

This electronic thesis or dissertation has been downloaded from the King's Research Portal at <https://kclpure.kcl.ac.uk/portal/>



MRI-informed and machine-learning assisted PET image reconstruction

Bland, James

Awarding institution:
King's College London

The copyright of this thesis rests with the author and no quotation from it or information derived from it may be published without proper acknowledgement.

END USER LICENCE AGREEMENT



Unless another licence is stated on the immediately following page this work is licensed

under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International

licence. <https://creativecommons.org/licenses/by-nc-nd/4.0/>

You are free to copy, distribute and transmit the work

Under the following conditions:

- Attribution: You must attribute the work in the manner specified by the author (but not in any way that suggests that they endorse you or your use of the work).
- Non Commercial: You may not use this work for commercial purposes.
- No Derivative Works - You may not alter, transform, or build upon this work.

Any of these conditions can be waived if you receive permission from the author. Your fair dealings and other rights are in no way affected by the above.

Take down policy

If you believe that this document breaches copyright please contact librarypure@kcl.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

MRI-Informed and Machine-Learning Assisted PET Image Reconstruction



James Robert Bland

Supervisor: Prof. Andrew Reader

Dr Claudia Prieto

Department of Biomedical Engineering

King's College London

This thesis is submitted for the degree of

Doctor of Philosophy

October 2019

Abstract

Positron emission tomography (PET) allows an unparalleled insight into a wide range of medical conditions including neurological disorders, oncology and cardiovascular diseases. Despite PET's clinical utility, further improvements can still be made by tackling the susceptibility of reconstructed PET images to noise, particularly for low count PET projection data. In addition, the limited spatial resolution of the PET system leads to partial volume effects (PVEs), which introduce bias in regional quantification. The recent advent of clinically available simultaneous PET-MR scanners enables the simultaneous acquisition of complementary information within a clinically practical workflow. Reparameterising the PET reconstruction process through algorithms such as the machine learning inspired kernel method (KEM), utilises the readily available MR information and has been shown to aid in the suppression of noise and PVEs. However, uncertainty still persists in the capability of MR-informed methods to handle inconsistencies between the PET-MR datasets. Therefore, this thesis aims to demonstrate the major denoising power provided by the MR-informed kernel method, to show the effectiveness of modified KEM implementations in addressing its limitations, and to assess the performance of multiple KEM methods relative to other state-of-the-art MR-informed methods. An additional research direction, also originating from the machine learning field, is the inclusion of convolutional neural networks (CNNs) in PET reconstruction as a powerful means for image denoising. Further investigation and development is presented for unrolled neural networks, and their performance relative to traditional PET reconstruction methodologies. This thesis therefore identifies the key benefits

presented by the inclusion of MR-information into the PET reconstruction process, highlights how best to include MR into regularised or reparametrised PET image reconstruction, and explores the potential advantages of incorporating CNNs into PET reconstruction.

Acknowledgements

First off I would like to thank primary supervisor Andrew Reader, for his unwavering support, direction, (highly detailed) feedback, and overall contribution that he has made to my development as an independent scientific researcher. Andrew's dedication and time commitment to his group members elevates him far beyond the typical academic PI. I am highly indebted to the other group member's of the PET reconstruction team, Abolfazl, Martin, Sam and Casper, who have provided invaluable feedback, technical support, and friendship throughout the course of my PhD. I would also like to thank my secondary superior Claudia Prieto Vasquez, in addition to Paul Marsden, Jo Hajnal, and Rob Eckersley for their insightful comments and discussion.

Embarking on this PhD journey has also led to me gaining a whole host of fantastic friends and colleagues. I would particularly like to thank Aditi, Elis and Jessica, for being freakn' awesome :D (and maybe Gemma). I would also like to thank my not so new friends, who have put up with my poverty and my constant whining of said poverty.

My deepest gratitude goes to Weiwei, for her academic and emotional support that continues to bring me peace and joy. Lastly, I would like to thank my parents, Denise and Bob, and my sister, Georgina, for being a constant source of good in my life and a tireless supporter of me.

This PhD was funded by the King's College London and Imperial College London EPSRC Centre for Doctoral Training in Medical Imaging under Grant EP/L015226/1.

Table of contents

List of figures	8
List of tables	13
Repeated Symbols	14
1 Introduction to PET Imaging and the use of Prior Information	16
1.1 Introduction to PET Imaging	16
1.2 Limitations of PET Imaging	21
1.3 MR-Informed PET Reconstruction	23
1.4 Deep Learning in Medical Imaging	27
1.5 Motivation and Plan of Thesis	28
2 PET Imaging Background and Theory	31
2.1 PET Data Acquisition	31
2.2 Object Representation	38
2.3 Data Representation	39
2.4 Statistically Based Iterative Reconstruction	45
2.5 Inclusion of Prior Information	47
2.6 Iterative Solutions	50
2.7 PET Image Quality	56

2.8	Anatomically-Informed PET Reconstruction	59
2.9	Synergistic PET-MR Reconstruction	63
2.10	Deep learning in PET Reconstruction	64
3	MR-Informed Kernel Method for Low Dose PET Imaging	68
3.1	The MR-Informed Kernel Method	69
3.2	Theory	71
3.3	Simulated Phantoms and Patient Data	76
3.4	Hyperparameter Selection	79
3.5	3D Simulated Phantom Study	85
3.6	3D Real Data Study	91
3.7	Discussion	94
3.8	Limitations	98
3.9	Real Patient Case Study	99
3.10	Summary	101
4	The MR-Informed Kernel Method With Spatially Compact Basis Functions	103
4.1	Spatially Compact Basis Functions	104
4.2	PET Reconstruction Using MR Voxel Sizes	108
4.3	Simulated Phantoms and Patient Data	110
4.4	Hyperparameter Selection	112
4.5	3D Simulated Phantom Study	114
4.6	3D Real Data Study	121
4.7	Discussion	124
4.8	Limitations	129
4.9	Summary	129

5	Comparison of MR-Informed Reconstruction Methods	131
5.1	Reparameterisation vs Regularisation	132
5.2	Theory	134
5.3	Simulated Phantoms and Patient Data	139
5.4	Hyperparameter Selection	141
5.5	Simulation Study	144
5.6	Real Data Studies	155
5.7	Discussion	160
5.8	Limitations	165
5.9	Summary	166
6	Machine Learning Inspired Solutions for PET Reconstruction	167
6.1	Theory	171
6.2	Simulated Phantoms and Patient Data	176
6.3	2D Simulated Phantom Study	178
6.4	3D Simulated Phantom Study	185
6.5	Real Data Studies	189
6.6	Discussion	191
6.7	Limitations	195
6.8	Summary	196
7	Conclusions and Future Work	197
7.1	Summary	197
7.2	Limitations	199
7.3	Future Work	201
	Publications	203

Table of contents	7
References	205
Appendix A MLEM Theory	234
Appendix B MR Resolution Reconstruction of Noise Free Data	240
Appendix C Extended Evaluation of Simulated Dataset	241

List of figures

1.1	MRI (top row) and PET (bottom row) images shown at their native resolution	20
2.1	PET image acquisition, from positron annihilation, to a measured sinogram dataset.	35
2.2	Resolution degrading factors of PET imaging.	36
2.3	Schematic illustrating span 11 ($5 + 6$) axial compression across the 64 rings of the Biograph mMR scanner.	37
2.4	Variation in PET sensitivity in the axial and transaxial plane.	44
2.5	Schematic showing photon attenuation, random coincidences and scatter coincidences.	45
2.6	Summary of a subset of potential functions, their gradient, and their formulation.	49
2.7	The EM and gradient ascent algorithms solving a 1D objective function. . .	53
3.1	Schematic showing the construction of the MR-informed basis function, with and without the inclusion of the spatial Gaussian term.	76
3.2	Example basis functions extracted from the patient dataset described in §3.3.2 using the MR-informed kernel method.	76
3.3	Repeated line search along the hyperparameter values for each kernel hyperparameter, using a 2D simulated phantom.	83

3.4	Hyperparameter line search along the hyperparameter values for each of the kernel hyperparameter, using real 3D FDG data	84
3.5	Reconstructed KEM images using the maximum and minimum kernel hyperparameter value across the range of each hyperparameter.	85
3.6	A 2D transverse slice of the MLEM and KEM reconstructed images are shown for a range of reduced dose levels (from 100% to 1% of the full count data).	87
3.7	A 2D zoomed in coronal slice of the MLEM and KEM reconstructed images are shown for a range of reduced dose levels (from 100% to 1% of the full count data).	88
3.8	Whole brain and tumour NRMSE calculated for the reduced dose reconstruction of the 3D simulated dataset.	89
3.9	NRMSE (%) vs number of iterations for reduced dose KEM, simulated data.	90
3.10	Coefficient of variation versus mean value for a white matter region and the right caudate region, using simulated data.	90
3.11	Horizontal intensity profile through the tumour region in the frontal white matter region, simulated data.	91
3.12	KEM and MLEM reconstructed images of the 3D real patient dataset resampled at varying count levels (100% to 1%).	93
3.13	KEM and MLEM reconstructed images of the 3-D real patient dataset resampled at varying count levels (100% to 1%).	93
3.14	Whole brain NRMSE between the stated reconstruction method and the MLEM reconstruction of the full dose data	94
3.15	Coefficient of variation versus mean value for a white matter region and the right caudate, real data.	94

3.16	PET image and corresponding T1w MR image of a patient with unilateral temporal lobe epilepsy.	100
3.17	Reading results using low dose reconstructed images for a patient with unilateral temporal lobe epilepsy.	101
4.1	Schematic showing the construction of an MR-informed basis function. . .	107
4.2	PET and MR phantom constructed from the BrainWeb MR phantom. . . .	111
4.3	Basis functions of either the conventional kernel method (KEM k NN), or the proposed spatially compact kernel method for two different MR regions. . .	114
4.4	Vertical line profiles of the noise free difference images, through a tumour containing slice.	116
4.5	NRMSE versus iteration number for a number of kernel variations, applied to the high count simulated data.	117
4.6	NRMSE versus iteration number for a number of kernel variations, applied to the low count simulated data.	118
4.7	Mean value of tumour regions versus the standard deviation of white matter region for the simulated high count	119
4.8	Reconstructions of the three tumour regions for high count simulated data. .	119
4.9	Reconstructed images of the high count simulated data for a tumour containing slice.	120
4.10	Reconstructed images of the low count simulated data for a tumour containing slice.	121
4.11	Mean value of tumours versus the standard deviation of white matter region for the real data.	122
4.12	Reconstructions of the three tumour regions for real data.	123
4.13	Reconstructed images of the real data with the augmented tumours, for the tumour containing slice.	124

5.1	Schematic showing the categorisation of the different MR-informed methods under comparison.	134
5.2	Schematic showing the formation of the 3 MR-informed kernel methods under comparison.	136
5.3	Schematic showing the formation of the 3 MR-informed MAP methods under comparison.	137
5.4	Simulated FDG PET phantom based on the BrainWeb phantom, with the four PET-unique high intensity regions added.	140
5.5	Real patient FDG dataset, augmented with simulated tumour regions, to produce genuine regions of PET-MR mismatch.	141
5.6	NRMSE and SSIM trade-off curves for the MR-informed methods applied to high count simulated data.	148
5.7	Images of the MR-informed methods applied to the high counts simulated data.	148
5.8	The tumour mean value vs white matter standard deviation, for the MR-informed methods applied to high count simulated data.	150
5.9	NRMSE and SSIM trade-off curves for the MR-informed methods applied to simulated low counts data.	152
5.10	Reconstructed images of the MR-informed methods applied to low count simulated data.	152
5.11	The tumour mean value vs white matter standard deviation for the MR-informed methods applied to low count simulated data.	154
5.12	Tumour mean vs white matter standard deviation for the high count real data.	157
5.13	Reconstructed images for the high count real data.	158
5.14	Tumour mean vs white matter standard deviation for the low count real data.	159
5.15	Reconstructed images for the low count real data.	160
6.1	Schematic of the Noise2Noise framework.	175

6.2	Schematic showing the general architecture of the unrolled network reconstruction algorithm.	179
6.3	Schematic of the U-Net used in chapter 6.	181
6.4	Training RMSE curves for the unrolled and U-Net architectures.	182
6.5	Whole brain bias vs standard deviation calculated over 10 noise realisations	183
6.6	Single slice images, for the 2D simulated dataset reconstructed either by post-smoothed MLEM (at 100 iterations), a post-reconstruction U-NET or an unrolled network.	184
6.7	Training RMSE curves of Noise2Noise U-Net	186
6.8	Whole brain bias vs standard deviation curve for U-Net and CGD algorithm applied to simulated dataset	187
6.9	Images of the post-reconstruction U-Net applied to a test (simulated patient) dataset.	188
6.10	Images of the post-reconstruction U-Net applied to a test (real patient) dataset.	190
6.11	Internal activation images of the post-reconstruction U-Net applied to a test (real patient) dataset.	191
B.1	Vertical line profiles of the noise free images reconstructed at MR resolution, through a tumour containing slice.	240
C.1	Bias, standard deviation images and trade-off curves for the MR-informed methods applied to high count data.	243
C.2	Bias and standard deviation images and trade-off curves for the MR-informed methods applied to low count simulated data.	244

List of tables

1.1	Summary of the sensitivity and spatial resolution properties of the major medical imaging modalities.	17
3.1	Kernel hyperparameters that minimise the whole brain NRMSE for the 2D simulated and 3D patient scan datasets.	82
4.1	Algorithm and hyperparameters of the reconstruction methods investigated, for both simulated and real data.	115
5.1	Comparison of 6 MR-informed methods in terms of feature vectors, and PET/MR derived weights.	138
5.2	Summary of each MR-informed method's parameters, for both high and low count datasets.	143

Repeated Symbols

Roman Symbols

α	Kernel basis function coefficients
β	Penalty hyperparameter
χ	Intensity-spatial composite feature vector
δ	Small positive offset
ε	Noise vector
η_j	Neighbouring voxel's of voxel j
μ	Attenuation coefficients
Ω	Objective function
ω	High dimensional feature vector weights
Φ	Network learnable parameters
ϕ	High dimensional feature vector
ψ	Inter-voxel potential function
ρ	Complete data
σ_r	Spatial standard deviation
σ_{MR} or σ_f	Standard deviation of MR derived feature space
σ_{PET}	Standard deviation of PET derived feature space
θ	Object's spatial (voxel) coefficients
ζ	Continuous object distribution
A	System matrix

B	Object's spatial (voxel) basis functions
b	Additive background counts
f	MR intensity PET feature vector
g	Neural network
g_r	Regularisation network
K	Kernel spatial basis functions
k	Number of k-nearest neighbours
L	Poisson likelihood
m	Measured discrete projection data
n	Number of iterations
N_j	Neighbourhood length
N_p	Number of projection bins
N_v	Number of voxels
P	Image space PSF
p	Patch size
q	Modelled discrete projection data
R	Gibb's energy penalty term
r	Spatial coordinates
u	Prior image
w	Penalty weights
x	MR voxel coefficients
Y	Attenuation factors
Z	Normalisation factors

Chapter 1

Introduction to PET Imaging and the use of Prior Information

1.1 Introduction to PET Imaging

Positron emission tomography (PET) imaging is a highly versatile molecular imaging modality that enables a wide variety of biochemical processes to be observed and quantified. Molecules of a given chemical structure are selected to target particular physiological functions. The chosen molecules are then labelled with radioisotopes to allow external localisation. These radiolabelled molecules are referred to as radiotracers. Radiolabelled analogues of endogenous compounds are common choices of radiotracers. Patients undergo an intravenous injection (or rarely inhalation) of radiotracers before imaging. Through imaging, the spatial-temporal distribution of radiotracer uptake and the associated physiological function is observed. PET imaging differs from other nuclear medicine imaging modalities, such as single photon emission computed tomography (SPECT), through its use of positron emitting radioisotopes. After nuclear decay, the ejected positron undergoes annihilation with a local electron, producing two anti-parallel gamma photons. The inherent localisation provided by the anti-parallel photons is responsible for the increased detection sensitivity of PET

scanners' in comparison to SPECT. Approximately 5% of emitted photons are detected in PET systems, whereas the detection sensitivity of SPECT systems is an order of magnitude lower (dependent on the choice of collimator). The limited detection sensitivity of nuclear medicine and PET imaging systems is eclipsed by their high chemical sensitivity and the specificity of particular radiotracers. Chemical sensitivity is defined here as the quantity of molecular probe required to distinguish its signal from the background signal. The high chemical sensitivity of PET enables trace quantities (nanograms [1]) of radiotracer to be used, allowing the observation of biological systems without perturbation. This greatly exceeds the chemical sensitivity of alternative modalities, such as magnetic resonance imaging (micrograms to milligrams [1]) and contrast enhanced CT (not well characterised), as shown in Table 1.1.

Modalities	Contrast Agent	Sensitivity (moles L ⁻¹)	Spatial Resolution
MRI	gadolinium	10 ⁻³ – 10 ⁻⁵	25 – 100 μm
CT	iodinated media	10 ⁻¹ – 10 ⁻²	50 – 100 μm
PET	radiotracers	10 ⁻¹¹ – 10 ⁻¹²	1 – 2 mm
SPECT	radiotracers	10 ⁻¹⁰ – 10 ⁻¹¹	1 – 2 mm
Ultrasound	microbubbles	10 ⁻¹²	50 – 500 μm

Table 1.1 Summary of the sensitivity and spatial resolution properties of the major medical imaging modalities [2, 3, 1]. Sensitivity values for ultrasound and CT are not well characterised. Ultrasound images are also depth limited.

Through the appropriate choice of the radiolabelled molecule's chemical structure, numerous physiological and pharmacological processes can be monitored. The vast majority of PET scans (over 95%) use a radiolabelled glucose analogue, fluorodeoxyglucose ([¹⁸F]FDG), the uptake of which is closely linked to glucose metabolism. The moderate half life (110 mins) of the ¹⁸F radioisotope allows off-site production of the isotope, and minimises the patient dose sustained in comparison to longer half life isotopes. The widespread development of PET imaging clinically is due to [¹⁸F]FDG's efficacy in oncological cases, where it has been found to accumulate within malignant tumours [4]. In addition, modern day PET scanners are

generally combined with computed tomography (CT) that provides an accompanying high resolution anatomical image. These dual modality PET-CT systems aid tumour localisation and PET image quantification (through the calculation of attenuation correction factors) in comparison to the preceding stand-alone PET systems. Through accurately modelling the image acquisition process (including attenuation, normalisation and nuclear decay) the measured counts can be directly related to the quantity of radiotracer per unit volume (Bq ml^{-1}). These values of radioactivity concentration (conc_{img}) can then be normalised to account for the patients weight (M) and injected dose (ID) through calculation of the standard uptake value (SUV):

$$SUV = \frac{\text{conc}_{img}}{ID/M} \quad (1.1)$$

$[^{18}\text{F}]$ FDG PET-CT imaging is used in the diagnosis, staging and treatment monitoring of numerous cancers [5], including lymphoma [6, 7], lung [8], colorectal [9], head and neck [10], and oesophageal [11, 12]. The inclusion of PET imaging in the clinical workup for oncological cases has demonstrated a 30% change in patient management across all cancers [13]. Specifically, for head and neck cancers $[^{18}\text{F}]$ FDG PET imaging has played an increasing role in patient diagnosis and treatment planning [14, 15].

In addition to oncology, PET imaging is used for the assessment of myocardial viability ($[^{18}\text{F}]$ FDG) and perfusion ($[^{13}\text{N}]$ ammonia, $[^{15}\text{O}]$ water) for patients with ischaemic heart failure; detection of the site of focal infection ($[^{18}\text{F}]$ FDG); and neurological cases [16]. This work focuses on neurological PET, in which PET provides a useful clinical and research tool for neurodegenerative diseases (Alzheimer's, Parkinson's and Huntington's disease), neurooncology and epilepsy, amongst other conditions. Once again, the predominant clinical radiotracer in PET neuroimaging is $[^{18}\text{F}]$ FDG. For instance in Alzheimer's cases, the degree of dementia has been established to correlate with focal hypometabolism in parietal, temporal and/or frontal lobes and diffuse hypometabolism in the associative cortex [17, 18]. In addition to $[^{18}\text{F}]$ FDG a host of more specialised tracers are also utilised in neurological PET within

both a clinical and research setting. For PET imaging of amyloid plaques, [^{18}F]florbetapir and [^{18}F]florbetaben are used to investigate atypical patients suffering from unexplained memory loss, to confirm an Alzheimer's dementia diagnosis [19, 20]. The radiotracer [^{18}F]fluoroDOPA is used to image dopamine synthesis in the study of pre-synaptic brain function, and can be used for the early identification of patients suffering from Parkinson's disease [21]. The dopamine system plays a major role in neurodegenerative disorders such as Parkinson's and Huntington's disease that impair a patient's motor and cognitive function [22]. [^{11}C]raclopride and [^{18}F]fallypride are used for imaging D2-D3 and extrastriatal D2 dopamine receptors respectively [23]. Benzodiazepine (GABA_A) receptor imaging via the radiotracer [^{11}C]flumazenil can be used to localise epileptogenic regions [24]. Protein synthesis imaging via the amino acid based [^{11}C]methionine is used for neurooncology cases.

The utility of PET imaging is not confined to disease diagnosis, but has recently been extended to longitudinal studies to determine the treatment efficacy and optimal patient management [25, 6, 10]. In such studies, which are generally oncology focused, patients undergo an initial baseline PET scan, followed by an interim PET scan after completing a reduced chemotherapy course. Patients with an abnormal interim PET scan have their treatment escalated (more aggressive chemotherapy or radiotherapy) to improve the likelihood of cure, whereas patients with a clear interim PET scan have their treatment adapted to reduce the chances of side-effects.

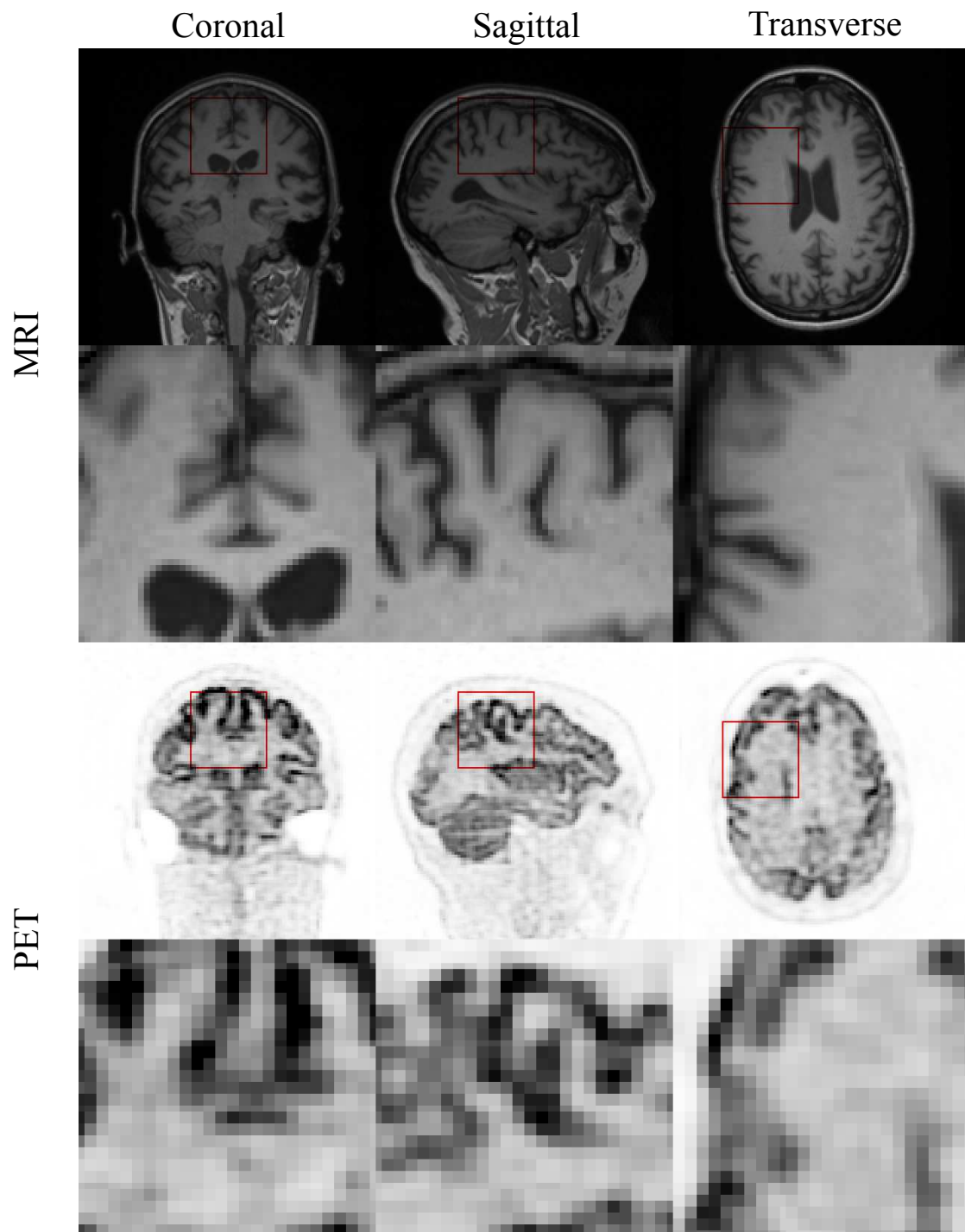


Fig. 1.1 T1 MPRAGE MRI (top row) and [^{18}F]FDG PET (bottom row) images shown at their native resolution, illustrating the respective image quality of the two modalities

1.2 Limitations of PET Imaging

PET imaging provides clinically distinct information to anatomical based modalities, which can shape patient diagnosis and management across a wide breadth of pathologies. Despite the clinical utility of [^{18}F]FDG PET, issues can arise due to the non specific nature of [^{18}F]FDG metabolism. This leads to PET imaging of [^{18}F]FDG being susceptible to false positives due to inflammation, infection and muscle use, in addition to other high metabolism processes. This can alter clinical interpretation and regional quantification. For instance, the use of [^{18}F]FDG PET-CT to identify pulmonary nodules in lung carcinoma cases provides a high diagnostic sensitivity of 96.8%, whereas the corresponding specificity for malignancy is only 77.8% [26]. Similar studies have also highlighted the limited specificity of [^{18}F]FDG PET uptake [27]. To address this issue alternative radiotracers can be used (in conjunction with FDG), which target biological processes that have greater specificity. Radiotracers such as ^{68}Ga -DOTATATE (which binds to somatostatin receptors) have demonstrated superior specificity and coronary imaging over [^{18}F]FDG for atherosclerotic inflammation [28, 29]. In the case of neurooncology, [^{11}C]methionine [30, 31] and [^{18}F]fluororethyltyrosine (FET) are radiolabelled amino acids that implicitly highlight cell proliferation (by targeting protein synthesis) and have been shown to provide superior tumour extent for low to intermediate grade gliomas (in comparison to FDG). It is also standard clinical practice to require patients to fast in preparation of their PET scan, to reduce the high metabolism of glucose in irrelevant tissue regions.

The predominant limitation of PET imaging (and nuclear medicine) is its poor image quality in comparison to anatomical modalities such as CT or MRI. This is primarily due to the intrinsic relationship between the PET image's signal to noise ratio (SNR) and the effective patient dose. For a fixed scan time period, a higher injected activity will lead to an higher number of detected counts and a reconstructed image with a higher SNR, but this is at the expense of an increased patient dose. Pragmatically, the detected count

level could also be increased through increased scan time, however this has drawbacks of reduced patient throughput, and is generally impractical. Clinical dose levels are chosen in adherence to guidelines that require all considerations are made to minimise the patient dose without compromising the PET image's diagnostic quality [32]. In reality, no ideal dose level exists and a compromise is made in terms of patient dose level and diagnostic image quality. For clinical scans the injected activity lies within the range 150 - 400 MBq, which corresponds to an whole body effective dose between 3-8 mSv for an adult [32]. Within the clinical dose range the noise present in the measured dataset is evident and visibly degrades the reconstructed image. The limited detector sensitivity of current clinical PET scanners exacerbates the situation further. For oncology cases, noise present in the PET image increases the chance of false-positives for tumour identification. In particular, tumour malignancy is in part decided from the maximum pixel standard uptake value (SUV_{max}) within a ROI, with higher SUV_{max} values increasing the risk of malignancy. High levels of noise within the specified ROI will make this SUV_{max} value less reliable [33]. TNM (tumour nodule metastasis) staging and determining the extent of tumours may also be effected by high levels of noise. For the assessment of neurodegenerative diseases, high levels of noise will impact the regional quantification within small regions of interest. Furthermore, a radiologist's ability to recognise spatial patterns (e.g. metabolism asymmetry), which is essential for the diagnosis of some neurological conditions, may be negatively affected by noise degraded images. It should also be noted that there is currently no known "safe" radiation level, below which the stochastic effects of radiation exposure are inconsequential. Therefore, the capability to reduce the radiation exposure, whilst maintaining or improving the diagnostic quality of the corresponding PET image would be beneficial in reducing the risk to the patients. A lower radiation dose will also reduce the dose routinely received by radiation workers.

A second limitation of PET scanners is their poor spatial resolution (≈ 4 mm for post-reconstruction smoothed clinical [^{18}F]FDG) in comparison to anatomical modalities (≈ 1 mm), which leads to visible blurring of the image. The PET scanner's limited spatial resolution is due to the finite detector size, photon acollinearity and positron range (amongst other factors). The voxel size in the reconstructed image (≈ 2 mm) is of comparable extent to the PET system's spatial resolution, leading to partial volume effects (PVEs), such as the spill over of small high intensity regions to neighbouring voxels. A second minor factor is the tissue fraction effect, due to one voxel being comprised of multiple different tissues [34]. These PVEs introduce bias in the regional quantification of the PET image [35]. As previously discussed, regional quantification is clinically important especially in the assessment of neurodegenerative diseases. Blurring artefacts due to PVEs will also artificially increase the extent of tumour regions. Possibly of greater consequence is the effect of PVEs in the identification of necrotic tumours or necrotic nodes that appear as cold regions with only a thin metabolically active border. The spill in of activity from the border into the cold region due to PVEs severely impedes the diagnostic value of [^{18}F]FDG PET [36, 37]. Low intensity tumours and cysts can also result from some cancers [38], for which the diagnostic value of PET is affected by PVEs.

This problem of degraded image quality can be approached from multiple research directions within the fields of data acquisition and image reconstruction, with this work addressing the later. Two methods that can address the limited image quality of PET (for a fixed patient dose and scan time) are the inclusion of MR anatomical information and/or machine learning (e.g. deep learning) into the PET reconstruction process.

1.3 MR-Informed PET Reconstruction

The recent development of simultaneous PET-MR scanners such as Siemens Biograph mMR [39, 40] or GE Healthcare's SIGNA PET-MR [41], readily provides the capability to incor-

porate co-registered MR information simply into the clinical PET work-up. In comparison to the current clinical "gold standard" of PET-CT, PET-MR scanners possess numerous benefits. Anatomical MR images, such as T1 or T2 weighted, provide complementary structural information with excellent soft tissue contrast, which generally lies in closer agreement with the PET radiotracer distribution in comparison to the equivalent CT image. This agreement between the structural MR and the extent of the PET radiotracer, aids the localisation of the PET signal. Consequently, PET-MR has shown diagnostic advantages for some cancers, such as prostate cancer where [^{18}F]choline PET-MR has an increased detection rate (86% vs 76%) and an increased sensitivity for lesion detection within the prostate gland (100% vs 67%) in comparison to PET-CT [42]. Although, the converse trend is true for the detection of other forms of lesions, such as small pulmonary nodules [43, 44], for which PET-CT has a higher detection rate than PET-MR. On balance [^{18}F]FDG PET-MR and PET-CT provide comparable diagnostic information (when MR is used for the anatomical framework only) [45].

Other potential advantages of PET-MR systems include the non-ionising nature of MR imaging, which reduces the radiation dose to approximately half that of the equivalent PET-CT scan. This is of heightened importance for paediatric patients and could possibly extend the use of PET-MR systems to clinical screening [46]. The positron range within a PET-MR scanner is also reduced due to the motion of charged particles within a magnetic field [47]. For high energy positron emitting isotopes such as ^{82}Rb the magnetic field confinement provided in PET-MR systems improves the spatial resolution FWHM from 5.24 mm to 3.90 mm at 3T [48]. Although, for the ubiquitous FDG isotope the resultant improvement in spatial resolution is minimal [49]. The simultaneous acquisition of PET-MR also enables the extraction of corresponding motion vector fields (e.g. to account for respiratory motion), which can be incorporated into the reconstruction process to reduce the prominence of image blurring artefacts [50–53]. Motion correction is important in compensating for both voluntary

and involuntary motion [54]. Imaging protocols that are subject to respiratory motion, such as thoracic PET imaging can greatly benefit from motion correction in terms of detectability and quantification [55]. Involuntary motion characteristics of some neurodegenerative disorders may lead to bias between healthy controls and patients [56, 57]. An alternative benefit of simultaneous PET-MR systems is the capability to perform multiple MR sequences within the duration of a PET scan (20-30mins). For instance during a simultaneous PET-MR acquisition an image-derived input function can be obtained through using a co-registered TOF MR angiography (to determine an arterial region of interest) in combination with the PET image. TOF MR angiography refers to an MR sequence used to visualise blood vessels that does not require a contrast agent. This provides a non invasive alternative to arterial cannulization, which can be used for dynamic PET data quantification [58]. PET imaging can also be combined with mutliparametric imaging sequences such as diffusion weighted imaging, perfusion MRI and MR spectroscopy [59]. PET-MR's enhanced functional anatomical correspondence is particularly advantageous for neurological [^{18}F]FDG PET scans, where the radiotracer distribution is well delineated between the white and grey matter boundaries. For conditions such as Alzheimer's disease, PET-MR provides unique advantages in improving early and differential diagnosis [60]. Lastly, as with the dual modality PET-CT scanner, the dual modality PET-MR scanner provides a number of practical advantages, such as implicit image registration.

In comparison to sequential PET-MR systems, simultaneous PET-MR systems provide clear advantages such as increased patient throughput, and minimal adjustment for co-registration. The clearest benefit of simultaneous PET-MR is the guarantee of observing the patient in the same state with both modalities. Sequential PET-MR images may have a reduced correspondence due to potential changes in the patient's physiological or mental state. This is particularly true for neurological studies [61]. Secondary benefits of simultaneous

PET-MR systems include the easier application of MRI-based PVC and MRI-informed reconstruction methods to the simultaneously acquired PET datasets [62, 63].

The close correspondence between the [^{18}F]FDG radiotracer distribution and the soft tissue structure in MR images, makes MR images highly suitable to regularise the PET reconstruction process. Within the field of PET reconstruction, numerous MR-informed algorithms have been proposed to reduce noise in the reconstructed image whilst preserving anatomical boundaries [64, 65]. Anatomical based regularisation also helps to negate PVEs, caused by a combination of resolution degrading factors and the limited spatial resolution of the PET scanner [66, 67]. The reduction of PVEs through the incorporation of MR information can be further extended through reconstruction of the PET image at MR resolution [68–70]. The improvements garnered by these MR-informed methods have been evaluated in terms of the clinically pertinent evaluation metrics of regional quantification and tumour detection [71]. Enhanced regional quantification of small neurological regions, such as the striatum, is of particular importance considering the use of regional hypometabolism to assess conditions including Parkinson’s [72, 73] and Alzheimer’s [74, 75] disease. For these MR visible regions the use of MR information produced significantly better results over conventional reconstruction methods in terms of quantification [65]. In the task-based assessment of lesion detection MR-informed emission tomography (ET) reconstruction methods also demonstrated superior performance over conventional reconstruction methods, for cases where the lesion boundary is present in the anatomical image [76–79]. However, for the case of no anatomical lesion boundary the utility of MR-informed reconstruction methods appears minimal [80], and may even suppress the reconstruction of hypermetabolic lesions [81, 64]. This is a major pitfall for the visual diagnostic interpretation of PET images, in particular for oncological cases. Changes in tumour extent, SUV_{max} , and shape due to incorrect anatomical guidance may negatively impact tumour detection and grading, and consequently the patient’s management.

1.4 Deep Learning in Medical Imaging

The implicit inclusion of population statistics into PET image reconstruction via supervised machine learning methods provides an alternative avenue to address the current shortcomings within PET reconstruction, and medical imaging generally. Within the field of supervised machine learning, artificial neural networks, and in particular deep learning [82] has taken greater prominence. Deep learning has surpassed the previous state-of-the-art performance level in image classification [83], speech recognition [84], natural language processing [85, 86] and strategy games [87]. Recent advancements in the field of computer vision have been accompanied with a trend towards convolutional neural networks (CNNs) [88] with deeper [89], and more complex [90, 91] architectures. Deep learning's recent success in image classification (and other autonomous challenges) has led to it being used in a variety of more applied tasks including traffic flow [92], semantic segmentation [93], astronomy [94] and fault diagnosis of rotating machinery [95], to name but a few. In the general field of image processing, deep learning methodologies have shown state of the art performance for image denoising [96–98], super resolution [99–101], style transfer [102] and deconvolution [103]. Directing our attention to the field of medical imaging in particular, the use of deep learning methods has become common place in image analysis [104–108] image restoration and more recently image reconstruction [109–111]. Such methods have only recently emerged in the more specialised discipline of PET image reconstruction [112–116]. Based on the strong performance of deep learning within the broader field of medical imaging, its inclusion into the PET reconstruction process may further improve the reconstructed image, with respect to the clinically pertinent evaluation metrics of regional quantification, noise level, and tumour identification.

1.5 Motivation and Plan of Thesis

This thesis investigates the use of both MR information and supervised machine learning methods for the improvement and development of PET image reconstruction. A major issue of including prior information into the reconstruction process is the possibility of biasing the reconstructed image, due to the lack of correspondence between the true PET distribution and the prior distribution. This thesis therefore looks into both MR-informed and deep learning inspired PET reconstruction methods, and seeks to improve the performance of such methods for the case of differing PET-MR structures, and real patient data, respectively. More specifically, the following contributions are presented in this thesis:

- The MR-informed kernel method [117] applied to low dose PET datasets.
- Modifications to the MR-informed kernel method to address its limitations in the described scenario of differing PET-MR structures.
- A comparison of multiple kernel inspired PET reconstruction methods with state-of-the-art MR-informed regularised methods, for the case of limited PET-MR correspondence.
- Suitable configurations of unrolled CNNs for minimising testing error on 2D simulated PET datasets.
- Validation of the Noise2Noise framework on real patient 3D PET datasets [118, 119] for post-reconstruction CNNs.

To summarise, this work identifies the key benefits presented by the inclusion of MR-information into the PET reconstruction process, highlights how best to include MRI into regularised or reparametrised PET image reconstruction, and explores the potential advantages of incorporating CNNs into PET reconstruction. The organisation of this thesis is as follows.

Chapter 2 describes the physical processes underpinning PET data acquisition. This is followed by the mathematical background and theory behind the statistical reconstruction methodologies that are used in both a clinical and research setting. The inclusion of prior information into the reconstruction algorithm is discussed. Lastly, a brief overview of the development of deep learning in PET reconstruction is presented.

Chapter 3 introduces the MR-informed kernel method with greater theoretical detail. The kernel expectation maximisation (KEM) algorithm is employed for the novel application of patient dose reduction, down to a limit of 1% of the original dose level. The performance of the KEM method is evaluated for both simulated and real patient 3D datasets, and compared to conventional reconstruction algorithms. The limitations of the MR-informed KEM method with respect to differing PET-MR structures is also highlighted.

Chapter 4 proposes an adaptation to the preceding MR-informed KEM method, altering either the algorithm or the feature vector used to form the kernel matrix. The KEM method is also extended to reconstruct the PET image at the native MR voxel size, further reducing PVEs. Comparisons are made between the proposed KEM method and the conventional KEM method, evaluated in terms of their capability to recover unique PET features, whilst also improving overall image quality metrics. Both methods are investigated using 3D simulated and real data studies.

Chapter 5 compares a number of kernel and maximum a posteriori (MAP) methodologies, with the aim of identifying methods that enable a favourable trade-off between the suppression of noise and the retention of unique features present in the PET data. The reconstruction methods investigated were: the MR-informed conventional and spatially-compact kernel methods; the MR-informed Bowsher and Gaussian MR-guided MAP methods; and the PET-MR-informed hybrid kernel and anato-functional MAP methods. The trade-off between improving the reconstruction of the whole brain region and the PET-unique regions was investigated for all methods, using both simulated and real 3D datasets.

Chapter 6 investigates unrolled neural networks for the reconstruction of simulated 2D PET data. The presented network was shown to simultaneously improve both image quality and rapidly accelerate reconstruction times, in comparison to post-smoothed MLEM and a post-reconstruction U-Net. The unrolled network's number of modules and denoising layers were varied to stratify the relative importance of each. As a final investigation the post-reconstruction U-Net was used to reconstruct real patient 3D PET datasets, through training using only noisy training labels.

Chapter 7 summarises the overall contributions presented in this thesis. Further comparison is made between the MR-informed and machine learning inspired reconstruction methodologies. Potential limitations of this work are highlighted, along with suggestions for future directions.

Chapter 2

PET Imaging Background and Theory

This chapter provides a brief overview of the underlying principles and theory that govern the data acquisition and image reconstruction processes of a PET system. Limitations in a PET system's sensitivity and resolution in addition to the stochastic nature of radioactive decay degrades the acquired projection data. Statistical reconstruction methods provide multiple benefits in atoning for these limitations, justifying their use in clinical practice. However, further improvements to PET image quality and clinical efficacy can be made through the inclusion of prior information into the reconstruction process. Both anatomical and deep learning based priors show promise in this regard. The advancements made in these fields shall also be reviewed.

2.1 PET Data Acquisition

The phenomenon of nuclear decay underpins both PET and nuclear medicine imaging systems. Nuclear medicine techniques such as SPECT, use gamma-emitting radioisotopes that decay and produce tissue penetrating radiation. The radiation is detected externally to the patient, from which the original distribution of a radioisotope within the patient is estimated. PET imaging systems are restricted to the use of positron emitting radioisotopes. The most

common radioisotope used in clinical PET imaging is ^{18}F , which mainly undergoes decay via the β^+ decay mode (branching fraction 97%). The nuclear equation for this positron emitting decay mode is:



where e^+ is a positron and ν_e is the corresponding neutrino. The ejected positron (e^+) interacts with the surrounding medium, reducing its kinetic energy to thermal levels before undergoing annihilation with an electron (Figure 2.1, row 1). The positron ejected from a decaying ^{18}F nucleus can have a range of energies (peak energy of 0.633MeV) and will traverse a positron range (path length of a positron with peak energy is 2.4 mm in water) before annihilation [120, 121]. The positron electron annihilation produces two antiparallel photons of energy 511keV (the mass energy of an electron). The two photons are detected within a specified time interval, allowing the photons to be attributed to the same decay event. Therefore, the event can be localised to lie within a connecting line (or tube) of response (LOR) between the two detectors.

The paired detected signals (referred to as counts) for each LOR can be stored either in sinogram (multiple sinograms for 3D), or list-mode data formats. For sinogram datasets, the detected counts are binned into time intervals (for dynamic PET scans), or summed across the whole scan duration (for static PET scans). Each LOR in the sinogram dataset is specified by its angular orientation (ϕ) and its perpendicular distance from the gantry's centre (s) [122]. Figure 2.1 illustrates how a uniformly emitting point source placed off centre within a simplified 2D PET scanner will produce a sinusoidal shaped trace. A 2D sinogram of the Shepp Logan phantom is also shown for illustrative purposes [123]. Recent PET scanners also have improved detector temporal resolution, down to 210 ps for the Siemens Biograph Vision PET-CT [124]. This short time resolution allows the delay in arrival between two detected photons from the same decay event to be calculated. This facilitates further localisation of the event to lie within a short segment of the LOR.

This detector capability is referred to as time of flight (TOF). However, for the detector temporal resolution currently available in production scanners, the use of TOF presents only minimal visual improvement in comparison to conventional (non-TOF) methods where the event probability along a LOR is uniform. Rapid progress is being made to further improve detector and scintillator temporal resolution, with current research detector systems now capable of sub 100 ps timing resolution, providing a 5 fold increase in signal to noise ratio (SNR) over non-TOF systems [125, 126]. The gain in SNR for a TOF system relative to a corresponding non-TOF system is expressed through the below relationship between the patients diameter (D), the coincidence timing resolution (τ) and the speed of light (c):

$$\frac{SNR_{TOF}}{SNR_{non-TOF}} = \sqrt{\frac{D}{c\tau}} \quad (2.2)$$

Where $c \times \tau$ is the distance over which the decay can be localised along the LOR. This thesis will only consider non-TOF PET systems such as the Siemens Biograph mMR, and will not account for TOF in the reconstruction methodologies presented.

The technical merit of PET imaging over single photon imaging stems from the localisation of events to lie within a LOR, without the need for collimators. This implicit localisation is based on the assumption of the detected photons being collinear. In reality, the majority of annihilation photons deviate from collinearity (with a FWHM of 0.5° [127]) due to the net momentum of the positron electron system prior to annihilation. The non-collinear angle can result in a photon landing on an adjacent detector, causing the activated LOR to spatially deviate from the event location (as shown in Fig. 2.2, row 2). The magnitude of this resolution degrading effect is proportional to the bore size. There are also a number of other physical factors that degrade the spatial resolution of the PET system. The positron range, i.e. the displacement between the nuclear decay and the annihilation event caused by the positron's motion (Fig. 2.2, row 1), can lead to the LOR not including the decay event location. The extent of the positron range is highly dependent on the kinetic energy

spectrum of a given radioisotope [128]. The finite detector width (d) also limits the resolution of the reconstructed image. The spatial resolution of the detectors varies along the transaxial direction of the gantry from approximately $d/2$ at the centre, to d at the edge. By contrast, the spatial resolution of the detectors in the axial direction remains relatively fixed, and is dependent on the intra-ring separation (width) and the axial compression used. The parallax error caused by detector depth also degrades the spatial resolution in a spatially variant manner along off-centre lines of response. Depth of interaction (DOI) detectors have been proposed as a potential solution, which can more accurately determine the LOR, whilst retaining the stopping potential of the original scintillation crystal [129]. Additional resolution degrading factors include inter-crystal Compton scattering as shown in Figure 2.2, row 4. Some DOI detector designs also use Compton scattering to enhance detection efficiency as well as reduce parallax errors [130].

The entirety of the real data used in this thesis was acquired on the Siemens Biograph mMR scanner. The axial FOV of the scanner is 25.8cm and the transaxial FOV is 59.4 cm [131]. It is a 3D PET system, enhancing its detection sensitivity relative to 2D PET systems, with a maximum ring difference of 60 permitted for LORs. The proprietary scanner software applies span 11 axial compression (as illustrated in Fig. 2.3), reducing the number of sinogram planes from 4084 to 837. Higher span values ease the memory requirements and computational burden of image reconstruction, however this is at the expense of degraded contrast and spatial resolution. A shift variant PSF is required to accurately model axial compression in the system matrix. Other forms of data compression have been used in previous scanner generations, such as sinogram mashing, which reduces the angular sampling and trimming, which reduces the number of sinogram bins (bin truncation). The precise resolution properties of the Siemens mMR scanner have been reported in the literature [39].

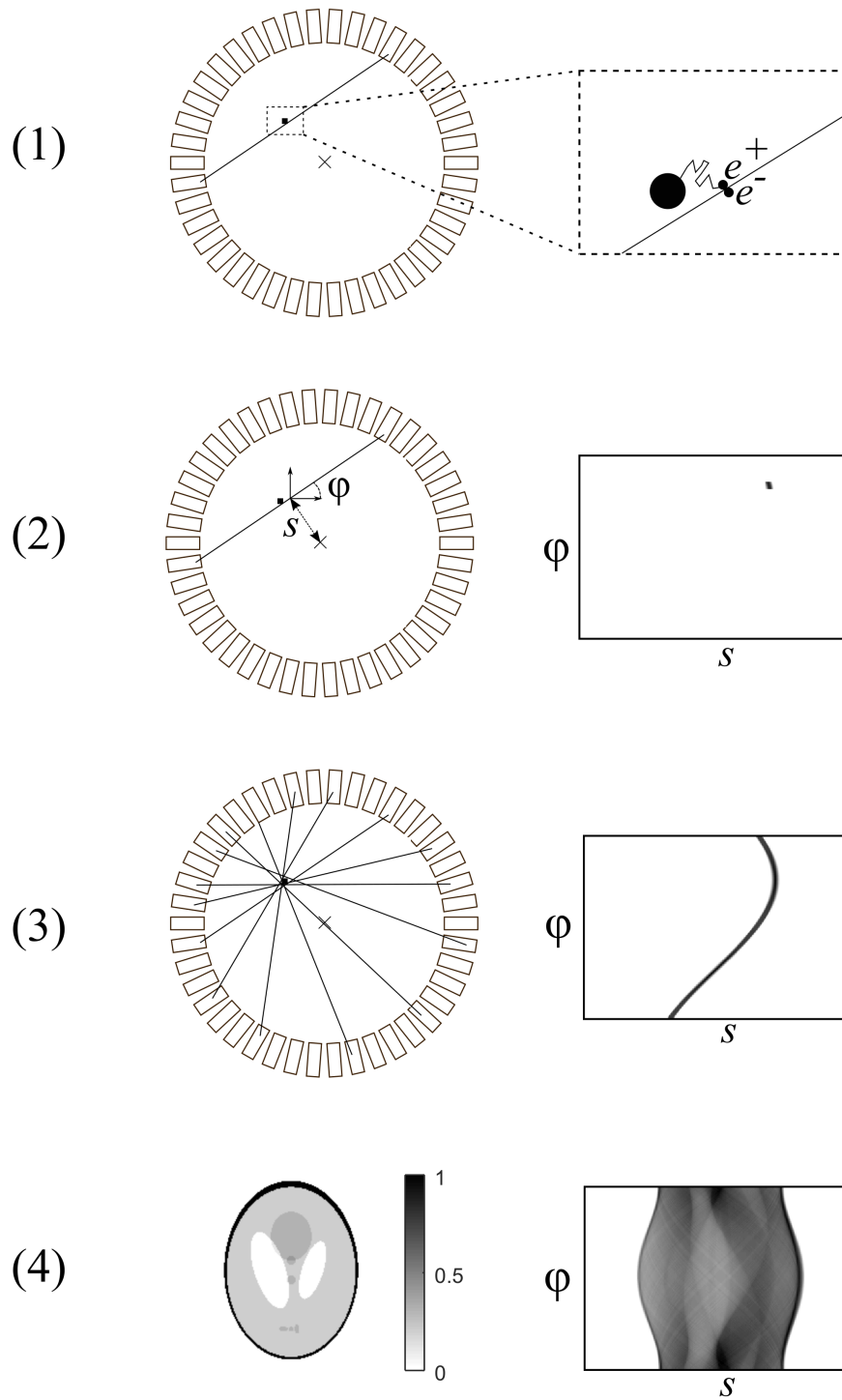


Fig. 2.1 (1) A point source undergoes nuclear decay emitting a positron that undergoes annihilation with an electron, producing two anti parallel photons. (2) The angle (φ) and distance (s) specify each LOR. The resultant sinogram trace for a single decay event along a given LOR is shown. (3) Point source is shown undergoing multiple nuclear decays events followed by annihilation events. The resultant sinogram trace is shown on the right. (4) Shepp Logan phantom [123] and the corresponding sinogram are shown.

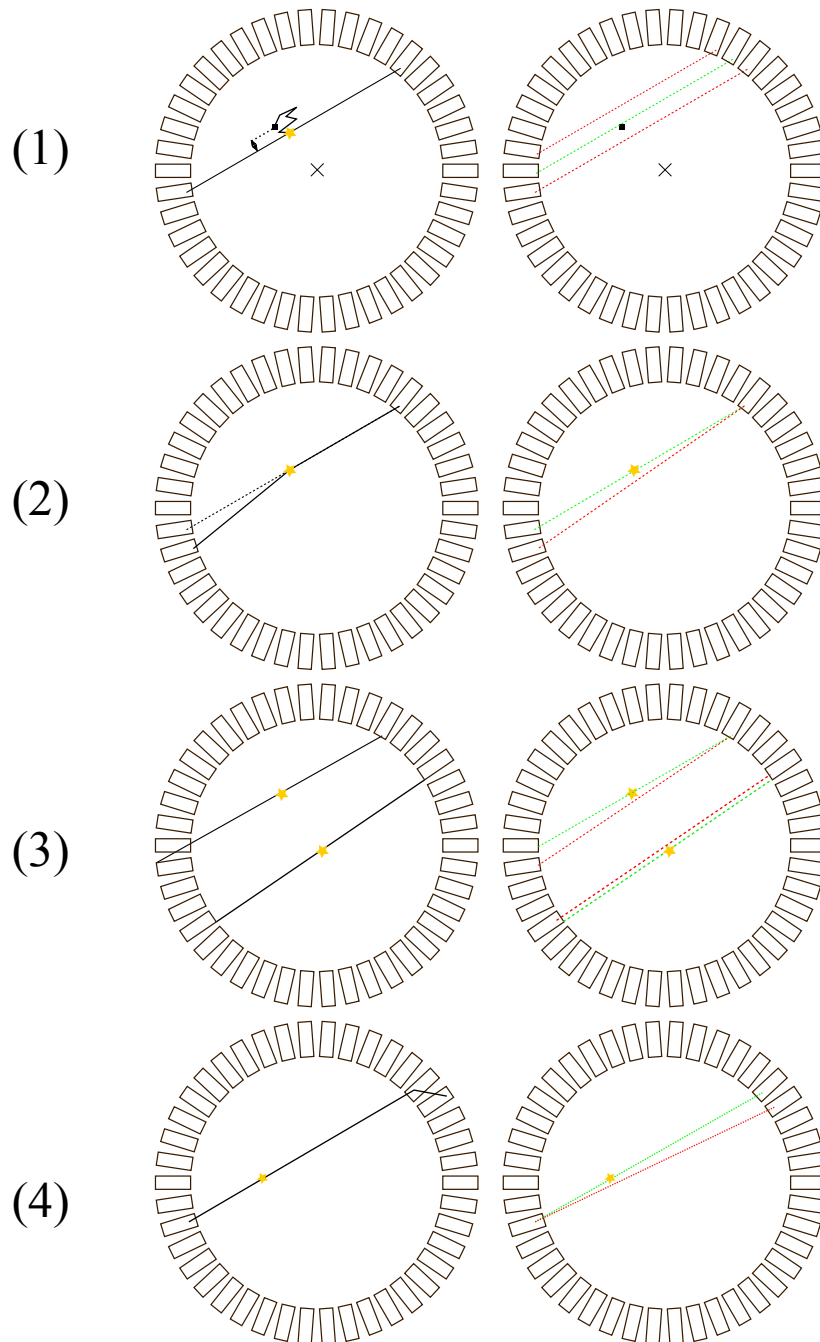


Fig. 2.2 Resolution degrading factors of a current PET system. Left hand column illustrates possible sites of nuclear decay or annihilation events and the corresponding pair of photons. The right hand side represents the ideal LOR in green, and the actual resolution degrading LOR in red that no longer intersects the decay or annihilation event location. (1) Positron range between the nuclear decay event and the positron electron annihilation. (2) The impact of photon acollinearity. (3) Possible variation in event location (parallax error) due to detector width and detector depth. (4) Intercrystal scattering.

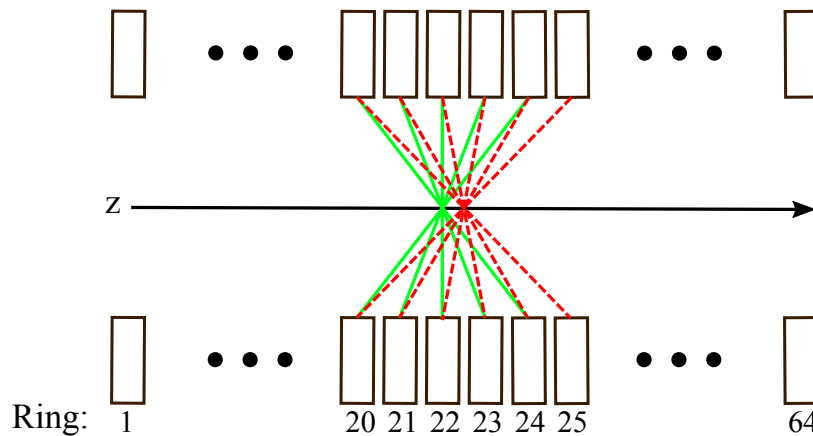


Fig. 2.3 Schematic illustrating span 11 ($5 + 6$) axial compression across the 64 rings of the Biograph mMR scanner. The 5 green solid lines are combined to contribute to the direct sinogram for ring 22. The 6 red dashed lines are combined to form an inter ring sinogram that lies between rings 22 and 23. A span of 121 produces direct sinogram planes only, seeing as the maximum ring difference allowed is 60.

A second important factor in the acquisition of PET data, is the physical sensitivity of the system. The system's sensitivity can be subdivided into the solid angle encompassed by detectors surrounding the radioactive source (geometric efficiency), the stopping power of the scintillation crystal, and the quantum efficiency of the detector. Recent efforts have been made to maximise the physical sensitivity of PET systems, through extending their physical extent [132], or combining coincidence detection with Compton scattering [133]. As mentioned above, increased scintillator depth is important for improved stopping power, as is the choice of scintillation material such as BGO, LSO and LYSO. The number of scintillation photons produced by the scintillator is also important due to its relation with energy and temporal resolution. Lastly other physical factors, such as gaps between scintillation crystals and detector dead time at high count levels can also effect the physical sensitivity of the system.

In addition to the true coincidence counts discussed above, where two photons are detected within a given time interval that originate from the same radioactive event, random and scattered coincidence events are also detected as shown in Figure 2.5. Random coincidence

events (Figure 2.5, panel 3) occur when two photons from two separate decays are detected within the coincidence time window. The associated LOR from the detected photons, is unlikely to pass close to the original location of either decay event. Scatter coincidence events (Figure 2.5, panel 4) occur when one (or both) of the two photons emitted from a single decay event is scattered into a different detector element. Scattering of photons to lie outside the original LOR, also attenuates the number of counts detected along the original LOR. These attenuation effects can lead to notable artefacts in clinical studies if not properly accounted for, such as anomalously high uptake at the skin or within the lungs of a patient (due to the lack of attenuating tissue in these regions).

2.2 Object Representation

The PET image reconstruction problem seeks to determine the original continuous distribution of a radioactive object (ζ) from a finite number of projection measurements ($\mathbf{q}$). Only time-independent static PET imaging is described in this work, where the object model can be fully described by $\zeta(\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ represents the spatial parameters. To accurately represent the continuous object distribution ($\zeta(\boldsymbol{\theta})$) would require infinite sampling across all projection coordinates. In reality a discrete representation of the object is used, with the appropriate spatial resolution chosen for a given projection sampling. The vast majority of PET reconstruction algorithms model the object distribution using cubic voxel basis functions in a regular 3D lattice:

$$\zeta(\mathbf{r}, \boldsymbol{\theta}) = \sum_j^{N_v} \theta_j B(\mathbf{r} - \mathbf{r}_j) \quad (2.3)$$

$$\zeta(\mathbf{r}, \boldsymbol{\theta}) = \sum_j^{N_v} \theta_j \Pi(|\mathbf{r} - \mathbf{r}_j| \leq \frac{v}{2}) \quad (2.4)$$

where the voxel basis function (B) describes a top hat like function (Π) and v represents the voxel width. Other choices of linear spatial basis functions include Gaussian convolution

kernel sieve [134], overlapping blob shaped basis functions [135] and regions of interest [136]. The use of non-linear basis functions has been mainly restricted to dynamic PET imaging for kinetic modelling, where many biological parameters of interest are inherently non-linear [137].

2.3 Data Representation

The PET imaging system can be modelled in terms of a linearly shift variant relationship between the continuous object distribution and the measured projection data. Although it should be noted that the assumption of linearity can become questionable at high count levels, due to detector dead time [138]. For static PET imaging the linear relationship between the continuous object distribution and the modelled projection data ($\mathbf{q}$) can be expressed as:

$$q_i(s, \varphi) = \int^{LOR_i} \zeta(\mathbf{r}, \boldsymbol{\theta}) dl \quad (2.5)$$

For a discrete modelled sinogram dataset and an object described in terms of voxel basis functions the above equation can be re-expressed in terms of matrix-vector multiplication:

$$\mathbf{q} = \mathbf{H}\boldsymbol{\theta} \quad (2.6)$$

where $\mathbf{H}$ represents the Radon transform, which performs a line integral along each LOR [139]. For the remainder of this thesis $\boldsymbol{\theta}$ represents the discrete voxel coefficients, and $\mathbf{q}$ represents the discrete modelled sinogram data. For the discrete object representation, the Radon transform sums the voxel values that lie along a given LOR. Each LOR corresponds to a row of $\mathbf{H}$ with the elements assigned a fractional value dependent on the the intersection length of the LOR through the voxel. Multiple ray tracing algorithms exist that differ in the chosen interpolation kernel; a popular choice is the Siddon algorithm [140].

In addition to the geometric Radon transform, other physical processes must also be included into the forward model to accurately represent the acquisition process of a real PET scanner. The above equation is therefore generalised:

$$\mathbf{q} = \mathbf{A}\boldsymbol{\theta} \quad (2.7)$$

where $\mathbf{A}$ is the system matrix, which models the acquisition process (including the Radon transform). The spatially variant physical sensitivity along each LOR of the PET scanner is modelled through element-wise scaling factors in the sinogram domain, referred to as normalisation factors (NFs). The normalisation factors themselves can be divided in a component based manner (component based normalisation) into time independent geometric factors and time dependent detector efficiencies [141, 142]. The detector efficiencies accounts for the acquisition (count statistic) dependent dead time and the temperature dependent detector gain. The geometric factors account for the variation in geometric sensitivity between the centre and the edge of the FOV, as well as block gaps between scintillator crystals. As shown in Fig. 2.4, the sensitivity of a (3D) PET system varies strongly along the axial direction, with the greatest proportion of emitted photons detected from a source located at the centre of the scanner. This axial variation in sensitivity can be reduced through adding collimators between each detector ring (septas). However, the majority of new PET systems do not use ring collimators, in order to maximise the overall number of detected counts. Detector sensitivity can also vary to a lesser extent in the transaxial direction, as shown in Fig. 2.4 (4). The attenuation of photons along the LOR are also modelled by element wise scaling factors in accordance to the Beer-Lambert law, producing attenuation factors (AFs) for each LOR. The attenuation factor is approximately spatially invariant with respect to the location of the annihilation event along the LOR. This is illustrated in the equation below

(and in Fig. 2.5), through considering the attenuation factor ($y(r)$) for a chosen LOR in 1D:

$$y(r) = \exp\left(-\int_0^R \mu(r)dr\right) \times \exp\left(-\int_R^L \mu(r)dr\right) = \exp\left(-\int_0^L \mu(r)dr\right) \quad (2.8)$$

where R is the location of the annihilation event along the LOR, μ is the attenuation coefficient, and L is the length of the LOR. This is in contrast to SPECT imaging where the attenuation factors along each LOR of the collimator are spatially (depth) dependent. For PET-CT imaging systems the μ -maps are generally determined through applying a bilinear or trilinear energy correction scaling to the CT image (originally in Hounsfield units), blurred to PET resolution. CT is useful for determining the μ -maps of PET photons as CT quantifies the attenuation of photons, albeit at a different energy. On the other hand determining the attenuation μ -map for PET-MR systems is less straightforward as MR signals are not related to the attenuation of photons. Accurate attenuation correction is of heightened importance for neurological PET-MR, due to the presence of the highly attenuating skull. Multiple solutions to the problem of PET-MR attenuation correction have been proposed in the literature, with varying success:

- **MR segmentation:** A predefined MR sequence (e.g. Dixon) is acquired, from which all voxels are individually assigned a tissue class and the corresponding predetermined attenuation coefficients [143, 144]. The assignment is made independent of any spatial correlation. Short echo time sequences can also be used such as UTE [145] and ZTE [146] to determine the location of bone in the patient, although currently such scans are impractical for whole body imaging.
- **MR-CT databases:** Large numbers of MR-CT pairs are used to form an atlas consisting of a template MR with a corresponding CT image (or attenuation labels). The template MR can then be wrapped to fit the acquired MR image, via a series of deformations. The same deformations are then applied to the CT image to attain the patient specific CT

image (attenuation map) [147–150]. Recent efforts have sought to exploit population databases through deep learning based methodologies that synthesise the corresponding CT image [151–153].

- Emission data: The attenuation map can be derived simultaneously with the emission image, through alternately estimating each until consistent, generally referred to as MLAA [154–156]. Such methods are highly underdetermined and prone to cross-talk, and therefore generally additional constraints are added to the dataset through penalty terms and TOF information [157]. Through use of TOF information, the attenuation map can be determined up to a constant scaling factor [158]. Further improvements can be made through the inclusion of prior anatomical information [159].

The reciprocal form of both AF and NF can also be used for data correction, as opposed to system modelling. The data correction versions are called attenuation correction factors (ACFs) and normalisation correction factors (NCFs). Returning to the system modelling context, both the AFs and NFs can be represented as diagonal matrices $\mathbf{Y}$ and $\mathbf{Z}$ respectively. Resolution degrading factors, which lead to blurring and PVEs, can also be included (individually) into the system matrix either in projection space (inter-crystal scatter) or image space (positron range). The dominant resolution degrading factor in clinical PET scanners is due to photon acollinearity, which can be approximated as depth-independent and included in the projection space detector blurring matrix. In practice, these individual factors can be lumped together and approximated by an image space PSF ($\mathbf{P}$). Using an image space PSF also has the added benefit of being able to model spatially-variant blurring effects. This disregards the spatially variant nature of a PET system's resolution. The resolution degrading PET acquisition process can also be considered in the frequency domain, whereby the modulation transfer function is non uniform (and finite), resulting in the high frequency components being poorly acquired. Applying PSF modelling seeks to amplify the high frequency components of the modulation transfer function that were diminished during

acquisition. Including PSF modelling into the system matrix leads to images with enhanced hot lesion detection rate and regional quantification [160, 161], making it clinically popular. However, the use of PSF modelling makes the reconstruction process underdetermined, with multiple solutions existing that fit with the data. Depending on the size and shape of the PSF kernel this can result in non physical Gibb's like ringing artefacts in unregularised image reconstruction [162, 163]. The overall system matrix ($\mathbf{A}$) can now be factorised into the individual components discussed above [164]:

$$\mathbf{q} = \mathbf{YZHP}\boldsymbol{\theta} \quad (2.9)$$

The random and scatter coincidence events also affect the measured count distribution, distorting it from the true coincidence count distribution. The random and scatter coincidence events are typically included into the modelled projection data as an additive background vector $\mathbf{b}$:

$$\mathbf{q} = \mathbf{YZHP}\boldsymbol{\theta} + \mathbf{b} \quad (2.10)$$

The scattered counts term could be included into the system matrix as it is linear in $\boldsymbol{\theta}$ [127]. The scattered counts can be estimated using single scatter simulation, multiple energy windows or Monte Carlo simulation, in addition to other methods [165]. Energy windowing is a prominent method in SPECT imaging for scatter correction, however is of limited use in PET for the following reasons. The high energy of PET photons (511keV) means the main mode of interaction between the photon and scintillation material is Compton scattering. A narrow energy window would eliminate many true coincidences that underwent scatter within the scintillation crystal. This is compounded by the relatively poor energy resolution of scintillation materials used in PET e.g. (LSO) in comparison to SPECT (NaI). Therefore, a wide energy window is generally used in PET imaging, combined with iterative single scatter estimation scaled using tail fitting [166]. The random counts can be estimated using

either the delay window [167] or singles method ($2\tau S_1 S_2$, where S_1 is the singles count rate at detector 1). A variety of iterative reconstruction algorithms can be used to determine the modelled projection data ($\mathbf{q}$) and the object parameter distribution ($\boldsymbol{\theta}$) from the measured projection data ($\mathbf{m}$).

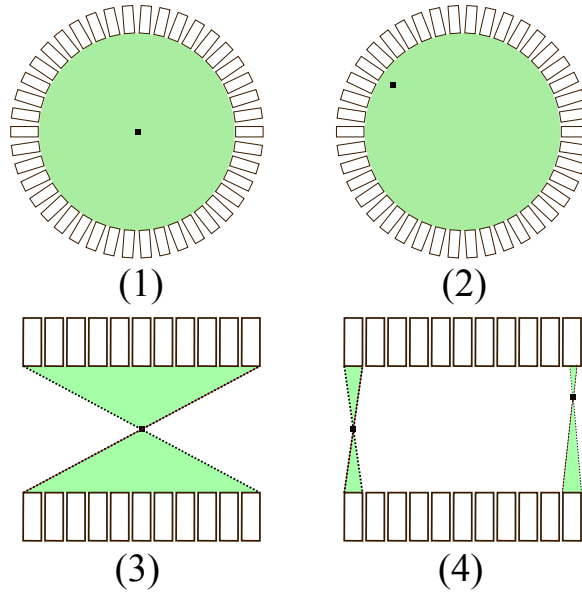


Fig. 2.4 (1) Radial sensitivity of PET scanner at centre of FOV. (2) Radial sensitivity of PET scanner at edge of FOV. (3) Axial sensitivity of PET scanner at centre of FOV. (4) Axial sensitivity of PET scanner at edge of FOV.

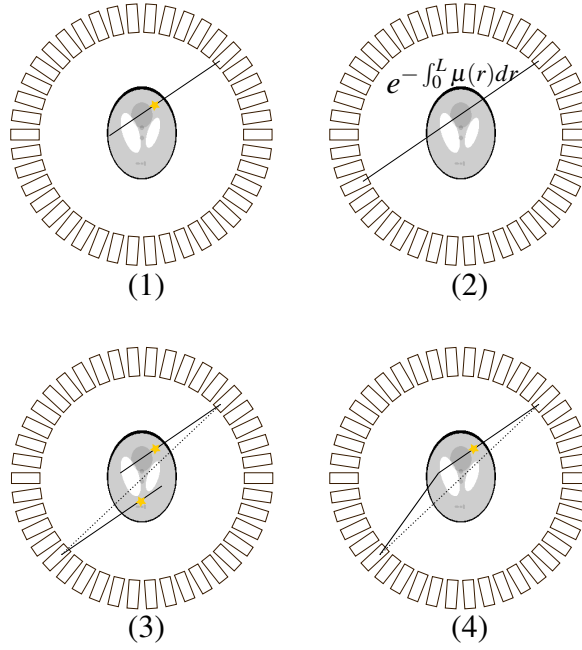


Fig. 2.5 (1 and 2) Schematic of attenuation along LOR, where L is the length of the LOR. (3) Schematic of random coincidence events. (4) Schematic of scattered coincidence events.

2.4 Statistically Based Iterative Reconstruction

Statistical PET image reconstruction seeks to find the set of parameters that maximise a statistically derived objective function, to provide an accurate object representation. Due to the discrete count like nature of radioactive decay the corresponding projection data (measured counts) is well modelled by the Poisson probability distribution:

$$\Pr(m_i|\hat{m}_i) = \frac{(\hat{m}_i)^{m_i} e^{-\hat{m}_i}}{m_i!} \quad (2.11)$$

where $\hat{m}_i$ is the expected number of counts in projection bin i , and m_i is the number of counts in projection bin i for a single noise realisation. This differs from the previously used term $\mathbf{q}$, which represents the modelled expected number of counts, dependent on the measured data. The Poisson probability distribution states the probability of obtaining a random number of counts m_i , given an expected number of counts $\hat{m}_i$. This probability distribution can be easily

extended over multiple projection bins, to find the total probability:

$$\Pr(\mathbf{m}|\hat{\mathbf{m}}) = \prod_i^{N_p} \frac{(\hat{m}_i)^{m_i} e^{-\hat{m}_i}}{m_i!} \quad (2.12)$$

For the inverse problem of PET imaging the single noise realisation is known, from which the expected projection data needs to be estimated ($\mathbf{q}$). As previously shown for the discrete voxel basis functions, the modelled projection data is linearly related to the voxel coefficients ($\boldsymbol{\theta}$). The Poisson likelihood for a parameter vector $\boldsymbol{\theta}$ and a measured projection dataset $\mathbf{m}$ is:

$$L(\boldsymbol{\theta}|\mathbf{m}) = \prod_i^{N_p} \frac{\left(\sum_j^{N_v} a_{ij}\theta_j + b_i\right)^{m_i} \exp\left(-\sum_j^{N_v} a_{ij}\theta_j - b_i\right)}{m_i!} \quad (2.13)$$

The natural logarithm is generally applied to the objective function as the location of the function's stationary points remains unchanged and it helps simplify the algebraic manipulation.

$$\Omega^{ML}(\boldsymbol{\theta}|\mathbf{m}) = \ln L(\boldsymbol{\theta}|\mathbf{m}) = \sum_i^{N_p} m_i \ln \left(\sum_j^{N_v} a_{ij}\theta_j + b_i \right) - \sum_j^{N_v} a_{ij}\theta_j - b_i \quad (2.14)$$

The $m_i!$ term is ignored, due to its independence with respect to $\boldsymbol{\theta}$. Using this objective function a set of voxel coefficients are sought that maximises the Poisson likelihood for a fixed measured projection dataset:

$$\boldsymbol{\theta}^{ML} = \arg \max_{\boldsymbol{\theta}} \sum_i^{N_p} m_i \ln q_i(\boldsymbol{\theta}) - q_i(\boldsymbol{\theta}) \quad (2.15)$$

The objective function can be re-expressed in the same form as the (negative) Kullback-Leiber distance plus a constant offset. Multiple alternative objective functions also exist, such as the (weighted) least squares objective function [168, 169]. However, the least squares objective function would incorrectly model the measured projection data using a Gaussian distribution.

An alternative to statistically based reconstruction methods are analytically based methods, which seek to invert the mathematical transform that relates the measurement and object

domains. The most popular analytical method is filtered back projection with reprojection to account for the missing oblique data [170, 171]. Data modelling as described in §2.3 cannot be incorporated into such methods, therefore projected data is generally precorrected to account for background counts and attenuation factors. After precorrection the data is no longer Poisson distributed, and has increased variance. A faster alternative uses 2D back projection for direct sinograms produced using Fourier rebinning (FORE) [172, 173]. FORE into direction sinograms can be followed by both analytical (2D radon) and statistical reconstruction methods [174]. Despite the early prominence of analytical based methods, the use of statistical iterative methods is now almost universal due to their advantages of a statistical noise model, system modelling and physical (i.e. non-negativity) constraints.

2.5 Inclusion of Prior Information

The maximum likelihood objective function can be extended to include prior knowledge about the object's parameter distribution (voxel coefficients). This can be achieved through re-expressing the objective function in terms of Bayes' theorem:

$$\Pr(\boldsymbol{\theta}|\mathbf{m}) = \frac{\Pr(\mathbf{m}|\boldsymbol{\theta})\Pr(\boldsymbol{\theta})}{\Pr(\mathbf{m})} \quad (2.16)$$

Again the natural logarithm is applied to simplify the algebraic manipulation, and the $\Pr(\mathbf{m})$ term can be disregarded. The resultant formulation can be used to describe the maximum *a posteriori* (MAP) objective function:

$$\Omega^{MAP}(\boldsymbol{\theta}|\mathbf{m}) = \Omega^{ML}(\boldsymbol{\theta}|\mathbf{m}) + \ln(\Pr(\boldsymbol{\theta})) \quad (2.17)$$

This MAP objective function could also be viewed as a penalised maximum likelihood (PML) objective function as discussed in [175], although the interpretation of individual elements

may change accordingly. The Gibb's energy representation can be used to describe the prior term:

$$\Pr(\boldsymbol{\theta}) = \frac{\exp(-\beta R(\boldsymbol{\theta}))}{J} \quad (2.18)$$

which when substituted in Eq. 2.17 gives:

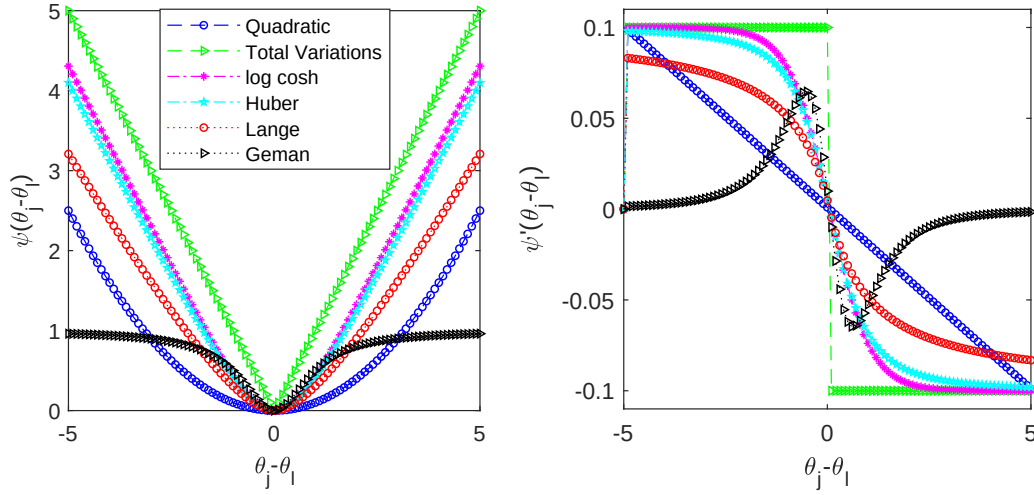
$$\Omega^{MAP}(\boldsymbol{\theta}|\mathbf{m}) = \Omega^{ML}(\boldsymbol{\theta}|\mathbf{m}) - \beta R(\boldsymbol{\theta}) \quad (2.19)$$

The above equation can be separated into a data fidelity term ($\Omega^{ML}(\boldsymbol{\theta}|\mathbf{m})$) and a regularising term ($R(\boldsymbol{\theta})$), with the relative trade-off between the two terms determined by the hyperparameter (β). Determining the appropriate value of β for a given dataset is still an area of active research. J is a normalising factor called the partition function, that is independent of $\boldsymbol{\theta}$ and can be disregarded. A hugely popular family of priors are based on the Markov random fields (MRF) model. Such priors determine the statistical distribution of a given voxel value based on its similarity with neighbouring voxels (cliques):

$$R(\boldsymbol{\theta}) = \sum_j^{N_v} \sum_{l \in \eta_j} w_{jl} \psi(\theta_j - \theta_l) \quad (2.20)$$

where η_j are the neighbouring voxels of voxel j , w_{jl} is the weighting factor between voxels j and l , and ψ is the inter-voxel potential function. A wide variety of potential functions have been developed in the PET reconstruction literature including relative difference [176], Huber [177], Lange [178], log-cosh [179], Geman [180], Kaipio [181] and total variation [182], to name but a few. A subset of these potential functions are shown in Fig. 2.6. The majority of such potential functions aim to suppress small differences in voxel values (likely due to noise), but allow large differences (likely due to regional uptake boundaries). A spatially variant level of regularisation ($\beta(\mathbf{r})$) could also be employed to retain fine image detail whilst still encouraging smoothness across uniform regions [183, 184]. The MRF prior formulation described above can also be extended to include anatomical information. An alternative

family of priors also exist that assumes spatial independence between voxels. These methods encourage similarity to a specified mean image on a voxel-wise basis, using for example the independent Gaussian model [185], or maximum entropy criterion [186].



Penalty Name	Formula
Quadratic	$(\theta_j - \theta_l)^2$
Total Variation	$ \theta_j - \theta_l $
log-cosh	$\log \cosh \left(\frac{\theta_j - \theta_l}{\delta} \right)$
Huber	$\delta^2 \left(\sqrt{1 + \left(\frac{\theta_j - \theta_l}{\delta} \right)^2} - 1 \right)$
Lange	$\delta \left(\frac{ \theta_j - \theta_l }{\delta} - \log \left(1 + \frac{ \theta_j - \theta_l }{\delta} \right) \right)$
Geman	$\frac{(\theta_j - \theta_l)^2}{\delta^2 + (\theta_j - \theta_l)^2}$

Fig. 2.6 Summary of a subset of potential functions, their gradient, and their formulation, as defined in the PET reconstruction literature. The hyperparameter δ has been assigned a value of 1, for all methods illustrated.

Prior information regarding the distribution of the object's voxel coefficients ($\Pr(\boldsymbol{\theta})$) can also be incorporated into the objective function through reparameterisation. As previously described in §2.2 the object's parameters (in this case voxel coefficients) $\boldsymbol{\theta}$ can be rewritten in terms of an alternative set of spatial basis functions ($\mathbf{K}$) and coefficients ($\boldsymbol{\alpha}$).

$$\boldsymbol{\theta} = \mathbf{K}\boldsymbol{\alpha} \quad (2.21)$$

which when substituted in Eq. 2.15 gives:

$$\boldsymbol{\alpha}^{ML} = \arg \max_{\boldsymbol{\alpha}} \sum_i^{N_p} m_i \ln ([\mathbf{A}\mathbf{K}\boldsymbol{\alpha}]_i) - [\mathbf{A}\mathbf{K}\boldsymbol{\alpha}]_i \quad (2.22)$$

The structure of the basis functions combined with the non negativity constraint on $\boldsymbol{\alpha}$ (due to the EM algorithm) restricts the potential solution space for the modelled object's parameters ($\boldsymbol{\theta}$). In a similar manner to regularised methods, the ideal basis functions will restrict the inclusion of noise, whilst allowing accurate changes in uptake due to regional boundaries. It is still unclear whether regularisation or reparameterisation is the best means to include prior information into the objective function for PET reconstruction [187].

2.6 Iterative Solutions

2.6.1 Expectation Maximisation

For the maximum likelihood objective function (Ω^{ML}) proposed in Eq. 2.15 there exists no closed form solution for estimating the parameter distribution $\boldsymbol{\theta}^{ML}$. Instead an iterative solution can be found through employing the expectation maximisation (EM) algorithm, which belongs to the family of functional substitution or optimisation transfer algorithms [188]. The EM algorithm can be broken down into two parts. First, a surrogate function is defined using the conditional expectation of a non-observed complete dataset, to minorise

the original function. The complete data (ρ_{ij}) provides the number of events at voxel j that are detected along LOR i . The complete data Poisson log likelihood objective function is defined as:

$$\Omega^{ML}(\hat{\boldsymbol{\rho}}|\boldsymbol{\rho}) = \sum_{i,j} \rho_{ij} \log(\hat{\rho}_{ij}) - \hat{\rho}_{ij} - \log(\rho_{ij}!) \quad (2.23)$$

where $\hat{\rho}_{ij}$ is the expected complete data and once more the factorial term $(\rho_{ij}!)$ can be discarded. Unfortunately, the complete data is unknown. We therefore seek to calculate the conditional expectation of the complete data Poisson log likelihood, which acts as a surrogate function of the complete data Poisson log likelihood. The conditional expectation of the complete data Poisson log likelihood is:

$$E [\Omega^{ML}(\hat{\boldsymbol{\rho}}|\boldsymbol{\rho})|\boldsymbol{\theta}^n, \mathbf{m}] = \sum_{i,j} E [\rho_{ij}|\boldsymbol{\theta}^n, m_i] \log(\hat{\rho}_{ij}) - \hat{\rho}_{ij} \quad (2.24)$$

where the conditional expectation of the complete data $(E [\rho_{ij}|\boldsymbol{\theta}^n, m_i])$ is defined as:

$$E [\rho_{ij}|\mathbf{m}_i, \boldsymbol{\theta}^n] = \frac{a_{ij}\theta_j^n m_i}{\sum_c a_{ic}\theta_c^n + b_i} \quad (2.25)$$

and the expectation of the complete data $(\hat{\rho}_{ij})$ is defined as:

$$\hat{\rho}_{ij} = a_{ij}\theta_j \quad (2.26)$$

Second, the surrogate function is maximised with respect to $\boldsymbol{\theta}$, bringing $\boldsymbol{\theta}$ closer to the true solution [189–191]. The surrogate objective function is easily solvable by setting the partial derivatives of Eq. 2.24 to zero. The arrived at closed form update equation is separable with respect to θ_j :

$$\boldsymbol{\theta}_{EM}^{n+1} = \frac{\boldsymbol{\theta}^n}{\mathbf{A}^\top \mathbf{1}} \mathbf{A}^\top \left(\frac{\mathbf{m}}{\mathbf{A}\boldsymbol{\theta}^n + \mathbf{b}} \right) \quad (2.27)$$

where division is an element-wise operation. This algorithm solves the maximum likelihood objective function, using the expectation maximisation algorithm, and consequently is referred to as maximum likelihood expectation maximisation (MLEM). A step wise derivation of the EM update equation (Eq. 2.27) can be found in Appendix A, as shown by Matthews [192]. The EM algorithm can alternatively be described in the context of the minorise-maximisation (MM) framework for constructing an optimisation algorithm [188, 193]. MM differs slightly to EM methods in that the EM algorithm relies on conditional expectations (of complete data) to construct the surrogate function. Whereas, the MM framework generally uses convexity and inequality based argument. The EM algorithm applied to the ML objective function has favourable properties such as a simple update formula, a non-negative solution, and monotonic convergence to the ML estimate. Despite the notable benefits of MLEM its use clinically is limited due to its slow convergence rate. In practical applications an approximate version is employed where only a reduced number of projections are used for each image update. The division of projection bins into subsets provides the algorithm its name, ordered subset expectation maximisation (OSEM). The rapid increase in convergence speed of OSEM relative to MLEM is at the expense of the theoretical convergence guarantee and it is no longer a true maximum likelihood estimator. Efforts have been made in the literature to provide convergent versions of ordered subset methods [194, 195]. Although currently the overriding majority of clinical PET scans are reconstructed using OSEM.

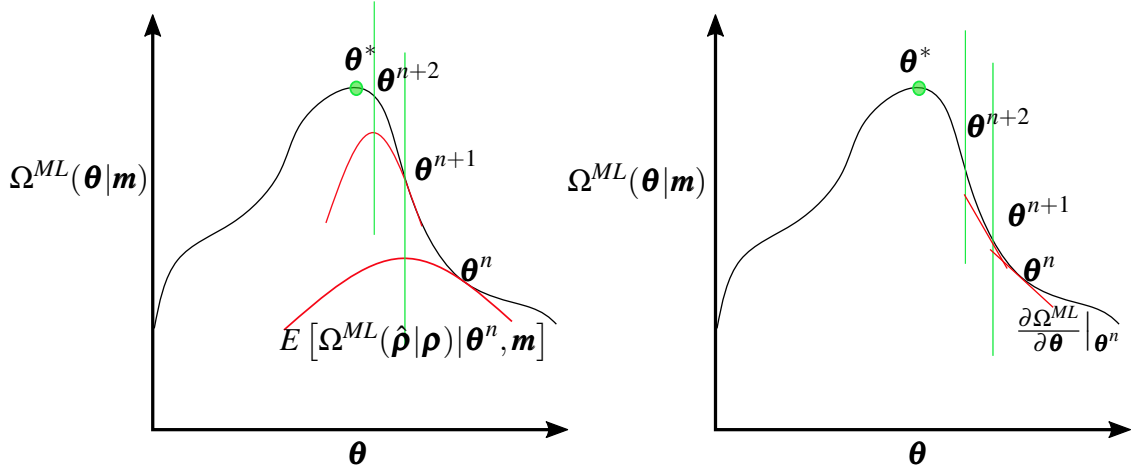


Fig. 2.7 Schematics showing the differences between the EM (left) and gradient ascent (right) algorithms used to solve an illustrative 1D objective function. For a 1D objective function, the coordinate ascent and gradient ascent algorithms will be equivalent.

For the ML objective function reparameterised in terms of alternative spatial basis functions a similar closed form update equation is found:

$$\alpha^{n+1} = \frac{\alpha^n}{K^T A^T \mathbf{1}} K^T A^T \left(\frac{m}{A K \alpha^n + b} \right) \quad (2.28)$$

The EM algorithm can also be employed for the MAP objective function, known as generalised EM (GEM). For the family of the MRF priors a closed form update solution cannot be found directly, due to the coupling between neighbouring voxels. An approximate solution is provided by the one step late algorithm [196], in which the object's parameter distribution (voxel coefficients) from the preceding iteration are used in the calculation of the prior derivatives. This allows a closed form update equation to be arrived at:

$$\theta^{n+1} = \frac{\theta^n}{A^T \mathbf{1} + \beta \frac{\partial R}{\partial \theta} | \theta^n} A^T \left(\frac{m}{A \theta^n + b} \right) \quad (2.29)$$

The simplicity of this algorithm has lead to its widespread use in the PET reconstruction research field [197–200]. However, the approximation employed in Eq. 2.29, means that the

algorithm no longer maximises the MAP objective function, and hence in general does not converge to a MAP solution. An additional weakness of this algorithm is its sensitivity to high β values, which can lead to the denominator tending to zero or negative values (i.e. a non-physical solution). For specific prior formulations, surrogate priors forms can be found that allow a separable closed form update solution to be found. The first example proposed in the PET literature was for the weighted MRF quadratic prior:

$$R(\boldsymbol{\theta}) = \frac{1}{4} \sum_j^{N_v} \sum_{l \in \eta_j} w_{jl} (\theta_j - \theta_l)^2 \quad (2.30)$$

The corresponding surrogate prior form for the above prior is:

$$R(\boldsymbol{\theta} | \boldsymbol{\theta}^n) = \frac{1}{2} \sum_j^{N_v} (\theta_j - \theta_j^{REG})^2 \sum_{l \in \eta_j} w_{jl} \quad (2.31)$$

$$\theta_j^{REG} = \frac{1}{2 \sum_{k \in \eta_j} w_{jk}} \sum_{l \in \eta_j} w_{jl} (\theta_j^n + \theta_l^n) \quad (2.32)$$

This surrogate prior form finds the voxel wise difference between the current update image (θ_j) and a weighted smoothed update image (θ_j^{REG}). Substituting the surrogate prior into Eq. 2.19, in addition to the surrogate function used for the ML objective function (Eq. 2.24) and maximising this overall surrogate objective function leads to the following update equation:

$$\theta_j^{n+1} = \frac{2B_j^n}{D_j^n + \sqrt{(D_j^n)^2 + 4C_j B_j^n}} \quad (2.33)$$

$$D_j^n = \sum_i^{N_p} a_{ij} - \frac{\beta}{2} \sum_{l \in \eta_j} w_{jl} (\theta_j^n + \theta_l^n) \quad (2.34)$$

$$B_j^n = \theta_{j,EM}^{n+1} \sum_j^{N_v} a_{ij} \quad (2.35)$$

$$C_j = \beta \sum_{l \in \eta_j} w_{jl} \quad (2.36)$$

2.6.2 Gradient-Based and Coordinate-Ascent Algorithms

A popular means for solving parameter optimisation problems generally is the use of gradient-based algorithms. The gradient of the objective function with respect to the object's parameter distribution vector, is used as the search direction, along which a line search can be performed to find the appropriate step size. Examples of such algorithms include steepest ascent, although this suffers from slow convergence. The conjugate gradient algorithm [168] can ameliorate slow convergence issues, especially in combination with preconditioning [201]. The pre-conditioned conjugate gradient algorithm is presented below for a least squares objective function (in accordance with [202]), where $\mathbf{M}$ is the pre-conditioner and the step size λ can be calculated analytically:

$$\Omega^{LS} = \|\mathbf{A}\boldsymbol{\theta} - \mathbf{m}\|_2^2 \quad (2.37)$$

$$\mathbf{g}^n = -\nabla \Omega^{LS}(\boldsymbol{\theta}^n) \quad (2.38)$$

$$\mathbf{p}^n = \mathbf{M}\mathbf{g}^n \quad (2.39)$$

$$\gamma^n = \begin{cases} 0, & n = 0 \\ \frac{\langle \mathbf{g}^n, \mathbf{p}^n \rangle}{\langle \mathbf{g}^{n-1}, \mathbf{p}^{n-1} \rangle}, & n > 0 \end{cases} \quad (2.40)$$

$$\mathbf{d}^n = \mathbf{p}^n + \gamma^n \mathbf{d}^{n-1} \quad (2.41)$$

$$\lambda^n = \frac{\langle \mathbf{d}^n, \mathbf{g}^n \rangle}{\langle \mathbf{d}^n, \mathbf{A}^\top \mathbf{A} \mathbf{d}^n \rangle} \quad (2.42)$$

$$\boldsymbol{\theta}^{n+1} = \boldsymbol{\theta}^n + \lambda^n \mathbf{d}^n \quad (2.43)$$

A considerable limitation of these gradient-based methods is the lack of an inbuilt non-negativity constraint, although solutions do exist [168, 203, 204]. The EM algorithm can also be re-written as gradient ascent method, allowing faster convergence through using an adaptive step size [205].

Coordinate-ascent (CA) algorithms are conceptually simple and straightforward to implement, whereby each object parameter (voxel coefficient) is updated in turn to maximise the overall objective function [169, 206]. The 1D optimisation along each variable allows the non-negativity constraint to be easily enforced. These methods also present fast convergence rates if provided with a good initial estimate of the image. However, the 1D maximisation of the objective function, with respect to each voxel, leads to numerous costly forward back projects being required. Therefore, alternative approximations for the log likelihood objective function have been proposed in the literature [207]. Recently, these methods have dipped in popularity, possibly due to the increase parallelism available with modern day GPUs, which can lead to rapid speed improvements in competing algorithms. A generalisation of CA to grouped-coordinate ascent enables greater parallelism and also results in faster convergence rates, in comparison to conventional CA methods [208]. Despite the progress made via statistical based reconstruction methods, PET image quality remains inferior to anatomical based imaging modalities. Further improvements to PET image quality are still needed to maximise the clinical efficacy of such scans.

2.7 PET Image Quality

Assessment of PET image quality can be evaluated in terms of image properties, and/or human observer tasks [209]. Despite the image properties (or observer tasks) being unique, they are generally interlinked, with improvements in one normally coming at the expense of another. The most important image properties are spatial resolution, image contrast and stochastic noise, due to their direct impact on regional quantification and lesion detection.

These image quality metrics are dependent on both the acquisition and reconstruction of PET data. The physical limitations of current PET systems degrade the acquired projection data, which in turn affects the resultant image quality. The reconstruction process (algorithm choice, hyperparameter selection) also affects the resultant image quality. Through the appropriate choice of reconstruction algorithm, possible limitations in the acquired data can be overcome to some extent.

The spatial resolution of the PET system is limited by the trade-off between particular aspects of the scanner design and PET physics, as discussed in §2.1. For instance, the spatial resolution can be further improved through a reduction in crystal size, but this will be at the expense of photon sensitivity. Therefore, the physical parameters of the PET system are chosen to provide the best trade-off in terms of overall image quality. The resultant PVEs due to the limited PET resolution negatively affects quantification and detection of small lesions (in particular necrotic tumours). Grey matter is also susceptible to PVEs due to the relatively thin cerebral cortex, which is only a few millilitres across in some dimensions. Due to the issues surrounding unregularised PSF modelling alternative solutions are required for reducing the PVEs.

Stochastic noise is due to the statistical variation in nuclear decay, which obeys a Poisson distribution. The nature of this distribution requires the expected number of events to equal the variance. Therefore, the signal to noise ratio (SNR) along each LOR is proportional to the square root of a source's activity.

$$\text{SNR} = \frac{\hat{m}}{\sigma_m} \propto \sqrt{\hat{m}} \quad (2.44)$$

The dose level a clinical patient is subject too tries to balance the level of the noise present in the reconstructed image with the level of patient exposure. For low activity levels, the variability between the number of measured counts and the expected number of counts leads to notable high frequency noise. The limited sensitivity of the PET system further compounds

the images SNR. As shown in §2.1 the sensitivity of the PET system is dependent on the geometric layout of detectors, scintillator depth, detector quantum efficiency and dead time. The use of statistical reconstruction methods combined with early termination is currently used as a clinical solution to minimise the visible noise present in the reconstructed image. As previously discussed, statistical methods allow simple inclusion of the system model, background counts, and the correct statistical model. However, such methods can lead to variation in image quality depending on the some what arbitrary choice of termination criteria. Low iteration numbers (generally chosen in practise) lead to poor recovery of high spatial frequencies [210], poor contrast and spatially variant convergence. However, the level of high frequency noise present in the reconstructed image is reduced. Conversely, high iteration numbers (providing a near convergent solution), ensure all measurable spatial frequencies are accurately recovered uniformly across the image. However, this is at the expense of increased random noise in the image. Most clinical images also undergo post-reconstruction smoothing using a Gaussian kernel with a FWHM of around 4 mm. This reduces the visible impact of noise in the reconstructed image, but once more at the expense of spatial resolution. A desirable reconstruction algorithm would enable improvements in spatial resolution, PVEs and image noise.

Human observer performance studies are an alternative, possibly more clinically pertinent means for assessing image quality. In such studies, multiple images are assessed in relation to a specific task, whereby the images presented have used different imaging systems, conditions, or reconstruction algorithms. Two popular means for undertaking such task-based observer studies include receiver operator characteristics (ROC) studies and the use of the Hotelling observer [211–213]. For the comparison of multiple reconstruction methods, using human observers (radiologists in the case of PET-MR images) would be very time consuming given the evaluation required of multiple patients and reconstruction methodologies. Ideally, each scan would also be read by multiple radiologists to account for inter-person variation in the

reading of PET scans. Automating the task-based observer process is still an area of active research within medical imaging, even for the relatively simple task of tumour detection [214, 215].

In general the image quality of PET images remains inferior to anatomical based imaging modalities. Further advancements in the PET reconstruction process are required to improve the clinical efficacy of PET imaging. With the recent development of simultaneous PET-MR scanners, PET datasets combined with anatomical MR will become more frequent. This could encourage the use of MR-informed reconstruction methods clinically. The inclusion of anatomical information into the PET reconstruction process has been shown to produce images with superior noise reduction and spatial resolution.

2.8 Anatomically-Informed PET Reconstruction

As previously discussed in §2.7, there are two dominant factors that degrade the reconstructed PET image quality: 1) high Poisson noise in the PET data, a consequence of the limited injected dose and limited sensitivity of current clinical PET scanners; 2) poor spatial resolution, due to the finite detector size, photon acollinearity and positron range (amongst other factors), which leads to PVEs. Utilising anatomical (MR) information within the reconstruction process, could be one solution to addressing these issues. High resolution autoradiography studies have shown a strong correlation between neuroreceptor mapping PET radiotracers and MR derived anatomical structure [216]. However a direct structural correlation between PET and MR is not always explicit, due to PET's relatively poor spatial resolution, misregistration errors, high noise levels present in PET images, and possibly most importantly the choice of radiotracer.

Anatomical imaging modalities, such as CT or MRI, produce images with superior SNR and spatial resolution ($\approx 1\text{mm}$) in comparison to PET. The superior image quality of anatomical images, alongside their close structural correspondence to some radiotracer distributions

(in particular FDG), has encouraged many researchers to use anatomical information to overcome the limitations of PET [71, 58]. Anatomical information can be either included directly into the reconstruction process (anatomically-informed), or post reconstruction (generally through the application of a spatially variant filter). The anatomically informed reconstruction methods will be discussed first, in chronological order. These methods can be subdivided into requiring segmentation, or being segmentation free, and are generally based on the regularisation or reparameterisation objective functions highlighted in §2.5.

Segmentation based models identify regional boundaries using the side anatomical information. Discontinuities (jumps) in the radiotracer distribution are then permitted in the vicinity of these boundaries, through assigning small weights to the regularisation term. One example is the weighted quadratic penalty, with the weights determined by the boundary locations [217, 218]. Alternatively, segmented boundary information can be used in the formation of line sites, which represent the likelihood of a boundary between different regions [180, 219, 220]. The cross correlation between PET and anatomically-derived line sites can also be used to reduce the likelihood of anatomically induced artefacts [62]. For ROIs specified directly from the segmented image, computationally simpler ML formulations can be used to determine the mean ROI parameter values [136]. This was then extended to a tissue composition model where all voxels of a given tissue type were modelled to be Gaussian distributed around a global tissue mean [221, 222]. Tissue segmentation maps can also be used to determine the fraction of tissue types (white matter, grey matter and CSF) that each voxel is comprised of. Anatomical based MAP formulations can be combined with this information to accurately determine the variation of grey matter intensity values within a specified region, whilst also suppressing noise [77]. More involved tissue weighted voxel composition formulations also exist in the literature [223], in which the regional support (and number of regions) are also estimated alongside the mean voxel distribution.

Segmentation-free priors provide a direct means for including anatomical information into the reconstruction process. This simplicity aids their recent popularity in comparison to their segmentation counterparts. The Bowsher method was the foremost example of such methods, and is still used as a benchmark for aspiring novel MAP reconstruction methods [224]. Given a corresponding MR image, the Bowsher method only assigns non-zero weights to neighbouring voxels of similar intensity. As such, voxels within the same anatomical region are encouraged to have similar intensity values, without the need for explicit segmentation. Multiple MAP methods that involve local neighbourhood spatial varying weights have since been proposed in the literature [225, 226]. Two other means for including side anatomical information into the reconstruction process is via parallel level sets (PLS) or information theoretic based priors. PLS determines the dot product between spatial gradients across the two modalities. Dissimilar gradients decrease the regularisation weighting factor, reducing the chances of artefacts in the reconstructed PET image [227–229]. Information theoretic priors seek a joint histogram that maximises the mutual information or joint entropy formulation [199, 230]. For two MR voxels with similar intensities to each other, their corresponding PET voxels will be encouraged to also have similar values to one another. Such methods do not enforce any spatial regularisation however, and can lead to spurious high intensity pixels. The mutual information prior has also been shown to lead to increased bias, resulting in the joint entropy formulation being generally preferred [231].

As discussed in §2.5, prior information can be used to constrain the reconstructed image via reparameterisation instead of regularisation [187]. Such methods were first introduced into the PET reconstruction community, through kernel derived basis functions [117, 81]. In a similar manner to preceding regularised priors [224, 226] structural information is extracted from the associated anatomical image on a local neighbourhood level. A more in depth background to the kernel method is presented in chapter 3, alongside applications within the field of tomographic reconstruction. Besides the kernel method, other neighbourhood

based priors such as population learnt dictionaries [232] and supervoxels [68] have been used in a reparameterisation fashion. Learnt dictionaries [233] have also been applied in a regularisation fashion.

MR-informed correction methods for partial volume effects and image noise can also be applied in a post reconstruction manner. Despite within reconstruction regularisation generally providing superior performance to post reconstruction filtering, post reconstruction methods provide clear practical benefits in that they can be applied offline to already reconstructed datasets [234]. Additional steps can also be taken to improve the performance of such methods, like data pre-whitening. A brief overview of these post reconstruction methods is now presented.

A wide variety of post-reconstruction denoising filters have been developed within image processing (e.g. Wiener, Hann, Metz [235]). Aside from image space, filtering can also be applied to the sinogram data pre reconstruction [236, 237], in the Fourier domain (to compensate for missing data) and to a sparse domain of choice [238, 239]. Image filtering has also been incorporated during the PET reconstruction process using the Metz filter [240]. Despite the popular use of filters post FBP reconstruction, the Gaussian filter remains dominant for iteratively reconstructed data. Some more involved filtering methods have shown promise in the image processing field, such as anisotropic diffusion [241, 242], non local means [243], and block matching 3D [244]. The non local means (NLM) algorithm seeks to determine each voxel value through a weighted mean of the neighbouring voxel values. The similarity weights are determined based on the Euclidean distance between the corresponding feature vectors. The NLM algorithm has demonstrated improved denoising power over Gaussian and anisotropic filtering methods, which can be heightened further through the inclusion of anatomical information [245]. The previously discussed NLM kernel method uses the same NLM denoising weights, but includes them during reconstruction. HYPR processing is a second popular anisotropic filter that has also been adapted for the

kernel method [246, 247], although only for dynamic PET datasets. Similar to the NLM filter, multiple existing filters have been readily adapted to include MR information, which in turn generally improves the PVC power of such methods [248, 249].

Through correctly modelling the finite PET system resolution and the tissue fraction effect, the impact of PVEs can be minimised. The need for PVC in PET was originally addressed for phantoms of known size and distribution, through the application of correction factors dependent on the object size and shape [250]. Multiple extensions to this work have subsequently been developed (discussed here solely in the context of neurological PET). Voxel wise PVC has been implemented through the application of recovery coefficients using both a two and three compartmental tissue model [251, 252]. ROI based PVC has also been developed through the geometric transfer matrix (GTM), which is applied to parcellated PET datasets. The GTM links the extent of the MR defined ROIs to the imaged (blurred) ROIs. The matrix can then be solved through single value decomposition [253]. Both image space and sinogram space models of the PSF have been investigated [254]. It should be noted that the GTM provides ROI based PVC analysis whereas the compartment methods listed provide voxel based PVC analysis. Overviews of such PVC methods are provided in the literature [255, 253, 34]. The above PVC methods have been included in widely used neuroimaging software, such as FreeSurfer [256].

2.9 Synergistic PET-MR Reconstruction

An extension of MR-informed PET reconstruction algorithms, is to reconstruct the PET and MR datasets simultaneously using a shared objective function. Such methods enable the use of PET information to regularise the MR image, as well as the more conventional use of MR information to regularise the PET image. One advantage of such methods is the ability to improve the reconstruction of MR images with high acceleration factors, and PET images suffering from poor count statistics. These techniques are of particular relevance, due to the

increasing use of simultaneous PET-MR scanners. The generalised objective function for simultaneous PET-MR reconstruction is as follows [228]

$$\Omega(\boldsymbol{\theta}, \mathbf{x} | \mathbf{m}, \mathbf{t}) = \Omega^{PET}(\boldsymbol{\theta} | \mathbf{m}) + \Omega^{MR}(\mathbf{x} | \mathbf{t}) - \beta R(\boldsymbol{\theta}, \mathbf{x}) \quad (2.45)$$

The Ω^{PET} objective function is a data fidelity term such as the maximum likelihood. The noise model for MR data can be well approximated as Gaussian, and therefore a least squares model is generally chosen for the MR data fidelity objective function (Ω^{MR}), where $\mathbf{t}$ is the measured k space data. The final term is a penalty term between the reconstructed PET and MR image, to encourage similarity across the images. Most penalty formulations are based on matching image gradients, including PLS [228], and total variation based priors [257–259]. Varying weight values within a local neighbourhood have also been proposed as a penalty term [260], where an alternating optimisation algorithm was used that updated the MR and PET images in turn.

2.10 Deep learning in PET Reconstruction

Numerous forms of regularising penalty functions have been suggested in the reconstruction literature, ranging from enforcing certain image space attributes [261–263, 116], similarity with anatomical information [199, 230, 225] (a subset of which are discussed in chapter 5), and more recently learnt regularisers [264, 232]. Despite the success of well-established priors that use "hand crafted" filters, the appropriate choice of prior information and β hyperparameter is difficult in practice and somewhat arbitrary. Learnt regularisation allows the choice of kernels to be determined from a representative training dataset. The combination of learnt kernels with nonlinear functions can be used to denoise the current PET image. This is achieved through a series of convolutions that extract features from the image, and thresholds that remove or select certain features. Through applying a series of such layers to

a given input image, high level features can be extracted. These principles underlie the use of convolutional neural networks (CNNs) for medical image reconstruction and denoising. Such networks have found success when applied to the medical imaging modalities of CT [110, 265, 266] and MRI [267–271]. This thesis focuses on the imaging modality of PET, for which deep learning methodologies have recently been included into the reconstruction process [272, 273, 110, 116, 274, 275, 112, 115, 276, 113]. Neural networks can be incorporated into the PET image reconstruction process at multiple stages: in lieu of an inverse to the imaging process (direct) e.g. [277, 278] ; in combination with a conventional reconstruction algorithm (unrolled) e.g. [276, 272]; and post-reconstruction e.g. [279, 280]. Within the field of PET reconstruction unrolled networks have shown superior performance to post-reconstruction networks [275]. Such networks are generally comprised of three main components:

1. Regularising network that takes the current update image and produces an appropriate prior distribution.
2. Data fidelity term that updates the current update image in accordance with the measured data, in this work the EM update is used.
3. Image fusion layer that combines the EM update image with the prior image distribution, using the MAP formulation [185].

In addition to the different reconstruction processes that a CNN can model, a second highly important variable is the chosen network architecture. A whole host of different network architectures have been proposed in the literature [83, 89, 108, 90, 91]. Choosing the optimal architecture for a given task and dataset is not straightforward. The U-Net architecture [108], originally proposed for semantic segmentation of medical images, has shown recent promise for PET image denoising both as a post-reconstruction [116] and unrolled [272] approach. A U-Net can be viewed as being comprised of two halves. The first half, consists of a number of "encoding" layers that reduces the spatial extent of the input image, whilst increasing the

number of features. Compressing the network in this manner encodes the most pertinent structural information into the latent feature space. The second half of the U-Net "decodes" the latent space features through increasing the spatial extent of the features, and combines them to construct the output image. A U-Net consists of multiple down-sampling (encoding) and up-sampling (decoding) layers, allowing it to extract structural information at multiple resolutions. Other popular architectures use residual blocks [90] (skip connections) and inception modules [91].

The above description of CNNs requires a series of training images to supervise the learning of the internal network parameters. In natural image classification human observers can be used to assign classification labels, across a vast training database. For regression based image denoising networks, short and long exposure photographic images can be used to train the network. However, a prominent issue that adversely affects the use of deep learning methods within the PET imaging field, is the limited availability and size of standardised and labelled training and testing datasets. As a result of this, medical imaging neural networks are generally trained on small databases, e.g. 30 images (U-Net) [108]. By comparison, large scale neural networks that have been developed for image classification, such as GoogLeNet, are trained on datasets consisting of 1.2 million images [91]. Determining the necessary extent of the training database is made more difficult by the opaque nature of neural networks, and the non trivial task of identifying the features identified by the network, although recent efforts have been made to address this issue [281, 282]. A potential solution for the lack of accessible medical imaging datasets is the use of simulated or physical (e.g. NEMA [283] or Hoffman brain [284]) phantom datasets [112, 280]. This however, leads to difficulties owing to the differences in count levels, physical effects, and appearance between simulated and real datasets. Broadly speaking, this is due to the test dataset not lying within the domain of the training dataset, and hence the prior distribution does not correctly represent the test dataset. A recent development that could help address a lack of training data is the use of

noisy training labels (Noise2Noise). This concept was first addressed for denoising natural images, for a variety of noise models [118], and has recently been extended and applied to MRI and CT datasets [119]. However, no validation of this training framework has yet been demonstrated for PET datasets.

Chapter 3

MR-Informed Kernel Method for Low Dose PET Imaging

Within the field of PET imaging a key research objective is to reduce the required patient dose whilst still maintaining a clinically informative and robust scan. A typical PET scan (excluding CT) performed in the U.K. will impart an effective dose of 3-8 mSv on the patient, in addition to a lower (but more frequent) dose to the healthcare providers [285, 32, 286]. By comparison, the average U.K. background radiation dose is 2.4 mSv annually. Although the received patient dose is always sought to be minimised, the received patient dose is non-negligible. Reducing the patient dose below the current clinical range leads to low count measured projection data that is more susceptible to Poisson noise, and this may adversely effect the interpretation of the reconstructed image.

The ability to reduce the required injected dose for a clinical PET scan has numerous benefits. The ability to reduce patient dose can reduce the patients' cancer risk, the cost per scan, and would widen the demographic of healthy volunteers and patients who can undergo PET scans. Ultra-low PET doses 1% - 0.1% would expose the patient to radiation levels comparable to that of a transatlantic flight (0.08mSv) [287], potentially enabling PET imaging to be used for clinical screening in a similar manner to breast mammography [288].

Alternatively, through improving the reconstruction of low count datasets a reduction in patient scan time could be achieved, if the current injected dose level is maintained. The shorter scan duration would allow greater patient throughput, and correspondingly a reduced cost per scan. Previous work in the area of reduced dose PET imaging, has investigated the diagnostic quality of reduced dose PET data reconstructed using the conventional MLEM algorithm. This was undertaken for the applications of detecting oncological lesions [289], diagnosis of Alzheimer’s disease [290], and lung cancer screening [291]. These papers illustrate the clinical relevance of reduced dose PET scans, due to their capability of maintaining diagnostic utility while reducing patient dose levels. Inspired by the relative success shown by using the MLEM reconstruction algorithm for reduced count data and the denoising power of MR-informed reconstruction methods over MLEM [65, 230, 228, 199, 223], we investigate the potential low dose capabilities of the MR-informed kernel method [81][117].

In this chapter the re-parameterised MR-informed kernel method is re-introduced, with greater detail given to its machine learning origin. The denoising power of the kernel method is applied to reduced count data (as low as 1% of the original count level) and assessed through 3D simulation and resampled real data studies in comparison to post-smoothed MLEM. The basis of this work has previously been published during the period of the author’s PhD studies [292]. The diagnostic value of the reduced count kernel method was also evaluated by an expert neuroscientist within a retrospective pilot study consisting of a five person cohort, comprised of four patients with Alzheimer’s disease and one patient with temporal lobe epilepsy.

3.1 The MR-Informed Kernel Method

The kernel trick originates from the field of machine learning ([148]) where it is used, generally speaking, to determine the similarity (inner product) between a set of data-points in a higher dimensional (transformed) space, without the need to explicitly know the mapping

required, or undertake calculations in this higher dimensional space. The inner product between the data-points in this higher dimensional space is instead calculated through applying a particular choice of similarity kernel to the inner product of the data-points in the original, lower dimensional space. A popular use of the kernel trick is for extending the capabilities of support vector machines (SVMs), a supervised linear classifier, to non-linear classification [293, 294]. Kernel-SVM enables data that may not be linearly separable in its original vector space, to be transformed into an implicitly defined higher dimensional space in which the data may be separable through calculating the inner product in this higher dimensional space. The kernel trick was recently introduced into the PET reconstruction community [117] (referred to as the kernel method). In the original paper the kernel method was applied to dynamic PET data, using composite reconstructed images formed from sinograms summed over multiple time-points to devise the voxel-wise feature vectors. Following this nominal paper the kernel method has since been used in a wide range of tomographic reconstruction problems including MR-guided PET [81, 295, 296], fluorescence molecular tomography [297, 298], and diffuse optical tomography [299], in addition to dynamic PET [117, 247, 300, 301]. The kernel method can also be easily related to the nonlocal means filter (NLM) [243] used in (post-reconstruction) image denoising. Similarly to the kernel method, NLM filtering uses a k -nearest neighbourhood (k NN) search across a given feature space to determine the voxels that contribute to a given basis function. In this work we shall be focusing on the MR-informed kernel method, where the kernel trick is used to implicitly transform MR ($\mathbf{x}$) defined feature vectors into a higher dimensional representation and calculate their similarity (via the inner product) to reparameterise the reconstructed PET image ($\boldsymbol{\theta}$). The underlying theory of the MR-informed kernel method will now be presented for completeness.

3.2 Theory

A low-dimensional feature vector ($\mathbf{f}_j(\mathbf{x})$) can be extracted from an image (or group of images) for each voxel, where the index j runs over all voxels in the image. This feature vector describes a particular aspect (or feature) of the image. For the original application of the kernel method to dynamic PET data, the feature vector represented the long term temporal variation of each voxel. This was achieved through forming a temporal-series of PET images from composite sinograms (summed over multiple dynamic time frames). The feature vector for each voxel was the voxel intensities for all (composite) time-points. For the MR-informed kernel method, the feature vector is used to describe low-level structural information of the MR image, such as edges. These feature vectors are implemented as spatial patches of intensity values surrounding each voxel [302]. In both cases, it is assumed that the PET image (θ) (or images in the case of dynamic PET) can be represented as a non-linear function (Γ) of these low-dimensional features. This work shall focus on the MR-informed kernel method, where the PET image is described in terms of some non-linear function of the MR ($\mathbf{x}$) derived features (such as edges):

$$\theta_j = \Gamma(\mathbf{f}_j(\mathbf{x})) \quad (3.1)$$

Unfortunately, the form of this non-linear function is unknown. One possible solution would be to use a convolutional neural network to approximate this continuous function within a bounded domain, in accordance with the universal approximation theorem [303]. However, this requires separate training and testing datasets. A more elegant solution is to seek a higher dimensional representation of the feature vectors, which the PET image (θ) can be described as a linear combination of:

$$\theta_j = \boldsymbol{\omega}^\top \boldsymbol{\phi}(\mathbf{f}_j) \quad (3.2)$$

where $\boldsymbol{\phi}(\mathbf{f}_j)$ is the high-dimensional representation of the feature vectors, and $\boldsymbol{\omega}$ is the vector consisting of the weighting factors across each component of this high dimensional vector. This can be re-expressed in terms of summation notation:

$$\theta_j = \sum_{l=1}^N \omega_l \phi_l(\mathbf{f}_j) \quad (3.3)$$

Each coefficient of the voxel basis functions (θ_j) is represented by the same linear combination of the elements comprising its corresponding high-dimensional feature vector representation ($\boldsymbol{\phi}(\mathbf{f}_j)$). In other words $\boldsymbol{\omega}$ is the same for all j . The weighting factors ($\boldsymbol{\omega}$) can also be expressed in terms of these high-dimensional basis functions as:

$$\boldsymbol{\omega} = \sum_{l=1}^{N_v} \alpha_l \boldsymbol{\phi}(\mathbf{f}_l) \quad (3.4)$$

Substituting Eq. 3.4 into Eq. 3.2 leads to:

$$\theta_j = \sum_{l=1}^{N_v} \alpha_l \sum_{k=1}^N \phi_k(\mathbf{f}_l) \phi_k(\mathbf{f}_j) \quad (3.5)$$

$$\theta_j = \sum_{l=1}^{N_v} \alpha_l \kappa(\mathbf{f}_l, \mathbf{f}_j) \quad (3.6)$$

where the inner product between the higher dimensional representation of the feature vectors can found through the appropriate choice of kernel $\kappa(\mathbf{f}_l, \mathbf{f}_j) \equiv \boldsymbol{\phi}(\mathbf{f}_l)^\top \boldsymbol{\phi}(\mathbf{f}_j)$ applied to the feature vectors in the original space. Equation 3.6 can be expressed in terms of matrix notation as:

$$\boldsymbol{\theta} = \mathbf{K} \boldsymbol{\alpha} \quad (3.7)$$

The PET image ($\boldsymbol{\theta}$) can now be considered as a linear combination of the columns (or spatial basis functions) of $\mathbf{K}$. In the field of PET reconstruction, two main forms of kernel have been adopted, the Gaussian kernel [117] (radial basis function) and the HYPR kernel

[247, 304, 246]. The HYPR kernel however, has been evaluated purely for dynamic PET reconstruction, and therefore shall not be considered further in this work. The initial PET reconstruction paper on the kernel method [117], used a Gaussian kernel of the form:

$$\kappa(\mathbf{f}_l, \mathbf{f}_j) = \exp\left(\frac{-\|\mathbf{f}_l - \mathbf{f}_j\|_2^2}{2\sigma_f^2}\right) \quad (3.8)$$

where σ_f is a hyperparameter used to tune the width of the Gaussian similarity kernel. The Gaussian kernel has since been extended to include an additional spatial component [305], expressed as follows:

$$\kappa(\mathbf{f}_l, \mathbf{r}_l, \mathbf{f}_j, \mathbf{r}_j) = \exp\left(\frac{-\|\mathbf{f}_l - \mathbf{f}_j\|_2^2}{2\sigma_f^2}\right) \exp\left(\frac{-\|\mathbf{r}_l - \mathbf{r}_j\|_2^2}{2\sigma_r^2}\right) \quad (3.9)$$

where $\mathbf{r}_j$ is the spatial location of the voxel and σ_r is the corresponding hyperparameter controlling the Gaussian width. A schematic illustrating the formation of the kernel matrix, using the kernels presented in Eq. 3.8 and Eq. 3.9 is shown in Figure 3.1. As stated previously in §2.5 the kernel derived basis functions can be used to reparameterise the Poisson likelihood objective function. The resultant update formula for the basis function coefficients ($\boldsymbol{\alpha}$) is repeated here:

$$\boldsymbol{\alpha}^{n+1} = \frac{\boldsymbol{\alpha}^n}{\mathbf{K}^\top \mathbf{A}^\top \mathbf{1}} \mathbf{K}^\top \mathbf{A}^\top \left(\frac{\mathbf{m}}{\mathbf{A} \mathbf{K} \boldsymbol{\alpha}^n + \mathbf{b}} \right) \quad (3.10)$$

where $\mathbf{A}$ is the system matrix, $\mathbf{m}$ are the measured projection data, n is the iteration number, and $\mathbf{b}$ are the additive background randoms and scatters. The inclusion of the kernel derived basis functions into the iterative PET reconstruction algorithm shall be referred to as KEM throughout this work. The use of the EM algorithm constrains the kernel coefficients to be non-negative. The reconstructed PET image at each iteration is defined as:

$$\boldsymbol{\theta}^n = \mathbf{K} \boldsymbol{\alpha}^n \quad (3.11)$$

It should be noted that the kernel matrix (and all reparameterised spatial basis functions) must be row normalised, in an analogous manner to which column weights are normalised for penalty based functions. Without row normalisation, voxels that contribute to fewer basis functions due to the asymmetric nature k NN selection, will be poorly reconstructed.

In the previous implementations of the kernel method, the spatial basis functions that comprise the kernel matrix ($\mathbf{K}$) were generally sparsified to reduce the computational burden and to remove correlations between highly dissimilar voxels (in terms of spatial or feature space values). Firstly the voxels contributing to each basis function are spatially restricted to lie within a local neighbourhood of the central voxel (η_j).

$$K_{jl} = \begin{cases} \kappa(\mathbf{f}_j, \mathbf{f}_l), & \text{if } l \in \eta_j \\ 0, & \text{otherwise} \end{cases} \quad (3.12)$$

As such, elements of the kernel matrix that lie outside of the spatial neighbourhood are set to zero, removing spatially dissimilar voxels. Secondly, within this confined spatial neighbourhood only voxels that are one of the k nearest neighbours (k NN) of the central voxel j in feature space are selected to contribute to the given basis function. The proximity of voxels in feature space is calculated using the Euclidean distance between the two feature vectors.

$$K_{jl} = \begin{cases} \kappa(\mathbf{f}_j, \mathbf{f}_l), & \text{if } \mathbf{f}_l \in k\text{NN of } \mathbf{f}_j \\ 0, & \text{otherwise} \end{cases} \quad (3.13)$$

Further means of sparsifying the kernel matrix have also been suggested in [117], such as the ε -ball method, however they have not since been considered by other researchers. In addition to the normalisation of the kernel matrix, the feature vectors themselves are generally normalised by their standard deviations (SD_m) for each feature vector element m , to aid in

the consistent selection of the σ_f hyperparameter:

$$\bar{f}_{j,m} = \frac{f_{j,m}}{SD_m(\mathbf{f})} \quad (3.14)$$

where the standard deviation SD_m quantifies how each feature vector element m varies across all basis functions $j = 1 \rightarrow N_v$. The work presented in this chapter uses the Gaussian kernel to determine the kernel matrix, with patch based feature vectors extracted from the co-registered MR image. Each basis function is restricted to a local spatial neighbourhood and sparsified to only include contributions from the central voxel's k NNs in feature space. The kernel matrix is row normalised and the feature vectors are normalised by their standard deviations.

Although the kernel method was derived from the kernel trick predominantly used in machine-learning fields, it can also be simply explained in terms of building the PET image from a set of MR-informed spatial basis functions. These basis functions can be described as a local similarity matrix between the central voxel (of each neighbourhood) and the surrounding voxels. Quantifying the voxels similarity (assessed in terms of feature vectors) extracts structural information from the MR image. Therefore, a PET image constructed through a weighted sum of these MR-informed basis functions is confined to have a similar structure to that of the MR image. Examples of some of the spatial basis functions produced by the kernel method when applied to a real patient MR image are shown in Figure 3.2.

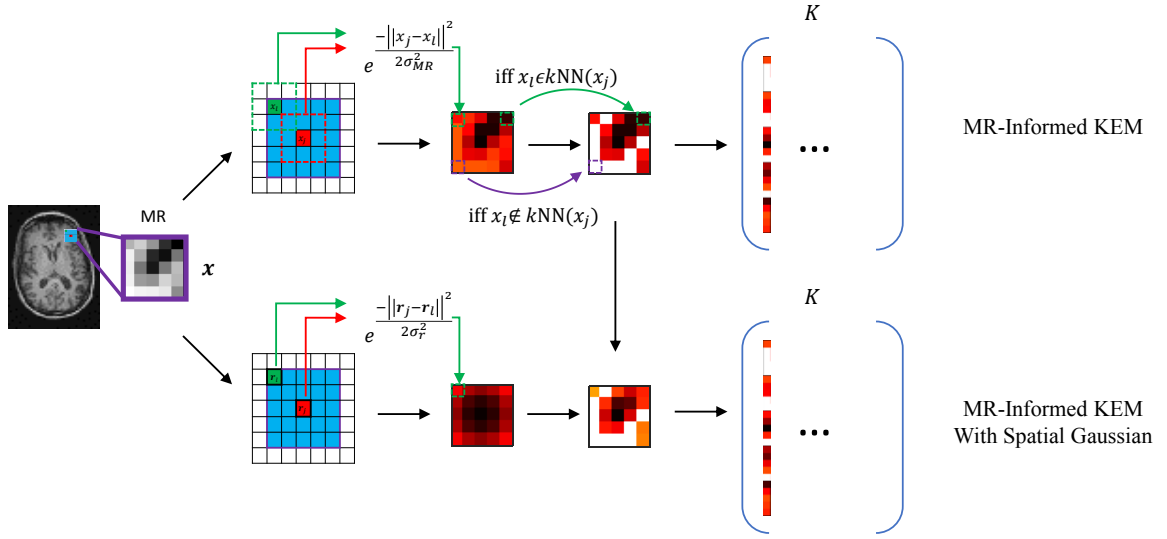


Fig. 3.1 Schematic showing the construction of the MR-informed basis function, with and without the inclusion of the spatial Gaussian term.

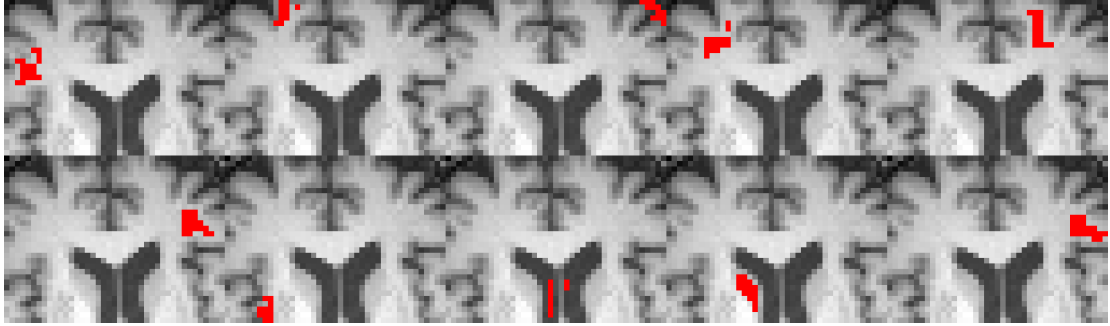


Fig. 3.2 Example basis functions extracted from the patient dataset described in §3.3.2 using the MR-informed kernel method. The hyperparameters used to construct the basis functions are those stated in Table 3.1, with the number of kNN set to 100 to make the basis functions easier to see.

3.3 Simulated Phantoms and Patient Data

3.3.1 2D and 3D Simulated Phantom

The segmented MR BrainWeb phantom was used as the basis of the PET activity phantom and attenuation μ -map for a 3D [^{18}F]FDG PET scan [306]. A corresponding T1w phantom

was also constructed from the same BrainWeb dataset. The PET phantom was constructed, projected into sinogram space, and reconstructed, using an internal software framework (APIRL) [131] designed to replicate the Siemens Biograph mMR scanner [39], and based on the Siddon projection algorithm [140]. As such, the 3D PET phantom had a grid size of $344 \times 344 \times 127$ and a voxel size of $2.08626 \times 2.08626 \times 2.03125 \text{mm}^3$. The grey and white matter regions of the PET activity phantom were assigned intensities to match the expected radio-tracer uptake of [^{18}F]FDG in neurological scans [64, 307, 308]. A high intensity spherical tumour was also added in the white matter of the frontal-left region, which was not present in the associated MR image (a profile through the tumour region is presented in Figure 3.11). The tumour region is smooth and has a peak voxel intensity of 0.34 arb. units, whereas in comparison the grey matter voxels have an intensity of 0.31 arb. units. The size and intensity of the tumour region was chosen to make it challenging to recover, but still visible in the 100% count MLEM reconstruction. The μ -map used three tissue classes of bone, soft tissue and air, with values assigned in accordance to scaled CT attenuation maps [198].

To simulate the data acquisition process the activity phantom was projected into span 11 sinograms. A 4.5 mm point spread function (PSF) was applied in the forward model, to simulate the overall effects of photon acollinearity, positron range, and finite crystal width [309]. The expected randoms sinograms were modelled as appropriately scaled uniform intensity sinograms. The expected scatter sinograms were simulated through applying a 1D smoothing kernel along each projection of the appropriately scaled activity sinograms. The ACFs were found in accordance to Eq. 2.8, through forward projecting the μ -map. The NCFs were found in accordance with component based normalisation, by calculating the product of scan specific (crystal efficiencies and detector dead times) and scan independent factors (geometric factors, crystal interference and axial effects) [142]. The scan specific factors were based on a typical mMR normalisation file, for the simulated dataset. The

summation of the expected activity, randoms and scatter sinograms are referred to as the expected prompts sinograms. The expected prompts sinograms are subsequently scaled with respect to the normalisation and attenuation factors. The expected prompts sinograms were also scaled to produce an expected count level of 5×10^8 . The expected randoms and scatter projection data were each simulated with a count level equal to 20% of the prompts. A Poisson distributed random vector was then obtained from the expected prompts sinograms to simulate the measured prompts sinograms. The acquisition process of the PET projection data was repeated at reduced count levels of 50%, 25%, 10%, 5%, and 1%, relative to the original prompts count level. Gaussian white noise was added to the T1 MR phantom. A transverse slice of the PET-MR phantom was used to construct a 2-D phantom, following the same process as described above. The prompts count level for the 2D dataset is 3.3×10^6 counts.

3.3.2 3D PET Patient Scan Dataset

An [^{18}F]FDG scan was obtained on a Siemens Biograph mMR PET-MR scanner, alongside various simultaneously acquired MR sequences. The patient was scanned as part of an Alzheimer's disease study, with similar structural patterns observed in the PET image and MR image. No tumour regions were observed in either image. The total prompt counts were 4.69×10^8 , with a scan duration of 23 min. The tracer activity at time of injection (81 min prior to start of image acquisition) was 229 MBq. In accordance to the dose conversion factors stated in [285], the patient's effective dose from the injection of PET radiotracers is 4.4mSv. A T1-weighted gradient echo (MPRAGE) sequence image was acquired simultaneously with the PET scan. The T1w MR image was down-sampled to PET resolution, from which the kernel basis functions were calculated. The MR-based patient μ -map was produced from the vendor supplied ultra-short echo (UTE) sequence in combination with a tissue probability map (Dixon) sequence. The UTE sequence allows bone structure to be located and assigned a

separate tissue class. The raw prompts data were randomly resampled (without replacement), producing resampled PET data from 1% to 50% of the original count level. The DICOM files were converted to interfile format through the JSRecon script. The ACFs were calculated directly from e7 tools produced attenuation file. The NCFs were calculated via component based normalisation, with the component factors provided by the e7 tools normalisation file. Randoms and scatters estimates were also obtained using e7 tools via the delayed coincidence method and the multi-slice single scatter simulations, respectively. The 2D single scatter estimate was extended to 3D through inverse single slice re-binning (iSSRB) [198]. Both the 2D simulated and 3D real datasets were used for determining the appropriate values of the KEM hyperparameters to minimise the difference between low count KEM reconstructions, and the full count MLEM reconstruction.

3.4 Hyperparameter Selection

The performance of the kernel method is highly dependent on the five following hyperparameters: neighbourhood size (N_j), patch size (p), number of k NN (k), spatial standard deviation (σ_r) and feature space standard deviations (σ_f). Thus, to find the optimal hyperparameters for the kernel method requires fully evaluating the 5-dimensional hyperparameter space, which is impractical for 3D datasets. The appropriate hyperparameter value is also dependent on the specific task in hand, therefore the hyperparameters values selected in this chapter based on minimising whole brain normalised root mean square error (NRMSE) for low count real patient datasets are unlikely to match the hyperparameters chosen for alternative tasks. Previous investigation into the optimal hyperparameter values for the MR-informed kernel method has been presented by Hutchcroft et al. [81], in which discrete values of each hyperparameter were evaluated, presenting results focusing on the neighbourhood size and the number of k -nearest neighbours in feature space. Similarly, work by Novosad and Reader [295] also evaluated a subset of the hyperparameters for discrete values, presenting

results for the neighbourhood size and feature space standard deviation. A greater level of hyperparameter evaluation was undertaken in the complementary work using the kernel method for fluorescence molecular tomography [298, 297] and diffuse optical tomography [299]. Due to the lack of consensus on the best value for each kernel hyperparameter, and the variation of these optimal hyperparameter values with application, an evaluation of all the kernel hyperparameters is presented in this work.

Two rounds of hyperparameter selection were undertaken in this work to try and obtain the best kernel hyperparameters for the application of using KEM on very low count data, such that the KEM reconstructed images replicate the results provided by the MLEM algorithm using the full count dataset. In the initial hyperparameter search, each kernel hyperparameter undergoes a repeated line-search across an appropriate range of values (as indicated in Table 3.1), with the other four hyperparameters values randomly selected. The line search is repeated 50 times for each hyperparameter. The hyperparameter values are assessed through comparing the resultant PET image (θ) to the ground truth (θ^{GT}) image in terms of NRMSE:

$$NRMSE = \sqrt{\frac{\sum_j^{N_v} (\theta_j - \theta_j^{GT})^2}{\sum_k^{N_v} (\theta_k^{GT})^2}} \times 100(\%) \quad (3.15)$$

The range and mean NRMSE values of each hyperparameter are presented in Fig. 3.3, for the hyperparameters values investigated. For this initial hyperparameters search a 2D simulated dataset is used, the specifications for which are presented in §3.3.1. The hyperparameters corresponding to the minimum NRMSE value along each hyperparameter curve are then selected as the starting kernel hyperparameters in the second hyperparameter selection round.

Due to the inherent differences between simulated and real datasets a second hyperparameter selection process is undertaken using PET patient scans (real data). The kernel method is now implemented in 3D and applied to a 1% count level real dataset, and compared via whole brain NRMSE to the MLEM reconstructed image of the 100% count dataset, with

both images undergoing post-smoothing (FWHM = 4mm). This level of post-reconstruction smoothing matches that generally used in clinical scans [289] and reduces the amount of noise present in the full count MLEM reconstruction. A simple 1D hyperparameter line search is used to select the optimal hyperparameters for the real dataset. The parameter values are initiated to those that minimised the NRMSE for the simulated dataset. Next, the algorithm selects one of the five hyperparameters under investigation (e.g. σ_f), undergoes a 1D line search to minimise the whole brain NRMSE with respect to this hyperparameter value, whilst keeping the other hyperparameters fixed (i.e. coordinate descent). The value of the selected hyperparameter is then updated, and the NRMSE minimisation step is repeated with respect to another selected hyperparameter. Random ordering of selecting which hyperparameters to update was used to reduce the susceptibility of the minimisation algorithm to local minima. The stopping criteria of the hyperparameter selection algorithm required all 5 hyperparameters to remained unchanged for a single update cycle (all 5 hyperparameters updated). After the chosen hyperparameters were found, the second hyperparameter selection process using real patient data was repeated, feeding a different 1% subset of the PET data to the kernel method. The chosen hyperparameters remained unchanged. The hyperparameter ranges investigated and the selected hyperparameters that minimised the whole brain NRMSE are shown in Table 3.1, for both the 2D simulated and 3D real datasets. Figure 3.3 shows the variation (specifically the maximum, minimum and mean) in NRMSE values with respect to each hyperparameter for the 2D simulated dataset. Figure 3.4 shows the 1D hyperparameter line curves using the real 3D data, for each hyperparameter investigated and for the final cycle of hyperparameter updates. Figure 3.5 illustrates the visual variation of the KEM reconstructed images with respect to each hyperparameter, applied to the PET patient scan. Each hyperparameter is shown at its minimum and maximum value, with the remaining hyperparameters fixed to the optimal 3D hyperparameters as stated in Table 3.1.

Due to the nature of the optimisation algorithm used, the selected hyperparameters may only correspond to a local minimum in NRMSE, and may not minimise the NRMSE globally across the entire hyperparameter space. As such, alternative hyperparameter combinations may exist in the hyperparameter ranges evaluated that may provide improved NRMSE reduction in comparison to those chosen. In addition, the choice of whole brain NRMSE as an error metric leads to the selection of hyperparameters that produce large detailed basis functions that accurately maintain shared PET-MR structures and suppress noise. However, these basis functions, although ideal for minimising whole brain NRMSE, may lead to (and are subsequently shown to) suppress features unique to the PET data (PET unique regions). It should also be noted that for practical implementation the neighbourhood size was limited to 7 (as opposed to 9), due to computational speed and memory size allocation. As shown in Figure 3.4, the difference in whole brain NRMSE between the neighbourhood sizes of 7 and 9 is minimal. The number of k -nearest neighbours was also varied with the count level of the PET data, providing a tuning hyperparameter for the amount of MR constraint required. In other words, higher count PET images were reconstructed using a lower k NN value.

Hyperparameters	2D Dataset		3D Dataset	
	Range	Selected	Range	Selected
k	10 – 60	50	10 – 100	100
N_j	3 – 11	11	3 – 9	9
p	1 – 7	1	1 – 5	3
σ_f	0.1 – 10	0.5	0.1 – 20	20
σ_r	1 – 50	10	1 – 20	20

Table 3.1 Kernel hyperparameters that minimise the whole brain NRMSE for the 2D simulated and 3D patient scan datasets, as evaluated in §3.4. N_j and p are sampled over the range stated for the odd numbers only. The k hyperparameter is sampled in intervals of 10. σ_f and σ_s are unevenly sampled in the given range, with a greater number of lower values investigated.

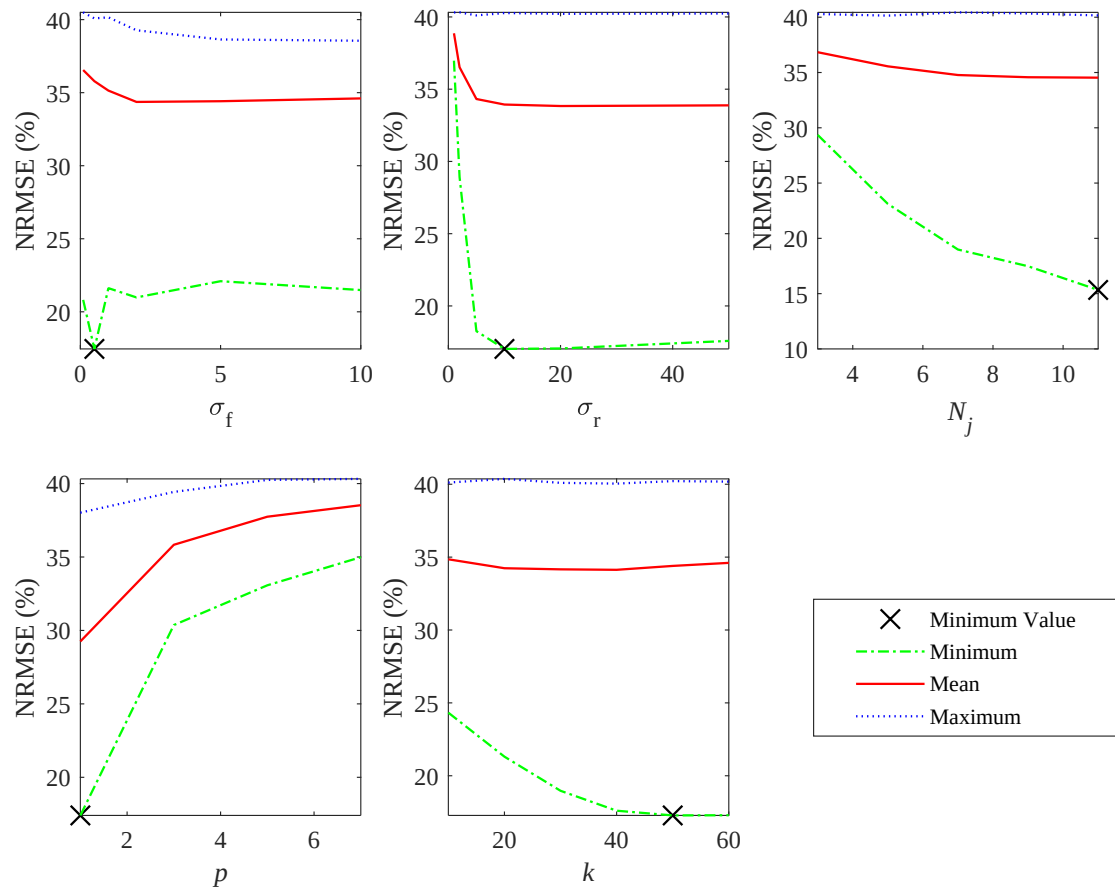


Fig. 3.3 Repeated line search along the hyperparameter values for each kernel hyperparameter, using a 2D simulated phantom. Each line search is repeated for 50 realisations, with the other four hyperparameters randomly selected for each line search. NRMSE refers to whole brain NRMSE.

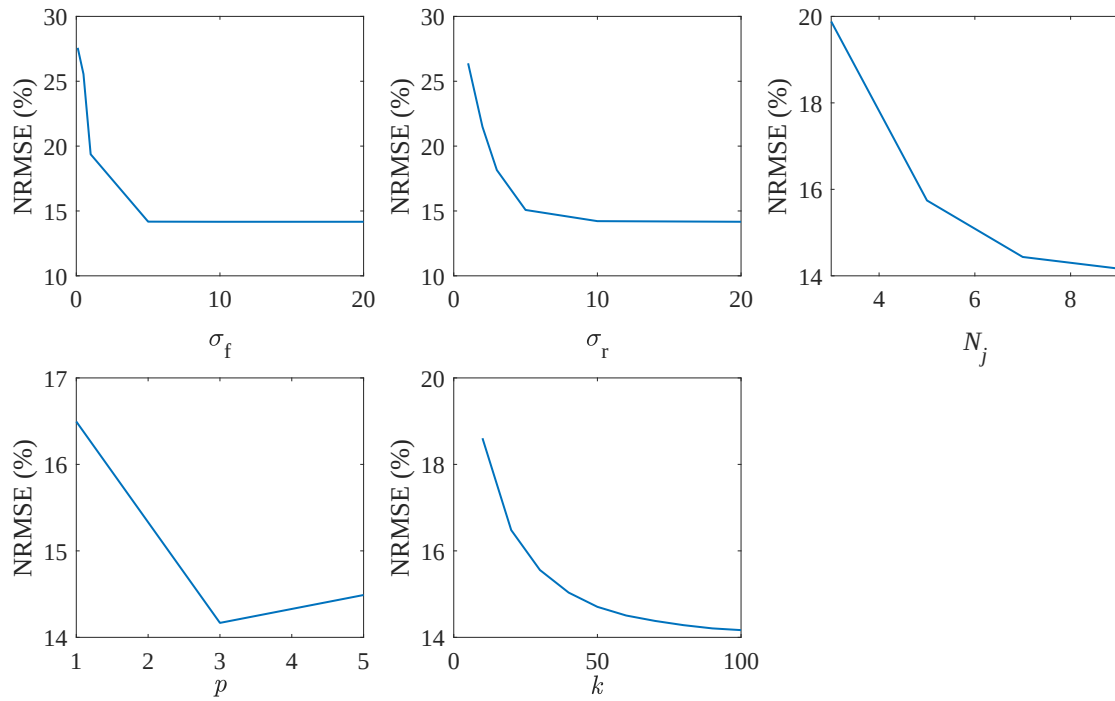


Fig. 3.4 Hyperparameter line search along the hyperparameter values for each of the kernel hyperparameter, using real 3D FDG data. The hyperparameters were initialized using the values determined from the 2D setup. The stopping criteria of the hyperparameter selection algorithm required all 5 hyperparameters to remained unchanged for a single update cycle (all 5 hyperparameters updated). These graphs show the 1D hyperparameter search for each hyperparameter for the last cycle over all hyperparameters.

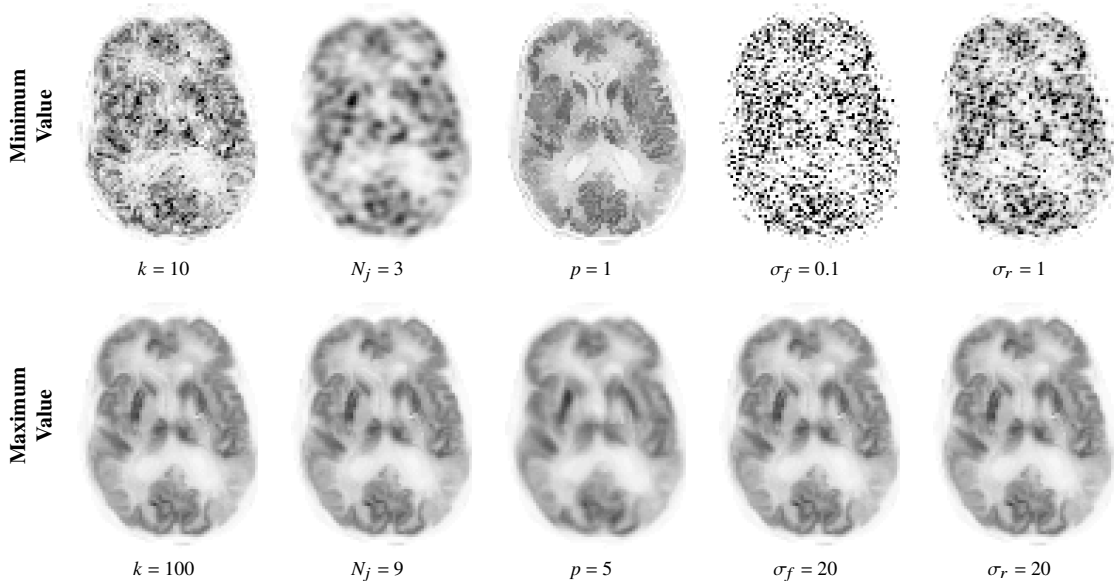


Fig. 3.5 Reconstructed KEM images using the maximum and minimum kernel hyperparameter value in the range of each hyperparameter. The top image (for each hyperparameter) is reconstructed using the smallest kernel hyperparameter value in the range, and the bottom image is reconstructed using the largest kernel hyperparameter. The remaining fixed kernel hyperparameters are the 3D chosen hyperparameters stated in Table 3.1.

3.5 3D Simulated Phantom Study

3.5.1 Experimental Methods

A 3D simulation study was performed to compare the MR-informed KEM method to MLEM, using the $[^{18}\text{F}]\text{FDG}$ phantom described in §3.3.1. Both methods are investigated for datasets of varying count levels (100%-1%) of the original 5×10^8 count level. The images are reconstructed up to 300 iterations, with the impact of post-reconstruction smoothing explored for both methods. In this study the metric of whole-brain NRMSE is used to evaluate the reconstructed image quality (and denoising power), using the noise-free MLEM image as the ground truth. The reconstructed noise-free MLEM image was chosen as the ground truth, to allow a greater level of comparison between the simulated and real data studies (in which the

noisy MLEM reconstructed image of the full counts dataset was used as an effective ground truth). The change in image voxel values between successive iterations is also investigated in terms of percentage difference (referred to as $NRMSE - PD$) to highlight the convergence trends of each method:

$$NRMSE - PD(n) = \sqrt{\frac{\sum_j^{N_v} (\theta_j^n - \theta_j^{n-20})^2}{\sum_k^{N_v} (\theta_k^{GT})^2}} \times 100(\%) \quad (3.16)$$

This metric indicates how much the average voxel changes (relative to the ground truth) between its value at the current iteration (n), and its value at an earlier iteration number ($n - 20$). In addition regional error metrics of mean ($Mean_{ROI}$) and coefficient of variation (CV_{ROI}) are also investigated within a specified ROI, with the CV_{ROI} defined as:

$$CV_{ROI}(\boldsymbol{\theta}) = \frac{SD_{ROI}(\boldsymbol{\theta})}{Mean_{ROI}(\boldsymbol{\theta})} \quad (3.17)$$

where $Mean_{ROI}(\boldsymbol{\theta})$ is the mean voxel value and $SD_{ROI}(\boldsymbol{\theta})$ is the standard deviation within a ROI. The right caudate and white matter ROIs were segmented manually, through applying a threshold to the simulated phantom (right caudate only), within a restricted spatial region.

3.5.2 Results

The reconstructed KEM and MLEM images from this simulation study are shown in Fig. 3.6 for a tumour containing slice, at 100 iterations. The hyperparameters used for KEM are stated in Table 3.1, with the number of kNN used for each count level increased with the level of denoising required. A zoomed in coronal slice is shown in Fig. 3.7, highlighting the high intensity tumour region via the red arrowhead. The variation in whole-brain NRMSE (Eq. 3.15) of both methods is shown in Fig. 3.8 for differing count levels, amount of post-reconstruction smoothing, and iteration number. Both the low dose reconstruction and the full dose reference image are smoothed to the level stated on the axis, prior to the

NRMSE calculation. The variation of NRMSE with respect to iteration number is more clearly presented in Figure 3.9 (columns A and B). Figure 3.9 (columns C and D) show the difference in voxel values between iterations (in steps of 20), averaged over all the voxels in the image and presented as a percentage difference (Eq. 3.16). These curves indicate the proximity of each method to convergence, and show how this varies with count level.

Each method is also assessed in terms of the trade-off between the mean ($Mean_{ROI}(\theta)$) and coefficient of variation (CV_{ROI}) metrics, evaluated over a white matter and right caudate ROI, as shown in Figure 3.10 (the ROIs are highlighted on the right hand side of the figure). A horizontal profile through the tumour region is also shown in Figure 3.11, for the count levels from 100% to 25%. The low count level profiles are not shown as they are dominated by noise.

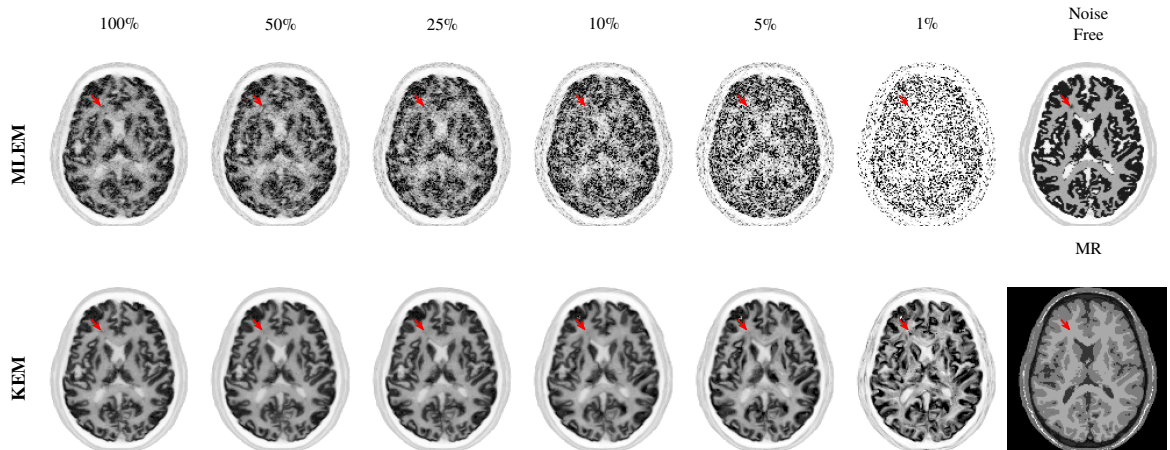


Fig. 3.6 A 2D transverse slice of the MLEM and KEM reconstructed images are shown for a range of reduced dose levels (from 100% to 1% of the full count data), at 100 iterations. The datasets are constructed from the simulated 3D phantom. For the kernel method, the number of kNN was increased with decreasing count level, $kNN = 10, 25, 25, 50, 50$, and 50 , for count levels from 100% to 1% respectively. The high intensity tumour present in the PET data is not present in the T1 image. The location of the tumour is indicated by the red arrow.

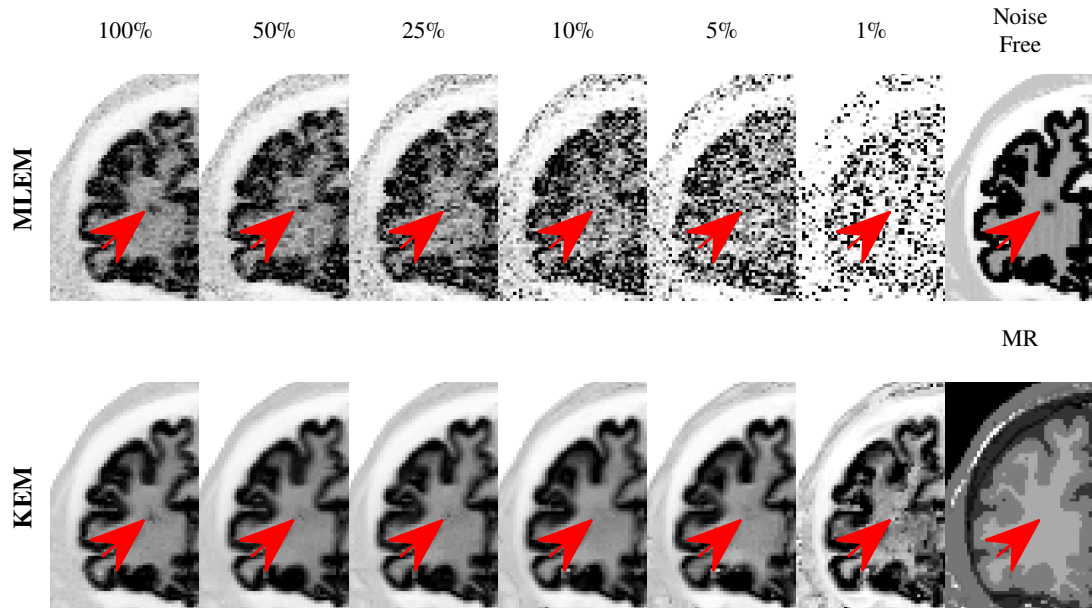


Fig. 3.7 A 2D zoomed in coronal slice of the MLEM and KEM reconstructed images are shown for a range of reduced dose levels (from 100% to 1% of the full count data), to highlight the high intensity tumour region. The datasets are constructed from the simulated 3D phantom. For the kernel method, the number of k NN was increased with decreasing count level, k NN = 10, 25, 25, 50, 50, and 50, for count levels form 100% to 1% respectively. The high intensity tumour present in the PET data is not present in the T1 image. The location of the tumour is indicated by the red arrow.

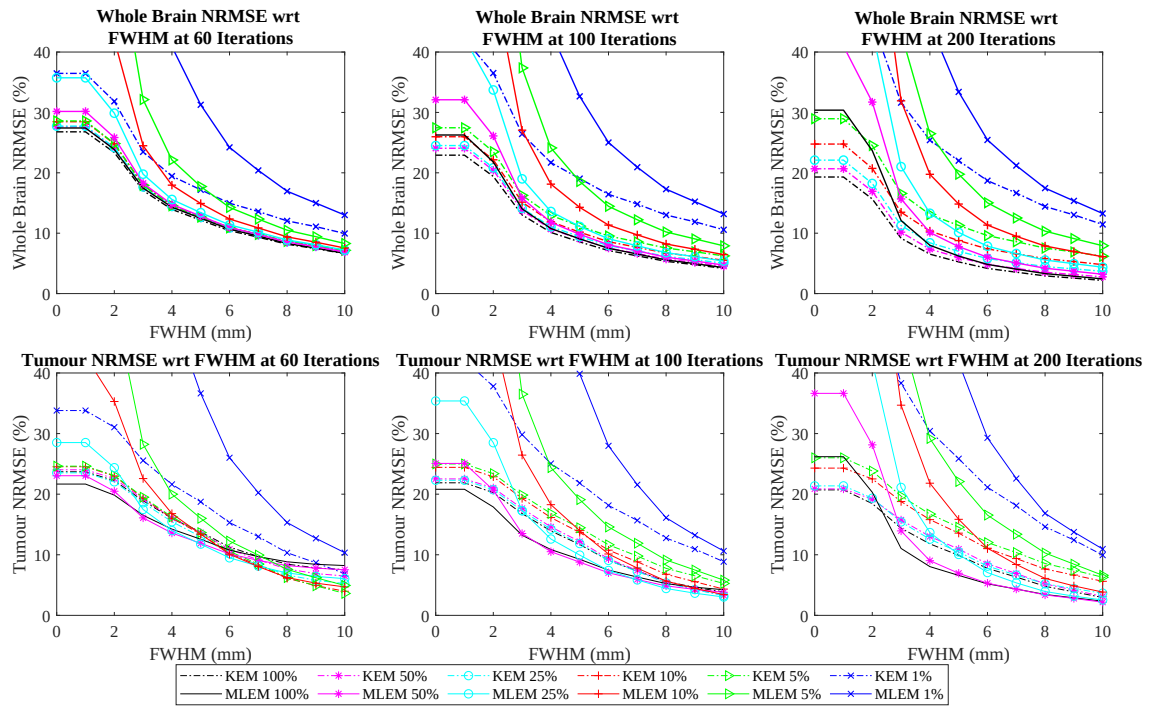


Fig. 3.8 Whole brain (top row) and tumour (bottom row) NRMSE calculated for the reduced dose reconstruction of the 3D simulated dataset, using the MLEM reconstruction of the noise free data at 300 iterations as the ground truth. Both the reduced dose reconstruction and the full dose reference image are smoothed by the specified FWHM along the x-axis. The results are shown for different iteration numbers.

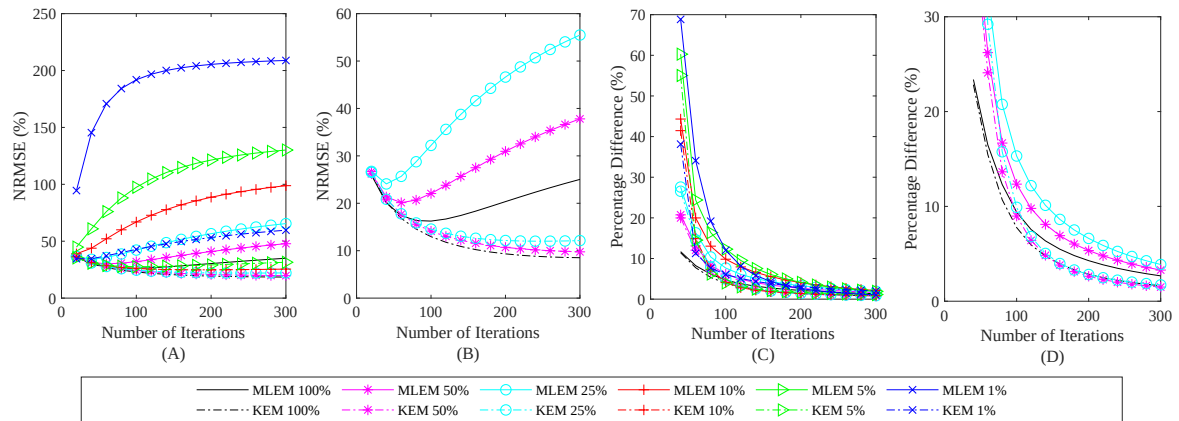


Fig. 3.9 Column (A) shows the variation in whole brain NRMSE with respect to iteration number. Column (B) shows the variation in whole brain NRMSE for the higher count data only (for greater clarity). Column (C) shows the change in voxel values in steps of 20 iterations, averaged over all voxels in the image and presented as a percentage difference (Eq. 3.16). Column (D) presents the same data as column (C) but for the higher count data only (for greater clarity).

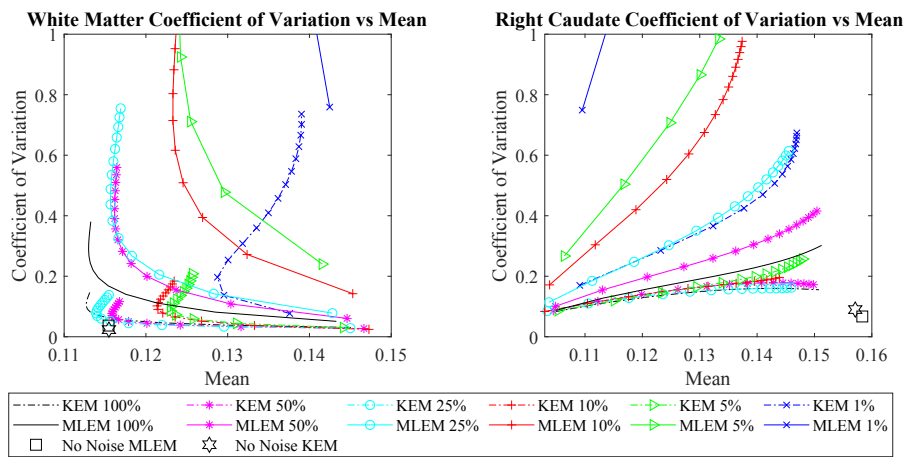


Fig. 3.10 Coefficient of variation versus mean value for a white matter region and the right caudate region, calculated from the KEM and MLEM reconstructed images of the simulated data. The corresponding value for the MLEM reconstruction of the no noise data at 300 iterations is shown by the single black square, with the KEM reconstruction of the no noise data shown by the black star. The ground truth value for the caudate region and the white matter region are 0.306 and 0.116, respectively. A 2D slice of the 3D ROIs are highlighted in red on the right hand side images.

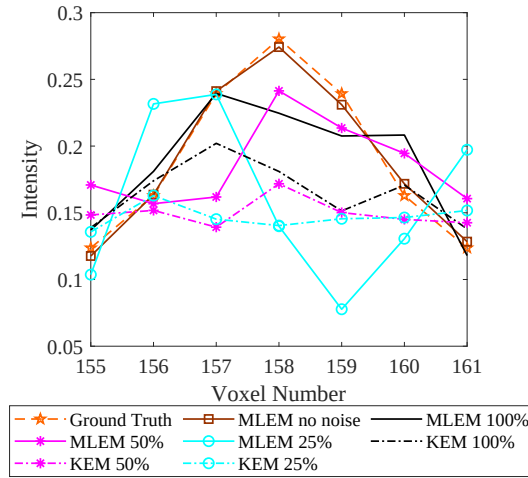


Fig. 3.11 Horizontal intensity profile through the tumour region in the frontal white matter region. The profiles for all methods are taken from the reconstructed images at 100 iterations.

3.6 3D Real Data Study

3.6.1 Experimental Methods

A similar set of experimental methods are now undertaken using a PET patient scan (real data), which underwent post-acquisition sub-sampling to produce a series of datasets with a range of count levels (100% to 1%) as described in §3.3.2. Through repeating the methodology and error analysis in a similar manner as in the simulation studies, similarities between the simulated and real data studies can be easily analysed. In consideration of this, the KEM hyperparameters used in the real data study match those used for each count level in the simulated data study.

In an extension to the simulated studies methodology, an optimal level of post-reconstruction smoothing was investigated (for each method at each count level) for the real data reconstructed images. The level of post-reconstruction smoothing was selected to minimise the whole brain NRMSE between the reduced dose image and the full dose MLEM image (refer-

ence image). A 4mm FWHM post-reconstruction smoothing was applied to the reference MLEM image, matching the smoothing width used for mMR datasets as described in previous literature [39, 40]. For each reconstructed image, the optimal level of post-reconstruction smoothing was calculated through applying a series of smoothing widths to the reduced dose image. An interpolation function was then used to find the smoothing width that minimized the NRMSE.

3.6.2 Results

The reduced dose reconstructed images for the KEM and MLEM methods with no post-reconstruction smoothing are shown in Figure 3.12. The reconstructed images of each method, with the optimal level of post-reconstruction smoothing applied (as defined in §3.6.1) is shown in Figure 3.13. In a similar manner to the simulation studies, the whole brain NRMSE between the reduced dose data images and the reference MLEM image is investigated (as shown in Figure 3.14), with respect to post reconstruction smoothing and iteration number. As before, the low dose reconstruction and the full dose MLEM reconstructions are both smoothed according to the level indicated on the x-axis. Figure 3.15 shows the coefficient of variation versus mean values for the right caudate and white matter regions.

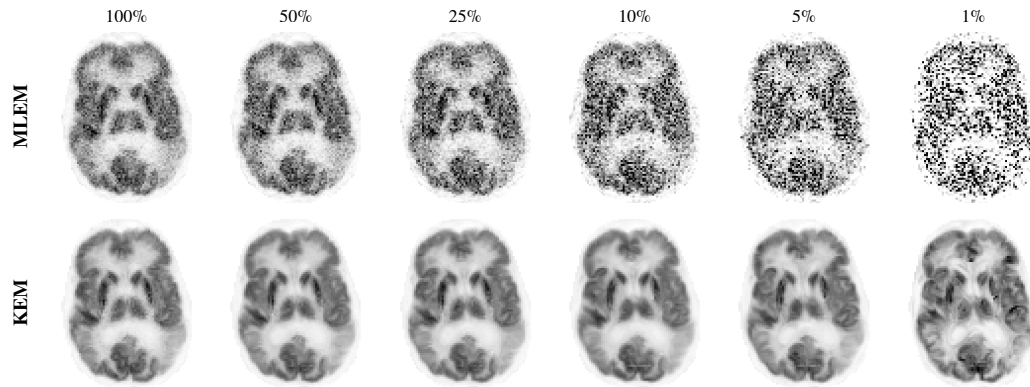


Fig. 3.12 KEM and MLEM reconstructed images of the 3D real patient dataset resampled at varying count levels (100% to 1%). All reconstructed images are shown at 100 iterations. The kernel hyperparameters used are stated in Table 3.1. The number of k NN was increased with decreasing count level, k NN = 10, 25, 25, 50, 50, and 50, for count levels 100% to 1% respectively.

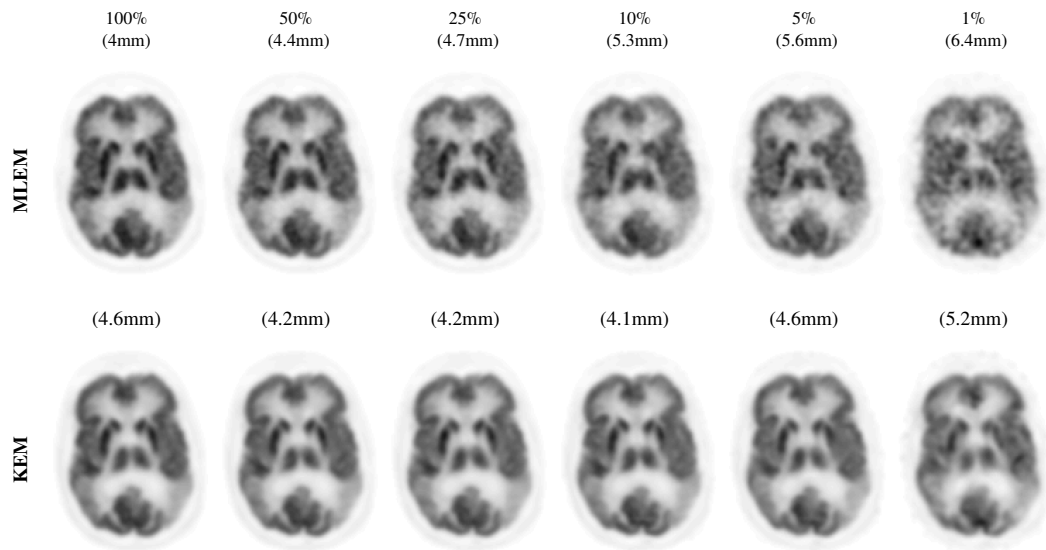


Fig. 3.13 KEM and MLEM reconstructed images of the 3-D real patient dataset resampled at varying count levels (100% to 1%). All images have undergone post reconstruction smoothing, with the FWHM stated above each image. The width of the FWHM was chosen to minimize the NRMSE between the reduced dose image and the full count MLEM data, with 4 mm post reconstruction smoothing. All reconstructed images are shown at 100 iterations. The kernel hyperparameters used are stated in Table 3.1. The number of k NN was increased with decreasing count level, k NN = 10, 25, 25, 50, 50, and 50, for count levels 100% to 1% respectively.

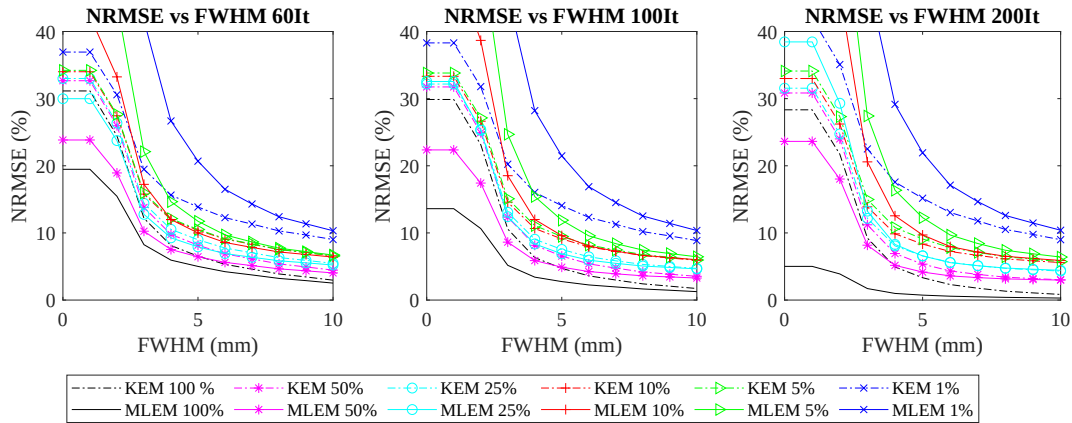


Fig. 3.14 Whole brain NRMSE between the reduced dose reconstruction (at stated iteration numbers) and the MLEM reconstruction of the full dose data at 300 iterations. Both the reduced dose reconstruction and the full dose reference are smoothed by the FWHM specified along the x-axis. These results use the real patient dataset.

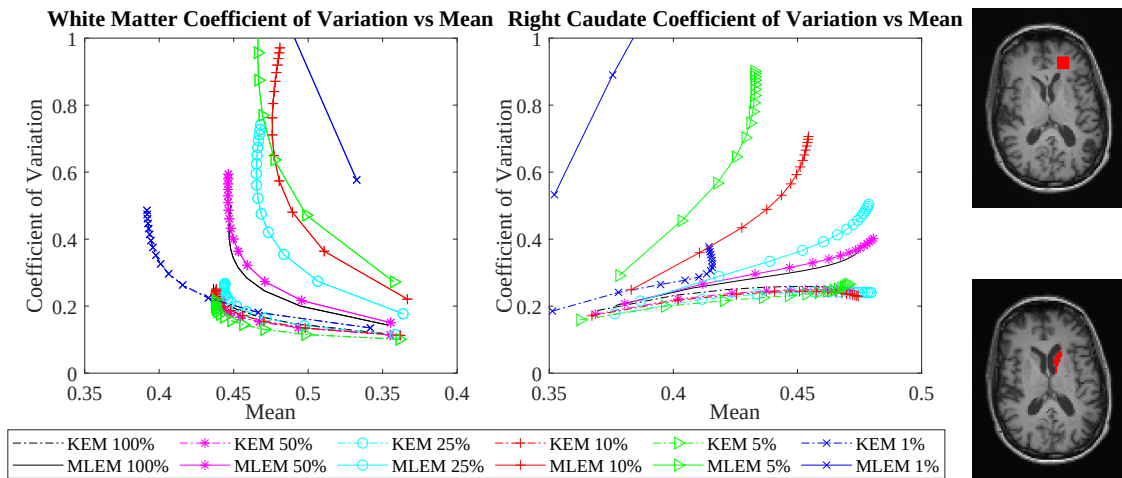


Fig. 3.15 Coefficient of variation versus mean value for a white matter region and the right caudate, calculated from the KEM and MLEM reconstructions of the real data. The ROIs are highlighted on the MR image on the right hand side.

3.7 Discussion

This chapter has presented a novel application of the MR-informed kernel method, evaluating the kernel method's accuracy in reconstructing very low dose PET datasets in comparison

to post-smoothed MLEM. Both methods were investigated using simulated and real patient datasets acquired on the Siemens Biograph mMR, for an [^{18}F]FDG scan. As expected, the use of low dose datasets severely degrades the reconstructed images, when reconstructed using MLEM. The MLEM images of low dose datasets with no post-reconstruction smoothing applied were very susceptible to noise, for both simulated and real PET FDG data, as shown by Figure 3.6 and Figure 3.12. In comparison, the KEM images of the low dose data are visually less noisy, and retain many detailed structures present in the MLEM reconstructed images of the noise free or high count datasets.

The use of post-reconstruction smoothing can be seen to majorly benefit the MLEM reconstructed images of low dose datasets, in which high levels of smoothing reduce the level of noise in the image, as shown in Figure 3.13. But, without the use of MR structural information, the uniformly smoothed MLEM images have blurred anatomical boundaries, and increased PVEs. The smoothing of PET intensities across regional boundaries is detrimental to the clinical efficacy of PET images. For instance an overly-blurred MLEM image will bias the quantification of MR-visible regions that are important in the management of neurological diseases including Parkinson's, epilepsy, and Alzheimer's disease [76, 72, 75, 74, 310]. Additionally, such smoothing may alter the regional intensity patterns used in the assessment of dementia. This highlights the advantages of using the KEM methods for reconstructing low-dose datasets, due to the KEM reconstructed images closer visual similarity to the full dose image. Features such as the caudate, putamen and thalamus are clear in the KEM reconstructed images to a reduced dose level of 5%, which is not achievable for MLEM.

However, for both the KEM and MLEM reconstructed images of the 1% dataset, visual differences can be seen compared to the full dose reconstruction, e.g. the increased asymmetry of the thalamus and the poor recovery of the putamen. This illustrates a potential limitation for the level of dose reduction achievable using the kernel method. As expected, the count level of the PET data cannot be reduced indefinitely without effecting the reconstructed

images. The application of post reconstruction smoothing, shown in Figure 3.13, can be seen to partially resolve this issue by smoothing away the artificial looking MRI structures from the KEM images, although increased levels of smoothing negatively impact the quantification accuracy as discussed above.

In the 3D simulation study §3.5 the KEM reconstructed image of 100% of the data provides the lowest whole brain NRMSE (as shown in Figure 3.8) irrespective of post reconstruction smoothing. The KEM reconstruction of 10% of the data has comparable whole brain NRMSE to the MLEM reconstruction of 100% of the data (for simulated data, at 100 iterations). The NRMSE of the KEM reconstructed images are also very closely grouped (between the 100%-5% dose levels), demonstrating that the kernel method is relatively insensitive to reduced dose (down to the 5% dose level). However, the increase in post reconstruction smoothing level unsurprisingly reduces the improvement of the kernel method over MLEM, at the expense of spatial resolution. Similar trends are also seen for the real data NRMSE in Figure 3.14, where both the kernel method and MLEM are compared to the MLEM image of 100% dose data, at 300 iterations. In Figure 3.14, the NRMSE curves for the kernel reconstructed images are also tightly grouped as was the case for the simulated data. The use of the MLEM reconstructed image of 100% count data as the real data reference image makes it unfair to directly compare the KEM and MLEM NRMSE values. This is particularly true for high count levels, and for low levels of post reconstruction smoothing, where the MLEM reconstructed images will inherently be more like the chosen reference image. For these cases, the simulated data provide a more reliable assessment of the performance of the different methods.

Figure 3.9 (columns A and B) shows the variation of the KEM and MLEM whole brain NRMSE curves with respect to iteration number, with the minimum kernel NRMSE value outperforming the minimum MLEM NRMSE value for a given count level. MLEM shows a steep rise in NRMSE values at high iteration numbers, in comparison the KEM NRMSE

values at high iteration numbers remain closer to their minimum value. This indicates the kernel method's relative insensitivity to iteration number (for high count data). For reconstructed images of lower count data, the minimum NRMSE value occurs at earlier iteration numbers. Figure 3.9 (columns C and D) show that the kernel method tends to converge faster than the MLEM reconstruction for the same count data, with the kernel method reconstructions for any count level (except 1% data), converging at a similar rate.

The kernel method is outperformed by the MLEM method for the reconstruction of the high intensity tumour region in the simulated data, both in terms of NRMSE and visually. Figure 3.11 shows the suppression of the tumour region when reconstructed by the kernel method in comparison to the MLEM method. The lower count levels are not shown due to the dominant noise present in the reconstructed images. The suppression of PET unique regions by the kernel method in comparison to MLEM, agrees with the findings of [81] and demonstrates a clear limitation in the MR-guided kernel method in the recovery of features unique to the PET data.

The regional error metrics show a clear reduction in noise (coefficient of variation) for the KEM reconstructed images of both the simulated and real datasets, as shown in Figure 3.10 and Figure 3.15 respectively. For the right caudate, the KEM mean values of the reduced dose simulated data are shown to be closer to the true value compared to MLEM at reduced doses. There is a similar pattern between the simulated and real data curves of the right caudate region, from which it can be inferred that the KEM mean values of the reduced dose real data may be closer to the true mean value. For real data, the KEM mean value of the right caudate for the different dose level reconstructed images (5%-100%) are within 5.7% of one another, again indicating the insensitivity of the kernel method to reduced dose data. For the white matter region, the KEM mean values for real data are also closely grouped within 5.9% of one another, between the 100% and 5% data levels. These mean values also lie very

close to the mean value of the full count MLEM white matter ROI, indicating the accurate quantification of this region down to the 5% count level.

3.8 Limitations

The MR-informed kernel method has demonstrated improvements in the reconstruction of low dose data from both a visual and quantitative standpoint. However, although visually little impact is seen in the kernel reconstruction of reduced dose data (down to 5%), small deviations in the NRMSE values for the real data reconstructions are seen for all reduced count levels. The impact of the slight variation in the KEM reconstructions of reduced dose data should be assessed by a clinician, to determine to what count level the low count kernel images remain diagnostically comparable to the full count reconstructions.

In addition, the MR-informed kernel method is adversely affected through incorrect anatomical guidance, as shown here and in the literature [81][311]. The robustness of the kernel method has been demonstrated to some extent by Baikejiang et al [299], but incorrect anatomical guidance still leads to evident visual blurring. The kernel method can be susceptible to the following forms of anatomical misguidance at any given region in the image: 1) a PET feature that is not present in the MR, 2) an MR feature that is not present in the PET, 3) features in both the PET and MR which do not match. These three cases are now considered in turn.

A PET feature that is not present in the MR leads to suppression of the PET unique feature by the kernel method. Choosing hyperparameters that result in smaller basis functions will reduce the suppression of such features, but at the expense of increased noise. Similarly, for penalised methods (e.g. the Bowsher prior [223]), the over-smoothing of PET unique features can be reduced through varying the hyperparameter value. In both cases, reducing the influence of the anatomical image reduces the noise suppression properties.

A MR feature that is not present in the PET, will have only a limited impact on the KEM reconstruction method. MR features which are not present in the PET data will lead to inappropriate PET basis functions, however the coefficients for these basis functions are estimated from the PET data only. Therefore, provided such basis functions are located within homogeneous regions of the PET image, the additional MR features will not have an adverse impact on the reconstructed PET image. Similarly, for penalised methods, mismatching MR features will just cause constrained regions of smoothing in homogeneous PET regions.

Finally, for the case of PET and MR features that do not match, the PET data will be blurred by the MR-shaped basis functions. This issue can again be mitigated through the selection of hyperparameters that produce small, compact basis functions, which can better represent unique PET features. To extend the kernel method to other applications, such as oncology and non-FDG tracers would require these issues to be addressed.

3.9 Real Patient Case Study

This work presented on the MR-informed kernel method has contributed to a broader clinician led study, in which a range of reconstruction algorithms were assessed at reduced count levels for the diagnosis of Alzheimer's disease and temporal lobe epilepsy patients [312]. Clinical assessment and diagnosis of such scans is generally based on recognising broad FDG patterns. For instance, a temporal lobe epilepsy patient would expect to see a hypo metabolism in the associated temporal lobe, as shown in Figure 3.16. The reconstructed PET images were evaluated in terms of global image quality, diagnostic confidence and pattern recognition for multiple patient cases. The full dose PET and corresponding MR images are shown in Figure 3.16, for a temporal lobe epilepsy patient. The variation in global quality and diagnostic confidence for this patient are shown in Figure 3.17, each metric scored on a scale from 0 to 3. The investigated MR-informed methods of KEM and MAPEM can be seen to outperform MLEM, down to a dose level of 5% for this patient for both metrics. Averaging

across all the patient scans examined in this study, the preliminary findings indicated that the inclusion of MR information, via methods such as KEM, allowed a notable dose reduction to be achieved without effecting the diagnostic power of the PET images for dementia and epilepsy patients.

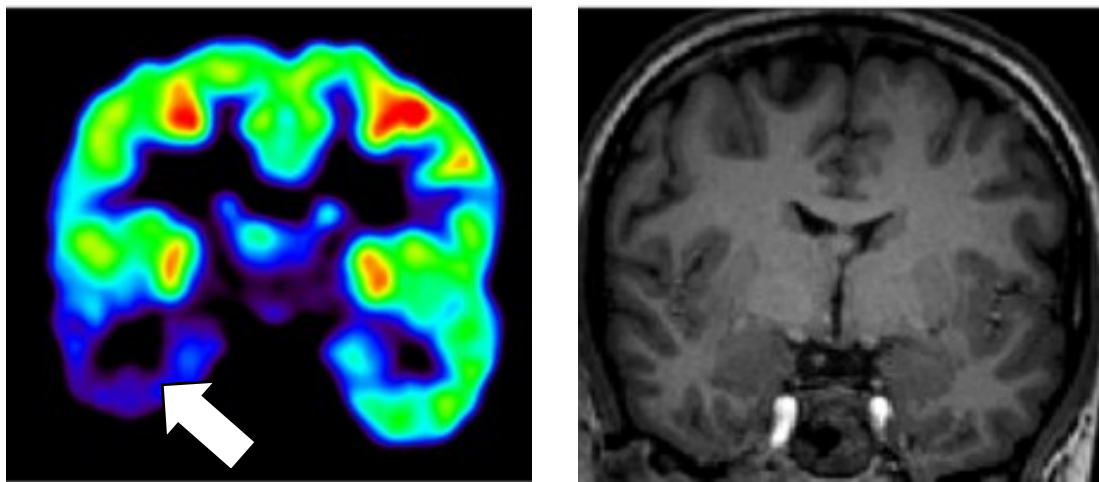


Fig. 3.16 [^{18}F]FDG PET image (left) and corresponding T1w MR image (right) of a patient with unilateral temporal lobe epilepsy (hypometabolism temporal lobe indicated by the arrow). The PET image is of the full dose (100%) dataset, reconstructed using the MLEM algorithm with post-reconstruction smoothing

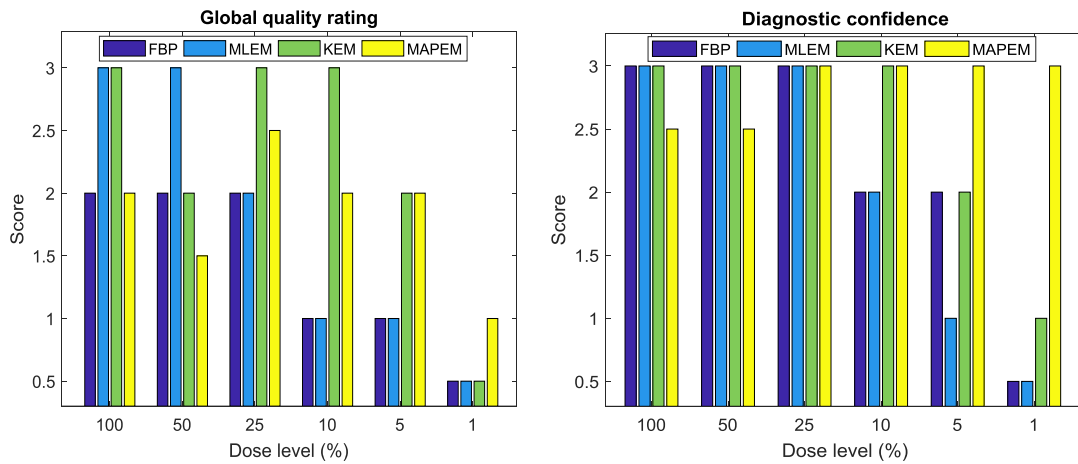


Fig. 3.17 Reconstructions of 100%, 25%, 10%, 5% and 1% dose levels for a patient with unilateral temporal lobe epilepsy. Reading results assessed in terms of global quality (left) and diagnostic confidence (right) for different dose levels and reconstruction methods on a scale of 0-3.

3.10 Summary

In the majority of neurological scans there is a large degree of structural similarity between the functional PET image and the anatomical MR image. The inclusion of MR information into the PET reconstruction via the kernel method is therefore a highly attractive option for reducing noise and PVEs in comparison to MLEM, whilst maintaining shared PET-MR boundaries. In this chapter the MR-informed kernel method was investigated in terms of its capability to maintain image quality for reconstructing PET data at a substantially reduced count level. The KEM and MLEM methods were evaluated over a range of count levels (100% to 1%), for both simulated and real datasets. The impact of post-reconstruction and iteration number was also assessed. The kernel method demonstrated the capability to reconstruct impressively low count datasets without impairing the image quality, for instance the KEM reconstructed images for 10% of the data had comparable whole brain NRMSE to the MLEM reconstruction of 100% of the data (for simulated data, at 100 iterations). For

regional metrics, the kernel method at reduced count levels (corresponding to reduced dose levels) attained a reduced coefficient of variation and more accurate mean values compared to MLEM. For real data, the KEM mean value of the right caudate for the different dose level reconstructed images (5%-100%) are within 6% of one another, and very similar to the mean value of the MLEM reconstruction of the full count dataset. Despite the clear benefits of the kernel method, its success is highly dependent on the close correspondence between the PET and MR structures. For the simulated study the MR-informed kernel method clearly suppressed high intensity PET tumours, making it unsuitable for oncology cases. Whereas for the real data study and the clinician led study of dementia patients, the kernel method using 10% of the data demonstrated comparable images and diagnostic utility to that of the full count MLEM images.

Chapter 4

The MR-Informed Kernel Method With Spatially Compact Basis Functions

The results presented in Chapter 3 show the excellent denoising power of the MR-informed kernel method incorporated into the expectation maximisation algorithm (KEM) for PET reconstruction. The simplicity of re-parameterising the maximum likelihood objective function through kernel derived basis functions is a clear benefit in comparison to more involved regularised methods [313]. The MR-informed KEM method also requires one fewer hyperparameter in comparison to penalty based methods, although practical schemes have been devised to determine the β regularisation parameter for an approximately uniform image resolution [314].

However, the MR-informed KEM method is currently limited in its potential clinical utility due to its susceptibility to PET-MR mismatches, as shown in Chapter 3. These suppressed PET-unique features may be of considerable clinical importance, particularly for oncology cases. The MR-informed KEM method implemented in Chapter 3 also down-sampled the MR image to PET resolution before extracting MR structures. This may not be the best means for using the complementary MR information, and further improvements in the reconstructed images could be attained through using the MR image in its native resolution.

This Chapter seeks to address some of the shortcomings of the previously presented KEM method through: (1) introducing an alternative means for constructing the kernel derived spatial basis functions to reduce suppression of PET-unique regions, and (2) investigating the advantages and disadvantages of reconstructing the PET image with MR size voxels using spatial basis functions. This Chapter is based on work previously published by the author during the course of his PhD studies [69].

4.1 Spatially Compact Basis Functions

The previously considered MR-informed KEM method selects the contributing voxels for each basis function based on their k NN in feature space. In keeping with other previous implementations of the MR-informed KEM method [81, 295, 305], the implemented method in Chapter 3 only used MR intensity patches or single voxels in the feature vector formation. In this chapter the MR-informed kernel method is extended to include the positional vector of the voxel in a composite feature vector description. The intensity and positional components of the feature vector are each weighted in accordance with their respective standard deviations. The relative values of the standard deviations therefore determines the relative importance between the intensity and positional components of the feature vector. This new composite feature vector (χ) is described below as:

$$\chi_j = \begin{bmatrix} \frac{f_j}{\sigma_f} \\ \frac{r_j}{\sigma_r} \end{bmatrix} \quad (4.1)$$

Where j is the voxel index, $\mathbf{f}$ is the intensity based feature vector, $\mathbf{r}$ is the positional vector, σ_f is the standard deviation that determines the width of the Gaussian kernel acting on the intensity based feature vectors, and σ_r is the standard deviation that determines the width of the Gaussian kernel acting on the positional based feature vectors. The standard

deviation values σ_f and σ_r are user selected, to define the extent of the kernel and the relative importance of the spatial and intensity feature vectors. Using the composite feature vector formulation, the process by which the basis functions are constructed remains the same as previously described in §3. For both the conventional intensity only and composite feature vector cases each voxel is compared to its neighbouring voxels within a predefined spatial neighbourhood in terms of their (composite) feature vectors. The k -nearest neighbours to this central voxel in (composite) feature space, measured in terms of Euclidean distance are then selected to contribute to the basis function. As the composite feature vector now includes a positional term, spatially proximal voxels are preferred over distant voxels, although this is also dependent on their intensity (patch) values. The relative importance between the spatial location and intensity (patch) value in the selection of contributing voxels is determined by the standard deviation values. For instance a small value of σ_r (i.e. $\sigma_r \ll \sigma_f$), will make the resulting basis functions very sensitive to the positional vector $\mathbf{r}$, and will therefore likely produce compact basis functions. Alternatively, if $\sigma_f \ll \sigma_r$ then the selection of contributing voxels to each basis function will be more sensitive to the voxel's intensity (or patch) value, producing basis functions that clearly highlight structures in the MR. The Euclidean distance between two voxels, j and k , evaluated in the composite feature space is given by:

$$\|\mathbf{x}_j - \mathbf{x}_k\|_2^2 = \frac{1}{\sigma_f^2} \|\mathbf{f}_j - \mathbf{f}_k\|_2^2 + \frac{1}{\sigma_r^2} \|\mathbf{r}_j - \mathbf{r}_k\|_2^2 \quad (4.2)$$

The above equation (Eq. 4.2) is also the argument of the exponential term used to calculate the weights of the contributing voxels for a given basis function. As such, the Gaussian exponential term for the composite feature vector is given by:

$$\kappa(\mathbf{x}_j, \mathbf{x}_l) = \exp\left(\frac{-\|\mathbf{x}_l - \mathbf{x}_j\|_2^2}{2}\right) = \exp\left(\frac{-\|\mathbf{f}_l - \mathbf{f}_j\|_2^2}{2\sigma_f^2}\right) \exp\left(\frac{-\|\mathbf{r}_l - \mathbf{r}_j\|_2^2}{2\sigma_r^2}\right) \quad (4.3)$$

The product of exponentials is the same set of exponential terms presented in Eq. 3.9, however the expression has now been arrived at more elegantly, in keeping with the original Gaussian kernel form Eq. 3.8 but using the composite feature vector (χ). The important aspect of the composite feature vector is its ability to produce more compactly shaped basis functions, through the k NN selection in composite feature space, without reducing the number of voxels contributing to the basis function. By comparison, the spatial Gaussian weights originally presented in [305], are of minor benefit, as it reduces the number of voxels contributing to the basis function and hence its overall denoising capability.

The spatially compact basis functions presented above, which were arrived at through employing the composite feature vector can also be attained in an alternative manner, which shall now be explained. Instead of sparsifying the basis functions through selecting the k NN in feature space, an alternative form of sparsification is used whereby the kernel weights are initially calculated for all voxels within the spatial neighbourhood, and then only a fixed number of voxels with largest weights are selected to contribute to the basis function. This KEM method is therefore called KEM largest value sparsification (LVS). In this implementation the weights are calculated using the product of Gaussians (Eq. 3.9), which includes the spatial term and intensity term. The feature vector (f) expressed in Eq. 3.9 is the original intensity only feature vector. When discussing the compact basis functions in terms of the KEM LVS method only this form of intensity only feature vector will be used, which is the same as used in the conventional k NN KEM method. The resulting spatially compact basis functions of the KEM LVS method are identical to the basis functions produced by the composite feature vectors. However, due to the simpler sparsification process of the KEM LVS method over the composite feature vector, it is this implementation that is used throughout this chapter (and thesis). This is purely an implementation detail as the resultant basis functions from either method are identical. Therefore, all reference to feature vectors for the remainder of this chapter will be referring to the intensity only feature

vector, not the composite feature vector. The compact basis functions and conventional KEM basis functions shall be distinguished through their sparsification methods, LVS and k NN respectively. To further clarify this point the modified KEM method using spatially compact basis functions will be referred to as KEM LVS, whereas the conventional KEM method shall be referred to as KEM k NN. A schematic showing the structural differences, and differences in formation between the conventional KEM, the spatial Gaussian KEM, and the spatially compact KEM LVS is shown in Fig. 4.1. It should be noted that the two methods compared in this chapter are KEM with spatial Gaussian and KEM LVS.

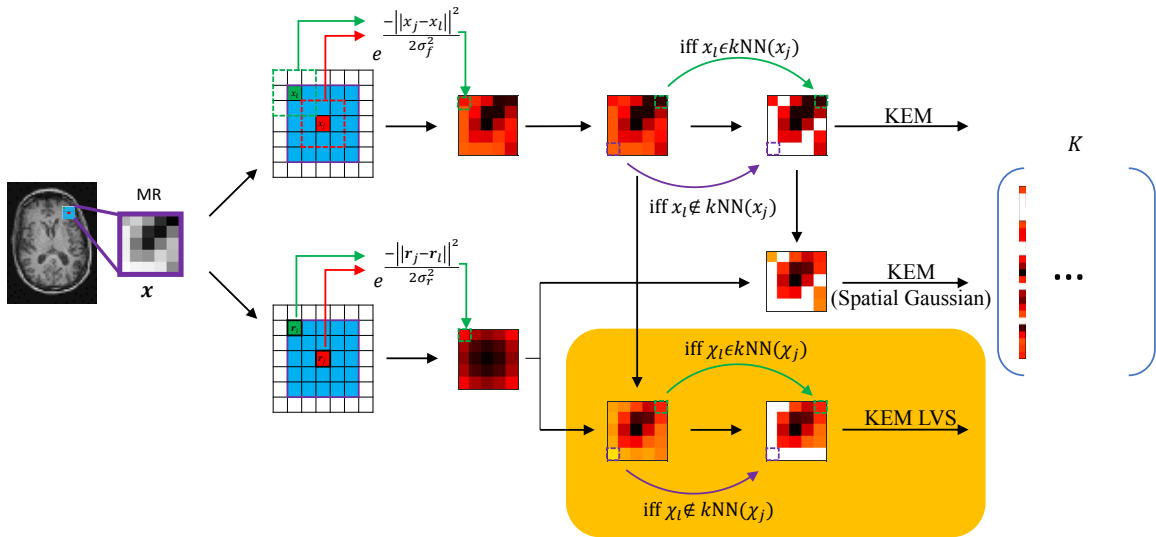


Fig. 4.1 Schematic showing the construction of an MR-informed basis function. For the kernel method with and without the inclusion of the spatial Gaussian term, and the proposed spatially compact KEM LVS method

The added flexibility provided by the KEM LVS formulation maintains the ability of the conventional k NN kernel method to extract structures from the MR image in regions such as the grey-white matter boundary. But in uniform MR regions the positional vector will determine the selection of contributing voxels, due to the lack of variation in MR intensity values in such regions. The more compact nature of these basis functions in uniform regions will aid in the reconstruction of PET-unique features, normally heavily suppressed by the

conventional KEM method. As such, KEM LVS improves the recovery of PET-unique features whilst retaining the edge-preserving and denoising properties of conventional KEM method.

Similarities can also be drawn between the MR-informed KEM LVS basis functions and the spatial basis functions constructed through applying simple linear iterative clustering (SLIC) supervoxels [315–317] to an MR image. The SLIC algorithm uses a distance measure that weights both image intensity and spatial proximity (very similar to that of Eq. 4.2) for assigning individual voxels to supervoxel clusters. As previously stated in §2.8 supervoxel basis functions have previously been used to re-parameterise the PET reconstruction objective function, and has produced comparable results to the KEM method [68].

4.2 PET Reconstruction Using MR Voxel Sizes

The MR-informed KEM method is now extended by using MRI voxels sizes ($\approx 1\text{mm}^3$) to build the basis functions from. These basis functions comprised of MRI sized voxels are then used to reconstruct the PET image at MR voxel size. In this chapter the terminology of MR resolution refers to images comprised from MR size voxels, and PET resolution refers to images comprised of PET size voxels. Using a smaller voxel size for the basis functions provides the added benefits of allowing smaller MR visible structures to be extracted, in addition to reducing the level of PVEs that may lead to regional quantification errors. Related research has also identified key quantification improvements that are achieved through using MR voxel sizes in combination with MR-informed MAP based PET reconstruction methods [70]. To attain these same benefits with the MR-informed KEM method, the MR image will now be used at its native resolution, with the FOV increased to account for the larger FOV of the PET scan. From this MR image, MR voxel size spatial basis functions ($\mathbf{K}'$) can be extracted using the kernel method. The matrices and vectors comprised from MR size voxels shall be represented by a ' to differentiate them from their PET size voxels counterparts. A

PET image can be constructed from combining these basis functions comprised of MR size voxels with appropriate coefficients (α') applied:

$$\theta' = K' \alpha' \quad (4.4)$$

where (θ') is the PET image using MR size voxels. This high resolution PET image is then resampled to low (PET) resolution to allow easy incorporation into the original KEM update equation. A trilinear resampling operator (R) is used for this, to calculate each high resolution voxel's contribution to the eight neighbouring low resolution voxels.

$$\theta = R \theta' \quad (4.5)$$

where (θ) is the PET image comprised of PET size voxels. The update equation to determine the coefficients for the MR voxel size basis functions (from which the MR voxel size PET image can be calculated) is given by:

$$\alpha'^{n+1} = \frac{\alpha'^n}{K'^T R^T A^T \mathbf{1}} K'^T R^T A^T \left(\frac{m}{A R K' \alpha'^n + b} \right) \quad (4.6)$$

It should be noted that reconstructing the PET image to a spatial resolution that exceeds the limited PET data spatial resolution has risks. The sudden cut-off in the measured spatial frequencies leads to non-physical oscillations in reconstructed PET image that may become apparent in the PET image at MR voxel size.

The work in this chapter will focus on the presented developments to the kernel method, including the formation of spatially compact basis functions, and constructing the basis functions from MR size voxels.

4.3 Simulated Phantoms and Patient Data

4.3.1 3D Simulated Phantom

A 3D PET-MR phantom was constructed from the BrainWeb segmented T1 MR phantom following a similar procedure as described in §3.3.1, unless otherwise stated below. Both the PET and MR phantoms were constructed at MR voxel size with a grid size of $688 \times 688 \times 254$ and a voxel size of $1.0431 \times 1.0431 \times 1.0156 \text{ mm}^3$. High intensity lesions (A, B and C shown in Figure 4.2) were added to the white matter of the PET phantom (not present in the MR) with non-uniform intensity values. The max intensity values of tumours A, B and C (for PET images at MR voxel size) were $4\times$, $5\times$ and $6\times$ the intensity of the white matter, respectively. The PET phantom (constructed using MR size voxels) was projected into a span 11 sinogram, and rescaled to a high count level with a total prompts count of 1×10^8 , and a lower count level with a total prompts count of 1×10^7 . Images reconstructed at PET resolution have a voxel size of $2.08626 \times 2.08626 \times 2.03125 \text{ mm}^3$ and a matrix size of $344 \times 344 \times 127$, equivalent to the specifications of the standard reconstruction from the Siemens Biograph mMR scanner.

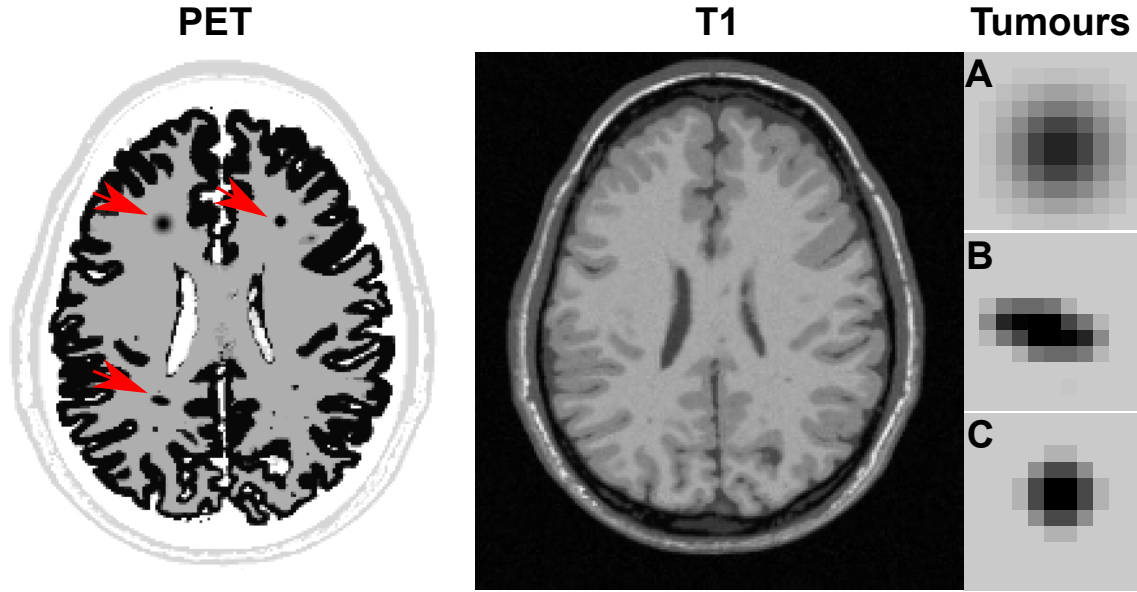


Fig. 4.2 PET phantom (left), MR phantom (center) constructed from the BrainWeb MR phantom [306]. The added PET unique structures are high intensity lesions labelled A, B, and C (right). The PET phantom is constructed using MR size voxels.

4.3.2 3D PET Patient Scan Dataset

An [^{18}F]FDG scan was undertaken on a Siemens Biograph mMR PET-MR with a total prompt count of 4.69×10^8 and a scan duration of 23 min. The tracer activity at time of injection (81 min prior to start of image acquisition) was 229 MBq. A T1 MPRAGE scan provided the anatomical image, from which the kernel basis functions were derived at the native MR voxel size ($1.055 \times 1.055 \times 1.1 \text{ mm}^3$) over the PET field of view, or resampled to the PET resolution ($2.08626 \times 2.08626 \times 2.03125 \text{ mm}^3$). The data acquisition and reconstruction process was the same as in §3.3.2, unless otherwise stated. Due to the patient's medical condition, the dataset does not contain any PET positive MR negative features. To increase the utility of the dataset in assessing the methods under investigation for their ability to recover PET-unique regions whilst simultaneously suppressing noise, genuine known PET-MR mismatches are added to the dataset. This was achieved through augmenting the dataset with simulated tumours, of equal size and shape to those used in the simulated data. The

simulated tumour-only images were forward projected and scaled by the attenuation and normalisation factors to produce the expected tumour sinograms, from which a noisy Poisson realisation is obtained (measured tumour sinograms). The measured tumour-only sinograms are added to the prompts data. The randoms and scatter sinograms were not updated to account for the effects of the added tumour regions.

4.4 Hyperparameter Selection

The σ_f and σ_r hyperparameters for kernel methods were set to large fixed values in keeping with the results in Chapter 3 (§3.4) and the preceding literature [81]. The same standard deviation hyperparameter values are used for both conventional and proposed kernel methods, allowing a comparison which at least ensures that the performance of the conventional kernel method is fairly represented. Alternative hyperparameter values for the proposed kernel method could improve its performance still further. It should be noted that as σ_r and σ_f are set to large values, the kernel weights (calculated using Eq. 4.3) will be relatively fixed around the value of 1, independent of the intensity or positional vectors, producing approximately flat basis functions. In comparison to chapter 3, fewer numbers of k NN are used in the implementation of the kernel methods. The reduced number of k NN used in this chapter is to aid the kernel methods ability to recover the PET-unique regions, which is the focus of this chapter. The same number of k NN are used for both KEM k NN and KEM LVS. An increased number of k NN is used for the MR-resolution reconstructions due to their smaller voxel size.

As stated above KEM LVS improves the reconstruction of PET-unique regions present in MR-uniform regions, in comparison to the conventional KEM method. However, through using the conventional k NN KEM method and employing a narrower spatial Gaussian (small σ_r) (Eq. 3.9) a more localised basis function will be produced which will aid in the recovery of PET-unique regions. Similarly the conventional KEM method can be employed within

a reduced spatial neighbourhood (N_j is small), as another means of localising the basis functions and thus improving the reconstruction of PET-unique regions. Examples of such basis functions are provided in Figure 4.3. Although the conventional KEM method with a small σ_r or N_j does indeed localise the basis functions, it also restricts the extent of the basis function in general. Therefore, the denoising power of the conventional KEM method is lost to some extent. Taken to the extreme case (for illustrative purposes only), the conventional KEM method could be implemented with a σ_r value of 0, and a N_j value of 1. Both would result in basis functions consisting of the central voxel only. This has brought us back to the original voxel representation of MLEM, which as shown in the previous chapter is highly susceptible to noise. The proposed KEM LVS method, allows a more localised representation of the basis function to be attained, whilst maintaining its spatial extent and the number of contributing voxels. This preserves the denoising and edge-preserving power of the kernel method to a greater extent. A variety of other hyperparameter choices and KEM implementations (in addition to the ones just discussed) are provided in Figure 4.3, showing the basis functions formed in both a uniform MR region, and a white-grey matter boundary MR region.

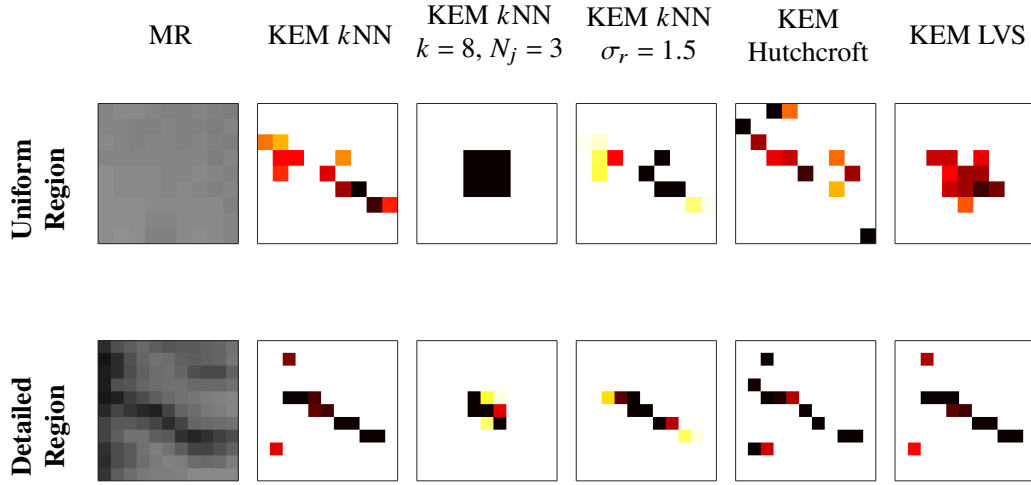


Fig. 4.3 Basis functions derived using either the conventional kernel method (KEM kNN), or the proposed spatially compact kernel method (KEM LVS). The impact of the different implementations on basis function shape is shown for a uniform (top row) and structured (bottom row) MR region. The method developed by Hutchcroft et al. [81] is also shown for comparison. The hyperparameters for each method are, unless otherwise indicated in each subfigure: $\sigma_f = 5$, $\sigma_r = 30$, $N_j = 9$, $p = 3$, and $k = 10$. Only the proposed spatially compact kernel method can deliver compact basis functions in uniform MR regions and also structured basis functions in detail containing MR regions for the same hyperparameters.

4.5 3D Simulated Phantom Study

4.5.1 Experimental Methods

In this chapter, several kernel implementations are analysed, the hyperparameters of which are stated in Table 4.1. The hyperparameter variation allows the resolution and sparsification changes to be systematically assessed, stratifying their individual and combined effects on the recovery of PET unique structures. The hyperparameter variation also ensures that the capabilities of the conventional KEM method are accurately represented. For instance, the conventional KEM method implementation that uses only 2 contributing voxels ($kNN = 2$) illustrates its performance in the limit of very small basis functions. As stated in §3.4 there is little consensus on the correct hyperparameters choice for the KEM method, and this varies

with the desired use. For instance in §3.4, $p = 1$ performed best for simulated data but $p = 3$ performed best for real data (where there is no ground truth data). In consideration of this, the KEM methods are initially investigated using both $3 \times 3 \times 3$ patches [81, 297] and single voxel feature vectors [295, 311, 296].

Algorithm	Voxel Size	k NN or LVS	Neighbourhood (N_j)	Patch Size (p)	Number of Voxels (k)
MLEM	PET	N/A	1	1	1
KEM	PET	k NN	5	1,3	6
KEM	PET	LVS	5	1,3	6
KEM	PET	k NN	3	1,3	3
KEM	MR	k NN	7	1,3	16
KEM	MR	LVS	7	1,3	16

Table 4.1 Algorithm and hyperparameters of the reconstruction methods investigated, for both simulated and real data. A increased neighbourhood size and number of contributing voxels were used for the MR voxel size kernel methods due to their smaller voxel size.

All methods are evaluated in terms of the trade-off between the image denoising power and the accurate reconstruction of PET-unique high intensity regions. The following error metrics of NRMSE (as previously defined in Eq. 3.15) and ROI mean $Mean_{ROI}(\boldsymbol{\theta})$ and ROI standard deviation $SD_{ROI}(\boldsymbol{\theta})$ were calculated, where the ground truth phantom ($\boldsymbol{\theta}_{GT}$) using MR size voxels is the reference image for the NRMSE calculation. The NRMSE and mean values were calculated for the three tumour regions to assess the ability of each KEM implementation to accurately recover PET unique regions. The whole brain NRMSE and white matter standard deviation were used to indicate the noise reduction properties of all methods. Images reconstructed at PET resolution were resampled to MR voxel size through trilinear interpolation for the calculation of the error metrics. All reconstructed images in this paper are shown at MR voxel size.

4.5.2 Results

The KEM LVS and KEM k NN methods were initially investigated using noise free data, to highlight the bias caused by each method across the PET unique regions. The difference between the reconstructed KEM method and the (noise-free) MLEM image are shown for a vertical profile across the image in Figure 4.4. This profile encompasses two PET unique regions (highlighted) that lie between voxel numbers 300-310 and 370-380, as such particular attention should be paid to large deviations from the ground truth (0 value) in these regions.

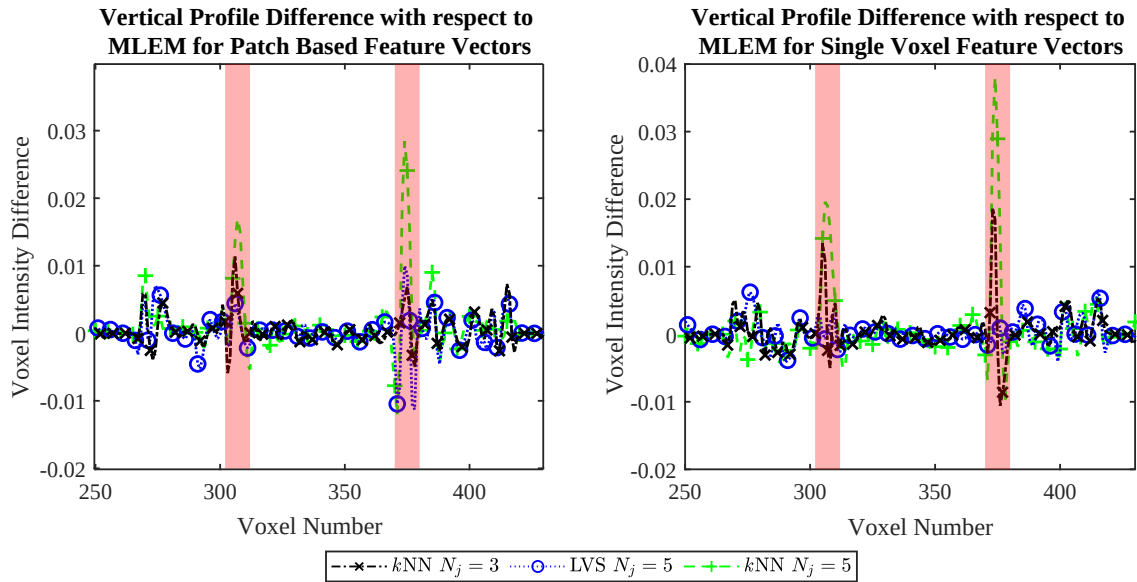


Fig. 4.4 Vertical line profiles of the difference between the reconstructed noise free simulated data after 500 iterations and the noise free MLEM image, through a tumour containing slice. The highlighted regions correspond to the top and bottom tumours on the left side of the brain, shown in Figure 4.2. Three PET resolution KEM method implementations are compared for both single voxel and patch feature vectors, the additional hyperparameter values are stated in Table 4.1.

Extending this initial comparison, the whole brain and combined tumour region NRMSE values are presented in Figure 4.5 for a wide variety of KEM hyperparameter values used in the reconstruction of the high count (10^8) simulated data. The corresponding results for the low count (10^7) simulated dataset are shown in Figure 4.6. Both single voxel and patch-based feature vector KEM methods return similar results for the reconstruction of the PET unique

structures in terms of NRMSE. In order to reduce the level of repetition and improve clarity the remainder of this chapter focuses on the single voxel feature vector implementation of the KEM method only.

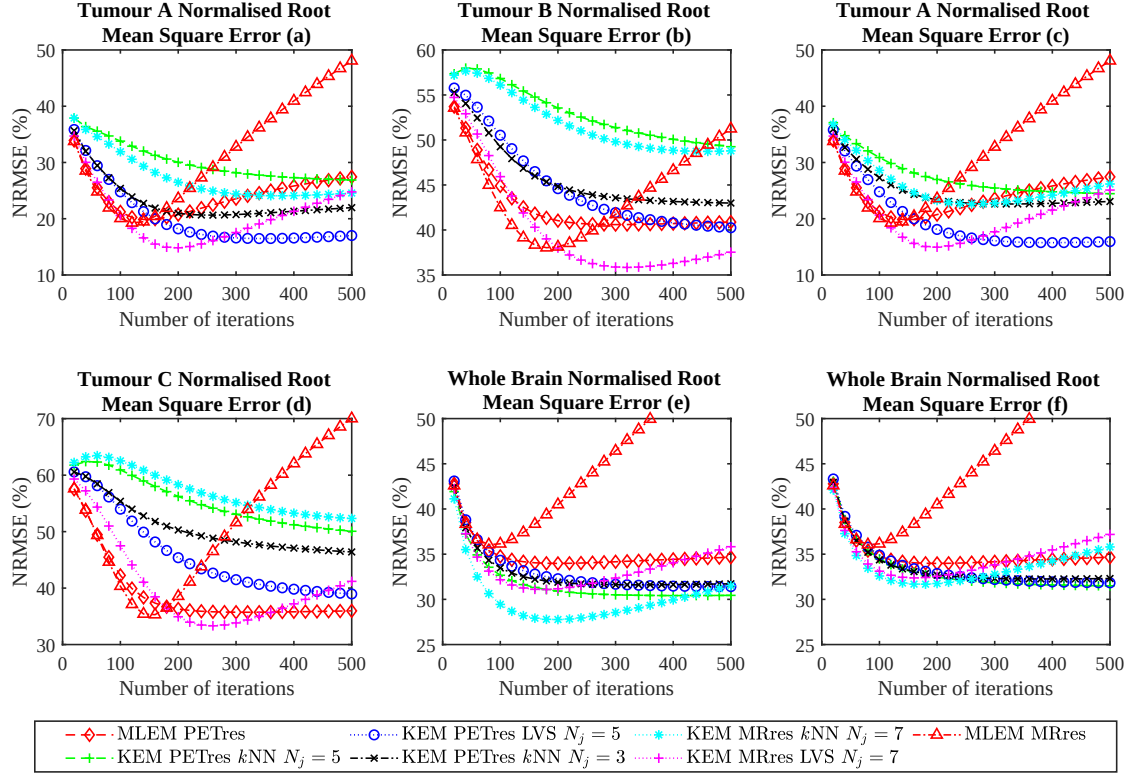


Fig. 4.5 NRMSE versus iteration number for the kernel variations stated in Table 4.1, for the higher count data (1×10^8). Figures shown for the three tumour regions of the simulated data shown in Figure 4.2, and the whole brain region. (a), (b), (d), and (e) First two columns use single voxel feature vectors in the kernel matrix formation and the (c) and (f) last column uses patch-based feature vectors.

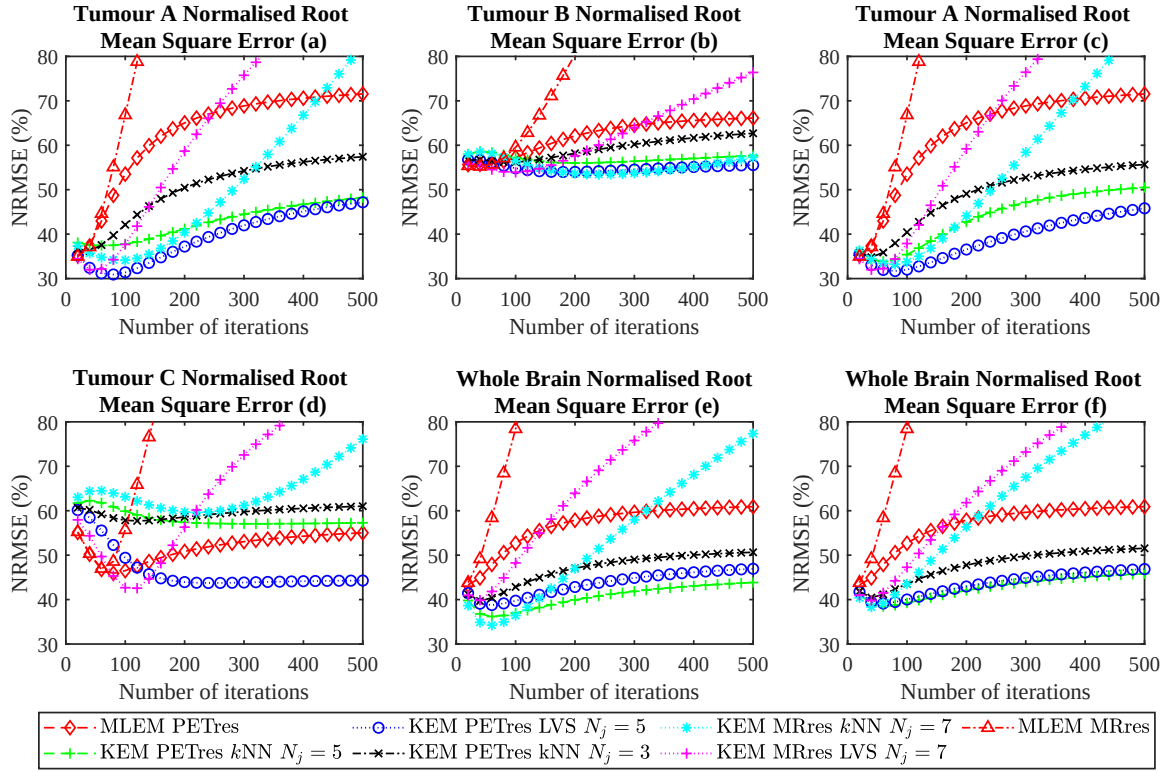


Fig. 4.6 NRMSE versus iteration number for the kernel variations stated in Table 4.1, for the lower count data (1×10^7). Figures shown for the three tumour regions of the simulated data shown in Figure 4.2, and the whole brain region. (a), (b), (d), and (e) First two columns use single voxel feature vectors in the kernel matrix formation and the (c) and (f) last column uses patch-based feature vectors.

Additional metrics are also investigated to enable a simple comparison between the high-count simulated dataset and the real patient dataset (both of which produce high count images). Figure 4.7 shows the tumour mean value with respect to the spatial standard deviation of a chosen white matter background region, for the higher count simulated data. Figure 4.8 shows the reconstructed images of the high count dataset for each tumour region at a fixed noise level. Figures 4.9 and 4.10 show the whole brain reconstructed images for the high count and low count simulated data, respectively.

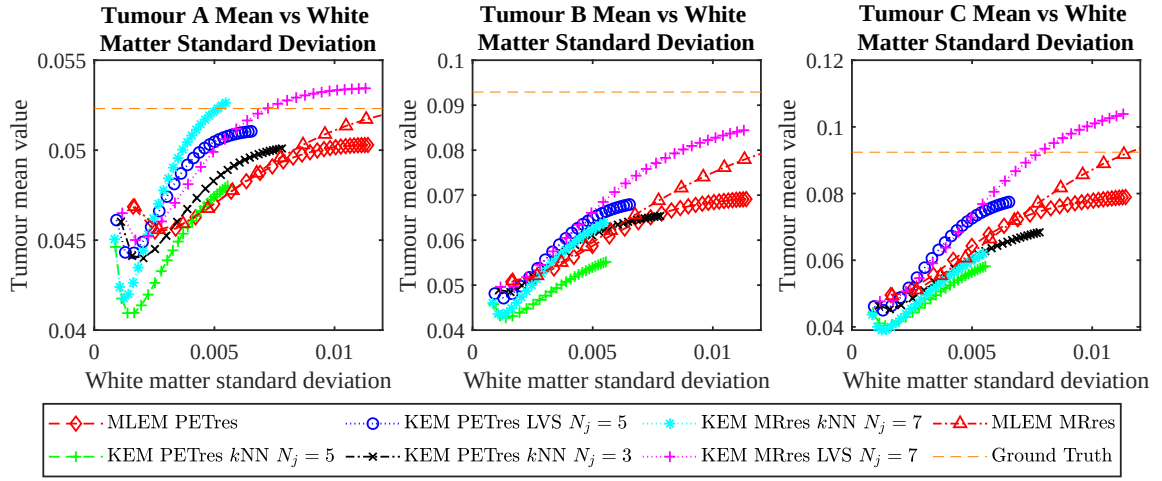


Fig. 4.7 Mean value of tumour regions A, B, and C (shown in Figure 4.2) versus the standard deviation of white matter region for the simulated high count (1×10^8) data. Ground truth mean values are labelled for reference.

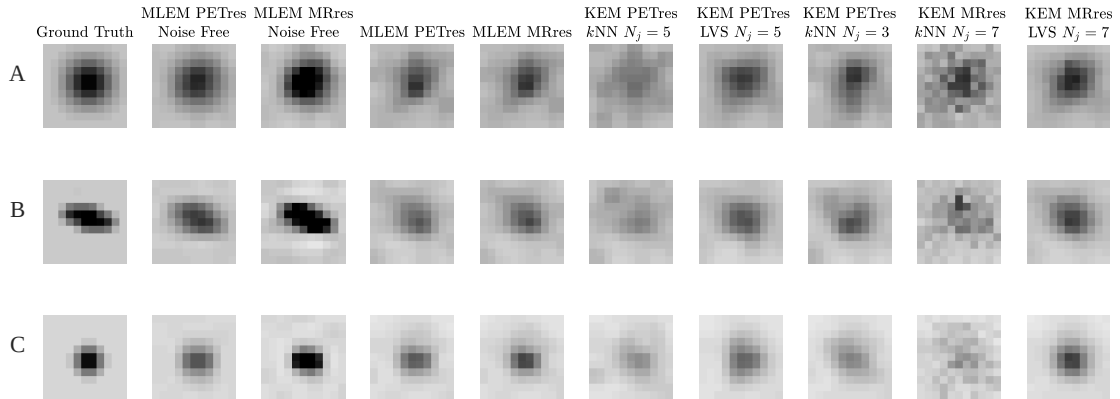


Fig. 4.8 Reconstructions of the three tumour regions for high count (1×10^8) simulated data, shown for each of the kernel methods (single voxel feature vectors) stated in Table 4.1, at a fixed noise level. Ground truth of the tumour regions are shown in the first column. The colour thresholds are different for each tumour, with a maximum intensity of 0.13, 0.15 and 0.2 for tumours A, B and C respectively.

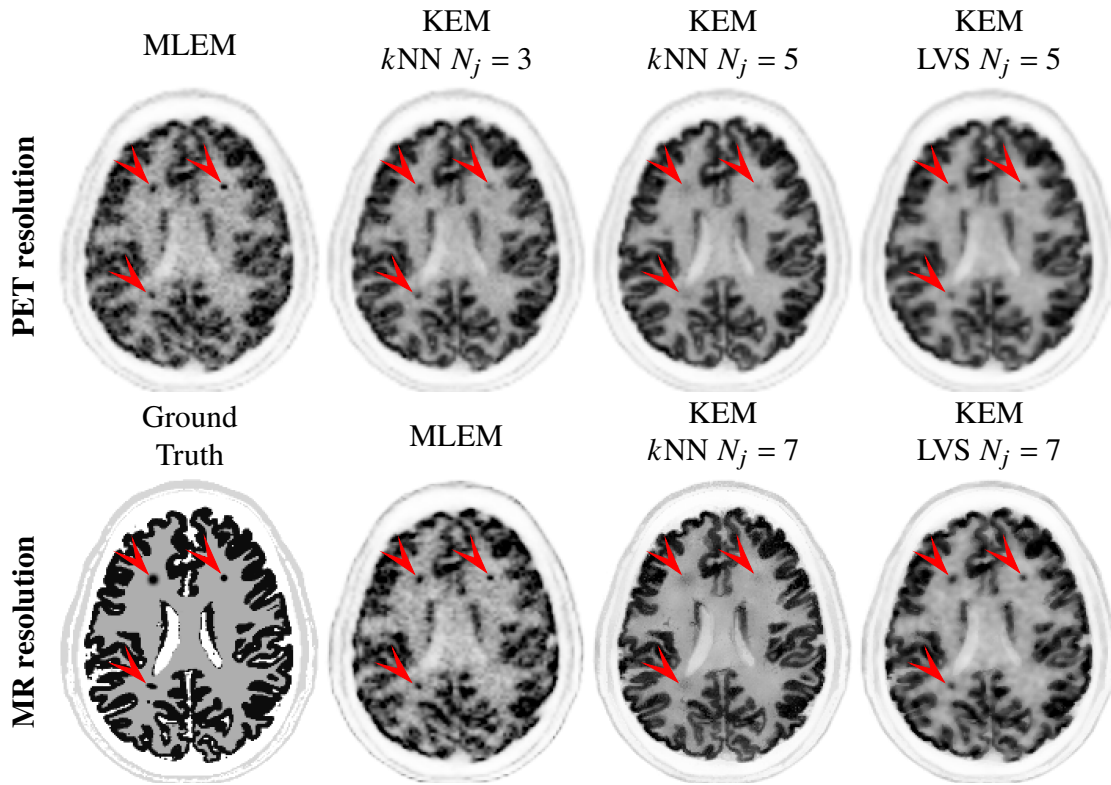


Fig. 4.9 Reconstructed images of the high count (1×10^8) simulated data for a tumour containing slice, for each of the kernel implementations (single voxel feature vectors) stated in Table 4.1. All images are shown at a fixed iteration number of 100. Red arrows are centered on the tumour regions.

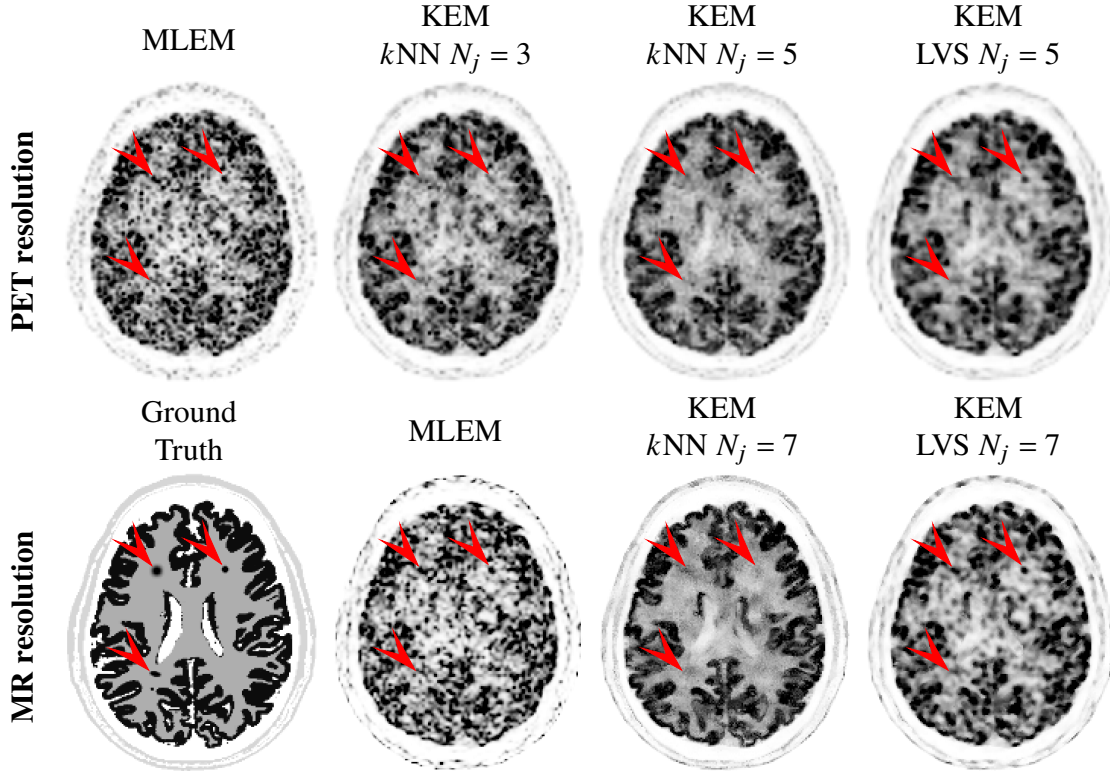


Fig. 4.10 Reconstructed images of the low count (1×10^7) simulated data for a tumour containing slice, for each of the kernel implementations (single voxel feature vectors) stated in Table 4.1. All images are shown at a fixed iteration number of 100. Red arrows are centered on the tumour regions.

4.6 3D Real Data Study

4.6.1 Experimental Methods

The real data was reconstructed for MLEM and the (single voxel feature vector) KEM implementations stated in Table 4.1, for up to 500 iterations. The reconstructed images were assessed in terms of tumour mean vs white matter standard deviation, and the visual appearance of the overall image and the synthetic tumour regions.

4.6.2 Results

Figure 4.11 shows the tumour mean vs white matter standard deviation curves, for the three simulated high intensity tumours added to the real patient dataset to provide known regions of PET-MR mismatch. An estimated ground truth mean value for each region is also shown on the curve. Figure 4.12 shows the reconstructed tumour regions for each reconstruction method at a fixed noise level. Lastly, the whole brain reconstructed images are shown for each method, for a tumour containing slice, at a fixed iteration number.

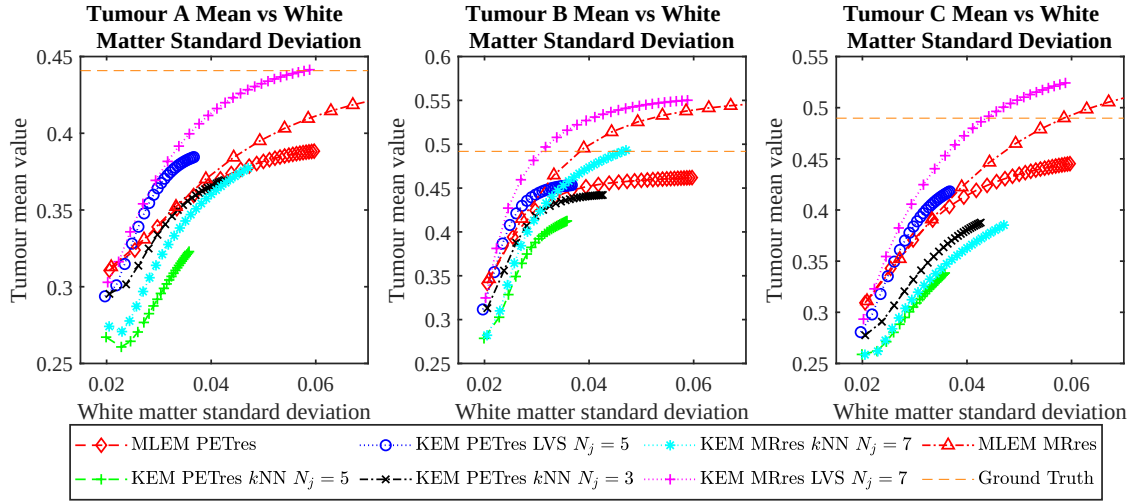


Fig. 4.11 Mean value of tumours A, B, and C versus the standard deviation of white matter region for the real data. The ground truth mean values were calculated from the real data through summing the simulated tumours mean to the mean value of the MLEM reconstruction of the white matter background region at 500 iterations. The ground truth values are only approximate, due to the large spatial standard deviation of the white matter background region (approximately 0.06) that the measured value is subject to.

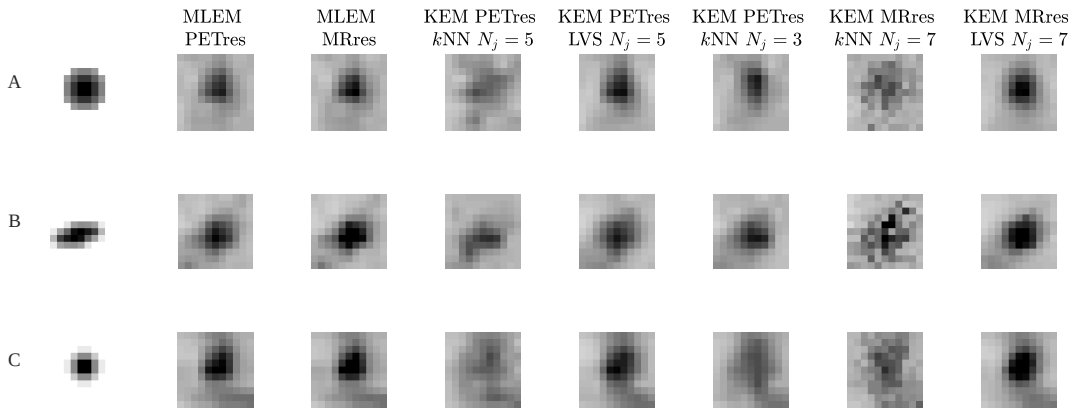


Fig. 4.12 Reconstructions of the three tumour regions for real data, shown for each of the kernel methods stated in Table 4.1, at a fixed noise level. The shape of each of the tumour regions are shown in the first column. The colour thresholds are different for each tumour, with a maximum intensity of 0.7, 0.8 and 0.6 for tumours A, B and C respectively.

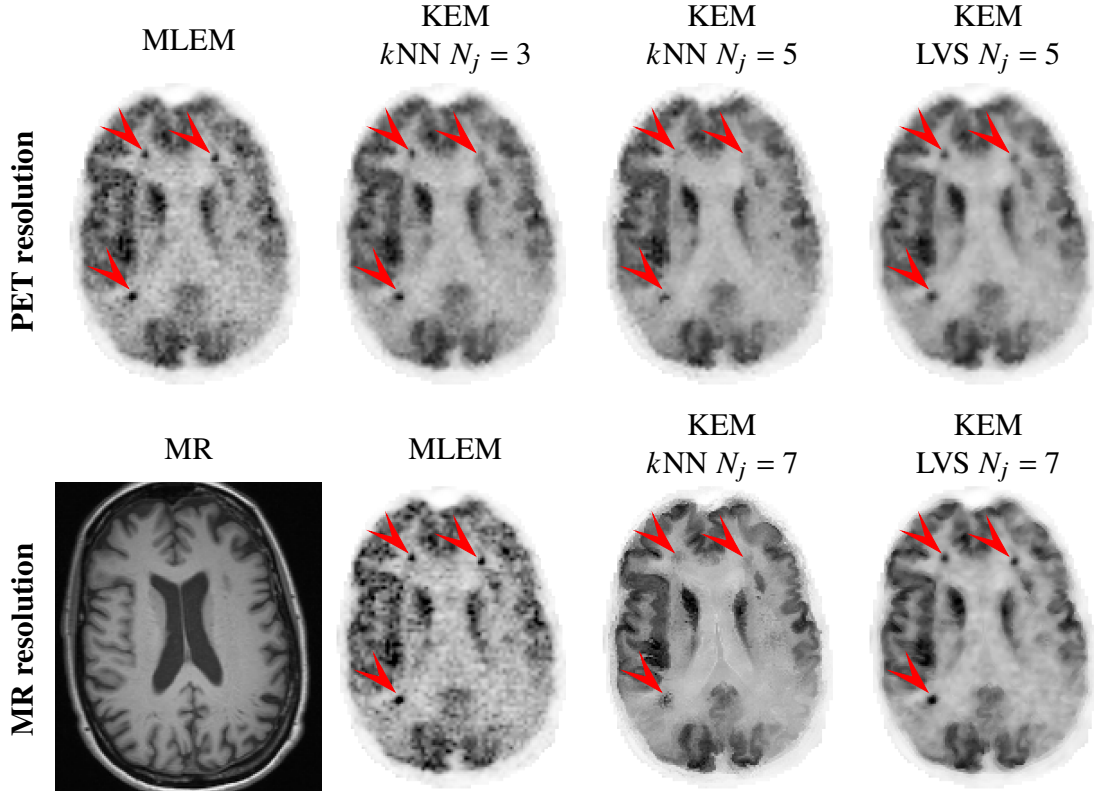


Fig. 4.13 Reconstructed images of the real data with the augmented tumours, for the tumour containing slice. All images are shown at 100 iterations. Red arrows are centered on the tumour regions

4.7 Discussion

Figure 4.4 shows that the proposed KEM LVS method clearly reduces the suppression of the PET-unique regions in comparison to the kNN KEM method with the same neighbourhood size, for PET resolution reconstructed images of noise-free data. The vertical difference profile is repeated for both single voxel and patch ($3 \times 3 \times 3$) feature vectors. Irrespective of the feature vector used, similar trends are shown in both, with the profiles of the kNN method deviating to a greater extent than the LVS method's profiles. Figure 4.5 shows the improved whole brain denoising properties of the conventional kNN PET resolution KEM method in comparison to MLEM, for the high count data (1×10^8). The conventional KEM

method achieves a lower whole brain NRMSE in comparison to MLEM for all iterations, for both single voxel and patch based feature vectors, reinforcing the results shown in Chapter 3. The level of noise reduction provided by the KEM method is furthered (at lower iterations) through constructing the basis functions from MR size voxels, due to greater recovery of detailed structures present in the ground truth, that are blurred at PET resolution. However, the k NN based KEM method demonstrates an increased error for the reconstruction of all PET unique features, relative to MLEM. The k NN KEM method's error in reconstructing tumour regions can be reduced through decreasing the neighbourhood size ($N_j = 3$) and number of contributing voxels ($k = 2$), but this is at the expense of increased whole brain NRMSE. The NRMSE of the tumour regions can also be reduced relative to the conventional KEM method, through using the KEM LVS method. At PET resolution, the KEM LVS method reconstructs the tumour regions with a comparable NRMSE to MLEM, whilst providing a reduced whole brain NRMSE. The LVS kernel method clearly outperforms the conventional kernel method in accurately reconstructing the tumour regions, although the whole brain NRMSE is notably lower for the conventional k NN kernel method. The KEM LVS method using MR size voxels leads to the lowest tumour NRMSE for all tumours in comparison to all other kernel methods, however the value is highly iteration number dependent. The performance of the KEM LVS methods also appears relatively invariant to the feature vector used (also shown in Figure 4.4), in both cases the PET resolution KEM LVS method achieves a low fixed NRMSE value for the whole brain and tumour regions.

The NRMSE curves of the reduced count data (1×10^7), Figure 4.6, further demonstrates the improvement of the proposed LVS kernel method at PET resolution relative to the conventional kernel method and MLEM. The increased noise present in this simulated dataset heightens the importance of the kernel method's noise suppression, leading to the LVS kernel method outperforming MLEM for all tumours in terms of NRMSE for both single voxel and patch-based feature vectors. However, all kernel and MLEM reconstructions

constructed at MR voxel size are shown to be highly susceptible to noise and tend further away from the ground truth at high iterations.

Returning to the high counts dataset, in Figure 4.7 the MR voxel size kernel reconstructions are shown to overshoot the ground truth mean value. This overshoot and undershoot of the ground truth value can also be seen for the MR resolution reconstruction of the noise free data, as shown in B. These oscillations occur near sharp intensity changes, and is likely due to Gibb's ringing. Therefore, although the MR voxel size LVS kernel method produces the lowest background noise level for a given mean value of tumour A, the MR voxel size kernel reconstruction methods are of little utility for quantitative reconstructions. A comparison of the PET resolution reconstruction methods, shows that for a given noise level the proposed PET resolution LVS kernel method attains a mean value closer to the ground truth than the conventional kernel method, and MLEM, for all three tumours. In addition, its tumour mean value tends to a very similar value to that achieved from MLEM, for all three tumours.

Visually the KEM LVS method produces reconstructions of the tumour regions that are similar to the MLEM reconstruction in terms of shape and intensity, as shown in Figure 4.8 at equal noise levels for the high count level data. In comparison, the k NN kernel method leads to notable blurring and distortion of the tumour regions, resulting in tumour regions that are visually dissimilar to the MLEM reconstructions. The reconstructed images are shown in Figure 4.9 (for high count data) and Figure 4.10 (for low count data), at a fixed iteration number for a tumour containing slice. In Figure 4.10, attention should be paid to tumour C which is still clearly visible for the KEM LVS method at PET resolution, which is not the case for the majority of other methods compared. In particular for the MLEM reconstruction tumour C is difficult to distinguish from the background noise. The MLEM reconstruction at MR voxel size suffers from noise to a greater extent than the PET resolution reconstruction, as shown in Figure 4.10. This highlights the potential risk of using an

unconstrained reconstruction method at MR voxel size, with the increased high intensity noise causing greater uncertainty in tumour detection.

In summary of the simulated data results, the proposed KEM LVS method at PET resolution reconstructs the PET unique structures with a reduced deviation from the ground truth (for noise free simulations) and a lower NRMSE (for noisy data), in comparison to the conventional k NN kernel method. Whilst at the same time the proposed KEM LVS method retains, to a large extent, the denoising properties and cortical detail present in the images reconstructed with the k NN kernel method. The mean value of the PET unique structures achieved for a given noise level can be further enhanced at MR voxel size; however, the mean values can be larger than the predicted ground truth.

For the augmented real dataset complementary results to the simulated dataset are observed. In Figure 4.11, the PET resolution KEM LVS method attains the highest tumour mean value for a given standard deviation, in comparison to all other PET resolution reconstruction methods. The tumour mean values of the KEM LVS method are also similar to that attained by MLEM for tumours A and B, indicating the accurate reconstruction of these PET-unique regions. These results are in agreement with the corresponding simulated data results (Figure 4.7), which also show the superior performance of the KEM LVS method, in comparison to all other PET resolution reconstruction methods. The ground truth mean values, shown in Figure 4.11, were calculated from the real data through summing the simulated tumours mean to the mean value of the white matter background region from the MLEM reconstructed image at 500 iterations. The ground truth values are only approximate, due to the large spatial standard deviation of the white matter background region that the measured value is subject to. Conversely to the simulated results, the mean value of the MLEM reconstructed image is slightly higher than the PET resolution LVS method for tumour region C. This demonstrates the shortcomings of the KEM LVS method for reconstructing small high intensity PET unique regions. These small high intensity PET unique regions would be better reconstructed

using MLEM, although due to the noise present in the MLEM reconstructions these regions could be confused with noise spikes, or blurred by post reconstruction smoothing. Further developments to the MR-informed kernel method such as the inclusion of PET data [296], shall be discussed in §5. The mean value of the tumour regions reconstructed with the MR voxel size kernel LVS method in Figure 4.11 outperforms the other kernel implementations and MLEM; however as shown in the simulation studies, the quantitative accuracy is limited. This is due to unconstrained reconstructed images at MR voxel size being more susceptible to increased levels of noise, and possible Gibbs-like artefacts which can alter the regional quantification. The use of kernel basis functions will reduce the impact of both these effects, however this depends on the size of the basis functions, and their shape.

Figure 4.12 shows that the conventional k NN kernel method at PET and MR voxel sizes visually suppresses the tumour regions in comparison to MLEM, at matched noise levels. The KEM LVS method's reconstruction of the tumour regions at PET resolution are visually closer to those reconstructed using MLEM. The improved performance of the KEM LVS method is also visually evident in Figure 4.13, in which the KEM LVS method reconstructs the PET-unique features in a similar manner to MLEM whilst maintaining the noise reduction and detail preserving properties of the kernel method. Figure 4.13 also highlights the visual improvements gained through reconstructing using MR size voxels, as shown by the MR-resolution KEM k NN image.

Both simulated and real data results demonstrate the benefit of the spatially compact KEM LVS method for the reconstruction of PET unique structures, relative to the conventional kernel method. In both datasets, there is a clear trade-off between the spatially compact and conventional kernel methods, with the former achieving more accurate reconstruction of PET unique structures, and the later achieving further reduction in whole brain NRMSE.

4.8 Limitations

This chapter has demonstrated the success of a proposed adaptation to the kernel method in reducing the suppression of PET-MR mismatches, where the PET feature is present in a uniform MR region. However, as discussed in Chapter 3 there is a variety of other PET-MR mismatches not investigated here, these include MR structures not present in the PET, and structures in both the PET and MR that do not match. The kernel method is practically insensitive to MR structures in uniform PET regions, as the incorrectly shaped basis functions will be correctly weighted using the PET data only. But, differing PET and MR structures will lead to distortion of the PET structure in accordance to the shape of the MR. To accurately reconstruct contradicting PET-MR structures requires further investigation and development of the kernel method. In addition, the KEM LVS and other MR-informed kernel methods need to be assessed with respect to state-of-the-art MAPEM methods, to identify the relative merits of these two branches for including MR information into PET reconstruction. Based on the preceding work undertaken, evaluating these methods in respect to the trade-off between PET-unique feature reconstruction and noise suppression would be of particular interest. Lastly, the proposed kernel method should be used to reconstruct real data which has genuine PET-MR mismatches (e.g., tumours only present in the PET), to assess the full utility of the proposed method.

4.9 Summary

As demonstrated in this chapter and previously shown in the literature, the conventional MR-informed kernel method may lead to excessive smoothing of structures that are only present in the PET data. This chapter makes two main contributions in order to address this problem: first, we exploit the potential of the MR-informed kernel method to form more spatially compact basis functions which are able to preserve PET-unique structures, and

second, we consider reconstruction at the native MR voxel size. The former contribution of using spatially compact basis functions demonstrated clear improvements in the recovery of PET unique structures compared to the conventional kernel method for both simulated and real data. For all sizes of PET unique structures evaluated, the spatially compact kernel method showed quantitative and visual improvement in comparison to the conventional kernel method whilst maintaining to a large extent the kernel method's improved noise reduction properties relative to MLEM. This provides an improved balance between noise reduction, and recovery of PET unique structures. Further developments to the MR-informed kernel method are required to eliminate the bias introduced by the kernel method in the reconstruction of PET unique structures, in particular for very small high intensity regions. Extending the kernel method to MR voxel size resulted in visually more appealing images, especially in terms of grey-white matter delineation, however the quantification of tumour regions was negatively effected due to the presence of Gibbs-like artefacts. Future work may look to assess the current performance of the KEM LVS method with respect to other MR-informed KEM and MAPEM based methods, or alternatively extend the kernel method further to prevent the suppression of all forms of PET-unique regions discussed above.

Chapter 5

Comparison of MR-Informed Reconstruction Methods

Numerous image reconstruction methodologies for PET imaging have been developed that incorporate MR information into the reconstruction process via regularisation (PML or MAP). The kernel method as presented in chapter 3 is an alternative method for including MR information into the reconstruction process. Due to developments made to the kernel method (chapter 4 and [311]) a variety of kernel implementations are also present in the literature. These kernel methods are part of a broader cohort of algorithms that re-parameterise the objective function, constraining the image to be comprised from a finite number of MR-informed spatial basis functions.

Both regularised and reparameterised MR-informed reconstruction methods produce images with suppressed noise and reduced partial volume effects. However, as discussed in §4.8 little comparison has been made between methods that include MR-information via re-parameterisation as opposed to the more well-established method of regularisation. Further investigation of MR-informed methods is also required in general to determine the potentially negative influence of including MR information in the reconstruction of PET features that are MR negative (not present in the MR image). In particular, this chapter

focuses on the trade-off between whole brain image quality and the faithful recovery of structures unique to the PET data. The ideal MR-informed PET reconstruction method would universally improve the reconstructed images, for shared and discrepant PET-MR regions alike. The current PET update image θ^n can also be included into the guidance process (referred to in this work as PET-MR-informed), and has been proposed as one potential solution for ameliorating PET-MR differences. A subset of PET-MR-informed methods are therefore also included in the comparison of MR-informed methods. This Chapter is based on work previously published by the author during the course of his PhD studies [318].

5.1 Reparameterisation vs Regularisation

The methods under comparison in this chapter include MR information via a spatial similarity matrix within a local spatial neighbourhood. These methods can all be referred to as MR-informed. These MR-informed methods can be divided into two categories:

- Reparameterising the objective function with an MR-informed kernel matrix.
- Regularising the objective function with MR-informed weights (of a quadratic inter-voxel potential function).

The MR(only)-informed kernel methods under comparison are those discussed in Chapters 3 and 4, which shall be referred to as KEM and KEM LVS, respectively. The MR(only)-informed regularisation methods under comparison are the Gaussian similarity kernel [319, 302] and the asymmetric Bowsher prior [224], due to their enduring popularity, and ability to match the performance of more involved MR-informed methods [64, 71, 65]. As shown in [64] the inclusion of anatomical information into the weighting factors results in well recovered PET-MR boundaries irrespective of the inter-voxel potential function used. Therefore, the regularised methods under comparison in this chapter shall use a weighted

quadratic potential function, for simplicity. The numerous benefits of MR-informed regularised and reparameterised reconstruction methods have been illustrated in chapters 2 - 4, however such methods are rarely compared side by side. In addition, the inclusion of MR-information into the reconstruction process has the tendency to suppress or blur high intensity lesions that are not present in the MR image (MR negative) [81, 64]. This is a major pitfall for the visual diagnostic interpretation of PET images, in particular for oncological cases, where PET imaging has been shown to improve the diagnosis and subsequent treatment of cancers [5]. Although, it should be recognised that such MR-negative lesions are uncommon. One potential avenue currently under investigation for reducing the suppression of these high intensity PET-unique regions is to extend the guidance to include the reconstructed PET image at each iteration. This concept was implemented firstly via MAP (regularisation) [226, 320, 321] and has recently been extended to a kernel (re-parameterisation) [311, 296] implementation. Such methods shall be referred to as PET-MR-informed, from which this work shall compare the anato-functional (a MAP method) [320], and the hybrid kernel method (HKEM) [311]. The language used to describe such methods is illustrated in a Venn diagram, as shown in Fig. 5.1.

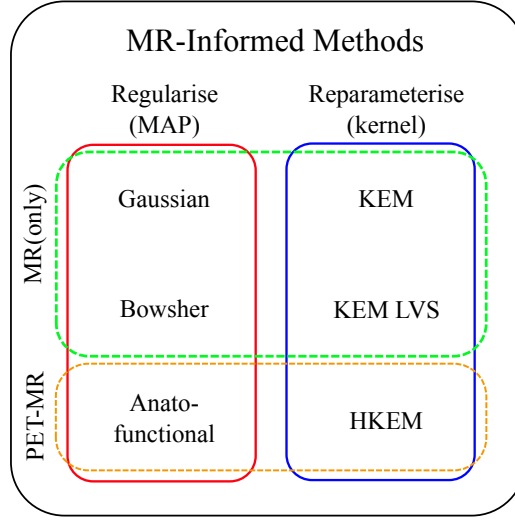


Fig. 5.1 Schematic showing the categorisation of the different MR-informed methods under comparison.

Alternative PET-MR-informed methods are present in the literature as discussed in §2.8. These include the joint Shannon entropy [199, 231] and parallel level sets priors [227, 228]. Whilst such methods remain active within the PET reconstruction field, they do not lie within the same group of algorithms where the weights can be expressed in the form of Eq. 5.5. Also such methods have been implemented in a regularised fashion only and lack a corresponding reparameterisation method for comparison.

5.2 Theory

Whilst the majority of the underlying theory has been discussed in previous chapters a concise summary of the methods under comparison shall be presented here. The Gaussian, Bowsher and anato-functional MAP methods under comparison are all implemented using a weighted quadratic prior. The MAP objective function and the weighted quadratic prior are shown below:

$$\boldsymbol{\theta}^{MAP} = \arg \max_{\boldsymbol{\theta}} \sum_i^{N_p} m_i \ln ([\mathbf{A}\boldsymbol{\theta}]_i) - [\mathbf{A}\boldsymbol{\theta}]_i - \beta R(\boldsymbol{\theta}) \quad (5.1)$$

$$R(\boldsymbol{\theta}) = \sum_j^{N_v} \sum_{l \in \eta_j} w_{jl} \psi(\theta_j - \theta_l) = \sum_j^{N_v} \sum_{l \in \eta_j} w_{jl} (\theta_j - \theta_l)^2 \quad (5.2)$$

Where $\mathbf{A}$ is the system matrix R is the prior term, β is the hyperparameter, w_{jl} are the MR-informed weighting factors and ψ is the potential function. The De Pierro surrogate form of the prior term is used to give a separable update solution for the MAP objective function, as shown in §2.5 Eq. 2.33-2.36. The De Pierro algorithm guarantees convergence to the maxima of the MAP objective function unlike alternative algorithms such as one step late. The KEM, KEM LVS and HKEM kernel methods under comparison are all implemented using a simple reparameterisation of the maximum likelihood objective function. The kernelised objective function is shown below:

$$\boldsymbol{\alpha}^{ML} = \arg \max_{\boldsymbol{\alpha}} \sum_i^{N_p} m_i \ln([\mathbf{AK}\boldsymbol{\alpha}]_i) - [\mathbf{AK}\boldsymbol{\alpha}]_i \quad (5.3)$$

$$\boldsymbol{\theta}^{KEM} = \mathbf{K}\boldsymbol{\alpha}^{ML} \quad (5.4)$$

where $\mathbf{K}$ is the kernel basis matrix and $\boldsymbol{\alpha}$ is the basis function coefficients. For the six MR-informed methods under comparison, the MAP weighting factors and kernel basis functions used are strongly associated with one another. Therefore, the MAP weighting factors and the kernel basis functions discussed can all be generalised as a Gaussian term, followed by a k NN sparsification. This formalisation is used to unify and highlight the similarities of such methods. The generalised Gaussian term is :

$$w_{jl}^n \text{ or } K_{jl}^n = \exp \left(-\frac{\|\mathbf{z}_j(\mathbf{x}, \boldsymbol{\theta}^n, \mathbf{r}) - \mathbf{z}_l(\mathbf{x}, \boldsymbol{\theta}^n, \mathbf{r})\|_2^2}{2} \right) \quad (5.5)$$

where the correct expression of $\mathbf{z}(\mathbf{x}, \boldsymbol{\theta}^n, \mathbf{r})$ is given in Table 5.1 for each method. A subset of these methods undergo a k NN sparsification of the weighting factors (or basis functions), as described in Table 5.1. The resultant MAP weighting factors or KEM basis functions are also shown in Table 5.1. Schematics illustrating how the kernel basis functions and MAP

weighting factors are produced for each method under comparison are shown in Fig. 5.2 and Fig. 5.3 respectively.

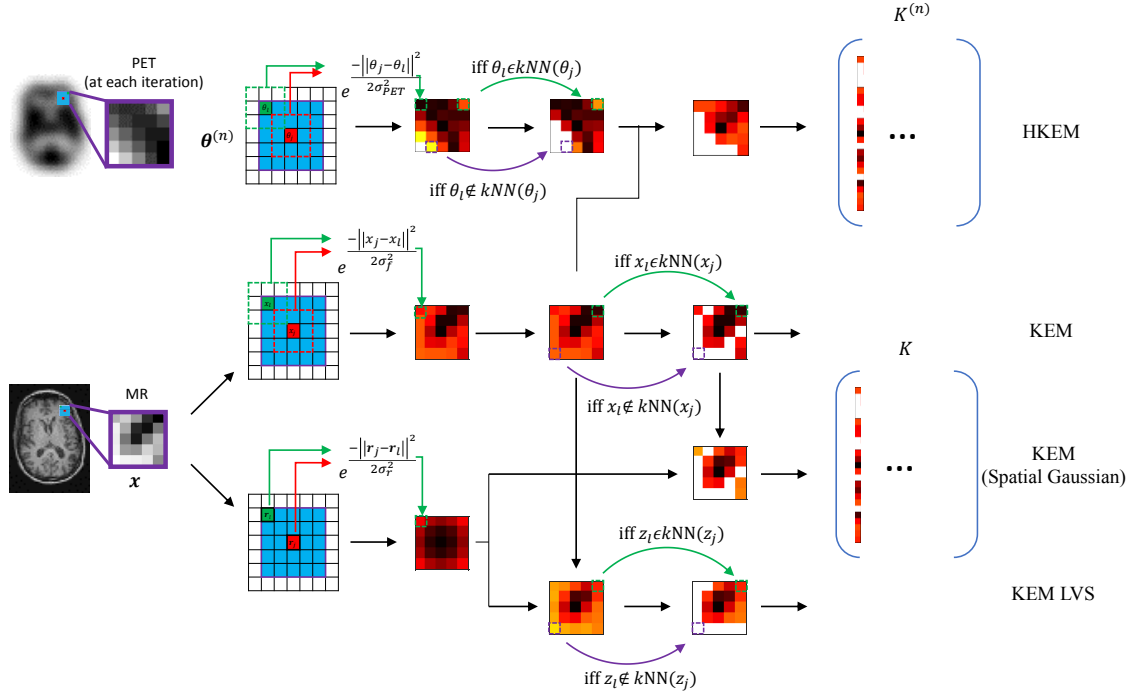


Fig. 5.2 Schematic showing the formation of the 3 MR-informed kernel methods under comparison. The neighbourhood comparison is shown for the generalised patch case and z for the KEM LVS method corresponds to its composite feature vector as defined in Table 5.1

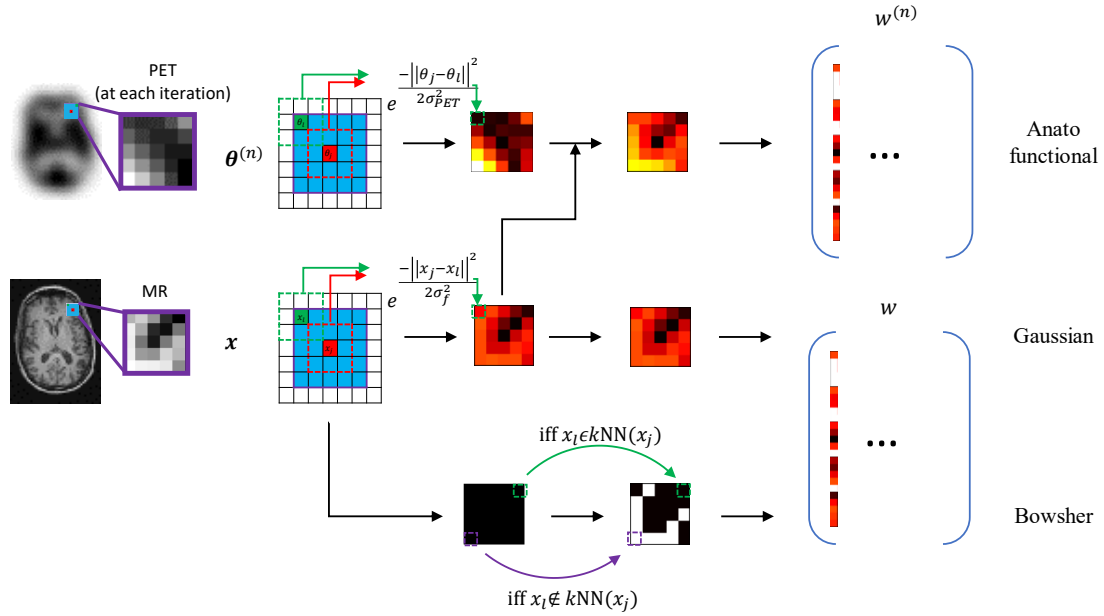


Fig. 5.3 Schematic showing the formation of the 3 MR-informed MAP methods under comparison. The neighbourhood comparison is shown for the generalised patch case.

Method	$\mathbf{z}_j(\mathbf{x}, \boldsymbol{\theta}^n, \mathbf{r})$	MR derived weights	PET derived weights	k -nearest neighbours
Regularisation				
Gaussian	$\frac{x_j}{\sigma_{MR}}$	$\exp\left(-\frac{\ x_j - x_j\ ^2}{2\sigma_{MR}^2}\right)$	1	use all neighbours
Bowsher	0	1	1	w.r.t x_j
Anato-functional	$\left[\frac{x_j}{\sigma_{MR}}, \frac{\theta_j^n}{\sigma_{PET}}\right]$	$\exp\left(-\frac{\ x_j - x_j\ ^2}{2\sigma_{MR}^2}\right)$	$\exp\left(-\frac{\ \theta_j^n - \theta_j^n\ ^2}{2\sigma_{PET}^2}\right)$	uses all neighbours
Reparameterisation				
KEM	$\frac{x_j}{\sigma_{MR}}$	$\exp\left(-\frac{\ x_j - x_j\ ^2}{2\sigma_{MR}^2}\right)$	1	w.r.t x_j
KEM LVS	$\left[\frac{x_j}{\sigma_{MR}}, \frac{\mathbf{r}_j}{\sigma_r}\right]$	$\exp\left(-\frac{\ x_j - x_j\ ^2}{2\sigma_{MR}^2}\right) \exp\left(-\frac{\ \mathbf{r}_j - \mathbf{r}_j\ _2^2}{2\sigma_r^2}\right)$	1	w.r.t $\left[\frac{x_j}{\sigma_{MR}}, \frac{\mathbf{r}_j}{\sigma_r}\right]$
HKEM	$\left[\frac{x_j}{\sigma_{MR}}, \frac{\boldsymbol{\theta}_j^n}{\sigma_{PET}}\right]$	$\exp\left(-\frac{\ x_j - x_j\ ^2}{2\sigma_{MR}^2}\right)$	$\exp\left(-\frac{\ \boldsymbol{\theta}_j^n - \boldsymbol{\theta}_j^n\ ^2}{2\sigma_{PET}^2}\right)$	w.r.t $\boldsymbol{\theta}_j^n$

Table 5.1 The 6 MR-informed methods under comparison are shown in this table, with the appropriate feature vector of the generalised Gaussian (Eq. 5.5) stated for each method. An representative kernel basis function or MAP weights are shown at the end of each row, with darker colours corresponding to higher intensity values.

5.3 Simulated Phantoms and Patient Data

5.3.1 3D Simulated Phantom

A 3D simulated [^{18}F]FDG PET phantom was constructed with a voxel-side length of 1 mm, based on the BrainWeb [306] segmented MR database. As before, the phantom is constructed in a similar manner to §3.3.1. However, real PET images have more structural variation than the piece-wise constant simulated phantoms used in the preceding chapters. To discourage overly piece-wise constant images, Gaussian smoothed random structures were incorporated into the simulated PET phantom in this chapter, in accordance to Eq. 5.6 [322], producing more varied tissue structure:

$$\theta_{rand} = \theta(1 + H(2G(rand, \sigma_{sm}) - 1)) \quad (5.6)$$

where $rand$ are uniformly distributed random numbers between 0 and 1 for each image voxel, G corresponds to convolution with a Gaussian kernel of width σ_{sm} , and H is an amplitude parameter. Four high intensity structures (lesions) were added to the PET phantom (only). The two smaller lesions are located in the white matter, whereas the larger lesions are positioned across MR boundaries that differ (Fig. 5.4). Profiles through each of these PET-unique regions are also shown in Fig. 5.4, with the intensity of Tumour C and D equal to three times the intensity of the grey matter. The PET phantom was projected into a span 11 sinogram, and rescaled to a high count level with a total prompts count of 10^8 , and a lower count level with a total prompts count of 10^7 . This was repeated for 10 noise realisations. The simulated sinograms were reconstructed at mMR resolution (voxel size of $2.09 \times 2.09 \times 2.03\text{mm}^3$), with resolution modelling (PSF) included in the reconstruction.

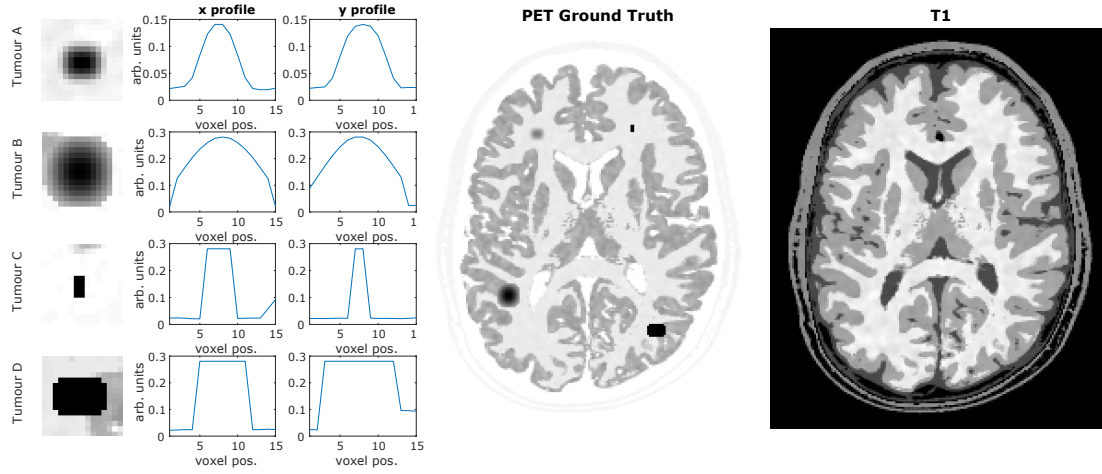


Fig. 5.4 (left) Simulated PET-unique high intensity regions (Tumours A, B, C and D) that differ in structure to the corresponding regions in the T1 MR phantom. (centre) Simulated FDG PET phantom based on the BrainWeb phantom, with the four PET-unique high intensity regions added. (right) T1 image. All images shown at MR resolution. The grey matter intensity value shown in the tumour profiles is approximately 0.1 arbitrary units.

5.3.2 3D PET Patient Scan Dataset

Real [^{18}F]FDG data from a patient scan was reconstructed using all seven methods previously considered. The dataset was acquired from the Siemens Biograph mMR simultaneous PET-MR scanner, allowing simple acquisition of co-registered PET-MR data. The [^{18}F]FDG scan had a total prompt count of 4.69×10^8 and a scan duration of 23 minutes. The tracer activity at time of injection (81 minutes prior to start of image acquisition) was 229 MBq. The data acquisition and reconstruction process is the same as in §3.3.2, unless otherwise stated. A T1 MPRAGE scan provided the anatomical image, which was resampled to the PET resolution ($2.08626 \text{ mm} \times 2.08626 \text{ mm} \times 2.03125 \text{ mm}$) for use in anatomical guidance. Four simulated tumours were added to the FDG dataset to give known PET-unique regions relative to the corresponding MR structure. The intensity profiles of the tumour regions are shown in Fig. 5.5 (the PET intensity of the grey matter is approximately 0.4 arb. units). This augmented

dataset is reconstructed at the normal full (100%) count level and then also resampled to a lower count level of 10% (relative to the total prompt count level).

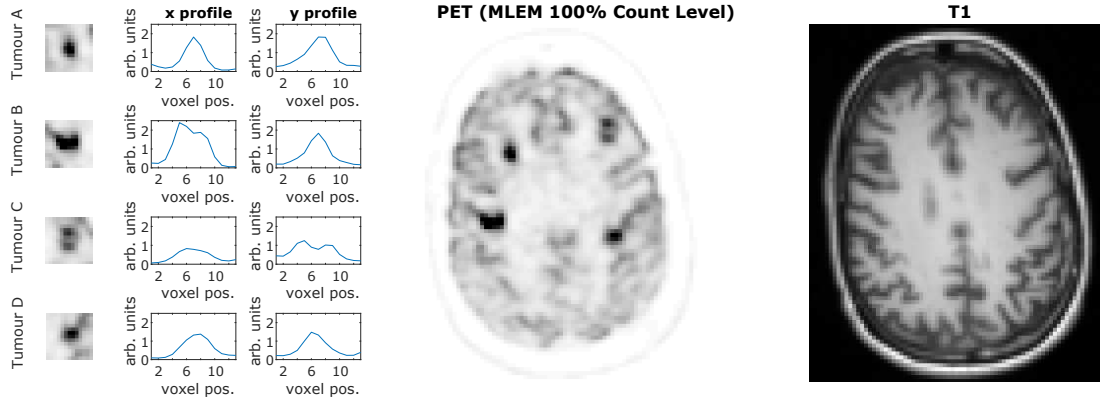


Fig. 5.5 Real patient FDG dataset, augmented with simulated tumour regions, to produce genuine regions of PET-MR mismatch. (left) Tumour regions and profiles from the MLEM reconstructed image for the 100% count level. (middle) MLEM reconstructed image of the whole brain region, shown at 300 iterations for the 100% count level. (right) T1 MPRAGE image. All images shown at PET resolution. The grey matter PET intensity value is approximately 0.4 arb. units.

5.4 Hyperparameter Selection

All methods underwent a restricted grid search of the parameter range, for both the high count and low count simulated datasets. All images were reconstructed up to 300 iterations, with images resized to the original 1 mm resolution of the ground truth phantom for the calculation of the error metrics. The structural similarity index (SSIM) [323] was used to determine the similarity between the reconstructed images and the ground truth, as is generally perceived to provide a closer numerical representation of visual perception. SSIM has previously been used in the PET and image processing literature to evaluate novel methodologies [227, 324, 249], and to assess structural similarity between multi-modality images as part of a proposed methodology[301], amongst other uses. The simplified formulation of SSIM

employed in this work is:

$$SSIM(\boldsymbol{\theta}, \boldsymbol{\theta}^{GT}) = \frac{(2\bar{\boldsymbol{\theta}}^{GT}\bar{\boldsymbol{\theta}} + C_1)(2\boldsymbol{\sigma}_{\boldsymbol{\theta}^{GT}\boldsymbol{\theta}} + C_2)}{(\bar{\boldsymbol{\theta}}^2 + (\bar{\boldsymbol{\theta}}^{GT})^2 + C_1)(\boldsymbol{\sigma}_{\boldsymbol{\theta}}^2 + \boldsymbol{\sigma}_{\boldsymbol{\theta}^{GT}}^2 + C_2)} \quad (5.7)$$

where $\bar{\boldsymbol{\theta}}$ refers to the spatial local mean, $\boldsymbol{\sigma}_{\boldsymbol{\theta}}$ refers to the local standard deviation, and $\boldsymbol{\sigma}_{\boldsymbol{\theta}^{GT}\boldsymbol{\theta}}$ is the cross-covariance between the image $\boldsymbol{\theta}^{GT}$ and $\boldsymbol{\theta}$. C_1 and C_2 are constant values dependent on the dynamic range of the image. The parameters for each reconstruction method were compared for their ability to improve the SSIM metric, for regions where the PET and MR phantom notably differ (PET unique regions), and for the whole brain region. The accurate reconstruction of the whole brain region normally comes at the expense of the accurate reconstruction of PET unique regions. The SSIM of the whole brain region is plotted against the SSIM of the PET unique regions to produce a trade-off curve. The parameter values that produced the SSIM trade-off curve that was closest to the SSIM value of 1 for both the whole brain and PET unique regions, i.e. produced the largest sum of the unique SSIM squared and whole brain SSIM squared, was selected for each reconstruction method. The chosen parameters for each reconstruction method are presented in Table 5.2. It should be noted that the performance of the PET-MR-informed methods of HKEM, and anato-functional were found to be highly sensitive to the chosen parameters (in particular $\boldsymbol{\sigma}_{PET}$) and the noise level of the PET image under reconstruction.

Parameters for High Count Data				Parameters for Low Count Data			
Method	Fixed	Varied	Chosen	Fixed	Varied	Chosen	
Regularisation	Gaussian	N/A	$\sigma_{MR} = 0.01 - 5$	N/A	$\sigma_{MR} = 0.01 - 5$	$\sigma_{MR} = 0.5$	
			$\beta = 0 - 10^6$		$\beta = 0 - 10^7$		
	Bowsher	N/A	$k = 0 - 60$	N/A	$k = 0 - 60$	$k = 40$	
			$\beta = 0 - 10^6$		$\beta = 0 - 10^7$		
Reparameterisation	Anato-functional	$\sigma_{MR} = 0.1$	$\sigma_{PET} = 0.001 - 0.5$	$\sigma_{PET} = 0.01$	$\sigma_{PET} = 0.001 - 0.5$	$\sigma_{PET} = 0.5$	
			$\beta = 0 - 10^6$		$\beta = 0 - 10^7$		
	KEM	N/A	$\sigma_{MR} = 0.01 - 5$	N/A	$\sigma_{MR} = 0.01 - 5$	$\sigma_{MR} = 0.5$	
			$k = 0 - 60$		$k = 0 - 60$		
HKEM	KEM LVS	$\sigma_s = 40$	$\sigma_{MR} = 0.01 - 5$	$\sigma_s = 40$	$\sigma_{MR} = 0.01 - 5$	$\sigma_{MR} = 0.5$	
			$k = 0 - 60$		$k = 0 - 60$		
		$\sigma_{MR} = 0.1$	$\sigma_{PET} = 0.001 - 0.5$	$\sigma_{MR} = 0.5$	$\sigma_{PET} = 0.001 - 0.5$	$\sigma_{PET} = 0.5$	
		$k_{MR} = 124$	$k_{PET} = 0 - 124$	$k_{MR} = 124$	$k_{PET} = 124$		

Table 5.2 Summary of each MR-informed method’s parameters. Each method underwent a restructured grid search, using a 10^8 and 10^7 count 3D simulated dataset, to determine the chosen parameters, which gave the best whole brain to PET-unique region SSIM tradeoff. The range of values over which the grid search was evaluated, and the chosen parameters are stated.

5.5 Simulation Study

This chapter focuses on the trade-off between whole brain image quality and the faithful recovery of structures unique to the PET data, evaluated in terms of structural similarity (SSIM) normalised root mean square error (NRMSE), bias and mean values.

5.5.1 Experimental Methods

The chosen parameter ranges for each reconstruction method were employed to reconstruct 10 noise realisations of the simulated data, for both the 10^8 and 10^7 count levels. The reconstructed images were assessed using a variety of error and image quality metrics for individual voxels and across regions of interest (ROI), to determine which MR-informed method provided the best trade-off between the reconstruction of the whole brain and PET-unique regions. The values calculated for SSIM (Fig. 5.6 and 5.9) are now for multi-noise realisation data and were averaged across all noise realisations. The voxel wise metrics of mean, bias and standard deviation Fig. C.1 and 5.9) are calculated across the multiple noise realisations (N_{noise}) in accordance to the following equations:

$$\bar{\theta}_j = \frac{1}{N_{noise}} \sum_m^{N_{noise}} \theta_{j,m} \quad (5.8)$$

$$Bias_j = \bar{\theta}_j - \theta_j^{GT} \quad (5.9)$$

$$\sigma_j = \sqrt{\frac{1}{N_{noise}} \sum_m^{N_{noise}} (\theta_{j,m} - \bar{\theta}_j)^2} \quad (5.10)$$

where $\bar{\theta}_j$ now represents the mean value of θ_j averaged over multiple noise realisations. ROI based error metrics of bias and standard deviation are also calculated across multiple noise realisations. These metrics are an extension of the metrics above; the metric value (squared) is now summed over each voxel within the ROI and normalised according to the ground truth. ROI bias vs standard deviation curves are presented for the whole brain region, and each of

the PET unique regions (Fig. 5.6, 5.9, C.1 and C.2).

$$Bias_{ROI}(\%) = \sqrt{\frac{\sum_j^{N_{ROI}} (\bar{\theta}_j - \theta_j^{GT})^2}{\sum_j^{N_{ROI}} (\theta_j^{GT})^2}} \times 100 \quad (5.11)$$

$$\sigma_{ROI,1}(\%) = \sqrt{\frac{\sum_m^{N_{noise}} \sum_j^{N_{ROI}} (\bar{\theta}_j - \theta_{j,m})^2}{N_{noise} \sum_j^{N_{ROI}} (\theta_j^{GT})^2}} \times 100 \quad (5.12)$$

$$NRMSE(\%) = \sqrt{Bias_{ROI}^2 + \sigma_{ROI,1}^2} = \sqrt{\frac{\sum_m^{N_{noise}} \sum_j^{N_{ROI}} (\theta_{j,m} - \theta_j^{GT})^2}{N_{noise} \sum_j^{N_{ROI}} (\theta_j^{GT})^2}} \times 100 \quad (5.13)$$

Alternative error metrics that only depend on a single noise realisation are also included to allow direct comparison with the real data studies (Fig. 5.8, 5.11, 5.12 and 5.14). The tumour mean values were compared to the white matter standard deviation (calculated using an eroded ROI) to allow the trade-off between the retention of PET unique features and noise suppression of each method to be assessed. These single noise realisation error metrics are:

$$\bar{\theta}_{ROI} = \frac{1}{N_{ROI}} \sum_j^{N_{ROI}} \theta_j \quad (5.14)$$

$$\sigma_{ROI,2} = \sqrt{\frac{1}{N_{ROI}} \sum_j^{N_{ROI}} (\theta_j - \bar{\theta}_{ROI})^2} \quad (5.15)$$

However, the tumour mean vs white matter standard deviation metric is incomplete and does not penalise the distortion of tumour regions, bias of the cortical regions, or the smoothing of different intensities across regional boundaries (PVEs). To address the issue of PVEs, the mean value of the white matter, grey matter and cerebrospinal fluid (CSF) regions (using the full non-eroded ROI including edges) are also examined. These regional mean values are plotted against their corresponding standard deviation values. The regional standard deviation

values used for these plots are also calculated using the full non-eroded ROI. Methods that reduce PVEs across these regional boundaries should result in a reduction in white matter and CSF mean values, and an increase in the grey matter mean value relative to unsmoothed MLEM, (whilst over-smoothing methods will produce the converse trends). Improved reconstruction methods should therefore ideally lead to a fixed tumour mean relative to unsmoothed MLEM, reduced white matter standard deviation relative to unsmoothed MLEM, and follow the aforementioned regional mean trends.

5.5.2 Results

For most of the figures presented either the regularisation parameter β is increased (for MAP methods), or the number of k -nearest neighbours is increased (for kernel methods) along each curve. The original full neighbourhood implementation of HKEM [311] is equivalent to the last point on each HKEM curve ($k_{PET} = 124$). The popular reconstruction method of MLEM is also shown for comparison, at varying levels of post reconstruction smoothing.

Simulation Studies High Counts

Figure 5.6 shows a series of trade-off curves for all reconstruction methods under investigation, applied to the 10 noise realisations of the high count (10^8) simulated dataset. The bias vs standard deviation plot for the whole brain region (Fig. 5.6 column 1), demonstrates the improvements of all regularised and reparameterised MR-informed and PET-MR-informed reconstruction methods over post-smoothed MLEM, across the whole brain region where the majority of the MR structure matches the PET. The use of MR information in these methods, enables smoothing across MR uniform regions (reducing noise), whilst preserving MR boundaries and reducing PVEs. Of particular interest in this work is the ability to maintain these positive attributes of MR-informed methods, whilst also maintaining genuine high intensity features (tumour regions) present only in the PET data and not the MR. To

address this question, both NRMSE and SSIM curves are presented in Fig. 5.6 (columns 2 and 3 respectively), showing the trade-off between the accurate reconstruction of the whole brain region vs the PET unique regions (averaged over the four PET-unique regions). Both the NRMSE and SSIM curves (Fig. 5.6) show that the anato-functional method achieves the best trade-off relative to the other investigated methods, although the improvement is less apparent in terms of NRMSE. The KEM LVS and HKEM method also perform well in terms of the SSIM trade-off between the PET-unique and whole brain regions, whereas if evaluated in terms of the NRMSE trade-off the Bowsher method performs more favourably. The PET-unique region bias vs whole brain region standard deviation trade-off curve (Fig. 5.6 column 4) also demonstrates the improved performance of the anato-functional and KEM LVS methods, both attaining a reduced bias (for the PET unique regions B and D) for a fixed whole brain standard deviation (image noise level) in comparison to post-smoothed MLEM. The points indicated by arrows on the NRMSE curves (Fig. 5.6 column 2) indicate the parameters that produce approximately equivalent whole brain NRMSE. The corresponding images with fixed whole brain NRMSE are shown in Fig. 5.7, in which all MR-informed methods can be seen to yield a sharper reconstruction of the regions with shared PET-MR structure, in comparison to post-smoothed MLEM. Additionally, the anato-functional method can be seen to perform favourably in reconstructing PET-unique regions with a relatively high, constant intensity value and well-defined boundaries (e.g. Tumour D, Fig. 5.7 bottom row).

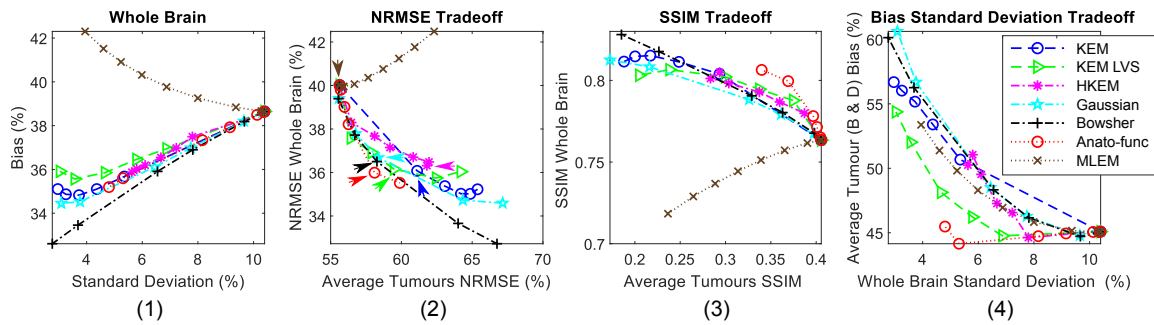


Fig. 5.6 All the results shown are for the high counts (10^8) simulated dataset, over multiple noise realisations. (1) Bias vs standard deviation for the whole brain region. (2) Whole brain NRMSE vs the PET-unique region NRMSE (averaged over the four PET-unique regions). (3) Whole brain SSIM vs the PET-unique region SSIM (averaged over the four PET-unique regions). (4) PET-unique region bias (averaged over tumour regions B and D only) vs whole brain standard deviation. The methods investigated are shown for increasing β values (MAP) or k -nearest neighbours (KEM). MLEM is shown for increasing levels of post reconstruction smoothing.

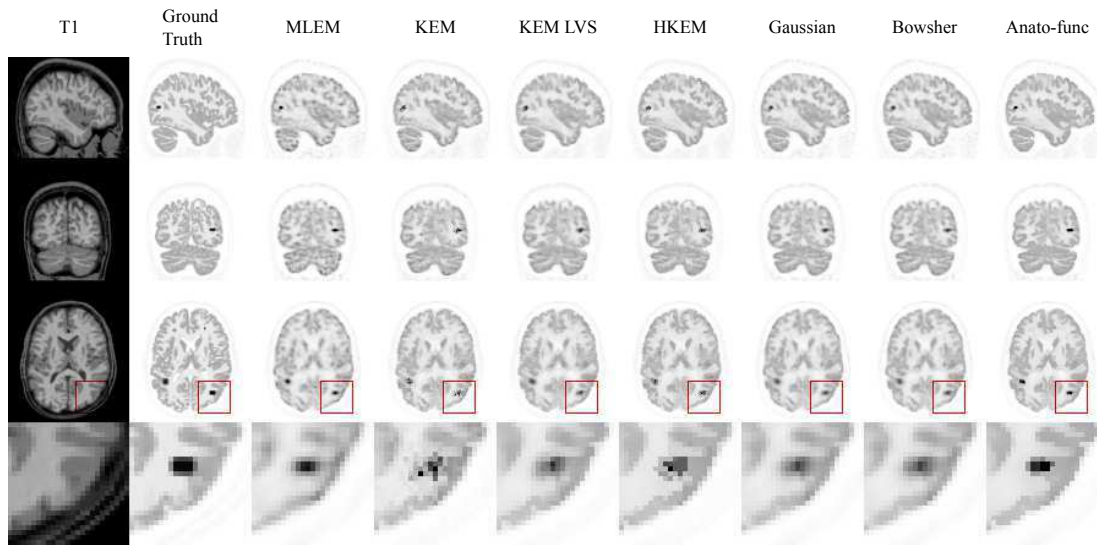


Fig. 5.7 All the results shown are for the high counts (10^8) simulated dataset for a single noise realisation, at PET resolution. Reconstructed images for each of the methods under investigation. A zoomed in transverse region is shown along the bottom row. All MR-informed methods are shown at approximately fixed whole brain NRMSE, as indicated by the arrows on the NRMSE curves, Fig. 5.6 column 2.

The ability of each reconstruction method to suppress image noise, without in turn suppressing the PET-unique regions was also assessed through tumour mean vs white matter standard deviation curves (Fig. 5.8). These graphs show the improvement attained by the anato-functional, HKEM and to lesser extent KEM LVS methods relative to the other MR(only)-informed reconstruction methodologies for the high intensity tumour regions B and D. The anato-functional and HKEM methods can achieve mean values similar to that of unsmoothed MLEM for the high intensity PET-unique regions, whilst markedly reducing the standard deviation across the white matter region. All MR-informed and PET-MR-informed methods also lead to a reduction in the white matter mean and an increase in the grey matter mean in comparison to MLEM (Fig. 5.8, bottom row), as expected for methods that reduce the PVEs in these regions. This demonstrates the capability of MR informed methods to counteract PVEs that cause spill-out effects for high intensity regions. The PET-unique regions of A and C appear to have insufficient intensity to really benefit from inclusion of PET information in the PET-MR-informed methods, with both the MR-informed and PET-MR-informed methods producing similar tumour mean vs white matter standard deviation curves (Fig. 5.8, top and middle rows), as such the information that can be drawn from the PET-unique regions A and C is limited.

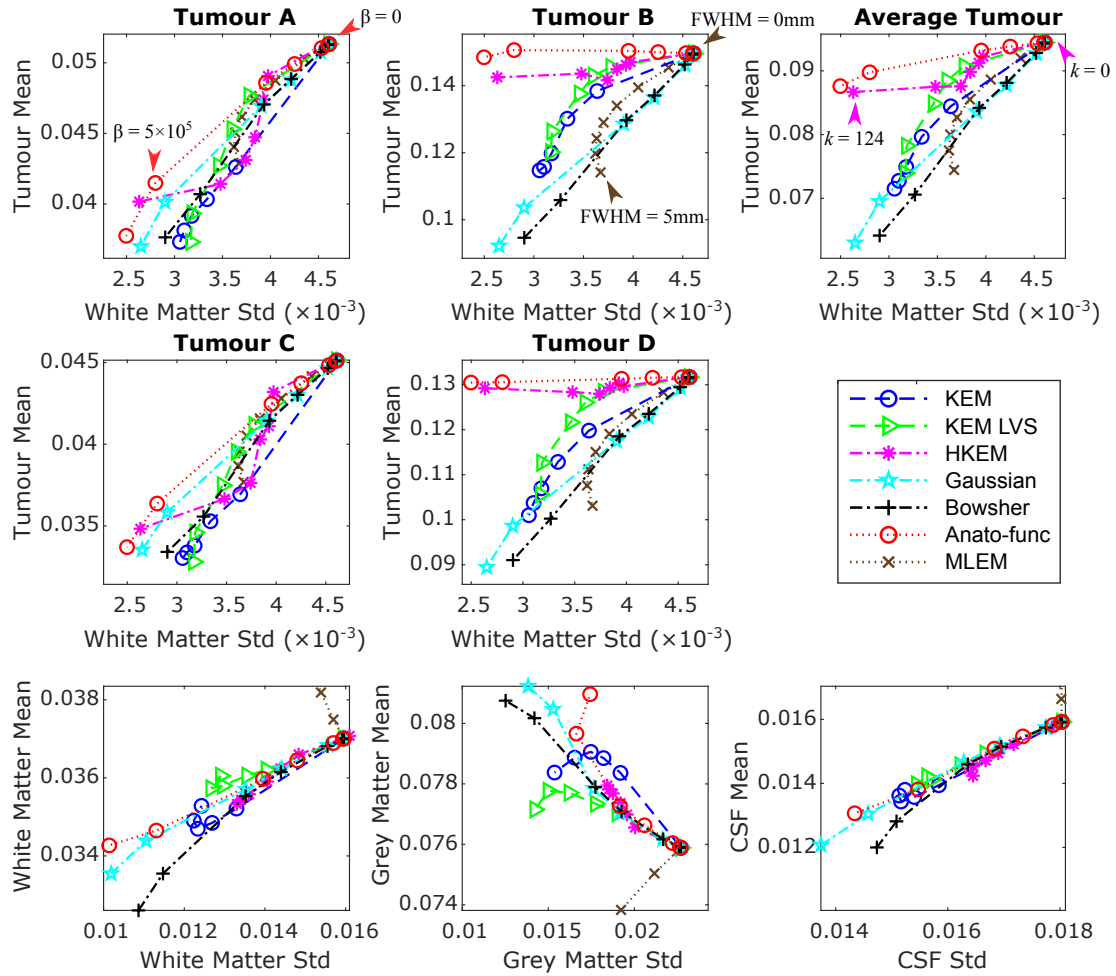


Fig. 5.8 All the results shown are for the high counts (10^8) simulated dataset, averaged over the multiple noise realisations. (top and middle row) The tumour mean value vs white matter standard deviation, for all methods under consideration, for each PET-unique region. The mean and standard deviation values were calculated within the specified (eroded) ROIs at PET resolution, and then averaged across the multiple noise realisations. The MLEM noise free mean values for the different tumour regions are: 0.0624 for Tumour A, 0.155 for Tumour B, 0.0575 for Tumour C and 0.137 for Tumour D. (bottom row) Regional mean value vs regional standard deviation using the full ROI including edges. The expected trend for methods that reduce PVEs is a reduction in the white matter and CSF mean, and an increase in the grey matter mean. The β (MAP) or k (KEM) parameter values are increased along each curve. MLEM is also shown for increasing level of post-reconstruction smoothing.

Simulation Studies Low Counts

Figure 5.9 (columns 2 and 3) shows the NRMSE and SSIM trade-off curves between the PET-unique (averaged over the four PET-unique regions) and whole brain regions, for the low count (10^7) simulated dataset. Both the NRMSE and SSIM plots show KEM LVS performing better than the other reconstruction methods, achieving both an improvement in whole brain and PET unique region image quality in comparison to unsmoothed MLEM. This trend can also be clearly seen in the bias (for tumour regions B and D) vs whole brain standard deviation plot (Fig. 5.9 column 4), with KEM LVS achieving the lowest tumour bias for a fixed whole brain noise level. The Bowsher method also performs well, when evaluated in terms of NRMSE trade-off (Fig. 5.9 column 2), but less so if evaluated in terms of SSIM trade-off (Fig. 5.9, column 3) or bias vs standard deviation trade-off (Fig. 5.9, column 4) where instead the HKEM method performs better than the other methods excluding KEM LVS. The points indicated by arrows on the NRMSE curves (Fig. 5.9, column 2) correspond to the parameters that produce approximately equivalent whole brain NRMSE, with the reconstructed images shown in Fig. 5.10. In Figure 5.10, the KEM LVS and HKEM methods show a marginally sharper recovery of the PET-unique regions in comparison to the remaining methods.

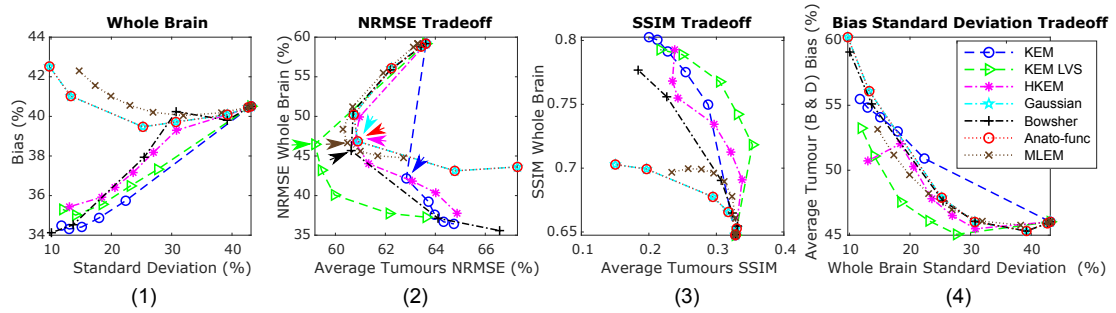


Fig. 5.9 All the results shown are for the low counts (10^7) simulated dataset, using multiple noise realisations. (1) Bias vs standard deviation for the whole brain region. (2) Whole brain NRMSE vs the PET-unique region NRMSE (averaged over the four PET-unique regions). (3) Whole brain SSIM vs the PET-unique region SSIM (averaged over the four PET-unique regions). (4) PET-unique region bias (averaged over tumour regions B and D only) vs whole brain standard deviation. Each method is shown for increasing β values (MAP) or k -nearest neighbours (KEM). MLEM is shown for increasing levels of post reconstruction smoothing.

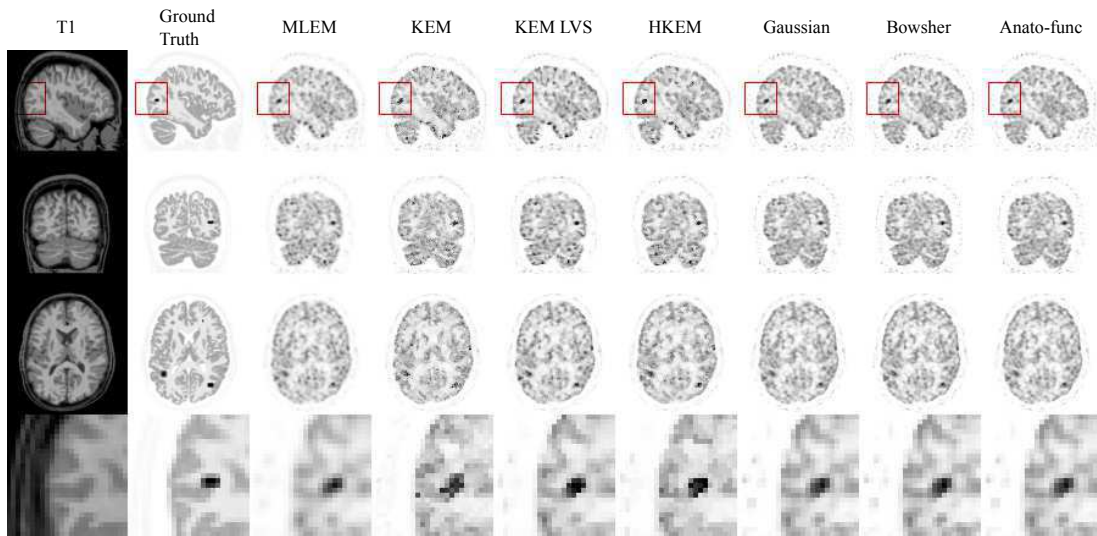


Fig. 5.10 All the results shown are for the low counts (10^7) simulated dataset, for single noise realisation, at PET resolution. Reconstructed images for each of the methods under investigation. A zoomed in sagittal region is shown along the bottom row. All MR-informed methods are shown at approximately fixed whole brain NRMSE, as indicated by arrows on the NRMSE curves, Fig. 5.9 column 2. MLEM is shown with post reconstruction smoothing with a FWHM of 3.5mm.

Figure 5.11 shows the tumour mean vs white matter region standard deviation for the low count dataset, with the KEM LVS and HKEM methods performing well in comparison to the other methods investigated. The two high intensity PET unique regions (B and D), particularly demonstrate the noise suppression abilities of the KEM LVS and HKEM methods, for a fixed tumour mean value. The KEM method also achieves a reduced white matter standard deviation for a relatively fixed PET-unique region mean value in Fig. 5.11, despite visually its PET-unique regions being more distorted than the corresponding KEM LVS and HKEM PET-unique regions (as shown in Fig. 5.10). This highlights the limitations of using ROI based error metrics. Figure 5.11 (bottom row) shows an increase in the white matter mean value and a reduction in the grey matter mean value for the Gaussian and anato-functional methods due to an increased smoothing across the white-grey matter boundary (PVEs). Therefore, the Gaussian and anato-functional methods (with their low count chosen parameters) have lost their PVE correction capabilities that were demonstrated in the high counts simulation studies (Fig. 5.8). Additional figures for both the high and low count simulation studies can be found in appendix C, and are referred to as Fig. C.1 and Fig. C.2 respectively.

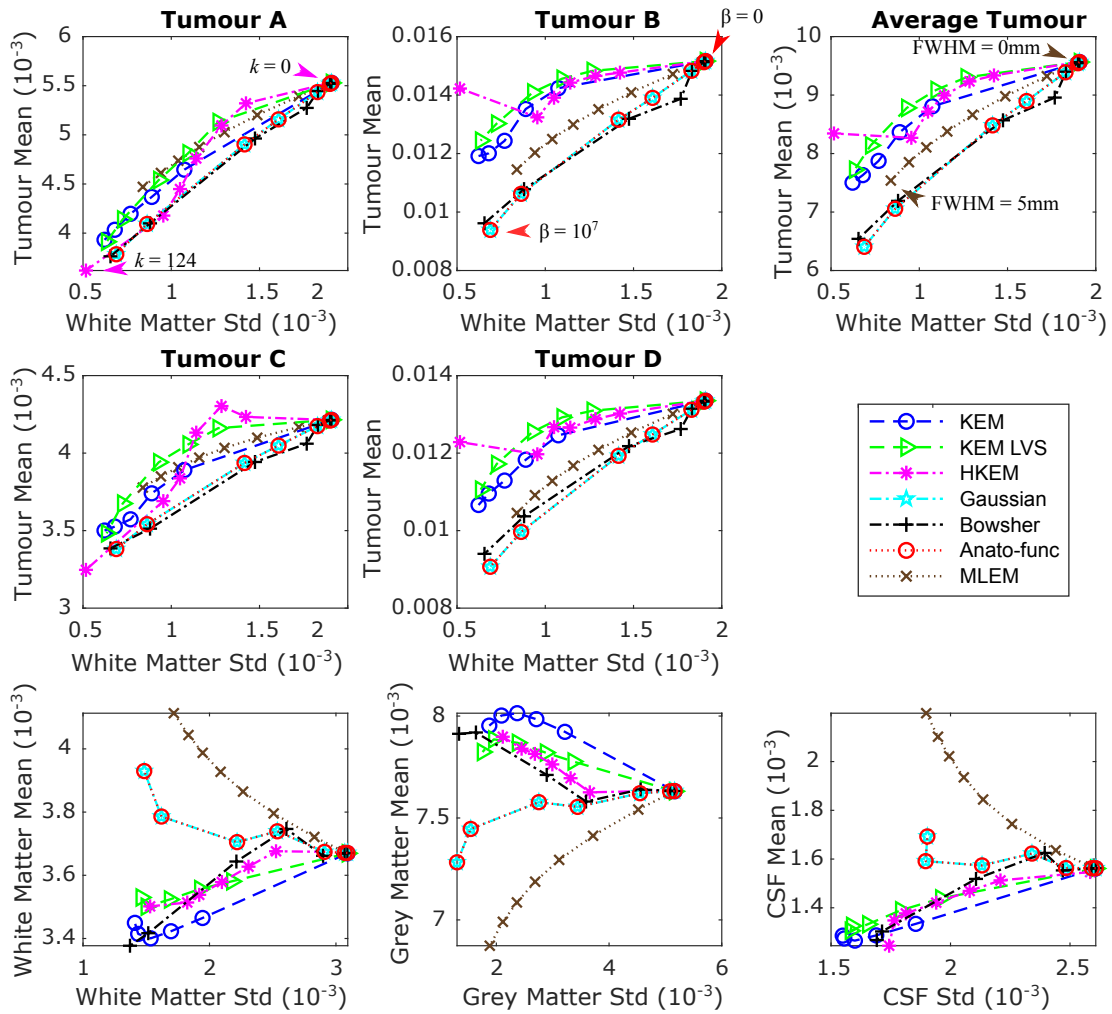


Fig. 5.11 All the results shown are for the low counts (10^7) simulated dataset, averaged over multiple noise realisations. (top and middle row) The tumour mean value vs white matter standard deviation, for all methods under consideration, for each PET-unique region. The mean and standard deviation values were calculated within the specified ROIs at PET resolution, and then averaged across the multiple noise realisations. The MLEM noise free mean values for the different tumour regions are: 0.00624 for Tumour A, 0.0155 for Tumour B, 0.00575 for Tumour C and 0.0137 for Tumour D. (bottom row) Regional mean value vs regional standard deviation using the full ROI including edges. The expected trend for methods that reduce PVEs is a reduction in the white matter and CSF mean, and an increase in the grey matter mean. β (MAP) or k (KEM) parameter values were increased along each curve. MLEM is also shown for increasing level of post-reconstruction smoothing.

5.6 Real Data Studies

5.6.1 Experimental Methods

The parameters used for the 100% and 10% count level augmented FDG real data studies were those chosen for the high and low count simulation studies respectively. However, the σ_{PET} parameters employed by the anato-functional and HKEM methods were re-selected for each count level, due to the dependence of σ_{PET} on the PET image intensity. As previously evaluated in the simulated data studies, the mean value of each PET-unique region was compared to the standard deviation of an eroded white matter region, in order to assess the ability of each method to simultaneously recover the PET-unique region whilst suppressing noise in an (assumed) approximately uniform region. The white matter mask was extracted from the T1 image using FSL [325]. The mean value of the white matter, grey matter and cerebrospinal fluid (CSF) regions (using the full non-eroded ROI including edges) are also examined for the real data studies, to help investigate the PVE properties of each reconstruction method.

5.6.2 Results

100% Count Level Augmented FDG Dataset

Figure 5.12 shows the tumour mean vs white matter standard deviation trade-off plots (top and middle row) in addition to the white matter, grey matter and CSF regional mean plots (bottom row), for the augmented FDG dataset (Fig. 5.5), at the 100% count level. The anato-functional method is shown to perform well for the high count real data studies, achieving the lowest white matter standard deviation (noise level) for a fixed tumour mean value (for three out of the four tumour regions). The KEM LVS and HKEM methods also perform better than the remaining MR-informed methods when evaluated in terms of tumour mean vs white matter standard deviation trade-off (Fig. 5.12). Figure 5.12 (bottom row) shows

that all MR-informed methods produce images with a reduced white matter mean, and an increased grey matter mean for moderate levels of regularisation (or reparameterisation). These trends correspond to a reduction in PVEs, in particular for the KEM and Bowsher methods. Figure 5.13 shows the reconstructed images for each of the MR-informed methods, with the parameters chosen to present an approximately fixed white matter standard deviation (as indicated by the arrowheads in Figure 5.12). Figure 5.13 also shows the improved definition and delineation between the white and grey matter when using the Bowsher and KEM methods.

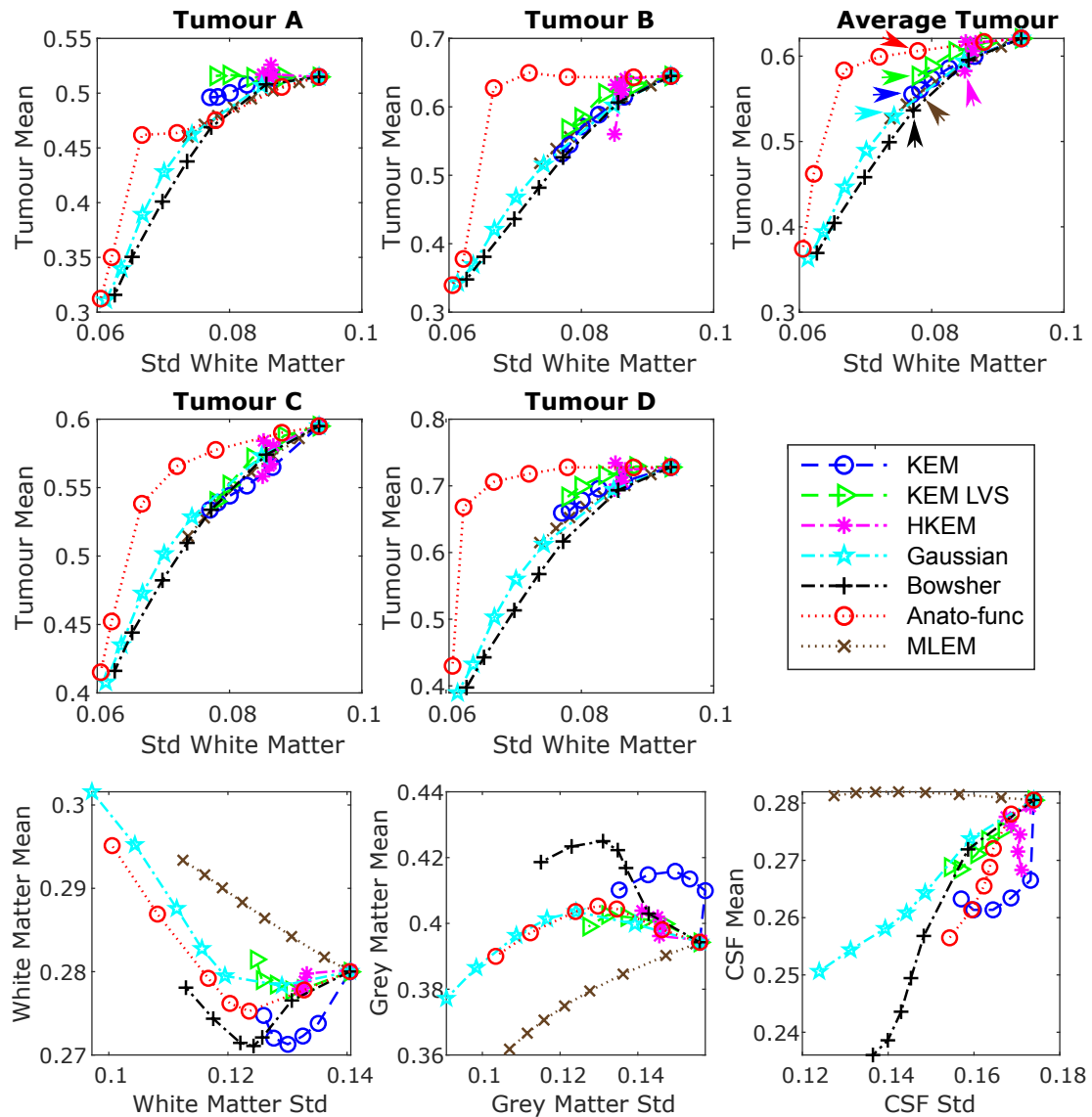


Fig. 5.12 All results shown are for the 100% count level augmented real dataset. (top and middle rows) Tumour mean vs white matter (eroded mask) standard deviation. The arrows indicate the mean vs standard deviation trade-off of the chosen parameters for the images shown in Fig. 5.13. (bottom row) Regional mean value vs regional standard deviation using the full ROI including edges. The expected trend for methods that reduce PVEs is a reduction in the white matter and CSF mean, and an increase in the grey matter mean.

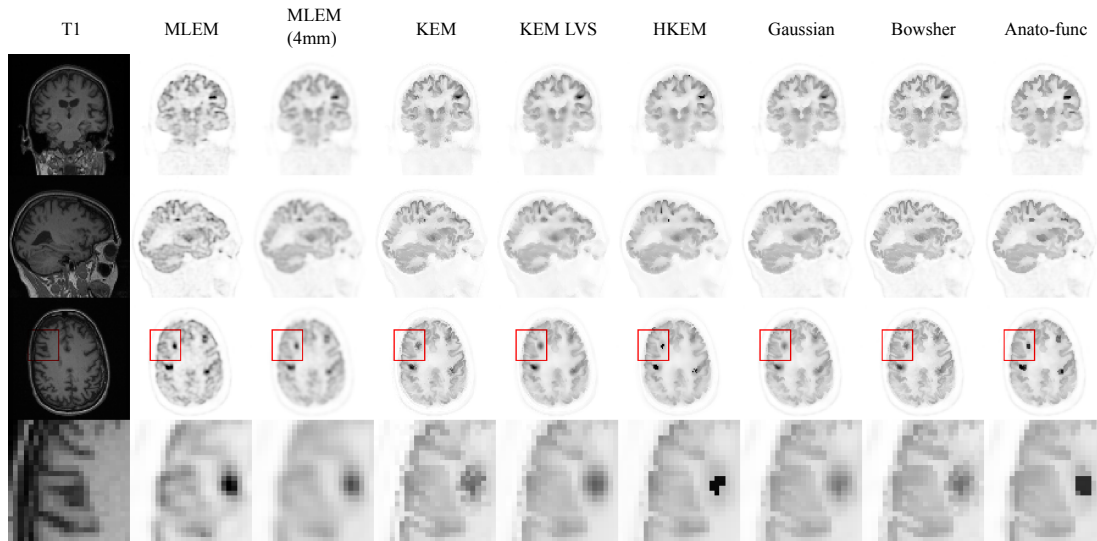


Fig. 5.13 The reconstructed images are shown for each of the reconstruction methods investigated, applied to the 100% count level augmented real dataset. A zoomed in transverse region is shown along the bottom row. All MR-informed methods are shown at approximately fixed white matter standard deviation, as indicated by the arrows in Fig. 5.12. MLEM is shown with post-reconstruction smoothing (FWHM 4mm).

10% Count Level Augmented FDG Dataset

Figure 5.14 shows the tumour mean values vs white matter standard deviation plots (top and middle row) in addition to the white matter, grey matter and CSF regional mean plots (bottom row), for the augmented FDG dataset at the 10% count level. KEM LVS, HKEM and to a lesser extent the anato-functional method perform well in maintaining the mean value of the PET-unique regions, whilst reducing the white matter standard deviation (Fig. 5.14 , top and middle rows). Figure 5.14 (bottom row) shows that the anato-functional and Gaussian methods lead to an increase in the white matter mean value and a decrease in the grey matter mean value. This trend demonstrates that the anato-functional and Gaussian methods have lost their PVE reduction capabilities that were demonstrated in the high count real data studies (Fig. 5.12). Figure 5.15 shows the reconstructed images for each of the MR-informed methods, with the parameters chosen to present an approximately fixed white matter standard deviation (as indicated by the arrowheads in Figure 5.14).

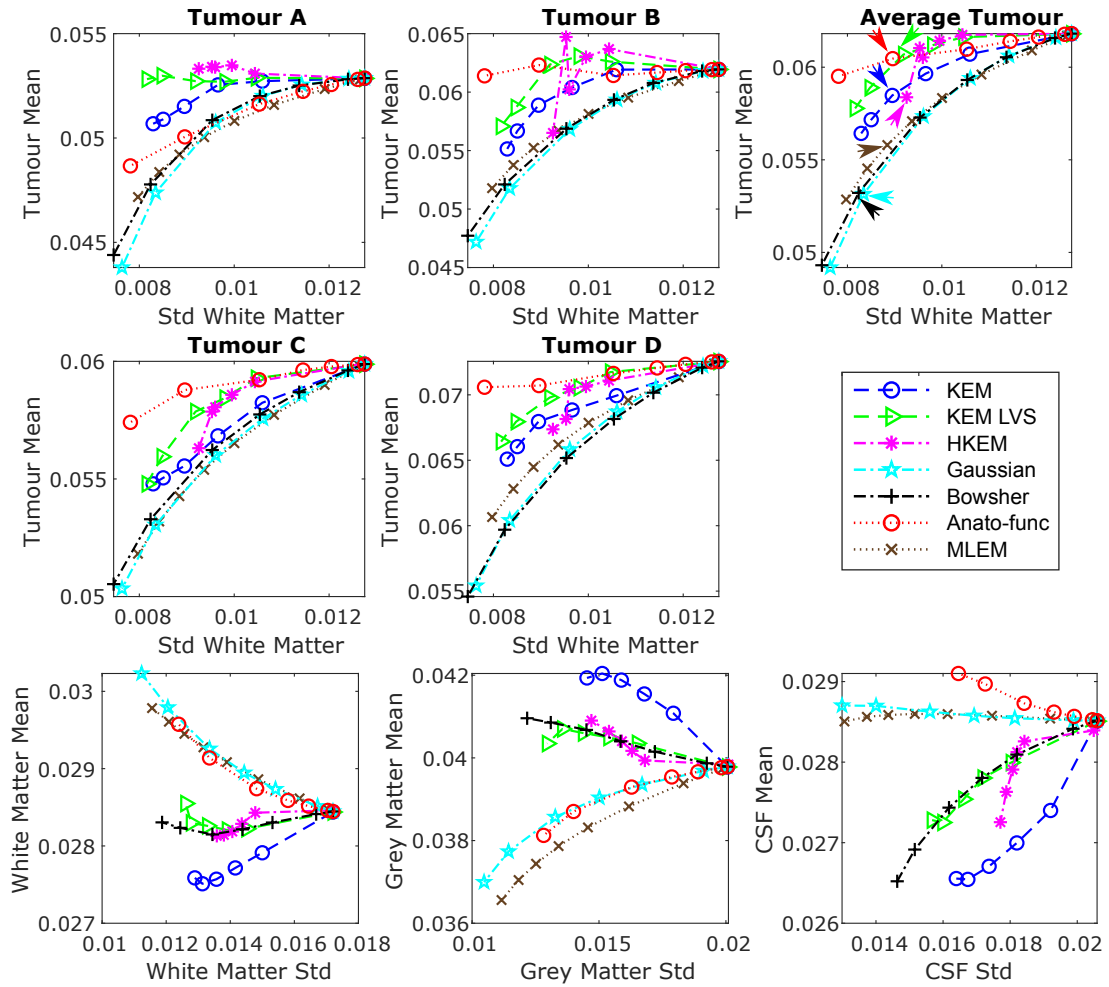


Fig. 5.14 All results shown are for the 10% count level augmented real dataset. (top and middle row) Tumour mean vs white matter (eroded mask) standard deviation. The arrows indicate the mean vs standard deviation trade-off of the chosen parameters for the images shown in Fig. 5.15 . (bottom row) Regional mean value vs regional standard deviation using the full ROI including edges. The expected trend for methods that reduce PVEs is a reduction in the white matter and CSF mean, and an increase in the grey matter mean.

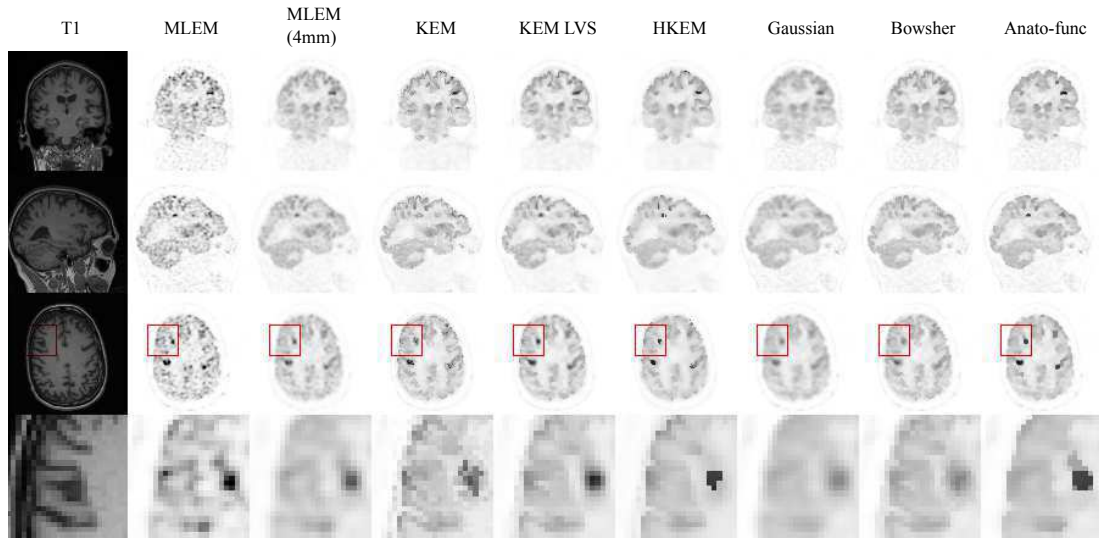


Fig. 5.15 The reconstructed images are shown for each of the reconstruction methods investigated, applied to the 10% count level augmented real dataset. A zoomed in transverse region is shown along the bottom row. All MR-informed methods are shown at approximately fixed white matter standard deviation, as indicated by arrows in, Fig. 5.14. MLEM is shown with post-reconstruction smoothing (FWHM 4mm).

5.7 Discussion

5.7.1 Simulation Studies

The results from the high count simulation studies shows that the anato-functional method excels in concurrently improving the reconstruction of the whole brain and PET-unique regions, when assessed via the majority of metrics presented in Fig. 5.6-5.8. The KEM LVS and HKEM methods also perform well in terms of the SSIM trade-off (Fig. 5.6) between the PET-unique and whole brain regions, and achieve a reduced white matter standard deviation for a fixed tumour mean value (Fig. 5.8). The Bowsher method achieves the lowest whole brain NRMSE (Fig. 5.6) and notably reduces PVEs, however it clearly suppresses the mean values of the PET-unique regions (Fig. 5.8). The remaining MR-informed methods are not the foremost method in respect to any of the presented metrics. The superior performance of

the anato-functional method over HKEM can be attributed to the added flexibility provided by tuning the β regularisation parameter. This is important due to the lack of PET structure present in the reconstructed PET image at low iterations, as such, the weighting factors (for the anato-functional method) and basis functions (for HKEM) will be strongly influenced by the MR structure only, producing reconstructions similar to MR-informed methods. Thus, a low level of regularisation for the anato-functional method is beneficial for high count datasets to prevent PET unique features being suppressed at low iterations. For HKEM, fewer voxels can be selected to contribute to the basis function (lower kNN) to prevent PET unique features being suppressed. As can be seen in Fig. 5.6 column 4 the HKEM method performs similarly to the anato-functional method for $kNN = 5$, however for larger values of kNN the tumour regions are generally more biased. Therefore, the added flexibility of the regularised anato-functional method improves its recovery of PET-unique regions as well as the shared PET-MR regions, in comparison to HKEM. Although, for high regularisation levels (high β) the anato-functional methods will also suppress the PET-unique regions. Despite the beneficial attributes of the anato-functional method over MR(only)-informed methods, it should be noted that the inclusion of PET information for both HKEM and anato-functional does degrade their noise reduction ability for regions with differing PET and MR structures as shown in the standard deviation images and bias vs standard deviation trade-off plots for the PET-unique regions (Fig. C.1). In addition, the effective reduction of regularisation that occurs for these PET-MR-informed methods at the PET-unique regions may leave the corresponding reconstructed images more vulnerable to Gibbs like artefacts [163, 326]. This is clearly a downside of the PET-MR-informed methods, however the increase in standard deviation (noise) is only observed in these particular regions (where the PET and MR structures differ) and at large β values. Both improvements in the noise reduction properties of PET-MR-informed methods, and more effective suppression of Gibbs like artefacts can be achieved through selecting a larger σ_{PET} parameter (as done for the low

count simulations), resulting in broader PET similarity kernels. However, this will impair the discussed PET-MR-informed methods capability of reconstructing PET-unique regions.

The results from the low count simulation studies show that the KEM LVS method performs better than the other investigated methods in terms of the majority of metrics evaluated, such as NRMSE trade-off, SSIM trade-off (Fig. 5.9 columns 2 and 3) and tumour bias vs tumour standard deviation trade-off (Fig. C.2). The lack of dependence of the KEM LVS method on PET data helps suppress the noise in the image, whilst the compactness of the basis functions enables improved recovery of the PET unique regions (relative to the other MR-informed methods). The HKEM method also performs well when evaluated in terms of SSIM trade-off, tumour bias vs whole brain standard deviation trade-off (Fig. 5.9, columns 3 and 4), or tumour mean vs white matter standard deviation (Fig. 5.11). Both the KEM and Bowsher methods lead to large reductions in whole brain NRMSE (Fig. 5.9, columns 2) and PVEs (Fig. 5.11, bottom row), however this is at the expense of suppression and possible deformation of the tumour regions (as shown in Fig. 5.10 and Fig. C.2 bias-std trade-off and profiles). In contrast to the high count simulation studies, the anato-functional method performs poorly in terms of all the metrics presented (Fig. 5.9-5.11) for the low count simulation studies. The noise present in the low count dataset clearly has a negative impact on the anato-functional method, which fails to outperform the rival methods as it did for the high count dataset, (Fig. 5.9 vs Fig. 5.6). For the anato-functional method, to preserve PET only edges requires the value of σ_{PET} to be relatively small. However, for noisy PET data, most of the PET voxel values will be quite different to the central voxel value, leading to most of the weighting factors going to zero. This limits the extent of regularisation for the anato-functional method resulting in noisy reconstructed images. This problem can be resolved by using a larger value of σ_{PET} (as chosen for the low count simulations), but this limits the influence of the PET update image, and hence the recovery of PET-unique regions. By comparison, the HKEM method uses the k NN sparsification to select voxel to contribute

to the basis function, based on their PET voxel value. The k NN sparsification allows PET features to be extracted, whilst allowing a large σ_{PET} value to be selected. Therefore, a fixed number of voxels are selected to contribute to each basis function, effectively fixing the level of regularisation (irrespective of the noise present in the PET image). This allows the HKEM method to achieve an improved trade-off in comparison to the anato-functional method. A further comparison of the high and low count simulation studies (Fig. 5.7 vs Fig. 5.10) shows a major change in the Gaussian (and anato-functional) method's capability to delineate white and grey matter regions. This is due to the different σ_{MR} values selected (as shown in Table 5.2), for the different count levels. For the high counts dataset, a smaller value of σ_{MR} is selected (in comparison to the low count dataset value), emphasising differences in MR voxel values, and thus enhancing edges in the reconstructed PET image. Whereas, for the low counts dataset, a larger value of σ_{MR} was selected, leading to similar weighting values for all MR voxels as differences in the MR voxels are effectively ignored, producing spatially smooth PET images. This trend is also observed for the high and low count real data studies, where the same σ_{MR} parameters are used as in the corresponding simulation study.

5.7.2 Real Data Studies

For the high count simulation and real data studies similar trends are observed. In both studies, the anato-functional method followed by the KEM LVS and HKEM methods perform better than the remaining MR-informed methods when evaluated in terms of tumour mean vs white matter standard deviation trade-off (Fig. 5.8 and Fig. 5.12). In particular, the anato-functional method achieves the lowest white matter standard deviation for a fixed tumour mean for three out of the four tumour regions assessed (Fig. 5.8 and 5.12, top and middle row), whilst also reducing PVEs with respect to post-smoothed MLEM (Fig. 5.8 and Fig. 5.12, bottom row). Therefore, the inclusion of PET information into the reconstruction process via the anato-functional method (and HKEM) appears to be beneficial

in the high count simulated and real patient datasets investigated. All MR-informed methods investigated also demonstrate improved PVE correction for both simulation and real data studies, in comparison to post-smoothed MLEM. Visual support of the improved grey-white matter delineation is shown in Fig. 5.13. For the low count simulation and real data studies, similar trends are also observed in terms of accurately reconstructing PET unique regions, whilst concurrently suppressing noise, with the KEM LVS and to a lesser extent the HKEM method performing better than the other methods investigated. In both simulated and real data studies the KEM LVS and HKEM methods achieve similar PVE reduction and noise suppression properties (Fig. 5.11 and Fig. 5.14, bottom row) to the Bowsher method, whilst reconstructing the PET-unique regions with an increased mean value (Fig. 5.11 and Fig. 5.14, top and middle row) relative to Bowsher. The performance of the anato-functional method however, does differ slightly between the simulation and real data studies, with the anato-functional method achieving an improved tumour mean vs white matter standard deviation trade-off in the real data study (Fig. 5.11 vs Fig. 5.14). This change can be attributed to the σ_{PET} parameter, which was re-selected for the real data studies. Excluding the anato-functional method, a similar trend is observed between the simulated and real data studies for the remaining MR-informed methods in terms of tumour mean vs white matter standard deviation (Fig. 5.11 and Fig. 5.14, top and middle row). The regional mean values for the MR-informed methods also show similar results between the simulated and real data studies (Fig. 5.11 and Fig. 5.14, bottom row). In both cases, the KEM method provides the best reduction in PVEs (reducing the white matter mean and increasing the grey matter mean), whereas the anato-functional and Gaussian methods show the converse trend due to increased PVEs. The over-smoothing of the Gaussian and anato-functional methods, in comparison to the other MR-informed methods investigated (such as Bowsher) can be seen in Fig. 5.15.

5.8 Limitations

This chapter has compared a selection of MR-informed PET reconstruction methods that are prominent in the literature. Unsurprisingly, a number of alternative MR-informed methods discussed in §2.8 could have been included and may lead to improved results depending on the metrics evaluated. However, existing comparison literature has shown the comparable and possibly improved performance of Bowsher over alternative methods such as parallel level sets [64] and joint entropy [65]. It should also be noted that such methods lie outside the family of methods evaluated in this work, which include MR-information through local neighbourhood similarity matrices. In addition, further comparison could be made with deep learning based reconstruction methods, that have recently shown impressive denoising results [116].

The image quality measures of NRMSE and SSIM used in this chapter (and thesis) to evaluate the proposed methods are useful for representing performance of quantification tasks. However, the ability to interpret such error metrics as potential changes in patient diagnosis and management is limited. The ultimate evaluation of medical image quality is through task-based observer studies, using either a human observer or a representative mathematical observer. Two popular means for undertaking such task-based observer studies include receiver operating characteristic (ROC) curves and the use of the Hotelling observer [235, 327, 212, 213]. For the presented comparison study the use of human observers (radiologists in the case of PET-MR images) would be very time consuming given the evaluation required of multiple patients and reconstruction methodologies. Ideally, each scan would also be read by multiple radiologists to account for inter-person variation in the reading of PET scans. Automating the task-based observer process is still an area of active research within medical imaging, even for the relatively simple task of tumour detection [214, 215].

5.9 Summary

Three regularised and three re-parameterised MR-informed PET reconstruction methodologies have been compared. The capability of each method to reduce PVEs and the noise present in the reconstructed image without biasing the recovery of genuine high intensity PET-MR mismatched regions was investigated. For the high count dataset, the anato-functional method provided the best SSIM and NRMSE whole brain to PET-unique region trade-off relative to the other methods investigated. The inclusion of the current PET image through the calculation of the weighting factors, as presented by the anato-functional MAP method, enables PVE correction and noise suppression to be attained in regions of matching PET-MR structure, whilst also reconstructing PET-unique regions with a similar bias to unsmoothed MLEM. For the low count simulated dataset, the use of spatially-compact basis functions (KEM LVS) achieved the best SSIM and NRMSE trade-off for the reconstruction of PET-unique and whole brain regions, outperforming the other methods investigated. The HKEM method also performed well for the simulated and real datasets in terms of tumour mean vs white matter standard deviation, relative to the remaining MR-informed methods. The presence of noise in the low count dataset (of similar intensity to the PET-unique regions) lead to the inclusion of noise in the reconstructed PET image for the anato-functional (and the HKEM to a lesser extent) method if certain parameters values were selected. Therefore, at low count levels the inclusion of PET information into the reconstruction process could be more beneficial if integrated in an alternative manner. To conclude, for the reconstruction of noisy data multiple MR-informed methods produce favourable whole brain vs PET-unique region trade-off curves, very comfortably outperforming the whole brain denoising of post-smoothed MLEM, for a fixed PET-unique region SSIM.

Chapter 6

Machine Learning Inspired Solutions for PET Reconstruction

Chapter 5 presented a comparison of some of the most prominent reconstruction methods in the field of MR-informed PET reconstruction. These PET reconstruction methods were assessed in terms of their ability to improve the overall PET image quality, whilst maintaining features that are present only in the PET data. All considered MR-informed methods were shown to provide major denoising power relative to the popular (clinically representative) method of post-reconstruction smoothed MLEM. However, in order to reduce the suppression of PET-unique features, the level of regularisation employed for these MR-informed methods is limited. Consequently, due to the low regularisation levels used the improvement in overall image quality lies beneath the full potential of regularised methods. In light of the limited improvements currently attainable via regularised reconstruction methodologies, substantive advances in the field of PET reconstruction may still be realised. One avenue for unlocking such potential could be utilising population statistics via supervised machine learning methodologies. As described in chapter 2 there exists a wide variety of deep learning inspired solutions to problems such as image denoising [96, 98], super resolution [99–101], style transfer [102] and deconvolution [103], which recently have shown state-of-the-art

performance for medical image reconstruction. In particular neural networks have been shown to be highly beneficial in regards to: (1) reduction of the administered patient dose (PET or CT) without impairing the overall image quality or diagnostic value of the scan [328, 329, 112–116] and (2) acceleration of the reconstruction process, even approaching real time [330, 274, 331]. As such, this chapter initially seeks to combine the previous advances demonstrated by neural networks (image denoising and acceleration) to produce an unrolled network that achieves both benefits simultaneously.

The unrolled (or variational) network architecture enables the advanced prior knowledge of the system matrix to be incorporated into the reconstruction process alongside learnt denoising features [268]. The inclusion of the system matrix allows for far more compact networks comprised of fewer learnable parameters in comparison to direct networks that must effectively learn the system matrix [269]. This leads to multiple benefits in terms of improving the network's generalisation to unseen datasets and reducing the risk of overfitting, which most networks are susceptible to if limited training data is used. Unrolled networks have also shown recent success in the reconstruction literature, outperforming post-reconstruction networks for a variety of imaging modalities, e.g. MRI [270, 332, 269, 268], CT [333–335] and PET [275, 272, 276]. The superior image reconstruction achieved through including regularising information into the reconstruction process, as opposed to post-reconstruction, matches the results found for conventional iterative reconstruction algorithms [234]. However, difficulties arise in training unrolled network architectures on 3D imaging datasets due to their vast memory and time requirements, in combination with the large system matrix. A common solution to reduce the computational burden and improve generalisability by training on image patches is not possible for unrolled networks, due to the presence of the system matrix in the network. Research in reducing these memory requirements for unrolled networks is ongoing [336].

Moreover, the neural network component of unrolled architectures is not purely limited to image denoising but can also reduce the number of (time consuming) image-space to projection-space transforms required [274]. Therefore, unrolled neural networks hold the potential to both enhance the reconstructed image quality, and accelerate the reconstruction process in comparison to conventional iterative reconstruction methods. The aim of this chapter is to identify the extent to which unrolled neural networks (and the inclusion of data fidelity terms) can denoise, enhance and accelerate the reconstruction process. Only limited investigation into this form of unrolled network has currently been presented in the PET reconstruction literature [276, 272, 275]. Therefore this work will investigate in greater detail how the efficacy of unrolled networks varies with the number of data fidelity layers, and whether such fidelity layers are required at all. The impact of the number of regularising network layers, and the overall depth of the network is also explored, to stratify the importance of such hyperparameters. Initially, this chapter investigates unrolled networks trained from a simulated 2D PET database. The unrolled network is compared to post-reconstruction smoothed MLEM and a post-reconstruction U-Net.

Despite the benefits of using an unrolled network architecture for image regression as demonstrated in the literature, the endemic issue of how to suitably train a neural network for real PET datasets persists. PET imaging is particularly problematic in this regard, due to the lack of appropriate training labels. An ideal set of training labels would have improved spatial resolution and reduced stochastic noise in comparison to current clinical practice. To attain such PET datasets would require the patient (and supporting staff) to receive a disproportionately high radiation dose, and a PET system with greater spatial resolution capabilities than currently available. Unsurprisingly, such clinical datasets don't currently exist. The use of alternative training labels from pre-clinical PET, auto-radiography, simulated or physical phantom studies (e.g. NEMA [283] or Hoffman brain [284]) can provide images with increased spatial resolution and SNR but are of limited use due to the

sensitivity of networks to variations in the training and testing domains. Training networks with an unrepresentative dataset can lead to unpredictable and incorrect results.

Previous real data implementations of deep learning inspired PET reconstruction methodologies have been mainly confined to converting low count images to high counts images through count rebinning [116, 337, 113, 338, 115, 339, 272, 273]. Alternative implementations have used MR-informed reconstructed images as the training labels [331], or trained predominantly on simulated datasets before "fine tuning" on real datasets [280]. In contrast, a recent image processing study proposed to train an image regression network through the use of noisy labels, referred to as Noise2Noise [118]. The capability to train networks using noisy labels only, greatly facilitates the application of deep learning networks to medical imaging datasets. Subsequent work has extended the Noise2Noise concept and applied it to the medical imaging datasets of MRI and CT [119]. The remainder of this work seeks to apply the Noise2Noise framework to a 3D PET real patient dataset. This framework has currently only been employed in a post-reconstruction fashion, and as previously discussed unrolled networks are exceptionally memory demanding if applied to 3D datasets. Therefore the previously described post-reconstruction U-Net was extended to 2.5D (limited 3rd dimension) and used in validating the Noise2Noise framework for PET imaging. The U-Net was investigated for both 3D simulated and real patient datasets via the use of noisy training labels.

In comparison to the preceding research based chapters, the results presented in this chapter are less comprehensive due to unavoidable time constraints. A subset of this work has been presented at the 2019 IEEE Nuclear Science Symposium and Medical Imaging Conference [340].

6.1 Theory

6.1.1 Unrolled Networks

The objective function for PET reconstruction can be subdivided into a data fidelity term $\Omega(\boldsymbol{\theta}|\mathbf{m})$ and a regularisation term $R(\boldsymbol{\theta})$ weighted by a hyperparameter β [196].

$$\boldsymbol{\theta}^{MAP} = \arg \max_{\boldsymbol{\theta}} \Omega(\boldsymbol{\theta}|\mathbf{m}) - \beta R(\boldsymbol{\theta}) \quad (6.1)$$

An unrolled neural network includes the data fidelity image update within its architecture. For PET image reconstruction the popular MLEM update can be used to provide the data fidelity update [276]. The unrolled network employed in this work acts directly on the first MLEM image update, $\boldsymbol{\theta}^1$, and the overall unrolled network is represented by $g(\boldsymbol{\theta}^1, \mathbf{m}; \Phi)$, where $\mathbf{m}$ is the projection data and Φ are learnable network parameters. The unrolled network can be decomposed into modules, where module n receives an image $\boldsymbol{\theta}^n$ and outputs the next update image $\boldsymbol{\theta}^{n+1}$. Each module consists of: a data fidelity layer that produces the next EM update $\boldsymbol{\theta}_{EM}^n$; an internal regularising network (g_r) that produces a corresponding prior image $\mathbf{u}^{n+1}$; and a fusion layer that combines the prior and update images. This fusion of $\boldsymbol{\theta}_{EM}^n$ and $\mathbf{u}^{n+1}$ is elegantly achieved via the Levitan and Herman MAP formulation [185]. The corresponding equations for the prior image and update image respectively are:

$$\mathbf{u}^{n+1} = g_r(\boldsymbol{\theta}^n; \Phi^n) \quad (6.2)$$

$$\boldsymbol{\theta}^{n+1} = \arg \max_{\boldsymbol{\theta}} \Omega^{ML}(\boldsymbol{\theta}^n|\mathbf{m}) - \beta^n \|\boldsymbol{\theta}^n - \mathbf{u}^{n+1}\|_2^2 \quad (6.3)$$

where $\Omega^{ML}(\boldsymbol{\theta}^n|\mathbf{m})$ is the Poisson log likelihood and the hyperparameter β is now expressed as a learnable parameter that can vary with each module. The latter equation will now be re-written to explicitly include the Poisson log likelihood term, and shall be expressed in index notation. An EM-surrogate for the Poisson log likelihood function [313, 66] is

employed to allow a separable update equation to be found:

$$\begin{aligned} \Omega^{ML}(\boldsymbol{\theta}^n | \mathbf{m}) - \beta^n \|\boldsymbol{\theta}^n - \mathbf{u}^{n+1}\|_2^2 = \\ \sum_i^{N_p} m_i \log([\mathbf{A}\boldsymbol{\theta}]_i + b_i) - ([\mathbf{A}\boldsymbol{\theta}]_i + b_i) - \beta^n \sum_j^{N_v} (\theta_j - u_j^{n+1})^2 \end{aligned} \quad (6.4)$$

$$\Omega^{ML}(\boldsymbol{\theta}^n | \mathbf{m}) - \beta^n \|\boldsymbol{\theta}^n - \mathbf{u}^{n+1}\|_2^2 \geq \sum_j^{N_v} a_j (\theta_{j,EM}^n \log(\theta_j) - \theta_j) - \beta^n (\theta_j - u_j^{n+1})^2 \quad (6.5)$$

where

$$a_j = \sum_i^{N_p} a_{ij} \quad (6.6)$$

and $\theta_{j,EM}^n$ is defined as:

$$\theta_{j,EM}^n = \frac{\theta_j^n}{a_j} \sum_i^{N_p} a_{ij} \frac{m_i}{[\mathbf{A}\boldsymbol{\theta}^n]_i + b_i} \quad (6.7)$$

The $\boldsymbol{\theta}$ variable that maximises the above Eq. 6.3 can be found by equating the derivative of Eq. 6.5 (right hand side) to zero. After rearranging the following equation is found:

$$2\beta^n \theta_j^2 + (a_j - 2\beta^n u_j^{n+1})\theta_j - a_j \theta_{j,EM}^n = 0 \quad (6.8)$$

The solution to the quadratic equation is given by:

$$\theta_j^{n+1} = \frac{-(a_j - 2\beta^n u_j^{n+1}) + \sqrt{(a_j - 2\beta^n u_j^{n+1})^2 + 8\beta^n a_j \theta_{j,EM}^n}}{4\beta^n} \quad (6.9)$$

Through Taylor expanding Eq. 6.9 it can be shown that $\theta_j^{n+1} \rightarrow \theta_{j,EM}^n$ as $\beta \rightarrow 0$. In addition to the forward expression for each layer present in the network, the partial derivative of each layer with respect to each input is required to correctly implement the back-propagation algorithm required to train the learnable parameters present. The overall network is trained through minimising a loss function, between the network's output and the set of training

labels. The L2 loss function is employed in this work and is shown below:

$$\Phi = \arg \min_{\Phi} \frac{1}{N_t} \sum_{t=1}^{N_t} \|g(\theta_t^1, m_t; \Phi) - \theta_t^{GT}\|_2^2 \quad (6.10)$$

where N_t is the number of training samples, Φ are the network's internal parameters and θ_t^{GT} is the ground truth training label for the input image θ_t^1 .

6.1.2 Noise2Noise Real Data Training

The Noise2Noise framework [118] provides a means to train networks in an unsupervised manner. It has been shown to provide effective denoising power for Gaussian, Poisson and Bernoulli noise models. The Noise2Noise framework is based on the assumptions that the noise realisations used are independent and zero-mean. Such work has since been extended in the medical imaging domain [119], in which it is used to denoise images in a post-reconstruction manner. The work by [119] also extends the Noise2Noise framework through developing a consensus loss function and presents promising validation results for accelerated MRI and low dose CT images. The consensus loss function is yet to be investigated using PET modality data.

In this section a CNN (g) with internal parameters (Φ) is used to denoise a reconstructed PET image (θ). The training of the CNN's internal parameters can be expressed in terms of a loss function, here we once again use the L2 loss function with ground truth image labels:

$$\Phi = \arg \min_{\Phi} \frac{1}{N_t} \sum_{t=1}^{N_t} \|g(\theta_t; \Phi) - \theta_t^{GT}\|_2^2 \quad (6.11)$$

where as before θ^{GT} represents the ground truth training labels. The noisy reconstructed image can be constructed from the ground truth image θ^{GT} and an additive noise term ϵ :

$$\theta = \theta^{GT} + \epsilon \quad (6.12)$$

Instead of the above loss function (Eq. 6.11), the Noise2Noise framework seeks to train the network parameters (Φ) using noisy data labels:

$$\Phi = \arg \min_{\Phi} \frac{1}{N_t} \sum_{t=1}^{N_t} \|g(\theta_t^{GT} + \epsilon_{t1}; \Phi_1) - (\theta_t^{GT} + \epsilon_{t2})\|_2^2 \quad (6.13)$$

Two independent noise realisations can be easily acquired from PET imaging datasets through resampling the projection data without replacement into independent sinograms. Each count is randomly assigned into dataset 1 or 2, partitioning the data into two half count noise realisations. Both noise realisations are then separately reconstructed to produce two noise independent images. These two independent noisy reconstructed images are represented by $\theta_1 \equiv \theta^{GT} + \epsilon_1$ and $\theta_2 \equiv \theta^{GT} + \epsilon_2$. A schematic providing an overview of the Noise2Noise training and testing framework is provided in Fig. 6.1 The dependency of zero-mean noise requires the use of linear reconstruction methods such as FBP. Methods such as MLEM are non-linear and lead to positively biased mean values in low count regions. The consensus loss function extends the Noise2Noise framework to combine the two noise realisations in the output of the network. The above network learnable parameter optimisation is replaced with:

$$\begin{aligned} \Phi = \arg \min_{\Phi} \frac{1}{N_t} \sum_{t=1}^{N_t} & \frac{1}{2} \|g(\theta_t^{GT} + \epsilon_{t1}; \Phi_1) - (\theta_t^{GT} + \epsilon_{t2})\|_2^2 + \\ & \frac{1}{2} \|g(\theta_t^{GT} + \epsilon_{t2}; \Phi_2) - (\theta_t^{GT} + \epsilon_{t1})\|_2^2 - \\ & \frac{1}{4} \|g(\theta_t^{GT} + \epsilon_{t1}; \Phi_1) - g(\theta_t^{GT} + \epsilon_{t2}; \Phi_2)\|_2^2 \end{aligned} \quad (6.14)$$

Two additional regularising terms are added to this loss function. The first is the L2 norm of the learnable parameter values. The second is the L2 norm of the difference between the noisy full counts reconstructed image and the output image. The effective level of image denoising can be controlled by varying the level of the second regularisation term, via a hyperparameter β_r . Small values of β_r encourage the output image (as expressed in Eq.

6.15) to be similar to the full counts input image, whilst large values of β_r allow the output image to be dissimilar (high levels of regularisation).

$$\theta = \frac{g(\theta^{GT} + \epsilon_1; \Phi_1) + g(\theta^{GT} + \epsilon_2; \Phi_2)}{2} \quad (6.15)$$

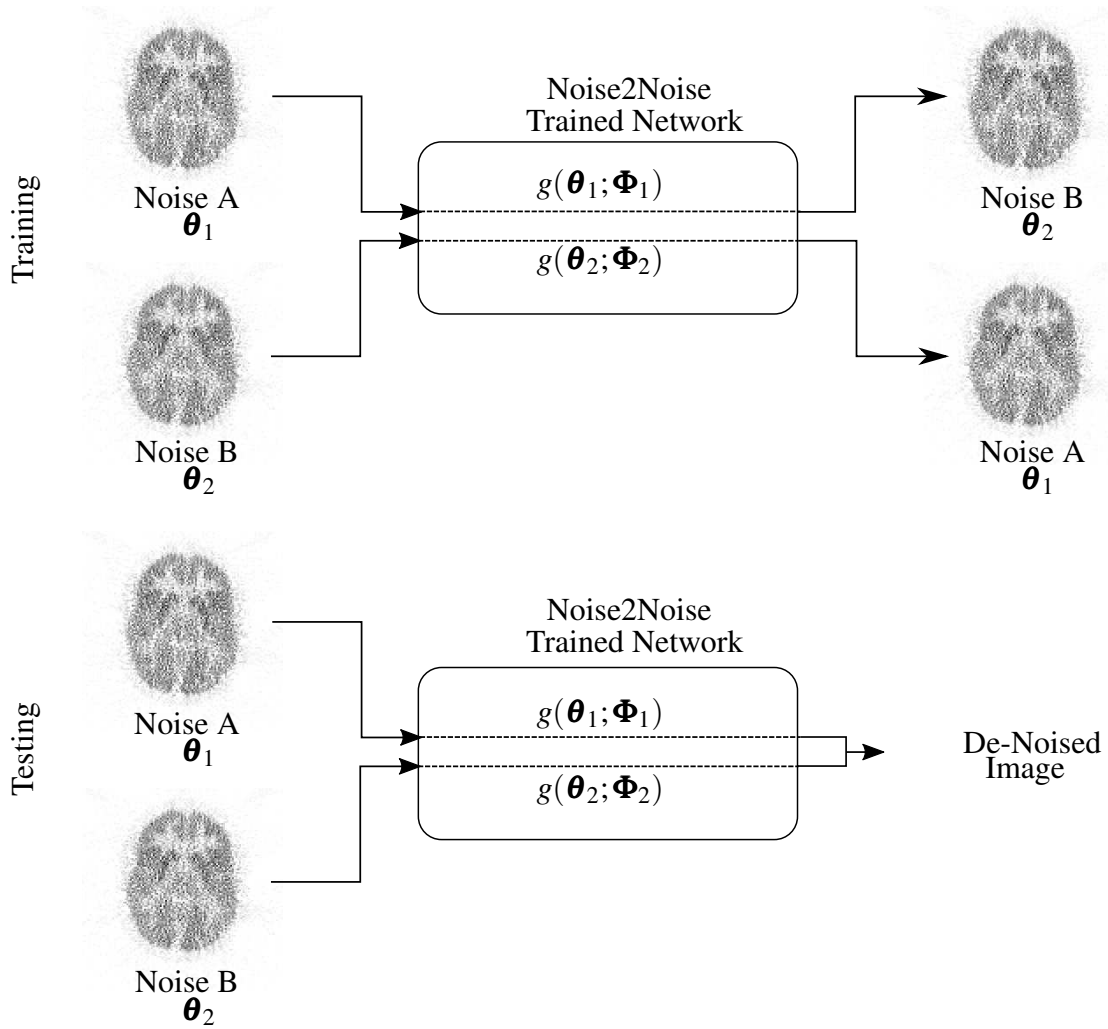


Fig. 6.1 Schematic of the Noise2Noise framework, showing how it used to train a neural network using unlabelled data, and test (predict) also using unlabelled data.

6.2 Simulated Phantoms and Patient Data

6.2.1 2D Simulated Phantom

The network was trained and evaluated using a simulated [^{18}F]FDG PET dataset. The simulated PET dataset was constructed using the BrainWeb database that consists of 20 3D segmented MR images. The MR images were transformed to [^{18}F]FDG PET phantoms as previously presented in chapter 5. These PET phantoms were repositioned and resized in accordance to the voxel and matrix size of the Siemens Biograph mMR scanner. The simulated database was split in to training and testing categories, consisting of 16 and 4 phantoms respectively. 41 transverse slices were automatically selected from each 3D phantom focusing on the brain, and ignoring slices consisting of empty space or the neck/shoulders of the patient. These slices were treated as individual 2D images, producing an overall database of 820 images. These images were then projected into sinogram space, with Poisson noise applied (10^6 counts), to produce the noisy projection data, in accordance with the previous chapters simulated data. All images were reconstructed without PSF modelling in this chapter.

6.2.2 3D Simulated Phantom

In contrast to the above simulated dataset, each phantom was treated as a single 3D dataset, and accordingly reconstructed. Only one phantom was reserved for the testing database, with the remaining 19 phantoms comprising the training database. These 3D phantoms were projected into sinogram space, with Poisson noise applied (2×10^8) counts, to produce noisy projection data, in accordance with previous chapters simulated data. For the Noise2Noise training schemes these sinograms were then split into two (1×10^8) datasets that were individually reconstructed using a least squares objective function minimised via preconditioned conjugate gradient decent (CGD), due to the requirement for zero mean error [169, 202].

The CGD algorithm (as shown in §2.6) is combined with an EM style pre-conditioner, to improve convergence speeds [178, 204, 168].

$$\mathbf{C}^n = \text{diag}\left\{\frac{\theta_j^n + \delta}{\sum_i a_{ij}}\right\} \quad (6.16)$$

As presented in [341] a small positive value is added to ensure the pre-conditioner matrix is positive definite ($\delta \approx 0.01 \times \text{max value}$). Due to the increased convergence speed of the pre-conditioned CGD algorithm a reduced number of updates (50) were used in comparison to MLEM (300). The reconstructed image datasets underwent a number of pre-processing steps including cropping, centering and intensity scaling with respect to the maximum voxel intensity in each image dataset. Sixty image slices from each 3D image dataset were automatically selected, focusing on the brain, and ignoring slices consisting of empty space or the neck/shoulders of the patient.

6.2.3 3D PET Patient Scan Datasets

The PET cohort dataset consists of 34 patients that underwent a [^{18}F]FDG PET-MR scan on the Siemens biograph mMR scanner. The prompts counts level of the acquired PET datasets varied from 8.1×10^8 to 2.3×10^8 . Two of the patient datasets were discarded due to irregularities, with an additional dataset reserved for testing the supervised U-Net. Therefore, the training database consisted of the remaining 31 patient datasets. The full count sinogram datasets were re-binned (without replacement) into two half count sinogram datasets. These two independent noise realisations were reconstructed using the CGD algorithm as described above. The reconstructed 3D image datasets underwent a number of pre-processing steps, similar to that described for the simulated data. Sixty image slices from each 3D image dataset were automatically selected, focusing on the brain, again in a similar manner to the simulated data study.

6.3 2D Simulated Phantom Study

6.3.1 Experimental Methods

The proposed unrolled network architecture is shown in Fig. 6.2. Each module of the network consists of a EM update layer, a regularisation network, and a fusion layer, which combines the EM update image and the prior distribution in accordance to the Levitan Herman update equation (Eq. 6.9). The relationship between these three components is also shown in Fig. 6.2. The unrolled network is implemented in 2D, with the (uncropped) image dimensions equivalent to that of the Siemens mMR scanner. The regularising network is based on a residual network structure, with the input image also linked to the output via a skip connection. 64 convolutional kernels are used in each convolution layer of size 5×5 . Each convolution (Conv) layer is followed by a batch normalisation (BN) layer, and a leaky rectified linear unit (ReLU) layer. The proposed network was implemented using MATLAB's R2019a Deep Learning Toolbox. The Adam (adaptive moment estimation) algorithm is used for the optimisation of the network learnable parameters [342]. All training used a batch size of 10, for 100 epochs. A learning rate of 0.0001 and an L2 learnable parameter regularisation of 0.0001 were used. The network is trained end-to-end, with the output of module $n - 1$ connected to the input of module n . The first EM update is used as in the input image to the network.

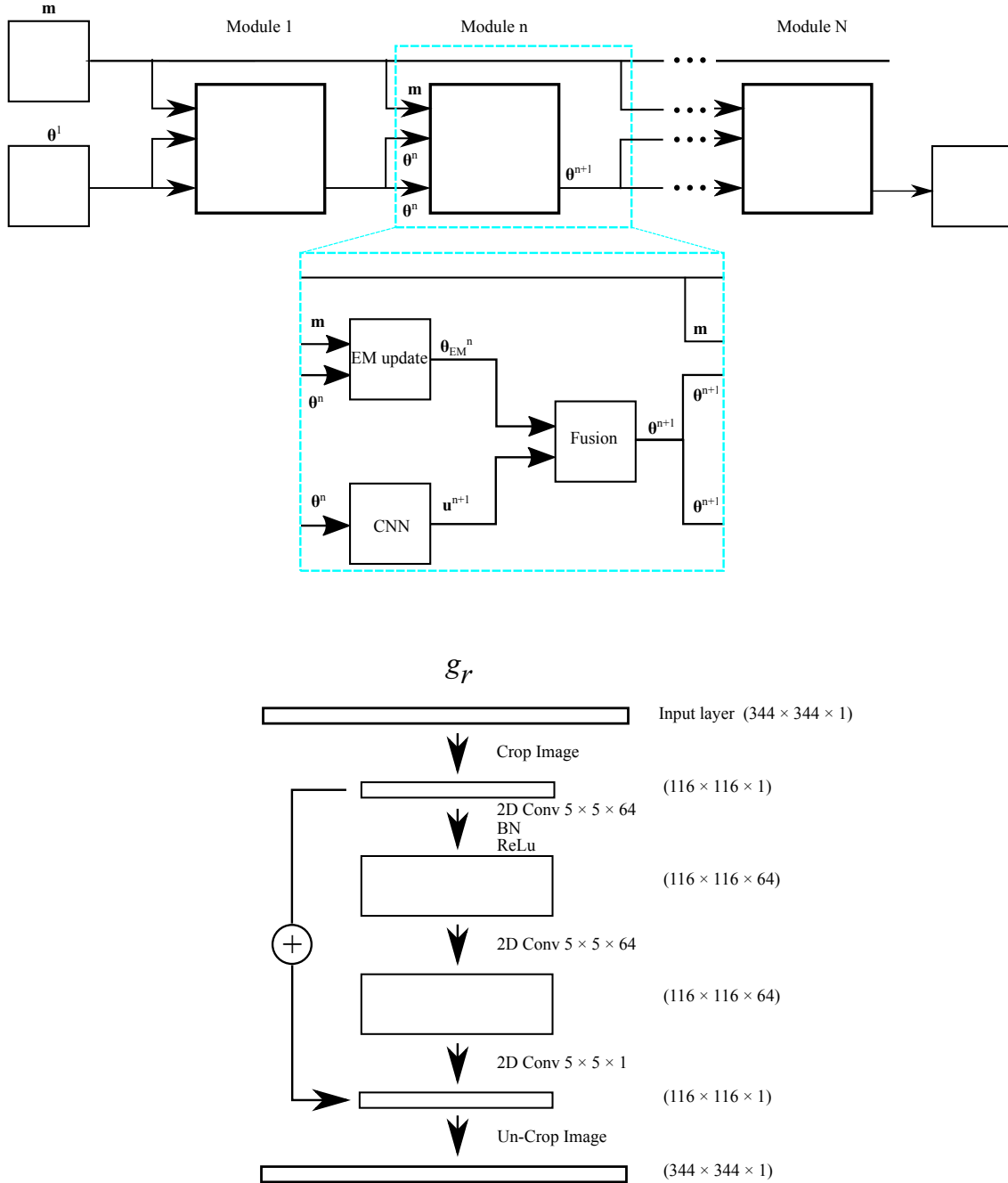


Fig. 6.2 Schematic showing the general architecture of the un-rolled network reconstruction algorithm, and the architecture of the CNN used in the internal regularising network (g_r). The network used a kernel filter width of 5 and was trained end-to-end for 100 epochs with a batch size of 10.

The networks were trained using noise free MLEM reconstructed images at 100 iterations as training labels. To fully evaluate the potential of unrolled networks their architecture shall be investigated with respect to the number of modules and the number of layers in the internal regularising network (g_r). The number of modules (each of which contains an EM update) is varied between 1, 2, 4 and 8. In addition, the number of the Conv-BN-ReLU layers present in the regularising network is varied between 2, 4 and 8. The regularising network directly acting on the input image without any data fidelity (EM update) layer is also investigated. Post-reconstruction smoothed MLEM and a post-reconstruction U-Net were also investigated to provide a suitable comparison. The architecture of the U-Net employed is shown in Fig. 6.3. The U-Net architecture chosen uses addition based skip connections that have proved popular in the PET reconstruction literature [338], although this does differ from the concatenation layers used in the original U-Net formulation [108]. The U-Net employs 3×3 kernels, in keeping with its implementation in the literature. The methods discussed are evaluated in terms of whole brain bias vs standard deviation trade-off as previously described in chapter 5, calculated using 10 noise realisations. The noise free MLEM reconstructed images were used as the ground truth for the calculation of these metrics.

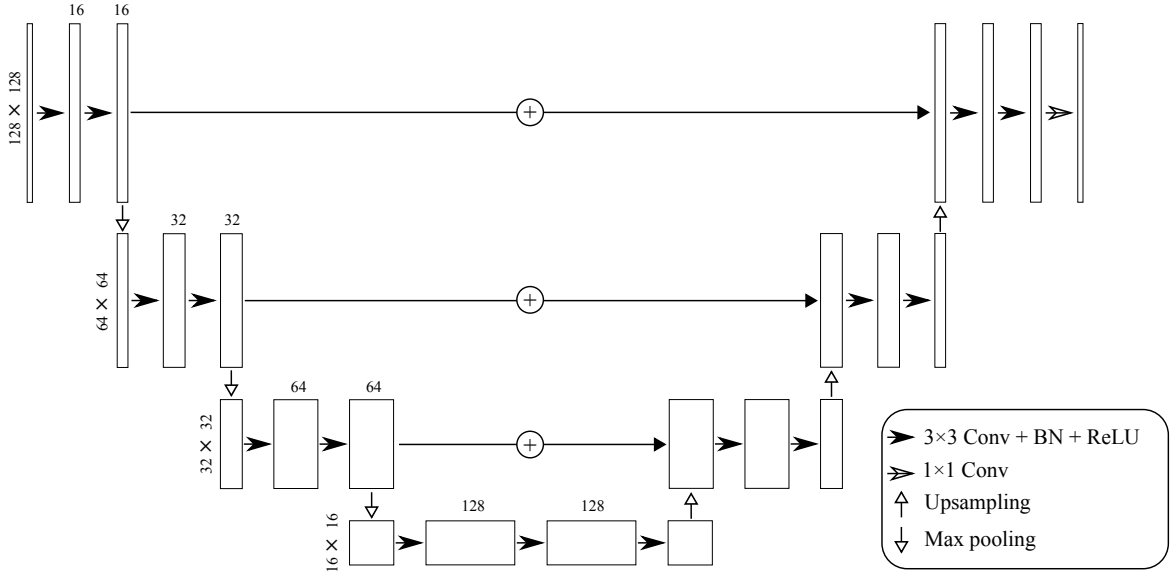


Fig. 6.3 Schematic of the post-reconstruction U-Net architecture used in this chapter. The 2D U-Net used for comparison with the unrolled network is fed each image slice individually. Whereas the Noise2Noise U-Net (2.5D) is fed three neighbouring slices at a time and trained to output the noise free central slice.

6.3.2 Results

The training loss curves of an unrolled network investigated and the post-reconstruction U-Net are shown in Fig. 6.4. All results and metrics are based solely on reconstructed images from the testing database. Figure 6.5 shows the bias vs standard deviation trade-off curve of the unrolled network for varying number of modules, and varying number of convolutional layers (within the regularisation network). The majority of unrolled network architectures are shown to outperform both post-smoothed MLEM and the post-reconstruction U-Net. Representative images for each reconstruction method are shown in Fig. 6.6.

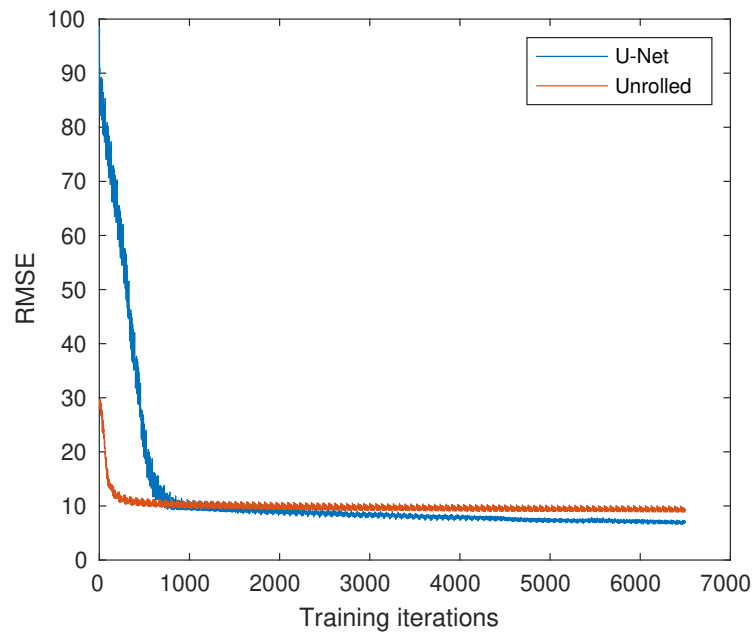


Fig. 6.4 Training RMSE curves for a representative unrolled network (4 modules with 4 layers) and the post-reconstruction U-Net architectures.

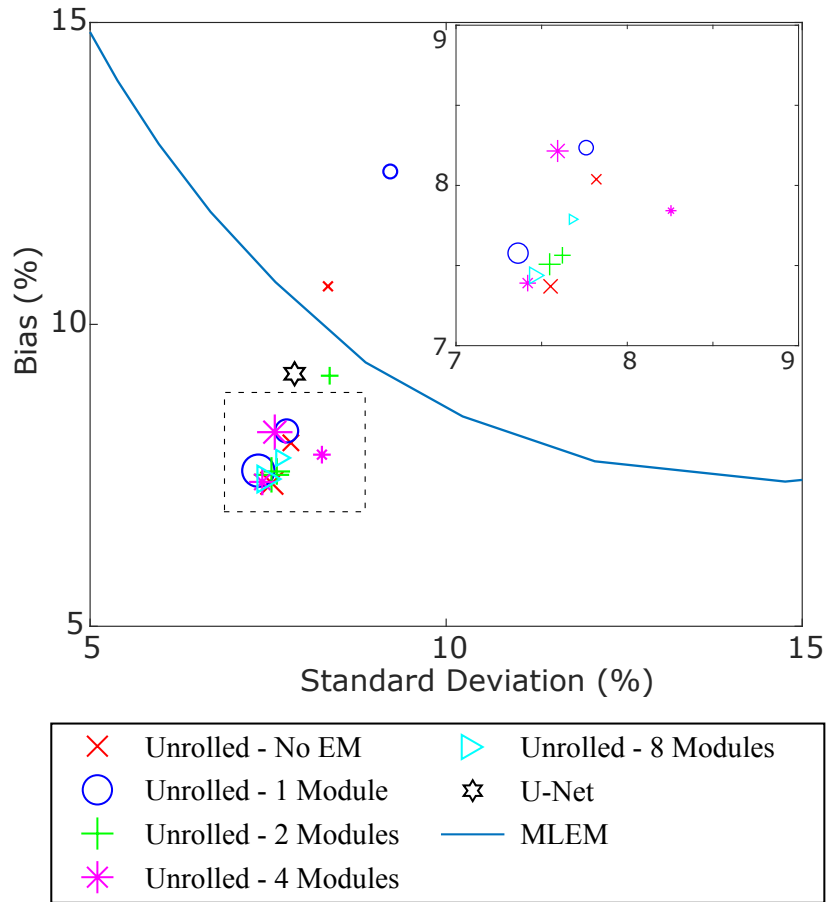


Fig. 6.5 Whole brain bias vs standard deviation calculated over 10 noise realisations for: post smoothed MLEM (at 100 iterations); post-reconstruction U-Net; varying unrolled networks with increasing numbers of modules (indicated by marker); and increasing number of layers within the regularisation network (indicated by increasing marker size). The unrolled network with no EM update layers is also shown. A zoomed in region is shown in the top right hand corner.

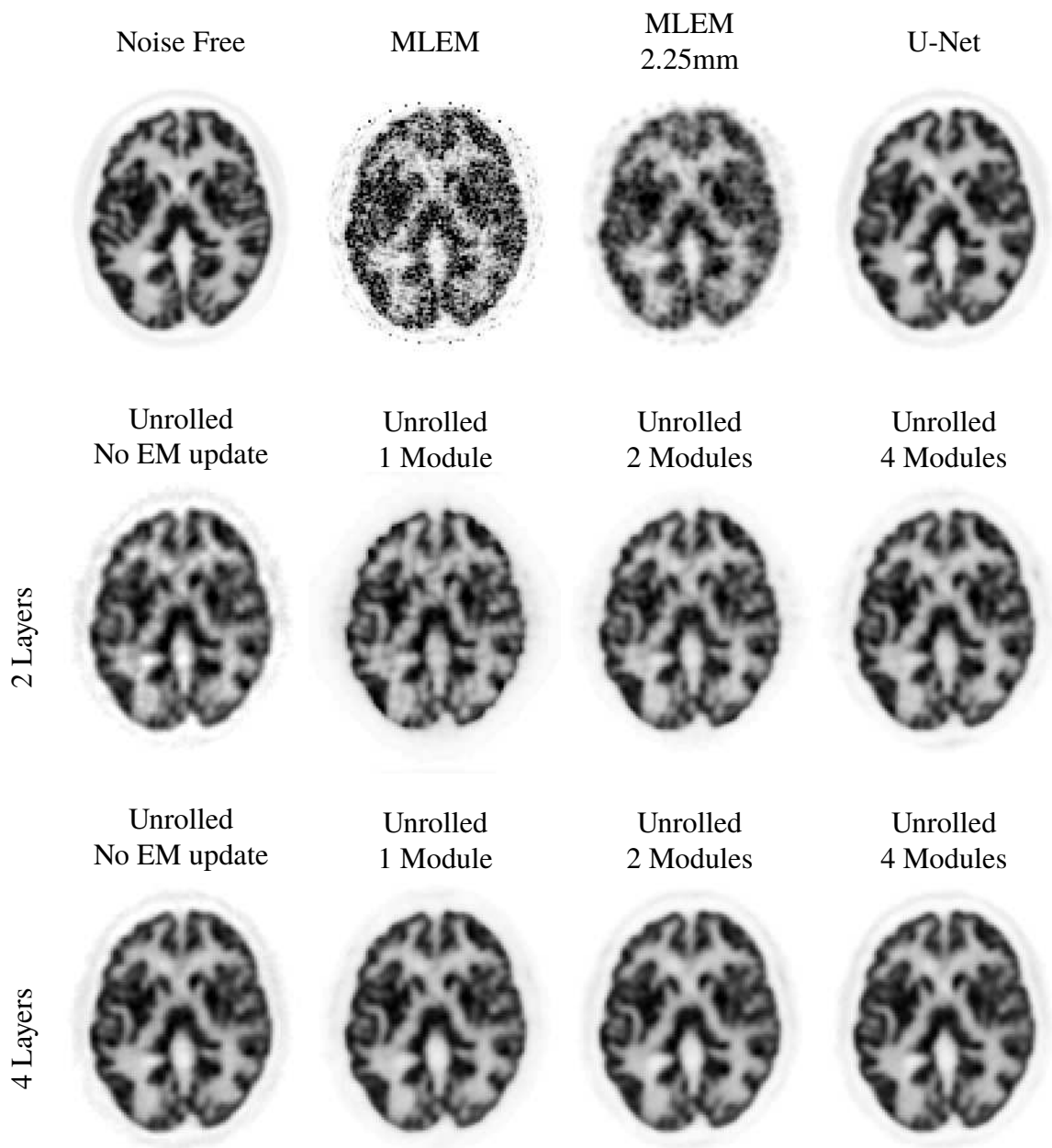


Fig. 6.6 Single slice images, for the 2D simulated dataset reconstructed either by post-smoothed MLEM (at 100 iterations), a post-reconstruction U-Net or an unrolled network with varying number of layers and modules.

6.4 3D Simulated Phantom Study

6.4.1 Experimental Methods

In the second part of this chapter the post-reconstruction U-Net employed above is now used to investigate the possible benefits of using the Noise2Noise framework on PET imaging data. The capability to train networks using noisy data labels could lead to the widespread application of deep learning methodologies on PET imaging datasets. In comparison to the U-Net used in §6.3, the U-Net trained using the Noise2Noise framework acts on multiple input slices to produce a single output slice. Through extending the U-Net in this manner, greater structural information can be extracted from the images, particularly in the transverse direction.

As described in §6.1.2 the Noise2Noise framework trains one half count reconstruction (input image) to match another independent half count reconstruction (training labels), as shown in Fig. 6.1. The network was trained for a total of 100 epochs, with a batch size of 10 for the first 50 epochs and a batch size of 30 for the remaining 50 epochs. A learning rate of 10^{-6} was used as well as an L2 parameter regularisation value of 10^{-3} . The hyperparameter β_r was varied producing U-Nets with varying levels of denoising power. Once more, the results are evaluated in terms of whole brain bias vs standard deviation trade-off as described in chapter 5, calculated using 50 noise realisations. The ground truth phantom was used as the ground truth for the calculation of these metrics. The Noise2Noise trained U-Net was compared to post-smoothed CGD. All results shown are for the single test dataset.

6.4.2 Results

Figure 6.7 shows the training RMSE curves of the Noise2Noise trained U-Net for the simulated (and real) datasets. This training curve is much noisier than the corresponding curve for the U-Net trained using noise-free image labels Fig. 6.4, due to the presence

of noise in the training labels. The effect of this could be reduced through using larger batch sizes, although not implemented here due to the relatively small training dataset used. Figure 6.8 shows the bias vs standard deviation trade-off curve of the U-Net trained via the Noise2Noise framework for varying levels of regularisation. Its denoising power is shown to vary with respect to the regularisation hyperparameter and tends to the unsmoothed CGD value. Images representing the hyperparameters indicated in Fig. 6.8 are shown in Fig. 6.9.

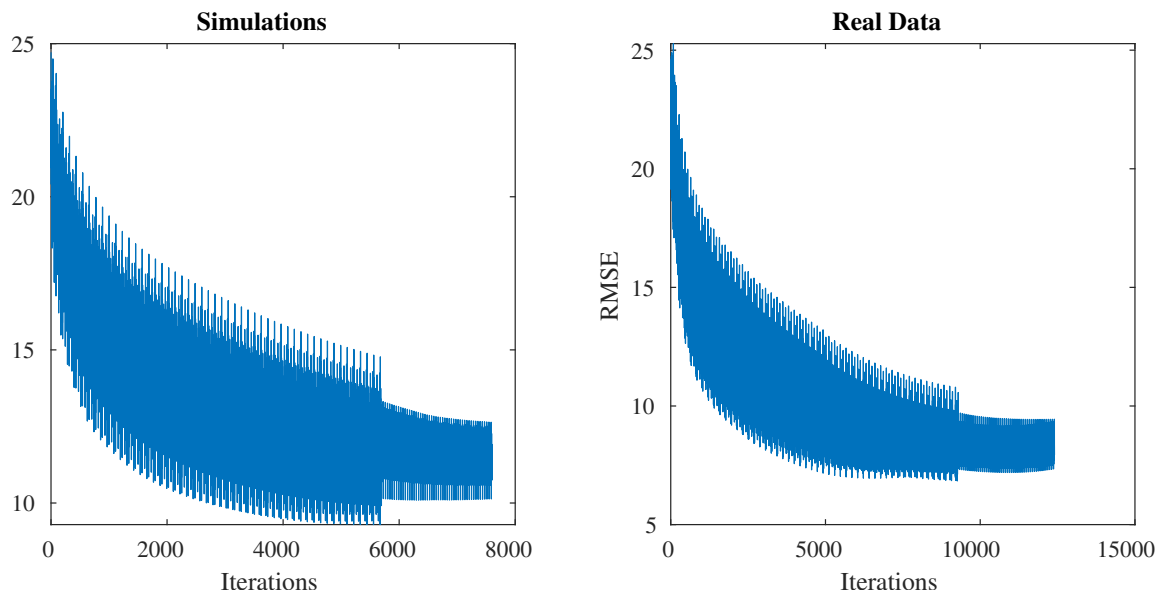


Fig. 6.7 Training RMSE curves of the Noise2Noise trained U-Net, for both the simulated (left), and real patient (right) datasets. Iteration in this figure refers to training iteration.

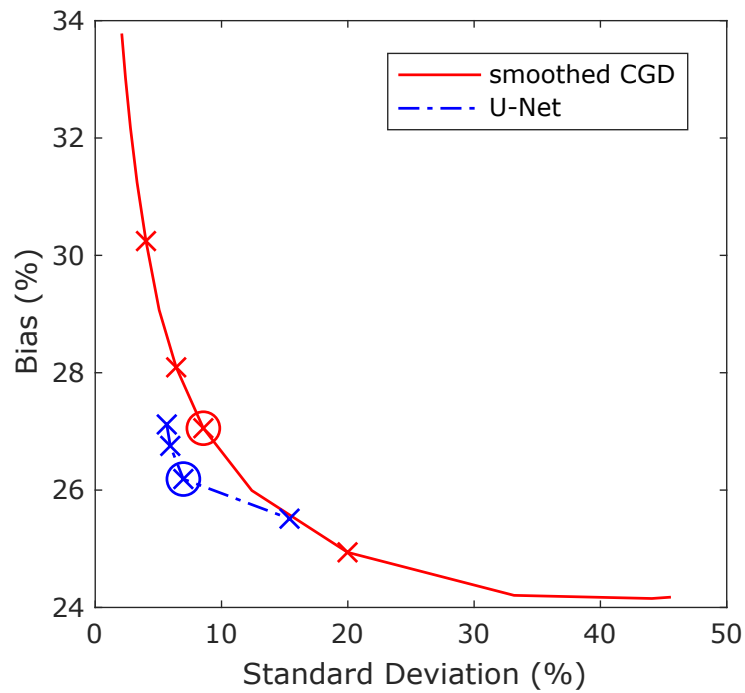


Fig. 6.8 Whole brain bias vs standard deviation curve for U-Net and CGD algorithm applied to a simulated dataset. Curves are shown for varying β_r parameter (U-Net) and varying level of post reconstruction smoothing (CGD). The metrics were calculated using 50 noise realisations of the test dataset. The data-points marked with crosses are shown in Fig. 6.9. The data-points marked with circles are the minimum NRMSE point for each curve.

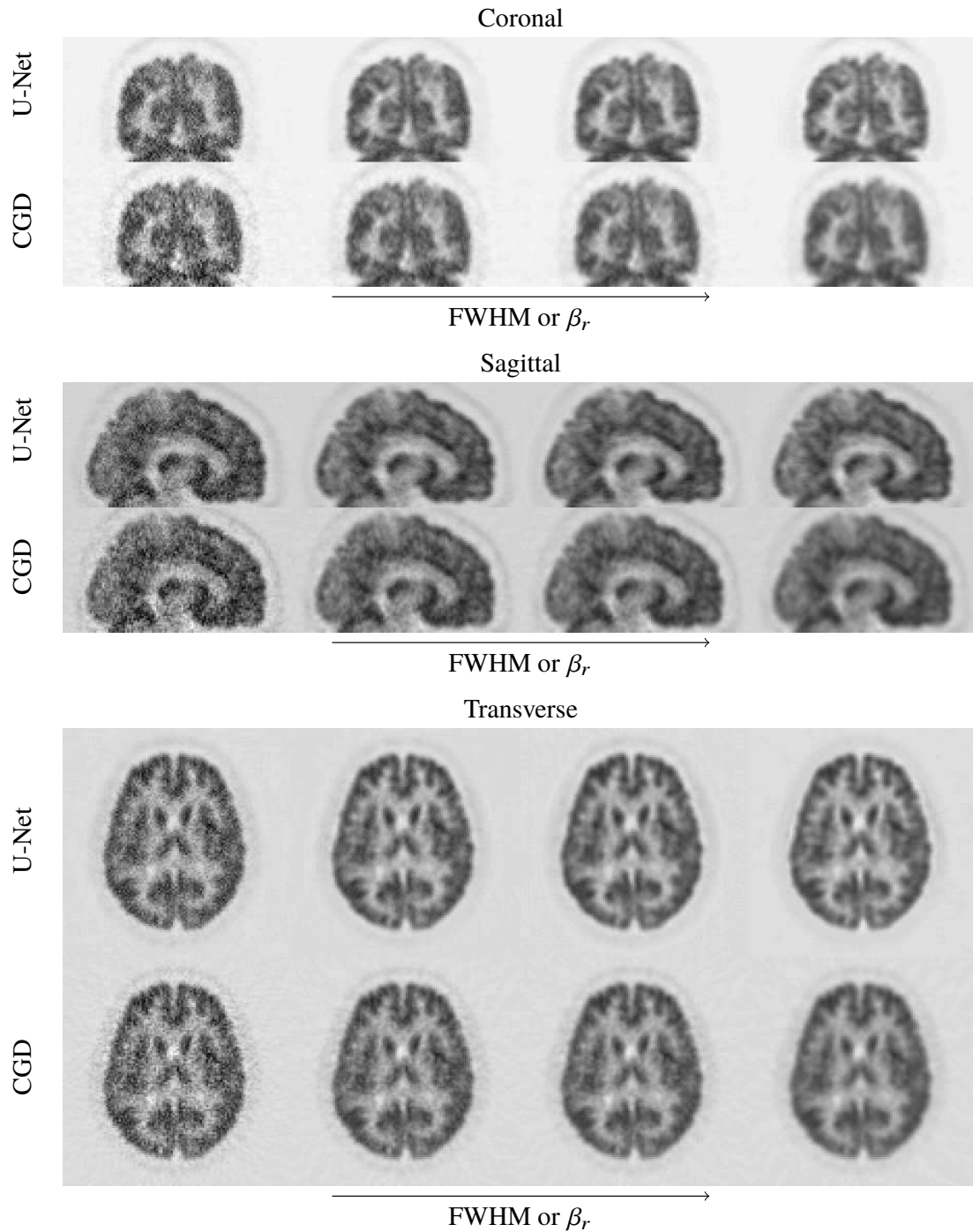


Fig. 6.9 Coronal, sagittal and transverse images of the post-reconstruction U-Net applied to a test (unseen simulated patient) dataset. The full count CGD reconstructed image with increasing levels of post-reconstruction smoothing is also shown for comparison.

6.5 Real Data Studies

6.5.1 Experimental Methods

The U-Net described in the Noise2Noise simulations is now applied in a post-reconstruction manner to real patient PET datasets. As with the simulated dataset, three adjacent slices were fed to the network, to estimate the central slice. The training parameters used were the same as for the simulated dataset. All results shown are for the single test dataset. As no ground truth image is known for the real dataset, only example images are shown in comparison to post-smoothed CGD.

6.5.2 Results

Figure 6.10 shows the Noise2Noise trained U-Net for varying levels of β_r in comparison to post-smoothed CGD. The visual improvements gained appear relatively minimal, although the caudate and putamen regions show improved contrast. Figure 6.11 also shows a subset of the activation images produced by passing the test dataset through the trained network.

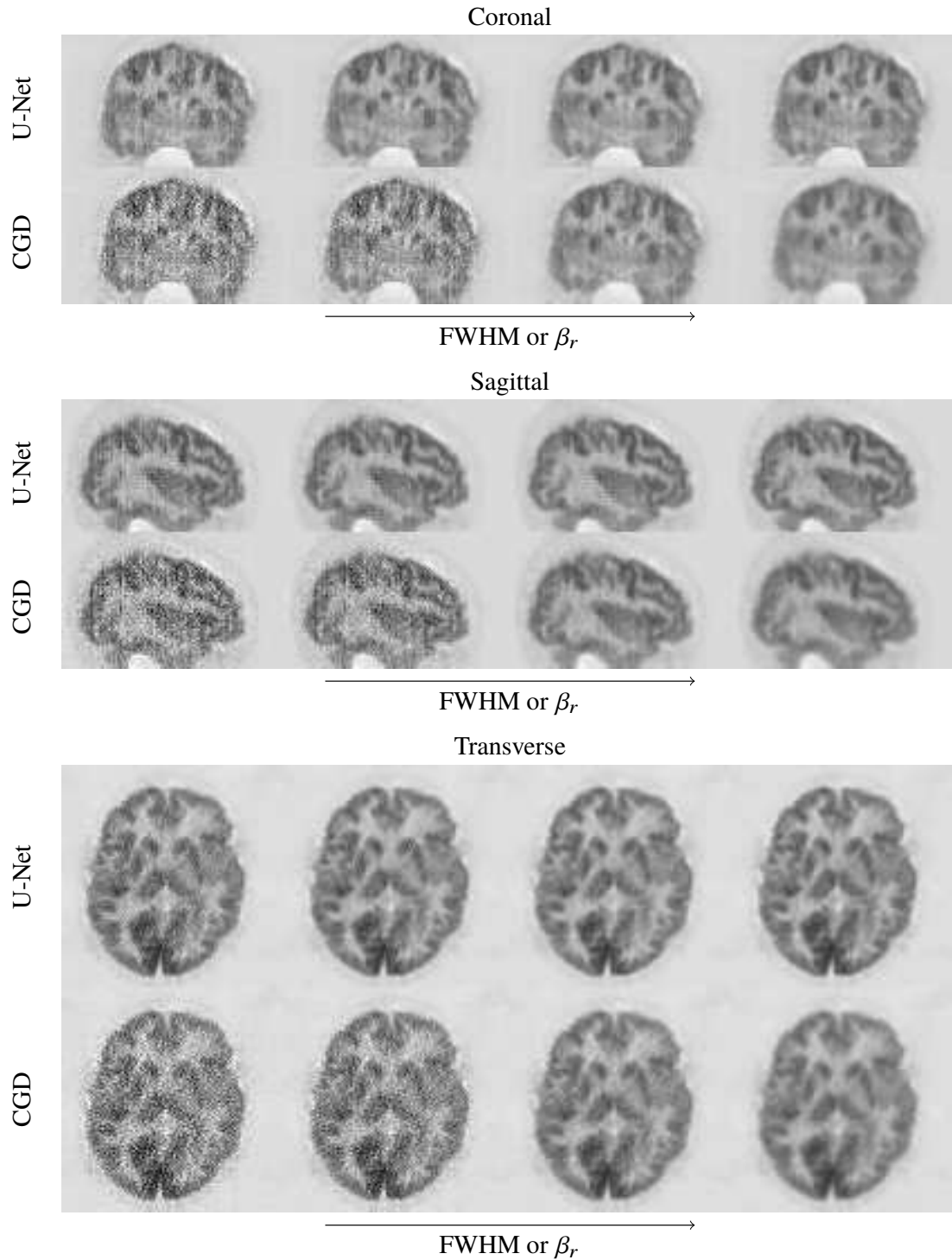


Fig. 6.10 Coronal, sagittal and transverse images of the post-reconstruction U-Net applied to a test (unseen real patient) dataset. The full count CGD reconstructed image with increasing levels of post-reconstruction smoothing is also shown for comparison.

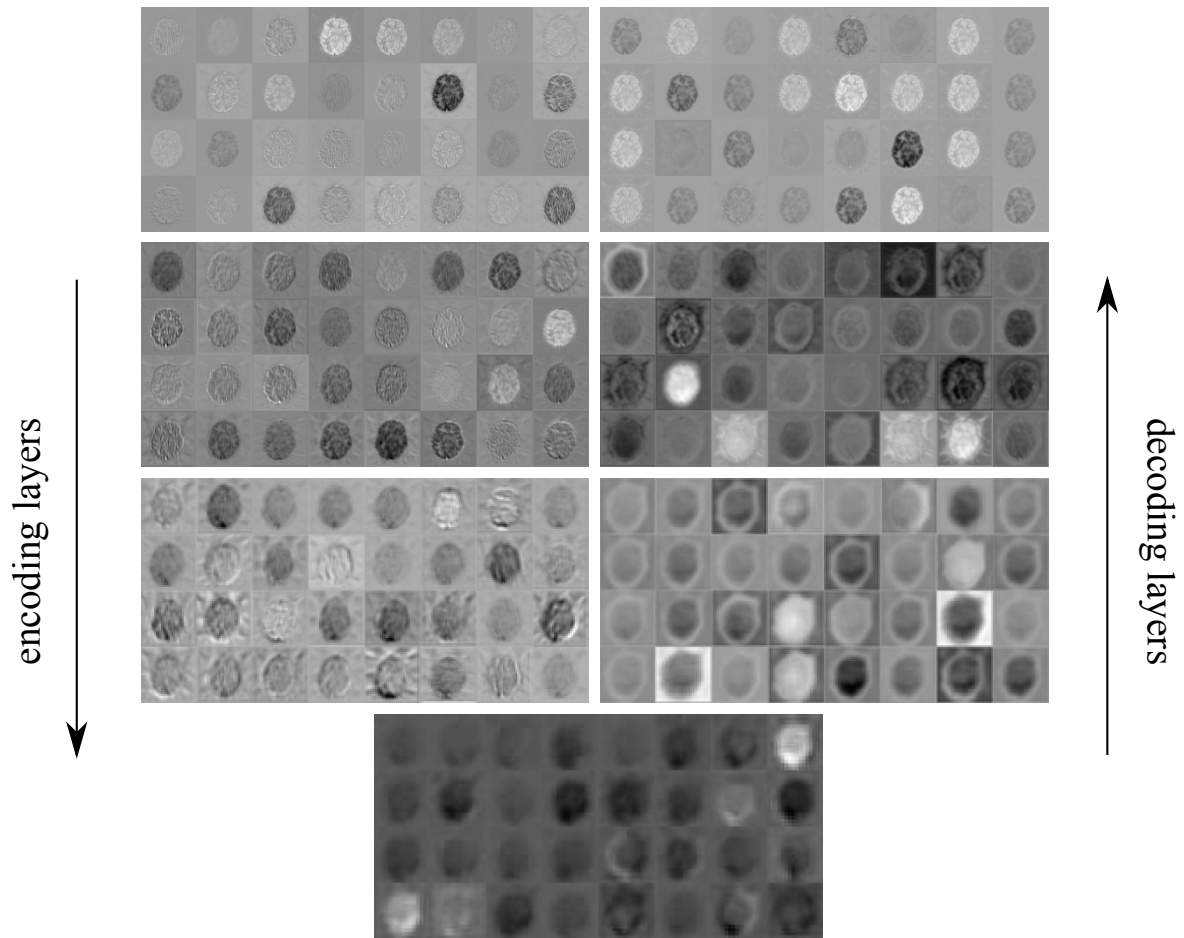


Fig. 6.11 Internal activation images of the post-reconstruction U-Net applied to a test (real patient) dataset. The activation images are shown for both encoding and decoding layers.

6.6 Discussion

6.6.1 Unrolled Networks

Figure 6.5 shows the whole brain bias vs standard deviation curve, for post smoothed MLEM, various unrolled networks with increasing numbers of modules, and with increasing numbers of convolutional layers. This figure also includes an unrolled network with no EM update layers (effectively a residual network applied directly to the first EM update image), and a post-reconstruction U-Net. The results in Fig. 6.5 demonstrate the unrolled network's

superior performance (when implemented with more than 1 module and 2 convolutional layers) , relative to post-smoothed MLEM and the post-reconstruction U-Net. The unrolled network produces images of increased image quality as shown in Fig. 6.6 and with reduced noise for a fixed bias level. The unrolled network consisting of 2 modules each containing 4 layers led to an accelerated reconstruction time by a factor of 16 in comparison to either post-reconstruction method, reducing the total run time from 16.8s to 1.04s for a set of 20 2D images. Therefore unrolled networks are able to simultaneously achieve both an accelerated reconstruction time and an improved image quality, relative to MLEM and to post-reconstruction networks.

Figure 6.5 also shows that as the number of layers and modules increase, the bias and standard deviation values tend to a roughly fixed point. This shows that for sufficient number of parameters, the network is now predominantly limited by the relatively small training database and not the network structure (modules and layers). In fact, further increasing the number of network layers or modules may reduce the network's performance, due to overfitting and poor generalisation when applied to the separate test dataset. This issue is exacerbated if the training data itself is limited and poorly represents the population distribution. The inclusion of the data fidelity term throughout an unrolled network may reduce the tendency of such networks to overfit for unseen data. That said, alternative network architectures may still outperform this point, if their architecture enables them to extract a greater amount of insight from the training database.

Rather counterintuitively removing the EM update layer completely from the unrolled network architecture is shown to produce equally good results to the other unrolled network variations, provided there is a suitable number of layers in the regularisation network (> 2). This is most likely due to the lack of variation in the simulated dataset, and the similarity of the testing dataset. In such situations even simple network architectures can produce favourable results. Real patient PET scans will present a much wider range in variation in

comparison to the simulated dataset used. The simplistic structure of the network without the EM update layer may also aid in network training, especially over the limited number of epochs investigated.

It also should be noted that the training RMSE curve of the U-Net outperforms all the unrolled network architectures, as shown in Fig. 6.4. However, the U-Nets capability to generalise to the unseen test data appears inferior to the majority of unrolled network variations, as shown by Fig. 6.5. A limited number of alternative parameter choices for the U-Net were also investigated but still suffered from this overfitting problem. A simple solution to this issue, not undertaken here, is to train the network simultaneously on a validation set, and terminate training when the validation error begins to increase. Through using a validation dataset to address the overfitting issue of the U-Net its performance may exceed that of the unrolled network investigated so far.

Additional work is required to replicate these results with real data, where the increased inter-dataset variability is likely to improve the performance achieved by unrolled networks built from multiple EM update-denoising modules. Although as previously stated, there remains many currently unsolved challenges in attaining a suitable ground truth for real PET studies.

The unrolled network investigated in this work provided major speed enhancements and improved the overall image quality, in comparison to iterative reconstruction methods. However, the inclusion of the projection based updates into the network architecture leads to inhibitive large memory requirements, long training times and other training complexities, which are exacerbated when training on real patient 3D datasets. In addition, the Noise2Noise framework has currently only been shown for networks acting on already reconstructed images. As such, the post reconstruction U-Net architecture is used for the remainder of this chapter, which focuses on the Noise2Noise framework.

6.6.2 Noise2Noise

Figure 6.8 shows the bias vs standard deviation curve for the post-reconstruction U-Net trained using noisy labels only (in accordance to the Noise2Noise framework), and compared to post-smoothed CGD. The performance of the U-Net exceeds that of the post-smoothed CGD with increasing values of β_r . At small values of β_r the U-Net's performance tends to that of the post-smoothed CGD. Figure 6.9 shows the corresponding images of the U-Net and post-smoothed CGD for the parameters indicated in Fig. 6.8. The visual improvement of the U-Net over the CGD is marginal, however slight increases in contrast can be seen in the caudate and putamen regions. Therefore although these numerical and visual results are not outstanding they do show that the Noise2Noise framework can be applied to noisy PET datasets (in 3D), which has not previously been demonstrated in the literature [119]. These superior results for the Noise2Noise trained U-Net were achieved with limited investigation of the U-Net architecture, therefore other architectures may produce preferential results. Although the parameters chosen in this set up match those in the literature [338].

One particularly challenging issue of 3D PET datasets is the variation in noise level, both intra-dataset and inter-dataset. Intra-dataset noise is due to the spatial variation in the radio-tracer distribution and scanner sensitivity. Additionally, a PET system's spatial resolution varies spatially, requiring the use of a spatially variant point spread function to accurately model. The capacity of the network to effectively denoise 3D datasets demonstrates its ability to cope with these challenges, even without the ground truth training labels. Inter-patient variation is generally more extensive for real patient datasets in comparison to simulations. Further validation on real data (partially shown in the real data studies §6.5) is required to illustrate the capability of such networks to train effectively on real PET datasets.

Figure 6.10 shows the reconstructed images for both the Noise2Noise trained U-Net and post-reconstruction smooth CGD algorithm, applied to a real [^{18}F]FDG patient dataset. The U-Net was trained entirely from real patient datasets in accordance to the Noise2Noise

framework. Once more the Noise2Noise trained U-Net provides (marginal) visual improvement relative to post-smoothed CGD. Unfortunately, the real data results cannot be easily quantified given the lack of a ground truth. The activation images of the U-Net, shown in Fig. 6.11 show the structures and features that the U-Net extracts from the noisy image, and combines to form the denoised image. The results presented here illustrate the potential of the Noise2Noise framework for training PET image denoising networks. However, due to the limited time spent investigating alternative architectures and hyperparameters the results presented are unlikely to represent the full power of the method.

6.7 Limitations

The work encompassed by this chapter has only been investigated to a preliminary extent. As such, there are numerous limitations, and clear need for more extensive future work. The unrolled networks presented in §6.3 provide evaluation of the impact on image quality due to variation in the number of layers and the number of denoising modules. The unrolled network's architecture employed in this work borrows extensively from other literature in the PET reconstruction field [276, 275]. A comparison between the various unrolled networks present in the literature is required to clarify the optimal network architecture, and the respective advantages of each method. Showing where such unrolled networks stand in relation to the more well established MR-informed MAP methods (i.e. chapter 5), would also help validate the utility of deep learning within the PET reconstruction field. Limited results have also been shown in the literature which extend the unrolled architecture to 3D datasets [275] and real patient datasets [272]. Such advances should also be included in the presented unrolled network. A second limitation of the presented unrolled network (in addition to the literature networks), is the use of a relatively small and synthetic training database. This reduces the capability of such networks to generalise effectively to clinical

datasets. As discussed earlier in this chapter, PET imaging lacks extensive training databases highlighting the potential of the Noise2Noise framework.

The Noise2Noise framework shown in this chapter on real patient 3D PET datasets shows promising results. However, the current improvement between the Noise2Noise framework and post-smoothed CGD is minimal. More extensive investigation is required into the architecture of the network used, and the training process (i.e. number of epochs and batch size) to maximise the power of the Noise2Noise framework. Increasing the training database, and extending the network to 3D (as opposed to 2.5D) would also likely lead to improvements in the output image quality. Lastly, the Noise2Noise trained network also needs to be compared to post-smoothed MLEM and MAP based reconstruction.

6.8 Summary

The preliminary findings in this chapter show the power of unrolled networks in denoising simulated PET images, in comparison to popular iterative reconstruction methods and a post-reconstruction U-Net. However, the direct application of such networks to real PET data is difficult due to the lack of appropriate ground truth. The use of the Noise2Noise framework, which has previously been evaluated for other imaging modalities, shows promise in providing a suitable means for training networks using real patient 3D PET imaging data. The real data results presented in this chapter show only marginal improvement in comparison to post-reconstruction smoothing, however alternative architectures and increased training datasets are likely to improve these results further.

Chapter 7

Conclusions and Future Work

7.1 Summary

The emergence of simultaneous PET-MR has provided the capability to acquire corresponding MRI images in a clinically practical manner. The inclusion of MR images as a form of prior information into the PET reconstruction process provides many opportunities to improve the reconstruction of count-limited and spatial resolution limited PET datasets. A second source of prior information is present in large population datasets, which can be utilised through supervised machine learning methods. Deep learning methods in particular have shown recent state-of-the-art performance in multiple computer vision tasks predicated by their widespread uptake in the research community and beyond. This thesis provides original contributions for both MR-informed PET reconstruction methodologies, and machine learning inspired reconstruction methodologies, focusing on the development, validation and comparison of these PET reconstruction methods. The main contributions of this thesis include:

- Investigation of the MR-informed kernel method for reconstructing extremely low count PET datasets, as low as 1% the original clinical count level. The MR-informed kernel method provides a simple framework to include MR-information into the

reconstruction process, without the requirement for tuning a MAP regularisation hyperparameter β . The KEM reconstructed images using 10% of the data had comparable whole brain NRMSE to the MLEM reconstructed images using 100% of the data. Such low dose capabilities of the MR-informed kernel method have previously not been demonstrated in the literature for clinical PET datasets.

- Development of the MR-informed kernel method to include a composite spatial-intensity feature vector (KEM LVS), and to produce reconstructed images at native MR voxel size. The composite feature vector enhanced the kernel method's ability to recover high intensity features that were unique to the PET data (i.e. not present in the MR), providing reduced image noise (across the whole brain) whilst maintaining the mean value of PET-unique regions.
- Comparison and clarification of the numerous proposed MR-informed PET reconstruction strategies. This work investigated two distinct branches of PET reconstruction, comparing multiple prominent MAP and kernel based methods, for the case of both shared as well as discrepant features in the PET and MR datasets. This work also identified the relative performance ranking of each method, and the level of regularisation permissible with unbiased reconstruction of the PET-unique features. For the high count datasets investigated, the anato-functional method provided the best SSIM and NRMSE whole brain to PET-unique region trade-off. Whereas, for the low count simulated datasets, KEM LVS achieved the best SSIM and NRMSE trade-off.
- Investigation into unrolled image reconstruction networks and the relative performance benefits of numerous architecture variants including number of regularisation layers, and number of data fidelity update layers. This work showed the power of unrolled networks for denoising and accelerating PET reconstruction, with unrolled networks

achieving a 16 times speed increase and a reduced whole brain NRMSE, in comparison to MLEM and a post-reconstruction U-Net.

- Validation of the Noise2Noise network training framework on real patient 3D PET datasets. This contribution showed the ability to train regression based neural networks on noisy PET labels.

7.2 Limitations

MR-informed PET reconstruction is devised on the premise of shared structural information between the PET and MR datasets. For the vast majority of neurological datasets, the structural information present in [^{18}F]FDG PET images will closely match the corresponding co-registered MR image. Through using appropriate levels of MR-informed reparameterisation, PET counts levels can be reduced to as low as 5%, whilst maintaining the visual image quality achieved by reconstructing 100% of the data using MLEM. However, for MR-informed strategies to be used clinically they need to be robust to PET-MR structural mismatch. Comparing a series of prominent MR-informed methods shows that low levels of regularisation (or reparameterisation) can be used to attain improved image quality without suppressing PET-unique features. However, a clear limitation in this finding is the level of regularisation that can be applied before biasing PET-unique features. High levels of regularisation are important for suppressing noise and reducing PVEs that affect regional quantification. As stated in chapter 6 alternative sources of prior information could enable dramatic image quality improvements, without negatively biasing the image.

An important factor in the evaluation of all reconstruction methods is the choice of hyperparameter values. For the comparison of MR-informed methods in chapter 5 hyperparameters were selected through a grid search, with hyperparameters selected that optimise a chosen metric. It should be noted that the accuracy of the chosen hyperparameters is limited by the

hyperparameter grid spacing, with the possibility of improved performances being attained through using hyperparameter values outside those evaluated. Improved performance of methods may also be achieved through increased hyperparameter sampling, however this is at the expense of increased computational time. Due to the complexity of the system model and the computational cost for 3D datasets few other hyperparameter optimisation techniques were practically feasible. The optimal hyperparameters selected are also dependent on the chosen metric. Unsurprisingly different metrics will lead to different optimal hyperparameters, which may not be optimal when assessed using a variety of alternative metrics. Selecting the optimal hyperparameters for real patient datasets is also troublesome due to the lack of a ground truth image to assess the reconstructed images. Physical phantoms such the NEMA [283] or Hoffman [284] phantom, or pre-clinical autoradiography studies could be used to attain hyperparameters in place of simulated phantoms. However, differences may still arise between the real patient dataset and the phantom with respect to count level and structures present, that will alter the optimal hyperparameters.

Numerical image quality metrics have been predominantly used throughout this thesis to assess the reconstruction methods under evaluation. Such metrics are useful for assessing regional quantification and noise properties. However, the clinical relevance of numerical image quality improvements is hard to quantify in terms of patient outcome. ROC curves and other forms of task based image assessment could be used to extend the presented comparison of MR-informed methods. MR-informed reconstruction methods could be assessed in terms of tumour detection, quantification or staging, whereby for instance regularised methods could increase clinicians' confidence in identifying genuine high intensity lesions. A wide variety of other task based assessments are available, including pattern recognition, diagnostic confidence, and patient outcomes. The difficulty in such forms of image assessment mainly lies in finding a number of highly skilled clinicians willing to undertake such time consuming tasks.

7.3 Future Work

A natural extension of the comparison work presented in chapter 5 would be to include deep learning inspired methodologies, a subset of which were investigated in chapter 6. This would allow the denoising power of recently developed deep learning methods to be placed in the wider context of MR-informed PET reconstruction. Establishing the relative performance of such methods is important, in order to justify the recent trend to prefer deep learning over model based iterative reconstruction methods. The development of unrolled networks could potentially combine the individual benefits of post-reconstruction networks and model based iterative reconstruction. Further comparisons could also include training networks using the Noise2Noise framework in addition to noise free training labels. This would help determine the utility of the Noise2Noise framework through quantifying the difference between using noisy and noise free training labels.

The preliminary work in deep learning (chapter 6) has many possible avenues that could be undertaken. The inclusion of population based prior information can also be combined with patient specific MR information, possibly providing improvements in image quality that exceeds what MR-informed and deep learning inspired methods can provide individually. This could lead to extremely low dose PET imaging becoming a clinical reality, or aid MR-informed regularisation whilst retaining genuine PET-unique regions.

Further validation is required for using noisy labels to train deep neural networks on real PET datasets. Factors of particular importance include the network's architecture, currently only a U-Net (with a limited parameter search) has been investigated but other architectures such as ResNet, and unrolled networks could respond more favourably to noisy training labels. The network should also be extended to 3D, allowing the receptive field of the network to encompass the entire image volume. Lastly, the network should be trained with an increased number of [^{18}F]FDG datasets. Large number of training samples are likely to

be of even greater importance than with networks trained using the ground truth, due to the need to effectively average out the noise present in the data labels.

Publications

This thesis is based on the following publications submitted during the course of my PhD:

- Bland, J., Mehranian, A., Belzunce, M. A., Ellis, S., McGinnity, C. J., Hammers, A., & Reader, A. J. (2018). MR-Guided Kernel EM Reconstruction for Reduced Dose PET Imaging. *IEEE transactions on radiation and plasma medical sciences*, 2(3), 235–243.
- Bland, J., Belzunce, M. A., Ellis, S., McGinnity, C. J., Hammers, A., & Reader, A. J. (2018). Spatially Compact MR-Guided Kernel EM for PET Image Reconstruction. *IEEE transactions on radiation and plasma medical sciences*, 2(5), 470-482.
- Bland, J., Mehranian, A., Belzunce, M. A., da Costa-Luis, C., Ellis, S., McGinnity, C. J., Hammers, A., & Reader, A. J. (2019). Intercomparison of MR-Informed Methods for PET Image Reconstruction. *Medical Physics*, (in press).

The following work was presented at the specified conferences throughout the course of my PhD:

- Bland, J., Mehranian, A., Belzunce, M. A., Ellis, S., McGinnity, C. J., Hammers, A., & Reader, A. J. (2017). MR-Derived Spatial Basis Functions for Reduced Dose PET Image Reconstruction . *6th conference on PET/MR and SPECT/MR, (PSMR)*
- Bland, J., Mehranian, A., Belzunce, M. A., Ellis, S., McGinnity, C. J., Hammers, A., & Reader, A. J. (2017). MR-Resolution Kernel Method for PET Reconstruction. *IEEE Nuclear Science Symposium and Medical Imaging Conference, (NSS/MIC)*

- Belzunce, M. A., Mehranian, A., Bland, J., & Reader, A. J. (2017). ML and MAP PET reconstruction with MR-voxel sizes for simultaneous PET-MR. *IEEE Nuclear Science Symposium and Medical Imaging Conference, (NSS/MIC)* .

In this work I provided discussion and feedback for the MR-voxel size resampling and reconstruction.

- Bland, J., Mehranian, A., Belzunce, M. A., da Costa-Luis, C., Ellis, S., McGinnity, C. J., Hammers, A., & Reader, A. J. (2018). Intercomparison of MR-Informed Methods for PET Image Reconstruction. *IEEE Nuclear Science Symposium and Medical Imaging Conference, (NSS/MIC)*

- Mehranian, A., Bland, J., McGinnity, C. J., Hammers, A., and Reader, A. J. (2019). Clinical assessment of MR-assisted PET image reconstruction algorithms for low-dose brain PET imaging. *IEEE Nuclear Science Symposium and Medical Imaging Conference, (NSS/MIC)*.

In this work I performed the MR-informed kernel reconstruction of the low dose datasets, and supported the study generally.

- Bland, J., Mehranian, A., da Costa-Luis, C., & Reader, A. J. (2019). Multi-Module Deep Learning for Enhanced and Accelerated PET Image Reconstruction. *IEEE Nuclear Science Symposium and Medical Imaging Conference, (NSS/MIC)*

References

- [1] Tarik F Massoud and Sanjiv S Gambhir. Molecular imaging in living subjects: seeing fundamental biological processes in a new light. *Genes and Development*, 17(5):545–580, 2003.
- [2] Moritz F. Kircher and Jürgen K. Willmann. Molecular body imaging: MR imaging, CT, and US. Part I. Principles. *Radiology*, 263(3):633–643, 2012.
- [3] Alexander L. Klibanov, Peter T. Rasche, Michael S. Hughes, Jolette K. Wojdyla, Karen P. Galen, James H. Wible, and Gary H. Brandenburger. Detection of Individual Microbubbles of Ultrasound Contrast Agents: Imaging of Free-Floating and Targeted Bubbles. *Investigative Radiology*, 39(3):187–195, 2004.
- [4] P Som, H L Atkins, D Bandoypadhyay, J S Fowler, A R Macgregor, K Matsui, Z H Oster, and D F Sacker. A Fluorinated Glucose Analog , 2-fluoro-2-deoxy-D-glucose (F-18): Nontoxic Tracer for RapidTumor Detection ofacute orchronic. *Journal of Nuclear Medicine*, 21:670–675, 1980.
- [5] A. Zhu, D. Lee, and H. Shim. Metabolic PET iamging in cancer detection and therapy response. *Semin. Oncol*, 38(1):55–69, 2012.
- [6] John Radford, Tim Illidge, Nicholas Counsell, Barry Hancock, Ruth Pettengell, Peter Johnson, Jennie Wimperis, Dominic Culligan, Bilyana Popova, Paul Smith, Andrew McMillan, Alison Brownell, Anton Kruger, Andrew Lister, Peter Hoskin, Michael O’Doherty, and Sally Barrington. Results of a Trial of PET-Directed Therapy for Early-Stage Hodgkin’s Lymphoma. *New England Journal of Medicine*, 372(17):1598–1607, 2015.
- [7] Sally F. Barrington, N. George Mikhaeel, Lale Kostakoglu, Michel Meignan, Martin Hutchings, Stefan P. Mueller, Lawrence H. Schwartz, Emanuele Zucca, Richard I. Fisher, Judith Trotman, Otto S. Hoekstra, Rodney J. Hicks, Michael J. O’Doherty, Roland Hustinx, Alberto Biggi, and Bruce D. Cheson. Role of imaging in the staging and response assessment of lymphoma: Consensus of the international conference on malignant lymphomas imaging working group. *Journal of Clinical Oncology*, 32(27):3048–3058, 2014.
- [8] David R. Baldwin and Matthew E.J. Callister. The British Thoracic Society guidelines on the investigation and management of pulmonary nodules. *Thorax*, 70(8):794–798, 2015.

- [9] J. Brush, K. Boyd, F. Chappell, F. Crawford, M. Dozier, E. Fenwick, J. Glanville, H. McIntosh, A. Renehan, D. Weller, and M. Dunlop. The value of FDG positron emission tomography/computerised tomography (PET/CT) in pre-operative staging of colorectal cancer: A systematic review and economic evaluation. *Health Technology Assessment*, 15(35), 2011.
- [10] Hisham Mehanna, Christopher C. McConkey, Martin Max Robinson, Andrew G J Hartley, Christopher Nutting, Ned Powell, Hoda Al Booz, Martin Max Robinson, Alison F Smith, M Sc, Peter Hall, Janet Dunn, Wai-Lup Wong, Christopher C. McConkey, Joy K. Ragnab, Martin Max Robinson, Andrew G J Hartley, Christopher Nutting, Ned Powell, Hoda Al Booz, Martin Max Robinson, Alison F Smith, Peter Hall, and Janet Dunn. PET-CT surveillance versus neck dissection in advanced head and neck cancer. *New England Journal of Medicine*, 220(9):449, 2016.
- [11] Fidelma L. Flanagan, Farrokh Dehdshti, Barry A. Siegel, Diana D. Trask, Sudhir R. Sundareshan, G. Alexander Patterson, and Joel D. Cooper. Staging of Esophageal Cancer with 18F-Fluorodeoxyglucose Positron Emission Tomography. *American Journal of Roentgenology*, (168):417–424, 1997.
- [12] L. Goense, P. S. N. van Rossum, J. B. Reitsma, M. G. E. H. Lam, G. J. Meijer, M. van Vulpen, J. P. Ruurda, and R. van Hillegersberg. Diagnostic Performance of 18F-FDG PET and PET/CT for the Detection of Recurrent Esophageal Cancer After Treatment with Curative Intent: A Systematic Review and Meta-Analysis. *Journal of Nuclear Medicine*, 56(7):995–1002, 2015.
- [13] B. E. Hillner, D. Liu, R. E. Coleman, A. F. Shields, I. F. Gareen, L. Hanna, S. H. Stine, and B. A. Siegel. The National Oncologic PET Registry (NOPR): Design and Analysis Plan. *Journal of Nuclear Medicine*, 48(11):1901–1908, 2007.
- [14] Akram Al-Ibraheem, Andreas Buck, Bernd Joachim Krause, Klemens Scheidhauer, and Markus Schwaiger. Clinical applications of FDG PET and PET/CT in head and neck cancer. *Journal of Oncology*, 2009, 2009.
- [15] Richard J. Wong. Current status of FDG-PET for head and neck cancer. *Journal of Surgical Oncology*, 97(8):649–652, 2008.
- [16] ICSCNM/RCR. Evidence-based indications for the use of PET-CT in the United Kingdom 2016 - The Royal College of Radiologists, Royal College of Physicians of London, Royal College of Physicians and Surgeons of Glasgow, Royal College of Physicians of Edinburgh, British. *Royal College of Physicians (RCP); Royal College of Radiologists*, 2016.
- [17] Daniel H. S. Silverman, Gary W Small, Carol Y Chang, Carolyn S Lu, Michelle A Kung De Aburto, Johannes Czernin, Stanley I Rapoport, Gene E Alexander, Mark B Schapiro, William J Jagust, John M Hoffman, Kathleen A Welsh-bohmer, Christopher M Clark, Mony J De Leon, Michael E Phelps, and Patient Page. Positron Emission Tomography in Evaluation of Dementia: Regional Brain Metabolism and Long-term Outcome Daniel. *JAMA.*, 286(17):2120–2127, 2001.
- [18] Lisa Mosconi, Rachel Mistur, Remigiusz Switalski, Wai Hon Tsui, Lidia Glodzik, Yi Li, Elizabeth Pirraglia, Susan De Santi, Barry Reisberg, Thomas Wisniewski, and

- Mony J. De Leon. FDG-PET changes in brain glucose metabolism from normal cognition to pathologically verified Alzheimer's disease. *European Journal of Nuclear Medicine and Molecular Imaging*, 36(5):811–822, 2009.
- [19] K. A. Johnson, S. Minoshima, N. I. Bohnen, K. J. Donohoe, N. L. Foster, P. Herscovitch, J. H. Karlawish, C. C. Rowe, M. C. Carrillo, D. M. Hartley, S. Hedrick, V. Pappas, and W. H. Thies. Appropriate Use Criteria for Amyloid PET: A Report of the Amyloid Imaging Task Force, the Society of Nuclear Medicine and Molecular Imaging, and the Alzheimer's Association. *Journal of Nuclear Medicine*, 54(3):476–490, 2013.
- [20] Christopher M Clark, Julie A Schneider, Barry J Bedell, Thomas G Beach, Warren B Bilker, Mark A Mintun, Michael J Pontecorvo, Franz Hefti, Alan P Carpenter, Matthew L Flitter, Michael J Krautkramer, Hank F Kung, R Edward Coleman, Carl H Sadowsky, Eric M Reiman, Simone P Zehntner, and Daniel M Skovronsky. Use of Florbetapir-PET for Imaging. *Journal of the American Medical Association*, 305(3):275–283, 2011.
- [21] Pekka Jokinen, Hans Helenius, Elina Rauhala, Anna Brück, Olli Eskola, and Juha O. Rinne. Simple ratio analysis of 18F-fluorodopa uptake in striatal subregions separates patients with early parkinson disease from healthy controls. *Journal of Nuclear Medicine*, 50(6):893–899, 2009.
- [22] A. Varrone and C. Halldin. New developments of dopaminergic imaging in Parkinson's disease. *Nuclear Medicine and Molecular Imaging*, 56(1):68–82, 2012.
- [23] Kentaro Hatano, Kiichi Ishiwata, and Philip Elsinga. PET Tracers for Imaging of the Dopaminergic System. *Current Medicinal Chemistry*, 13(18):2139–2153, 2006.
- [24] Y. F. Tai and P. Piccini. Applications of positron emission tomography (PET) in neurology. *Journal of Neurology, Neurosurgery and Psychiatry*, 75(5):669–676, 2004.
- [25] Peter Johnson, Massimo Federico, Amy Kirkwood, Alexander Fosså, Leanne Berkahn, Angelo Carella, Francesco D'Amore, Gunilla Enblad, Antonella Franceschetto, Michael Fulham, Stefano Luminari, Michael O'Doherty, Pip Patrick, Thomas Roberts, Gamal Sidra, Lindsey Stevens, Paul Smith, Judith Trotman, Zaid Viney, John Radford, and Sally Barrington. Adapted Treatment Guided by Interim PET-CT Scan in Advanced Hodgkin's Lymphoma. *New England Journal of Medicine*, 374(25):2419–2429, 2016.
- [26] Michael K. Gould, Courtney C Maclean, Ware G Kuschner, Chara E Rydzak, and Douglas K Owens. Accuracy of Positron Emission Tomography for Diagnosis of Pulmonary Nodules. *JAMA : the journal of the American Medical Association*, 285(7), 2001.
- [27] Paul Cronin, Ben A. Dwamena, Aine Marie Kelly, and Ruth C. Carlos. Solitary Pulmonary Nodules: Meta-analytic Comparison of Cross-sectional Imaging Modalities for Diagnosis of Malignancy. *Radiology*, 246(3):772–782, 2008.
- [28] JHF Rudd, Ea Warburton, TD Fryer, Ha Jones, JC Clark, N Antoun, P Johnström, Ap Davenport, PJ Kirkpatrick, BN Arch, JD Pickard, and PL Weissberg. Imaging Atherosclerotic Plaque Inflammation With [18 F]-Fluorodeoxyglucose Positron Emission Tomography. 2002.

- [29] Jason M Tarkin, Francis R Joshi, Nicholas R Evans, Mohammed M Chowdhury, Nichola L Figg, Aarti V Shah, Lakshi T Starks, Abel Martin-Garrido, Roido Manavaki, Emma Yu, Rhoda E Kuc, Luigi Grassi, Roman Kreuzhuber, Myrto A Kostadima, Mattia Frontini, Peter J Kirkpatrick, Patrick A Coughlin, Deepa Gopalan, Tim D Fryer, John R Buscombe, Ashley M Groves, Willem H Ouwehand, Martin R Bennett, Elizabeth A Warburton, Anthony P Davenport, and James H F Rudd. Detection of Atherosclerotic Inflammation by ^{68}Ga -DOTATATE PET Compared to ^{18}F FDG PET Imaging. *Journal of the American College of Cardiology*, 69(14):1774–1791, apr 2017.
- [30] P Bustany, M Chatel, J M Derlon, F Darcel, P Sgouropoulos, F Soussaline, and A Syrota. Brain tumor protein synthesis and histological grades: a study by positron emission tomography (PET) with C^{11} -L-Methionine. *Journal of neuro-oncology*, 3(4):397–404, 1986.
- [31] Gary J. R. Cook, M. N. Maisey, and I. Fogelman. Normal variants, artefacts and interpretative pitfalls in PET imaging with 18-fluoro-2- deoxyglucose and carbon-11 methionine. *European Journal of Nuclear Medicine*, 26(10):1363–1378, 1999.
- [32] WL Wong, B Sanghera, and B Cronin. Recommendations for cross-sectional imaging in cancer management. Technical report, The Royal College of Radiologists, London, 2014.
- [33] Ayesha S Bryant, Sara J Pereira, Daniel L Miller, and Robert James Cerfolio. Satellite Pulmonary Nodule in the Same Lobe (T4N0) Should Not Be Staged as IIIB Non – Small Cell Lung Cancer. *Annals of Thoracic Surgery*, 82:1808–1814, 2006.
- [34] Kjell Erlandsson, Irène Buvat, P. Hendrik Pretorius, Benjamin A. Thomas, and Brian F. Hutton. A review of partial volume correction techniques for emission tomography and their applications in neurology, cardiology and oncology. *Physics in Medicine and Biology*, 57(21), 2012.
- [35] C C Meltzer, P E Kinahan, P J Greer, T E Nichols, C Comtat, M N Cantwell, M P Lin, and J C Price. Comparative evaluation of MR-based partial-volume correction schemes for PET. *Journal of nuclear medicine : official publication, Society of Nuclear Medicine*, 40(12):2053–65, 1999.
- [36] Peymaneh Abadi, Allan Johansen, Christian Godballe, and Oke Gerke. F-FDG PET / CT to differentiate malignant necrotic lymph node from benign cystic lesions in the neck. *Annals of Nuclear Medicine*, 31(2):101–108, 2017.
- [37] Stephan K Haerle, Klaus Strobel, Nader Ahmad, Alex Soltermann, Daniel T Schmid, and Sandro J Stoeckli. Contrast-enhanced ^{18}F -FDG-PET / CT for the assessment of necrotic lymph node metastases. *Head & neck*, 33(3):324–329, 2011.
- [38] Charles Marcus, Pat W Whitworth, Rathan M Subramaniam, and Subramaniam Rm. PET/CT in the Management of Thyroid Cancers. *American Journal of Roentgenology*, 202(6):1316–1329, 2014.

- [39] Gaspar Delso, S. Furst, Björn Jakoby, Ralf Ladebeck, Carl Ganter, Stephan G. Nekolla, Markus Schwaiger, Sibylle I. Ziegler, Sebastian Fürst, Björn Jakoby, Ralf Ladebeck, Carl Ganter, Stephan G. Nekolla, Markus Schwaiger, and Sibylle I. Ziegler. Performance Measurements of the Siemens mMR Integrated Whole-Body PET/MR Scanner. *Journal of Nuclear Medicine*, 52(12):1914–1922, dec 2011.
- [40] Anna M Karlberg, Oddbjørn Sæther, Live Eikenes, and Pål Erik Goa. Quantitative comparison of PET performance-Siemens Biograph mCT and mMR. *EJNMMI physics*, 3(1):5, dec 2016.
- [41] Ronald Boellaard, Ivo Rausch, Thomas Beyer, Gaspar Delso, Maqsood Yaqub, Harald H. Quick, and Bernhard Sattler. Quality control for quantitative multicenter whole-body PET/MR studies: A NEMA image quality phantom study with three current PET/MR systems. *Medical physics*, 42(10):5961–5969, 2015.
- [42] A. Piccardo, F. Paparo, R. Picazzo, M. Naseri, P. Ricci, A. Marziano, L. Bacigalupo, E. Biscaldi, G. A. Rollandi, F. Grillo-Ruggieri, and M. Farsad. Value of fused 18F-Choline-PET/MRI to evaluate prostate cancer relapse in patients showing biochemical recurrence after EBRT: Preliminary results. *BioMed Research International*, 2014, 2014.
- [43] L. M. Sawicki, J. Grueneisen, C. Buchbender, B. M. Schaarschmidt, B. Gomez, V. Ruhlmann, A. Wetter, L. Umutlu, G. Antoch, and P. Heusch. Comparative Performance of 18F-FDG PET/MRI and 18F-FDG PET/CT in Detection and Characterization of Pulmonary Lesions in 121 Oncologic Patients. *Journal of Nuclear Medicine*, 57(4):582–586, 2016.
- [44] Hersh Chandarana, Laura Heacock, Rajan Rakheja, Linda R. DeMello, John Bonavita, Tobias K. Block, Christian Geppert, James S. Babb, and Kent P. Friedman. Pulmonary Nodules in Patients with Primary Malignancy: Comparison of Hybrid PET/MR and PET/CT Imaging. *Radiology*, 268(3):874–881, 2013.
- [45] Claudio Spick, Ken Herrmann, and Johannes Czernin. 18F-FDG PET/CT and PET/MRI Perform Equally Well in Cancer: Evidence from Studies on More Than 2,300 Patients. *Journal of Nuclear Medicine*, 57(3):420–430, 2016.
- [46] Habib Zaidi and Alberto Del Guerra. An outlook on future design of hybrid PET/MRI systems. *Medical Physics*, 38(10):5667–5689, 2011.
- [47] Bruce E. Hammer, Nelson L. Christensen, and Brain G. Heil. Use of a magnetic field to increase the spatial resolution of positron emission tomography. *Medical Physics*, 21(12), 1994.
- [48] Hao Peng and Craig S. Levin. Study of PET intrinsic spatial resolution and contrast recovery improvement for PET/MRI systems. *Physics in Medicine and Biology*, 57(9), 2012.
- [49] N. Jon Shah, Hans Herzog, Christoph Weirich, Lutz Tellmann, Joachim Kaffanke, Liliana Caldeira, Elena Rota Kops, Syed M. Qaim, Heinz H. Coenen, and Hidehiro Iida. Effects of magnetic fields of up to 9.4 T on resolution and contrast of PET images as measured with an MR-BrainPET. *PLoS ONE*, 9(4), 2014.

- [50] S. Y. Chun, T. G. Reese, J. Ouyang, B. Guerin, C. Catana, X. Zhu, N. M. Alpert, and G. El Fakhri. MRI-Based Nonrigid Motion Correction in Simultaneous PET/MRI. *Journal of Nuclear Medicine*, 53(8):1284–1291, 2012.
- [51] Richard Manber, Kris Thielemans, Brian F. Hutton, Anna Barnes, Sébastien Ourselin, Simon Arridge, Celia O’Meara, Simon Wan, and David Atkinson. Practical PET respiratory motion correction in clinical PET/MR. *Journal of Nuclear Medicine*, 56(6):890–896, 2015.
- [52] Ashley Gillman, Jye Smith, Paul Thomas, Stephen Rose, and Nicholas Dowson. PET motion correction in context of integrated PET/MR: Current techniques, limitations, and future projections: Current. *Medical Physics*, 44(12):e430–e445, 2017.
- [53] Christian F. Baumgartner, Christoph Kolbitsch, Daniel R. Balfour, Paul K. Marsden, Jamie R. McClelland, Daniel Rueckert, and Andrew P. King. High-resolution dynamic MR imaging of the thorax for respiratory motion correction of PET using groupwise manifold alignment. *Medical Image Analysis*, 18(7):939–952, 2014.
- [54] Moritz Blume, Axel Martinez-Möller, Andreas Keil, Nassir Navab, and Magdalena Rafecas. Joint reconstruction of image and motion in gated positron emission tomography. *IEEE Transactions on Medical Imaging*, 29(11):1892–1906, 2010.
- [55] Nikolaos Dikaio, David Izquierdo-Garcia, Martin J Graves, Venkatesh Mani, Zahi A Fayad, and Tim D Fryer. MRI-based motion correction of thoracic PET: initial comparison of acquisition protocols and correction strategies suitable for simultaneous PET/MRI systems. *Eur Radiol*, 22(2):439–446, 2012.
- [56] Hans Herzog, Lutz Tellmann, Isabelle Stangier, Elena Rota Kops, Kay Bente, Christian Boy, Uwe Pietrzyk, Rene Hurlemann, and Roger Fulton. Motion artifact reduction on parametric PET images of neuroreceptor binding. *Journal of Nuclear Medicine*, 46(6):1059–1065, 2005.
- [57] Yasuhiko Ikari, Tomoyuki Nishio, Yoko Makishi, Yukari Miya, Kengo Ito, Robert A. Koeppe, and Michio Senda. Head motion evaluation and correction for PET scans with 18F-FDG in the Japanese Alzheimer’s disease neuroimaging initiative (J-ADNI) multi-center study. *Annals of Nuclear Medicine*, 26(7):535–544, aug 2012.
- [58] Yuankai Zhu and Xiaohua Zhu. MRI-Driven PET Image Optimization for Neurological Applications. *Frontiers in Neuroscience*, 13(July):1–15, 2019.
- [59] Andrew B. Rosenkrantz, Kent Friedman, Hersh Chandarana, Amy Melsaether, Linda Moy, Yu Shin Ding, Komal Jhaveri, Luis Beltran, and Rajan Jain. Current status of hybrid PET/MRI in oncologic imaging. *American Journal of Roentgenology*, 206(1):162–172, 2016.
- [60] Xin Y. Zhang, Zhen L. Yang, Guang M. Lu, Gui F. Yang, and Long J. Zhang. PET/MR Imaging: New Frontier in Alzheimer’s Disease and Other Dementias. *Frontiers in Molecular Neuroscience*, 10(November):1–12, 2017.

- [61] Marcelo Queiroz, Felipe De Galiza BarBosa, Carlos alBerTo BuchPiGuel, and GuiDo Giovanni Cerri. Positron emission tomography/magnetic resonance imaging (PET/MRI): An update and initial experience at HC-FMUSP. *Rev. Assoc. Med. Bras.*, 64(1):71–84, 2018.
- [62] X Ouyang, W H Wong, Johnson V E, X Hu, and C T Chen. Incorporation of Correlated Structural Images in PET image reconstruction. *IEEE Trans Med Imaging*, 13(4):222–225, 1994.
- [63] Andrew J. Reader, Peter J. Julyan, Heather Williams, David L. Hastings, and Jamal Zweit. EM Algorithm System Modeling by Image-Space Techniques for PET Reconstruction. *IEEE Transactions on Nuclear Science*, 50(5), 2003.
- [64] Georg Schramm, Martin Holler, Ahmadreza Rezaei, Kathleen Vunckx, Florian Knoll, Kristian Bredies, Fernando Boada, and Johan Nuyts. Evaluation of Parallel Level Sets and Bowsher’s Method as Segmentation-Free Anatomical Priors for Time-of-Flight PET Reconstruction. *IEEE Transactions on Medical Imaging*, 0062(c):1–14, 2017.
- [65] Kathleen Vunckx, Ameya Atre, Kristof Baete, Anthonin Reilhac, Christophe M. Deroose, Koen Van Laere, and Johan Nuyts. Evaluation of three MRI-based anatomical priors for quantitative PET brain imaging. *IEEE Transactions on Medical Imaging*, 2012.
- [66] Hesheng Wang and Baowei Fei. An MR image-guided, voxel-based partial volume correction method for PET images. *Medical Physics*, 39(1):179–194, 2012.
- [67] David Izquierdo-Garcia, John R. Davies, Martin J. Graves, James H.F. Rudd, Jonathan H. Gillard, Peter L. Weissberg, Tim D. Fryer, and Elizabeth A. Warburton. Comparison of methods for magnetic resonance-guided [18-f] fluorodeoxyglucose positron emission tomography in human carotid arteries: Reproducibility, partial volume correction, and correlation between methods. *Stroke*, 40(1):86–93, 2009.
- [68] Jieqing Jiao, Pawel Markiewicz, Ninon Burgos, David Atkinson, Brian Hutton, Simon Arridge, and Sebastien Ourselin. Detail-preserving PET reconstruction with sparse image representation and anatomical priors. *Information Processing in Medical Imaging*, pages 540–551, 2015.
- [69] James Bland, Martin A. Belzunce, Sam Ellis, Colm J. McGinnity, Alexander Hammers, and Andrew J. Reader. Spatially Compact MR-Guided Kernel EM for PET Image Reconstruction. *IEEE Transactions on Radiation and Plasma Medical Sciences*, 2(5):470–482, 2018.
- [70] Martin A. Belzunce, Abolfazl Mehranian, and Andrew J. Reader. Enhancement of Partial Volume Correction in MR-guided PET Image Reconstruction by using MRI Voxel Sizes. *IEEE Transactions on Radiation and Plasma Medical Sciences*, C:1–1, 2018.
- [71] Bing Bai, Quanzheng Li, and Richard M. Leahy. Magnetic resonance-guided positron emission tomography image reconstruction. *Seminars in Nuclear Medicine*, 2013.

- [72] Claude Nahmias, E. Stephen Garnett, Gunter Firnau, and Anthony Lang. Striatal dopamine distribution in Parkinsonian patients during life. *Journal of the Neurological Sciences*, 1985.
- [73] I. Cheng Chang, Kun Han Lue, Hung Jen Hsieh, Shu Hsin Liu, and Chih Hao K. Kao. Automated striatal uptake analysis of 18F-FDOPA PET images applied to Parkinson's disease patients. *Annals of Nuclear Medicine*, 2011.
- [74] W. E. Klunk, J. C. Price, C. A. Mathis, N. D. Tsopelas, B. J. Lopresti, S. K. Ziolko, W. Bi, J. A. Hoge, A. D. Cohen, M. D. Ikonomic, J. A. Saxton, B. E. Snitz, D. A. Pollen, M. Moonis, C. F. Lippa, J. M. Swearer, K. A. Johnson, D. M. Rentz, A. J. Fischman, H. J. Aizenstein, and S. T. DeKosky. Amyloid Deposition Begins in the Striatum of Presenilin-1 Mutation Carriers from Two Unrelated Pedigrees. *Journal of Neuroscience*, 2007.
- [75] L. Mosconi and P. F. McHugh. FDG- and amyloid-PET in Alzheimer's disease: Is the whole greater than the sum of the parts?, 2011.
- [76] Kristof Baete, Johan Nuyts, Wim Van Paesschen, Paul Suetens, and Patrick Dupont. Anatomical-based FDG-PET reconstruction for the detection of hypo-metabolic regions in epilepsy. *IEEE Transactions on Medical Imaging*, 23(4):510–519, 2004.
- [77] Kristof Baete, Johan Nuyts, Koen Van Laere, Wim Van Paesschen, Sarah Ceysens, Liesbet De Ceuninck, Olivier Gheysens, Annemarie Kelles, Jimmy Van Den Eynden, Paul Suetens, and Patrick Dupont. Evaluation of anatomy based reconstruction for partial volume correction in brain FDG-PET. *NeuroImage*, 23(1):305–317, 2004.
- [78] Philippe P. Bruyant, H. C. Gifford, Gene Gindi, and Michael A. King. Investigation of observer-performance in MAP-EM reconstruction with anatomical priors and scatter correction for lesion detection in 67Ga images. In *IEEE Nuclear Science Symposium & Medical Imaging Conference*, 2003.
- [79] Andre Lehovitch, Howard S. Gifford, Peter B. Schneider, Shayne Squires, Robert Licho, Michael A. King, Philippe P. Bruyant, and Gene Gindi. Impact on Reader Performance for Lesion-Detection/Localization Tasks of Anatomical Priors in SPECT Reconstruction. *IEEE Transactions on Medical Imaging*, 28(9):1459–1467, 2009.
- [80] S. Kulkarni, P. Khurd, I. Hsiao, L. Zhou, and G. Gindi. A Channelized Hotelling Observer Study of Lesion Detection in SPECT MAP Reconstruction Using Anatomical Priors. *Physics in Medicine and Biology*, 52(12):3601–3617, 2007.
- [81] Will Hutchcroft, Guobao Wang, Kevin T. Chen, Ciprian Catana, and Jinyi Qi. Anatomically-aided PET reconstruction using the kernel method. *Physics in Medicine and Biology*, 61(18):6668–6683, 2016.
- [82] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep Learning. *Nature*, 521, 2015.
- [83] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. ImageNet Classification with Deep Convolutional Neural Networks. In *NIPS 2012: Neural Information Processing Systems, Lake Tahoe, Nevada.*, 2012.

- [84] Geoffrey Hinton, Li Deng, Dong Yu, George Dahl, Abdel Rahman Mohamed, Navdeep Jaitly, Andrew Senior, Vincent Vanhoucke, Patrick Nguyen, Tara Sainath, and Brian Kingsbury. Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *IEEE Signal Processing Magazine*, 29(6):82–97, 2012.
- [85] Sepp Hochreiter and Jurgen Schmidhuber. Long short-term memory. *Neural Computation*, 9(8):1735–1780, 1997.
- [86] Ilya Sutskever, James Martens, and Geoffrey Hinton. Generating text with recurrent neural networks. In *ICML'11 Proceedings of the 28th International Conference on International Conference on Machine Learning*, pages 1017–1024, 2011.
- [87] David Silver, Aja Huang, Chris J Maddison, Arthur Guez, Laurent Sifre, George Van Den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, Sander Dieleman, Dominik Grewe, John Nham, Nal Kalchbrenner, Ilya Sutskever, Timothy Lillicrap, Madeleine Leach, and Koray Kavukcuoglu. Mastering the game of Go with deep neural networks and tree search. *Nature*, 529(7585):484–489, 2016.
- [88] Yann LeCun, Leon Bottou, Yoshua Bengio, and Parick Haffner. Gradient-Based Learning Applied to Document Recognition. *proc. OF THE IEEE*, 1998.
- [89] Karen Simonyan and Andrew Zisserman. Very Deep Convolutional Networks for Large-Scale Image Recognition. pages 1–14, 2014.
- [90] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2016-Decem:770–778, 2016.
- [91] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 07-12-June:1–9, 2015.
- [92] Yisheng Lv, Yanjie Duan, Wenwen Kang, Zhengxi Li, and Fei Yue Wang. Traffic Flow Prediction with Big Data: A Deep Learning Approach. *IEEE Transactions on Intelligent Transportation Systems*, 16(2):865–873, 2015.
- [93] Liang Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L. Yuille. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(4):834–848, 2018.
- [94] Yushi Chen, Zhouhan Lin, Xing Zhao, Gang Wang, and Yanfeng Gu. Deep learning-based classification of hyperspectral data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(6):2094–2107, 2014.
- [95] Feng Jia, Yaguo Lei, Jing Lin, Xin Zhou, and Na Lu. Deep neural networks: A promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data. *Mechanical Systems and Signal Processing*, 72-73:303–315, 2016.

- [96] Junyuan Xie, Linli Xu, and Enhong Chen. Image denoising and inpainting with deep neural networks. *Advances in Neural Information Processing Systems*, 1:341–349, 2012.
- [97] Chunwei Tian, Yong Xu, Lunke Fei, and Ke Yan. Deep Learning for Image Denoising: A Survey. In *International Conference on Genetic and Evolutionary Computing*, 2018.
- [98] Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang. Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising. *IEEE Transactions on Image Processing*, 26(7):3142–3155, 2017.
- [99] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Image Super-Resolution Using Deep Convolutional Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(2):295–307, 2016.
- [100] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. Accurate Image Super-Resolution Using Very Deep Convolutional Networks. In *The IEEE Conference on Computer Vision and Pattern Recognition*, pages 1646–1654, 2016.
- [101] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew P Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, and Others. Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network. In *The IEEE Conference on Computer Vision and Pattern Recognition*, volume 2, page 4, 2017.
- [102] Leon A Gatys, Alexander S. Ecker, and Matthias Bethge. Image style transfer using convolutional neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2414–2423, 2016.
- [103] Li Xu, Jimmy S.J. Ren, Ce Liu, and Jiaya Jia. Deep convolutional neural network for image deconvolution. *Advances in Neural Information Processing Systems*, 2(January):1790–1798, 2014.
- [104] Konstantinos Kamnitsas, Christian Ledig, Virginia F.J. Newcombe, Joanna P. Simpson, Andrew D. Kane, David K. Menon, Daniel Rueckert, and Ben Glocker. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. *Medical Image Analysis*, 36:61–78, 2017.
- [105] Mohammad Havaei, Axel Davy, David Warde-Farley, Antoine Biard, Aaron Courville, Yoshua Bengio, Chris Pal, Pierre Marc Jodoin, and Hugo Larochelle. Brain tumor segmentation with Deep Neural Networks. *Medical Image Analysis*, 35:18–31, 2017.
- [106] Sergio Pereira, Adriano Pinto, Victor Alves, and Carlos A. Silva. Brain Tumor Segmentation Using Convolutional Neural Networks in MRI Images. *IEEE Transactions on Medical Imaging*, 35(5):1240–1251, 2016.
- [107] Jose Dolz, Christian Desrosiers, and Ismail Ben Ayed. 3D fully convolutional networks for subcortical segmentation in MRI: A large-scale study. *NeuroImage*, 170(April 2017):456–470, 2018.

- [108] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Medical image computing and computer-assisted intervention - MICCAI 2015*, pages 234–241, 2015.
- [109] Hongyoon Choi. Deep Learning in Nuclear Medicine and Molecular Imaging: Current Perspectives and Future Directions. *Nuclear Medicine and Molecular Imaging*, 52(2):109–118, 2018.
- [110] Ge Wang, Jong Chu Ye, Senior Member, Klaus Mueller, Senior Member, and Jeffrey A Fessler. Image Reconstruction Is a New Frontier of Machine Learning. *IEEE Transactions on Medical Imaging*, 37(6):1289–1296, 2018.
- [111] Simon Arridge, Peter Maass, Ozan Öktem, and Carola-Bibiane Schönlieb. Solving inverse problems using data-driven models. *Acta Numerica*, 28:1–174, 2019.
- [112] Bao Yang, Leslie Ying, and Jing Tang. Artificial Neural Network Enhanced Bayesian PET Image Reconstruction. *IEEE Transactions on Medical Imaging*, 37(6):1297–1309, 2018.
- [113] Sydney Kaplan and Yang Ming Zhu. Full-Dose PET Image Estimation from Low-Dose PET Image Using Deep Learning: a Pilot Study. *Journal of Digital Imaging*, 3, 2018.
- [114] Yan Wang, Biting Yu, Lei Wang, Chen Zu, David S. Lalush, Weili Lin, Xi Wu, Jiliu Zhou, Dinggang Shen, and Luping Zhou. 3D conditional generative adversarial networks for high-quality PET image estimation at low dose. *NeuroImage*, 174(March):550–562, 2018.
- [115] K. Kim, D. Wu, K. Gong, J. Dutta, J.H. Kim, Y.D. Son, H.K. Kim, G. El Fakhri, and Q. Li. Penalized PET reconstruction using deep learning prior and local linear fitting. *IEEE Transactions on Medical Imaging*, 2018.
- [116] Junshen Xu, Enhao Gong, John Pauly, and Greg Zaharchuk. 200x Low-dose PET Reconstruction using Deep Learning. *arXiv:1712.04119*, 2017.
- [117] Guobao Wang and Jinyi Qi. PET image reconstruction using kernel method. *IEEE Transactions on Medical Imaging*, 34(1):61–71, 2015.
- [118] Jaakko Lehtinen, Jacob Munkberg, Jon Hasselgren, Samuli Laine, Tero Karras, Miika Aittala, and Timo Aila. Noise2Noise: Learning Image Restoration without Clean Data. (3), 2018.
- [119] Dufan Wu, Kuang Gong, Kyungsang Kim, and Quanzheng Li. Consensus Neural Network for Medical Imaging Denoising with Only Noisy Training Samples. pages 1–9, 2019.
- [120] Z. H. Cho, J. K. Chan, L. Ericksson, M. Singh, S. Graham, N. S. MacDonald, and Y. Yano. Positron ranges obtained from biomedically important positron emitting radionuclides. *Journal of Nuclear Medicine*, 16(12):1174–1176, 1975.

- [121] J. Cal-González, J. L. Herraiz, S. España, P. M.G. Corzo, J. J. Vaquero, M. Desco, and J. M. Udías. Positron range estimations with PeneloPET. *Physics in Medicine and Biology*, 58(15):5127–5152, 2013.
- [122] Frederic H Fahey. Data acquisition in PET imaging. *Journal of nuclear medicine technology*, 30(2):39–49, 2002.
- [123] L A Shepp and B F Logan. The Fourier reconstruction of a head section. *IEEE Transactions on Nuclear Science*, NS-21(3):21–43, 1974.
- [124] Joyce J van Sluis, Johan de Jong, Jenny Schaar, Walter Noordzij, Paul van Snick, Rudi Dierckx, Ronald Borra, Antoon Willemsen, and Ronald Boellaard. Performance characteristics of the digital Biograph Vision PET/CT system. *Journal of Nuclear Medicine*, page jnumed.118.215418, 2019.
- [125] S Vandenberghe, E Mikhaylova, E D Hoe, P Mollet, and J S Karp. Recent developments in time-of-flight PET. *EJNMMI Physics*, 3(3):1–30, 2016.
- [126] J. W. Cates and C. S. Levin. Evaluation of a TOF-PET detector design that achieves leq 100 ps coincidence time resolution. *2017 IEEE Nuclear Science Symposium and Medical Imaging Conference, NSS/MIC 2017 - Conference Proceedings*, 2018.
- [127] Simon R. Cherry, J. A. Sorenson, and M. E. Phelps. *Chapter 18 - Positron emission tomography*. 2012.
- [128] Jacobo Cal-González, Mailyn Pérez-Liva, Joaquín L. Herraiz, Juan J. Vaquero, Manuel Desco, and José M. Udías. Tissue-Dependent and Spatially-Variant Positron Range Correction in 3D PET. *IEEE Transactions on Medical Imaging*, 34(11):2394–2403, 2015.
- [129] Taiga Yamaya, Takayuki Mitsuhashi, Takahiro Matsumoto, Naoko Inadama, Fumihiko Nishikido, Eiji Yoshida, Hideo Murayama, Hideyuki Kawai, Mikio Suga, and Mitsuo Watanabe. A SiPM-based isotropic-3D PET detector X’tal cube with a three-dimensional array of 1 mm 3 crystals. *Physics in Medicine and Biology*, 56(21):6793–6807, 2011.
- [130] A Braem, M Chamizo Llatas, E Chesi, J G Correia, F Garibaldi, C Joram, S Mathot, E Nappi, M Ribeiro da Silva, F Schoenahl, J Séguinot, P Weilhammer, and H Zaidi. Feasibility of a novel design of high resolution parallax-free Compton enhanced PET scanner dedicated to brain research. *Physics in Medicine and Biology*, 49(12):2547–2562, jun 2004.
- [131] Martin A. Belzunce and Andrew J. Reader. Assessment of the impact of modeling axial compression on PET image reconstruction. *Medical Physics*, 44(10):5172–5186, 2017.
- [132] Simon R. Cherry, Terry Jones, Joel S. Karp, Jinyi Qi, William W. Moses, and Ramsey D. Badawi. Total-Body PET: Maximizing Sensitivity to Create New Opportunities for Clinical Research and Patient Care. *Journal of Nuclear Medicine*, 59(1):3–12, 2018.

- [133] Taiga Yamaya, Eiji Yoshida, Hideaki Tashima, Atsushi Tsuji, Kotaro Nagatsu, Mitsutaka Yamaguchi, Naoki Kawachi, Yusuke Okumura, Mikio Suga, and Katia Parodi. Whole gamma imaging (WGI) concept: simulation study of triple-gamma imaging. *Journal of Nuclear Medicine*, 58(supplement 1):152–152, may 2017.
- [134] Samuel Matej and Robert M Lewitt. Practical Considerations for 3-D construction Using Spherically Symmetric Volume Elements. *IEEE Trans Med Imaging*, 15(1), 1996.
- [135] Robert M Lewitt and Samuel Matej. Overview of methods for image reconstruction from projections in emission computed tomography. *Proceedings of the IEEE*, 91(10):1588–1611, 2003.
- [136] R E Carson. A maximum likelihood method for region-of-interest evaluation in emission tomography. *Journal of computer assisted tomography*, 10(4):654–63, 1986.
- [137] Andrew J Reader and Jeroen Verhaeghe. 4D image reconstruction for emission tomography. *Physics in Medicine and Biology*, 59:R371–R418, 2014.
- [138] T. F. Budinger. PET instrumentation: What are the limits? *Seminars in Nuclear Medicine*, 28(3):247–267, 1998.
- [139] Johann Radon. On the Determination of Functions From Their Integral Values Along Certain Manifolds. Technical Report 4, 1986.
- [140] Robert L. Siddon. Fast calculation of the exact radiological path for a three-dimensional CT array. *Medical Physics*, 12(2):252–255, mar 1985.
- [141] E.J. Hoffman, T.M. Guerrero, G. Germano, W.M. Digby, and M. Dahlbom. PET system calibrations and corrections for quantitative and spatially accurate images. *IEEE Transactions on Nuclear Science*, 36(1):1108–1112, 1989.
- [142] M. A. Belzunce and A. J. Reader. Time-invariant component-based normalization for a simultaneous PET-MR scanner. *Physics in Medicine and Biology*, 61(9):3554–3571, 2016.
- [143] Vincent Keereman, Roel Van Holen, Pieter Mollet, and Stefaan Vandenberghe. The effect of errors in segmented attenuation maps on PET quantification. *Medical Physics*, 38(11):6010–6019, oct 2011.
- [144] Axel Martinez-Möller, Michael Souvatzoglou, Gaspar Delso, Ralph A Bundschuh, Christophe Chefd’hotel, Sibylle I Ziegler, Nassir Navab, Markus Schwaiger, and Stephan G Nekolla. Tissue classification as a potential approach for attenuation correction in whole-body PET/MRI: evaluation with PET/CT data. *Journal of nuclear medicine* :, 50(4):520–6, apr 2009.
- [145] Vincent Keereman, Yves Fierens, Tom Broux, Yves De Deene, Max Lonnew, and Stefaan Vandenberghe. MRI-based attenuation correction for PET/MRI using ultra-short echo time sequences. *Journal of nuclear medicine : official publication, Society of Nuclear Medicine*, 51(5):812–8, may 2010.

- [146] Florian Wiesinger, Laura I. Sacolick, Anne Menini, Sandeep S. Kaushik, Sangtae Ahn, Patrick Veit-Haibach, Gaspar Delso, and Dattesh D. Shanbhag. Zero TEMR bone imaging in the head. *Magnetic Resonance in Medicine*, 75(1):107–114, jan 2016.
- [147] Ninon Burgos, M. Jorge Cardoso, Kris Thielemans, Marc Modat, Stefano Pedemonte, John Dickson, Anna Barnes, Rebekah Ahmed, Colin J. Mahoney, Jonathan M. Schott, John S. Duncan, David Atkinson, Simon R. Arridge, Brian F. Hutton, and Sebastien Ourselin. Attenuation Correction Synthesis for Hybrid PET-MR Scanners: Application to Brain Studies. *IEEE Transactions on Medical Imaging*, 33(12):2332–2341, dec 2014.
- [148] Thomas Hofmann, Bernhard Schölkopf, and Alexander J. Smola. Kernel methods in machine learning. *Annals of Statistics*, 36(3):1171–1220, 2008.
- [149] Elena Rota Kops and Hans Herzog. Alternative methods for attenuation correction for PET images in MR-PET scanners. In *2007 IEEE Nuclear Science Symposium Conference Record*, pages 4327–4330. IEEE, 2007.
- [150] Ian B. Malone, Richard E. Ansorge, Guy B. Williams, Peter J. Nestor, T. Adrian Carpenter, and Tim D. Fryer. Attenuation correction methods suitable for brain imaging with a PET/MRI scanner: A comparison of tissue atlas and template attenuation map approaches. *Journal of Nuclear Medicine*, 52(7):1142–1149, 2011.
- [151] Kuang Gong and Quanzheng Li. Attenuation Correction for Brain PET imaging using Deep Neural Network based on Dixon and ZTE MR images arXiv : 1712 . 06203v1 [physics . med-ph] 17 Dec 2017. pages 1–14, 2017.
- [152] Fang Liu, Hyungseok Jang, Richard Kijowski, Tyler Bradshaw, and Alan B Mcmillan. Deep learning Mr imaging-based attenuation correction for PET/MR imaging. *radiology.rsna.org n Radiology*, 286(2), 2018.
- [153] Dong Nie, Xiaohuan Cao, Yaozong Gao, Li Wang, and Dinggang Shen. Estimating CT Image from MRI Data Using 3D Fully Convolutional Networks. In *International Workshop on Deep Learning in Medical Image Analysis*, pages 170–178. Springer, Cham, oct 2016.
- [154] Y. Censor, D. E. Gustafson, A. Lent, and H. Tuy. A New Approach to the Emission Computerized Tomography Problem: Simultaneous Calculation of Attenuation and Activity Coefficients. *IEEE Transactions on Nuclear Science*, 26(2):2775–2779, apr 1979.
- [155] J. Nuyts, P. Dupont, S. Stroobants, R. Benninck, L. Mortelmans, and P. Suetens. Simultaneous maximum a posteriori reconstruction of attenuation and activity distributions from emission sinograms. *IEEE Transactions on Medical Imaging*, 18(5):393–403, may 1999.
- [156] Alaleh Rashidnasab, Alexandre Bousse, Beverley F. Holman, Brian F. Hutton, and Kris Thielemans. Joint reconstruction of activity and attenuation in dynamic PET. *2016 IEEE Nuclear Science Symposium, Medical Imaging Conference and Room-Temperature Semiconductor Detector Workshop, NSS/MIC/RTSD 2016*, 2017-Janua(1):1–3, 2017.

- [157] André Salomon, Andreas Goedicke, Bernd Schweizer, Til Aach, and Volkmar Schulz. Simultaneous Reconstruction of Activity and Attenuation for PET/MR. *IEEE Transactions on Medical Imaging*, 30(3):804–813, mar 2011.
- [158] Michel Defrise, Ahmadreza Rezaei, and Johan Nuyts. Time-of-flight PET data determine the attenuation sinogram up to a constant. *Physics in Medicine and Biology*, 57(4):885–899, 2012.
- [159] Abolfazl Mehranian and Habib Zaidi. Joint Estimation of Activity and Attenuation in Whole-Body TOF PET/MRI Using Constrained Gaussian Mixture Models. *IEEE Transactions on Medical Imaging*, 34(9):1808–1821, sep 2015.
- [160] Vladimir Y Panin, Frank Kehren, H Rothfuss, Christian Michel, and Michael Casey. PET reconstruction with system matrix derived from point source measurements. *IEEE NSS and MIC Conference Record*, 25(7):2483–2487, 2004.
- [161] Dan J. Kadrmas, Michael E. Casey, Noel F. Black, James J. Hamill, Vladimir Y. Panin, and Maurizio Conti. Experimental comparison of lesion detectability for four fully-3D PET reconstruction schemes. *IEEE Transactions on Medical Imaging*, 28(4):523–534, 2009.
- [162] Arman Rahmim, Jinyi Qi, and Vesna Sossi. Resolution modeling in PET imaging: Theory, practice, benefits, and pitfalls. *Medical Physics*, 40(6):1–15, 2013.
- [163] Johan Nuyts. Unconstrained image reconstruction with resolution modelling does not have a unique solution. *EJNMMI Physics*, 1, 2014.
- [164] Richard M. Leahy and Jinyi Qi. Statistical approaches in quantitative positron emission tomography. *Statistics and Computing*, 10(2):147–165, 2000.
- [165] Habib Zaidi and Kenneth F. Koral. Scatter modelling and compensation in emission tomography. *European Journal of Nuclear Medicine and Molecular Imaging*, 31(5):761–782, may 2004.
- [166] K. Thielemans, R. M. Manjeshwar, Ch Tsoumpas, and F. P. Jansen. A new algorithm for scaling of PET scatter estimates using all coincidence events. *IEEE Nuclear Science Symposium Conference Record*, 5:3586–3590, 2007.
- [167] Michael E. Casey and E.J. Hoffman. A Technique to Reduce Noise in Accidental Coincidence Measurements and Coincidence Efficiency Calibration. *Journal of computer assisted tomography*, 10(5), 1986.
- [168] L. Kaufman. Maximum likelihood, least squares, and penalized least squares for PET. *IEEE Transactions on Medical Imaging*, 12(2):200–214, jun 1993.
- [169] Jeffrey A Fessler. Penalized Weighted Least-Squares Image Reconstruction for Positron Emission Tomography. Technical Report 2, 1994.
- [170] James G Colsher. Fully three-dimensional positron emission tomography. *Physics in Medicine and Biology*, 25:103–115, 1980.

- [171] P.E. Kinahan and J.G. Rogers. Analytic 3D image reconstruction using all detected events. *IEEE Transactions on Nuclear Science*, 36(1):964–968, 1989.
- [172] Michel Defrise, P E Kinahan, D W Townsend, C Michel, M Sibomana, and D F Newport. Exact and Approximate Rebinning. *IEEE Transactions on Medical Imaging*, 16(2):145–157, 1997.
- [173] Samuel Matej and Ivan G. Kazantsev. Fourier-based reconstruction for fully 3-D PET: Optimization of interpolation parameters. *IEEE Transactions on Medical Imaging*, 25(7):845–854, 2006.
- [174] Claude Comtat, Paul E. Kinahan, Jeffrey A. Fessler, Thomas Beyer, David W. Townsend, Michel Defrise, and Christian Michel. *Clinically feasible reconstruction of 3D whole-body PET/CT data using blurred anatomical labels*, volume 47. 2002.
- [175] Tom Hebert and Richard Leahy. A Generalized EM Algorithm for 3-D Bayesian Reconstruction from Poisson Data Using Gibbs Priors. *IEEE Transactions on Medical Imaging*, 8(2):194–202, jun 1989.
- [176] J. Nuyts, D. Beque, P. Dupont, L. Mortelmans, D. Bequé, P. Dupont, and L. Mortelmans. A concave prior penalizing relative differences for maximum-a-posteriori reconstruction in emission tomography. *IEEE Transactions on Nuclear Science*, 49(1 I):56–60, feb 2002.
- [177] Peter J. Huber. *Robust Statistics*, volume 82. 1981.
- [178] K Lange. Convergence of EM image reconstruction algorithms with Gibbs smoothing. *IEEE Tr. Med. Im.*, 9(4):439–446, 1990.
- [179] Peter J Green. Bayesian Reconstructions From Emission Tomography Data Using a Modified EM Algorithm. *IEEE TRANSACTIONS ON MEDICAL IMAGING*, 9, 1990.
- [180] Stuart Geman and Donald Geman. Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-6(6):721–741, nov 1984.
- [181] J. P. Kaipio, V. Kolehmainen, M. Vauhkonen, and E. Somersalo. Inverse problems with structural prior information. *Inverse Problems*, 15(3):713–729, jun 1999.
- [182] Leonid I. Rudin, Stanley Osher, and Emad Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1-4):259–268, nov 1992.
- [183] Michael Hintermüller and M. Monserrat Rincon-Camacho. Expected absolute value estimators for a spatially adapted regularization parameter choice rule in L1-TV-based image restoration. *Inverse Problems*, 26(8), 2010.
- [184] F. Knoll, Y. Dong, C. Langkammer, M. Hintermuller, and R. Stollberger. Total Variation Denoising with Spatially Dependent Regularization 1. In *Proc. Intl. Soc. Mag. Reson. Med.* 18, 2010.

- [185] Emanuel Levitan and Gabor T. Herman. A Maximum a Posteriori Probability Expectation Maximization Algorithm for Image Reconstruction in Emission Tomography. *IEEE Transactions on Medical Imaging*, 6(3):185–192, sep 1987.
- [186] J. Nunez and J. Llacer. A fast Bayesian reconstruction algorithm for emission tomography with entropy prior converging to feasible images. *IEEE Transactions on Medical Imaging*, 9(2):159–171, jun 1990.
- [187] Michael Elad, Peyman Milanfar, and Ron Rubinstein. Analysis versus synthesis in signal priors. In *European Signal Processing Conference*, 2006.
- [188] Kenneth Lange, David R. Hunter, and Ilsoon Yang. Optimization Transfer Using Surrogate Objective Functions. *Journal of Computational and Graphical Statistics*, 9(1):1–20, 2000.
- [189] A.P. Dempster, N.M. Laird, and Donald B Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society Series B Methodological*, 39(1):1–38, 1977.
- [190] L. A. Shepp and Y. Vardi. Maximum Likelihood Reconstruction for Emission Tomography. *IEEE Transactions on Medical Imaging*, 1982.
- [191] Kenneth Lange and Richard Ellis Carson. EM Reconstruction Algorithms for Emission and Transmission Tomography. *Journal of computer assisted tomography*, 8(2):306–316, 1984.
- [192] Julian Matthews. private communication, 2019.
- [193] David R. Hunter and Kenneth Lange. A Tutorial on MM Algorithms. *American Statistician*, 58(1):30–37, 2004.
- [194] Alvaro R. De Pierro and Michel Eduardo Beleza Yamagishi. Fast EM-like methods for maximum "a posteriori" estimates in emission tomography. *IEEE Transactions on Medical Imaging*, 20(4):280–288, 2001.
- [195] Ing-Tsung Hsiao, Anand Rangarajan, Parmeshwar Khurd, and Gene Gindi. An accelerated convergent ordered subsets algorithm for emission tomography. *Physics in Medicine and Biology*, 49(11):2145–2156, jun 2004.
- [196] Peter J Green. On Use of the EM Algorithm for Penalized Likelihood Estimation. *Journal of the Royal Statistical Society*, 52(3):443–452, 1990.
- [197] Eugenio Rapisarda, Luca Presotto, Elisabetta De Bernardi, Maria Carla Gilardi, and Valentino Bettinardi. Optimized Bayes variational regularization prior for 3D PET images. *Computerized Medical Imaging and Graphics*, 38(6):445–457, 2014.
- [198] Abolfazl Mehranian, Habib Zaidi, and Andrew J. Reader. MR-guided joint reconstruction of activity and attenuation in brain PET-MR. *NeuroImage*, 162(June):276–288, 2017.
- [199] Jing Tang and Arman Rahmim. Bayesian PET image reconstruction incorporating anato-functional joint entropy. *Physics in medicine and biology*, 54(23):7063–75, dec 2009.

- [200] Jian Zhou, Jean Louis Coatrieux, Alexandre Bousse, Huazhong Shu, and Limin Luo. A Bayesian MAP-EM algorithm for PET image reconstruction using wavelet transform. *IEEE Transactions on Nuclear Science*, 54(5):1660–1669, 2007.
- [201] L Kaufman. Implementing and Accelerating the \mbox{EM} Algorithm for Positron Emission Tomography. *\mbox{IEEE Trans. Medical Imaging}*, M1-6(6):37–51, 1987.
- [202] Jeffrey A. Fessler and Scott D. Booth. Conjugate-gradient preconditioning methods for shift-variant PET image reconstruction. *IEEE Transactions on Image Processing*, 8(5):688–699, 1999.
- [203] Erkan Ü Mumcuoglu, Richard M Leahy, and Simon R Cherry. Bayesian reconstruction of PET images: methodology and performance analysis. *Physics in Medicine and Biology*, 41(9):1777–1807, sep 1996.
- [204] Erkan Mumcuoğlu, Richard Leahy, Zhenyu Zhou, and Simon R. Cherry. Fast Gradient-Based Methods for Bayesian Reconstruction of Transmission and Emission PET Images. *IEEE Transactions on Medical Imaging*, 13(4):687–701, 1994.
- [205] Babak A Ardekani, Michael Braun, Brian F Hutton, Iwao Kanno, and Hidehiro Iida. Minimum cross-entropy reconstruction of PET images using prior anatomical information. *Physics in Medicine & Biology*, 41:2497–2517, 1996.
- [206] J. Zheng, S.S. Saquib, K. Sauer, and C.A. Bouman. Parallelizable Bayesian tomography algorithms with rapid, guaranteed convergence. *IEEE Transactions on Image Processing*, 9(10):1745–1759, 2000.
- [207] K. Sauer and C. Bouman. A local update strategy for iterative reconstruction from projections. *IEEE Transactions on Signal Processing*, 41(2):534–548, 1993.
- [208] J.A. Fessler, E.P. Ficaro, N.H. Clinthorne, and K. Lange. Grouped-coordinate ascent algorithms for penalized-likelihood transmission image reconstruction. *IEEE Transactions on Medical Imaging*, 16(2):166–175, 1997.
- [209] Simon R. Cherry, James A. Sorenson, and Michael E. Phelps. Image Quality in Nuclear Medicine. In *Physics in Nuclear Medicine*. 2012.
- [210] Dong Zeng, Qi Xie, Wenfei Cao, Jiahui Lin, Shanli Zhang, Jing Huang, Zhaoying Bian, and Deyu Meng. Low-dose dynamic cerebral perfusion computed tomography reconstruction via intrinsic tensor sparsity regularization. 0062(c):1–10, 2017.
- [211] Charles E. Metz. ROC Methodology in Radiologic Imaging. *Investigative Radiology*, 21(9):720–733, 1986.
- [212] Kyle J. Myers and Harrison H. Barrett. Addition of a channel mechanism to the ideal-observer model. *Journal of the Optical Society of America A*, 4(12):2447, 1987.
- [213] H H Barrett, J Yao, J P Rolland, and K J Myers. Model observers for assessment of image quality. *Proceedings of the National Academy of Sciences of the United States of America*, 90(21):9758–65, 1993.

- [214] Moritz Schwyzer, Daniela A. Ferraro, Urs J. Muehlematter, Alessandra Curioni-Fontecedro, Martin W. Huellner, Gustav K. von Schulthess, Philipp A. Kaufmann, Irene A. Burger, and Michael Messerli. Automated detection of lung cancer at ultralow dose PET/CT by deep neural networks – Initial results. *Lung Cancer*, 126, 2018.
- [215] S. K. Lakshmanaprabu, Sachi Nandan Mohanty, K. Shankar, N. Arunkumar, and Gustavo Ramirez. Optimal deep learning model for classification of lung cancer on CT images. *Future Generation Computer Systems*, 92, 2019.
- [216] Gene Gindi, Mindy Lee, Anand Rangarajan, and I. George Zubal. Bayesian Reconstruction of Functional Images Using Anatomical Information as Priors. *IEEE Transactions on Medical Imaging*, 1993.
- [217] Jeffrey A Fessler, Neal H Clinthorne, and W Leslie Rogers. Regularized Emission Image Reconstruction Using Imperfect Side Information. *October*, 39(5):1464–1471, 1992.
- [218] Adam M Alessio and Paul E Kinahan. Improved quantitation for PET/CT image reconstruction with system modeling and anatomical priors. *Medical Physics*, 33(11), 2006.
- [219] Valen E. Johnson, W H Wong, X. P. Hu, and C. T. Chen. Bayesian restoration of PET images using Gibbs priors. In *Progress in clinical and biological research*, pages 15–28, 1991.
- [220] R. Leahy and X. Yan. Incorporation of anatomical MR data for improved functional imaging with PET. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 511 LNCS:105–120, 1991.
- [221] Srikanth Sastry and Richard E. Carson. Multimodality bayesian algorithm for image reconstruction in positron emission tomography: A tissue composition model. *IEEE Transactions on Medical Imaging*, 16(6):750–761, 1997.
- [222] B. Lipinski, H. Herzog, E. Rota Kops, W. Oberschelp, and H. W. Müller-Gärtner. Expectation maximization reconstruction of positron emission tomography images using anatomical magnetic resonance information. *IEEE Transactions on Medical Imaging*, 16(2):129–136, 1997.
- [223] James E. Bowsher, Valen E. Johnson, Timothy G. Turkington, Ronald J. Jaszcak, Carey E. Floyd, and R. Edward Coleman. Bayesian reconstruction and use of anatomical a priori information for emission tomography. *IEEE Transactions on Medical Imaging*, 1996.
- [224] J.E. Bowsher, H Yuan, L.W. Hedlund, T.G. Turkington, G. Akabani, A. Badea, W.C. Kurylo, C.T. Wheeler, G.P. Cofer, M.W. Dewhirst, and G.A. Johnson. Utilizing MRI Information to Estimate F18-FDG Distributions in Rat Flank Tumors. *IEEE Symposium Conference Record Nuclear Science 2004.*, 4(C):2488–2492, 2004.
- [225] Rebekka Loeb, Nassir Navab, and Sibylle I. Ziegler. Direct Parametric Reconstruction Using Anatomical Regularization for Simultaneous PET/MRI Data. *IEEE Transactions on Medical Imaging*, 34(11):2233–2247, nov 2015.

- [226] Nguyen, Van-Giang and Soo-Jin Lee. Incorporating Anatomical Side Information Into PET Reconstruction Using Nonlocal Regularization. *IEEE Transactions on Image Processing*, 22(10):3961–3973, oct 2013.
- [227] Matthias Joachim Ehrhardt and Simon R. Arridge. Vector-valued image processing by parallel level sets. *IEEE Transactions on Image Processing*, 23(1):9–18, 2014.
- [228] Matthias Joachim Ehrhardt, Kris Thielemans, Luis Pizarro, David Atkinson, Sébastien Ourselin, Brian F Hutton, and Simon R Arridge. Joint reconstruction of PET-MRI by exploiting structural similarity. *Inverse Problems*, 31(1):015001, 2015.
- [229] Yu-Jung Tsai, Georg Schramm, Sangtae Ahn, Alexandre Bousse, Simon Arridge, Johan Nuyts, Brian F. Hutton, Charles W. Stearns, and Kris Thielemans. Benefits of Using a Spatially-variant Penalty Strength with Anatomical Priors in PET Reconstruction. *IEEE Transactions on Medical Imaging*, PP(c):1–1, 2019.
- [230] Sangeetha Somayajula, Christos Panagiotou, Anand Rangarajan, Quanzheng Li, Simon R Arridge, and Richard M Leahy. Theoretic Anatomical Priors. *Health (San Francisco)*, 30(3):537–549, 2011.
- [231] Johan Nuyts. The use of mutual information and joint entropy for anatomical priors in emission tomography. In *2007 IEEE Nuclear Science Symposium Conference Record*, pages 4149–4154. IEEE, 2007.
- [232] Marzieh S Tahaei and Andrew J Reader. Patch-based image reconstruction for PET using prior-image derived dictionaries. *Physics in Medicine and Biology*, 61(18):6833–6855, 2016.
- [233] Shuhang Chen, Huafeng Liu, Pengcheng Shi, and Yunmei Chen. Sparse representation and dictionary learning penalized image reconstruction for positron emission tomography. *Phys. Med. Biol*, 60, 2015.
- [234] J. Nuyts, K. Baete, D. Beque, and P. Dupont. Comparison between MAP and post-processed ML for incorporating anatomical knowledge in emission tomography. 24(5), 2005.
- [235] C. E. Metz and R. N. Beck. Quantitative effects of stationary linear image processing on noise and resolution of structure in radionuclide images. *Journal of Nuclear Medicine*, 15(3):164–170, 1974.
- [236] T H Farquhar, Student Member, A Chatziioannou, G Chinn, M Dahlbom, Senior Member, and E J Hoffman. An Investigation of Filter Choice for Filtered Back-Projection Reconstruction in PET T. *IEEE Transactions on Nuclear Science*, 45(3):1133–1137, 1998.
- [237] Sari Peltonen, Uygur Tuna, Enrique Sánchez-Monge, and Ulla Ruotsalainen. PET sinogram denoising by block-matching and 3D filtering. *IEEE Nuclear Science Symposium Conference Record*, pages 3125–3129, 2012.
- [238] Joel S. Karp, Gerd Muehllehner, and Robert M. Lewitt. Constrained Fourier Space Method for Compensation of Missing Data in Emission Computed Tomography. *IEEE Transactions on Medical Imaging*, 7(1):21–25, 1988.

- [239] Hossein Arabi and Habib Zaidi. Improvement of image quality in PET using post-reconstruction hybrid spatial-frequency domain filtering. *Physics in Medicine and Biology*, 63(21), 2018.
- [240] M Jacobson, R Levkovitz, A Ben-Tal, K Thielemans, T Spinks, D Belluzzo, E Pagani, V Bettinardi, MC Gilardi, A Zverovich, and G Mitra. Enhanced 3D PET OSEM reconstruction using inter-update Metz filtering. *Phys. Med. Biol.*, 45(8):2417, 2000.
- [241] Omer Demirkaya. Post-reconstruction filtering of positron emission tomography whole-body emission images and attenuation maps using nonlinear diffusion filtering. *Academic Radiology*, 11(10):1105–1114, 2004.
- [242] Pietro Perona and Jitendra Malik. Scale-space and edge detection using anisotropic diffusion. *IEEE Transactions On Pattern Analysis and Machine Intelligence*, 12(7):629–639, 1990.
- [243] A. Buades and B. Coll. A non-local algorithm for image denoising. *Computer Vision and Pattern*, 2(0):60–65, 2005.
- [244] Kostadin Dobov, Alessandro Foi, and Karen Egiazarian. Video denoising by sparse 3D transform-domain collaborative filtering. In *15th European Signal Processing Conference*, volume 7, pages 145–149, 2007.
- [245] Chung Chan, Roger Fulton, Robert Barnett, David Dagan Feng, and Steven Meikle. Postreconstruction Nonlocal Means Filtering of Whole-Body PET with an Anatomical Prior. *IEEE Transactions on Medical Imaging*, 33(3):636–650, 2014.
- [246] Bradley T Christian, Nicholas T Vandehey, John M Floberg, and Charles A Mistretta. Dynamic PET denoising with HYPR processing. *Journal of nuclear medicine : official publication, Society of Nuclear Medicine*, 51(7):1147–1154, 2010.
- [247] Benjamin Spencer, Jinyi Qi, Ramsey D. Badawi, and Guobao Wang. Dynamic PET Image reconstruction for parametric imaging using the HYPR kernel method. *2017 SPIE Medical Imaging Proceedings*, 10132, 2017.
- [248] Kaiming He, Jian Sun, and Xiaoou Tang. Guided image filtering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(6):1397–1409, 2013.
- [249] Jianhua Yan, Jason Chu Shern Lim, and David W. Townsend. MRI-guided brain PET image filtering and partial volume correction. *Physics in Medicine and Biology*, 60(3):961–976, 2015.
- [250] Edward J. Hoffman, Sung Cheng Huang, and Michael E. Phelps. Quantitation in positron emission computed tomography: 1. effect of object size, 1979.
- [251] C C Meltzer, J P Leal, H S Mayberg, Wagner H N Jr, and J M Frost. Correction of PET data for partial volume effects in human cerebral cortex by MR imaging. *Journal of computer assisted tomography*, 14(4):561–70, 1990.
- [252] H W Muller-Gartner, J M Links, J L Prince, P R Brakeman, R N Bryan, E Mcveigh, J P Leal, C Davatzikos, and J J Frost. Measurement of Radiotracer Concentration in Brain Gray-Matter Using Pet (Gm-Pet) - Mri-Based Correction for Partial Volume Effect. *Eur J Nucl Med*, 19(8):645, 1992.

- [253] O G Rousset, Y Ma, and A C Evans. Correction for partial volume effects in PET: principle and validation. *Journal of nuclear medicine : official publication, Society of Nuclear Medicine*, 39(5):904–11, 1998.
- [254] Vincent Frouin, Claude Comtat, Anthonin Reilhac, and Marie-Claude Grégoire. Correction of partial-volume effect for PET striatal imaging: fast implementation and study of robustness. *Journal of Nuclear Medicine*, 43(12):1715–26, 2002.
- [255] Habib Zaidi, Torsten Ruest, Frederic Schoenahl, and Marie-Louise Montandon. Comparative assessment of statistical brain MR image segmentation algorithms and their impact on partial volume correction in PET. *NeuroImage*, 32(4):1591–1607, oct 2006.
- [256] Douglas N. Greve, David H. Salat, Spencer L. Bowen, David Izquierdo-Garcia, Aaron P. Schultz, Ciprian Catana, J. Alex Becker, Claus Svarer, Gitte M. Knudsen, Reisa A. Sperling, and Keith A. Johnson. Different partial volume correction methods lead to different conclusions: An 18 F-FDG-PET study of aging. *NeuroImage*, 2016.
- [257] Florian Knoll, Martin Holler, Thomas Koesters, Ricardo Otazo, Kristian Bredies, and Daniel K Sodickson. Joint MR-PET Reconstruction Using a Multi-Channel Image Regularizer. *IEEE Transactions on Medical Imaging*, 36(1):1–16, jan 2017.
- [258] Ying Zhang and Xiaoqun Zhang. PET-MRI joint reconstruction with common edge weighted total variation regularization. *Inverse Problems*, 34, 2018.
- [259] Abolfazl Mehranian, Martin A. Belzunce, Claudia Prieto, Alexander Hammers, and Andrew J. Reader. Synergistic PET and SENSE MR Image Reconstruction Using Joint Sparsity Regularization. *IEEE Transactions on Medical Imaging*, 37(1):20–34, 2018.
- [260] Abolfazl Mehranian, Martin A. Belzunce, Colm J. McGinnity, Aurelien Bustin, Claudia Prieto, Alexander Hammers, and Andrew J. Reader. Multi-modal synergistic PET and MR reconstruction using mutually weighted quadratic priors. *Magnetic Resonance in Medicine*, 81(3):2120–2134, mar 2019.
- [261] V Y Panin, Student Member Zeee, G L Zeng, Member Zeee, G T Gullberg, and Senior Member Zeee. Total Variation Regulated EM Algorithm. *IEEE Transactions on Nuclear Science*, 46(6):2202–2210, 1999.
- [262] Xingjian Yu, Chenye Wang, Hongjie Hu, and Huafeng Liu. Low dose PET image reconstruction with total variation using alternating direction method. *PLoS ONE*, 11(12):1–19, 2016.
- [263] Peter J. Huber. Robust Estimation of a Location Parameter. *Ann. Math. Stat.*, 35(1):73–101, 1964.
- [264] Jing Tang, Yanhua Wang, Rutao Yao, and Leslie Ying. Sparsity-Based Pet Image Reconstruction Using Mri Learned Dictionaries. (1):1087–1090, 2014.
- [265] Kyong Hwan Jin, Michael T. McCann, Emmanuel Froustey, and Michael Unser. Deep Convolutional Neural Network for Inverse Problems in Imaging. *IEEE Transactions on Image Processing*, 26(9):4509–4522, 2017.

- [266] Hu Chen, Yi Zhang, Weihua Zhang, Peixi Liao, Ke Li, Jiliu Zhou, and Ge Wang. a Low-dose CT via convolutional neural network. *Biomedical Optics Express*, 8(2):679, 2017.
- [267] Shanshan Wang, Zhenghang Su, Leslie Ying, Xi Peng, Shun Zhu, Feng Liang, Dagan Feng, and Dong Liang. Accelerating magnetic resonance imaging via deep learning. *Proceedings - International Symposium on Biomedical Imaging*, 2016-June:514–517, 2016.
- [268] Kerstin Hammernik, Teresa Klatzer, Erich Kobler, Michael P. Recht, Daniel K. Sodickson, Thomas Pock, and Florian Knoll. Learning a variational network for reconstruction of accelerated MRI data. *Magnetic Resonance in Medicine*, 79(6):3055–3071, 2018.
- [269] Hemant K. Aggarwal, Merry P. Mani, and Mathews Jacob. MoDL: Model Based Deep Learning Architecture for Inverse Problems. *IEEE Transactions on Medical Imaging*, PP(c):1, 2018.
- [270] Jo Schlemper, Jose Caballero, Joseph V Hajnal, Anthony N Price, and Daniel Rueckert. A Deep Cascade of Convolutional Neural Networks for Dynamic MR Image Reconstruction. *IEEE Transactions on Medical Imaging*, 37(2):491–503, 2018.
- [271] Yoseob Han, Jaejun Yoo, Hak Hee Kim, Hee Jung Shin, Kyunghyun Sung, and Jong Chul Ye. Deep learning with domain adaptation for accelerated projection-reconstruction MR. *Magnetic Resonance in Medicine*, 80(3):1189–1205, 2018.
- [272] Kuang Gong, Dufan Wu, Quanzheng Li, Kyungsang Kim, Jaewon Yang, Georges El Fakhri, and Youngho Seo. EMnet: an unrolled deep neural network for PET image reconstruction. *Proc. SPIE 10948, Medical Imaging 2019: Physics of Medical Imaging*, 1094853(March):185, 2019.
- [273] Kevin T. Chen, Enhao Gong, Fabiola Bezerra de Carvalho Macruz, Junshen Xu, Athanasia Boumis, Mehdi Khalighi, Kathleen L. Poston, Sharon J. Sha, Michael D. Greicius, Elizabeth Mormino, John M. Pauly, Shyam Srinivas, and Greg Zaharchuk. Ultra-Low-Dose 18 F-Florbetaben Amyloid PET Imaging Using Deep Learning with Multi-Contrast MRI Inputs. *Radiology*, 290(3):649–656, 2019.
- [274] Jieqing Jiao and Sebastien Ourselin. Fast PET reconstruction using Multi-scale Fully Convolutional Neural Networks. pages 1–8, 2017.
- [275] Kuang Gong, Massachusetts General Hospital, Dufan Wu, Massachusetts General Hospital, Jaewon Yang, Sun Tao, and Mass General Hospital. MAPEM-Net : An Unrolled Neural Network for Fully 3D PET Image Reconstruction. Number June, 2019.
- [276] Hongki Lim, Il Yong Chun, Yuni K Dewaraja, and Jeffrey A. Fessler. Improved low-count quantitative PET reconstruction with a recurrent neural network. pages 1–11, 2019.
- [277] Bo Zhu, Jeremiah Z. Liu, Stephen F. Cauley, Bruce R. Rosen, and Matthew S. Rosen. Image reconstruction by domain-transform manifold learning. *Nature*, 555(7697):487–492, 2018.

- [278] Ida Häggström, C. Ross Schmidlein, Gabriele Campanella, and Thomas J. Fuchs. DeepPET: A deep encoder-decoder network for directly solving the PET reconstruction inverse problem. *Medical Image Analysis*, 54:253–262, 2018.
- [279] Casper O. da Costa-Luis and Andrew J. Reader. Deep Learning for Suppression of Resolution-Recovery Artefacts in MLEM PET Image Reconstruction. *2017 IEEE Nuclear Science Symposium and Medical Imaging Conference, NSS/MIC 2017 - Conference Proceedings*, pages 7–9, 2018.
- [280] Kuang Gong, Jiahui Guan, Chih-Chieh Liu, and Jinyi Qi. PET Image Denoising Using a Deep Neural Network Through Fine Tuning. *IEEE Transactions on Radiation and Plasma Medical Sciences*, (1):1–1, 2018.
- [281] Alexander Mordvintsev, Christopher Olah, and Mike Tyka. DeepDream - a code example for visualizing Neural Networks, 2015.
- [282] James R. Clough, Ilkay Oksuz, Esther Puyol-Anton, Bram Ruijsink, Andrew P. King, and Julia A. Schnabel. Global and Local Interpretability for Cardiac MRI Classification. (June), 2019.
- [283] National Electrical Manufacturers Association. NEMA Standards Publication NU 2–2007, Performance measurements of positron emission tomographs. Technical report, 2007.
- [284] E. J. Hoffman, P. D. Cutler, W. M. Digby, and J. C. Mazziotta. 3-D phantom to simulate cerebral blood flow and metabolic images for PET. *IEEE Transactions on Nuclear Science*, 37(2):616–620, 1990.
- [285] ICRP. ICRP 106 Publication. Radiation Dose to Patients from Radiopharmaceuticals. *Annals of the ICRP*, 38(1-2):21–24, 2008.
- [286] Aruna Kaushik, Abhinav Jaimini, Madhavi Tripathi, Maria D’Souza, Rajnish Sharma, Anupam Mondal, Anil K. Mishra, and Bilikere S. Dwarakanath. Estimation of radiation dose to patients from 18FDG whole body PET/CT investigations using dynamic PET scan protocol. *Indian Journal of Medical Research*, 142(6):721–731, 2015.
- [287] UK Government. Ionising radiation: dose comparisons, 2011.
- [288] R Edward Hendrick. Radiation Doses and Cancer Risks from Breast Imaging Studies. *Radiology*, 257(1):246–253, 2010.
- [289] Ferdinand Seith, Holger Schmidt, Julia Kunz, Thomas Kuestner, Sergios Gatidis, Konstantin Nikolaou, Christian la Fougère, and Nina F Schwenzer. *Simulation of tracer dose reduction in ¹⁸F-FDG-Positron emission tomography / magnetic resonance imaging (PET/MRI): Effects on oncologic reading, image quality and artifacts*. 2017.
- [290] K. Herholz, R. Evans, J. Anton-Rodriguez, R. Hinz, and J. C. Matthews. The effect of 18F-florbetapir dose reduction on region-based classification of cortical amyloid deposition. *European Journal of Nuclear Medicine and Molecular Imaging*, 41(11):2144–2149, 2014.

- [291] Joshua D Schaefferkoetter, Jianhua Yan, Therese Anna Soderlund, David W Townsend, Maurizio Conti, John Kit Chung Tam, Ross Soo, and Ivan Tham. Quantitative Accuracy and Lesion Detectability of Low-Dose FDG-PET for Lung Cancer Screening. *Journal of nuclear medicine : official publication, Society of Nuclear Medicine*, pages 1–26, 2016.
- [292] James Bland, Abolfazl Mehranian, Martin A. Belzunce, Sam Ellis, Colm J. McGinnity, Alexander Hammers, and Andrew J. Reader. MR-Guided Kernel EM Reconstruction for Reduced Dose PET Imaging. *IEEE Transactions on Radiation and Plasma Medical Sciences*, in press(3):1–1, 2017.
- [293] E Boser, Isabelle M Guyon, and N Vapnik. A Training Algorithm for Optimal Margin Classifiers. In *COLT '92 Proceedings of the fifth annual workshop on Computational learning theory*, pages 144–152, 1992.
- [294] Vladimir Vapnik. *The Nature of Statistical Learning Theory*. 1995.
- [295] Philip Novosad and Andrew J Reader. MR-guided dynamic PET reconstruction with the kernel method and spectral temporal basis functions. *Physics in Medicine and Biology*, 2016.
- [296] Daniel Deidda, Student Member, Nicolas Karakatsanis, Senior Member, Philip M Robson, Nikos Efthimiou, Zahi A Fayad, Robert G Aykroyd, Charalampos Tsoumpas, and Senior Member. Effect of PET-MR Inconsistency in the Kernel Image Reconstruction Method. *IEEE Transactions on Radiation and Plasma Medical Sciences*, PP:1, 2018.
- [297] Rehemai Baikejiang, Yue Zhao, Brett Z Fite, Katherine W Ferrara, and Changqing Li. Anatomical image-guided fluorescence molecular tomography reconstruction using kernel method. *Journal of biomedical optics*, 22(5), 2017.
- [298] Yue Zhao, Rehemai Baikejiang, and Changqing Li. Application of kernel method in fluorescence molecular tomography. In Fred S. Azar and Xavier Intes, editors, *Proc. SPIE 10057, Multimodal Biomedical Imaging XII*, volume 100570P. International Society for Optics and Photonics, feb 2017.
- [299] Rehemai Baikejiang, Wei Zhang, and Changqing Li. Gaussian kernel based anatomically-aided diffuse optical tomography reconstruction. In *Proc. SPIE 10059, Optical Tomography and Spectroscopy of Tissue XII*, volume 1005912, page 1005912, feb 2017.
- [300] Guobao Wang. High Temporal-Resolution Dynamic PET Image Reconstruction Using a New Spatiotemporal Kernel Method. *IEEE Transactions on Medical Imaging*, PP(c):1, 2018.
- [301] Kuang Gong, Jinxiu Cheng-Liao, Guobao Wang, Kevin T. Chen, Ciprian Catana, and Jinyi Qi. Direct Patlak Reconstruction from Dynamic PET Data Using the Kernel Method with MRI Information Based on Structural Similarity. *IEEE Transactions on Medical Imaging*, 0062(1):1–11, apr 2017.

- [302] Guobao Wang, Jinyi Qi, Guobao Wang, and Jinyi Qi. Penalized likelihood PET image reconstruction using patch-based edge-preserving regularization. *IEEE Transactions on Medical Imaging*, 31(12):2194–2204, dec 2012.
- [303] Andrew R Barron. Universal Approximation Bounds for Superpositions of a Sigmoidal Function. *IEEE Transactions on Information Theory*, 39(3):930–945, 1993.
- [304] Ju-Chieh (Kevin) Cheng, Julian Matthews, Vesna Sossi, Jose Anton-Rodriguez, André Salomon, and Ronald Boellaard. Incorporating HYPR de-noising within iterative PET reconstruction (HYPR-OSEM). *Phys. Med. Biol*, 62, 2017.
- [305] Sam Ellis and Andrew J. Reader. Kernelised EM Image Reconstruction for Dual-Dataset PET Studies. *2016 IEEE Nuclear Science Symposium, Medical Imaging Conference*, pages 1–3, oct 2016.
- [306] Chris Cocosco, Vasken Kollokian, Remi K Kwan, G Bruce Pike, and Alan C Evans. BrainWeb : Online Interface to a 3D MRI Simulated Brain Database. *3-rd International Conference on Functional Mapping of the Human Brain*, 5(4):S425, 1997.
- [307] Sam Ellis and Andrew J. Reader. Penalized maximum likelihood simultaneous longitudinal PET image reconstruction with difference-image priors. *Medical Physics*, 45(7):3001–3018, 2018.
- [308] Chih-Chieh Liu and Jinyi Qi. Higher SNR PET Image Prediction using A Deep Learning Model and MRI Image. *Phys. Med. Biol*, 2019.
- [309] Martin A. Belzunce, Abolfazl Mehranian, Z. Chalampalakakis, and Andrew J. Reader. Evaluation of shift-invariant image-space PSFs for the Biograph mMR PET Scanner. In *PSMR Conference, Lisbon*,, 2017.
- [310] William J. Jagust, Dan Bandy, Kewei Chen, Norman L. Foster, Susan M. Landau, Chester A. Mathis, Julie C. Price, Eric M. Reiman, Daniel Skovronsky, and Robert A. Koeppe. The Alzheimer’s Disease Neuroimaging Initiative positron emission tomography core. *Alzheimer’s and Dementia*, 2010.
- [311] D. Deidda, N. Karakatsanis, P. M. Robson, N. Efthimiou, Z. A. Fayad, R. G. Aykroyd, and C. Tsoumpas. Hybrid PET-MR listmode kernelized expectation maximization reconstruction for quantitative PET images of the carotid arteries. In *NSS/MIC/RTSD, Atlanta, Georgia*, page (in press), 2017.
- [312] Abolfazl Mehranian, James Bland, Colm J. McGinnity, Alexander Hammers, and Andrew J. Reader. Clinical assessment of MR-assisted PET image reconstruction algorithms for low-dose brain PET imaging. In *IEEE Nuclear Science Symposium Conference Record*, 2019.
- [313] Alvaro R. De Pierro. Modified expectation maximization algorithm for penalized likelihood estimation in emission tomography. *IEEE Transactions on Medical Imaging*, 14(1):132–137, 1995.
- [314] Jeffrey A. Fessler and W. Leslie Rogers. Spatial resolution properties of penalized-likelihood image reconstruction: Space-invariant tomographs. *IEEE Transactions on Image Processing*, 5(9):1346–1358, 1996.

- [315] Muqing Zhou, Zhaoguo Wu, Difan Chen, and Ya Zhou. An improved vein image segmentation algorithm based on SLIC and Niblack threshold method. In Xinggang Lin and Jesse Zheng, editors, *International conference on optical instruments and technology*, volume 9045, page 90450D. SPIE-INT SOC OPTICAL ENGINEERING, 1000 20TH ST, PO BOX 10, BELLINGHAM, WA 98227-0010 USA, dec 2013.
- [316] Maroua Mehri, Nabil Sliti, Pierre Héroux, Petra Gomez-Krämer, Najoua Essoukri Ben Amara, and Rémy Mullot. Use of SLIC superpixels for ancient document image enhancement and segmentation. In *Document recognition and retrieval XXII*, volume 9402, page 940205, feb 2015.
- [317] X.D. Bai, Z.G. Cao, Y. Wang, M.N. Ye, and L. Zhu. Image segmentation using modified SLIC and Nyström based spectral clustering. *Optik - International Journal for Light and Electron Optics*, 125(16):4302–4307, aug 2014.
- [318] James Bland, Abolfazl Mehranian, Martin A Belzunce, Casper da Costa-Luis, Sam Ellis, Colm J McGinnity, Alexander Hammers, and Andrew J Reader. Intercomparison of MR-Informed Methods for PET Image Reconstruction. *Medical physics*, (in press), 2019.
- [319] D. Kazantsev, S. Ourselin, B. F. Hutton, K. J. Dobson, A. P. Kaestner, W. R. B. Lionheart, P. J. Withers, P. D. Lee, and S. R. Arridge. A novel technique to incorporate structural prior information into multi-modal tomographic reconstruction. *Inverse Problems*, 30(6):065004, 2014.
- [320] Abolfazl Mehranian, Martin A. Belzunce, Flavia Niccolini, Marios Politis, Claudia Prieto, Federico Turkheimer, Alexander Hammers, and Andrew J. Reader. PET image reconstruction using multi-parametric anato-functional priors. *Physics in Medicine and Biology*, 62(15):5975–6007, jul 2017.
- [321] Daniil Kazantsev, Alexandre Bousse, Stefano Pedemonte, Simon R. Arridge, Brian F. Hutton, and Sebastien Ourselin. Edge preserving bowsher prior with nonlocal weighting for 3D spect reconstruction. *Proceedings - International Symposium on Biomedical Imaging*, pages 1158–1161, 2011.
- [322] Casper da Costa-Luis. BrainWeb-based multimodal models of 20 normal brains, 2019.
- [323] Zhou Wang ; A.C. Bovik ; H.R. Sheikh ; E.P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004.
- [324] Julian Rasch, Eva Maria Brinkmann, and Martin Burger. Joint reconstruction via coupled Bregman iterations with applications to PET-MR imaging. *Inverse Problems*, 34(1), 2018.
- [325] Stephen M. Smith, Mark W. Woolrich, Saad Jbabdi, Brian Patenaude, Christian Beckmann, Mark Jenkinson, Michael Chappell, Salima Makni, and Timothy Behrens. Bayesian analysis of neuroimaging data in FSL. *NeuroImage*, 45(1):S173–S186, 2008.
- [326] Simon Stute and Claude Comtat. Practical considerations for image-based PSF and blobs reconstruction in PET. *Physics in medicine and biology*, 58(11):3849–3870, 2013.

- [327] Charles E. Metz. Basic principles of ROC analysis. *Seminars in Nuclear Medicine*, 1978.
- [328] Dufan Wu, Kyungsang Kim, Georges El Fakhri, and Quanzheng Li. A Cascaded Convolutional Neural Network for X-ray Low-dose CT Image Denoising. 2017.
- [329] Hu Chen, Yi Zhang, Weihua Zhang, Peixi Liao, Ke Li, Jiliu Zhou, and Ge Wang. a Low-dose CT via convolutional neural network. *Biomedical Optics Express*, 8(2):679, 2017.
- [330] Lishui Cheng, Sangtae Ahn, Steven G Ross, Hua Qian, and Bruno De Man. Accelerated Iterative Image Reconstruction Using a Deep Learning Based Leapfrogging Strategy. *International Conference on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine*, (June 2018):715–20, 2017.
- [331] D. S. Rigie, Georg Schramm, T. Vahle, T. Shepard, Johan Nuyts, and F. E. Boada. Approximating MRI-Based Anatomically Guided PET Reconstruction with a Convolutional Neural Network. In *IEEE Nuclear Science Symposium & Medical Imaging Conference (NSS/MIC)*, 2019.
- [332] Yan Yang, Jian Sun, Huibin Li, and Zongben Xu. Deep ADMM-Net for compressive sensing MRI. *Advances in Neural Information Processing Systems*, (Nips):10–18, 2016.
- [333] Hu Chen, Yi Zhang, Yunjin Chen, Jungfeng Zhang, Weihua Zhang, Huaiqiang Sun, Yang Lv, Peixi Liao, Jiliu Zhou, and Ge Wang. LEARN: Learned Experts’ Assessment-Based Reconstruction Network for Sparse-Data CT. *IEEE Trans Med Imaging*, 37(6):1333–1347, 2018.
- [334] Ji He, Yan Yang, Yongbo Wang, Dong Zeng, Zhaoying Bian, Hao Zhang, Jian Sun, Zongben Xu, and Jianhua Ma. Optimizing a Parameterized Plug-and-Play ADMM for Iterative Low-Dose CT Reconstruction. *IEEE Transactions on Medical Imaging*, 38(2):371–382, 2019.
- [335] Dufan Wu, Kyungsang Kim, Georges El Fakhri, and Quanzheng Li. Iterative Low-dose CT Reconstruction with Priors Trained by Artificial Neural Network. *IEEE Transactions on Medical Imaging*, 36(12):2479–2486, 2017.
- [336] Dufan Wu, Kyungsang Kim, and Quanzheng Li. Computationally Efficient Deep Neural Network for Computed Tomography Image Reconstruction. *Medical Physics*, 2019.
- [337] Lei Xiang, Yu Qiao, Dong Nie, Le An, Weili Lin, Qian Wang, and Dinggang Shen. Deep auto-context convolutional neural networks for standard-dose PET image estimation from low-dose PET/MRI. *Neurocomputing*, 267:406–416, 2017.
- [338] Kuang Gong, Jiahui Guan, Kyungsang Kim, Xuezhu Zhang, Jaewon Yang, Youngho Seo, Georges El Fakhri, Jinyi Qi, and Quanzheng Li. Iterative PET Image Reconstruction Using Convolutional Neural Network Representation. *IEEE Transactions on Medical Imaging*, PP(c):1, 2018.

- [339] Yan Wang, Biting Yu, Lei Wang, Chen Zu, David S. Lalush, Weili Lin, Xi Wu, Jiliu Zhou, Dinggang Shen, and Luping Zhou. 3D conditional generative adversarial networks for high-quality PET image estimation at low dose. *NeuroImage*, 174(March):550–562, 2018.
- [340] James Bland, Abolfazl Mehranian, Casper Costa-Luis, and Andrew J. Reader. Multi-Module Deep Learning for Enhanced and Accelerated PET Image Reconstruction. In *IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)*, 2019.
- [341] Jinyi Qi and Richard M Leahy. Iterative reconstruction techniques in emission computed tomography. *Physics in Medicine and Biology*, 51(15):R541–R578, 2006.
- [342] Diederik P Kingma and Jimmy Lei Ba. ADAM: A METHOD FOR STOCHASTIC OPTIMIZATION. In *International Conference on Learning Representations*, pages 1–15, 2015.

Appendix A

MLEM Theory

This chapter presents a more in depth derivation of the MLEM update formula presented in chapter 2, Eq. 2.27. The indices i and j correspond to the projection bins and the voxel locations, respectively. The symbols used in this appendix will follow those used in the thesis, but with some exceptions as indicated below.

- m_i - number of counts in bin i .
- $\hat{m}_i$ - expected number of counts in bin i .
- θ_j - number of events at voxel j .
- $\hat{\theta}_j$ - expected number of events at voxel j . This is different from normal notation where θ (without hat) normally is the expected number of events !
- ρ_{ij} - number of counts in bin i due to events at voxel j (complete data).
- $\hat{\rho}_{ij}$ - expected number of counts in bin i due to events at voxel j

A number of mathematical definitions listed below (without proof) are required in the derivation of MLEM.

- Bayes' theorem:

$$\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)} = \frac{\Pr(B|A)\Pr(A)}{\Pr(B)} \quad (\text{A.1})$$

- Probability mass function for a multivariate binomial distribution:

$$\Pr(n_1, \dots, n_k) = \frac{N!}{n_1! \dots n_k!} \Pr(n_1)^{n_1} \dots \Pr(n_k)^{n_k} \quad (\text{A.2})$$

- Expectation for a multivariate binomial distribution:

$$E[n_i | \sum_i n_i = N] = N \times \Pr(n_i) \quad (\text{A.3})$$

- Sum of independent Poisson random variables is Poisson

The derivation of MLEM is based on the maximisation of the complete data Poisson likelihood objective function. The complete data (ρ_{ij}) states the number of events at voxel j that are detected at projection bin i . The complete data simplifies the maximisation of the Poisson likelihood however is unknown. The complete data Poisson likelihood is:

$$\Pr(\rho_{ij} | \hat{\rho}_{ij}) = \frac{\hat{\rho}_{ij}^{\rho_{ij}} \exp(-\hat{\rho}_{ij})}{\rho_{ij}!} \quad (\text{A.4})$$

where

$$\hat{\rho}_{ij} = a_{ij} \hat{\theta}_j \quad (\text{A.5})$$

The above definition can be derived from a geometric argument, whereby a fraction of events at voxel j will be detected in projection bin i in accordance to the system matrix a_{ij} . Assuming the complete data variables are independent, the probability mass function for a specific complete data random variable is given by the products of the individual

probabilities:

$$\Pr(\boldsymbol{\rho}|\hat{\boldsymbol{\rho}}) = \prod_{ij} \frac{\hat{\rho}_{ij}^{\rho_{ij}} \exp(-\hat{\rho}_{ij})}{\rho_{ij}!} \quad (\text{A.6})$$

The above objective function is simplified through application of the log function. We seek to maximise this objective function (Poisson log likelihood) with respect to the object expectation vector ($\hat{\boldsymbol{\theta}}$). The factorial term is dropped due to its independence with respect to $\hat{\boldsymbol{\theta}}$. The switch from probability to likelihood is simply abiding by terminology. The probability states how likely a specific random variable is given known expectation values. Whereas the likelihood states how likely an expectation value is given a set of measured random variables.

$$L(\hat{\boldsymbol{\rho}}|\boldsymbol{\rho}) = \prod_{ij} \frac{\hat{\rho}_{ij}^{\rho_{ij}} \exp(-\hat{\rho}_{ij})}{\rho_{ij}!} \quad (\text{A.7})$$

$$\Omega^{ML}(\hat{\boldsymbol{\rho}}|\boldsymbol{\rho}) = \sum_{i,j} \rho_{ij} \log(\hat{\rho}_{ij}) - \hat{\rho}_{ij} - \log(\rho_{ij}!) \quad (\text{A.8})$$

The $\rho!$ term can be ignored due to its independence with respect to $\hat{\boldsymbol{\theta}}$. The complete data random variable is unknown. But the conditional expectation of the complete data can be found. Notice that the conditional expectation of the complete data $E(\rho_{ij}|\hat{\boldsymbol{\rho}}, m_i)$ is different from the expectation of the complete data $E(\rho_{ij}) \equiv \hat{\rho}_{ij}$.

$$E \left[\Omega^{ML}(\hat{\boldsymbol{\rho}}|\boldsymbol{\rho}) | \hat{\boldsymbol{\theta}}^k, \mathbf{m} \right] = \sum_{i,j} E \left[\rho_{ij} | \hat{\boldsymbol{\theta}}^k, m_i \right] \log(\hat{\rho}_{ij}) - \hat{\rho}_{ij} \quad (\text{A.9})$$

The expectation only acts on random variables i.e. ρ_{ij} returning the average value of the random variable if repeated an infinite number of times. As $\hat{\rho}_{ij}$ is already a number, applying the expectation to it has no effect. The maximum of the conditional expectation of the Poisson log likelihood is found by taking the derivative of the expected conditional Poisson log likelihood with respect to $\hat{\boldsymbol{\theta}}$.

$$\frac{\partial E \left[\Omega^{ML}(\hat{\boldsymbol{\rho}}|\boldsymbol{\rho}) | \hat{\boldsymbol{\theta}}^k, \mathbf{m} \right]}{\partial \hat{\theta}_j} = \sum_i \frac{E \left[\rho_{ij} | \hat{\boldsymbol{\theta}}^k, m_i \right]}{\hat{\theta}_j} - a_{ij} \quad (\text{A.10})$$

$$0 = \sum_i \frac{E[\rho_{ij}|\hat{\boldsymbol{\theta}}^k, m_i]}{\hat{\theta}_j} - a_{ij} \quad (\text{A.11})$$

$$\hat{\theta}_j = \frac{\sum_i E[\rho_{ij}|\hat{\boldsymbol{\theta}}^k, m_i]}{\sum_i a_{ij}} \quad (\text{A.12})$$

It should be noted that if the background counts are also included in this formulation of the complete data (i.e. ordinary Poisson) then the simple formula above can no longer be found. From the above equation we now have to derive the conditional expectation of the complete data. A number of constraints can also be imposed in a similar manner to before, based on physical considerations.

$$m_i = \sum_{j=1}^{N_v} \rho_{ij} \quad (\text{A.13})$$

i.e. the counts in projection bin i due to voxel j , summed over all voxels will equal the total number of counts measured in bin i . And as previously defined earlier through geometric reasoning:

$$\hat{\rho}_{ij} = a_{ij} \hat{\theta}_j \quad (\text{A.14})$$

Through combining these two definitions a third definition can be arrived at:

$$\hat{m}_i = \sum_j a_{ij} \hat{\theta}_j \quad (\text{A.15})$$

The derivation of the conditional expectation of the complete data requires the use of Bayes theorem as stated above.

$$\Pr(\rho_{i1} \dots \rho_{iN_v} | m_i, \hat{\boldsymbol{\theta}}^k) = \frac{\Pr(m_i | \rho_{i1} \dots \rho_{iN_v}, \hat{\boldsymbol{\theta}}^k) \Pr(\rho_{i1} \dots \rho_{iN_v} | \hat{\boldsymbol{\theta}}^k)}{\Pr(m_i | \hat{\boldsymbol{\theta}}^k)} \quad (\text{A.16})$$

The first probability term on the right hand side $\Pr(m_i | \rho_{i1} \dots \rho_{iN_v}, \hat{\boldsymbol{\theta}}^k)$ can be ignored. Given a known complete dataset $\rho_{i1} \dots \rho_{iN_v}$ the measured data m_i is known in accordance to the

definition above. Therefore its probability is 1.

$$\Pr(m_i | \rho_{i1} \dots \rho_{iN_v}, \hat{\boldsymbol{\theta}}^k) = 1 \quad (\text{A.17})$$

Expressing the second and third probabilities in terms of the Poisson distribution (and assuming independence) will lead to (after some manipulation) to Eq. A.21

$$\Pr(\rho_{i1} \dots \rho_{iN_v} | \hat{\boldsymbol{\theta}}^k) = \prod_{j=1}^{N_v} \frac{\hat{\rho}_{ij}^{\rho_{ij}} \exp(-\hat{\rho}_{ij})}{\rho_{ij}!} \quad (\text{A.18})$$

$$\Pr(m_i | \hat{\boldsymbol{\theta}}^k) = \frac{\hat{m}_i^{m_i} \exp(-\hat{m}_i)}{m_i!} \quad (\text{A.19})$$

$$\Pr(\rho_{i1} \dots \rho_{iN_v} | m_i, \hat{\boldsymbol{\theta}}^k) = \frac{m_i!}{\hat{m}_i^{m_i} \exp(-\hat{m}_i)} \prod_{j=1}^{N_v} \frac{\hat{\rho}_{ij}^{\rho_{ij}} \exp(-\hat{\rho}_{ij})}{\rho_{ij}!} \quad (\text{A.20})$$

$$\Pr(\rho_{i1} \dots \rho_{iN_v} | m_i, \hat{\boldsymbol{\theta}}^k) = \frac{m_i!}{\prod_j^{N_v} \rho_{ij}!} \prod_j^{N_v} \left(\frac{\hat{\rho}_{ij}}{\hat{m}_i} \right)^{\rho_{ij}} \quad (\text{A.21})$$

We have now arrived at the probability mass function for the multivariate binomial distribution (Eq. A.2). The conditional expectation for the individual complete data random vectors can be deduced from the above equation, given the definition of the expectation of a multivariate binomial distribution defined earlier Eq.A.3.

$$E[\rho_{ij} | \sum_j \rho_{ij} = m_i, \hat{\boldsymbol{\theta}}^k] = m_i \times \Pr(\rho_{ij} | \hat{\boldsymbol{\theta}}^k) \quad (\text{A.22})$$

$$E[\rho_{ij} | \sum_j \rho_{ij} = m_i, \boldsymbol{\theta}^k] = m_i \frac{\hat{\rho}_{ij}}{\hat{m}_i} \quad (\text{A.23})$$

$$E[\rho_{ij} | \sum_j \rho_{ij} = m_i, \boldsymbol{\theta}^k] = m_i \frac{a_{ij} \hat{\theta}_j^k}{\sum_j^{N_v} a_{ij} \hat{\theta}_j^k} \quad (\text{A.24})$$

$$\hat{\theta}_j = \frac{\sum_i m_i \frac{a_{ij} \hat{\theta}_j^k}{\sum_j^{N_v} a_{ij} \hat{\theta}_j^k}}{\sum_i a_{ij}} \quad (\text{A.25})$$

$$\hat{\theta}_j = \frac{\hat{\theta}_j^k}{\sum_i a_{ij}} \sum_i a_{ij} \frac{m_i}{\sum_j^{N_v} a_{ij} \hat{\theta}_j^k} \quad (\text{A.26})$$

We have now arrived at the MLEM update formula.

Appendix B

MR Resolution Reconstruction of Noise Free Data

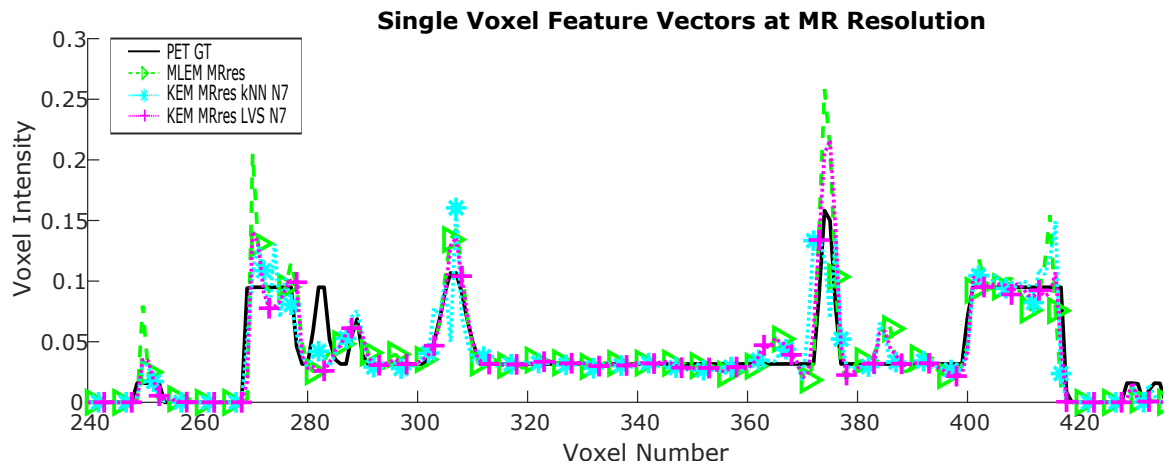


Fig. B.1 Vertical line profiles of the MR resolution reconstruction methods acting on noise free simulated data, after 500 iterations, through a tumour containing slice. Three MR resolution methods are compared for single voxel feature vectors.

Appendix C

Extended Evaluation of Simulated Dataset

For the high count (10^8) simulation studies, Fig. C.1 shows for each tumour region: the bias images, standard deviation images, the bias vs standard deviation trade-off curves, and the tumour profiles through the mean image. For the tumour profiles, the selected parameters of each method correspond to a fixed whole brain NRMSE level shown in Fig. 5.6 column 2. All error metrics in this figure are calculated from 10 noise realisations. In Fig. C.1 (bias images and bias vs standard deviation plots), the MR(only)-informed methods (KEM, KEM LVS, Gaussian and Bowsher) can be seen to increasingly bias the PET-unique regions for increasing values of β or k , in comparison to un-smoothed MLEM. Through the incorporation of PET information into the weight calculation the PET-MR-informed methods, in particular the anato-functional method achieves a reduced bias for the more intense PET-unique regions B and D (Fig. C.1, bias images), in comparison to the MR(only)-informed methods. For the bias standard deviation trade-off curves of the PET-unique regions B and D, the anato-functional method attains values similar to that of MLEM for the first five β values, the markers for which are therefore superimposed on top of one another. Consequently, the bias and standard deviation images of the anato-functional method have a

very similar pattern to that of un-smoothed MLEM for the first five β values. These β values for the anato-functional method still lead to a reduction in the whole brain region NRMSE values, explaining the favourable NRMSE trade-off curves shown in Fig. 5.6 (columns 2 and 3). Considering the MR(only)-informed methods, the KEM LVS methods provides the best bias-standard deviation trade-off curves for three out of the four PET-unique regions, although the improvement is minimal and cannot be seen in the bias or standard deviation images.

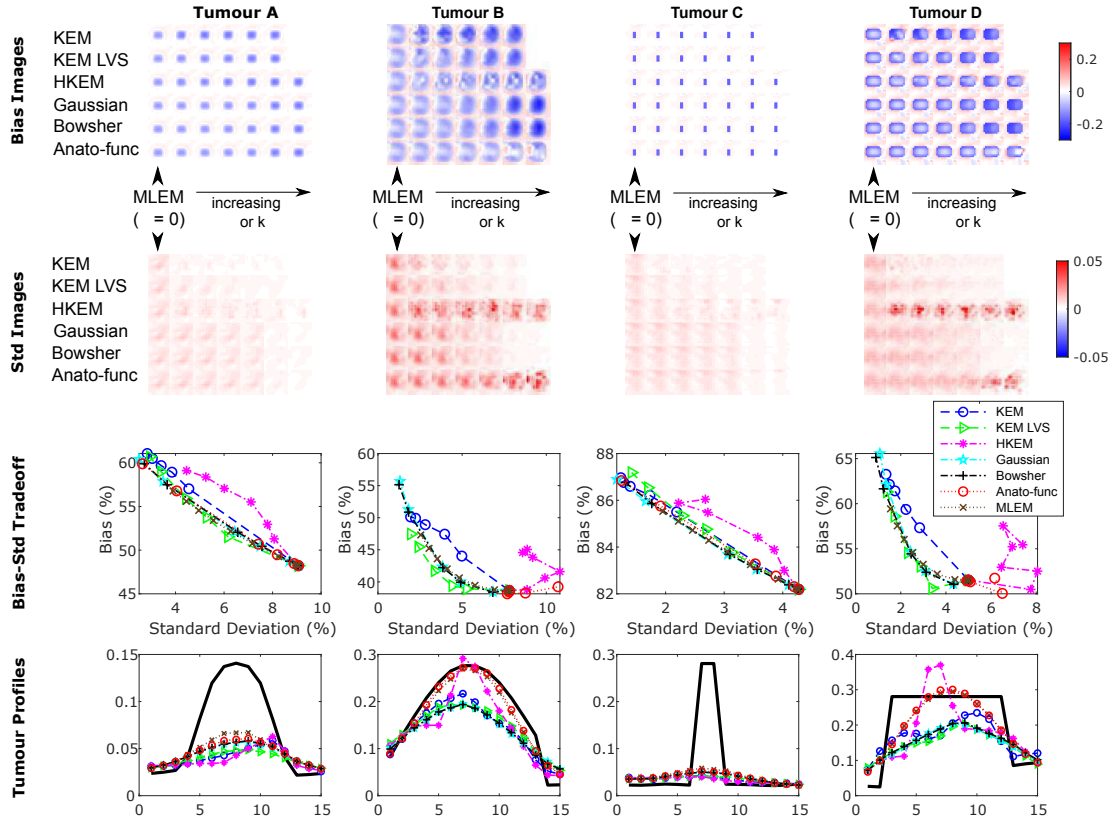


Fig. C.1 All the results shown are for the high counts (10^8) simulated dataset, using multiple noise realisations. Bias Images: bias images for each of the PET-unique regions, shown for increasing level of β (MAP) or k (KEM) for each reconstruction method. Std Images: standard deviation images for each of the PET-unique regions, shown for increasing level of β or k for each reconstruction method. Bias-Std Trade-off: Bias vs standard deviation plots for each of the PET-unique regions, for increasing levels of β (MAP) or k (KEM) along each curve. Tumour Profiles: tumour profiles through the mean image (averaged across noise realisations) of each reconstruction method (approximately fixed NRMSE), for each PET-unique region.

For the low count (10^7) simulation studies, Fig. C.2 shows for each tumour region: the bias images, standard deviation images, the bias vs standard deviation trade-off curves, and the tumour profiles through the mean image. For the tumour profiles, the selected parameters of each method correspond to a fixed whole brain NRMSE level shown in Fig. 5.9 column 2. In this figure the KEM LVS method outperforms all other compared methods in terms of the bias vs standard deviation plots for three out of the four PET unique regions (and provides

the 2nd best bias vs standard deviation trade-off for the remaining PET unique region). The tumour profiles (bottom row) show the KEM LVS and HKEM method performing similarly to unsmoothed MLEM for the high intensity tumour regions B and D, whereas all MAP based methods result in a more suppressed tumour profiles in comparison to MLEM.

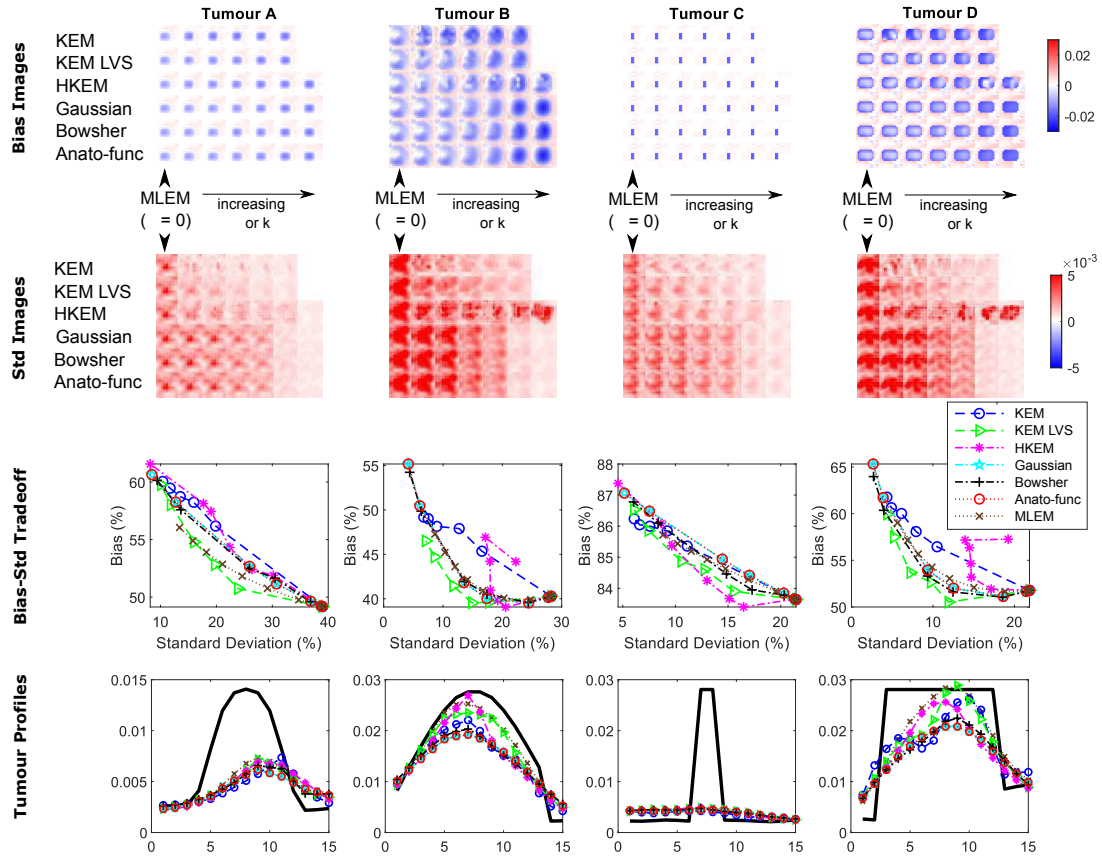


Fig. C.2 All the results shown are for the low counts (10^7) simulated dataset, using multiple noise realisations. Bias Images: bias images for each of the PET-unique regions, shown for increasing level of β (MAP) or k (KEM) for each reconstruction method. Std Images: standard deviation images for each of the PET-unique regions, shown for increasing level of β or k for each reconstruction method. Bias-Std Trade-off: bias vs standard deviation plots for each of the PET-unique regions, for increasing levels of β (MAP) or k (KEM) along each curve. Tumour profiles: tumour profiles through the mean image of each reconstruction method (the selected parameters for which correspond to an approximately fixed whole brain NRMSE), for each PET-unique region. The MLEM profile has been taken from the mean image of MLEM with post reconstruction smoothing applied (FWHM 3.5mm).