Correcting for Patient Breathing Motion in PET Imaging

by

Teaghan O'Briain
B.Sc., University of Victoria, 2020

A Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of

MASTER OF SCIENCE

in the Department of Physics and Astronomy

© Teaghan O'Briain, 2022
University of Victoria

All rights reserved. This Thesis may not be reproduced in whole or in part, by photocopying or other means, without the permission of the author.

Correcting for Patient Breathing Motion in PET Imaging

by

Teaghan O'Briain

B.Sc., University of Victoria, 2020

Supervisory Committee

Dr. Magdalena Bazalova-Carter, Supervisor

(Department of Physics and Astronomy)

Dr. Carlos Uribe, Committee Member

(Department of Physics and Astronomy)

ABSTRACT

Positron emission tomography (PET) requires imaging times that last several minutes long. Therefore, when imaging areas that are prone to respiratory motion, blurring effects are often observed. This blurring can impair our ability to use these images for diagnostics purposes as well for treatment planning. While there are methods that are used to account for this effect, they often rely on adjustments to the imaging protocols in the form of longer scan times or subjecting the patient to higher doses of radiation. This dissertation explores an alternative approach that leverages state-of-the-art deep learning techniques to align the PET signal acquired at different points of the breathing motion. This method does not require adjustments to standard clinical protocols; and therefore, is more efficient and/or safer than the most widely adopted approach. To help validate this method, Monte Carlo (MC) simulations were conducted to emulate the PET imaging process, which represent the focus of our first experiment. The next experiment was the development and testing of our motion correction method.

A clinical four-ring PET imaging system was modelled using GATE (v. 9.0). To validate the simulations, PET images were acquired of a cylindrical phantom, point source, and image quality phantom with the modeled system and the experimental procedures were also simulated. The simulations were compared against the measurements in terms of their count rates and sensitivity as well as their image uniformity, resolution, recovery coefficients, coefficients of variation, contrast, and background variability. When compared to the measured data, the number of true detections in the MC simulations was within 5%. The scatter fraction was found to be (31.1 $\pm$ 1.1)% and (29.8 $\pm$ 0.8)% in the measured and simulated scans, respectively. Analyzing the measured and simulated sinograms, the sensitivities were found to be 10.0 cps/kBq and 9.5 cps/kBq, respectively. The fraction of random coincidences were

19% in the measured data and 25% in the simulation. When calculating the image uniformity within the axial slices, the measured image exhibited a uniformity of (0.015 ± 0.005), while the simulated image had a uniformity of (0.029 ± 0.011). In the axial direction, the uniformity was measured to be (0.024 ± 0.006) and (0.040 ± 0.015) for the measured and simulated data, respectively. Comparing the image resolution, an average percentage difference of 2.9% was found between the measurements and simulations. The recovery coefficients calculated in both the measured and simulated images were found to be within the EARL ranges, except for that of the simulation of the smallest sphere. The coefficients of variation for the measured and simulated images were found to be 12% and 13%, respectively. Lastly, the background variability was consistent between the measurements and simulations, while the average percentage difference in the sphere contrasts was found to be 8.8%. The code used to run the GATE simulations and evaluate the described metrics has been made available (https://github.com/teaghan/PET_MonteCarlo).

Next, to correct for breathing motion in PET imaging, an interpretable and unsupervised deep learning technique, FlowNet-PET, was constructed. The network was trained to predict the optical flow between two PET frames from different breathing amplitude ranges. As a result, the trained model groups different retrospectively-gated PET images together into a motion-corrected single bin, providing a final image with similar counting statistics as a non-gated image, but without the blurring effects that were initially observed. As a proof-of-concept, FlowNet-PET was applied to anthropomorphic digital phantom data, which provided the possibility to design robust metrics to quantify the corrections. When comparing the predicted optical flows to the ground truths, the median absolute error was found to be smaller than the pixel and slice widths, even for the phantom with a diaphragm movement of 21 mm. The improvements were illustrated by comparing against images without motion and com-

puting the intersection over union (IoU) of the tumors as well as the enclosed activity and coefficient of variation (CoV) within the no-motion tumor volume before and after the corrections were applied. The average relative improvements provided by the network were 54%, 90%, and 76% for the IoU, total activity, and CoV, respectively. The results were then compared against the conventional retrospective phase binning approach. FlowNet-PET achieved similar results as retrospective binning, but only required one sixth of the scan duration. The code and data used for training and analysis has been made publicly available (https://github.com/teaghan/FlowNet_PET).

The encouraging results provided by our motion correction method present the opportunity for many possible future applications. For instance, this method can be transferred to clinical patient PET images or applied to alternative imaging modalities that would benefit from similar motion corrections. When applied to clinical PET images, FlowNet-PET would provide the capability of acquiring high quality images without the requirement for either longer scan times or subjecting the patients to higher doses of radiation. Accordingly, the imaging process would likely become more efficient and/or safer, which would be appreciated by both the health care institutions and their patients.

Contents

List of Tables

List of Figures

Acknowledgements

I would like to give a special thank you to Carlos Uribe and Ioannis Sechopoulos who each provided a unique set of expertise that were vital to the success of this project. Their willingness to take time out of every week to help guide our work is very much appreciated.

This project would not have been possible without the support from Magdalena Bazalova-Carter, who was always willing to contribute her endless amount of knowledge and strong leadership. I consider myself lucky to have had such an understanding and personable supervisor who also exhibits excellence in her field.

While working on this project, Christian Michel passed away. The work done in this dissertation could not have been possible without the generous contribution of his time and kindness, which he was always willing to share.

I would like to thank Antoi Meeuwis and Daphne Lobeek at Radboud University Medical Centre for acquiring the experimental measurements used to validate the simulations. I am also very grateful for the assistance provided by Judson Jones and Antje Schulte from Siemens Healthineers, which helped us build an understanding of the scanner model as well as the reconstruction code.

Paul Segars from Duke University provided ongoing support for the XCAT phantom and graciously allowed for the training and tests sets to be made publicly available. Lastly, I would like to express my gratitude to Erik Aarntzen who provided the clinical activity distributions that we used to define the range of activities for the digital phantoms.

Dedication

For the late Christian Michel.

Chapter 1

Introduction

Photography in the 19th century often required exposure times that were several minutes long. If that were the case today, my family portraits would represent a blurred mess of four brothers wrestling and a helpless mother trying to tear us apart. Luckily, modern digital photography has progressed to have exposures that last a fraction of a second rather than minutes. However, not all imaging modalities have benefited from these types of developments. For instance, Positron Emission Tomography (PET) is an imaging technique often used to diagnose and track the progression of cancer as well as design treatment plans for radiation therapy. Rather than relying on the optical light around us, PET depends on a radioactive decay process to generate the "light" used to produce images. For reasons that will be made more clear in this dissertation, the "exposure" times in PET still last on the order of minutes rather than fractions of a second. While the imaging subject in PET is unlikely to be attacked by an older brother, the patient will hopefully be breathing, which imposes the same limitation: imaging a moving object results in blurring. Now, rather than an irritated mother, you may encounter a disappointed oncologist who cannot determine where exactly a cancerous tumor is located and how large it really is. It is difficult to state which situation is worse; I personally would like to avoid an angry mother, however, the families of lung cancer patients would likely prefer a well-informed oncologist.

Luckily, this dissertation provides a way for us to now have both.

The task that we aimed to solve is how to produce PET images that are largely free of the breathing motion artifacts seen when adopting standard PET procedures. Of course, there are other methods to accomplish this, such as Retrospective Phase Binning (RPB). In this scenario, the PET image is only created from the data acquired when the patient's breath is at a particular phase of the breathing cycle, while the rest of the data is disregarded. If this sounds inefficient, it is. In order to produce high quality images, the imaging protocol has to be adjusted to either include longer imaging times or subject the patient to more radiation - neither of which are desirable. You may ask, "why not break the image data into frames like a video and align the frames to produce an image that is largely free of these blurring effects?" To this, I would respond, "that sounds like a wonderful idea for a thesis."

The solution is obviously a bit more involved. For instance, how do you perform the alignments of the separate frames? Additionally, how do you validate that the alignments are correct when applied to patient data? These questions provided the motivation for my thesis and will be explored in-depth throughout this dissertation. To explain briefly here – in keeping with the current trend of modern day science – the problem of aligning the images was posed as a nail for the all-encompassing hammer of machine learning. Namely, the PET data was split into several frames where each frame had a limited amount of motion within it. A neural network was then constructed to predict the pixel-wise shift between each PET frame, providing the ability to shift all of the acquired data to be aligned with each other. Therefore, 100% of the data is utilized to produce a PET image free of breathing motion artifacts.

The task of validating these corrections was accomplished by using digital phantoms rather than clinical data. Adopting digital phantoms has several advantages such as (1) having precise knowledge of the anatomical structures within the patient

and (2) the ability to simulate patients without motion to compare against. Furthermore, to improve the realism of these digital phantoms, they can be coupled with Monte Carlo (MC) simulations of a clinical imaging system. If the PET scanner is modelled precisely, these MC simulations can provide accurate representations of the physical processes involved in the imaging procedure and closely resemble images acquired from the clinic.

Accordingly, the objective of this dissertation is to (1) develop and validate a MC pipeline for a clinical PET system and (2) engineer a neural network that is capable of correcting for breathing motion in PET imaging. The former task was accomplished by modelling a Siemens PET scanner and comparing the simulated data to clinical measurements. These evaluations resulted in a close match between the two in terms of the detector alignment and several image quality metrics. To solve the second task, an unsupervised learning algorithm was designed and applied to digital phantom data as a proof-of-concept. Comparing the results of our method to RPB, our approach produced similar quality images, but required a much shorter scan duration. Importantly, since the method is unsupervised, it can be transferred to clinical data with limited human intervention.

Before explaining the details of how these tasks were carried out, a background on PET Imaging, Monte Carlo methods, and the adopted machine learning procedures are provided in Chapter 2. In Chapter 3, a study is presented on the development of a MC pipeline designed to emulate a clinical scanner. This study includes several tests where the simulations are validated against experimental measurements. Chapter 4 includes a second study where the proposed motion correction method is applied to digital phantom data in a proof-of-concept. A discussion on these experiments and future applications of this method are given in Chapter 5. Lastly, the claims and results are summarized in Chapter 6 along with a discussion on the clinical

implications of our method.

Chapter 2

Background

2.1 Positron Emission Tomography

This chapter will provide some background information on PET imaging that is necessary to understand its use in this dissertation. Additionally, the motivation for our work will be illuminated by explaining how the PET imaging process occurs.

2.1.1 The Physics

Luckily for the reader, the theory behind PET is actually quite fascinating, at least in comparison to, say, computed tomography (CT) or – heaven forbid – ultrasound imaging. This modality adopts modern physics in a way that impacts our ability to – among other tasks – detect cancer and quantify the metabolic activity of cancer cells within patients. To explain how this is done, we must begin with radioactive decay. Specifically, when a nucleus is neutron deficient, it can undergo what is known as β^+-decay. In this decay process, a proton (p) decays into a neutron (n), an electron neutrino (ν_e), and a positron (β^+):

$$p \rightarrow n + \beta^+ + \nu_e \ . \tag{2.1}$$

Therefore, an unstable isotope reduces its number of protons by one, but increases

in the number of neutrons:

$$ {}^{M}_{Z}X \rightarrow {}_{Z-1}{}^{M}X + \beta^{+} + v_e \ . \tag{2.2} $$

An important quantity for PET is the rate that the radionuclide decays, known as the activity, $\mathcal{A}$. This activity is proportional to the number of remaining parent nuclei, N, with a proportionality constant called the decay constant, λ. The activity can then be written as

$$ \mathcal{A} = -\frac{dN}{dt} = \lambda N \ . \tag{2.3} $$

The solution to this differential equation provides an equation for the number of remaining radionuclides after time, t:

$$ N(t) = N_0 e^{-\lambda t} \ , \tag{2.4} $$

where N_0 is the initial number of radionuclides at time $t = 0$. Equation 2.4 provides physical intuition for how the decay constant can be understood. Namely, the decay constant is related to the the half-life, $t_{1/2}$, which is time that it takes for a population of radioisotopes to reach half of the original population. This relationship can be written as

$$ \lambda = \frac{\ln(2)}{t_{1/2}} \tag{2.5} $$

Before the reader loses hope in this theory section being remotely interesting, this is where the beauty of the physics arises. For the purposes of PET, the importance of the β^{+}-decay process lies within one of the biproducts: the positron. The positron is the antimatter counterpart of an electron; it has nearly identical properties except

that its charge is positive rather than negative. After the positron is produced, it travels and loses kinetic energy (primarily through Coulomb interactions) until it reaches the end of its range and interacts with an electron. Importantly, if an antimatter particle comes into contact with its matter counterpart, the two annihilate each other and the only remaining trace of their prior existence is the energy that was once contained in their mass. This energy now exists as in the form of two photons. Why two? Well, we established how energy conservation is achieved, but momentum also needs to be conserved. In more detail, since the positron loses most of its kinetic energy prior to annihilation, the initial momentum is close to zero, and therefore, the final momentum should be too. Hence, two photons are required to carry an equal amount of energy in opposite directions to conserve the momentum within this process. Since the rest-mass energy of both the electron and positron is 511 keV, these annihilation photons also have an energy of 511 keV. The process of β^+-decay producing two annihilation photons is illustrated in the left diagram of Figure 2.1. While there are some subtle details regarding the kinetic energy of the positron not being quite zero and the photons not travelling directly in opposite directions, these are not all that relevant to the main objective of this thesis.

According to this decay and annihilation process, two 511 keV photons travelling in opposite directions are a signature of the accumulation of the radionuclide. Of course, we cannot detect these photons at their source, but we can surround the patient with rings of detectors that aim to register *coincidence detections*. Coincidence detections are two photons that are both absorbed by the detector crystals within a certain amount of time (often around 4 ns). By drawing a line of response (LOR) between the two detectors, the location of the annihilation event can be limited to being somewhere along this line. This annihilation coincidence detection (ACD) process is outlined in the right diagram of Figure 2.1.

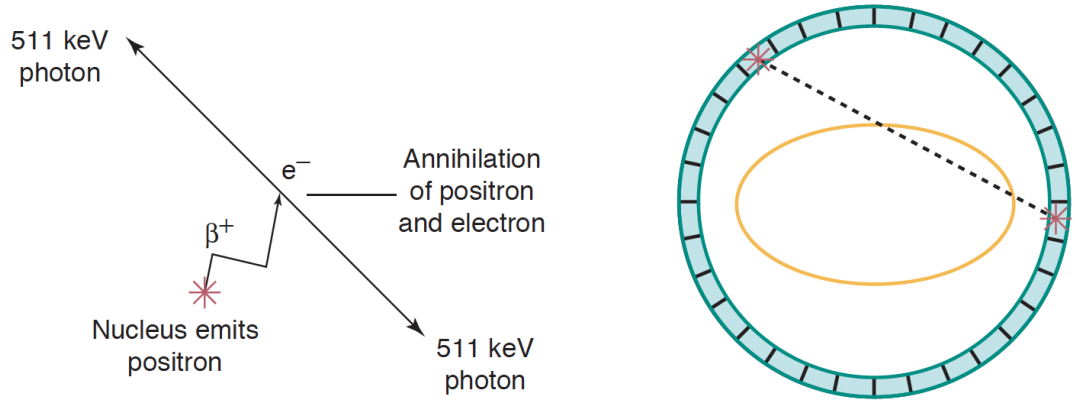


Figure 2.1: Annihilation Coincidence Detection: Following β^+-decay, the positron annihilates with an electron and two 511 keV photons are produced with 180 degree separation (**left**). These photons are detected almost simultaneously within the ring of detectors, which registers a coincidence detection (**right**). The annihilation is presumed to have occurred along the line of response (LOR) connecting these two detectors. Obtained from [1].

Detections are made through the use of scintillation crystals coupled to photomultiplier tubes (PMTs). The PMTs produce a signal that reflects the position and time of each detection along with the deposited energy. The timing information is used to identify detections that are in coincidence, whereas the energy signal can be useful to remove photons that are not annihilation photons or have been scattered. The type of material used for the scintillation crystals is chosen based on several factors including the (1) mass density, (2) time required to emit light, (3) linear attenuation of 511 keV photons, and (4) energy conversion efficiency. A common choice for these crystals is lutetium oxyorthosilicate (Lu2SiO4O, abbreviated as LSO) activated with cerium, which is the crystal material used in the scanner investigated in Chapter 3. The relevant properties of LSO are shown in Table 2.1.

With a large accumulation of activity within the patient, there are many β^+-decays constantly occurring, making the ACD process prone to detecting *true coincidences*, *scatter coincidences*, as well as *random coincidences*. A true coincidence is when

Table 2.1: A summary of the physical properties of LSO.

Property	Value
Mass density	7.4 g/cm^3
Linear attenuation of 511 keV photons	11.4 mm
Light output	30000 photons/MeV
Decay time	40 ns

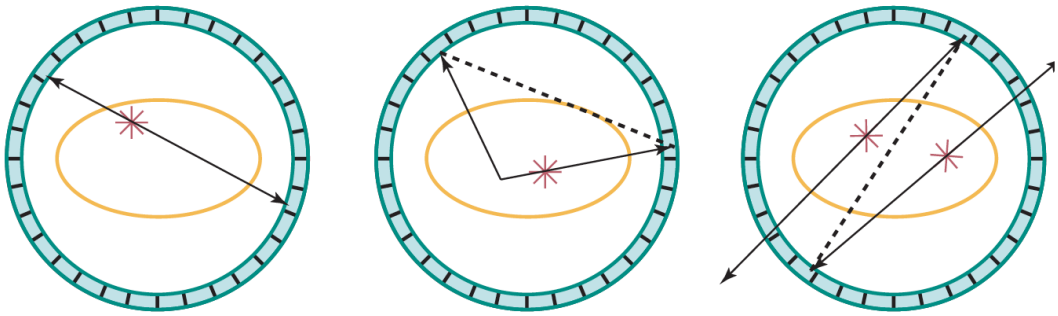


Figure 2.2: The different types of coincidence detections: true (**left**), scatter (**centre**), and random (**right**). The photon trajectories are represented by solid arrows, whereas the dotted lines represent the lines of response (LORs). Obtained from [1].

the emitted photons come from the same β^+-decay and do not scatter. When one or both of these photons scatter before being registered, this is known as a scatter coincidence. Alternatively, when the simultaneously detected photons are emitted from different nuclear decays, these are called random coincidences. These three scenarios are illustrated in Figure 2.2, which makes it evident that the LORs from both scatter and random coincidences do not correctly include the actual location of the decay event(s).

Another factor that greatly impacts the data acquisition is the attenuation of the annihilation photons. In the energy ranges seen in PET, this attenuation is often the result of interactions such as photoelectric absorption, Rayleigh scattering, Compton scattering, or pair production. The probability of a photon travelling a distance of x without interacting with a material with a linear attenuation of μ is $e^{-\mu x}$. In PET, both photons have to escape the patient for a coincidence to be registered.

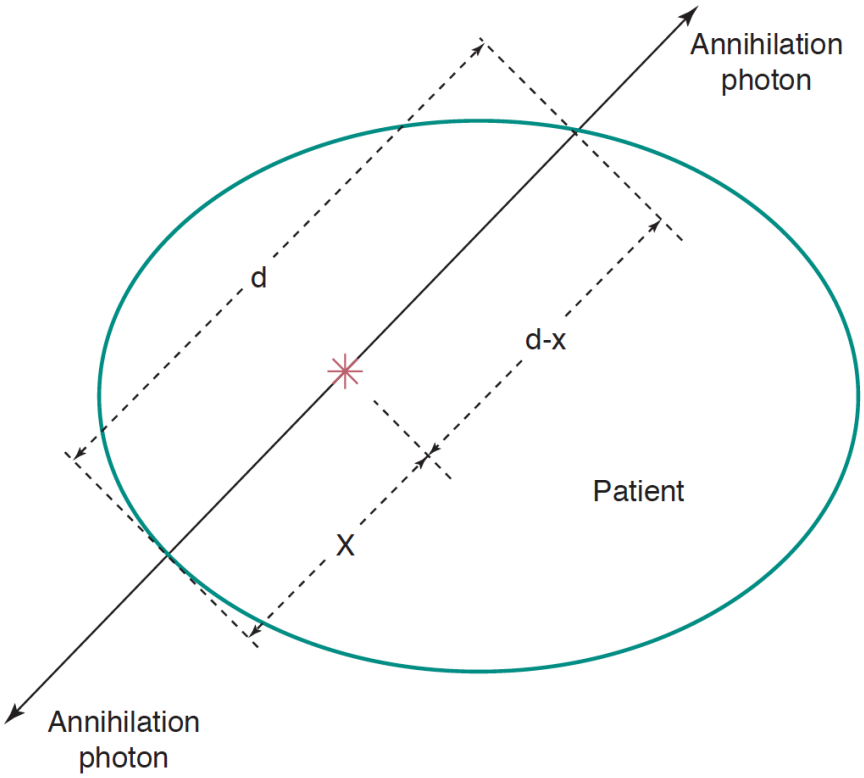


Figure 2.3: Attenuation of the annihilation photons. Obtained from [1].

For instance, as illustrated in Figure 2.3, if the patient is uniform and the combined path length of the two photons through the patient is d, then the probability of both photons escaping is

$$(e^{-\mu x}) \cdot (e^{-\mu(d-x)}) = e^{-\mu d} \, . \tag{2.6}$$

The effect of attenuation on PET images is mainly loss of information, especially in the areas that are surrounded by attenuating material. However, if the attenuating properties within the patient are estimated accurately, the attenuation effect is highly predictable. Accordingly, prior to the PET data collection, a CT image is often acquired in order to create an attenuation map (or μ-map). This μ-map allows for the attenuation to be estimated throughout the patient to facilitate an attenuation correction. These systems that combined PET and CT are known as PET/CT

scanners.

Correcting for scatter and random coincidences as well as attenuation will be discussed in Section 2.1.2. However, we have yet to address a particularly important question: "how does any of this translate into cancer imaging?" Up until now, we have discussed how β^+-decay creates 511 keV photons that can be simultaneously registered by detectors surrounding the patient. Determining the LOR between these detections starts to provide some insight as to where the decay event occured within the patient, but why does this decay occur within tumors and not elsewhere? Well, β^+-decay does not tend to naturally occur at high rates within humans anywhere - within cancer or otherwise. However, β^+ emitters can be injected into a patient; the amount of activity administered to the patient is called the *injected activity*. If these radioisotopes are disproportionately taken up by cancer cells, then the resulting decay events will be traced back to the tumor. Luckily, the radiochemical (^{18}F)2-fluoro-2-deoxy-D-glucose (FDG for short) has been developed for this specific purpose. Namely, ^{18}F is a β^+ emitter, which is substituted into the glucose molecule, FDG. Since cancer cells are working so hard to kill us, they absorb glucose at a high rate to support their metabolically demanding activity. Therefore, FDG satisfies two important requirements for a PET radiopharmaceutical in oncology: (1) it is a β^+-emitter and (2) it is disproportionately taken up by cancer cells (as well as other cells with high metabolic activity, of course). Some relevant properties of ^{18}F are listed in Table 2.2. There are other advantages of using FDG over alternative radiopharmaceuticals, but these are not particularly relevant to our project; FDG is the best option for our purposes, so we use it.

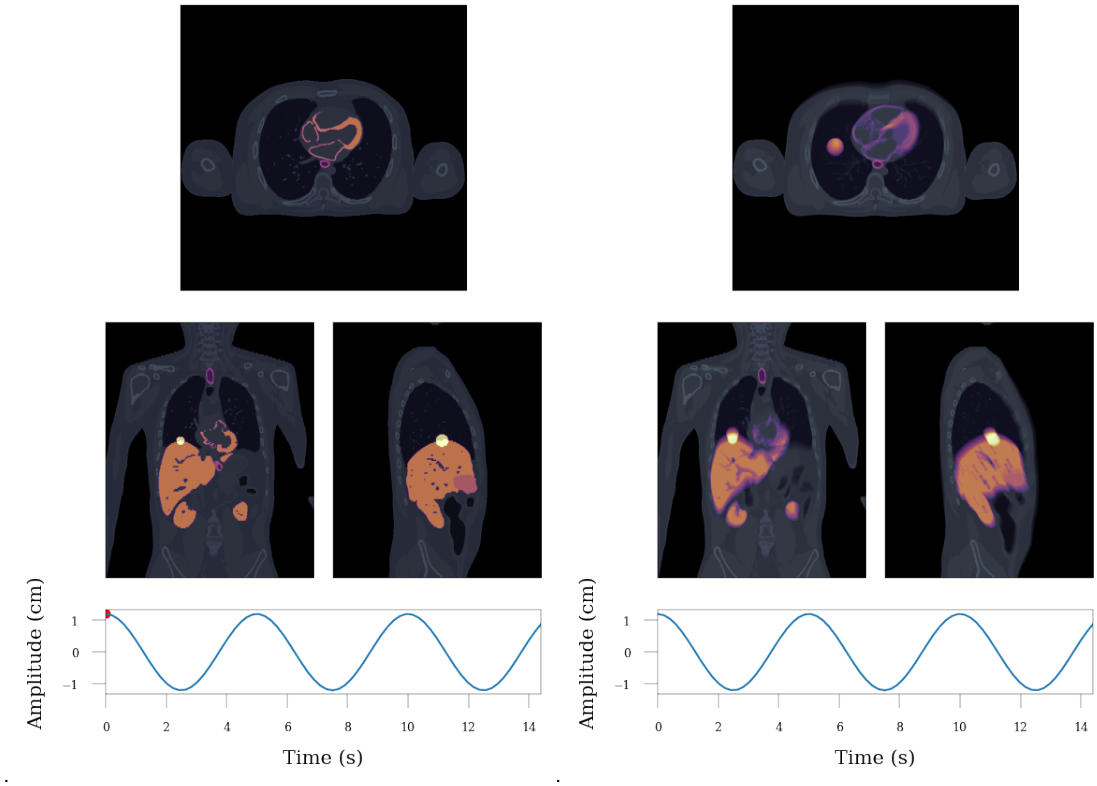
One aspect of FDG that *is* relevant to this project is its half-life of 109.7 minutes. As discussed previously, the half-life of a radioisotope relates its activity to the number of remaining isotopes. In PET, a higher activity means that signal can be

Table 2.2: A summary of the decay properties of ^{18}F.

Property	Value
Half-life	109.7 min
Specific activity	75-185 MBq/μmole
Mode of generation	Cyclotron
β^+ energy	0.635 MeV
Injected activity for adults	150-750 MBq

acquired more quickly, but also subjects the patient to a more harmful dose of radiation. Therefore, there is a trade-off between these two that needs to be optimized. Importantly, FDG does not decay too slowly, making it possible to produce diagnostic quality images in a reasonable amount of time. That being said, the acquisition times are still on the order of minutes rather than fractions of a second - and here lies the problem we aim to solve. Namely, during this acquisition period, the patient will hopefully be breathing; otherwise, the oncologist's decision to pursue PET imaging may be questioned. As shown in Figure 2.4, this breathing motion results in blurring in the final PET image. This is especially true when imaging areas that are more prone to respiratory motion; therefore, this problem is particularly relevant to cancers such as those found in the lungs, esophagus, pancreas, and liver.

The blurring in PET images provides a multitude of issues in oncology. For instance, oncologists will have a more difficult time tracking the progress of the cancer and determining how effective prior treatments have been. This is because the size of a tumor is difficult to determine when the tumor is blurred in the PET image. Furthermore, planning treatments becomes more difficult when radiologists cannot accurately determine how large the tumor is or where it is exactly located. Consequently, producing PET images free of these blurring effects would be beneficial to oncology. Existing methods that aim to solve this problem will be discussed in Chapter 4 along with the development of our own method. Spoiler alert: our method is - objectively - awesome.



(a) (b)

Figure 2.4: Breathing motion artifacts in PET: (a) a GIF of a breathing patient across three breath cycles and (b) the motion averaged activity distribution.

2.1.2 Image Reconstructions

To understand how PET images are generated from the coincidence detections, it is useful to first understand how the data is organized. This will also be relevant to some of the experiments shown in Chapter 3. Recall that when two detections of 511 keV photons occur within a predefined time window, they are determined to be in coincidence. The line that joins these two detectors represents the LOR. Within a single plane, at each angle, ϕ, around the the imaged object, $f(x, y)$, a projection, $p(s, \phi)$, is constructed by computing the line integrals of all parallel LORs (located at a given s from the midpoint) [2]. A *sinogram* is formed by organizing the projections into rows - each row corresponding to a different ϕ. Figure 2.5 shows how a single

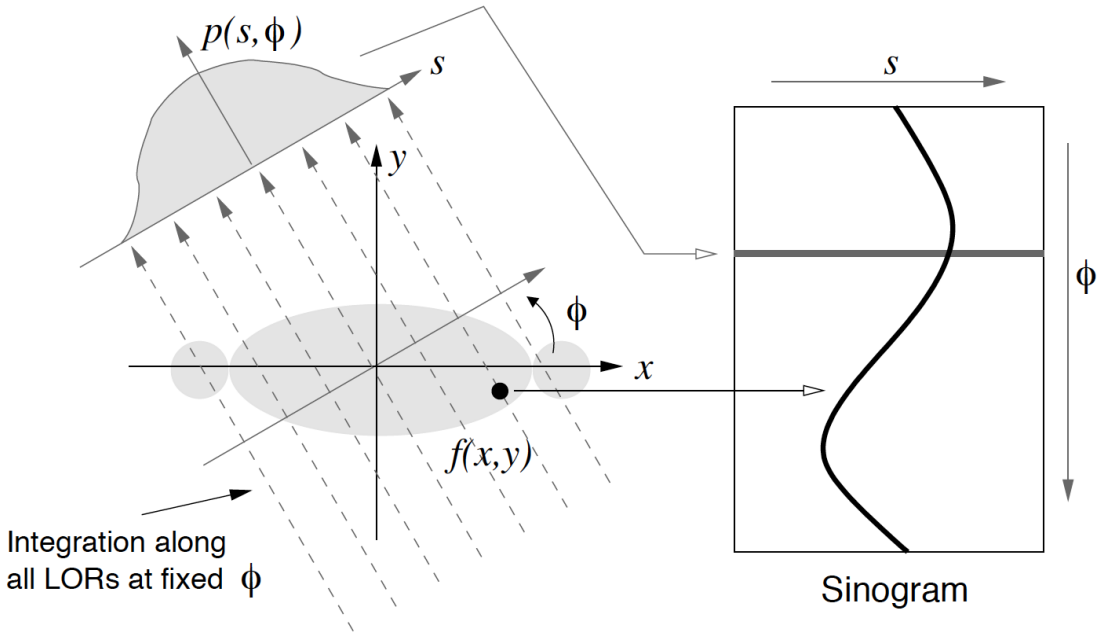
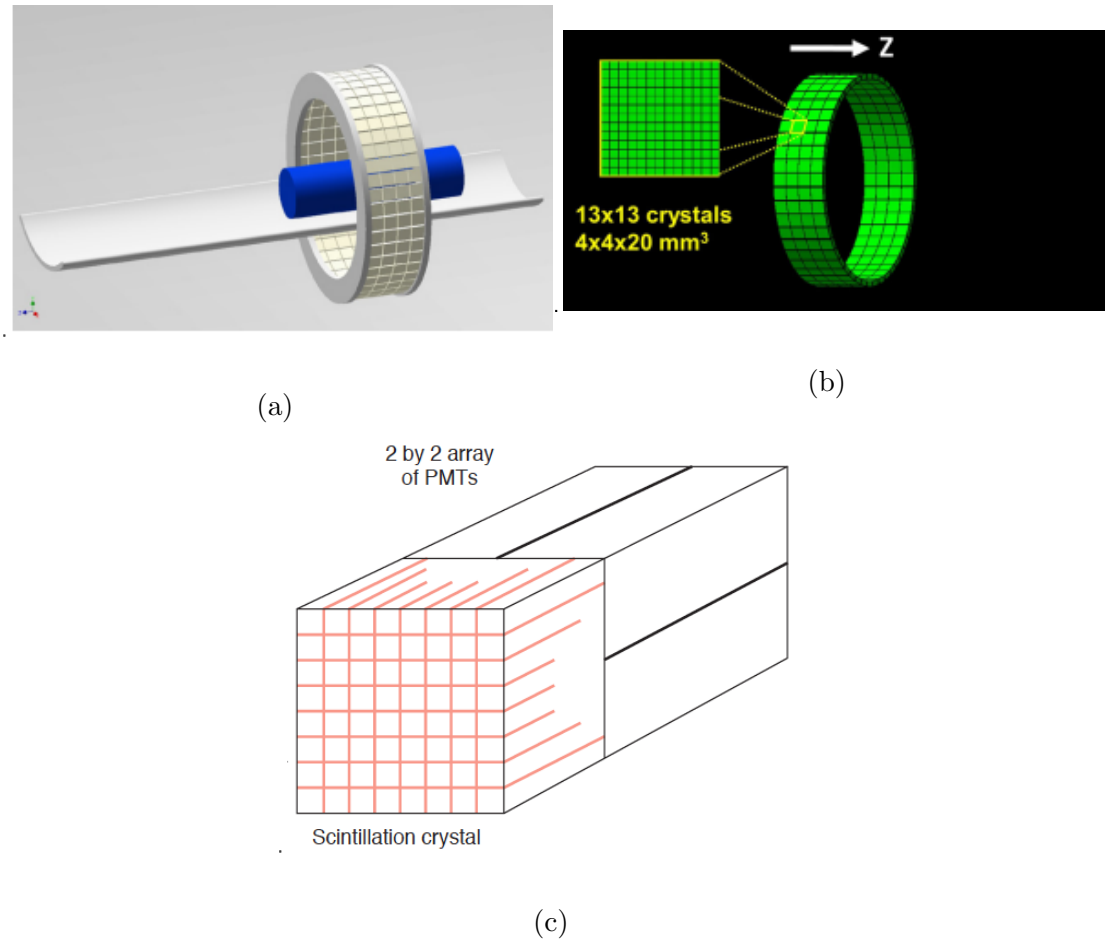


Figure 2.5: A 2D object, $f(x, y)$, is imaged by collecting detections exterior to the object. At a given angle, ϕ, a projection, $p(s, \phi)$, is formed by integration along all parallel LORs (*i.e.*, along each s). Each projection represents a row in the resulting sinogram. Obtained from [2].

point in the object creates a sinusoid in the sinogram; of course, an object is made up of many "points" that will produce an array of sinusoids sumperimposed on each other. By repeating this process for multiple planes, an axial dimension is added to create a three dimensional (3D) sinogram, where each voxel of the sinogram represents the number of counts along a particular LOR connecting two unique detector crystals.

There can be many rings of detectors in a PET system, which means that coincidences can be registered between detectors that are not within the same axial plane. For instance, Figure 2.6 shows the crystal orientation of the scanner investigated in Chapter 3. Note that this scanner is referred to as a "4-ring" scanner, however, each of these rings consists of 13 rows of crystals, providing a total of 52 rows of crystals. The registering of coincidences between detectors that are out-of-plane from each other gives rise to what is called 3D PET acquisition. In this scenario, a single detec-



(a) (b) (c)

Figure 2.6: The (a) orientation of the rings in the Siemens Biograph mCT scanner (obtained from [3]) where each block in the rings is composed of a 13×13 array of crystals (obtained from [4]) that are connected to (c) four PMTs (obtained from [1]).

tor can simultaneously register coincidences with detectors in the same axial plane as well as detectors offset from its axial plane. Furthermore, additional sinogram *segments* can be generated by "viewing" the patient at different angles. This notion is depicted in Figure 2.7 where five segments are created by viewing the patient at five different angles. The combined segments are called a *Michelogram* named after the late Christian Michel, who graciously provided support for the work shown in Chapter 3. Michelograms are often still referred to as sinograms and the terms are used interchangeably.

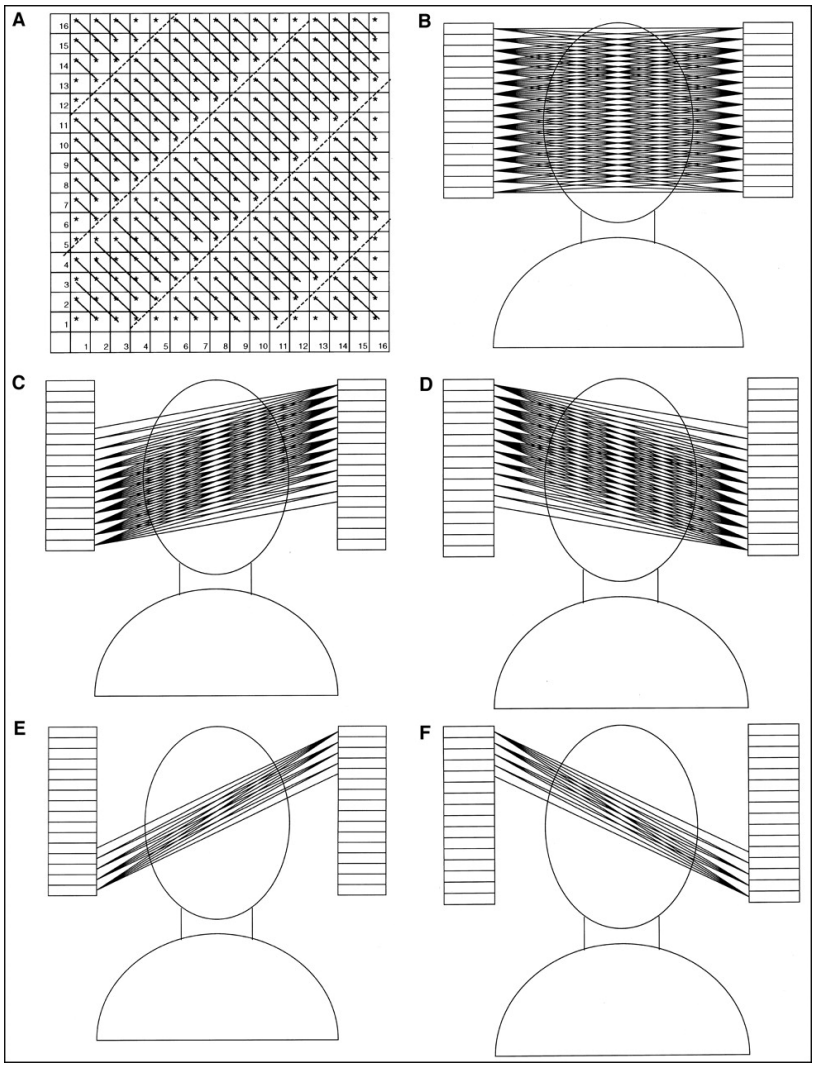


Figure 2.7: (A) A Michelogram created from 16 rings of PET detector. The Michelogram includes five segments by imaging the patient at five different angles (B-F). Obtained from [5].

While, in theory, sinograms contain all of the information present in PET images, they are difficult to interpret and infer meaning from. Therefore, it is convenient to use this sinogram data to estimate the distribution of activity within the patient. As evident in Figure 2.1, a single coincidence detection only provides a LOR that the original annihilation event likely occurred along, but the exact location is unknown. However, when a large enough number of these coincidences are detected at different

angles, the locations where a large amount of activity has accumulated become more clear. The process of estimating the activity distribution from the sinogram is called *image reconstruction*.

Rather than going through the history of the different image reconstruction algorithms, we will just briefly discuss the most relevant algorithm in PET and the only one used in this study: ordered subset expectation maximization (OSEM) [6]. The expectation maximization is a process that starts with a crude estimate of the image (*i.e.*, the activity distribution within the patient) and iteratively improves that estimate by (i) computing what the forward projection (the sinogram) would look like for that activity distribution, (ii) comparing this estimated sinogram to the measured sinogram data, (iii) adjusting this projection, (iv) back-projecting the new sinogram estimate to a new image, and (v) use this new estimate of the image as the starting point of the next iteration. The comparisons in step (ii) are accomplished by calculating the ratio of the measured counts along an LOR to the estimated counts. Rather than performing this optimization process on the entirety of the projection data all at once, the process is sped up by sorting the projections into equally sized subsets. Accordingly, the resulting image from the first subset is used as the starting estimate for the second subset, and so on.

During this iterative process, several corrections can be applied; most notably, corrections for scatter and random coincidences as well as attenuation corrections. The random coincidences are estimated by collecting *delayed coincidences* that follow the same conditions as normal coincidence detections, but with a time delay offset (typically around 1 μs). Accordingly, these delayed coincidences can only occur if the two events *are not* true coincidences. Subtracting the estimate of random coincidences from the sinogram data reduces the magnitude of the effect of random coincidence, but the statistical nature of this noise remains. Similarly, the amount of scatter in

different locations within the patient can be estimated from CT images and subtracted from the measured data. This CT image also provides the necessary information to estimate the amount of signal that has been attenuated; therefore, this attenuated signal can be artificially enhanced to better resemble the true activity distribution within the patient.

The image reconstructions conducted for this dissertation were accomplished using the e7-tools provided by Siemens. This toolkit includes command line programs that allow for several data-processing tasks to be accomplished including the creation of μ-maps from CT images and the reconstruction of PET images from sinogram data using the OSEM algorithm. To perform these reconstructions, the e7-tools require the sinogram data, a CT image, and a normalization file. The normalization file consists of information necessary to account for the varying sensitivity of a scanner across the sinogram and this quality control is updated daily. The e7-tools expect all of these files to be in the Interfile format, which is often used when exchanging nuclear medicine data and consists of a header file with information describing what is stored in a separate data file. However - when retrieved from the clinical scanner - files such as the sinogram data are often stored in Digital Imaging and Communications in Medicine (DICOM) format, which is the standard for the management of medical imaging data. Luckily, the e7-tools come equipped with the a JavaScript utility JSRecon, which automatically converts DICOM files to the Interfile-Format.

With all of the correct files in the appropriate format, the e7-tools first creates a μ-map from the CT image. Then, the sinogram data can be used to estimate the activity distribution within the imaged object using the OSEM algorithm. This process includes applying the normalization factor and correcting for scatter and random coincidences as well as using the μ-map to correct for attenuation. The resulting PET image is stored in Interfile format, which is easily loaded into memory

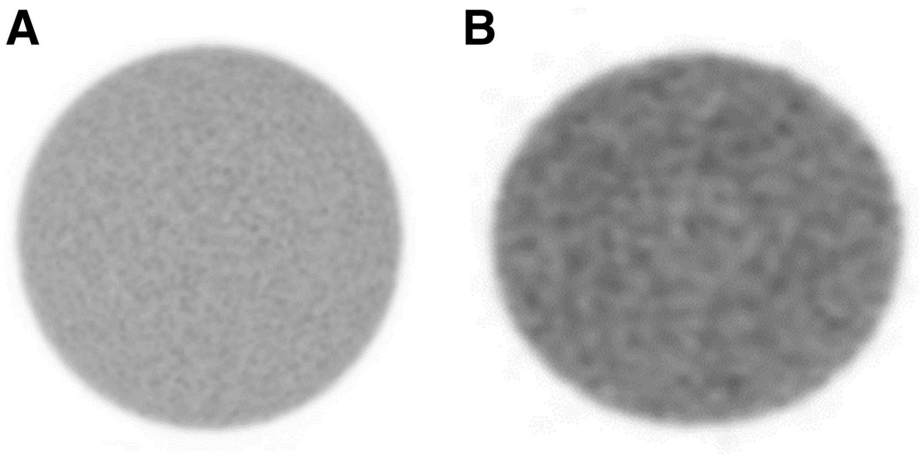


Figure 2.8: Image uniformity: Image A is judged to have excellent uniformity, whereas Image B is judged to be good. Obtained from [7].

using Python for further analysis.

2.1.3 Image Analysis

On this note, there are several image analysis metrics that are useful for assessing the quality of the images. These include image uniformity, resolution, and activity recovery. The image uniformity evaluates how consistently a scanner can reproduce the activity within the entire field of view. As will be shown in Chapter 3, this is often assessed by imaging a large volume of uniform activity and determining how the average activity in the image varies in different regions across the entire volume. For instance, Figure 2.8 shows how a degradation in uniformity visually effects the image quality. Image resolution is used to assess how accurately a scanner can distinguish between two closely located objects and - correspondingly - how well details of individual objects can be observed. The effect of resolution degradation is depicted in Figure 2.9. This resolution is often determined by calculating the spread of measured activity when imaging a very confined activity concentration.

The activity recovery is a comparison of the estimated activity concentration to the injected activity concentration, and is often measured using the National Electri-

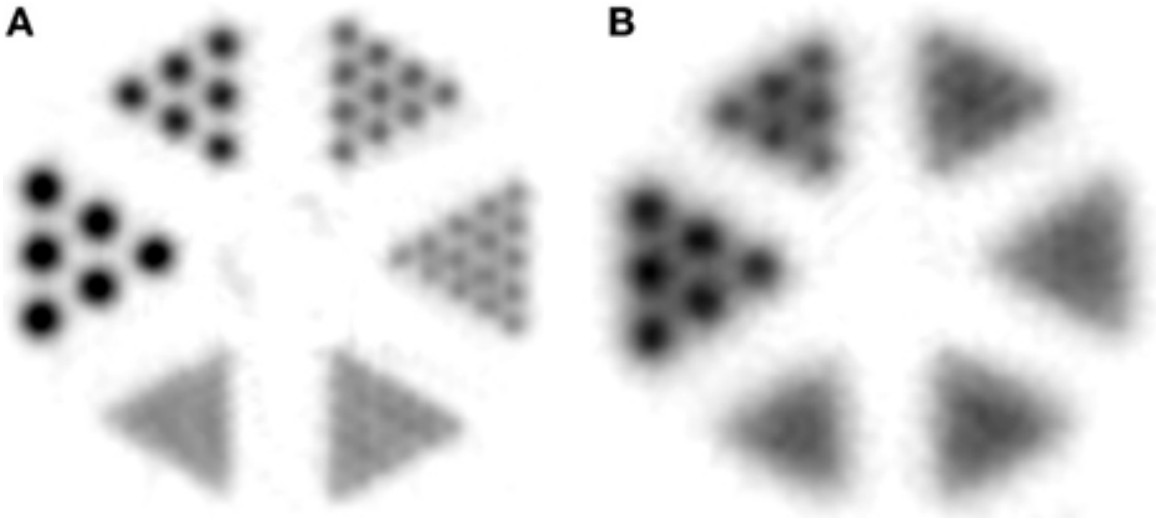


Figure 2.9: Image resolution: a degradation in resolution is seen from image A to image B. Obtained from [8].

cal Manufacturers Association (NEMA) International Electrotechnical Commission (IEC) image quality phantom (shown in Figure 2.10). This phantom consists of a large compartment that roughly resembles the shape of a human body and can be filled with a background activity. Within this compartment are six fillable spheres and a cylindrical insert through the center of the phantom. The recovery of activity is measured for each of these spheres to estimate how accurately the scanner and reconstruction algorithm can estimate the true activity concentration within tumors of different sizes. Understanding this relation allows clinicians to make a better inference of the true activity within a volume using the estimated activity. Similarly, this same phantom can be used to compare the contrast of the sphere vs. background activities for different sphere sizes. Finally, to measure the amount of noise in the image relative to the activity, the coefficient of variation is used, which is the ratio of the standard deviation in activity to the average activity within a particular region of interest. While these metrics are often useful to determine the quality of the images produced by a scanner and reconstruction algorithm, in Chapter 3, they were used to compare the images produced by our Monte Carlo pipeline to images obtained from a

Figure 2.10: The NEMA IEC phantom: a body phantom with a lung insert and six fillable spheres with various sizes. Obtained from [9]

clinical scanner. These comparisons provided an avenue to verify that the simulations accurately resembled experimental measurements.

2.2 Monte Carlo Simulations

Running PET simulations has many advantages. For instance, these simulations can help optimize clinical protocols and assess the impact of scanner parameters on the resulting image quality. The advantage of using simulations is that these types of tests can be run without the associated costs and effort required to take physical measurements. Monte Carlo (MC) methods provide one way to accurately simulate the processes involved in PET and are often considered the "gold standard" for these types of simulations. In our project, MC simulations were used to create PET images

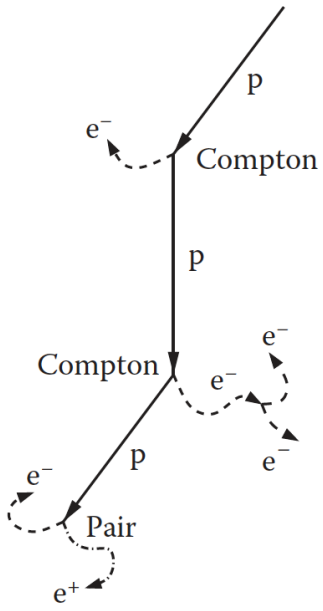


Figure 2.11: Monte Carlo particle tracking: An example of a particle history starting with a primary photon p (straight line) that undergoes two Compton interactions followed by a pair production. These interactions create secondary photons, electrons e^- (dashed lines) and positrons e^+ (dashed dotted line). Obtained from [10].

of high fidelity to test our motion correction method.

The idea behind MC simulations is to emulate the statistical nature of physical interactions by defining probabilities associated with different interactions. To accomplish this, each particle is tracked in steps; at the end of each step the particle may or may not undergo a particular interaction depending on several parameters including the type of particle and its energy as well as the type of medium it is travelling in. Furthermore, there are several different types of interactions that can occur and - depending on the interaction - the energy of the original particle will decrease and/or new particles will be created, which is illustrated in Figure 2.11.

When tracking a photon, the probability of an interaction occurring is directly related to the attenuation coefficient of the material, μ, which is dependent on the photon energy. For instance, for a photon with energy E travelling through a homogeneous medium with a linear attenuation coefficient of $\mu(E)$, the probability $p(s)$ of

the photon interacting after travelling a path length s within the medium is

$$p(s)ds = \mu(E)e^{\mu(E)s}ds \ . \tag{2.7}$$

The average distance a photon travels before interacting is known as the *mean free path length* and can be written as

$$\langle s \rangle = \int_0^\infty s \ p(s) \ ds = \frac{1}{\mu(E)} \ . \tag{2.8}$$

Therefore, the number of mean free path lengths in a step size of s is given by

$$\lambda = \frac{s}{\langle s \rangle} = \mu(E)s \ , \tag{2.9}$$

which provides an alternative representation of the interaction probability:

$$p(\lambda)d\lambda = e^{-\lambda}d\lambda \ . \tag{2.10}$$

This form has the advantage that it can be applied to heterogeneous materials, such that

$$\lambda = \sum_i \mu_i(E)s_i \ , \tag{2.11}$$

where s_i is the step length of the i^{th} step and μ_i is the linear attenuation coefficient of the material through which that step is taken. The cumulative distribution function of Equation 2.10 is given by

$$P(\lambda) = \int_0^\lambda p(\lambda')d\lambda' = 1 - e^{-\lambda} \ . \tag{2.12}$$

This allows the number of mean free path lengths, λ_1, to be sampled by generating

a uniform random number, ξ_1, from the half-open interval $[0, 1)$:

$$\lambda_1 = -\ln(1 - \xi_1) \; . \tag{2.13}$$

In this scenario, the photon would travel λ_1 mean free path lengths before interacting. Of course, the linear attenuation coefficient often has contributions from photoelectric absorption (μ_A), Raleigh scatter (μ_R), Compton scatter (μ_C), and pair production (μ_P):

$$\mu(E) = \mu_A(E) + \mu_R(E) + \mu_C(E) + \mu_P(E) \; . \tag{2.14}$$

To determine which type of interaction occurs, the range [0,1] can be divided into the different interactions:

$$[P_0, P_1] : \text{Photoelectric Absorption}$$
$$[P_1, P_2] : \text{Rayleigh Scatter}$$
$$[P_2, P_3] : \text{Compton Scatter}$$
$$[P_3, P_4] : \text{Pair Production} \; .$$

where

$$
\begin{aligned}
P_0 &= 0 \\
P_1 &= \frac{\mu_A}{\mu} \\
P_2 &= P_1 + \frac{\mu_R}{\mu} \\
P_3 &= P_2 + \frac{\mu_C}{\mu} \\
P_4 &= 1
\end{aligned}
\tag{2.15}
$$

Accordingly, by sampling a second uniform random number, ξ_2, from the interval $[0, 1]$, whichever interaction range includes this number will be chosen as the interaction type. The probabilities of the different interactions occurring are often determined empirically. Therefore, tracking the interactions and their products allows for accurate estimates of the statistical nature of the real-world physical processes.

A similar approach could be taken to track charged particles like electrons and positrons, however, the mean free path length for these particles is often so short that it would require simulating a huge number of interactions. This would be possible, but highly impractical. Therefore, alternative methods have been developed that leverage the fact that these interactions typically result in a small amount of transferred energy with small scattering angles. The condensed history (CH) method splits the interactions into hard and soft collisions. Soft collisions occur when the energy loss is determined to be small (again, by sampling a random number) and this energy is distributed evenly across the step length. Furthermore, the accumulation of all of the small scattering angles of these soft collisions is emulated with one single, larger scattering angle sampled from a predetermined distribution. Conversely, the hard collisions are simulated in a similar manner as the photon interactions. The threshold energy between soft and hard collisions as well as the scattering angle dis-

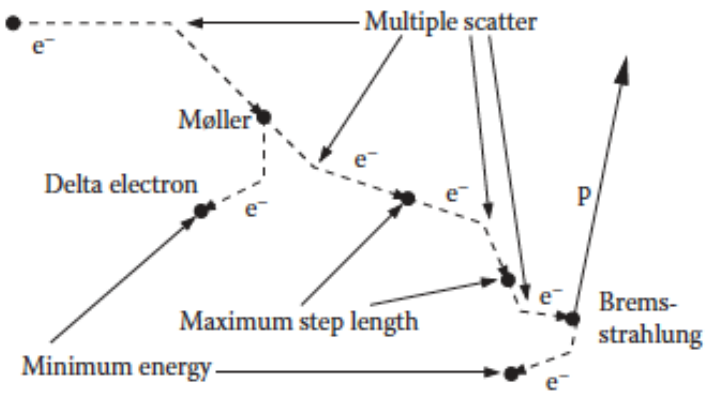


Figure 2.12: Monte Carlo electron tracking: A primary electron e^- (dashed line) undergoes a soft interaction followed by a hard interaction (Møller), which produces secondary electrons (delta electrons). Several soft interactions occur before a bremsstrahlung production event produces another secondary electron and photon p (straight lines). Obtained from [10].

tributions are estimated from empirical measurements. An example of an electron history is shown in Figure 2.12.

Due to the large number of particles and interactions present in these simulations, the computations can take a notoriously long time. Several parameters can be chosen to facilitate faster simulation times, but often come with the cost of lower simulation accuracy. For instance, there is an energy threshold for particles that defines the end of that particle's trajectory. Setting a lower threshold energy improves the accuracy, but comes at the cost of longer simulation times. Additionally, decreasing the step size of the particles also improves the simulation accuracy, but suffers from a similar efficiency limitation. The trade-off between efficiency and accuracy for these particular parameters will not be addressed in this dissertation; however, the relation between this trade-off and other aspects of MC simulations *is* explored in Chapter 3

To produce a Monte Carlo simulation of a PET system, it is necessary to define the geometry and materials in both the patient as well as the scanner. Additionally, the possible physics involved in the relevant types of interactions has to be well under-

stood. Luckily, there are tools built for this purpose that make this process relatively easy. For instance, the Geant4 toolkit [11] is useful for generating MC simulations in high energy physics. Furthermore, the Geant4 application for tomographic emission (GATE)[12, 13, 14] scripting language is useful for running PET simulations since it provides a user-friendly interface for the Geant4 toolkit. GATE allows the user to define the geometry and materials within the scanner as well as the requirements for detections to be registered. The material composition of the patient are defined by creating a *phantom*, whereas the activity distribution is the set by creating a *source*, which resembles the properties of the initially emitted particles. These definitions are described in more detail in Chapter 3. Assuming the difficulties of hard-coding the necessary physics are taken care of within Geant4 and GATE, the accuracy of these PET simulations largely depends on the resolution of the digital patient and how well the system and patient are emulated. Assessing this accuracy provided the motivation for the experiments shown in Chapter 3.

2.3 Neural Networks

In the field of machine learning, neural networks (NNs) have been designed to translate a given input into a desired output. NNs form this translation by having multiple layers, where each layer performs an operation on the output of the previous layer. The layers contain neurons that learn minor, independent tasks. When combined with the tasks of other neurons, a meaningful result is produced. This concept can be related to the human brain; hence, the name, neural networks.

One common way of orienting the neurons in the network layers is through the use of convolutional layers. These layers consist of weights, w, which are oriented in a series of filters and each filter can contain a bias term, b. A 2D convolutional operation of a single filter is shown in Figure 2.13. Formally, the output from the s^{th}

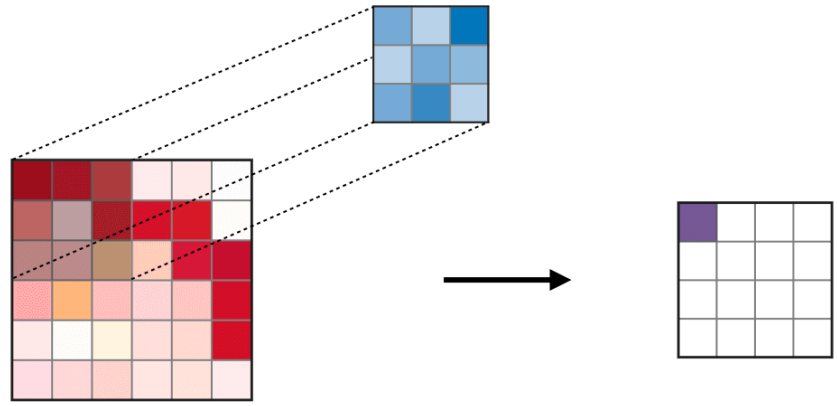


Figure 2.13: An animation of a single 2D convolutional filter operating on an input to produce a feature map. Obtained from [15].

filter in the l^{th} layer operating on the previous layer output, $\mathbf{h}^{(l-1)}$, can be summarized as

$$\mathbf{h}_s^{(l)} = \mathbf{g}(\mathbf{w_s}^{(l)} \otimes \mathbf{h}^{(l-1)} + \mathbf{b_s}^{(l)}) \ ,$$

where $\mathbf{g}$ is an activation function, which can be used to allow the NN to learn non-linear relationships. Currently, the most common activation used in NNs is the Rectified Linear Unit (ReLU) activation:

$$g(z) = \begin{cases} 0 & \text{for } z < 0 \\ z & \text{for } z \geq 0 \ . \end{cases} \tag{2.16}$$

When applied to images, the filters extract features from the inputs and - by compiling multiple layers together - it is possible to extract more complex representations of the original input, which are often referred to as feature maps. As seen in Figure 2.13, when the filter is shifted by more than one pixel before performing the next operation, the output of that layer is downsampled. The number of pixels that the filter moves in between operations is called the stride length.

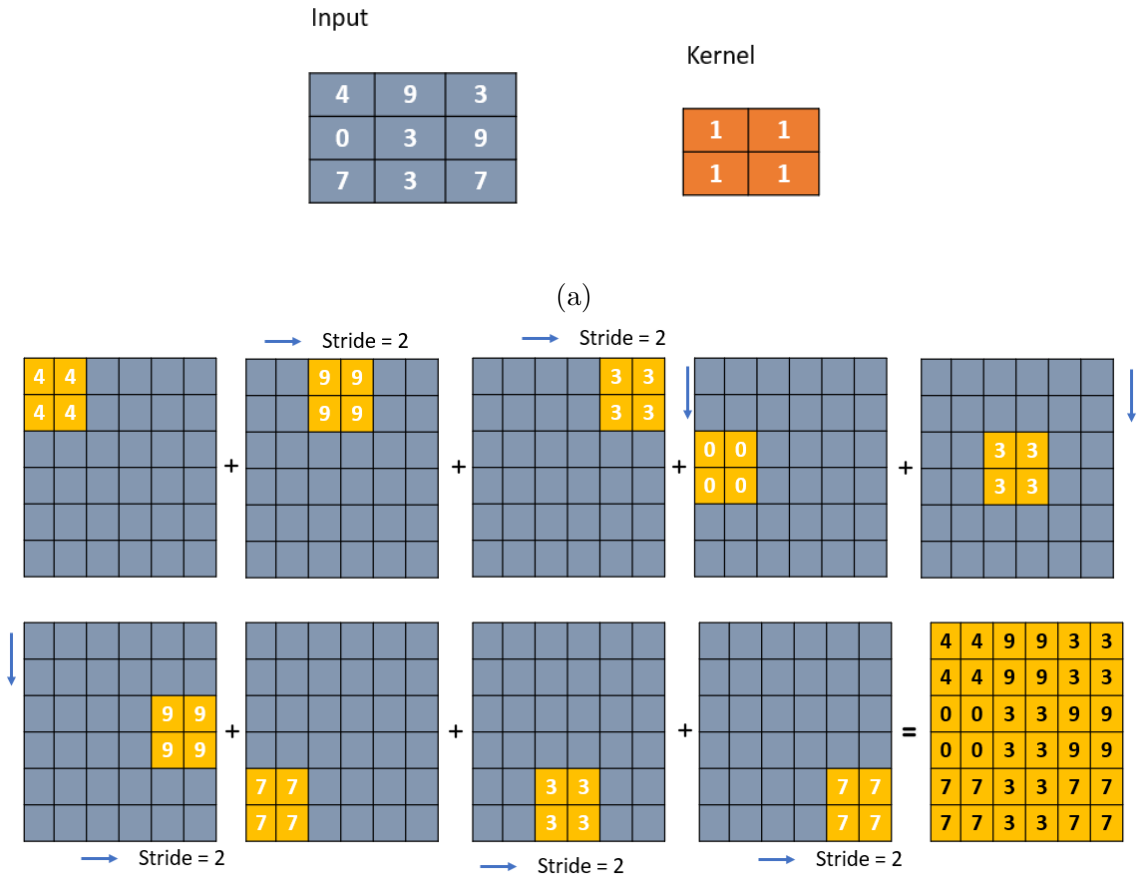


Figure 2.14: Transposed convolution: (a) the input and the filter (or kernel), (b) the output is computed by applying the filter to each of the input pixels and adding the results together with a stride length of two. Obtained from [16].

When developing a convolutional neural network (CNN), it can be useful to have an operation that upsamples the feature maps as well. One type of layer that accomplishes this is the transposed convolutional layer, which is outlined in Figure 2.14. In this scenario, each weight in the filter disperses the individual pixel values within the input layer to a broader area and these dispersed values are added together to form the output.

Through training, the network learns which features are useful to extract and how they can be combined to create meaningful relationships between different aspects of

the image. In other words, the network learns an optimal set of the weights and biases of each layer. This learning is achieved by having the network try to accomplish a given task on a set of training data.

The training data typically consists of paired inputs (often images) and outputs, which could be other images or labels associated with the input images. In supervised learning, the task is to have the NN translate the inputs to the outputs. Training is accomplished by having the NN compute a forward-propagation by making a prediction given an input. Initially, the weights are essentially randomly set, so this prediction will likely be quite poor. This prediction is compared to the known output (or target) through a loss function, which evaluates how far the prediction is from the target. Backpropagation is then accomplished by computing the gradient of the loss function with respect to the network weights. Accordingly, the weights are adjusted to reduce this loss, so that the next time this input is seen by the NN, the prediction will hopefully be closer to the target. By iteratively computing forward-propagations, backpropagations, and weight adjustments over the entire training set, the network learns to generalize its predictive capabilities over time. Once an adequate level of performance is achieved, the NN weights are fixed and the network can be applied to new inputs to perform predictions.

Supervised learning can only be accomplished when you have access to labels associated with each input sample. However, obtaining labels can be quite cumbersome, which provides the necessity for *unsupervised* learning, where the aim is to learn a task without having the target labels. As you might expect, defining the objective to be accomplished by the network in unsupervised learning is often not as straightforward. Nevertheless, this type of training procedure is preferred when access to quality labels is limited. This is often the case for large datasets of clinical PET images, which is why we applied unsupervised learning techniques to accomplish the

task of respiratory motion correction in Chapter 4.

The machine learning applied in this thesis was all done using Python 3.7 and the PyTorch deep learning library (v. 1.10.0) [17]. PyTorch allows users to easily design the framework of a NN with predefined layers (such as convolutional layers) as well as the capability of designing custom layer operations. Similarly, commonly used loss functions are already defined within the toolkit, but customizing new loss functions can be easily accomplished. The optimization algorithms used during the learning stage are also simple to utilize. Accordingly, the necessary steps required to train a NN in PyTorch are as follows: (1) define the architecture of the NN, (2) choose an optimization method, (3) use the NN to perform an estimate using a training sample (or a batch of training samples), (4) evaluate the loss function, (5) back-propagate the loss through the NN weights, (6) update the NN weights using the optimizer, and (7) iteratively repeat steps (3)-(6) over the entire training set until the network reaches an adequate level of performance. It is worth noting that each of these steps can be written in just a few lines of code using PyTorch. Lastly, PyTorch is constructed to work on both CPUs and Graphics Processing Units (GPUs), which can make the training procedure much more efficient. This is especially important when applying a NN to 3D PET images because performing 3D convolutional operations is much more computationally intensive than 1D or 2D convolutions. Accordingly, training a 3D CNN would not be feasible without the use of GPUs, especially since these types of projects require many iterations of trial an error until a final model is chosen. For this reason, I am thankful that I did not undertake this project 10 years ago!

Before applying these exciting tools to correct for respiratory motion in PET imaging, we will return to MC simulations. Namely, the next chapter covers an experiment where a MC simulation pipeline is developed and the results are compared against experimental measurements. Validating the simulation results will then allow

for these simulations to be used to produce PET images to test our motion correction method.

Chapter 3

In-Depth Monte Carlo Modeling and Validation of a Clinical PET System

3.1 Introduction

Positron emission tomography (PET) is an imaging technique that detects radiation emitted from an administered radionuclide, allowing for in vivo assessment of physiological processes. PET is used in the fields of cardiology and neurology, and is especially useful in oncology when the radiopharmaceutical targets cancerous cells [18, 19, 20]. The use of Monte Carlo (MC) simulations in the study of PET imaging has a wide range of applications. For instance, these simulations can help optimize clinical protocols and assess the impact of scanner parameters on the resulting image quality [21, 22, 4]. The advantage of using simulations is that these types of tests can be run without the associated costs of taking physical measurements.

The ability to relate the simulated results to clinical systems is possible due to the potential for realism of MC simulations [23]. Additionally, while obtaining a large amount of real patient data can be difficult, the simulations only come at the cost of computational power and time. Therefore, large realistic datasets can be generated providing many benefits, especially with regards to machine learning applications. For instance, MC simulations have been used to generate PET data that were then

used to train a deep learning network that performed image reconstructions based on sinogram data [24]. One main advantage of simulating the data is that there is an inherent "ground truth" since the exact circumstances regarding the imaged object are known. Therefore, creating a fine-tuned training set such as this would provide an ideal setting for the development of a proof of concept for an unsupervised training framework before attempting to apply the novel method to real patient data. In this scenario, while the training process would not require access to the truth, validating the performance post-training would be much more robust. However, before generating large training sets, the simulations must be validated against experimental measurements to ensure that they accurately resemble clinical data.

One toolkit that is useful in generating MC simulations in high energy physics is Geant4 [11], which consists of software that is used in many fields, including medical imaging. For the purposes of PET, the GEANT4 application for tomographic emission (GATE)[12, 13, 14] scripting language is useful since it provides a user-friendly interface for the Geant4 toolkit. GATE has been used to investigate several PET scanner models for a variety of purposes. For instance, novel PET scanner scintillation crystals were simulated using GATE to determine their impact on performance characteristics [21]. Simulations have also been validated against measurements by comparing the sensitivity, scatter fraction, count rate performance, and spatial resolution [25, 26]. Additionally, GATE simulations were validated in terms of their coincidence count rates, scatter fraction, and noise equivalent count rates [27]. GATE simulation of two Siemens scanners were previously validated and compared in terms of the spatial resolution, sensitivity, and scatter fraction [28]. The Siemens Biograph mCT scanner was previously simulated using GATE for the purpose of comparing the results to another simulation package and against experimental results [3]. Another simulation of the Biograph mCT scanner was developed to test a cost-effective

solution to extend the axial field of view in PET, but these simulations were not validated against experimental measurements [4]. Lastly, Geant4 simulations were validated when modelling a whole-body PET scanner prototype that was constructed from 4-layer depth-of-interaction detectors [22].

To develop a large PET dataset for the purpose of testing deep learning data manipulation techniques, a GATE simulation of a PET system was developed. Due to the wide accessibility of a Biograph 40 mCT PET/CT system (Siemens Healthineers), this scanner was modelled so that future developments can be transferred to the available clinical data. Therefore, the objective of this paper was to validate that our GATE simulations of the system could accurately resemble experimental measurements from a clinical scanner. This study includes metrics previously used in MC validation studies, but also introduces a new method to estimate the sensitivity for validation purposes. Additionally, the code is made publicly available to allow readers to easily implement and use the simulations and analyses shown (https://github.com/teaghan/PET_MonteCarlo).

3.2 Materials and Methods

First, physical images were acquired of several phantoms with the scanner and then these experimental set-ups were replicated within GATE. Using the acquired data, the detector geometry, count rates, sensitivity, image uniformity, image resolution, recovery coefficients, coefficients of variation, hot sphere contrast, and background variability were compared between the measurements and simulations.

3.2.1 PET Scanner and Software

A Biograph 40 PET/CT (Siemens Healthineers) scanner was used to obtain the experimental measurements and was modelled in GATE. The geometry of this PET

system consists of lutetium oxyorthosilicate (LSO) crystals with dimensions of 3.98 mm × 3.98 mm and a thickness of 20 mm. The crystals are oriented in blocks - each block consists of a 13×13 array of crystals and a set of 48 blocks are distributed around the center of the system to form a ring with a diameter of 842 mm. This scanner contains 4 rings, therefore, providing a total of 32,448 detector crystals with an axial field of view of 216 mm.

For all measurements and simulations, fluorodeoxyglucose (F-18) was used as the radionuclide to be imaged. The acquired data was processed using the prototype software JS-Recon e7-tools (Version VG62B, Siemens Healthineers) to produce the sinograms. Following clinical protocol, images were reconstructed using the non-TOF OSEM algorithm with 3 iterations 24 subsets and applying a Gaussian filter with a full-width at half-maximum (FWHM) of 5 mm. Each image had 109 slices with a slice thickness of 2 mm and a matrix size of 200 × 200 pixels with a uniform pixel width of 4 mm. However, to be consistent with National Electrical Manufacturers Association (NEMA) NU 2-2007 protocol[29], the point source images were reconstructed with 600 × 600 pixels to keep the pixel width (1.36 mm) less than one third of the image resolution.

3.2.2 PET measurements

Experimental measurements were acquired at Radboud University Medical Center in the Netherlands. Two different phantoms were scanned: a uniform cylinder and the NEMA International Electrotechnical Commission (IEC) image quality phantom. For each scan, the data was acquired in list-mode to enable retrospective analysis. Additionally, resolution measurements of a point source were acquired from the Siemens documentation [30].

For the uniform cylinder, a 20 cm long high-density polyethylene cylindrical phan-

tom with an inner diameter of 20 cm and outer diameter of 21 cm was filled with an activity concentration of 12.35 kBq/mL. This phantom was placed with its long axis in the axial direction and at the center of the PET/CT scanner, then imaged for 10 minutes.

The resolution measurements were obtained by following the resolution protocol outlined in NEMA NU 2-2007 [29]. This protocol emulates a point source by using a capillary tube with an inside diameter of less than 1 mm and an outside diameter of less than 2 mm; the activity within the tube extended less than 1 mm. The activity was chosen so that the rate of random coincidences was less than 5% of the total event rate and the source was imaged at six different locations until at least 1×10^5 counts were acquired. These six locations included three transverse positions at both the center of the axial FOV and one-fourth of the axial FOV from the center of the FOV. Measured from the center of the FOV, the three transverse locations included (0 cm, 1 cm), (0 cm, 10 cm), and (10 cm, 0 cm), where the first dimension was measured horizontally within the scanner.

The NEMA IEC phantom consists of a compartment that mimics the general shape of the human body along with six fillable spheres (inner diameters 37, 28, 22, 17, 13, and 10 mm) and a cylindrical insert (diameter of 51mm) through the center of the phantom. The housing of the phantom is made from a plexiglass material and is approximately 3 mm thick. Measurements of the NEMA phantom were acquired by filling the spheres with a solution that had an activity concentration of 21,040 Bq/mL and a background solution of 2,260 Bq/mL. Scans from two bed positions were acquired for 300 seconds per bed position, and the second position was used for comparisons since the spheres were more centered in the axial direction.

3.2.3 PET Simulations

The PET imaging process was simulated with GATE (v. 9.0)[12, 13, 14] and a diagram of this process is shown in Figure 3.1. In GATE, the source is defined as the distribution of radioactive atoms, which is the F-18 radionuclide in PET imaging. Defining the distribution of the source was accomplished by creating a voxelized arrangement of activity values where each voxel in our source had a size of 2 mm × 2 mm × 2 mm.

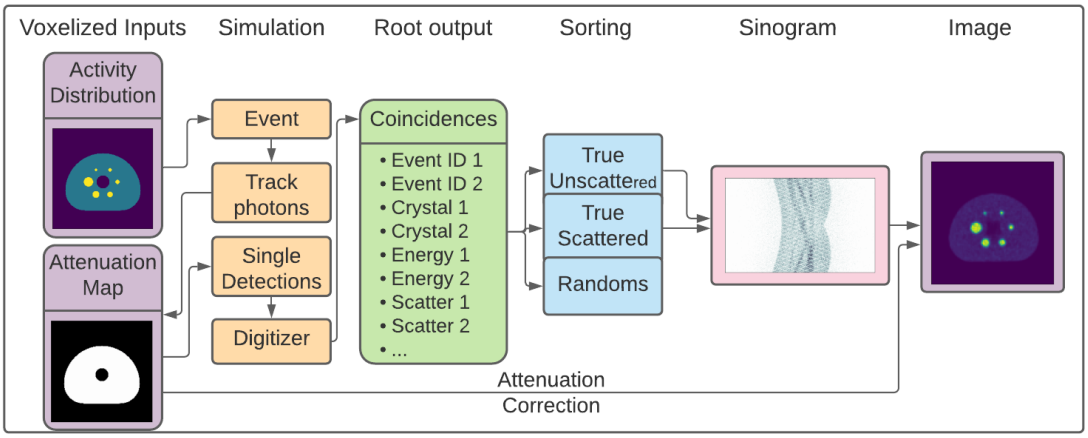


Figure 3.1: A schematic diagram of the simulation process.

A voxelized phantom was also created with each voxel being assigned a material. The available materials were defined separately by providing their chemical composition, density, and state (gas, liquid, or solid). The materials used in the voxelized phantoms for these experiments are listed in Table 3.1. While the use of a voxelized source and phantom was not necessary for simple geometries, such as the cylindrical phantom, doing so allowed for the method to be extended to more complex geometries such as the NEMA image quality phantom or phantoms including anatomical structures. Conversely, when simulating a point source, the voxels were too large to accurately model its physical dimensions. Therefore, the shape and orientation of the

source were defined within GATE by depicting the source as a cylinder and specifying its radius (0.145 mm) and length (0.5 mm) as well as its center location. The point source was contained within a glass cylinder with an outer diameter of 0.575 mm and a length of 125 mm.

Table 3.1: A list of the GATE materials used in the voxelized phantoms.

Material	Density (g/cm^3)	Constituents (compound or mass fraction)	State
Air	0.00129	N (0.755), O (0.232), Ar (0.0128), C (0.000124)	Gas
Water	1.00	H_2O	Liquid
Polyethylene	0.96	CH_2	Solid
Glass	2.50	Na (0.102), Ca (0.051), Si (0.248), O (0.599)	Solid
Plexiglass	1.19	H (0.080), C (0.600), O (0.320)	Solid

Within GATE, the photons that are detected by the crystals are processed by the digitizer, which applies any desired blurring effects and defines the criteria for coincidences to be registered. To replicate the variations in the detection process observed in real scanners, several factors were applied to the simulated photon detections within the crystals. Based on guidance from the manufacturer, a Gaussian blurring with a FWHM of 500 ps was applied, defining the temporal resolution. To provide a source of energy blurring to the detected pulses, each crystal was randomly assigned an energy resolution between 15% and 35% at a reference energy of 511 keV, which are the default values found within the GATE documentation. In agreement with the physical scanner, single events were accepted in the energy window [435 keV - 650 keV], and two events were considered to be a coincident detection if they were recorded within 4.0625 ns of each other and separated by at least five crystals. The same criteria were applied to the delayed coincidences, but with an offset of 500 ns, which provided an estimate for the distribution of random coincidences in the

simulation.

To speed up the simulation process, each MC simulation was split into multiple jobs, which were run in parallel on the remote clusters of Compute Canada (www.computecanada.ca) providing access to 28,000 compute cores. Each job used a single core and ran for 1-2 hours, but the number of jobs used for a single simulation depended on the amount of activity in the phantom. For instance, the point source simulations only required 100 jobs each whereas the uniform cylinder required 2000 jobs. Once all of the jobs were completed, the coincidence detections were binned into a sinogram, matching the format of the sinograms developed within the JS-Recon prototype software. This allowed for the sinograms from both the measurements and simulations to be reconstructed in an identical manner using this same software. For the simulations, rather than creating separate sinograms for the prompt and delayed coincidences and then performing a correction for the random events, the random coincidences were identified in the sorting of the data and disregarded. However, recording the delayed coincidences allowed for direct comparisons to be made with the delayed detections in the measurements.

Similar to the measurements, the simulations were reconstructed using the non-TOF OSEM algorithm with 3 iterations and 24 subsets using a post-reconstruction Gaussian filter with a FWHM of 5 mm. Furthermore, the same normalization files that were used to reconstruct the measurements were used for the simulations. Lastly, to produce the attenuation maps used in the reconstruction of the simulated sinograms, the materials within the voxelized phantoms were converted into the corresponding attenuation coefficients that were expected by the JS-Recon code. During the reconstruction process, these were then converted into the attenuation coefficients for 511 keV photons.

3.2.4 Validation tests

To mitigate potential biases, validation metrics were chosen such that they could be applied identically to both the measurements and the simulations. Furthermore, since the objective was to verify the accuracy of the detections, agreement between the sinograms provided a better quantitative assessment of the simulations, though the reconstructed images were still needed to calculate some standard validation metrics. Aside from the image reconstructions, the analysis was performed in Python 3.7.

Comparing count rates, sensitivity, and image uniformity

Using the measured and simulated data from the uniform cylindrical phantom scan, the sinograms were generated - correcting for the random coincidences. The alignment of the detector crystals was then verified by comparing the distribution of the detections within the resulting sinograms. More explicitly - for this scanner and software - the sinogram (or Michelogram) [5] consists of nine segments: the first segment contains the lines of response (LORs) within each axial plane, while the remaining segments represent the LORs at increasingly large angles (in both directions) from different transaxial planes. Therefore, each voxel within the sinogram represents the total number of detections along a particular LOR. Due to the spacing of the crystals and blocks - as well as the exposed fraction of the crystals at different LOR angles - the detections are more likely to occur along certain LORs. As a result, the peaks and gaps seen in the sinogram are specific to the orientation of the scanner crystals, which can be seen in Figure 3.2. Since the distribution of crystals is symmetric around the bore of the scanner, investigating the distribution of these detections along the axial direction is more informative (i.e., comparing the number of counts in each plane). This was accomplished by summing the number of detections in each plane of the sinogram and comparing the results. Note that, for this comparison, it was impor-

tant that the phantom in the simulation was well aligned (within the scanner) to the measured phantom. Therefore, the source was defined as a geometrical shape (similar to the point source previously mentioned) rather than a voxelized volume to easily rotate and shift the phantom slightly. To be consistent with the proposed pipeline - for all subsequent comparisons - the cylindrical phantom was defined as a voxelized volume.

The total number of true detections as well as the fraction of random and scattered coincidences were also compared. To have a consistent method for estimating the number of random detections, a coincidence offset of 500 ns was used. To estimate the scatter fraction, a method that utilizes the cylindrical phantom sinogram was adopted [31]. Briefly, within a single plane, the distribution of the true and scattered detections was modelled by two separate curves. The parameters of those curves were fit to the actual distribution of the total counts in the data by first taking the sum of the two curves and convolving this with a point-spread-function (FWHM = 4 mm). An example of this is shown in Figure 3.3. This fitting procedure was performed after the nine segments of the sinogram were summed together, without correcting for random detections, to create a total of 109 slices rather than the original 621 slices. Using the optimized curves, the scatter fraction was then estimated by taking the ratio between the number of scattered and total detections within the central 60% of the cylindrical phantom (in the radial direction). Furthermore, to disregard any disagreements that may have arisen due to the ends of the phantom not being modelled correctly, this metric was only calculated for the central 60% of the axial planes and the average was reported.

The sinogram sensitivity [32] was then measured using the previously described summed representation of the sinogram after correcting for the random detections. To calculate the sensitivity, the sum was taken along the projection and bin axes,

dividing by the scan duration to produce a single vector of 109 elements. Effectively, this vector is the average number of true detections per second in each plane computed over the entire duration of the scan. Comparing this to the total activity within a given plane provided the sensitivity of each plane. Again, this metric was computed for the planes in the central 60% of the cylinder. Comparing the sensitivities in the sinogram-space – rather than using the NEMA protocol – provided a more robust validation of the simulations since there was no further data manipulation made on the detections other than the sorting of coincidences.

Following the reconstruction of the two sinograms, the image uniformity was calculated both within each slice and between slices [33]. More specifically, within each slice, five circular regions of interest (ROIs) with a diameter of 30 mm were drawn. One of these ROIs was located at the center of the image and another at each of the four cardinal locations placed 10 mm from the edge of the phantom. Using the average pixel intensity within each ROI, the integral uniformity within the slice and the integral axial uniformity were calculated [33]. These metrics were computed for the central 60% of the phantom (in the axial direction), and the averages and standard deviations were calculated.

Image Resolution

Using the reconstructed images of the point source, profiles in all three directions were taken through the peak of the activity distribution. The maximum of each profile was estimated by fitting a parabola to the peak and the two nearest neighboring pixels. The FWHM of the profile was then measured by linearly interpolating between adjacent pixels at half the maximum. The average transverse and axial FWHM measurements were determined at 1 cm and 10 cm from the center in the transverse direction.

The NEMA Image Quality phantom

Similar to how PET systems are analyzed to determine their ability to image tumors of different sizes [34, 35], the NEMA image quality phantom was utilized in this study to compare the resulting statistics between the measurements and the simulations. Following the EANM Research Ltd (EARL) procedure [34], the recovery coefficients (RCs) were determined by selecting a volume of interest (VOI) within each sphere that included all pixels above half the maximum intensity, correcting for the background. Using the pixel values within this VOI, the RCs were calculated by finding the ratio between the activity concentration within each sphere in the image and the administered activity. This was done for both the maximum and mean concentrations. The RCs from the measurements and simulations were then compared against each other as well as against the EARL standards [34]. To mitigate partial volume artifacts mostly affecting the smallest spheres, five separate simulations were computed - randomly shifting the spheres within the range of half the voxel size in each direction. The EARL procedure was also followed in order to calculate the coefficient of variation (CoV) in both images. Lastly, using the same images, the NEMA NU 2-2007 protocol was followed to calculate both the hot sphere contrast and background variability for each sphere size [29].

3.3 Results

3.3.1 Cylindrical phantom comparisons

Figure 3.2 shows a comparison of the total number of detections within each plane of the sinograms for the uniform cylindrical phantom. In panel (d) of this figure, the locations where a high signal is accumulated matched between the measured and simulated data, which implies that the detecting surfaces of the crystals are properly

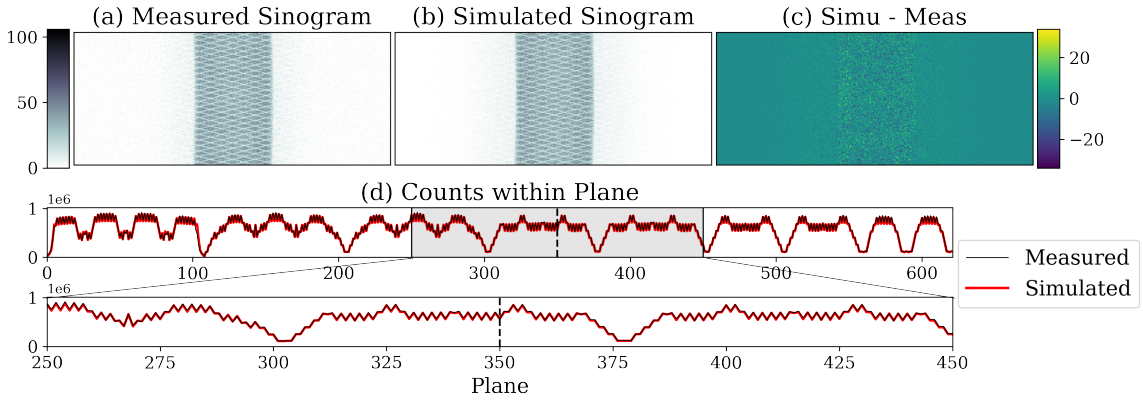


Figure 3.2: The 350th plane (in the axial direction) of the uniform cylindrical phantom sinograms from both the (a) measured and (b) simulated data, along with (c) the difference between the two sinograms. In each frame, the color scales are in units of counts. A similar agreement was found in all planes as shown by the similarity in (d) the total number of counts found in each plane along the axial direction. The counts in the 250-450 planes were rescaled to more clearly visualize where detections were and were not being acquired.

aligned in the simulation. Similarly, alignment was seen for the areas with little or no signal, which correspond to geometrical gaps in the detectors. These results provide confidence that the scanner is well represented geometrically in the MC simulation.

3.3.2 Counting statistics

When imaging the cylindrical phantom, the total number of true detections in the measured and simulated data were found to be 3.67×10^8 and 3.47×10^8, respectively, reflecting a 5% difference. By using the number of delayed detections, the fraction of random coincidences was found to be 19% and 25% in the measured and simulated data, respectively. An example of the fitted scattered and primary curves for a single plane is shown in Figure 3.3. The average scatter fraction across the planes were found to be $(31.1 \pm 1.1)\%$ and $(29.8 \pm 0.8)\%$ for the measured and simulated scans, respectively.

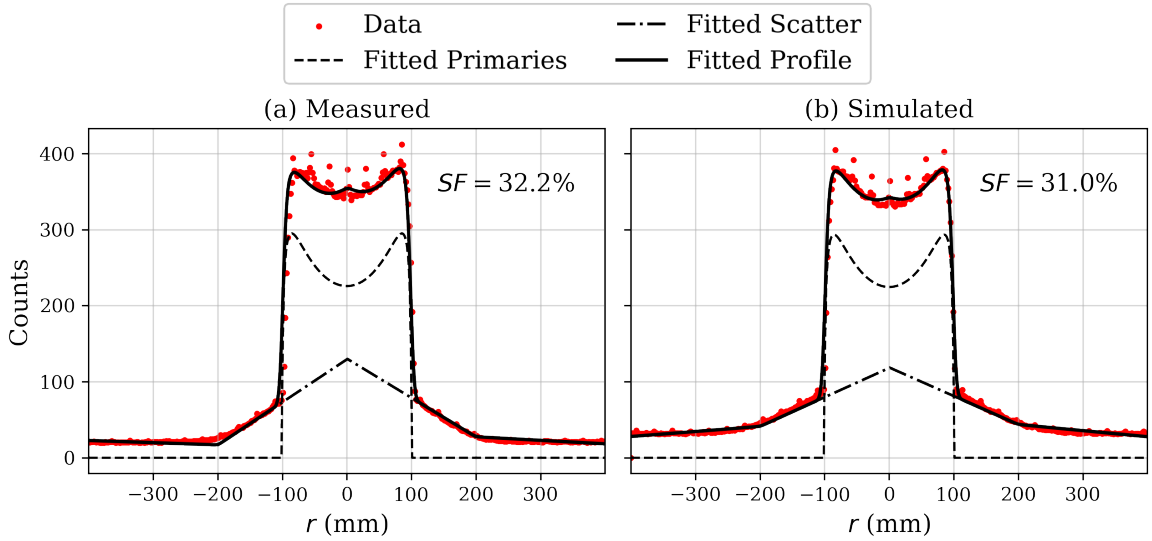


Figure 3.3: Results from the method used to calculate the scatter fraction in the central plane of the two sinograms ((a) measured and (b) simulated) of the cylindrical uniform phantom. The scatter fraction is the ratio between the fitted scatter and profile, averaged over the central 60% of radial values within the phantom.

3.3.3 Sensitivity

The sensitivity calculated plane-by-plane in the measured and the simulated sinograms is shown in Figure 3.4, exemplifying agreement in shape, but a systematic shift of 5% in the magnitude. The global sensitivities across this region in the measured and simulated sinograms were found to be 10.0 cps/kBq and 9.5 cps/kBq, respectively.

3.3.4 Uniformity

Comparing the ROIs within the slices of the resulting images demonstrated an integral uniformity of (0.015 ± 0.005) and (0.029 ± 0.011) for the measured and simulated data, respectively. Similarly, by comparing each ROI across all slices, the axial integral uniformities were found to be (0.024 ± 0.006) and (0.040 ± 0.015) for the measured and simulated data, respectively.

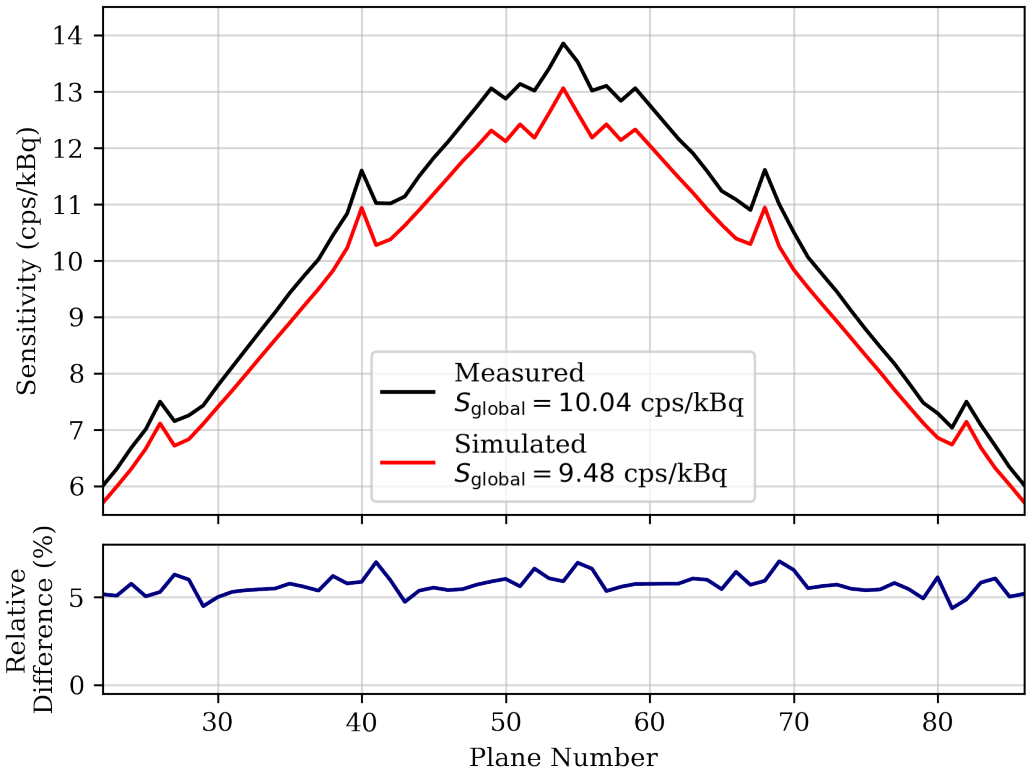


Figure 3.4: A plane-by-plane comparison of the sensitivities found in the measured and simulated sinograms. The comparison was made for the central 60% of the phantom. The global sensitivity across this same section of the phantom is shown.

3.3.5 Point source comparisons

Image Resolution

The calculated resolutions are summarized in Table 3.2, where the average percentage difference between the resolution in the measured and simulated images was found to be 2.9%.

Table 3.2: A summary of the resolution measurements using the images of the point source.

	Measured FWHM (mm)	Simulation FWHM (mm)
Transverse (@ 1 cm)	5.9	5.9
Transverse (@ 10 cm)	6.0	6.2
Axial (@ 1 cm)	5.5	5.7
Axial (@ 10 cm)	6.0	6.3

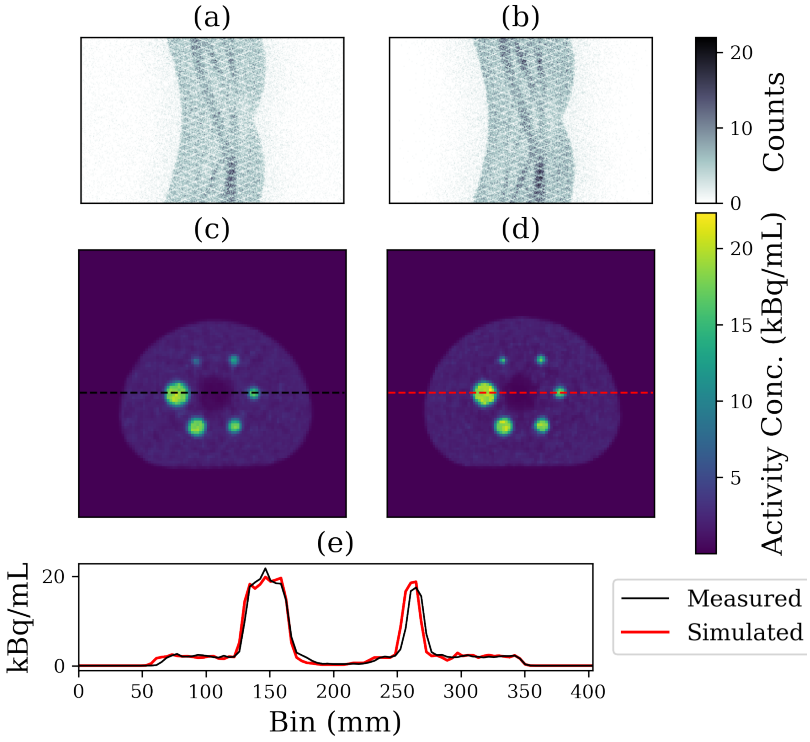


Figure 3.5: The NEMA IEC body image quality phantom. A single slice from the sinograms and reconstructed images of both the (a & c) measured and (b & d) simulated data. Profiles are shown crossing two of the spheres in the images (e).

3.3.6 NEMA phantom image quality comparisons

Recovery coefficients and coefficients of variation

The measured and simulated data from the image quality phantom showed similar characteristics in both the sinogram- and image-space (Figure 3.5). To compare the two quantitatively, the RCs for each sphere in the phantom are shown in Figure 3.6. In general, there is agreement between the two sets of RC values aside from the two smallest spheres, and only the smallest sphere in the simulation provided an RC outside of the EARL range. The CoV for the measured and simulated images were both found to be 12%. As a reference, this value is below the 15% maximum threshold set in the EARL procedure.

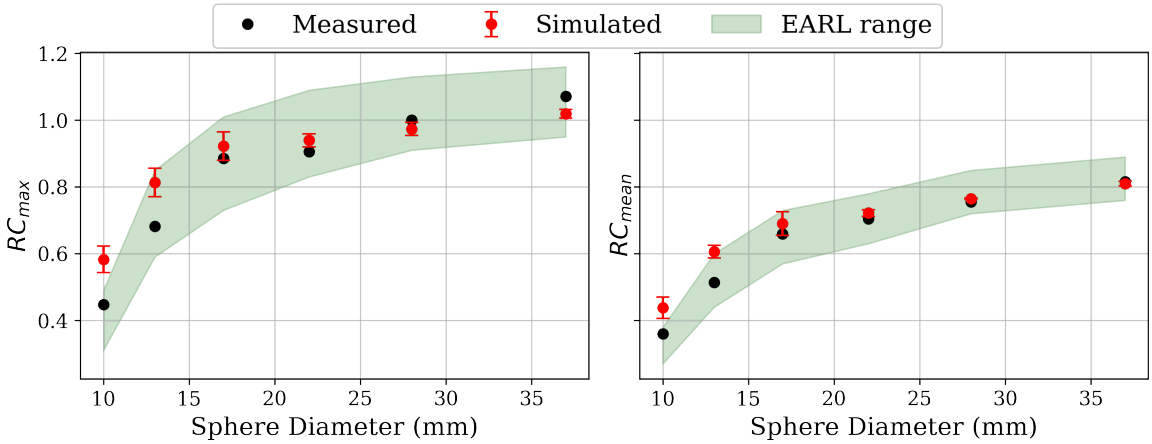


Figure 3.6: Comparisons of the recovery coefficients between the measured and simulated data for each of the spheres in the NEMA image quality phantom. As a reference, the suggested ranges outlined by EARL are also shown.

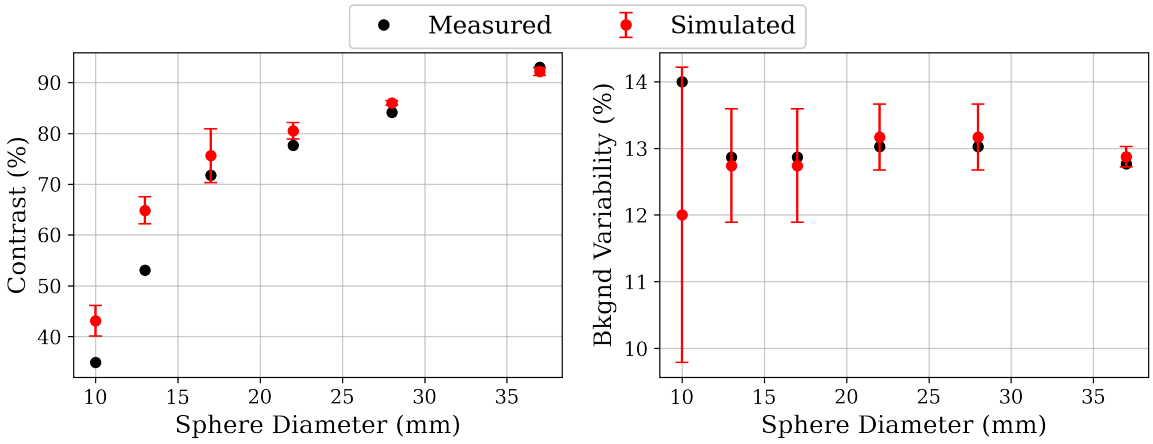


Figure 3.7: Comparisons of the hot sphere contrasts and background variabilities between the measured and simulated data for each sphere size in the NEMA image quality phantom.

Contrast and Background Variability

A comparison of the hot sphere contrasts and background variabilities is shown in Figure 3.7. The average percentage difference between the measured and simulated images was found to be 8.8% and 3.4% for the contrast and variability, respectively.

3.4 Discussion

An accurate distribution of detector crystals was confirmed by comparing the number of detections along the various lines of response when scanning a cylindrical phantom. Additionally, the scatter fractions were found to agree, however the total number of true detections differed by 5% and the rate of random detections in the simulation was found to be lower than in the measured data. The global scanner sensitivities were found to be within 5% of each other and the image uniformity was consistent between the measured and simulated images. Similarly, the image resolution in the measured and simulated images were found to be consistent within 5%.

Furthermore, there were some disagreements in the RC values and hot sphere contrasts for the smaller spheres, especially the smallest sphere. The EARL ranges shown in Figure 3.6 are based on experimental measurements taken from a large collection of PET systems [35]. Due to the partial volume effects, a decrease in measured activity as the sphere size decreases is observed, but this trend was less evident in the simulations. It is possible that this is the result of a systematic occurrence exemplified in both the EARL ranges and the measurements. For instance, if the spheres are typically not completely filled with the solution, then the measured activity concentration within the spherical VOI will be lower, and this effect would be more substantial for smaller spheres. This type of systematic error would not be present in the simulations because the spheres are modelled by a homogeneous spherical activity distribution fully encompassing the sphere volume. Since the RCs and hot sphere contrast met-

rics are correlated, the source of the discrepancies seen in the small spheres for both metrics is likely the same. Lastly, the CoVs in the measured and simulated images were both found to agree with the EARL standards.

While, in general, many of the presented metrics showed agreement between the measurements and simulations, there were several disagreements that may have illuminated limitations in either the simulation process or the adopted validation methods themselves. It is possible that some of these disagreements are simply a consequence of the voxelized phantom not properly resembling the physical phantom used in the measurements. For example, this could be due to the resolution limits of the voxels used in the digital phantom or the housing components not being perfectly modeled. A visible example of the phantoms in the measurements and simulation not matching perfectly can be seen in Figure 3.5. Here, the phantom in the measurement is rotated slightly around the plane axis, which results in the profile across the measured image being slightly shorter than the simulation profile. Enhancing the resolution of the phantom would better model small structures such as the 10 mm sphere in the NEMA IEC phantom; however, this would result in an increase in computation time. Our goal was to present a single pipeline for all of our simulations and limit the variations we made across simulations. Therefore, the $(2\times2\times2)$ mm^3 voxel size was chosen as a trade-off between having reasonable computation times and some limitations as shown in the RC values for the small spheres.

The simulation times varied depending on the total activity in the phantom used. For instance, the point source simulations took less than 100 hours on a single CPU core, while the cylindrical phantom required approximately 2000 hours. However, due to the access to such large computer clusters, over 1000 jobs could be launched simultaneously, reducing the real-world computation down to several hours. While clinical patient scans are almost always more efficient to carry out than Monte Carlo

simulations, the value in using simulations is that the variability in the scanning conditions are user-defined. For instance, accessing the true activity distribution within a digital patient is possible, which can be imperative to some machine learning tasks. Therefore, whether or not the simulations are worth the time required to run them largely depends on the accessibility to large computer clusters and the requirement of ground truth data.

Lastly, generating an idealized machine learning training set is desired to train a model that can improve certain data processing tasks, such as image reconstruction or artifact correction. Although the fidelity requirements depend on the task at hand, it is likely that these simulated images are of sufficient realism compared to real-world PET images for machine learning purposes. In any case, the simplified nature of even the most sophisticated current digital patient phantoms probably introduce larger deviations from real patient images than the disagreements shown in this work. However, a simulated dataset would provide an excellent source for developing a proof-of-concept or for the initial training of a network that can then be fine-tuned using clinical images.

As much as possible, the code to run these simulations was made publicly available (https://github.com/teaghan/PET_MonteCarlo). This includes all scripts required to run the GATE simulations as well as the Python analysis code used to obtain the metrics included in the validations. However, due to the proprietary nature of the code required to sort the GATE results into sinograms and reconstruct the images, it is not possible to make these available.

3.5 Conclusions

A GATE simulation process was developed to mimic the PET imaging process of the Biograph 40 mCT PET/CT system (Siemens Healthineers). The PET Monte

Carlo simulations were validated against experimental measurements using several phantoms and showed agreement in the detector alignment, counting statistics, sensitivity, image uniformity, and image spatial resolution. The recovery coefficients, coefficients of variation, hot sphere contrasts, and background variabilities were also found to be in agreement, though there were discrepancies found when comparing the RC values and contrasts for smaller spheres.

Chapter 4

FlowNet-PET: Unsupervised Learning to Perform Respiratory Motion Correction in PET Imaging

4.1 Introduction

Positron emission tomography (PET) is used to measure the distribution of a radionuclide administered to the patient, which provides an assessment of the physiological processes associated with the uptake of that radionuclide. When a cancer-targeting radiopharmaceutical (such as fluorodeoxyglucose (FDG) labeled with ^{18}F) is chosen, PET can be used to identify cancer cells with high metabolic activity [36]. However, PET is limited by the rate of uptake of the radiopharmaceutical in the tissue and by the radioactive decay of the radionuclide involved, which make long acquisition times (2-4 min/bed position in clinical routine) necessary to produce diagnostic quality images. Due to patient motion during the acquisition, PET images represent the motion-averaged activity distribution over several minutes, and therefore, motion-induced artifacts are common [37]. This is especially true when imaging areas that are more prone to movement due to breathing, such as those in the lower thorax and upper abdomen. The end result is poorer image quality and a reduction in apparent uptake, which can impair diagnostics and cancer localization for patients with lung cancer, esophageal cancer, pancreatic cancer, and liver lesions [38, 39, 40].

To account for motion, retrospective phase binning (RPB) is commonly used to produce respiratory-gated PET images [41]. In this scenario, the patient's breathing pattern is monitored using an external device [42, 43, 44] or data-driven methods [45, 46, 47] and the events detected during a single phase of the breathing motion are used to reconstruct the image while the remaining detections are disregarded. Although this mitigates the majority of the breathing motion, the main disadvantage of this technique is the poor efficiency of the data acquisition, which results in noisy images. Solving this issue often requires a higher injected activity, resulting in a higher patient effective dose and imaging costs. Alternatively, the acquisition times can be increased (6 min/bed position), which decreases patient throughput and increases patient discomfort. Neither of these solutions are ideal.

Alternatively, registration between the different phases of the breathing motion using a global non-rigid registration method have been proposed [48]. This has the advantage that all of the detected events are used to generate a single image. Unfortunately, registration of noisy images (i.e., individual gated ones) can be challenging. As a result, these methods often depend on the registration of matching gated low-dose computed tomography images, which can be misaligned with the gated PET images [49]. Furthermore, the details of these methods have not been made publicly available.

Registration of PET images from different phases or amplitude ranges of the breathing cycle can also be framed as an optical flow estimation problem; similar to learning the pixel-wise shift (or *optical flow*) between subsequent frames in a video. Estimating the optical flow between gated PET images has been accomplished before using the Lucas and Kanade algorithm [50, 51]. In this scenario, to develop a solvable system of equations, the optical flow is assumed to be constant in small local volumes and the least-squares method is used to solve for the flow in each small volume.

To accomplish this, the spatio-temporal derivatives in the PET image data must be calculated, which requires smoothing of the data to compensate for the effect of noise.

Alternatively, optical flow estimation has been an area of focus within computer vision where the goal is to allow the model to *learn* how to produce an accurate optical flow when provided with two image frames. This can be accomplished using supervised learning with convolutional neural networks (CNNs)[52, 53]. In this case, the ground truth optical flows between all of the training samples are known *a priori*, which provides the network with a direct reference to compare the predictions against. Since access to ground truth optical flows is not realistic for many applications, unsupervised methods have also been developed [54]. Accessing ground truth optical flows for clinical PET data is impossible, and therefore, unsupervised techniques are required to train a network on patient data.

In this paper, we propose FlowNet-PET, which is a modification of these prior works in computer vision and is capable of performing the registration between three dimensional (3D) PET images acquired during different amplitude ranges of the breathing motion. Once trained, the model groups a set of gated PET images into a motion-corrected single bin, providing a final image without the blurring effects that were initially observed. The method was applied to simulated data; therefore, the work described in this paper reflects a proof-of-concept for the proposed framework. Importantly, due to the unsupervised nature of the training, FlowNet-PET can be easily extended to clinical patient images in future work. Additionally, having a neural network estimate the optical flow rather than apply the transformation itself provides a more interpretable model, which is preferable in clinical applications. To improve the accessibility of this method, the code and datasets have been made publicly available along with the training and test sets (https://github.com/teaghan/FlowNet_PET).

4.2 Materials and Methods

4.2.1 The Training Data

To facilitate this proof-of-concept, a lung cancer dataset of 300 phantoms was generated using the 4D extended cardiac-torso (XCAT) anthropomorphic digital phantom [55] with (2×4×4) mm^3 voxels. The parameters for each XCAT phantom were varied randomly. These included the gender, axial sections included in the field of view, transaxial shifts, scaling factors in each direction, size and location of the lung lesion, extent of the diaphragm motion, and extent of the chest expansion. The distributions of each of these randomized parameters are shown in Table 4.1. Additionally, the activity concentration in each organ and tumor was varied based on activity distributions observed in clinical data.

Table 4.1: The distributions of the randomized XCAT parameters used to generate the training data.

Parameter	Distribution
Gender	Male / Female
Axial section in image	Any FOV including the thoracic region
Transaxial shifts	Uniform(-4 cm, 4 cm)
Axial scaling factor	Uniform(0.7, 1.3)
Transaxial scaling factor	Uniform(0.8, 1.2)
Lesion diameter	Uniform(1 cm, 3 cm)
Lesion location	Within the lungs
Axial diaphragm motion	Uniform(0.9 cm, 2.1 cm)
Chest expansion	0.7 × (Diaphragm motion)

The dataset of 300 XCAT phantoms was split into training (270) and validation (30) sets (the test sets are explained in Section 4.3). For each phantom, ten frames spaced across a single breathing cycle were created, each frame having (108×152×152) pixels representing the activity distribution throughout the phantom. To emulate a simple representation of PET data acquisition, these distributions were used to sample a random number (between 1×10^6 and 9×10^6) of counts per frame. Since

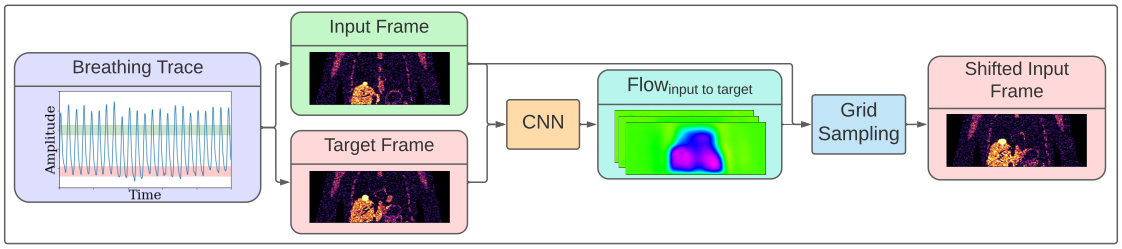


Figure 4.1: A schematic diagram of how a single PET frame is aligned to another frame using FlowNet-PET.

each phantom consisted of ten frames, this provided 90 paired PET samples to train on per phantom and a total of 24300 training samples.

The intention of creating a large training set of digital phantoms with variability was to provide a simple modeling of some of the variations that could be seen in a dataset of clinical patient data; the goal was not to necessarily recreate real data perfectly. This is because the method is unsupervised and can be retrained (or fine-tuned) on a dataset of clinical gated patient images, which would intrinsically include the necessary variability. However, the XCAT data provides a valuable dataset for the development and initial evaluation of the FlowNet-PET motion correction method for several reasons: (1) although the training was unsupervised, having access to the ground truth optical flows provided a robust source of validation when comparing models and reporting the accuracy of the final model, (2) knowing the true activity distribution within the phantom made it possible to track the pixels that belonged to different parts of the phantom (for instance, which pixels belonged to the tumor before and after the corrections were applied), and (3) the same phantom could be used to produce images with and without motion, providing a ground truth image to compare the corrected images against.

4.2.2 The FlowNet-PET Framework

FlowNet-PET was constructed to accomplish the task of predicting the optical flow between two PET images acquired from different amplitude bins of the breathing motion. These two frames are referred to as the input and target PET frames. The estimated flow can then be applied to the input frame through a grid sampling procedure to align the detected counts with the target PET frame. This procedure is outlined in Figure 4.1.

An in-depth description of the CNN architecture used in FlowNet-PET is provided in Section 4.7.1 of the Appendices, which includes a detailed diagram of the network (Figure 4.10). In brief, the network takes two PET frames as inputs and subsequently downsamples the images into a feature map through a series of 3D convolutions. The feature map is then upsampled in steps back to the original resolution of the images. At each step, the optical flow is predicted - then enhanced - until an optical flow of the original resolution is produced. The grid sampling procedure then uses the optical flow to index pixels in the input PET frame and place them in new locations within the shifted PET frame. A more detailed explanation on the grid sampling is provided in Section 4.7.2 of the Appendices. The method was implemented using Python 3.7 and PyTorch 1.10.0.

4.2.3 Training FlowNet-PET

To make the method transferable to clinical patient data, an unsupervised procedure was adopted for training [54]. Namely, rather than comparing the predicted optical flows to the ground truths, a pixel-wise comparison was made between the target frame and the shifted input frame. This objective function is referred to as the photometric loss term, which is explained in detail in Section 4.7.3 of the Appendices. Due to the noise found in PET images, performing a pixel-wise comparison could re-

sult in the alignment of the noise characteristics rather than the general distribution of the PET detections. Therefore, the unsupervised learning procedure was adjusted slightly such that the target and shifted frames were blurred with a Gaussian filter before being compared in the loss function. Upon investigation, a filter size of (15×15×15) voxels with a standard deviation of (1.8×3.6×3.6) mm was found to produce the best results for this blurring. By minimizing this photometric loss, the network learned to produce optical flows that successfully shifted the signal in the input frame to match the target frame.

Another development that is unique to our method is how we produced invertible optical flows. In PET images, the pixel intensities can be related to the number of detected coincidences in a particular location. Allowing the network to artificially change the number of counts would remove the quantitative imaging advantage of PET. Therefore, to restrict the network to produce optical flows that do not remove or multiply individual counts, we added a second loss term that constrained the network to produce one-to-one pixel correspondences. In short, the optical flows predicted by the network were constrained to be invertible by predicting the flow between the two frames in both directions, where one flow should reverse the effect of the other. The final loss function that was minimized throughout training was a combination of the photometric and invertibility loss terms. More details regarding the formalism of the invertibility constraint are provided in Section 4.7.3 of the Appendices.

4.3 Experiments

4.3.1 Model Selection

Throughout training, not only were the two loss terms evaluated on both the training and validation sets, but the accuracy of the optical flows was evaluated for the validation set using a mean-absolute-error metric. This later evaluation was not used for

training; it helped assess the true performance of the model both during training and after the training had finished. Having this robust metric allowed for the selection of network and training parameters such as the model architecture, learning rate, loss weights, number of training iterations, and image smoothing parameters. While these tests were too extensive to include in this paper, it is worth mentioning that these parameters were optimized by comparing the predicted optical flows against the ground truth optical flows within the validation set. The training progress of the final model is shown in Figure 4.12 in Section 4.7.3 of the Appendices.

4.3.2 Testing on XCAT Frames

The network was first applied to a test set resembling the same format of the data used to train the network. This test set consisted of six phantoms that were created with identical parameters except for the extent of breathing motion, where the maximum diaphragm motion was varied from 9 mm to 21 mm (the same range as the training data). Each phantom had ten PET frames spaced across a single breathing cycle and a lung lesion with a diameter of 25 mm was placed at the top edge of the diaphragm where the largest amount of motion takes place. Additionally, there were 12×10^6 counts spread out equally across the ten frames using the same sampling method as the training set. The accuracy of the optical flow estimates was assessed by predicting the flow from peak inhalation to the end of exhalation and comparing this prediction to the ground truth flow, only considering pixels where activity was present.

Using this same test set, nine of the ten frames from each phantom were shifted using FlowNet-PET to be aligned with the tenth frame and then summed together. This provided a "corrected image" that was compared to an image produced without motion. The improvements were quantified by creating a mask that selected a volume of interest (VOI) containing the tumor pixels in the no-motion, uncorrected, and cor-

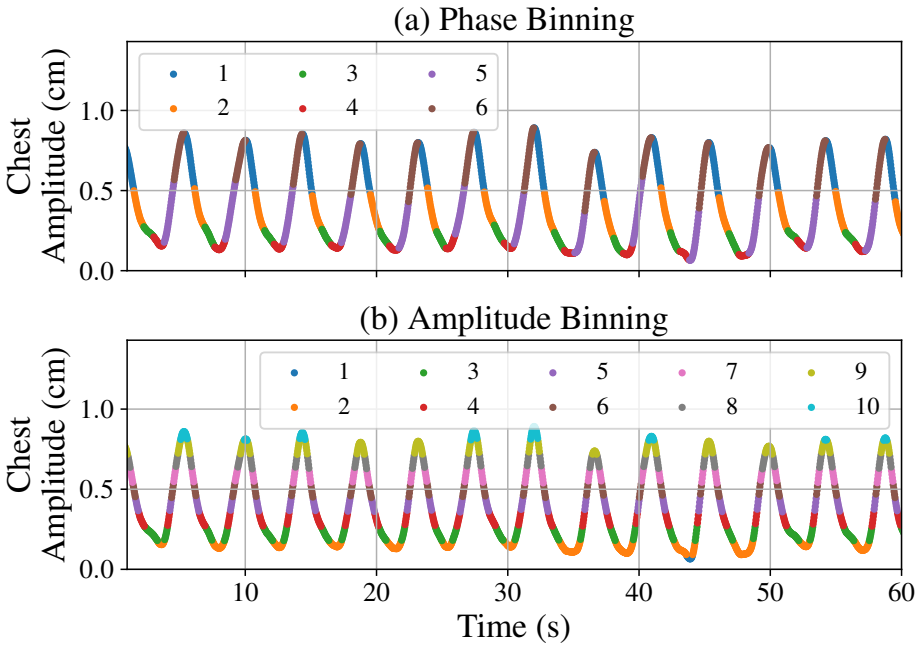


Figure 4.2: Time points in a clinical breathing trace binned based on (a) phase and (b) amplitude.

rected images. Using these masks, three metrics were computed: (1) the intersection over union (IoU) of each mask with the no-motion mask, (2) the activity enclosed in each image using the no-motion mask, and (3) the coefficient of variation (CoV) in each image using the no-motion mask. The IoU between two VOIs is defined as the ratio of the volume of overlap to the volume of the combined VOIs. The CoV is the ratio of the standard deviation to the mean within the VOI. The relative improvements for each of these three metrics were calculated as the percentage of the original residual in the metric (between the no-motion image and the non-corrected image) that was resolved with the FlowNet-PET corrections.

4.3.3 Comparisons to Retrospective Phase Binning

To compare the improvements against those achieved with retrospective phase binning (RPB) [41], another test set was generated to introduce motion within each bin. For this, the breathing amplitude was varied within the same range as before, but it was

divided over 24 time frames rather than ten. Additionally, the PET data collection was modeled by using a clinical breathing trace and the number of counts in each frame was proportional to the time spent in that frame. To create the RPB image, the breathing pattern was split into six equally-spaced phase bins and the detected events in the most quiescent phase were selected as the RPB image. To produce the FlowNet-PET-corrected image, the breathing pattern was split into ten equally-sized amplitude bins between the minimum and maximum extent of the breathing motion. Amplitude binning was chosen to limit the motion within each bin [56, 57, 58]. The difference between binning the time points based on the phase and amplitude is visualized in Figure 4.2. To easily compare against RPB, the FlowNet-PET-corrected image was created by shifting nine of these ten bins to align with the bin that had the largest overlap with the quiescent phase. Finally, a separate image without motion was created in the quiescent phase. Each of these images (uncorrected, RPB, FlowNet-PET-corrected, and no motion) had a total of 12×10^6 counts; however, it is important to note that, in theory, the RPB image required a scan duration that was six times longer in order to "acquire" the same number of counts found in the other images. As before, the IoU, number of enclosed counts, and the CoV of the tumors in each image were compared.

4.3.4 Comparisons using Monte Carlo Data

Monte Carlo (MC) simulations of PET imaging were used to generate images whose characteristics more accurately emulated those of clinical patient images. These simulations were accomplished using our previously-validated pipeline of a Siemens Biograph mCT scanner [59]. The RPB, FlowNet-PET-corrected, and no-motion images were simulated to involve 100 seconds of scan time; however, the RPB required a total scan duration of 600 seconds to acquire the same amount of signal.

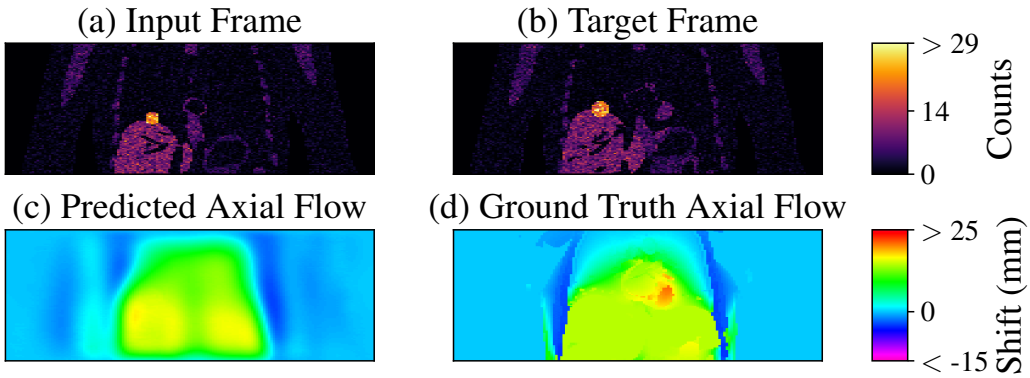


Figure 4.3: Coronal views of (a & b) two XCAT frames from the same phantom, (c) the predicted axial optical flow, and (d) the ground truth flow between the two frames.

To obtain images that were more representative of the raw counts, the images were reconstructed without decay correction and no calibration factor was applied. As a result, we used the images in units of counts instead of activity concentrations. Six phase binned attenuation maps were generated for the attenuation correction and the phase with the largest overlap was used for the attenuation correction of each PET image.

The MC simulations helped determine if the discrepancies found between the RPB method and the corrections provided by FlowNet-PET would be mitigated by the blurring due to the actual PET imaging procedure. Through the simulation process, the exact locations of the pixels corresponding to the tumor activity were lost; therefore, these images were used for a qualitative comparison between the two methods.

4.4 Results

4.4.1 Testing on XCAT Frames

An example of two XCAT frames (full inhalation and full exhalation) from the test phantom with a maximum diaphragm shift of 15 mm is shown in Figure 4.3. The pre-

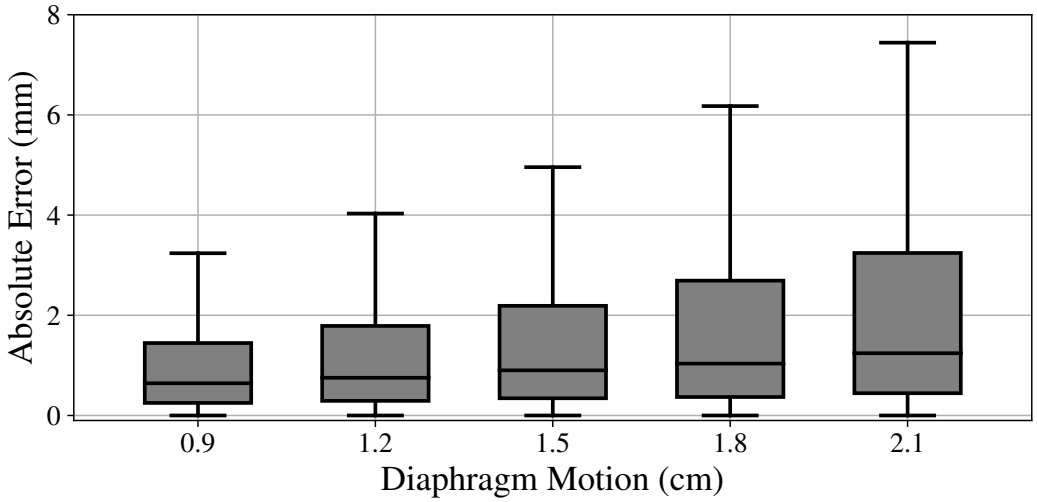


Figure 4.4: Distributions of the absolute residual errors between the predicted and true optical flows from the maximum inhale frame to maximum exhale for a variety of simulated motion extents. Each box extends from the lower to upper quartile values of the data, with a line at the median. The whiskers show the range of 1.5× the interquartile range beyond the lower and upper quartiles.

dicted and ground truth optical flows between the two frames are also presented, showing some over-smoothing present in the FlowNet-PET prediction. The breathing motion also resulted in shifts in the lateral and anterior-posterior directions (not shown) that caused the tumor to move in and out of the coronal views. The absolute residual errors in the optical flow predictions are summarized in Figure 4.4. In all cases, the median absolute residual error was found to be smaller than the slice thickness (2 mm) and the pixel width (4 mm) of the images. In fact, 75th percentile of the residual errors were all within the pixel width of 4 mm.

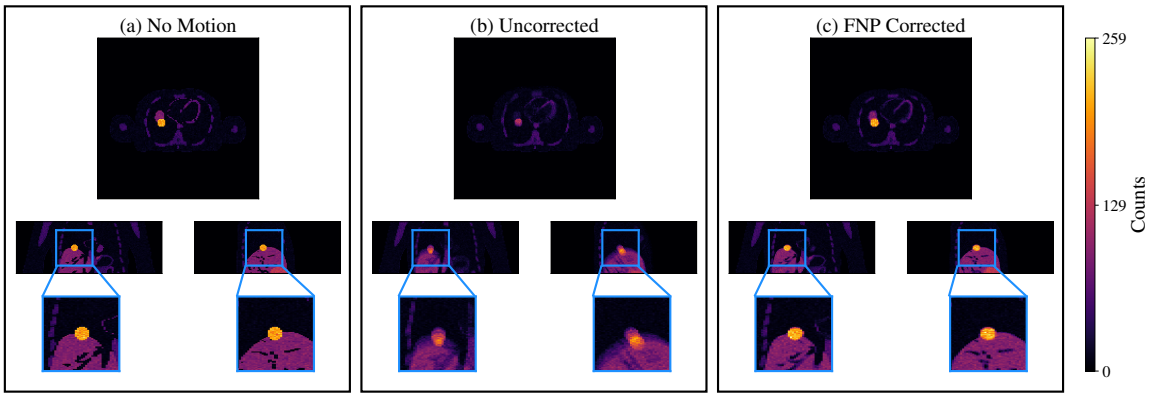


Figure 4.5: An example of an XCAT PET image with (a) no motion, (b) motion, and (c) the motion correction accomplished by FlowNet-PET.

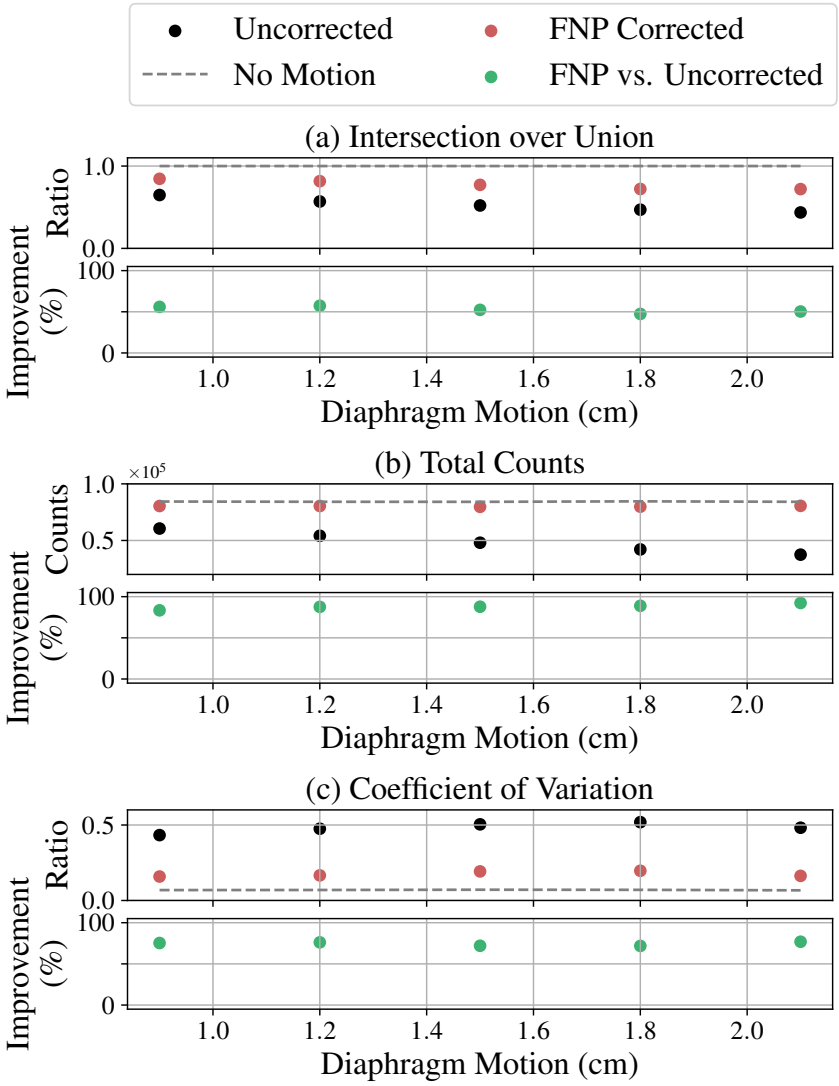


Figure 4.6: Image comparisons with the no motion image in terms of the (a) intersection over union, (b) total number of enclosed counts, and (c) coefficient of variation both before and after the FlowNet-PET (FNP) corrections were applied. The relative improvements are shown for a range of breathing motion amplitudes.

Figure 4.5 shows the sum of all ten frames from the test phantom with a maximum diaphragm shift of 21 mm before and after the FlowNet-PET correction was applied. A summary of the correction performance is shown in Figure 4.6 along with the improvements relative to the uncorrected images. Across the range of tested breathing amplitudes, the average relative improvements were 54%, 90%, and 76%

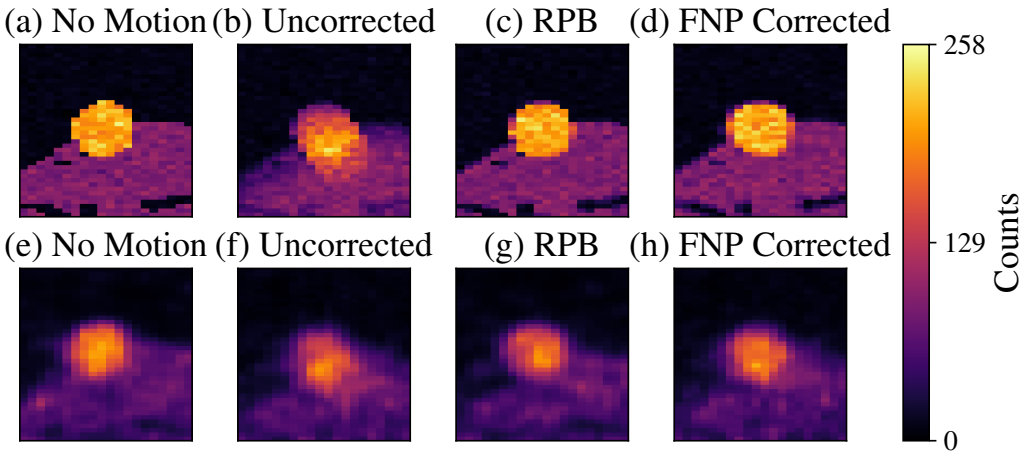


Figure 4.7: Sagittal views of the tumor in the XCAT phantom (a & e) without motion, (b & f) with motion, (c & g) retrospectively phase binned, and (d & h) with motion but corrected for with FlowNet-PET. The top row of images depicts the results when applied to the binned XCAT phantoms, whereas the bottom row is applied to Monte Carlo data.

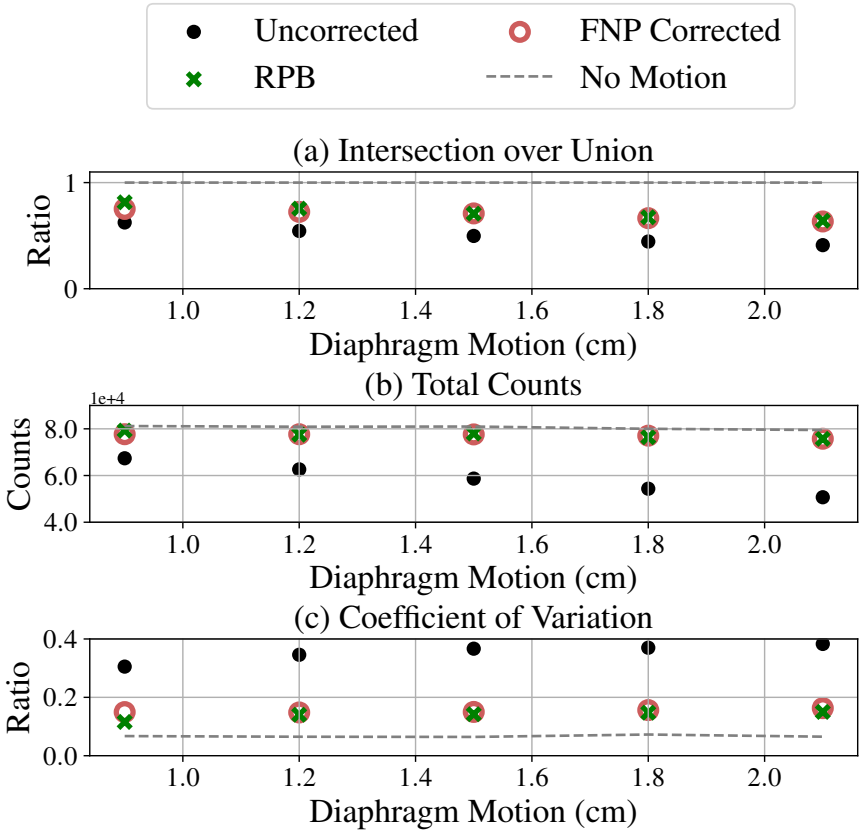


Figure 4.8: Motion correction metrics comparing retrospective phase binning (RPB) to the corrections provided by FlowNet-PET (FNP).

for the IoU, total activity, and CoV, respectively.

4.4.2 Comparisons to Retrospective Phase Binning

The top row of Figure 4.7 shows a visual comparison of the improvements provided by RPB and FlowNet-PET when applied to the binned XCAT PET frames with a diaphragm shift of 18 mm. As shown in Figure 4.8, both methods provide a similar correction performance. For instance, the average relative improvements with FlowNet-PET vs. RPB were IoU: 40% vs. 44%, total activity: 83% vs. 85%, and CoV: 71% vs. 75%, respectively. However, RPB involved a scan time that was six times longer than that of the image corrected with FlowNet-PET.

4.4.3 Comparisons using Monte Carlo Data

The bottom row of Figure 4.7 shows a visual comparison of the improvements provided by RPB and FlowNet-PET when applied to the MC-simulated XCAT phantom with a diaphragm shift of 18 mm. Both methods are found to largely mitigate the breathing motion artifacts. As a qualitative assessment, Figure 4.9 shows the profiles drawn through the center of the tumor. As expected, the counts in both the RPB and FlowNet-PET profiles are more condensed compared to the uncorrected image, creating tumor profiles with a higher peak.

4.5 Discussion

In PET imaging, the pixel intensities are related to the number of counts detected by a particular pair of detectors. Therefore, it is reasonable to assume that – when shifting the input frame using the grid sampler – individual counts should not be removed or multiplied. This type of constraint is not enforced in the traditional FlowNet framework, since the typical datasets are constructed from images of pixelated color

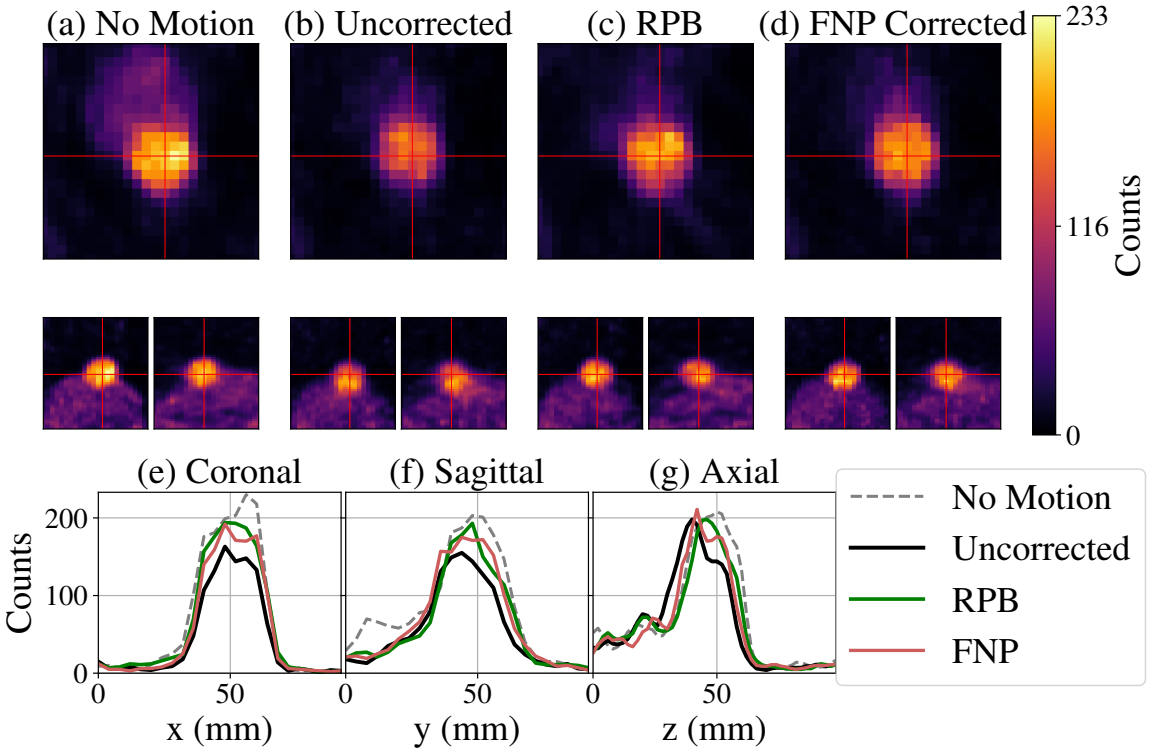


Figure 4.9: Profiles drawn through the center of the tumor in the Monte Carlo-generated PET images (a) with no motion, (b) with motion, (c) retrospectively phase binned, and (d) with motion but corrected with FlowNet-PET.

intensities. However, to constrain the network to not artificially add or subtract detections in FlowNet-PET, the invertibility loss was optimized during training.

Additionally, in previous applications of FlowNet frameworks, a smoothness loss term has been used to constrain the network to produce optical flows that are spatially consistent [54]. This loss term was attempted in FlowNet-PET, however, the performance was found to degrade. Interestingly, the effect of producing smooth optical flows was still accomplished through the use of the invertibility constraint, and as a result, a smoothness loss term was not included in our work. Evaluations like this and the choices outlined in Section 4.3.1 are only possible due to the access to ground truth optical flows. Therefore, developing FlowNet-PET with XCAT data is ideal and these choices would be transferred to the model applied to clinical

data.

To perform the motion correction with FlowNet-PET, the PET data was binned based on the amplitude of the breathing motion. Amplitude binning was chosen to limit the amount of motion within each bin, which is visualized in Figure 4.2. It can be seen that amplitude binning results in a similar amount of motion within each bin, whereas phase binning results in less motion in some bins and more motion in others. As a result, when adopting RPB, it is logical to use phase binning since only the detections acquired within a single bin (with a small amount of motion) are selected and the remainder of the data is disregarded. It is also evident in Figure 4.2 that a few of the amplitude bins account for a small fraction of the total scan time. This is not an issue for FlowNet-PET because these bins will only have a small number of counts in their images and have a minimal contribution to the overall corrected image. The choice of using ten amplitude bins was made to provide the majority of the bins with a high enough signal-to-noise yet still limit the amount of motion within each bin. However, if FlowNet-PET was used in a clinical setting, the number of bins could be left as a parameter to be chosen by the clinician since this does not require retraining the model.

An example of a predicted optical flow was compared against the ground truth flow in Figure 4.3 to exemplify a limitation of this method. Namely, FlowNet-PET is found to produce optical flows that are over-smoothed, which is a characteristic of the FlowNet architecture [60] and is likely exaggerated by using an unsupervised objective. For instance, if the network was trained with supervised learning – comparing the predictions to the ground truth flows - the optical flows that FlowNet-PET produced would likely be much closer to the ground truths. However, performing such supervised learning with clinical data would be very challenging, if not impossible.

As seen in Figure 4.5, the inaccuracies in the optical flow estimates do not trans-

late into noticeable errors in the corrected images. This is not surprising because the unsupervised learning objective was a comparison between the shifted and target frames in the image-space. Nevertheless, there are a few solutions that could help overcome the limitation of producing over-smoothed optical flows. For example, FlowNet-PET could be initially trained with a supervised objective function by using the XCAT phantoms, then fine-tuned with clinical data using an unsupervised learning objective. Another solution would be to adopt a different framework that produces optical flows by iteratively updating the flow matrix [53]. These iterative methods have been found to produce sharper optical flows compared to frameworks such as FlowNet.

Another limitation that can be observed in Figure 4.3 is that FlowNet-PET does not predict the heart motion accurately. Evidently, since the breathing motion is used to separate the frames in the training set, the network has learned to ignore heart motion because it is not correlated with the phase of the breathing motion. This type of interpretability of the framework potentially enhances the applicability of the method. This is because - rather than having the neural network apply the transformation internally - the model produces an optical flow, which is human interpretable. The optical flow is then applied to the input frame through a grid sampling algorithm that can also be easily understood. In this sense, having a network that produces the optical flows rather than applying the transformations internally is desirable. We leave the door open to training FlowNet-PET with cardiac-gated data.

As seen in Figure 4.6, the relative improvements obtained with FlowNet-PET were fairly constant across the tested breathing amplitudes. Of the metrics used to evaluate image correction, the enclosed activity is likely the best representation of tumor depiction accuracy. For instance, if a small amount of activity is left outside of the true tumor volume, this could have a large effect on the IoU, yet it will not

substantially change the image quality. Conversely, this will only impact the enclosed activity metric when the activity left out of the true volume is substantial.

To introduce some variability in the training set, several XCAT parameters were varied randomly. However, when applying the method to clinical data, the necessary variations would be intrinsic to the training set. Additionally, newly acquired images could be added to the training set to continually update the network. For an unsupervised model such as FlowNet-PET, separating the training and test data may be interesting for a proof-of-concept; however, there is no need to keep the two separate once the method is being applied in a clinical setting. Furthermore, continually fine-tuning the network on new data could be made specific for each institution that adopts the method. In this scenario, the network could be updated based on newly acquired data from that particular clinic; therefore, the network would be more specifically tuned to the protocols and scanners of that institution.

When comparing FlowNet-PET to RPB, the performance metrics showed similar results for both methods and the images shown in Figure 4.7 had indistinguishable differences. Furthermore, as seen in Figure 4.7, the PET spatial resolution limitations were found to be more substantial than any visual differences in the images produced by the two methods. To easily compare the RPB and FlowNet-PET images, the amplitude bin that had the largest overlap with the quiescent phase was chosen as the target. However, when adopting FlowNet-PET in a clinical setting, it would likely make sense to choose the amplitude bin with the largest number of counts as the target bin.

When applying the network to the MC simulated data, even though the model was not trained on these types of images, FlowNet-PET was still effective at reducing the motion artifacts. It can be expected that, when trained on clinical images, the network would be even more effective since it would have learned from images that

resulted from the same imaging process.

Another intuitive way to register the PET data would be to align the raw detections before the images are reconstructed. This could either be accomplished in sinogram-space or with the raw list-mode data. However, it is unlikely that performing the registration in these domains could be accomplished through the use of optical flows. This is because the use of optical flows assumes that the shifts are localized, which requires the signal to be continuous. The signal in sinograms are continuous within each segment of the sinogram, but – in order to accurately depict the breathing motion of the patient – some of the detected events found in the input sinogram frame would have to be shifted to another segment to be properly aligned with the target frame. This type of shift would no longer be localized; therefore, current optical flow predictive methods would likely fail. Furthermore, the shifts in the list-mode data would be even less localized since the detections are not yet organized to be continuous.

It is important to note some of the limitations of this method. Most importantly, FlowNet-PET is capable of correcting for the motion between the different amplitude bins, but cannot correct the motion within the individual bins. Therefore, the method is limited by the gating method that is adopted. In addition to not correcting for cardiac motion, any movement of the patient within the scanner will not be corrected for properly. In this scenario, the movement would result in the blurring within one or more of the binned images and this blurring would remain present after the corrections were applied. Therefore, when applied to clinical data, it would likely be beneficial to have individual time points flagged as unusable when these types of circumstances arise. Furthermore, FlowNet-PET is limited by the signal-to-noise ratio of the individual PET frames; having fewer counts in the individual frames will result in poorer predictions. Accordingly, using the frame with the largest number

of counts as the target bin would provide the best starting point to perform the corrections.

The applicability of this method is not limited to PET; the FlowNet-PET framework could be applied to low-dose CT images as well as other imaging modalities. Furthermore, it could be used dynamically during radiation therapy. For instance, amplitude-binned low-dose CT images could be acquired prior to treatment to predict the optical flow between the different bins. Then, during treatment, only the patient's breathing would be monitored in order to determine the current and next bin. Using the optical flow between these two bins, the treatment volume could be traced to its new location within the patient. Since the optical flow predictions were made prior to the treatment, this process would be efficient and could be applied in real-time.

The code required to train FlowNet-PET, the analysis code, and datasets have been made publicly available (https://github.com/teaghan/FlowNet_PET).

4.6 Conclusion

FlowNet-PET was developed to correct for breathing motion in PET imaging using unsupervised learning. This framework was applied to XCAT phantom data to illustrate a proof-of-concept. FlowNet-PET produced interpretable optical flows that were used to shift the amplitude-gated images into a single bin. The mitigation of breathing motion artifacts was shown both qualitatively and quantitatively. Lastly, when compared to the retrospective phase binning method, FlowNet-PET was found to provide similar results, but only required one sixth of the scan duration. Since the training is unsupervised, the method is transferable to clinical data without the need for human intervention.

4.7 Appendices

4.7.1 The Network Architecture

A schematic diagram of our CNN is shown in Figure 4.10. The network architecture was based on "FlowNet Simple" [52] where the 2D convolutions were replaced with 3D ones to account for PET images. Transitioning the network from a 2D-CNN to a 3D-CNN comes with substantial costs in memory use and computation time. Therefore, the network was simplified to consist of fewer convolutional layers and filters, both of which can be inferred from Figure 4.10. As seen in this Figure, the network takes two PET frames as inputs and subsequently downsamples the images through a series of 3D convolutions. Each of the downsampling blocks consists of two 3D convolutional layers; both layers use a filter size of (3×3×3) pixels, but the first layer uses a stride length of two pixels in all three directions whereas the second layer uses a stride length of one pixel. The convolutions are each followed by a ReLU activation. Following the downsampling blocks, the feature map is then upsampled in stages, where each upsampling operation includes a single transposed convolutional layer with a filter size of (4×4×4) pixels and a stride length of two pixels, which is followed by a ReLU activation.

At each intermediate resolution, there is a flow estimator, which consists of a single convolutional layer with a filter size of (5×5×5) pixels and a stride length of one pixel. Each flow estimator receives the intermediate outputs produced throughout the network that are at the matching resolution, including the upsampled lower resolution flow. The combination of these inputs is clearer in Figure 4.10. The idea behind these skip connections is to provide the later parts of the network with information that has potentially been lost in previous operations and make the learning process easier [61]. The final outputs of the network are the optical flows estimated at each resolution,

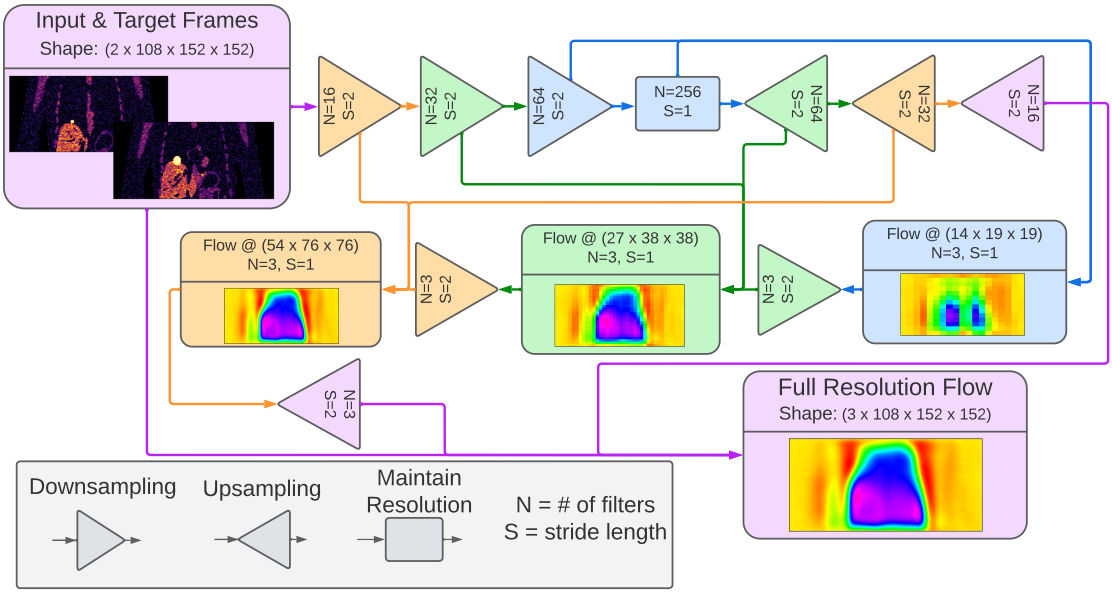


Figure 4.10: A diagram of the CNN architecture used in FlowNet-PET. Each colored shape represents a convolutional block of the network and the arrows represent the connections between them. The color depicts the resolution of the output of that block, whereas the shape describes the type of operation (downsampling, upsampling, or constant resolution). For the blocks that produce outputs with intuitive representations, sagittal slices of the 3D image are shown; the remaining outputs are non-intuitive feature maps.

including the original resolution of the input frame. This allows for comparisons to be made between the shifted input frame and target frame at various resolutions, which also helps provide a smoother learning process [52].

4.7.2 Grid Sampling

Understanding the training objectives outlined below requires a brief background on the grid sampling procedure, T, that performs the transformation to align the input frame, I_{inp}, with the target frame, I_{tgt}. An example of how this operation works for a small 2D grayscale image is shown in Figure 4.11. In the case of 3D images, the baseline grid of pixel locations, G_0, consists of a 3D grid, where each pixel contains a unique combination of x, y, z pixel locations. The forward flow from the input to

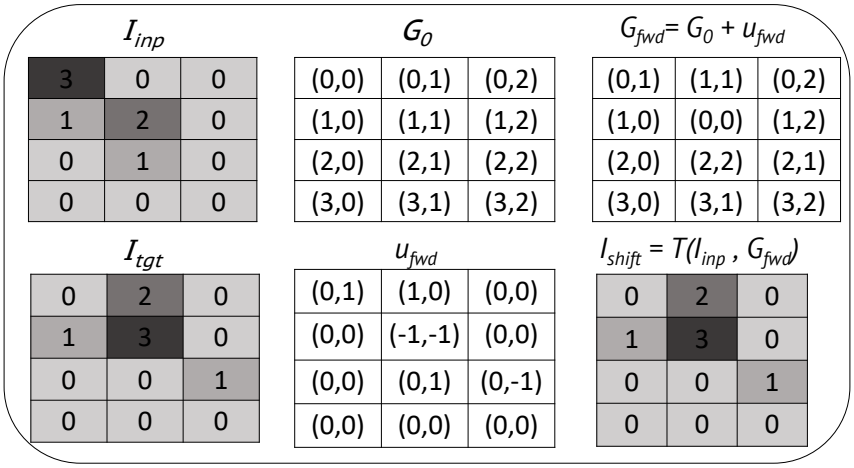


Figure 4.11: An example of the grid sampling procedure for a small 2D grayscale image.

the target frame, u_{fwd}, is used to adjust the pixel locations in the grid to produce a new grid, $G_{fwd} = G_0 + u_{fwd}$. The transformation is then performed by sampling I_{inp} using G_{fwd} :

$$I_{shift} = T\left(I_{inp},\ G_{fwd}\right) . \tag{4.1}$$

4.7.3 Training the Model

The objective function used to train FlowNet-PET consists of two loss terms: the photometric loss and invertibility loss. For the photometric loss, a Charbonnier penalty function was chosen to mitigate the effects of outliers [62]. Namely, when comparing a shifted frame to its target frame, this loss term can be summarized as

$$\rho_{photo}(u_{fwd}) = \frac{1}{N}\sum_{i=0}^{N}\left(\left(I^i_{shift} - I^i_{tgt}\right)^2 + \epsilon^2\right)^{\alpha} , \tag{4.2}$$

for N pixels in each frame. Based on previous work [54], α and ϵ were chosen to be 0.25 and 1×10^{-9}, respectively.

Counter to the photometric loss term, the invertibility constraint operates on

the pixel grid rather than the images themselves. In more detail, the invertibility constraint is enforced by also predicting the backward flow from the target to the input frame, u_{bwd}. Then, by using $G_{bwd} = G_0 + u_{bwd}$ in the grid sampling procedure to sample G_{fwd}, a cycled grid is produced:

$$G_{cyc} = T\left(G_{fwb},\ G_{bwd}\right) . \tag{4.3}$$

Accordingly, if G_{cyc} is constrained to return to G_0, then u_{fwd} is invertible and u_{bwd} is its inverse. This constraint is enforced by adopting the invertibility loss term, which can be formalized as

$$\rho_{inv}(u_{fwd},\ u_{bwd}) = \frac{1}{3N}\sum_{i=0}^{N}\sum_{j=1}^{3}\left(G_{cyc}^{i,j} - G_0^{i,j}\right)^2 , \tag{4.4}$$

where $j = 1, 2, 3$ represent the $x,\ y,\ z$ coordinates.

Other implementations of FlowNet have utilized a similar cycle-consistency where the constraint is imposed on the cycled input frame [63]. However, this does not guarantee a one-to-one pixel correspondence since there could be multiple pixels with the same intensity, allowing cycle-consistency to be achieved without ensuring that the flow itself is invertible. With our method, since the baseline grid has a unique combination of $x,\ y,\ z$ locations at every pixel, the only possible way for the constraint of $G_{cyc} = G_0$ to be met is if u_{fwd} is invertible and u_{bwd} is its inverse

The final loss function that was minimized through training was the sum of the photometric and the invertibility loss terms

$$\mathcal{L}(u_{fwd},\ u_{bwd}) = \rho_{photo}(u_{fwd}) + \lambda\rho_{inv}(u_{fwd},\ u_{bwd}) , \tag{4.5}$$

where a weighting factor of $\lambda = 1000$ was multiplied to the invertibility term, which helped balance the scaling of the gradients of each loss term. This objective

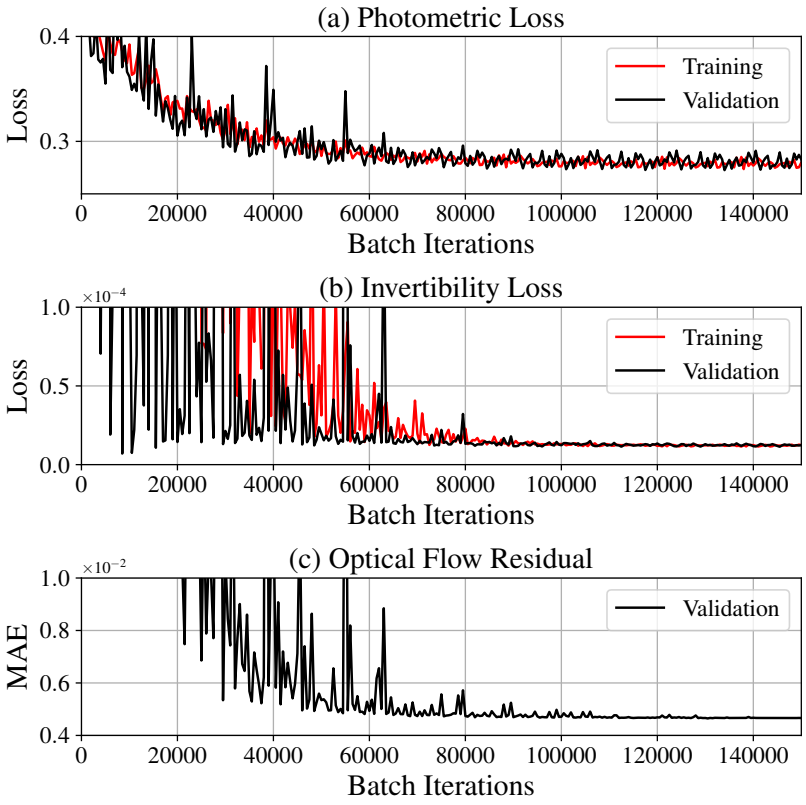


Figure 4.12: The progress of the CNN performance evaluated during training. The (a) photometric loss and (b) invertibility loss were evaluated on the training and validation sets, whereas the (c) mean absolute error between the predicted and ground truth optical flows were evaluated only on the validation set. This later metric was only used to track the true performance of the network and not for the training.

was optimized using the Adam optimizer [64] with an initial learning rate of 0.0003, which was decayed by a factor of 0.7 every 8000 iterations. The model was trained for a total of 1.5×10^5 iterations, using one sample per iteration. The two loss terms were evaluated on the training and validation sets throughout the training of the model and are shown in Figure 4.12. Also shown in this plot is the mean absolute error evaluated between the predicted and ground truth optical flows in the validation set, which was not used for training.

Chapter 5

Discussion and Future Work

The initial motivation for developing a MC pipeline for PET imaging was to create a large set of images that would then be used to train the motion correction neural network. By using the XCAT phantom in these simulations, the PET detections acquired from a breathing patient could be accurately emulated and the ground truth activity distributions (with and without motion) would be known. Therefore, a network like FlowNet-PET could be trained on the sinogram representations of the data instead of the images. As described in Section 2.1.2, the reconstruction algorithms depend on the amount of signal in the sinograms. As a result, reconstructing several images from low signal-to-noise (S/N) sinograms and adding these images together produces a less accurate image than reconstructing an image from a single high S/N sinogram. Therefore, performing the corrections in the sinogram-space is more ideal since the corrected data could be combined *before* reconstructing the image. However, as noted in Section 4.5, the lack of continuity in the sinogram information makes it difficult (if not impossible) to apply algorithms like FlowNet to this type of data. For this reason, FlowNet-PET was trained on PET images, removing the necessity for a large MC-produced dataset. Rest assured, the extensive validations of our MC pipeline were not a waste of the reader's time: using these simulations, Figure 4.9 shows a qualitative assessment of the performance of our method.

Despite the limited use of the MC simulations, it actually turns out that we produced something that likely has clinical relevance. The results shown in Chapter 4 provide encouragement that FlowNet-PET can perform accurate respiratory motion corrections without the requirement for higher injected activities and/or longer scan durations. In addition to performing accurate motion corrections without protocol adjustments, this method has two notable features: (1) the results are human-interpretable and (2) the training is unsupervised. The interpretability is accomplished by having the CNN produce optical flows rather than perform the corrections internally. This should mitigate some of the current skepticism surrounding neural networks: having to trust a *black box* to perform adjustments and/or inference on the measured data. Instead, by predicting optical flows, the motion that the CNN has detected can be easily interpreted, understood, and - in the case of our proof-of-concept - validated. Additionally, adopting an unsupervised learning procedure improves the transferability of this method to clinical patient images and to other imaging modalities. This feature provides a wide-range of possibilities for future applications of the method.

Based on the theme of this dissertation, the most obvious application would be to transfer the training to clinical PET images. For instance, our collaborators at Radboud University Medical Center have acquired a large set of patient PET images while monitoring the patients' breathing motion. This data could be used as a new training dataset for FlowNet-PET. The detections would have to be split based on a particular binning procedure to produce images that represent the activity distributions at different stages of the breathing motion. By training FlowNet-PET on these images, the network would learn from data that intrinsically contains the variations that would be encountered when applied to new clinical data. Depending on the desired flexibility and generalizability, this approach could be made specific

to a particular institution/scanner or extended to a large set of data acquired from several clinics and scanner models.

When transferring the method to clinical data, there are several approaches that could be adopted. The most straight-forward approach would be to retrain the model from scratch on new data. This would resemble the same training procedure performed in Chapter 4 (only with a new dataset) and would likely perform well. Another option would be to take a pretrained and validated FlowNet-PET model that has been trained on XCAT data (such as the one presented in Chapter 4) and fine-tune it on clinical images. This approach has the advantage of starting from a network that is validated against ground truth data, which is impossible to do with clinical images. Since the more significant weight adjustments would be performed on the XCAT data, the fine-tuning should be performed with a much lower learning rate. This would help avoid having the network adjusted so substantially in the fine-tuning stage that the pretraining on XCAT data would be undone.

As mentioned in Section 4.5, a third approach to transferring the method to clinical PET images would be to have FlowNet-PET pretrained on XCAT data using supervised learning, then performing the fine-tuning on clinical images using the unsupervised learning procedure shown in Chapter 4. This option is the most technical and would likely require more experimentation than the other two approaches. This is because, when switching from the pretraining to fine-tuning, multiple discontinuities are introduced in the training: both the training data and the learning objective are being adjusted. Therefore, in addition to using a lower learning rate, it would likely be wise to slowly transition the training from XCAT to clinical images. Once the pretraining is completed, this could be accomplished by initially performing a small percentage of the training iterations on clinical images while still performing the rest of the iterations on XCAT data with the supervised learning objective. The

percentage of iterations performed on the clinical images could be slowly increased until 100% of the training is being performed on clinical images with the unsupervised learning objective.

When transferring these types of methods to clinical images, unexpected challenges often arise. Applying the method to XCAT data was an important first step because the results could be validated in a robust manner; however, this dataset was likely more refined than a dataset acquired from a clinical setting. One difficulty that might arise is in the binning of the PET data. With the XCAT data, it was possible to hand-choose a relatively clean breathing pattern to use for modelling the breathing motion, which made it easy to bin the data. This would likely not be the case with some of the clinical data and finding the most optimal binning procedure would require some experimentation. Another difficult aspect of transferring the method to clinical images is how to validate that the method is performing accurate corrections. For this, it would likely make sense to have retrospectively gated images to compare against. Of course, similar to Section 4.4.2, the FlowNet-PET-corrected images could be generated with only a portion of the overall scan detections to make these comparisons with the RPB images one-to-one in terms of the number of counts in the images.

As mentioned, FlowNet-PET could also be applied to other imaging modalities. For instance, the network could learn the registration between binned low-dose CT images or perhaps combined PET/CT images. In the case of combining PET/CT images, if the two modalities were binned in the same manner, the network would have additional information (*i.e.*, another image channel) to learn the optical flows from. However, these corrections would depend on the accuracy of the the alignment between the PET and CT images. Essentially, assuming the transitions are continuous, any set of data binned based on regular motion could be used in the training

and application of FlowNet-PET.

Furthermore, as mention in Section 4.5, FlowNet-PET could be used to predict organ motion prior to radiation therapy treatment, and these predictions could be utilized by the treatment planning system to dynamically track the treatment volume during treatment. Additionally, FlowNet-PET could be used to register PET/CT images from one visit to another. Having these images aligned would allow oncologists to more easily compare images from separate hospital visits and assess the progression of the cancer and/or the effectiveness of prior treatments. However, this application would be limited by the amount of physiological changes (e.g. weight and tumor size) between visits - changes that are not the result of physical motion/shifts.

Whichever FlowNet-PET applications are undertaken, it will be beneficial to have a starting point to work from, which is provided by making the code and datasets used in Chapter 4 publicly available (https://github.com/teaghan/FlowNet_PET). I hope this repository and the discussion on the possible future applications of FlowNet-PET will be helpful to those who are interested.

Chapter 6

Conclusion

A Monte Carlo pipeline was developed to emulate the PET imaging process of Biograph 40 mCT PET/CT system (Siemens Healthineers). These simulations were validated against experimental measurements in terms of the scanner geometry and several image quality metrics. Images generated using this pipeline helped validate the unsupervised learning method for correcting respiratory motion, FlowNet-PET. This network produced interpretable optical flows that were used to shift the amplitude-gated images into a single bin. FlowNet-PET was trained on anthropomorphic digital phantom data, which allowed the corrections to be validated against ground truth data. When comparing the results to retrospectively phase binned images, FlowNet-PET provided comparable results, but only required one sixth of the scan duration. This work represents an important building block to a potential clinical application of FlowNet-PET, where the method can be applied to improve the efficiency and safety of PET imaging.

Bibliography

[1] Jerrold Bushberg, Anthony Seibert, Edwin Leidholdt, and John Boone. *The Essential Physics of Medical Imaging*. Lippincott Williams & Wilkins, 2012.

[2] Robert E. Henkin. *Nuclear Medicine, Vols 1 and 2, 2nd ed.* Mosby Elsevier, 2006.

[3] Jonathan K. Poon, Magnus L. Dahlbom, Michael E. Casey, Jinyi Qi, Simon R. Cherry, and Ramsey D. Badawi. Validation of the simset simulation package for modeling the siemens biograph mct pet scanner. *Physics in medicine & biology*, 60(3):N35–N45, 2015.

[4] Sara A. Zein, Nicolas A. Karakatsanis, Mohammad Issa, Amin A. Haj-Ali, and Sadek A. Nehmeh. Physical performance of a long axial field-of-view pet scanner prototype with sparse rings configuration: A monte carlo simulation study. *Medical physics (Lancaster)*, 47(4):1949–1957, 2020.

[5] Frederic H. Fahey. Data acquisition in pet imaging. *Journal of nuclear medicine technology*, 30(2):39–49, 2002.

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

[7] Frederic Fahey, Paul Christian, Katherine Zukotynski, Briana Sexton-Stallone, Christina Kiss, Bonnie Clarke, Arzu Onar-Thomas, and Tina Y. Poussaint. Use of a qualification phantom for pet brain imaging in a multicenter consortium: A collaboration between the pediatric brain tumor consortium and the snmmi clinical trials network. *Journal of Nuclear Medicine*, 60(5):677–682, 2019.

[8] Claudia Kuntner and David Stout. Quantitative preclinical pet imaging: opportunities and challenges. *Frontiers in Physics*, 2, 2014.

[9] Nema iec pet body phantom set. https://capintec.com/product/nema-iec-pet-body-phantom-set/. Accessed: 2022-06-05.

[10] Joao Seco and Frank Verhaegen, editors. *Monte Carlo techniques in radiation therapy*. CRC/Taylor & Francis, Boca Raton, 2013.

[11] J. Allison, K. Amako, J. Apostolakis, H. Araujo, P. Arce, M. Asai, D. Axen, S. Banerjee, G. Barrand, F. Behner, L. Bellagamba, J. Boudreau, L. Broglia, A. Brunengo, S. Chauvie, R. Chytracek, G. Cooperman, G. Cosmo, P. Degtyarenko, A. Dell'Acqua, G. Depaola, D. Dietrich, A. Feliciello, C. Ferguson, H. Fesefeldt, G. Folger, F. Foppiano, S. Garelli, S. Giani, R. Giannitrapani, D. Gibin, J. J. Gómez Cadenas, I. González, G. Gracia Abril, W. Greiner, V. Grichine, A. Grossheim, P. Gumplinger, K. Hashimoto, H. Hasui, A. Heikkinen, A. Howard, V. Ivanchenko, A. Johnson, F. W. Jones, J. Kallenbach, N. Kanaya, M. Kawabata, S. Kelner, P. Kent, T. Kodama, R. Kokoulin, H. Kurashige, T. Lampén, V. Lara, V. Lefebure, F. Lei, M. Liendl, F. Longo, S. Magni, M. Maire, E. Medernach, K. Minamimoto, Y. Morita, R. Nartallo, P. Nieminen, T. Nishimura, K. Ohtsubo, M. Okamura, S. O'Neale, K. Paech, A. Pfeiffer, M. G. Pia, F. Ranjard, A. Rybin, G. Santin, T. Sasaki, N. Savvas, Y. Sawada, S. Scherer, S. Sei, D. Smith, N. Starkov, H. Stoecker, J. Sulkimo, S. Tanaka,

E. Safai Tehrani, M. Tropeano, P. Truscott, H. Uno, P. Urban, A. Walkden, W. Wander, H. Weber, J. P. Wellisch, T. Wenaus, D. C. Williams, D. Wright, H. Yoshida, and D. Zschiesche. Geant4—a simulation toolkit. *Nuclear instruments & methods in physics research. Section A, Accelerators, spectrometers, detectors and associated equipment*, 506(3):250–303, 2003.

[12] S. Jan, G. Santin, D. Strul, S. Staelens, K. Assié, D. Autret, S. Avner, R. Barbier, M. Bardiès, P. M. Bloomfield, D. Brasse, V. Breton, P. Bruyndonckx, I. Buvat, A. F. Chatziioannou, Y. Choi, Y. H. Chung, C. Comtat, D. Donnarieix, L. Ferrer, S. J. Glick, C. J. Groiselle, D. Guez, P-F Honore, S. Kerhoas-Cavata, A. S. Kirov, V. Kohli, M. Koole, M. Krieguer, D J van der Laan, F. Lamare, G. Largeron, C. Lartizien, D. Lazaro, M. C. Maas, L. Maigne, F. Mayet, F. Melot, C. Merheb, E. Pennacchio, J. Perez, U. Pietrzyk, F. R. Rannou, M. Rey, D. R. Schaart, C. R. Schmidtlein, L. Simon, T. Y. Song, J-M Vieira, D. Visvikis, R. V. d. Walle, E. Wieërs, and C. Morel. Gate: a simulation toolkit for pet and spect. *Physics in medicine & biology*, 49(19):4543–4561, 2004.

[13] S. Jan, D. Benoit, E. Becheva, T. Carlier, F. Cassol, P. Descourt, T. Frisson, L. Grevillot, L. Guigues, L. Maigne, C. Morel, Y. Perrot, N. Rehfeld, D. Sarrut, D. R. Schaart, S. Stute, U. Pietrzyk, D. Visvikis, N. Zahra, and I. Buvat. Gate v6: a major enhancement of the gate simulation platform enabling modelling of ct and radiotherapy. *Physics in medicine & biology*, 56(4):881–901, 2011.

[14] David Sarrut, Manuel Bardiès, Nicolas Boussion, Nicolas Freud, Sébastien Jan, Jean-Michel Létang, George Loudos, Lydia Maigne, Sara Marcatili, Thibault Mauxion, Panagiotis Papadimitroulas, Yann Perrot, Uwe Pietrzyk, Charlotte Robert, Dennis R. Schaart, Dimitris Visvikis, and Irène Buvat. A review of the use and potential of the gate monte carlo simulation code for radiation ther-

apy and dosimetry applications. *Medical physics (Lancaster)*, 41(6):064301–n/a, 2014.

[15] Afshine Amidi and Shervine Amidi. *Convolutional Neural Networks Cheatsheet*, (https://stanford.edu/-shervine/teaching/cs-230/cheatsheet-convolutional-neural-network) 2020.

[16] Mathematics in transposed convolution explained. https://medium.com/@kvkumar1993/mathematics-in-transposed-convolution-explained-vignesh-kumar-korakki-bf133c74958. Accessed: 2022-05-19.

[17] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. Pytorch: An imperative style, high-performance deep learning library. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems 32*, pages 8024–8035. Curran Associates, Inc., 2019.

[18] Frank M. Bengel, Bernhard Permanetter, Martin Ungerer, Stephan Nekolla, and Markus Schwaiger. Non-invasive estimation of myocardial efficiency using positron emission tomography and carbon-11 acetate – comparison between the normal and failing human heart. *European Journal of Nuclear Medicine*, 27(3):319–326, 2000.

[19] A. Moretti, A. Gorini, and R. F. Villa. Affective disorders, antidepressant drugs and brain metabolism. *Molecular psychiatry*, 8(9):773–785, 2003.

[20] Michael E. Spieth and Darcy L. Kasner. A tabulated summary of the fdg pet literature. *The Journal of nuclear medicine (1978)*, 43(3):441–441, 2002.

[21] Amir Ghabrial, Daniel Franklin, and Habib Zaidi. A monte carlo simulation study of the impact of novel scintillation crystals on performance characteristics of pet scanners. *Physica medica*, 50:37–45, 2018.

[22] Abdella M. Ahmed, Andrew Chacon, Harley Rutherford, Go Akamatsu, Akram Mohammadi, Fumihiko Nishikido, Hideaki Tashima, Eiji Yoshida, Taiga Yamaya, Daniel R. Franklin, Anatoly Rosenfeld, Susanna Guatelli, and Mitra Safavi-Naeini. A validated geant4 model of a whole-body pet scanner with four-layer doi detectors. *Physics in Medicine & Biology*, 65(23):235051–235051, 2020.

[23] Gunjan Kayal, Maxime Chauvin, Alex Vergara-Gil, Naomi Clayton, Ludovic Ferrer, Tumelo Moalosi, Peter Knoll, Lara Struelens, and Manuel Bardiès. Generation of clinical 177lu spect/ct images based on monte carlo simulation with gate. *Physica medica*, 85(4):24–31, 2021.

[24] Ida Haggstrom, C. R. Schmidtlein, Gabriele Campanella, and Thomas J. Fuchs. Deeppet: A deep encoder–decoder network for directly solving the pet image reconstruction inverse problem. *Medical image analysis*, 54:253–262, 2019.

[25] C. R. Schmidtlein, Assen S. Kirov, Sadek A. Nehmeh, Yusuf E. Erdi, John L. Humm, Howard I. Amols, Luc M. Bidaut, Alex Ganin, Charles W. Stearns, David L. McDaniel, and Klaus A. Hamacher. Validation of gate monte carlo simulations of the ge advance/discovery ls pet scanners. *Medical physics (Lancaster)*, 33(1):198–208, 2006.

[26] F. Lamare, A. Turzo, Y. Bizais, C. C. L. Rest, and D. Visvikis. Validation of a monte carlo simulation of the philips allegro/gemini pet systems using gate. *Physics in medicine & biology*, 51(4):943–962, 2006.

[27] P. Gonias, N. Bertsekas, N. Karakatsanis, G. Saatsakis, A. Gaitanis, D. Nikolopoulos, G. Loudos, L. Papaspyrou, N. Sakellios, X. Tsantilas, A. Daskalakis, P. Liaparinos, K. Nikita, A. Louizi, D. Cavouras, I. Kandarakis, and G. S. Panayiotakis. Validation of a gate model for the simulation of the siemens biograph™ 6 pet scanner. *Nuclear instruments & methods in physics research. Section A, Accelerators, spectrometers, detectors and associated equipment*, 571(1):263–266, 2007.

[28] N. Karakatsanis, N. Sakellios, N. X. Tsantilas, N. Dikaios, C. Tsoumpas, D. Lazaro, G. Loudos, C. R. Schmidtlein, K. Louizi, J. Valais, D. Nikolopoulos, J. Malamitsi, J. Kandarakis, and K. Nikita. Comparative evaluation of two commercial pet scanners, ecat exact hr+ and biograph 2, using gate. *Nuclear instruments & methods in physics research. Section A, Accelerators, spectrometers, detectors and associated equipment*, 569(2):368–372, 2006.

[29] National Electrical Manufacturers Association. *NEMA Standards Publication NU 2–2007, Performance measurements of positron emission tomographs*, 2007.

[30] Siemens Medical Solutions USA, Inc., Hoffman Estates, Illinois. *Biograph TruePoint PET-CT: System Specifications*, 2008.

[31] H. W. de Jong, M. Lubberink, H. Watabe, H. Iida, and A. A. Lammertsma. A method to measure pet scatter fractions for daily quality control: Sf measurement method for daily qc. *Medical physics (Lancaster)*, 36(10):4609–4615, 2009.

[32] L. Le Meunier, F. Mathy, and P. D. Fagret. Validation of a pet monte-carlo simulator with random events and dead time modeling. *IEEE transactions on nuclear science*, 50(5):1462–1468, 2003.

[33] Osama R Mawlawi, David W Jordan, James R Halama, Charles R Schmidtlein, and Wesley W Wooten. Pet/ct acceptance testing and quality assurance: The report of aapm task group 126, 2019, 2019.

[34] Ronald Boellaard, Mike J. O'Doherty, Wolfgang A. Weber, Felix M. Mottaghy, Markus N. Lonsdale, Sigrid G. Stroobants, Wim J. G. Oyen, Joerg Kotzerke, Otto S. Hoekstra, Jan Pruim, Paul K. Marsden, Klaus Tatsch, Corneline J. Hoekstra, Eric P. Visser, Bertjan Arends, Fred J. Verzijlbergen, Josee M. Zijlstra, Emile F. I. Comans, Aiaan A. Lammertsma, Anne M. Paans, Antoon T. Willemsen, Thomas Beyer, Aneas Bockisch, Cornelia Schaefer-Prokop, Dominique Delbeke, Richard P. Baum, Arturo Chiti, and Bernd J. Krause. Fdg pet and pet/ct: Eanm procedure guidelines for tumour pet imaging: version 1.0. *European journal of nuclear medicine and molecular imaging*, 37(1):181–200, 2010.

[35] Anes Kaalep, Terez Sera, Wim Oyen, Bernd J. Krause, Arturo Chiti, Yan Liu, and Ronald Boellaard. Eanm/earl fdg-pet/ct accreditation - summary results from the first 200 accredited imaging systems. *European journal of nuclear medicine and molecular imaging*, 45(3):412–422, 2018.

[36] RL Wahl. *Principles and Practice of PET and PET/CT*. Lippincott Williams & Wilkins, 2nd ed. edition, 2008.

[37] Medhat M. Osman, Christian Cohade, Yuji Nakamoto, and Richard L. Wahl. Respiratory motion artifacts on PET emission images obtained using CT attenuation correction on PET-CT. *European Journal of Nuclear Medicine and Molecular Imaging*, 30(4):603–606, January 2003.

[38] Willem Grootjans, Rick Hermsen, Erik H.F.M.van der Heijden, Olga C.J. Schuurbiers-Siebers, Eric P. Visser, Wim J.G. Oyen, and Lioe-Fee de Geus-Oei. The impact of respiratory gated positron emission tomography on clinical staging and management of patients with lung cancer. *Lung Cancer*, 90(2):217–223, November 2015.

[39] C. S. van der Vos, W. Grootjans, A. P. W. Meeuwis, C. H. Slump, W. J. G. Oyen, L.-F. F. de Geus-Oei, and E. P. Visser. Comparison of a free-breathing CT and an expiratory breath-hold CT with regard to spatial alignment of amplitude-based respiratory-gated PET and CT images. *Journal of Nuclear Medicine Technology*, 42(4):269–273, October 2014.

[40] Charlotte S van der Vos, Willem Grootjans, Dustin R Osborne, Antoi P W Meeuwis, James J Hamill, Shelley Acuff, Lioe-Fee de Geus-Oei, and Eric P Visser. Improving the spatial alignment in PET/CT using amplitude-based respiration-gated PET and respiration-triggered CT. *J. Nucl. Med.*, 56(12):1817–1822, December 2015.

[41] D Didierlaurent, S Ribes, H Batatia, C Jaudet, L O Dierickx, S Zerdoud, S Brillouet, O Caselles, and F Courbon. The retrospective binning method improves the consistency of phase binning in respiratory-gated PET/CT. *Phys. Med. Biol.*, 57(23):7829–7841, December 2012.

[42] Sadek A. Nehmeh, Yusuf E. Erdi, Clifton C. Ling, Kenneth E. Rosenzweig, Heiko Schoder, Steve M. Larson, Homer A. Macapinlac, Olivia D. Squire, and John L. Humm. Effect of respiratory gating on quantifying pet images of lung cancer. *The Journal of nuclear medicine (1978)*, 43(7):876–881, 2002.

[43] Sadek A. Nehmeh, Yusuf E. Erdi, Kenneth E. Rosenzweig, Heiko Schoder, Steve M. Larson, Olivia D. Squire, and John L. Humm. Reduction of respi-

ratory motion artifacts in pet imaging of lung cancer by respiratory correlated dynamic pet: Methodology and comparison with respiratory gated pet. *The Journal of nuclear medicine (1978)*, 44(10):1644–1648, 2003.

[44] Yusuf E. Erdi, Sadek A. Nehmeh, Tinsu Pan, Alexander Pevsner, Kenneth E. Rosenzweig, Gikas Mageras, Ellen D. Yorke, Heiko Schoder, Wendy Hsiao, Olivia D. Squire, Phil Vernon, Jonathan B. Ashman, Hassan Mostafavi, Steven M. Larson, and John L. Humm. The ct motion quantitation of lung lesions and its impact on pet-measured suvs. *The Journal of nuclear medicine (1978)*, 45(8):1287–1292, 2004.

[45] Paul J Schleyer, Michael J O'Doherty, Sally F Barrington, and Paul K Marsden. Retrospective data-driven respiratory gating for PET/CT. *Physics in Medicine and Biology*, 54(7):1935–1950, mar 2009.

[46] Florian Büther, Iris Ernst, Mohammad Dawood, Peter Kraxner, Michael Schäfers, Otmar Schober, and Klaus P. Schäfers. Detection of respiratory tumour motion using intrinsic list mode-driven gating in positron emission tomography. *European journal of nuclear medicine and molecular imaging*, 37(12):2315–2327, 2010.

[47] Adam L. Kesner and Claudia Kuntner. 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 (Lancaster)*, 37(10):5550–5559, 2010.

[48] Glafkos Havariyoun, Eleni Kalogianni, Danielle Levart, Benjamin Corcoran, N. Mulholland, and Gill Vivian. Q.freeze for respiratory motion correction in pet/ct: Assessment and optimisation with a 4d pet/ct phantom and patient

data. In *Annual Congress of the European Association of Nuclear Medicine 2016*, 10 2016.

[49] Axel Martinez-Möller, Michael Souvatzoglou, Nassir Navab, Markus Schwaiger, and Stephan G. Nekolla. Artifacts from misaligned ct in cardiac perfusion pet/ct studies: Frequency, effects, and potential solutions. *Journal of Nuclear Medicine*, 48(2):188–193, 2007.

[50] M Dawood, N Lang, Xiaoyi Jiang, and K.P Schafers. Lung motion correction on respiratory gated 3-d pet/ct images. *IEEE transactions on medical imaging*, 25(4):476–485, 2006.

[51] M. Dawood, M. Fieseler, F. Büther, X. Jiang, and K. P. Schäfers. A multi-resolution optical flow based approach to respiratory motion correction in 3d pet/ct images. In *Medical Biometrics*, Lecture Notes in Computer Science, pages 314–322. Springer Berlin Heidelberg, Berlin, Heidelberg.

[52] Philipp Fischer, Alexey Dosovitskiy, Eddy Ilg, Philip Häusser, Caner Hazirbas, Vladimir Golkov, Patrick van der Smagt, Daniel Cremers, and Thomas Brox. Flownet: Learning optical flow with convolutional networks. *CoRR*, abs/1504.06852, 2015.

[53] Zachary Teed and Jia Deng. *RAFT: Recurrent All-Pairs Field Transforms for Optical Flow*, pages 402–419. Computer Vision – ECCV 2020. Springer International Publishing, Cham, 2020.

[54] Jason J. Yu, Adam W. Harley, and Konstantinos G. Derpanis. Back to basics: Unsupervised learning of optical flow via brightness constancy and motion smoothness. In *Computer Vision - ECCV 2016 Workshops, Part 3*, 2016.

[55] W. P. Segars, G. Sturgeon, S. Mendonca, Jason Grimes, and B. M. W. Tsui. 4d xcat phantom for multimodality imaging research. *Medical physics (Lancaster)*, 37(9):4902–4915, 2010.

[56] M.D Olsen, Jeffrey R., Ph.D Lu, Wei, B.S Hubenschmidt, James P., B.S Nystrom, Michelle M., M.S Klahr, Paul, M.D Bradley, Jeffrey D., Ph.D Low, Daniel A., and M.D Parikh, Parag J. Effect of novel amplitude/phase binning algorithm on commercial four-dimensional computed tomography quality. *International journal of radiation oncology, biology, physics*, 70:243–252, 2008.

[57] A. F. Abdelnour, S. A. Nehmeh, T. Pan, J. L. Humm, P. Vernon, H. Schöder, K. E. Rosenzweig, G. S. Mageras, E. Yorke, S. M. Larson, and Y. E. Erdi. Phase and amplitude binning for 4d-ct imaging. *Physics in medicine & biology*, 52(12):3515–3529, 2007.

[58] Mohammad Dawood, Florian Büther, Norbert Lang, Otmar Schober, and Klaus P. Schäfers. Respiratory gating in positron emission tomography: A quantitative comparison of different gating schemes: Respiratory gating schemes in pet: A quantitative comparison. *Medical physics (Lancaster)*, 34(7):3067–3076, 2007.

[59] T. O'Briain, C. Uribe, I. Sechopoulos, C. Michel, and M. Bazalova-Carter. In depth modelling and validation of a clinical pet system monte carlo simulation. *Medical Physics, revised*, 2022.

[60] Stefano Savian, Mehdi Elahi, and Tammam Tillo. *Optical Flow Estimation with Deep Learning, a Survey on Recent Advances*, pages 257–287. Deep Biometrics. Springer International Publishing, Cham, 2020.

[61] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, June 2016.

[62] Deqing Sun, Stefan Roth, and Michael J. Black. A quantitative analysis of current practices in optical flow estimation and the principles behind them. *International journal of computer vision*, 106(2):115–137, 2013;2014;.

[63] Simon Meister, Junhwa Hur, and Stefan Roth. Unflow: Unsupervised learning of optical flow with a bidirectional census loss. *CoRR*, abs/1711.07837, 2017.

[64] Diederik Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *International Conference on Learning Representations*, 12 2014.