ShapeMamba-EM: Fine-Tuning Foundation Model with Local Shape Descriptors and Mamba Blocks for 3D EM Image Segmentation

Ruohua Shi[1,2,3], Qiufan Pang[1], Lei Ma[1,2,3], Lingyu Duan[1,2,4], Tiejun Huang[1,2,3], and Tingting Jiang[1,2,3]✉

[1] National Engineering Research Center of Visual Technology, School of Computer Science, Peking University, Beijing, China
[2] State Key Laboratory of Multimedia Information Processing, School of Computer Science, Peking University, Beijing, China
[3] National Biomedical Imaging Center, Peking University, Beijing, China
[4] Peng Cheng Laboratory, Shenzhen, China
{shiruohua@,pqf@stu.,lei.ma@,lingyu@,tjhuang@,ttjiang@}pku.edu.cn

Abstract. Electron microscopy (EM) imaging offers unparalleled resolution for analyzing neural tissues, crucial for uncovering the intricacies of synaptic connections and neural processes fundamental to understanding behavioral mechanisms. Recently, the foundation models have demonstrated impressive performance across numerous natural and medical image segmentation tasks. However, applying these foundation models to EM segmentation faces significant challenges due to domain disparities. This paper presents *ShapeMamba-EM*, a specialized fine-tuning method for 3D EM segmentation, which employs adapters for long-range dependency modeling and an encoder for local shape description within the original foundation model. This approach effectively addresses the unique volumetric and morphological complexities of EM data. Tested over a wide range of EM images, covering five segmentation tasks and 10 datasets, ShapeMamba-EM outperforms existing methods, establishing a new standard in EM image segmentation and enhancing the understanding of neural tissue architecture.

Keywords: 3D EM image segmentation · State space model · Local shape descriptor.

1 Introduction

Electron microscopy (EM) allows the imaging of neural tissue at a resolution sufficient to resolve individual synapses and fine neural processes. Therefore, the segmentation of EM images plays a pivotal role in the realm of biological research, offering profound insights into the inner mechanisms underlying behavior and helping drive future theoretical experiments [22,20,11]. Many excellent algorithms and datasets have emerged [2,12,29,26,25,24].

Recent advancements in computer vision have spurred breakthroughs by the foundation models, such as the Segment anything model (SAM), which has

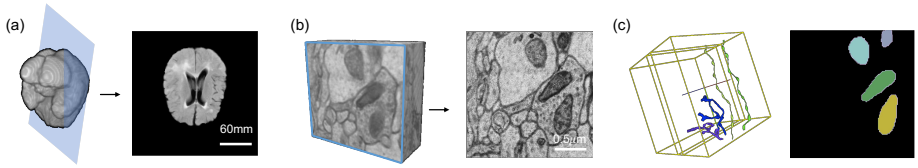

Fig. 1. Illustration of the medical data and EM data. (a) MRI edema image from the BraTS2021 dataset. (b) EM mitochondria data from MitoEM-R dataset. (c) 3D and 2D segmentation results of (b). The boundaries of instances share a similar local shape, and the scope of the instance spans the entire volume.

achieved promising zero-shot segmentation performance on a variety of natural image datasets [14]. However, their performance significantly declines when applied to some downstream tasks, like biological and medical images, primarily due to the substantial disparity between the two image domains.

An effective method to enhance the generalizability of the foundation models to downstream tasks lies in fine-tuning. Recently, numerous methods applied to medical datasets have demonstrated impressive efficacy [4,28,31,32]. For instance, Med-SA [30] maintains the pre-trained SAM parameters frozen while integrating LoRA modules to the designated positions. MedSAM [18] has achieved this by fine-tuning the decoder with 1.1 million masks, enabling SAM's application in medical imaging. In addition to these methods trained on 2D medical images, some methods propose solutions for 3D images. 3DSAM-adapter [6] proposes a holistically designed scheme for transferring SAM from 2D to 3D for promptable medical image segmentation. SAM-Med3D [28] reformulates SAM to a thorough 3D architecture trained on a comprehensively processed large-scale volumetric medical dataset.

However, the medical foundation models above cannot be directly applied to EM data. Taking Fig. 1 as an example, although both medical and EM images are all volumetric grayscale images, EM images have significantly higher resolution (about 10^5 times), which results in more noise. Besides, the objects to be segmented in EM images have relatively consistent local features and are widely distributed across the spatial domain. This distribution, combined with the inherent anisotropy, intensifies the challenges of EM segmentation. Consequently, EM images require customized methods to address these challenges. On the other hand, medical images have grayscale and volumetric data, which are closer to EM images than to natural images, so the fine-tuned medical SAM models are more suitable for EM data than SAM. Therefore, this paper proposes a fine-tuning method specifically designed for EM image segmentation, based on a 3D medical foundation model, named **ShapeMamba-EM**. ShapeMamba-EM first modifies the original foundation model to enhance the tuning efficiency, and then adds two novel modules targeting the local morphological features and long-range dependencies in EM data as shown in Fig. 1 (b,c).

Specifically, for the selection of the foundation model, we opted for the currently largest model trained on 3D medical images, SAM-Med3D [28], which

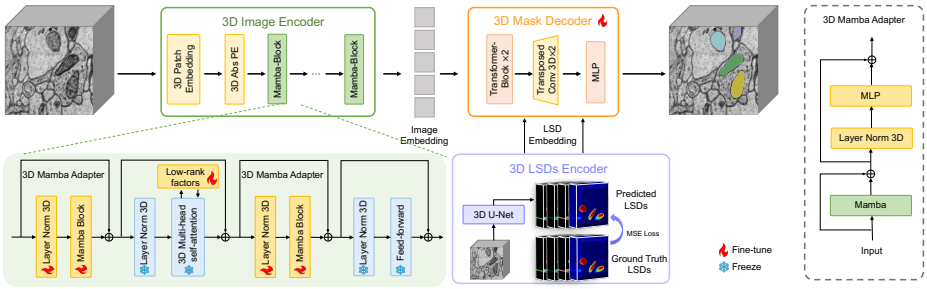

Fig. 2. The overall architecture of ShapeMamba-EM. The image encoder is updated with FacT. The volumetric or temporal information is effectively incorporated via a set of 3D Mamba adapters. The mask decoder is fully fine-tuned and modified to recover the prediction resolution. The LSDs are trained by the 3D U-Net network.

consists of three parts: image encoder, mask decoder, and prompt encoder. Inspired by existing work on medical data, we leverage FacT [13] to tune the image encoder module for retaining most pre-trained weights while only updating lightweight weight increments. Furthermore, the prompt encoder is removed because crafting appropriate prompts is a challenging task for EM data and automatic segmentation has shown promise.

Besides modifying the original model, we tackle the challenges of *local morphological features* and *long-range dependencies* for EM data with two novel modules. Firstly, we find pixel-wise prediction alone insufficient. To accurately segment objects with similar *local morphological features* and imperfect edges, we implement a 3D U-Net architecture [5] to predict the Local Shape Descriptors (LSDs) [23] to enhance the boundary prediction. Secondly, to address the challenge of *long-range dependency* inherent in EM object analysis, we draw inspiration from the recent innovation in Mamba [7] which utilizes state space sequential models (SSMs) [8,19]. These models excel at extracting long-dependency information with reduced computational burden and lower memory consumption. Consequently, we propose the integration of *3D Mamba Adapters* into the image encoder. Through extensive experimentation, we demonstrate the superior performance of the ShapeMamba-EM framework across a broad spectrum of EM images, spanning five segmentation tasks and 10 datasets.

2 Method

2.1 Overview

The overall framework of our proposed ShapeMamba-EM is illustrated in Fig. 2. We introduce an innovative model that enhances the SAM-Med3D core framework for 3D EM segmentation tasks. Specifically, we have augmented the SAM-Med3D architecture by incorporating the FacT approach into the 3D Multi-head

self-attention model. This integration enables a more effective and efficient fine-tuning process. Furthermore, we introduce *3D Mamba Adapters* designed to tackle the challenges of long-range dependency of segmentation objects. Besides, a 3D U-Net network is incorporated into the *3D Mask Decoder* to capture the Local Shape Descriptors of the segmentation objects. These two designs address the inherent limitations of SAM-Med3D in EM segmentation, thereby increasing segmentation accuracy and efficiency.

2.2 SAM-Med3D

In this paper, we adopt SAM-Med3D [28] as our foundation model, due to its excellent ability to extract volumetric data features by training on a large amount of 3D medical data. It is designed based on the segment anything model (SAM) [14] for volumetric medical imaging by incorporating a 3D structure for direct spatial information capture. SAM-Med3D employs a 3D convolution for patch embedding, uses a 3D positional encoding extended from SAM's 2D version, and inputs these to 3D attention blocks. These blocks, enhanced with 3D relative positional encodings, are part of the Multi-Head Self-Attention module, facilitating direct spatial detail capture. The prompt encoder uses 3D position encodings for sparse prompts and 3D convolutions for dense prompts, while the 3D Mask Decoder integrates 3D upscaling with 3D transposed convolution.

2.3 Parameter-efficient fine-tuning of 3D image encoder

In order to effectively extract image features, Med-SAM3D's 3D Image Encoder comprises a substantial portion of network parameters. Fine-tuning all these weights is computationally intensive. Previous research has shown that PETL techniques can achieve adaptation performance similar to full fine-tuning but with significantly fewer network parameters updated [21,13,27]. In this regard, we adopt FacT [13], a SOTA parameter efficient transfer learning (PETL) technique, that can obtain comparable or superior performance compared to other PETL methods while introducing a smaller number of trainable parameters.

2.4 3D Mamba Adapter

In 3D EM images, the objects are widely distributed yet dense. To address this long-range dependencies challenge in EM images, we concentrate on enhancing the representational capacity of the encoder in neural network architectures through the incorporation of the Mamba layer, which is designed to capture long-range dependencies within the input data. Specifically, we design the *3D Mamba Adapter* as illustrated in Fig. 2 (right), which substitutes the self-attention module in the transformer architecture with the more efficient Mamba layer. Each 3D Mamba Adapter consists of a *3D Layer Norm* operation with a *Mamba Block* [7]. This enables both multi-scale and global feature modeling while maintaining a high efficiency during training and inference.

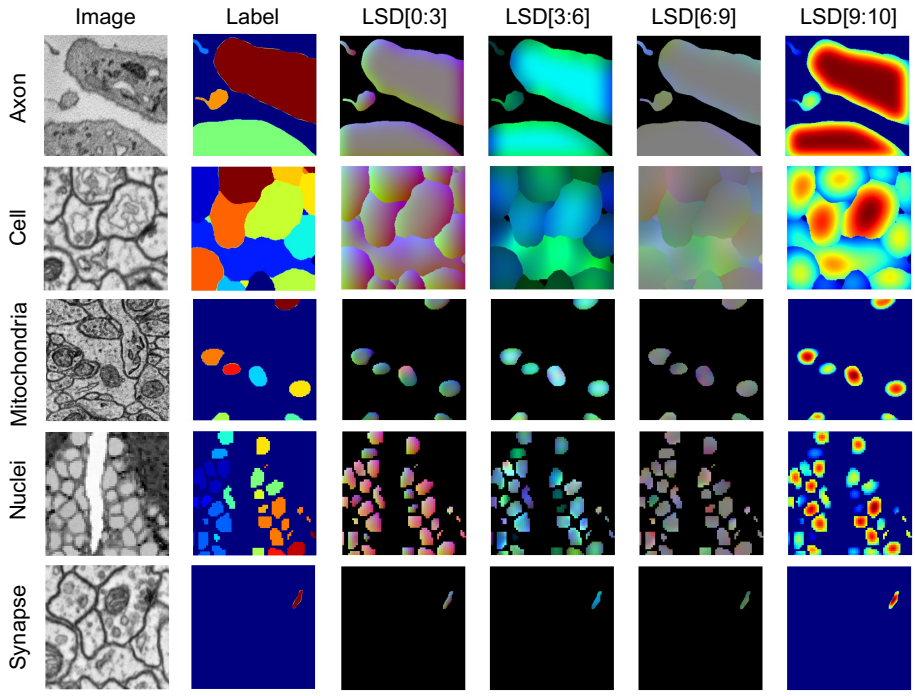

Fig. 3. Visualizations of LSDs for different segmentation tasks. From left to right: the EM image, segmentation labels, and components of the LSDs.

2.5 3D Local Shape Descriptor Encoder

Although recent deep learning networks have made good progress in EM image segmentation, a neural network might focus only on a few center voxels to detect objects and achieve high accuracy during training, especially if trained using a voxel-wise loss. However, this strategy might fail in rare cases where boundary evidence is ambiguous. The objects to be segmented in EM images, such as axons, cells, and mitochondria, have regular local shapes. Inspired by [23], to enhance the model's learning of the local shapes objects in EM images, We first define the Local Shape Descriptors (LSDs) of a segment $i \in \{1, \ldots, l\}$ under a given voxel v. we intersect the segment $\mathrm{y}(v)$ underlying a voxel $v \in \Omega$ with a 3D ball of radius σ centered at v to obtain a subset of voxels $S_v \subset \Omega$, formally given as:

$$S_v = \left\{ v' \in \Omega \,|\mathrm{y}(v) = \mathrm{y}\left(\nu'\right), |\, v - v'|_2^2 \leq \sigma \right\}. \tag{1}$$

The LSD $\mathrm{lsd}^{\mathrm{y}} : \Omega \mapsto \mathbb{R}^{10}$ for a voxel v is a concatenation of the size, center offset and coordinate covariance, that is:

$$\mathrm{lsd}^{\mathrm{y}}(\boldsymbol{v}) = (\underbrace{\mathrm{s}\left(S_v\right)}_{\text{size}}, \underbrace{\mathrm{m}\left(S_v\right) - v}_{\text{center offset}}, \underbrace{\mathrm{c}\left(S_v\right)}_{\text{covariance}}). \tag{2}$$

Task	**Axon**		**Cell**						**Synapse**			
Dataset	Gauy		ISBI2012		SNEMI3D		CREMI		CREMI		EM-R50	
Metrics	Dice	mAP	Dice	mAP	Dice	mAP	Dice	mAP	Dice	mAP	Dice	mAP
U3D-BCD	0.790	0.801	0.932	0.930	0.964	0.961	0.952	0.960	0.814	0.823	0.764	0.775
SwinU	0.722	0.746	0.932	0.935	0.962	0.958	0.943	0.950	0.821	<u>0.829</u>	0.733	0.749
nnU-Net	0.788	0.789	<u>0.971</u>	<u>0.966</u>	<u>0.965</u>	<u>0.962</u>	0.947	0.951	<u>0.828</u>	0.814	0.760	0.758
MA-SAM	0.539	0.621	0.763	0.742	0.874	0.878	0.836	0.845	0.475	0.464	0.527	0.539
3DSAMA	0.474	0.599	0.518	0.569	0.371	0.428	0.531	0.594	0.01	0.01	0.01	0.01
MSAM3D	0.716	0.743	0.858	0.862	0.931	0.927	0.943	0.948	0.611	0.591	0.711	0.718
w/o M	0.735	0.750	0.947	0.951	0.962	0.965	<u>0.954</u>	<u>0.966</u>	0.815	0.823	0.779	0.780
w/o L	<u>0.791</u>	<u>0.796</u>	0.939	0.942	0.959	0.958	0.942	0.951	0.799	0.810	<u>0.780</u>	<u>0.801</u>
Ours	**0.809**	**0.827**	**0.958**	**0.951**	**0.974**	**0.972**	**0.965**	**0.973**	**0.834**	**0.865**	**0.792**	**0.817**

Table 1. Quantitative outcomes of methods applied to segmentation tasks for axons, synapses, and cells. Bold and underlined numbers denote the 1st and 2nd scores.

where $\mathrm{s}(S_v) = |S_v|$ is the size of S_v, $\mathrm{m}(S_v)$ is the covariance of its coordinates and $\mathrm{c}(S_v)$ is the mean coordinates:

$$\mathrm{m}(S_v) = \frac{1}{\mathrm{s}(S_v)} \sum_{v \in S_v} v, \quad \mathrm{c}(S_v) = \frac{1}{\mathrm{s}(S_v)} \sum_{v \in S_v} (v - \mathrm{m}(S_v)) (v - \mathrm{m}(S_v))^{\mathrm{T}}. \quad (3)$$

Therefore, for a 3D image, LSDs are represented as a ten-dimensional embedding. This encapsulation includes: LSD[0:3] for the mean offset; LSD[3:6] and LSD[6:9] delineating the covariance of coordinates, with LSD[3:6] capturing the diagonal entries and LSD[6:9] the off-diagonals; and finally, LSD[9:10] reflecting the size, quantified as the number of voxels within the intersected Gaussian. Some visualization examples of LSDs are shown in Fig. 3. We use $\mathrm{lsd}^y(v)$ to formulate an auxiliary learning task that complements the prediction of affinities. For that, we use the 3D U-Net network to learn the $\mathrm{lsd}^{\mathrm{x}} : \Omega \mapsto \mathbb{R}^{10}$ directly from raw data x, and take it as the embedding to the 3D Emage Decoder network.

3 Experiments and Results

We extensively evaluate our method on five EM image segmentation tasks, covering 10 datasets, i.e., axon segmentation, cell segmentation, mitochondria segmentation, synapse segmentation, and nuclei segmentation. We first conduct comparisons with state-of-the-art EM image segmentation methods and SAM fine-tuning methods, and then provide ablation studies to analyze our method.

3.1 Datasets and Experimental Settings

We conduct a series of extensive experiments on 5 segmentation tasks with 10 datasets to evaluate the performance of our method. Here we briefly introduce

Task	**Mitochondria**										**Nuclei**	
Dataset	Gauy		Kasthuri++		Lucchi++		MitoEM-H		MitoEM-R		NucMM-Z	
Metrics	Dice	mAP	Dice	mAP	Dice	mAP	Dice	mAP	Dice	mAP	Dice	mAP
U3D-BCD	0.564	0.529	0.889	0.831	0.880	0.753	0.746	0.773	0.775	0.844	0.879	0.894
SwinU	0.472	0.443	0.904	0.861	0.869	0.874	0.779	0.822	0.803	<u>0.867</u>	0.866	0.837
nnU-Net	0.528	0.501	0.859	0.872	0.856	0.829	0.807	0.830	0.825	0.864	0.907	0.894
MA-SAM	0.349	0.315	0.769	0.773	0.754	0.740	0.692	0.701	0.723	0.744	0.839	0.858
3DSAMA	0.145	0.188	0.582	0.668	0.645	0.682	0.671	0.710	0.786	0.802	0.863	0.897
MSAM3D	0.537	0.521	0.902	0.878	0.715	0.728	0.711	0.714	0.822	0.835	0.878	0.895
w/o M	0.572	<u>0.598</u>	0.951	0.920	0.906	<u>0.910</u>	0.817	<u>0.825</u>	<u>0.846</u>	0.852	<u>0.910</u>	0.903
w/o L	<u>0.586</u>	0.591	0.942	<u>0.922</u>	<u>0.915</u>	0.908	<u>0.820</u>	0.812	0.839	0.844	0.904	<u>0.899</u>
Ours	**0.612**	**0.603**	**0.968**	**0.936**	**0.940**	**0.954**	**0.847**	**0.877**	**0.852**	**0.930**	**0.915**	**0.907**

Table 2. Quantitative outcomes of methods applied to segmentation tasks for mitochondrion and nucleus. Bold and underlined numbers denote the 1st and 2nd scores.

them, and more details are shown in the supplementary materials. Specifically, the *Axon Segmentation Task* uses the Gauy dataset [9], the *Cell Segmentation Task* uses three datasets: ISBI2012 [2], SNEMI3D [15] CREMI [1], the *Mitochondira Segmentation Task* uses five datasets: MitoEM-R [29], MitoEM-H [29], Kasthuri++ [3], Lucchi++ [3], and Gauy [9], the *Nuclei Segmentation Task* uses NucMM-Z [17] dataset, ans the *Synapse Segmentation Task* uses two datasets: CREMI [1] and EM-R50 [16]. For the evaluation, we evaluate the methods using mean 3D Average Precision (mAP) [29] and Dice scores at the instance level.

We comprehensively compare our proposed method against a suite of cutting-edge algorithms. These include recent successful approaches for biomedical data segmentation utilizing CNN architectures, such as U3D-BCD [29] and nnU-Net [12], alongside Transformer-based architecture, SwinUNETR [10] (SwinU). Additionally, we examine fine-tuning methods based on SAM, specifically MA-SAM [4] and 3DSAM-Adapter [6] (3DSAMA). Our evaluation also extends to direct fine-tuning using Med-SAM3D (MSAM3D), supplemented by an ablation study to assess the impact of our novel 3D Mamba Adapter and 3D LSD Encoder Module. During the training of fine-tuning methods, we independently trained the model for each task and employed Binary Cross-Entropy (BCE) loss during training. The experiments are conducted training on 8 NVIDIA A800 GPUs. More details of the experiments including the splits of the training and testing data, σ in LSD generation are shown in the supplementary materials.

3.2 Quantitative and qualitative segmentation results

Qualitative results. The quantitative results in Tab. 1 and Tab. 2 underscore the superior performance of our proposed method in precise EM segmentation across all five tasks. Compared to both CNN-based, Transformer-based, and fine-tuning based methods, the proposed ShapeMamba-EM demonstrates competitive even higher performance. We showcase several predictions in Fig. 4.

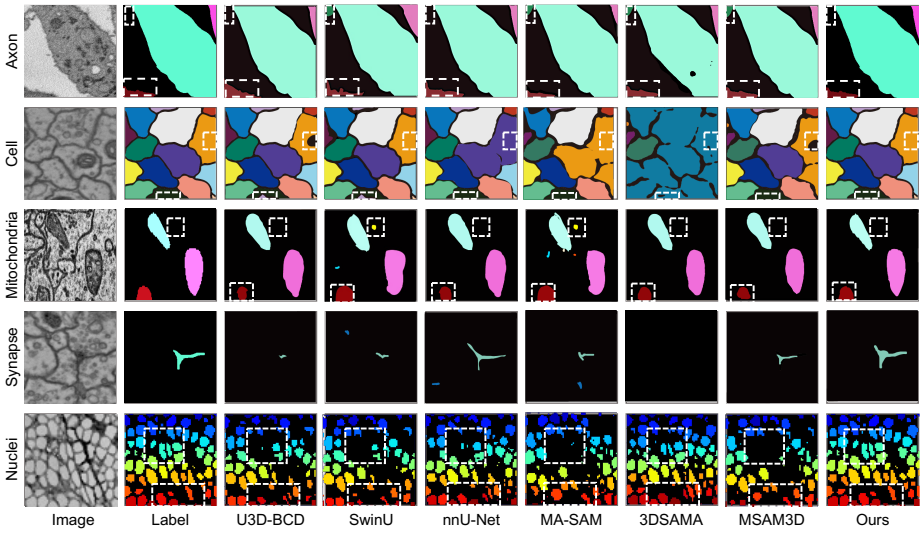

Fig. 4. The visualization of segmentation results.

More comparisons of other SOTA methods for each dataset are provided in the supplementary material.

Comparisions of fine-tuning methods. In the comparative analysis of fine-tuning algorithms, it indicates that those based on Med-SAM3D reflects significant improvement over those based solely on SAM (such as MA-SAM and 3DSAM-Adapter). Such progress underscores the effectiveness of leveraging medical data in refining the model's ability to generalize to EM data. In Fig. 4, we observe that SAM-based methods demonstrate poor performance in segmenting small objects and dense cells, which strongly requires volumetric spatial information. Particularly for the 3D SAM-Adapter method, it seldom predicts the synapse. This limitation might stem from its reliance on the quality of prompts during training. Considering that the number of positive prompts in synapse datasets is substantially lower than that of negative prompts, the model faces challenges in learning useful information effectively. Furthermore, its performance in cell segmentation is hindered by the static number of prompts.

Ablation study for 3D Mamba Adapter and 3D LSD Encoder. We further evaluate the effectiveness of the 3D Mamba Adapter (w/o M) and the 3D LSD Encoder (w/o L). Results indicate an average increase of 5% with the use of the 3D Mamba Adapter and 8% with the 3D LSD Encoder. Additionally, we observe that the impact of the 3D Mamba Adapter on the nuclei segmentation task is relatively minor, potentially attributable to the smaller size of the images and objects to be segmented. Moreover, compared to methods that utilize multi-head self-attention modules, such as MSAM3D and 3DSAMA, the experiments confirm Mamba's superior performance. The segmentation results of the models with and without 3D Mamba Adapter are shown in the supplementary material.

Limitations. This paper focuses on fine-tuning foundation models for EM images. Due to space constraints, we did not compare with other Mamba-based methods as they are 2D models and do not support 3D segmentation. Additionally, we used the 3D U-Net for LSD estimation because the original method uses it. We plan to explore more advanced models in future work.

4 Discusion and Conclusion

In summary, ShapeMamba-EM offers an innovative fine-tuning method for EM segmentation, leveraging a 3D medical foundation model to address unique challenges of EM data such as high resolution and complex tissues. It surpasses traditional models by integrating 3D Mamba Adapters and Local Shape Descriptors Encoder, improving accuracy and efficiency. Extensive tests on diverse EM datasets highlight its effectiveness in high-resolution image segmentation, setting new benchmarks. This work aims to enhance EM segmentation and showcase the adaptability of medical foundation models for in-depth biological studies.

Acknowledgments. This work was supported by the Natural Science Foundation of China under contract 62088102, the Beijing Natural Science Foundation (Grant No. JQ24023), and the Beijing Municipal Science & Technology Commission Project (No.Z231100006623010). We also acknowledge the Biomedical Computing Platform of National Biomedical Imaging Center and High-Performance Computing Platform of Peking University for providing computational resources.

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

References

1. CREMI Challenge, `https://cremi.org/`
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., Liu, T., Seyedhosseini, M., Tasdizen, T., Kamentsky, L., Burget, R., Uher, V., Tan, X., Sun, C., Pham, T.D., Bas, E., Uzunbas, M.G., Cardona, A., Schindelin, J., Seung, H.S.: Crowdsourcing the creation of image segmentation algorithms for connectomics. Frontiers in Neuroanatomy **9**, 142 (2015)
3. Casser, V., Kang, K., Pfister, H., Haehn, D.: Fast mitochondria detection for connectomics. In: Proceedings of the Third Conference on Medical Imaging with Deep Learning. vol. 121, pp. 111–120 (06–08 Jul 2020)
4. Chen, C., Miao, J., Wu, D., Yan, Z., Kim, S., Hu, J., Zhong, A., Liu, Z., Sun, L., Li, X., et al.: MA-SAM: Modality-agnostic SAM adaptation for 3D medical image segmentation. arXiv preprint arXiv:2309.08842 (2023)
5. Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O.: 3D U-Net: learning dense volumetric segmentation from sparse annotation. In: Medical Image Computing and Computer-Assisted Interventino, 21, 2016, Proceedings, Part II 19. pp. 424–432 (2016)

6. Gong, S., Zhong, Y., Ma, W., Li, J., Wang, Z., Zhang, J., Heng, P.A., Dou, Q.: 3DSAM-adapter: Holistic adaptation of SAM from 2D to 3D for promptable medical image segmentation. arXiv preprint arXiv:2306.13465 (2023)
7. Gu, A., Dao, T.: Mamba: Linear-time sequence modeling with selective state spaces. arXiv preprint arXiv:2312.00752 (2023)
8. Gu, A., Goel, K., Re, C.: Efficiently modeling long sequences with structured state spaces. In: International Conference on Learning Representations (2022)
9. Guay, M.D., Emam, Z.A., Anderson, A.B., Aronova, M.A., Pokrovskaya, I.D., Storrie, B., Leapman, R.D.: Dense cellular segmentation for EM using 2D–3D neural network ensembles. Scientific Reports **11**(1), 2561 (2021)
10. Hatamizadeh, A., Nath, V., Tang, Y., Yang, D., Roth, H.R., Xu, D.: Swin UNETR: Swin transformers for semantic segmentation of brain tumors in MRI images. In: International MICCAI brainlesion workshop. pp. 272–284 (2021)
11. Hulse, B.K., Haberkern, H., Franconville, R., Turner-Evans, D., Takemura, S.y., Wolff, T., Noorman, M., Dreher, M., Dan, C., Parekh, R., et al.: A connectome of the drosophila central complex reveals network motifs suitable for flexible navigation and context-dependent action selection. Elife **10** (2021)
12. Isensee, F., Jaeger, P.F., Kohl, S.A., Petersen, J., Maier-Hein, K.H.: nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nature Methods **18**(2), 203–211 (2021)
13. Jie, S., Deng, Z.H.: Fact: Factor-tuning for lightweight adaptation on vision transformer. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 37, pp. 1060–1068 (2023)
14. Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.Y., Dollar, P., Girshick, R.: Segment anything. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 4015–4026 (October 2023)
15. Lee, K., Zung, J., Li, P., Jain, V., Seung, H.S.: Superhuman accuracy on the SNEMI3D connectomics challenge. arXiv preprint arXiv:1706.00120 (2017)
16. Lin, Z., Wei, D., Jang, W.D., Zhou, S., Chen, X., Wang, X., Schalek, R., Berger, D., Matejek, B., Kamentsky, L., et al.: Two stream active query suggestion for active learning in connectomics. In: Proceedings of European Conference on Computer Vision. pp. 103–120 (2020)
17. Lin, Z., Wei, D., Petkova, M.D., Wu, Y., Ahmed, Z., Zou, S., Wendt, N., Boulanger-Weill, J., Wang, X., Dhanyasi, N., et al.: NucMM dataset: 3D neuronal nuclei instance segmentation at sub-cubic millimeter scale. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 164–174 (2021)
18. Ma, J., He, Y., Li, F., Han, L., You, C., Wang, B.: Segment anything in medical images. Nature Communications **15**(1), 654 (2024)
19. Ma, J., Li, F., Wang, B.: U-Mamba: Enhancing long-range dependency for biomedical image segmentation. arXiv preprint arXiv:2401.04722 (2024)
20. Motta, A., Berning, M., Boergens, K.M., Staffler, B., Beining, M., Loomba, S., Hennig, P., Wissler, H., Helmstaedter, M.: Dense connectomic reconstruction in layer 4 of the somatosensory cortex. Science **366**(6469), eaay3134 (2019)
21. Qiao, Y., Yu, Z., Wu, Q.: VLN-PETL: Parameter-efficient transfer learning for vision-and-language navigation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 15443–15452 (2023)
22. Schneider-Mizell, C.M., Gerhard, S., Longair, M., Kazimiers, T., Li, F., Zwart, M.F., Champion, A., Midgley, F.M., Fetter, R.D., Saalfeld, S., et al.: Quantitative neuroanatomy for connectomics in drosophila. Elife **5**, e12059 (2016)

23. Sheridan, A., Nguyen, T.M., Deb, D., Lee, W.C.A., Saalfeld, S., Turaga, S.C., Manor, U., Funke, J.: Local shape descriptors for neuron segmentation. Nature Methods **20**(2), 295–303 (2023)
24. Shi, R., Bi, K., Du, K., Ma, L., Fang, F., Duan, L., Jiang, T., Huang, T.: PS-Net: Human perception-guided segmentation network for em cell membrane. Bioinformatics **39**(8), btad464 (2023)
25. Shi, R., Duan, L., Huang, T., Jiang, T.: Evidential uncertainty-guided mitochondria segmentation for 3D EM images. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 38, pp. 4847–4855 (2024)
26. Shi, R., Wang, W., Li, Z., He, L., Sheng, K., Ma, L., Du, K., Jiang, T., Huang, T.: U-RISC: an annotated ultra-high-resolution electron microscopy dataset challenging the existing deep learning algorithms. Frontiers in Computational Neuroscience **16**, 842760 (2022)
27. Tu, C.H., Mai, Z., Chao, W.L.: Visual query tuning: Towards effective usage of intermediate representations for parameter and memory efficient transfer learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 7725–7735 (2023)
28. Wang, H., Guo, S., Ye, J., Deng, Z., Cheng, J., Li, T., Chen, J., Su, Y., Huang, Z., Shen, Y., et al.: SAM-Med3D. arXiv preprint arXiv:2310.15161 (2023)
29. Wei, D., Lin, Z., Franco-Barranco, D., Wendt, N., Liu, X., Yin, W., Huang, X., Gupta, A., Jang, W.D., Wang, X., et al.: MitoEM dataset: Large-scale 3D mitochondria instance segmentation from EM images. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 66–76 (2020)
30. Wu, J., Fu, R., Fang, H., Liu, Y., Wang, Z., Xu, Y., Jin, Y., Arbel, T.: Medical SAM Adapter: Adapting segment anything model for medical image segmentation. arXiv preprint arXiv:2304.12620 (2023)
31. Zhang, Y., Hu, S., Jiang, C., Cheng, Y., Qi, Y.: Segment anything model with uncertainty rectification for auto-prompting medical image segmentation. arXiv preprint arXiv:2311.10529 (2023)
32. Zhang, Y., Shen, Z., Jiao, R.: Segment anything model for medical image segmentation: Current applications and future directions. Computers in Biology and Medicine p. 108238 (2024)