Blaze3DM: Integrating Triplane Representation with Diffusion for Solving 3D Inverse Problems in Medical Imaging

Jia He[1,2], Bonan Li[3], Ge Yang[1,2,‡], Ziwen Liu[1,‡]

[1] School of Artificial Intelligence, University of Chinese Academy of Sciences
[2] State Key Laboratory of Multimodal Artificial Intelligence Systems, Institute of Automation, Chinese Academy of Sciences
[3] School of Electronic, Electrical and Communication Engineering, University of Chinese Academy of Sciences
hejia2020@ia.ac.cn, libonan@ucas.ac.cn, ge.yang@ia.ac.cn, liuziwen@ucas.ac.cn

Abstract. Solving inverse problems, such as image restoration and reconstruction, is essential in medical imaging. Recently, research on deep learning-based methods for solving 3D data problems has become a focus in the field. Existing diffusion models achieve excellent reconstruction quality but face challenges with volume inconsistency and high computational costs when dealing with 3D medical images. To overcome these challenges, we propose Blaze3DM, a novel approach that combines triplane neural fields with a diffusion model for effective 3D medical image reconstruction. Blaze3DM leverages compact, data-dependent triplane embeddings to ensure volume consistency and significantly improve the computational efficiency of the diffusion model. Furthermore, we introduce a guidance-based sampling method for zero-shot 3D inverse problem solving, enabling Blaze3DM to generate high-fidelity 3D volumes from limited, low-quality 2D slices. We evaluate Blaze3DM on various 3D inverse problem tasks across multiple imaging modalities, including sparse-view CT, limited-angle CT, compressed-sensing MRI, and MRI isotropic super-resolution. The experimental results demonstrate that Blaze3DM not only achieves state-of-the-art reconstruction performance but also markedly improves computational efficiency by approximately 22 to 40 times. Code is available at: https://github.com/Jenn-He/Blaze3DM.

Keywords: Medical Imaging · Diffusion Model · Inverse Problem.

1 Introduction

The medical image inverse problem involves reconstructing complete image information from limited observed measurements, which is challenging due to its ill-posed nature. Representative applications of this include sparse-view computed tomography (SV-CT) [8, 29], limited-angle computed tomography (LA-CT) [8, 29], and compressed-sensing magnetic resonance imaging (CS-MRI) [9, 10]. In 3D

scenarios, the inverse problems expand to include issues like resolution degradation in the slice dimension, leading to tasks such as MRI isotropic reconstruction (ZSR-MRI) [8]. The high dimensionality and complex manifold structure of 3D medical images make it difficult to achieve high-quality image reconstruction with high computational efficiency. Therefore, it is critical to develop a compact and efficient representation of 3D medical images. Such representation can be used not only for solving 3D inverse problems but also for many other 3D medical image analysis applications.

Nowadays, diffusion models (DMs) [13, 17, 28, 23] have demonstrated excellent performance in image generation by gradually generating samples from noise. Recent studies[1, 5, 18, 25, 32] have explored using diffusion models to address reconstruction and super-resolution challenges. These works leverage diffusion processes to progressively refine images, ultimately yielding high-quality reconstructions. [6] further combined diffusion sampling with the Krylov subspace to enhance the efficiency of solving large-scale inverse problems. However, in 3D scenarios, these methods only infer slice by slice and stack individual slices processed by 2D backbone models.

Recently, DiffusionMBIR [7] and TPDM [16], have both utilized diffusion models as the generation backbone in 3D scenarios. DiffusionMBIR[7] employs the model-based iterative reconstruction (MBIR) method [14] and diffusion model to solve zero-shot medical inverse problem. It models stacked slices with additional Total Variation (TV) inter-slice constraints along the Z-axis, but still struggles to model long-range dependencies across the 3D distribution. Consequently, TPDM [16] makes further improvements by modeling the 3D data distribution as a product of two perpendicular 2D plane distributions and performs posterior-based sampling [5] alternatively in perpendicular directions to solve the 3D inverse problem. While these methods can produce high-quality results, they pose substantial challenges in computational costs.

To address these limitations, we leverage neural field[35, 21, 20, 33], especially triplane neural fields[2, 27, 30, 34] to efficiently and expressively represent 3D medical volumes. By constructing triplane embeddings and using a powerful diffusion model to learn the complex distribution of these triplane representations, we achieve expressive and effective 3D volume generation with rich geometric details, ensuring volumetric consistency and information compactness. In summary, our study has the following contributions:

- Blaze3DM is the first to use triplane neural fields for 3D medical image modeling. This representation improves volumetric consistency and boosts diffusion model efficiency.
- We present a guidance-based sampling method for the triplane diffusion model. By incorporating conditions related to degraded images and transformations, our model performs zero-shot 3D inverse problem solving and generates high-quality volumes.
- Experiments on four 3D medical tasks (SV-CT, LA-CT, CS-MRI, ZSR-MRI) show Blaze3DM achieves high performance and efficiency, with a speedup of 22 to 40 times over existing methods.

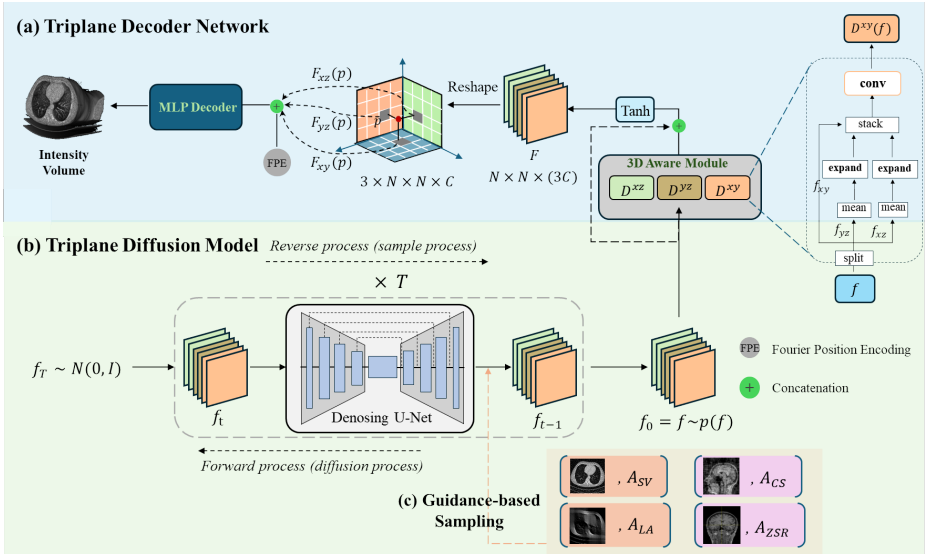

Fig. 1. The diagram of Blaze3DM. (a) The Triplane Decoder Network decodes the triplane embedding f into the volume intensity I using a 3D-aware module and a lightweight MLP decoder. (b) The Triplane Diffusion Model, based on a diffusion UNet, learns the triplane distribution for unconditional or conditional generation. (c) Guidance-based sampling uses random slices of the measurement volume and the degradation transform for zero-shot 3D volume restoration.

2 Methodology

2.1 Overview

Blaze3DM comprises two core components: the Triplane Decoder Network and the Triplane Diffusion Model. The training process has two stages. First, we construct triplane embeddings by jointly optimizing the data-dependent embeddings with the shared Triplane Decoder Network on high-quality medical volumes. Second, we train a powerful Triplane Diffusion Model to learn the distribution of these high-quality triplane embeddings. During inference, the Triplane Diffusion Model generates high-quality triplanes from limited degraded slices using a guidance-based sampling method, then the generated triplane is decoded into a 3D volume by the Triplane Decoder Network. The diagram is shown in Fig 1.

2.2 Triplane Decoder Network

Let $I \in \mathbb{R}^{H \times W \times D}$ denote the intensity of the medical volume V, where H, W, and D represent the dimensions of the volume along X, Y and Z axes, respectively. For each volume V, we define its corresponding triplane representation as $f \in \mathbb{R}^{N \times N \times 3C}$, where N and C denote the resolution and channel size. The position p of the neural field ranges from $[-1, 1]$.

3D Aware Module. The triplane representation consists of three axis-aligned, orthogonal 2D feature maps $f = \{f_{xy}, f_{yz}, f_{xz}\}$, corresponding to the XY, YZ, and XZ planes, respectively. To model interactions between these planes, we employ a 3D-aware module (3DAM) [34], which ensures each hyperplane absorbs information from the others, updating f to F. As shown in Fig.1 (a), 3DAM includes three 2D convolution networks, $\{D_\phi^{xy}, D_\phi^{yz}, D_\phi^{xz}\}$, each of which processes its corresponding hyperplane. For the XY plane, $F_{xy} = \tanh(D_\phi^{xy}(f) + f_{xy})$, with similar operations for the YZ and XZ planes, yielding $F = \{F_{xy}, F_{yz}, F_{xz}\}$. The final representation at position p is obtained by concatenating the feature maps: $F(p) = F_{xy}(p) \oplus F_{yz}(p) \oplus F_{xz}(p)$.

MLP Decoder. After obtaining representation $F(p)$ at position p, we apply a lightweight MLP network D_ϕ^{mlp} to predict volume intensity $\bar{I}(p)$. To overcome spectral bias [26] and extract finer details, we integrate Fourier Position Encoding (FPE) [24] to enhance the MLP's capacity and replace ReLU activation functions with Sinusoidal functions [22] to better capture complex and periodic patterns:

$$\bar{I}(p) = D_\phi^{mlp}([F(p), \mathrm{FPE}(p)]). \tag{1}$$

Triplane Fitting. We take high-quality 3D medical volumes as training dataset $\{V_1, V_2, ..., V_M\}$. The triplane fitting process involves jointly optimizing both the data-dependent triplane embeddings $\{f_1, f_2, ..., f_M\}$ and the shared decoder D_ϕ to reconstruct the original volume intensity. In each training step, we randomly select one volume V_i and then randomly sample a batch of spatial points to form training pairs $\{p, I_i(p)\}$. The reconstruction loss function is the mean square error (MSE) loss between predicted intensity $\bar{I}_i(p)$ and ground truth $I_i(p)$.

$$\mathcal{L}_{rec} = \mathbb{E}_{p \in V_i} \| I_i(p) - \bar{I}_i(p) \|^2. \tag{2}$$

Loss Regularization. To ensure the triplane representation can be well learned by diffusion models or decoded into images, we added regularization terms to constrain the characteristics of triplane embeddings.

First, we use total variation (TV)[3] regularization to enforce smooth textures in the triplane embeddings, reducing the difficulty of fitting the distribution for the generation backbone:

$$\mathcal{L}_{TV} = \mathbb{E}_{p \in V_i} \left(TV(f_{i,xy}) + TV(f_{i,yz}) + TV(f_{i,xz}) \right). \tag{3}$$

Second, we apply L_2 regularization to prevent extreme values in the triplane embeddings, reducing the impact of outliers on the diffusion model:

$$\mathcal{L}_{L_2} = \mathbb{E}_{p \in V_i} \left(\|f_{i,xy}(p)\|^2 + \|f_{i,yz}(p)\|^2 + \|f_{i,xz}(p)\|^2 \right). \tag{4}$$

Third, we introduce Explicit Intensity Regularization (EIR) to prevent overfitting to a fixed volume size and to enable arbitrary-size volume decoding, inspired by NFD [27]:

$$\mathcal{L}_{EIR} = \mathbb{E}_{p \in V_i} \left(\|\bar{I}_i(p) - \bar{I}_i(p + w)\|^2 \right), \quad w \in U(p), \tag{5}$$

Algorithm 1 Guidance-based Sampling

Require: Pretrained decoder D_{ϕ^*}, degradation transformation $\mathcal{A}$, measurement y, gradient scale λ, guidance slice count γ, fixed variance $\Sigma_t = \beta_t I$.

$f_T \sim \mathcal{N}(0, \sigma_1^2 I)$
for $t = T$ **to** 1 **do**
  $f_{t-1} \leftarrow$ sample from $\mathcal{N}(\mu_\theta(f_t), \Sigma_t)$
  $V_\gamma \leftarrow$ random select γ slices from degraded volume
  $f_{t-1} \leftarrow f_{t-1} + \lambda \Sigma_t \nabla_{f_{t-1}} \mathbb{E}_{p \in V_\gamma} \left\| \mathcal{A}(D_{\phi^*}(\hat{f}_0(f_{t-1}(p)))) - y(p) \right\|^2$
end for
$x = D_{\phi^*}(f_0)$
return x

where $U(p)$ denotes the spatial neighborhood of position p. The overall training objective can be described as:

$$f_i^*, \phi^* = \underset{f_i, \phi}{argmin} \left\{ \mathcal{L}_{rec} + \lambda_1 \mathcal{L}_{TV} + \lambda_2 \mathcal{L}_{L_2} + \lambda_3 \mathcal{L}_{EIR} \right\}, \tag{6}$$

where $\lambda_1, \lambda_2, \lambda_3$ are the regularization weights.

2.3 Triplane Diffusion Model

We use diffusion model as the generative backbone network to learn the distribution of triplane embeddings of high-quality volumes, and perform unconditional/conditional generation for medical volume reconstruction.

Diffusion Model Training. We adopt the diffusion procedure of DDPM [13] with a fixed variance schedule $\{\beta_t\}_{t=1}^T$ and follow its training objective to predict the noise ϵ in certain timesteps. However, the triplane representation f consists of three 2D planes f_{xy}, f_{yz}, and f_{xz}, each corresponding to different spatial orientations and exhibiting distinct distribution patterns. To avoid gradient explosion issues [34] when jointly optimizing these three distinct components, we decompose the training loss into three separate terms:

$$\mathcal{L}_{df} = \mathbb{E}_{t,\epsilon} \left[\|\epsilon_{xy} - \epsilon_\theta(f_t, t)_{xy}\|^2 + \|\epsilon_{yz} - \epsilon_\theta(f_t, t)_{yz}\|^2 + \|\epsilon_{xz} - \epsilon_\theta(f_t, t)_{xz}\|^2 \right] \tag{7}$$

Guidance-based Sampling. Considering a forward model for an imaging system $y = \mathcal{A}(x)$, the image inverse problem means estimating the unknown image x given limited measurement y, where $\mathcal{A}$ denotes the degradation transformation. We follow [11] to formalize the inverse problem as a posterior estimation problem. Based on DPS [5], we propose a guidance-based sampling method that additionally corrects predicted f_{t-1} towards the manifold $y = \mathcal{A}(\hat{x_0}) = \mathcal{A}(D_{\phi^*}(\hat{f}_0(f_{t-1})))$, which can be achieved through gradient descent:

$$f_{t-1} \leftarrow f_{t-1} + \lambda \Sigma_{t-1} \nabla_{f_{t-1}} \|\mathcal{A}(D_{\phi^*}(\hat{f}_0(f_{t-1})) - y\|^2, \tag{8}$$

where $\hat{f}_0$ is estimated using the Tweedie formula[12, 15], $\hat{f}_0(f_{t-1}) = E[f_0|f_{t-1}] = \frac{1}{\bar{\alpha}_t}(x_t - \sqrt{1-\bar{\alpha}_t}\epsilon_t)$. Our homoscedastic noise design incorporates the noise parameter into the guidance step size λ, thereby eliminating the need for covariance matrix inversion. We set $\Sigma_t = \beta_t I$. To reduce computational costs, we randomly select γ slices from the degraded volume (XY slices for SV-CT, LA-CT, CS-MRI, and XZ/YZ slices for ZSR-MRI task) to perform efficient guidance. The pseudo-algorithm for our guidance-based sampling is depicted in Algorithm.1.

3 Experiments

We conduct experiments on four classic 3D medical inverse problem tasks: SV-CT, LA-CT, CS-MRI, and ZSR-MRI. Our forward model utilizes TPDM's slice-independent degradation assumption. In the formula $y = Ax + n$, A indicates partial sampling in the sinogram for SV-CT and LA-CT tasks, Poisson disk sampling in k-space for CS-MRI tasks, and resolution down-sampling in the slice dimension for ZSR-MRI. The n denotes additive measurement noise. To compare fairly with DiffusionMBIR [7] and TPDM[16], we specifically perform uniform 36-view downsampling for SV-CT, 0-90°downsampling for LA-CT, retrospective $\times 8$ acceleration Poisson sub-sampling for CS-MRI, and $4\times$ anisotropy along Z-axis for ZSR-MRI task.

3.1 Dataset and Implementation Details

Dataset. For CT tasks, we take public AAPM-CT dataset [19] with 10 contrast-enhanced abdominal CT volumes. We follow [16] to use the same 9 volumes for training and 1 for testing. The volumes are sized at $256\times256\times256$ with about 2 mm^3 voxels. For MRI tasks, we use public human brain IXI-MRI dataset[1], taking 200 T1-weighted MRI for training and 10 volumes for testing. The volumes are resized to $256\times256\times256$ after removing black slices, with near 1 mm^3 voxels.

Implementation Details. (1)Triplane Fitting: The triplane size is set to $128\times 128\times 32$. The weights in Eq.(6) are set to $\lambda_1 = 0.01, \lambda_2 = 0.001, \lambda_3 = 1$. We take Adam optimizer ($\beta_1 = 0.9, \beta_2 = 0.999$) with a learning rate of $1e-3$. The triplane fitting is trained for 8k steps for CT and 40k for MRI. (2) Diffusion Model Training: We take guided diffusion[11] as model architecture, which is compatible with NFD [27] pretrained checkpoint on the triplane embeddings of ShapeNet dataset [4], making it useful for model initialization. The learning rate is $1e-5$ and the batch size is set to 1. The diffusion model trains for 200k steps for CT and 600k for MRI. (3) Guidance-based sampling: The diffusion process includes 1000 steps with a uniform noise schedule. We set $\gamma = 16$ with 256 slices per volume, yielding a 6.25% slice sampling rate per iteration. A larger γ is theoretically better, so we chose the maximum feasible value within our GPU memory limits (24G). The guidance scale is set to $\lambda = 6$ for CT tasks, and $\lambda = 10$ for MRI tasks.

[1] `http://brain-development.org/ixi-dataset/`

Methods	SV-CT		LA-CT	
	PSNR ↑	Average SSIM ↑	PSNR ↑	Average SSIM ↑
DiffusionMBIR [7]	34.78±0.08	0.858±0.001	34.64±0.08	0.852±0.001
TPDM [16]	38.25±0.06	0.949±0.001	38.02±0.07	0.948±0.001
Blaze3DM (Ours)	**38.39**±0.03	**0.951**±0.001	**38.13**±0.04	**0.950**±0.001

Table 1. Quantitative evaluation (PSNR, Average SSIM) of SV-CT and LA-CT. ↑ indicates higher is better. Std is upon 3 repeated experiments.

Methods	CS-MRI		ZSR-MRI	
	PSNR ↑	Average SSIM ↑	PSNR ↑	Average SSIM ↑
DiffusionMBIR [7]	37.21±0.05	0.935±0.002	N/W	
TPDM [16]	39.18±0.04	0.951±0.001	38.42±0.04	0.946±0.001
Blaze3DM (Ours)	**39.35**±0.02	**0.953**±0.001	**38.86**±0.03	**0.950**±0.001

Table 2. Quantitative evaluation (PSNR, Average SSIM) of CS-MRI and ZSR-MRI. N/W: Not Working. ↑ indicates higher is better. Std is upon 3 repeated experiments.

3.2 Results of Inverse Problem Solving

SV-CT and LA-CT. The quantitative results of 36-views SV-CT and 90° LA-CT in Tab.1 demonstrate that our Blaze3DM outperforms DiffusionMBIR [7] and TPDM [16] in both PSNR and SSIM[31] metrics. The visualization results of SV-CT and LA-CT are shown in Fig.2 and Fig.3 respectively. Despite TV regularization, DiffusionMBIR tends to produce layer artifacts along the slice direction. TPDM samples appear overly smooth, potentially losing structures in low-contrast areas. In contrast, Blaze3DM ensures good 3D consistency and provides superior details that closely match ground truth.

CS-MRI and ZSR-MRI. The quantitative evaluation is shown in Tab.2. DiffusionMBIR exhibits layered artifacts in both XZ and YZ planes due to insufficient constraints. While TPDM shows better consistency, it suffers from several jagged artifacts in the edge regions of XY plane, which may be caused by alternative reconstruction in perpendicular directions. In contrast, the Blaze3DM method demonstrates excellent detail in 3D space and achieves state-of-the-art performance.

Efficiency Analysis. We compare the inference efficiency of Blaze3DM, DiffusionMBIR, and TPDM on four tasks, evaluated by Wall-clock time and FLOPs (Floating Point Operations per second), as shown in Tab.3. For single-volume reconstruction, DiffusionMBIR performs slice-by-slice inference along the Z-axis, taking approximately 1 day. TPDM uses two perpendicular diffusion models for slice-by-slice inference, taking nearly 1.5 days. In contrast, Blaze3DM uses a

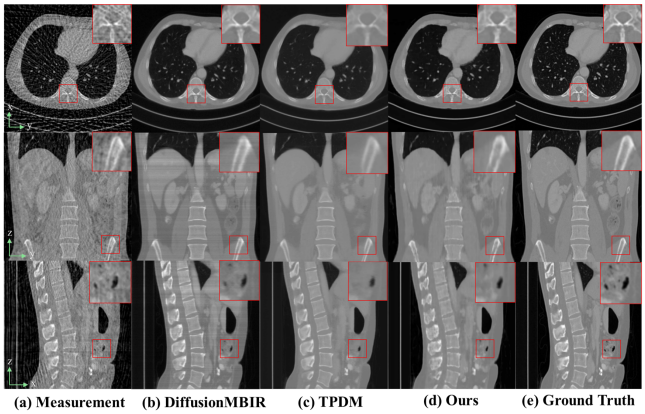

Fig. 2. 36-view SV-CT reconstruction results of the test volume of AAPM CT dataset. (**First row**: axial plane; **Second row**: coronal plane; **Third row**: sagittal plane)

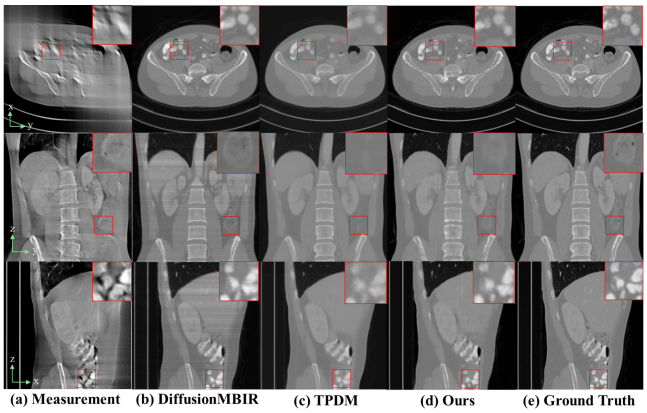

Fig. 3. 90° LA-CT reconstruction results of the test volume of AAPM CT dataset. (**First row**: axial plane; **Second row**: coronal plane; **Third row**: sagittal plane)

triplane diffusion model for a single inference and takes only 1 hour,thereby reducing the time complexity from $O(S)$ to $O(1)$, where S denotes the number of slices. Overall, Blaze3DM is nearly 22× faster than DiffusionMBIR and 40× faster than TPDM, while requiring approximately half the FLOPs.

4 Conclusion

We propose Blaze3DM, an efficient 3D medical volume generative model that addresses volumetric inconsistency, reconstruction artifacts, and low efficiency. Using a triplane neural field and a diffusion model, it enables high-fidelity syn-

Methods	Time (h)	Memory (G)	FLOPs (G)
DiffusionMBIR	18.0±2	16.2±0.2	663.3
TPDM	32.1±3	21.5±0.3	661.0
Blaze3DM	**0.8±0.2**	22.1±0.3	**349.7**

Table 3. Inference time, memory costs and FLOPs for single volume reconstruction.

thesis and zero-shot inverse problem solving. Experiments show Blaze3DM offers competitive performance with significant speedup in inference time.

Acknowledgements. This work is supported in part by the National Natural Science Foundation of China (grant 92354307), the National Key Research and Development Program of China (grant 2024YFF0729202), the Strategic Priority Research Program of the Chinese Academy of Sciences (grant XDA0460305), the Fundamental Research Funds for the Central Universities (grant E3E45201X2) and the Young Scientists Fund of the National Natural Science Foundation of China (Grand No. 12401675).

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

References

1. Abu-Hussein, S., Tirer, T., Giryes, R.: Adir: Adaptive diffusion for image reconstruction. arXiv preprint arXiv:2212.03221 (2022)
2. Chan, E.R., Lin, C.Z., Chan, M.A., Nagano, K., Pan, B., De Mello, S., Gallo, O., Guibas, L.J., Tremblay, J., Khamis, S., et al.: Efficient geometry-aware 3d generative adversarial networks. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 16123–16133 (2022)
3. Chan, T.F., Esedoglu, S., Park, F.E., Yip, A.M.: Total variation image restoration: Overview and recent developments. In: Handbook of Mathematical Models in Computer Vision (2006)
4. Chang, A.X., Funkhouser, T., Guibas, L., Hanrahan, P., Huang, Q., Li, Z., Savarese, S., Savva, M., Song, S., Su, H., et al.: Shapenet: An information-rich 3d model repository. arXiv preprint arXiv:1512.03012 (2015)
5. Chung, H., Kim, J., Mccann, M.T., Klasky, M.L., Ye, J.C.: Diffusion posterior sampling for general noisy inverse problems. In: The International Conference on Learning Representations (2023)
6. Chung, H., Lee, S., Ye, J.C.: Decomposed diffusion sampler for accelerating large-scale inverse problems. arXiv preprint arXiv:2303.05754 (2023)
7. Chung, H., Ryu, D., McCann, M.T., Klasky, M.L., Ye, J.C.: Solving 3d inverse problems using pre-trained 2d diffusion models. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 22542–22551 (2023)
8. Chung, H., Sim, B., Ryu, D., Ye, J.C.: Improving diffusion models for inverse problems using manifold constraints. Advances in Neural Information Processing Systems **35**, 25683–25696 (2022)

9. Chung, H., Sim, B., Ye, J.C.: Come-closer-diffuse-faster: Accelerating conditional diffusion models for inverse problems through stochastic contraction. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 12413–12422 (2022)
10. Chung, H., Ye, J.C.: Score-based diffusion models for accelerated mri. Medical image analysis **80**, 102479 (2022)
11. Dhariwal, P., Nichol, A.: Diffusion models beat gans on image synthesis. Advances in neural information processing systems **34**, 8780–8794 (2021)
12. Efron, B.: Tweedie's formula and selection bias. Publications of the American Statistical Association **106**(496), 1602–1614 (2011)
13. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. Advances in neural information processing systems **33**, 6840–6851 (2020)
14. Katsura, M., Matsuda, I., Akahane, M., Sato, J., Akai, H., Yasaka, K., Kunimatsu, A., Ohtomo, K.: Model-based iterative reconstruction technique for radiation dose reduction in chest ct: comparison with the adaptive statistical iterative reconstruction technique. European radiology **22**, 1613–1623 (2012)
15. Kim, K., Ye, J.: Noise2score: Tweedie's approach to self-supervised image denoising without clean images. In: Advances in Neural Information Processing Systems. pp. 864–874 (2021)
16. Lee, S., Chung, H., Park, M., Park, J., Ryu, W.S., Ye, J.C.: Improving 3d imaging with pre-trained perpendicular 2d diffusion models. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 10710–10720 (2023)
17. Li, B., Zhang, Z., Nie, X., Han, C., Hu, Y., Qiu, X., Guo, T.: Styo: Stylize your face in only one-shot. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 39, pp. 4625–4633 (2025)
18. Li, B., Zhang, Z., Yang, X., Wang, X.: Coser: Towards consistent dense multiview text-to-image generator for 3d creation. In: Proceedings of the Computer Vision and Pattern Recognition Conference (CVPR). pp. 2880–2890 (June 2025)
19. McCollough, C.H., Bartley, A.C., Carter, R.E., Chen, B., Drees, T.A., Edwards, P., Holmes III, D.R., Huang, A.E., Khan, F., Leng, S., et al.: Low-dose ct for the detection and classification of metastatic liver lesions: results of the 2016 low dose ct grand challenge. Medical physics **44**(10), e339–e352 (2017)
20. Mildenhall, B., Srinivasan, P.P., Tancik, M., Barron, J.T., Ramamoorthi, R., Ng, R.: Nerf: Representing scenes as neural radiance fields for view synthesis. Communications of the ACM **65**(1), 99–106 (2021)
21. Mittal, A.: Neural radiance fields: Past, present, and future. arXiv preprint arXiv:2304.10050 (2023)
22. Molaei, A., Aminimehr, A., Tavakoli, A., Kazerouni, A., Azad, B., Azad, R., Merhof, D.: Implicit neural representation in medical imaging: A comparative survey. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 2381–2391 (2023)
23. Nichol, A.Q., Dhariwal, P.: Improved denoising diffusion probabilistic models. In: International conference on machine learning. pp. 8162–8171 (2021)
24. Papa, S., Valperga, R., Knigge, D.M., Kofinas, M., Lippe, P., Sonke, J.J., Gavves, E.: How to train neural field representations: A comprehensive study and benchmark. arXiv preprint arXiv:2312.10531 (2023)
25. Qiu, X., Han, C., Zhang, Z., Li, B., Guo, T., Nie, X.: Diffbfr: Bootstrapping diffusion model for blind face restoration. In: Proceedings of the 31st ACM International Conference on Multimedia. pp. 7785–7795 (2023)

26. Rahaman, N., Baratin, A., Arpit, D., Draxler, F., Lin, M., Hamprecht, F., Bengio, Y., Courville, A.: On the spectral bias of neural networks. In: International conference on machine learning. pp. 5301–5310 (2019)
27. Shue, J.R., Chan, E.R., Po, R., Ankner, Z., Wu, J., Wetzstein, G.: 3d neural field generation using triplane diffusion. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 20875–20886 (2023)
28. Song, J., Meng, C., Ermon, S.: Denoising diffusion implicit models. In: International Conference on Learning Representations (2020)
29. Song, Y., Shen, L., Xing, L., Ermon, S.: Solving inverse problems in medical imaging with score-based generative models. arXiv preprint arXiv:2111.08005 (2021)
30. Wang, T., Zhang, B., Zhang, T., Gu, S., Bao, J., Baltrusaitis, T., Shen, J., Chen, D., Wen, F., Chen, Q., et al.: Rodin: A generative model for sculpting 3d digital avatars using diffusion. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 4563–4573 (2023)
31. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE transactions on image processing **13**(4), 600–612 (2004)
32. Whang, J., Delbracio, M., Talebi, H., Saharia, C., Dimakis, A.G., Milanfar, P.: Deblurring via stochastic refinement. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 16293–16303 (2022)
33. Wu, Q., Li, Y., Xu, L., Feng, R., Wei, H., Yang, Q., Yu, B., Liu, X., Yu, J., Zhang, Y.: Irem: high-resolution magnetic resonance image reconstruction via implicit neural representation. In: Medical Image Computing and Computer Assisted Intervention. pp. 65–74 (2021)
34. Wu, R., Liu, R., Vondrick, C., Zheng, C.: Sin3dm: Learning a diffusion model from a single 3d textured shape. In: The Twelfth International Conference on Learning Representations (2024)
35. Xie, Y., Takikawa, T., Saito, S., Litany, O., Yan, S., Khan, N., Tombari, F., Tompkin, J., Sitzmann, V., Sridhar, S.: Neural fields in visual computing and beyond. Computer Graphics Forum **41** (2021)