Learning-Based Methods for Improved Simultaneous PET-MRI

Submitted in partial fulfillment of the requirements
of the degree of

Doctor of Philosophy

of the
Indian Institute of Technology Bombay, India
and
Monash University, Australia
by

Viswanath P. Sudarshan

Supervisors:

Suyash P. Awate
and
Zhaolin Chen

The course of study for this award was developed jointly by
Monash University, Australia and Indian Institute of Technology Bombay,
and was given academic recognition by each of them.
The programme was administered by the IITB-Monash Research Academy.

2021

Declaration

I declare that this written submission represents my ideas in my own words and where others' ideas or words have been included, I have adequately cited and referenced the original sources. I declare that I have properly and accurately acknowledged all sources used in the production of this report. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. I understand that any violation of the above will be a cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

Viswanath
IITB and Monash
11 February 2021

Abstract

Recent developments for simultaneous positron emission tomography (PET) and magnetic resonance imaging (MRI) systems have enabled simultaneous measurement of structure, function, and metabolic information. Such a multimodal system provides multidimensional opportunities for fundamental neuroscience as well as for clinical neuroimaging for early diagnosis/prognosis of diseases.

For neuroscience research as well as clinical neuroimaging, it is imperative to optimize three crucial inter-related factors: (i) patient safety (by reducing ionizing radiations), (ii) diagnostic information (by improving image quality and data analysis), and (iii) the cost-effectiveness of the imaging session. In certain cases, reducing acquisition time might be necessary to reduce patient comfort in case of elderly patients or patients with movement disorders. Improving PET and MRI image quality is critical for improved theranostics and clinical decision making. Reducing the radiotracer dose improves patient safety and enables longitudinal PET imaging for adults and high-risk populations such as children and pregnant women. For PET, there is a tradeoff between improving image quality and the number of photon-counts available for image reconstruction; reducing acquisition time in MRI (and thus PET), achieved by undersampling the data, affects the image quality of MRI due to aliasing artifacts and the image quality of PET due to the reduced number of photon-counts. Reducing the overall acquisition time in addition to improving patient comfort, eliminates patient-motion artifacts in the reconstructed images, which is unavoidable in patients with movement disorders or certain neurodegenerative diseases.

Learning-based methods for image quality improvement during image reconstruction or post-reconstruction have demonstrated improved accuracy for *ill-posed* inverse problems in comparison to traditional methods. The learning technique may depend on an already existent training dataset consisting of high-quality images (from several subjects or the longitudinal scans of the subject under consideration) or the imaging data (typically structural scans) obtained within the same imaging session. In the case of learning techniques based on previously acquired data, especially on different subjects/patients, a key

factor for their clinical utility is the robustness to unseen data as well as the quantification of the the risk involved in the prediction.

This dissertation tries to address research problems related to improving the image quality of PET and MRI images with the aim of reducing the amount of injected radiations and overall acquisition times. The dissertation involves developing robust learning-based algorithms to improve PET and MRI image quality under accelerated MRI acquisition and static low-dose PET scenarios. For dynamic functional PET imaging, the thesis explores improved PET image quality to achieve improved brain functions at higher temporal resolution functional PET imaging.

Table of Contents

List of Figures

List of Abbreviations

PET	Positron emission tomography
MRI	Magnetic resonance imaging
CT	X-ray computed tomography
PET-MRI	Simultaneous PET and MRI
PET-CT	Combined PET and CT
FDG	[^{18}F]-Fluorodeoxyglucose
T1w	T1 weighted MRI
T2w	T2 weighted MRI
UTE	Ultrashort Echotime
dMRI	Diffusion MRI
fMRI	Functional MRI
fPET	Functional PET
SNR	Signal to noise ratio
DNN	Deep neural network
OOD	Out of distribution
ICA	Independent component analysis
MAP	Maximum a posteriori
EM	Expectation-maximization

Chapter 1

Introduction

In vivo non-invasive multimodal imaging involving molecular and structural imaging has evolved over the past couple of decades to enable combined imaging of structure, function, and metabolism within a single scanning session. Structural/anatomical imaging modalities like magnetic resonance imaging (MRI) and X-ray computed tomography (CT) in combination with molecular imaging modalities like positron emission tomography (PET), single-photon emission CT (SPECT), enable the understanding of pathophysiology of various diseases in oncology, neurology, and cardiology. PET, with its high sensitivity (picomolar range), provides quantitative metabolic information regarding various pathologies at the molecular level and can be used for early diagnosis (in some cases, presymptomatic detection) of several diseases, optimization of drug development, therapy, and treatment monitoring (**??**). On the other hand, MRI, in addition to providing superior soft-tissue contrast in comparison to CT, can provide functional information via various imaging protocols as well as imaging sequences such as functional MRI (fMRI), diffusion MRI (dMRI), MR spectroscopy (MRS), etc.

The inherent limitations due to the fundamental contrast-generation mechanism and due to the limitations of the imaging hardware render it unlikely that a single imaging modality can satisfy ideal imaging characteristics such as high sensitivity and specificity, high spatial resolution, and achieving all the above without compromising patient safety (e.g., using toxic substances, the involvement of ionizing radiations, etc.). Hence, there have been sustained efforts to combine imaging modalities for improved clinical theranostics as well as medical research (**?**) leading to the development of multimodal imaging systems like PET-CT and PET-MRI. Unlike information fusion from two different standalone imaging modalities, multimodal imaging systems enable simultaneous and/or sequential imaging within a single imaging session, thus enabling improved image registration between the two images and consequently, enabling improved diagnosis. The

combination of structural and metabolic information provides complementary insights for various biological processes associated with different pathologies or with the normal functioning of the human body. The advent of the simultaneous PET-MRI systems has enabled synergistic PET and MRI data acquisition and joint analysis of the spatial and temporal data from both modalities which benefits both medical research and clinical diagnosis (**?**).

1.1 Simultaneous PET-MRI System

The initial efforts of combining PET and MRI commenced around the mid-1990s, even before the introduction of the combined PET-CT systems. Compared to the PET-CT systems, the simultaneous PET-MRI scanner has the advantages of reduced ionizing radiations (due to the non-ionizing nature of MRI), and increased flexibility through access to a wide variety of MRI contrasts (both structural and functional). MRI provides a range of imaging sequences to catering to specific tissue characteristics and functions such as dMRI, MRS, quantitative susceptibility mapping (QSM), and functional MRI (fMRI).

Before the advent of the simultaneous systems, clinicians/researchers had to rely on software-based fusion of the PET and MRI images from dedicated scanners with sequential imaging sessions spread over hours or even days. Such sequential scans spread over several hours might involve a change in the functioning or physiology of the subject being scanned in medical conditions such as stroke (**?**). Hence, it is useful and in some cases, necessary to simultaneously acquire PET and MRI imaging data. The currently available simultaneous PET-MRI scanners have come a long way since the development of the first proof-of-concept model called BrainPET that consisted of an MRI compatible PET insert into a 3 T MRI system to today's fully integrated whole-body PET-MRI scanner where the PET detectors are placed between the gradient coils and the radio frequency (RF) coils of the MRI (**?**). The major challenge in combining PET and MRI was the co-existence of PET and MRI hardware due to mutual interference. This is unlike the PET-CT system where the two imaging subsystems remained largely unaltered due to the sequential design (**?**). With the advancement in PET detector electronics, current simultaneous PET-MRI systems provide a unique and unparalleled opportunity in the acquisition and correlation analysis of the spatiotemporal signals arising from both the systems (**?**). Such a system is promising for the field of neuroscience research as well as clinical neuroimaging.

1.2 Simultaneous PET-MRI: Brain Imaging

In the general context of brain imaging, simultaneous PET-MRI offers the opportunity to provide insights into understanding the human brain in healthy individuals as well as individuals with pathology (**?**). For example, the joint analysis of PET and MRI spatiotemporal data can lead to a better understanding of the energy utilization in the brain (**?**). On the other hand, for clinical neuroimaging, MRI (which is already the preferred choice of imaging modality for brain tumor imaging) combined with PET can aid in the identification of pathological sites at an early or presymptomatic stage (**?**).

PET-MRI for neuroscience. The central aim of neuroscience is to investigate and understand the working of the human brain, neural basis of cognition and various pathophysiologies (**?**). Primarily, the utility of simultaneous PET-MRI for neuroscience is seen in using PET to understand the coupling between metabolism and blood flow. While functional MRI (fMRI) provides superior spatial and temporal information about blood flow, PET provides quantitative information about metabolism in the brain depending on the radiotracer used (e.g., glucose metabolism in the case of [^{1}8F]-fluorodeoxyglucose (FDG)). Thus, the simultaneous acquisition of PET and fMRI data can provide a deeper understanding of the physiology underlying the contrast generation mechanism in fMRI (**??**). Recent works have focussed on improving the temporal resolution of dynamic FDG-PET using a continuous infusion of the radiotracer (called functional PET (fPET)) (**??????**).

The continuous infusion technique improves temporal resolution by maintaining the glucose concentration in the blood plasma. Current studies have demonstrated the identification of brain activations at a temporal resolution of ≤ 10 seconds achieved with a combination of bolus and continuous infusion to improve the signal to noise ratio (SNR) (**?**). However, even with the combination of bolus and continuous infusion protocol, there exists a substantial gap in the temporal resolution of fPET compared to that of fMRI. It is challenging to obtain brain mappings at a higher temporal resolution without significantly increasing the dose. The continuous infusion technique enables both functional and metabolic connectivity for a single subject at rest as well as while performing specific tasks (**?**).

In addition to the above, simultaneous PET-MRI imaging enables improved pharmacokinetic studies coupled with arterial spin labeling, studies involving neurotransmission and receptor expression (**?**). Such a system allows for investigation into several fundamental neuronal and physiological aspects such as the effect of drugs on blood flow opening up the possibilities for precision medicine(**?**).

PET-MRI for clinical neuroimaging. The combined information of PET and MRI provides improved sensitivity and desired quantitative accuracy than either independent

modalities crucial for understanding various pathologies and brain-related disorders detailed as follows. (i) For brain tumors, the superior specificity of PET with the availability of target-specific radiotracers, can aid in better theranostics by improved grading of tumors, and improved tumor localization, which is, in turn, critical during radiotherapy; (ii) for dementia/neurodegeneration, PET enables differential diagnosis of dementia as well as predictions at a presymptomatic stage; (iii) in certain cases of epilepsy, the MRI findings might be negative while certain specific PET radiotracers can identify the epileptic foci. For all the cases stated above, the high specificity and sensitivity of PET needs to be supported with the high-resolution MRI scans for the tasks of image enhancement, motion correction, partial volume correction, especially for the localization of the site of the pathology (**???**). Furthermore, PET can be combined with fMRI studies to monitor either the progression or regression (in case of treatment) of a disease(**?**). Importantly, as the above-stated disorders are more common in elderly patients or patients with difficulty in staying still for a long time, the combined structural and functional information in a single imaging session improves patient comfort.

1.3 Challenges in PET-MRI Neuroimaging

The simultaneous acquisition of PET-MRI data also necessitates the development of optimized neuroimaging protocols. For example, in neuroscience research, although a reduction in overall acquisition time is not as critical as in the clinical setting, longer acquisition times result in enhanced patient motion, which in turn increases motion artifacts in the images. For neuroimaging studies spanning over an hour, intermittent MRI scans such as navigator sequences are needed for continuous patient motion estimation. Another important aspect is the tradeoff between acquisition time and image quality for both PET and MRI.

Given the stochasticity of the photons in PET imaging, the noise in PET images increases with a reduced number of photons (due to a reduction in dosage or acquisition time). Reduction in dosage aligns with the principle of *as low as reasonably achievable* (**?**), and improves patient safety. Reducing the involvement of ionizing radiations in PET can potentially lay the foundation for longitudinal PET scans for subjects suspected with specific pathologies even at a presymptomatic stage. Currently, PET is not the preferred modality of choice for radiation-sensitive populations such as children and pregnant women (**?**). Another essential aspect to be considered is patient comfort. As mentioned earlier, shorter acquisitions improve the imaging experience for patients with movement-related disorders and/or patients who cannot lay still for the entire duration

of the scan (e.g, pediatric oncology). In such cases, reducing the scan time will reduce the dependency on anesthesia/sedatives. Hence, the development of optimized PET-MRI imaging protocols should enable the generation of high-quality images meeting diagnostic standards, while reducing radiation-dose (patient safety) and acquisition time (patient comfort). Additionally, the optimized protocols should also allow for cost-effectiveness of the system via increased throughput.

For the MRI subsystem, acquisition times can be reduced by eliminating redundant acquisition of contrasts and/or undersampling the MRI k-space. Typically, the longer acquisition time per MRI sequence (around 3–4 minutes) combined with the typical clinical routine of obtaining multiple MRI contrasts allows the accumulation of a sufficient number of photons from the PET scan for improved PET image quality. However, the accelerated MRI acquisition (by undersampling the k-space) will reduce the time per bed position, degrading PET image quality, and increasing the demand for radiotracer dose. Hence, to improve patient safety and comfort, there is a need for improving PET and MRI signal detection (through hardware improvements) and processing (through computational solutions)(**?**). The hardware-related developments such as improved PET detector technology and the introduction of time-of-flight imaging technologies have indeed improved the sensitivity of PET-MRI and reduced partial volume effects, respectively. On the other hand, developments in the area of medical image computing techniques, specifically, the use of machine-learning-based priors and deep neural networks (DNNs) have demonstrated working proofs of concepts to enhance (i) PET image quality from PET data with a dramatic dose reduction factor (DRF) and (ii) MRI image quality from severely undersampled MRI data.

1.4 Machine Learning for PET and MRI Image Quality Improvement

In the general context of image quality improvement techniques for medical images, including both within-reconstruction as well as post-reconstruction methods, statistical image reconstruction methods with increasingly complex models have been developed to capture finer image details. Initial prior-based methods explored simple Markov random field (MRF) based priors such as quadratic prior and edge-preserving Huber function within a Bayesian inference framework. Subsequently, the incorporation of patch-based methods, given their success in natural images, for medical images, improved over classical methods by providing robustness to noisy data and preserving higher-order image features like textural patterns (**?**). Recently, with the increased parallel-computation

power, medical imaging has witnessed the translation of several machine-learning-based techniques, including deep neural networks (DNNs), to clinics and research centres. As mentioned earlier in this section, the DNN-based techniques are now being extended to address problems specific to medical imaging modalities like improving PET image quality under low-dose/low-photon-count conditions, and improving MRI image quality under severely undersampled MRI k-space data.

Specifically, for PET, methods relying on dictionaries and neural networks have been shown effective in replacing the traditional/classical methods that perform image reconstruction/restoration without explicit prior modeling (**???**). Recent works for the task of predicting standard-dose PET (SD-PET images from low-dose PET (LD-PET) images employ a DNN learned in a supervised or unsupervised setting achieve a dose reduction factor (DRF) in the range of 4–100 ×(**???**). The DNN based works were preceded by other machine learning-based approaches such as regression forests (**?**), dictionary learning (**?**).

On the other hand, for MRI, sparse representations using wavelets or dictionaries proved effective in achieving improved image quality at a higher acceleration rate in comparison to other classical approaches that relied on image-gradient-based methods. However, compressed sensing (CS) in MRI demands acquisition sequences that are non-Cartesian and practically difficult to implement (**?**). Acceleration in routine clinical studies follows Cartesian sampling strategies with/without parallel imaging. Higher acceleration under Cartesian schemes deviates from the criteria of incoherence mandated by CS-MRI. To this end, DNN-based solutions have been proposed for de-aliasing MRI images arising from increased acceleration in MRI acquisition using Cartesian sequences (**??**).

Although the learning-based models have demonstrated improvements over traditional/classical approaches by providing improved accuracy in imaging scenarios giving rise to images with very low SNR or high level of artifacts, one major challenge obstructing the translation of such methods is the validation of the robustness of the method. Typically, the DNN models developed for PET or MRI image enhancement are trained and tested on similar datasets. For instance, both training and testing are performed on PET and MRI data collected from individuals with healthy brain anatomy and glucose metabolism. Thus, for data samples that are unseen during training, called out-of-distribution (OOD) data, relying on the outputs of the pre-trained learning-based methods involves the risk of inaccurate predictions during inference (actual testing). Most works focus on algorithms that improve the accuracy of the prediction and seldom address the issue of (i) robustness of the developed algorithm to unseen data acquisitions and (ii) han-

dling the inaccurate predictions by quantifying the uncertainty in prediction. Hence, for reliable translation, it is imperative to evaluate the robustness of the proposed machine learning-based methods to newer OOD data acquisitions. Furthermore, because it is difficult to account for all types of OOD data acquisitions, incorporation of mechanisms to quantify the risks underlying the developed models/frameworks can potentially act as a proxy for residual error maps, given the non-existence of reference images. Providing information about uncertainty during inference can help the scientists/radiologists make crucial decisions when high uncertainty in the output is reported, e.g., (i) whether there is a need to repeat the scan and (ii) augmentation of the existing training set to improve the accuracy of the model to such unseen cases.

In this way, with the combined developments in both detector hardware and learning-based solutions for improved image quality, the idea of PET imaging involving radiation-levels comparable to annual radiation-levels from radionuclides that are part of our daily life could become a reality (**?**) with the added advantage of faster scans. This thesis focusses on developing learning-based methods for image quality improvement in simultaneous PET-MRI neuroimaging under different acquisition scenarios explained in the following chapters.

1.5 Organization of the Thesis

The work presented in this dissertation describes the novel image reconstruction and image restoration methods developed in the context of simultaneous PET-MRI systems that address some of the existing challenges mentioned in the previous sections. The thesis is organized as follows.

Chapter 2 provides a brief background of PET-MRI followed by a short literature review on the prior works on image quality improvement for both PET and MRI images obtained from dedicated scanners as well as the simultaneous PET-MRI system. The chapter explores both (i) within-reconstruction and (ii) post-reconstruction enhancement techniques. Chapter 3 describes a novel joint PET-MRI dictionary method developed for Bayesian reconstruction of both PET and MRI images under different image acquisition scenarios. Chapter 4 describes a novel DNN framework developed for enabling ultra-low-dose PET imaging that is robust to OOD PET acquisitions. Chapter 5 details an image restoration method leveraging MRI information for improved temporal resolution in fPET imaging, which is a novel neuroimaging application for improved mapping of brain function. Chapter 6 summarizes the developed models and concludes the thesis indicating directions for future work.

Chapter 2

Background and Related Work

As introduced in the previous chapter, current simultaneous PET-MRI systems enable the synergistic acquisition of both structural and functional information and provide the opportunity for multidimensional neuroimaging unparalleled by other combined systems (**?**). The superior specificity of PET combined with the excellent soft-tissue contrast and functional imaging capability offered by MRI provides the opportunity to understand several biological processes in neuroscience and underlying pathophysiology of several diseases in neurology, cardiology, and oncology. For example, in neuroscience, we can gain insights into energy consumption while resting and while performing specific tasks (**??**). The simultaneous acquisition of PET and MRI data allows joint modeling of spatial as well as temporal data which is *not* possible in sequential systems. Furthermore, the combination of PET and MRI can potentially provide improved insights in several neuropsychological processes such as changes in hemodynamic responses and quantitative changes in radiotracer uptake, etc. (**?**). On the other hand, for clinical neurology, the simultaneous data from PET and MRI potentially enables improved theranostics. For example, the combined measurements of tumor vasculature and tracer uptake can further our knowledge of cancer biology and tumor proliferation, and more importantly, enable monitoring of the individual's response to therapy. The current state of simultaneous PET-MRI systems results from continuous developments in the field of medical imaging over the past few decades.

2.1 Historical Overview

The underlying physics of both PET and MRI finds its origins in the investigations into nuclear spins in the late 1920s and 1930s. While the idea of nuclear spin motivated Dirac to comment on the "discrepancies" in his equations (**?**), I Rabi discovered that magnetic

resonance occurred when external energy is provided to a nuclear spin system near the Larmor frequency (**?**). Several years later, the first human PET scan was achieved by E Hoffman, M Ter-Pogossian, and M Phelps at Washington University who built the first PET camera (**?**), followed by the introduction of the first whole-body PET scanner in 1977(**?**). Subsequently, several groups were motivated to study the energy consumption in the human brain, and the first results on quantification of cerebral glucose metabolic rate using FDG-PET were reported in **?**.

For MRI, the initial years since the discovery of nuclear magnetic resonance (NMR) was mainly employed for the analysis of chemical compounds (NMR spectroscopy). However, the discovery that relaxation times differed across tissues by R Damadian in the 1970s provided the necessary motivation to use MRI to study diseases (**?**). P Lauterbur later generated the first 2D and 3D MRI images using spatially encoded gradients, subsequently leading to the first scan of the human body using MRI in 1977 (shortly after the advent of CT in 1973 by G Hounsfield). Around the same time, P Mansfield proposed the technique of echo-planar imaging (EPI), which was realized on scanners almost after a decade. The EPI and other related techniques enabled several MRI modalities with clinical application such as dMRI and fMRI (**?**). While functional brain PET imaging was being conducted using various radiotracers such as FDG and H_2O[15], it was observed that MRI can also detect a substantial change in blood flow. This marked the birth of fMRI. Since then, fMRI has become a cornerstone of neuroimaging due to its non-ionizing radiations, and superior spatial and temporal resolution, compared to other functional imaging techniques.

Over the last four decades, although the initial years saw the improvements of the dedicated PET and MRI scanners, there was a need to combine PET (and also other molecular imaging systems) with structural imaging modalities due to its low resolution and poor SNR in the images (**?**). Although software-based image fusion/registration and post-processing was pursued at various clinics and research labs, there were several drawbacks to this approach due to inaccurate image registration, change in patient physiology etc. The introduction of the combined PET-CT system changed the the landscape of medical imaging in the late 1990s. The success of PET-CT and the known benefits of combining MRI with PET encouraged the development of the PET-MRI scanners. The first commercial simultaneous PET-MRI scanner became available in 2011. Since then, PET-MRI systems have opened up new research opportunities and enhanced clinical diagnosis for several imaging applications. However, the full potential of PET-MRI has not yet been achieved.

2.1.1 PET-MRI Systems

The earliest works on combining emission tomography with structural modalities like CT was realized with the introduction of the simultaneous SPECT and CT in the 1990s (**?**). The efforts to combine PET and CT were ongoing in parallel, following which, in 2001, the first commercial PET-CT was introduced.

The combination of PET and CT into one single instrument enabled obtaining metabolic and structural information in a single imaging session. The existence of CT in the same imaging system solved the issue of attenuation correction for PET data to a large extent and also eliminated the use of transmission scans within the PET subsystem(**?**). The subsequent years witnessed large-scale adoption of the PET-CT systems to such an extent that dedicated PET scanners were hardly produced. In addition to addressing the issue of attenuation correction, CT scans could also be used for their diagnostic value. Hence, PET-CT became widely adopted across clinics for whole-body oncology (**?**). The significantly shorter acquisition times in CT (compared to PET) rendered the development of simultaneous PET-CT systems unnecessary. Hence, in the design of PET-CT systems, the PET and CT subsystems are positioned one behind the other with a common patient bed. Consequently, there is a possibility of substantial patient movement between acquisitions further resulting in artifacts in the reconstructed images (**?**).

Given the superior soft-tissue contrast and resolution of MRI compared to CT, the efforts to combine PET and MRI started around the mid-1990s. Unlike PET-CT, combining PET and MRI faced several challenges. In the combined PET-CT system, the PET and CT subsystems are relatively more similar to their dedicated counterparts. However, for combining PET and MRI into a single system, the interference of the magnetic field and the performance of the PET electronics (involving photomultiplier tubes (PMTs)) proved to be a major hindrance. The search for the ideal MRI compatible PET detectors gave rise to different generations of the PET-MRI detectors: (i) avalanche photodiodes- (APDs) based PET detectors were insensitive to large magnetic fields but demonstrated a substantial loss in signal gain compared to PMTs; (ii) silicon photomultipliers (siPMs) based detectors operating in Geiger mode enabled time-of-flight (TOF) imaging, and finally, recent developments in (iii) digital siPMs (dSiPMs) while providing the desired gain, can adversely affect MRI images (**?**). In the state-of-the-art PET-MRI scanners, while the PET system enables an isotropic spatial resolution of 3–4 mm, the 3 T MRI system can easily achieve an isotropic resolution of around 1 mm.

Simultaneous PET-MRI offers several advantages over the PET-CT systems. First, the non-ionizing radiations from MRI make it the modality of choice for longitudinal scans, as well as for imaging of radiation-risky populations such as children and pregnant

women. A typical PET-CT scan might induce a radiation dose of 14 mSv to 32 mSv with the the contribution from CT going up to or more than≥ 50% (**?**). Secondly, MRI offers a wide variety of contrasts ranging from structural images (e.g., T1w, T2w) to functional images (e.g., diffusion MRI (dMRI), fMRI, MRS). Thirdly, the truly simultaneous acquisition of PET and MRI data (as opposed to the sequential acquisition in PET-CT) provides opportunities for improved accuracy in patient motion correction. Finally, MRS in combination with PET can potentially enable spatial matching of biochemicals, which in turn helps in inferring metabolic status for various pathologies.

2.1.2 PET-MRI for Brain Imaging

The PET-MRI imaging is well-suited for brain imaging where typically linear image registration algorithms suffice for image registration between PET and MRI (**??**). Scans that are conducted on separate (or dedicated) scanners involve the issue of differential positioning of the patient. Hence, active research for brain imaging focuses on improving data processing pipelines that lead to better joint analysis of the PET and MRI data leveraging complementary information. The joint analysis of spatiotemporal data from PET and MRI brain imaging is aided by the fact that typically, spatial alignment of the PET and MRI images is not a challenging task (in comparison to imaging of other body parts) due to reduced motions in the head. However, for longer PET scans (spanning several minutes) it is difficult to control patient movement and hence intermittent MRI scans involving navigator sequences aid in accurate motion correction during image reconstruction (**?**). For attenuation correction, the development of ultrashort echo time (UTE) sequence has provided a non-ionizing replacement for attenuation correction and has been shown to perform with desired clinical accuracy for brain imaging(**?**).

Besides improved motion and attenuation correction, the simultaneous/non-simultaneous acquisition of MRI from a single system can address the inherent shortcomings in the PET images such as partial volume errors and low SNR of the PET images. Prior to the arrival of simultaneous PET-MRI systems, partial volume correction techniques relied on segmentation of the MRI image than image-registration-based techniques. Given the spatial co-existence of the PET and MRI images, segmentation-free approaches have come to the fore to address image quality improvement tasks reducing the propagation of errors due to the MRI segmentation algorithms (**???**). In addition to improvements due to spatial alignment, the joint spatiotemporal analysis of dynamic PET and MRI data is useful in applications such as tracer kinetic studies, drug development, and several other aspects of neuroscience research. Unlike brain imaging in PET-MRI, whole-body PET-MRI imaging still requires hardware-related and data-processing-related

optimization across multiple bed positions, improved information from MRI for PET attenuation, motion, and scatter correction.

This dissertation focuses on methods and algorithms developed in the context of brain imaging. Specifically, we address the issue of PET image quality improvement using MRI information under different conditions. To this end, we consider only FDG-PET images in this thesis. As explained in the following sections, several acquisition scenarios give rise to PET images with poor SNR in addition to the inherent limitations of resolution and partial volume effect. This thesis discusses novel learning-based models which exploit anatomical information from MRI to improve PET image quality. Although the developed models are generic enough to be easily extended to other anatomies, optimization of the methods to specific anatomies is indeed necessary.

2.2 Towards Ultra-Low-Dose and Accelerated PET-MRI Brain Imaging

As mentioned in the previous section, the simultaneous acquisition of PET and MRI data provides opportunities for optimization of the scanning protocol and various data processing techniques by leveraging the complementary information at various stages of data flow e.g., dynamic motion correction, attenuation correction, simultaneous fMRI and functional PET studies, etc. Besides hardware and imaging protocol developments for brain imaging using PET-MRI, active research now focuses on further optimization such as (i) further reduction in ionizing radiations from PET (low-dose imaging), (ii) faster scans to improve patient comfort as well as scanner throughput.

2.2.1 Towards low-dose PET imaging

Typical PET acquisitions involve providing a bolus injection of the radiotracer to the subject and allowing a certain uptake period (usually 30–60 minutes) outside of the scanner before the PET data is acquired. Typically, PET data is acquired in the form of list-mode data spanning the entire scan duration, say 20 minutes. The list-mode data entail the continuous set of recorded coincidence events across the PET detectors. Subsequently, during the PET image reconstruction, the acquired list-mode data is then grouped into several bins of desired time interval T_{bin} resulting in either a dynamic sequence of PET images or a single static PET image.

A static PET image represents a single PET image that represents an integral of radiotracer activity over the entire scan duration. A static PET image can be reconstructed by specifying a long T_{bin} that spans the entire duration of PET data acquisition. Typically,

the long acquisition time for PET indeed leads to better image quality (in terms of SNR) in comparison to multiple short fragments T_{bin} over the entire scan duration. However, we lose the quantitative information related to the dynamic uptake of the tracer, and hence, static PET provides only semi-quantitative information such as the standardized uptake value (SUV) maps. On the other hand, dynamic PET image reconstruction can be performed by dividing the overall acquisition time into desired multiple T_{bin} intervals. The length of T_{bin} determines the temporal resolution for the dynamic PET image sequence. Because the PET photon-counts follow a Poisson distribution, a lower T_{bin} corresponds to higher temporal resolution but poorer SNR in PET data. Another factor of consideration is that the uptake duration (from time of radiotracer injection to the beginning of the scan) contributes to loss in information regarding the true underlying dynamics of the radiotracer from the time of injection. The study of the distribution of the radiotracer right from the start of injection provides a quantitative picture of the radiotracer metabolism while the subject is at rest or performing a certain task. To circumvent this issue, a new imaging protocol involving continuous infusion of the radiotracer has been recently studied for functional PET imaging (Section 2.3).

2.2.2 Towards accelerated MRI

It is imperative to optimize the PET-MRI acquisition and processing for improved throughput of the scanner. Along with certain commonly acquired structural scans, certain studies might benefit from the acquisition of other functional studies such as fMRI and/or dMRI. While dosage optimization is crucial for reducing induced ionizing radiations to the patients' safety and can potentially enabling PET imaging for children and pregnant women, acceleration of MRI acquisition is needed for several reasons described as follows. First, typical acquisition protocols in simultaneous PET-MRI imaging acquire multiple contrasts of MRI images, e.g., T1w, T2w, UTE, Dixon contrast (in case of whole-body imaging) (**?**), etc. The multi-contrast MRI images are of use in various stages of a typical PET image reconstruction and processing pipelines, e.g., creating attenuation maps (**?**), estimating patient motion with high temporal resolution (**?**), PET image restoration (**??**), etc. Second, as mentioned above, certain neuroimaging studies acquire fMRI or dMRI images simultaneously with PET images (**?**), where the fMRI and dMRI scan acquisitions can be much longer than typical MRI scans and longer than the PET acquisition. Third, to better estimate patient motion (a critical issue in PET imaging (**?**), anatomical MRI scans are repeated throughout the PET imaging session. However, the temporal resolution of the motion estimates is limited by the MRI sequence acquisition times. In general, faster scanning in MRI and PET, reduces the risk of motion artifacts in

the reconstructed images of patients. Fourth, acquiring multiple MRI contrasts can often exceed the time required for a PET scan, in a single bed position, risking significant idle time for the PET imaging component (**?**). Typically, accelerated MRI data acquisition is achieved by either (i) undersampling the k-space data or (ii) parallel imaging using multiple channels, detailed in the paragraph below. However, both these strategies for accelerated MRI data acquisition result in the degradation of image quality of the reconstructed MRI images.

Undersampled MRI data acquisition. For MRI, the image resolution is dependent on the number of points in the k-space. Hence, for high-resolution images, dense sampling of k-space is required, consequently leading to longer acquisition times. Typical MRI acquisitions involve sampling of the k-space in the Cartesian domain. Hence, the major determining factor of acquisition time is the number of phase-encoding lines. Faster data acquisition is achieved by skipping certain pre-determined phase-encoding lines. In this case, the acceleration factor is proportional to the number of skipped phase-encoding lines. Although the implementation of the Cartesian undersampling scheme is more convenient in comparison to non-Cartesian trajectories (e.g., radial or spiral), the achievable acceleration factor without significantly degrading the reconstructed image quality is limited(**?**). This is due to coherent artifacts resulting from Cartesian undersampling. The realization of non-Cartesian trajectories such as radial or spiral trajectories provides greater acceleration without the introduction of coherent artifacts. A comprehensive theoretical framework about the same is presented in **??**. For simplicity, this dissertation uses only the easily-implementable Cartesian trajectory(**?**). Nevertheless, as explained in subsequent chapters, the methods developed can be extended to other non-Cartesian trajectory acquisition strategies for MRI.

Parallel imaging with undersampled MRI data acquisition. Parallel-imaging, i.e, simultaneous acquisition of MRI data using multiple receiver coils/channels improves the *ill-posedness* of the inverse problem associated with undersampled MRI data acquisition by obtaining more measurements. In the parallel imaging scenario, multiple set of coils is used to image the same anatomical region simultaneously and each coil generates the full FOV image based on its own sensitivity(**??**). The final MRI image is reconstructed using the sensitivity encoded images obtained from each coil as well as the inherent sensitivity of the coils. For example, if two channels are available for parallel imaging and if the k-space data is undersamped by a factor of two, consequently, an acceleration factor of 2 $\times$ is achieved. Nevertheless, to determine the value of the reconstructed image at a particular voxel location, the measurement from the two channels can be leveraged. This improves the *ill-posedness* of the inverse problem (image reconstruction) to be solved.

However, the improvement depends on the amount of acceleration as well as the number of coils used for imaging. Parallel imaging techniques such as sensitivity encoding for MRI (SENSE) (**?**), GRAPPA (**?**), or a combination of both (**?**) aid in the reduction of scan time without affecting image quality.

2.2.3 Raw-Data Corrections

Attenuation and Scatter Correction. The acquired PET data needs to be corrected for attenuation and scatter of photons while traveling through the biological tissues as well as the scanner hardware. Although estimation of attenuation coefficients using MRI is a challenging problem, through the development of MRI imaging sequences such as UTE, substantial progress has been made in the area of estimating accurate attenuation maps for brain imaging (**??**). While the estimation of attenuation maps for whole-body PET-MRI remains an active research area, for applications involving adult brain imaging with an intact skull, attenuation correction can be assumed to be "finally solved to a level of accuracy sufficient for the majority of clinical applications and/or to a level comparable to that seen in PET/CT, i.e. overall uncertainties seen with PET/MRI are no worse than those seen in PET/CT and were considered to be clinically acceptable."(**??**). Similarly, scatter correction "has not been a major issue to be addressed for neurological PET/MR imaging"(**?**). However, improved accuracy of attenuation and scatter correction algorithms result in improved quantification in PET images.

Motion Correction. One of the major advantages of the simultaneous MRI acquisition is the ability to provide information about patient motion, which in turn aids in PET image improvement. Typical PET-MRI acquisitions last for about 20 minutes or in the case of continuous infusion studies, up to an hour or more. Hence, it is common for the subject/patient being scanned to move their heads over the course of the scan in addition to other actions/activities such as coughing, talking or falling asleep (**?**). For neuroimaging, given the enclosure of the brain within the skull, a slight movement of the head does not affect static PET images. For dynamic PET imaging, prolonged motion could affect image registration and cause blurring effects in the PET images (**?**). Nevertheless, the effect of head motion for brain imaging is less severe than studies involving involuntary organ motion, such as in cardiac or pulmonary imaging.

2.2.4 Image Quality Improvement

The intrinsic spatial resolution of radioactivity in PET, typically 4-6 mm **?**, is far lower than the anatomical resolution in MRI, typically 1-1.5 mm. Spatial resolution in

PET is limited by positron range, non-collinearity of the annihilation photons, scatter inside the scintillation crystals, finite crystal dimension, interaction depth, etc. PET reconstruction is challenged by the stochasticity in gamma-ray emission, and thus, it typically relies on *statistical prior* models on the MRI and PET images. The low spatial resolution results in partial volume effects. That is, activity in a particular voxel location contains activity from neighboring tissues as well. As a result, this causes under-estimation of high activity (due to spill-out) which is critical in clinical diagnosis. PET images also typically have lower SNR (**???**) because of the high levels of stochasticity in the photon-emission processes. The photon-counts in PET follow a Poisson distribution with additional noise components resulting from the scanner electronics (randoms etc.).

Classical reconstruction methods like the expectation-maximization (EM) based maximum-likelihood (ML) estimation for PET reconstruction **?**, and several others use a Poisson noise distribution. Over the years, several methods for improving PET images have been developed. These works address the problems of both partial volume effects and noise removal using different computational modeling techniques. Primarily, these techniques can be divided into (i) within-reconstruction techniques and (ii) post-reconstruction techniques. While the post-reconstruction techniques model the image enhancement task as a regularized deconvolution problem that involves the point spread function (PSF) of the PET system, the within-reconstruction techniques model propose a regularized reconstruction model which involves the PET forward model. In both cases, many works have explored the use of anatomical information from MRI to achieve improved results compared to non-regularized schemes as well as regularization schemes depending on the PET image alone. Specifically, the methods developed and that are included in this thesis focus on leveraging statistical dependencies across the FDG-PET and structural MRI images (e.g., T1w and T2w). For the purpose of development of different image enhancement algorithms within the scope of this thesis, we assume that the PET data has been corrected using the typical scanner-provided estimates for attenuation, scatter, and motion correction. As mentioned in the previous section, several studies have indicated that for neurological studies, the problems related to attenuation, scatter, and motion correction are less critical for adult brain imaging (the focus of this thesis), in comparison to whole-body PET-MRI(**??**).

2.3 Functional- PET (fPET) Imaging

The human brain accounts for roughly one-fifth of total-body energy consumption (**?**). Deep structural and functional insights into the understanding of the human brain can

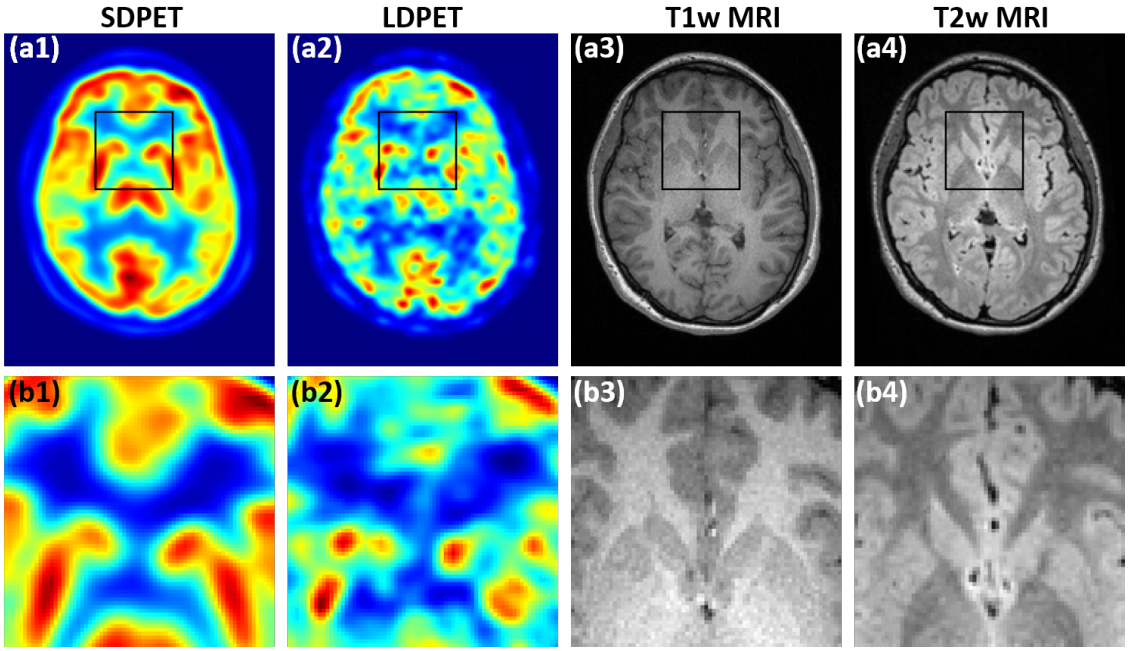

Figure 2.1: **Visualization of low-dose and standard-dose FDG-PET images along with corresponding T1w and T2w MRI images.** Top: **(a1)**:standard-dose PET image (SD-PET); **(a2)**:low-dose PET image (LDPET) at DRF = 180×; **(a3) and (a4)**: T1w and T2w MRI images; Bottom: **(b1)–(b4)**: corresponding zoomed region of interest. The zoomed region of interest is highlighted using the black rectangle box in (a1)–(a4).

provide clarity in terms of neural basis of cognition and various pathophysiologies (**?**). The understanding of the relationship between functional and metabolic connectivity can reveal intricate details about neuronal responses. Consequently, the defects therein can potentially predict brain disorders such as Alzheimer's disease. Although FDG-PET was being used to study cerebral glucose metabolism (which in turn relates to synaptic transmission) since the late 1970s, the emergence of fMRI in the 1990s marked a paradigm shift in studying functions of the brain. Unlike PET, fMRI does not involve ionizing radiations and provides superior spatial as well as temporal resolution. The contrast in fMRI images is blood oxygenation level-dependent (BOLD). However, BOLD-fMRI provides only an indirect and non-quantitative measure of neuronal activity. The BOLD contrast is dependent on several factors such as cerebral blood flow, volume, and oxygenation. Furthermore, physiological factors such as breathing, heart rate, blood-flow volume, etc. affect the BOLD-fMRI signal. Consequently, there is increased variability in the functional analysis of individual subjects using fMRI. For example, multiple runs of the same fMRI protocol for experiments involving short task durations have shown increased vari-

ability in SNR (**?**). Hence, understanding the dependency between neuronal activity and the measured signal is challenging **?**. Furthermore, comparing fMRI data across healthy individuals and individuals with pathology might be misleading due to the noise introduced by physiological confounding factors. On the other hand, although FDG-PET can provide a quantitative estimate of the rate of glucose metabolism, the static PET scans provide an integral effect of glucose utilization over a long period of time.

As mentioned in Section 2.2.1, dynamic PET imaging using a bolus administration of radiotracer provides an opportunity to study dynamic uptake of the tracer in the brain. However, conventional bolus injection FDG PET scans are not sensitive to cerebral metabolic changes over an extended time duration due to the lack of sustained supply of FDG to the brain (**?**). To circumvent this problem, **?** used a continuous infusion radiotracer infusion approach, together with dynamic PET scanning, to achieve enhanced sensitivity for tracking dynamic radiotracer uptake. This constant infusion approach using FDG was labeled as 'functional' PET, to highlight similarities to the fMRI technique. Subsequent research using fPET methodology has shown promising results for isolating functional brain areas during external tasks and at rest (**????**) Despite the improvement in temporal resolution in comparison to the conventional bolus approach, the temporal resolution of fPET remains substantially lower than that of fMRI, which is of the order of seconds. Hence, even with the developments provided by the novel imaging technique, the current temporal resolution of fPET (around 20–60 seconds) limits the opportunity to use fPET for detailed investigations of brain metabolic responses to rapidly switching tasks and brain stimulation paradigms.

The simultaneous PET-MRI system provides a unique opportunity to understand the complex relationship underlying glucose metabolism and blood flow. While PET imaging can provide information about metabolic connectivity, fMRI provides information about functional connectivity. Thus, with the advent of the fPET imaging technique, the interrelationship between functional connectivity and metabolic connectivity can be studied. To this end, while one study found stronger agreement between functional activation and metabolic activity (**?**), another study reported that there are differences in functional activations obtained via fMRI and the glucose uptake over time (**?**). Another study in this direction has already established that the activation maps recovered from dynamic PET imaging and fMRI provide partially complementary information (**?**).

In the following sections, we will describe specific problems related to image quality enhancement of PET and MRI arising from different acquisition scenarios discussed in the previous sections. We will also describe the prior works for such problems and their limitations motivating the methods developed as part of this thesis.

2.4 Image Reconstruction

PET and MRI images differ in structure, contrast, and resolution because of fundamental differences in the imaging mechanisms underlying these two modalities. For instance, compared to MRI images, PET images have significantly larger voxel sizes because of the underlying limitations of PET detectors. Figure 2.1 shows examples of a standard-dose FDG-PET (SD-PET) 2.1(a1), low-dose FDG-PET (LD-PET) 2.1(a2) with 180× reduced counts and the corresponding spatially co-registered T1w and T2w MRI images (2.1(a3) and 2.1(a4)), respectively. Although being different modalities, statistical dependencies are indeed present across the PET and MRI images, depending on the radiotracer employed for PET imaging. For example, FDG-PET shows greater anatomical correspondence with structural MRI images in comparison to other radiotracers such as amyloid PET.

The anatomical information can be leveraged either by segmenting the MRI image or by relying on co-registration between the PET and MRI images. Given the simultaneous acquisition of PET and MRI within a single system, employing segmentation-free methods are preferred to avoid error propagation from inaccurate segmentation estimates (**?**). Furthermore, the anatomical MRI prior based PET reconstruction methods can address both the issues of partial volume effects as well as the noise in PET images. However, the PET raw data is not always readily accessible and hence, the post-reconstruction image enhancement techniques are still of interest. Finally, although in the context of PET-MRI, improvement of PET images is undoubtedly obtainable via MRI images, recent studies have shown that MRI images can also benefit from joint reconstruction techniques. Over the past few years, there has been an increased interest in simultaneously reconstructing PET and MRI images by exploiting the dependencies across both images to enhance the quality of both PET and MRI images. This emphasized the need for joint modeling of priors that leverage dependencies across PET and MRI dependencies. However, while modeling joint priors, it is important to carefully model the dependencies across the PET and MRI images to minimize the cross-modality artifacts.

Prior-based PET image reconstruction Prior based image reconstruction in medical imaging has been studied for the past few decades. Let u represent the PET activity image, and v represent the structural MRI image. Let operators $\mathcal{H}$ and $\mathcal{F}$ represent the forward operators corresponding to PET and MRI imaging, respectively. Let y represent the acquired PET data and z represent the acquired MRI data. In the context of medical image reconstruction using priors, the optimal PET image reconstruction given a high-

quality MRI image is obtained by

$$u^{\text{optimal}} := \arg\min_{u} \mathcal{D}_{\text{PET}}(\mathcal{H}u; y) + \alpha\mathcal{R}(u, v) \tag{2.1}$$

where $\mathcal{D}_{\text{PET}}(\cdot)$ denotes the data-fidelity term corresponding to the PET forward problem and $\mathcal{R}(\cdot)$ denotes the regularization function that depends on u and v. Early methods exploit such dependencies to address problems specific to PET imaging (**??**), e.g., to correct for attenuation and partial-volume effects, to improve SNR, etc. Equation 2.4 can be modified to accommodate MRI reconstruction with the joint PET-MRI penalty function $\mathcal{R}(\cdot)$ in which case the reconstruction problem becomes

$$v^{\text{optimal}} := \arg\min_{v} \mathcal{D}_{\text{MRI}}(\mathcal{F}v; z) + \alpha\mathcal{R}(u, v). \tag{2.2}$$

Here, $\mathcal{D}_{\text{MRI}}(\cdot)$ represents the data-fidelity term corresponding to the MRI forward problem. The above mentioned single-modality reconstruction problem can be further extended to incorporate joint reconstruction of PET and MRI images as

$$(u^{\text{optimal}}, v^{\text{optimal}}) := \arg\min_{u,v} \mathcal{D}_{\text{PET}}(\mathcal{H}u; y) + \mathcal{D}_{\text{MRI}}(\mathcal{F}v; z) + \alpha\mathcal{R}(u, v) \tag{2.3}$$

We will now discuss methods proposed in the literature for individual PET and MRI reconstruction followed by joint reconstruction priors along with their advantages and disadvantages.

2.4.1 PET Reconstruction from Noisy Data

Clinics to-date still employ the classical MLEM **?** reconstruction method or its variants, which use statistical modeling of PET measurements. The MLEM based PET reconstruction is noisy, and hence a post-reconstruction Gaussian smoothing operation is performed for better visualization. Methods in **?**, **?**, **?**, employ simple Markov Random Field (MRF) priors such as quadratic or Huber prior or use structural MRI images to improve PET image quality. While quadratic MRF priors might result in over-smoothed reconstructions with loss of features, edge-preserving priors such as Huber or TV might result in staircase-like artifacts. Another class of MRF priors involves introducing regularity on a patch rather than local image gradients. Since resolution loss is inherent in PET images due to detector limitations, priors learned from PET images alone might be ineffective in denoising without loss of information **?**. Mutual information-based prior was also reported in **?**, which is based on the joint histogram of PET image and the corresponding MRI image, but does not enforce neighborhood constraints. In **?**, the penalties for a patch of voxels in PET is learned from a fixed MRI image, and similarly, in **?**, a patch-based dictionary learned from MRI image was used to fit to PET data.

2.4.2 MRI reconstruction from Undersampled Data

Acceleration of MRI k-space data acquisition continues to be an active area of research (**????**). As mentioned earlier, routine clinical scans typically employ Cartesian trajectories, with (i) dense sampling near the center of the k-space and (ii) sparser sampling in the high-frequency regions (**??**). As described in the Introduction section, many works exploit the sparsity of the signal representations using wavelet frames (**??**), dictionaries (**???**), or apply TV-based constraints (**???**). Recent works rely on the alignment of image gradients across multiple MRI contrasts (**??**). On the other hand, patch-based methods use sparse dictionary representations (**??**) to achieve higher undersampling rates under low-SNR conditions. **?** gives a detailed account of undersampled MRI reconstruction in the clinic.

2.4.3 Joint PET-MRI Reconstruction

Pioneering work by **?** models gradient-based dependencies (fine-scale, since they operate on immediate neighboring voxels) to perform joint reconstruction of (simulated) PET-MRI images and improve the quality of both PET and accelerated-MRI reconstructions. The method proposed in **?** was named parallel level sets (PLS), where the regularization term penalizes misaligned image gradients in PET-MRI at every voxel. In the same work, two variants of PLS, namely, linear PLS (LPLS) and quadratic PLS (QPLS) were described. For the pair of real-valued images $f(\cdot)$ and $g(\cdot)$ on spatial domain Ω, the general PLS regularization term involves a penalty of the form

$$R(f, g) := \int_{x \in \Omega} D(\nabla f(x), \nabla g(x)) dx, \tag{2.4}$$

where $D(\cdot, \cdot)$ denotes the dissimilarity between the images gradients $\nabla f(x)$ and $\nabla g(x)$. **?** report that while LPLS leads to sharper images, QPLS, which employs the squared norms, is more stable and can reconstruct shared edges better than LPLS. For QPLS, they define

$$D(\nabla f(x), \nabla g(x)) := [1 + |\nabla f(x)|^2 |\nabla g(x)|^2 - \langle \nabla f(x), \nabla g(x) \rangle^2]^{0.5} \tag{2.5}$$

For LPLS, they define

$$D(\nabla f(x), \nabla g(x)) := |\nabla f(x)||\nabla g(x)| - \langle \nabla f(x), \nabla g(x) \rangle. \tag{2.6}$$

For both QPLS and LPLS, they use regularized gradient norms to compute $|\nabla f(x)|$ and $|\nabla g(x)|$. Later methods (**??**) further enhanced the gradient-based prior models to improve *in vivo* PET-MRI joint reconstruction. Similarly, **?** proposed total generalized variation (TGV) that penalizes, at every spatial location, the sum of the singular values of the

Jacobian of the 2-vector-valued PET-MRI image. Effectively, TGV penalizes misalignment in gradients, as in PLS. Both TGV and PLS introduced regularization terms that depend on the magnitude and orientations of the image gradients at corresponding locations in the PET and MRI images. Very recently, **?** proposed another TV-based prior, based on joint sparsity regularization (JSR). Gradient-based priors are best suited for piecewise-smooth images and can fail to model textural patterns in images (**??**). On the other hand, patch-based prior models, including dictionary models, can preserve textures and edges better than gradient-based priors, even under low SNR, as demonstrated in several denoising applications (**??????**). Although patch-based priors have improved independent reconstructions of PET (**??**) and MRI (**??**), methods using patch-based priors for joint reconstruction are virtually absent in the literature. Patch-based methods are able to capture higher-order features such as texture because they operate at a larger context than fine-scale features such as image gradients. Moreover, given the distinctive appearances of PET and MRI images, enforcing the alignment of image gradients has a higher risk of infusing cross-modality artifacts compared to patch-level dependencies. Thus, while patch-based methods potentially improve the robustness to noisy data, and joint modeling aids in image quality improvement of PET and MRI, a joint patch-based method can potentially be robust to noise and reduce cross-modality artifacts by enforcing patch-level dependencies across PET and MRI instead of dependencies at the voxel-level.

2.5 Image Enhancement for Static Low-dose PET

Post-reconstruction PET image quality enhancement techniques range from traditional filtering using Gaussian smoothing to complex DNN architectures. While works using statistical models to exploit information from the noisy PET image alone, improve over Gaussian smoothing **??**, works that jointly model dependencies across co-registered PET and MRI images show substantial improvement in PET image quality **?**. Recently, **?** proposed an unsupervised model for PET image denoising by employing a conditional deep image prior (DIP) that uses the subject's anatomical MRI or CT as the input to the DNN. These methods focused on denoising the PET image and do not involve a reduction in the dose.

Post-reconstruction image quality enhancement with learning-based models. Initial work led by **??** showed that learning-based approaches, such as regression forests and sparse modeling using dictionaries can synthesize SD-PET images from LD-PET images at a DRF of around 4×.

For the same DRF, (i) **?** proposed a DNN that used an auto-context strategy to estimate patches in the SD-PET image based on the patches in the input set of LD-PET and T1w MRI images; and (ii) **?** employed a generative adversarial network (GAN) based framework. The input to the generator of the GAN was a single channel image obtained by fusing the multimodal MRI images and the LD-PET images. **?** used a standard ResNet architecture to learn a mapping between the noisy LD-PET and SD-PET images (DRF = 1) that included fine-tuning of the DNN using a VGG-based loss **?**. For substantially higher DRF, pioneering work by **?** showed that by using a 2.5D-style PET-only input to the DNN that estimates the residual between the LD-PET image and the reference SD-PET image, it is possible to achieve a DRF of around 200×. Subsequently, the work in **?** showed that, with a similar architecture and training strategy, the inclusion of MRI images as multi-channel input produced better image quality than using PET images alone. The work in **?** showed that learning a mapping between the LD-PET sinogram data and the SD-PET sinogram data can lead to some improvement in the reconstructed SD-PET images in comparison to the strategy of learning the mapping from LD-PET to SD-PET in the spatial image domain. However, as mentioned earlier, the measured raw sinogram data might not always be accessible. On the other hand, we can easily obtain an estimate of the sinogram of the PET images using a simulated system matrix generated using the knowledge of scanner geometry and simulation tools such as GATE **?**. The retrospectively obtained sinogram approximates the sinogram that is obtained after typical data correction steps applied to the PET raw data.

The prior work discussed above employ loss functions either exclusively in the spatial image space, or exclusively in the sinogram space, but *not* both. Several DNN-based methods for undersampled MRI reconstruction have shown that including a transform domain (k-space) loss function in addition to the image space loss function improves the quality of reconstructed images at higher undersampling of MRI data **??**. Additionally, the proposed methods including those for PET and MRI reconstruction, do *not* evaluate the models for robustness to out-of-distribution (OOD) data in new acquisitions, which are important for clinical translation. Further typical PET reconstruction methods seldom quantify the uncertainty in the prediction provided by the DNNs. Modeling uncertainty in DNNs can potentially (i) inform the radiologist about the imperfections in reconstructions that may be crucial in clinical decision making or subsequent automated post-processing of reconstructed images; and (ii) provide improved performance when the DNN is presented with OOD data **??**.

2.6 Improved Brain Activation at High Temporal Resolution Low-Dose fPET

Analysis of fPET data is challenging because of low SNR and partial volume errors in the reconstructed PET images (**?**). Recent works have improved the SNR in fPET by applying a combined bolus and continuous infusion of radiotracer during fPET data acquisition (**??**). However, the statistical power of these experimental approaches is still relatively low when compared with fMRI. To mitigate this issue, spatial denoising of the reconstructed PET images is performed prior to the functional analysis of the brain using techniques such as independent component analysis (ICA). Gaussian smoothing is widely used as a post-reconstruction spatial and temporal smoothing operation for functional neuroimaging analyses (**?????**). However, the Gaussian kernel acts as a low-pass filter. Therefore, it further worsens the partial volume errors in fPET images; this can cause errors in the localization and quantification of brain functional activations at high-temporal resolution fPET imaging. MRI-based PET reconstruction methods have shown substantial improvement in PET image quality compared to conventional methods (**???**). For instance, several studies have explored post-reconstruction PET image enhancement using anatomical information from structural MRI (**???**) to perform partial volume correction and image deblurring (**??**). These methods propose a regularized deconvolution problem formulation using MRI information to address the PET denoising and partial volume error problems. Works involving anatomical information as regularization terms include penalties that model the statistical dependencies across the PET and MRI images, such as joint entropy regularization.

? proposed a variant of the well-known Bowsher prior (**?**), modeled as prior information in the reconstruction process. The Bowsher prior, in principle, is a weighted Markov random field (MRF) model that promotes delineation of PET image voxels that are dissimilar according to the intensities in the spatially co-registered MRI image. The weights are computed based on a similarity metric (e.g. absolute difference) evaluated on the structural image. Subsequently, **?** proposed an asymmetrical variant of the original Bowsher prior and demonstrated that the asymmetrical version yielded PET image reconstruction with improved bias-variance trade-off in comparison to other image gradient-based priors such as parallel level sets (**?**) and compared to the originally proposed Bowsher prior.

2.7 Problems Addressed in this Dissertation

The research problems (RPs) covered in the scope of this thesis are listed here and the aims associated with the particular task follow.

- **RP 1 (Chapter 3).** Existing MRI-informed image reconstruction methods for PET suffer from cross-modality artifacts.
  Aims. Joint modeling of the PET-MRI patch-based dictionary prior modeling to improve PET image quality and reduce cross-modality artifacts.

- **RP 2 (Chapter 3).** Joint modeling of PET-MRI image-gradient-based priors suffer from (i) loss in textural features in the images, (ii) cross-modality artifacts, and (iii) are sensitive to perturbations in the noise level of the test data.
  Aim. Joint PET-MRI patch-based dictionary prior to improve joint PET-MRI reconstruction by restoring both edges and textural features in the images and also reduce cross-modality artifacts. (Chapter 3)

- **RP 3 (Chapter 4).** For static LD-PET to SD-PET image synthesis, DNN-based methods are not (i) robust to newer OOD PET data acquisitions and (ii) do not quantify uncertainty in the predicted images.
  Aim. (i) Modeling imaging physics-based information into the loss function of the DNNs to improve the robustness of the DNN to newer OOD acquisitions and (ii) modeling the per-voxel heteroscedasticity in the training data to make uncertainty estimation part of the DNN framework.

- **RP 4 (Chapter 5).** Classical image filtering approaches fail to identify brain metabolic activations at high temporal resolution fPET experiments
  Aim. Develop an image denoising model to leverage anatomical information from structural MRI to remove noise in the fPET frames in addition to removal of partial volume effects. The restored spatiotemporal information from fPET images can be used to identify improved estimates of source components via ICA.

Figure 2.2 illustrates the context of this thesis within the entire PET-MRI imaging framework.

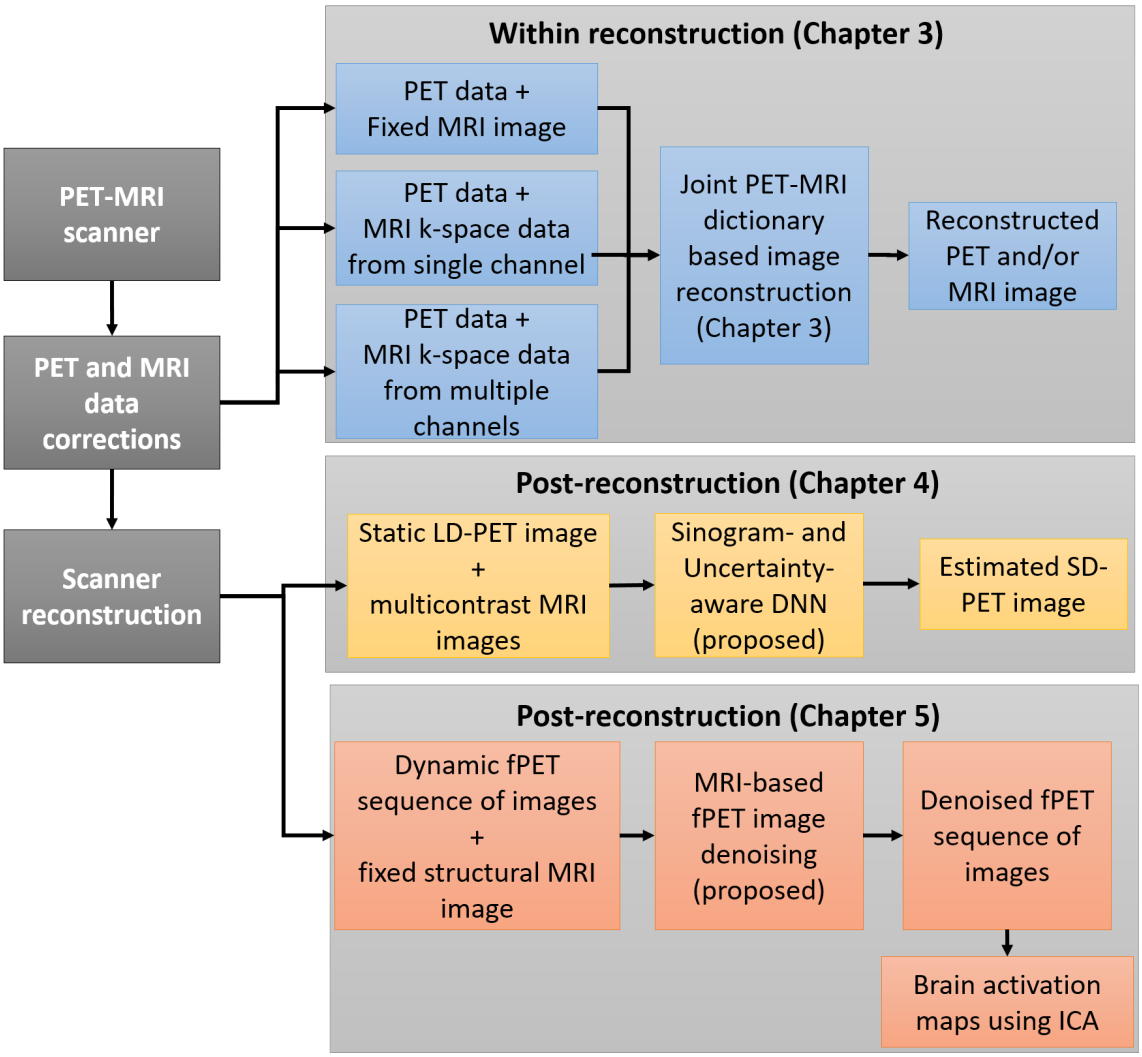

Figure 2.2: **The context of problems addressed in this thesis within the imaging workflow.** The models developed in this thesis address the research problems described in Section 2.7.

Chapter 3

Joint Dictionary based PET and MRI Reconstruction

3.1 Introduction

To overcome the limitations of the image-gradient-based joint priors for PET and MRI reconstruction, we propose a novel joint patch-based prior. Specifically, we propose a novel *joint PET-MRI patch-based dictionary* model for the joint reconstruction of PET-MRI images. We formulate the joint dictionary model within a Markov random field (MRF) framework that is designed to model inter-voxel dependencies.

This chapter describes the proposed novel generative models based on a joint PET-MRI dictionary for the (i) reconstruction of PET images with a fixed MRI image and the joint reconstruction of PET and MRI images with (ii) undersampled MRI data acquired from single channel, and (iii) undersampled MRI data acquired from multiple channels (parallel MRI).

The developed joint PET-MRI patch-based prior learns inter-modality patch-level (higher-order) dependencies, rather than voxel-wise gradient-based (fine-scale) dependencies, while preserving intra-modality textural patterns in the images. In this work, we propose an expectation maximization (EM) based optimization for the joint reconstruction problem, within a Bayesian inference framework. The proposed framework can accommodate a variety of sampling patterns for accelerated MRI k-space data acquisition, which can be implemented easily on commercial PET-MRI scanners. The results, on both simulated phantom, simulated BrainWeb data, as well as *in vivo* human brain data, demonstrate improved image quality and reduced reconstruction errors in both PET and MRI images compared with existing methods, both qualitatively and quantitatively.

The chapter is organized as follows. Section 3.2 describes (i) our proposed joint PET-MRI dictionary prior model and (ii) the Bayesian reconstruction framework for reconstruction of PET data with different MRI acquisition scenarios, relying on EM optimization. Section 3.3 shows empirical results, including quantitative and qualitative evaluation, on simulated and *in vivo* data. Section 3.4 summarizes the work and describes possible directions for future work.

Previous works on developing priors for the joint reconstruction of PET-MRI are based on image gradients, and thus promote piecewise-constant / smooth results. This can lead to staircasing artifacts or cartoonish appearances. Unlike such gradient-based models, patch-based models can capture the realistic textural patterns and contextual features in the image.

3.2 Methods

We describe our joint PET-MRI patch-based dictionary prior, followed by the likelihood models for PET and MRI, and finally, our Bayesian framework for joint reconstruction using EM.

3.2.1 Joint-Dictionary PET-MRI Prior Model: Formulation

Let U and V represent, respectively, the co-registered PET activity image and the complex-valued MRI image, each containing I voxels. We propose a MRF prior model on the image pair (U, V). For every voxel i, the MRF neighborhood comprises all other voxels within a radius of d_U mm in the PET image and a radius of d_V mm in the MRI image. For this work, we set $d_U := d_V := 4$ mm. Let operators $\mathcal{R}_i^{\text{PET}}$ and $\mathcal{R}_i^{\text{MRI}}$ extract square patches of side length $L = (d_U + 1)$ mm, centered at voxel i, from the PET and MRI respectively, and subsequently vectorize the patches. The spatial locations underlying the square patches form a clique in the MRF. Consider overlapping patches centered at all possible spatial locations at a distance of $\max(d_U, d_V)$ mm away from the boundary of the PET image and the MRI image, to give us K such patches from each image each lying completely inside the image. Let the k-th concatenated patch be $\begin{pmatrix} \mathcal{R}_k^{\text{PET}} U \\ \mathcal{R}_k^{\text{MRI}} V \end{pmatrix}$. We propose that every such joint patch can be represented by a sparse linear combination of J joint-patch templates, called atoms, which form the columns of the joint dictionary A. The joint dictionary A consists of a concatenation of the PET dictionary (A^U) and the MRI dictionary (A^V) arranged as $A = \begin{pmatrix} A^U \\ A^V \end{pmatrix}$. Atoms in the PET dictionary A^U have non-negative components, consistent with the values in the PET image U. We propose to constrain the MRI dictionary atoms to be non-negative valued, modeling patterns in the patches of

the MRI-magnitude image that is, typically, clinically more relevant than the MRI-phase image. The phase image is more corrupted, has abrupt discontinuities resulting from phase wrapping, and fails to capture anatomically relevant information. To augment the dictionary fit, we introduce a complex-valued image Φ that models the per-voxel phase (unit magnitude). Let C_k be the coefficients underlying the sparse linear combination. We model the MRF prior on (U, V) using the associated Gibbs energy $G(U, V; A)$ as the sum of potential functions across all $L \times L$ sized cliques. Thus, we design

$$G(U,V;A) := \alpha \sum_{k=1}^{K} \min_{C_k \geq 0, \Phi} \Big(\beta \|\mathcal{R}_k^{\text{PET}} U - A^U C_k\|_2^2 + (1-\beta)\|\mathcal{R}_k^{\text{MRI}} V - \mathcal{R}_k^{\text{MRI}} \Phi \odot A^V C_k\|_2^2 + \lambda \|C_k\|_1 \Big), \tag{3.1}$$

where $\odot$ denotes the Hadamard product and $C_k \geq 0$ enforces non-negativity on each component of C_k. This formulation enforces a non-informative (flat) prior on the phase image Φ. Parameter $\alpha > 0$ controls the overall prior strength by acting on every term in Equation 3.1. Parameter $\beta > 0$ balances the quality of fit with respect to the PET and MRI images; in this way, β also effectively accounts for the differences in the intensity ranges across the PET and MRI modalities; β is analogous to such a parameter in **?**. Parameter $\lambda > 0$ promotes the sparsity of the coefficients. We tune these parameters empirically using a validation set.

3.2.2 Joint-Dictionary PET-MRI Prior Model: Learning Strategy

We learn the joint dictionary A from a set of T uncorrupted co-registered image pairs of PET activity and non-negative-valued MRI-magnitude, i.e., $\{(\mathring{U}^t, \mathring{V}^t)\}_{t=1}^T$. We formulate the dictionary learning problem as the constrained optimization problem:

$$A^{\text{opt.}} := \arg\min_A \min_C \sum_t^T \sum_k^K \beta \|\mathcal{R}_k^{\text{PET}} \mathring{U}^t - A^V C_k^t\|_2^2 + (1-\beta)\|\mathcal{R}_k^{\text{MRI}} \mathring{V}^t - A^V C_k^t\|_2^2 + \lambda \|C_k^t\|_1, \tag{3.2}$$

such that all atoms A_j satisfy the constraints: $\|A_j^U\|_2^2 + \|A_j^V\|_2^2 \leq 1$, $A_j^U \geq 0$, and $A_j^V \geq 0$. We want the dictionary representations AC_k^t to have non-negative intensities and, hence, we enforce non-negative constraints on elements of A and C. The unit-norm constraint avoids trivial solutions, e.g., atom norms tending to ∞ while coefficients C_k^t tending to zero. We use patches that have standard deviations above noise levels in both PET and MRI images for the purpose of training the joint dictionary. The convex combination

of the weights β and $(1-\beta)$ for the PET and MRI dictionary fit terms, respectively, is consistent with the unit-norm constraint on the atoms. When fixing one of the two variables A and C, our formulation leads to convex sub-problems with convex constraint sets. Hence, we solve the objective function by alternating minimization for A and C, using gradient-descent updates followed by projections onto the constraint space. The updates adaptively select the step size to ensure that the objective function reduces at each step. Specifically, we reduce the step size by 50% when the objective function value reduces while we increase the value by 10% if the objective function value increases.

3.2.3 Likelihood Model for PET Data

Consider a circular arrangement of D detector pairs around the subject. We model the measured coincident gamma events at a particular detector pair d, by the random variable Y_d. Our aim is to reconstruct the mean activity U_i at each location i. The PET system matrix H can be obtained for a known configuration of the detector pairs. The system matrix H is of dimension $D \times I$, where every element of the matrix $H_{d,i}$ models the probability that the activity occurring at location i is detected in d-th detector pair. For the simulation studies, we neglect the effect of the scatter, random noise, and attenuation correction. Then, the observed data Y_d at tube d is drawn from Poisson($\sum_i^I U_i H_{d,i}$). For the observed PET data $Y := \{Y_d\}_{d=1}^D$, the likelihood model $P(Y|U)$ is

$$P(Y|U) = \prod_{d=1}^{D} P(Y_d| \sum_{i}^{I} U_i H_{d,i}). \tag{3.3}$$

3.2.4 Likelihood Model for Undersampled MRI Data

Let $\mathcal{F}$ represent the Fourier-encoding operator. Let $\mathcal{S}$ represent the k-space undersampling operator that selects the phase-encoding lines in k-space, leading to the complex-valued data matrix Z. We model measurement noise as additive independent identically distributed (i.i.d.) complex Gaussian with zero mean and variance σ^2. Thus, the likelihood model for MRI is

$$P(Z|V, \sigma^2) := \zeta \exp(-\|Z - \mathcal{S}\mathcal{F}V\|_F^2/\sigma^2), \tag{3.4}$$

where $\|\cdot\|_F$ is the Frobenius norm and ζ is the normalizing factor.

3.2.5 Likelihood Model for Parallel-MRI

The MRI likelihood model in Equation 3.4 can be extended for parallel-MRI as follows. Let L coils with coil sensitivity maps $\{S_l\}_{l=1}^L$ acquire data $Z := \{Z_l\}_{l=1}^L$. Similar to

the single-coil case, we consider i.i.d. complex Gaussian noise in k-space, with mean 0 and variance σ_p^2. Thus, the likelihood for parallel-MRI is

$$P(Z|V,\sigma_p^2) := \zeta_p \prod_{i=1}^{L} \exp(-\|Z_l - \mathcal{S}\mathcal{F}(S_l \odot V)\|_2^2/\sigma_p^2) \tag{3.5}$$

where $\odot$ denotes the Hadamard product and ζ_p is the normalizing constant.

3.2.6 A Novel Bayesian Joint PET-MRI Reconstruction: Formulation

We leverage our joint-dictionary prior model and the likelihood models for PET and MRI to formulate an EM based framework for joint reconstruction of PET and MRI images.

Let random variable $\widetilde{Y} := \{\widetilde{Y}_{d,i}\}_{d=1,i=1}^{D,I}$ denote the (unknown) number of counts $\widetilde{Y}_{d,i}$ at detector pair d resulting from the activity at spatial location i. We model each $\widetilde{Y}_{d,i}$ as a hidden random variable. Thus, the complete data $Y^{\text{complete}} := \widetilde{Y} \cup Y$. We can also interpret the complete data as the set of random variables $\widetilde{Y}_{d,i}$ under the constraint $\sum_i^I \widetilde{Y}_{d,i} = Y_d$, where random variables Y_d model the observed data. EM considers U and V as parameters. The complete-data likelihood is

$$P(Y^{\text{complete}}|U) := \prod_{d=1}^{D} \prod_{i=1}^{I} P(\widetilde{Y}_{d,i}|U_i H_{d,i}) \tag{3.6}$$

that is defined on the subspace where $\sum_i^I \widetilde{Y}_{d,i} = Y_d$. Given the acquired MRI data Z, the acquired PET data Y, and the learned joint dictionary A, we reconstruct the PET and MRI images U and V as the maximum-a-posteriori (MAP) estimates

$$\arg\max_{U,V} P(U,V|Y,Z,A,\sigma^2) \tag{3.7}$$

$$= \arg\max_{U,V} P(Y|U)P(Z|V,\sigma^2)P(U,V|A). \tag{3.8}$$

During EM optimization, in the $(m+1)$-th iteration, with parameters values U^m and V^m, EM designs a functional

$$\begin{aligned} &Q(U,V;U^m,V^m) \\ &:= \mathbb{E}_{P(\widetilde{Y}|U^m,V^m,Y,Z)}\left[\log(P(Y^{\text{complete}}|U)P(Z|V,\sigma^2)P(U,V|A))\right], &(3.9)\\ &= \mathbb{E}_{P(\widetilde{Y}|U^m,Y)}\left[\log P(\widetilde{Y}|U)\right] + \log P(Z|V,\sigma^2) + \log P(U,V|A), &(3.10)\end{aligned}$$

keeping the constraint $\sum_{i=1}^{I} \widetilde{Y}_{d,i} = Y_d$.

The log-likelihood term (first term) in the functional for MRI is

$$\mathcal{L}_{\text{MRI}}(V|Z,\sigma^2) := \log(P(Z|V,\sigma^2)) = -\|Z - \mathcal{S}\mathcal{F}V\|_F^2/\sigma^2 + \log(\zeta). \tag{3.11}$$

For PET, using $\mathbb{E}\left[\widetilde{Y}_{d,i}\right] = U_i H_{d,i}$ and other properties of Poisson distributions under the constraint $\sum_{i=1}^{I} \widetilde{Y}_{d,i} = Y_d$, the expectation term (third term) in the functional simplifies to

$$\begin{aligned}
&\mathcal{L}_{\text{PET}}(U|Y) \\
&:= \sum_{d=1}^{D}\sum_{i=1}^{I} -U_i H_{d,i} + \log(U_i H_{d,i})\mathbb{E}_{P(\widetilde{Y}|U^m,Z)}\left[\widetilde{Y}_{d,i}\right] - f(Y) &(3.12)\\
&= \sum_{d=1}^{D}\sum_{i=1}^{I} -U_i H_{d,i} + Y_d\left(\frac{U_i^m H_{d,i}}{\sum_i^I U_i^m H_{d,i}}\right)\log(U_i H_{d,i}) - f(Y), &(3.13)
\end{aligned}$$

where the term $f(Y)$ depends only on the observed PET data Y, and is independent of any parameter.

Thus, in a simplified form, the sum of the two fidelity-terms and the regularization-parameter-weighted prior-term (ignoring the constant terms) is

$$Q(U,V;U^m) = \mathcal{L}_{\text{PET}}(U|Y) + \mathcal{L}_{\text{MRI}}(V|Z,\sigma^2) + -G(U,V;A). \tag{3.14}$$

At iteration $m+1$, EM updates the estimates of the reconstructed images to U^{m+1} and V^{m+1} given by

$$\begin{aligned}
&\arg\max_{U,V\geq 0}\max_{\Phi,C\geq 0}\Big\{ - \|Z - \mathcal{S}\mathcal{F}V\|_F^2/\sigma^2 \\
&+ \sum_{d=1}^{D}\sum_{i=1}^{I} -U_i H_{d,i} + Y_d\left(\frac{U_i^m H_{d,i}}{\sum_i^I U_i^m H_{d,i}}\right)\log(U_i H_{d,i}) \\
&- \alpha\sum_{k=1}^{K}\beta\|\mathcal{R}_k^{\text{PET}}U - A^U C_k\|_2^2 + (1-\beta)\|\mathcal{R}_k^{\text{MRI}}V - \mathcal{R}_k^{\text{MRI}}\Phi\odot A^V C_k\|_2^2 + \lambda\|C_k\|_1\Big\}. &(3.15)
\end{aligned}$$

3.2.7 Joint PET-MRI Reconstruction: Optimization Strategy

We begin by initializing the PET-MRI images U and V to their maximum-likelihood estimates. Within each iteration of EM, we optimize using alternating minimization for parameters Φ, C, U, and V, in that sequence.

Given current estimate V, we can estimate the phase image Φ at every voxel. The optimization function for Φ reduces to

$$\arg\min_{\Phi} \sum_{k=1}^{K} \|\mathcal{R}_k^{\text{MRI}} V - \mathcal{R}_k^{\text{MRI}} \Phi \odot A^V C_k\|_2^2. \tag{3.16}$$

Irrespective of the dictionary fits $A^V C_k$, the optimal value at i-th voxel in Φ is the phase that rotates the real value at that component in $A^V C_k$ which corresponds to the i-th voxel to match the complex value at the i-th voxel in V. Hence, we get updates for Φ in closed form as the phase component of the complex-valued image V. This strategy was first proposed in **?**.

Next, given U, V, and Φ, we estimate C_k as follows. Let operator $\mathcal{M}$ take the complex-valued image V and output an image with values as the magniudes of the complex values in V. We fit the dictionary to the magnitude images $\mathcal{M}V$ and U as

$$\arg\max_{C \geq 0} \sum_{k=1}^{K} \|\mathcal{R}_k^{\text{PET}} U - A^U C_k\|_2^2 + \|\mathcal{R}_k^{\text{MRI}} \mathcal{M} V - A^V C_k\|_2^2 + \|C_k\|_1. \tag{3.17}$$

For U, we use iterative gradient-ascent updates followed by projections onto the constraint space modeling the non-negativity constraint, with adaptive step size. For V, we use iterative gradient ascent, where we adaptively select the step size to ensure that the objective function reduces at each step.

3.2.8 Joint PET-MRI Reconstruction: Algorithm Summary

Our EM-based joint reconstruction algorithm is summarized below.

1. **Training.** Learn joint dictionary A using training set $\{(\mathring{U}^t, \mathring{V}^t)\}_{t=1}^{T}$, as described in Section 3.2.2.

2. **Input.** PET-MRI data (Y, Z) and learned dictionary A. Initialization of the reconstructed PET image U^0 using the MLEM algorithm. Initialization of the reconstructed MRI images using the zero-filled inverse Fourier transform (ZF-IFT), represented by the complex-valued image V^0 and phase component Φ^0.

   2.1. Set iteration number $m \leftarrow 0$.

   2.2. During iteration $(m + 1)$ of the EM algorithm, at the E step: Define the $Q(.)$ function as in Equation 3.14.

2.3. During iteration $(m+1)$ of the EM algorithm, at the M step: Perform iterative projected gradient descent until convergence, based on the objective function in 3.15, as follows.

2.3.1. Estimate Φ^{m+1} by closed-form update using V^{m+1}, as described earlier in this section.

2.3.2. Estimate C by iterative projected gradient descent until convergence, keeping (U, V, Φ) fixed to their latest update.

2.3.3. Estimate U^{m+1} by iterative projected gradient descent until convergence, keeping (V, Φ, C) fixed to their latest update.

2.3.4. Estimate V^{m+1} by iterative projected gradient descent until convergence, keeping (U, Φ, C) fixed to their latest update.

2.4. Set iteration number $m \leftarrow m+1$.

2.5. If $\frac{(\|U^m - U^{m-1}\|_F^2 + \|V^m - V^{m-1}\|_F^2)^{0.5}}{(\|U^{m-1}\|_F^2 + \|V^{m-1}\|_F^2)^{0.5}} \leq 1e^{-3}$, then terminate and go to Step 2.6. Otherwise, increment m by 1 and repeat the EM iterations by going to Step 2.2.

2.6. **Output.** Jointly-reconstructed PET and MRI images V^m and U^m.

3.2.9 Joint PET-MRI Reconstruction: Extension to Parallel-MRI

Given the PET-MRI measurements $(Y, \{Z_l\}_{l=1}^L)$, and the learned joint dictionary A, we formulate the joint reconstruction as the maximum-a-posteriori (MAP) estimate

$$\arg\max_{U,V} P(U, V|Y, \{Z_l\}_{l=1}^L, A, \sigma_p^2) = \arg\max_{U,V} P(Y|U) P(\{Z_l\}_{l=1}^L|V, \sigma_p^2) P(U, V|A). \tag{3.18}$$

The log-likelihood term for parallel-MRI becomes

$$\mathcal{L}_{\text{pMRI}}(V|\{Z_l\}_{l=1}^L, \sigma_p^2) := \log(P(\{Z_l\}_{l=1}^L|V, \sigma_p^2)) = -\sum_{l=1}^{L} \left(\|Z_l - \mathcal{S}\mathcal{F}(S_l \odot V)\|_2^2 / \sigma_p^2 \right) + \log(\zeta_p). \tag{3.19}$$

Hence, the functional in Equation 3.14 can be modified for parallel-MRI as

$$\mathcal{L}_{\text{PET}}(U|Y) + \mathcal{L}_{\text{pMRI}}(V|\{Z_l\}_{l=1}^L, \sigma_p^2) + -G(U, V; A). \tag{3.20}$$

Subsequent EM updates are modified accordingly. In this case, we initialize the MRI image V using zero-filled inverse Fourier transform followed by SENSE (**?**). Similar to joint PET-MRI reconstruction with single-coil MRI, we initialize the PET image U using MLEM.

3.2.10 PET-only Reconstruction: With Fixed MRI image

The reconstruction scenario discussed in the previous sections can be modified for the reconstruction of only PET data, given a fixed MRI magnitude image V^{fixed} and PET data $\{Y_d\}_{d=1}^{D}$. Hence, for the PET-only reconstruction, the the MAP estimate is

$$\arg\max_{U} P(U|V^{\text{fixed}}, Y, A) = \arg\max_{U} P(Y|U)P(U, V^{\text{fixed}}|A). \tag{3.21}$$

For the case of PET-only reconstruction, the well-known one-step-late (OSL) image reconstruction strategy **?** can be used with our joint dictionary prior. The OSL-based updates for the PET image reconstruction estimate, for each voxel i, is given by

$$U_i^{m+1} := \left(\sum_{d=1}^{D} \frac{U_i^m H_{d,i} Y_d}{\sum_{i=1}^{I} U_i^m H_{d,i}}\right) \Big/ \left(\frac{\partial G(U, V^{\text{fixed}}, A)}{\partial U_i}\Big|_{U_i = U_i^m} + \sum_{d=1}^{D} H_{i,d}\right), \forall i, \tag{3.22}$$

where we evaluate $\partial G(U, V^{\text{fixed}}, A)/\partial U_i$ at $U_i = U_i^m$ by (i) fitting the dictionary A to the image pair (U^m, V^{fixed}), as dictated within $G(U, V^{\text{fixed}}, A)$, to produce optimal coefficients $\{C_i^*\}_{i=1}^{I}$, and then (ii) taking the partial derivative of $(1-\beta)\|\mathcal{R}_i^{\text{PET}} U - A^U C_i\|_2^2$ with respect to U_i. EM iterations stop when the relative change in the estimates U^m and U^{m+1} is small.

3.3 Experiments and Results

PET-only Reconstruction. We evaluate the one-sided PET-only reconstruction problem on simulated phantom data where the MRI image is assumed to be fixed/static. Given the fixed MRI image, the joint dictionary trained on noiseless PET and MRI data, the task is to reconstruct noisy PET data. We compare our PET reconstruction framework using joint-dictionary (jd), qualitatively and quantitatively, with five different methods: (i) MLEM-based reconstruction; (ii) Huber-loss MRF priors, similar to TV priors penalizing magnitudes of intensity gradients; (iii) reconstructing PET with MRI-image patch-based dictionaries learned from MRI image alone (MRI-Dict); (iv) reconstructing PET using LPLS (**?**); and (v) reconstructing PET using joint-TV (JTV) (**?**).

Joint PET-MRI reconstruction (single-coil MRI). We evaluate our Bayesian joint-reconstruction framework for PET and single-channel MRI on (i) simulated PET-MRI datasets using BrainWeb data (**??**), including fully-sampled and undersampled MRI data, and (ii) an *in vivo* PET-MRI cohort with undersampled MRI data. We compare our JD framework, qualitatively and quantitatively, with five different methods: (i) independently reconstructing PET and MRI without priors, i.e., ZF-IFT for MRI, and MLEM-based reconstruction for PET; (ii) independently reconstructing PET and MRI with Huber-

loss MRF priors; (iii) reconstructing PET and MRI independently with patch-based dictionaries (ID), one learned for modeling MRI images and another learned for modeling PET images; (iv) jointly reconstructing PET and MRI using LPLS (**?**); and (v) jointly reconstructing PET and MRI using QPLS(**?**).

Joint PET-MRI reconstruction (parallel-MRI). We evaluate our Bayesian joint-reconstruction framework for PET and parallel-MRI using *in vivo* PET-MRI data with 2× undersampled MRI data. In this case, we compare our JD method with (1) independent reconstructions of MRI and PET: (i) without the use of priors; (ii) using Huber-loss MRF prior, and (2) joint reconstruction using JSR prior **?** relying on joint gradient-based sparsity, extending the TV prior.

For all the three scenarios, we evaluate the quality of the reconstructed image ($\widehat{x}$), with respect to the ground truth image (x), based on the (i) relative root mean squared error (RRMSE), defined as $\|\widehat{x} - x\|_F / \|x\|_F$, and (ii) structural similarity index (SSIM) (**?**). We optimize the hyperparameters underlying all methods to minimize the RRMSE on a validation set, as detailed later.

3.3.1 PET-only Reconstruction: Results With Simulated Phantom.

For the simulated phantom (Figure 3.1(a1)-(a2)), we sufficiently blur the PET image to reproduce the *lower resolution in PET* **?**, relative to MRI. The MLEM reconstruction (Figure 3.1(a3)) retains a lot of the noise compared to prior-based methods. EM with the HuberMRF prior (Figure 3.1(a4)) gets rid of most of the random noise. JTV and PLS (Figure 3.1(b3),(b4)) leverage the anatomical structure in the MRI, encouraging edges in the PET reconstruction to occur at the same spatial locations as the edges in the MRI image. They improve over Huber MRF prior, but the gradient-based penalty limits the quality of reconstruction of the (i) blue circular blobs, (ii) red circular outside rim, and (iii) red parallel bars in the center. Using a MRI-patch statistical model to reconstruct PET images (PET patch intensities being significantly smoother than MRI patch intensities) results in overfitting of the dictionary to the noise (Figure 3.1(b2)). Our reconstruction (Figure 3.1(b1)) using a joint patch-based dictionary model maintains both fine-scale regularity, in the form of smoothness, and larger-scale regularity by preservation of structures like the straightness and separability of the red bars, circularity of the blue blobs, and the continuity in the red outer ring. Our reconstruction has much smaller residual magnitudes (Figure 3.1(c1)) compared to all other methods, and is closest to the ground truth qualitatively and quantitatively (Figure 3.1).

3.3.2 Joint PET-MRI Reconstruction: Results on Two Simulated BrainWeb (BW) Datasets

For BrainWeb-based simulations, we consider two sets of data: (i) a single subject's brain volume from the original BrainWeb dataset (**?**), which we call as the **BW1** dataset, and (ii) a set of 20 subjects' brain volumes with anatomical models (**?**), which we call as the **BW20** dataset. We evaluate all methods under two different scenarios: (i) using PET-MRI data simulated using a *single corruption-free* MRI image provided in BW1, where we additionally simulate lesions in PET and MRI, and (ii) using *multiple noisy* MRI images in BW20, where we learn dictionaries from a small subset of subjects and subsequently use it to reconstruct undersampled and lower-quality data in the rest of the subjects.

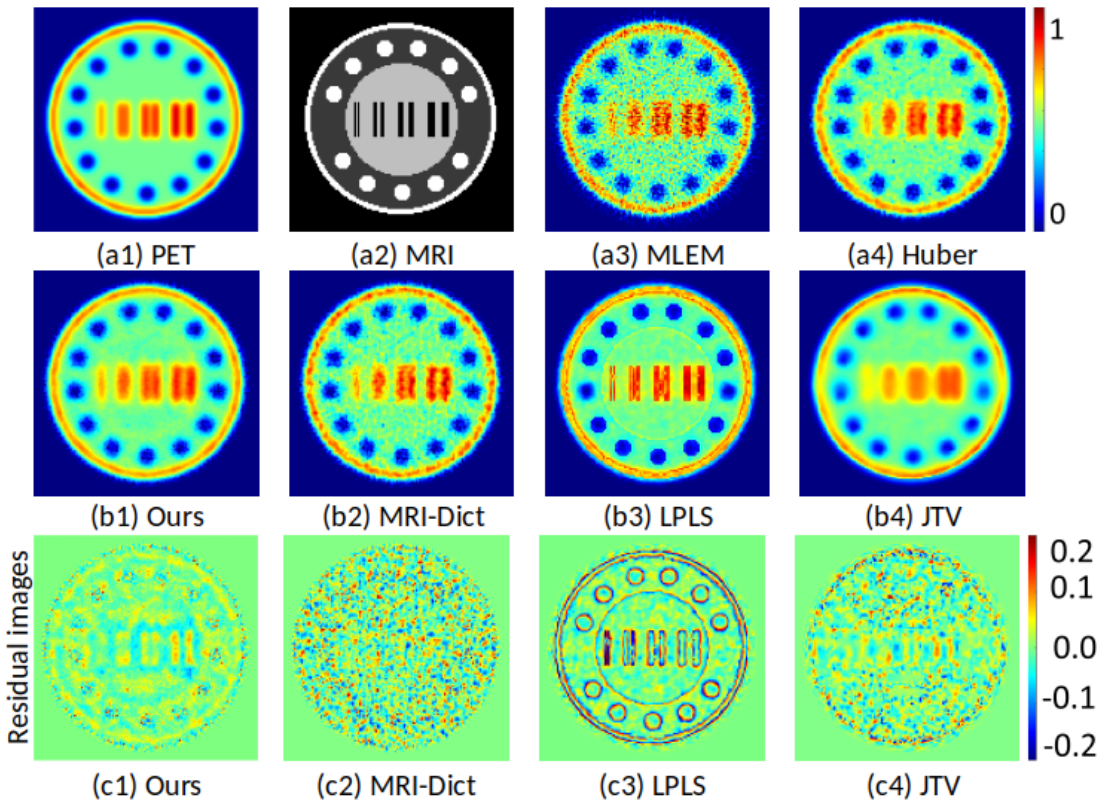

Figure 3.1: **PET-only Reconstruction: Results on Simulated Phantom. (a1)-(a2)** PET-MRI ground truth. **(a3),(a4),(b1)–(b4)** PET reconstructions using various methods. **(c1)–(c4)** Residual (reconstructed - truth) images for the results in (b1)–(b4). **RRMSE: Ours 0.06**, MRI-Dict 0.10, PLS 0.09, JTV 0.08, HuberMRF 0.09, MLEM 0.18. **SSIM: Ours 0.92**, MRI-Dict 0.81, PLS 0.86, JTV 0.90, HuberMRF 0.84, MLEM 0.61.

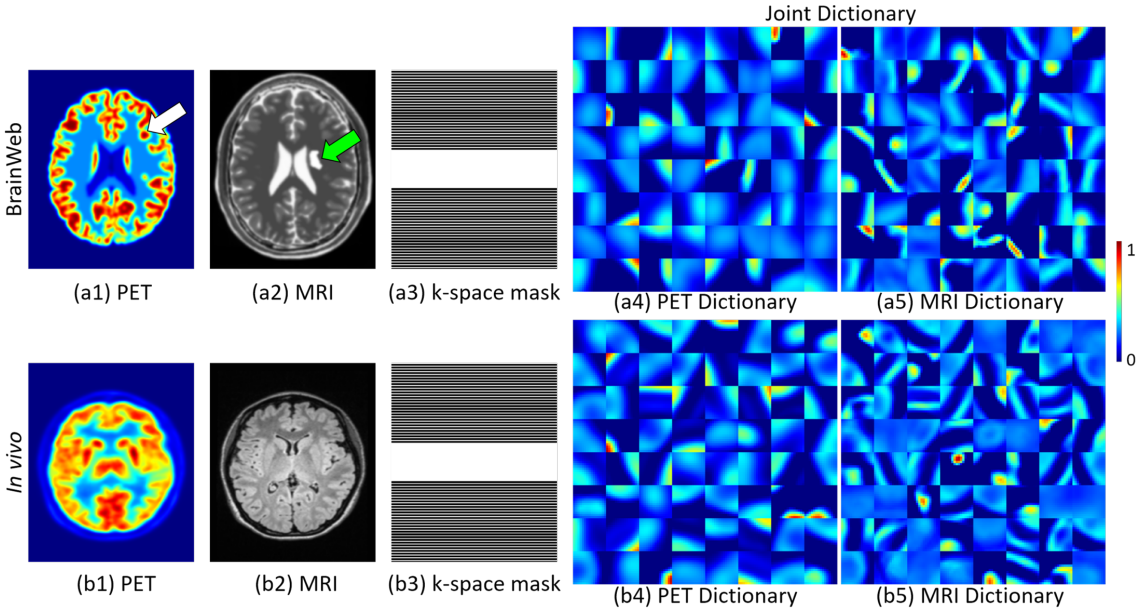

Figure 3.2: **PET-MRI Data and Joint PET-MRI Dictionary.** *Simulated PET-MRI Images from BW1.* (a1)-(a2) PET and MRI images, with simulated lesions. (a3) MRI k-space undersampling mask(2×), (a4)-(a5) *Joint dictionary* A, comprising concatenated A^U and A^V, learned from BW1 PET-MRI images (without simulated lesions): (a4) 64 atoms in A^U representing PET image patches, in a one-to-one correspondence to the 64 MRI atoms in (a5), (a5) 64 atoms in A^V representing MRI image patches; for better visualization, we use a colormap different from grayscale . *In vivo PET-MRI Images.* (b1)-(b2) PET and MRI images. (b3) MRI k-space undersampling mask(2×), (b4)-(b5) *Joint dictionary* learned from *in vivo* PET-MRI images, analogous to (a4)-(a5).

3.3.3 Simulating PET-MRI Images with Lesions

Simulating MRI Images and Data. For the BW1 dataset, we use T2-weighted MRI images (Figure 3.2 (a2)) having an isotropic voxel size of 1 mm^3. For the BW20 dataset, we use T1-weighted MRI images (Figure 3.5 (b1)) having an isotropic voxel size of 1 mm^3, with the MRI simulator parameters as SFLASH (spoiled FLASH) sequence with TR 22 milliseconds, TE 9.2 milliseconds, and flip angle 30 degrees. For the BW1 dataset, we simulate fully-sampled and undersampled k-space data. For the BW20 dataset, we simulate undersampled k-space data. We undersample MRI k-space (undersampling factor 2×) using a standard Cartesian undersampling scheme in Figure 3.2 (a3) that is easily implementable on the scanner; the sampling is dense around the center of the k-space, compared to the high-frequency regions. For MRI data acquisition, we added

i.i.d. zero-mean complex Gaussian noise to the k-space data. We use the same noise level (i.e., standard deviation) for both the cases of fully-sampled k-space and undersampled k-space, such that the RRMSE for the fully-sampled IFT reconstructed image is around 0.14.

Simulating PET Images and Data. For BW1 and BW20 datasets, we use (i) the anatomical models of the normal brains provided in the datasets and (ii) the radiotracer uptake ratios described in **??** to simulate the corresponding PET activity images (Figures 3.2 (a1) and 3.5(a1)). We model the PET system matrix using strip integrals provided in `github.com/JeffFessler/mirt`. We simulate around 10^6 counts for the PET data using a 2D scanner with 180 degree coverage and 215 radial lines. In addition, we smooth the resulting PET activity image by a Gaussian filter of standard deviation 2 mm to model the point spread function of a typical PET scanner. For all the simulations, we ignore scatter, attenuation, and scanner random effects (a more accurate model that includes all these effects would benefit all the methods). This approach is similar to simulation studies in **??**. We control the Poisson noise level such that the standard MLEM reconstruction gives a RRMSE of 0.25, to mimic practical SNR values. For the BW20 dataset, we simulate two different noise levels for the PET data such that the standard MLEM reconstruction gives a RRMSE of 0.25 and 0.16. This mimics practical scenarios where radiotracer dosage might vary across subjects. The different noise levels also cater to applications with low-dose PET imaging that produce low SNR images.

Simulating Lesions in PET-MRI. For BW1, we simulate modality-specific appearances of lesions in both PET and MRI images (Figure 3.2 (a1)–(a2)), indicated by the arrows in the MRI image (near the ventricles), and the PET image (near the cortex).

3.3.4 Dictionary Learning

For the BW1 dataset, we train the dictionaries in ID and JD by (i) taking fully-sampled noisy k-space data, (ii) using that to reconstruct PET-MRI using the Huber prior, (iii) and then taking a few slices for training; we ensure that the slices we use to evaluate all reconstruction methods are at least 10 mm away from the training slices. Figures 3.2(a4)-(a5) show the joint dictionary A learned from training slices in the simulated PET-MRI data. In comparison with the MRI dictionary atoms, the PET dictionary atoms depict patterns with more gradual spatial variation. Unlike the BW1 dataset, the BW20 dataset provides data from multiple subjects and, thus, we learn the joint dictionary A using mid-brain slices of one of the subjects as the training set.

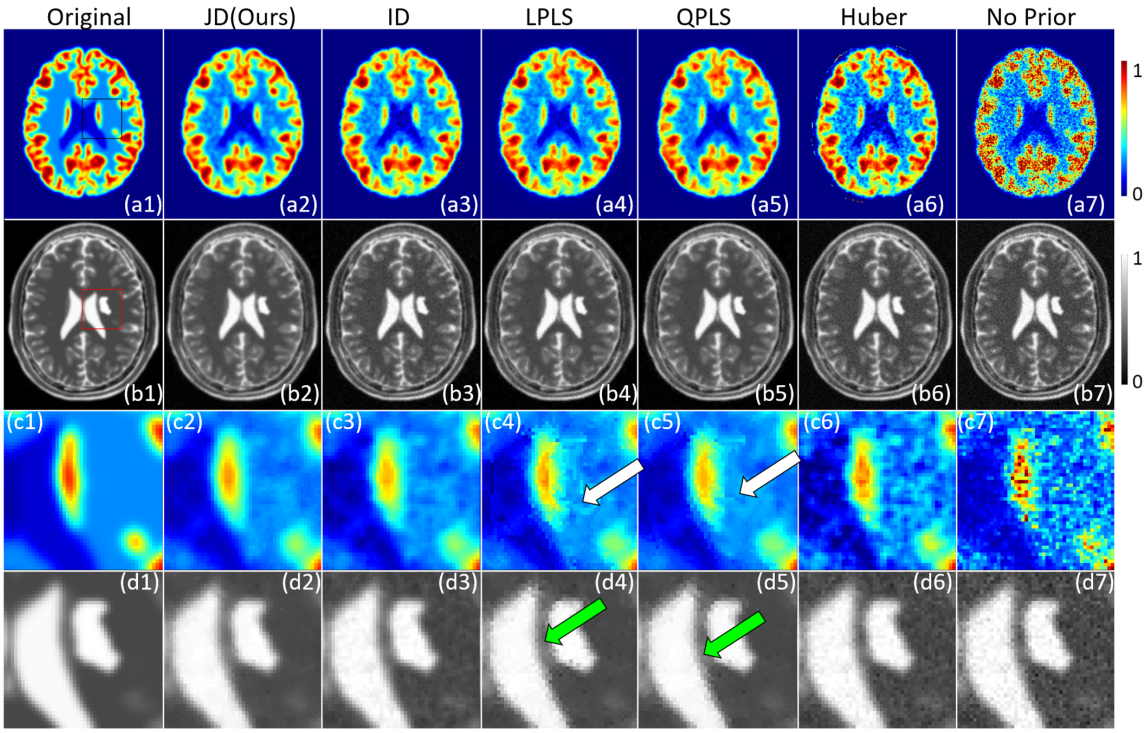

Figure 3.3: **Results on Simulated PET-MRI using BW1: With Fully-Sampled MRI.** Ground truth images: (a1) PET, and (b1) MRI. Joint PET-MRI reconstructions: using **our JD prior** for (a2) **PET** [RRMSE: 0.08] and (b2) **MRI** [0.06]; using **LPLS prior** for (a4) **PET** [0.10] and (b4) **MRI** [0.07]; using **QPLS prior** for (a5) **PET** [0.09] and (b5) **MRI** [0.07]. Independent PET and MRI reconstructions: using **ID prior** for (a3) **PET** [0.09] and (b3) **MRI** [0.07]; using **Huber prior** for (a6) **PET** [0.15] and (b6) **MRI** [0.11]; without using any prior for (a7) **PET** [0.25] and (b7) **MRI** [0.14].

3.3.5 Hyperparameter Tuning

For the BW1 dataset, we tune all hyperparameters for all methods to minimize RRMSE on the evaluation set. For the BW20 dataset, we tune all hyperparameters for all methods to minimize RRMSE on one other subject (validation set), different from the training set. Subsequently, we use these parameter values on the evaluation set.

3.3.6 Joint PET-MRI Reconstruction: Results With Fully-Sampled MRI (BW1)

Figure 3.3 shows PET-MRI reconstructions, from fully-sampled MRI. The prior-less reconstructions (Figure 3.3(a7)–(d7)), i.e., using ZF-IFT for MRI and MLEM for PET, are unable to eliminate the noise. Independent reconstructions of PET and MRI using the Huber-loss based prior (Figure 3.3(a6)–(d6)) remove a significant amount of noise, while

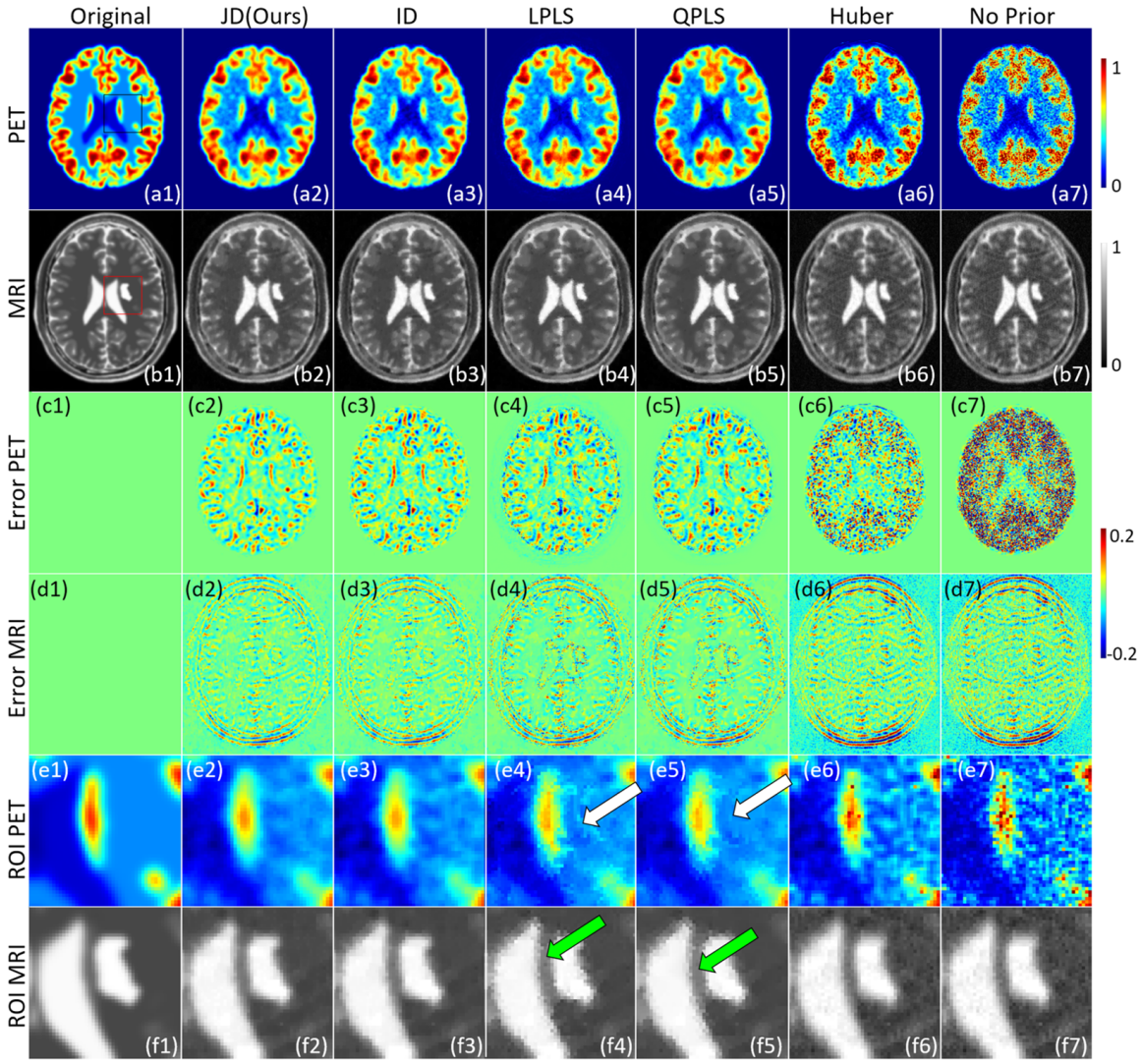

Figure 3.4: **Joint PET-MRI Reconstruction: Results on Simulated PET-MRI using BW1: With 2× Undersampled MRI.** Ground truth images: (a1) PET, and (b1) MRI. Joint PET-MRI reconstructions: using **our JD prior** for (a2) **PET** [RRMSE: 0.08] and (b2) **MRI** [0.08]; using **LPLS prior** for (a4) **PET** [0.10] and (b4) **MRI** [0.09]; using **QPLS prior** for (a5) **PET** [0.09] and (b5) **MRI** [0.08]. Independent PET and MRI reconstructions: using **ID prior** for (a3) **MRI** [0.10] and (b3) **MRI** [0.09]; using **Huber prior** for (a6) **PET** [0.10] and (b6) **MRI** [0.12]; without using any prior for (a7) **PET** [0.25] and (b7) **MRI** [0.14].

preserving edges (in both PET and MRI). Joint PET-MRI reconstruction using LPLS and QPLS priors (Figure 3.3(a4)–(d4),(a5)–(d5)) further improve the MRI reconstruction, with reduced noise and better edge recovery. The reconstructions with the LPLS, and QPLS priors show staircasing effects, which are typical for methods penalizing gradient magnitudes. The LPLS and QPLS priors also show cross-modality artifacts in the region around the lesion, in the PET reconstructions (Figure 3.3(a4)-(a5),(c4)-(c5)), which

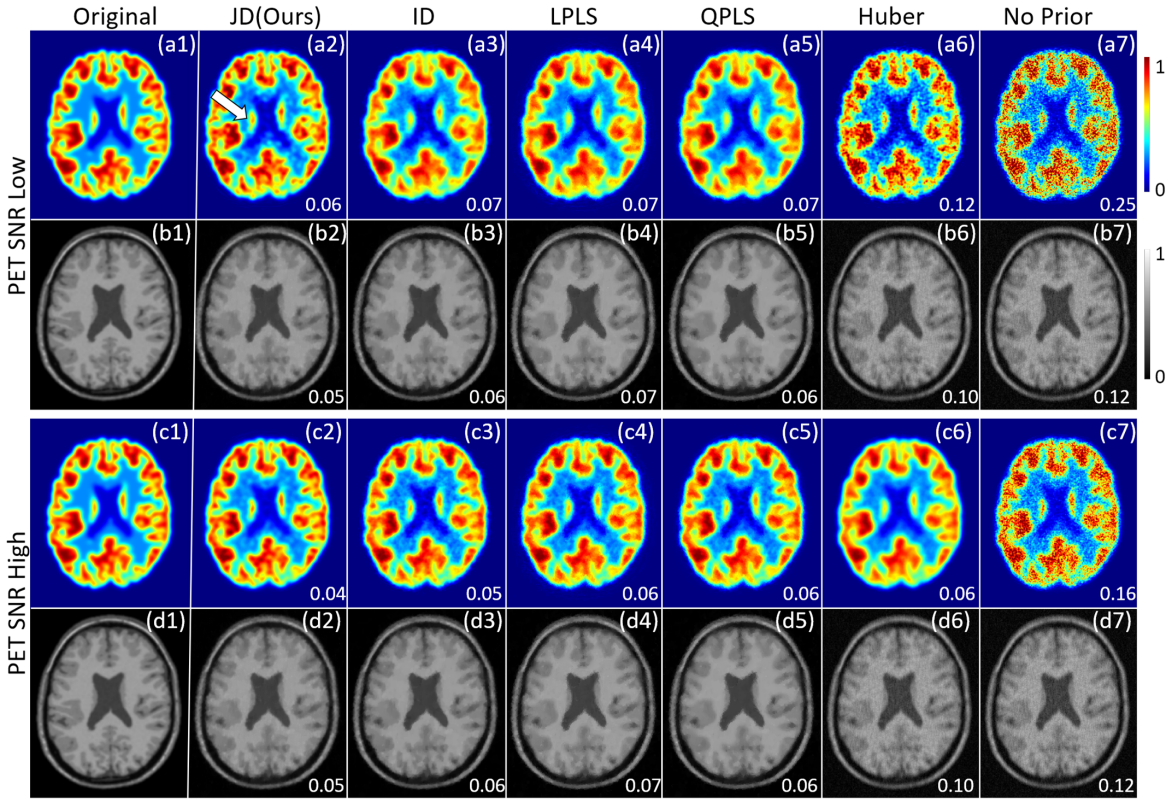

Figure 3.5: **Joint PET-MRI Reconstruction: Results on Simulated PET-MRI using BW20: With 2× Undersampled MRI.** Two different SNR levels in PET data: (a1)–(a7) and (b1)–(b7) **PET SNR low**; (c1)–(c7) and (d1)–(d7) **PET SNR high.** Ground truth images: (a1),(c1) PET, and (b1),(d1) MRI. Joint PET-MRI reconstructions: (a2)–(d2) using **our JD prior**; (a4)–(d4) using **LPLS prior**; (a5)–(d5) using **QPLS prior**. Independent PET and MRI reconstructions: (a3)–(d3) using **ID prior**; (a6)–(d6) using **Huber prior**; (a7)–(d7) without using any prior. RRMSE values are provided at the bottom-right in each image.

is more pronounced for LPLS; this observation is consistent with that in **?** where QPLS performed better at reconstructing shared edges. On the other hand, the patch-based ID prior (Figure 3.3(a3)–(d3)) is able to remove noise further without introducing staircasing effects. In comparison, our patch-based JD prior produces results (Figure 3.3(a2)–(d2)) that (i) are closer to the ground truth in terms of spatial regularity and, more importantly, (ii) preserve modality-specific textural features. The RRMSE values (Figure 3.3) indicate that, compared to other methods, our JD shows an improvement in the reconstructed PET images, without loss of reconstruction quality in the MRI images.

3.3.7 Joint PET-MRI Reconstruction: Results with UnderSampled MRI (BW1 and BW20)

Figure 3.4 shows PET-MRI reconstructions with undersampled MRI data, using the undersampling scheme in Figure 3.2(a3). The independent reconstructions for PET-MRI without priors (Figure 3.4(a7)–(f7)) and with the Huber prior (Figure 3.4(a6)–(f6)) are unable to get rid of the aliasing artifacts in MRI. The Huber prior performs better at noise reduction in both PET and MRI, similar to what we observed earlier in the fully-sampled case. The patch-based ID prior (Figure 3.4(a3)–(f3)) does reduce noise further in PET and MRI, but is unable to get rid of undersampling artifacts in MRI. Clearly, joint reconstructions, using QPLS, LPLS, and JD, are better in terms of (i) reducing aliasing artifacts in MRI reconstructions (Figure 3.4(b2),(b4)-(b5)), (ii) reducing noise in both PET and MRI reconstructions (Figure 3.4(a2),(a4)-(a5),(b2),(b4)-(b5)), and (iii) weaker spatial correlations in the residual images (Figure 3.4(c2),(c4)-(c5),(d2),(d4)-(d5)). LPLS and QPLS tend to introduce jagged edges, similar to the fully-sampled case. The zoomed insets for LPLS (Figure 3.4(e4)-(f4)) and QPLS (Figure 3.4(e5)-(f5)) clearly show cross-modality artifacts between PET and MRI. The ID based reconstruction for PET (Figure 3.4(a3)) shows smoother variations than LPLS and QPLS. However, for MRI, ID still retains most of the aliasing artifacts compared to the joint priors. On the other hand, our JD improves over ID by modeling the inter-modality and intra-modality patch regularity to reconstruct textural patterns in both modalities.

Figure 3.5 shows PET-MRI reconstructions, for the BW20 dataset, at *two different noise levels for PET*; here, while the MRI noise level remains the same across these experiments, we do employ different instance of the noise in the MRI data. In this dataset (Figure 3.5), the aliasing artifacts in MRI are less severe compared to the BW1 dataset (Figure 3.4). Hence, the patch-based ID for MRI reconstruction(Figure 3.5 (b3)), performs better, in terms of removing aliasing artifacts, over Huber-based MRI reconstructions, and is comparable in image quality to LPLS and QPLS MRI reconstructions(Figure 3.5 (b4)-(b5)). Both patch based ID and JD yield smoother MRI reconstructions compared to LPLS and QPLS methods, while our JD prior resulted in the the lowest RRMSE for the MRI image. On the other hand, with regards to PET image reconstruction, our JD seems to retrieve most of the contrast and performs better denoising. In comparison, other methods find it difficult to reconstruct the original contrast and textural patterns as well, especially in the gray matter around the ventricles (indicated by arrow in Figure 3.4 (a2)).

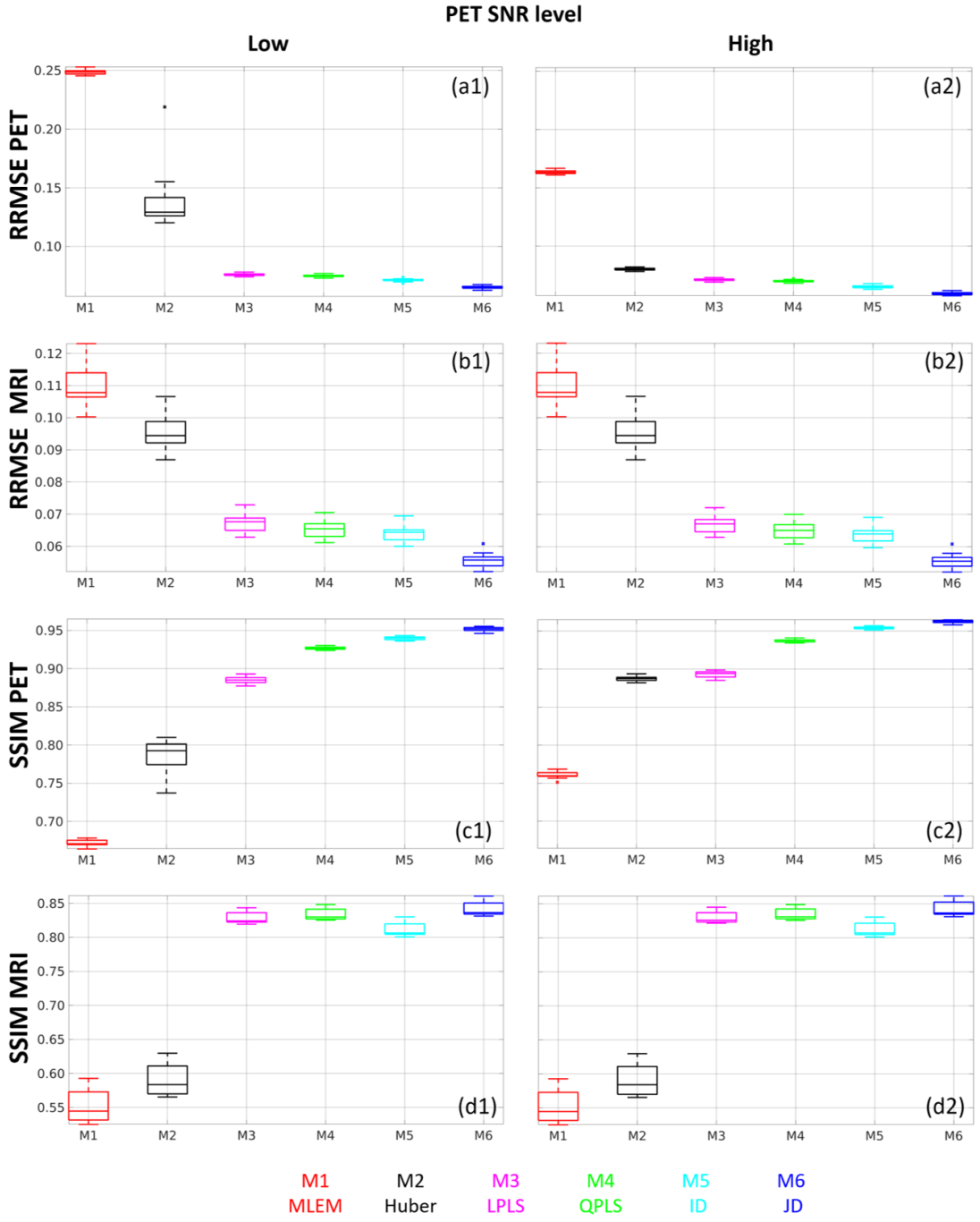

Figure 3.6: **Results on Simulated PET-MRI using BW20: With 2× Undersampled MRI;** RRMSE and SSIM values for PET and MRI images, on 10 mid-brain slices across 8 subjects in the evaluation set of BW20. Two Different Noise Levels in PET. (a1)–(d1) **PET SNR low**; (a2)–(d2) **PET SNR high.** RRMSE values for all the methods for: (a1) **PET** with low SNR PET data, (a2) **PET** with high SNR PET data; (b1) **MRI** with low SNR PET data, (b2) **MRI** with high SNR PET data. SSIM values for all the methods for: (c1) **PET** with low SNR PET data, (c2) **PET** with high SNR PET data; (d1) **MRI** with low SNR PET data, (d2) **MRI** with high SNR PET data.

From the experiments described above, for both BW1 and BW20, we see that the joint reconstruction methods of LPLS, QPLS, and JD help reduce aliasing artifacts in the MRI reconstructions better than other methods. This may stem from the ability of PET reconstructions, which are free of aliasing artifacts, to be able to guide MRI reconstructions towards reducing aliasing artifacts. In this way, joint reconstruction methods leverage the complementary information and the characteristics of PET and MRI images to improve the reconstruction of both modalities. Although ID yields smoother reconstructions for both PET and MRI without the staircasing effects or jagged edges, it is unable to (i) significantly reduce aliasing artifacts in MRI images, and (ii) retrieve original contrast in PET images. The results show that JD can overcome these problems by modeling a joint patch-based prior. In this way, JD shows significant improvement in the PET image quality as well as better textural recovery in both PET and MRI.

Quantitative Results (BW20). We perform quantitative evaluation of different reconstruction methods across multiple slices of different subjects for the BW20 dataset, for the *two different noise levels*. As mentioned before, for both ID and JD, we learn the dictionaries from a single subject and use it for reconstructing images from unseen data. This demonstrates the generalizability of the dictionary for practical scenarios. The box plots (Figure 3.6) for undersampled BrainWeb-based simulations indicate that the joint reconstructions outperform individual reconstructions in both PET and MRI. Our methods leads to a significant improvement in the quality of the PET reconstructions quantitated by RRMSE and SSIM both. Although the RRMSE values of ID based MRI reconstructions is comparable to that of QPLS MRI reconstructions, in terms of SSIM, QPLS is significantly better than ID.

3.3.8 Joint PET-MRI Reconstruction: Results on *in vivo* Datasets With MRI

We now describe the details of the *in vivo* dataset followed by qualitative and quantitative results.

3.3.9 Dataset

We acquire simultaneous brain PET-MRI data for 25 subjects using a Siemens mMR 3-Tesla scanner. The average FDG dose across subjects was around 220 MBq. T2-weighted MRI images were acquired with 2× Cartesian undersampling scheme using the inbuilt protocol with the following acquisition parameters: TR 5000 milliseconds, TE 395 milliseconds, TI 1800 milliseconds, and acquisition time 352 seconds. The corre-

sponding PET acquisition time was 30 minutes. We reconstruct images using the proprietary software tools on the scanner and other third-party software tools (**?**). The native PET resolution was $2.086 \times 2.086 \times 2.032$ mm^3 and the native MRI resolution was 1 mm^3 isotropic. The reconstruction pipeline includes standard pre-processing steps for MRI data and PET data, e.g., attenuation and scatter correction, registration, etc. The MRI UTE sequence images were used for attenuation correction (**?**). We use these reconstructed PET and MRI images as ground truth. We generate PET sinogram data from the scanner provided images using strip integrals mentioned before. We retrospectively undersample MRI k-space, like the strategy in other recent methods (**?**). We use a Cartesian undersampling scheme (Figure 3.2(b3)), where the sampling strategy is similar to that used for simulating PET-MRI in Section 3.3.2, resulting in 2× undersampling. For the *in vivo* PET-MRI data, we evaluate all the methods on *three different noise levels*. We use noise levels such that the standard MLEM based reconstruction results in RRMSEs of 0.42, 0.25, and 0.16. As mentioned before, this pertains to scenarios where the radiotracer dosage (and hence the PET image SNR) might vary across subjects, which depends on several factors such as acquisition protocols and body weight etc. Similar to BW20 data simulation, we use different noise instances for MRI acquisition across PET SNR levels.

For parallel-MRI, we use gradient-echo T2w MRI and FDG-PET from 5 subjects with 12 head coils. We use reconstructions produced by the on-scanner software as ground truth (Figure 3(a1)-(b1)), and then retrospectively obtain MRI-PET measurements with parallel MRI with 2× k-space undersampling.

3.3.10 Dictionary Learning

The joint dictionary A learned from *in vivo* PET-MRI slices appears in Figure 3.2(b4)-(b5). For the *in vivo* data, we learn the joint dictionary on five mid-brain slices of a single subject (training set). Subsequently, we tune the hyperparameters (for all the methods) by reconstructing PET-MRI images for another subject (validation set; different from the training set) to minimize the RRMSE. Finally, we use the learned dictionary and the tuned hyperparameters to reconstruct the PET-MRI images from 25 other subjects (evaluation set).

3.3.11 Joint PET-MRI Reconstruction: Results with Undersampled MRI (*In Vivo*)

Figure 3.7 shows the reconstructions of *in vivo* images for PET and undersampled k-space MRI *in vivo* data. The undersampling sheme is shown in Figure 3.2(b1). The

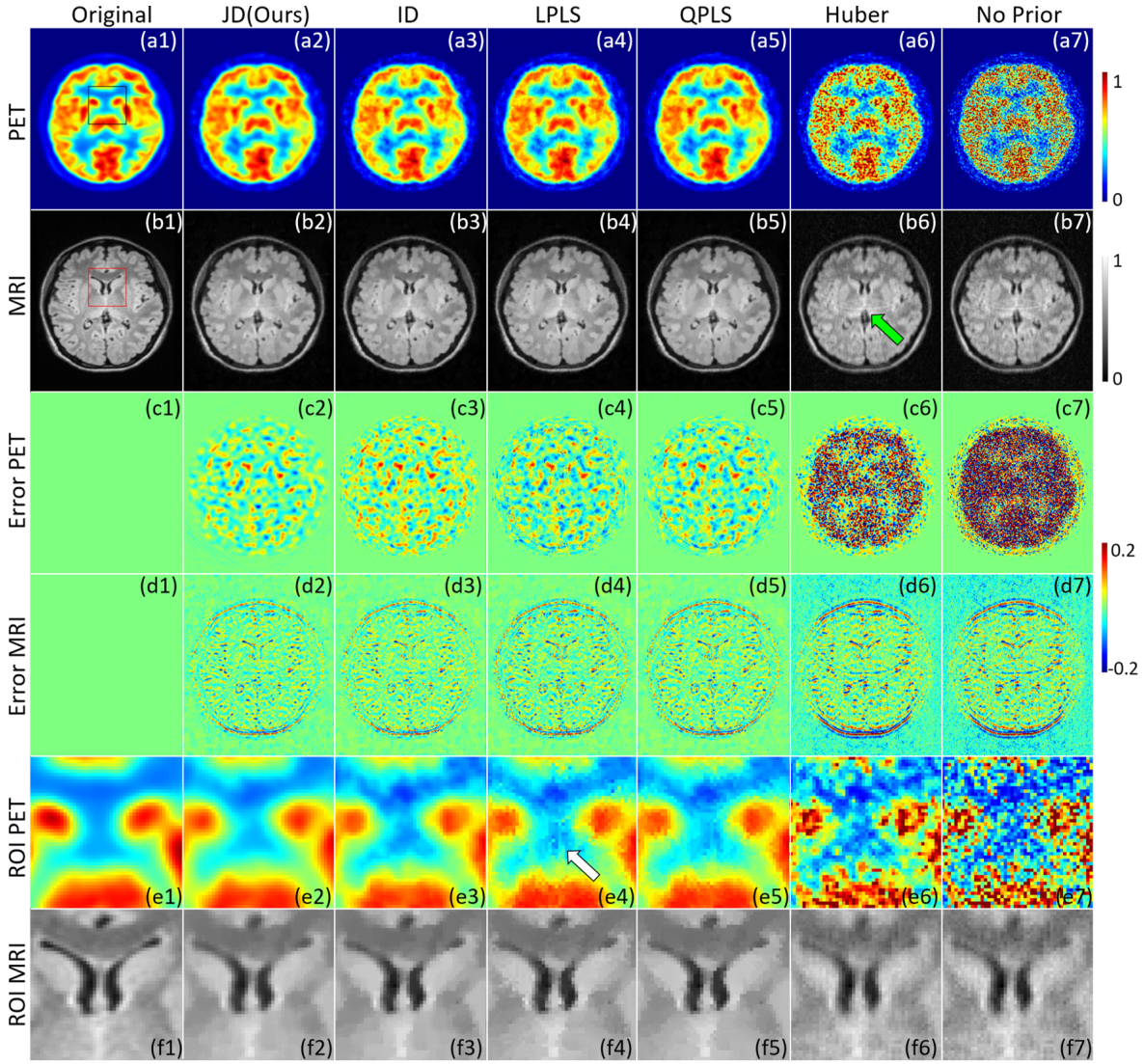

Figure 3.7: **Joint PET-MRI Reconstruction: Results on *in vivo* PET-MRI: (2× Undersampled MRI).** Ground truth images: (a1) PET, and (b1) MRI. Joint PET-MRI reconstructions: using **our JD prior** for (a2) **PET** [RRMSE: 0.08] and (b2) **MRI** [0.07]; using **LPLS prior** for (a4) **PET** [0.10] and (b4) **MRI** [0.08]; using **QPLS prior** for (a5) **PET** [0.09] and (b5) **MRI** [0.08]. Independent PET and MRI reconstructions: using **ID prior** for (a3) **MRI** [0.11] and (b3) **PET** [0.09]; using **Huber prior** for (a6) **MRI** [0.11] and (b6) **PET** [0.09]; without using any prior for (a7) **MRI** [0.14] and (b7) **PET** [0.14].

in vivo PET and MRI ground-truth images (Figure 3.7(a1)-(b1)) have larger voxel sizes and richer texture compared to the BrainWeb images in Figures 3.3, 3.4 and Figure 3.5. The Huber prior (Figure 3.7(a6)) removes noise over the MLEM reconstruction (Figure 3.7(a7)). The LPLS and QPLS priors for PET reconstruction (Figure 3.7(a4)-(a5)) are unable to recover the original regularity (smoothness) and tend to produce cartoon-

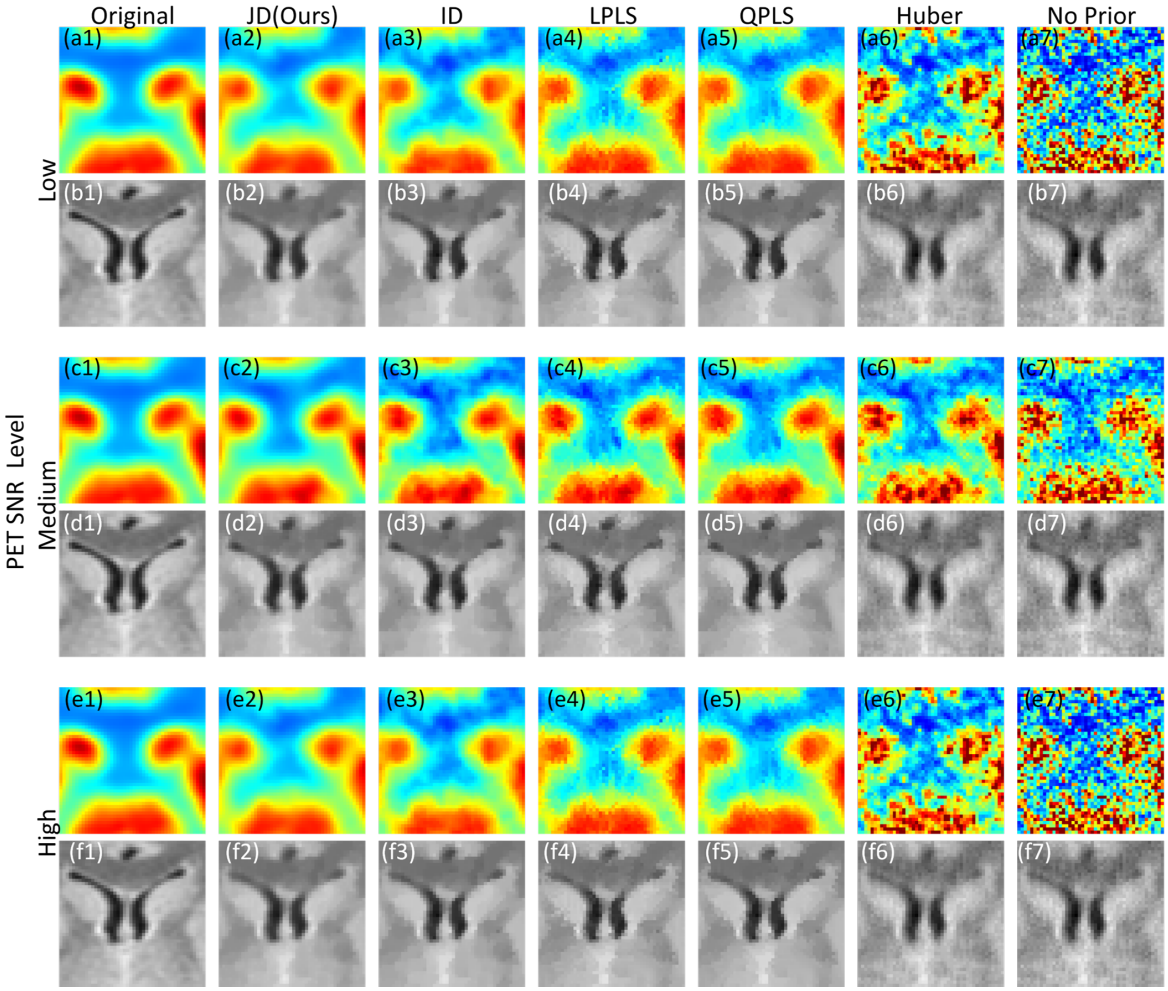

Figure 3.8: **Joint PET-MRI Reconstruction: Results on *In vivo* PET-MRI: Three different SNR levels in PET.** Zoomed regions from ground truth: (a1),(c1),(e1) PET, and (b1),(d1),(f1) MRI. PET SNR levels: (a1)–(a7) and (b1)–(b7) low (c1)–(c7) and (d1)–(d7) medium (e1)–(e7) and (f1)–(f7) high (a2)–(f2) using **our JD prior**; (a4)–(f4) using **LPLS prior**; (a5)–(f5) using **QPLS prior**. Independent PET and MRI reconstructions: (a3)–(f3) using **ID prior**; (a6)–(f6) using **Huber prior**; (a7)–(f7) without using any prior.

ish reconstructions. The ID based reconstruction for PET(Figure 3.7(a3)) does recover smoothly varying structures, but overall, there seems to be a loss in terms of contrast compared to the corresponding PET reconstruction using JD prior (Figure 3.7(a2)). Our method performs much better at reducing noise and preserving textural regularity. We do observe the spatial correlations in the residual image (Figure 3.7(c2)), which is typical for regularized estimates. Nevertheless, compared to the LPLS and the QPLS residual images (Figure 3.7(c4)-(c5)), our method has reduced magnitudes and also lesser spatial correlations (Figure 3.7(c2) and (d2)). For MRI, the independently reconstructed images, without any prior and with the Huber prior (Figure 3.7(b6),(b7)), retain aliasing artifacts

(indicated by arrow in Figure 3.7(b6)); nevertheless, the Huber prior performs better at reducing noise. The LPLS and QPLS reconstructions perform significantly better for MRI images compared to Huber and prior-less reconstructions; in this case, the ID based MRI reconstruction produces images comparable in visual quality to the joint priors.

Figure 3.8 shows zoomed regions of interest (same as in Figure 3.7), for both PET and MRI images across three different SNR levels of PET data (increasing top to bottom). The zoomed inset images corresponding to the the low-SNR case (Figure 3.8(a1)–(a7) and (b1)–(b7)) are the same as the zoomed insets from Figure 3.7. Joint prior methods as well as ID based MRI reconstructions, across different noise realizations, are of comparable image quality. However, significant difference in PET image reconstructions is seen among all methods, especially for the low-SNR case(Figure 3.8(a1)–(a7)). LPLS and QPLS fail to recover the smooth edges for the PET image reconstructions at low-SNR(Figure 3.8(a4)-(a5)). The ID-based reconstruction for PET shows loss of contrast at low-SNR PET data, whereas the image quality is comparable to JD at high-SNR. We observe that all the methods work reasonably well in the high-SNR case (Figures 3.8 (e1)–e(7)). At low SNR, Huber still retains noise, whereas ID and the joint priors: LPLS, QPLS, and JD, remove significant amount of noise.

Quantitative Results. Figure 3.9 shows the RRMSE and SSIM plots for PET-MRI reconstructions for mid-brain slices of 25 subjects from the evaluation set. We present quantitative analysis across three varying SNR levels of PET data (same as described before), and with 2× k-space undersampling. As mentioned before, we use the hyperparameters tuned on a validation set for our evaluation set. For all the methods, as SNR increases, for PET reconstructions, we see a reduction in RRMSE values and increase in SSIM values. As expected, there is little difference between PLS based methods and the patch-based ID and JD reconstructions for MRI. Clearly, our JD prior based reconstructions for both PET and MRI provide (i) lower RRMSE values and (ii) higher SSIM values, in comparison to all other methods across all noise levels.

3.3.12 Effect of Increasing the Prior / Regularization-Term Weight

Figure 3.10 shows the effect of modifying the prior-weight, α, for the joint reconstruction priors considered in this work, i.e., LPLS, QPLS and JD. Although all methods provide smooth reconstructions for higher value of α (Figure 3.10 (a3)–(f3)), the reconstructions with minimum RRMSE are shown in Figure 3.10 (a2)–(f2). As expected, for lower values of α (Figure 3.10 (a1)–(f1)), there is considerable noise in the reconstructed images. The LPLS prior based reconstructions (Figure 3.10 (e1)–(e3) and (f1)–(f3)) provide sharper images with increasing α, but also tend to show increased staircasing ef-

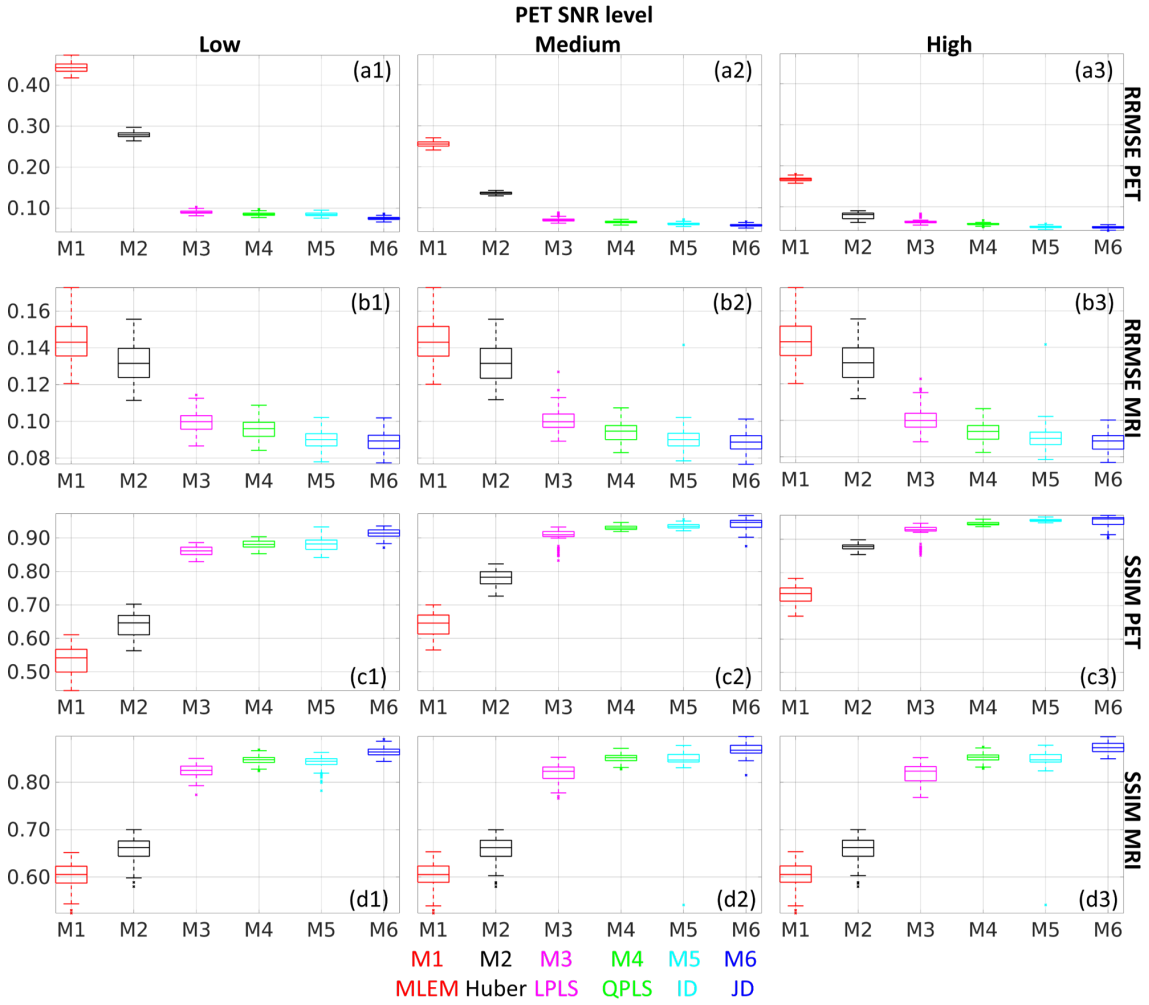

Figure 3.9: **Results on *In vivo* PET-MRI: With 2× Undersampled MRI; Three Different Noise Levels in PET.** RRMSE and SSIM values for PET and MRI images, on 10 mid-brain slices across 25 subjects in the evaluation set. (a1)–(d1) **PET SNR low**; (a2)–(d2) **PET SNR medium**; (a3)–(d3) **PET SNR high.** RRMSE values for all the methods for: (a1) **PET** with low SNR PET data, (a2) **PET** with medium SNR PET data, (a3) **PET** with high SNR PET data (b1) **MRI** with low SNR PET data, (b2) **MRI** with medium SNR PET data, (b3) **MRI** with high SNR PET data SSIM values for all the methods for: (c1) **PET** with low SNR PET data, (c2) **PET** with medium SNR PET data, (c3) **PET** with high SNR PET data (d1) **MRI** with low SNR PET data, (d2) **MRI** with medium SNR PET data, (d3) **MRI** with high SNR PET data

fect. Moreover, the strength of the cross-modality artifact in the PET images (from the MRI-specific lesion), increases as α increases. Compared to LPLS, the QPLS based reconstructions (Figure 3.10 (c1)–(c3) and (d1)–(d3)) are less prone to these artifacts. On

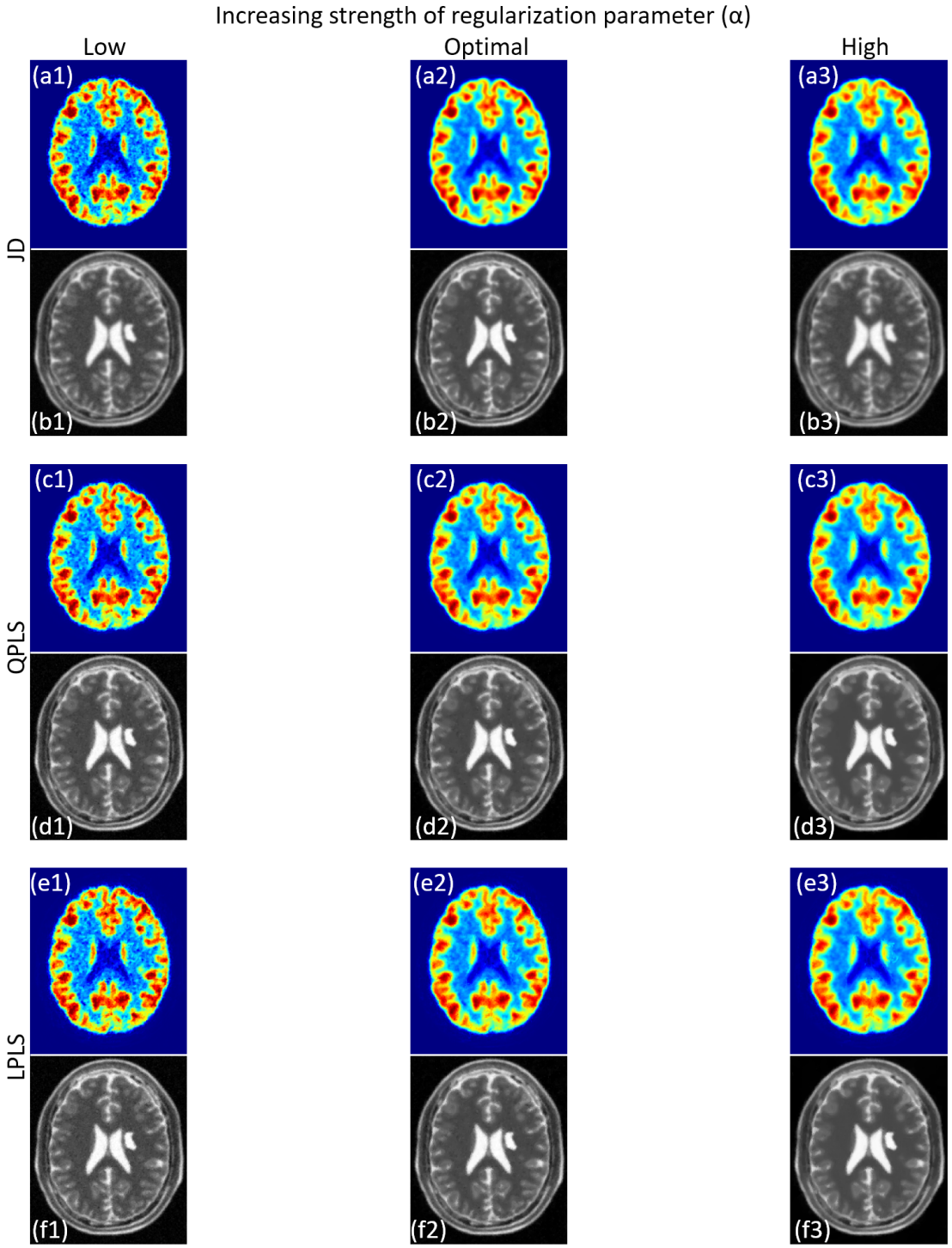

Figure 3.10: **Effect of Increasing the Prior / Regularization Term Weight (α). Comparison among joint reconstruction priors.** Joint PET-MRI reconstructions: using **our JD prior** for PET: (a1)–(a3) and MRI: (b1)–(b3); using **QPLS prior** for PET: (c1)–(c3) and MRI: (d1)–(d3); using **LPLS prior** for PET: (e1)–(e3) and MRI: (f1)–(f3);

the other hand, our JD based reconstructions (Figure 3.10 (a1)–(a3) and (b1)–(b3)) show

increased smoothing as α increases. For higher values of α, over-regularization is seen across all methods; for JD this leads to increasing blur in the images; for both LPLS and QPLS, this leads to increasing staircasing effects.

3.3.13 Joint PET and Parallel-MRI Reconstruction: Results on Undersampled Parallel-MRI (*In Vivo*)

We use gradient-echo T2w MRI and FDG-PET for 5 subjects in a Siemens mMR Biograph scanner, with 12 head coils. We use reconstructions produced by the on-scanner software as ground truth (Figure 3.11(a1)-(b1)), and then retrospectively obtain MRI-PET measurements with parallel MRI with 2× k-space undersampling. The *in vivo* images are far from being piecewise constant, and exhibit weaker inter-modality dependencies between image gradients. TV based priors are prone to cartoonish reconstructions (Figure 3.11(b3)), especially in PET, compared to our reconstructions (Figure 3.11(a2)-(b2)) that are more regular and natural, and better at preserving edges, higher-order features such as texture, and modality specific features. The inter-pixel correlations, at a larger scale, are more prominent in the joint regularization techniques (Figure 3.11(d2)-(d3)) compared to the Huber-prior residual image (Figure 3.11(d3)), resulting in structural bias in the reconstructed images. Nevertheless, the joint priors yield better SNR-bias trade-off compared to the Huber prior based reconstructions (Figure 3.11(b2)-(b4)) and (Figure 3.11(f2)-(f4)).

Quantitative Results. For *in vivo* data, we reconstruct 20 slices over 4 subjects different from the one used to learn the dictionary. Our joint-dictionary model typically gives improved RRMSE (lower) and SSIM (higher) compared to other methods (Figure 3.12), for both the simulated BrainWeb images and *in vivo* images. While JD and JSR perform similarly on BrainWeb MRI, our JD prior performs much better on BrainWeb simulated PET images. Huber and JSR priors perform better for BrainWeb simulations (where the true images are closer to being piecewise constant), compared to *in vivo* images.

3.4 Discussion

In this work, we present a novel framework for the reconstruction of PET and accelerated-MRI data, for simultaneous PET-MRI scanners, using a joint patch-based dictionary. Our joint reconstruction formulation models the PET-MRI joint dictionary as a MRF, leading to a Bayesian inference framework. PET reconstruction combined with three different MRI data acquisition scenarios were handled: (i) scanner-reconstructed fixed MRI image; (ii) undersampled k-space data from single channel; and (iii) undersampled k-space

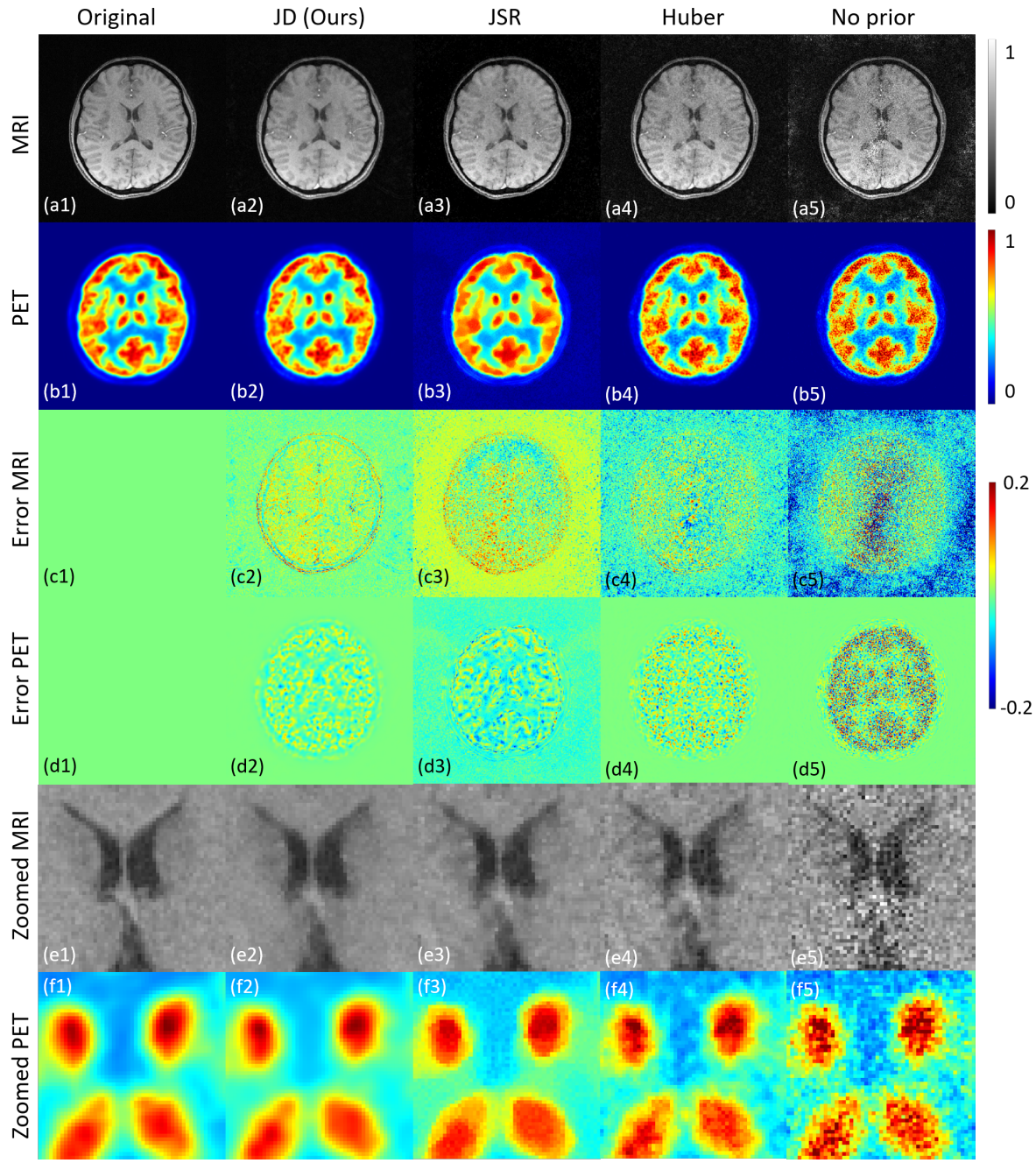

Figure 3.11: **Joint PET-MRI Reconstruction (parallel-MRI): Results on *in vivo* PET-MRI: (2× Undersampled MRI).** *Ground truth* for **(a1)** MRI, **(b1)** PET. *Joint reconstructions with* **our JD prior** for **(a2)** MRI [RRMSE: 0.07], **(b2)** PET [0.04]; and *with JSR prior* for **(a3)** MRI [0.10], **(b3)** PET [0.09]. *Independent reconstructions with Huber prior* for **(a4)** MRI [0.15], **(b4)** PET [0.07]; and *without prior* **(a5)** MRI [0.27], **(b5)** PET [0.13]. **(c1)–(c5)** and **(d1)–(d5)** Residuals. **(e1)–(e5)** and **(f1)–(f5)** Zoomed insets.

data from multiple channels. Furthermore, the proposed framework is capable of handling various Cartesian undersampling schemes for k-space data acquisition. We provide

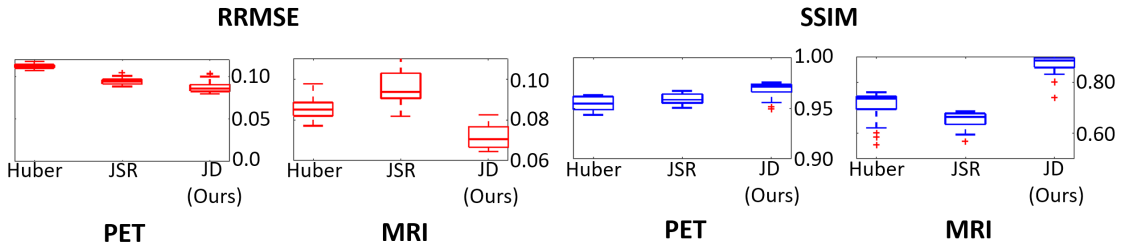

Figure 3.12: **Joint PET-MRI Reconstruction (parallel-MRI): Results on *in vivo* PET-MRI:** (2× **Undersampled MRI).** *In vivo*: 20 slices in 4 subjects.

the detailed theoretical framework using EM to solve the PET-MRI joint-reconstruction optimization problem. To the best of our knowledge, this work is the first to propose a patch-based dictionary prior for the joint reconstruction problem.

Our patch-based joint dictionary models intra-modality and inter-modality textural patterns and produces improved reconstructions over state of the art, especially under increased noise levels. It is well known that gradient-norm-based penalties can be prone to staircasing artifacts and cartoonish reconstructions. The qualitative and quantitative results in the simulated datasets (BW1 and BW20) demonstrate the superior performance of our framework compared to existing methods. In the simulated datasets, the dependencies across the PET and MRI images, at fine and coarse scales both, are stronger compared to the *in vivo* PET-MRI images. Hence, LPLS and QPLS perform better in the simulated BrainWeb cases compared to the *in vivo* cases. Unlike the simulated datasets, the *in vivo* images depict richer textural features where patch-based dictionary models offer better regularity in the reconstructed images, especially at higher noise levels. The texture modeling and larger contextual information (through patch-based modeling) tends to help dictionary methods alleviate cross-modality artifacts, compared to existing methods that model fine-scale gradient-based dependencies. We find that joint modeling and reconstruction of PET and MRI has the ability to improve the reconstruction quality of both PET and MRI images. For example, while the ID prior finds it difficult to remove aliasing artifacts in MRI images (especially in the BW1 dataset experiment, Figure 3.4), the joint priors (LPLS, QPLS, and JD), do much better by using the aliasing-free information in the PET modality. On the other hand, the joint modeling helps improve PET reconstructions by leveraging anatomical information in the MRI images.

Our experiments on the simulated BW20 dataset, on which we evaluate all methods across varying PET SNR levels, demonstrate the generalizability of dictionaries (for both ID and JD) learned from a small training to a different larger evaluation set. As part of future work, further experiments are indeed needed for evaluating generalizability across

dosage levels, patient anomalies, etc. Compared to the hand-crafted prior models, e.g., LPLS and QPLS, learning-based methods, like ID and JD, require some training data for learning the prior model. For instance, neural network based models have proven to be very powerful at learning complex regression models for some medical image reconstruction tasks, but require very large amounts of training data for model learning. Nevertheless, all models, whether hand-crafted or learning-based, to require a validation set for tuning hyperparameters. Our method has a few more hyperparameters in comparison to the LPLS and QPLS priors, i.e., the number of atoms in the dictionary, the patch size, and the sparsity parameter λ. Tuning a larger number of hyperparameters requires a larger validation set, in general. Nevertheless, we find that our reconstruction results are fairly robust to small perturbations in all these hyperparameters.

In the case of joint reconstruction of PET and parallel-MRI, for the same acceleration factor as the single-coil case, the SENSE reconstructed images do not suffer from aliasing artifacts. Hence, for the same acceleration factor, in parallel-MRI, the problem is to perform reconstruction with enhanced noise removal in both PET and MRI. Our joint dictionary shows superior image quality to the joint TV method by restoring smoother PET images and MRI images with improved textural information.

Computational Cost. We implement our joint-dictionary-based joint PET-MRI reconstruction framework in MATLAB R2018a. We run our experiments on a Intel Xeon E5-2680 2.5 GHz workstation. The average time to reconstruct an *in vivo* dataset (whole brain) is around 5 hours. We observe that the algorithm typically requires around 5 EM iterations and within each EM iteration, the algorithm typically requires less than 30 iterations of projected gradient descent for updating each of the variables (U, V, C) within each M-step. As is typical of patch-based techniques, our method is more computationally intensive compared to the image gradient-based techniques. Nevertheless, there is scope for substantial acceleration by parallelization. Similar to several patch-based methods, our JD-based formulation faces increased computational cost compared to the gradient-based methods. However, the dictionary model can be parallelizable to reduce computation time. The current evaluation is on a cohort of 25 patients. In the future, we plan to evaluate on larger datasets.

Dictionary-based methods have been successful in (i) accurate reconstructions of MRI images with highly accelerated data acquisition, e.g., **?**, and (ii) restoring images under low SNR, e.g., **??**. These characteristics of dictionary-based models indicate a strong potential of our joint-dictionary framework towards PET-MRI imaging with lower-dose PET scanning combined with faster MRI scanning. Future directions can focus on

joint reconstruction with lower doses in PET and highly accelerated multi-contrast parallel MRI.

Chapter 4

Deep Learning based Robust to OOD Ultra-Low-Dose PET

4.1 Introduction

Chapter 2 described how lowering the PET radiotracer dose degrades image quality and affecting diagnostic requirements. Hence, synthesizing PET images that match the image quality of standard-dose PET (SD-PET) is of importance. As in the previous chapter, we restrict our discussion to FDG-PET. Chapter 2 also briefly discussed several approaches proposed in the literature to enable low-dose imaging. These techniques were either (i) within reconstruction, or (ii) post-reconstruction PET image enhancement techniques.

This chapter proposes a novel framework for the synthesis of SD-PET images from multimodal data that exploits the underlying physics of the imaging system and also models the per-voxel heteroscedasticity in the training data to predict SD-PET images from LD-PET images (DRF $\approx 180\times$) and robustness to OOD acquisitions of PET data. To the best of our knowledge, this work is the first work to model both imaging physics as well as uncertainty for the task of medical image synthesis using DNNs. Thus, we propose a sinogram- and uncertainty- based DNN (suDNN) framework. Our results on a cohort of 28 subjects with *in vivo* PET acquisitions demonstrate improved (i) quality of the reconstructed images and (ii) robustness of the learned model in reconstructing OOD PET data in comparison to state-of-the-art methods.

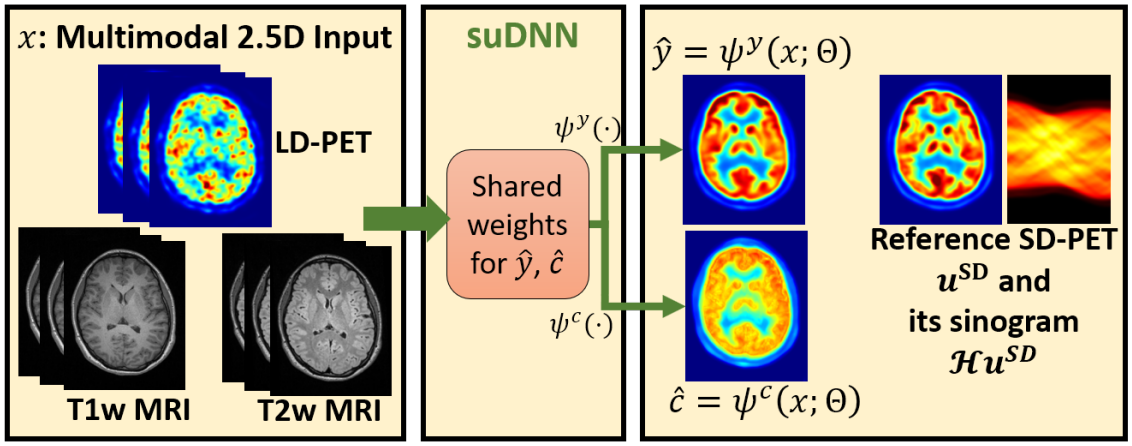

Figure 4.1: **Illustration of the proposed suDNN framework**. The input images are multimodal PET and MRI for the 2.5D-style training scheme. The outputs of the network are (i) the mean predicted image $\widehat{y}$, and (ii) the image $\widehat{c}$ modeling the per-voxel variances in the residuals between the predicted image and the reference SD-PET image.

4.2 Methods

We elucidate the underlying mathematical model along with the network architecture and the training strategy used for the proposed suDNN framework that estimates SD-PET images using the multimodal input data.

4.2.1 Model

Let the set of acquired LD-PET images for N subjects be $\{u_n^{\text{LD}}\}_{n=1}^N$. Similarly, the set of T1w MRI, T2w MRI images, and the SD-PET images are denoted by $\{v_n^{\text{T1}}\}_{n=1}^N$, $\{v_n^{\text{T2}}\}_{n=1}^N$, and $\{u_n^{\text{SD}}\}_{n=1}^N$, respectively. For the $n-$th subject, the images $u_n^{\text{LD}}, v_n^{\text{T1}}, v_n^{\text{T2}}$, and u_n^{SD}, each containing I voxels, are spatially co-registered. Our task is to predict an estimate of the SD-PET image, given the triplet of images $x_n := \{u_n^{\text{LD}}, v_n^{\text{T1}}, v_n^{\text{T2}}\}$ via a regression function $\Psi(\cdot;\Theta)$, parameterized by Θ. To solve the above mentioned regression problem, we use the training set consisting of T subjects given by $\mathcal{T} := \{x_n \cup u_n^{\text{SD}}\}_{n=1}^T$ where $T < N$. We realize the multidimensional regression function $\Psi(\cdot;\Theta)$ as a DNN that takes the triplet x_n as input and produces as output (i) an estimate of the SD-PET image $\widehat{y}_n$ and (ii) the associated per-voxel uncertainty maps $\widehat{c}_n$, as formulated in the subsequent subsections. Figure 4.1 shows the proposed uncertainty-aware deep learning framework suDNN detailed below.

4.2.2 Uncertainty-aware Physics-based Loss Function

To solve the multidimensional regression problem stated above, oftentimes, while modeling the loss function for the DNN, the residual between the output of the network $\widehat{y}$ and the reference image u^{SD} is assumed to be an isotropic standard Gaussian distribution with zero mean and a single fixed standard deviation across all voxels. This assumption does *not* account for heteroscedasticity of the per-voxel residuals between the predicted PET image and the reference SD-PET image. We show that modeling the per-voxel heteroscedasticity in the residual images can potentially lead to the robustness of the learned model to newly acquired OOD PET test data. Thus, for the n-th subject, we model the output of suDNN as a pair consisting of (i) the predicted PET image $\widehat{y}_n$ and (ii) the image $\widehat{c}_n$ modeling the per-voxel variances in the residuals between the predicted image and the reference SD-PET image. We propose a DNN that is based on a U-Net architecture(**?**) with a dropout model (**?**) in its bottleneck layer, to solve the regression problem. The proposed suDNN differs from the standard U-Net by employing: (i) a 2.5D-style (similar to the strategy in (**?**)) input with multimodal images as multi-channel input and (ii) dual-channel output, where one channel estimates the mean image $\widehat{y}_n$ and the other channel estimates the variance image $\widehat{c}_n$, inspired by (**?**). That is, we predict both $\widehat{y}_n$ and $\widehat{c}_n$ using a DNN with split-head producing two images as output (one image per head) as shown in Figure 4.1. Let the proposed DNN be represented by

$$\Psi(\cdot;\Theta) := (\psi^{y}(\cdot;\Theta), \psi^{c}(\cdot;\Theta)). \tag{4.1}$$

Therefore, for a given multimodal input x_n, the outputs of the DNN are given by $\widehat{y}_n := \psi^{y}(x_n;\Theta)$ and $\widehat{c}_n := \psi^{c}(x_n;\Theta)$.

For improved robustness to OOD acquisition, within the loss function of the DNN, we propose to enforce similarity in the (i) spatial image domain and (ii) sinogram domain modeling the PET detector geometry. Further, we model the per-voxel variances in the image space that consequently models the associated uncertainty in the sinogram space as well. Let the uncertainty-aware loss in the image space be $\mathcal{L}_{\text{U}}$ and the uncertainty-aware PET-physics-based loss in the sinogram space be $\mathcal{L}_{\text{P}}$. The overall loss function of the DNN $\mathcal{L}_{\text{PU}}$ is a weighted combination of the two losses $\mathcal{L}_{\text{U}}$ and $\mathcal{L}_{\text{P}}$, described subsequently.

Uncertainty-aware Image Space Loss $\mathcal{L}_{\text{U}}$. Let $[\widehat{y}_n]_i$ represent the i-th voxel in the image $\widehat{y}_n$, and let $[\widehat{c}_n]_i$ represent the i-th voxel in the image $\widehat{c}_n$. We model a Gaussian likelihood in the image space with a per-voxel mean $\widehat{y}_n$ and per-voxel variance $\widehat{c}_n$. Thus, after taking the negative log-likelihood of the Gaussian probability distribution function

and ignoring the constant terms the image space loss is

$$\mathcal{L}_{\mathrm{U}}(\Theta) := \sum_{n=1}^{N} \sum_{i=1}^{I} \frac{([\widehat{y}_n]_i - [u_n^{\mathrm{SD}}]_i)^2}{[\widehat{c}_n]_i + \epsilon} + \log([\widehat{c}_n]_i + \epsilon), \tag{4.2}$$

where ϵ is a small positive real-valued constant for numerical stability. Equation 4.2 consists of two components: (i) the residual mean-squared error term $([\widehat{y}_n]_i - [u_n^{\mathrm{SD}}]_i)^2$ scaled by the variance $[\widehat{c}_n]_i$, and (ii) the penalty term $\log([\widehat{c}_n]_i + \epsilon)$ corresponding to the per-voxel variance $[\widehat{c}_n]_i$ which penalizes large values of $[\widehat{c}_n]_i$. We enforce positivity on the elements of $\widehat{y}_n$ using ReLU activation function in the final layer of the head modeling ψ^y. Unlike ψ^y, the output values from the head ψ^c are unconstrained . Thus, positivity of $\widehat{c}_n$ is implicitly enforced by employing an exponentiation layer as the final layer of ψ^c. The two terms in Equation 4.2 prevent the predicted per-voxel variance $[\widehat{c}_n]_i$ from tending to zero or infinity. As in **?**, we do *not* require explicit supervision for learning $\widehat{c}_n$. Rather, it is learned implicitly using the loss in Equation 4.2.

Uncertainty-aware Sinogram Space Loss $\mathcal{L}_{\mathrm{P}}$. To model the loss in the sinogram space, let operator $\mathcal{H}$ model the PET forward transformation (linear operator to produce sinogram) associated with the PET image acquisition for each transaxial slice. Operator $\mathcal{H}$ takes a 2D image with I voxels and produces a sinogram with L discrete elements. Given that the per-voxel residual $(\widehat{y}_n - u_n^{\mathrm{SD}})$ in the image space is modeled using a Gaussian distribution, the per-element residual in the sinogram space also follows a Gaussian distribution. Similarly, given that $\widehat{c}_n$ models the heteroscedasticity of the residuals across the voxels in the image space, we model the per-element variances in the sinogram-space residual $(\mathcal{H}\widehat{y}_n - \mathcal{H}u_n^{\mathrm{SD}})$ by the image $\mathcal{H}c_n$. For simplicity, we exclude modeling the covariances between the elements in the sinogram space resulting from the dependencies introduced by the sinogram operator $\mathcal{H}$. Thus, we propose a physics-based loss term

$$\mathcal{L}_{\mathrm{P}}(\Theta) := \sum_{n=1}^{N} \sum_{l=1}^{L} \frac{([\mathcal{H}\widehat{y}_n]_l - [\mathcal{H}u_n^{\mathrm{SD}}]_l)^2}{[\mathcal{H}\widehat{c}_n]_l + \tau} + \log([\mathcal{H}\widehat{c}_n]_l + \tau), \tag{4.3}$$

where τ is a small positive real-valued constant for numerical stability.

Overall loss function $\mathcal{L}_{\mathrm{PU}}$. We propose to optimize the set of parameters Θ of our DNN by minimizing the overall loss function consisting of uncertainty-aware loss functions in both the image-space and the sinogram-space given by

$$\mathcal{L}_{\mathrm{PU}} := \mathcal{L}_{\mathrm{U}} + \lambda \mathcal{L}_{\mathrm{P}}, \tag{4.4}$$

where λ is a positive real-valued free parameter that controls the weight of the phyics-based loss $\mathcal{L}_{\mathrm{P}}$. In this work, we tune the value of λ using a validation set.

4.2.3 DNN Architecture and Learning Strategy

We employ a U-net architecture comprising an encoder and a decoder that share a symmetric structure (with skip connections across encoder and decoder) each consisting of 3 convolutional blocks that downsample/upsample the feature-maps by a factor of 2 using max-pooling. The initial layer at the beginning of the encoder uses a 1×1 convolutional layer to fuse the multimodal LD-PET and MRI images to produce a single-channel image. We use standard batch-normalization (**?**) and ReLU activation function after every convolutional layer. The bottleneck layer is followed by a dropout layer for regularization (**?**). The output of the penultimate convolutional block in the decoder is fed into two identical convolutional blocks, one for predicting the mean ($\widehat{y_n}$) and one for predicting the variance ($\widehat{c_n}$). The proposed network was trained using Adam (**?**) optimizer with ℓ_2 regularization on the weights, for 500 epochs, with an initial learning rate $\gamma = 3e^{-5}$. We employ a cosine annealing strategy for scheduling γ. We employ a dropout-probability value of $1e^{-3}$ during training as well as inference.

4.3 Experiments and Results

This section describes the *in vivo* data acquired for this work, followed by the description of the baseline methods used for comparison. Subsequently, we describe the experiments used to evaluate the robustness of all the methods to OOD degradations in the input data. Finally, we perform an ablation study to analyze the contribution of each component (multimodal input, individual loss terms) in the proposed suDNN framework.

4.3.1 In Vivo Data

We collected PET and MRI data from a cohort of 28 healthy individuals on a 3T Siemens Biograph mMR system. The average dose administered for each subject was ≈230 MBq. The MRI contrast images, i.e., ultra-short echo time (UTE), T1 MPRAGE, and T2-SPC, were acquired during the PET scan. The SD-PET image was reconstructed using counts obtained over a duration of 30 minutes, starting 55 minutes after the administration of the tracer. The total number of useful counts over the 30-minute duration used for reconstruction of the SD-PET image was around 600×10^6. To simulate the LD-PET data, we randomly selected around 3.4×10^6 counts, spread uniformly over the scan duration, resulting in a DRF of around 180×. For attenuation correction, pseudo-CT maps generated using the UTE images (**?**) were employed. Both the SD-PET and LD-PET images were reconstructed using proprietary software using ordinary-Poisson ordered-subset expectation-maximization (OP-OSEM) algorithm with 3 iterations and 21

subsets, along with point spread function (PSF) modeling and a post-reconstruction Gaussian smoothing. The software produced reconstructed PET images of voxel sizes $2.09 \times 2.09 \times 2.03$ mm^3. The voxel size for the reconstructed MRI images was 1 mm^3 isotropic. For each subject, the LD-PET, SD-PET and the T2w MRI images were registered (using rigid spatial deformation) and resampled to the T1w MRI image space using ANTS (**?**). For the task of predicting SD-PET images from the input set of LD-PET, T1w MRI, and T2w MRI images, we randomly selected 20 subjects for training, 2 subjects for validation, and the remaining for testing. For each subject, we obtained 100 transaxial slices (around 70 slices within the cerebrum and around 30 slices in the cerebellum).

4.3.2 Baseline-methods for Comparison

We evaluate the performance of the proposed suDNN in comparison to recently proposed *five* other DNN-based methods for SD-PET prediction. For a fair comparison, we incorporate a 2.5D-style training scheme for all other methods (similar to ours). The baseline methods, say M1 to M5, are as follows.

- **M1: Conditional DIP.** M1 corresponds to an unsupervised method based on conditional DIP as in (**?**). As the method is unsupervised, it does *not* rely on any training data. As mentioned earlier, the input to the DNN is the structural MRI image. For this method, we use the validation set to tune the optimal number of epochs, to maximize the SSIM between the predicted PET image and the reference SD-PET image.

- **M2: Unimodal ResNet with perceptual loss.** M2 is similar to the framework proposed in (**?**). This method uses the PET image (unimodal) as input, with a standard ResNet architecture, and employs a perceptual loss that is based on features obtained from a VGG network trained on natural images.

- **M3 and M4: 2.5D unimodal and multimodal U-net, respectively.** Both the methods M3 and M4 use the architecture described in (**?**). While M4 uses PET and multi-contrast MRI images as multi-channel input (**?**), M3 takes only the PET images as input (unimodal) (**?**). Both M3 and M4 estimate the residuals between the input LD-PET and the reference SD-PET image.

- **M5: Multi-channel GAN with fused input.** M5 is similar to the GAN-based network described in (**?**). The multi-channel input consists of PET and multi-contrast MRI images. Due to the non-availability of diffusion-weighted MRI images for our dataset, we use only the T1w and T2w MRI images for training. This framework

initially models a learnable 1×1 convolution layer to produce a fused image that becomes the input to the generator of the GAN.

While methods M3–M5 were proposed to achieve DRFs in the range 4–200, M1 and M2 do *not* focus on dose reduction, but on denoising instead. For quantitative evaluation of predicted SD-PET images across all methods, we use (i) peak signal to noise ratio (PSNR) and (ii) structural similarity index (SSIM) (**?**) with respect to the reference SD-PET images.

4.3.3 Simulation of OOD PET data

For training all the DNNs discussed in this work, we use the LD-PET images. However for evaluation of the robustness of all the DNNs in dealing with OOD PET acquisition, we generate two additional sets of test data at increasing degradation levels in the input LD-PET data, namely very low-dose (vLD-PET) and ultra-low-dose (uLD-PET). We generate the OOD test set consisting of vLD-PET and uLD-PET as follows. We retrospectively (i) forward-projected the LD-PET images using the PET forward operator (to produce) $\mathcal{H}$, (ii) generated two differently-scaled instances of the Poisson distribution using the generated sinogram from (i) as the mean, and (iii) performed OSEM reconstruction to get the input vLD-PET and uLD-PET images. The scaling factors for the generated OOD input LD-PET images are 0.1 and 0.01 corresponding to vLD-PET and uLD-PET, respectively. After OSEM reconstruction, the PSNR values, averaged across the test set, between the reference SD-PET image and the set of input images LD-PET, vLD-PET, and uLD-PET were 21, 17, and 13, respectively.

4.3.4 Results: Qualitative and Quantitative

Figure 4.2 shows the predicted images from different methods across three different variations of the LD-PET data, for a representative subject. The input PET images LD-PET, vLD-PET, and uLD-PET are shown in Figures 4.2 (a2), (b2), and (d2), respectively, with Figures 4.2 (a1), (b1), and (d1) showing the corresponding sinograms. The DIP-based M1 (Figure 4.2 (a3), (b3), (d3)) denoises the input LD-PET image. However, as expected, it performs poorly in predicting the FDG uptake in the reference SD-PET image. Unlike M1, the ResNet-based M2 (Figure 4.2 (a4), (b4), (d4)) is designed to predict the activity in the reference SD-PET image. However, even with the LD-PET input, it is unable to produce images with accurate textural features. M2 relies on a standard ResNet architecture that employs short-range skip connections compared to longer-range, hierarchically designed skip connections in the U-net architecture. Methods M3 (Figure 4.2

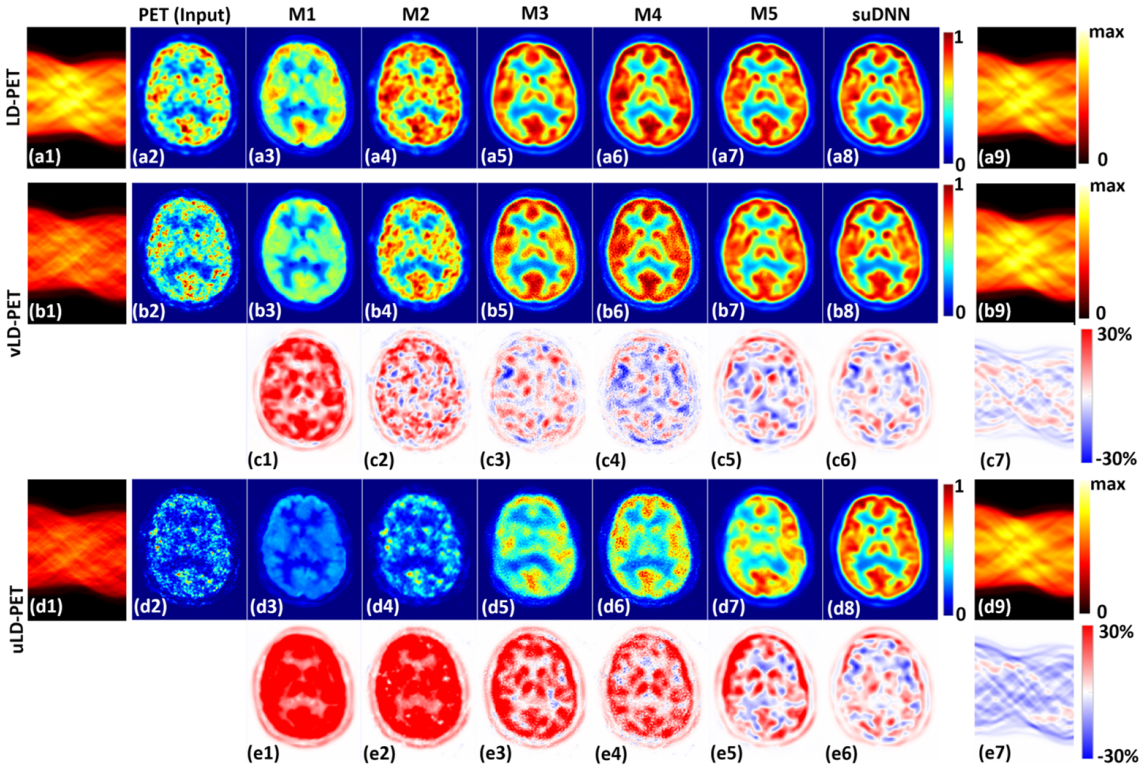

Figure 4.2: **Results for input PET images and sinograms for low-dose (LD) PET (row a), very LD-PET (rows b and c), and ultra LD-PET (rows d and e) respectively**. Panels (a1-a2) show the input LD-PET, (b1-b2) vLD-PET, and (d1-d2) uLD-PET sinograms and images; panels (a3-a8) the predicted images for all methods for LD-PET; panels (b3-b8) and (c1-c6) the predicted images and corresponding residual images (with respect to SD-PET) for vLD-PET; panels (d3-d8) and (e1-e6) the predicted images and corresponding residual images for uLD-PET as input; panels (a9-b9) and (d9) the sinograms of the predicted images (panels (a8, b8, and d8)); and panels (c7) and (e7) show the residuals of the predicted sinograms in comparison to the reference SD-PET sinogram.

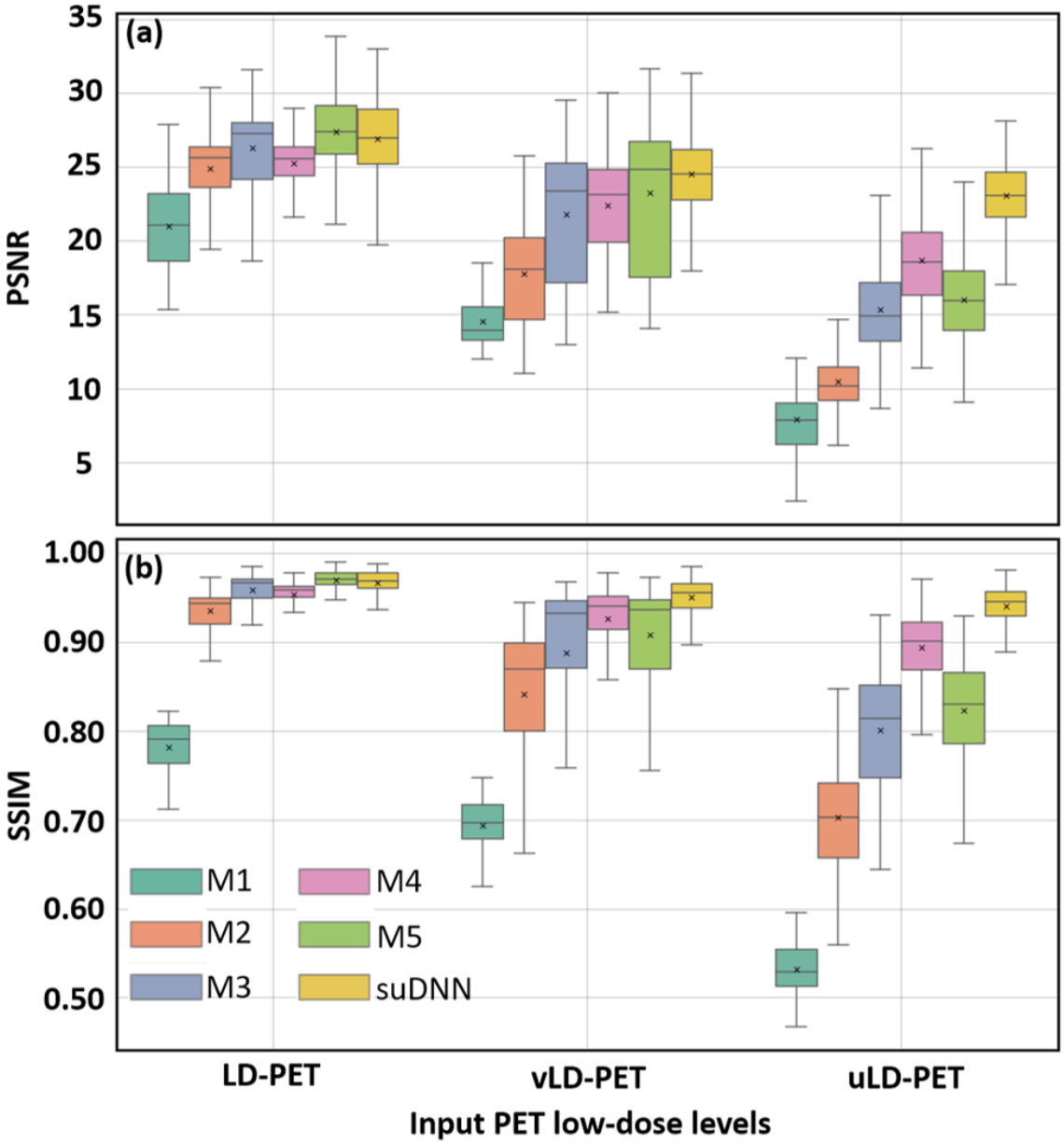

Figure 4.3: **Quantitative evaluation of the methods for three different levels of degradation of the input PET data: LD-PET, vLD-PET, and uLD-PET.** (a) PSNR and (b) SSIM values for the predicted PET images on 100 brain slices for each test set.

(a5), (b5) (d5)) and M4 (Figure 4.2 (a6), (b6), (d6)), which rely on predicting the residual images as output, produce realistic SD-PET images when using LD-PET as the input. However, when using vLD-PET and uLD-PET as inputs, both M3 and M4 show some residual noise in the images despite reasonably recovering the contrast and texture similar to the SD-PET image. M4 improves over the loss in contrast shown by M3, emphasizing the contribution of the multimodal MRI input. M5, which is GAN-based, shows superior performance with LD-PET (Figure 4.2 (a7)), showing little degradation (in terms of contrast and certain structures like the sulci and gyri) with vLD-PET (Figure 4.2 (b7)),

and does *not* predict the desired texture and contrast when using uLD-PET as input (Figure 4.2 (d7)). On the other hand, our suDNN method shows superior prediction across varying levels of input (Figure 4.2 (a8),(b8), and (d8)). Our method shows more realistic texture and contrast, and reduced magnitude in residual images (Figure 4.2 (c6) and (e6)), in comparison to other baselines. For our suDNN, the sinograms of the predicted images (Figures 4.2 (a9), (b9), and (d9)) demonstrate little difference in appearance in compar-

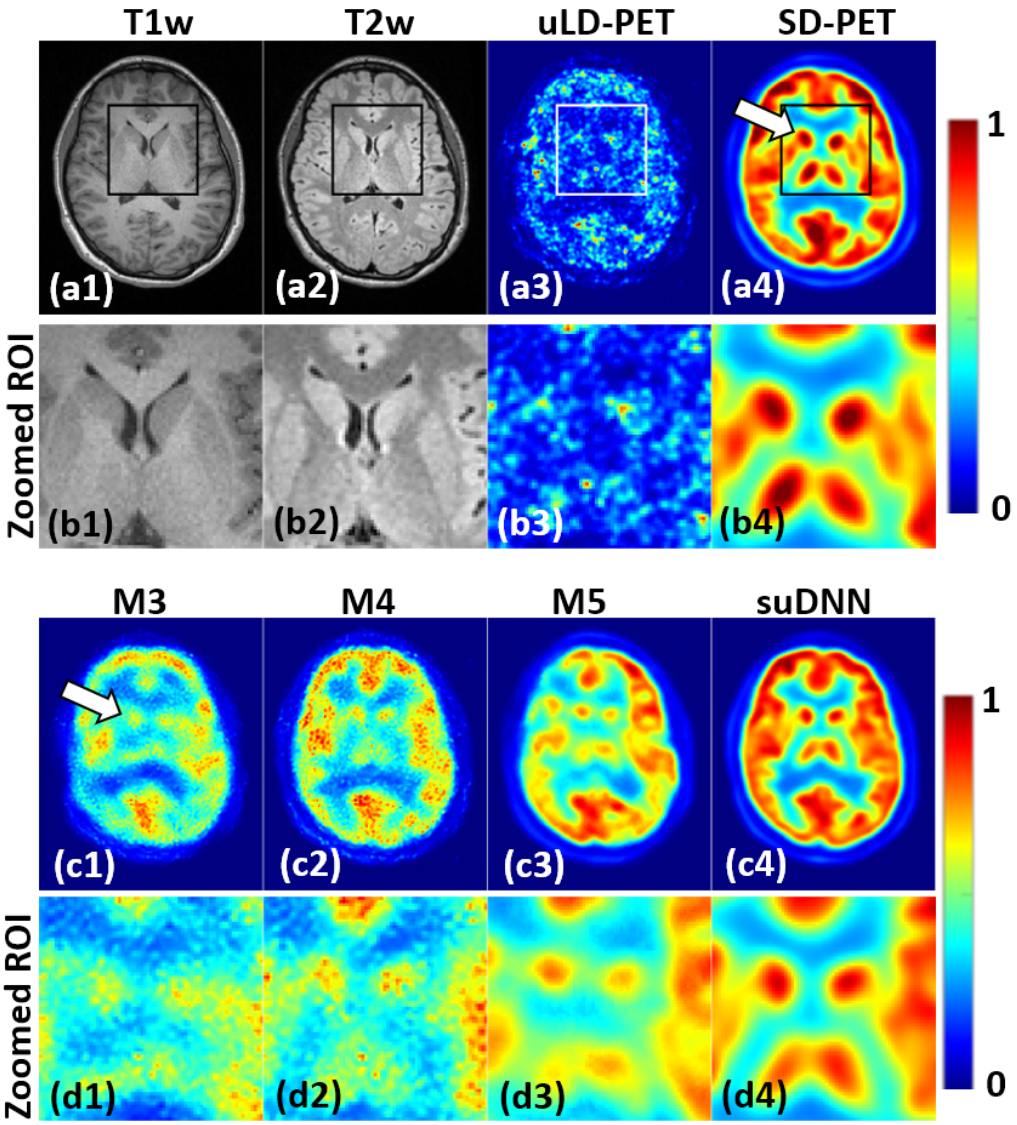

Figure 4.4: **Visual inspection of the zoomed ROIs of the input, reference, and predicted images for the case of uLD-PET. (a1)–(a3)**: input T1w MRI, T2w MRI, and uLD-PET image images. **(a4)**: reference SD-PET image. **(c1)–(c4)**: predicted images from the methods M3–M5 and the proposed suDNN method. **(d1)–(d4)** corresponding zoomed regions.

ison to each other which is in agreement with the image quality of the predicted images obtained with different low-dose inputs. The residual images between the sinograms of the predicted images and that of the reference image SD-PET corresponding to the inputs vLD-PET and uLD-PET are shown in Figure 4.2 (c7) and (d7).

Figures 4.3 (a) and (b) show quantitative plots with PSNR and SSIM values obtained for 100 slices of every subject from the test set for different PET inputs: LD-PET, vLD-PET, and uLD-PET. As the input quality degrades, all methods show a drop in performance. Nevertheless, our method shows the most graceful degradation ($\approx$ 3.5 dB with uLD-PET). On the other hand, the other methods show a severe loss in their performance with uLD-PET (e.g., $\approx$ 10 dB (M5), 7 dB (M4), and 11 (M3)). A similar trend can be observed in the SSIM plot (Figures 4.3 (b)). While our method shows degradation of $\approx$ 0.02 with uLD-PET as input with respect to LD-PET as the input, other methods show a severe decrease in SSIM values with uLD-PET ($\approx$ 0.13 (M5), 0.04 (M4), and 0.1 (M3)). Thus, with LD-PET as input, the performance of suDNN is comparable to M3–M5; nevertheless, as the input degrades, our method significantly outperforms all other methods demonstrating substantially higher robustness/insensitivity to OOD data. We conducted paired *t*-test for SSIM and PSNR values for all methods for the three low-dose inputs. The improvement using our suDNN method was found to be statistically significant ($p \ll 0.001$) in comparison to all other methods (M1–M5) at all inputs LD-PET, vLD-PET, and uLD-PET.

For the results corresponding to uLD-PET input in Figure 4.2, we carefully analyze the predicted images along with the input and the reference images. The zoomed region of interest (ROI) includes the caudate, putamen, and thalamus. The caudate nucleus shows hyperintensity in the SD-PET image (highlighted using the white arrow in Figure 4.4 (a4)) which is *not* the case in the uLD-PET image (Figure 4.4 (a3), (b3)). The unimodal DNN M3 (Figure 4.4 (c1), (d1)) severely underestimates the uptake in the caudate and the thalamus regions. Although our suDNN (Figure 4.4 (c4), (d4)) provides the best estimate of the predicted images, other multimodal DNN methods M4, M5 (Figure 4.4 (c2), (d2) and (c3), (d3)) do show some recovery of the hyperintensity in the caudate and thalamus regions compared to M3. This demonstrates the importance of including the MRI structural images (Figure 4.4 (a1), (b1) and (a2), (b2)) that distinctly show the subcortical nuclei in the cerebrum.

4.3.5 Ablation Study

We perform an ablation study to analyze the contribution from different components in the proposed DNN. To this end, consistent with the prior works in this domain, we

found that using (i) 2.5D input-based training scheme, and (ii) multimodal information as multi-channel input to the DNN, provided substantially improved results in comparison to using 2D-only training and using unimodal input (without MRI information). Moreover, as evident from the results in Figures 4.2–4.4, methods that rely on predicting the residual between the LD-PET and the SD-PET images, are *not* robust to OOD acquisitions.

Hence, to evaluate the importance of multiple components in the proposed suDNN framework, we evaluate *four* DNNs Ablation-DNN1–Ablation-DNN4, as part of the ablation study, described here.

- **Ablation-DNN1: 2.5D Unimodal U-net.** We define a basic DNN which includes a U-net architecture similar to (**?**) with a unimodal input but with a modified output such that it predicts the PET image instead of the residual between the input LD-PET and the reference SD-PET image (as in M3). Ablation-DNN1 is trained using the 2.5D scheme, penalizing the mean-squared error in the image space, say $\mathcal{L}_{\mathrm{I}}(\widehat{y}, u^{\mathrm{SD}})$, between the predicted and the reference images.

- **Ablation-DNN2: 2.5D Multimodal U-net.** We modify the DNN Ablation-DNN1 by replacing the unimodal input with a multimodal input including multi-contrast MRI images, retaining the same loss function as Ablation-DNN1.

- **Ablation-DNN3: 2.5D Multimodal U-net with manifold loss.** In addition to the loss $\mathcal{L}_{\mathrm{I}}$ (as in Ablation-DNN2), this DNN includes a learned manifold-based loss $\mathcal{L}_{\mathrm{E}}(\widehat{y}, u^{\mathrm{SD}})$ (similar to the perceptual loss in (**?**) or the manifold-based loss in (**?**)); thus, the total loss is $\mathcal{L}_{\mathrm{I}} + \lambda_{\mathrm{E}}\mathcal{L}_{\mathrm{E}}$, where $\lambda_{\mathrm{E}} \in \mathbb{R}^{+}$ controls the weight of the loss term $\mathcal{L}_{\mathrm{E}}$. The learned-manifold based loss relies on learning an autoencoder trained using the set of SD-PET images. The loss function $\mathcal{L}_{\mathrm{E}}$ penalizes the differences between the encodings obtained by applying the encoder (from learned autoencoder) to the predicted PET and reference SD-PET images. That is, $\mathcal{L}_{\mathrm{E}}(\widehat{y}, u^{\mathrm{SD}}; \Phi_{\mathrm{E}}) := \|\Phi_{\mathrm{E}}(\widehat{y}) - \Phi_{\mathrm{E}}(u^{\mathrm{SD}})\|_F^2$, where $\|\cdot\|_F$ represents the Frobenius tensor norm.

- **Ablation-DNN4: 2.5D Multimodal U-net with physics-based loss.** Instead of the the learned-manifold loss in Ablation-DNN3 , Ablation-DNN4 uses a sinogram-space loss $\mathcal{L}_{\mathrm{S}}$ given as $\mathcal{L}_{\mathrm{S}} := \|\mathcal{H}\widehat{y} - \mathcal{H}\widehat{u}^{\mathrm{SD}}\|_F^2$. Thus, the total loss for Ablation-DNN4 is $\mathcal{L}_{\mathrm{I}} + \lambda_{\mathrm{S}}\mathcal{L}_{\mathrm{S}}$, where $\lambda_{\mathrm{S}} \in \mathbb{R}^{+}$ controls the strength of $\mathcal{L}_{\mathrm{S}}$.

The free parameters $(\lambda_{\mathrm{E}}, \lambda_{\mathrm{S}})$ were automatically tuned to $(2e^{-3}, 3e^{-3})$ using the validation set.

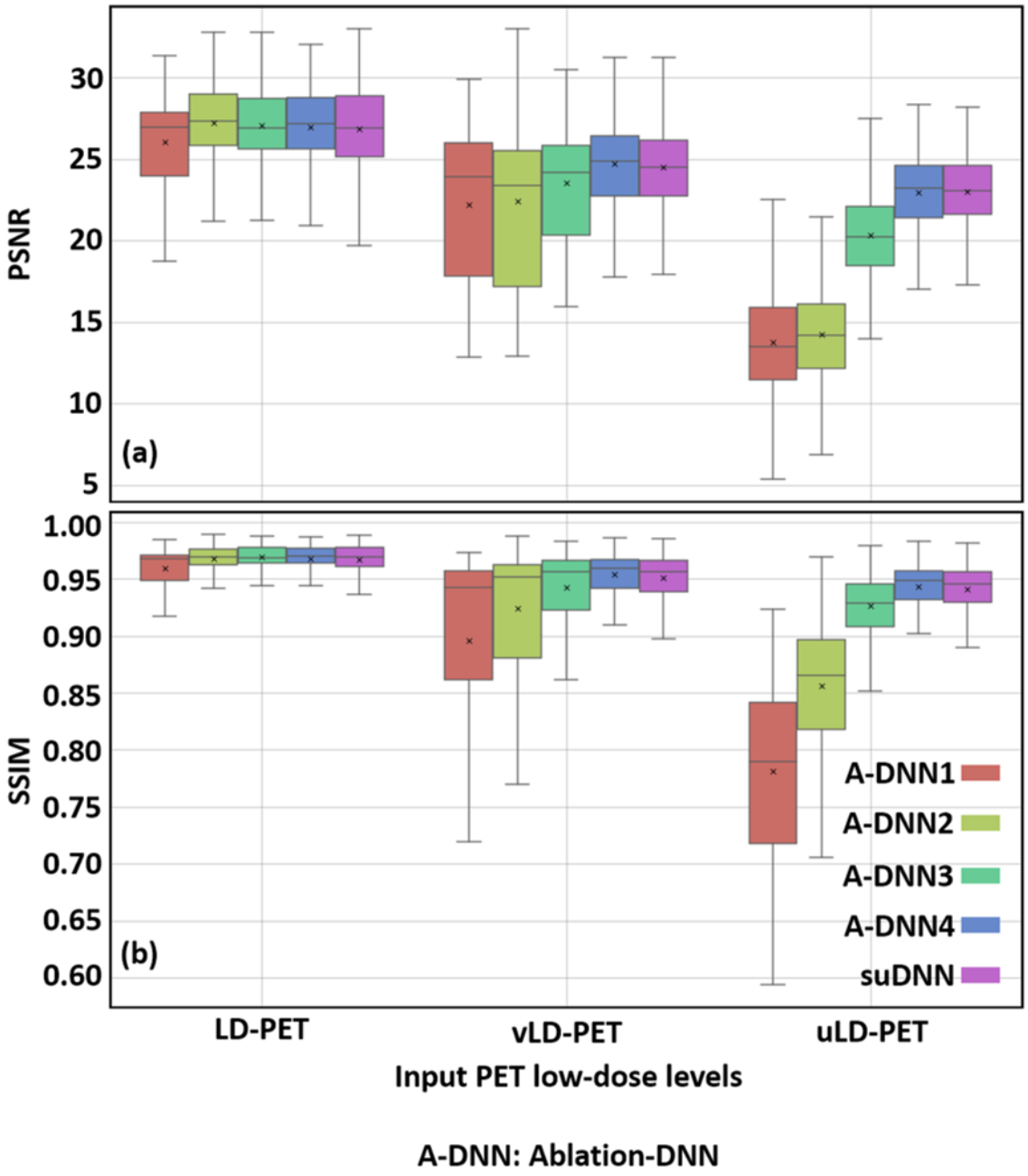

Figure 4.5: **Ablation Study: Quantitative evaluation for the ablation study at 3 different levels of degradation of the input PET data: LD-PET, vLD-PET, and uLD-PET.** **(a)** PSNR and **(b)** SSIM values for predicted SD-PET images, on 100 brain slices in every test set.

Figure 4.5 shows quantitative evaluation of the DNNs in the ablation study: Ablation-DNN1–Ablation-DNN4 and suDNN, for the input PET images LD-PET, vLD-PET, and uLD-PET. Similar to the results in Figure 4.3, DNNs with a multimodal input improve substantially over DNNs with unimodal input (suDNN and Ablation-DNN2–Ablation-DNN4 better than Ablation-DNN1). Inclusion of the learned manifold-based loss $\mathcal{L}_{\mathrm{E}}(\cdot)$, in addition to the image-space loss $\mathcal{L}_{\mathrm{I}}(\cdot)$, for Ablation-DNN3 provides improved robustness over Ablation-DNN2 and Ablation-DNN1. Further, Ablation-DNN4, that includes a physics-based loss instead of the learned manifold-based loss in Ablation-DNN3, shows significant improvement over Ablation-DNN3 with vLD-PET and uLD-PET. Finally, the proposed suDNN that models uncertainty estimation in both image and

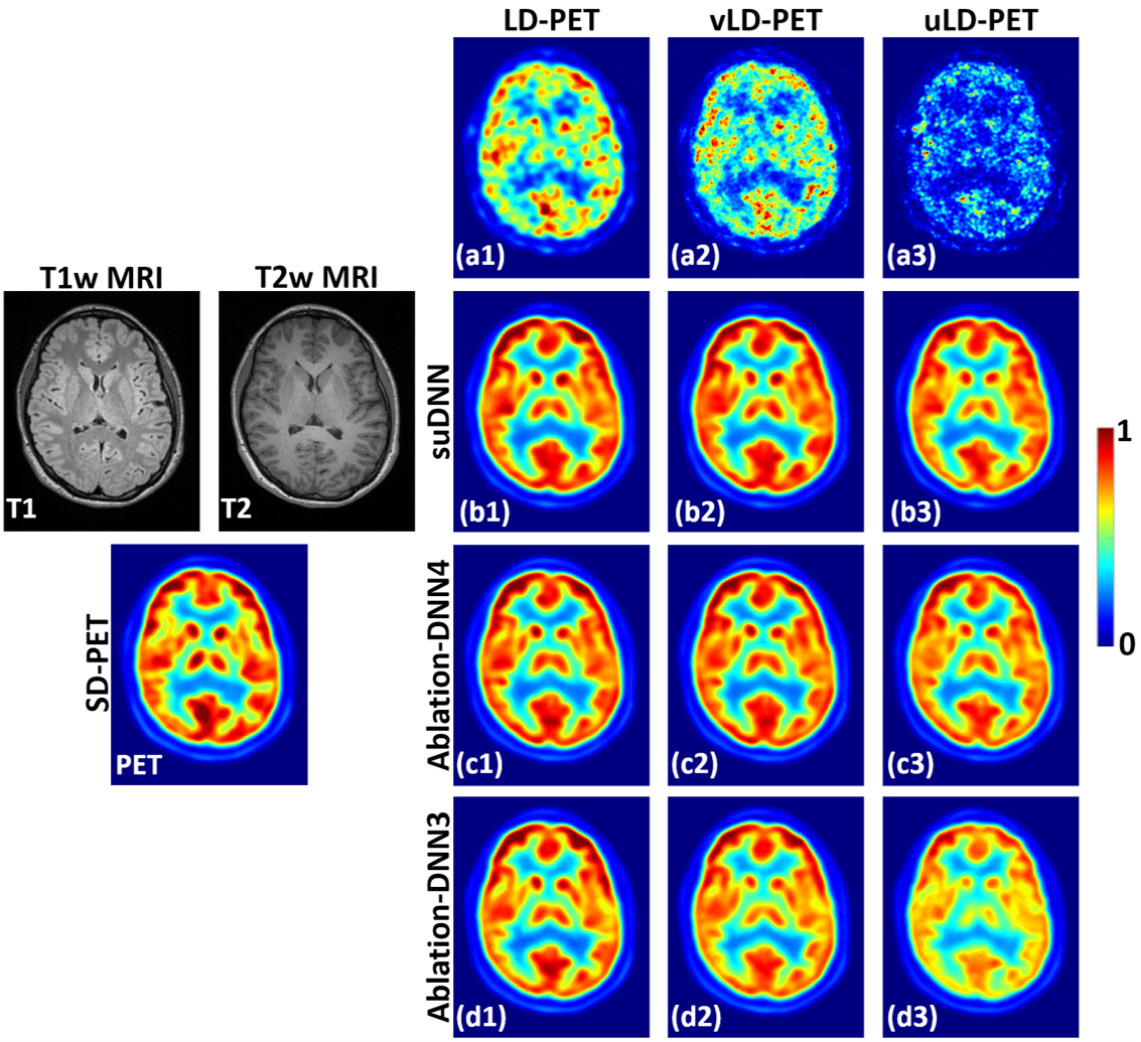

Figure 4.6: **Results of DNNs in the ablation study with input PET images: LD-PET, vLD-PET, uLD-PET.** Variations in input PET **(a1)–(a3)**: LD-PET, vLD-PET, and uLD-PET respectively. Predicted images using varying levels of PET input from: **(b1)–(b3)**: suDNN; **(c1)–(c3)**: Ablation-DNN4; **(d1)–(d3)**: Ablation-DNN3.

sinogram space, provides comparable performance to Ablation-DNN4 with the added advantage of providing the uncertainty maps detailed in Section 4.3.6.

Figure 4.6 shows the predicted PET images from three best-performing (based on Figure 4.5) DNNs in the ablation study: Ablation-DNN3, Ablation-DNN4, and suDNN, for the input PET images LD-PET, vLD-PET, uLD-PET. For the LD-PET input, although the predicted SD-PET image from Ablation-DNN3 (Figure 4.6 (d1)) is closer to that of suDNN and Ablation-DNN4 (Figure 4.6 (b1) and (c1)), Ablation-DNN3 shows substantial degradation with uLD-PET as input (Figure 4.6 (d3)). Clearly, with vLD-PET and uLD-PET as input, the predicted PET images from Ablation-DNN4 (Figure 4.6 (c2) and (c3)) show improved recovery of contrast and texture compared to the corresponding images from Ablation-DNN3(Figure 4.6 (d3) and (d4)). This emphasizes the contribution of the sinogram-based loss function in comparison to the encoder-based loss $L_{\mathrm{E}}(\widehat{y}, u^{\mathrm{SD}}; \Phi_{\mathrm{E}})$ for Ablation-DNN3. As the input LD-image degrades (to vLDPET and uLD-PET), we see a degradation of the predicted images from all the DNNs. However, all the three DNNs (Ablation-DNN3, Ablation-DNN4, and suDNN) are able to recover the overall structural regularity even with uLD-PET as input. The predicted images for suDNN and Ablation-DNN4 for the uLD-PET input are comparable (Figure 4.6 (b3) and (c3)). This reinstates that modeling the uncertainty does not affect the predicted PET image quality.

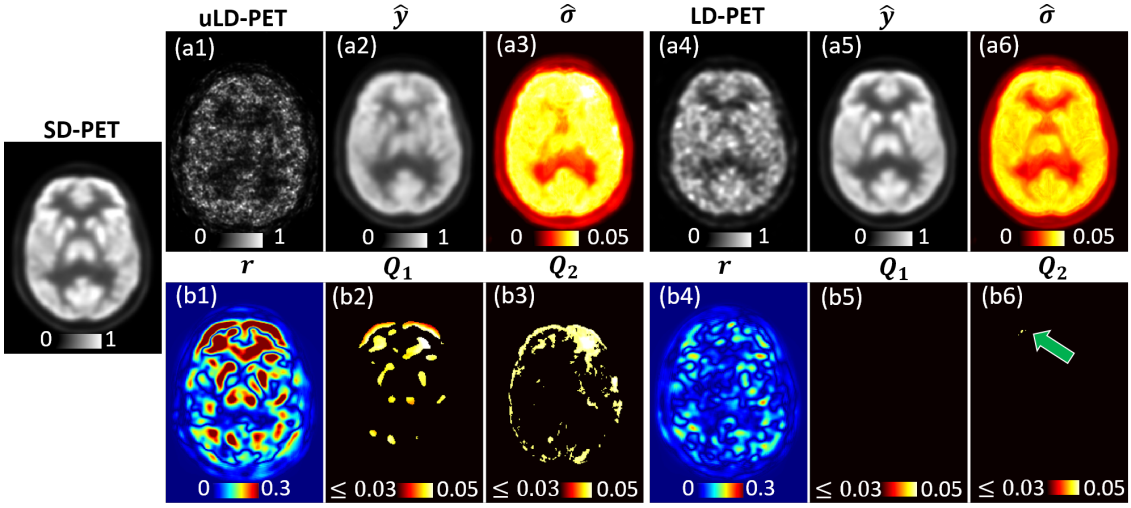

Figure 4.7: **Utility of Uncertainty Maps. Columns 1–3: uLD-PET as input and columns 4–6: LD-PET as input. (a1) and (a4)**: Input images uLD-PET and LD-PET. **(a2) and (a5)**: Predicted PET images $\widehat{y}$. **(a3) and (a6)**: Predicted per-voxel standard deviation image $\sqrt{\widehat{c}}$. **(b1) and (b4)**: Absolute residual images $r = |u^{\mathrm{SD}} - \hat{y}|$. Quantification maps **(b2), (b5)**: $Q_1(\widehat{\sigma}; r, \delta_{\mathrm{R}})$ and **(b3), (b6)**: $Q_2(\widehat{\sigma}; \delta_{\mathrm{U}})$.

The proposed suDNN provides additional information in the form of uncertainty maps as detailed in the following section.

4.3.6 Utility of Uncertainty Maps

We now analyze the uncertainty maps produced by the proposed suDNN with the inputs uLD-PET and LD-PET, and how to extract useful information from the same. For the input PET images uLD-PET and LD-PET (Figures 4.7 (a1) and (a4)), the network produces the predicted images (Figures 4.7 (a2) and (a5) respectively), along with the per-voxel variances $\widehat{c}$. For improved visualization, we show the standard deviation $\widehat{\sigma} = \sqrt{\widehat{c}}$ (per-voxel square-root) maps (Figures 4.7 (a3) and (a6)). We define two global thresholds, one for the predicted uncertainty image δ_{U} and another for the computed residual image δ_{R}, to identify pixels with high residuals and high uncertainty, respectively. That is, while pixel locations with absolute residual values $r \geq \delta_{\mathrm{R}}$ indicate sub-optimal reconstruction, pixel locations with $\widehat{\sigma} \geq \delta_{\mathrm{U}}$ indicate predictions with high uncertainty. Subsequently, we obtain two binary masks BM1 and BM2 by applying the threshold values δ_{R} and δ_{U} on r and $\widehat{\sigma}$, respectively. Through empirical analysis, we fix the values for the global thresholds (δ_{R}, δ_{U}) to be $(0.25, 0.03)$, respectively. Finally, to improve the utility of the uncertainty maps, we generate two quantification maps: (i) $Q_1(\widehat{\sigma}; r, \delta_{\mathrm{R}})$ (Figures 4.7 (b2) and (b5)), obtained by applying the binary mask, BM1, on $\widehat{\sigma}$, and (ii) $Q_2(\widehat{\sigma}; \delta_{\mathrm{U}})$ (Figures 4.7 (b2) and (b5)), obtained by applying the binary mask, BM2, on $\widehat{\sigma}$. As expected, the map Q_1 with the LD-PET input has substantially fewer non-zero values, compared to the map Q_1 obtained with uLD-PET as input. A similar trend is observed for the map Q_2. That is, as expected, prediction by the DNN with uLD-PET as input shows higher uncertainty compared to the prediction with LD-PET as input. Importantly, the high-intensity values in the map $Q1$ agree with the high-intensity values in the map $Q2$; this implies that regions with high residual values correspond to regions with high uncertainty in the predicted images. In this way, the map Q_2 (available at inference), might act as a proxy for the prediction error while inferring a PET reconstruction from test data.

4.4 Discussion

This work presents a novel sinogram- and uncertainty- aware DNN framework for the prediction of SD-PET images from ultra LD-PET images in simultaneous PET-MRI systems. Our imaging-physics-based modeling allows for accurate prediction of SD-PET images from LD-PET images with 180× dose reduction. Further, the proposed DNN

demonstrated improved robustness to newer OOD PET acquisitions (vLD-PET and uLD-PET images). The modeling of per-voxel heteroscedasticity within the loss-function of the DNN enabled uncertainty quantification in the predicted images.

By comparing the performance of our proposed DNN across state-of-the-art methods (Figures 4.2 and 4.3) we demonstrated the robustness to practical (OOD) degradations in the data. The DNN models relying on multimodal information (M4, M5, and suDNN) synthesize SD-PET images of comparable image quality with LD-PET as input. However, methods that rely on predicting the residual between the input LD-PET and the reference SD-PET images are *not* robust to OOD data (vLD-PET and uLD-PET as input). Although the multimodal GAN-based method (M5) shows substantial improvement over the unimodal (M3) and the multimodal (M4) residual-predicting U-net DNNs with vLD-PET as input, it underestimates the SD-PET contrast with uLD-PET.

The proposed suDNN framework is the first work, to the best of our knowledge, to include a PET physics-based loss function for the enhancement of uLD-PET images. This is analogous to the works in undersampled MRI reconstruction where the utility of including penalties in the k- space is known to improve the performance of the DNN(**?**). The ablation study (Figures 4.5 and 4.6) showed that modeling a transform-domain loss function like the encoder loss (Ablation-DNN3) or the physics-based loss function (Ablation-DNN4 and suDNN) provides robustness to OOD data. However, the sinogram-space loss (Ablation-DNN4 and suDNN) provided predicted images with improved accuracy in comparison to the manifold loss in Ablation-DNN3.

For all the DNNs considered in this work, the use of multi-contrast MRI images as multi-channel input provides a substantial improvement over unimodal PET-only inputs (Figure 4.4) and agrees with other works for this task (**??**). The qualitative comparison of predicted SD-PET images in Figure 4.6 suggests that DNNs with multimodal input leverages the anatomical information from the input MRI images. To understand the contribution of the multimodal inputs in comparison to the unimodal (PET-only) inputs, we analyze the feature maps obtained from initial layers (second) of the DNN, trained with unimodal and multimodal inputs, while maintaining the same network architecture. Figure 4.8 shows four feature maps (out of 64) from both the networks. The feature maps obtained using the multimodal inputs show anatomical features, either missing or distorted in the unimodal case. Additionally, we observed that networks that rely on predicting the residual can suffer from sub-optimal performance when the input image is degraded and *not* very similar to the images in the training set.

The proposed framework predicts SD-PET images with better accuracy and quantifies the uncertainty in the predictions that can aid clinical decision making. We demon-

strated the potential utility of the generated uncertainty maps (Figure 4.7) by defining global thresholds in terms of high residual and uncertainty values obtained in the experiments. Subsequent works should involve defining these threshold values in terms of clinically interpretable values, e.g., standardized uptake values (SUV).

The limitations of this study are as follows. First, the network uses a 2.5D-style input instead of training on actual 3D images. In the future, we plan to accommodate training using 3D images, which requires handling of a 3D system matrix, demanding high computational power. Secondly, although the quantitative improvements are shown using standard metrics such as PSNR and SSIM, for clinical validation, we plan to incorporate perceptual scores provided by radiologists as in (**??**). Thirdly, in the current form, the estimated SD-PET images are not representative of quantitative values such as SUV. The predicted images from our suDNN framework for the three inputs LD-PET, vLD-PET, and uLD-PET are not entirely different from each other in terms of structure, contrast and other features in the images. The estimated structure and regularity in the predicted SD-PET images even with uLD-PET as input can be attributed to the fact that the learned model relies heavily on the input MRI images for structural regularity than the input PET images (evident in Figure 4.8). Future work calls for the incorporation of improved handling of the LD-PET and SD-PET images to produce images that reflect the deviation in the number of counts in the input. For example, reporting relative SUV maps such as the change in SUV with respect to the SUV in a reference anatomical region.

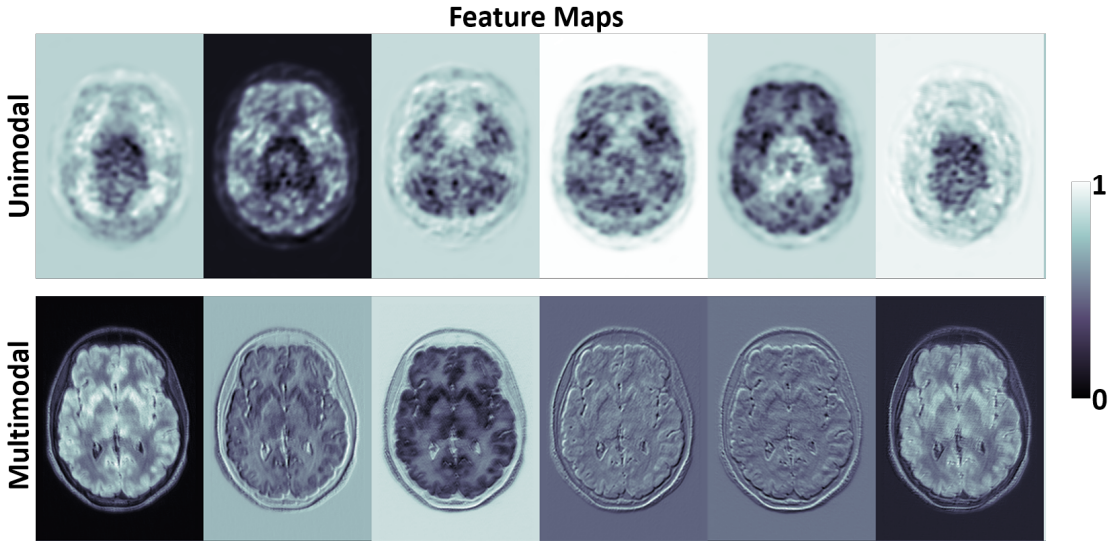

Figure 4.8: **Feature maps obtained from initial layers of the proposed network with unimodal (PET) and multimodal inputs (PET and multi-contrast MRI).** Six feature maps are shown out of 64 feature maps obtained at the output of the second layer of the DNN.

Finally, while the size of the dataset used in this work is larger than that of other works considered here, we plan to evaluate the proposed method on multiple cohorts, including data from individuals with healthy and pathological conditions.

In summary, using a DNN framework that is informed by the underlying imaging physics and that models uncertainty, a robust mapping from ultra-low-dose PET images to standard-dose PET images was achieved. The proposed DNN demonstrated robustness to unseen out-of-distribution PET acquisitions and provided an estimate of the underlying uncertainty of the prediction which facilitates a new paradigm of risk assessment in the application of DNNs to low dose PET image reconstruction. The method has the potential to dramatically improve the utility of ultra-low-dose PET imaging in diagnostic imaging, therapeutic monitoring, and in drug development research in oncology, neurology, and cardiology. Physics-inspired DNN-based reconstruction of ultra-low-dose PET scans has the potential to substantially expand the use of PET in longitudinal studies and imaging of radiation-sensitive populations including children and pregnant women.

Chapter 5

Incorporation of anatomical MRI knowledge for enhanced mapping of brain metabolism using functional PET

5.1 Introduction

The previous two chapters (Chapters 3 and 4) described novel image quality improvement methods developed in the context of static PET imaging. We now shift our focus to dynamic PET imaging for brain functional analysis. Chapter 2 briefly indicated the continuous-infusion-based fPET technique for dynamic brain mapping and the associated problem therein. Typically, functional analysis for PET and MRI involves standard filtering using Gaussian filtering followed by the application of ICA to identify brain activations related to resting-state or task-based activations. However, under extremely low SNR, Gaussian filtering or other edge-preserving denoising techniques can produce suboptimal denoised images consequently affecting the estimates provided by ICA. Hence, methods leveraging anatomical information from MRI are of interest.

This chapter describes an application of improved image quality of dynamic PET images in the context of fPET imaging. For this study, we hypothesized that accurate identification of brain metabolic activations could be obtained by filtering the fPET images using knowledge from the anatomical MRI image. We propose a modified version of the asymmetrical Bowsher prior that was proposed in **?** within a Bayesian image denoising context. The modified Bowsher prior is a weighted quadratic MRF prior with the PET image modeled as an MRF and the weights of the quadratic penalty obtained from the MRI image. The proposed modification corresponds to selecting the weights from the static MRI image in a model that uses patch-based similarity in the MRI image rather

than a similarity based on voxel-intensity alone. The proposed anatomical prior improves the identification of independent signal components from the fPET data by improving the PET SNR and reducing partial volume errors, as demonstrated in prior works (**???**). As a consequence of improved SNR in the fPET images, this work demonstrates improved identification of brain activation in response to external stimulation for both two-fold and three-fold reductions in the duration of the experiment duration to acquire the fPET data. The formulation of the prior model in this work differs from the one proposed in **??** in that the current model uses a smoothly-decaying function that in turn depends on the neighborhood voxel-locations in the static MRI image. Additionally, we incorporate patch-level differences (as opposed to voxel-level differences) to estimate the weights within the neighborhood of a voxel. The method is henceforth referred to as MRI-MRF prior and was validated using both simulated and in vivo visual task fPET datasets. The accuracy of the method was compared with conventional smoothing methods at both the subject and group level ICA, and the in vivo fPET dynamic data were downsampled to verify the robustness of the proposed method in response to reduced task stimulation durations.

5.2 Method

Let $\{u_o^t\}_{t=1}^T$ represent the uncorrupted (unobserved) dynamic sequence comprising T fPET images, each containing N voxels. Let $\{u^t\}_{t=1}^T$ represent the corresponding observed low-quality dynamic sequence comprising T fPET images. In this work, we consider the sequence $\{u^t\}_{t=1}^T$ as that output by the PET-MRI scanner using conventional model-based iterative methods such as maximum likelihood expectation-maximization (MLEM) (**?**). Typically, the PET detector PSF models cross-crystal effects during the gamma-ray co-incidence event, which is leveraged by the scanner during PET image reconstruction resulting in reduced partial volume errors. However, in comparison to MRI, the PET resolution is still much lower due to the fundamental limitation in the current PET detector technology. Therefore, in this work, we propose to denoise and improve the anatomical accuracy of the reconstructed PET images by relying on the MRI anatomical information modeled as a weighted MRF prior detailed below. The proposed method serves as a post-reconstruction image enhancement procedure that removes noise and reduces the partial volume errors by removing edge-artifacts.

Let v represent the fixed MRI image co-registered with the PET image sequence, i.e., each individual PET image u^t is spatially aligned to the fixed MRI image v. We model the PET images u^t as an MRF consisting of a neighborhood system $\mathcal{N} := \{\mathcal{N}_i\}_{i=1}^I$ where $\mathcal{N}_i$ denotes the neighbors of voxel i. In the MRF model, we consider square-shaped neigh-

borhoods with the square width corresponding/equivalent to L mm. We model the prior distribution on u_o^t given the fixed MRI image v, as an MRF $P(u_o^t|v)$. Because our goal is to perform image enhancement of the PET image, we consider a Gaussian likelihood model that penalizes the squared difference of the intensities in the uncorrupted (unobserved) image u_o^t and the corrupted observed image u^t. Thus, given the sequence of acquired and reconstructed fPET images, $\{u^t\}_{t=1}^T$, and the fixed MRI image v of the same subject, we formulate the restoration of the fPET image sequence $\{u_o^t\}_{t=1}^T$ as maximum-a-posteriori (MAP) estimation:

$$\arg\max_{u_o^t \geq 0} P(u_o^t|u^t, v) = \arg\max_{u_o^t \geq 0} P(u^t|u_o^t)P(u_o^t|v) \tag{5.1}$$

Thus, the negative-log posterior with the mean-squared error data fidelity term and the MRF prior becomes

$$\widehat{u_0^t} = \arg\min_{u_o^t \geq 0} ||u^t - u_0^t||_2^2 + \alpha R(u_0^t|v) \tag{5.2}$$

Here $R(\cdot)$ acts the MRI-guided MRF (MRI-MRF) regularization term, which incorporates the anatomical information from MRI image, v. $R(\cdot)$ is the negative logarithm of the probability distribution which is the Gibbs energy associated with the MRF in terms of the weighted squared difference between the voxel intensities within the neighborhood, where the weights are obtained from the static MRI image v.The parameter α determines the strength of the regularization $R(\cdot)$. The formulation in Equation 5.2 is generic and allows the incorporation of arbitrary prior models that enforce a certain type of regularity, e.g., piecewise smoothness, on the fPET images. In this work, we model $R(\cdot)$ as a modified version of the asymmetrical Bowsher prior presented by **?**. Specifically, $R(\cdot)$ is modelled as a weighted quadratic MRF function defined as, $R(u|v) = \sum_{i \in I} sum_{j \in I} w_{ij}(u_i - u_j)^2$. Here the weights w_{ij} are computed based on the intensity values from the co-registered MRI image, v, as

$$w_{ij} = \frac{\exp\left(-(||\mathcal{N}_i(v) - \mathcal{N}_j(v)||_1)/(2\sigma_w^2)\right)}{\left(\sum_j \exp\left(-(||\mathcal{N}_i(v) - \mathcal{N}_j(v)||_1)/(2\sigma_w^2)\right)\right)} \tag{5.3}$$

where the operator $\mathcal{N}_i(\cdot)$ extracts a vectorized isotropic 3D patch of volume L^3 mm^3 centered around voxel i, and the parameter σ_w determines the spatial pattern of weights within the patch in the neighborhood of voxel i. The parameter σ_w is indirectly dependent on the location as the neighborhood varies across different voxel locations i in the MRI image. The strategy of determining the weights w_{ij} in the MRF-based regularization term by relying on patch-difference norms have been used within the literature on patch-based denoising methods, first proposed on natural images in the works of **??** and on MRI images in the works of **???**. While a high value of σ_w leads to weights that are similar

for all the neighboring voxels, a low value of σ_w assigns higher weights to a few selected voxels in the neighborhood. The latter scenario leads to an extension of the strategy in the asymmetric Bowsher prior (**?**) that (i) enforces neighborhood weights to be binary (1 or 0) and (ii) designs weights only based on voxel-intensity differences (instead of patch differences). Our proposed strategy of using patch-based differences can provide additional robustness to noise, artifacts, and minor misalignments between the MRI and PET images while leading to better structure preservation, in ways that are similar to those studied in general image denoising (**?**). While iterative denoising algorithms, as in **??**, offer algorithms for data-driven tuning of the parameter σ_w to improve performance, in the application in this manuscript, where the weights only need to be precomputed once (for a particular subject, due to the MRI image v being static), we tune the parameter σ_w based on validation on simulated data (Section 5.3.1). The Bayesian optimization problem with the MRI-MRF prior in Equation 5.2 was solved using the limited memory BFGS method (L-BFGS) (**?**), with positivity constraints.

To perform spatial ICA, we construct a spatiotemporal data matrix, Y, using $\{u^t\}_{t=1}^T$, such that the dimension of Y is $T \times N$. ICA models Y as a linear combination of the underlying independent components: $Y = AS$, where S contains the independent components, and A is the mixing matrix. In the context of PET imaging, the measured PET data is affected by the blurring matrix, H (**??**), and the ICA model becomes

$$Y_0 H = AS_0 H \tag{5.4}$$

where Y_0 represents the spatiotemporal matrix constructed from the true PET signals, and S_0 models the true underlying independent components of Y_0. The matrix, H, acting on the spatial dimension, models the partial volume errors in PET measurements, and hence, the resultant independent components though the mixing operation, A.

The goal of fPET data analysis is to identify S_0 from Equation 5.4. Image denoising in the spatial domain is an important pre-processing step prior to application of the ICA algorithm. The characteristics of an ideal filter for estimation of the source components, S_0, would be to recover the signal without compromising the independence of the true underlying components. Typically, a Gaussian smoothing filter with a suitable width, specified by its full width at half maximum (FWHM) is used to reduce spatial noise, for example, during fMRI data analysis. However, performing a Gaussian smoothing can introduce additional bias in fPET images and the corresponding independent components due to worsening of the partial volume errors (**??**). Hence, this work proposes an MRI guided filtering scheme that can (i) perform denoising, as well as (ii) reduce partial volume errors, to provide an improved estimation of the underlying source components, S_0.

5.3 Data and Experiments

Both simulated and in vivo fPET and MRI data were used to validate the performance of the MRI-MRF prior. For comparison, the MRI-MRF prior processed fPET images were compared with those obtained using Gaussian smoothing with varying kernel sizes (specified by FWHM).

5.3.1 Simulated Data and Experiments

Continuous infusion of FDG PET activity was simulated for 60 minutes using a two-tissue compartment model involving the three kinetic parameters K_1, K_2 and K_3 and a fitted arterial input function from intravenous blood samples collected from our previous in vivo experimental data (**?**) to yield a total of 60 frames. PET image was simulated based on the T1w MNI template MRI image with an isotropic voxel resolution of 8 mm^3. The simulated FDG activity was corrected by the blood partition fraction and haematocrit using the same procedure as in our previous work (**?**). Brain regions were segmented into grey matter, white matter and the occipital cortices using the MNI structural atlas (**?**) using FSL (**?**). The MRI and PET images were simulated with an isotropic spatial resolution of 2 mm in the MNI space. The regional specific metabolic kinetic parameters used for the simulated dataset were $K_1 = 0.101$, $K_2 = 0.071$, $K_3 = 0.042$ for grey matter and $K_1 = 0.047$, $K_2 = 0.070$, $K_3 = 0.035$ for white matter, respectively (**?**). A visual task stimulus was simulated between 20 to 30 minutes in the visual cortex region similar to the *in vivo* experimental paradigm. During the visual stimulation period, the parameter K_3 in the occipital cortex was simulated to have a 20% increment.

The tomographic iterative GPU-based reconstruction toolbox (TIGRE) was used for PET image reconstruction (**?**). The PET images were forward projected, and Poisson noise was applied in the measurement space, to generate a high-dose dataset. Subsequently, we simulated dynamic low-dose PET data using the Poisson thinning approach (**?**) such that the low-dose data had a dose reduction factor (DRF) of 100 compared to that of the high-dose data. The generated PET images were smoothed using a 3D Gaussian filter with FWHM 2.35 mm to simulate the partial volume effect produced by the system. Finally, MLEM algorithm was used to reconstruct the PET images to yield 60 frames for both low and high dose datasets.

The ICA-specific pre-processing steps including spatial normalization and dimensionality reduction were performed as described in detail by **?** on the post-reconstruction smoothed images. In this work, we performed both subject-level and group-level ICA on fPET data. For group analysis, the spatiotemporal matrix from each subject was concate-

nated along the temporal dimension before the application of ICA (**?**). The pre-processed data, which was an estimate of Y_0, was then decomposed using an ICA unmixing algorithm in the FastICA toolbox (**??**).

5.3.2 *In Vivo* Data and Experiments

A cohort of five healthy volunteers were scanned for a visual task stimulus experiment using a 3T Siemens Biograph mMR (Siemens Healthineers, Erlangen, Germany) PET-MRI scanner, approved by the institute human ethics committee. The overall stimulation protocol consisted of three visual stimulation periods consisting of alternating periods of rest and task blocks. A detailed description of the experiment is provided in our earlier work in (**?**). The subjects rested for a period of 20 minutes to allow sufficient FDG accumulation in the brain, during which structural T1 MPRAGE MRI scans were acquired. Following this, the subjects viewed a circular flickering checkerboard stimulus for 10 minutes. The checkerboard was retained for a period of 120 seconds and subsequently, an intermittent 32 seconds on and 16 seconds off design was employed. Following the first task stimulation, which involved 3 blocks: rest, task, and rest, two other stimulation experiments, using the full checkerboard visual, were carried out. We used the PET data acquired during the first full checkerboard. Hence, the PET data for each subject was of 30-minute duration, including 10 minutes resting before the stimulation, 10 minutes of a full checkboard stimulation followed by another 10 minutes of rest (Figure 5.4(a)). The average dose of FDG given to each subject was 95.9 ± 5.9 MBq which was infused at a constant rate of 36 mL/hr over the 90-minute duration.

We reconstructed PET images from the list-mode data using two different values for the temporal bin-width (*Tbin*) of (i) 30 seconds for the low-dose PET images, and (ii) 3 minutes for high-dose PET images. The average dose for the corresponding low dose fPET images across the group of subjects was calculated to be 7.5 kBq/kg/frame. The PET data was corrected for attenuation using a pseudo-computed tomography (pCT) map (**??**). The corrected PET data sinograms were reconstructed using ordered subsets expectation maximization (OSEM) algorithm with 3 iterations and 21 subsets along with PSF modelling. The PET 3D volumes were reconstructed with voxel sizes of $3 \times 3 \times 2.03$ mm^3. For standard analysis, all the images were registered to the MNI-152 template. The high-dose PET images from the 3-minute binned data were used to register and resample the low-dose PET images with the T1w MRI (acquired at 1 mm^3 isotropic resolution) for each subject using ANTS (**?**).

We also undertook a comparison of the performance of the MRI-MRF and Gaussian filtering schemes when the duration of the task and resting blocks was reduced. This

analysis was carried out by downsampling the total number of low-dose fPET images reconstructed from the list mode data. Functional PET analyses were computed at both the subject-level and group-level for downsampling factors (DF) of 2 and 3 to simulate fPET images of duration 30 seconds but acquired at 1 minute and 1 minute 30 seconds (90 seconds) intervals, respectively. The downsampled PET images correspond to reduced task duration with a lower number of temporal frames.

5.3.3 Optimal Patch-Width Selection

The optimal kernel sizes for the Gaussian low pass filter and the optimal patch-width for the MRI-MRF prior, for processing the in vivo data were selected and validated using simulated data. We optimized the parameters to achieve high sensitivity without substantial loss of specificity using ICA computed activation maps. For computing the sensitivity and specificity values, the region of interest (ROI), occipital cortex, was obtained using the segmentation procedure as described in Section 5.3.1. The sensitivity and specificity performance metrics are defined as follows

$$\text{Sensitivity}[\%] = \frac{\text{number of activated voxels inside the ROI}}{\text{total number of voxels inside the ROI}} \times 100 \tag{5.5}$$

$$\text{Specificity}[\%] = \frac{\text{number of nonactivated voxels outside the ROI}}{\text{total number of voxels outside the ROI}} \times 100. \tag{5.6}$$

Both sensitivity and specificity metrics provide a quantitative assessment of the activation maps (z-score map) obtained from the different filtering operations. The metrics were computed by considering a voxel as activated if $|z| \geq 1.6$ and $|z| \geq 2.1$, for subject-level and group-level analysis respectively **?**. The parameter search-space for the MRI-MRF prior, includes varying values of the regularization parameter, patch length (α, L respectively). On the other hand, for the Gaussian kernel, we varied the FWHM parameter which in turn determines the kernel size. The parameter σ_w was determined empirically in this work. We found via simulation studies (Section 5.4.1) that the sensitivity and specificity values do not change dramatically for minor perturbations in the parameter σ_w. The effect of varying the σ_w around the empirically tuned value is demonstrated in the following section.

Table 5.1: **Comparison of sensitivity and specificity of MRI-MRF and Gaussian smoothing filters.** The sensitivity and specificity values at two different z-score threshold values are provided.

MRI-MRF				
L (mm)	Sensitivity		Specificity	
	$\|z\| \geq 1.6$	$\|z\| \geq 2.1$	$\|z\| \geq 1.6$	$\|z\| \geq 2.1$
10	78.7	61.6	91.3	97.2
14	87.7	79.9	92.3	94.8
18	83.3	73.9	89.8	94.4

Gaussian				
FWHM (mm)	Sensitivity		Specificity	
	$\|z\| \geq 1.6$	$\|z\| \geq 2.1$	$\|z\| \geq 1.6$	$\|z\| \geq 2.1$
11	77.5	60.2	89.6	94.5
13	75.5	61.1	89.6	95.4
15	64.2	49.1	93.3	94.7

Table 5.2: **Effect of varying the prior-weight parameter α for MRI-MRF prior.** Effect of varying α for a fixed patch-width L on sensitivity and specificity values for the visual-task related component obtained from the ICA (z-score threshold = 1.6)

MRI-MRF (L = 10 mm)					
α (x 1e^{-3})	5	10	15	20	25
Sensitivity	58.4	70.6	75.1	78.7	74.3
Specificity	93.6	95.4	91.3	91.4	92.7

MRI-MRF (L = 14 mm)					
α (x 1e^{-3})	5	10	15	20	25
Sensitivity	82.1	82.8	87.7	80.2	60.9
Specificity	92.2	90.4	92.5	90.5	91.4

5.4 Experiments and Results

5.4.1 Results for Simulated Data

Table 5.1 compares the sensitivity and specificity for both denoising schemes at different parameter configurations. For the MRI-MRF prior, the patch-length was varied from 10 mm to 18 mm which represented a varying patch size of 5 to 9 voxels in each direction, respectively. In the case of Gaussian filtering, the kernel size was determined by the full width at half maximum of the Gaussian function. The FWHM for Gaussian varied from 11 mm to 15 mm. It is to be noted that while parameter L (for the MRI-MRF prior) represents the width of the entire patch, FWHM (for the Gaussian filter) represents approximately half of the kernel-width. The parameter range chosen for the Gaussian smoothing is consistent with the Gaussian kernel widths used in the prior work (**?**). The sensitivity values for the MRI-MRF processed image are dramatically higher than that of the Gaussian smoothed images, whereas the specificity values are comparable between the two methods. For the fPET data analysis, a patch-length of 14 mm was chosen for the MRI-MRF prior. However, both the Gaussian kernels with FWHM 11 mm and 13 mm show similar sensitivity and specificity values. Therefore, the analysis using the Gaussian-filtered in vivo fPET data was undertaken using both the 11 mm and 13 mm FWHM filters.

Table 5.2 shows the effect of varying the free parameter α for a given patch-width L on the sensitivity and specificity values for the visual-task component obtained from the ICA. The sensitivity and specificity values for varying prior-weight α is shown for two different patch-widths $L = 10$ mm and 14 mm. The sensitivity and specificity values in Table 5.1 correspond to a z-score threshold value of 1.6 (i.e., $|z| \geq 1.6$).

Figure 5.1 shows the visual task specific activation for the reference noiseless fPET images, and for the three denoising schemes using the optimal parameters chosen from Table 5.1. The ICA activation map obtained using the noiseless images serves as the reference map (Figure 5.1 (a)). The activation map obtained from post-reconstruction filtered fPET images using the MRI-MRF prior (Figure 5.1 (b)) was closest to the reference activation map in the visual cortex. On the other hand, the activation maps obtained using

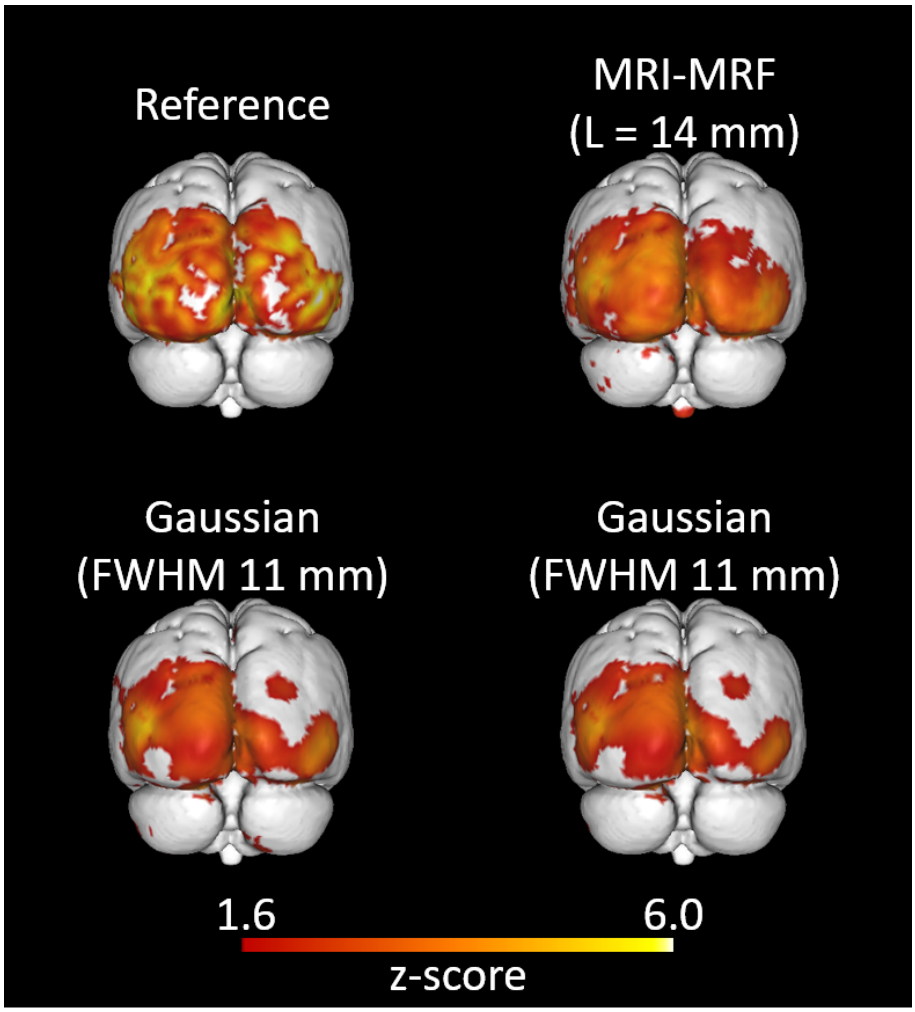

Figure 5.1: **Comparison of brain activation maps for simulated data.** Visualization of activation in the visual cortex using ICA on noiseless fPET images (a), MRI-MRF prior (b), Gaussian smoothing with FWHM 11 mm (c), and FWHM 13 mm (d).

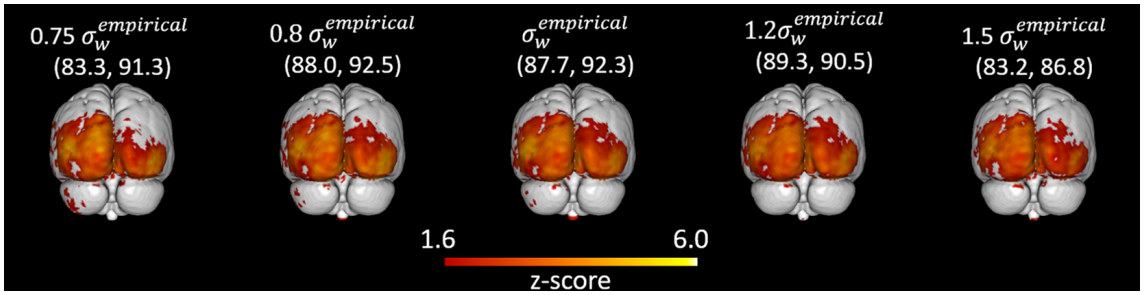

Figure 5.2: **Effect of varying the weights-parameter σ_w for MRI-MRF prior.** Analysis of the effect of varying σ_w around the empirically chosen value $\sigma_w^{\text{empirical}}$ on the activation maps for the MRI-MRF method. The values in the parenthesis represent the sensitivity and specificity values obtained for the component corresponding to the visual task from the ICA respectively.

Gaussian smoothing with both FWHMs yield suboptimal activation maps in the visual cortex with asymmetrical patterns.

Figure 5.2 shows the effect of varying the parameter σ_w in the MRI-MRF formulation. As mentioned earlier, the parameter σ_w aids in distribution of the weights for the MRF prior. Hence, a low value of σ_w corresponds to assigning higher weightage to only a selected few neighbour of the central voxel under consideration. On the other hand, a large value of σ_w corresponds to uniform distribution of the weights in the neighbourhood. In Figure 5.2, sensitivity values corresponding to the σ_w value used are mentioned. We see that at low σ_w, several false positive activations are seen with low sensitivity and for a high value of σ_w sensitivity is slightly reduced than the optimal value. Let the empirically determined value of σ_w to yield maximum sensitivity, be $\sigma_w^{\text{empirical}}$. From Figure 5.2, we see that for immediate values around the empirically determined σ_w, we observe only a slight variation in the sensitivity and specificity values. That is, the performance of the ICA algorithm is more sensitive to the value of α compared to σ_w.

5.4.2 Results for In Vivo Data

The results reported in this section are for the fPET images reconstructed using the list-mode data binned at T_{bin} = 30 s, and for DF = 1, 2 and 3.

Figure 5.3 shows the post-reconstruction filtered fPET images along with the subject's MRI image (Figure 5.3, left column) and the corresponding vendor provided low dose fPET image (Figure 5.3, second column). The denoised image using the MRI-MRF prior (Figure 5.3, third column) shows superior recovery of PET signals in different regions of the brain while removing substantial amount of noise. Specifically, the white and grey matter boundaries are well delineated, the shape of the ventricles has been re-

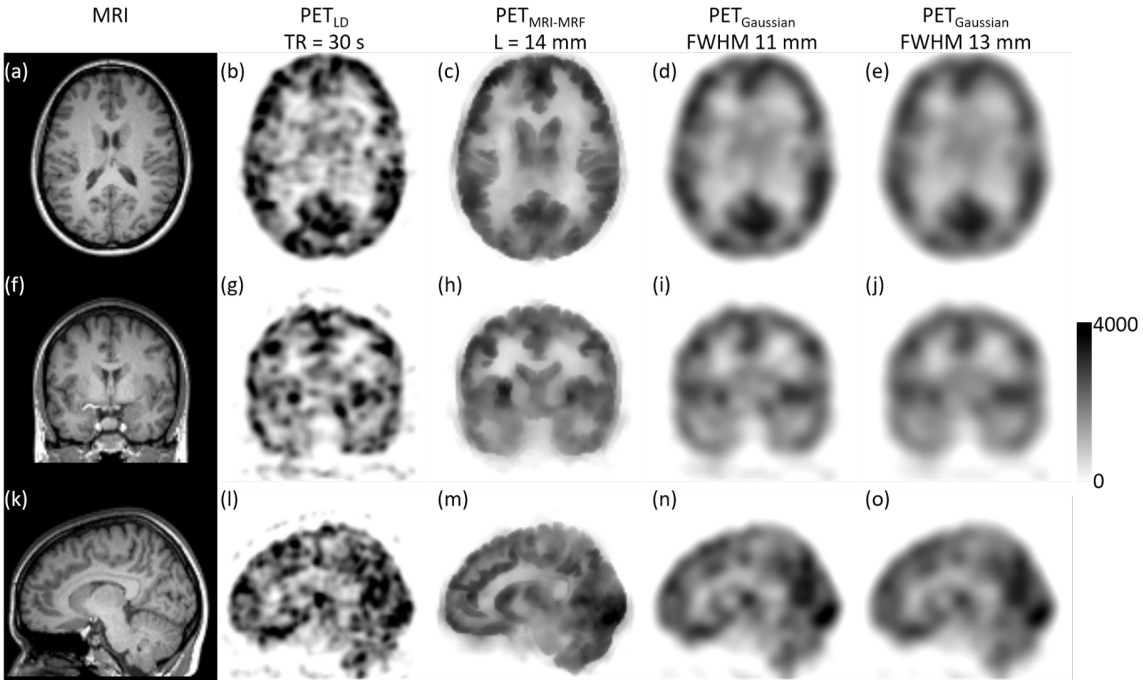

Figure 5.3: **Comparison of denoised *in vivo* images using different methods.** Assessment of the post-reconstruction smoothed fPET images with binning time of T_{bin} = 30 s. The subject's MRI image (a); the vendor reconstructed PET image (b); the filtered image using the MRI-MRF prior (L = 14 mm) (c); and using Gaussian kernels with FWHM (d) 11 mm, (e) 13 mm.

covered (which is not evident in the low dose PET image), and anatomical features in the gyri, sulci and details of the cortical folding (refer Figure 5.3m) have been restored. On the other hand, the denoised images using both Gaussian kernels (FWHM 11 mm and 13 mm) are heavily blurred and show substantial loss of anatomical details due to the partial volume errors (Figure 5.3, fourth and fifth columns).

Figure 5.4 shows the results of an individual subject-level fPET analysis obtained using different filtering techniques for a downsampling factor of one (i.e. DF=1, includes all list-mode data). The ICA activation maps corresponding to the visual task component along with the timecourses are calculated for each method. The timecourses are derived from the normalized (z-scored) spatial components following the methods in (**?**). The ICA timecourse for both Gaussian kernels (Figures 5.4d and 5.4f) are noisy and do not closely follow the experimental task paradigm. Moreover, the shape of the region of brain activation does not follow the known anatomical structure of the primary visual cortex but extends into adjacent neuroanatomical structures including the white matter, likely due to large partial volume errors. Conversely, the activation map obtained using the MRI-MRF prior (Figure 5.4c) shows localized activity near the visual cortex with a

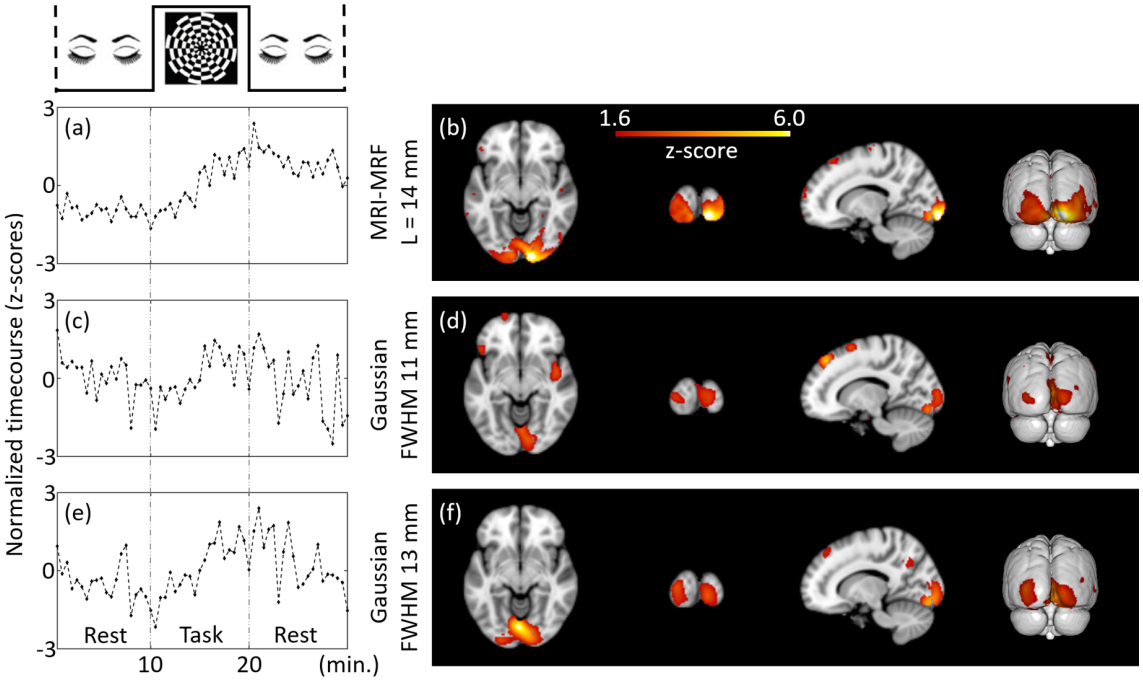

Figure 5.4: **Comparison of brain activation maps and timecourses for subject-level analysis with all the fPET frames:** T_{bin} = 30 s and DF = 1 at MNI coordinate (14, -94, -8). The independent components estimated from the filtered fPET images using different schemes are provided. The task paradigm is shown in (a). ICA maps and timecourses: top to bottom: MRI-MRF prior with L = 14 mm (b) and (c), Gaussian smoothing with FWHM = 11 mm (d) and (e), Gaussian smoothing with FWHM 13 mm (f) and (g).

significantly higher z-score within the visual cortex compared to both Gaussian kernels. The ICA timecourse for the MRI-MRF prior (Figure 5.4b) accords more closely with the experimental design with increased uptake during the visual task block. The comparison of the visual task components for the three methods for all brain sections is consistent with these observations.

Figure 5.5 compares the visual task components for MRI-MRF prior and the Gaussian smoothing with FWHM 13 mm across multiple axial slices showing brain activation. At DF = 1 (all frames included), for most of the axial slices, the MRI-MRF prior shows activation which is aligned with the visual cortex anatomy. In slices where there is increased activation shown by the MRI-MRF prior, results from Gaussian smoothing shows a large blob of activation around the visual cortex but extends to other anatomical regions.

The results for the group-level fPET analyses for the three filtering techniques using the complete list-mode dataset (DF = 1) are shown in Figure 5.6. A higher z-score range was observed for the group-level analyses compared to the single subject-level analysis. However, in contrast to the subject-level analysis, the timecourses estimated from all

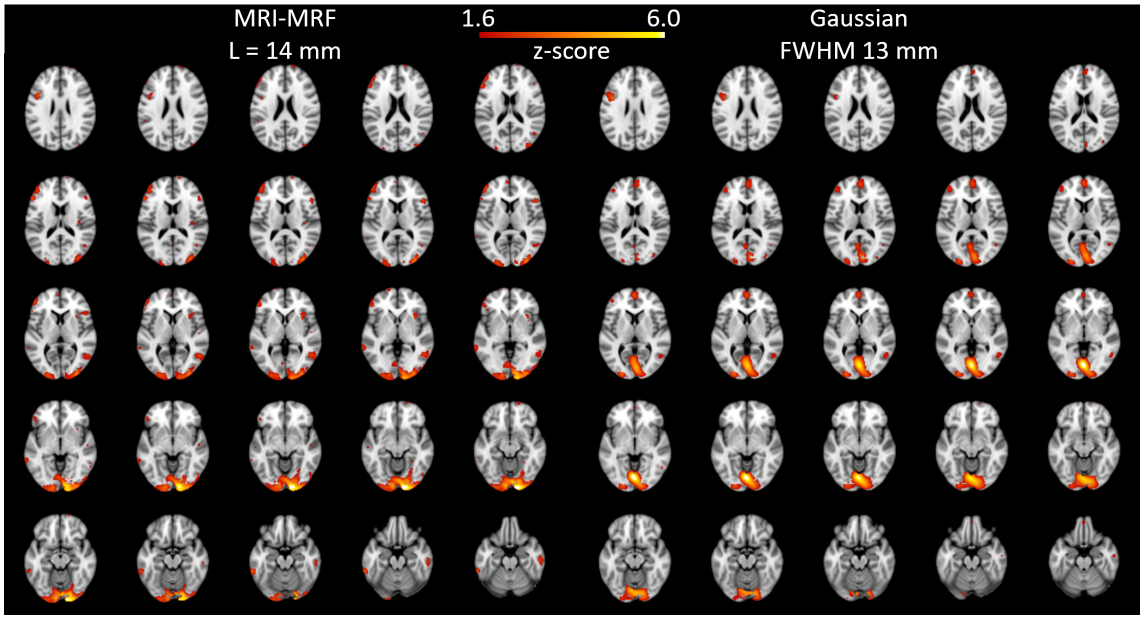

Figure 5.5: **Comparison of brain activation maps across all slices for subject-level analysis with all the fPET frames:** T_{bin} = 30 s and DF = 1 for multiple slices. Visual task component across slices for MRI-MRF (left) and Gaussian filter (right) with FWHM 13 mm.

methods (Figures 5.6b, 5.6d and 5.6f) at the group level recapitulated the experimental design paradigm. The activation map corresponding to the MRI-MRF prior followed the known neuroanatomical representation of the primary visual cortex and was consistent with the subject-level result. On the other hand, the activation maps using the two Gaussian kernels did not represent activation in the primary visual cortex and demonstrated diffuse cerebral metabolic activity into large adjacent anatomical regions including the white matter.

Figure 5.7 shows the subject-level fPET analyses for downsampling factors of two and three (DF = 2 and 3). Timepoints T1 and T2 represent the onset of the task and second resting block in the downsampled task paradigm. Plausible ICA activation maps were not generated using an 11mm FWHM Gaussian kernel for both DFs and therefore no results are included. The ICA timecourses during the task-block for the MRI-MRF filter demonstrated a steadier gradual increase, in agreement with the task paradigm, in comparison to the 13mm FWHM Gaussian kernel for DFs of 2 and 3 respectively (Figures 5.7a and 5.7e compared to Figures 5.7c and 5.7g respectively). The activation map axial view for DF = 3 did not reveal activation in the left hemisphere as was expected for the visual task (Figure 5.7h). However, for DF = 2 there was some activation in the left hemisphere visual cortex (Figure 5.7d) although it was not as widespread as for the fully sampled

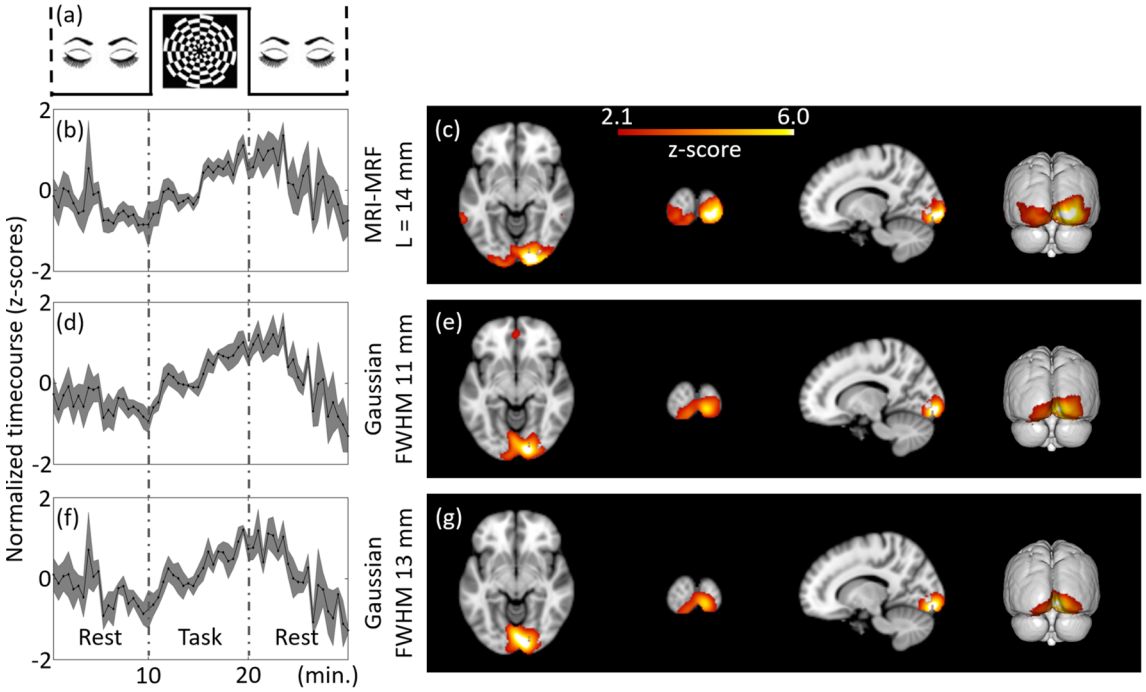

Figure 5.6: **Group-level estimation of brain activations using ICA:** T_{bin} = 30 s and DF = 1 at MNI coordinate (14, -94, -8). The independent components estimated from the filtered fPET images using different schemes are provided. The task paradigm followed for the group study is shown in (a). ICA maps and timecourses: top to bottom: MRI-MRF prior with L = 14 mm (b) and (c), Gaussian smoothing with FWHM = 11 mm (d) and (e), Gaussian smoothing with FWHM 13 mm (f and (g)).

dataset. On the other hand, the activation maps for the MRI-MRF prior (Figures 5.7b and 5.7f) showed spatial congruency across the three DFs, whilst the discrepancy between the z-scores for the MRI-MRF prior and the 13 mm FWHM Gaussian filter was largest for DF =3 compared to DF = 2 and 1.

Figure 5.8 shows the group-level fPET analyses at DF = 2 and DF = 3. In contrast to the group-level analysis for the fully sampled dataset where there was little difference between the activation maps estimated by the MRI-MRF method and the Gaussian kernel with FWHM 13 mm (Figure 5.6), the activation maps estimated for the group-level analyses for DF= 2 and DF = 3 showed marked differences. For both DF = 2 and 3, the ICA timecourses for the MRI-MRF prior (Figures 5.8a and 5.8e) showed agreement with the task experimental design with higher z-scores than for the 13mm FWHM Gaussian filter timecourses (Figures 5.8c and 5.8g). The activation maps show that while the MRI-MRF prior was able to resolve brain activation that was consistent with activation of the visual cortex (Figures 5.8b and 5.8f), at both DF = 2 and 3 the 13mm FWHM Gaussian filter was unable to resolve extended activation throughout the primary visual cortex (Figures 5.8d

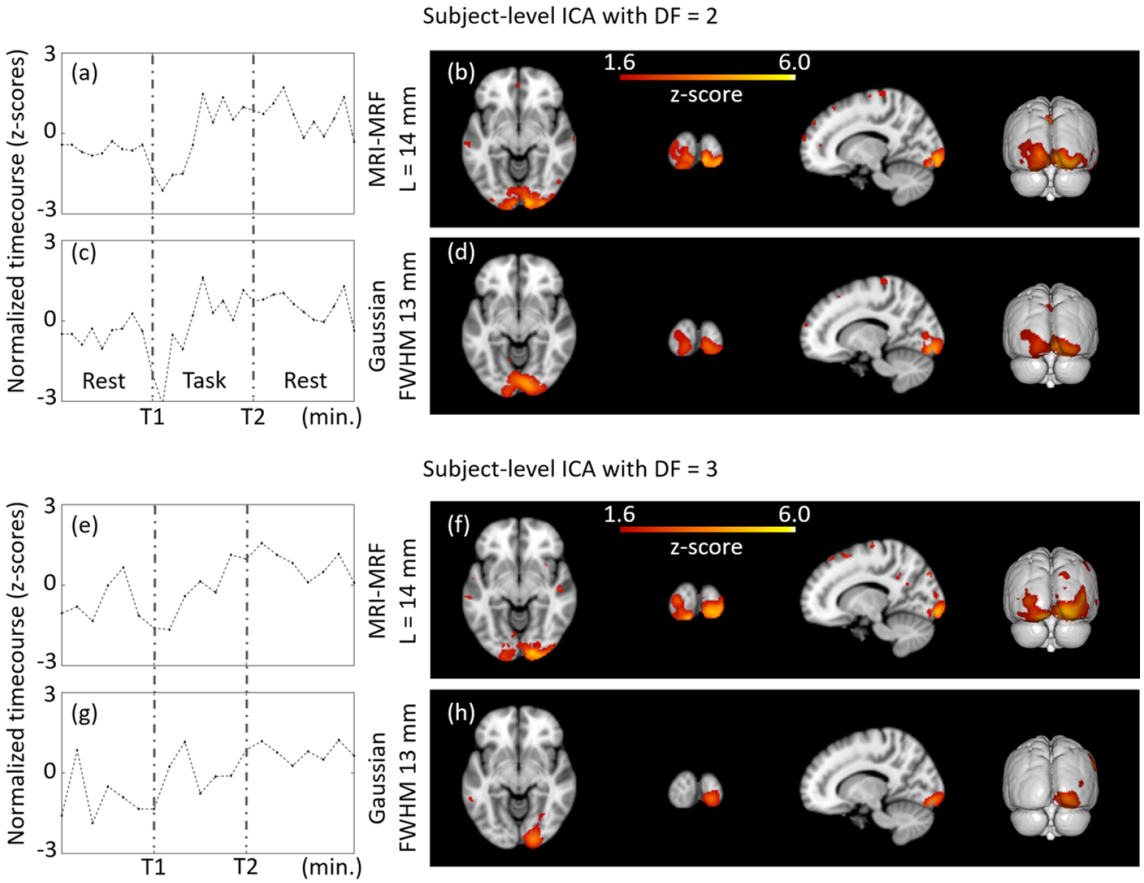

Figure 5.7: **Subject-level (representative) estimation of brain activations using the reduced task and resting blocks with DF = 2 and 3:** MNI coordinates (14, -94, -8). The independent components estimated from the filtered fPET images using different schemes are provided. DF = 2: MRI-MRF prior (b) and Gaussian kernel with FWHM 13 mm (d). DF = 3: MRI-MRF prior (f), and Gaussian smoothing with FWHM = 13 mm (h). The T1 and T2 represent the onsets of the task and second resting block, respectively.

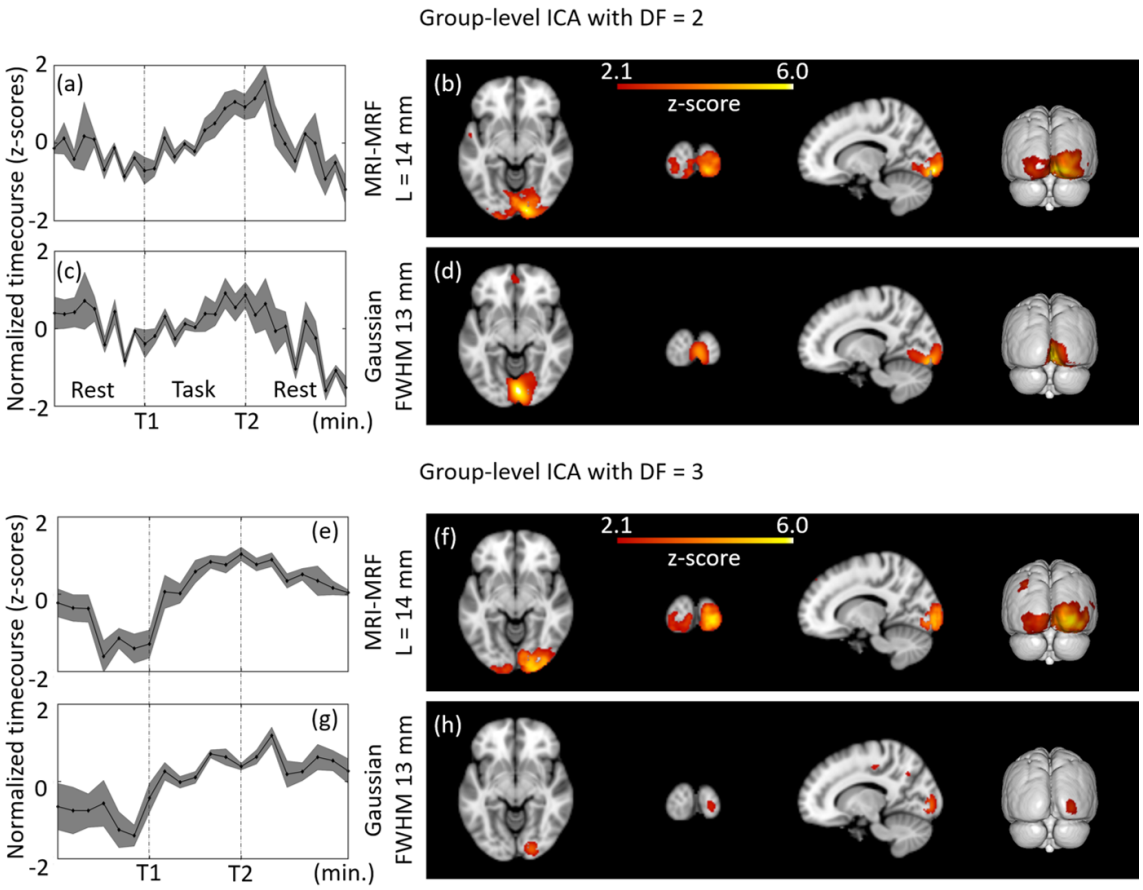

Figure 5.8: **Group-level estimation of brain activations using the reduced task and resting blocks with DF = 2 and 3:** MNI coordinates (14, -94, -8). The ICA components estimated from the filtered fPET images using different schemes are provided. DF = 2: MRI-MRF prior (b) and Gaussian kernel with FWHM 13 mm (d). DF = 3: MRI-MRF prior (f), and Gaussian smoothing with FWHM = 13 mm (h).

and 5.8h) with no activation in the left hemisphere for DF = 3 (Figure 5.8h). Conversely, the activation maps for the MRI-MRF prior were congruent across the subject-level and group-level analyses although greater consistency in the right hemisphere.

5.5 Discussion

We have proposed an MRI-assisted fPET processing framework for the analysis of task-related metabolic changes in the brain using high temporal resolution fPET data and for low-dose fPET brain mapping applications. We investigated the effect of using the anatomical information from a subject's MRI to denoise the fPET dataset to reduce the partial volume error in the PET images to increase the sensitivity of the ICA analysis. The PET image restoration problem was posed as a solution to a Bayesian optimization problem that was solved using L-BFGS due to its greater computational efficiency than gradient-descent-based optimization techniques.

This study compared the post-reconstruction filtered images from the MRI-MRF method and Gaussian smoothing with varying kernel sizes as well as the ICA activation maps from the fPET dataset using a visual stimulation task. Visual assessment of the post-reconstruction smoothed images showed that the MRI-MRF processed PET images recovered many features which were not readily observed in the conventional low dose PET images. The MRI-MRF filtered PET images revealed localized tracer uptake in the sub-cortical nuclei adjacent to the lateral ventricles (e.g., Figure 5.3c), whereas little or no uptake was apparent in the comparable low-dose and Gaussian denoised PET images. Furthermore, the level of Gaussian smoothing required to obtain plausible activations in the visual cortex rendered the fPET image hard to interpret visually as there was a substantial loss of features. The MRI-MRF method provides a balance between the visual interpretability of the fPET images together with improved resolution and sensitivity for functional analysis using ICA.

The task-based experimental design paradigm enabled meaningful comparison of the ICA timecourses obtained using the two filtering techniques, by inspection of the FDG uptake in the visual cortex during the visual stimulation task. The proposed methodology was able to achieve consistent activation maps at both the subject-level and group-level for DF = 1, 2, and 3. However, this was not true for the Gaussian smoothing kernels. Moreover, since the fPET data was acquired for an exogenous task-based stimulus, a good correlation between the subject-level and group-level activation maps was expected. In particular, the improved results for the individual subject-level analysis demonstrates

the benefit of the MRI-MRF method to enhance single subject-level analysis using low dose high temporal resolution fPET data with reduced task durations.

The study involving downsampling factors that demonstrated that the proposed processing pipeline could detect dynamic brain metabolic activity for visual task stimulation for approximately three minutes. However, this interpretation does assume that the FDG dosage per frame in the fPET images is consistent for the different downsampling factors. In practice, this would be achievable experimentally by altering the infusion protocol or slightly increasing the dosage of the radiotracer (**?**). The Gaussian smoothing technique failed to identify task-related ICA components for the shorter task durations (i.e., at higher DFs) due to reduced sensitivity.

Unlike fMRI, fPET images suffer from very low SNR, and hence the spatial denoising scheme must be carefully chosen to provide an optimal bias-variance trade-off. MRI-guided PET image denoising and deblurring have been extensively reported in the literature (**??**) with many solutions for post-reconstruction PET image enhancement. However, this work is the first to demonstrate the effectiveness of the MRI-based spatial denoising technique for dynamic fPET imaging, such that fPET images are both visually interpretable and produce accurate functional maps with improved temporal resolution. The high specificity and sensitivity of the algorithm also enabled single subject-level analyses along with reasonable visualization of the fPET images without loss of anatomical details. Traditional methods such as Gaussian smoothing perform averaging without consideration of the anatomical boundaries and hence the quantitative accuracy of FDG signals is degraded by partial volume errors. This was reflected in the diffuse visual activation maps obtained with the Gaussian filtering. Using edge-preserving denoising techniques such as anisotropic filtering would also yield suboptimal performance because of the poor SNR of the fPET images and the difficulty of distinguishing between tissue boundaries and noise.

The formulation of the MRI-MRF prior in this work is generic and allows for modeling of higher-level image features such as dictionary atoms. Nevertheless, the proposed filtering framework is efficient and computationally less expensive in comparison to other patch-based techniques, and hence, the framework is easier to adapt to other research and clinical applications.

Although the MRI-MRF prior in this work was applied in the image domain on post-reconstructed fPET images, the same could be applied within the image reconstruction process provided the PET list-mode data was accessible. Research using image restoration techniques in a reconstruction framework generally employ a Poisson noise model for the sinogram data and a system matrix composed of several matrix operations representing the point spread function, forward projection, attenuation correction, scatter cor-

rection, and back projection. Our work solved the image restoration problem in the image domain and employed a least-squares-type data term rather than a fixed noise-model in the image space. This was because the noise characteristics of the reconstructed PET images inherently depend on the reconstruction algorithm. For example, noise characteristics during filtered back-projection reconstruction depend upon the filter employed, whilst in maximum likelihood expectation maximization reconstruction and its variants, the noise characteristics depend on the number of iterations as well as the strength of the prior function.

In this work, we have applied an MRI assisted PET data processing method to identify brain metabolic activations using FDG-PET. This concept can further be applied to other types of PET experiments such as neuroreceptor and neuroinflammation tracer studies, where counts are statistically limited during dynamic imaging for kinetic modeling analyses. However, these tracers will have different binding targets and biodistribution in brain tissues compared with FDG-PET. Therefore, it is essential to consider appropriate MRI contrasts when designing the MRI-MRF prior model. For example, **?** investigated the correspondence of quantitative susceptibility mapping (QSM) MRI and [11C]-PK11195 PET for the localization of tracer associated with neuroinflammation.

The current work has several limitations. One of the limitations is the small sample size. In this work, we show proof of the principle for utilizing anatomical information for fPET data processing. Advanced statistical image restoration models such as joint patch-based techniques and neural networks may further improve the image quality for shorter image acquisition durations and potentially in future approach the temporal resolution offered by fMRI. However, the proposed framework is readily adaptable to use these techniques in the research context, although modeling higher statistical dependencies would increase the number of hyperparameters that were required to be tuned. A further limitation is that the MRI-MRF modeled as a Bowsher-like prior may be perceived as a technique that relies excessively on the anatomical modality. Although this may be relatively unimportant or in fact, beneficial in the case of tracers like FDG that are widely distributed throughout the brain, this may not be the case for other heterogeneously distributed tracers such as for amyloid PET imaging. More sophisticated image restoration models that maintain a balance between the PET and MRI features for each tracer may need to be incorporated at the cost of more computational time. The use of spatial regularization could be carefully extended to include a temporal smoothing constraint governed by studies in tracer kinetics. A comprehensive study of several MRI-PET joint priors in the context of dynamic functional PET denoising and analytical techniques is an important direction for future studies.

Chapter 6

Conclusions and Future Work

6.1 Summary

Brain imaging using simultaneous PET-MRI system has shown promise for advancing neuroscience and clinical neuroimaging. The combined information of structure, function, and metabolism is unparalleled among other multimodal imaging systems. Furthermore, compared to PET-CT, PET-MRI reduces radiation exposure due to the non-ionizing nature of MRI. However, there is scope for further reduction in the radiation. Firstly, given its high specificity, enabling ultra-low-dose PET scans will greatly benefit pediatric oncology and oncology for pregnant women and adults in their reproductive age. This is because of the risk of accumulation of dangerously high levels of ionizing radiations owing to longitudinal imaging (required in tumor progression or regression studies). Secondly, PET enables presymptomatic diagnosis of neurodegenerative diseases such as Alzheimer's disease (**?**). Finally, the reduced radiations will enable increased research studies for neuroscience and understanding several other biological processes. In Chapter 2, we described the utility of low-dose imaging in the context of functional imaging using PET. On the other hand, faster scans are required for cost-effectiveness, optimized imaging protocols, and more importantly, to reduce the effect of patient motion on the reconstructed images. Accelerated imaging, in addition to reducing patient motion artifacts, also benefits populations who find it difficult to lay still in the scanner for a long time (e.g., neuroscience studies) such as children, elderly subjects, and patients with movement disorders or epileptic seizures. For the reasons stated above, it is beneficial and in some cases, necessary to enable ultra-low-dose PET and faster PET-MRI imaging.

Through Chapters 3– 5, we described the novel models developed to some of the prevalent problems related to PET-MRI brain imaging indicated in Chapter 1 and 2. The main aim of these research problems was to achieve improved PET (and MRI) image

quality arising from various applications that demand low-dose as well as faster acquisitions. Both low-dose PET acquisitions and accelerated data acquisition by undersampling of the k-space generate images with low SNR and aliasing artifacts, respectively. The thesis covered both the paradigms of within-reconstruction as well as post-reconstruction image enhancement techniques summarized in the following sections. We now summarize the developed models along with their implications and indicate possible directions for future work.

6.2 Joint Patch-based Dictionary for PET and MRI Reconstruction

We developed a novel Bayesian image reconstruction framework for the three inverse problems: reconstruction of (i) PET data with a fixed MRI image, (ii) PET and undersampled MRI data from single-coil, and (iii) PET and undersampled MRI data from multiple coils. For all the three cases, our reconstruction framework models the PET-MRI joint dictionary as a MRF, leading to Bayesian inference. This work was the first to propose a jointly modeled patch-based dictionary for PET-MRI reconstruction. The results on simulated and *in vivo* data demonstrated that our model restores both edges and textural features in the images with reduced cross-modality artifacts in comparison to state-of-the-art image-gradient-based priors. We evaluated the robustness of our joint dictionary to varying noise-levels in the test data without having to retrain the joint dictionary. The degradation in the performance of our model at high noise-levels was the least compared to other priors. The last couple of years has seen the evolution of DNN-based priors for PET and MRI image improvement within the reconstruction routine (**??**). Future work calls for adapting DNN-based priors for joint PET-MRI reconstruction. The proposed framework is generic and can be easily adapted to different datasets: e.g., different undersampling schemes for k-space data acquisition, or different tracers for PET images. However, the dictionary atoms need to be adapted in the case of the involvement of new tracer. While we observe a substantial improvement in PET image quality at different noise-levels by employing the joint dictionary, the contribution of the joint dictionary in improving the de-aliasing of the MRI images is not very clear. Indeed all the joint priors discussed in this work (LPLS, QPLS, and the joint dictionary) show improvement over independent priors (depending on one modality only). The reduction of aliasing artifacts due to the joint prior modeling were comparable across all methods. Nevertheless, our joint dictionary prior shows slightly better textural features in the MRI images and reduces the piecewise-smooth results, which are typical of image-gradient-based priors.

6.3 Sinogram- and Uncertainty- aware DNN (suDNN) for Predicting SD-PET from uLDPET

In Chapter 4, we proposed a novel DNN framework which modeled (i) the PET imaging physics as well as (ii) the per-voxel heteroscedasticity in the training data, leading to the robustness of the DNN to OOD data and quantification of uncertainties in the predicted SD-PET images in the absence of ground-truth /reference images, respectively. Analogous works in MRI image enhancement (de-aliasing) have shown that modeling imaging specific information into the DNN loss function yielded improved accuracy. In a first proof-of-concept study, our suDNN demonstrated robustness to physiological perturbations in the incoming test data (OOD data). Furthermore, we observed that although modeling the per-voxel heteroscedasticity as part of the DNN loss function slightly degraded the performance of the DNN in comparison to using the losses in the image space and sinogram space alone, the per-voxel heteroscedasticity images and the derived quantification maps provide a rich source of information for improved interpretation of the predicted images, adding to the reliability of the predicted images using the DNN. However, in the current form, the predicted images from our suDNN framework is not quantitative, i.e., not representative of quantities such as SUV. For subsequent clinical validation, it is imperative to modify the network to produce outputs that reflect the reduced number of photon-counts in the input LD-PET image. Additionally, efforts to produce quantitative uncertainty maps needs to be pursued.

6.4 MRI-Assisted High Temporal Resolution fPET for Improved Brain Mapping

Both Chapters 3 and 4 covered the aspects of robustness to incoming data that is unseen during training. On the other hand, Chapter 5 presented a novel MRI-assisted fPET processing framework for functional analysis of fPET data at high temporal resolution and for low doses of radiotracer. Compared to traditional Gaussian smoothing, our approach yields visually interpretable PET images while increasing the sensitivity and anatomical accuracy of activation maps estimated using ICA. Through validation using simulated data, we have demonstrated that the MRI-MRF method is able to accurately estimate visual task-related brain activation maps even under poor SNR conditions. The application to *in vivo* fPET data demonstrated that the MRI-MRF prior achieves the detection of reduced task durations of approximately three minutes and provides an avenue for further increases in the temporal resolution and sensitivity of both single subject and group-level

brain metabolic mapping studies. Advanced image quality improvement techniques using DNNs can aid in identifying accurate brain mappings at a temporal resolution approaching that of fMRI.

6.5 Towards Low-Dose and Faster Whole-Body PET-MRI

Through several examples in this thesis, we have seen that neuroimaging using PET-MRI has seen several fast-paced developments. On the other hand, there are several challenges inherent to whole-body imaging using PET-MRI. To name a few, implementing UTE MRI sequences for attenuation correction is challenging due to magnetic field inhomogeneities (**?**), limitations of MRI sequences for lung imaging, and truncation artifacts for large fields of view (**?**).

In addition to the challenges mentioned here, longer acquisition time and cost of operation also contribute to the limited acceptance of PET-MRI for whole-body studies. Several improvements such as continuous moving table MRI and specialized MRI sequences for whole-body imaging have reduced overall scan-time. The acceleration of MRI acquisition makes PET the time-limiting factor and reduced acquisition time for PET imaging degrades image quality. Hence, the extension of learning-based methods for low-dose whole-body PET imaging as well as accelerated MRI acquisition is an active area of research. The optimization of imaging protocols and the development of image quality improvement techniques for simultaneous PET-MRI in whole-body oncology and cardiology could unleash the full potential of the imaging system.

For cardiology, PET-MRI scanners provide unique opportunities in theranostics of myocarditis and myocardial infarction using perfusion studies. In this context, learning-based image quality improvement in cardiac PET-MRI can aid in detecting and localizing subtle changes in morphology and tissue characterization. For clinical whole-body oncology involving longitudinal examination, the reduction in the amount of ionizing radiations from PET in addition to the superior localization of the pathology from MRI can improve theranostics in several malignancies such as liver metastases, breast cancer and prostate cancer.

Complementary to the developments in imaging protocol and hardware technology (e.g., total-body PET scanners), machine-learning-based solutions can help the faster acceptance of PET-MRI for whole-body imaging. Compared to neuroimaging, the variability of learning-based methods for whole-body applications is higher due to the combination of several factors such as larger variations in patient physiology and internal

movement of organs. To this end, research and development of learning-based methods for whole-body imaging should focus on improving the reliability of the proposed solution across patients and scanners. As shown as part of this thesis, modeling uncertainties within the learning-based frameworks improve risk-assessment during inference. Furthermore, given the absence of reference images, quantification of the underlying uncertainty might prove to be the metric defining the success/failure of the specific learning-based model.

Finally, the developed learning-based methods as part of this thesis are working proof-of-concept for neurological imaging applications under low-dose and/or shorter data acquisition time. The translation of these methods can help improve the effectiveness of PET-MRI in understanding the physiology and function of the human brain in healthy and pathological conditions, making it the modality of choice for scientific and diagnostic investigations of the brain. Further optimization in terms of simultaneous data acquisition and joint analysis for low-dose and accelerated PET and MRI will rely on active research in image quality enhancement techniques using learning-based techniques. Although there exist several proof-of-concept solutions with the aim of further reducing the dose and accelerated image acquisition with the aim of improving the accuracy of the estimated images, future works in this direction should focus on improving the reliability of the method when deployed across different sets of patient cohorts, different scanners, use of different PET tracers, etc. This is crucial for the clinical translation of learning-based methods. Furthermore, in the context of learning-based methods for medical imaging, the developed models in this thesis can be extended to imaging other anatomies leading to improved whole-body PET-MRI.

List of Publications

This dissertation is based on the following studies reported across journals and conferences.

Journals (Accepted)

- VP Sudarshan and GF Egan and Z Chen and SP Awate (2020 February). Joint PET-MRI Image Reconstruction using a Patch-Based Joint-Dictionary Prior.
  Med Imag Anal., 101669. https://doi.org/10.1016/j.media.2020.101669

Journals (Under review)

- VP Sudarshan and S Li and SD Jamadar and GF Egan and Z Chen and SP Awate (2020 August). Incorporation of Anatomical MRI Knowledge for Enhanced Mapping of Brain Metabolism using Functional PET.
  Neuroimage (Revised manuscript under review).

- VP Sudarshan and U Upadhyay and GF Egan and Z Chen and SP Awate (2020 October). Towards Ultra-Low-Dose PET using Deep Learning with Out-Of-Distribution Robustness.
  Med Imag Anal.(Provisional application filed, Revision under preparation).

Full-length Conferences

- VP Sudarshan, K Gupta, GF Egan, Z Chen, SP Awate (2019). Joint Reconstruction of PET + Parallel-MRI in a Bayesian Coupled-Dictionary MRF Framework.
  In: Int Conf Med Imag Comput Comput Assist-Intervention - MICCAI 2019. Lecture Notes in Computer Science, vol 11766. Springer. https://doi.org/10.1007/978-3-030-32248-9_5

- VP Sudarshan, Z Chen, SP Awate (2018). Joint PET+MRI Patch-Based Dictionary for Bayesian Random Field PET Reconstruction.
  In: Int Conf Med Imag Comput Comput Assist-Intervention - MICCAI 2018. Lecture Notes in Computer Science, vol 11070. Springer. https://doi.org/10.1007/978-3-030-00928-1_39

Short Abstracts

- VP Sudarshan and S Li and SD Jamadar and PGD Ward and GF Egan and SP Awate and Z Chen (2020). MRI-assisted High Temporal Resolution Dynamic FDG-PET Imaging for Assessing Brain Functions.
  In: 2020 Proc Int Soc Magn Res Med. (ISMRM)

- VP Sudarshan and A Fernandez and K Pawar and S Li and GF Egan and SP Awate and Z Chen (2019). Synthesis of Standard-Dose FDG PET Images from Low-Dose Acquisition using a Combination of Atlas and CNN-Based Method.
  In: 2019 Proc Int Soc Magn Res Med. (ISMRM)

- VP Sudarshan and GF Egan and Z Chen and SP Awate (2018). PET Image Reconstruction using Joint MR-PET Dictionary.
  In: 2018 IEEE Nucl Sci Symp Med Imag Conf Proc (NSS/MIC) (pp. 1-3).

Patents

- VP Sudarshan and U Upadhyay and GF Egan and Z Chen and SP Awate. Systems and Methods for Automatically Enhancing Low-Dose PET Images with Robustness to Out-Of-Distribution (OOD) Data. Provisional patent application no: 202021036972. Filing date: 27.08.2020.

Acknowledgements

Over the past few years i have had the opportunity to understand the challenges related to simultaneous PET-MRI systems and subsequently, develop, and work on models for PET and MRI image analysis. This was possible due to the contribution of many people spread across the globe. I would like to express my sincerest gratitude to all the people involved.

First of all, thanks to my supervisors, Prof. Suyash and Dr. Zhaolin, for guiding me and imparting the skills needed to tackle medical image analysis problems. They have continuously supported me in defining the research problems of interest, allowing me to get a broader perspective of conducting research.

I thoroughly enjoyed as well as learned a lot from my discussions with Prof. Suyash on various methods, formulations, and implementation of the problems. Most importantly, he instilled the discipline of implementing a solution only after i have studied the formulations on paper.

On the other hand, Dr. Zhaolin has always provided me with a holistic picture of the overall imaging process which has helped me in understanding the imaging system and the real-world challenges related to the data, in a better way.

Both have at several points emphasized the importance of time and that has helped me in planning manuscript submissions and eventually, carrying out thesis submission on time. I will definitely miss the whiteboard sessions with both of them. Specifically, the discussions have led to better experimental designs for the models developed.

I am indebted to Prof. Gary for providing a strong vision for the problems pursued within this thesis. His involvement in the project through his swift responses with attention to detail has kept me motivated. Thank you so much Sir.

Insightful discussions with my friends and colleagues, especially my co-authors, have been an integral part of my work. I would like to thank all of them who are spread across Mumbai, Melbourne, Munich, and Madras.

I would also want to extend my thanks to the IITB-Monash academy which enables the joint research collaboration between IITB and Monash.

Above all, thanks to my parents and wife for always having supported me. Thanks to their patience, especially these last four years which kept me motivated.

Viswanath
IITB and Monash
11 February 2021