Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence (IJCAI-25)
Special Track on AI and Social Good

Direct Estimation of Attenuation Information from Sinograms for Positron Emission Tomography Reconstruction

Prabath Hetti Mudiyanselage[1] , **Ruwan Tennakoon**[1] , **John Thangarajah**[1] , **Robert Ware**[2] and **Jason Callahan**[3]
[1]School of Computing Technologies, RMIT University, Australia
[2]Cyclotek (Aust) Pty Ltd., Australia
[3]Melbourne Theranostic Innovation Centre, Australia
s3929442@student.rmit.edu.au, {ruwan.tennakoon, john.thangarajah}@rmit.edu.au, rob.ware@me.com, jcallahan@mtic.net.au

Abstract

Positron Emission Tomography (PET) is a powerful imaging modality for assessing biochemical processes within the body. However, accurate image reconstruction is challenged by photon attenuation, particularly in dense structures such as bones, leading to quantification errors and reduced diagnostic confidence. Computed Tomography (CT) based attenuation correction is the standard approach but introduces additional radiation exposure, longer imaging times, and patient inconvenience, as well as potential registration errors, motion artifacts, and energy scaling inaccuracies. In this study, we propose a 3D U-Net based deep learning framework that directly estimates attenuation information from PET sinograms, eliminating the need for additional imaging modalities. Our approach integrates PET physics and employs custom skip connections to enhance cross-domain learning. We evaluate our model on a simulated brain dataset derived from real patient templates, achieving a Dice coefficient of 0.650 and an accuracy of 0.486 for bone structures. The clinical applicability of our method is further assessed by reconstructing PET images with the generated attenuation maps, yielding an MSE of 0.007 and an SSIM of 0.956, demonstrating strong structural consistency with CT-based attenuation correction. These results highlight the feasibility of performing PET image attenuation correction using PET sinograms alone, offering a promising alternative that reduces imaging time, radiation exposure, and patient burden while enabling faster and more efficient PET reconstruction.

1 Introduction

Positron Emission Tomography (PET) imaging is a functional imaging technique widely used for diagnosing, staging, treatment planning, and follow-up in oncology, cardiology, and neurology. Unlike Computed Tomography (CT) and Magnetic Resonance Imaging (MRI), which primarily provide structural information, PET focuses on assessing biological process functionality.

In PET imaging, a compound called a radiotracer, which consists of a radioactive isotope and a targeting molecule, is injected intravenously into the patient. The radiotracer accumulates in targeted regions and undergoes decay, emitting positrons. When a positron encounters a nearby electron, they annihilate each other, producing two high-energy gamma photons that travel in opposite directions at a 180° angle. The PET scanner detects incoming photons, recording an event only if two photons are recorded in detectors opposed at 180° from each other within a short time window. The recorded events are then converted into a *sinogram*, where the events are organized to represent the body's projections at various angles. Sinograms are converted into 3D PET images through image reconstruction, which employs methods such as filtered back-projection (FBP) and iterative reconstruction.

Some high-energy gamma photons emitted after positron decay may not reach the detectors because they are absorbed by dense bodily structures. This absorption decreases the number of photons detected compared to the actual emission (i.e., activity), a phenomenon known as attenuation. As a result, *the events recorded by a PET scanner is a composition of both the spatial distributions of radioactivity concentrations (λ, activity map) and linear attenuation coefficients (μ, attenuation map)*. Attenuation results in increased noise, distortion, and artifacts in PET images; therefore, correcting for attenuation is essential during the PET reconstruction process to ensure accurate and reliable imaging results.

Since attenuation is closely related to anatomical structures, combined PET/CT imaging systems are commonly used in clinical practice, with CT primarily assisting in attenuation correction. However, additional CT scans in PET/CT increase radiation exposure, extend imaging time, and contribute to patient discomfort. Even with advanced low-dose CT technology, patients still face additional radiation exposure, which also raises ethical considerations for vulnerable populations such as children and repeat patients [Xia *et al.*, 2012; Quinn *et al.*, 2016]. In medical research, studies often involve both patients and healthy volunteers, with participants typically undergoing multiple PET scans following a single radiotracer injection and repositioning the bed between ses-

sions. This requires multiple CT scans solely for attenuation correction, resulting in significant and avoidable radiation exposure. As an alternative, PET/MRI systems offer reduced radiation but come with challenges such as higher costs, increased complexity, and longer scan durations, which may cause claustrophobia in some patients. Additionally, MRI is incompatible with certain metallic implants, further limiting its applicability.

In addition to the patient care limitations mentioned above, the use of separate CT (or MRI) imaging sessions introduces several technical challenges, including registration issues, motion artifacts, classification errors, and scaling discrepancies. In practice, a single CT image is taken with the patient holding their breath at the start of the PET scan session, while the PET scan is performed with normal breathing over a longer period, especially during dynamic PET imaging, which observes the spatiotemporal distribution of radiotracers over an extended time. This discrepancy makes it difficult to correct for respiratory (mid-breath) or cardiac (mid-cardiac) motion. Additionally, registering the image with respect to a particular organ, such as the heart, can result in misregistration of other organs or structures, such as the liver, lungs, skeletal muscle, or bones. There are also registration issues due to rigid body motion during the scan, such as movement of the head, especially during full body PET imaging. The scanner's bed position movement between CT and PET imaging also contributes to this problem.

These limitations can be overcome if attenuation correction maps (μ-maps) are generated directly from the PET raw data or sinograms, eliminating the need for additional CT (or MRI) scans. Such an approach is feasible because the activity of specific 3D locations experiences varying attenuation depending on the gamma-ray emission direction, information that is inherently captured in the raw data and sinograms. In this study, we hypothesized that *a deep learning framework can be used to directly estimate the corresponding μ-maps from PET sinogram data*. Our experiments reveal that a conventional U-Net, widely used in medical imaging, is insufficient for generating μ-maps from PET sinogram data due to the inherent mismatch between input and output representations. To address this challenge, we draw inspiration from physical models of 3D PET image reconstruction and introduce skip connections[1] that explicitly guide the U-Net in capturing structural information from sinogram data. To validate the effectiveness of our approach, we perform a quantitative analysis comparing the generated pseudo-μ-maps with CT-based μ-maps and evaluate attenuation-corrected PET images reconstructed using pseudo-μ-maps against those corrected with CT-based μ-maps.

There have been some AI based efforts to generate pseudo-CT images from reconstructed PET images, which can be later thresholded to obtain μ-maps [Hashimoto *et al.*, 2021; Partin *et al.*, 2024; Dong *et al.*, 2019; Armanious *et al.*, 2020]. These studies typically use non attenuation corrected (NAC) PET images as input to generate synthetic CT images. PET image reconstruction which consists of a filtering step, may discard potentially useful information corresponding to CT images as noise. Additionally, the PET reconstruction process is time-consuming [Ma *et al.*, 2022], which can significantly extend the duration of dynamic and back-to-back imaging sessions.

Our research is a collaborative effort between academic and industrial stakeholders, with experts in artificial intelligence (AI) research from RMIT University who developed AI models and specialists in medical and radiology research from Cyclotek (Aust) Pty Ltd and the Melbourne Theranostic Innovation Centre (MTIC) who defined the problem and conducted the evaluation of the results. This interdisciplinary collaboration integrates AI techniques with clinical and research expertise in PET imaging. At present, the PET imaging center at MTIC relies on CT scans for attenuation correction in PET imaging. We aim to develop deep learning models that reduce reliance on CT imaging during PET scanning and gradually implement these advancements within the center. Following this study, we aim to evaluate the techniques developed here using real patient data from the center. As part of the next phase of this ongoing project, we plan to extend our models to process full-body PET imaging data acquired using the Siemens Quadra Vision PET scanner at MTIC. Ultimately, our research seeks to reduce radiation exposure during PET/CT scanning sessions while also addressing challenges such as motion artifacts. We plan to collaborate with more volunteers at MTIC under reduced radiation conditions to further advance PET-based medical research. Our initiative has the potential to enhance the impact of PET imaging in clinical practice, improve patient safety, and contribute to life-saving advancements in medical imaging, which aligns with the United Nations Sustainable Development Goal (SDG) of good health and well-being (SDG 3).

2 Background and Related Work

2.1 PET Data Acquisitions and Sinogram

In PET imaging, a biologically functional molecule chemically combined with a radioactive atom, known as a radiotracer [Foster *et al.*, 2014], is injected intravenously into the patient. For example, 18-F Fluorodeoxyglucose (18-F FDG) is a commonly used radiotracer that targets glucose metabolism. During the PET imaging session, the radioisotope decays to a stable state by emitting a positron over time. When an emitted positron interacts with a nearby electron within a short distance (0.35 mm for positrons emitted by an 18-F atom), two 511 keV high energy annihilation photons are produced as the positron and electron annihilate [Pohost and Dilsizian, 2019]. These photons travel in opposite directions at a 180° angle and, if not absorbed by surrounding structures, are detected by two photodetectors in the detector ring of the PET scanner. If two photons are detected within a short time window, typically within 10 ns, they are considered to originate from the same annihilation event and are classified as a true coincidence event [Vaquero and Kinahan, 2015]. The sinogram is a data representation that organizes raw coincidence events into a structured format, facilitating effective image reconstruction. Each entry in the sinogram corresponds to a specific line of response (LOR), defined by the angle of projection ϕ, and the distance from the center of

[1] i.e., skipping layers in the neural network

the scanner, s. As the PET scanner acquires data from multiple angles around the subject, the detected events are sorted into the sinogram based on their geometric parameters. This results in a two-dimensional matrix where rows represent different projection angles, and columns correspond to specific LORs at those angles.

2.2 AI for PET Attenuation Correction

The current clinical practice for obtaining attenuation maps involves thresholding CT images. Several AI based approaches have been proposed to generate pseudo-CT images from PET images, extracting anatomical structural information through an additional thresholding step. These methods typically use nonattenuation corrected (NAC) reconstructed PET images as the input. In some studies [Dong *et al.*, 2019; Armanious *et al.*, 2020], generative adversarial network (GAN) based frameworks are employed to generate synthetic CT images. In contrast, other studies [Hashimoto *et al.*, 2021; Liu *et al.*, 2018] utilize 2D U-Net and 3D U-Net based architectures [Partin *et al.*, 2024]. All of these approaches struggle to generalize across different regions of the body, scanner models, and pathological conditions due to their dependence on NAC PET images for training [McMillan and Bradshaw, 2021]. In another study [Arabi and Zaidi, 2020], time-of-flight (TOF) PET emission data was used to attempt attenuation correction directly in the sinogram domain using HighResNet architecture. However, TOF image acquisition support is not available in all PET scanners.

PET image reconstruction from sinograms is performed using algorithmic or iterative approaches, such as Ordered Subset Expectation Maximization (OSEM). Recently, several studies have explored direct PET image reconstruction from PET sinograms using deep learning models. In those works, the focus is on improving the computational efficiency of the reconstruction stage and they still rely on CT scans for attenuation correction information. The DeepPET [Häggström *et al.*, 2019] model and another study [Ma *et al.*, 2022] utilizes 2D autoencoder architectures, while ReconU-Net implements another variation of the 2D U-Net [Hashimoto and Ote, 2024] with custom skip connections. It has been demonstrated that AI models, such as DeepPET, can reduce the time required for typical PET reconstruction algorithms from 360 seconds to just 14 seconds [Ma *et al.*, 2022].

3 Proposed Method

In this study, our objective is to synthesize a pseudo-μ-map for a patient, representing air, soft tissue, and bone structures for attenuation correction, using PET sinogram data from the same imaging session as input. An overview of the proposed approach is in Figure 1.

3.1 Physics Inspired 3D U-Net Architecture

The input to our model is a 2D sinogram, $\mathbf{S}_{\phi,d,p} \in \mathbb{R}^{\Phi \times D \times P}$ where Φ represents the number of azimuthal angle bins, D corresponds to the number of transaxial distance bins, and P indicates the number of slices along the depth axis. The model output is a 3D pseudo-μ-map, μ, where each voxel, $\mu_{x,y,z}$, is classified as either air, soft tissue, or bone. A simple

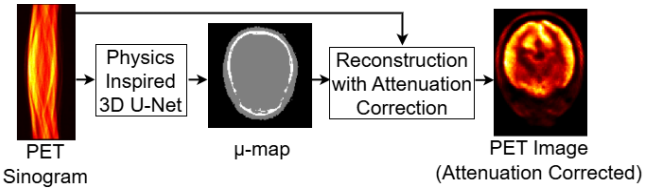

Figure 1: Overview of our proposed method to generate attenuation corrected PET images using PET sinograms alone. The corresponding μ-map is generated by our model without requiring additional CT images.

approach would be to employ a conventional U-Net, which is widely used in medical imaging, to transform the sinogram into a μ-map while training the model using a supervised loss function such as cross-entropy. However, our experiments indicate that this approach performs poorly (see Section 4 for details). The primary challenge lies in the difference in representation between the sinograms and μ-maps — each plane in sinograms is represented in terms of line-of-response (LOR) angles and distances from the center, whereas μ-maps exist in a spatial domain (height, width, depth). The relationship between μ-map, spatial distributions of radioactivity concentrations (λ, activity map), and sinogram is given by the Beer-Lambert law represented by [Kinahan *et al.*, 2003]:

$$\mathbf{S}(\phi,d,p) = \underbrace{\exp\left\{-\int \mu\left(\mathbf{x}_{\phi,d,p}(r)\right) dr\right\}}_{\text{Attenuation}} \times \underbrace{\int \lambda\left(\mathbf{x}_{\phi,d,p}(r)\right) dr}_{\text{Nonattenuated Sinogram}} \tag{1}$$

$$\mathbf{S}(\phi,d,p) = \exp\left\{-\mathcal{R}(\mu)\right\} \cdot \mathcal{R}(\lambda). \tag{2}$$

Here, $\mathbf{x}_{\phi,d,p}(r)$ parameterize the points along an LOR denoted by ϕ, d on plane p, and $\mathcal{R}(\cdot)$ is the Radon transform. As a result, a conventional model must not only learn to predict attenuation but also learn to reconstruct structural information from the sinogram (i.e., do the inverse Radon transform), adding significant complexity to the learning process.

To enhance the U-Net's ability to reconstruct structural information, we incorporate physical constraints of PET image reconstruction — specifically, inverse Radon transforms — directly into the skip connections.

Our custom skip connections consist of three main components, as shown in Equation 3. For each 3D U-Net level, l, we filter the k^{th} individual feature map, $\mathbf{F}_k^{(l)}$ using a ramp filter, denoted by $\mathcal{H}(\cdot)$, and then apply an inverse Radon transform, $\mathcal{R}^{-1}(\cdot)$, to obtain a feature map in the reconstructed spatial domain. These two operations together form the Filtered Back Projection (FBP) process [Defrise *et al.*, 2005]. Then, we upsample ($\mathcal{U}(\cdot)$) the resultant image by a factor of 1.8 and crop it to obtain the final spatial domain feature map, $\hat{\mathbf{F}}_k^{(l)}$.

$$\hat{\mathbf{F}}_k^{(l)} = \mathcal{U}\left(\mathcal{R}^{-1}\left(\mathcal{H}\left(\mathbf{F}_k^{(l)}\right)\right)\right) \tag{3}$$

The overall architecture of the proposed U-Net is illustrated in Figure 2. Our U-Net is inspired by 3D U-Net [Çiçek

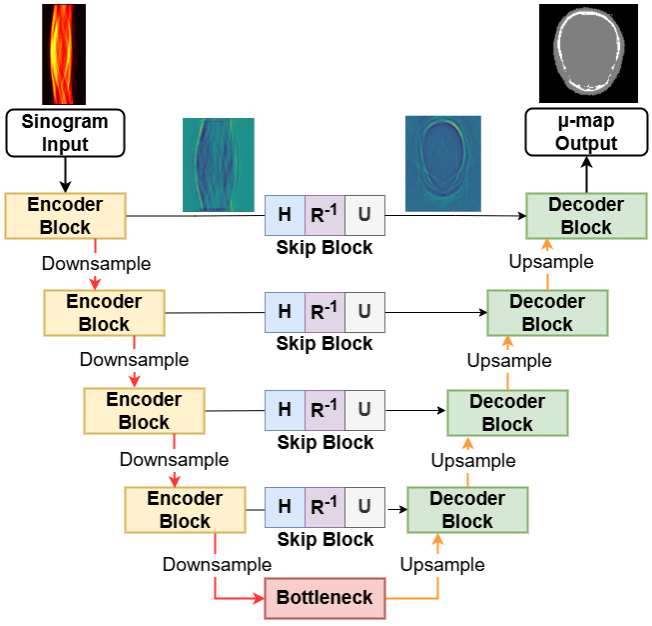

Figure 2: Overview of our 3D U-Net based network architecture for generating pseudo-μ-maps from PET sinograms using custom skip connections.

et al., 2016], which is typically used for 3D image segmentation, as well as 2D DeepPET [Häggström *et al.*, 2019] and 2D ReconU-Net [Hashimoto and Ote, 2024], which were developed to reconstruct PET images from attenuation-corrected sinograms.

We have selected the weighted cross-entropy loss [Ronneberger *et al.*, 2015], as shown in Equation 4, as the loss function for training our classification model.

$$\mathcal{L} = \frac{1}{|\Omega|} \sum_{\mathbf{x} \in \Omega} w(\mathbf{x}) \cdot \log(p_{\ell(\mathbf{x})}(\mathbf{x})) \tag{4}$$

Here, Ω represents the spatial domain of μ-maps, $w(\mathbf{x})$ is a weighting function at position $\mathbf{x}$, and $p_{\ell(\mathbf{x})}(\mathbf{x})$ denotes the predicted probability of the true label $\ell(\mathbf{x})$ at position $\mathbf{x}$. In brain attenuation maps, most areas are composed of air and soft tissues, while the region corresponding to bone is significantly smaller but exhibits a higher attenuation factor. The average voxel distribution in the training dataset for air, soft tissue, and bone was approximately 62.3%, 35.3%, and 2.4%, respectively. We used these values to determine the loss weights according to the normalized, inverse class frequencies.

4 Experiment and Results

4.1 Dataset

Currently, no publicly available datasets provide PET sinogram data with corresponding μ-maps for the same patients. To address this, we simulated a dataset using publicly available brain attenuation and activity templates derived from CT and MRI images of real patients [López-González *et al.*, 2020; Paredes-Pacheco *et al.*, 2021]. Simulated PET imaging is widely used to compensate for the scarcity of real patient data, particularly in raw sinogram format, as seen in studies like DeepPET [Häggström *et al.*, 2019] and ReconU-Net [Hashimoto and Ote, 2024]. SimPET, a Monte Carlo-based simulator, can model commercially available PET scanners and generate realistic sinograms. It has been validated against real datasets, showing that 98.09% of voxels exhibit differences of less than 10% when simulating the GE Discovery ST PET scanner [Paredes-Pacheco *et al.*, 2021]. SimPET has been successfully used in multiple studies to generate PET images from attenuation and activity templates derived from real patient data [López-González *et al.*, 2020; López-González *et al.*, 2019], making it well suited for our study.

The dataset [López-González *et al.*, 2020; Paredes-Pacheco *et al.*, 2021] comprises brain templates with simulated neurodegenerative diseases, derived from 25 healthy subjects. However, due to dimension mismatches in two attenuation maps, only 23 patient templates were included in our analysis. For each subject, in addition to the healthy brain activity maps, the dataset provides 29 variants featuring simulated dysfunction. These variants introduce five levels of hypometabolism (10%, 20%, 40%, 60%, and 80%) across six regions of interest (ROIs). ROIs 1–5 correspond to temporal lobe epilepsy (TLE), while ROI 6 represents Alzheimer's disease. The dataset was divided into training, validation, and holdout test sets, as shown in Table 1.

Sinogram generation was performed using the SimPET Monte Carlo simulator for the GE Discovery ST PET scanner. Each simulation represented a 5 minute PET scan using 18-F FDG radiotracer radiotracer at 22.2 MBq radioactive intensity applied for the brain, corresponding to approximately 10% of the full-body dose (245 MBq). The generated sinograms were also used to reconstruct both NAC PET images and AC PET images using the Ordered Subset Expectation Maximization (OSEM) algorithm with the STIR 4.0 toolkit [Thielemans *et al.*, 2012], following the default parameter template from SimPET.

Preprocessing: The 3D sinograms generated using the SimPET simulator, originally consisting of 23 oblique and direct planes with dimensions 210×249×576, were converted into standard 2D sinograms using the Single-Slice Rebinning (SSRB) algorithm [Defrise *et al.*, 2005] in the STIR 5.1 toolkit [Thielemans *et al.*, 2012]. This rebinning step reduced input data size and managed memory constraints efficiently. After rebinning, the resulting 2D sinograms had dimensions 210×249×47, where 210 represents azimuthal angle bins, 249 corresponds to transaxial distance bins, and 47 indicates

Segment	Patients	Simulated Sinograms
Training	14	416
Validation	4	124
Holdout Test	5	155
Total	**23**	**695**

Table 1: Dataset distribution across training, validation, and holdout test sets.

the number of slices along the depth axis. To further prepare the sinograms for model input, they were cropped to remove empty space and padded during data loading to standardized dimensions of 216×96×48 (height×width×depth). Additionally, the sinograms were normalized to the range [0,1] using a min-max scaler.

The CT-based μ-maps included regions below the brain (e.g., mouth and neck) absent in sinograms and reconstructed PET images due to their exclusion from activity maps. To ensure alignment, only the brain and skull regions were considered, and the μ-maps were cropped to 256×256×175 voxels. They were then downsampled to 128×128×48 voxels to match the pseudo-μ-maps, ensuring consistency with the GE Discovery ST PET scanner's reconstructed PET output. The downsampled CT-based μ-maps served as the ground truth for the model.

4.2 Implementation Details

All experiments were conducted on an Ubuntu system with dual Intel Xeon Silver 4208 CPUs, 128 GB of memory, an Nvidia Quadro 8000, and two Nvidia Quadro 6000 GPUs. Model implementation was done using PyTorch 2.5.1 [Ansel *et al.*, 2024], with TorchRadon 3.0 [Ronchetti, 2020] for FBP operations.

We used the Adam optimizer with the ReduceLROnPlateau learning rate scheduler, starting at 1e-5, and applied early stopping to prevent overfitting. The batch size was set to 8 to optimize GPU memory usage. Dropout rates of 0.1, 0.1, and 0.6 were applied to the encoder, bottleneck, and decoder layers, respectively, with L2 regularization (0.01) for all layers to enhance generalization. The full code for our experiments including the details of the network architecture, is available at https://github.com/prabathbr/PET_Attenuation_Estimation.

4.3 Baseline Model

The works in literature most relevant to ours focus on generating attenuation information for PET reconstruction use non-attenuation corrected (NAC) PET images as input with a 2D or 3D U-Net and regression loss function [Liu *et al.*, 2018; Hashimoto *et al.*, 2021; Partin *et al.*, 2024]. Therefore, we adapted this approach as our baseline. However, these prior works were conducted on proprietary datasets, and the corresponding model codes were not publicly available. Additionally, our dataset contains direct μ-maps instead of pseudo-CT images which makes the regression objective unsuitable. Consequently, we used a vanilla 3D U-Net architecture [Çiçek *et al.*, 2016] with the same loss function in Equation 4 to generate pseudo-μ-maps directly from NAC PET images. The model was fine-tuned to achieve optimal results.

During NAC PET image preprocessing, we removed the scanner gantry artifacts from the input NAC PET images using a circular mask and scaled them to align with the dimensions of the ground truth μ-maps.

4.4 Ablation Studies

The following two model variants were implemented as part of the ablation study: **A1) Direct Skip:** This model replaces the FBP-based skip connections in our proposed method with direct skip connections as in a conventional U-Net. **A2) No Skip:** This model removes the skip connections entirely making it a 3D encoder-decoder.

4.5 Evaluation Metrics

For Pseudo-μ-maps: We compared the generated pseudo-μ-maps with the CT-based ground truth μ-maps using classification accuracy, precision, and recall as evaluation metrics. We also calculated the Dice coefficient to measure the overlap between the generated image regions and the ground truth regions for each class.

For Reconstructed PET Images: To compare attenuation-corrected PET images generated using the proposed pseudo-μ-map and the CT-based μ-map, we evaluated mean squared error (MSE) and structural similarity index measure (SSIM). SSIM was computed using a sliding Gaussian window of size $p \times p$ with $p = 5$ to emphasize localized details and the dynamic range L was set to 5 considering range [Wang *et al.*, 2021]. To facilitate metric computation, reconstructed PET voxel values were converted to standardized uptake value ratio (SUVR), using the cerebellum as the reference region.

4.6 Results

Qualitative Results

We present qualitative comparisons of the generated pseudo-μ-maps from our model, the baseline model, and ablation studies against the ground truth in Figure 3. The ground truth, derived from CT images, inherently offers higher resolution and finer bone structure details than PET images. Our model closely replicates these structural features but at a lower resolution. The baseline model fails to capture finer details, such as air pockets (yellow solid arrows) and complex bone structures (cyan dotted arrows). Both ablation studies struggle to identify discontinuities in bone structures and accurately reconstruct soft tissue regions. However, the no skip ablation model preserves finer details better than the direct skip ablation model. We performed attenuation correction using pseudo-μ-maps obtaining reconstructed PET images, as shown in Figure 4. Overall, the images generated by our model appear visually similar to those produced using the CT-based ground truth, which preserves all structural information. In contrast, the non-attenuation corrected PET image exhibits noticeable differences.

Quantitative Comparison

We quantitatively compared the generated pseudo-μ-maps, and the results are shown in Table 2. We performed a Wilcoxon signed rank test with a significance level of p=0.05 to compare our proposed model with the baseline model across all evaluation metrics. The results indicate statistically significant differences, leading us to reject the null hypothesis. Our proposed model slightly outperforms the baseline for bone structures, which have a higher attenuation coefficient. The Dice coefficient increases from 0.639 to 0.650 and accuracy improves from 0.475 to 0.486. However, both models exhibit similar overall performance across all quantitative metrics. Our ablation studies highlight the advantage of custom skip connections. Compared to our proposed model (Dice: 0.650, Accuracy: 0.486 for bone structures), performance drops significantly when no skip connections are used,

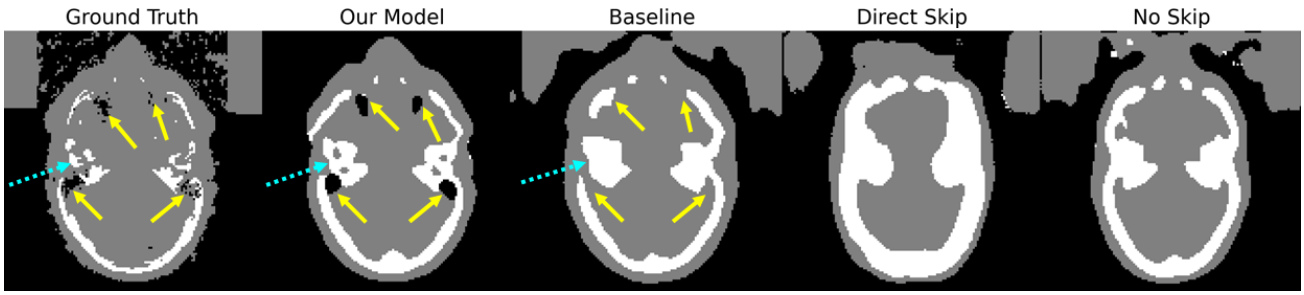

Figure 3: Comparison of generated pseudo-μ-maps in axial plane. Yellow solid arrows indicate air pockets, while cyan dotted arrows highlight complex bone structures. The baseline model struggles to capture finer details, whereas our model better preserves fine structural features, closely resembling the ground truth.

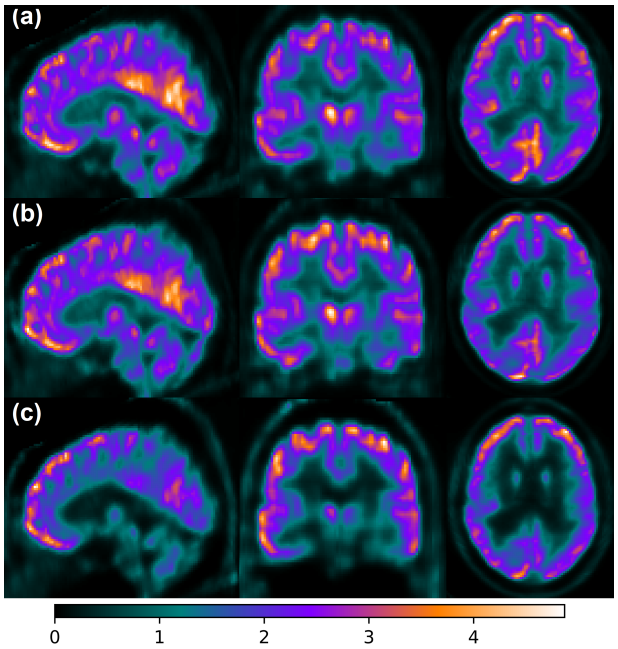

Figure 4: Visualization of SUVR in reconstructed PET images after attenuation correction using μ-maps over sagittal, coronal, and axial planes. The images are based on (a) CT-based ground truth, (b) our model, and (c) without attenuation correction.

resembling a 3D encoder-decoder (Dice: 0.414, Accuracy: 0.264). Direct skip connections further degrade performance (Dice: 0.356, Accuracy: 0.218). These findings suggest that skip connections without proper domain translation can be more detrimental than omitting them entirely. In the no skip case, the model learns domain translation solely through the bottleneck, which, while limiting, is still preferable to direct skip connections that fail to capture meaningful domain adaptation. We evaluated MSE and SSIM based on the SUVR values of reconstructed PET images attenuation corrected using both CT-based μ-maps and pseudo-μ-maps generated by our model. This analysis aimed to assess the clinical viability of our approach. The results show an MSE of 0.007 ± 0.004 and an SSIM of 0.956 ± 0.009, indicating high structural similarity and minimal error between the reconstructed images. These findings suggest that our method could serve as a reliable alternative for attenuation correction in PET imaging, with performance comparable to the CT-based ground truth.

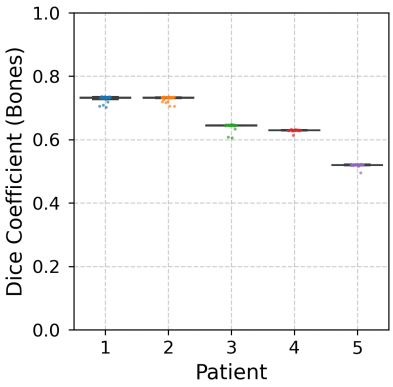

Figure 5: Dice coefficient variation for bone structure prediction across patients under different disease progression scenarios. The narrow IQR indicates that our proposed model maintains robustness in predicting bone structure despite disease progression.

4.7 Discussion

Robustness Testing

Robustness testing ensures deep learning models remain reliable, generalizable, and clinically accurate despite real-world variations. We validated our model using two approaches.

Our test dataset includes 30 variations of healthy and neurodegenerative disease-induced conditions per patient. This simulation replicates neurological pathology, such as the progression of Temporal Lobe Epilepsy or Alzheimer's disease, allowing us to assess the model's ability to adapt to different stages of disease evolution while maintaining reliable attenuation estimation. Figure 5 illustrates the variation in the Dice coefficient for bone structure prediction across patients (in the test set) from our proposed model across 30 different disease and region simulations. The narrow interquartile range (IQR) for each patient indicates that our model consistently predicts bone structure despite disease progression, demonstrating their robustness.

In order to evaluate the robustness of our model with unseen data, we modified an attenuation template by removing part of the skull to mimic incomplete skulls that may be encountered during PET scans due to surgical alterations or

	(a)			(b)	
Metric	Baseline	Our Model	stat (p-value)	Direct Skip	No Skip
Dice					
Bones	0.639 ± 0.084	$\mathbf{0.650 \pm 0.078}$	2651 (< 0.001)	0.356 ± 0.065	0.414 ± 0.076
Air	$\mathbf{0.961 \pm 0.011}$	0.953 ± 0.013	2966 (< 0.001)	0.924 ± 0.019	0.931 ± 0.023
Tissue	$\mathbf{0.876 \pm 0.028}$	0.857 ± 0.026	2330 (< 0.001)	0.703 ± 0.051	0.775 ± 0.055
Accuracy					
Bones	0.475 ± 0.091	$\mathbf{0.486 \pm 0.084}$	2762 (< 0.001)	0.218 ± 0.047	0.264 ± 0.060
Air	$\mathbf{0.924 \pm 0.020}$	0.911 ± 0.024	2989 (< 0.001)	0.860 ± 0.032	0.871 ± 0.040
Tissue	$\mathbf{0.781 \pm 0.044}$	0.750 ± 0.040	2331 (< 0.001)	0.544 ± 0.061	0.636 ± 0.074
Precision					
Bones	0.505 ± 0.072	$\mathbf{0.525 \pm 0.065}$	898 (< 0.001)	0.224 ± 0.048	0.294 ± 0.067
Air	$\mathbf{0.980 \pm 0.016}$	0.943 ± 0.044	0 (< 0.001)	0.970 ± 0.019	0.974 ± 0.023
Tissue	0.871 ± 0.051	$\mathbf{0.904 \pm 0.057}$	1032 (< 0.001)	0.749 ± 0.109	0.768 ± 0.109
Recall					
Bones	0.885 ± 0.141	0.867 ± 0.140	32.5 (< 0.001)	$\mathbf{0.895 \pm 0.099}$	0.727 ± 0.103
Air	0.942 ± 0.021	$\mathbf{0.966 \pm 0.021}$	622 (< 0.001)	0.885 ± 0.047	0.894 ± 0.056
Tissue	$\mathbf{0.883 \pm 0.029}$	0.822 ± 0.075	1432 (< 0.001)	0.670 ± 0.037	0.792 ± 0.031

Table 2: Comparison of evaluation metrics across models on the holdout test dataset. (a) Performance evaluation of our proposed model against the baseline. (b) Performance comparison of ablation studies to assess the impact of model modifications.

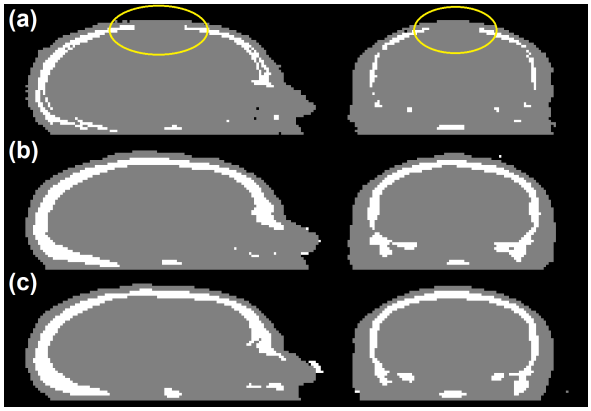

Figure 6: Alteration test conducted using (a) an attenuation template with a partially removed skull shown in yellow circle, used to simulate sinograms and NAC PET images, (b) the output μ-map from the baseline model, and (c) the output μ-map from our model. Both models fail to generalize to unseen deformations.

implants. These types of data augmentations were not used during training. We used sinograms and NAC PET images generated from this template as inputs to evaluate both our proposed model and the baseline model, respectively. However, both models failed to generate correct μ-maps, as shown in Figure 6. This result indicates that both data-driven approaches struggle with generalization in such scenarios.

Future Work

Our study is currently limited to brain PET images with a single radiotracer, 18-F FDG using a simulated dataset. Evaluating our approach on real full-body PET datasets and incorporating additional radiotracers would enhance its applicability, which we consider for future work. Another key limitation of our approach is its reliance on CT-based attenuation correction as ground truth. Because our model is trained on CT-derived μ-map, the generated μ-map may inherit modality-specific biases, limiting their ability to fully capture PET attenuation properties. CT-based attenuation correction requires energy scaling from 140 keV (X-ray) to 511 keV (γ photons), which can introduce systematic errors due to differences in attenuation between tissue and bone structures at these energy levels. This reliance on CT-based training data may reduce the model's generalizability across different attenuation conditions. Future work will explore physics-driven approaches to potentially mitigate these limitations.

5 Conclusion

In this work, we introduce a deep learning framework based on a 3D U-Net to directly estimate pseudo-μ-map from PET sinograms. Our approach, incorporating custom skip connections, eliminates the need for NAC reconstructed PET images, reducing preprocessing time and resource usage. Through rigorous evaluation, our model achieves a Dice coefficient of 0.650 and an accuracy of 0.486 for bone structures, outperforming the baseline 3D U-Net (Dice: 0.639, Accuracy: 0.475). Additionally, we demonstrate the clinical applicability of our method by using the generated pseudo-μ-map for attenuation correction in PET reconstruction, achieving results comparable to CT-based attenuation correction.

Despite its promising performance, our model faces challenges in handling unseen anatomical variations, such as missing structures due to the data-driven nature of our model. Additionally, because our training data uses CT-based μ-maps as ground truth, the generated maps may inherit CT-related biases. Future work will explore more physics-driven approaches to mitigate this limitation. Furthermore, while this study focuses on brain regions, we aim to extend our method to full-body PET imaging to enhance its broader applicability.

Ethical Statement

There are no ethical issues, as simulated datasets using publicly available templates are used in this paper.

Acknowledgments

This research is supported by Cyclotek (Aust) Pty Ltd. and an Australian Government Research Training Program (RTP) Scholarship. We also acknowledge the support provided by Dr. Jesús Silva-Rodríguez for the SimPET PET scanner simulator setup.

References

[Ansel *et al.*, 2024] Jason Ansel, Edward Yang, Horace He, et al. PyTorch 2: Faster Machine Learning Through Dynamic Python Bytecode Transformation and Graph Compilation. In *Proceedings of the 29th ACM International Conference on Architectural Support for Programming Languages and Operating Systems, Volume 2*, pages 929–947, New York, NY, USA, April 2024. Association for Computing Machinery.

[Arabi and Zaidi, 2020] Hossein Arabi and Habib Zaidi. Deep learning-guided estimation of attenuation correction factors from time-of-flight PET emission data. *Medical Image Analysis*, 64:101718, August 2020.

[Armanious *et al.*, 2020] Karim Armanious, Tobias Hepp, Thomas Küstner, et al. Independent attenuation correction of whole body [18F]FDG-PET using a deep learning approach with Generative Adversarial Networks. *EJNMMI Research*, 10(1):53, December 2020.

[Defrise *et al.*, 2005] Michel Defrise, Paul E. Kinahan, and Christian J. Michel. Image Reconstruction Algorithms in PET. In Dale L. Bailey, David W. Townsend, Peter E. Valk, and Michael N. Maisey, editors, *Positron Emission Tomography: Basic Sciences*, pages 63–91. Springer, London, 2005.

[Dong *et al.*, 2019] Xue Dong, Tonghe Wang, Yang Lei, et al. Synthetic CT generation from non-attenuation corrected PET images for whole-body PET imaging. *Physics in Medicine & Biology*, 64(21):215016, November 2019.

[Foster *et al.*, 2014] Brent Foster, Ulas Bagci, Awais Mansoor, et al. A review on segmentation of positron emission tomography images. *Computers in biology and medicine*, 50:76–96, July 2014.

[Hashimoto and Ote, 2024] Fumio Hashimoto and Kibo Ote. ReconU-Net: a direct PET image reconstruction using U-Net architecture with back projection-induced skip connection. *Physics in Medicine & Biology*, 69(10):105022, May 2024.

[Hashimoto *et al.*, 2021] Fumio Hashimoto, Masanori Ito, Kibo Ote, et al. Deep learning-based attenuation correction for brain PET with various radiotracers. *Annals of Nuclear Medicine*, 35(6):691–701, June 2021.

[Häggström *et al.*, 2019] Ida Häggström, C. Ross Schmidtlein, Gabriele Campanella, and Thomas J. Fuchs. DeepPET: A deep encoder–decoder network for directly solving the PET image reconstruction inverse problem. *Medical Image Analysis*, 54:253–262, May 2019.

[Kinahan *et al.*, 2003] Paul E. Kinahan, Bruce H. Hasegawa, and Thomas Beyer. X-ray-based attenuation correction for positron emission tomography/computed tomography scanners. *Seminars in Nuclear Medicine*, 33(3):166–179, July 2003.

[Liu *et al.*, 2018] Fang Liu, Hyungseok Jang, Richard Kijowski, et al. A deep learning approach for 18F-FDG PET attenuation correction. *EJNMMI Physics*, 5(1):24, December 2018.

[López-González *et al.*, 2019] Francisco Javier López-González, Alexis Moscoso, Nikos Efthimiou, et al. Spill-in counts in the quantification of 18F-florbetapir on Aβ-negative subjects: the effect of including white matter in the reference region. *EJNMMI Physics*, 6(1):27, December 2019.

[López-González *et al.*, 2020] Francisco J. López-González, Jesús Silva-Rodríguez, José Paredes-Pacheco, et al. Intensity normalization methods in brain FDG-PET quantification. *NeuroImage*, 222:117229, November 2020.

[Ma *et al.*, 2022] Ruiyao Ma, Jiaxi Hu, Hasan Sari, et al. An encoder-decoder network for direct image reconstruction on sinograms of a long axial field of view PET. *European Journal of Nuclear Medicine and Molecular Imaging*, 49(13):4464–4477, November 2022.

[McMillan and Bradshaw, 2021] Alan B. McMillan and Tyler J. Bradshaw. Artificial Intelligence–Based Data Corrections for Attenuation and Scatter in Position Emission Tomography and Single-Photon Emission Computed Tomography. *PET Clinics*, 16(4):543–552, October 2021.

[Paredes-Pacheco *et al.*, 2021] José Paredes-Pacheco, Francisco Javier López-González, Jesús Silva-Rodríguez, et al. SimPET—An open online platform for the Monte Carlo simulation of realistic brain PET data. Validation for 18F-FDG scans. *Medical Physics*, 48(5):2482–2493, May 2021.

[Partin *et al.*, 2024] Lauren Partin, Bruce Spottiswoode, Charles Hayden, et al. Deep learning-based CT-less attenuation correction of brain FDG PET. *Journal of Nuclear Medicine*, 65(supplement 2):242223, June 2024.

[Pohost and Dilsizian, 2019] Gerald M Pohost and Vasken Dilsizian. *Cardiac CT, PET and MR*. John Wiley & Sons, Incorporated, Newark, 2019.

[Quinn *et al.*, 2016] Brian Quinn, Zak Dauer, Neeta Pandit-Taskar, et al. Radiation dosimetry of 18F-FDG PET/CT: incorporating exam-specific parameters in dose estimates. *BMC Medical Imaging*, 16(1):41, December 2016.

[Ronchetti, 2020] Matteo Ronchetti. TorchRadon: Fast Differentiable Routines for Computed Tomography. *ArXiv*, abs/2009.14788, 2020.

[Ronneberger *et al.*, 2015] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Medical Image*

Computing and Computer-Assisted Intervention – MICCAI 2015, pages 234–241, Cham, 2015. Springer International Publishing.

[Thielemans *et al.*, 2012] Kris Thielemans, Charalampos Tsoumpas, Sanida Mustafovic, et al. STIR: software for tomographic image reconstruction release 2. *Physics in Medicine & Biology*, 57(4):867, January 2012.

[Vaquero and Kinahan, 2015] Juan José Vaquero and Paul Kinahan. Positron Emission Tomography: Current Challenges and Opportunities for Technological Advances in Clinical and Preclinical Imaging Systems. *Annual Review of Biomedical Engineering*, 17(1):385–414, December 2015.

[Wang *et al.*, 2021] Rui Wang, Hui Liu, Takuya Toyonaga, et al. Generation of synthetic PET images of synaptic density and amyloid from 18F-FDG images using deep learning. *Medical Physics*, 48(9):5115–5129, 2021.

[Xia *et al.*, 2012] Ting Xia, Adam M Alessio, Bruno De Man, et al. Ultra-low dose CT attenuation correction for PET/CT. *Physics in Medicine and Biology*, 57(2):309–328, January 2012.

[Çiçek *et al.*, 2016] Özgün Çiçek, Ahmed Abdulkadir, Soeren S. Lienkamp, et al. 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016*, volume 9901, pages 424–432, Cham, 2016. Springer International Publishing.