Qualitative and Quantitative Improvements for Positron Emission Tomography Using Different Motion Correction Methodologies

A Dissertation Presented for the

Doctor of Philosophy

Degree

The University of Tennessee, Knoxville

Tasmia Rahman Tumpa

December 2021

© by Tasmia Rahman Tumpa, 2021
All Rights Reserved.

Acknowledgments

I would like to express my sincere gratitude to my advisor Dr. Jens Gregor, for rendering his valuable guidance and mentoring over the five years of my doctoral journey. I will always be thankful for what I learned from him during this period.

I would like to thank Dr. Hairong Qi and Dr. Amir Sadovnik, for serving on my doctoral committee and providing their valuable feedback. The knowledge that I gathered from their academic courses has been a great motivation for my research. I am highly grateful to Dr. Dustin Osborne for all of his cooperation and constant support. My quest as a researcher in the medical field has been made easy by his expertise and helpfulness. I would also like to thank the people at the Graduate School of Medicine (University of Tennessee-Knoxville) for their assistance in acquiring the data for this research. I also wish to extend my humble gratitude for the Ronald Nutt Fellowship in Imaging Formation and Processing that I received during my graduate study. It has been a great honor and support to receive the fellowship.

I would like to dedicate this work to my family, who served as the greatest motivation in pursuing my PhD. Without the moral support from my mother Nasima Akhter, and my father Hafizur Rahman, it would not have been possible for me to reach this far. I am immensely indebted to my husband Tamim for always being there for me throughout this journey. I would like to thank my sister Mouri, and her family for lending all possible encouragement and morale-boosting. Lastly and most importantly, I would like to express my deepest gratitude to God, who has showered His blessings and allowed me to reach where I am today.

Abstract

Positron Emission Tomography (PET) data suffers from low image quality and quantitative accuracy due to different kinds of motion of patients during imaging. Hardware-based motion correction is currently the standard; however, is limited by several constraints, the most important of which is retroactive data correction. Data-driven techniques to perform motion correction in this regard are active areas of research. The motivation behind this work lies in developing a complete data-driven approach to address both motion detection and correction. The work first presents an algorithm based on the positron emission particle tracking (PEPT) technique and makes use of time-of-flight (TOF) information from the scanner to enhance the performance of motion detection. The algorithm functions entirely in raw coincidence event data space and facilitates the detection of all kinds of motion, including global and local. The application of the algorithm was first studied on respiratory motion correction both in phantom and clinical studies. The performance was validated against the current clinical standard- external pressure sensor band, and the detected motion signals by the proposed algorithm correlated well with the band. The algorithm was also compared against two state-of-the-art algorithms, and it was found to outperform both. Next, the application of the algorithm was explored in head-neck motion detection and therefrom motion correction. This work adopted two different paths for motion correction, namely, conventional and deep learning approaches. The former was implemented by proposing a variation of the PEPT algorithm for raw PET coincidence event data registration. The latter was explored by means of training a deep convolutional neural network for PET image data registration. Both approaches were validated against current clinical standards and found to produce comparable results with increased computational efficiency. The motivation behind this work as a whole was to achieve qualitative and quantitative improvements of the PET

data by implementing the proposed motion detection and correction approaches, as well as provide a detailed analytical study to support the ongoing research in this field.

Table of Contents

1 Introduction **1**

1.1 Motivation 1

1.2 Research Goal 2

1.3 Research Contribution 3

1.4 Overview 4

2 Background **5**

2.1 Positron Emission Tomography 5

2.1.1 Working Principle 5

2.1.2 Types of Coincidence Events 6

2.1.3 Time-of-flight (TOF) Information 8

2.1.4 Data Processing 8

2.2 Motion Artifacts during PET 9

2.3 Motion Correction Approaches 9

2.4 Related Work on Respiratory Motion 12

2.5 Related Work on Head-Neck Motion 14

2.6 Deep Learning in Image Registration 16

3 Motion Detection Algorithm: Application to Respiratory Motion Correction **18**

3.1 Time-of-Flight Weighted Positron Emission Particle Tracking (TOF-PEPT) . 19

3.1.1 Overview of the TOF-PEPT Algorithm 20

3.1.2 Implementation Details of the TOF-PEPT Algorithm 23

3.2 Performance Evaluation . . . 24
3.2.1 Validation and Evaluation . . . 24
3.2.2 Data . . . 28
3.2.3 Qualitative Evaluation of Motion Signals Estimated by TOF-PEPT against Other Methods . . . 29
3.2.4 Quantitative Evaluation of Motion Signals Estimated by TOF-PEPT against Other Methods . . . 36
3.2.5 Qualitative and Quantitative Evaluation of Motion Compensated Image . . . 40
3.2.6 Analysis of the Algorithm Parameters . . . 42
3.3 Summary . . . 42

4 Motion Detection Algorithm: Application to Head-Neck Motion Correction 50
4.1 Implementation Details of Motion Correction . . . 51
4.1.1 Static Frame Detection . . . 51
4.1.2 Point Source Detection: TOF-weighted Line Density Method . . . 52
4.1.3 Extraction of Transformation Parameters . . . 55
4.2 Performance Evaluation . . . 56
4.2.1 Validation and Evaluation . . . 56
4.2.2 Data . . . 57
4.2.3 Qualitative Analysis on the Estimated Motion and Quantitative Evaluation of Transformation Parameters . . . 57
4.2.4 Qualitative and Quantitative Comparison of Motion Corrected Images 59
4.3 Summary . . . 65

5 Deep Learning Based PET Motion Correction 70
5.1 Brief Overview of Image Registration . . . 70
5.2 Network Architecture . . . 71
5.3 Loss Functions . . . 72
5.4 Performance Evaluation . . . 74

5.4.1 Validation and Evaluation 74
5.4.2 Data 75
5.4.3 Qualitative and Quantitative Evaluation of the Neural Network Performance 75
5.4.4 Analysis of the Neural Network Parameters 83
5.4.5 Summary 83

6 Summary and Next Steps **86**
6.1 Summary of Research Contributions 86
6.2 Future Work 88

Bibliography **90**

Vita **105**

List of Tables

3.1 Detailed description of the experimental studies 30

3.2 Quantitative comparison of detected motion signals by the proposed TOF-PEPT method against the gold standard external pressure sensor band (Anzai) and the state-of-the-art data-driven method by center-of-mass (COM) 39

3.3 Performance of the proposed method under different conditions 41

3.4 Number of line-of-responses (LOR)s and iterations used by the algorithm with varying sizes of region-of-interest (ROI) 46

3.5 Number of line-of-responses (LOR)s and iterations used by the algorithm with a varying frame duration 46

4.1 Description of the experimental studies conducted 58

4.2 Mean Euclidean distance between reference and transformed coordinates . . 62

4.3 Quantitative evaluation of motion corrected image 66

5.1 Comparison of image registration performance 77

5.2 Comparison of performance for the application to motion correction 80

List of Figures

2.1 Illustration of the distribution of line-of-responses (LOR) from a positron emitter in the field-of-view. 7

2.2 Illustration of different types of coincidence events: (a) True events where two photons from a single annihilation event are detected; (b) Random events where two photons from two different annihilation events are detected as coincidence events; (c) Scattered events where one of the photons get scattered. 7

2.3 Illustration of respiratory gating approaches: (a) amplitude-based gating where the data are binned according to the range of amplitude (b) phase-based gating where the data are binned according to the breathing cycle. . 11

2.4 Illustration of phase-based gating approaches where the data are binned according to the breathing cycle into eight temporal windows. 11

2.5 Example of motion impact due to respiratory (top row) and head-neck (bottom row) motion: (a) A slice of lung PET data in the coronal plane from a motion blurred image volume due to respiratory motion (b) The same slice from a motion minimized subset of the data obtained by phase-based respiratory gating (c) A slice of brain PET data in the axial plane from a motion impacted image volume due to the random orientation of head (d) The same slice from a motion free frame of the data from the initial head position. The impact of motion can be seen in the arrow marked regions in the motion impacted images in (a) and (c) where the lesion in the lung area appears elongated, and there appears a loss of details in the brain area, respectively. 13

3.1 Illustration of the concept of time-of-flight (TOF): The true annihilation position lies somewhere along the line of response (LOR) joining the detector pair. If d is the distance between detector A and B, then Δd from TOF information allows estimating the true annihilation position with an uncertainty that follows a Gaussian distribution centered on the TOF bin with full width at half maximum (FWHM) corresponding to the scanner system timing resolution. 22

3.2 The flow diagram of the TOF-PEPT algorithm: Multiple frames of short duration are created and passed to the algorithm along with the user defined region-of-interest (ROI). The algorithm provides the location coordinate of the object within ROI for each sampling frame. The location as a function of time, provides an estimate of the actual motion. 26

3.3 For Phantom Study 1, comparison of the motion signals estimated by the TOF-PEPT versus the Anzai band: (a) Displacement in the Y direction for a 60-seconds time frame; (b) Displacement in the Y direction for the entire 3-minute scan. The vertical lines across the peak of the signals show the inspiration peaks. It can be seen that the phase, displacement, and inspiration peak time corresponded well. The Anzai band data magnitude was normalized by its maximum value to aid visualization. The TOF-PEPT data is shown in the original scale. 31

3.4 For Clinical Study 1, comparison of the motion signals estimated by the TOF-PEPT versus the Anzai band: (a) Displacement in the Y direction for a 60-seconds time frame; (b) Displacement in the Y direction for the entire 5-minute scan. The vertical lines across the peak of the signals show the inspiration peaks. It can be seen that the phase, displacement, and inspiration peak time corresponded well except for a few places where there were involuntary movements of the patient and issues of missing triggers from the Anzai band. The Anzai band data magnitude was normalized by its maximum value to aid visualization. The TOF-PEPT data is shown in the original scale. 32

3.5 For Phantom Study 2, comparison of motion signals estimated by TOF-PEPT, TOF-COM, and COM: (a) Time domain plots for the full 15-minute scan. Baseline shift was added to the TOF-COM and COM signals to aid visualization; (b)-(d) Time domain plots with frames of 60-seconds for the three different frequencies and amplitude ranges; (e)-(g) Fourier domain plots for the full 15-minute scan. 34
3.6 For Clinical Study 3, comparison of motion signals estimated by TOF-PEPT, TOF-COM, and COM: (a) Time domain plots; (b)-(d) Fourier domain plots. 35
3.7 For Clinical Study 5, comparison of motion signals estimated by TOF-PEPT, TOF-COM, and COM from left cardiac ventricle: (a) Time domain plots of the respiratory motion. To separate the respiratory motion from the cardiac motion, a low pass filter with a cutoff frequency of 0.5 Hz was applied; (b)-(d) Fourier domain plots 37
3.8 For Clinical Study 6, comparison of motion signals estimated by TOF-PEPT, TOF-COM, and COM from an internal lesion in the lung area: (a) Time domain plots; (b)-(d) Fourier domain plots 38
3.9 For Clinical Study 1. comparison of the (a) ungated and (b)-(e) amplitude gated images created by using motion signals estimated by TOF-PEPT, Anzai band, TOF-COM, and COM, respectively. Increased maximum SUV and sharper image quality are observed in the gated image by the TOF-PEPT algorithm. 43
3.10 For Clinical Study 6, comparison of the (a) ungated and (b)-(d) amplitude-gated images created by using motion signals estimated by TOF-PEPT, TOF-COM, and COM, respectively. The lesion area, along with its maximum SUV is indicated by an arrow in each image. (e) Result of Gaussian-fitted line profile data (a-d vertical lines). SUV can be seen to be the highest, and FWHM can be seen to be the lowest in the gated image by the TOF-PEPT algorithm. 44

3.11 Quantitative comparison of (a) correlation coefficient, (b) RMSE, and (c) normalized mean shift in peak location time for varying scaling factors of ROI circumradius. 45
3.12 Quantitative comparison of (a) correlation coefficient, (b) RMSE, and (c) normalized mean shift in peak location time for varying frame duration. . . 45

4.1 Illustration of the process of detecting static frames: (a) Estimated motion signal (b) Gradient norm of the forward different approximation of the motion signal. Based on the data-driven threshold, time frames were sorted into transitional frames and longer duration static frames. 53
4.2 Illustration of the process of Bresenham mapping: (a) A line of response (LOR) with its time-of-flight (TOF) center (b) Incrementing the voxels intercepted by the LOR following a Gaussian distribution centered on the TOF center and with full width at half maximum (FWHM) corresponding to the system timing resolution 54
4.3 Plots of the detected motion signals (along the principal direction of motion) and static frames: (a) Patient Study 1: Three static frames; (b) Patient Study 2: Four static frames; (c) Patient Study 3: Four static frames; (d) Patient Study 4: Three static frames; and (e) Patient Study 5: Four static frames. The gray colored regions show the transitional frames. 60
4.4 Point sources' locations in the axial plane (top row) (a) before and (b) after the application of the transformation and in the coronal plane (bottom row) (c) before and (d) after the application of the transformation. It can be seen that the point sources were aligned in different orientation in each different frame before and got aligned to the reference position after applying the transformation. 61

4.5 Comparison of uncorrected and motion corrected images for Patient Study 3 (Top row) and Patient Study 5 (Bottom row) in the axial plane from: (a, d) Original motion impacted image; (b, e) Corrected image by applying frame-by-frame manual landmark based registration; (c, f) Corrected image by applying the proposed automated listmode event data correction. 63
4.6 Comparison of uncorrected and motion corrected images for Patient Study 3 (Top row) in the sagittal plane and Patient Study 5 (Bottom row) in the coronal plane: (a, d) Original motion impacted image; Corrected image by applying frame-by-frame manual landmark based registration; (c, f) Corrected image by applying the proposed automated listmode event data correction. . 64
4.7 Motion corrected images showing the impact of attenuation correction for Patient Study 3 in (Top row) axial plane and (Bottom row) sagittal plane : (a, c) With attenuation correction (b, d) Without any attenuation correction. 67

5.1 Multi-step affine registration network: Initially, source and target images are concatenated and fed into the convolutional neural network. The network outputs transformation parameters which are passed to the spatial transform layer along with the source image. The layer outputs warped images. At the next step, the warped image is passed as the source image, and the process continues for t number of steps. 73
5.2 Convolutional neural network: Concatenated source and target images are passed to a series of convolutional and pooling layers. At the final layer, two convolution operations yield the transformation parameters. 73
5.3 Example of three static frames where the patient rested their head in three different positions. The figure shows a slice from 3D volume in the axial plane. 78
5.4 Example of neural network performance in warping moving source image. Rows from top to bottom show the sum of the three frames without any correction (left), correction with registration from the deep learning approach (DL-middle), and correction with registration using ANTs iterative algorithm (right), respectively, in the axial, coronal, and sagittal view. 78

5.5 Example of three static frames where the patient rested their head in three different positions. The figure shows a slice from 3D volume in the axial plane. 79
5.6 Example of neural network performance in warping moving source image. Rows from top to bottom show the sum of the three frames without any correction (left), correction with registration from the deep learning approach (DL-middle), and correction with registration using ANTs iterative algorithm (right), respectively, in the axial, coronal, and sagittal view. 79
5.7 A line profile (Right) drawn across a region in the brain area (Left). It can be seen that the peak to valley ratio is higher for motion corrected data with deep learning (DL) than ANTs. 81
5.8 A line profile (Right) drawn across a region in the brain area (Left). It can be seen that the peak to valley ratio is higher for motion corrected data with deep learning (DL) and ANTs. 81
5.9 Example of neural network performance in creating motion corrected listmode data and therefrom reconstructed image volume. Rows from top to bottom show the sum of the three frames without any correction (left), correction with registration from the deep learning approach (right), respectively, in the axial, coronal, and sagittal view. 82
5.10 The plot of loss versus epoch shows the neural network performance with changing step sizes of 1, 2, 3, and 4, respectively. It can be seen that with increasing step size, the network performance improved and saturated at step size 4 for the dataset used. 84
5.11 The plot shows the neural network performance with 4 different network architectures. Network 1: 5 levels of convolution with 32, 64, 128, 256, and 512 filter kernels; Network 2: 5 levels of convolution with 16, 32, 64, 128, and 256 filter kernels; Network 3: 4 levels of convolution with 32, 64, 128, and 256 filter kernels; and Network 4: 4 levels of convolution with 16, 32, 64, and 128 filter kernels. It can be seen that the network performed better with 5 levels of convolution, and the filter choice of case 1 performed comparatively better than case 2. 84

Chapter 1

Introduction

1.1 Motivation

Positron Emission Tomography (PET) is a non-invasive technique widely used in clinical applications to study organ and tissue functions. One of the major concerns with PET imaging is motion artifacts caused by different kinds of movements of the patients being scanned. These movements can be classified as periodic, non-rigid motion such as respiratory motion, cardiac motion, and irregular, rigid motion such as head-neck motion, whole body motion, etc. These movements result in a blurred reconstructed image with degraded image quality. This impacts the image analysis for diagnosis purposes such as evaluating standardized uptake value (SUV), measuring lesion intensity, size, location, etc. [19]

The use of external devices to measure motion signals from the patient is the most widely practiced approach. Pressure sensors, spirometer, temperature sensors, video cameras, position sensitive detectors, etc., are commonly used devices to monitor different kinds of patient motion. However, existing key issues include device cost and setup, necessary training to use the devices, regular maintenance, etc., and most importantly, they do not allow retroactive corrections of data. For a dataset where these devices were not used, there is no other way to correct the data except for some way to extract the motion information directly from the data itself. These limitations thus have shifted the focus towards data-driven methods for the detection and correction of motion during PET.

To date, several data-driven motion correction approaches have been studied. They differed in the way the motion detection and corrections were performed. The simplest way to perform data-driven motion correction is frame-by-frame (FBF) registration. The data is sorted into a sequence of frames of fixed or arbitrarily defined varying time duration. An image from each frame is then reconstructed, realigned to a reference frame, and summed together post hoc. However, with fixed or arbitrarily defined frame intervals, intra-frame motion does not get resolved and thus requires prior motion information.

Center-of-mass (COM), principal component analysis (PCA), spectral analysis, etc., are existing approaches that allow detecting motion directly from raw PET data. Most existing work has addressed global motion detection only, with little or no exploration into internal region/local motion detection. The motivation of this dissertation is to develop a robust, data-driven motion detection algorithm that can be applied to detect all kinds of patient movements directly from listmode data and study the application of the algorithm for respiratory and head-neck motion correction.

The next motivation lies in exploring computationally efficient methodologies to perform motion correction in a purely data-driven manner. With known motion information, correction is addressed by creating motion-free static frames and therefrom performing frame-by-frame registration. In most cases, the registration is performed in the image domain and typically by optimizing different similarity criteria between fixed and moving image frames. Image registration is a computationally expensive method and can take significant processing time depending on the complexity of the task. This work attempts to address the registration task with improved efficiency and without compromising accuracy.

1.2 Research Goal

The primary goal is to propose a robust, data-driven motion detection algorithm that can be used for different kinds of motion detection, from a small range of motion of small lesions due to breathing to a large range of motion of head-neck due to irregular body movements.

This work is then aimed at addressing motion correction with computational efficiency and reliable performance. With the known motion information from the proposed motion

detection algorithm, two different motion correction approaches will be explored, namely, conventional and deep learning techniques.

1.3 Research Contribution

This work explored complete data-driven motion detection and correction approaches for the quantitative and qualitative improvements of motion suffered PET data. The work first proposed a motion detection algorithm based on positron emission particle tracking (PEPT) [70, 71, 72] technique which is a well-established algorithm, mostly being used in industrial applications for measuring velocity, flow patterns, etc. The original PEPT algorithm has been modified to make its implementation entirely data-driven and compatible with clinical applications. The algorithm was further adapted to make use of time-of-flight (TOF) information provided by the modern PET scanning systems to allow improved performance of the movement detection. The resulting algorithm has been referred to as TOF-PEPT hereafter.

The next contribution has been the study of the application of the proposed motion detection approach for mainly two different kinds of motion correction: respiratory and head-neck motion. For respiratory motion, region specific motion tracking has been performed, and amplitude-based gating [64, 31] has been implemented to produce motion compensated data. For head-neck motion correction, a line density based PEPT method has been proposed to perform remapping of the raw data with the aid of external markers. The use of external markers facilitated the calculation of transformation parameters needed to perform remapping and thereby bypass the process of PET image registration. The proposed method allowed an automated way of correcting the raw PET data itself instead.

Finally, along with the proposed conventional approaches to perform motion correction, a deep learning based solution has been studied to train a neural network for the task of PET image registration, i.e. predicting transformation parameters. Image registration with a trained neural network is computationally less expensive compared to the conventional iterative approach. PET motion correction with deep neural networks has been addressed with limited scope. Thus, the motivation was to study the feasibility of the application of

deep learning approach for PET motion correction and compare its performance with the conventional iterative approach. The research contributions can be enumerated as below:

- A clinical implementation of a data-driven particle tracking algorithm with the use of TOF information to detect different kinds of patient movements.
- Study and validate the performance of the proposed motion detection algorithm in correcting PET data for respiratory motion.
- Study and validate the performance of the proposed motion detection algorithm in correcting PET data for head-neck motion by means of a line density based particle tracking algorithm with external markers.
- Proposal of a deep learning approach for PET motion correction and comparing its performance with the conventional iterative approach.

1.4 Overview

The rest of this dissertation is organized as follows: Chapter 2 gives a brief background of PET, motion artifacts and correction, and provides a literature review of the prior works on motion detection, motion correction, and PET image registration with deep learning. Chapter 3 presents a detailed description of the proposed TOF-PEPT algorithm for motion detection and its application to respiratory motion correction. Chapter 4 describes the application of the motion detection algorithm for head-neck motion correction. Chapter 5 presents a deep learning approach for motion correction by means of PET image registration. Lastly, Chapter 6 briefly summarizes the work presented here and the work slated for the future.

Chapter 2

Background

2.1 Positron Emission Tomography

Positron emission tomography (PET) is a nuclear medicine procedure that allows visualization of metabolic processes in the body, such as measuring glucose utilization in groups of cells. In PET, a biologically active compound labeled with a radioisotope, typically called radiopharmaceutical, radionuclide, or radioactive tracer, is applied to the patient by means of injection or inhalation, and the image of how this radioactive material is processed in different parts of the body is obtained by means of the PET scanner. This helps to diagnose and stage disease based on functional imaging unlike other modalities liked CT and MRI that typically are used only for anatomical reference. For instance, the imaging of cells with high metabolisms, such as liver, brain, and cancer cells, allows for the detection of abnormal tracer metabolism of the tissue being studied. For tumor or cancer treatment, PET is used for diagnosis and evaluating how a patient is responding to therapy treatment. The working principle, along with some key aspects related to PET, is described below.

2.1.1 Working Principle

The radioactive tracers used in PET are made by attaching radioactive isotope to chemical elements that are used by particular tissue or organ. The commonly used radioactive isotope in PET scanning are Fluorine-18 (18F), Carbon-11 (11C), Nitrogen-13 (13N), Oxygen-15

(15O), and they decay by positron emission with a varying half-life of approximately 110 minutes, 20 minutes, 10 minutes, and 2 minutes, respectively [61]. The most commonly used radioisotope is F-18 owing to its long half-life. The decay can be mathematically expressed as:

$$N(t) = N_0 \left(\frac{1}{2}\right)^{\frac{t}{t_{1/2}}} \tag{2.1}$$

Here, N_0 and N_t respectively denote initial and remaining quantity, and $t_{1/2}$ denotes the half-life. The radioactive material undergoes $\beta+$ decay via the emission of a positron. Positrons emitted from the nucleus of radioisotopes are anti-electrons that combine with electrons after traveling a short distance. This results in the conversion of their masses into equivalent energy ($E = mc^2$) through the emission of two oppositely directed 511-keV γ rays. This phenomenon is referred to as annihilation. The rays are detected by scintillation detectors of the PET scanner. When the two γ rays strike the opposing detectors within a small timing window (4-12 ns [87]), they are detected as a coincidence event. A line joining the two detectors allows localizing the annihilation event i.e. the positron emitter, and the line is typically referred to as the *line of response* (LOR). There can be millions of LORs detected within the field-of-view (FOV), as depicted in Figure 2.1.

2.1.2 Types of Coincidence Events

The coincident events can be classified into true events and background events. The latter further can be of two types: random events and scattered events [82, 87]. True events refer to the situation when two detected photons come from the same annihilation event. On the other hand, there can be a situation where one of the photons may get lost, and two photons from two different annihilation events are by chance detected within the same coincidence timing window. This results in random events, and the line joining them does not localize any of the two actual annihilation locations. Scattered events occur when two photons originate from the same annihilation event, but the line joining them does not localize the true annihilation position in this case as well, as one of the photons undergoes Compton scattering. Collectively all coincident events are called prompts. Figure 2.2 illustrates the phenomenons.

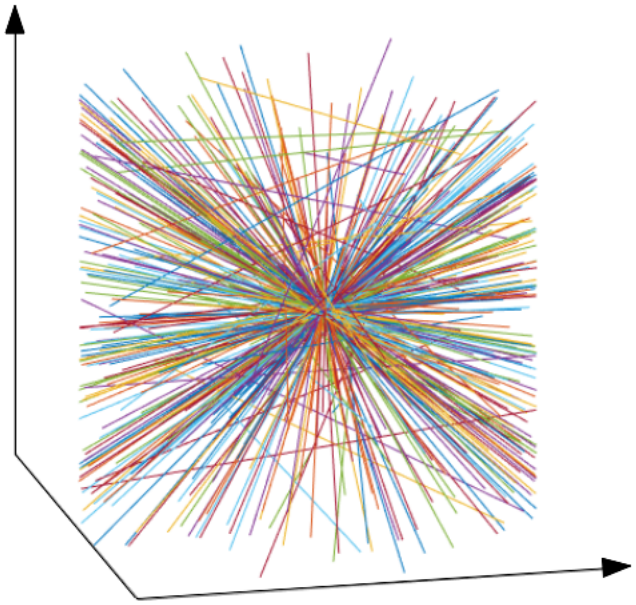


Figure 2.1: Illustration of the distribution of line-of-responses (LOR) from a positron emitter in the field-of-view.

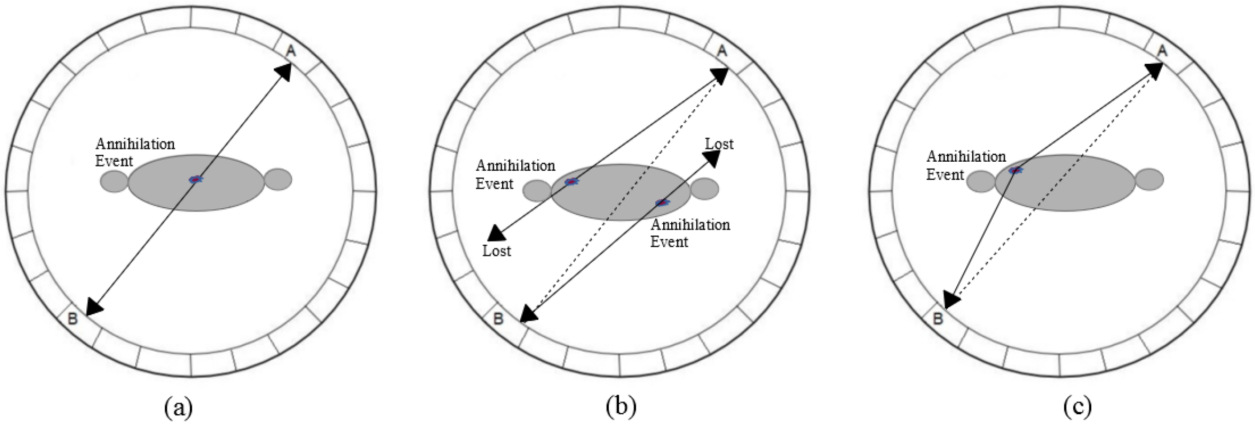


Figure 2.2: Illustration of different types of coincidence events: (a) True events where two photons from a single annihilation event are detected; (b) Random events where two photons from two different annihilation events are detected as coincidence events; (c) Scattered events where one of the photons get scattered.

2.1.3 Time-of-flight (TOF) Information

With the advances in PET scanning systems, it has been made possible to measure the difference between the time taken for each γ ray to reach the opposite detectors, what is known as time-of-flight (TOF) information. From this information, how far the annihilation location is from the center of a LOR can be measured. This provides an estimate of the true annihilation location, mostly referred to as TOF center/bin, and aids in visualizing PET data with an improved signal-to-noise ratio (SNR). Depending on the scanner system, the timing resolution can be as low as 215 ps [79].

2.1.4 Data Processing

For the whole duration of the PET scan, information associated with each coincidence event, such as index of the detector pairs, time-of-flight, etc., is collected as a chronologically sorted list of data, known as *listmode* data. This raw data is further organized into a two-dimensional accumulation array such that it provides a set of projection data. The format is referred to as a *sinogram*, where the two directions represent the perpendicular distance of LORs from the center of the scanner and the angle of the orientation of LORs, respectively. Applying Filtered Back Projection (FBP) or other image reconstruction techniques on the sinogram data, a cross-sectional image is reconstructed, which represents the distribution of the radioactivities within the body. To obtain a clinically accurate image, data correction is needed afterward, which includes attenuation correction, scatter correction, etc, [75]. While traveling within the body from the annihilation location to the detector, photons get absorbed or attenuated depending on the density of the tissue. This results in inaccurate quantification of uptake. To correct for this, attenuation through the patient for all LORs is determined either by a transmission scan for PET only scanning system or by a CT image for a combined PET/CT system.

2.2 Motion Artifacts during PET

A major concern with PET imaging with respect to artifacts caused by the motion of the patients. PET scanning can require significantly longer scan times (may take up to 90 minutes for the whole body scan) which makes it difficult for patients to hold their breath or remain steady. The motion due to breathing or irregular movements thus results in motion blurred image. This can be challenging for application in radiotherapy or oncology imaging. For quantitative evaluation of PET data, it is a common practice to use standardized uptake value (SUV), which can be expressed as [27]:

$$SUV = \frac{A_c}{D_c/w} \tag{2.2}$$

where A_c is the radioactivity concentration measured in Bq/mL, D_c is the decay corrected injected dose in Bq, and w is the patient's weight. This measure is helpful since it is normalized by the amount of injected dose with respect to patient size and thus reduces variability. However, with the presence of motion, the image-derived activity concentration gets affected. It was reported in [27] that for small lesions near the diaphragm, patient respiratory motion can result in an underestimation of SUV by 25% or more. Apart from the measure of SUV, motion impact may lead to a 2-fold overestimation of tumor/lesion size.

The issue with motion can be more challenging for a combined PET/CT scan that incorporates functional and anatomical information of organs from PET and CT scans, respectively. In combined PET/CT, CT scan data is to correct for attenuation in the PET images, and the spatial mismatch between PET and CT may lead to artifacts. Thus it is of significant importance to address motion impacts during PET scanning.

2.3 Motion Correction Approaches

Two important attributes of motion correction are discussed here for the ease of understanding the following chapters, namely, respiratory gating and frame-by-frame registration.

Respiratory gating refers to the process of sorting data into multiple gates or subsets with respect to different phases of the respiratory cycle, either based on time or magnitude [18]. Each subset is then associated with a small part of the total motion, and hence the effect of motion artifacts can be minimized. A motion-free or motion reduced image can be reconstructed for each gate [64]. The idea is to minimize motion impact within each small subset. Data from each gate/subset is then aligned to a reference position.

The method of gating can be classified into mainly two groups: amplitude-based gating and phase-based gating. The former sorts the data into multiple gates based on the amplitude of the breathing signal, whereas the latter does the sorting with respect to time in each breathing cycle [18, 34]. Data from corresponding time phases or amplitude ranges are summed together to create motion minimized data. Amplitude-based gating has been shown to perform better than phase-based since it takes into account the depth of respiration. Figure 2.3(a) illustrates the amplitude-based gating approach with an equal gate width, and Figure 2.3(b) shows the phase-based gating approach where data are sorted with respect to time in each cycle. Figure 2.4 shows an example from a patient study where the whole data was sorted into eight gates by dividing each breathing cycle into eight equal temporal windows.

However, with single gate image reconstruction, only a fraction of co-incident events are used, which results in degraded SNR in the reconstructed image. To avoid the problem stated, the most widely adopted practice is to use frame-by-frame registration. This is performed by sorting the listmode data into a sequence of frames of fixed or arbitrarily defined time duration and reconstructing image volume for each frame. Image from each frame is then realigned to the image from the reference frame. The registration is performed by optimizing different similarity criteria such as mutual information, cross-correlation, the sum of absolute difference, etc. The aligned frames are summed together post hoc to create a motion corrected image volume.

Figure 2.5 shows two examples of the impact of motion and its correction for respiratory and head-neck motion. Figure 2.5(a) shows a slice in the coronal plane from a motion impacted image volume, and Figure 2.5(b) shows the same slice from a motion-minimized subset of data by applying phase-based respiratory gating. It can be seen that the lesion

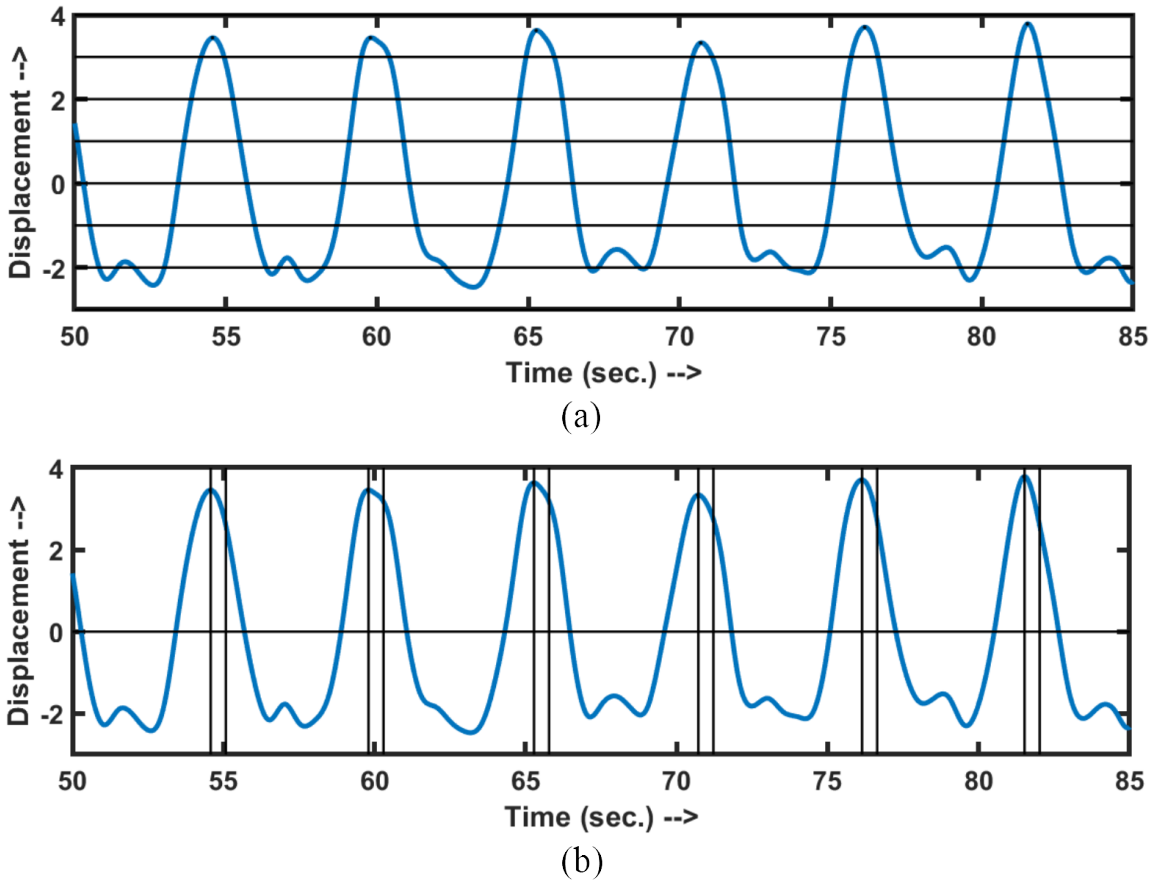


Figure 2.3: Illustration of respiratory gating approaches: (a) amplitude-based gating where the data are binned according to the range of amplitude (b) phase-based gating where the data are binned according to the breathing cycle.

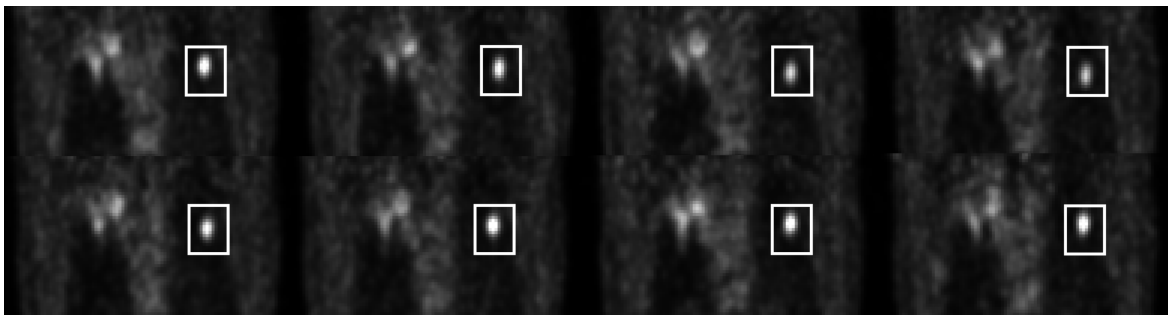


Figure 2.4: Illustration of phase-based gating approaches where the data are binned according to the breathing cycle into eight temporal windows.

in the lung area appears large/elongated in the summed image owing to the movement of the lesion with the breathing. This results in an overestimation of the lesion size and underestimation of the corresponding SUV. Figure 2.5(c) shows another example of motion impact due to head-neck motion for brain PET data. During the scanning period, the patient rested their head in different positions resulting in an overall motion blurred image, which can be seen from a slice in the axial plane. In contrast, another slice from the motion free frame of data from the initial head position (Figure 2.5(d)) allows the visualization of the brain area with more sharpness and details.

2.4 Related Work on Respiratory Motion

Prior work on respiratory motion correction primarily addressed the use of external devices. Dawood *et al.*[17], Qiao *et al.*[78], Lamare *et al.*[46], Liu *et al.*[51] presented methods using such devices. However, the recent focus shifted to data-driven methods because of several limitations associated with external devices, the most important of which is retroactive data corrections. In the absence of use of such a device, it is not possible to correct the data post hoc except for some data-driven techniques. Data-driven methods, furthermore, provide a means to determine lesion-specific (internal) motion, whereas, with external devices, it is only possible to observe global (external) motion. For the proper quantification of lesions, this ability is powerful since internal motion may not perfectly correlate with the external motion typically monitored.

Data-driven methods directly estimate motion from raw listmode data or reconstructed PET image/sinogram data. Some widely adopted techniques are Center of Mass (COM), Principal Component Analysis (PCA), spectral analysis, and Geometric Sensitivity Gating (GSG), etc. These are described next.

Klein *et al.*[39], Bundschuh *et al.*[10], Xu *et al.*[105] proposed motion estimation applying COM on sinogram data. The former two approaches tracked the axial component of the sinogram COM over short time segments, while the latter approach incorporated TOF information under different conditions. Thielemans *et al.*[90] applied PCA on down-sampled, short time frame sinograms, whereas Bertolli *et al.*[4, 5] proposed a method that automated

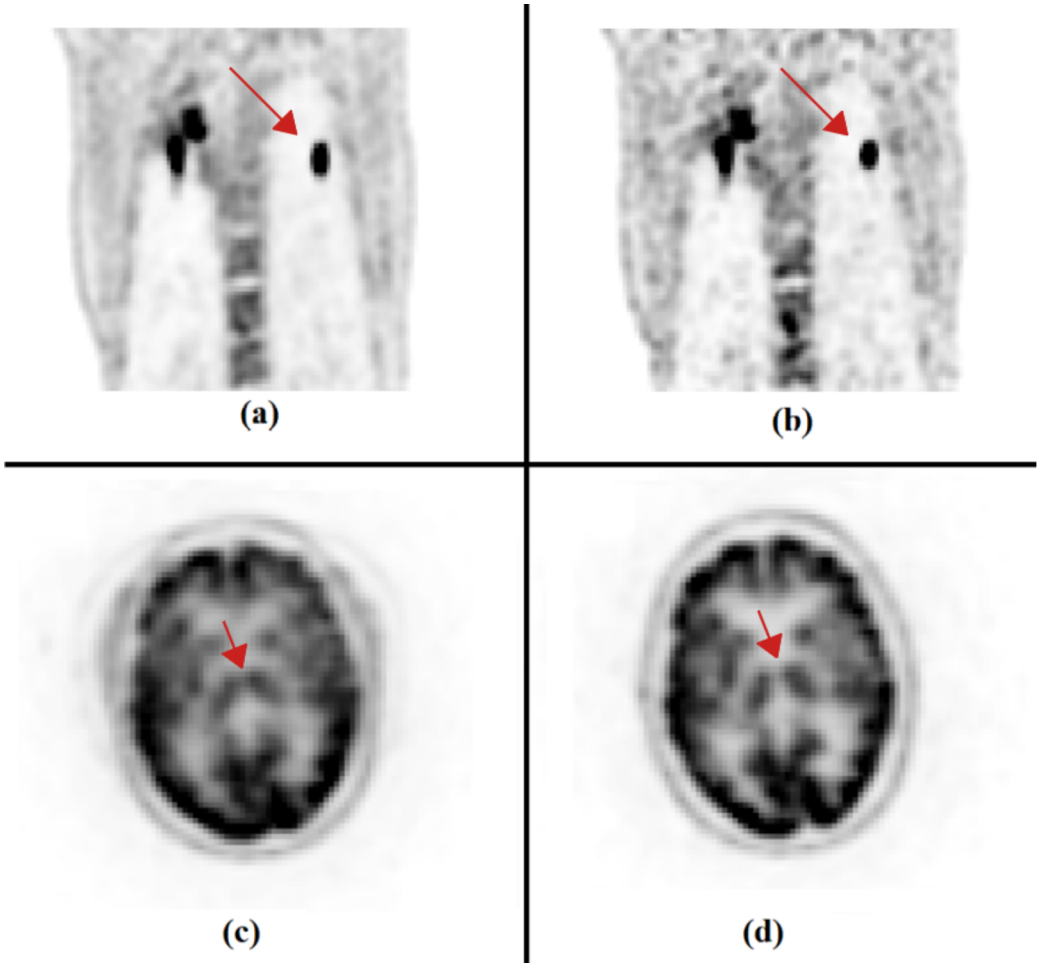


Figure 2.5: Example of motion impact due to respiratory (top row) and head-neck (bottom row) motion: (a) A slice of lung PET data in the coronal plane from a motion blurred image volume due to respiratory motion (b) The same slice from a motion minimized subset of the data obtained by phase-based respiratory gating (c) A slice of brain PET data in the axial plane from a motion impacted image volume due to the random orientation of head (d) The same slice from a motion free frame of the data from the initial head position. The impact of motion can be seen in the arrow marked regions in the motion impacted images in (a) and (c) where the lesion in the lung area appears elongated, and there appears a loss of details in the brain area, respectively.

this process. Later on, a similar kind of approach with Fourier Rebinning of TOF (FORET) was suggested by Wang *et al.*[97, 99, 98], where both COM and PCA were applied on the rebinned sinograms. Spectral analysis was studied by Visvikis *et al.*[96] in the form of Fourier analysis of time activity curves within an ROI. Kesner *et al.*[37, 36, 38] and Schleyer *et al.*[84, 85] studied variations thereof. He *et al.*[32], Büther *et al.*[11, 12] utilized the geometric sensitivity (GS) of the scanner and estimated the motion by observing the change in the total event count in each dynamically sorted sinogram frame, with the change in axial position. However, these methods require processing and reconstruction of multiple dynamic sinograms or images frames in advance. Also, some of these methods are limited by their ability to only estimate motion along specific directions, e.g., geometric sensitivity gating is restricted to estimating only axial motion. In this regard, motion estimation from raw listmode data allows motion detection in all three directions simultaneously.

Ren *et al.*[80], Lu *et al.* [53] studied COM based motion estimation by taking the average of the coordinates of coincidence events over a short duration. Either TOF center bin or center of LOR coordinates was considered as the coordinates of each event location. Feng *et al.*[25] later, proposed an algorithm where the uncertainty associated with the measurement of TOF information was taken into account as well. Salomon *et al.*[25] suggested another COM based approach, where motion was detected by tracking the average position of LORs in small overlapping ROIs and combining them. COM method has the advantage of being very straightforward and easy to implement. Nevertheless, with thousands to millions of LORs within an ROI, tracking the movement of a small lesion can be challenging, which has been shown in the next chapter.

2.5 Related Work on Head-Neck Motion

Unlike respiratory motion, which is not feasible to prevent other than holding the breath for an unreasonable amount of time, one simple approach to circumvent and minimize the impact of head-neck motion has been the use of head restraints, such as thermoplastic mask[29], head holders, the vacuum-lock bag[6], etc. However, they are not fully capable of restricting the motion completely, and some are uncomfortable for patients. The use of physical restraints

can also have a negative impact for dementia patients and should be avoided[101]. Thus other methods are needed to offer a better solution.

A widely used approach to motion correct the PET data is to extract the motion information using an external motion tracking system and correct the data accordingly. Several techniques have been studied to track the motion information using external hardware devices, such as video camera-based surveillance system[76], position-sensitive detectors[28], the Polaris tracking system[52], stereo vision based Tracoline tracking system[66], Microsoft Kinect[65], etc. These devices provide an estimate of motion information which is then used to correct the data either by repositioning LORs [35, 8, 60, 62] or image registration[62, 77]. Although external tracking systems are very useful and can collect motion information at a rate of several frames per second, the widespread use of these devices can be expensive to implement in every scanner system in terms of initial setup, necessary training, regular maintenance, etc.

In this regard as well, data-driven approaches offer substantial advantages and allow for correcting data from any imaging system regardless of whether external tracking systems were available or not. Frame-by-frame (FBF) image registration is the most widely practiced data-driven solution. Several research works have been proposed where the registration task was performed by optimizing mutual information[15, 74, 100], cross-correlation[15, 74, 50], the sum of absolute difference[50, 1], the standard deviation of the ratio of image volume[50, 1], etc. However, such approach suffers a trade-off between frame duration and accuracy/computational efficiency. For image noise minimization and accurate registration, a longer frame duration is favorable, but intra-frame motion can go unnoticed in such cases. Contrarily, with a shorter frame duration, it becomes computationally expensive due to the handling of multiple image reconstruction. Besides, without any prior knowledge of motion information, an intra-frame motion may remain unresolved due to arbitrarily defined frame boundaries. Thus having prior motion information and sorting the data into frames based on that alleviate these limitations.

Recent research works have proposed approaches to detect motion information directly from listmode data utilizing COM/PCA methods. Feng *et al.*[26], Thielemans *et al.*[91], Schleyer *et al.*[83] studied such approaches. However, motion correction was performed in

image space by means of FBF registration. Motion correction purely in raw event data domain approach has been explored in a limited capacity. Although Lu *et al.*[54] recently presented event-by-event motion correction, transformation parameters were obtained by FBF registration of reconstructed images. Iterative registration of images can be computationally expensive, and thus predicting transformation parameters in some other way can be beneficial to perform the data correction.

2.6 Deep Learning in Image Registration

Medical image registration plays a significant role in many clinical applications, such as segmentation, motion tracking, image reconstruction, etc. Image registration falls under different categories depending on input image type, deformation model, etc. Registration tasks can be classified into unimodal, multi-modal (PET-CT, PET-MRI, CT-MRI, etc.), intra-patient, patient-atlas, affine, rigid, deformable, etc. There exist several state-of-the-art traditional image registration approaches [2, 40, 92, 95]. These traditional approaches seek to find a numerical solution to an optimization problem and iterate over each image pair to solve the optimization problem. The computation can be very intensive, and thus in recent days, researchers have focused on exploiting the deep learning approach to replace the costly optimization task dedicated for a single pair of images with a trained model that will instead learn the underlying patterns of the registration task.

Till date, several research groups have explored many different approaches including Convolutional Neural Network (CNN) [21, 20, 3, 24, 47, 88], Generative Adversarial Network (GAN) [56, 23, 106], Reinforcement Learning (RL) [42, 55, 49] etc. One way of training a neural network is supervised learning which requires ground truth of the required transformation parameters [42, 88, 107, 81]. Either the networks are trained on synthetic images with known ground truths, or ground truth information is computed using other existing techniques. In routine clinical applications, accurate ground truth information is very difficult to gather, which makes supervise learning difficult. Thus, unsupervised and self-supervised learning have received special attention for medical image registration.

The introduction of Spatial Transform Network (STN) by Jaderberg *et al.* [33] paved a novel way for unsupervised image registration. It consisted of a neural network, a grid generator, and a sampler. The neural network was used to estimate a mapping between fixed and moving image pairs by learning features from them, the grid generator computed the sampling grid, and the sampler provided a warped/moved image by performing the sampling operation with interpolation. This allowed training the neural network in an end-to-end unsupervised manner. Several research works have been proposed later on with similar concepts. Li *et al.*[47], de Vos *et al.* [21, 20] proposed a Fully Convolutional Network (FCN), whereas Balakrishnan *et al.* [3] made use of a U-net-like architecture. These works were followed by several other research works that addressed different aspects, such as affine and deformable registration [20, 86], diffeomorphic registration [16, 43], reducing negative Jacobian determinant [44], encouraging invertibility [110, 86], etc. Also, to enhance the performance of the registration, different approaches, such as cascaded network [110, 20], multi-scale estimation [41, 24, 30], multi-step network [86], etc. have been proposed.

Most of these works addressed MRI and CT image registration. PET image registration has been so far addressed in a limited scope [104, 109, 108], with very few research works exploring the application of neural networks in PET motion correction [48]. PET images, in general, differ from other modalities, and thus study on the application of deep learning in PET motion correction is expected to be beneficial.

Chapter 3

Motion Detection Algorithm: Application to Respiratory Motion Correction

PET scan can take a longer time and patients may undergo different kinds of motion during that scanning period. A motion detection approach attempts to sort the whole scan data into multiple frames of short duration. Then, for each small frame, it is possible to track an object of interest within the field-of-view (FOV) or region-of-interest (ROI). Detecting the movement of the object from one frame to another gives a notion about motion information. In this chapter, a motion detection approach is presented, which tracks an object of interest in a data-driven manner. The algorithm is built upon the concept of tracking positron-emitting particles. The chapter provides a detailed insight into the working principle of the algorithm, analysis of its parameters, its application on the detection and thereby correction of respiratory motion, and a quantitative, and qualitative comparison study against the current clinical standard, and state-of-the-art methods with the aid of phantom and clinical studies.

Disclosure

This chapter is an adapted version of the manuscripts published in [93, 94].

3.1 Time-of-Flight Weighted Positron Emission Particle Tracking (TOF-PEPT)

Positron Emission Particle Tracking (PEPT) is a technique that can measure the location and speed of a particle labeled with a positron-emitting radioisotope. The most widely used PEPT algorithm was developed at the University of Birmingham using a modified PET scanner/gamma camera [70, 71, 72].

The algorithm determines the location of the radioactive particle in an iterative manner. As mentioned in Chapter 2, a radioactive particle decays by emitting a positron, and the emitted positron combines with an electron shortly, resulting in an annihilation event. The emitted pair of gamma rays is electronically identified as a coincidence event if detected within the coincidence timing window. A trajectory connecting the two detectors forms a line-of-response (LOR). In an ideal situation with a single radioactive particle with no radioactive background, all the trajectories should meet in a space that uniquely determines the location of the particle. However, in actuality, many LORs do not correspond to the true origin of events but instead get corrupted by delayed events, Compton scattering, etc. Nevertheless, all the uncorrupted LORs are expected to meet at a point in the space, which gives an estimate of the particle location. The PEPT algorithm has been designed to obtain the location estimate by iteratively discarding a fraction of corrupted LORs based on a distance threshold and determining the point in space for which the sum of distances to the remaining set of LORs is minimized.

The application of the PEPT algorithm has mostly been limited to industrial applications. Other than preliminary works[93, 67, 68, 69], a variant of the algorithm (PeTrack) was presented for use in external beam radiotherapy and cardiac PET studies [57, 14]. The method was based solely on external markers and required an initial estimate of the marker position along with several user-defined threshold parameters. In this work, the TOF-PEPT algorithm [94] has been proposed and studied as a new and unique data-driven approach by modifying the original PEPT algorithm to allow its implementation in a completely data-driven manner and incorporating the time-of-flight (TOF) information into the original

algorithm as an additional weighting factor to further improve the tracking performance. The next subsections provide a detailed description of the proposed algorithm.

3.1.1 Overview of the TOF-PEPT Algorithm

Let $L_i = L_1, L_2, \ldots L_N$ denote N number of LORs originating from a positron-emitting particle. If $A_i(x_A, y_A, z_A)$ and $B_i(x_B, y_B, z_B)$ denote the Cartesian coordinates of two detector pairs, then a LOR L_i, is represented by $\overrightarrow{A_iB_i}$. Furthermore, let $\delta(L_i, P)$ denote the perpendicular distance of L_i from the unknown (parameterized) particle location $P(x, y, z)$. An expression of $\delta(L_i, P)$ can then be derived by:

$$\delta(L_i, P) = \frac{|\overrightarrow{A_iB_i} \times \overrightarrow{A_iP}|}{|\overrightarrow{A_iB_i}|} \tag{3.1}$$

Let D denote the sum thereof over all LORs considered, which can be mathematically expressed as:

$$D = \sum_{L_i \in \Omega} \delta(L_i, P) \tag{3.2}$$

An estimate of the location, P_m, is then obtained by minimizing D.

$$P_m = argmin_P(D) \tag{3.3}$$

This can be achieved by solving following set of equations:

$$\frac{\partial D}{\partial x} = 0; \;\; \frac{\partial D}{\partial y} = 0; \;\; \frac{\partial D}{\partial z} = 0 \tag{3.4}$$

The proposed algorithm incorporated a weighting factor based on the TOF information. As mentioned in the previous Chapter 2, TOF information provides an estimate of the true annihilation location i.e., from where LORs were originated. If A and B are two opposing detectors, and t_A and t_B are the time taken for γ rays to reach them, respectively, then TOF provides the time difference $\Delta t = t_A - t_B$. From this, the distance of the annihilation origin from the LOR center can be calculated as follows [89]:

$$\Delta d = \frac{c\Delta t}{2} \tag{3.5}$$

where c is the speed of light, $3 \times 10^{10} cm/s$. The location of the TOF bin can be estimated along a LOR by using the following equations, considering TOF is positive toward A detector:

$$TOF_{bin} = \begin{cases} \frac{d}{2} - \Delta d & \text{if } \Delta t \geq 0 \\ \frac{d}{2} + \Delta d & \text{if } \Delta t < 0 \end{cases} \tag{3.6}$$

where d is the distance between detectors A and B. The location estimate, however, is not perfect, and there is an uncertainty associated with it that follows a Gaussian distribution centered on TOF_{bin} [89]. The corresponding full width at half maximum (FWHM) and standard deviation σ_T, depends on the timing resolution of the scanner system used. Figure 3.1 illustrates the geometry.

Let P_i denote the estimate of the TOF_{bin}, i.e. annihilation location for L_i. The weighting factor was then calculated as follows:

$$w_i \triangleq 1 + \max\left(0, 1 - \left(\frac{||P_i - P_m||}{\sqrt{2}\sigma_T}\right)^2\right) \in [1, 2] \tag{3.7}$$

The weighting factor was based on a first-order Taylor series expansion of a Gaussian function. It was chosen in a way so that a higher emphasis (up to 2x) was placed on LORs for which its annihilation location was close to the current estimate of P_m than LORs that are further away (up to 1x). Thus the weighting factors were in the range of [1, 2]. In this work, $\sigma_T = 6.53cm$ was used based on an idealized individual LOR timing resolution of 512 ps.

Thus, Equation 3.9 was modified as follows to incorporate the weighting, as well as solve the problem analytically:

$$D_w = \sum_{L_i \in \Omega} w_i \delta(L_i, P)^2 \tag{3.8}$$

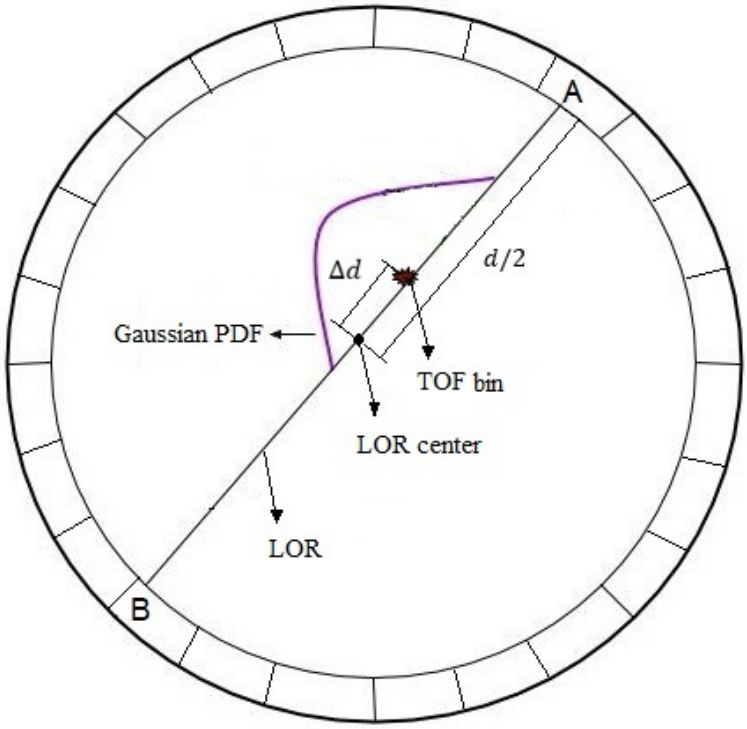


Figure 3.1: Illustration of the concept of time-of-flight (TOF): The true annihilation position lies somewhere along the line of response (LOR) joining the detector pair. If d is the distance between detector A and B, then Δd from TOF information allows estimating the true annihilation position with an uncertainty that follows a Gaussian distribution centered on the TOF bin with full width at half maximum (FWHM) corresponding to the scanner system timing resolution.

At the end of each iteration, the mean perpendicular distance of all LORs from P_m can be calculated as:

$$\overline{\delta}(P_m) = \sum_{L_i \in \Omega} \frac{\delta(L_i, P_m)}{N} \tag{3.9}$$

LORs which were farther than one standard deviation away from the mean distance of the set were discarded. The iteration stopped when the remaining number of LORs dropped below the number of LORs for which the TOF-based annihilation locations were within a sphere of radius σ_T surrounding the particle location estimate.

3.1.2 Implementation Details of the TOF-PEPT Algorithm

The whole listmode data was first chronologically divided into multiple frames of predefined frame duration. If T denotes the total scan duration, and t_s denotes the duration of each frame, then $M = T/t_s$ denotes the number of frames that were created. Each frame contained N number of LORs. The information related to which detector pair was hit by a LOR and its corresponding TOF was extracted from the listmode data. With the calculation of the Cartesian coordinates of the detector pair and TOF center, the TOF-PEPT algorithm was applied to the set of LORs within each time frame. Being applied to the M number of frames one by one, the algorithm then localized the object of interest at each sampling time point.

Lesion-specific motion detection required limiting the computation within a selected ROI, as millions of LORs originated from many locations within the patient are detected by the PET scanner. A user-based selection of ROI was thus incorporated to determine which LORs were to be taken into account during the computation. MATLAB [58] was used to design a graphical user interface (GUI) that allowed visualization of image volumes in DICOM (Digital Imaging and Communications in Medicine) format and select a box-shaped ROI around the lesion or point source. DICOM images were reconstructed using Biograph mCT (Siemens Healthineers, USA) software with a pixel size of 4.0728 mm $\times$ 4.0728 mm and a slice thickness of 2.0084 mm. The size of ROIs was chosen to be a few pixels (approximately 2-4) larger on each side than the object of interest by visual inspection. The coordinates of the ROI drawn were then mapped to the actual scanner geometry.

With the selected ROI, the algorithm determined whether the shortest/perpendicular distance of a LOR from the center of the ROI lies within the circumradius of the ROI. For LORs, which were farther away, were discarded before the beginning of the iterative search by the algorithm. Thus a candidate set of LORs was chosen based on the ROI selection to keep the motion tracking limited to the object of interest within the ROI. The whole implementation process has been explained by means of Figure 3.2 and Algorithm 1 outlined below.

Estimated particle location as a function of time provided the movement information along Left-Right (LR), Anterior-Posterior (AP), and Superior-Inferior (SI) directions respectively denoted by X, Y, and Z. Motion signal for the duration of the entire scan was thus obtained along each direction. The motion signals were then smoothed by applying a low pass filter. The cutoff frequency of the filter was determined from the frequency corresponding to the peak magnitude [80]. This varied from patient to patient and ranged from 0.167 Hz to 0.3 Hz for the clinical data studied.

3.2 Performance Evaluation

3.2.1 Validation and Evaluation

To evaluate the performance of the proposed motion detection algorithm against a gold standard, a clinically approved external pressure sensor band (Anzai Medical, Japan) was used, which provides respiratory motion information along with the information related to inspiration and expiration peaks. To further compare against state-of-the-art data-driven methods, the COM algorithms proposed by Ren *et al.*[80] were implemented. To ensure a fair comparison, similar ROI selection and restriction methodology were used for both COM and the TOF-PEPT algorithm. The comparisons were performed against COM using TOF information (referred to as TOF-COM) and COM without using TOF (referred to as COM). The former used an average of TOF center/bin within the FOV, whereas the latter made use of the average of LOR centers as the estimate of the particle location.

Algorithm 1 Locate an object of interest frame by frame

Input: User defined ROI
Input: M number of frames
Output: Object coordinates P_m at each sampling frame
  Initialize: $P_m \leftarrow$ Center of ROI
  Initialize: $N \leftarrow$ Number of initially selected LORs
  $\sigma_T \leftarrow$ Gaussian width corresponding to the system timing resolution
  $N_{TOF} \leftarrow$ Number of LORs for which their TOF center lies within a sphere of radius σ_T
  for $Frame = 1, 2, \ldots M$ **do**
    while $N \geq N_{TOF}$ **do**
      $D \leftarrow 0$
      for $i = 1, 2, \ldots N$ **do**
        $P_i \leftarrow$ TOF center estimate
        $L_i \leftarrow$ Line joining detector coordinates
        $\delta_i \leftarrow$ Perpendicular distance as a function of L_i and unknown location P
        $w_i \leftarrow$ Weight factor
        $D \leftarrow D + w_i(\delta_i)^2$
      end for
      Solve $\frac{\partial D}{\partial x} = 0$, $\frac{\partial D}{\partial y} = 0$, $\frac{\partial D}{\partial z} = 0$ to refine the estimate P_m
      $\overline{\delta} \leftarrow$ Mean perpendicular distance of all LORs from P_m
      $\sigma \leftarrow$ Standard deviation of perpendicular distances of all LORs from P_m
      Discard LOR if $\delta_i > \overline{\delta} + \sigma$
      $N \leftarrow$ Number of remaining LORs
    end while
  end for

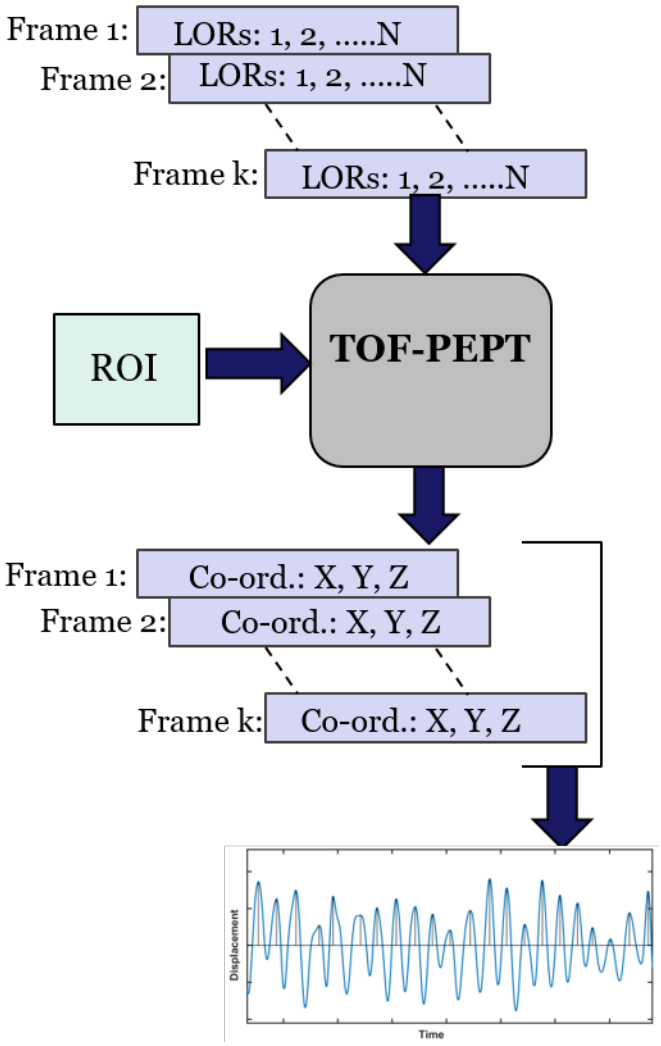


Figure 3.2: The flow diagram of the TOF-PEPT algorithm: Multiple frames of short duration are created and passed to the algorithm along with the user defined region-of-interest (ROI). The algorithm provides the location coordinate of the object within ROI for each sampling frame. The location as a function of time, provides an estimate of the actual motion.

The analysis was conducted by calculating: Pearson correlation coefficient [38], root mean square error (RMSE), and mean shift in time corresponding to the inspiration peaks. To conduct the comparison with the Anzai band, the tracked motion signals were resampled using the built-in *resample* function of MATLAB at the Anzai band sampling rate (20 ms). The correlation coefficient and RMSE were computed for the whole duration signals. The mean shift in time was evaluated by taking the average of the difference in time corresponding to inspiration peaks for each cycle of the detected and the Anzai band signals. Each shift was normalized by the corresponding cycle duration of the band data.

An additional quantitative comparison study was performed to evaluate the performance of the proposed TOF-weighting against the TOF-based LOR exclusion approach proposed in the PeTrack paper [57, 14]. Their approach discards LORs for which the distances between the TOF-based annihilation locations to the current estimate of the particle location exceed a user-defined threshold.

From motion signals detected by each method, respiratory gating was implemented to create motion compensated images and evaluate the performance. Amplitude-based gating was used and adapted from [31] in a way that sorted the data by keeping all events that lie within the baseline amplitude to 35% of the peak amplitude for each individual cycle. The objective was to create reconstructed images that correspond with the phase of the breathing cycle where the patient spent the greatest amount of time statistically (usually the expiration phase for most patients). With the chosen set of events, new listmode data was created and rebinned into sinograms using Biograph mCT software and the e7 processing tooling. The sinograms were then processed by the manufacturer provided OSEM algorithm (with 3 iterations, 24 subsets, 5 mm Gaussian filter, UltraHD (TOF+PSF)) found on the mCT Flow PET/CT platform to reconstruct image volume. The images were reconstructed with the same resolution as the DICOM images mentioned previously. To quantitatively assess the impact of motion correction, ROIs were drawn across several regions on the gated images using RadiAnt [59] and Inveon Research Workplace visualization software (Siemens Healthineers, USA).

Lastly, the two key user-defined parameters, namely, frame duration and ROI size, which control how many and which LORs are to be kept for each frame, respectively, were varied,

and the sensitivity and robustness of the TOF-PEPT algorithm were assessed with respect to their changes. The assessment focused on how the evaluative metrics described above varied with these parameters. Five different frame duration of 200, 400, 600, 800, and 1000 ms were used. The ROI was varied in size by controlling the radius of the circumscribing sphere using scaling factors of 1.5, 2, and 2.75.

3.2.2 Data

The performance of the algorithm was evaluated by means of two phantom and six clinical studies. In order to conduct the first phantom study, a respiratory phantom consisting of an elliptical disc was used, which rotated at a frequency of 0.18 Hz. A single radioactive point source was placed on the top of its platform, and an Anzai band was wrapped across the point source. The phantom had its larger movements in the Y direction with approximately 9-10 mm displacement from the initial position.

A servo motor with a small arm attached to it was used to conduct the second phantom study. A radioactive point source was attached to the arm. An Arduino board [45] was used to control the motor, and it was programmed to let the arm move in an up and down motion in the Y direction for 15 minutes at three different frequencies, namely, 0.167, 0.25, and 0.5 Hz with three different displacement ranges of about, respectively, 6, 3, and 1.5 mm and for 7.5, 5, and 2.5 minutes. No Anzai band was used for this study.

Clinical studies were performed on a 64-slice Biograph mCT Flow PET/CT with 64-bit listmode data acquisition. Patients were recruited in compliance with the University of Tennessee Graduate School of Medicine Institutional Review Board approved protocol (#3941). Three scenarios each with two clinical studies were examined. Clinical Studies 1 and 2 were performed using external radioactive point sources attached to the abdomen along with the Anzai band. Clinical Studies 3 and 4 made use of a single point source each but no Anzai band. For Clinical Studies 5 and 6, no external source was used, and the algorithm tracked internal motion directl from the left cardiac ventricle and a 1.5 cm3 lesion in the lung area, respectively.

For tracking the motion from any kind of region or point source, the algorithm basically confined its search within the selected ROI, and the location estimate was done by minimizing

the sum of squared distances to all LORs within the ROI. In order to estimate motion from multiple point sources/internal regions, individual ROI was selected for each, and motion estimation was performed independently by using the set of LORs within that ROI. Table 3.1 provides additional details for all studies.

To make the external point sources, zeolite beads (approximately 2 mm in diameter) were soaked in an 18F-fluorodeoxyglucose solution for nearly 10 minutes. Clinical Study 1 was conducted with three point sources, and they were placed on the center, left, and right side of the abdomen of the patient, approximately 13-18 cm apart from each other. Rest of the clinical studies for which a single point source was used, it was placed on the center of the abdomen.

3.2.3 Qualitative Evaluation of Motion Signals Estimated by TOF-PEPT against Other Methods

Phantom Study 1: With Point Source and Anzai Band

Figure 3.3 shows the detected motion signal by the proposed TOF-PEPT method along with the Anzai band for Phantom Study 1. It can be seen that the two signals correlated almost perfectly well for the entire scan duration. Triggers/timestamps that were provided by the Anzai band to indicate inspiration and expiration peaks aligned well with the TOF-PEPT driven signal peaks for the entire duration of the scan. Additionally, the range of movement could be realized with the data-driven approach, and it matched satisfactorily with the true range of movement.

Clinical Studies 1 and 2: With Point Sources and Anzai Band

Figure 3.4 shows the comparison of the motion signal estimated by the proposed TOF-PEPT method and Anzai band signal for Clinical Study 1. Between 20-30 second and 65-90 second, the patient either coughed or moved. That movement is apparent in both signals at around a similar time. In fact, the signal provided by the Anzai band appeared flattened, possibly owing to exceeding the threshold pressure of the sensor. On the contrary, it was possible to monitor the signal with the actual movement information by the data-driven TOF-PEPT.

Table 3.1: Detailed description of the experimental studies

Study	Anzai Band Used	Number of External Point Sources	Radioactivity (approx.) kBq/mL	Scan Duration (Minutes)	ROI Circumradius (mm)
Phantom Study 1	Yes	1	Point Sources: 74-185	3	42.00
Phantom Study 2	No	1		15	17.59
Clinical Study 1	Yes	3		5	26.02, 21.71, 24.43
Clinical Study 2	Yes	1		3	24.28
Clinical Study 3	No	1		3	23.32
Clinical Study 4	No	1		3	31.10
Clinical Study 5	No	–	–	3	57.28
Clinical Study 6	No	–	Lesion: 10-15	3	24.39

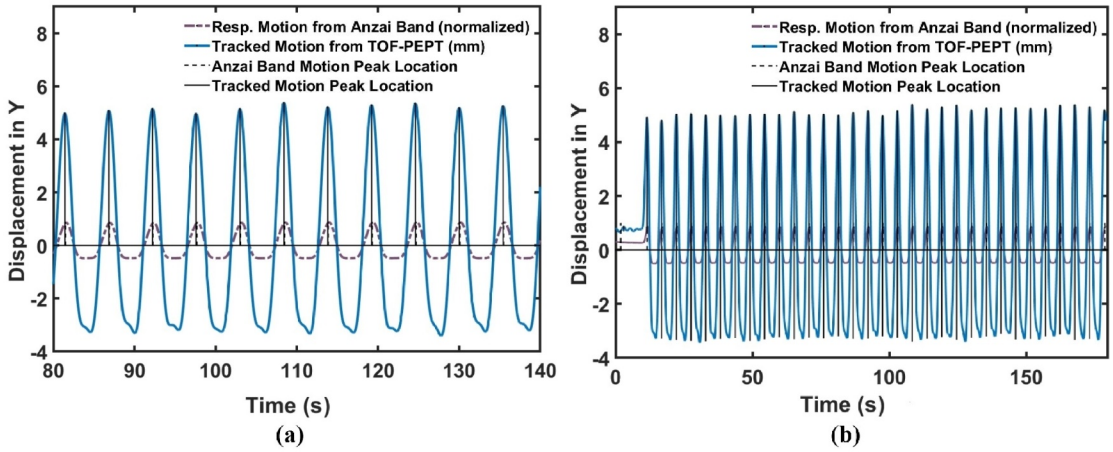


Figure 3.3: For Phantom Study 1, comparison of the motion signals estimated by the TOF-PEPT versus the Anzai band: (a) Displacement in the Y direction for a 60-seconds time frame; (b) Displacement in the Y direction for the entire 3-minute scan. The vertical lines across the peak of the signals show the inspiration peaks. It can be seen that the phase, displacement, and inspiration peak time corresponded well. The Anzai band data magnitude was normalized by its maximum value to aid visualization. The TOF-PEPT data is shown in the original scale.

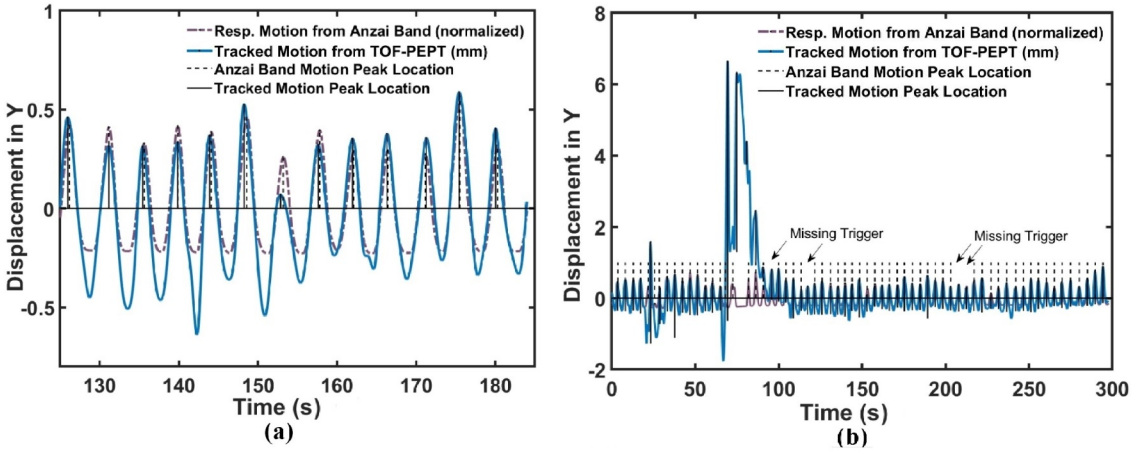


Figure 3.4: For Clinical Study 1, comparison of the motion signals estimated by the TOF-PEPT versus the Anzai band: (a) Displacement in the Y direction for a 60-seconds time frame; (b) Displacement in the Y direction for the entire 5-minute scan. The vertical lines across the peak of the signals show the inspiration peaks. It can be seen that the phase, displacement, and inspiration peak time corresponded well except for a few places where there were involuntary movements of the patient and issues of missing triggers from the Anzai band. The Anzai band data magnitude was normalized by its maximum value to aid visualization. The TOF-PEPT data is shown in the original scale.

Besides, the Anzai signal missed information related to inspiration and expiration peaks in a few places, as marked in the figure.

Phantom Study 2: With Point Source and No Anzai Band

For phantom study 2, the servo motor moved its arm along with the point source attached to it up and down at three different frequencies with three different amplitude ranges. The movement of the arm was estimated by both TOF-PEPT and COM algorithms. The detected motion signals were analyzed in the time and frequency domain. Figure 3.5 shows the time and Fourier domain analysis of the signals. The three user defined frequencies were correctly captured in the frequency response of the signal provided by the TOF-PEPT algorithm, with the signal exhibiting almost no noisy fluctuations. Although the TOF-COM algorithm captured the three frequencies as well, the signal appeared noisier both in time and frequency domain. With the COM algorithm, it was possible to detect only the two lowest frequencies. This can be seen to be pronounced during the last segment, where the phantom was made to move with the smallest amplitude range. Signals being noisier with both variations of the COM algorithms, the true range of displacement was successfully captured only by the TOF-PEPT for the entire duration.

Clinical Studies 3 and 4: With Point Source and No Anzai Band

Figure 3.6 shows a qualitative comparison of the estimated motion signals by different methods for Clinical Study 3. In this case, as well, the motion estimate obtained by the TOF-PEPT algorithm exhibited less noise compared to the COM algorithms. This is evident in both the time and Fourier domain signals. Studies with Clinical Study 4 provided similar results.

Clinical Studies 5 and 6: Lesions/Internal Regions and No Anzai Band

For Clinical Studies 5 and 6, motion signals were detected using data from internal regions selected within the patient, namely, left cardiac ventricle (LV) and an approximately $1.5cm^3$ lesion in the lung area, respectively. The intent was to realize the ability of the TOF-PEPT algorithm to detect internal, localized motion. Figure 3.7 and 3.8 shows the extracted signals.

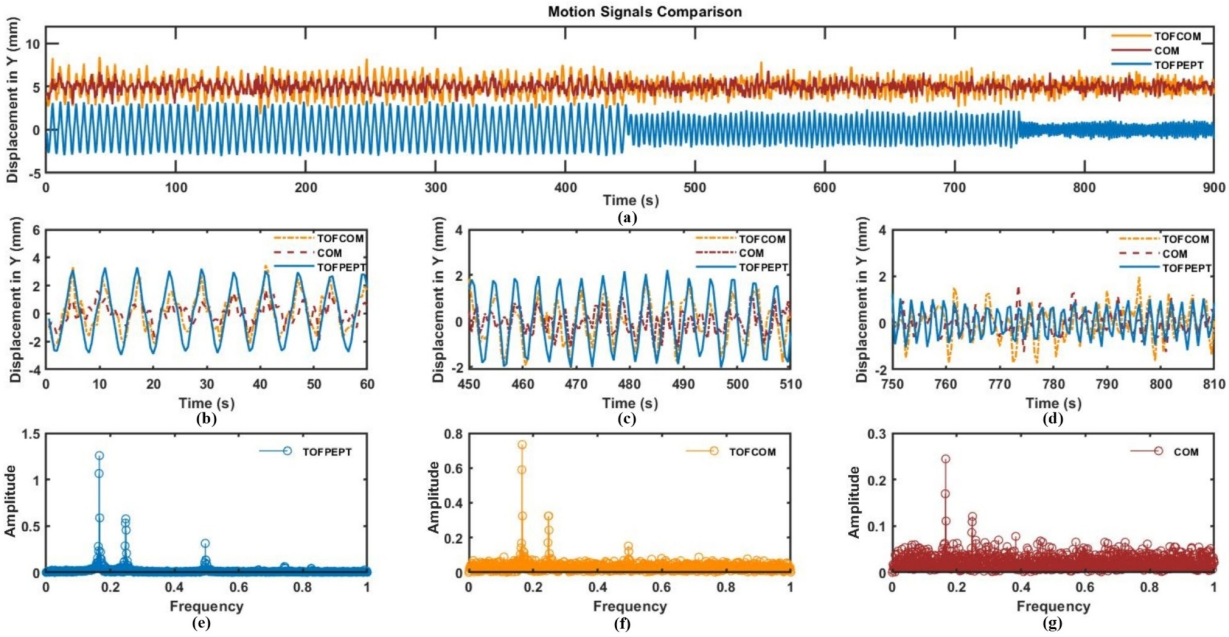


Figure 3.5: For Phantom Study 2, comparison of motion signals estimated by TOF-PEPT, TOF-COM, and COM: (a) Time domain plots for the full 15-minute scan. Baseline shift was added to the TOF-COM and COM signals to aid visualization; (b)-(d) Time domain plots with frames of 60-seconds for the three different frequencies and amplitude ranges; (e)-(g) Fourier domain plots for the full 15-minute scan.

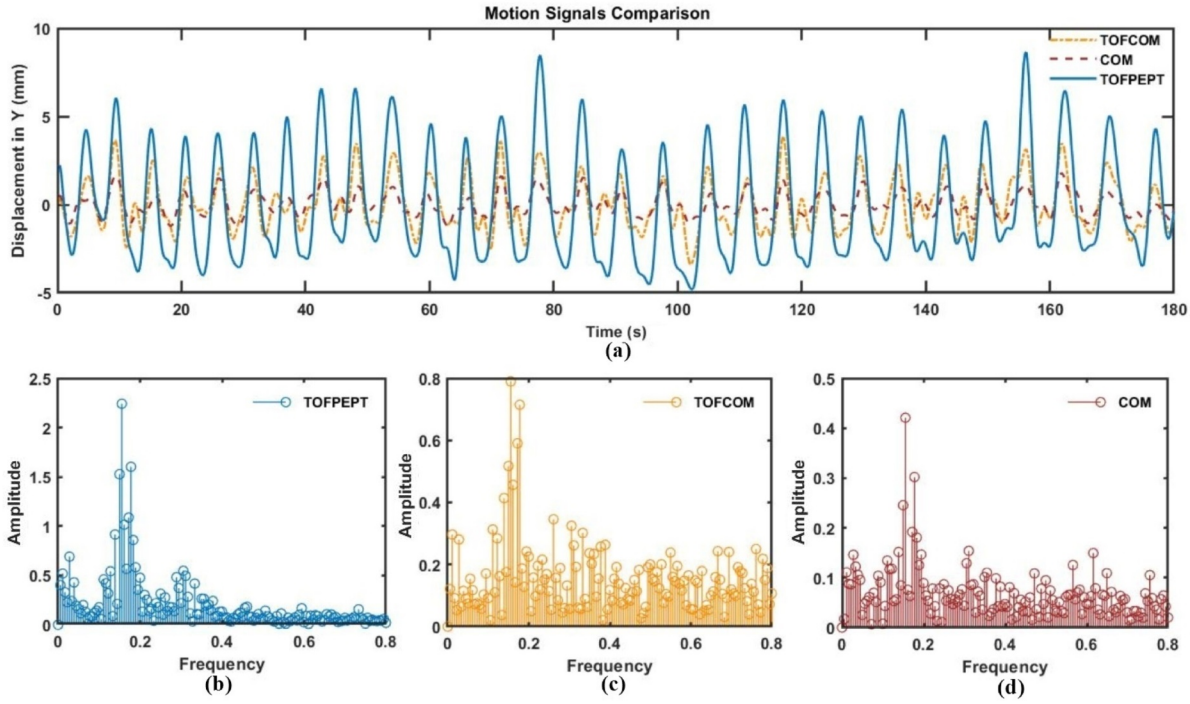


Figure 3.6: For Clinical Study 3, comparison of motion signals estimated by TOF-PEPT, TOF-COM, and COM: (a) Time domain plots; (b)-(d) Fourier domain plots.

The figures also show signals in the Fourier domain (3.7, 3.8 (b),(c)). All these results were compared against the COM algorithms.

For the left cardiac ventricle region tracking, the cardiac motion information was captured as well by the TOF-PEPT algorithm along with the respiratory motion. This can be realized from the two dominant frequency components at around 0.3 and 0.7 Hz in Figure 3.7 (b), which shows the frequency response. The cardiac motion information having higher frequency component got obscured in noise for the signals provided by the TOF-COM and COM algorithms.

For the lesion tracking, it was possible to obtain a very clean motion signal by the TOF-PEPT algorithm. Signals provided by the TOF-COM and COM algorithms were noisier with respect to the TOF-PEPT. Motion signal detected by only COM, in particular, was extremely noisy to extract any information out of it.

3.2.4 Quantitative Evaluation of Motion Signals Estimated by TOF-PEPT against Other Methods

A quantitative comparison of detected motion signal using the TOF-PEPT and state-of-the-art COM based data-driven techniques was made with the Anzai band, which is the current clinical standard. Pearson correlation coefficient, Root Mean Squared Error (RMSE), and mean shift in peak locations were used to perform the comparison. Table 3.2 shows the results for the motion signals tracked from the point source used for each study. A higher correlation coefficient and a lower RMSE and mean shift in peak locations indicate better performance. It can be seen that the correlation coefficients were the highest for the detected motion signals by the TOF-PEPT algorithm for every study, with the lowest RMSE and mean shift as well. On average, correlation coefficients were 13.5% and 38.7% higher than the TOF-COM and COM algorithms, respectively.

Table 3.3 provides a quantitative comparison of the implementation of the TOF-PEPT with/without TOF-based weighting (proposed in this work) and with/without TOF-based LOR exclusion (proposed in the PeTrack paper). The comparisons were made by the analysis of correlation with the Anzai band. It was observed that the method implemented with

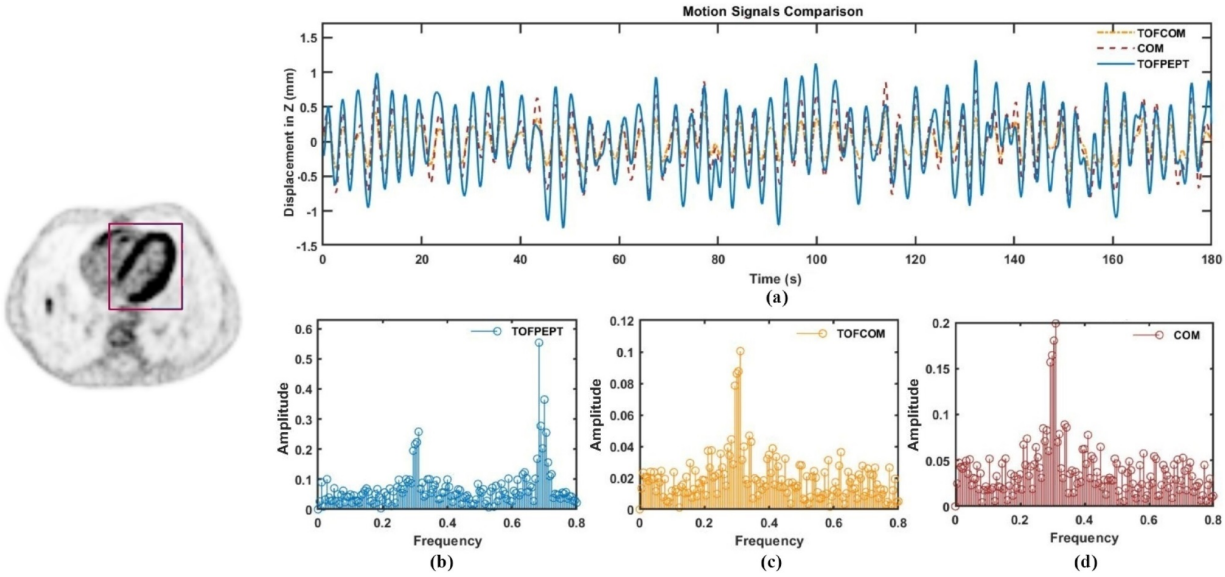


Figure 3.7: For Clinical Study 5, comparison of motion signals estimated by TOF-PEPT, TOF-COM, and COM from left cardiac ventricle: (a) Time domain plots of the respiratory motion. To separate the respiratory motion from the cardiac motion, a low pass filter with a cutoff frequency of 0.5 Hz was applied; (b)-(d) Fourier domain plots

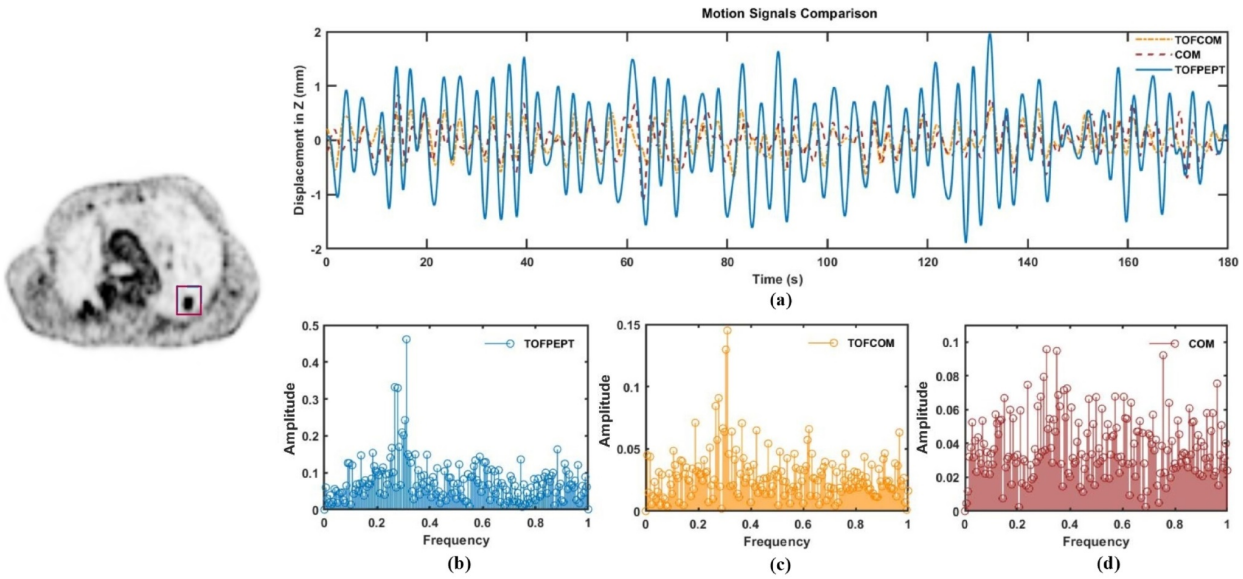


Figure 3.8: For Clinical Study 6, comparison of motion signals estimated by TOF-PEPT, TOF-COM, and COM from an internal lesion in the lung area: (a) Time domain plots; (b)-(d) Fourier domain plots

Table 3.2: Quantitative comparison of detected motion signals by the proposed TOF-PEPT method against the gold standard external pressure sensor band (Anzai) and the state-of-the-art data-driven method by center-of-mass (COM)

		Phantom	Clinical Study 1(1)	Clinical Study 1(2)	Clinical Study 1(3)	Clinical Study 2
Correlation Coefficients	TOF-PEPT	**0.995**	**0.966**	**0.94**	**0.947**	**0.976**
	TOF-COM	0.992	0.834	0.788	0.876	0.881
	COM	0.99	0.892	0.509	0.839	0.655
RMSE	TOF-PEPT	**0.035**	**0.096**	**0.116**	**0.142**	**0.092**
	TOF-COM	0.078	0.148	0.183	0.217	0.144
	COM	0.056	0.138	0.203	0.228	0.0.229
Mean Shift in peak locations	TOF-PEPT	**0.002**	**0.017**	**0.021**	**0.026**	**0.019**
	TOF-COM	0.006	0.061	0.072	0.043	0.045
	COM	0.008	0.046	0.124	0.043	0.0.077

TOF-based weighting and no TOF-based LOR exclusion resulted in the highest correlation coefficient values (shown in bold). The reverse scenario yielded the worst performance. For the low activity point source (2) used in Clinical Study 1, the TOF-based weighting allowed a 4.6% increase in correlation coefficient while the TOF-based LOR exclusion resulted in a 7.9% decrease in correlation. For PET scanners with better timing resolution i.e. lower uncertainty in TOF measurements, the improvements can be expected to be even better with TOF-weighting.

3.2.5 Qualitative and Quantitative Evaluation of Motion Compensated Image

In order to evaluate the motion compensated data, maximum standardized uptake values (SUV) were measured across lesions in the reconstructed images. Motion blurring results in an underestimation of SUV across lesions as mentioned earlier, which is expected to get resolved in a motion compensated image.

Figure 3.9 presents images reconstructed without any gating and with amplitude-based gating using estimated motion signals by different methods for Clinical Study 1. It can be seen from the gated images that the uptake increased across the marked lesion, with the maximum SUV being the highest in the TOF-PEPT image. The SUV was higher than the SUVs measured in the ungated image by about 16% and the other gated images by 4-10%. Across several other lesions, there was an increase of 14-39% in maximum SUV in the TOF gated data in comparison with the ungated data.

A similar comparison of ungated and gated reconstructed images was performed using lesion-specific motion information, and Figure 3.10 shows the comparison for Clinical Study 6. In the original ungated image, the lesion can be seen motion blurred and more elongated as a result. The motion gated images resulted in reduced motion blurring. This minimization of motion impact was higher with the TOF-PEPT method, as can be seen from the arrow marked lesion. The maximum SUV was higher in the gated image by the TOF-PEPT than the ungated, COM, and TOF-COM results by 8.7%, 7.1%, and 5.6%, respectively. This improved performance by TOF-PEPT can also be realized from the Gaussian-fitted line

Table 3.3: Performance of the proposed method under different conditions

TOF-based LOR Weighting	TOF-based LOR Exclusion	Correlation Coefficient				
		Phantom	Clinical Study 1 (1)	Clinical Study 1 (2)	Clinical Study 1 (3)	Clinical Study 2
No	No	0.9946	0.961	0.898	0.932	0.966
No	Yes	0.9945	0.958	0.827	0.896	0.941
Yes	**No**	**0.9946**	**0.966**	**0.940**	**0.947**	**0.976**
Yes	Yes	0.9945	0.959	0.839	0.902	0.944

profiles for all four methods. SUV was the highest, and FWHM across the lesion was the lowest for the TOF-PEPT based gated image. There was a 12.17% reduction in the measured FWHM in comparison with the ungated image.

3.2.6 Analysis of the Algorithm Parameters

Two user-defined parameters, namely, ROI size and frame duration, have been used in the proposed algorithm. These parameters were varied, and the resultant motion signals estimated by the TOF-PEPT were compared against the Anzai band. Figures 3.11 and 3.12 show the corresponding plots of the correlation coefficient, RMSE, and normalized mean shift in peak locations, respectively, for varying ROI size and frame duration. ROI size was varied by varying the circumradius. With increasing ROI radius, the performance of the TOF-PEPT algorithm can be seen to degrade. It is suspected that for too large an ROI, more background LORs are considered that do not originate from the point source/lesion. Thus noise is introduced. As for the frame duration, both short and long frame duration negatively impacted the performance. The former is likely a consequence of not having enough LORs to track the object of interest, while the latter is most likely due to making the sampling rate too low. When both the parameters were varied simultaneously, the trends appeared to be consistent.

Tables 3.4 and 3.5 show the number of LORs and iterations used by the algorithm with this varying parameter setup, namely ROI scaling factor and frame duration, respectively. This is shown only for a single point source tracking for Clinical Study 1 and 2. Both for increasing ROI size and frame duration, the number of LORs within each frame increased. Iteration count remained unaffected with the increase in frame duration and increased for the increase in ROI size. The trend remained the same for the other motion estimations as well.

3.3 Summary

The TOF-PEPT algorithm proposed here is a new data-driven approach for the estimation of motion during PET. It facilitates the motion detection of external point sources as well

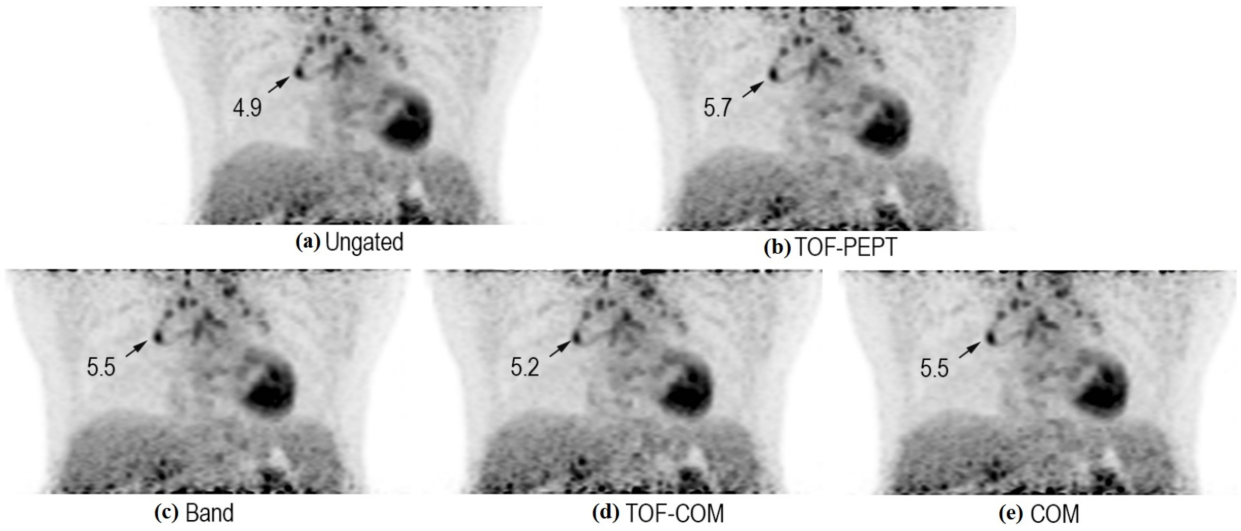


Figure 3.9: For Clinical Study 1. comparison of the (a) ungated and (b)-(e) amplitude gated images created by using motion signals estimated by TOF-PEPT, Anzai band, TOF-COM, and COM, respectively. Increased maximum SUV and sharper image quality are observed in the gated image by the TOF-PEPT algorithm.

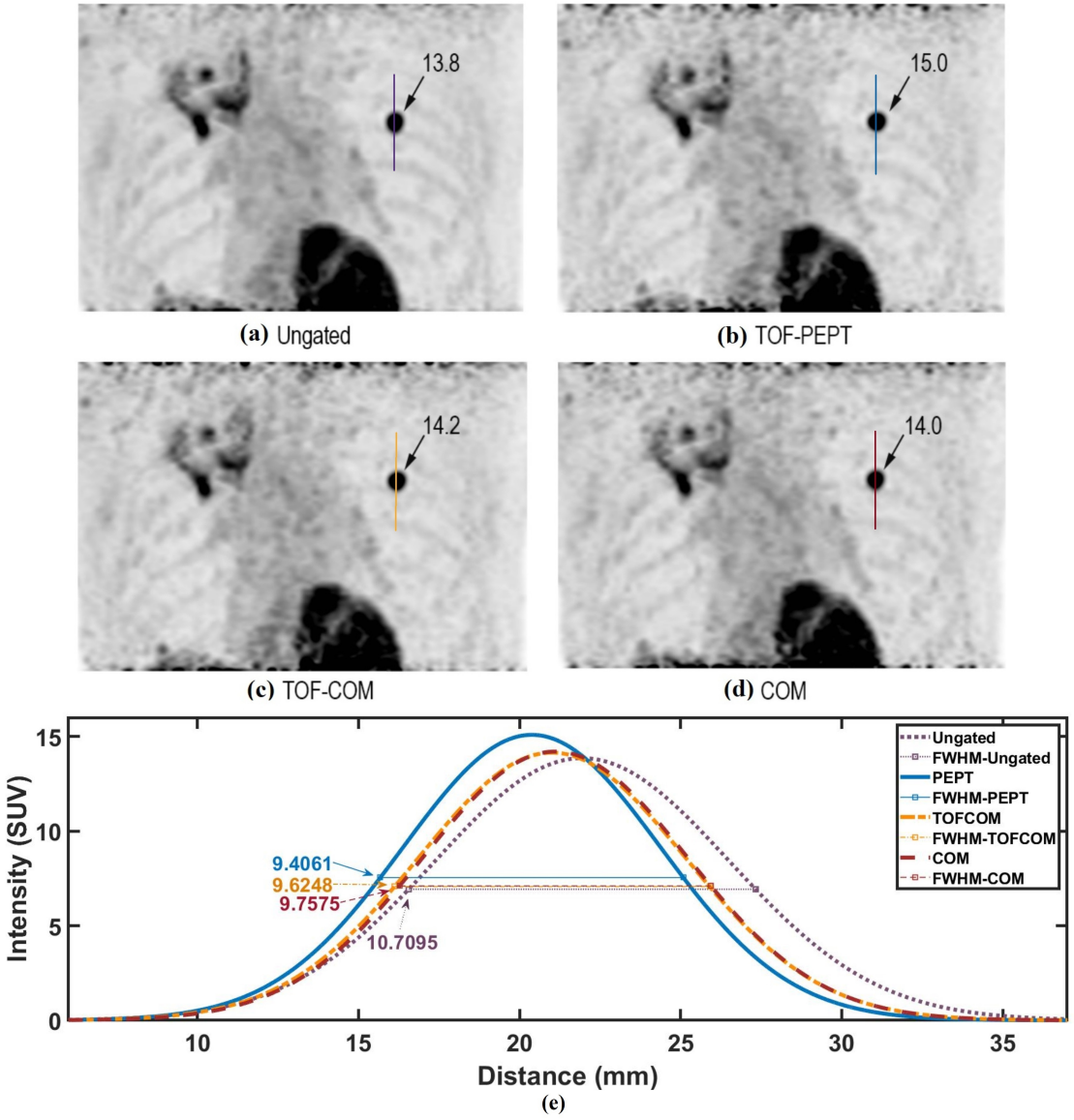


Figure 3.10: For Clinical Study 6, comparison of the (a) ungated and (b)-(d) amplitude-gated images created by using motion signals estimated by TOF-PEPT, TOF-COM, and COM, respectively. The lesion area, along with its maximum SUV is indicated by an arrow in each image. (e) Result of Gaussian-fitted line profile data (a-d vertical lines). SUV can be seen to be the highest, and FWHM can be seen to be the lowest in the gated image by the TOF-PEPT algorithm.

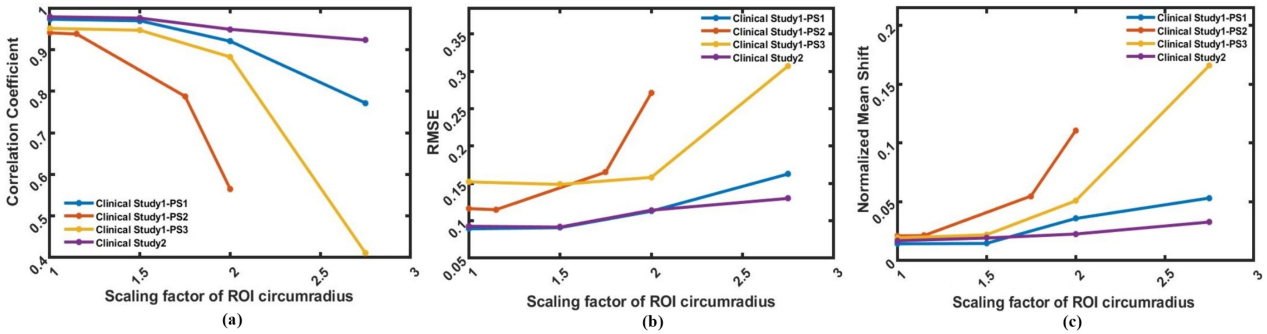


Figure 3.11: Quantitative comparison of (a) correlation coefficient, (b) RMSE, and (c) normalized mean shift in peak location time for varying scaling factors of ROI circumradius.

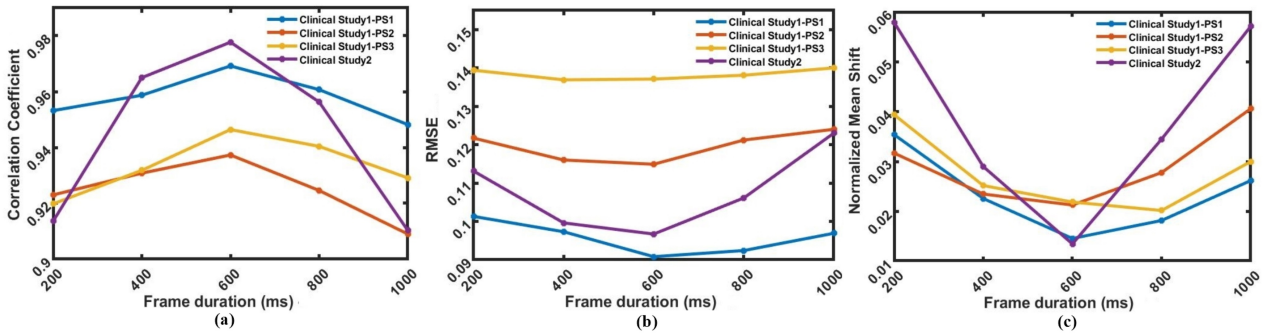


Figure 3.12: Quantitative comparison of (a) correlation coefficient, (b) RMSE, and (c) normalized mean shift in peak location time for varying frame duration.

Table 3.4: Number of line-of-responses (LOR)s and iterations used by the algorithm with varying sizes of region-of-interest (ROI)

ROI Scaling Factor	Clinical Study 1		Clinical Study 2	
	Number of LORs (approx.)	Number of Iterations	Number of LORs (approx.)	Number of Iterations
1	700	5	3000	4
1.5	1000	7	4000	4
2	1800	8	5000	5
2.75	2400	9	8000	7

Table 3.5: Number of line-of-responses (LOR)s and iterations used by the algorithm with a varying frame duration

Frame Duration	Clinical Study 1		Clinical Study 2	
	Number of LORs (approx.)	Number of Iterations	Number of LORs (approx.)	Number of Iterations
200	450	7	700	4
400	800	7	2000	4
600	1200	7	3000	4
800	1500	7	4000	4
1000	2000	7	5000	4

as lesions and other internal regions directly from raw listmode data. The performance of the algorithm was validated against the current clinical standard Anzai band along with two other state-of-the-art algorithms, namely COM and PeTrack.

It was possible to estimate motion from the TOF-PEPT algorithm with timing resolutions ranging from 200 ms to 1000 ms. With respect to the quantitative and qualitative measures studied in this work, the estimated motion signals were found to correlate well with the band while outperforming the state-of-the-art COM algorithms studied here to varying degrees. For the clinical studies, the motion signals estimated by the TOF-PEPT correlated with the band with correlation coefficients in the range of 0.94-0.97. In contrast, motion signals by the COM algorithms correlated with the band with correlation coefficients in the range of 0.51-0.92. While comparing the TOF-PEPT and the COM methods with no Anzai band as a reference, the motion signals derived by the TOF-PEPT algorithm suffered less noise than the COM estimates. It was realized in every case both with time and Fourier domain analysis. Furthermore, the range of motion amplitude observed in the signals detected by the TOF-PEPT algorithm was closer to the actual range of movement than the signals detected by the TOF-COM and COM algorithms.

Analysis of the motion compensated images showed that the signals provided by the TOF-PEPT algorithm performed better motion compensation compared to the COM algorithms. With all the gated images being subjected to similar noise characteristics and count loss by the use of amplitude-based gating method, gated images based on the TOF-PEPT algorithm resulted in a 6-9% increase in maximum SUV measured in the motion impacted lesion.

The robustness of the algorithm with changing parameter setup was studied with respect to ROI size and frame duration. Up to a 2-fold increase in the ROI size, the algorithm-provided motion signals correlated with the band with correlation coefficients in the range of 0.88-0.97 except for one case with the low activity point source. For frame duration of 200 ms to 1000 ms, the motion signals provided by the TOF-PEPT algorithm were found to be in good correlation with the band-provided signals with correlation coefficients in the range of 0.91-0.97. The robustness of the algorithm was thus studied and presented for a wide range of frame duration and ROI sizes.

The number of LORs available within each frame and the number of iterations needed to satisfy the chosen criterion varied with the ROI size and frame duration. With larger ROIs and frame duration, the number of LORs increased almost linearly, which was expected. The number of required iteration, however, was varied in an interesting way. It depended on how the mean perpendicular distance of all the LORs to the source particle/lesion varied with changing parameter setup. Increasing the frame duration did not affect the mean perpendicular distance and thus the number of iterations. On the other hand, mean distance increased in proportion to the increase in ROI size. With increasing mean distance, higher number of iterations were needed by the algorithm to discard distant LORs. Thus with increasing ROI size, the number of iterations increased, while increasing frame duration did not affect the iteration count.

The ability of the method in tracking the movement along all three axes at a time allowed studying the principal axis of movement and how it varied with respect to the location of the object of interest within the body. Point sources placed on the abdomen had a large range of movement in the Y direction. On the other hand, point sources on the side of the abdomen had significant movement in the Z direction. The lesion in the lung area and the myocardium regions showed dominant movements along the Z direction. This property can be beneficial for specific applications.

Two observations were made regarding the TOF-COM and COM algorithms while performing the comparison studies. The first one was that they exhibited more sensitivity to the size of ROI than the TOF-PEPT algorithm. In fact, for lesion motion tracking, the regular COM resulted in providing an extremely noisy signal. The size of ROI had to be decreased by means of further visual inspection to obtain a reasonable result. The second interesting finding was that in one case, the regular COM correlated better with the band than TOF-COM. It was expected that the use of TOF information would lead to better tracking performance. However, the possible reasoning behind this can be the use of the TOF center without consideration of its associated error/uncertainty. For the scanner used, the FWHM of the Gaussian associated with the TOF uncertainty was on the order of 6 cm, which was quite large compared to the 2 mm point source used.

The TOF-based LOR weighting proposed in this work was compared with the TOF-based LOR exclusion suggested for the PeTrack algorithm. It was found that for every case, the LOR exclusion resulted in degraded performance. In contrast, LOR weighting aided in boosting performance consistently. In particular, for the point source with relatively lower activity (Clinical Study 1-Point Source 2), the TOF-weighting enhanced the performance significantly. There was a 12.5% difference in correlation coefficients between the two approaches. It was speculated that since TOF has its own error in measurement, hard-thresholding based on the TOF either caused the removal of LORs that should have been considered or the inclusion of LORs that should have been discarded. On the other hand, the TOF-based weighting factor was designed in a way that does not mitigate the contribution of a specific LOR but enhances its contribution depending on its probability of actually being originated from the source of interest. With more advanced system with better timing resolution and more accurate TOF information, it is expected that the TOF-based weighting will allow more accurate motion estimate.

Most existing external tracking devices and data-driven methods consider only global motion detection, with very few addressing internal/localized lesion-specific motions. Motion correction based on global motion estimates may not fully correlate with individual lesion motion and thus offer appropriate correction. The presented TOF-PEPT algorithm is expected to provide a framework for more robust estimation of different kinds of motion data for use by motion correction algorithms.

Chapter 4

Motion Detection Algorithm: Application to Head-Neck Motion Correction

Head-neck motion correction is significantly important in brain PET imaging, where even the slightest movement can have an adverse impact on both image quality and quantitative assessments.

In this chapter, the application of the proposed motion detection algorithm has been studied for head-neck motion correction. A fully automated, data-driven methodology is proposed that can track the movement of the head from raw listmode data and therefrom perform event-based correction with the aid of externally placed radioactive markers.

The primary task involved sorting the whole duration listmode data into multiple motion-free static frames based on the motion data detected by the proposed TOF-PEPT algorithm. The first static position was considered to be the reference, and LORs within subsequent frames were aligned to it. To estimate the underlying rigid transformation parameters, radioactive point sources were used. They were placed on radiolucent glasses, and patients were asked to wear the glass. A TOF-weighted line density based particle tracking method was used to estimate the location of the point sources for each static frame. The rigid transformation parameters were then estimated from the singular value decomposition (SVD) [13] analysis of the point source coordinates.

Assuming the motion as a rigid body motion, the transformation from one position (x, y, z) to another (x', y', z') can be fully realized by a rotation along with translation as follows:

$$\begin{bmatrix} x' \\ y' \\ z \end{bmatrix} = \underbrace{\begin{bmatrix} R_{xx} & R_{xy} & R_{xz} & t_x \\ R_{yx} & R_{yy} & R_{yz} & t_y \\ R_{zx} & R_{zy} & R_{zz} & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix}}_{T} \begin{bmatrix} x \\ y \\ z \end{bmatrix} \tag{4.1}$$

where, $R_{xx}, R_{xy}, R_{xz}, R_{yx}, R_{yy}, R_{yz}, R_{zx}, R_{zy}, R_{zz}$ are rotational, and t_x, t_y, t_z are translational parameters. T denotes the whole transformation matrix. Spatial coordinates (x, y, z) of detector pair associated with each line-of-responses (LOR) can be re-positioned to the reference position (x', y', z') by applying the transformation matrix. With all the realigned LORs, it allows creating a motion corrected data. The following sections provide the entire detail of the implementation of the proposed approach and its application on head-neck motion detection and correction.

4.1 Implementation Details of Motion Correction

4.1.1 Static Frame Detection

Let M(t) =[X(t), Y(t), Z(t)] represents the estimated motion signal obtained by TOF-PEPT method. To determine when the patient moved, a forward difference approximation of the gradient, i.e.,$\nabla M(t) = M(t+1) - M(t)$, and its norm thereof was calculated as follows:

$$||\nabla M(t)||_2 = \sqrt{|X(t+1) - X(t)|^2 + |Y(t+1) - Y(t)|^2 + |Z(t+1) - Z(t)|^2} \tag{4.2}$$

The resulting signal was smoothed and then a threshold was determined from the mean absolute deviation (MAD) of the resulting signal to detect when the patient had significant movement. The threshold was set to be a scalar λ times the MAD. Mathematically, MAD

can be expressed as follows:

$$MAD = \frac{1}{n}\sum_{i=1}^{n}|S_i - \overline{S}| \quad \text{where, } S = ||\nabla M(t)||_2 \tag{4.3}$$

Here, n, S, and $\overline{S}$ represent the number of sampling data points, the data values of the signal of interest, and their average, respectively. The sampling time frames at which the resultant signal amplitudes were above the threshold were considered transitional and discarded afterward as the patient's head was in motion. The sampling time frames at which the threshold was not exceeded, were combined into longer duration frames, referred to as static time frames hereafter. Figure 4.1 presents an illustration. For the studies presented in this chapter, λ=3 was used. This choice allowed the detection of all discernible patient movements.

4.1.2 Point Source Detection: TOF-weighted Line Density Method

The proposed method is based on the line density based approach used for particle tracking [102, 7, 103]. The main idea lies in realizing the distribution of the LORs, i.e. positron emitters within the whole field-of-view (FOV), using a 3-dimensional accumulation array, i.e. a regularly spaced 3-dimensional grid with each voxel representing the total number of LORs that intercepted the corresponding voxel. All the voxel values were initialized by zeros and incremented for each LOR intercepting them. The process can be considered similar to back-projecting.

The accumulation array was specified using a Cartesian coordinate system with $200 \times 200 \times 109$ voxels and spatial resolution that matches the FOV for the scanner system used. To accumulate the voxels, Bresenham's algorithm [9] was implemented. The resultant array represented the distribution of the LORs within the whole FOV. Figure 4.2 is provided to aid in visualizing the accumulation array and its accumulation process. For each LOR, intercepting voxels were incremented with values corresponding to the Gaussian distribution associated with the time-of-flight (TOF) bin/center of each LOR. Voxels closer to the TOF bin/center were incremented with higher values than the voxels farther away. This allowed realizing the distribution with better precision.

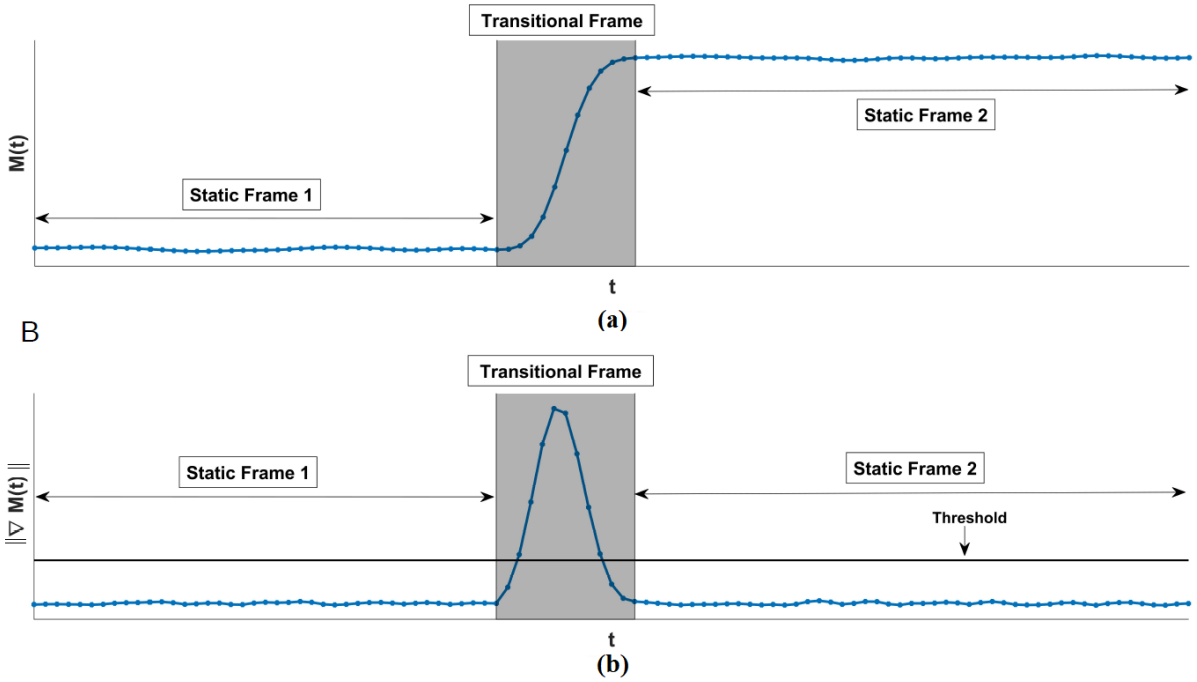


Figure 4.1: Illustration of the process of detecting static frames: (a) Estimated motion signal (b) Gradient norm of the forward different approximation of the motion signal. Based on the data-driven threshold, time frames were sorted into transitional frames and longer duration static frames.

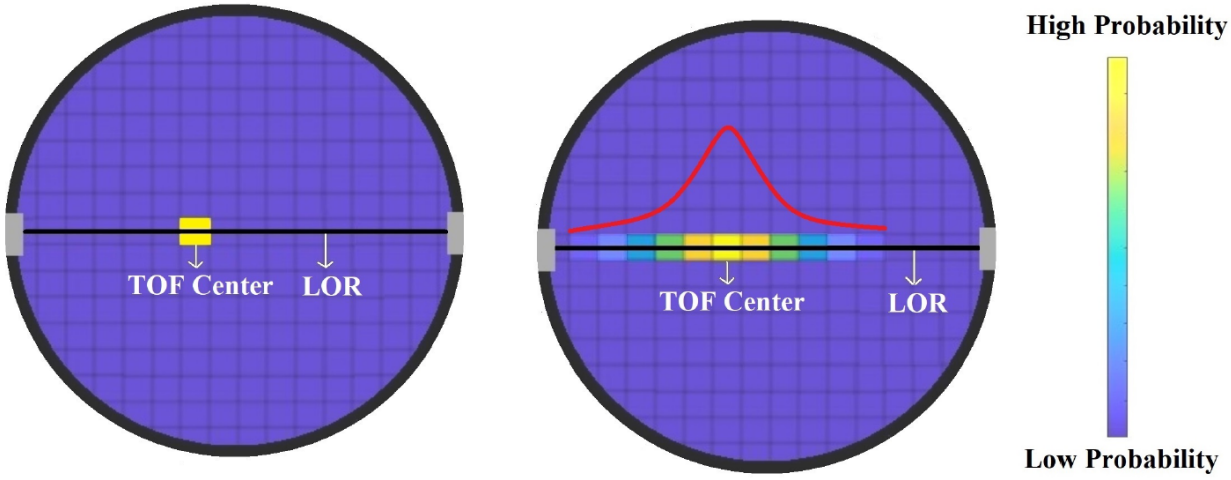


Figure 4.2: Illustration of the process of Bresenham mapping: (a) A line of response (LOR) with its time-of-flight (TOF) center (b) Incrementing the voxels intercepted by the LOR following a Gaussian distribution centered on the TOF center and with full width at half maximum (FWHM) corresponding to the system timing resolution

4.1.3 Extraction of Transformation Parameters

The point sources having high activity concentration were easily distinguishable in the accumulation array as regions with higher LOR distributions compared to the neighboring brain uptake regions. Assuming that N point sources produced N local maxima, The process of determining the location of point sources was performed one at a time in an iterative way. First, the coordinates of the voxel with the highest value were identified. That provided a location estimation of the corresponding point source. To further refine the estimation, a weighted average of the voxel coordinates within a $k \times k \times k$ neighborhood was taken with the voxel values as weights. If $(x_{max}, y_{max}, z_{max})$ denote the coordinates of the voxel with the highest count, and Ω denotes the neighborhood region centered on it, then the estimation of true location was made by:

$$\begin{aligned} x_{est} &= \frac{\sum_{i\in\Omega} x_i v_i}{\sum_{i\in\Omega} v_i} \\ y_{est} &= \frac{\sum_{i\in\Omega} y_i v_i}{\sum_{i\in\Omega} v_i} \\ z_{est} &= \frac{\sum_{i\in\Omega} z_i v_i}{\sum_{i\in\Omega} v_i} \end{aligned} \tag{4.4}$$

where (x_i, y_i, z_i) denotes the voxel coordinates in the neighborhood, and v_i denotes their corresponding values. At the end of each iteration, the neighborhood voxel values were reset to 0 to continue the search for the next point source. The kernel size k was chosen to be large enough to cover the uptake regions of the point sources yet small enough to not overlap with neighboring point sources' uptake region. Empirically, $k = 9$ was found to produce good results for the experimental setup used in this work.

This process was done for each static frame to determine how the point sources got repositioned from one frame to another. This allowed performing a singular value decomposition based analysis to estimate the rigid transformation parameters to align the LORs within each static frame to the reference frame. Covariance matrices were formed from the location coordinates of the point sources in the reference and subsequent static frames.

SVD of the matrices provided the rotational parameters, and translation parameters were calculated post hoc from the shift in the mean locations.

All the LORs in the static frames were repositioned according to the transformation parameters. The TOF information was also estimated from the repositioned LORs and updated accordingly. Repositioned LORs that fell outside the FOV were discarded. A new listmode data was then created with the remaining LORs. A final motion-corrected image was reconstructed therefrom.

4.2 Performance Evaluation

4.2.1 Validation and Evaluation

To evaluate the performance of the proposed automated LOR motion correction, motion corrected listmode data was rebinned into sinograms and images were reconstructed therefrom using the OSEM algorithm available on the Biograph mCT Flow PET/CT scanner. Reconstructions were carried out following the standard clinical protocol of the institution with 3 iterations, 24 subsets, and 5×5 Gaussian post-smoothing. All data processing and reconstruction were performed using the Siemens e7 processing tools (Siemens Healthineers, Knoxville).

To qualitatively and quantitatively evaluate the motion corrected images produced by the proposed approach, they were compared against a more conventional frame-by-frame image registration method. For each static frame, an image volume was reconstructed, and manual landmark based registration was implemented to align the image from each frame to the reference image. A single motion corrected image volume was produced by summing all the aligned image volumes. To perform landmark based registration, the point sources were used as landmarks. Regions-of-interests were drawn manually encompassing each point source and were subjected to thresholding (15% of maximum value) to create a binary mask. The centroid of each mask was used as landmark which corresponded to each point source.

To quantify the registration accuracy, thresholding was applied to images to create binary masks, and Dice index [22] was used as follows to measure the percentage of overlapping:

$$\text{Dice(A,B)} = \frac{2|A \cap B|}{|A| + |B|} \tag{4.5}$$

where A and B denote binary masks segmented from reference and registered images, respectively. A high value indicates a high degree of similarity, and $0 \leq \text{Dice(A,B)} \leq 1$.

In addition to the quantitative comparisons, two board-certified radiologists qualitatively reviewed the motion corrected image. Motion corrected images were reviewed and visually assessed by them to determine if they were reasonable and useful from a clinical perspective.

4.2.2 Data

For head-neck motion correction, 5 Brain PET imaging studies were carried out with the use of external radioactive markers in compliance with the University of Tennessee Institutional Review Board approved protocol (IRB #3941). Imaging was performed on a 64-slice Biograph mCT Flow PET/CT, and 64-bit listmode data were acquired. Patients were asked to randomly rest their heads in different positions and orientations during the scan period. Thus, for every patient study performed, there were different ranges of movements and multiple static motion free frames. Three or four external markers were placed on radiolucent glasses worn by the patient. They were around 2 mm in diameter, made from zeolite beads, and had radioactivity in the $74-185$ kBq ($2-5$ μCi) range. Table 4.1 provides the information related to the number of point sources used, total scan duration, and the total number of static frames for each study.

4.2.3 Qualitative Analysis on the Estimated Motion and Quantitative Evaluation of Transformation Parameters

Figure 4.3 shows detected motion signals using TOF-PEPT for all five clinical studies. They are shown in the direction along which the patient had the largest range of movement. Since the patients' head movement was mostly around Z and Y axes, the dominant movement was

Table 4.1: Description of the experimental studies conducted

Study	Number of External Point Sources	Scan Duration (Minutes)	Number of Static Frames
Patient Study 1	3	3	3
Patient Study 2	3	3	4
Patient Study 3	4	3	4
Patient Study 4	3	3	3
Patient Study 5	3	3	4

observed in the XY (axial), and ZX (coronal) planes for all the studies. The patient in study 1 moved twice. The head was rested on three different positions during the whole 3 minute scanning period, and there were thus three static frames. Patient studies 2, 3, 4, and 5 had four, four, three, and four static frames, respectively. The shaded grey areas in both figures show the transitional states. All the transitions were accurately identified by TOF-PEPT algorithm for all the patients.

Figures 4.4(a) and (c) show the location coordinates of the point sources in each static frame in the axial and coronal plane, respectively, for one patient study. For *frame 2* and *frame 3*, transformation parameters were derived to reposition the point sources in *frame 1*. The derived transformation parameters were then applied to the point sources in *frame 2* and *frame 3* to evaluate the performance of registration. Figure 4.4(b) and (d) show the point sources' location after applying the transformation parameters. To quantify the registration error, the mean Euclidean distance between transformed and reference location was calculated. Table 4.2 shows the results. It can be seen that the derived transformation parameters allowed registration with the mean distance within 1.15 ± 0.09 mm.

4.2.4 Qualitative and Quantitative Comparison of Motion Corrected Images

Figure 4.5 presents the comparison of the motion corrected images for Patient Study 3 (Top row) and Patient Study 5 (bottom row). The images are shown in the axial plane. The left column shows a slice of the original motion impacted image volume. It can be seen that the head was positioned in three different orientations, which resulted in significant blurring and loss of detail. The initial head position served as the first static reference frame. With the TOF-PEPT algorithm, each motion free static frame was identified and corrected. The middle column shows a slice from the image volume created by implementing manual frame-by-frame landmark registration. The right column shows results from the image volume reconstructed from the motion corrected listmode data created by applying the proposed automated data-driven correction. Figure 4.6 shows the sagittal and coronal views for Patient Study 3 and Patient Study 5, respectively. It can be seen that the shifted

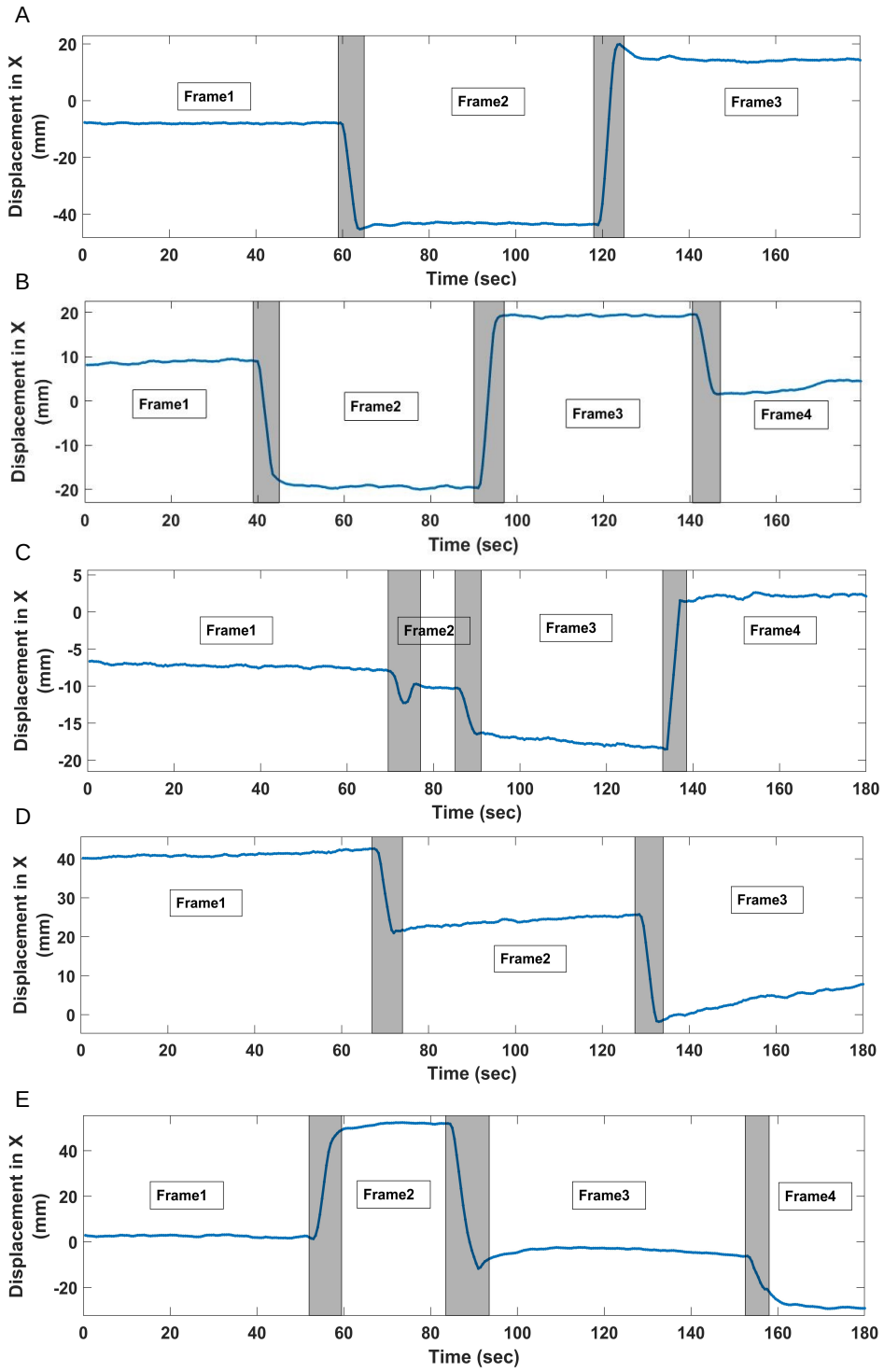


Figure 4.3: Plots of the detected motion signals (along the principal direction of motion) and static frames: (a) Patient Study 1: Three static frames; (b) Patient Study 2: Four static frames; (c) Patient Study 3: Four static frames; (d) Patient Study 4: Three static frames; and (e) Patient Study 5: Four static frames. The gray colored regions show the transitional frames.

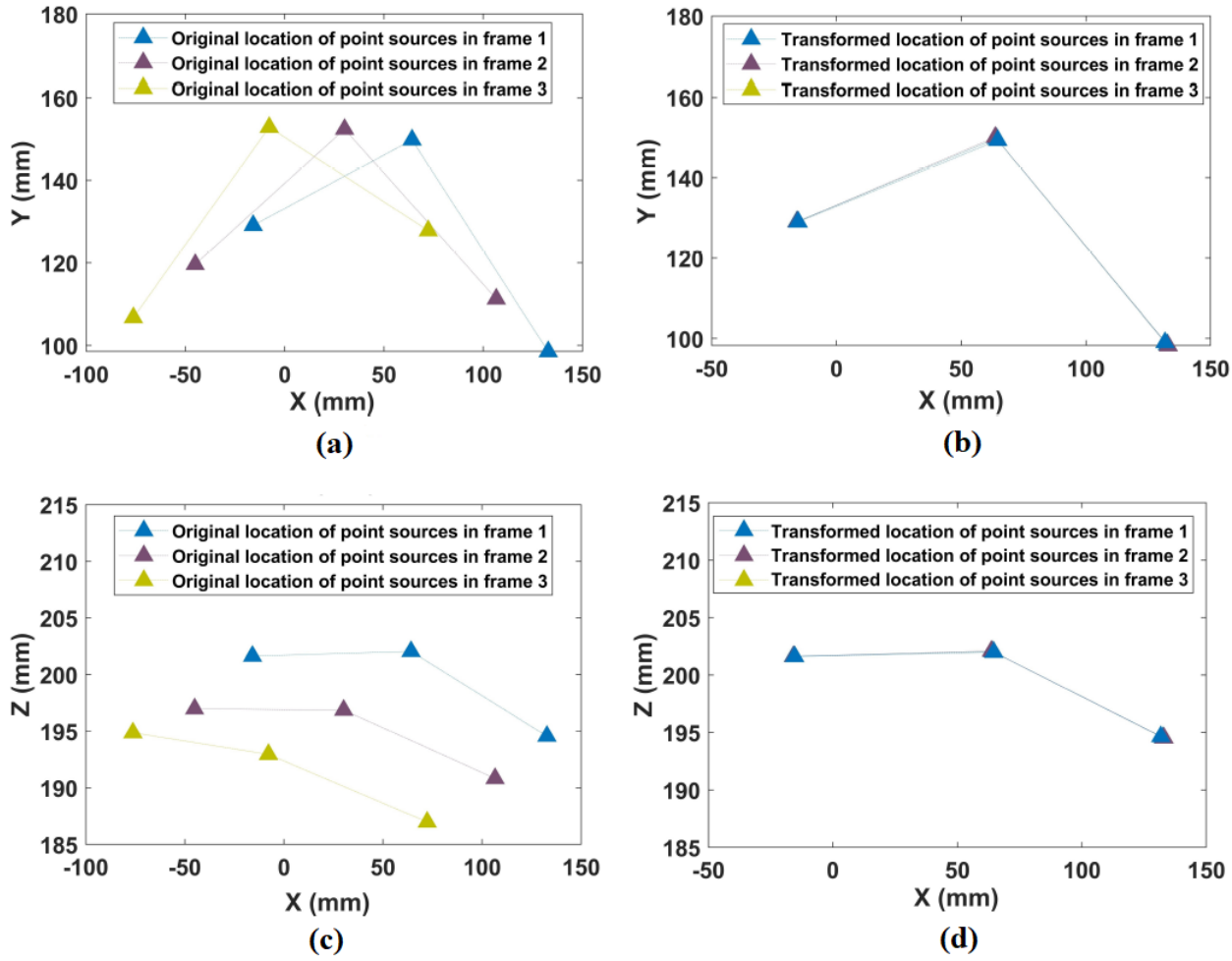


Figure 4.4: Point sources' locations in the axial plane (top row) (a) before and (b) after the application of the transformation and in the coronal plane (bottom row) (c) before and (d) after the application of the transformation. It can be seen that the point sources were aligned in different orientation in each different frame before and got aligned to the reference position after applying the transformation.

Table 4.2: Mean Euclidean distance between reference and transformed coordinates

Study	Mean Euclidean Distance (mm)		
	Frame 2	Frame 3	Frame 4
Patient Study 1	1.48	1.36	—
Patient Study 2	2.31	2.51	0.78
Patient Study 3	0.52	0.77	0.73
Patient Study 4	0.38	0.77	—
Patient Study 5	1.24	0.78	1.68
Mean	1.17	1.23	1.06

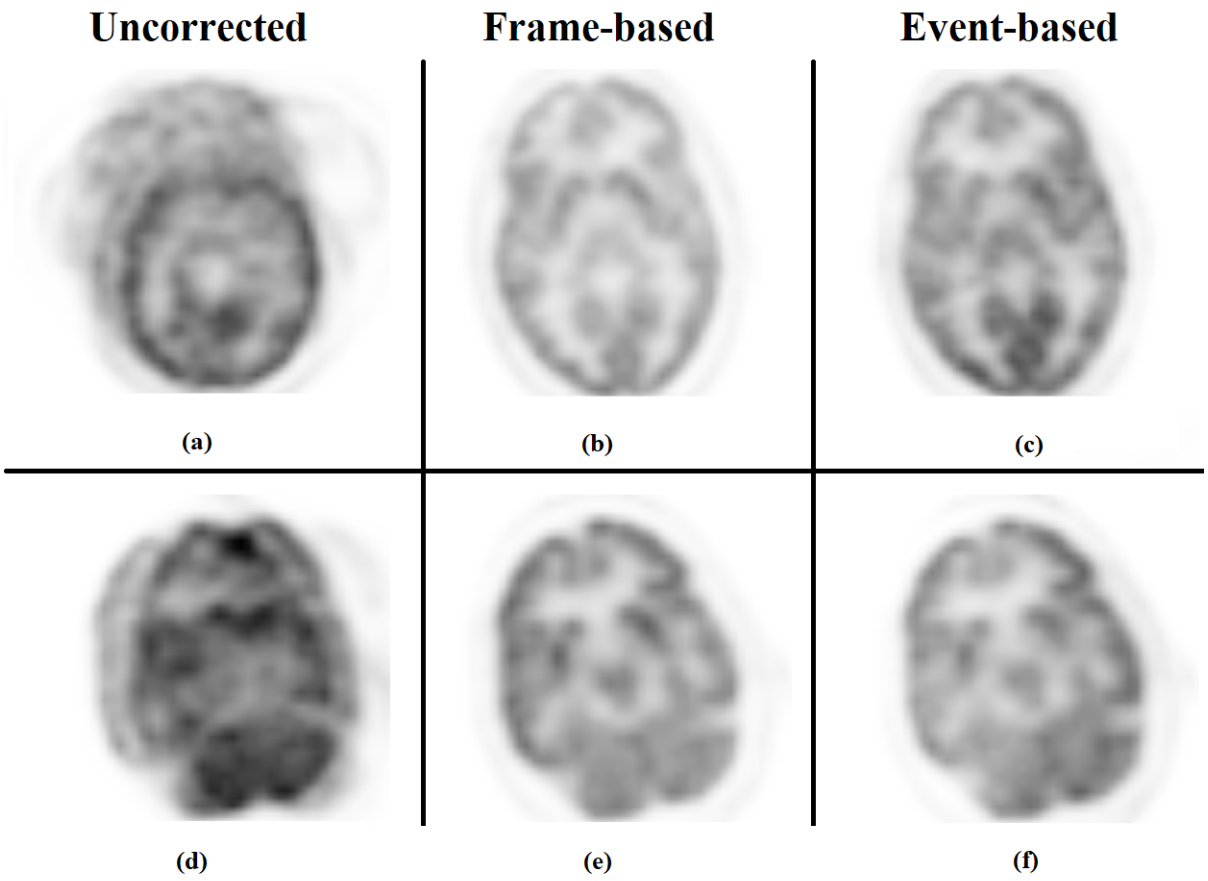


Figure 4.5: Comparison of uncorrected and motion corrected images for Patient Study 3 (Top row) and Patient Study 5 (Bottom row) in the axial plane from: (a, d) Original motion impacted image; (b, e) Corrected image by applying frame-by-frame manual landmark based registration; (c, f) Corrected image by applying the proposed automated listmode event data correction.

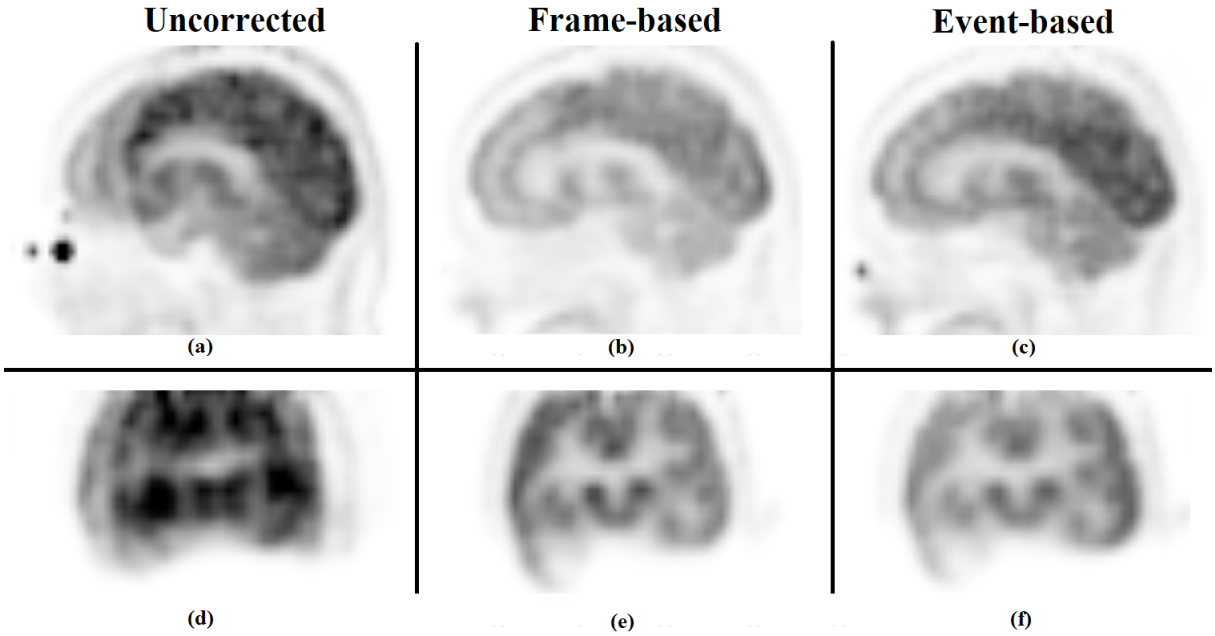


Figure 4.6: Comparison of uncorrected and motion corrected images for Patient Study 3 (Top row) in the sagittal plane and Patient Study 5 (Bottom row) in the coronal plane: (a, d) Original motion impacted image; Corrected image by applying frame-by-frame manual landmark based registration; (c, f) Corrected image by applying the proposed automated listmode event data correction.

head positions were aligned well to the reference position and the overall motion corrected images enhanced the image sharpness.

For quantitative comparison, a region-of-interest encompassing the brain was drawn, then was subjected to 15% and 25% thresholding to create binary masks and compute Dice index. Table 4.3 lists the resulting Dice indices for all five patient studies. It can be seen that the gold standard created for this study by implementing manual frame-by-frame registration and the proposed automated event based correction achieved similar results with numbers in the range of $0.965 \pm 0.024\%$ for the former and $0.961 \pm 0.027\%$ for the latter.

The research PET studies conducted for this work were acquired after diagnostic CT and PET, and patient positions had to be readjusted during research PET. This resulted in an attenuation mismatch between PET and CT and slight uptake changes. Reconstruction of the PET images without attenuation correction, as shown in Figure 4.7 for Patient Study 3, shows the removal of the uptake changes.

The qualitative review of the images by the radiologists indicated that the manually registered images showed comparable image quality to that of the automated algorithm. The difference that was noted between images from the two methodologies was the change in uptake pattern. In routine clinical practice, this mismatch would not occur as it would not be needed to readjust the patient position as had to be done for the research PET.

4.3 Summary

In this chapter, a fully automated, data-driven methodology has been presented that directly performed listmode data correction without the need for reconstructing and registering multiple image frames. Data correction was performed by event repositioning with the aid of external point sources as unique data points in order to find rotation and translation information. From motion detection to correction, the entire methodology was applied on raw listmode data with the TOF information.

The previously proposed motion detection algorithm, TOF-PEPT, utilized a sufficiently high time resolution (500 ms frames) to minimize intra-frame motion. The algorithm was able to identify all the significant patient movements for all patient studies, which allowed

Table 4.3: Quantitative evaluation of motion corrected image

Study	Threshold	Dice Index	
		Event Based	Frame Based
Patient Study 1	15%	0.970	0.974
	25%	0.920	0.938
Patient Study 2	15%	0.989	0.984
	25%	0.954	0.949
Patient Study 3	15%	0.993	0.995
	25%	0.980	0.985
Patient Study 4	15%	0.971	0.967
	25%	0.931	0.918
Patient Study 5	15%	0.973	0.980
	25%	0.924	0.954
Mean		0.961	0.965

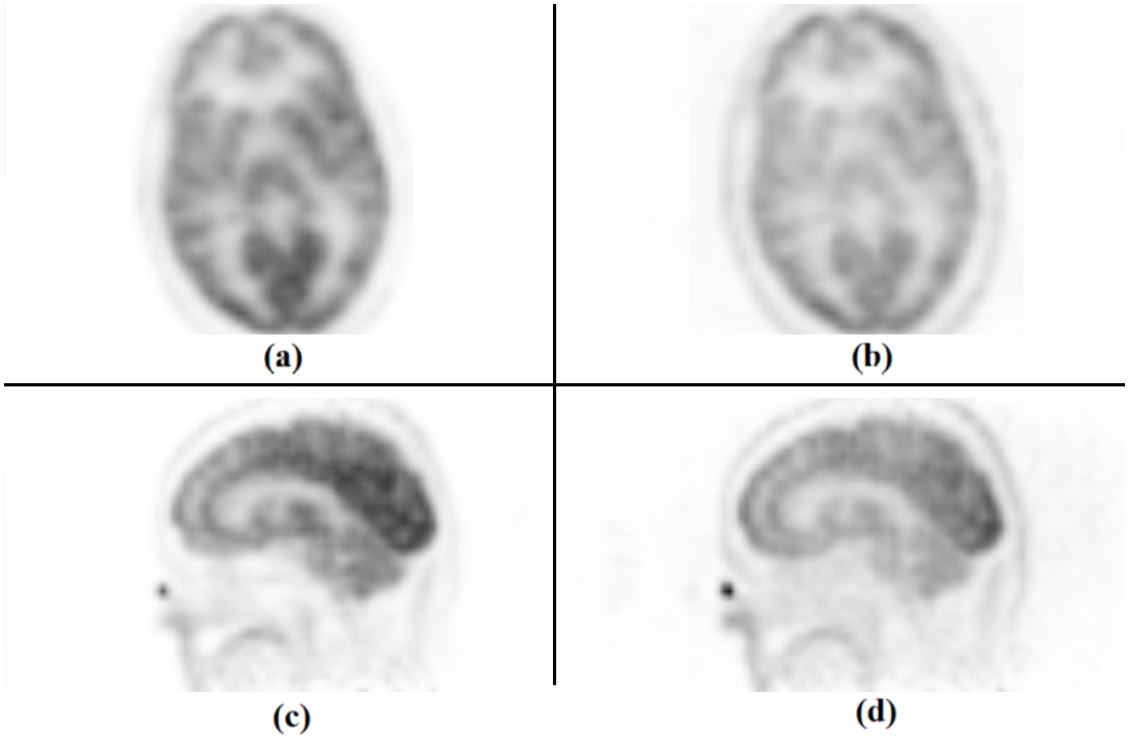


Figure 4.7: Motion corrected images showing the impact of attenuation correction for Patient Study 3 in (Top row) axial plane and (Bottom row) sagittal plane : (a, c) With attenuation correction (b, d) Without any attenuation correction.

to create completely motion free static frames. For each static frame, transformation parameters were estimated with a reasonable mean Euclidean distance between reference and transformed location. This allowed the repositioning of shifted LORs to the reference position with satisfactory accuracy.

A variation of the particle tracking technique has been proposed, following a line density based algorithm with TOF information in order to estimate the location of multiple unique point sources. The use of TOF allowed setting the voxels with higher values if closer to the origin of annihilation and thus have a better realization of the distribution of the activity uptake within the FOV.

For refined estimation of point source location, the algorithm made use of a kernel of user defined size to calculate the weighted average of voxel coordinates and reset them to zero during each iteration. The impact of the kernel size was studied, and it was observed that while using it for resetting voxel values, a smaller kernel size produced a higher error. The possible reason for this is that with the smaller kernel, voxels corresponding to one particular point source were not being reset properly while searching for another point source. Thus, kernel size should be large enough to encompass the uptake region of a point source and smaller than half the minimum distance between two point sources not to cause overlapping. The size of the kernel, however, did not result in much difference while using it for estimating the weighted average of voxel coordinates. Thus the same kernel was used for both purposes.

To have the external point sources distinguishable in the accumulation array, they are expected to have higher activity concentration compared to the brain area. This was conveniently achieved with reasonably low radioactive (2-5 μCi) point sources of small sizes (2 mm in diameter), which had very higher activity concentrations.

In order to evaluate and validate the performance of the proposed approach, a frame-by-frame landmark-based registration technique was implemented as a gold standard. Motion corrected images were created both from the proposed automatically corrected listmode data and by applying the frame based registration. The Dice similarity index results showed that both approaches achieved nearly identical performance. To note, the manual frame-based registration technique availed the motion information derived from the proposed motion

detection algorithm. Thus, it would not be typically possible to achieve similar performance with frame-based registration in the absence of such prior motion information.

Additional diagnostic review of the radiologists found the performance of the proposed approach comparable to the manual registration approach, except for the attenuation mismatch. With both the methods demonstrating similar performances, the results favor the proposed automated approach as it eliminates the need to reconstruct multiple image volumes and perform manual data correction with the added benefit of being able to automatically determine whether significant motion occurred during imaging and then correct the data accordingly.

Chapter 5

Deep Learning Based PET Motion Correction

PET image suffers motion artifacts due to various kinds of motion. Several motion correction approaches have been suggested to date that involves image registration. The deep learning approach in image registration has drawn significant attention and achieved remarkable performance in recent years. The aim is to explore a deep learning approach to deal with PET motion correction with the aid of image registration.

In this work, affine registration with a deep learning approach has been studied with a recurrent convolutional neural network. Output from a neural network is passed as input to the same network at the next step to refine the prediction of the network. In earlier research works, it was demonstrated that a multi-step recurrent network yields better performance [86], particularly for affine registration. A similar concept is implemented in this work with a few modifications. A detailed overview of the implemented approach is provided in the following sections.

5.1 Brief Overview of Image Registration

The image registration task involves mapping a moving source image I_{src} to a fixed target image I_{tgt}. I_{src} and I_{tgt} can be defined in the spatial domain $\Omega \in \mathbb{R}^{h \times w \times d}$. The objective is to obtain a mapping function f that will transform I_{src} to I_{tgt}, i.e., $f : I_{src} \rightarrow I_{tgt}$. If

I_{wrpd} denotes warped image, and ϕ denotes transformation parameters, then the warping operation can be expressed as:

$$I_{wrpd} = f(I_{src}, \phi) \tag{5.1}$$

The task of the neural network is to predict the transformation parameters ϕ by optimizing the network parameters θ in a way that minimizes the dissimilarities between warped and fixed images. If $\mathcal{S}$ denotes an image dissimilarity metric, then the network learns by optimizing:

$$\theta = \arg\min_{\theta} \mathcal{S}(I_{wrpd}, I_{tgt}) \tag{5.2}$$

If d denotes image dimensions, then the transformation parameters for affine registration include a transform matrix A such that $A \in \mathbb{R}^{d \times d}$ and translation vector B such that $B \in \mathbb{R}^{d \times 1}$.

5.2 Network Architecture

In this work, an unsupervised deep learning approach has been adopted by means of a multi-step convolutional neural network (CNN) [86] and a spatial transform layer [33]. For each pair of source and target images, it was concatenated and fed as input to the CNN. The CNN was made to predict transformation parameters ϕ, which were then passed to the transform layer along with the source image. The spatial transform layer consisted of a grid generator and a sampler. The grid generator determined the new mapping based on the predicted transformation parameters, and the sampler performed linear interpolation to warp the moving source image.

In multi-step network architecture, the network was trained in a recurrent approach. The warped image that came initially as output from the spatial transform layer was passed to the same network at the next step as the input source image, and the network then made another refined prediction from the updated image pair. Figure 5.1 shows an illustration of the network. This process can be continued for the number of steps predefined by the user. In this work, a 3-step prediction has been performed. The final transformation parameters were calculated by taking the composition of the parameters at each step. Expressed

mathematically, if A_1, A_2, and A_1 are the affine transform matrices, and t_1, t_2, and t_3 are the translation vectors at three steps, respectively, then the composition A' and t' can be obtained by:

$$\begin{aligned} A' &= A_3 A_2 A_1 \\ t' &= A_3 A_2 t_1 + A_3 t_2 + t_3 \end{aligned} \tag{5.3}$$

The architecture of CNN was inspired from the literature in [110] and subsection 5.4.4 provides a study as support of the network choice. The network consisted of a series of convolutional and pooling layers. Each convolution operation was performed with a kernel size 3, stride 1, and with Relu [63] activation function. Average pooling operation with kernel size 2 was performed after convolution in selected layers. At the final layer, two convolution operations were carried out to predict the transform matrix and translational parameters, respectively. The kernels were of input image size, and a linear activation function was used. Figure 5.2 shows the CNN architecture.

5.3 Loss Functions

Image Dissimilarity Loss

In this work, negative Normalized Cross Correlation (NCC) was used as the image dissimilarity loss function. The loss function ($\mathcal{L}_{img}$) can be expressed by [73]:

$$\mathcal{L}_{img} = -NCC(I_{wrpd}, I_{tgt}) = -\frac{\sum_{i\in\Omega}(I^i_{wrpd} - \overline{I_{wrpd}})(I^i_{tgt} - \overline{I_{tgt}})}{\sqrt{\sum_{i\in\Omega}(I^i_{wrpd} - \overline{I_{wrpd}})^2}\sqrt{\sum_{i\in\Omega}(I^i_{tgt} - \overline{I_{tgt}})^2}} \tag{5.4}$$

where, $\overline{I_{wrpd}}$ and $\overline{I_{tgt}}$ denote the mean of the warped and fixed image, respectively.

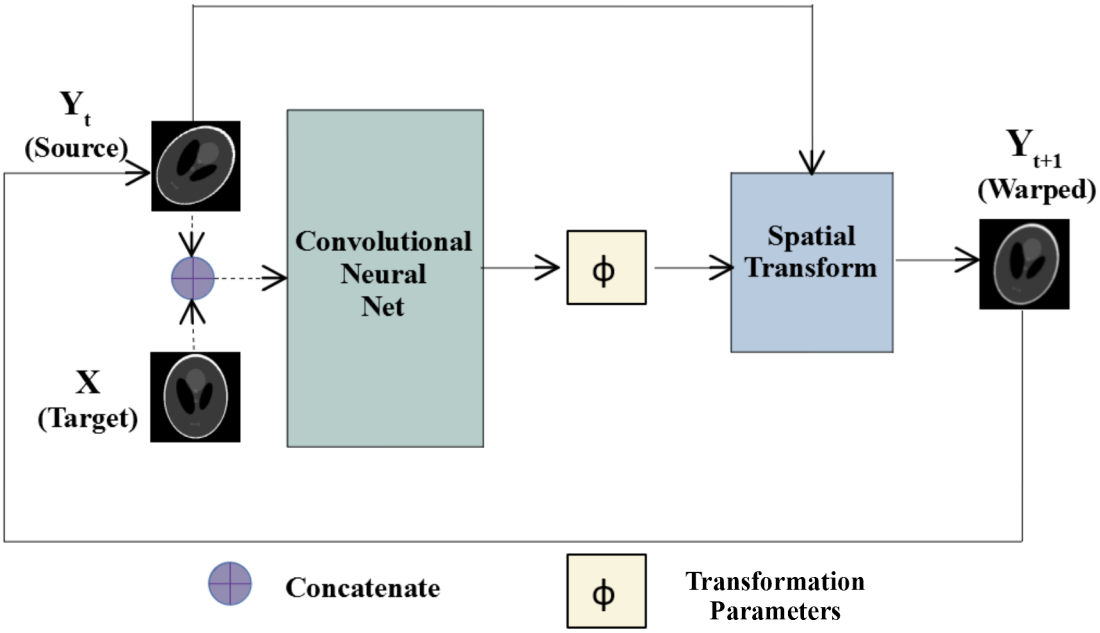


Figure 5.1: Multi-step affine registration network: Initially, source and target images are concatenated and fed into the convolutional neural network. The network outputs transformation parameters which are passed to the spatial transform layer along with the source image. The layer outputs warped images. At the next step, the warped image is passed as the source image, and the process continues for t number of steps.

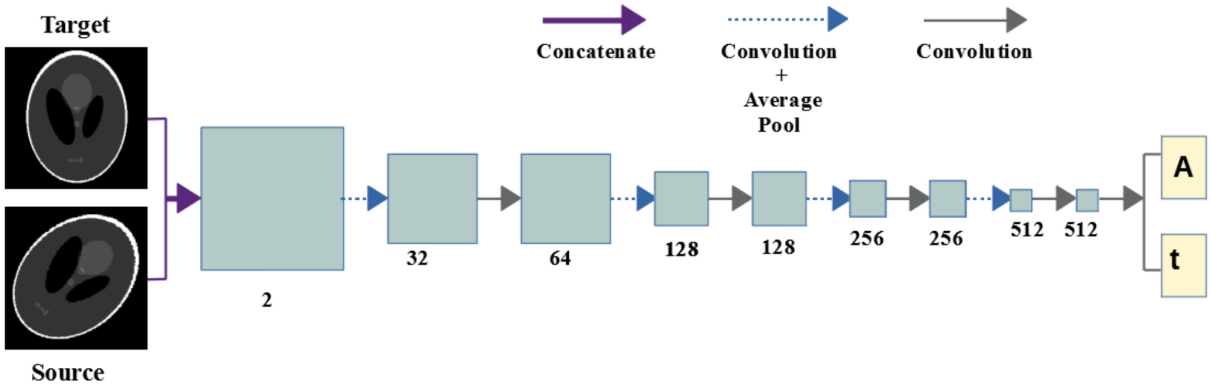


Figure 5.2: Convolutional neural network: Concatenated source and target images are passed to a series of convolutional and pooling layers. At the final layer, two convolution operations yield the transformation parameters.

Regularization Loss

The following loss function was used [86] to regularize the transformation prediction:

$$\mathcal{L}_{reg} = ||A' - I||_F^2 + ||b||_2^2 \tag{5.5}$$

Here, F denotes Frobenius norm of the difference between transform matrix A and identity matrix I, and b denotes translation vector. This loss is used to regularize the prediction of the transformation parameters so that they do not overshoot.

Total Loss

The total loss for the entire network can be expressed as follows:

$$\mathcal{L}^{total} = \mathcal{L}_{reg}\lambda_{reg} + \mathcal{L}_{img}\lambda_{img} \tag{5.6}$$

Here, λ_{reg} and λ_{img} denote the weighting factors for regularization and image dissimilarity loss, respectively. λ_{img} and λ_{reg} were set to 1 and 0.01, respectively. The choices were made empirically for this work, and a future work intends to explore this in-depth.

5.4 Performance Evaluation

5.4.1 Validation and Evaluation

To evaluate the performance of the neural network in registering image, Dice index as defined in chapter 4 was used to statistically measure the similarity between warped and target images. A higher dice index indicated better performance. Additionally, as a baseline, the conventional iterative affine registration algorithm from advanced normalization tools (ANTS) software package [2] was used to compare the performance of the registration and computation time.

To further extend the application of the neural network in motion correction and evaluate its performance, it was implemented on the patient study conducted and presented in

Chapter 4. Each patient study had three to four motion free static frames, which in the earlier work were registered to a reference frame with the aid of external point sources. Here, the registration task was carried out with the trained neural network. For each static motion free frame, 3D PET image volume was created, preprocessed, and passed to the neural network with the reference image volume. The network then provided the required transformation matrix along with a registered image as output from the spatial transform layer. The registered images were then summed together to produce an overall motion corrected image.

Lastly, the predicted transformation parameters from the neural network were used to correct the raw listmode data as described in the previous chapter. To be specific, transformation parameters for each static frame were applied on all the line-of-responses (LOR) within each frame to align them to the reference frame. With all the remapped LORs, motion corrected listmode data and therefrom motion corrected image volumes were created with Siemens e7 processing tools (Siemens Healthineers, Knoxville).

5.4.2 Data

There were in total 100 3D PET image volumes augmented from 5 different patient studies. The augmented dataset was created by creating random transformation parameters and repositioning the LORs accordingly in raw event data space. This resulted in augmented listmode files which were then histrogammed and sent to the reconstruction algorithm in order to create final image volumes. Images were resized to the dimension of $128 \times 128 \times 96$. The dataset was cross-validated with $4:1$ split ratio as training and test data. The neural network was trained for 100 epochs with 20 steps per epoch and a batch size of 4. The learning rate was set to $1e-4$.

5.4.3 Qualitative and Quantitative Evaluation of the Neural Network Performance

Mean Dice index and computational time were used as evaluative metrics. Table 5.1 shows the comparison of the neural network performance with respect to ANTs for the augmented

dataset. It can be seen that the neural network achieved a very similar performance compared to the conventional iterative algorithm with 20x faster computation time with GPU and 3x faster with CPU. For motion correction as well, the deep learning approach performed on par with the iterative approach. The network was trained in the high-performance computing (HPC) cluster of the University of Tennessee, Knoxville, with Tesla V100S GPU and Intel Xeon (E5-2670) CPU.

The performance of the neural network was tested on the actual clinical dataset as well, and the trained network was used to register images for the purpose of performing motion correction. Figures 5.4 and 5.6 show the qualitative evaluation of the network performance in motion correction. Figures 5.3 and 5.5 show the three motion free static frames, respectively, for two patient studies. *Frame 2* and *Frame 3* were registered to the reference frame *Frame 1* with the aid of the trained network. They were registered with ANTs as well to compare the performance of the neural network. The motion corrected images were compared with the uncorrected image in all three planes: axial, coronal, and sagittal. Image quality improvements with respect to the uncorrected image were readily apparent in the motion corrected images, with neural network and ANTs showing comparable performance. Table 5.2 shows the quantitative comparison between two methods by means of Dice index. Both from qualitative and quantitative perspectives, deep learning and conventional iterative approach yielded similar performance.

In order to further quantitatively evaluate the performance from a clinical perspective, one-dimensional line profiles were taken along the brain area in the motion corrected images, as well as in the uncorrected images to analyze the standardized uptake value (SUV). Figures 5.7 and 5.8 show the results with the measured SUV along the line. It can be seen that the peak-to-valley ratios of SUV were higher in the motion corrected images, which can be expected in a reduced motion blurred image. There were $30 - 60\%$ increases in the peak-to-valley ratio. The neural network can be seen to perform better or similar with respect to ANTs.

Lastly, with the transformation parameters provided by the trained neural network, the original raw listmode data were corrected. Figure 5.9 shows the uncorrected and motion corrected image volume reconstructed from the corrected raw listmode data with

Table 5.1: Comparison of image registration performance

Study	Mean Dice index		Mean Computational Time (Seconds)		
	ANTs	Deep Learning	ANTs	Deep Learning	
				GPU	CPU
Cross Validation 1	0.82	0.80	2.49	0.15	0.81
Cross Validation 2	0.85	0.84	1.96	0.16	0.96
Cross Validation 3	0.86	0.86	2.08	0.11	0.93
Cross Validation 4	0.91	0.88	2.61	0.10	0.80
Cross Validation 5	0.82	0.81	4.08	0.11	0.80
Mean	0.85	0.84	2.65	0.13	0.86

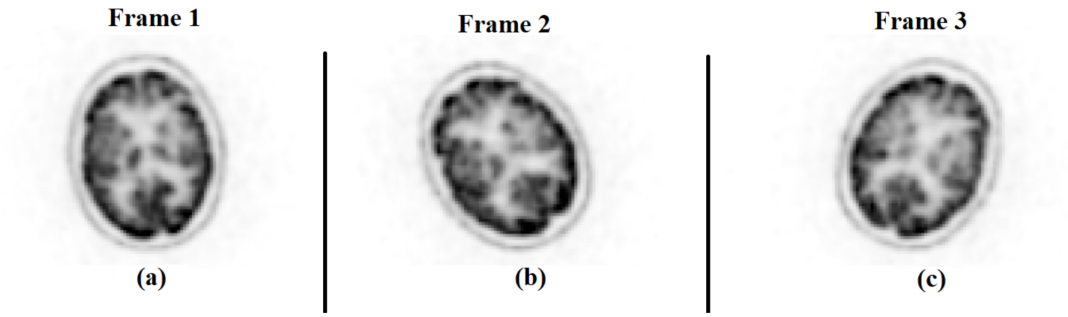


Figure 5.3: Example of three static frames where the patient rested their head in three different positions. The figure shows a slice from 3D volume in the axial plane.

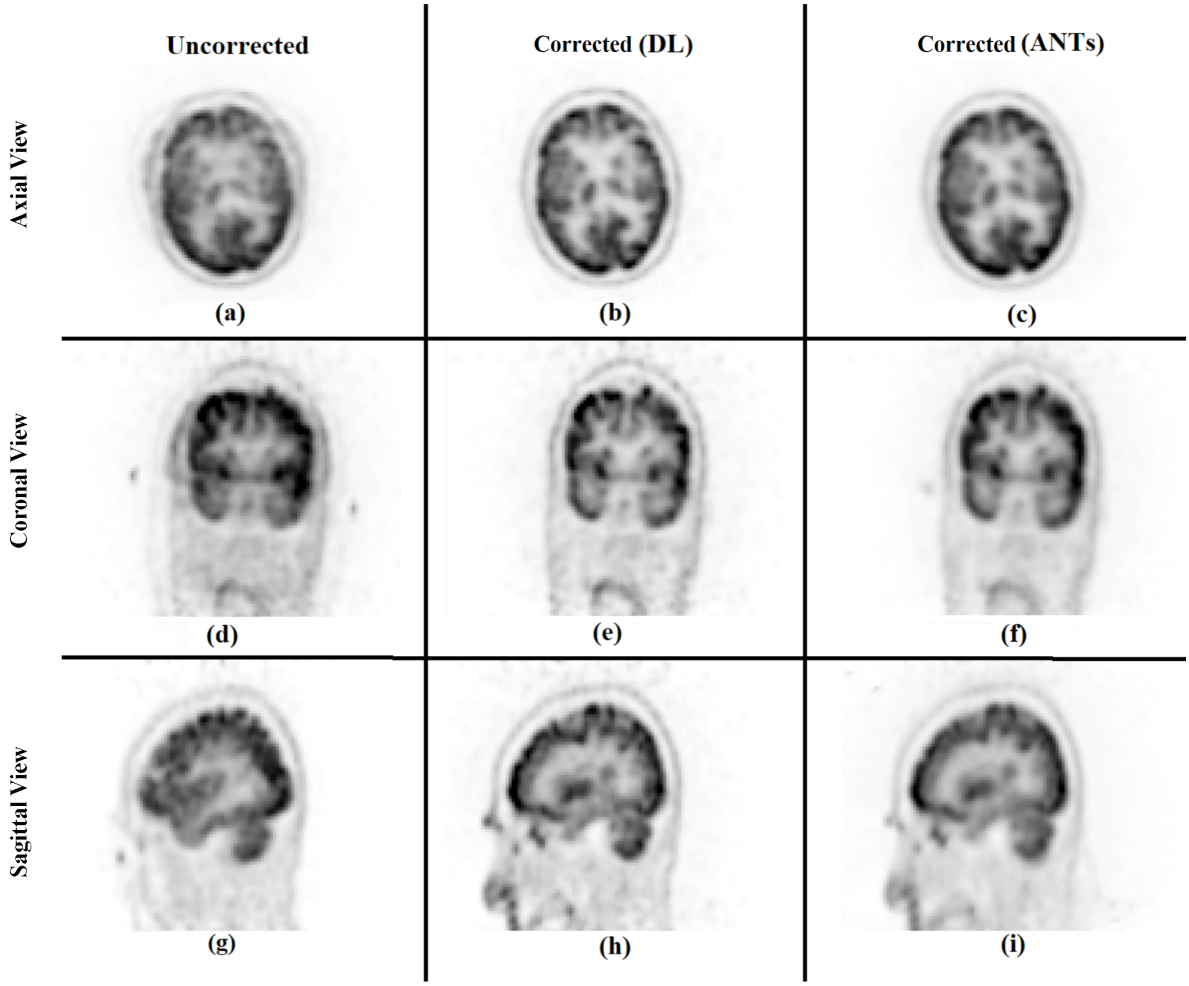


Figure 5.4: Example of neural network performance in warping moving source image. Rows from top to bottom show the sum of the three frames without any correction (left), correction with registration from the deep learning approach (DL-middle), and correction with registration using ANTs iterative algorithm (right), respectively, in the axial, coronal, and sagittal view.

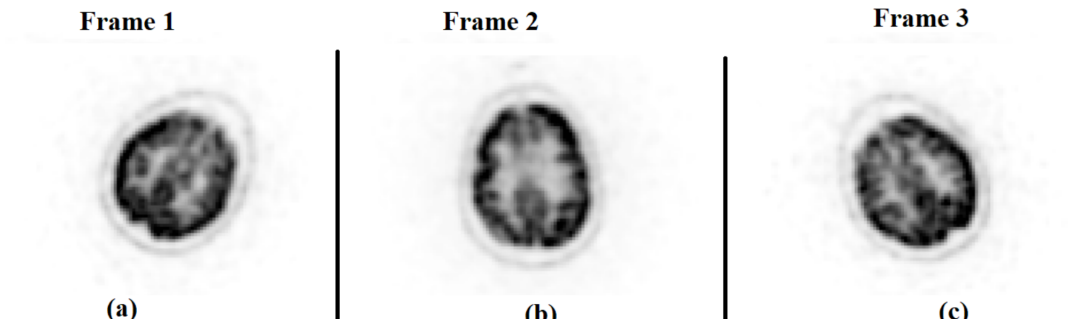


Figure 5.5: Example of three static frames where the patient rested their head in three different positions. The figure shows a slice from 3D volume in the axial plane.

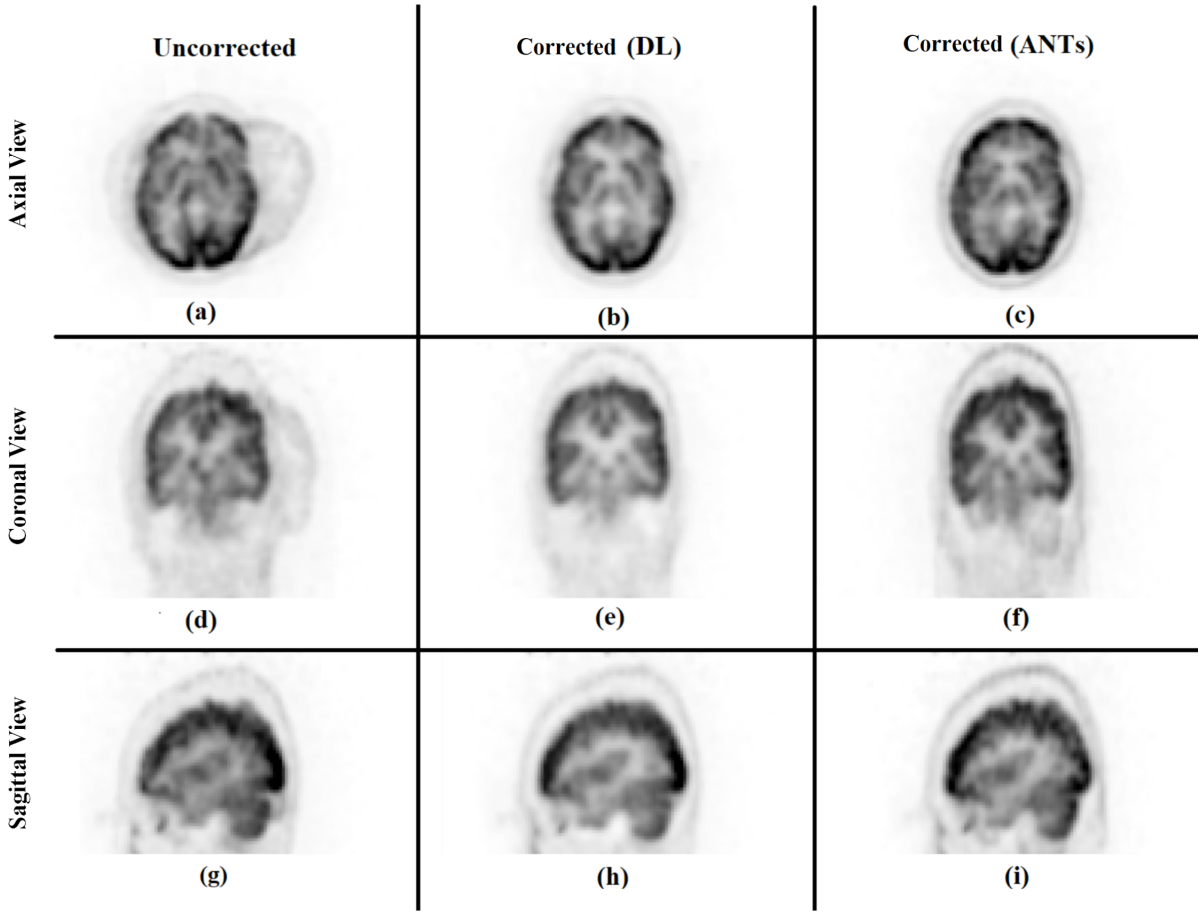


Figure 5.6: Example of neural network performance in warping moving source image. Rows from top to bottom show the sum of the three frames without any correction (left), correction with registration from the deep learning approach (DL-middle), and correction with registration using ANTs iterative algorithm (right), respectively, in the axial, coronal, and sagittal view.

Table 5.2: Comparison of performance for the application to motion correction

Study	Mean Dice index		Mean Computational Time (Seconds)	
	ANTs	Deep Learning	ANTs	Deep Learning (GPU)
Clinical Study 1	0.85	0.83	5.74	0.30
Clinical Study 2	0.79	0.82	3.62	0.32
Clinical Study 3	0.93	0.92	3.74	0.22
Clinical Study 4	0.93	0.94	6.62	0.20
Clinical Study 5	0.90	0.91	7.23	0.33
Mean	0.88	0.88	5.39	0.27

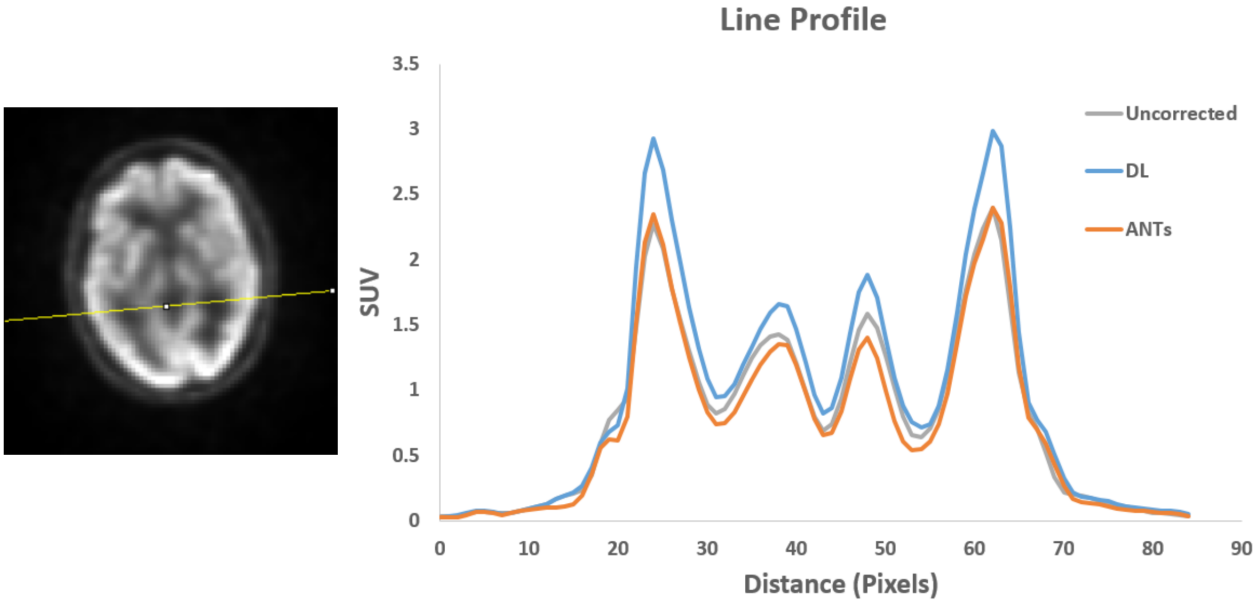


Figure 5.7: A line profile (Right) drawn across a region in the brain area (Left). It can be seen that the peak to valley ratio is higher for motion corrected data with deep learning (DL) than ANTs.

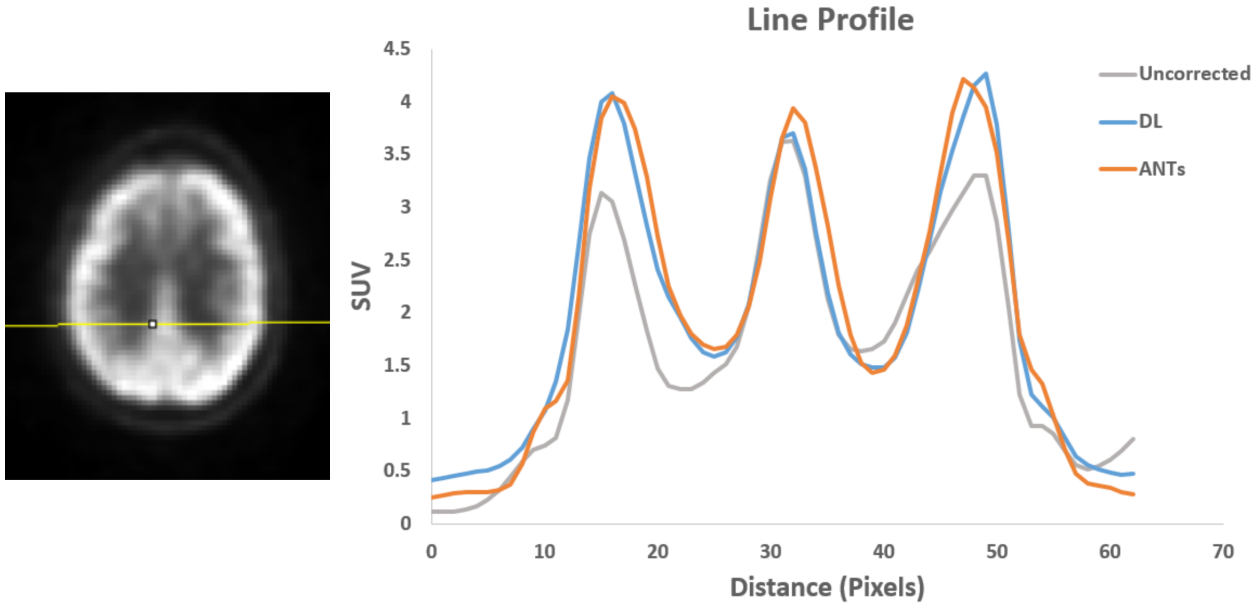


Figure 5.8: A line profile (Right) drawn across a region in the brain area (Left). It can be seen that the peak to valley ratio is higher for motion corrected data with deep learning (DL) and ANTs.

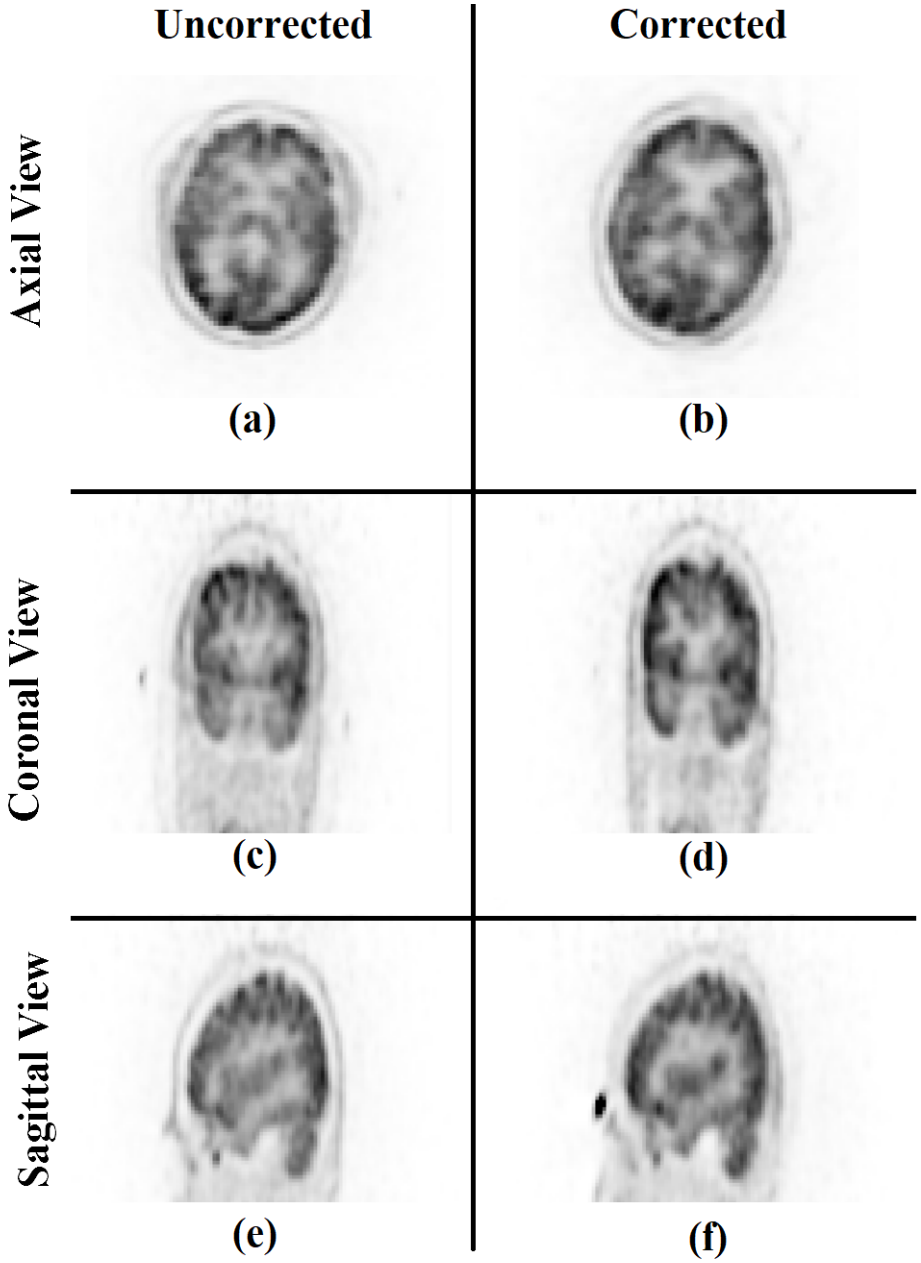


Figure 5.9: Example of neural network performance in creating motion corrected listmode data and therefrom reconstructed image volume. Rows from top to bottom show the sum of the three frames without any correction (left), correction with registration from the deep learning approach (right), respectively, in the axial, coronal, and sagittal view.

the transformation parameters prediction from the neural network. Motion corrected image volume can be seen to have sharper details compared to the uncorrected data.

5.4.4 Analysis of the Neural Network Parameters

There were two studies conducted to evaluate the network performance with respect to the choice of step size (defined in Section 5.2) and the network design.

Figure 5.10 presents the first study, which shows the loss during training versus epoch with varying step sizes from 1 to 4. It can be seen that the network learned faster and better with increasing step size, with that being saturated at step size 4. Thus in this work, the step size of 3 was chosen to train the network.

Figure 5.10 presents the performance of the network in image registration for four different network architectures. The network designs were respectively: 5 levels of convolution with 32, 64, 128, 256, and 512 filter kernels, 5 levels of convolution with 16, 32, 64, 128, and 256 filter kernels, 4 levels of convolution with 32, 64, 128, and 256 filter kernels, and 4 levels of convolution with 16, 32, 64, and 128 filter kernels. It can be seen that with 5 levels of convolution, the network learned better having deeper layers with more abstraction. The filter choice of case 1 performs comparatively better than case 2, which can be expected since the former learned more features.

5.4.5 Summary

This work presents the application of the deep learning approach for PET motion correction. A slightly modified version of a multi-step recurrent deep learning technique has been adopted to train a neural network for predicting affine transformation parameters. The network was trained on an augmented dataset, and the predicted parameters were passed to a spatial transform layer to get a warped/registered image.

The performance of the network was compared against conventional iterative algorithm from ANTs software. The registration accuracy was evaluated by means of Dice index, and it was found that both the methods resulted in comparable performance with the neural network working 3x faster with CPU and 20x faster with GPU.

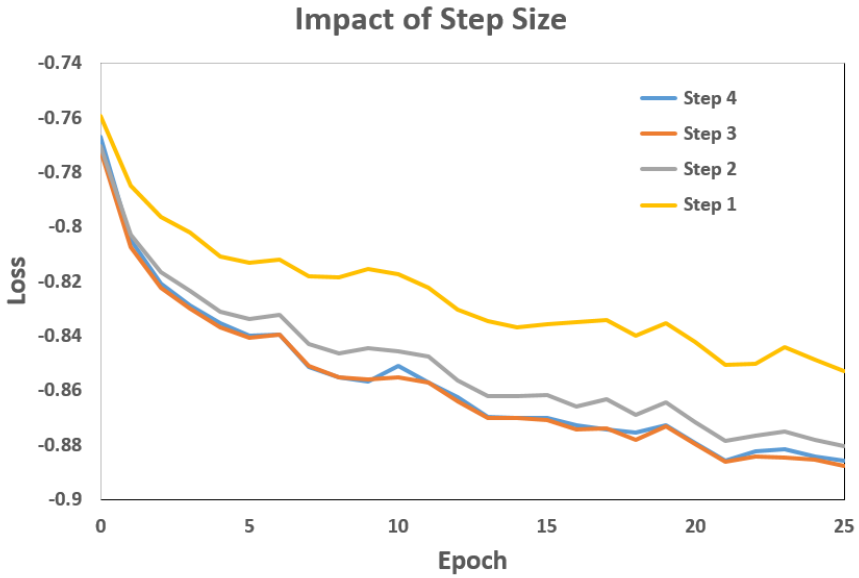


Figure 5.10: The plot of loss versus epoch shows the neural network performance with changing step sizes of 1, 2, 3, and 4, respectively. It can be seen that with increasing step size, the network performance improved and saturated at step size 4 for the dataset used.

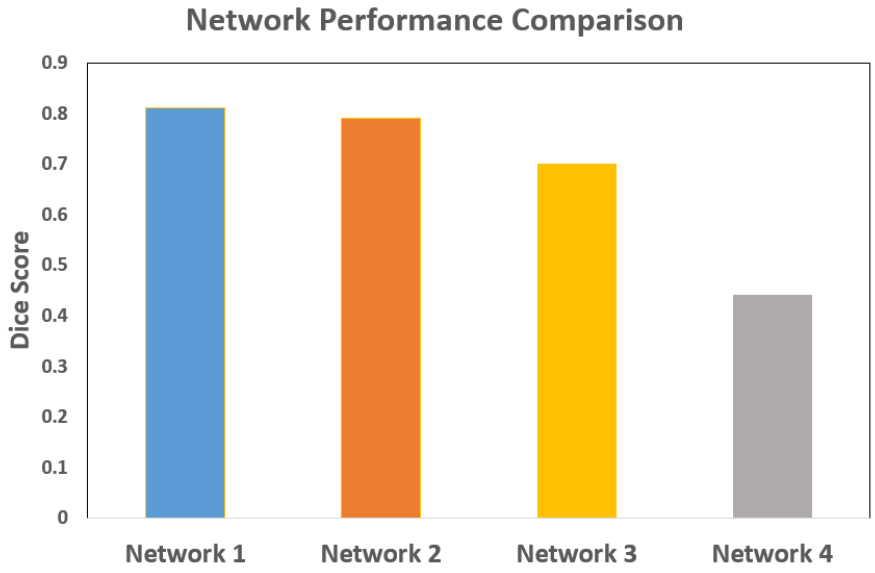


Figure 5.11: The plot shows the neural network performance with 4 different network architectures. Network 1: 5 levels of convolution with 32, 64, 128, 256, and 512 filter kernels; Network 2: 5 levels of convolution with 16, 32, 64, 128, and 256 filter kernels; Network 3: 4 levels of convolution with 32, 64, 128, and 256 filter kernels; and Network 4: 4 levels of convolution with 16, 32, 64, and 128 filter kernels. It can be seen that the network performed better with 5 levels of convolution, and the filter choice of case 1 performed comparatively better than case 2.

The performance of the application of deep learning in motion correction was further studied and evaluated by means of the original patient studies. Motion impacted data was identified using the proposed motion detection algorithm TOF-PEPT, and motion free static frames were created to perform frame-by-frame registration with the trained neural network. The aligned frames were summed together to create motion corrected data. To evaluate the motion corrected data, frame-by-frame registration was implemented with ANTs algorithm as well. Motion corrected images from the two methods were compared against the uncorrected data by means of Dice indices and line profiles. Both from qualitative and quantitative perspectives, it could be seen that the neural network yielded reliable and comparable performance with respect to the conventional iterative algorithm.

Additionally, an attempt was made to apply the predicted transformation parameters by the neural network on raw listmode data to align the shifted LORs within each static frame. An overall motion corrected listmode data with remapped LORs was created, which was then passed to the image reconstruction tools to create a motion corrected image volume. It was possible to achieve a reasonable correction with this approach, however further improvement with appropriate mapping from image dimensions to actual scanner dimensions and more accurate transformation prediction will allow producing a more clinically suitable reconstructed image data. This ability to correct the raw data itself is expected to be beneficial for other applications as well.

Furthermore, two additional studies were presented for the evaluation of the choice of network parameters. The first study was presented to show the reasoning behind the choice of step size of 3 in this work. The second study showed the network performance in image registration for four different network architectures. It was observed that the network choice with deeper layers with more filters performed better.

The intend of this overall work was to study the feasibility of the application of the deep learning technique in PET image motion correction and compare its performance with the conventional iterative approach. The findings from these studies are expected to allow exploring this approach for this kind of application in the future with more robust and improved performance.

Chapter 6

Summary and Next Steps

6.1 Summary of Research Contributions

This dissertation presents a complete data-driven solution to address motion detection during PET imaging and explores a combination of methodologies to perform motion correction.

The work first proposes a data-driven motion detection approach that extracts motion information entirely from raw data space. A modified and clinically suitable version of a positron emission particle tracking algorithm has been proposed. The algorithm incorporated the time-of-flight information from the scanning system to achieve enhanced performance. The algorithm was tested under different scenarios including the detection of different kinds of motion both internal and global with phantom and clinical studies. It was presented that the algorithm could detect both respiratory and head-neck motion with reliable and satisfactory performance. Performance accuracy was validated against the current clinical standard Anzai band and two variations of data-driven COM algorithms. Phantom studies aided in validating how the motion signals detected by the proposed method corresponded with the actual motion signal. The quantitative evaluation of the motion signals provided by different methods showed that signals detected by the TOF-PEPT algorithm correlated the best with the Anzai band with a higher correlation coefficient and a lower root mean squared error than the COM algorithms. Also, a combination of time and Fourier domain analysis was presented to demonstrate that the TOF-PEPT method allowed the detection of motion signals with lesser noise and better accuracy. Motion compensated images were

created and presented as well to realize the performance of the algorithm in minimizing the motion impacts. Improved standardized uptake values with improved and sharper images details were apparent for all the studies conducted.

The next contribution of this work was to study the application of the proposed algorithm in head-neck motion correction. The motivation was to propose a data-driven method that will allow performing the motion detection to correction in a complete automated manner in the raw listmode data space. The process included the detection of movements, creating motion free static frames post hoc, and aligning each frame to a reference position by estimating the corresponding transformation parameters. The estimation of transformation was achieved with the aid of external markers by proposing a line density based PEPT method that allowed to track the markers in each static frame. Singular value decomposition analysis on the tracked data aided in the transformation estimation. It was shown that transformation parameters were estimated with minimal error and later on were used to reposition the LORs of each static frame. With all the repositioned LORs, corrected listmode data and motion corrected image volumes were created. The performance of the motion correction was analyzed both qualitatively and quantitatively with respect to a conventional frame-by-frame registration. It was found that the proposed approach achieved comparable performance with the frame-based registration, and offered the added benefit of making the process completely automated with computational efficiency.

Finally, an unsupervised deep learning approach is presented to perform the task of 3D PET image registration for the purpose of motion correction. A multi-step learning approach was adopted. The network was trained on a dataset augmented from actual clinical data and tested on both the augmented and actual clinical datasets. The network performance in motion correction has been performed and presented both qualitatively and quantitatively. The performance was compared against conventional iterative algorithm from Advanced Normalization Tools (ANTs) software. It was found that the deep learning approach performed on par with the iterative approach with approximately twenty times faster with GPU and three times faster with CPU. Additional studies were presented on the choice of network parameters. Furthermore, the feasibility of the application of the proposed

deep learning approach in correcting raw listmode data itself was studied, and a reasonable correction was achieved.

The contribution of the work as a whole is expected to provide a complete framework to address the detection and correction of different kinds of motion that PET data encounters. This work presents a detailed insight into the proposed approaches and comprehensive analysis of the methods from different aspects which are believed to aid in the existing research of this field.

6.2 Future Work

With the advance of technology and continuing research, it is possible to expand the works presented here to a further extent. The planned next steps can be listed as follows:

- Motion detection: In this paper, a state-of-the-art photomultiplier (PMT) based PET/CT system has been used to acquire data. Next-generation SiPM-based PET system is expected to have improved TOF measurement capabilities with higher timing resolution. This should increase the accuracy of annihilation localization. The performance of the proposed motion detection algorithm with the time-of-flight weighting thus is believed to have better accuracy and performance. The future work will focus on the use of the data from the SiPM based system and evaluate the performance.

  Additional data corrections, such as attenuation correction and scatter correction, may help improve positional accuracy as well as the lower limits of detection. Incorporating such corrections along with further validation and characterization of the proposed approach across a wider range of imaging conditions is the focus of future work.

- Motion correction:

  This work addressed the raw data correction and application of the deep learning approach for affine motion. Future work could focus on expanding to address deformable registration for respiratory motion correction. Research is underway in a limited capacity to explore a combination of affine and deformable registrations of

4D PET data by means of a similar approach presented here with a U-net-like CNN architecture.

The application of the neural network in creating motion corrected raw listmode data with better accuracy and signal-to-noise ratio is the future focus. Additionally, it is intended to conduct the studies with a wide range of datasets and improve the performance with in-depth research on the choice of network architecture and parameter selection.

Bibliography

[1] Andersson, J. L. (1998). How to obtain high-accuracy image registration: application to movement correction of dynamic positron emission tomography data. *European Journal of Nuclear Medicine*, 25(6):575–586. 15

[2] Avants, B. B., Tustison, N., and Song, G. (2009). Advanced normalization tools (ANTs). *Insight j*, 2(365):1–35. 16, 74

[3] Balakrishnan, G., Zhao, A., Sabuncu, M. R., Guttag, J., and Dalca, A. V. (2018). An unsupervised learning model for deformable medical image registration. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 9252–9260. 16, 17

[4] Bertolli, O., Arridge, S., Stearns, C. W., Wollenweber, S. D., Hutton, B. F., and Thielemans, K. Data driven respiratory signal detection in PET taking advantage of time-of-flight data. In *2016 IEEE Nuclear Science Symposium, Medical Imaging Conference and Room-Temperature Semiconductor Detector Workshop (NSS/MIC/RTSD)*, pages 1–3. IEEE. 12

[5] Bertolli, O., Arridge, S., Stearns, C. W., Wollenweber, S. D., Hutton, B. F., and Thielemans, K. Improvement of the sign determination method for data-driven respiratory signal in TOF-PET. In *2017 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)*, pages 1–4. IEEE. 12

[6] Beyer, T., Tellmann, L., Nickel, I., and Pietrzyk, U. (2005). On the use of positioning aids to reduce misregistration in the head and neck in whole-body PET/CT studies. *Journal of Nuclear Medicine*, 46(4):596–602. 14

[7] Bickell, M., Buffler, A., Govender, I., and Parker, D. (2012). A new line density tracking algorithm for PEPT and its application to multiple tracers. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 682:36–41. 52

[8] Bloomfield, P. M., Spinks, T. J., Reed, J., Schnorr, L., Westrip, A. M., Livieratos, L., Fulton, R., and Jones, T. (2003). The design and implementation of a motion correction scheme for neurological PET. *Physics in Medicine and Biology*, 48(8):959. 15

[9] Bresenham, J. E. (1965). Algorithm for computer control of a digital plotter. *IBM Systems Journal*, 4(1):25–30. 52

[10] Bundschuh, R. A., Martinez-Moeller, A., Essler, M., Martinez, M. J., Nekolla, S. G., Ziegler, S. I., and Schwaiger, M. (2007). Postacquisition detection of tumor motion in the lung and upper abdomen using list-mode PET data: a feasibility study. *Journal of Nuclear Medicine*, 48(5):758–63. 12

[11] Buther, F., Dawood, M., Stegger, L., Wubbeling, F., Schafers, M., Schober, O., and Schafers, K. P. (2009). List mode-driven cardiac and respiratory gating in PET. *Journal of Nuclear Medicine*, 50(5):674–81. 14

[12] Büther, F., Vehren, T., Schäfers, K. P., and Schäfers, M. (2016). Impact of data-driven respiratory gating in clinical PET. *Radiology*, 281(1):229–238. 14

[13] Challis, J. H. (1995). A procedure for determining rigid body transformation parameters. *Journal of Biomechanics*, 28(6):733–737. 50

[14] Chamberland, M. J., deKemp, R. A., and Xu, T. (2020). Motion tracking of low-activity fiducial markers using adaptive region of interest with list-mode positron emission tomography. *Medical Physics*. 19, 27

[15] Costes, N., Dagher, A., Larcher, K., Evans, A. C., Collins, D. L., and Reilhac, A. (2009). Motion correction of multi-frame PET data in neuroreceptor mapping: simulation based validation. *Neuroimage*, 47(4):1496–1505. 15

[16] Dalca, A. V., Balakrishnan, G., Guttag, J., and Sabuncu, M. R. (2018). Unsupervised learning for fast probabilistic diffeomorphic registration. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 729–738. Springer. 17

[17] Dawood, M., Buther, F., Jiang, X., and Schafers, K. P. (2008). Respiratory motion correction in 3-d PET data with advanced optical flow algorithms. *IEEE Transactions on Medical Imaging*, 27(8):1164–1175. 12

[18] Dawood, M., Büther, F., Lang, N., Schober, O., and Schäfers, K. P. (2007). Respiratory gating in positron emission tomography: a quantitative comparison of different gating schemes. *Medical physics*, 34(7):3067–3076. 10

[19] Dawood, M., Lang, N., Jiang, X., and Schafers, K. P. (2006). Lung motion correction on respiratory gated 3-d PET/CT images. *IEEE Transactions on Medical Imaging*, 25(4):476–85. 1

[20] de Vos, B. D., Berendsen, F. F., Viergever, M. A., Sokooti, H., Staring, M., and Išgum, I. (2019). A deep learning framework for unsupervised affine and deformable image registration. *Medical Image Analysis*, 52:128–143. 16, 17

[21] de Vos, B. D., Berendsen, F. F., Viergever, M. A., Staring, M., and Išgum, I. (2017). End-to-end unsupervised deformable image registration with a convolutional neural network. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, pages 204–212. Springer. 16, 17

[22] Dice, L. R. (1945). Measures of the amount of ecologic association between species. *Ecology*, 26(3):297–302. 57

[23] Fan, J., Cao, X., Wang, Q., Yap, P.-T., and Shen, D. (2019a). Adversarial learning for mono-or multi-modal registration. *Medical image analysis*, 58:101545. 16

[24] Fan, J., Cao, X., Yap, P.-T., and Shen, D. (2019b). Birnet: Brain image registration using dual-supervised fully convolutional networks. *Medical Image Analysis*, 54:193–206. 16, 17

[25] Feng, T., Wang, J., Sun, Y., Zhu, W., Dong, Y., and Li, H. (2017). Self-gating: an adaptive center-of-mass approach for respiratory gating in PET. *IEEE Transactions on Medical Imaging*, 37(5):1140–1148. 14

[26] Feng, T., Yang, D., Zhu, W., Dong, Y., and Li, H. (2016). Real-time data-driven rigid motion detection and correction for brain scan with listmode PET. In *2016 IEEE Nuclear Science Symposium, Medical Imaging Conference and Room-Temperature Semiconductor Detector Workshop (NSS/MIC/RTSD)*, pages 1–4. IEEE. 15

[27] Fletcher, J. and Kinahan, P. (2010). PET/CT standardized uptake values (SUVs) in clinical practice and assessing response to therapy. *NIH Public Access*, 31(6):496–505. 9

[28] Goldstein, S. R., Daube-Witherspoon, M. E., Green, M. V., and Eidsath, A. (1997). A head motion measurement system suitable for emission computed tomography. *IEEE Transactions on Medical Imaging*, 16(1):17–27. 15

[29] Green, M. V., Seidel, J., Stein, S. D., Tedder, T. E., Kempner, K. M., Kertzman, C., and Zeffiro, T. A. (1994). Head movement in normal subjects during simulated PET brain imaging with and without head restraint. *Journal of Nuclear Medicine*, 35(9):1538–1546. 14

[30] Guo, Y., Bi, L., Ahn, E., Feng, D., Wang, Q., and Kim, J. (2020). A spatiotemporal volumetric interpolation network for 4d dynamic medical image. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4726–4735. 17

[31] Hamill, J. J. and Le Meunier, L. (2017). Optimal respiratory gating in medical imaging. US Patent 9,579,070. 3, 27

[32] He, J., O'Keefe, G. J., Gong, S. J., Jones, G., Saunder, T., Scott, A. M., and Geso, M. (2008). A novel method for respiratory motion gated with geometric sensitivity of the scanner in 3d PET. *IEEE Transactions on Nuclear Science*, 55(5):2557–2565. 14

[33] Jaderberg, M., Simonyan, K., Zisserman, A., et al. (2015). Spatial transformer networks. In *Advances in neural information processing systems*, pages 2017–2025. 17, 71

[34] Jani, S. S., Robinson, C. G., Dahlbom, M., White, B. M., Thomas, D. H., Gaudio, S., Low, D. A., and Lamb, J. M. (2013). A comparison of amplitude-based and phase-based positron emission tomography gating algorithms for segmentation of internal target

volumes of tumors subject to respiratory motion. *International Journal of Radiation Oncology* Biology* Physics*, 87(3):562–569. 10

[35] Jin, X., Mulnix, T., Gallezot, J.-D., and Carson, R. E. (2013). Evaluation of motion correction methods in human brain PET imaging—a simulation study based on human motion data. *Medical physics*, 40(10):102503. 15

[36] Kesner, A. L., Bundschuh, R. A., Detorie, N. C., Dahlbom, M., Ziegler, S. I., Czernin, J., and Silverman, D. H. (2009). Respiratory gated PET derived in a fully automated manner from raw PET data. *IEEE Transactions on Nuclear Science*, 56(3):677–686. 14

[37] Kesner, A. L., Chung, J. H., Lind, K. E., Kwak, J. J., Lynch, D., Burckhardt, D., and Koo, P. J. (2016). Validation of software gating: A practical technology for respiratory motion correction in PET. *Radiology*, 281(1):239–48. 14

[38] Kesner, A. L. and Kuntner, C. (2010). A new fast and fully automated software based algorithm for extracting respiratory signal from raw PET data and its comparison to other methods. *Medical physics*, 37(10):5550–5559. 14, 27

[39] Klein, G., Reutter, B., Botvinick, E., Budinger, T., and Huesman, R. (2001). Fine-scale motion detection using intrinsic list mode PET information. In *Proceedings IEEE Workshop on Mathematical Methods in Biomedical Image Analysis (MMBIA 2001)*, pages 71–78. IEEE. 12

[40] Klein, S., Staring, M., Murphy, K., Viergever, M. A., and Pluim, J. P. (2009). Elastix: a toolbox for intensity-based medical image registration. *IEEE Transactions on Medical Imaging*, 29(1):196–205. 16

[41] Krebs, J., Delingette, H., Mailhé, B., Ayache, N., and Mansi, T. (2019). Learning a probabilistic model for diffeomorphic registration. *IEEE Transactions on Medical Imaging*, 38(9):2165–2176. 17

[42] Krebs, J., Mansi, T., Delingette, H., Zhang, L., Ghesu, F. C., Miao, S., Maier, A. K., Ayache, N., Liao, R., and Kamen, A. (2017). Robust non-rigid registration through

agent-based action learning. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 344–352. Springer. 16

[43] Krebs, J., Mansi, T., Mailhé, B., Ayache, N., and Delingette, H. (2018). Unsupervised probabilistic deformation modeling for robust diffeomorphic registration. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, pages 101–109. Springer. 17

[44] Kuang, D. and Schmah, T. (2019). Faim–a convnet method for unsupervised 3d medical image registration. In *International Workshop on Machine Learning in Medical Imaging*, pages 646–654. Springer. 17

[45] Kushner, D. (2011). The making of arduino. *IEEE spectrum*, 26. 28

[46] Lamare, F., Fayad, H., Fernandez, P., and Visvikis, D. (2015). Local respiratory motion correction for PET/CT imaging: Application to lung cancer. *Medical Physics*, 42(10):5903–12. 12

[47] Li, H. and Fan, Y. (2018). Non-rigid image registration using self-supervised fully convolutional networks without training data. In *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, pages 1075–1078. IEEE. 16, 17

[48] Li, T., Zhang, M., Qi, W., Asma, E., and Qi, J. (2020). Motion correction of respiratory-gated PET images using deep learning based image registration framework. *Physics in Medicine & Biology*. 17

[49] Liao, R., Miao, S., de Tournemire, P., Grbic, S., Kamen, A., Mansi, T., and Comaniciu, D. (2016). An artificial agent for robust image registration. *arXiv preprint arXiv:1611.10336*. 16

[50] Lin, K.-P., Huang, S.-C., Yu, D.-C., Melega, W., Barrio, J. R., and Phelps, M. E. (1996). Automated image registration for fdopa PET studies. *Physics in Medicine and Biology*, 41(12):2775. 15

[51] Liu, C., Alessio, A. M., and Kinahan, P. E. (2011). Respiratory motion correction for quantitative PET/CT using all detected events with internal-external motion correlation. *Medical Physics*, 38(5):2715–23. 12

[52] Lopresti, B. J., Russo, A., Jones, W. F., Fisher, T., Crouch, D. G., Altenburger, D., and Townsend, D. W. (1999). Implementation and performance of an optical motion tracking system for high resolution brain PET imaging. *IEEE Transactions on Nuclear Science*, 46(6):2059–2067. 15

[53] Lu, Y., Gallezot, J.-D., Naganawa, M., Ren, S., Fontaine, K., Wu, J., Onofrey, J. A., Toyonaga, T., Boutagy, N., and Mulnix, T. (2019). Data-driven voluntary body motion detection and non-rigid event-by-event correction for static and dynamic PET. *Physics in Medicine and Biology*, 64(6):065002. 14

[54] Lu, Y., Naganawa, M., Toyonaga, T., Gallezot, J.-D., Fontaine, K., Ren, S., Revilla, E. M., Mulnix, T., and Carson, R. E. (2020). Data-driven motion detection and event-by-event correction for brain PET: Comparison with vicra. *Journal of Nuclear Medicine*, page jnumed. 119.235515. 16

[55] Ma, K., Wang, J., Singh, V., Tamersoy, B., Chang, Y.-J., Wimmer, A., and Chen, T. (2017). Multimodal image registration with deep context reinforcement learning. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 240–248. Springer. 16

[56] Mahapatra, D., Antony, B., Sedai, S., and Garnavi, R. (2018). Deformable medical image registration using generative adversarial networks. In *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, pages 1449–1453. IEEE. 16

[57] Manwell, S., Klein, R., Xu, T., and deKemp, R. A. (2020). Clinical comparison of the positron emission tracking (petrack) algorithm with the real-time position management system for respiratory gating in cardiac positron emission tomography. *Medical Physics*, 47(4):1713–1726. 19, 27

[58] MATLAB (2018). *9.7.0.1190202 (R2019b)*. The MathWorks Inc., Natick, Massachusetts. 23

[59] Medixant (2021). Radiant dicom viewer. 27

[60] Menke, M., Atkins, M. S., and Buckley, K. R. (1996). Compensation methods for head motion detected during PET imaging. *IEEE Transactions on Nuclear Science*, 43(1):310–317. 15

[61] Miele, E., Spinelli, G. P., Tomao, F., Zullo, A., De Marinis, F., Pasciuti, G., Rossi, L., Zoratto, F., and Tomao, S. (2008). Positron emission tomography (PET) radiotracers in oncology–utility of 18f-fluoro-deoxy-glucose (FDG)-PET in the management of patients with non-small-cell lung cancer (NSCLC). *Journal of Experimental & Clinical Cancer Research*, 27(1):1–10. 6

[62] Montgomery, A. J., Thielemans, K., Mehta, M. A., Turkheimer, F., Mustafovic, S., and Grasby, P. M. (2006). Correction of head movement on PET studies: comparison of methods. *Journal of Nuclear Medicine*, 47(12):1936–1944. 15

[63] Nair, V. and Hinton, G. E. (2010). Rectified linear units improve restricted boltzmann machines. In *Icml*. 72

[64] Nehmeh, S., Erdi, Y., Ling, C., Rosenzweig, K., Squire, O., Braban, L., Ford, E., Sidhu, K., Mageras, G., Larson, S., et al. (2002). Effect of respiratory gating on reducing lung motion artifacts in PET imaging of lung cancer. *Medical physics*, 29(3):366–371. 3, 10

[65] Noonan, P. J., Howard, J., Cootes, T. F., Hallett, W. A., and Hinz, R. (2012). Realtime markerless rigid body head motion tracking using the microsoft kinect. In *2012 IEEE Nuclear Science Symposium and Medical Imaging Conference Record (NSS/MIC)*, pages 2241–2246. IEEE. 15

[66] Olesen, O. V., Sullivan, J. M., Mulnix, T., Paulsen, R. R., Hojgaard, L., Roed, B., Carson, R. E., Morris, E. D., and Larsen, R. (2013). List-mode PET motion correction using markerless head tracking: Proof-of-concept with scans of human subject. *IEEE Transactions on Medical Imaging*, 32(2):200–209. 15

[67] Osborne, D., Acuff, S., Tumpa, T., and Hu, D. (2017). Respiratory motion correction using novel particle tracking techniques. *Journal of Nuclear Medicine*, 58(supplement 1):1362–1362. 19

[68] Osborne, D., hyeb Lee, S., Hu, D., and Acuff, S. (2015). Head and neck motion correction in PET imaging using new particle tracking techniques. *Journal of Nuclear Medicine*, 56(supplement 3):1822–1822. 19

[69] Osborne, D. R., Hu, D., and Lee, S. H. (2019). Motion correction for PET medical imaging based on tracking of annihilation photons. US Patent 10,255,684. 19

[70] Parker, D., Broadbent, C., Fowles, P., Hawkesworth, M., and McNeil, P. (1993). Positron emission particle tracking-a technique for studying flow within engineering equipment. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 326(3):592–607. 3, 19

[71] Parker, D., Dijkstra, A., Martin, T., and Seville, J. (1997). Positron emission particle tracking studies of spherical particle motion in rotating drums. *Chemical Engineering Science*, 52(13):2011–2022. 3, 19

[72] Parker, D. J. and Fan, X. (2008). Positron emission particle tracking—application and labelling techniques. *Particuology*, 6(1):16–23. 3, 19

[73] Penney, G. P., Weese, J., Little, J. A., Desmedt, P., Hill, D. L., et al. (1998). A comparison of similarity measures for use in 2-d-3-d medical image registration. *IEEE Transactions on Medical imaging*, 17(4):586–595. 72

[74] Perruchot, F., Reilhac, A., Grova, C., Evans, A. C., and Dagher, A. (2004). Motion correction of multi-frame PET data. In *IEEE Symposium Conference Record Nuclear Science 2004.*, volume 5, pages 3186–3190. IEEE. 15

[75] Phelps, M. E. (2000). Positron emission tomography provides molecular imaging of biological processes. *Proc Natl Acad Sci U S A*, 97(16):9226–33. 8

[76] Picard, Y. and Thompson, C. (1995). Digitized video subject positioning and surveillance system for PET. *IEEE Transactions on Nuclear Science*, 42(4):1024–1029. 15

[77] Picard, Y. and Thompson, C. J. (1997). Motion correction of PET images using multiple acquisition frames. *IEEE Transactions on Medical Imaging*, 16(2):137–144. 15

[78] Qiao, F., Pan, T., Clark, J. W., J., and Mawlawi, O. R. (2006). A motion-incorporated reconstruction method for gated PET studies. *Physics in Medicine and Biology*, 51(15):3769–83. 12

[79] Reddin, J. S., Scheuermann, J. S., Bharkhada, D., Smith, A. M., Casey, M. E., Conti, M., and Karp, J. S. (2018). Performance evaluation of the sipm-based siemens biograph vision PET/CT system. In *2018 IEEE Nuclear Science Symposium and Medical Imaging Conference Proceedings (NSS/MIC)*, pages 1–5. IEEE. 8

[80] Ren, S., Jin, X., Chan, C., Jian, Y., Mulnix, T., Liu, C., and Carson, R. E. (2017). Data-driven event-by-event respiratory motion correction using TOF PET list-mode centroid of distribution. *Physics in Medicine and Biology*, 62(12):4741–4755. 14, 24

[81] Rohé, M.-M., Datar, M., Heimann, T., Sermesant, M., and Pennec, X. (2017). Svf-net: Learning deformable image registration using shape matching. In *International Conference on Medical Image Computing and Computer-assisted Intervention*, pages 266–274. Springer. 16

[82] Saha, G. B. (2015). *Basics of PET imaging: physics, chemistry, and regulations*. Springer. 6

[83] Schleyer, P., Dunn, J., Reeves, S., Brownings, S., Marsden, P., and Thielemans, K. (2015). Detecting and estimating head motion in brain PET acquisitions using raw time-of-flight PET data. *Physics in Medicine and Biology*, 60(16):6441. 15

[84] Schleyer, P. J., O'Doherty, M. J., Barrington, S. F., and Marsden, P. K. (2009). Retrospective data-driven respiratory gating for PET/CT. *Physics in Medicine and Biology*, 54(7):1935–50. 14

[85] Schleyer, P. J., O'Doherty, M. J., and Marsden, P. K. (2011). Extension of a data-driven gating technique to 3d, whole body PET studies. *Physics in Medicine and Biology*, 56(13):3953. 14

[86] Shen, Z., Han, X., Xu, Z., and Niethammer, M. (2019). Networks for joint affine and non-parametric image registration. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4224–4233. 17, 70, 71, 74

[87] Shukla, A. K. and Kumar, U. (2006). Positron emission tomography: An overview. *Journal of medical physics/Association of Medical Physicists of India*, 31(1):13. 6

[88] Sokooti, H., De Vos, B., Berendsen, F., Lelieveldt, B. P., Išgum, I., and Staring, M. (2017). Nonrigid image registration using multi-scale 3d convolutional neural networks. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 232–239. Springer. 16

[89] Sorenson, J. A. and Phelps, M. E. (1987). *Physics in Nuclear Medicine*. Grune & Stratton New York. 20, 21

[90] Thielemans, K., Rathore, S., Engbrant, F., and Razifar, P. Device-less gating for PET/CT using pca. In *2011 IEEE Nuclear Science Symposium*, pages 3904–3910. IEEE. 12

[91] Thielemans, K., Schleyer, P., Dunn, J., Marsden, P. K., and Manjeshwar, R. M. Using pca to detect head motion from PET list mode data. In *2013 IEEE Nuclear Science Symposium and Medical Imaging Conference (2013 NSS/MIC)*, pages 1–5. IEEE. 15

[92] Thirion, J.-P. (1998). Image matching as a diffusion process: an analogy with maxwell's demons. 16

[93] Tumpa, T. R., Acuff, S. N., Gregor, J., Lee, S., Hu, D., and Osborne, D. R. (2018). Respiratory motion correction using a novel positron emission particle tracking technique: A framework towards individual lesion-based motion correction. In *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 5249–5252. IEEE. 18, 19

[94] Tumpa, T. R., Acuff, S. N., Gregor, J., Lee, S., Hu, D., and Osborne, D. R. (2021). A data-driven respiratory motion estimation approach for PET based on time-of-flight weighted positron emission particle tracking. *Medical Physics*, 48(3):1131. 18, 19

[95] Vercauteren, T., Pennec, X., Perchant, A., and Ayache, N. (2009). Diffeomorphic demons: Efficient non-parametric image registration. *NeuroImage*, 45(1):S61–S72. 16

[96] Visvikis, D., Barret, O., Fryer, T., Turzo, A., Lamare, F., Le Rest, C. C., and Bizais, Y. (2003). A posteriori respiratory motion gating of dynamic PET images. In *2003 IEEE Nuclear Science Symposium. Conference Record (IEEE Cat. No. 03CH37515)*, volume 5, pages 3276–3280. IEEE. 14

[97] Wang, J., Feng, T., and Tsui, B. M. Development and evaluation of data-driven respiratory gating methods with simulated list-mode PET data. In *2015 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)*, pages 1–4. IEEE. 14

[98] Wang, M., Guo, N., Schaefferkoetter, J., Yan, J., Hu, G., Zhang, H., El Fakhri, G., and Li, Q. (2015). Automatic respiratory gating of TOF-PET data using time-of-flight information. *Journal of Nuclear Medicine*, 56(supplement 3):427–427. 14

[99] Wang, M., Guo, N., Zhang, H., Elfhakri, G., Hu, G., and Li, Q. Retrospective data-driven respiratory gating for PET using tof information. In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 4520–4523. IEEE. 14

[100] Wardak, M., Wong, K.-P., Shao, W., Dahlbom, M., Kepe, V., Satyamurthy, N., Small, G. W., Barrio, J. R., and Huang, S.-C. (2010). Movement correction method for human brain PET images: application to quantitative analysis of dynamic 18f-fddnp scans. *Journal of Nuclear Medicine*, 51(2):210–218. 15

[101] Werner, P., Cohen-Mansfield, J., Braun, J., and Marx, M. S. (1989). Physical restraints and agitation in nursing home residents. *Journal of the American Geriatrics Society*, 37(12):1122–1126. 15

[102] Wiggins, C., Santos, R., and Ruggles, A. (2016). A novel clustering approach to positron emission particle tracking. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 811:18–24. 52

[103] Wiggins, C., Santos, R., and Ruggles, A. (2017). A feature point identification method for positron emission particle tracking with multiple tracers. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 843:22–28. 52

[104] Xia, K.-j., Yin, H.-s., and Wang, J.-q. (2019). A novel improved deep convolutional neural network model for medical image fusion. *Cluster Computing*, 22(1):1515–1527. 17

[105] Xu, J. and Tsui, B. M. (2015). Improved intrinsic motion detection using time-of-flight PET. *IEEE Transactions on Medical Imaging*, 34(10):2131–45. 12

[106] Yan, P., Xu, S., Rastinehad, A. R., and Wood, B. J. (2018). Adversarial image registration with application for mr and trus image fusion. In *International Workshop on Machine Learning in Medical Imaging*, pages 197–204. Springer. 16

[107] Yang, X., Kwitt, R., Styner, M., and Niethammer, M. (2017). Quicksilver: Fast predictive image registration–a deep learning approach. *NeuroImage*, 158:378–396. 16

[108] Yu, H., Jiang, H., Zhou, X., Hara, T., Yao, Y.-D., and Fujita, H. (2020). Unsupervised 3d PET-CT image registration method using a metabolic constraint function and a multi-domain similarity measure. *IEEE Access*, 8:63077–63089. 17

[109] Yu, H., Zhou, X., Jiang, H., Kang, H., Wang, Z., Hara, T., and Fujita, H. (2019). Learning 3d non-rigid deformation based on an unsupervised deep learning for PET/CT image registration. In *Medical Imaging 2019: Biomedical Applications in Molecular, Structural, and Functional Imaging*, volume 10953, page 109531X. International Society for Optics and Photonics. 17

[110] Zhao, S., Lau, T., Luo, J., Eric, I., Chang, C., and Xu, Y. (2019). Unsupervised 3d end-to-end medical image registration with volume tweening network. *IEEE Journal of Biomedical and Health Informatics*, 24(5):1394–1404. 17, 72

Vita

Tasmia Rahman Tumpa was born and brought up in Dhaka, Bangladesh. She received her B.Sc. degree in Electrical and Electronic Engineering from Bangladesh University of Engineering and Technology, Dhaka, Bangladesh, in September, 2015. She served as a lecturer in Daffodil International University, Dhaka, Bangladesh from September, 2015 to June, 2016. She joined the University of Tennessee, Knoxville in Fall, 2016 to pursue her PhD in Computer Engineering. During her graduate studies, she received the Ronald Nutt Fellowship in Imaging Formation and Processing.