SegNeuron: 3D Neuron Instance Segmentation in Any EM Volume with a Generalist Model

Yanchao Zhang[1,2], Jinyue Guo[1,3], Hao Zhai[1,2], Jing Liu[1], and Hua Han[1,2](✉)

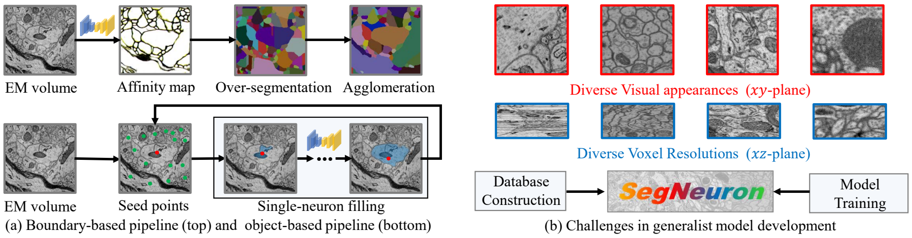
[1] Laboratory of Brain Atlas and Brain-inspired Intelligence, Key Laboratory of Brain Cognition and Brain-inspired Intelligence Technology, Institute of Automation, Chinese Academy of Sciences, Beijing, China
{hua.han}@ia.ac.cn
[2] School of Future Technology, University of Chinese Academy of Sciences, Beijing, China
[3] School of Artificial Intellengace, University of Chinese Academy of Sciences, Beijing, China

Abstract. Building a generalist model for neuron instance segmentation from electron microscopy (EM) volumes holds great potential to accelerate data processing and analysis in connectomics. However, the diversity in visual appearances and voxel resolutions present obstacles to model development. Meanwhile, prompt-based foundation models for segmentation struggle to achieve satisfactory performance due to the inherent complexity and volumetric continuity of neuronal structures. To address this, this paper introduces **SegNeuron**, a generalist model for dense neuron instance segmentation with strong zero-shot generalizability. To this end, we first construct a multi-resolution, multi-modality, and multi-species volume EM database, named **EMNeuron**, consisting of over **22 billion** voxels, with over **3 billion** densely labeled. On this basis, we devise a novel workflow to build the model with customized strategies, including pretraining via multi-scale Gaussian mask reconstruction, domain-mixing finetuning, and foreground-restricted instance segmentation. Experimental results on unseen datasets indicate that SegNeuron not only significantly surpasses existing generalist models, but also achieves competitive or even superior results with specialist models. Datasets, codes, and models are available at https://github.com/yanchaoz/SegNeuron.

Keywords: Connectomics · Neuron Segmentation · Volume Electron Microscopy · Deep Learning · Generalist Model

1 Introduction

Efficient and accurate neuron segmentation from electron microscopy (EM) volumes has become a bottleneck that hinders progress in connectomic analysis [1,27]. Consequently, automatic neuron segmentation methods based on deep neural networks have emerged as a solution, effectively trading computational resources for expert reconstruction time. As shown in Fig. 1(a), existing methods can be categorized as boundary- and object-based. Specifically, boundary-based



(a) Boundary-based pipeline (top) and object-based pipeline (bottom)

(b) Challenges in generalist model development

Fig. 1. Widely adopted pipelines and key challenges in generalist model development.

approaches [21,20,28] train models to predict descriptors for neuron boundaries (*e.g.*, affinity maps) and then employ graph-based agglomeration for instance segmentation [4,11]. In contrast, object-based approaches execute single-neuron filling by extending the trace area from the seed points iteratively [16,26]. While effective, these methods suffer from poor model generalization, which requires repetitive annotation, training, inference, and proofreading for new datasets.

Such cumbersome workflow could be streamlined and expedited with a generalist model that robustly determines the instance-level belonging of each voxel in any EM volume. Nonetheless, the diverse visual appearances and voxel resolutions present in EM data, as illustrated in Fig. 1(b), due to variations in species, tissues, sample preparation protocols, and imaging techniques [25,15,27,36], pose challenges in model development. Recently, visual foundation models via prompt engineering have garnered significant attention for segmentation tasks of natural and biomedical images [18,2,22]. While they seem to seamlessly integrate into instance-based pipelines, the complexity of neural structures and the homogenization of EM images bring difficulties in identifying neuron instances. More importantly, those 2D models fail to tackle the extensive neuron splitting and merging in 3D space. Fortunately, neuronal membranes exhibit consistent characteristics across datasets, making it possible to develop a generalist model to predict boundary descriptors (*i.e.*, affinity maps) for 3D neuron reconstruction.

In this paper, we introduce **SegNeuron**, a boundary-based neuron segmentation model, generalized across diverse data distributions and spatial resolutions. Such strong adaptability hinges on a large-scale database and customized training strategies (see Fig. 2). Specifically, we construct a heterogeneous and nonreductant volumetric EM dataset, *i.e.*, **EMNeuron**, containing over 22 billion voxels in total. Combining publicly labeled datasets with crowdsourced annotations, we get fine-grained instance-level annotations of over 3 billion voxels after preprocessing. On this basis, we introduce a Gaussian noise addition-recovery proxy task for model pretraining. This novel technique builds mask reconstruction without distribution distortion in a multi-scale manner, enabling robust representation learning from unlabeled EM volumes. Moreover, the HOG feature is employed as an additional target to enhance the extraction of high-frequency features. The pretrained model is then finetuned on the labeled dataset to predict affinity maps. To prompt general feature extraction, we introduce frequency

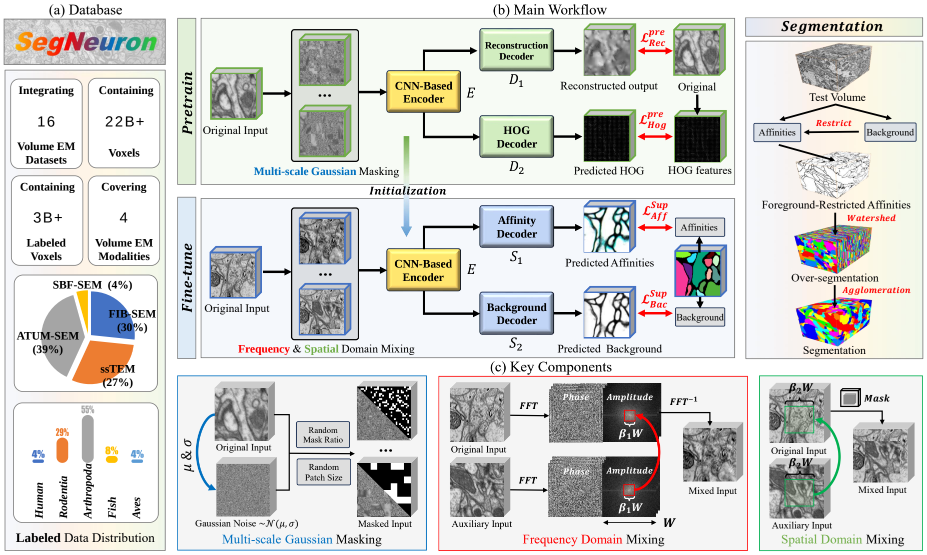


Fig. 2. A large-scale database and customized training strategies for SegNeuron.

and spatial domain mixing to generate training data with style-augmented appearances and mixed voxel resolutions. Finally, we restrict predicted affinities via foreground masks to remove noisy values for graph-based segmentation.

Qualitative and quantitative results illustrate the superior performance and strong generalizability of SegNeuron on both in- and out-of-distribution datasets. All key components are proven effective through extensive ablation studies.

2 Method

In the paper, we aim to train a generalist model for neuron segmentation, capable of predicting voxel affinities along the x-, y-, and z-axes for any EM volume. Here, we provide a formal definition of affinity maps. For a given EM image $I \in \mathbb{R}^{D\times H\times W}$, the three-channel affinity map $A \in \{0,1\}^{3\times D\times H\times W}$ used to characterize neuron boundaries is generated from dense instance annotations $G \in \mathbb{N}^{D\times H\times W}$. Each affinity value indicates whether two adjacent voxels belong to the same object along the $x/y/z$ axis (1 for belonging and 0 for not). Furthermore, the foreground mask is denoted by $F \in \{0,1\}^{D\times H\times W}$, where 1 represents the neuronal region and 0 represents the plasma membrane and myelin in G.

2.1 Database Construction

EMNeuron integrates 16 volume EM datasets (4 in-house, 12 publicly available), covering diverse species, sample preparation protocols, imaging techniques, and voxel resolutions. Detailed information is shown in Table 1 and Fig. 2(a). Dataset

Table 1. Details of EMNeuron. Underlined items represent in-house datasets.

Dataset	Modality	Res.(nm)	Total	Labeled	Dataset	Modality	Res.(nm)	Total	Labeled
		(x, y, z)	voxels	voxels			(x, y, z)	voxels	voxels
1.ZFinch[19]	SBF-SEM	9,9,20	3635M	131M	9.HBrain[25]	FIB-SEM	8,8,8	3072M	844M
2.ZFish[23]	SBF-SEM	9,9,20	1674M	-	10.FIB25[29]	FIB-SEM	8,8,8	312M	312M
3.vEM1	ATUM-SEM	8,8,50	1205M	157M	11.Minnie[7]	ssTEM	8,8,40	2096M	-
4.vEM2	ATUM-SEM	8,8,30	1329M	281M	12.Pinky[10]	ssTEM	8,8,40	1165M	117M
5.vEM3	ATUM-SEM	8,8,40	1301M	253M	13.FAFB[36]	ssTEM	8,8,40	2625M	577M
6.MitoEM[32]	ATUM-SEM	8,8,30	1048M	-	14.Basil[7]	ssTEM	8,8,40	23M	23M
7.H01[27]	ATUM-SEM	8,8,30	1166M	118M	15.Harris[12]	others	6,6,50	30M	30M
8.Kasthuri[17]	ATUM-SEM	6,6,30	1526M	478M	16.vEM4	others	8,8,20	45M	45M

1∼13 is used for model development, and 14∼16 for evaluation. Notably, the overall size of most datasets reaches the petabyte level, making it impractical to utilize all for training. Therefore, we first select representative and informative unlabeled areas from each dataset and integrate them with the labeled parts to construct our dataset. To avoid ambiguous feature learning caused by inconsistent annotation styles, we conduct comprehensive data cleaning and transforming, which includes adjusting tangential resolutions (x, y) to 6∼9 nm, unifying membrane thickness, and masking out unlabeled myelin and glial regions.

2.2 Model Training

Pretraining via Multi-scale Gaussian Mask Reconstruction Mask reconstruction is a well-established proxy task that learns general representations from large-scale unlabeled data. The key is to design appropriate masking strategies for EM data to eliminate information from the input. Transformer-based models make this easy by directly deleting selected patches or replacing them with mask tokens [14,33]. However, given the diverse resolutions and highly differentiated structures in EM datasets, operating in a single-scale manner with a fixed patch size limits the network's ability to capture boundary information at different levels. Another straightforward idea is to set masked pixels to zero/mean value. Unfortunately, this process leads to global statistics distortion and severe distribution changes between original and masked EM input.

To address these issues, we propose an architecture-agnostic masked image modeling framework tailored for volumetric EM datasets: **masking voxels with random Gaussian noise in a multi-scale manner**, as shown in Fig. 2(c). Specifically, we first partition input data equally into 3D spatial patches. The patch size is randomly generated based on input dimensions, prompting the network to learn multi-scale representations from the EM database with a diverse voxel resolution. Patches for masking are chosen randomly according to the masking ratio, and their pixel values are replaced with random values drawn from a Gaussian distribution $\mathcal{N}(\mu, \sigma)$ that matches the grayscale distribution of the original input. This strategy effectively alleviates global statistics distortion in masked inputs and prompts better modeling of local low-frequency information. The masked input is then fed into the network consisting of an encoder E and two decoders D_1, D_2, as in Fig. 2(b). D_1 is utilized to reconstruct the original input. For D_2, we adopt the histogram of oriented gradients (HOG) [8] as an additional prediction target, which proved to be effective in extracting high-frequency features [31,6]. The overall loss function $\mathcal{L}^{pre}$ can be expressed as

$$\mathcal{L}^{pre} = \mathcal{L}^{pre}_{Rec} + \lambda_1 \mathcal{L}^{pre}_{Hog} = \mathcal{L}_{MSE}(E(D_1(\tilde{I}, I))) + \lambda_1 \mathcal{L}_{MSE}(E(D_2(\tilde{I}, H))), \quad (1)$$

where $\tilde{I}$ represents the masked input, H is the calculated HOG feature, $\mathcal{L}_{MSE}$ denotes mean squared error loss, and λ_1 is the weight coefficient.

Domain-mixing Finetuning As illustrated in Fig. 2(b), we first initialize the encoder E in the segmentation network with pretrained weights. Subsequently, the entire model is finetuned in a supervised manner on a high-quality, heterogeneous EM database with diverse visual appearances and voxel resolutions.

To encourage learning of generalized knowledge for neuron affinities and avoid overfitting to dataset-specific biases, we introduce two customized mixing strategies in the frequency and spatial domains respectively (Fig. 2(c)). **In the frequency domain**, the amplitude spectrum represents low-level features such as texture and appearance, while the phase spectrum captures higher-level content such as shape and boundary. Hence, for each layer of the 3D input, we preserve the phase spectrum and replace the low-frequency amplitude component with that of the sampled auxiliary input, determined by frequency mixing ratio β_1. This operation preserves discriminative boundary information while significantly enriching the style and texture of the training data [34]. **In the spatial domain**, we extend CutMix [35] to the neuron segmentation task, generating new inputs with mixed semantic content based on spatial mixing ratio β_2. Consequently, we integrate data with diverse appearances and voxel resolutions into a unified input, which forces the network to extract domain-invariant features of affinity relationships. During training, we predict both neuron affinities and foreground masks for those mixed inputs via two decoders of the segmentation network, denoted as S_1 and S_2. The supervised loss function $\mathcal{L}^{sup}$ is given by

$$\mathcal{L}^{sup} = \mathcal{L}^{sup}_{Aff} + \lambda_2 \mathcal{L}^{sup}_{Fg} = L_{CE}(E(S_1(\hat{I}, A))) + \lambda_2 L_{CE}(E(S_2(\hat{I}, F))), \quad (2)$$

where $\hat{I}$ is the mixed input, L_{CE} denotes the cross entropy loss, and λ_2 denotes the weight assigned to the foreground segmentation task.

Foreground-restricted Segmentation To mask noise values, we filter background voxels in each channel of the predicted affinity map a using the predicted foreground mask f, as $a^r_{c,d,h,w} \equiv \min(a_{c,d,h,w}, f_{d,h,w})$. Such foreground-restricted affinity map a^r then serves as input for graph-based segmentation (Fig. 2(b)).

3 Experiments

3.1 Datasets and Evaluation Metrics

We employ the vEM4 (8, 8, 20 nm, in-house), Basil (8, 8, 40 nm, public) [7], and Harris (6, 6, 50 nm, public) [12] datasets for **out-of-distribution** evaluation,

Table 2. Qualitative results of different network architectures and pretraining schemes.

Methods			vEM4		Basil		Harris	
Architectures	Params/ FLOPs	Pretraining schemes	VI ↓	ARE ↓	VI ↓	ARE ↓	VI ↓	ARE ↓
UNETR[13]	115M/ 334G	from scratch	1.1190	0.1490	1.8955	0.4489	3.0017	0.3209
		SimSiam[5]	1.1059	0.1492	1.8446	0.4161	2.8935	**0.2734**
		MAE[14]	1.0829	0.1391	1.8145	**0.4130**	2.9475	0.3032
		Ours	**1.0493**	**0.1369**	**1.8002**	0.4182	**2.8000**	0.2877
SwinUNETR[30]	62M/ 197G	from scratch	1.0944	0.1679	1.5552	0.3509	2.5673	0.2736
		SimSiam[5]	1.0825	**0.1624**	1.5092	0.3562	2.5210	0.2367
		Ours	**1.0748**	0.1628	**1.4662**	**0.3131**	**2.4239**	**0.2252**
PNI-Net[21]	33M/ 317G	from scratch	0.9984	0.1480	0.9018	0.1782	0.9615	0.1319
		SimSiam[5]	0.9804	0.1393	0.8878	0.1765	0.8316	0.1052
		Ours	**0.9674**	**0.1295**	**0.8479**	**0.1647**	**0.8314**	**0.0934**
MNet[9]	40M/ 471G	from scratch	0.9096	0.1211	0.8437	0.1543	1.0788	0.1244
		Ours(SegNeuron)	**0.8655**	**0.1022**	**0.7719**	**0.1531**	**1.0221**	**0.1170**

which are unseen during model development and exhibit different appearances and voxel resolutions. Additionally, we reserve a small portion of the labeled database for **in-distribution** evaluation, including two annotated subvolumes from FIB25 (8, 8, 8 *nm*, public) [29], a proofread subvolume from Kasthuri (6, 6, 30 *nm*, public) [17], and two proofread subvolumes cropped from the FAFB (8, 8, 40 *nm*, public) [36]. Two common voxel-level metrics are used to evaluate the reconstruction performance: the variation of information (VI) [24] and adapted Rand error (ARE) [3]. Lower values in both metrics indicate higher segmentation quality, with the background being ignored during evaluation.

3.2 Implementation Details

We optimize all models using Adam with a learning rate of 1e-3, and a batch size of 8 for 400,000 iterations on 4 NVIDIA RTX V100 GPUs. Skip connections are disabled during pre-training. Instance results are obtained by solving the multicut problem with the Kernighan-Lin solver [4]. The masking ratio ranges between 0.5 and 0.7, and the mixing ratios β_1 and β_2 are 0.005 and 0.5. The weighting coefficient λ_1 in pretraining and λ_2 in finetuning are set to 5 and 1. The following data augmentation is used in finetuning: reflection, rotation, photometric perturbation, Gaussian noise/blur, Cutout, and anisotropic scaling. Please refer to the supplementary material and released codes for more details.

3.3 Results

Network Architectures and Pretraining Schemes For architecture, we consider the following alternatives in volumetric segmentation: UNETR [13], a network that adopts Vision Transformer as encoder; SwinUNETR [30], a segmentation model leveraging Swin Transformer as encoder; PNI-Net [21], a CNN-based model widely used in neuron segmentation; and MNet [9], a CNN-based network with mesh architecture. For pretraining schemes, we compare our

Table 3. Quantitative comparison of the generalist models.

Methods		vEM4		Basil		Harris	
		VI ↓	ARE ↓	VI ↓	ARE ↓	VI ↓	ARE ↓
2D	SAM[18]	2.8579	0.8113	2.3677	0.7881	2.0374	0.5252
	MicroSAM[2]	3.9972	0.9310	3.3483	0.8891	2.1904	0.5755
	SegNeuron	**0.4028**	**0.0839**	**0.6749**	**0.0922**	**0.5229**	**0.1063**
3D	SAM[18]	5.5215	0.8985	4.6624	0.9482	5.1034	0.8538
	MicroSAM[2]	5.7512	0.9640	4.8111	0.9655	4.4460	0.6566
	SegNeuron	**0.8655**	**0.1022**	**0.7719**	**0.1531**	**1.0221**	**0.1170**

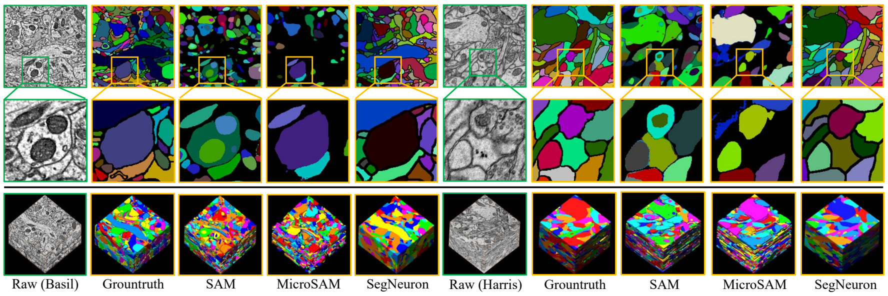


Fig. 3. 2D (top) and 3D (bottom) visual comparison of the generalist models.

method with MAE [14] and SimSiam [5]. Quantitative results on three out-of-distribution datasets provided in Table 2 support the following conclusions: (a) The proposed pretraining method demonstrates effectiveness across various network architectures and achieves consistent improvements over alternative strategies. (b) Transformer-based UNETR struggles to model diverse voxel resolutions and thus exhibits poor performance in affinities segmentation. SwinUNETR alleviates this problem through a heuristic structure, however it still lags behind CNN-based models by a wide margin in performance. (c) Due to the significant anisotropy present in the Harris dataset, PNI-Net with a 2.5D structure shows superior adaptability; Benefiting from a mesh architecture, MNet achieves optimal evaluation results on vEM4 and Basil datasets. In consideration of all factors, we adopt MNet as the architecture for SegNeuron in subsequent experiments.

Comparison with Generalist Models We consider the following generalist models for neuron reconstruction on out-of-distribution datasets: SAM [18], a prompt-based foundation model for segmentation, and MicroSAM [2], a fine-tuned version of SAM on EM data. Both methods first perform inference on every layer with grid-point prompts and non-maximum suppression. Then, an overlap-based connection algorithm is used to obtain the instance results in 3D space. Quantitative results in Table 3 demonstrate that in neuron segmentation, whether in 2D or 3D space, SegNeruon significantly outperforms existing generalist models, *i.e.*, with an average gain of **400%** on VI and **600%** on ARE. Moreover, visual comparisons are provided in Fig. 3. The segmentation results of

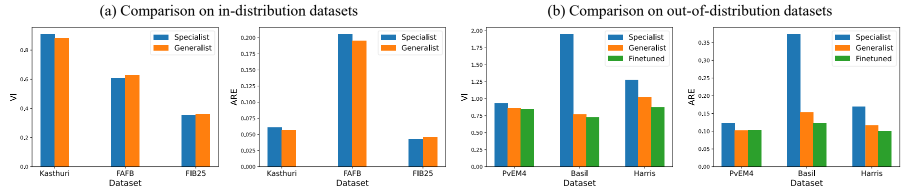


Fig. 4. Quantitative comparison with specialist models.

SegNeuron exhibit exceptional accuracy and spatial continuity. In contrast, SAM tends to generate false positive predictions in organelles, and while MicroSAM mitigates this after training on EM data, it also introduces more false negatives and even causes a decrease in overall performance. In conclusion, the practical value of SAM and MicroSAM in zero-shot neuron segmentation is limited, whereas SegNeuron effectively bridges this gap with strong generalizability.

Comparison with Specialist Models As shown in Fig. 4(a), SegNeuron achieves competitive segmentation performance with the specialist models on in-distribution datasets. This demonstrates that training a generalist model on heterogeneous datasets using the proposed strategies does not cause in-distribution performance degradation. For out-of-distribution evaluations, quantitative results in Fig. 4(b) suggest that SegNeuron can effectively extract general features of neuron boundaries on unseen distributions rather than overfitting the training data. Furthermore, we achieve further performance improvements and obtain the best results on out-of-distribution datasets by finetuning SegNeruon under the same settings. This indicates that SegNeuron can also serve as a powerful foundational model, providing excellent pretraining for new datasets.

Table 4. Abaltaion study for the generalist model development.

Methods		vEM4		Basil		Harris	
		VI ↓	ARE ↓	VI ↓	ARE ↓	VI ↓	ARE ↓
Database	-	0.9480	0.1241	0.9929	0.2288	1.2300	0.1298
	w/ preprocessing	**0.9338**	**0.1217**	**0.9625**	**0.1805**	**1.1036**	**0.1246**
Pretraining	-	0.9338	0.1217	0.9625	0.1805	1.1036	0.1246
	zero mask	0.9377	0.1173	0.8644	0.1854	1.0888	**0.1107**
	mean mask	0.9191	0.1197	0.9199	0.1921	1.1024	0.1194
	Gaussian mask	0.9241	**0.1102**	**0.8213**	**0.1647**	**1.0600**	0.1141
	Gaussian mask w/o multi-scale	**0.9119**	0.1168	0.9029	0.1713	1.0940	0.1306
	Gaussian mask w/o HOG loss	0.9341	0.1190	0.8446	0.1651	1.0882	0.1155
Finetuning	-	0.9241	0.1102	0.8213	0.1647	1.0600	0.1142
	w/ frequency mixing	0.9031	0.1112	0.7821	0.1544	1.0563	0.1197
	w/ spatial mixing	0.9058	0.1133	0.8208	0.1617	**0.9935**	**0.1126**
	w/ spatial & frequency mixing	**0.8655**	**0.1022**	**0.7719**	**0.1531**	1.0221	0.1170
Segmentation	-	**0.8655**	**0.1022**	**0.7719**	**0.1531**	**1.0221**	**0.1170**
	w/o foreground restriction	0.8879	0.1056	0.8069	0.1620	1.0414	0.1182

Ablations We validate the effectiveness of key components through ablation experiments on unseen datasets. Results in Table 4 show that: (a) Our pretraining strategy effectively leverages multi-scale perception and reduces statistics distortion, making it well-suited for neuron segmentation. (b) The mixing strategies in spatial and frequency domains encourage domain-invariant learning and enhance model generalization. (c) Preprocessing and foreground-restricted segmentation significantly improve performance without extra overhead.

4 Conclusion

This paper proposes SegNeuron, a neuron instance segmentation model trained on large-scale heterogeneous EM datasets with strong zero-shot generalization capabilities. We believe the released model can significantly simplify existing workflows and accelerate the scientific analysis of connectomics.

Acknowledgments. This work was supported by the STI 2030–Major Projects under Grant 2021ZD0204500 and Grant 2021ZD0204503, 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. Abbott, L.F., Bock, D.D., Callaway, E.M., et al.: The mind of a mouse. Cell **182**(6), 1372–1376 (2020)
2. Archit, A., Nair, S., Khalid, N., et al.: Segment anything for microscopy. bioRxiv pp. 2023–08 (2023)
3. Arganda-Carreras, I., Turaga, S.C., Berger, D.R., et al.: Crowdsourcing the creation of image segmentation algorithms for connectomics. Frontiers in Neuroanatomy **9**, 142 (2015)
4. Beier, T., Pape, C., Rahaman, N., et al.: Multicut brings automated neurite segmentation closer to human performance. Nature Methods **14**(2), 101–102 (2017)
5. Chen, X., He, K.: Exploring simple siamese representation learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognitio. pp. 15750–15758 (2021)
6. Chen, Y., Huang, W., Zhou, S., et al.: Self-supervised neuron segmentation with multi-agent reinforcement learning. In: Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence. pp. 609–617 (2023)
7. Consortium, M., Bae, J.A., Baptiste, M., et al.: Functional connectomics spanning multiple areas of mouse visual cortex. BioRxiv pp. 2021–07 (2021)
8. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: 2005 IEEE computer society conference on computer vision and pattern recognition. vol. 1, pp. 886–893. Ieee (2005)

9. Dong, Z., He, Y., Qi, X., et al.: Mnet: Rethinking 2d/3d networks for anisotropic medical image segmentation. In: Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence. pp. 870–876 (2022)
10. Dorkenwald, S., Turner, N.L., Macrina, T., et al.: Binary and analog variation of synapses between cortical pyramidal neurons. Elife **11**, e76120 (2022)
11. Funke, J., Tschopp, F., Grisaitis, W., et al.: Large scale image segmentation with structured loss based deep learning for connectome reconstruction. IEEE Transactions on Pattern Analysis and Machine Intelligence **41**(7), 1669–1680 (2018)
12. Harris, K.M., Spacek, J., Bell, M.E., et al.: A resource from 3d electron microscopy of hippocampal neuropil for user training and tool development. Scientific data **2**(1), 1–19 (2015)
13. Hatamizadeh, A., Tang, Y., Nath, V., et al.: Unetr: Transformers for 3d medical image segmentation. In: Proceedings of the IEEE/CVF winter conference on applications of computer vision. pp. 574–584 (2022)
14. He, K., Chen, X., Xie, S., et al.: Masked autoencoders are scalable vision learners. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognitio. pp. 16000–16009 (2022)
15. Hildebrand, D.G.C., Cicconet, M., Torres, R.M., et al.: Whole-brain serial-section electron microscopy in larval zebrafish. Nature **545**(7654), 345–349 (2017)
16. Januszewski, M., Kornfeld, J., Li, P.H., et al.: High-precision automated reconstruction of neurons with flood-filling networks. Nature Methods **15**(8), 605–610 (2018)
17. Kasthuri, N., Hayworth, K.J., Berger, D.R., et al.: Saturated reconstruction of a volume of neocortex. Cell **162**(3), 648–661 (2015)
18. Kirillov, A., Mintun, E., Ravi, N., et al.: Segment anything. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 4015–4026 (2023)
19. Kornfeld, J., Benezra, S.E., Narayanan, R.T., et al.: Em connectomics reveals axonal target variation in a sequence-generating network. Elife **6**, e24364 (2017)
20. Lee, K., Lu, R., Luther, K., Seung, H.S.: Learning and segmenting dense voxel embeddings for 3d neuron reconstruction. IEEE Transactions on Medical Imaging **40**(12), 3801–3811 (2021)
21. Lee, K., Zung, J., Li, P., et al.: Superhuman accuracy on the snemi3d connectomics challenge. arXiv:1706.00120 (2017)
22. Ma, J., He, Y., Li, F., et al.: Segment anything in medical images. Nature Communications **15**(1), 654 (2024)
23. Motta, A., Berning, M., Boergens, K.M., et al.: Dense connectomic reconstruction in layer 4 of the somatosensory cortex. Science **366**(6469), eaay3134 (2019)
24. Nunez-Iglesias, J., Kennedy, R., Parag, T., et al.: Machine learning of hierarchical clustering to segment 2d and 3d images. PloS one **8**(8), e71715 (2013)
25. Scheffer, L.K., Xu, C.S., Januszewski, M., et al.: A connectome and analysis of the adult drosophila central brain. Elife **9**, e57443 (2020)
26. Schmidt, M., Motta, A., Sievers, M., Helmstaedter, M.: Roboem: automated 3d flight tracing for synaptic-resolution connectomics. Nature Methods **21**(5), 908–913 (2024)
27. Shapson-Coe, A., Januszewski, M., Berger, D.R., et al.: A petavoxel fragment of human cerebral cortex reconstructed at nanoscale resolution. Science **384**(6696), eadk4858 (2024)
28. Sheridan, A., Nguyen, T.M., Deb, D., et al.: Local shape descriptors for neuron segmentation. Nature Methods pp. 1–9 (2022)
29. Takemura, S.y., Aso, Y., Hige, T., et al.: A connectome of a learning and memory center in the adult drosophila brain. Elife **6**, e26975 (2017)

30. Tang, Y., Yang, D., Li, W., et al.: Self-supervised pre-training of swin transformers for 3d medical image analysis. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognitio. pp. 20730–20740 (2022)
31. Wei, C., Fan, H., Xie, S., et al.: Masked feature prediction for self-supervised visual pre-training. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognitio. pp. 14668–14678 (2022)
32. Wei, D., Lin, Z., Franco-Barranco, D., 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. Springer (2020)
33. Xie, Z., Zhang, Z., Cao, Y., et al.: Simmim: A simple framework for masked image modeling. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognitio. pp. 9653–9663 (2022)
34. Yang, Y., Soatto, S.: Fda: Fourier domain adaptation for semantic segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognitio. pp. 4085–4095 (2020)
35. Yun, S., Han, D., Oh, S.J., et al.: Cutmix: Regularization strategy to train strong classifiers with localizable features. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 6023–6032 (2019)
36. Zheng, Z., Lauritzen, J.S., Perlman, E., et al.: A complete electron microscopy volume of the brain of adult drosophila melanogaster. Cell **174**(3), 730–743 (2018)