Self-supervised Axial Super-Resolution for Volume Microscopy via Diffusion-Guided Structure Distillation

Bohao Chen[1,2], Yanchao Zhang[2,3], Yanan Lv[2,4], Hua Han[2,3], and Xi Chen[2]

[1] School of Advanced Interdisciplinary Sciences, University of Chinese Academy of Sciences, Beijing, China
[2] State Key Laboratory of Brain Cognition and Brain-inspired Intelligence Technology, Institute of Automation, Chinese Academy of Sciences, Beijing, China
xi.chen@ia.ac.cn
[3] School of Future Technology, University of Chinese Academy of Sciences, Beijing, China
[4] School of Artificial Intellengace, University of Chinese Academy of Sciences, Beijing, China

Abstract. Anisotropic resolution remains a fundamental challenge in 3D microscopy, where axial resolution is significantly lower than lateral resolution due to physical limitations. To address this, we propose a self-supervised volume super-resolution (VSR) framework named Diffusion to Resolution (D2R), which leverages 2D diffusion priors to enhance axial resolution without requiring high-resolution (HR) volume as supervision. D2R consists of three stages: (1) learning biological priors via a 2D diffusion model trained on high-resolution XY slices, (2) generating pseudo-HR lateral (XZ/YZ) volumes through cross-plane fusion, and (3) performing stable structure distillation to train a 3D VSR network. To further improve VSR quality, we introduce Axial Enhancement Network (AENet), a 3D VSR model incorporating lightweight channel attention to enhance fine details while maintaining inter-slice continuity. Extensive experiments on FIB-SEM datasets demonstrate that D2R-AENet outperforms state-of-the-art self-supervised methods in both image similarity and membrane segmentation accuracy, achieving performance close to supervised approaches. These results validate the effectiveness of our framework in high-fidelity volumetric reconstruction under practical conditions where HR references are unavailable. Codes are available at https://github.com/hmzawz2/D2R-models.

Keywords: Isotropic reconstruction · Diffusion models · Volume Microscopy.

1 Introduction

Recent advances in 3D microscopy, including both light microscopy (LM) and electron microscopy (EM), have enabled high-resolution visualization of spatial biological structures, capturing intricate cellular and subcellular architectures.

However, a fundamental limitation of 3D microscopy imaging is its inherent anisotropic resolution, where the axial resolution is significantly lower than the lateral resolution due to physical limitations. This anisotropy hinders accurate morphological analysis and downstream tasks such as membrane segmentation and neuron reconstruction. To address this issue, volume super-resolution (VSR) techniques have been explored to enhance axial resolution, reconstructing high-resolution (HR) volumes from axially low-resolution (LR) 3D microscopy data.

Traditional interpolation methods, such as linear and cubic interpolation, provide fast but blurry axial reconstruction, especially on noisy or complex biological structures. Recently, deep learning-based methods have made great advances in VSR of microscopy volumes. Supervised methods such as SRUNet[5], generate HR and LR volumes along the axial direction using isotropic EM data to train 3D UNet[16] network, demonstrating the feasibility of recovering HR volumes from LR inputs. However, isotropic HR volumes are challenging to obtain in practice, limiting its applicability. Recent studies predominantly utilize 2D convolution networks or generative models to restore independent lateral (XZ/YZ) slices of volumes, achieving VSR along lateral directions. IsoRecon[3] leveraged a Cycle-GAN framework to train lateral super-resolution models, offering a data-driven approach to isotropic reconstruction. With diffusion models emerging as powerful generative methods for image restoration tasks, DiffuseIR[15] and the methods proposed in [8,9], offer better VSR quality empowered by diffusion models. However, since the reconstructed slices are independently processed, both Cycle-GAN and diffusion-based methods struggle to preserve 3D continuity, often leading to misalignment artifacts and excessive inference time. Deep frame interpolation methods, such as STDIN[18] and vEMDiffuse-i[11], have notably enhanced volume resolution and structure continuity along the axial direction (Z-axis). However, these methods only perform reliably under fixed super-resolution scaling factors or with HR volumes as supervision. Although vEMDiffuse-a [11] supports arbitrary scaling factors without HR volumes, its axial reconstructions suffer from lateral-axial distribution shift, leading to incorrect details. These limitations underscore the need for a self-supervised VSR training framework that intrinsically aligns cross-plane distributions while preserving 3D continuity.

To address these limitations, we propose Diffusion-to-Resolution (D2R), a training framework for VSR in volumetric imaging. For microscopy data with spatially isotropic structures, D2R first learns biological priors by training a 2D diffusion model on XY slices, then transfers these priors to recover XZ/YZ slices and synthesize pseudo-high-resolution (pseudo-HR) volumes, ensuring consistent structural fidelity across planes. Subsequently, a 3D VSR model is trained for volume reconstruction. During training, randomly introduced structure errors from diffusion sampling are gradually eliminated through average loss optimization, leading the model to learn a reliable structure transformation. Additionally, we introduce Axial Enhancement Network (AENet), a VSR network with channel attention to enhance details while preserving inter-slice continuity by 3D convolution. Experiments on synthetic anisotropic EM datasets demonstrate that

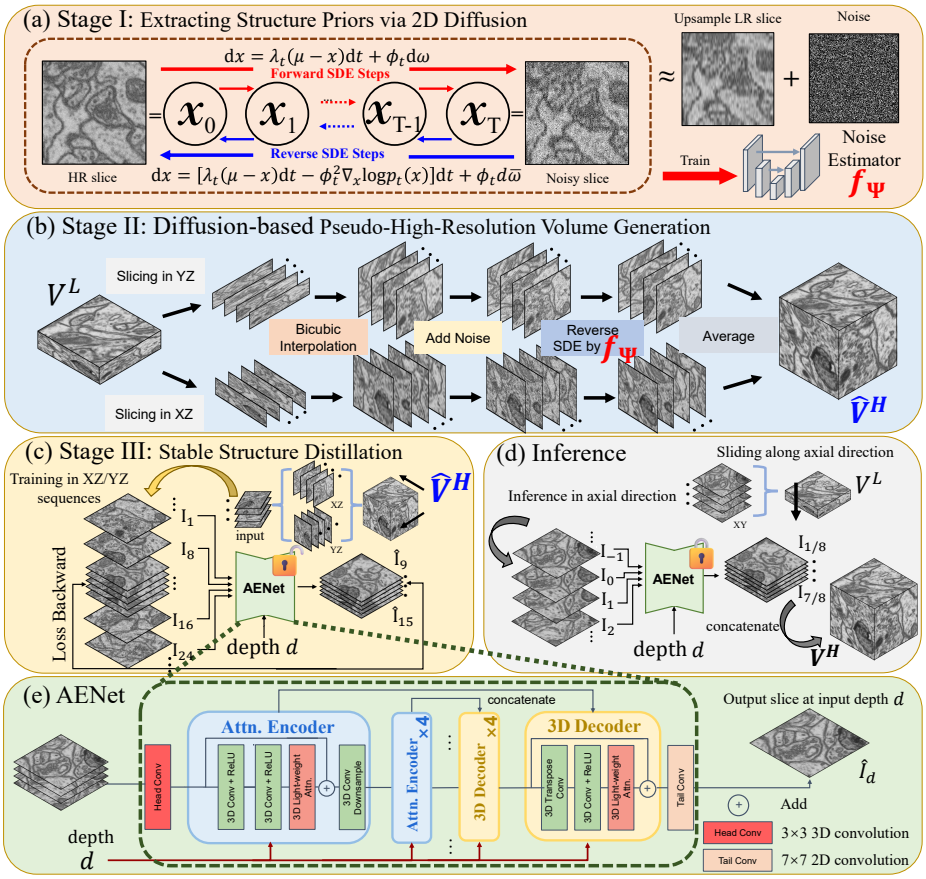

Fig. 1. An overview of the proposed D2R training framework and the AENet.

D2R-AENet achieves superior performance in similarity metrics for 8× VSR, with results showing notable improvements in membrane segmentation tasks.

Our contributions are summarized as follows: (1) We propose D2R, a self-supervised training framework that leverages 2D diffusion priors for 3D VSR task. (2) We introduce AENet, a novel 3D VSR network designed to enhance details and maintain spatial consistency across slices. (3) We evaluate synthetic anisotropic volumetric microscopy data, showing our method outperforms other VSR methods in reconstruction quality and membrane segmentation accuracy.

2 Proposed Method

As shown in Fig. (1), our D2R framework enables self-supervised VSR through three stages: (I) structure prior extraction, (II) pseudo-HR volume generation, and (III) stable structure distillation. A 2D diffusion model is first trained on XY

slices to capture biological priors (Stage I), and then applied slice-by-slice to the XZ/YZ planes of LR volumes to produce a pseudo-HR volume $\hat{V}^H$ (Stage II). To address inconsistencies from independent 2D processing, the VSR network is trained solely on the more reliable XZ/YZ sequences of $\hat{V}^H$ (Stage III). The trained model is then applied to the input LR volume for HR inference.

2.1 Stage I: Extracting Structure Priors via 2D Diffusion

Consider an image degradation process approximated by a stochastic differential equation (SDE), where a high-resolution image $I^H = x_0$ gradually deteriorates into a noisy version x_T, modeled as $x_T \approx \text{upsample}(I^L) + \epsilon$ with $\epsilon \sim \mathcal{N}(0, \delta^2)$. This degradation process follows the SDE:

$$\mathrm{d}x = \lambda_t(\mu - x)\mathrm{d}t + \phi_t \mathrm{d}\omega, \tag{1}$$

where ω refers to a standard Wiener process. λ_t and ϕ_t control the speed of mean reversion and stochastic volatility, respectively. To construct a closed-form solution of Eq. (1), the state x_t follows a Gaussian distribution with mean $m_t(x) = \mu + (x_0 - \mu)e^{-\bar{\lambda}_t}$ and variance $n_t = \delta^2(1 - e^{-2\bar{\lambda}_t})$. The reverse process reconstructs x_0 by solving:

$$\mathrm{d}x = \left[\lambda_t(\mu - x) - \phi_t^2 \nabla_x \log p_t(x)\right] \mathrm{d}t + \phi_t \mathrm{d}\bar{\omega}, \tag{2}$$

where $\bar{\omega}$ denotes a reverse-time Wiener process. $\nabla_x \log p_t(x)$ is estimated during training via conditional scores $\nabla_x \log p_t(x|x_0) = -(x_t - m_t(x))/n_t$. To stabilize training, we reparameterize $x_t = m_t(x) + \sqrt{n_t}\sigma_t$ and approximate σ_t using a network $f_\Psi(\cdot)$. We compute the Euclidean distance between the predicted noise and the ground-truth noise, follows a likelihood objective to train f_Ψ:

$$\mathcal{L}(\phi) = \sum_{t=0}^{T} \gamma_t \mathbb{E}\left[\|x_t - (\mathrm{d}x_t)_{f_\Psi} - x^*_{t-1}\|\right], \tag{3}$$

where $(\mathrm{d}x_t)_{f_\Psi}$ denotes the reverse-time SDE in Eq. (2) and its score is predicted by the noise network $f_\Psi(\cdot)$. $x^*_{t-1} = \text{argmin}_{x^*_{t-1}}[-(x_{t-1}|x_t, x_0)]$. For a comprehensive mathematical derivation and other technical details, please refer to [12].

2.2 Stage II: Diffusion-based Pseudo-HR Volume Generation

The trained diffusion model synthesizes pseudo-high-resolution (pseudo-HR) volumes by slice-wise reconstruction. For i-th low-resolution slice (I^L_i) in each orthogonal plane (XZ/YZ), the restoration is as follows:

$$\hat{I}^H_i = f_\Psi\left(\text{upsample}(I^L_i, k) + \epsilon\right), \quad \epsilon \sim \mathcal{N}(0, \delta^2) \tag{4}$$

where $f_\Psi(\cdot)$ represents the diffusion model from Stage I, and $\text{upsample}(\cdot, k)$ applies cubic interpolation to upsample the image along the Z-axis by a factor of

k. The reconstructed slices set on both directions $\{\hat{I}_1^H, ..., \hat{I}_n^H\}_{\{XZ/YZ\}}$ are then concatenated along respective axes and averaged to form the pseudo-HR volume:

$$\hat{V}^H = \frac{1}{2}\left(\text{concat}(\{\hat{I}_1^H, ..., \hat{I}_n^H\}_{XZ}) + \text{concat}(\{\hat{I}_1^H, ..., \hat{I}_n^H\}_{YZ})\right). \tag{5}$$

The pseudo-HR volume $\hat{V}^H$ serves as the training data for Stage III, where diffusion-sampling errors are progressively filtered out during the training of the VSR model, leading to stable structural transformations across slices in inference.

2.3 Stage III: Stable Structure Distillation

Since the XY plane lacks explicit constraints, in Stage III, we solely train VSR networks on lateral planes (XZ/YZ) of pseudo-HR volumes to avoid introducing artifacts in XY direction. The network g_θ optimizes:

$$\theta^* = \underset{\theta}{\text{argmin}}\ \mathbb{E}_{(V^L, \hat{V}^H)} \left\| g_\theta(V^L) - \hat{V}^H \right\|, \tag{6}$$

which forces g_θ to learn consensus structural transformations across noisy samples, suppressing artifacts while distilling stable structures cross slices like organelles and membranes.

Although the D2R training framework is applicable to any VSR methods, here, we introduce Axial Enhancement Network (AENet), which performs VSR through 3D convolutions that explicitly enforce spatial continuity and details.

AENet Details By predicting an intermediate slice $\hat{I}_d$ from a 4-layer input sequence at relative depths d, AENet enables flexible VSR scaling factors r through slice-wise estimation $\hat{I}_{i/r} = g_\theta(I_{1:4}, i/r), i \in \{1, \ldots, r-1\}$. As shown in Fig. 1(e), it employs 3D convolutions with relative depth encoding to enforce axial continuity, with a lightweight 3D channel attention module [2] enhances fine details as follows:

$$f_o = \sigma(W \cdot \text{pool}(f_i + b)) \odot f_i \tag{7}$$

where f_i and f_o denote input/output 3D features, $W \in \mathbb{R}^{C\times C}$ and $b \in \mathbb{R}^C$ are learnable parameters. As AENet employs 3D ResNet as backbone, it equips with multiscale skip connections to facilitate feature fusion, and the decoder refines predictions via transposed convolutions. This design ensures high-fidelity reconstruction while maintaining computational efficiency in VSR task.

Loss Function The loss function combines loss for structural fidelity and loss for high-frequency details:

$$L_{\text{total}} = L_1 + \lambda_{\text{SSIM}} L_{\text{SSIM}} + \lambda_{\text{FFL}} L_{\text{FFL}}, \tag{8}$$

where L_1 and L_{SSIM} preserve low-frequency structures and L_{FFL} [6] prioritize high-frequency detail recovery through gradient-aware focal modulation. In our experiments, we set $\lambda_{\text{SSIM}} = 1$ and $\lambda_{\text{FFL}} = 100$, balancing the low-frequency and high-frequency details in microscopy slices.

Inference As illustrated in Fig. 1(d), during inference, the trained AENet model processes the input LR volume V^L along the Z-axis using a sliding window. At each step, it takes a sequence of adjacent XY slices (e.g., 4 in AENet) and inserts $(r-1)$ intermediate slices between the central slice pair, where r is the super-resolution factor. The original and predicted slices are then concatenated to reconstruct the output HR volume V^H.

3 Experiments

3.1 Datasets and implement details

To simulate realistic anisotropic EM imaging, we simulate LR volumes by retaining every r-th axial slice ($r = 8$) from the original HR volumes, preserving voxel independence and original noise distribution. This slice sampling strategy avoids structure blending and noise suppression in average downsampling [11]. Experiments use two public FIB-SEM datasets: **FIB-25** [17]: Drosophila visual circuits (10 nm isotropic resolution); **EPFL** [1]: Mouse hippocampus (5 nm isotropic resolution). Downsampling yields LR volumes with 80 nm (FIB-25) and 40 nm (EPFL) axial resolution. Each dataset is partitioned into training (70%), validation (15%), and test (15%), with non-overlapping subvolumes reserved for resolution analysis and downstream tasks. During testing, all methods reconstruct subvolumes to fully assess restoration capability from XY/XZ/YZ planes. AENet is trained using ADAM [7], with an initial learning rate of 2×10^{-4}, halved when training plateaus. After 60 epochs, the best model on validation dataset is selected. All experiments run on a server with an Nvidia V100 GPU.

3.2 Comparison with VSR Methods

We evaluate AENet against bicubic interpolation (as baseline) and other SOTA VSR methods, including methods trained with HR volumes as supervision (SRUNet [5], vEMDiffuse-i[11]) and methods that perform VSR using only LR volumes (IsoVEM [4], vEMDiffuse-a[11] and Lee et al.[8], list as self-supervised methods). To validate the adaptability of the D2R framework, we use it to train SRUNet (D2R-SRUNet) and propose two variants of AENet: Sup-AENet, trained with HR volume supervision, and D2R-AENet, which is trained with D2R framework.

Quantitative Evaluation Table 1 presents the evaluation results based on image similarity and estimated resolution of reconstructed subvolumes. Restoration results demonstrate our D2R framework is highly adaptable, enabling self-supervised training while maintaining strong performance. Notably, both D2R-SRUNet and D2R-AENet achieve results comparable to their supervised counterparts. In contrast, vEMDiffuse[11] show notable degradation when transitioning from supervised (vEMDiffuse-i) to self-supervised (vEMDiffuse-a), underscoring the effectiveness of our D2R framework in training VSR networks.

Table 1. Quantitative evaluation on FIB-25 and EPFL datasets with supervised (Sup.) and self-supervised (Self-sup.) methods. Bold/underline indicate best/second-best.

Dataset	Type	Methods	PSNR(↑)			SSIM(↑)			Reso. (nm)↓
			XY	XZ	YZ	XY	XZ	YZ	
FIB-25	Sup.	SRUNet [5]	<u>27.01</u>	<u>27.24</u>	<u>27.25</u>	<u>0.6942</u>	<u>0.7193</u>	<u>0.7073</u>	<u>45.49</u>
		vEMDiffuse-i [11]	25.86	26.16	26.16	0.6292	0.6331	0.6185	57.75
		Sup-AENet (ours)	**27.89**	**28.24**	**28.24**	**0.7202**	**0.7442**	**0.7315**	**44.31**
	Self-sup.	Baseline	23.05	22.69	22.69	0.5183	0.5056	0.4917	57.11
		IsoVEM [4]	24.67	24.96	24.96	0.5952	0.6332	0.6161	54.68
		Lee et al. [8]	25.53	25.87	25.87	0.5700	0.6088	0.5939	59.08
		vEMDiffuse-i [11]	23.38	23.74	23.74	0.5072	0.5192	0.5033	63.46
		D2R-SRUNet	<u>27.01</u>	<u>27.24</u>	<u>27.24</u>	<u>0.6895</u>	<u>0.7045</u>	<u>0.6919</u>	**43.93**
		D2R-AENet (ours)	**27.64**	**27.83**	**27.83**	**0.7023**	**0.7070**	**0.6930**	<u>46.73</u>
EPFL	Sup.	SRUNet [5]	<u>25.79</u>	<u>26.18</u>	<u>26.18</u>	<u>0.6335</u>	<u>0.6837</u>	<u>0.6619</u>	<u>23.80</u>
		vEMDiffuse-i [11]	24.88	25.41	25.41	0.5501	0.5986	0.5733	27.05
		Sup-AENet (ours)	**26.35**	**26.89**	**26.90**	**0.6402**	**0.6861**	**0.6634**	**21.97**
	Self-sup.	Baseline	22.22	22.29	22.30	0.4328	0.4464	0.4261	31.73
		IsoVEM [4]	23.43	23.89	23.89	0.4968	0.5582	0.5290	27.63
		Lee et al. [8]	24.38	24.92	24.92	0.5035	0.5654	0.5396	27.64
		vEMDiffuse-a [11]	23.25	23.73	23.73	0.4965	0.5302	0.5056	30.60
		D2R-SRUNet	<u>25.54</u>	<u>25.99</u>	<u>25.99</u>	<u>0.6187</u>	**0.6491**	**0.6470**	<u>24.95</u>
		D2R-AENet (ours)	**26.43**	**26.63**	**26.63**	**0.6245**	<u>0.6445</u>	<u>0.6365</u>	**22.95**

The proposed AENet achieves superior performance in both supervised and self-supervised setting, outperforming all VSR methods that rely on supervision and significantly surpassing existing self-supervised approaches when integrated with the D2R framework as D2R-AENet. Resolution analysis via Fourier Shell Correlation (FSC) [14] further confirms that D2R-trained models reach resolution levels comparable to supervised methods, with D2R-SRUNet even exceeding its supervised counterpart in some cases.

Membrane Segmentation Comparison. To assess the impact of VSR on downstream analysis, we evaluate membrane segmentation performance using a public pre-trained model [20] without fine-tuning. The model is applied to reconstructed volumes, and segmentation accuracy is measured via Intersection over Union (IoU)[21], adapted Rand error (ARE)[10], and Variation of Information (VoI)[13], with membrane segmentation on ground truth volumes as reference. As shown in Table 2, AENet consistently achieves top-tier segmentation performance across both training settings. Similar to Sec.3.2, vEMDiffuse undergoes notable performance drops when shifting from supervised to self-supervised training, whereas AENet maintains stability. These results further validate the robustness of our approach in VSR and its effectiveness in producing high-fidelity reconstructions for downstream tasks.

Table 2. Membrane segmentation accuracy with supervised (Sup.) and self-supervised (Self-sup.) methods using different metrics. Bold/underline indicate best/second-best.

Type	Methods	FIB-25			EPFL		
		IoU (↑)	ARE (↓)	VoI (↓)	IoU (↑)	ARE (↓)	VoI (↓)
Sup.	SRUNet[5]	0.6153	0.3504	0.6696	0.6968	0.2732	0.5704
	vEMDiffuse-i[11]	<u>0.6472</u>	<u>0.3149</u>	<u>0.6129</u>	<u>0.7533</u>	<u>0.2199</u>	<u>0.4918</u>
	Sup-AENet (ours)	**0.6874**	**0.2753**	**0.5566**	**0.7604**	**0.2133**	**0.4808**
Self-sup.	Baseline	0.5383	0.4304	0.7689	0.5337	0.4387	0.7834
	vEMDiffuse-a[11]	0.5818	0.3850	0.7149	0.6663	0.3053	0.6238
	IsoVEM[4]	0.5970	0.3709	0.7005	0.6541	0.3191	0.6480
	Lee et.al[8]	0.6143	0.3445	0.6406	<u>0.7093</u>	<u>0.2591</u>	<u>0.5413</u>
	D2R-SRUNet	**0.6779**	**0.2844**	**0.5693**	0.6847	0.2840	0.5823
	D2R-AENet (ours)	<u>0.6676</u>	<u>0.2936</u>	<u>0.5795</u>	**0.7462**	**0.2261**	**0.4999**

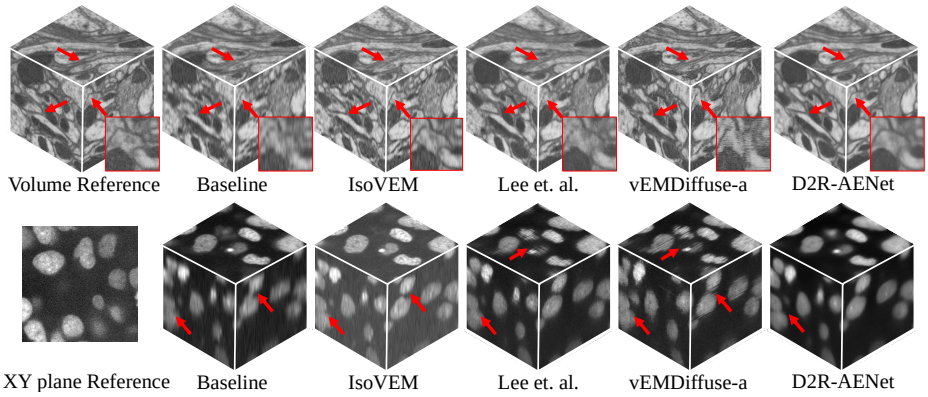

Fig. 2. Visual comparison of VSR results across XY/XZ/YZ planes of self-supervised methods on EM (upper) and LM (lower) datasets. Red arrows indicate hallucination.

Visual Comparisons We provide visual comparisons of reconstruction results on both EM and LM datasets. As shown in Fig. 2, the D2R-AENet effectively restores membranes with minimal artifacts on EM subvolumes (from FIB-25), outperforming baseline and prior VSR methods. On zebrafish retina LM data [19], we enhance the nuclei channel by increasing lateral resolution to match the axial resolution (10×), refining nuclear structures while suppressing noise. These results demonstrate the robustness of our approach in different modalities.

3.3 Ablation Study on Pseudo-HR Generation Scales

Existing self-supervised VSR methods, like vEMDiffuse-a, degrade due to distribution shift between axial and lateral directions in LR training volumes. D2R framework mitigates this by aligning lateral and axial distributions (Stage I-II) before training VSR networks (Stage III). To evaluate the impact of lateral

Table 3. VSR quality under different pseudo-HR volumes scales (k in Eq.(4))

Upsampling Scale (k)	PSNR-XY	PSNR-XZ	PSNR-YZ	SSIM-XY	SSIM-XZ	SSIM-YZ
1× (No upsample)	26.12	26.60	26.60	0.6074	0.6561	0.6375
2×	26.27	26.77	26.77	0.6174	0.6650	0.6452
4×	26.47	26.98	26.98	0.6190	0.6713	0.6532
8×	26.50	27.01	27.01	0.6185	0.6742	0.6535

distribution of pseudo-HR training volumes, we train D2R-AENet with varying upsampling scales k on downsampled EPFL data and conduct ablation experiments on a smaller EPFL test set. As shown in Table 3, slice similarity improves steadily from $k = 1$ to $k = 4$ before plateauing at higher scales, demonstrating that effective stable structure distillation requires neither strict alignment with isotropic resolution nor excessive upsampling; instead, capturing essential structure translation patterns is sufficient. We set $k = 8$ in all D2R experiments.

4 Conclusion

In this work, we propose a VSR training framework named D2R that enables high-quality self-supervised training through 2D diffusion priors. Our D2R framework effectively trains VSR networks without HR volume supervision, achieving performance close to supervised counterparts in both resolution and membrane segmentation. Additionally, we introduce AENet, which is designed for VSR tasks. When integrated with D2R, D2R-AENet outperforms existing self-supervised methods with reliable details. Both similarity metrics and downstream results validate the effectiveness of our method in real world VSR tasks.

Acknowledgments. This work was supported by the Scientific research instrument and equipment development project of Chinese Academy of Sciences under Grant PTYQ2025TD0002, and the National Natural Science Foundation of China under Grant 32171461.

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

References

1. Aurelien, L., Yunpeng, L., Carlos, B., Pascal, F.: https://www.epfl.ch/labs/cvlab/data/data-em/ (2013), accessed: 2024-08-16
2. Choi, M., Kim, H., Han, B., Xu, N., Lee, K.M.: Channel attention is all you need for video frame interpolation. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 34, pp. 10663–10671 (2020)
3. Deng, S., Fu, X., Xiong, Z., Chen, C., Liu, D., Chen, X., Ling, Q., Wu, F.: Isotropic reconstruction of 3d em images with unsupervised degradation learning. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd

International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part V 23. pp. 163–173. Springer (2020), https://https://github.com/sydeng99/IsoRecon
4. He, J., Zhang, Y., Sun, W., Yang, G., Sun, F.: Isovem: Isotropic reconstruction for volume electron microscopy based on transformer. bioRxiv pp. 2023–11 (2023)
5. Heinrich, L., Bogovic, J.A., Saalfeld, S.: Deep learning for isotropic super-resolution from non-isotropic 3d electron microscopy. In: Medical Image Computing and Computer-Assisted Intervention- MICCAI 2017: 20th International Conference, Quebec City, QC, Canada, September 11-13, 2017, Proceedings, Part II 20. pp. 135–143. Springer (2017)
6. Jiang, L., Dai, B., Wu, W., Loy, C.C.: Focal frequency loss for image reconstruction and synthesis. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 13919–13929 (2021)
7. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization (2014)
8. Lee, K., Jeong, W.K.: Reference-free isotropic 3d em reconstruction using diffusion models. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 235–245. Springer (2023)
9. Lee, K., Jeong, W.K.: Reference-free axial super-resolution of 3d microscopy images using implicit neural representation with a 2d diffusion prior. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 593–602. Springer (2024)
10. Liu, T., Jones, C., Seyedhosseini, M., Tasdizen, T.: A modular hierarchical approach to 3d electron microscopy image segmentation. Journal of neuroscience methods **226**, 88–102 (2014)
11. Lu, C., Chen, K., Qiu, H., Chen, X., Chen, G., Qi, X., Jiang, H.: Diffusion-based deep learning method for augmenting ultrastructural imaging and volume electron microscopy. Nature Communications **15**(1), 4677 (2024)
12. Luo, Z., Gustafsson, F.K., Zhao, Z., Sjölund, J., Schön, T.B.: Image restoration with mean-reverting stochastic differential equations. arXiv preprint arXiv:2301.11699 (2023)
13. Meilă, M.: Comparing clusterings—an information based distance. Journal of multivariate analysis **98**(5), 873–895 (2007)
14. Nieuwenhuizen, R.P., Lidke, K.A., Bates, M., Puig, D.L., Grünwald, D., Stallinga, S., Rieger, B.: Measuring image resolution in optical nanoscopy. Nature methods **10**(6), 557–562 (2013)
15. Pan, M., Gan, Y., Zhou, F., Liu, J., Zhang, Y., Wang, A., Zhang, S., Li, D.: Diffuseir: Diffusion models for isotropic reconstruction of 3d microscopic images. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 323–332. Springer (2023)
16. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18. pp. 234–241. Springer (2015)
17. Takemura, S.y., Xu, C.S., Lu, Z., Rivlin, P.K., Parag, T., Olbris, D.J., Plaza, S., Zhao, T., Katz, W.T., Umayam, L., et al.: Synaptic circuits and their variations within different columns in the visual system of drosophila. Proceedings of the National Academy of Sciences **112**(44), 13711–13716 (2015)
18. Wang, Z., Sun, G., Li, G., Shen, L., Zhang, L., Han, H.: Stdin: Spatio-temporal distilled interpolation for electron microscope images. Neurocomputing **505**, 188–202 (2022)

19. Weigert, M., Schmidt, U., Boothe, T., Müller, A., Dibrov, A., Jain, A., Wilhelm, B., Schmidt, D., Broaddus, C., Culley, S., et al.: Content-aware image restoration: pushing the limits of fluorescence microscopy. Nature methods **15**(12), 1090–1097 (2018)
20. Zhang, Y., Guo, J., Zhai, H., Liu, J., Han, H.: Segneuron: 3d neuron instance segmentation in any em volume with a generalist model. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 589–600. Springer (2024), https://github.com/yanchaoz/SegNeuron
21. Zhou, D., Fang, J., Song, X., Guan, C., Yin, J., Dai, Y., Yang, R.: Iou loss for 2d/3d object detection. In: 2019 international conference on 3D vision (3DV). pp. 85–94. IEEE (2019)