MICCAI

A Boundary-aware Cold-Diffusion Model for Electron Microscopy Segmentation

Muge Qi, Ruohua Shi, Yu Cai, Liuyuan He, Wenyao Wang ✉, and Lei Ma ✉

National Biomedical Imaging Center, Peking University, Beijing, China
`1361519095@qq.com`, `{shiruohua, 2406799035, liyhe, wangwenyao, lei.ma}@pku.edu.cn`

Abstract. The advancement of electron microscopy (EM) imaging technology has expanded its applications in life science research, making the automation of EM image analysis a key focus in biomedical imaging. As a core task in EM image analysis, semantic segmentation has garnered significant attention, and convolutional neural networks (CNNs) have been extensively studied, currently emerging as the mainstream method. However, existing methods still face several unresolved challenges. One issue arises from the convolution process, which makes it difficult to efficiently balance global and local information, thus limiting further improvements in segmentation accuracy. Another issue stems from the nature of CNNs, which aim to establish an optimal mapping between images and labels, achieving high accuracy in in-domain data segmentation but at the cost of a noticeable performance drop on out-of-domain data. In this paper, we explore the potential of diffusion probabilistic models (DPMs), known for their exceptional image modeling capabilities, to address these challenges. Specifically, we introduce a diffusion probabilistic model for the semantic segmentation of EM images, which we call EM-Cold-SegDiffusion (ECSD). We adopt a cold or deterministic diffusion framework to achieve higher inference efficiency and a more deterministic segmentation process. Additionally, by introducing an edge-sensitive loss function, we significantly enhance both training efficiency and model performance. Experimental results on common EM segmentation tasks demonstrate that ECSD outperforms mainstream models, offering a promising and superior solution for EM segmentation.

Keywords: Electron microscopy · Semantic segmentation · Cold diffusion

1 Introduction

By capturing biological ultrastructural details, electron microscopy (EM) has provided unprecedented insights into life science, making it now one of the most critical tools for understanding biological complexity[17]. To uncover underlying biological phenomena through quantitative EM image analysis, precise segmentation is a prerequisite(Figure 1). In recent years, deep learning methods have emerged as the most promising approach for EM segmentation[3]. However, two

key issues that have persisted in EM segmentation tasks remain unresolved. First, models lack effective integration of local and global information, limiting further improvements in segmentation accuracy. Second, models are highly dependent on training data, making it difficult to generalize to data and tasks outside of the training conditions.

Convolutional neural networks (CNNs) achieved great success in the ISBI 2012 neuron segmentation challenge and have remained a primary focus in the field of EM segmentation[2]. However, CNNs are limited by the design of their convolutional kernels, which can only focus on context within a certain range around the pixels, thus constraining the model's spatial receptive field. This limitation results in insufficient perception of long-range and global information, creating a bottleneck in segmentation accuracy. UNet revolutionized the field by proposing an encoder-decoder architecture with skip connections, successfully integrating image features at different scales, which significantly improved segmentation accuracy[20]. The DeepLab series adopted dilated convolutions to expand the receptive field without increasing too much computational cost, also achieving notable results[8]. A network proposed by Lee et al. incorporated an auxiliary task of predicting long-range affinities during pretraining, allowing the model to capture a broader context, and for the first time, surpassing human annotators in the accuracy of EM neuron segmentation[14]. The segmentation performance improvements of the mentioned models largely stem from their ability to perceive a broader context within the image. However, due to the inherent limitations of convolution operations, CNNs struggle to efficiently perceive and integrate global contextual information.

More than pursuing segmentation accuracy, improving the generalization ability of models has become an increasingly urgent demand in EM research[9]. While expanding the training data scale is the most fundamental solution to enhancing generalization, it is often not feasible due to practical constraints. Because EM annotation is a well-known labour intensive work. To address the challenge of performance degradation on out-of-domain (OOD) data, several approaches have incorporated domain adaptation techniques[19, 4]. Additionally, models trained with self-supervised learning can leverage transfer learning, allowing them to train with fewer labeled data, thereby alleviating the burden of extensive data annotation. For example, MitoNet combines self-supervised learning with large-scale supervised training to achieve general mitochondrial segmentation[10]. Recently, emerging foundational models such as Segment anything model (SAM)[13] have demonstrated general image segmentation capabilities and have been adapted for domain-specific tasks like electron microscopy segmentation[15, 1, 26, 21]. However, these cutting-edge approaches primarily function as interactive annotation tools, and their automatic segmentation capabilities remain limited—often even falling short of specialized segmentation models. This highlights the need to reconsider the inherent limitations of CNN-based or transformer-based regression models in segmentation tasks, as they rely on regression to map images to features and ultimately to labels within the training data (Figure 1 Top). This reliance fundamentally limits their performance

on out-of-domain data, as they struggle to generalize to unseen images, particularly when these images belong to distributions that differ significantly from those seen during training. To date, no effective solution has been found to fully address this challenge, leaving a critical gap in the field that hinders progress in EM image segmentation.
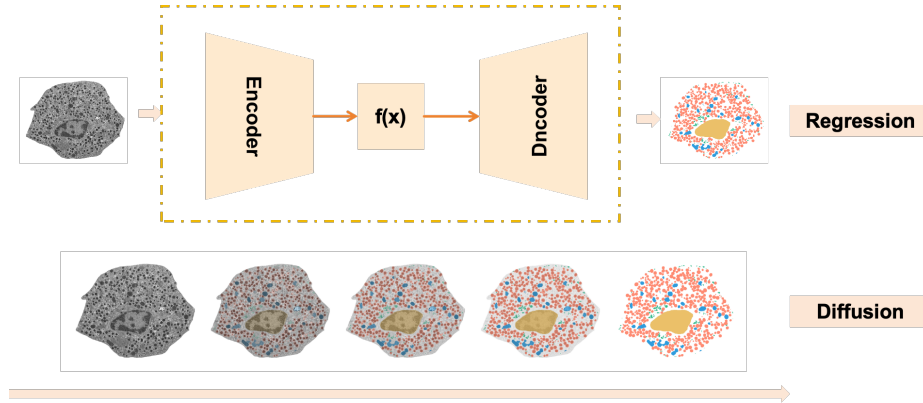

Fig. 1. Image segmentation process. Top: Regression-based models is a "one-step" process, which maps raw images to segmentation masks. Bottom: Diffusion-based models is a iterative process, sculpturing images to masks.

In this paper, we attempt to approach the EM segmentation problem from the perspective of diffusion probabilistic models (DPMs), which is well know for its excellent generation ability and is now one of the hottest topics in artificial intelligence[12]. Our decision is based on the following considerations. During the denoising and diffusion process, DPM is able to simultaneously capture both the global structure and pixel-level details of the image data. Additionally, its iterative nature offers stable training schemes and stronger robustness, underscoring its potential for better generalization. Recently, numerous DPM-based segmentation models have demonstrated excellent performance in biomedical image segmentation[18, 22, 23]. Given the deterministic characteristics of EM segmentation and the computational costs involved, we developed our EM-Cold-SegDiffusion (ECSD) based on previous model, cold-seg diffusion model[24], and explored its application in EM segmentation (Figure 1 Bottom).

Experiments show that directly applying the cold-seg diffusion model to EM segmentation leads to suboptimal results. The main challenge lies in the complexity of unlabeled EM images, with significant variation in organelles' morphology. To address this, we introduced time-dependent boundary regularization to help the model focus on contours[5]. This modification significantly improved performance in mitochondrial and multi-class segmentation tasks, enhancing accuracy, generalization, and convergence speed. We believe diffusion models have great potential to become a major focus in EM segmentation and beyond.

2   Method

2.1   Overview of EM-Cold-Seg-Diffusion

Our method is built upon the Cold Seg Diffusion framework [24], which consists of two main processes: diffusion and denoising (Figure 2).
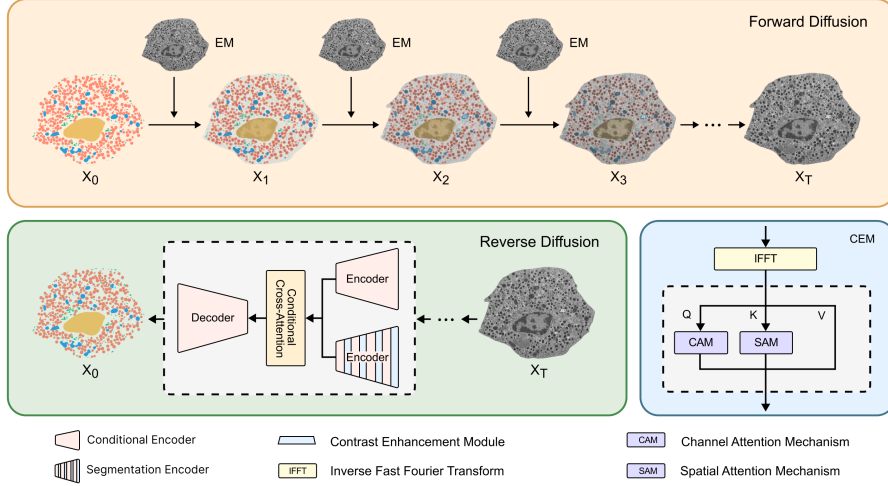

Fig. 2. Overview of EM-Cold-Seg-Diffusion. Forward diffusion process: EM segmentation masks are progressively degraded to corresponding images by gradually adding raw images. Reverse denoising process: The raw EM images are iteratively denoised to generate the final segmentation masks.

Diffusion Process: EM mask images are progressively degraded to EM original images, defined as a Markov chain (Figure 2 Top):

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

Where x_0 represents the mask image (input label), while x_t denotes the degraded image at time step t. The final degraded image x_T is the EM image at the end of the diffusion process. The predefined degradation rate β_t controls the amount of noise added at each step, and z is the deterministic degradation operator (non-Gaussian noise) used to degrade the image.

Denoising Process: The original segmentation mask is progressively restored using an improved ResUNet network R_θ (Figure 2 Bottom):

$$p_\theta(x_{0:T-1}|x_T) := \prod_{t=1}^{T} p_\theta(x_{t-1}|x_t)$$

Where $p_\theta(x_{t-1}|x_t)$ is the conditional distribution of the restored image at time step $t-1$, given the degraded image x_t. The improved ResUNet network R_θ parameterized by θ is used to perform the denoising process. The framework's core components are inherited directly from [24],and we briefly introduce their functions here. The Contrast Enhancement Module (CEM) enhances encoder features in the frequency domain by applying learnable filters for frequency weighting. The enhanced features are then reconstructed using the Inverse Fast Fourier Transform (IFFT) and further refined by integrating the Channel Attention Mechanism (CAM) and Spatial Attention Mechanism (SAM) to highlight critical features. Conditional Cross-Attention Module (CCAM) in Figure 2 generates queries (Q), keys (K), and values (V) from the segmentation encoder and the conditional encoder, respectively. It utilizes cross-modal attention to guide the network to focus on the target organelle regions in the EM images that need to be segmented.

2.2   Boundary-Aware Cold Diffusion Enhancement

Image edge information is extracted explicitly to guide the distribution of learning weights at different time steps (Figure 3).
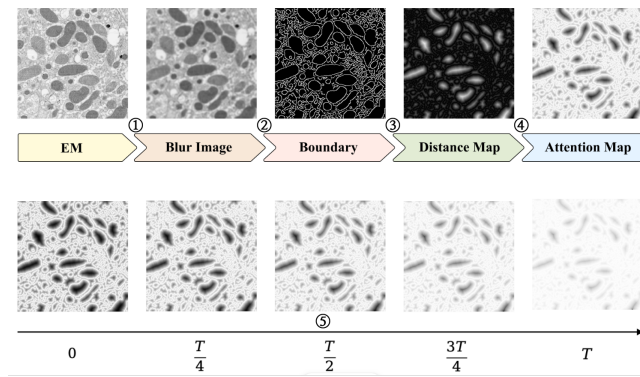

Fig. 3. Boundary-Aware Cold Diffusion Enhancement. Top: The calculation process of the boundary attention map. From left to right: image blurring, Canny edge detection, distance transform, and generation of the edge attention map. Bottom: Time-dependent boundary attention. As the denoising process progresses from early to late stages, more local context is incorporated into the learning process.

Gaussian Blur: Considering the distinct texture and contrast properties of EM images, we employed a Gaussian blur filter to selectively suppress non-membrane structures while enhancing the visibility of organelle membranes during edge detection (Figure 3 - ①).

Canny Edge Detection: The input image is normalized to the range $[0, 255]$, and edges are extracted using adaptive thresholds (Figure 3 - ②):

$$\text{edge} = \text{Canny}(x_{\text{blur}}, \tau_{\text{low}}, \tau_{\text{high}})$$

Canny is a common edge detection method used to locate edge pixels in images[6], x_{blur} denotes the Gaussian-blurred image, τ_{low} and τ_{high} are two hyperparameter thresholds.

Distance Transform and Normalization: Then, the Euclidean distance transformation algorithm[11] is used to calculate the L2 distance of each pixel to the nearest non-zero pixel. The grayscale output shows pixel-wise distances to edges. (Figure 3 - ③):

$$\text{dist}_x = \text{DT}(\text{edge})$$

An inverse distance weight map is then generated (Figure 3 - ④):

$$\text{inv\_dist}_x = \frac{S \cdot 1.1415 - \text{dist}_x}{\max(\text{dist}_x)}$$

Where S denotes the image size, and 1.1415 serves as a correction coefficient for the L_2 norm, which is $\sqrt{2}$ inherited from Dermosegdiff [5].

Time-Adaptive Boundary Attention: A time-decay factor is introduced to dynamically adjust the focus on boundary regions (Figure 3 - ⑤):

$$W_{\text{att}} = \text{inv\_dist}_x^{\left(\gamma \cdot \frac{T-t}{T}\right)^\gamma}$$

where t is the current time step, T is the total number of steps, and $\gamma = 0.8$ controls the shape of the decay curve. Boundary enhancement is disabled for samples with low foreground occupancy (area $< 1\%$).

Composite Loss Function: The composite loss function $\mathcal{L}_{\text{total}}$ is defined as:

$$\mathcal{L}_{\text{total}} = \frac{1}{m} \sum_{i=1}^{m} \underbrace{\|x_{i,0} - f(x_{i,t}, t)\|^2}_{\text{Basic Prediction Loss}} \odot \underbrace{W_{\text{att}}}_{\text{Boundary Attention Weights}}$$

3   Experiments and Results

To validate the effectiveness of our method in EM segmentation, we tested it on common tasks of increasing difficulty: mitochondrial segmentation and multi-class organelle segmentation. Additionally, we conducted an ablation study to highlight the importance of our specific loss function in achieving efficient EM segmentation.

3.1   Datasets and evaluation metrics

For mitochondrial segmentation, we used the Lucchi++ datasets [7]. For multi-class segmentation, we utilized the beta-seg dataset[16] to segment mitochondria, Golgi apparatus, cell nuclei, and insulin secretory granules. In the ablation study, we compared our modified model with the original cold-seg-diffusion

model. Segmentation performance was evaluated using the Jaccard index and Dice coefficient.

3.2   Implementation Details

The Gaussian blur, Canny and DT methods are implemented by the OpenCV library.

For model training, we follow Cold-Seg-Diffusion [24] and set the diffusion steps $T = 50$. Input patches are cropped to 512×512 for Lucchi++ and 256×256 for BetaSeg, both with a batch size of 2. We use the AdamW optimizer with an initial learning rate of 4×10^{-5}, weight decay of 1×10^{-6}, and a cosine warmup schedule.

To enhance the model's generalization ability and mitigate overfitting, several data processing techniques are applied during training. On one hand, boundary-overlapping sampling is utilized to strengthen boundary feature learning, ensuring the model focuses more on critical boundary information. On the other hand, data augmentation techniques such as random flipping, cropping, and rotation are implemented to enrich the diversity of training samples.

The entire method is implemented based on the PyTorch framework and trained on a workstation equipped with 8 NVIDIA Tesla A100 GPUs to ensure efficient computation.

3.3   Mitochondrial segmentation results

GobletNet[25] is a recently published work that achieved state-of-the-art performance across EM segmentation tasks. We compare our model with it and original Cold-Seg Diffusion model on the segmentation of the Lucchi++ dataset. Quantitative results show that our model outperforms GobletNet in Jaccard index and Dice coefficient Table 1. The segmentation results demonstrate that mistakes made by other models are corrected by our model (Figure 4). It worth to be noted that the original Cold-Seg Diffusion model is not as good as our model, even lower than GobletNet. And to achieve table performance, Cold-Seg Diffusion model need 130 training epochs and our model only need 70 epochs.

Table 1. Comparison of Segmentation Performance

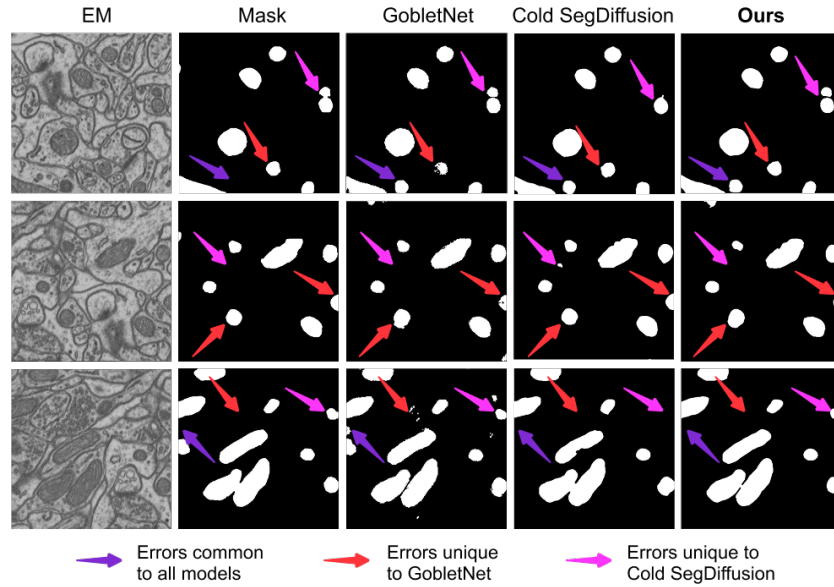| Method	Lucchi++		Betaseg	
	Jaccard	Dice	Jaccard	Dice
SAM[25]	68.23	81.11	41.57	48.99
SwinUNet[25]	–	–	39.16	47.37
GobletNet[25]	81.66	89.91	69.74	80.25
Cold-Seg-Diffusion	79.73	88.72	66.12	76.20
Our method	**87.11**	**93.11**	**73.92**	**83.09**

Fig. 4. Segmentation performance on Lucchi++ dataset. Three cases of segmentation results. From left to right, raw EM images, ground truths, GobletNet, Cold Seg Diffusion and our model. The arrows of different colors in the figure represent distinct types of errors, as indicated in the legend.

3.4   Multi-class organelles segmentation results

Multi-class segmentation is regarded more difficult than individual semantic segmentation. But the results follow similar trend. Our model exhibits better evaluation index and segmentation results (Table 1 and Figure 5). Similarly, the original model exhibit unsatisfactory results, while our model with added boundary regularization significantly improves its performance. Additionally, to achieve the results in Table 1, the original model requires 140 epochs to train, while our model converges to optimal performance in just 80 epochs.

4   Discussion and Conclusion

In this study, we treat electron microscopy (EM) image segmentation as a special type of generative task, and through cold diffusion, we achieve the generation of segmentation results from the raw images. To the best of our knowledge, this is the first attempt to apply a diffusion model to solve the EM segmentation task, and the experimental results demonstrate the advantages of this approach in terms of segmentation performance. However, this work has not yet fully exploited the capabilities of diffusion models. Given the rapid development and flexibility of diffusion models, we believe they have the potential to be extended
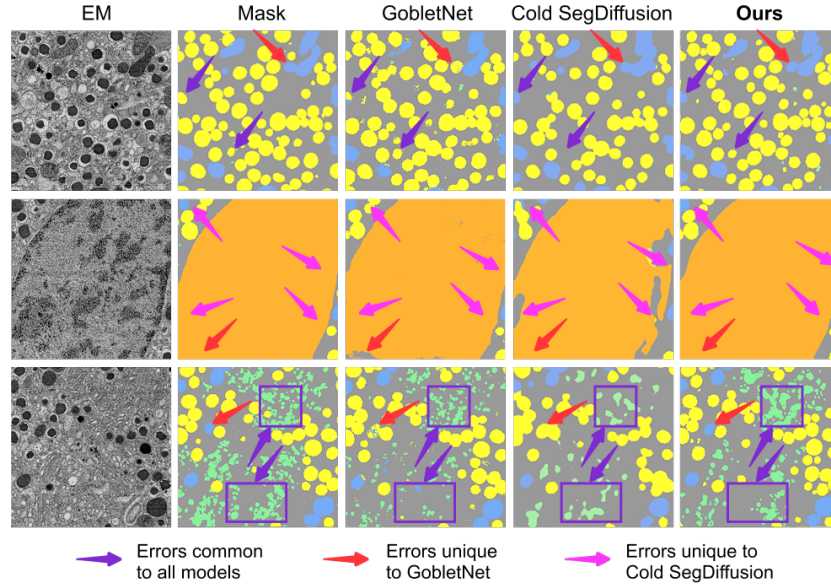
Fig. 5. Segmentation performance on Betaseg dataset. The arrows of different colors in the figure represent distinct types of errors, as indicated in the legend.

to other EM tasks, including recognition, registration, and more, providing a possible foundation for building general models in the EM domain.

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

References

1. Archit, A., Freckmann, L., Nair, S., Khalid, N., Hilt, P., Rajashekar, V., Freitag, M., Teuber, C., Buckley, G., von Haaren, S., et al.: Segment anything for microscopy. Nature Methods pp. 1–13 (2025)
2. Arganda-Carreras, I., Turaga, S.C., Berger, D.R., Cireşan, D., Giusti, A., Gambardella, L.M., Schmidhuber, J., Laptev, D., Dwivedi, S., Buhmann, J.M., et al.: Crowdsourcing the creation of image segmentation algorithms for connectomics. Frontiers in neuroanatomy **9**, 152591 (2015)
3. Aswath, A., Alsahaf, A., Giepmans, B.N., Azzopardi, G.: Segmentation in large-scale cellular electron microscopy with deep learning: A literature survey. Medical image analysis **89**, 102920 (2023)

4. Bermúdez-Chacón, R., Márquez-Neila, P., Salzmann, M., Fua, P.: A domain-adaptive two-stream u-net for electron microscopy image segmentation. In: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018). pp. 400–404. IEEE (2018)

5. Bozorgpour, A., Sadegheih, Y., Kazerouni, A., Azad, R., Merhof, D.: Dermosegdiff: A boundary-aware segmentation diffusion model for skin lesion delineation. In: International workshop on predictive intelligence in medicine. pp. 146–158. Springer (2023)

6. Canny, J.: A computational approach to edge detection. IEEE Transactions on pattern analysis and machine intelligence (6), 679–698 (1986)

7. Casser, V., Kang, K., Pfister, H., Haehn, D.: Fast mitochondria detection for connectomics. In: Medical Imaging with Deep Learning. pp. 111–120. PMLR (2020)

8. Chen, L.C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L.: Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. IEEE transactions on pattern analysis and machine intelligence **40**(4), 834–848 (2017)

9. Conrad, R., Narayan, K.: Cem500k, a large-scale heterogeneous unlabeled cellular electron microscopy image dataset for deep learning. Elife **10**, e65894 (2021)

10. Conrad, R., Narayan, K.: Instance segmentation of mitochondria in electron microscopy images with a generalist deep learning model trained on a diverse dataset. Cell Systems **14**(1), 58–71 (2023)

11. Fabbri, R., Costa, L.D.F., Torelli, J.C., Bruno, O.M.: 2d euclidean distance transform algorithms: A comparative survey. ACM Computing Surveys (CSUR) **40**(1), 1–44 (2008)

12. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. Advances in neural information processing systems **33**, 6840–6851 (2020)

13. Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.Y., et al.: Segment anything. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 4015–4026 (2023)

14. Lee, K., Zung, J., Li, P., Jain, V., Seung, H.S.: Superhuman accuracy on the snemi3d connectomics challenge. arXiv preprint arXiv:1706.00120 (2017)

15. Ma, J., He, Y., Li, F., Han, L., You, C., Wang, B.: Segment anything in medical images. Nature Communications **15**(1), 654 (2024)

16. Müller, A., Schmidt, D., Xu, C.S., Pang, S., D'Costa, J.V., Kretschmar, S., Münster, C., Kurth, T., Jug, F., Weigert, M., et al.: 3d fib-sem reconstruction of microtubule–organelle interaction in whole primary mouse β cells. Journal of Cell Biology **220**(2), e202010039 (2020)

17. Peddie, C.J., Genoud, C., Kreshuk, A., Meechan, K., Micheva, K.D., Narayan, K., Pape, C., Parton, R.G., Schieber, N.L., Schwab, Y., et al.: Volume electron microscopy. Nature Reviews Methods Primers **2**(1), 51 (2022)

18. Rahman, A., Valanarasu, J.M.J., Hacihaliloglu, I., Patel, V.M.: Ambiguous medical image segmentation using diffusion models. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 11536–11546 (2023)

19. Roels, J., Hennies, J., Saeys, Y., Philips, W., Kreshuk, A.: Domain adaptive segmentation in volume electron microscopy imaging. In: 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019). pp. 1519–1522. IEEE (2019)

20. 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)

21. Wan, J., Li, W., Adhinarta, J.K., Banerjee, A., Sjostedt, E., Wu, J., Lichtman, J., Pfister, H., Wei, D.: Trisam: Tri-plane sam for zero-shot cortical blood vessel segmentation in vem images. arXiv preprint arXiv:2401.13961 (2024)
22. Wu, J., Fu, R., Fang, H., Zhang, Y., Yang, Y., Xiong, H., Liu, H., Xu, Y.: Medsegdiff: Medical image segmentation with diffusion probabilistic model. In: Medical Imaging with Deep Learning. pp. 1623–1639. PMLR (2024)
23. Wu, J., Ji, W., Fu, H., Xu, M., Jin, Y., Xu, Y.: Medsegdiff-v2: Diffusion-based medical image segmentation with transformer. In: Proceedings of the AAAI conference on artificial intelligence. vol. 38, pp. 6030–6038 (2024)
24. Yan, P., Li, M., Zhang, J., Li, G., Jiang, Y., Luo, H.: Cold segdiffusion: A novel diffusion model for medical image segmentation. Knowledge-Based Systems **301**, 112350 (2024)
25. Zhou, Y., Li, L., Wang, C., Song, L., Yang, G.: Gobletnet: Wavelet-based high-frequency fusion network for semantic segmentation of electron microscopy images. IEEE Transactions on Medical Imaging (2024)
26. Zhuo, Z., Belevich, I., Leinonen, V., Jokitalo, E., Malm, T., Sierra, A., Tohka, J.: Segment anything for dendrites from electron microscopy. arXiv preprint arXiv:2411.02562 (2024)