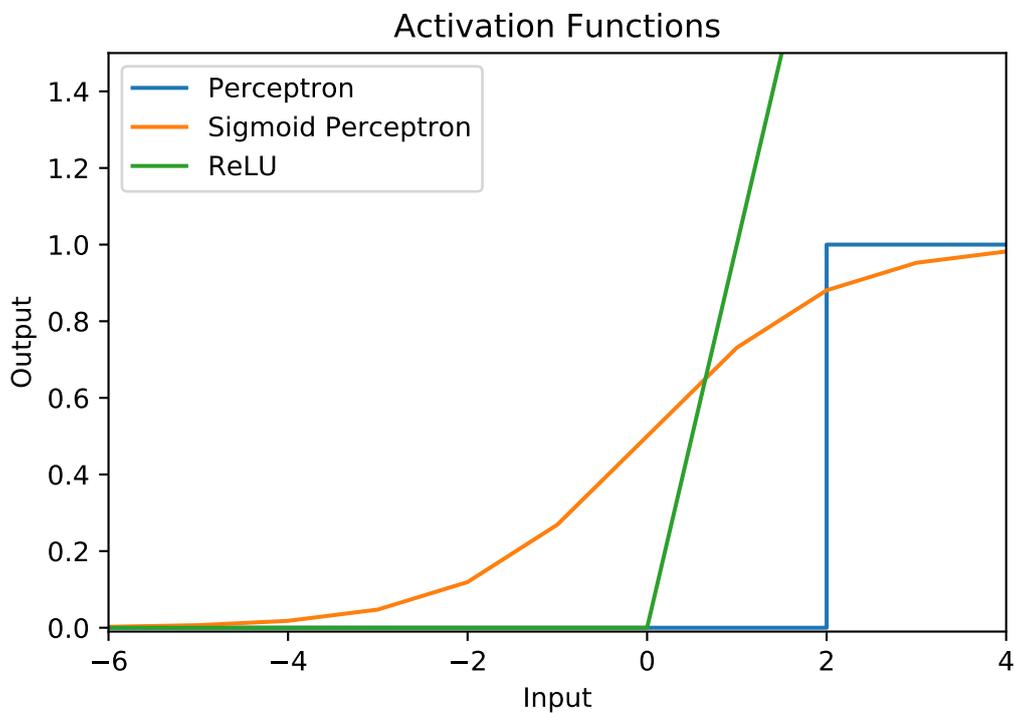


Figure 2.10: Example activation functions.

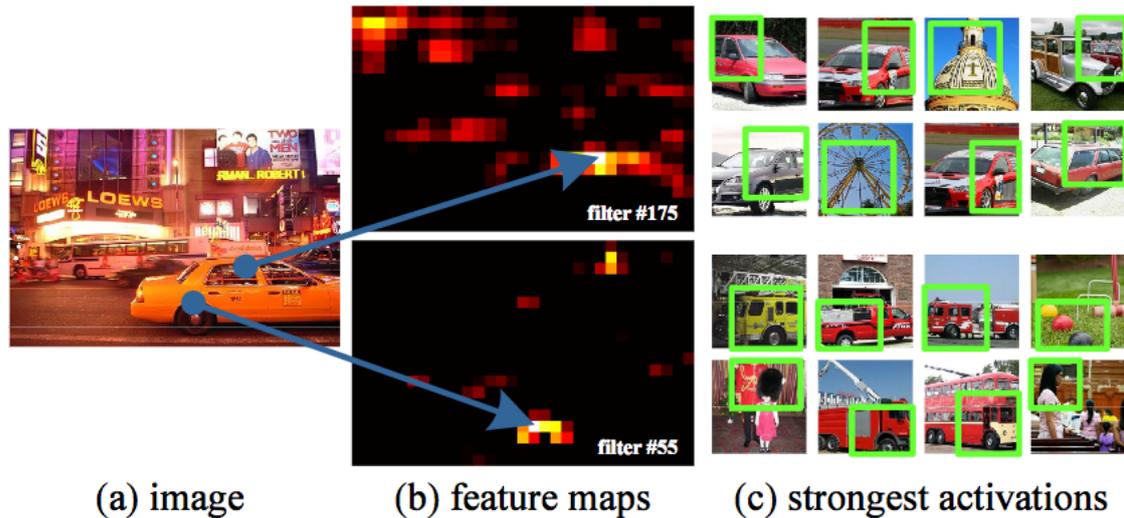


Figure 2.11: Example feature maps with sample activations (He et al., 2014). The upper feature map appears to detect car windows, whilst the lower seems to detect round objects.

Convolutional Neural Networks

Convolutional neural networks (ConvNets) are neural networks specifically suited for data represented by arrays, making them well suited for image recognition⁴⁰. Whilst the name seems complex, it is derived from *discrete convolution*, the process of combining and/or converting discrete signals to form a separate output. This function is performed by *feature maps*, an integral part of ConvNets, displayed in figure 2.11.

ConvNets became popular for image recognition after the network developed by Krizhevsky et al. (2012) won the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) in 2012 (LeCun et al., 2015) with a Top-5 error rate⁴¹ of 15.3% (second place returning 26.3%).

Advantages over Existing Methods

The Support Vector Machine (SVM) (Boser et al., 1992), a widely used ML classification methodology, works on the understanding that data assumed to fall into one of two classes can be mapped to a high-dimensional space in such a way that these classes can be linearly separated. In simpler terms, if the features of each exemplar are plotted, a straight line/hyperplane can be drawn between the classes. This assumes similarity between neighbouring examples, a perfectly feasible proposition in simple tasks (particularly those plottable on a two dimensional diagram), but exceedingly complex and restrictive when faced with the very high dimensionality and myriad variations possible between ostensibly similar images from the same class. This similarity results in an inability to disambiguate irrelevant features, restricting classifier efficacy to materials highly similar to

⁴⁰Still images being easily reducible to Red Green Blue (RGB) vectors of pixel values

⁴¹The rate of the correct category being in the top 5 results returned by the classifier. In this case, from 1000 possible categories.

training exemplars (Bengio et al., 2006). One needs only to consider the variety of backgrounds and lighting levels available within posed portraits (i.e. ‘passport photo’ style pictures of an individual’s face) to see the incredibly large variety inherent within even a constrained image type. LeCun et al. (2015) point out that image recognition classifiers need to be able to ignore irrelevant features such as lighting, position and orientation, giving the example of identifying the difference between a wolf and Samoyed dog. At the pixel level, two pictures of the latter could vary considerably in pose, lighting and background, whilst pictures of either breed in similar poses and conditions could appear quite similar.

Unlike the aforementioned disambiguation limitations, neural networks are not constrained to linear separators when mapping classes, giving far greater flexibility when dealing with features. With sufficient depth and nodes, neural networks are capable of learning intricate feature combinations, making them both less sensitive to irrelevant features seen across images and also more sensitive to small (but relevant) class-specific features (LeCun et al., 2015).

Multi-layer, non-linear machine learning architectures (commonly referred to as ‘Deep Learning’), occupy a suitable middle ground for automated image recognition. Reliance upon annotated training datasets determines that it is a supervised learning methodology. Indeed, humans are required to design networks, observe outputs and extensively tune variables. However, unlike the supervised methods previously described, deep learning methods effectively generate their own feature extractors, meaning less specific annotations (such as those already created by investigators) suffice. No exposure to materials during training is required. In effect, deep learning in our context is lightly supervised, and doesn’t require prior domain knowledge from its trainers. The practicality of this approach is further improved by the free availability of tools for building networks, such as TensorFlow (Abadi et al., 2015) and Caffe (Jia et al., 2014).

Rather than having trainers flag or highlight specific features within images or crafting highly feature-specific datasets, networks are provided with large quantities of labelled data (i.e. n labels per image rather than specific coordinates of features), with networks designed, tested and tuned around what *might* work in a process of trial and error. Training is often an exercise in quantity, with large volumes of data used to overcome a lack of specifically labelled features. The model learns common elements between samples of like classes, and vice-versa. Methods for augmenting small or otherwise limited datasets exist. For example, Krizhevsky et al. (2012) utilised random sampling of image regions (“patches”), horizontal reflections and RGB intensity shifts as a means to reduce overfitting in their ConvNet. The scale of data being utilised becomes clear when one considers the ILSVRC dataset consisted of 1.2 million images, with augmentation increasing this count to around 2.5 billion.

Limitations

Such an approach is obviously limited by processing capacity (the aforementioned ConvNet utilised 650,000 nodes and 60 million parameters), but the architecture itself is extremely

suiting to parallel processing - hence the sudden rush to massively parallel Graphics Processing Unit (GPU)⁴² infrastructure once costs became more reasonable. Another danger, however, comes directly from the opaque nature of the models' inherent networks. Being 'deep', it is impossible for trainers to fully understand exactly what the networks are observing and learning. Commercial facial analysis products based upon neural networks have been shown to underperform on non-white (and particularly non-male) subjects (Buolamwini and Gebru, 2018), indicating a likely unconscious bias by developers and dataset designers during development - effectively reproducing the previously mentioned limitations of supervised learning.

Another, more technical issue is efficiency - an obvious temptation during development and testing is to simply add layers and nodes to a network as a means of improving precision and/or recall. Whilst improving accuracy is important, increased complexity *will* increase computational and memory costs. Developers need to be mindful of their networks' deployability, particularly in the triage scenarios described in section 2.2.2. In this chapter we detail a classifier workflow designed around existing practices, but an obvious first step in improving performance would be to merge all modules into what hopefully constitutes a more efficient, single network. This option was not available at time of development due to dataset and processing limitations - primarily due to an ironic lack of innocent child imagery and *lawful* pornography, but also due to a more traditional lack of computing infrastructure.

2.6.7 Defining NSFW, Pornography, and CEM

Effective training of classifiers requires clear and objective labelling schemas with agreed terminology. An issue we have thus far avoided is the definition of CEM - as mentioned previously, no existing automated CEM detection methods actually seek the *concept*, instead merely measuring similarity to pre-existing, *known* data.

The definition of pornography, meanwhile, remains debatable in myriad contexts, often reduced to variations detailing materials produced and/or collected for sexual gratification. The broad range of human desires, kinks and peccadilloes means it is impossible to definitively predict what may or may not meet such a benchmark, with the oft quoted "*I know it when I see it*", made in reference to hard core pornography, perhaps best summarising the level of confusion regarding an all-encompassing definition of pornography itself.

Hard core pornography

"I shall not today attempt further to define the kinds of material I understand to be embraced within that shorthand description (hard core pornography); and perhaps I could never succeed in intelligibly doing so. But I know it when I see it..." *Justice Stewart (Jacobellis v. Ohio, 1964)*

⁴²Whilst not necessarily as fast as traditional CPUs, GPUs (aka 'video cards') are designed to work with parallel workloads, a common feature in tasks such as texture mapping.

Such ambiguity appears to have been accepted commercially. Figure 2.12 displays Microsoft Azure’s Computer Vision API providing confidence scores for “adult content”, but also for “racy” (seemingly an alternate term for Not Safe For Work (NSFW)). Such an approach appears particularly relevant for dealing with sensitivities regarding CEM - the sample image shown can be reasonably viewed as mildly ‘racy’, but beyond the relative absence of clothing, it is difficult to disambiguate this image from an innocent holiday picture. Any datasets for use in this context would benefit from a stronger focus on tangible aspects and classes (e.g. ‘swimsuit’) rather than ambiguous concepts such as NSFW.

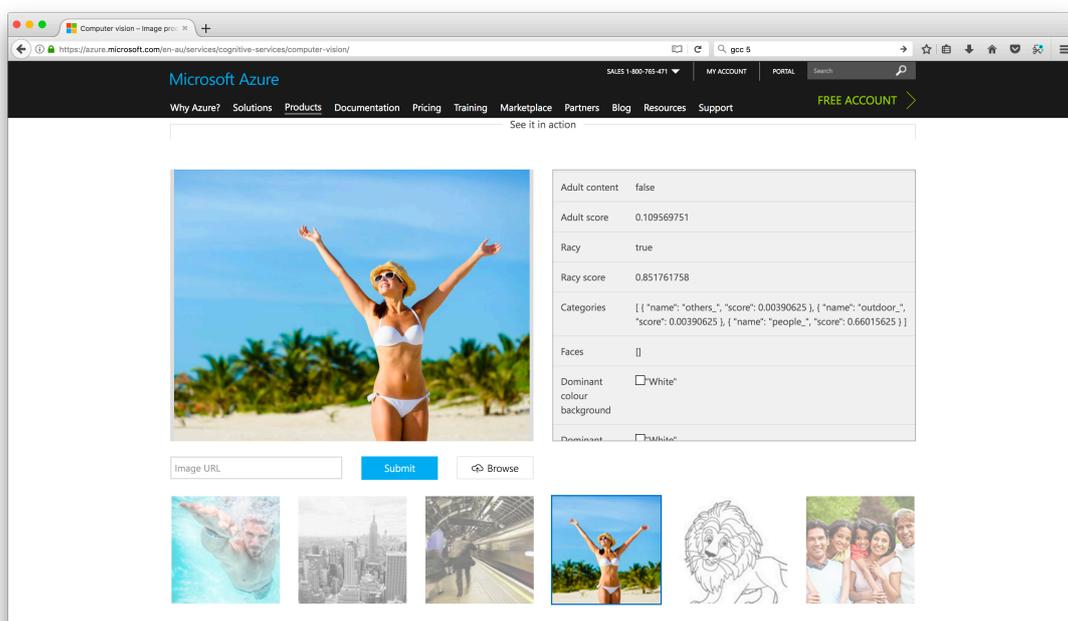


Figure 2.12: Microsoft Azure Computer Vision API demo (Microsoft, 2017) - note inclusion of ‘Adult Content’ and ‘Racy’

Perhaps surprisingly (given the common view of legislation being overlong, ambiguous and difficult to understand), the Australian legal definition of CEM, whilst wordy, is relatively clear, as detailed within *Defining CEM* (page 61).

Legislative definitions are adequate for establishing evidentiary proofs and informing any subsequent criminal charges, but the simplistic nature has effectively been deemed inadequate for the purposes of informing sentencing in many jurisdictions, where judicial officers request itemised counts of all CEM against taxonomies such as Child Exploitation Tracking System (CETS) and COPINE, summarised in tables 2.7 and 2.8. CETS is the current *de facto* standard across Australia. Both primarily serve as means for measuring offence severity, a useful metric in sentencing and investigation prioritisation.

Defining CEM

Child Exploitation Material (CEM) is the preferred term of law enforcement and partner agencies for the more colloquial ‘child pornography’. This is due to a perception that the term ‘pornography’ legitimises sexual exploitation of children by using a phrase common with lawful sexual content.

CEM and ‘child pornography’ are legislatively defined and can therefore differ between jurisdictions. Australian State and Territory legislation has gradually aligned over the last decade, with Commonwealth legislation serving as a ‘standard’ definition and law if and when required.

Commonwealth (Australia) legislation^a defines “child pornography material” as meaning:

- (a) material that depicts a person, or a representation of a person, who is, or appears to be, under 18 years of age and who:
 - (i) is engaged in, or appears to be engaged in, a sexual pose or sexual activity (whether or not in the presence of other persons); or
 - (ii) is in the presence of a person who is engaged in, or appears to be engaged in, a sexual pose or sexual activity;

and does this in a way that reasonable persons would regard as being, in all the circumstances, offensive; or

- (b) material the dominant characteristic of which is the depiction, for a sexual purpose, of:
 - (i) a sexual organ or the anal region of a person who is, or appears to be, under 18 years of age; or
 - (ii) a representation of such a sexual organ or anal region; or
 - (iii) the breasts, or a representation of the breasts, of a female person who is, or appears to be, under 18 years of age;

in a way that reasonable persons would regard as being, in all the circumstances, offensive; or

- (c) material that describes a person who is, or is implied to be, under 18 years of age and who:
 - (i) is engaged in, or is implied to be engaged in, a sexual pose or sexual activity (whether or not in the presence of other persons); or
 - (ii) is in the presence of a person who is engaged in, or is implied to be engaged in, a sexual pose or sexual activity;

and does this in a way that reasonable persons would regard as being, in all the circumstances, offensive; or

- (d) material that describes:
 - (i) a sexual organ or the anal region of a person who is, or is implied to be, under 18 years of age; or
 - (ii) the breasts of a female person who is, or is implied to be, under 18 years of age;

and does this in a way that reasonable persons would regard as being, in all the circumstances, offensive.

^a *Criminal Code Act (Cth) 1995, 473.1 - Definitions*

Category	Summary (abbreviated)
1	No Sexual Activity (restricted to nudity/suggestiveness)
2	Sex acts, either solo or between children
3	Non-penetrative sex acts between child(ren) and adult(s)
4	Penetrative sex acts between child(ren) and adult(s)
5	Sadism/Bestiality/Child Abuse
6	Animated/Virtual/Anime
7	Non-Illegal (typically part of series including CEM)
8	Adult pornography
9	Ignorable

Table 2.7: Child Exploitation Tracking System (CETS), abbreviated. Refer Appendix B.1 for complete version

Level	Name
1	Indicative
2	Nudist
3	Erotica
4	Posing
5	Erotic Posing
6	Explicit Erotic Posing
7	Explicit Sexual Activity
8	Assault
9	Gross Assault
10	Sadistic/bestiality

Table 2.8: Combating Paedophile Information Networks in Europe (COPINE), abbreviated. Refer Appendix B.1 for complete version

2.6.8 Applying Machine Learning to CEM Detection

By drawing their ‘learning’ from data rather than pre-defined features, ML methodologies are actually well suited for tasks with complicated or otherwise difficult to define concepts. Pornography/Not Safe For Work (NSFW) detectors are commonly implemented in corporate environments within e-mail scanners and internet proxy monitors, frequently operating in conjunction with supplementary materials such as black and white lists of known bad/innocuous sites. Typically such products focus upon adult materials ranging from swimsuit modelling to hard core pornography, due largely to (a) demand for such products in light of workplace harassment laws, and (b) the readily accessible and largely lawful nature of such materials for classifier training and development.

Moustafa (2015) slightly modified and combined the AlexNet and GoogLeNet networks to create a pornographic image classifier, achieving around 94% accuracy.

Of most relevance to our work is the approach taken by Vitorino et al. (2018), who created a two tiered CNN for CEM detection - the first step being the less sensitive task of general pornography detection, followed by a second, more sensitive (from the ethical and legal perspectives) step in child detection - limited by access to data.

This rather short list of published work represents an exhaustive survey of publicly accessible research into deep learning based CEM detection. Deep learning is a data ‘hungry’ methodology, relying upon sufficient quantities of individual examples to detect features and relationships unique to the provided classes/categories.

2.6.9 An Unfortunate Dearth of Data

As mentioned previously, research in the field of online child exploitation is largely unworkable within academia and vast swathes of industry due to the ethical, legal and welfare implications of accessing, possessing and/or viewing such materials.

Realistically, the only organisations *intentionally* gathering, *labelling* and *sharing* (to whichever extent) CEM as part of their *core* ‘business’ are law enforcement agencies, making them an obvious point of contact and collaboration. Caetano et al. (2016) used the Pornography-2K dataset (Moreira et al., 2016) (itself an extension of the dataset produced by Avila et al. (2013)) for training an adult porn classifier, but were reduced to indirect access to data from one hard drive from one case being conducted by the Brazilian Federal Police for training, testing and validating their CEM classifier. This limitation is entirely understandable and beyond the control of the authors, but we assert that such a tight focus runs the risk (if not guarantee) of overfitting⁴³, due largely to (a) suspect/offender tastes, predilections and methodologies, and (b) temporal ‘staleness’ due to trends and fashions - not only in terms of offending, but fashions and appearances of persons and objects viewed within imagery and multimedia.

⁴³The inadvertent design or tuning of a classifier to perform best on the sample dataset rather than the actual population

Grajeda et al. (2017) don't report any pornographic (lawful or otherwise) datasets within their survey of digital forensic datasets, with most image/multimedia sets gravitating towards more 'traditional' DF topics such as steganography and device (e.g. camera) forensics.

Suprisingly, this dearth of CEM also extends to lawful materials. In their survey of adult pornography detection methodologies, Ries and Lienhart (2014) mention the absence of shared, publicly available databases of *adult* pornography, leading to the conclusion that individual research in most cases "*can't be quantitatively compared*".

Researchers therefore are forced to create their own corpora, even for lawful materials. Avila et al. (2013) create and use a pornography image dataset in order to test the application of their concept detection system ('BossaNova') in pornography detection. Generated by extracting frames from pornographic and non-pornographic movies, the authors extended the corpus' research value by intentionally classifying innocuous content according to the difficulty of disambiguation with porn. They also focus upon multi-ethnic content across genres. The authors have made this corpus freely available for research purposes (subject to a licensing agreement), and the corpus has since been used in further research by Caetano et al. (2016); Moustafa (2015). The dataset itself was extended (Moreira et al., 2016), adding further content and overcoming a possible limitation caused by the original version's reliance upon specialised pornography websites for the 'pornographic' content.

Simulating CEM is a possible approach. Sae-Bae et al. (2014) were forced to use explicit adult images for training and validating elements of their automated child pornography detection system, with a limited (105 image) corpus of 'explicit-like' images of children for validating their overall performance. We are uncertain how the authors were able to objectively measure the 'likeness' of their images to CEM, beyond an obvious focus on materials depicting persons with youthful appearances or made to look like minors. In either case, this may be adequate for more mature minors, but of limited value for younger persons - almost guaranteeing a corresponding bias in any resulting tools.

An approach we considered as part of our research is the combination of an off the shelf pornography detector with a custom age-detection tool. However, this also presented issues with data. Chatzis et al. (2016) identified the absence of a standard test database when researching facial features (in particular, face to iris ratio) as a means for identifying children within images. In particular, no 'ground truth' system with confirmed ages was found to be available. Instead, a collection of 75 publicly available images of persons with known ages was used - a sample of which indicated a strong bias towards images at least capable of portrait-style cropping (i.e. reduction to a passport-style image restricted to the subject's face from approximately directly ahead). Sensing a lack of suitable datasets whilst researching automated face-based age and gender estimation, Eidinger et al. (2014) assembled a corpus of approx. 26,580 age and gender labelled images of 2,284 subjects. Critically, the images are "in the wild" - i.e. with unpredictable variations in conditions such as lighting, poses, and background (Wang et al., 2013). Eight age groups are provided,

but unfortunately for CEM identification purposes in the Australian context⁴⁴, one of the labelled age groups is ‘15-20’, making it of limited use in disambiguating near-legal and legal ages.

In their further work on the topic of age and gender classification, Levi and Hassner (2015) summarise the challenges of data corpora succinctly - gathering a labelled image set of ages and genders either requires access to private information, or sufficient resources to undertake a tedious, time consuming labelling exercise. Assembling a CEM corpus represents a tedious, time consuming and also psychologically harmful extension to this challenge.

To date, we have been unable to identify any record of a documented CEM dataset being made available for research, beyond the ad-hoc, ‘per case’ arrangement directly made between researchers and law enforcement. This data ‘drought’, whilst understandable, is nonetheless ironic given that many law enforcement organisations are flooded, unable to keep up with the volumes of such materials being seized.

2.7 Conclusions

Existing DF research is heavily focused upon efficiency gains through either reducing the amount of data analysed or improving the presentation of data to analysts - for example, through the clustering of like content.

Such approaches are of obvious and proven value, though with differing impacts depending upon the nature of data being sought and the software tools available to analysts. We respectfully posit that automated identification of materials of interest is a critical, complementary means for not only dealing with workloads but also ensuring analyst welfare. Unfortunately, such automation is currently limited to the sharing of hashsets, an inflexible and dated approach limited to identification of previously viewed materials.

A common element of the methodologies within Section 2.3 is that whilst useful in assisting reviewers in sorting, visualising and arranging data, they ultimately rely upon manual review. This reduces workloads and accelerates review up to (but not including) the point of actually examining files - a major shortcoming when considering the stressors and dangers listed within Section 2.5. The only way to automate this step of the analysis process is to effectively teach an algorithm to recognise content, in a robust and sustainable manner. Simple examples and proofs of concept are helpful as one-off tools, but fail to enable standardisation and discourage development of integrated tools (the absence of both being identified challenges in DF).

In order to achieve standardisation within DF, one needs to construct an ontology, codifying what is sought.

This can arguably be done at two levels:

1. **Macro:** What activities and entities exist in our environment? What is the landscape?
2. **Micro:** What are the characteristics of those individual items/activities?

⁴⁴the age of ‘adulthood’ in terms of CEM within Australia being 18 years

In Chapter 3 we address the *macro* by introducing the Tor-use Motivation Model (TMM), a two dimensional schema capable of recording the activity ('what') and motivation ('why') of criminal behaviour online. In Chapter 4 we provide an exemplar of the *micro* through the Majura Schema, a CEM ontology capable of supporting not only machine learning, but also cross-jurisdictional knowledge sharing through a focus on tangible features rather than abstract concepts.

In Chapter 5 we apply our research through the introduction of Monte Carlo Filesystem Search (MCFS), an automated crawl strategy specifically designed to meet the challenges of on-site triage - the algorithm is unsupervised and 'learns' from the encountered landscape, reducing analyst workload *and* time spent on premises.

Chapter 3

Understanding and Categorising Online Criminal Activity

Introducing the Tor Motivation Model

This chapter investigates the structure and content of a popular *dark web* as a case study to aid the efficient, automated identification of materials of interest to law enforcement. Exhaustive examinations of networks such as the World-Wide Web (WWW) have been shown to be expensive, inefficient and ultimately fruitless, particularly when dealing with dynamic content (Chakrabarti et al., 1999). Focused crawls are far more efficient, and have the added bonus of avoiding blanket crawls and their perceived association with mass surveillance. They do, however, require an underlying understanding of topics, both of interest and ignorable, in order to make informed decisions about likely locations of value. An ontology capable of representing and recording all online criminal activity is therefore warranted.

The ‘ontologisation’ of criminal activity, with a particular focus on electronic/online actions, is perhaps not as simple as it sounds. One could read through all relevant legislation, itemising each offence and applicable proofs, but such an approach would result in an overly prescriptive and restrictive ‘map’. Firstly, the detection and interdiction of crime does not require the establishing of all relevant proofs against a specific offence - this is the purpose of an investigation. Secondly, legislation isn’t static, with offences introduced, amended and withdrawn on a near-constant basis. Thirdly, legislation is jurisdictionally specific, reducing (if not eliminating) the value of sharing any resulting schemas. Australia alone could possibly produce nine ‘correct’ schemas - one for each State and Territory, plus the Commonwealth.

We have already discussed the varying definitions of “dark”, “deep”, “invisible” and “hidden” webs in popular culture, industry and academia in Section 2.3.4. In the interests of brevity and clarity, for the purposes of this Chapter, we regard “dark webs” as networks supporting the *non-coordinatability of traits* (Wallace, 1999) - in other words, systems enabling remote communications *without* requiring the provision of features capable of being used to de-anonymise users.

3.1 Methodology & Scope

We selected The Onion Router (Tor) as our “dark web” exemplar because of its popular association with illegal activities such as narcotics trafficking, a perception due largely to the rise (and fall) of the “Silk Road” - an infamous online marketplace already detailed within chapter 2.

In order to construct our motivation model we initially undertook a limited crawl of Tor, manually labelling 500 randomly selected pages using an author defined controlled vocabulary (refer Appendix A.1 for labels). Richly descriptive, the results proved of limited value, with a lack of specificity leading to inconsistent results across otherwise highly similar materials. During the labelling process, we observed that whilst many of the sites’ topics appeared dissimilar, their motivations for existence were tightly aligned. For example, many sites were specifically aiming for financial profit via online marketplaces of otherwise dissimilar products or services. We therefore introduced the idea of a structured model, separating the topic(s) discussed from the motivation for discussion.

This chapter is focused upon improving techniques available to law enforcement approaches for monitoring and investigating illicit behaviour on Tor. We regard technical exploits and general methods for undermining or circumventing the security of Tor as outside of scope for this research, for reasons including:

- **Ethical:** Undermining a protocol used for lawful (and ethical) purposes will negatively impact users’ use of the network, particularly in cases where such users’ personal safety and liberty is placed at risk.
- **Legal:** Law enforcement’s use of technical attacks (and surveillance in general) is regulated by the laws of their relevant jurisdiction. Such legislation generally requires a level of suspicion/belief of specific illegal behaviour, and limits surveillance to specific and associated entities. It is unlikely that widespread, ‘blanket’ police surveillance of a service such as Tor would be allowed within a jurisdiction such as Australia.
- **Practical:** The open-source nature of the Tor Project’s software allows for the free publication, critical appraisal and distribution of software hacks and patches. Wang et al. (2011) and also Ling et al. (2013) freely detail such vulnerabilities, with subsequent fixes (where required) readily distributed. Such an unstructured and open process makes exploits’ lifecycles difficult to anticipate, further compounding the ‘arms race’ of patching and cracking.

Our interest in anonymous networks is limited to the efficient identification of illegal behaviour - a focused crawler for law enforcement, capable of limiting the impact of additional latencies associated with Tor. Where possible, ‘value adding’ features such as attempts at identifying the geographical location of illegal behaviour will be examined.

The absence of ‘value added’ information limits the use of previously proposed taxonomies / topic groupings (refer to Table 2.4) for obtaining an in-depth understanding of sites and their associated purposes. The use of *ethics* as a differentiator (Guitton, 2013)

Topic	Motivation/Purpose
Drugs/Narcotics	Education & Training
Extremism	File Sharing
Finance	Forum
Hacking	General
Identification/Credentials	Information Sharing/Reportage
Intellectual Property/Copyright Materials	Marketplace/For Sale
Other - Not of interest	Recruitment/Advocacy
Pornography - Adult	System/Placeholder
Pornography - Child	
Pornography - Illicit or Illegal	
Search Engine/Index	
Unclear	
Violence	
Weapons	

Figure 3.1: Tor-use Motivation Model

moves toward a model more suitable for law enforcement use, but the inclusion of distasteful yet legal¹ content adds potential complexity without corresponding improvements in application. Furthermore, the *prioritisation* of scarce investigative and technical resources requires a deeper understanding of a site’s underlying motivations, beyond the topic itself. The immediate threat of offending should make a website hosting, sharing and inciting the production of ‘real world’ Child Exploitation Material (CEM) of greater concern to law enforcement than a site hosting ‘fantasy’ texts.

We took elements of the taxonomies and topic groupings listed in Section 2.3.4, dropping or enhancing topics according to their relative interest to law enforcement. We then combined these topics with *motivations*, in order to provide relevant information to classifiers with minimal additional complexity. Figure 3.1 summarises the proposed topics and motivations, with detailed explanations provided in Tables 3.1 and 3.2, respectively.

The presence of ‘Adult Pornography’ as a distinct category does contradict our focus on illegal materials. The inclusion was debated, and was ultimately retained in order to remain consistent with the Child Exploitation Tracking System (CETS), a grading scale currently in use by Joint Anti Child Exploitation Team (JACET)s across multiple Australian jurisdictions.

3.2 Conducting the Crawl

Our research on focused crawl strategies (refer Chapter 2) influenced the logical representation of our crawl as a tree, with seed sites forming a first branch from a virtual root. All subsequent outgoing links were added to the tree as children of the referring nodes, with the following rules applied for security and performance reasons:

- All files required to successfully render content (e.g. images within HTML pages) were immediately downloaded at time of accessing the initial document; and

¹for example, non-violent racial discrimination

Topic	Explanation
Drugs/Narcotics	Illegal drugs/chemical compounds for consumption/ingestion - either via blanket unlawfulness (e.g. proscribed drugs) or via unlawful access (e.g. prescription-only/restricted medications sold without lawful accessibility).
Extremism	Illegal or ‘of concern’ levels of extremist ideology. Note this does not provide blanket coverage of <i>fundamentalist</i> ideologies and dogma - only those associated with illegal acts. Socialist/anarchist/religious materials (for example) will not be included unless inclusive or indicative of associated illegal conduct, such as hate crimes.
Finance	Any monetary/currency/exchangeable materials. Includes carding (sale of stolen credit card details), Bitcoin, Litecoin etc.
Hacking	Materials relating to the <i>illegal</i> access to or alteration of data and/or electronic services.
Identification / Credentials	Materials used for providing/establishing identification with third parties. Examples include passports, driver licenses and login credentials.
Intellectual Property / Copyright Materials	Otherwise lawful materials stored, transferred or made available without consent of their legal rights holders.
Other - Not of interest	Material not of interest to law enforcement - e.g. personal sites, Facebook mirrors.
Pornography - Adult	Lawful, ethical pornography (i.e. involving only consenting adults).
Child Exploitation	Child abuse materials (aka child pornography), including ‘fantasy’ fiction materials, CGI. Also includes the provision/offering of child abuse materials and/or activities.
Pornography - Illicit or Illegal	Illegal pornography NOT including children/child abuse. Includes bestiality, stolen/vengeance porn, hidden cameras etc.
Search Engine / Index	Site providing links/references to other sites/services. Referred to as a ‘nexus’ by Moore and Rid (2016)
Unclear	Unable to completely establish topic of material.
Violence	Materials relating to violence against persons or property.
Weapons	Materials specifically associated with materials and/or items for use in violent acts against persons or property. Examples include firearms and bomb-making ingredients.

Table 3.1: TMM Topics

Motivation	Explanation
Education & Training	Materials providing instruction - e.g. ‘how to’ guides
File Sharing	General file sharing, typically (but not limited to) movie/image sharing
Forum	Sites specifically designed for multiple users to communicate as peers
General	Materials not covered by the other motivations. Typically, materials of a nature not of interest to law enforcement. For example, personal biography sites.
Information Sharing / Reportage	Journalism/reporting on topics. Can include biased coverage, but obvious propaganda materials are covered by <i>Recruitment/Advocacy</i> .
Marketplace / For Sale	Services/goods for sale, regardless of means of payment.
Recruitment / Advocacy	Propaganda
System/Placeholder	Automatically generated content, not designed for any identifiable purpose other than diagnostics - e.g. “It Works” message provided by default by Apache2

Table 3.2: TMM Motivations

- Any outbound links previously observed as children of other documents were ignored. A move to a Directed Acyclic Graph (DAG) based data structure for keeping track of such links is proposed for future research.

3.2.1 Legal and Ethical Considerations

Possessing, accessing and causing the transmission of child pornography are criminal offences within the Commonwealth of Australia and the State of Victoria, where this research was undertaken. In addition to ethical clearance, permission for conducting this crawl was obtained from The (Commonwealth) Minister for Justice, Attorney General’s Department, and the Australian Federal Police (AFP). Relevant state authorities were notified through existing joint agency agreements.

We followed established conventions for ethically conducting the crawl. For example:

- **Robots:** Our crawler respected `robots.txt` files, including rate limits.
- **Rate Limits:** In addition to rate limits specified within the robots file, the crawler itself was set to not revisit a domain within two minutes. Additionally, three consecutive *get* rejections/failures from a domain were treated as an indication of a rate limitation, with the domain subsequently blocked internally for at least 24 hours. Subsequent rejection was then treated as a dead link, with the domain dropped from consideration.
- **Populating Forms:** The crawler makes no attempt to generate user credentials, complete *captchas* or otherwise obtain access to restricted sites.

We acknowledge these self-imposed limitations possibly led to under-reporting of the relative quantity of illicit content. However, in the case of populating forms, a material proportion of sites’ content and nature were ascertainable from the login screen.

3.2.2 Implementation and Security

The previously detailed limiting of the crawl to readily available materials made the data stored within the system of relatively low value to external parties. Nevertheless, we regarded the possibility of inadvertent or involuntary sharing of CEM to be an unacceptable risk, particularly if our crawler’s identity and location were identified via a malicious site or exploit. Security was therefore regarded as the principal design consideration.

The datastore was hosted in a MongoDB(MongoDB inc, n.d.) database, secured with SSL/TLS using random passwords, upon a RAID-10 array encrypted using LUKS. Remote access was limited (via an iptables firewall) to ssh and MongoDB, with all passwords based on SHA-512 hashes of 100,000+ character random streams. Database access required validated X.509 certificates. IP based restrictions were also implemented using tools such as *fail2ban*(Fail2Ban.org, 2016) and blacklists of known bad IPs. Any unsuccessful login attempts resulted in the remote address being immediately banned for several hours. The server’s swap partition was encrypted with a random key assigned upon boot, reducing the

likelihood of passwords being recoverable in the event of the physical server being stolen or misplaced.

The crawler agent operated from a separate host, with access to the Tor network via a locally hosted squid (Squid Project, 2015) proxy. Remote access was limited to SSH (authenticated via public/private key), with an iptables firewall closing all other inbound ports.

In summary, the content of the crawl (including cached data) was never co-hosted with the crawler, increasing protection from external parties even beyond the anonymity provided by Tor.

3.2.3 Seeding and Steering the Crawl

We bootstrapped the crawler with several readily available lists of known hidden services, sourced from advertised Tor indexing services and lists, as listed within Table 3.3. No filtering of sites was made, beyond removal of duplicate entries. We emphasise that whilst sites hosting CEM or other illicit activities were identified within the seed corpus, we **do not** allege or infer that **any** of the aforementioned services knowingly included sites hosting such materials within their indices, nor that the listed sites hosted such materials at the time of their inclusion. For ethical reasons, we have chosen not to detail which source(s) contained links to such sites.

Source	Sites
https://ahmia.fl	4,989
http://darkspider.info/onionlist.txt http://darkspider.info/uponionlist.txt	12,760
https://github.com/kenorb/cicada-2014/blob/master/stage11/scripts/onions-list.txt	3,205
Total Unique:	15,503

Table 3.3: Tor Bootstrap sources.

Unlike the prior work mentioned in Section 2.3.3, no attempt was made to steer the crawler towards any particular genre or topic. The crawler was not discouraged from leaving Tor. In fact, no ‘safe’ domains were defined, with the crawler encouraged to examine all links and domains encountered during the crawl, subject to the previously detailed self-imposed limitations.

3.2.4 Labelling

We developed a review application in Java for labelling content acquired during the crawl. Files were rendered within an integrated browser, with executable content and out-links disabled, and rendering content (e.g. `img src` tags in HTML) redirected to local copies. The user was presented with the content, plus categories, motivations and languages for ease of labelling. Figure 3.2 displays the application user interface displaying a *not of interest* document - in this instance, a system/placeholder.

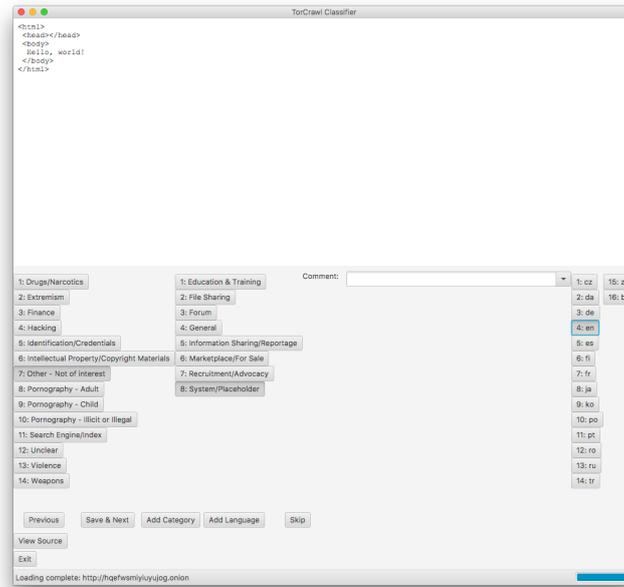


Figure 3.2: TMM Labelling App

3.3 Results - Applying the TMM

Of the original 15,503 unique *.onion* URLs used to bootstrap the crawl, 4,089 were found to be up and responding at the time of the crawler’s request. The crawler was allowed to run on a randomised selection basis from 12 April 2016 to 01 July 2016, making 1,210,089 successful (HTTP code 200) GET requests from 1,313,714 attempts. 1,155,549 unique URLs were accessed from 42,013 unique domains. 408,216 of these were Tor URLs, from 7,954 Tor virtual domains (i.e. *.onion*).

HTTP **Content-Type** headers were found to be unreliable, with incorrect values often recorded by the servers. Hence, all content returned by successful HTTP GET requests were automatically tested for content type using the Apache TIKA project (The Apache Software Foundation, n.d.). Figure 3.3 displays the top 10 media types for WWW sites, along with their relative popularity within Tor sites crawled.

The high frequency of plain text files within the Tor network appears to largely consist of ‘keep out’ messages. Akin to a `robots.txt` file, these appear to be used to inform the reader that their presence is not welcome on the site. Interestingly, a second use appears to be server diagnostics and monitoring, with files consisting of JSON formatted strings listing usage statistics and uptime appearing regularly throughout the labelling process.

Unique Content

We observed a relatively higher incidence of repeated content within Tor sites, as compared with WWW domains encountered during the crawl. Figure 3.4 shows that whilst individual WWW domains typically contain more pages than those on Tor, the ratio of non-unique pages is lower than on Tor based sites.

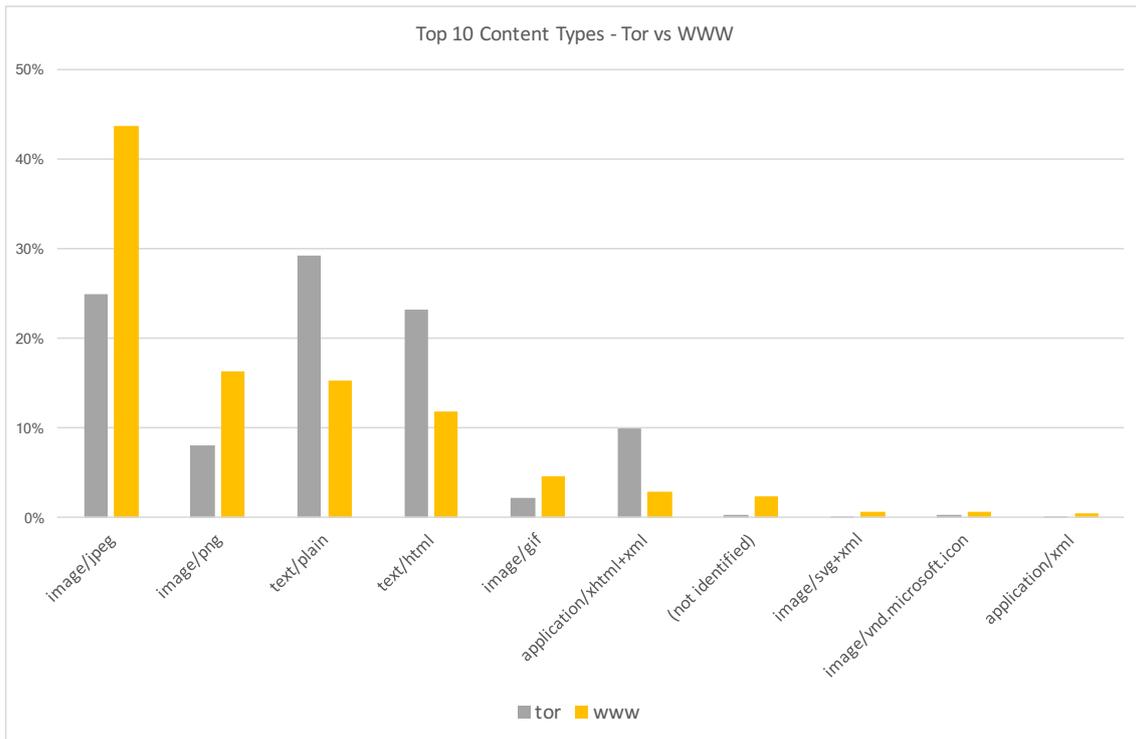


Figure 3.3: Top 10 media/content types - Tor vs WWW

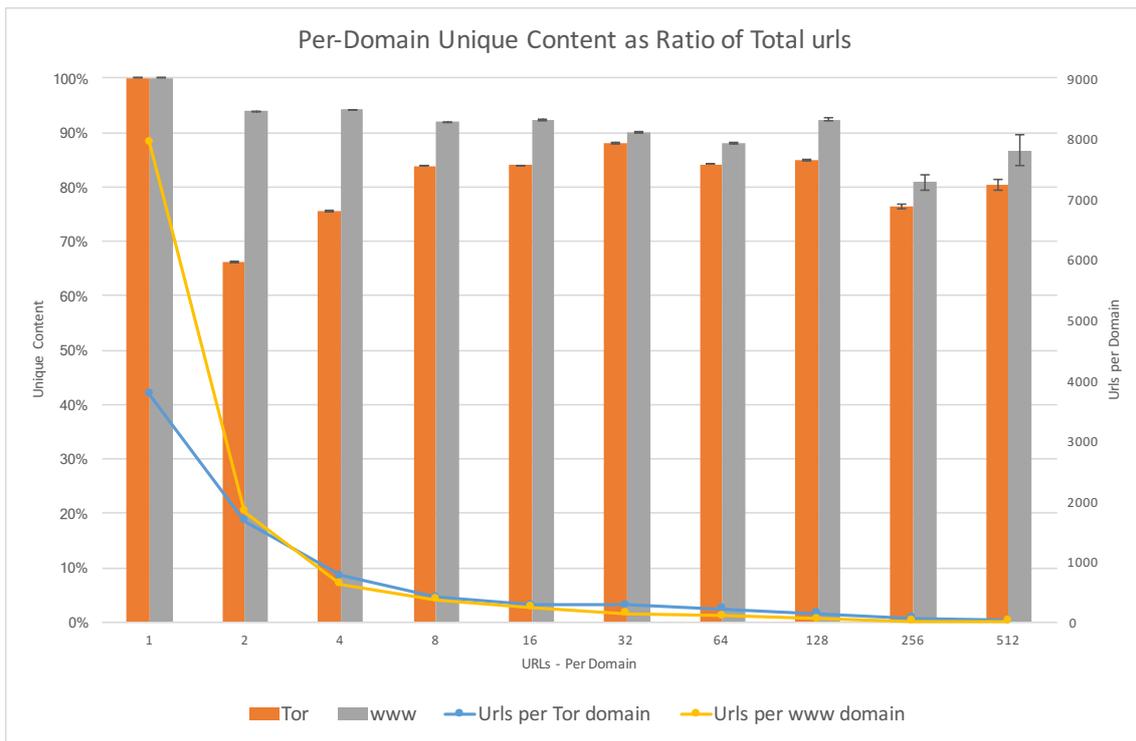


Figure 3.4: Site sizes and actual unique content - Tor vs WWW

3.3.1 Seed Sites

Figure 3.5 (table 3.4 refers) shows categories for sites detected as English language from the seed websites. Some key points:

1. Almost a third of unique sites are not of particular interest from a legal/ethical perspective;
2. Finance focused websites (typically Bitcoin tumblers and escrow services) outnumber drug/narcotics oriented sites;
3. Sites hosting blatantly illegal materials such as CEM sites appear within directories and indices readily available on the WWW; and
4. Only one extremist site was observed, possibly indicating:
 - (a) a move away from ‘broadcast’ style radicalisation and communication by such organisations; or
 - (b) a lack of preparedness to ‘advertise’ such sites on such services.

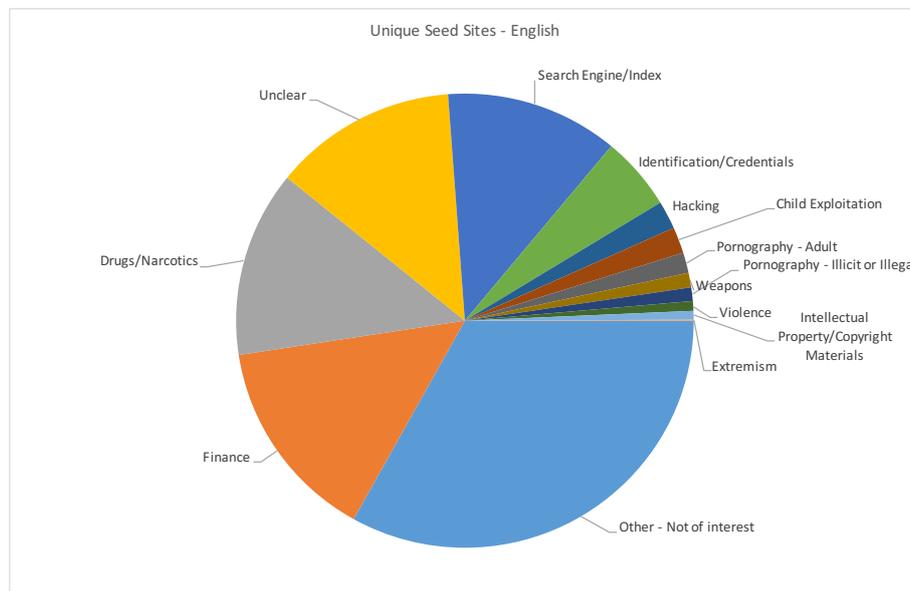


Figure 3.5: Labelled Categories - Unique English language sites

The absence of `robots.txt` files or logins/captchas on sites hosting CEM leads to several possible conclusions:

1. technical naiveté regarding the avoidance of accidental or incidental detection;
2. indifference at such detection, possibly due to complete faith in the anonymisation offered by Tor; and/or
3. a conscious decision to permit scraping/crawling, allowing for possible inclusion in search engines and other such directories.

3.3.2 All Sites

We labelled 2,419 unique documents (by SHA-1 value) from the crawl, using the TMM. Due to the presence of excessively large (in terms of page count - refer Figure 3.4) domains

Category	Unique Sites Observed	Percentage
Other - Not of interest	479	33.08%
Finance	210	14.50%
Drugs/Narcotics	192	13.26%
Unclear	188	12.98%
Search Engine/Index	177	12.22%
Identification/Credentials	76	5.25%
Hacking	29	2.00%
Pornography - Child	27	1.86%
Pornography - Adult	21	1.45%
Weapons	15	1.04%
Pornography - Illicit or Illegal	14	0.97%
Violence	10	0.69%
Intellectual Property/Copyright Materials	9	0.62%
Extremism	1	0.07%

Table 3.4: Unique English language seed sites - by category

within Tor, we chose to report and measure the occurrences of categories on a *per domain* basis. Figure 3.6 (Table 3.5 refers) shows the relative occurrence of categories within Tor. Note that whilst a number of sites were observed to contain multiple categories (for example, a series of sites offering search/indexing services exclusively focused upon narcotics vendors), the relative complexity of these combinations was excessive to report, with most combinations being infrequent (and therefore statistically insignificant given our sample size). For this reason, we chose to report on a per-category basis.

Of note:

- Approximately 40% of domains within Tor are *prima facie* not of interest to law enforcement. When including *unclear* content, over half is not of interest;
- Amongst the “of interest” categories, *finance* is the most prevalent, appearing on 16% of domains;
- In terms of sexually focused materials, CEM and illicit/illegal pornography occur on roughly similar terms of 1.75%, occurring almost 25% more frequently than lawful adult pornography.

Figure 3.7 (Table 3.6 refers) displays the ratio of domains displaying particular motivations, reflecting the dominance of products and services for sale within the network. Interestingly, the number of placeholder and system generated messages appears high, though one suspects that many of these may be diagnostic messages for hosts serving other applications and protocols through different, non-advertised ports.

The value of the Tor-use Motivation Model (TMM) becomes more apparent when categories and motivations are combined. Figures 3.8 and 3.9 show the motivations observed around domains categorised as serving drug/narcotics, child pornography, and finance related content.

Category	Domains
Other - Not of interest	1281
Finance	476
Search Engine/Index	396
Unclear	255
Drugs/Narcotics	213
Identification/Credentials	96
Hacking	54
Child Exploitation	53
Pornography - Illicit or Illegal	52
Pornography - Adult	42
Intellectual Property/Copyright Materials	25
Weapons	20
Violence	10
Extremism	2

Table 3.5: (Virtual) domains by category

Motivation	Domains
Marketplace/For Sale	1058
System/Placeholder	645
General	599
File Sharing	207
Forum	163
Education & Training	111
Information Sharing/Reportage	89
Recruitment/Advocacy	59

Table 3.6: (Virtual) domains by motivation

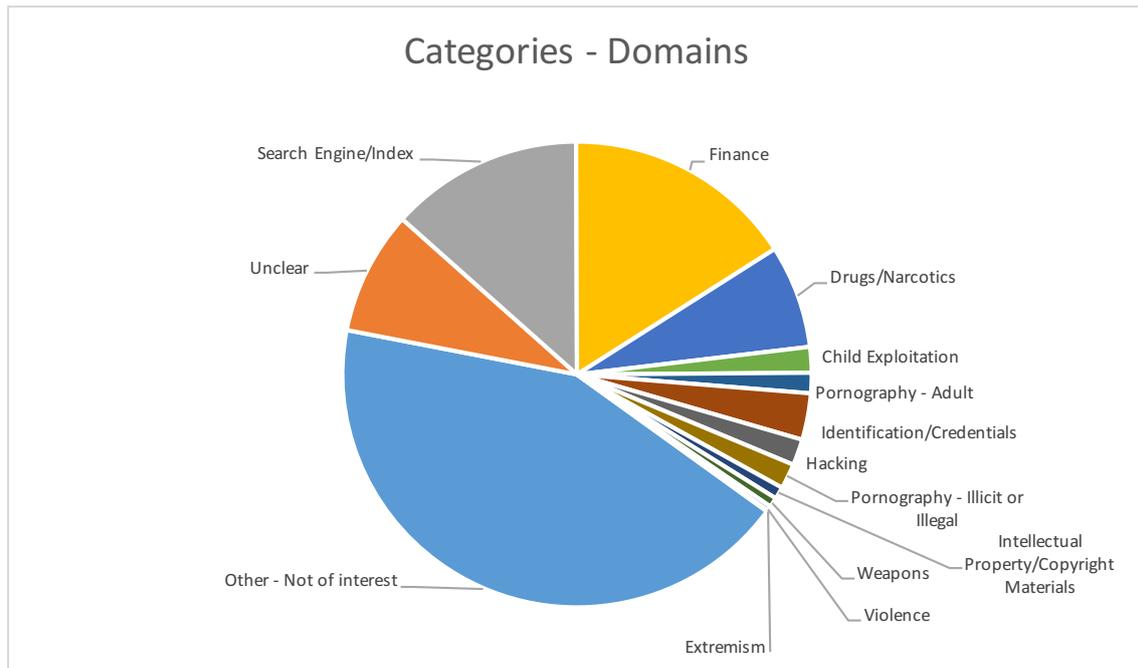


Figure 3.6: (Virtual) domains by category

Unsurprisingly given popular reportage of sites such as the aforementioned Silk Road, drug/narcotics related content (Figure 3.8) strongly trends towards commerce, with approximately 82% of drug/narcotic related domains purely designed to sell. Including the relatively small percentage of domains offering other facilities such as fora and information sharing, the total proportion exceeds 90%. Some fora were also observed, usually discussing locations and/or vendors selling, but also as places for discussion of specific drugs as alternative medications for mental illness.

Finance related domains were observed to be extremely heavily skewed towards commerce, with over 95% of sites offering products/services for sale. These largely consisted of Bitcoin ‘tumbling’ (aka laundering) sites, exchanges, and also services appearing to offer a ‘get rich quick’ scheme based upon a claimed flaw within the Bitcoin protocol. Stolen credit cards, bank/PayPal accounts and store gift cards dominated other services.

Domains relating to child exploitation (Figure 3.9) proved surprising. Whereas we anticipated a large proportion of file sharing sites, only 52% of identified child exploitation related domains provided such a service, amongst which over half required payment. About a quarter were forums, consisting of online discussions conducted by self-confessed paedophiles regarding predatory practices and the collection/sharing of child pornography. One site advertised itself as a ‘support’ forum, though in terms of supporting and normalising *paedophilia* rather than aiding persons avoid such behaviour and actions.

By way of comparison, other illicit/illegal pornography associated domains (refer figure 3.10) skewed heavily towards commerce, with 92% of domains offering materials for sale. Contrastingly, just over half the domains offering adult pornography (refer 3.11) did so on a ‘user pays’ basis, with 42% of such domains offering materials for free.

Of particular concern were three domains (5%) providing education and training for child exploitation, particularly in terms of grooming vulnerable children. Whilst these

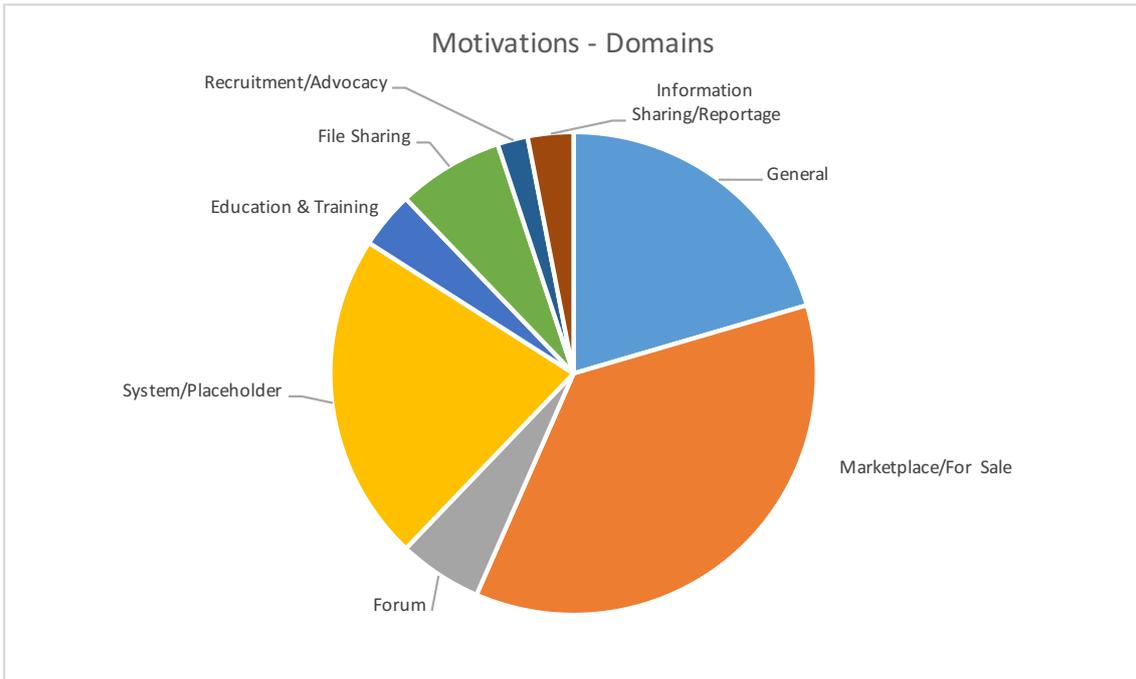


Figure 3.7: (Virtual) domains by motivation

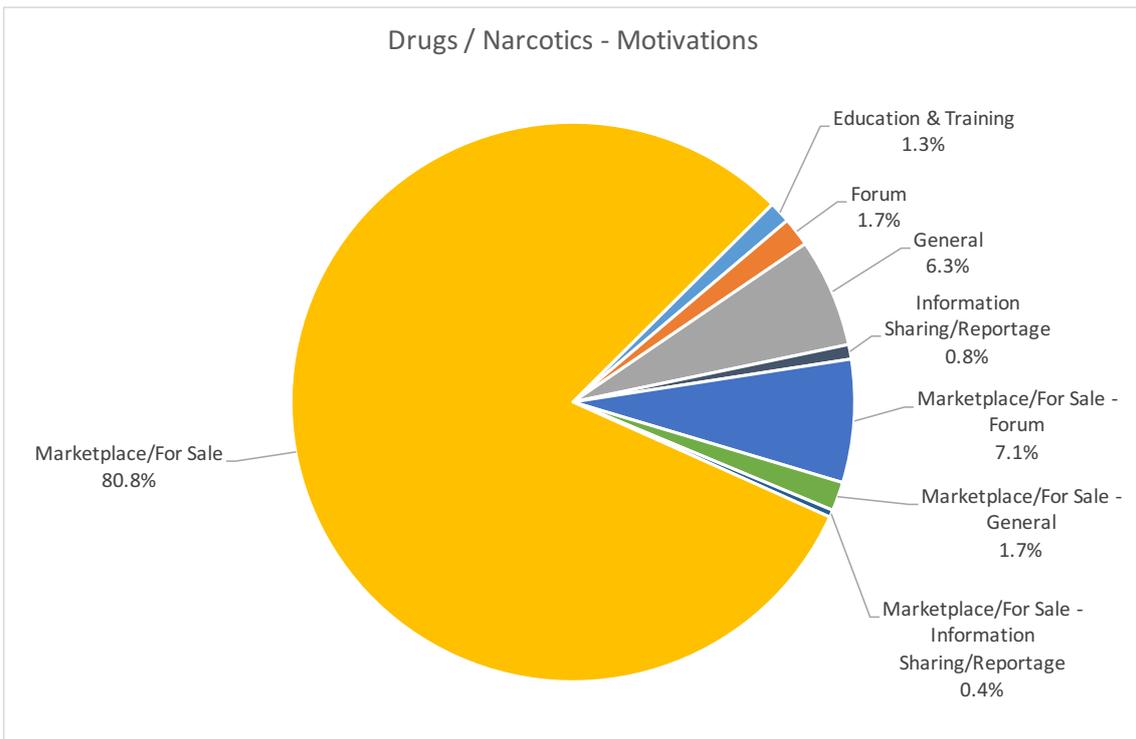


Figure 3.8: (Virtual) domains categorised as Drugs/Narcotics related - motivations

sites weren't observed to provide any materials beyond what one would regard as basic knowledge or 'common sense' (for want of a better term), the ability to anonymously train future offenders is a major issue for law enforcement.

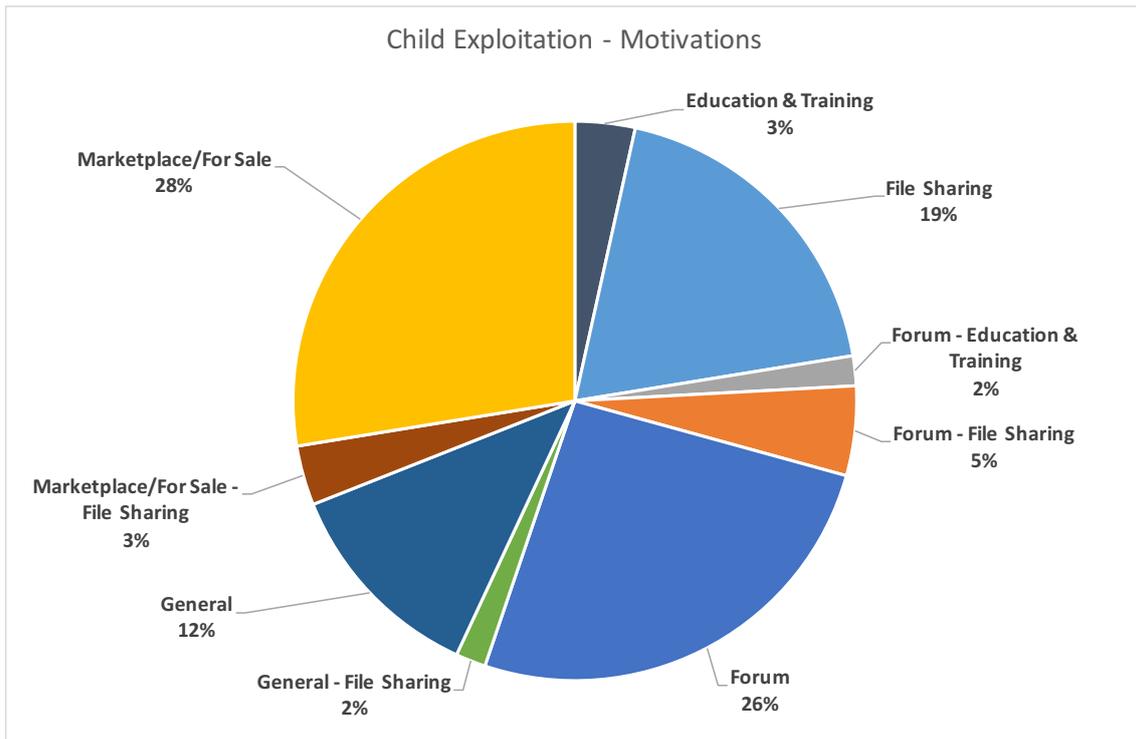


Figure 3.9: (Virtual) domains categorised as Child Exploitation related - motivations

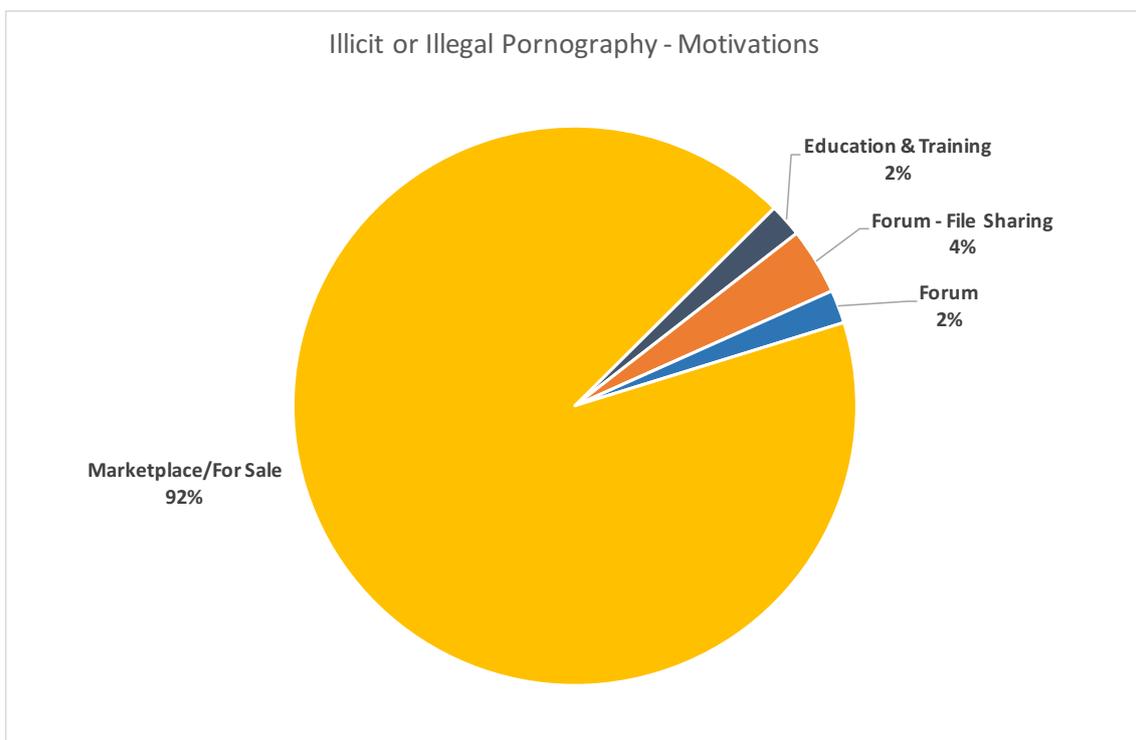


Figure 3.10: (Virtual) domains categorised as Illicit/Illegal Pornography related - motivations

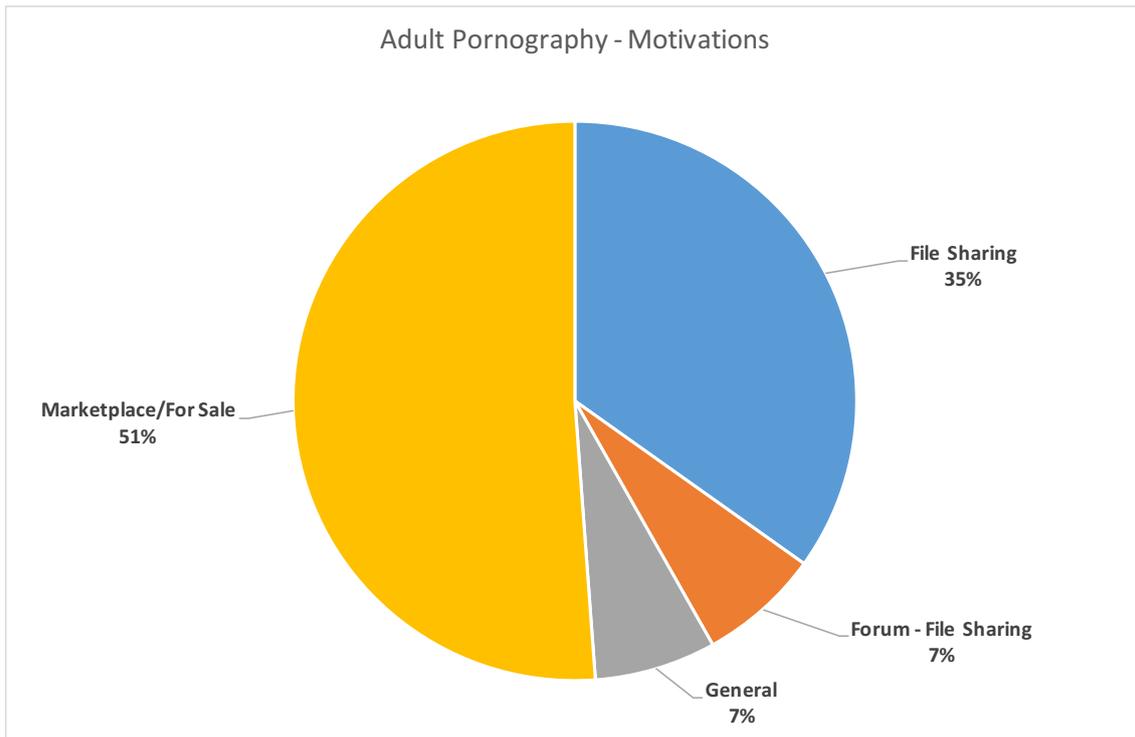


Figure 3.11: (Virtual) domains categorised as adult (i.e. legal) pornography related - motivations

3.4 Conclusions

Our development and testing of the TMM in this chapter demonstrates the limitations of re-using ontologies designed with differing goals. The TMM enabled the generation of a far more granular and objective snapshot of illicit activity online than any previous work. Schematisation from a law enforcement perspective gave the opportunity to disambiguate ‘questionable’ activity from the outright illegal, with the inclusion of motivation providing a means for simplification without loss of accuracy. This resulted in some perhaps unexpected results, particularly that approximately a third of Tor hidden sites are not of interest to law enforcement - a contradiction to the popular perception of dark webs simply being tools for enabling drug dealing, terrorism and the like. We observed a hitherto unreported prominence of illicit finance, particularly laundering services for cryptocurrencies such as Bitcoin - existing schemas failing to reflect such activities by grouping them within generic ‘other’ categories.

The TMM was designed specifically to avoid possible oversights such as the one just described - efforts to annotate and eventually automatically classify without the use of a robust and appropriate schema are doomed to failure. Our results prove TMM’s robustness and flexibility, particularly through its ability to reflect unanticipated subjects such as the provision of ‘how-to’ manuals relating to the grooming and exploitation of children. Despite the name, the TMM is easily ported to other networks, including the WWW.

Chapter 4 moves from the ‘macro’ TMM to focus upon a specific crime type. We introduce the Majura Schema, a schema emphasising visible attributes and features as a

means to enable the automated identification and classification of CEM against *existing* scales of severity.

Chapter 4

Recognising & Classifying Child Exploitation Materials

“...I know it when I see it...”
(*Jacobellis v. Ohio*, 1964)

The challenges and dangers of increased workloads and exposure to offensive materials have already been discussed in section 2.5. This combination (particularly the latter) represents a clear threat to practitioner health and welfare, making it arguably the most important issue currently facing Digital Forensics (DF) within law enforcement.

An obvious mitigation strategy is to augment manual review with automated detection. Automated recognition and classification of electronic materials is hardly a novel concept - document clustering, described in Chapter 2, groups textual materials by their commonalities, effectively identifying topics. For image recognition, open source tools and commercial ‘as a service’ products such as openCV (OpenCV Team, 2018) and Microsoft Azure Computer Vision API (Microsoft, 2017), respectively, are readily available internationally with ample ‘off the shelf’ configurations for common applications.

A significant contributor to broader success within fields of research is the existence of agreed, ‘ground truth’ corpora - not only as a convenient means for developing and testing hypotheses, but also as a means for objectively comparing performance. In Section 2.6.9 we detailed the ‘dearth’ of data within Digital Forensics, where shareable corpora don’t exist beyond the Enron dataset (Klimt and Yang, 2004) (an email archive of a failed energy trading corporation) and some specialist datasets usually associated with classes of devices¹ rather than themes. There currently are no ground truth datasets in DF.

Corpora such as the *ImageNet Large Scale Visual Recognition Challenge* dataset (herein referred to as ‘ImageNet’) provide a ground truth within the field of computer vision. A collection of several million still images freely available to researchers at zero cost, the ImageNet corpus has become a de facto standard for testing image recognition tools and methodologies. The presence of such a standard has in turn supported the creation of what now is an active ecosystem of researchers with easily *measurable and comparable* outputs

¹e.g. Cell phones, files, disk images - a good source for such corpora is <http://digitalcorpora.org/>

- at the time of writing, an online search for ‘ImageNet’ (excluding citations and patents) returns 28,800 hits (*Google Scholar*, n.d.). Beyond some collections of task specific DF datasets focused upon operating systems, devices and/or relevant scenarios, no equivalent public corpus of measurable scale for topics such as Child Exploitation Material (CEM), violent extremism (herein referred to as *offensive materials*, or indeed, any broader topic of interest to DF exists.

The number of automated detection techniques and technologies currently in use across the bulk of DF is manifestly inadequate, particularly when dealing with previously unseen offensive materials. The lack of commercial incentives equivalent to other information retrieval challenges plays a part, but the absence of a ground truth dataset is arguably the greatest restriction - the only objective means to develop, test and compare performance is to assemble a one-off representative corpus and run implementations of each candidate tool or methodology. This is an inefficient approach, with health risks and legal restrictions making such experiments near impossible outside law enforcement.

In this chapter we outline the design and development of a three stage classifier for the automated recognition and classification of CEM imagery against the Child Exploitation Tracking System (CETS) scale, a measure of severity commonly used in prosecutions across Australia. We observed the classifier to struggle with most elements of CETS, due largely to the dominant use of abstract concepts (e.g. ‘sadism’) rather than specific activities or elements. We then introduce the *Majura Schema*, an age-agnostic pornography labelling schema capable of providing sufficient granularity to allow effective Machine Learning (ML) training, whilst also avoiding jurisdictionally-specific characteristics.

4.1 Ensuring Safety

As evidenced by the research discussed within Section 2.5.1, exposure to CEM is a known, acknowledged source of stress with detrimental impacts on reviewer health. The *extent* of the dangers, however, appears to have been historically underestimated. Investigators with regular, unavoidable exposure to such activities (such as members of a Joint Counter Terrorism Team (JCTT) and Joint Anti Child Exploitation Team (JACET) within Australia) are psychologically screened prior to, during and at the completion of deployment (the timing of which is organisationally, rather than member, defined). In the case of the Australian Federal Police (AFP), attendance at such counselling and screening is mandatory. ‘Exposure’ was typically defined as direct observation of offensive materials (imagery, multimedia, and to a lesser extent, textual), loosely measured as $exposure \approx \frac{quantity \times severity}{time}$. The folly of such a simple quantifier is exposed by the numerous factors such as *secondary victimhood* listed within Section 2.5.1.

All CEM corpora detailed within this chapter were constructed from annotations generated by investigators as part of normal duties, and only made accessible for this work at the conclusion of all analysis. All direct interaction with images/movies specifically for the development and testing of the automated classifier was conducted by psychologically cleared AFP personnel. In total, the number of CEM (or other offensive) images

directly viewed solely for the purposes of these experiments is less than 50 over the course of approximately one year.

Running experiments via off-premises shared infrastructure such as the CSIRO’s Bragg cluster (Ho, 2017) was seen as out of scope, given the necessary granting of root level access to administrators. As a result, we limited our experiments to an on-premises, restricted access server hosting a single, consumer grade gaming Graphics Processing Unit (GPU). Given this use of shared infrastructure and involvement of non AFP personnel in the experiments, the decision was made to minimise (if not completely remove) *any* possibility of inadvertent/intentional access to the source materials and underlying concepts. The following procedures were followed upon receipt of data, prior to upload to the processing server:

1. **Obfuscation:** Filenames were replaced with the MD5 hash of the files’ contents - many files’ names being sufficiently explicit and descriptive to cause concern with respect to the distress associated with textual content (refer Section 2.5.1); and
2. **Encryption:** Files were encrypted at rest. An integrated decryption module was developed with the training/validation software, ensuring unencrypted imagery was only ever present immediately at time of processing, and even then only within volatile memory².

Where unexpected results were observed, individual files were reviewed by the author in isolation from the remainder of the team. Feedback given was restricted to simple ‘label is correct/incorrect’, with actual content not discussed.

4.1.1 CEM Corpora

With the exception of stage 1 (the pre-trained ‘off the shelf’ classifier already provided good performance), all training and validation of the classifier was carried out using a corpus constructed from 13 cases held by AFP Digital Forensics systems at the time (February-March 2017) and annotated using the CETS annotation system (summarised in section 2.6.7 and detailed within table B.1). Given the need for annotations/labelling to have been completed by investigators, this tended to correspond with items having been seized during the final quarter of 2016. Whereas cases were drawn from geographically disparate locations (a majority of data coming from the AFP Sydney, Canberra, Melbourne and Perth offices), the risk remains that some matters may have unintentional similarities due to common sources and elements (e.g. two offenders having been members of the same sharing group). This risk was mitigated by (a) the geographical spread of cases used, (b) the removal of duplicate material during the ingestion process via the renaming of files by MD5 value (refer Section 4.1), and (c) the use of these filenames for selecting test and validation sets³. The risk of perceptually identical images with differing MD5

²Unavailable for this experiment due to the use of shared facilities, further hardening is available through encryption of the swap partition with a single-use key randomly generated at boot, as used for experiments within chapter 3

³Filenames starting with ‘e’ were placed in the validation set. Note that cryptographic digests are specifically designed to not elect or ‘hint’ at underlying data, hence making the assignment pseudorandom.

values remains, but given the quantity of images included within the dataset, this risk was perceived as acceptably low.

The test corpus is taken from an entirely separate, fully annotated case made available to the authors approximately three months after initial ‘ingestion’, containing a relatively similar distribution of CEM categories.

4.1.2 External/‘Simulated’ Corpora

The case used as the test corpus did not contain adult (i.e. lawful) pornography. The authors of Caetano et al. (2016) kindly provided permission to utilise their pornography dataset, but after several reviews by AFP members, the data was found to be unsuitable for this experiment - for want of a better term, the images depicted within the corpus didn’t appear to be ‘extreme’ enough to act as a proxy for what is being encountered within typical online child exploitation investigations within Australia. As a result, a mix of relevant imagery drawn from discussion fora and commercial websites was assembled by AFP staff and used instead. Innocent/ignorable materials typically encountered during investigations were simulated using a subset of the ImageNet corpus (Russakovsky et al., 2015).

An entirely separate ‘Tor’ corpus was generated by extracting all images gathered as part of the crawl detailed in Chapter 3. This is used for testing detection and classification techniques on what in terms of content is a completely distinct dataset, nevertheless skewed towards illegal and ‘of interest’ materials.

Figure 4.1 displays the relative counts of each corpus.

External Corpora			
Corpus	Count	Source	Comment
Adult Pornography	323383	Tumblr accounts advertising pornography (lawful)	Soft and hard core pornography observed. Some images appear to be part of sequences, resulting in ‘clean’ (in isolation) materials being included.
TorCrawl	535305	Tor - undirected crawl	Some pornography and CEM observed, but corpus not annotated.
ImageNet	649357	ImageNet Fall 2011 ⁴ (partial download)	No annotations made for this project, but dataset extensively observed and annotated elsewhere. Some imagery of male genitalia observed - appears related to medical research.

Table 4.1: External corpora unique file counts and descriptions

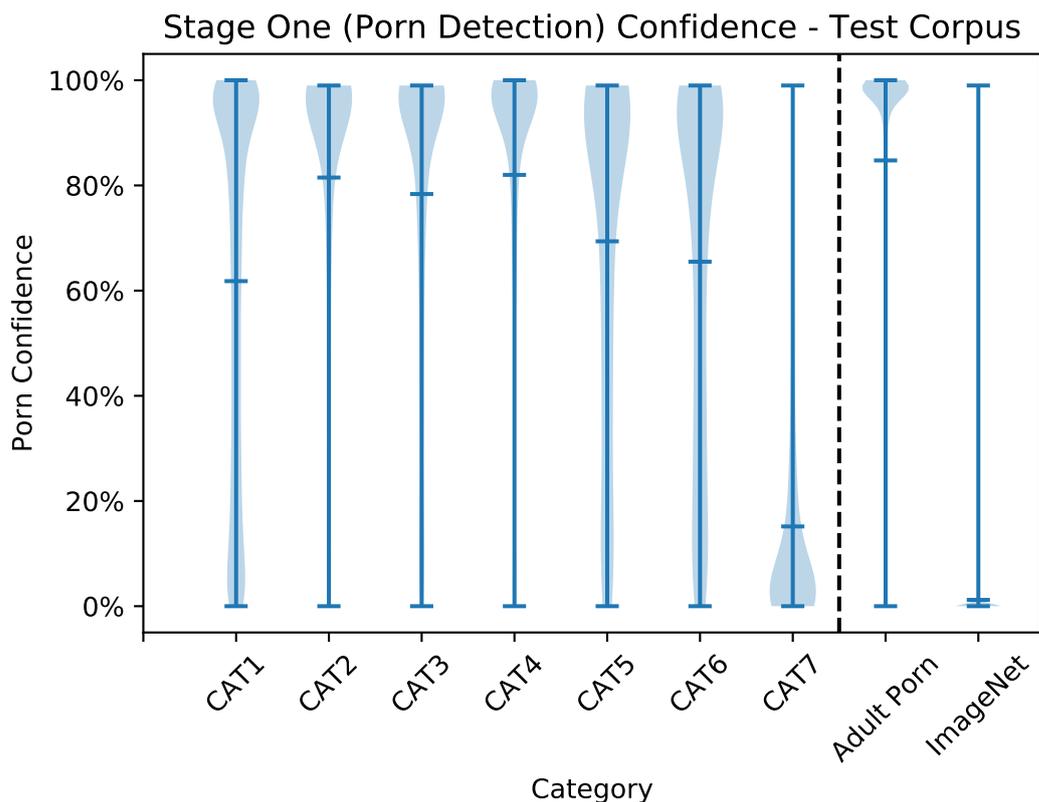


Figure 4.2: OpenNSFW pornography confidences - Test Corpus, CETS (refer Table B.1)

For legibility, discussion of the technical implementation of module two and three is contained within colorboxes. Pre-training and fine-tuning of these modules is discussed in boxes Pre-Training and Fine Tuning, respectively.

Module One: Is It pornography?

Automated detection of pornographic materials is a well-established commercial enterprise, with myriad products readily available for use in applications such as e-mail filtering. We therefore chose to evaluate an existing product for use in this stage.

OpenNSFW (Mahadeokar et al., 2016) is an open-source Caffe (Jia et al., 2014) based classifier for automatically detecting Not Safe For Work (NSFW) imagery, and due to its technical similarities with the intended architecture, was selected as the first candidate. A detailed discussion of the classifier’s design and training is available at https://github.com/yahoo/open_nsfw. The existing classifier was converted to a tensorflow model using the Caffe to Tensorflow convertor (Dasgupta, 2017).

Figure 4.2 summarises the confidences reported by the classifier across the test corpus - the strong performance in disambiguating pornography and CEM from innocent materials made it an ideal first step in culling ‘not of interest’ materials from the process.

Any imagery identified as pornographic (confidence score ≥ 0.8 , as per the authors’ advice) is passed to module two.

Module Two: Are There Children Present?

Module two is designed for detecting children within CEM. A binary child detector classifier was trained using $\frac{15}{16}$ (selected using the first character of each image’s MD5 digest) of the training corpus for applicable CETS categories, adopting a VGG ConvNet architecture (Simonyan and Zisserman, 2014).

Module Two Implementation

As with the discussion by Chollet (2016), a VGG-16 network pre-trained on the ImageNet 1000 class dataset was taken, the top stack of fully connected layers removed, and replaced with a fresh (untrained) 3-layer fully connected binary classifier. Figure 4.3 displays the original and fine tuned VGG-16 architecture. The first two fully connected layers have 512 units each with ReLU activations, and the third layer has two units (one for $isChild = True$ class, the other for $isChild = False$ class) with a softmax activation. The loss function of the classifier is binary cross-entropy, optimised on a labelled training set of images with and without children (extracted from adult pornography/CEM). A dropout with $p = 0.5$ is applied to the input of the 1st fully connected layer during training (but not during evaluation or image scoring).

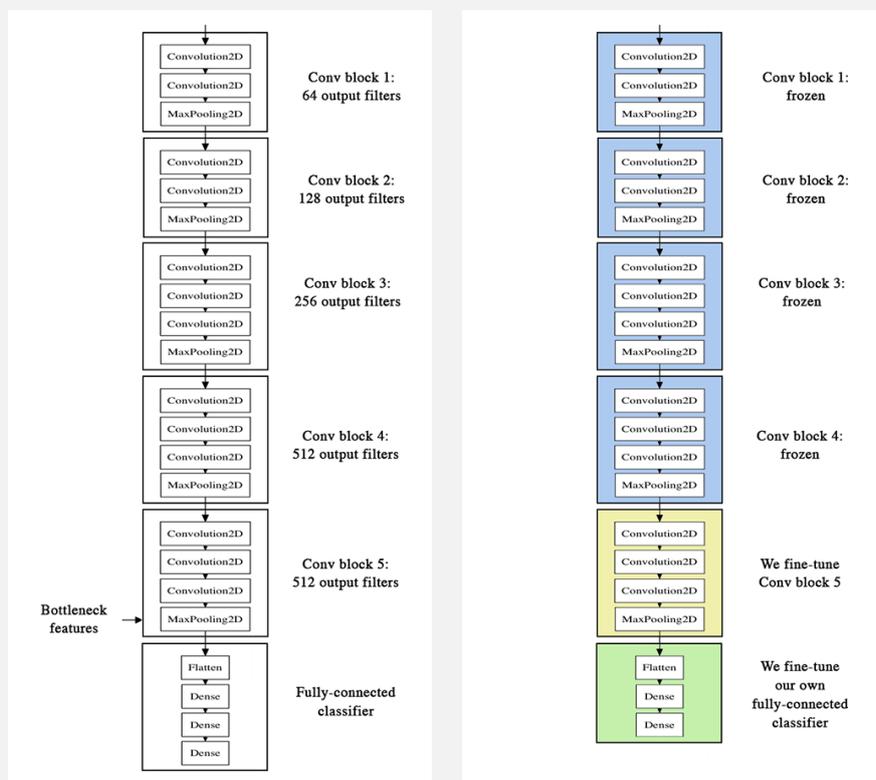


Figure 4.3: Original VGG-16 architecture (left), and fine-tuning to the relevant dataset (right)(Chollet, 2016)

As with step one, an arbitrary ‘isChild’ confidence score ≥ 0.8 was selected, with its performance observed to be adequate. In a search/triage situation, the workflow can cease here. Otherwise, all images meeting this threshold are subjected to step three.

Module Three: What CEM Category?

The third module determines the CETS⁵ category, reflecting the current workflow in use during typical online child exploitation investigations. Given the theoretically unlimited range of styles and representations, CAT6 (Animated/virtual) was regarded as ‘out of scope’ for training and testing the classifier.

Module Three Implementation

This module has a similar architecture to the previous classifier, with the only difference being in the structure of the fully connected classifier on top of the convolution layer stack. We again use a 3-layer fully connected classifier block, but the first two layers’ unit counts are doubled to 1024 each in order to give the classifier more expressive power for learning and distinguishing the largely abstract concepts present across the CETS schema. The third (top) layer has six units, reflecting CETS categories 1-5 and 7 (category 6 being excluded from this experiment), and uses softmax activation. The loss function for this classifier is weighted categorical cross-entropy, allowing to compensate for class imbalance in the training set.

The module is a multi-class classifier. At this time the category with the highest score is treated as the ‘winner’ regardless of confidence level. An obvious, simple extension may be to recognise confusion by introducing a ‘floor’ confidence - if no classes cross, the image is deemed ‘unclear’.

Although designed and implemented in complete isolation, it appears the design’s leveraging of existing pornography detection, combined with novel classifiers around elements of CEM, loosely correlates with that of Vitorino et al. (2018). Unlike their classifier, however, this classifier also disambiguates CEM categories, though admittedly with mixed success.

Why *Pornography* first?

The reasoning for placing pornography/NSFW detection as the first step in the workflow was twofold: Firstly, it replicates the typical process currently undertaken by reviewers. Secondly, as previously mentioned, the challenge of automated detection of sexual materials has already been addressed, albeit in the context of adult participants. We initially used an off the shelf solution to assess suitability, and as Figure 4.2 shows, strong performance was observed - particularly considering that to our knowledge, the model was not trained on CEM.

An alternate approach could be to place child detection as the first step, followed immediately by pornography/NSFW. This is a valid approach, with the byproduct of

⁵Refer tables 2.7 and B.1 for category summaries and full descriptions, respectively

providing the operator access to a list of files containing depictions of minors - useful information to have for victim identification purposes. We were not confident of the efficacy of this approach for practical reasons. Firstly, access to images of children *other* than those already within in our CEM dataset was limited, and as with Chatzis et al. (2016), we were unable to locate a standard corpus of such material. As a result, our best approach for training the child detection module was to use CEM as a ‘positive’ class, and all other materials as ‘negative’.

Pre-Training

As detailed in Section 4.1.3, modules two and three both consist of two stacked parts: a *feature extractor* consisting of several stacked convolutional layer blocks, producing bottleneck features, and a *classifier* block of fully connected layers. The weights of the feature extractor are initialised to the weights of the VGG-16 CNN network, pre-trained to classify images from the ImageNet 1000 classes dataset. The weights of the classifier block are initialised randomly.

During the pre-training stage, all convolutional layers in the feature extractor are frozen, with only the fully connected classifier’s weights allowed to be updated. The images fed into the model are re-scaled to 224x224 pixels with RGB channel values re-scaled by a factor 1/255 to be within [0, 1] range. The training images are then augmented via a number of random transformations such as zooming, shearing, flipping horizontal and/or vertical shifting, helping prevent overfitting by increasing variation between images ^a. The rescaled training images are fed in mini-batches of 50, augmented on the fly, and the model is pre-trained for 100 epochs using Adam optimisation (Kingma and Ba, 2014) with a learning rate of 10^{-3} . A validation set is used to estimate the out-of-sample loss during training, and early stopping is used to prevent overfitting to the training set. A snapshot of the model is saved after every 10 epochs. Once training is complete, the snapshot with the best validation loss is kept as the final model.

^aRe-scaling is conducted at time of inference/prediction, but no augmentation is carried out after the pre-training phase

Fine-Tuning

After pre-training, the weights of the top convolutional block in the feature extractor are unfrozen, and the whole model is fine-tuned for 100 more epochs with a reduced learning rate of 10^{-4} . Again, validation loss is evaluated and a snapshot of the model saved every 10 epochs, with the snapshot recording the best validation loss kept as the final model.

The validation Receiver Operating Characteristic (ROC)^a plots of both ‘isChild’ (module 2) and CETS (module 3) models after pre-training and fine-tuning steps are shown in Figure 4.4. Pre-training already yields decent classifiers, while fine-tuning results in noticeable further improvement, especially for the binary ‘isChild’ classifier.

^aA plot displaying binary classifier performance in terms of true positive (y axis) vs false positive (x axis) as the confidence threshold is tuned. The greater the area under curve, the better the performance.

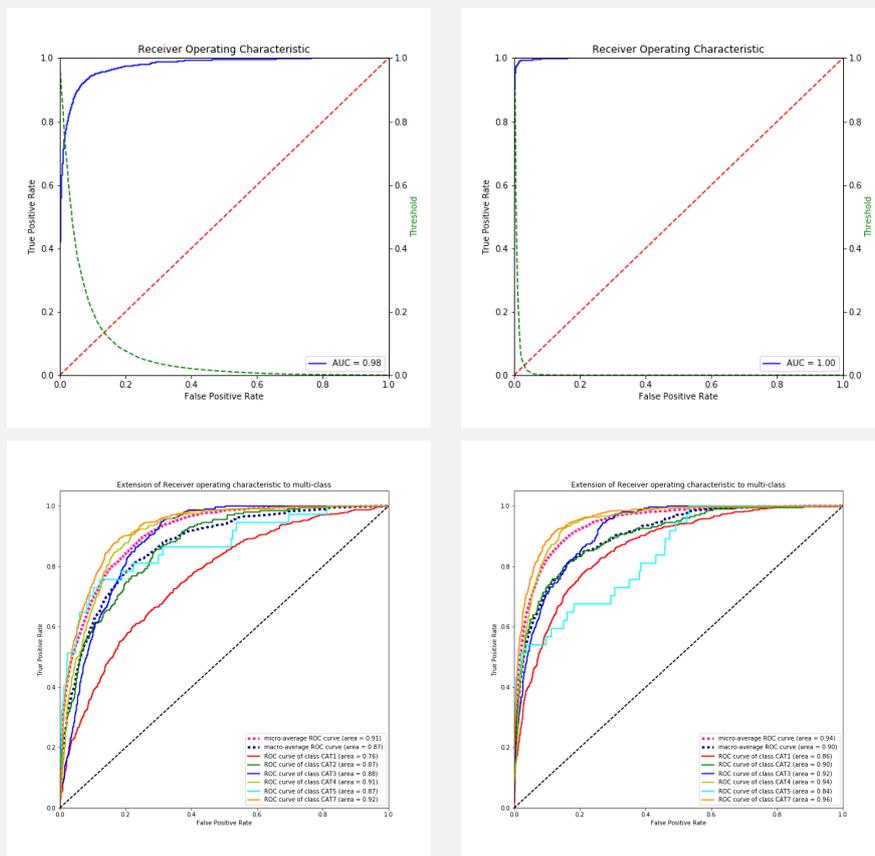


Figure 4.4: Validation set Receiver Operating Curves (ROC). Top Row: Module 2(‘isChild’). Bottom row: module 3(multi-class CETS). After pre-training (left) and fine-tuning (right).

4.1.4 Occlusion maps

The ROC curves of the trained ‘isChild’ and CETS classifiers indicate good out-of-sample performance, but one needs to make sure the features learned by the classifiers are indeed useful and generalisable to previously unseen images. As mentioned previously, it is impossible to completely understand what a classifier has learned, but a means to at least gain some understanding is required in order to detect any possible accidental features such as color palette or other superficial peculiarity common to both the training and validation sets. The risk of such accidental features is particularly significant, given our self-imposed limitation on examining the CEM corpora.

Occlusion maps (Zeiler and Fergus, 2013) are one method for gaining an understanding of what a CNN classifier has learned to use when scoring images. These are generated by systematically obscuring (occluding) different parts of an image, observing changes in the classifier’s scores. Collating these changes allows distinct areas of the image to be individually assessed for ‘value’ to the classifier. In this context we are using the term ‘occlusion map’ specifically to refer to a heat map of classification scores resulting from successively occluding parts of the image from the classifier.

Figure 4.5 demonstrates occlusion maps of benign images, in each instance featuring both an adult and a child, generated using module 2 (‘isChild’). Both faces are clearly visible, but high $isChild = True$ scores (denoted by red) correspond to the area around the child’s face, with the bulk of the adult’s face scored not significantly different to the neutral background. This indicates that at least in this instance, the classifier has learned to distinguish children from adults using facial features.

Interestingly the first image within Figure 4.5 appears to indicate well defined cheekbones as ‘youthful’, given the ‘blip’ in scoring around the adult’s left cheek. The third image appears not to obtain sufficient data from background participants in the image, indicating that perhaps the assembled dataset lacks samples of children outside the focus/foreground - a potential capability gap when considering poorly shot and/or lit photography.

Limitations

All corpora used within this chapter are based entirely upon still imagery.

As with Vitorino et al. (2018), we see the automated classification of animated/movie materials as an obvious step for expansion. ‘Movie’ materials were received, and a process of extraction (based upon every n th frame⁶) was utilised for use within training. However, this process was aborted and data not used due to the issue of labelling accuracy on a *per frame* basis. Note that multimedia files are classified/annotated according to the most extreme category observed during playback, rather than by a ratio of observed categories. Therefore, a minutes long movie file could be reasonably classified on the basis of an event taking several seconds, making disambiguation from false positives (and negatives) challenging - particularly if a low sample rate such as one frame per second is used.

⁶An attempt to use keyframes was also made, but failed due to what appeared to be codec related inconsistencies

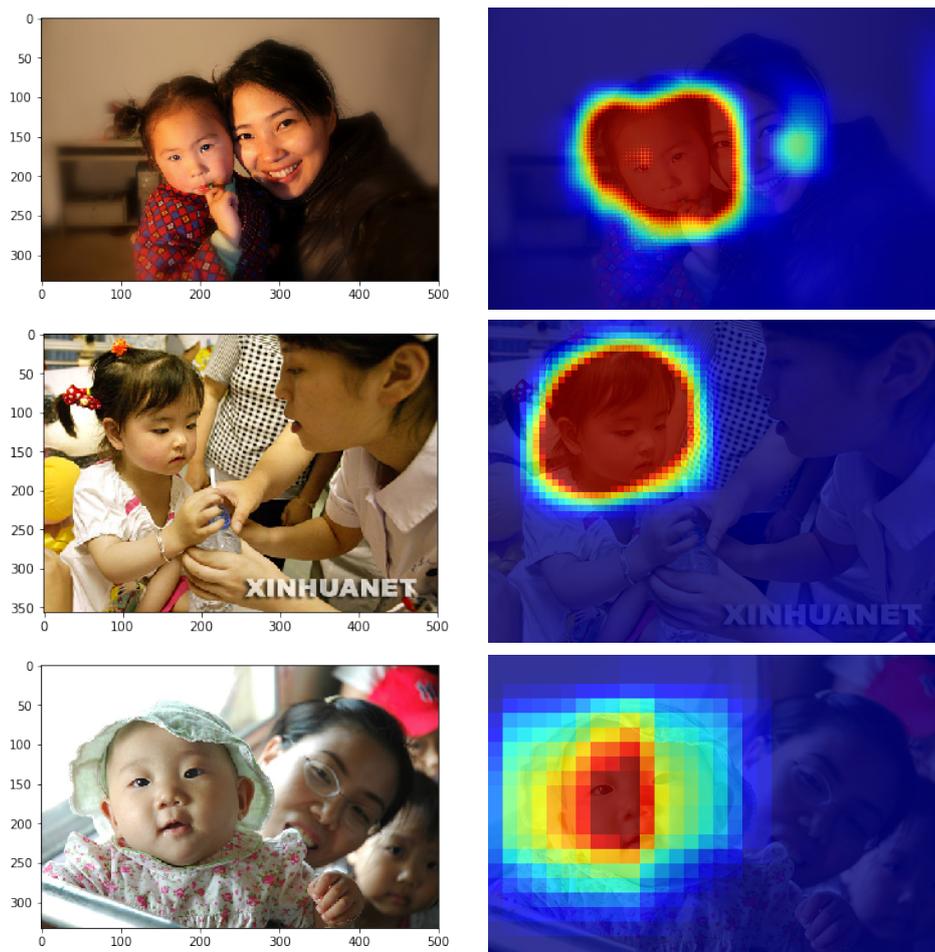


Figure 4.5: Original images and the corresponding occlusion maps for ‘isChild’ classifier. The values of $isChild = True$ scores are shown in color: red for high scores, blue for low scores. The overall $isChild = True$ scores for the images are: 0.980 for the top image, 0.997 for the middle image, and 1.0 for the bottom image

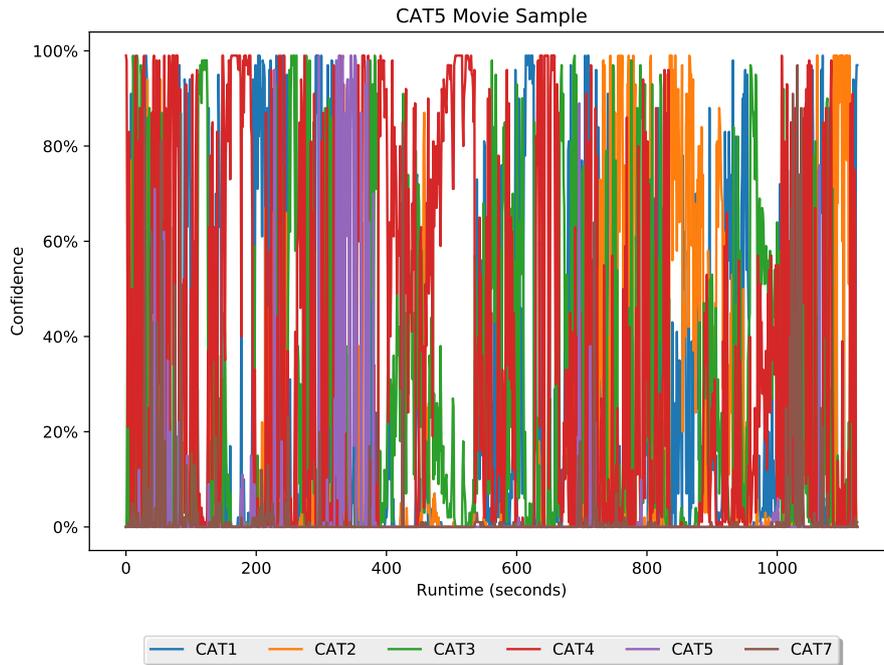


Figure 4.6: CAT5 movie example- 1 frame per second extraction/classification

By way of example, Figure 4.6 (plotted as a stacked area chart in Figure 4.7) shows a typical CAT5 film, with confidence scores plotted for every second throughout runtime. Where a frame is deemed not to contain CEM, all confidences are plotted as zero. In this instance, the ‘correct’ category dominates for less than 100 seconds of runtime (approx $300 \Rightarrow 400$ seconds, or $< 10\%$ of total frames), otherwise fading in with the ‘noise’ of indistinct categories. Such distinct sampling (typically ‘per frame / per n seconds’) is computationally slow, unreliable and wasteful due to the lack of context information being passed between frames. An approach capable of maintaining knowledge between frames (such as recurrent neural networks) would be better suited for this task, but may require specialised training/test data due to the relatively distinct domain.

Bulk manual review of materials was strictly out of bounds for this project, negating the option to generate more appropriate training data for movies.

Thus, despite what can only be described as a ‘wealth’ of data being available to researchers, due to the aforementioned infrastructure limitations, further experimentation was impracticable.

4.2 Experiments

The sheer size of the Tor and ImageNet corpora made complete manual annotation impractical. The classifier experiments were therefore split into triage scenarios (where only the top results are reviewed for CEM) and a complete scenario. Thus the models were used to classify:

1. the Tor imagery corpus for CEM content, with manual review of the top 10 results;

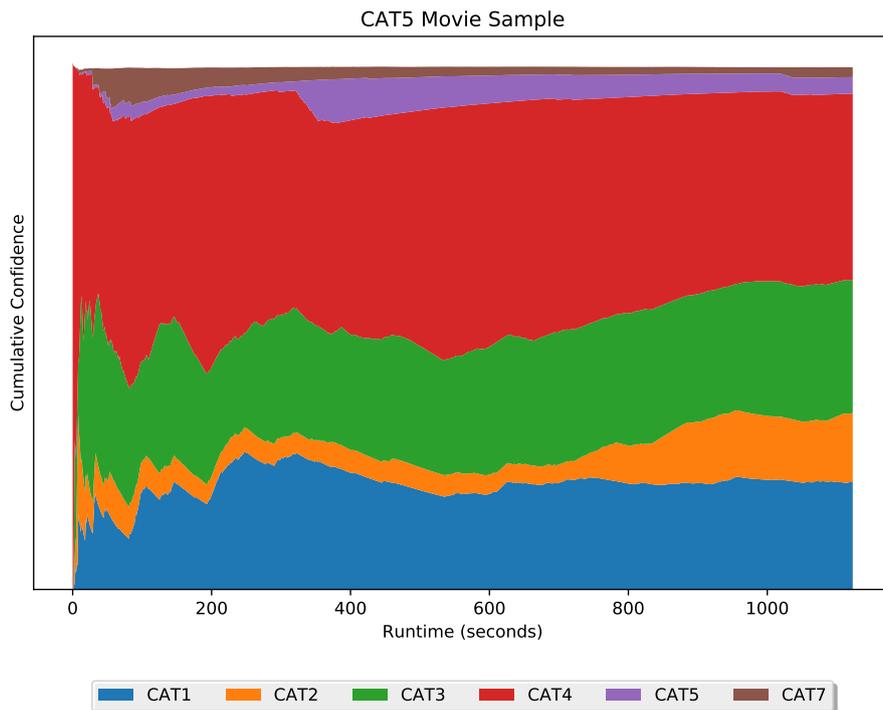


Figure 4.7: Figure 4.6 as stacked area plot. Note increased CAT5 area near 400 seconds

2. the ImageNet corpus for CEM content, also with manual review of the top 10 results; and
3. the test corpus for CEM content *and* CETS categories, providing the combined results.

4.2.1 Tor Imagery

Table 4.2 shows the top ten results, together with manual review. Where an image is either difficult to view (being a small thumbnail) or taken from an angle impossible to confidently estimate participant age, it is listed as ‘difficult’ together with the *likely* (according to the reviewer) age.

Of the ten images, all five definite CEM images are correctly identified, with the two likely CEM images also classified as CEM. One adult pornography image is misclassified (at number 10), together with two likely adult pornography images. Of the next ten images (not shown), all were sexually explicit, with one (the thirteenth image in the entire ranked set) obvious CEM.

4.2.2 ImageNet

Table 4.3 effectively lists what can go wrong when testing classifiers on a dataset that in all likelihood does not contain *any* material of the class(es) sought. No CEM or obviously *pornographic* material was observed throughout our relatively limited review of ImageNet’s content, though explicit nudity was observed. We believe these relate to medical imagery, as most such images depicted what appeared to be skin conditions such as rashes or lesions.

Image	Porn	Child	Manual Review	Result
1	0.99	1.0	CEM	✓
2	0.99	1.0	Difficult - Likely CEM	?
3	0.99	1.0	CEM	✓
4	0.99	1.0	CEM	✓
5	0.99	1.0	CEM	✓
6	0.99	1.0	Difficult - likely Adult	?
7	0.99	1.0	Difficult - likely Adult	?
8	0.99	1.0	Difficult - Likely CEM	?
9	0.99	1.0	CEM	✓
10	0.99	1.0	Adult	×

Table 4.2: Tor Triage scenario - Top 10 images, ranked by pornography and child confidence

On this test, seven incorrect results were observed, though only two could be described as ‘blatantly’ wrong. The remainder included what are best defined as ‘reasonable’ mistakes. Figure 4.8 displays examples of obviously wrong (Images 1, 6) and reasonably wrong (7, 8), respectively.



Figure 4.8: Images 1, 6, 7 and 8 of the ImageNet ‘Top 10’ (refer Table 4.3)

On the whole, however, the classifier worked well as a filter for non-CEM materials. Table 4.4 shows that Stage One classifies 0.12% ($\frac{800}{649,357}$) of images as pornography, and of these, 71% ($\frac{568}{800}$) as containing a child/children, making a CEM false positive rate of 0.09%.

4.2.3 Test Corpus

The classifier was run over the AFP sourced test corpus, with the quantities of images correctly identified as ‘passing through’ the three stages observed. Figure 4.9 (detailed in Table 4.4) shows the relative results for each category, plus the Adult pornography and ImageNet corpora representing CAT8 and CAT9 (Ignorable), respectively.

Image	Porn	Child	Manual Review	Result
1	0.99	1.0	Possible human foetus	×
2	0.99	1.0	Male genitalia (age unclear)	?
3	0.99	1.0	Arm with red sores	×
4	0.99	1.0	Maritime organism (flesh coloured)	×
5	0.99	1.0	Female genitalia (age unclear)	?
6	0.99	1.0	Birds in human hands	×
7	0.99	1.0	Human hand with rash	×
8	0.99	1.0	Frog/toad on rock (flesh coloured)	×
9	0.99	1.0	Male genitalia (likely adult)	?
10	0.99	1.0	Sores/rash on neck	×

Table 4.3: Triage scenario - Top 10 ImageNet results

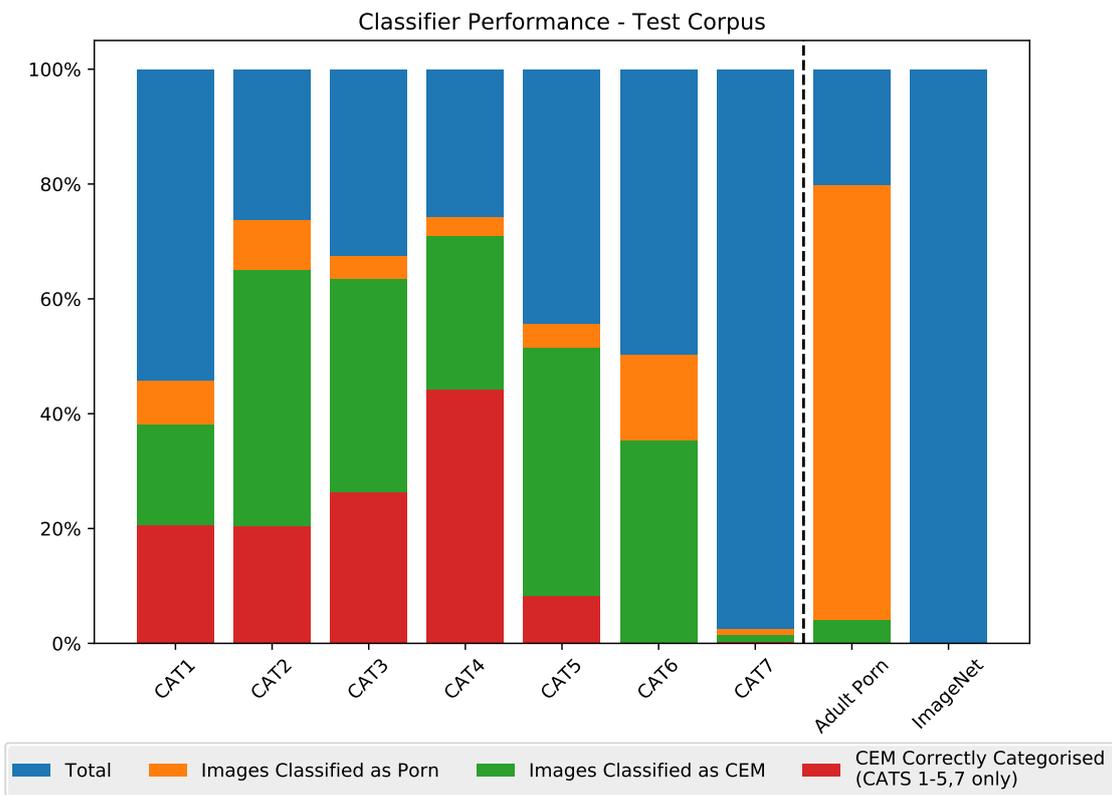


Figure 4.9: Classifier performance on test corpus. Note ‘CAT6’ category not implemented

4.3 Classifier Results

Top and bottom twenty results for the classifier and skin tone filter (for comparison) are available for each test corpus within the appendices. Whilst all CEM has been redacted and remaining materials blurred, please note these results include explicit pornographic imagery.

CAT	Total	Classified as Porn	Classified as CEM	Classified as CAT	%Porn	%CEM	%CAT
CAT1	15124	6929	5769	3098	45.8%	38.1%	20.5%
CAT2	976	719	635	199	73.7%	65.1%	20.4%
CAT3	1805	1217	1147	473	67.4%	63.5%	26.2%
CAT4	3029	2248	2149	1339	74.2%	70.9%	44.2%
CAT5	241	134	124	20	55.6%	51.5%	8.3%
CAT6	657	330	232	0	50.2%	35.3%	n/a
CAT7	790	20	12	2	2.5%	1.5%	0.3%
Adult Porn	323383	258186	13302	0	79.8%	4.1%	0.0%
ImageNet	649357	800	568	0	0.1%	0.1%	0.0%

Table 4.4: Test Corpus Classifier Results (Percentages shown are cumulative, not percentage)

Our results show that it is possible to train ConvNets to reliably identify CEM, including disambiguation from adult pornography. Module one of the classifier identified around 60%-70% of CATs 2-5 as pornography - adequate for triage purposes, but materially less than the approximately 80% of adult pornography correctly classified. Whilst the data used for training OpenNSFW has never been publicised, it is clear that CEM and ‘extreme’ materials such as bestiality and sadomasochism were either underutilised or not utilised at all⁷, resulting in a corresponding underperformance - particularly with category five materials. Category seven’s near-complete ignorance by the classifier (2.5% correctly classified compared to 50% random chance) is more indicative of the category’s conflicting attributes than any issue with module one, due to its existence as *indicative* rather than illegal, but not ignorable (another category on the scale).

Module two performed very strongly, in turn identifying around 80% of such files as containing children. Figure 4.10 displays the ‘isChild’ confidences of images passed through the module, against their actual category. The module is notably quite strongly confident on most categories, with the only vagueness really coming from category seven (the aforementioned ‘contradictory’ category), which only had a very small sample of twenty images. The images recorded against ImageNet are false positives, largely dominated by what appears to be medical-related imagery of male genitalia interspersed with images such as those already displayed in figure 4.8.

Our selection of 0.8 as the threshold for modules one and two is validated by the results, with the ROC plots in figure 4.11 showing optimal thresholds near that value, consistent with the validation set results in Fine Tuning.

Taking a different approach, the classifier was very effective at filtering out lawful materials. The combination of modules one and two results in around 4% ($\frac{13302}{323383}$) of lawful pornographic materials being misclassified as CEM, and a negligible number of false positives (0.09%, or $\frac{568}{649357}$) arise from ImageNet’s ‘clean’ imagery. Figure 4.10 demonstrates the strength of module two’s child detection, with the vast bulk of materials showing very strong confidence in the absence of children.

⁷This is not a criticism of the authors. We emphasise this absence is entirely understandable and reasonable!

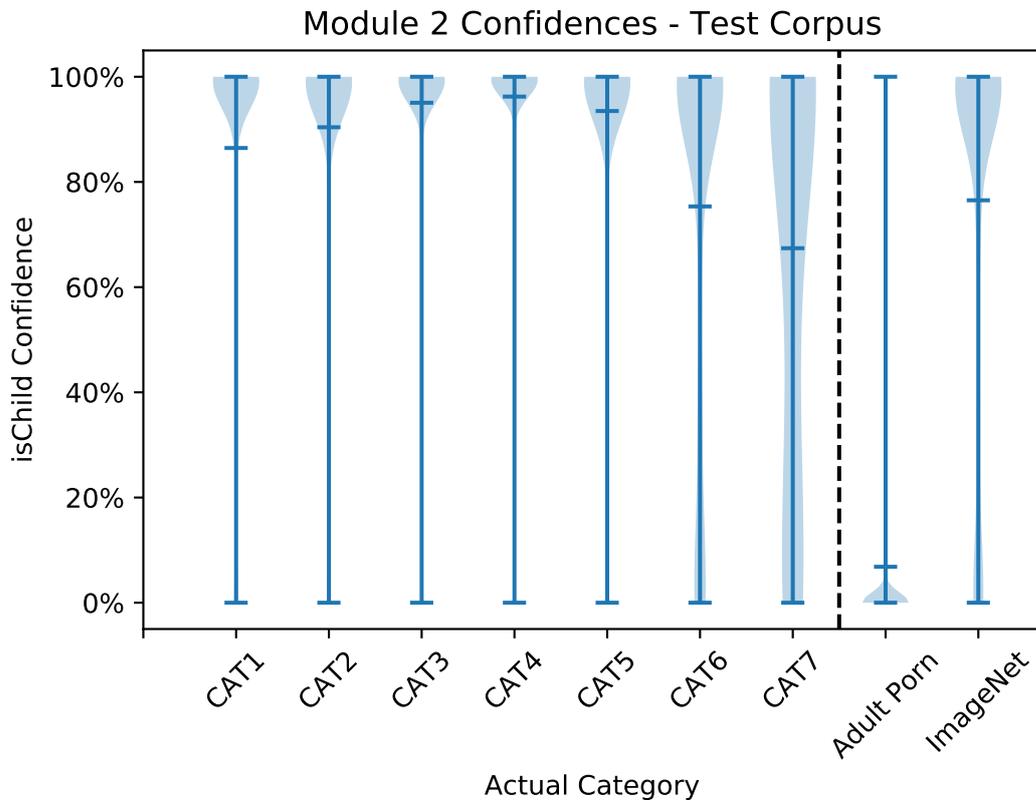


Figure 4.10: Module Two confidences for images passed from module one (isPorn confidence ≥ 0.8)

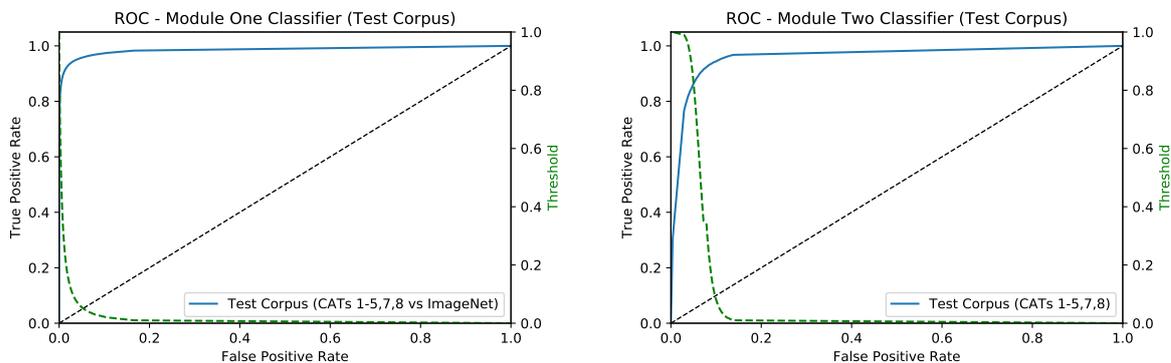


Figure 4.11: Test corpus receiver operating characteristic plots (ROCs). Left: Module 1 ('isPorn'/OpenNSFW). Right: Module 2 (isChild)

Categorisation of images against CETS is problematic. Outperforming random chance ($\frac{1}{6}$, given six implemented CETS classes), overall performance remains far too poor to be deemed ‘acceptable’ as a stand-alone reliable classifier in a legal context. Performance with CAT7 in particular proved disappointing, with the hierarchical nature of the classifier resulting in errors propagating through all stages. CAT1’s underperformance in module one demonstrates the propagation issue, with module three typically being around 50% confident around correct classification (refer figure 4.12), but only receiving a relatively small proportion of materials. Possible reasons for under-performance within specific categories are detailed further in Section 4.4.

Contrastingly, the classifier showed good performance on CAT4 imagery, far better than other categories. The best performer in module one (refer figure 4.2) and consistent in module two (figure 4.10), CAT4 is module three’s strongest performing category, with a majority of images *correctly* classified with confidence beyond 50%. Significantly, this is arguably the most strictly defined CETS category, requiring only the presence of sexual penetration - a clearly definable concept. Figure 4.13 displays module three’s occlusion maps for the first image (sorted by SHA-1 value) from each CETS category in the test corpus. Of all examples, the CAT4 image draws the most value from a tightly focused section of the image, with the remainder being of very little (or no) value. Contrastingly, CAT2 obtains very little information from the image in totality, with only a slight increase in value around the upper center. Most remaining categories rely upon the near entirety of the image - in the case of CAT7, erroneously so (the module misclassifying the image as CAT3).

4.4 Limitations of Existing Schemas

Note: In-depth discussion of CETS categories is limited, given the particularly offensive nature of these concepts.

Whilst difficult to quantify, a reasonable hypothesis is that the largely abstract nature of CETS categories greatly increases the complexity of training effective machine learning tools. CAT4, being reliant solely upon sexual penetration, is the most objective illegal CETS category. Whilst impossible to measure, we posit that the classifier’s results (refer Figure 4.9) loosely reflect the level of objectivity - CAT4 being the most ‘objective’ (relying solely upon the presence of sexual penetration), and CATS 1 and 7 being the least. CAT7 can include individually lawful images occurring as part of series containing CEM, making it impossible to accurately categorise without broader context.

CAT1 is less broad, but can still include ‘sexualised’ or suggestive imagery - examples of which may appear socially acceptable when depicting adults, but offensive (if not outright illegal) with children.

Whilst not as broad as CAT7, the severity of offending inherent in CAT5 makes its rather broad remit particularly challenging. Beyond the presence of children, there really aren’t any ‘common’ visual elements across the category.

CAT6, being focused solely on virtual/animated materials, effectively forces an overlap with other categories. Not a concern at time of CETS’ inception, this has potential to

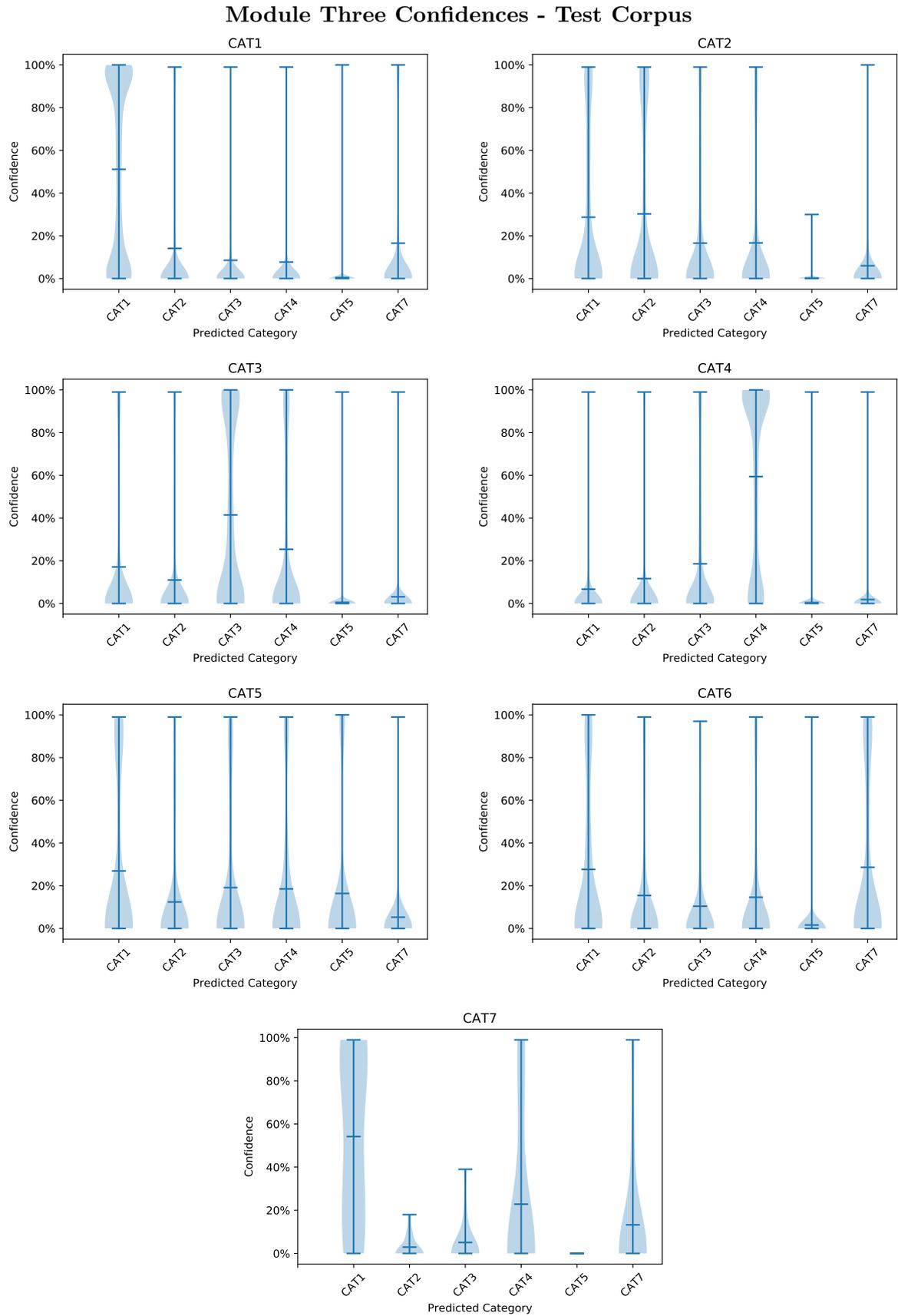


Figure 4.12: Module Three Classification Confidences - Per Test Corpus Category

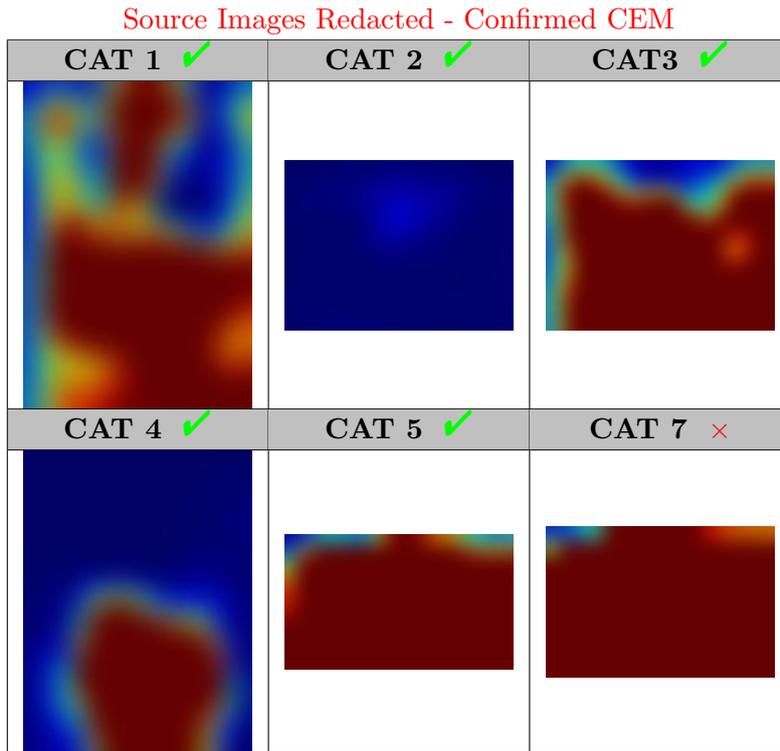


Figure 4.13: Module 3 (CETS) sample occlusion maps. ✓ denotes correct classification.

become an issue as the quality and realism of CGI renderings improves. Whilst not thought to be present in material quantities within the test corpus, we have observed some CGI CEM within the training corpus of a quality sufficient to be initially mistaken as real-world at first glance.

This move towards visual ambiguity will eventually result in CAT6 largely becoming redundant. Whereas the reason for differentiation is understandable (‘real’ vs ‘simulated’ victims), this distinction will be further muddled by the emergence of ‘deepfake’ materials - deep learning based software used to ‘learn’ a target’s face and use it to replace existing actor(s) in real footage. A particular reported use is that of creating simulated celebrity pornography, such as that shown in Figure 4.14.

Classifier Performance and CETS Limitations - A Summary

Category	Summary
CAT1	Problematic under this architecture, at least whilst using off-the-shelf pornography/‘NSFW’ detection. This category can include sexually suggestive posing, which does <i>not</i> appear to be a focus for products such as OpenNSFW. Whilst nudity is detectable in adults (as shown in the adult pornography corpus), the module underperforms with children. This is most likely due to the OpenNSFW classifier <i>not</i> having been trained on any underage materials.
CAT2	Very good performance modules one and two (second only to CAT4), but module three struggles to disambiguate with CAT1 (refer figure 4.12), only correctly classifying approximately a third of examples seen (refer table 4.4.

Category <i>cont.</i>	Summary <i>cont.</i>
CAT3	Despite very strong (94% correct) performance by module two, performance on this category suffered due to middling performance by modules one (approx. 67%) and three (approx. 41%). Module three does show promise, however, providing performance well in excess of random selection ($\frac{1}{6} \approx 17\%$).
CAT4	Best performing category for this classifier. As shown in figure 4.12, performance in module three is particularly strong, with a majority of images showing extremely strong confidences. We speculate this is due to the non-abstract nature of the class (being focused upon a physically depictable act).
CAT5	Arguably the most problematic category in terms of breadth and visual ‘uniqueness’. Underperforms in module one, most likely due to an absence of ‘extreme’ porn such as sado-masochism and bestiality in the OpenNSFW training corpora. Module two performed quite strongly (approx. 92% correct), but module three tended to slightly trend towards CAT1 - possibly due to a visual overlap between ‘solo urination’ (CAT1) and ‘bodily fluids’ (CAT5).
CAT6	Surprisingly accurate results, considering this category concerns only anime/virtual materials. Whereas module one’s performance is consistent with random selection, module two shows stronger performance (though still less than CATs1-5), correctly selecting approximately two thirds of materials encountered. It is impossible to measure performance beyond this stage, due to the absence of a CAT6 classifier under the current architecture.
CAT7	Extremely problematic under this architecture. An image is ‘indicative’ and non-illegal, but also <i>of interest</i> due to the presence of an ‘ignorable’ elsewhere in CETS (CAT9 ⁸). The category itself is contradictory when used in isolation, and is best suited for identifying co-located materials.

Table 4.5: CETS category limitations and advantages - a summary.

4.5 Towards a ‘Base’ for Cooperation

Franqueira et al. (2017) recognise the absence of a “recognised scale of indecency levels and a taxonomy of terms” as a challenge in the investigation of online child exploitation. We concur - building training/validation/test corpora around individual jurisdictions’ definitions is wasteful, with the unfortunate side-effect of actively *discouraging* collaboration. An alignment of international jurisdictions’ definition of child exploitation is unlikely within the foreseeable future, but as alluded to in section 2.6.7, an agreeable *taxonomy* of the relevant components (i.e. pornography) seems readily achievable.

⁸Not required within this classifier architecture)



Figure 4.14: Still image taken from 'Deepfake' Katy Perry pornography video (Quach, 2018). *Image redacted for ethical reasons*

4.5.1 Defining Child Exploitation Imagery

A key challenge in establishing a *lingua Franca* of child exploitation is its reliance upon defining legislation.

The 'vagueness' of legislation such as that described in Section 2.6 is in direct response to the unpredictable nature of offender tastes, proclivities and methodologies - codifying specific behaviours runs the risk of unintentional consequences, such as loopholes. Typically, law enforcement agencies use varying scales to quantify materials identified and their respective severity. Table B.1 (page 163) displays the full CETS scale, as used by the AFP in online child exploitation investigations - of note, most categories are quite broad in terms of activities *capable* of being depicted, with the exception of CAT4 - penetration being a narrow, definitive concept when compared with descriptors such as 'suggestive' posing.

The classifier's performance largely reflects this fact. Occlusion maps of CAT4 images indicated a heavy (if not complete) focus upon regions depicting sexual penetration, whilst other categories would perhaps score genitalia, breasts (or lack thereof) as a relevant characteristic.

4.5.2 Building a *Concrete* Taxonomy

A key issue in developing the classifier was the generation of an adequately sized, representative corpus of materials. Considering the only difference between 'adult' and 'child' pornography is participant age, we made the conscious decision to build our taxonomy around adult pornography, both for safety and practicality purposes - there is a *lot* of adult pornography available online.

An AFP DF practitioner was tasked with downloading several thousand adult pornography images he deemed approximately representative of what is typically encountered

within investigations. CEM was not used at this stage, due to the inclusion of non law-enforcement reviewers - instead, adult pornography with ‘similar’ posing, scenarios etc was used.

Four reviewers (three law enforcement, one academia) then assembled and reviewed a random selection of several hundred of the aforementioned images. A ‘round table’ was then conducted about each image, with attributes deemed relevant for law enforcement recorded and subsequently arranged into broad categories.

Of particular concern, the schema needed to meet three separate requirements:

1. The ability to be mapped into established CEM scales such as CETS;
2. Simplicity to a level allowing reliable use without *reasonable* conflicts between labellers’ annotations (i.e. different answers both being ‘right’);
3. A capability to record visually disparate participant attributes, such as race, ethnicity and gender; and
4. The avoidance of jurisdictionally specific terms and definitions - for example, a ‘child’ could not be defined due to the changing meaning of the term internationally⁹.

The recording of race, ethnicity and gender (item 3) was proposed as a quality assurance measure - recording attributes such as gender and race/ethnicity is not for direct use by any subsequent classifiers, but rather as a quality-assurance measure to help avoid inadvertent race/ethnicity or gender biases (discussed in section 2.6.5) or the gender-based equivalent.

In either case, the classification of race/ancestry proved challenging, particularly when considering the range of possible options. Whilst organisations such as the US census bureau record five possible races (United States Census Bureau, 2018)¹⁰, the Australian Bureau of Statistics takes a more granular approach in their census preparations, with 320 possible responses for ‘ancestry’ (Australian Bureau of Statistics, 2016) (‘race’ not being recorded *per se*). Obviously, such delineations are made more in response to statistical need rather than visual representation, and each approach presents dangers in oversimplification and unnecessary complexity, respectively. Due to the unreliability of recorded data (refer Section 4.5.3) race/ethnicity was dropped from the schema.

⁹The intention being to allow ‘bolt on’ taxonomy elements in such situations.

¹⁰White, Black or African American, American Indian or Alaska Native, Asian, Native Hawaiian or Other Pacific Islander

Pornography	
<i>Is the image/material pornographic?</i>	
Pornographic	Depicts nudity or other sexual concepts
<i>Suggestive</i> ¹²	<i>Depicts activity alluding to or ‘teasing’ sexual concepts, without explicit display</i>
Not Pornographic	Does not depict nudity or any other sexual/adult concepts

Table 4.6: Majura Schema - Pornography

Nudity	
<i>What are the levels of nudity visible within this image?</i>	
Nudity	Complete and/or partial nudity are visible. In this instance, ‘nudity’ is consistent with Western, corporate standards i.e. visible genitalia and/or buttocks. Visible nipples are regarded as ‘nudity’ when on breasts.
Suggestive	Clothing, revealing and/or posing of a suggestive nature. Examples include ‘side boob’, revealing cleavage, nudity with genitalia behind improvised coverage, lingerie pictures - in summary, NSFW.
Nil	No nudity visible.

Table 4.7: Majura Schema - Nudity

After numerous iterations, version 1.0 of the Majura¹¹ schema was agreed, with provisions for recording:

- pornography (table 4.6);
- nudity (table 4.7);
- penetration (table 4.8);
- BDSM (table 4.9);
- props (table 4.10);
- virtual/animation (table 4.11);
- bodily fluids (table 4.12); and
- participants (table 4.13).

4.5.3 Testing the Schema

Six labellers (three male, three female) were asked to annotate a selection of 49 images taken from the corpora detailed in Section 4.1.2. Whilst all have experience working in a law enforcement environment, only two have worked as investigators within this field. The images were selected to represent a broad spectrum of lawfully available materials, including ‘traditional’ hard-core and soft-core pornography, bestiality, ‘extreme’ pornography, parody materials and innocent/non-sexual. The images were printed in a large

¹¹Named after the AFP National Forensics facility

Penetration	
<i>Is penetration visible? Any form of penetration can be included (the nature of the item/limb performing the penetration is irrelevant)</i>	
Oral	Oral penetration- for example: penis to mouth, sex toys, props etc.
Vaginal	Vaginal penetration: penis to vagina, cunnilingus, sex toys/props
Anal	Anal penetration: penis to anus, anal cunnilingus, sex toys/props
Other	Penetration of other human/animal orifices, both ‘natural’ and ‘manufactured’ - for example, nostrils, wounds.
None	No penetration visible within image.

Table 4.8: Majura Schema - Penetration

BDSM	
<i>Violent, aggressive, derogatory or otherwise physically painful/submissive behaviour for gratification.</i>	
Bondage	The use of restraints (including weighing down of limbs) to maintain physical control of participants.
Domination	The overpowering or other control over participants, without the use of restraints. Often includes physically aggressive sexual interaction.
Sadism	The infliction of physical pain upon others for apparent sexual gratification.
Masochism	The infliction of physical pain for the recipient’s apparent sexual gratification.
None	No BDSM (or similar) present

Table 4.9: Majura Schema - BDSM

Props	
<i>Are props (i.e. mechanical or inanimate items) depicted being used in a sexual or suggestive manner.</i>	
Sex Toy	Item(s) appearing to be commercially manufactured and designed to be used in a sexual manner
Other	Items appearing to have been improvised for use as sex toys. For example: vegetables, gloves.
None	No props observed

Table 4.10: Majura Schema - Props

Virtual	
<i>Is the image/video animated, CGI or otherwise ‘simulated’?</i>	
Yes	The entire image (or the main focus) is CGI or otherwise animated. This does NOT include backgrounds (e.g. ‘green screens’) or cutaways.
No	The entire image (or the main focus) isn’t animated/simulated.

Table 4.11: Majura Schema - Virtual

Bodily Fluids	
<i>Are bodily fluids (e.g. blood, semen, spit, urine) visible?</i>	
Yes- self/non interactive	Bodily fluids are visible, but are clearly not in contact with participants, or are only present on the generating person(s).
Yes - interactive	Bodily fluids are visible either present on 'receiving' participants, or clearly en route. For example, 'facials', 'money shot', 'Bukkake'
No	No bodily fluids visible within the image.

Table 4.12: Majura Schema - Bodily Fluids

Participants	
<i>Describe the participants and their interactions. Select all that apply. Interactions needn't be penetrative (this is recorded in another topic).</i>	
Male_Female	Male(s) and female(s) visibly interacting/in contact with one another.
Female_Female	Multiple females visibly interacting/in contact with one another.
Male_Male	Male(s) visibly interacting/in contact with one another.
Male_Transgender	Male(s) and transgender person(s) (visibly inconsistent genital configuration/appearance) visibly interacting/in contact with one another.
Female_Transgender	Female(s) and transgender persons (visibly inconsistent genital configuration/appearance) visibly interacting/in contact with one another.
Female	Female(s) not visibly interacting with other persons.
Male	Male(s) not visibly interacting with other persons.
Transgender	Transgender person(s) not visibly interacting with other people. NOTE: Use the appearance of visibly inconsistent genital configuration/appearance as a guide.
Animal_Male	Animal(s) and male(s) visibly interacting/in contact with one another.
Animal_Female	Animal(s) and female(s) visibly interacting/in contact with one another.
Animal_Transgender	Animal(s) and transgender person(s) (visibly inconsistent genital configuration/appearance) visibly interacting/in contact with one another.
None	No people are visible within this image.

Table 4.13: Majura Schema: Participants

thumbnail format next to a table detailing the annotation schema, with reviewers invited to circle the attributes relevant to each. A report of the combined results is included in this thesis as Appendix C. Some inconsistencies were observed regarding definitions of ‘porn’, but significantly, recordings of race (refer Section 4.5.3) were found to be extremely inconsistent across labellers.

The labelling schema (with aforementioned changes) was then ported to a browser based application, and fourteen digital forensic practitioners were invited to annotate a selection from the full adult pornography corpus. The labellers were allowed to work within the same office and discuss images (if required), but the actual images shown to each person were randomised to avoid collaboration and ‘group think’. 3438 unique images were annotated - 3420 by individual users, 15 by two users, and one by three.

In particular, several inadvertently vague and/or difficult questions became readily apparent:

Race/Ancestry

As mentioned previously, the schema also included the option to record race/ancestry - not for direct use by any subsequent classifiers, but rather as a quality-assurance measure to help avoid inadvertent racial bias. This was dropped relatively early in the process, due to labellers’ reported difficulties confidently identifying and recording racial characteristics, particularly in participants only partially depicted in low-quality imagery.

Pornography

An immediate issue was identified in the identification of ‘pornography’ - the definition thereof being extremely difficult to objectively and consistently apply. A cluster of images consistent with sexuality or suggestiveness was observed, but not (in the labellers’ opinion) constituting porn - of the 49 images originally assessed, 14 involved labeller disagreement - 5 with 5 vs 1, and 9 with 4 vs 2 split of viewer opinion. For example, figure 4.15 is certainly suggestive, but the absence of visible genitalia or sexual posing led to strong disagreement over its interpretation as ‘porn’, particularly as the person shown is an adult male. We originally considered broadening the definition of ‘pornography’ to include Not Safe For Work (NSFW) but reviewer feedback indicated this would swing too far towards *inclusion* of borderline materials.

We instead added a ‘suggestive’ attribute within the pornography section, akin to the ‘Racy’ attribute offered by MS Azure Computer Vision API discussed in section 2.6.7. *Prima facie* this has provided annotators with a more comfortable middle ground, with the side-effect of allowing *context* to be introduced into otherwise abstract concepts - ‘racy’ involving adults being of little interest to law enforcement, but of great concern when children are involved.

‘Suggestive’ nudity remains in the schema for this version - on the whole, labeller feedback indicated a preference for the additional granularity.



Figure 4.15: Male, nude (Choi, n.d.). Is it pornography?

‘Virtual’ Imagery

Probably the most unanticipated inconsistency with the schema’s results concerned animated and virtual imagery. During the second, more thorough labelling session, users reported difficulty in declaring an image ‘virtual’ - particularly in cases where otherwise ‘real’ images contained captions, and in one unanticipated series, images containing ‘thought’ and ‘speech’ bubbles. Animated/virtual characters were observed interacting with ‘real world’ participants, in a manner not dissimilar to ‘augmented reality’ scenarios.

Originally, the schema demanded an ‘all or nothing’ approach to animation, with an image assumed to either be animated, or not. This was unable to accurately record such mixed images. As a result, the schema was updated to require the main focus (as defined by the viewer) to be animated/CG for the image to be labelled ‘virtual’.

‘Negative’ Labels

The schema includes ‘negative’ labels for all categories, in an attempt to ensure disambiguation between items not present vs. questions not asked/answers not recorded. Strong feedback was received from users regarding these labels, with their recommendations tending to be either:

1. Remove the negative, due to confusion and unnecessary labeller workload; or
2. The relevant question’s default result is ‘None/No’ - i.e., if the question is asked and the user *does not* select any options.

This was considered a reasonable request, but should be regarded as an implementation consideration rather than an integral design feature.

Feedback regarding the schema’s simplicity and completeness was positive, and the data could be comfortably annotated.

4.6 Skin Tone Analysis - A Final Critique

We have already discussed the limitations of skin tone analysis in Section 2.6.5, but previous research has been limited to anecdotal reports rather than quantified examples. Table

4.14 shows the skin tone percentages for the training and test corpora assembled for this research (refer Section 4.1.1).

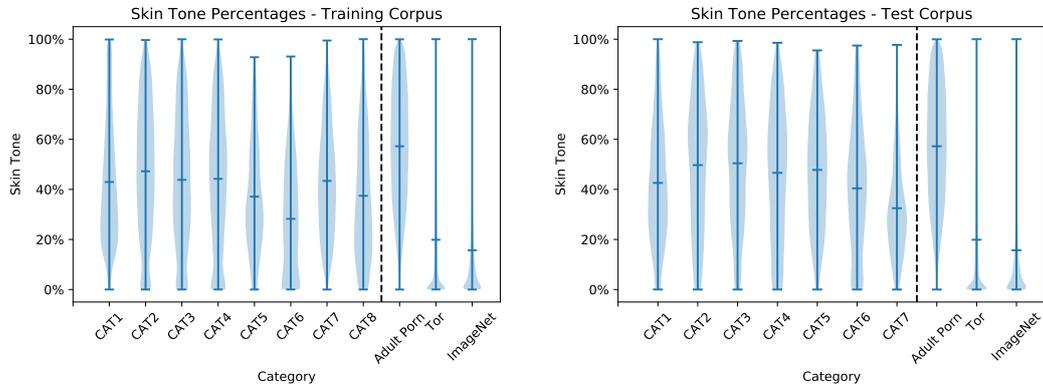


Table 4.14: Skin Tone percentages (with median) calculated using the ‘Uniform Daylight Illumination’ algorithm (Kovac et al., 2003) per CETS category (Refer tables 2.7 and B.1 for category summaries and full descriptions, respectively) vs adult porn, Tor and ImageNet - training and test corpora

The plots show that unsurprisingly, skin tone alone can’t be reliably used as a disambiguator from adult pornography, nor between CEM categories. However, they show that (a) CEM tends to involve lower skin tone percentages than adult pornography, but also (b) consistent with Vitorino et al. (2018), the more extreme categories of CEM (particularly categories 5 & 6) tend to involve less skin as a proportion of the image, but such distributions appear to be largely dictated by the downloading user. Being based on a single criminal investigation, the test corpus probably reflects an individual’s or small group’s tastes, proclivities and perhaps even methodologies (different applications and sources reflecting their users’ biases). It is entirely feasible that relative percentages of skin tone between categories will change on a case by case basis.

With regards to ImageNet, the examples of high skin-tone percentages come from what appears to be medical imagery. Table B.27 (page 263) displays the top twenty ImageNet images, sorted by skin tone percentage. Of these twenty, only ten actually display human skin, with the remainder appearing to be cellular images generated via microscopy. None depict what a reasonable person would define as pornography (or even nudity), despite each displaying 100% skin tone. Table B.31 (page 278) displays the equivalent images from the TorCrawl corpus - only one of the twenty actually shows people, and even then, only tangentially with sodium lighting lending a skin tone-esque hue to the picture. The remainder appear to be textures, simple icons and what appear to be cryptographic strings on skin tone coloured backgrounds (perhaps as part of a captcha style facility).

Table B.35 (page 292) shows that the top adult pornography corpus results, whilst close (99.96%), don’t achieve such complete domination by skin tones.

4.7 Conclusions

In this chapter we demonstrate the limitations of current research into automated detection of CEM. Long considered unreliable (with a tendency to under-report ‘extreme’ categories (Vitorino et al., 2018)), we have quantified the limitations of skin tone detection in categorising and disambiguating CEM from adult pornography. We also display the high degrees of variability between users’ collections, reflecting the impact of individual tastes upon the metric. In summary, we show skin tone detection to be a method of last resort, best kept in reserve for when other methods are unavailable.

We document the introduction of a three-stage deep learning classifier trained and validated on data from multiple, isolated, ‘real world’ criminal cases, and test its performance on an unprecedented combination of multiple, thematically distinct corpora including data from a completely separate investigation, Tor imagery, ImageNet and also adult pornography. Performance adequate for triage purposes is observed across all test corpora, particularly in CEM detection. However, poorer performance was observed during testing than during training and validation - a typical characteristic of overfitting. This approach would definitely benefit from a wider, less temporally and case specific dataset.

The three-stage architecture propagates errors. Whilst we show module two (in particular) to perform strongly on sexually explicit materials, performance suffers due to modules two and three only reviewing materials passed by prior classifiers. This makes module one (pornography detection) particularly important, and whilst OpenNSFW worked extremely well considering its ‘off the shelf’ nature, there is ample room for improvement, particularly in regards to ‘extreme’ pornography. Module two performed strongly, especially considering the limited dataset and processing infrastructure restricting further tuning.

Critically, the CETS scale was shown to be of limited use in machine learning, due largely to the abstract, context-heavy nature of most categories. CAT7 (‘indicative’ materials) is probably impossible to implement *without* providing a classifier access to co-located materials and relevant case data, both of which are out of scope for this research. CAT5 would benefit from either tighter, less ambiguous definitions and/or the re-training of module one with access to more ‘extreme’ materials.

Finally, this chapter introduces and tests the Majura schema, an age-agnostic pornography annotation schema specifically designed to support collaborative development of ML tools and techniques within a traditionally under-researched area. The schema is focused upon visible features rather than abstract concepts and measures of severity, taking its design inspiration from the strong performance of the classifier on CAT4 (sexual penetration) imagery. By providing a jurisdictionally independent ‘lingua franca’ for annotation, we provide a convenient means for researchers and law enforcement to share ‘prior work’ without denying immediate applicability to local work practices and case law. This research represents the beginnings of long-term foundations for improved data exploitation and information retrieval within law enforcement.

An obvious step beyond automated detection of individual items is the efficient search of their host media. In the next chapter we introduce Monte Carlo Filesystem Search (MCFS), a crawl strategy specifically designed for efficient search of electronic media such

as HDDs. Shown to outperform unguided methods, MCFS accelerates the search process during preservation and collection *without* supervision, effectively emulating domain knowledge without substantial memory or processing overheads.

Chapter 5

Accelerating Search

Introducing Monte-Carlo Filesystem
Search

5.1 Introduction

In chapter 4 we introduced a convolutional neural network (ConvNet) based Child Exploitation Material (CEM) detector/classifier as proof of automated content recognition's viability within Digital Forensics (DF). In this chapter, we introduce a means to implement such tools to not only reduce human exposure to offensive materials, but also as a means to accelerate search.

Investigations hinge on the ability to access relevant data in a lawful and timely fashion, with court issued search warrants used as the primary means for physically accessing, examining and seizing items as evidence (discussed in more depth within section 2.2.1). Once issued and physical access obtained, a warrant holder is faced with the challenge of examining all potentially relevant things (including electronic devices) within a target premises before being able to seize items and/or copy data. With regard to electronic media, this analysis will involve manual browsing of data (perhaps with the targeting of specific features), or the calculation of cryptographic hashes with subsequent comparisons against pre-established hash sets of known files. Both processes are resource intensive from a computational, bandwidth, and/or human perspective. The impact of inefficient search is increased due to both practical and legislative reasons:

- **Practical** - Examinations are carried out on premises, with performance heavily reliant upon suspects' hardware. Section 3G of the *Crimes Act 1914* (Cth) specifically allows for the obtaining of assistance in the form of "persons assisting" and "constables assisting", "as is reasonable and necessary in the circumstances". Ignoring availability and expense, such quantities of persons and equipment are often unworkable during search warrant execution, where anticipated workloads, safety and security can vary unpredictably, as can the quality and performance of available infrastructure.

- **Legislative** - *Crimes Act 1914* (Cth) Section 23C places strict limits (typically 4 hours) on the time allowed between arrest and laying of charges or release, making any increase in relevant information available to investigators during this time extremely valuable.

Beyond business requirements, search within investigations differs further from ‘conventional’ search, due to factors including:

- **Obfuscation** - Users will often hide or obfuscate their data, particularly in situations involving shared access and ‘public’ computers;
- **Lack of knowledge** - Investigators have limited knowledge of the data being sought, typically in the form of a third party tip-off regarding an isolated incident;
- **Lack of access** - Investigators don’t have access to the target user’s data, behaviours or devices prior to search. Index based searchers such as Spotlight (OS X) and Search/Find (Windows) rely upon the software building up knowledge of data and behaviours throughout use, rather than being ‘switched on’ whenever required; and
- **Risk/impact of loss** - Data preservation is critical within digital investigations, with the loss of evidential data unacceptable. Writeblockers can help avoid inadvertent overwriting or introduction of data, but the stability and reliability of the physical storage infrastructure is often uncertain. The ‘minimal footprint’ approach therefore extends beyond avoiding data alteration to also ensuring minimal wear on often fragile storage devices.

In the context of triage, the only way to increase search performance *without* compromising coverage (i.e. by ignoring unlikely devices or logical locations) is to improve the prioritisation of files for examination. This challenge is quite similar to that faced by search providers when indexing vast networks, and is a problem currently addressed at that scale by effective crawl strategies. Whilst we are dealing with geographically and logically smaller landscapes, the problem remains the same as that we started addressing in chapter 3: namely, *can we use our recognition of materials encountered (in this case, known ‘of interest’ and ‘ignorable’ files) to guide a more efficient search?*

5.2 Developing a File System Crawl Strategy

Initially designed and implemented for use in World-Wide Web (WWW) indexing and search, crawl strategies are a relevant means for accelerating search within file systems - a task made more urgent by growth in domestic media storage capacities. Web search is effectively file search writ large, at a scale where arbitrary walks most likely would never complete without the aid of vast infrastructure. Even if successful, the dynamic nature of content means much of the crawled data would be out of date prior to each crawl’s completion.

Unlike file system search, web crawlers can leverage an efficient linking mechanism, in the form of HTML links within pages - it standing to reason that HTML linked content is

more likely than not to be topically linked to an examined page. Local files typically don't contain such a convenient linking mechanism, with the file system structure providing the most obvious source of clustering.

On a more positive note, file systems typically follow a simple tree structure rather than the eponymous web structure of the WWW. Symbolic links ('shortcuts' on Windows based systems) and pipes can emulate the bidirectional nature of Directed Acyclic Graph (DAG)s, but such files can be easily identified and treated accordingly. Therefore, a crawler need only deal with tree structures.

A consideration for any searcher working within an investigative framework is that of minimal footprint - i.e. ensuring data preservation by avoiding unnecessary use of possibly evidentiary infrastructure, minimising wear and the corresponding risks of errors or failure. We can therefore reasonably assume that the crawler shouldn't conduct a complete sweep of a filesystem's structure prior to commencing in-depth examinations.

5.3 Monte-Carlo Tree Search

Monte Carlo Tree Search (MCTS) methods are algorithms designed to produce progressively more effective decision outcomes by successive refinement of a randomly expanding search tree. Utilised predominantly in the arena of automated game playing (e.g. *Go* being a popular target), MCTS can be applied in any domain appropriate for tree-based representation (Browne et al., 2012). Rather than mapping an entire landscape of potential moves and their eventual outcomes, sample moves in a game tree are simulated and assessed for their utility (value). Where nothing is known of particular moves' utility, selections are made at random. Once some outcomes are known, branches can be selected on the basis of expected high utility (i.e. exploitation), or through the desire to evaluate further outcomes (i.e. exploration). Once a leaf (outcome) is identified and assessed, the result is then propagated back up the tree. Each node records its visit count and overall score for child nodes, allowing subsequent node selections to occur without traversing entire branches. Importantly, this also makes the average scores of moves immediately accessible to the operator - no additional processing is required to observe nodes' values as estimated at any point in time. This also makes the approach valuable in situations where a complete examination of all outcomes is not required - the algorithm can be set to run for a defined period, with accuracy improving the longer it is allowed to continue.

Multi-Armed Bandit

A problem whereby a finite set of resources can be spent through a number of competing options with unknown or partially known attributes. These attributes can become clearer through selection, though at a cost of available resources. Named after "one armed bandits", the slang for poker/slot machines, due to the approach's similarity to that of a gambler seeking an 'optimal' combination of bets to make on a selection of such machines.

By *MCTS* we refer to the Upper Confidence Bounds for Trees (UCT) algorithm first proposed by Kocsis and Szepesvári (2006). Described as "the most popular algorithm

Weight/Bias	Summary
Exploration	‘Inquisitive’ search, where the crawler places a greater emphasis on navigating widely through the available landscape, with the potential pay-off of greater rewards in future.
Exploitation	A ‘greedy’ search, whereby the crawler immediately focuses upon an area returning positive results, remaining at that location until the value is exhausted. Potentially greater future payoffs are deferred in the interests of immediate reward.

Table 5.1: Exploration vs Exploitation

in the MCTS family” by Browne et al. (2012), this algorithm employs the *UCB1* policy (Auer et al., 2002) as the allocation strategy for selecting nodes within the tree, treating each such selection as a *multi-armed bandit* problem. Specifically, a node n is selected so as to maximise:

$$score_{UCT}(n) = \frac{sum(n)}{count(n)} + 2C_p \sqrt{\frac{2 \log(count(parent(n)))}{count(n)}}$$

where C_p is a constant used to weight exploration vs exploitation (summarised in table 5.1), $sum(n)$ is the sum of all scores achieved by previous traversals through node n , $count(n)$ is the visit count for the particular node and $parent(n)$ denotes the parent of the node. Note that the first term in the equation is simply the average result from previous iterations through the node, and the second term accounts for the variance in that estimate. Tiebreak situations are resolved through random selection.

The balancing of exploration vs exploitation within search algorithms can be a great challenge in its own right. UCB1 uses the constant C_p as a means to trade off exploration against exploitation, with Kocsis et al. (2006) identifying $C_p = \frac{1}{\sqrt{2}}$ as satisfying the Hoeffding-Azuma¹ inequality.

MCTS, combined with UCT, has been the subject of research aiming to improve its efficacy, largely in the field of two player games (Enzenberger et al., 2010; Gelly et al., 2006). Single player games have also been considered (Schadd et al., 2012). File system search can be regarded as a single player game of finding the optimal score (the target files) within the shortest clock time or iteration count, with the investigative and physical nature of the search being taken into account, namely:

- **I/O:** Where will the bottlenecks be within this search? Given the intentionally lightweight memory and processor requirements of MCTS, how will the host system’s speed impact on file examinations and scoring?

¹A concentration inequality for bounded random variables - to simplify, a means to measure the probability of future deviations from the mean/median of a random sequence - in this case, bounded between 0 and 1. This assumes a *martingale*, being a random sequence whereby the probability of future outputs is not affected by past results - akin to an unbiased roulette wheel.

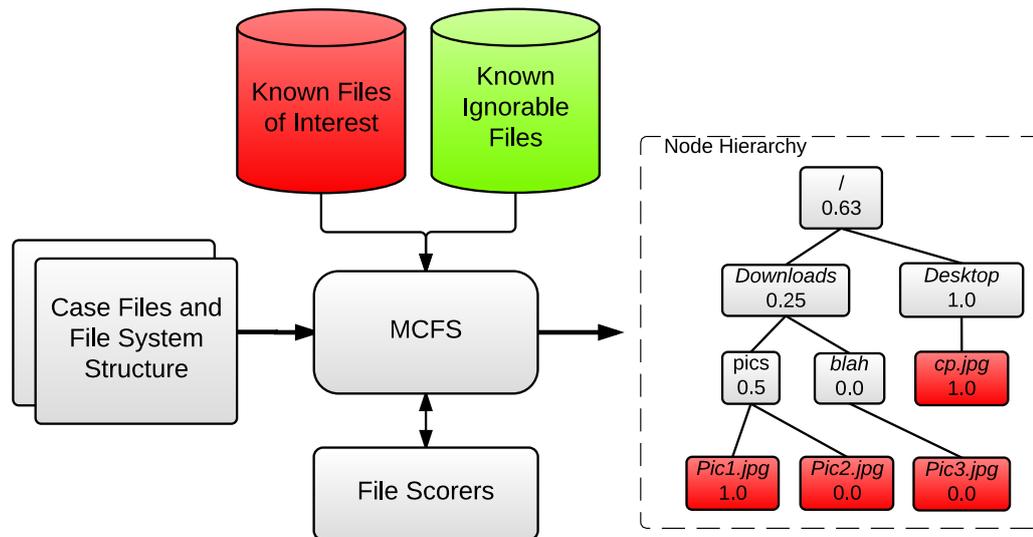


Figure 5.1: MCFS inputs and outputs - a simplified view.

- **Intention:** Is this search seeking to identify *all* items of potential investigative value, *as many* items of investigative value as possible within a set timeframe, or *any* items of investigative value in as short a period as possible?
- **Clustering:** The efficacy of MCTS relies upon similar outcomes being clustered within the tree structure, so that performing a series of “good moves” ultimately increases the percentage of successful outcomes in that part of the tree (and vice versa). Given our treatment of identifying items of interest as successful outcomes, we are operating under the assumption that at least some degree of clustering will be inherent within the candidate file system(s). This is a reasonable assumption in the situation of a hierarchical file system, given the typical use of a directory structure is precisely to ensure clustering and co-location of like information.

5.4 Monte-Carlo Filesystem Search

Monte Carlo Filesystem Search (MCFS) enhances MCTS for the file search problem through the use of *virtual branches*, inline container (archive file) expansion and both metadata and file content based scorers. Figure 5.1 summarises the MCFS approach in terms of required inputs and outputs of the system, while Algorithm 1 defines the actual search procedure used.

5.4.1 File System Structures & Removing Redundant Visits

Unlike potential moves within games such as chess and Go, extremely deep (virtually infinite) tree structures should not be assumed to occur within file systems. Most architectures support subdirectories several thousand iterations deep, but such structures will typically become unwieldy and impractical – particularly for human users. On the other hand, larger datastores and archives may require such depth as a means of maintaining manageability and compatibility across legacy systems.

Even limiting our target usage scenario to domestic devices such as PCs, tablets and mobile phones still involves a very high degree of variance in anticipated file system size, bandwidth and complexity. Individual investigators' experience with and knowledge of particular devices' usage may lead to some degree of predictability, but this has been undermined by the increasing integration of offsite/cloud-based storage offerings such as Google Drive[®] and Microsoft SkyDrive[®]. For example, the live examination of a domestic PC can currently be reasonably assumed to involve the analysis of up to an average of 2-3 storage devices' worth of data at reasonable bus speeds. However, such a PC (or indeed tablet/mobile phone) can also transparently leverage cloud-based file stores of virtually unlimited storage capacity, access to which can be made during search warrant execution if the investigator suspects that data may constitute evidential material.²

File systems can be far less symmetric than the tree representation of potential moves within the game of Go. For this reason, during search we record whether all the children of a particular node have been completely examined (i.e. all subsequent paths to a leaf node traversed) and therefore the branch is exhausted. This ensures efficiency by avoiding repeat iterations of known sequences, but would not be practicable in situations where dynamic (or extremely large) trees are anticipated.

5.4.2 Treating Content Directories Differently with Virtual Branches

A key feature of the MCTS algorithm is that unvisited nodes' values are infinite, forcing every child of a branch to be visited prior to any repeat visits being undertaken. This presents a challenge within file systems, whereby a branch (i.e. directory) can contain any ratio of files (leaves) to subdirectories (branches). The algorithm's walk will be statistically pushed to either depth or breadth-first depending upon this ratio.

Our customisation of the MCTS algorithm (refer Algorithm 1) is to create a new virtual subdirectory (termed a *virtual branch*) within each directory³ when it is first expanded, separating all subdirectories from content files. Thus we effectively move all nodes containing content files to the leaves of the directory structure. The resulting MCFS algorithm will no longer be prevented from entering a sub-branch of the filesystem tree simply because the root of that sub-branch is co-located with a large number of content files.

Once the virtual branch has been created it is treated as a standard subdirectory for all subsequent selections. We acknowledge this approach introduces potentially redundant

²Section 3L(1) *Crimes Act 1914* (Cth)

³This applies to all nodes with readily identifiable children. For example, archive files (e.g. *tar*, *zip*) are treated as branches.

Algorithm 1 Monte-Carlo Filesystem Search

input: filesystem root directory - $n_{\langle \text{root} \rangle}$
parameter: node (directory/file) scoring function - $score()$
output: score for each node (directory/file)

MAIN()

for *iteration* in 1 ... *limit* **do**

 MCFS($n_{\langle \text{root} \rangle}$)

MCFS(*node: n*)

if $\neg isBranch(n)$ **then**

exhausted(*n*) $\leftarrow true$

return $score(n)$

else

if $containsBothFiles\&\Subdirectories(n)$ **then**

 MoveFilesToNewSubDir(*n*)

candidates $\leftarrow \{i \in children(n) \mid \neg exhausted(i)\}$

if $|candidates| == 0$ **then**

exhausted(*n*) $\leftarrow true$

return *null*

else

child $\leftarrow \arg \max_{i \in candidates} score(i)$ **

val $\leftarrow MCFS(child)$

if *val* $\neq null$ **then**

sum(*n*) $+= val$

count(*n*) $++$

return *val*

MoveFilesToNewSubDir(*node: n*)

files $\leftarrow \{i \in children(n) \mid \neg isDirectory(i)\}$

subdirs $\leftarrow \{i \in children(n) \mid isDirectory(i)\}$

newdir $\leftarrow new\ directory(files)$

node $\leftarrow new\ directory(subdirs \cup \{newdir\})$

*‘Branch’ in this context refers to any node capable of containing children - for example, archive files are treated in the same manner as directories

**If maximum scores are tied then select child either randomly or using the *Tiebreak* rule

processing to the algorithm, since MCTS needn’t perform a full assessment of every node within a tree. However, the disambiguation of files and directories within most file systems is a fast process, not anticipated to materially impact performance in most cases.

5.5 Integrating Domain Knowledge and Heuristics

MCTS ostensibly appears a good method to rapidly assess a tree, but total reliance on random selection can be inefficient where readily apparent biases exist. Such “domain knowledge” is a readily available performance enhancer in guided search, particularly within specialised searches such as those within child protection investigations. An investigator with rudimentary IT knowledge will instinctively check locations known to have housed files of interest in past investigations (e.g. “Downloads” within a user’s home directory). An obvious method for improving our automated approach is to introduce such domain knowledge to the node selection process, with an eye to remaining consistent with

our desire for minimising processing and memory requirements. As detailed within Section 2.6.1, a large proportion of CEM files shared on P2P networks have very descriptive names - presumably to make them more desirable to online file sharers.

The efficacy of introducing domain knowledge within Monte-Carlo based search has already been shown by Chaslot et al. (2008), with Progressive Bias (PB) used as a means for balancing predicted against actual results. The use of a diminishing bias does give us the ability to push the crawler without risking being overrun by potentially inaccurate knowledge.

PB has three key characteristics:

- **Bias:** A bias within the range $[0,1]$ is calculated and added to each potentially selected node’s MCFS result, pushing the crawler towards pre-identified “successful” areas;
- **Progression:** The PB score is reduced as the crawler continues, thereby reducing the bias’ influence as our results tree grows;
- **Training:** Under PB pre-determined series of n iterations are used to identify potential clusters of good results, with this training used to inform the bias.

The use of a training phase within our usage scenario is both problematic and unnecessary. Firstly, such activity seemingly contradicts our stated goals of speed and efficiency, but more significantly, PB appears designed for very large landscapes such as Go, where player moves have a cascading effect on future options. Our scenario typically involves smaller, finite landscapes with independent moves, whereby incorrect selections simply delay (rather than negate) success. There are no “losing” moves *per se*, meaning a learning phase will get limited feedback.

We added two prioritiser mechanisms to MCFS, assigning scores based on domain knowledge to each node encountered during the crawl⁴:

- **Biased MCFS (MCFS-B):** MCFS-B (below) isn’t completely dissimilar to the learning phase proposed by Chaslot et al. (2008), but we chose instead to implement file metadata (i.e. directory name / file name) based similarity measures as low cost sources of bias in this work. Thus we bias the node selection score of the Upper Confidence Bounds for Trees algorithm (discussed in Section 5.3) by adding a term as follows:

$$score_{MCFS-B}(n) = score_{UCT}(n) + \frac{bias(n)}{count(n) + 1}$$

where $bias(n)$ denotes a file metadata-based similarity measure in the range $[0,1]$.

- **Boosted MCFS (MCFS-BO):** MCFS-BO is a variation on MCFS-B, whereby similarity is used to *multiply* node selection scores, rather than being added. The intention is to maintain the bias as a *ratio* of the underlying UCT algorithm.

$$score_{MCFS-BO}(n) = score_{UCT}(n) \times (bias(n) + 1)$$

⁴Unlike file scores, node scores are only used during the playout phase and don’t carry over into the base MCFS algorithm.

- **Tiebreak Only (MCFS-TB):** Whilst aiming to be lightweight, metadata analysis still imposes a computational performance penalty on the MCFS selection process. Instead of requiring calculation during every step in the process, in this scenario we limit the use of similarity scores to tiebreak situations - at least once for each node (through the parent branch's expansion phase).

5.5.1 Scoring Nodes

A challenge in treating file search as a game is how to score the results: If the searcher only wants to find a particular file, then secure hashing is reliable, ensuring false positives/negatives remain negligible. MD5 hashsets are prevalent across law enforcement, leading to the algorithm's selection as default, binary scorer.

Beyond files of interest, the crawler should be able to minimise interaction with known ignorable files - for example, commercial binaries and operating system components.

The aims of speed, efficiency and accuracy inevitably require compromise. Secure hashes such as MD5 are of established value with datasets readily available, but provide limited *similarity* values - a file either is or isn't identical to a known value. The crawler is denied any feedback until a known file (either of interest or ignorable) is encountered, reducing MCFS to a random walk. Giving the crawler some information, however minor, can inform the search, improving performance at minimal cost.

Three *file* scoring methods are proposed for use with MCFS:

- **Simple Scorer** - For this approach, known *files of interest* are given the score 1, known *ignorable files* 0, and *unknown files* 0.5⁵

$$score(f) = \begin{cases} 1 & \text{if } f \in \text{KnownOfInterest} \\ 0 & \text{if } f \in \text{KnownIgnorables} \\ 0.5 & \text{otherwise} \end{cases}$$

where f denotes a non-directory file.

- **Type Of Interest Scorer** - Scoring function granularity is increased by taking into account the *file type*. A list of *file types of interest* is generated (e.g. all multimedia files), and two weights are introduced: $\theta^{(TOI)} \in [0, 1]$ as the score for all files of those types, and $\theta^{other} < \theta^{TOI}$ for all other files. In this way, image files could for example be scored higher than text files. The value(s) assigned to the weights can be set directly by the user or optimised over a training set (as is performed within experiments detailed later within this document).

$$score(f) = \begin{cases} 1 & \text{if } f \in \text{KnownOfInterest} \\ 0 & \text{if } f \in \text{KnownIgnorables} \\ \theta^{(TOI)} & \text{if } type(f) \in \text{TypesOfInterest} \\ \theta^{(other)} & \text{otherwise} \end{cases}$$

- **Similarity-based Scorers** - Granularity of the scoring function is further increased by taking into account the similarity between *file content* and that of known files of interest. In this case the θ^{TOI} weight is replaced by the formula below, where the $\theta^{(min)}$ is the minimum score assigned to files with a *type of interest*, and the maximum similarity between the file and any file amongst the known *files of interest* is used to increase this score up to a maximum of 1:

$$\theta^{TOI}(f) = \theta^{(min)} + (1 - \theta^{(min)}) * MaxSim(f)$$

where

$$MaxSim(f) = \max_{f' \in \text{KnownOfInterest}} similarity(f, f')$$

The particular content-based similarity measure used will depend on the application domain. To date, image similarity measurements have been used within experiments.

In accordance with the proofs established by Kocsis et al. (2006), all scorers are restricted to returning values in the range [0,1].

Selection Criteria	Prioritisers	Tokenizers
Best First	{ Multinomial Naive Bayes Logistic Regression Cosine Similarity }	{ 2 to 3gram (no stemming) 2 to 4gram (no stemming) Non-word character splitter (with stemming) }
MCFS		
MCFS-B		
MCFS-TB		

Figure 5.2: Selection criteria, Prioritisers and Tokenizers

5.6 Experiments

5.6.1 Dataset

Thirty three separate electronic storage devices⁶ were identified as suitable for experiments and in keeping with the agreed conditions (listed below). All of the devices' contents were preserved, acquired and analysed by Australian Federal Police (AFP) DF Melbourne Office members during 21 investigations into possession and online trading of CEM from 2006 to 2013. All electronic media seized by the AFP are acquired as forensic images using laboratories and procedures accredited to ISO 17025, as assessed by National Association of Testing Authorities, Australia (NATA), with identification and classification of CEM undertaken by trained investigators. Internal and external review (including defence counsel rights to cross-examination) ensure a high level of accuracy in the classification of materials.

We sought and obtained access to the preserved data associated with the aforementioned media. Access to this data was subject to conditions including the following:

- **Finalised Matters Only:** Only materials relating to finalised matters (i.e. those having been dealt with by relevant court(s) and no longer subject to appeal) could be used.
- **Restricted Movement:** Materials (including CEM) couldn't leave AFP systems.
- **Restricted Access:** No persons beyond authorised AFP personnel could access the data.

An experimental harness was then established in order to conduct simulated tests using separate configurations of file scoring (refer Section 5.5.1), node scoring and the MCFS algorithm itself. Batch crawls were conducted, with items' corresponding forensic images (all in Expert Witness Format (EWF)) mounted to a CentOS 6.5 virtual machine using Libewf(libyal, n.d.). Volumes located on all active partitions (as identified using the mmls tool from The Sleuth Kit(Carrier, n.d.) were in turn mounted to the host Operating System (OS). A subsequent C# based crawler running on Windows 7 was used in tandem with MountImage Pro(GetData, 2018) for generating PhotoDNA hashes. In summary, the following data was generated or obtained during these crawls:

⁵An arbitrary figure selected for being midway between the two values

⁶As measured at the device level - i.e. a computer containing 2 hard disk drives is counted as 2 items

#	Description	# Files	# Directories ⁹	Avg File Depth	# Files of Interest
A1	External HDD	1861	228	2.77	17 (0.913%)
A2	USB memory Stick	67	1	0.01	10 (14.925%)
A3	Internal HDD	135215	3591	6.63	883 (0.653%)
A4	Internal HDD	87476	8465	9.32	58 (0.066%)
A5	Internal HDD	130318	14966	7.27	22 (0.017%)
A6	Internal HDD	71225	5687	7.72	102 (0.143%)
A7	Internal HDD	533444	78914	24.63	4 (0.001%)
A8	Internal HDD	46917	19664	15.79	9 (0.019%)
A9	External HDD	170196	6904	2.45	108 (0.063%)
A10	Internal HDD	200064	27577	7.89	467 (0.233%)
A11	USB Memory Stick	2405	131	9.22	25 (1.040%)
A12	USB Memory Stick	497	57	6.00	30 (6.036%)
A13	USB Memory Stick	2343	245	4.43	11 (0.469%)

Table 5.2: Device summary - Training Corpus

- **Cryptographic Hashes:** Message Digest 5 (MD5) and Secure Hash Algorithm 1 (SHA-1) hashes;
- **PhotoDNA:** Image similarity hashes (where applicable);
- **Archive contents:** Archive files are parsed, with all child files/directories treated as per ‘normal’ files/directories within the filesystem;
- **File types:** File content type, as identified by the Windows OS;
- **File size:** File length/size;
- **File name:** File name;
- **File path:** Path/parent node of file;

Data obtained during these crawls was merged within a MongoDB database, forming the basis for simulated crawls to be run on the resulting metadata. The resulting dataset was temporally split⁷ into a training set consisting of six external devices (two external HDDs and four USB memory sticks) and eight internal HDDs, and a test set consisting of four USB memory sticks, two external HDDs, and four internal HDDs. These sets are detailed in tables 5.2 and 5.3, respectively.

In line with the desire to maintain user-end simplicity and user friendliness, classification and segregation of devices was limited to their physical nature - external (e.g. USB memory sticks) and internal (e.g. SATA HDDs located *within* devices). This is a legacy of the item labelling system in use at that time within DF. Arguably, a more relevant demarcation in our context would be ‘bootable’ or ‘data only’, referring to the presence (or absence) of an operating system and applications, with their associated directory structures. Identification of operating systems/bootable media could be programmatically performed, but this was regarded as out of scope for these experiments, due largely to the aim of requiring minimal (if any) domain knowledge for basic operation.

⁷According to the assigned unique, incremental item identifier

⁹Not including root

#	Description	# Files	# Directories ⁹	Avg File Depth	# Files of Interest
B1	Internal HDD	1039	108	1.91	3 (0.289%)
B2	Internal HDD	2352	213	4.36	11 (0.468%)
B3	USB memory stick	5056	55	1.32	80 (1.582%)
B4	USB memory stick	998	15	1.27	1 (0.100%)
B5	USB memory stick	3720	47	1.32	79 (2.124%)
B6	Internal HDD	242618	47523	7.11	5 (0.002%)
B7	External HDD	32332	2904	4.88	17 (0.053%)
B8	USB memory stick	2507	481	2.97	4 (0.160%)
B9	External HDD	4251	59	1.64	44 (1.035%)
B10	Internal HDD	114509	14114	6.78	5 (0.004%)

Table 5.3: Device summary - Test Corpus

5.6.2 File Scoring

A historic version of the AFP Child Exploitation Tracking System (CETS) database from 2012 was used as our corpus of MD5 and PhotoDNA hashes (refer 5.5.1) of known files of interest (FOI). Beyond providing the schema detailed within Chapter 4, the broader CETS project includes a database of investigator annotations gathered during CEM investigations.

The National Software Reference Library (NSRL) v2.43(National Institute of Standards and Technology, n.d.) is an internationally used dataset of file hashes and basic metadata associated with “known, traceable software applications” - in other words, ‘ignorable’ files in all instances other than intellectual property matters. The version used contained a collection of 36,108,466 unique hashes taken from 123,298,485 files. The NSRL only includes base filenames (rather than full default installation paths), limiting its use in training domain knowledge.

We use PhotoDNA, a proprietary algorithm made available to law enforcement by Microsoft as our image similarity measure. The PhotoDNA algorithm generates hashes based upon features extracted from candidate images, making it a useful tool for identifying resized/cropped or otherwise altered versions of known images of interest. We note that similarity measurement is based upon edit distances between hashes - beyond performance considerations, a key advantage of this approach is the one-way nature of the hashing algorithm¹⁰. Unfortunately (from a research perspective), the algorithm itself appears unavailable for review or analysis by outside parties, making independent assessment difficult. PhotoDNA’s use by agencies such as Victoria Police and the AFP (the aforementioned CETS database including PhotoDNA hashes) does make it suitable within our experiments, given it is a readily implementable enhancement for existing search methodologies.

Skin tone detection is a simple method for identifying exposed skin, a common (though by no means integral) component of CEM and pornography. Whilst insufficient for reliable CEM search (the shortcomings are described in some detail within Section 2.6.5), a bias towards multiple files containing high concentrations of skin-colored pixels could at least indicate the presence of sexually explicit materials. We therefore implemented a skin

¹⁰Carrying sensitive data to target premises is ill-advised for most investigations!

Scorer	Summary
Cosine similarity	A measure of similarity between two vectors - in this case, populated by our path <i>n</i> grams. Similarity is measured according to the quantity of common non-zero entries. This approach doesn't penalize vectors of disparate non-zero lengths - crucial when one considers the larger quantity of terms in the 'files of interest' vector.
Multinomial Naive Bayes	An approach whereby the classifier constructs a probability model of trained terms - e.g. if 100% of instances involving term <i>x</i> are associated with class <i>y</i> , the model will always assign <i>x</i> against <i>y</i> in isolation. The approach is 'naive' as no attempt is made to understand correlations of terms/attributes. 'Multinomial' Naive Bayes measures the <i>frequency</i> of attributes rather than simple presence.
Logistic regression	This method predicts probabilities of classes (to be specific, the default class in a choice of two). Unlike Naive Bayes, probabilities are log transformed and attempts made to establish optimal weights for identified features. Based upon the implementation by LingPipe API - see http://alias-i.com/lingpipe/docs/api/com/aliasi/stats/LogisticRegression.html .

Table 5.4: Node Scorer algorithms explained

tone percentage scorer using the CvAdaptiveSkinDetector within OpenCVSharp (Shimat, n.d.), whereby an image's score would be boosted from the type of interest score by a ratio built on the skin tone percentage.

No convenient, off the shelf methods for achieving similarity scores or skin tone detection within movie files were identified during the design and execution of experiments around Monte Carlo Filesystem Search (MCFS), with the only viable methods identified requiring a level of processor and/or memory capacity incompatible with the stated goals of avoiding such overheads. The absence of scorer granularity for such files does represent a limitation within this work. Interestingly, the experiment results (detailed later within this work) confirm the MCFS algorithm's outperforming baseline searches even in devices with a higher proportion of movie files.

5.6.3 Node Scoring

A combined list of unique file names & paths of 191,249 manually labelled CEM files and 50,045 ignorable files was provided to us by Victoria Police (Australia), sourced from cases unrelated to the devices used for testing. We gathered a list of filenames and paths of interest to train three metadata based file prioritisers, based upon a Cosine similarity scorer, a Multinomial Naive Bayes (NB) classifier, and a logistic regression scorer (refer table 5.4). The cosine similarity and NB scorers were trained with the paths tokenized either by splitting on non-word characters (and then stemming), 2-3 and 3-4 character grams. We created a basic model of 48 patterns observed in ignorable and child exploitation filenames for use as features in creating the logistic regression scorer.

5.6.4 Test Approaches

A number of scenarios were selected with the aim of providing an holistic sample of typical search scenarios, particular to the field of forensic image search:

- **Approach 1: Uninformed Search** - Files are examined in an uninformed fashion - in this case, depth-first (akin to the Unix/POSIX *find* tool).
- **Approach 2: Informed Search** - A best-first crawl based on metadata information, as discussed within Section 2.6.1.
- **Approach 3: MCFS** - File examination ordering is directed by the MCFS algorithm, MCFS-B and MCFS-TB.

The purpose of these scenarios is twofold: to establish the value of MCFS as a search methodology, and to assess the algorithm's ability to cope with varying file system hierarchies and distributions of files of interest.

Each test was allowed to run until every item's file of interest (FOI)¹¹ was identified by the software.

5.6.5 Evaluation

Performance is measured as the number of files examined within each device (expressed as a percentage of total files within the device) prior to the first and all files of interest being located, averaged across all devices within the datasets.

5.7 Calibrating Performance on the Training Set

The performance of an approach such as MCFS relies heavily upon the tuning of inputs according to the anticipated landscape. We therefore assessed four main inputs, as detailed below:

5.7.1 Balancing Exploration vs Exploitation

As discussed in Section 5.3, C_p is used as a means for balancing the tradeoff between exploration and exploitation of known positive results. Low values emphasise exploitation, higher values exploration.

An initial trial run of all items within our training corpus was conducted using the standard MD5 scorer, with the C_p variable set in the range [0.2,2.0] with increments of 0.2. This reflects findings observed within earlier tests on simulated datasets. We also included the value $C_p = \frac{1}{\sqrt{2}} \approx 0.71$ to assess whether the settings suggested by Kocsis et al. (2006) apply within our intended landscape.

Searches across devices tended to perform best with a low (0.2) C_p value, leading us to add C_p values of 0.1 and 0.01 in an effort to identify the best performer. Figures 5.3 and 5.4 show the results, with $C_p = 0.2$ performing best when seeking the first file of interest,

¹¹A file identified within the CETS dataset as containing child pornography

and $C_p = 0.1$ for all found. $C_p = 0.1$ was observed outperforming $C_p = 0.2$ at the point of third file of interest found, leading to a number of possibilities:

- The use of weights for file types introduces a risk of excessive bias, particularly during early stages of the crawl;
- Files of interest tend to be tightly (but not necessarily absolutely) clustered. The risk of individual outliers being overlooked within individual cases remains real, but is tempered by the use of unique cryptographic hashes.¹²

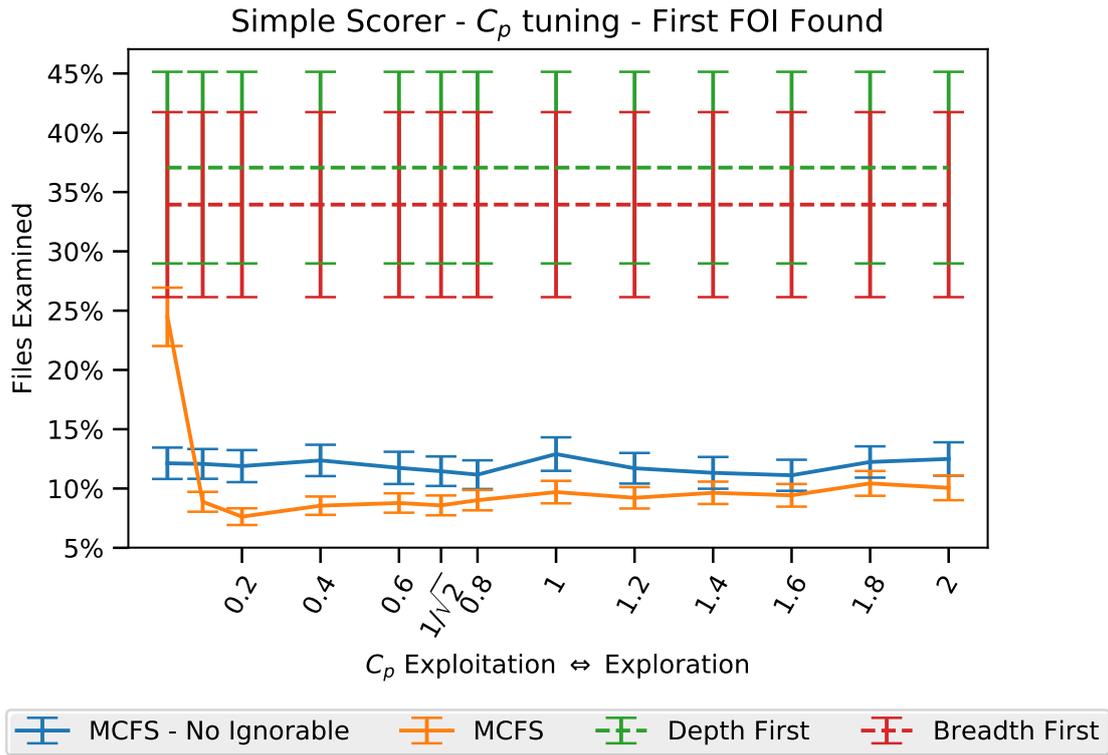


Figure 5.3: Training Corpus - Testing known ignorables and tuning exploration vs exploitation - First File of Interest Found. Refer Table D.1 (Appendix D.1) for tabulated results

On the whole, $C_p = \frac{1}{\sqrt{2}}$ proved disappointing within this context, being regularly outperformed by lower values - unsurprising when one considers the different landscape within this usage scenario. In terms of impracticality of measurement, Kocsis and Szepesvári (2006) faced an effectively infinite tree with good and bad game moves having a cascading effect through branches. In this use case, the challenge is to make selections throughout a large but measurable tree in an efficient (if not near optimal) basis.

5.7.2 Known Ignorables

The introduction of a known ignorable hashset needs to provide a performance improvement sufficient to at least offset the additional memory and processing overheads. We

¹²File classifications listed within CETS are themselves reviewed, leading us to assess the risk of biased manual classifications/labels as marginal to low

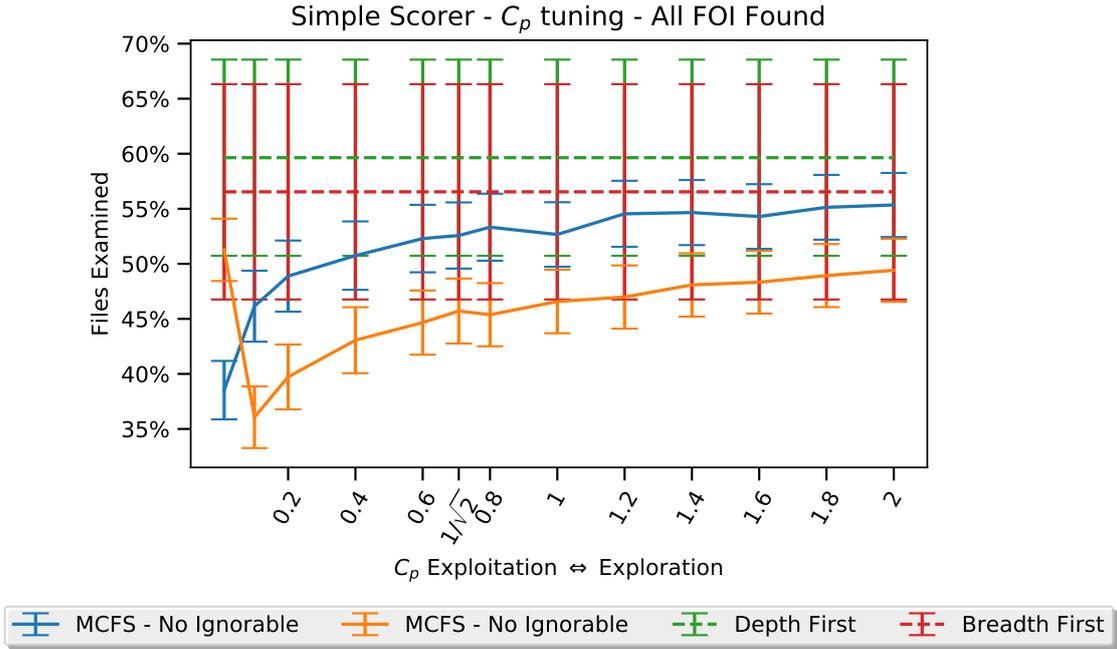


Figure 5.4: Training Corpus - Testing known ignorables and tuning exploration vs exploitation - All Files of Interest Found. Refer Table D.2 (Appendix D.1) for tabulated results

therefore conducted a set of parallel tests with our simple scorer. Figures 5.3 and 5.4 show the results, with a near consistent outperformance when the known ignorable hashset was used.

The exception to this rule occurs when C_p approaches absolute exploitation, effectively reproducing best first search by minimising the impact of visit count on node selection. A crawler strongly biased towards exploitation will be heavily influenced by nodes scored early within the crawl. The sheer size of the NSRL dataset, plus the fact that it includes common operating system files, means a crawler is more likely to encounter ignorable files than files of interest - particularly during early stages of a search, when the crawler is effectively performing a random walk.

5.7.3 Optimising File Scorer Weights for Unknown Files

The use of file types of interest (TOI) within the crawler in effect amounts to a specialised but ultimately probabilistic ‘guess’ as to a file’s value to the searcher. A search for CEM is far more likely to find value within multimedia files than it is within raw text, *vice-versa* within fraud matters.

Given our research focus on CEM, file types of interest (TOI) within this work are defined as any formats capable of being rendered as still images or video, with θ^{TOI} and θ^{other} tested in the range [0.01,0.99], in line with the file score range of [0,1]. The optimisation of θ^{TOI} and θ^{other} had a positive impact on performance, particularly in later stages of search - figure 5.5 showing a comparatively smaller spread of outperformance against the binary scorer, as against figure 5.6.

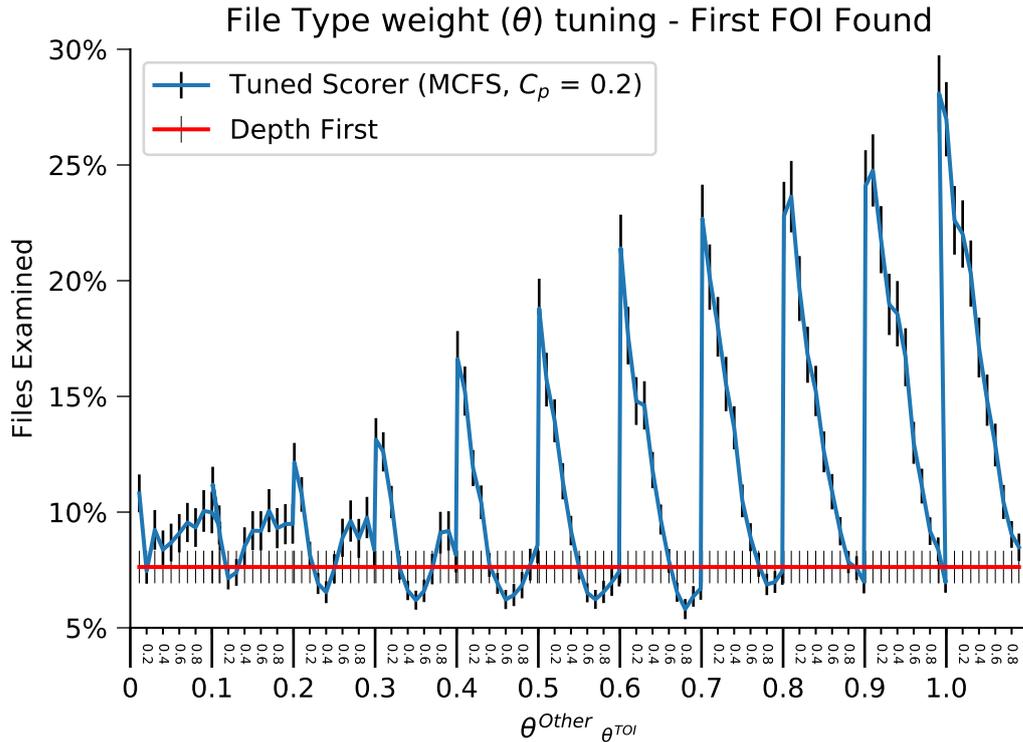


Figure 5.5: Training Corpus - File type weight tuning - First File of Interest found. x axis major ticks denote θ^{other} , minor ticks θ^{TOI} . Depth first results shown as comparison.

A definitive, ‘optimal’ θ combination couldn’t be readily identified within the training corpus, reflecting the variability of structure between devices. Generally speaking, assigning a lower-range value to θ^{other} and a mid-range value to θ^{TOI} provides the best overall performance, with performance decreasing as either θ approaches the upper or lower bounds. Figures 5.7 and 5.9 show the distribution of combinations’ performance, *generally* following the aforementioned rule. It should be noted, however, that whilst these plots do show the relative performance means for each configuration, many results are extremely similar (if not interchangeable) once standard error is taken into account, as depicted within figures 5.5 and 5.6.

Granular scorers (PhotoDNA and skin tone detection) saw similar impacts from θ tuning, though interestingly the skin tone based scorer’s performance is less impacted than that utilising PhotoDNA. This is *possibly* related to the independent nature of skin tone detection, not relying upon previously observed data for estimating unknown files’ value.

A level of noise is visible within the distributions, particularly that for first FOI found (Figure 5.7). Furthermore, a large proportion of combinations’ results fall within each another’s standard error. It is reasonable to say that these experiments would most likely benefit from a larger corpus, but the general trends and performance are visible.

For example, the best performing $\theta^{TOI} / \theta^{other}$ combination when seeking the first file of interest was using 0.8/0.6. Three other combinations’ (0.5/0.3, 0.6/0.4, 0.7/0.5) results

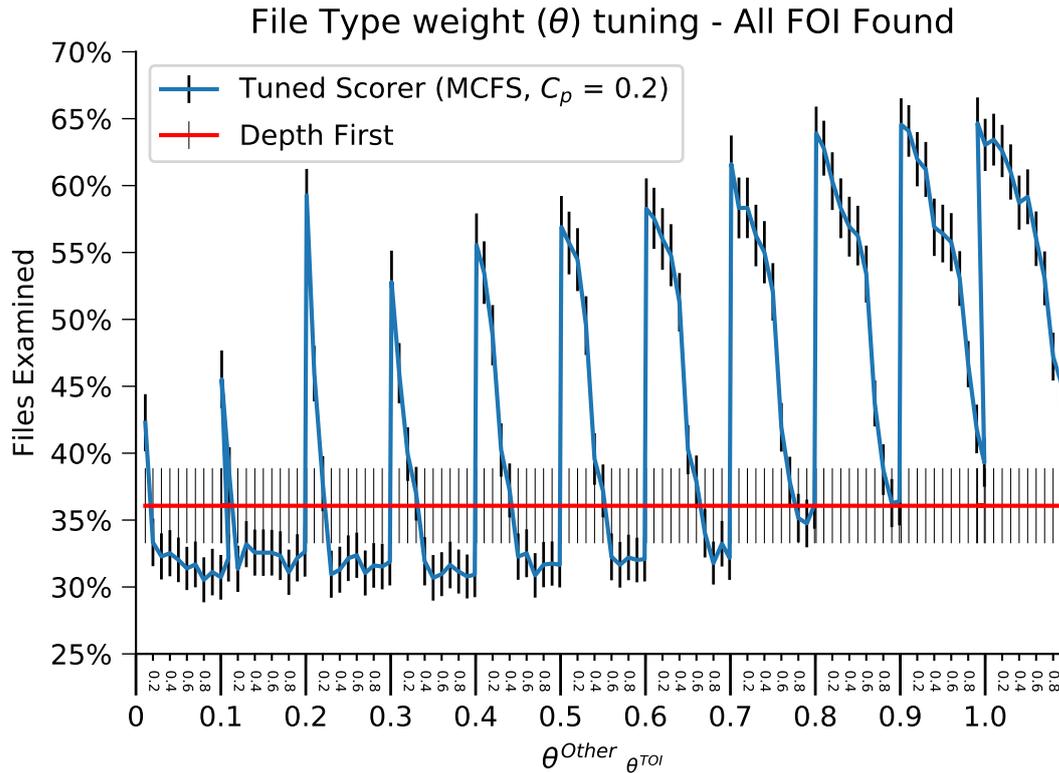


Figure 5.6: Training Corpus - File type weight tuning - All Files of Interest found. x axis major ticks denote θ^{other} , minor ticks θ^{TOI} . Depth first results shown as comparison.

fell within the standard error. The difference between weights appears more influential than the underlying values themselves.

Less definitive combinations were observed when examining performance for all FOI found. In this case, 0.9/0.3 was the best combination, but a further 32 combinations fell within the standard error.

The best performing combinations for PhotoDNA were 0.2/0.1 for first hit and 0.1/0.1 for all FOI found. As with the simple file scorer, fourteen different file type weight combinations fall within the standard error of the optimal, largely around a combination of very low default values with moderate type of interest values.

The skin tone based scorer saw best results using 0.2/0.2 both for first hit *and* all FOI found. Interestingly, two scorers within the standard error of the optimal for all FOI found contain optimal weight combinations where type of interest values are *lower* than the default (0.2/0.3 and 0.1/0.2).

We only use the optimal file type weight combinations for validation, but the repeated observation of multiple configurations falling within the standard error of optimal combinations leads to the conclusion that this aspect of our experiments would benefit from a larger corpus for testing and evaluation.

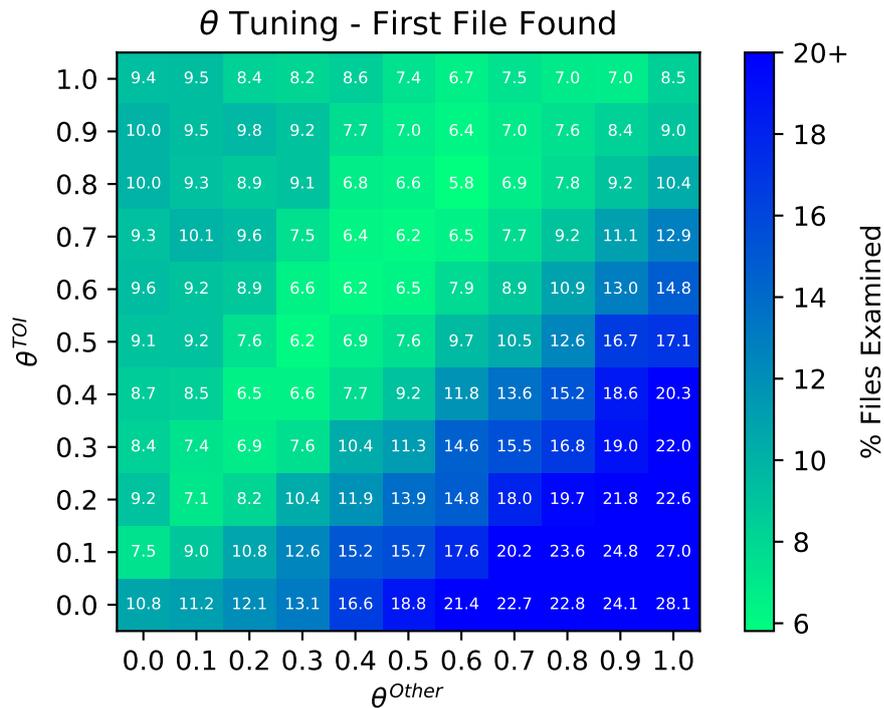


Figure 5.7: Training Corpus - File type weight tuning - First File of Interest found

5.7.4 Selecting Prioritisers for Informed Search

Cosine similarity combined with non-word character splitting performed best of our available informed search scorers. On the whole, our informed search scorers performed poorly, with only cosine similarity based scorers performing to an adequate standard. Figure 5.10 shows the results - whilst our informed search performed adequately (all FOI being found after an average of approximately 44% of files present), our MCFS based scorers easily outperformed the baseline (depth first) scorers. In fact, MCFS-B and MCFS-TB using a PhotoDNA scorer located all FOI faster than the informed search found could locate the first FOI.

5.8 Findings and Discussion

We conducted a series of experiments comparing MCFS based crawls across the test dataset (refer table 5.3), using well-established informed and uninformed searches used as benchmarks. Optimal file type weight combinations were established using a training dataset, as detailed within Section 5.7.

5.8.1 File Scorers

Figure 5.11 displays our optimal searchers' performance across the test corpus, with results shown for first hit and all found. All scorers easily outperform depth-first search - by at least two thirds across internal devices, and at least around one third on external devices.

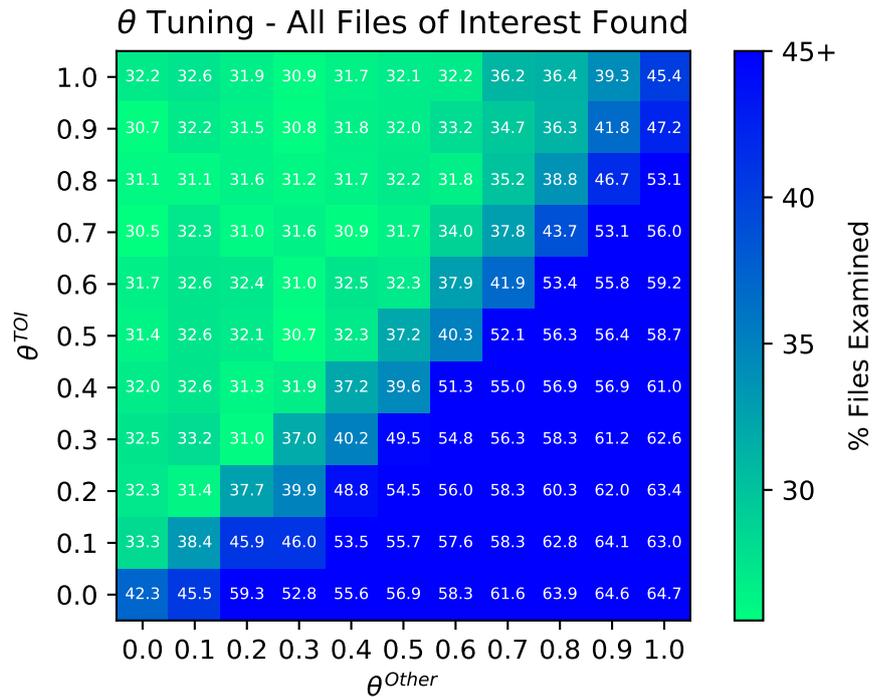


Figure 5.8: Training Corpus - File type weight tuning - All Files of Interest found

5.8.2 Informed Search vs MCFS

Domain knowledge enhanced MCFS implementations underperformed expectations. Whilst outperforming best-first search in all instances, Figure 5.11 shows MCFS-B only marginally outperformed MCFS in some instances, but never to a degree outside standard error for both. MCFS-TB returns marginal improvement when using PhotoDNA, but otherwise displays poor performance when compared with MCFS.

5.8.3 Internal vs External Devices

We anticipated varying results between internal and external devices, given the tendency of internal devices being used for operating system installations (and therefore maintaining associated file system structures). Figures 5.12 and 5.13 show the contrast between the two.

As an example, B2, an internal HDD with comparatively few files and directories, contains a tight FOI cluster. The crawls reflect this, with steep lines reflecting the rapid identification of all remaining FOI once the first is encountered.

The crawl for B3, a USB memory stick with FOI located across numerous locations, shows the impact of such a topology. Domain knowledge performed poorly on this device, with the best first search performing worst of all scorers shown. Depth first search initially performs strongly with the crawler initially striking a cluster of FOI, but progress slows dramatically after approximately half of all FOI are located. The PhotoDNA scorer based crawl closely tracks the depth first approach until this point, whereupon it continues its

Granular θ Tuning - SkinTone and PhotoDNA Scorers

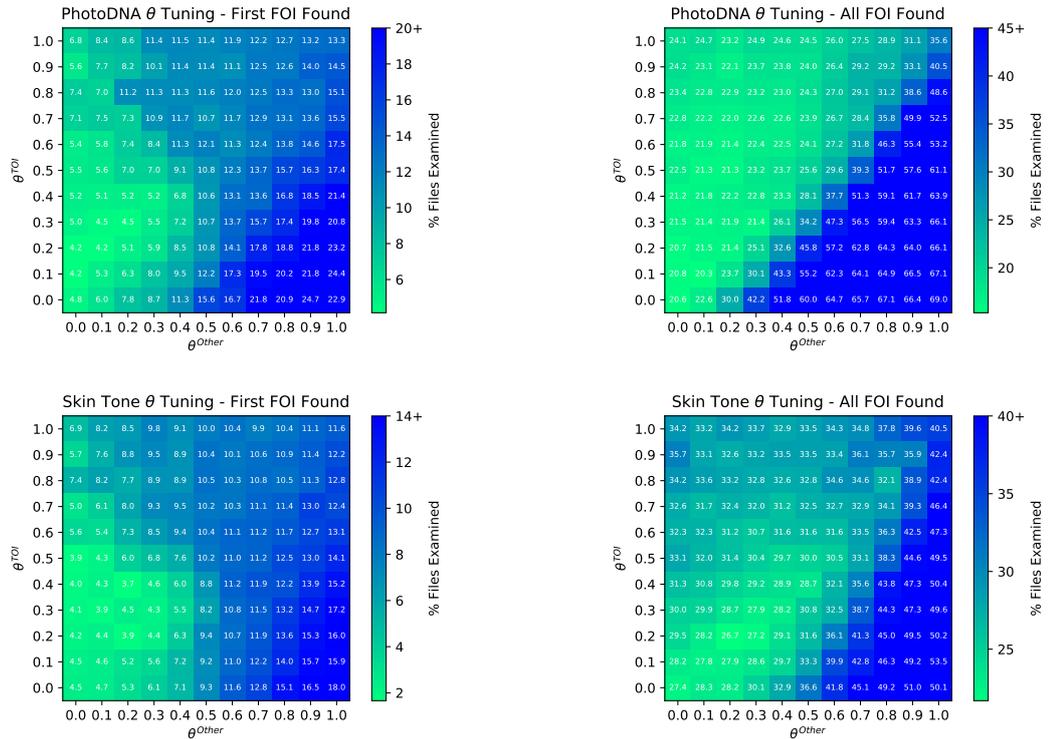


Figure 5.9: Training Corpus - Granular scorer file type weight tuning. Refer Appendices D.3 and D.4 for tabulated data

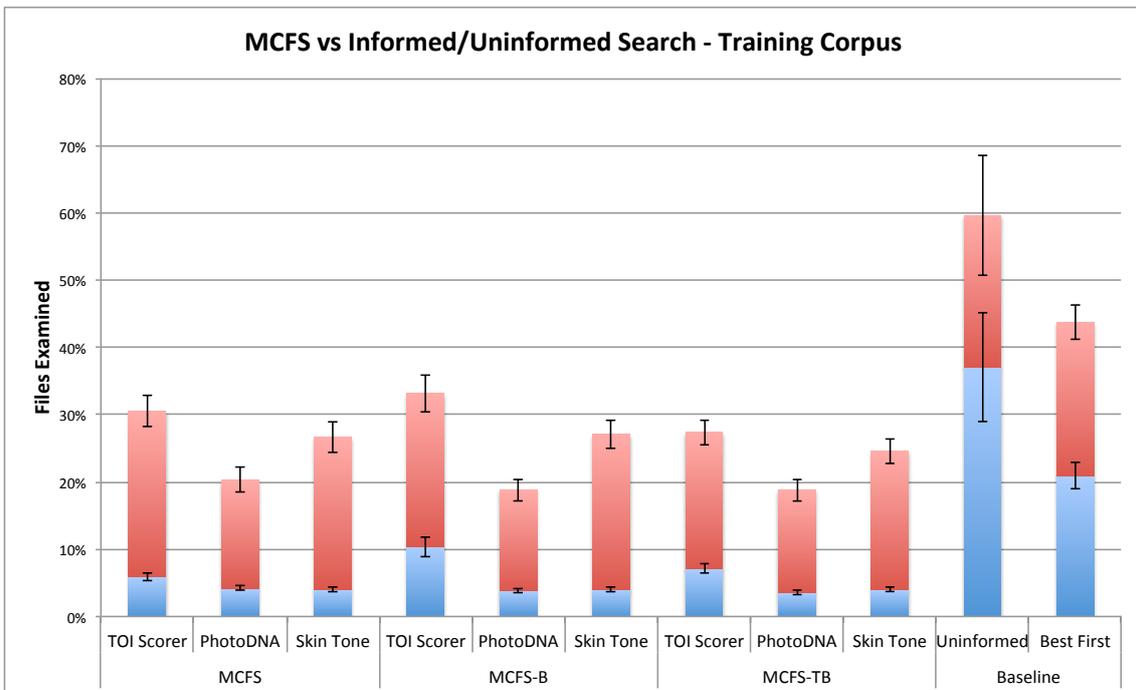


Figure 5.10: Training Corpus - MCFS vs Uninformed and Informed Search, First/All FOI Found

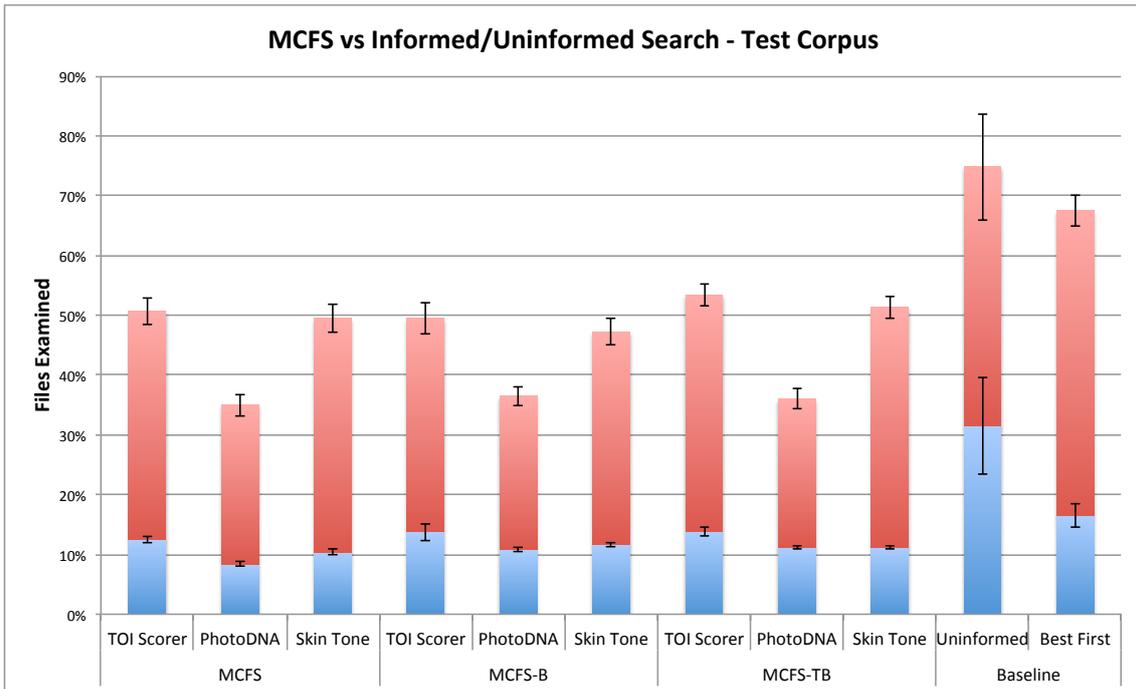


Figure 5.11: Test Corpus - MCFS vs Uninformed and Informed Search, First/All FOI Found

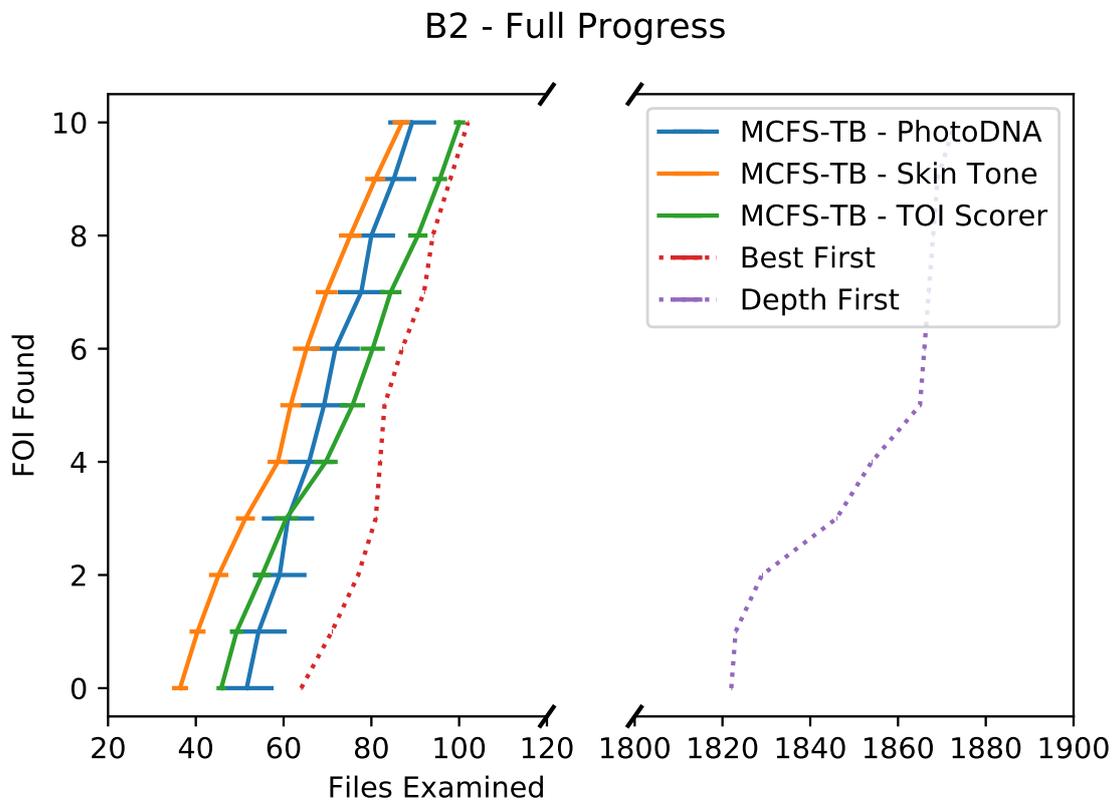


Figure 5.12: Example internal device, full crawl progress. $C_p = 0.1$. Note use of ‘broken’ x axis to allow inclusion of depth first search.

# FOI Found	Scorer	Weights	
		Type of Interest θ^{TOI}	Other Files θ^{other}
1st	Type of Interest	0.8	0.6
All	Type of Interest	0.7	0.01
1st	PhotoDNA	0.2	0.01
All	PhotoDNA	0.1	0.1
1st	Skin Tone	0.2	0.2
All	Skin Tone	0.2	0.2

Table 5.5: Training set best performing file type weight combinations. Note: Numerous configurations perform within standard error of each top performing combination. Refer Appendices for tabulated data.

steep climb to locating approximately $\frac{3}{4}$ of all FOI. It comfortably outperforms the other crawlers, reflecting strong feedback available to the scorer.

In its basic form, MCFS has shown itself to be an effective method for accelerating file system search, particularly in larger, complex tree structures. Search efficiency can be improved by around a third compared to uninformed (depth/breadth first) search, with minimal processing and memory overhead. Whilst less effective when searching the comparatively simpler file system structures seen within external devices, search performance still is materially improved.

Our non-binary, similarity based file scorers performed strongly. Of particular note, PhotoDNA performed extremely well, identifying all FOI in less than half the time taken by uninformed search. We have no doubt that the high level of performance is at least partially due to comparatively high scores given to unknown CEM imagery encountered during the crawl, indicating a high level of robustness within the algorithm.

Skin tone based scoring shows promise, providing some performance improvement when using MCFS-B. Whilst not as effective as PhotoDNA in this instance, the lack of reliance upon known FOI similarity datasets makes this scorer suitable in situations where such information is unavailable.

The utility of domain knowledge was lower than expected when combined with MCFS, though we draw encouragement from the outperformance of equivalent best-first searches. It would appear that our text-based features suffer from a lack of knowledge of our target devices, indicating a high degree of variability even within a field as seemingly specialised as CEM (an issue already discussed within chapter 4). Whilst good domain knowledge has been shown to improve performance in Monte-Carlo based search, the underlying use of random selection and online learning means the algorithm is less susceptible to issues such as overfitting than classifiers purely based upon feature matrices as in Marturana and Tacconi (2013).

The approach taken with MCFS differs from related work in this field by prioritising the order of files subjected to ‘deep’ examination based upon file system hierarchies and learned experience, rather than attempting to predict file value according to previously identified features. We have shown that not unlike the findings of Roussev and Quates (2012), MCFS can learn from inter-source similarities in the form of known files of interest

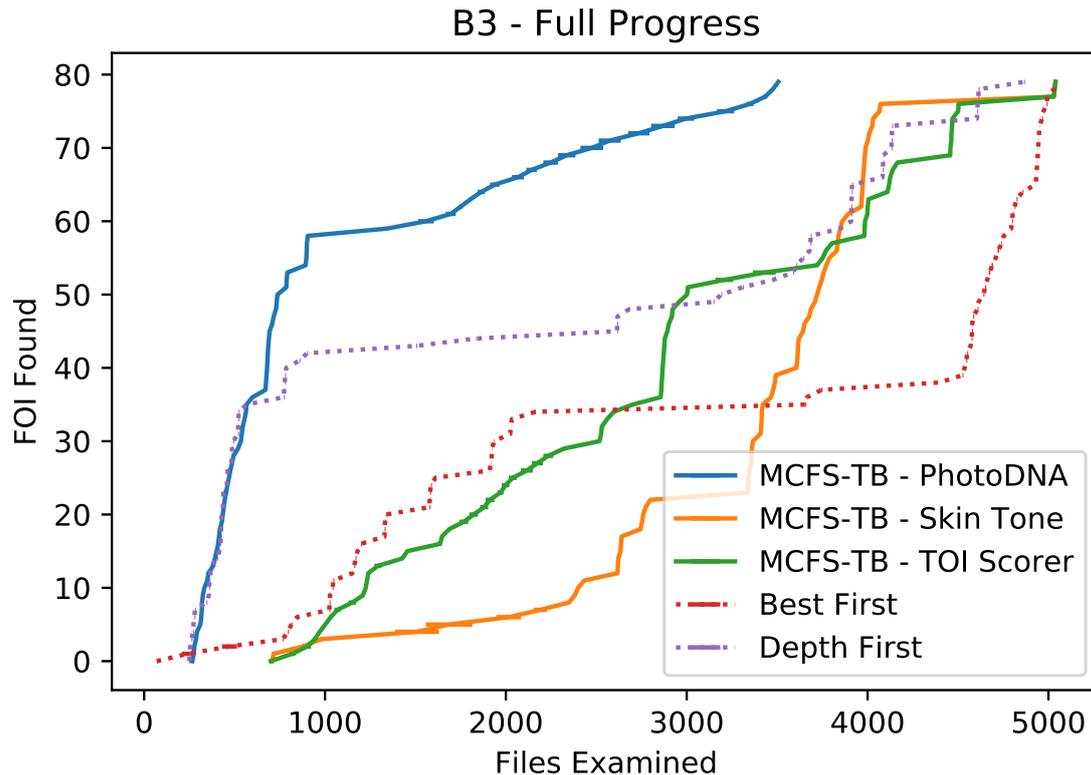


Figure 5.13: Example external device - full crawl progress

seen in previous investigations. With ‘always on’ scoring, analysts can observe search evolution and override the process when obvious clusters of interest become apparent. On the other hand, the unsupervised nature of the approach removes analysts from the burden of continuous monitoring and manual interaction, allowing for multiple examinations to be conducted simultaneously - a substantial performance improvement in live, on-site examinations.

MCFS is potentially a very powerful tool when used in conjunction with the clustering tools proposed by Beebe et al. (2011). Indexing and similarity hashing have already been identified as resource intensive during analysis - online clustering and indexing could greatly improve the performance of this approach by providing analysts with relevant results *early* in the crawl, rather than at the end of batch processing.

Data reduction approaches such as those proposed by Richard and Roussev (2006); Ferraro and Russell (2004) could greatly benefit from the MCFS approach, particularly due to the ability to identify *locations* of interest for analyst review.

5.9 Implementing MCFS

A common DF practitioner complaint already discussed within Chapter 2 is the unavailability of tools implementing emerging research. Implementation was a key consideration in the design of MCFS, principally to ensure availability. The algorithm is simple and portable, capable of implementation in languages such as C# and Python. It is

lightweight, with the largest memory overhead arising from the need to store visit counts and scores. Random selection is the main consumer of computational power, though the impact relies entirely upon the selected Random Number Generator (RNG). We utilised Java's `SecureRandom` function during our experiments, but only due to a need to ensure randomness between threads (multiple instances being run in parallel for efficiency). Realistically, any lightweight RNG should suffice in this role.

5.9.1 Demo Crawler

A demonstration application was written as part of the AFP's Project Stonefish as a proof of concept for both the MCFS algorithm and the classifier detailed within Chapter 4. Dubbed the 'Stonefish Crawler', the application was designed principally as a triage tool, the key considerations being (in no particular order):

Support: Whereas the original experimental code was written in Java, `C#` was selected as the target language/environment for this project - `C#` being the preferred development language within the AFP.

Compatibility: DF within the AFP utilises a standard hardware/software configuration for their field kit, with the established and preferred methodology for examinations remaining direct connection to storage media via hardware write blockers, allowing physical level access via a 'known good' hardware/software configuration with ample memory and processing capacity. Boot discs are the first fallback, providing physical level access to embedded or otherwise non-removable media (such as the storage media used in Apple's MacBook series). Live examination can be performed using the target devices 'as is', but this is typically the approach of last resort. No 'one size fits all' language capable of supporting all configurations currently exists. Given the need to support `C#`, `.NET Core` (with runtime environments available for Windows, Linux and MacOS) was the obvious choice for development. A GUI was ruled 'out of scope' on compatibility and performance grounds - `.NET Core` not offering a GUI library, and graphics requiring additional processing and configuration to ensure correct vision even on low resolution systems.

Speed: Materials of interest need to be found at least as quickly as existing depth/breadth first approaches. Given the triage scenario, a measure utilising 'first past the post' is most suitable.

Efficiency: As explained previously, the DF hardware configuration provides ample memory and processor capacity, with interface speeds typically being the limiting factor during triage. The need to run on potentially obsolete hardware negates this assumption, leading to a need to emphasise frugality rather than convenience.

Figure 5.14 shows the initial 'splash' screen, including a simple interface for selecting target drives for search. Users can elect to search multiple drives, in which case a virtual root node is established with the selected drives as first branches - treating them as a single

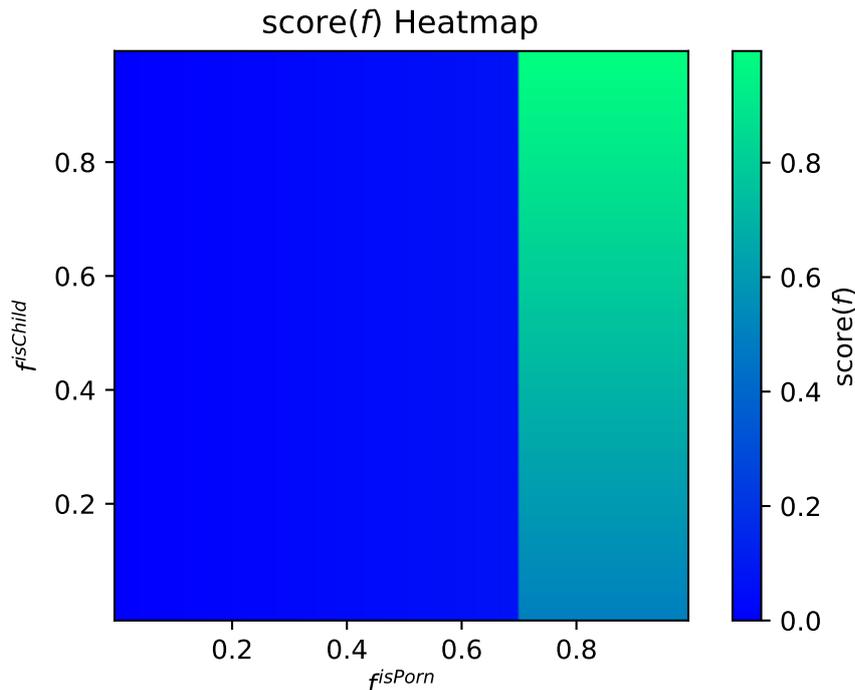


Figure 5.15: Stonefish Crawler $score(f)$ heatmap (assumes $\theta^{other} = 0.8$)

$$score(f) = \begin{cases} 1 & \text{if } f \in \text{KnownOfInterest} \\ 0 & \text{if } f \in \text{KnownIgnorables} \\ \text{if } type(f) \in \text{Image} & \text{if } f^{(isPorn)} \begin{cases} \geq 0.7, \text{ then } 0.5 + (\frac{f^{(isChild)}}{2}) \\ \text{else } (\frac{f^{(isPorn)}}{2}) \times (1 - \theta^{TOI}) \end{cases} \\ \theta^{other} & \text{otherwise} \end{cases}$$

A heatmap demonstrating the application of this scorer is shown in Figure 5.15.

Search Configurations

The Stonefish crawler was also designed to demonstrate the flexibility of MCFS, with three search levels available:

1. **Quick:** The crawler only examines parseable files (images), using file extensions to identify content type.
2. **Medium:** The crawler checks filenames and sizes against ‘of interest’ and ignorable hashsets. File extensions used for identifying file type.
3. **Thorough:** Digest values are calculated for all files and checked against hashsets. File types are determined using file headers.

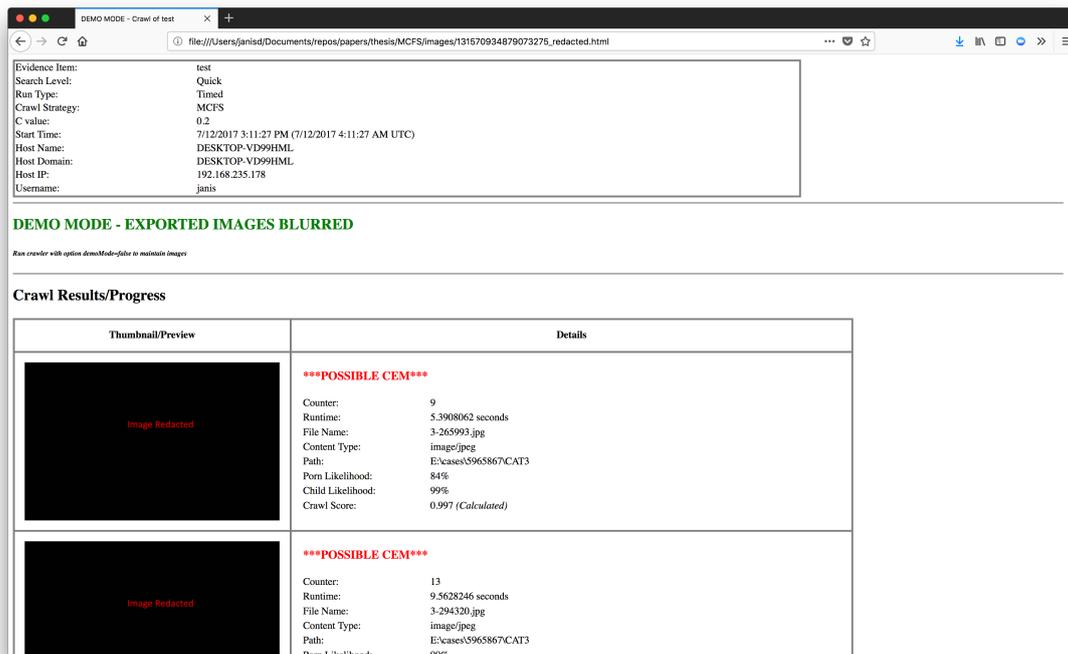


Figure 5.16: Stonefish Crawler Demo - HTML Report (*Images redacted*)

All levels utilise the classifier for scoring. Calculation of cryptographic digests does increase processing requirements, but we found the overhead to be minimal due to candidate files already being read for processing (the digest and image processing threads sharing the same buffer).

Reporting Results

Being a terminal based application, the Stonefish crawler gives live results via standard output (i.e. text on the screen). Lightweight and dynamic, it is an output method more suited for transient information such as updates rather than reports. We therefore chose to output detailed crawl information to static HTML - a file format capable of being rendered by the web browsers included within all major operating system distributions over the last decade. All HTML elements used in the output are from standards introduced and largely unchanged since the 1990s, again helping ensure compatibility. No Javascript, CSS or dynamic elements are included for performance and compatibility reasons.

Figure 5.16 shows a report excerpt. Certainly not a visually dynamic format, the file nonetheless can serve as a convenient DF report and contemporaneous note of the search. Image thumbnails (redacted, but shown in correct size) can be clicked to open the original source in the default viewer. The report can be read and updated during the crawl, with updates rendered via the browser's 'refresh' command. The thumbnail binaries are embedded within the HTML file, making the report readily portable. A cryptographic hash of the report can be output to screen as a means for proving integrity, in a manner akin to EWF files.

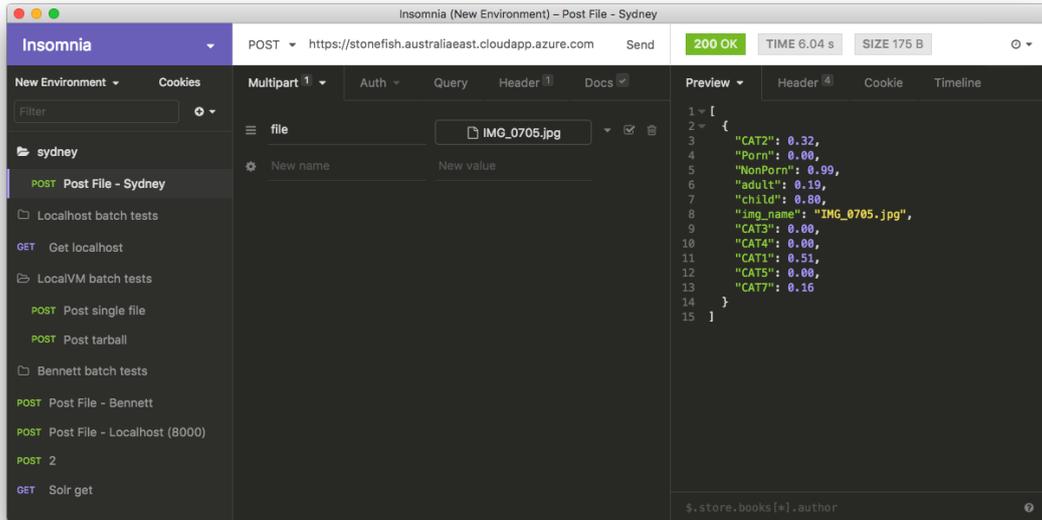


Figure 5.17: Stonefish Classifier Server - sample interaction

The crawler was designed to run on basic hardware - in fact, an Android based version has been implemented for use on tablets and mobile phones. The presence of GPUs or higher-end CPUs can not be assumed. Instead, we chose to implement a client/server configuration, sending reduced images to a remote server.

5.9.2 Classifier Server

We hosted the classifier within a Python based web server, accessible through RESTful calls via HTTPS. Clients connect to the service via a `post` call, with the file uploaded as a multipart form field. Responses are returned via JSON¹³. The server was run on a dual Xeon PC¹⁴ hosted within Monash University. Network latency was anticipated to be a constraint for the service's success, but we observed a 4G wi-fi hub to provide adequate speeds - a demonstration run in Barton, ACT (an urban area approximately 700kms from Monash University) saw processing take less than two seconds per image¹⁵. Bandwidth emerged to be minor issue, largely due to the application resizing images to a maximum dimension of 224 pixels (i.e. $length^{long} = 224$ pixels, $length^{short} = \text{original short edge} \times \frac{224}{\text{original long edge}}$).

Figure 5.17 shows a sample interaction with the server, with the classifier confidences being 0.0 for pornography and 0.8 for containing a child. Note that due to the client software used, the image is not resized prior to transmission, resulting in a delayed (6 seconds) turnaround.

For workflow compatibility purposes, the server is also capable of processing digest values - instead (or in addition to) uploading the sample file, a query string containing the digest type and value can be included in the `post` call. The service was populated with

¹³For demonstration purposes redundant, negative values are included - e.g. both porn & nonPorn values

¹⁴GPUs were unavailable for this experiment

¹⁵The server also supports faster batch processing via archive files



Figure 5.18: IMG_0705.jpg - refer Figure 5.17

‘known ignorable’ MD5 values from the NSRL collection. Interestingly, the digest lookups service required a greater memory footprint than the classifier, with the Python script holding all digests in a set for speed of lookup. Even though these sets were cached (using `pickle`) for loading speed, loading the sets still took over a minute. Populating the sets typically took more than five minutes, leading to delays initiating the service. Whereas this performance could no doubt be improved through memory management, parallel processing or faster storage, this does demonstrate potential downsides, particularly as the database of known files grows.

Flexibility and Affordability

Being Python based, the classifier server is also highly portable. Cloud-based services (including GPU instances) have been utilised on occasion, with a bash script now capable of installing all dependencies, downloading source code (from a repository at Monash University) and running the server within twenty five minutes. If extended to PowerShell (all testing has been conducted on MS Azure), it is entirely feasible that high-end classifier servers could be created and destroyed for triage on a per-search warrant basis by individual analysts, costing approximately AUD\$2 per hour.

5.10 Conclusion

In this chapter we have not only introduced and proven the efficacy of Monte Carlo Filesystem Search as a device search accelerator, but we have also demonstrated the viability of the CEM classifier from Chapter 4 as an applied tool rather than just an abstract research concept. This combination moves DF search from an exclusive focus upon known items of interest to include *likely* items of interest, representing a major increase in capability without a corresponding requirement for resources.

Chapter 6

Conclusions

This dissertation summarises our efforts at improving the efficiency and safety of Digital Forensics (DF) analysis by establishing the viability of automation, with a specific focus upon the *accessibility* of outputs to practitioners worldwide. Our focus upon technical, legal and procedural integration supports automation throughout the investigation lifecycle - informing what evidence is being sought, the nature of that evidence, and the means to identify, preserve and collect that evidence in the most efficient and safe manner possible.

To summarise our findings in regards to the research questions:

Can criminality be ontologised? Yes, though with care. We have demonstrated the viability of ‘criminal’ ontologies and schemas, but also show the dangers of repurposing those designed for other reasons - even when using the same source material.

Chapter 3 introduces the Tor-use Motivation Model (TMM), whose recording of individual sites’ purpose *and* motivation provides a great deal of accuracy without ambiguity, allowing for more granular recording of likely sites of interest without bloated and vague labels.

Chapter 4 introduces the Majura Schema, a means of objectively annotating sexually explicit materials. By avoiding subjective terms such as ‘child abuse’, we overcome the limitations of existing Child Exploitation Material (CEM) schemas such as Child Exploitation Tracking System (CETS) and Combating Paedophile Information Networks in Europe (COPINE). The Majura Schema is also compatible with the use of adult pornography as a proxy for CEM - removing developer exposure to such materials, but also making a far larger and more readily accessible corpus (internet pornography) applicable to what is thus far a specialised, isolated and under-resourced field.

The implementation of these ontologies demonstrates the viability of machine readable languages within DF - particularly those capable of use in multiple jurisdictions. This directly addresses the ‘dearth of data’ we describe within Section 2.6.9, which in turn helps overcome the shortage of reliable, accessible tools detailed within Section 2.5. To summarise, this simplifies the task of sharing DF data with wider investigations tools (and vice-versa) - in this instance, ‘what’ are we searching for?

Can offensive materials be automatically recognised and classified reliably, with minimal labelling? Yes, but with caveats. In Chapter 4 we report on the design and implementation of a deep learning based CEM classifier - evolving automated recognition within DF from similarity with previously seen materials to the underlying concepts themselves. The prototype is demonstrably more than adequate for triage, but also reflects the limitations of the CETS schema when used for training data.

Our ‘ontologisation’ of criminal behaviour informs the automated definition of what types of data investigators are seeking. This research question takes the next logical step, and deals with recognising such materials when encountered. Whilst certainly an improvement on prior methods, the existing automated detection methodologies detailed within Section 2.6 are insufficient for use beyond rudimentary triage systems. By being limited to either basic CEM detection (i.e. materials of a sexual nature depicting children) or based upon simulated and/or limited datasets, their reliability and scope are insufficient for further use, their outputs unable to be mapped against scales such as CETS and COPINE for use in prosecution. The need for manual annotation therefore remains unaddressed, leaving investigators and analysts open to the same risks discussed within Section 2.5.1.

The methods used in designing and implementing our classifiers are novel and robust. We have proven automated classifiers to be a feasible approach for efficiency and practitioner safety purposes.

Can automated classifiers prioritise search for evidentiary electronic materials?

Yes. The previous research questions build towards the integrated use of automated detection and classification throughout the investigative lifecycle. In Section 2.2.1 we detail how within Australia, search warrants tend to be the primary means of evidence preservation and collection. Unfortunately, these tend to be time and resource limited affairs, conducted in highly variable safety and infrastructure conditions. An obvious performance and safety improvement is parallel search across multiple devices - an option typically limited by access to sufficient personnel. Here, we seek to discover if an unattended search can at least closely emulate an experienced practitioner’s performance - i.e. how quickly (in terms of files examined) can evidentiary material be located, and can it be done in an efficient (processor and memory) manner?

In chapter 5 we introduce the Monte Carlo Filesystem Search (MCFS), a lightweight and modular crawl strategy specifically designed for accelerating file system searches. We prove the algorithm’s performance with extensive searches of ‘real world’ evidentiary data, and then demonstrate the implementation of an MCFS crawler utilising the CEM classifier from chapter 4.

Beyond experimental performance, the crawler implementation proves the feasibility of classifiers and crawl strategies as tools within DF. Different configurations support local, remote and hybrid processing, enabling the use of a wide range of hardware and operating systems.

This work makes a major contribution to the sustainability of DF within law enforcement, particularly whilst the field grapples with the impacts of rapid growth and exposure to harmful materials. A common practitioner complaint (detailed within this work) is of

research outputs not translating to an operational context. Correspondingly, all outputs from this research are readily deployable, being compatible with existing technical and jurisdictional frameworks.

6.1 Future Work

Our research represents a foundation for the development and exploitation of automated classification and search methodologies within DF. Given time and resource constraints, many elements and possible variants of the work detailed within this document couldn't be examined. Future work arising from this research includes:

Understanding and Categorising Online Criminal Activity

- The development of specialised schemas structures for specific crime types could aid automated ranking and prioritisation of items for law enforcement akin to the Majura Schema's role in CEM investigations.
- Annotations generated during the Tor-use Motivation Model (TMM)'s development could be re-used to develop text-based classifiers for automatically detecting and prioritising sites of interest on Tor. The classification of embedded and linked elements such as thumbnails and multimedia could also be used, either in isolation or in ensemble.
- In terms of refining TMM, we intend to evaluate the inclusion of 'normalisation' as a motivation - i.e. the specific treatment of behaviours or attitudes typically seen as abhorrent or frowned upon by the general public, but treated as 'normal' or 'acceptable' between proponents. Typically such materials would be recorded as recruitment/advocacy, but our observance of several such sites (particularly in the child exploitation area) appearing to be written specifically for persons of a like view (as opposed to possible 'recruits') makes this a possible addition to the model.

Recognising & Classifying CEM

- The CEM classifier is limited to still images. Useful as a prototype or proof of concept, this overlooks a large proportion of such materials typically encountered by law enforcement. A classifier capable of analysing movies is essential for the viability of this approach.
- The CEM classifier's three module architecture is susceptible to false results in earlier stages. An alternative approach is to create a single classifier capable of recognising and contrasting innocent material, pornography and CEM, possibly with mapping back to existing schemas. This approach was deemed out of scope due to limited computing capacity rather than research considerations, and remains a preferred option in future implementations.

- The Majura schema’s use in developing a ‘proxy’ corpus of adult pornography in place of CEM is untested, but a promising approach in evolving the safety of developers and annotations in this field. *An initiative addressing this shortcoming is currently underway within the AFP, with volunteers currently being assembled for a 10,000 image trial using a browser based annotation package.*

Accelerating Search

- MCFS specifically benefits from the additional feedback provided by increased scorer granularity. Implementations such as that using our prototype CEM classifier show promise, but extensive additional tuning will be required to achieve optimal performance.
- Automated tuning of parameters *during* crawls is a promising direction for improving performance, allowing crawler behaviour to adjust according to encountered landscapes such as operating systems, file type clusters and average directory depths.
- The crawl of Tor in chapter 3 was set up as a tree, with duplicate links simply dropped from subsequent pages. All selections were randomly made, with the intention to run simulated crawls using MCFS. This didn’t eventuate due to time limitations, and numerous questions remain regarding how one deals with a network where hosts seem to drop in and out on regular intervals. Ultimately, though, the very representation of a web is problematic for MCFS, as a crawler could theoretically get caught in infinite loops whenever a link back to a previously visited site is encountered. The use of Directed Acyclic Graphs (DAGs) is a possible solution, but again, will require further work to establish value.

6.2 Practical Impacts

On a practical level, the most immediate improvement to research in this field is the removal of legal, ethical and health risks associated with research into automated recognition of offensive materials. Ethical & legal clearance is feasible within Australia for research into the violent imagery and movies typically associated with terrorist and extremist groups, though subject to some restrictions regarding distribution and viewership. Even disregarding ethical clearance, CEMs are a more complex matter. Many jurisdictions have criminalised the accessing and/or possession of CEM, often as a strict & absolute offence - i.e. mens rea needn’t be proven, with minimal (or no) exemptions. Commonwealth (Australian) legislation does provide exceptions for developers of detection/security software and also those acting in the ‘public good’, but interestingly, ‘research’ in this field requires written permission from the relevant minister (in the case of this thesis, the then Commonwealth Minister for Justice). The development of a secure offensive data storage and processing facility would be of great benefit, by giving researchers ‘hands off’ access

to such materials for training, testing and validating detection and classification methodologies. Such a data ‘airlock’, funded by the AFP is currently under construction and is scheduled for implementation at Monash University during 2018.

The research detailed within this dissertation directly influenced the establishment of Project Stonefish, an international joint law enforcement, industry and academic initiative aimed directly at reducing health & safety risks caused by exposure to offensive materials.

Appendix A

Chapter 3 Appendices

A.1 Initial 500 Page Tag/Label Combinations

Categories			Count
Failed to Render/Error			86
Foreign	Login Page	Unclear	42
Foreign	Unclear		40
image			28
captcha	image		13
Failed to Render/Error	Not of Interest		12
Unclear			10
captcha			9
image	narcotics		8
Not of Interest			8
Blog - Personal			8
Foreign	Login Page		7
news			7
Login Page	Unclear		6
image	weapons		6
Data Repository/Share site	Unclear		5
Marketplace	Of Interest	narcotics	5
Index/Directory			5
Marketplace	Unclear		4
Failed to Render/Error	Unclear		4
Not of Interest	System File		4
software			4
Data Repository/Share site			4
Login Page	Marketplace		3
Data Repository/Share site	Foreign	Unclear	3
Advocacy	Not of Interest		3
Diagnostics	Not of Interest		3
Unclear	image		3
Not of Interest	image		3
Forum	Mental health		3
Marketplace	narcotics		3
Foreign	Marketplace	narcotics	3

Categories			Count
System File			3
image	Not of Interest		2
Blog - Personal	Not of Interest	Personal Site	2
Marketplace	Pharmaceuticals		2
Get Rich Quick	bitcoin		2
Hosting/Service Provider	Tor		2
Child Pornography	Login Page	Of Interest	2
Not of Interest	Unclear		2
Advocacy			2
Tor	software		2
Adult Pornography	Data Repository/Share site		2
Hosting/Service Provider			2
Index/Directory	Tor		2
Data Repository/Share site	software		2
Electronics	Marketplace		2
Marketplace	Of Interest	weapons	2
Data Repository/Share site	Tor	software	2
Forum	Of Interest	narcotics	2
bitcoin	image		2
Index/Directory	System File		2
Forum	Unclear		2
Unclear	captcha		2
Login Page	Marketplace	Unclear	2
Advocacy	Books	Data Repository/Share site	2
Data Repository/Share site	Photographs		1
Advocacy	Human Rights	privacy	1
Marketplace	narcotics	Of Interest	1
captcha	Data Repository/Share site		1
bitcoin	Currency Mining/Generation		1

Categories				Count
Copyrighted materials	Data Repository/Share site	Foreign	movies	1
Hosting/Service Provider	Tor	encryption		1
Marketplace	captcha			1
Foreign	Marketplace	Unclear		1
Blog - Personal	Foreign			1
Forum	Login Page			1
Electronics	Foreign	Marketplace		1
Adult Pornography	Data Repository/Share site	Unclear		1
Forum	Not of Interest			1
Failed to Render/Error	Foreign	Index/Directory		1
Data Repository/Share site	Failed to Render/Error			1
Blog - Personal	Not of Interest	news		1
Marketplace	Network Status/-Diagnostics			1
bitcoin	news			1
Online Chat	software			1
Network Status/-Diagnostics	Not of Interest			1
Hosting/Service Provider	Tor	email		1
Copyrighted materials	Data Repository/Share site	movies	music	1
Education/Manual	Not of Interest	software		1
Education/Manual	Foreign	Not of Interest		1
Data Repository/Share site	Login Page			1
Network Status/-Diagnostics				1
litecoin	Search Engine			1
Advocacy	Wikileaks	news		1
Data Repository/Share site	Education/Manual	software	Tor	1
Index/Directory	Tor	news		1
Search Engine	Tor			1
Education/Manual	Tor	software		1

Categories				Count
Credentials/ Identification	Credit Cards	Marketplace		1
Advocacy	Wikileaks			1
Foreign	Index/Directory			1
Foreign	Login Page	Marketplace	Unclear	1
Foreign	Forum	Unclear		1
Data Repository/ Share site	Marketplace			1
Credit Cards	image			1
Not of Interest	software			1
Hacking Services	Marketplace			1
Failed to Render/ Error	Foreign			1
Gambling/gaming				1
Counterfeit	Credentials/ Identification	Marketplace		1
Login Page	captcha			1
Child Pornography	Data Repository/ Share site	Forum	Of Interest	1
Books	Credit Cards	Data Repository/ Share site	Forum	1
Network Status/ Diagnostics	Tor			1
Books	Hacking Services	news		1
captcha	software			1
Data Repository/ Share site	Not of Interest	software		1
wikipedia article				1
Hosting/Service Provider	Index/Directory	software		1
Books	Hacking Services			1
Gore/Offensive Imagery	image			1
Advocacy	Not of Interest	privacy		1
news	Not of Interest			1
Credentials/ Identification	Hacking Services	Marketplace		1
Foreign	Login Page	captcha		1
Hosting/Service Provider	image			1

Categories				Count
bitcoin				1
Foreign	wikipedia article			1
Data Repository/Share site	Login Page	Of Interest		1
Foreign	Forum	Of Interest		1
Foreign	Unclear	image		1
Adult Pornography	Foreign			1
Forum	Index/Directory	bitcoin		1
Financial	Login Page			1
Data Repository/Share site	Not of Interest	music		1
AUD	bitcoin			1
Foreign	Not of Interest			1
Hosting/Service Provider	email	encryption		1
Counterfeit	Credit Cards	Marketplace		1
Electronics	Failed to Render/Error	Marketplace		1
Hitman	Marketplace			1
Clothing	Marketplace			1
Education/Manual	wikipedia article			1
Index/Directory	Not of Interest	System File		1
Advocacy	Fundraising	privacy	software	1
Hacking Services	Of Interest			1
Financial	bitcoin	laundering/ financial anonymiser		1
Index/Directory	bitcoin			1
Data Repository/Share site	Failed to Render/Error	Unclear		1
Education / Manual	Encryption / Cryptography	wikipedia article		1
currency	Forum	narcotics		1
Login Page				1
Counterfeit	currency	Forum	Index/Directory	1
Marketplace	image			1
Login Page	Of Interest	narcotics		1
Child Pornography	Data Repository/Share site	Of Interest		1

Categories				Count
Copyrighted materials	Data Repository/Share site	music		1
Of Interest	bitcoin	lauderin / financial anonymiser		1
Hacking Services	Marketplace	Of Interest		1
currency	euro	image		1
Credentials/Identification	Marketplace			1
Hosting/Service Provider	advertising/pay per click	bitcoin		1
Credit Cards	Encryption/Cryptography	Forum	Hacking Services	1
Forum	Hacking Services	Of Interest		1
Failed to Render/Error	Foreign	Forum		1
Data Repository/Share site	encryption	software		1
Foreign	Login Page	Marketplace	Of Interest	1
Data Repository/Share site	music			1
Fundraising	Tor	software		1
Not of Interest	Religious			1
Forum	Not of Interest	Personal Site		1
Of Interest	narcotics			1
Foreign				1
Foreign	Of Interest	narcotics		1
litecoin				1
Credit Cards	Marketplace	Of Interest		1
Forum	Login Page	Unclear		1
Get Rich Quick	Of Interest	bitcoin		1
Foreign	encryption			1
escrow				1
Index/Directory	Unclear			1
Email	Hosting/Service Provider			1
Adult Pornography	Child Pornography	Login Page	Of Interest	1
Not of Interest	Online Chat			1
Foreign	Forum			1
Index/Directory	Not of Interest			1

Categories				Count
Failed to Render/Error	Forum	Of Interest	narcotics	1
Foreign	Of Interest			1
bitcoin	currency mining			1
Foreign	Login Page	Of Interest		1
Login Page	Marketplace	Of Interest		1
Credit Cards	Forum			1
Child Pornography	Of Interest			1
Encryption/ Cryptography	Foreign	Unclear		1
Encryption/ Cryptography	Forum			1
software	wikipedia article			1
Books	Copyrighted materials	Data Repository/Share site		1
Login Page	Marketplace	narcotics		1
Forum				1
Hosting/Service Provider	music			1
Blog - Personal	Education/Manual	Personal Site	software	1
Child Pornography	image			1
Seized Site (FBI etc)	image			1

Table A.1: Raw unstructured labelling data - initial 500 page categorisation

Appendix B

Chapter 4 Appendices

B.1 CETS & COPINE Scales

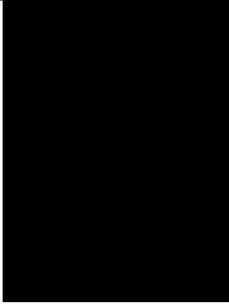
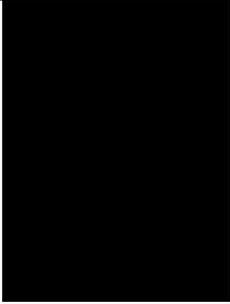
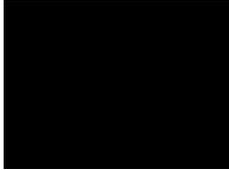
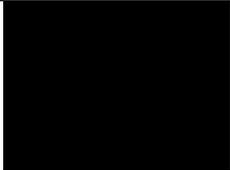
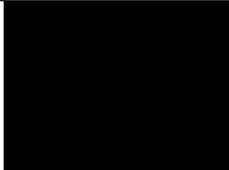
Category	CETS Classification	Guide
1	CEM - No Sexual Activity	Depictions of Children with No Sexual Activity - Nudity, surreptitious images showing underwear, nakedness, sexually suggestive posing, explicit emphasis on genital areas, solo urination.
2	CEM - Solo \ Sex Acts between children	Solo masturbation by a child or non penetrative sex acts between children. Includes the use of penetrative sex toys by the victim (if offender is using toy is Cat 4).
3	CEM - Adult Non-Penetrative	Non-Penetrative Sexual Activity between Child(ren) and Adult(s). Mutual masturbation and other non-penetrative sexual activity.
4	CEM - Child \ Adult Penetrate	Penetrative Sexual Activity between Child(ren) and Adult(s) - including, but not limited to, intercourse, cummingus and fellatio.
5	CEM - Sadism \ Bestiality \ Child Abuse	Sadism, Bestiality or Humiliation (urination, defecation, vomit, bondage etc) or Child Abuse as per CCA 1995.
6	CEM - Animated or Virtual	Anime, cartoons, comics and drawings depicting children engaged in sexual poses or activity.
7	CEM - Non-illegal \ Indicative	Non-illegal child material (believed to form part of a series containing CEM). Includes images of circumcision being performed.
8	Adult Pornography	All pornographic material not considered CEM related.
9	Ignorable	Banners and other non-objectionable graphics useful for establishing proportionality. System files and unrelated images - holiday snaps, landscape, family photos, etc.
0	Unchecked	Material not yet assigned a category.
	<i>If in doubt (about age) make it Category 8 - Adult. If undecided between two categories - make it the lower category.</i>	

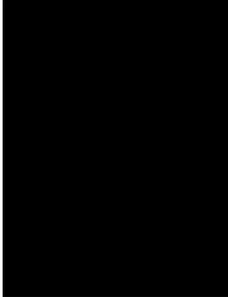
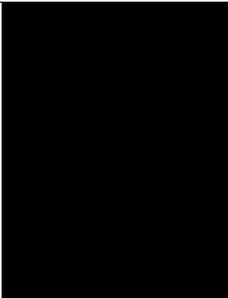
Table B.1: CETS, including AFP guideline for labelling/annotation of files in child exploitation investigations

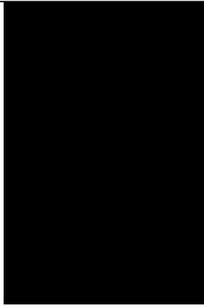
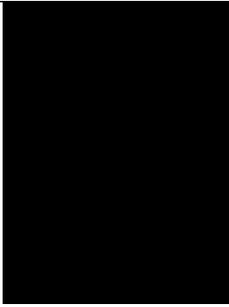
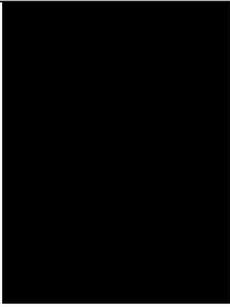
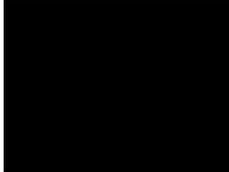
Level	Name	Description of Picture Qualities
1	Indicative	Non-erotic and non-sexualised pictures showing children in their underwear, swimming costumes, etc. from either commercial sources or family albums; pictures of children playing in normal settings, in which the context or organisation of pictures by the collector indicates inappropriateness
2	Nudist	Pictures of naked or semi-naked children in appropriate nudist settings, and from legitimate sources
3	Erotica	Surreptitiously taken photographs of children, in play areas or other safe environments showing either underwear or varying degrees of nakedness
4	Posing	Deliberately posed pictures of children fully, partially clothed or naked (where the amount, context and organisation suggests sexual interest)
5	Erotic Posing	Deliberately posed pictures of fully, partially clothed or naked children in sexualised or provocative poses
6	Explicit Erotic Posing	Emphasising genital areas where the child is either naked, partially or fully clothed
7	Explicit Sexual Activity	Involves touching, mutual and self-masturbation, oral sex and intercourse by child, not involving an adult
8	Assault	Pictures of children being subjected to a sexual assault, involving digital touching, involving an adult
9	Gross Assault	Grossly obscene pictures of sexual assault, involving penetrative sex, masturbation or oral sex involving an adult
10	Sadistic/bestiality	Pictures showing a child being tied, bound, beaten, whipped or otherwise subjected to, something that implies pain b. Pictures where an animal is involved in, some form of sexual behaviour with a child

Table B.2: Combating Paedophile Information Networks in Europe (COPINE) paedophile picture typology. (Taylor et al., 2001)

B.2 Test Corpus CAT1 Skin Tone Results

TEST_CAT1: Top 20 Images			
Image	Details	Image	Details
 Redacted: Annotated as CEM by AFP	Rank: 1 Skin Tone: 99.99% Porn: 0.99 Child: 0.0 CAT1: 0.01 CAT2: 0.0 CAT3: 0.98 CAT4: 0.0 CAT5: 0.0 CAT7: 0.0	 Redacted: Annotated as CEM by AFP	Rank: 2 Skin Tone: 99.98% Porn: 0.88 Child: 0.99 CAT1: 0.22 CAT2: 0.04 CAT3: 0.1 CAT4: 0.6 CAT5: 0.0 CAT7: 0.01
 Redacted: Annotated as CEM by AFP	Rank: 3 Skin Tone: 99.93% Porn: 0.99 Child: 0.98 CAT1: 0.99 CAT2: 0.0 CAT3: 0.0 CAT4: 0.0 CAT5: 0.0 CAT7: 0.0	 Redacted: Annotated as CEM by AFP	Rank: 4 Skin Tone: 99.92% Porn: 0.99 Child: 0.34 CAT1: 0.01 CAT2: 0.12 CAT3: 0.85 CAT4: 0.0 CAT5: 0.0 CAT7: 0.0
 Redacted: Annotated as CEM by AFP	Rank: 5 Skin Tone: 99.90% Porn: 0.99 Child: 0.0 CAT1: 0.03 CAT2: 0.81 CAT3: 0.06 CAT4: 0.0 CAT5: 0.0 CAT7: 0.08	 Redacted: Annotated as CEM by AFP	Rank: 6 Skin Tone: 99.89% Porn: 0.89 Child: 0.99 CAT1: 0.14 CAT2: 0.0 CAT3: 0.12 CAT4: 0.72 CAT5: 0.0 CAT7: 0.0

TEST_CAT1: Top 20 Images			
Image	Details	Image	Details
 <p>Redacted: Annotated as CEM by AFP</p>	<p>Rank: 7</p> <p>Skin Tone: 99.88%</p> <p>Porn: 0.99</p> <p>Child: 0.99</p> <p>CAT1 0.05</p> <p>CAT2 0.89</p> <p>CAT3 0.01</p> <p>CAT4 0.02</p> <p>CAT5 0.0</p> <p>CAT7 0.0</p>	 <p>Redacted: Annotated as CEM by AFP</p>	<p>Rank: 8</p> <p>Skin Tone: 99.85%</p> <p>Porn: 0.99</p> <p>Child: 0.99</p> <p>CAT1 0.99</p> <p>CAT2 0.0</p> <p>CAT3 0.0</p> <p>CAT4 0.0</p> <p>CAT5 0.0</p> <p>CAT7 0.0</p>
 <p>Redacted: Annotated as CEM by AFP</p>	<p>Rank: 9</p> <p>Skin Tone: 99.76%</p> <p>Porn: 0.96</p> <p>Child: 0.75</p> <p>CAT1 0.98</p> <p>CAT2 0.0</p> <p>CAT3 0.0</p> <p>CAT4 0.0</p> <p>CAT5 0.0</p> <p>CAT7 0.01</p>	 <p>Redacted: Annotated as CEM by AFP</p>	<p>Rank: 10</p> <p>Skin Tone: 99.68%</p> <p>Porn: 0.99</p> <p>Child: 0.01</p> <p>CAT1 0.09</p> <p>CAT2 0.0</p> <p>CAT3 0.78</p> <p>CAT4 0.11</p> <p>CAT5 0.0</p> <p>CAT7 0.0</p>
 <p>Redacted: Annotated as CEM by AFP</p>	<p>Rank: 11</p> <p>Skin Tone: 99.51%</p> <p>Porn: 0.99</p> <p>Child: 0.99</p> <p>CAT1 0.96</p> <p>CAT2 0.0</p> <p>CAT3 0.0</p> <p>CAT4 0.0</p> <p>CAT5 0.0</p> <p>CAT7 0.02</p>	 <p>Redacted: Annotated as CEM by AFP</p>	<p>Rank: 12</p> <p>Skin Tone: 99.41%</p> <p>Porn: 0.91</p> <p>Child: 0.99</p> <p>CAT1 0.88</p> <p>CAT2 0.0</p> <p>CAT3 0.0</p> <p>CAT4 0.0</p> <p>CAT5 0.0</p> <p>CAT7 0.11</p>

TEST_CAT1: Top 20 Images			
Image	Details	Image	Details
 <p>Redacted: Annotated as CEM by AFP</p>	<p>Rank: 13 Skin Tone: 99.38% Porn: 0.48 Child: 0.99 CAT1: 0.84 CAT2: 0.01 CAT3: 0.0 CAT4: 0.0 CAT5: 0.0 CAT7: 0.13</p>	 <p>Redacted: Annotated as CEM by AFP</p>	<p>Rank: 14 Skin Tone: 99.36% Porn: 0.97 Child: 0.66 CAT1: 0.97 CAT2: 0.01 CAT3: 0.0 CAT4: 0.0 CAT5: 0.0 CAT7: 0.0</p>
 <p>Redacted: Annotated as CEM by AFP</p>	<p>Rank: 15 Skin Tone: 99.30% Porn: 0.66 Child: 0.99 CAT1: 0.94 CAT2: 0.01 CAT3: 0.0 CAT4: 0.0 CAT5: 0.0 CAT7: 0.03</p>	 <p>Redacted: Annotated as CEM by AFP</p>	<p>Rank: 16 Skin Tone: 99.27% Porn: 0.33 Child: 0.34 CAT1: 0.0 CAT2: 0.0 CAT3: 0.0 CAT4: 0.0 CAT5: 0.0 CAT7: 0.99</p>
 <p>Redacted: Annotated as CEM by AFP</p>	<p>Rank: 17 Skin Tone: 99.25% Porn: 0.42 Child: 0.77 CAT1: 0.26 CAT2: 0.05 CAT3: 0.01 CAT4: 0.0 CAT5: 0.13 CAT7: 0.53</p>	 <p>Redacted: Annotated as CEM by AFP</p>	<p>Rank: 18 Skin Tone: 99.23% Porn: 0.99 Child: 0.64 CAT1: 0.01 CAT2: 0.21 CAT3: 0.0 CAT4: 0.0 CAT5: 0.0 CAT7: 0.76</p>

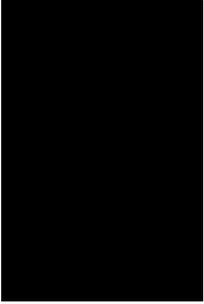
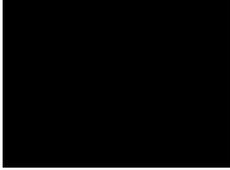
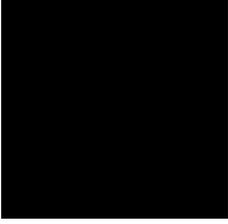
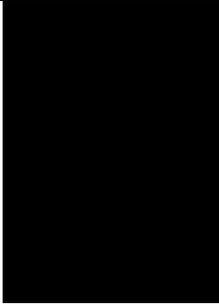
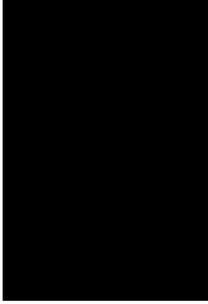
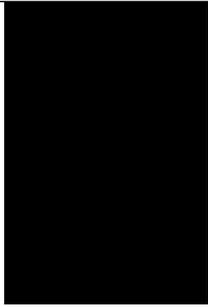
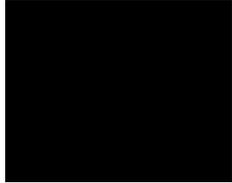
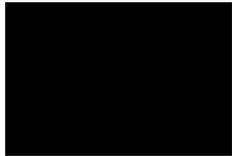
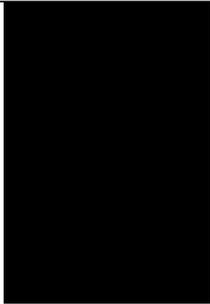
TEST_CAT1: Top 20 Images			
Image	Details	Image	Details
	Rank: 19 Skin Tone: 99.22% Porn: 0.13 Child: 0.0 CAT1 0.75 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.24		Rank: 20 Skin Tone: 99.14% Porn: 0.63 Child: 0.99 CAT1 0.0 CAT2 0.0 CAT3 0.91 CAT4 0.0 CAT5 0.0 CAT7 0.08
Redacted: Annotated as CEM by AFP		Redacted: Annotated as CEM by AFP	

Table B.3: Test corpus CAT1 top 20 results by skin tone percentage

TEST_CAT1: Bottom 20 Images			
Image	Details	Image	Details
 Redacted: Annotated as CEM by AFP	Rank: 1 Skin Tone: 0.00% Porn: 0.63 Child: 1.0 CAT1 0.02 CAT2 0.97 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.0	 Redacted: Annotated as CEM by AFP	Rank: 2 Skin Tone: 0.00% Porn: 0.37 Child: 0.99 CAT1 0.31 CAT2 0.0 CAT3 0.64 CAT4 0.0 CAT5 0.0 CAT7 0.02
 Redacted: Annotated as CEM by AFP	Rank: 3 Skin Tone: 0.00% Porn: 0.63 Child: 1.0 CAT1 0.55 CAT2 0.11 CAT3 0.31 CAT4 0.0 CAT5 0.0 CAT7 0.0	 Redacted: Annotated as CEM by AFP	Rank: 4 Skin Tone: 0.00% Porn: 0.0 Child: 0.99 CAT1 0.88 CAT2 0.1 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.0
 Redacted: Annotated as CEM by AFP	Rank: 5 Skin Tone: 0.00% Porn: 0.6 Child: 0.99 CAT1 0.31 CAT2 0.66 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.01	 Redacted: Annotated as CEM by AFP	Rank: 6 Skin Tone: 0.00% Porn: 0.4 Child: 0.99 CAT1 0.64 CAT2 0.3 CAT3 0.01 CAT4 0.0 CAT5 0.03 CAT7 0.0

TEST_CAT1: Bottom 20 Images			
Image	Details	Image	Details
 Redacted: Annotated as CEM by AFP	Rank: 7 Skin Tone: 0.00% Porn: 0.52 Child: 0.99 CAT1 0.05 CAT2 0.94 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.0	File Missing	Rank: 8 Skin Tone: 0.00% Porn: 0.01 Child: 0.99 CAT1 0.4 CAT2 0.0 CAT3 0.27 CAT4 0.16 CAT5 0.01 CAT7 0.13
 Redacted: Annotated as CEM by AFP	Rank: 9 Skin Tone: 0.00% Porn: 0.58 Child: 1.0 CAT1 0.5 CAT2 0.48 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.0	 Redacted: Annotated as CEM by AFP	Rank: 10 Skin Tone: 0.00% Porn: 0.1 Child: 1.0 CAT1 0.66 CAT2 0.0 CAT3 0.32 CAT4 0.0 CAT5 0.0 CAT7 0.0
 Redacted: Annotated as CEM by AFP	Rank: 11 Skin Tone: 0.00% Porn: 0.16 Child: 1.0 CAT1 0.0 CAT2 0.01 CAT3 0.23 CAT4 0.73 CAT5 0.0 CAT7 0.0	 Redacted: Annotated as CEM by AFP	Rank: 12 Skin Tone: 0.00% Porn: 0.76 Child: 0.99 CAT1 0.65 CAT2 0.0 CAT3 0.13 CAT4 0.15 CAT5 0.0 CAT7 0.04

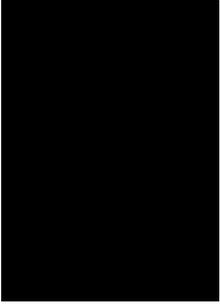
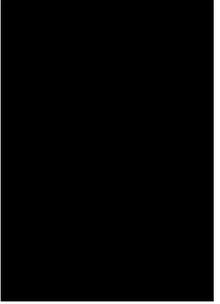
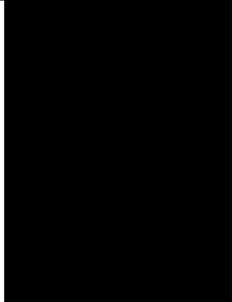
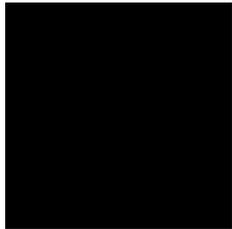
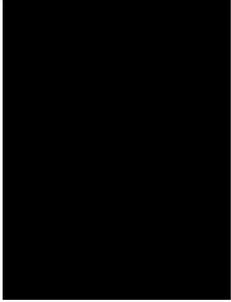
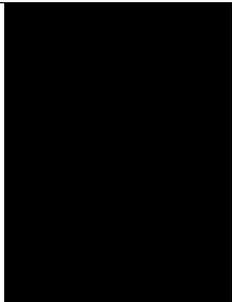
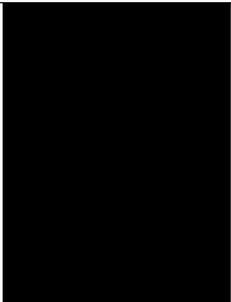
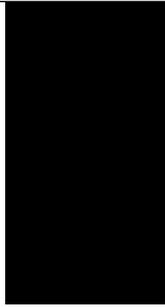
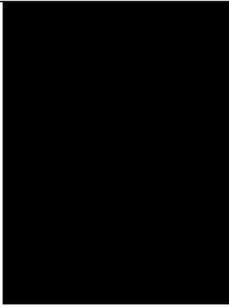
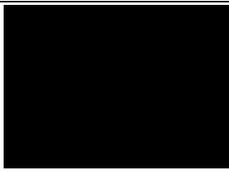
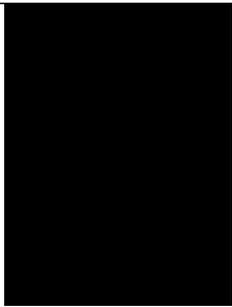
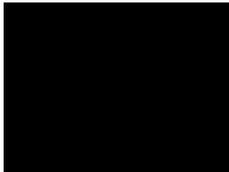
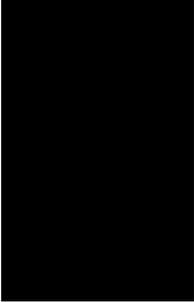
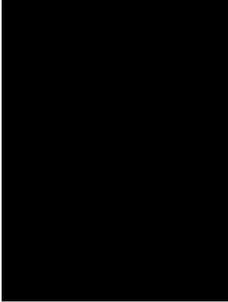
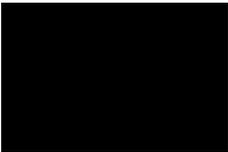
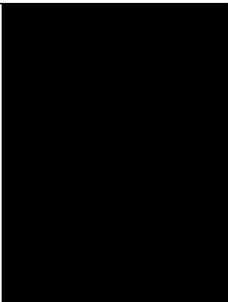
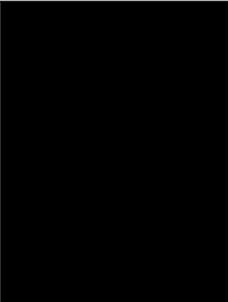
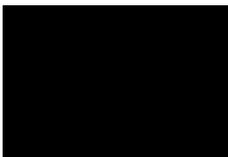
TEST_CAT1: Bottom 20 Images			
Image	Details	Image	Details
	Rank: 19 Skin Tone: 0.00% Porn: 0.25 Child: 1.0 CAT1 0.02 CAT2 0.97 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.0		Rank: 20 Skin Tone: 0.00% Porn: 0.92 Child: 1.0 CAT1 0.99 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.0
Redacted: Annotated as CEM by AFP		Redacted: Annotated as CEM by AFP	

Table B.4: Test corpus CAT1 bottom 20 results by skin tone percentage

B.3 Test Corpus CAT1 Classifier Results

TEST_CAT1: Top 20 Images			
Image	Details	Image	Details
 Redacted: Annotated as CEM by AFP	Rank: 1 Skin Tone: 63.52% Porn: 1.0 Child: 0.01 CAT1 0.0 CAT2 0.0 CAT3 0.99 CAT4 0.0 CAT5 0.0 CAT7 0.0	 Redacted: Annotated as CEM by AFP	Rank: 2 Skin Tone: 36.45% Porn: 0.99 Child: 1.0 CAT1 0.99 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.0
 Redacted: Annotated as CEM by AFP	Rank: 3 Skin Tone: 38.96% Porn: 0.99 Child: 1.0 CAT1 0.99 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.0	 Redacted: Annotated as CEM by AFP	Rank: 4 Skin Tone: 18.78% Porn: 0.99 Child: 1.0 CAT1 0.13 CAT2 0.0 CAT3 0.13 CAT4 0.73 CAT5 0.0 CAT7 0.0
 Redacted: Annotated as CEM by AFP	Rank: 5 Skin Tone: 58.70% Porn: 0.99 Child: 1.0 CAT1 0.21 CAT2 0.0 CAT3 0.54 CAT4 0.23 CAT5 0.0 CAT7 0.0	 Redacted: Annotated as CEM by AFP	Rank: 6 Skin Tone: 35.23% Porn: 0.99 Child: 1.0 CAT1 0.97 CAT2 0.01 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.0

TEST_CAT1: Top 20 Images			
Image	Details	Image	Details
 Redacted: Annotated as CEM by AFP	Rank: 7 Skin Tone: 21.99% Porn: 0.99 Child: 1.0 CAT1 0.97 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.02	 Redacted: Annotated as CEM by AFP	Rank: 8 Skin Tone: 70.01% Porn: 0.99 Child: 1.0 CAT1 0.03 CAT2 0.0 CAT3 0.96 CAT4 0.0 CAT5 0.0 CAT7 0.0
 Redacted: Annotated as CEM by AFP	Rank: 9 Skin Tone: 56.24% Porn: 0.99 Child: 1.0 CAT1 0.01 CAT2 0.98 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.0	 Redacted: Annotated as CEM by AFP	Rank: 10 Skin Tone: 78.30% Porn: 0.99 Child: 1.0 CAT1 0.99 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.0
 Redacted: Annotated as CEM by AFP	Rank: 11 Skin Tone: 63.08% Porn: 0.99 Child: 1.0 CAT1 0.99 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.0	 Redacted: Annotated as CEM by AFP	Rank: 12 Skin Tone: 38.43% Porn: 0.99 Child: 1.0 CAT1 0.16 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.83

TEST_CAT1: Top 20 Images			
Image	Details	Image	Details
 Redacted: Annotated as CEM by AFP	Rank: 13 Skin Tone: 37.62% Porn: 0.99 Child: 1.0 CAT1 0.33 CAT2 0.65 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.0	 Redacted: Annotated as CEM by AFP	Rank: 14 Skin Tone: 34.33% Porn: 0.99 Child: 1.0 CAT1 0.96 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.03
 Redacted: Annotated as CEM by AFP	Rank: 15 Skin Tone: 45.86% Porn: 0.99 Child: 1.0 CAT1 0.0 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.99	 Redacted: Annotated as CEM by AFP	Rank: 16 Skin Tone: 49.72% Porn: 0.99 Child: 1.0 CAT1 0.99 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.0
 Redacted: Annotated as CEM by AFP	Rank: 17 Skin Tone: 62.89% Porn: 0.99 Child: 1.0 CAT1 0.99 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.0	 Redacted: Annotated as CEM by AFP	Rank: 18 Skin Tone: 25.79% Porn: 0.99 Child: 1.0 CAT1 0.19 CAT2 0.79 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.01

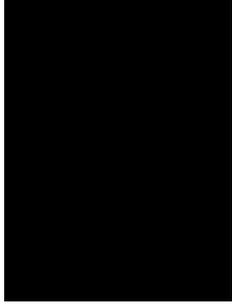
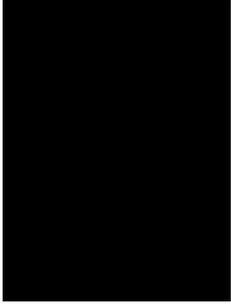
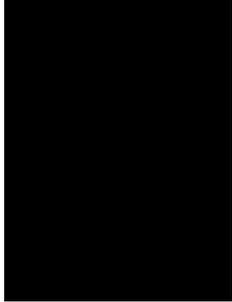
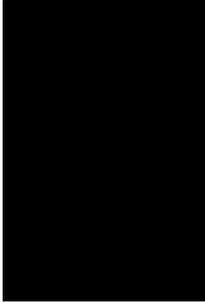
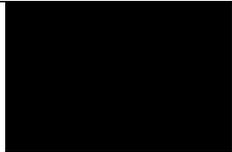
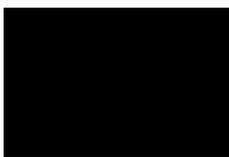
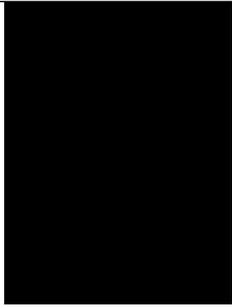
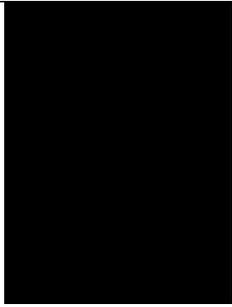
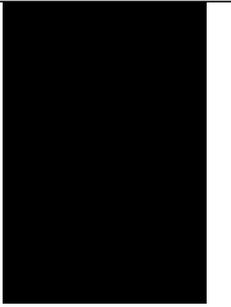
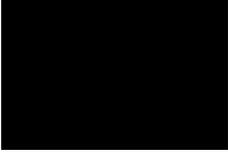
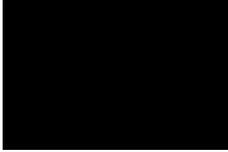
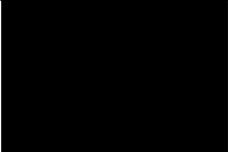
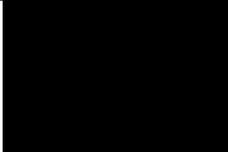
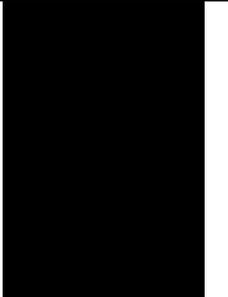
TEST_CAT1: Top 20 Images			
Image	Details	Image	Details
	Rank: 19 Skin Tone: 43.33% Porn: 0.99 Child: 1.0 CAT1 0.99 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.0		Rank: 20 Skin Tone: 60.09% Porn: 0.99 Child: 1.0 CAT1 0.88 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.11 CAT7 0.0
Redacted: Annotated as CEM by AFP		Redacted: Annotated as CEM by AFP	

Table B.5: Test corpus CAT1 top 20 results by classifier (porn,child)

TEST_CAT1: Bottom 20 Images			
Image	Details	Image	Details
	<p>Rank: 1</p> <p>Skin Tone: 0.15%</p> <p>Porn: 0.0</p> <p>Child: 0.0</p> <p>CAT1 0.0</p> <p>CAT2 0.0</p> <p>CAT3 0.15</p> <p>CAT4 0.84</p> <p>CAT5 0.0</p> <p>CAT7 0.0</p>		<p>Rank: 2</p> <p>Skin Tone: 11.66%</p> <p>Porn: 0.0</p> <p>Child: 0.0</p> <p>CAT1 0.67</p> <p>CAT2 0.0</p> <p>CAT3 0.0</p> <p>CAT4 0.0</p> <p>CAT5 0.0</p> <p>CAT7 0.32</p>
Redacted: Annotated as CEM by AFP		Redacted: Annotated as CEM by AFP	
	<p>Rank: 3</p> <p>Skin Tone: 13.85%</p> <p>Porn: 0.0</p> <p>Child: 0.0</p> <p>CAT1 0.0</p> <p>CAT2 0.0</p> <p>CAT3 0.0</p> <p>CAT4 0.0</p> <p>CAT5 0.0</p> <p>CAT7 0.99</p>		<p>Rank: 4</p> <p>Skin Tone: 12.56%</p> <p>Porn: 0.0</p> <p>Child: 0.0</p> <p>CAT1 0.47</p> <p>CAT2 0.0</p> <p>CAT3 0.0</p> <p>CAT4 0.0</p> <p>CAT5 0.0</p> <p>CAT7 0.52</p>
Redacted: Annotated as CEM by AFP		Redacted: Annotated as CEM by AFP	
	<p>Rank: 5</p> <p>Skin Tone: 69.72%</p> <p>Porn: 0.0</p> <p>Child: 0.0</p> <p>CAT1 0.85</p> <p>CAT2 0.01</p> <p>CAT3 0.0</p> <p>CAT4 0.0</p> <p>CAT5 0.0</p> <p>CAT7 0.13</p>		<p>Rank: 6</p> <p>Skin Tone: 9.44%</p> <p>Porn: 0.0</p> <p>Child: 0.0</p> <p>CAT1 0.56</p> <p>CAT2 0.0</p> <p>CAT3 0.0</p> <p>CAT4 0.0</p> <p>CAT5 0.01</p> <p>CAT7 0.41</p>
Redacted: Annotated as CEM by AFP		Redacted: Annotated as CEM by AFP	

TEST_CAT1: Bottom 20 Images			
Image	Details	Image	Details
 Redacted: Annotated as CEM by AFP	Rank: 7 Skin Tone: 14.78% Porn: 0.0 Child: 0.0 CAT1 0.0 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.99	 Redacted: Annotated as CEM by AFP	Rank: 8 Skin Tone: 27.36% Porn: 0.0 Child: 0.0 CAT1 0.75 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.24
 Redacted: Annotated as CEM by AFP	Rank: 9 Skin Tone: 19.17% Porn: 0.0 Child: 0.0 CAT1 0.1 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.89	 Redacted: Annotated as CEM by AFP	Rank: 10 Skin Tone: 10.11% Porn: 0.0 Child: 0.0 CAT1 0.0 CAT2 0.0 CAT3 0.06 CAT4 0.02 CAT5 0.0 CAT7 0.9
 Redacted: Annotated as CEM by AFP	Rank: 11 Skin Tone: 16.87% Porn: 0.0 Child: 0.0 CAT1 0.01 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.98	 Redacted: Annotated as CEM by AFP	Rank: 12 Skin Tone: 12.52% Porn: 0.0 Child: 0.0 CAT1 0.0 CAT2 0.02 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.97

TEST_CAT1: Bottom 20 Images			
Image	Details	Image	Details
 Redacted: Annotated as CEM by AFP	Rank: 13 Skin Tone: 7.58% Porn: 0.0 Child: 0.0 CAT1: 0.0 CAT2: 0.0 CAT3: 0.0 CAT4: 0.0 CAT5: 0.0 CAT7: 0.99	 Redacted: Annotated as CEM by AFP	Rank: 14 Skin Tone: 16.25% Porn: 0.0 Child: 0.0 CAT1: 0.0 CAT2: 0.0 CAT3: 0.0 CAT4: 0.0 CAT5: 0.0 CAT7: 0.99
 Redacted: Annotated as CEM by AFP	Rank: 15 Skin Tone: 12.89% Porn: 0.0 Child: 0.0 CAT1: 0.07 CAT2: 0.0 CAT3: 0.0 CAT4: 0.0 CAT5: 0.0 CAT7: 0.92	 Redacted: Annotated as CEM by AFP	Rank: 16 Skin Tone: 8.12% Porn: 0.0 Child: 0.0 CAT1: 0.0 CAT2: 0.0 CAT3: 0.0 CAT4: 0.0 CAT5: 0.0 CAT7: 0.99
 Redacted: Annotated as CEM by AFP	Rank: 17 Skin Tone: 16.09% Porn: 0.0 Child: 0.0 CAT1: 0.0 CAT2: 0.0 CAT3: 0.0 CAT4: 0.0 CAT5: 0.0 CAT7: 0.99	 Redacted: Annotated as CEM by AFP	Rank: 18 Skin Tone: 21.15% Porn: 0.0 Child: 0.0 CAT1: 0.0 CAT2: 0.0 CAT3: 0.0 CAT4: 0.0 CAT5: 0.0 CAT7: 0.99

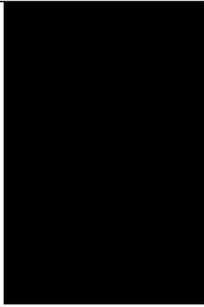
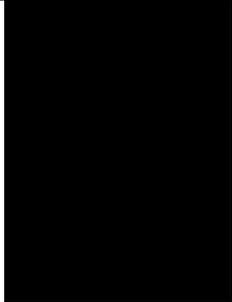
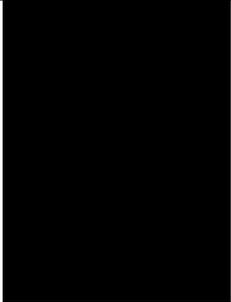
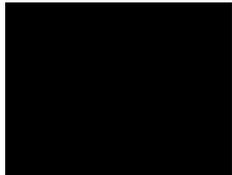
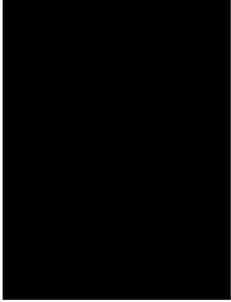
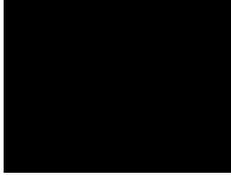
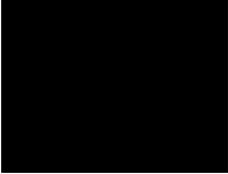
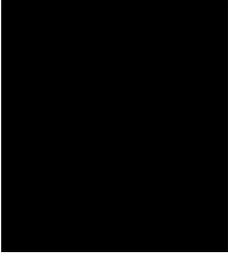
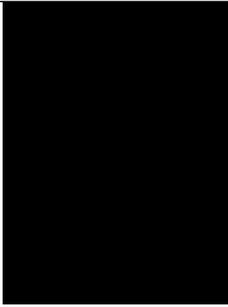
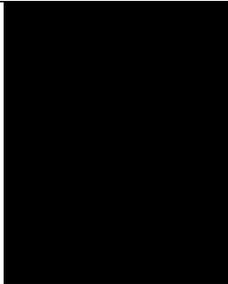
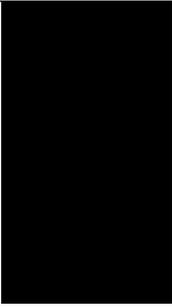
TEST_CAT1: Bottom 20 Images			
Image	Details	Image	Details
	Rank: 19 Skin Tone: 23.57% Porn: 0.0 Child: 0.0 CAT1 0.0 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.99		Rank: 20 Skin Tone: 48.99% Porn: 0.0 Child: 0.0 CAT1 0.0 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.99
Redacted: Annotated as CEM by AFP		Redacted: Annotated as CEM by AFP	

Table B.6: Test corpus CAT1 bottom 20 results by classifier (porn,child)

B.4 Test Corpus CAT2 Skin Tone Results

TEST_CAT2: Top 20 Images			
Image	Details	Image	Details
 Redacted: Annotated as CEM by AFP	Rank: 1 Skin Tone: 98.79% Porn: 0.98 Child: 0.99 CAT1 0.04 CAT2 0.0 CAT3 0.14 CAT4 0.8 CAT5 0.0 CAT7 0.0	 Redacted: Annotated as CEM by AFP	Rank: 2 Skin Tone: 98.67% Porn: 0.97 Child: 0.99 CAT1 0.12 CAT2 0.0 CAT3 0.04 CAT4 0.81 CAT5 0.0 CAT7 0.0
 Redacted: Annotated as CEM by AFP	Rank: 3 Skin Tone: 98.58% Porn: 0.49 Child: 0.99 CAT1 0.18 CAT2 0.55 CAT3 0.25 CAT4 0.0 CAT5 0.0 CAT7 0.0	 Redacted: Annotated as CEM by AFP	Rank: 4 Skin Tone: 98.39% Porn: 0.9 Child: 0.03 CAT1 0.0 CAT2 0.97 CAT3 0.01 CAT4 0.0 CAT5 0.0 CAT7 0.0
 Redacted: Annotated as CEM by AFP	Rank: 5 Skin Tone: 97.55% Porn: 0.93 Child: 0.99 CAT1 0.0 CAT2 0.0 CAT3 0.8 CAT4 0.06 CAT5 0.0 CAT7 0.12	 Redacted: Annotated as CEM by AFP	Rank: 6 Skin Tone: 97.44% Porn: 0.99 Child: 0.99 CAT1 0.15 CAT2 0.07 CAT3 0.66 CAT4 0.1 CAT5 0.0 CAT7 0.0

TEST_CAT2: Top 20 Images			
Image	Details	Image	Details
 Redacted: Annotated as CEM by AFP	Rank: 7 Skin Tone: 97.43% Porn: 0.93 Child: 0.99 CAT1 0.98 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.0	 Redacted: Annotated as CEM by AFP	Rank: 8 Skin Tone: 96.98% Porn: 0.99 Child: 0.99 CAT1 0.01 CAT2 0.0 CAT3 0.94 CAT4 0.0 CAT5 0.0 CAT7 0.03
 Redacted: Annotated as CEM by AFP	Rank: 9 Skin Tone: 96.79% Porn: 0.93 Child: 0.3 CAT1 0.02 CAT2 0.94 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.02	 Redacted: Annotated as CEM by AFP	Rank: 10 Skin Tone: 96.51% Porn: 0.99 Child: 0.87 CAT1 0.01 CAT2 0.79 CAT3 0.16 CAT4 0.01 CAT5 0.0 CAT7 0.0
 Redacted: Annotated as CEM by AFP	Rank: 11 Skin Tone: 96.42% Porn: 0.99 Child: 0.99 CAT1 0.0 CAT2 0.0 CAT3 0.96 CAT4 0.03 CAT5 0.0 CAT7 0.0	 Redacted: Annotated as CEM by AFP	Rank: 12 Skin Tone: 96.23% Porn: 0.99 Child: 0.34 CAT1 0.0 CAT2 0.7 CAT3 0.18 CAT4 0.1 CAT5 0.0 CAT7 0.0

TEST_CAT2: Top 20 Images			
Image	Details	Image	Details
 <p>Redacted: Annotated as CEM by AFP</p>	<p>Rank: 13</p> <p>Skin Tone: 95.88%</p> <p>Porn: 0.99</p> <p>Child: 0.57</p> <p>CAT1 0.0</p> <p>CAT2 0.01</p> <p>CAT3 0.98</p> <p>CAT4 0.0</p> <p>CAT5 0.0</p> <p>CAT7 0.0</p>	 <p>Redacted: Annotated as CEM by AFP</p>	<p>Rank: 14</p> <p>Skin Tone: 95.83%</p> <p>Porn: 0.99</p> <p>Child: 1.0</p> <p>CAT1 0.0</p> <p>CAT2 0.0</p> <p>CAT3 0.03</p> <p>CAT4 0.96</p> <p>CAT5 0.0</p> <p>CAT7 0.0</p>
 <p>Redacted: Annotated as CEM by AFP</p>	<p>Rank: 15</p> <p>Skin Tone: 95.48%</p> <p>Porn: 0.89</p> <p>Child: 0.99</p> <p>CAT1 0.78</p> <p>CAT2 0.21</p> <p>CAT3 0.0</p> <p>CAT4 0.0</p> <p>CAT5 0.0</p> <p>CAT7 0.0</p>	 <p>Redacted: Annotated as CEM by AFP</p>	<p>Rank: 16</p> <p>Skin Tone: 95.12%</p> <p>Porn: 0.87</p> <p>Child: 0.99</p> <p>CAT1 0.99</p> <p>CAT2 0.0</p> <p>CAT3 0.0</p> <p>CAT4 0.0</p> <p>CAT5 0.0</p> <p>CAT7 0.0</p>
 <p>Redacted: Annotated as CEM by AFP</p>	<p>Rank: 17</p> <p>Skin Tone: 94.83%</p> <p>Porn: 0.99</p> <p>Child: 0.99</p> <p>CAT1 0.0</p> <p>CAT2 0.0</p> <p>CAT3 0.99</p> <p>CAT4 0.0</p> <p>CAT5 0.0</p> <p>CAT7 0.0</p>	 <p>Redacted: Annotated as CEM by AFP</p>	<p>Rank: 18</p> <p>Skin Tone: 94.71%</p> <p>Porn: 0.9</p> <p>Child: 0.99</p> <p>CAT1 0.0</p> <p>CAT2 0.0</p> <p>CAT3 0.99</p> <p>CAT4 0.0</p> <p>CAT5 0.0</p> <p>CAT7 0.0</p>

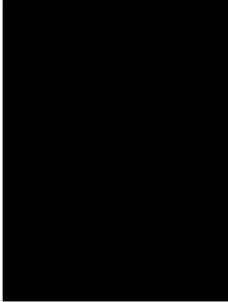
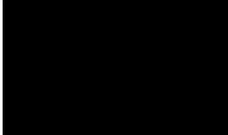
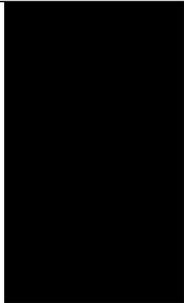
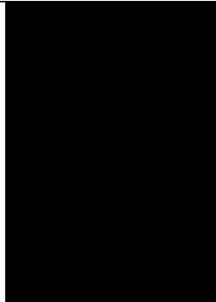
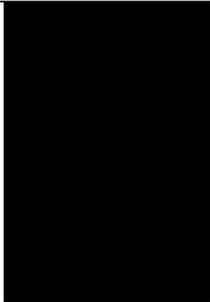
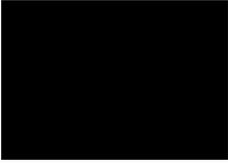
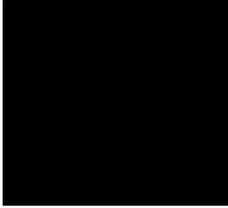
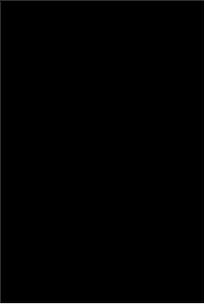
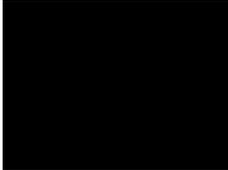
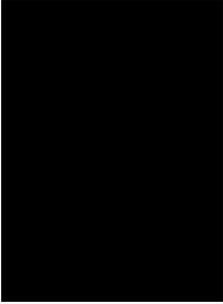
TEST_CAT2: Top 20 Images			
Image	Details	Image	Details
	Rank: 19 Skin Tone: 94.70% Porn: 0.94 Child: 0.99 CAT1 0.71 CAT2 0.26 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.01		Rank: 20 Skin Tone: 94.48% Porn: 0.97 Child: 0.99 CAT1 0.82 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.15
Redacted: Annotated as CEM by AFP		Redacted: Annotated as CEM by AFP	

Table B.7: Test corpus CAT2 top 20 results by skin tone percentage

TEST_CAT2: Bottom 20 Images			
Image	Details	Image	Details
 Redacted: Annotated as CEM by AFP	Rank: 1 Skin Tone: 0.00% Porn: 0.97 Child: 1.0 CAT1 0.0 CAT2 0.88 CAT3 0.02 CAT4 0.09 CAT5 0.0 CAT7 0.0	 Redacted: Annotated as CEM by AFP	Rank: 2 Skin Tone: 0.00% Porn: 0.98 Child: 1.0 CAT1 0.0 CAT2 0.0 CAT3 0.78 CAT4 0.21 CAT5 0.0 CAT7 0.0
 Redacted: Annotated as CEM by AFP	Rank: 3 Skin Tone: 0.00% Porn: 0.89 Child: 0.99 CAT1 0.01 CAT2 0.0 CAT3 0.97 CAT4 0.0 CAT5 0.0 CAT7 0.0	 Redacted: Annotated as CEM by AFP	Rank: 4 Skin Tone: 0.00% Porn: 0.56 Child: 1.0 CAT1 0.0 CAT2 0.0 CAT3 0.14 CAT4 0.83 CAT5 0.0 CAT7 0.0
 Redacted: Annotated as CEM by AFP	Rank: 5 Skin Tone: 0.00% Porn: 0.34 Child: 1.0 CAT1 0.0 CAT2 0.1 CAT3 0.01 CAT4 0.88 CAT5 0.0 CAT7 0.0	 Redacted: Annotated as CEM by AFP	Rank: 6 Skin Tone: 0.00% Porn: 0.49 Child: 1.0 CAT1 0.0 CAT2 0.0 CAT3 0.23 CAT4 0.76 CAT5 0.0 CAT7 0.0

TEST_CAT2: Bottom 20 Images			
Image	Details	Image	Details
	Rank: 7 Skin Tone: 0.00% Porn: 0.34 Child: 1.0 CAT1 0.0 CAT2 0.87 CAT3 0.11 CAT4 0.01 CAT5 0.0 CAT7 0.0		Rank: 8 Skin Tone: 0.00% Porn: 0.97 Child: 1.0 CAT1 0.0 CAT2 0.0 CAT3 0.0 CAT4 0.99 CAT5 0.0 CAT7 0.0
Redacted: Annotated as CEM by AFP		Redacted: Annotated as CEM by AFP	
	Rank: 9 Skin Tone: 0.00% Porn: 0.8 Child: 1.0 CAT1 0.85 CAT2 0.02 CAT3 0.11 CAT4 0.0 CAT5 0.0 CAT7 0.0		Rank: 10 Skin Tone: 0.00% Porn: 0.96 Child: 1.0 CAT1 0.0 CAT2 0.99 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.0
Redacted: Annotated as CEM by AFP		Redacted: Annotated as CEM by AFP	
	Rank: 11 Skin Tone: 0.00% Porn: 0.71 Child: 1.0 CAT1 0.0 CAT2 0.99 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.0		Rank: 12 Skin Tone: 0.00% Porn: 0.44 Child: 1.0 CAT1 0.0 CAT2 0.77 CAT3 0.21 CAT4 0.0 CAT5 0.0 CAT7 0.0
Redacted: Annotated as CEM by AFP		Redacted: Annotated as CEM by AFP	

TEST_CAT2: Bottom 20 Images			
Image	Details	Image	Details
 Redacted: Annotated as CEM by AFP	Rank: 13 Skin Tone: 0.00% Porn: 0.94 Child: 1.0 CAT1: 0.0 CAT2: 0.36 CAT3: 0.1 CAT4: 0.52 CAT5: 0.0 CAT7: 0.0	 Redacted: Annotated as CEM by AFP	Rank: 14 Skin Tone: 0.00% Porn: 0.85 Child: 0.99 CAT1: 0.01 CAT2: 0.97 CAT3: 0.0 CAT4: 0.0 CAT5: 0.0 CAT7: 0.0
 Redacted: Annotated as CEM by AFP	Rank: 15 Skin Tone: 0.00% Porn: 0.78 Child: 1.0 CAT1: 0.0 CAT2: 0.99 CAT3: 0.0 CAT4: 0.0 CAT5: 0.0 CAT7: 0.0	 Redacted: Annotated as CEM by AFP	Rank: 16 Skin Tone: 0.00% Porn: 0.86 Child: 1.0 CAT1: 0.0 CAT2: 0.0 CAT3: 0.26 CAT4: 0.73 CAT5: 0.0 CAT7: 0.0
 Redacted: Annotated as CEM by AFP	Rank: 17 Skin Tone: 0.00% Porn: 0.71 Child: 1.0 CAT1: 0.0 CAT2: 0.0 CAT3: 0.85 CAT4: 0.14 CAT5: 0.0 CAT7: 0.0	 Redacted: Annotated as CEM by AFP	Rank: 18 Skin Tone: 0.00% Porn: 0.16 Child: 1.0 CAT1: 0.01 CAT2: 0.77 CAT3: 0.0 CAT4: 0.21 CAT5: 0.0 CAT7: 0.0

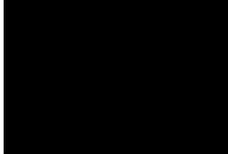
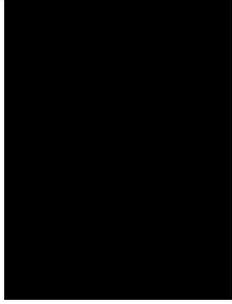
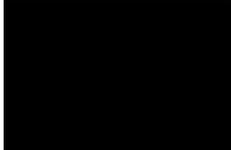
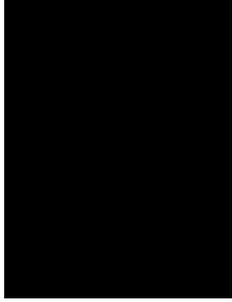
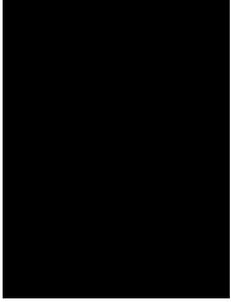
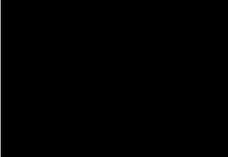
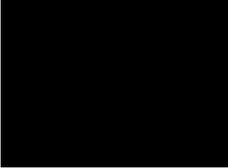
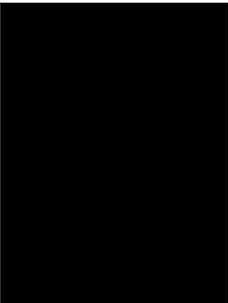
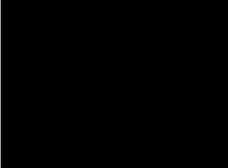
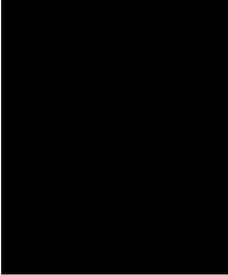
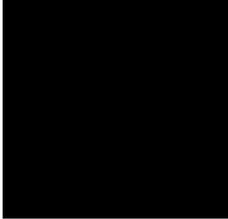
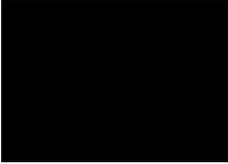
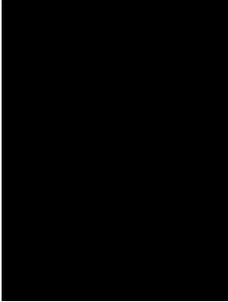
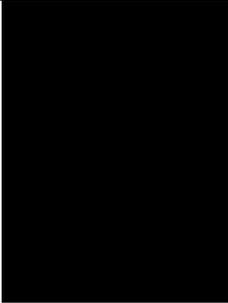
TEST_CAT2: Bottom 20 Images			
Image	Details	Image	Details
	Rank: 19 Skin Tone: 0.00% Porn: 0.89 Child: 1.0 CAT1 0.0 CAT2 0.9 CAT3 0.01 CAT4 0.07 CAT5 0.0 CAT7 0.0		Rank: 20 Skin Tone: 0.00% Porn: 0.08 Child: 0.99 CAT1 0.0 CAT2 0.0 CAT3 0.12 CAT4 0.86 CAT5 0.0 CAT7 0.0
Redacted: Annotated as CEM by AFP		Redacted: Annotated as CEM by AFP	

Table B.8: Test corpus CAT2 bottom 20 results by skin tone percentage

B.5 Test Corpus CAT2 Classifier Results

TEST_CAT2: Top 20 Images			
Image	Details	Image	Details
 Redacted: Annotated as CEM by AFP	Rank: 1 Skin Tone: 58.79% Porn: 0.99 Child: 1.0 CAT1: 0.0 CAT2: 0.83 CAT3: 0.0 CAT4: 0.15 CAT5: 0.0 CAT7: 0.0	 Redacted: Annotated as CEM by AFP	Rank: 2 Skin Tone: 42.28% Porn: 0.99 Child: 1.0 CAT1: 0.0 CAT2: 0.0 CAT3: 0.53 CAT4: 0.46 CAT5: 0.0 CAT7: 0.0
 Redacted: Annotated as CEM by AFP	Rank: 3 Skin Tone: 4.90% Porn: 0.99 Child: 1.0 CAT1: 0.65 CAT2: 0.0 CAT3: 0.26 CAT4: 0.07 CAT5: 0.0 CAT7: 0.0	 Redacted: Annotated as CEM by AFP	Rank: 4 Skin Tone: 46.91% Porn: 0.99 Child: 1.0 CAT1: 0.34 CAT2: 0.0 CAT3: 0.05 CAT4: 0.57 CAT5: 0.0 CAT7: 0.01
 Redacted: Annotated as CEM by AFP	Rank: 5 Skin Tone: 59.98% Porn: 0.99 Child: 1.0 CAT1: 0.02 CAT2: 0.0 CAT3: 0.81 CAT4: 0.15 CAT5: 0.0 CAT7: 0.0	 Redacted: Annotated as CEM by AFP	Rank: 6 Skin Tone: 53.57% Porn: 0.99 Child: 1.0 CAT1: 0.58 CAT2: 0.0 CAT3: 0.02 CAT4: 0.38 CAT5: 0.0 CAT7: 0.0

TEST_CAT2: Top 20 Images			
Image	Details	Image	Details
 Redacted: Annotated as CEM by AFP	Rank: 7 Skin Tone: 5.25% Porn: 0.99 Child: 1.0 CAT1 0.69 CAT2 0.01 CAT3 0.23 CAT4 0.01 CAT5 0.0 CAT7 0.03	 Redacted: Annotated as CEM by AFP	Rank: 8 Skin Tone: 91.40% Porn: 0.99 Child: 1.0 CAT1 0.38 CAT2 0.6 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.0
 Redacted: Annotated as CEM by AFP	Rank: 9 Skin Tone: 43.49% Porn: 0.99 Child: 1.0 CAT1 0.09 CAT2 0.86 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.03	 Redacted: Annotated as CEM by AFP	Rank: 10 Skin Tone: 6.58% Porn: 0.99 Child: 1.0 CAT1 0.0 CAT2 0.99 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.0
 Redacted: Annotated as CEM by AFP	Rank: 11 Skin Tone: 18.56% Porn: 0.99 Child: 1.0 CAT1 0.02 CAT2 0.76 CAT3 0.16 CAT4 0.04 CAT5 0.0 CAT7 0.0	 Redacted: Annotated as CEM by AFP	Rank: 12 Skin Tone: 81.41% Porn: 0.99 Child: 1.0 CAT1 0.0 CAT2 0.0 CAT3 0.01 CAT4 0.98 CAT5 0.0 CAT7 0.0

TEST_CAT2: Top 20 Images			
Image	Details	Image	Details
 Redacted: Annotated as CEM by AFP	Rank: 13 Skin Tone: 55.44% Porn: 0.99 Child: 1.0 CAT1 0.97 CAT2 0.02 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.0	 Redacted: Annotated as CEM by AFP	Rank: 14 Skin Tone: 53.43% Porn: 0.99 Child: 1.0 CAT1 0.99 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.0
 Redacted: Annotated as CEM by AFP	Rank: 15 Skin Tone: 62.15% Porn: 0.99 Child: 1.0 CAT1 0.0 CAT2 0.84 CAT3 0.0 CAT4 0.15 CAT5 0.0 CAT7 0.0	 Redacted: Annotated as CEM by AFP	Rank: 16 Skin Tone: 55.98% Porn: 0.99 Child: 1.0 CAT1 0.0 CAT2 0.99 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.0
 Redacted: Annotated as CEM by AFP	Rank: 17 Skin Tone: 53.78% Porn: 0.99 Child: 1.0 CAT1 0.0 CAT2 0.99 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.0	 Redacted: Annotated as CEM by AFP	Rank: 18 Skin Tone: 39.45% Porn: 0.99 Child: 1.0 CAT1 0.0 CAT2 0.0 CAT3 0.07 CAT4 0.92 CAT5 0.0 CAT7 0.0

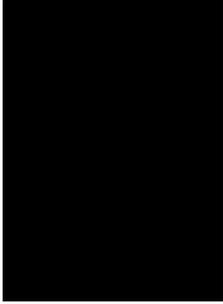
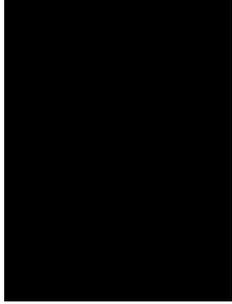
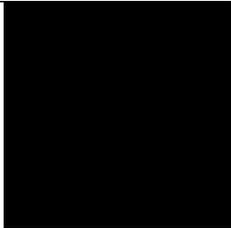
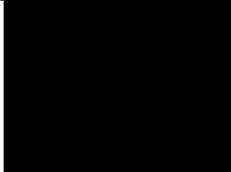
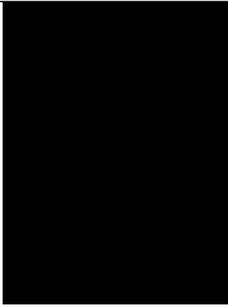
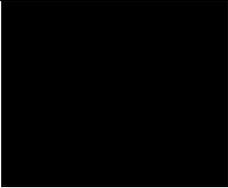
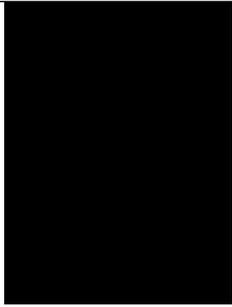
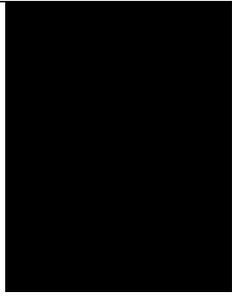
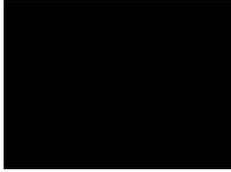
TEST_CAT2: Top 20 Images			
Image	Details	Image	Details
	Rank: 19 Skin Tone: 61.15% Porn: 0.99 Child: 1.0 CAT1 0.42 CAT2 0.54 CAT3 0.03 CAT4 0.0 CAT5 0.0 CAT7 0.0		Rank: 20 Skin Tone: 34.69% Porn: 0.99 Child: 1.0 CAT1 0.01 CAT2 0.98 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.0
Redacted: Annotated as CEM by AFP		Redacted: Annotated as CEM by AFP	

Table B.9: Test corpus CAT2 top 20 results by classifier (porn,child)

TEST_CAT2: Bottom 20 Images			
Image	Details	Image	Details
	Rank: 1 Skin Tone: 8.48% Porn: 0.0 Child: 0.89 CAT1 0.0 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.99		Rank: 2 Skin Tone: 8.61% Porn: 0.0 Child: 0.99 CAT1 0.0 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.99
Redacted: Annotated as CEM by AFP		Redacted: Annotated as CEM by AFP	
	Rank: 3 Skin Tone: 2.78% Porn: 0.0 Child: 0.99 CAT1 0.51 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.48		Rank: 4 Skin Tone: 17.90% Porn: 0.0 Child: 0.99 CAT1 0.01 CAT2 0.0 CAT3 0.01 CAT4 0.0 CAT5 0.01 CAT7 0.96
Redacted: Annotated as CEM by AFP		Redacted: Annotated as CEM by AFP	
	Rank: 5 Skin Tone: 25.78% Porn: 0.0 Child: 1.0 CAT1 0.0 CAT2 0.0 CAT3 0.0 CAT4 0.99 CAT5 0.0 CAT7 0.0		Rank: 6 Skin Tone: 17.80% Porn: 0.01 Child: 0.9 CAT1 0.01 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.98
Redacted: Annotated as CEM by AFP		Redacted: Annotated as CEM by AFP	

TEST_CAT2: Bottom 20 Images			
Image	Details	Image	Details
 Redacted: Annotated as CEM by AFP	Rank: 7 Skin Tone: 10.72% Porn: 0.01 Child: 0.96 CAT1 0.0 CAT2 0.0 CAT3 0.02 CAT4 0.0 CAT5 0.0 CAT7 0.96	 Redacted: Annotated as CEM by AFP	Rank: 8 Skin Tone: 18.78% Porn: 0.01 Child: 0.99 CAT1 0.11 CAT2 0.73 CAT3 0.09 CAT4 0.0 CAT5 0.0 CAT7 0.04
 Redacted: Annotated as CEM by AFP	Rank: 9 Skin Tone: 72.54% Porn: 0.01 Child: 0.99 CAT1 0.22 CAT2 0.0 CAT3 0.39 CAT4 0.38 CAT5 0.0 CAT7 0.0	File Missing	Rank: 10 Skin Tone: 77.27% Porn: 0.01 Child: 0.99 CAT1 0.02 CAT2 0.0 CAT3 0.4 CAT4 0.54 CAT5 0.0 CAT7 0.02
 Redacted: Annotated as CEM by AFP	Rank: 11 Skin Tone: 42.23% Porn: 0.01 Child: 0.99 CAT1 0.0 CAT2 0.0 CAT3 0.15 CAT4 0.04 CAT5 0.0 CAT7 0.78	 Redacted: Annotated as CEM by AFP	Rank: 12 Skin Tone: 9.39% Porn: 0.01 Child: 1.0 CAT1 0.11 CAT2 0.62 CAT3 0.03 CAT4 0.17 CAT5 0.0 CAT7 0.05

TEST_CAT2: Bottom 20 Images			
Image	Details	Image	Details
	<p>Rank: 13</p> <p>Skin Tone: 11.02%</p> <p>Porn: 0.02</p> <p>Child: 0.99</p> <p>CAT1 0.06</p> <p>CAT2 0.0</p> <p>CAT3 0.23</p> <p>CAT4 0.68</p> <p>CAT5 0.0</p> <p>CAT7 0.01</p>		<p>Rank: 14</p> <p>Skin Tone: 30.89%</p> <p>Porn: 0.02</p> <p>Child: 1.0</p> <p>CAT1 0.46</p> <p>CAT2 0.0</p> <p>CAT3 0.12</p> <p>CAT4 0.36</p> <p>CAT5 0.0</p> <p>CAT7 0.04</p>
Redacted: Annotated as CEM by AFP		Redacted: Annotated as CEM by AFP	
	<p>Rank: 15</p> <p>Skin Tone: 24.71%</p> <p>Porn: 0.02</p> <p>Child: 1.0</p> <p>CAT1 0.0</p> <p>CAT2 0.0</p> <p>CAT3 0.0</p> <p>CAT4 1.0</p> <p>CAT5 0.0</p> <p>CAT7 0.0</p>		<p>Rank: 16</p> <p>Skin Tone: 41.21%</p> <p>Porn: 0.03</p> <p>Child: 0.99</p> <p>CAT1 0.0</p> <p>CAT2 0.0</p> <p>CAT3 0.67</p> <p>CAT4 0.09</p> <p>CAT5 0.0</p> <p>CAT7 0.22</p>
Redacted: Annotated as CEM by AFP		Redacted: Annotated as CEM by AFP	
	<p>Rank: 17</p> <p>Skin Tone: 7.51%</p> <p>Porn: 0.03</p> <p>Child: 1.0</p> <p>CAT1 0.01</p> <p>CAT2 0.0</p> <p>CAT3 0.95</p> <p>CAT4 0.03</p> <p>CAT5 0.0</p> <p>CAT7 0.0</p>		<p>Rank: 18</p> <p>Skin Tone: 5.21%</p> <p>Porn: 0.03</p> <p>Child: 1.0</p> <p>CAT1 0.04</p> <p>CAT2 0.0</p> <p>CAT3 0.46</p> <p>CAT4 0.49</p> <p>CAT5 0.0</p> <p>CAT7 0.0</p>
Redacted: Annotated as CEM by AFP		Redacted: Annotated as CEM by AFP	

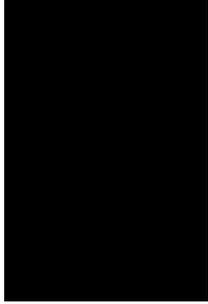
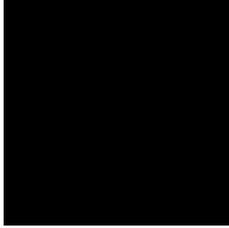
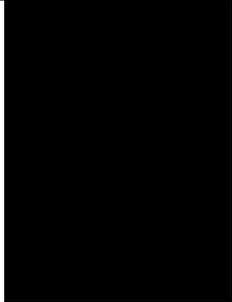
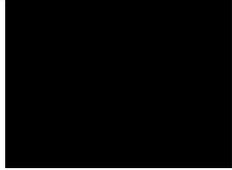
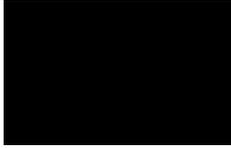
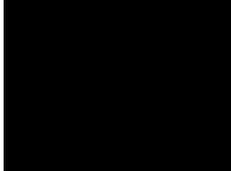
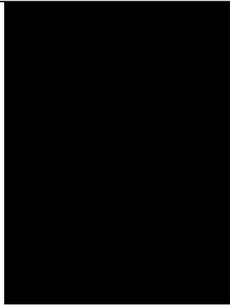
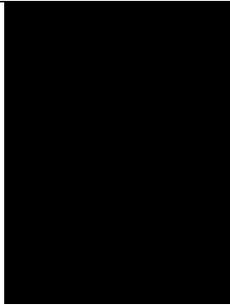
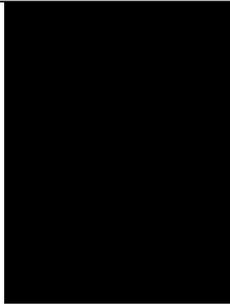
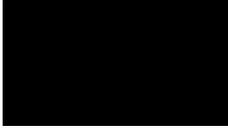
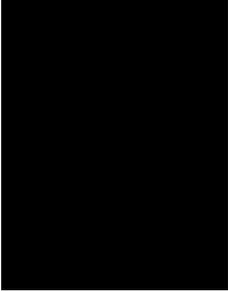
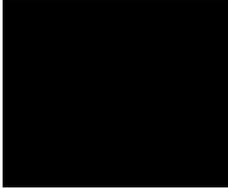
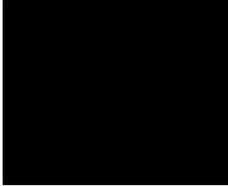
TEST_CAT2: Bottom 20 Images			
Image	Details	Image	Details
	Rank: 19 Skin Tone: 9.05% Porn: 0.03 Child: 1.0 CAT1 0.97 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.02		Rank: 20 Skin Tone: 26.89% Porn: 0.04 Child: 0.19 CAT1 0.19 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.77
Redacted: Annotated as CEM by AFP		Redacted: Annotated as CEM by AFP	

Table B.10: Test corpus CAT2 bottom 20 results by classifier (porn,child)

B.6 Test Corpus CAT3 Skin Tone Results

TEST_CAT3: Top 20 Images			
Image	Details	Image	Details
 Redacted: Annotated as CEM by AFP	Rank: 1 Skin Tone: 99.29% Porn: 0.99 Child: 0.34 CAT1: 0.0 CAT2: 0.91 CAT3: 0.03 CAT4: 0.04 CAT5: 0.0 CAT7: 0.0	 Redacted: Annotated as CEM by AFP	Rank: 2 Skin Tone: 98.62% Porn: 0.99 Child: 0.99 CAT1: 0.0 CAT2: 0.01 CAT3: 0.38 CAT4: 0.59 CAT5: 0.0 CAT7: 0.0
 Redacted: Annotated as CEM by AFP	Rank: 3 Skin Tone: 98.56% Porn: 0.99 Child: 0.99 CAT1: 0.0 CAT2: 0.0 CAT3: 0.11 CAT4: 0.88 CAT5: 0.0 CAT7: 0.0	 Redacted: Annotated as CEM by AFP	Rank: 4 Skin Tone: 98.35% Porn: 0.99 Child: 0.0 CAT1: 0.03 CAT2: 0.96 CAT3: 0.0 CAT4: 0.0 CAT5: 0.0 CAT7: 0.0
 Redacted: Annotated as CEM by AFP	Rank: 5 Skin Tone: 98.13% Porn: 0.94 Child: 0.99 CAT1: 0.0 CAT2: 0.0 CAT3: 0.99 CAT4: 0.0 CAT5: 0.0 CAT7: 0.0	 Redacted: Annotated as CEM by AFP	Rank: 6 Skin Tone: 98.07% Porn: 0.99 Child: 0.88 CAT1: 0.24 CAT2: 0.0 CAT3: 0.02 CAT4: 0.7 CAT5: 0.0 CAT7: 0.03

TEST_CAT3: Top 20 Images			
Image	Details	Image	Details
 Redacted: Annotated as CEM by AFP	Rank: 7 Skin Tone: 97.87% Porn: 0.99 Child: 0.99 CAT1 0.0 CAT2 0.0 CAT3 0.99 CAT4 0.0 CAT5 0.0 CAT7 0.0	 Redacted: Annotated as CEM by AFP	Rank: 8 Skin Tone: 97.13% Porn: 0.96 Child: 0.99 CAT1 0.07 CAT2 0.04 CAT3 0.86 CAT4 0.01 CAT5 0.0 CAT7 0.0
 Redacted: Annotated as CEM by AFP	Rank: 9 Skin Tone: 97.05% Porn: 0.74 Child: 0.99 CAT1 0.85 CAT2 0.14 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.0	 Redacted: Annotated as CEM by AFP	Rank: 10 Skin Tone: 96.88% Porn: 0.67 Child: 0.99 CAT1 0.2 CAT2 0.78 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.0
 Redacted: Annotated as CEM by AFP	Rank: 11 Skin Tone: 96.77% Porn: 0.91 Child: 1.0 CAT1 0.03 CAT2 0.96 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.0	 Redacted: Annotated as CEM by AFP	Rank: 12 Skin Tone: 96.53% Porn: 0.23 Child: 0.99 CAT1 0.92 CAT2 0.0 CAT3 0.04 CAT4 0.01 CAT5 0.0 CAT7 0.01

TEST_CAT3: Top 20 Images			
Image	Details	Image	Details
 Redacted: Annotated as CEM by AFP	Rank: 13 Skin Tone: 96.51% Porn: 0.14 Child: 0.99 CAT1 0.1 CAT2 0.0 CAT3 0.62 CAT4 0.25 CAT5 0.0 CAT7 0.0	 Redacted: Annotated as CEM by AFP	Rank: 14 Skin Tone: 96.45% Porn: 0.11 Child: 0.99 CAT1 0.1 CAT2 0.0 CAT3 0.66 CAT4 0.22 CAT5 0.0 CAT7 0.0
 Redacted: Annotated as CEM by AFP	Rank: 15 Skin Tone: 96.35% Porn: 0.99 Child: 0.99 CAT1 0.0 CAT2 0.0 CAT3 0.0 CAT4 0.98 CAT5 0.0 CAT7 0.0	 Redacted: Annotated as CEM by AFP	Rank: 16 Skin Tone: 96.26% Porn: 0.96 Child: 0.99 CAT1 0.0 CAT2 0.0 CAT3 0.0 CAT4 0.99 CAT5 0.0 CAT7 0.0
 Redacted: Annotated as CEM by AFP	Rank: 17 Skin Tone: 96.09% Porn: 0.99 Child: 0.98 CAT1 0.96 CAT2 0.0 CAT3 0.03 CAT4 0.0 CAT5 0.0 CAT7 0.0	 Redacted: Annotated as CEM by AFP	Rank: 18 Skin Tone: 96.03% Porn: 0.39 Child: 0.51 CAT1 0.0 CAT2 0.0 CAT3 0.81 CAT4 0.18 CAT5 0.0 CAT7 0.0

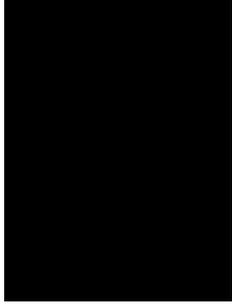
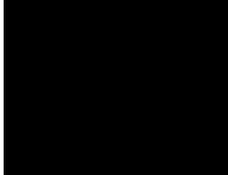
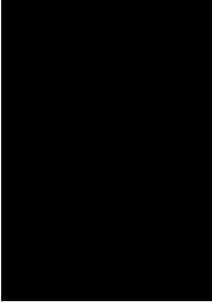
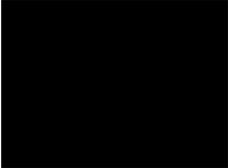
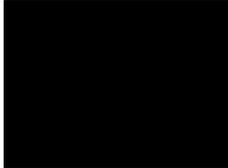
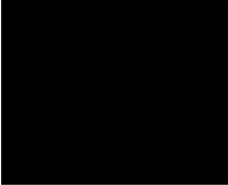
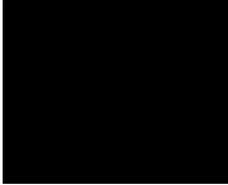
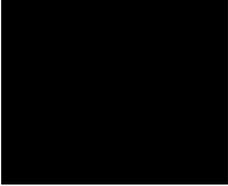
TEST_CAT3: Top 20 Images			
Image	Details	Image	Details
	Rank: 19 Skin Tone: 95.95% Porn: 0.87 Child: 0.99 CAT1 0.02 CAT2 0.97 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.0		Rank: 20 Skin Tone: 95.83% Porn: 0.86 Child: 0.99 CAT1 0.0 CAT2 0.0 CAT3 0.76 CAT4 0.23 CAT5 0.0 CAT7 0.0
Redacted: Annotated as CEM by AFP		Redacted: Annotated as CEM by AFP	

Table B.11: Test corpus CAT3 top 20 results by skin tone percentage

TEST_CAT3: Bottom 20 Images			
Image	Details	Image	Details
 Redacted: Annotated as CEM by AFP	Rank: 7 Skin Tone: 0.00% Porn: 0.99 Child: 1.0 CAT1 0.0 CAT2 0.0 CAT3 0.12 CAT4 0.87 CAT5 0.0 CAT7 0.0	 Redacted: Annotated as CEM by AFP	Rank: 8 Skin Tone: 0.00% Porn: 0.73 Child: 1.0 CAT1 0.0 CAT2 0.0 CAT3 0.1 CAT4 0.89 CAT5 0.0 CAT7 0.0
 Redacted: Annotated as CEM by AFP	Rank: 9 Skin Tone: 0.00% Porn: 0.59 Child: 1.0 CAT1 0.0 CAT2 0.72 CAT3 0.0 CAT4 0.26 CAT5 0.0 CAT7 0.0	 Redacted: Annotated as CEM by AFP	Rank: 10 Skin Tone: 0.00% Porn: 0.05 Child: 1.0 CAT1 0.01 CAT2 0.0 CAT3 0.89 CAT4 0.08 CAT5 0.0 CAT7 0.0
 Redacted: Annotated as CEM by AFP	Rank: 11 Skin Tone: 0.00% Porn: 0.99 Child: 1.0 CAT1 0.0 CAT2 0.0 CAT3 0.71 CAT4 0.27 CAT5 0.0 CAT7 0.0	 Redacted: Annotated as CEM by AFP	Rank: 12 Skin Tone: 0.00% Porn: 0.6 Child: 1.0 CAT1 0.03 CAT2 0.0 CAT3 0.5 CAT4 0.46 CAT5 0.0 CAT7 0.0

TEST_CAT3: Bottom 20 Images			
Image	Details	Image	Details
 Redacted: Annotated as CEM by AFP	Rank: 13 Skin Tone: 0.00% Porn: 0.97 Child: 1.0 CAT1: 0.2 CAT2: 0.0 CAT3: 0.2 CAT4: 0.58 CAT5: 0.0 CAT7: 0.0	 Redacted: Annotated as CEM by AFP	Rank: 14 Skin Tone: 0.00% Porn: 0.45 Child: 1.0 CAT1: 0.12 CAT2: 0.0 CAT3: 0.86 CAT4: 0.0 CAT5: 0.0 CAT7: 0.0
 Redacted: Annotated as CEM by AFP	Rank: 15 Skin Tone: 0.00% Porn: 0.97 Child: 1.0 CAT1: 0.3 CAT2: 0.0 CAT3: 0.26 CAT4: 0.42 CAT5: 0.0 CAT7: 0.0	 Redacted: Annotated as CEM by AFP	Rank: 16 Skin Tone: 0.01% Porn: 0.2 Child: 1.0 CAT1: 0.0 CAT2: 0.0 CAT3: 0.3 CAT4: 0.68 CAT5: 0.0 CAT7: 0.0
 Redacted: Annotated as CEM by AFP	Rank: 17 Skin Tone: 0.01% Porn: 0.98 Child: 1.0 CAT1: 0.0 CAT2: 0.0 CAT3: 0.47 CAT4: 0.52 CAT5: 0.0 CAT7: 0.0	 Redacted: Annotated as CEM by AFP	Rank: 18 Skin Tone: 0.02% Porn: 0.97 Child: 1.0 CAT1: 0.0 CAT2: 0.0 CAT3: 0.33 CAT4: 0.64 CAT5: 0.0 CAT7: 0.0

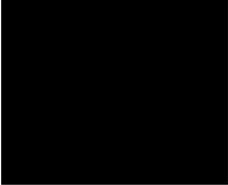
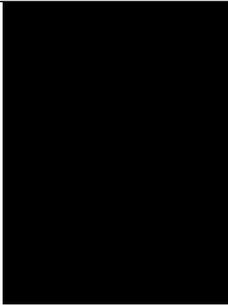
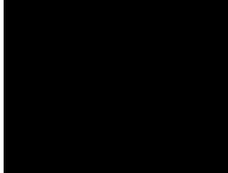
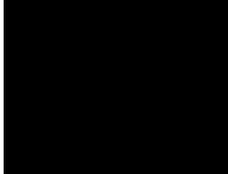
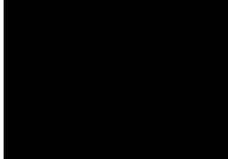
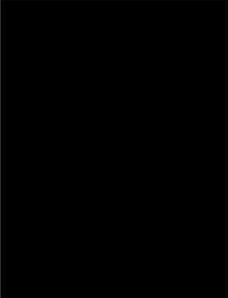
TEST_CAT3: Bottom 20 Images			
Image	Details	Image	Details
	Rank: 19 Skin Tone: 0.02% Porn: 0.88 Child: 1.0 CAT1 0.24 CAT2 0.0 CAT3 0.49 CAT4 0.25 CAT5 0.0 CAT7 0.0		Rank: 20 Skin Tone: 0.04% Porn: 0.81 Child: 0.99 CAT1 0.0 CAT2 0.0 CAT3 0.08 CAT4 0.91 CAT5 0.0 CAT7 0.0
Redacted: Annotated as CEM by AFP		Redacted: Annotated as CEM by AFP	

Table B.12: Test corpus CAT3 bottom 20 results by skin tone percentage

TEST_CAT3: Top 20 Images			
Image	Details	Image	Details
 Redacted: Annotated as CEM by AFP	Rank: 7 Skin Tone: 25.15% Porn: 0.99 Child: 1.0 CAT1 0.0 CAT2 0.0 CAT3 0.99 CAT4 0.0 CAT5 0.0 CAT7 0.0	 Redacted: Annotated as CEM by AFP	Rank: 8 Skin Tone: 60.87% Porn: 0.99 Child: 1.0 CAT1 0.0 CAT2 0.0 CAT3 0.99 CAT4 0.0 CAT5 0.0 CAT7 0.0
 Redacted: Annotated as CEM by AFP	Rank: 9 Skin Tone: 80.62% Porn: 0.99 Child: 1.0 CAT1 0.0 CAT2 0.0 CAT3 0.99 CAT4 0.0 CAT5 0.0 CAT7 0.0	 Redacted: Annotated as CEM by AFP	Rank: 10 Skin Tone: 55.67% Porn: 0.99 Child: 1.0 CAT1 0.0 CAT2 0.0 CAT3 0.99 CAT4 0.0 CAT5 0.0 CAT7 0.0
 Redacted: Annotated as CEM by AFP	Rank: 11 Skin Tone: 49.64% Porn: 0.99 Child: 1.0 CAT1 0.0 CAT2 0.0 CAT3 0.92 CAT4 0.07 CAT5 0.0 CAT7 0.0	 Redacted: Annotated as CEM by AFP	Rank: 12 Skin Tone: 58.35% Porn: 0.99 Child: 1.0 CAT1 0.0 CAT2 0.0 CAT3 0.99 CAT4 0.0 CAT5 0.0 CAT7 0.0

TEST_CAT3: Top 20 Images			
Image	Details	Image	Details
 Redacted: Annotated as CEM by AFP	Rank: 13 Skin Tone: 37.02% Porn: 0.99 Child: 1.0 CAT1 0.0 CAT2 0.0 CAT3 0.99 CAT4 0.0 CAT5 0.0 CAT7 0.0	 Redacted: Annotated as CEM by AFP	Rank: 14 Skin Tone: 57.05% Porn: 0.99 Child: 1.0 CAT1 0.15 CAT2 0.0 CAT3 0.78 CAT4 0.06 CAT5 0.0 CAT7 0.0
 Redacted: Annotated as CEM by AFP	Rank: 15 Skin Tone: 60.96% Porn: 0.99 Child: 1.0 CAT1 0.0 CAT2 0.0 CAT3 0.99 CAT4 0.0 CAT5 0.0 CAT7 0.0	 Redacted: Annotated as CEM by AFP	Rank: 16 Skin Tone: 56.12% Porn: 0.99 Child: 1.0 CAT1 0.04 CAT2 0.0 CAT3 0.39 CAT4 0.56 CAT5 0.0 CAT7 0.0
 Redacted: Annotated as CEM by AFP	Rank: 17 Skin Tone: 39.21% Porn: 0.99 Child: 1.0 CAT1 0.9 CAT2 0.02 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.07	 Redacted: Annotated as CEM by AFP	Rank: 18 Skin Tone: 45.27% Porn: 0.99 Child: 1.0 CAT1 0.0 CAT2 0.0 CAT3 0.99 CAT4 0.0 CAT5 0.0 CAT7 0.0

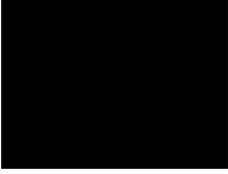
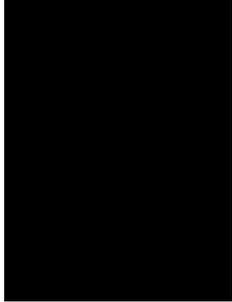
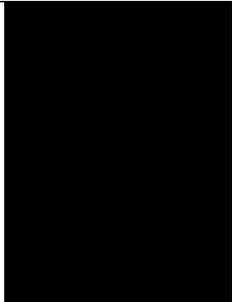
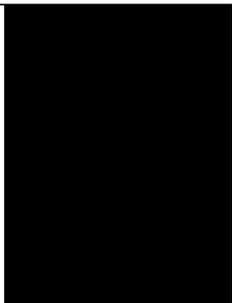
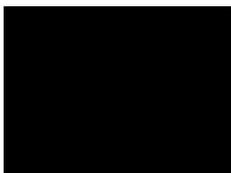
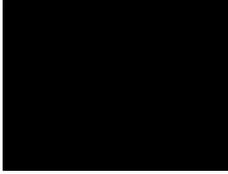
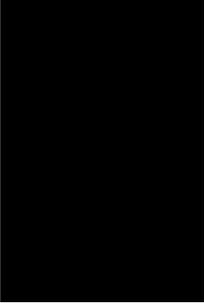
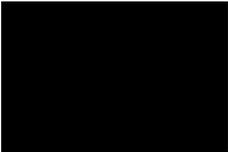
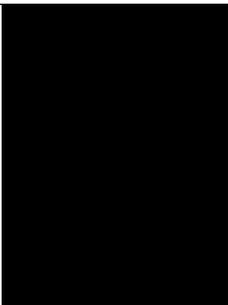
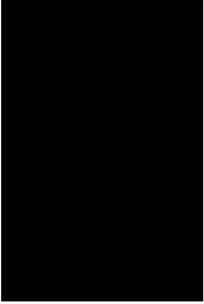
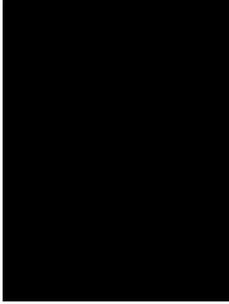
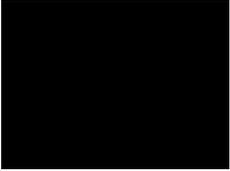
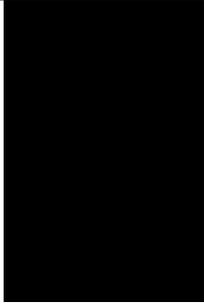
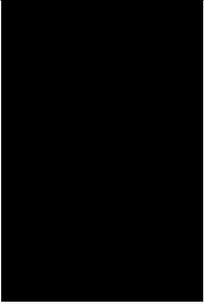
TEST_CAT3: Top 20 Images			
Image	Details	Image	Details
	Rank: 19 Skin Tone: 42.04% Porn: 0.99 Child: 1.0 CAT1 0.0 CAT2 0.0 CAT3 0.0 CAT4 0.99 CAT5 0.0 CAT7 0.0		Rank: 20 Skin Tone: 70.12% Porn: 0.99 Child: 1.0 CAT1 0.0 CAT2 0.0 CAT3 0.0 CAT4 0.99 CAT5 0.0 CAT7 0.0
Redacted: Annotated as CEM by AFP		Redacted: Annotated as CEM by AFP	

Table B.13: Test corpus CAT3 top 20 results by classifier (porn,child)

TEST_CAT3: Bottom 20 Images			
Image	Details	Image	Details
 Redacted: Annotated as CEM by AFP	Rank: 1 Skin Tone: 14.44% Porn: 0.0 Child: 0.99 CAT1 0.0 CAT2 0.0 CAT3 0.06 CAT4 0.93 CAT5 0.0 CAT7 0.0	 Redacted: Annotated as CEM by AFP	Rank: 2 Skin Tone: 54.77% Porn: 0.0 Child: 1.0 CAT1 0.0 CAT2 0.0 CAT3 0.64 CAT4 0.34 CAT5 0.0 CAT7 0.0
 Redacted: Annotated as CEM by AFP	Rank: 3 Skin Tone: 11.46% Porn: 0.0 Child: 1.0 CAT1 0.0 CAT2 0.0 CAT3 0.0 CAT4 0.99 CAT5 0.0 CAT7 0.0	File Missing	Rank: 4 Skin Tone: 44.33% Porn: 0.0 Child: 1.0 CAT1 0.01 CAT2 0.0 CAT3 0.87 CAT4 0.0 CAT5 0.0 CAT7 0.09
 Redacted: Annotated as CEM by AFP	Rank: 5 Skin Tone: 20.03% Porn: 0.0 Child: 1.0 CAT1 0.0 CAT2 0.0 CAT3 0.68 CAT4 0.31 CAT5 0.0 CAT7 0.0	 Redacted: Annotated as CEM by AFP	Rank: 6 Skin Tone: 80.91% Porn: 0.01 Child: 0.25 CAT1 0.54 CAT2 0.0 CAT3 0.12 CAT4 0.12 CAT5 0.0 CAT7 0.19

TEST_CAT3: Bottom 20 Images			
Image	Details	Image	Details
 Redacted: Annotated as CEM by AFP	Rank: 7 Skin Tone: 26.19% Porn: 0.01 Child: 0.43 CAT1 0.13 CAT2 0.0 CAT3 0.55 CAT4 0.09 CAT5 0.0 CAT7 0.22	 Redacted: Annotated as CEM by AFP	Rank: 8 Skin Tone: 33.31% Porn: 0.01 Child: 0.99 CAT1 0.01 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.98
 Redacted: Annotated as CEM by AFP	Rank: 9 Skin Tone: 28.20% Porn: 0.01 Child: 0.99 CAT1 0.88 CAT2 0.0 CAT3 0.0 CAT4 0.02 CAT5 0.0 CAT7 0.07	 Redacted: Annotated as CEM by AFP	Rank: 10 Skin Tone: 24.68% Porn: 0.01 Child: 0.99 CAT1 0.0 CAT2 0.0 CAT3 0.99 CAT4 0.0 CAT5 0.0 CAT7 0.0
 Redacted: Annotated as CEM by AFP	Rank: 11 Skin Tone: 55.94% Porn: 0.01 Child: 1.0 CAT1 0.43 CAT2 0.0 CAT3 0.0 CAT4 0.55 CAT5 0.0 CAT7 0.0	 Redacted: Annotated as CEM by AFP	Rank: 12 Skin Tone: 16.61% Porn: 0.01 Child: 1.0 CAT1 0.0 CAT2 0.0 CAT3 0.95 CAT4 0.04 CAT5 0.0 CAT7 0.0

TEST_CAT3: Bottom 20 Images			
Image	Details	Image	Details
 Redacted: Annotated as CEM by AFP	Rank: 13 Skin Tone: 67.66% Porn: 0.01 Child: 1.0 CAT1 0.01 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.98	 Redacted: Annotated as CEM by AFP	Rank: 14 Skin Tone: 26.08% Porn: 0.01 Child: 1.0 CAT1 0.0 CAT2 0.0 CAT3 0.0 CAT4 0.99 CAT5 0.0 CAT7 0.0
 Redacted: Annotated as CEM by AFP	Rank: 15 Skin Tone: 32.85% Porn: 0.02 Child: 0.86 CAT1 0.02 CAT2 0.0 CAT3 0.89 CAT4 0.02 CAT5 0.0 CAT7 0.05	 Redacted: Annotated as CEM by AFP	Rank: 16 Skin Tone: 40.91% Porn: 0.02 Child: 0.94 CAT1 0.68 CAT2 0.0 CAT3 0.1 CAT4 0.0 CAT5 0.0 CAT7 0.2
 Redacted: Annotated as CEM by AFP	Rank: 17 Skin Tone: 14.56% Porn: 0.02 Child: 0.99 CAT1 0.01 CAT2 0.84 CAT3 0.0 CAT4 0.1 CAT5 0.0 CAT7 0.02	 Redacted: Annotated as CEM by AFP	Rank: 18 Skin Tone: 45.29% Porn: 0.02 Child: 0.99 CAT1 0.0 CAT2 0.0 CAT3 0.86 CAT4 0.13 CAT5 0.0 CAT7 0.0

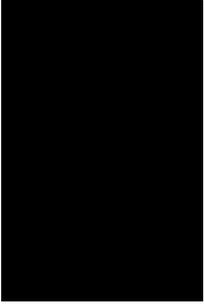
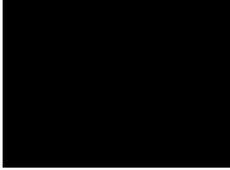
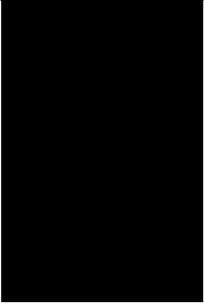
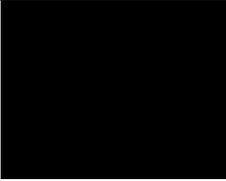
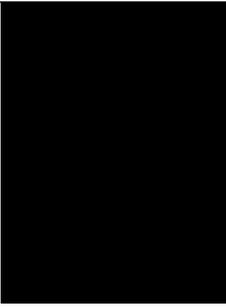
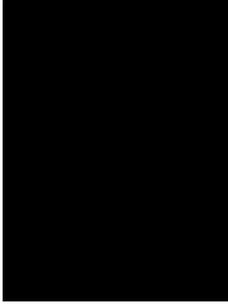
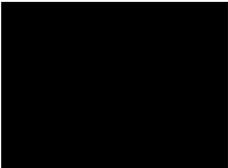
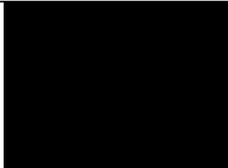
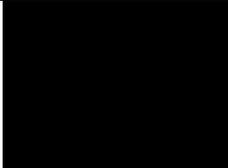
TEST_CAT3: Bottom 20 Images			
Image	Details	Image	Details
	Rank: 19 Skin Tone: 74.90% Porn: 0.02 Child: 0.99 CAT1 0.33 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.66		Rank: 20 Skin Tone: 76.80% Porn: 0.02 Child: 1.0 CAT1 0.06 CAT2 0.85 CAT3 0.01 CAT4 0.05 CAT5 0.0 CAT7 0.0
Redacted: Annotated as CEM by AFP		Redacted: Annotated as CEM by AFP	

Table B.14: Test corpus CAT3 bottom 20 results by classifier (porn,child)

B.8 Test Corpus CAT4 Skin Tone Results

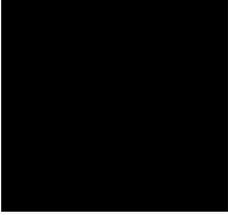
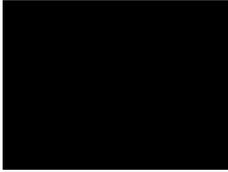
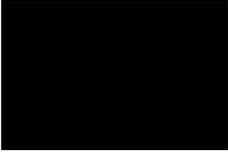
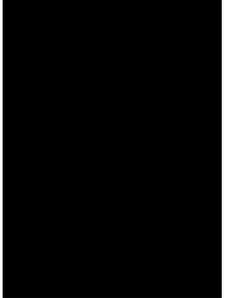
TEST_CAT4: Top 20 Images			
Image	Details	Image	Details
 Redacted: Annotated as CEM by AFP	Rank: 1 Skin Tone: 98.55% Porn: 0.99 Child: 0.96 CAT1: 0.0 CAT2: 0.99 CAT3: 0.0 CAT4: 0.0 CAT5: 0.0 CAT7: 0.0	 Redacted: Annotated as CEM by AFP	Rank: 2 Skin Tone: 98.30% Porn: 0.86 Child: 0.99 CAT1: 0.0 CAT2: 0.0 CAT3: 0.99 CAT4: 0.0 CAT5: 0.0 CAT7: 0.0
 Redacted: Annotated as CEM by AFP	Rank: 3 Skin Tone: 98.17% Porn: 0.99 Child: 0.98 CAT1: 0.0 CAT2: 0.0 CAT3: 0.07 CAT4: 0.92 CAT5: 0.0 CAT7: 0.0	 Redacted: Annotated as CEM by AFP	Rank: 4 Skin Tone: 98.09% Porn: 0.78 Child: 0.36 CAT1: 0.01 CAT2: 0.58 CAT3: 0.27 CAT4: 0.12 CAT5: 0.0 CAT7: 0.0
 Redacted: Annotated as CEM by AFP	Rank: 5 Skin Tone: 98.07% Porn: 0.93 Child: 0.99 CAT1: 0.0 CAT2: 0.0 CAT3: 0.08 CAT4: 0.91 CAT5: 0.0 CAT7: 0.0	 Redacted: Annotated as CEM by AFP	Rank: 6 Skin Tone: 97.99% Porn: 0.99 Child: 0.95 CAT1: 0.08 CAT2: 0.02 CAT3: 0.88 CAT4: 0.0 CAT5: 0.0 CAT7: 0.0

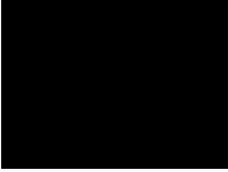
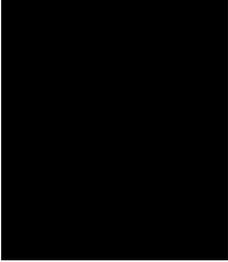
TEST_CAT4: Top 20 Images			
Image	Details	Image	Details
 Redacted: Annotated as CEM by AFP	Rank: 7 Skin Tone: 97.96% Porn: 0.99 Child: 0.99 CAT1 0.0 CAT2 0.0 CAT3 0.25 CAT4 0.74 CAT5 0.0 CAT7 0.0	 Redacted: Annotated as CEM by AFP	Rank: 8 Skin Tone: 97.96% Porn: 0.99 Child: 0.92 CAT1 0.08 CAT2 0.05 CAT3 0.85 CAT4 0.0 CAT5 0.0 CAT7 0.0
 Redacted: Annotated as CEM by AFP	Rank: 9 Skin Tone: 97.80% Porn: 0.97 Child: 1.0 CAT1 0.0 CAT2 0.0 CAT3 0.09 CAT4 0.9 CAT5 0.0 CAT7 0.0	 Redacted: Annotated as CEM by AFP	Rank: 10 Skin Tone: 97.14% Porn: 0.99 Child: 0.99 CAT1 0.27 CAT2 0.04 CAT3 0.0 CAT4 0.6 CAT5 0.0 CAT7 0.08
 Redacted: Annotated as CEM by AFP	Rank: 11 Skin Tone: 97.05% Porn: 0.99 Child: 0.96 CAT1 0.0 CAT2 0.0 CAT3 0.0 CAT4 0.99 CAT5 0.0 CAT7 0.0	 Redacted: Annotated as CEM by AFP	Rank: 12 Skin Tone: 96.84% Porn: 0.9 Child: 0.8 CAT1 0.0 CAT2 0.0 CAT3 0.99 CAT4 0.0 CAT5 0.0 CAT7 0.0

TEST_CAT4: Top 20 Images			
Image	Details	Image	Details
 Redacted: Annotated as CEM by AFP	Rank: 13 Skin Tone: 96.80% Porn: 0.96 Child: 0.99 CAT1 0.04 CAT2 0.03 CAT3 0.83 CAT4 0.07 CAT5 0.0 CAT7 0.01	 Redacted: Annotated as CEM by AFP	Rank: 14 Skin Tone: 96.78% Porn: 0.99 Child: 1.0 CAT1 0.0 CAT2 0.0 CAT3 0.71 CAT4 0.28 CAT5 0.0 CAT7 0.0
 Redacted: Annotated as CEM by AFP	Rank: 15 Skin Tone: 96.62% Porn: 0.99 Child: 1.0 CAT1 0.0 CAT2 0.0 CAT3 0.11 CAT4 0.88 CAT5 0.0 CAT7 0.0	 Redacted: Annotated as CEM by AFP	Rank: 16 Skin Tone: 96.47% Porn: 0.99 Child: 0.94 CAT1 0.0 CAT2 0.0 CAT3 0.0 CAT4 0.99 CAT5 0.0 CAT7 0.0
 Redacted: Annotated as CEM by AFP	Rank: 17 Skin Tone: 96.36% Porn: 0.82 Child: 1.0 CAT1 0.02 CAT2 0.0 CAT3 0.86 CAT4 0.1 CAT5 0.0 CAT7 0.0	 Redacted: Annotated as CEM by AFP	Rank: 18 Skin Tone: 96.19% Porn: 0.85 Child: 0.99 CAT1 0.0 CAT2 0.0 CAT3 0.99 CAT4 0.0 CAT5 0.0 CAT7 0.0

TEST_CAT4: Top 20 Images			
Image	Details	Image	Details
	Rank: 19 Skin Tone: 96.18% Porn: 0.81 Child: 0.99 CAT1 0.0 CAT2 0.0 CAT3 0.14 CAT4 0.85 CAT5 0.0 CAT7 0.0		Rank: 20 Skin Tone: 96.07% Porn: 0.79 Child: 0.99 CAT1 0.0 CAT2 0.0 CAT3 0.03 CAT4 0.96 CAT5 0.0 CAT7 0.0
Redacted: Annotated as CEM by AFP		Redacted: Annotated as CEM by AFP	

Table B.15: Test corpus CAT4 top 20 results by skin tone percentage

TEST_CAT4: Bottom 20 Images			
Image	Details	Image	Details
 Redacted: Annotated as CEM by AFP	Rank: 7 Skin Tone: 0.00% Porn: 0.92 Child: 1.0 CAT1 0.0 CAT2 0.0 CAT3 0.37 CAT4 0.62 CAT5 0.0 CAT7 0.0	 Redacted: Annotated as CEM by AFP	Rank: 8 Skin Tone: 0.00% Porn: 0.92 Child: 1.0 CAT1 0.84 CAT2 0.13 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.02
 Redacted: Annotated as CEM by AFP	Rank: 9 Skin Tone: 0.00% Porn: 0.85 Child: 1.0 CAT1 0.0 CAT2 0.0 CAT3 0.0 CAT4 0.99 CAT5 0.0 CAT7 0.0	 Redacted: Annotated as CEM by AFP	Rank: 10 Skin Tone: 0.00% Porn: 0.11 Child: 1.0 CAT1 0.0 CAT2 0.0 CAT3 0.02 CAT4 0.97 CAT5 0.0 CAT7 0.0
 Redacted: Annotated as CEM by AFP	Rank: 11 Skin Tone: 0.00% Porn: 0.79 Child: 1.0 CAT1 0.0 CAT2 0.0 CAT3 0.53 CAT4 0.46 CAT5 0.0 CAT7 0.0	 Redacted: Annotated as CEM by AFP	Rank: 12 Skin Tone: 0.00% Porn: 0.62 Child: 1.0 CAT1 0.0 CAT2 0.0 CAT3 0.89 CAT4 0.1 CAT5 0.0 CAT7 0.0

TEST_CAT4: Bottom 20 Images			
Image	Details	Image	Details
 Redacted: Annotated as CEM by AFP	Rank: 13 Skin Tone: 0.00% Porn: 0.98 Child: 1.0 CAT1 0.0 CAT2 0.0 CAT3 0.2 CAT4 0.79 CAT5 0.0 CAT7 0.0	 Redacted: Annotated as CEM by AFP	Rank: 14 Skin Tone: 0.00% Porn: 0.23 Child: 0.99 CAT1 0.0 CAT2 0.0 CAT3 0.65 CAT4 0.34 CAT5 0.0 CAT7 0.0
 Redacted: Annotated as CEM by AFP	Rank: 15 Skin Tone: 0.00% Porn: 0.82 Child: 0.99 CAT1 0.01 CAT2 0.0 CAT3 0.02 CAT4 0.95 CAT5 0.0 CAT7 0.0	 Redacted: Annotated as CEM by AFP	Rank: 16 Skin Tone: 0.00% Porn: 0.97 Child: 1.0 CAT1 0.0 CAT2 0.0 CAT3 0.0 CAT4 0.99 CAT5 0.0 CAT7 0.0
 Redacted: Annotated as CEM by AFP	Rank: 17 Skin Tone: 0.00% Porn: 0.92 Child: 1.0 CAT1 0.01 CAT2 0.0 CAT3 0.41 CAT4 0.57 CAT5 0.0 CAT7 0.0	 Redacted: Annotated as CEM by AFP	Rank: 18 Skin Tone: 0.00% Porn: 0.99 Child: 1.0 CAT1 0.0 CAT2 0.0 CAT3 0.0 CAT4 0.99 CAT5 0.0 CAT7 0.0

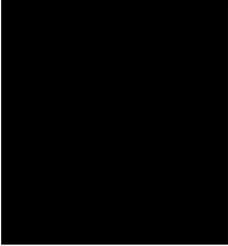
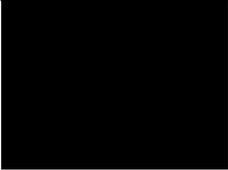
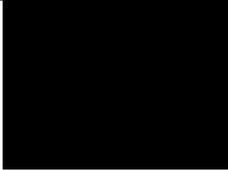
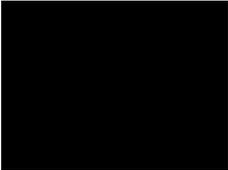
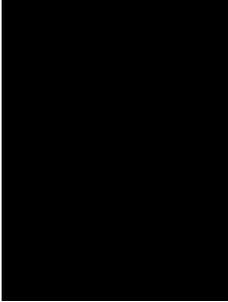
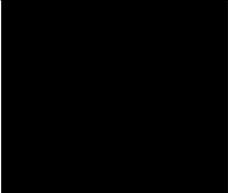
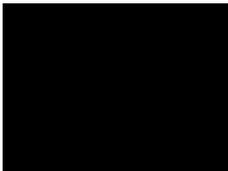
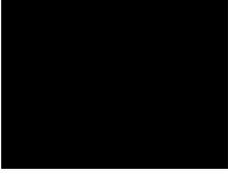
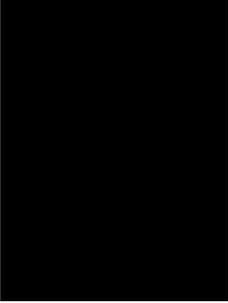
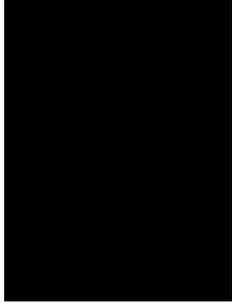
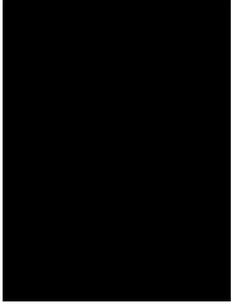
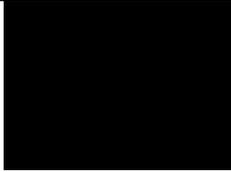
TEST_CAT4: Bottom 20 Images			
Image	Details	Image	Details
	Rank: 19 Skin Tone: 0.00% Porn: 0.14 Child: 1.0 CAT1 0.0 CAT2 0.0 CAT3 0.79 CAT4 0.19 CAT5 0.0 CAT7 0.0		Rank: 20 Skin Tone: 0.00% Porn: 0.96 Child: 1.0 CAT1 0.0 CAT2 0.0 CAT3 0.12 CAT4 0.86 CAT5 0.0 CAT7 0.0
Redacted: Annotated as CEM by AFP		Redacted: Annotated as CEM by AFP	

Table B.16: Test corpus CAT4 bottom 20 results by skin tone percentage

B.9 Test Corpus CAT4 Classifier Results

TEST_CAT4: Top 20 Images			
Image	Details	Image	Details
 Redacted: Annotated as CEM by AFP	Rank: 1 Skin Tone: 69.09% Porn: 1.0 Child: 0.4 CAT1 0.0 CAT2 0.0 CAT3 0.0 CAT4 0.99 CAT5 0.0 CAT7 0.0	 Redacted: Annotated as CEM by AFP	Rank: 2 Skin Tone: 82.33% Porn: 0.99 Child: 1.0 CAT1 0.0 CAT2 0.0 CAT3 0.0 CAT4 0.99 CAT5 0.0 CAT7 0.0
 Redacted: Annotated as CEM by AFP	Rank: 3 Skin Tone: 58.12% Porn: 0.99 Child: 1.0 CAT1 0.0 CAT2 0.0 CAT3 0.0 CAT4 0.99 CAT5 0.0 CAT7 0.0	 Redacted: Annotated as CEM by AFP	Rank: 4 Skin Tone: 82.83% Porn: 0.99 Child: 1.0 CAT1 0.0 CAT2 0.0 CAT3 0.0 CAT4 0.99 CAT5 0.0 CAT7 0.0
 Redacted: Annotated as CEM by AFP	Rank: 5 Skin Tone: 8.37% Porn: 0.99 Child: 1.0 CAT1 0.0 CAT2 0.0 CAT3 0.01 CAT4 0.98 CAT5 0.0 CAT7 0.0	 Redacted: Annotated as CEM by AFP	Rank: 6 Skin Tone: 43.34% Porn: 0.99 Child: 1.0 CAT1 0.0 CAT2 0.0 CAT3 0.0 CAT4 0.99 CAT5 0.0 CAT7 0.0

TEST_CAT4: Top 20 Images			
Image	Details	Image	Details
 Redacted: Annotated as CEM by AFP	Rank: 7 Skin Tone: 54.35% Porn: 0.99 Child: 1.0 CAT1 0.0 CAT2 0.0 CAT3 0.0 CAT4 0.99 CAT5 0.0 CAT7 0.0	 Redacted: Annotated as CEM by AFP	Rank: 8 Skin Tone: 68.96% Porn: 0.99 Child: 1.0 CAT1 0.0 CAT2 0.0 CAT3 0.83 CAT4 0.16 CAT5 0.0 CAT7 0.0
 Redacted: Annotated as CEM by AFP	Rank: 9 Skin Tone: 84.41% Porn: 0.99 Child: 1.0 CAT1 0.0 CAT2 0.0 CAT3 0.02 CAT4 0.97 CAT5 0.0 CAT7 0.0	 Redacted: Annotated as CEM by AFP	Rank: 10 Skin Tone: 63.94% Porn: 0.99 Child: 1.0 CAT1 0.0 CAT2 0.0 CAT3 0.02 CAT4 0.97 CAT5 0.0 CAT7 0.0
 Redacted: Annotated as CEM by AFP	Rank: 11 Skin Tone: 65.82% Porn: 0.99 Child: 1.0 CAT1 0.0 CAT2 0.0 CAT3 0.0 CAT4 0.99 CAT5 0.0 CAT7 0.0	 Redacted: Annotated as CEM by AFP	Rank: 12 Skin Tone: 16.79% Porn: 0.99 Child: 1.0 CAT1 0.0 CAT2 0.0 CAT3 0.02 CAT4 0.97 CAT5 0.0 CAT7 0.0

TEST_CAT4: Top 20 Images			
Image	Details	Image	Details
	Rank: 13 Skin Tone: 74.42% Porn: 0.99 Child: 1.0 CAT1 0.0 CAT2 0.0 CAT3 0.0 CAT4 0.99 CAT5 0.0 CAT7 0.0		Rank: 14 Skin Tone: 67.13% Porn: 0.99 Child: 1.0 CAT1 0.0 CAT2 0.0 CAT3 0.03 CAT4 0.96 CAT5 0.0 CAT7 0.0
Redacted: Annotated as CEM by AFP		Redacted: Annotated as CEM by AFP	
	Rank: 15 Skin Tone: 34.77% Porn: 0.99 Child: 1.0 CAT1 0.0 CAT2 0.0 CAT3 0.0 CAT4 0.99 CAT5 0.0 CAT7 0.0		Rank: 16 Skin Tone: 42.92% Porn: 0.99 Child: 1.0 CAT1 0.0 CAT2 0.0 CAT3 0.0 CAT4 0.99 CAT5 0.0 CAT7 0.0
Redacted: Annotated as CEM by AFP		Redacted: Annotated as CEM by AFP	
	Rank: 17 Skin Tone: 24.65% Porn: 0.99 Child: 1.0 CAT1 0.0 CAT2 0.0 CAT3 0.09 CAT4 0.9 CAT5 0.0 CAT7 0.0		Rank: 18 Skin Tone: 73.66% Porn: 0.99 Child: 1.0 CAT1 0.0 CAT2 0.0 CAT3 0.02 CAT4 0.97 CAT5 0.0 CAT7 0.0
Redacted: Annotated as CEM by AFP		Redacted: Annotated as CEM by AFP	

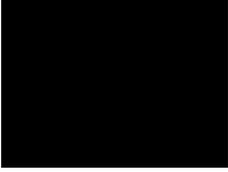
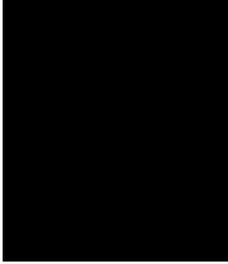
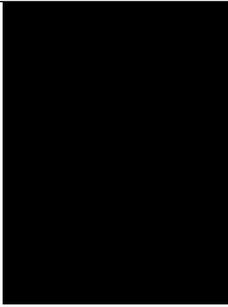
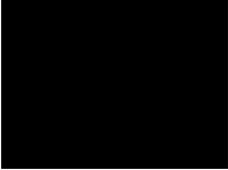
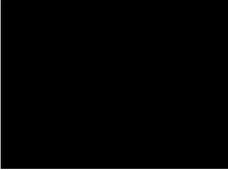
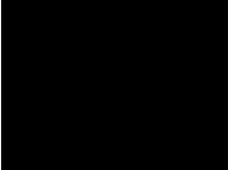
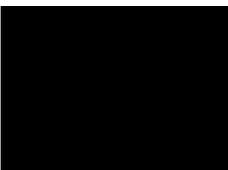
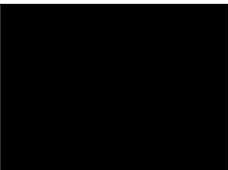
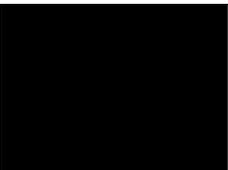
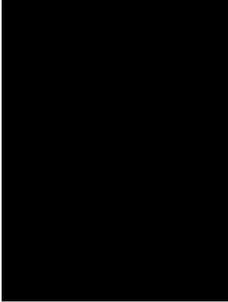
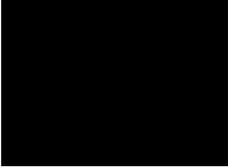
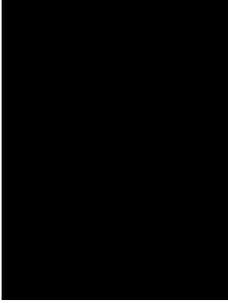
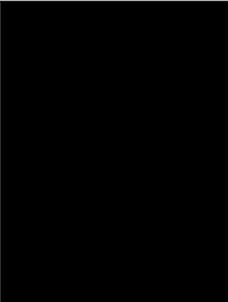
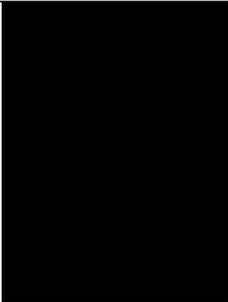
TEST_CAT4: Top 20 Images			
Image	Details	Image	Details
 Redacted: Annotated as CEM by AFP	Rank: 19 Skin Tone: 38.36% Porn: 0.99 Child: 1.0 CAT1 0.01 CAT2 0.0 CAT3 0.97 CAT4 0.0 CAT5 0.0 CAT7 0.0	 Redacted: Annotated as CEM by AFP	Rank: 20 Skin Tone: 80.68% Porn: 0.99 Child: 1.0 CAT1 0.0 CAT2 0.0 CAT3 0.0 CAT4 0.99 CAT5 0.0 CAT7 0.0

Table B.17: Test corpus CAT4 top 20 results by classifier (porn,child)

TEST_CAT4: Bottom 20 Images			
Image	Details	Image	Details
 <p>Redacted: Annotated as CEM by AFP</p>	<p>Rank: 1</p> <p>Skin Tone: 22.30%</p> <p>Porn: 0.0</p> <p>Child: 0.98</p> <p>CAT1: 0.0</p> <p>CAT2: 0.0</p> <p>CAT3: 0.0</p> <p>CAT4: 0.11</p> <p>CAT5: 0.0</p> <p>CAT7: 0.87</p>	 <p>Redacted: Annotated as CEM by AFP</p>	<p>Rank: 2</p> <p>Skin Tone: 76.10%</p> <p>Porn: 0.0</p> <p>Child: 0.99</p> <p>CAT1: 0.0</p> <p>CAT2: 0.0</p> <p>CAT3: 0.14</p> <p>CAT4: 0.85</p> <p>CAT5: 0.0</p> <p>CAT7: 0.0</p>
 <p>Redacted: Annotated as CEM by AFP</p>	<p>Rank: 3</p> <p>Skin Tone: 3.60%</p> <p>Porn: 0.0</p> <p>Child: 0.99</p> <p>CAT1: 0.0</p> <p>CAT2: 0.02</p> <p>CAT3: 0.02</p> <p>CAT4: 0.95</p> <p>CAT5: 0.0</p> <p>CAT7: 0.0</p>	 <p>Redacted: Annotated as CEM by AFP</p>	<p>Rank: 4</p> <p>Skin Tone: 27.84%</p> <p>Porn: 0.0</p> <p>Child: 0.99</p> <p>CAT1: 0.0</p> <p>CAT2: 0.0</p> <p>CAT3: 0.0</p> <p>CAT4: 0.99</p> <p>CAT5: 0.0</p> <p>CAT7: 0.0</p>
 <p>Redacted: Annotated as CEM by AFP</p>	<p>Rank: 5</p> <p>Skin Tone: 37.57%</p> <p>Porn: 0.0</p> <p>Child: 0.99</p> <p>CAT1: 0.0</p> <p>CAT2: 0.0</p> <p>CAT3: 0.59</p> <p>CAT4: 0.4</p> <p>CAT5: 0.0</p> <p>CAT7: 0.0</p>	 <p>Redacted: Annotated as CEM by AFP</p>	<p>Rank: 6</p> <p>Skin Tone: 19.39%</p> <p>Porn: 0.0</p> <p>Child: 0.99</p> <p>CAT1: 0.0</p> <p>CAT2: 0.03</p> <p>CAT3: 0.0</p> <p>CAT4: 0.95</p> <p>CAT5: 0.0</p> <p>CAT7: 0.0</p>

TEST_CAT4: Bottom 20 Images			
Image	Details	Image	Details
 Redacted: Annotated as CEM by AFP	Rank: 7 Skin Tone: 14.65% Porn: 0.0 Child: 0.99 CAT1 0.0 CAT2 0.0 CAT3 0.96 CAT4 0.03 CAT5 0.0 CAT7 0.0	 Redacted: Annotated as CEM by AFP	Rank: 8 Skin Tone: 10.53% Porn: 0.0 Child: 0.99 CAT1 0.0 CAT2 0.0 CAT3 0.39 CAT4 0.6 CAT5 0.0 CAT7 0.0
 Redacted: Annotated as CEM by AFP	Rank: 9 Skin Tone: 11.25% Porn: 0.0 Child: 0.99 CAT1 0.0 CAT2 0.01 CAT3 0.0 CAT4 0.97 CAT5 0.0 CAT7 0.0	 Redacted: Annotated as CEM by AFP	Rank: 10 Skin Tone: 17.56% Porn: 0.0 Child: 0.99 CAT1 0.1 CAT2 0.0 CAT3 0.05 CAT4 0.17 CAT5 0.0 CAT7 0.67
 Redacted: Annotated as CEM by AFP	Rank: 11 Skin Tone: 13.70% Porn: 0.0 Child: 0.99 CAT1 0.0 CAT2 0.0 CAT3 0.0 CAT4 0.99 CAT5 0.0 CAT7 0.0	 Redacted: Annotated as CEM by AFP	Rank: 12 Skin Tone: 15.18% Porn: 0.0 Child: 0.99 CAT1 0.0 CAT2 0.0 CAT3 0.1 CAT4 0.89 CAT5 0.0 CAT7 0.0

TEST_CAT4: Bottom 20 Images			
Image	Details	Image	Details
File Missing	Rank: 13 Skin Tone: 36.81% Porn: 0.0 Child: 1.0 CAT1 0.61 CAT2 0.11 CAT3 0.2 CAT4 0.0 CAT5 0.0 CAT7 0.06	 Redacted: Annotated as CEM by AFP	Rank: 14 Skin Tone: 21.35% Porn: 0.0 Child: 1.0 CAT1 0.0 CAT2 0.0 CAT3 0.58 CAT4 0.41 CAT5 0.0 CAT7 0.0
 Redacted: Annotated as CEM by AFP	Rank: 15 Skin Tone: 21.32% Porn: 0.0 Child: 1.0 CAT1 0.0 CAT2 0.0 CAT3 0.0 CAT4 0.99 CAT5 0.0 CAT7 0.0	 Redacted: Annotated as CEM by AFP	Rank: 16 Skin Tone: 22.08% Porn: 0.0 Child: 1.0 CAT1 0.0 CAT2 0.0 CAT3 0.64 CAT4 0.35 CAT5 0.0 CAT7 0.0
 Redacted: Annotated as CEM by AFP	Rank: 17 Skin Tone: 10.98% Porn: 0.0 Child: 1.0 CAT1 0.03 CAT2 0.0 CAT3 0.75 CAT4 0.2 CAT5 0.0 CAT7 0.0	 Redacted: Annotated as CEM by AFP	Rank: 18 Skin Tone: 10.95% Porn: 0.0 Child: 1.0 CAT1 0.0 CAT2 0.0 CAT3 0.45 CAT4 0.54 CAT5 0.0 CAT7 0.0

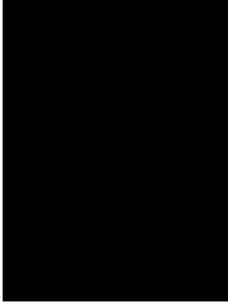
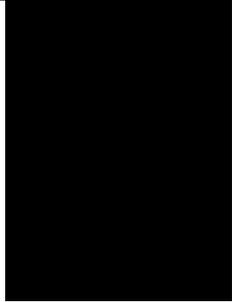
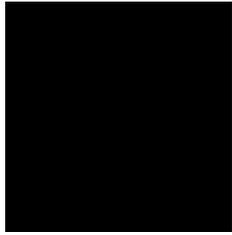
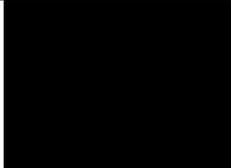
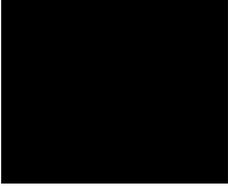
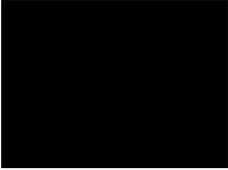
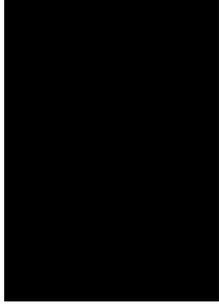
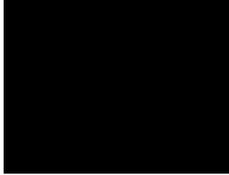
TEST_CAT4: Bottom 20 Images			
Image	Details	Image	Details
	Rank: 19 Skin Tone: 15.26% Porn: 0.0 Child: 1.0 CAT1 0.01 CAT2 0.0 CAT3 0.21 CAT4 0.77 CAT5 0.0 CAT7 0.0		Rank: 20 Skin Tone: 4.15% Porn: 0.0 Child: 1.0 CAT1 0.0 CAT2 0.0 CAT3 0.25 CAT4 0.74 CAT5 0.0 CAT7 0.0
Redacted: Annotated as CEM by AFP		Redacted: Annotated as CEM by AFP	

Table B.18: Test corpus CAT4 bottom 20 results by classifier (porn,child)

B.10 Test Corpus CAT5 Skin Tone Results

TEST_CAT5: Top 20 Images			
Image	Details	Image	Details
	<p>Rank: 1</p> <p>Skin Tone: 95.53%</p> <p>Porn: 0.69</p> <p>Child: 0.99</p> <p>CAT1 0.0</p> <p>CAT2 0.0</p> <p>CAT3 0.98</p> <p>CAT4 0.01</p> <p>CAT5 0.0</p> <p>CAT7 0.0</p>		<p>Rank: 2</p> <p>Skin Tone: 94.17%</p> <p>Porn: 0.15</p> <p>Child: 0.95</p> <p>CAT1 0.99</p> <p>CAT2 0.0</p> <p>CAT3 0.0</p> <p>CAT4 0.0</p> <p>CAT5 0.0</p> <p>CAT7 0.0</p>
Redacted: Annotated as CEM by AFP		Redacted: Annotated as CEM by AFP	
	<p>Rank: 3</p> <p>Skin Tone: 93.91%</p> <p>Porn: 0.04</p> <p>Child: 0.99</p> <p>CAT1 0.42</p> <p>CAT2 0.0</p> <p>CAT3 0.48</p> <p>CAT4 0.03</p> <p>CAT5 0.0</p> <p>CAT7 0.06</p>		<p>Rank: 4</p> <p>Skin Tone: 93.49%</p> <p>Porn: 0.98</p> <p>Child: 0.99</p> <p>CAT1 0.0</p> <p>CAT2 0.0</p> <p>CAT3 0.4</p> <p>CAT4 0.59</p> <p>CAT5 0.0</p> <p>CAT7 0.0</p>
Redacted: Annotated as CEM by AFP		Redacted: Annotated as CEM by AFP	
	<p>Rank: 5</p> <p>Skin Tone: 93.12%</p> <p>Porn: 0.08</p> <p>Child: 0.99</p> <p>CAT1 0.94</p> <p>CAT2 0.0</p> <p>CAT3 0.0</p> <p>CAT4 0.0</p> <p>CAT5 0.0</p> <p>CAT7 0.04</p>		<p>Rank: 6</p> <p>Skin Tone: 92.33%</p> <p>Porn: 0.99</p> <p>Child: 0.99</p> <p>CAT1 0.91</p> <p>CAT2 0.0</p> <p>CAT3 0.0</p> <p>CAT4 0.07</p> <p>CAT5 0.0</p> <p>CAT7 0.0</p>
Redacted: Annotated as CEM by AFP		Redacted: Annotated as CEM by AFP	

TEST_CAT5: Top 20 Images			
Image	Details	Image	Details
 Redacted: Annotated as CEM by AFP	Rank: 7 Skin Tone: 89.26% Porn: 0.95 Child: 0.91 CAT1 0.0 CAT2 0.0 CAT3 0.95 CAT4 0.03 CAT5 0.0 CAT7 0.0	 Redacted: Annotated as CEM by AFP	Rank: 8 Skin Tone: 88.48% Porn: 0.17 Child: 0.99 CAT1 0.86 CAT2 0.0 CAT3 0.04 CAT4 0.05 CAT5 0.0 CAT7 0.03
 Redacted: Annotated as CEM by AFP	Rank: 9 Skin Tone: 88.13% Porn: 0.92 Child: 0.96 CAT1 0.99 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.0	 Redacted: Annotated as CEM by AFP	Rank: 10 Skin Tone: 87.66% Porn: 0.92 Child: 0.99 CAT1 0.93 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.02 CAT7 0.03
 Redacted: Annotated as CEM by AFP	Rank: 11 Skin Tone: 87.66% Porn: 0.92 Child: 0.99 CAT1 0.93 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.02 CAT7 0.03	 Redacted: Annotated as CEM by AFP	Rank: 12 Skin Tone: 86.79% Porn: 0.04 Child: 0.99 CAT1 0.01 CAT2 0.0 CAT3 0.09 CAT4 0.19 CAT5 0.0 CAT7 0.68

TEST_CAT5: Top 20 Images			
Image	Details	Image	Details
	Rank: 13 Skin Tone: 86.55% Porn: 0.96 Child: 0.99 CAT1 0.87 CAT2 0.02 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.1		Rank: 14 Skin Tone: 86.10% Porn: 0.95 Child: 1.0 CAT1 0.0 CAT2 0.0 CAT3 0.06 CAT4 0.93 CAT5 0.0 CAT7 0.0
Redacted: Annotated as CEM by AFP		Redacted: Annotated as CEM by AFP	
	Rank: 15 Skin Tone: 85.95% Porn: 0.97 Child: 0.97 CAT1 0.92 CAT2 0.05 CAT3 0.01 CAT4 0.0 CAT5 0.0 CAT7 0.0		Rank: 16 Skin Tone: 85.83% Porn: 0.24 Child: 0.99 CAT1 0.7 CAT2 0.0 CAT3 0.19 CAT4 0.08 CAT5 0.0 CAT7 0.01
Redacted: Annotated as CEM by AFP		Redacted: Annotated as CEM by AFP	
	Rank: 17 Skin Tone: 85.37% Porn: 0.16 Child: 1.0 CAT1 0.78 CAT2 0.0 CAT3 0.06 CAT4 0.09 CAT5 0.02 CAT7 0.03		Rank: 18 Skin Tone: 84.01% Porn: 0.16 Child: 0.99 CAT1 0.01 CAT2 0.0 CAT3 0.28 CAT4 0.0 CAT5 0.03 CAT7 0.66
Redacted: Annotated as CEM by AFP		Redacted: Annotated as CEM by AFP	

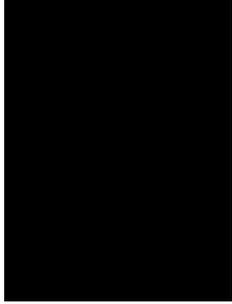
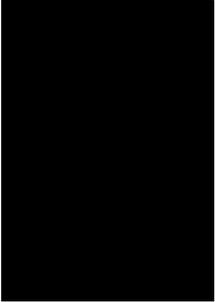
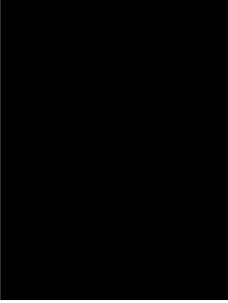
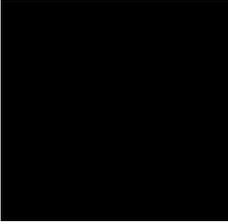
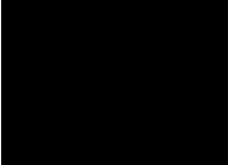
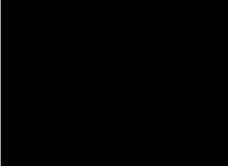
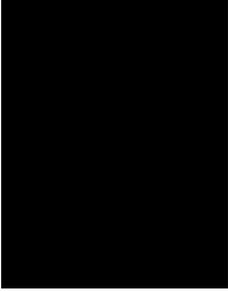
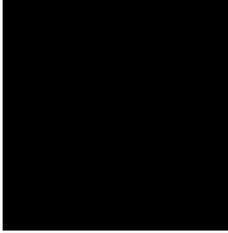
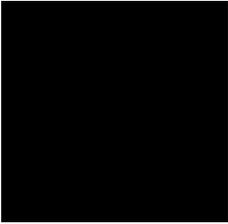
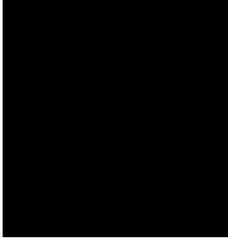
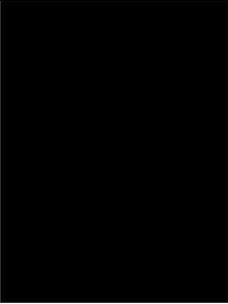
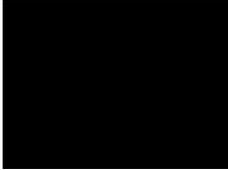
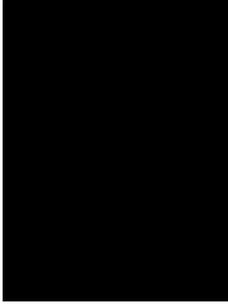
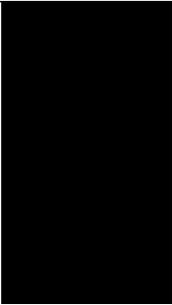
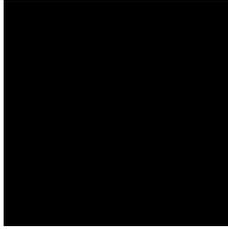
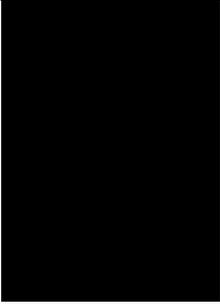
TEST_CAT5: Top 20 Images			
Image	Details	Image	Details
	Rank: 19 Skin Tone: 83.13% Porn: 0.7 Child: 0.93 CAT1 0.05 CAT2 0.1 CAT3 0.0 CAT4 0.03 CAT5 0.0 CAT7 0.79		Rank: 20 Skin Tone: 83.12% Porn: 0.21 Child: 0.99 CAT1 0.0 CAT2 0.0 CAT3 0.0 CAT4 0.99 CAT5 0.0 CAT7 0.0
Redacted: Annotated as CEM by AFP		Redacted: Annotated as CEM by AFP	

Table B.19: Test corpus CAT5 top 20 results by skin tone percentage

TEST_CAT5: Bottom 20 Images			
Image	Details	Image	Details
 <p>Redacted: Annotated as CEM by AFP</p>	<p>Rank: 1</p> <p>Skin Tone: 0.00%</p> <p>Porn: 0.76</p> <p>Child: 1.0</p> <p>CAT1 0.24</p> <p>CAT2 0.26</p> <p>CAT3 0.22</p> <p>CAT4 0.01</p> <p>CAT5 0.0</p> <p>CAT7 0.23</p>	<p>File Missing</p>	<p>Rank: 2</p> <p>Skin Tone: 0.00%</p> <p>Porn: 0.35</p> <p>Child: 1.0</p> <p>CAT1 0.0</p> <p>CAT2 0.24</p> <p>CAT3 0.36</p> <p>CAT4 0.38</p> <p>CAT5 0.0</p> <p>CAT7 0.0</p>
 <p>Redacted: Annotated as CEM by AFP</p>	<p>Rank: 3</p> <p>Skin Tone: 1.04%</p> <p>Porn: 0.94</p> <p>Child: 1.0</p> <p>CAT1 0.09</p> <p>CAT2 0.0</p> <p>CAT3 0.89</p> <p>CAT4 0.01</p> <p>CAT5 0.0</p> <p>CAT7 0.0</p>	 <p>Redacted: Annotated as CEM by AFP</p>	<p>Rank: 4</p> <p>Skin Tone: 1.25%</p> <p>Porn: 0.04</p> <p>Child: 0.99</p> <p>CAT1 0.64</p> <p>CAT2 0.0</p> <p>CAT3 0.0</p> <p>CAT4 0.0</p> <p>CAT5 0.0</p> <p>CAT7 0.35</p>
 <p>Redacted: Annotated as CEM by AFP</p>	<p>Rank: 5</p> <p>Skin Tone: 5.68%</p> <p>Porn: 0.74</p> <p>Child: 1.0</p> <p>CAT1 0.0</p> <p>CAT2 0.0</p> <p>CAT3 0.7</p> <p>CAT4 0.28</p> <p>CAT5 0.0</p> <p>CAT7 0.0</p>	 <p>Redacted: Annotated as CEM by AFP</p>	<p>Rank: 6</p> <p>Skin Tone: 6.15%</p> <p>Porn: 0.9</p> <p>Child: 0.81</p> <p>CAT1 0.08</p> <p>CAT2 0.88</p> <p>CAT3 0.02</p> <p>CAT4 0.0</p> <p>CAT5 0.0</p> <p>CAT7 0.0</p>

TEST_CAT5: Bottom 20 Images			
Image	Details	Image	Details
 Redacted: Annotated as CEM by AFP	Rank: 7 Skin Tone: 6.30% Porn: 0.0 Child: 0.95 CAT1 0.01 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.97	 Redacted: Annotated as CEM by AFP	Rank: 8 Skin Tone: 6.98% Porn: 0.03 Child: 0.99 CAT1 0.0 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 1.0 CAT7 0.0
 Redacted: Annotated as CEM by AFP	Rank: 9 Skin Tone: 8.23% Porn: 0.07 Child: 0.8 CAT1 0.82 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.17	 Redacted: Annotated as CEM by AFP	Rank: 10 Skin Tone: 9.86% Porn: 0.99 Child: 0.99 CAT1 0.0 CAT2 0.0 CAT3 0.02 CAT4 0.96 CAT5 0.0 CAT7 0.0
 Redacted: Annotated as CEM by AFP	Rank: 11 Skin Tone: 10.44% Porn: 0.0 Child: 0.63 CAT1 0.33 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.65	 Redacted: Annotated as CEM by AFP	Rank: 12 Skin Tone: 11.14% Porn: 0.97 Child: 0.01 CAT1 0.71 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.28

TEST_CAT5: Bottom 20 Images			
Image	Details	Image	Details
 Redacted: Annotated as CEM by AFP	Rank: 13 Skin Tone: 13.13% Porn: 0.64 Child: 1.0 CAT1: 0.02 CAT2: 0.21 CAT3: 0.0 CAT4: 0.75 CAT5: 0.0 CAT7: 0.0	 Redacted: Annotated as CEM by AFP	Rank: 14 Skin Tone: 13.77% Porn: 0.89 Child: 1.0 CAT1: 0.0 CAT2: 0.0 CAT3: 0.01 CAT4: 0.98 CAT5: 0.0 CAT7: 0.0
 Redacted: Annotated as CEM by AFP	Rank: 15 Skin Tone: 14.88% Porn: 0.37 Child: 0.99 CAT1: 0.99 CAT2: 0.0 CAT3: 0.0 CAT4: 0.0 CAT5: 0.0 CAT7: 0.0	 Redacted: Annotated as CEM by AFP	Rank: 16 Skin Tone: 15.00% Porn: 0.49 Child: 0.99 CAT1: 0.0 CAT2: 0.0 CAT3: 0.09 CAT4: 0.88 CAT5: 0.0 CAT7: 0.01
 Redacted: Annotated as CEM by AFP	Rank: 17 Skin Tone: 15.18% Porn: 0.16 Child: 1.0 CAT1: 0.15 CAT2: 0.0 CAT3: 0.0 CAT4: 0.0 CAT5: 0.0 CAT7: 0.84	 Redacted: Annotated as CEM by AFP	Rank: 18 Skin Tone: 15.38% Porn: 0.99 Child: 0.94 CAT1: 0.31 CAT2: 0.64 CAT3: 0.0 CAT4: 0.0 CAT5: 0.0 CAT7: 0.03

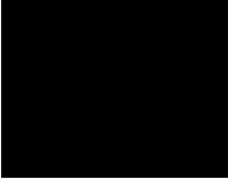
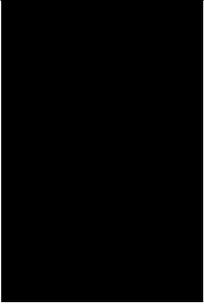
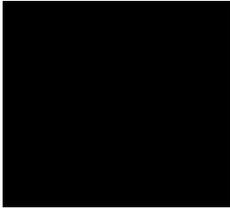
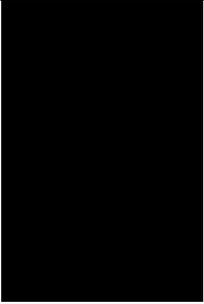
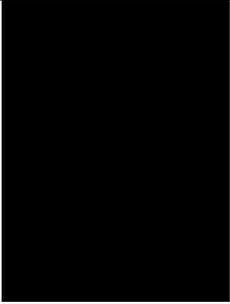
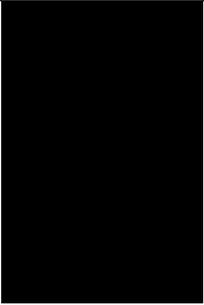
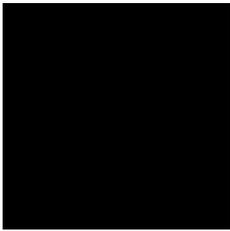
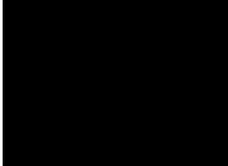
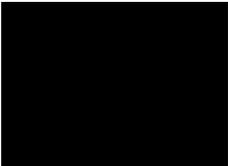
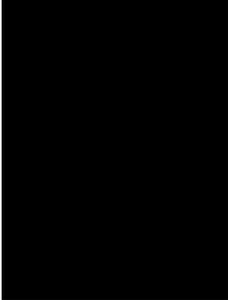
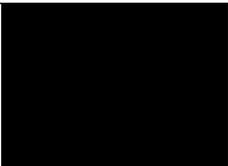
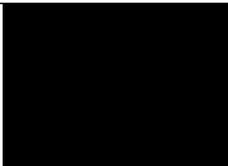
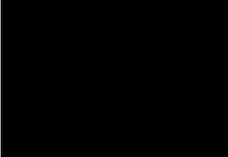
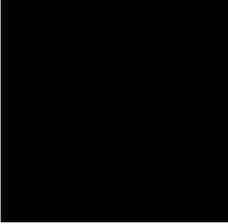
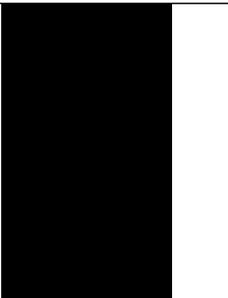
TEST_CAT5: Bottom 20 Images			
Image	Details	Image	Details
	Rank: 19 Skin Tone: 15.79% Porn: 0.99 Child: 0.88 CAT1 0.27 CAT2 0.71 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.01		Rank: 20 Skin Tone: 16.15% Porn: 0.8 Child: 0.99 CAT1 0.0 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 1.0 CAT7 0.0
Redacted: Annotated as CEM by AFP		Redacted: Annotated as CEM by AFP	

Table B.20: Test corpus CAT5 bottom 20 results by skin tone percentage

B.11 Test Corpus CAT5 Classifier Results

TEST_CAT5: Top 20 Images			
Image	Details	Image	Details
 Redacted: Annotated as CEM by AFP	Rank: 1 Skin Tone: 40.09% Porn: 0.99 Child: 1.0 CAT1 0.01 CAT2 0.0 CAT3 0.92 CAT4 0.06 CAT5 0.0 CAT7 0.0	 Redacted: Annotated as CEM by AFP	Rank: 2 Skin Tone: 48.10% Porn: 0.99 Child: 1.0 CAT1 0.99 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.0
 Redacted: Annotated as CEM by AFP	Rank: 3 Skin Tone: 40.09% Porn: 0.99 Child: 1.0 CAT1 0.01 CAT2 0.0 CAT3 0.92 CAT4 0.06 CAT5 0.0 CAT7 0.0	 Redacted: Annotated as CEM by AFP	Rank: 4 Skin Tone: 54.88% Porn: 0.99 Child: 1.0 CAT1 0.03 CAT2 0.0 CAT3 0.85 CAT4 0.11 CAT5 0.0 CAT7 0.0
 Redacted: Annotated as CEM by AFP	Rank: 5 Skin Tone: 40.01% Porn: 0.99 Child: 1.0 CAT1 0.01 CAT2 0.0 CAT3 0.94 CAT4 0.04 CAT5 0.0 CAT7 0.0	 Redacted: Annotated as CEM by AFP	Rank: 6 Skin Tone: 74.71% Porn: 0.99 Child: 1.0 CAT1 0.0 CAT2 0.99 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.0

TEST_CAT5: Top 20 Images			
Image	Details	Image	Details
 Redacted: Annotated as CEM by AFP	Rank: 7 Skin Tone: 18.19% Porn: 0.99 Child: 1.0 CAT1 0.0 CAT2 0.11 CAT3 0.02 CAT4 0.86 CAT5 0.0 CAT7 0.0	 Redacted: Annotated as CEM by AFP	Rank: 8 Skin Tone: 67.70% Porn: 0.99 Child: 1.0 CAT1 0.1 CAT2 0.0 CAT3 0.89 CAT4 0.0 CAT5 0.0 CAT7 0.0
 Redacted: Annotated as CEM by AFP	Rank: 9 Skin Tone: 73.35% Porn: 0.99 Child: 1.0 CAT1 0.99 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.0	 Redacted: Annotated as CEM by AFP	Rank: 10 Skin Tone: 74.05% Porn: 0.99 Child: 1.0 CAT1 0.95 CAT2 0.0 CAT3 0.04 CAT4 0.0 CAT5 0.0 CAT7 0.0
 Redacted: Annotated as CEM by AFP	Rank: 11 Skin Tone: 43.27% Porn: 0.99 Child: 1.0 CAT1 0.0 CAT2 0.99 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.0	 Redacted: Annotated as CEM by AFP	Rank: 12 Skin Tone: 61.57% Porn: 0.99 Child: 1.0 CAT1 0.16 CAT2 0.0 CAT3 0.83 CAT4 0.0 CAT5 0.0 CAT7 0.0

TEST_CAT5: Top 20 Images			
Image	Details	Image	Details
 Redacted: Annotated as CEM by AFP	Rank: 13 Skin Tone: 47.01% Porn: 0.99 Child: 1.0 CAT1: 0.0 CAT2: 0.0 CAT3: 0.0 CAT4: 0.99 CAT5: 0.0 CAT7: 0.0	 Redacted: Annotated as CEM by AFP	Rank: 14 Skin Tone: 69.26% Porn: 0.99 Child: 1.0 CAT1: 0.0 CAT2: 0.0 CAT3: 0.0 CAT4: 0.0 CAT5: 0.99 CAT7: 0.0
 Redacted: Annotated as CEM by AFP	Rank: 15 Skin Tone: 44.18% Porn: 0.99 Child: 1.0 CAT1: 0.08 CAT2: 0.0 CAT3: 0.09 CAT4: 0.0 CAT5: 0.82 CAT7: 0.0	 Redacted: Annotated as CEM by AFP	Rank: 16 Skin Tone: 72.76% Porn: 0.99 Child: 1.0 CAT1: 0.04 CAT2: 0.94 CAT3: 0.0 CAT4: 0.0 CAT5: 0.0 CAT7: 0.01
 Redacted: Annotated as CEM by AFP	Rank: 17 Skin Tone: 58.12% Porn: 0.99 Child: 0.99 CAT1: 0.87 CAT2: 0.0 CAT3: 0.0 CAT4: 0.0 CAT5: 0.1 CAT7: 0.01	 Redacted: Annotated as CEM by AFP	Rank: 18 Skin Tone: 92.33% Porn: 0.99 Child: 0.99 CAT1: 0.91 CAT2: 0.0 CAT3: 0.0 CAT4: 0.07 CAT5: 0.0 CAT7: 0.0

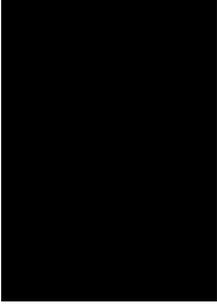
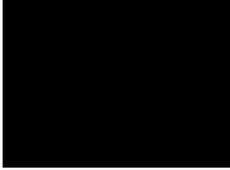
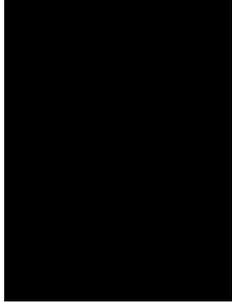
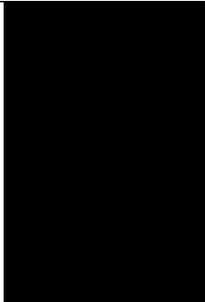
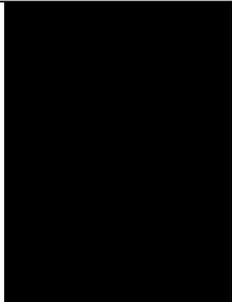
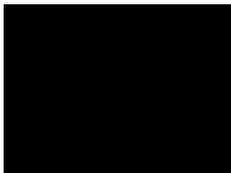
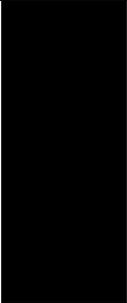
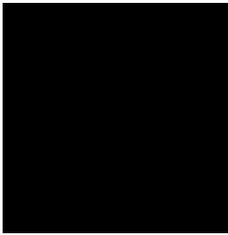
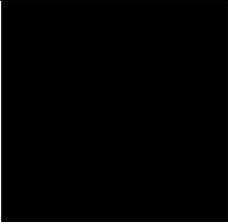
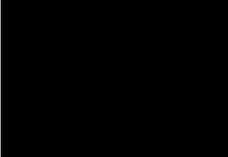
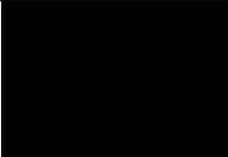
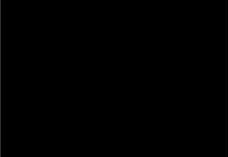
TEST_CAT5: Top 20 Images			
Image	Details	Image	Details
	Rank: 19 Skin Tone: 67.55% Porn: 0.99 Child: 0.99 CAT1 0.0 CAT2 0.0 CAT3 0.99 CAT4 0.0 CAT5 0.0 CAT7 0.0		Rank: 20 Skin Tone: 39.79% Porn: 0.99 Child: 0.99 CAT1 0.79 CAT2 0.01 CAT3 0.0 CAT4 0.0 CAT5 0.14 CAT7 0.03
Redacted: Annotated as CEM by AFP		Redacted: Annotated as CEM by AFP	

Table B.21: Test corpus CAT5 top 20 results by classifier (porn,child)

TEST_CAT5: Bottom 20 Images			
Image	Details	Image	Details
	<p>Rank: 1</p> <p>Skin Tone: 10.44%</p> <p>Porn: 0.0</p> <p>Child: 0.63</p> <p>CAT1 0.33</p> <p>CAT2 0.0</p> <p>CAT3 0.0</p> <p>CAT4 0.0</p> <p>CAT5 0.0</p> <p>CAT7 0.65</p>		<p>Rank: 2</p> <p>Skin Tone: 6.30%</p> <p>Porn: 0.0</p> <p>Child: 0.95</p> <p>CAT1 0.01</p> <p>CAT2 0.0</p> <p>CAT3 0.0</p> <p>CAT4 0.0</p> <p>CAT5 0.0</p> <p>CAT7 0.97</p>
Redacted: Annotated as CEM by AFP		Redacted: Annotated as CEM by AFP	
	<p>Rank: 3</p> <p>Skin Tone: 69.28%</p> <p>Porn: 0.0</p> <p>Child: 0.99</p> <p>CAT1 0.08</p> <p>CAT2 0.0</p> <p>CAT3 0.02</p> <p>CAT4 0.0</p> <p>CAT5 0.0</p> <p>CAT7 0.89</p>		<p>Rank: 4</p> <p>Skin Tone: 42.99%</p> <p>Porn: 0.01</p> <p>Child: 1.0</p> <p>CAT1 0.0</p> <p>CAT2 0.0</p> <p>CAT3 0.0</p> <p>CAT4 0.0</p> <p>CAT5 0.98</p> <p>CAT7 0.0</p>
Redacted: Annotated as CEM by AFP		Redacted: Annotated as CEM by AFP	
	<p>Rank: 5</p> <p>Skin Tone: 17.35%</p> <p>Porn: 0.02</p> <p>Child: 0.91</p> <p>CAT1 0.0</p> <p>CAT2 0.0</p> <p>CAT3 0.0</p> <p>CAT4 0.0</p> <p>CAT5 0.0</p> <p>CAT7 0.99</p>		<p>Rank: 6</p> <p>Skin Tone: 74.26%</p> <p>Porn: 0.02</p> <p>Child: 0.99</p> <p>CAT1 0.0</p> <p>CAT2 0.0</p> <p>CAT3 0.99</p> <p>CAT4 0.0</p> <p>CAT5 0.0</p> <p>CAT7 0.0</p>
Redacted: Annotated as CEM by AFP		Redacted: Annotated as CEM by AFP	

TEST_CAT5: Bottom 20 Images			
Image	Details	Image	Details
 Redacted: Annotated as CEM by AFP	Rank: 7 Skin Tone: 74.15% Porn: 0.03 Child: 0.99 CAT1 0.0 CAT2 0.0 CAT3 0.97 CAT4 0.02 CAT5 0.0 CAT7 0.0	 Redacted: Annotated as CEM by AFP	Rank: 8 Skin Tone: 6.98% Porn: 0.03 Child: 0.99 CAT1 0.0 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 1.0 CAT7 0.0
 Redacted: Annotated as CEM by AFP	Rank: 9 Skin Tone: 86.79% Porn: 0.04 Child: 0.99 CAT1 0.01 CAT2 0.0 CAT3 0.09 CAT4 0.19 CAT5 0.0 CAT7 0.68	 Redacted: Annotated as CEM by AFP	Rank: 10 Skin Tone: 93.91% Porn: 0.04 Child: 0.99 CAT1 0.42 CAT2 0.0 CAT3 0.48 CAT4 0.03 CAT5 0.0 CAT7 0.06
 Redacted: Annotated as CEM by AFP	Rank: 11 Skin Tone: 1.25% Porn: 0.04 Child: 0.99 CAT1 0.64 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.35	 Redacted: Annotated as CEM by AFP	Rank: 12 Skin Tone: 46.61% Porn: 0.05 Child: 0.0 CAT1 0.0 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.99

TEST_CAT5: Bottom 20 Images			
Image	Details	Image	Details
 Redacted: Annotated as CEM by AFP	Rank: 13 Skin Tone: 43.26% Porn: 0.05 Child: 0.99 CAT1 0.02 CAT2 0.76 CAT3 0.01 CAT4 0.0 CAT5 0.0 CAT7 0.19	 Redacted: Annotated as CEM by AFP	Rank: 14 Skin Tone: 67.80% Porn: 0.06 Child: 0.36 CAT1 0.0 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.99 CAT7 0.0
 Redacted: Annotated as CEM by AFP	Rank: 15 Skin Tone: 8.23% Porn: 0.07 Child: 0.8 CAT1 0.82 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.17	 Redacted: Annotated as CEM by AFP	Rank: 16 Skin Tone: 47.78% Porn: 0.07 Child: 0.97 CAT1 0.13 CAT2 0.25 CAT3 0.5 CAT4 0.0 CAT5 0.0 CAT7 0.1
 Redacted: Annotated as CEM by AFP	Rank: 17 Skin Tone: 23.48% Porn: 0.07 Child: 0.97 CAT1 0.0 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.99	 Redacted: Annotated as CEM by AFP	Rank: 18 Skin Tone: 93.12% Porn: 0.08 Child: 0.99 CAT1 0.94 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.04

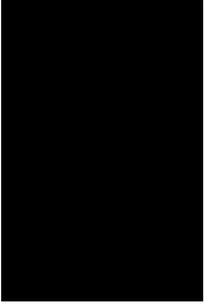
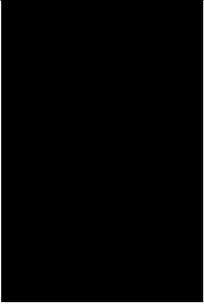
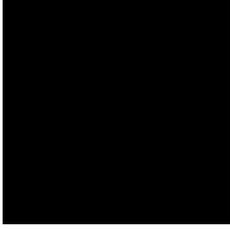
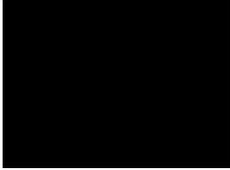
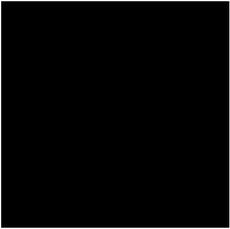
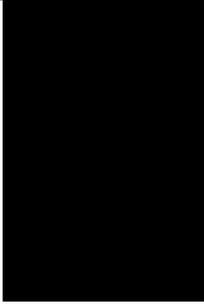
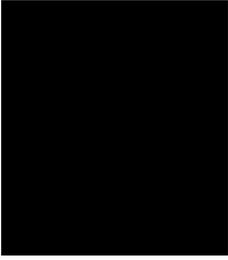
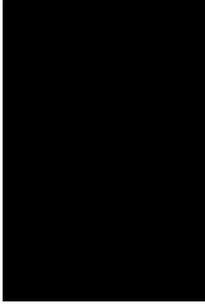
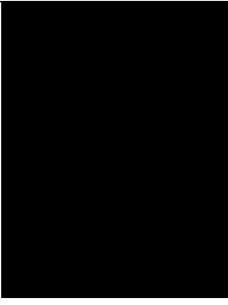
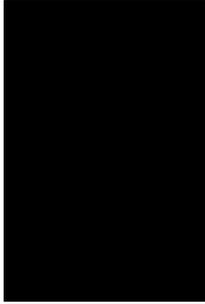
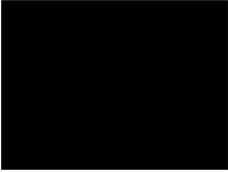
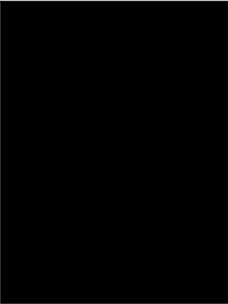
TEST_CAT5: Bottom 20 Images			
Image	Details	Image	Details
	Rank: 19 Skin Tone: 73.33% Porn: 0.09 Child: 0.99 CAT1 0.99 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.0		Rank: 20 Skin Tone: 67.48% Porn: 0.1 Child: 0.99 CAT1 0.0 CAT2 0.0 CAT3 0.99 CAT4 0.0 CAT5 0.0 CAT7 0.0
Redacted: Annotated as CEM by AFP		Redacted: Annotated as CEM by AFP	

Table B.22: Test corpus CAT5 bottom 20 results by classifier (porn,child)

B.12 Test Corpus CAT7 Skin Tone Results

TEST_CAT7: Top 20 Images			
Image	Details	Image	Details
 Redacted: Annotated as CEM by AFP	Rank: 1 Skin Tone: 97.71% Porn: 0.22 Child: 0.99 CAT1: 0.99 CAT2: 0.0 CAT3: 0.0 CAT4: 0.0 CAT5: 0.0 CAT7: 0.0	 Redacted: Annotated as CEM by AFP	Rank: 2 Skin Tone: 96.45% Porn: 0.76 Child: 0.98 CAT1: 0.0 CAT2: 0.0 CAT3: 0.0 CAT4: 0.0 CAT5: 0.0 CAT7: 0.99
 Redacted: Annotated as CEM by AFP	Rank: 3 Skin Tone: 93.79% Porn: 0.42 Child: 0.0 CAT1: 0.04 CAT2: 0.0 CAT3: 0.0 CAT4: 0.0 CAT5: 0.0 CAT7: 0.95	 Redacted: Annotated as CEM by AFP	Rank: 4 Skin Tone: 92.26% Porn: 0.52 Child: 0.11 CAT1: 0.0 CAT2: 0.8 CAT3: 0.13 CAT4: 0.05 CAT5: 0.0 CAT7: 0.0
 Redacted: Annotated as CEM by AFP	Rank: 5 Skin Tone: 89.78% Porn: 0.68 Child: 0.67 CAT1: 0.01 CAT2: 0.0 CAT3: 0.0 CAT4: 0.0 CAT5: 0.0 CAT7: 0.98	 Redacted: Annotated as CEM by AFP	Rank: 6 Skin Tone: 87.63% Porn: 0.06 Child: 0.02 CAT1: 0.03 CAT2: 0.0 CAT3: 0.0 CAT4: 0.0 CAT5: 0.0 CAT7: 0.96

TEST_CAT7: Top 20 Images			
Image	Details	Image	Details
 Redacted: Annotated as CEM by AFP	Rank: 7 Skin Tone: 86.56% Porn: 0.12 Child: 0.99 CAT1 0.08 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.91	 Redacted: Annotated as CEM by AFP	Rank: 8 Skin Tone: 85.78% Porn: 0.62 Child: 0.99 CAT1 0.71 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.28
 Redacted: Annotated as CEM by AFP	Rank: 9 Skin Tone: 85.24% Porn: 0.68 Child: 0.97 CAT1 0.0 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.99	 Redacted: Annotated as CEM by AFP	Rank: 10 Skin Tone: 83.97% Porn: 0.03 Child: 0.99 CAT1 0.11 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.88
 Redacted: Annotated as CEM by AFP	Rank: 11 Skin Tone: 83.84% Porn: 0.21 Child: 0.99 CAT1 0.98 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.01	 Redacted: Annotated as CEM by AFP	Rank: 12 Skin Tone: 83.10% Porn: 0.06 Child: 0.34 CAT1 0.85 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.14

TEST_CAT7: Top 20 Images			
Image	Details	Image	Details
 Redacted: Annotated as CEM by AFP	Rank: 13 Skin Tone: 82.60% Porn: 0.98 Child: 0.0 CAT1: 0.55 CAT2: 0.0 CAT3: 0.0 CAT4: 0.0 CAT5: 0.0 CAT7: 0.44	 Redacted: Annotated as CEM by AFP	Rank: 14 Skin Tone: 82.55% Porn: 0.05 Child: 0.99 CAT1: 0.0 CAT2: 0.0 CAT3: 0.0 CAT4: 0.0 CAT5: 0.0 CAT7: 0.99
 Redacted: Annotated as CEM by AFP	Rank: 15 Skin Tone: 81.26% Porn: 0.18 Child: 0.05 CAT1: 0.99 CAT2: 0.0 CAT3: 0.0 CAT4: 0.0 CAT5: 0.0 CAT7: 0.0	 Redacted: Annotated as CEM by AFP	Rank: 16 Skin Tone: 80.58% Porn: 0.73 Child: 0.58 CAT1: 0.14 CAT2: 0.0 CAT3: 0.0 CAT4: 0.0 CAT5: 0.0 CAT7: 0.85
 Redacted: Annotated as CEM by AFP	Rank: 17 Skin Tone: 80.54% Porn: 0.03 Child: 0.67 CAT1: 0.02 CAT2: 0.1 CAT3: 0.31 CAT4: 0.0 CAT5: 0.0 CAT7: 0.55	 Redacted: Annotated as CEM by AFP	Rank: 18 Skin Tone: 80.53% Porn: 0.21 Child: 0.98 CAT1: 0.65 CAT2: 0.0 CAT3: 0.0 CAT4: 0.0 CAT5: 0.0 CAT7: 0.34

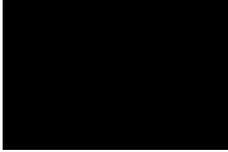
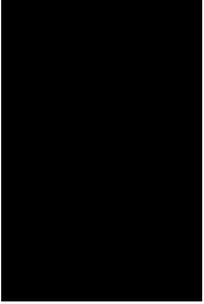
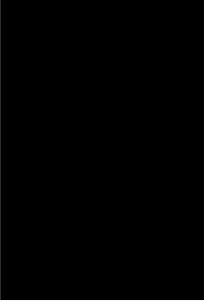
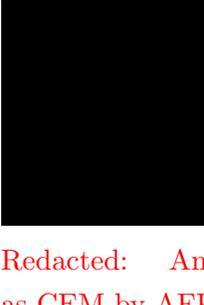
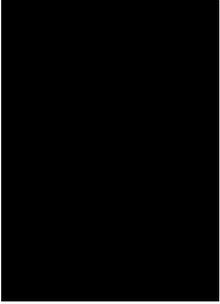
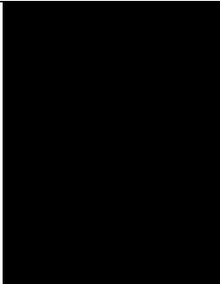
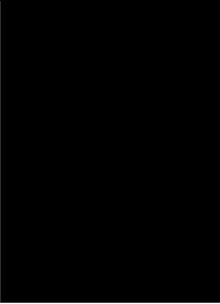
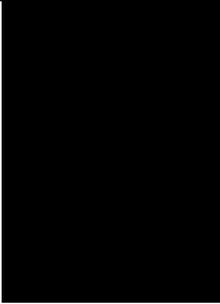
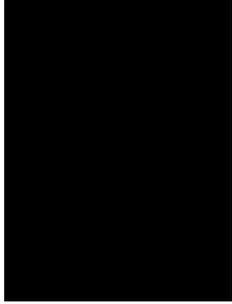
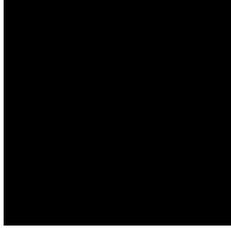
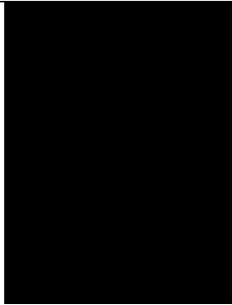
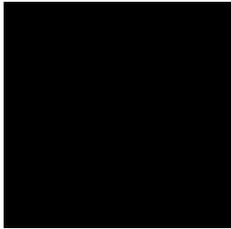
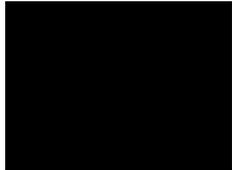
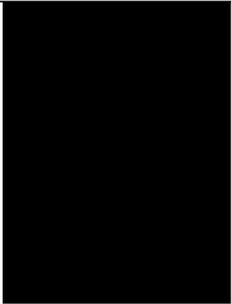
TEST_CAT7: Top 20 Images			
Image	Details	Image	Details
	Rank: 19 Skin Tone: 80.35% Porn: 0.07 Child: 0.82 CAT1 0.0 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.99		Rank: 20 Skin Tone: 79.80% Porn: 0.0 Child: 0.99 CAT1 0.0 CAT2 0.0 CAT3 0.01 CAT4 0.98 CAT5 0.0 CAT7 0.0
Redacted: Annotated as CEM by AFP		Redacted: Annotated as CEM by AFP	

Table B.23: Test corpus CAT7 top 20 results by skin tone percentage

TEST_CAT7: Bottom 20 Images			
Image	Details	Image	Details
 Redacted: Annotated as CEM by AFP	Rank: 1 Skin Tone: 0.00% Porn: 0.16 Child: 1.0 CAT1 0.01 CAT2 0.98 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.0	 Redacted: Annotated as CEM by AFP	Rank: 2 Skin Tone: 0.00% Porn: 0.37 Child: 1.0 CAT1 0.0 CAT2 0.84 CAT3 0.01 CAT4 0.14 CAT5 0.0 CAT7 0.0
 Redacted: Annotated as CEM by AFP	Rank: 3 Skin Tone: 0.00% Porn: 0.0 Child: 1.0 CAT1 0.03 CAT2 0.96 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.0	 Redacted: Annotated as CEM by AFP	Rank: 4 Skin Tone: 0.00% Porn: 0.12 Child: 0.51 CAT1 0.12 CAT2 0.75 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.11
 Redacted: Annotated as CEM by AFP	Rank: 5 Skin Tone: 0.00% Porn: 0.0 Child: 0.98 CAT1 0.26 CAT2 0.05 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.68	 Redacted: Annotated as CEM by AFP	Rank: 6 Skin Tone: 0.00% Porn: 0.03 Child: 1.0 CAT1 0.7 CAT2 0.02 CAT3 0.22 CAT4 0.03 CAT5 0.0 CAT7 0.0

TEST_CAT7: Bottom 20 Images			
Image	Details	Image	Details
 Redacted: Annotated as CEM by AFP	Rank: 7 Skin Tone: 0.00% Porn: 0.21 Child: 0.7 CAT1 0.0 CAT2 0.98 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.01	 Redacted: Annotated as CEM by AFP	Rank: 8 Skin Tone: 0.00% Porn: 0.15 Child: 0.99 CAT1 0.15 CAT2 0.0 CAT3 0.1 CAT4 0.72 CAT5 0.0 CAT7 0.01
 Redacted: Annotated as CEM by AFP	Rank: 9 Skin Tone: 0.00% Porn: 0.0 Child: 0.99 CAT1 0.37 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.61	 Redacted: Annotated as CEM by AFP	Rank: 10 Skin Tone: 0.00% Porn: 0.96 Child: 0.99 CAT1 0.98 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.0
 Redacted: Annotated as CEM by AFP	Rank: 11 Skin Tone: 0.01% Porn: 0.02 Child: 0.99 CAT1 0.14 CAT2 0.0 CAT3 0.84 CAT4 0.0 CAT5 0.0 CAT7 0.0	 Redacted: Annotated as CEM by AFP	Rank: 12 Skin Tone: 0.03% Porn: 0.12 Child: 0.99 CAT1 0.0 CAT2 0.0 CAT3 0.08 CAT4 0.91 CAT5 0.0 CAT7 0.0

TEST_CAT7: Bottom 20 Images			
Image	Details	Image	Details
 Redacted: Annotated as CEM by AFP	Rank: 13 Skin Tone: 0.03% Porn: 0.01 Child: 1.0 CAT1 0.0 CAT2 0.11 CAT3 0.87 CAT4 0.0 CAT5 0.0 CAT7 0.0	 Redacted: Annotated as CEM by AFP	Rank: 14 Skin Tone: 0.14% Porn: 0.08 Child: 0.87 CAT1 0.0 CAT2 0.0 CAT3 0.0 CAT4 0.94 CAT5 0.0 CAT7 0.03
 Redacted: Annotated as CEM by AFP	Rank: 15 Skin Tone: 0.35% Porn: 0.17 Child: 1.0 CAT1 0.09 CAT2 0.0 CAT3 0.63 CAT4 0.01 CAT5 0.24 CAT7 0.02	 Redacted: Annotated as CEM by AFP	Rank: 16 Skin Tone: 1.41% Porn: 0.01 Child: 1.0 CAT1 0.0 CAT2 0.99 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.0
 Redacted: Annotated as CEM by AFP	Rank: 17 Skin Tone: 1.48% Porn: 0.05 Child: 0.97 CAT1 0.04 CAT2 0.0 CAT3 0.01 CAT4 0.01 CAT5 0.0 CAT7 0.92	 Redacted: Annotated as CEM by AFP	Rank: 18 Skin Tone: 1.49% Porn: 0.0 Child: 0.99 CAT1 0.03 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.96

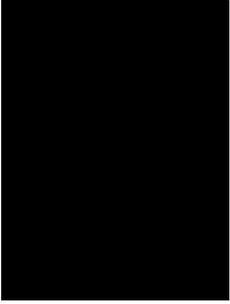
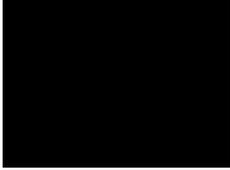
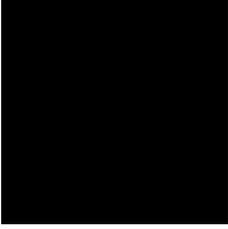
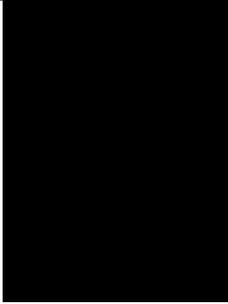
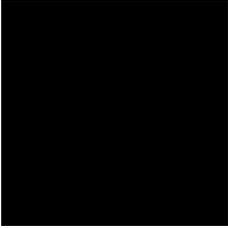
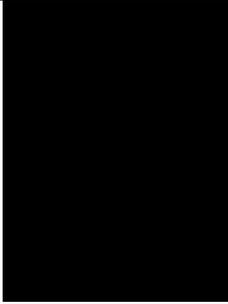
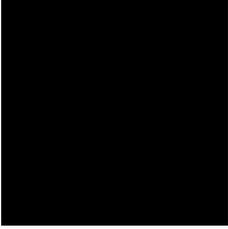
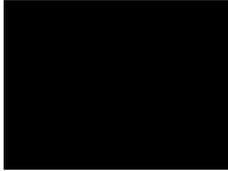
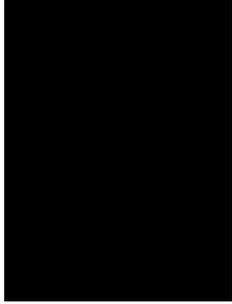
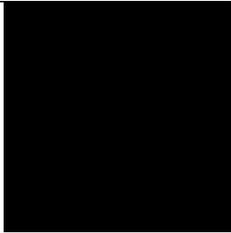
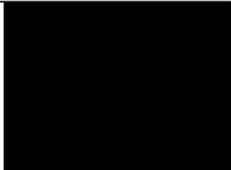
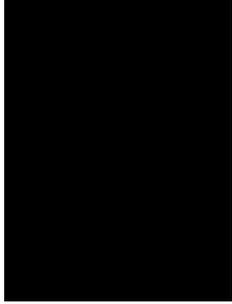
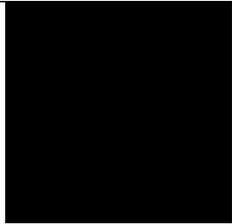
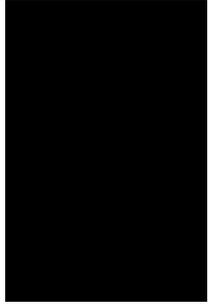
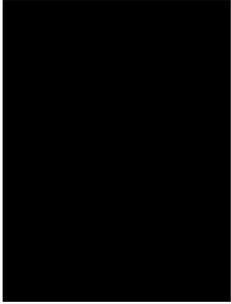
TEST_CAT7: Bottom 20 Images			
Image	Details	Image	Details
	Rank: 19 Skin Tone: 1.81% Porn: 0.02 Child: 1.0 CAT1 0.99 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.0		Rank: 20 Skin Tone: 1.83% Porn: 0.04 Child: 0.99 CAT1 0.02 CAT2 0.0 CAT3 0.01 CAT4 0.0 CAT5 0.0 CAT7 0.96
Redacted: Annotated as CEM by AFP		Redacted: Annotated as CEM by AFP	

Table B.24: Skin Tone - Test corpus CAT7 bottom 20 results

B.13 Test Corpus CAT7 Classifier Results

TEST_CAT7: Top 20 Images			
Image	Details	Image	Details
 <p>Redacted: Annotated as CEM by AFP</p>	<p>Rank: 1</p> <p>Skin Tone: 15.89%</p> <p>Porn: 0.99</p> <p>Child: 1.0</p> <p>CAT1 0.0</p> <p>CAT2 0.0</p> <p>CAT3 0.02</p> <p>CAT4 0.97</p> <p>CAT5 0.0</p> <p>CAT7 0.0</p>	 <p>Redacted: Annotated as CEM by AFP</p>	<p>Rank: 2</p> <p>Skin Tone: 38.68%</p> <p>Porn: 0.99</p> <p>Child: 0.99</p> <p>CAT1 0.99</p> <p>CAT2 0.0</p> <p>CAT3 0.0</p> <p>CAT4 0.0</p> <p>CAT5 0.0</p> <p>CAT7 0.0</p>
 <p>Redacted: Annotated as CEM by AFP</p>	<p>Rank: 3</p> <p>Skin Tone: 26.04%</p> <p>Porn: 0.99</p> <p>Child: 0.99</p> <p>CAT1 0.38</p> <p>CAT2 0.08</p> <p>CAT3 0.39</p> <p>CAT4 0.12</p> <p>CAT5 0.0</p> <p>CAT7 0.0</p>	 <p>Redacted: Annotated as CEM by AFP</p>	<p>Rank: 4</p> <p>Skin Tone: 58.53%</p> <p>Porn: 0.98</p> <p>Child: 1.0</p> <p>CAT1 0.94</p> <p>CAT2 0.0</p> <p>CAT3 0.0</p> <p>CAT4 0.0</p> <p>CAT5 0.0</p> <p>CAT7 0.05</p>
 <p>Redacted: Annotated as CEM by AFP</p>	<p>Rank: 5</p> <p>Skin Tone: 61.08%</p> <p>Porn: 0.98</p> <p>Child: 1.0</p> <p>CAT1 0.78</p> <p>CAT2 0.0</p> <p>CAT3 0.19</p> <p>CAT4 0.0</p> <p>CAT5 0.0</p> <p>CAT7 0.0</p>	 <p>Redacted: Annotated as CEM by AFP</p>	<p>Rank: 6</p> <p>Skin Tone: 82.60%</p> <p>Porn: 0.98</p> <p>Child: 0.0</p> <p>CAT1 0.55</p> <p>CAT2 0.0</p> <p>CAT3 0.0</p> <p>CAT4 0.0</p> <p>CAT5 0.0</p> <p>CAT7 0.44</p>

TEST_CAT7: Top 20 Images			
Image	Details	Image	Details
 Redacted: Annotated as CEM by AFP	Rank: 7 Skin Tone: 23.56% Porn: 0.97 Child: 0.99 CAT1 0.29 CAT2 0.0 CAT3 0.0 CAT4 0.66 CAT5 0.0 CAT7 0.02	 Redacted: Annotated as CEM by AFP	Rank: 8 Skin Tone: 0.00% Porn: 0.96 Child: 0.99 CAT1 0.98 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.0
 Redacted: Annotated as CEM by AFP	Rank: 9 Skin Tone: 13.82% Porn: 0.94 Child: 0.99 CAT1 0.0 CAT2 0.0 CAT3 0.0 CAT4 0.99 CAT5 0.0 CAT7 0.0	 Redacted: Annotated as CEM by AFP	Rank: 10 Skin Tone: 72.84% Porn: 0.94 Child: 0.52 CAT1 0.18 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.8
 Redacted: Annotated as CEM by AFP	Rank: 11 Skin Tone: 41.15% Porn: 0.92 Child: 0.61 CAT1 0.02 CAT2 0.0 CAT3 0.93 CAT4 0.0 CAT5 0.0 CAT7 0.03	 Redacted: Annotated as CEM by AFP	Rank: 12 Skin Tone: 35.44% Porn: 0.89 Child: 0.95 CAT1 0.29 CAT2 0.18 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.52

TEST_CAT7: Top 20 Images			
Image	Details	Image	Details
	<p>Rank: 13</p> <p>Skin Tone: 32.28%</p> <p>Porn: 0.86</p> <p>Child: 0.99</p> <p>CAT1 0.88</p> <p>CAT2 0.09</p> <p>CAT3 0.0</p> <p>CAT4 0.0</p> <p>CAT5 0.0</p> <p>CAT7 0.0</p>		<p>Rank: 14</p> <p>Skin Tone: 56.91%</p> <p>Porn: 0.86</p> <p>Child: 0.06</p> <p>CAT1 0.25</p> <p>CAT2 0.0</p> <p>CAT3 0.0</p> <p>CAT4 0.0</p> <p>CAT5 0.0</p> <p>CAT7 0.74</p>
Redacted: Annotated as CEM by AFP		Redacted: Annotated as CEM by AFP	
	<p>Rank: 15</p> <p>Skin Tone: 62.70%</p> <p>Porn: 0.85</p> <p>Child: 0.04</p> <p>CAT1 0.0</p> <p>CAT2 0.0</p> <p>CAT3 0.0</p> <p>CAT4 0.0</p> <p>CAT5 0.0</p> <p>CAT7 0.99</p>		<p>Rank: 16</p> <p>Skin Tone: 52.15%</p> <p>Porn: 0.84</p> <p>Child: 0.98</p> <p>CAT1 0.0</p> <p>CAT2 0.0</p> <p>CAT3 0.0</p> <p>CAT4 0.0</p> <p>CAT5 0.0</p> <p>CAT7 0.99</p>
Redacted: Annotated as CEM by AFP		Redacted: Annotated as CEM by AFP	
	<p>Rank: 17</p> <p>Skin Tone: 35.36%</p> <p>Porn: 0.84</p> <p>Child: 0.3</p> <p>CAT1 0.11</p> <p>CAT2 0.0</p> <p>CAT3 0.0</p> <p>CAT4 0.0</p> <p>CAT5 0.0</p> <p>CAT7 0.88</p>		<p>Rank: 18</p> <p>Skin Tone: 62.92%</p> <p>Porn: 0.82</p> <p>Child: 0.0</p> <p>CAT1 0.04</p> <p>CAT2 0.01</p> <p>CAT3 0.0</p> <p>CAT4 0.0</p> <p>CAT5 0.0</p> <p>CAT7 0.94</p>
Redacted: Annotated as CEM by AFP		Redacted: Annotated as CEM by AFP	

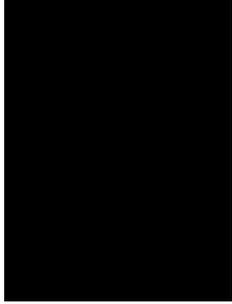
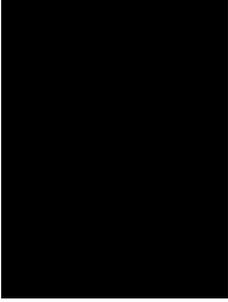
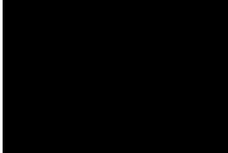
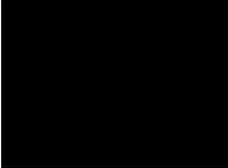
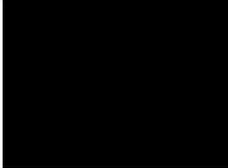
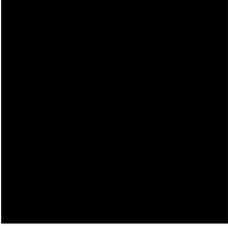
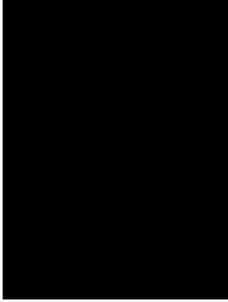
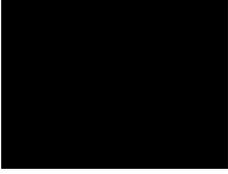
TEST_CAT7: Top 20 Images			
Image	Details	Image	Details
	Rank: 19 Skin Tone: 62.25% Porn: 0.81 Child: 0.99 CAT1 0.97 CAT2 0.0 CAT3 0.01 CAT4 0.0 CAT5 0.0 CAT7 0.01		Rank: 20 Skin Tone: 54.09% Porn: 0.8 Child: 0.09 CAT1 0.08 CAT2 0.09 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.82
Redacted: Annotated as CEM by AFP		Redacted: Annotated as CEM by AFP	

Table B.25: Test corpus CAT7 top 20 results by classifier (porn,child)

TEST_CAT7: Bottom 20 Images			
Image	Details	Image	Details
	Rank: 1 Skin Tone: 26.29% Porn: 0.0 Child: 0.0 CAT1 0.0 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.99		Rank: 2 Skin Tone: 9.95% Porn: 0.0 Child: 0.0 CAT1 0.55 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.44
Redacted: Annotated as CEM by AFP		Redacted: Annotated as CEM by AFP	
	Rank: 3 Skin Tone: 9.39% Porn: 0.0 Child: 0.0 CAT1 0.0 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.99		Rank: 4 Skin Tone: 27.42% Porn: 0.0 Child: 0.0 CAT1 0.0 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.99
Redacted: Annotated as CEM by AFP		Redacted: Annotated as CEM by AFP	
	Rank: 5 Skin Tone: 34.84% Porn: 0.0 Child: 0.0 CAT1 0.25 CAT2 0.0 CAT3 0.19 CAT4 0.0 CAT5 0.0 CAT7 0.54		Rank: 6 Skin Tone: 44.07% Porn: 0.0 Child: 0.0 CAT1 0.21 CAT2 0.07 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.7
Redacted: Annotated as CEM by AFP		Redacted: Annotated as CEM by AFP	

TEST_CAT7: Bottom 20 Images			
Image	Details	Image	Details
 Redacted: Annotated as CEM by AFP	Rank: 7 Skin Tone: 26.66% Porn: 0.0 Child: 0.0 CAT1 0.0 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.99	 Redacted: Annotated as CEM by AFP	Rank: 8 Skin Tone: 33.95% Porn: 0.0 Child: 0.01 CAT1 0.0 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.99
 Redacted: Annotated as CEM by AFP	Rank: 9 Skin Tone: 3.65% Porn: 0.0 Child: 0.01 CAT1 0.61 CAT2 0.0 CAT3 0.11 CAT4 0.06 CAT5 0.0 CAT7 0.19	 Redacted: Annotated as CEM by AFP	Rank: 10 Skin Tone: 23.99% Porn: 0.0 Child: 0.04 CAT1 0.0 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.99
 Redacted: Annotated as CEM by AFP	Rank: 11 Skin Tone: 2.99% Porn: 0.0 Child: 0.04 CAT1 0.75 CAT2 0.0 CAT3 0.07 CAT4 0.08 CAT5 0.0 CAT7 0.08	 Redacted: Annotated as CEM by AFP	Rank: 12 Skin Tone: 14.14% Porn: 0.0 Child: 0.04 CAT1 0.95 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.04

TEST_CAT7: Bottom 20 Images			
Image	Details	Image	Details
 Redacted: Annotated as CEM by AFP	Rank: 13 Skin Tone: 3.31% Porn: 0.0 Child: 0.06 CAT1 0.5 CAT2 0.0 CAT3 0.09 CAT4 0.23 CAT5 0.0 CAT7 0.15	 Redacted: Annotated as CEM by AFP	Rank: 14 Skin Tone: 27.08% Porn: 0.0 Child: 0.06 CAT1 0.0 CAT2 0.18 CAT3 0.07 CAT4 0.0 CAT5 0.0 CAT7 0.72
 Redacted: Annotated as CEM by AFP	Rank: 15 Skin Tone: 16.38% Porn: 0.0 Child: 0.1 CAT1 0.49 CAT2 0.0 CAT3 0.0 CAT4 0.26 CAT5 0.0 CAT7 0.24	 Redacted: Annotated as CEM by AFP	Rank: 16 Skin Tone: 3.91% Porn: 0.0 Child: 0.11 CAT1 0.68 CAT2 0.0 CAT3 0.17 CAT4 0.04 CAT5 0.0 CAT7 0.09
 Redacted: Annotated as CEM by AFP	Rank: 17 Skin Tone: 22.10% Porn: 0.0 Child: 0.12 CAT1 0.23 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.75	 Redacted: Annotated as CEM by AFP	Rank: 18 Skin Tone: 13.66% Porn: 0.0 Child: 0.16 CAT1 0.08 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.91

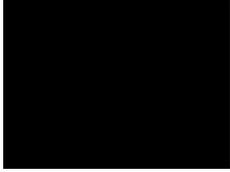
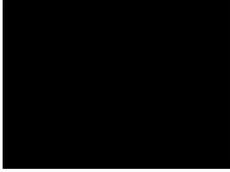
TEST_CAT7: Bottom 20 Images			
Image	Details	Image	Details
	Rank: 19 Skin Tone: 9.26% Porn: 0.0 Child: 0.16 CAT1 0.11 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.88		Rank: 20 Skin Tone: 9.26% Porn: 0.0 Child: 0.16 CAT1 0.11 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.88
Redacted: Annotated as CEM by AFP		Redacted: Annotated as CEM by AFP	

Table B.26: Test corpus CAT7 bottom 20 results by classifier (porn,child)

B.14 ImageNet corpus Skin Tone Results

WARNING
CONTAINS/MAY CONTAIN SEXUALLY EXPLICIT
IMAGERY

ImageNet: Top 20 Images			
Image	Details	Image	Details
	Rank: 1 Skin Tone: 100.00% Porn: 0.0 Child: 0.92 CAT1 0.97 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.02		Rank: 2 Skin Tone: 100.00% Porn: 0.09 Child: 0.92 CAT1 0.3 CAT2 0.0 CAT3 0.33 CAT4 0.07 CAT5 0.27 CAT7 0.01
	Rank: 3 Skin Tone: 100.00% Porn: 0.1 Child: 0.71 CAT1 0.22 CAT2 0.06 CAT3 0.6 CAT4 0.08 CAT5 0.0 CAT7 0.0		Rank: 4 Skin Tone: 100.00% Porn: 0.36 Child: 0.99 CAT1 0.3 CAT2 0.07 CAT3 0.54 CAT4 0.04 CAT5 0.0 CAT7 0.02
	Rank: 5 Skin Tone: 100.00% Porn: 0.15 Child: 0.05 CAT1 0.99 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.0		Rank: 6 Skin Tone: 100.00% Porn: 0.02 Child: 0.61 CAT1 0.7 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.01 CAT7 0.28

ImageNet: Top 20 Images			
Image	Details	Image	Details
	Rank: 7 Skin Tone: 100.00% Porn: 0.04 Child: 0.99 CAT1 0.0 CAT2 0.0 CAT3 0.98 CAT4 0.0 CAT5 0.0 CAT7 0.0		Rank: 8 Skin Tone: 100.00% Porn: 0.0 Child: 0.03 CAT1 0.56 CAT2 0.03 CAT3 0.0 CAT4 0.0 CAT5 0.02 CAT7 0.37
	Rank: 9 Skin Tone: 100.00% Porn: 0.03 Child: 0.99 CAT1 0.87 CAT2 0.0 CAT3 0.09 CAT4 0.0 CAT5 0.0 CAT7 0.01		Rank: 10 Skin Tone: 100.00% Porn: 0.02 Child: 0.99 CAT1 0.99 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.0
	Rank: 11 Skin Tone: 100.00% Porn: 0.06 Child: 0.87 CAT1 0.08 CAT2 0.0 CAT3 0.9 CAT4 0.0 CAT5 0.0 CAT7 0.0		Rank: 12 Skin Tone: 100.00% Porn: 0.77 Child: 0.0 CAT1 0.01 CAT2 0.0 CAT3 0.72 CAT4 0.25 CAT5 0.0 CAT7 0.0
	Rank: 13 Skin Tone: 100.00% Porn: 0.01 Child: 0.02 CAT1 0.72 CAT2 0.0 CAT3 0.07 CAT4 0.02 CAT5 0.0 CAT7 0.17		Rank: 14 Skin Tone: 100.00% Porn: 0.24 Child: 0.89 CAT1 0.04 CAT2 0.0 CAT3 0.91 CAT4 0.03 CAT5 0.0 CAT7 0.0

ImageNet: Top 20 Images			
Image	Details	Image	Details
	Rank: 15 Skin Tone: 100.00% Porn: 0.0 Child: 0.99 CAT1 0.0 CAT2 0.0 CAT3 0.04 CAT4 0.95 CAT5 0.0 CAT7 0.0		Rank: 16 Skin Tone: 100.00% Porn: 0.02 Child: 0.0 CAT1 0.05 CAT2 0.0 CAT3 0.94 CAT4 0.0 CAT5 0.0 CAT7 0.0
	Rank: 17 Skin Tone: 100.00% Porn: 0.0 Child: 0.99 CAT1 0.0 CAT2 0.0 CAT3 0.99 CAT4 0.0 CAT5 0.0 CAT7 0.0		Rank: 18 Skin Tone: 100.00% Porn: 0.03 Child: 1.0 CAT1 0.0 CAT2 0.0 CAT3 0.99 CAT4 0.0 CAT5 0.0 CAT7 0.0
	Rank: 19 Skin Tone: 100.00% Porn: 0.02 Child: 0.67 CAT1 0.0 CAT2 0.0 CAT3 0.99 CAT4 0.0 CAT5 0.0 CAT7 0.0		Rank: 20 Skin Tone: 100.00% Porn: 0.1 Child: 0.74 CAT1 0.19 CAT2 0.0 CAT3 0.65 CAT4 0.11 CAT5 0.01 CAT7 0.01

Table B.27: ImageNet top 20 results by skin tone percentage

ImageNet: Bottom 20 Images			
Image	Details	Image	Details
	Rank: 1 Skin Tone: 0.00% Porn: 0.0 Child: 0.2 CAT1 0.0 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.99		Rank: 2 Skin Tone: 0.00% Porn: 0.0 Child: 0.0 CAT1 0.93 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.06
	Rank: 3 Skin Tone: 0.00% Porn: 0.09 Child: 0.99 CAT1 0.62 CAT2 0.0 CAT3 0.2 CAT4 0.09 CAT5 0.0 CAT7 0.07		Rank: 4 Skin Tone: 0.00% Porn: 0.0 Child: 0.99 CAT1 0.24 CAT2 0.0 CAT3 0.64 CAT4 0.04 CAT5 0.0 CAT7 0.05
	Rank: 5 Skin Tone: 0.00% Porn: 0.01 Child: 1.0 CAT1 0.97 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.02		Rank: 6 Skin Tone: 0.00% Porn: 0.02 Child: 0.99 CAT1 0.45 CAT2 0.0 CAT3 0.06 CAT4 0.27 CAT5 0.0 CAT7 0.2
	Rank: 7 Skin Tone: 0.00% Porn: 0.01 Child: 0.99 CAT1 0.15 CAT2 0.0 CAT3 0.0 CAT4 0.01 CAT5 0.0 CAT7 0.82		Rank: 8 Skin Tone: 0.00% Porn: 0.25 Child: 0.99 CAT1 0.0 CAT2 0.0 CAT3 0.0 CAT4 0.98 CAT5 0.0 CAT7 0.0

ImageNet: Bottom 20 Images			
Image	Details	Image	Details
	Rank: 9 Skin Tone: 0.00% Porn: 0.0 Child: 0.98 CAT1 0.99 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.0		Rank: 10 Skin Tone: 0.00% Porn: 0.0 Child: 0.0 CAT1 0.49 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.5
	Rank: 11 Skin Tone: 0.00% Porn: 0.01 Child: 0.99 CAT1 0.02 CAT2 0.0 CAT3 0.84 CAT4 0.1 CAT5 0.0 CAT7 0.01		Rank: 12 Skin Tone: 0.00% Porn: 0.0 Child: 1.0 CAT1 0.31 CAT2 0.0 CAT3 0.67 CAT4 0.0 CAT5 0.0 CAT7 0.0
	Rank: 13 Skin Tone: 0.00% Porn: 0.0 Child: 0.34 CAT1 0.51 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.48		Rank: 14 Skin Tone: 0.00% Porn: 0.02 Child: 0.99 CAT1 0.52 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.47
	Rank: 15 Skin Tone: 0.00% Porn: 0.0 Child: 0.99 CAT1 0.02 CAT2 0.0 CAT3 0.57 CAT4 0.38 CAT5 0.0 CAT7 0.0		Rank: 16 Skin Tone: 0.00% Porn: 0.0 Child: 0.97 CAT1 0.35 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.64

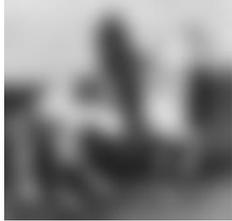
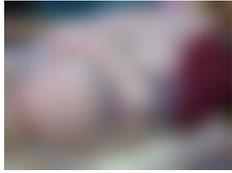
ImageNet: Bottom 20 Images			
Image	Details	Image	Details
	Rank: 17 Skin Tone: 0.00% Porn: 0.22 Child: 0.99 CAT1 0.38 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.04 CAT7 0.57		Rank: 18 Skin Tone: 0.00% Porn: 0.0 Child: 0.99 CAT1 0.25 CAT2 0.0 CAT3 0.56 CAT4 0.07 CAT5 0.0 CAT7 0.09
	Rank: 19 Skin Tone: 0.00% Porn: 0.14 Child: 1.0 CAT1 0.02 CAT2 0.06 CAT3 0.0 CAT4 0.9 CAT5 0.0 CAT7 0.0		Rank: 20 Skin Tone: 0.00% Porn: 0.11 Child: 0.97 CAT1 0.96 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.02

Table B.28: ImageNet bottom 20 results by skin tone percentage

B.15 ImageNet Corpus Classifier Results

WARNING
CONTAINS/MAY CONTAIN SEXUALLY EXPLICIT
IMAGERY

ImageNet: Top 20 Images			
Image	Details	Image	Details
	Rank: 1 Skin Tone: 25.41% Porn: 0.99 Child: 1.0 CAT1 0.0 CAT2 0.0 CAT3 0.69 CAT4 0.3 CAT5 0.0 CAT7 0.0		Rank: 2 Skin Tone: 69.08% Porn: 0.99 Child: 1.0 CAT1 0.03 CAT2 0.0 CAT3 0.96 CAT4 0.0 CAT5 0.0 CAT7 0.0
	Rank: 3 Skin Tone: 57.37% Porn: 0.99 Child: 0.99 CAT1 0.35 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.63		Rank: 4 Skin Tone: 69.98% Porn: 0.99 Child: 0.99 CAT1 0.0 CAT2 0.0 CAT3 0.97 CAT4 0.02 CAT5 0.0 CAT7 0.0
	Rank: 5 Skin Tone: 94.34% Porn: 0.99 Child: 0.99 CAT1 0.01 CAT2 0.0 CAT3 0.02 CAT4 0.93 CAT5 0.0 CAT7 0.02		Rank: 6 Skin Tone: 24.57% Porn: 0.99 Child: 0.99 CAT1 0.77 CAT2 0.0 CAT3 0.12 CAT4 0.09 CAT5 0.0 CAT7 0.0

ImageNet: Top 20 Images			
Image	Details	Image	Details
	Rank: 7 Skin Tone: 24.11% Porn: 0.99 Child: 0.99 CAT1 0.0 CAT2 0.0 CAT3 0.17 CAT4 0.74 CAT5 0.0 CAT7 0.07		Rank: 8 Skin Tone: 70.14% Porn: 0.99 Child: 0.99 CAT1 0.0 CAT2 0.0 CAT3 0.88 CAT4 0.11 CAT5 0.0 CAT7 0.0
	Rank: 9 Skin Tone: 82.38% Porn: 0.99 Child: 0.99 CAT1 0.02 CAT2 0.0 CAT3 0.92 CAT4 0.04 CAT5 0.0 CAT7 0.0		Rank: 10 Skin Tone: 56.32% Porn: 0.99 Child: 0.99 CAT1 0.0 CAT2 0.0 CAT3 0.94 CAT4 0.04 CAT5 0.0 CAT7 0.0
	Rank: 11 Skin Tone: 58.86% Porn: 0.99 Child: 0.99 CAT1 0.0 CAT2 0.0 CAT3 0.44 CAT4 0.55 CAT5 0.0 CAT7 0.0		Rank: 12 Skin Tone: 52.82% Porn: 0.99 Child: 0.99 CAT1 0.0 CAT2 0.0 CAT3 0.03 CAT4 0.96 CAT5 0.0 CAT7 0.0

ImageNet: Top 20 Images			
Image	Details	Image	Details
	Rank: 13 Skin Tone: 55.99% Porn: 0.99 Child: 0.99 CAT1 0.0 CAT2 0.0 CAT3 0.04 CAT4 0.92 CAT5 0.0 CAT7 0.02		Rank: 14 Skin Tone: 51.01% Porn: 0.99 Child: 0.99 CAT1 0.0 CAT2 0.0 CAT3 0.0 CAT4 0.99 CAT5 0.0 CAT7 0.0
	Rank: 15 Skin Tone: 64.12% Porn: 0.99 Child: 0.99 CAT1 0.99 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.0		Rank: 16 Skin Tone: 7.80% Porn: 0.99 Child: 0.99 CAT1 0.0 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.99
	Rank: 17 Skin Tone: 24.59% Porn: 0.99 Child: 0.99 CAT1 0.0 CAT2 0.0 CAT3 0.69 CAT4 0.29 CAT5 0.0 CAT7 0.0		Rank: 18 Skin Tone: 22.54% Porn: 0.99 Child: 0.99 CAT1 0.09 CAT2 0.0 CAT3 0.12 CAT4 0.25 CAT5 0.0 CAT7 0.52

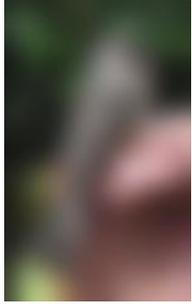
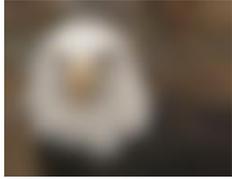
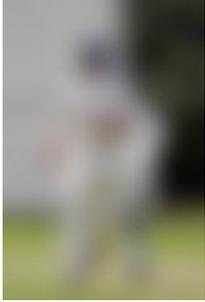
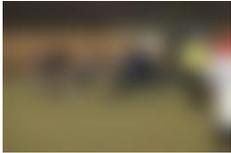
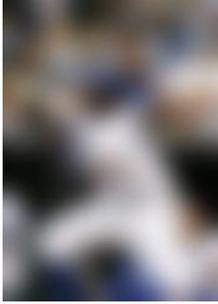
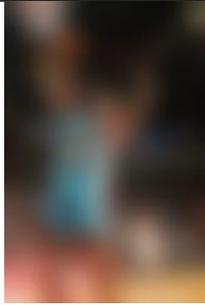
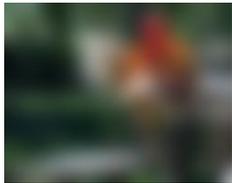
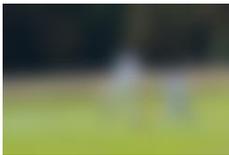
ImageNet: Top 20 Images			
Image	Details	Image	Details
	Rank: 19 Skin Tone: 25.59% Porn: 0.99 Child: 0.99 CAT1 0.22 CAT2 0.0 CAT3 0.25 CAT4 0.22 CAT5 0.0 CAT7 0.3		Rank: 20 Skin Tone: 58.64% Porn: 0.99 Child: 0.99 CAT1 0.0 CAT2 0.0 CAT3 0.14 CAT4 0.85 CAT5 0.0 CAT7 0.0

Table B.29: ImageNet top 20 results by classifier (porn,child)

ImageNet: Bottom 20 Images			
Image	Details	Image	Details
	Rank: 1 Skin Tone: 20.21% Porn: 0.0 Child: 0.0 CAT1 0.6 CAT2 0.0 CAT3 0.0 CAT4 0.03 CAT5 0.0 CAT7 0.34		Rank: 2 Skin Tone: 21.35% Porn: 0.0 Child: 0.0 CAT1 0.97 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.02
	Rank: 3 Skin Tone: 42.11% Porn: 0.0 Child: 0.0 CAT1 0.0 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.99		Rank: 4 Skin Tone: 2.49% Porn: 0.0 Child: 0.0 CAT1 0.02 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.96
	Rank: 5 Skin Tone: 5.89% Porn: 0.0 Child: 0.0 CAT1 0.07 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.92		Rank: 6 Skin Tone: 46.97% Porn: 0.0 Child: 0.0 CAT1 0.0 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.99

ImageNet: Bottom 20 Images			
Image	Details	Image	Details
	Rank: 7 Skin Tone: 15.02% Porn: 0.0 Child: 0.0 CAT1 0.02 CAT2 0.0 CAT3 0.0 CAT4 0.5 CAT5 0.0 CAT7 0.47		Rank: 8 Skin Tone: 3.14% Porn: 0.0 Child: 0.0 CAT1 0.85 CAT2 0.04 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.09
	Rank: 9 Skin Tone: 2.66% Porn: 0.0 Child: 0.0 CAT1 0.0 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.99		Rank: 10 Skin Tone: 18.35% Porn: 0.0 Child: 0.0 CAT1 0.93 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.05
	Rank: 11 Skin Tone: 46.07% Porn: 0.0 Child: 0.0 CAT1 0.16 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.83		Rank: 12 Skin Tone: 64.97% Porn: 0.0 Child: 0.0 CAT1 0.04 CAT2 0.79 CAT3 0.0 CAT4 0.02 CAT5 0.0 CAT7 0.13

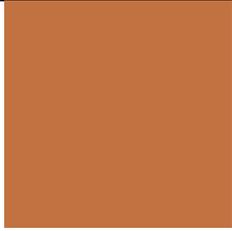
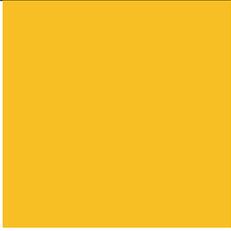
ImageNet: Bottom 20 Images			
Image	Details	Image	Details
	Rank: 13 Skin Tone: 10.42% Porn: 0.0 Child: 0.0 CAT1 0.0 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.99		Rank: 14 Skin Tone: 12.98% Porn: 0.0 Child: 0.0 CAT1 0.32 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.67
	Rank: 15 Skin Tone: 22.74% Porn: 0.0 Child: 0.0 CAT1 0.0 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.99		Rank: 16 Skin Tone: 7.34% Porn: 0.0 Child: 0.0 CAT1 0.29 CAT2 0.61 CAT3 0.01 CAT4 0.03 CAT5 0.0 CAT7 0.04
	Rank: 17 Skin Tone: 6.51% Porn: 0.0 Child: 0.0 CAT1 0.0 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.99		Rank: 18 Skin Tone: 5.48% Porn: 0.0 Child: 0.0 CAT1 0.97 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.02

ImageNet: Bottom 20 Images					
Image	Details		Image	Details	
	Rank:	19		Rank:	20
	Skin Tone:	0.27%		Skin Tone:	9.96%
	Porn:	0.0		Porn:	0.0
	Child:	0.0		Child:	0.0
	CAT1	0.77		CAT1	0.68
	CAT2	0.0		CAT2	0.0
	CAT3	0.0		CAT3	0.0
	CAT4	0.0		CAT4	0.0
	CAT5	0.0		CAT5	0.0
	CAT7	0.2		CAT7	0.31

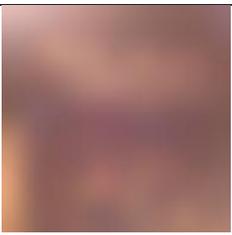
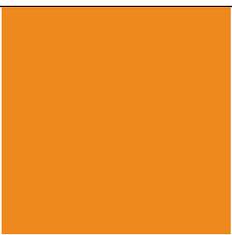
Table B.30: ImageNet bottom 20 results by classifier (porn,child)

B.16 TorCrawl corpus Skin Tone Results

WARNING
CONTAINS/MAY CONTAIN SEXUALLY EXPLICIT
IMAGERY

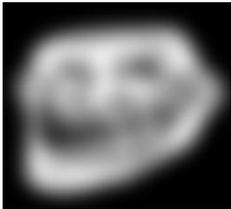
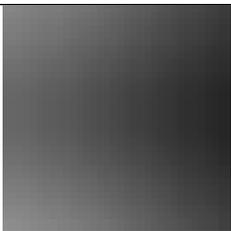
TorCrawl: Top 20 Images			
Image	Details	Image	Details
	Rank: 1 Skin Tone: 100.00% Porn: 0.0 Child: 0.66 CAT1 0.75 CAT2 0.0 CAT3 0.08 CAT4 0.04 CAT5 0.0 CAT7 0.11		Rank: 2 Skin Tone: 100.00% Porn: 0.01 Child: 0.82 CAT1 0.0 CAT2 0.0 CAT3 0.08 CAT4 0.0 CAT5 0.0 CAT7 0.9
	Rank: 3 Skin Tone: 100.00% Porn: 0.17 Child: 0.01 CAT1 0.88 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.11		Rank: 4 Skin Tone: 100.00% Porn: 0.07 Child: 0.66 CAT1 0.63 CAT2 0.01 CAT3 0.11 CAT4 0.02 CAT5 0.01 CAT7 0.19
	Rank: 5 Skin Tone: 100.00% Porn: 0.0 Child: 0.02 CAT1 0.19 CAT2 0.0 CAT3 0.4 CAT4 0.3 CAT5 0.0 CAT7 0.09		Rank: 6 Skin Tone: 100.00% Porn: 0.0 Child: 0.99 CAT1 0.92 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.07

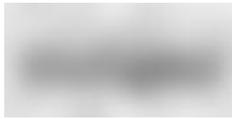
TorCrawl: Top 20 Images			
Image	Details	Image	Details
	Rank: 7 Skin Tone: 100.00% Porn: 0.0 Child: 0.93 CAT1 0.54 CAT2 0.01 CAT3 0.01 CAT4 0.0 CAT5 0.0 CAT7 0.4		Rank: 8 Skin Tone: 100.00% Porn: 0.0 Child: 0.86 CAT1 0.6 CAT2 0.04 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.35
	Rank: 9 Skin Tone: 100.00% Porn: 0.0 Child: 0.0 CAT1 0.57 CAT2 0.0 CAT3 0.02 CAT4 0.22 CAT5 0.0 CAT7 0.17		Rank: 10 Skin Tone: 100.00% Porn: 0.03 Child: 0.08 CAT1 0.91 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.08
	Rank: 11 Skin Tone: 100.00% Porn: 0.0 Child: 0.03 CAT1 0.02 CAT2 0.0 CAT3 0.0 CAT4 0.88 CAT5 0.0 CAT7 0.09		Rank: 12 Skin Tone: 100.00% Porn: 0.0 Child: 0.99 CAT1 0.0 CAT2 0.0 CAT3 0.99 CAT4 0.0 CAT5 0.0 CAT7 0.0

TorCrawl: Top 20 Images			
Image	Details	Image	Details
	Rank: 13 Skin Tone: 100.00% Porn: 0.05 Child: 1.0 CAT1 0.93 CAT2 0.0 CAT3 0.02 CAT4 0.0 CAT5 0.0 CAT7 0.02		Rank: 14 Skin Tone: 100.00% Porn: 0.0 Child: 0.07 CAT1 0.98 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.01
	Rank: 15 Skin Tone: 100.00% Porn: 0.05 Child: 0.98 CAT1 0.93 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.06		Rank: 16 Skin Tone: 100.00% Porn: 0.0 Child: 0.88 CAT1 0.36 CAT2 0.07 CAT3 0.43 CAT4 0.02 CAT5 0.0 CAT7 0.1
	Rank: 17 Skin Tone: 100.00% Porn: 0.12 Child: 0.36 CAT1 0.41 CAT2 0.01 CAT3 0.1 CAT4 0.0 CAT5 0.0 CAT7 0.45		Rank: 18 Skin Tone: 100.00% Porn: 0.03 Child: 0.23 CAT1 0.93 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.06

TorCrawl: Top 20 Images			
Image	Details	Image	Details
	Rank: 19 Skin Tone: 100.00% Porn: 0.13 Child: 0.05 CAT1 0.74 CAT2 0.01 CAT3 0.05 CAT4 0.01 CAT5 0.01 CAT7 0.16		Rank: 20 Skin Tone: 100.00% Porn: 0.01 Child: 0.25 CAT1 0.26 CAT2 0.02 CAT3 0.0 CAT4 0.0 CAT5 0.01 CAT7 0.68

Table B.31: TorCrawl top 20 results by skin tone percentage

TorCrawl: Bottom 20 Images			
Image	Details	Image	Details
	Rank: 1 Skin Tone: 0.00% Porn: 0.02 Child: 0.0 CAT1 0.53 CAT2 0.06 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.39		Rank: 2 Skin Tone: 0.00% Porn: 0.0 Child: 0.0 CAT1 0.01 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.98
	Rank: 3 Skin Tone: 0.00% Porn: 0.0 Child: 1.0 CAT1 0.0 CAT2 0.0 CAT3 0.96 CAT4 0.02 CAT5 0.0 CAT7 0.0		Rank: 4 Skin Tone: 0.00% Porn: 0.01 Child: 0.0 CAT1 0.61 CAT2 0.05 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.32
	Rank: 5 Skin Tone: 0.00% Porn: 0.02 Child: 0.42 CAT1 0.0 CAT2 0.0 CAT3 0.4 CAT4 0.06 CAT5 0.0 CAT7 0.53		Rank: 6 Skin Tone: 0.00% Porn: 0.0 Child: 1.0 CAT1 0.0 CAT2 0.0 CAT3 0.99 CAT4 0.0 CAT5 0.0 CAT7 0.0
	Rank: 7 Skin Tone: 0.00% Porn: 0.01 Child: 0.0 CAT1 0.32 CAT2 0.2 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.46		Rank: 8 Skin Tone: 0.00% Porn: 0.0 Child: 0.97 CAT1 0.99 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.0

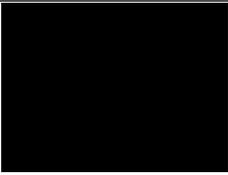
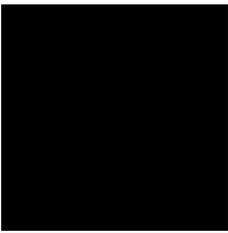
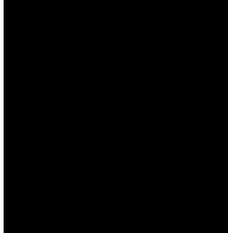
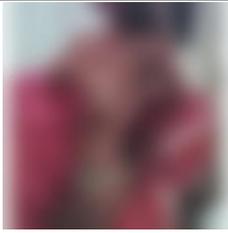
TorCrawl: Bottom 20 Images			
Image	Details	Image	Details
	Rank: 9 Skin Tone: 0.00% Porn: 0.0 Child: 0.0 CAT1 0.52 CAT2 0.16 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.3		Rank: 10 Skin Tone: 0.00% Porn: 0.01 Child: 0.92 CAT1 0.0 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.99
	Rank: 11 Skin Tone: 0.00% Porn: 0.02 Child: 0.0 CAT1 0.31 CAT2 0.03 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.63		Rank: 12 Skin Tone: 0.00% Porn: 0.0 Child: 0.97 CAT1 0.05 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.94
	Rank: 13 Skin Tone: 0.00% Porn: 0.02 Child: 0.0 CAT1 0.25 CAT2 0.01 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.73		Rank: 14 Skin Tone: 0.00% Porn: 0.01 Child: 0.98 CAT1 0.64 CAT2 0.03 CAT3 0.19 CAT4 0.06 CAT5 0.0 CAT7 0.06
	Rank: 15 Skin Tone: 0.00% Porn: 0.0 Child: 0.08 CAT1 0.41 CAT2 0.0 CAT3 0.0 CAT4 0.06 CAT5 0.0 CAT7 0.51		Rank: 16 Skin Tone: 0.00% Porn: 0.04 Child: 0.0 CAT1 0.2 CAT2 0.07 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.72

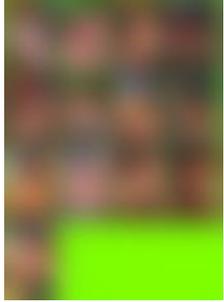
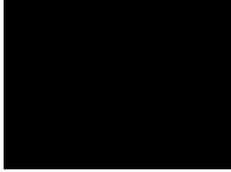
TorCrawl: Bottom 20 Images			
Image	Details	Image	Details
	Rank: 17 Skin Tone: 0.00% Porn: 0.04 Child: 0.0 CAT1 0.22 CAT2 0.04 CAT3 0.01 CAT4 0.0 CAT5 0.01 CAT7 0.7		Rank: 18 Skin Tone: 0.00% Porn: 0.0 Child: 0.83 CAT1 0.0 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.98
	Rank: 19 Skin Tone: 0.00% Porn: 0.0 Child: 0.07 CAT1 0.29 CAT2 0.31 CAT3 0.0 CAT4 0.14 CAT5 0.0 CAT7 0.23		Rank: 20 Skin Tone: 0.00% Porn: 0.0 Child: 0.01 CAT1 0.99 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.0

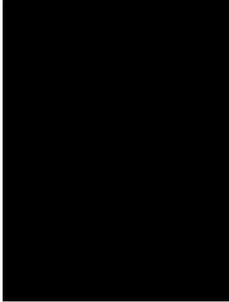
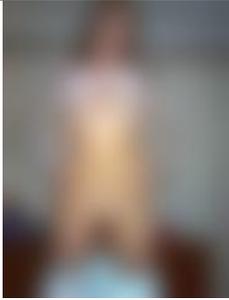
Table B.32: TorCrawl bottom 20 results by skin tone percentage

B.17 TorCrawl Corpus Classifier Results

WARNING
CONTAINS/MAY CONTAIN SEXUALLY EXPLICIT
IMAGERY

TorCrawl: Top 20 Images			
Image	Details	Image	Details
 Redacted: Confirmed CEM	Rank: 1 Skin Tone: 48.26% Porn: 0.99 Child: 1.0 CAT1 0.0 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.99 CAT7 0.0	 Redacted: Confirmed CEM	Rank: 2 Skin Tone: 55.83% Porn: 0.99 Child: 1.0 CAT1 0.0 CAT2 0.0 CAT3 0.29 CAT4 0.69 CAT5 0.0 CAT7 0.0
 Redacted: Confirmed CEM	Rank: 3 Skin Tone: 74.18% Porn: 0.99 Child: 1.0 CAT1 0.0 CAT2 0.0 CAT3 0.0 CAT4 0.99 CAT5 0.0 CAT7 0.0	 Redacted: Possible CEM	Rank: 4 Skin Tone: 44.43% Porn: 0.99 Child: 1.0 CAT1 0.06 CAT2 0.0 CAT3 0.04 CAT4 0.56 CAT5 0.28 CAT7 0.03
	Rank: 5 Skin Tone: 37.25% Porn: 0.99 Child: 1.0 CAT1 0.0 CAT2 0.01 CAT3 0.0 CAT4 0.98 CAT5 0.0 CAT7 0.0		Rank: 6 Skin Tone: 53.55% Porn: 0.99 Child: 1.0 CAT1 0.0 CAT2 0.0 CAT3 0.01 CAT4 0.98 CAT5 0.0 CAT7 0.0

TorCrawl: Top 20 Images			
Image	Details	Image	Details
	Rank: 7 Skin Tone: 24.68% Porn: 0.99 Child: 1.0 CAT1: 0.66 CAT2: 0.0 CAT3: 0.0 CAT4: 0.0 CAT5: 0.0 CAT7: 0.33		Rank: 8 Skin Tone: 68.77% Porn: 0.99 Child: 1.0 CAT1: 0.05 CAT2: 0.0 CAT3: 0.0 CAT4: 0.18 CAT5: 0.0 CAT7: 0.75
	Rank: 9 Skin Tone: 38.18% Porn: 0.99 Child: 1.0 CAT1: 0.99 CAT2: 0.0 CAT3: 0.0 CAT4: 0.0 CAT5: 0.0 CAT7: 0.0		Rank: 10 Skin Tone: 0.00% Porn: 0.99 Child: 1.0 CAT1: 0.0 CAT2: 0.76 CAT3: 0.0 CAT4: 0.0 CAT5: 0.0 CAT7: 0.21
	Rank: 11 Skin Tone: 13.90% Porn: 0.99 Child: 1.0 CAT1: 0.65 CAT2: 0.0 CAT3: 0.01 CAT4: 0.0 CAT5: 0.0 CAT7: 0.31	 Redacted: Confirmed CEM	Rank: 12 Skin Tone: 72.92% Porn: 0.99 Child: 1.0 CAT1: 0.02 CAT2: 0.29 CAT3: 0.12 CAT4: 0.56 CAT5: 0.0 CAT7: 0.0

TorCrawl: Top 20 Images			
Image	Details	Image	Details
 <p>Redacted: Confirmed CEM</p>	<p>Rank: 13 Skin Tone: 6.25% Porn: 0.99 Child: 1.0 CAT1: 0.3 CAT2: 0.0 CAT3: 0.0 CAT4: 0.01 CAT5: 0.0 CAT7: 0.67</p>	 <p>Redacted: Possible CEM</p>	<p>Rank: 14 Skin Tone: 68.35% Porn: 0.99 Child: 1.0 CAT1: 0.91 CAT2: 0.0 CAT3: 0.08 CAT4: 0.0 CAT5: 0.0 CAT7: 0.0</p>
	<p>Rank: 15 Skin Tone: 60.66% Porn: 0.99 Child: 1.0 CAT1: 0.0 CAT2: 0.0 CAT3: 0.0 CAT4: 0.99 CAT5: 0.0 CAT7: 0.0</p>		<p>Rank: 16 Skin Tone: 26.37% Porn: 0.99 Child: 1.0 CAT1: 0.98 CAT2: 0.0 CAT3: 0.0 CAT4: 0.0 CAT5: 0.0 CAT7: 0.0</p>
	<p>Rank: 17 Skin Tone: 6.06% Porn: 0.99 Child: 1.0 CAT1: 0.18 CAT2: 0.38 CAT3: 0.08 CAT4: 0.33 CAT5: 0.0 CAT7: 0.0</p>	 <p>Redacted: Confirmed CEM</p>	<p>Rank: 18 Skin Tone: 12.55% Porn: 0.99 Child: 1.0 CAT1: 0.13 CAT2: 0.74 CAT3: 0.0 CAT4: 0.09 CAT5: 0.0 CAT7: 0.02</p>

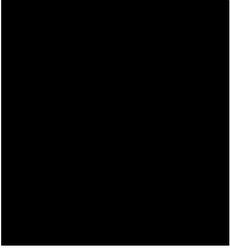
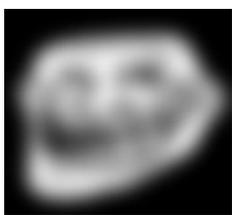
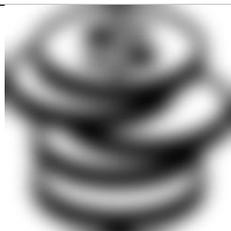
TorCrawl: Top 20 Images			
Image	Details	Image	Details
 <p>Redacted: Confirmed CEM</p>	<p>Rank: 19 Skin Tone: 41.08% Porn: 0.99 Child: 1.0 CAT1 0.01 CAT2 0.91 CAT3 0.0 CAT4 0.05 CAT5 0.0 CAT7 0.0</p>		<p>Rank: 20 Skin Tone: 90.02% Porn: 0.99 Child: 1.0 CAT1 0.0 CAT2 0.99 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.0</p>

Table B.33: TorCrawl top 20 results by classifier (porn,child)

TorCrawl: Bottom 20 Images			
Image	Details	Image	Details
	Rank: 1 Skin Tone: 1.77% Porn: 0.0 Child: 0.0 CAT1: 0.99 CAT2: 0.0 CAT3: 0.0 CAT4: 0.0 CAT5: 0.0 CAT7: 0.0		Rank: 2 Skin Tone: 5.86% Porn: 0.0 Child: 0.0 CAT1: 0.99 CAT2: 0.0 CAT3: 0.0 CAT4: 0.0 CAT5: 0.0 CAT7: 0.0
	Rank: 3 Skin Tone: 0.18% Porn: 0.0 Child: 0.0 CAT1: 0.64 CAT2: 0.0 CAT3: 0.03 CAT4: 0.0 CAT5: 0.0 CAT7: 0.32		Rank: 4 Skin Tone: 8.42% Porn: 0.0 Child: 0.0 CAT1: 0.99 CAT2: 0.0 CAT3: 0.0 CAT4: 0.0 CAT5: 0.0 CAT7: 0.0
	Rank: 5 Skin Tone: 0.00% Porn: 0.0 Child: 0.0 CAT1: 0.01 CAT2: 0.0 CAT3: 0.0 CAT4: 0.0 CAT5: 0.0 CAT7: 0.98		Rank: 6 Skin Tone: 36.63% Porn: 0.0 Child: 0.0 CAT1: 0.02 CAT2: 0.0 CAT3: 0.89 CAT4: 0.04 CAT5: 0.0 CAT7: 0.03
	Rank: 7 Skin Tone: 100.00% Porn: 0.0 Child: 0.0 CAT1: 0.57 CAT2: 0.0 CAT3: 0.02 CAT4: 0.22 CAT5: 0.0 CAT7: 0.17		Rank: 8 Skin Tone: 0.00% Porn: 0.0 Child: 0.0 CAT1: 0.99 CAT2: 0.0 CAT3: 0.0 CAT4: 0.0 CAT5: 0.0 CAT7: 0.0

TorCrawl: Bottom 20 Images			
Image	Details	Image	Details
	Rank: 9 Skin Tone: 0.00% Porn: 0.0 Child: 0.0 CAT1 0.52 CAT2 0.16 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.3		Rank: 10 Skin Tone: 3.00% Porn: 0.0 Child: 0.0 CAT1 0.1 CAT2 0.0 CAT3 0.0 CAT4 0.84 CAT5 0.0 CAT7 0.04
	Rank: 11 Skin Tone: 0.00% Porn: 0.0 Child: 0.0 CAT1 0.64 CAT2 0.0 CAT3 0.28 CAT4 0.05 CAT5 0.0 CAT7 0.01		Rank: 12 Skin Tone: 0.00% Porn: 0.0 Child: 0.0 CAT1 0.0 CAT2 0.0 CAT3 0.0 CAT4 0.94 CAT5 0.0 CAT7 0.04
	Rank: 13 Skin Tone: 0.00% Porn: 0.0 Child: 0.0 CAT1 0.21 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.77		Rank: 14 Skin Tone: 0.00% Porn: 0.0 Child: 0.0 CAT1 0.99 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.0
	Rank: 15 Skin Tone: 0.00% Porn: 0.0 Child: 0.0 CAT1 0.0 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.99		Rank: 16 Skin Tone: 0.00% Porn: 0.0 Child: 0.0 CAT1 0.86 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.13

TorCrawl: Bottom 20 Images			
Image	Details	Image	Details
	Rank: 17 Skin Tone: 25.33% Porn: 0.0 Child: 0.0 CAT1 0.01 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.98		Rank: 18 Skin Tone: 0.00% Porn: 0.0 Child: 0.0 CAT1 0.32 CAT2 0.0 CAT3 0.0 CAT4 0.59 CAT5 0.0 CAT7 0.08
	Rank: 19 Skin Tone: 0.06% Porn: 0.0 Child: 0.0 CAT1 0.7 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.29		Rank: 20 Skin Tone: 2.38% Porn: 0.0 Child: 0.0 CAT1 0.91 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.06

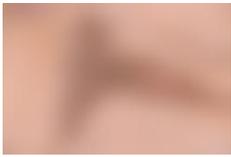
Table B.34: TorCrawl bottom 20 results by classifier (porn,child)

B.18 Adult Pornography corpus Skin Tone Results

WARNING
CONTAINS/MAY CONTAIN SEXUALLY EXPLICIT
IMAGERY

TEST_CAT8: Top 20 Images			
Image	Details	Image	Details
	Rank: 1 Skin Tone: 99.96% Porn: 0.92 Child: 0.0 CAT1 0.07 CAT2 0.0 CAT3 0.67 CAT4 0.0 CAT5 0.0 CAT7 0.24		Rank: 2 Skin Tone: 99.96% Porn: 0.78 Child: 0.0 CAT1 0.63 CAT2 0.06 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.28
	Rank: 3 Skin Tone: 99.94% Porn: 0.72 Child: 0.0 CAT1 0.94 CAT2 0.02 CAT3 0.03 CAT4 0.0 CAT5 0.0 CAT7 0.0		Rank: 4 Skin Tone: 99.94% Porn: 0.99 Child: 0.03 CAT1 0.3 CAT2 0.0 CAT3 0.69 CAT4 0.0 CAT5 0.0 CAT7 0.0
	Rank: 5 Skin Tone: 99.94% Porn: 0.99 Child: 0.0 CAT1 0.99 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.0		Rank: 6 Skin Tone: 99.94% Porn: 0.93 Child: 0.01 CAT1 0.97 CAT2 0.0 CAT3 0.01 CAT4 0.01 CAT5 0.0 CAT7 0.0

TEST_CAT8: Top 20 Images			
Image	Details	Image	Details
	Rank: 7 Skin Tone: 99.94% Porn: 0.99 Child: 0.93 CAT1 0.03 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.95		Rank: 8 Skin Tone: 99.93% Porn: 0.99 Child: 0.09 CAT1 0.99 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.0
	Rank: 9 Skin Tone: 99.93% Porn: 0.98 Child: 0.0 CAT1 0.86 CAT2 0.0 CAT3 0.13 CAT4 0.0 CAT5 0.0 CAT7 0.0		Rank: 10 Skin Tone: 99.93% Porn: 0.99 Child: 0.0 CAT1 0.36 CAT2 0.0 CAT3 0.0 CAT4 0.58 CAT5 0.0 CAT7 0.04
	Rank: 11 Skin Tone: 99.93% Porn: 0.99 Child: 0.0 CAT1 0.93 CAT2 0.0 CAT3 0.06 CAT4 0.0 CAT5 0.0 CAT7 0.0		Rank: 12 Skin Tone: 99.92% Porn: 0.99 Child: 0.0 CAT1 0.95 CAT2 0.0 CAT3 0.04 CAT4 0.0 CAT5 0.0 CAT7 0.0

TEST_CAT8: Top 20 Images			
Image	Details	Image	Details
	Rank: 13 Skin Tone: 99.92% Porn: 0.99 Child: 0.0 CAT1 0.0 CAT2 0.99 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.0		Rank: 14 Skin Tone: 99.92% Porn: 0.99 Child: 0.0 CAT1 0.38 CAT2 0.6 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.0
	Rank: 15 Skin Tone: 99.91% Porn: 0.92 Child: 0.0 CAT1 0.78 CAT2 0.02 CAT3 0.02 CAT4 0.03 CAT5 0.0 CAT7 0.13		Rank: 16 Skin Tone: 99.91% Porn: 0.97 Child: 0.0 CAT1 0.7 CAT2 0.17 CAT3 0.02 CAT4 0.0 CAT5 0.0 CAT7 0.07
	Rank: 17 Skin Tone: 99.91% Porn: 0.98 Child: 0.0 CAT1 0.92 CAT2 0.0 CAT3 0.05 CAT4 0.0 CAT5 0.0 CAT7 0.0		Rank: 18 Skin Tone: 99.91% Porn: 0.99 Child: 0.0 CAT1 0.99 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.0

TEST_CAT8: Top 20 Images			
Image	Details	Image	Details
	Rank: 19 Skin Tone: 99.91% Porn: 0.94 Child: 0.0 CAT1 0.02 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.95		Rank: 20 Skin Tone: 99.91% Porn: 0.97 Child: 0.0 CAT1 0.03 CAT2 0.0 CAT3 0.96 CAT4 0.0 CAT5 0.0 CAT7 0.0

Table B.35: Test corpus CAT8 top 20 results by skin tone percentage

TEST_CAT8: Bottom 20 Images			
Image	Details	Image	Details
File Missing	Rank: 1 Skin Tone: 0.00% Porn: 0.01 Child: 0.99 CAT1 0.4 CAT2 0.0 CAT3 0.27 CAT4 0.16 CAT5 0.01 CAT7 0.13		Rank: 2 Skin Tone: 0.00% Porn: 0.99 Child: 0.99 CAT1 0.97 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.02
	Rank: 3 Skin Tone: 0.00% Porn: 0.34 Child: 0.99 CAT1 0.1 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.89		Rank: 4 Skin Tone: 0.00% Porn: 0.19 Child: 0.03 CAT1 0.05 CAT2 0.94 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.0
File Missing	Rank: 5 Skin Tone: 0.00% Porn: 0.01 Child: 0.99 CAT1 0.4 CAT2 0.0 CAT3 0.27 CAT4 0.16 CAT5 0.01 CAT7 0.13		Rank: 6 Skin Tone: 0.00% Porn: 0.97 Child: 0.0 CAT1 0.0 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.98

TEST_CAT8: Bottom 20 Images			
Image	Details	Image	Details
	Rank: 7 Skin Tone: 0.00% Porn: 0.08 Child: 0.0 CAT1 0.0 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.99		Rank: 8 Skin Tone: 0.00% Porn: 0.81 Child: 0.99 CAT1 0.06 CAT2 0.93 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.0
	Rank: 9 Skin Tone: 0.00% Porn: 0.7 Child: 0.99 CAT1 0.0 CAT2 0.02 CAT3 0.0 CAT4 0.96 CAT5 0.0 CAT7 0.0		Rank: 10 Skin Tone: 0.00% Porn: 0.9 Child: 0.11 CAT1 0.91 CAT2 0.06 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.01
	Rank: 11 Skin Tone: 0.00% Porn: 0.15 Child: 0.0 CAT1 0.0 CAT2 0.75 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.24		Rank: 12 Skin Tone: 0.00% Porn: 0.58 Child: 0.0 CAT1 0.03 CAT2 0.86 CAT3 0.0 CAT4 0.08 CAT5 0.0 CAT7 0.0

TEST_CAT8: Bottom 20 Images			
Image	Details	Image	Details
	Rank: 13 Skin Tone: 0.00% Porn: 0.96 Child: 0.0 CAT1 0.0 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.99		Rank: 14 Skin Tone: 0.00% Porn: 0.18 Child: 0.99 CAT1 0.0 CAT2 0.0 CAT3 0.02 CAT4 0.93 CAT5 0.0 CAT7 0.02
	Rank: 15 Skin Tone: 0.00% Porn: 0.95 Child: 0.34 CAT1 0.12 CAT2 0.74 CAT3 0.03 CAT4 0.02 CAT5 0.0 CAT7 0.07		Rank: 16 Skin Tone: 0.00% Porn: 0.37 Child: 0.54 CAT1 0.02 CAT2 0.0 CAT3 0.01 CAT4 0.42 CAT5 0.0 CAT7 0.52
	Rank: 17 Skin Tone: 0.00% Porn: 0.96 Child: 0.37 CAT1 0.17 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.82		Rank: 18 Skin Tone: 0.00% Porn: 0.27 Child: 0.05 CAT1 0.73 CAT2 0.12 CAT3 0.0 CAT4 0.01 CAT5 0.0 CAT7 0.13

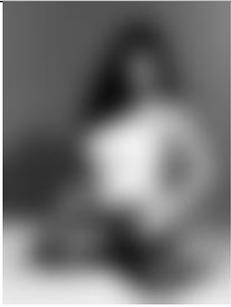
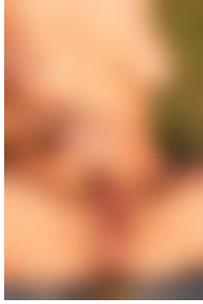
TEST_CAT8: Bottom 20 Images			
Image	Details	Image	Details
	Rank: 19 Skin Tone: 0.00% Porn: 0.06 Child: 0.0 CAT1 0.0 CAT2 0.0 CAT3 0.26 CAT4 0.73 CAT5 0.0 CAT7 0.0		Rank: 20 Skin Tone: 0.00% Porn: 0.13 Child: 0.0 CAT1 0.27 CAT2 0.54 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.17

Table B.36: Test corpus CAT8 bottom 20 results by skin tone percentage

B.19 Adult Pornography Corpus Classifier Results

WARNING
CONTAINS/MAY CONTAIN SEXUALLY EXPLICIT
IMAGERY

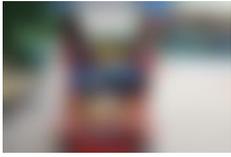
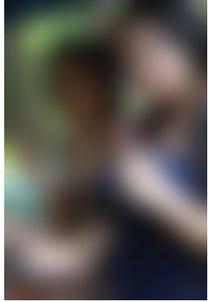
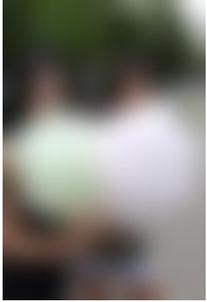
TEST_CAT8: Top 20 Images			
Image	Details	Image	Details
	Rank: 1 Skin Tone: 84.48% Porn: 1.0 Child: 0.99 CAT1 0.0 CAT2 0.0 CAT3 0.05 CAT4 0.94 CAT5 0.0 CAT7 0.0		Rank: 2 Skin Tone: 86.24% Porn: 1.0 Child: 0.99 CAT1 0.0 CAT2 0.0 CAT3 0.29 CAT4 0.7 CAT5 0.0 CAT7 0.0
	Rank: 3 Skin Tone: 92.48% Porn: 1.0 Child: 0.65 CAT1 0.0 CAT2 0.0 CAT3 0.1 CAT4 0.89 CAT5 0.0 CAT7 0.0		Rank: 4 Skin Tone: 88.69% Porn: 1.0 Child: 0.06 CAT1 0.0 CAT2 0.0 CAT3 0.38 CAT4 0.61 CAT5 0.0 CAT7 0.0
	Rank: 5 Skin Tone: 83.88% Porn: 1.0 Child: 0.0 CAT1 0.05 CAT2 0.0 CAT3 0.08 CAT4 0.0 CAT5 0.0 CAT7 0.85		Rank: 6 Skin Tone: 78.98% Porn: 1.0 Child: 0.0 CAT1 0.02 CAT2 0.34 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.63

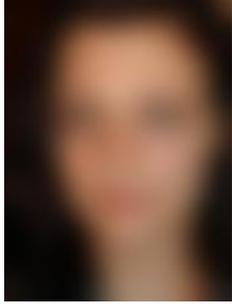
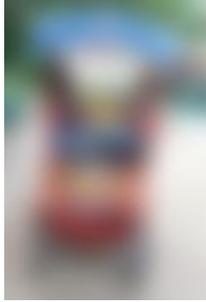
TEST_CAT8: Top 20 Images			
Image	Details	Image	Details
	Rank: 7 Skin Tone: 93.00% Porn: 1.0 Child: 0.0 CAT1 0.0 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.99		Rank: 8 Skin Tone: 79.08% Porn: 1.0 Child: 0.0 CAT1 0.13 CAT2 0.02 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.84
	Rank: 9 Skin Tone: 86.51% Porn: 1.0 Child: 0.0 CAT1 0.0 CAT2 0.0 CAT3 0.16 CAT4 0.83 CAT5 0.0 CAT7 0.0		Rank: 10 Skin Tone: 95.79% Porn: 1.0 Child: 0.0 CAT1 0.0 CAT2 0.99 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.0
	Rank: 11 Skin Tone: 97.44% Porn: 1.0 Child: 0.0 CAT1 0.08 CAT2 0.89 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.01		Rank: 12 Skin Tone: 89.47% Porn: 1.0 Child: 0.0 CAT1 0.0 CAT2 0.99 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.0

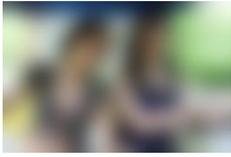
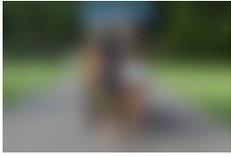
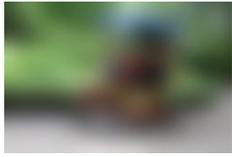
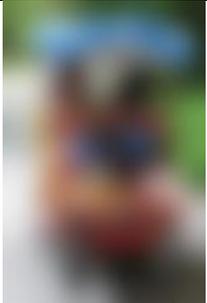
TEST_CAT8: Top 20 Images			
Image	Details	Image	Details
	Rank: 13 Skin Tone: 99.20% Porn: 1.0 Child: 0.0 CAT1 0.26 CAT2 0.57 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.15		Rank: 14 Skin Tone: 91.45% Porn: 1.0 Child: 0.0 CAT1 0.0 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.99
	Rank: 15 Skin Tone: 76.27% Porn: 1.0 Child: 0.0 CAT1 0.69 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.3		Rank: 16 Skin Tone: 95.02% Porn: 1.0 Child: 0.0 CAT1 0.0 CAT2 0.99 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.0
	Rank: 17 Skin Tone: 95.95% Porn: 1.0 Child: 0.0 CAT1 0.03 CAT2 0.0 CAT3 0.13 CAT4 0.82 CAT5 0.0 CAT7 0.0		Rank: 18 Skin Tone: 97.63% Porn: 1.0 Child: 0.0 CAT1 0.0 CAT2 0.0 CAT3 0.01 CAT4 0.98 CAT5 0.0 CAT7 0.0

TEST_CAT8: Top 20 Images					
Image	Details		Image	Details	
	Rank:	19		Rank:	20
	Skin Tone:	82.06%		Skin Tone:	98.31%
	Porn:	1.0		Porn:	1.0
	Child:	0.0		Child:	0.0
	CAT1	0.0		CAT1	0.0
	CAT2	0.0		CAT2	0.99
	CAT3	0.0		CAT3	0.0
	CAT4	0.0		CAT4	0.0
	CAT5	0.0		CAT5	0.0
	CAT7	0.99		CAT7	0.0

Table B.37: Test corpus CAT8 top 20 results by classifier (porn,child)

TEST_CAT8: Bottom 20 Images			
Image	Details	Image	Details
	Rank: 1 Skin Tone: 30.56% Porn: 0.0 Child: 0.0 CAT1 0.03 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.95		Rank: 2 Skin Tone: 9.22% Porn: 0.0 Child: 0.0 CAT1 0.85 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.14
	Rank: 3 Skin Tone: 21.43% Porn: 0.0 Child: 0.0 CAT1 0.0 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.99		Rank: 4 Skin Tone: 9.42% Porn: 0.0 Child: 0.0 CAT1 0.78 CAT2 0.18 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.03
	Rank: 5 Skin Tone: 17.73% Porn: 0.0 Child: 0.0 CAT1 0.0 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.99		Rank: 6 Skin Tone: 21.00% Porn: 0.0 Child: 0.0 CAT1 0.0 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.99

TEST_CAT8: Bottom 20 Images			
Image	Details	Image	Details
	Rank: 7 Skin Tone: 50.71% Porn: 0.0 Child: 0.0 CAT1 0.8 CAT2 0.0 CAT3 0.03 CAT4 0.0 CAT5 0.08 CAT7 0.07		Rank: 8 Skin Tone: 26.42% Porn: 0.0 Child: 0.0 CAT1 0.0 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.99
	Rank: 9 Skin Tone: 10.39% Porn: 0.0 Child: 0.0 CAT1 0.87 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.12		Rank: 10 Skin Tone: 10.78% Porn: 0.0 Child: 0.0 CAT1 0.05 CAT2 0.12 CAT3 0.0 CAT4 0.0 CAT5 0.01 CAT7 0.79
	Rank: 11 Skin Tone: 2.83% Porn: 0.0 Child: 0.0 CAT1 0.02 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.97		Rank: 12 Skin Tone: 80.86% Porn: 0.0 Child: 0.0 CAT1 0.0 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.99

TEST_CAT8: Bottom 20 Images			
Image	Details	Image	Details
	Rank: 13 Skin Tone: 3.96% Porn: 0.0 Child: 0.0 CAT1: 0.43 CAT2: 0.0 CAT3: 0.0 CAT4: 0.0 CAT5: 0.0 CAT7: 0.56		Rank: 14 Skin Tone: 22.56% Porn: 0.0 Child: 0.0 CAT1: 0.0 CAT2: 0.0 CAT3: 0.0 CAT4: 0.0 CAT5: 0.0 CAT7: 0.99
	Rank: 15 Skin Tone: 29.12% Porn: 0.0 Child: 0.0 CAT1: 0.0 CAT2: 0.0 CAT3: 0.0 CAT4: 0.0 CAT5: 0.0 CAT7: 0.99		Rank: 16 Skin Tone: 5.83% Porn: 0.0 Child: 0.0 CAT1: 0.1 CAT2: 0.0 CAT3: 0.0 CAT4: 0.0 CAT5: 0.0 CAT7: 0.89
	Rank: 17 Skin Tone: 2.92% Porn: 0.0 Child: 0.0 CAT1: 0.98 CAT2: 0.0 CAT3: 0.0 CAT4: 0.0 CAT5: 0.0 CAT7: 0.01		Rank: 18 Skin Tone: 9.50% Porn: 0.0 Child: 0.0 CAT1: 0.57 CAT2: 0.0 CAT3: 0.0 CAT4: 0.0 CAT5: 0.0 CAT7: 0.42

TEST_CAT8: Bottom 20 Images			
Image	Details	Image	Details
	Rank: 19 Skin Tone: 11.31% Porn: 0.0 Child: 0.0 CAT1 0.0 CAT2 0.18 CAT3 0.0 CAT4 0.02 CAT5 0.0 CAT7 0.78		Rank: 20 Skin Tone: 31.42% Porn: 0.0 Child: 0.0 CAT1 0.01 CAT2 0.0 CAT3 0.0 CAT4 0.0 CAT5 0.0 CAT7 0.98

Table B.38: Test corpus CAT8 bottom 20 results by classifier (porn,child)

Appendix C

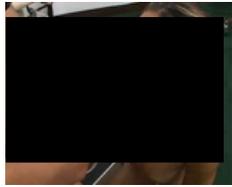
Annotation Schema Test v1

Results

WARNING
CONTAINS/MAY CONTAIN SEXUALLY EXPLICIT
IMAGERY

Image

1



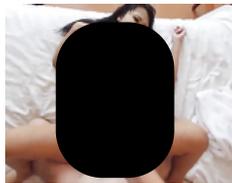
Male 1	Male 2	Male 3	Female 1	Female 2	Female 3
Pornographic	Pornographic	Pornographic	Pornographic	Pornographic	Pornographic
Nudity	Nudity	Nudity	Nudity	Nudity	Nudity
Female/Female	Female	Female/female	Female/female	Female/Female	Female/feamle
	White				Black
		Asian			
Unkown race			Unknown race	Unkown race	
Oral penetration		Oral penetration		Oral penetration	Oral penetration
Sex toy	Sex Toy: Other	Sex toy	Sex toy	Sex toy	Sex toy
	Domination				

2



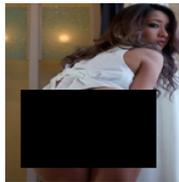
Pornographic	Pornographic	Pornographic	Pornographic	Pornographic	Pornographic
Nudity	Nudity	Nudity	Nudity	Nudity	Nudity
Female	Female	Female	Female	Female	Female
Unkown race	Unknown race	Unknown race	Unknown race	Unknown race	Unknown race
Vaginal penetration	Vaginal penetration	Vaginal penetration	Vaginal penetration		Vaginal penetration
				Anal penetration	
Sex toy: Other	Sex toy: Other	Sex toy: Other	Sex toy: Other	Sex toy: Other	Sex toy: Other
			Sadism/masochism		

3



Pornographic	Pornographic	Pornographic	Pornographic	Pornographic	Pornographic
Nudity	Nudity	Nudity	Nudity	Nudity	Nudity
Male/Female	Male/Female	Male/female	Male/female	Male/Female	Male/Female
Asian	Asian	Asian	Asian	Asian	Asian
Vaginal penetration	Vaginal penetration	Vaginal penetration		Vaginal penetration	Vaginal penetration

4



Pornographic		Pornographic	Not pornographic	Pornographic	Pornographic
Nudity	No nudity	No nudity	No nudity	Nudity	Nudity
Female	Female	Female	Female	Female	Female
Asian	Asian	Asian	Asian	Asian	Asian

5



Not pornographic					
No nudity					
	Male	Male	Male	Male	

Image



	Male 1	Male 2	Male 3	Female 1	Female 2	Female 3
			Male/male			
		White	White		White	
	Unknown race	Unknown race		Unknown race	Unknown race	

6



Pornographic Nudity					
	Female				
	Male				
Male/Female	Male/Female	Male/female	Male/female	Male/female	Male/female
Male/male	Male/male	Male/male	Male/male	Male/male	Male/male
White	White	White	White	White	White
	Unknown race				
Oral penetration					

7



Not pornographic No nudity					
			Male	Male	
	Animal/male	Animal/male			
	White	White	White	White	White
Unknown race	Unknown race				

8



Pornographic Nudity					
Male/male	Male/male	Male/male	Male/male	Male/male	Male/male
White	White	White	White	White	White
Anal penetration					

9



Pornographic Nudity					
	Female				
	Male				
Male/female	Male/female	Male/female	Male/female	Male/female	Male/female
	Female/female	Female/female	Female/female	Female/female	Female/female
White	White	White	White	White	White
Black	Black	Black	Black	Black	Black
Other penetration	Oral penetration	Oral penetration	Oral penetration	Oral penetration	Oral penetration

Image

10



	Male 1	Male 2	Male 3	Female 1	Female 2	Female 3
Pornographic Nudity	Pornographic Nudity	Pornographic Nudity	Pornographic Nudity	Pornographic Nudity	Pornographic Nudity	Pornographic Nudity
		Female	Female	Female		Female
		Male				
Male/female White	Male/female White	Male/female White	Male/female White	Male/female White	Male/female White	Male/female White
		Domination Sadism/masochism				
Excretion	Excretion	Excretion	Excretion	Excretion	Excretion	Excretion

11



Pornographic Nudity	Pornographic Nudity	Not pornographic No nudity	Not pornographic Nudity	Not pornographic No nudity	Not pornographic No nudity
Male	Male	Male	Male	Male	
Asian	Asian	Asian	Asian	Asian	

12



Pornographic Nudity	Pornographic Nudity	Pornographic	Pornographic Nudity	Pornographic Nudity	Pornographic Nudity
	Female				
	Male				
Male/female White	Male/female White	Male/female White	Male/female White	Male/female White	Male/female White
	Unknown race				
Oral penetration	Oral penetration	Oral penetration	Oral penetration	Oral penetration	Oral penetration
Anal penetration	Anal penetration	Anal penetration	Anal penetration	Anal penetration	Anal penetration
Sex toy	Sex toy	Sex toy	Sex toy	Sex toy	Sex toy
	Domination				

13



Not pornographic No nudity	Pornographic	Not pornographic No nudity	Not pornographic No nudity	Pornographic	Not pornographic No nudity
	Female				
Female/female Black	Female/female Black	Female/female Black	Female/female Black	Female/female Black	Female/female Black
	Unknown race				



Not pornographic No nudity	Not pornographic	Not pornographic No nudity	Not pornographic No nudity	Pornographic	Pornographic No nudity
----------------------------	------------------	----------------------------	----------------------------	--------------	------------------------

Image

14



Male 1	Male 2	Male 3	Female 1	Female 2	Female 3
	Female Male				
Male/female Asian	Male/female Asian	Male/female Asian		Unknown race Unknown race	Male/female Asian

15



Not pornographic No nudity	Not pornographic Nudity	Not pornographic Nudity	Not pornographic No nudity	Not pornographic	Not pornographic Nudity
Female White	Female White Unknown race	Female White	Female White	Female White	Female White
					Oral penetration

16



Pornographic Nudity	Pornographic Nudity	Pornographic Nudity	Pornographic Nudity	Pornographic Nudity	Pornographic Nudity
	Female Male				
Male/female White Oral penetration					
		Other penetration Bondage		Bondage	
Sadism/Masochism					Domination Sadism/masochism

17



Not pornographic No nudity	Not pornographic	Not pornographic No nudity	Not pornographic No nudity	Pornographic Nudity	Pornographic No nudity
		Animal/female Animal/male			
	Unknown race	Unknown race			White
Bondage Domination	Bondage	Bondage	Bondage	Bondage Domination Sadism/masochism	Bondage Domination Sadism/masochism
Virtual	Virtual	Virtual	Virtual	Virtual	Virtual

18



Not pornographic No nudity	Not pornographic	Not pornographic No nudity	Not pornographic No nudity	Not pornographic	Not pornographic No nudity
	Animal/female	Animal/female			Female

Image	Male 1	Male 2	Male 3	Female 1	Female 2	Female 3
18		Unknown race	Unknown race			White
19	Not pornographic No nudity	Not pornographic Female Male	Not pornographic No nudity	Not pornographic No nudity	Not pornographic No nudity	Not pornographic No nudity
	Male/female White	Male/female White	Male/female White	Unknown race	White	White
20	Pornographic Nudity	Pornographic Nudity Female Male	Pornographic Nudity	Pornographic Nudity	Pornographic Nudity	Pornographic Nudity
	Male/female Male/male	Male/female Male/male	Male/female Male/male	Male/female White Black	Male/female White Black	Male/female White Black
	Unknown race		Black Unknown race Oral penetration			
	Vaginal penetration Anal penetration	Vaginal penetration Anal penetration Domination	Vaginal penetration	Vaginal penetration Anal penetration	Vaginal penetration Anal penetration	Vaginal penetration Anal penetration
21	Not pornographic No nudity Female	Not pornographic Female	Not pornographic Female	Not pornographic No nudity Female White	Not pornographic No nudity Female	Pornographic No nudity Female
	Unknown race	Unknown race	Unknown race		Unknown race	Unknown race
22	Not pornographic No nudity	Not pornographic	Not pornographic No nudity	Not pornographic No nudity	Not pornographic	Not pornographic No nudity

Image

	Male 1	Male 2	Male 3	Female 1	Female 2	Female 3
23	Not pornographic	Not pornographic	Not pornographic	Not pornographic	Pornographic	Pornographic
	No nudity	No nudity	No nudity	No nudity	No nudity	No nudity
	Male/female	Male/female	Male/female	Male/female	Male/female	Male/female
	White	White	White	White	White	White
	Unknown race	Unknown race	Unknown race		Unknown race	



24	Not pornographic					
	Nudity	Nudity	No nudity	Nudity	Nudity	Nudity
	Female	Female	Female	Female	Female	Female
	White	White	White	White	White	White
	Unknown race	Unknown race	Unknown race	Unknown race		



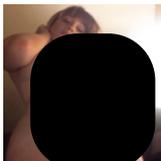
25	Pornographic	Pornographic	Not pornographic	Not pornographic	Pornographic	Pornographic
	Nudity	Nudity	Nudity	No nudity	No nudity	Nudity
	Female	Female	Female	Female	Female	Female
	Black	Black	Black	Black	Black	Black



26	Pornographic	Pornographic	Pornographic	Pornographic	Pornographic	Pornographic
	No nudity					
	Male/female	Male/female	Male/female	Male/female	Male/female	Male/female
	White	White	White	White	White	White
	Excretion	Excretion	Excretion	Excretion	Excretion	Excretion



27	Pornographic	Pornographic	Not pornographic	Not pornographic	Pornographic	Pornographic
	Nudity	Nudity	Nudity	Nudity	Nudity	Nudity
	Transgender	Transgender	Transgender	Transgender	Transgender	Transgender
	White	White	White	White	White	White
	Unknown race					



Image



Male 1

Male 2

Male 3

Female 1

Female 2

Female 3

28



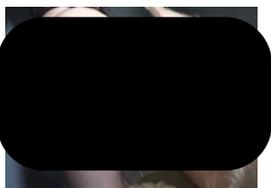
Pornographic	Pornographic	Not pornographic	Not pornographic	Pornographic	Pornographic
Nudity		Nudity	Nudity	Nudity	Nudity
Female	Female	Female	Female	Female	Female
Black	Black	Black	Black	Black	Black

29



Pornographic	Pornographic	Not pornographic	Pornographic	Pornographic	Pornographic
Nudity		Nudity	Nudity	Nudity	Nudity
Male/female	Male/female	Male/female		Male/female	Male/female
White	White	White	White	White	White
					Unknown race
Bondage		Bondage	Bondage	Bondage	Bondage
Domination	Domination	Domination	Domination	Domination	Domination
	Sadism/Masochism	Sadism/masochism			

30



Not pornographic	Pornographic	Pornographic	Pornographic	Pornographic	Pornographic
Nudity	Nudity	Nudity	No nudity	Nudity	Nudity
	Female				
	Animal/female	Animal/female	Animal/female	Animal/female	Animal/female
			White	White	
Unknown race	Unknown race	Unknown race			Unknown race
	Sadism/masochism				

31



Not pornographic		Not pornographic	Not pornographic	Not pornographic	Not pornographic
Nudity	Nudity	Nudity	Nudity	Nudity	Nudity
	Female	Female			
	Male				
				White	White
Unknown race	Unknown race	Unknown race	Unknown race	Unknown race	Unknown race



Pornographic	Pornographic	Pornographic	Pornographic	Pornographic	Pornographic
Nudity	Nudity	Nudity	Nudity	Nudity	Nudity

Image

32



Male 1	Male 2	Male 3	Female 1	Female 2	Female 3
Male/female	Male/female	Male/female	Male/female	Male/female	Male/female
	White	White	White	White	White
Unknown race					Unknown race
Vaginal penetration		Oral penetration	Oral penetration	Oral penetration	
Anal penetration	Anal penetration	Vaginal penetration	Vaginal penetration	Vaginal penetration	Vaginal penetration
Sex toy	Sex toy	Anal penetration	Anal penetration	Anal penetration	Anal penetration
		Sex toy	Sex toy	Sex toy	Sex toy

33



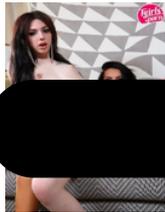
Pornographic	Pornographic	Pornographic	Pornographic	Pornographic	Pornographic
Nudity		Nudity	Nudity	Nudity	Nudity
Male	Male	Male	Male	Male	Male
	White			White	
Unknown race		Unknown race	Unknown race		Black

34



Pornographic	Pornographic	Pornographic	Pornographic	Pornographic	Pornographic
Nudity		Nudity	Nudity	Nudity	Nudity
Male	Male	Male	Male	Male	Male
Black	Black	Black	Black	Black	Black

35



Pornographic	Pornographic	Pornographic	Pornographic	Pornographic	Pornographic
Nudity	Nudity	Nudity	Nudity	Nudity	Nudity
	Male/female				
Transgender	Transgender	Transgender	Transgender	Transgender	Transgender
White	White	White	White	White	White
Anal penetration	Anal penetration	Oral penetration	Anal penetration	Anal penetration	Anal penetration

36



Pornographic	Pornographic	Pornographic	Pornographic	Pornographic	Pornographic
Nudity	Nudity	Nudity	Nudity	Nudity	Nudity
			Female		
Animal/female	Animal/female	Animal/female	Animal/female	Animal/female	Animal/female
White	White	White	White	White	White
Anal penetration	Anal penetration		Anal penetration	Anal penetration	Anal penetration
	Sadism/masochism				

Image

Male 1

Male 2

Male 3

Female 1

Female 2

Female 3

37



Not pornographic					
No nudity					
Male	Male	Male	Male	Male	Male
White	White	White	White	White	White

38



Not pornographic					
No nudity					
Female	Female	Female	Female	Female	Female
White	White	White	White	White	White

39



Pornographic	Pornographic	Pornographic	Pornographic	Pornographic	Pornographic
Nudity	Nudity	Nudity	Nudity	Nudity	Nudity
Female	Female	Female	Female	Female	Female
White	White	White	White	White	White
Bondage	Bondage	Bondage	Bondage	Bondage	Bondage
				Domination	Domination
		Sadism/masochism	Sadism/masochism	Sadism/masochism	Sadism/masochism

40



Not pornographic					
Nudity	Nudity	Nudity	Nudity	Nudity	Nudity
Female	Female	Animal/female	Female	Female	Female
White	White	White	White	White	White

41



Not pornographic					
No nudity					
Female	Female	Female	Female	Female	Female
White	White	White		Unknown race	White

Image



Male 1

Male 2

Male 3

Female 1

Female 2

Female 3

42



Pornographic Nudity Female	Pornographic Nudity Female	Pornographic Nudity Male/female	Pornographic Nudity Female White	Pornographic Nudity White	Pornographic Nudity Male/female
Unknown race	Unknown race	Unknown race	White	Unknown race	Unknown race
Vaginal penetration	Anal penetration Sex toy	Vaginal penetration	Vaginal penetration Sex toy	Vaginal penetration	Vaginal penetration
Sex toy: Other	Sex toy: Other	Sex toy: Other	Sex toy: Other	Sex toy: Other	Sex toy: Other
Bondage		Bondage	Bondage	Bondage	Bondage
Sadism/Masochism		Sadism/masochism	Sadism/masochism	Sadism/masochism	Sadism/masochism

43



Not pornographic No nudity Female White	Not pornographic No nudity Female White	Not pornographic Nudity Female White	Not pornographic No nudity Female White	Not pornographic No nudity Female White	Not pornographic No nudity Female White
--	--	---	--	--	--

44

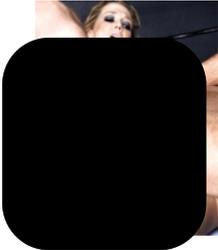


Pornographic Nudity Female	Pornographic Nudity Female	Pornographic Nudity Female	Pornographic Nudity Female White	Pornographic Nudity Female White	Pornographic Nudity Female
Unknown race	Unknown race Sex toy: Other	Unknown race			Unknown race
Bondage	Bondage Domination Sadism/masochism	Bondage Domination Sadism/masochism	Bondage Sadism/masochism	Bondage Domination Sadism/masochism	Bondage

45



Pornographic Nudity Female	Pornographic Nudity Female	Pornographic Nudity Female	Pornographic Nudity Female	Pornographic Nudity Female White	Pornographic Nudity Female
Unknown race	Unknown race	Unknown race	Unknown race		Unknown race
Anal penetration	Anal penetration	Anal penetration		Anal penetration	Anal penetration
Sex toy		Sex toy	Sex toy	Sex toy	Sex toy

Image	Male 1	Male 2	Male 3	Female 1	Female 2	Female 3	
46		Not pornographic No nudity Female White Virtual	Not pornographic No nudity Female	Not pornographic Nudity Female White Virtual	Not pornographic No nudity Female White	Not pornographic No nudity Female White	Not pornographic No nudity Female White
47		Pornographic Nudity Female White Unknown race	Pornographic Nudity Female White Bondage Domination Sadism/masochism	Pornographic Nudity Female White Bondage Domination Sadism/masochism	Pornographic Nudity Female White Bondage Domination Sadism/masochism	Pornographic Nudity Female White Bondage Domination Sadism/masochism	Pornographic Nudity Female White Bondage
48		Pornographic Nudity Female Male Male/female Male/male White Vaginal penetration Anal penetration Domination Sadism/masochism	Pornographic Nudity Male/female Male/male White Vaginal penetration Anal penetration	Pornographic Nudity Male/female Male/male White Vaginal penetration Anal penetration	Pornographic Nudity Male/female Male/female White Vaginal penetration Anal penetration	Pornographic Nudity Male/female Male/female White Vaginal penetration Anal penetration	Pornographic Nudity Male/female Male/female White Vaginal penetration Anal penetration
49		Pornographic Nudity Male/male Unknown race	Pornographic Nudity Male/male Unknown race	Pornographic Nudity Male/male Unknown race	Pornographic Nudity Male/male Unknown race	Pornographic Nudity Male/male Unknown race	

Appendix D

Chapter 5 Appendices

D.1 Training Corpus C_p Tuning Results

Table D.1: C_p tuning - MD5 Scorer, First FOI found

Scorer	C_p	Runs	Min	Max	Avg	Std Dev	Std Err
MD5 Scorer	0.01	2552	.04%	96.73%	24.48%	28.07%	2.46%
MD5 Scorer	0.1	1942	.02%	38.76%	8.88%	9.57%	.84%
MD5 Scorer	0.2	260	.03%	28.3%	7.63%	8.03%	.7%
MD5 Scorer	0.4	260	.12%	31.34%	8.55%	8.82%	.77%
MD5 Scorer	0.6	260	.04%	39.15%	8.78%	9.32%	.82%
MD5 Scorer	$\frac{1}{\sqrt{2}}$	260	.02%	45.88%	8.58%	9.53%	.84%
MD5 Scorer	0.8	260	.02%	42.16%	9.02%	9.77%	.86%
MD5 Scorer	1	260	.02%	49.32%	9.7%	10.81%	.95%
MD5 Scorer	1.2	260	.05%	39.21%	9.22%	10.3%	.9%
MD5 Scorer	1.4	260	.01%	43.07%	9.64%	10.75%	.94%
MD5 Scorer	1.6	260	.06%	43.56%	9.43%	10.88%	.95%
MD5 Scorer	1.8	260	.05%	51.9%	10.43%	11.92%	1.05%
MD5 Scorer	2	260	.01%	48.37%	10.06%	11.88%	1.04%
MD5 Scorer (no ignorable files)	0.01	2200	.04%	73.28%	12.13%	15.08%	1.32%
MD5 Scorer (no ignorable files)	0.1	2207	.07%	59.92%	12.07%	14.32%	1.26%
MD5 Scorer (no ignorable files)	0.2	260	.09%	70.24%	11.89%	15.4%	1.35%
MD5 Scorer (no ignorable files)	0.4	260	.03%	62.31%	12.37%	15.06%	1.32%
MD5 Scorer (no ignorable files)	0.6	260	.17%	72.77%	11.74%	15.49%	1.36%
MD5 Scorer (no ignorable files)	$\frac{1}{\sqrt{2}}$	260	.07%	58.24%	11.46%	14.18%	1.24%
MD5 Scorer (no ignorable files)	0.8	260	.08%	59.94%	11.17%	13.8%	1.21%
MD5 Scorer (no ignorable files)	1	260	.08%	58.19%	12.9%	16.1%	1.41%
MD5 Scorer (no ignorable files)	1.2	260	.04%	59.79%	11.71%	14.72%	1.29%
MD5 Scorer (no ignorable files)	1.4	260	.06%	62.31%	11.32%	15.27%	1.34%
MD5 Scorer (no ignorable files)	1.6	260	.03%	63.65%	11.12%	14.94%	1.31%
MD5 Scorer (no ignorable files)	1.8	260	.16%	63.8%	12.24%	15.04%	1.32%
MD5 Scorer (no ignorable files)	2	260	.05%	64.55%	12.49%	16.03%	1.41%

Table D.2: C_p tuning - MD5 Scorer, All FOI found

Scorer	C_p	Runs	Min	Max	Avg	Std Dev	Std Err
MD5 Scorer	0.01	2552	.64%	99.8%	51.27%	32.29%	2.83%
MD5 Scorer	0.1	1942	1.18%	98.51%	36.07%	31.96%	2.8%
MD5 Scorer	0.2	260	1.95%	98.51%	39.72%	33.57%	2.94%
MD5 Scorer	0.4	260	3.08%	99.1%	43.06%	34.2%	3.%
MD5 Scorer	0.6	260	3.58%	98.64%	44.66%	33.23%	2.91%
MD5 Scorer	$\frac{1}{\sqrt{2}}$	260	4.36%	99.1%	45.71%	33.6%	2.95%
MD5 Scorer	0.8	260	4.66%	98.92%	45.38%	32.7%	2.87%
MD5 Scorer	1	260	5.19%	98.96%	46.58%	32.9%	2.89%
MD5 Scorer	1.2	260	5.7%	98.58%	46.98%	32.76%	2.87%
MD5 Scorer	1.4	260	6.04%	99.07%	48.08%	32.78%	2.87%
MD5 Scorer	1.6	260	6.45%	98.8%	48.33%	32.47%	2.85%
MD5 Scorer	1.8	260	6.84%	99.26%	48.93%	32.67%	2.87%
MD5 Scorer	2	260	7.14%	99.22%	49.42%	32.65%	2.86%
MD5 Scorer (no ignorable files)	0.01	2200	.71%	98.51%	38.52%	30.26%	2.65%
MD5 Scorer (no ignorable files)	0.1	2207	1.96%	99.04%	46.15%	36.76%	3.22%
MD5 Scorer (no ignorable files)	0.2	260	2.9%	97.77%	48.88%	36.83%	3.23%
MD5 Scorer (no ignorable files)	0.4	260	6.27%	99.33%	50.75%	35.42%	3.11%
MD5 Scorer (no ignorable files)	0.6	260	8.58%	98.99%	52.29%	34.96%	3.07%
MD5 Scorer (no ignorable files)	$\frac{1}{\sqrt{2}}$	260	9.47%	98.89%	52.57%	34.27%	3.01%
MD5 Scorer (no ignorable files)	0.8	260	10.03%	99.28%	53.33%	34.69%	3.04%
MD5 Scorer (no ignorable files)	1	260	10.25%	99.07%	52.66%	33.4%	2.93%
MD5 Scorer (no ignorable files)	1.2	260	11.38%	99.34%	54.54%	34.21%	3.%
MD5 Scorer (no ignorable files)	1.4	260	11.77%	99.02%	54.66%	33.7%	2.96%
MD5 Scorer (no ignorable files)	1.6	260	11.87%	99.24%	54.3%	33.47%	2.94%
MD5 Scorer (no ignorable files)	1.8	260	12.%	98.9%	55.13%	33.55%	2.94%
MD5 Scorer (no ignorable files)	2	260	12.31%	99.05%	55.35%	33.19%	2.91%

D.2 Training Corpus MD5 Scorer Parameter Tuning Results

Table D.3: Training Corpus - MD5 Scorer (with ignorable file hashset) Parameter Tuning Results - First Hit ($C_p = 0.1$)

Default Score	TOI Score	C param	Runs	Min	Max	Avg	Std Dev	Std Err
0.01	0.01	0.2	260	.03%	48.61%	10.81%	13.18%	.82%
0.01	0.1	0.2	260	.04%	33.97%	7.45%	8.73%	.54%
0.01	0.2	0.2	260	.01%	53.44%	9.23%	13.82%	.86%
0.01	0.3	0.2	260	.01%	51.66%	8.38%	13.37%	.83%
0.01	0.4	0.2	304	.05%	50.01%	8.68%	13.17%	.82%
0.01	0.5	0.2	268	.04%	49.64%	9.09%	13.35%	.83%
0.01	0.6	0.2	264	.01%	49.32%	9.55%	13.65%	.85%
0.01	0.7	0.2	260	.02%	49.14%	9.33%	13.49%	.84%
0.01	0.8	0.2	267	.03%	57.34%	10.05%	14.55%	.9%
0.01	0.9	0.2	260	.%	56.87%	9.99%	14.53%	.9%
0.01	0.99	0.2	260	.04%	56.39%	9.44%	13.62%	.84%
0.1	0.01	0.2	272	.06%	47.64%	11.18%	12.48%	.77%
0.1	0.1	0.2	275	.06%	33.09%	8.95%	9.94%	.62%
0.1	0.2	0.2	260	.%	29.85%	7.14%	7.8%	.48%
0.1	0.3	0.2	260	.01%	38.1%	7.39%	9.31%	.58%
0.1	0.4	0.2	260	.01%	50.71%	8.52%	13.34%	.83%
0.1	0.5	0.2	260	.05%	50.06%	9.2%	13.57%	.84%
0.1	0.6	0.2	260	.02%	56.71%	9.19%	13.71%	.85%
0.1	0.7	0.2	260	.03%	59.84%	10.06%	15.%	.93%
0.1	0.8	0.2	260	.02%	58.18%	9.31%	13.98%	.87%
0.1	0.9	0.2	260	.05%	49.16%	9.48%	14.%	.87%
0.1	0.99	0.2	260	.01%	60.86%	9.5%	13.87%	.86%
0.2	0.01	0.2	260	.02%	50.11%	12.15%	13.6%	.84%
0.2	0.1	0.2	260	.%	52.81%	10.77%	12.08%	.75%
0.2	0.2	0.2	260	.02%	35.2%	8.17%	8.96%	.56%
0.2	0.3	0.2	260	.02%	31.42%	6.91%	7.29%	.45%
0.2	0.4	0.2	260	.04%	26.62%	6.54%	7.6%	.47%
0.2	0.5	0.2	260	.02%	41.83%	7.55%	10.2%	.63%
0.2	0.6	0.2	260	.05%	50.22%	8.89%	13.29%	.82%
0.2	0.7	0.2	260	.04%	58.8%	9.61%	14.37%	.89%
0.2	0.8	0.2	260	.04%	56.31%	8.86%	13.58%	.84%
0.2	0.9	0.2	260	.02%	60.01%	9.77%	14.39%	.89%
0.2	0.99	0.2	260	.04%	49.06%	8.37%	12.37%	.77%
0.3	0.01	0.2	260	.02%	88.48%	13.13%	14.99%	.93%

Training Corpus - MD5 Scorer (with ignorable file hashset) Parameter Tuning Results - First Hit (continued)

Default Score	TOI Score	C_p	Runs	Min	Max	Avg	Std Dev	Std Err
0.3	0.1	0.2	260	.02%	67.48%	12.6%	13.66%	.85%
0.3	0.2	0.2	260	.02%	49.13%	10.44%	11.24%	.7%
0.3	0.3	0.2	260	.03%	32.08%	7.59%	8.22%	.51%
0.3	0.4	0.2	260	.03%	26.87%	6.62%	6.73%	.42%
0.3	0.5	0.2	260	.02%	32.84%	6.2%	6.61%	.41%
0.3	0.6	0.2	260	.03%	28.36%	6.6%	7.83%	.49%
0.3	0.7	0.2	260	.04%	43.95%	7.54%	10.61%	.66%
0.3	0.8	0.2	260	.03%	67.11%	9.11%	14.53%	.9%
0.3	0.9	0.2	260	.02%	60.27%	9.19%	13.73%	.85%
0.3	0.99	0.2	260	.03%	55.77%	8.16%	12.27%	.76%
0.4	0.01	0.2	260	.01%	90.91%	16.62%	19.47%	1.21%
0.4	0.1	0.2	260	.08%	80.62%	15.23%	17.11%	1.06%
0.4	0.2	0.2	260	.09%	46.61%	11.9%	12.56%	.78%
0.4	0.3	0.2	260	.%	45.09%	10.42%	11.72%	.73%
0.4	0.4	0.2	260	.03%	36.49%	7.72%	8.28%	.51%
0.4	0.5	0.2	260	.01%	43.28%	6.92%	7.27%	.45%
0.4	0.6	0.2	260	.02%	32.84%	6.22%	6.68%	.41%
0.4	0.7	0.2	260	.03%	32.77%	6.42%	7.7%	.48%
0.4	0.8	0.2	260	.04%	64.63%	6.84%	9.15%	.57%
0.4	0.9	0.2	260	.01%	59.61%	7.73%	11.01%	.68%
0.4	0.99	0.2	260	.%	60.68%	8.58%	12.73%	.79%
0.5	0.01	0.2	260	.03%	90.44%	18.8%	20.73%	1.29%
0.5	0.1	0.2	260	.02%	87.75%	15.72%	18.69%	1.16%
0.5	0.2	0.2	260	.03%	69.96%	13.95%	14.92%	.93%
0.5	0.3	0.2	260	.11%	51.34%	11.33%	12.61%	.78%
0.5	0.4	0.2	260	.04%	38.16%	9.2%	10.26%	.64%
0.5	0.5	0.2	260	.03%	35.82%	7.6%	8.18%	.51%
0.5	0.6	0.2	260	.01%	28.8%	6.51%	6.63%	.41%
0.5	0.7	0.2	260	.04%	29.5%	6.23%	6.63%	.41%
0.5	0.8	0.2	260	.01%	31.17%	6.55%	7.75%	.48%
0.5	0.9	0.2	260	.01%	60.98%	6.97%	9.47%	.59%
0.5	0.99	0.2	260	.%	66.43%	7.42%	10.19%	.63%
0.6	0.01	0.2	260	.02%	87.37%	21.4%	23.41%	1.45%
0.6	0.1	0.2	260	.04%	87.45%	17.63%	20.01%	1.24%
0.6	0.2	0.2	260	.04%	77.12%	14.8%	16.66%	1.03%
0.6	0.3	0.2	260	.04%	84.55%	14.61%	16.82%	1.04%
0.6	0.4	0.2	260	.01%	47.8%	11.81%	12.64%	.78%

Training Corpus - MD5 Scorer (with ignorable file hashset) Parameter Tuning Results - First Hit (continued)

Default Score	TOI Score	C_p	Runs	Min	Max	Avg	Std Dev	Std Err
0.6	0.5	0.2	260	.02%	46.14%	9.68%	10.44%	.65%
0.6	0.6	0.2	260	.04%	42.04%	7.89%	8.26%	.51%
0.6	0.7	0.2	260	.04%	25.88%	6.54%	6.64%	.41%
0.6	0.8	0.2	260	.03%	63.19%	5.81%	7.0%	.43%
0.6	0.9	0.2	260	.01%	28.58%	6.36%	7.16%	.44%
0.6	0.99	0.2	260	.05%	56.03%	6.7%	7.87%	.49%
0.7	0.01	0.2	260	.03%	90.44%	22.68%	23.63%	1.47%
0.7	0.1	0.2	260	.02%	90.01%	20.15%	22.76%	1.41%
0.7	0.2	0.2	260	.04%	88.56%	18.01%	20.79%	1.29%
0.7	0.3	0.2	260	.01%	82.84%	15.53%	18.93%	1.17%
0.7	0.4	0.2	260	.%	78.49%	13.64%	15.01%	.93%
0.7	0.5	0.2	260	.03%	48.15%	10.48%	11.38%	.71%
0.7	0.6	0.2	260	.02%	37.3%	8.91%	9.77%	.61%
0.7	0.7	0.2	260	.01%	45.49%	7.72%	8.37%	.52%
0.7	0.8	0.2	260	.02%	62.8%	6.87%	7.44%	.46%
0.7	0.9	0.2	260	.05%	66.17%	6.98%	7.65%	.47%
0.7	0.99	0.2	260	.02%	64.57%	7.45%	9.55%	.59%
0.8	0.01	0.2	260	.01%	90.65%	22.79%	23.81%	1.48%
0.8	0.1	0.2	260	.02%	89.71%	23.63%	24.89%	1.54%
0.8	0.2	0.2	260	.01%	87.79%	19.66%	22.63%	1.4%
0.8	0.3	0.2	260	.%	85.92%	16.8%	19.51%	1.21%
0.8	0.4	0.2	260	.02%	80.79%	15.24%	17.49%	1.08%
0.8	0.5	0.2	260	.%	59.67%	12.6%	14.21%	.88%
0.8	0.6	0.2	260	.04%	57.73%	10.88%	12.3%	.76%
0.8	0.7	0.2	260	.03%	40.33%	9.17%	9.83%	.61%
0.8	0.8	0.2	260	.01%	29.72%	7.84%	7.68%	.48%
0.8	0.9	0.2	260	.05%	67.92%	7.58%	8.28%	.51%
0.8	0.99	0.2	260	.02%	66.66%	6.99%	8.12%	.5%
0.9	0.01	0.2	260	.01%	90.14%	24.09%	24.91%	1.54%
0.9	0.1	0.2	260	.16%	90.57%	24.76%	25.14%	1.56%
0.9	0.2	0.2	260	.05%	88.01%	21.77%	23.38%	1.45%
0.9	0.3	0.2	260	.01%	87.79%	18.98%	21.33%	1.32%
0.9	0.4	0.2	260	.03%	87.37%	18.57%	22.73%	1.41%
0.9	0.5	0.2	260	.04%	85.96%	16.69%	20.2%	1.25%
0.9	0.6	0.2	260	.01%	72.76%	12.99%	14.62%	.91%
0.9	0.7	0.2	260	.02%	45.7%	11.12%	12.1%	.75%
0.9	0.8	0.2	260	.%	46.25%	9.18%	9.64%	.6%

Training Corpus - MD5 Scorer (with ignorable file hashset) Parameter Tuning Results - First Hit (continued)

Default Score	TOI Score	C_p	Runs	Min	Max	Avg	Std Dev	Std Err
0.9	0.9	0.2	260	.02%	34.15%	8.39%	8.32%	.52%
0.9	0.99	0.2	260	.01%	30.62%	6.96%	7.16%	.44%
0.99	0.01	0.2	260	.05%	90.4%	28.09%	26.56%	1.65%
0.99	0.1	0.2	260	.02%	89.24%	26.98%	25.87%	1.6%
0.99	0.2	0.2	260	.01%	88.39%	22.61%	23.88%	1.48%
0.99	0.3	0.2	260	.04%	88.95%	22.02%	23.44%	1.45%
0.99	0.4	0.2	260	.03%	87.96%	20.31%	23.04%	1.43%
0.99	0.5	0.2	260	.02%	87.32%	17.11%	20.82%	1.29%
0.99	0.6	0.2	260	.14%	87.96%	14.84%	17.84%	1.11%
0.99	0.7	0.2	260	.01%	69.91%	12.9%	14.88%	.92%
0.99	0.8	0.2	260	.02%	47.94%	10.44%	11.83%	.73%
0.99	0.9	0.2	260	.02%	46.67%	9.03%	9.3%	.58%
0.99	0.99	0.2	260	.%	62.71%	8.48%	9.51%	.59%

Table D.4: Training Corpus - MD5 Scorer (without ignorable file hashset) Parameter Tuning Results - First Hit($C_p = 0.1$)

Default Score	TOI Score	C_p	Runs	Min	Max	Avg	Std Dev	Std Err
0.01	0.01	0.2	260	.02%	72.%	12.24%	15.45%	.96%
0.01	0.1	0.2	260	.03%	29.27%	7.51%	8.51%	.53%
0.01	0.2	0.2	260	.04%	54.08%	8.84%	14.06%	.87%
0.01	0.3	0.2	260	.07%	51.69%	8.42%	13.35%	.83%
0.01	0.4	0.2	260	.01%	50.87%	9.3%	13.65%	.85%
0.01	0.5	0.2	260	.04%	49.61%	9.79%	13.95%	.86%
0.01	0.6	0.2	260	.02%	49.36%	10.35%	14.17%	.88%
0.01	0.7	0.2	260	.01%	63.17%	10.83%	15.56%	.97%
0.01	0.8	0.2	260	.%	57.76%	11.69%	15.63%	.97%
0.01	0.9	0.2	260	.04%	69.01%	12.09%	16.28%	1.01%
0.01	0.99	0.2	260	.03%	75.42%	12.47%	17.49%	1.08%
0.1	0.01	0.2	260	.%	92.09%	16.%	22.3%	1.38%
0.1	0.1	0.2	260	.05%	68.6%	11.54%	14.5%	.9%
0.1	0.2	0.2	260	.01%	36.72%	7.96%	9.32%	.58%
0.1	0.3	0.2	260	.01%	53.78%	9.48%	14.25%	.88%
0.1	0.4	0.2	260	.05%	51.56%	8.58%	13.28%	.82%
0.1	0.5	0.2	260	.%	49.97%	9.49%	13.55%	.84%
0.1	0.6	0.2	260	.05%	56.98%	10.02%	14.49%	.9%
0.1	0.7	0.2	260	.01%	60.65%	10.45%	15.11%	.94%

*Training Corpus - MD5 Scorer (without ignorable file hashset) Parameter Tuning Results
- First Hit (continued)*

Default Score	TOI Score	C_p	Runs	Min	Max	Avg	Std Dev	Std Err
0.1	0.8	0.2	260	.04%	58.2%	10.99%	15.3%	.95%
0.1	0.9	0.2	260	.07%	75.8%	11.46%	16.38%	1.02%
0.1	0.99	0.2	260	.04%	66.62%	12.95%	17.61%	1.09%
0.2	0.01	0.2	260	.04%	91.37%	17.03%	23.68%	1.47%
0.2	0.1	0.2	260	.%	92.14%	15.38%	21.84%	1.35%
0.2	0.2	0.2	260	.%	70.8%	12.5%	15.87%	.98%
0.2	0.3	0.2	260	.06%	38.91%	8.19%	9.58%	.59%
0.2	0.4	0.2	260	.05%	53.81%	9.22%	14.21%	.88%
0.2	0.5	0.2	260	.06%	51.41%	8.69%	13.32%	.83%
0.2	0.6	0.2	260	.04%	49.95%	9.24%	13.44%	.83%
0.2	0.7	0.2	260	.04%	55.28%	9.99%	14.39%	.89%
0.2	0.8	0.2	260	.03%	57.72%	10.44%	14.83%	.92%
0.2	0.9	0.2	260	.04%	63.02%	11.26%	15.35%	.95%
0.2	0.99	0.2	260	.04%	59.41%	11.03%	15.46%	.96%
0.3	0.01	0.2	260	.02%	90.67%	18.84%	24.67%	1.53%
0.3	0.1	0.2	260	.02%	90.7%	17.46%	23.92%	1.48%
0.3	0.2	0.2	260	.04%	92.2%	15.61%	21.68%	1.34%
0.3	0.3	0.2	260	.01%	74.29%	11.93%	15.38%	.95%
0.3	0.4	0.2	260	.03%	36.67%	7.64%	8.66%	.54%
0.3	0.5	0.2	260	.05%	53.81%	8.65%	13.89%	.86%
0.3	0.6	0.2	260	.01%	51.57%	8.63%	13.33%	.83%
0.3	0.7	0.2	260	.04%	50.71%	9.41%	13.56%	.84%
0.3	0.8	0.2	260	.03%	49.54%	9.65%	13.62%	.84%
0.3	0.9	0.2	260	.01%	63.24%	10.29%	15.24%	.95%
0.3	0.99	0.2	260	.02%	62.5%	10.91%	15.13%	.94%
0.4	0.01	0.2	260	.02%	94.4%	23.18%	27.69%	1.72%
0.4	0.1	0.2	260	.02%	90.68%	18.52%	23.79%	1.48%
0.4	0.2	0.2	260	.06%	91.47%	17.79%	22.13%	1.37%
0.4	0.3	0.2	260	.%	92.74%	15.56%	22.86%	1.42%
0.4	0.4	0.2	260	.1%	73.25%	12.57%	15.65%	.97%
0.4	0.5	0.2	260	.04%	34.71%	7.59%	8.63%	.54%
0.4	0.6	0.2	260	.03%	53.82%	9.1%	14.07%	.87%
0.4	0.7	0.2	260	.03%	55.03%	8.61%	13.43%	.83%
0.4	0.8	0.2	260	.01%	49.9%	9.21%	13.67%	.85%
0.4	0.9	0.2	260	.01%	62.72%	9.54%	14.31%	.89%
0.4	0.99	0.2	260	.01%	61.71%	10.61%	15.08%	.94%
0.5	0.01	0.2	260	.01%	95.53%	28.47%	31.22%	1.94%
0.5	0.1	0.2	260	.%	96.83%	23.61%	27.85%	1.73%

*Training Corpus - MD5 Scorer (without ignorable file hashset) Parameter Tuning Results
- First Hit (continued)*

Default Score	TOI Score	C_p	Runs	Min	Max	Avg	Std Dev	Std Err
0.5	0.2	0.2	260	.01%	95.95%	19.63%	25.41%	1.58%
0.5	0.3	0.2	260	.01%	91.42%	17.4%	23.96%	1.49%
0.5	0.4	0.2	260	.05%	93.78%	15.75%	22.44%	1.39%
0.5	0.5	0.2	260	.02%	60.68%	11.65%	14.62%	.91%
0.5	0.6	0.2	260	.01%	37.17%	7.98%	9.1%	.56%
0.5	0.7	0.2	260	.%	53.77%	9.3%	14.19%	.88%
0.5	0.8	0.2	260	.01%	51.56%	8.72%	13.42%	.83%
0.5	0.9	0.2	260	.05%	54.81%	9.39%	13.79%	.86%
0.5	0.99	0.2	260	.04%	49.59%	9.65%	13.85%	.86%
0.6	0.01	0.2	260	.02%	97.44%	32.32%	32.59%	2.02%
0.6	0.1	0.2	260	.01%	94.58%	27.37%	30.02%	1.86%
0.6	0.2	0.2	260	.02%	94.41%	24.25%	28.63%	1.78%
0.6	0.3	0.2	260	.01%	90.78%	18.87%	24.31%	1.51%
0.6	0.4	0.2	260	.02%	91.67%	16.37%	22.41%	1.39%
0.6	0.5	0.2	260	.02%	92.33%	15.15%	21.49%	1.33%
0.6	0.6	0.2	260	.03%	62.13%	11.76%	14.94%	.93%
0.6	0.7	0.2	260	.04%	36.54%	7.62%	8.71%	.54%
0.6	0.8	0.2	260	.03%	53.77%	9.07%	14.01%	.87%
0.6	0.9	0.2	260	.01%	51.64%	8.63%	13.49%	.84%
0.6	0.99	0.2	260	.02%	50.84%	9.18%	13.46%	.83%
0.7	0.01	0.2	260	.04%	97.47%	34.61%	34.15%	2.12%
0.7	0.1	0.2	260	.06%	98.22%	32.17%	32.99%	2.05%
0.7	0.2	0.2	260	.09%	94.37%	27.29%	30.28%	1.88%
0.7	0.3	0.2	260	.02%	94.41%	23.36%	27.39%	1.7%
0.7	0.4	0.2	260	.04%	93.24%	20.61%	25.86%	1.6%
0.7	0.5	0.2	260	.05%	92.31%	16.54%	23.44%	1.45%
0.7	0.6	0.2	260	.02%	93.8%	15.84%	22.61%	1.4%
0.7	0.7	0.2	260	.05%	63.17%	12.3%	15.02%	.93%
0.7	0.8	0.2	260	.03%	36.61%	7.95%	9.17%	.57%
0.7	0.9	0.2	260	.01%	53.79%	9.16%	14.08%	.87%
0.7	0.99	0.2	260	.04%	51.73%	8.59%	13.35%	.83%
0.8	0.01	0.2	260	.15%	97.66%	36.9%	34.81%	2.16%
0.8	0.1	0.2	260	.%	98.52%	34.18%	34.76%	2.16%
0.8	0.2	0.2	260	.03%	97.45%	30.53%	32.93%	2.04%
0.8	0.3	0.2	260	.05%	96.73%	30.11%	32.1%	1.99%
0.8	0.4	0.2	260	.02%	94.4%	22.9%	27.52%	1.71%
0.8	0.5	0.2	260	.05%	91.71%	20.47%	25.23%	1.56%
0.8	0.6	0.2	260	.02%	91.09%	16.55%	22.61%	1.4%

Training Corpus - MD5 Scorer (without ignorable file hashset) Parameter Tuning Results - First Hit (continued)

Default Score	TOI Score	C_p	Runs	Min	Max	Avg	Std Dev	Std Err
0.8	0.7	0.2	260	.01%	92.43%	15.77%	22.61%	1.4%
0.8	0.8	0.2	260	.02%	73.86%	12.23%	15.39%	.95%
0.8	0.9	0.2	260	.02%	36.81%	7.99%	9.31%	.58%
0.8	0.99	0.2	260	.03%	54.09%	9.2%	14.21%	.88%
0.9	0.01	0.2	260	.02%	98.88%	38.99%	35.63%	2.21%
0.9	0.1	0.2	260	.07%	97.68%	35.04%	34.26%	2.12%
0.9	0.2	0.2	260	.12%	98.93%	32.99%	34.24%	2.12%
0.9	0.3	0.2	260	.03%	97.44%	27.75%	31.22%	1.94%
0.9	0.4	0.2	260	.03%	97.44%	27.19%	30.47%	1.89%
0.9	0.5	0.2	260	.07%	92.55%	21.91%	26.95%	1.67%
0.9	0.6	0.2	260	.03%	91.85%	19.47%	24.68%	1.53%
0.9	0.7	0.2	260	.04%	92.3%	17.65%	23.31%	1.45%
0.9	0.8	0.2	260	.04%	91.98%	14.9%	21.74%	1.35%
0.9	0.9	0.2	260	.06%	63.15%	12.03%	14.96%	.93%
0.9	0.99	0.2	260	.02%	29.85%	7.82%	8.78%	.54%
0.99	0.01	0.2	260	.02%	99.09%	41.94%	35.87%	2.22%
0.99	0.1	0.2	260	.01%	98.88%	37.89%	35.46%	2.2%
0.99	0.2	0.2	260	.01%	98.34%	35.93%	35.14%	2.18%
0.99	0.3	0.2	260	.%	97.7%	33.61%	34.32%	2.13%
0.99	0.4	0.2	260	.01%	95.58%	30.97%	32.82%	2.04%
0.99	0.5	0.2	260	.08%	97.43%	26.5%	30.67%	1.9%
0.99	0.6	0.2	260	.03%	95.09%	23.37%	28.01%	1.74%
0.99	0.7	0.2	260	.01%	90.82%	19.04%	23.99%	1.49%
0.99	0.8	0.2	260	.04%	91.98%	17.55%	24.17%	1.5%
0.99	0.9	0.2	260	.03%	93.64%	15.53%	22.86%	1.42%
0.99	0.99	0.2	260	.%	65.97%	12.29%	15.41%	.96%

Table D.5: Training Corpus - MD5 Scorer (with ignorable file hashset) Parameter Tuning Results - All FOI Found ($C_p = 0.2$)

Default Score	TOI Score	C_p	Runs	Min	Max	Avg	Std Dev	Std Err
0.01	0.01	0.1	210	1.19%	98.6%	42.28%	34.42%	2.13%
0.01	0.1	0.1	210	.87%	98.51%	33.33%	28.64%	1.78%
0.01	0.2	0.1	210	.76%	98.51%	32.29%	27.77%	1.72%
0.01	0.3	0.1	210	.74%	98.51%	32.52%	27.79%	1.72%
0.01	0.4	0.1	210	.73%	98.51%	32.03%	26.83%	1.66%
0.01	0.5	0.1	210	.71%	97.01%	31.39%	26.05%	1.62%
0.01	0.6	0.1	210	.72%	98.51%	31.69%	27.59%	1.71%

Training Corpus - MD5 Scorer (with ignorable file hashset) Parameter Tuning Results - All FOI Found (continued)

Default Score	TOI Score	C_p	Runs	Min	Max	Avg	Std Dev	Std Err
0.01	0.7	0.1	210	.71%	98.51%	30.54%	27.1%	1.68%
0.01	0.8	0.1	210	.64%	98.51%	31.12%	28.03%	1.74%
0.01	0.9	0.1	210	.81%	98.51%	30.73%	26.91%	1.67%
0.01	0.99	0.1	210	.83%	98.51%	32.17%	28.37%	1.76%
0.1	0.01	0.1	210	1.72%	99.81%	45.52%	34.78%	2.16%
0.1	0.1	0.1	210	1.08%	98.51%	38.38%	33.32%	2.07%
0.1	0.2	0.1	210	1.04%	98.51%	31.36%	27.83%	1.73%
0.1	0.3	0.1	210	.99%	98.51%	33.2%	27.73%	1.72%
0.1	0.4	0.1	210	.96%	98.51%	32.57%	27.71%	1.72%
0.1	0.5	0.1	210	.93%	98.51%	32.57%	27.93%	1.73%
0.1	0.6	0.1	210	.91%	98.51%	32.56%	27.44%	1.7%
0.1	0.7	0.1	210	.68%	98.51%	32.34%	29.29%	1.82%
0.1	0.8	0.1	210	.67%	98.51%	31.1%	27.21%	1.69%
0.1	0.9	0.1	210	.69%	98.51%	32.18%	28.42%	1.76%
0.1	0.99	0.1	210	.72%	98.51%	32.63%	29.99%	1.86%
0.2	0.01	0.1	210	2.33%	99.78%	59.26%	31.93%	1.98%
0.2	0.1	0.1	210	1.83%	99.75%	45.95%	33.42%	2.07%
0.2	0.2	0.1	210	1.06%	98.51%	37.72%	33.25%	2.06%
0.2	0.3	0.1	210	1.04%	98.51%	30.95%	28.3%	1.75%
0.2	0.4	0.1	210	1.03%	98.51%	31.29%	27.72%	1.72%
0.2	0.5	0.1	210	1.01%	98.51%	32.15%	27.47%	1.7%
0.2	0.6	0.1	210	.93%	98.51%	32.37%	27.25%	1.69%
0.2	0.7	0.1	208	.94%	98.51%	31.05%	27.1%	1.68%
0.2	0.8	0.1	210	.89%	97.01%	31.61%	26.85%	1.67%
0.2	0.9	0.1	200	.9%	98.51%	31.53%	27.52%	1.71%
0.2	0.99	0.1	201	.86%	98.51%	31.87%	28.32%	1.76%
0.3	0.01	0.1	200	1.38%	99.8%	52.79%	37.78%	2.34%
0.3	0.1	0.1	200	1.27%	99.78%	45.98%	36.26%	2.25%
0.3	0.2	0.1	200	1.16%	99.7%	39.93%	32.26%	2.0%
0.3	0.3	0.1	200	1.12%	98.51%	36.97%	32.33%	2.0%
0.3	0.4	0.1	200	1.07%	98.51%	31.92%	29.05%	1.8%
0.3	0.5	0.1	200	1.06%	98.51%	30.69%	27.44%	1.7%
0.3	0.6	0.1	200	1.01%	98.51%	30.96%	26.44%	1.64%
0.3	0.7	0.1	200	.96%	98.51%	31.64%	28.07%	1.74%
0.3	0.8	0.1	200	.95%	98.51%	31.15%	26.48%	1.64%
0.3	0.9	0.1	200	.92%	98.51%	30.79%	26.42%	1.64%
0.3	0.99	0.1	200	.82%	98.51%	30.93%	27.19%	1.69%
0.4	0.01	0.1	200	1.56%	99.7%	55.56%	38.08%	2.36%

Training Corpus - MD5 Scorer (with ignorable file hashset) Parameter Tuning Results - All FOI Found (continued)

Default Score	TOI Score	C_p	Runs	Min	Max	Avg	Std Dev	Std Err
0.4	0.1	0.1	200	1.46%	99.77%	53.5%	37.73%	2.34%
0.4	0.2	0.1	200	1.3%	99.77%	48.82%	36.25%	2.25%
0.4	0.3	0.1	200	1.19%	99.33%	40.24%	32.14%	1.99%
0.4	0.4	0.1	200	1.17%	98.51%	37.22%	32.51%	2.02%
0.4	0.5	0.1	200	1.09%	98.51%	32.27%	27.89%	1.73%
0.4	0.6	0.1	200	1.07%	98.51%	32.52%	28.85%	1.79%
0.4	0.7	0.1	200	1.02%	98.51%	30.87%	26.77%	1.66%
0.4	0.8	0.1	200	.97%	98.51%	31.67%	27.29%	1.69%
0.4	0.9	0.1	200	.99%	98.51%	31.75%	26.63%	1.65%
0.4	0.99	0.1	200	.93%	98.51%	31.65%	27.17%	1.69%
0.5	0.01	0.1	200	2.16%	99.81%	56.9%	37.52%	2.33%
0.5	0.1	0.1	200	1.86%	99.86%	55.71%	37.79%	2.34%
0.5	0.2	0.1	200	1.61%	99.77%	54.47%	37.88%	2.35%
0.5	0.3	0.1	200	1.42%	99.78%	49.53%	35.4%	2.2%
0.5	0.4	0.1	200	1.3%	98.51%	39.56%	31.04%	1.93%
0.5	0.5	0.1	200	1.17%	98.51%	37.18%	31.87%	1.98%
0.5	0.6	0.1	200	1.14%	98.51%	32.25%	28.23%	1.75%
0.5	0.7	0.1	200	1.07%	98.51%	31.66%	27.5%	1.71%
0.5	0.8	0.1	200	1.02%	98.51%	32.2%	27.41%	1.7%
0.5	0.9	0.1	200	1.02%	98.51%	32.02%	26.88%	1.67%
0.5	0.99	0.1	200	.99%	98.51%	32.08%	26.86%	1.67%
0.6	0.01	0.1	200	3.36%	99.73%	58.27%	36.52%	2.26%
0.6	0.1	0.1	200	2.9%	99.71%	57.55%	36.85%	2.29%
0.6	0.2	0.1	200	2.4%	99.73%	56.01%	37.1%	2.3%
0.6	0.3	0.1	200	1.87%	99.6%	54.8%	37.38%	2.32%
0.6	0.4	0.1	200	1.55%	98.51%	51.28%	35.24%	2.19%
0.6	0.5	0.1	200	1.3%	98.51%	40.26%	29.67%	1.84%
0.6	0.6	0.1	200	1.25%	98.51%	37.85%	31.77%	1.97%
0.6	0.7	0.1	200	1.21%	98.51%	34.03%	29.35%	1.82%
0.6	0.8	0.1	200	1.14%	98.51%	31.79%	25.84%	1.6%
0.6	0.9	0.1	200	1.04%	98.51%	33.23%	27.46%	1.7%
0.6	0.99	0.1	200	1.01%	98.51%	32.18%	26.58%	1.65%
0.7	0.01	0.1	200	3.3%	99.84%	61.6%	34.55%	2.14%
0.7	0.1	0.1	200	3.73%	99.84%	58.33%	36.62%	2.27%
0.7	0.2	0.1	200	3.29%	99.8%	58.33%	36.44%	2.26%
0.7	0.3	0.1	200	3.11%	99.74%	56.27%	37.08%	2.3%
0.7	0.4	0.1	200	2.58%	99.84%	55.02%	37.48%	2.32%
0.7	0.5	0.1	200	1.84%	98.51%	52.06%	34.61%	2.15%

Training Corpus - MD5 Scorer (with ignorable file hashset) Parameter Tuning Results - All FOI Found (continued)

Default Score	TOI Score	C_p	Runs	Min	Max	Avg	Std Dev	Std Err
0.7	0.6	0.1	200	1.61%	98.51%	41.93%	29.28%	1.82%
0.7	0.7	0.1	200	1.26%	98.51%	37.84%	30.48%	1.89%
0.7	0.8	0.1	200	1.16%	98.51%	35.15%	29.2%	1.81%
0.7	0.9	0.1	200	1.02%	98.51%	34.75%	28.73%	1.78%
0.7	0.99	0.1	200	1.07%	98.51%	36.15%	28.97%	1.8%
0.8	0.01	0.1	200	4.21%	99.8%	63.9%	32.29%	2.%
0.8	0.1	0.1	200	4.82%	99.71%	62.8%	33.06%	2.05%
0.8	0.2	0.1	200	3.65%	99.43%	60.33%	34.77%	2.16%
0.8	0.3	0.1	200	3.29%	99.6%	58.3%	36.04%	2.24%
0.8	0.4	0.1	200	3.19%	99.65%	56.93%	36.04%	2.24%
0.8	0.5	0.1	200	2.99%	98.99%	56.26%	35.92%	2.23%
0.8	0.6	0.1	200	2.49%	98.79%	53.39%	34.25%	2.12%
0.8	0.7	0.1	200	1.88%	98.51%	43.71%	27.68%	1.72%
0.8	0.8	0.1	200	1.58%	98.51%	38.77%	30.5%	1.89%
0.8	0.9	0.1	200	1.33%	98.51%	36.27%	29.%	1.8%
0.8	0.99	0.1	205	1.28%	98.51%	36.39%	28.85%	1.79%
0.9	0.01	0.1	213	3.1%	99.87%	64.56%	31.7%	1.97%
0.9	0.1	0.1	215	4.17%	99.85%	64.08%	31.05%	1.93%
0.9	0.2	0.1	227	3.83%	99.78%	61.97%	32.68%	2.03%
0.9	0.3	0.1	232	3.26%	99.64%	61.2%	33.16%	2.06%
0.9	0.4	0.1	239	3.72%	99.84%	56.88%	34.55%	2.14%
0.9	0.5	0.1	236	3.53%	98.99%	56.41%	34.74%	2.15%
0.9	0.6	0.1	234	3.19%	98.99%	55.77%	34.97%	2.17%
0.9	0.7	0.1	239	3.02%	99.04%	53.06%	33.%	2.05%
0.9	0.8	0.1	233	2.62%	99.63%	46.65%	27.79%	1.72%
0.9	0.9	0.1	246	2.%	98.51%	41.81%	29.27%	1.82%
0.9	0.99	0.1	251	1.59%	98.51%	39.31%	29.34%	1.82%
0.99	0.01	0.1	238	4.09%	99.78%	64.65%	31.17%	1.93%
0.99	0.1	0.1	227	4.04%	99.88%	63.04%	31.44%	1.95%
0.99	0.2	0.1	233	4.01%	99.84%	63.43%	31.2%	1.94%
0.99	0.3	0.1	243	4.07%	99.87%	62.58%	31.48%	1.95%
0.99	0.4	0.1	251	3.9%	99.8%	61.02%	33.26%	2.06%
0.99	0.5	0.1	247	4.08%	98.99%	58.73%	32.58%	2.02%
0.99	0.6	0.1	260	3.8%	99.5%	59.16%	33.03%	2.05%
0.99	0.7	0.1	256	3.3%	99.04%	56.03%	32.85%	2.04%
0.99	0.8	0.1	258	3.08%	99.63%	53.07%	32.45%	2.01%
0.99	0.9	0.1	267	3.36%	98.51%	47.22%	28.83%	1.79%
0.99	0.99	0.1	299	2.79%	98.51%	45.36%	28.75%	1.78%

Table D.6: Training Corpus - MD5 Scorer (without ignorable file hashset) Parameter Tuning Results - All FOI Found ($C_p = 0.2$)

Default Score	TOI Score	C_p	Runs	Min	Max	Avg	Std Dev	Std Err
0.01	0.01	0.1	210	1.17%	98.53%	44.71%	35.99%	2.23%
0.01	0.1	0.1	210	1.07%	98.51%	34.56%	29.03%	1.8%
0.01	0.2	0.1	210	1.09%	98.51%	33.03%	27.3%	1.69%
0.01	0.3	0.1	210	1.07%	98.51%	34.3%	28.62%	1.78%
0.01	0.4	0.1	210	1.01%	98.51%	33.98%	27.7%	1.72%
0.01	0.5	0.1	210	.9%	97.01%	33.08%	26.15%	1.62%
0.01	0.6	0.1	210	.89%	98.51%	33.91%	27.02%	1.68%
0.01	0.7	0.1	210	.89%	98.51%	33.99%	27.07%	1.68%
0.01	0.8	0.1	210	.85%	98.51%	35.99%	28.08%	1.74%
0.01	0.9	0.1	210	.91%	98.51%	35.06%	27.98%	1.74%
0.01	0.99	0.1	210	.86%	98.51%	36.35%	29.75%	1.85%
0.1	0.01	0.1	210	1.17%	99.82%	44.69%	34.72%	2.15%
0.1	0.1	0.1	210	1.37%	98.61%	44.69%	35.9%	2.23%
0.1	0.2	0.1	210	1.23%	98.51%	35.1%	29.14%	1.81%
0.1	0.3	0.1	210	1.15%	98.51%	33.4%	27.29%	1.69%
0.1	0.4	0.1	210	1.11%	98.51%	33.24%	27.37%	1.7%
0.1	0.5	0.1	210	1.04%	98.51%	33.46%	27.64%	1.71%
0.1	0.6	0.1	210	1.01%	98.51%	34.55%	26.98%	1.67%
0.1	0.7	0.1	210	1.01%	97.01%	35.11%	27.99%	1.74%
0.1	0.8	0.1	210	.94%	98.51%	32.77%	26.51%	1.64%
0.1	0.9	0.1	210	.96%	98.51%	36.86%	28.47%	1.77%
0.1	0.99	0.1	210	.9%	98.51%	37.04%	29.17%	1.81%
0.2	0.01	0.1	210	1.61%	99.8%	50.79%	38.57%	2.39%
0.2	0.1	0.1	210	1.55%	99.83%	45.96%	35.35%	2.19%
0.2	0.2	0.1	210	1.37%	98.9%	44.89%	36.2%	2.24%
0.2	0.3	0.1	210	1.34%	98.51%	34.18%	28.3%	1.76%
0.2	0.4	0.1	210	1.28%	98.51%	33.84%	27.26%	1.69%
0.2	0.5	0.1	210	1.19%	98.51%	34.04%	27.33%	1.69%
0.2	0.6	0.1	210	1.11%	98.51%	33.9%	26.85%	1.67%
0.2	0.7	0.1	205	1.04%	98.51%	34.05%	26.85%	1.67%
0.2	0.8	0.1	201	1.%	98.51%	35.76%	28.69%	1.78%
0.2	0.9	0.1	200	.98%	98.51%	33.99%	27.34%	1.7%
0.2	0.99	0.1	200	.94%	98.51%	34.68%	28.02%	1.74%
0.3	0.01	0.1	200	2.16%	99.81%	59.59%	39.31%	2.44%
0.3	0.1	0.1	200	1.86%	99.86%	54.%	39.47%	2.45%
0.3	0.2	0.1	200	1.59%	99.69%	46.78%	35.44%	2.2%
0.3	0.3	0.1	200	1.52%	98.96%	45.61%	36.71%	2.28%
0.3	0.4	0.1	200	1.47%	98.51%	34.84%	28.97%	1.8%

*Training Corpus - MD5 Scorer (without ignorable file hashset) Parameter Tuning Results
- All FOI Found (continued)*

Default Score	TOI Score	C_p	Runs	Min	Max	Avg	Std Dev	Std Err
0.3	0.5	0.1	200	1.32%	98.51%	33.79%	27.86%	1.73%
0.3	0.6	0.1	200	1.2%	98.51%	33.92%	26.57%	1.65%
0.3	0.7	0.1	200	1.18%	98.51%	34.84%	28.44%	1.76%
0.3	0.8	0.1	200	1.11%	98.51%	34.59%	27.07%	1.68%
0.3	0.9	0.1	200	1.04%	98.51%	35.24%	28.47%	1.77%
0.3	0.99	0.1	200	1.02%	98.51%	35.95%	28.95%	1.8%
0.4	0.01	0.1	200	3.45%	99.81%	65.2%	38.66%	2.4%
0.4	0.1	0.1	200	2.74%	99.84%	62.77%	38.84%	2.41%
0.4	0.2	0.1	200	2.2%	99.86%	54.52%	39.13%	2.43%
0.4	0.3	0.1	200	1.87%	99.77%	47.03%	35.44%	2.2%
0.4	0.4	0.1	200	1.71%	98.86%	46.41%	37.01%	2.3%
0.4	0.5	0.1	200	1.51%	98.51%	34.63%	28.53%	1.77%
0.4	0.6	0.1	200	1.48%	98.51%	33.87%	27.94%	1.73%
0.4	0.7	0.1	200	1.38%	98.51%	34.21%	27.75%	1.72%
0.4	0.8	0.1	200	1.33%	98.51%	34.1%	26.85%	1.67%
0.4	0.9	0.1	200	1.31%	98.51%	34.92%	27.64%	1.71%
0.4	0.99	0.1	200	1.14%	98.51%	34.09%	27.36%	1.7%
0.5	0.01	0.1	200	5.62%	99.86%	67.69%	37.11%	2.3%
0.5	0.1	0.1	200	5.69%	99.58%	66.52%	38.45%	2.38%
0.5	0.2	0.1	200	3.58%	99.88%	64.85%	38.8%	2.41%
0.5	0.3	0.1	200	2.86%	99.82%	59.31%	37.35%	2.32%
0.5	0.4	0.1	200	2.44%	99.86%	48.46%	35.7%	2.21%
0.5	0.5	0.1	200	1.96%	98.94%	46.09%	36.77%	2.28%
0.5	0.6	0.1	200	1.75%	98.51%	35.07%	29.09%	1.8%
0.5	0.7	0.1	200	1.62%	98.51%	34.59%	28.49%	1.77%
0.5	0.8	0.1	200	1.47%	98.51%	34.03%	27.41%	1.7%
0.5	0.9	0.1	200	1.42%	98.51%	34.27%	27.45%	1.7%
0.5	0.99	0.1	200	1.29%	98.51%	34.11%	26.46%	1.64%
0.6	0.01	0.1	200	6.24%	99.89%	78.28%	29.78%	1.85%
0.6	0.1	0.1	200	6.26%	99.89%	81.3%	29.52%	1.83%
0.6	0.2	0.1	200	6.25%	99.87%	66.96%	37.24%	2.31%
0.6	0.3	0.1	200	5.51%	99.78%	65.74%	38.67%	2.4%
0.6	0.4	0.1	200	29.69%	99.92%	83.94%	23.12%	1.43%
0.6	0.5	0.1	200	5.76%	99.8%	75.51%	29.98%	1.86%
0.6	0.6	0.1	200	2.14%	98.67%	47.12%	37.02%	2.3%
0.6	0.7	0.1	200	1.07%	98.51%	35.05%	30.22%	1.87%
0.6	0.8	0.1	200	.85%	98.51%	34.62%	28.97%	1.8%
0.6	0.9	0.1	200	.83%	98.51%	33.42%	27.26%	1.69%

*Training Corpus - MD5 Scorer (without ignorable file hashset) Parameter Tuning Results
- All FOI Found (continued)*

Default Score	TOI Score	C_p	Runs	Min	Max	Avg	Std Dev	Std Err
0.6	0.99	0.1	200	.79%	98.51%	34.29%	27.74%	1.72%
0.7	0.01	0.1	200	28.51%	99.93%	89.71%	17.94%	1.11%
0.7	0.1	0.1	200	27.41%	99.93%	88.95%	19.47%	1.21%
0.7	0.2	0.1	200	31.23%	99.93%	87.69%	20.87%	1.29%
0.7	0.3	0.1	200	28.39%	99.93%	86.98%	21.83%	1.35%
0.7	0.4	0.1	200	28.92%	99.93%	85.74%	22.51%	1.4%
0.7	0.5	0.1	200	29.31%	99.92%	85.54%	22.44%	1.39%
0.7	0.6	0.1	200	7.84%	99.92%	78.15%	28.86%	1.79%
0.7	0.7	0.1	200	2.54%	99.02%	48.15%	36.93%	2.29%
0.7	0.8	0.1	200	1.12%	98.51%	35.23%	30.34%	1.88%
0.7	0.9	0.1	200	.92%	98.51%	33.13%	27.86%	1.73%
0.7	0.99	0.1	200	.83%	98.51%	34.93%	30.37%	1.88%
0.8	0.01	0.1	200	38.77%	99.93%	90.81%	15.99%	.99%
0.8	0.1	0.1	200	31.91%	99.93%	90.77%	16.28%	1.01%
0.8	0.2	0.1	200	25.35%	99.93%	89.19%	18.77%	1.16%
0.8	0.3	0.1	200	20.98%	99.93%	88.13%	21.04%	1.3%
0.8	0.4	0.1	203	30.44%	99.93%	87.72%	21.11%	1.31%
0.8	0.5	0.1	207	29.69%	99.93%	87.04%	21.5%	1.33%
0.8	0.6	0.1	206	30.14%	99.93%	85.94%	22.4%	1.39%
0.8	0.7	0.1	213	33.75%	99.91%	84.09%	20.53%	1.27%
0.8	0.8	0.1	200	3.51%	99.28%	49.38%	36.44%	2.26%
0.8	0.9	0.1	200	1.23%	98.51%	36.31%	30.62%	1.9%
0.8	0.99	0.1	200	2.77%	98.51%	34.83%	27.94%	1.73%
0.9	0.01	0.1	218	9.81%	99.93%	84.34%	26.54%	1.65%
0.9	0.1	0.1	231	9.79%	99.93%	83.99%	26.61%	1.65%
0.9	0.2	0.1	236	9.79%	99.93%	83.53%	27.4%	1.7%
0.9	0.3	0.1	235	9.69%	99.92%	81.97%	28.85%	1.79%
0.9	0.4	0.1	243	9.78%	99.93%	81.94%	29.21%	1.81%
0.9	0.5	0.1	247	9.76%	99.93%	80.34%	28.7%	1.78%
0.9	0.6	0.1	244	9.76%	99.92%	77.8%	29.58%	1.83%
0.9	0.7	0.1	242	9.79%	99.92%	74.46%	31.81%	1.97%
0.9	0.8	0.1	232	9.78%	99.77%	62.67%	32.69%	2.03%
0.9	0.9	0.1	237	7.77%	99.36%	50.84%	34.89%	2.16%
0.9	0.99	0.1	235	2.96%	98.51%	37.91%	29.04%	1.8%
0.99	0.01	0.1	225	12.6%	99.93%	84.57%	25.78%	1.6%
0.99	0.1	0.1	233	12.52%	99.93%	83.74%	26.19%	1.62%
0.99	0.2	0.1	240	12.82%	99.93%	84.81%	25.98%	1.61%
0.99	0.3	0.1	237	12.85%	99.93%	84.01%	26.8%	1.66%

*Training Corpus - MD5 Scorer (without ignorable file hashset) Parameter Tuning Results
 - All FOI Found (continued)*

Default Score	TOI Score	C_p	Runs	Min	Max	Avg	Std Dev	Std Err
0.99	0.4	0.1	241	12.3%	99.93%	82.94%	27.75%	1.72%
0.99	0.5	0.1	261	12.65%	99.93%	82.68%	28.32%	1.76%
0.99	0.6	0.1	253	12.68%	99.93%	82.2%	28.74%	1.78%
0.99	0.7	0.1	258	12.68%	99.93%	80.79%	28.51%	1.77%
0.99	0.8	0.1	271	12.66%	99.93%	78.62%	29.23%	1.81%
0.99	0.9	0.1	288	12.58%	99.96%	72.25%	29.4%	1.82%
0.99	0.99	0.1	291	12.36%	99.11%	55.91%	32.23%	2.0%

D.3 PhotoDNA Scorer Parameter Tuning Results

Table D.7: Training Corpus - PhotoDNA Scorer (with ignorable file hashset) Parameter Tuning Results - First Hit

Default Score	TOI Score	C param	Runs	Min	Max	Avg	Std Dev	Std Err
0.01	0.01	0.2	260	.08%	25.37%	4.83%	5.1%	.45%
0.01	0.1	0.2	260	.02%	19.1%	4.16%	4.26%	.37%
0.01	0.2	0.2	260	.03%	26.87%	4.15%	4.52%	.4%
0.01	0.3	0.2	260	.08%	32.84%	5.01%	6.12%	.54%
0.01	0.4	0.2	260	.08%	29.62%	5.23%	6.73%	.59%
0.01	0.5	0.2	260	.03%	28.36%	5.47%	6.69%	.59%
0.01	0.6	0.2	260	.01%	59.33%	5.45%	7.43%	.65%
0.01	0.7	0.2	260	.06%	73.24%	7.06%	12.33%	1.08%
0.01	0.8	0.2	260	.08%	74.85%	7.37%	13.53%	1.19%
0.01	0.9	0.2	260	.04%	74.45%	5.58%	10.22%	.9%
0.01	0.99	0.2	260	.04%	75.25%	6.82%	12.06%	1.06%
0.1	0.01	0.2	260	.04%	28.52%	6.%	6.58%	.58%
0.1	0.1	0.2	260	.08%	21.67%	5.28%	5.31%	.47%
0.1	0.2	0.2	260	.04%	16.42%	4.22%	4.07%	.36%
0.1	0.3	0.2	260	.01%	22.05%	4.53%	4.83%	.42%
0.1	0.4	0.2	260	.07%	26.87%	5.1%	5.35%	.47%
0.1	0.5	0.2	260	.04%	38.23%	5.64%	7.48%	.66%
0.1	0.6	0.2	260	.07%	66.8%	5.83%	8.3%	.73%
0.1	0.7	0.2	260	.07%	76.06%	7.54%	13.87%	1.22%
0.1	0.8	0.2	260	.05%	74.85%	7.%	13.3%	1.17%
0.1	0.9	0.2	260	.02%	75.45%	7.68%	16.11%	1.41%
0.1	0.99	0.2	260	.04%	75.05%	8.42%	15.28%	1.34%
0.2	0.01	0.2	260	.04%	38.67%	7.82%	10.35%	.91%
0.2	0.1	0.2	260	.08%	37.98%	6.25%	7.42%	.65%
0.2	0.2	0.2	260	.03%	21.86%	5.12%	5.39%	.47%
0.2	0.3	0.2	260	.02%	15.32%	4.46%	4.07%	.36%
0.2	0.4	0.2	260	.02%	26.84%	5.21%	5.21%	.46%
0.2	0.5	0.2	260	.01%	46.88%	6.97%	9.42%	.83%
0.2	0.6	0.2	260	.05%	69.82%	7.36%	12.13%	1.06%
0.2	0.7	0.2	260	.01%	74.45%	7.29%	13.62%	1.19%
0.2	0.8	0.2	260	.01%	75.65%	11.16%	20.15%	1.77%
0.2	0.9	0.2	260	.03%	75.45%	8.21%	14.7%	1.29%
0.2	0.99	0.2	260	.%	74.85%	8.63%	16.16%	1.42%
0.3	0.01	0.2	260	.01%	45.45%	8.73%	11.02%	.97%
0.3	0.1	0.2	260	.04%	40.54%	7.97%	10.2%	.89%
0.3	0.2	0.2	260	.04%	29.96%	5.91%	6.87%	.6%

Training Corpus - PhotoDNA Scorer (with ignorable file hashset) Parameter Tuning Results - First Hit (continued)

Default Score	TOI Score	C param	Runs	Min	Max	Avg	Std Dev	Std Err
0.3	0.3	0.2	260	.02%	16.35%	5.54%	5.2%	.46%
0.3	0.4	0.2	260	.05%	17.91%	5.2%	4.5%	.39%
0.3	0.5	0.2	260	.05%	46.08%	7.02%	9.67%	.85%
0.3	0.6	0.2	260	.05%	72.43%	8.39%	14.49%	1.27%
0.3	0.7	0.2	260	.%	76.86%	10.94%	19.5%	1.71%
0.3	0.8	0.2	260	.04%	76.86%	11.33%	19.58%	1.72%
0.3	0.9	0.2	260	.01%	75.05%	10.09%	18.46%	1.62%
0.3	0.99	0.2	260	.%	75.25%	11.38%	19.61%	1.72%
0.4	0.01	0.2	260	.05%	79.04%	11.26%	14.98%	1.31%
0.4	0.1	0.2	260	.06%	44.3%	9.48%	11.1%	.97%
0.4	0.2	0.2	260	.11%	40.31%	8.55%	10.37%	.91%
0.4	0.3	0.2	260	.11%	27.44%	7.16%	7.35%	.64%
0.4	0.4	0.2	260	.03%	28.37%	6.83%	7.19%	.63%
0.4	0.5	0.2	260	.%	69.62%	9.07%	12.59%	1.1%
0.4	0.6	0.2	260	.%	76.46%	11.29%	19.19%	1.68%
0.4	0.7	0.2	260	.02%	75.25%	11.68%	19.31%	1.69%
0.4	0.8	0.2	260	.01%	75.65%	11.33%	19.25%	1.69%
0.4	0.9	0.2	260	.05%	75.05%	11.41%	19.35%	1.7%
0.4	0.99	0.2	260	.02%	76.66%	11.53%	20.04%	1.76%
0.5	0.01	0.2	260	.01%	86.26%	15.62%	20.62%	1.81%
0.5	0.1	0.2	260	.04%	59.45%	12.21%	14.18%	1.24%
0.5	0.2	0.2	260	.04%	43.26%	10.82%	12.6%	1.11%
0.5	0.3	0.2	260	.04%	58.12%	10.69%	13.23%	1.16%
0.5	0.4	0.2	260	.03%	57.34%	10.61%	14.28%	1.25%
0.5	0.5	0.2	260	.09%	75.45%	10.8%	18.03%	1.58%
0.5	0.6	0.2	260	.05%	75.65%	12.07%	18.94%	1.66%
0.5	0.7	0.2	260	.03%	76.66%	10.66%	18.57%	1.63%
0.5	0.8	0.2	260	.06%	75.05%	11.61%	19.79%	1.74%
0.5	0.9	0.2	260	.02%	75.45%	11.39%	19.4%	1.7%
0.5	0.99	0.2	260	.04%	76.66%	11.41%	19.68%	1.73%
0.6	0.01	0.2	260	.01%	90.44%	16.72%	21.17%	1.86%
0.6	0.1	0.2	260	.01%	87.92%	17.29%	22.16%	1.94%
0.6	0.2	0.2	260	.01%	80.79%	14.12%	18.42%	1.62%
0.6	0.3	0.2	260	.02%	67.4%	13.73%	17.66%	1.55%
0.6	0.4	0.2	260	.03%	76.66%	13.14%	19.67%	1.73%
0.6	0.5	0.2	260	.11%	75.65%	12.26%	19.51%	1.71%
0.6	0.6	0.2	260	.02%	77.06%	11.28%	18.92%	1.66%
0.6	0.7	0.2	260	.07%	75.45%	11.75%	19.43%	1.7%

Training Corpus - PhotoDNA Scorer (with ignorable file hashset) Parameter Tuning Results - First Hit (continued)

Default Score	TOI Score	C param	Runs	Min	Max	Avg	Std Dev	Std Err
0.6	0.8	0.2	260	.03%	76.26%	12.04%	19.77%	1.73%
0.6	0.9	0.2	260	.04%	76.26%	11.11%	19.24%	1.69%
0.6	0.99	0.2	260	.08%	74.65%	11.89%	19.89%	1.74%
0.7	0.01	0.2	260	.08%	90.35%	21.79%	25.69%	2.25%
0.7	0.1	0.2	260	.11%	85.06%	19.47%	23.95%	2.1%
0.7	0.2	0.2	260	.17%	84.89%	17.84%	23.63%	2.07%
0.7	0.3	0.2	260	.06%	82.16%	15.67%	21.77%	1.91%
0.7	0.4	0.2	260	.09%	75.86%	13.56%	19.93%	1.75%
0.7	0.5	0.2	260	.03%	75.45%	13.74%	20.06%	1.76%
0.7	0.6	0.2	260	.04%	76.26%	12.37%	19.58%	1.72%
0.7	0.7	0.2	260	.07%	75.45%	12.86%	19.56%	1.72%
0.7	0.8	0.2	260	.08%	76.26%	12.49%	19.65%	1.72%
0.7	0.9	0.2	260	.03%	74.85%	12.52%	19.44%	1.7%
0.7	0.99	0.2	260	.06%	78.07%	12.24%	20.26%	1.78%
0.8	0.01	0.2	260	.%	89.76%	20.92%	25.21%	2.21%
0.8	0.1	0.2	260	.06%	90.4%	20.2%	24.63%	2.16%
0.8	0.2	0.2	260	.04%	85.87%	18.83%	24.03%	2.11%
0.8	0.3	0.2	260	.18%	85.96%	17.37%	23.94%	2.1%
0.8	0.4	0.2	260	.05%	75.45%	16.77%	22.66%	1.99%
0.8	0.5	0.2	260	.04%	75.65%	15.65%	20.98%	1.84%
0.8	0.6	0.2	260	.04%	75.45%	13.8%	19.94%	1.75%
0.8	0.7	0.2	260	.04%	75.05%	13.11%	19.85%	1.74%
0.8	0.8	0.2	260	.05%	75.45%	13.34%	19.56%	1.72%
0.8	0.9	0.2	260	.09%	76.06%	12.56%	19.78%	1.73%
0.8	0.99	0.2	260	.05%	76.26%	12.68%	20.1%	1.76%
0.9	0.01	0.2	260	.17%	89.84%	24.74%	27.91%	2.45%
0.9	0.1	0.2	260	.%	89.59%	21.78%	26.12%	2.29%
0.9	0.2	0.2	260	.08%	88.56%	21.76%	25.86%	2.27%
0.9	0.3	0.2	260	.04%	88.9%	19.81%	26.66%	2.34%
0.9	0.4	0.2	260	.01%	87.24%	18.55%	25.02%	2.19%
0.9	0.5	0.2	260	.12%	82.2%	16.33%	22.72%	1.99%
0.9	0.6	0.2	260	.03%	76.06%	14.58%	20.38%	1.79%
0.9	0.7	0.2	260	.09%	76.66%	13.57%	20.19%	1.77%
0.9	0.8	0.2	260	.05%	76.46%	13.02%	19.57%	1.72%
0.9	0.9	0.2	260	.07%	76.86%	14.03%	19.94%	1.75%
0.9	0.99	0.2	260	.08%	75.45%	13.17%	19.72%	1.73%
0.99	0.01	0.2	260	.08%	90.06%	22.9%	25.7%	2.25%
0.99	0.1	0.2	260	.12%	89.24%	24.36%	27.54%	2.42%

Training Corpus - PhotoDNA Scorer (with ignorable file hashset) Parameter Tuning Results - First Hit (continued)

Default Score	TOI Score	C param	Runs	Min	Max	Avg	Std Dev	Std Err
0.99	0.2	0.2	260	.12%	88.31%	23.2%	27.06%	2.37%
0.99	0.3	0.2	260	.16%	88.01%	20.83%	26.9%	2.36%
0.99	0.4	0.2	260	.04%	87.45%	21.42%	27.08%	2.38%
0.99	0.5	0.2	260	.03%	77.55%	17.39%	22.99%	2.02%
0.99	0.6	0.2	260	.08%	76.61%	17.47%	22.31%	1.96%
0.99	0.7	0.2	260	.01%	75.45%	15.52%	20.75%	1.82%
0.99	0.8	0.2	260	.03%	76.26%	15.05%	20.48%	1.8%
0.99	0.9	0.2	260	.04%	75.45%	14.5%	21.2%	1.86%
0.99	0.99	0.2	260	.04%	76.26%	13.3%	20.06%	1.76%

Table D.8: Training Corpus - PhotoDNA Scorer (with ignorable file hashset) Parameter Tuning Results - All FOI Found

Default Score	TOI Score	C param	Runs	Min	Max	Avg	Std Dev	Std Err
0.01	0.01	0.1	260	.56%	88.06%	20.63%	21.52%	1.89%
0.01	0.1	0.1	260	.52%	92.54%	20.81%	22.73%	1.99%
0.01	0.2	0.1	260	.53%	89.55%	20.74%	23.11%	2.03%
0.01	0.3	0.1	260	.51%	91.73%	21.53%	24.46%	2.15%
0.01	0.4	0.1	260	.53%	94.03%	21.25%	23.01%	2.02%
0.01	0.5	0.1	260	.54%	94.03%	22.48%	24.98%	2.19%
0.01	0.6	0.1	260	.54%	91.04%	21.75%	22.38%	1.96%
0.01	0.7	0.1	411	.56%	91.36%	22.81%	25.12%	2.2%
0.01	0.8	0.1	388	.56%	91.08%	23.45%	24.23%	2.13%
0.01	0.9	0.1	260	.55%	95.37%	24.19%	26.06%	2.29%
0.01	0.99	0.1	260	.64%	91.04%	24.14%	25.66%	2.25%
0.1	0.01	0.1	260	.55%	92.54%	22.61%	21.71%	1.9%
0.1	0.1	0.1	260	.55%	91.04%	20.31%	20.86%	1.83%
0.1	0.2	0.1	260	.53%	94.03%	21.46%	23.31%	2.04%
0.1	0.3	0.1	260	.52%	92.53%	21.42%	23.83%	2.09%
0.1	0.4	0.1	260	.51%	94.03%	21.77%	24.78%	2.17%
0.1	0.5	0.1	260	.54%	94.03%	21.28%	23.36%	2.05%
0.1	0.6	0.1	260	.53%	89.55%	21.88%	23.4%	2.05%
0.1	0.7	0.1	260	.54%	86.57%	22.24%	24.21%	2.12%
0.1	0.8	0.1	260	.55%	94.03%	22.79%	23.6%	2.07%
0.1	0.9	0.1	260	.57%	90.74%	23.14%	24.64%	2.16%
0.1	0.99	0.1	260	.59%	91.49%	24.7%	26.58%	2.33%
0.2	0.01	0.1	260	.6%	91.04%	30.04%	25.46%	2.23%
0.2	0.1	0.1	260	.59%	91.04%	23.67%	22.43%	1.97%

Training Corpus - PhotoDNA Scorer (with ignorable file hashset) Parameter Tuning Results - All FOI Found (continued)

Default Score	TOI Score	C param	Runs	Min	Max	Avg	Std Dev	Std Err
0.2	0.2	0.1	260	.49%	94.03%	21.44%	23.07%	2.02%
0.2	0.3	0.1	260	.55%	92.54%	21.88%	24.06%	2.11%
0.2	0.4	0.1	260	.55%	92.61%	22.21%	24.11%	2.11%
0.2	0.5	0.1	260	.54%	93.12%	21.31%	22.59%	1.98%
0.2	0.6	0.1	260	.5%	90.45%	21.44%	22.62%	1.98%
0.2	0.7	0.1	260	.56%	91.04%	22.02%	22.33%	1.96%
0.2	0.8	0.1	260	.58%	90.46%	22.9%	23.44%	2.06%
0.2	0.9	0.1	260	.54%	85.07%	22.07%	22.09%	1.94%
0.2	0.99	0.1	260	.61%	89.55%	23.23%	23.86%	2.09%
0.3	0.01	0.1	260	.94%	96.51%	42.24%	28.84%	2.53%
0.3	0.1	0.1	260	.64%	92.54%	30.11%	24.91%	2.18%
0.3	0.2	0.1	260	.6%	88.06%	25.07%	22.72%	1.99%
0.3	0.3	0.1	260	.58%	85.07%	21.4%	20.94%	1.84%
0.3	0.4	0.1	260	.56%	92.54%	22.81%	24.36%	2.14%
0.3	0.5	0.1	260	.55%	91.74%	23.16%	24.89%	2.18%
0.3	0.6	0.1	260	.58%	91.18%	22.41%	23.66%	2.08%
0.3	0.7	0.1	260	.57%	91.54%	22.59%	23.52%	2.06%
0.3	0.8	0.1	260	.58%	92.02%	23.17%	24.84%	2.18%
0.3	0.9	0.1	260	.6%	90.65%	23.75%	23.9%	2.1%
0.3	0.99	0.1	260	.6%	91.04%	24.86%	25.48%	2.23%
0.4	0.01	0.1	260	1.5%	97.28%	51.83%	28.%	2.46%
0.4	0.1	0.1	260	1.05%	91.46%	43.33%	26.53%	2.33%
0.4	0.2	0.1	260	.7%	91.04%	32.59%	24.91%	2.18%
0.4	0.3	0.1	260	.67%	94.03%	26.08%	22.77%	2.%
0.4	0.4	0.1	260	.56%	88.06%	23.29%	22.33%	1.96%
0.4	0.5	0.1	260	.57%	91.16%	23.71%	23.7%	2.08%
0.4	0.6	0.1	260	.58%	80.6%	22.52%	21.74%	1.91%
0.4	0.7	0.1	260	.6%	89.55%	22.6%	22.87%	2.01%
0.4	0.8	0.1	260	.6%	92.54%	23.%	23.66%	2.08%
0.4	0.9	0.1	260	.6%	92.53%	23.78%	24.53%	2.15%
0.4	0.99	0.1	260	.63%	89.55%	24.57%	24.67%	2.16%
0.5	0.01	0.1	263	4.83%	98.52%	60.%	26.17%	2.3%
0.5	0.1	0.1	260	3.87%	98.18%	55.23%	27.64%	2.42%
0.5	0.2	0.1	260	1.54%	96.23%	45.82%	28.87%	2.53%
0.5	0.3	0.1	260	.83%	92.54%	34.15%	24.14%	2.12%
0.5	0.4	0.1	260	.66%	85.07%	28.12%	24.03%	2.11%
0.5	0.5	0.1	260	.63%	92.54%	25.58%	25.54%	2.24%
0.5	0.6	0.1	260	.6%	92.14%	24.13%	24.14%	2.12%

Training Corpus - PhotoDNA Scorer (with ignorable file hashset) Parameter Tuning Results - All FOI Found (continued)

Default Score	TOI Score	C param	Runs	Min	Max	Avg	Std Dev	Std Err
0.5	0.7	0.1	260	.6%	93.75%	23.93%	23.53%	2.06%
0.5	0.8	0.1	260	.62%	92.54%	24.35%	25.%	2.19%
0.5	0.9	0.1	260	.63%	92.54%	23.98%	25.02%	2.19%
0.5	0.99	0.1	260	.62%	89.55%	24.54%	24.45%	2.14%
0.6	0.01	0.1	266	5.37%	99.48%	64.72%	24.23%	2.13%
0.6	0.1	0.1	261	4.54%	98.57%	62.26%	25.94%	2.28%
0.6	0.2	0.1	265	5.32%	99.02%	57.17%	26.76%	2.35%
0.6	0.3	0.1	260	3.5%	96.24%	47.31%	28.19%	2.47%
0.6	0.4	0.1	260	1.13%	95.52%	37.7%	24.14%	2.12%
0.6	0.5	0.1	260	.73%	92.54%	29.56%	25.24%	2.21%
0.6	0.6	0.1	263	.67%	91.13%	27.16%	26.35%	2.31%
0.6	0.7	0.1	361	.6%	94.22%	26.71%	27.24%	2.39%
0.6	0.8	0.1	370	.61%	94.09%	27.02%	28.45%	2.5%
0.6	0.9	0.1	260	.63%	95.52%	26.38%	25.77%	2.26%
0.6	0.99	0.1	260	.65%	92.37%	25.95%	25.3%	2.22%
0.7	0.01	0.1	266	4.97%	99.28%	65.69%	25.24%	2.21%
0.7	0.1	0.1	270	4.93%	99.05%	64.15%	25.46%	2.23%
0.7	0.2	0.1	271	4.87%	99.26%	62.78%	25.38%	2.23%
0.7	0.3	0.1	282	4.97%	96.72%	56.45%	27.17%	2.38%
0.7	0.4	0.1	269	4.9%	95.91%	51.3%	27.11%	2.38%
0.7	0.5	0.1	263	2.61%	96.53%	39.29%	26.16%	2.29%
0.7	0.6	0.1	260	.86%	96.41%	31.84%	26.37%	2.31%
0.7	0.7	0.1	282	.76%	88.05%	28.42%	25.99%	2.28%
0.7	0.8	0.1	273	.69%	93.96%	29.13%	28.4%	2.49%
0.7	0.9	0.1	274	.65%	91.64%	29.22%	27.8%	2.44%
0.7	0.99	0.1	279	.65%	93.15%	27.54%	27.41%	2.4%
0.8	0.01	0.1	280	5.42%	99.42%	67.07%	24.42%	2.14%
0.8	0.1	0.1	278	5.11%	98.75%	64.85%	26.07%	2.29%
0.8	0.2	0.1	280	5.13%	98.36%	64.28%	25.14%	2.2%
0.8	0.3	0.1	282	3.21%	99.04%	59.37%	26.75%	2.35%
0.8	0.4	0.1	277	5.15%	95.52%	59.08%	25.02%	2.19%
0.8	0.5	0.1	285	3.04%	91.23%	51.74%	26.52%	2.33%
0.8	0.6	0.1	286	3.06%	95.85%	46.31%	29.49%	2.59%
0.8	0.7	0.1	280	1.45%	89.55%	35.78%	25.48%	2.23%
0.8	0.8	0.1	287	.85%	92.54%	31.2%	27.28%	2.39%
0.8	0.9	0.1	287	.7%	93.51%	29.24%	28.16%	2.47%
0.8	0.99	0.1	302	.68%	95.43%	28.94%	27.23%	2.39%
0.9	0.01	0.1	270	5.58%	98.3%	66.36%	24.95%	2.19%

Training Corpus - PhotoDNA Scorer (with ignorable file hashset) Parameter Tuning Results - All FOI Found (continued)

Default Score	TOI Score	C param	Runs	Min	Max	Avg	Std Dev	Std Err
0.9	0.1	0.1	270	6.11%	99.%	66.51%	24.95%	2.19%
0.9	0.2	0.1	268	5.72%	99.26%	64.03%	25.87%	2.27%
0.9	0.3	0.1	272	5.96%	98.66%	63.31%	25.8%	2.26%
0.9	0.4	0.1	267	4.98%	96.59%	61.73%	24.77%	2.17%
0.9	0.5	0.1	266	4.92%	97.63%	57.56%	25.46%	2.23%
0.9	0.6	0.1	266	3.35%	95.16%	55.45%	26.54%	2.33%
0.9	0.7	0.1	264	2.93%	97.04%	49.91%	27.33%	2.4%
0.9	0.8	0.1	277	2.81%	95.51%	38.57%	26.13%	2.29%
0.9	0.9	0.1	276	1.08%	94.29%	33.11%	27.82%	2.44%
0.9	0.99	0.1	299	.81%	95.83%	31.09%	28.26%	2.48%
0.99	0.01	0.1	275	5.5%	98.8%	69.03%	23.21%	2.04%
0.99	0.1	0.1	276	4.96%	99.%	67.09%	25.5%	2.24%
0.99	0.2	0.1	283	5.11%	98.94%	66.13%	25.42%	2.23%
0.99	0.3	0.1	282	6.53%	99.05%	66.11%	25.36%	2.22%
0.99	0.4	0.1	285	6.03%	97.68%	63.89%	25.67%	2.25%
0.99	0.5	0.1	285	5.08%	96.99%	61.14%	26.39%	2.31%
0.99	0.6	0.1	289	5.96%	95.42%	53.17%	26.92%	2.36%
0.99	0.7	0.1	303	3.41%	96.75%	52.45%	27.75%	2.43%
0.99	0.8	0.1	312	3.47%	94.03%	48.61%	27.22%	2.39%
0.99	0.9	0.1	320	3.33%	90.96%	40.46%	27.55%	2.42%
0.99	0.99	0.1	329	1.83%	95.79%	35.58%	27.49%	2.41%

D.4 Skin Tone Scorer Parameter Tuning Results

Table D.9: Training Corpus - Skin Tone Scorer (with ignorable file hashset) Parameter Tuning Results - First Hit

Default Score	TOI Score	C param	Runs	Min	Max	Avg	Std Dev	Std Err
0.01	0.01	0.2	260	.07%	19.4%	4.55%	4.4%	.39%
0.01	0.1	0.2	260	.04%	22.14%	4.46%	4.33%	.38%
0.01	0.2	0.2	260	.04%	20.9%	4.21%	4.29%	.38%
0.01	0.3	0.2	260	.01%	34.07%	4.06%	4.84%	.42%
0.01	0.4	0.2	260	.02%	31.76%	4.03%	5.02%	.44%
0.01	0.5	0.2	260	.05%	42.02%	3.91%	4.97%	.44%
0.01	0.6	0.2	260	.02%	64.59%	5.61%	10.23%	.9%
0.01	0.7	0.2	260	.05%	73.64%	5.01%	9.48%	.83%
0.01	0.8	0.2	260	.03%	77.67%	7.39%	15.54%	1.36%
0.01	0.9	0.2	260	.01%	74.65%	5.68%	12.12%	1.06%
0.01	0.99	0.2	260	.06%	77.06%	6.85%	15.92%	1.4%
0.1	0.01	0.2	260	.06%	23.88%	4.66%	5.01%	.44%
0.1	0.1	0.2	260	.08%	32.84%	4.64%	5.36%	.47%
0.1	0.2	0.2	260	.02%	23.88%	4.41%	4.41%	.39%
0.1	0.3	0.2	260	.02%	14.4%	3.91%	3.63%	.32%
0.1	0.4	0.2	260	.05%	28.59%	4.3%	4.88%	.43%
0.1	0.5	0.2	260	.07%	56.54%	4.27%	6.7%	.59%
0.1	0.6	0.2	260	.02%	66.4%	5.43%	11.44%	1.%
0.1	0.7	0.2	260	.04%	73.04%	6.08%	10.64%	.93%
0.1	0.8	0.2	260	.06%	75.65%	8.23%	17.73%	1.56%
0.1	0.9	0.2	260	.07%	76.86%	7.56%	15.28%	1.34%
0.1	0.99	0.2	260	.01%	75.65%	8.25%	17.99%	1.58%
0.2	0.01	0.2	260	.05%	28.36%	5.34%	6.31%	.55%
0.2	0.1	0.2	260	.1%	25.37%	5.2%	5.08%	.45%
0.2	0.2	0.2	260	.04%	14.94%	3.87%	3.91%	.34%
0.2	0.3	0.2	260	.05%	25.37%	4.51%	4.69%	.41%
0.2	0.4	0.2	260	.1%	15.69%	3.66%	3.31%	.29%
0.2	0.5	0.2	260	.06%	40.85%	6.%	8.57%	.75%
0.2	0.6	0.2	260	.06%	68.01%	7.31%	14.43%	1.27%
0.2	0.7	0.2	260	.07%	75.86%	8.03%	16.31%	1.43%
0.2	0.8	0.2	260	.04%	76.46%	7.69%	15.88%	1.39%
0.2	0.9	0.2	260	.05%	76.26%	8.84%	18.16%	1.59%
0.2	0.99	0.2	260	.03%	74.45%	8.53%	17.71%	1.55%
0.3	0.01	0.2	260	.1%	33.03%	6.08%	7.85%	.69%
0.3	0.1	0.2	260	.06%	34.33%	5.6%	6.64%	.58%
0.3	0.2	0.2	260	.04%	17.41%	4.41%	4.57%	.4%

*Training Corpus - Skin Tone Scorer (with ignorable file hashset) Parameter Tuning Results
- First Hit (continued)*

Default Score	TOI Score	C param	Runs	Min	Max	Avg	Std Dev	Std Err
0.3	0.3	0.2	260	.05%	17.91%	4.31%	4.3%	.38%
0.3	0.4	0.2	260	.07%	16.42%	4.63%	4.54%	.4%
0.3	0.5	0.2	260	.08%	43.86%	6.78%	9.95%	.87%
0.3	0.6	0.2	260	.03%	70.82%	8.51%	16.27%	1.43%
0.3	0.7	0.2	260	.04%	76.06%	9.35%	18.88%	1.66%
0.3	0.8	0.2	260	.%	77.06%	8.86%	18.5%	1.62%
0.3	0.9	0.2	260	.%	77.46%	9.47%	19.73%	1.73%
0.3	0.99	0.2	260	.04%	77.87%	9.76%	19.54%	1.71%
0.4	0.01	0.2	260	.07%	43.06%	7.12%	9.42%	.83%
0.4	0.1	0.2	260	.02%	36.71%	7.17%	8.57%	.75%
0.4	0.2	0.2	260	.07%	26.97%	6.26%	6.82%	.6%
0.4	0.3	0.2	260	.09%	19.4%	5.51%	5.31%	.47%
0.4	0.4	0.2	260	.04%	26.87%	6.01%	6.16%	.54%
0.4	0.5	0.2	260	.03%	51.31%	7.59%	11.09%	.97%
0.4	0.6	0.2	260	.05%	76.06%	9.36%	18.02%	1.58%
0.4	0.7	0.2	260	.%	75.25%	9.54%	19.13%	1.68%
0.4	0.8	0.2	260	.04%	77.67%	8.86%	18.36%	1.61%
0.4	0.9	0.2	260	.04%	77.67%	8.94%	17.72%	1.55%
0.4	0.99	0.2	260	.04%	76.06%	9.13%	18.5%	1.62%
0.5	0.01	0.2	260	.01%	45.41%	9.35%	12.32%	1.08%
0.5	0.1	0.2	260	.02%	41.95%	9.19%	12.26%	1.08%
0.5	0.2	0.2	260	.1%	37.35%	9.43%	11.53%	1.01%
0.5	0.3	0.2	260	.01%	41.85%	8.22%	9.88%	.87%
0.5	0.4	0.2	260	.11%	52.52%	8.78%	11.84%	1.04%
0.5	0.5	0.2	260	.06%	76.46%	10.16%	16.93%	1.48%
0.5	0.6	0.2	260	.08%	74.85%	10.43%	19.08%	1.67%
0.5	0.7	0.2	260	.01%	75.65%	10.22%	19.07%	1.67%
0.5	0.8	0.2	260	.06%	79.28%	10.49%	19.37%	1.7%
0.5	0.9	0.2	260	.%	76.66%	10.42%	19.29%	1.69%
0.5	0.99	0.2	260	.06%	76.26%	9.96%	19.39%	1.7%
0.6	0.01	0.2	260	.08%	72.81%	11.58%	16.06%	1.41%
0.6	0.1	0.2	260	.12%	48.29%	11.%	14.34%	1.26%
0.6	0.2	0.2	260	.04%	51.71%	10.67%	13.95%	1.22%
0.6	0.3	0.2	260	.13%	56.34%	10.85%	14.47%	1.27%
0.6	0.4	0.2	260	.07%	76.06%	11.23%	18.16%	1.59%
0.6	0.5	0.2	260	.02%	76.46%	10.97%	19.23%	1.69%
0.6	0.6	0.2	260	.04%	77.87%	11.11%	19.4%	1.7%
0.6	0.7	0.2	260	.05%	75.05%	10.32%	18.18%	1.59%

*Training Corpus - Skin Tone Scorer (with ignorable file hashset) Parameter Tuning Results
- First Hit (continued)*

Default Score	TOI Score	C param	Runs	Min	Max	Avg	Std Dev	Std Err
0.6	0.8	0.2	260	.06%	76.06%	10.27%	19.25%	1.69%
0.6	0.9	0.2	260	.04%	75.05%	10.12%	19.15%	1.68%
0.6	0.99	0.2	260	.04%	75.86%	10.42%	19.19%	1.68%
0.7	0.01	0.2	260	.17%	75.29%	12.76%	18.52%	1.62%
0.7	0.1	0.2	260	.09%	72.86%	12.16%	18.12%	1.59%
0.7	0.2	0.2	260	.08%	59.36%	11.88%	16.95%	1.49%
0.7	0.3	0.2	260	.02%	71.83%	11.52%	17.91%	1.57%
0.7	0.4	0.2	260	.01%	75.05%	11.92%	18.99%	1.67%
0.7	0.5	0.2	260	.04%	75.25%	11.19%	18.42%	1.62%
0.7	0.6	0.2	260	.01%	76.86%	11.19%	19.3%	1.69%
0.7	0.7	0.2	260	.04%	75.65%	11.13%	19.24%	1.69%
0.7	0.8	0.2	260	.05%	76.06%	10.78%	19.27%	1.69%
0.7	0.9	0.2	260	.11%	76.46%	10.63%	19.36%	1.7%
0.7	0.99	0.2	260	.02%	74.85%	9.94%	18.27%	1.6%
0.8	0.01	0.2	260	.17%	76.23%	15.14%	21.44%	1.88%
0.8	0.1	0.2	260	.06%	75.07%	14.01%	19.66%	1.72%
0.8	0.2	0.2	260	.08%	74.04%	13.62%	21.09%	1.85%
0.8	0.3	0.2	260	.08%	75.86%	13.16%	19.72%	1.73%
0.8	0.4	0.2	260	.05%	75.05%	12.15%	20.59%	1.81%
0.8	0.5	0.2	260	.15%	75.65%	12.49%	20.3%	1.78%
0.8	0.6	0.2	260	.02%	77.46%	11.69%	19.48%	1.71%
0.8	0.7	0.2	260	.06%	77.06%	11.41%	19.38%	1.7%
0.8	0.8	0.2	260	.11%	75.25%	10.51%	19.16%	1.68%
0.8	0.9	0.2	260	.04%	75.65%	10.88%	19.06%	1.67%
0.8	0.99	0.2	260	.06%	75.86%	10.39%	19.13%	1.68%
0.9	0.01	0.2	260	.18%	76.44%	16.54%	21.24%	1.86%
0.9	0.1	0.2	260	.03%	75.86%	15.67%	22.23%	1.95%
0.9	0.2	0.2	260	.12%	77.51%	15.33%	22.48%	1.97%
0.9	0.3	0.2	260	.02%	75.25%	14.74%	23.75%	2.08%
0.9	0.4	0.2	260	.14%	76.66%	13.92%	21.32%	1.87%
0.9	0.5	0.2	260	.04%	78.27%	12.96%	20.81%	1.82%
0.9	0.6	0.2	260	.04%	77.06%	12.66%	20.35%	1.78%
0.9	0.7	0.2	260	.1%	76.46%	12.96%	20.3%	1.78%
0.9	0.8	0.2	260	.03%	76.06%	11.34%	19.65%	1.72%
0.9	0.9	0.2	260	.02%	78.47%	11.44%	19.27%	1.69%
0.9	0.99	0.2	260	.08%	76.66%	11.09%	19.33%	1.69%
0.99	0.01	0.2	260	.12%	74.95%	17.99%	23.59%	2.07%
0.99	0.1	0.2	260	.01%	76.95%	15.9%	23.46%	2.06%

Training Corpus - Skin Tone Scorer (with ignorable file hashset) Parameter Tuning Results - First Hit (continued)

Default Score	TOI Score	C param	Runs	Min	Max	Avg	Std Dev	Std Err
0.99	0.2	0.2	260	.02%	76.66%	15.97%	23.84%	2.09%
0.99	0.3	0.2	260	.08%	76.74%	17.22%	25.22%	2.21%
0.99	0.4	0.2	260	.02%	78.47%	15.2%	23.95%	2.1%
0.99	0.5	0.2	260	.12%	75.25%	14.13%	22.06%	1.93%
0.99	0.6	0.2	260	.12%	76.06%	13.12%	20.49%	1.8%
0.99	0.7	0.2	260	.02%	75.45%	12.42%	20.2%	1.77%
0.99	0.8	0.2	260	.11%	76.06%	12.83%	20.32%	1.78%
0.99	0.9	0.2	260	.11%	75.25%	12.25%	19.66%	1.72%
0.99	0.99	0.2	260	.03%	77.26%	11.61%	19.43%	1.7%

Table D.10: Training Corpus - Skin Tone Scorer (with ignorable file hashset) Parameter Tuning Results - All Found

Default Score	TOI Score	C param	Runs	Min	Max	Avg	Std Dev	Std Err
0.01	0.01	0.1	260	.6%	97.01%	27.43%	26.15%	2.29%
0.01	0.1	0.1	260	.64%	98.51%	28.21%	26.77%	2.35%
0.01	0.2	0.1	260	.62%	98.51%	29.5%	28.26%	2.48%
0.01	0.3	0.1	260	.67%	98.51%	30.04%	29.71%	2.61%
0.01	0.4	0.1	260	.65%	98.51%	31.26%	30.21%	2.65%
0.01	0.5	0.1	260	.67%	98.51%	33.05%	32.12%	2.82%
0.01	0.6	0.1	260	.69%	98.51%	32.26%	30.83%	2.7%
0.01	0.7	0.1	390	.72%	98.51%	32.65%	31.26%	2.74%
0.01	0.8	0.1	362	.7%	98.51%	34.2%	32.79%	2.88%
0.01	0.9	0.1	260	.74%	98.59%	35.72%	33.9%	2.97%
0.01	0.99	0.1	260	.73%	98.51%	34.17%	32.84%	2.88%
0.1	0.01	0.1	260	.62%	97.01%	28.29%	27.53%	2.41%
0.1	0.1	0.1	260	.63%	98.51%	27.82%	26.87%	2.36%
0.1	0.2	0.1	260	.63%	98.51%	28.19%	26.97%	2.37%
0.1	0.3	0.1	260	.64%	98.51%	29.87%	29.39%	2.58%
0.1	0.4	0.1	260	.64%	98.51%	30.8%	30.17%	2.65%
0.1	0.5	0.1	260	.66%	98.51%	31.96%	30.84%	2.7%
0.1	0.6	0.1	260	.68%	97.01%	32.32%	31.37%	2.75%
0.1	0.7	0.1	260	.68%	98.51%	31.71%	30.23%	2.65%
0.1	0.8	0.1	260	.67%	98.51%	33.6%	32.44%	2.84%
0.1	0.9	0.1	260	.74%	98.51%	33.06%	31.85%	2.79%
0.1	0.99	0.1	260	.76%	98.51%	33.18%	31.6%	2.77%
0.2	0.01	0.1	260	.63%	98.51%	28.19%	26.77%	2.35%
0.2	0.1	0.1	260	.63%	98.51%	27.91%	27.28%	2.39%

*Training Corpus - Skin Tone Scorer (with ignorable file hashset) Parameter Tuning Results
- All Found (continued)*

Default Score	TOI Score	C param	Runs	Min	Max	Avg	Std Dev	Std Err
0.2	0.2	0.1	260	.66%	98.51%	26.69%	26.06%	2.29%
0.2	0.3	0.1	260	.65%	98.51%	28.69%	27.66%	2.43%
0.2	0.4	0.1	260	.67%	98.51%	29.83%	28.87%	2.53%
0.2	0.5	0.1	260	.69%	98.51%	31.39%	30.7%	2.69%
0.2	0.6	0.1	260	.71%	98.51%	31.23%	29.96%	2.63%
0.2	0.7	0.1	260	.74%	98.51%	32.37%	30.77%	2.7%
0.2	0.8	0.1	260	.71%	98.51%	33.19%	31.75%	2.78%
0.2	0.9	0.1	260	.75%	97.01%	32.6%	31.3%	2.75%
0.2	0.99	0.1	260	.74%	98.51%	34.25%	32.55%	2.85%
0.3	0.01	0.1	260	.63%	95.52%	30.14%	26.39%	2.31%
0.3	0.1	0.1	260	.64%	97.01%	28.6%	26.67%	2.34%
0.3	0.2	0.1	260	.65%	98.51%	27.24%	25.55%	2.24%
0.3	0.3	0.1	260	.67%	98.51%	27.88%	26.9%	2.36%
0.3	0.4	0.1	260	.68%	98.51%	29.25%	27.07%	2.37%
0.3	0.5	0.1	260	.65%	98.51%	30.39%	28.91%	2.54%
0.3	0.6	0.1	260	.7%	97.01%	30.69%	29.03%	2.55%
0.3	0.7	0.1	260	.73%	98.51%	32.01%	30.87%	2.71%
0.3	0.8	0.1	260	.73%	98.51%	32.84%	31.37%	2.75%
0.3	0.9	0.1	260	.75%	98.51%	33.16%	31.74%	2.78%
0.3	0.99	0.1	260	.75%	98.51%	33.69%	30.87%	2.71%
0.4	0.01	0.1	260	.7%	98.51%	32.85%	28.87%	2.53%
0.4	0.1	0.1	260	.67%	98.51%	29.73%	26.41%	2.32%
0.4	0.2	0.1	260	.68%	98.51%	29.09%	26.11%	2.29%
0.4	0.3	0.1	260	.68%	98.51%	28.24%	25.69%	2.25%
0.4	0.4	0.1	260	.68%	98.51%	28.91%	26.8%	2.35%
0.4	0.5	0.1	260	.68%	98.51%	29.66%	26.83%	2.35%
0.4	0.6	0.1	260	.72%	98.51%	31.55%	29.02%	2.55%
0.4	0.7	0.1	260	.72%	98.51%	31.22%	28.86%	2.53%
0.4	0.8	0.1	260	.76%	98.51%	32.63%	30.59%	2.68%
0.4	0.9	0.1	260	.76%	98.51%	33.49%	31.%	2.72%
0.4	0.99	0.1	260	.76%	98.51%	32.92%	30.59%	2.68%
0.5	0.01	0.1	260	.79%	98.51%	36.61%	30.32%	2.66%
0.5	0.1	0.1	260	.75%	97.01%	33.27%	27.64%	2.42%
0.5	0.2	0.1	260	.74%	98.51%	31.62%	26.66%	2.34%
0.5	0.3	0.1	260	.71%	98.51%	30.83%	25.77%	2.26%
0.5	0.4	0.1	260	.7%	98.51%	28.66%	26.04%	2.28%
0.5	0.5	0.1	260	.7%	98.51%	29.97%	27.02%	2.37%
0.5	0.6	0.1	260	.72%	98.51%	31.56%	28.02%	2.46%

*Training Corpus - Skin Tone Scorer (with ignorable file hashset) Parameter Tuning Results
- All Found (continued)*

Default Score	TOI Score	C param	Runs	Min	Max	Avg	Std Dev	Std Err
0.5	0.7	0.1	260	.75%	98.51%	32.47%	29.3%	2.57%
0.5	0.8	0.1	260	.76%	98.51%	32.8%	29.63%	2.6%
0.5	0.9	0.1	260	.74%	98.51%	33.49%	30.9%	2.71%
0.5	0.99	0.1	260	.77%	98.51%	33.53%	30.42%	2.67%
0.6	0.01	0.1	260	1.59%	98.51%	41.79%	29.05%	2.55%
0.6	0.1	0.1	264	1.06%	98.51%	39.93%	30.8%	2.7%
0.6	0.2	0.1	260	.86%	98.51%	36.07%	30.1%	2.64%
0.6	0.3	0.1	260	.79%	98.51%	32.54%	27.68%	2.43%
0.6	0.4	0.1	260	.75%	98.51%	32.14%	26.8%	2.35%
0.6	0.5	0.1	260	.73%	98.51%	30.48%	26.68%	2.34%
0.6	0.6	0.1	260	.75%	98.51%	31.61%	26.99%	2.37%
0.6	0.7	0.1	361	.78%	98.51%	32.68%	27.58%	2.42%
0.6	0.8	0.1	372	.76%	98.51%	34.58%	30.1%	2.64%
0.6	0.9	0.1	260	.79%	98.51%	33.42%	30.57%	2.68%
0.6	0.99	0.1	260	.79%	98.51%	34.32%	31.%	2.72%
0.7	0.01	0.1	260	4.92%	98.51%	45.09%	27.49%	2.41%
0.7	0.1	0.1	260	4.23%	98.51%	42.79%	28.02%	2.46%
0.7	0.2	0.1	262	2.56%	98.51%	41.28%	28.7%	2.52%
0.7	0.3	0.1	270	1.29%	99.%	38.7%	29.17%	2.56%
0.7	0.4	0.1	260	.93%	97.01%	35.64%	28.21%	2.47%
0.7	0.5	0.1	265	.88%	98.51%	33.06%	26.4%	2.32%
0.7	0.6	0.1	263	.82%	98.51%	33.54%	27.14%	2.38%
0.7	0.7	0.1	273	.8%	98.51%	32.93%	27.54%	2.42%
0.7	0.8	0.1	277	.83%	98.51%	34.64%	29.29%	2.57%
0.7	0.9	0.1	274	.78%	98.51%	36.13%	31.38%	2.75%
0.7	0.99	0.1	276	.83%	97.01%	34.81%	31.24%	2.74%
0.8	0.01	0.1	283	3.95%	98.51%	49.19%	26.6%	2.33%
0.8	0.1	0.1	284	2.68%	97.84%	46.32%	26.15%	2.29%
0.8	0.2	0.1	296	5.43%	97.01%	45.%	26.37%	2.31%
0.8	0.3	0.1	287	3.01%	98.51%	44.28%	28.14%	2.47%
0.8	0.4	0.1	290	2.71%	98.51%	43.79%	27.84%	2.44%
0.8	0.5	0.1	295	2.4%	98.13%	38.3%	26.76%	2.35%
0.8	0.6	0.1	287	1.16%	98.51%	36.3%	28.51%	2.5%
0.8	0.7	0.1	288	.98%	98.51%	34.05%	27.4%	2.4%
0.8	0.8	0.1	278	.91%	97.01%	32.12%	26.6%	2.33%
0.8	0.9	0.1	276	.87%	98.51%	35.72%	29.72%	2.61%
0.8	0.99	0.1	274	.85%	98.51%	37.82%	33.01%	2.89%
0.9	0.01	0.1	278	3.49%	98.51%	51.%	26.5%	2.32%

*Training Corpus - Skin Tone Scorer (with ignorable file hashset) Parameter Tuning Results
- All Found (continued)*

Default Score	TOI Score	C param	Runs	Min	Max	Avg	Std Dev	Std Err
0.9	0.1	0.1	285	5.3%	98.67%	49.17%	26.46%	2.32%
0.9	0.2	0.1	272	4.09%	99.42%	49.46%	27.2%	2.39%
0.9	0.3	0.1	283	3.07%	98.51%	47.35%	28.89%	2.53%
0.9	0.4	0.1	277	3.87%	98.88%	47.34%	26.56%	2.33%
0.9	0.5	0.1	274	3.86%	98.51%	44.64%	27.43%	2.41%
0.9	0.6	0.1	280	3.26%	98.51%	42.52%	26.04%	2.28%
0.9	0.7	0.1	297	2.83%	97.01%	39.27%	26.55%	2.33%
0.9	0.8	0.1	292	1.6%	98.51%	38.87%	28.81%	2.53%
0.9	0.9	0.1	279	1.04%	98.51%	35.89%	27.52%	2.41%
0.9	0.99	0.1	285	.98%	98.51%	39.61%	31.19%	2.74%
0.99	0.01	0.1	274	3.93%	99.96%	50.08%	25.7%	2.25%
0.99	0.1	0.1	277	6.12%	99.71%	53.54%	27.08%	2.38%
0.99	0.2	0.1	278	3.97%	98.51%	50.23%	26.27%	2.3%
0.99	0.3	0.1	277	6.7%	98.51%	49.6%	28.42%	2.49%
0.99	0.4	0.1	272	3.93%	98.51%	50.37%	27.28%	2.39%
0.99	0.5	0.1	307	4.01%	98.25%	49.48%	25.87%	2.27%
0.99	0.6	0.1	290	6.68%	98.51%	47.31%	25.53%	2.24%
0.99	0.7	0.1	309	3.52%	98.51%	46.35%	25.19%	2.21%
0.99	0.8	0.1	291	3.08%	98.51%	42.35%	25.28%	2.22%
0.99	0.9	0.1	298	3.17%	98.51%	42.39%	28.58%	2.51%
0.99	0.99	0.1	319	2.26%	98.51%	40.48%	28.78%	2.52%

D.5 Best First Results

Table D.11: Best First Test Results - First File Of Interest Found

Metadata Scorer	Tokenizer	Run Count	Min	Max	Avg	Std Dev	Std Err
Cosine Similarity	Non word character splitter (with stemming)	260	.6%	61.96%	20.97%	22.45%	1.97%
Cosine Similarity	2-3gram (fold case)	260	.32%	73.62%	26.36%	24.83%	2.18%
Cosine Similarity	2-4gram (fold case)	260	.32%	73.62%	26.37%	24.83%	2.18%
Log Regression Feature Scorer		780	.39%	92.13%	27.72%	28.5%	2.5%
Multinomial Naive Bayes	Non word character splitter (with stemming)	260	1.41%	96.32%	45.8%	39.14%	3.43%
Multinomial Naive Bayes	2-4gram (fold case)	260	.96%	94.11%	47.32%	31.87%	2.8%
Multinomial Naive Bayes	2-3gram (fold case)	260	.96%	94.11%	47.32%	31.88%	2.8%

Table D.12: Best First Test Results - All Files Of Interest Found

Metadata Scorer	Tokenizer	Run Count	Min	Max	Avg	Std Dev	Std Err
Cosine Similarity	2-3gram (fold case)	260	1.53%	99.14%	50.3%	29.94%	2.63%
Cosine Similarity	2-4gram (fold case)	260	1.53%	99.14%	50.29%	29.94%	2.63%
Cosine Similarity	Non word character splitter (with stemming)	260	2.42%	97.37%	43.77%	29.13%	2.56%
Log Regression Feature Scorer		780	12.86%	99.53%	61.35%	31.87%	2.8%
Multinomial Naive Bayes	2-3gram (fold case)	260	4.93%	98.93%	72.43%	27.18%	2.38%
Multinomial Naive Bayes	2-4gram (fold case)	260	4.93%	98.93%	72.44%	27.19%	2.38%
Multinomial Naive Bayes	Non word character splitter (with stemming)	260	6.96%	99.03%	67.96%	29.77%	2.61%

D.6 Test Corpus C_p Tuning Results

Table D.13: Test Corpus C_p tuning - MD5 Scorer, First FOI found

Scorer	C_p	Runs	Min	Max	Avg	Std Dev	Std Err
MD5 Scorer	0.01	977	.47%	98.77%	33.64%	29.32%	3.28%
MD5 Scorer	0.1	687	.19%	45.52%	14.29%	14.84%	1.66%
MD5 Scorer	0.2	80	.19%	48.8%	13.4%	15.81%	1.77%
MD5 Scorer	0.4	80	.23%	47.74%	12.52%	15.19%	1.7%
MD5 Scorer	0.6	80	.%	52.36%	13.12%	16.59%	1.86%
MD5 Scorer	$\frac{1}{\sqrt{2}}$	80	.28%	57.07%	12.71%	16.14%	1.8%
MD5 Scorer	0.8	80	.05%	50.91%	13.4%	16.74%	1.87%
MD5 Scorer	1	80	.03%	51.2%	13.37%	16.55%	1.85%
MD5 Scorer	1.2	80	.09%	47.26%	12.85%	16.02%	1.79%
MD5 Scorer	1.4	80	.12%	59.29%	14.%	17.96%	2.01%
MD5 Scorer	1.6	80	.04%	61.89%	13.51%	17.42%	1.95%
MD5 Scorer	1.8	80	.09%	57.75%	13.38%	17.53%	1.96%
MD5 Scorer	2	80	.26%	53.22%	13.12%	17.04%	1.9%
MD5 Scorer (no ignorable files)	0.01	725	.32%	54.38%	14.94%	17.74%	1.98%
MD5 Scorer (no ignorable files)	0.1	772	.36%	57.27%	15.04%	18.7%	2.09%
MD5 Scorer (no ignorable files)	0.2	80	.03%	53.99%	13.73%	17.04%	1.91%
MD5 Scorer (no ignorable files)	0.4	80	.36%	59.38%	14.19%	18.15%	2.03%
MD5 Scorer (no ignorable files)	0.6	80	.09%	58.81%	14.98%	18.83%	2.11%
MD5 Scorer (no ignorable files)	$\frac{1}{\sqrt{2}}$	80	.02%	56.3%	13.82%	17.06%	1.91%
MD5 Scorer (no ignorable files)	0.8	80	.06%	57.07%	13.86%	17.09%	1.91%
MD5 Scorer (no ignorable files)	1	80	.07%	58.81%	14.33%	17.75%	1.98%
MD5 Scorer (no ignorable files)	1.2	80	.03%	58.52%	15.28%	18.88%	2.11%
MD5 Scorer (no ignorable files)	1.4	80	.3%	61.6%	15.21%	19.03%	2.13%
MD5 Scorer (no ignorable files)	1.6	80	.17%	62.66%	14.33%	17.97%	2.01%
MD5 Scorer (no ignorable files)	1.8	80	.22%	57.56%	14.23%	18.37%	2.05%
MD5 Scorer (no ignorable files)	2	80	.26%	53.99%	14.13%	17.41%	1.95%

Table D.14: Test Corpus C_p tuning - MD5 Scorer, All FOI found

Scorer	C_p	Runs	Min	Max	Avg	Std Dev	Std Err
MD5 Scorer	0.01	977	3.23%	99.37%	47.56%	31.69%	3.54%
MD5 Scorer	0.1	687	4.3%	54.57%	24.88%	16.54%	1.85%
MD5 Scorer	0.2	80	10.67%	55.92%	26.57%	14.64%	1.64%
MD5 Scorer	0.4	80	12.5%	58.71%	29.55%	14.33%	1.6%
MD5 Scorer	0.6	80	9.38%	59.77%	29.57%	14.85%	1.66%
MD5 Scorer	$\frac{1}{\sqrt{2}}$	80	12.46%	60.92%	29.75%	14.86%	1.66%
MD5 Scorer	0.8	80	10.81%	61.41%	30.26%	15.43%	1.72%
MD5 Scorer	1	80	10.28%	61.89%	31.13%	15.54%	1.74%
MD5 Scorer	1.2	80	9.55%	63.04%	31.37%	15.98%	1.79%
MD5 Scorer	1.4	80	11.58%	63.43%	32.22%	15.95%	1.78%
MD5 Scorer	1.6	80	12.16%	63.81%	32.4%	15.48%	1.73%
MD5 Scorer	1.8	80	10.92%	64.77%	33.04%	15.93%	1.78%
MD5 Scorer	2	80	10.5%	65.64%	32.8%	15.56%	1.74%
MD5 Scorer (no ignorable files)	0.01	725	3.53%	59.48%	28.42%	19.89%	2.22%
MD5 Scorer (no ignorable files)	0.1	772	6.8%	62.18%	33.31%	19.39%	2.17%
MD5 Scorer (no ignorable files)	0.2	80	12.23%	64.49%	36.07%	18.16%	2.03%
MD5 Scorer (no ignorable files)	0.4	80	11.76%	66.31%	38.05%	16.65%	1.86%
MD5 Scorer (no ignorable files)	0.6	80	11.02%	67.76%	39.12%	17.25%	1.93%
MD5 Scorer (no ignorable files)	$\frac{1}{\sqrt{2}}$	80	13.91%	68.05%	39.36%	16.88%	1.89%
MD5 Scorer (no ignorable files)	0.8	80	10.93%	67.85%	38.85%	16.27%	1.82%
MD5 Scorer (no ignorable files)	1	80	12.3%	67.28%	39.73%	16.39%	1.83%
MD5 Scorer (no ignorable files)	1.2	80	13.31%	68.14%	40.29%	16.54%	1.85%
MD5 Scorer (no ignorable files)	1.4	80	12.63%	67.66%	40.16%	16.65%	1.86%
MD5 Scorer (no ignorable files)	1.6	80	12.22%	68.43%	40.28%	16.2%	1.81%
MD5 Scorer (no ignorable files)	1.8	80	15.66%	68.62%	40.06%	15.96%	1.78%
MD5 Scorer (no ignorable files)	2	80	15.43%	68.62%	39.68%	15.16%	1.7%

D.7 Test Corpus MD5 Scorer Parameter Tuning Results

Table D.15: Test Corpus - MD5 Scorer (with ignorable file hashset) Parameter Tuning Results - First Hit

Default Score	TOI Score	C param	Runs	Min	Max	Avg	Std Dev	Std Err

Table D.16: Test Corpus - MD5 Scorer (without ignorable file hashset) Parameter Tuning Results - First Hit

Default Score	TOI Score	C_p	Runs	Min	Max	Avg	Std Dev	Std Err

Table D.17: Test Corpus - MD5 Scorer (with ignorable file hashset) Parameter Tuning Results - All FOI Found

Default Score	TOI Score	C_p	Runs	Min	Max	Avg	Std Dev	Std Err
0.01	0.01	0.2	80	9.01%	61.6%	31.65%	17.85%	2.0%
0.01	0.1	0.2	80	2.84%	65.64%	23.8%	21.31%	2.38%
0.01	0.2	0.2	80	1.61%	65.35%	22.66%	24.22%	2.71%
0.01	0.3	0.2	100	1.36%	62.95%	20.81%	23.06%	2.58%
0.01	0.4	0.2	80	1.0%	60.83%	19.59%	22.57%	2.52%
0.01	0.5	0.2	85	1.19%	57.56%	19.25%	21.58%	2.41%
0.01	0.6	0.2	80	.99%	57.27%	19.29%	20.96%	2.34%
0.01	0.7	0.2	80	.96%	56.4%	19.55%	20.98%	2.35%
0.01	0.8	0.2	86	1.05%	55.92%	19.79%	20.25%	2.26%
0.01	0.9	0.2	85	.89%	55.53%	19.77%	20.02%	2.24%
0.01	0.99	0.2	85	.92%	56.4%	20.1%	20.15%	2.25%
0.1	0.01	0.2	80	10.03%	77.09%	46.54%	19.54%	2.18%
0.1	0.1	0.2	80	9.48%	57.56%	25.58%	15.49%	1.73%
0.1	0.2	0.2	80	2.48%	61.5%	22.55%	21.23%	2.37%
0.1	0.3	0.2	80	1.46%	63.72%	21.64%	23.53%	2.63%
0.1	0.4	0.2	80	1.32%	61.02%	20.34%	22.65%	2.53%
0.1	0.5	0.2	80	1.2%	58.52%	19.21%	22.17%	2.48%
0.1	0.6	0.2	80	1.1%	58.13%	19.24%	21.92%	2.45%
0.1	0.7	0.2	80	1.06%	56.4%	19.03%	21.13%	2.36%
0.1	0.8	0.2	80	.98%	55.15%	19.29%	20.46%	2.29%
0.1	0.9	0.2	80	.89%	58.04%	19.75%	20.73%	2.32%
0.1	0.99	0.2	80	.99%	54.67%	19.76%	20.26%	2.27%
0.2	0.01	0.2	80	15.09%	90.51%	58.12%	23.62%	2.64%
0.2	0.1	0.2	80	15.22%	72.28%	46.07%	19.5%	2.18%
0.2	0.2	0.2	80	10.0%	56.3%	25.9%	15.78%	1.76%

Test Corpus - MD5 Scorer (with ignorable file hashset) Parameter Tuning Results - All FOI Found (continued)

Default Score	TOI Score	C_p	Runs	Min	Max	Avg	Std Dev	Std Err
0.2	0.3	0.2	80	2.12%	61.21%	21.53%	20.3%	2.27%
0.2	0.4	0.2	80	1.38%	62.95%	20.83%	23.09%	2.58%
0.2	0.5	0.2	80	1.16%	60.92%	19.8%	22.68%	2.54%
0.2	0.6	0.2	80	1.13%	58.33%	19.09%	21.79%	2.44%
0.2	0.7	0.2	80	1.19%	57.46%	19.17%	21.49%	2.4%
0.2	0.8	0.2	80	.96%	57.27%	18.88%	20.72%	2.32%
0.2	0.9	0.2	80	1.05%	55.44%	18.96%	20.61%	2.3%
0.2	0.99	0.2	80	1.11%	56.79%	18.91%	20.15%	2.25%
0.3	0.01	0.2	80	28.23%	90.48%	61.95%	20.59%	2.3%
0.3	0.1	0.2	80	21.6%	88.5%	55.85%	22.88%	2.56%
0.3	0.2	0.2	80	16.33%	71.22%	43.92%	16.87%	1.89%
0.3	0.3	0.2	80	9.31%	55.92%	26.29%	15.41%	1.72%
0.3	0.4	0.2	80	2.1%	61.69%	21.36%	21.1%	2.36%
0.3	0.5	0.2	80	1.45%	62.18%	20.35%	23.12%	2.58%
0.3	0.6	0.2	80	1.29%	60.83%	19.3%	22.82%	2.55%
0.3	0.7	0.2	80	1.28%	58.33%	19.05%	22.19%	2.48%
0.3	0.8	0.2	80	1.02%	57.65%	19.04%	21.32%	2.38%
0.3	0.9	0.2	80	1.14%	57.27%	19.42%	21.31%	2.38%
0.3	0.99	0.2	80	.95%	55.53%	19.01%	20.45%	2.29%
0.4	0.01	0.2	80	33.38%	90.27%	64.92%	19.1%	2.14%
0.4	0.1	0.2	80	24.57%	88.82%	57.23%	20.94%	2.34%
0.4	0.2	0.2	80	22.66%	87.52%	53.8%	21.45%	2.4%
0.4	0.3	0.2	80	14.24%	70.64%	43.48%	16.65%	1.86%
0.4	0.4	0.2	80	7.95%	56.11%	25.96%	14.87%	1.66%
0.4	0.5	0.2	80	2.08%	60.35%	20.3%	19.77%	2.21%
0.4	0.6	0.2	80	1.33%	62.37%	20.33%	23.48%	2.63%
0.4	0.7	0.2	80	1.2%	59.77%	19.04%	22.48%	2.51%
0.4	0.8	0.2	80	.99%	58.33%	18.93%	21.93%	2.45%
0.4	0.9	0.2	80	1.21%	57.65%	18.87%	21.2%	2.37%
0.4	0.99	0.2	80	1.11%	56.59%	19.27%	20.86%	2.33%
0.5	0.01	0.2	80	37.8%	92.05%	70.02%	17.35%	1.94%
0.5	0.1	0.2	80	25.77%	89.9%	59.55%	18.86%	2.11%
0.5	0.2	0.2	80	25.04%	87.9%	55.27%	21.25%	2.38%
0.5	0.3	0.2	80	23.17%	81.72%	52.07%	20.33%	2.27%
0.5	0.4	0.2	80	19.13%	70.07%	42.29%	16.28%	1.82%
0.5	0.5	0.2	80	11.73%	55.82%	26.78%	14.39%	1.61%
0.5	0.6	0.2	80	2.17%	59.87%	20.28%	19.64%	2.2%
0.5	0.7	0.2	80	1.43%	62.18%	20.84%	23.11%	2.58%

Test Corpus - MD5 Scorer (with ignorable file hashset) Parameter Tuning Results - All FOI Found (continued)

Default Score	TOI Score	C_p	Runs	Min	Max	Avg	Std Dev	Std Err
0.5	0.8	0.2	80	1.18%	59.96%	19.92%	22.36%	2.5%
0.5	0.9	0.2	80	1.36%	58.9%	19.63%	21.7%	2.43%
0.5	0.99	0.2	80	1.29%	57.07%	19.37%	20.81%	2.33%
0.6	0.01	0.2	80	25.34%	92.%	72.74%	17.35%	1.94%
0.6	0.1	0.2	80	25.81%	89.37%	64.5%	15.62%	1.75%
0.6	0.2	0.2	80	24.53%	88.2%	57.7%	19.29%	2.16%
0.6	0.3	0.2	80	22.66%	87.4%	54.17%	20.9%	2.34%
0.6	0.4	0.2	80	19.77%	79.93%	51.2%	20.7%	2.31%
0.6	0.5	0.2	80	17.18%	70.36%	41.8%	16.68%	1.86%
0.6	0.6	0.2	80	9.28%	56.79%	26.37%	15.4%	1.72%
0.6	0.7	0.2	80	2.09%	63.43%	21.74%	20.6%	2.3%
0.6	0.8	0.2	80	1.72%	62.18%	20.87%	22.87%	2.56%
0.6	0.9	0.2	80	1.52%	60.44%	21.31%	22.04%	2.46%
0.6	0.99	0.2	80	1.39%	58.04%	20.51%	21.01%	2.35%
0.7	0.01	0.2	80	25.26%	93.04%	75.91%	17.67%	1.98%
0.7	0.1	0.2	80	25.55%	90.06%	67.88%	14.33%	1.6%
0.7	0.2	0.2	80	24.02%	89.05%	60.25%	18.53%	2.07%
0.7	0.3	0.2	80	23.94%	87.66%	55.64%	20.93%	2.34%
0.7	0.4	0.2	80	22.92%	87.39%	53.2%	21.06%	2.35%
0.7	0.5	0.2	80	21.3%	77.5%	48.82%	19.15%	2.14%
0.7	0.6	0.2	80	18.24%	70.26%	40.91%	17.26%	1.93%
0.7	0.7	0.2	80	10.59%	57.46%	26.53%	14.76%	1.65%
0.7	0.8	0.2	80	2.33%	62.75%	22.9%	20.38%	2.28%
0.7	0.9	0.2	80	1.74%	61.69%	22.04%	22.57%	2.52%
0.7	0.99	0.2	80	1.52%	60.25%	21.25%	21.57%	2.41%
0.8	0.01	0.2	80	23.3%	94.48%	77.05%	17.65%	1.97%
0.8	0.1	0.2	80	24.91%	91.6%	65.53%	19.12%	2.14%
0.8	0.2	0.2	80	24.11%	88.91%	64.93%	16.09%	1.8%
0.8	0.3	0.2	80	24.7%	91.24%	59.76%	18.87%	2.11%
0.8	0.4	0.2	80	23.68%	87.32%	52.39%	22.36%	2.5%
0.8	0.5	0.2	80	21.51%	87.09%	50.29%	21.23%	2.37%
0.8	0.6	0.2	80	19.01%	67.54%	46.69%	18.67%	2.09%
0.8	0.7	0.2	80	16.92%	70.45%	40.61%	18.15%	2.03%
0.8	0.8	0.2	80	4.68%	57.56%	27.3%	14.93%	1.67%
0.8	0.9	0.2	80	3.05%	64.%	23.94%	20.75%	2.32%
0.8	0.99	0.2	80	1.52%	62.56%	22.18%	22.16%	2.48%
0.9	0.01	0.2	80	23.55%	95.59%	71.19%	24.06%	2.69%
0.9	0.1	0.2	80	23.38%	92.46%	66.38%	19.46%	2.18%

Test Corpus - MD5 Scorer (with ignorable file hashset) Parameter Tuning Results - All FOI Found (continued)

Default Score	TOI Score	C_p	Runs	Min	Max	Avg	Std Dev	Std Err
0.9	0.2	0.2	80	23.34%	90.92%	56.52%	22.66%	2.53%
0.9	0.3	0.2	80	24.57%	91.84%	54.11%	22.06%	2.47%
0.9	0.4	0.2	80	23.89%	92.53%	52.55%	22.09%	2.47%
0.9	0.5	0.2	80	23.98%	91.23%	51.36%	22.12%	2.47%
0.9	0.6	0.2	80	22.87%	86.78%	49.6%	20.73%	2.32%
0.9	0.7	0.2	80	20.96%	66.7%	45.75%	17.83%	1.99%
0.9	0.8	0.2	80	19.39%	70.26%	40.67%	17.23%	1.93%
0.9	0.9	0.2	80	4.27%	58.23%	27.47%	15.75%	1.76%
0.9	0.99	0.2	80	3.22%	63.14%	25.43%	20.02%	2.24%
0.99	0.01	0.2	80	25.09%	96.9%	69.7%	26.22%	2.93%
0.99	0.1	0.2	80	24.45%	92.54%	60.08%	23.87%	2.67%
0.99	0.2	0.2	80	23.77%	95.26%	55.66%	22.17%	2.48%
0.99	0.3	0.2	80	22.96%	92.36%	54.18%	22.34%	2.5%
0.99	0.4	0.2	80	23.85%	93.43%	51.76%	20.31%	2.27%
0.99	0.5	0.2	80	23.85%	92.3%	50.7%	20.93%	2.34%
0.99	0.6	0.2	80	23.26%	90.98%	50.22%	21.39%	2.39%
0.99	0.7	0.2	80	22.96%	85.26%	48.09%	19.02%	2.13%
0.99	0.8	0.2	80	22.32%	69.32%	45.73%	18.04%	2.02%
0.99	0.9	0.2	80	14.07%	70.74%	37.26%	18.18%	2.03%
0.99	0.99	0.2	80	4.46%	58.33%	28.39%	15.58%	1.74%

Table D.18: Test Corpus - MD5 Scorer (without ignorable file hashset) Parameter Tuning Results - All FOI Found

Default Score	TOI Score	C_p	Runs	Min	Max	Avg	Std Dev	Std Err
0.01	0.01	0.2	80	7.23%	62.66%	33.77%	20.11%	2.25%
0.01	0.1	0.2	80	2.89%	63.62%	25.21%	21.49%	2.4%
0.01	0.2	0.2	80	1.59%	66.03%	24.97%	24.14%	2.7%
0.01	0.3	0.2	80	1.32%	61.98%	22.05%	22.71%	2.54%
0.01	0.4	0.2	80	1.22%	59.67%	20.6%	22.09%	2.47%
0.01	0.5	0.2	80	1.24%	59.58%	19.93%	21.83%	2.44%
0.01	0.6	0.2	80	1.05%	56.98%	20.33%	21.04%	2.35%
0.01	0.7	0.2	80	1.14%	57.27%	20.9%	21.01%	2.35%
0.01	0.8	0.2	80	1.18%	56.4%	21.46%	20.66%	2.31%
0.01	0.9	0.2	80	1.3%	54.86%	21.82%	20.29%	2.27%
0.01	0.99	0.2	80	1.21%	54.76%	21.46%	20.66%	2.31%
0.1	0.01	0.2	80	13.95%	99.38%	69.2%	30.58%	3.42%
0.1	0.1	0.2	80	10.08%	62.18%	34.36%	20.02%	2.24%
0.1	0.2	0.2	80	2.63%	64.87%	25.54%	22.24%	2.49%

Test Corpus - MD5 Scorer (without ignorable file hashset) Parameter Tuning Results - All FOI Found (continued)

Default Score	TOI Score	C_p	Runs	Min	Max	Avg	Std Dev	Std Err
0.1	0.3	0.2	80	1.48%	65.64%	24.56%	23.81%	2.66%
0.1	0.4	0.2	80	1.23%	61.98%	21.92%	22.73%	2.54%
0.1	0.5	0.2	80	1.28%	58.71%	20.39%	21.71%	2.43%
0.1	0.6	0.2	80	1.03%	58.71%	20.%	21.32%	2.38%
0.1	0.7	0.2	80	1.15%	57.27%	20.31%	21.04%	2.35%
0.1	0.8	0.2	80	.91%	56.88%	20.95%	20.84%	2.33%
0.1	0.9	0.2	80	1.21%	55.34%	21.27%	20.53%	2.3%
0.1	0.99	0.2	80	1.27%	56.02%	21.%	20.43%	2.28%
0.2	0.01	0.2	80	25.81%	99.69%	75.87%	27.88%	3.12%
0.2	0.1	0.2	80	15.01%	99.32%	69.94%	30.65%	3.43%
0.2	0.2	0.2	80	11.78%	62.18%	34.64%	18.91%	2.11%
0.2	0.3	0.2	80	2.01%	66.12%	25.16%	21.64%	2.42%
0.2	0.4	0.2	80	1.58%	65.64%	24.52%	23.69%	2.65%
0.2	0.5	0.2	80	1.33%	63.72%	22.39%	23.32%	2.61%
0.2	0.6	0.2	80	1.31%	59.87%	20.58%	21.95%	2.45%
0.2	0.7	0.2	80	1.32%	58.33%	19.54%	21.73%	2.43%
0.2	0.8	0.2	80	.93%	58.23%	20.67%	21.3%	2.38%
0.2	0.9	0.2	80	1.09%	56.02%	21.25%	20.72%	2.32%
0.2	0.99	0.2	80	1.22%	55.15%	20.83%	20.43%	2.28%
0.3	0.01	0.2	80	29.68%	99.77%	77.07%	25.05%	2.8%
0.3	0.1	0.2	80	25.04%	99.76%	75.21%	27.95%	3.12%
0.3	0.2	0.2	80	17.05%	99.29%	70.71%	29.49%	3.3%
0.3	0.3	0.2	80	7.99%	64.39%	34.51%	18.5%	2.07%
0.3	0.4	0.2	80	2.51%	66.51%	25.72%	22.36%	2.5%
0.3	0.5	0.2	80	1.47%	65.16%	24.73%	23.86%	2.67%
0.3	0.6	0.2	80	1.34%	63.04%	22.06%	22.87%	2.56%
0.3	0.7	0.2	80	1.29%	60.35%	20.28%	22.54%	2.52%
0.3	0.8	0.2	80	1.06%	58.33%	20.12%	21.39%	2.39%
0.3	0.9	0.2	80	1.16%	57.27%	20.7%	21.29%	2.38%
0.3	0.99	0.2	80	1.18%	56.21%	20.61%	20.69%	2.31%
0.4	0.01	0.2	80	32.53%	99.76%	79.73%	21.06%	2.35%
0.4	0.1	0.2	80	33.55%	99.76%	77.5%	23.9%	2.67%
0.4	0.2	0.2	80	30.19%	99.71%	76.34%	26.61%	2.97%
0.4	0.3	0.2	80	16.16%	99.31%	70.88%	28.66%	3.2%
0.4	0.4	0.2	80	11.29%	62.37%	35.42%	17.69%	1.98%
0.4	0.5	0.2	80	2.75%	66.41%	25.48%	21.75%	2.43%
0.4	0.6	0.2	80	1.39%	65.54%	24.52%	23.86%	2.67%
0.4	0.7	0.2	80	1.23%	62.08%	21.95%	22.79%	2.55%
0.4	0.8	0.2	80	1.13%	59.77%	20.55%	22.1%	2.47%

Test Corpus - MD5 Scorer (without ignorable file hashset) Parameter Tuning Results - All FOI Found (continued)

Default Score	TOI Score	C_p	Runs	Min	Max	Avg	Std Dev	Std Err
0.4	0.9	0.2	80	1.34%	59.%	20.15%	21.84%	2.44%
0.4	0.99	0.2	80	1.19%	57.46%	20.54%	21.15%	2.36%
0.5	0.01	0.2	80	50.3%	99.77%	84.12%	15.83%	1.77%
0.5	0.1	0.2	80	44.77%	99.74%	80.86%	19.36%	2.17%
0.5	0.2	0.2	80	41.58%	99.76%	78.93%	22.04%	2.46%
0.5	0.3	0.2	80	34.44%	99.76%	77.17%	25.66%	2.87%
0.5	0.4	0.2	80	16.33%	99.32%	72.82%	27.66%	3.09%
0.5	0.5	0.2	80	10.33%	63.23%	35.24%	17.84%	1.99%
0.5	0.6	0.2	80	2.51%	67.95%	26.61%	22.89%	2.56%
0.5	0.7	0.2	80	1.67%	65.26%	24.81%	23.81%	2.66%
0.5	0.8	0.2	80	1.39%	62.27%	22.04%	22.73%	2.54%
0.5	0.9	0.2	80	1.32%	60.44%	21.08%	22.46%	2.51%
0.5	0.99	0.2	80	1.17%	59.%	20.11%	21.71%	2.43%
0.6	0.01	0.2	80	58.9%	99.76%	84.87%	15.15%	1.69%
0.6	0.1	0.2	80	58.9%	99.75%	84.7%	15.2%	1.7%
0.6	0.2	0.2	80	50.3%	99.74%	82.26%	17.83%	1.99%
0.6	0.3	0.2	80	42.05%	99.77%	79.66%	21.17%	2.37%
0.6	0.4	0.2	80	35.25%	99.76%	77.31%	25.%	2.79%
0.6	0.5	0.2	80	27.%	99.33%	73.69%	26.85%	3.%
0.6	0.6	0.2	80	11.13%	65.64%	36.84%	18.25%	2.04%
0.6	0.7	0.2	80	2.49%	68.72%	26.7%	22.5%	2.52%
0.6	0.8	0.2	80	1.54%	65.93%	24.85%	23.75%	2.66%
0.6	0.9	0.2	80	1.37%	62.37%	22.21%	22.64%	2.53%
0.6	0.99	0.2	80	1.12%	60.73%	20.68%	21.94%	2.45%
0.7	0.01	0.2	80	57.65%	99.77%	84.98%	15.37%	1.72%
0.7	0.1	0.2	80	58.23%	99.77%	84.79%	15.42%	1.72%
0.7	0.2	0.2	80	57.07%	99.77%	84.83%	15.19%	1.7%
0.7	0.3	0.2	80	60.35%	99.77%	83.98%	15.87%	1.77%
0.7	0.4	0.2	80	31.68%	99.77%	79.8%	21.03%	2.35%
0.7	0.5	0.2	80	33.5%	99.76%	78.22%	24.32%	2.72%
0.7	0.6	0.2	80	28.19%	99.32%	74.09%	26.42%	2.95%
0.7	0.7	0.2	80	15.41%	65.35%	36.94%	16.4%	1.83%
0.7	0.8	0.2	80	2.8%	69.01%	27.02%	22.25%	2.49%
0.7	0.9	0.2	80	1.57%	65.45%	25.37%	24.23%	2.71%
0.7	0.99	0.2	80	1.17%	63.04%	22.54%	23.08%	2.58%
0.8	0.01	0.2	80	61.69%	99.79%	85.23%	15.26%	1.71%
0.8	0.1	0.2	80	55.82%	99.75%	85.%	15.48%	1.73%
0.8	0.2	0.2	80	54.76%	99.77%	84.86%	15.43%	1.73%
0.8	0.3	0.2	80	59.38%	99.75%	84.97%	15.02%	1.68%

Test Corpus - MD5 Scorer (without ignorable file hashset) Parameter Tuning Results - All FOI Found (continued)

Default Score	TOI Score	C_p	Runs	Min	Max	Avg	Std Dev	Std Err
0.8	0.4	0.2	80	50.98%	99.76%	83.61%	16.53%	1.85%
0.8	0.5	0.2	80	48.38%	99.76%	80.71%	19.69%	2.2%
0.8	0.6	0.2	80	38.18%	99.77%	78.46%	23.36%	2.61%
0.8	0.7	0.2	80	29.04%	99.32%	74.58%	25.66%	2.87%
0.8	0.8	0.2	80	13.32%	67.66%	38.27%	16.75%	1.87%
0.8	0.9	0.2	80	2.68%	70.64%	27.92%	22.88%	2.56%
0.8	0.99	0.2	80	1.69%	66.41%	25.53%	23.84%	2.67%
0.9	0.01	0.2	80	27.85%	99.76%	84.92%	16.41%	1.83%
0.9	0.1	0.2	80	27.3%	99.76%	84.74%	16.61%	1.86%
0.9	0.2	0.2	80	60.25%	99.77%	85.47%	14.9%	1.67%
0.9	0.3	0.2	80	59.58%	99.79%	85.11%	15.17%	1.7%
0.9	0.4	0.2	80	59.38%	99.75%	85.14%	15.%	1.68%
0.9	0.5	0.2	80	60.92%	99.78%	84.87%	14.97%	1.67%
0.9	0.6	0.2	80	51.28%	99.76%	81.4%	18.84%	2.11%
0.9	0.7	0.2	80	39.5%	99.73%	79.52%	22.23%	2.49%
0.9	0.8	0.2	80	33.97%	99.32%	75.34%	24.28%	2.71%
0.9	0.9	0.2	80	9.49%	68.14%	39.37%	16.58%	1.85%
0.9	0.99	0.2	80	3.2%	69.68%	28.92%	22.52%	2.52%
0.99	0.01	0.2	80	55.15%	99.75%	85.47%	15.56%	1.74%
0.99	0.1	0.2	80	27.93%	99.8%	83.9%	18.26%	2.04%
0.99	0.2	0.2	80	56.69%	99.8%	85.56%	15.02%	1.68%
0.99	0.3	0.2	80	58.04%	99.78%	85.44%	15.03%	1.68%
0.99	0.4	0.2	80	61.21%	99.77%	85.42%	14.9%	1.67%
0.99	0.5	0.2	80	60.73%	99.76%	85.28%	14.82%	1.66%
0.99	0.6	0.2	80	52.55%	99.77%	84.59%	15.52%	1.74%
0.99	0.7	0.2	80	53.66%	99.76%	81.75%	18.4%	2.06%
0.99	0.8	0.2	80	37.59%	99.67%	79.61%	21.89%	2.45%
0.99	0.9	0.2	80	34.69%	99.35%	75.01%	23.76%	2.66%
0.99	0.99	0.2	80	14.7%	68.43%	40.77%	16.75%	1.87%

References

- Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viégas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y. and Zheng, X. (2015). TensorFlow: Large-scale machine learning on heterogeneous systems. Software available from tensorflow.org.
URL: <https://www.tensorflow.org/>
- Abbasi, A. and Chen, H. (2007). Affect intensity analysis of dark web forums, *2007 IEEE Intelligence and Security Informatics*, pp. 282–288.
- Al-Rowaily, K., Abulaish, M., Haldar, N. A.-H. and Al-Rubaian, M. (2015). Bisal – a bilingual sentiment analysis lexicon to analyze dark web forums for cyber security, *Digital Investigation* **14**: 53 – 62.
URL: <http://www.sciencedirect.com/science/article/pii/S1742287615000870>
- Alex, M. E. and Kishore, R. (2017). Forensics framework for cloud computing, *Computers Electrical Engineering* **60**: 193 – 205.
URL: <http://www.sciencedirect.com/science/article/pii/S0045790617302689>
- Anwar, T. and Abulaish, M. (2012). Identifying cliques in dark web forums - an agglomerative clustering approach, *Intelligence and Security Informatics (ISI), 2012 IEEE International Conference on*, pp. 171–173.
- Auer, P., Cesa-Bianchi, N. and Fischer, P. (2002). Finite-time analysis of the multiarmed bandit problem, *Machine Learning* **47**(2-3): 235–256.
- Australian Bureau of Statistics (2016). Ancestry 1st response/ancestry 2nd response/ancestry multi response (2901.0 - census dictionary, 2011).
URL: <http://www.abs.gov.au/ausstats/abs@.nsf/Lookup/2901.0Chapter602011>
- Australian Federal Police (2017). Policing for a safer australia - strategy for future capability.
URL: <https://www.afp.gov.au/sites/default/files/PDF/strategy-for-future-capability.pdf>

- Avila, S., Thome, N., Cord, M., Valle, E. and de A. Araújo, A. (2013). Pooling in image representation: The visual codeword point of view, *Computer Vision and Image Understanding* **117**(5): 453 – 465.
URL: <http://www.sciencedirect.com/science/article/pii/S1077314212001737>
- Barratt, M. J. (2012). Silk road: Ebay for drugs, *Addiction* **107**(3): 683–683.
URL: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1360-0443.2011.03709.x>
- Beebe, N. and Dietrich, G. (2007). A new process model for text string searching, in P. CRAIGER and S. SHENOI (eds), *Advances in Digital Forensics III*, Vol. 242 of *IFIP — The International Federation for Information Processing*, Springer New York, pp. 179–191.
- Beebe, N. L. and Clark, J. G. (2005). A hierarchical, objectives-based framework for the digital investigations process, *Digital Investigation* **2**(2): 147 – 167.
URL: <http://www.sciencedirect.com/science/article/pii/S1742287605000307>
- Beebe, N. L. and Clark, J. G. (2007). Digital forensic text string searching: Improving information retrieval effectiveness by thematically clustering search results, *Digital Investigation* **4**, **Supplement**(0): 49 – 54.
URL: <http://www.sciencedirect.com/science/article/pii/S1742287607000412>
- Beebe, N. L., Clark, J. G., Dietrich, G. B., Ko, M. S. and Ko, D. (2011). Post-retrieval search hit clustering to improve information retrieval effectiveness: Two digital forensics case studies, *Decision Support Systems* **51**(4): 732 – 744. Recent Advances in Data, Text, and Media Mining; Information Issues in Supply Chain and in Service System Design.
URL: <http://www.sciencedirect.com/science/article/pii/S0167923611000388>
- Beebe, N. L. and Liu, L. (2014). Clustering digital forensic string search output, *Digital Investigation* **11**(4): 314–322.
- Bengio, Y., Delalleau, O. and Roux, N. L. (2006). The curse of highly variable functions for local kernel machines, in Y. WEISS, B. SCHÖLKOPF and J. C. PLATT (eds), *Advances in Neural Information Processing Systems 18*, MIT Press, pp. 107–114.
URL: <http://papers.nips.cc/paper/2810-the-curse-of-highly-variable-functions-for-local-kernel-machines.pdf>
- Biryukov, A., Pustogarov, I., Thill, F. and Weinmann, R.-P. (2014). Content and popularity analysis of tor hidden services, *Proceedings of the 2014 IEEE 34th International Conference on Distributed Computing Systems Workshops, ICDCSW '14*, IEEE Computer Society, Washington, DC, USA, pp. 188–193.
- Biryukov, A., Pustogarov, I. and Weinmann, R. (2013). Content and popularity analysis of tor hidden services, *CoRR* **abs/1308.6768**.
URL: <http://arxiv.org/abs/1308.6768>
- BitTorrent inc (n.d.). Bittorrent.
URL: <https://www.bittorrent.com>

- Borys, S. (2017). Federal budget 2017 afp to get 321mfundingboosttohireextrapersonnel.
URL: <http://www.abc.net.au/news/story-streams/federal-budget-2017/2017-05-08/federal-budget-2017-afp>
- Boser, B. E., Guyon, I. M. and Vapnik, V. N. (1992). A training algorithm for optimal margin classifiers, *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*, COLT '92, ACM, New York, NY, USA, pp. 144–152.
URL: <http://doi.acm.org/10.1145/130385.130401>
- Brown, J., Fielding, J. and Grover, J. (1999). Distinguishing traumatic, vicarious and routine operational stressor exposure and attendant adverse consequences in a sample of police officers, **13**.
- Browne, C., Powley, E., Whitehouse, D., Lucas, S., Cowling, P., Rohlfshagen, P., Tavener, S., Perez, D., Samothrakis, S. and Colton, S. (2012). A survey of monte carlo tree search methods, *IEEE Transactions on Computational Intelligence and AI in Games* **4**(1): 1–43.
- Buolamwini, J. and Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification, in S. A. Friedler and C. Wilson (eds), *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*, Vol. 81 of *Proceedings of Machine Learning Research*, PMLR, New York, NY, USA, pp. 77–91.
URL: <http://proceedings.mlr.press/v81/buolamwini18a.html>
- Caetano, C., Avila, S., Schwartz, W. R., Guimarães, S. J. F. and de A. Araújo, A. (2016). A mid-level video representation based on binary descriptors: A case study for pornography detection, *Neurocomputing* **213**: 102 – 114. Binary Representation Learning in Computer Vision.
URL: <http://www.sciencedirect.com/science/article/pii/S0925231216307111>
- Carrier, B. (2003). Defining digital forensic examination and analysis tools using abstraction layers, *International Journal of digital evidence* **1**(4): 1–12.
- Carrier, B. (n.d.). The sleuth kit (tsk) and autopsy: Open source digital forensics tools.
URL: <https://www.sleuthkit.org/>
- Carrier, B. D. and Spafford, E. H. (2006). Categories of digital investigation analysis techniques based on the computer history model, *Digital Investigation* **3**, **Supplement**(0): 121 – 130. The Proceedings of the 6th Annual Digital Forensic Research Workshop (DFRWS '06).
URL: <http://www.sciencedirect.com/science/article/pii/S1742287606000739>
- Carrier, B. and Spafford, E. H. (2003). Getting physical with the digital investigation process, *International Journal of Digital Evidence* **2**(2).
- CBurnett (n.d.). An example artificial neural network with a hidden layer.
- Chakrabarti, S., Punera, K. and Subramanyam, M. (2002). Accelerated focused crawling through online relevance feedback, *Proceedings of the 11th international conference on World Wide Web*, ACM, pp. 148–159.

- Chakrabarti, S., Van den Berg, M. and Dom, B. (1999). Focused crawling: a new approach to topic-specific web resource discovery, *Computer Networks* **31**(11): 1623–1640.
- Chaslot, G. M. J., Winands, M. H., van den Herik, H. J., Uiterwijk, J. W. and Bouzy, B. (2008). Progressive strategies for monte-carlo tree search, *New Mathematics and Natural Computation* **4**(03): 343–357.
- Chatzis, V., Panagiotopoulos, F. and Mardiris, V. (2016). Face to iris area ratio as a feature for children detection in digital forensics applications, *2016 Digital Media Industry Academic Forum (DMIAF)*, pp. 121–124.
- Chen, H. (2012). *Dark Web: Exploring and Data Mining the Dark Side of the Web*, Vol. 30 of *Integrated Series in Information Systems*, Springer Science+Business Media.
- Chen, H., Chung, W., Qin, J., Reid, E., Sageman, M. and Weimann, G. (2008). Uncovering the dark web: A case study of jihad on the web, *Journal of the American Society for Information Science and Technology* **59**(8): 1347–1359.
URL: <http://dx.doi.org/10.1002/asi.20838>
- Choi, E. a. B. (n.d.). Youtube channel of 'bigboichoi'.
URL: <https://gaming.youtube.com/watch?v=ONTyad2k7fc>
- Chollet, F. (2016). Building powerful image classification models using very little data.
URL: <https://blog.keras.io/building-powerful-image-classification-models-using-very-little-data.html>
- Cohen, M. (2008). Pyflag – an advanced network forensic framework, *Digital Investigation* **5**, **Supplement**(0): S112 – S120. The Proceedings of the Eighth Annual {DFRWS} Conference.
URL: <http://www.sciencedirect.com/science/article/pii/S1742287608000418>
- Culley, A. (2003). Computer forensics: past, present and future, *Information Security Technical Report* **8**(2): 32 – 36.
URL: <http://www.sciencedirect.com/science/article/pii/S1363412703002048>
- da Cruz Nassif, L. and Hruschka, E. (2013). Document clustering for forensic analysis: An approach for improving computer inspection, *Information Forensics and Security, IEEE Transactions on* **8**(1): 46–54.
- Dalins, J., Tyshetskiy, Y., Wilson, C., Carman, M. J. and Boudry, D. (2018). Laying foundations for effective machine learning in law enforcement. majura – a labelling schema for child exploitation materials, *Digital Investigation* .
URL: <http://www.sciencedirect.com/science/article/pii/S1742287618301555>
- Dalins, J., Wilson, C. and Carman, M. (2015). Monte-carlo filesystem search - a crawl strategy for digital forensics, *Digit. Investig.* **13**(C): 58–71.
URL: <http://dx.doi.org/10.1016/j.diin.2015.04.002>

- Dalins, J., Wilson, C. and Carman, M. (2018). Criminal motivation on the dark web: A categorisation model for law enforcement, *Digital Investigation* .
URL: <http://www.sciencedirect.com/science/article/pii/S174228761730333X>
- Dasgupta, S. (2017). Caffe to tensorflow.
URL: <https://github.com/ethereon/caffe-tensorflow>
- Decherchi, S., Tacconi, S., Redi, J., Leoncini, A., Sangiacomo, F. and Zunino, R. (2009). Text clustering for digital forensics analysis, in Á. Herrero, P. Gastaldo, R. Zunino and E. Corchado (eds), *Computational Intelligence in Security for Information Systems*, Vol. 63 of *Advances in Intelligent and Soft Computing*, Springer Berlin Heidelberg, pp. 29–36.
- Diligenti, M., Coetzee, F., Lawrence, S., Giles, C. L., Gori, M. et al. (2000). Focused crawling using context graphs., *VLDB*, pp. 527–534.
- Dingledine, R., Mathewson, N. and Syverson, P. (2004). Tor: The second-generation onion router, *Proceedings of the 13th Conference on USENIX Security Symposium - Volume 13*, SSYM'04, USENIX Association, Berkeley, CA, USA, pp. 21–21.
URL: <http://dl.acm.org/citation.cfm?id=1251375.1251396>
- Drain, L. (2015). Bikini selfie (facebook page of lauren drain).
URL: <https://www.facebook.com/LaurenDrain/photos/a.658326280852872.1073741826.322607391091431/>
- Edelmann, R. J. (2010). Exposure to child abuse images as part of one's work: possible psychological implications, *The Journal of Forensic Psychiatry & Psychology* **21**(4): 481–489.
URL: <https://doi.org/10.1080/14789940903540792>
- Eidinger, E., Enbar, R. and Hassner, T. (2014). Age and gender estimation of unfiltered faces, *IEEE Transactions on Information Forensics and Security* **9**(12): 2170–2179.
- eMule Project (n.d.). emule project.net - official emule homepage. downloads, help, documentaiton, news.
URL: <https://www.emule-project.net/home/perl/general.cgi?l=1>
- Enzenberger, M., Muller, M., Arneson, B. and Segal, R. (2010). Fuego—an open-source framework for board games and go engine based on monte carlo tree search, *IEEE Transactions on Computational Intelligence and AI in Games* **2**(4): 259–270.
- Fail2Ban.org (2016). Fail2ban.
URL: http://www.fail2ban.org/wiki/index.php/Main_Page
- Farivar, C. (2013). Just a month after shutdown, silk road 2.0 emerges.
URL: <https://arstechnica.com/information-technology/2013/11/just-a-month-after-shutdown-silk-road-2-0-emerges/>
- Fei, B., Eloff, J., Venter, H. and Olivier, M. (2005). Exploring forensic data with self-organizing maps, in M. Pollitt and S. Shenoï (eds), *Advances in Digital Forensics*, Vol.

- 194 of *IFIP — The International Federation for Information Processing*, Springer US, pp. 113–123.
URL: http://dx.doi.org/10.1007/0-387-31163-7_10
- Ferraro, M. M. and Russell, A. (2004). Current issues confronting well-established computer-assisted child exploitation and computer crime task forces, *Digital Investigation* **1**(1): 7 – 15.
- Fielding, R. T. (1994). Maintaining distributed hypertext infostructures: Welcome to mom-spider’s web, *Comput. Netw. ISDN Syst.* **27**(2): 193–204.
URL: [http://dx.doi.org/10.1016/0169-7552\(94\)90133-3](http://dx.doi.org/10.1016/0169-7552(94)90133-3)
- Florescu, D., Levy, A. Y. and Mendelzon, A. O. (1998). Database techniques for the world-wide web: A survey, *SIGMOD record* **27**(3): 59–74.
- Franqueira, V. N., Bryce, J., Mutawa, N. A. and Marrington, A. (2017). Investigation of indecent images of children cases: Challenges and suggestions collected from the trenches, *Digital Investigation* .
URL: <http://www.sciencedirect.com/science/article/pii/S1742287617302669>
- Fu, T., Abbasi, A. and Chen, H. (2010). A focused crawler for dark web forums, *J. Am. Soc. Inf. Sci. Technol.* **61**(6): 1213–1231.
URL: <http://dx.doi.org/10.1002/asi.v61:6>
- Garfinkel, S. L. (2010). Digital forensics research: The next 10 years, *Digital Investigation* **7**, **Supplement**(0): S64 – S73. The Proceedings of the Tenth Annual {DFRWS} Conference.
URL: <http://www.sciencedirect.com/science/article/pii/S1742287610000368>
- Gelly, S., Wang, Y., Munos, R. and Teytaud, O. (2006). Modification of UCT with Patterns in Monte-Carlo Go, *Technical Report RR-6062*, INRIA.
- GetData (2018). Computer forensics software: Mount encase images and dd images.
URL: <http://www.mountimage.com/>
- Google (n.d.). Download your data - google account help.
URL: <https://support.google.com/accounts/answer/3024190?hl=en>
- Google Scholar (n.d.).
URL: https://scholar.google.com.au/scholaras_Fvis=1&q=imagenet&hl=en&as_Fsdt=1,5
- Gordon, A. D. (1996). Null models in cluster validation, in W. Gaul and D. Pfeifer (eds), *From Data to Knowledge*, Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 32–44.
- Grajeda, C., Breitingner, F. and Baggili, I. (2017). Availability of datasets for digital forensics – and what is missing, *Digital Investigation* **22**: S94 – S105.
URL: <http://www.sciencedirect.com/science/article/pii/S1742287617301913>
- Gribbin, C. (2017). Revenge porn Facebook teaming up with government to stop nude photos ending up on messenger, instagram.

URL: <http://www.abc.net.au/news/2017-11-02/facebook-offers-revenge-porn-solution/9112420>

Guidance Software (2018). Encase forensic software.

URL: <https://www.guidancesoftware.com/encase-forensic>

Guitton, C. (2013). A review of the available content on tor hidden services: The case against further development, *Computers in Human Behavior* **29**(6): 2805 – 2815.

URL: <http://www.sciencedirect.com/science/article/pii/S0747563213002690>

He, K., Zhang, X., Ren, S. and Sun, J. (2014). Spatial pyramid pooling in deep convolutional networks for visual recognition, *CoRR* **abs/1406.4729**.

URL: <http://arxiv.org/abs/1406.4729>

Ho, T. (2017). Csiro accelerator cluster - bragg.

URL: <https://confluence.csiro.au/display/SC/CSIRO+Accelerator+Cluster+-+Bragg>

Home Office (United Kingdom) (2012). Definition of policing by consent.

URL: <https://www.gov.uk/government/publications/policing-by-consent/definition-of-policing-by-consent>

Horswell, J. and Fowler, C. (2004). Associative evidence - the locard exchange principle, in J. Horswell (ed.), *The Practice of Crime Scene Investigation*, CRC Press, chapter 2, pp. 45–55.

Ieong, R. S. C. (2006). Forza - digital forensics investigation framework that incorporate legal issues, *Digit. Investig.* **3**: 29–36.

URL: <http://dx.doi.org/10.1016/j.diin.2006.06.004>

Iliou, C., Kalpakis, G., Tsikrika, T., Vrochidis, S. and Kompatsiaris, I. (2016). Hybrid focused crawling for homemade explosives discovery on surface and dark web, *2016 11th International Conference on Availability, Reliability and Security (ARES)*, pp. 229–234.

IMDB.com (n.d.). Demolition man photo gallery.

URL: <https://www.imdb.com/title/tt0106697/mediaviewer/rm3746957824>

Internet Archive (2015). Official limewire website.

URL: <https://web.archive.org/web/20150206024529/http://www.limewire.com/>

Islam, M., Watters, P. A. and Yearwood, J. (2011). Real-time detection of children’s skin on social networking sites using markov random field modelling, *Information Security Technical Report* **16**(2): 51 – 58. Social Networking Threats.

URL: <http://www.sciencedirect.com/science/article/pii/S1363412711000550>

Ith, T. (2015). Microsoft’s photodna: Protecting children and businesses in the cloud.

URL: <https://news.microsoft.com/features/microsofts-photodna-protecting-children-and-businesses-in-the-cloud/>

- Jacobellis v. Ohio* (1964). Supreme Court of the United States of America.
URL: <https://www.law.cornell.edu/supremecourt/text/378/184>
- Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S. and Darrell, T. (2014). Caffe: Convolutional architecture for fast feature embedding, *arXiv preprint arXiv:1408.5093*.
- Kent, K., Chevalier, S., Grance, T. and Dang, H. (2006). *Guide to Integrating Forensic Techniques into Incident Response*, number 800-86 in *Special Publication (NIST SP)*, NIST.
- Kingma, D. P. and Ba, J. (2014). Adam: A method for stochastic optimization, *CoRR* **abs/1412.6980**.
URL: <http://arxiv.org/abs/1412.6980>
- Klimt, B. and Yang, Y. (2004). The enron corpus: A new dataset for email classification research, in J.-F. Boulicaut, F. Esposito, F. Giannotti and D. Pedreschi (eds), *Machine Learning: ECML 2004*, Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 217–226.
- Klinger, Evan; Starkweather, D. (2010). phash - the open source perceptual hash library.
URL: <http://phash.org/>
- Klinger, Evan; Starkweather, D. (2018). phash demo.
URL: <http://www.phash.org/cgi-bin/phash-demo-new.cgi>
- Kocsis, L. and Szepesvári, C. (2006). Bandit based monte-carlo planning, *In: ECML-06. Number 4212 in LNCS*, Springer, pp. 282–293.
- Kocsis, L., Szepesvári, C. and Willemson, J. (2006). Improved Monte-Carlo Search, *Technical Report 1*, University of Tartu, Estonia.
- Kornblum, J., Grohne, H. and Ol, T. (2018). ssdeep - fuzzy hashing program.
URL: <https://ssdeep-project.github.io/ssdeep/>
- Kovac, J., Peer, P. and Solina, F. (2003). Human skin color clustering for face detection, *The IEEE Region 8 EUROCON 2003. Computer as a Tool.*, Vol. 2, pp. 144–148 vol.2.
- Krawetz, N. (2013). Kind of like that.
URL: <http://www.hackerfactor.com/blog/index.php?/archives/529-Kind-of-Like-That.html>
- Krizhevsky, A., Sutskever, I. and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks, *Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1, NIPS'12*, Curran Associates Inc., USA, pp. 1097–1105.
URL: <http://dl.acm.org/citation.cfm?id=2999134.2999257>
- Latapy, M., Magnien, C. and Fournier, R. (2013). Quantifying paedophile activity in a large p2p system, *Information Processing & Management* **49**(1): 248 – 263.
URL: <http://www.sciencedirect.com/science/article/pii/S0306457312000283>

- LeCun, Y., Bengio, Y. and Hinton, G. (2015). Deep learning, *Nature* **521**: 436 EP –.
URL: <http://dx.doi.org/10.1038/nature14539>
- Levi, G. and Hassner, T. (2015). Age and gender classification using convolutional neural networks, *2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 34–42.
- L’huillier, G., Ríos, S. A., Alvarez, H. and Aguilera, F. (2010). Topic-based social network analysis for virtual communities of interests in the dark web, *ACM SIGKDD Workshop on Intelligence and Security Informatics*, ACM, p. 9.
- Li, Z., Arlwais, S., Xie, Y., Yu, F. and Wang, X. (2013). Finding the linchpins of the dark web: a study on topologically dedicated hosts on malicious web infrastructures, *IEEE Symposium on Security and Privacy, 2013*, IEEE.
- libyal (n.d.). Libewf git repository.
URL: <https://github.com/libyal/libewf>
- Ling, Z., Luo, J., Yu, W., Fu, X., Jia, W. and Zhao, W. (2013). Protocol-level attacks against tor, *Computer Networks* **57**(4): 869 – 886.
URL: <http://www.sciencedirect.com/science/article/pii/S1389128612003799>
- Loader, I. (2016). In search of civic policing: Recasting the ‘peelian’ principles, *Criminal Law and Philosophy* **10**(3): 427–440. Copyright - Springer Science+Business Media Dordrecht 2016; Document feature - ; Last updated - 2016-10-11.
URL: <https://search-proquest-com.ezproxy.lib.monash.edu.au/docview/1810748417?accountid=12528>
- Mahadeokar, J., Farfade, S., Kamat, A. R. and Kappeler, A. (2016). Open nsfw model.
URL: https://github.com/yahoo/open_nsfw
- Martini, B. and Choo, K.-K. R. (2012). An integrated conceptual digital forensic framework for cloud computing, *Digital Investigation* **9**(2): 71 – 80.
URL: <http://www.sciencedirect.com/science/article/pii/S174228761200059X>
- Marturana, F. and Tacconi, S. (2013). A machine learning-based triage methodology for automated categorization of digital media, *Digital Investigation* **10**(2): 193 – 204. Triage in Digital Forensics.
- Micro, T. (2018). Tlsh - trend micro locality sensitive hash (git repository).
URL: <https://github.com/trendmicro/tlsh>
- Microsoft (2015). Photodna cloud service.
URL: <https://www.microsoft.com/en-us/photodna>
- Microsoft (2017). Computer vision - image processing & analytics — microsoft azure.
URL: <https://azure.microsoft.com/en-au/services/cognitive-services/computer-vision/>
- MongoDB inc (n.d.). Open source document database — mongodb.
URL: <https://www.mongodb.com/>

- Moore, D. and Rid, T. (2016). Cryptopolitik and the darknet, *Survival* **58**(1): 7–38.
- Moreira, D., Avila, S., Perez, M., Moraes, D., Testoni, V., Valle, E., Goldenstein, S. and Rocha, A. (2016). Pornography classification: The hidden clues in video space–time, *Forensic Science International* **268**: 46 – 61.
URL: <http://www.sciencedirect.com/science/article/pii/S0379073816304169>
- Moustafa, M. (2015). Applying deep learning to classify pornographic images and videos, *CoRR* **abs/1511.08899**.
URL: <http://arxiv.org/abs/1511.08899>
- National Institute of Standards and Technology (n.d.). National software reference library (nsrl).
URL: <https://www.nist.gov/software-quality-group/national-software-reference-library-nsrl>
- Nuix (n.d.). Investigation, cybersecurity, information governance and e-discovery software — nuix.
URL: <https://www.nuix.com/>
- OpenCV Team (2018). Opencv.
URL: <https://opencv.org/>
- Oxford English Dictionary (2018). "forensic, adj. and n.".
URL: www.oed.com/view/Entry/73107
- Palmer, G. (ed.) (2001). *A Road Map for Digital Forensics Research 2001*, Digital Forensics Research Workshop.
- Panchenko, A., Beaufort, R. and Fairon, C. (2012). Detection of child sexual abuse media on p2p networks: normalization and classification of associated filenames, *Proceedings of the LREC Workshop on Language Resources for Public Security Applications*.
- Peersman, C., Schulze, C., Rashid, A., Brennan, M. and Fischer, C. (2016). icop: Live forensics to reveal previously unknown criminal media on p2p networks, *Digital Investigation* **18**: 50 – 64.
URL: <http://www.sciencedirect.com/science/article/pii/S1742287616300779>
- Petroni, Jr., N. L., Walters, A., Fraser, T. and Arbaugh, W. A. (2006). Fatkit: A framework for the extraction and analysis of digital forensic data from volatile system memory, *Digital Investigation* **3**(4): 197 – 210.
URL: <http://www.sciencedirect.com/science/article/pii/S1742287606001228>
- Pogue, C. (2011). Sniper forensics - part 1: A brief history lesson, <https://www.trustwave.com/Resources/SpiderLabs-Blog/Sniper-Forensics—Part-1—A-Brief-History-Lesson/>.

- Pollitt, M. M. (1995). Computer forensics: An approach to evidence in cyberspace, *Proceedings of the 18th National Information Systems Security Conference*, Vol. 2, National Institute of Standards and Technology (NIST) National Computer Security Center, pp. 487–491.
- Pollitt, M. M. (2013). Triage: A practical solution or admission of failure, *Digital Investigation* **10**(2): 87 – 88. Triage in Digital Forensics.
URL: <http://www.sciencedirect.com/science/article/pii/S1742287613000030>
- Powell, M. B., Cassematis, P., Benson, M. S., Smallbone, S. and Wortley, R. (2014). Police officers' perceptions of the challenges involved in internet child exploitation investigation, *Policing: An International Journal* **37**(3): 543–557.
URL: <https://doi.org/10.1108/PIJPSM-08-2013-0080>
- Powell, M., Cassematis, P., Benson, M., Smallbone, S. and Wortley, R. (2015). Police officers' perceptions of their reactions to viewing internet child exploitation material, *Journal of Police and Criminal Psychology* **30**(2): 103–111.
URL: <https://doi.org/10.1007/s11896-014-9148-z>
- Prosise, C., Mandia, K. and Pepe, M. (2003). *Incident Response & Computer Forensics*, 2 edn, McGraw-Hill, Inc., New York, NY, USA.
- Quach, K. (2018). Fyi: There's now an ai app that generates convincing fake smut vids using celebs' faces.
URL: https://www.theregister.co.uk/2018/01/25/ai_fake_skin_flicks/
- Quick, D. and Choo, K.-K. R. (2013a). Digital droplets: Microsoft skydrive forensic data remnants, *Future Generation Computer Systems* **29**(6): 1378 – 1394. Including Special sections: High Performance Computing in the Cloud Resource Discovery Mechanisms for P2P Systems.
URL: <http://www.sciencedirect.com/science/article/pii/S0167739X13000265>
- Quick, D. and Choo, K.-K. R. (2013b). Dropbox analysis: Data remnants on user machines, *Digital Investigation* **10**(1): 3 – 18.
URL: <http://www.sciencedirect.com/science/article/pii/S174228761300011X>
- Quick, D. and Choo, K.-K. R. (2014). Google drive: Forensic analysis of data remnants, *Journal of Network and Computer Applications* **40**: 179 – 193.
URL: <http://www.sciencedirect.com/science/article/pii/S1084804513002051>
- Quick, D. and Choo, K.-K. R. (2016). Big forensic data reduction: digital forensic images and electronic evidence, *Cluster Computing* **19**(2): 723–740.
URL: <https://doi.org/10.1007/s10586-016-0553-1>
- Reith, M., Carr, C. and Gunsch, G. (2002). An examination of digital forensic models, *International Journal of Digital Evidence* **1**(3): 1–12.

- Richard, III, G. G. and Rousev, V. (2006). Next-generation digital forensics, *Commun. ACM* **49**(2): 76–80.
- Ries, C. X. and Lienhart, R. (2014). A survey on visual adult image recognition, *Multimedia Tools and Applications* **69**(3): 661–688.
URL: <https://doi.org/10.1007/s11042-012-1132-y>
- Rousev, V. (2010). Data fingerprinting with similarity digests, in K.-P. Chow and S. Shenoi (eds), *Advances in Digital Forensics VI*, Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 207–226.
- Rousev, V. and Quates, C. (2012). Content triage with similarity digests: The {M57} case study, *Digital Investigation* **9**, **Supplement**(0): S60 – S68. The Proceedings of the Twelfth Annual {DFRWS} Conference 12th Annual Digital Forensics Research Conference.
- Rousev, V., Quates, C. and Martell, R. (2013). Real-time digital forensics and triage, *Digital Investigation* **10**(2): 158 – 167. Triage in Digital Forensics.
- Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A. C. and Fei-Fei, L. (2015). ImageNet Large Scale Visual Recognition Challenge, *International Journal of Computer Vision (IJCV)* **115**(3): 211–252.
- Russell, S. and Norvig, P. (2009). *Artificial Intelligence: A Modern Approach*, 3rd edn, Prentice Hall Press, Upper Saddle River, NJ, USA, chapter Artificial Neural Networks, pp. 727–737.
- Sabbah, T., Selamat, A., Selamat, M. H., Ibrahim, R. and Fujita, H. (2016). Hybridized term-weighting method for dark web classification, *Neurocomputing* **173**(Part 3): 1908 – 1926.
URL: <http://www.sciencedirect.com/science/article/pii/S092523121501396X>
- Sae-Bae, N., Sun, X., Sencar, H. T. and Memon, N. D. (2014). Towards automatic detection of child pornography, *2014 IEEE International Conference on Image Processing (ICIP)*, pp. 5332–5336.
- Scanlon, J. R. and Gerber, M. S. (2014). Automatic detection of cyber-recruitment by violent extremists, *Security Informatics* **3**(1): 5.
URL: <https://doi.org/10.1186/s13388-014-0005-5>
- Schadd, M. P., Winands, M. H., Tak, M. J. and Uiterwijk, J. W. (2012). Single-player monte-carlo tree search for samegame, *Knowledge-Based Systems* **34**(0): 3–11.
- Seigfried-Spellar, K. C. (2017). Assessing the psychological well-being and coping mechanisms of law enforcement investigators vs. digital forensic examiners of child pornography investigations, *Journal of Police and Criminal Psychology* .
URL: <https://doi.org/10.1007/s11896-017-9248-7>

- Shapiro, L. G. (n.d.). Object and concept recognition for content-based image retrieval.
URL: http://imagedatabase.cs.washington.edu/groundtruth/barcelona2/barcelona2_031.gif
- Shimat (n.d.). opencvsharp git repository.
URL: <https://github.com/shimat/opencvsharp/releases>
- Simonyan, K. and Zisserman, A. (2014). Very deep convolutional networks for large-scale visual recognition.
URL: <http://www.robots.ox.ac.uk/vgg/research/verydeep/>
- Spideroak (2015).
URL: <https://spideroak.com/>
- Squid Project (2015). squid: Optimising web delivery.
URL: <http://www.squid-cache.org/>
- Steel, C. M. (2009). Child pornography in peer-to-peer networks, *Child Abuse & Neglect* **33**(8): 560 – 568.
URL: <http://www.sciencedirect.com/science/article/pii/S0145213409001604>
- Stephenson, P. (2003). A comprehensive approach to digital incident investigation, *Information Security Technical Report* **8**(2): 42 – 54.
- Stoffel, K., Cotofrei, P. and Han, D. (2010). Fuzzy methods for forensic data analysis, *Soft Computing and Pattern Recognition (SoCPaR), 2010 International Conference of*, pp. 23–28.
- Supreme Court of Victoria (Australia) (2015). Supreme court (general civil procedure) rules 2015 (sr no 103 of 2015), http://www5.austlii.edu.au/au/legis/vic/num_regs/sccpr2015n103o2015514/.
- Taylor, M., Holland, G. and Quayle, E. (2001). Typology of paedophile picture collections, *The Police Journal* **74**(2): 97–107.
URL: <https://doi.org/10.1177/0032258X0107400202>
- Teelink, S. and Erbacher, R. F. (2006). Improving the computer forensic analysis process through visualization, *Commun. ACM* **49**(2): 71–75.
URL: <http://doi.acm.org/10.1145/1113034.1113073>
- The Apache Software Foundation (n.d.). Apache tika.
URL: <https://tika.apache.org>
- The Open Group (2017). Posix.1-2017, *Technical Report 7*, IEEE and The Open group.
- Tibshirani, R., Walther, G. and Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **63**(2): 411–423.
URL: <http://dx.doi.org/10.1111/1467-9868.00293>

- Tor Project (2017). Tor project git repository.
URL: <https://git.torproject.org/tor.git>
- United States Census Bureau (2018). Race.
URL: <https://www.census.gov/topics/population/race/about.html>
- United States of America vs Ross William Ulbricht* (2015). United States District Court.
URL: <https://www.scribd.com/doc/283722300/Ross-Ulbricht-Sentencingfullscreenfromembed>
- Violanti, J. M. and Aron, F. (1995). Police stressors: Variations in perception among police personnel, *Journal of Criminal Justice* **23**(3): 287 – 294.
URL: <http://www.sciencedirect.com/science/article/pii/004723529500012F>
- Vitorino, P., Avila, S., Perez, M. and Rocha, A. (2018). Leveraging deep neural networks to fight child pornography in the age of social media, *Journal of Visual Communication and Image Representation* **50**: 303 – 313.
URL: <http://www.sciencedirect.com/science/article/pii/S1047320317302377>
- Wallace, K. A. (1999). Anonymity, *Ethics and Information Technology* **1**(1): 21–31.
URL: <http://dx.doi.org/10.1023/A:1010066509278>
- Wang, H., Kang, B. and Kim, D. (2013). Pfw: A face database in the wild for studying face identification and verification in uncontrolled environment, *2013 2nd IAPR Asian Conference on Pattern Recognition*, pp. 356–360.
- Wang, X., Luo, J., Yang, M. and Ling, Z. (2011). A potential http-based application-level attack against tor, *Future Generation Computer Systems* **27**(1): 67 – 77.
URL: <http://www.sciencedirect.com/science/article/pii/S0167739X10000713>
- Westlake, B., Bouchard, M. and Frank, R. (2017). Assessing the validity of automated webcrawlers as data collection tools to investigate online child sexual exploitation, *Sexual Abuse* **29**(7): 685–708.
URL: <https://doi.org/10.1177/1079063215616818>
- X-Ways Software AG (n.d.). X-ways forensic.
URL: <https://www.x-ways.net/forensics/index-m.html>
- Xu, J., Chen, H., Zhou, Y. and Qin, J. (2006). On the topology of the dark web of terrorist groups, in S. Mehrotra, D. Zeng, H. Chen, B. Thuraisingham and F.-Y. Wang (eds), *Intelligence and Security Informatics*, Vol. 3975 of *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, pp. 367–376.
URL: http://dx.doi.org/10.1007/11760146_32
- Yang, C. C., Tang, X. and Thuraisingham, B. M. (2010). An analysis of user influence ranking algorithms on dark web forums, *ACM SIGKDD Workshop on Intelligence and Security Informatics*, ISI-KDD '10, ACM, New York, NY, USA, pp. 10:1–10:7.
URL: <http://doi.acm.org/10.1145/1938606.1938616>

- Yang, L., Liu, F., Kizza, J. and Ege, R. (2009). Discovering topics from dark websites, *Computational Intelligence in Cyber Security, 2009. CICS '09. IEEE Symposium on*, pp. 175–179.
- Zeiler, M. D. and Fergus, R. (2013). Visualizing and understanding convolutional networks, *CoRR* **abs/1311.2901**.
URL: <http://arxiv.org/abs/1311.2901>

Glossary

AFP Australian Federal Police. iv, vii, 1, 4, 5, 8, 9, 11, 13, 15, 26, 45, 71, 84–86, 98, 106–108, 125, 127, 140, 150

Bibcam A term commonly associated with CEM files. Origins unknown to author. 47

CEM Child Exploitation Material . iv, v, vii, x, 1, 8–11, 17, 21, 22, 26–28, 30, 34, 37, 44–49, 51, 54, 55, 59–61, 63–65, 69, 71, 72, 76, 82, 84, 86, 90–92, 99, 100, 112, 113, 115, 122, 125, 127, 128, 131, 138, 141, 145, 147–149, 371, *Glossary*: CEM

CEM An alternative term for Child Pornography. iv, 1, 8, 17, 61, 69, 84, 115, 147, 371

CETS Child Exploitation Tracking System. xi, xiii, xiv, 11, 43, 60, 62, 69, 84, 114, 127, 141, 147, 160, 161

CFC Context Focused Crawler. 34

Child Pornography Refer *Defining CEM* (page 61) for Commonwealth (Australia) definition, as per *Criminal Code Act (Cth) 1995, 473.1 - Definitions*. 25, 42, 61

COPINE Combating Paedophile Information Networks in Europe. xi, xiii, xiv, 11, 60, 62, 147, 160–162

CT Counter Terrorism. 26

DAG Directed Acyclic Graph. 56, 117, 149

DF Digital Forensics. iv, vii, 1, 7–11, 13–16, 19, 20, 24, 26–29, 31, 32, 42, 43, 65, 83, 84, 106, 115, 125, 126, 139, 140, 143, 145, 147, 148

DFRWS Digital Forensic Research Workshop. 14, 16

DIPL Digital Investigation Process Language. 17

DNS Domain Name System. 39

DOM Document Object Model. 34

EEDI End-to-End Digital Investigation. 17

- EFW** An industry standard format for storing acquired data in a forensically sound manner. Refer https://forensicswiki.org/wiki/ASR_Data%27s_Expert_Witness_Compression_Format. 125, 143
- FAT** A file system originating in MS-DOS systems, now updated (FAT32) and often seen in use on removable media such as USB memory sticks. 19
- GPU** Graphics Processing Unit, aka 'graphics card', aka 'video card'. A specialised circuit/chip designed specifically to accelerate rendering of graphics, typically in gaming. The highly parallel nature of GPUs makes them well suited for ML tasks such as neural network training and inference. 58, 85, 96
- Hamming Distance** A measure of of similarity between strings of *equal* length, whereby the 'distance' between the strings is measured by the minimum number of characters *changed* to make them identical. Insertions/deletions are not allowed. 50, 52
- HDD** Hard Disk Drive. Persistent storage devices for saving data. Traditionally denotes media based upon spinning platters, but colloquially can refer to any persistent storage device usually (but not always) physically located/connected in a permanent/non-removable manner. 24, 25, 126, 127, 135
- JACET** Joint State and Federal Police teams assigned to online child exploitation investigations. 69, 84
- JCTT** Joint Intelligence, State and Federal Police Counter Terrorism teams. 84
- JSON** JavaScript Object Notation. A data interchange format akin to XML, built largely around name/value pairs and lists/arrays. Refer <https://www.json.org/>. As implied, originally developed for use with JavaScript, but has achieved wide support and is largely seen as language-independent. 73, 144
- MAR** mutual assistance request. 21
- MCFS** Monte Carlo Filesystem Search. v, xvii, 10, 11, 65, 119, 120, 122–125, 128, 129, 134–142, 145, 148, 149
- MCTS** Monte Carlo Tree Search. 117–121, 373
- MD5** Message Digest 5. 22, 49, 50, 52, 126, 127, 145
- mens rea** 'Guilty mind' - effectively a person's *intention* to commit a crime. 7, 149
- ML** Machine Learning. 54, 57, 60, 84
- NAT** Network Address Translation. 39
- NATA** National Association of Testing Authorities, Australia. 125

- NIST** National Institute of Standards and Technology. 16, 18
- NSFW** Not Safe For Work. 59, 60, 88, 91, 108, 111
- NSRL** National Software Reference Library. 127, 131
- OS** Operating System. 23, 34, 41, 42, 125, 126
- P2P** Peer to Peer, a non-centralised, hierarchically flat network architecture avoiding the use of central hubs, servers etc. All participants act as peers, sharing and relaying searches etc. as required. Whilst practical for other uses, P2P is commonly associated with online file sharing applications/protocols such as BitTorrent. 22
- Progressive Bias** A method for influencing the playout stage of Monte Carlo Tree Search (MCTS) by biasing results towards branches previously observed to provide better results, as part of a training stage. 122
- PTHC** Pre Teen Hard Core. 47
- PTSC** Pre Teen Soft Core. 47
- RNG** Random Number Generator. 140
- SHA-1** Secure Hash Algorithm 1. 22, 49, 50, 126
- SSD** Solid State Drive. Persistent storage devices based upon integrated circuitry rather than moving media such as platters. 23
- TMM** Tor-use Motivation Model. iv, 9, 11, 65, 76, 81, 147, 148
- Tor** Tor/'The Onion Router' is a protocol (and associated freely available software package) employing online relays for the purposes of ensuring user anonymity. Tor can be used as a 'one way' anonymiser, concealing client access to a known internet node (such as a www site). However, Tor also supports 'hidden sites', being websites and/or associated services hosted within its router network, thereby providing two-way anonymity. Whilst an acronym, only the 'T' in 'Tor' is capitalised. xvi, 4, 10, 35, 36, 68, 72–76, 81
- write blocker** A software/hardware device allowing read access to media, whilst blocking writes. 22
- WWW** World-Wide Web. xvi, 32–35, 39, 67, 73, 74, 116, 117