Sparse3Diff: A Diffusion Framework for 3D Reconstruction from Sparse 2D Slices in Volumetric Optical Imaging

Hyun Jung Lee[1†], Eunjung Jo[1†], Minjoo Lim[1], Young-Han Son[1], Bogyeong Kang[1], Hyeonyeong Nam[1], Ji-Hoon Jeong[2], Dong-Hee Shin[1], and Tae-Eui Kam[1]

[1] Korea University, Seoul 02841, South Korea
{hyulee,jeju1993do}@korea.ac.kr
[2] Chungbuk National University, Cheongju 28644, South Korea

Abstract. Volumetric optical imaging is an essential tool for understanding various biological processes. However, due to the inherent limitations, such as long imaging time, volume scanning techniques reduce volumetric information into sparse 2D slices. Although many deep learning methods attempt to reconstruct 3D volumes from sparse slices, they struggle with out-of-distribution (OOD) data, which arises from the diversity of biological structures and the limited structural information in sparse slices. To overcome these challenges, we propose Sparse3Diff, a novel diffusion-based framework that reconstructs high-fidelity 3D volumes from sparse 2D slices. Sparse3Diff incorporates a sparse slice-guided position-aware diffusion process that utilizes sparse slices as guidance and conditions on z-position to maintain structural coherence along the z-axis. Additionally, to achieve stable reconstruction under sparse OOD data, we propose a self-alignment strategy that enables the model to be gradually fine-tuned by leveraging its own inferred slices as self-guidance. Experimental results demonstrate that even with sparse OOD data, the Sparse3Diff achieves accurate 3D reconstruction and remains robust across various scanning datasets.

Keywords: Volumetric optical imaging · 3D reconstruction · Diffusion model · Out-of-distribution data · Generative modeling

1 Introduction

Volumetric optical imaging has significantly advanced our understanding of dynamic biological processes, such as blood flow dynamics [6] and large-scale neuronal activities [17]. However, acquiring full volumetric data is challenging due to limitations in imaging techniques such as long imaging times [1, 12, 18, 25, 26], low temporal resolution [2, 3], and the risk of cellular damage [10, 15]. To mitigate these challenges, volume scanning methods often reduce the full volumetric

†These authors contributed equally to this work.

information into sparse 2D slices [5, 11, 14, 22]. Thus, these scanning techniques produce inaccurate 3D structures due to missing intermediate slices between sparse slices, underscoring the need for accurate 3D reconstruction from sparse slices.

Various deep learning (DL)-based methods [10, 11] have been proposed to reconstruct 3D volumes from sparse slices. Recurrent-MZ [10] trains a model using full volumetric data. However, it lacks a strategy for adapting to out-of-distribution (OOD) datasets, arising from the wide variation in biological structures and the limited availability of full volumetric data. MicroDiffusion [11] takes a different approach by training only on sparse slices through a two-stage process. It first generates a coarse 3D volume using implicit neural representations (INR) [20] and then refines it with a diffusion model [7]. However, the limited structural information in sparse slices results in inaccuracies in the initial coarse reconstruction, which can be amplified during refinement. Moreover, similar to Recurrent-MZ, MicroDiffusion lacks a mechanism for aligning with OOD data, which limits its practical applicability. Therefore, an effective method is needed to reconstruct 3D volumes from sparse slices with limited structural information while remaining robust against OOD data.

In this paper, we introduce Sparse3Diff, a novel diffusion-based framework for reconstructing 3D volumes from sparse 2D slices by generating intermediate slices between sparse slices. To address the limited information in sparse slices, it is essential to develop a robust pre-trained model, which can be achieved through our sparse slice-guided position-aware diffusion process. In this diffusion process, the sparse slices remain constant, preserving their pixel values at each diffusion time step, thereby enhancing structural consistency along the z-axis. Additionally, by leveraging classifier-free guidance (CFG) [8], we integrate z-position information into the model to ensure that the generated slices accurately correspond to their respective z-positions. To align the pre-trained model to sparse OOD data, we introduce a self-alignment strategy that fine-tunes the pre-trained model to effectively leverage sparse slices. The self-alignment strategy consists of two stages: (1) the pre-trained model infers intermediate slices between the sparse slices, and (2) the model is fine-tuned by performing the sparse slice-guided position-aware diffusion process using the previously inferred slices as self-guidance. In this stage, the loss is computed only on the original sparse slices. This strategy effectively aligns the pre-trained model to perform robustly on sparse OOD data. Evaluations on two scanning datasets demonstrate that our proposed Sparse3Diff surpasses existing methods, exhibiting robustness across diverse scanning datasets. In summary, our contributions are:

1. We propose a sparse slice-guided position-aware diffusion process that iteratively provides guidance by preserving pixel values of sparse slices while utilizing CFG to condition the z-position, enhancing structural coherence along the z-axis.
2. We introduce a self-alignment strategy that fine-tunes the pre-trained model on sparse OOD data by using inferred slices as self-guidance through our proposed diffusion process, enhancing its robustness in sparse OOD data.

3. Our proposed Sparse3Diff outperforms existing methods, demonstrating robust and accurate performance across various scanning datasets.

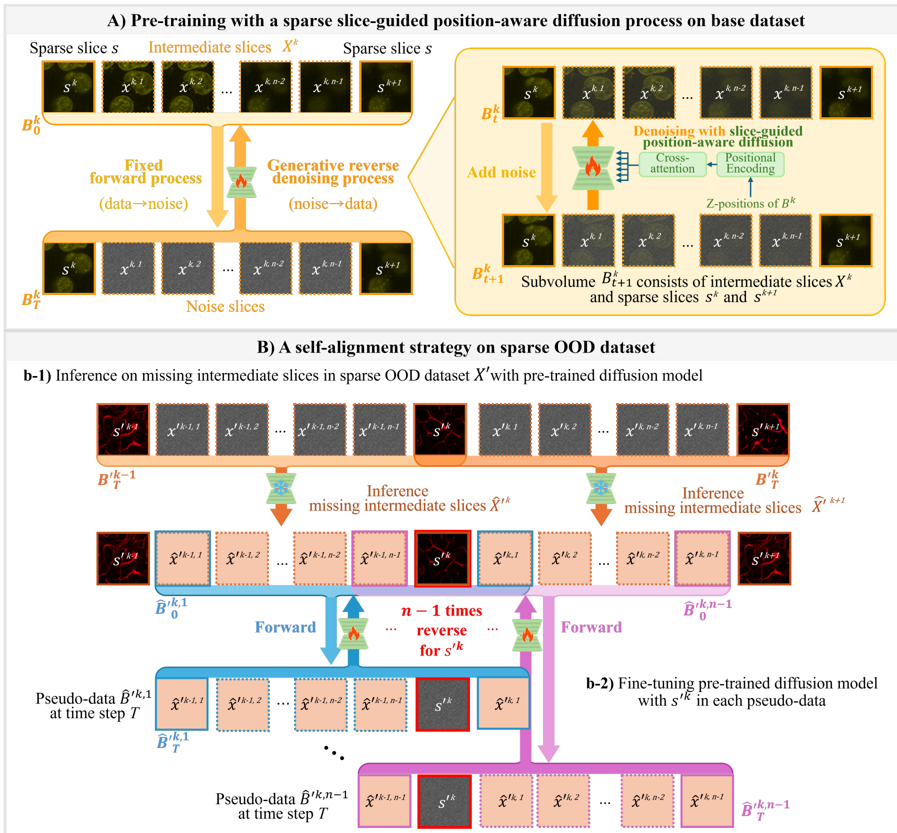

Fig. 1. Overview of Sparse3Diff framework. A) Pre-training the diffusion model with a sparse slice-guided position-aware diffusion process on the base dataset and B) aligning the pre-trained model with a self-alignment strategy on the sparse OOD dataset.

2 Method

2.1 Problem formulation

The 3D volume $\mathbb{B}$ consists of a sequence of consecutive subvolumes B. Each subvolume B^k consists of two consecutive sparse slices, s^k and s^{k+1}, along with the intermediate slices X^k between them, expressed as $B^k = [s^k, X^k, s^{k+1}]$. The sparse slices are sampled from $\mathbb{B}$ at fixed intervals of n, forming the set $\mathbb{S}$. The intermediate slices are denoted as $X^k = \{x^{k,1}, x^{k,2}, \ldots, x^{k,n-1}\}$, consisting of $n-1$ slices. Our goal is to reconstruct the complete 3D volume $\mathbb{B}$ from the sparse slices $\mathbb{S}$. To achieve this, a model f estimates the missing intermediate slices $\hat{X}$ within each subvolume. By integrating all reconstructed subvolumes, the fully sampled 3D volume is obtained.

2.2 Sparse slice-guided position-aware diffusion process

To address the limited information in sparse slices, we introduce a sparse slice-guided position-aware diffusion process to develop a robust pre-trained diffusion model. Unlike conventional diffusion models [7, 21], which generate a single image from noise, our approach directly learns the underlying distribution of intermediate slices together with their corresponding z-positions and sparse slices. Specifically, the sparse slices are preserved at their original pixel values at each diffusion time step, providing iterative guidance, with the forward and reverse processes applied only to the intermediate slices [19]. The forward and reverse processes are defined as follows:

$$q(X_t|X_{t-1}) := \mathcal{N}\left(X_t; \sqrt{1-\beta_t}X_{t-1}, \beta_t\mathbf{I}\right), \tag{1}$$

$$p_\theta(X_t|X_{t-1}) = \mathcal{N}(X_{t-1}; \mu_t, \beta_t\mathbf{I}), \tag{2}$$

where β indicates the noise schedule, and $t \in [0, T]$ denotes the diffusion time steps. An important aspect of our approach is that the sparse slices used as guidance remain unchanged throughout the diffusion process, serving as a stable anchor for generating the intermediate slices.

To improve the model's capability for generating slices that accurately correspond to their positions along the z-axis, we incorporate the z-position of B^k as a conditioning factor c^k using a CFG method. This approach enables high-quality conditional generation by jointly training a conditional and an unconditional diffusion model. Specifically, the z-positions of slices $Z^k \in \mathbb{R}^{n+1}$ in B^k are encoded using sinusoidal positional encoding [23]. The resulting embeddings are concatenated to form c^k, with each embedding defined as follows:

$$c^k_{j,2i} = \sin\left(\frac{z^k_j}{10000^{\frac{2i}{d_c}}}\right), \quad c^k_{j,2i+1} = \cos\left(\frac{z^k_j}{10000^{\frac{2i}{d_c}}}\right), \tag{3}$$

for $j \in \{1, 2, \ldots, n+1\}$ and $i \in \{1, 2, \ldots, \frac{d_c}{2}\}$, where z_j represents the z-position of the j-th slice in B^k, and d_c as the embedding dimension.

During training, with a fixed probability, the condition factor c is randomly discarded from (B, c) to train the unconditional model, while the conditional model incorporates c using cross-attention. Instead of the conventional noise level ϵ-prediction approach, we employ V-loss [13] to train both models, allowing the model to preserve structural details more effectively. The model is updated by minimizing V-loss which measures the difference between the predicted value

$\hat{V}_\theta(X_t, c)$ and the target $V(X_t)$. Here, $V(X_t)$ represents a combination of noise ϵ and the original intermediate slices X_0, formulated as follows:

$$\mathcal{L} = \lambda_t \|V(X_t) - \hat{V}_\theta(X_t, c)\|, \quad V(X_t) = \sqrt{\alpha_t}\,\epsilon - \sqrt{1-\bar{\alpha}_t}\,X_0. \tag{4}$$

where λ indicates the weight of noise. With the pre-trained model, 3D reconstruction is achieved by generating intermediate slices during inference following the sampling process of CFG. The predicted noise for each conditional and unconditional model is linearly combined, as shown in

$$\tilde{\epsilon}_\theta(X_t) = (1+w)\epsilon_\theta(X_t, c) - w\epsilon_\theta(X_t). \tag{5}$$

By leveraging this process, the model enforces z-axis consistency, ensuring coherent alignment of the inferred slices, which improves the quality of the 3D reconstruction.

2.3 Self-alignment strategy for sparse OOD data

The significant diversity of biological structures between 3D volumetric imaging datasets presents a major challenge in addressing sparse OOD data. We conduct experiments in a setting where only sparse slices are available, with no intermediate slices included in the dataset. To effectively handle such cases, we introduce a self-alignment strategy specifically tailored for sparse OOD data as shown in Fig. 1(B). The proposed self-alignment strategy consists of two steps: (1) inferring intermediate slices, and (2) aligning the pre-trained diffusion model to sparse OOD data.

Step 1: Inferring intermediate slices. In each iteration of Step 1, a sparse slice s^k is given as a reference point where $k \in \{2, 3, \ldots, M-1\}$. We identify two adjacent sparse slices, one before s^{k-1}, and one after s^{k+1} the reference sparse slice. The pre-trained diffusion model f is employed to infer the intermediate slices $\hat{X}^{k-1}$ and $\hat{X}^k$ between the sparse slice and its adjacent slices.

Step 2: Aligning the pre-trained diffusion model to sparse OOD data. Next, the inferred intermediate slices $\hat{X}^{k-1}$ and $\hat{X}^k$ are used as self-guidance to fine-tune the pre-trained diffusion model f with our sparse slice-guided position-aware diffusion process. To achieve this, for each s^k, we construct a pseudo-dataset containing $n-1$ pseudo-data $\hat{B}^{k,l} \in \mathbb{R}^{n+1\times H\times W}$ with $l \in \{1, 2, \ldots, n-1\}$. Each pseudo-data comprises consecutive inferred slices along with the sparse slice s^k. This approach establishes a position-dependent spatial context by ensuring that each pseudo-data presents a sparse slice s^k at a different position. By learning the spatial characteristics and distribution of s^k in this position-dependent manner, the model achieves improved alignment, even with sparse OOD data. The self-guided fine-tuning process involves computing a V-loss only between the sparse slice s^k_t and its corresponding slice produced by the diffusion process for each time step. For each s^k, the model is updated $n-1$ times with the use of pseudo-dataset. After completing these steps, a new reference point is selected, and the updated model generates a new pseudo-dataset to further fine-tune the model f. Once the fine-tuning process with

Table 1. Results of the 3D reconstruction on the base dataset. CFG indicates classifier-free guidance. Bold denotes the best results, while underline denotes the second-highest results.

	SSIM ↑			PSNR ↑		
Method	DNA	Membrane	Golgi	DNA	Membrane	Golgi
Recurrent-MZ	0.448	0.330	0.470	22.75	23.36	19.41
Hash	0.671	0.725	0.617	26.07	27.78	21.70
INR	<u>0.695</u>	0.803	0.669	**27.43**	30.49	23.66
MicroDiffusion	0.672	0.769	0.712	26.48	27.29	24.88
DDPM	0.683	0.901	0.700	23.91	28.89	23.47
Ours	**0.745**	<u>0.933</u>	<u>0.819</u>	26.49	<u>31.89</u>	<u>28.30</u>
Ours+CFG	**0.745**	**0.938**	**0.827**	<u>26.55</u>	**32.70**	**29.21**

Table 2. Results of the 3D reconstruction on the sparse OOD dataset. CFG indicates classifier-free guidance and SA denotes the self-alignment strategy. Bold denotes the best results, while underline denotes the second-highest results.

	SSIM ↑			PSNR ↑		
Method	Dendrite	Vasculature	Neuron	Dendrite	Vasculature	Neuron
Recurrent-MZ	0.264	0.143	0.237	22.52	16.43	22.90
Hash	0.481	0.292	0.350	25.05	17.09	<u>23.86</u>
INR	0.571	0.336	**0.367**	25.53	17.70	23.40
MicroDiffusion	0.529	0.359	<u>0.366</u>	25.56	17.70	**24.11**
DDPM	0.476	0.403	0.292	25.89	19.36	22.84
Ours	0.591	<u>0.410</u>	0.307	<u>27.30</u>	19.90	22.75
Ours+CFG	<u>0.593</u>	<u>0.410</u>	0.310	27.24	<u>19.96</u>	23.17
Ours+CFG+SA	**0.627**	**0.424**	0.335	**27.43**	**20.13**	23.59

the self-alignment strategy is completed, a final inference step is performed to generate the full 3D volume. This alignment process maximizes the use of sparse slices, enabling the model to generate high-fidelity volumetric reconstruction with sparse OOD data, and enhancing its applicability to real-world scenarios.

3 Experiments

3.1 Experimental setup

Datasets. In this study, we conducted experiments on two datasets: the Allen Cell dataset [24] and the MicroDiffusion dataset [11]. The Allen Cell dataset, used as the base dataset for pre-training the diffusion model, comprises 3D data of DNA, membranes, and the Golgi apparatus. The MicroDiffusion dataset contains 3D data of dendrites, vasculature, and neurons and serves as sparse OOD data. Both datasets were obtained using the widely adopted laser scanning method.
Implementation details. To evaluate the robustness of Sparse3Diff across various scanning datasets, we simulated the use of a non-diffracting beam scanning following the procedure described in MicroDiffusion [4] and generated a synthetic non-diffracted beam dataset from the laser-scanned dataset. Unless otherwise specified, all experiments were conducted on the laser scanning dataset. In all experiments, we used an interval of $n = 6$.

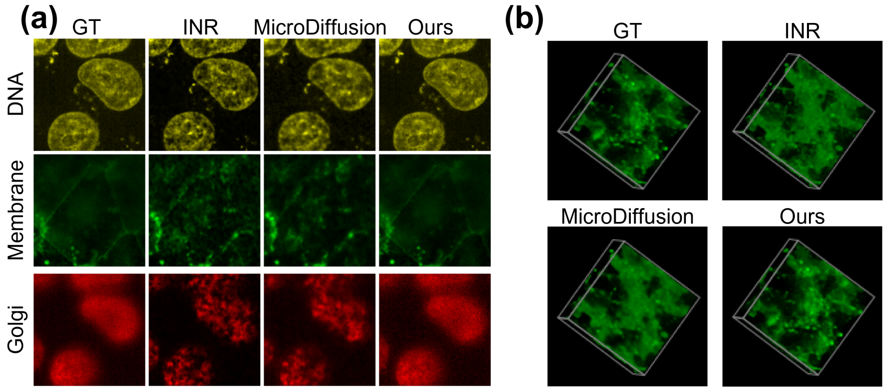

Fig. 2. Visualization results on the base dataset (a) 2D visualizations and (b) 3D visualizations of the membrane.

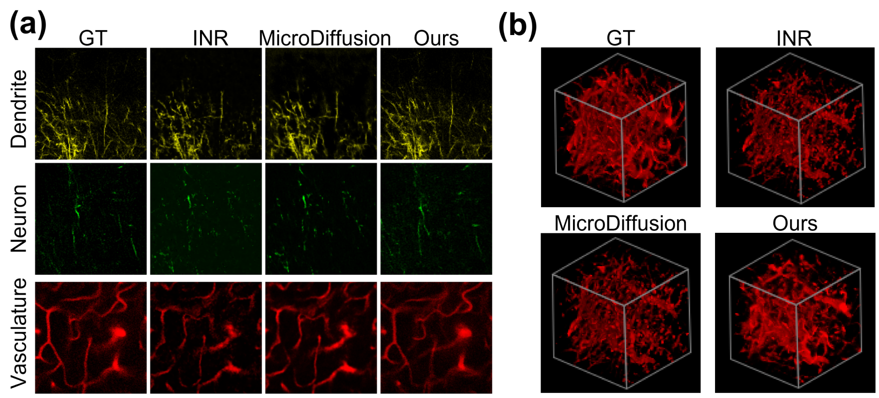

Fig. 3. Visualization results on the sparse OOD dataset (a) 2D visualizations and (b) 3D visualizations of the vasculature.

Competing methods and metrics. To evaluate the 3D reconstruction performance of Sparse3Diff, we compared it against several methods: the RNN-based Recurrent-MZ [10], INR [20], MicroDiffusion [11], Hash encoding [16], and DDPM [7]. For the evaluation metrics, we used structural similarity index map (SSIM) and peak signal-to-noise ratio (PSNR) [9].

3.2 Experimental results

Performance evaluation on base dataset. We compared the 3D reconstruction performance of Sparse3Diff against competing methods on three objects, as shown in Table 1. Under the sparse setting, the competing methods struggle to achieve structurally consistent reconstruction. Notably, this result sug-

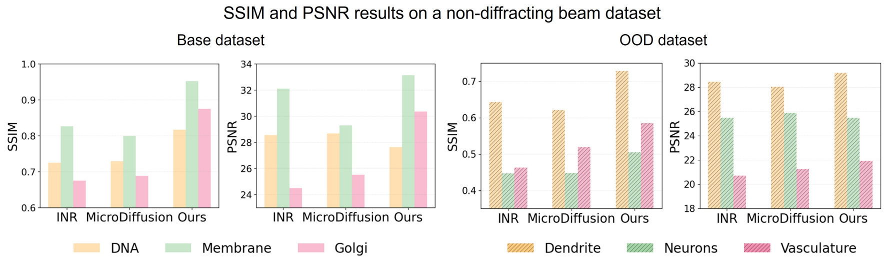

Fig. 4. Performance comparison between our proposed method and competing methods on a non-diffracting beam dataset.

gests that applying the diffusion model to the INR result may contribute to MicroDiffusion's performance drop, possibly amplifying INR's errors. In contrast, Sparse3Diff effectively reconstructs the 3D volume by leveraging a sparse slice-guided position-aware diffusion process. Its performance improves further, surpassing other comparison models with an average increase of 13.38% in SSIM and 12.32% in PSNR. This superior performance is illustrated in Fig. 2, where the reconstructed results of our method show higher similarity to the ground truth (GT).

Performance evaluation on sparse OOD dataset. To evaluate the effectiveness of our self-alignment strategy, we conducted experiments on the sparse OOD dataset. As shown in Table 2, our method, which leverages the pre-trained model, outperforms competing methods. The incorporation of the self-alignment strategy further enhances overall performance, yielding improvements of up to 13.96% in SSIM and 10.53% in PSNR. These improvements are illustrated in Fig. 3, demonstrating that the proposed self-alignment strategy effectively aligns the model with sparse OOD data.

Performance evaluation on non-diffracting beam dataset. We conducted an additional experiment to evaluate the performance of the proposed method in a non-diffracting beam dataset. As illustrated in Fig. 4, our method outperforms existing methods in the non-diffracting beam dataset. These findings demonstrate that our model achieves greater robustness across various scanning datasets compared to competing methods.

4 Conclusion

In this paper, we propose a novel diffusion-based framework, Sparse3Diff, for high-fidelity 3D reconstruction from sparse 2D slices. To ensure structural consistency along the z-axis, we incorporate a sparse slice-guided position-aware diffusion process. Moreover, to align the pre-trained model to sparse OOD data, we propose a self-alignment strategy. Experimental results on both the base and sparse OOD datasets demonstrate that Sparse3Diff outperforms existing methods in both reconstruction accuracy and structural coherence while remaining practical in sparse OOD data. Furthermore, our additional study demonstrates the robustness of our proposed method across various scanning datasets. These findings highlight the potential of Sparse3Diff in revolutionizing 3D reconstruction technologies with high-quality results and robust adaptability across diverse applications.

Acknowledgments. This work was supported by the National Research Foundation of Korea (NRF) grant funded by MSIT (Ministry of Science and ICT) (No. RS-2023-00212498). Furthermore, it was supported by MSIT, Korea, under the ITRC (Information Technology Research Center) support program (IITP-2025-RS-2024-00436857) supervised by the IITP (Institute for Information & Communications Technology Planning & Evaluation). Additionally, this work was supported by grants from the IITP—specifically, the Artificial Intelligence Graduate School Program at Korea University (No. RS-2019-II190079). Finally, this work was also supported by the Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (RS-2024-00457882, Artificial Intelligence Research Hub Project).

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

References

1. Baumgart, E., Kubitscheck, U.: Scanned light sheet microscopy with confocal slit detection. Optics express **20**(19), 21805–21814 (2012)
2. Boden, A., Pennacchietti, F., Coceano, G., Damenti, M., Ratz, M., Testa, I.: Volumetric live cell imaging with three-dimensional parallelized resolft microscopy. Nature Biotechnology **39**(5), 609–618 (2021)
3. Fanous, M.J., Popescu, G.: Ganscan: continuous scanning microscopy using deep learning deblurring. Light: Science & Applications **11**(1), 265 (2022)
4. Garini, Y., Vermolen, B.J., Young, I.T.: From micro to nano: recent advances in high-resolution microscopy. Current opinion in biotechnology **16**(1), 3–12 (2005)
5. Guo, Y., Huang, Y., Li, J., Wang, L., Yang, Z., Liu, J., Peng, X., Yan, W., Qu, J.: Deep penetration microscopic imaging with non-diffracting airy beams. Membranes **11**(6), 391 (2021)
6. Guo, Y., Wang, L., Luo, Z., Zhu, Y., Gao, X., Weng, X., Wang, Y., Yan, W., Qu, J.: Dynamic volumetric imaging of mouse cerebral blood vessels in vivo with an ultralong anti-diffracting beam. Molecules **28**(13), 4936 (2023)
7. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. Advances in neural information processing systems **33**, 6840–6851 (2020)
8. Ho, J., Salimans, T.: Classifier-free diffusion guidance. In: NeurIPS 2021 Workshop on Deep Generative Models and Downstream Applications (2021), https://openreview.net/forum?id=qw8AKxfYbI
9. Hore, A., Ziou, D.: Image quality metrics: Psnr vs. ssim. In: 2010 20th international conference on pattern recognition. pp. 2366–2369. IEEE (2010)
10. Huang, L., Chen, H., Luo, Y., Rivenson, Y., Ozcan, A.: Recurrent neural network-based volumetric fluorescence microscopy. Light: Science & Applications **10**(1), 62 (2021)

11. Hui, M., Wei, Z., Zhu, H., Xia, F., Zhou, Y.: Microdiffusion: Implicit representation-guided diffusion for 3d reconstruction from limited 2d microscopy projections. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11460–11469 (2024)
12. Im, K.B., Han, S., Park, H., Kim, D., Kim, B.M.: Simple high-speed confocal line-scanning microscope. Optics express **13**(13), 5151–5156 (2005)
13. Lin, S., Liu, B., Li, J., Yang, X.: Common diffusion noise schedules and sample steps are flawed. In: Proceedings of the IEEE/CVF winter conference on applications of computer vision. pp. 5404–5411 (2024)
14. Massaro, G., Giannella, D., Scagliola, A., Lena, F.D., Scarcelli, G., Garuccio, A., Pepe, F.V., D'Angelo, M.: Light-field microscopy with correlated beams for high-resolution volumetric imaging. Scientific Reports **12**(1), 16823 (2022)
15. Morales-Curiel, L.F., Gonzalez, A.C., Castro-Olvera, G., Lin, L.C., El-Quessny, M., Porta-de-la Riva, M., Severino, J., Morera, L.B., Venturini, V., Ruprecht, V., et al.: Volumetric imaging of fast cellular dynamics with deep learning enhanced bioluminescence microscopy. Communications biology **5**(1), 1330 (2022)
16. Müller, T., Evans, A., Schied, C., Keller, A.: Instant neural graphics primitives with a multiresolution hash encoding. ACM transactions on graphics (TOG) **41**(4), 1–15 (2022)
17. Prevedel, R., Yoon, Y.G., Hoffmann, M., Pak, N., Wetzstein, G., Kato, S., Schrödel, T., Raskar, R., Zimmer, M., Boyden, E.S., et al.: Simultaneous whole-animal 3d imaging of neuronal activity using light-field microscopy. Nature methods **11**(7), 727–730 (2014)
18. Ren, Y.X., Wu, J., Lai, Q.T., Lai, H.M., Siu, D.M., Wu, W., Wong, K.K., Tsia, K.K.: Parallelized volumetric fluorescence microscopy with a reconfigurable coded incoherent light-sheet array. Light: Science & Applications **9**(1), 8 (2020)
19. Saharia, C., Ho, J., Chan, W., Salimans, T., Fleet, D.J., Norouzi, M.: Image super-resolution via iterative refinement. IEEE transactions on pattern analysis and machine intelligence **45**(4), 4713–4726 (2022)
20. Sitzmann, V., Martel, J., Bergman, A., Lindell, D., Wetzstein, G.: Implicit neural representations with periodic activation functions. Advances in neural information processing systems **33**, 7462–7473 (2020)
21. Song, J., Meng, C., Ermon, S.: Denoising diffusion implicit models. arXiv preprint arXiv:2010.02502 (2020)
22. Tan, X.J., Kong, C., Ren, Y.X., Lai, C.S., Tsia, K.K., Wong, K.K.: Volumetric two-photon microscopy with a non-diffracting airy beam. Optics Letters **44**(2), 391–394 (2019)
23. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. Advances in neural information processing systems **30** (2017)
24. Viana, M.P., Chen, J., Knijnenburg, T.A., Vasan, R., Yan, C., Arakaki, J.E., Bailey, M., Berry, B., Borensztejn, A., Brown, E.M., et al.: Integrated intracellular organization and its variations in human ips cells. Nature **613**(7943), 345–354 (2023)
25. Wang, W., Lemke, F., Wapler, M.C., Wallrabe, U., Czarske, J.W., Koukourakis, N.: 3d-scanning microscopy with adaptive lenses and prisms for zebrafish studies. Journal of Optical Microsystems **1**(2), 024501–024501 (2021)
26. Winter, P.W., Shroff, H.: Faster fluorescence microscopy: advances in high speed biological imaging. Current opinion in chemical biology **20**, 46–53 (2014)