Lesion Synthesis Toolbox: Development and validation of dedicated software for synthesis of realistic lesions in raw PET and CT patient images

by

Hanif Gabrani-Juma, B.Eng

A thesis submitted to the Faculty of Graduate and Postdoctoral Affairs in partial fulfillment of the requirements for the degree of

Masters of Applied Science

in

Biomedical Engineering

Carleton University
Ottawa, Ontario

Abstract

PET is one of the most sensitive clinical tools for early detections of small tumours, but its performance is dependent on multiple factors including image quality and human perception. Current methods relying on physical and numerical image quality phantoms are inadequate for evaluating clinical task-based performance, such as limits of lesion detection, due to lack of physiologic realism of the images.

In this work, we describe development and validation of the Lesion Synthesis Toolbox, an easy-to-use software to synthetize well-characterized, user-defined lesions in real patient PET data prior to image reconstruction and in corresponding CT images on GE Discovery line of PET/CT scanners.

We employ synthetic lesions in a preliminary perception study to estimate the limits-of-detection (LOD) in representative clinical setting. Finally, we describe a custom developed, web-based tool for conducting perception studies using synthetic lesions of varying parameters to characterize the LOD. We conclude with a discussion of how these tools can be employed in future studies to compare LOD between imaging methods, image generation algorithms, observers, display types, and computer-based lesion detection solutions.

Acknowledgements

I wish to extend my gratitude to Ran Klein and Andy Adler for guiding me through this project. This work would not have been possible without the amazing people that have supported me throughout the journey.

Table of Contents

List of Tables

List of Illustrations

List of Abbreviations

2D – two dimensional

3D – Three dimensional

AI – Artificial Intelligence

BSREM – Block sequential regularized expectation maximization

CD – Contrast detail

CMV – Center most voxel

CT – Computed Tomography

DICOM – Digital Imaging and Communications in Medicine

FBP- Filtered back projection

^{18}F-FDG – Fluorine-18 labelled fluorodeoxyglucose

FTP – File transfer protocol

FOV – Field of view

FWHM – full width half maximum

^{68}Ge – 68Gallium

GE – General Electric

GUI – Graphic user interface

HU – Hounsfield units

I/O – Input/output

LYSO - Lutetium-yttrium oxyorthosilicate

LOD – Limits of detection

LOR – line of response

PACS – Picture archive and communication systems

PET – Positron Emission Tomography

PHI – Patient health information

PSF – Point spread function

MAFC – Multiple alternative forced choice

MIP – Maximum intensity projection

MLEM – Maximum Likelihood Expectation Maximization

MRI – Magnetic resonance imaging

MRN – Medical record number

OHSN-REB – Ottawa Health Science Network Research Ethics Board

OSEM – Ordered subset expectation maximization

^{82}Rb – Rubidium

ROI – Region of interest

ROC – Receiver operating characteristic

REB – Research and ethics board

TOF – Time of flight

WB – Whole-Body

Chapter 1: Introduction

Positron emission tomography (PET) is a sensitive 3D medical imaging modality used for a broad spectrum of medical and research applications. It quantifies physiologic function by measuring the concentrations of radioactive-labelled biological compounds within the scanners field of view (FOV). A biological compound that is used by the body (e.g. glucose) is labeled with a positron-emitting isotope, creating a tracer (e.g. ^{18}F-fluorodeoxyglucose). The tracer is administered to the patient, distributed through the body and is accumulated by target tissues. As the isotope undergoes radioactive decay the resulting emission can be detected by a PET camera. The PET system utilizes various image reconstruction algorithms on the detected decay events to produce tomographic volume images quantifying the regional tracer concentrations within the body, corresponding to the biologic process being imaged.

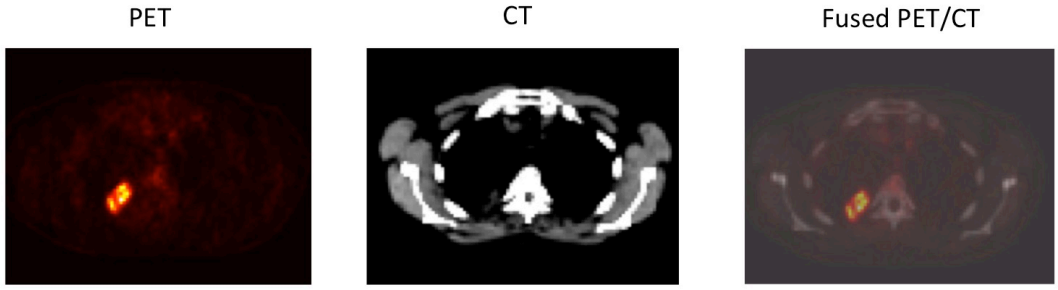

Figure 1: Cross-section of example patient ^{18}F-FDG PET scan (left) with low-dose CT (center) and their fused image (right) enabling anatomical localization of PET lesions.

Due to its functional imaging capabilities, PET has found many applications as a clinical diagnostic and research tool. Within the domain of oncology, hybrid PET and x-ray computed tomography (CT) systems are routinely used for the diagnosis, staging,

treatment planning and monitoring of cancer. In the context of PET/CT systems, a low-dose (i.e. low tube current to reduce radiation dose to patient) CT is used for physics-based corrections of the PET image, as an anatomical image for determining the exact location of PET lesions within the body and for incremental diagnostic information. Example PET and CT images are shown in Figure 1.

A well-known limitation of PET is its relatively low spatial resolution, which can underrepresent small or faint lesions. Newer, more sophisticated PET image reconstruction techniques aim to suppress noise and improve the detectability of smaller and more faint lesions, potentially increasing the limits of human perception for lesion detection.

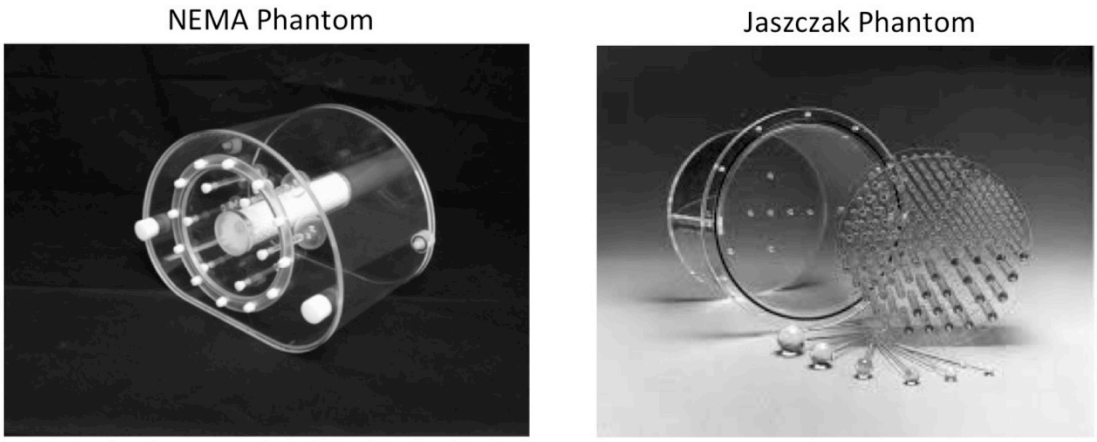

Figure 2: Two commonly used image quality phantoms used to evaluate the spatial resolution and uptake profiles of a PET scanner: The NEMA 2012 Hot-Spheres Phantom (Left), and The Jaszczak Image Quality Phantom (Right).

At present, medical imaging scientists primarily rely on physical and numerical phantom experiments and on subjective comparisons of image quality in patient images to characterize advancements in image reconstruction techniques [1], but these evaluation

techniques have important limitations. Physical phantoms (shown in Figure 2) are simplistic in design and are typically imaged in near-perfect imaging scenarios (e.g. in the absence of patient motion from breathing) and therefore do not emulate clinical images in all their complexity (i.e. non-homogenous, asymmetric structures, variability in structure size, orientation, shape). Patient images, conversely, rarely have reliable ground truths associated with them [2], [3], against which image derived metrics can be compared. Deriving ground truth information from the images themselves is conceptually flawed, especially for faint structures around the spatial resolution of the system, but comparison between alternative image generating techniques may be appropriate as demonstrated in Figure 3, where changes in reconstruction parameters lead to varying lesion conspicuousness. Nevertheless, absolute ground truth such as presence/absence of a lesion and properties of lesions (e.g. size and accumulated tracer concentrations) are not readily measurable from the images alone, especially when the lesion parameters are at the margins of system performance. Consequently, the true underlying characteristics, such as the true physiologic uptake, of real lesions cannot be known [4].

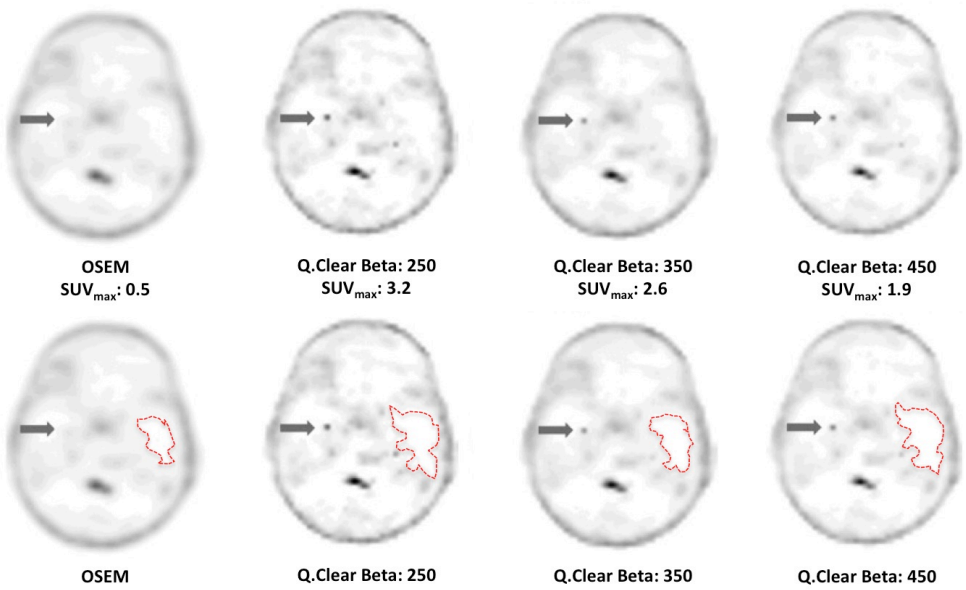

Figure 3: Example of how the true underlying activity in patient images is ultimately unknown. Four reconstructions of the same measured emission data applying four different image generation techniques. The first column uses a traditional OSEM 2 iteration 32-subset reconstruction. Images in the 3 columns to the right are generated using a BSREM reconstruction with varying beta values. The top row demonstrates the BSREM reconstruction showing lesions not visible in OSEM. The bottom row shows the delineation (dashed red) of a structure within the image. Note the variance in lesion uptake and structure delineation with image generation techniques. Image adapted from [5].

In this work we sought to address these limitations by creating realistic, well characterized user-defined lesions within actual patient data and to demonstrate how these synthetic data may be used in future perception studies to evaluate detectibly of lesions across alternative image generation techniques, image display, and observer methodologies. Synthetic lesion and patient data must be combined at early phases in the image generation process, so that down-stream processes (i.e. the image reconstruction algorithm used) may be interrogated for their effect on lesion perception. An example application utilizes a dataset of synthetic lesions inserted in raw PET data prior to image

reconstruction to compare the effect of different image reconstruction parameters on the ability of radiologists to perceive small and faint lesions within a patient.

Therefore, the primary objectives of this research were to: (1) develop methods to embed synthetic lesions with known physical characteristics into raw PET projection data prior to image reconstruction, and (2) develop methods to embed synthetic lesions in corresponding patient CT data. Following validation of the lesion synthesis methods, secondary objectives were (3) to develop a library of PET/CT patient images with synthetic lesions using clinically available image reconstruction techniques, and (4) demonstrate the utility of synthetic lesions through a perception study to estimate the limits of perception for the task specific goal of lesion detection. In this work we aspire to develop an objective benchmark tool with which to test image reconstruction technologies, human observers and even artificially intelligent machine observers.

Elements of this work have been accepted as an oral for presentation at:

- Medical Imaging Perception Society Annual Meeting 2019
    - Salt Lake City, Utah, July 14 – 17, 2019

Chapter 2: Background

2.1 PET as a medical imaging tool

Positron emission tomography (PET) is the imaging modality of choice by physicians for non-invasive assessment of cancerous diseases [6]. Because tracers can be developed to specifically target a physiologic process of interest, minute quantities of tracer accumulations on the order of nano- and even pico-moles can be detected [7]. Hundreds of tracers have been developed for applications including oncology, cardiology and neurology [7]–[9]. The applications of new radiotracers remain an active field of research and development. Modern PET systems are hybrid systems incorporating x-ray computed tomography (CT) or magnetic resonance imaging (MRI) that are used for performing physical corrections of measured PET images (i.e. attenuation correction) and as complimentary anatomical imaging.

PET generates images by measuring the radiopharmaceutical (radiotracer) concentration within the scanners field of view. PET is routinely used as the standard of practice given the sensitivity to detect early disease, specific biomarker targeting and the ability to represent the physiologic function (functional imaging) of the anatomy being imaged.

2.2 Physics of PET

PET radiotracers utilize unstable isotopes that decay by positron emission (β^+) into a more stable form. This phenomenon can be described in the equation one below, where X and Y represent the parent and daughter isotopes respectively, A and Z represents the atomic mass and number of the parent isotope (X), and ν_e represents the electron neutrino released along with β^+ during decay.

$$ {}^{A}_{Z}X \rightarrow {}^{A}_{Z-1}Y + \beta^{+} + \nu_{e} \qquad 1 $$

A positron is a positively charged electron and is an example of antimatter. The emitted positrons travel through the surrounding medium, losing energy with inelastic collisions with surrounding electrons until the positrons kinetic energy is reduced to a point where it combines with an electron and subsequently they annihilate. The annihilation produces two co-linear gamma rays (511 keV photons) propagating in opposite directions (co-linear). While most photons will be absorbed (attenuated) by the surrounding media (i.e. patient's body), some photon pairs traverse the neighbouring tissues and exit the body where they can be detected by specially designed, high-density PET detectors.

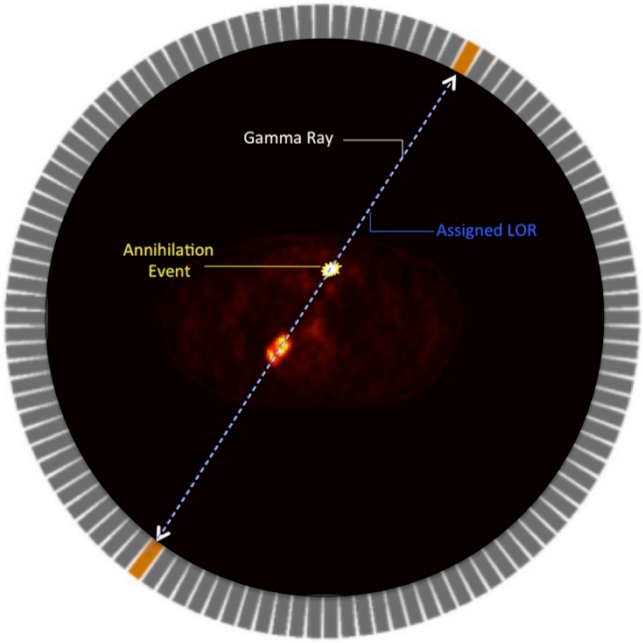

Figure 4: Illustration of annihilation event producing two-colinear photons that penetrate the patient and are detected by the detectors of the PET camera, resulting in a true coincidence event.

To detect photon pairs a typical 3D PET scanner is equipped with several rings each comprised of several scintillation crystal detectors. Photons that are absorbed by the

crystal cause the crystal to shine (scintillate). Scintillation detectors convert the energy of a gamma ray into visible light. The intensity of light generated by the scintillation detector is directly proportional to the energy of the gamma ray detected. The crystals are traditionally coupled to photomultiplier tubes or more recently to solid-state detectors to generate electrical signals that are registered by a computer to record a detected 511 keV photon event. Advanced PET systems with updated hardware can record the time of each detected event down to sub-nanosecond precision. Coincidence events are extracted by comparing detection times within a coincidence-timing windows (typically < 10 nanoseconds) dependant on crystal detector and electronics characteristics [10]–[12].

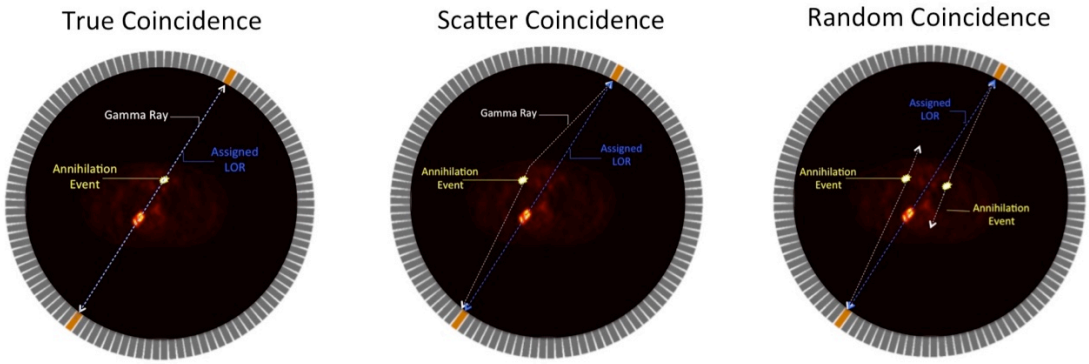

Figure 5: Types of detectable coincidence event Include: true coincidence event (left), scatter coincidence event (middle) and random coincidence event (right)

Pairs of photons that are detected within a narrow coincidence-timing window are considered as coincidence event that originated along the line of response (LOR) between the two detectors. If the photons originated along the LOR the detection event is referred to as a true event. If one or both photons scatter en route the detector, the resulting LOR may not align with the origin of the radiation, and these events are referred to as scatter events. Photon pairs from two separate decay events simultaneously activating two

detectors within the coincidence-timing window are termed random events [11], [12]. Scatter and random events degrade image quality and therefore are explicitly estimated and corrected for during image reconstruction.

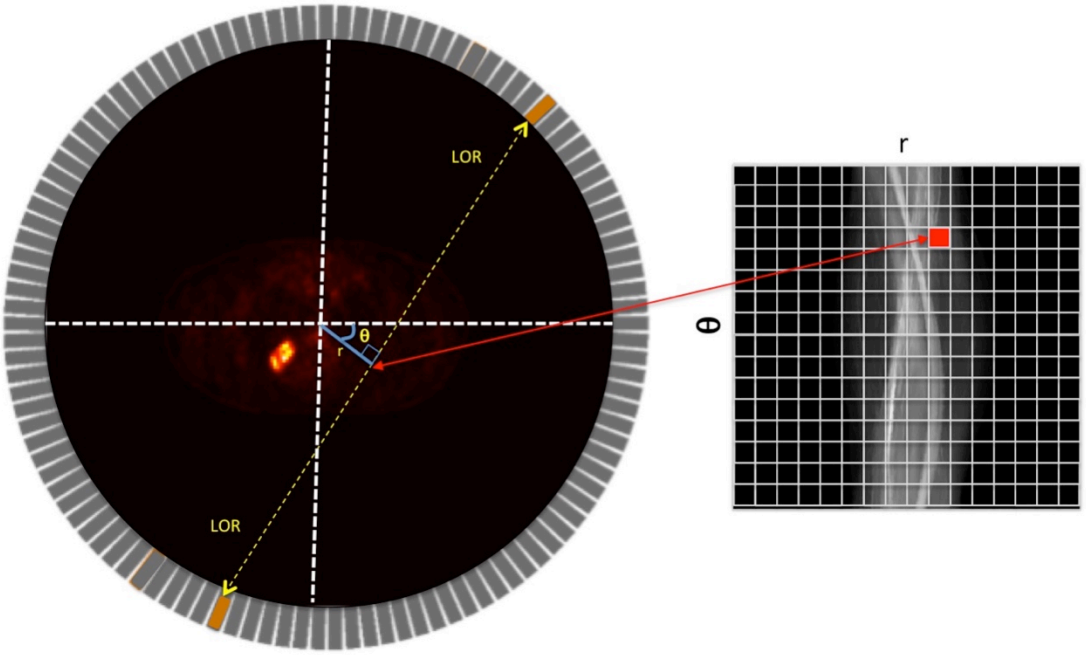

Figure 6: Illustration of the relationship between line of response (LOR) and the corresponding elements in the sinogram where r and θ represents the distance and angel of LOR from the origin.

Each LOR is assumed to correspond to the location of an annihilation event despite the possibility of the coincidence event resulting from scatter and/or random events. Detected events may be recorded as LORs listed one-by-one in a long file, which is referred to as list-mode data acquisition. The sum of all recorded events along each LOR can be binned into histograms that represent each LOR by its unique detector pair combinations. Histograms can be reformatted to more convenient representations of LOR angular and radial positions within the scanners FOV, referred to as sinograms or projections. Most image reconstruction algorithms generate 3D images from sinograms which have first

been corrected for physical and instrument effects [11]. Hybrid systems such as PET/CT and PET/MR offer the advantage of producing co-registered (aligned) anatomical data that can be conveniently used for the attenuation and scatter data correction of sinograms.

2.3 Imaging process

Patient preparation prior to imaging includes instructions to follow during imaging, and administration of applicable medications/radiotracers. In cases with short-lived tracers (e.g. ^{82}Rb, Half-life = 76 seconds), the patient is positioned prior to administration of the tracer. More typically (e.g. ^{18}F, Half-life = 110 minutes) the tracer is administered to the patient before (e.g. 60 minutes) the patient is positioned on the scanner bed for imaging. Once the patient is ready for imaging, they will be positioned on the scanner bed, typically in the supine position with arms above the head to reduce attenuation and scatter to organs in the thorax and abdomen. Using a scout scan (e.g. 2D x-ray or fast MR sequence), the technologist will configure the desired region(s) of the body to image (e.g. eyes to mid-thighs). After applying necessary positional adjustments, both a PET and anatomical imaging (3D CT or MRI) are acquired. The anatomical imaging is subsequently used to derive 511 keV photon attenuation images [11]. Depending on imaging parameters and type of exam, a typical PET scan lasts several minutes, and up to an hour.

2.4 PET image Types

Emission data (detected coincidences) are typically recorded in list-mode format, which can then be replayed in various formats. Within the context of clinical oncology, PET data is primarily acquired using a static image acquisition protocol in which the average

distribution over the total scan time is used to generate a single volumetric image. Occasionally respiratory-gating signals (acquired from external respiratory tracking hardware) are utilized to generate cine images at several respiratory phases such that a movie of respiratory motion can be visualized. Other applications may use cardiac-gating where cardiac motion can be visualized. Dynamic imaging, a movie sequence over time since tracer administration, is also increasingly used to visualize changes in tracer distribution over time and to subsequently quantify the rates of physiologic function using specialized image analysis software [13].

2.4.1 Static Whole-Body Acquisition

Static imaging is the most basic image acquisition mode. During a static acquisition all the true coincidence events detected are used to reconstruct a single image. A voxel within a static PET image represents the average tissue radiotracer uptake or activity concentration over the duration of the scan. The work described in this thesis focuses on static image acquisition protocols. Static imaging protocols can be used in a step and shoot fashion, where the bed with the patient is gradually moved further into the scanner, to image patient regions exceeding the scanners physical axial coverage. Static images are typically on the order of 2 to 10 minutes per bed position, where images of each bed positions can be stitched together.

2.5 Emission Data Correction

Prior to image reconstruction, the acquired emission data must be corrected for physical effects and measurement errors. Detector efficiency, dead time, random, scatter,

attenuation and isotope decay corrections are all applied in clinical routine prior to image reconstruction, and are detailed below:

2.5.1 Detector Efficiency Normalization

Normalization correction accounts for the effective variations in individual detector sensitivities along each LOR. The variability in detector sensitivity, gaps where detection cannot occur, and varying angles of incidents of each LOR with the detector (parallax effect) create non-uniformities within the acquired data. The detector normalization algorithm uses a calibration file that is stored on the scanner and is updated routinely to account for changes in detector efficiency over the life of the machine.

The detector normalization calibration file is generated using a cylindrical phantom experiment where a uniform radiation source is placed within the FOV. To mitigate dead-time effects, the source is made to have relatively low activity concentrations, however to ensure adequate image quality and count statistics long acquisition times (>12 hrs) are used. Typically, long-lived isotopes (e.g. ^{18}F or ^{68}Ge) are used. During this process each detector is exposed to the same uniform source and the number of counts detected by each coincidence pair of detectors is recorded. The measured counts are scaled to the average of all detectors to generate normalization factors for each individual detector. An efficiency normalization map specific to the scanners geometry can then be used to correct detector efficiency variations [14].

2.5.2 Dead-Time Correction

Detector dead-time is a phenomenon that refers to the time the system needs to process a detected event before another event can be detected and is due to the finite amount of

time required for the system to respond to a photon detection event. A dead-time loss occurs when a photon passes through a detector while the system is still processing a previously detected event. This limitation of detector technology hinders the PET system from accurately recording event-rates, resulting in a non-linear fall-off relationship between the activity within the scanners FOV and the detected count rate.

Two dead-time loss models may apply to PET; paralyzable and non-paralyzable models [14]. In the paralyzable model, when multiple events are detected before the initial event has been processed, all events are discarded. In the non-paralyzable model, when multiple events are detected before the initial event has been processed, the initial event continues to be processed and events detected during the processing period are discarded. When using a paralyzable model the observed count rates will approach zero as the true count rate (activity within the FOV) increases. In a non-paralyzable approach the count rate for a given system plateaus for which scaling factors can be used to correct dead-time losses [14]. In practice, dead-time is usually a combination of paralyzable and non-paralyzable effects and is measured empirically to derive a correction function.

2.5.3 Random Coincidence Correction

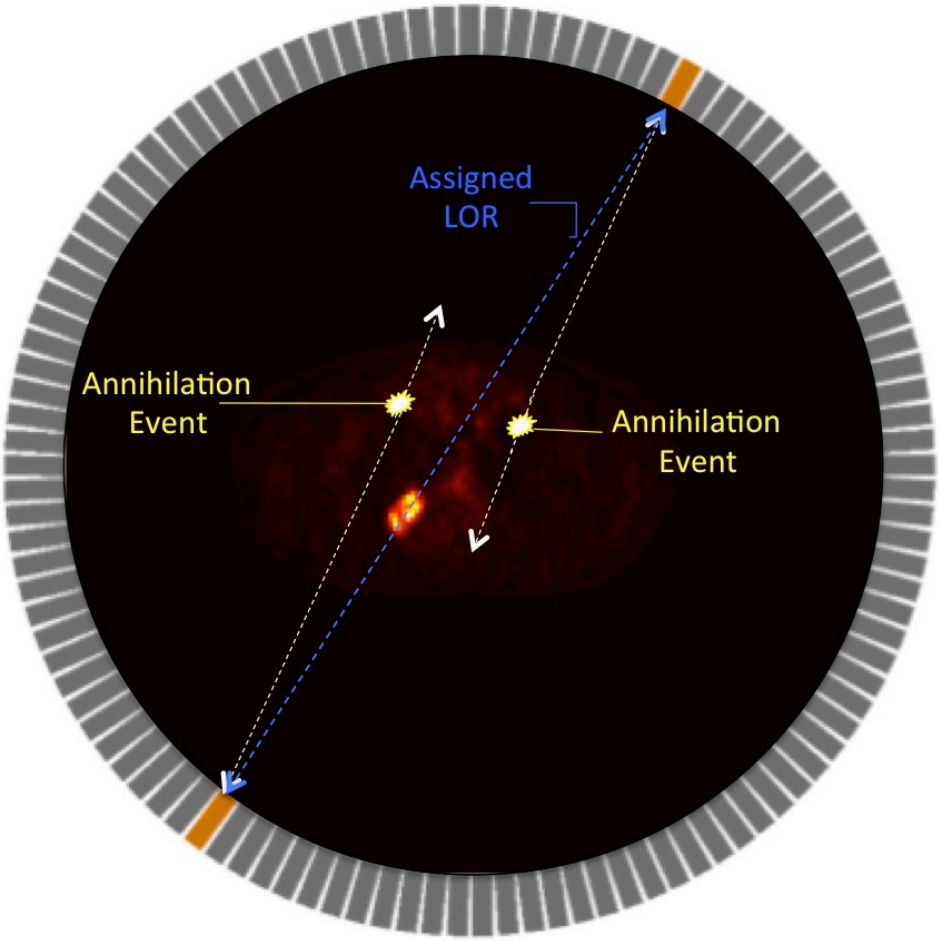

Figure 7: Illustration of Random Coincidence Event. Two annihilation events (yellow), gamma ray paths (white) and the incorrect assigned LOR (blue).

Random coincidence correction is required when photon pairs from two separate decay events simultaneously activate two detectors within the coincidence-timing window. These events, if not removed from the measured data, generate uniform background activity within the image that distorts the true activity distribution within the image and diminishes the quantitative accuracy of image measurements. There are two methods typically employed to estimate random coincidence rates:

Using a delayed window approach, a second coincidence window is used after each detected event. The second window is longer than the initial coincidence-timing window and occurs slightly delayed to the initial window. Coincidences recorded during this delayed-window are random by definition, as the time delay entails two separate decay events.

Estimated random coincidence rates can also be calculated from the count rates of each LOR detector pair using the following equation:

$$N_R = 2\tau A^2 \qquad 2$$

The variable N_R represents the expected number of random events for the LOR with A recorded counts and coincidence acceptance time window of length τ [14]. This equation is a probability of two separate decay events being recorded within the allotted coincidence-timing window.

Both techniques determine the number of random coincidences for each set of detector pairs. The random events are subtracted from the total number of events detected [14]. As clearly demonstrated by the equation above, random coincidences increase in quadrature with activity in the FOV, and typically represent 20-50% of the detected events in routine clinical studies.

2.5.4 Scatter Correction

Scatter events are a result of a photon path change due to interaction(s) with the medium that it passes through. If not corrected for, scatter events will propagate into the reconstructed volume increasing background counts, decreasing contrast and increasing image noise. In modern PET systems, scatter photons are partially filtered out using

energy discrimination as they are detected, and any remaining scatter is estimated and subtracted from the projection data prior to image reconstruction.

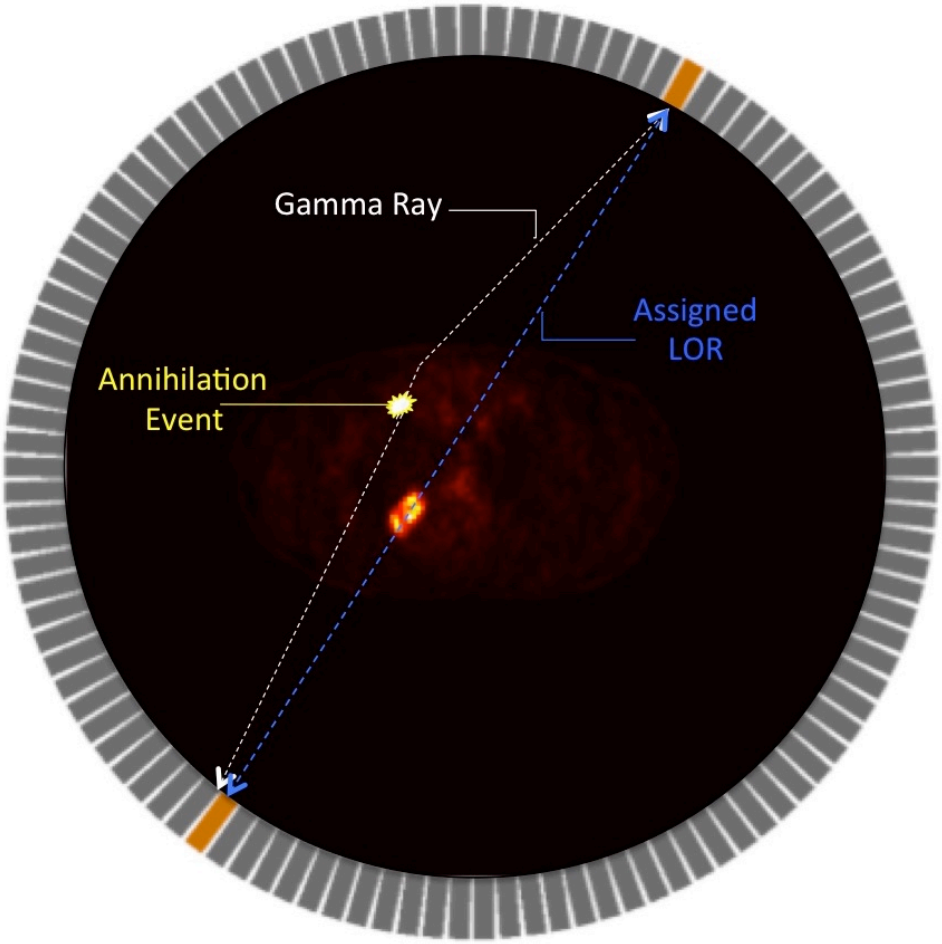

Figure 8: Illustration of Scatter Coincidence Event. Annihilation event (yellow), True gamma ray path (white) and the incorrect assigned LOR (blue).

Since all photons have an initial energy of 511 keV, scattered photons can be filtered during acquisition if their recorded energy is below a predetermined threshold energy level. However due to the limited energy resolution of scintillation detectors, this method cannot discard all scatter events.

While several approaches to scatter correction exist, most commonly, scatter is directly estimated using the number of emitted photons and information from the attenuation map (generated from CT or MR) using the using the Klein-Nishina method to model Compton scatter. A sinogram of the estimates scattered counts detected by the PET system is subtracted from the detected counts [14]. Scatter counts are proportional to the activity in the FOV and is dependent on the activity distribution and the composition of the scattering medium (i.e. the patient anatomy).

2.5.5 Attenuation Correction

Prior to detection, photons travel through neighbouring tissues where the energy of the photon can be absorbed, decreasing the measured count rate. This phenomenon is referred to as attenuation. Attenuation translates into an underestimation of activity in the reconstructed volume, particularly in deep structures. The likelihood that a pair of photons will exit the medium (body) and strike a pair of detectors at distance D apart is given by:

$$P_{Co} = P_a P_b = e^{\int_0^d \mu(x)dx} e^{\int_d^D \mu(x)dx} = e^{\int_0^D \mu(x)dx} \qquad 3$$

Where d and D-d represent the distances that each annihilation photon must travel from point of origin, through the medium, and to their corresponding detectors, and $\mu(x)$ represents the 511 keV linear attenuation coefficients (e.g. cm^{-1} units) of the medium at point x along the LOR. An important property that is revealed by the above equation, and that is exploited for attenuation correction, is that the total attenuation along a LOR is not dependent on the position of the origin of the photon along the LOR, d. Nevertheless,

total attenuation along the LOR must be measured accurately and used to correct acquired data.

Transmission scans are acquired using the CT in hybrid PET/CT systems and the x-ray CT image attenuation values are scaled to estimate 511 keV photon linear attenuation factors, at the voxel level. With MR this process is less straightforward, but in general involves classifying regions of the MR image into a small set of tissue types (e.g. air, soft tissue and bone) for which characteristic 511 keV photon linear attenuation values are predetermined. Attenuation factors are calculated along each LOR by integrating the linear attenuation values along the LOR. Attenuation effects in the emission data are corrected by dividing the detected counts by their LOR attenuation factor [14].

2.6 Image Reconstruction

Following normalization, dead-time, random coincidence, scatter and attenuation correction of the projection data a sinogram is produced for reconstruction of a 3D image volume. The reconstructed volume is typically corrected for isotope decay such that image voxels represent tracer concentration in units of Bq/cc at time of scan start.

Image reconstruction is an inverse problem, where the measured image is calculated from photons observed by the detectors in the PET system. While the PET system measures annihilation photons for a given LOR, we desire an image representing tracer distribution with precise spatial coordinates within the measured medium. Image reconstruction produces 3D tomographic slices of the measured medium. Typically image reconstruction algorithm are categorized as either analytical (i.e. filtered back projection) or iterative methods (e.g. ordered subset expectation maximization algorithm) [15] .

2.6.1 Filtered back projection

In this analytical technique, an image matrix is first defined for all valid LOR. This can be achieved by drawing imaginary lines between detector pairs through each intersecting voxel in the image matrix. For each valid LOR, the number of detected events by the detector pair is added to each voxel that is intersected by the LOR. During this process, the number of detected events measured by detector pairs is projected back along the line from which they had originated, resulting in a superposition of back-projections.

This simple approach results in a fast estimation of the activity distribution of the measured medium. The voxels in the image volume are typically blurry representation of the true underlying activity distribution. The blurring effect is attributed to the equal distribution of detected counts along the estimated line of origin and is proportional to the distance from the source activity. The blurring of the image can be corrected by applying a deconvolution filter in the Fourier domain. A ramp filter is typically used to amplify high spatial frequencies within the data. To mitigate amplification of noise, a low pass filter is also typically employed. The process is repeated for all projections within the sinogram. An inverse Fourier transform is applied to return the image spatial coordinates and the resulting image, in theory, should be an exact analytical solution of the inverse problem.

While this approach is widely praised for its low computational cost, it is not appropriate for all data type conditions. FBP is sensitive to noise and has therefore been shown to fail in cases where there is wide variability in activity concentration within the FOV, or a low count-density image is acquired. Additionally, this method is prone to image artifacts [14], [16]. Finally, FBP is limited in its ability to model non-linear and spatially variant

elements of the imaging process accurately resulting in suboptimal image resolution, contrast and quantitative accuracy. Consequently, FBP has mostly fallen out of favour for more advanced iterative image reconstruction techniques.

2.6.2 Maximum Likelihood Expectation Maximization

Alternative methods to FBP have been sought to minimize image artifacts and improve image quality by including more complete models of the imaging physics and imaging system. These methods mostly employ iterative reconstruction algorithms based on maximum-likelihood expectation maximization (MLEM). The iterative algorithm starts with a blank (or uniform) image matrix as an initial estimate, which is forward projected and compared to measured projection data. The residual between the estimate projection data and the measured projection data is used to adjust the image estimate. The adjustments are formulated to monotonically reduce the error between the projection data with each iteration. The forward-projection of the estimated image and adjustments are repeated until the residual between the estimated image and the measured data is sufficiently small or a predetermined number of iterations are reached.

When the residual between the measured and estimated projection data is minimized, the statistical likelihood is maximized, and the expectation maximization algorithm iteratively maximizes this likelihood under a Poisson data model. The MLEM algorithm accounts for statistical noise within the projection data by assuming a Poisson distribution (representative of counting statistics). Because Poisson noise is inversely related to the number of counts detected in each projection, accounting for this noise model improves resulting images for low-count density acquisitions.

By incorporating the system geometry and detector efficiency, into the forward projector the detection probability of a gamma ray for a specific LOR can be introduced, further improving the accuracy of the reconstructed image. Almost any aspects of the system can be added to the forward projector to improve modeling accuracy of the physics and instrumentation, including, most notably, the system point-spread function.

The term iterative reconstruction comes from multiple projection and back projection steps needed to achieve higher local accuracy compared to FBP. This technique is less prone to producing streaking artifacts and also can reduce activity spill over. The iterative nature of this method results in longer reconstruction times due to the need for more computations. Since noise is amplified with the number of iterations applied, a combination of stopping criteria, regularization methods, or filtering is applied to limit excessive noise in the reconstructed volume.

2.6.3 Ordered subset expectation maximization

The ordered subset expectation maximization (OSEM) algorithm is an accelerated variant of MLEM that is employed clinically to manage image reconstruction times [17]. Compared to MLEM, the two techniques are implemented in a similar fashion employing iterative projection and back-projection steps, but only a fraction of the angular data is used on each step. The OSEM algorithm exploits redundancies between neighbouring projections. It divides the projection data into evenly separated subsets to reduce the computation time required by a factor approximately equal to the number of subsets initialized [17].

Similar effects to MLEM have been observed with the OSEM algorithm regarding the relationship between the amplification of noise and the number of iterations applied, and

therefore a combination of stopping criteria, regularization methods, or filtering are applied to limit excessive noise in the reconstructed volume. As with most modern PET installations OSEM is applied clinically within the Division of Nuclear Medicine at The Ottawa Hospital, and is therefore primarily used in this work.

2.6.4 Block Sequential Regularized Expectation Maximization

A possible incremental improvement over OSEM is the inclusion of *a priori* constraints on the image such as smoothing, noise level or anatomical information. A macro-parameter balances between maximizing expectation with MLEM and maximizing a priori constraints. One such algorithm is the block sequential regularized expectation maximization (BSREM) that applies a Bayesian penalized-likelihood constraint on the reconstructed image to minimize noise locally, allowing the algorithm to reach full image convergence without excessively amplifying noise in the image [1]. The scientific community has interest in this algorithm as it claims to improve contrast recovery and limit the noise which in-turn improve the detection of small lesions where otherwise noise and lack of contrast may hinder detection[1], [18].

The BSREM algorithm controls noise at higher iterations by applying a relative difference penalty within a localized region of the previous iteration when comparing the estimated projection (subset) to the measured projection data. The BSREM algorithm uses a 2 iteration 24 subset OSEM reconstruction as the initial estimate and refines subsequent iterations using a noise penalization controlled by the term denoted as β ranging from 0 (OSEM alone) and 1000 [18].

Clinically the BSREM algorithm is offered by General Electric Health Care (GE) marketed as Q.Clear. As of January 2019, the division of Nuclear Medicine at The

Ottawa Hospital does not apply the BSREM algorithm for clinical interpretation, but are evaluating it, as a possible alternative to OSEM is certain applications.

2.7 Image quality measures as a motivation for synthetic lesions

When it comes to images, and specifically medical images, there are many measures of image quality to compare alternative image generation/image acquisition techniques. Common metrics include noise measurements, spatial resolution, target to background ratios, etc. A complex interplay between these image quality metrics, lesion properties and observer performance exists which ultimately influences the limits of detectable lesions. Hence, in conventional practice utilizing phantom experiments and patient data it is unclear how to optimize image quality for optimal lesion detection. An alternative approach is to characterize lesion detection performance under varying conditions through a series of lesion detection experiments.

To characterize lesion detection, observer lesion detection performance must be measured relative to known lesion properties. These metrics can be used to characterize the limits of lesion detection in absolute terms, thus enabling comparison between competing PET image reconstruction technologies, image display technologies, image rendering methods and observers of these image, be them human or machine. Lesion detection characterization depends on a reliable reference truth of the presence and characteristics of test lesions at levels that challenge the limits of perception. Obtaining reference truth from reconstructed clinical images is unreliable since the detection of lesions is limited by the technologies used to reconstruct, render and display the image, and by the observer, especially in small and faint lesion which are challenging to detect.

Furthermore, due to performance limitations of the above factors and of the imaging system, characterization of the lesions (e.g. size and intensity) from the reconstructed images can be inaccurate. To overcome these challenges, we sought to create clinically representative lesions, with user specified characteristics (i.e. size, shape, location, intensity) in raw PET patient data prior to image reconstruction. We term this approach "lesion synthesis".

2.8 Introduction to Lesion Synthesis

In the context of this work, lesions refer to abnormal growth of tumour cells in an organ or region of the body. In PET, hot lesions refer to focal spots in the with abnormally elevated image intensity (tracer activity) relative to surrounding normal tissues. Lesion synthesis refers to the insertion of an artificial lesion with known physical characteristics into real medical image data. Because synthetic lesions can be controlled, they can be useful to evaluate medical imaging tools and image-processing techniques, or as an educational resource to train readers [19].

Different approaches to lesion synthesis have been demonstrated at almost every step of the imaging process including subject preparation, PET acquisition, image reconstruction and post-reconstruction stages as illustrated in Figure 9. These approaches range from invasively embedding physical sources within a biological specimen (black) to retrospectively painting fake digital lesions on reconstructed images (purple). The synthetic lesion is intended to serve as a reference truth by which the performance of an image analysis related task can be evaluated. In the context of this work we are primarily concerned with the task of lesion detection and characterizing performance by

determining the limits-of-detection (LOD) as a function of lesion (e.g. size, activity) and image (e.g. activity, noise) properties. Because LOD are objective and quantitative metrics of task-based performance, they can serve as an objective and quantitative measure by which image generating technologies can be compared and benchmarked.

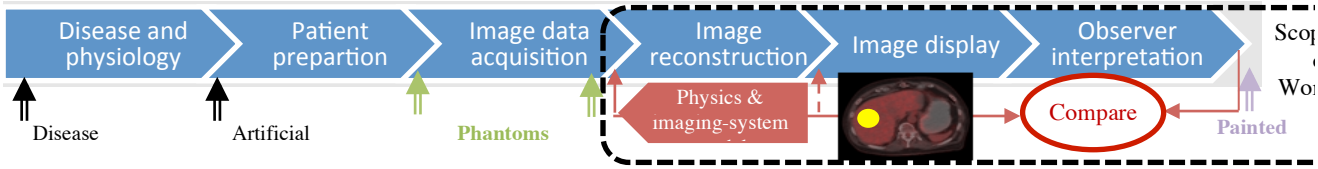

Figure 9: Steps in medical imaging for lesion detection and opportunities for inserting known lesions ranging from diseases models and implanted sources (black arrows), physical and numerical phantom experiments (green arrows), synthetic lesion in patient data (red arrows) and painted lesions on reconstructed images (purple).

An absence of high-quality reference truth to probe the LOD exists because there are no clinical tests that exceed the sensitivity of PET/CT for imaging early cancers; hence surrogate experiments have been resorted to. Lesions have been introduced into living animal subjects using implantable radioactive sources [20] or malignant cell lines[21] (black arrows in Figure 9), but cannot be ethically used in human studies and may not produce realistic data due to their invasive nature. Furthermore, animal experiments, are expensive, tedious and require a high level of expertise which limits their widespread use. Phantom experiments (physical [19], [20] or numerical [24], green in Figure 1) are advantageous as the reference truth can be explicitly designed to probe specific lesion parameters, but they are often criticized for their simplicity which does not accurately represent clinical data in all its complexity (e.g. complex biological structures, artifacts). Furthermore, physical phantoms often lack variability, producing predictable images that

are ill-suited for perception studies[19], [25]. Libraries of human labeled patient images (e.g. expert panels) (purple in Figure 9), are advantageous for their clinical realism, but can be tedious to obtain and they are not useful for optimizing steps leading up to image generation[26]. Furthermore, the reference truth is limited by the accuracy of the viewers and becomes exceedingly unreliable as lesions become more difficult to perceive, as they approach the LOD [27]. Surrogate markers such as other imaging modalities, histopathology, or outcomes are also of limited value for optimizing lesion detection with PET because they are less sensitive than PET for tumor detection, they do not correlate sufficiently with PET, and they may accurately portray lesion properties.

The method of choice to embed artificial lesions into existing medical image data is ultimately determined by the goal of the research study or the level of complexity needed to accurately represent the lesions. The end goal in lesion synthesis is to embed a synthetic lesion with known characteristics such as size, shape, location and intensity with adequate blending or system modeling to achieve realism and to avoid obvious artifacts.

Digital lesion synthesis is a compromise that inserts artificial lesion data into real patient data (red workflow in Figure 9). Lesion synthesis has been successfully applied for characterizing lesion perception tasks in various imaging modalities[25], including few PET studies[19], [22], [28], [29].

2.8.1 Voxel Based Synthetic Image Generation

The most obvious approach for lesion synthesis is in image voxel space where lesion values replace or augment patient voxel values. The voxel based synthetic image is

generated using an intensity distribution map specified by the user. Voxels in the intensity distribution map correspond to desired intensity values representing the simulated object in the same units as the patient image (i.e. HU for CT, or Bq/cc for PET). Typically blending, smoothing, and image texture modelling is applied to the lesion image to mimic image quality characteristics seen in real images [19], and to avoid obvious stitching artifacts. This technique may be applied to limited image regions to add features (i.e. structures or lesions) with user-defined attributes. Image reconstruction is not required for voxel-based lesion synthetic and therefore this technique cannot be used to investigate image reconstruction effects on lesion perception. Furthermore, because the patient image properties are estimated and not absolutely known it is not always possible to accurately represent lesion qualities such as lesion to background ratio with respect to the true uptake of the background (i.e. organ) opposed to the estimated background in the resulting image. Blending of tumour [30] images and patient images is attractive due to its simplicity. The biggest drawback of using a blending technique is that it is a post-reconstruction method and therefore cannot test the impact of alternative image reconstruction methods on detectability. The second drawback is that accurate modeling of texture and noise in image space may not be trivial, producing unrealistic lesions, stitching artifacts, which can adversely impact realism.

2.8.2 Lesion synthesis in raw PET data

A more promising approach simulates individual lesions using the patient specific material maps and models the imaging physics and the scanner system to generate raw detector counts (projection data) for the lesion alone, which are then, combined with the raw patient data before image reconstruction. If the same reconstruction method as for the

original patient data is used, the patient image (i.e. almost identical texture and noise) can be reproduced with the addition of well-characterized lesions. Lesions geometries are defined in image space and are then projected into sinogram, histogram or LOR space by simulating the imaging physics and instrumentation processes. Simulations vary in complexity and level of detail, but as a minimum require a lesion activity distribution map (source of photons) and a material map of the FOV to simulate photon interaction with the medium and photon detection. Noise models are also commonly added. The resulting data from the simulation is an estimate of the expected projection data that would be recorded by a physical system had only the lesion activity been present in the patient.

2.8.2.1 Simulation techniques

The commercial development of PET technologies has benefited greatly from information gained from simulation studies. Simulations to generate synthetic datasets (e.g. patient models) with user-defined characteristics (i.e. lesions and organ properties) have enabled to investigate physical effects (e.g. scatter) that are often difficult to observe empirically [4], [24]. Simulations with varying scanner designs have been used to optimize design considerations such as scintillator crystal size and geometry. While sophisticated simulations enable in depth investigation of a specific phenomenon, the associated computational cost can be extensive. Generally, as the simulation environment increases in complexity, greater computational power is needed. The degree of complexity to which a simulation model is defined is in part dependant on the application; therefore, simplified simulation tools that are often favoured, especially when many simulations are required and/or when computer resources are limited. In the

context of nuclear medicine simulation tools can be generally classified as either particle tracking or analytical simulations [24], [31], [32].

2.8.2.1.1 Photon Tracking

Predicting the response of complex systems using deterministic models is often impossible or inaccurate because they may not relate the range of possible outcomes for a system with many variables and probabilities. Monte Carlo simulations have been used for a wide range of applications since the 1940s to predict behaviour of complex systems including particle tracking, weather models, and gambling. Named after the famous casino, Stanislaw Olam realized that he could predict the likelihood of winning a game of solitaire by simulating many individual games and summarizing the distribution of observed outcomes to generate a probability of winning [33]. In the context of nuclear imaging, the inputs are distribution of activity and materials in the FOV, and the outcomes are the number of detected photons in each detector [34], [35].

Particle tracking simulations are the most precise simulation techniques and have been applied to a variety of high-energy physics applications [4]. Using Monte Carlo simulation techniques individual photon tracks are simulated as photon interactions are modeled as random probabilities of occurrence along their path. With detailed interaction models and large number of photons simulated, precise simulations of the imaging systems can be achieved. A common drawback of particle tracking simulations can be attributed to the large computational cost and technical complexity in setting up the simulations [4], [34].

The phenomena modelled by photon tracking simulation packages, often rely on complex random processes and that cannot be estimated using simple mathematics (i.e. filter).

General-purpose Monte Carlo photon tracking simulation packages have been tailored specifically towards nuclear imaging applications, and offer sufficient accuracy for modeling scanner geometries, characterizing data contaminants and characterizing effects such as scatter [4].

Typically, a Monte Carlo based photon tracking PET simulators comprise of a random number generator, predetermined probability distributions for particles interacting with predefined materials and methods to sample the probability distributions [32]. The user specifies the geometry of the PET system characterizing the orientation and composition of materials (i.e. detectors) [36]. Subject data volumes to be simulated are defined by the user specifying the activity and object material as either voxel-based matrices or geometric shapes. The simulation initializes random emissions based on the activity distribution map and then tracks individual particles by sampling a predefined probability distribution as they interact with surrounding materials until each particle has been either absorbed or detected, or it has exited the FOV. Simulations require billions of photons to be tracked, resulting in lengthy simulation times [4], [37].

2.8.2.1.2 Analytical Simulation

Contrary to tracking an individual photon through materials and its resulting interactions, an analytical simulator offers the ability to mathematically model various phenomena (including scanner geometry, physical effects, and detector efficiency) through mathematical models of the probability of where photons originating at a specific location are likely to arrive [4], [24]. Analytical simulations model the average probability of photon interactions without needing to simulate many millions of individual photons, and are therefore much more computationally efficient, albeit

potentially at the expense of simulation accuracy. Analytical simulation packages offer the advantage of generating multiple realizations of the data with relatively low computational cost.

The complexity of a typical analytical simulator is dependant on desired phenomenon to investigate. Like particle-tracking simulators, analytical simulators comprise of a scanner detector geometry model, a user defined activity map (3D volume) describing the object to simulate, and a materials map defining the patient and imaging hardware configuration [4], [24]. Using the scanner detector geometry model, valid LOR corresponding to voxels from the activity distribution image are sampled. This technique is commonly referred to as forward projection when mean estimations of sinogram projections are generated without modeling noise and the realism of physical effects may be limited.

Noise is often added to the noiseless sinogram facilitating the generation of multiple realizations via bootstrapping [38]. Physical effects that can be analytically modeled can also be incorporated into the analytical simulation. From the user defined attenuation map (3D volume) describing the density of materials that comprise the simulated object, attenuation for each LOR can be calculated and applied so that the simulation considers the effects of attenuation on the generated sinogram.

Similarly, other phenomena such as detector efficiency can be applied by considering a detector efficiency factor for each corresponding detector, the inverse of the detector efficiency correction applied prior to image reconstruction as described in the Data Correction section of this work.

2.8.3 Perceptual Studies

Throughout the history of medical imaging, the perception and image quality of medical images has remained a critical focus of the medical imaging community [18], [25], [39], [40]. Readers of medical images are concerned with the effect that image generation algorithms, image-viewing conditions, or image rendering (e.g. scaling, color mapping) has on the performance of completing the task at hand [19], [25]. These and other factors can influence diagnostic accuracy that impact patient care and therefore may also concern medical-legal labiality. A previous study clearly demonstrated that 20% of (^{18}F-FDG avid) lesions imaged were not perceivable in fused FDG PET/CT images when compared to standard interpretation of the PET and CT images side-by-side [19]. Other sources of perceptual bias must also be taken into consideration when optimizing a clinical workflow for task specific goals (e.g. effects of image reconstruction parameters on lesion detection).

Perceptual studies in medical imaging began in the late 1940s with x-ray imaging using physical phantoms [41]. In one of these early studies, a phantom consisting of a 2-demnsional array of disks with varying levels of attenuation was imaged and observers were asked to identify the lowest contrast disk they could discern. This was repeated for disks of varying diameters, to characterize the limits of detection, deriving contrast-detail (CD) diagrams [42]. This simple technique posed some limitations; (1) a fixed test pattern enabled the observer to know exactly what to look for and where, producing a recognition or memory bias, (2) this technique could not quantify the probability that the observer was correct, and (3) the CD patterns did not represent clinical images and therefore their applicability to clinical tasks was not clear. With these limitations in mind, researchers have explored alternatives to CD diagrams employing studies using images

with known lesions (real or synthetic) in which observers are tasked with reporting whether a lesion is present of absent. Performance metrics such as sensitivity, specificity, and true/false positive/negative rates can be computed by comparison of observer response to the reference truth.

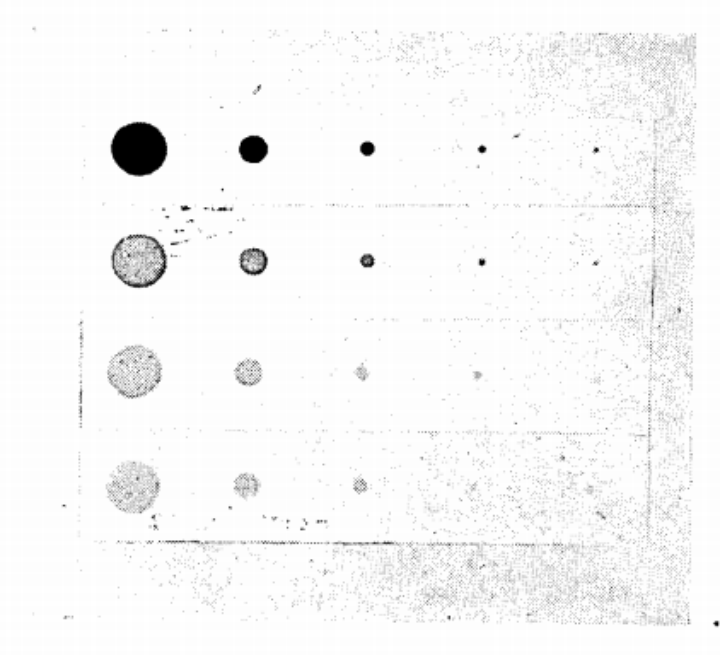

Figure 10: X-ray image of a CD Diagram test pattern used in the 1948 study to evaluate the limit of detectability for spheres with varying attenuation contrast and size [42].

It is well understood however that perception accuracy is dependent on multiple image variables including target-to-background contrast, target size and image noise, and that a single set of performance metrics is not sufficient to summarize performance. The receiver operating characteristic (ROC)[43] curve plots the true-positive rate as a function of the false-positive rate as a metric of interest (e.g. target size) varies over a predefined range (ideally spanning from unperceivable to obvious). A common application utilizing ROC curves, presents a series of images to the observer for them to

determine the presence/absence of a lesion and their degree of confidence on an ordinal scale (e.g. scale of 0 to 10). The ROC curve is generated by plotting the true- and false-positive rates at each reported level of confidence. Alternatively, ROC curves can be produced as a function of target size, contrast, or any other controlled variable.

In the experiment, study participants may be asked to determine the location of the target within a set of images. If the reported location is accurate (within a predefined distance of the reference truth) the decision is scored as correct and false otherwise. Limitations using this method include (1) the decisional variability between observers is high as the metric of evaluation is subjective and can vary between observers and (2) perception experiments can take considerable time, especially when lesions are not obvious, resulting in observer fatigue which may influence observer performance and limit the sample size.

An alternative experimental paradigm is the multiple-alternative forced choice (MAFC)[39] experiment which presents the observer with M number of images at any given time and the observer must select the single image in which they perceive the target signal (i.e. lesion). Forced choice is favourable as evaluation time can be a factor of 10 less than the same study implemented as a lesion search analysis [39], enabling a much larger sample-size in the same session time. Because of the larger sample size, the probability of the correct response can be determined without the need for the observer to report their confidence. An ROC curve can likewise be produced using MAFC for any controlled variable [43]. A commonly sited limitation of MAFC is that the operator can guess between multiple alternative images without clearly validating where in the image space they perceive the lesion, and therefore MAFC has not been used widely.

2.8.4 Task-based optimization

Accurate and realistic lesion synthesis offers the opportunity to evaluate lesion perception in an objective and quantitative manner. Using metrics of limits of detections researchers can optimize all aspects of PET imaging from image reconstruction to image interpretation for the specific goal of detecting subtle lesions reliably. Limits of detection (LOD) with respect to lesion size, lesion-to-background contrast and image noise characterize the smallest and faintest lesion that can be perceived by the observer. LOD can be used to guide various aspects of medical imaging including:

- Determining the LOD of human viewers.
- Determining the LOD of artificial intelligent (AI) viewers.
- Objectively comparing between human and AI viewers.
- Optimization of image display and human viewer conditioning (e.g. fatigue, lighting) [25]
- Optimization of image reconstruction and data correction techniques.

2.8.5 Thesis Overview

While previous work on lesion synthesis in raw PET data has been demonstrated in the literature, tools for use by non-experts are not available to the broader research community. The primary objective of this work was to develop and validate a complete software solution enabling lesion synthesis with known physical characteristics (i.e. size, shape, intensity, and orientation) into ^{18}F-FDG patient scans prior to image reconstruction and into the corresponding CT images. These tools are intended to be easy to use including turn key deployment for work with compatible PET/CT systems. To achieve

realism, state of the art methods, compatible with clinical systems and reconstruction techniques were explored.

Using these techniques, a preliminary database of patient PET/CT images with labelled synthetic lesions was generated. The library of synthetic lesions was used to investigate the LOD of study participants as a baseline estimate of expected values in subsequent studies. To accomplish this, a preliminary perception study was deployed within our lab in Division of Nuclear Medicine.

Application of these techniques will enable researchers to investigate aspects of lesion detection that require precise gold truth. These methods can be used investigate optimization of clinical workstations for the task-specific goal of lesion detection or training an artificial observer.

Chapter 3 discusses the describe implementation and evaluation of photon tracking and analytical simulators for the synthesis of lesions. The chapter discusses advantages and disadvantages of each technique and further describes subsequent image processing workflow to generate a patient image with a synthetic lesion. Monte Carlo techniques employed a model of the clinical PET/CT system (GE Discovery 710), used to generate projection planes of synthetic lesions. An analytical simulation alternative is then described, validated and compared. Methods to combined synthetic projection data with real patient projection data prior to image reconstruction are described, as is the subsequent image reconstruction workflow. Numerical phantoms and their corresponding validation scans were analyzed to validate the entire lesion synthesis workflow. Additionally, voxel-based method to embed and blend synthetic lesions into corresponding patient CT data are described.

Chapter 4 describes the design of the perception study deployed within the Division of Nuclear Medicine at The Ottawa Hospital. Lesion synthesis techniques were applied to generate a library of patient PET/CT images with labelled synthesis lesions. The developed dataset was used to determine the smallest and most faint lesions that human observers could discern to determine the LOD of study participants. User defined characteristics used to develop the lesion library are also described in this chapter. It is hypothesized that as lesion size and contrast decreases lesion detection rates will decrease with relatively small advantage for experienced observers.

Chapter 5 summarizes the work to date and discusses its potential impact on future research and applications.

Chapter 3: The Lesion Synthesis Toolbox

3.1 Introduction

This chapter covers development of a fully integrated lesion synthesis toolbox that can be used by non-expert users to produce synthetic lesions in actual patient data. The chapter begins describing lesion synthesis in PET using two alternative methods: (1) photon tracking lesion simulation and (2) analytical lesion simulation. The chapter then goes on to compare between these two methods. Lesion synthesis in CT is also discussed. The chapter concludes with a complete description of the workflow and how the toolbox can be applied for various research applications.

3.2 Patient Data

In this work, clinical patient data was acquired using a GE Discovery 710 series hybrid PET/CT system at The Ottawa Hospital, Regional Cancer Centre. Image acquisition followed typical image acquisition practices with image acquisition being performed approximately 60 minutes after FDG administration (5 MBq/kg, up to 444 MBq). After positioning the patient on the scanner bed an x-ray scout scan (120 kVp, 10 mA) was performed, and the scan range was designated as required for the clinical application. A helical CT scan followed immediately (120 kVp, auto mA, 0.9 second rotation time, 0.984:1 pitch) and reconstructed on the console using full FOV, non-iterative reconstruction. With the patient remaining in the same position on the bed the PET scan followed.

Whole-body (WB) scans were acquired using a step-and-shoot mode, where the bed position within the FOV was moved axially every 2.5 minutes to cover the desired patient

range (typically 8 bed positions for an "eyes-to-thighs" scan) despite the scanner having only a 15.4 cm axial field of view. Bed positions were overlapped by 3.6 cm to compensate for lower sensitivity at extreme axial slices of the PET system.

For each patient scan the following data files were collected for further processing:

- Reconstructed CT Scan - for attenuation correction of the PET data during reconstruction, to define lesion location and to simulate photon interactions.
- PET Histogram File - the raw emission data measured by the PET system and used to reconstruct patient activity.
- PET Geometric Correction File - used to maintain uniform pixel size across transverse planes and to correct for detector efficiency.
- PET Normalization File - used to correct data considering detector efficiency.
- PET calibration File - used to calibrate PET images into absolute units [Bq/cc].
- Reconstructed Image File – used to define lesion location and sample background activity. (This file is optional, as it can also be reconstructed offline from the files listed above using the image reconstruction research toolbox.)

Images were reconstructed using a vendor provided image reconstruction research toolbox (REGRECON5, General Electric Healthcare, Waukesha, WI) for off-line image processing, which is numerically equivalent to the clinical reconstruction software on the PET console. Off-line image reconstruction is essential for routine use as it eliminates the need to continuously move lesion appended raw emission files onto the clinical system console for image reconstruction, which complicates workflow and may interfere with the clinical workflows. The reconstruction toolbox ran on Matlab R2017a under either

Linux or Windows environments. Input from our development team to the vendor was used to improve compatibility, correct errors and improve documentation of the Toolbox for broader use by the research community through a collaborative research agreement. Patient data was acquired under an Ottawa Hospital Research Ethics Board approved study (REB#20150509) and no explicit patient consent was required.

3.3 The Lesion Synthesis Workflow

The Lesion Synthesis Toolbox consists of several sequential steps that greatly automate the lesion synthesis process. The workflow is illustrated in Figure 11, starting with patient data retrieval from the scanner console computer, definition of the lesion in the patient image space, simulating the lesion emission data, combining it with the respective patient data, reconstructing the patient images with the synthetic lesion and completing in recording the synthetic lesion properties as a reference truth.

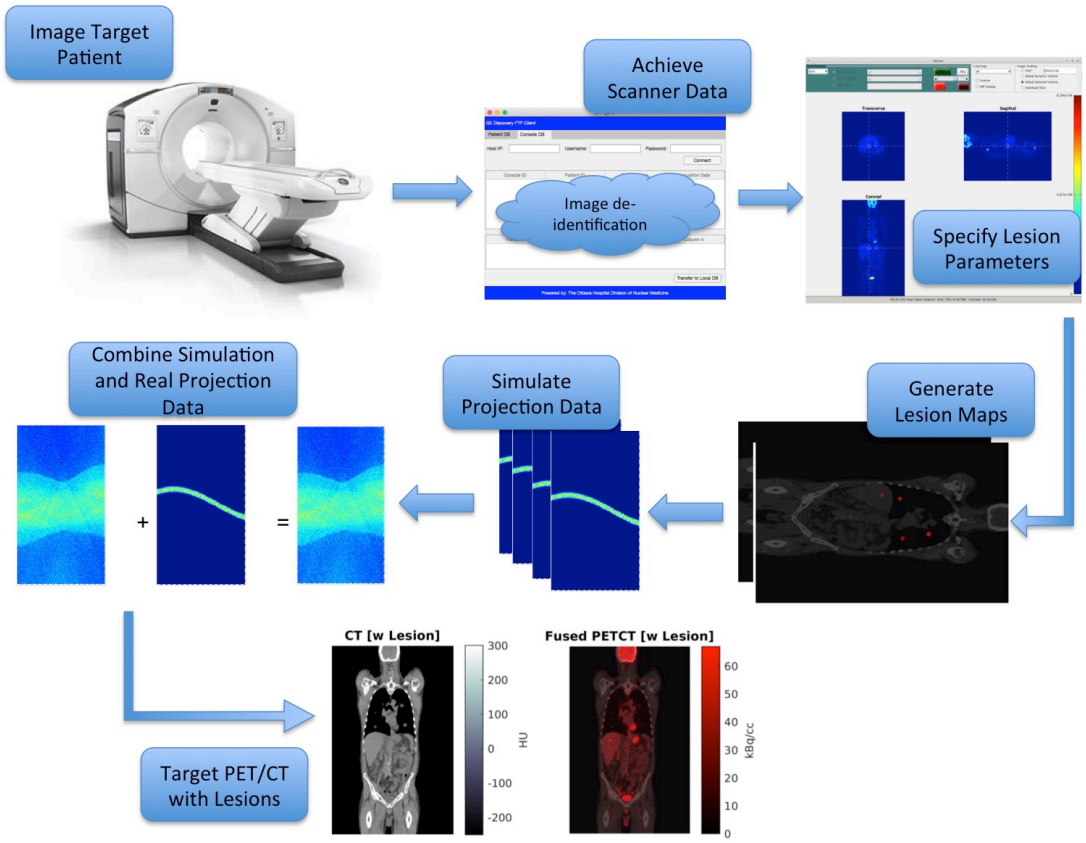

Figure 11: Lesion synthesis workflow comprising the Lesion Synthesis Toolbox.

3.3.1 GE Discovery Data Retriever (DiscoveryDR)

Patient data resides temporarily (several days) on the scanner console computer in a proprietary database and file structure. Because the scanner is heavily used clinically, research must not interfere with the clinical workflow or system serviceability. Therefore, we sought to develop a tailored remote file transfer application to extract patient data from the scanner and feed it to the lesion synthesis toolbox.

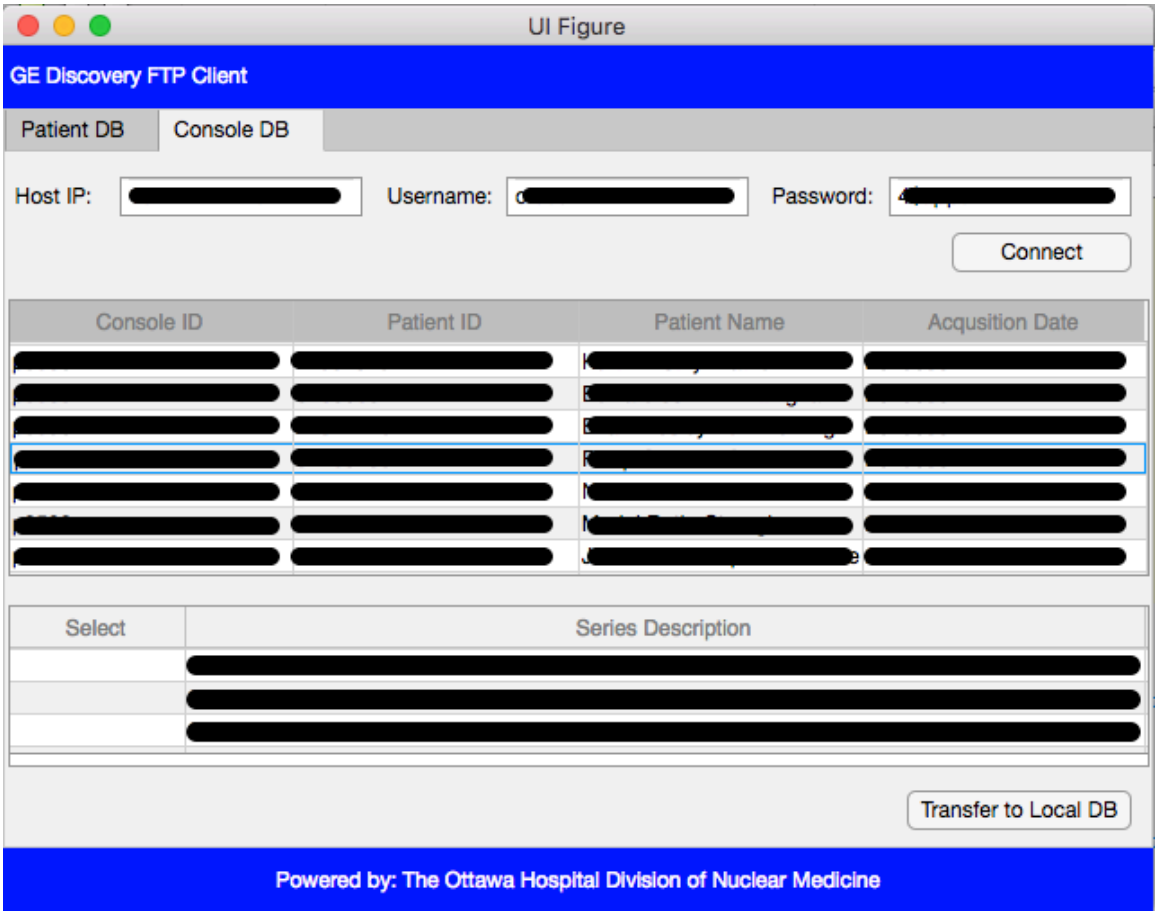

Figure 12: Screenshot of the DiscoveryDR Client compatible with all GE Discovery 600/690/710 Series Scanners.

The DiscoveryDR client GUI was developed using the MATLAB App Designer. The user specifies scanner credentials (i.e. Scanner IP address, Username, Password) while on The Ottawa Hospital Network. Using a MATLAB implementation of a file-transfer protocol (FTP), the client is able to recursively interrogate the file system. Sinogram files for each patient are stored in a master head directory. Folders and files within the head directory are labelled with random letters and numbers to protect patient privacy. Within the head directory, each folder represents a unique patient study. Within a study folder, several folders can reside each containing data from separate scans (i.e. scout scan, CT,

PET) associated to the specific patient and study. For each scan directory in the study directory, a single file in DICOM format is transferred and read by the DiscoveryDR client to identify the study, patient and scan name for which it belongs to. The entire study directory is queried to populate a list of available studies and scans organized by patient name or ID. The time required to query the entire master head directory ranges from 2-5 minutes depending on network traffic and the number of studies available on the console.

Once the list of available studies has been populated, the user can select the study (or individual files) for export to the local machine. File transfer is performed using FTP and files are stored on a local directory in a patient-study directory hierarchy.

After each study is successfully transferred, the files are de-identified (or anonymized) to remove all patient identifiers. A master list linking de-identified studies to patient medical record numbers (MRN) is updated.

The DiscoveryDR was developed to either allow the user to retrieve individual studies through a user interface for browsing available data on the console shown in Figure 12, or by enabling the user to perform batch file extraction where patient specific identifiers (such as MRN) are provided in table format in a Microsoft Excel file.

3.3.2 Image data de-identification

To protect patient identity and to mitigate memory bias in perception research participants, DICOM image data was de-identified by removing all Patient Health Information (PHI) including names, medical record number, addresses, telephone numbers and insurance numbers. The developed DiscoveryDR client performs de-identification immediately after data retrieval from the scanner console.

3.3.3 Lesion definition

Lesion images are defined as activity distribution maps in the image space. Using a custom interactive volumetric viewer, target lesion locations are defined on the patient PET/CT data (Figure 13). For each a set of lesion sizes, lesion to background contrast, and CT HU are specified. Multiple values can be specified to generate the lesion multiple times with varying parameter.

An open-source volumetric viewer (View4D) was used for the user to define target lesion location within the patient image space. Lesion centres were returned as Cartesian coordinates and along with lesion intensity, lesion size, and lesion shape parameters were used to generate activity distribution maps [44]. In this thesis work the viewer was used as follows: the user pressed the s-key on the keyboard to indicate a spherical lesion at the viewer crosshairs with 10.0 mm diameter. The viewer display is updated with a contour of the lesion at the specified location. Using the '<' and '>' keys, the user can increase or decrease the lesion size by 0.5 mm increments. More advanced lesion shapes such as random blobs were also implemented; but we elected to use simple spheres to simplify interpretation of these validation results.

A binary lesion mask image with the same dimensions as the reconstructed target PET image is generated indicating voxels belonging to the lesion. The patient reconstructed PET image is sampled using the lesion mask to measure the background activity intensity, which together with the desired lesion contrast ratio determines the lesion activity to simulate. The lesion activity maps are encoded as indices corresponding to activity levels list in a separate look up table, both of which are used by the simulation

toolboxes discussed in the next section. The lesion activity maps, binary masks, and the specified lesion characteristics (size, location, intensity) are archived for further processing.

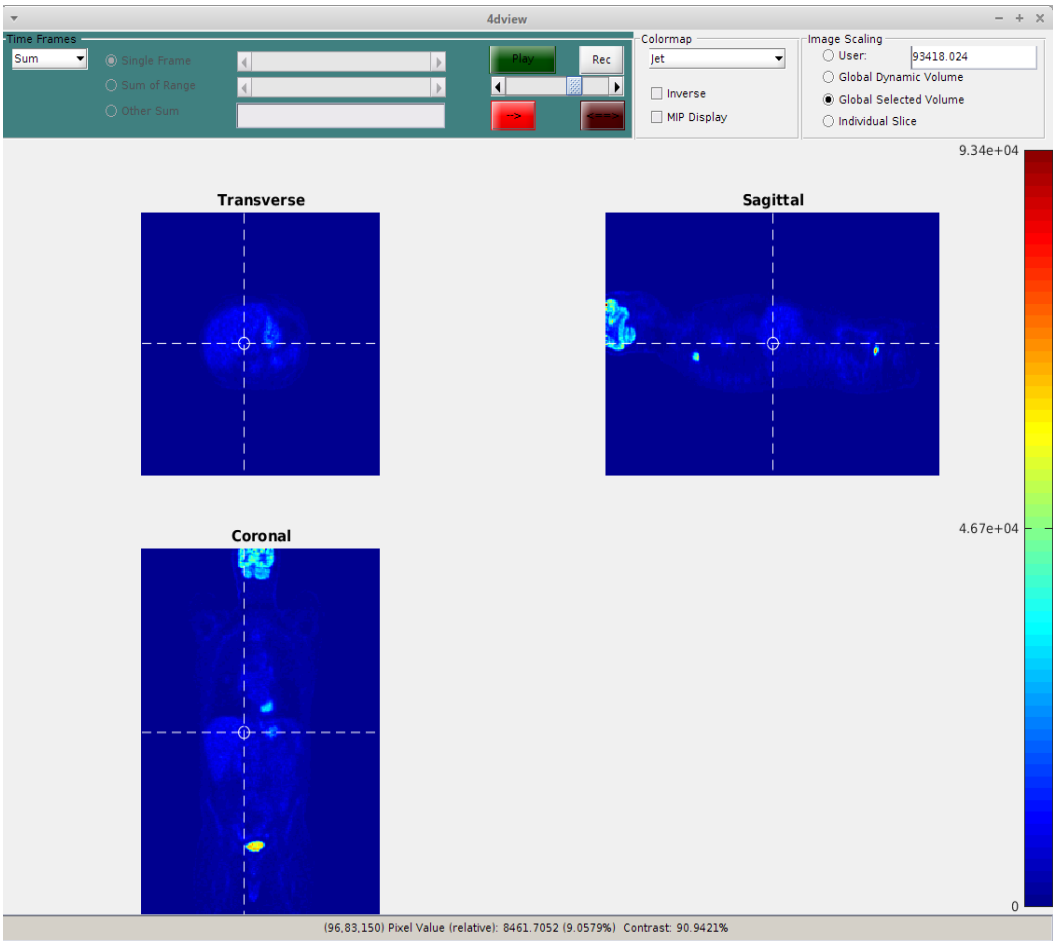

Figure 13: Screen capture of the volume viewer where a 10 mm diameter liver lesion (white circle contour) was defined in the target PET image.

3.3.4 Lesion simulation

Lesion simulation consists of using an activity distribution map within the scanner field of view and modeling the emission process to produce the projection data measured by the scanner. Different emission data simulators are available that range in complexity. In

this work we considered the use of both a Monte Carlo simulator and analytical forward projector alternatives to generate simulated emission data to combine with clinical patient emission data, as shown in conceptual block diagram in Figure 14.

Lesion simulation can generate either list-mode event data or binned sinogram data. While list-mode data provides maximum flexibility to reconstruct different types of images (i.e. dynamic, gated or static), list-mode files are large, cumbersome and are not routinely stored in clinical routine. Sinogram files in contrast, are routinely maintained and in the context of clinical FDG studies have minimal loss of fidelity over list-mode files, since acquisitions are short (2-3 minutes per bed position) and therefore do not contain meaningful dynamic data anyway. Respiratory-gated images are possibly of interest in certain cases but are of secondary importance and were not considered in the context of this work.

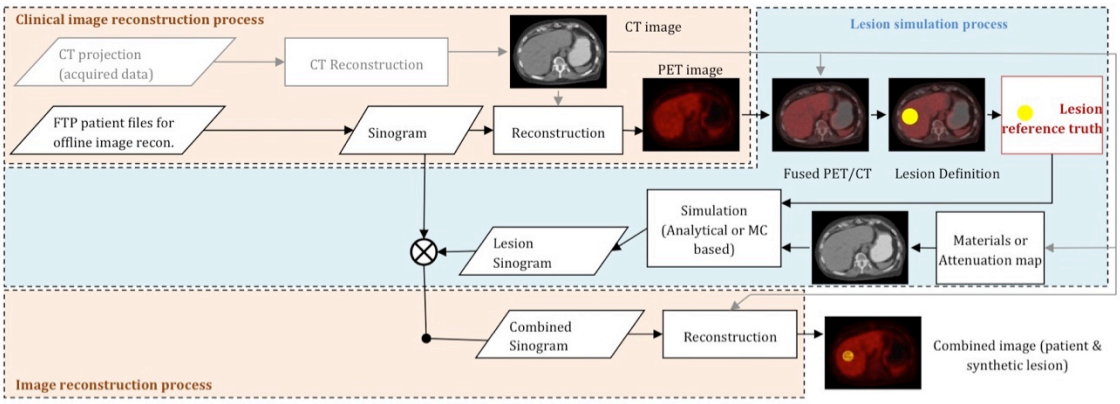

Figure 14: Lesion synthesis conceptual block diagram.

3.3.5 Lesion and Patient Data Merging

The simulated emission projection data was converted into the GE sinogram format for further processing. Proprietary read/write tools provided by GE were used to read patient

sinogram files and to add simulated lesion counts to the patient PET sinogram and to save the combined file; thereby emulating the addition of lesion activity within the patient. This step is indicated as ⊗ in Figure 14. The addition of simulated and patient sinograms is demonstrated in Figure 15.

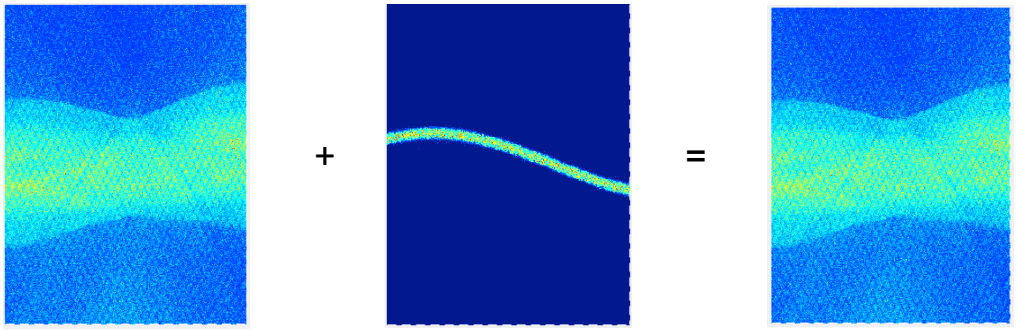

Figure 15: Trans-axial sinogram slice of the target patient PET true coincidence counts (left), trans-axial sinogram slice of the simulated lesion emission data (middle), trans-axial sinogram slice of the combined patient and simulation emission data for image reconstruction (right). Note that simulated emission data is on a lower scale than patient data and therefore is not apparent within the combined data.

3.3.6 Image reconstruction

Simulated emission data was reconstructed using the GE Recon Toolbox using the image reconstruction protocol implemented clinically: OSEM reconstruction with point-spread function modeling, 2 iterations and 24 subsets. As opposed to clinical routine, post-reconstruction filtering and time-of-flight reconstruction were not always performed, and these are specified as appropriate throughout this chapter. Images were saved in DICOM format compatible with picture archive and communication systems (PACS) within our institution.

3.3.7 Lesion records

For each image lesion information could be stored in an accompanying text file and/or in the image header. The information contained the coordinates, shape, size, activity and HU values of the lesion. In addition, lesion mask images could be saved in DICOM segmentation format and could be sent to PACS where the lesion masks could then be superimposed on the PET/CT images in the clinical image viewing system (Hybrid Viewer and Hybrid 3D, Hermes Medical Solutions, Stockholm, Sweden).

3.4 Lesion Synthesis using Photon Tracking – SimSET

A virtual model of the GE Discovery 710 PET scanner was designed using geometry and detector crystal composition specific to the scanner. The virtual PET scanner was designed in SimSET, a general-purpose nuclear imaging Monte Carlo package (University of Washington, Seattle, WA).

The virtual PET scanners system parameters were configured to model typical 3D acquisitions. The virtual PET scanner consisted of 13,824 lutetium-yttrium oxyorthosilicate (LYSO) scintillation crystals configured as a 70 cm diameter ring with 15 cm axial lengths. The 13,824 LYSO detectors were organized into four detector block rings, with 64 blocks per ring, with each block consisting of with 6 and 9 crystals in the axial and trans axial dimensions respectively. All simulations were configured to simulate static acquisitions with a scan time of 2.5 minutes per bed position using ^{18}F-FDG, as per our clinical routine. True, scatter and random coincidences were simulated, however only true coincidence events were used for validation purposes.

Desired activity and attenuation distribution of simulated objects were specified using 192x192x47 volumetric image matrices. Indices of the activity image corresponded to the

specified activity index table in units of Ci/cc. Indices of the attenuation image corresponded to an attenuation index table with preconfigured materials within the simulation package (e.g. soft tissue, bone, air, lungs).

The SimSET simulation tool generated a 3D sinogram representing the number of events measured by each LYSO detector pair with dimensions $13{,}824^2$. By mitigating impossible detector combinations and performing a proprietary sinogram-meshing algorithm reduced the sinogram into the proprietary 3D GE sinogram with dimensions $381 \times 553 \times 288$ for image reconstruction using non-TOF methods. These dimensions correspond to 381 LOR angles, offset distance and ring combinations.

3.4.1 Validation of Virtual Scanner Geometry

The simulation must produce the desired lesion characteristics such as position, size, and orientation do not deviate when embedding simulated projection data with raw patient projection data. Accurate lesion synthesis requires that synthetic lesions are properly positioned in raw patient data, corresponding to precise placement post image reconstruction. Therefore, when designing the virtual PET system, the placement of each detector in 3D space and the sequential numbering of virtual LYSO detectors must correspond with those of the physical scanner

Detector alignment of the virtual scanner was investigated with a custom developed numerical phantom. The numerical phantom was generated using an image matrix size of $192 \times 192 \times 47$ voxels corresponding to a 700 mm trans-axial FOV and 154 mm-axial FOV. A uniformly distributed (24 MBq/cc) hollow cube with dimension $400 \times 400 \times 10$ mm and 20 mm wall thickness was placed in image slices 4-6 (of 47). A uniformly distributed (12 MBq) circle with a radius of ~18 mm was placed inside the bottom right

corner of the hollow cube. Background activity was set to zero. No attenuation simulation was included, by setting the material in the entire field of view to air.

A reference of the expected simulation generated sinogram was generated using a vendor provided forward projector as used in the GE OSEM image reconstruction algorithm. Sinograms from the simulation were visually compared with those of the forward projector. Likewise, reconstructed images from the SimSET simulation were visually compared to the numerical phantom (Figure 18).

3.4.1.1 Results

The SimSET simulation elapsed 19.8 minutes and simulated ~120 million decay events associated with 2.5 minute simulated acquisitions time. The simulation tracked ~240 million photons, resulting in ~1.84 million detected true coincidence events. Comparison images of the SimSET simulation produced sinogram and forward projection produced sinograms are shown in Figure 16. The two sinograms appear to agree in pattern, confirming proper geometry and detector indices in the simulation configuration. As expected, the SimSET produced sinogram clearly shows the presence of noise associated with photon counting statistics, while noise is absent in the forward projected sinogram.

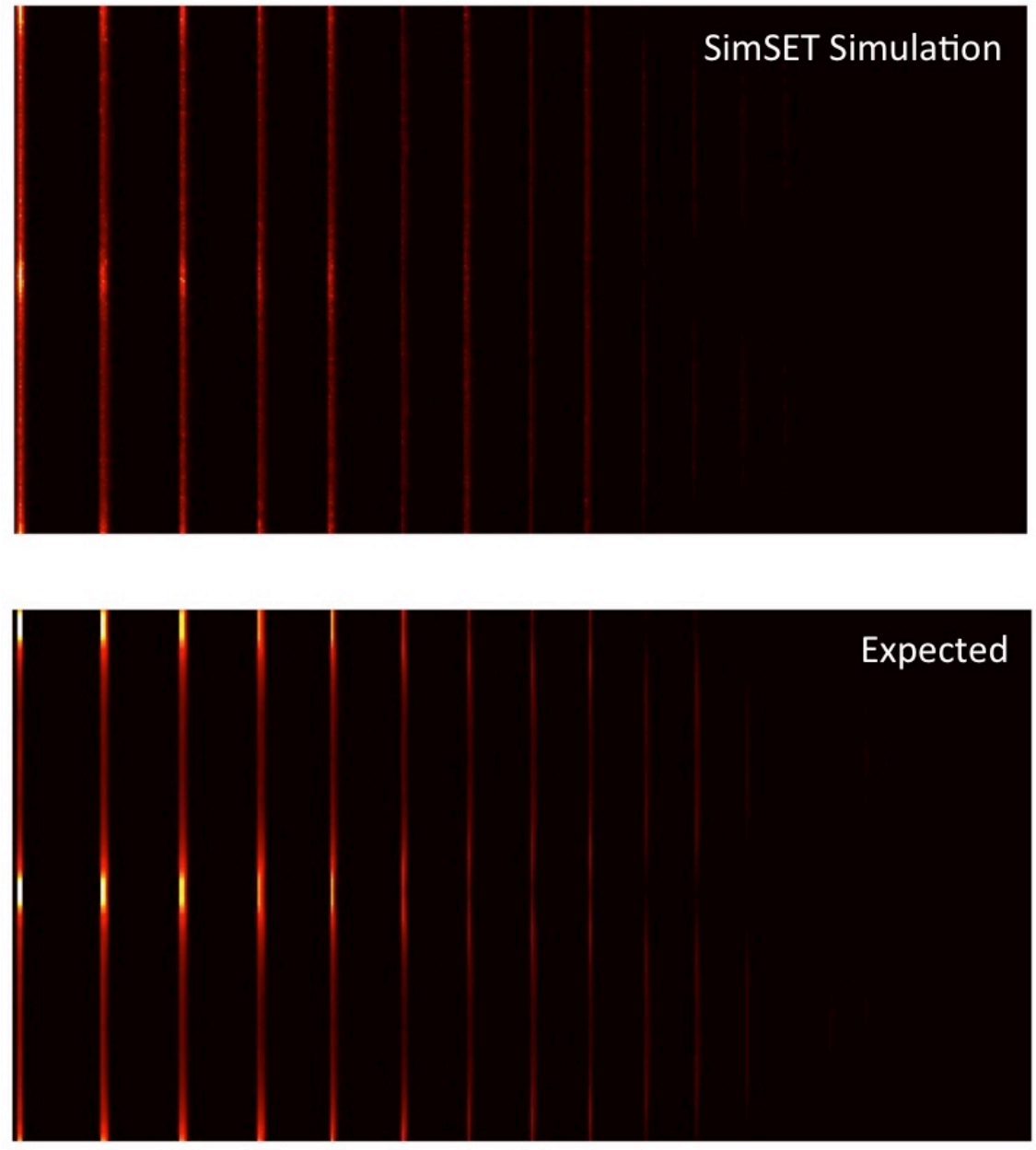

Figure 16: Maximum intensity projection (MIPs) in the transverse direction of the proprietary GE sinogram produced using SimSET simulated counts (top) and the expected GE sinogram produced using a forward projection of the numerical input phantom. The forward projector is vendor provided as is used in their reconstruction algorithm. Images were scaled individually.

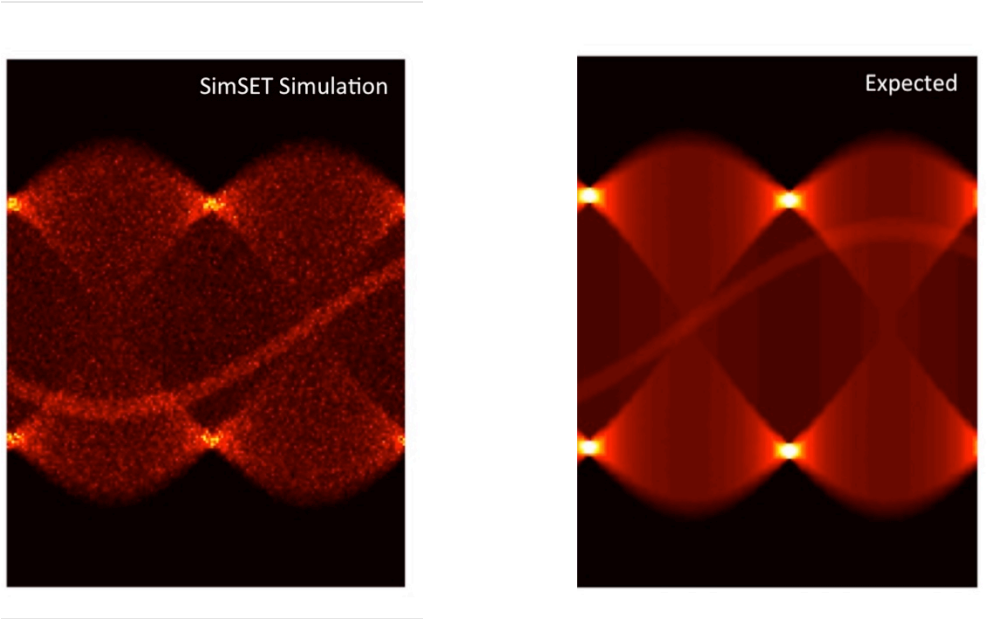

Figure 17: Maximum intensity projection (MIPs) in the sagittal direction of the proprietary GE sinogram produced using SimSET simulated counts (top) and the expected GE sinogram produced using a forward projection of the numerical input phantom.

Figure 18 demonstrates that the reconstructed images also agreed closely with the simulation model with regards to orientation and size. As expected reconstructed images had the presence of noise associated with counting statistics and was blurry due to loss of spatial resolution consistent with the PET system design.

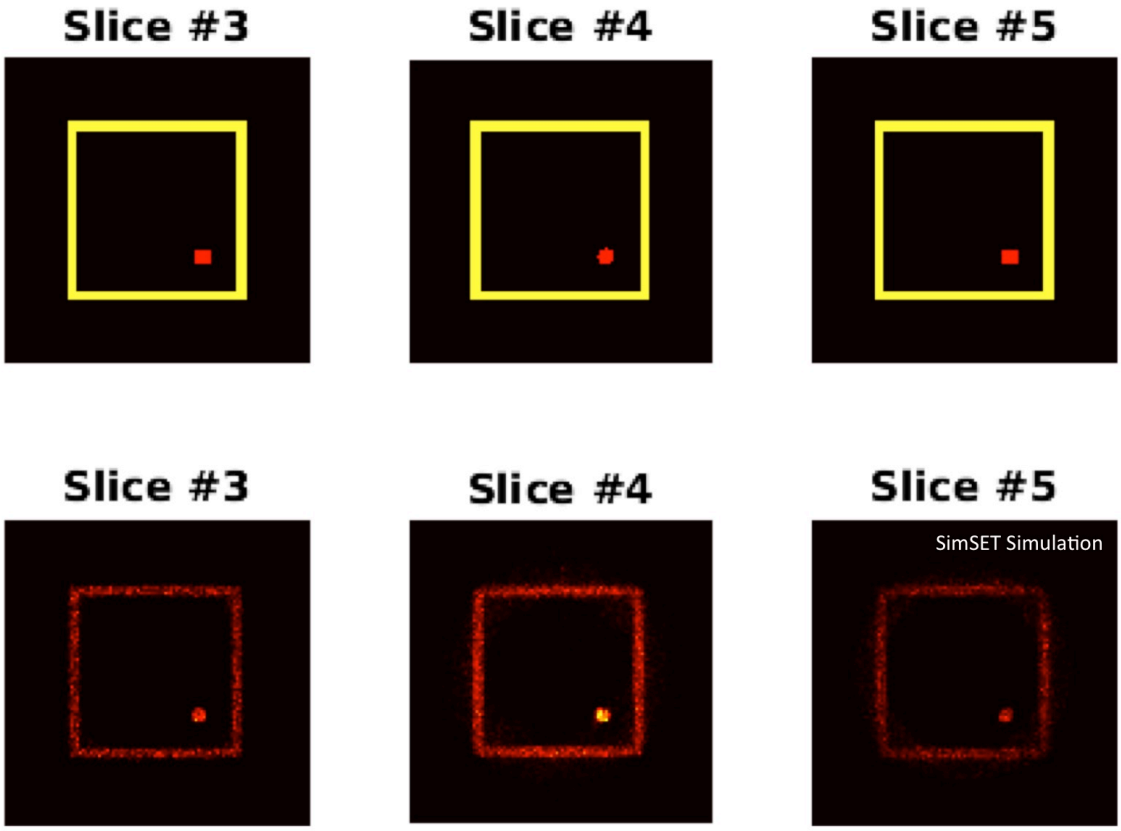

Figure 18: Trans-axial slices of the input phantom (top) and reconstructed SimSET simulation (bottom) showing good agreement with regards to orientation and size of the image objects.

3.4.2 Validation of Patient Attenuation Modeling

As the simulated lesion counts are added to the patient raw detected counts prior to image reconstruction, the modelling of patient attenuation must be considered to negate the effects of the attenuation correction by the image reconstruction algorithm. SimSET enables users to specify material distribution maps of the simulated object to consider the effect of attenuation in the simulation. To model the patient in SimSET, we considered mapping the patient CT to predefined materials in SimSET.

The patient attenuation map is derived using a voxel-by-voxel approach where the intensity in HU on the CT is mapped to a material predefined in SimSET [45]. SimSET then considers photon interactions with resulting materials during the photon tracking process to model individual photon attenuation. The patient CT was mapped with 7 pre-defined SimSET materials as detailed in Table 1.

Table 1: CT Hounsfield Unit Range to SimSET Material for Attenuation Modelling

Patient CT HU	SimSET Material
HU < -700	Air
-700 ≤ HU < -250	Lung
- 250 ≤ HU < -50	Fat
-50 ≤ HU < 50	Water
50 ≤ HU < 100	Blood
100 ≤ HU ≤ 200	Muscle
HU > 200	Bone

Modeling of patient attenuation within SimSET was evaluated with a simple numerical phantom consisting of a uniformly distributed phantom defined on a 192×192×47 image matrix spanning the 700 mm trans-axial by 154 mm axial FOV. The phantom (lesion) consisted of a 70 mm diameter by 10 mm long cylinder centered on the scanner long axis with uniform activity concentration of 12 MBq/cc per voxel.

Using Table 1, the patient CT was converted into the SimSET attenuation material map shown in Figure 19. The phantom emission was simulated to produce 3D sinograms, which were then reconstructed without time-of-flight or post-reconstruction smoothing. The reconstructed image volume was then compared to the input phantom to evaluate uniformity and activity concentration accuracy associated with accurate attenuation modeling (and subsequent correction during image reconstruction).

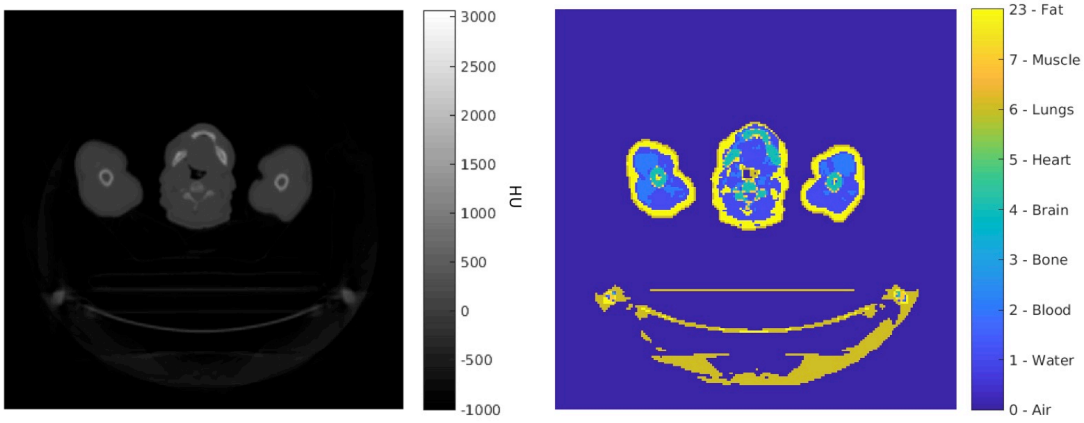

Figure 19: Trans-axial cross section of the target patient CT in HU (left) translated into a SimSET materials map (right) where a color represents material indices.

3.4.2.1 Results

The 2.5 minute simulated acquisitions elapsed over 110 minutes, producing 1.15 billion simulated decay events. The simulation tracked 2.3 billion photons, resulting in 259,401 detected true coincidence events illustrated in Figure 20 and corresponding to the expected sinogram shape of an axially centered uniform cylinder.

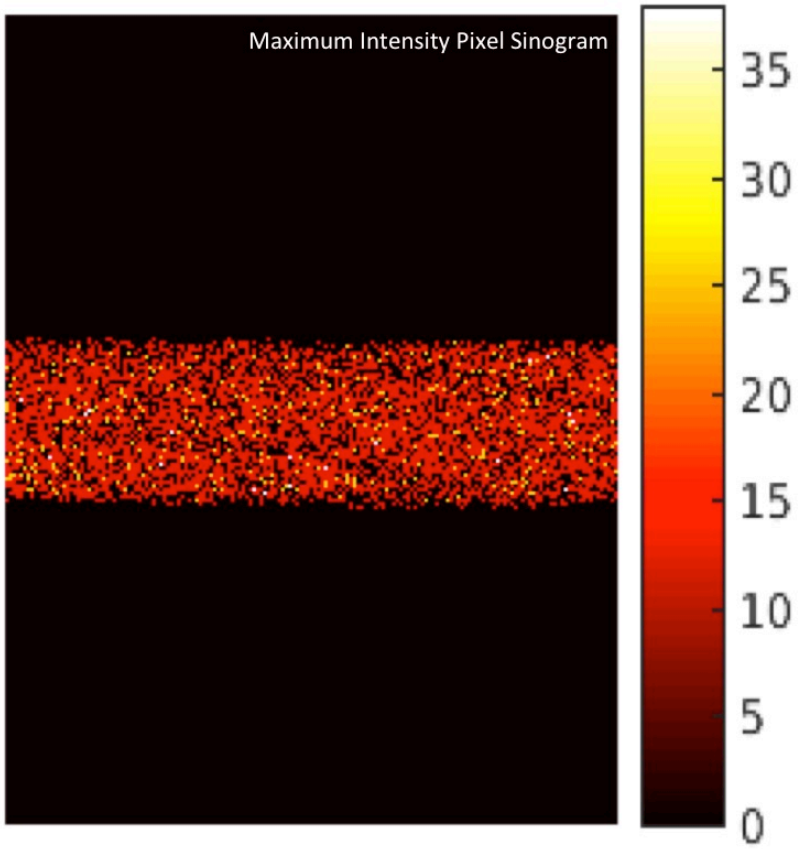

Figure 20: Maximum intensity pixel (MIP) image of the SimSET simulated sinogram. Note the realistic noise characteristics in simulated events associated with counting statistics.

Qualitative visual evaluation of the reconstructed image with attenuation correction clearly exhibits artifacts attributed to inaccurate attenuation modeling regardless of whether attenuation correction was applied or not. The resulting reconstructed image did not match with the corresponding input phantom.

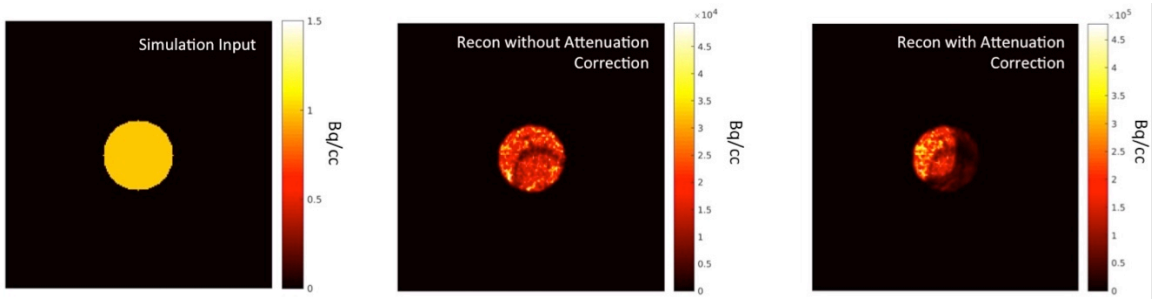

Figure 21: Transaxial slice of simulated disk (left) and after emission simulation and reconstructed without attenuation correction (centre) and with attenuation correction (right).

3.4.3 Discussion

We evaluated using the SimSET photon tracking simulation package to simulate physical phenomenon associated with PET imaging including photon emission, photon interactions and photon detection. Consequently, realistic statistics including noise could be produced. Furthermore, this framework (at least in theory) offers the ability to model lower order phenomena such as randoms and scatter that could further improve the simulation accuracy.

The numerical phantom described in Validation of Virtual Scanner Geometry proved especially useful as the square boarder and asymmetric properties enabled validation of the object orientation and sizing. These results validated appropriate configuration of the virtual detectors corresponding to the physical scanner. Even a slight inconsistency between the simulated and physical scanner produced clear artifacts in the reconstructed image, which would not necessarily be visible with a circularly symmetrical phantom. By subtracting the simulated image from its reconstructed counterpart, we were able to easily validate that the virtual PET scanner configuration agreed with the physical scanner, as no positional or rotational biases existed between the two (images not shown).

Despite the promises of using a photon-tracking simulation for lesion synthesis, several important challenges presented themselves during this work. Firstly, the attenuation modelling of the patient using SimSET predefined materials proved to be insufficient for our purposes as the attenuated corrected image from the image reconstruction workflow did not produce a uniform cylinder as expected from the simulated input numerical phantom. This can ultimately be attributed to the number of materials or more likely, inaccurate material definitions or simulation configuration. While the SimSET package does offer the ability to configure materials, this was beyond the scope of this work and we relied on predefined materials, hence we suspect in appropriate configuration of the simulation, which we were not able to resolve. Regardless, challenges with the attenuation simulation clearly demonstrated to us potential limitations of this approach including susceptibility to CT image noise and beam-hardening artifacts.

The simulation times we experienced in the validation studies proved to be a practical challenge. In our validation studies, we limited our numerical phantoms to only a few slices to achieve practical simulation times while maintaining realism. However, simulating entire numerical phantom volumes (e.g. WB patient scan) would potentially results in simulation times on the order of days (without the use of a computing cluster), which would pose a practical limitation on the number of synthetic lesions that could be generated.

While the SimSET simulation package can model TOF data, GE Discovery TOF sinogram file formats and implementation of the Reconstruction Toolbox do not provide a simple method to combine simulated and patient data. Hence, even with the vendors

assistance, we were not able to implement lesion synthesis in TOF PET which limits our ability to create synthetic lesions in images that are equivalent to current clinical practice.

3.4.4 Conclusion

The simulated geometry of the physical system being modelled was validated. The limitations associated with modelling of patient attenuation within the pre-defined materials within the SimSET, the associated technical constraints to combine TOF data and long simulation times warrant alternative simulation methods to be considered.

3.5 Lesion Synthesis – Analytical

Like Monte Carlo based lesion synthesis an activity distribution map and patient attenuation information are required for lesion synthesis. Activity maps were volumetric images in units of Bq/cc and attenuation was derived by the toolbox for CT images [29]. Phantom lesion images were generated programmatically (MATLAB), and patient inserted lesions were defined (location, geometry, shape and contrast for each lesion) graphically using an in-house developed image viewer[1]. Lesion image volumes were defined as radionuclide activity concentrations [Bq/cc] and were then processed in a fully automated workflow to generate reconstructed images of the patient PET with embedded synthetic lesions.

3.5.1 Methods – Lesion Insertion in PET

Iterative image reconstruction, such as the OSEM algorithm included in the PET Reconstruction toolbox, include a forward projector that estimates the detected emission

data from an estimated activity distribution map. This functionality is essentially the same as a simulator but is optimized by the vendor for the specific scanner geometry and computational efficiency. For the purpose of lesion synthesis, we therefore appropriated the forward projector for lesion emission simulation.

Use of the vendor provided forward projector has several benefits: (1) it inherently supports TOF simulation, (2) it accepts activity concentration images as inputs, (3) the output is in GE sinogram format that can be easily added to the patient sinogram, and (4) attenuation, geometric efficiency and scanner resolution simulations are performed in an identical manner as during reconstruction and therefore are perfectly matched to the correction processes.

Using this tool, we employed a technique that specifies the activity distribution and positional location of the lesion with respect to the reconstructed target patient image [29]. The projector generates TOF projections for one phi angle at a time taking into account specific detector efficiencies (intrinsic and geometric), patient attenuation and scanner resolution (PSF) [29]. During the image reconstruction process, TOF patient projections were combined with simulated TOF lesion projections.

3.5.2 Methods - Validation Experiments

Three validation experiments were devised to evaluate the placement (location), geometry, shape, contrast recovery, attenuation modelling and bed-stitching capabilities of the lesion synthesis in PET workflow.

A single bed-position, 2.5 minute, PET scan was acquired using a GE Discovery 710 PET/CT in our clinic in the absence of activity or attenuating medium in the scanner

FOV, generating an empty PET scan. The empty scan data was used to validate phantom data alone – without embedding synthetic lesion data in patient data.

Validation of embedding synthetic lesions into real PET and CT data was performed using an 8-bed-position, 2.5-minute/bed positions, ^{18}F-FDG patient PET scan acquired in the clinic.

Using the lesion synthesis PET workflow, TOF projection data was generated for the numerical phantom, and combined with projection data from the PET scan for image reconstruction and analysis. The raw projection data was reconstructed using TOF-OSEM with 2 iterations and 24 subsets with point spread function correction (VUE Point FX Sharp IR).

3.5.2.1 Validation Study 1 – Validation of PET Simulation Geometry

A custom hot-sphere numerical phantom with geometries inspired by the NEMA IEC Body PET phantom was developed, comprised of eight spheres with diameters 2, 5, 10, 13, 17, 22, 28, and 37 mm. A cylinder with radius of 280 mm and length of 78 mm was generated for slices 13 through 34 enclosing the spherical volumes to act as background activity. Background activity was initialized as 50 MBq/cc. The spheres were located at 30 mm radius from the image centre at 45-degree intervals along the trans-axial plane. Spheres activities were ×1.5 background. A 10 mm radius hollow cylinder (no activity) was centered on the image long axis, as illustrated in Figure 22.

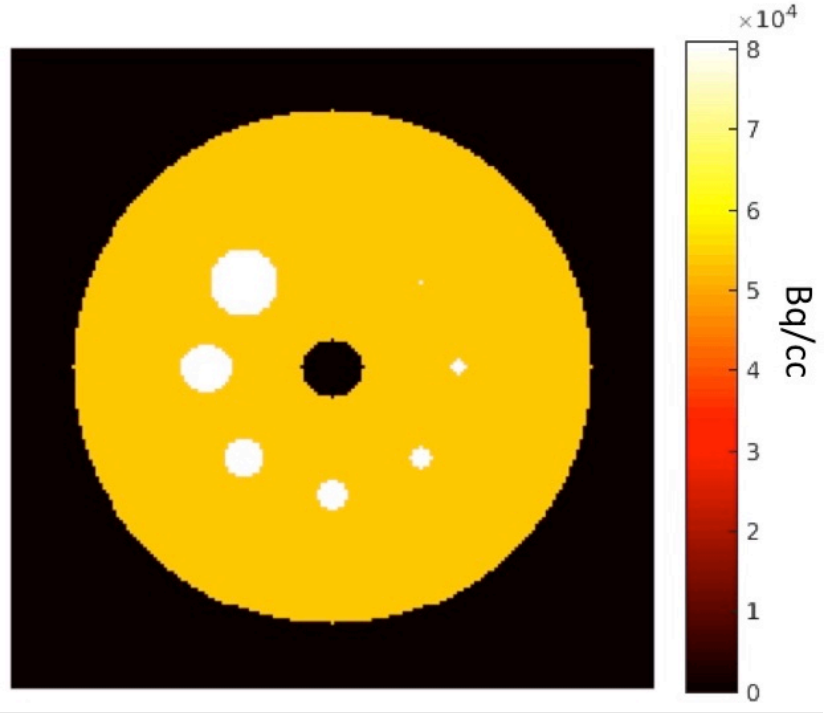

Figure 22: Cross section of NEMA hot-spheres numerical phantom used as an input to the simulation for validation of simulation geometry.

Projection data was generated for the numerical phantom and combined with projection data from the empty PET scan and then reconstructed error images between the reconstructed volume and input numerical phantoms were generated and evaluated for differences.

The error associated with each voxel was calculated using equation 4, where $\delta_{i,j,k}$ represents the error as a percentage, x and y represents voxel intensity at indices of i,j,k for the input numerical phantom and reconstructed image respectively.

$$\boldsymbol{\delta_{i,j,k} = \frac{y_{i,j,k} - x_{i,j,k}}{x_{i,j,k}}} \qquad \boldsymbol{4}$$

The activity in each simulated sphere was sampled from the reconstructed image deriving a curve characterising the relationship between sphere size and associated uptake error.

Average regional differences in each sphere were evaluated using a region of interest (ROI) defined by the sphere mask in the input numerical phantom (i.e. sampling of the entire sphere volume).

3.5.2.2 Validation Study 2 – Validation of Activity Linearity

The custom numerical phantom illustrated in Figure 23 was developed to evaluate the linearity of the activity range of the virtual PET scanner. A cylinder with radius of 280 mm and length of 69 mm was generated between slices 13 and 34 enclosing the spherical volumes to act as background activity. The background activity was defined as 50 MBq/cc. A cylindrical band along the length of the background activity cylinder was generated with an inner diameter of 140 mm and width of 35 mm. The band was placed centred within the phantom with an intensity ×1.5 background.

Two sizes of spheres were simulated: 7 spheres 22.5 mm and 15 sphere 15 mm diameter. Contrast for both set of spheres ranged from ×1 to ×2 background.

As with the hot-sphered phantom, the phantom was forward projected, reconstructed, and an error image was generated. The mean activity in each sphere was sampled and plotted as a function of the the corresponding simulated sphere activity concentration.

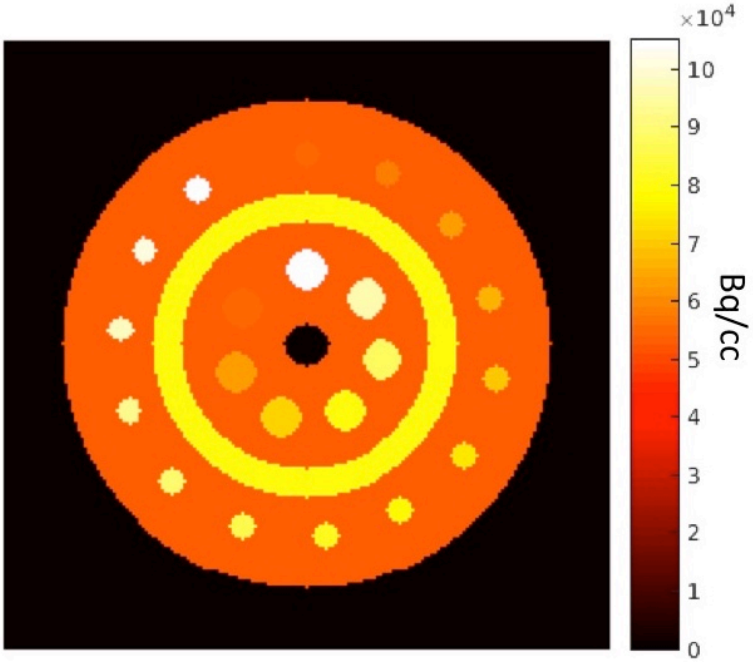

Figure 23: Cross sectional slice of contrast numerical phantom for input into PET simulation.

3.5.2.3 Validation Study 3 – Validation of Attenuation Modeling and Bed Stitching Uniformity

A numerical phantom was designed to evaluate several effects simultaneously including attenuation modeling and stitching of multiple bed positions. A uniformly distributed cylinder (5 kBq/cc) with radius of 280 mm was generated in all trans-axial planes along 8 bed positions (98 cm length), centered on the scanner long axis. By comparison, patient activity concentrations were as high as 70 kBq/cc and ~ 25 kBq/cc on average.

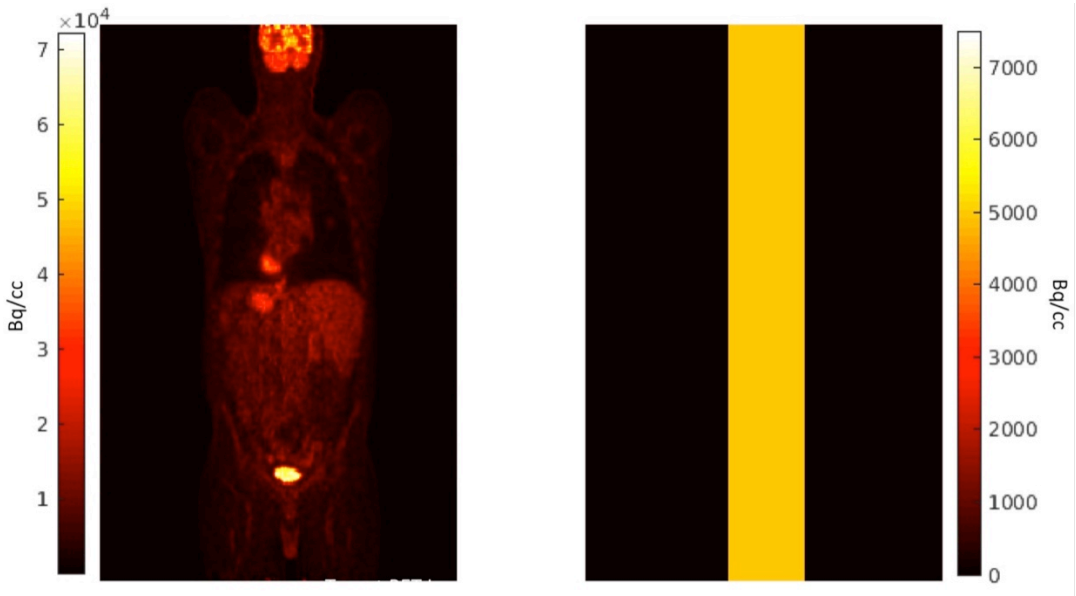

Figure 24: Coronal plane of target patient PET image (left) and numerical phantom (right) that was simulated within the patient evaluating the attenuation modelling and bed stitching of the Lesion Synthesis Toolbox.

TOF Projection planes were generated for the numerical phantom using the patient CT data to simulate attenuation as if the rod activity was present within the patient. Attenuated projection planes were combined with raw patient projection planes acquired using an 8-bed clinical protocol reconstructed without filtering.

The activity of the simulated rod was extracted by subtracting the patient image (without the synthetic activity and reconstructed with the same parameters) and the rod in the difference image was evaluated for uniformity across all trans-axial slices. Errors less than ±5% of the simulated rod activity concentration (i.e. 0.25 kBq/cc) were considered acceptable.

3.5.3 Results

3.5.3.1 Validation Study 1 – Validation of PET Simulation Geometry

Visual analysis of residual and percent error images indicated perfect alignment of simulated structures, with largest magnitude errors focused around the edges of objects (Figure 25). Sampled activities in spheres ≥17 mm diameter agreed well (<5% error) with simulated activities and precipitously underestimated activates in smaller spheres, consistent with partial volume effects (Figure 26).

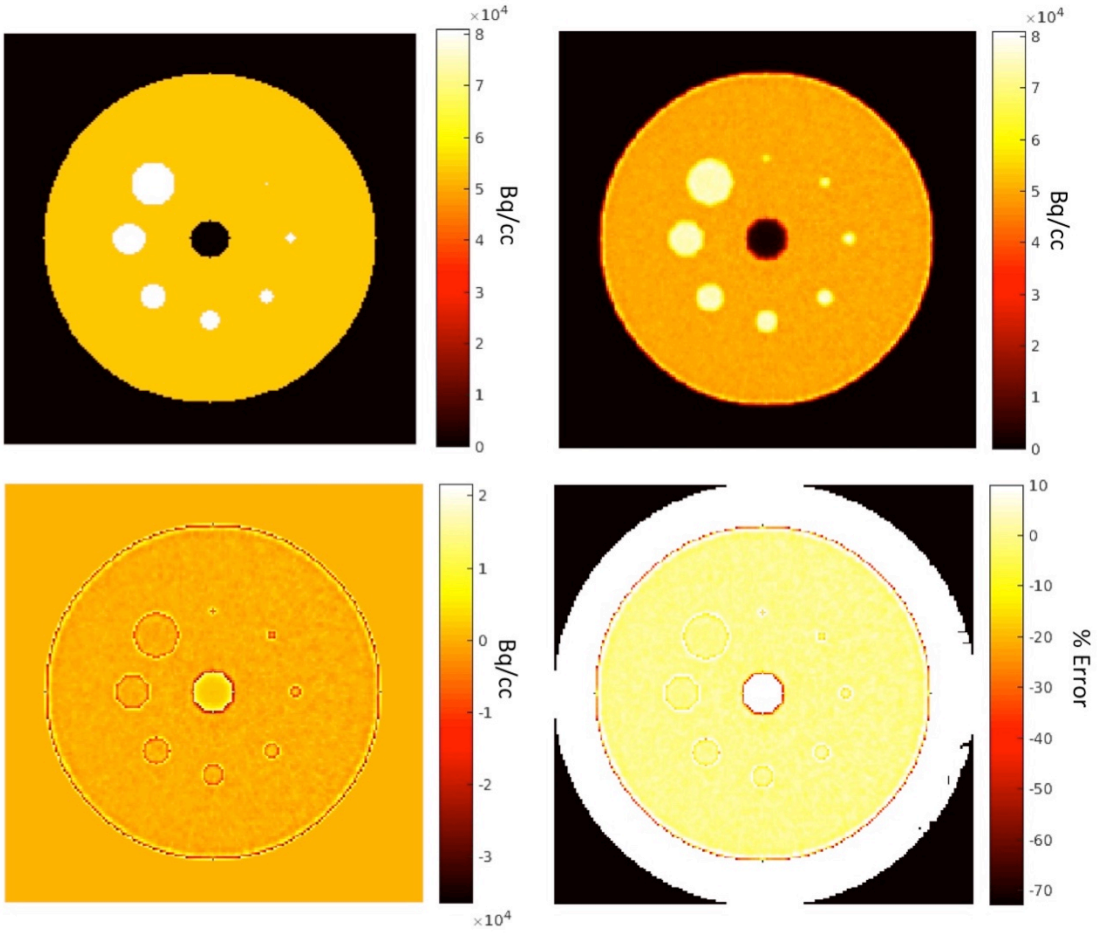

Figure 25: Trans-axial slice of the hot-sphere input (top-left), reconstructed image (top-right), residual image (bottom-left) and percent error image (bottom-right). Note the largest errors are around the edges of simulated objects. All colorbars except bottom-right are in units of Bq/cc.

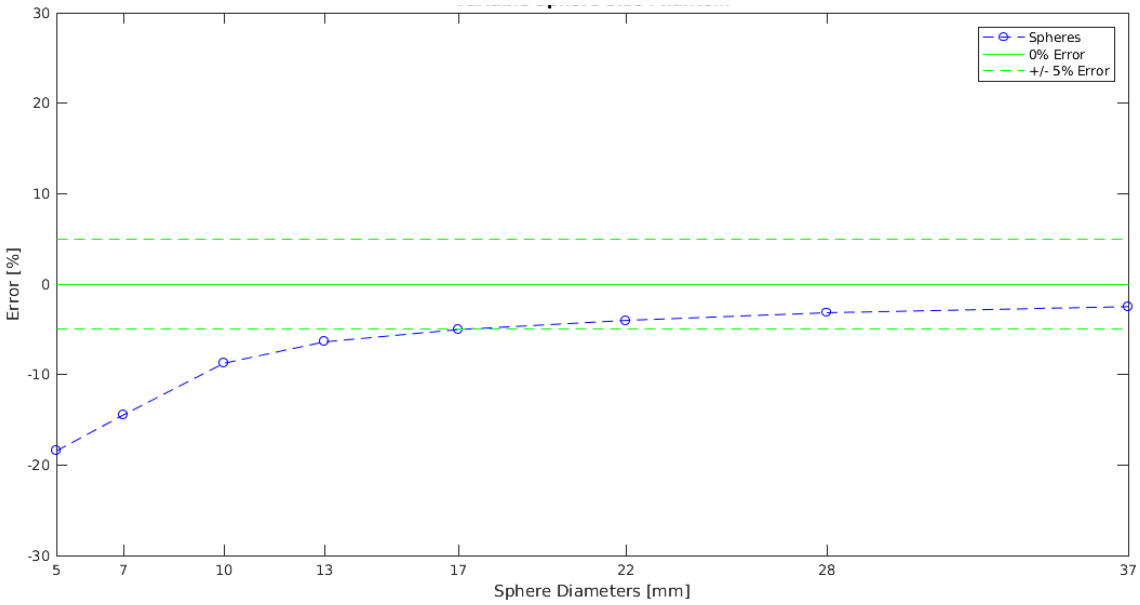

Figure 26: A plot illustrating the relationship of the errors between the specified spheres in the numerical phantom and the reconstruction of the simulated numerical phantom as a function of sphere size. Note the errors approach zero and remain within the ±5% threshold for spheres ≥17 mm.

3.5.3.2 Validation Study 2 – Validation of Virtual Simulator Linearity

Visual interpretation of contrast phantom residual and percent error image volumes exhibited similar patterns to hot-sphere phantom, where image structures were perfectly aligned with simulated structures and largest sampled intensity errors focused around the edges of structures. See Figure 27.

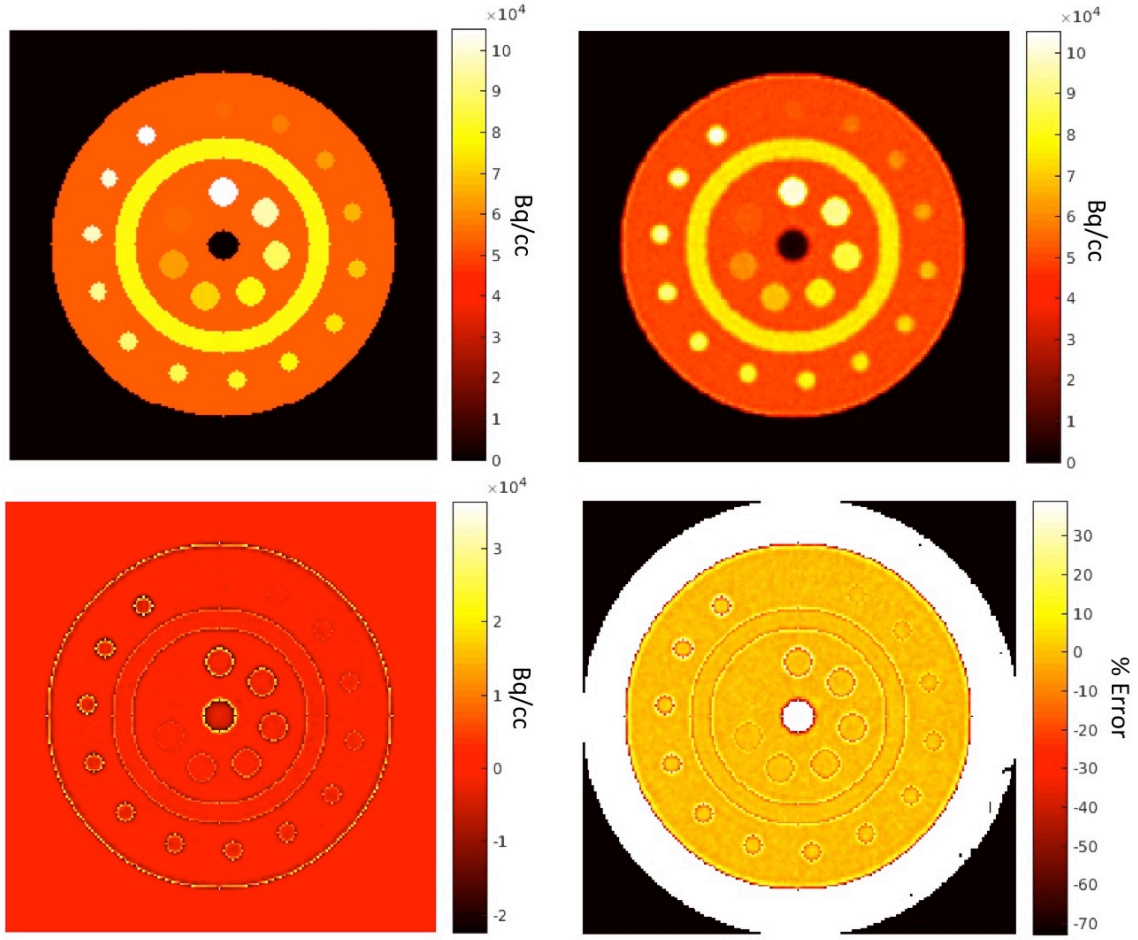

Figure 27: Trans-axial slice of the contrast input phantom (top-left), reconstructed image (top-right), residual image (bottom-left) and percent error image (bottom-right). Note the most significant errors are around the periphery of simulated objects. All image units, except bottom-right are in units of Bq/cc.

All sphere sampled activities were within 5% of simulated activity. Large spheres (22.5 mm diameter) had negligible error in image activity at any intensity, while the activities of smaller spheres (15 mm diameter) were positively biased with activity over estimation growing linearly with contrast. These results are consistent with partial volume effects as

background activity spills into the sphere ROI and with PSF reconstruction associated ringing artifact “moving” edge activity into the center of small lesions.

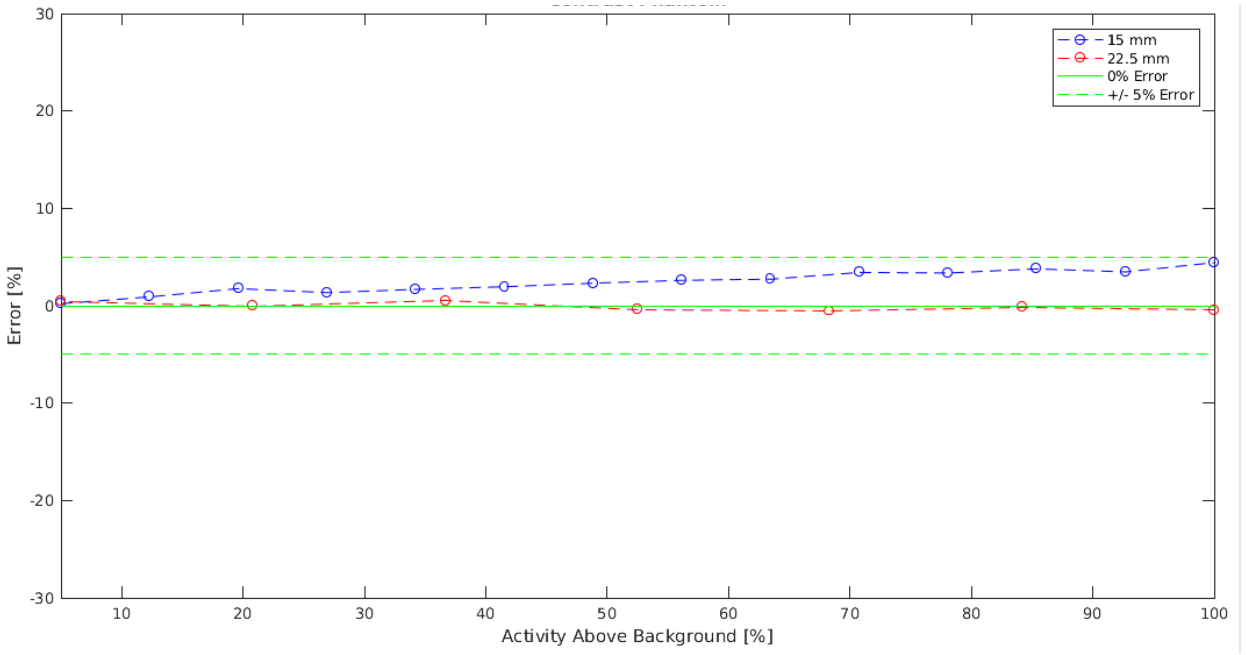

Figure 28: A plot illustrating the quantification error between the numerical phantom and the reconstructed simulation of the numerical phantom. The associated errors for 15 mm spheres (blue) and 22.5 mm spheres (red) as a function of the specified contrast above background. Note the errors are within the ±5% acceptable range.

3.5.3.3 Validation Study 3 – Validation of Attenuation Modeling and Bed Stitching Uniformity

The simulated hot rod was successfully embedded into the patient scan as illustrated in Figure 29, correctly positioned in the center of the FOV passing through each trans-axial plane and increasing the image activity.

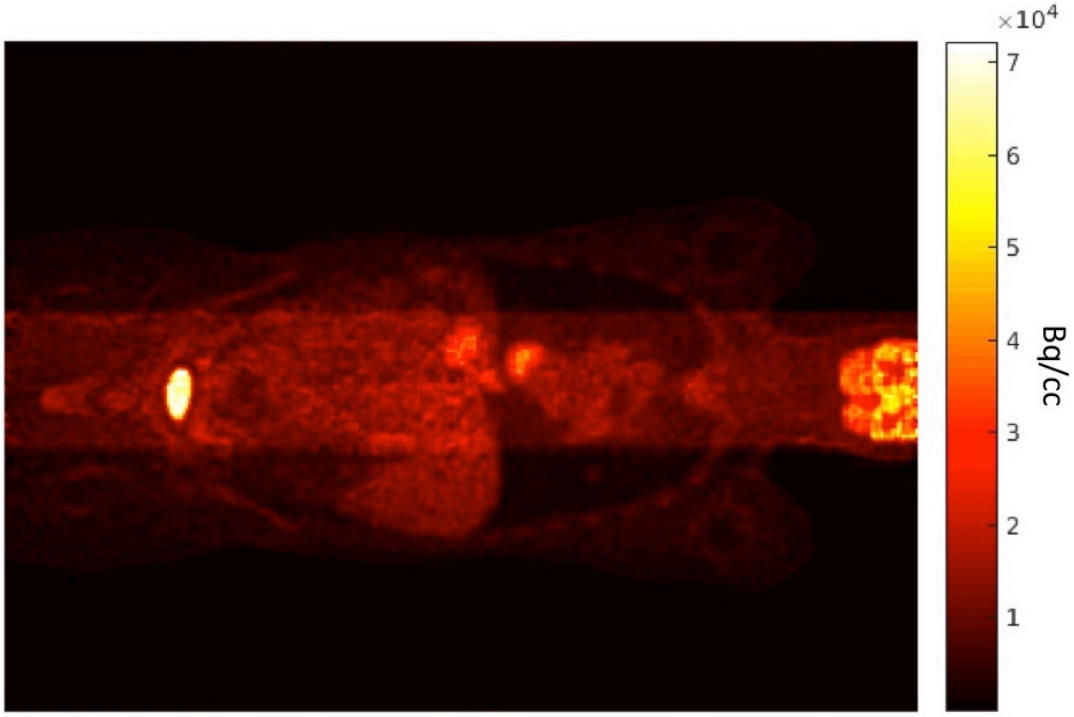

Figure 29: A coronal plane of the reconstructed PET volume generated by adding the sinograms of the simulated uniformly distributed numerical rod phantom with the sinograms of the target patient PET data prior to image reconstruction.

The simulated rod activity image was extracted from the combined patient-rod image by subtracting the original patient PET scan (Figure 30). The extracted rod was uniformly distributed along all trans-axial planes as expected, with no obvious bed-stitching artifacts present in the image. Average rod ROI intensities underestimated the simulated activity with mean activity error of -1.6 ± 1.2% as shown in Figure 31, but did not exceed the ±5% error threshold, nor did they exhibit a clear spatial pattern associated with bed-stitching problems.

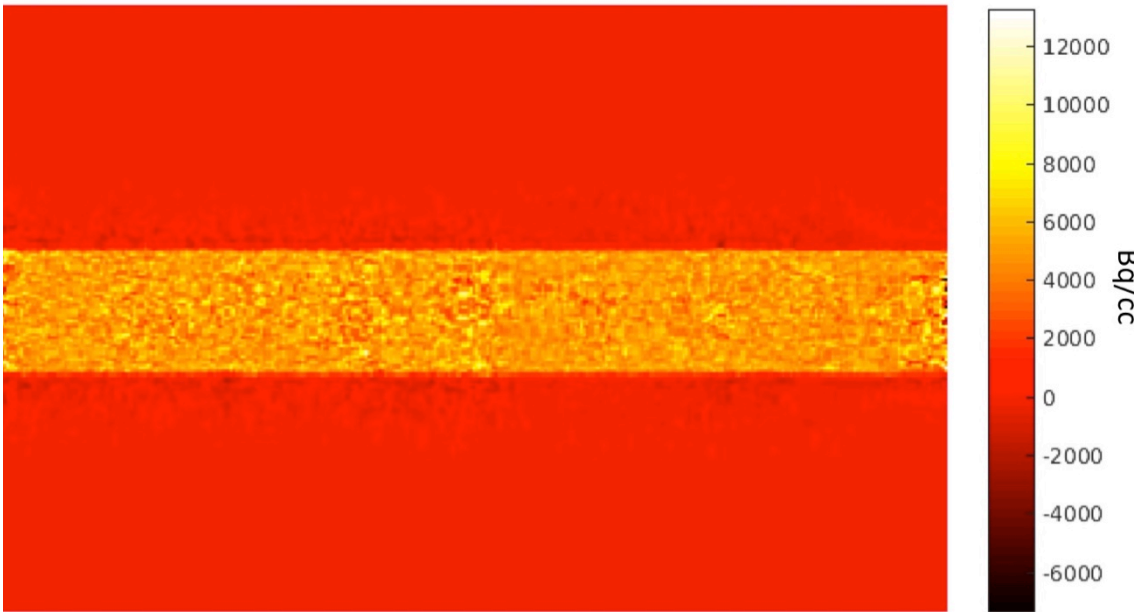

Figure 30: Illustration of the extracted uniformly distributed hot rod numerical phantom. The image was derived by subtracting the reconstructed patient volume with the embedded phantom and the original target patient PET volume. The phantom appears to be uniformly distributed where noise can be attributed from the reconstruction.

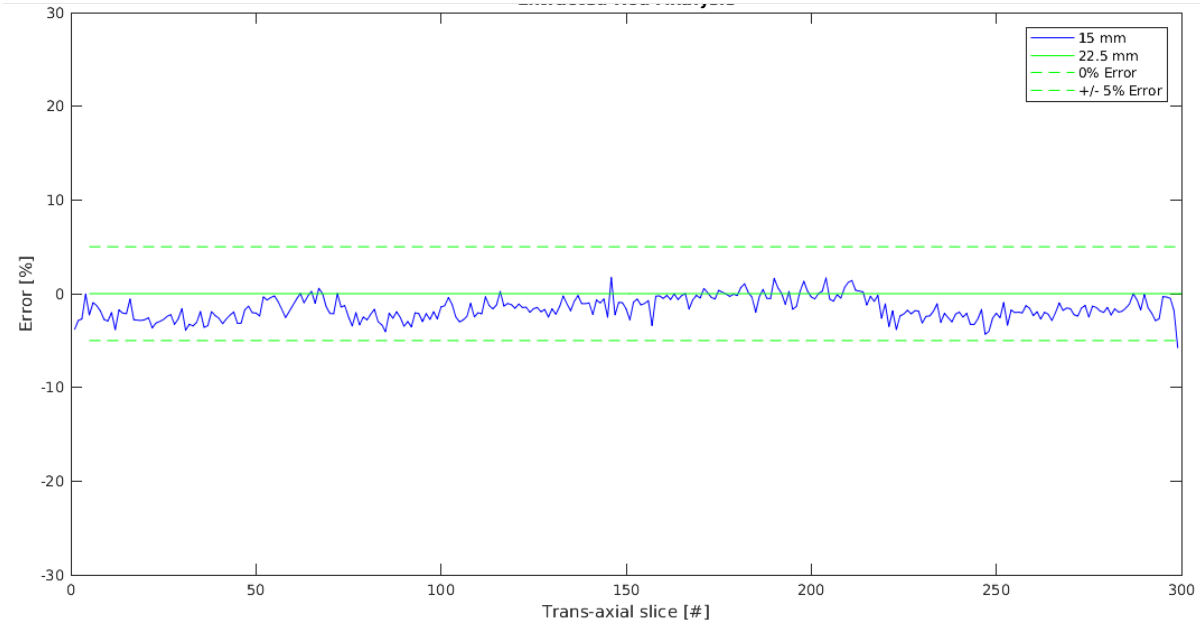

Figure 31: Mean activity error for each trans-axial slice of the extracted rod from the reconstructed PET image compared to the input numerical phantom. Note the error remains within the ±5% error threshold.

3.5.4 Discussion

The developed analytical simulation package enables the simulation of lesions using scanner geometry identical to the PET system in our clinic. Images were reconstructed using TOF OSEM with 2 iterations and 24 subsets, with point spread function correction as is used in clinical practice. The reconstruction settings did not include trans-axial and post image reconstruction smoothing to simply analysis and interpretation of biases, and because these are easily achievable as required during image visualization.

Images generated using the simulation package exhibited similar quantification biases for sampled spheres as experienced with physical PET phantom experiments, and these biases are consistent with partial volume effects due to the finite spatial resolution of the PET scanner. Our simulation studies reported underestimation as high as 20% for small spheres (<10 mm diameter), but underestimation errors were <5% for spheres with diameters ≥13mm.

The simulation package exhibited near perfect activity linearity for large spheres (22.5 mm diameter), as biases remained negligible as the activity was varied. Linearity trended towards an overestimation bias for lesion <15 mm as the specified intensity increased, however all sampled errors remained <5%.

The realism of simulated lesions was not evaluated in this work, because at these early stages of validating the workflow it was more pertinent to validate the quantitative accuracy of the simulation using simple geometric shapes and uniform activity

concentrations. Nevertheless, our workflow enables simulating non-uniform, asymmetric distorted sphere-based lesions which may be more representative of real lesions. This workflow also facilitates the usage of real lesion geometries delineated from a database of real PET lesions.

Because this workflow embedded lesion in sinogram space, before image reconstruction, lesion size and activity response may be different with other image reconstruction algorithms (e.g. BSREM). In this work we validated lesion accuracy using OSEM as it is not only our clinical standard, but also the industry standard for PET image reconstruction algorithms. Future studies to compare image reconstruction algorithms could use this toolbox not only to characterize lesion image response with lesion activity and size, but also to power perception research.

3.5.5 Conclusion

Analytical simulations offer a reliable and vendor supported approach to embedding lesions in raw PET projections. The developed workflow can be used to generate synthetic lesions with known reference truth for subsequent research including: optimization of image reconstruction parameters used in a clinical setting for the task-specific goal of lesion detection, investigation the limits of human perception and artificial intelligence-based observers.

3.6 Lesion Insertion in CT

Typically, when reading PET images, physicians pair the corresponding CT to investigate size, anatomical location and tissue densities of suspected lesions, therefore the generation of lesions in the corresponding CT is pertinent to a clinically realistic dataset.

However, as the purpose of this work is to establish the limits of detection for PET data using various image reconstruction techniques, simulation of lesions in CT data is of secondary importance.

CT lesion synthesis was performed using the image pixel substitution method. Binary maps specified during the PET lesion simulation process outlining the lesions geometry are used to generate lesions in the CT.

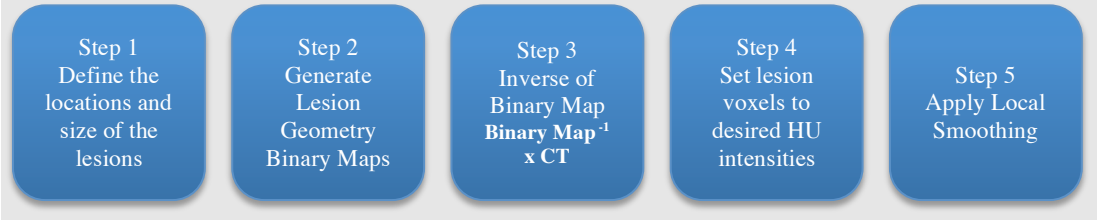

Figure 32: Lesion Synthesis in CT Workflow.

The voxels in the target patient CT associated with the lesion to embed were set to 0 by multiplying the target patient CT by the inverse of the binary lesion map. Subsequently lesion voxels in the target CT were given intensities corresponding to the desired lesion intensity in HU. A Gaussian filter with a standard deviation of 1 pixel was applied to the regions in the immediate location of the lesion to blend synthetic lesions with neighboring tissues and structures, making lesion edges less conspicuous. This method was previously described by D'Alessandro *et al* and validated to produce realistic lesion in CT images [19].

3.7 Demonstration of Lesion Synthesis in Patient PET and CT

Using the PET lesion synthesis workflow an arbitrary patient was selected from our clinical database to demonstrate lesion synthesis using the developed workflow.

Four spherical lesions were manually defined using the user interface: three lesions were within the lungs and one was specified in the liver. Lesions embedded into the CT were uniformly distributed with intensities of 50 HU. Lesion activity maps specified background to contrast ratios of 1:2.5 to 1:4.0 in increments of 0.5 as shown in Figure 33 (left).

3.7.1 Results

Simulated PET projection planes from the four simulated lesions were successfully combined with patient raw projection data, Figure 33 (right). Similarly, the four lesions were embedded into the target patient CT successfully. As expected with a voxel-substitution method, lesions were present with the specified intensities, Figure 33 (middle). Visual interpretation of image data presented no localization variability of lesions between PET and CT data, Figure 33 (right).

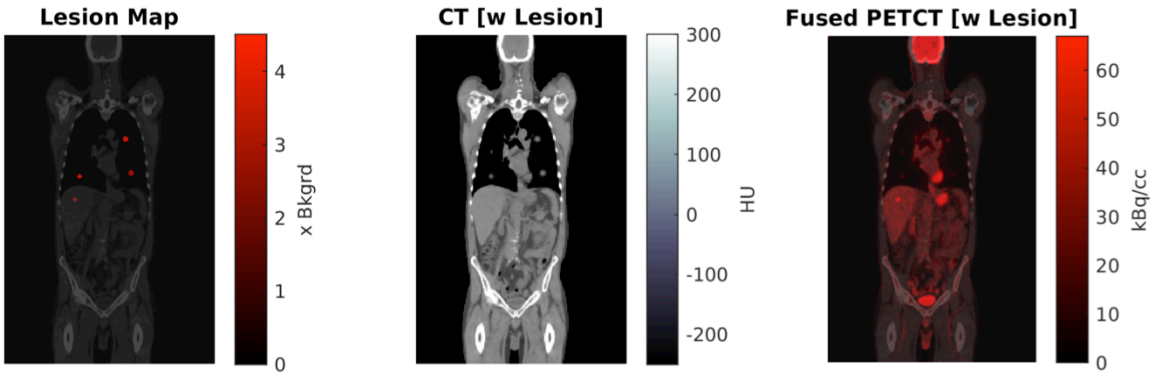

Figure 33: Application of lesion synthesis in PET and CT workflow. Lesion geometry with intensities representing the desired contrast for PET [left], Patient CT with embedded simulated lesions [middle], and Fused PET (with embedded lesions) and CT (lesion absent) [right].

3.8 Summary

The lesion synthesis toolbox is a complete platform for inserting well-characterized lesions into PET-CT. It is a standalone tool that integrates with GE 600 and 700 series PET-CT systems, but in principle the modular system design can be adapted to support other vendors and scanner models. Reconstruction is numerically identical to the GE's clinical algorithm but enables offline processing that does not interfere with clinical workflow. Resulting files including PET and CT images and lesion segmentation files are compatible with DICOM standard and are readable by third party viewer software. Because key aspects of the workflow are integrated, and automated batch lesion synthesis is possible and can be utilized to generate large lesion datasets with relatively little effort.

Chapter 4: Towards Large Scale Clinical Perception Studies

4.1 Introduction

This chapter demonstrates a practical application of synthetic lesions by evaluating the limits of detection (LOD) by human viewers, which is the task specific goal of lesion detection in clinical PET. We develop methods to ultimately deploy preliminary perceptual studies for characterizing an individual's LOD in PET using synthetic lesions. The Lesion Synthesis Toolbox described in the previous chapter is used to generate a library of patient images with synthetic lesions of varying size and intensity within in the liver (abdominal) region. In the first portion of the chapter using a single patient, a small sample size of lesions and observers are used to approximate the LOD as a baseline. In a second, follow-up perception study, this estimate is subsequently used to generate a larger lesion database for evaluating LOD in small lesions with contrast levels surpassing the upper-most contrast limit probed in the preliminary study. Based on our experience from these first two perception studies, custom perception-study software was designed to conduct future perception studies with greater efficiency, to reduce observer fatigue and to enable interrogation of LOD with greater precision. The chapter concludes with a conceptual design of a practical perception-study using the LOD estimates from the preliminary studies and the perception-study software to effectively evaluate the LOD.

4.2 Database of Candidate Patients for Lesion Synthesis

To generate a database of patient images with synthetic PET lesions we sought patient images clear of existing disease (lesions), or patients with only localized disease (e.g. non-metastatic disease with only an isolated primary tumour). Candidate patients were

screened from patients imaged within the previous month in our clinic (December 2018 – January 2019), under an Ottawa Hospital Research Ethics Board approved study (REB #20150509) where no explicit patient consent was required. Using the clinical PACs system, a research trainee, selected 15 candidate patients with disease absent in the mediastinum and/or the abdominal regions. Candidate patient images were further screened by a physician with dual certification in nuclear medicine and radiology specializing in PET/CT and with over 5 years experience. Of the 15 initial patients, 3 patients had no disease present in the mediastinum, 8 patients had no disease in the abdominal region, and 4 patients did not have disease in either region. Using the developed GE Discovery Data Retriever, raw data files for each of the candidate patients were transferred from the clinical console to the Lesion Synthesis Toolbox processing suite for subsequent processing.

4.3 Preliminary Perception Trial – Estimating Human LOD

The first perception study sought to roughly estimate the LOD of human observers to inform subsequent studies regarding relevant lesion parameters that challenge detectability by human observers. These limits can hence enable more efficient and precise sampling of the LOD with few lesions around the LOD at small parameter increments.

4.3.1 Methods

4.3.1.1 Database of Patients with Synthetic Lesions

Synthetic liver lesions of varying size and lesion-to-background ratio were generated in an arbitrary subset of four patients. The liver was selected as the lesion site because it

generally has uniform and moderate FDG uptake and is commonly used as a reference region for reporting lesion intensity [46]. Lesions were generated at a resolution of 256×256 corresponding to a 700 mm trans-axial FOV. Each lesion was manually placed within the liver so that no more than one lesion appeared in any trans-axial slice. Within a single patient 9 possible lesions with varying background to contrast ratios were generated. Lesions diameters varied between patients from 2.5 mm to 10.0 mm in increments of 2.5 mm and lesion-to-background ratios varied from 1:1.25 to 1:2.25 in increments of 0.25, resulting in a total of 36 lesion variants.

Images were reconstructed using the clinical image reconstruction parameters with 6.4 mm full-width at half-maximum (FWHM) Gaussian smoothing applied post-reconstruction.

4.3.2 Research and Ethics Board Application

In addition to the patient image database used for embedding lesions, a separate research and ethics board application approved by The Ottawa Health Science Network Research Ethics Board (OHSN-REB) concerning perception study participants in this study (REB # 20180722-01H).

4.3.2.1 Perception Experiment

Using the preliminary library of synthetic liver lesions, the LOD were evaluated for two inexperienced image readers (i.e. non-clinicians). Volumetric PET images with synthetic lesions (without corresponding CT) were shown using a custom volumetric viewer. The observer noted Cartesian coordinates of the center most voxel (CMV) of suspected lesions. The study administrator confirmed localization if the suspected lesion CMV was within 2 voxel (7 mm) of the true lesion location. True-positive response rates were

summarized by lesion size and contrast level and the transition zone in true-positive rates was considered a rough estimate of the individual's LOD. Observers were not limited in the number of responses and were made aware of the total number of lesions present in each image. Observers could use all viewer functionalities including upper image intensity limit. A screen capture of the viewer with crosshairs centered on a perceived lesion is demonstrated in Figure 34.

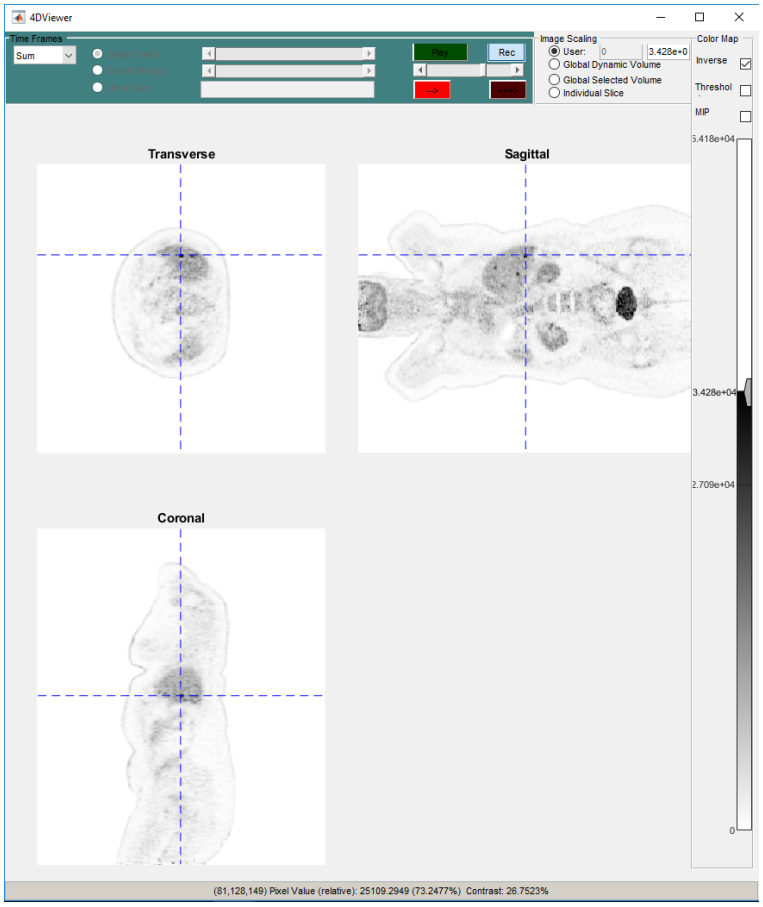

Figure 34: Screen capture of custom volumetric viewer with crosshairs on one of nine 10mm lesions in the volumetric image. The Cartesian coordinates corresponding to the location specified by the crosshairs is shown on the bottom of the figure. The observer elected to lower the image intensity upper threshold to improve lesion visualization in the liver (colorbar on right).

4.3.3 Results

Similar LOD were resolved for both inexperienced observers. Observer A and B correctly localized 11/36 and 12/36 synthetic lesions respectively, indicating no statistically significant difference ($p>0.8$). Results are summarized in Table 2 and Table 3 by lesion size and lesion-to-background respectively.

Table 2: Summary of preliminary perceptual trial outlining the number of synthetic lesions found in each variant as a function of lesion size

	Observer A	**Observer B**
2.5 mm lesion	0/9	0/9
5.0 mm lesion	0/9	0/9
7.5 mm lesion	5/9	6/9
10.0 mm lesion	6/9	6/9
Total Found	11/36	12/36

Observer A was able to discern lesions with diameter ≥ 7.5 mm and contrast ≥ 1:2.25 . Similarly, observer B was able to discern lesions with diameter ≥ 7.5 mm and contrast ≥ 1:2. Both observers were not able to successfully detect lesions ≤ 5.0 mm at any simulated contrast level.

Table 3: Summary of preliminary perceptual trial outlining the number of synthetic lesions found in each variant as a function of lesion size

(Above Background)	**Observer A**	**Observer B**
1:1.25	0/4	0/4
1:1.50	0/4	0/4
1:1.75	0/4	0/4
1:2.00	1/4	2/4
1:2.25	2/4	2/4
1:2.50	2/4	2/4
1:2.75	2/4	2/4
1:3.00	2/4	2/4

1:3.25	2/4	2/4
Total Found	11/36	12/36

4.3.4 Discussion

This is a preliminary study to estimate a human observer's LOD for liver lesions with known characteristics in clinical PET data. Although we only investigated responses of two novice observers, their results agreed closely with each other, strengthening our confidence in these results. Our results found that observers were not able to discern lesions that were 5.0 mm diameter or smaller in size regardless of their contrast with respects to the limits employed in this study. We hypothesize that these results reflect the spatial resolution of PET (~ mm FWHM) and post image reconstruction smoothing (~7 mm FWHM final image resolution) [47] and that brighter lesions (e.g. 1:5 lesion-to-background contrast) may still be discernible at these sizes.

From our lesion synthesis validation studies (Figure 26) we found that spheres with a diameter of 5.0 mm or less were discernable with the resulting image underestimating the true activity concentration by only 20%. Hence, we hypothesize that due to the introduction of patient emission and CT image related noise in the pre-reconstruction data, error is propagated in to the reconstructed image that renders lesions ≤ 5.0 mm more difficult to discern over the baseline noise level in the patient data.

While, higher intensity lesions are expected to become perceivable, this small study was not able to elucidate at what level. Furthermore, observers were able to localize lesions with diameters of 7.5 mm and greater, but it is unclear where the true LOD lies between 5.0 mm and 7.5 mm diameter lesions and is to be addressed in subsequent studies with a finer lesion size step size.

Finally, we did not evaluate the LOD using a clinical image reading workstation. These systems are designed with more functionality, advanced image-rendering capabilities and run on more advanced image display hardware that may contribute to improved lesion detection. Nevertheless, we opted for a simplified viewer that can be tailored to provide commonly used functionality to enable perception of inconspicuous lesions while removing other functionality that could distract the observer from the task and/or waste time.

4.3.5 Conclusion

The LOD in the liver were evaluated to be approximately 7.5 mm diameter with a minimum background to contrast ratios ranging 1:2.0-1:3.25. Subsequent studies of lesions larger than 7.5 mm should reduce the parameter step sizes within a more restricted range to achieve more precise LOD estimates. Studies looking to evaluate perception of smaller lesions should consist of higher lesion-to-background contrast. The number of test images per variant should also be increased to improve confidence in these results.

4.4 Second Perception Study

The preliminary perception study estimated the LOD for sub-centimeter lesions with background to contrast ratios ≤ 1:3.25, and therefore did not characterize the LOD for small lesions, which were not perceivable at these contrast levels. Therefore, we sought to design a follow-up perception study evaluating the LOD for small liver lesions using higher background to contrast ratios.

4.4.1 Research and Ethics Board Application

This work was part of a research and ethics board application approved by The Ottawa Health Science Network Research Ethics Board (OHSN-REB) concerning perception study participants in the proposed study (REB # 20180722-01H).

4.4.1.1 Database of Patients with Synthetic Lesions

Synthetic liver lesions, varied in size and lesion-to-background ratio, were generated in four candidate PET studies. Lesions were generated at a resolution of 256×256 corresponding to a 700 mm trans-axial FOV. A single lesion was manually placed within the liver. Lesions diameters varied from 2.5 mm to 10.0 mm in increments of 2.5 mm and lesion-to-background ratios varied from 1:2.0 to 1:6.0 in increments of 0.5, resulting in a total of 36 lesion variants.

Images were reconstructed using the clinical image reconstruction parameters with 6.4 mm full-width at half-maximum (FWHM) Gaussian smoothing applied post-reconstruction. Anonymized patient PET data with synthetic lesions were uploaded to the clinical workstation at the Department (Hermes Medical Solutions, Stockholm, Sweden).

4.4.2 Perception Experiment

Volumetric PET images with synthetic lesions (without corresponding CT) were shown using clinical workstation within the Department of Nuclear Medicine (Hybrid Viewer, Hermes Medical Solutions) and observers were asked to identify the center most voxel of a target lesion within the liver using the subject viewer (labelled TCS-PET only). Once the observer was satisfied with the response, the study administer selected the confirmation viewer tab (labelled TCS) to confirm lesion localization on the reregistered activity distribution map generated during lesion synthesis.

An example PET study with a synthetic liver lesion displayed using the clinical workstation configured as the subject viewer, is shown in Figure 35. A fused image of the PET and the lesion mask (reference truth) is shown in Figure 36 (confirmation viewer), which can be used to visually evaluate accurate lesion detection (bottom set of images). In both images the study subject successfully placed the crosshairs on a 7.5 mm diameter synthetic lesion in the liver.

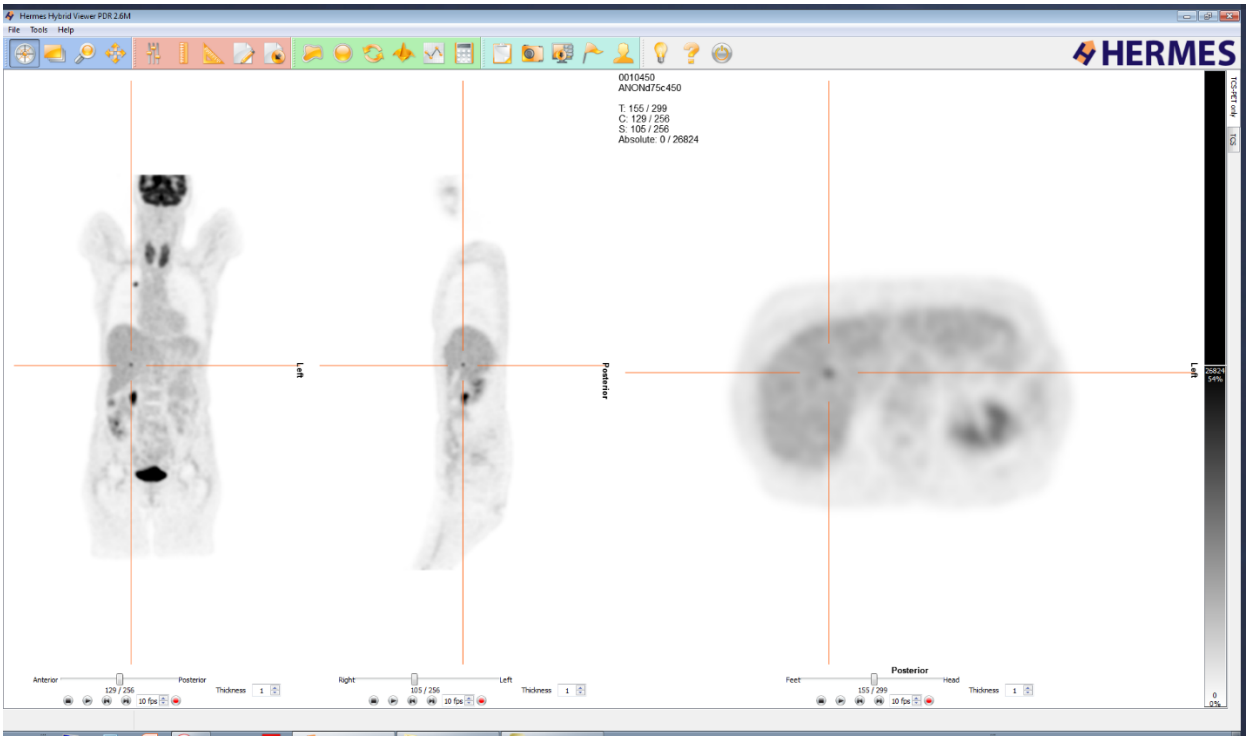

Figure 35: Screen capture from clinical workstation – subject viewer (Hybrid Viewer, Hermes Medical Solutions). The images correspond to the PET image with the crosshairs on a 7.5 mm diameter synthetic lesion in the liver.

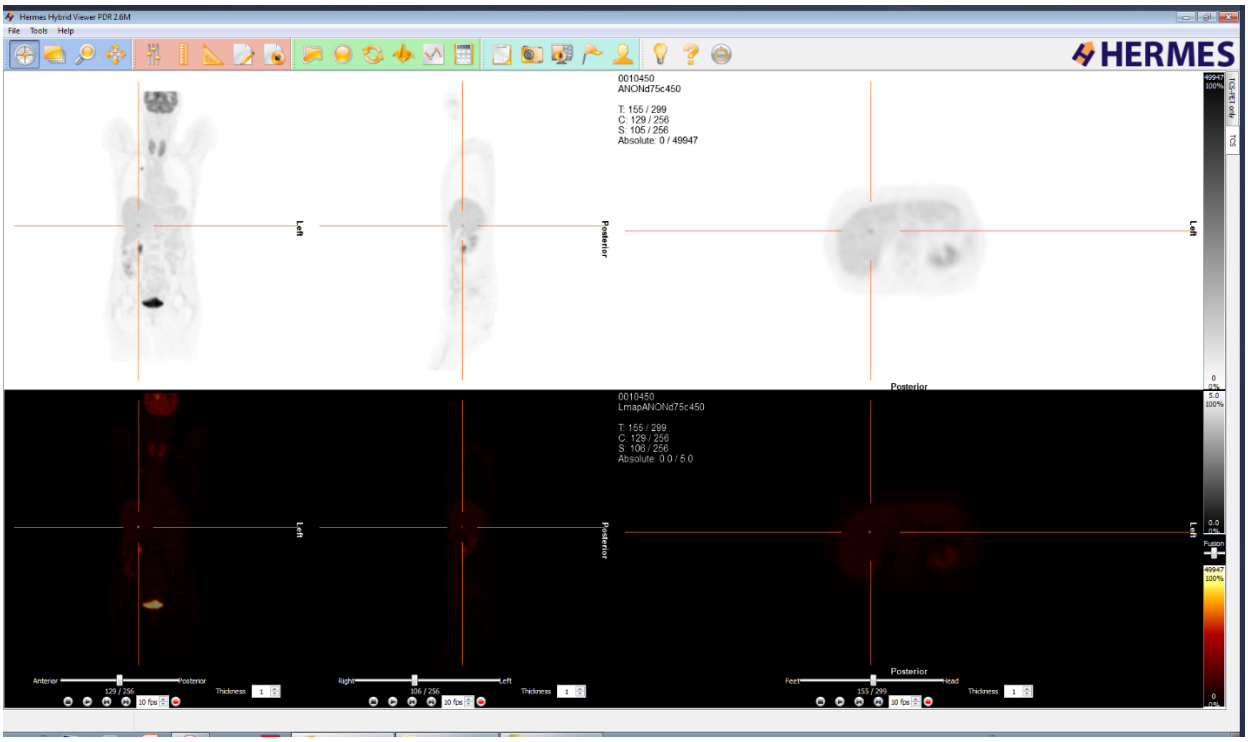

Figure 36: Screen capture from clinical workstation – confirmation viewer (Hybrid Viewer, Hermes Medical Solutions). The top row corresponds to the PET image with the crosshairs on a 7.5 mm diameter synthetic lesion in the liver. The bottom row corresponds to the activity distribution map generated during lesion synthesis.

4.5 Results

Table 4: Summary results of the second perceptual trial outlining the minimum background to contrast ratio of synthetic lesions found in each variant as a function of lesion size

Lesion Diameter (mm)	Observer A – Min. Contrast Ratio Detected	Observer B – Min. Contrast Ratio Detected
2.5	1:5.5	1:5.0
5.0	1:4.5	1:4.5
7.5	1:2.5	1:2.0
10.0	1:2.0	1:2.0

Similar LOD were resolved for both inexperienced observers. Observer A and B found 23/36 and 25/36 synthetic lesions respectively, indicating no statistically significant difference ($p>0.62$). Results are summarized in Table 4 by lesion size and lesion-to-background. Visual representation of the interplay between lesion detectability, lesion-to-background ratio and lesion size is illustrated in Figure 37.

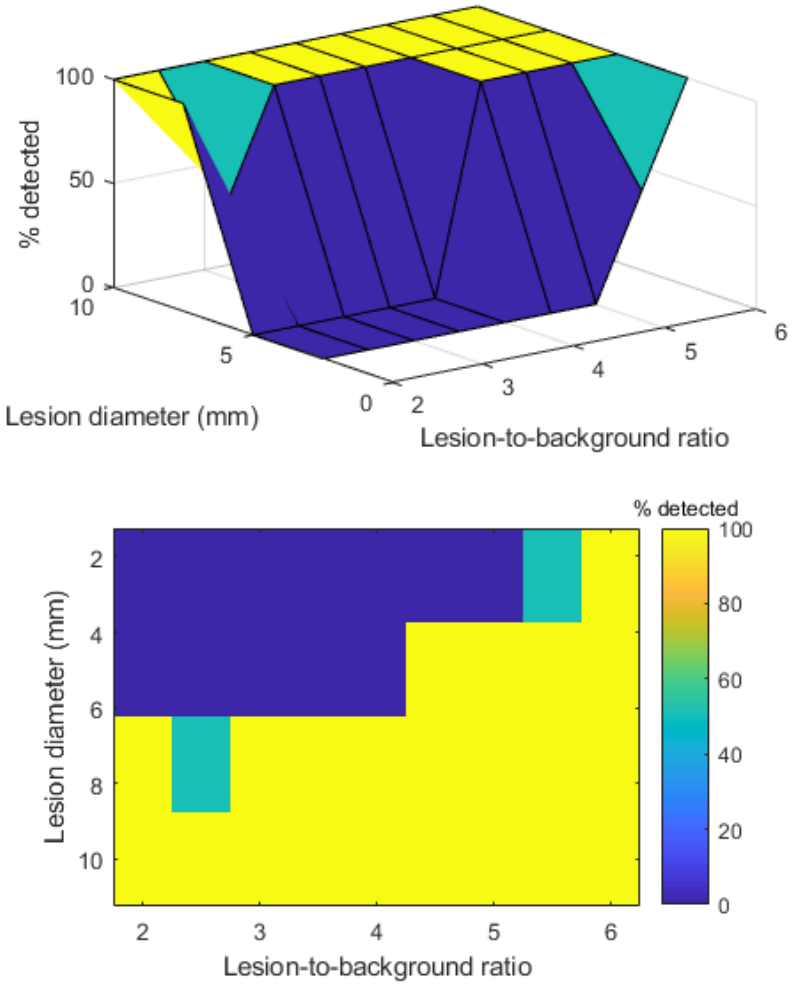

Figure 37: Summary of secondary lesion localization study results. 3D visualization of the interplay between lesion diameter, lesion-to-background ratio and the detectability of lesions as a percentage

(top). 2D representation of the interplay between lesion diameter, lesion-to-background ratio and detectability as a percentage (bottom)

4.6 Discussion

In the second study to estimate a human observers' LOD, we included higher intensity small liver synthetic lesions, overcoming a clear limitation of the preliminary study. Although we only investigated responses of two inexperienced observers, their results agreed closely with each other, further strengthening our confidence in these results. Furthermore, for larger lesions (≥ 7.5 mm diameter), the results of this study agreed closely with those of the preliminary study with regards to LOD.

Our results found that observers were indeed able to discern lesions that were smaller than the image spatial resolution (≤ 5.0 mm diameter in size) with sufficient background-to-contrast, justifying generating a database with synthetic lesions over a large range of lesion sizes and intensities.

Previous work has characterized the quantitative imaging accuracy of image reconstruction techniques for sub-centimeter lesions, however it did not evaluate the limits of detectability of these lesions by human observers [48] Other studies have also evaluated the detectability of lesions with respect to image generation techniques [29], [49], [50], but these were limited to lesions ≥10 mm diameter. To our knowledge our work demonstrates the first study to evaluate the LODs for sub-centimeter lesions, which are more relevant for detection of early cancer and small metastases – which are of paramount clinical significance.

These experiments clearly demonstrated the ease by which well characterized lesions can be embedded in raw patient PET data prior to image reconstruction to generate a

clinically realistic lesion dataset with a reliable reference truth that was not previously readily available. Therefore, these results highlight the incremental insight from synthetic lesions over numerical phantoms alone for determining clinically representative LODs.

4.7 Summary of perception studies

The results of these early perception studies clearly demonstrate the interplay between lesion size and intensities with regards to detectability. Lesions on the order of the image spatial resolution and smaller quickly become more difficult to discern above the image noise level requiring greater contrast with the background to be detectable. These finding clearly demonstrate the need to characterize LOD as a function of both size and intensity as a bare minimum. Other metrics such as image noise, background intensity, and anatomical location may also prove to be important factors.

During observer evaluation, we noted the time required to localize a lesion was dependent on size and contrast. The time elapsed during localization may be a useful surrogate metric for perceptibility and/or observer confidence to consider in studies comparing between alternative image generation/viewing techniques. We hypothesize that we can reduce the time required to localize a lesion by providing the observer with a single two-dimensional slice transecting the target lesion, as opposed to forcing the observer to locate lesion in a volume image by interactive scrolling through image slices, mitigating observer fatigue. By decreasing response times, the number of datapoint that can be measured in a single study session can be increased for greater precision and statistical power.

Contrary to the original CD diagram study design (Figure 10), the test pattern in these studies was not fixed and therefore observers did not know where in the image space to expect the synthetic lesion. While in these studies observers were tasked with finding spherical lesions, subsequent studies could also consider including asymmetric, heterogonous lesions. However, the advantage of spherical lesions is that they can be simply characterized and localization of target lesions relative to the reference truth can be accurately assesses.

When evaluating an individual's LOD the intervals between lesion sizes and activities should be of high fidelity (small intervals) to achieve a precise LOD metrics. This work did not evaluate the variability LOD within different anatomical regions (i.e. the lungs or breast). However, we hypothesize that anatomical regions (i.e. lungs or liver) would have varying physiological uptake of the tracer. For example, a lesion with given size and contrast in the liver may be more conspicuous than the same lesion in the lung as there are more event counts (and less noise) and more homogenous activity distribution in the liver region. Additionally, future studies should probe lesion detection in multiple anatomical regions to achieve a more complete understanding of LOD in a clinical context.

Clinical image interpretation should ideally be performed in a controlled environment consisting of specialized calibrated diagnostic medical image displays and darkened lighting. The tools available on the clinical workstations however can distract observers and possibly decrease throughput as they encompass a full suite of clinical functionality not necessary when conducting perception studies. When working with clinical workstations useful metrics such as the time required to localize the lesion are not

automatically recorded. In our preliminary studies, we found majority of the time required to localize a lesion was attributed to (1) loading the volumetric image onto the display, as PACs systems archive images over the hospital network and (2) scrolling through the entire image volume, searching for a lesion. These factors ultimately affect the throughput, where study administrators must choose between having sufficient sample size or reasonable study durations (< 1 hour). As such the two preliminary studies inspired optimization of the study design resulting in the development of a custom web-based Image Perception Suite.

4.8 Perception Study Deployment Tool

As clinical workstations may be locked down preventing the installation of additional software for security reasons, a custom image perception suite was developed as a web-based application hosted on a local server, allowing study participants to perform the study using the web browser on the clinical workstation.

The web application was developed using the MATLAB App Developer (R2017b) and hosted using the MATLAB Web App Server. The application functioned as a standalone application on the server machine.

During development we consulted with experienced nuclear medicine clinicians to discuss basic controls needed when localizing lesions in a clinical environment. Ultimately, it was decided that observers must have the ability to control the upper and lower color scale limits. Furthermore, we elected to use a predefined colormap as used clinically (inverse of the linear gray scale) as the effect of colormaps on lesion localization is unknown.

On launching the application, the user is a required to login with user credentials, which are used to authenticate the study participant, log their identity and to enforce participation according to the study protocol (e.g. number of experiment sessions) (see Figure 38). On the following screen the user is tasked with selecting the study from a list of studies and is provided with a set of instructions and/or reviewing the informed consent form. Figure 39 illustrates the instructions for the participating in the lesion localization study.

Figure 38: Screen capture of the developed Image Perception Suite at the log on screen .

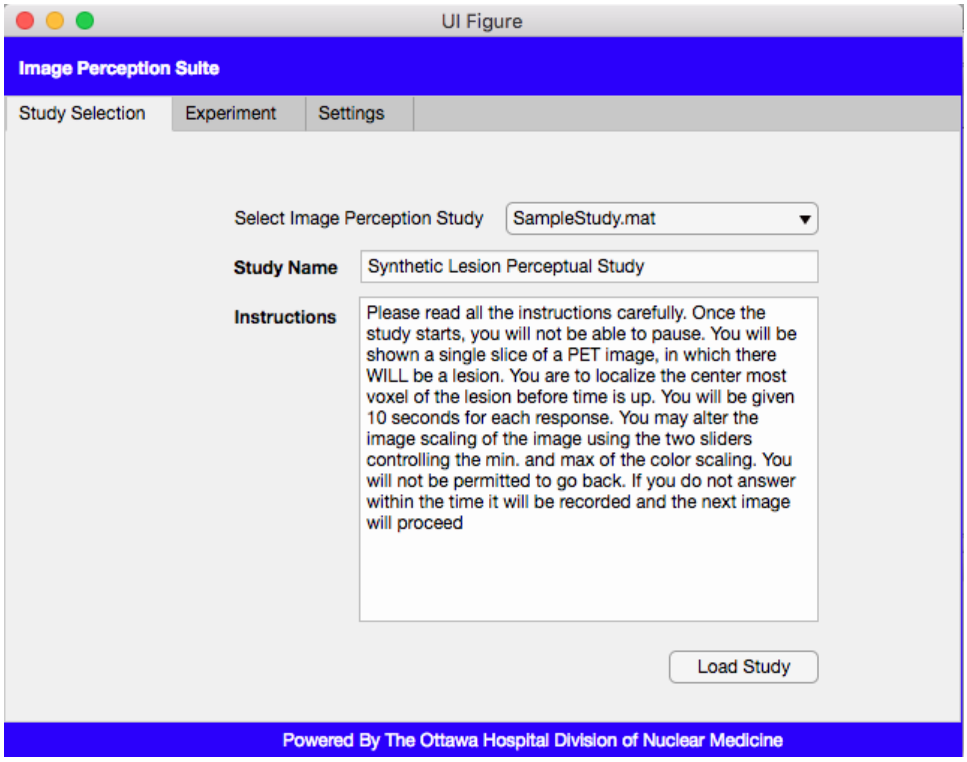

Figure 39: Screen capture of the Image Perception Suite software displaying the instruction for the selected perception study.

Once the study has been selected, a series of images are presented to the participants in automated order. In Figure 40, the suite displays a single PET transaxial slice containing the target lesion and records the coordinates where the observer has localized (by mouse click) the lesion. The Image Perception Suite limits response times to a pre-set time interval (count down timer at top of the figure). If time elapses, the lesion is interpreted as undetected by the observer. While the image is displayed the viewer can modify the image color scale limits using two sliders corresponding to lower and upper limits.

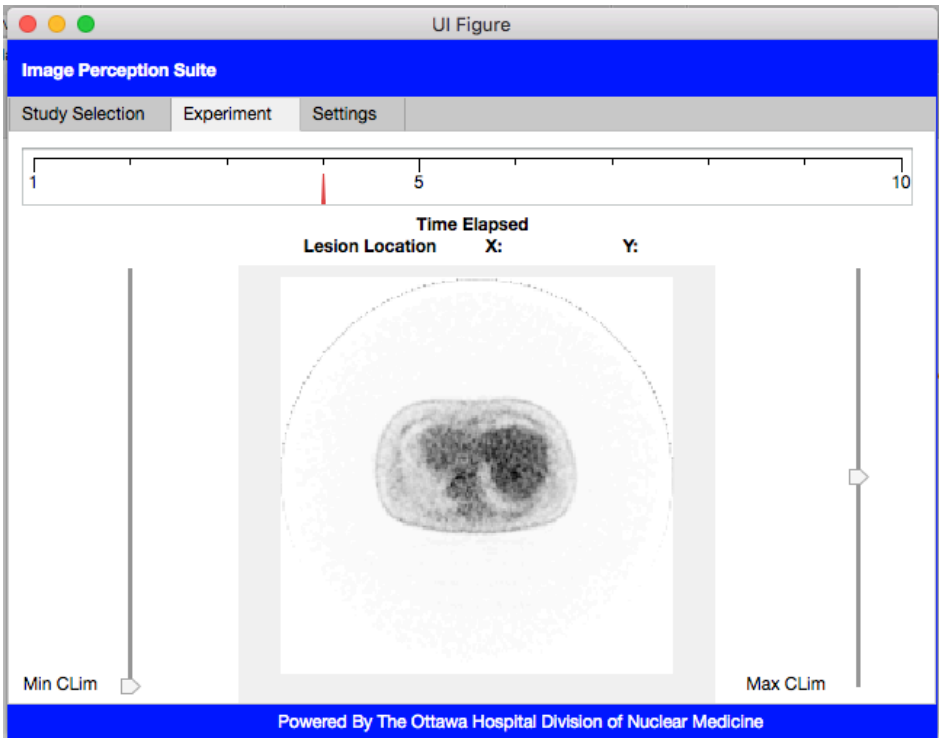

Figure 40: Screen capture of the custom developed Image Perception Suite showing a sample trans-axial slice displayed using the clinical color map settings.

The Image Perception Suite can be configured to display images in either a predefined order or in random order. The application programming interface was designed to accommodate image sequencing in an adaptive manner, based on previous responses. The study terminates when either all images have been displayed or when a pre-defined study time limit has been reached, depending on the study settings.

The following information is recorded throughout the study session for subsequent analysis:

- Image identifier (which can be linked to the image/lesion properties)
- Time to response (seconds)
- Indicated lesion location (image coordinates)

- Color scale limits

4.9 Summary

This work demonstrates the application of the Lesion Synthesis Toolbox for perception studies to characterize the LODs effectively. Preliminary perception studies provided insight towards (1) approximate LOD for PET liver lesions, (2) adequately defining a range of background-to-contrast ratios for synthetic lesions and (3) effectively conducting perception studies. The Image Perception Suite was developed to deploy large-scale perception studies within a clinical setting in an efficient manner and eliminates the need for supervision by a study administrator. These contributions along with the Lesion Synthesis Toolbox and access to clinical data enable practical implementation of large-scale perception studies in PET and PET/CT.

Chapter 5: Summary

Positron emission tomography (PET) is a highly sensitive non-invasive 3D medical imaging technique, used to quantify physiologic function by measuring the radioactive decay of synthesized radioactive biological compounds. Hybrid PET/CT has found many applications as a clinical diagnostic tool within the domain of oncology, routinely used for the diagnosis, staging, treatment planning and monitoring of cancer. However, the relatively low spatial resolution of resulting PET images, poses a source of influence in patient care/outcome as small or faint lesions can be underrepresented. Newer, more sophisticated PET image reconstruction techniques aim to suppress noise and improve the detectability of smaller and more faint lesions, potentially increasing the limits of human perception for the task-specific goal of lesion detection.

The advancement from simple analytical image reconstruction techniques (i.e. filtered back projection) to more advance iterative reconstructions (i.e. OSEM/BSREM) allow the employment of system modeling techniques with the goal of ultimately improving quantitative accuracy and image quality, which are typically evaluated using (1) physical phantoms such as the NEMA 2012 hot-sphere and Jaszczak phantom and (2) labelled patient image databases to evaluate the image quality and perception of known physiologic features. While these techniques may be intuitive, they each pose unique limitations. Physical phantoms are often overly simplified and do not model biological patterns, and patient images often lack reliable, marginally perceptible, ground truths [2], [3].

To address the medical imaging community desire for more precise evaluation of image generation techniques, the use of emission data simulation is increasingly being used to power more clinically relevant task-based performance metrics such, as lesion limits of detection (LOD).

However, there exists a disconnect between open source/community supported simulation techniques and proprietary image generation software used in clinical scenarios that is throttling implementation or synthetic emission data for broader application. The complexity of implementing simulation data is technically onerous requiring collaboration with PET vendors. Additionally, the use of simulation data is often critiqued for insufficient realism and variety compared to real patient data. Therefore, the evaluation of image generation techniques for the purposes of characterizing LOD in PET images is a complicated and unresolved problem that can be tackled using integrated software that can be distributed to a broader research community lacking specific technical expertise.

5.1 Contributions

Chapter 2 of this thesis provides background information on PET and the need for synthetic lesions and Chapter 3 describes the development of the Lesion Synthesis Toolbox capable of embedding synthetic lesion in raw PET data prior to image reconstruction and corresponding CT data. The chapter describes a custom developed tool comprising an integrated lesion synthesis workflow including the GE Discovery Data Retriever, and PET/CT viewer for defining lesions, lesion emission and detection simulation, combining of synthetic and patient data, image reconstruction, and

aggregating data to assemble a synthetic lesion database. The chapter explores implementation of two alternative lesion simulators: Monte Carlo, photon tracking (SimSET) and an analytical, forward projector. A detailed model of our clinical PET scanner (GE Discovery 710 PET/CT) was developed in SimSET and used to simulate synthetic emission from within the patient. The simulation data was successfully manipulated to conform with proprietary data formats to conform to clinically available image reconstruction software. The SimSET scanner model was validated for geometric consistency with the vendor reconstruction software, but ultimately the use of SimSET was abandoned. Our evaluation of SimSET demonstrated severe attenuation artifacts indicating inaccurate modeling of photon attenuation for reasons that are not completely clear, significant technical challenges for merging time-of-flight data and exceedingly long simulation execution times.

Subsequently the chapter describes an alternative, analytical approach for the synthesis of emission data employing the forward projector from the vendor provided image reconstruction algorithm. Following validation, the analytical approach was able to successfully model the scanner geometry, patient attenuation modeling, and to incorporate time-of-flight data modeling into raw patient PET data. Furthermore, simulation times were on the order of minutes on a modern desktop computer enabling production of many scripted lesion variants.

The chapter concludes by describing an image matrix substitution method for embedding synthetic lesion into the corresponding patient CT and demonstrating a full system integration test. By synthesizing lesion in both PET emission data and CT attenuation

data we are able to increase the clinical realism of lesions, which are typically viewed on both modalities together - to our knowledge this is a world first.

In Chapter 4 of this thesis we take preliminary steps towards applying synthetic lesions to investigate lesion perception in PET and to determine the limits of lesion detection. We describe a process to identify candidate patient data from a clinical PET/CT database in our clinic. We describe a preliminary perceptual trial used to estimate the LOD of liver lesions in terms of lesion size and contrast. The preliminary trial found the LOD to be approximately 7.5 mm diameter with a minimum contrast to background ratio ranging from 1:2.0-1:3.25 above background. The secondary perception study confirmed that small lesions (≤ 5.0mm) with high background to contrast ratios are discernable, where the LOD of both observers were very similar with 1:5.5, 1:4.5, 1:2.5, 1:2.0 contrast visible in lesions 2.5, 5.0, 7.5, 10.0 mm respectively. Based on our experience in these preliminary trials we draw conclusions to optimize future perception studies to minimize operator fatigue and obtain a more precise LOD. Chapter 4 concludes by building on the findings from the preliminary trial to design a subsequent LOD study using a custom developed image perception research software package. The perception research software runs as a web application accessible from any web browser and therefore can be run on active clinical workstations without the need to install software. Furthermore, the software is intended to be easily adaptable to run multiple, different image perception trials and manages input data images, contains image visualization and manipulation tools, and captures participant responses to a text file format for subsequent analysis by researchers.

The body of this work developed a vendor-supported toolbox for generating synthetic lesions within raw PET and reconstructed CT patient data for the purposes of evaluating image perception, and infrastructure for conducting perception research. Hence, we created the necessary infrastructure for studying the task-based performance of lesion detection and characterizing system-operator performance in terms of the limits of detectable lesions. We demonstrate usage of this tool by employing a small preliminary perception study to estimate the LOD of lesions within the liver. We also describe methods for subsequent follow up studies to obtain a more precise characterization of the LOD.

5.2 Limitations

This document primarily concerns with the development and implementation of new technologies and not so much with their demonstrated application, as such it has several limitations that must be addressed.

5.2.1 Realism of Synthetic Lesions

For the purposes of the preliminary perceptual study, spherical lesions with uniformly distributed contrast factors were used. A major limitation of this method to generate synthetic lesions is the realism when compared to real lesions seen in clinical scenarios. It is common that, lesions may present themselves with necrotic cores where intensity of the lesion increases radially outward, or lesions may be non-symmetrically shaped resembling “blobs” on reconstructed PET images. While this is a legitimate concern for some research questions, for the purposes of a baseline LOD we chose to simplify analysis and eliminate any shaped based detection bias by using perfect spheres. The

lesion synthesis toolbox developed however facilitates the generation of non-uniform shapes, which can be evaluated in subsequent studies after a baseline understanding of LOD has been established.

5.2.2 Modelling of Imaging Physics

The analytical approach used to synthetize lesions is limited in the physics it models. The approach does not model the associated scatter, random or dead time associated with the detection of emission data. Similarly, the employed simulation method does not model the scan duration and therefore cannot model low count rate acquisitions (and corresponding noise) associated with reduced scan times. One approach to mitigate this limitation to achieve a more complete simulation package is to use more advanced physics simulations such as Monte Carlo (e.g. SimSET). However, these Monte Carlo simulations require many more computations (and execution time) and their benefit to lesion realism remains unclear. We hypothesize that in relatively small and low intensity lesions, which produce relatively few detected events compared to the patient scan, these second order effects (i.e. scatter, randoms and deadtime) are negligible with regards to lesion realism and quantitative accuracy. Nevertheless, to conclusively address these concerns lesion realism perception studies must be conducted with expert viewers.

5.2.3 Other Applications

The premise of this work is that databases of patient PET images with synthetic lesions are applicable to a wide range of applications required to propel the field forward. An application that was repeatedly mentioned throughout was a task-based performance metric based on the limits of detectable (LOD) lesions as a benchmark tool to compare

between alternative technologies. LOD can be applied to objectively compare between image generation techniques (types of image reconstruction, reconstruction parameters and filters). Similarly, LOD can also evaluate the effect image viewing conditions (e.g. room light intensity), image-rendering conditions (e.g. fused PET/CT vs. side-by-side) have on the LOD, and image display technologies (e.g. flat panel display vs. virtual reality). These concepts can also be extended to compare between viewers to demonstrate the effectiveness of training and experience and to compare between human- and machine-observers.

Large libraries of synthetic lesions with an absolute ground truth can be generated relatively easily and used to train a machine-observer to detect lesions and to compare detection accuracy and LOD for trained machine observers. These (anonymized) image libraries can be made available to communities of researchers to enable data driven development of machine-observers, as is successfully demonstrated by the rapid advancement classification and segmentation applications in general images.

With user friendly GUI based workflows like the Lesion Synthesis Toolbox, researchers can synthesize datasets with synthetic lesions for any application without the need for a technical expertise in simulations or programming experience ultimately facilitating image perception research.

5.3 Conclusion

This thesis describes three methods to embedding lesion in patient data. The resulting work validates the development of the Lesion Synthesis Toolbox for embedding lesions into raw PET patient data prior to image reconstruction and the corresponding

reconstructed patient CT. The toolbox was used to estimate the LOD in a preliminary baseline perceptual study. The resulting work lays the foundations for subsequent studies to investigate a complete clinical characterization of the limits of detection towards improved image generation technologies, image display techniques and the use of artificial intelligence machine observers.

References

[1] E. J. Teoh, D. R. McGowan, R. E. Macpherson, K. M. Bradley, and F. V. Gleeson, "Phantom and Clinical Evaluation of the Bayesian Penalized Likelihood Reconstruction Algorithm Q.Clear on an LYSO PET/CT System," *J. Nucl. Med.*, vol. 56, no. 9, pp. 1447–1452, Sep. 2015.

[2] "A phantom design for assessment of detectability in PET imaging - Wollenweber - 2016 - Medical Physics - Wiley Online Library." [Online]. Available: https://aapm-onlinelibrary-wiley-com.proxy.library.carleton.ca/doi/full/10.1118/1.4960365. [Accessed: 16-Feb-2019].

[3] "Practical no-gold-standard evaluation framework for quantitative imaging methods: application to lesion segmentation in positron emission tomography." [Online]. Available: https://www.spiedigitallibrary.org/journals/Journal-of-Medical-Imaging/volume-4/issue-1/011011/Practical-no-gold-standard-evaluation-framework-for-quantitative-imaging-methods/10.1117/1.JMI.4.1.011011.full?casa_token=RSoWtLBp_8YAAAAA%3aT9UWycP9oQbgFpLlIwvPo88oYpRLQAJiHsPJsxeqZGircbL9Q8fDUKNPQUa-U3xVVDV0DnVEPynL&SSO=1. [Accessed: 16-Feb-2019].

[4] "(PDF) ASIM: An analytic PET simulator." [Online]. Available: https://www.researchgate.net/publication/288252691_ASIM_An_analytic_PET_simulator. [Accessed: 16-Feb-2019].

[5] J. Lantos, E. S. Mittra, C. S. Levin, and A. Iagaru, "Standard OSEM vs. regularized PET image reconstruction: qualitative and quantitative comparison using phantom data and various clinical radiopharmaceuticals," *Am. J. Nucl. Med. Mol. Imaging*, vol. 8, no. 2, pp. 110–118, Apr. 2018.

[6] A. Selva-O'Callaghan *et al.*, "Conventional Cancer Screening versus PET/CT in Dermatomyositis/Polymyositis," *Am. J. Med.*, vol. 123, no. 6, pp. 558–562, Jun. 2010.

[7] S. L. Pimlott and A. Sutherland, "Molecular tracers for the PET and SPECT imaging of disease," *Chem. Soc. Rev.*, vol. 40, no. 1, pp. 149–162, 2011.

[8] M. Schwaiger and H.-J. Wester, "How many PET tracers do we need?," *J. Nucl. Med. Off. Publ. Soc. Nucl. Med.*, vol. 52, no. Suppl, pp. 36S-41S, 2011.

[9] S. S. Gambhir, "Molecular imaging of cancer with positron emission tomography," *Nat. Rev. Cancer*, vol. 2, no. 9, pp. 683–693, Sep. 2002.

[10] R. E. Schmitz, A. M. Alessio, and P. E. Kinahan, "The Physics of PET/CT scanners," p. 16.

[11] D. Townsend, "Physical Principles and Technology of Clinical PET Imaging," vol. 33, no. 2, p. 13, 2004.

[12] "Positron emission tomography: An overview." [Online]. Available: https://www-ncbi-nlm-nih-gov.proxy.library.carleton.ca/pmc/articles/PMC3003889/. [Accessed: 16-Feb-2019].

[13] R. Klein, “Initial Steps to Tracer Kinetic Modeling and MBF Quantification,” *Ann. Nucl. Cardiol.*, 2018.
[14] “Positron Emission Tomography - Basic Sciences | Dale L. Bailey | Springer.” [Online]. Available: https://www.springer.com/gp/book/9781852334857. [Accessed: 16-Feb-2019].
[15] M. Lubberink, R. Boellaard, and A. A. Lammertsma, “Quantitative Comparison of Analytic and Iterative Reconstruction Methods in 2- and 3-Dimensional Dynamic Cardiac 18F-FDG PET,” p. 9.
[16] “The image quality of FBP and MLEM reconstruction - IOPscience.” [Online]. Available: https://iopscience.iop.org/article/10.1088/0031-9155/55/11/012/meta. [Accessed: 16-Feb-2019].
[17] H. M. Hudson and R. S. Larkin, “Accelerated image reconstruction using ordered subsets of projection data,” *IEEE Trans. Med. Imaging*, vol. 13, no. 4, pp. 601–609, 1994.
[18] “The Q.Clear PET reconstruction algorithm: Evaluation using the NEMA IQ phantom.” [Online]. Available: http://jnm.snmjournals.org/content/57/supplement_2/1973?related-urls=yes&legid=jnumed;57/supplement_2/1973. [Accessed: 08-Nov-2017].
[19] B. D’Alessandro *et al.*, “Synthetic Positron Emission Tomography-Computed Tomography Images for Use in Perceptual Studies,” *Semin. Nucl. Med.*, vol. 41, no. 6, pp. 437–448, Nov. 2011.
[20] M. B. Abbott, L. R. Furenlid, D. W. Wilson, G. D. Stevenson, J. M. Woolfenden, and H. H. Barrett, “An Implantable Synthetic SPECT Lesion: A Bridge from Phantom to Reality,” *J. Nucl. Med.*, vol. 48, no. 11, pp. 1796–1799, Nov. 2007.
[21] C.-P. Day, G. Merlino, and T. Van Dyke, “Preclinical Mouse Cancer Models: A Maze of Opportunities and Challenges,” *Cell*, vol. 163, no. 1, pp. 39–53, Sep. 2015.
[22] G. El Fakhri, S. Surti, C. M. Trott, J. Scheuermann, and J. S. Karp, “Improvement in Lesion Detection with Whole – Body Oncologic TOF - PET,” *J. Nucl. Med. Off. Publ. Soc. Nucl. Med.*, vol. 52, no. 3, pp. 347–353, Mar. 2011.
[23] S. D. Wollenweber, A. M. Alessio, and P. E. Kinahan, “A phantom design for assessment of detectability in PET imaging: A phantom design for PET detectability,” *Med. Phys.*, vol. 43, no. 9, pp. 5051–5062, Aug. 2016.
[24] “PETSTEP: Generation of synthetic PET lesions for fast evaluation of segmentation methods - ScienceDirect.” [Online]. Available: https://www-sciencedirect-com.proxy.library.carleton.ca/science/article/pii/S1120179715003130. [Accessed: 01-Aug-2018].
[25] E. A. Krupinski, “Current perspectives in medical image perception,” *Atten. Percept. Psychophys.*, vol. 72, no. 5, Jul. 2010.
[26] S. S. Yip and H. J. W. L. Aerts, “Applications and limitations of radiomics,” *Phys. Med. Biol.*, vol. 61, no. 13, pp. R150–R166, Jul. 2016.
[27] B. T. Vincent, “A tutorial on Bayesian models of perception,” *J. Math. Psychol.*, vol. 66, pp. 103–114, Jun. 2015.

[28] B. Berthon *et al.*, “PETSTEP: Generation of synthetic PET lesions for fast evaluation of segmentation methods,” *Phys. Medica PM Int. J. Devoted Appl. Phys. Med. Biol. Off. J. Ital. Assoc. Biomed. Phys. AIFB*, vol. 31, no. 8, pp. 969–980, Dec. 2015.
[29] K. A. Wangerin, S. Ahn, S. Wollenweber, S. G. Ross, P. E. Kinahan, and R. M. Manjeshwar, “Evaluation of lesion detectability in positron emission tomography when using a convergent penalized likelihood image reconstruction method,” *J. Med. Imaging*, vol. 4, no. 1, Jan. 2017.
[30] M. T. Madsen, K. S. Berbaum, A. N. Ellingson, B. H. Thompson, B. F. Mullan, and R. T. Caldwell, “A new software tool for removing, storing, and adding abnormalities to medical images for perception research studies,” *Acad. Radiol.*, vol. 13, no. 3, pp. 305–312, Mar. 2006.
[31] R. L. Harrison, S. D. Vannoy, D. R. Haynor, S. B. Gillispie, M. S. Kaplan, and T. K. Lewellen, “Preliminary Experience With The Photon History Generator Module Of A Public-domain Simulation System For Emission Tomography,” in *1993 IEEE Conference Record Nuclear Science Symposium and Medical Imaging Conference*, San Francisco, CA, USA, 1993, pp. 1154–1158.
[32] I. Buvat and I. Castiglioni, “Monte Carlo simulations in SPET and PET,” *Q. J. Nucl. Med.*, vol. 46, no. 1, pp. 48–61, Mar. 2002.
[33] R. ECKHARDT, “Stan Ulam, John von Neumann, and the Monte Carlo Method,” *Los Alamos Sci. Spec. Issue*, vol. 15, pp. 131–137, 1987.
[34] “GATE: a simulation toolkit for PET and SPECT - IOPscience.” [Online]. Available: https://iopscience.iop.org/article/10.1088/0031-9155/49/19/007/meta. [Accessed: 16-Feb-2019].
[35] “SimSET User Guide Home Page.” [Online]. Available: https://depts.washington.edu/simset/html/user_guide/user_guide_index.html. [Accessed: 09-Apr-2019].
[36] “SimSET User Guide Home Page.” [Online]. Available: https://depts.washington.edu/simset/html/user_guide/user_guide_index.html. [Accessed: 09-Apr-2019].
[37] “Acceleration of SimSET photon history generation - IEEE Conference Publication.” [Online]. Available: https://ieeexplore-ieee-org.proxy.library.carleton.ca/abstract/document/1239680. [Accessed: 16-Feb-2019].
[38] “Resampling estimates of precision in emission tomography - IEEE Journals & Magazine.” [Online]. Available: https://ieeexplore-ieee-org.proxy.library.carleton.ca/abstract/document/41486. [Accessed: 16-Feb-2019].
[39] A. E. Burgess, “Visual Perception Studies and Observer Models in Medical Imaging,” *Semin. Nucl. Med.*, vol. 41, no. 6, pp. 419–436, Nov. 2011.
[40] J. E. Gillam and M. Rafecas, “Monte-Carlo simulations and image reconstruction for novel imaging scenarios in emission tomography,” *Nucl. Instrum. Methods Phys. Res. Sect. Accel. Spectrometers Detect. Assoc. Equip.*, vol. 809, pp. 76–88, Feb. 2016.

[41] L. H. Garland, “On the scientific evaluation of diagnostic procedures,” *Radiology*, vol. 52, no. 3, pp. 309–328, Mar. 1949.
[42] A. Rose, “The Sensitivity Performance of the Human Eye on an Absolute Scale*,” *JOSA*, vol. 38, no. 2, pp. 196–208, Feb. 1948.
[43] C. E. Metz, “Basic principles of ROC analysis,” *Semin. Nucl. Med.*, vol. 8, no. 4, pp. 283–298, Oct. 1978.
[44] “View4D - File Exchange - MATLAB Central.” [Online]. Available: https://www.mathworks.com/matlabcentral/fileexchange/44600-view4d. [Accessed: 04-Apr-2019].
[45] Croteau, E, R. Klein, J. M. Renaud, M. Premaratne, and R. A. Dekemp, “Chapter 2: Cardiovascular PET-CT,” in *Hybrid Imaging in Cardiovascular Medicine*, 1st ed., .
[46] R. L. Wahl, H. Jacene, Y. Kasamon, and M. A. Lodge, “From RECIST to PERCIST: Evolving Considerations for PET Response Criteria in Solid Tumors,” *J. Nucl. Med.*, vol. 50, no. Suppl_1, pp. 122S-150S, May 2009.
[47] H. J. Yoon, Y. J. Jeong, H. J. Son, D.-Y. Kang, K.-Y. Hyun, and M.-K. Lee, “Optimization of the spatial resolution for the GE discovery PET/CT 710 by using NEMA NU 2-2007 standards,” *J. Korean Phys. Soc.*, vol. 66, no. 2, pp. 287–294, Jan. 2015.
[48] “Point-spread function reconstructed PET images of sub-centimeter lesions are not quantitative | SpringerLink.” [Online]. Available: http://link.springer.com.proxy.library.carleton.ca/article/10.1186/s40658-016-0169-9. [Accessed: 02-Feb-2017].
[49] J. Qi, “Comparison of Lesion Detection and Quantification in MAP Reconstruction with Gaussian and Non-Gaussian Priors,” *International Journal of Biomedical Imaging*, 2006. [Online]. Available: https://www.hindawi.com/journals/ijbi/2006/087567/abs/. [Accessed: 08-Apr-2019].
[50] J. Nuyts *et al.*, “Performance of MAP Reconstruction for Hot Lesion Detection in Whole-Body PET/CT: An Evaluation With Human and Numerical Observers,” *IEEE Trans. Med. Imaging*, vol. 28, no. 1, pp. 67–73, Jan. 2009.