Evaluating Denoising Diffusion Probabilistic Models for PET Denoising: A Critical Analysis in a Limited Data Regime

Bram Popelier*, Ide Van den Borre*, Simon DeKeyser†, Vic De Ridder†, Aleksandra Pižurica*
*Department of Telecommunications and Information Processing, Group for Artificial Intelligence and Sparse Modelling
Faculty of Engineering and Architecture, Ghent University, Belgium
† Nuclivision.com
{Bram.Popelier; Ide.VandenBorre; Aleksandra.Pizurica}@UGent.be

***Abstract*—Positron Emission Tomography (PET) is a key imaging technique in clinical practice offering unique functional insights. However, its broader applicability is limited by radiation exposure and lengthy scan times, and reducing either degrades image quality by increasing noise. Over recent years, U-Net-based convolutional neural networks have become dominant approaches for PET denoising. These methods show good performance in denoising, but they offer limited uncertainty handling due to their deterministic nature and are prone to oversmoothing. This paper critically evaluates the applicability of denoising diffusion probabilistic models (DDPMs) for low-dose PET image denoising, particularly when training data is limited. We adapted a conditional DDPM architecture to the PET context and compared its performance to a U-Net baseline on clinical PET data. The conditional DDPM was both deployed as a standalone method and in a hybrid-ensemble format where it was placed in series with the baseline U-Net. The conditional DDPM, while promising in theory, introduced unrealistic features in some outputs and was computationally intensive. The hybrid model, intended to combine both strengths, underperformed due to high sample variability in DDPM outputs. Contrary to recent studies suggesting DDPM superiority, our experiments demonstrate that DDPMs underperform relative to U-Net, showing inferior PSNR and SSIM and introducing notable artifacts. Our findings highlight the importance of training dataset size and quality for DDPM effectiveness and provide practical guidelines regarding the trustworthiness of diffusion models for clinical PET denoising.**

***Index Terms*—PET, DDPM, U-Net, image denoising**

I. Introduction

Objective and Research Methodology

Positron Emission Tomography (PET) plays an indispensable role in modern clinical diagnostics, with applications spanning oncology, cardiology, and neurology. Clinical demand is rising steadily, yet current PET protocols require long scan times (15–30 minutes) and expose patients to high radiation doses (8–30 mSv), creating bottlenecks in imaging infrastructure and raising safety concerns [1, 7]. This challenge is especially critical for radiosensitive populations such as pediatric and pregnant patients [10, 6]. Market projections reflect this pressure, with the global PET scanner market growing rapidly and becoming increasingly integrated into mainstream diagnostics [3]. Reducing scan time and dose is therefore both a clinical and economic imperative—but doing so inevitably increases Poisson noise, degrading image quality and motivating the need for robust denoising methods.

Deep learning models, particularly U-Net architectures [4], have shown strong performance in denoising PET images. However, these models are deterministic and often prone to oversmoothing, potentially suppressing subtle diagnostic features. In contrast, Denoising Diffusion Probabilistic Models (DDPMs) are generative and stochastic, allowing them to model complex noise distributions, preserve finer structures, and generate uncertainty estimates—making them conceptually attractive for clinical imaging tasks.

Previous work has demonstrated promising results using 3D DDPMs trained on large synthetic or clinical datasets with substantial computational resources [12, 2, 14, 13]. However, their applicability under realistic constraints—limited training data, clinical PET input only, and 2D models—remains underexplored.

In this work, we perform a critical comparative evaluation of DDPM-based denoising under constrained clinical conditions. We adapt a lightweight 2D conditional DDPM to perform single-slice PET denoising using real clinical data from the UDPET2024 challenge [11] and benchmark it against a strong U-Net baseline. To further explore the potential of diffusion models, we implement a hybrid cascade architecture in which a pre-trained U-Net provides an initial denoised estimate that is refined by the DDPM. This setup is inspired by the cascaded architecture proposed in Xie et al. [12], but our work extends it by integrating an ensemble strategy: we generate multiple DDPM outputs per input and apply fusion techniques to enhance robustness and enable uncertainty-aware inference. While the DDPM-based approaches did not surpass the U-Net in this setting, our analysis offers new insights into the current limitations of DDPMs and practical considerations for their use in data-limited PET reconstruction scenarios.

II. Background and related work

Technical Limitations of Existing Methods

PET image denoising is particularly challenging due to the heterogeneous nature of noise characteristics across differ-

ent patients and scanning conditions, including variations in tracer injection timing, occasional infiltration at injection sites, differences in patient body mass index, weight-based dosing protocols, and scanner sensitivity variations across imaging centers [12]. Traditional denoising methods, including non-local means filtering and conventional neural networks, have shown limited success in addressing the complex noise characteristics of low-dose, short-acquisition PET images. These approaches often struggle with the non-stationary nature of PET noise and frequently introduce artifacts or over-smooth clinically relevant features. While recent advances in deep learning, particularly U-Net architectures, have shown promise for medical image denoising, they are inherently deterministic and produce a single output per input. As a result, they do not offer any estimate of reconstruction uncertainty—a limitation that may be critical in clinical applications, where understanding the confidence in the denoised output could enhance diagnostic decision-making.

Denoising Diffusion Probabilistic Models

Denoising Diffusion Probabilistic Models (DDPMs) are a class of generative models that approach denoising through a probabilistic framework involving iterative refinement steps [5, 8]. Unlike deterministic architectures such as U-Net, DDPMs learn the underlying data distribution and generate outputs via a stochastic sampling process. This enables them to model complex, non-stationary noise distributions and, in theory, preserve fine anatomical details that may be over-smoothed by conventional networks.

DDPMs are based on a two-stage diffusion process: a forward (noising) process that progressively corrupts an image with Gaussian noise, and a reverse (denoising) process that aims to recover the original image through learned steps. The forward process gradually adds Gaussian noise to a clean image x_0 over T steps as follows:

$$q(x_t|x_{t-1}) = \mathcal{N}(x_t \mid \sqrt{1-\beta_t}\, x_{t-1}, \beta_t I),$$

$$q(x_{1:T}|x_0) = \prod_{t=1}^{T} q(x_t|x_{t-1}),$$

where β_t is a variance schedule linearly increasing from $\beta_1 = 10^{-4}$ to $\beta_T = 2.0 \times 10^{-2}$ and $T = 1000$.

The reverse process is modeled as a learned conditional distribution:

$$p_\theta(x_{t-1}|x_t, c) = \mathcal{N}(x_{t-1} \mid \mu_\theta(x_t, c, t), \sigma_t^2 I),$$

where $c = x_{\text{low}}$ is the conditioning low-dose PET slice, and μ_θ is predicted by a neural network (commonly called an ϵ-predictor). The network $f_\theta([x_t, c], t)$ is trained to estimate the noise ϵ that was added at each step. The training objective minimizes the mean squared error between the predicted and true noise:

$$\mathcal{L}_{\text{simple}} = \mathbb{E}_{x_0,t,\epsilon}\left[\|\epsilon - \epsilon_\theta(x_t, c, t)\|^2\right].$$

Thus, the model learns to reverse the forward diffusion process by progressively denoising noisy latent variables conditioned on the input.

However, from a clinical perspective, the stochastic nature of DDPMs introduces challenges: rather than multiple plausible outputs, what is required is a single reliable reconstruction accompanied by information about uncertainty. To make this practically useful, we generate multiple denoised outputs and aggregate them using ensemble strategies. This allows us to produce a single consolidated reconstruction along with a pixel-wise uncertainty estimate that can highlight areas where the model's predictions are less stable.

Recent studies have applied DDPMs to PET denoising and reported promising results, typically using 3D architectures trained on large clinical datasets [12, 2, 14, 13]. These approaches result in excellent denoising capabilities, but are extremely computationally demanding. In contrast, our work investigates whether useful denoising results can be achieved using a more lightweight 2D DDPM trained on substantially less clinical data. We also explore a hybrid approach that combines a deterministic U-Net with a DDPM-based refinement step, and propose ensemble-based uncertainty estimation. While our setup differs from existing work in terms of dimensionality, training scale, and architecture, it offers insight into how DDPMs behave under more constrained, clinically realistic conditions.

III. Proposed Methods

This work develops and assesses three learning-based approaches for low-dose PET image denoising: a baseline U-Net, a conditional Denoising Diffusion Probabilistic Model (DDPM), and a hybrid cascade model. The objective is to restore high-quality PET images from count-reduced acquisitions while preserving diagnostic value.

Dataset

The research utilized the publicly available Ultra-Low Dose PET Imaging (UDPET) Challenge 2024 dataset, comprising full-body PET scans from 1,447 subjects (Siemens Biograph Vision Quadra: $n = 387$; United Imaging uEXPLORER: $n = 1,060$). Images are provided at simulated dose levels from 1% to 100% in NIFTI format (in SUV units). Low-count images were generated by sinogram domain downsampling, ensuring realistic Poisson statistics. A subset of 50 FDG-PET volumes from the Siemens Vision Quadra scanner was used, focusing on the 25% (low-dose) vs. 100% (full-dose) count levels. For the 2D models, 3D volumes were sliced into 2D numpy arrays.

Baseline U-Net

A standard 2D U-Net served as a baseline for supervised denoising. It was trained on $\approx 53k$ 644×644 training slices to map a low-count 25% axial PET slice, x_{low}, to an estimated full-count slice, $\hat{x}_{high}$, by minimizing the mean absolute error ($\mathcal{L}_{L1}$):

$$\mathcal{L}_{L1} = \frac{1}{N}\sum_{i=1}^{N} ||f_\theta(x_{low}^{(i)}) - x_{high}^{(i)}||_1$$

The L1 loss was chosen for its ability to preserve sharp edges and SUV contrast. The U-Net has a four-level encoder-decoder topology ($\sim 3M$ parameters) with residual connections.

Conditional Denoising Diffusion Probabilistic Model (DDPM)

To evaluate the potential of diffusion-based denoising in PET imaging, we implemented a conditional DDPM (code adapted from [9]) tailored for slice-wise processing. The model follows the standard forward and reverse diffusion formulation, where Gaussian noise is progressively added and then removed through a learned denoising network (see Figure 1). Conditioning is applied at each reverse step by concatenating the current noisy latent variable x_t with the corresponding low-dose PET slice x_{low} along the channel dimension, resulting in a combined input $[x_t, x_{\text{low}}]$.

The denoising network is an ϵ-predictor U-Net tasked with estimating the noise added at each diffusion step. Our implementation uses a deeper U-Net architecture with approximately 10.7 million parameters. It employs Swish activations and incorporates self-attention layers in the bottleneck to better capture long-range dependencies and global anatomical context.

The model was trained using $\approx 52k$ 128×128 information dense PET patches extracted from full-body scans. Training minimizes the mean squared error between the predicted and true noise components at each timestep. A schematic overview of the conditional diffusion process is shown in Figure 1.

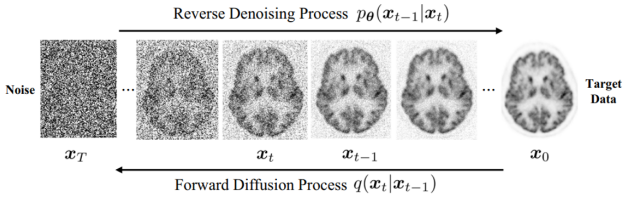

Fig. 1. Schematic representation of the diffusion process.

Hybrid U-Net-DDPM Cascade

We implemented a hybrid denoising approach that combines a deterministic U-Net with a conditional DDPM to exploit the strengths of both models. This cascade architecture was inspired by Xie et al. [12], who proposed using a U-Net to generate an initial prior followed by refinement steps using a conditional DDPM. We adopt this setup as a starting point but extend it with an ensemble approach to exploit the stochastic nature of the DDPM for uncertainty interpretability. Concretely, the DDPM refinement is done as follows.

First, we re-noise the U-Net output x_{UNet} to an intermediate timestep $t_{\text{start}} = 100$ using the forward diffusion equation:

$$x_{t_{\text{start}}} = \sqrt{\overline{\alpha}_{t_{\text{start}}}}\, x_{\text{UNet}} + \sqrt{1 - \overline{\alpha}_{t_{\text{start}}}}\, \epsilon, \quad \epsilon \sim \mathcal{N}(0, I)$$

Then, we apply 100 reverse diffusion steps to refine this prior using the DDPM, conditioned on the original low-dose input x_{low}:

$$x_0 = \text{DDPM}_\theta([x_{t_{\text{start}}}, x_{\text{low}}],\, t_{\text{start}} \to 0)$$

To aggregate the $K = 16$ DDPM refinements per input patch, we evaluated five ensemble fusion methods. The simple mean and simple median directly average or select the central tendency of the outputs. Beyond these, we implemented a trimmed mean (robust mean), which discards extreme samples outside the interquartile range before averaging. The consistency-weighted mean assigns higher weights to samples that are closer to the ensemble median, emphasizing agreement across samples. Finally, the guided ensemble combines the consistency-weighted output with the simple mean, using the original low-count input as a soft guidance signal, with a mixing factor $\alpha = 0.3$. These methods not only aim to stabilize the final prediction but also allow estimation of voxel-wise uncertainty by analyzing sample variability.

A schematic overview of this hybrid refinement approach is shown in Figure 2.

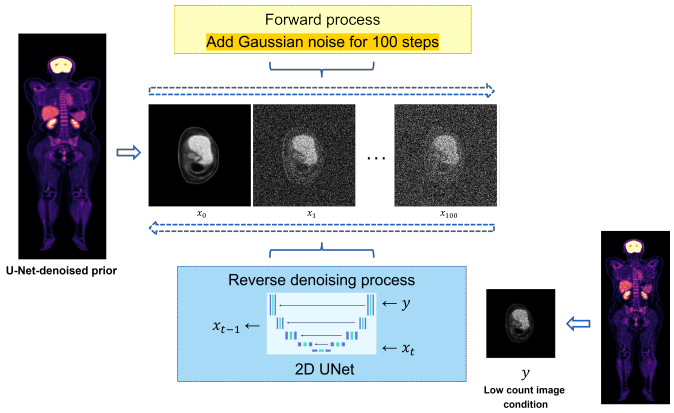

Fig. 2. Schematic representation of the hybrid model that combines both the U-Net model for generating priors and the DDPM for further refinement.

Performance Metrics and Experimental Framework

Model performance was assessed using Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM). Experiments were run on NVIDIA A10G GPUs (24GB VRAM) using a PyTorch framework and MONAI functions. Processing was limited to 2D slices.

IV. Experimental Results and Discussion

We evaluated the performance of the baseline U-Net, the conditional DDPM, and the hybrid ensemble model for low-dose PET denoising. Performance was assessed both quantitatively using PSNR and SSIM, and qualitatively through visual inspection.

Conditional DDPM

To ensure a consistent and focused evaluation, both the conditional DDPM and the U-Net baseline were tested on the same 1,500 information-dense 128×128 patches, extracted from held-out clinical PET slices selected for their high diagnostic relevance. As summarized in Table I, the U-Net clearly outperformed the DDPM in both PSNR and SSIM. The DDPM not only failed to improve upon the low-dose

input but also degraded reconstruction quality in the majority of cases. In contrast, the U-Net consistently enhanced image quality relative to the input across most patches.

Qualitative results (Figure 3) confirm the quantitative findings. The DDPM reduced noise in many patches, but it exhibited critical failure modes, such as the appearance of hallucinated structures—particularly in regions with high tracer uptake. These artifacts raise concerns regarding clinical applicability. Furthermore, the model exhibited high inter-sample variability when generating multiple reconstructions from identical inputs, indicating instability likely due to undertraining, limited dataset size, and the inherent challenges of learning a 3D imaging technique with a 2D architecture.

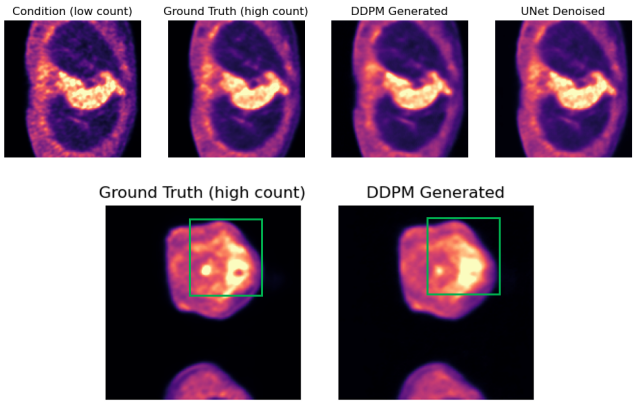

Fig. 3. Conditional DDPM results. Top four figures: DDPM denoised output together with the low-count input, the high-count ground truth and the U-Net denoised result. Bottom two figures: illustration where the DDPM struggles to preserve high-activity structures.

TABLE I
PATCH-LEVEL DENOISING PERFORMANCE (MEAN ± SD)

Model	**PSNR (dB)**	**SSIM**
Low-count input	34.10 ± 5.29	0.9198 ± 0.0982
U-Net	36.29 ± 4.86	0.9449 ± 0.0870
DDPM	30.78 ± 7.73	0.7732 ± 0.1880

These failure modes were not reported in earlier DDPM-based PET denoising work [12, 2, 14], likely because those studies used 3D architectures and larger training datasets. Our DDPM was trained on limited clinical data and implemented in 2D, potentially limiting its ability to capture spatial context and converge fully. The observed inter-sample variability from the DDPM further suggests training instability under these constraints.

Hybrid Cascade and Ensemble Refinement

The hybrid cascade model was designed to combine the speed and stability of a U-Net with the generative capabilities of a DDPM. Each U-Net denoised patch was re-noised and refined using 100 DDPM reverse steps. To improve robustness and enable uncertainty estimation, we generated $n = 16$ DDPM samples per patch and applied five ensemble strategies: simple mean, simple median, trimmed (robust) mean, a guided ensemble, and a consistency-weighted mean.

We evaluated this hybrid approach on 1,000 information-dense test patches. Across all ensemble strategies, performance was similar and remained significantly below that of the U-Net baseline. While the U-Net achieved a PSNR of 36.27 ± 7.33 dB, the ensemble outputs yielded PSNR values ranging from 31.95 ± 6.39 dB to 32.07 ± 6.42 dB—representing a degradation of approximately 4.2 to 4.4 dB. Similar drops were observed in SSIM. These results indicate that the ensemble strategy, despite averaging multiple DDPM outputs, was not sufficient to recover the structural fidelity achieved by the baseline model.

Qualitative inspection (Figure 4) confirmed this degradation. All ensemble variants exhibited systematic over-smoothing, often blurring high-uptake regions and reducing detail. This can be attributed to the high variability between individual DDPM outputs, likely due to limited training data, a 2D architecture lacking spatial context, and thus underconverged DDPM training. Averaging across inconsistent outputs dampens fine structures, undermining the intended benefit of the hybrid design.

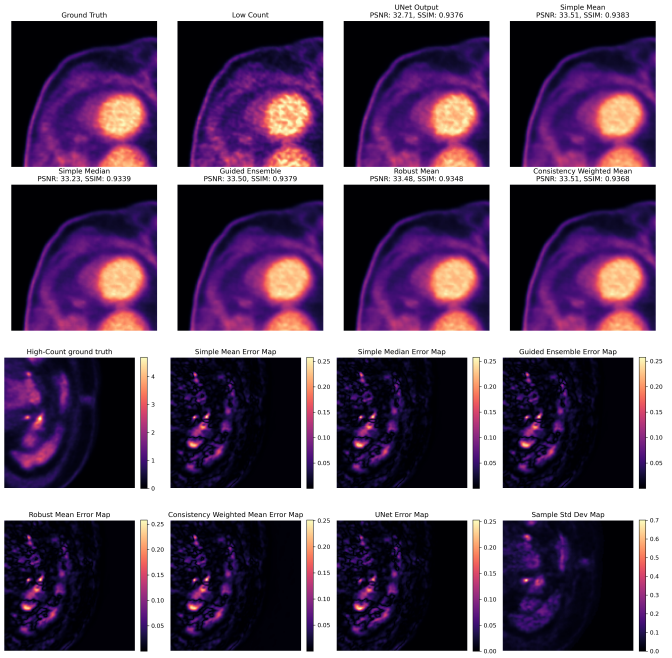

Fig. 4. Qualitative results of the hybrid-ensemble model. Top eight figures: results of the five ensemble strategies with the respective ground truth and U-Net denoised output. Bottom eight figures: variance maps of the denoising method outputs and the respective ground truth in SUV values.

Comparative Discussion

Table II summarizes the quantitative performance of all evaluated models. In our experiments on a subset of the UD-PET2024 dataset, the U-Net baseline consistently improved PSNR and SSIM across the majority of test cases, with

TABLE II
COMPREHENSIVE PERFORMANCE COMPARISON ACROSS ALL EVALUATED METHODS.

Method	PSNR Change [dB]	SSIM Change
U-Net	$+1.52 \pm 2.01$	-0.045 ± 0.148
DDPM	-3.32 ± 9.37	-0.147 ± 0.212
Hybrid Ensemble	-4.25 to -4.37	-0.135 to -0.147

no hallucination artifacts observed. In contrast, the conditional DDPM and the hybrid ensemble model not only failed to improve performance but degraded it—both in terms of PSNR/SSIM and visual fidelity. The DDPM, in particular, introduced structural artifacts and exhibited high sample variability, indicating instability during training.

These findings diverge from those reported in Gong et al. [2] and Yu et al. [14], where DDPMs achieved state-of-the-art results for PET denoising. However, those studies relied on substantially larger and more specific datasets: Gong et al. trained on 120 + 140 PET volumes, which were used to evaluate methods, while Yu et al. employed a 3D DDPM trained on large-scale total-body PET data (317 full-body volumes for training) with 3D spatial context and longer training durations on 8 Nvidia A100 GPUs. Xie et al. [12] achieved superior results with their dose-aware 3D cascade approach which also incorporated 320 full-body PET volumes in multiple dose levels. In contrast, our implementation used a 2D DDPM trained from scratch without priors on a much smaller dataset, highlighting the challenges of using diffusion models under constrained clinical conditions.

Taken together, these comparisons suggest that DDPMs are sensitive to training stability, dataset size, and architectural choices. These findings also highlight that, since PET is inherently a 3D imaging modality, denoising methods can benefit significantly from architectures that exploit the full 3D spatial context. Our results do not contradict prior work but rather contextualize its limitations: under data-scarce and computationally limited scenarios, deterministic models such as U-Net currently offer more robust and reliable denoising for clinical PET.

Finally, while PSNR and SSIM remain useful benchmarks, our findings also emphasize their limitations. The hallucination artifacts introduced by the DDPM were not always reflected in these metrics, reinforcing the need for qualitative assessment in evaluating medical image reconstruction.

V. CONCLUSION

In this work, we explored and critically evaluated conditional DDPM's for low-dose PET denoising in direct comparison to a robust and widely used deterministic baseline, the U-Net. The U-Net demonstrated strong performance, providing reliable improvements in image quality from low-dose PET data, and producing artifact-free reconstructions in clinically feasible inference times. Its simplicity, stability, and effectiveness highlight it as a practical choice for immediate clinical deployment.

In contrast, the conditional DDPM, despite its conceptual appeal due to uncertainty modeling capabilities and flexibility in capturing complex data distributions, did not perform well under our limited clinical data and 2D setting. The DDPM consistently showed inferior denoising quality compared to the baseline, substantial variability between samples, and clinically problematic artifacts in high activity regions. These performance issues likely stemmed from training instability, the limited size of the training available dataset, and the use of a simpler 2D architecture rather than a computationally more intensive, yet spatial aware, 3D approach.

We also evaluated a hybrid ensemble model, combining the speed and robustness of U-Net with DDPM-based ensemble refinement strategies. However, the hybrid approach did not yield any improvements over the standalone U-Net. Instead, averaging multiple DDPM outputs in an ensemble introduced systematic over-smoothing due to the high variance and instability of individual DDPM samples. This result underscores the critical importance of stable and well-converged generative models as a prerequisite for successful ensemble refinement.

Overall, while DDPMs offer attractive theoretical advantages and hold significant promise demonstrated in previous literature, our study highlights critical practical limitations under realistic clinical conditions with limited training data and computational resources. These results suggest that, under constrained conditions, simpler and more stable architectures such as the U-Net remain more reliable and suitable for near-term clinical application. Future research should focus on developing robust training strategies, employing larger and more diverse datasets, and rigorously validating generative diffusion models in clinically relevant scenarios to bridge the gap between theoretical promise and practical clinical utility.

VI. ACKNOWLEDGMENTS

This work was supported by the Special Research Fund of Ghent University under Grant BOF/IOP/2022/038, and by the Flanders AI Research Programme under Grant 174B09119.

BIBLIOGRAPHY

[1] Esma A. Akin et al. *Optimizing Oncologic FDG-PET/CT Scans to Decrease Radiation Exposure*. URL: https://www.imagewisely.org/Imaging-Modalities/Nuclear-Medicine/Optimizing-Oncologic-FDG-PETCT-Scans.

[2] Kuan Gong et al. “PET image denoising based on denoising diffusion probabilistic model”. In: *European Journal of Nuclear Medicine and Molecular Imaging* 51.2 (2023), pp. 358–368. DOI: 10.1007/s00259-023-06417-8.

[3] Grand View Research. *PET Scanners Market Size, Share & Trends Report, 2030*. 2024. URL: https://www.grandviewresearch.com/industry-analysis/pet-scanners-market-report.

[4] Fumio Hashimoto et al. “Deep learning-based PET image denoising and reconstruction: a review”. In: *Radiological Physics and Technology* 17.1 (Feb. 2024), pp. 24–46. DOI: 10.1007/s12194-024-00780-3. URL: https://doi.org/10.1007/s12194-024-00780-3.

[5] Jonathan Ho, Ajay Jain, and Pieter Abbeel. “Denoising diffusion probabilistic models”. In: *arXiv (Cornell University)* (Jan. 2020). DOI: 10.48550/arxiv.2006.11239. URL: https://arxiv.org/abs/2006.11239.

[6] Makoto Hosono et al. “Cumulative radiation doses from recurrent PET–CT examinations”. In: *British Journal of Radiology* 94.1126 (July 2021), p. 20210388. ISSN: 0007-1285. DOI: 10.1259/bjr.20210388. eprint: https://academic.oup.com/bjr/article-pdf/94/1126/20210388/54580790/bjr.20210388.pdf. URL: https://doi.org/10.1259/bjr.20210388.

[7] Davin Korstjens. “PET scan volumes continue to grow”. In: *AuntMinnie* (2024). Accessed May 2025. URL: https://www.auntminnie.com/clinical-news/molecular-imaging/article/15665651/pet-scan-volumes-continue-to-grow.

[8] Alexander Quinn Nichol and Prafulla Dhariwal. “Improved denoising diffusion probabilistic models”. In: *arXiv (Cornell University)* (Jan. 2021). DOI: 10.48550/arxiv.2102.09672. URL: https://arxiv.org/abs/2102.09672.

[9] Junbo Peng et al. “CBCT-Based synthetic CT image generation using conditional denoising diffusion probabilistic model”. In: *Medical Physics* 51.3 (2024), pp. 1847–1859. DOI: https://doi.org/10.1002/mp.16704. eprint: https://aapm.onlinelibrary.wiley.com/doi/pdf/10.1002/mp.16704. URL: https://aapm.onlinelibrary.wiley.com/doi/abs/10.1002/mp.16704.

[10] Elizabeth Robbins. “Radiation risks from imaging studies in children with cancer”. In: *Pediatric Blood & Cancer* 51.4 (2008), pp. 453–457. DOI: https://doi.org/10.1002/pbc.21599. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/pbc.21599. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/pbc.21599.

[11] *Ultra-Low Dose PET Imaging Challenge 2024*. https://udpet-challenge.github.io/. Accessed: 2024-11.

[12] Huidong Xie et al. “DDPET-3D: Dose-Aware Diffusion Model for 3D Ultra Low-Dose PET Imaging”. In: *arXiv preprint arXiv:2311.04248* (2023). DOI: 10.48550/arXiv.2311.04248.

[13] Boxiao Yu et al. “PET Image Denoising Based on 3D Denoising Diffusion Probabilistic Model: Evaluations on Total-Body Datasets”. In: *Medical Image Computing and Computer Assisted Intervention – MICCAI 2024*. Ed. by Marius George Linguraru et al. Cham: Springer Nature Switzerland, 2024, pp. 541–550. ISBN: 978-3-031-72104-5.

[14] Boxiao Yu et al. “Robust whole-body PET image denoising using 3D diffusion models: evaluation across various scanners, tracers, and dose levels”. In: *European Journal of Nuclear Medicine and Molecular Imaging* (Feb. 2025). DOI: 10.1007/s00259-025-07122-4. URL: https://doi.org/10.1007/s00259-025-07122-4.