WiD-PET: PET Image Reconstruction from Low-Dose Data Using a Wavelet-Informed Diffusion Model with Fast Inference

Qingcheng Lyu[1], Tong Chen[1], Yiran Wang[1], Erjian Guo[1], and Luping Zhou[1]✉

University of Sydney

Abstract. Reconstruction of standard-dose Positron Emission Tomography is vital for clinical diagnosis, while recent diffusion-denoising probabilistic models offer strong generative capabilities, when applied to this task, they often struggle with fine detail recovery, slow inference, and inadequate cross-slice continuity in 3D volumes. To overcome these issues, we introduce WiD-PET, which employs a wavelet transform to produce smaller wavelet-transformed inputs, and thereby reduces inference time to 10% of that required by the DDPM model. Additionally, a high-frequency enhancer is adopted for reconstructing fine and rich image details. Moreover, a spatial consistency feature extractor and spatial consistency attention are implemented to enhance cross-slice continuity in 3D PET reconstructions. Evaluations across dose levels (1/20, 1/50, and 1/100) reveal that WiD-PET consistently achieves superior reconstruction quality, detail preservation, and inference efficiency. Project page: `https://github.com/SwingM/WiD.git`.

Keywords: Positron Emission Tomography (PET) · Diffusion Model · Image Reconstruction · Ultra-low Doses · Wavelet-transformation

1 Introduction

Positron emission tomography (PET) is essential for assessing human metabolism, yet high-quality PET imaging requires high radiotracer doses, increasing radiation risks. Lower dose leads to a lower signal-to-noise ratio of the PET image, thereby, reducing diagnostic accuracy, especially for ultra-low-dose levels (e.g., 1/50 or 1/100). Reconstruction of standard-dose PET (SPET) from low-dose PET (LPET) is crucial.

Some GAN-based methods have been proposed for PET image reconstruction, including GAN-based PET synthesis [9], cycle-GAN approaches [21,22], and specialized variants. For example, EA-GAN [20] incorporates an edge detector to improve edge quality, and AR-GAN [12] uses adaptive rectification with spectral constraints. Other innovations include PCC-GAN [4] with point context clustering, StillGAN [13] with structure and illumination constraints, and MRI-CGAN [2,18] that leverages MRI as auxiliary information. However,

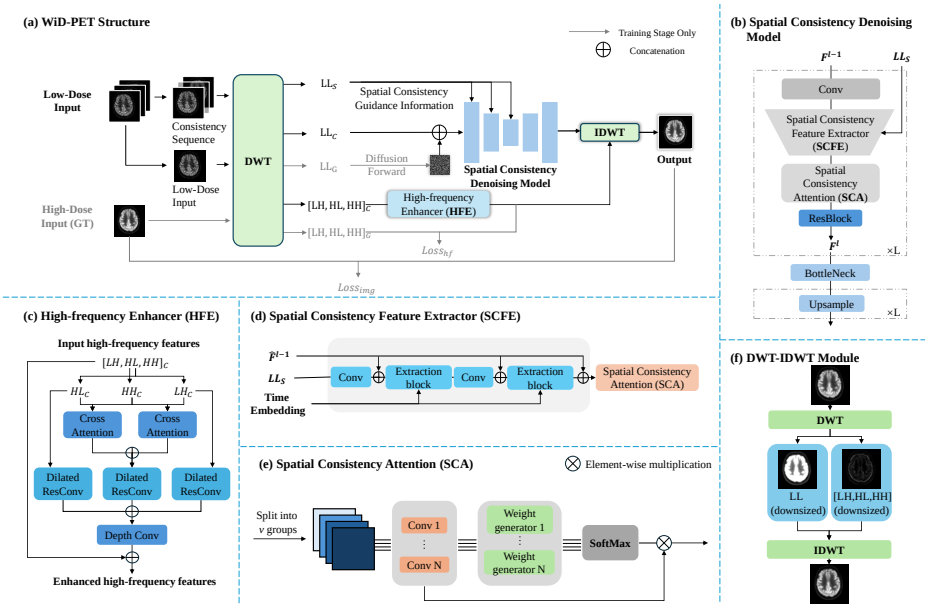


Fig. 1. Overall workflow of WiD-PET. (a) illustrates the complete process, while (b)–(f) depicts individual modules. Please see Section 2.2 for more explanations.

GAN-based methods often suffer from adversarial training instability and mode collapse [14,16], resulting in low discriminability of the generated samples.

Diffusion Denoisingn Probabilistic Models (DDPM) [8] have recently surpassed GAN-based approaches in medical image reconstruction, and many works in the field are published [3,6,15]. While effective, when applied to SPET reconstruction, diffusion models face at least three key challenges. **First**, their iterative sampling process is computationally expensive, particularly for high-resolution inputs, resulting in significantly slower inference speeds compared to GAN-based models. **Second**, traditional diffusion models uniformly denoise in the image space, often failing to recover fine details, as textures and edges in ultra-low dose PET (UDPET) images—corresponding to high-frequency components—are particularly prone to degradation. **Third**, models originally designed for 2D images, typically process slices independently, neglecting spatial consistency across slices and resulting in incoherent reconstructions.

To address the limitations of existing diffusion-based models for dose-less PET imaging, we propose a novel framework, the **Wavelet-informed Diffusion Model with Fast Inference (WiD-PET)**, which introduces several innovative contributions. **First**, by utilizing wavelet transformations and a high-frequency enhancer (HFE) module, WiD-PET focuses on reconstructing fine image details and synthesizing SPET images through the integration of these enhanced details with the global contrast produced by DDPM. **Second**, processing wavelet-transformed inputs allows WiD-PET to focus on critical image features

and reduces inference time by an impressive 90% compared to the 2D-DDPM. **Third**, to ensure spatial coherence across 3D PET slices, WiD-PET incorporates a spatial consistency feature extractor (SCFE) and spatial consistency attention (SCA), enabling seamless and accurate reconstructions. Extensive testing on various dose levels (1/20, 1/50, and 1/100) confirms WiD-PET's ability to deliver superior reconstruction quality, detail preservation, and outstanding efficiency.

2 Method

2.1 Preliminary

Denoising Diffusion Probabilistic Models (DDPM) provide a framework for image reconstruction [8]. Let the SPET and LPET image sets be $\mathcal{X}^G = \{\mathbf{x}_i^G\}_{i=1}^N$ and $\mathcal{X}^L = \{\mathbf{x}_i^L\}_{i=1}^N$, respectively, with $\mathbf{x}_i^G, \mathbf{x}_i^L \in \mathbb{R}^{H\times W\times C}$. In our framework, the generation of $\mathbf{x}^G$ is conditioned on the LPET image and a spatial sequence S (adjacent slices before and after current slice), denoted as $\mathcal{C} = \{\mathbf{c}_i\}_{i=1}^N$. Here, we omit the subscript i for brevity and denote the timestep as t. In a conditional setting (cDDPM), the forward process adds noise as:

$$q(\mathbf{x}_t^G \mid \mathbf{x}_{t-1}^G) = \mathcal{N}(\mathbf{x}_t^G; \sqrt{\alpha_t}\mathbf{x}_{t-1}^G, \beta_t\mathbf{I}), \tag{1}$$

with $\alpha_t = 1 - \beta_t$. The reverse process denoises iteratively:

$$p_\theta(\mathbf{x}_{t-1}^G \mid \mathbf{x}_t^G, \mathbf{c}) = \mathcal{N}\left(\mathbf{x}_{t-1}^G; \mu_\theta(\mathbf{x}_t^G, t, \mathbf{c}), \Sigma_\theta(\mathbf{x}_t^G, t, \mathbf{c})\right). \tag{2}$$

Wavelet Transformation decomposes an input image $\psi \in \mathbb{R}^{H\times W\times C}$ into frequency bands $\mathbf{LL}, \mathbf{LH}, \mathbf{HL}, \mathbf{HH} \in \mathbb{R}^{\frac{H}{2}\times\frac{W}{2}\times C}$ via:

$$f_{\text{dwt}}(\psi) = \{\mathbf{LL}, \mathbf{LH}, \mathbf{HL}, \mathbf{HH}\}, \tag{3}$$

2.2 Overview

An overview of the proposed WiD-PET framework is illustrated in Figure 1(a). First, SPET images, along with the LPET condition and its adjacent slices (spatial sequence), are decomposed using the Haar 2D discrete wavelet transform (DWT) module, leading to the low-frequency component $\mathbf{LL}_x$ and the high-frequency components $[\mathbf{LH}, \mathbf{HL}, \mathbf{HH}]_x$, where the subscript $x = g, c, S$ indicates the ground-truth SPET image, the LPET condition, and the spatial sequence of LPET, respectively. The low-frequency component $\mathbf{LL}_g$ undergoes a forward diffusion process, then it is combined with the condition $\mathbf{LL}_c$ and passed to the denoising model (Fig. 1(b)), $\mathbf{LL}_S$ is also used as a condition to provide spatial consistency guidance. The spacial consistency guidance is processed through a spatial consistency feature extractor block (SCFE) followed by a spacial consistency attention module (SCA) to steer the denoising process effectively. At the other hand, the high-frequency components are refined using a High-frequency Enhancer (HFE) to restore fine details and textures. Finally, the enhanced high- and reconstructed low-frequency components are re-composed back to the reconstructed image using the inverse discrete wavelet transform (IDWT).

2.3 Fast Wavelet-informed Diffusion Architecture

Our framework consists of a DWT-IDWT module, a Spatial Consistency Denoising Model, and a High-frequency Enhancer module.

DWT-IDWT Module: As aforementioned, our approach employs a DWT-IDWT module (Fig. 1(f)) to decompose the input SPET ground truth, LPET condition, and spatial sequence S into low-frequency ($\mathbf{LL}_g$, $\mathbf{LL}_c$, $\mathbf{LL}_S$) and high-frequency ($[\mathbf{LH}, \mathbf{HL}, \mathbf{HH}]_g$, $[\mathbf{LH}, \mathbf{HL}, \mathbf{HH}]_c$, $[\mathbf{LH}, \mathbf{HL}, \mathbf{HH}]_S$) components. This decomposition reduces each component's size from $C \times H \times W$ to $C \times \frac{H}{2} \times \frac{W}{2}$, effectively scaling down the input by a factor of 4. Using wavelet-decomposition modules has been demonstrated to efficiently reduce the input image's scale, thereby lowering inference time demands [5,10,11]. Here, the low-frequency component of SPET ($\mathbf{LL}_g$) undergoes forward diffusion and is fused with LPET's low-frequency component ($\mathbf{LL}_c$) to reconstruct the global contrast via the **denoising model**, while LPET's high-frequency components are enhanced through the **HFE** for fine details. Finally, the outputs are recombined through the IDWT.

Spatial Consistency Denoising Model: The denoising model reconstructs noise-free images from low-frequency inputs using a self-attention U-Net with spatial sequence guidance $\mathbf{LL}_S$ (Fig. 1(b)). leveraging spatial information is critical for PET image generation. However, previous works have been insufficient in this regard, as they either employed simple convolutional networks or directly fed adjacent slices into the denoising model [7,19]. Our approach introduces a spatial consistency-informed denoising model featuring two novel modules: the spatial consistency feature extractor (SCFE) and the spatial consistency attention (SCA). These modules exploit adjacent slices to better capture spatial continuity, overcoming limitations of existing 2D-based diffusion and GAN models. Each encoder layer in the denoising model comprises a convolution block, the SCFE module, the SCA module, and a Resblock, while the bottleneck and decoder layers remain consistent with those in DDPM. The SCFE and SCA modules are detailed in the following sections.

Spatial Consistency Feature Extractor (SCFE)(Fig. 1(d)) takes three inputs: the convolution of the output from the previous encoder layer $\hat{F}^{l-1} = \mathrm{CONV}(F^{l-1})$, the spatial consistency guidance $\mathbf{LL}_S$, and the time embedding of the current time step. When $l = 1$, $F^{l-1} = F^0 = \mathrm{CONCAT}(\mathbf{LL}_\mathbf{g}^\mathbf{noise}, \mathbf{LL}_\mathbf{c})$ ($\mathbf{LL}_\mathbf{g}^\mathbf{noise}$ is the noise-corrupted version of $\mathbf{LL}_g$). SCFE employs a residual convolutional network for processing. To begin, the spatial sequence $\mathbf{LL}_S$ is aligned with F^{l-1} using a 1×1 convolution and then concatenated with $\hat{F}^{l-1}$). A specialized extraction block, consisting of multiple convolutional layers, extracts spatial consistency information by performing feature fusion and iterative noise adjustment. This extraction operation is applied twice within the SCFE to ensure robust feature refinement. At the final stage, the extracted spatial consistency features are integrated with the original input feature $\hat{F}^{l-1}$ through a residual connection, preserving the input's integrity while enhancing it with the extracted spatial information. Then the features are passed to SCA for further processing.

Spatial Consistency Attention (SCA) (Fig. 1(e)) enhances spatial feature extraction by leveraging multi-scale convolutional kernels and normalized attention weights. The input spatial features are first divided into v subgroups along the channel dimension, where v corresponds to the number of slices in the spatial sequence. Each subgroup is then processed by convolutional blocks with kernels of varying sizes to capture multi-resolution features effectively. The outputs of these convolutions are passed through a global pooling layer, followed by a weight generator consisting of fully connected layers, and finally normalized using a SoftMax layer to produce attention weights. The normalized attention weights are applied to the corresponding sub-groups of the convolutional outputs, and the weighted features are subsequently multiplied and refined. The weighted outputs from all sub-groups are concatenated as the module output.

High-frequency Enhancer: Employing a residual convolutional structure, our High-frequency Enhancer is a lightweight module aiming at enhancing details of LPET by integrating information across different high-frequency components. It takes $[\mathbf{LH}, \mathbf{HL}, \mathbf{HH}]_c$ as the input. First, the pairs $[\mathbf{LH}, \mathbf{HL}]_c$, and $[\mathbf{HL}, \mathbf{HH}]_c$ are processed through cross-attention blocks to capture interactions between high-frequency components. Next, both $\mathbf{LH}_c$, $\mathbf{HL}_c$, and the concatenation of cross-attention outputs undergo dilated convolution blocks. The resulting features are concatenated, passed through a depth convolution layer, and is concatenated with $[\mathbf{LH}, \mathbf{HL}, \mathbf{HH}]_c$ via a residual link.

2.4 Loss Function

We design a comprehensive loss function that incorporates multiple components to optimize the reconstruction process. We defined the noise loss as $\mathcal{L}_{\text{noise}} = \|\mathbf{n} - \hat{\mathbf{n}}\|_2^2$, minimizes the discrepancy between the Gaussian noise $\mathbf{n}$ added during the forward diffusion process and the predicted noise $\hat{\mathbf{n}}$. The reconstruction loss is given by $\mathcal{L}_{\text{img}} = \|\mathbf{x}^G - \hat{\mathbf{x}}^G\|_1$, where $\mathbf{x}^G$ is the SPET ground-truth and $\hat{\mathbf{x}}^G$ is the reconstruction PET image.

To enhance fine details and preserve high-frequency information, we introduce a high-frequency loss $\mathcal{L}_{\text{hf}}$, which combines an ℓ_2 loss for the high-frequency components with total variation (TV) regularization, defined as:

$$\mathcal{L}_{\text{hf}} = \lambda_{\ell_2} \|\mathbf{x}_{\text{hf}} - \hat{\mathbf{x}}_{\text{hf}}\|_2^2 + \lambda_{\text{tv}} \, \text{TV}(\hat{\mathbf{x}}_{\text{hf}}), \tag{4}$$

where λ_{ℓ_2} and λ_{tv} are the weights balancing the ℓ_2 loss and total variation loss. Here $\mathbf{x}_{\text{hf}}$ is the concatenation of the high-frequency components of the SPET image, served as the ground-truth, and $\hat{\mathbf{x}}_{\text{hf}}$ represents the high-frequency output.

The overall loss is then defined as:

$$\mathcal{L} = \mathcal{L}_{\text{noise}} + \mathcal{L}_{\text{hf}} + \mathcal{L}_{\text{img}}. \tag{5}$$

3 Experiments and Results

Dataset: The dataset used in this study is from the Ultra Low-Dose PET Imaging (UDPET) challenge [17] and consists of 560 ^{18}F-FDG PET scans with cor-

responding low-dose images acquired at various time reductions. Data were collected from 230 Siemens and 330 United Imaging subjects. Each brain scan has a resolution of $128 \times 128 \times 128$ with pixel spacings of $[1.65\,\text{mm}, 1.65\,\text{mm}, 1.65\,\text{mm}]$ (Siemens) and $[1.667\,\text{mm}, 1.667\,\text{mm}, 2.886\,\text{mm}]$ (United Imaging), and are used for testing and validation across protocols and scanner types. These brain images were split into training and test sets in a 3:1 ratio.

Experiment Details: We utilizes 1/20, 1/50, and 1/100 doses (ultra-low dose) of LPET images as the input, corresponding to normal-dose SPET as the ground truth. The framework is implemented using the PyTorch library and executed on an NVIDIA RTX 3090 GPU. During the training stage, random cropping and flipping are applied to the input images as data augmentation. The initial learning rate is 1×10^{-4} and the decay factor is 0.8 for every 5000 iterations.

Evaluation Metrics: To assess reconstruction quality, we compare our method with baselines using PSNR, SSIM, and NMSE. PSNR measures pixel differences (higher is better), SSIM evaluates perceptual quality, and NMSE quantifies overall error (lower is better). To further assess image detail recovery, we employ **Gradient Loss**, which emphasizes local structures and edges through first-order intensity changes, and **Brenner Gradient Loss**, which captures intricate textures and sharpness using second-order variations. Meanwhile, we measure the inference time of the model by calculating the average time to infer a full SPET 3D image from whole brain LPET slices.

3.1 Experimental Results

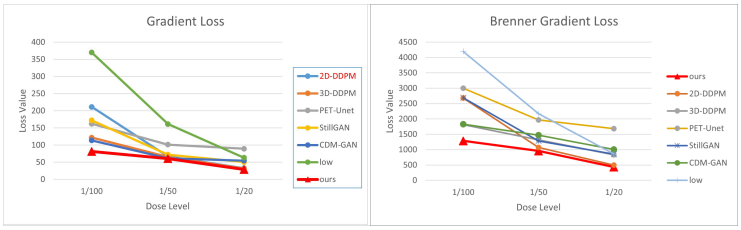
Table 1 shows a quantitative comparison of our method against baselines: LPET, 2D-DDPM[8], 3D-DDPM(cWDM) [5], Still-GAN[13], CDM-GAN[7], and PET-Unet[1]. These were chosen for their publicly available code.

Table 1. Quantitative comparison results.

Methods	1/100 dose			1/50 dose			1/20 dose			Inference Time (s/128 slices)
	PSNR	SSIM	NMSE ($\times 10^{-4}$)	PSNR	SSIM	NMSE ($\times 10^{-4}$)	PSNR	SSIM	NMSE ($\times 10^{-4}$)	
Low dose[17]	15.46 ± 5.75	0.46 ± 0.007	22.02 ± 9.78	21.70 ± 5.54	0.70 ± 0.004	22.73 ± 7.04	22.72 ± 5.18	0.77 ± 0.001	4.02 ± 3.96	–
2D-DDPM[8]	22.68 ± 4.82	0.76 ± 0.003	7.27 ± 1.77	25.86 ± 3.10	0.86 ± 0.002	5.62 ± 0.45	26.11 ± 2.69	0.92 ± 0.001	3.27 ± 0.77	268.00
3D-DDPM(cWDM)[5]	25.16 ± 4.39	0.85 ± 0.005	14.10 ± 6.10	27.40 ± 6.19	0.90 ± 0.001	4.20 ± 1.11	28.81 ± 2.85	0.93 ± 0.001	6.10 ± 0.81	108.05
Still-GAN[13]	23.61 ± 3.94	0.83 ± 0.003	9.69 ± 9.35	24.48 ± 2.66	0.85 ± 0.002	7.04 ± 1.91	25.63 ± 2.60	0.90 ± 0.001	3.96 ± 0.64	26.12
CDM-GAN[7]	23.95 ± 4.47	0.84 ± 0.005	12.62 ± 11.16	26.80 ± 2.73	0.85 ± 0.001	3.80 ± 0.39	28.84 ± 2.61	0.91 ± 0.001	3.12 ± 0.60	25.24
Pet-Unet[1]	23.30 ± 3.09	0.81 ± 0.003	8.43 ± 0.63	24.33 ± 2.62	0.84 ± 0.002	6.29 ± 0.40	27.93 ± 2.60	0.91 ± 0.001	2.60 ± 0.77	**2.02**
WiD-PET(Ours)	**26.68 ± 3.08**	**0.89 ± 0.001**	**3.60 ± 0.58**	**27.82 ± 2.53**	**0.91 ± 0.001**	**3.11 ± 0.35**	**29.93 ± 2.55**	**0.94 ± 0.001**	**1.85 ± 0.56**	20.24
p-value (Pair-T with the second-best)	6e-4	3e-4	6e-4	3e-4	1e-4	1e-5	3e-5	1e-5	3e-5	–

Overall Quality As shown in Table 1, all comparison methods significantly improve the quality of LPET images. Among them, our proposed method consistently achieves superior performance across 1/20, 1/50, and 1/100 dose levels. Taking the 1/100 dose level (ultra-low dose) as an example, our WiD-PET significantly enhances the PSNR/SSIM/NMSE values from 15.46/0.46/22.02 of LPET to 26.68/0.89/3.60 in the reconstructed SPET images. This represents a substantial improvement over the second-best performer, 3D-DDPM, which achieves

25.16/0.85/14.10. Paired t-test results confirm that the improvements in our method are statistically significant ($p<0.05$). Meanwhile, among 2D methods, although the general-purpose 2D diffusion-based model 2D-DDPM outperforms the 2D GAN-based StillGAN, both fall short compared to the specialized 2D-PET reconstruction model, CDM-GAN. In contrast, WiD-PET demonstrates superior performance at enhancing high-frequency details while leveraging spatial consistency across slices to reduce artifacts and improve visual smoothness.



(a) Gradient Loss Comparison (b) Brenner Gradient Loss Comparison

Fig. 2. Comparison of (a) Gradient Loss and (b) Brenner Gradient Loss.

Detail Recovery To assess high-frequency detail preservation, we evaluate **gradient loss** and **Brenner gradient loss**, which measure local structure and edge recovery via first- and second-order differences. As shown in Fig. 2, WiD-PET achieves the lowest losses across all dose levels. At moderate doses (1/20 and 1/50), diffusion-based models (2D-DDPM, 3D-DDPM, and WiD-PET) outperform GAN-based methods in capturing fine details. However, at the ultra-low dose (1/100), both vanilla 2D- and 3D-DDPM struggle—especially 2D-DDPM, which shows a significantly higher loss. In contrast, WiD-PET's tailored strategies substantially enhance detail recovery.

Inference Speed Our WiD-PET demonstrates remarkable efficiency during inference. As shown in Table 1, we achieved a **10-fold** speedup in 2D-DDPM, making it **5 times** faster than the wavelet-informed 3D-DDPM. This speed surpasses both GAN-based models and is second only to Pet-Unet, which, despite its faster inference, delivers significantly inferior reconstruction quality.

Visual Comparison A visual comparison in Fig. 3 further confirms that our reconstructions are sharper and more detailed, closely resembling the ground truth. These findings align with our quantitative analysis.

Ablation studies: We conducted ablation studies (Table 2) to assess the contributions of individual components. Starting from a vanilla 2D-DDPM (Base Model), adding the wavelet transform with the HFE improved reconstruction quality. Incorporating Spatial-consistency guidance via the SCFE and SCA modules further enhanced performance, and the introduction of a high-frequency loss term provided additional gains. The full integration of these components in WiD-PET yields the best overall results, validating the contribution of each module.

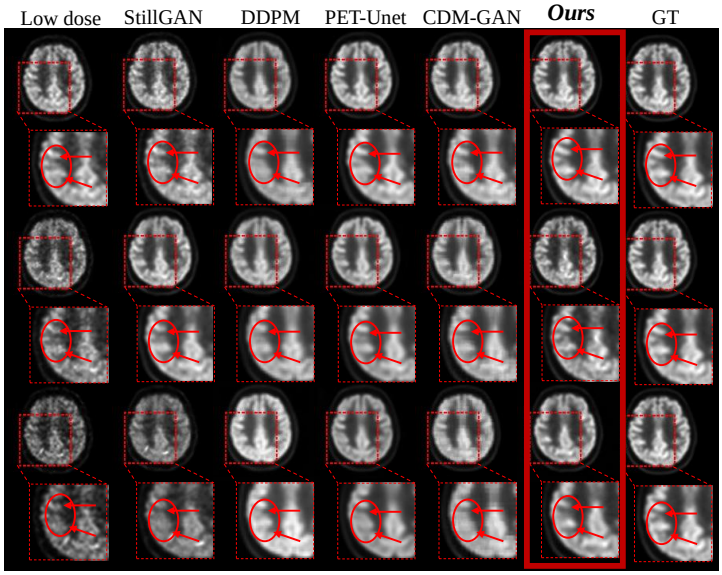


Fig. 3. Visual comparison of reconstruction results of 2D-methods.

Table 2. Results of ablation experiments on both dose levels from the UDPET dataset. We removed the wavelet transformation components, SCFE, and high-frequency loss from the base diffusion model to assess the performance impact.

Components				1/100 dose			1/50 dose			1/20 dose		
Base Model	Wavelet Transform and Enhancer	Spatial Guidance	High-frequency Loss	PSNR	SSIM	NMSE ($\times 10^{-4}$)	PSNR	SSIM	NMSE ($\times 10^{-4}$)	PSNR	SSIM	NMSE ($\times 10^{-4}$)
✓	·	·	·	22.68	0.76	7.27	25.86	0.86	5.62	26.11	0.82	3.27
✓	✓	·	·	25.19	0.79	6.50	25.26	0.86	5.47	28.33	0.71	2.30
✓	✓	✓	·	25.41	0.79	5.81	25.28	0.79	5.18	28.58	0.84	2.24
✓	✓	·	✓	25.91	0.86	4.63	26.16	0.87	4.57	28.69	0.92	2.22
✓	✓	✓	✓	**26.68**	**0.89**	**3.60**	**27.82**	**0.91**	**3.11**	**29.93**	**0.94**	**1.85**

4 Conclusion

In this paper, we tackle three challenges in SPET reconstruction from LPET inputs with our novel wavelet-informed fast diffusion framework. By integrating Haar discrete wavelet transformation and a high-frequency enhancer, we largely reduce inference time and improve detail recovery. Additionally, our spatial consistency modules (SCFE and SCA) enhance spatial continuity. Our proposed method, WiD-PET, consistently outperforms state-of-the-art approaches across different dose levels, demonstrating superior quality and generalizability.

Disclosure of Interest

The authors declare that they have no conflict of interest.

References

1. Chen, K., Gong, E., de Carvalho Macruz, F., Xu, J., Boumis, A., Khalighi, M., Poston, K., Sha, S., Greicius, M., Mormino, E., Pauly, J., Srinivas, S., Zaharchuk, G.: Ultra-low-dose 18f-florbetaben amyloid pet imaging using deep learning with multi-contrast mri inputs. Radiology **290**(3), 649–656 (Mar 2019). https://doi.org/10.1148/radiol.2018180940, epub 2018 Dec 11. Erratum in: Radiology. 2020 Sep;296(3):E195. doi: 10.1148/radiol.2020202527
2. Choudhury, C., Goel, T., Tanveer, M.: A coupled-gan architecture to fuse mri and pet image features for multi-stage classification of alzheimer's disease. Information Fusion **109**, 102415 (2024). https://doi.org//10.1016/j.inffus.2024.102415, https://www.sciencedirect.com/science/article/pii/S1566253524001933
3. Cui, J., Xie, Y., Guo, N., Feng, Y., Li, Q.: Pet image denoising using consistent denoising diffusion model. Journal of Nuclear Medicine **65**(supplement 2), 241799–241799 (2024), https://jnm.snmjournals.org/content/65/supplement_2/241799
4. Cui, J., Wang, Y., Wen, L., Zeng, P., Wu, X., Zhou, J., Shen, D.: Image2points: A 3d point-based context clusters gan for high-quality pet image reconstruction. In: ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). p. 1726–1730. IEEE (Apr 2024). https://doi.org/10.1109/icassp48485.2024.10446360, http://dx.doi.org/10.1109/ICASSP48485.2024.10446360
5. Friedrich, P., Durrer, A., Wolleb, J., Cattin, P.C.: cwdm: Conditional wavelet diffusion models for cross-modality 3d medical image synthesis (2024), https://arxiv.org/abs/2411.17203
6. Gong, K., Johnson, K., El Fakhri, G., Li, Q., Pan, T.: Pet image denoising based on denoising diffusion probabilistic model. European Journal of Nuclear Medicine and Molecular Imaging **51**(2), 358–368 (2024), https://doi.org/10.1007/s00259-023-06417-8
7. Han, Z., Wang, Y., Zhou, L., Wang, P., Yan, B., Zhou, J., Wang, Y., Shen, D.: Contrastive diffusion model with auxiliary guidance for coarse-to-fine pet reconstruction (2023), https://arxiv.org/abs/2308
8. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models (2020), https://arxiv.org/abs/2006.11239
9. Islam, J., Zhang, Y.: Gan-based synthetic brain PET image generation. Brain Informatics **7**(1), 3 (Mar 2020). https://doi.org/10.1186/s40708-020-00104-2, https://doi.org/10.1186/s40708-020-00104-2
10. Li, J., Cheng, B., Chen, Y., Gao, G., Shi, J., Zeng, T.: Ewt: Efficient wavelet-transformer for single image denoising. Neural Networks **177**, 106378 (2024). https://doi.org/https://doi.org/10.1016/j.neunet.2024.106378, https://www.sciencedirect.com/science/article/pii/S0893608024003022
11. Li, Z.N., Chen, X.H., Guo, S.N., Wang, S.Q., Pun, C.M.: Wavenhancer: Unifying wavelet and transformer for image enhancement. Journal of Computer Science and Technology **39**(2), 336–345 (2024), https://doi.org/10.1007/s11390-024-3414-z
12. Luo, Y., Zhou, L., Zhan, B., Fei, Y., Zhou, J., Wang, Y., Shen, D.: Adaptive rectification based adversarial network with spectrum constraint for high-quality pet image synthesis. Medical Image Analysis **77**, 102335 (2022). https://doi.org/https://doi.org/10.1016/j.media.2021.102335, https://www.sciencedirect.com/science/article/pii/S1361841521003807

13. Ma, Y., Liu, J., Liu, Y., Fu, H., Hu, Y., Cheng, J., Qi, H., Wu, Y., Zhang, J., Zhao, Y.: Structure and illumination constrained gan for medical image enhancement. IEEE Transactions on Medical Imaging **40**(12), 3955–3967 (2021). https://doi.org/10.1109/TMI.2021.3101937
14. Metz, L., Poole, B., Pfau, D., Sohl-Dickstein, J.: Unrolled generative adversarial networks. CoRR **abs/1611.02163** (2016), http://arxiv.org/abs/1611.02163
15. Pan, S., Wang, T., Qiu, R.L., Axente, M., Chang, C.W., Peng, J., Patel, A.B., Shelton, J., Patel, S.A., Roper, J., et al.: 2d medical image synthesis using transformer-based denoising diffusion probabilistic model. Physics in Medicine & Biology **68**(10), 105004 (2023), https://doi.org/10.1088/1361-6560/acca5c
16. Salimans, T., Goodfellow, I., Zaremba, W., Cheung, V., Radford, A., Chen, X.: Improved techniques for training gans (2016), https://arxiv.org/abs/1606.03498
17. Shi, K., Guo, R., Xue, S., Rominger, A., Li, B.: Ultra-low dose pet imaging challenge 2022 (2022), https://zenodo.org/records/6361846
18. Sun, H., Jiang, Y., Yuan, J., Wang, H., Liang, D., Fan, W., Hu, Z., Zhang, N.: High-quality pet image synthesis from ultra-low-dose pet/mri using bi-task deep learning. Quantitative Imaging in Medicine and Surgery **12**(12), 5326 (2022), https://doi.org/10.21037/qims-22-116
19. Xie, H., Gan, W., Zhou, B., Chen, M.K., Kulon, M., Boustani, A., Spencer, B.A., Bayerlein, R., Ji, W., Chen, X., Liu, Q., Guo, X., Xia, M., Zhou, Y., Liu, H., Guo, L., An, H., Kamilov, U.S., Wang, H., Li, B., Rominger, A., Shi, K., Wang, G., Badawi, R.D., Liu, C.: Dose-aware diffusion model for 3d low-dose pet: Multi-institutional validation with reader study and real low-dose data (2024), https://arxiv.org/abs/2405.12996
20. Yu, B., Zhou, L., Wang, L., Shi, Y., Fripp, J., Bourgeat, P.: Ea-gans: Edge-aware generative adversarial networks for cross-modality mr image synthesis. IEEE Transactions on Medical Imaging **38**(7), 1750–1762 (2019). https://doi.org/10.1109/TMI.2019.2895894
21. Zhao, K., Zhou, L., Gao, S., Wang, X., Wang, Y., Zhao, X., Wang, H., Liu, K., Zhu, Y., Ye, H.: Study of low-dose pet image recovery using supervised learning with cyclegan. Plos one **15**(9), e0238455 (2020), https://pubmed.ncbi.nlm.nih.gov/32886683/
22. Zhou, L., Schaefferkoetter, J.D., Tham, I.W., Huang, G., Yan, J.: Supervised learning with cyclegan for low-dose fdg pet image denoising. Medical Image Analysis **65**, 101770 (2020). https://doi.org//10.1016/j.media.2020.101770, https://www.sciencedirect.com/science/article/pii/S1361841520301341