A New Paradigm for Low-dose PET/CT Reconstruction with Mamba-powered Progressive Network and Physics-informed Consistency

Zixin Tang[1], Caiwen Jiang[1], Zhiming Cui[1], Dinggang Shen[1,2,3(✉)]

[1] School of Biomedical Engineering & State Key Laboratory of Advanced Medical Materials and Devices, ShanghaiTech University, Shanghai 201210, China
dgshen@shanghaitech.edu.cn
[2] Shanghai United Imaging Intelligence Co., Ltd., Shanghai 200030, China
[3] Shanghai Clinical Research and Trial Center, Shanghai 201210, China

Abstract. Positron Emission Tomography (PET) is a powerful imaging technique but involves radiation exposure due to the use of radioactive tracers. A promising solution to mitigate this risk is reconstructing standard-dose PET (SPET) from low-dose PET (LPET). Previous studies have primarily focused on attenuation-corrected PET data; however, the attenuation correction process can amplify noise and artifacts, especially in low-dose scenarios. Additionally, PET scans are often paired with CT scans for attenuation correction, further contributing to radiation exposure. To address these challenges, we propose a new paradigm that reconstructs Attenuation-Corrected SPET (AC SPET) and standard-dose CT (SCT) images from the original Non-Attenuation-Corrected LPET (NAC LPET)) and low-dose CT (LCT) data through a collaborative reconstruction framework. Key components of our proposed method include: (1) a coarse-to-fine learning strategy, wherein specialized reconstruction basis is initially built by processing each modality individually, followed by Domain Adapters to facilitate cross-modal feature correlation; (2) a hybrid Mamba-powered Expert Network that effectively captures long-range dependencies between different regions of whole-body PET/CT images; and (3) a Physics-informed Mutual Loss function to enforce consistency between the PET and CT domains, ensuring robust and reliable reconstruction results. Extensive experiments on the collected dataset demonstrate that our model achieves diagnostic-quality reconstruction while significantly reducing radiation exposure.

Keywords: Low-dose PET/CT reconstruction · Mamba-based network · Mutual information · PET attenuation correction.

1 Introduction

Positron Emission Tomography (PET) is a sensitive functional imaging technique that visualizes biological processes within the human body. However, PET imaging requires the injection of radioactive tracers, which exposes patients to

radiation that may increase the risk of cancer [9]. To mitigate such risk, a promising approach—complementary to improvement in imaging hardware—is to reconstruct high-quality standard-dose PET (SPET) images from the low-dose PET (LPET) counterparts.

Early methods for LPET reconstruction are primarily filtering-based techniques, such as non-local means [1] and bilateral filters [7]. While easy to implement, these methods often produce overly smoothed images and exhibit suboptimal performance. Recently, deep learning-based methods have marked breakthroughs for medical image reconstruction [10, 11, 20, 22]. However, most of them rely on traditional convolution-based architectures [19], which are inherently limited in capturing long-range, global dependencies in 3D volumetric data. Moreover, these techniques typically perform reconstruction on PET data that has already undergone attenuation correction. The attenuation correction process can exacerbate the accumulation of noise and artifacts in LPET images, making it challenging to eliminate these distortions through post-processing algorithms.

Therefore, incorporating attenuation correction into the reconstruction algorithm, where Attenuation-Corrected SPET (AC SPET) images are reconstructed from the Non-Attenuation-Corrected LPET (NAC LPET) data, better aligns with actual clinical workflows. However, due to the fact that PET attenuation correction relies on calculating attenuation coefficients from corresponding CT images, it also introduces additional radiation due to the required CT scanning [14]. One potential solution to reduce the radiation risk from CT is to reconstruct standard-dose CT (SCT) images from low-dose CT (LCT) data [2]. In the context of CT, "low-dose" refers to adjustments in scanning parameters, such as a lower tube current [16]. While previous research on LCT reconstruction has primarily focused on single-modality CT imaging, limited attention has been given to reducing radiation from CT in the context of PET/CT imaging.

Considering all factors, we propose a collaborative PET/CT reconstruction algorithm that reconstructs AC SPET and SCT images from the original NAC LPET and LCT images, simultaneously reducing radiation exposure from both modalities across the entire imaging process. Specifically, our method employs a coarse-to-fine learning framework, where a single-modal reconstruction basis is first developed by processing PET and CT independently, followed by the integration of Domain Adapters to enhance cross-modal feature alignment. Within this framework, we utilize a hybrid Mamba-powered Expert Network to better extract long-range dependency from whole-body PET/CT images with a large field-of-view (FOV). Additionally, we design a Physics-informed Mutual Loss function to reinforce the object-specific consistency between the PET and CT domains, ensuring robust and accurate reconstruction results.

The contributions of our work can be summarized as follows: (1) proposing a novel PET/CT reconstruction paradigm by incorporating attenuation correction to better align with clinical practice and minimize radiation risks from both scans; (2) developing a dual-modality Mamba-based framework with physics-based constraints to facilitate complementary reconstruction; and (3) validating the effectiveness of the proposed method through extensive experiments.

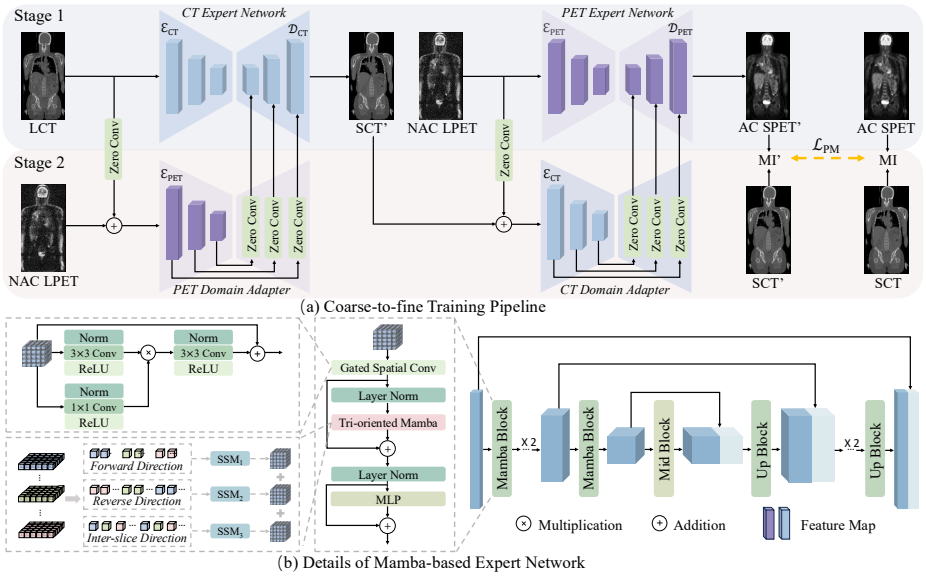

Fig. 1. Overview of the proposed method. (a) Illustration of the coarse-to-fine training framework. (b) The detailed structure of the Mamba-based Expert Network.

2 Methodology

The pipeline of the proposed framework is presented in Fig. 1 (a). Given the input LCT and NAC LPET, they are fed into the CT Expert Network and PET Expert Network, respectively, for reconstruction in Stage 1. In Stage 2, we first integrate a PET Domain Adapter to enhance the reconstruction of LCT, and then symmetrically use the refined SCT' to enhance the reconstruction of PET through the CT Domain Adapter. A Physics-informed Mutual Loss ($\mathcal{L}_{\text{PM}}$) is introduced to enforce consistency between the AC SPET' and SCT'. Below, we provide a detailed explanation of each component within the framework.

2.1 Coarse-to-fine Learning Strategy

PET images reflect the internal metabolic condition, while CT images accurately represent the anatomical structure. To effectively harness the complementary nature of these modalities while preserving their distinct characteristics, we propose a coarse-to-fine learning strategy by first achieving single-modal reconstruction and then introducing cross-modal feature alignment to enhance the synergy between the two modalities.

In the first stage, the input LCT and NAC LPET are processed through two dedicated networks for single-modal reconstruction. By independently processing the PET and CT images, we obtain sufficiently trained expert networks, denoted as $\mathcal{E}_{\text{CT}}$, $\mathcal{D}_{\text{CT}}$, and $\mathcal{E}_{\text{PET}}$, $\mathcal{D}_{\text{PET}}$. This allows each modality to be handled

according to its specific characteristics, serving as the foundation for the subsequent stages. In the second stage, to enhance the cross-modal feature alignment, we draw inspiration from ControlNet [25] and introduce a Domain Adapter to extract and fuse multi-scale features from the two imaging domains. First, the NAC LPET is used to enhance the reconstruction of CT images. The trained $\mathcal{E}_{\mathrm{CT}}$ is frozen in this step, while the $\mathcal{D}_{\mathrm{CT}}$ network remains trainable. The trained $\mathcal{E}_{\mathrm{PET}}$ serves as the PET Domain Adapter to extract metabolic-related features from the NAC LPET. The input to this process is a fusion of the NAC LPET and zero-convolved LCT. The features extracted at each $\mathcal{E}_{\mathrm{PET}}$ layer are then passed through zero-convolution layers and forwarded to the corresponding layers of $\mathcal{D}_{\mathrm{CT}}$. This allows the CT reconstruction to benefit from additional metabolic information derived from PET while preserving the anatomical features learned during Stage 1, ensuring more accurate alignment between the PET and CT modalities.

A similar approach is then applied to the PET reconstruction, where the refined SCT' is introduced as a supplement, in line with the clinical requirement for attenuation correction. The coarse-to-fine strategy effectively integrates the modality-specific enhancements with cross-modal refinement, ultimately leading to improved reconstructions for both PET and CT images.

2.2 Expert Modeling with Mamba Blocks

Considering the 3D nature of PET and CT images, long-range feature dependencies within the volumetric data are inherently complex. Conventional convolution-based networks are limited in their ability to extract long-range information from high-resolution 3D medical images due to the locality of their receptive fields. To address the challenges of long-sequence modeling, inspired by previous studies [6, 23], we propose a Mamba-based Expert Network, which retains the U-shaped architecture but integrates Mamba blocks to replace the original convolutional blocks in the encoder.

As shown in Fig. 1 (b), each Mamba block begins with a Gated Spatial Convolution (GSC) module. The GSC module plays a key role in controlling the flow of information through a gating mechanism [13], which selectively enhances or suppresses spatial features based on their relevance. This ensures that the network only focuses on the most informative features, thereby reducing redundancy in feature representation. To capture long-range dependencies from 3D features, the network integrates a Tri-orientated Mamba layer, which computes feature dependencies from three directions: forward, reverse, and inter-slice. The input features are first flattened into these three sequences for feature interaction and then summed up to merge the features into a fused representation. This approach enhances the network's ability to process complex volumetric data, ensuring comprehensive local and global contextual information modeling.

The decoder side maintains a typical design with up-sampling blocks, which consists of a residual connection to merge features learned in previous layers and transposed convolutions to recover spatial dimensions. The overall design

enables the network to fully exploit spatial features across the entire 3D volume, capturing both local details and global relationships.

2.3 Mutual Consistency for PET/CT Alignment

While the coarse-to-fine learning strategy facilitates reconstruction by sharing complementary information between the two modalities, it is essential to maintain physical consistency between the anatomical details provided by CT and the metabolic information provided by PET. To achieve this, we introduce a Physics-informed Mutual Loss ($\mathcal{L}_{\mathrm{PM}}$) to encourage the network to align both modalities in the final PET refinement stage.

The physical consistency is measured by calculating the mutual information [18] between paired PET and CT, which reflects the degree of statistical dependence and shared information between the two modalities. By minimizing the discrepancy in mutual information between the predicted and ground-truth PET-CT pairs, our proposed Physics-informed Mutual Loss enforces subject-specific consistency, thereby improving the accuracy and reliability of the reconstructed images. Formally, the $\mathcal{L}_{\mathrm{PM}}$ can be expressed as:

$$\mathcal{L}_{\mathrm{PM}} = \left| I(\hat{P}, \hat{C}) - I(P, C) \right|, \tag{1}$$

where $\hat{P}$ and $\hat{C}$ represent the predicted PET and CT images, while P and C denote the ground-truth PET and CT images, respectively. $I(.,.)$ represents the mutual information between the two modalities. To also account for voxel-wise accuracy, the final loss function is a weighted combination of three components: $\mathcal{L}_1$ loss, $\mathcal{L}_2$ loss, and the Physics-informed Mutual Loss.

$$\mathcal{L}_{total} = \lambda_1 \mathcal{L}_1 + \lambda_2 \mathcal{L}_2 + \lambda_3 \mathcal{L}_{\mathrm{PM}}, \tag{2}$$

where λ_1, λ_2, and λ_3 are the weighting factors used to control the contribution of each loss term to the final objective. By minimizing this total loss, the network is encouraged to generate high-quality PET images with voxel-wise accuracy to the ground-truth and also anatomical consistency with the CT.

3 Experiments

3.1 Dataset and Implementation Details

We collect 251 samples from The Cancer Imaging Archive (TCIA) [4] Non-Small Cell Lung Cancer (NSCLC) collection [12], with 200 samples used for training, 26 for validation, and 25 for testing. Each sample includes co-registered whole-body scans of SCT, SPET, and NAC SPET. We follow the sparse-sampling strategy to simulate LCT [8], where the SCT data is projected into a sinogram using the Radon transform and then uniformly downsampled by selecting every 5th

[4] https://www.cancerimagingarchive.net/collections/

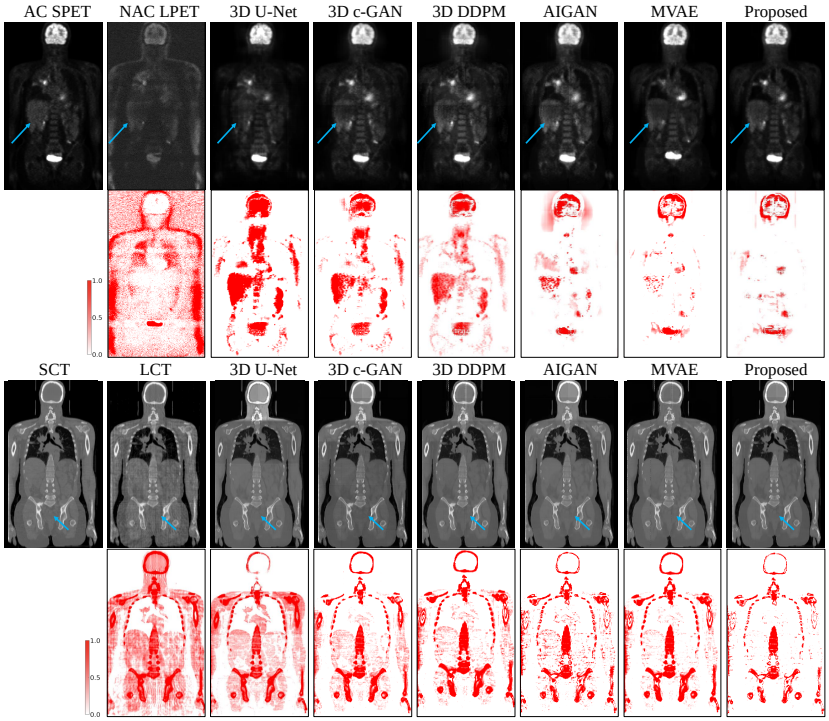

Fig. 2. Visual comparison of reconstructed PET (1st row) and CT (3rd row) images produced by six different methods. The corresponding error maps between the generated results and the ground truth are shown in the 2nd and 4th rows, respectively. Key regions of interest are indicated by the blue arrows.

projection. This sparse sinogram is subsequently reconstructed using the inverse Radon transform with the reduced-angle projections. To simulate the reduced signal and increased noise characteristic of NAC LPET, Poisson noise is added to the sinogram of NAC SPET [15], with the intensity scaled by a factor of 100, mimicking the effects of reduced radiation exposure during data acquisition. The simulation approaches have demonstrated strong correlation with real low-dose outcomes, enabling our experiments to closely approximate real-world scenarios. For consistency across all scans in our experiments, we resample the data to a voxel spacing of $2\times2\times2\ mm^3$ and re-scaled the intensity values to the range [0, 1]. Each scan is then cropped into patches of size $128\times128\times128$ voxels to fit within the memory limitations. To achieve full FOV output, consecutive outputs from each sample are stitched into a single volume, with overlapping regions averaged to preserve consistency.

Our model is implemented using the PyTorch framework and trained on an Nvidia Tesla A100 80GB GPU. The weighting factors λ_1, λ_2, and λ_3 in Eq. 2 are initialized to 0.4, 0.4, and 0.2, respectively. We employ the Adam optimizer with

Table 1. Quantitative comparison of our method with five state-of-the-art generation methods, in terms of PSNR [dB] and SSIM [%].

Method	PET		CT	
	SSIM	PSNR	SSIM	PSNR
3D U-Net [3]	90.75 (2.35)	31.52 (2.98)	91.62 (1.76)	31.81 (2.02)
3D c-GAN [21]	93.36 (1.63)	33.59 (2.04)	93.86 (1.51)	33.60 (1.98)
3D DDPM [24]	94.25 (1.21)	34.41 (1.73)	95.72 (0.90)	35.78 (1.53)
AIGAN [5]	95.19 (1.40)	35.33 (1.86)	96.05 (1.16)	36.11 (1.94)
MVAE [17]	95.84 (1.29)	35.87 (1.97)	96.90 (0.95)	36.45 (1.47)
Proposed	**96.73 (1.08)**	**36.44 (1.86)**	**97.32 (0.84)**	**37.19 (1.33)**

a learning rate of 0.0001. Validation is performed every 5 epochs, and the best model is selected based on its performance on the validation set. The quantitative evaluation of the model's performance is carried out using two metrics: Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM).

3.2 Comparison Experiments

We compare our proposed method with five state-of-the-art approaches, including single-modality methods: 3D U-Net [4], 3D-cGAN [21], and 3D DDPM [24]; and dual-modality methods: AIGAN [5] and MVAE [17]. All comparison methods, as well as our proposed method, are implemented using a consistent data processing pipeline. The quantitative results are presented in Table 1, while the qualitative results are shown in Fig. 2.

Quantitative Evaluation The results presented in Table 1 highlight the superior performance of our method compared to five state-of-the-art approaches. Our method achieves the highest PSNR and SSIM values for both PET and CT reconstructions, with PET SSIM at **96.73%** and CT SSIM at **97.32%**. Similarly, for PSNR, our method attains **36.44 dB** for PET and **37.19 dB** for CT, outperforming the closest competitor, MVAE, by margins of 0.57 dB and 0.74 dB, respectively. Notably, our method also exhibits the smallest standard deviations across all metrics, demonstrating its stability and consistency. These improvements can be attributed to the proposed collaborative reconstruction framework, which effectively leverages cross-modal information and advanced feature modeling. The results underscore the robustness of our approach, making it a promising solution for real-world clinical applications.

Qualitative Evaluation The visualization results of the reconstructed PET and CT images using different methods are shown in Fig. 2. Our proposed method demonstrates its capability to reconstruct PET and CT images with the least noise while retaining detailed texture. This is particularly attributed to the incorporation of the Mamba-based Expert Network, which effectively captures long-range dependencies across the entire 3D volume, leading to clearer and more accurate reconstructions. Meanwhile, the error maps for the PET and CT images reconstructed by our method exhibit the lightest color, reflecting a closer

Table 2. Quantitative results of ablation study, in terms of PSNR [dB] and SSIM [%].

Method	PET		CT	
	SSIM	PSNR	SSIM	PSNR
M1	93.63 (1.86)	34.45 (2.43)	95.81 (1.09)	36.94 (1.78)
M2	94.58 (1.19)	35.09 (2.27)	96.64 (0.92)	37.62 (1.56)
M3	96.15 (1.13)	36.13 (1.95)	**97.32 (0.84)**	**38.19 (1.33)**
M4	**96.73 (1.08)**	**36.44 (1.86)**	**97.32 (0.84)**	**38.19 (1.33)**

match to the AC SPET and SCT images. This indicates that combining single-modal reconstruction with cross-modal refinement, the coarse-to-fine learning strategy significantly improves the reconstruction quality. Together, these observations validate the effectiveness of our proposed method and highlight its superior performance over state-of-the-art approaches in terms of both image quality and accuracy.

3.3 Ablation Study

To evaluate the effectiveness of each proposed strategy, we conduct ablation studies with four models: 1) M1: Base model without Mamba blocks; 2) M2: Mamba-based Expert Model without the coarse-to-fine strategy; 3) M3: Mamba-based Expert Model with the coarse-to-fine strategy but without the Physics-informed Mutual Loss; and 4) M4: Mamba-based Expert Model with both the coarse-to-fine strategy and the mutual consistency loss. All methods are tested under the same experimental settings, with quantitative results shown in Table 2.

The results in Table 2 show that including Mamba blocks (M2) leads to notable improvements in reconstruction quality, as evidenced by the higher SSIM and PSNR values for both PET and CT images. This enhancement can be attributed to the Mamba blocks' ability to capture long-range dependencies, which is crucial to preserving fine details and global structures. Introducing the coarse-to-fine strategy (M3) further refines the reconstruction, particularly in SSIM for CT (97.32%). This suggests that the coarse-to-fine approach effectively leverages complementary PET and CT information, resulting in more accurate and stable reconstructions. The inclusion of the Physics-informed Mutual Consistency loss (M4) further improves PET reconstruction, achieving the highest SSIM and PSNR values among all configurations. Note that CT reconstruction is completed within the M3 setting, which is why the CT metrics remain unchanged in M4. This highlights the critical role of aligning anatomical and metabolic features between the two modalities, ensuring more accurate and consistent reconstructions. These findings demonstrate the positive contribution of each component to enhancing PET/CT reconstruction quality, with the full model (M4) achieving the best overall results.

4 Conclusion

In this paper, we propose a novel approach to reduce radiation exposure from both PET and its accompanying CT scans during image acquisition by collaboratively reconstructing NAC LPET and LCT images. To achieve this, we design a coarse-to-fine learning framework that effectively leverages PET and CT's complementarity for more stable and accurate reconstruction. Furthermore, within this framework, we introduce a Mamba-based Expert Network and a Physics-based Mutual Loss to capture long-range dependencies in whole-body PET/CT volumes and enhance consistency between the PET and CT domains, thereby improving reconstruction performance. Extensive experiments validate the effectiveness of our approach, offering a promising solution for achieving more efficient and safer PET/CT acquisition.

Acknowledgments. This work was supported in part by National Natural Science Foundation of China (grant numbers 82441023, U23A20295, 62131015, 82394432), the China Ministry of Science and Technology (S20240085, STI2030-Major Projects-2022ZD0209000, STI2030-Major Projects-2022ZD0213100), Shanghai Municipal Central Guided Local Science and Technology Development Fund (No. YDZX20233100001001), The Key R&D Program of Guangdong Province, China (grant number 2023B0303040001), and HPC Platform of ShanghaiTech University.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Buades, A., Coll, B., Morel, J.M.: A non-local algorithm for image denoising. In: Proc. of CVPR. vol. 2, pp. 60–65. IEEE (2005)
2. Chen, H., Li, Q., Zhou, L., Li, F.: Deep learning-based algorithms for low-dose CT imaging: A review. European Journal of Radiology **172**, 111355 (2024)
3. Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O.: 3d u-net: Learning dense volumetric segmentation from sparse annotation. In: Ourselin, S., Joskowicz, L., Sabuncu, M.R., Unal, G., Wells, W. (eds.) Proc. of MICCAI. pp. 424–432. Springer International Publishing, Cham (2016)
4. Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O.: 3D U-Net: learning dense volumetric segmentation from sparse annotation. In: Proc. of MICCAI. pp. 424–432. Springer (2016)
5. Fu, Y., Dong, S., Niu, M., Xue, L., Guo, H., Huang, Y., Xu, Y., Yu, T., Shi, K., Yang, Q., et al.: AIGAN: Attention–encoding integrated generative adversarial network for the reconstruction of low-dose CT and low-dose PET images. Medical Image Analysis **86**, 102787 (2023)
6. Gu, A., Dao, T.: Mamba: Linear-time sequence modeling with selective state spaces. arXiv preprint arXiv:2312.00752 (2023)
7. Hofheinz, F., Langner, J., Beuthien-Baumann, B., Oehme, L., Steinbach, J., Kotzerke, J., van den Hoff, J.: Suitability of bilateral filtering for edge-preserving noise reduction in PET. EJNMMI Research **1**, 1–9 (2011)

8. Hu, Z., Gao, J., Zhang, N., Yang, Y., Liu, X., Zheng, H., Liang, D.: An improved statistical iterative algorithm for sparse-view and limited-angle CT image reconstruction. Scientific Reports **7**(1), 10747 (2017)
9. Huang, B., Law, M.W.M., Khong, P.L.: Whole-body PET/CT scanning: estimation of radiation dose and cancer risk. Radiology **251**(1), 166–174 (2009)
10. Jiang, C., Pan, Y., Cui, Z., Nie, D., Shen, D.: Semi-supervised standard-dose PET image generation via region-adaptive normalization and structural consistency constraint. IEEE Transactions on Medical Imaging **42**(10), 2974–2987 (2023)
11. Jiang, C., Pan, Y., Liu, M., Ma, L., Zhang, X., Liu, J., Xiong, X., Shen, D.: PET-diffusion: Unsupervised PET enhancement based on the latent diffusion model. In: Proc. of MICCAI. pp. 3–12. Springer (2023)
12. Kinahan, P., Muzi, M., Bialecki, B., Herman, B., Coombs, L.: Data from the ACRIN 6668 Trial NSCLC-FDG-PET (Version 2) [data set] (2019). https://doi.org/10.7937/tcia.2019.30ilqfcl
13. Liu, H., Dai, Z., So, D., Le, Q.V.: Pay attention to MLPs. In: Ranzato, M., Beygelzimer, A., Dauphin, Y., Liang, P., Vaughan, J.W. (eds.) Proc. of NeurIPS. vol. 34, pp. 9204–9215. Curran Associates, Inc. (2021)
14. Martí-Climent, J.M., Prieto, E., Morán, V., Sancho, L., Rodríguez-Fraile, M., Arbizu, J., García-Velloso, M.J., Richter, J.A.: Effective dose estimation for oncological and neurological PET/CT procedures. EJNMMI Research **7**, 1–8 (2017)
15. Mokri, S., Saripan, M., Abd Rahni, A., Nordin, A., Hashim, S., Marhaban, M.: PET image reconstruction incorporating 3d mean-median sinogram filtering. IEEE Transactions on Nuclear Science **63**(1), 157–169 (2016)
16. Mostafapour, S., Greuter, M., van Snick, J.H., Brouwers, A.H., Dierckx, R.A., van Sluis, J., Lammertsma, A.A., Tsoumpas, C.: Ultra-low dose CT scanning for PET/CT. Medical Physics **51**(1), 139–155 (2024)
17. Pinton, N.J., Bousse, A., Cheze-Le-Rest, C., Visvikisc, D.: Multi-branch generative models for multichannel imaging with an application to PET/CT synergistic reconstruction. IEEE Transactions on Radiation and Plasma Medical Sciences (2025)
18. Pluim, J.P., Maintz, J.A., Viergever, M.A.: Mutual-information-based registration of medical images: a survey. IEEE Transactions on Mdical Imaging **22**(8), 986–1004 (2003)
19. Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional networks for biomedical image segmentation. In: Proc. of MICCAI. pp. 234–241. Springer (2015)
20. Tang, Z., Jiang, C., Cui, Z., Shen, D.: Hf-resdiff: High-frequency-guided residual diffusion for multi-dose pet reconstruction. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 372–381. Springer (2024)
21. Wang, Y., Yu, B., Wang, L., Zu, C., Lalush, D.S., Lin, W., Wu, X., Zhou, J., Shen, D., Zhou, L.: 3d conditional generative adversarial networks for high-quality PET image estimation at low dose. Neuroimage **174**, 550–562 (2018)
22. Xiang, L., Chen, Y., Chang, W., Zhan, Y., Lin, W., Wang, Q., Shen, D.: Deep-learning-based multi-modal fusion for fast mr reconstruction. IEEE Transactions on Biomedical Engineering **66**(7), 2105–2114 (2018)
23. Xing, Z., Ye, T., Yang, Y., Liu, G., Zhu, L.: Segmamba: Long-range sequential modeling mamba for 3D medical image segmentation. In: Proc. of MICCAI. pp. 578–588. Springer (2024)
24. Yu, B., Ozdemir, S., Dong, Y., Shao, W., Shi, K., Gong, K.: PET image denoising based on 3d denoising diffusion probabilistic model: Evaluations on total-body datasets. In: Proc. of MICCAI. pp. 541–550. Springer (2024)

25. Zhang, L., Rao, A., Agrawala, M.: Adding conditional control to text-to-image diffusion models. In: Proc. of ICCV. pp. 3836–3847 (2023)