Structure-Adaptive Sparse Diffusion in Voxel Space for 3D Medical Image Enhancement

Hongxu Jiang[1], Fei Li[2], Boxiao Yu[3], Ying Zhang[4], Kaleb Smith[5], Kuang Gong[3], and Wei Shao[2⋆]

[1] Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL, USA
[2] Department of Medicine, University of Florida, Gainesville, FL, USA
[3] Department of Biomedical Engineering, University of Florida, Gainesville, FL, USA
[4] Research Computing, University of Florida, Gainesville, FL, USA
[5] NVIDIA, Santa Clara, CA, USA
{hongxu.jiang, boxiao.yu, yingz, weishao}@ufl.edu, fei.Li@medicine.ufl.edu, kgong@bme.ufl.edu, kasmith@nvidia.com

Abstract. Three-dimensional (3D) medical image enhancement, including denoising and super-resolution, is critical for clinical diagnosis in CT, PET, and MRI. Although diffusion models have shown remarkable success in 2D medical imaging, scaling them to high-resolution 3D volumes remains computationally prohibitive due to lengthy diffusion trajectories over high-dimensional volumetric data. We observe that in conditional enhancement, strong anatomical priors in the degraded input render dense noise schedules largely redundant. Leveraging this insight, we propose a sparse voxel-space diffusion framework that trains and samples on a compact set of uniformly subsampled timesteps. The network predicts clean data directly on the data manifold, supervised in velocity space for stable gradient scaling. A lightweight Structure-aware Trajectory Modulation (STM) module recalibrates time embeddings at each network block based on local anatomical content, enabling structure-adaptive denoising over the shared sparse schedule. Operating directly in voxel space, our framework preserves fine anatomical detail without lossy compression while achieving up to 10× training acceleration. Experiments on four datasets spanning CT, PET, and MRI demonstrate state-of-the-art performance on both denoising and super-resolution tasks. Our code is publicly available at: https://github.com/mirthAI/sparse-3d-diffusion.

Keywords: Sparse diffusion · Voxel space · Medical image enhancement.

1 Introduction

Three-dimensional (3D) medical imaging modalities, such as computed tomography (CT), magnetic resonance imaging (MRI), and positron emission tomography (PET), are central to modern clinical diagnosis and treatment planning [10].

⋆ Corresponding author.

In practice, however, image acquisition is often constrained by radiation dose, scan duration, and hardware limitations, resulting in low signal-to-noise ratio and limited spatial resolution. These degradations obscure subtle anatomical structures and compromise diagnostic confidence, motivating extensive research into 3D image enhancement tasks, including denoising [28,6] and super-resolution [17,30]. Early enhancement approaches operated in a slice-by-slice or 2.5D manner, often resulting in through-plane inconsistencies due to limited volumetric context. Fully 3D convolutional neural networks [2,23] and generative adversarial networks [25,29] were later introduced, but convolutional models often produce over-smooth details and adversarial methods remain difficult to optimize [25].

Diffusion probabilistic models [9,27] have recently emerged as a powerful alternative for image enhancement, offering more stable training, better distribution coverage, and improved perceptual fidelity [13,28]. More recently, flow matching-based methods [15,16] achieve high-fidelity enhancement with far fewer sampling steps by directly learning a continuous velocity field along smoother probability paths [21]. Despite these advances, scaling such models to high-resolution 3D data remains challenging, and existing approaches still reduce computational cost through compressed representations, including VAE-learned latents [32,13], wavelet decompositions [5,31], or downsampled inputs [1,4]. While efficient, such compressions inevitably discard high-frequency anatomical details (e.g., cortical folds and lung fissures), imposing representational bottlenecks that fundamentally limit enhancement quality. A natural alternative is returning to the original voxel space, thereby preserving full spatial resolution and avoiding information loss. However, existing models learn to predict noise or velocity targets that reside in complex, high-dimensional spaces far from the data manifold [14,20], requiring both prolonged training to capture such mappings and hundreds of iterative denoising steps to recover a clean volume, making direct 3D voxel-space generation prohibitively expensive.

To mitigate this overhead, we revisit the problem from a conditional modeling perspective. In enhancement tasks, the degraded input already encodes strong anatomical priors, fundamentally reducing the learning complexity compared to unconditional generation. The key is to choose a prediction target that preserves this advantage: predicting noise or velocity maps the network output away from the data manifold, discarding the structural proximity between the degraded input and the clean target. We instead adopt clean data x_0 prediction, which keeps the network output on the data manifold throughout training, allowing the model to directly refine the conditioning prior rather than learn a detour through high-variance noise or velocity spaces.

This closer alignment between prediction target and enhancement objective yields smoother denoising behaviors in which neighboring timesteps produce highly similar outputs [3,7,19], naturally motivating sparse diffusion on a subset of representative timesteps [18,24]. While [12] showed that a fixed sparse schedule suffices for 2D conditional generation, extending it to patch-based 3D diffusion reveals a further challenge: high-resolution patches across diverse anatomical re-

gions (e.g., homogeneous liver tissue versus intricate vascular structures) exhibit varied denoising dynamics that a single global schedule cannot accommodate.

To this end, we propose a structure-aware sparse diffusion framework that operates directly in 3D voxel space for medical image enhancement. The framework combines three complementary designs: x_0-prediction with velocity supervision for stable on-manifold learning, sparse diffusion that eliminates redundant denoising steps, and a Structure-aware Trajectory Modulation (STM) module that recalibrates the time embedding according to local anatomical complexity, enabling structure-adaptive denoising at each step. Together, these components achieve up to 10× faster training while delivering state-of-the-art performance across CT, PET, and MRI denoising and super-resolution benchmarks.

The main contributions are as follows: 1) We propose a sparse diffusion framework in 3D voxel space that combines x_0-prediction with velocity supervision and timestep subsampling, achieving up to 10× faster training while improving enhancement quality. 2) We introduce Structure-aware Trajectory Modulation (STM), which leverages tri-planar structural encoding and block-wise time embedding modulation to adapt denoising trajectories to local anatomical structure. 3) Extensive experiments on four datasets spanning CT, PET, and MRI deliver state-of-the-art performance on both denoising and super-resolution tasks.

2 Methodology

2.1 Overview

The proposed framework, illustrated in Fig. 1, performs structure-adaptive sparse diffusion directly in voxel space for 3D medical image enhancement. High-resolution volumes are first partitioned into overlapping 3D patches to enable memory-efficient processing while preserving local spatial context. During diffusion, the model takes the noisy patch x_t and degraded input x_{cond} as joint inputs and predicts the clean target x_0, supervised by a velocity loss for stable on-manifold learning. Training operates on a subset of uniformly subsampled timesteps, forming a sparse trajectory that eliminates redundant denoising steps. A Structure-aware Trajectory Modulation (STM) module further extracts structural cues from x_{cond} to recalibrate time embeddings at each UNet block, adapting the denoising behavior to local anatomical complexity.

2.2 Sparse Voxel-Space Diffusion

Let $x_0 \in \mathbb{R}^{D\times H\times W}$ denote a high-quality 3D medical volume and $y \in \mathbb{R}^{D\times H\times W}$ its degraded input. To enable tractable processing under GPU memory constraints, we extract overlapping 3D patches $\{x_0^{(i)}, y^{(i)}\}_{i=1}^{N}$ of size $d \times h \times w$ from the volume pair. The forward diffusion process corrupts x_0 via $x_t = \sqrt{\bar{\alpha}_t}\, x_0 + \sqrt{1-\bar{\alpha}_t}\, \epsilon$, where $\bar{\alpha}_t = \prod_{s=1}^{t}(1-\beta_s)$ and $\epsilon \sim \mathcal{N}(0, \mathbf{I})$.

***x_0*-Prediction.** In conditional enhancement, the degraded input y already provides strong anatomical priors, restricting the problem to local refinement rather

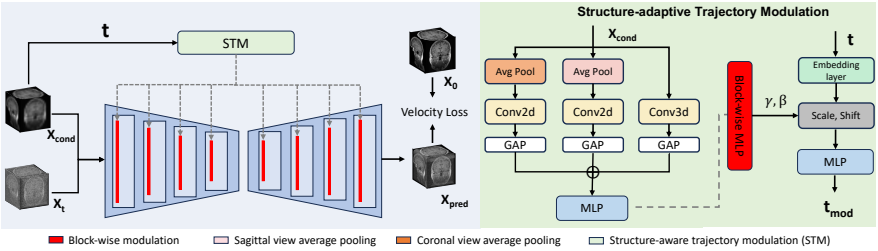

Fig. 1. Overview of the proposed framework.

than unconditioned generation. Rather than regressing noise ϵ or velocity v, which reside in high-variance spaces off the data manifold, we let the denoising network f_θ directly predict the clean data: $\hat{x}_0 = f_\theta(x_t, y, t)$.

Sparse Timestep Scheduling. The x_0-prediction formulation yields smoother denoising trajectories with high inter-step redundancy, as neighboring noise levels produce highly similar outputs. We exploit this by uniformly subsampling K timesteps from the full schedule of T steps (e.g., T=1000) to form a sparse schedule $\mathcal{S} = \{s_1, \ldots, s_K\}$ where $K \ll T$. (we use K=5; see ablation in table 2). Both forward and reverse processes then operate exclusively on $\mathcal{S}$, whereas conventional models require the full T steps.

Velocity Supervision. While x_0-prediction keeps the network output on the data manifold, directly supervising in x_0-space leads to unbalanced gradient magnitudes across timesteps: the loss scale vanishes near t=0 and explodes near t=T [24,9]. We therefore reformulate the supervision in velocity space. Substituting $\epsilon = (x_t - \sqrt{\bar{\alpha}_t}\, x_0)/\sqrt{1-\bar{\alpha}_t}$ into the velocity definition $v_t = \sqrt{\bar{\alpha}_t}\, \epsilon - \sqrt{1-\bar{\alpha}_t}\, x_0$ yields $v_t = (\sqrt{\bar{\alpha}_t}\, x_t - x_0)/\sqrt{1-\bar{\alpha}_t}$. We apply the same conversion to the network prediction $\hat{x}_0$ to obtain:

$$\hat{v}_t = \frac{\sqrt{\bar{\alpha}_t}\, x_t - \hat{x}_0}{\sqrt{1-\bar{\alpha}_t}}. \tag{1}$$

This retains the benefit of on-manifold prediction while producing uniformly scaled gradients across all timesteps in $\mathcal{S}$, stabilizing training under our sparse schedule. The training objective is given by:

$$\mathcal{L}_{\text{base}} = \mathbb{E}_{x_0,\, t\sim\mathcal{U}(\mathcal{S}),\, \epsilon\sim\mathcal{N}(0,\mathbf{I})} \left[\|v_t - \hat{v}_t\|_2^2 \right]. \tag{2}$$

2.3 Structure-aware Trajectory Modulation

The sparse diffusion process described above applies a globally uniform denoising schedule across all patches regardless of their content. In practice, however, 3D medical volumes exhibit pronounced anatomical heterogeneity: patches containing fine structural detail (e.g., bronchial trees, cortical folds) require more gradual denoising, whereas homogeneous regions (e.g., soft tissue, background) can tolerate coarser adjustments. To address this, we propose Structure-aware

Trajectory Modulation (STM), which dynamically conditions the denoising process on local anatomical content, enabling structure-adaptive denoising behavior across the shared sparse schedule.

Multi-structure Encoder. To capture structural context from complementary spatial dimensions, we design a lightweight multi-structure encoder E_ϕ as shown in Fig. 1. A volumetric branch extracts holistic 3D structural features from the conditioning input x_{cond} through 3D convolutions followed by global average pooling (GAP). In parallel, two planar branches operate on mean projections of x_{cond} along the coronal and sagittal axes, each processed by 2D convolutions and GAP to capture orientation-specific 2D structural patterns that may be diluted in purely volumetric pooling. The three branch outputs are concatenated and fused through an MLP to produce a compact structural representation:

$$\mathbf{c} = \text{MLP}\Big(\Big[E_\phi^{\text{3D}}(x_{\text{cond}});\ E_\phi^{\text{cor}}(\tilde{x}_{\text{cond}}^{\text{cor}});\ E_\phi^{\text{sag}}(\tilde{x}_{\text{cond}}^{\text{sag}})\Big]\Big) \in \mathbb{R}^{d_c}, \tag{3}$$

where $\tilde{x}_{\text{cond}}^{\text{cor}}$ and $\tilde{x}_{\text{cond}}^{\text{sag}}$ denote mean projections along the anterior–posterior and lateral axes, respectively.

Block-wise Trajectory Modulation. The structural representation $\mathbf{c}$ modulates the denoising process by recalibrating the time embedding at every residual block of the UNet. Each block contains a dedicated MLP that maps $\mathbf{c}$ to block-specific affine parameters:

$$(\gamma_k, \beta_k) = \text{split}(\text{MLP}_k(\mathbf{c}))\,, \tag{4}$$

which then modulate the time embedding $\mathbf{e}_t$ to produce a structure-conditioned variant:

$$\mathbf{t}_{\text{mod}}^{(k)} = (1 + \gamma_k) \odot \mathbf{e}_t + \beta_k, \tag{5}$$

where $\mathbf{e}_t$ is obtained from timestep t via a sinusoidal embedding layer and k indexes the residual blocks of the UNet. The block-wise design allows each network depth to learn its own structure-time interaction, as shallow blocks may prioritize coarse anatomical layout while deeper blocks attend to fine-grained detail. Since the time embedding governs how aggressively the network denoises at each step, modulating it with structural information effectively reshapes the denoising trajectory: the same physical timestep induces different denoising strengths depending on local anatomical complexity, enabling structurally distinct patches to follow individualized denoising paths over the shared sparse schedule.

3 Experimental Results

Datasets. We evaluate the proposed approach on two 3D medical image enhancement tasks: image denoising and image super-resolution, using four publicly available datasets that span CT, PET, and MRI modalities. For CT image denoising, we use the lung LDCT-and-Projection-data dataset [22] with image sizes ranging from 512×512×300 to 512×512×400. We used 50 patients in total, with 40 for training and 10 for testing. For PET image denoising, we use brain

FDG PET dataset [8] with an image size of 256×256×89. We used 70 patients for training and 18 for evaluation. For CTA image super-resolution, the AortaSeg24 dataset [11] were used with image sizes ranging from 350×350×500 to 520×520×800. We used 50 patients for training and 10 for evaluation. For MRI super-resolution, we use UHB FCD Lesion Brain MRI Dataset [26] with image sizes of 320×320×208. We used 120 scans in total, with 96 for training and 24 for testing. All volumes were normalized to [-1, 1].

Model Evaluation. We evaluate enhancement quality using four 3D metrics: PSNR, SSIM, MS-SSIM, and HFEN. We compare against four representative 3D diffusion baselines: 3D LDM [13] (latent-space), 3D WDM [5] (wavelet-domain), 3D DDPM [28], and 3D DDIM [27], covering the dominant paradigms of 3D medical image diffusion. Models are trained with learning rate 1×10^{-4} and batch size 4 on two NVIDIA B200 GPUs. Full configurations are available in our repository.

Quantitative Results. Table 1 presents quantitative results across all datasets and tasks. Our method consistently achieves the best or second-best scores on every metric. On lung CT denoising, our model improves PSNR by +0.57 dB and SSIM by +0.016 over the strongest baseline, while on aorta CTA it surpasses 3D DDIM by +0.46 dB PSNR with uniformly better SSIM, MS-SSIM, and HFEN, indicating more faithful recovery of fine vascular structures. A consistent improvement is observed across all remaining datasets as well. We attribute these gains to the complementary strengths of the framework: operating in voxel space avoids the representational loss of latent compression, sparse scheduling eliminates redundant steps that cause repeated smoothing, and STM adapts the denoising behavior to local anatomical complexity, preserving fine details in structurally rich regions while allowing more aggressive denoising elsewhere.

Table 1. Quantitative comparison on denoising and 4× super-resolution tasks across four datasets. Best results are in **red**, second-best are blue. Across all datasets, our method's improvements over the second-best baseline are statistically significant under a paired Wilcoxon signed-rank test ($p < 0.01$).

	Lung CT Denoising				Brain PET Denoising			
Method	**PSNR↑**	**SSIM↑**	**MS-SSIM↑**	**HFEN↓**	**PSNR↑**	**SSIM↑**	**MS-SSIM↑**	**HFEN↓**
3D LDM	28.89	0.7380	0.9430	0.0308	35.43	0.9605	0.9872	0.0179
3D WDM	29.54	0.7578	0.9532	0.0283	36.27	0.9624	0.9884	0.0167
3D DDPM	29.25	0.7414	0.9426	0.0291	36.59	0.9621	0.9890	0.0159
3D DDIM	31.25	0.8073	0.9613	0.0230	36.85	0.9646	0.9896	0.0156
Ours	**31.82**	**0.8232**	**0.9644**	**0.0215**	**37.08**	**0.9660**	**0.9899**	**0.0149**
	Aorta CTA Super-Resolution				**Brain MRI Super-Resolution**			
Method	**PSNR↑**	**SSIM↑**	**MS-SSIM↑**	**HFEN↓**	**PSNR↑**	**SSIM↑**	**MS-SSIM↑**	**HFEN↓**
3D LDM	40.24	0.9765	0.9951	0.0111	33.96	0.9613	0.9908	0.0207
3D WDM	40.66	0.9772	0.9952	0.0107	34.45	0.9645	0.9912	0.0200
3D DDPM	40.72	0.9742	0.9947	0.0107	33.68	0.9592	0.9903	0.0217
3D DDIM	41.52	0.9798	0.9957	0.0096	35.62	0.9700	0.9930	0.0172
Ours	**41.98**	**0.9827**	**0.9965**	**0.0089**	**35.95**	**0.9724**	**0.9934**	**0.0163**

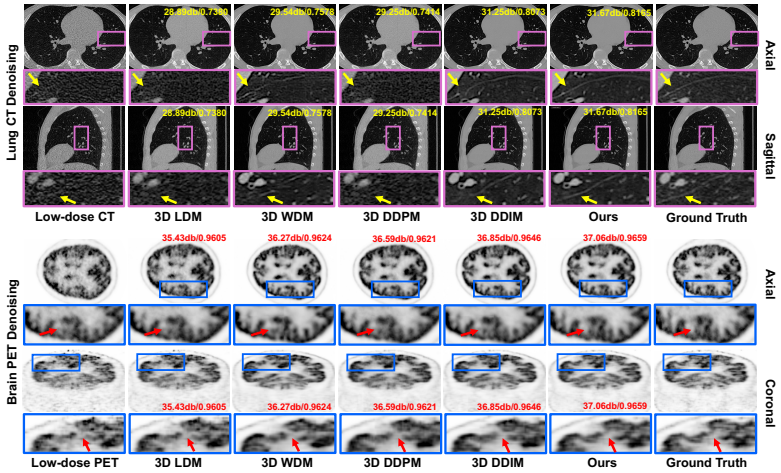

Fig. 2. Qualitative denoising results on lung CT and brain PET. Purple and blue boxes highlight lung fissures and subtle anatomical boundaries, respectively.

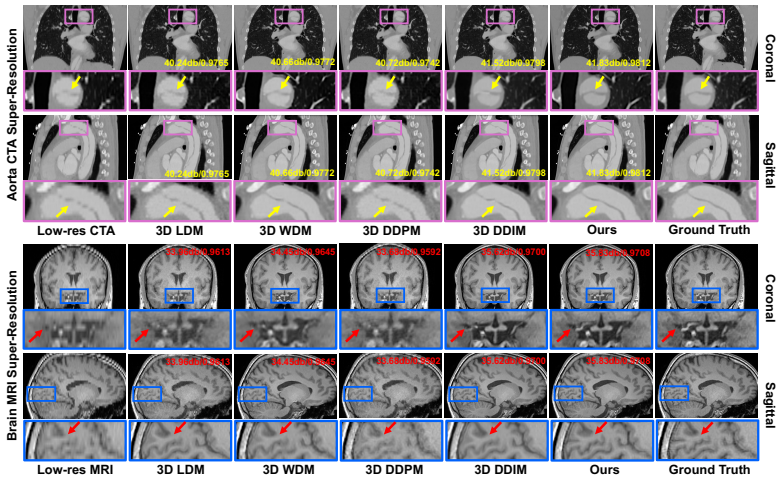

Fig. 3. Qualitative 4× super-resolution results on aorta CTA and brain MRI. Highlighted regions show vessel boundaries and cortical structures.

Qualitative Results. Fig. 2 presents denoising comparisons on lung CT and brain PET. In CT, our method reconstructs the sharpest lung fissure lines (purple boxes), closely matching the normal-dose reference, while baselines exhibit blurring or structural loss. In PET, our approach recovers clearer anatomical boundaries (blue boxes) where baselines suffer from oversmoothing. Fig. 3 shows 4× super-resolution results on aorta CTA and brain MRI. Our method produces more distinct vessel boundaries in CTA and preserves sharper gray–white matter

interfaces in MRI, whereas baselines fail to resolve fine cortical folds and aortic branches. Across both tasks, our proposed method consistently improves fine anatomical detail recovery.

Computational Efficiency. We evaluate training efficiency by comparing convergence curves against 3D DDIM on both tasks. As shown in Fig. 4, our method achieves up to 10× faster convergence on denoising and super-resolution, reaching higher PSNR at the same iteration count while the baseline remains far from its final performance. This confirms that sparse voxel-space diffusion with STM substantially reduces the computational cost of 3D diffusion without sacrificing enhancement quality.

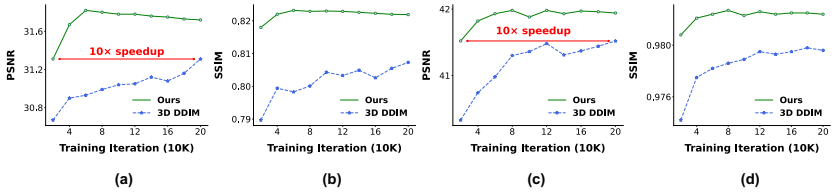

Fig. 4. Convergence curves for lung CT denoising and aorta CTA 4× super-resolution. Our method reaches the baseline's final performance with up to 10× fewer iterations.

Ablation Study. As shown in Table 2, the performance peaks at K=5, with further steps degrading quality due to repeated noise injection–denoising cycles that oversmooth fine structures, confirming that a compact sparse schedule is both sufficient and optimal for conditional enhancement. Table 3 isolates the contribution of proposed components. Switching from ϵ and v-prediction to x_0-prediction with velocity supervision dramatically improves all metrics. Adding sparse diffusion scheduling improves performance while reducing training time by $\sim$10×. Finally, incorporating STM yields further gains at no additional training cost, validating that structure-adaptive modulation and sparse diffusion are complementary.

Table 2. Ablation on diffusion step count K for lung CT denoising (PSNR / SSIM). Best in **red**, second in blue.

Method / Step#	3	5	10	50	100
3D DDIM	30.69 / .7897	31.01 / .8020	31.25 / .8073	30.56 / .7872	30.28 / .7839
Ours	31.05 / .7968	**31.82** / **.8232**	31.69 / .8194	31.41 / .8140	30.98 / .7988

Table 3. Component ablation on lung CT denoising. x_0+v: x_0-prediction with velocity supervision. Sparse: sparse timestep scheduling (K=5). Best in **red**, second in blue.

x_0+v	Sparse	STM	PSNR↑	SSIM↑	MS-SSIM↑	HFEN↓	Train Time↓
✗	✗	✗	19.42	.4639	.5087	0.0812	1×
✓	✗	✗	31.25	.8073	.9613	0.0230	1×
✓	✓	✗	31.56	.8137	.9624	0.0222	∼0.1×
✓	✓	✓	**31.82**	**.8232**	**.9644**	**0.0215**	∼0.1×

4 Conclusion

We presented a sparse voxel-space diffusion framework for 3D medical image enhancement. The key insight is that strong anatomical priors in the degraded input make dense noise schedules largely redundant. Exploiting this, our framework operates on as few as five timesteps while STM adapts denoising to local anatomical structure. Experiments across CT, PET, and MRI demonstrate up to 10× training acceleration with consistent improvements in reconstruction quality. A promising future direction is to combine our structure-adaptive sparse scheduling with flow matching for more efficient volumetric enhancement.

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

References

1. Bieder, F., Wolleb, J., Durrer, A., Sandkuehler, R., Cattin, P.C.: Memory-efficient 3d denoising diffusion models for medical image processing. In: Medical Imaging with Deep Learning. pp. 552–567. PMLR (2024) 2
2. Chaudhari, A.S., Fang, Z., Kogan, F., Wood, J., Stevens, K.J., Gibbons, E.K., Lee, J.H., Gold, G.E., Hargreaves, B.A.: Super-resolution musculoskeletal mri using deep learning. Magnetic resonance in medicine **80**(5), 2139–2154 (2018) 2
3. Choi, J., Lee, J., Shin, C., Kim, S., Kim, H., Yoon, S.: Perception prioritized training of diffusion models. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 11472–11481 (2022) 2
4. Dorjsembe, Z., Pao, H.K., Odonchimed, S., Xiao, F.: Conditional diffusion models for semantic 3d brain mri synthesis. IEEE Journal of Biomedical and Health Informatics **28**(7), 4084–4093 (2024) 2
5. Friedrich, P., Wolleb, J., Bieder, F., Durrer, A., Cattin, P.C.: Wdm: 3d wavelet diffusion models for high-resolution medical image synthesis. In: MICCAI workshop on deep generative models. pp. 11–21. Springer (2024) 2, 6
6. Gao, Q., Li, Z., Zhang, J., Zhang, Y., Shan, H.: Corediff: Contextual error-modulated generalized diffusion model for low-dose ct denoising and generalization. IEEE Transactions on Medical Imaging **43**(2), 745–759 (2023) 2
7. Go, H., Lee, Y., Lee, S., Oh, S., Moon, H., Choi, S.: Addressing negative transfer in diffusion models. Advances in Neural Information Processing Systems **36**, 27199–27222 (2023) 2

8. 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) 6
9. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. Advances in neural information processing systems **33**, 6840–6851 (2020) 2, 4
10. Hussain, S., Mubeen, I., Ullah, N., Shah, S.S.U.D., Khan, B.A., Zahoor, M., Ullah, R., Khan, F.A., Sultan, M.A.: Modern diagnostic imaging technique applications and risk factors in the medical field: a review. BioMed research international **2022**(1), 5164970 (2022) 1
11. Imran, M., Krebs, J.R., Sivaraman, V.B., Zhang, T., Kumar, A., Ueland, W.R., Fassler, M.J., Huang, J., Sun, X., Wang, L., et al.: Multi-class segmentation of aortic branches and zones in computed tomography angiography: The aortaseg24 challenge. arXiv preprint arXiv:2502.05330 (2025) 6
12. Jiang, H., Imran, M., Zhang, T., Zhou, Y., Liang, M., Gong, K., Shao, W.: Fast-ddpm: Fast denoising diffusion probabilistic models for medical image-to-image generation. IEEE Journal of Biomedical and Health Informatics (2025) 2
13. Khader, F., Müller-Franzes, G., Tayebi Arasteh, S., Han, T., Haarburger, C., Schulze-Hagen, M., Schad, P., Engelhardt, S., Baeßler, B., Foersch, S., et al.: Denoising diffusion probabilistic models for 3d medical image generation. Scientific Reports **13**(1), 7303 (2023) 2, 6
14. Li, T., He, K.: Back to basics: Let denoising generative models denoise. arXiv preprint arXiv:2511.13720 (2025) 2
15. Lipman, Y., Chen, R.T., Ben-Hamu, H., Nickel, M., Le, M.: Flow matching for generative modeling. In: 11th International Conference on Learning Representations, ICLR 2023 (2023) 2
16. Liu, X., Gong, C., Liu, Q.: Flow straight and fast: Learning to generate and transfer data with rectified flow. In: The Eleventh International Conference on Learning Representations (ICLR) (2023) 2
17. Liu, X., Xie, Y., Liu, C., Cheng, J., Diao, S., Tan, S., Liang, X.: Diffusion probabilistic priors for zero-shot low-dose ct image denoising. Medical Physics **52**(1), 329–345 (2025) 2
18. Luo, X., Xie, Y., Qu, Y., Fu, Y.: Skipdiff: Adaptive skip diffusion model for high-fidelity perceptual image super-resolution. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 38, pp. 4017–4025 (2024) 2
19. Ma, Q., Ning, X., Liu, D., Niu, L., Zhang, L.: Decouple-then-merge: Finetune diffusion models as multi-task learning. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 23281–23291 (2025) 2
20. Ma, Z., Xu, R., Zhang, S.: Pixelgen: Pixel diffusion beats latent diffusion with perceptual loss. arXiv preprint arXiv:2602.02493 (2026) 2
21. Martin, S., Gagneux, A., Hagemann, P., Steidl, G.: Pnp-flow: Plug-and-play image restoration with flow matching. In: International Conference on Learning Representations. vol. 2025, pp. 45466–45492 (2025) 2
22. Moen, T.R., Chen, B., Holmes III, D.R., Duan, X., Yu, Z., Yu, L., Leng, S., Fletcher, J.G., McCollough, C.H.: Low-dose ct image and projection dataset. Medical physics **48**(2), 902–911 (2021) 5
23. Pham, C.H., Tor-Díez, C., Meunier, H., Bednarek, N., Fablet, R., Passat, N., Rousseau, F.: Multiscale brain mri super-resolution using deep 3d convolutional networks. Computerized Medical Imaging and Graphics **77**, 101647 (2019) 2
24. Salimans, T., Ho, J.: Progressive distillation for fast sampling of diffusion models. In: International Conference on Learning Representations (2022) 2, 4

25. Sánchez, I., Vilaplana Besler, V.: Brain mri super-resolution using generative adversarial networks. In: International conference on Medical Imaging with Deep Learning: Amsterdam, 4-6th July 2018. pp. 1–8 (2018) 2
26. Schuch, F., Walger, L., Schmitz, M., David, B., Bauer, T., Harms, A., Fischbach, L., Schulte, F., Schidlowski, M., Reiter, J., et al.: An open presurgery mri dataset of people with epilepsy and focal cortical dysplasia type ii. Scientific Data **10**(1), 475 (2023) 6
27. Song, J., Meng, C., Ermon, S.: Denoising diffusion implicit models. In: International Conference on Learning Representations (2021) 2, 6
28. 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: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 541–550. Springer (2024) 2, 6
29. Zhang, K., Hu, H., Philbrick, K., Conte, G.M., Sobek, J.D., Rouzrokh, P., Erickson, B.J.: Soup-gan: Super-resolution mri using generative adversarial networks. Tomography **8**(2), 905–919 (2022) 2
30. Zhang, T., Jiang, H., Gong, K., Shao, W.: Geodesic diffusion models for medical image-to-image generation. arXiv preprint arXiv:2503.00745 (2025) 2
31. Zheng, J., He, M., Tang, X., Wang, X., Cao, T., Zeng, T., Zhang, L., You, C.: 3d wavelet latent diffusion model for whole-body mr-to-ct modality translation. arXiv preprint arXiv:2507.11557 (2025) 2
32. Zhu, L., Xue, Z., Jin, Z., Liu, X., He, J., Liu, Z., Yu, L.: Make-a-volume: Leveraging latent diffusion models for cross-modality 3d brain mri synthesis. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 592–601. Springer (2023) 2