CT-Conditioned Diffusion Prior with Physics-Constrained Sampling for PET Super-Resolution

Liutao Yang[1,2], Zi Wang[1], Peiyuan Jing[1,3], Xiaowen Wang[1], Javier A. Montoya-Zegarra[3], Kuangyu Shi[4], Daoqiang Zhang[2], and Guang Yang[1,5,6⋆]

[1] Department of Bioengineering and Imperial-X, Imperial College London, London, UK
{l.yang23,g.yang}@imperial.ac.uk
[2] College of Artificial Intelligence, Nanjing University of Aeronautics and Astronautics, Nanjing, China
[3] School of Engineering, Zurich University of Applied Sciences, Winterthur, Switzerland
[4] Department of Nuclear Medicine, Inselspital, University of Bern, Bern, Switzerland
[5] National Heart and Lung Institute, Imperial College London, London, UK
[6] Cardiovascular Research Centre, Royal Brompton Hospital, London, UK

Abstract. PET super-resolution is highly under-constrained because paired multi-resolution scans from the same subject are rarely available, and effective resolution is determined by scanner-specific physics (e.g., PSF, detector geometry, and acquisition settings). This limits supervised end-to-end training and makes purely image-domain generative restoration prone to physically inconsistent structures when anatomical and physical constraints are weak. We formulate PET super-resolution as posterior inference under heterogeneous system configurations and propose a CT-conditioned diffusion framework with physics-constrained sampling. During training, a conditional diffusion prior is learned from high-quality PET/CT pairs using cross-attention for anatomical guidance, without requiring paired LR–HR PET data. During inference, measurement consistency is enforced through a scanner-aware forward model with explicit PSF effects and gradient-based data-consistency refinement. Under both standard and OOD settings, the proposed method consistently improves experimental metrics and lesion-level clinical relevance indicators over strong baselines, while improving measurement consistency and structural fidelity.

Keywords: PET Super-resolution · Diffusion models · Inverse problems.

1 Introduction

Positron emission tomography (PET) is essential for functional imaging in oncology, neurology, and cardiology, but image quality is fundamentally limited

⋆ Corresponding author.

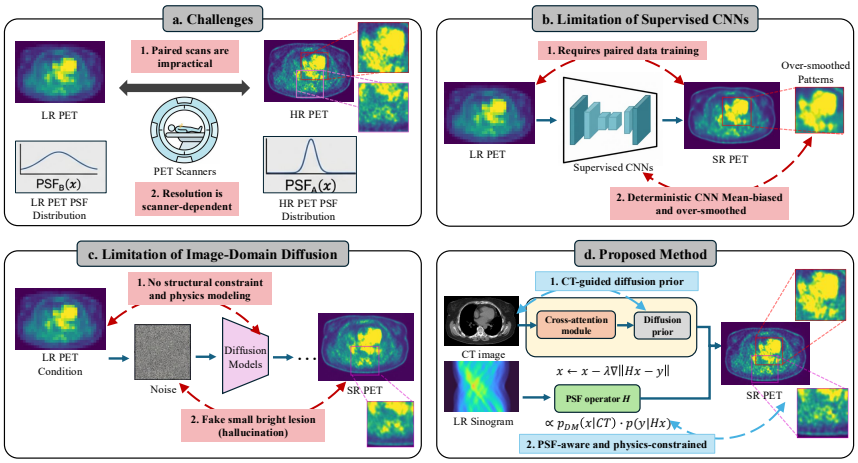

Fig. 1. Conceptual motivation and method positioning for PET super-resolution. (a) PET super-resolution is ill-posed due to scanner-dependent degradation and limited paired data. (b) Supervised CNN mapping depends on paired LR–HR data and tends to over-smooth details. (c) Image-domain diffusion can recover texture but may introduce physically inconsistent structures without structural/physics constraints. (d) Our method combines CT-conditioned priors with scanner-aware, PSF-aware consistency for faithful and physically consistent reconstruction.

by counting statistics, detector response, and acquisition geometry [14, 15, 3]. Increasing resolution or reducing dose/time amplifies noise and blur, yielding a persistent sensitivity–resolution trade-off [15, 3].

PET super-resolution is highly ill-posed because degradation is scanner dependent (e.g., PSF and acquisition configuration), not a fixed downsampling rule. Paired multi-resolution scans are therefore scarce, especially across heterogeneous systems, making supervised LR→HR mapping difficult to generalize. Classical and iterative reconstruction methods enforce physics and Poisson statistics [15, 13], while supervised deep models improve appearance but remain sensitive to domain shift and hallucination under mismatch [12, 11, 21].

Diffusion-based PET reconstruction combines learned priors with data consistency [10, 18, 4, 17, 20, 1, 9], but three gaps remain for PET super-resolution: limited cross-modal anatomical guidance, weak modeling of heterogeneous scanner operators, and predominantly image-domain restoration with limited measurement domain guarantees [11, 21]. Because reconstructed appearance entangles system response and post-processing, robust PET super-resolution should be constrained by the low-quality measurement and its scanner-specific forward model. To address these limitations, we reformulate PET super-resolution as posterior inference under heterogeneous scanner configurations. The proposed framework combines a CT-conditioned diffusion prior with physics-constrained sampling: CT cross-attention provides anatomical guidance to mitigate physically inconsistent restoration artifacts, while PSF-aware measurement-domain

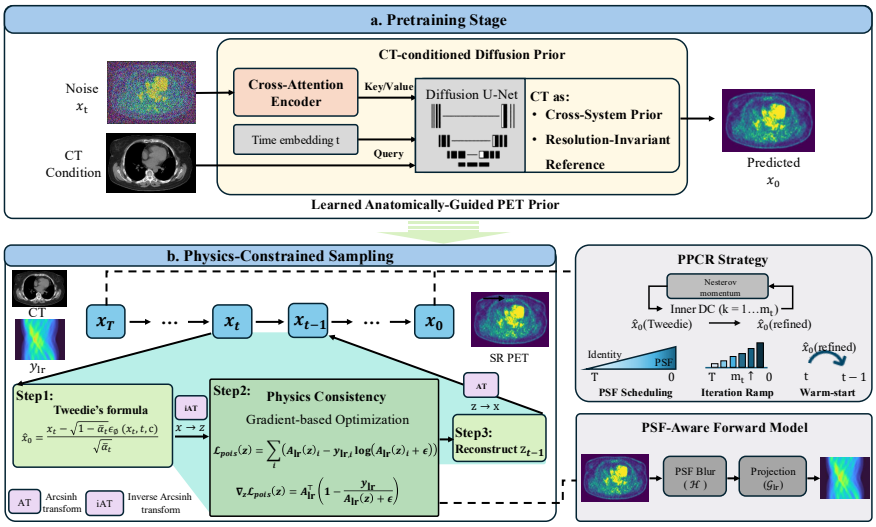

Fig. 2. Overview of the proposed CT-guided diffusion framework. **Pretraining (a):** learn a CT-conditioned diffusion prior from high-quality PET/CT. **Inference (b):** start from noisy PET, iteratively denoise, and apply measurement-domain data-consistency updates. **Progressive Physics-Constrained Refinement (PPCR):** progressively enforces physics constraints (late activation and coarse-to-fine PSF) to produce a scanner-consistent high-resolution PET output.

data consistency constrains samples to scanner-consistent solutions. This design targets the key failure modes of prior methods by jointly improving structural fidelity, physical plausibility, and robustness across standard and out-of-distribution degradation settings.

The main contributions of this work are summarized as follows:

- We formulate PET super-resolution as a posterior inference problem that progressively constrains the solution space using a CT-conditional diffusion prior and PSF-based data consistency.
- We highlight the distinction between underlying activity estimation and reconstructed image appearance, enabling physically consistent reconstruction without paired multi-resolution datasets.
- We demonstrate that the proposed framework outperforms supervised end-to-end models and image-domain diffusion approaches in both structural fidelity and quantitative accuracy.

2 Methods

2.1 Problem Formulation

Let $z \in \mathbb{R}^N$ denote the tracer activity distribution in physical units. For scanner configuration k, PET measurements follow:

$$y_k \sim \text{Poisson}(A_k z + b_k), \tag{1}$$

where A_k denotes the scanner-specific forward operator and b_k represents background contributions (e.g., random and scatter events). In the following, the low-quality acquisition is denoted by y_{lr} and its forward operator by A_{lr}.

Rather than learning an explicit mapping between resolutions, we formulate PET super-resolution as posterior inference:

$$p(z \mid y_{\text{lr}}, c) \propto p_\phi(z \mid c)\, p(y_{\text{lr}} \mid z), \tag{2}$$

where $p_\phi(z \mid c)$ is a learned structural prior, with c denoting the co-registered CT anatomy used for conditioning; and $p(y_{\text{lr}} \mid z)$ is defined by scanner-specific physics (including PSF effects, projection geometry, and Poisson noise statistics).

2.2 Score-Based CT-Conditional Prior

We model $p_\phi(z \mid c)$ using a conditional diffusion model [10, 18]. Diffusion models learn the score function $\nabla_x \log p_\phi(x \mid c)$ of the data distribution in a normalized image space.

Given a physical activity map z, we map it to model space via an arcsinh transform:

$$x = \frac{\text{asinh}(z/s)}{\kappa}. \tag{3}$$

Here, κ is a scaling constant. The arcsinh transform stabilizes PET's high dynamic range: it behaves approximately logarithmically at high uptake while remaining well behaved near zero.

During training, Gaussian noise is added to x, and a neural network $\epsilon_\phi(x_t, t, c)$ is trained to predict the injected noise. This is equivalent to learning the conditional score under the diffusion framework [18].

CT conditioning is incorporated via cross-attention in a Transformer-style formulation [19]. CT is encoded into a feature pyramid; cross-attention at the 8×-downsampled encoder, bottleneck, and decoder blocks uses PET features as queries and resolution-matched CT features as keys/values, adding the attended context residually. The prior is learned exclusively from high-quality PET/CT pairs, without requiring degradation modeling.

2.3 Physics-Constrained Likelihood Modeling

We separate the linear measurement operator from the expected measurement:

$$A_{\text{lr}} = \mathcal{G}_{\text{lr}} \mathcal{H}, \qquad \bar{y}_{\text{lr}}(z) = A_{\text{lr}} z + b_{\text{lr}} = \mathcal{G}_{\text{lr}}(\mathcal{H} z) + b_{\text{lr}}, \tag{4}$$

Here, $\mathcal{H}$ is the PSF operator, $\mathcal{G}_{\text{lr}}$ is the LR measurement operator, and b_{lr} models random/scatter background. As the PSF enters only inference-time DC,

measured scanner-specific or spatially varying PSFs and their adjoints can be substituted without retraining.

Under a Poisson measurement model, the negative log-likelihood (up to constants) is:

$$\mathcal{L}_{\text{Pois}}(z) = \sum_i \left(\bar{y}_{\text{lr}}(z)_i - y_{\text{lr},i} \log\left(\bar{y}_{\text{lr}}(z)_i + \epsilon\right)\right), \tag{5}$$

where $\epsilon > 0$ is a small numerical constant. Its gradient is computed analytically as:

$$\nabla_z \mathcal{L}_{\text{Pois}}(z) = A_{\text{lr}}^\top \left(1 - \frac{y_{\text{lr}}}{\bar{y}_{\text{lr}}(z) + \epsilon}\right) = \mathcal{H}^\top \mathcal{G}_{\text{lr}}^\top \left(1 - \frac{y_{\text{lr}}}{\bar{y}_{\text{lr}}(z) + \epsilon}\right). \tag{6}$$

Since the background is constant, it does not enter the gradient; this explicit adjoint form matches the implementation.

2.4 Tweedie-Guided Sampling and Progressive Physics-Constrained Refinement (PPCR)

Reverse diffusion sampling can be interpreted as approximate posterior sampling under a learned score prior [18, 4].

At timestep t, the diffusion model predicts noise, from which a denoised estimate can be obtained via Tweedie's formula [6]:

$$\hat{x}_0 = \frac{x_t - \sqrt{1 - \bar{\alpha}_t}\, \epsilon_\phi(x_t, t, c)}{\sqrt{\bar{\alpha}_t}}. \tag{7}$$

The quantity $\hat{x}_0$ serves as a Tweedie's estimate of the clean image under the learned prior.

To incorporate measurement information, we refine this Tweedie estimate using likelihood-gradient descent in activity space:

$$z^{\text{refined}} = z^{\text{Tweedie}} - \eta_t \nabla_z \mathcal{L}_{\text{Pois}}(z^{\text{Tweedie}}). \tag{8}$$

We combine the Tweedie update with Progressive Physics-Constrained Refinement (PPCR), which schedules no-PSF to full-PSF data consistency (DC), increases inner iterations, and uses Nesterov acceleration with warm-start [7, 2]. The prior proposes HR candidates; PSF-aware DC matches their degraded predictions to the LR PET rather than creating resolution itself.

Formally, let m_t denote the number of inner DC iterations and let $\mathcal{H}_t$ denote the scheduled PSF operator (identity in early steps, full PSF in late steps). The DC update at timestep t is:

$$z_t^{(k+1)} = z_t^{(k)} - \eta_t \nabla_z \mathcal{L}_{\text{Pois}}^{\mathcal{H}_t}(z_t^{(k)}), \quad k = 1, \ldots, m_t, \tag{9}$$

with m_t increasing over t. Early no-PSF DC stabilizes coarse uptake, whereas late full-PSF DC enforces measurement-faithful HR refinement [18, 4, 17].

Table 1. Degradation settings for standard and out-of-distribution (OOD) evaluation.

Setting	FWHM	Dose	Angular	Radial	Target Spacing	SR Factor
Standard	8 mm	10%	×2	×2	8 mm	×4
OOD	12 mm	5%	×3	×2	12 mm	×6

3 Experiments

3.1 Experimental Setup

Dataset Experiments were conducted on the FDG-PET-CT-Lesions dataset [8] (TCIA). For this study, 120 cases were randomly split into 100/10/10 for training/validation/testing. For supervised comparison methods, paired LR–HR training data are generated using simulated LR data from the training split. Our method is trained without simulated LR–HR PET pairs. All methods are then evaluated on the same unified simulated test settings for fair comparison. Detailed simulation settings are described in Degradation Protocol.

Degradation Protocol LR measurements were generated by a physics-based pipeline, rather than plain downsampling, with Gaussian PSF blur, projection geometry, angular/radial rebinning, and dose reduction. Table 1 defines standard SR×4 and OOD SR×6 settings. Their 8/12 mm values are effective image-domain degradations, not intrinsic detector resolutions.

Implementation Details A 2D CT-conditioned diffusion UNet was pretrained as the PET prior. DDIM used 50 steps: no-PSF DC at 1–35, full-PSF DC at 36–50, 2–20 linearly ramped inner iterations, Nesterov $\mu = 0.9$, warm-start $\alpha = 0.3$, and DC step size 0.05. Code and configurations are available at: https://github.com/yan4243/CT-Conditioned-Diffusion-PET.

3.2 Comparison with State-of-the-Art

Compared Methods The proposed method was compared against representative supervised and unsupervised PET super-resolution approaches. **Supervised methods** include UNet [16], SRCNN [5], and DeepSR [12]. **Unsupervised methods** include ScorePET [17], PETDM [11], and SRDPM [21]. ScorePET uses a PET score prior with likelihood DC; our posterior targets heterogeneous scanner-dependent LR observations using CT conditioning, PSF-aware degradation, and PPCR. All supervised models were retrained on the same training split. We implemented the unsupervised and diffusion-based methods using the configurations in their respective publications.

Table 2. Quantitative comparison under standard (8 mm, SR×4) and OOD (12 mm, SR×6) settings. Best results are **bolded**; second-best are underlined.

	8 mm (SR×4)			**12 mm (SR×6)**		
Method	PSNR↑	SSIM↑	NMSE↓	PSNR↑	SSIM↑	NMSE↓
LR	36.63±4.56*	0.887±0.04*	0.09±0.05*	34.98±3.75*	0.849±0.05*	0.09±0.05*
DeepSR [12]	42.49±3.70*	0.957±0.02*	0.03±0.03*	37.27±3.31*	0.896±0.04*	0.06±0.05*
UNet [16]	39.50±4.55*	0.929±0.05*	0.06±0.07*	36.35±3.81*	0.877±0.07*	0.09±0.10*
SRCNN [5]	42.75±3.70*	0.960±0.02*	0.03±0.03*	37.33±3.32*	0.899±0.04*	0.06±0.05*
SRDPM [21]	40.30±3.29*	0.923±0.03*	0.04±0.05*	36.74±3.08*	0.877±0.05*	0.08±0.05*
PETDM [11]	41.32±3.27*	0.924±0.03*	0.04±0.05*	36.84±3.13*	0.879±0.05*	0.08±0.05*
ScorePET [17]	43.97±3.84*	0.970±0.02*	0.02±0.02*	37.85±3.35*	0.912±0.04*	0.05±0.04*
Ours	**46.84±3.08**	**0.980±0.01**	**0.01±0.01**	**40.66±2.86**	**0.945±0.03**	**0.03±0.03**

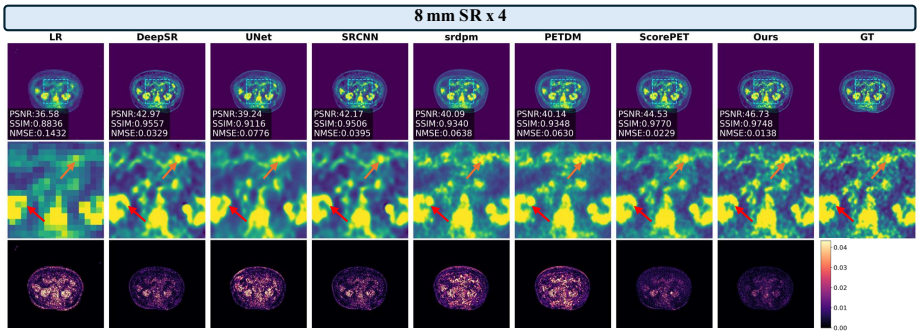

Fig. 3. Qualitative comparison under the standard setting (8 mm, SR×4).

Quantitative and Qualitative Evaluation Reconstruction performance was evaluated using PSNR, SSIM, and NMSE. Quantitative results for both standard and OOD settings are summarized in Table 2. Under standard (8 mm, SR×4), ScorePET is the strongest baseline, while under OOD (12 mm, SR×6) all baselines degrade, especially supervised methods. Our method remains best in both settings, indicating stronger fidelity and robustness under increased degradation. Qualitative examples are shown only for the standard setting in Fig. 3, while OOD comparisons are reported in Table 2. CNN-based methods over-smooth lesion boundaries, and diffusion baselines recover texture but show less structural stability; in contrast, our method preserves finer details with better quantitative consistency.

Clinical Relevance Analysis To assess clinical relevance beyond whole-image metrics, we further performed lesion-level analysis (Fig. 4) on two lung cancer cases. Across both lesions, our method shows the smallest SUVmax/SUVmean

Table 3. Ablation study under the standard degradation setting (8 mm, SR×4). "CT cond." indicates the conditioning strategy. Best results are **bolded**.

	Components				Metrics		
Variant	CT cond.	DC	PSF	Sched.	PSNR↑	SSIM↑	NMSE↓
w/o DC	CrossAttn				43.92±3.45	0.962±0.02	0.018±0.02
w/o PSF	CrossAttn	✓		✓	45.11±3.26	0.971±0.02	0.014±0.01
w/o PPCR	CrossAttn	✓	✓		45.78±3.18	0.975±0.01	0.012±0.01
CT Concat	Concat	✓	✓	✓	45.94±3.15	0.976±0.01	0.011±0.01
Ours	CrossAttn	✓	✓	✓	**46.84±3.08**	**0.980±0.01**	**0.010±0.01**

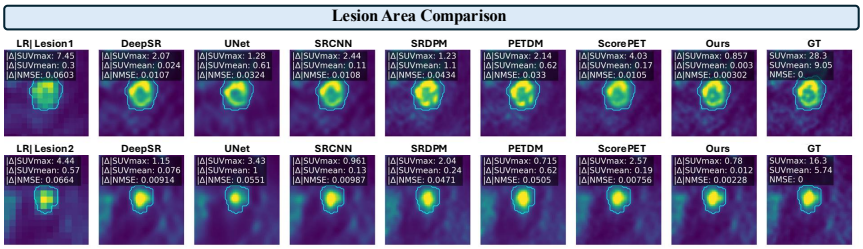

Fig. 4. Lesion-level quantitative comparison on two lung cancer cases. Each lesion reports ΔSUVmax, ΔSUVmean, and lesion NMSE within the lesion mask; lower values indicate better consistency.

deviations and the lowest lesion NMSE among compared methods, indicating improved lesion-specific quantitative fidelity beyond visual enhancement.

3.3 Ablation Study

We conducted ablation experiments under the standard setting (8 mm, SR×4) to evaluate the contributions of DC, PSF modeling, PPCR, and CT conditioning (Table 3). Evaluated variants include **w/o DC** (remove physics-based data consistency), **w/o PSF** (keep DC but remove explicit PSF modeling), **w/o PPCR** (keep DC/PSF but disable progressive scheduling), and **CT Concat** (replace cross-attention with feature concatenation).

Removing DC causes the largest drop, showing that measurement consistency is the dominant factor. Removing PSF or PPCR further degrades performance, confirming the importance of physically accurate blur modeling and progressive enforcement. Replacing cross-attention with concatenation also reduces performance, indicating that cross-attention provides stronger CT-guided structural conditioning.

4 Conclusion

We proposed physics-constrained diffusion sampling with a CT-conditioned PET prior for PET super-resolution. The method improves over supervised and diffusion baselines under standard and OOD degradations. Ablations confirm the roles of data consistency, PSF modeling, and CT cross-attention. As a methods paper, the study focuses on the operator-based framework under controlled 2D evaluation; validation with measured scanner PSFs and fully 3D clinical pipelines is outside the current scope.

Acknowledgements This study was supported by the Basic Research Program of Jiangsu (BK20251400), the Jiangsu Funding Program for Excellent Postdoctoral Talent, UKRI grants MC/PC/21013, EP/Z002206/1, MR/U506710/1, and MR/V023799/1; ERC IMI (101005122); H2020 (952172); the Royal Society (IEC/NSFC/211235), the NVIDIA Academic Hardware Grant Program, Boehringer Ingelheim (SABER), the NIHR Imperial Biomedical Research Centre (RDA01), the Wellcome Leap Dynamic Resilience program (co-funded by Temasek Trust), and the Imperial College London Seeds for Success Fund.

Disclosure of Interests The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Bae, S., Yu, B., Lee, J.S., Gong, K.: PET image reconstruction with diffusion priors and half-quadratic splitting. In: Medical Imaging 2025: Physics of Medical Imaging. p. 99. SPIE (2025). https://doi.org/10.1117/12.3047504
2. Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J.: Distributed optimization and statistical learning via the alternating direction method of multipliers. Foundations and Trends in Machine Learning **3**(1), 1–122 (2011). https://doi.org/10.1561/2200000016
3. Cherry, S.R., Jones, T., Karp, J.S., Qi, J., Moses, W.W., Badawi, R.D.: Total-body PET: Maximizing sensitivity to create new opportunities for clinical research and patient care. Journal of Nuclear Medicine **59**(1), 3–12 (2018). https://doi.org/10.2967/jnumed.116.184028
4. Chung, H., Kim, J., McCann, M.T., Klasky, M.L., Ye, J.C.: Diffusion posterior sampling for general noisy inverse problems. arXiv preprint arXiv:2209.14687 (2022), https://arxiv.org/abs/2209.14687
5. Dong, C., Loy, C.C., He, K., Tang, X.: Learning a deep convolutional network for image super-resolution. In: Computer Vision – ECCV 2014. pp. 184–199. Springer (2014). https://doi.org/10.1007/978-3-319-10593-2_13
6. Efron, B.: Tweedie's formula and selection bias. Journal of the American Statistical Association **106**(496), 1602–1614 (2011). https://doi.org/10.1198/jasa.2011.tm11181

7. Friedman, J., Hastie, T., Tibshirani, R.: Regularization paths for generalized linear models via coordinate descent. Journal of Statistical Software **33**(1), 1–22 (2010). https://doi.org/10.18637/jss.v033.i01
8. Gatidis, S., Hepp, T., Früh, M., La Fougère, C., Nikolaou, K., Pfannenberg, C., Schölkopf, B., Küstner, T., Cyran, C., Rubin, D.: A whole-body FDG-PET/CT dataset with manually annotated tumor lesions. Scientific Data **9**(1), 601 (2022). https://doi.org/10.1038/s41597-022-01718-3
9. Hashimoto, F., Gong, K.: PET image reconstruction using deep diffusion image prior. IEEE Transactions on Medical Imaging **45**(6), 2628–2638 (2026). https://doi.org/10.1109/TMI.2026.3659792
10. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. In: Advances in Neural Information Processing Systems. vol. 33, pp. 6840–6851. Curran Associates, Inc. (2020)
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: Medical Image Computing and Computer Assisted Intervention – MICCAI 2023. pp. 3–12. Springer (2023). https://doi.org/10.1007/978-3-031-43907-0_1
12. Özaltan, C.K., Türkölmez, E., Namer, I.J., Çiçek, A.E., Aksoy, S.: Deep convolutional neural networks for PET super-resolution. In: Medical Imaging 2024: Image Processing. p. 113. SPIE (2024). https://doi.org/10.1117/12.3007549
13. Qi, J., Leahy, R.M.: Iterative reconstruction techniques in emission computed tomography. Physics in Medicine & Biology **51**(15), R541–R578 (2006). https://doi.org/10.1088/0031-9155/51/15/R01
14. Rahmim, A., Zaidi, H.: PET versus SPECT: Strengths, limitations and challenges. Nuclear Medicine Communications **29**(3), 193–207 (2008). https://doi.org/10.1097/MNM.0b013e3282f3a515
15. Reader, A.J., Zaidi, H.: Advances in PET image reconstruction. PET Clinics **2**(2), 173–190 (2007). https://doi.org/10.1016/j.cpet.2007.08.001
16. Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. pp. 234–241. Springer (2015). https://doi.org/10.1007/978-3-319-24574-4_28
17. Singh, I.R.D., Denker, A., Barbano, R., Kereta, Ž., Jin, B., Thielemans, K., Maass, P., Arridge, S.: Score-based generative models for PET image reconstruction. Machine Learning for Biomedical Imaging **2**(Generative Models), 547–585 (2024). https://doi.org/10.59275/j.melba.2024-5d51
18. Song, Y., Sohl-Dickstein, J., Kingma, D.P., Kumar, A., Ermon, S., Poole, B.: Score-based generative modeling through stochastic differential equations. In: International Conference on Learning Representations (2021)
19. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. In: Advances in Neural Information Processing Systems. vol. 30, pp. 5998–6008. Curran Associates, Inc. (2017)
20. Webber, G., Mizuno, Y., Howes, O.D., Hammers, A., King, A.P., Reader, A.J.: Likelihood-scheduled score-based generative modeling for fully 3D PET image reconstruction. IEEE Transactions on Medical Imaging **44**(11), 4445–4456 (2025). https://doi.org/10.1109/TMI.2025.3576483
21. Zhou, Z., Song, T.A., Yang, F., Lois, C., Johnson, K., Dutta, J.: Super resolution Tau PET using a diffusion probabilistic model. Journal of Nuclear Medicine **65**(Supplement 2), 242184 (Jun 2024), https://jnm.snmjournals.org/content/65/supplement_2/242184.abstract