MR-Guided PET Image Restoration for Neurological Applications

Marzieh S. Tahaei

Department of Biomedical Engineering,

McGill University

A thesis submitted to McGill University in partial fulfilment of the requirements of the degree of Doctor of Philosophy

Abstract

PET and MR provide complementary information for diagnosis and treatment planning for patients as well as for studying the biological and pathological processes in the human brain. Therefore many PET studies are also coupled with MR imaging.

PET images suffer from high noise and low spatial resolution compared to T1-weighted images. The main objective of this dissertation is to develop methods that use superior anatomical information inherent in the subject's T1-weighted MR image to help reduce noise and improve spatial resolution in PET images.

To this end, we first propose a PET reconstruction method in which the PET image is re-parameterized using patch-based basis functions, learned from the subject's T1-weighted MR image. Also, a reconstruction algorithm that imposes sparsity on the parameters is proposed and used to find a sparse set of coefficients for the MR-based basis functions. The results suggest that the proposed method can provide considerable improvement in the quality of reconstructed images compared to the standard penalized likelihood framework.

Post reconstruction methods are of great interest in the analysis of PET data as they can be applied to large multicentre datasets in which raw measurement data is not available. Therefore, as the second objective of this dissertation, we develop a dual re-parameterization that restores PET images after they are reconstructed using a conventional method. This dual re-parameterization uses kernels extracted from the T1-weighted MR image without segmentation and the median filtered PET image to represent the restored image. An algorithm for estimating the coefficients of these basis functions is proposed. Moreover, a fast non-iterative method that uses guided filtering to restore PET images. The proposed methods are compared to other denoising methods and the results show improvement in the image quality in terms of the error within the brain and the bias and variance of PET-specific lesions.

The accuracy of denoising and partial volume correction methods depends on the

correct tuning of hyperparameters which are usually sensitive to the noise level and the amount of blurring in the image. For large multicentre datasets, this information may not be available. In this context, as the third objective of the thesis, a blind PET image restoration method is proposed. In the proposed framework, a deep neural network is trained on simulated PET images with different characteristics such as activity level, amount of smoothing, and the number of reconstruction iterations. Two models are trained, one with and one without a T1-weighted MR image as input. In addition, data inherent uncertainty is estimated in the proposed model. The resulting uncertainty map can be used to predict regions that may have a higher amount of error in the reconstructed image. The results of the test simulation data show significant improvement in image quality compared to the proposed guided filtering method. The application of the proposed method to real data shows promising generalizability across scanners and radiotracers.

Résumé

La tomographie par émission de positrons (TEP) et l'imagerie par résonance magnétique (IRM) fournissent des informations complémentaires pour le diagnostic et la planification du traitement des patients, ainsi que pour l'étude des processus biologiques et pathologiques du cerveau humain. Par conséquent, de nombreuses études qui se basent sur la TEP ont également recours à l'IRM.

L'objectif principal de cette thèse est de développer des méthodes qui utilisent des informations anatomiques extraites à partir de l'IRM pondérée en T1 dans le but d'aider la réduction du bruit et l'amélioration de la résolution spatiale des images TEP. Premièrement, nous proposons une méthode de reconstruction dans laquelle l'image TEP est re-paramétrée à l'aide de fonctions de base par patch préalablement entraînée sur l'image IRM pondérée en T1 du sujet. En outre, un algorithme de reconstruction qui impose des paramètres épars est proposé et utilisé pour trouver un ensemble de coefficients pour les fonctions de base en IRM. Les résultats suggèrent que la méthode proposée apporte une amélioration considérable quant à la qualité des images reconstruites par rapport à la méthode standard qui utilise la pénalité de la vraisemblance. Les méthodes de post-reconstruction sont d'un grand intérêt dans l'analyse des données TEP. Elles peuvent être appliquées à de grands ensembles de données multicentriques dans lesquels les mesures brutes ne sont pas disponibles. Par conséquent, comme deuxième objectif de cette thèse, nous développons une méthode duale de re-paramétrisation qui restaure les images TEP après leur reconstruction à l'aide d'une méthode conventionnelle. Cette re-paramétrisation duale utilise à la fois des noyaux extraits de l'image IRM pondérée en T1 sans segmentation et de l'image médiane TEP pour représenter l'image restaurée. Un algorithme d'estimation des coefficients de ces fonctions de base est proposé. En outre, une approche similaire de la re-paramétrisation duale est utilisée pour développer une méthode rapide et non itérative qui utilise un filtrage guidé pour la restauration des images TEP.

Les méthodes proposées sont comparées à d'autres méthodes de réduction de bruit. Les résultats montrent une amélioration de la qualité de l'image en termes d'erreur cérébrale ainsi que du biais et de la variance des lésions spécifiques à la TEP. La précision des méthodes de de réduction de bruit et de correction partielle du volume dépend du réglage des hyper-paramètres. Ces derniers sont généralement sensibles au niveau du bruit et à la quantité de flou présent dans l'image. Pour les grands ensembles de données multicentriques, ces informations peuvent ne pas être disponibles. Dans ce contexte, comme troisième objectif de la thèse, une méthode de restauration d'images TEP en aveugle est proposée. Dans cette méthode, un réseau est entraîné sur des images TEP simulées présentant différentes caractéristiques telles que le niveau d'activité, la quantité de lissage et le nombre d'itérations de reconstruction. Deux modèles sont générés, l'un comprenant en entrée les images IRM pondérées en T1 et l'autre non. De plus, l'incertitude inhérente aux données est estimée dans les modèles proposés. La carte d'incertitude résultante peut être utilisée pour prédire les régions susceptibles de générer une plus grande quantité d'erreur dans l'image reconstruite. Les résultats des données de simulation de test montrent une amélioration significative de la qualité de l'image. L'application de la méthode proposée à des données réelles montre des résultats prometteurs quant à la généralisabilité entre scanners et radiotraceurs.

Acknowledgements

The endeavour and kindness of the many people I have been blessed to live and work with made finishing this dissertation possible. I wish to express my gratitude to:

My supervisor Dr. Louis Collins for his invaluable guidance and his great understanding: You have been a tremendous mentor for me and I wish to thank you for encouraging my research, for allowing me to grow as a research scientist and for your invaluable advice on my research and future career.

My co-supervisor Dr. Andrew Reader for his continuous support and constructive suggestions: I have learned a lot from you and without your precise comments this work would have been of much lesser quality. I wish you the very best life has to offer you.

My advisory committee Dr. Joelle Pineau, Dr. Pedro Rosa-Neto and Dr. Brandon Xia for their guidance throughout my program.

My lab mates and my graduate fellows Azar Zandifar, Mahsa Dadar, Philip Novosad, Paul Gravel and Vladmir Fonov for the encouraging discussions and sharing fun moments in this long journey. Special thanks to Houssem Gueziri for translating my abstract to French. I also wish to thank Pina Sorrini and Nancy Abate for navigating me through the department requirements.

My parents: I owe it all to you. No words can ever describe my gratitude for your unconditional love and support and I am humble before you.

My brother and sister Maedeh and Mohammad for your long, humorous and kind phone calls.

My beloved husband Vahid for being with me every step of this experience: I am extremely fortunate to have you by my side and without your patience and encouragement I would not have been able to finish.

Finally and above all, I would like to thank God almighty for his infinite mercy and for giving me purpose in life.

This thesis is dedicated to
my parents Fatemeh and Ali,
my husband Vahid,
and to my children
Yahya and Zoha.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Positron emission tomography (PET) imaging is a powerful tool used for diagnosis, staging and treatment planning of diseases, and also understanding biological processes in the body. PET images inherently suffer from higher amounts of noise and have a relatively low resolution compared to other modalities such as anatomical MR. The amount of noise and the resolution of the reconstructed PET images depend on many factors including the radiotracer type and dose, reconstruction algorithm, the number of iterations at which the reconstruction method is stopped, the scanner model, and the acquisition protocol.

In neurological studies, often brain PET images are accompanied with T1-weighted MR images. This MR image can be from the same PET-MR scanner or it can be acquired at a different time with an MR-only scanner and then registered to the PET image.

In region-based analysis, the MR image is needed to define the region of interest (ROI) since for many radiotracers, the information inferred from the PET image might not be enough for the correct estimation of the ROI. For voxel-wise analysis, T1-weighted MR images are needed for the alignment of images of different subjects to a common space.

The availability of T1-weighted MR images in these settings together with its rich anatomical information about different tissue types have motivated the use of the subject's registered MR images for different corrections to be performed on PET measurement data such as attenuation correction and motion correction (Hofmann et al. 2009, Schreibmann et al. 2010, Ullisch et al. 2012, Catana et al. 2011).

Numerous methods have also been proposed that use the rich anatomical information in the T1-weighted MR image for partial volume correction effect (PVE) correction and noise reduction in PET images. These methods can be based on preprocessing of the sinogram data, regularization within the reconstruction and post-filtering of the reconstructed images (Stayman & Fessler 2000, Tang & Rahmim 2009, Wang & Qi 2012, Turkheimer et al. 2008, Chan et al. 2010).

The advantage of parameter estimation and modelling within reconstruction is that the noise in the raw measurement data is well modelled with a Poisson distribution and thus, parameters are estimated using a correct model of the noise. If done properly, modelling within reconstruction, results in im-ages that are better in terms of quantification (Kolthammer 2013, Zhou et al. 2009, Yan et al. 2009). In this context, the first aim of the project is to develop a regularization method for PET image reconstruction using advanced image processing techniques to reduce the noise and improve the boundary information in the resulting PET images.

On the other hand, for many PET studies, usually access to raw data is not possible and hence post reconstruction methods are preferable. Furthermore, different centres may use different reconstruction methods/parameters, which means that providing reconstructed images with only necessary corrections (scanner-dependent corrections such as attenuation and normalization) should yield better inter-centre consistency. Also, post reconstruction methods can be tuned to the needs of each study. Therefore, for multicentre studies, denoising after reconstruction is either preferred or is the only option. The second objective of the dissertation therefore, is to develop a

post reconstruction method that reduces the noise in PET images while improving the anatomical boundaries using the subject registered T1-weighted MR image.

In very large multicentre studies, accurate information about the image resolution and noise characteristics may not be available. In such cases, a blind partial volume correction and denoising methods can be useful in improving the quality of PET images. The third objective of this dissertation is to develop a method that can restore PET images without any knowledge about the radiotracer type or dose, the scanner or the number of iterations used for reconstruction. We try to achieve this challenging goal by simulating many images with different characteristics and use a machine learning framework to learn from this simulated data.

1.1 Scientific Objectives

The objective of this dissertation was to develop and validate novel methods for improving positron emission tomography (PET) image quantification using complementary anatomical information from MRI data. This was achieved by:

1. Reducing the noise in PET images within the reconstruction through novel regularization techniques. The proposed regularization method leverage from complementary anatomical information from MRI data to reduce the noise while improving the resolution of the resulting images.

2. Reducing the noise while improving the boundary information in PET images after reconstruction. The proposed framework works best in combination with conventional PVE correction methods to both improve resolution and reduce the noise in images.

3. Performing simultaneous PET partial volume effect (PVE) correction and denoising when no prior knowledge about the noise level, the radioactivity type, the dose or the scanner's point spread function (PSF) is available.

1.2 Dissertation outline

This dissertation is written in a manuscript-based style. The following summarizes the content of each chapter.

Chapter 2: Background

This chapter provides a background to the main topics that are covered in this thesis. Firstly, a brief introduction to MR physics is provided. Then PET data physics and acquisition process are introduced. Also, the derivations of the most commonly used PET reconstruction methods, both with and without regularization, are presented. These derivations are central to the first manuscript (Chapter 3). Then some of the well-known approaches for MR-Guided partial volume correction on reconstructed images are presented. This section is closely related to Chapter 4 of the dissertation in which a novel method for improving the quality of PET images is proposed. The remaining of this chapter provides an introduction to deep learning, convolutional neural networks and different regularization strategies used to avoid over-fitting in these networks. This section is related to the content of the third manuscript 3 (Chapter 5).

Chapter 3: PET reconstruction using prior image derived dictionaries

This chapter is based on a manuscript published in Physics in Medicine and Biology. In this chapter a novel reconstruction method that uses patch-based dictionaries learned from a prior image is presented. The performance of the proposed method is evaluated using simulation and real data.

Chapter 4: Two Novel PET Image Restoration Methods Guided by PET-MR Kernels

This chapter is based on a manuscript published on MR-guided PET image restoration. The two proposed post-reconstruction methods for restoring PET images using the subject's registered T1-weighted MR image is presented. The sensitivity of the

methods to change in noise level is elaborated and the performance of the method is compared to a state-of-the-art denoising method.

Chapter 5: Blind PET Image Restoration

This chapter is based on a manuscript in preparation on blind PET image restoration in which a deep learning framework for restoring PET images is proposed. The proposed method uses an Inception-Res architecture to lean restoring PET images from simulated training data. The proposed model is trained both with and without the T1-weighted MR image.

Chapter 6: Conclusion and future work

This chapter provides a discussion on the findings of the thesis and makes suggestions for the future direction of the work.

1.3 Original Contributions

The following lists the original contribution of this dissertation:

- A simplified derivation of MAP EM with quadratic penalty and positivity constraint
- Developing a re-parametrization method based on learning a patch-based dictionary from a prior image and finding coefficients using conventional MLEM
- Proposing a l_1 regularized Poisson log-likelihood to impose sparsity on the likelihood parameters. Solving this objective function using alternate direction methods of multipliers
- Imposing sparsity on the coefficients of patch-based dictionary learned from a prior image for PET reconstruction using the method proposed above
- Proposing a post-reconstruction method based on the re-parametrization of a PET image by kernel-based basis functions extracted from the median filtered PET image and the registered MR image.

- Proposing a method for finding the coefficients of these basis functions while imposing sparsity of PET-based coefficients in a least-squares objective function
- PET image restoration using MR-based and PET-based filters with twicing method
- Proposing a deep learning framework for restoring PET images with no prior information
- Incorporating aleatoric uncertainty estimation in the proposed deep learning framework to help clinicians and researchers identifying potentially erroneous regions in the restored image

1.4 Contributions of Authors

I am the first author of all the manuscripts included in this dissertation. The initial idea, development, implementation, experimentation and evaluation of the proposed methods as well as drafting the manuscript was performed by the author. The contribution of other co-authors included supervision of the research, instructive discussions and revising the manuscripts. The following list summarizes the contributions of the co-authors for each manuscript:

1.4.1 Chapter 3: Patch-based image reconstruction for PET using prior-image derived dictionaries

- **Authors**: Marzieh S Tahaei and Andrew J Reader
- **Contributions**: Proposing and developing of the method: Marzieh S. Tahaei; Implementation, testing, and validation: Marzieh S. Tahaei; Manuscript preparation and revision: Marzieh Tahaei and Andrew J. Reader

1.4.2 Chapter 4: Two novel PET image restoration methods guided by PET-MR kernels

- **Authors**: Marzieh S. Tahaei, Andrew J. Reader and D. Louis Collins

- **Contributions**: Proposing and developing of the methods: Marzieh S. Tahaei; Implementation, testing, and validation: Marzieh S. Tahaei; Manuscript preparation and revision: Marzieh Tahaei, D.Louis Collins and Andrew J. Reader

1.4.3 Chapter 5: Blind PET Image Restoration

- **Authors**: Marzieh S. Tahaei, Andrew J. Reader and D. Louis Collins

- **Contributions**: Proposing and developing of the method: Marzieh S. Tahaei; Implementation, testing, and validation: Marzieh S. Tahaei; Manuscript preparation and revision: Marzieh Tahaei, D. Louis Collins and Andrew J. Reader

1.5 Published and submitted journal articles

The following is a list of manuscripts published and submitted during my Ph.D.:

- Tahaei, Marzieh S., Andrew J. Reader, and D. Louis Collins. "Blind PET Image Restoration: A CNN with Uncertainty Estimation". In preparation for submission to IEEE Transaction in Medical Imaging.

- Tahaei, Marzieh S., Andrew J. Reader, and D. Louis Collins. "Two novel PET image restoration methods guided by PET-MR kernels: Application to brain imaging." Medical Physics 46.5 (2019): 2085-2102.

- Tahaei, Marzieh S., and Andrew J. Reader. "Patch-based image reconstruction for PET using prior-image derived dictionaries." Physics in Medicine and Biology 61.18 (2016): 6833.

1.6 Published conference proceedings and abstracts

The following is a list of accepted conference papers and abstract during my PhD:

- Marzieh S. Tahaei, Andrew J. Reader, and D. Louis Collins. "MR-Guided Blind PET Image Restoration Using Convolutional Networks", MED-NIPS workshop at NIPS 2018.

- Marzieh S. Tahaei, Andrew J. Reader, and D. Louis Collins. "MR-Guided PET Image Restoration for Neurological Applications", 2016 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC).

- Marzieh S. Tahaei, Andrew J. Reader, "Combining different variance reduction approaches for PET image reconstruction", IEEE Medical Imaging Conference (MIC), Seattle, WA (Oct. 2014).

Chapter 2

Background

This chapter provides the background information for the manuscripts presented in the dissertation and is organized as follows. It provides a brief introduction to PET and MR physics in sections 1 and 5 respectively. Sections 2 and 3 focus on PET image reconstruction and are central to the methods proposed in Chapter 3. Section 4 describes some well-known partial volume correction methods and is related to the methods proposed in Chapter 4. Section 5 and 6 introduce machine learning and deep learning respectively to provide some background information for the framework proposed in Chapter 5.

2.1 PET Physics

Positron emission tomography(PET) is a powerful in vivo molecular imaging technology used for diagnosis, therapy monitoring, disease staging and studying the human body. In PET imaging the subject under study is injected with a positron-emitting radiotracer of interest and is placed in the PET scanner field of view. PET imaging then provides a 4D spatiotemporal image of the resulting radioactivity distribution in the body. The radiotracers used in PET imaging are usually labelled compounds designed to provide quantitative measures of some biological processes such as glucose metabolism or binding potential (i.e. the affinity of the injected molecule to a

specific receptor). With new PET radiotracers being developed, PET provides endless potential for tracking compounds of interest in vivo. The following section gives a description of PET scan physics.

2.1.1 Positron emission

A radiotracer is a molecule with biologically interesting characteristics that is labeled with a radionuclide. The radionuclides used in PET imaging are positron decaying radioisotopes. These radioisotopes undergo radioactive decay: the proton in the unstable nuclei of the radioisotope is converted to a neutron while releasing a positron (the antiparticle of the electron) and a neutrino. The positron travels a few millimeters until it loses its momentum and annihilates with an electron, emitting two back-to-back 511 keV (kilo-electron volts) photons. These photons are then detected by the PET scanner.

To limit radioactivity exposure to the subjects, radionuclides used for PET imaging have relatively short half-lives. Table 1 shows some of the most widely used radionuclides in PET. Values in this table are taken from (Bailey 1996). Any molecule, once labeled with a positron-emitting radioisotope can be traced using PET. However, by far, the most common radiotracer used both in clinic and research is [^{18}F]FDG (fluorodeoxyglucose), a glucose analogue (Hess et al. 2014). FDG has applications in oncology, neurology, cardiology, and infectious diseases.

Radionuclide	Half-life (mins)	Maximum Energy (Mev)	Positron Range in water (FWHM in mm)
^{18}F	109.8	0.64	1.0
^{11}C	20.3	0.96	1.1
^{13}N	9.97	1.19	1.4
^{15}O	2.03	1.70	1.5

Table 2.1: Adapted from (Bailey 1996), properties of widely used radio-isotopes for PET imaging.

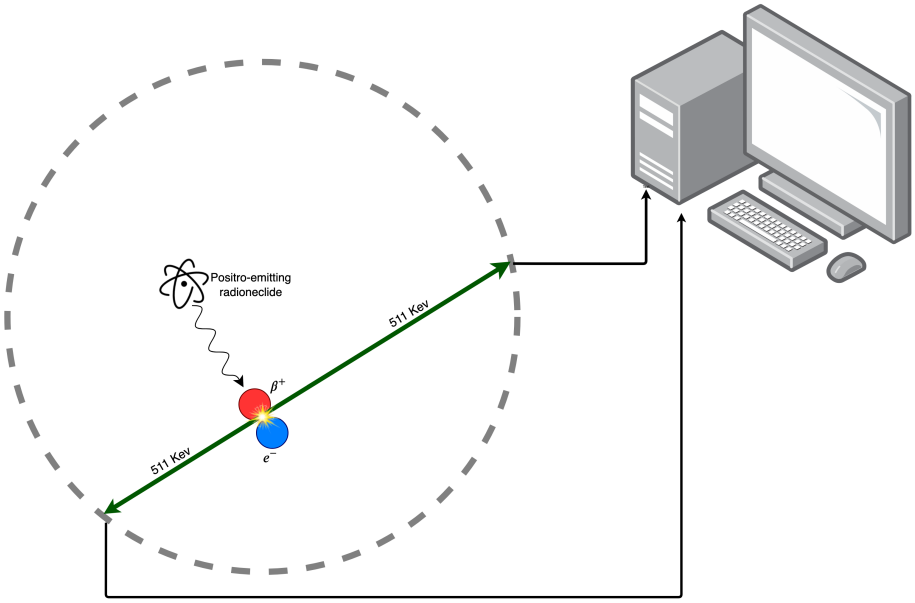


Figure 2.1: Schema of a PET acquisition process. This figure shows the annihilation event that leads to emission of two back-to-back photons. These photons are detected by the scanner and are determined as coincidence event. All of these events are stored in list-mode file format and then processed to reconstruct the image.

2.1.2 Coincidence events

Detectors in a PET scanner consist of scintillation crystals coupled to a photo-detector. When a photon resulting from positron-electron annihilation reaches the scintillation crystals, its energy is proportionally converted to visible light. This light is then detected by a photo-detector. If two of such detections occur within a short time window T, they are recorded as a coincidence event, which means they are likely photons emitted from the same positron-electron annihilation.

The line connecting the two detectors associated with a coincidence event is called the line of response (LOR). In other words, somewhere along the line of response lies the annihilation site that resulted in the two detected photons. Figure 2.1 shows an annihilation event leading to a coincidence event that is detected and stored by the

scanner's computer.

2.1.3 Storing PET data

There are two ways to store PET measurement data: sinograms and list mode. A sinogram contains the counts of events detected along each LOR. In 2D, an LOR in sinogram space is parameterized by its angle and its distance from centre. Figure2.2 shows the formation of the singoram image of a point of interest as well as the sinogram of a real brain and the corresponding reconstructed image. Figure 2.2 A-B shows that a point in image space is equivalent to a Sine function in sinogram space.

In the list-mode format, however, the coordinates of each coincidence event along with additional information such as the time of detection of each photon and its energy are recorded sequentially.

List mode offers greater retention of timing information, avoidance of binning of event attributes, faster reconstruction and more precise motion correction (Rahmim et al. 2005).

2.1.4 Corrections in PET measurement data

Not all coincidence events result from an ideal emission of a photon pair from a single positron-electron annihilation event. In fact, coincidence events recorded by a PET scanner can be divided into four different categories:

- True events: Each true event is a result of two back-to-back photons emitted from a single annihilation event. This is what we want to measure.

- Scattered events: A scattered event occurs when one or both photons emitted from the annihilation site undergo Compton scattering. If the energy loss due to scattering is large enough to be identified, these events can be discarded. Otherwise, they are mistakenly considered as true events and need to be corrected.

- Random Events: A random event occurs when two photons from two unrelated annihilation events reach the detectors in the same coincidence window and are

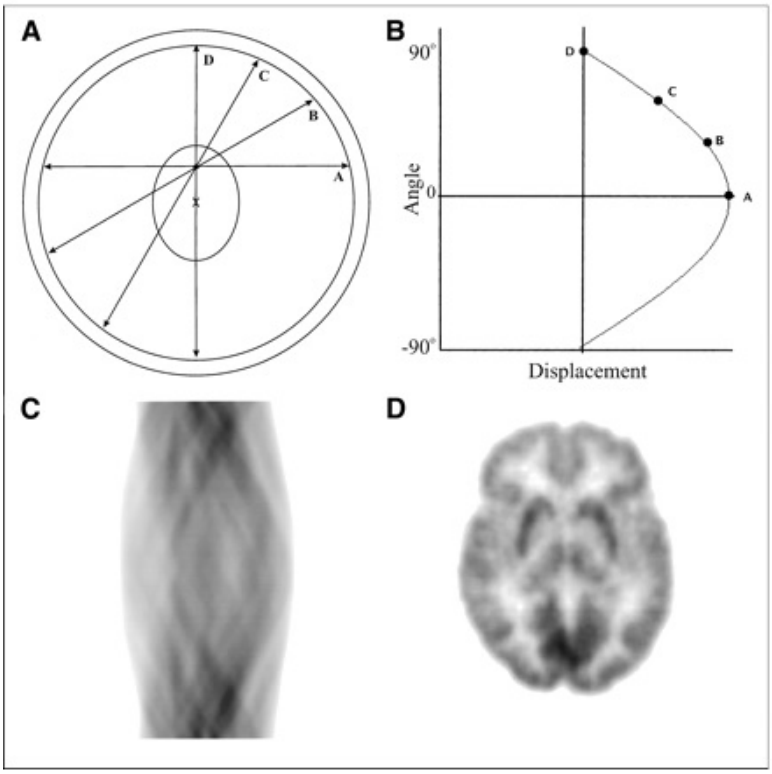


Figure 2.2: A) A point of interest in the scanner field of view and four LORs (A-D) associated with 4 annihilation events occurring at this point. The associated LORs are passing through this point. B) The sinogram image of these four detected events. In the sinogram, the orientation angle is on y-axis and displacement from the centre is on the x-axis. C) An example sinogram of a more complex object (a brain). D) The reconstructed image of the sinogram data in C. This figure is taken from (Fahey 2002). Permission for reuse of figure granted by publishers.

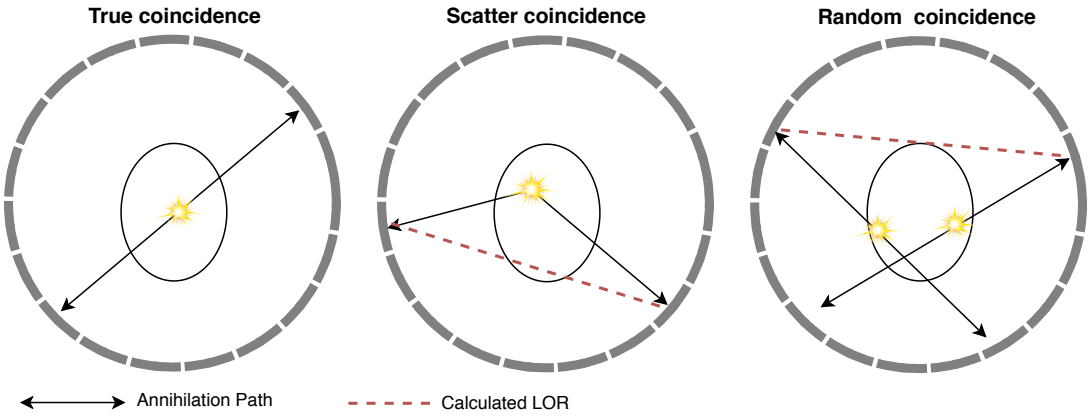


Figure 2.3: Types of coincidence events.

recorded as a single coincidence event.

Scattered and random events lead to mispositioning of the positron-electron annihilation site, which results in degradation of the reconstructed image. Figure 2.3 illustrates these different types of coincidence events.

One popular approach for scatter corrections is to fit a Gaussian function to the detected radioactivity outside the object in sinogram space (Lercher & Wienhard 1994). This approach is based on the fact that any detected activity outside the object is only due to scattering. Scatter distribution is estimated by fitting the tail of a Gaussian function to the activity outside the object. Monte Carlo simulation is another approach for scatter estimation used especially for whole-body PET data.

Correction for random events is based on single count rates (Brasse et al. 2005). The number of random events in a given LOR connecting detector 1 and detector 2 is related to the number of single counts associated with these two detectors using the following formula:

$$N_r = 2TC_1C_2 \tag{2.1}$$

where C_1 and C_2 are the number of single counts detected in detector 1 and 2 respectively and T is the coincidence timing window.

In addition to random and scattered events, there are a few undesirable effects that contribute to the error in PET measurement data. One is the positron range, which is the difference in the location of positron emission and that of annihilation. The positron range leads to a loss in the resolution of the reconstructed image. This effect, as well as other factors that contribute to loss of resolution, including detector inefficiency and tissue fraction effect can be corrected both within reconstruction using point spread function (PSF) modelling and after reconstruction using partial volume effect correction methods. Both of these approaches will be discussed in section 2.4.

Also, in PET scanning, there are variations in crystal and detector efficiencies as

well as geometric-based LOR sensitivity. These factors lead to inhomogeneous detection in the scanner field of view. The correction of these effects is called normalization. Normalization factors can be obtained from scanning a uniform radioactivity source, e.g. from a ^{68}Ge cylindrical phantom. The normalization coefficient factor for a given LOR is proportional to the inverse of the counts measured in the uniformity scan for the given LOR (Defrise et al. 1991).

Moreover, photons emitted from the subject go through the body until they hit the scanner detectors. This leads to attenuation in the resulting measurement. Depending on where the photon is emitted and what medium it passes through until it reaches the scanner, the amount of attenuation is different. It is crucial for reconstructed images to account for this effect. This correction is usually done using anatomical information obtained from a structural scan in PET-CT and PET-MR devices (Xu et al. 1991), or a transmission scan in PET-only devices (Dekemp & Nahmias 1994). In transmission scans, first, a radioactive source (usually ^{68}Ge/^{68}Ga) is placed in the scanner close to the ring and a (blank) scan is acquired. Then the same is done with the subject placed in the scanner before injection of the radiotracer (this is the so-called transmission scan). The ratio of the blank scan to the transmission scan is then used to obtain attenuation correction factors.

2.2 PET image reconstruction

2.2.1 Analytic reconstruction

In analytic methods, raw sinogram data needs to be corrected before reconstruction. The corrections are first started by estimating the scatter and random sinograms and subtracting them from the raw sinogram data. Then the attenuation correction factors that are obtained from the structural image or transmission scan are applied to the sinogram. Finally, normalization factors are applied to the resulting sinogram. The corrected sinogram is then used to produce the reconstructed image.

Analytical PET reconstruction methods are based on the central-section theorem which states that:

For a 2D function f, the 1D Fourier transform of the projection of f at an angle is equivalent to a 1-D central slice through the 2D Fourier transform of f at the same angle.

Filtered back projection (FBP) (Mersereau & Oppenheim 1974) is the analytical PET reconstruction method based on the above theorem while including a ramp filter. This method consists of the following step:

Algorithm 1: Filtered back-projection

for *each projection angle* **do**
    Take 1D Fourier transform of the parallel projections
    Multiply the result by a ramp filter
    Take the inverse of 1D Fourier transform
    Backproject the result into the image
end

FBP is simple, fast and can produce reasonably good images, especially for high count data. For these reasons FBP has been used for reconstructing PET images for many years. However, with increasing computer power, iterative reconstruction techniques, with more accurate models of the scanning process and more flexibility for reducing noise and increasing resolution, have become became popular. This category of methods is discussed in the next section.

2.2.2 Statistical reconstruction

Statistical iterative reconstruction algorithms are a class of algorithms that are developed to reconstruct PET images while incorporating an accurate model of the PET measurement data. With statistical reconstruction methods, one can model the noise statistics of the count measurements as well as physical effects in the acquisition process, within the reconstruction algorithm. These can include accounting for normalization, attenuation, scatter and random corrections within reconstruction (Hutton & Baccarne 1998), modelling the partial volume effect by incorporating point spread

function in the system matrix (Reader et al. 2003) and even estimation of kinetic parameters directly from PET data. In addition, reducing the noise can also be performed within reconstruction using regularization methods (Zhou et al. 2009, Yan et al. 2009, Matthews et al. 2010, Langs et al. 2013).

In the following subsection, a detailed derivation of the most known statistical reconstruction method for PET, namely expectation maximization image reconstruction is explained.

2.2.2.1 Poisson log-likelihood maximization

Let $x = [x_1 ... x_J]$ be the reconstructed image and $m = [m_1 ... m_I]$ be the PET measurement data.[1]. The system matrix A is a $I \times J$ matrix that models the PET measurement process. In other words, the system matrix reflects how PET measurement data is obtained from the object within the scanner, and hence should include the geometry of the tomograph as well as normalization and attenuation factors. A can be factorized as:

$$A = NLG \tag{2.2}$$

where N and L are diagonal $I \times I$ matrices modelling normalization and attenuation given by inverses of the correction factors. G is an $I \times J$ matrix that performs line integrals with respect to the scanner geometry. Each element a_{ij} in the system matrix indicates the probability that a coincidence event within voxel j is detected in projection bin i.

PET measurement data m is count data and hence follows a random Poisson distribution:

$$p(m_i|\bar{m}_i) = \frac{e^{-\bar{m}_i}\bar{m}_i{}^{m_i}}{m_i!} \tag{2.3}$$

Assuming that the system matrix A represents the PET scanner, the mean of num-

[1] In reconstruction, the object f is usually represented using discretized voxels as basis functions. However voxels are only one way of an infinite number of ways by which the object can be represented. This is called re-parameterization and will be discussed in section 2.3.3

ber of counts detected in bin i can be modelled as:

$$\bar{m_i} = \sum_{j=1}^{J} a_{ij} x_j \tag{2.4}$$

Therefore, the log-likelihood of the parameter vector x given the measurement data can be written as:

$$L(x|m) = \Pi_{i=1}^{I} \frac{e^{-\Sigma_{j=1}^{J} a_{ij} x_j} (\sum_{j=1}^{J} a_{ij} x_j)^{m_i}}{m_i!} \tag{2.5}$$

$$\ln(L(x|m)) = \sum_{i=1}^{I} (m_i ln(\sum_{j=1}^{J} a_{ij} x_j) - \sum_{j=1}^{J} a_{ij} x_j - ln(m_i!)) \tag{2.6}$$

The last term in the above equation ($ln(m_i!)$) does not depend on x and hence can be dropped when maximizing with respect to x.

In order to maximize a given likelihood, one often takes the derivative of the objective function with respect to the parameter of interest, i.e x, and sets this derivative to zero. Poisson log-likelihood however, is too complicated and setting the derivative to zero does not lead to a closed-form solution. A well-known approach to solve such problems is to use expectation maximization (EM).

2.2.2.2 Expectation Maximization

In expectation maximization (EM), instead of maximizing the log-likelihood function itself, we find its expectation with respect to a latent variable conditioned on some estimate of the parameter of interest and on the measured data, this is called ”E” step. We then try to maximize this expectation, the ”M” step and hence the name ”EM”. We continue updating the expectation and maximizing it in an iterative manner until convergence.

The first step is to design a latent variable z, where at iteration t, z^t is a function of the current estimate of parameter of interest x^t and the measurement data m and can be written as $z^t = f(m, x^t)$. The variable z should be designed in a way that knowing

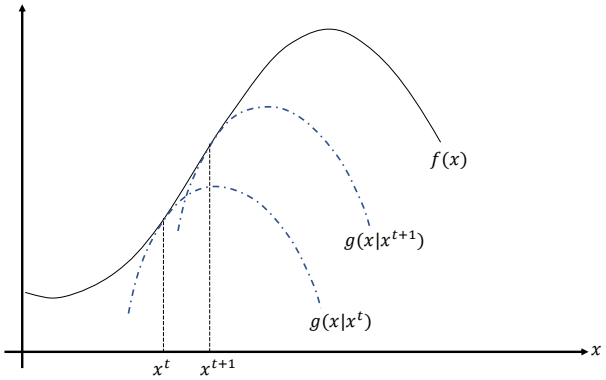


Figure 2.4: In each step the surrogate function of the objective function is estimated and then maximized leading moving the parameter of interest one step toward the optimum.

z in addition to the measurement data m would simplify the problem, ideally to a closed-form solution such as: $x^{t+1} = g(m,z)$.

But how can one estimate z when x is unknown? The solution is to start to initialize x^t with an arbitrary value and then in iterative manner estimate z^t and update x^{t+1} accordingly. Now at each step t, we first estimate the log-likelihood using z^t in the expectation step. This estimation is then maximized with respect to x and a new x^{t+1} is obtained. This step is called maximization.

$$x^{t+1} = g(m,z^t) = g(m,f(m,x^t)) \tag{2.7}$$

In fact in each step, EM maximizes the conditional log-likelihood of the latent variable given the current estimate of the parameter of interest and the measured data m.

2.2.2.3 EM from majorization-maximization point of view

Majorization-maximization (MM) or optimization transfer, is an iterative optimization method where in each step a surrogate of the original function is maximized and by doing that the estimation of the solution becomes one step closer to the optimum.

Like EM, MM consists of two steps. First a surrogate function that locally approximates the objective function around a given point x^t is estimated. The surrogate func-

tion should have two characteristics:

- Its value at x^t should be equal to the value of the objective function at x^t
- It should be smaller than the objective function elsewhere (it should lower-bound the objective function)

In the second step, called the maximization step, this function is maximized and x^t is updated accordingly. Figure 2.4 shows this process as x moves upward towards the optimum of the objective function. In fact, EM is a special case of MM algorithm in which the objective function is the log likelihood of the parameters of interest.

2.2.2.4 MLEM derivation for PET reconstruction

In order to be able to find a closed-form solution for the Poisson log-likelihood in PET reconstruction, we define $z = \{z_{ij}\}$ as follows:

z_{ij} is the number of counts in sinogram bin i that come from voxel j[2].

If z were known, m_i could easily be obtained by summing over all voxels that contributed to bin i:

$$m_i = \sum_{j=1}^{J} z_{ij} \tag{2.8}$$

$$\tag{2.9}$$

Also, the model of the mean of the latent variable and the model of the mean of measurement data can be obtained as:

$$\bar{z_{ij}} = a_{ij} x_j \tag{2.10}$$

$$\bar{m_i} = \sum_{j=1}^{J} \bar{z_{ij}} \tag{2.11}$$

Now the Poisson log-likelihood of x given z can be written as:

[2]The definition of z and thus the resulting surrogate function are based on (Reader & Verhaeghe 2014).

$$L(x|z) = \Pi_{j=1}^{J} \Pi_{i=1}^{I} \frac{e^{-a_{ij}x_j}(a_{ij}x_j)^{z_{ij}}}{z_{ij}!} \tag{2.12}$$

$$lnL(x|z) = \sum_{j=1}^{J} \sum_{i=1}^{I} -a_{ij}x_j + z_{ij} ln(a_{ij}x_j) - ln(z_{ij}!) \tag{2.13}$$

Now we take the derivative with respect to x_j and set it to zero :

$$\frac{\partial lnL(x|z)}{\partial x_j} = \sum_{i=1}^{I} [-a_{ij} + \frac{z_{ij}}{x_j}] \tag{2.14}$$

$$x_j = \frac{\sum_{i=1}^{I} z_{ij}}{\sum_{i=1}^{I} a_{ij}}$$

Knowing z will simplify the solution to the above. Now the question is, how can one estimate z in the expectation step? Since in each step of EM, an estimate of x at iteration t is known, we can use this information to provide an estimation of z. In fact the ratio of z_{ij} over m_i should be equal to the ratio of the model of the mean of z_{ij} over the model of the mean of m_i, substituting these mean values from equations 2.10 and 2.11 we have:

$$\frac{z_{ij}^t}{m_i} = \frac{a_{ij}{x_j}^t}{\sum_{j=1}^{J} \bar{z_{ij}^t}} \tag{2.15}$$

Therefore z_{ij}^t can be estimated using the following:

$$z_{ij}^t = \frac{a_{ij}{x_j}^t}{\sum_{j=1}^{J} \bar{z_{ij}}^t} m_i \tag{2.16}$$

By substituting this estimation into the solution from eq. 2.14, we obtain the following:

$$x_j^{t+1} = \frac{\sum_{i=1}^{I} \frac{a_{ij}{x_j}^t}{\sum_{j=1}^{J} \bar{z_{ij}}} m_i}{\sum_{i=1}^{I} a_{ij}} = \frac{x_j^t}{\sum_{i=1}^{I} a_{ij}} \sum_{i=1}^{I} a_{ij} \frac{m_i}{\sum_{j=1}^{J} a_{ij} x_j^t} \tag{2.17}$$

This equation is known as the EM update.

2.2.2.5 Poisson log-likelihood in terms of EM update

In (Reader & Verhaeghe 2014), the conditional expectation of the Poisson log-likelihood of z given x^t in eq. 2.13 was rearranged and rewritten as a function of the EM update:

$$G(x;x^t,m) = \sum_{j=1}^{J} s_j(x_j^{EM} ln(x_j) - x_j) \tag{2.18}$$

where $s_j = \sum_{i=1}^{I} a_{ij}$ and x^{EM} is x^{t+1} from eq. 2.18. s is called the sensitivity image.

In essence this formula can be regarded as a surrogate function of the log likelihood of the parameters given measurement data m. Hence, from the optimization transfer perspective optimising this function iteratively leads to maximization of the log likelihood of image.

This rearrangement shortens and simplifies the formulation of the log likelihood and thus will be used hereafter in our derivation of regularized EM.

2.3 Reducing the noise within reconstructions

PET image reconstruction is an ill-posed and ill-conditioned inverse problem. Hence the resulting images of applying MLEM on PET measurement data are very noisy. Therefore regularization of the objective function is often needed to control the variance in the resulting images.

In this section, an overview of very simple regularization methods are briefly presented. The content of this section are mainly from a conference record accepted in the Nuclear Science Symposium and Medical Imaging Conference in 2014 [3].

2.3.1 Early termination

Among PET reconstruction methods, MLEM is now frequently used in clinical practice due to a visually improved image quality and their superiority to conven-

[3] Tahaei, Marzieh S., and Andrew J. Reader. "Combining different variance reduction approaches for PET image reconstruction." 2014 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC). IEEE, 2014.

tional methods in lesion detection (Wells et al. 2000). However, limited count data combined with the ill-conditioned system matrix in PET can result in images with high variance.

A common practice for reducing the variance in the reconstructed images is the early termination of the MLEM algorithm. The main drawback of this method is that given the non-linearity of the EM algorithm, different pixels converge at different rates and thus the resulting image suffers from spatially variant characteristics.

2.3.2 MAPEM

The *maximum a posteriori* (MAP) probability estimation is a Bayesian framework in which different prior functions can be used to maximize the posterior probability distribution. MAP enables faster convergence by reducing the ill-conditioning of the PET reconstruction problem. More importantly, through the use of a variety of penalty functions, MAP provides images with the desired characteristics. Although similar to early termination, if not run until convergence MAP also suffers from spatially variant characteristics. The log MAP formulation for PET reconstruction can be written as:

$$ln(L_{a\,posteriori}(x|m)) = ln(L(x|m)) + ln(Prior(x)), \tag{2.19}$$

Using the Gibbs distribution $\frac{1}{Z}\exp(-\beta U(x))$ as the prior probability of image x, the MAP formulation can be simplified to the following:

$$ln(L_{a\,posteriori}(x|m)) = ln(L(x|m)) - \beta U(x) \tag{2.20}$$

where β is the regularization hyperparameter and controls the relative importance of the penalty term and the constant Z has been discarded since it has no impact on the optimization. For the purpose of this section, a weighted quadratic function is used as

$U(x)$ to penalize large differences in the intensities between the neighbouring pixels.

$$U(x) = \sum_{j=1}^{J} \sum_{k \in N(j)} w_{jk}(x_j - x_k)^2. \tag{2.21}$$

$\{N(j)\}$ is the set of k nearest neighbours of pixel j.

In an optimization transfer framework, one can view the EM algorithm as maximizing the following surrogate function for $ln(L(x|m))$:

$$G(x;x^t) = \sum_{j=1}^{J} s_j(x_j^{EM} ln(x_j) - x_j), \tag{2.22}$$

where $s_j = \sum_{i=1}^{I} a_{ij}$ is the j'th pixel of the sensitivity image and x_j^{EM} is derived from EM update and is a function of x^t, i.e., the current x.

Now, using De Pierro's decoupling rule (De Pierro 1995), the following surrogate function also reported in (Wang & Qi 2012) is obtained for $U(x)$:

$$-\beta \sum_{j=1}^{J} \sum_{k \in Neib(j)} 2w_{jk}((x_j - \frac{x_j^n + x_k^n}{2})^2 + (x_k - \frac{x_j^n + x_k^n}{2})^2).$$

We borrow the following inequality from (Wang & Qi 2012) and use it to obtain a surrogate function for the penalty term.

$$(x_j - x_k)^2 \leq 2(x_j - \frac{x_j^n + x_k^n}{2})^2 + 2(x_k - \frac{x_j^n + x_k^n}{2})^2 \tag{2.23}$$

Using 2.23, the surrogate function for the regularization term becomes:

$$H(x;x^n) = \sum_{j=1}^{J} \sum_{k \in Neib(j)} w_{jk}(2(x_j - \frac{x_j^n + x_k^n}{2})^2 + 2(x_k - \frac{x_j^n + x_k^n}{2})^2) \tag{2.24}$$

Then the surrogate function for MAP is obtained by adding the surrogates for the log

likelihood and the penalty function.

$$Q(x;x^n,m) = G(x;x^n,m) - \beta H(x;x^n) \tag{2.25}$$

Imposing a non-negativity constraint on x, the Lagrangian of this constrained objective function can be written as:

$$\begin{aligned} Maximize \quad & G(x;x^n,m) - \beta H(x;x^n) + \lambda_j x_j \\ Subject\ to \quad & \lambda_j \geq 0, \quad x_j \geq 0 \quad \forall j \\ & \lambda_j x_j = 0 \quad \forall j \end{aligned} \tag{2.26}$$

where λ is the Lagrange multiplier. We now solve this constrained optimization problem using Karush–Kuhn–Tucker conditions (KKT) (Kuhn & Tucker 1951, Karush 1939). Taking the derivative with respect to x_j, the KKT conditions lead to solving the following equations:

$$\frac{\partial G(x;x^n,m)}{\partial x_j} - \beta \frac{\partial H(x;x^n)}{\partial x_j} + \lambda_j = 0 \tag{2.27}$$

$$\begin{aligned} Subject\ to \quad & \lambda_j \geq 0, \quad x_j \geq 0 \quad \forall j \\ & \lambda_j x_j = 0 \quad \forall j \end{aligned} \tag{2.28}$$

The derivatives are calculated as follows:

$$\begin{aligned} \frac{\partial H(x;x^n)}{\partial x_j} &= \sum_{k \in Neib(j)} 4 w_{jk} (x_j - \tfrac{x_j^n + x_k^n}{2}) + \sum_{k \in Neib(j)} 4 w_{jk} (x_j - \tfrac{x_j^n + x_k^n}{2}) \\ &= \sum_{k \in Neib(j)} 8 w_{jk} (x_j - \tfrac{x_j^n + x_k^n}{2}) \end{aligned} \tag{2.29}$$

$$\frac{\partial G(x;x^n,m)}{\partial x_j} = \sum_{j=1}^{J} s_j (\frac{x_j^{EM}}{x_j} - 1) \tag{2.30}$$

$$\frac{s_j x_j^{EM}}{x_j} - s_j - \beta \frac{\partial H(x, x_k)}{\partial x_j} + \lambda_j = 0 \tag{2.31}$$

We multiply both sides by x_j. Considering the third condition in KKT formulation($\lambda_j x_j = 0$), we have

$$\begin{aligned} & s_j x_j^{EM} - s_j x_j - x_j \beta \frac{\partial H(x, x_k)}{x_j} \\ = & s_j x_j^{EM} - s_j x_j - \beta x_j^2 \textstyle\sum_{k \in Neib(j)} w_{jk} + x_j \frac{\beta}{2} \sum_{k \in Neib(j)} w_{kj}(x_j^n + x_k^n) = 0 \end{aligned} \tag{2.32}$$

Rearranging the above formula, we have:

$$x_j^2(\beta \sum_{k \in Neib(j)} w_{jk}) + x_j(s_j - \frac{\beta}{2} \sum_{k \in Neib(j)} w_{kj}(x_j^n + x_k^n)) - s_j x_j^{EM} = 0 \tag{2.33}$$

The equation has two solutions, of which only one is positive (and therefore satisfies the primary feasibility condition in the KKT formulation). Hence, the update rule becomes:

$$\begin{aligned} x_j^{n+1} &= \frac{2 s_j x_j^{EM}}{B + \sqrt{B^2 + 4\beta s_j x_j^{EM} \sum_{k \in Neib(j)} w_{jk}}}, \\ \text{where} \qquad B &= s_j - \frac{\beta}{2} \sum_{k \in Neib(j)} w_{kj}(x_j^n + x_k^n) \end{aligned} \tag{2.34}$$

2.3.3 Reparameterization

In reparameterization, in preference to the trivial voxel representation of the object, non-trivial basis functions like truncated Gaussians, blobs, B-splines etc., are used to represent the object (Lewitt 1992, 1990, Jacobs et al. 1998, Wang & Qi 2015, Jiao et al. 2015). Hence, the reconstruction method aims to find the coefficients for these basis functions.

An object in the scanner can be represented by a linear superposition of a set of

basis functions:

$$f(r) = \sum_h \alpha_h B_h(r), \tag{2.35}$$

where f is the object and α_h is the coefficient for the h'th basis B_h function. One very simple example of possible basis functions are discrete truncated Gaussian functions, where the discretized object can be approximately represented as:

$$f_j = \sum_{h \in Neib(j)} \alpha_h N_{truncated}(j; h, \sigma^2), \tag{2.36}$$

$$N_{truncated}(j, h, \sigma^2) = \begin{cases} N(j; h, \sigma^2) & \text{if } h \in Neib(j) \\ 0 & \text{otherwise} \end{cases}$$

where $N(j; h, \sigma^2)$ is a discrete Gaussian distribution function with mean h and variance σ^2. The use of overlapping Gaussians as basis functions while imposing non-negativity on the coefficients result in smoothed images.

2.4 PVE correction

The partial volume effect is the degradation of spatial resolution in medical images leading to cross-contamination between nearby regions with different activities. In PET images PVE is an important limiting factor in accurate quantification of PET radiotracer distribution. Two main factors that degrade the resolution (Erlandsson et al. 2012):

- The limited resolution of the PET scanner related to the image acquisition process, the scanner design and reconstruction process.

- The continuous radioactivity distribution is represented using a grid of voxels.

Each voxel value is equal to the mean uptake value within that voxel. For example when the voxel lies on the boundary of two different tissues, the heterogeneity in the radiotracer uptake in that single voxel is lost. This is referred to as tissue fraction effect.

Most partial volume correction methods try to resolve the loss of resolution imposed by the former which results from the acquisition and the reconstruction process. For a given region of interest (ROI), if the surrounding tissue has a lower radioactivity, PVE is manifested as a spill-out of radioactivity from the ROI to the surrounding tissue leading the under-estimation of the activity within the ROI. On the other hand, if the surrounding tissue has a higher activity, this activity may spill-in to ROI causing over-estimation of the activity in ROI.

The loss of resolution in PET images is usually characterized by a point spread function (PSF). The PSF is the image of a point source in the scanner field of view and is usually modelled using a Gaussian function. The PSF of a given scanner is spatially variant with increasing width transaxially towards the detector ring (Moses 2011). However, for brain studies, due to the small size of the brain, usually a spatially-invariant PSF is sufficient to describe the loss of resolution in the image.

For neurological applications, T1-weighted MR images with high-resolution anatomical details and a good soft tissue contrast have been used for partial volume correction for decades. With simultaneous PET/MR scanners becoming more available, more precise alignment of PET and MR images become possible which make use of MR-guided PVE correction more appealing. Unfortunately, as of today, PVE in PET images is usually overlooked and is not routinely corrected in research and clinical practice.

PVE correction methods can be applied both within and after reconstruction. In the following some of most known PVE corrections are briefly explained.

2.4.1 Post reconstruction PVE correction

2.4.1.1 Deblurring

The aim of PVE correction is to restore the resolution by reversing the effect of the PSF (Erlandsson et al. 2012). The most trivial approach for post-reconstruction PVE correction is deblurring. Assuming that the PVE has a similar effect of a applying a convolution on the PET image, Richardson-Lucy (Richardson 1972, Lucy 1974) and Van-Cittert deconvolution (van Cittert 1931) methods can be used to correct for partial volume effect.

The problem with these approaches is that the ill-conditioned nature of deconvolution leads to noise amplification in resulting images. To control noise amplification, iterative methods are usually stopped before convergence which results in incomplete recovery of PVE.

2.4.1.2 Müller-Gärtner

Müller-Gärtner (Müller-Gärtner et al. 1992) is a method for correcting PVE in a single region, typically the grey matter. First the brain T1-weighted MR image is segmented into grey matter (GM), white matter (WM) and cerebro spinal fluid (CSF). The CSF is assumed to have no radioactivity. Then a mask for WM is obtained by setting the voxels in the WM to the mean activity in WM and zero elsewhere. The mean activity in WM is estimated from an erroded WM mask. The resulting WM image is convolved with the PSF and subtracted from the original image. A binary GM mask is also obtained by setting the values in grey matter to 1 and zero elsewhere. The GM binary mask is also convolved with the PSF.

$$f_{GM}(x) = \frac{f(x) - [C_{WM}(x)M_{WM}] \circledast h(x)}{C_{GM} \circledast h(x)} \tag{2.37}$$

where x represents spatial location and f_{GM} is the PVE corrected Gray matter image, C_{WM} is the WM binary mask, C_{GM} is the GM binary mask. M_{WM} is the mean

activity in WM and $h(x)$ is the PSF.

2.4.1.3 Geometric transfer matrix (GTM)

Rousset et al. (1998), developed a method to account for partial volume effects in multiple regions. They begin by segmenting an MR image into n regions $R = [r_1, ..., r_n]$. Assuming that the activity in each region is homogeneous, the partial volume effect can be modelled by multiplying a Matrix $W \in \mathbb{R}^{n \times n}$ by the regional mean values of the ground truth image.

$$R^{\text{PVE}} = WR^{\text{truth}} \tag{2.38}$$

Where each element in matrix W (W_{ij}) represents the fractional contribution (or contamination) of each region r_j into region r_i. W is computed by convolving the mask of each region with the PSF function.

The true mean values, thus can be obtained multiplying the inverse of W with the mean values obtained from the image.

$$\hat{R} = G^{-1}R^{\text{PVE}} \tag{2.39}$$

Since this method assumes homogeneous activity in each region, the end result is the mean uptake values of the predefined regions and not a voxel-based image.

2.4.1.4 Region-based voxel-wise correction

In the region-based voxel-wise correction method (Yang et al. 1996), for an image x, first the GTM method is used to find the mean value of n regions, then these mean values are assigned to each region resulting in an image with constant values in each region:

$$t = \sum_{i=1}^{n} \hat{R}_i P_i \tag{2.40}$$

where P_i is a mask image in which the voxels in the i-th region are equal to 1 and 0 elsewhere.

The resulting image along with the PSF is used to obtain the corrected voxel based image using the following:

$$\hat{x} = x\frac{t}{t \circledast h} \tag{2.41}$$

The result is an entire voxel-based PVE corrected image with within tissue variability. This method is well suited for voxel-based analysis.

2.4.2 Within reconstruction PSF Modelling

A distinct approach to reduce the PVE in the reconstructed image is to model PSF in the reconstruction algorithm. This is done by including a model of the PSF in image space in the system matrix :

$$A = NLGH \tag{2.42}$$

where N, L and G are the matrices defined for eq. 2.2. H is a $J \times J$ matrix in which each column k is a Gaussian function modelling the PSF at voxel k (Reader et al. 2003).

2.5 Magnetic resonance imaging

Magnetic resonance imaging (MRI) is a non-invasive medical imaging technique used for the quantification of human anatomy and function. MRI does not expose the subject to ionizing radiation and through different settings of radio-frequency pulses and gradients, an MR scanner is able to produce images of different contrasts.

2.5.1 MRI Physics

MR is based on nuclear magnetic resonance (NMR) phenomenon which states that nuclei, in the presence of a strong magnetic field, produce electromagnetic signals when the strong field is perturbed by a weaker magnetic field.

Protons in hydrogen atoms that are by far the most abundant atom in the living

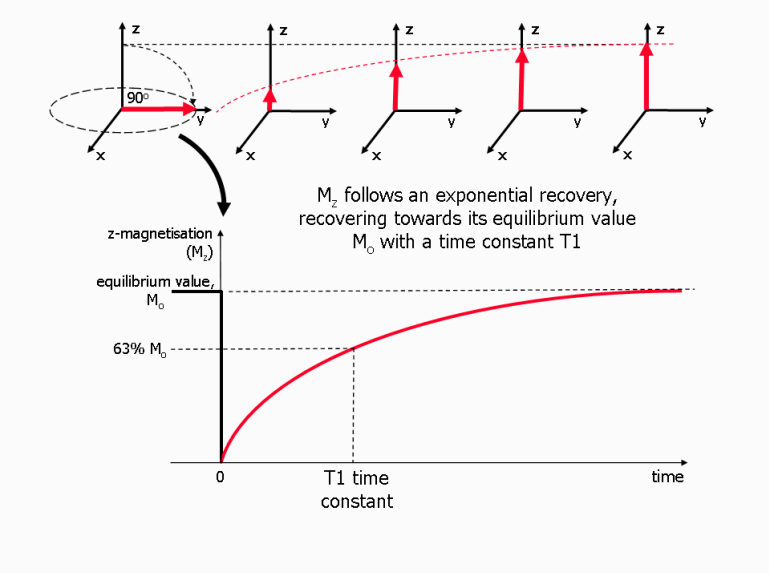


Figure 2.5: Illustration of T1 relaxation: When an RF pulse is applied to the net magnetization, it changes it from equilibrium, i.e. parallel with B0, to a 90 degree angle with B0 which in turn sets the longitudinal component of the net magnetic signal to zero. Once the RF pulse is stopped, the net magnetization recovers back to its equilibrium exponentially at the time constant T1. This figure is taken from (Ridgway 2010). Permission for reuse of figure granted by publishers.

organisms possess an intrinsic property called spin which results in a small magnetic field for each proton. Spins of protons are randomly oriented and hence the net magnetic field of the body is zero. An MR scanner provides a strong magnetic field B_0 in the scanner field of view. When the subject is placed in this magnetic field, the protons in his or her body become parallel/anti-parallel with the magnetic field B_0 and reach a new equilibrium. In this equilibrium state, the number of parallel spins is slightly higher than the anti-parallel ones, resulting in a non-zero net magnetic field proportional to the B_0 strength. This process is called magnetization. The strength of B_0 has a direct effect on the size of net magnetization and hence the signal detected by the scanner.

During image acquisition, transmitter coils in the scanner generate a radiofrequency (RF) pulse perpendicular to B_0 which moves the direction of spins away from the di-

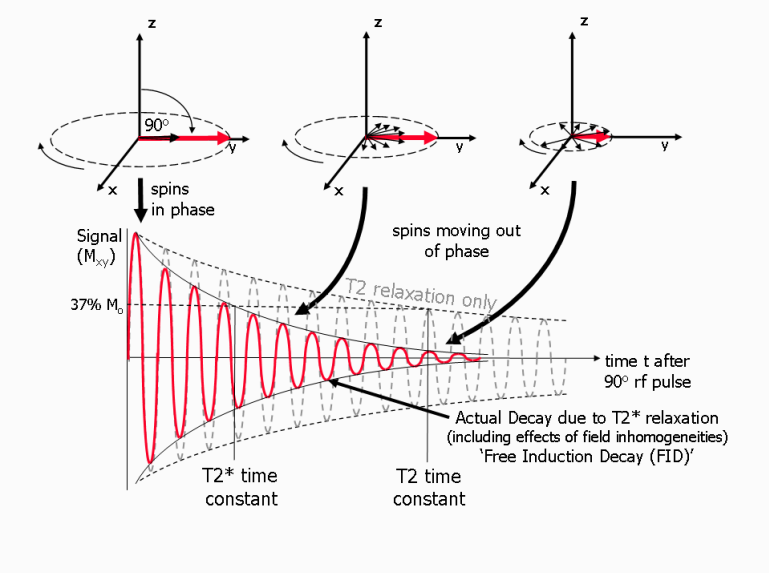


Figure 2.6: Illustration of T2 relaxation: When an RF pulse is applied to the net magnetization, it changes it from equilibrium, i.e. parallel with B0, to a 90 degree angle with B0. when the RF pulse is stopped the transverse component of excited protons decays to zero. The resulting signal is called free induction decay (FID). The decay time of FID is called T2*. T2* is smaller than T2 since it is also affected by de-phasing due to magnetic field inhomogeneities. T2 can be obtained by removing The effect of these additional de-phasing by application of a 180^0 refocusing pulse. This figure is taken from (Ridgway 2010). Permission for reuse of figure granted by publishers.

rection of magnetization by an angle called the *flip angle*. The frequency at which RF is generated is called Larmor or resonant frequency ω_0 and is determined by:

$$\omega_0 = \gamma B_0 \tag{2.43}$$

where γ is the constant gyromagnetic ratio equal to 42.6 MHz/tesla for protons.

Longer or larger RF pulses lead to a larger flip angles, and this process is called *excitation*. The excited net magnetization has two components: one component is parallel to B_0, known as the longitudinal component and the other component is perpendicular to B_0 and is known as the transverse component. The transverse component rotates

at the Larmor frequency leading to a small current in the receiver coils in the scanner.

When the RF signal is turned off, the protons spins return to their equilibrium state parallel to B_0 in a process known as *relaxation*.

There are two relaxation processes, each related to one of the mentioned components. One is the recovery of the longitudinal component of magnetization. This process occurs exponentially with a time constant named the spin-lattice relaxation time (T1) and creates a small current in the scanner. See Fig. 2.5.

The other relaxation process is associated with the transverse component. It is the decay of the transverse component while rotating around B_0 and producing a decaying oscillating magnetic field know as free induction decay (FID). FID is also detected by the scanner receiving coils. The decay of the transverse component is exponential with a time constant known as T2. Figure 2.6 shows this process.

2.5.2 Spatial Localization

The MR signal received by the scanner comes from the entire body. For the spatial localization of the MR signal, gradient coils in the scanner are used to apply three magnetic field gradients in the x, y and z directions. Slice selection is the first step of MR signal localization and is achieved by applying a gradient in the z-direction. Then, for a given slice in the z-direction, the particular frequency associated with the spins in this slice is chosen and an RF pulse is used to excite protons only in the given slice.

Once the particular slice is selected, then phase-encoding and frequency encoding are used to localize pixels within the slice. In phase-encoding a gradient is applied in the vertical direction right after the RF pulse. This gradient changes the frequency of the protons during relaxation, leading to a phase shift along the Y direction. Once the gradient is turned off proton spins in the same row will have the same phase.

For localization along the horizontal axis, frequency encoding is used by applying a gradient in the horizontal direction while recording RF. This changes the Larmor frequency of the protons along the horizontal direction.

2.5.3 Image formation

During frequency encoding, the recorded signal is used to fill a matrix named k-space. Each point in k-space is the sum of the RF signal for a particular gradient field. The inverse Fourier transform of K-space gives the reconstructed MR image.

Different combinations of pulse and gradient fields (called sequences) like spin echo and gradient echo will result in images of different contrast such as T1-weighted, T2-weighted, proton density and FLAIR. Also MR contrast agents can be used to increase the contrast of the images even further. The most common compound used as contrast agent in MR imaging in Gadolinium.

2.6 Machine learning

Machine learning is the framework that tries to solve different computational problems by learning model parameters from data. In machine learning, once a model is designed for the task in hand, it uses training data to estimate (learn) the optimum parameters. In supervised machine learning, which is the focus of this section, the training set consists of a paired set of input observation x and their corresponding target output y. The model is trained to find the relationship between x and y by finding parameters that lead to the minimization of a loss function on the training set (i.e training loss). Once trained, we expect the model to make accurate predictions on the previously unseen input samples called the test data. This prediction using a trained model is called inference. The ability of the model to accurately predict the target for the test data set is called generalization.

2.6.1 Dictionary learning

In dictionary learning (Aharon et al. 2006), the data vector is represented using a superposition of a number of basis vectors called atoms. These atoms form a dictionary and are usually learned from the data itself. When the dictionary size (i.e., the number of atoms) is larger than the dimensionality of the data, the dictionary is

called over-complete. In an over-complete dictionary, the atoms are not orthogonal and hence allow multiple representations of the same object (i.e., non-uniqueness).

When solving inverse problems in an overcomplete basis space, a sparsity constraint is often imposed on the coefficients of atoms to force a unique solution. This means that among multiple solutions, the one that has the fewest number of non-zero coefficients is selected. This is also referred to as sparse coding.

The combination of over-completeness with sparsity has shown to provide flexible representation that is robust to noise in different domains such as computer vision and speech.

Mathematicaly speaking, let us assume that n data points of size L are stored in the matrix $X \in \mathbb{R}^{n \times L}$. An example of these data points is the set of overlapping patches extracted from an image. Each data point x can then be sparsely represented as a linear combination of these dictionary atoms:

$$x = \sum_{k=1}^{K} \alpha_k D_k \tag{2.44}$$

where α_k is the coefficient for the k'th atom. The goal is to learn a dictionary $D \in \mathbb{L} \times \mathbb{K}$ that is composed of K atoms while imposing sparsity on the coefficient vector α. This problem can therefore be formulated as:

$$\arg\min_{\boldsymbol{\alpha}_k} \quad \|\mathbf{X} - \mathbf{D}\boldsymbol{\alpha}\|_F^2 + \lambda \|\alpha_i\|_0 , \tag{2.45}$$

where F is the Frobenius norm and λ is the hyper-parameter controlling sparsity. The above minimization problem is Np-hard. Therefore, approximate solutions are usually considered. K-SVD (Aharon et al. 2006) is the most well-known approach to solve this problem. For image processing application, images are often decomposed to over-lapping patches and the dictionary is learned from them to represent each patch in the image.

2.6.2 Neural Networks

A neural network is a computational tool roughly modelled after the neurobiology of the brain. In a neural network, a neuron is the basic computational unit which takes an input vector x and outputs a signal. The output is a nonlinear function (also known as the activation function) of the affine transformation of input: $F(Wx+b)$. A neural network or Multilayer perceptron (MLP) (Rumelhart et al. 1985) is simply a network of neurons arranged in layers. Each layer contains a number of neurons and the output of each neuron is fed as input to neurons in the next layer. The output of neuron i in layer l therefore can be written as: s

$$y_{il} = F(W_{il}y_{l-1} + b_{il}) \tag{2.46}$$

where y_{l-1} is the output of neurons in layer $l-1$. W_{il} is the weight vector connecting the neurons in layer l-1 to neuron i in layer l and b_{il} is the bias for neuron i in layer l. Fig.2.7 shows an example architecture for a 3-layer neural network.

To solve a machine learning problem using MLP, the first step is to design an architecture. An architecture of an MLP specifies the number of layers, the number of neurons in each layer and the activation function of each neuron. Once the architecture is specified, a predefined objective function is minimized on training data to learn the weights and biases in the network.

An example of a simple objective is the mean squared error. To increase the generalization power of the model a regularization term also known as weight decay (Krogh & Hertz 1992) parameter is added to the objective function. The following equation shows the objective function with the regularization term for the network in Fig.2.7:

$$J(W,b) = \left[\frac{1}{N}\sum_{i=1}^{N} J(W,b;x^{(i)},y^{(i)})\right] + \frac{\lambda}{2}\sum_{l=1}^{n_l-1}\sum_{i=1}^{s_l}\sum_{j=1}^{s_{l+1}}\left(W_{ji}^{(l)}\right)^2 \tag{2.47}$$

$$J(W,b;x,y) = \frac{1}{2}\left\|h_{W,b}(x) - y\right\|^2$$

Where s_l is the number of units in layer l and n_l is the number of layers in the network. $W_{ji}^{(l)}$ is the weight associated with the connection between unit j in layer l , and unit i in layer $l+1$. The first term is the mean squared error and the second term is the weight decay regularization. The weight decay parameter λ controls the relative importance of the two terms.

During training, usually the objective function is minimized iteratively using a gradient descent approach. To calculate the gradient of the loss function with respect to the weights in the MLP, backpropagation algorithm (Rumelhart et al. 1985) is used. In back-propagation first the loss function on the training data is computed and then chain rule is used to determine the partial derivative of the loss with respect to network parameters in each layer in a recursive manner.

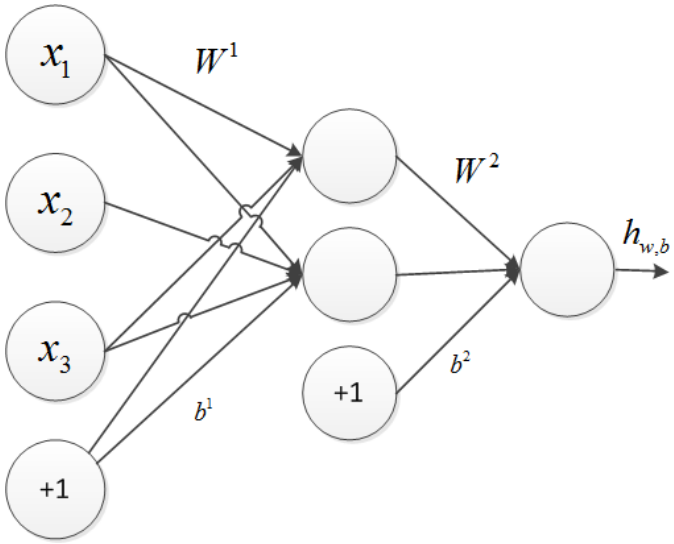


Figure 2.7: A 3-layer MLP architecture.

According to gradient descent algorithm in each iteration the weights and biases of all layers are updated using the following rules:

$$\begin{aligned} W_{ij}^{(l)} &= W_{ij}^{(l)} - \alpha \frac{\partial}{\partial W_{ij}^{(l)}} J(W,b), \\ b_i^{(l)} &= b_i^{(l)} - \alpha \frac{\partial}{\partial b_i^{(l)}} J(W,b), \end{aligned} \tag{2.48}$$

Where α is the learning rate. Computing the gradient of objective function leads to the following equations:

$$\begin{aligned} \frac{\partial}{\partial W_{ij}^{(l)}} J(W,b;x,y) &= a_j^{(l)}.\delta_i^{(l+1)}, \\ \frac{\partial}{\partial b_i^{(l)}} J(W,b;x,y) &= \delta_i^{(l+1)}, \end{aligned} \tag{2.49}$$

$\delta_i^{(l)}$ indicates the contribution of unit i in layer l to the error in the output layer and is computed as follows:

$$\delta_i^{(l)} = \left(\sum_{j=1}^{s_{l+1}} W_{ji}^{(l)} \delta_j^{(l+1)} \right) f'(z_i^{(l)}), \tag{2.50}$$

where $z_i^{(l)}$ is the weighted sum of inputs to unit i in layer l.

In each step of the gradient descent algorithm, the gradient is computed on the entire training data (batch) which is computationally expensive. In practice however, in each step the gradient is estimated and updated using a randomly selected smaller portion of training data called mini-batch. This method is known as stochastic gradient descent. In training, an epoch is one pass through the training data in small mini-batches.

2.7 Deep Learning

The conventional MLP architectures were often shallow with one or two hidden layers (i.e layers between the input and output layers). In the last decade, with the introduction of Deep Belief Networks (Hinton et al. 2006), training deeper models became possible. Since then, Deep learning (Bengio 2009, Bengio et al. 2007) has become an active area of research where new deep architectures and algorithms for training them are being developed.

Deep learning uses a hierarchical composition of a nonlinear transformation of data to learn more abstract features. Recent studies demonstrate the superiority of deep architectures over other machine learning techniques on a variety of tasks, e.g., computer vision (Boureau et al. 2010), speech recognition (Collobert & Weston 2008) and medical imaging (Litjens et al. 2017, Havaei et al. 2017).

One of the most important characteristics of deep learning is its compactness: a function that can be compactly (with less degree of freedom) represented by a deep architecture, may need an exponential number of components when represented by a shallow architecture.

Before this decade, researchers had little success in training deep architectures. This was due to the fact that finding the parameters of a deep architecture is a highly non-convex optimization problem. Gradient-based approaches like backpropagation used to solve this problem were often stuck in local minima. Moreover, these methods were facing the problem of vanishing gradients, i.e., the gradients that are propagated backward would rapidly diminish as the depth of the network increases and hence the gradients have a very small influence in changing the weight in the beginning layers.

2.7.1 The *deep* revolution

In 2006, Hinton et al. (2006) proposed a greedy layer-wise training for *deep* feed-forward networks that addressed the vanishing gradient problem. In their proposed method, generative models (restricted Boltzmann machines) were used as the layers of the network. They pre-trained the network one layer at a time starting from the first layer in an unsupervised manner. This led to a good initialization of the network parameters. Then they used the highest level of the stacked Boltzmann machines as an input to a supervised learning algorithm. This would enable the deep network to do a supervised learning tasks. The parameters of all the layers are then fine tuned using a gradient descent approach. In the fine tuning step all the layers are treated as a single model. Note that, for the pre-training step, unlabelled data can also be utilized which

is very beneficial since obtaining labelled data is often expensive. Pre-training with unlabelled data would lead to better local minima than random initialization used in conventional neural networks (Bengio et al. 2007).

Another well-known variant of deep networks with layer-wise greedy training described above, was proposed soon after by Bengio et al. (2007). This model used sparse denoising auto-encoders as building blocks of the network instead of restricted Boltzmann machines. An auto-encoder (AE) (Bourlard & Kamp 1988) is a one-hidden-layer neural network that first encodes the input vector into a hidden representation and then the hidden representation is mapped back to the input vector.

To prevent the network from learning a trivial identity mapping, certain constraints are imposed on the hidden representation. The first constraint is sparsity. The idea of sparse representation was inspired by the evidence that the neural activities in the brain seem to be sparse (Vincent et al. 2008).

In addition to the sparsity constraint, another approach to prevent from learning identity, is to use the denoising criteria. In this approach the goal of AE is changed from merely reconstruction of the input to cleaning the partially corrupted input. This would lead to both more interesting and more robust representations of data(Vincent et al. 2010). To do so, they first corrupt the input vector by introducing noise to it. In (Vincent et al. 2008), the stochastic corruption process consists of randomly setting some of the inputs to zero. For real valued input vectors, additive isotropic Gaussian noise is usually applied. The trained AE's are used as building blocks of the deep network. More precisely the representation found in one AE is used as the input to the next AE. Like deep belief networks, each AE was pre-trained using unlabelled data and then a supervised layer was added on the top and the whole model was fine tuned using labelled data. Figure 2.8 shows this process.

Meanwhile, in 2010, a data set of millions of labelled images called "ImageNet" was published by Fei-Fei Li's lab at Standford University (Deng et al. 2009). ImageNet

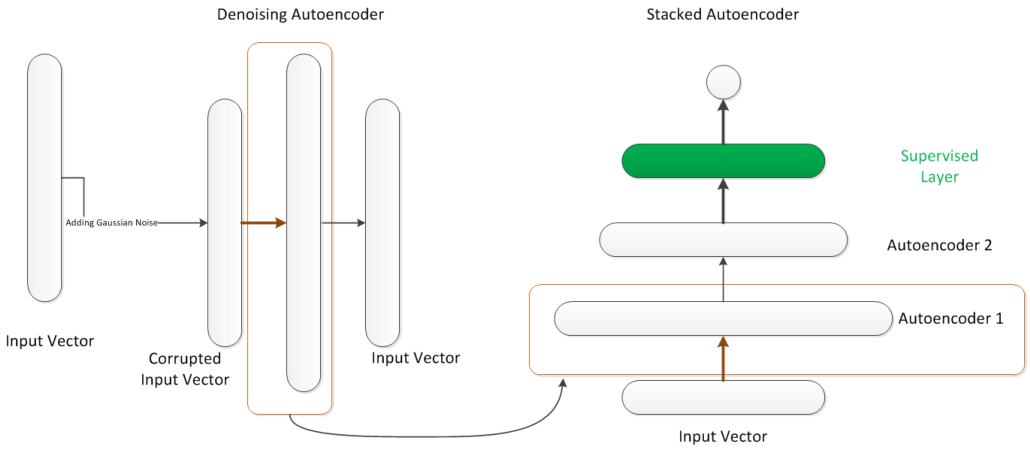


Figure 2.8: Stacked Autoencoder: Input data is corrupted with an additive gaussian noise and is forwarded to an autoencoder. Once the autoencoder is trained the weight vectors of the first layer are considered as one layer in the deep network. The procedure is done again to generate the second layer. Then a 2-layer MLP is added on the top to do the supervised learning

consists of many examples of name-made objects and natural scenes, all labelled by hand to see if machine learning methods can achieve the same labelling. This dataset allowed training of models with huge number of parameters possible.

Since 2011, the availability of such large datasets, the significant increase in speed using GPU's, as well as the use of rectified linear unit for activation, have made the training of deep convolutional network without "pre-training" and only using back-propagation possible. In fact in recent years many variants of convolutional networks have been proposed and achieved state-of-the-art performance in many applications. AlexNet (Krizhevsky et al. 2012), residual-Net (He et al. 2016), inception-Net (Szegedy et al. 2015) and U-Net (Ronneberger et al. 2015) are some of the well-known CNN architectures. The following subsection provides a brief description of convolutional networks.

2.7.2 Convolutional Neural Networks

Inspired by the biological processes in the visual cortex, a variation of MLP was introduced (LeCun et al. 1989) which leveraged spatial correlation in the image to constrain the network architecture. This led to a dramatic decrease in the number of parameters that a network was required to learn during training. By using shared weights, it enabled deeper networks to be trained using backpropagation.

2.7.2.1 CNN architecture

Convolutional neural networks (CNN) are neural networks where at least one layer is a convolution layer. A convolution layer uses convolution instead of the general matrix-vector multiplication and therefore the only parameters to be learned are the convolution filters.

In a 2D convolution layer, both the input and the output have 3 dimensions, namely, width, height, and depth. For example, a 100x100 RGB image given as input to a convolution layer has input width and height of 100 and its depth is 3 (RGB channels). Each filter in the convolution layer is parameterized by the weights (filter parameters). These weights are used to perform convolutions on the input. In CNNs, each filter can be viewed as a feature extractor. Hence each output of a convolution layer is called a feature map.

Each filter in the convolution layer is parameterized by the weights (filter parameters). These weights are used to perform convolutions on the input. A convolution layer with 64 convolution filters will produce an output of depth 64. Each of these 64 outputs is called a feature map. The k'th feature map (o_k)in the output of a convolution layer is obtained as:

$$O_k = F(b_k + \sum_{r}^{R} W_{rk} * X_r), \tag{2.51}$$

where x_r is the r'th channel (or feature map) in the input. b_k is the bias term associated to feature map k. W_{rk} is the filter associated with the input channel r and F is the

activation function.

Therefore, each feature map is the output of application of R different convolution kernels on R input channels summed together along with by a bias term and used as input to an activation function. Both bias and activation functions are pixel-wise operations.

As we go deeper in the network, more abstract and composite features are extracted from the data. For complicated computer vision tasks, we are usually interested in extracting many feature maps from each layer.

While the weights in each filter are learned by minimizing the loss function, each convolution layer has several important hyper-parameters that need to be specified prior to training. We have mentioned some of these parameters: filter height, width, and depth (number of output feature maps). Another important parameter for each filter is padding. Padding identifies how the filter will perform convolution on the boundary of the input. One option is to zero pad the input. Another option is to discard boundary pixels. Doing so will decrease the dimension of the input by filter size -1 pixels in each dimension.

Another parameter is stride, which identifies how the convolution filters slide on the input channel or feature map. For example, a stride of two pixels means that the convolution filter center traverses the entire image in steps of 2 pixels. Applying a filter with stride 2 will reduce the spatial dimension of the output feature map by a factor of 2.

Since the convolution filter size is usually much smaller than the image, the number of parameters to be learned in each layer is small. Reducing the number of parameters in each layer allows for deeper networks to be learned and fewer parameters means that the network is less prone to over-fitting. Also, the small size of the filters reduces the number of operations when computing the output and hence improves the computational efficiency of CNNs.

There are two additional advantages to CNNs. Due to the repetitive nature of convolution, CNN in combination of pooling is invariant to translation of the input image; this is an interesting property for many computer vision applications. And finally, as long as the filter input is larger than the filter, input images of any size can be used both during training and also when the network is used for inference.

In the following sub-sections, some advances in deep learning technologies that have contributed to CNN's great performance are briefly described.

2.7.2.2 Rectified linear Units

A Rectified Linear Unit (ReLU) is the following activation function:

$$f(x) = max(0, x) \tag{2.52}$$

Examples of other activations used include the logistic sigmoid and hyperbolic tangent. However, ReLU activations have become increasingly popular in training deep network in recent years.

Despite being non-differentiable and unbounded, ReLU has a few interesting properties that make it suitable as an activation function in a deep network:

- ReLU is very efficient to compute.
- ReLU is less prone to vanishing gradient problem. This is because in contrast to sigmoid and tangent hyperbolic, ReLU is not a saturating function.
- By forcing all negative values to zero, it makes the model sparse. Sparse models are known to have better generalization ability.

One problem with ReLU activation is that in rare cases neurons may get stuck on the negative side and hence do not contribute to the activation. This problem is known as dying ReLU. This problem is usually mitigated by using small learning rates or

other variants of ReLU like leaky ReLU. Leaky ReLU (Maas et al. 2013) has a very small positive gradient for negative values.

2.7.2.3 Filter factorization

Since convolution is a linear operation, stacking multiple convolutions of small filters together is equivalent to a convolution with a larger filter. For example, 2 layers of 3×3 convolutions are equivalent to one 5×5 convolution. Cascading multiple small filters instead of one large filter is called filter factorization. The benefit of filter factorization is that using multiple layers of small size filters has fewer parameters than a single convolution with the large filter. In the above example, two layers of m 3×3 filters for an input of n channels has $n \times 3 \times 3 \times 2 \times m = 18nm$ parameters whereas the latter network with one 5×5 filter has $n \times 5 \times 5 \times m = 25nm$ parameters. Using filter factorization can lead to enormous parameter reduction in very deep networks with lots of large filters. In practice, a ReLU activation layer is added after each step of factorization (each convolution layer). This reduction in the number of parameters combined with the non-linearity added after each convolution layer can make the model training easier and less susceptible to over-fitting (Szegedy et al. 2016).

2.7.2.4 Batch Normalization

In deep networks, during the training the parameters of the network change, this leads to the change of distribution of layer's inputs, as each layer's input is a function of the parameters of the previous layer. This change of the layer's input distribution is known as the covariance shift. Covariance shift is undesirable since the network needs to continuously adapt to these changes during training. One way to deal with this issue is standardization.

Standardization of the input data is an old practice (Wiesler & Ney 2011, Aksoy & Haralick 2001) used in different types of neural networks. In standardization, a linear transformation is applied to the data so that the mean of the transformed data is 0 and its standard deviation is 1. Standardization of the input to the first layer (i.e. training

data) has been long used in machine learning as it has been shown to improve the convergence rate during training (Wiesler & Ney 2011).

The idea in batch normalization (Ioffe & Szegedy 2015) is to standardize inputs of the intermediate layers of the network. This is done for each batch of data during training and hence the name batch normalization.

Batch normalization is done using two simplifications. 1) Instead of whitening the features in each layer which involves the computation of the covariance matrix, each scalar feature is standardized independently by subtracting the mean and dividing by the standard deviation across the training data. After this standardization, to preserve the capacity of the network in representation a scale and shift transformation is applied. These scale and shift parameters are also trainable and ensure that the representation power of the network is maintained. In fact, by applying proper scale and shift values the network is able to restore the original input if that is a good thing to do in terms of loss minimization.

The second simplification is that the aforementioned mean and variance are computed across each mini-batch rather than the whole training data.

In CNNs, batch normalization is usually applied to the output of each layer before applying the activation function. Also, when included in a CNN, batch normalization normalizes all the elements at different locations in a feature map together and hence one pair of scale and shift values are learned per feature map.

As opposed to training, the dependency of activation on mini-batches is neither necessary nor of interest during inference, as one would like the output to solely and deterministically depend on input. Therefore, the mean and variance during inference are computed across the whole test data. The following lists the main benefits of batch normalization

- In batch normalization, the transformation applied to each training sample is dependent on the other samples in a given mini-batch and hence is no longer

deterministic. This improves the generalization of the network.

- Batch normalization stabilizes the distribution of the input to each layer of the network. Normalization of input to each layer reduces the sensitivity to weight initialization.

- Batch normalization enables higher learning rates by making the network less sensitive to the parameter scale. Increasing the learning rate leads to faster convergence.

2.7.2.5 Residual Learning

In deep networks, as we go deeper in layers the feature extracted by the networks become more abstract. As mentioned earlier in Deep learning section 2.7 it is known that the depth of the network plays an important role in the success of deep learning. However, it has been shown that in a plain neural network, as the network becomes deeper the loss becomes saturated and then increases rapidly (He et al. 2016). This phenomenon is called network degradation. In fact, unlike the common expectation that deeper networks tend to over-fit due to an increase in the number of parameters, the deeper network has shown to have an increase in the training loss.

In theory, by adding identity layers to make a network deeper, the training error of the deeper network should not be greater than the shallower one. He et al. (2016), argue that the degradation problem is due to difficulties in approximating identity maps by non-linear layers. They try to address the degradation problem using residual networks.

Let us assume that the ideal mapping from input to output is $H(x)$. In a residual block, instead of learning $H(x)$, the layers of the network are explicitly forced to learn the residual, i.e. $F(x) = H(x) - x$. This is done by adding a skip identity connection as shown in Figure 2.9.

In fact, residual connections help to mitigate the problem of vanishing gradients by

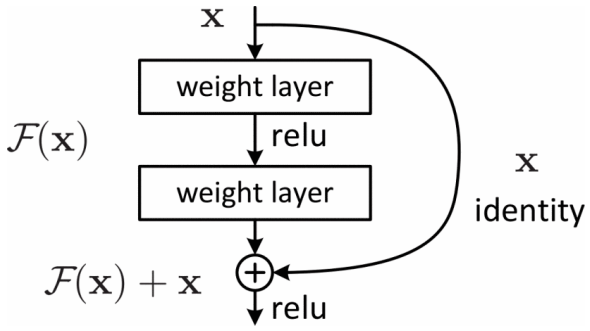


Figure 2.9: A residual block with an identity connection. Image is taken from (He et al. 2016)©[2016]IEEE. Permission for reuse of figure granted by publishers.

propagating the error through skip connections back to the early layers of the network.

ResNet (He et al. 2016) used a very deep network with 152 layers and achieved first place in the ImageNET Large Scale Visual Recognition Challenge (ILSVRC) (Russakovsky et al. 2015). Since then, residual connections and skip connections have been used in many other CNN variants including Inception-v4 (Szegedy et al. 2017) and U-Net(Ronneberger et al. 2015).

2.8 Summary

Noise and PVE are two important sources of error in PET quantification. In this section a basic background for understanding the noise and PVE in PET images and the common approaches to mitigate them were presented. In the following chapter (Chapter 3) a reconstruction method is proposed to reduce the noise and enhance the resolution of PET imaged using the subject's registered MR image. In chapter 4, two post reconstruction methods for PET image restoration are proposed. The proposed methods also benefit from the subject's registered MR image. In Chapter 5, two deep learning models are trained on simulation data to restore PET images both with and without a corresponding T1-weighted MR image.

Chapter 3

PET reconstruction using prior image derived dictionaries

Preface

In this chapter, a novel MR-guided method for reconstructing PET images is proposed. In this method, instead of the popular Bayesian approach for regularization of the reconstruction problem, a novel patch based re-parameterization framework is utilized. In this framework, a patch based dictionary extracted form the subject's T1-weighted MR image is used to represent the reconstructed PET image. Also, an algorithm for sparse Poisson log-likelihood is proposed and used for imposing sparsity on the coefficients of the dictionary atoms.

This work has been published as: Tahaei, Marzieh S., and Andrew J. Reader. "Patch-based image reconstruction for PET using prior-image derived dictionaries." Physics in Medicine & Biology 61.18 (2016): 6833.

Patch-Based Image Reconstruction for PET using Prior-Image Derived Dictionaries

Marzieh S. Tahaei[1] and Andrew J. Reader[2]

[1] McConnell Brain Imaging Centre, Montreal Neurological Institute, McGill University, Montreal, Canada

[2]Division of Imaging Sciences and Biomedical Engineering, Department of Biomedical Engineering, King's College London, St. Thomas' Hospital, London, UK

Abstract

In PET image reconstruction, regularization is often needed to reduce the noise in the resulting images. Patch-based image processing techniques have recently been successfully used for regularization in medical image reconstruction through a penalized likelihood framework. Re-parameterization within reconstruction is another powerful regularization technique in which the object in the scanner is re-parameterized using coefficients for spatially-extensive basis vectors. In this work, a method for extracting patch-based basis vectors from the subject's MR image is proposed. The coefficients for these basis vectors are then estimated using the conventional MLEM algorithm. Furthermore, using the alternating direction method of multipliers (ADMM), an algorithm for optimizing the Poisson log-likelihood while imposing sparsity on the parameters is also proposed. This novel method is then utilized to find sparse coefficients for the patch-based basis vectors extracted from the MR image. The results indicate the superiority of the proposed methods to patch-based regularization using the penalized likelihood framework.

Keywords: Positron emission tomography(PET), re-parameterization, image reconstruction, anatomical prior, patch representation , hybrid PET-MR, sparsity, dictionary learning.

3.1 Introduction

PET image reconstruction seeks to estimate a 3D image (i.e. object representation) of the radiotracer concentration from line of response measurements. Statistical algorithms for PET image reconstruction have become increasingly popular due to their ability to model the noise in the measurement data and also their flexibility in modelling the physical characteristics of the scanning process. Among iterative methods, the maximum likelihood expectation maximization (MLEM) method and its variations (e.g OSEM (Green 1990)) are now widely being used in clinical practice as they provide a visually improved image quality and can have better performance in lesion detection than conventional methods (Wells et al. 2000).

PET reconstruction is an ill-conditioned inverse problem. Hence, a regularization approach is required to restrict the space of feasible solutions. During the last decade many efforts have been taken to reduce the noise of the reconstructed images, spanning from the use of simple spatial smoothness priors, through more robust edge preserving smoothing priors to the recent methods of patch-based priors.

The literature on regularization in PET reconstruction can be divided into two main categories: i) including a prior function in the objective using a maximum a posteriori (MAP) framework and ii) representing the radioactive distribution as a superposition of spatial basis vectors such as blobs (re-parameterization).

Most regularization methods in the PET reconstruction literature focus on designing appropriate prior functions to be utilized in a MAP framework. Fessler (1997), Nuyts et al. (2002, 2003), Wang & Qi (2012), Ahn et al. (2015). Many of these MAP-based approaches benefit from the boundary information from magnetic resonance imaging (MRI) and computed tomography (CT) images in order to encourage smoothness only within anatomical structures and thus preserve edges (Ardekani et al. 1996, Bowsher et al. 1996, Rangarajan et al. 2000, Cheng-Liao & Qi 2011, Somayajula et al. 2005, Tang & Rahmim 2009, Rangarajan et al. 2000).

Patch-based image processing techniques have become increasingly popular in the last decade. They leverage two interesting characteristics of natural images: the intrinsic redundancy and the locally structured patterns in images. These characteristics allow image patches to be sparsely represented (very few non-zero coefficients are needed to represent their content) as a superposition of some learned over-complete set of basis vectors. Sparse representation of image patches has proven to achieve good performance in a variety of image processing tasks like denoising and restoration (Buades et al. 2005, Dabov et al. 2007). Their robustness to noise soon motivated the medical imaging community to use them for medical image reconstruction.

The authors in (Chen et al. 2015) and (Tang et al. 2014) have proposed PET image reconstruction methods that use sparse representation of image patches using a dictionary learned from anatomical images. These methods are based on a Bayesian (or MAP) framework where the regularized objective function is composed of a data fidelity term and a sparse representation error term. A maximum a posteriori expectation maximization (MAPEM) or an approximate one-step-late MAPEM (Green 1990) method is utilized to estimate the image.

On the other hand, some of the recent anatomical-based regularization methods such as kernel-based regularization in (Wang & Qi 2015, Novosad & Reader 2016) and the recent super-voxel method presented in (Jiao et al. 2015) fall into the re-parameterization category and hence the conventional MLEM algorithm can be used to reconstruct the image. These methods have shown promising results in PET image reconstruction. However these methods work best in situations where the functional activity within tissues is homogeneous. This work tries to address this issue using a patch-based re-parameterization framework. To do so, the radioactivity distribution in the field of view is modelled as a composition of overlapping patches, where each patch is represented as a superposition of learned dictionary atoms.

Motivated by non-local patch-based denoising techniques, a novel patch-based ba-

sis function extraction method from a prior images is proposed. In this method, the patches from the prior image are first clustered into C sets and for each cluster a dictionary is learned from the patches in that cluster. The resulting dictionaries are then used as basis vectors to represent patches in the PET image. In order to estimate the coefficients for these basis vectors, only a re-parameterization of the Poisson log likelihood is required and hence the image can be reconstructed using the conventional MLEM algorithm.

Furthermore, another important contribution of this work is to propose an algorithm for optimizing the re-parameterized Poisson log-likelihood objective function while imposing sparsity on the coefficients. An ADMM method is utilized to solve this l_1 penalized likelihood function. The proposed algorithm is then used to impose a sparsity penalty on the coefficients of the patch-based basis vectors extracted from the subject's T1-weighted MR image and the effect of the sparsity parameter is explored using simulation data.

In order to enhance readability, throughout the paper, bold capital letters are used to indicate matrices, bold lower-case letters represent column vectors and all non-bold letters denote scalar values.

3.2 Background

3.2.1 Sparse patch-based representation of images

Let us assume that $\mathbf{D} \in \mathbb{R}^{M \times N}$ is a dictionary of N atoms (basis vectors) learned from a database of overlapping patches of size M (i.e. $\sqrt{M} \times \sqrt{M}$ for a 2D image) taken from the image under consideration. The dictionary is learned in such a way that each patch $\mathbf{p}_k$ within the database can be represented as a sparse superposition of the dictionary atoms, i.e. $\mathbf{p}_k \approx \mathbf{D}\boldsymbol{\alpha}_k$, where $\boldsymbol{\alpha}_k \in \mathbb{R}^N$ is the coefficient vector representing the patch $\mathbf{p}_k$.

For a given dictionary, finding the sparse representation (coding) of the image $\mathbf{x}$ firstly involves extracting overlapping patches from the image. Let $\mathbf{R}_k \in \mathbb{R}^{M \times J}$ be a

binary matrix operator that extracts the patch $\mathbf{p}_k$ at location k from the image $\mathbf{x}$:

$$\mathbf{p}_k = \mathbf{R}_k \mathbf{x} \tag{3.1}$$

The transpose of the matrix $\mathbf{R}_k$ operating on the patch $\mathbf{p}_k$, gives an image in which $\mathbf{p}_k$ is put back into the position of patch k and zero-pads the image elsewhere. This procedure is illustrated in figure 1.

Once all the patches are extracted, for each patch $\mathbf{p}_k$, the sparse coding algorithm finds the set of coefficients $\boldsymbol{\alpha}_k$ for the dictionary atoms which minimizes the representation error while imposing sparsity on the coefficient vector. This is usually done using the following objective function for each patch:

$$\arg\min_{\boldsymbol{\alpha}_k} \quad \|\mathbf{p}_k - \mathbf{D}\boldsymbol{\alpha}_k\|_2^2, \text{ such that } \|\boldsymbol{\alpha}_k\|_0 < T \tag{3.2}$$

where T is the sparsity parameter, indicating that fewer than T atoms must be used. Each pixel in the reconstructed image $\mathbf{x}$ is then obtained by averaging reconstructed patches that cover it together:

$$\mathbf{x} \approx \sum_{k=1}^{K} \frac{\mathbf{R}_k^T(\mathbf{D}\boldsymbol{\alpha}_k)}{\mathbf{q}} \tag{3.3}$$

where $\mathbf{q} = \sum_{k=1}^{K} \mathbf{R}_k^T \vec{\mathbf{1}}$ is a vector of size J in which each elements q_j indicates the number of patches contributing to pixel j. Here and throughout the paper, *vector* by *vector* multiplication and division should be interpreted as element-wise multiplication and division respectively.

3.2.2 MLEM

Let $\mathbf{m} = \{m_1 \ldots m_I | m_i \in \mathbb{Z}^+ \cup \{0\}\}$ be a vector in which each element m_i contains the number of counts accumulated in bin i during a PET scan. Because of the counting nature of the annihilation events, m_i can be well modelled by a Poisson distribution

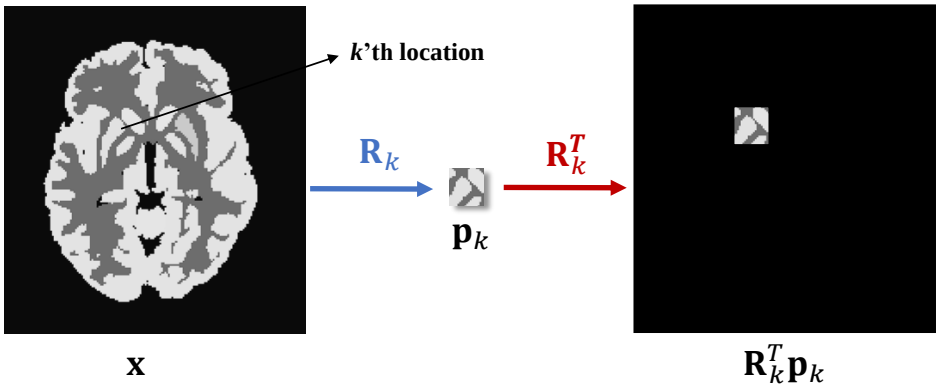


Figure 3.1: Schematic illustration of the patch-extraction operator and its transpose. Note that the order of pixels arranged in vector $\mathbf{x}$ depends on the method used to store the 2-D/3-D matrices.

with mean $\bar{m}_i$:

$$Pr(m_i|\bar{m}_i) = \frac{e^{-\bar{m}_i}\bar{m}_i^{m_i}}{m_i!} \tag{3.4}$$

The mean of the measurement data can be modelled as an affine transform of the modelled mean activity distribution $\mathbf{x} \in \mathbb{R}^J$ within the scanner through the system matrix $\mathbf{A} = \{a_{i,j}\} \in \mathbb{R}^{I\times J}$ as :

$$\bar{\mathbf{m}} \approx \mathbf{A}\mathbf{x} + \mathbf{r}, \tag{3.5}$$

where I and J are the number of PET measurements and image pixels respectively. $\mathbf{r}$ is the model of the random and scatter events. Then the log likelihood of the true activity distribution of $\mathbf{x}$ given $\mathbf{m}$ can be written as:

$$L(\mathbf{x}|\mathbf{m}) = \sum_{i=1}^{I}(m_i \ln(\sum_{j=1}^{J} a_{i,j}x_j) - \sum_{j=1}^{J} a_{i,j}x_j). \tag{3.6}$$
$$\text{such that} \quad \mathbf{x} \geq 0$$

$L(\mathbf{x}|\mathbf{m})$ can be maximized using the iterative MLEM algorithm. MLEM starts with a non-zero initialization of $\mathbf{x}$ and converges to the solution using the following update

rule:

$$\mathbf{x}^{n+1} = \frac{\mathbf{x}^n}{\mathbf{s}} \mathbf{A}^T \frac{\mathbf{m}}{\mathbf{A}\mathbf{x}^n + \mathbf{r}} \tag{3.7}$$

where n is the iteration number and $\mathbf{s} = \mathbf{A}^T \vec{\mathbf{1}}$ is the sensitivity image. From an optimization transfer perspective, in each iteration the MLEM method maximizes the following surrogate function for $L(\mathbf{x}|\mathbf{m})$ (Reader & Verhaeghe 2014):

$$G(\mathbf{x}|\mathbf{x}^n, \mathbf{m}) = \sum_{j=1}^{J} s_j (x_j^{EM} \ln(x_j) - x_j), \tag{3.8}$$

where $\mathbf{x}^{EM}$is equal to $\mathbf{x}^{n+1}$ from the EM update in Eq.3.7, therefore being a function of $\mathbf{x}^n$ and $\mathbf{m}$.

3.2.3 MAPEM

The MAP formulation for PET reconstruction can be written as:

$$L_{aposteriori}(\mathbf{x}|\mathbf{m}) = L(\mathbf{x}|\mathbf{m}) + ln(Prior(\mathbf{x})), \tag{3.9}$$

Assuming that the prior function for image $\mathbf{x}$ follows a Gibbs distribution $\frac{1}{Z}\exp(-\beta U(\mathbf{x}))$, one can write the MAP formulation as:

$$L_{aposteriori}(\mathbf{x}|\mathbf{m}) = \quad L(\mathbf{x}|\mathbf{m}) - \beta U(\mathbf{x}) \tag{3.10}$$

3.2.3.1 MAP with quadratic penalty (Q-MAP)

In this method, a weighted quadratic function is used as $U(\mathbf{x})$ to penalize large differences in the intensities between the neighbouring pixels:

$$U(\mathbf{x}) = \sum_{j=1}^{J} \sum_{k \in Neib(j)} w_{jk} (x_j - x_k)^2 \tag{3.11}$$

The resulting MAP objective function ($L_{aposteriori}$) can be solved using an optimization transfer framework (De Pierro's decoupling rule) (De Pierro 1995). In this work a

Gaussian function is used to determine the weights. See (Tahaei & Reader 2014) for a full derivation of the algorithm.

3.2.3.2 MAP with relative difference penalty (RD-MAP)

The problem with the quadratic penalty is that it can strongly smooth edges in the reconstructed image. One simple way to address this problem is to use the following relative difference penalty function (Nuyts et al. 2002) :

$$U(\mathbf{x}) = \sum_{j=1}^{J} \sum_{k \in Neib(j)} \frac{(x_j - x_k)^2}{(x_j + x_k) + \gamma |x_j - x_k|} \tag{3.12}$$

where γ is a parameter that controls the degree of edge-preservation. In our implementation, the resulting MAP objective function is maximized using the one-step-late approach (Green 1990).

3.2.3.3 Sparse patch-based MAPEM (Sparse PB-MAPEM)

One way to incorporate a patch-based representation for regularization in PET reconstruction is through the MAPEM framework. The penalty function in this model can be defined as minimizing representation error of patches in the image using a fixed dictionary while imposing sparsity on the coefficients. This leads to minimizing the following negative penalized log-likelihood objective function (Chen et al. 2015):

$$\begin{aligned} & \min \quad -\lambda L(\mathbf{x}|\mathbf{m}) + R_{sparsity}([\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_K]|\mathbf{x}) \qquad (3.13) \\ = \quad & \min \quad -\lambda \textstyle\sum_{i=1}^{I} (m_i ln(\sum_{j=1}^{J} a_{i,j} x_j) - \sum_{j=1}^{J} a_{i,j} x_j) + \sum_{k=1}^{K} (\|\mathbf{R}_k \mathbf{x} - \mathbf{D}\boldsymbol{\alpha}_k\|_2^2 + \mu \|\boldsymbol{\alpha}_k\|_0) \end{aligned}$$

where μ controls the sparsity. The above optimization can be solved in an alternating manner by (i) fixing $\mathbf{x}$ and solving for the $\boldsymbol{\alpha}_k$'s and then (ii) fixing the $\boldsymbol{\alpha}_k$'s and solving for $\mathbf{x}$ iteratively. In (Chen et al. 2015), problem (i) is solved using the orthogonal matching pursuit (OMP) algorithm (Tropp et al. 2007). OMP is a fast greedy algorithm. It

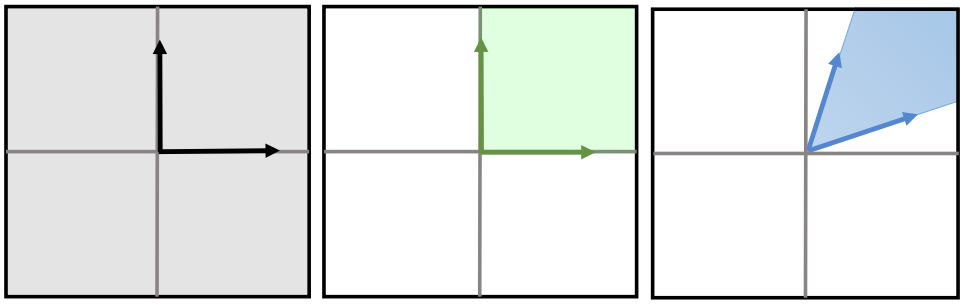


Figure 3.2: The span of two basis vectors in a 2-D space. Left: the span of two orthogonal basis vectors. Middle: the span of two orthogonal basis vectors with non-negative coefficients. Right: the span of two non-orthogonal basis vectors with non-negative coefficients.

starts with an empty set of selected atoms and at each iteration finds an unselected atom from $\mathbf{D}$ that best matches the current residual. This atom is then added to the set, and the residual as well as the coefficients are updated using this updated set of dictionary atoms. The iterations are continued until the residual is less than a specified error tolerance called ε.

To solve problem (ii), an optimization transfer framework is used(De Pierro 1995). See (Chen et al. 2015) for a detailed derivation of the algorithm. In (Chen et al. 2015) , the dictionary is learned from patches extracted from CT images. In this study however, for the sparse PB-MAPEM method the dictionary is learned from the subject's registered MR image using a Matlab implementation of the K-SVD algorithm (`http://www.cs.technion.ac.il/~elad/software/`).

3.2.4 Re-parameterized MLEM

Considering that the radioactivity distribution is modelled as a vector $\mathbf{x} \in \mathbb{R}^J$, one can represent the vector $\mathbf{x}$ as a superposition of some basis vectors:

$$\mathbf{x} = \sum_{l=1}^{L} \theta_l \boldsymbol{\phi}_l = \boldsymbol{\Phi}\boldsymbol{\theta} \tag{3.14}$$

where $\mathbf{\Phi} \in R^{J \times L}$ is a matrix in which the l'th column is the $l'th$ basis vector: $\mathbf{\Phi} = \{\boldsymbol{\phi}_1, ..., \boldsymbol{\phi}_L\}$ and $\boldsymbol{\theta}$ is the coefficient vector in which each θ_l is the coefficient for the l'th basis vector.

The are infinitely many choices of basis vectors. For a given set of basis vectors the coefficient vectors are considered as the representation of the radioactivity distribution in the given basis. In this setting, an obvious way to achieve regularization is to impose constraints on the coefficient vector and hence limit the space of possible representations of $\mathbf{x}$ for the given basis set $\mathbf{\Phi}$. This is usually achieved through a Bayesian approach.

Now, let us assume that for a given non-negative set $\mathbf{\Phi}$, one is only allowed to use non-negative coefficients to represent $\mathbf{x}$. With this new non-negativity constraint the span of the basis vectors is dependent on the shape of the basis vectors. Hence one can achieve different levels of regularization depending on the shape of the basis vectors. For a fixed number of basis vectors, the less the orthogonality between the basis vectors, the higher the regularization (the more limited the span of the possible radioactivity distributions). Figure 3.2 illustrates how the use of non-orthogonal basis vectors with non-negative coefficients can restrict the space of solutions.

This is very important for PET image reconstruction since the Poisson log-likelihood objective of the radioactivity distribution given the measurement data inherently imposes a non-negativity constraint on the mean of the radioactivity distribution. Therefore, one can achieve regularization by choosing non-negative basis vectors and using the traditional MLEM algorithm. This use of non-trivial basis vectors is considered as one of the categories of re-parameterization in the PET reconstruction literature. Different choices of overlapping and non-overlapping basis vectors have been explored in a number of studies. The use of blobs, Gaussians and anatomical regions of interest (ROIs) as well as the recent kernel-based method are among these methods(Lewitt 1992, 1990, Jacobs et al. 1998, Wang & Qi 2015, Jiao et al. 2015). In the

re-parameterization framework the MLEM update rule becomes:

$$\boldsymbol{\theta}^{n+1} = \frac{\boldsymbol{\theta}^n}{\boldsymbol{\Phi}^T \mathbf{A}^T \vec{\mathbf{1}}} \boldsymbol{\Phi}^T \mathbf{A}^T \frac{\mathbf{m}}{\mathbf{A}\boldsymbol{\Phi}\boldsymbol{\theta}^n + \mathbf{r}} \tag{3.15}$$

3.3 Proposed Method

In the following sub-section, a framework for incorporating a patch-based representation as the re-parameterization of the whole image using matrix multiplication is presented. In the remainder of this section, novel methods for estimating the coefficients and extracting the basis vectors for this patch-based re-parameterization framework are proposed.

3.3.1 General framework: Patch-based representation as image re-parameterization

Consider a general case in which each patch is represented using a dictionary associated with it. This dictionary can be identical for all patches or it can be chosen according to the structure or location of the patch. Let's assume that $\mathbf{D}_k = [[\mathbf{D}_k]_1 ... [\mathbf{D}_k]_N]$ is the dictionary matrix with $[\mathbf{D}_k]_d$ representing the d'th column vector (or dictionary atom) used for the representation of patch k (we will talk about how this dictionary can be learned later on), i.e., $\mathbf{p}_k \approx \mathbf{D}_k \boldsymbol{\alpha}_k$. The basis set $\boldsymbol{\Phi}_k$ can be defined by replacing each atom $[\mathbf{D}_k]_d$ in $\mathbf{D}_k$ by $\mathbf{R}_k^T [\mathbf{D}_k]_d$:

$$\boldsymbol{\Phi}_k = [\mathbf{R}_k^T [\mathbf{D}_k]_1, \mathbf{R}_k^T [\mathbf{D}_k]_2 \dots \mathbf{R}_k^T [\mathbf{D}_k]_N] \tag{3.16}$$

By generating $\boldsymbol{\Phi}_1$ to $\boldsymbol{\Phi}_K$ from $\mathbf{D}_1$ to $\mathbf{D}_K$ respectively and concatenating them together we obtain a huge sparse matrix called $\boldsymbol{\Phi}$:

$$\boldsymbol{\Phi} = [[\boldsymbol{\Phi}_1], ..., [\boldsymbol{\Phi}_K]]. \tag{3.17}$$

Equivalently, the vector $\boldsymbol{\theta}$ is obtained by concatenating $\boldsymbol{\alpha}_1$ to $\boldsymbol{\alpha}_k$'s together :

$$\boldsymbol{\theta}^T = [[\boldsymbol{\alpha}_1]^T, ..., [\boldsymbol{\alpha}_k]^T]^T \tag{3.18}$$

where $\boldsymbol{\alpha}_k$ is the sparse coefficient vector for representing the patch $\mathbf{p}_k$.

As mentioned before, the final image is formed by combining sparsely represented patches together. One can easily show that $\boldsymbol{\Phi\theta} = \sum_{k=1}^{K} \mathbf{R}_k^T(\mathbf{D}_k\boldsymbol{\alpha}_k)$. Hence, by normalizing for overlapping patches, the reconstructed image can be written in matrix multiplication form as:

$$\mathbf{x} \approx \mathbf{Q^{-1}\Phi\theta}, \tag{3.19}$$

where $\mathbf{Q} \in R^{J\times J}$ is a diagonal normalization matrix with diagonal elements identical to those of $\mathbf{q}$ in Eq.3.3. For notational convenience, hereafter we refer to $\mathbf{Q^{-1}\Phi}$ as $\boldsymbol{\Phi'}$.

$$\mathbf{x} \approx \boldsymbol{\Phi'\theta}, \tag{3.20}$$

3.3.2 Extracting the basis vectors from a prior image

In this section a new basis extraction method which uses patches from a prior image to represent a patch in the PET image is proposed. This method is motivated by many novel non-local denoising techniques in the image processing literature including the non-local means and BM3D algorithms (Kervrann & Boulanger 2006, Dabov et al. 2007, Chatterjee & Milanfar 2012).

In the proposed method, instead of learning one large dictionary from the prior image to represent every patch in the PET image, one can first cluster similar patches together and then learn a separate dictionary from each cluster. The idea of clustering prior to dictionary learning has been previously used in image denoising (Chatterjee & Milanfar 2009) and shown to improve the performance. Using this strategy in the aforementioned re-parameterization framework, can reduce the number of basis vec-

tors needed to represent each patch in the PET image. This in turn reduces the number of parameters to be estimated within the MLEM algorithm.

Once the dictionaries are learned, for every patch $\mathbf{p}_k$ in the PET image, the cluster to which the corresponding patch in the anatomical image (i.e. the patch in the prior image, such as an MR image, at location k) belongs to is selected and the dictionary learned from this cluster is used to represent $\mathbf{p}_k$.

The problem with this approach, i.e. using a dictionary associated with the same location in the prior image, is that in order to be able to use the boundary information from the prior image the intensity values on the sides of that boundary cannot have opposite relative contrast compared to those of the selected patch in the PET image. This is due to the implicit non-negativity constraint imposed by the MLEM algorithm on the basis vectors and the coefficients.

An example is a T1-weighted MR image for reconstructing fluorodeoxyglucose (FDG) images. Since the contrast in PET grey matter and white matter are inverted a dictionary associated with cortical boundary regions in a T1-weighted MR image cannot represent a patch in these regions in the PET image. One simple way to address this issue is to modify the prior image beforehand. For the case of a T1-weighted MR image used in this study, this is done by segmenting the MR image into white matter and grey matter and changing the intensity of grey matter to a value greater than the maximum intensity in the white matter using the following:

$$\forall j \in \text{grey-matter region} \qquad y_j = aw^*, a > 1 \tag{3.21}$$

where y is the modified MR image from which the patches are extracted for dictionary learning and w^* is the maximum intensity in the white-matter.

Once the opposite contrast issue is resolved, the basis vectors for PET reconstruction can be extracted from the image. To do so, first the resulting modified T1-weighted

MR image y is decomposed to overlapping patches of size $\sqrt{M} \times \sqrt{M}$. The patch decomposition performed on the PET and the modified MR image should be identical so every patch $\mathbf{p}_k$ from the PET image has one and only one correspondence at the same location and of the same size in the MR image. Every patch is then normalized by subtracting the minimum and dividing by the maximum. Let us call the normalized patch from the modified MR image at location k, $\mathbf{w}_k$. The normalized patches are then clustered into C distinct clusters $\{\mathbf{U}_1, ..\mathbf{U}_c, ...\mathbf{U}_C\}$ using the k-means algorithm. For each cluster index c, a dictionary $\mathbf{D}_c$ is learned from the patches in the cluster using the non-negative matrix factorization algorithm (described in the following section). Then for every patch $\mathbf{p}_k$ in the PET image the dictionary atoms learned from the cluster to which the normalized patch from location k in the MR image belongs is used to represent the patch.

Moreover, in order to ensure that the set of basis vectors is able to represent structures not represent in the prior image, an all-one vector rescaled to unit length is added to the dictionary:

$$\mathbf{p}_k \approx [\mathbf{D}_{c'}, \frac{\vec{\mathbf{1}}}{\sqrt{M}}]\boldsymbol{\alpha}_k \quad c' \text{ is the index of the cluster containing patch } \mathbf{w}_k \tag{3.22}$$

Note that adding this constant atom to the set of basis vectors, is expected to work well for relatively small patches and is empirically confirmed by simulating lesions in the PET data (see Simulation Studies section).

It is also worth noting that because of the normalization of the patches before basis extraction, the relative contrast imposed by the prior image (e.g. the modified MR image) will not force a similar contrast on the PET image. Therefore for the case of using the modified T1-weighted MR, as an example, the algorithm is not sensitive to the value of a in Eq.3.21, as long as a is greater than 1. In our implementation a was

set to 2. The procedure of extracting patch-based basis vectors from the subject's MR image is illustrated in figure 3.3.

3.3.2.1 Dictionary learning using sparse non-negative matrix factorization

Because of the implicit non-negativity constraint within the Poisson log likelihood objective function, a non-negative matrix factorization (NMF) is used for dictionary learning to ensure the non-negativity of the re-parameterized image. Let $\mathbf{U}_c$ be a matrix where each column is a patch of the modified MR image associated with cluster c. The goal of NMF is to factorize $\mathbf{U}_c$ into two non-negative matrices $\mathbf{D}$ and $\mathbf{E}$: $\mathbf{U}_c \approx \mathbf{DE}$. $\mathbf{D}$ can be viewed as the dictionary and columns of $\mathbf{E}$ can be considered as the coefficient vectors for patches in cluster c.

$$\begin{aligned} \text{Minimize} \quad & _{\mathbf{D},\mathbf{E}} = \|\mathbf{U} - \mathbf{DE}\|_F \\ \textit{such that} \quad & \mathbf{D} \geq 0, \quad \mathbf{E} \geq 0 \end{aligned} \tag{3.23}$$

where $\|.\|_F$ denotes the Frobenius norm. While the non-negativity constraint in NMF naturally favours sparsity, several studies in recent years have tried to impose a sparsity penalty on the data fidelity objective to enhance sparsity in a controllable manner (Hoyer 2004, Peharz & Pernkopf 2012) . For the purpose of this study, the NMF with the l_0 constraint proposed in (Peharz & Pernkopf 2012) is used. In (Peharz & Pernkopf 2012), multiplicative update rules are used to update the dictionary matrix in each step. In this study this algorithm is referred to as NMFl_0. The authors have also made a Matlab implementation of this method available online (`https://www.spsc.tugraz.at/tools/`). We have utilized this implementation for learning dictionaries from each cluster.

3.3.3 Estimating the coefficient vector

In this section, two methods for estimating the coefficient vector are presented. The first is the conventional re-parameterized MLEM method, and the latter is a novel

method for finding a sparse coefficient vector using ADMM.

Clustered patch-based MLEM (C-PB-MLEM)

By substituting $\mathbf{x}$ with $\mathbf{\Phi'\theta}$ in the likelihood function L (which can only be done if $\mathbf{\Phi'\theta} \geq 0$), a re-parameterized ML objective is obtained which can be maximized using the MLEM algorithm. We call this method clustered patch-based MLEM (C-PB-MLEM) which is simply a re-parameterization of the MLEM algorithm with clustered patch-based basis vectors extracted from a prior image. Therefore the update rule for C-PB-MLEM can be written as:

$$\boldsymbol{\theta}^{n+1} = \frac{\boldsymbol{\theta}^{n}}{\mathbf{\Phi'}^{T}\mathbf{A}^{T}\vec{\mathbf{1}}}\mathbf{\Phi'}^{T}\mathbf{A}^{T}\frac{\mathbf{m}}{\mathbf{A}\mathbf{\Phi'}\boldsymbol{\theta}^{n}+\mathbf{r}} \tag{3.24}$$

Note that when a large number of clusters is used, the patches belonging to each cluster are very similar and hence only a few dictionary atoms are needed to represent the patches belonging to that cluster. On the other hand, when a small number of clusters is used, each cluster may contain patches with diverse structures and hence a larger dictionary may be required to represent the patches belonging to each cluster. For the latter case, imposing sparsity constraint on the coefficient vector can lead to a more stable representation of the object. An algorithm for imposing a sparsity in a log likelihood objective function is proposed in the following section.

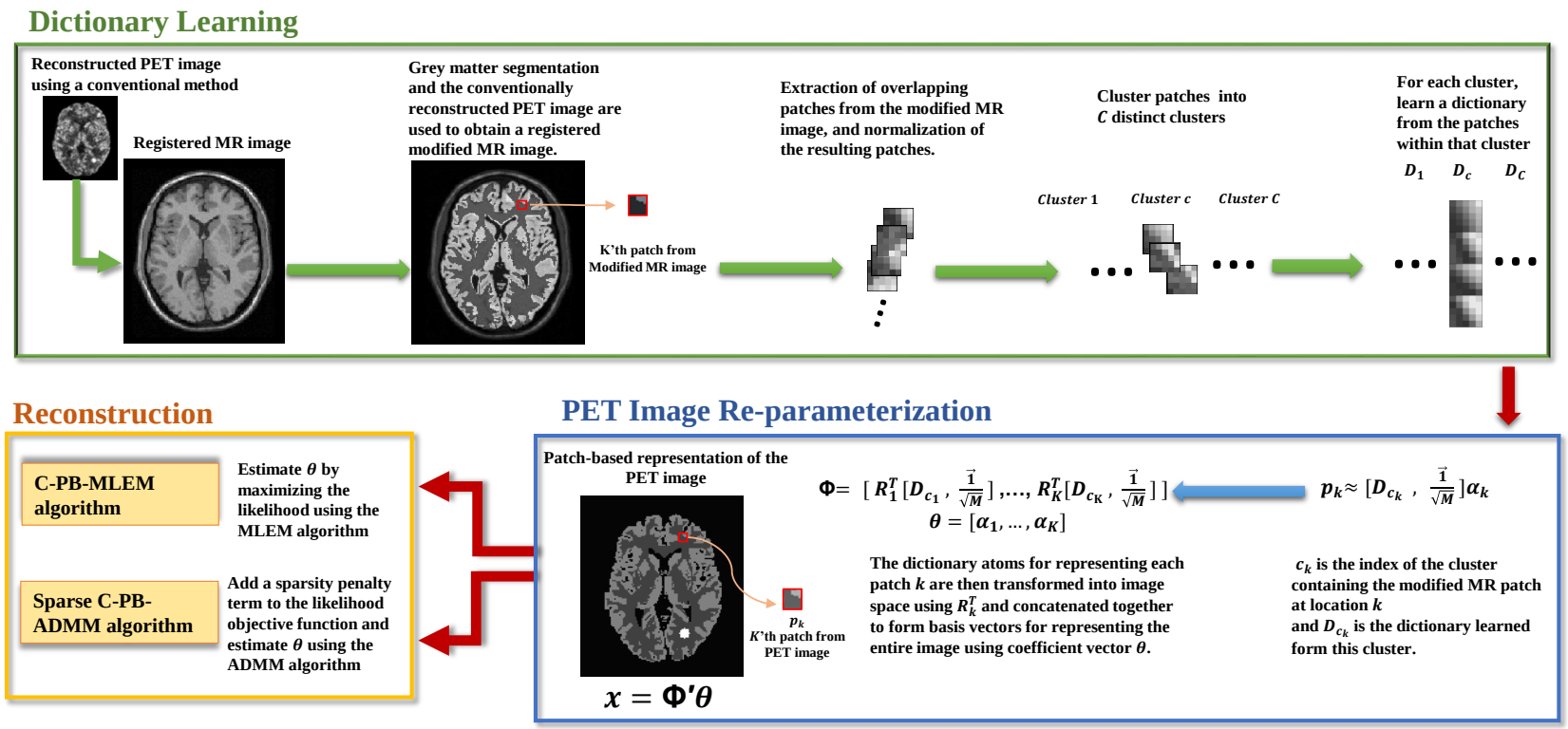


Figure 3.3: Illustration of the proposed method

3.3.3.1 Sparse clustered patch-based ADMM (sparse C-PB-ADMM)

For over-complete dictionaries, the non-negativity of $\mathbf{\Phi}'$ and $\boldsymbol{\theta}$ along with the over-completeness of the representation naturally favours sparse coefficient vectors. However, in order to enhance sparsity of the coefficient vector, one can add a sparsity penalty term to the negative log-likelihood objective. By setting the penalty function to l_1 (as a convex approximation of l_0) , the solution can be achieved by minimizing the following objective function:

$$\begin{aligned} \arg\min_{\boldsymbol{\theta}} \quad & -L(\boldsymbol{\theta}|\mathbf{m}) + \beta \left\|\boldsymbol{\theta}\right\|_1, \\ & \textit{such that} \quad \boldsymbol{\theta} \geq 0 \end{aligned} \tag{3.25}$$

Using Eq.3.8, the surrogate function for this objective function can be written as:

$$\begin{aligned} M(\boldsymbol{\theta}|\boldsymbol{\theta}^n, \mathbf{m}) = -\sum_{l=1}^{L} s'_l(\theta_l^{EM} ln(\theta_l) - \theta_l) + \beta \left\|\boldsymbol{\theta}\right\|_1, \\ \mathbf{s}' = \mathbf{\Phi'}^T \mathbf{A}^T \vec{\mathbf{1}}, \\ \boldsymbol{\theta}^{EM} = \frac{\boldsymbol{\theta}^n}{\mathbf{s}'} \mathbf{\Phi'}^T \mathbf{A}^T \frac{\mathbf{m}}{\mathbf{A}\mathbf{\Phi'}\boldsymbol{\theta}^n + \mathbf{r}} \end{aligned} \tag{3.26}$$

3.3.3.2 The algorithm for sparse re-parameterized MAP

To benefit from ADMM, we split the variable in the data fidelity term and the sparsity penalty term by introducing an auxiliary variable $\mathbf{z}$ and writing the optimization in Eq.3.26 in the following constrained setting:

$$\begin{aligned} \min \quad & -L(\boldsymbol{\theta}|\mathbf{m}) + \beta \left\|\mathbf{z}\right\|_1, \\ \text{subject to} \quad & \mathbf{z} = \boldsymbol{\theta}, \boldsymbol{\theta} \geq 0. \end{aligned} \tag{3.27}$$

Algorithm 2: ADMM for sparse ML

Initialize: $n = 0, \boldsymbol{\theta}^0 = \vec{\mathbf{1}}, \mathbf{y}^0 = \vec{\mathbf{0}}, \mathbf{z}^0 = \vec{\mathbf{0}}, s = \boldsymbol{\Phi}'^T \mathbf{A}^T \vec{\mathbf{1}}, \rho^0 = 1$;
while *not converged* **do**
  while *not converged* **do**
    $\boldsymbol{\theta}^{EM} = \frac{\boldsymbol{\theta}^n}{\mathbf{s}} \boldsymbol{\Phi}'^T \mathbf{A}^T \frac{\mathbf{m}}{\mathbf{A}\boldsymbol{\Phi}'\boldsymbol{\theta}^n + \mathbf{r}} \quad \mathbf{b} = \mathbf{s} - \rho^n(\mathbf{z}^n - \mathbf{y}^n)$,
    $\boldsymbol{\theta}^{n+1} = \frac{2\mathbf{s}\boldsymbol{\theta}^{EM}}{\mathbf{b} + \sqrt{\mathbf{b}^2 + 4\rho^n \mathbf{s}\boldsymbol{\theta}^{EM}}}$,
  end
  $\mathbf{z}^{n+1} = \max((\boldsymbol{\theta}^{n+1} + \mathbf{y}^n) - \frac{\beta}{\rho}, \mathbf{0})$
  $\mathbf{y}^{n+1} = \mathbf{y}^n + (\boldsymbol{\theta}^{n+1} - \mathbf{z}^{n+1})$
  ρ^{n+1} is updated according to Eq. 3.35
  $n = n + 1$
end

The augmented Lagrangian for the above equation can be written as :

$$\min_{\mathbf{z} \in \mathbb{R}^+} \quad -L(\boldsymbol{\theta}|\mathbf{m}) + \beta \|\mathbf{z}\|_1 + \frac{\rho}{2} \|\boldsymbol{\theta} - \mathbf{z} + \mathbf{y}\|_2^2 \tag{3.28}$$

where $\mathbf{y}$ is the Lagrange multiplier vector and ρ is a scalar penalty parameter. ADMM is used to solve the above optimization using the following iterative procedure:

$$\boldsymbol{\theta}^{n+1} = \arg\min_{\boldsymbol{\theta}} \; -L(\boldsymbol{\theta}|\mathbf{m}) + \frac{\rho}{2} \|\boldsymbol{\theta} - \mathbf{z}^n + \mathbf{y}^n\|_2^2 \tag{3.29}$$

$$\mathbf{z}^{n+1} = \arg\min_{\mathbf{z}} \; \beta \|\mathbf{z}\|_1 + \frac{\rho}{2} \|\boldsymbol{\theta}^{n+1} - \mathbf{z} + \mathbf{y}^n\|_2^2 \tag{3.30}$$

$$\mathbf{y}^{n+1} = \mathbf{y}^n + (\boldsymbol{\theta}^{n+1} - \mathbf{z}^{n+1}) \tag{3.31}$$

Solving subproblem 3.29 : the surrogate function used for ML estimation is based on the EM algorithm and can be formulated as:

$$\begin{aligned} \boldsymbol{\theta}^{n+1} &= \arg\min_{\boldsymbol{\theta}} \quad \sum_{l=1}^{L} s'_l(\theta_l - \theta_l^{EM} ln(\theta_l)) + \frac{\rho}{2} \|\boldsymbol{\theta} - \mathbf{z}^n + \mathbf{y}^n\|_2^2 \\ &= \sum_{l=1}^{L} s'_l(\theta_l - \theta_l^{EM} ln(\theta_l)) + \frac{\rho}{2} \sum_{l=1}^{L} (\theta_l - (\mathbf{z}_l^n - \mathbf{y}_l^n))^2 \end{aligned} \tag{3.32}$$

Taking the derivative with respect to θ_l and setting it to zero leads to the following

equation;

$$\rho\theta_l^2 + \theta(s_l' - \rho(z_l^n - y_l^n)) - s_l'\theta_l^{EM} = 0 \tag{3.33}$$

Solving this equation with respect to θ_l results in the following update:

$$\begin{aligned} \theta_l^{n+1} &= \frac{2s_l'\theta_l^{EM}}{b + \sqrt{b^2 + 4\rho s_l'\theta_l^{EM}}} \\ \text{where} \quad b &= s_l' - \rho(z_l^n - y_l^n) \end{aligned} \tag{3.34}$$

Solving subproblem 3.30 : this is an l_1-regularized non-negative least squares minimization which can be solved using the non-negative variant of the soft-thresholding operator (Daubechies et al. 2003) defined as:

$$S_\beta^+(\boldsymbol{\theta}^{n+1} + \mathbf{y}^n, \frac{1}{\rho}) \quad = max((\boldsymbol{\theta}^{n+1} + \mathbf{y}^n) - \tfrac{\beta}{\rho}, 0)$$

The penalty parameter ρ of the ADMM algorithm is adaptively changed for each iteration based on the method proposed in (Boyd et al. 2011):

$$\rho^{n+1} = \begin{cases} T\rho^n & \text{if } \left\|\mathbf{e}_{primal}^n\right\|_2 > \mu \left\|\mathbf{e}_{dual}^n\right\|_2 \\ \rho^n / T' & \text{if } \left\|\mathbf{e}_{dual}^n\right\|_2 > \mu \left\|\mathbf{e}_{primal}^n\right\|_2 \\ \rho^n & \text{otherwise} \end{cases} \tag{3.35}$$

where,

$$\mathbf{e}_{primal}^n = (\boldsymbol{\theta}^n - \mathbf{z}^n), \quad \mathbf{e}_{dual}^n = -\rho(\mathbf{z}^n - \mathbf{z}^{n-1}) \tag{3.36}$$

are the primal and dual residuals.

3.3.4 Implementation

3.3.4.1 Obtaining the modified MR image

The T1-weighted MR image is non-rigidly registered to a template space to perform the white matter (WM), grey matter (GM) and cerebrospinal fluid (CSF) segmentation. The resulting segmentations are then registered back to the native space of the MR image. The WM and GM segmentations are then used to modify the T1-weighted MR image using Eq.3.21. Then mutual information based rigid registration is used to register the modified T1-weighted MR image to the PET image reconstructed using MLEM (or OSEM). The resulting MR image is then re-sampled to the PET space to have the same voxel size. The registrations and segmentations are performed using the MINC toolkit (`http://www.bic.mni.mcgill.ca/ServicesSoftware/ServicesSoftwareMincToolKit`).

3.3.4.2 Parameter selection

One can note that when increasing the number of clusters, patches in each cluster become more similar and hence fewer dictionary atoms are needed to represent the patches associated with that cluster. In other words, the number of dictionary atoms needed to represent a cluster is inversely related to the total number of clusters.

On the other hand, the number of dictionary atoms needed to describe the underlying structure within a database of patches is directly related to the size of each patch. For simplicity, in our implementation, the number of atoms learned from every cluster is identical. Using this simplification, and considering the aforementioned information one can estimate the number of atoms to be learned from each cluster as a function of patch size and the total number of clusters:

$$N = \frac{M}{C}\delta \tag{3.37}$$

where N is the number of atoms learned per cluster, M is the number of pixels or voxels

in the patch (patch size), C is the total number of clusters and δ is a constant parameter. We found that setting δ to 20 works well for the experiments performed in this study.

3.4 Simulation Studies

For quantitative evaluation of the proposed methods the ensemble-based normalized root mean squared error (n-RMSE) is computed for each pixel within an ROI across the realizations and the results are averaged within the ROI.

$$\text{Mean pixel-based n-RMSE within ROI} = \frac{1}{|ROI|}\sum_{j\in\text{ROI}}\frac{\sqrt{\sum_{r=1}^{R}(x_j^r - x_j^{true})^2/R}}{x_j^{true}} \tag{3.38}$$

where R is the number of realizations, $\mathbf{x}^r$ is the reconstructed image of the r'th realization of the measurement data and $\mathbf{x}^{true}$ is the ground truth. $|ROI|$ is the number of pixels in the ROI.

The mean contrast recovery coefficient (CRC) of the lesions embedded in the simulation data is used as another figure of merit. The following shows the formula used for obtaining the CRC of the r'th realization :

$$CRC^r = \frac{|\frac{1}{|L|}\sum_{j\in L}x_j^r - \frac{1}{|B|}\sum_{j\in B}x_j^r|}{|\frac{1}{|L|}\sum_{j\in L}x_j^{true} - \frac{1}{|B|}\sum_{j\in B}x_j^{true}|} \tag{3.39}$$

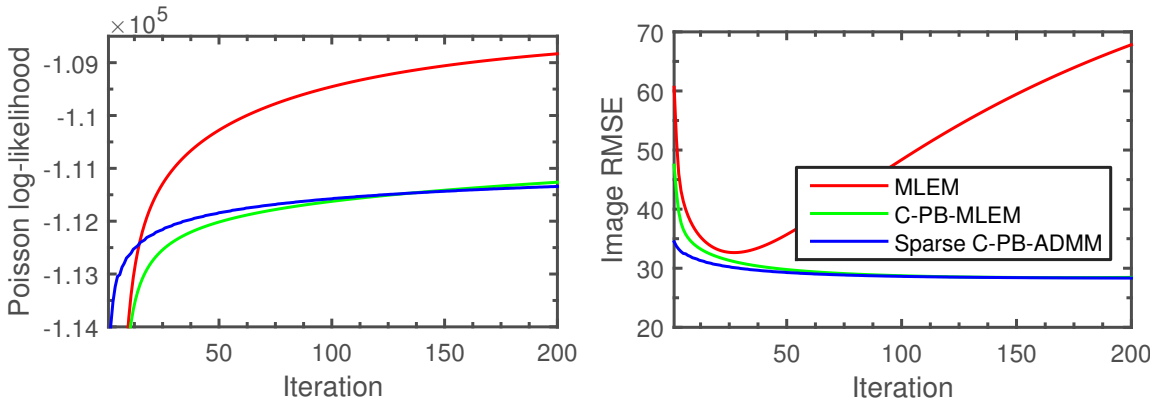
where L is the lesion and B is the background region.

3.4.1 2D phantom

3.4.1.1 Setup

A simulation study was performed using the BrainWeb phantom (`http://brainweb.bic.mni.mcgill.ca/brainweb`). One transverse slice of a realistically simulated T1-weighted MR image was re-sampled to PET image space.

For PET data simulation, the corresponding segmented brain was also re-sampled



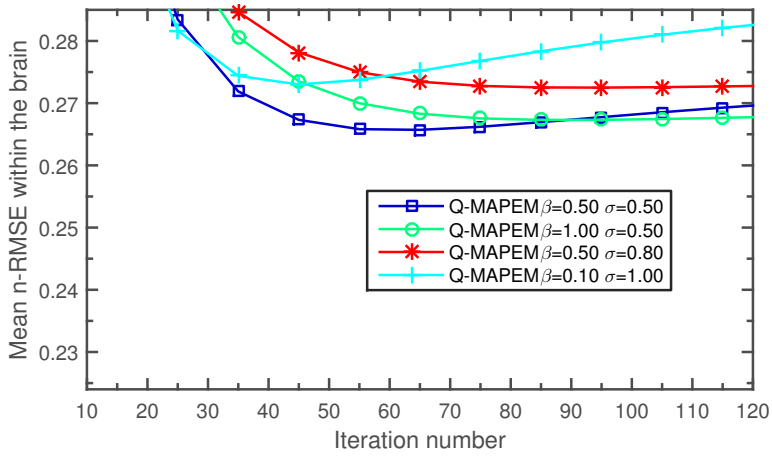
(a) (b)

Figure 3.4: a)The log-likelihood of the estimated parameters given a noisy realization of the 2D phantom sinogram data for MLEM, C-PB-MLEM and sparse C-PB-ADMM as a function of iteration. b) The RMSE of the estimated parameters as a function of iteration.

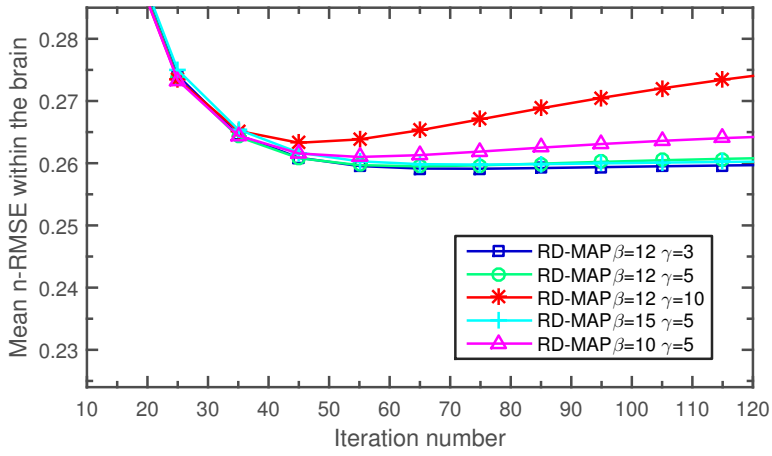
to the PET space. Then radioactivity distribution values from a PET SORTEO (Reilhac et al. 2006) (`http://sorteo.cermep.fr/home.php`) [^{18}F]fludeoxyglucose (FDG) simulation were used to find a realistic count rate in grey matter and white matter of the corresponding transverse slice.

In order to investigate the performance of the proposed method in regions where the PET and MR images do not agree, three lesions were embedded in the PET ground truth. More specifically, two hot lesions with 50% activity increase relative to their background (WM and GM) were embedded in the ground truth. The area of the hot lesion in the GM was 404 pixels and the area of the lesion in WM was 505 pixels. One cold lesion with 70% activity decrease relative to its background (GM) and total area of 395 pixels was also added to the image. The exact known boundaries of the lesions were used for calculation of the CRC values. The MR slice used for dictionary learning as well as the ground truth PET with embedded lesions are shown in figure 3.7.

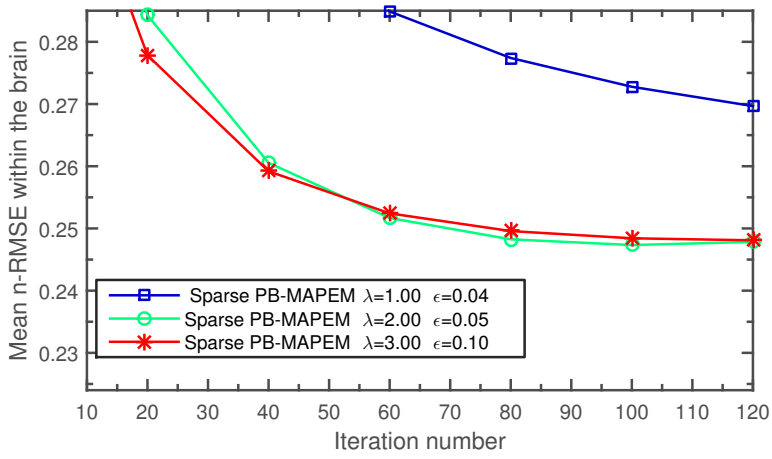
The ground truth was then forward projected using a 2D scanner model with 256 radial bins and 288 azimuthal angles. Linear attenuation coefficients equal to 0.099



(a)

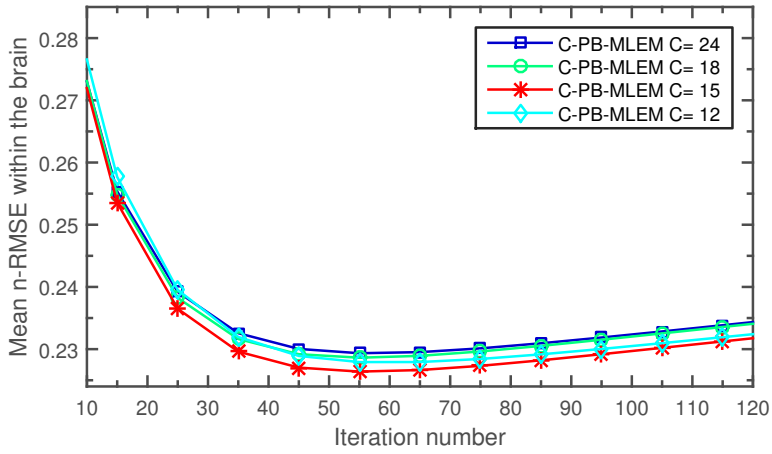


(b)

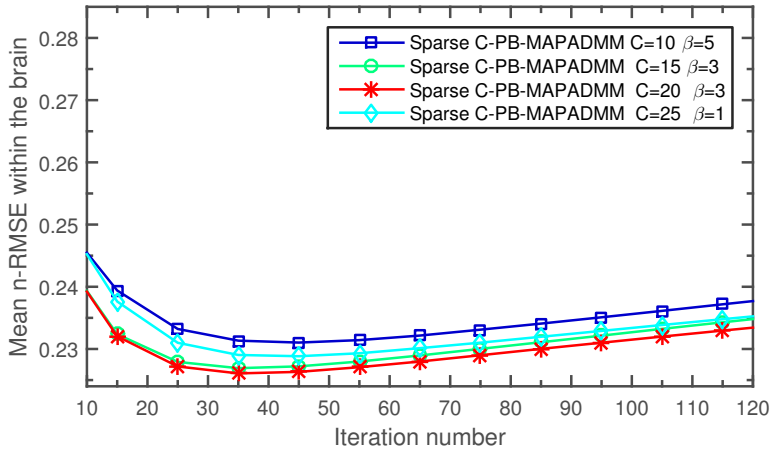


(c)

Figure 3.5: Mean pixel-based n-RMSE as a function of iteration within the brain for different parameters of five different methods: (a) Q-MAP, (b)RD-MAP and (c)sparse PB-MAPEM.



(a)



(b)

Figure 3.6: Mean pixel-based n-RMSE as a function of iteration within the brain for different parameters of five different methods: (a) Q-MAP, (b)RD-MAP, (c)sparse PB-MAPEM, (a) C-PB-MLEM and (b) sparse C-PB-ADMM.

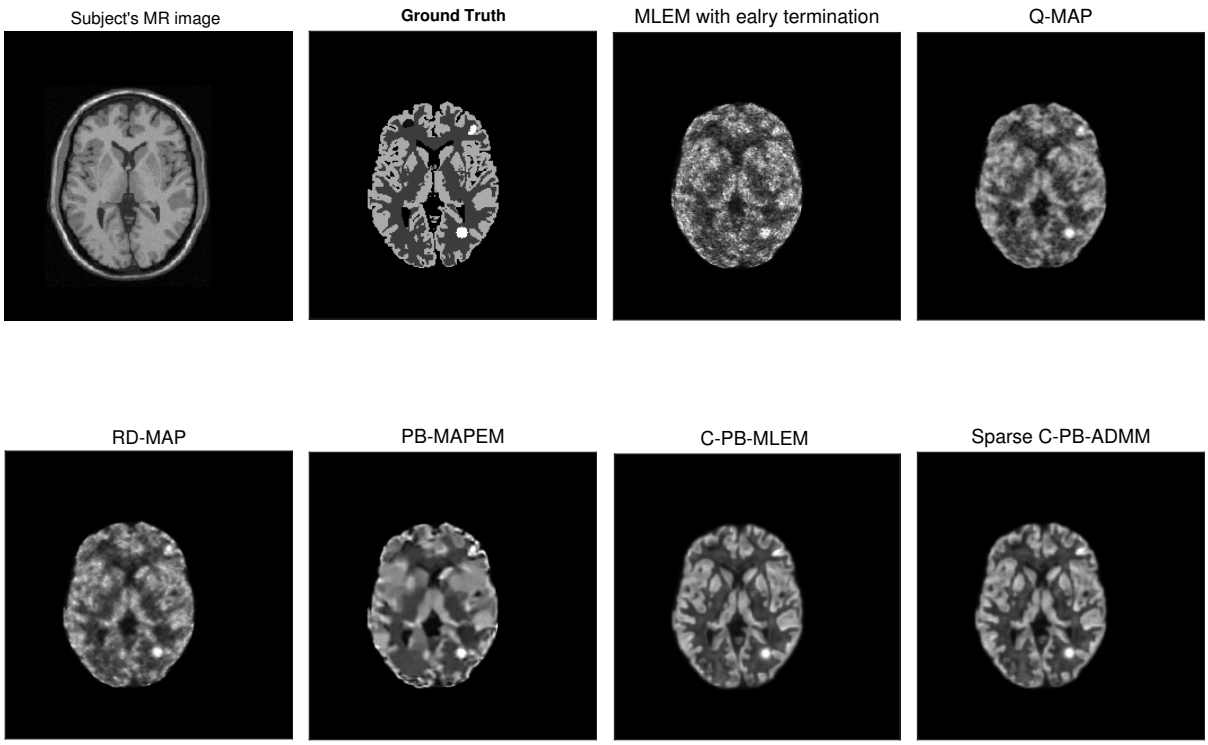


Figure 3.7: Sample reconstructed images using different methods at their best iteration and with tuned parameters.

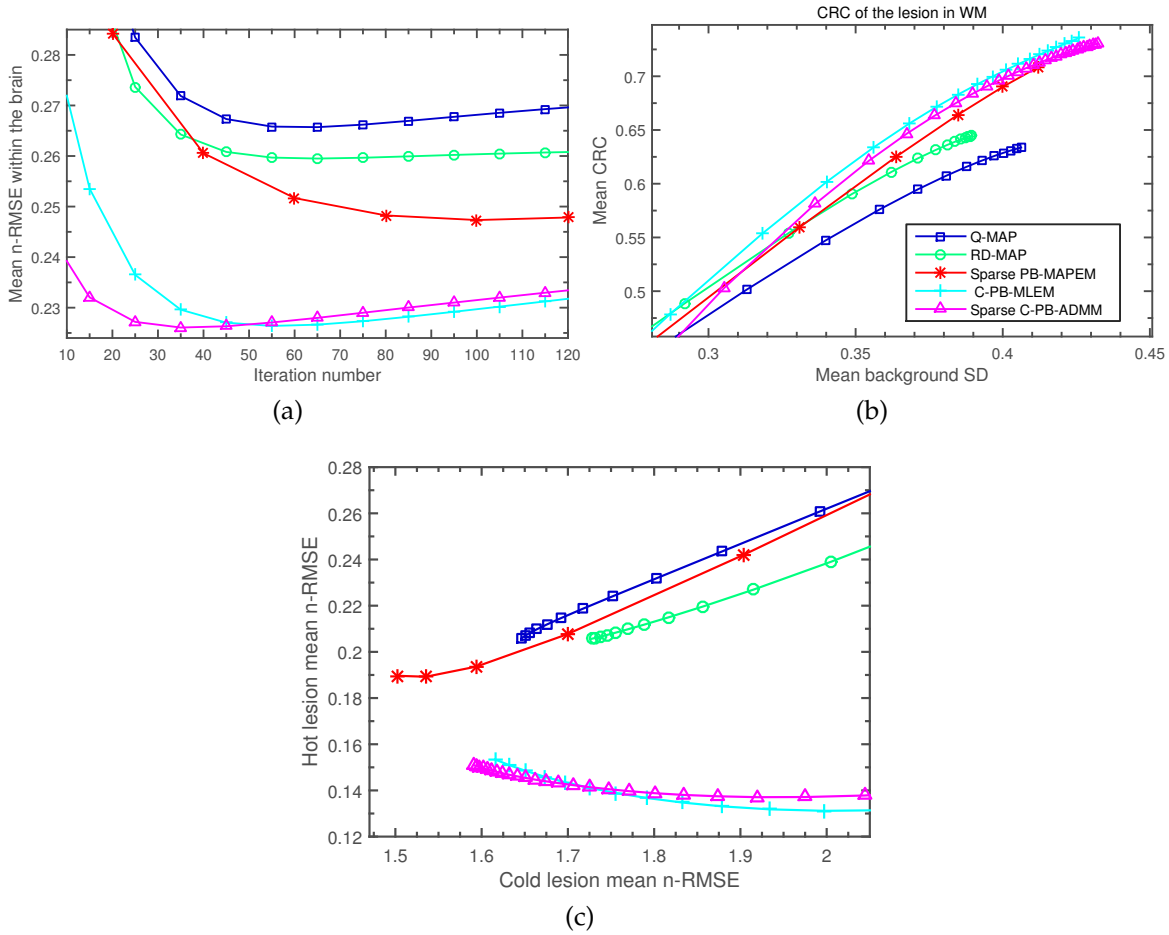


Figure 3.8: Different figures of merit for Q-MAP, RD-MAP, sparse PB-MAPEM, C-PB-MLEM and sparse C-PB-ADMM with their best parameters, leading to minimum n-RMSE within the brain. a) Mean n-RMSE within the brain as a function of iteration. b) Mean CRC of the lesion in WM as a function of mean standard deviation in WM. c) Mean n-RMSE of the hot lesion in GM versus mean n-RMSE of the cold lesion in GM. Mean pixel-based figures of merit for Q-MAP, RD-MAP, sparse PB-MAPEM, C-PB-MLEM and sparse C-PB-ADMM with their best parameters.

cm^{-1} for GM and WM, 0.095 cm^{-1} for fat, CSF, skin and connective tissue, 0.14 cm^{-1} for skull and 0 cm^{-1} for the air were also modelled in the forward projection. A heavily smoothed sinogram with total activity equal to 25 percent of the forward projected true image was added to the forward projected image to provide a simplified simulation of scatter and randoms estimates. Poisson noise was then introduced to generate 20 realizations each having an expected total number of events equal to 300,000. Corrections for attenuation, scatter and random events are performed using their exact mean values in the reconstruction algorithm. All images were represented by 256×256 pixels of side length 1.219 mm.

The same T1-weighted MR image used to generate the PET data was used to extract basis functions. Before that, MINC tools (`https://github.com/BIC-MNI`) were utilized in order to obtain tissue segmentation.

3.4.1.2 Results

The proposed methods were tested for different patch sizes. The patch size of 6 by 6 produced images with lowest total RMSE and thus was used in our experiments.

Figure 3.4 left shows the log likelihood of the parameters of interest given a noisy simulation of sinogram data as a function of iteration. For MLEM the parameters of interest are the image values x whereas for the C-PB-MLEM and sparse C-PB-ADMM methods the parameters of interest are the coefficients in θ. The number of clusters and hence the number of dictionaries learned from the modified MR image is 15. The β parameter for sparse C-PB-ADMM is set to 0.003.

This figure shows that the convergence rate of sparse C-PB-ADMM is higher than that of C-PB-MLEM. Once the dictionary is learned, the time needed to perform 50 iterations of C-PB-MLEM, C-PB-ADMM and MLEM on a machine with two 2.80GHz 10-core processors (Intel Xenon 2.80) and 64 GB system memory is 7, 9.3 and 2.1 minutes respectively. Note that even though the computational cost of each iteration of sparse C-PB-ADMM is higher compared to C-PB-MLEM because of its faster conver-

gence rate, fewer iterations are needed to achieve the minimum RMSE value (as shown in figure 3.4 right)

In this work, the performance of the proposed methods is compared to Q-MAP, RD-MAP and sparse PB-MAPEM methods. n-RMSE and mean CRC values are computed across 20 realizations. Figure 3.6 shows the effect of different parameters on these methods. To enhance readability, only a few of the tested parameters including the ones leading to lowest mean pixel-based n-RMSE within the brain are shown in the figure.

Figure 3.6 top left shows the effect of the σ value of a 7 by 7 truncated Gaussian kernel used for weight calculation and the β penalty parameters in Q-MAP. In order to tune the parameters, the σ value was ranged from 0.05 to 3 and the β value was varied from 0.01 to 5.

Figure 3.6 top right shows the effect of β and γ parameters for RD-MAP method. Both β and γ parameter were varied from 1 to 20. Figure 3.6 middle row shows the effect of λ and ε (error tolerance for the OMP method) parameters used in sparse PB-MAPEM on the accuracy of the reconstructed images. For sparse PB-MAPEM, as suggested by (Chen et al. 2015), the number of dictionary atoms is set to 144 and the patch size is set to 7 b 7. The λ value was tested for values of 1, 1.5, 2,2.5 and 3. The ε parameter was varied from 0.01 to 0.5.

In the bottom row, the right figure shows the effect of the number of clusters on the C-PB-MLEM algorithm. The left figure shows the influence of the number of clusters and the sparsity parameter β on the accuracy of the sparse-PB-ADMM method.

For both C-PB-MLEM and sparse C-PB-ADMM, the number of dictionary atoms learned per cluster is set according to the Eq. 3.37. As mentioned earlier, the sparsity parameter in the NMF-l_0 algorithm was set to 1/10th of the number of atoms learned per cluster. Both methods were tested for number of clusters C equal to 5, 10, 15, 20 and 25. The tested β value for sparse C-PB-ADMM ranged from 0.001 to 0.05.

Figure 3.7 shows the reconstruction results of a noisy realization using MLEM after 20 iterations (leading to minimum mean n-RMSE within the brain), Q-MAP, RD-MAP, sparse PB-MAPEM, the proposed C-PB-MLEM and sparse-C-PB-ADMM methods with their best parameters and at their best iteration found based on the lowest mean pixel-based n-RMSE within the brain. We can see that using the proposed methods substantially improved the image quality over the conventional MLEM, Q-MAP, RD-MAP and sparse PB-MAPEM.

Figure 3.8, shows the performance of different regularization methods with their best parameters based on the minimum mean pixel-based n-RMSE within the brain. The top left figure shows the mean pixel-based n-RMSE within the brain as a function of iteration. The top right figure shows the mean CRC of the lesion in white matter versus background noise across multiple realizations. The bottom figure shows the mean pixel-based n-RMSE within the hot lesion in the grey matter versus the mean pixel-based n-RMSE within the cold lesion in the grey matter.

The results indicate that both C-PB-MLEM and sparse C-PB-ADMM outperform the other regularization techniques. Sparse C-PB-ADMM gives a slightly lower best mean pixel-based n-RMSE value within the brain at the best iteration number compared to C-PB-MLEM, while the opposite is the case for mean CRC of the hot lesion in white matter. The mean pixel-based n-RMSE of the hot versus cold lesion shows that sparse PB-MAPEM provides a lower error within the cold lesion while the error in hot lesion is higher than that of C-PB-MLEM. Sparse C-PB-ADMM seems to provide a slightly better n-RMSE in the hot versus cold lesion compared C-PB-MLEM method. In summary C-PB-MLEM and sparse C-PB-ADMM seem to have a very similar performance.

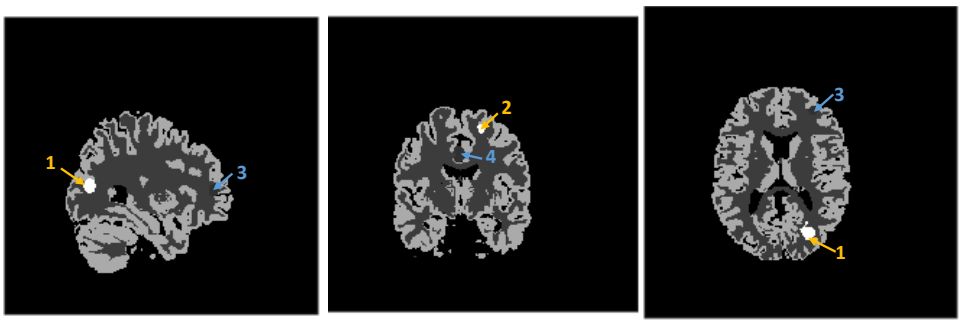


Figure 3.9: 3D FDG phantom with 4 embedded lesions.

3.4.2 3D phantom

3.4.2.1 Setup

In order to explore the extendibility of the proposed method to 3D data, the same T1-weighted MR image used in the previous section is utilized. This time however, instead of one slice, the whole 3D volume is used to generate a 3D FDG phantom. Four lesions are simulated in the FDG PET ground truth (shown in figure 3.9). Lesions 2, 3 and 4 are embedded in the GM and follow the cortical boundaries while lesion 1 is embedded in the WM. In lesions 1 and 2 the activity is increased by 50% relative to their background (WM and GM respectively). Lesions 3 and 4 have 70% activity reduction relative to the activity in the GM. The total volumes of lesions 1, 2, 3 and 4 are 515, 260, 283 and 553 voxels respectively.

The same procedure described in the previous section but in 3D is used to generate highly noisy sinogram data of this 3D phantom with 10 million counts. The size of the image is 256 by 256 by 207 voxels of 1.219mm length. The forward projection is based on the HRRT scanner geometry consisting of 2209 sinograms with 288 view angles and 256 projection bins per view.

The patch size is set to 4 by 4 by 4 voxels. In order to decrease the number of patches extracted from the image and hence reduce the computational complexity, the distance between corresponding voxels in the neighbouring patches is set to two voxels.

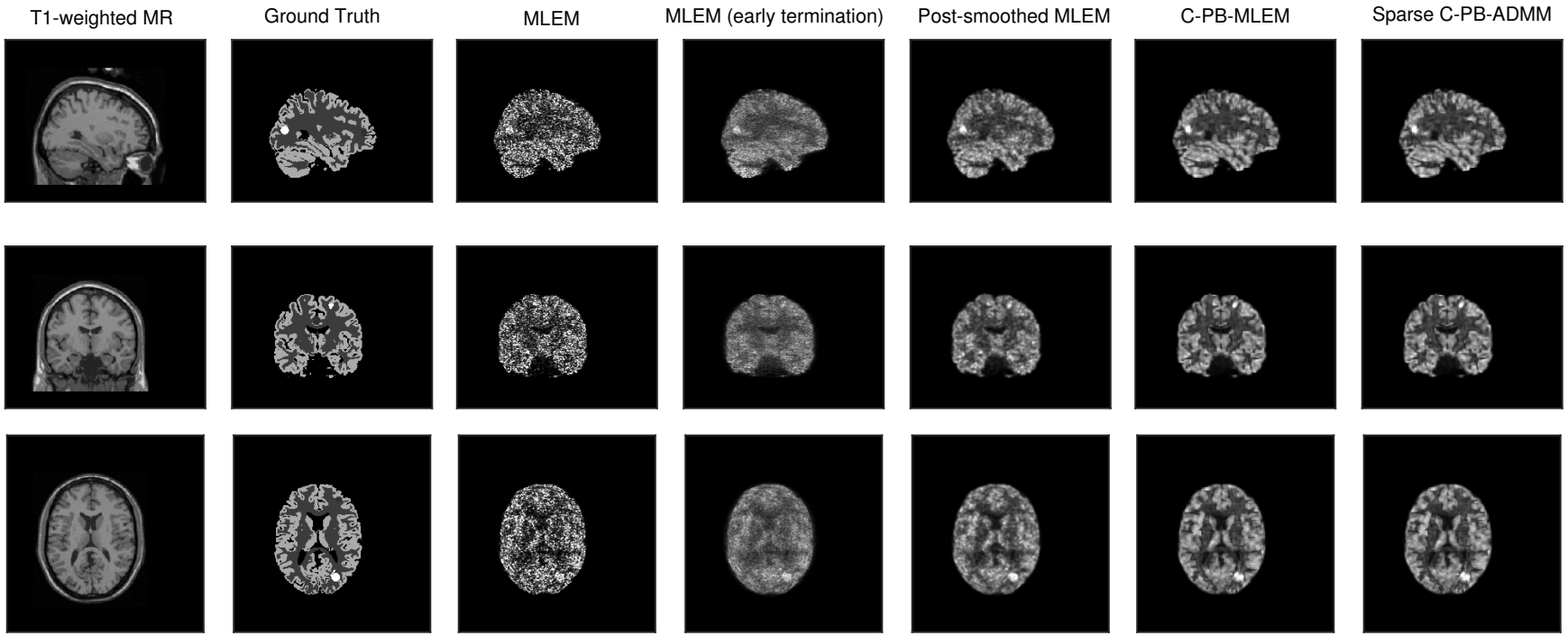


Figure 3.10: One example realization of the simulated 3D FDG data reconstructed using different methods shown for sagittal, coronal and transverse views. All methods are shown at their best iteration leading to minimum RMSE value.

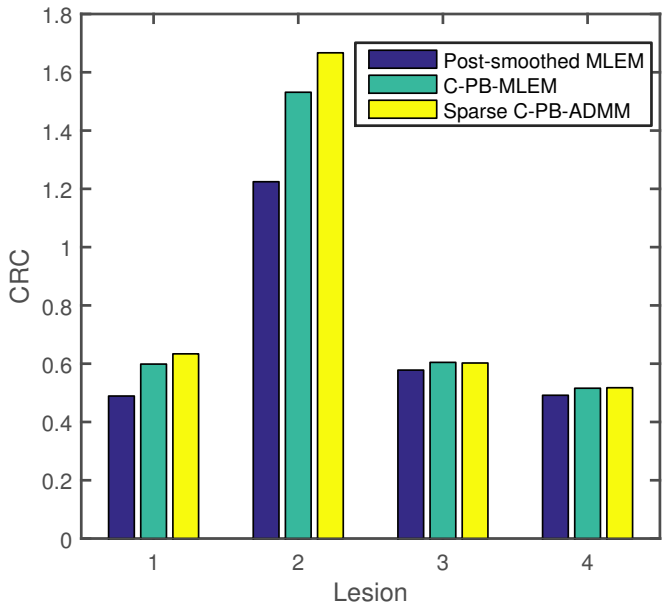


Figure 3.11: The CRC values of the 4 lesions indicated in figure 3.9 using post-smoothed MLEM after 60 iteration, C-PB-MLEM and sparse C-PB-ADMM . All methods are shown at their best iteration leading to minimum RMSE value.

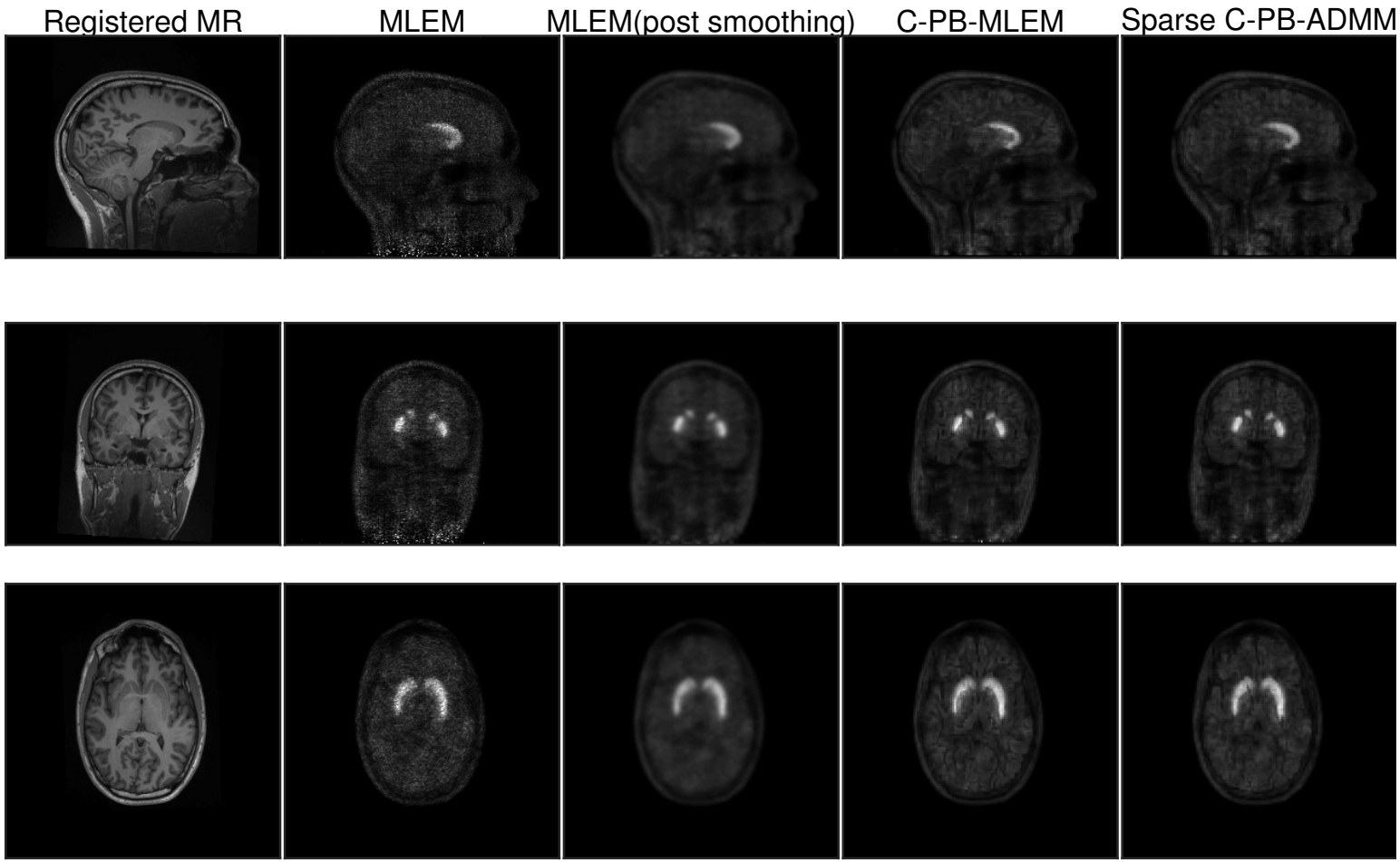


Figure 3.12: Reconstructed images of a 40 minute frame [^{11}C]raclopride scan of the real subject by different algorithms. Each column shows the resulting slices of a 3-D image of an algorithm in transverse, sagittal, and coronal views.

Figure 3.10 shows the transverse, sagittal, and coronal views of the reconstructed image of the noisy simulation using different methods. MLEM is shown after 50 iterations as well as after 10 iterations leading to minimum RMSE. Post-smoothed MLEM with a Gaussian kernel of size 5 by 5 by 5 voxels is tested for different σ values ranging from 0.5 to 3 and number of iterations ranging from 10 to 80. The one leading to minimum RMSE is shown here (σ equal to 1 voxel at iteration 60). The number of clusters used for C-PB-MLEM and sparse C-PB-ADMM is 20 and 15 respectively. The number of atoms learned per cluster is found based on Eq.3.37. C-PB-MLEM is shown after 35 iterations and sparse C-PB-ADMM is shown after 20 iterations. These iteration numbers were selected as they achieve minimum mean n-RMSE within the brain. Both C-PB-MLEM and sparse C-PB-ADMM significantly improve the image quality and provide a good contrast in the lesions.

To further evaluate the ability of the methods in improving the contrast in the lesions, the CRC values of the 4 lesions at the mentioned parameters and number of iterations (leading to minimum RMSE) are show in figure 3.11. For all four lesions the proposed methods outperform post-smoothed MLEM. Using sparse C-PB-ADMM leads to a greater CRC value compared to C-PB-MLEM for all lesions except for lesion 3, where the CRC value is slightly smaller.

3.5 Application to real subject data

In this section, the MLEM, C-PB-MLEM and sparse C-PB-ADMM methods are used to reconstruct images from real data. The HRRT list-mode data is obtained from a 370 MBq injection of [^{11}C]raclopride to a healthy human participant. The inter-frame motion was negligible and so not corrected for. Attenuation and normalization factors as well as scatter and randoms estimates obtained from the manufacturer's software were incorporated within the reconstruction. The duration of the reconstructed frame is 40 minutes starting 20 minutes after injection. The subject provided written informed consent and the ethics were approved under the McGill Faculty of Medicine

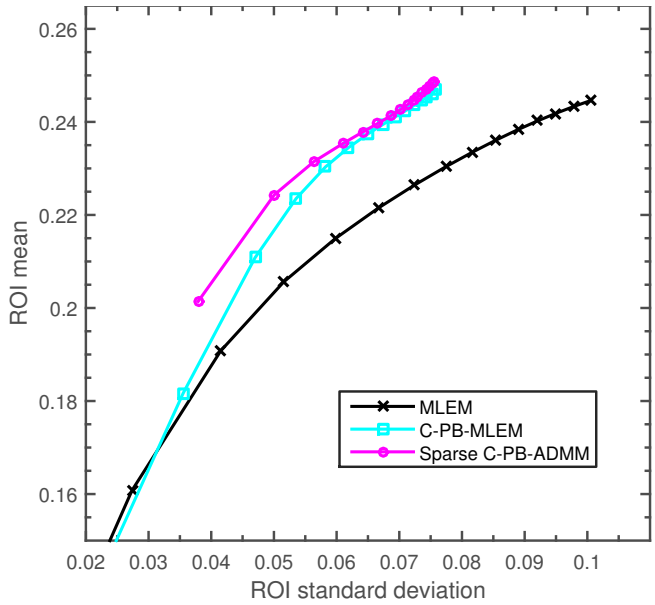


Figure 3.13: The mean activity versus the standard deviation in the right caudate in the real data reconstructed using MLEM and C-PB-MLEM and sparse C-PB-ADMM methods for increasing iterations.

Institutional Review Board.

For patch-based basis vector extraction, the process described in the implementation section is used to obtain a registered re-sampled modified MR image from the subject's T1-weighted MR image. Then overlapping patches of size 4 by 4 by 4 with corresponding voxels in the neighbouring patches having a distance of 2 voxels, are extracted and used to obtain basis vectors. The number of clusters in C-PB-MLEM and C-PB-ADMM is set to 15 and 20 respectively. The β value for sparse C-PB-ADMM is set to 0.003.

Figure 3.12 shows the registered re-sampled T1-weighted MR image as well as reconstructed images of this data using MLEM, C-PB-MLEM and sparse C-PB-ADMM algorithms. MLEM is shown after 50 iterations. Post-smoothed MLEM is obtained by applying a Gaussian kernel of size 5 by 5 by 5 and σ equal to 1 voxel to the reconstructed image at iteration 50. The images reconstructed using C-PB-MLEM and

saprse C-PB-ADMM are shown after 25 iterations. Figure 3.13 compares the mean activity in the right caudate excluding the edges versus standard deviation within this region with increasing iteration for MLEM, C-PB-MLEM and sparse C-PB-ADMM. One can note that both proposed methods can deliver a higher ROI mean within this region for every standard deviation.

3.6 Conclusion

In this work, a novel re-parameterization framework which uses patch-based basis functions learned from a registered prior image for PET image reconstruction is devised. In addition, a method for finding sparse coefficients for a re-parameterized Poisson log-likelihood objective function is proposed. The two approaches are then combined together to find sparse coefficients for patch-based basis functions. The proposed methods are validated by computer simulations. The results indicate that the proposed methods can achieve better quality images compared to the RD-MAP, Q-MAP and sparse PB-MAPEM approaches.

The core of this work is to use non-negative dictionaries extracted from prior images to parameterize the PET image reconstruction task. Although this work only looks into the use of T1-weighted MR images for basis vector extraction, once corrected for inverted intensities (if needed), any prior image with enough boundary information can be utilized in this framework. It is important to note that the proposed framework does permit flexibility, as demonstrated by the reconstruction of features not present in the MR (the lesions in the ground truth PET distribution) -so abnormal features in the radioactivity distribution can still be successfully reconstructed.

Note that there is a trade off between the amount of structural detail and the insensitivity of the method to misalignment. This trade off can be controlled by the number of clusters. In fact, in order to avoid sensitivity to mis-registration, one can decrease the number of clusters while increasing the number of atoms in each cluster. In the extreme case, an identical dictionary learned from the entire image can be utilized to

represent each and every patch in the image at the expense of increased variance. In the other extreme, we can partition the prior image patches into many clusters; leading to many dictionaries each with few atoms. This will result in low variance at the expense of increase in bias and becoming sensitive to slight mis-registrations.

Recently, advanced non-linear registration techniques have been used to construct 4D PET atlases (Bieth et al. 2013). A future study investigating the use of a PET atlas for basis extraction and reconstructing PET images of the same radiotracer would be of great interest.

Acknowledgement

This work was supported in part by the Natural Sciences and Engineering Research Council (NSERC) (RGPIN 387067-10), the Canada Research Chairs program and the UK Engineering and Physical Sciences Research Council (grant EP/M020142/1).

3.7 Appendix: Alternate direction method of multipliers

The alternate direction method of multipliers (ADMM) is a powerful optimization method introduced by (Glowinski & Marroco 1975) that has recently gained attention in many image processing applications such as image denoising and restoration, e.g. (Figueiredo & Bioucas-Dias 2010, Afonso et al. 2010, Matakos et al. 2013). Through the variable splitting method, ADMM allows the optimization of the sum of two or more functionals to be done separately in an alternating manner. For example let us consider $f(\mathbf{x}) + g(\mathbf{Ax})$ as the objective function to be optimized. By introducing an auxiliary variable $\mathbf{z} = \mathbf{Ax}$ the problem can be written as a constrained minimization :

$$\begin{aligned} \min \quad & f(\mathbf{x}) + g(\mathbf{z}) \\ \text{such that} \quad & \mathbf{Ax} - \mathbf{z} = 0 \end{aligned} \tag{3.40}$$

ADMM uses the augmented Lagrangian method to relax the above problem to :

$$L_p(\mathbf{x}, \mathbf{z}, \mathbf{u}) = f(\mathbf{x}) + g(\mathbf{z}) + \mu^T(\mathbf{Ax} - \mathbf{z}) + \frac{\rho}{2} \|\mathbf{Ax} - \mathbf{z}\|^2 \tag{3.41}$$

where μ is the Lagrangian multiplier and $\rho > 0$ is the penalty parameter. ADMM then solves the above optimization problem in an iterative alternating manner as follows:

$$\mathbf{x}^{k+1} = \arg\min_{\mathbf{x}} \quad L_p(\mathbf{x}, \mathbf{z}^k, \mu^k) \tag{3.42}$$

$$\mathbf{z}^{k+1} = \arg\min_{\mathbf{z}} \quad L_p(\mathbf{x}^{k+1}, \mathbf{z}, \mu^k) \tag{3.43}$$

$$\mu^{k+1} = \mu^k + \rho(\mathbf{Ax}^{k+1} - \mathbf{z}^{k+1}) \tag{3.44}$$

For a thorough review of the ADMM method see (Boyd et al. 2011).

References

Afonso, M. V., Bioucas-Dias, J. M. & Figueiredo, M. A. (2010), 'Fast image recovery using variable splitting and constrained optimization', *Image Processing, IEEE Transactions on* **19**(9), 2345–2356.

Ahn, S., Ross, S. G., Asma, E., Miao, J., Jin, X., Cheng, L., Wollenweber, S. D. & Manjeshwar, R. M. (2015), 'Quantitative comparison of osem and penalized likelihood image reconstruction using relative difference penalties for clinical pet', *Physics in medicine and biology* **60**(15), 5733.

Ardekani, B. A., Braun, M., Hutton, B. F., Kanno, I. & Iida, H. (1996), 'Minimum cross-entropy reconstruction of pet images using prior anatomical information', *Physics in Medicine and Biology* **41**(11), 2497.

Bieth, M., Lombaert, H., Reader, A. J. & Siddiqi, K. (2013), Atlas construction for dynamic (4d) pet using diffeomorphic transformations, *in* 'Medical Image Computing and Computer-Assisted Intervention–MICCAI 2013', Springer, pp. 35–42.

Bowsher, J. E., Johnson, V. E., Turkington, T. G., Jaszczak, R. J., Floyd, C. E. & Coleman, R. E. (1996), 'Bayesian reconstruction and use of anatomical a priori information for emission tomography', *Medical Imaging, IEEE Transactions on* **15**(5), 673–686.

Boyd, S., Parikh, N., Chu, E., Peleato, B. & Eckstein, J. (2011), 'Distributed optimization and statistical learning via the alternating direction method of multipliers', *Foundations and Trends® in Machine Learning* **3**(1), 1–122.

Buades, A., Coll, B. & Morel, J.-M. (2005), A non-local algorithm for image denoising, *in* 'Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on', Vol. 2, IEEE, pp. 60–65.

Chatterjee, P. & Milanfar, P. (2009), 'Clustering-based denoising with locally learned dictionaries', *Image Processing, IEEE Transactions on* **18**(7), 1438–1451.

Chatterjee, P. & Milanfar, P. (2012), 'Patch-based near-optimal image denoising', *Image Processing, IEEE Transactions on* **21**(4), 1635–1649.

Chen, S., Liu, H., Shi, P. & Chen, Y. (2015), 'Sparse representation and dictionary learning penalized image reconstruction for positron emission tomography', *Physics in medicine and biology* **60**(2), 807.

Cheng-Liao, J. & Qi, J. (2011), 'Pet image reconstruction with anatomical edge guided level set prior', *Physics in medicine and biology* **56**(21), 6899.

Dabov, K., Foi, A., Katkovnik, V. & Egiazarian, K. (2007), 'Image denoising by sparse 3-d transform-domain collaborative filtering', *Image Processing, IEEE Transactions on* **16**(8), 2080–2095.

Daubechies, I., Defrise, M. & De Mol, C. (2003), 'An iterative thresholding algorithm for linear inverse problems with a sparsity constraint', *arXiv preprint math/0307152* .

De Pierro, A. R. (1995), 'A modified expectation maximization algorithm for penalized likelihood estimation in emission tomography.', *IEEE Transactions on Medical Imaging* **14**(1), 132.

Fessler, J. A. (1997), Grouped coordinate descent algorithms for robust edge-preserving image restoration, *in* 'Optical Science, Engineering and Instrumentation'97', International Society for Optics and Photonics, pp. 184–194.

Figueiredo, M. A. & Bioucas-Dias, J. M. (2010), 'Restoration of poissonian images using alternating direction optimization', *Image Processing, IEEE Transactions on* **19**(12), 3133–3145.

Glowinski, R. & Marroco, A. (1975), 'Sur l'approximation, par e'le'ments finis d'ordre un, et la re?solution, par pe'nalisation-dualite' d'une classe de proble'mes de dirichlet non line'aires', *ESAIM: Mathematical Modelling and Numerical Analysis-Modélisation Mathématique et Analyse Numérique* **9**(R2), 41–76.

Green, P. J. (1990), 'On use of the em for penalized likelihood estimation', *Journal of the Royal Statistical Society. Series B (Methodological)* pp. 443–452.

Hoyer, P. O. (2004), 'Non-negative matrix factorization with sparseness constraints', *The Journal of Machine Learning Research* **5**, 1457–1469.

Jacobs, F., Matej, S. & Lewitt, R. (1998), 'Image reconstruction techniques for pet', *Dep. Radiol., Univ. Pennsylvania, Philadelphia, Tech. Rep. MIPG245* .

Jiao, J., Markiewicz, P., Burgos, N., Atkinson, D., Hutton, B., Arridge, S. & Ourselin, S. (2015), Detail-preserving pet reconstruction with sparse image representation and anatomical priors, *in* 'Information Processing in Medical Imaging', Springer, pp. 540–551.

Kervrann, C. & Boulanger, J. (2006), 'Optimal spatial adaptation for patch-based image denoising', *Image Processing, IEEE Transactions on* **15**(10), 2866–2878.

Lewitt, R. M. (1990), 'Multidimensional digital image representations using generalized kaiser–bessel window functions', *JOSA A* **7**(10), 1834–1846.

Lewitt, R. M. (1992), 'Alternatives to voxels for image representation in iterative reconstruction algorithms', *Physics in Medicine and Biology* **37**(3), 705.

Matakos, A., Ramani, S. & Fessler, J. A. (2013), 'Accelerated edge-preserving image restoration without boundary artifacts', *IEEE transactions on image processing: a publication of the IEEE Signal Processing Society* **22**(5), 2019.

Novosad, P. & Reader, A. J. (2016), 'Mr-guided dynamic pet reconstruction with the kernel method and spectral temporal basis functions', *Physics in medicine and biology* **61**(12), 4624.

Nuyts, J., Bequé, D., Dupont, P. & Mortelmans, L. (2002), 'A concave prior penalizing relative differences for maximum-a-posteriori reconstruction in emission tomography', *Nuclear Science, IEEE Transactions on* **49**(1), 56–60.

Nuyts, J., Fessler, J. et al. (2003), 'A penalized-likelihood image reconstruction method for emission tomography, compared to postsmoothed maximum-likelihood with matched spatial resolution', *Medical Imaging, IEEE Transactions on* **22**(9), 1042–1052.

Peharz, R. & Pernkopf, F. (2012), 'Sparse nonnegative matrix factorization with ? 0-constraints', *Neurocomputing* **80**, 38–46.

Rangarajan, A., Hsiao, T. & Gindi, G. (2000), 'A bayesian joint mixture framework for the integration of anatomical information in functional image reconstruction', *Journal of Mathematical Imaging and Vision* **12**(3), 199–217.

Reader, A. J. & Verhaeghe, J. (2014), '4d image reconstruction for emission tomography', *Physics in medicine and biology* **59**(22), R371.

Reilhac, A., Evans, A. C., Gimenez, G. & Costes, N. (2006), 'Creation and application of a simulated database of dynamic [18f] mppf pet acquisitions incorporating inter-individual anatomical and biological variability', *Medical Imaging, IEEE Transactions on* **25**(11), 1431–1439.

Somayajula, S., Asma, E. & Leahy, R. M. (2005), Pet image reconstruction using anatomical information through mutual information based priors, *in* 'Nuclear Science Symposium Conference Record, 2005 IEEE', Vol. 5, IEEE, pp. 2722–2726.

Tahaei, M. S. & Reader, A. J. (2014), Combining different variance reduction approaches for pet image reconstruction, *in* '2014 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)', pp. 1–6.

Tang, J. & Rahmim, A. (2009), 'Bayesian pet image reconstruction incorporating anato-functional joint entropy', *Physics in medicine and biology* **54**(23), 7063.

Tang, J., Wang, Y., Yao, R. & Ying, L. (2014), Sparsity-based pet image reconstruction using mri learned dictionaries, *in* 'Biomedical Imaging (ISBI), 2014 IEEE 11th International Symposium on', IEEE, pp. 1087–1090.

Tropp, J., Gilbert, A. C. et al. (2007), 'Signal recovery from random measurements via orthogonal matching pursuit', *Information Theory, IEEE Transactions on* **53**(12), 4655–4666.

Wang, G. & Qi, J. (2012), 'Penalized likelihood pet image reconstruction using patch-based edge-preserving regularization', *Medical Imaging, IEEE Transactions on* **31**(12), 2194–2204.

Wang, G. & Qi, J. (2015), 'Pet image reconstruction using kernel method', *Medical Imaging, IEEE Transactions on* **34**(1), 61–71.

Wells, R. G., King, M. A., Simkin, P. H., Judy, P. F., Brill, A. B., Gifford, H. C., Licho, R., Pretorius, P. H., Schneider, P. B. & Seldin, D. W. (2000), 'Comparing filtered backprojection and ordered-subsets expectation maximization for small-lesion detection and localization in^ 6^ 7ga spect', *Journal of Nuclear Medicine* **41**(8), 1391–1399.

Chapter 4

Two Novel PET Image Restoration Methods Guided by PET-MR Kernels

Preface

In the previous chapter a novel PET reconstruction method was proposed that was able to reconstruct images with lower noise and better boundary information. However, the proposed method was only applicable in situation were PET measurement data is available.

In this chapter two novel post reconstruction methods for restoring PET images are proposed. Both methods use the kernel extracted from the subjects registered MR image and the median filtered PET image to guide the restoration process.

This work has been published as: Tahaei, Marzieh S., Andrew J. Reader, and D. Louis Collins. "Two novel PET image restoration methods guided by PET-MR kernels: Application to brain imaging." Medical physics 46.5 (2019): 2085-2102.

Two Novel PET Image Restoration Methods Guided by PET-MR Kernels: Application to Brain Imaging

Marzieh S. Tahaei

McConnell Brain Imaging Centre, Montreal Neurological Institute, Montreal, Canada

Dept. of Biomedical Engineering, McGill University, Montreal, Canada

marzieh.tahaei@mail.mcgill.ca

Andrew J. Reader

Dept. of Biomedical Engineering, Division of Imaging Sciences and Biomedical Engineering, King's College London, St. Thomas' Hospital, London, Uk

D. Louis Collins

McConnell Brain Imaging Centre, Montreal Neurological Institute, Montreal, Canada

Dept. of Biomedical Engineering, McGill University, Montreal, Canada

Abstract

Purpose: Post-reconstruction PET image restoration methods that take advantage of available anatomical information can play an important role in accurate quantification of PET images. However, when using anatomical information, the resulting PET image may lose resolution in certain regions where the anatomy does not agree with the change in functional activity. In this work this problem is addressed by using both MR and filtered PET images to guide the denoising process.

Methods: In this work, two novel post-reconstruction methods for restoring PET images using the subject's registered T1-weighted MR image, are proposed. The first method is based on a representation of the image using basis functions extracted from T1-weighted MR and filtered PET image. The coefficients for these basis functions are estimated using a sparsity-penalized least squares objective function. The second method is a non-iterative fast method that uses guided kernel filtering in combination with twicing to restore the noisy PET image. When applied after conventional PVE correction, these methods can be considered as voxel-based MR-guided partial volume

effect (PVE) correction methods.

Results: Using simulation analyses of [^{18}F]FDG PET images of the brain with lesions, the proposed methods are compared to other denoising methods through different figures-of-merit. The results show promising improvements in image quality as well as reduction in bias and variance of the lesions. We also show the application of the second method on real [^{18}F]FDG data.

Conclusion: Two methods for restoring PET images were proposed. The methods were evaluated on simulation and real brain images. Most MR-guided PVE correction methods are only based on segmented T1-weighted images and their accuracy is very sensitive to segmentation errors, especially in regions of abnormalities and lesions. However, both proposed methods can use the T1-weighted image without segmentation. The simplicity and the very low computational cost of the second method make it suitable for clinical applications and large data studies. The proposed methods can be naturally extended to PVE correction and denoising of other functional modalities using corresponding anatomical information.

Key words: PET, MR, denoising, post-reconstruction, sparsity, PVE correction, basis functions, twicing, multi-centre studies, brain imaging

4.1 Introduction

PET imaging plays an important role in the diagnosis, treatment and drug development for different neurological disorders. In addition, using the growing variety of radiotracers, quantitative PET imaging is now considered as a powerful and unique tool for understanding the human brain in health and disease.

Two of the sources of error in quantification of PET images include high amount of noise and low spatial resolution. Low SNR is mainly due to the low number of counts imposed by constraints on administrated activity dose, combined with limited scanner sensitivity. SNR is also affected by the acquisition time and the type of radiotracer (Strother et al. 1990, Mandelkern 1995, Carlier et al. 2014). The partial volume

effect (PVE) is due to positron range, photon-non-colinearity, limited scanner spatial resolution, tissue fraction effects and the reconstruction process (Rousset et al. 2007).

PET images reconstructed using conventional algorithms suffer from increasing amounts of noise with increasing iterations (Barrett et al. 1994). In practice noise is usually minimized by the early termination of the reconstruction algorithm or by applying Gaussian post-smoothing, which leads to further loss of resolution in the resulting images. Two approaches have been considered to address this issue: one is to use sophisticated reconstruction algorithms which incorporate regularization to reduce the noise in the resulting image (Ardekani et al. 1996, Fessler 1997, Rangarajan et al. 2000, Nuyts et al. 2003, Tahaei & Reader 2016), and the other is to apply post-reconstruction methods on PET images reconstructed using conventional algorithms (Turkheimer et al. 2008, Boussion et al. 2009).

Despite a vast body of research to develop new regularization methods that reduce the noise within reconstruction, most of these advances have not been translated into practice. One reason for the lack of use of these methods in practice is that the measurement data format in many of the current scanners are proprietary and hence only the software provided by the scanner manufacturer can be used to reconstruct images (Thielemans et al. 2012). Therefore, applying a more sophisticated reconstruction algorithm on the raw measurement data obtained from the scanner may not be feasible. In such cases, state-of-the-art post-reconstruction denoising methods can be used to reduce the noise in reconstructed PET images while preserving the resolution.

Over the past several years, advances in computing infrastructure and neuroimaging technologies have led to growth in large data repositories of research and clinical PET images. Using post-reconstruction methods for larger data sets of PET images is of great interest for several reasons: i) post reconstruction methods can be applied to reconstructed images in large multicentre studies for which raw measurement data is not available; ii) in order to improve inter-centre consistency, in many multi-centres

studies, centres are required to provide images reconstructed using a common, off the shelf reconstruction algorithm (e.g., the ordered subsets expectation maximization algorithm) (Jagust et al. 2010). Therefore, the noise in these images can only be reduced using a post-reconstruction method. iii) Computational complexity of the post-reconstruction methods is usually much lower than the integration of noise reduction techniques with reconstruction methods.

Many clinical and research PET acquisitions are accompanied by MR images of the same subject and methods that leverage the detailed anatomical information in MR images for PET image restoration (i.e., for PVE correction and denoising) are interesting. The motivation for using MR images for this purpose is that similar MR intensities correspond to similar tissues, and these tissues have a tendency to have similar functions. In fact, many PVE correction methods have used the subject's co-registered and segmented MR to reduce the PVE in the resulting images (Müller-Gärtner et al. 1992, Rousset et al. 1998, Bataille et al. 2007, Frouin, Comtat, Reilhac & Grégoire 2002).

In addition to PVE correction, anatomical information has been previously used in denoising PET images (Turkheimer et al. 2008, Chan et al. 2010, Yan et al. 2015). One important challenge in using anatomical images as a guide image for denoising PET images is that the latter may suffer from degraded resolution in regions where the anatomy does not agree with the functional activity in the image. For example a hot spot in the functional image that does not have an anatomical homologue may lose boundary information in the denoised image when using anatomy-constrained methods. This work tries to address these issues by proposing two post-reconstruction frameworks to incorporate both the subject's T1-weighted MR image and the median-filtered PET images in guiding the image denoising process. In the first method, two kernel matrices, one extracted from the MR image and the other extracted from the median-filtered PET image are used as basis functions for PET image re-parameterization. An algorithm to estimate the coefficients of these basis functions is proposed. In order

Table 4.1: Summary of the meaning of some of the symbols and abbreviations used in the text

Symbol	Description
$\mathbf{x}$	True image
$\hat{\mathbf{x}}$	The estimated denoised image
$\mathbf{y}$	Noisy image
I	Number of voxels in $\mathbf{y}$
L	Neighbourhood lenght of the kernel
h	smoothness parameter of the kernel
$\mathbf{K}$	Kernel matrix of size $I \times I$
$\mathbf{W}$	Normalized kernel matrix
$\mathbf{M}$	Normalized kernel extracted from the registered MR image
$\mathbf{P}$	Normalized kernel extracted from the median-filtered PET image
$\boldsymbol{\beta}$	Coefficient vector for P
$\boldsymbol{\alpha}$	Coefficient vector for M
λ	Sparsity parameter for Method 1
$\mathcal{G}_i$	Guided kernel means(GKM) applied to the i'th voxel in the image
$\mathcal{S}_i$	Guided kernel sum (GKS) applied on the i'th voxel in the image

to avoid memory problems due to the large size of the kernel matrix for each 3D image, an implicit matrix-vector multiplication is performed using an operation similar to filtering.

While this method has a rigid mathematical framework, its iterative nature can lead to an excessive computational cost. Therefore, in order to improve the speed of the image denoising, a second very fast non-iterative method for denoising PET images is presented. This heuristic method is composed of two simple steps. In the first step, the subject's registered MR image is used as the guide image for denoising the PET image. In the second step, the median-filtered PET image is used as the guide image to denoise the residual. The denoised residual is then added to the denoised image from the first step to recover signals specific to the PET image. We also show that these methods, when applied after a conventional deblurring algorithm, can restore the PET image by reducing the noise as well as PVE.

4.2 Methodology

In this section a detailed description of the proposed methods is given. Hereafter in the paper, bold capital letters are used to indicate matrices, bold lower-case letters represent column vectors and non-bold letters denote scalar values. Table 4.1 summarises the symbols and abbreviations used throughout this paper.

Let $\mathbf{y} \in \mathbb{R}^I$ be a vector of size I representing the voxles of the noisy reconstructed image. The goal is to devise a restoration algorithm to obtain an estimate of the true image $\mathbf{x}$ from $\mathbf{y}$. Let us call this estimate $\hat{\mathbf{x}}$.

4.2.1 Guided kernel means (GKM)

In 3D filtering, the estimate of a voxel value $\hat{x}_i$ in the restored image is obtained by a weighted average of other voxels in the image.

The weight associated with voxel y_j in this average is proportional to the similarity of the voxel y_i to y_j as indicated by $k_{i,j}$.

$$\hat{x}_i = \frac{1}{\sum_{j\in\Lambda_i} k_{i,j}} \sum_{j\in\Lambda_i} y_j k_{i,j} : \tag{4.1}$$

where Λ_i is neighbourhood window of size $L \times L \times L$ voxels centred at voxel i. The weight $k_{i,j}$ can be a function of the spatial locations of voxels i and j, the intensities of the two voxels and/or the neighbourhood around them. In fact, many image filtering methods in the literature (e.g. Gaussian, bilateral and non-local means filtering methods) only differ in the choice of this kernel function (Milanfar 2013).

Applying the above filter is equivalent to minimizing the following weighted least squares objective function (Milanfar 2013):

$$\hat{x}_i = \arg\min_{x_i} \quad \sum_{j\in\Lambda_i} (x_i - y_j)^2 k_{i,j} \tag{4.2}$$

The weights in the filtering operation, indicated as $k_{i,j}$ can be viewed as the parameters

of the denoising method. In an ideal case, one would like to obtain the weights from the noise free ground truth image. Therefore, in practice, many denoising algorithms compute the weights from a pre-filtered image, or a so-called *guide image* to improve the estimate (Milanfar 2013).

This is in particular very important for PET images with high levels of noise. For these images, a registered MR image from the same subject (resampled to the PET image space) has very low noise compared to the PET image and therefore can be used as the guide image to obtain these parameters.

Assuming that the guide image has a relatively low amount of noise compared to the PET image, the following simple yet effective kernel function is used in this work:

$$k_{i,j} = e^{-\frac{(v_i - v_j)^2}{2h^2}} \tag{4.3}$$

where v_i is the intensity of the i'th voxel in the *guide* image that corresponds to the spatial location of y_i in the PET image and h is the smoothing parameter of the filter. Note that, as opposed to patch-based filtering methods (such as non-local means) where the squared difference between patches need to be calculated, Eq.(4.3) only depends on the squared difference between *single* voxel intensities and hence is fast to compute. We refer to this method as guided kernel filtering (GKM), where $\mathcal{G}_i$ is a voxel-wise operation on voxel y_i defined using equation 4.1 and 4.3:

$$\hat{x_i} = \mathcal{G}_i(y_i) = \frac{1}{\sum_{j \in \Lambda_i} e^{-\frac{(v_i - v_j)^2}{2h^2}}} \sum_{j \in \Lambda_i} y_j e^{-\frac{(v_i - v_j)^2}{2h^2}} \tag{4.4}$$

4.2.2 Filtering in matrix notation

Let $\mathbf{K}$ be a symmetric positive semi-definite matrix of size $I \times I$. If voxel j is in the neighbourhood of voxel i identified by λ_i, then the value of $k_{i,j}$ corresponds to the weight in the GKM algorithm; otherwise $k_{i,j}$ is set to zero. Note that, since in practice

the size of the neighbourhood is much smaller than the size of the image, the resulting kernel matrix $\mathbf{K}$ is usually very sparse. Applying GKM on the image $\mathbf{y}$ is identical to multiplying the row-normalized kernel matrix by the noisy image (Milanfar 2013):

$$\hat{\mathbf{x}} = \begin{bmatrix} \mathcal{G}_1(y_1) \\ . \\ . \\ . \\ \mathcal{G}_I(y_I) \end{bmatrix} = \frac{\mathbf{Ky}}{\mathbf{K}\vec{\mathbf{1}}} = \mathbf{Wy}, \tag{4.5}$$

where $\vec{\mathbf{1}}$ is an all-one vector of size I.

We refer to this row-normalized kernel matrix as $\mathbf{W}$. Even though matrix $\mathbf{W}$ is very sparse, storing the matrix for a 3D image may not be practical. In fact for a 3D image of size $256 \times 256 \times 207$ voxels and for a neighbourhood size of $7 \times 7 \times 7$ voxels, the matrix can have more than 4.6×10^9 non-zero elements (i.e. $I \times L^3$). Therefore, instead of explicit construction of this matrix the image $\hat{\mathbf{x}}$ is estimated by applying Eq.(4.4) on each voxel of noisy image $\mathbf{y}$.

4.2.3 Re-parametrization

Re-parameterization of PET images within reconstruction have been used as a means of regularization (Tahaei & Reader 2016, Wang & Qi 2015, Lewitt 1992, 1990, Jacobs et al. 1998, Jiao et al. 2015). In re-parametrization, the denoised image $\hat{\mathbf{x}}$ is represented as a superposition of some basis vectors $\mathbf{A}_1$ to $\mathbf{A_l}$:

$$\hat{\mathbf{x}} = \sum_{l=1}^{L} \theta_l \mathbf{A}_l = \mathbf{A}\boldsymbol{\theta} \tag{4.6}$$

where $\mathbf{A}$ is a matrix of size $I \times L$ in which the l'th column A_l is the $l'th$ basis vector. $\boldsymbol{\theta}$ is the coefficient vector in which each element θ_l is the coefficient for the l'th basis vector.

The choice of basis function can range from blobs (Lewitt 1992), to sophisticated structural-based basis functions, e.g. patch-based dictionaries learned from MR data

(Tahaei & Reader 2016). Row-normalized kernel matrices have also been used for re-parametrization of PET images within reconstruction (Wang & Qi 2015, Novosad & Reader 2016). When using a row-normalized kernel matrix extracted from a guide image for re-parameterization, each column i (i.e. basis function i) indicates the similarity of voxel i to other voxels in the guide image.

Once the basis matrix is calculated (or learned) from the guide image, the coefficient vector $\boldsymbol{\theta}$ is estimated by optimizing a pre-defined objective function. For reconstruction this objective function is usually the log likelihood of the coefficient vector given the measurement data. For post-reconstruction denoising, the coefficient vector can be estimated by minimizing the representation error (i.e., the squared difference between the noisy image and the re-parameterized image).

4.2.4 Proposed Method 1: PET-MR Guided kernel re-parameterization

In this section, a novel method for denoising PET images using the subject's MR image, registered and resampled to the PET image space, is proposed. In this method, the signal in the denoised PET image is re-parameterized using *two sets* of basis functions, one obtained from the registered MR image and the other obtained from the PET image:

$$\hat{\mathbf{x}} = \mathbf{M}\boldsymbol{\alpha} + \mathbf{P}\boldsymbol{\beta} \tag{4.7}$$

where $\mathbf{M}$ is a basis matrix calculated from the MR image and $\mathbf{P}$ is the basis matrix calculated from the median-filtered PET image. $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are the coefficient vectors for these basis matrices respectively. A detailed description of how these basis matrices are obtained and how the coefficient vectors are estimated is provided in the following sections.

4.2.4.1 Basis formation

Before basis matrix formation, both the median-filtered PET image and the registered MR image are normalized by their maximum value so that each image contains values between 0 and 1. Once the images are normalized, the GKM formulation in Eq. (4.3) is used to construct two kernel matrices, one from the registered MR image and the other from the median-filtered PET image. The resulting matrices are then row-normalized to obtain $\mathbf{M}$ and $\mathbf{P}$ respectively. As mentioned before, $\boldsymbol{\alpha} \in \mathbb{R}^I$ and $\boldsymbol{\beta} \in \mathbb{R}^I$ are the coefficient vectors for these basis matrices. In the following section, an algorithm to estimate the coefficient vectors is presented. This re-parameterization framework is illustrated in Fig. 4.1.

4.2.4.2 Problem formulation

The coefficients for basis matrices $\mathbf{M}$ and $\mathbf{P}$ can be estimated using a least-squares objective function. Note that the median-filtered PET image does not contain detailed boundary information due to low resolution (and high noise) in the PET image. Therefore, one would like the restored image to contain minimal signal represented by the PET basis matrix, i.e., only in regions where the PET image does not agree with the MR image (e.g., a lesion in PET not present in the MR). One way to control the contribution of the PET basis matrix in the final restored image is to impose a sparsity penalty on the coefficients of this matrix. This is done by adding an l_1 norm penalty to the least squares objective:

$$\underset{\boldsymbol{\alpha},\boldsymbol{\beta}}{\arg\min} \quad \|\mathbf{y} - (\mathbf{M}\boldsymbol{\alpha} + \mathbf{P}\boldsymbol{\beta})\|_2^2 + \lambda \|\boldsymbol{\beta}\|_1 \tag{4.8}$$

where λ is a hyper parameter that controls the sparsity of $\boldsymbol{\beta}$.

4.2.4.3 Algorithm

The above optimization problem can be solved in an alternating manner by fixing $\boldsymbol{\beta}$ and solving for $\boldsymbol{\alpha}$ and then fixing for $\boldsymbol{\alpha}$ and solving for $\boldsymbol{\beta}$ iteratively.

4.2.4.4 Subproblem 1:solving for $\boldsymbol{\alpha}$

Fixing β leads to the following problem:

$$\arg\min_{\boldsymbol{\alpha}} \quad \|\mathbf{y} - (\mathbf{M}\boldsymbol{\alpha} + \mathbf{P}\boldsymbol{\beta})\|_2^2 \tag{4.9}$$

Considering that $\mathbf{M}\vec{\mathbf{1}} = \vec{\mathbf{1}}$, this least squares problem that can be solved using, for example, the simultaneous algebraic reconstruction technique (SART) (Andersen & Kak 1984). The SART update rule for the above objective function is as follows:

$$\boldsymbol{\alpha}^{k+1} = \boldsymbol{\alpha}^k + \frac{\mathbf{M}^T \frac{((\mathbf{y}-\mathbf{P}\boldsymbol{\beta})-\mathbf{M}\boldsymbol{\alpha}^k)}{\mathbf{M}\vec{\mathbf{1}}}}{\mathbf{M}^T\vec{\mathbf{1}}} \tag{4.10}$$

4.2.4.5 Subproblem 2: solving for $\boldsymbol{\beta}$

Fixing $\boldsymbol{\alpha}$ and solving for $\boldsymbol{\beta}$ leads to a large-scale lasso problem. We use the fast iterative shrinkage-thresholding algorithm (FISTA) (Beck & Teboulle 2009) to solve this problem. FISTA is shown in algorithm 3. The parameter c should be set to a value greater than the largest eigenvalue of $\mathbf{P}\mathbf{P}^T$ which is obtained using the power iteration method (Epperson 2013). S_λ is the soft-thresholding operator defined as (Daubechies et al. 2003):

$$S_\lambda(x) = \text{sign}(x)\ \max(|x| - \lambda, 0) \tag{4.11}$$

Algorithm 3: FISTA

Initialize: $t_0 = 1$; $\mathbf{z}_0 = \boldsymbol{\beta}_0$;

while *Not converged* **do**

$k = k + 1$

$\boldsymbol{\beta}^k = S_{\lambda/c}(\mathbf{z}^k + (1/c)\mathbf{P}^t((\mathbf{y} - \mathbf{M}\alpha) - \mathbf{P}\mathbf{z}^k))$

$t^{k+1} = \frac{1+\sqrt{1+4(t^k)^2}}{2}$

$\mathbf{z}^{k+1} = \boldsymbol{\beta}^k + \frac{t^k-1}{t^{k+1}}(\boldsymbol{\beta}^k - \boldsymbol{\beta}^{k-1})$

end

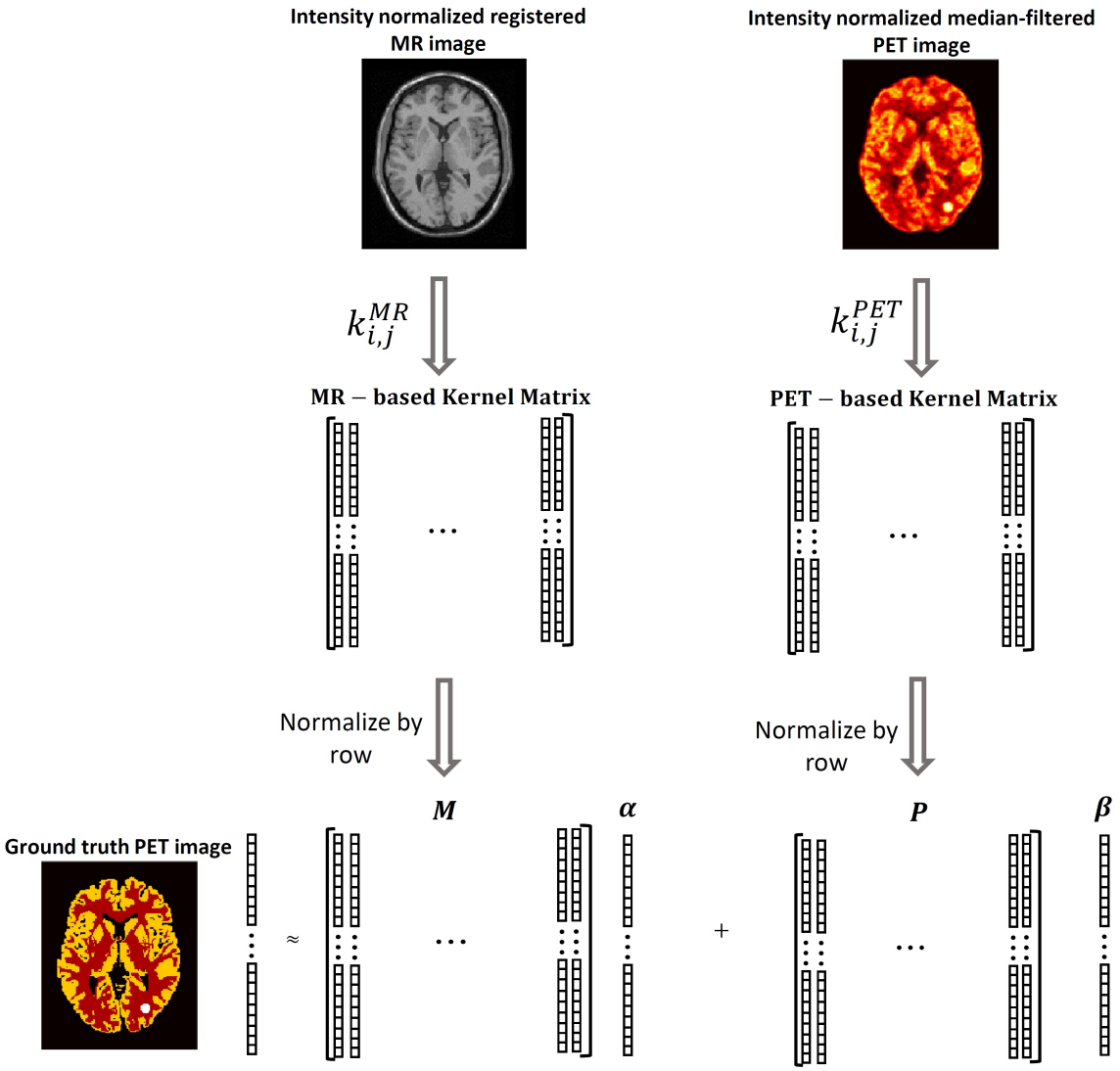


Figure 4.1: Illustration of the guided kernel re-parameterization in proposed Method 1. Two row-normalized intensity based kernel matrices are obtained, one from the normalized subject's registered MR image and the other from the normalized median-filtered PET image separately. These two matrices **M** and **P** are then used as basis matrices to re-parametrize the denoised PET image. In order to estimate the denoised image, the coefficient vectors for these basis matrices, namely α and β, have to be estimated by minimizing a sparsity-penalized least squares objective function.

4.2.4.6 Implicit matrix-vector multiplication

As previously mentioned, computing the row-normalized kernel matrix $\mathbf{W}$ for a 3D image is highly demanding on memory. However, both SART and FISTA involve a multiplication of either this matrix or its transpose by a vector. In the following we show how multiplying $\mathbf{W}$ and $\mathbf{W}^T$ by an arbitrary vector $\boldsymbol{\theta}$ can be performed implicitly.

Let us define the operation $\mathcal{S}_i$ as the guided kernel sum (GKS) of the i'th voxel of image $\boldsymbol{\theta}$ using $\mathbf{K}$ as:

$$\mathcal{S}_i(\theta_i) = \sum_{j \in \Lambda_i} \theta_j k_{i,j} \tag{4.12}$$

This is similar to GKM but without normalization. Applying this operator on every voxel i in the image is equivalent to multiplying the kernel matrix by the image:

$$\begin{bmatrix} \mathcal{S}(\theta_1) \\ . \\ . \\ . \\ \mathcal{S}(\theta_J) \end{bmatrix} = \mathbf{K}\boldsymbol{\theta} \tag{4.13}$$

Using the GKS operator, basis matrix-vector multiplications can be performed implicitly as follows:

$$\mathbf{W}\boldsymbol{\theta} = \text{diag}^{-1}[\mathbf{K}\vec{\mathbf{1}}]\mathbf{K}\boldsymbol{\theta} = \frac{\mathbf{K}\boldsymbol{\theta}}{\mathbf{K}\vec{\mathbf{1}}} = \frac{\begin{bmatrix} \mathcal{S}_1(\theta_1) & . & . & . & \mathcal{S}_I(\theta_I) \end{bmatrix}^T}{\begin{bmatrix} \mathcal{S}_1(1) & . & . & . & \mathcal{S}_I(1) \end{bmatrix}^T} \tag{4.14}$$

This is equivalent to GKM. Now, considering that the kernel matrix is a symmetric matrix, multiplication of the transpose of the basis matrix $\mathbf{W}$ by vector $\boldsymbol{\theta}$ can be written

as:

$$\begin{aligned} \mathbf{W}^T\boldsymbol{\theta} = \quad & \mathbf{K}^T(\mathrm{diag}^{-1}[\mathbf{K}\vec{\mathbf{1}}])^T\boldsymbol{\theta} = \\ \mathbf{K}\frac{\boldsymbol{\theta}}{\mathbf{K}\vec{\mathbf{1}}} = \quad & \left[\mathcal{S}_1(\frac{\theta_1}{\mathcal{S}_1(1)}) \ . \ . \ . \ \mathcal{S}_I(\frac{\theta_I}{\mathcal{S}_I(1)})\right]^T \end{aligned} \tag{4.15}$$

Here, the image is first normalized by $\left[\mathcal{S}_1(1) \ . \ . \ . \ \mathcal{S}_I(1)\right]^T$ and then GKS is applied to the normalized image.

4.2.5 Proposed Method 2: PET-MR GKM with twicing

In this section, a very fast and simple alternative approach for denoising using weights obtained from the registered MR and median-filtered PET images is presented.

4.2.5.1 Twicing

The idea of twicing was first proposed by Tukey in the 1970s to improve the estimate in denoising algorithms (Tukey 1977). In this method, the residual image (obtained as the difference between the original image and the estimated denoised image) is denoised and then added to the estimated image:

$$\begin{aligned} \mathbf{r} &= \mathbf{y} - \mathrm{filter}(\mathbf{y}) \\ \hat{\mathbf{x}} &= \mathrm{filter}(\mathbf{y}) + \mathrm{filter}(\mathbf{r}) \end{aligned} \tag{4.16}$$

If the residual contains some signal, denoising the residual helps to remove the noise and to recover that signal, and hence adding this filtered residual to the denoised image may improve the estimate. In principle, this process can be carried out for more than one iteration. Increasing the number of iterations will increase the noise in the resulting image.

4.2.5.2 Algorithm for Method 2

The rationale behind Method 2 is very similar to Method 1: in order to improve the boundary information in the restored image, we would like the parameters of GKM, i.e., the elements of $\mathbf{K}$, to be estimated from a noise-free image with detailed anatomical information. Hence using the MR image as the guide image in GKM seems very appealing.

However, since the kernel is obtained from the MR image, it may also remove some of the underlying signal in regions where PET and MR do not agree, e.g. a lesion in the PET that is not present in MR. This is where twicing plays an important role. The residual image contains some of the underlying signal specific to PET as well as noise. Denoising the residual using weights obtained from a median-filtered PET image helps to remove the noise from the residual, hence adding this denoised residual can recover part of the lost PET signal.

$$\begin{aligned} &\text{Step 1}: && \hat{x}_i^{\text{MR}} = \mathcal{G}_i^{\text{MR}}(y_i) \\ &\text{Step 2}: && \hat{x}_i^{\text{Final}} = \hat{x}_i^{\text{MR}} + \mathcal{G}_i^{\text{PET}}(y_i - \hat{x}_i^{\text{MR}}) \end{aligned} \tag{4.17}$$

$\mathcal{G}_i^{MR}$ and $\mathcal{G}_i^{PET}$ are defined by the GKM operator given in Eq. (4.4), in which the $k_{i,j}$ elements are obtained from the registered MR and the median-filtered PET image respectively. This method is illustrated in Fig.4.2.

Note that, similar to the proposed Method 1, before applying the GKM, both the median-filtered PET image and the registered MR image are normalized by their maximum intensity. Normalization brings both the median-filtered PET image and the MR image to be in a range between 0 and 1. This way the smoothing parameter of the kernel h can be chosen irrespective of the intensity range in the guide image. This enables us to use the same value of h for both the median filtered PET and the MR image.

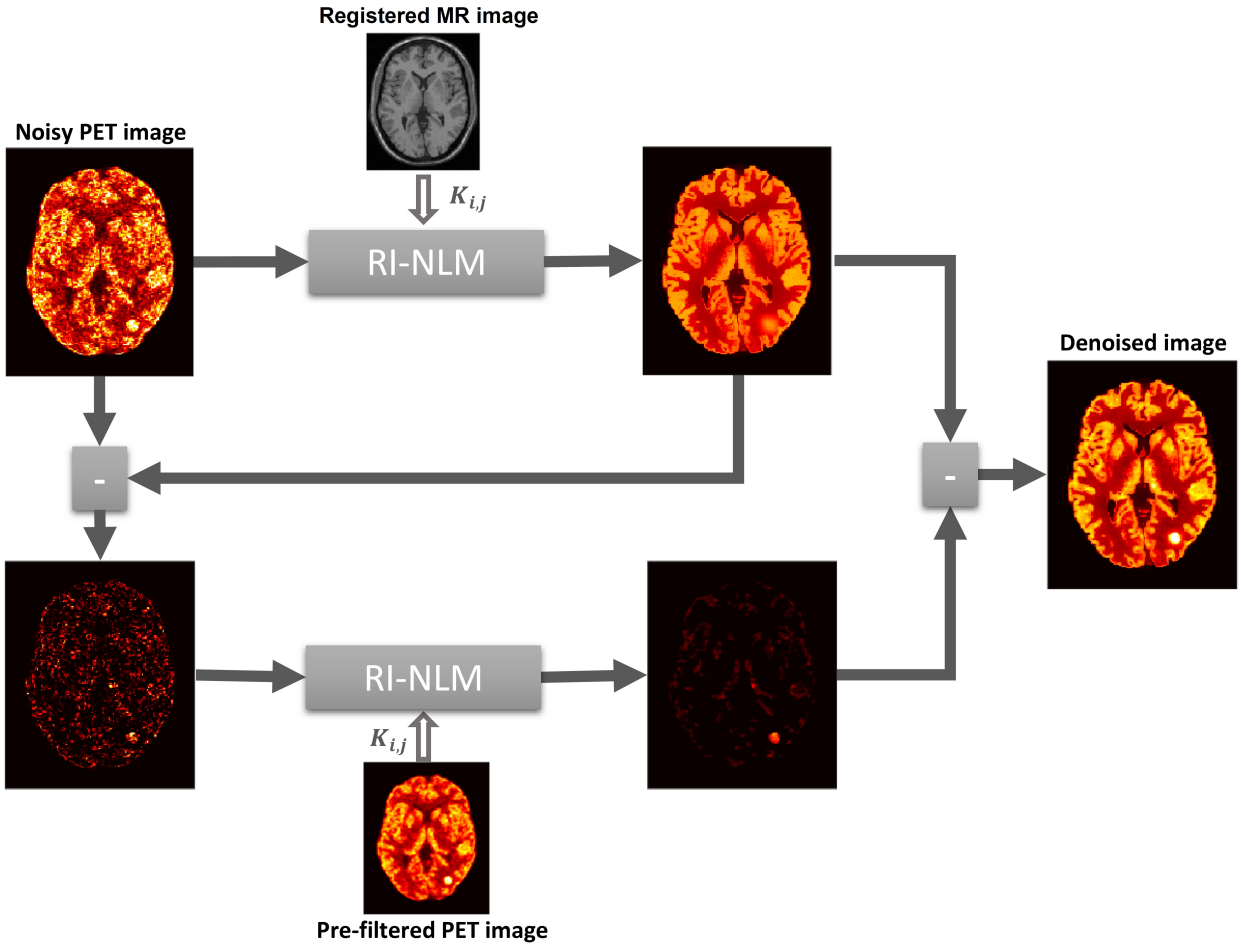


Figure 4.2: Illustration of proposed Method 2 (GKM with twicing): First the registered MR image is used as the guide image to denoise the PET image. The residual is then computed and denoised using a median-filtered PET image as the guide image. The denoised residual is then added to the denoised PET image.

4.2.6 Combination with PVE correction

In addition to noise, PVE in PET images is another very important source of error in PET quantification. When the point spread function (PSF) associated with PVE is known, deblurring algorithms can be used to improve the resolution of the PET image.

One issue with many iterative deblurring algorithms is that the ill-conditioned nature of deblurring leads to noise amplification in the resulting image(Mignotte & Meunier 2000, Teo et al. 2007). However, both of the proposed methods are able to remove a high amount of noise in the image. Therefore, applying these methods after deblurring can effectively reduce the normally encountered noise amplifications while still improving the boundary information.

4.2.6.1 Deblurring

For PVE correction, the Richardson-Lucy deblurring algorithm (Lucy 1974, Richardson 1972) with the exact known stationary point spread function was applied to the reconstructed image to obtain the deblurred image $\mathbf{u}$. The update rule of Richardson-Lucy algorithm is shown below:

$$\mathbf{u}^{n+1} = \mathbf{u}^{n}\mathbf{C}^{\mathbf{T}}\frac{\mathbf{y}}{\mathbf{C}\mathbf{u}^{n}} \tag{4.18}$$

where $\mathbf{C}$ is the blurring matrix in which each column $\mathbf{c}_i$ indicates the PSF centred at voxel i.

4.3 Experiments and results

In this section the performance of the proposed methods is evaluated using both simulation and real data. The proposed methods are compared with Block-matching and 4D filtering (BM4D) method (Maggioni et al. 2013) and median filtering. BM4D is an extension of BM3D (Dabov et al. 2007) denoising method to volumetric data. In BM3D, the image is first decomposed to patches and these patches are clustered based on similarity. This is followed by a 3-D collaborative Wiener filtering applied to

each cluster. The resulting filtered patches are then aggregated to form the denoised image. In median filtering each voxel in the denoised image is simply the median of the neighbourhood around it.

4.3.1 Simulation

Simulated [^{18}F]fluorodeoxyglucose(FDG) PET images were used to evaluate the performance of the proposed methods. An MR image was obtained from the BrainWeb project (Cocosco et al. 1997) (`http:brainweb.bic.mni.mcgill.ca/brainweb`). In order to simulate HRRT PET images, the corresponding BrainWeb segmentation was re-sampled to the PET image space ($256 \times 256 \times 207$ isotropic voxels of $1.22mm \times 1.22mm \times 1.22mm$). Then radioactivity values from PET SORTEO (Frouin et al. 2002) (`http://sorteo.cermep.fr/home.php`) for grey matter (GM) and white matter (WM) were used to generate a realistic [^{18}F]FDG radioactivity distribution at roughly MRI resolution.

Moreover, in order to evaluate the performance of the method in capturing information specific to the PET image, three lesions were added to the ground truth. Lesion 1 was added in the WM with 50% activity increase relative to WM. Lesion 2 is embedded in the GM with 50% activity increase relative to the GM. Lesion 3 was embedded in GM with 70% activity reduction relative to GM. Lesion 1 is a sphere with a radius of 5 voxels located in the WM. Lesion 2 has a radius of 4 voxels and is located across the GM and WM boundary. Lesion 3 is located in the GM and follows the cortical boundaries. The volumes of lesions 1, 2 and 3 are 515, 389 and 297 voxels respectively.

Figure 4.3 shows the resulting ground truth PET along with the registered MR image. The resulting ground truth was forward projected using an HRRT scanner model to obtain a sinogram data with 256 radial bins, 288 azimuthal angles and 104 projection planes. Appropriate linear attenuation coefficient factors similar to (Zaidi et al. 2007) were also included in the forward projection. In order to simulate scatter and randoms, a smoothed sinogram with total activity equal to one quarter of the projected

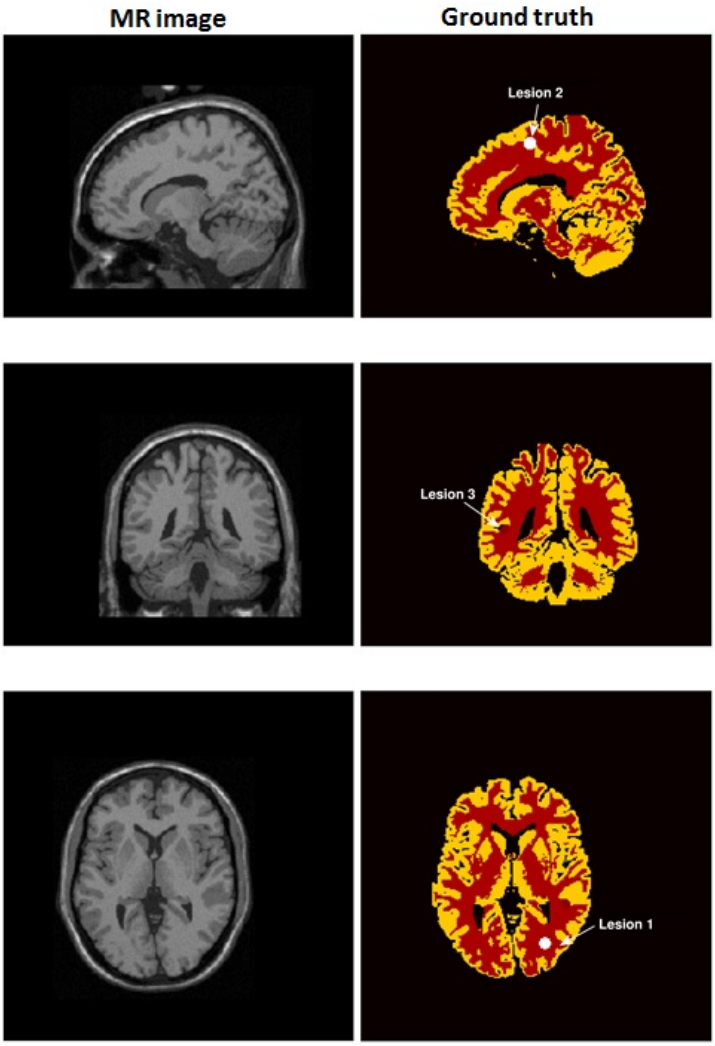


Figure 4.3: The simulated ground truth PET image and the corresponding registered T1-weighted MR image. The dimension of the image corresponds to an HRRT scanner field of view.

ground truth was added (Tahaei & Reader 2016).

The proposed methods were evaluated at different noise levels, the resulting sinogram was scaled so that the expected number of counts will be equal to 50M (noise level 3),100M (noise level 2) and 150M (noise level 1). Note that these values are smaller than the number of counts normally acquired for a brain scan on an HRRT scanner. Poisson noise was then introduced to the resulting sinograms to generate 5 noisy realizations from each noise level. The exact known values for attenuation

and scatter and random events were used for corrections within the reconstruction of the noisy measurement data. The OSEM (Ordered subset expectation maximization)(Hudson & Larkin 1994) algorithm with 16 subsets and iterations varying from 1 to 10 were used to reconstruct the PET images.

In each iteration of the alternating optimization in proposed Method 1, both SART and FISTA were terminated when the relative change of the objective function was less than 0.02. Simulation analysis showed that running the alternating minimization loop for 10 iterations works well for different noise levels; more iterations lead to increased noise in the resulting images.

Both proposed methods were tested with neighbourhood sizes ranging from $5 \times 5 \times 5$ to $13 \times 13 \times 13$ voxels and the smoothing parameter h ranging 0.02 to 0.06. Method 1 was also tuned for sparsity parameter λ ranging from 0 to 0.02.

4.3.1.1 Combination with PVE correction

The process described in the simulation section was used to create an FDG ground truth. This time however, the ground truth was smoothed using a Gaussian kernel of size $7 \times 7 \times 7$ voxels and FWHM equal to 4mm. The resulting ground truth was then forward projected and Poisson noise was introduced to the resulting sinogram (trues + scatter and randoms) to generate 5 realizations. Scatter and random evens were also modelled with the same approximation mentioned as in the previous section. The expected number of counts in each of the resulting sinograms was 300×10^6. For PVE correction, 10 iterations of the Richardson-Lucy algorithm with the known PSF function (FWHM=4mm) was used. The proposed methods were then applied to the PVE corrected images. The methods were tested for the same range of parameters mentioned in the previous section.

4.3.2 Figures of merit

In order to assess the performance of the methods in reducing the noise, the ensemble-based normalized root mean squared error (n-RMSE) is obtained for each voxel within

a ROI and across multiple realizations. The resulting voxel-based values are then averaged within the ROI.

$$\text{Mean voxel-based n-RMSE within ROI} = \frac{1}{|ROI|}\sum_{i\in \text{ROI}} \frac{\text{RMSE}_i}{x_i} \tag{4.19}$$

where $\text{RMSE}_i = \sqrt{\frac{\sum_{r=1}^{R}(\hat{x}_i^r - x_i)^2}{R}}$ is the root mean squared error of voxel i across R realizations.

Also, bias of the mean of the lesion, as well as the standard deviation of the mean of the lesions normalized by their true mean, computed across realizations are used to assess the performance of the proposed method. The known boundaries from the simulation are used to determine the lesions and the background regions.

Also, for quantitative evaluation of the proposed methods in regions where the functional activity does not agree with the anatomy, the mean contrast recovery coefficient (CRC) of the lesions specific to PET across multiple realizations was used. The CRC value of the lesion in one realization can be obtained as follows:

$$\text{CRC} = \frac{|\text{Mean anctivity in the lesion} - \text{Mean activity in the background}|}{|\text{known mean activity in the lesion} - \text{known mean activity in the background}|} \tag{4.20}$$

where the known mean activity is obtained from the ground truth.

For Method 1, there is trade-off between nRMSE and the contrast of the PET specific lesions. Therefore, to choose the optimum set of parameters, the following ratio was used:

$$\text{CRC to error ratio} = \frac{\overline{\text{CRC}_{\text{L1}}} \times \overline{\text{CRC}_{\text{L2}}} \times \overline{\text{CRC}_{\text{L3}}}}{\text{Mean n-RMSE within the brain}} \tag{4.21}$$

where $\overline{\text{CRC}_{\text{Li}}}$ indicates the mean CRC value across realizations for lesion Li.

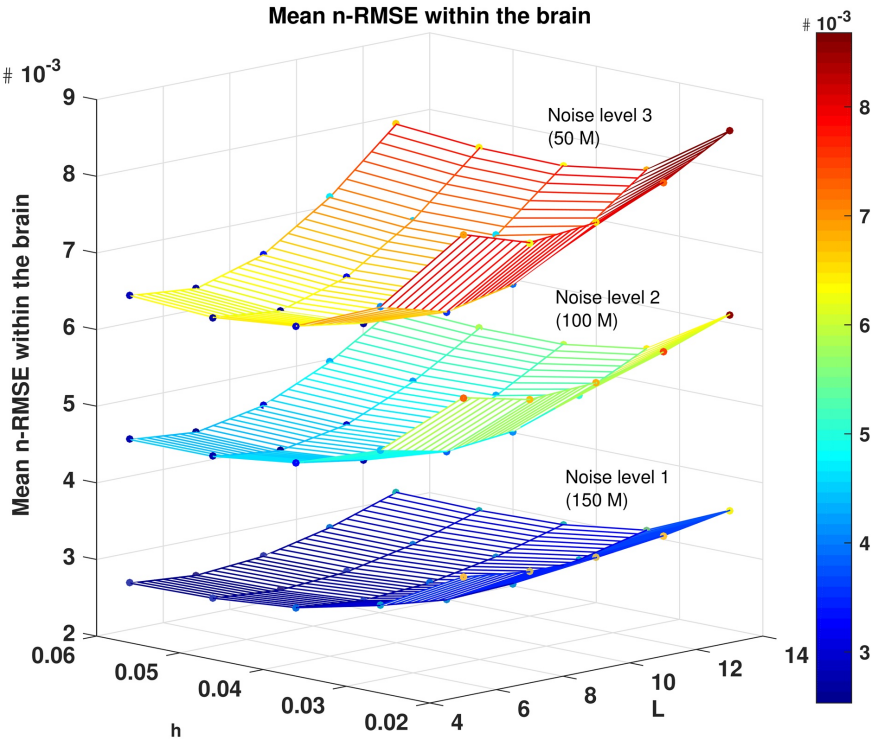


Figure 4.4: The mean n-RMSE within the brain as a function of smoothing parameter h and neighbourhood length L for the Method 2 shown for 3 different noise levels at their best iteration number (i.e., the OSEM iteration leading to minimum mean n-RMSE).

The SNR of the restored image is another figure of merit used in our analysis:

$$\mathrm{SNR}_{\hat{\mathbf{x}}^r} = \frac{\|\hat{\mathbf{x}}^r\|^2}{\|\hat{\mathbf{x}}^r - \mathbf{x}\|^2} \tag{4.22}$$

4.3.3 Results

Figure 4.4 shows the effect of smoothing parameter h and the neighbourhood length L on the mean voxel based n-RMSE within the brain for different noise levels. One can see that the 3 surfaces associated with 3 different noise levels follow a similar trend and hence the optimal set of parameters for obtaining minimum mean n-RMSE does not seem to be sensitive to the number of counts in the measurement data. Also, this

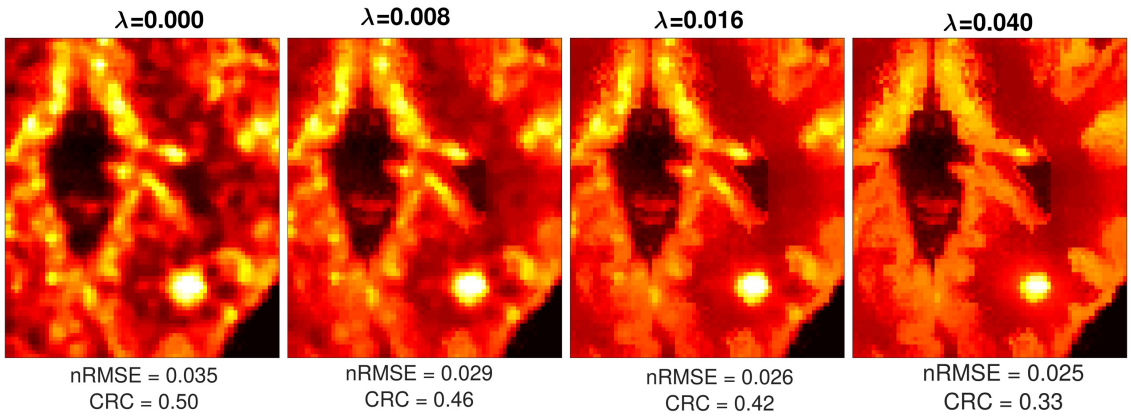


Figure 4.5: Transverse views of the simulation data showing the effect of increasing sparsity parameter λ of proposed Method 1 on image quality along with mean n-RMSE and mean CRC value of the lesion in WM for multiple realizations. By increasing the sparsity parameter the resulting images become less noisy; this leads to improvements in the nRMSE values within the brain at the expense of decrease in the CRC of the lesion in WM.

figure shows that by decreasing the noise level the parameter surface becomes flatter. This indicates that for lower noise levels, the mean n-RMSE value is less sensitive to the choice of parameters.

Figure 4.5 shows axial views of the denoised image with increasing value of λ. Increasing sparsity of the PET-based coefficients by increasing the value of λ can lead to less noise in the resulting image. As indicated at the bottom of each image, this increase will lead to lower nRMSE value within the brain at the expense of a decrease in the mean CRC value.

Table 4.2 shows the parameters leading to maximum CRC to nRMSE ratio for both methods. For Method 1, higher noise levels require a slight increase in the sparsity parameter λ to achieve the optimal value.

Figures 4.6 and 4.7 show the performance of the proposed methods for noise level 2 (100 M) and noise level 3 (50 M) respectively. Noise level 1 (150 M) is not shown here; applying the methods on higher count levels leads to images with higher quality.

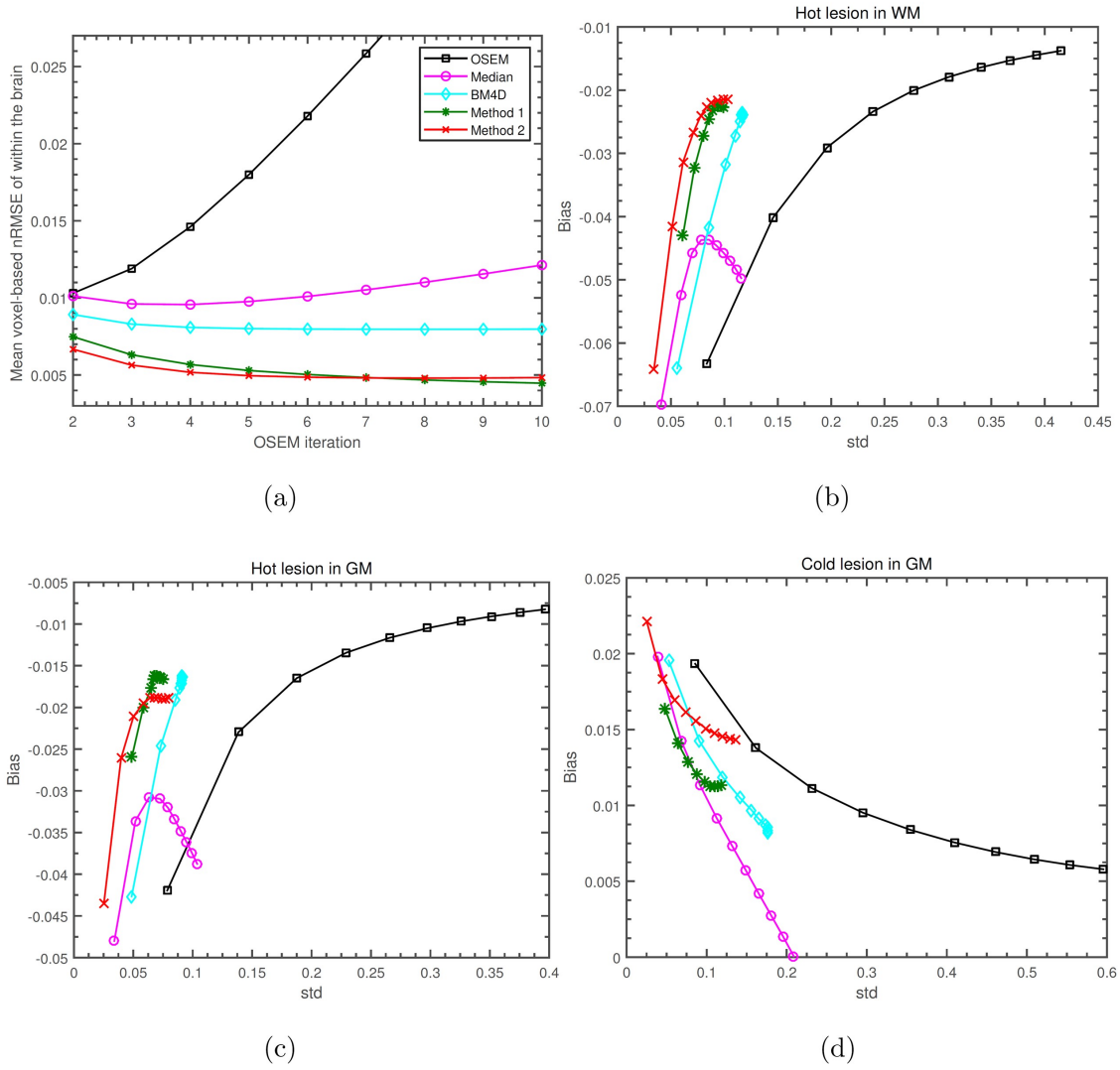


Figure 4.6: Different figures of merit for OSEM with no post-processing, median post-filtering, BM4D, Method 1 and Method 2 for noise level 2 (100 M counts). a) Mean n-RMSE within the brain as a function of iteration. Bias versus normalized standard deviation of the mean activity in the (b) hot lesion in the WM, (c) hot lesion in the GM and (d) cold lesion in the GM. Each data point corresponds to the figure of merit obtained from images that were reconstructed using a specific number of iterations of OSEM algorithm.

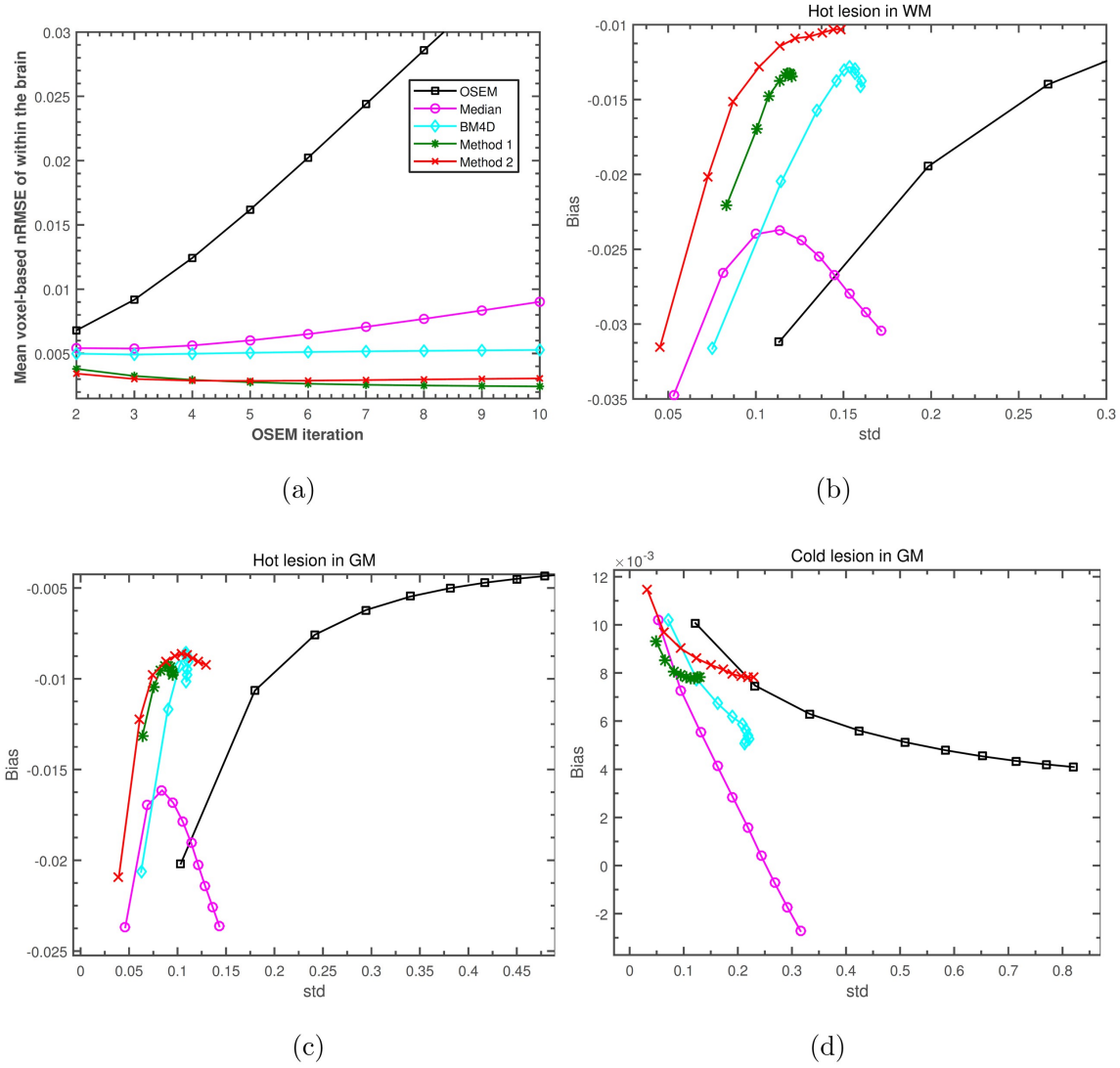


Figure 4.7: Different figures of merit for OSEM with no post-processing, median post-filtering, BM4D, Method 1 and Method 2 for noise level 3 (50 M counts). a) Mean n-RMSE within the brain as a function of iteration. Bias versus normalized standard deviation of the mean activity in the (b) hot lesion in the WM, (c) hot lesion in the GM and (d) cold lesion in the GM.

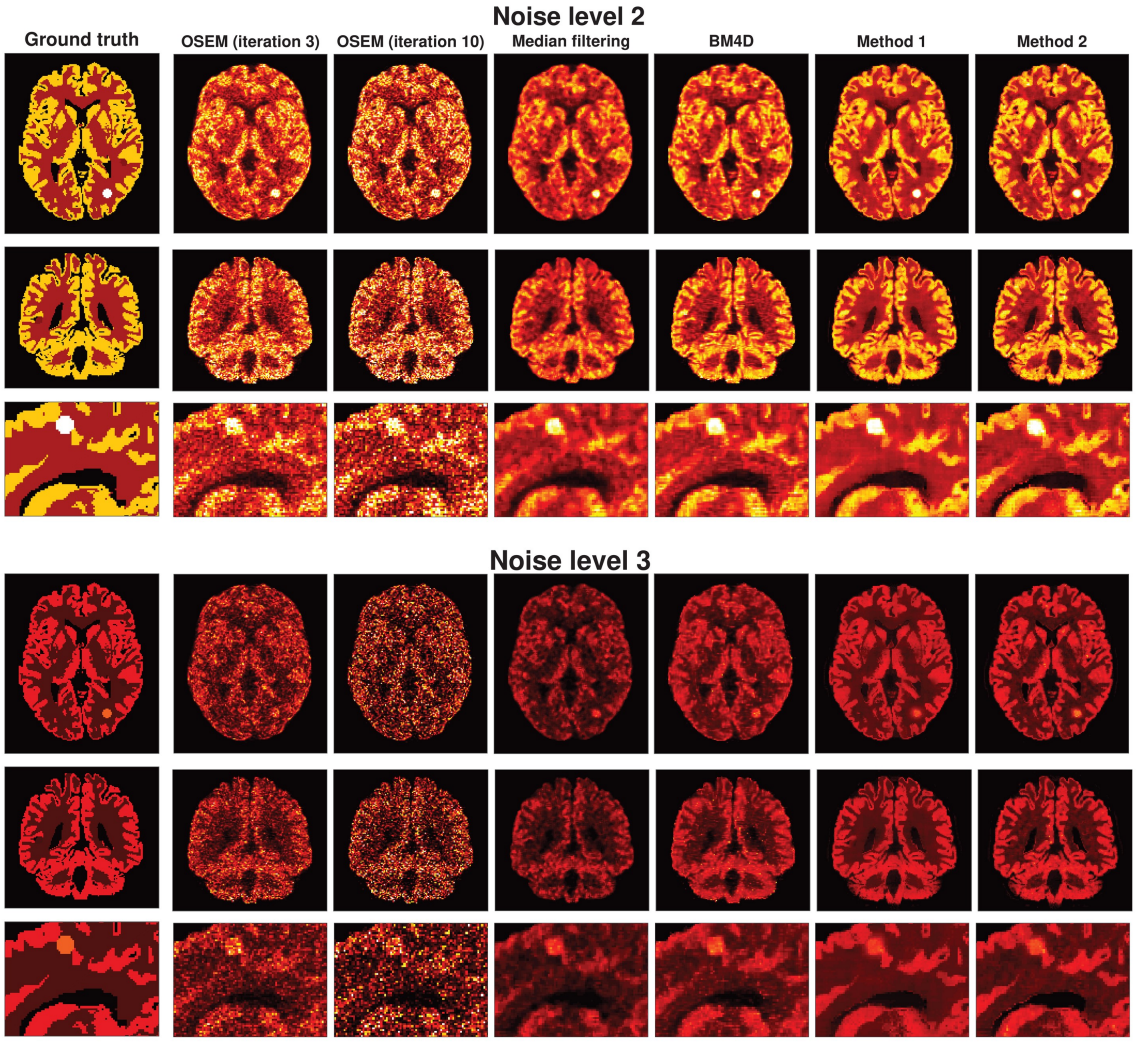


Figure 4.8: An example of our FDG simulation reconstructed using OSEM and then denoised using different methods. The images are shown for transverse, coronal and sagittal views. The ground truth image is show in Fig. 4.3. The axial and coronal views are cropped to focus on the brain. The sagittal view is zoomed for better visualization. All methods are shown with parameters indicated in Table 4.2.

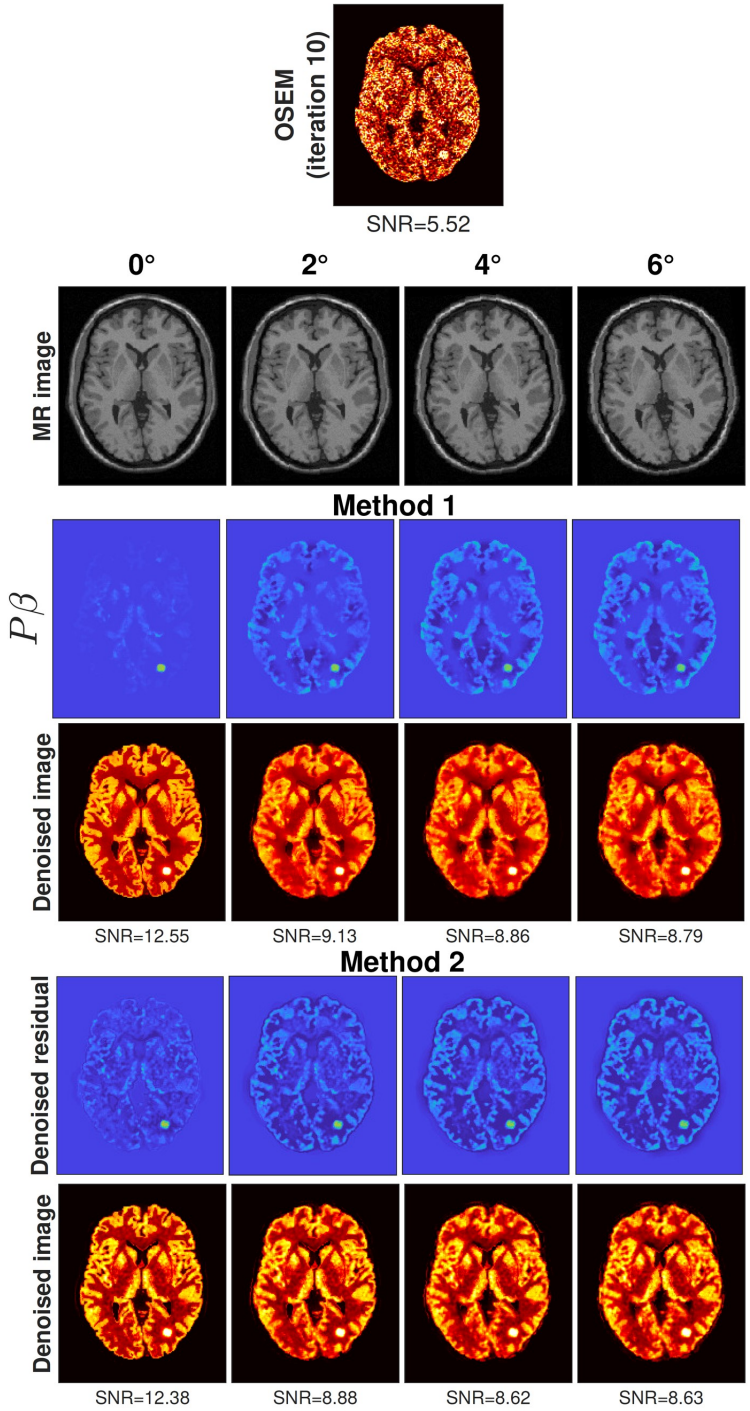


Figure 4.9: Example of applying Method 1 and 2 using misaligned MR images. The top row shows the noisy image (with Level 1 noise) on which the methods are applied. The MR image is rotated around z axis by 0, 2, 4 and 6 degrees. For each method the top row shows PET-based component of the final image. The SNR value is shown at the bottom of each image.

Table 4.2: Selected parameters for proposed methods

Method	Noise level (Counts)	OSEM iteration	**L**	**h**	**λ**	**n-RMSE**
Method 1	1 (150 M)	10	11	0.03	0.006	0.0022
	2 (100 M)	10	11	0.03	0.008	0.0040
	3 (50 M)	10	11	0.03	0.008	0.0059
Method 2	1 (150 M)	10	11	0.03	-	0.0026
	2 (100 M)	10	11	0.03	-	0.0044
	3 (50 M)	10	11	0.03	-	0.0063

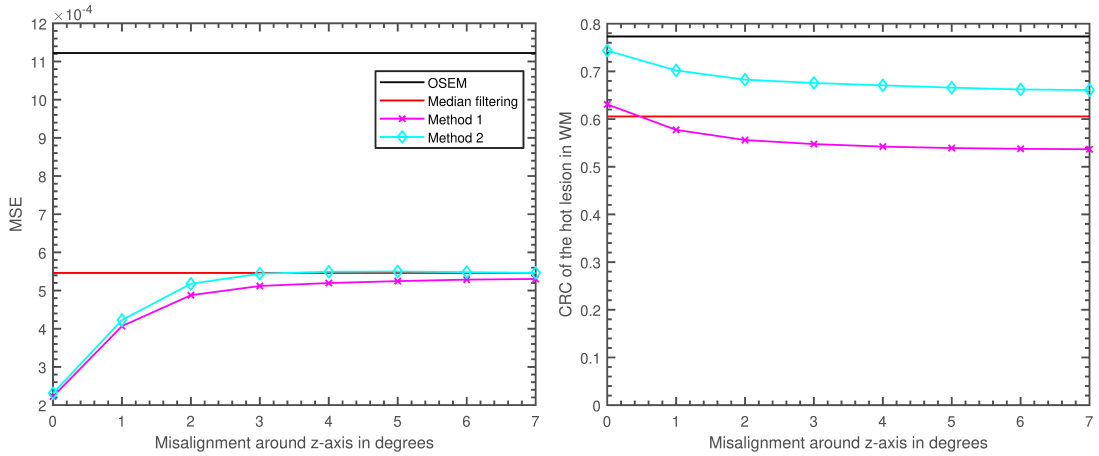


Figure 4.10: Left: mean squared error (MSE) within the brain as a function of rotational misalignment between PET and MR for Method 1 and Method 2. The MSE values for OSEM with no denoising and median filtering are also plotted for reference. Right: The CRC value of the lesion in white matter as as a function of rotational misalignment between PET and MR for Method 1 and Method 2.

The proposed methods are compared with OSEM with no post processing, median filtering and the BM4D algorithm (Maggioni et al. 2013) through different figures of merit. The plots show the proposed methods with parameters reported in Table 4.2. For BM4D algorithm the MATLAB implementation by the authours (`https://www.cs.tut.fi/~foi/GCF-BM3D/`) with a Gaussian noise assumption and automatic noise standard deviation estimation was used.

Figure 4.6a shows the mean n-RMSE value within the brain as a function of number of iterations of OSEM. One can clearly see that the proposed methods provide a lower n-RMSE value compared to median filtering and the BM4D algorithm. Figure 4.6b-d show the bias versus normalized standard deviation of the mean activity for the 3 lesions. For the hot lesion in WM (lesion 1) and the hot lesion in the GM (lesion 2) proposed Method 1 delivers a smaller magnitude bias for any given standard deviation as compared to other methods. However, the bias of the cold lesion in the GM (lesion 3) has a larger magnitude after Method 2 than the other denoising methods.

Figure 4.7 shows the same figures of merit for noise level 2. Note that in Fig.4.7a, after 3 iterations of OSEM, Method 1 provides a lower mean n-RMSE value than Method 2.

Figure 4.8 shows a reconstructed image of one sample realization after 10 iterations of OSEM (column 2) and the resulting images after applying different methods for noise level 2 (top images) and noise level 3 (bottom images). Column 1 shows early terminated OSEM after 3 iterations.

One can see that for both noise levels the proposed methods provide less noisy images with improved cortical boundaries. Also, while the images denoised using Method 1 and Method 2 are very similar, the latter appears to be slightly noisier than the former, when looking at the white matter of images in Fig.4.8.

4.3.3.1 Sensitivity to mis-registration

In this section the sensitivity of the proposed methods to mis-registration between PET and MR is evaluated.

One simulated FDG image at noise level 1 (see table 4.2) and after 10 iterations of OSEM, as described in section III, was used as the noisy PET image in this section. The corresponding MR image was rotated around z axis while keeping the PET image intact to simulate mis-registration between PET and MR images. The rotation angle was changed from 0 to 7 degrees and for each angle both proposed methods were applied to the noisy PET image.

Figure 4.9 shows the noisy PET, the rotated MR image as well as the result of application of Method 1 and Method 2 on the noisy image. Also, the PET-based components of the final denoised images for both Methods are illustrated. For Method 1, the PET-based component corresponds to $\mathbf{P}\boldsymbol{\beta}$ after convergence and for Method 2 the PET-based component is the denoised residual in the second step.

In the case of perfect alignment between PET and MR (0 degree rotation), the PET-based component for both methods is mostly confined to the PET specific lesions, with other regions being near zero. As the misalignment increases, the PET-based component becomes more pronounced. This means that in the boundaries where the PET and MR become more inconsistent with misalignment, the median filtered PET image has more weight in guiding the noise reduction. Therefore, considering that the median filtered PET image is noisier than the T1-weighted MR image, for both methods, increasing the misalignment leads to a decrease in the SNR value.

In Fig.4.10, the mean squared error(MSE) within the brain and the CRC of the lesion in WM is plotted as a function of misalignment in degrees. These values are also plotted for OSEM with no denoising and after median filtering for reference. The MSE increases and then almost plateaus after 3 degrees rotation. Also note that while Method 1 provides a slightly lower MSE value, the CRC value in the Method 1 is lower

than that of Method 2.

4.3.3.2 Combination with PVE correction

Figure 4.11 shows the mean n-RMSE within the brain for Method 2 when applied alone or after PVE correction as a function of the sparsity parameter and neighbourhood length. This figure indicates that the sensitivity of this figure of merit to these parameters is similar for both PVE-corrected and non-PVE corrected images. The parameters leading to minimum mean n-RMSE within the brain are shown in Table 4.3.

Different figures of merit for noisy images with no post-processing and also with PVE correction using the Richardson-Lucy Method are shown in Fig.4.12. Method 2 is shown for parameters indicated in Table 4.3. The results indicate that Method 2 when applied after PVE correction provides a measurable improvement in mean n-RMSE within the brain. The BM4D method, however, does not further improve the image when applied after PVE correction; this might be due to the changes in noise structure after PVE-correction. As shown in Fig. 4.12, PVE correction reduces the magnitude of the bias in all 3 lesions. Method 2 when applied after PVE-correction provides the lowest bias for all standard deviations for both the hot lesion in the GM and the hot lesion in the WM. However, the bias is higher for the cold lesion in GM compared to other denoising methods.

Figure 4.13 shows an example of a noisy image restored using Richardson-Lucy PVE correction algorithm with a combination of median filtering, BM4D and the proposed methods. One can see that noticeable increase in resolution can be obtained by applying proposed methods after Richardson-Lucy deblurring. Method 2 seems to provide more refined boundaries compared to Method 1 when applied on images with PVE.

4.3.4 Application to Real Data

List-mode data was obtained from injection of approximately 185 MBq [^{18}F]FDG to a healthy human participant scanned on a Siemens HRRT scanner. The OSEM with

Table 4.3: Parameters leading to minimum mean n-RMSE within the brain

Method	OSEM iteration	**h**	**L**	**n-RMSE**
Method 2	10	13	0.03	0.0269
PVE correction + Method 2	9	13	0.03	0.0225

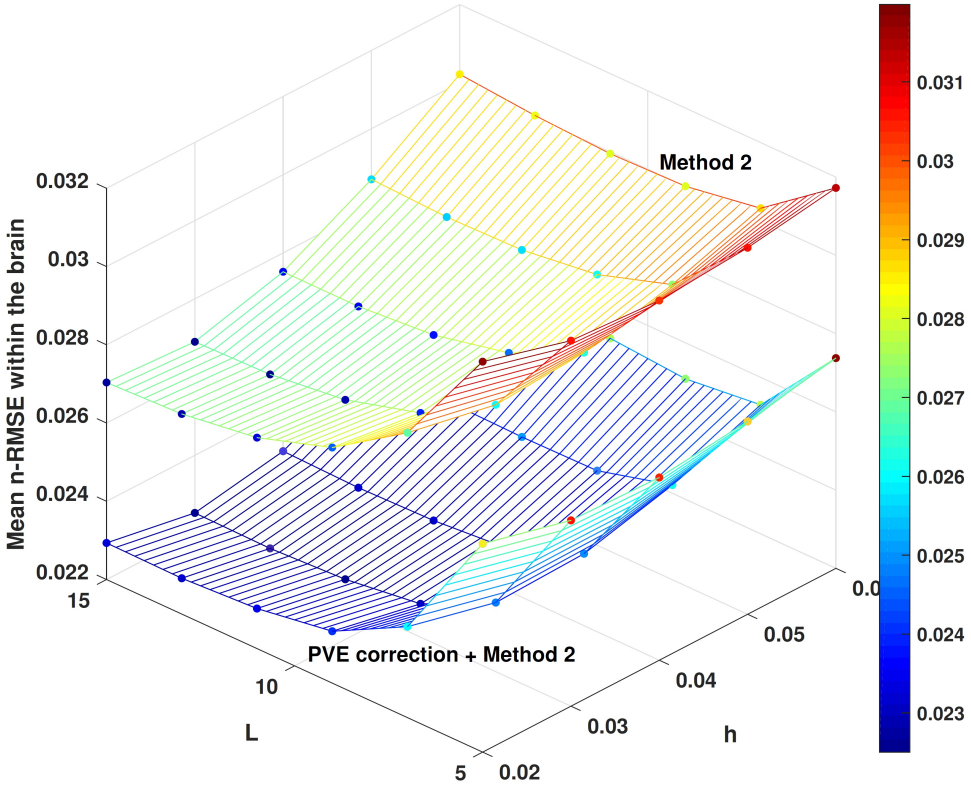


Figure 4.11: The normalized mean n-RMSE within the brain as a function of smoothing parameter h and neighbourhood length L for the Method 2 applied to images without PVE correction (Method 2) and applied to images after PVE correction (PVE correction + Method 2).

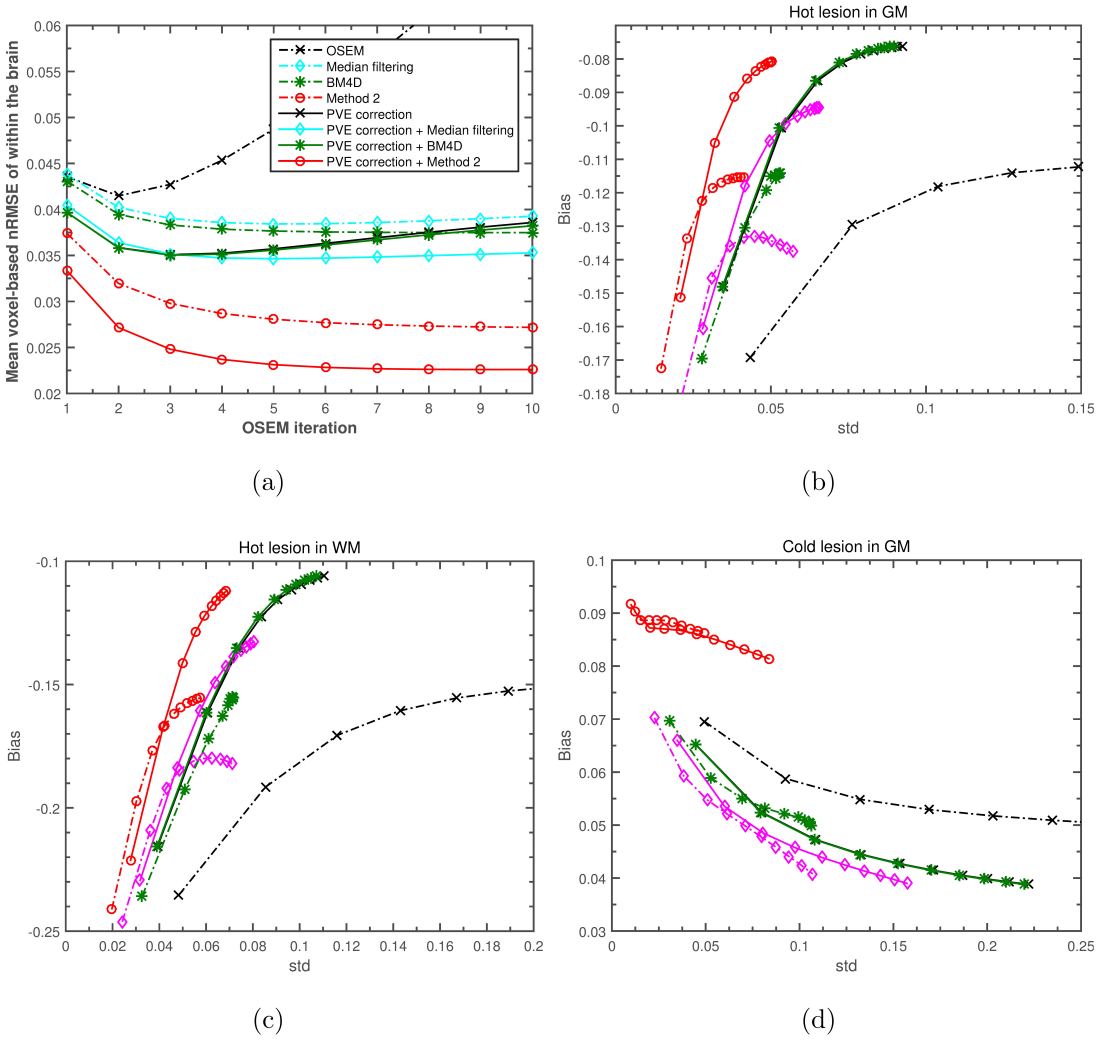


Figure 4.12: Different figures of merit for different methods when applied after OSEM reconstruction with no PVE correction as well as when PVE correction is applied after reconstruction. The parameters used for Method 2 are shown in Table 4.3. a) Mean n-RMSE within the brain as a function of iteration. Bias versus normalized standard deviation of the mean activity in the (b) hot lesion in the WM, (c) hot lesion in the GM and (d) cold lesion in the GM. Note that applying BM4D after PVE correction provides almost no improvement over applying PVE correction alone in the resulting figures of merit.

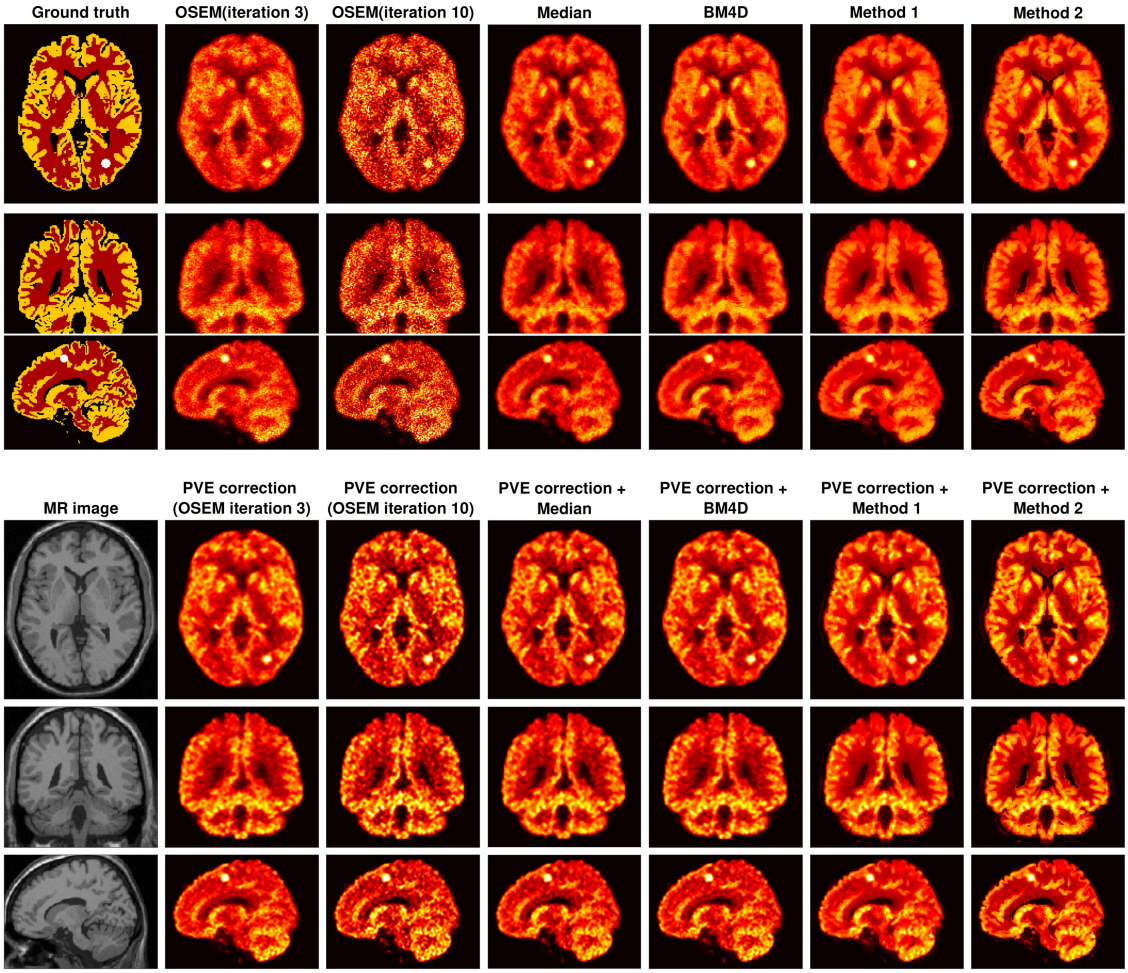


Figure 4.13: An example of a reconstructed FDG simulation. The images are cropped for better visualization. Top figure shows the noisy reconstructed image after 3 and 10 iterations of OSEM and when median filtering, BM4D and the proposed methods are is applied to the image after optimal number of iteration. The bottom figure shows when these methods are applied after PVE-correction.

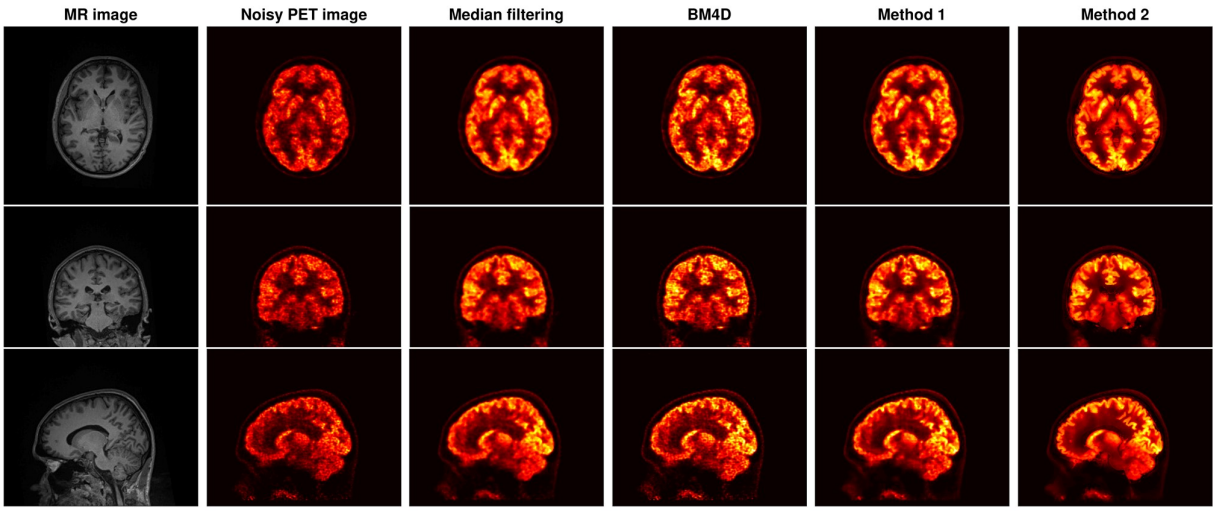


Figure 4.14: A real HRRT [^{18}F]FDG scan after 10 iterations of OSEM with 16 subsets as well as the denoised image after applying median filtering, BM4D and the proposed methods.

resolution modelling algorithm with 16 subsets and 10 iterations was used to reconstruct the image. The duration of the reconstructed frame was 40 minutes starting 20 minutes after injection. A total of about 340 million counts were recorded in this scan.

Figure 4.14 shows the resulting image in the transverse, coronal and sagittal views. Since resolution modelling was used within reconstruction the point spread function is small (comparable to the spatial resolution of 1.44 mm of the HRRT scanner when using 3D OSEM) (Olesen et al. 2009) and hence PVE correction may not be necessary. The result of applying median filtering, BM4D and the proposed methods on the reconstructed image is also shown in this figure. One can clearly see that the proposed methods provide a less noisy image with improved cortical boundaries.

4.4 Discussion and Conclusion

Two novel methods for denoising PET images were proposed. The proposed Method 1 (PET-MR guided kernel re-parameterization) relies on the re-parameterization of the PET image using row-normalized kernel matrices obtained from co-registered MR and median-filtered PET images. An algorithm to estimate the coefficients of these basis

functions while imposing sparsity on coefficients for the PET-based basis functions was proposed.

In the proposed Method 2 (PET-MR GKM with twicing), we use the subject's MR for guided kernel filtering of the noisy image. The residual is then denoised using the median-filtered PET as the guide image. The denoised image and the denoised residual are then added together to obtain the final image. Due to its non-iterative nature, Method 2 is much faster than Method 1.

While both Method 1 and Method 2 rely on PET and MR-based kernels, they use them differently. In Method 1 the sparsity parameter controls the weight of information between the MR and PET images to guide the denoising process. This parameter needs to be small to ensure that PET unique information does not suffer from strong blurring in the resulting image. In Method 2 however, since only the PET-based filter is applied to the residual, the method is not prone to severe blurring of the PET specific regions.

We have shown through simulation that the proposed methods are sensitive to misregistration between PET and T1-weighted MR images. Misalignment between PET and MR in our simulation led to noisier images. Also, a 6 degree misalignment around the z-axis can be considered as an example of a situation where there is noticeable inconsistency between PET and MR. Severe PET or MR unique abnormalities may have a similar effect. We have shown that these inconsistencies did not lead to serious artefacts in the resulting images. This is due to the fact that we are using both PET and MR in guiding our denoising process, therefore when MR information is inconsistent, both methods (given that the sparsity parameter for Method 1 is small) implicitly rely on PET to guide the denoising process in inconsistent regions.

Both methods can deal with high noise levels. Therefore, as also demonstrated in the simulation, it is suggested to run OSEM for more iterations than used in conventional practice in order to reduce the bias in the resulting image. We also show

that the proposed methods, when applied after a conventional deblurring algorithm, can be considered as a voxel-based MR guided PVE-correction method. Therefore the proposed methods should be of great interest for voxel-based brain studies.

In this work two types of simulation were used to evaluate our methods, one without and one with PVE. We use simulation without PVE to explore the capability of the method in reducing the noise. However, real PET images also suffer from PVE. In presence of non-corrected PVE in the PET image, the consistency between PET and MR boundaries is reduced which will lead to less refined boundaries after restoration. Applying a simple deconvolution on the noisy image will highly correlate the noise, resulting in artefacts in the deblurred images. In this case, proposed Method 2 provides a lower overall error (mean nRMSE within the brain) than proposed Method 1.

One limitation of the proposed methods is that for a given set of parameters, the performance of the methods in PET-specific regions depends on the size and the activity of that region. For example, for a very small PET specific lesion, a smaller neighbourhood size will lead to improved contrast in the lesion, at the expense of increased noise in the whole brain.

One advantage of the proposed methods compared to many other MR guided PVE correction methods in the literature is that both proposed methods can work without the need for segmentation of the MR image, which is an important source of error in PVE correction (Frouin et al. 2002).

Note that in proposed Method 1, by including the forward model in the least squares objective function, this method can easily be extended to a reconstruction algorithm in which coefficients for the basis functions are estimated from the PET measured data. The proposed denoising framework can be used for denoising any functional modality using corresponding anatomical images, e.g. CT for SPECT and PET, T1-weighted MR for fMRI. Furthermore, images of other radiotracers with higher

image quality could potentially also be used to guide the denoising process, similarly to what was proposed for reconstruction by Ellis et al(Ellis et al. 2018).

This work only evaluated the use of the proposed methods on a simulated FDG brain image and one real FDG dataset. Further application and analysis of the proposed methods on more real data is needed to ensure their applicability to other PET radiotracers as well as to other applications e.g. whole body PET or preclinical PET where the method might be more sensitive to PET-MRI registration issues.

Acknowledgment

The authors would like to acknowledge Dr Caroline Paquette for providing the real [^{18}F]FDG data and the T1-weighted MR data.

This study was funded in part by the Canadian Institute of Health Research (CIHR MOP-111169), les Fonds de Research Santé Québec Pfizer Innovation fund, National Science and Engineering Research Council of Canada (NSERC 4140438 - 2012), the EPSRC (EP/M020142/1) and the Wellcome EPSRC Centre for Medical Engineering at King's College London (WT 203148/Z/16/Z). We also wish to thank the Famille Louise and André Charron for financial support.

References

Andersen, A. H. & Kak, A. C. (1984), 'Simultaneous algebraic reconstruction technique (sart): a superior implementation of the art algorithm', *Ultrasonic imaging* **6**(1), 81–94.

Ardekani, B. A., Braun, M., Hutton, B. F., Kanno, I. & Iida, H. (1996), 'Minimum cross-entropy reconstruction of PET images using prior anatomical information', *Physics in Medicine and Biology* **41**(11), 2497.

Barrett, H. H., Wilson, D. W. & Tsui, B. M. (1994), 'Noise properties of the EM algorithm. I. Theory', *Physics in medicine and biology* **39**(5), 833.

Bataille, F., Comtat, C., Jan, S., Sureau, F. & Trebossen, R. (2007), 'Brain PET partial-volume compensation using blurred anatomical labels', *IEEE Transactions on Nuclear Science* **54**(5), 1606–1615.

Beck, A. & Teboulle, M. (2009), 'A fast iterative shrinkage-thresholding algorithm for linear inverse problems', *SIAM journal on imaging sciences* **2**(1), 183–202.

Boussion, N., Cheze Le Rest, C., Hatt, M. & Visvikis, D. (2009), 'Incorporation of wavelet-based denoising in iterative deconvolution for partial volume correction in whole-body PET imaging', *European journal of nuclear medicine and molecular imaging* **36**(7), 1064–1075.

Carlier, T., Ferrer, L., Necib, H., Bodet-Milin, C., Rousseau, C. & Kraeber-Bodéré, F.

(2014), 'Clinical necr in 18f-fdg PET scans: optimization of injected activity and variable acquisition time. relationship with snr', *Physics in medicine and biology* **59**(21), 6417.

Chan, C., Fulton, R., Feng, D. D. & Meikle, S. (2010), Median non-local means filtering for low snr image denoising: Application to PET with anatomical knowledge, *in* 'IEEE Nuclear Science Symposuim & Medical Imaging Conference', IEEE, pp. 3613–3618.

Cocosco, C. A., Kollokian, V., Kwan, R. K.-S., Pike, G. B. & Evans, A. C. (1997), Brainweb: Online interface to a 3D MRI simulated brain database, *in* 'NeuroImage', Citeseer.

Dabov, K., Foi, A., Katkovnik, V. & Egiazarian, K. (2007), 'Image denoising by sparse 3-d transform-domain collaborative filtering', *IEEE Transactions on image processing* **16**(8), 2080–2095.

Daubechies, I., Defrise, M. & De Mol, C. (2003), 'An iterative thresholding algorithm for linear inverse problems with a sparsity constraint', *arXiv preprint math/0307152* .

Ellis, S., Mallia, A., McGinnity, C. J., Cook, G. J. R. & Reader, A. J. (2018), 'Multitracer guided pet image reconstruction', *IEEE Transactions on Radiation and Plasma Medical Sciences* **2**(5), 499–509.

Epperson, J. F. (2013), *An introduction to numerical methods and analysis*, John Wiley & Sons.

Fessler, J. A. (1997), Grouped coordinate descent algorithms for robust edge-preserving image restoration, *in* 'Optical Science, Engineering and Instrumentation'97', International Society for Optics and Photonics, pp. 184–194.

Frouin, V., Comtat, C., Reilhac, A. & Gregoire, M.-C. (2002), 'Correction of partial-volume effect for PET striatal imaging: fast implementation and study of robustness', *Journal of Nuclear Medicine* **43**(12), 1715–1726.

Frouin, V., Comtat, C., Reilhac, A. & Grégoire, M.-C. (2002), 'Correction of partial-volume effect for PET striatal imaging: Fast implementation and study of robustness', *Journal of Nuclear Medicine* **43**(12), 1715–1726. cited By 99.

Hudson, H. M. & Larkin, R. S. (1994), 'Accelerated image reconstruction using ordered subsets of projection data', *IEEE transactions on medical imaging* **13**(4), 601–609.

Jacobs, F., Matej, S. & Lewitt, R. (1998), 'Image reconstruction techniques for PET', *Dep. Radiol., Univ. Pennsylvania, Philadelphia, Tech. Rep. MIPG245* .

Jagust, W. J., Bandy, D., Chen, K., Foster, N. L., Landau, S. M., Mathis, C. A., Price, J. C., Reiman, E. M., Skovronsky, D., Koeppe, R. A. et al. (2010), 'The Alzheimer's disease neuroimaging initiative positron emission tomography core', *Alzheimer's & Dementia* **6**(3), 221–229.

Jiao, J., Markiewicz, P., Burgos, N., Atkinson, D., Hutton, B., Arridge, S. & Ourselin, S. (2015), Detail-preserving PET reconstruction with sparse image representation and anatomical priors, *in* 'Information Processing in Medical Imaging', Springer, pp. 540–551.

Lewitt, R. M. (1990), 'Multidimensional digital image representations using generalized kaiser–bessel window functions', *JOSA A* **7**(10), 1834–1846.

Lewitt, R. M. (1992), 'Alternatives to voxels for image representation in iterative reconstruction algorithms', *Physics in Medicine and Biology* **37**(3), 705.

Lucy, L. B. (1974), 'An iterative technique for the rectification of observed distributions', *The astronomical journal* **79**, 745.

Maggioni, M., Katkovnik, V., Egiazarian, K. & Foi, A. (2013), 'Nonlocal transform-domain filter for volumetric data denoising and reconstruction', *IEEE transactions on image processing* **22**(1), 119–133.

Mandelkern, M. A. (1995), 'Nuclear techniques for medical imaging: positron emission tomography', *Annual Review of Nuclear and Particle Science* **45**(1), 205–254.

Mignotte, M. & Meunier, J. (2000), 'Three-dimensional blind deconvolution of spect images', *IEEE Transactions on Biomedical Engineering* **47**(2), 274–280.

Milanfar, P. (2013), 'A tour of modern image filtering: New insights and methods, both practical and theoretical', *IEEE Signal Processing Magazine* **30**(1), 106–128.

Müller-Gärtner, H. W., Links, J. M., Prince, J. L., Bryan, R. N., McVeigh, E., Leal, J. P., Davatzikos, C. & Frost, J. J. (1992), 'Measurement of radiotracer concentration in brain gray matter using positron emission tomography: MRI-based correction for partial volume effects', *Journal of Cerebral Blood Flow & Metabolism* **12**(4), 571–583.

Novosad, P. & Reader, A. J. (2016), 'MR-guided dynamic PET reconstruction with the kernel method and spectral temporal basis functions', *Physics in medicine and biology* **61**(12), 4624.

Nuyts, J., Fessler, J. et al. (2003), 'A penalized-likelihood image reconstruction method for emission tomography, compared to postsmoothed maximum-likelihood with matched spatial resolution', *Medical Imaging, IEEE Transactions on* **22**(9), 1042–1052.

Olesen, O. V., Sibomana, M., Keller, S. H., Andersen, F., Jensen, J., Holm, S., Svarer, C. & Højgaard, L. (2009), Spatial resolution of the HRRT PET scanner using 3D-OSEM PSF reconstruction, *in* 'Nuclear Science Symposium Conference Record (NSS/MIC), 2009 IEEE', IEEE, pp. 3789–3790.

Rangarajan, A., Hsiao, T. & Gindi, G. (2000), 'A Bayesian joint mixture framework for the integration of anatomical information in functional image reconstruction', *Journal of Mathematical Imaging and Vision* **12**(3), 199–217.

Richardson, W. H. (1972), 'Bayesian-based iterative method of image restoration', *JOSA* **62**(1), 55–59.

Rousset, O., Ma, Y. & Evans, A. (1998), 'Correction for partial volume effects in PET: Principle and validation', *Journal of Nuclear Medicine* **39**(5), 904–911. cited By 414.

Rousset, O., Rahmim, A., Alavi, A. & Zaidi, H. (2007), 'Partial volume correction strategies in PET', *PET clinics* **2**(2), 235–249.

Strother, S., Casey, M. & Hoffman, E. (1990), 'Measuring PET scanner sensitivity: relating countrates to image signal-to-noise ratios using noise equivalents counts', *IEEE transactions on nuclear science* **37**(2), 783–788.

Tahaei, M. S. & Reader, A. J. (2016), 'Patch-based image reconstruction for PET using prior-image derived dictionaries', *Physics in Medicine and Biology* **61**(18), 6833.

Teo, B.-K., Seo, Y., Bacharach, S. L., Carrasquillo, J. A., Libutti, S. K., Shukla, H., Hasegawa, B. H., Hawkins, R. A. & Franc, B. L. (2007), 'Partial-volume correction in PET: validation of an iterative postreconstruction method with phantom and patient data', *Journal of Nuclear Medicine* **48**(5), 802–810.

Thielemans, K., Tsoumpas, C., Mustafovic, S., Beisel, T., Aguiar, P., Dikaios, N. & Jacobson, M. W. (2012), 'Stir: software for tomographic image reconstruction release 2', *Physics in medicine and biology* **57**(4), 867.

Tukey, J. W. (1977), 'Exploratory data analysis'.

Turkheimer, F. E., Boussion, N., Anderson, A. N., Pavese, N., Piccini, P. & Visvikis, D.

(2008), 'PET image denoising using a synergistic multiresolution analysis of structural (MRI/CT) and functional datasets', *Journal of Nuclear Medicine* **49**(4), 657–666.

Wang, G. & Qi, J. (2015), 'PET image reconstruction using kernel method', *IEEE transactions on medical imaging* **34**(1), 61–71.

Yan, J., Lim, J. C.-S. & Townsend, D. W. (2015), 'MRI-guided brain PET image filtering and partial volume correction', *Physics in medicine and biology* **60**(3), 961.

Zaidi, H., Montandon, M.-L. & Alavi, A. (2007), 'Advances in attenuation correction techniques in PET', *PET clinics* **2**(2), 191–217.

Chapter 5

Blind PET Image Restoration

Preface The proposed methods in the previous chapter, work best after applying a conventional de-blurring algorithm for which the point spread function (PSF) in known a priori. For large multi-centre datasets, information about the PSF may not be available. In this chapter we try to address this by training a machine learning model on PET simulation data with different characteristics. These characteristics which describe the noise and PVE levels include PSF, activity level and iteration number of OSEM used to reconstruct the image. The goal is to train a single model that can be applied to images coming from different scanners and different radiotracers.

This work is in preparation for submission to Transactions on Medical Imaging as: Tahaei, Marzieh S., Andrew J. Reader, and D. Louis Collins. "Blind PET Image Restoration: A 3D CNN with Uncertainty Estimation.", Submitted to IEEE Transaction on Medical Imaging.

Blind PET Image Restoration: A 3D CNN with Uncertainty Estimation

Marzieh S. Tahaei[1], Andrew J. Reader[2], D. Louis Collins[1]

[1] McConnell Brain Imaging Centre, Montreal Neurological Institute, McGill University, Montreal, Canada

[2]Division of Imaging Sciences and Biomedical Engineering, Department of Biomedical Engineering, King's College London, St. Thomas' Hospital, London, UK

Abstract

PET images are subject to severe resolution degradation and noise in the scan and reconstruction process. Denoising and partial volume correction methods often have several hyper-parameters for which accurate tuning requires knowledge of the amount of noise and partial volume effect (PVE) in the data. Often, this information is either not available or cannot be easily quantified in large multicentre datasets in which an image may come from a different scanner with different radiotracers, acquisition protocols and reconstruction parameters. Through learning from simulations with varying levels of noise and PVE, a hyper parameter-free framework for restoring PET images is proposed. Our method uses a deep learning Inception-Res architecture to restore 3D patches of brain PET images while estimating the aleatoric uncertainty of each voxel in the patch. Two models are trained with and without the corresponding T1-weighted MR images. The proposed models are evaluated on simulated test data with different figures of merit. The results show significant improvements in image quality and low error in regions where PET is different from MR. Applying our model on real PET data of different radiotracers and from different scanners demonstrates the generalizability of the network to unseen distributions.

Key words: PET-MR, denoising, PVE correction, deep learning, Convolutional neural networks, Aleatoric uncertainty, Inception-Res networks

Introduction

Through the use of different radiotracers, PET imaging provides a unique tool to understand the human brain in health and disease. However, limitations in radiation dose due to the potential health risks of radioactivity, along with poor scanner sensitivity, lead to a large amount of noise in reconstructed PET images. In the last two decades, many methods have been developed to reduce noise in PET images by incorporating different prior information in the objective function within reconstruction (Ardekani et al. 1996, Rangarajan et al. 2000, Nuyts et al. 2003, Tahaei & Reader 2016). These methods can only be applied to raw PET measurement data, which is rarely available in large multicentre datasets.

Post reconstruction methods (Turkheimer et al. 2008, Chan et al. 2010, Tahaei et al. 2019) are another category of methods that can be applied to PET images after they are reconstructed using conventional reconstruction methods such as Filtered Back Projection and ordered subset expectation maximization (OSEM) (Hudson & Larkin 1994).

Such methods usually have one or more hyper-parameters, and their choice often depends on the level of noise in the image which itself depends on the injected radioactivity dose, scan duration and the number of iterations of the reconstruction method. Accurate tuning of these hyper-parameters can only be done for simulation data for which the ground truth is available. For real images, however, due to lack of ground truth for in vivo data, the choice of these hyper-parameters requires manual optimization and is often non-optimal.

In addition to noise, partial volume effects (PVE) are another very important source of error in PET image quantification. PVE is affected by the positron range of the radiotracer, the photon non-collinearity effect, along with the limited resolution of the scanner and the tissue fraction effects. PVE is quantified using a point spread function (PSF).

Most PVE correction methods, whether those that use the PET information only or those that use added anatomical information, need an estimation of the PSF to be known. For large multicentre datasets in which images are acquired using different scanners in different centres, the PSF function associated with the image in hand may not be available. Also the choice of reconstruction algorithm and the number of iterations after which the reconstruction is terminated has a direct effect on the resolution of the resulting image. In fact, early termination of OSEM is a common practice to control the noise at the expense of loss of resolution.

In natural image processing, blind deconvolution is a class of algorithms that aims to deconvolve a degraded image without knowing the PSF a priori (Ayers & Dainty 1988, Chan & Wong 1998, Levin et al. 2009). These methods work in situations where blurring is the main degradation in the image. For PET images, however, because of the high amount of noise present in reconstructed images, the applicability of these methods is very limited.

Motivated by the above discussion, this work aims to propose a post-reconstruction framework that is blind; meaning that the method does not require any prior knowledge about image characteristics such as radiotracer, activity levels, PSF nor iteration number of the reconstruction algorithm.

This challenging goal is achieved by learning the parameters from simulated training data generated for a reasonable range of the possible factors contributing to noise and PVE. Our proposed models are trained both with and without additional anatomical information.

Also, to ensure that using anatomical information does not impose a bias in PET on regions that do not follow anatomical boundaries, many PET specific lesions are embedded in the training data. We show that, unlike anatomically guided methods, by doing this, the network is able to maintain contrast of these PET-specific regions.

An important contribution of the proposed work is that aleatoric uncertainty in

the data is also estimated as an output of the proposed network. We also show that aleatoric uncertainty at each voxel is highly correlated with the error. Therefore, its estimation provides useful insight for test data for which ground truth is not known. The generalizability of the model is further investigated by applying the method on real data obtained from different radiotracers and different scanners.

The paper is organized as follows: after introduction, related work is discussed and a detailed description of the proposed model is presented in section I. In section II the proposed framework with different variations is evaluated using simulation and real data. Finally, a concluding summary and a discussion of the proposed framework are presented in Section IV.

5.1 Methodology

5.1.1 Related work

Deep learning methods use hierarchical compositions of nonlinear transformations of data to learn the distribution of interest. In the last decade, they have proven to be superior to other machine learning techniques in many different tasks. Among deep learning frameworks, convolutional neural networks (CNN) have gained increasing popularity in solving computer vision problems including image denoising and deblurring. CNNs have also been successful in solving various medical image processing problems such as segmentation (Ronneberger et al. 2015, Çiçek et al. 2016, Kamnitsas et al. 2017), denoising (Gondara 2016, Yang et al. 2018) and reconstruction (Zhu et al. 2018, Yang et al. 2017, Gong et al. 2018) of medical images of different modalities.

In the context of image denoising and deblurring, Jain & Seung (2009) showed that a CNN provides promising performance in reducing noise in natural images. Mao et al. (2016) used a CNN with deconvolution layers and skip connections for different image restoration tasks such as denoising and super-resolution. They showed that the use of skip connections in very deep architectures enables the network to recover

high-frequency components of the image.

Zhang et. al. used residual learning and batch normalization in a CNN to speed up training and improve performance for multiple tasks, including Gaussian denoising, single image super-resolution, and JPEG image deblocking (Zhang et al. 2017).

These are a few of many CNN models proposed for image denoising, deblurring, and super-resolution (Dong et al. 2015, Zhang et al. 2017a, Nimisha et al. 2017, Zhang et al. 2018). The recent trend in this regard is to use more advanced loss functions such as perceptual loss obtained from a pre-trained model and to use a generative adversarial framework to improve the resolution of the resulting images (Johnson et al. 2016, Yang et al. 2018).

For PET imaging, however, attention of the community has been mainly focused on using different variants of CNNs for estimation of standard-dose PET images from low-dose reconstructed PET data (Xiang et al. 2017, Xu et al. 2017, Wang et al. 2018, Kaplan & Zhu 2018, Chen et al. 2018). These methods are motivated by the fact that noise in PET images is highly anti-correlated with the dose. But there is no direct relationship between PVE and dose and hence these methods do not tend to improve the resolution of the resulting images. Training in these studies is based on real paired images with both low and standard-dose, obtained either by scanning the subjects twice or by sub-sampling from PET measurement data to produce low-dose images from standard-dose data. While training on real data is very appealing since no simulation can exactly mimic the actual physics of imaging, it enforces a few limitations. In fact, the scarcity of the real data limits the generalization of the network to specific radiotracer types, dose levels and other factors associated with the training data. To compensate, simulation data allows us to generate unlimited data with variabilities in different factors and hence is more generalizable.

5.1.2 Proposed Method

Let $\mathbf{y}$ be the reconstructed noisy and resolution degraded PET image obtained by applying an iterative reconstruction algorithm on PET measurement data. Our goal is to devise an image restoration method that estimates the ground truth image $\mathbf{x}$ using $\mathbf{y}$ with or without the registered T1-weighted MR image $\mathbf{m}$. We call this estimate $\hat{\mathbf{x}}$. The proposed method should be hyper-parameter-free; i.e., be able to reduce the noise and PVE in a given PET image without any prior information about the type of radiotracer, its dose (activity level), number of iterations or PSF. An Inception-Res network (Szegedy et al. 2017) is used to achieve this goal. In the following, a brief overview of the different components of the proposed model is presented.

5.1.2.1 Inception network

An Inception network (Szegedy et al. 2015) is composed of multiple Inception modules stacked together. An Inception module uses multiple kernels of different sizes in parallel and concatenates the resulting feature maps together. This is inspired by a model of the primate visual cortex, where different filters are used for processing the visual input. The use of multiple filters in parallel enables each module of the network to capture features at different scales. It also makes the network less sensitive to the choice of filter size.

One problem with this approach is that convolutions are computationally expensive and using multiple convolutions in parallel will lead to computational explosion in a deep Inception network. In order to address this problem, in each Inception module, before applying convolutions of different sizes, 1×1 convolutions are performed. Each 1×1 convolution in a CNN layer acts as a voxel-wise weighted linear composition across feature maps resulting in a single feature map.

5.1.2.2 Residual learning

It has been shown that deeper networks, despite their increased learning capacity, tend to have higher training errors than shallower counterparts. This phenomenon

is known as network degradation. Residual learning (He et al. 2016, Srivastava et al. 2015, Mao et al. 2016) has been proposed to address this issue in deep networks. In residual learning, the network layers (or blocks) are explicitly forced to learn the mapping from the input to the residual signal, i.e. if $H(\mathbf{y})$ is the desired mapping from input $\mathbf{y}$ to output $\mathbf{x}$, then instead of learning $H(\mathbf{y})$, $H(\mathbf{y}) - \mathbf{x}$ is learned. The rationale for doing this is that learning the mapping of the residual is easier than learning the mapping from input to output, similar to the way that learning a zero mapping is easier than learning an identity mapping.

5.1.2.3 Heteroscedastic Aleatoric Uncertainty

Uncertainty estimation is an interesting ongoing trend in deep learning. Uncertainty estimation for medical images is particularly important since an incorrect information in data may lead to an incorrect clinical decision or an incorrect conclusion from a pharmaceutical study. Uncertainty can be divided into two categories: aleatoric uncertainty and epistemic uncertainty. Aleatoric uncertainty is the data-inherent uncertainty due to lack of information in the data. In computer vision applications this can be the noise or the blurriness in the image. Epistemic uncertainty, on the other hand, is the uncertainty associated with the model (i.e., structure and parameters) rather than the data.

There are two types of aleatoric uncertainties: homoscedastic uncertainty assumes that the uncertainty is fixed for each datum while in heteroscedastic uncertainty each input dimension can have a different uncertainty value. For medical images, the amount of uncertainty raised during the measurement process is voxel-dependent. Therefore for PET image restoration, we aim to estimate heteroscedastic uncertainty as it provides an estimate of uncertainty for each voxel in the image

In deep learning heteroscedastic uncertainty can be easily learned from training data in an unsupervised manner, by making small changes to the network and loss function (Kendall & Gal 2017). This is achieved by allowing the network not only to

estimate the target, but also the uncertainty of the generated target.The uncertainty of a restored voxel is considered as the standard deviation (σ) of its estimation and can be learned from the training data using the following loss:

$$L(\theta|x,y) = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{2\sigma(x_i)^2}\|f(x_i)-y_i\|_2^2 + \frac{1}{2}\log\sigma(x_i)^2 \tag{5.1}$$

Note that by dividing the mean squared error by σ^2, a larger error in the estimated restored image leads to a larger σ value and hence higher uncertainty. Also, the second term in Eq. 5.1 is added to prevent the network from outputting infinite uncertainty everywhere in the image to minimize the loss.

5.1.3 Generating simulation data

In medical imaging, due to the in-vivo nature of quantification, ground truth images of real subject scans are extremely difficult, if not impossible, to acquire. Therefore, simulation is often used in PET imaging to generate ground truth functional maps from an anatomical image. Here segmented MR images are used to generate ground truth functional maps which are then combined with an accurate model of the PET scanning process to generate PET images. Training, validation and test simulation data were obtained using 20 T1-weighted MR images from BrainWEB (`http://brainweb.bic.mni.mgill.ca`) (Aubert-Broche et al. 2006) as well as their white matter and grey matter segmentations. Figure. 5.1 shows the pipeline designed to generate simulated PET data from these images. In short, the grey matter and white matter segmentations provided by BrainWEB are used to drive the spatial distribution of the functional signal that is input to a PET scanner model to simulate sinogram acquisition and PET image reconstruction. Details of the simulation process are described below.

5.1.3.1 Anatomical and functional simulation

Our goal is to generate multiple examples of registered MRI and PET data that can be used to train the network so that the model is equivariant with respect to the

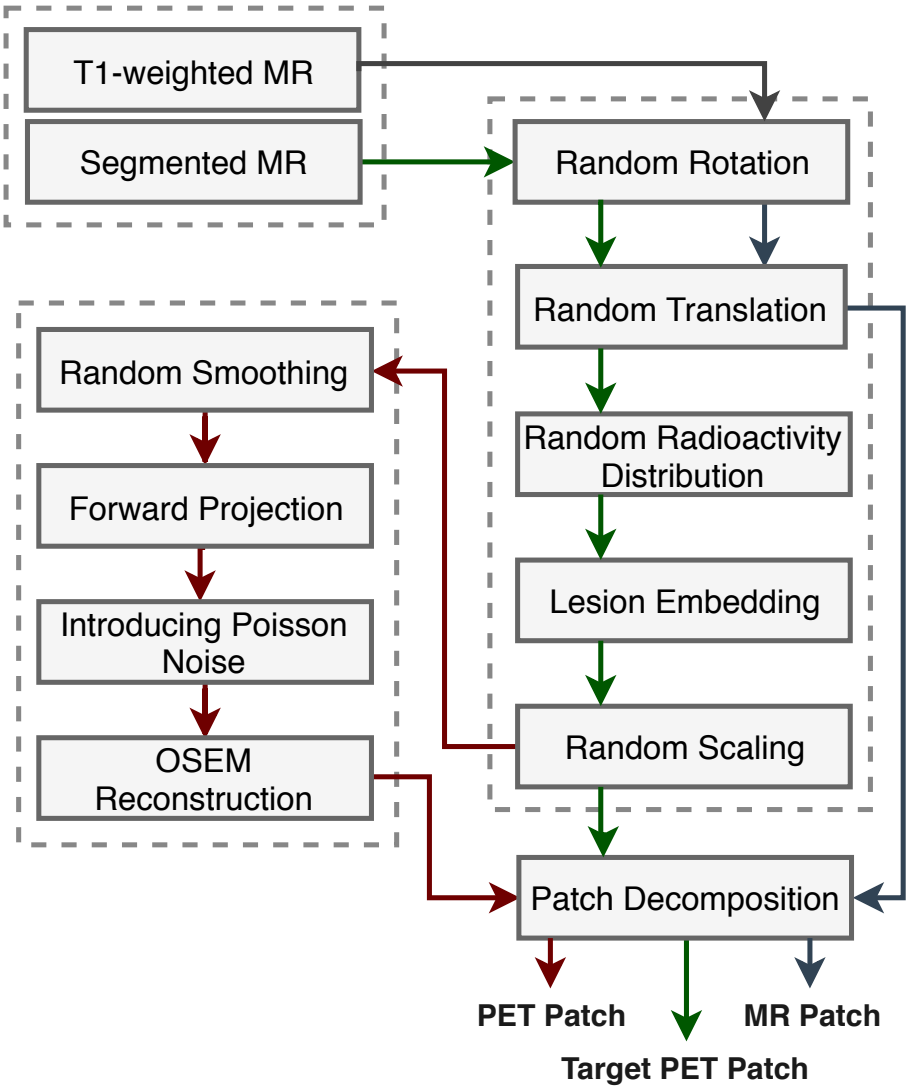


Figure 5.1: Pipeline for simulating training data. First, data augmentation is used to obtain ground truth pseudo PET images from BrainWEB anatomical segmentations. Then this ground truth image is smoothed and transferred into sinogram space and noise is introduced to the the resulting sinogram data. OSEM is then used to reconstruct sinogram data to generate reconstructed PET images. Finally, the anatomical T1-weighted MR, the pseudo ground truth PET and its noisy reconstruction are decomposed into 3D patches.

anatomy and the type and the amount of injected radiotracer. Also, the model should be able to preserve PET specific signal absent in the anatomy.

First, both simulated T1-weighted volume and segmentation maps are resampled

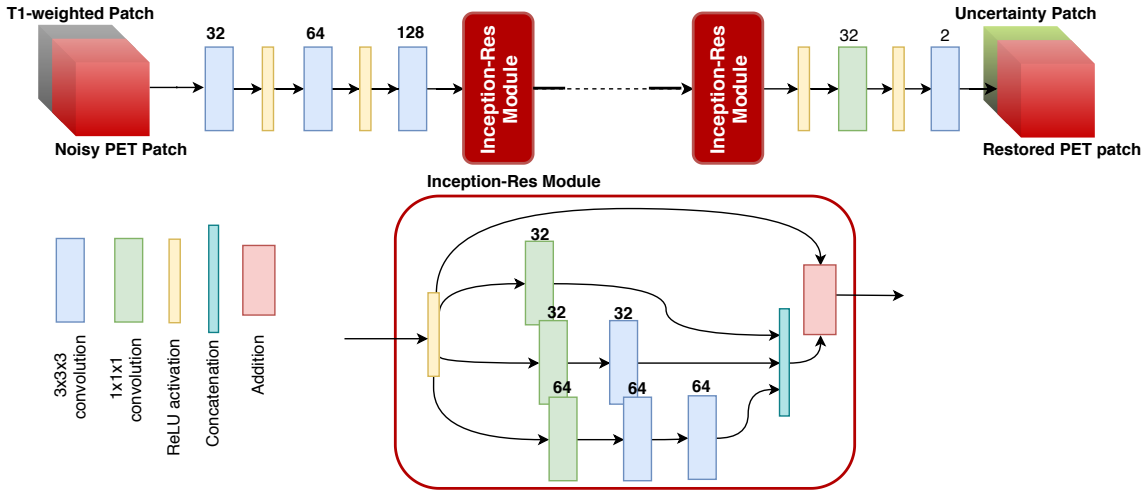


Figure 5.2: The proposed architecture for blind MR-guided PET restoration using Inception-Res network while estimating the uncertainty.

onto a 1.22mmx1.22mmx1.22mm grid (equivalent to the high resolution research tomograph(HRRT) (Schmand et al. 1998) voxel size). Data augmentation is used to generate multiple anatomical volumes from each subject in the BrainWEB dataset and also to help achieve equivariance with respect to anatomy. For augmentation, we applied random rotations (range 10 to 10 degrees) and translations (-10 to 10 voxels) in all 3 directions (x, y and z), to generate an augmentation of each T1-weighted MR volume. The same augmentation transformations are then applied to the corresponding 3D grey and white matter segmentation maps used to generate new corresponding functional data.

While one may want to use the functional values proportional to a real PET image to simulate a functional map, we would like our model to be generalizable to different functional maps resulting from different radiotracers in different conditions. Therefore, to achieve equivariance with respect to functional activity patterns, we set the mean value of grey matter to a random value between 0 and 12. Another random value between 0 and 4 is selected as the value for white matter. This way most

functional maps have grey matter with higher activity than white matter which is the common pattern for most radiotracers. Nonetheless, this random activity selection also allows generating functional maps with higher activity in the white matter.

As previously mentioned, for our MR-guided model, we want to ensure that the network is able to maintain PET-specific signals that are absent in the anatomical image. To achieve this, we added 500 lesions to each functional map obtained from the previous step while keeping the *T1-weighted MR data* intact. For generating each lesion, a random voxel in the brain was selected as the central location of the lesion. The lesion intensity was obtained by multiplying the intensity value of this voxel by a random number between 0 and 5. The resulting intensity value was then used to add a circular shape lesion with a random size (radius from 1 to 30 voxels) to the selected location. In order to train a model that is able to restore images of different count levels, the image is randomly scaled to obtain 100M to 400M counts.

5.1.3.2 PET simulation

The resulting simulated ground truth functional maps are smoothed with a Gaussian filter with a random FWHM ranging from 1mm to 5mm to mimic varying partial volume effects. The 3D functional map created above is forward-projected through an HRRT scanner model to sinogram space with radial bins, 288 azimuthal angles, and 104 projection planes. Linear attenuation coefficient factors from (Zaidi et al. 2007) were incorporated in the forward projection. Scatters and randoms events are simulated by a smoothed sinogram with activity one-quarter of the total activity and are added to the projected ground-truth image as in (Tahaei & Reader 2016). Poisson noise is then introduced to the resulting sinogram data. The resulting sinogram is then used to reconstruct the noisy PET image using ordinary Poisson Ordered Subset Expectation-Maximization algorithm (OP-OSEM) with 16 subsets and varying iterations (ranging from 4 to 12).

Out of the 20 subjects in BrainWEB, 18 are used to generate 162 training volumes

by repeating the above procedure 9 times. From the remaining 2 MR volumes, one is used to generate 9 validation volumes and one is used to generate 18 test volumes.

For training and validation, the resulting noisy PET images, their corresponding ground truth functional maps and the T1-weighted MR images are cropped and then decomposed to non-overlapping patches. Since the network is fully convolutional, an arbitrary patch size can be used to train and test. For training and validation data, images are decomposed to non-overlapping patches of size 15×15×15. For test data, to reduce the checker-board effect of the resulting images we use over-lapping patches with a small gap of 2 voxels. But in order to limit the computation and memory cost, a smaller patch size of $9 \times 9 \times 9$ voxels was used. Each noisy patch along with its corresponding MR patch is given to the network as a two-channel input. The residual patch (ground truth functional map - noisy PET image) is given to the network as the output to match. The network can be run without MR guidance; the only change to the network is removing one channel from input (the MR) while keeping the rest of the network intact.

Method	**Mean PSNR**	**Mean SSIM**	**Mean Lesions NMAE**
None (original noisy image)	24.95	0.83	0.29
GKM (no prior knowledge)	26.45	0.88	0.33
PET only Inception (no prior knowledge)	26.42	0.87	0.25
MR-guided Inception (no prior knowledge)	27.67	0.92	0.24
MR-guided Inception + Uncertainty (no prior knowledge)	27.65	0.91	0.24

Table 5.1: Mean peak signal to noise ratio (PSNR), mean structural similarity index (SSIM) and mean lesions normalized mean absolute error (NMAE) for the original noisy test images and after application of GKM, PET only Inception, MR-guided Inception (without uncertainty) and MR-guided Inception with uncertainty.

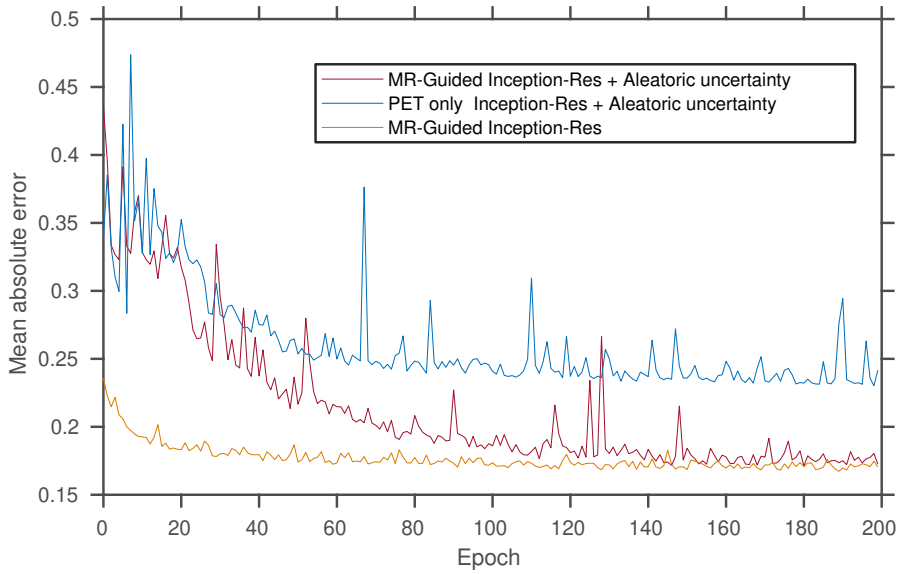


Figure 5.3: Mean absolute error loss of randomly selected validation patches for PET only reconstruction and for MRguided reconstruction (both with and without aleatoric uncertainty) as a function of epoch.

5.1.4 Real data

In order to examine the ability of the proposed model to generalize to real data, it is tested on PET images of three different radiotracers from real subjects. For all real PET datasets, the iterative reconstruction algorithm provided by the scanner was used for reconstruction. All frames were linearly registered together and then averaged to provide a static image. Then MINC tools software (http://bic-mni.github.io/) were used to linearly register T1-weighted MR to the PET image. Then for each image, the resulting MR along with the PET image is decomposed to $9 \times 9 \times 9$ overlapping patches with a gap equal to 2 voxels. The patches are first standardized and then fed to the trained network. The resulting patches are then composed to obtain the restored image and the uncertainty map. All images are also normalized by total activity to produce SUVR images.

5.1.5 Network Architecture and implementation details

The architecture of the network and each Inception-Res module is shown in Fig. 5.2. Since we have volumetric PET and MR data, we use 3D convolution layers in our network. In our network, Inception modules are stacked on the top of a three-layer stem. In addition to residual connections in each block, we also use residual learning, meaning that our labelled training data is the residual image (i.e., ground truth minus noisy patch).

All convolution layers in our network have padding and stride equal to 1 to retain the size of the patch throughout layers. In all hidden layers, ReLu was used as the activation function and the method of (He et al. 2015) was used for weight initialization. To reduce the number of parameters, $5 \times 5 \times 5$ convolutions in the Inception-Res module are factorized to two $3 \times 3 \times 3$ convolutions. The total number of trainable parameters for a network with 6 Inception modules is $\approx$ 1.8M. During training, in each epoch, for each image, 2/3 of the patches were randomly selected resulting in 60k 3D patches. The batch size was set to 50 patches. Adam with the learning rate equal to 0.0002 was used for training. In each setting, the model was run for 200 epochs and the epoch with the lowest loss on the validation set was used as the final model used on test and real data. Training for 200 epochs took 36 hours on an NVIDIA GTX TITAN X GPU. This architecture was implemented in Keras (Chollet 2015) with a Tensorflow backend.

5.1.6 Evaluation

To understand the effect of uncertainty estimation on performance of the network, an ablation study is performed by removing uncertainty estimation from the MR-guided model while keeping the rest of the network unchanged. The loss function for the model with aleatoric uncertainty is eq.5.1 and for the model without aleatoric uncertainty estimation is mean squared error.

The proposed framework is also compared to our recent guided kernel-means with twicing (GKM) method (Tahaei et al. 2019), an MR-guided PET image denoising method.

5.1.7 Figures of merit

In order to quantify the performance of the method in restoring the ground truth image in simulation data, peak signal to noise ratio (PSNR) and structural similarity index (SSIM) and normalized mean absolute error (NMAE) are used. PSNR is defined as:

$$PSNR = 20\ log_{10} \frac{max(\mathbf{x})}{\|\hat{\mathbf{x}} - \mathbf{x}\|_2} \tag{5.2}$$

Here, $\mathbf{x}$ is the ground truth image, $\hat{\mathbf{x}}$ is the noisy image and max($\mathbf{x}$) is the maximum intensity value in the ground truth image. The higher the PSNR the better the quality of the image. SSIM is obtained using the following formulation:

$$SSIM = \frac{(2\mu_{\mathbf{x}}\mu_{\hat{\mathbf{x}}} + c_1)(2\,\mathrm{cov}_{\mathbf{x}\hat{\mathbf{x}}} + c_2)}{(\mu_{\mathbf{x}}^2 + \mu_{\hat{\mathbf{x}}}{}^2 + c_1)(\sigma_{\mathbf{x}}{}^2 + \sigma_{\hat{\mathbf{x}}}{}^2 + c_2)} \tag{5.3}$$

where μ_x and $\mu_{\hat{x}}$ is the mean intensity in $\mathbf{x}$ and $\hat{\mathbf{x}}$ respectively. $\sigma_{\mathbf{x}}$ and $\sigma_{\hat{x}}$ are the standard deviation of voxels in $\mathbf{x}$ and $\hat{\mathbf{x}}$. $\mathrm{cov}_{x\hat{\mathbf{x}}}$ is the covariance of $\mathbf{x}$ and $\hat{\mathbf{x}}$. c_1 and c_2 are small constants to stabilize division by a small denominator. The higher the SSIM value, the better the image quality.

Also, the performance of the method in retaining PET specific information is evaluated using the MAE of lesion voxels after normalization by the ground truth:

$$\text{Lesions NMAE} = \frac{1}{N} \sum_{i=1}^{N} \frac{|\hat{x}_i - x_i|}{x_i} \tag{5.4}$$

N is the total number of lesion voxels in the image. We refer to this value as lesions normalized mean absolute error (NMAE). Lower values of Lesions NMAE are better.

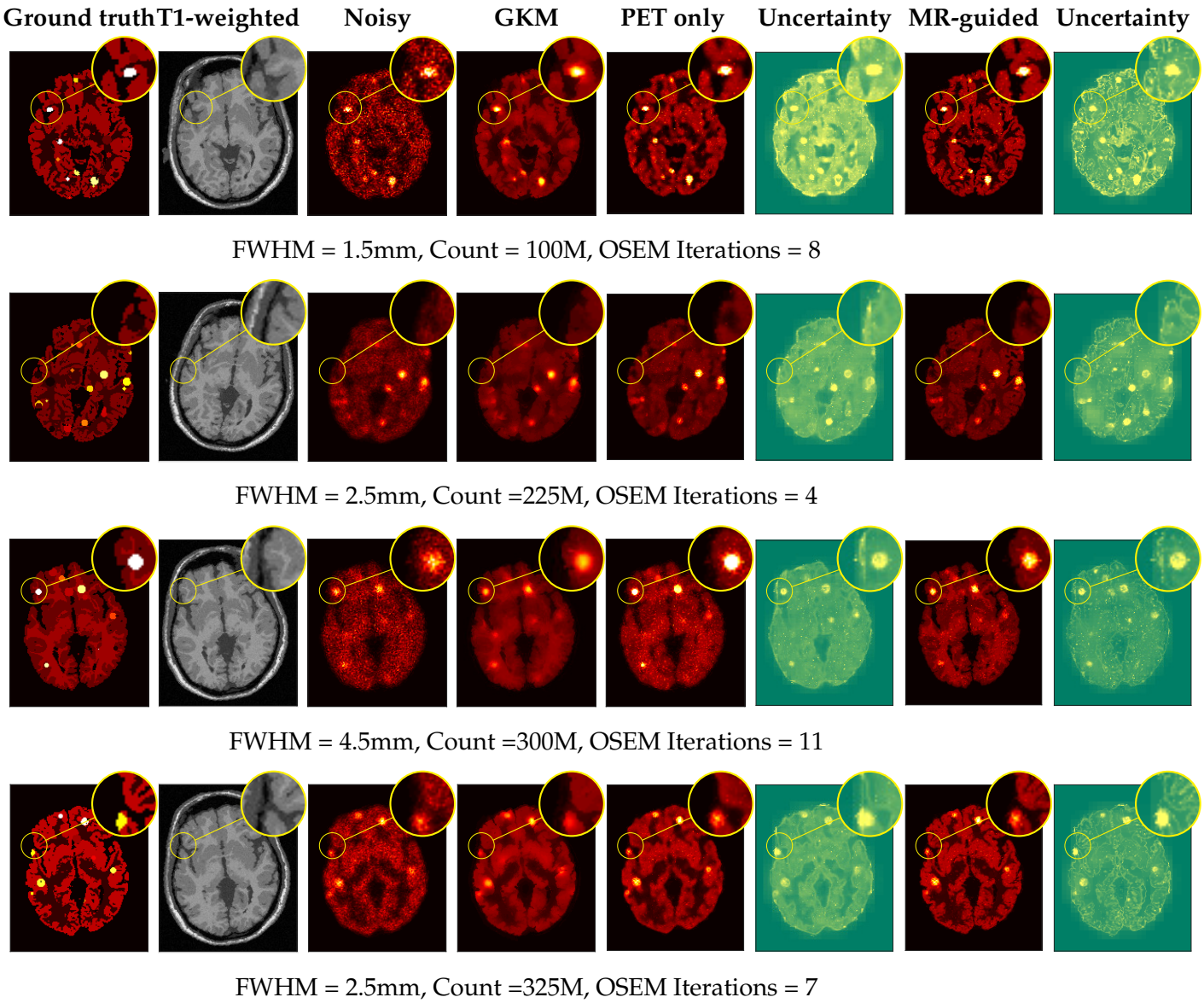


Figure 5.4: Axial view of four simulated test images from a single subject with different FWHM values, count levels and different number of OSEM iterations used for reconstruction. Each row shows one slice through an example volume. The characteristics of the image are shown at the bottom of each row. All PET images are normalized to same intensity scale for visualization and comparison. For better visualization a region of each image is magnified by a factor of 2.5. Both uncertainty maps for each sample are shown in the same scale.

5.2 Results

Figure 5.3 shows the trajectory of mean absolute loss of patches of the validation data during training. As one would expect, using T1-weighted MR as input reduces

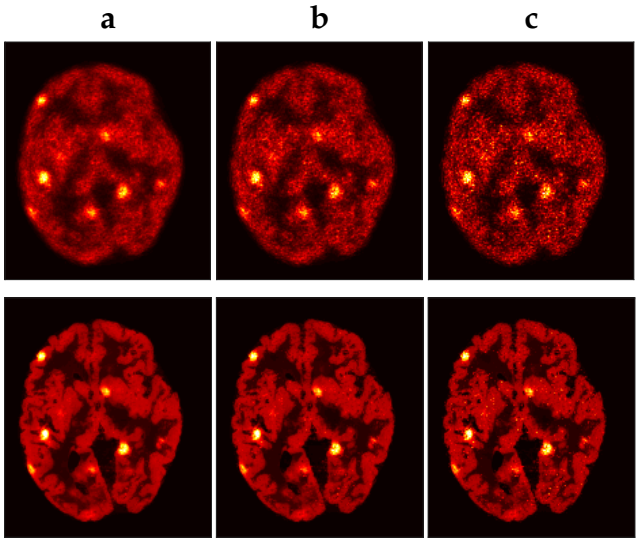


Figure 5.5: The figures shows a simulated test image with increasing iterations of OSEM used for reconstruction (3, 5, 9 OSEM iterations for columns a, b and c respectively). The top row shows the original images with no processing and the bottom figure shows the resulting images after applying the proposed method (MR-guided Inception-Res + Uncertainty)

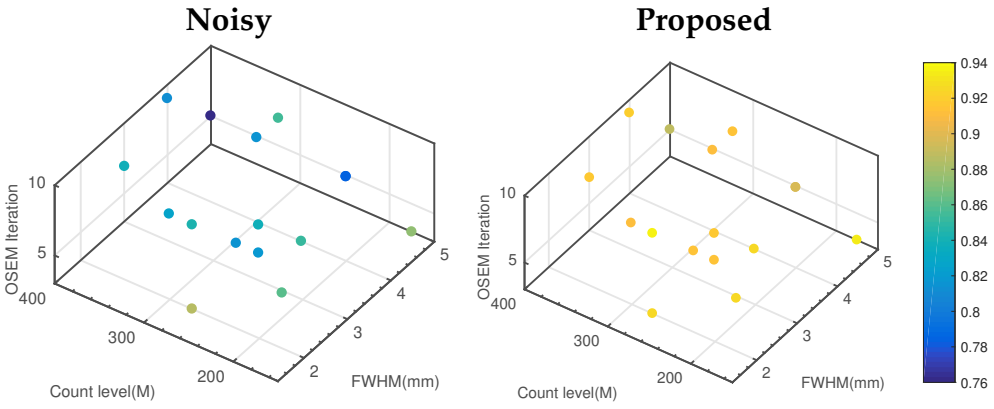


Figure 5.6: SSIM values for test images with no post processing (left figure) and after Inception-Res Net+ Uncertainty (right figure). This figure shows how the proposed method is able to improve image quality for different count levels, iteration numbers of OSEM and PVE levels.

the error on the validation set. After 150 epochs, the difference in MAE of the validation set with and without aleatoric uncertainty becomes very small.

Table 5.1 shows the mean of different figures of merit on test images after patch

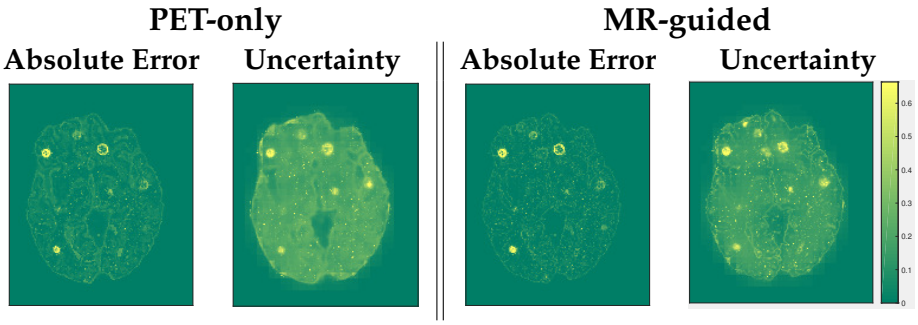


Figure 5.7: The uncertainty maps and the absolute error (Ground truth - restored image) for the proposed methods without and with T1-weighted MR as input.

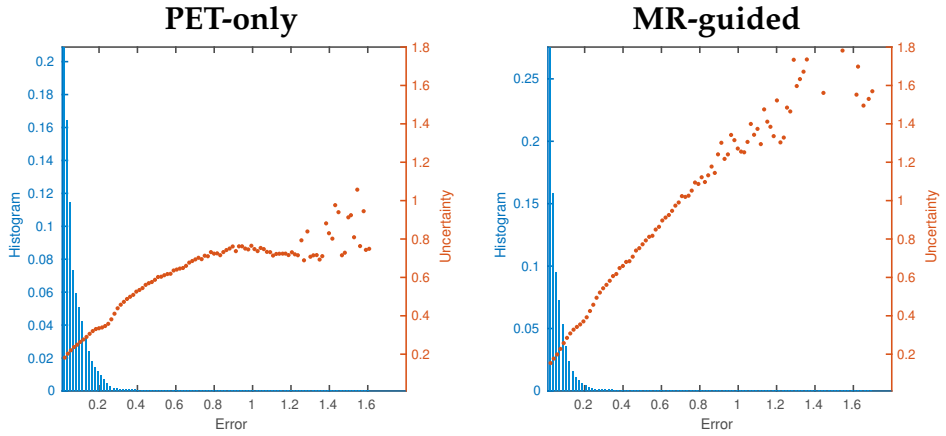


Figure 5.8: Relationship between uncertainty and error. In black is the histogram of error values for all voxels in the test data for PET-only (left) and MR-guided (right) methods. The mean uncertainty for each bin is shown in red. Notice the linear relationship between uncertainty and error value for the proposed MR-guided Inception net.

recomposition for the proposed method with and without MR-guided anatomical information.

Applying the proposed method on noisy and resolution degraded simulated PET data both without and with MR information considerably improved the image quality as quantified by PSNR (paired t-test, $p < 10^{-6}$) and SSIM (paired t-test, $p < 10^{-7}$). The PSNR and SSIM of MR-guided Inception-Res (with and without uncertainty estimation) are also significantly (paired t-test, $p < 10^{-4}$) higher than GKM.

Interestingly, by embedding PET-specific lesions in the training data, the proposed

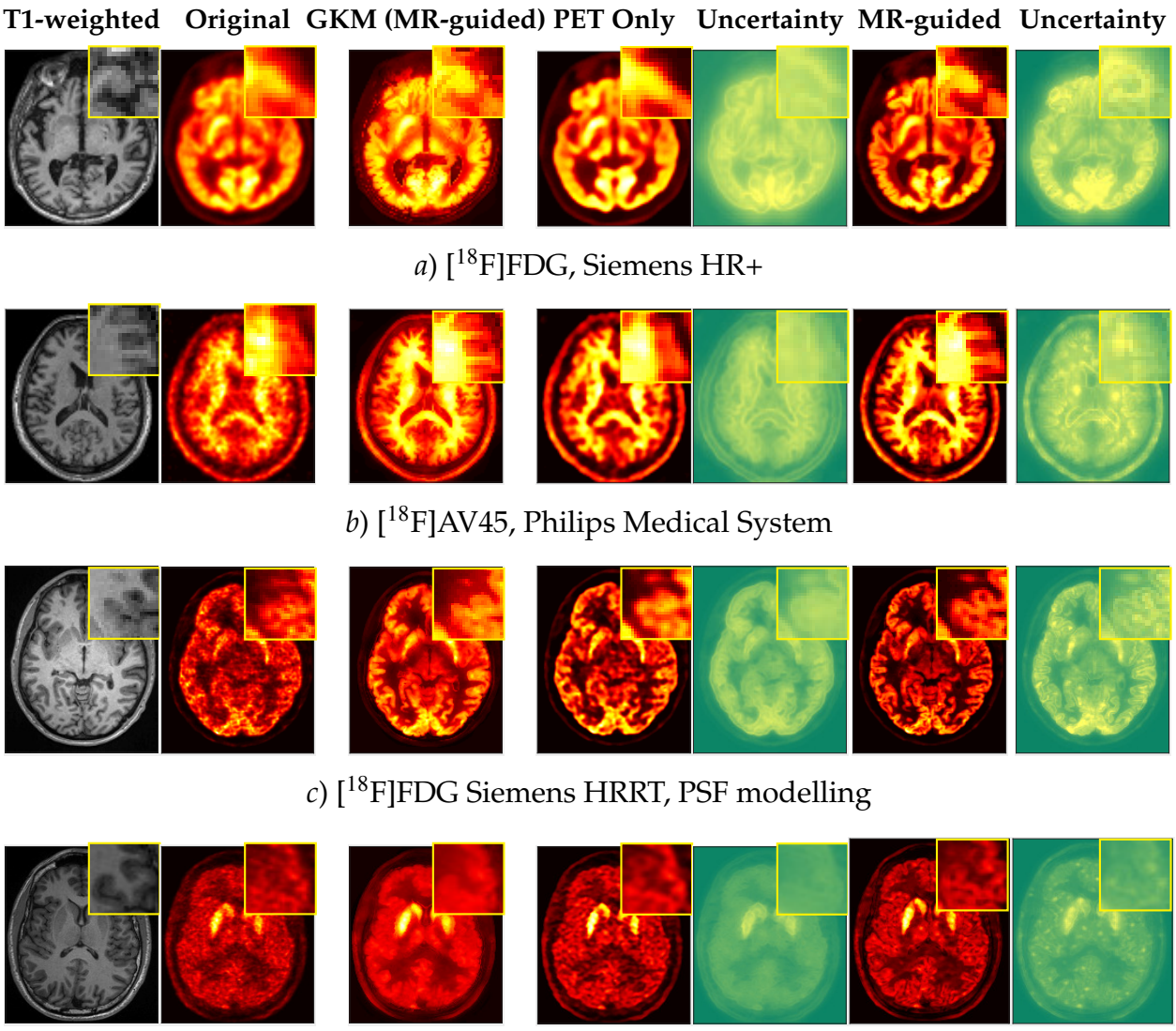


Figure 5.9: Four real PET images along with the subjects' registered and resampled T1-weighted MR images. Images *a* and *b* are from ADNI data set. They are both reconstructed using the OSEM algorithm. The exact number of iterations and the FWHM of the PSF function associated with these images are unknown. For images *c* and *d*, OSEM with resolution modelling with 16 subsets and 10 iterations was used to reconstruct the images. Uncertainty maps are normalized between 0 and 1.

methods learn to preserve PET specific information while improving the boundary information. This is reflected in the lesion NMAE value of the proposed methods being lower than that for degraded images with no-processing (paired t-test,both PET

only and MR-guided $p < 10^{-4}$) and after GKM (paired t-test, PET only $p < 10^{-5}$, MR-guided $p < 10^{-3}$).

There is no significant (paired t-test) change in performance of the proposed methods with and without aleatoric uncertainty as indicated by PSNR, SSIM or lesion NMAE values.

Figure 5.4 shows four images of random realizations of the test subject with different FWHM values, count levels and number of iterations of OSEM. The results of GKM and the proposed method with and without additional anatomical information, as well as their aleatoric uncertainty maps, are shown. In each image a small region is zoomed to better visualize the performance. Both proposed models improve image quality. While GKM tends to provide less noisy images, it does so at the expense of losing contrast in lesions.

Note that even without MR-guided anatomical information the proposed method considerably improves the boundary information of the noisy degraded image. In aleatoric uncertainty images, uncertainty tends to be higher on the boundaries and in the hot lesions.

In Fig. 5.5 a sample test volume (FWHM=2.5mm, count=225M) reconstructed using 3, 5 and 9 iterations of OSEM is shown before and after of applying MR-guided Inception network with uncertainty estimation. The proposed method results in images of similar quality for different number of iterations. This result suggest that the proposed model is almost invariant to the number of iterations used for reconstruction, at least in this example.

In Fig. 5.6, the SSIM value of 18 test images as a function of iteration number as well as count level and smoothing level are plotted before and after of applying MR-guided Inception with uncertainty estimation. This figure shows how applying the proposed method improves the quality of images with different characteristics.

Figure 5.7 shows, as an example, the aleatoric uncertainty of the methods with and

without T1-weighted MR image for the third row of Fig.5.4 along with the corresponding absolute error images. One can see that aleatoric uncertainty is higher in the model trained without MR information and that aleatoric uncertainty appears highly correlated with the error. Indeed, the correlation coefficient of error and uncertainty values of all voxels in test volumes was 0.80 and 0.81 for PET-only and MR-guided models respectively. Therefore, by learning uncertainty from training data, the network is able to predict where the error will be higher in the reconstructed images – an important point for images that may be interpreted clinically. To further investigate the relationship between uncertainty estimates and reconstruction error, we first discretized error values for all voxels in test data into 100 bins. We then measured the mean error for all voxels belonging to each bin. 5.8 plots the error histogram and the uncertainty estimates. In both models, for error values ranging from 0.0 to 0.8, uncertainty and error values seem to be highly correlated. However, for values above .08 (which only contribute to less than 0.06% of the voxels) the correlation is less. The small proportion of these values suggest that these are probably single (noise-like) voxels in the uncertainty map or the restored image.

5.2.0.1 Real data

Figure 5.9 shows the result of applying the algorithm to four real MR/PET datasets using different scanners and PET tracers. The first two rows show [^{18}F]FDG and florbetapir ([^{18}F]AV-45 images of two MCI subjects from the Alzheimer's Disease Neuroimaging Initiative (ADNI) (www.loni.ucla.edu),respectively. These images were scanned using Siemens and Philips scanners. The last two rows are images of [^{18}F]FDG and [^{11}C]raclopride acquired in the same centre using a Siemens HRRT scanner. One can see that Both GKM and the proposed MR-guided network benefit from anatomical information to provide a noticeable improvement in image quality especially in boundary regions. Nonetheless, the resulting images from GKM tend to have lower contrast than those of the proposed method. Also, we can see that even when no anatomical

information is provided (PET only network), the proposed model is able to improve the boundary information of the noisy degraded image. Note that even when a known PSF function is used to correct for PVE (the last two columns) by PSF modelling within reconstruction, image contrast and boundary information can be further improved by applying the proposed methods (both with and without the MR).

5.3 Discussion and Conclusion

In this work, an Inception-Res network was used to restore noisy and resolution degraded PET images. The network was trained using simulation data obtained by varying factors contributing to PVE and noise, with and without corresponding T1-weighted MR patches. The network was tested on both simulation data and real data from different scanners and tracers. Evaluation of the methods on simulation and real data showed that, even when no anatomical information is available, the trained model can considerably improve image quality. In many PET studies, subjects also undergo an MR acquisition and thus T1-weighted MR can be used to provide the network with detailed anatomical information. We showed that using the registered T1-weighted MR patches as another input to the network can further improve the quality, especially at the boundaries. By embedding many PET specific lesions in training data, our MR-guided model can learn to preserve and even improve the quality of lesion information in the restored image.

Patch decomposition plays an important role in the generalization of the proposed framework to variant structures .In fact, unlike many medical imaging applications such as classification, localization or segmentation where a global mapping of the image is required to find the solution, PVE correction and denoising using small neighborhoods is usually sufficient for correct quantification. Therefore, by patch decomposition, the network is less prone to overfitting to subject anatomies seen in the training data. The effect of using relatively small patches on the generalizability of the model is an area for future work.

In this work an isotropic Gaussian function was used to simulate PVE. While this may be fairly accurate for brain studies due to the small size of the brain with respect to the field of view (FOV), for whole body PET image restoration, using non-stationary smoothing would lead to more realistic simulations. Nonetheless, the patch-based and blind nature of the model enables it to recover patches independently. Hence, the current model should also work for images where the patches near the scanner ring have a lower resolution than those in the centre of FOV.

High correlation of aleatoric uncertainty with mean absolute error on test data suggest that it can be considered as a measure of the error in the quantification of a region of interest in real images. Since the target of our model was simulated *ground truth* patches, our uncertainty estimation reflects all of the various factors that contribute to error in the acquisition and reconstruction process.

Also this high correlation of uncertainty with the error may suggest that, as shown by (Kendall & Gal 2017), estimating only aleatoric uncertainty in a deep model compensates for the lack of implementing epistemic uncertainty.

There are some limitations to this method. First, we have only applied the method to brain images where a rigid registration can be used to align the PET and MR data. In other regions of the body, non-linear registration will be needed. However, we have shown that even without anatomical information from MR, the method works well.

Second, it is important to note that for PET imaging with count measurement data, the noise in the image is data-inherent with the higher noise level in higher activity regions. Therefore one should pay attention in analyzing uncertainty estimation in the resulting images. For example, a hot lesion will have higher uncertainty than a cold lesion due to the nature of count measurement data. To resolve this issue one may be encouraged to use the coefficient of variation (CV). For simulation studies, however, due to random intensity selection, some regions in the simulated image may have zero or near-zero values for which CV is either uncomputable or very unstable.

This may also extend to real data.

Overall, we proposed a CNN to blindly restore PET images and trained it on simulation data. The results show that the proposed model is able to generalize well to real images of different radiotracers with different noise and degradation characteristics. Moreover, our proposed framework provides uncertainty maps to help with visual inspection of potential errors in clinic and research. The use of uncertainty estimates in statistical analysis of PET datasets will be an interesting area for further work.

Acknowledgment

The authors would like to acknowledge Dr Caroline Paquette for providing the real [^{18}F]FDG data and the T1-weighted MR data and Dr A Dagher for providing the real [^{11}C]raclopride data and the T1-weighted MR data.

References

Ardekani, B. A., Braun, M., Hutton, B. F., Kanno, I. & Iida, H. (1996), 'Minimum cross-entropy reconstruction of PET images using prior anatomical information', *Physics in Medicine and Biology* **41**(11), 2497.

Aubert-Broche, B., Griffin, M., Pike, G. B., Evans, A. C. & Collins, D. L. (2006), 'Twenty new digital brain phantoms for creation of validation image data bases', *IEEE transactions on medical imaging* **25**(11), 1410–1416.

Ayers, G. & Dainty, J. C. (1988), 'Iterative blind deconvolution method and its applications', *Optics letters* **13**(7), 547–549.

Chan, C., Fulton, R., Feng, D. D. & Meikle, S. (2010), Median non-local means filtering for low snr image denoising: Application to PET with anatomical knowledge, *in* 'IEEE Nuclear Science Symposuim & Medical Imaging Conference', IEEE, pp. 3613–3618.

Chan, T. F. & Wong, C.-K. (1998), 'Total variation blind deconvolution', *IEEE transactions on Image Processing* **7**(3), 370–375.

Chen, K. T., Gong, E., de Carvalho Macruz, F. B., Xu, J., Boumis, A., Khalighi, M., Poston, K. L., Sha, S. J., Greicius, M. D., Mormino, E. et al. (2018), 'Ultra–low-dose 18f-florbetaben amyloid PET imaging using deep learning with multi-contrast MRI inputs', *Radiology* **290**(3), 649–656.

Chollet, F. (2015), 'keras', https://github.com/fchollet/keras.

Çiçek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T. & Ronneberger, O. (2016), 3D u-net: learning dense volumetric segmentation from sparse annotation, *in* 'International conference on medical image computing and computer-assisted intervention', Springer, pp. 424–432.

Dong, C., Loy, C. C., He, K. & Tang, X. (2015), 'Image super-resolution using deep convolutional networks', *IEEE transactions on pattern analysis and machine intelligence* **38**(2), 295–307.

Gondara, L. (2016), Medical image denoising using convolutional denoising autoencoders, *in* '2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW)', IEEE, pp. 241–246.

Gong, K., Guan, J., Kim, K., Zhang, X., Yang, J., Seo, Y., El Fakhri, G., Qi, J. & Li, Q. (2018), 'Iterative PET image reconstruction using convolutional neural network representation', *IEEE transactions on medical imaging* **38**(3), 675–685.

He, K., Zhang, X., Ren, S. & Sun, J. (2015), Delving deep into rectifiers: Surpassing human-level performance on imagenet classification, *in* 'Proceedings of the IEEE international conference on computer vision', pp. 1026–1034.

He, K., Zhang, X., Ren, S. & Sun, J. (2016), Deep residual learning for image recognition, *in* 'Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition', pp. 770–778.

Hudson, H. M. & Larkin, R. S. (1994), 'Accelerated image reconstruction using ordered subsets of projection data', *Medical Imaging, IEEE Transactions on* **13**(4), 601–609.

Jain, V. & Seung, S. (2009), Natural image denoising with convolutional networks, *in* 'Advances in neural information processing systems', pp. 769–776.

Johnson, J., Alahi, A. & Fei-Fei, L. (2016), Perceptual losses for real-time style transfer and super-resolution, *in* 'European conference on computer vision', Springer, pp. 694–711.

Kamnitsas, K., Ledig, C., Newcombe, V. F., Simpson, J. P., Kane, A. D., Menon, D. K., Rueckert, D. & Glocker, B. (2017), 'Efficient multi-scale 3D cnn with fully connected crf for accurate brain lesion segmentation', *Medical image analysis* **36**, 61–78.

Kaplan, S. & Zhu, Y.-M. (2018), 'Full-dose PET image estimation from low-dose PET image using deep learning: a pilot study', *Journal of digital imaging* pp. 1–6.

Kendall, A. & Gal, Y. (2017), What uncertainties do we need in bayesian deep learning for computer vision?, *in* 'Advances in neural information processing systems', pp. 5574–5584.

Levin, A., Weiss, Y., Durand, F. & Freeman, W. T. (2009), Understanding and evaluating blind deconvolution algorithms, *in* '2009 IEEE Conference on Computer Vision and Pattern Recognition', IEEE, pp. 1964–1971.

Mao, X., Shen, C. & Yang, Y.-B. (2016), Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections, *in* 'Advances in neural information processing systems', pp. 2802–2810.

Nimisha, T. M., Kumar Singh, A. & Rajagopalan, A. N. (2017), Blur-invariant deep learning for blind-deblurring, *in* 'Proceedings of the IEEE International Conference on Computer Vision', pp. 4752–4760.

Nuyts, J., Fessler, J. et al. (2003), 'A penalized-likelihood image reconstruction method for emission tomography, compared to postsmoothed maximum-likelihood with matched spatial resolution', *Medical Imaging, IEEE Transactions on* **22**(9), 1042–1052.

Rangarajan, A., Hsiao, T. & Gindi, G. (2000), 'A bayesian joint mixture framework for the integration of anatomical information in functional image reconstruction', *Journal of Mathematical Imaging and Vision* **12**(3), 199–217.

Ronneberger, O., Fischer, P. & Brox, T. (2015), U-net: Convolutional networks for biomedical image segmentation, *in* 'International Conference on Medical Image Computing and Computer-Assisted Intervention', Springer, pp. 234–241.

Schmand, M., Eriksson, L., Casey, M., Andreaco, M., Melcher, C., Wienhard, K., Flugge, G. & Nutt, R. (1998), 'Performance results of a new doi detector block for a high resolution pet-lso research tomograph hrrt', *IEEE Transactions on Nuclear Science* **45**(6), 3000–3006.

Srivastava, R. K., Greff, K. & Schmidhuber, J. (2015), Training very deep networks, *in* 'Advances in neural information processing systems', pp. 2377–2385.

Szegedy, C., Ioffe, S., Vanhoucke, V. & Alemi, A. A. (2017), Inception-v4, inception-resnet and the impact of residual connections on learning., *in* 'AAAI', Vol. 4, p. 12.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V. & Rabinovich, A. (2015), Going deeper with convolutions, *in* 'Proceedings of the IEEE conference on computer vision and pattern recognition', pp. 1–9.

Tahaei, M. S. & Reader, A. J. (2016), 'Patch-based image reconstruction for PET using prior-image derived dictionaries', *Physics in Medicine and Biology* **61**(18), 6833.

Tahaei, M. S., Reader, A. J. & Collins, D. L. (2019), 'Two novel PET image restoration methods guided by PET-MR kernels: Application to brain imaging', *Medical physics* **46**(5), 2085–2102.

Turkheimer, F. E., Boussion, N., Anderson, A. N., Pavese, N., Piccini, P. & Visvikis, D.

(2008), 'PET image denoising using a synergistic multiresolution analysis of structural (MRI /CT) and functional datasets', *Journal of Nuclear Medicine* **49**(4), 657–666.

Wang, Y., Yu, B., Wang, L., Zu, C., Lalush, D. S., Lin, W., Wu, X., Zhou, J., Shen, D. & Zhou, L. (2018), '3D conditional generative adversarial networks for high-quality PET image estimation at low dose', *NeuroImage* **174**, 550–562.

Xiang, L., Qiao, Y., Nie, D., An, L., Lin, W., Wang, Q. & Shen, D. (2017), 'Deep auto-context convolutional neural networks for standard-dose PET image estimation from low-dose PET/MRI', *Neurocomputing* **267**, 406–416.

Xu, J., Gong, E., Pauly, J. & Zaharchuk, G. (2017), '200x low-dose PET reconstruction using deep learning', *arXiv preprint arXiv:1712.04119* .

Yang, G., Yu, S., Dong, H., Slabaugh, G., Dragotti, P. L., Ye, X., Liu, F., Arridge, S., Keegan, J., Guo, Y. et al. (2017), 'Dagan: deep de-aliasing generative adversarial networks for fast compressed sensing MRI reconstruction', *IEEE transactions on medical imaging* **37**(6), 1310–1321.

Yang, Q., Yan, P., Zhang, Y., Yu, H., Shi, Y., Mou, X., Kalra, M. K., Zhang, Y., Sun, L. & Wang, G. (2018), 'Low-dose CT image denoising using a generative adversarial network with wasserstein distance and perceptual loss', *IEEE transactions on medical imaging* **37**(6), 1348–1357.

Zaidi, H., Montandon, M.-L. & Alavi, A. (2007), 'Advances in attenuation correction techniques in PET', *PET clinics* **2**(2), 191–217.

Zhang, K., Zuo, W., Chen, Y., Meng, D. & Zhang, L. (2017), 'Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising', *IEEE Transactions on Image Processing* **26**(7), 3142–3155.

Zhang, K., Zuo, W., Gu, S. & Zhang, L. (2017a), Learning deep cnn denoiser prior for image restoration, *in* 'Proceedings of the IEEE conference on computer vision and pattern recognition', pp. 3929–3938.

Zhang, K., Zuo, W. & Zhang, L. (2018), 'Ffdnet: Toward a fast and flexible solution for cnn-based image denoising', *IEEE Transactions on Image Processing* **27**(9), 4608–4622.

Zhu, B., Liu, J. Z., Cauley, S. F., Rosen, B. R. & Rosen, M. S. (2018), 'Image reconstruction by domain-transform manifold learning', *Nature* **555**(7697), 487.

Chapter 6

Discussion, Conclusion and Future Work

Throughout the last two decades, PET imaging has been widely used in clinical practice and research as a tool for understanding and diagnosis of different brain disorders such as Alzheimer's disease, Parkinson's disease and schizophrenia (Heiss 2009, Drzezga et al. 2014, Levivier et al. 2004, Klunk et al. 2004). PET images have high noise and poor spatial resolution when compared to MRI or CT. The availability of T1-weighted MR images in many PET studies, has motivated many researchers to develop methods that use the structural information from MR images to improve the quality of PET images.

This dissertation focused on the development of new methods that use a subject's MR image to improve quantification in PET images by reducing noise and improving spatial resolution. Two distinct approaches have been considered in this regard: regularization within reconstruction and restoration after reconstruction. The following sections (Section 6.1 and 6.2) provide a summary of the proposed within and after reconstruction methods. This is followed by a discussion about different aspects of the proposed methods in sections 6.3 to 6.5. Section 6.6 outlines the future directions of

this research and the dissertation in concluded in Section 6.7.

6.1 Regularization within reconstruction

The advantage of regularization within reconstruction is that it correctly accounts for the Poisson noise in the count measurement data. Hence, in cases where we have access to the raw measurement data, denoising within reconstruction may reduce the bias and variance of the resulting noisy images, in comparison to denoising after reconstruction, at which point the noise properties of the image are hard to model accurately.

In light of the above argument, Chapter 3 focuses on the development of a new reconstruction method that leverages information from anatomical images through a re-parameterization framework.

Re-parameterization can be considered as a regularization approach for image reconstruction. Instead of adding a penalty term to the objective function to penalize images of undesired characteristics, in re-parameterization, the object of interest (e.g. the radiotracer distribution) is represented by a reduced and/or constrained set of parameters, to deliver estimates with desirable characteristics.

From this perspective, I proposed to use a patch-based re-parameterization framework for PET reconstruction: dictionaries are learned from patches of a prior image and the learned dictionary atoms are then used as basis functions to represent the reconstructed image. The coefficients for these basis function are then estimated from PET measurement data.

In addition, I proposed a method for maximizing the Poisson log-likelihood penalized by an l_1 norm. By adding an l_1 norm to our objective function we impose sparsity on the learned parameters of interest. I combined this latter method with the aforementioned re-parameterized method to find sparse coefficients for patch-based dictionary atoms learned from a prior image within reconstruction. We extensively compared the method with other MAPEM regularization methods and showed that

the proposed methods can provide better images in terms of a variety of figures of merit.

6.2 Restoration of reconstructed images

Within-reconstruction regularization methods cannot be easily used in many research and clinical settings since PET measurement data format is usually proprietary and not available. Therefore, post-reconstruction methods with applicability to reconstructed images without raw data are of great interest. Chapters 4 and 5 present novel post-reconstruction PET image restoration methods that use complementary anatomical information from corresponding MRI data.

6.2.1 Two novel PET image restoration methods guided by PET-MR kernels: Application to brain imaging

Two novel methods for restoring reconstructed PET images were proposed. In the first proposed method, normalized kernel matrices obtained from the median filtered PET image and the registered MR image were used as basis functions to represent the restored PET image. I proposed an iterative algorithm for estimating the coefficients of these basis functions while imposing sparsity on these coefficients. This method is called guided kernel re-parametrization.

In the second method, first the registered MR image is used for guided kernel filtering where the kernels are extracted from MR image. The residual image is then obtained and filtered using the median filtered PET image as the guide image. The filtered residual is then added to the filtered image from the first step. This method is called guided kernel means (GKM) with twicing.

The two proposed methods were applied to PET images after applying a conventional de-blurring algorithm. Both methods were evaluated on simulation data using different figures of merit. Both proposed methods showed considerable improvements in image quality compared to state of the art denoising methods. Among the

proposed methods, the GKM with twicing can be easily implemented, is very fast and has very low memory and processing requirements.

6.2.2 Blind PET Image Restoration: A 3D CNN with Uncertainty Estimation

Chapter 5 shows the potential of deep learning architectures to tackle the challenging problem of blind restoration. In this context, a deep CNN was proposed for restoring PET images. For training this CNN, a pipeline was proposed that simulated 3D noisy PET patches from segmented MR images. In this pipeline, PET images with different characteristics were simulated using randomly chosen levels of radioactivity distribution, acquisition blurring and noise as well as different numbers of iterations of OSEM. Two separate models were trained, one with and one without registered MR images as input. An Inception-Res architecture was used to restore 3D PET image patches.

Also a map of voxel level aleatoric uncertainty was learned in the last layer of the network. This uncertainty map can be used in both research and the clinic as a tool for predicting erroneous regions. The above models, evaluated using test simulation data, show significant improvement in image quality compared to the GKM with twicing method when no information about the PSF is available. Also, the application of the method on real data showed the ability of the method to generalize to unseen scanners and radiotracers.

6.3 From specificity to generalizability

The evolution of the methods in this dissertation can be viewed as a journey from specificity to generalizability. The set of parameters for the reconstruction method proposed in Chapter 3 are specific to the radiotracer type and dose, scanner sensitivity and resolutions. There is no guarantee that the same set of parameters will work in different settings. Also, the methods are only applicable to images for which raw

measurement data is available. Therefore such reconstruction methods can only be utilized either by scanner vendors to provide reconstruction software for their scanners or for the few research-dedicated scanners for which the measurement data type is not proprietary to the scanner.

In Chapter 4, the proposed methods are designed to be applied on images reconstructed using conventional algorithms and thus eliminate the need to access the raw measurement data. We show through simulation that the optimal set of parameters for the two proposed post-reconstruction methods are not sensitive to noise level. But a correct estimation of the PSF function is needed for the methods to work well.

In Chapter 5, through learning from many images with different characteristics, a method that is generalizable to different radiotracer types, doses and resolutions is proposed. This is only achieved at the expense of a long training with many simulated images. Once trained, however, the inference is fast. We also show that through learning from many images, the network can perform very well even without anatomical information.

6.4 No segmentation is required

All the proposed methods in Chapters 4 and 5 can be regarded as both denoising and PVE correction. But unlike the majority of the MR-based PVE correction methods in the literature where segmentation of the MR volume is needed prior to PVE correction, the proposed frameworks use registered T1-weighted MR images without pre-processing. This is an advantage, as segmentation errors are known to be an important source of error in MR-guided PVE corrections (Frouin et al. 2002).

6.5 Bias towards anatomy

PET imaging tries to represent the physiological and pathophysiological process in the body at a molecular level e.g. metabolism, enzyme activity, protein accumulation, and receptor binding. T1-weighted MR on the other hand has an excellence contrast

for brain tissue therefore providing complementary information to PET (Lasocki & Hicks 2019).

There are a few challenges in incorporating anatomical information for PET image enhancement.One challenge is that using anatomical information for PET image restoration leads to loss of resolution in regions where PET and MR are in disagreement. This can be In this dissertation, for each method, different strategies are taken to mitigate this issue. The ability of the proposed methods in retaining PET-specific information are carefully evaluated by embedding PET-specific information in simulation data and measuring the amount of error and loss of contrast in these regions.

In Chapter 3, adding an all-one vector to the learned dictionary from the MR image allows for PET-specific information to be reconstructed. In Chapter 5, both proposed methods try to preserve PET-specific information by incorporating two components for re-parameterization: median filtered PET and registered MR . Therefore the method used the median filtered PET image to represent PET-specific signals. In Chapter 5, by embedding many PET-specific lesions (500) in each simulated brain volume we ensure that the model learns to maintain PET-specific information from training data.

With the advances in imaging technologies and the emergence of hybrid PET-MR scanners, there is an increasing opportunity to perform simultaneous PET and MR imaging. With a simultaneous acquisition, PET and MR images may be better aligned, making the registration of T1-weighted MR to PET more accurate. However, as of today the majority of PET-MR images come from separate scanners and hence registration is a very important first step before using MR images for correction of PET images. For brain studies which are the focus of this dissertation, a rigid registration with only 6 parameters is usually adequate to align images from the two modalities. However, although a rigid registration may seem an easy task there is no guarantee that the resulting images will be perfectly aligned. In fact rigid registration of specific

radiotracers with very localized distributions such as [^{11}C]raclopride can be challenging.

All MR-guided methods are based on the assumption that the function and anatomy are consistent in most regions and hence the proposed methods will work best if this assumption holds. But what if there is a small mis-alignment between PET and MR images? A small mis-alignment between PET and MR can be analogous to inconsistency between PET and MR images in many regions. As discussed above, all three proposed methods allow for PET specific signals to be retained in the restored PET image and hence the resulting images will not have a huge bias or artefact due to mis-registration. This was shown and discussed in detail for the proposed methods in Chapter 4.

6.6 Importance of PET restoration for quantification in research

6.6.1 Sample size

In order to be able to discover the difference between two conditions, the effect size has to be greater than the signal to noise ratio of the parameter of interest in the population. This is especially important in PET since it is an expensive modality and the size of the dataset in each group is usually very small. Therefore, reducing the noise in PET images can decrease the number of samples required for testing a specific hypothesis.

6.6.2 Disease prediction

Prediction of brain disorders in early stages plays an important role in improving treatment choices as well as reducing the cost of testing newly developed drugs by eliminating false positives. Today, with the availability of large datasets such as Alzheimer's Disease Neuroimaging Initiative (ADNI), there is a great potential in in-

ferring features from multi-parametric data to improve diagnosis and treatment of dementia (Jack Jr et al. 2013). However, prior to feature extraction for disease prediction using PET imaging data, reducing the noise and improving the resolution of these images is necessary as it can affect the prediction accuracy.

6.6.3 Statistical analysis

In voxel-wise analysis, and ROI analysis for small regions, reducing the noise without sacrificing the resolution is an important step before performing the statistical tests.

Many of the MR-guided PVE correction methods assume homogeneous activity in each anatomical segment and hence applying them to PET images before voxel-wise analysis may lead to loss of information in small regions. The proposed methods in this dissertation mitigate this problem by reducing the noise and improving the resolution without relying on this strong homogeneity assumption and hence can be utilized prior to voxel-wise analysis.

6.7 Future work

6.7.1 Minimal smoothing when voxels matter

In voxel-wise analysis PET images from different subjects are all registered to a common template space by using the transformation obtained from registering their corresponding MR images to the template space.

Assuming that images are aligned, then a voxel-wise statistical test is performed on the images from the two groups and the resulting significance values are corrected for multiple comparisons. However, prior to statistical testing, often a heavy Gaussian smoothing is applied to each image. The following lists the main reasons for smoothing:

- to reduce the effect of anatomical variability between subjects

- to improve the signal to noise ratio in PET images
- to make voxel values more consistent with the central-limit theorem which is the implicit assumption on which common statistical tests and corrections for multiple comparisons are usually based (Mikl et al. 2008, Friston et al. 1995, 1996)

In brain studies often a Gaussian with a FWHM of about 8mm is used to smooth PET images. This heavy smoothing may lead to loss of information in small regions. This in turn can limit the power of PET imaging to find pathological differences specific to very small regions. In this dissertation we developed methods that, in contrast to Gaussian smoothing, can reduce the noise while improving the spatial resolution.

Also, the high computational power available today, has made the application of permutation tests possible. Permutation statistical tests, as opposed to parametric ones, are not based on the central limit theorem and hence do not require pre-smoothing.

Therefore combining the proposed methods in this dissertation with the state-of-the-art non-linear registration methods with small between-subject variability may reduce the amount of smoothing needed and open new opportunities for exploratory studies in very small regions.

6.7.2 Re-using the proposed frameworks for other modalities

All the proposed frameworks in this dissertation have the potential to be used for restoring a lower resolution, noisy modality using a higher resolution and less noisy modality of the same subject as the guide image. Some examples of such potential include restoring fMRI using T1-weighted images and SPECT using CT.

6.7.3 Multi-modality guided restoration

Also all the proposed frameworks can be easily extended to include multiple guide images. For example, we can use both T1-weighted and T2-weighted images of the subject to guide the restoration process.

In the proposed method in Chapter 3, this can be done by learning dictionaries from patches extracted from different modalities (one dictionary per modality) and using the learned dictionaries in combination to re-parametrize the PET image. For methods in Chapter 4, a single kernel can be obtained as a function of multiple modalities and be used to represent the image. In Chapter 5, multiple modalities along the noisy PET image can be fed to the network as input channels.

6.7.4 A deep learning model with missing guide images

In Chapter 5, in the proposed deep learning framework, we proposed two models, one with MR and one without MR. A *single* model that can restore images both with and without MR could be interesting: If an MR image is available for a subject the model uses that to improve the restoration, and if not, it restores the PET image probably with less refined boundaries. This can be done by considering PET images without MR as incomplete data and imputing the missing part of data (MR image) either explicitly through training another network or implicitly within one framework.

6.8 Conclusion

High noise and low spatial resolution in PET images may lead to incorrect quantification of radioactivity concentrations. This in turn can lead to wrong statistical analysis in research and even misdiagnosis in the clinic.

In this dissertation three distinct approaches that use anatomically guided information to improve the quality of PET images were proposed. The main challenge in devising automatically guided methods is to retain PET specific information in regions where the function and anatomy are in disagreement. As discussed, all the proposed methods carefully try to address this problem through different approaches.

Incorporating the proposed methods in image processing pipelines can lead to more accurate quantification especially in small regions.

Bibliography

Aharon, M., Elad, M., Bruckstein, A. et al. (2006), 'K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation', *IEEE Transactions on signal processing* **54**(11), 4311.

Aksoy, S. & Haralick, R. M. (2001), 'Feature normalization and likelihood-based similarity measures for image retrieval', *Pattern recognition letters* **22**(5), 563–582.

Bailey, D. L. (1996), Quantification in 3D positron emission tomography, PhD thesis, University of Surrey.

Bengio, Y. (2009), 'Learning deep architectures for ai', *Foundations and trends® in Machine Learning* **2**(1), 1–127.

Bengio, Y., Lamblin, P., Popovici, D. & Larochelle, H. (2007), 'Greedy layer-wise training of deep networks', *Advances in neural information processing systems* **19**, 153.

Boureau, Y.-L., Bach, F., LeCun, Y. & Ponce, J. (2010), Learning mid-level features for recognition, *in* 'Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on', IEEE, pp. 2559–2566.

Bourlard, H. & Kamp, Y. (1988), 'Auto-association by multilayer perceptrons and singular value decomposition', *Biological cybernetics* **59**(4-5), 291–294.

Brasse, D., Kinahan, P. E., Lartizien, C., Comtat, C., Casey, M. & Michel, C. (2005),

'Correction methods for random coincidences in fully 3D whole-body pet: impact on data and image quality', *Journal of nuclear medicine* **46**(5), 859–867.

Catana, C., Benner, T., van der Kouwe, A., Byars, L., Hamm, M., Chonde, D. B., Michel, C. J., El Fakhri, G., Schmand, M. & Sorensen, A. G. (2011), 'MRI-assisted PET motion correction for neurologic studies in an integrated MR-PET scanner', *Journal of Nuclear Medicine* **52**(1), 154–161.

Chan, C., Fulton, R., Feng, D. D. & Meikle, S. (2010), Median non-local means filtering for low snr image denoising: Application to PET with anatomical knowledge, *in* 'IEEE Nuclear Science Symposuim & Medical Imaging Conference', IEEE, pp. 3613–3618.

Collobert, R. & Weston, J. (2008), A unified architecture for natural language processing: Deep neural networks with multitask learning, *in* 'Proceedings of the 25th international conference on Machine learning', ACM, pp. 160–167.

De Pierro, A. R. (1995), 'A modified expectation maximization algorithm for penalized likelihood estimation in emission tomography.', *IEEE Transactions on Medical Imaging* **14**(1), 132.

Defrise, M., Townsend, D., Bailey, D., Geissbuhler, A. & Jones, T. (1991), 'A normalization technique for 3D PET data', *Physics in Medicine & Biology* **36**(7), 939.

Dekemp, R. A. & Nahmias, C. (1994), 'Attenuation correction in PET using single photon transmission measurement', *Medical physics* **21**(6), 771–778.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K. & Fei-Fei, L. (2009), ImageNet: A Large-Scale Hierarchical Image Database, *in* 'CVPR09'.

Drzezga, A., Barthel, H., Minoshima, S. & Sabri, O. (2014), 'Potential clinical applica-

tions of PET/MR imaging in neurodegenerative diseases', *Journal of Nuclear Medicine* **55**(Supplement 2), 47S–55S.

Erlandsson, K., Buvat, I., Pretorius, P. H., Thomas, B. A. & Hutton, B. F. (2012), 'A review of partial volume correction techniques for emission tomography and their applications in neurology, cardiology and oncology', *Physics in Medicine & Biology* **57**(21), R119.

Fahey, F. H. (2002), 'Data acquisition in PET imaging', *Journal of nuclear medicine technology* **30**(2), 39–49.

Friston, K. J., Holmes, A. P., Poline, J., Grasby, P., Williams, S., Frackowiak, R. S. & Turner, R. (1995), 'Analysis of fMRI time-series revisited', *Neuroimage* **2**(1), 45–53.

Friston, K. J., Holmes, A., Poline, J.-B., Price, C. J. & Frith, C. D. (1996), 'Detecting activations in PET and fMRI: levels of inference and power', *Neuroimage* **4**(3), 223–235.

Frouin, V., Comtat, C., Reilhac, A. & Grégoire, M.-C. (2002), 'Correction of partial-volume effect for PET striatal imaging: fast implementation and study of robustness', *Journal of Nuclear Medicine* **43**(12), 1715–1726.

Havaei, M., Davy, A., Warde-Farley, D., Biard, A., Courville, A., Bengio, Y., Pal, C., Jodoin, P.-M. & Larochelle, H. (2017), 'Brain tumor segmentation with deep neural networks', *Medical image analysis* **35**, 18–31.

He, K., Zhang, X., Ren, S. & Sun, J. (2016), Deep residual learning for image recognition, *in* 'Proceedings of the IEEE conference on computer vision and pattern recognition', pp. 770–778.

Heiss, W.-D. (2009), 'The potential of PET/MR for brain imaging', *European journal of nuclear medicine and molecular imaging* **36**(1), 105–112.

Hess, S., Blomberg, B. A., Zhu, H. J., Høilund-Carlsen, P. F. & Alavi, A. (2014), 'The pivotal role of fdg-pet/ct in modern medicine', *Academic radiology* **21**(2), 232–249.

Hinton, G. E., Osindero, S. & Teh, Y.-W. (2006), 'A fast learning algorithm for deep belief nets', *Neural computation* **18**(7), 1527–1554.

Hofmann, M., Pichler, B., Schölkopf, B. & Beyer, T. (2009), 'Towards quantitative PET/MRi: a review of MR -based attenuation correction techniques', *European journal of nuclear medicine and molecular imaging* **36**(1), 93–104.

Hutton, B. F. & Baccarne, V. (1998), 'Efficient scatter modelling for incorporation in maximum likelihood reconstruction', *European journal of nuclear medicine* **25**(12), 1658–1665.

Ioffe, S. & Szegedy, C. (2015), 'Batch normalization: Accelerating deep network training by reducing internal covariate shift', *arXiv preprint arXiv:1502.03167* .

Jack Jr, C. R., Knopman, D. S., Jagust, W. J., Petersen, R. C., Weiner, M. W., Aisen, P. S., Shaw, L. M., Vemuri, P., Wiste, H. J., Weigand, S. D. et al. (2013), 'Tracking pathophysiological processes in alzheimer's disease: an updated hypothetical model of dynamic biomarkers', *The Lancet Neurology* **12**(2), 207–216.

Jacobs, F., Matej, S. & Lewitt, R. (1998), 'Image reconstruction techniques for pet', *Dep. Radiol., Univ. Pennsylvania, Philadelphia, Tech. Rep. MIPG245* .

Jiao, J., Markiewicz, P., Burgos, N., Atkinson, D., Hutton, B., Arridge, S. & Ourselin, S. (2015), Detail-preserving PET reconstruction with sparse image representation and anatomical priors, *in* 'Information Processing in Medical Imaging', Springer, pp. 540–551.

Karush, W. (1939), 'Minima of functions of several variables with inequalities as side constraints', *M. Sc. Dissertation. Dept. of Mathematics, Univ. of Chicago* .

Klunk, W. E., Engler, H., Nordberg, A., Wang, Y., Blomqvist, G., Holt, D. P., Bergström, M., Savitcheva, I., Huang, G.-F., Estrada, S. et al. (2004), 'Imaging brain amyloid in alzheimer's disease with pittsburgh compound-b', *Annals of neurology* **55**(3), 306–319.

Kolthammer, J. A. (2013), Quantitative Positron Emission Tomography for Estimation of Absolute Myocardial Blood Flow, PhD thesis, Case Western Reserve University.

Krizhevsky, A., Sutskever, I. & Hinton, G. E. (2012), Imagenet classification with deep convolutional neural networks, *in* 'Advances in neural information processing systems', pp. 1097–1105.

Krogh, A. & Hertz, J. A. (1992), A simple weight decay can improve generalization, *in* 'Advances in neural information processing systems', pp. 950–957.

Kuhn, H. W. & Tucker, A. W. (1951), 'Nonlinear programming, in (j. neyman, ed.) proceedings of the second berkeley symposium on mathematical statistics and probability'.

Langs, G., Hanbury, A., Menze, B. & Müller, H. (2013), Visceral: Towards large data in medical imaging—challenges and directions, *in* 'Medical Content-Based Retrieval for Clinical Decision Support', Springer, pp. 92–98.

Lasocki, A. & Hicks, R. J. (2019), 'How we read: the combined use of mri and novel pet tracers for the characterisation and treatment planning of masses in neuro-oncology', *Cancer Imaging* **19**(1), 57.

LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W. & Jackel, L. D. (1989), 'Backpropagation applied to handwritten zip code recognition', *Neural computation* **1**(4), 541–551.

Lercher, M. J. & Wienhard, K. (1994), 'Scatter correction in 3-d pet', *IEEE transactions on medical imaging* **13**(4), 649–657.

Levivier, M., Massager, N., Wikler, D., Lorenzoni, J., Ruiz, S., Devriendt, D., David, P., Desmedt, F., Simon, S., Van Houtte, P. et al. (2004), 'Use of stereotactic PET images in dosimetry planning of radiosurgery for brain tumors: clinical experience and proposed classification', *Journal of Nuclear Medicine* **45**(7), 1146–1154.

Lewitt, R. M. (1990), 'Multidimensional digital image representations using generalized kaiser–bessel window functions', *JOSA A* **7**(10), 1834–1846.

Lewitt, R. M. (1992), 'Alternatives to voxels for image representation in iterative reconstruction algorithms', *Physics in Medicine and Biology* **37**(3), 705.

Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., Van Der Laak, J. A., Van Ginneken, B. & Sánchez, C. I. (2017), 'A survey on deep learning in medical image analysis', *Medical image analysis* **42**, 60–88.

Lucy, L. B. (1974), 'An iterative technique for the rectification of observed distributions', *The astronomical journal* **79**, 745.

Maas, A. L., Hannun, A. Y. & Ng, A. Y. (2013), Rectifier nonlinearities improve neural network acoustic models, *in* 'Proc. icml', Vol. 30, p. 3.

Matthews, J., Angelis, G., Kotasidis, F., Markiewicz, P. & Reader, A. (2010), Direct reconstruction of parametric images using any spatiotemporal 4d image based model and maximum likelihood expectation maximisation, *in* 'Nuclear Science Symposium Conference Record (NSS/MIC), 2010 IEEE', IEEE, pp. 2435–2441.

Mersereau, R. M. & Oppenheim, A. V. (1974), 'Digital reconstruction of multidimensional signals from their projections', *Proceedings of the IEEE* **62**(10), 1319–1338.

Mikl, M., Mareček, R., Hluštík, P., Pavlicová, M., Drastich, A., Chlebus, P., Brázdil, M. & Krupa, P. (2008), 'Effects of spatial smoothing on fMRI group inferences', *Magnetic resonance imaging* **26**(4), 490–503.

Moses, W. W. (2011), 'Fundamental limits of spatial resolution in pet', *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **648**, S236–S240.

Müller-Gärtner, H. W., Links, J. M., Prince, J. L., Bryan, R. N., McVeigh, E., Leal, J. P., Davatzikos, C. & Frost, J. J. (1992), 'Measurement of radiotracer concentration in brain gray matter using positron emission tomography: MRI-based correction for partial volume effects', *Journal of Cerebral Blood Flow & Metabolism* **12**(4), 571–583.

Rahmim, A., Cheng, J.-C., Blinder, S., Camborde, M.-L. & Sossi, V. (2005), 'Statistical dynamic image reconstruction in state-of-the-art high-resolution pet', *Physics in Medicine & Biology* **50**(20), 4887.

Reader, A. J., Julyan, P. J., Williams, H., Hastings, D. L. & Zweit, J. (2003), 'Em algorithm system modeling by image-space techniques for PET reconstruction', *Nuclear Science, IEEE Transactions on* **50**(5), 1392–1397.

Reader, A. J. & Verhaeghe, J. (2014), '4d image reconstruction for emission tomography', *Physics in medicine and biology* **59**(22), R371.

Richardson, W. H. (1972), 'Bayesian-based iterative method of image restoration', *JOSA* **62**(1), 55–59.

Ridgway, J. P. (2010), 'Cardiovascular magnetic resonance physics for clinicians: part i', *Journal of cardiovascular magnetic resonance* **12**(1), 71.

Ronneberger, O., Fischer, P. & Brox, T. (2015), U-net: Convolutional networks for

biomedical image segmentation, *in* 'International Conference on Medical image computing and computer-assisted intervention', Springer, pp. 234–241.

Rousset, O. G., Ma, Y. & Evans, A. C. (1998), 'Correction for partial volume effects in pet: principle and validation', *Journal of Nuclear Medicine* **39**(5), 904–911.

Rumelhart, D. E., Hinton, G. E. & Williams, R. J. (1985), Learning internal representations by error propagation, Technical report, California Univ San Diego La Jolla Inst for Cognitive Science.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A. C. & Fei-Fei, L. (2015), 'ImageNet Large Scale Visual Recognition Challenge', *International Journal of Computer Vision (IJCV)* **115**(3), 211–252.

Schreibmann, E., Nye, J. A., Schuster, D. M., Martin, D. R., Votaw, J. & Fox, T. (2010), 'Mr-based attenuation correction for hybrid pet-mr brain imaging systems using deformable image registration', *Medical physics* **37**(5), 2101–2109.

Stayman, J. W. & Fessler, J. A. (2000), 'Regularization for uniform spatial resolution properties in penalized-likelihood image reconstruction', *Medical Imaging, IEEE Transactions on* **19**(6), 601–615.

Szegedy, C., Ioffe, S., Vanhoucke, V. & Alemi, A. A. (2017), Inception-v4, inception-resnet and the impact of residual connections on learning, *in* 'Thirty-First AAAI Conference on Artificial Intelligence'.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V. & Rabinovich, A. (2015), Going deeper with convolutions, *in* 'Proceedings of the IEEE conference on computer vision and pattern recognition', pp. 1–9.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J. & Wojna, Z. (2016), Rethinking the inception architecture for computer vision, *in* 'Proceedings of the IEEE conference on computer vision and pattern recognition', pp. 2818–2826.

Tang, J. & Rahmim, A. (2009), 'Bayesian PET image reconstruction incorporating anato-functional joint entropy', *Physics in medicine and biology* **54**(23), 7063.

Turkheimer, F. E., Boussion, N., Anderson, A. N., Pavese, N., Piccini, P. & Visvikis, D. (2008), 'PET image denoising using a synergistic multiresolution analysis of structural (MRI/CT) and functional datasets', *Journal of Nuclear Medicine* **49**(4), 657–666.

Ullisch, M. G., Scheins, J. J., Weirich, C., Kops, E. R., Celik, A., Tellmann, L., Stöcker, T., Herzog, H. & Shah, N. J. (2012), 'Mr-based PET motion correction procedure for simultaneous MR -PET neuroimaging of human brain', *PloS one* **7**(11), e48149.

van Cittert, P. H. (1931), 'Zum einfluß der spaltbreite auf die intensitätsverteilung in spektrallinien. ii', *Zeitschrift für Physik* **69**(5), 298–308.
URL: *https://doi.org/10.1007/BF01391351*

Vincent, P., Larochelle, H., Bengio, Y. & Manzagol, P.-A. (2008), Extracting and composing robust features with denoising autoencoders, *in* 'Proceedings of the 25th international conference on Machine learning', ACM, pp. 1096–1103.

Vincent, P., Larochelle, H., Lajoie, I., Bengio, Y. & Manzagol, P.-A. (2010), 'Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion', *The Journal of Machine Learning Research* **11**, 3371–3408.

Wang, G. & Qi, J. (2012), 'Penalized likelihood PET image reconstruction using patch-based edge-preserving regularization', *Medical Imaging, IEEE Transactions on* **31**(12), 2194–2204.

Wang, G. & Qi, J. (2015), 'PET image reconstruction using kernel method', *Medical Imaging, IEEE Transactions on* **34**(1), 61–71.

Wells, R. G., King, M. A., Simkin, P. H., Judy, P. F., Brill, A. B., Gifford, H. C., Licho, R., Pretorius, P. H., Schneider, P. B. & Seldin, D. W. (2000), 'Comparing filtered back-projection and ordered-subsets expectation maximization for small-lesion detection and localization inˆ 6ˆ 7ga spect', *Journal of Nuclear Medicine* **41**(8), 1391–1399.

Wiesler, S. & Ney, H. (2011), A convergence analysis of log-linear training, *in* 'Advances in Neural Information Processing Systems', pp. 657–665.

Xu, E., Mullani, N. A., Gould, K. L. & Anderson, W. (1991), 'A segmented attenuation correction for pet', *J Nucl Med* **32**(1), 161–165.

Yan, J., Planeta-Wilson, B., Gallezot, J. & Carson, R. E. (2009), Initial evaluation of direct 4D parametric reconstruction with human pet data, *in* 'Nuclear Science Symposium Conference Record (NSS/MIC), 2009 IEEE', IEEE, pp. 2503–2506.

Yang, J., Huang, S., Mega, M., Lin, K., Toga, A., Small, G. & Phelps, M. (1996), 'Investigation of partial volume correction methods for brain FDG PET studies', *IEEE Transactions on Nuclear Science* **43**(6), 3322–3327.

Zhou, Y., Tang, J. & Wong, D. (2009), Direct 4D parametric image reconstruction with plasma input and reference tissue models in reversible binding imaging, *in* 'Nuclear Science Symposium Conference Record (NSS/MIC), 2009 IEEE', IEEE, pp. 2516–2522.