Collaborative Topology and Connectivity Learning for EM Neuron Segmentation

Haoyuan Shi[1,2], Xiaoyu Liu[1,2], Yinda Chen[1,2], Jingcheng Xie[1,2], and Zhiwei Xiong[1,2(✉)]

[1] MoE Key Laboratory of Brain-inspired Intelligent Perception and Cognition, University of Science and Technology of China

[2] Anhui Province Key Laboratory of Biomedical Imaging and Intelligent Processing, Institute of Artificial Intelligence, Hefei Comprehensive National Science Center

haoyuan.shi@mail.ustc.edu.cn, zwxiong@ustc.edu.cn

Abstract. Accurate neuron segmentation in electron microscopy (EM) images is fundamental for mapping neural circuits and advancing neuroscience research. While current affinity-based paradigms have achieved notable progress, they persistently struggle with blurred membranes and intricate topologies, yielding severe split and merge errors. To tackle this, we propose a two-stage collaborative topology and connectivity learning framework that explicitly injects EM-specific semantic priors into affinity prediction. Specifically, we leverage neuron skeletons for macroscopic topological guidance and distance transform maps for precise connectivity constraints. Since the direct aggregation of these auxiliary features frequently provokes negative transfer in ambiguous regions, our framework strategically decouples multi-task extraction from feature fusion. In Stage-I, we employ an Affinity-Collaborative Weighting Mechanism (ACWM) to stabilize joint multi-task feature extraction. Crucially, rather than applying a naive hard fusion, Stage-II introduces a cascaded refinement scheme featuring a Block-wise Representation Routing (BRR) strategy. BRR adaptively evaluates and routes only the most reliable feature streams for block-level conditional fusion, inherently preventing noise propagation. To satisfy the rigorous efficiency demands of terabyte-scale inference, we compress the Stage-I backbone via gradient-based structural filter pruning. Extensive experiments across three benchmarks demonstrate that our method establishes state-of-the-art topological correctness while securing a highly efficient deployment architecture. Code is available at https://github.com/vshy-dream/CTCL-NeuronSeg.

Keywords: Electron microscopy · Neuron segmentation · Collaborative learning.

1 Introduction

Accurate neuron segmentation serves as the cornerstone of connectomics analysis, which is fundamental for deciphering brain-wide neural circuits [26, 8, 1, 24]. While recent advances in EM technologies have facilitated the acquisition of

massive volumetric neuronal data at nanometer resolution, densely intertwined synapses and ambiguous cellular boundaries still impede robust automated neuron segmentation [4, 3, 7]. To navigate these issues, affinity-based methods have emerged as the mainstream paradigm for large-scale connectomics [16, 18, 9, 5]. By predicting 3D graphs that quantify spatial continuity between adjacent voxels, these methods successfully capture local structural relationships. Nevertheless, relying solely on local affinities inherently lacks explicit macroscopic structural guidance, commonly yielding (i) erroneous splits along elongated processes and (ii) incorrect merges across blurred membranes.

A practical way to enrich representations under tight compute budgets is auxiliary learning. For instance, Local Shape Descriptor (LSD) [23] predicts local geometric attributes to provide discriminative signals for boundary disambiguation, yet it does not explicitly encode global topology or connectivity constraints, so affinity prediction can still fail in topologically intricate regions.

To bridge this gap, we introduce two heterogeneous EM-specific features that capture distinct yet complementary aspects of neuron shapes, i.e., neuron skeleton maps and distance transform (DT) maps. As illustrated in Fig.1(a), a skeleton map capturing the medial-axis architecture to enforce long-range continuity and suppress splits [19, 10], and a DT map measuring proximity to membranes to supply a continuous separation-strength field that penalizes connectivity leaks and mitigates merges [21, 27].

However, directly aggregating these heterogeneous auxiliary features via naive hard fusion inevitably provokes negative transfer, as noisy predictions in densely intertwined regions can easily mislead the primary affinity map. To conquer this, we design a deployable, two-stage systematic collaborative framework. First, an uncertainty-driven Affinity-Collaborative Weighting Mechanism (ACWM) stabilizes the joint extraction of multi-task features, dynamically balancing topology and connectivity learning. Second, to prevent noise propagation during feature integration, we introduce a cascaded refinement scheme featuring a Block-wise Representation Routing (BRR) strategy. Crucially, unlike classical Mixture-of-Experts (MoE) that introduces computationally heavy independent branches, our BRR adaptively evaluates and routes the semantically grounded feature streams (raw image, skeleton, DT) derived by a single shared backbone. Operating at the block level, BRR selectively gates only the most reliable streams for conditional fusion, effectively suppressing negative transfer. Finally, to satisfy the rigorous efficiency demands of terabyte-scale connectomics, we compress the ACWM backbone by 40% via gradient-based structural filter pruning, eliminating inter-feature redundancy and achieving a lightweight deployment architecture.

In summary, the main contributions of this work are as follows: (1) We propose a collaborative framework that integrates global skeletons and precise DT maps as explicit EM semantic priors to resolve severe split and merge errors in affinity prediction. (2) We design a two-stage pipeline featuring an uncertainty-driven ACWM for stable multi-task extraction, and a BRR module that inherently prevents negative transfer without the massive overhead

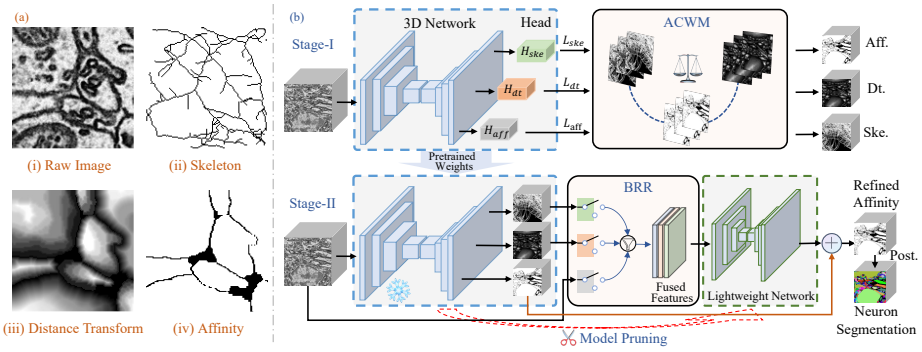

Fig. 1. (a) Visualization of EM-specific semantic priors. The raw image, distance transform, and affinity are visualized through a single section, while the skeleton is visualized through projection mapping within a volume. (b) Architecture of our two-stage collaborative learning framework. Stage-I employs ACWM to stabilize joint heterogeneous features extraction. Stage-II utilizes the BRR module to adaptively gate and route only the most reliable feature streams for conditional fusion, yielding refined affinities. The shared backbone is further compressed via pruning for efficient deployment.

of traditional MoE. (3) Enhanced by gradient-based structural filter pruning (yielding a 40% backbone compression), our computationally lightweight solution establishes state-of-the-art performance across three connectomics benchmarks (AC3/4, CREMI, Wafer4) with exceptional inference efficiency.

2 Method

2.1 Overview

Given a raw EM subvolume $V \in \mathbb{R}^{Z\times H\times W}$, our goal is to predict an affinity graph $A \in [0,1]^{3\times Z\times H\times W}$ for downstream agglomeration. As illustrated in Fig. 1, we propose a sequentially optimized two-stage framework. Stage-I jointly extracts affinity, skeleton, and distance-transform targets from a shared backbone, and employs an Affinity-Collaborative Weighting Module (ACWM) to stabilize optimization. Stage-II introduces a cascaded refinement scheme equipped with Block-wise Representation Routing (BRR) to conditionally fuse these heterogeneous representations, inherently averting negative transfer in ambiguous regions. To satisfy the strict throughput requirements of terabyte-scale EM volumes, we further apply structural filter pruning to the Stage-I backbone.

2.2 Stage-I: Collaborative Learning

In the first stage, our framework extracts heterogeneous representations from a shared backbone. To explicitly model macroscopic structures and boundaries for collaborative learning, we construct three complementary supervision targets from the instance label volume $Y \in \mathbb{N}^{Z\times H\times W}$.

Affinity Map. We predict a 1-neighborhood affinity graph $A^\star \in \{0,1\}^{3\times Z\times H\times W}$ along $\{z,y,x\}$. Given spatial offsets $\delta\in\{(1,0,0),(0,1,0),(0,0,1)\}$, the ground-truth for voxel p is defined as $A^\star_\delta(p) = \mathbb{I}[Y(p) \neq 0 \wedge Y(p) = Y(p+\delta)]$. The predicted affinity $\hat{A}$ is optimized via Binary Cross-Entropy (BCE) loss: $\mathcal{L}_{\text{aff}} = -\mathbb{E}[\, A^\star_\delta \log \hat{A}_\delta + (1-A^\star_\delta)\log(1-\hat{A}_\delta)\,]$.

Neuron Skeleton Map. To embed long-range topological continuity, we construct a global binary skeleton target $S^\star$. It is worth noting that the 3D skeleton computation is a one-time offline preprocessing step applied strictly to annotated training data. During inference on massive unlabeled EM volumes, this step is fully bypassed, thereby preserving terabyte-scale scalability. For each instance k, we compute the medial-axis skeleton S_k via a topology-preserving thinning algorithm [17]. To mitigate fragmentation induced by anisotropic EM resolution, we dilate S_k with a 3×3×3 kernel K and aggregate them into $S^\star = \max_k(S_k \oplus K)$. The prediction $\hat{S}$ is supervised by BCE loss: $\mathcal{L}_{\text{ske}} = -\mathbb{E}[\, S^\star \log \hat{S} + (1-S^\star)\log(1-\hat{S})\,]$.

Distance Transform Map. To impose an explicit boundary-proximity penalty against erroneous merges, we generate a continuous connectivity field $T^\star$. For instance k with voxel set I_k and boundary ∂I_k, the interior boundary distance is $D_k(p) = \min_{q\in\partial I_k} \|p-q\|_2$. To eliminate size bias from varying neuron calibers, we normalize it as $\tilde{D}_k(p) = D_k(p)/(\max_{u\in I_k} D_k(u)+\epsilon)$ where $\epsilon = 10^{-10}$. Aggregating these yields $T^\star(p) = \mathbb{I}[Y(p)\geq 1]\cdot \tilde{D}_{Y(p)}(p)$. The prediction $\hat{T}$ is optimized via Mean Squared Error (MSE) loss: $\mathcal{L}_{\text{dt}} = \mathbb{E}[\,(\hat{T}-T^\star)^2\,]$.

Affinity-Collaborative Weighting Mechanism. Jointly predicting these heterogeneous representations from a shared backbone naturally introduces gradient interference, as auxiliary representations exhibit varying signal-to-noise ratios across ambiguous boundaries. To stabilize optimization during this collaborative learning stage, we formulate the ACWM objective:

$$\mathcal{L}_{\text{ACWM}} = \frac{1}{2\sigma^2}\Big(\mathcal{L}_{\text{aff}} + \lambda_{\text{ske}}\mathcal{L}_{\text{ske}} + \lambda_{\text{dt}}\mathcal{L}_{\text{dt}}\Big) + \log\sigma, \tag{1}$$

where $\sigma > 0$ is a learnable global collaboration confidence inspired by multi-task uncertainty weighting [14], and $\lambda_{\text{ske}}, \lambda_{\text{dt}}$ are learnable uncertainty variables (initialized to 0.5) that automatically balance the auxiliary supervision. While ACWM *stabilizes* this multi-task extraction, negative transfer during subsequent feature *fusion* is further addressed by the cascaded refinement in Stage-II.

2.3 Stage-II: Cascaded Refinement

While ACWM stabilizes backbone training, the auxiliary semantic priors remain unevenly reliable across ambiguous boundaries. To prevent downstream refinement gradients from disrupting the well-calibrated representations of Stage-I, we decouple their training. Specifically, we freeze the optimized Stage-I backbone and design a cascaded refinement scheme featuring a BRR module and a pruned lightweight backbone.

Block-wise Representation Routing. Operating dynamically at the local block level, BRR receives a feature set $\mathcal{F} = \{F_{\text{raw}}, F_{\text{ske}}, F_{\text{dt}}\}$ derived from the Stage-I backbone, where F_{raw} denotes the implicit features from the raw image. Rather than a standard engineering combination, this design facilitates a synergistic cross-modal fusion, utilizing explicit structural priors to spatially guide and filter the implicit features. To adjudicate the region-specific discriminative capacity of each stream, BRR compresses them into global context vectors via spatial average pooling: $g_i = \mathbb{E}_{d,h,w}[F_i(d,h,w)]$. A normalized scoring function, $s = \text{softmax}(W \cdot [g_{\text{raw}}; g_{\text{ske}}; g_{\text{dt}}])$ (with W being a learned projection), evaluates the relative confidence of each stream. To rigorously filter out deceptive cues, BRR conditionally routes only the Top-K streams based on s, where K is determined empirically via subsequent ablations. The selected streams, indexed by $\mathcal{K}^*$, are concatenated and integrated via a $1 \times 1 \times 1$ convolution $\mathcal{C}$: $F_{\text{fused}} = \mathcal{C}\left(\text{Concat}(\{F_i : i \in \mathcal{K}^*\})\right)$. Finally, a lightweight network $\mathcal{N}'$ refines F_{fused} to yield the residual correction over the Stage-I affinity $\hat{A}$: $A_{\text{final}} = \hat{A} + \mathcal{N}'(F_{\text{fused}})$. This gating mechanism inherently dictates that only semantically and structurally reliable representations condition the final affinity.

Gradient-based Structural Filter Pruning. To satisfy the rigorous latency constraints of terabyte-scale connectomics, we compress the Stage-I backbone via structural filter pruning. Since affinity prediction constitutes the sole deployment objective, the pruning criterion is exclusively governed by the affinity loss $\mathcal{L}_{\text{aff}}$. Bypassing unstructured weight masking, we evaluate the importance of entire 3D filters by tracking their accumulated gradient magnitudes. Specifically, the importance score $S_{l,i}$ for the i-th filter $W_{l,i}$ in the l-th layer is defined as the expected L_2-norm of its affinity-driven gradients over N_{cal} calibration steps:

$$S_{l,i} = \frac{1}{N_{\text{cal}}} \sum_{t=1}^{N_{\text{cal}}} \left\| \frac{\partial \mathcal{L}_{\text{aff}}^{(t)}}{\partial W_{l,i}} \right\|_2 . \tag{2}$$

To avoid brittle threshold tuning, we enforce a fixed global compression budget (e.g., 40% reduction). Capitalizing on the hierarchical nature of EM representations, shallow layers are pruned conservatively to preserve high-frequency boundary textures, whereas deeper abstract layers are compressed more aggressively. By physically discarding filters with the lowest $S_{l,i}$, this protocol explicitly strips away the auxiliary branches, which become dispensable during inference. Ultimately, this yields a physically compressed, lightweight Stage-I model without altering the discrete routing logic of the Stage-II cascaded refinement.

3 Experiments and Results

3.1 Datasets and Metrics

We evaluate our framework on three heterogeneous EM benchmarks: AC3/4 [13] (mouse cortex, anisotropic 3×3×29 nm), Wafer4 [20] (mouse cortex, 8×8×35

Table 1. Quantitative comparison of different methods on the AC3/4, Wafer4, and CREMI datasets. 'Super.' denotes the Superhuman method [16]. 'S-I' and 'S-II' represent Stage-I and Stage-II of our approach, respectively. Best results are **bolded** and second-best are underlined.

Method	AC3/4		Wafer4		CREMI-A		CREMI-B		CREMI-C	
	VOI	ARAND	VOI	ARAND	VOI	ARAND	VOI	ARAND	VOI	ARAND
Super. [16]	1.0305	0.1794	0.6176	0.0411	0.6396	0.0887	0.7755	0.0482	1.1579	0.1793
MALA [9]	1.1338	0.1664	0.6133	0.0361	0.6336	0.0846	0.8501	0.0407	1.1739	0.1621
UNETR [11]	1.5376	0.1484	0.7630	0.0453	0.6937	0.1089	1.0388	0.0542	1.2344	0.1865
PEA [12]	1.0502	0.2093	0.5930	0.0341	0.6264	0.0909	0.7847	0.0407	1.1914	0.1689
APViT [25]	0.9764	0.0775	0.7039	0.0362	0.7041	0.1169	0.7807	0.0319	1.1180	**0.1102**
LSD [23]	0.9129	0.0927	0.5598	0.0262	0.6101	**0.0697**	0.8047	0.1224	1.0653	0.1495
CAD [20]	0.8835	0.0808	0.5587	0.0302	0.5653	0.0788	0.6844	**0.0297**	1.0597	0.1487
Ours (S-I)	0.7887	0.0695	0.5281	0.0283	0.5642	0.0709	0.7275	0.0329	1.0288	0.1443
Ours (S-II)	**0.7196**	**0.0586**	**0.5193**	**0.0259**	**0.5556**	0.0701	**0.6830**	0.0309	**1.0106**	0.1433

nm), and CREMI [6] (Drosophila brain, varying complexities A-C). To rigorously assess cross-sample generalization on AC3/4, we train on the first 80 sections of AC4, employ the last 20 sections as a separate validation set for model tuning, and test on the first 100 sections of AC3. For Wafer4 and CREMI, the initial 100 and subsequent 25 sections constitute the train and test splits, respectively. Performance is quantified via Adapted Rand Error (ARAND) [2] and Variation of Information (VOI) [22]. Crucially, VOI is the primary metric as it precisely diagnoses topological correctness. Lower values denote superior performance.

3.2 Implementation Details

Our framework is trained in a strictly sequential manner to optimize memory efficiency and representation stability, and we adopt the widely recognized Superhuman [16] as the shared backbone, which utilizes lightweight parallel output heads to ensure the architecture introduces minimal training overhead. The Stage-II cascaded refinement employs a structurally identical but distinctively initialized encoder-decoder. Models are trained using the Adam optimizer (β_1=0.9, β_2=0.999, learning rate 1×10^{-4}) [15] for 400k iterations on a single NVIDIA RTX 3090. Following standard practice [9], the Waterz algorithm [9] is utilized as post-processing to obtain neuron instances from the affinity maps.

3.3 Comparison with State-of-the-art Methods

Quantitative Segmentation Results. We conduct a comprehensive comparison against several state-of-the-art methods, including Superhuman [16], MALA [9], UNETR [11], PEA [12], APViT [25], LSD [23] and CAD [20], as detailed in Table 1. Remarkably, on the AC3/4 dataset, our Stage-I alone already establishes new state-of-the-art performance in both VOI and ARAND. The subsequent integration of the BRR module in Stage-II further boosts the performance, achieving an exceptional VOI of 0.7196 and an ARAND of 0.0586,

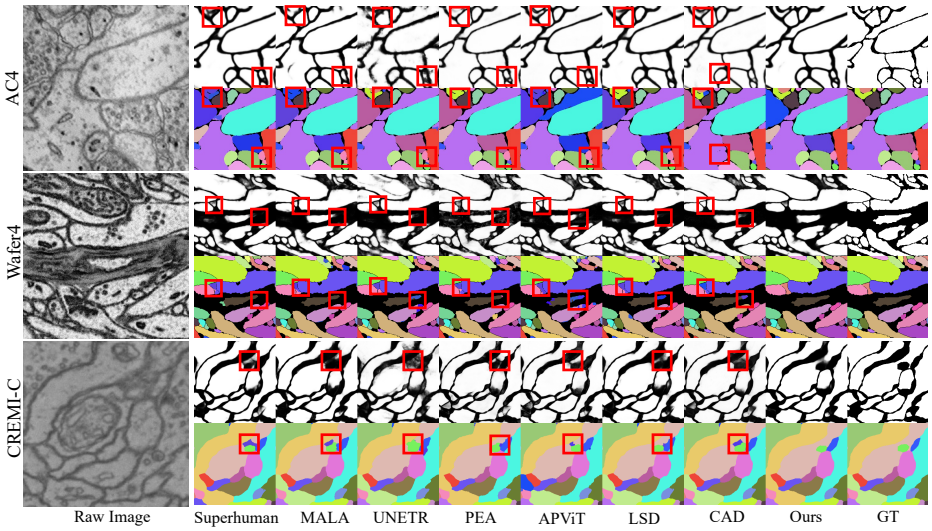

Fig. 2. The 2D visualization results on various datasets. In each dataset, the first row displays the affinity map, while the second row shows the corresponding instance segmentation result. Red boxes indicate the regions of mis-segmentation.

yielding relative error reductions of approximately 19% and 28% compared to the second-best approach. This robust performance trend is consistent across all evaluated benchmarks.

Notably, our framework demonstrates the most pronounced gains on highly challenging datasets characterized by densely intertwined neural structures (e.g., AC3/4, Wafer4, and CREMI-C), underscoring the effectiveness of our collaborative topology and connectivity learning paradigm.

Qualitative Segmentation Results. We present the 2D and 3D visual comparisons in Fig. 2 and 3, respectively. As shown in Fig. 2, our method produces sharper boundaries compared to existing baselines, effectively suppressing mis-segmentation errors. The 3D visualization segmentation, depicted in Fig. 3, further highlights the ability of our method to preserve intricate topological continuity, particularly in densely intertwined regions where the baseline methods fragment or merge neurites erroneously.

3.4 Ablation Study

We ablate our core components on the AC3/4 dataset. As detailed in Table 2, incorporating precise connectivity ($\mathcal{L}_{\text{dt}}$) and global topology ($\mathcal{L}_{\text{ske}}$) priors substantially reduces segmentation errors, while integrating ACWM further boosts performance by mitigating multi-task gradient conflicts. To effectively fuse these harmonized representations, our cascaded refinement strategy proves to be critical (Table 3). Specifically, routing the top K=2 semantic streams via BRR

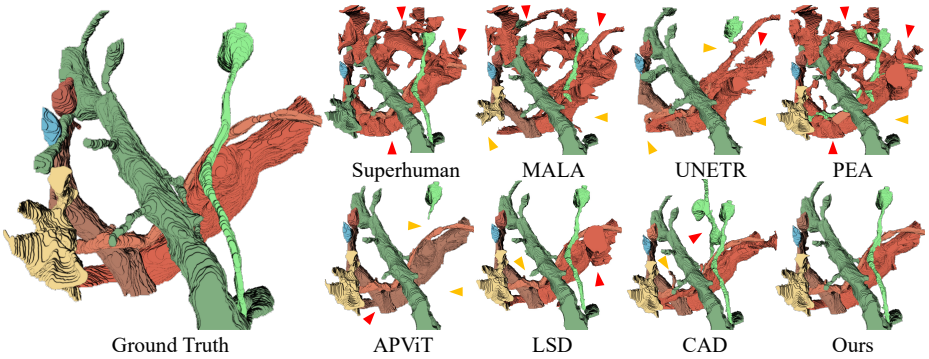

Fig. 3. The 3D visualization results on AC3/4 dataset. The red arrows indicate over-merged structures, and the orange arrows highlight over-split structures.

Table 2. Ablation study for the effectiveness of EM-specific features and ACWM on AC3/4 dataset.

L_{aff}	L_{dt}	L_{ske}	$ACWM$	VOI	ARAND
✓				1.0305	0.1794
✓	✓			0.8505	0.0865
✓		✓		0.8427	0.0827
✓	✓	✓		0.8096	0.0706
✓	✓	✓	✓	**0.7887**	**0.0695**

Table 3. Ablation study for the effectiveness of BRR and the selection of K on AC3/4 dataset.

BRR	K=3	K=2	K=1	VOI	ARAND
✓	✓			0.7887	0.0695
✓			✓	0.7427	0.0697
✓		✓		0.8392	0.0812
✓				**0.7299**	**0.0655**

Table 4. Ablation study for the effectiveness of gradient-based filter pruning on the AC3/4 dataset. Parameters (Param.) are measured in millions, FLOPs are measured in Giga Multiply-Accumulate operations, and Latency in milliseconds.

Stage	Pruning	Param.	FLOPs	Latency	VOI	ARAND
Stage-I	✗	1.47771	87.34	19.49	0.7887	0.0695
	✓	0.88239	64.00	16.94	0.8845	0.0801
Stage-II	✗	1.47767	87.31	19.38	0.7299	0.0655
	✓	0.88235	63.98	16.92	**0.7196**	**0.0586**

outperforms both naive hard fusion (K=3) and single-stream reliance (K=1), demonstrating its capacity to filter deceptive regional cues and prevent negative transfer. Furthermore, Table 4 reveals a synergistic effect of structural filter pruning within our cascaded pipeline, where Stage-II refines Stage-I's intermediate outputs into final results, with its reported metrics reflecting the pruned model. Although pruning marginally degrades the standalone Stage-I representation, its integration into the complete two-stage pipeline paradoxically elevates overall accuracy. We attribute this to reduced inter-feature redundancy, which acts as a potent structural regularizer for the BRR module. By compelling the refinement network to exploit only essential cues, this synergy yields a highly accurate yet computationally lightweight architecture.

4 Conclusion

We propose a two-stage collaborative learning framework for EM neuron segmentation that explicitly models topology and connectivity. By employing ACWM for stable multi-task extraction and BRR for reliable feature routing, our method effectively suppresses negative transfer in ambiguous regions. Furthermore, our gradient-based structural filter pruning establishes a lightweight deployment architecture without compromising representational capacity. Extensive evaluations demonstrate our approach achieves state-of-the-art topological correctness across diverse benchmarks, providing a robust and highly efficient engine for terabyte-scale connectomics analysis. While effective on these standard benchmarks, current datasets may lack the extreme challenges of brain-scale long-range continuity, making massive-scale evaluation a crucial future direction.

Acknowledgments. This work was supported by the National Natural Science Foundation of China under Grant 624B2137.

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., Denk, W., Dulac, C., Fairhall, A.L., Fiete, I., Harris, K.M., Helmstaedter, M., Jain, V., et al.: The mind of a mouse. Cell **182**(6), 1372–1376 (2020)
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. Chen, Y., Huang, W., Liu, X., Deng, S., Chen, Q., Xiong, Z.: Learning multi-scale consistency for self-supervised electron microscopy instance segmentation. In: ICASSP. pp. 1566–1570. IEEE (2024)
4. Chen, Y., Huang, W., Zhou, S., Chen, Q., Xiong, Z.: Self-supervised neuron segmentation with multi-agent reinforcement learning. In: IJCAI. pp. 609–617 (2023)
5. Chen, Y., Shi, H., Liu, X., Shi, T., Zhang, R., Liu, D., Xiong, Z., Wu, F.: Tokenunify: Scalable autoregressive visual pre-training with mixture token prediction. arXiv preprint arXiv:2405.16847 (2024)
6. CREMI: Miccai challenge on circuit reconstruction from electron microscopy images. https://cremi.org/ (2016)
7. Deng, S., Chen, Y., Huang, W., Zhang, R., Xiong, Z.: Unsupervised domain adaptation for em image denoising with invertible networks. IEEE Transactions on Medical Imaging (2024)
8. Dorkenwald, S., McKellar, C.E., Macrina, T., Kemnitz, N., Lee, K., Lu, R., Wu, J., Popovych, S., Mitchell, E., Nehoran, B., et al.: Flywire: online community for whole-brain connectomics. Nature methods **19**(1), 119–128 (2022)

9. Funke, J., Tschopp, F., Grisaitis, W., Sheridan, A., Singh, C., Saalfeld, S., Turaga, S.C.: 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)
10. Goodman, S., Zagon, I.: The neural cell spectrin skeleton: a review. American Journal of Physiology-Cell Physiology **250**(3), C347–C360 (1986)
11. Hatamizadeh, A., Tang, Y., Nath, V., Yang, D., Myronenko, A., Landman, B., Roth, H.R., Xu, D.: Unetr: Transformers for 3d medical image segmentation. In: WACV. pp. 574–584 (2022)
12. Huang, W., Deng, S., Chen, C., Fu, X., Xiong, Z.: Learning to model pixel-embedded affinity for homogeneous instance segmentation. In: AAAI. vol. 36, pp. 1007–1015 (2022)
13. Kasthuri, N., Hayworth, K.J., Berger, D.R., Schalek, R.L., Conchello, J.A., Knowles-Barley, S., Lee, D., Vázquez-Reina, A., Kaynig, V., Jones, T.R., et al.: Saturated reconstruction of a volume of neocortex. Cell **162**(3), 648–661 (2015)
14. Kendall, A., Gal, Y., Cipolla, R.: Multi-task learning using uncertainty to weigh losses for scene geometry and semantics. In: CVPR. pp. 7482–7491 (2018)
15. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
16. Lee, K., Zung, J., Li, P., Jain, V., Seung, H.S.: Superhuman accuracy on the snemi3d connectomics challenge. arXiv preprint arXiv:1706.00120 (2017)
17. Lee, T.C., Kashyap, R.L., Chu, C.N.: Building skeleton models via 3-d medial surface axis thinning algorithms. CVGIP: graphical models and image processing **56**(6), 462–478 (1994)
18. Li, Z., Yang, X., Liu, J., Hong, B., Zhang, Y., Zhai, H., Shen, L., Chen, X., Liu, Z., Han, H.: Deepmulticut: Deep learning of multicut problem for neuron segmentation from electron microscopy volume. IEEE Transactions on Pattern Analysis and Machine Intelligence **46**(12), 8696–8714 (2024)
19. Liao, M., Wan, G., Du, B.: Joint learning neuronal skeleton and brain circuit topology with permutation invariant encoders for neuron classification. In: AAAI. vol. 38, pp. 197–205 (2024)
20. Liu, X., Cai, M., Chen, Y., Zhang, Y., Shi, T., Zhang, R., Chen, X., Xiong, Z.: Cross-dimension affinity distillation for 3d em neuron segmentation. In: CVPR. pp. 11104–11113. IEEE Computer Society (2024)
21. Mishchenko, Y., Hu, T., Spacek, J., Mendenhall, J., Harris, K.M., Chklovskii, D.B.: Ultrastructural analysis of hippocampal neuropil from the connectomics perspective. Neuron **67**(6), 1009–1020 (2010)
22. Nunez-Iglesias, J., Kennedy, R., Parag, T., Shi, J., Chklovskii, D.B.: Machine learning of hierarchical clustering to segment 2d and 3d images. PloS one **8**(8), e71715 (2013)
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. Silbereis, J.C., Pochareddy, S., Zhu, Y., Li, M., Sestan, N.: The cellular and molecular landscapes of the developing human central nervous system. Neuron **89**(2), 248–268 (2016)
25. Sun, R., Luo, N., Pan, Y., Mai, H., Zhang, T., Xiong, Z., Wu, F.: Appearance prompt vision transformer for connectome reconstruction. In: IJCAI. pp. 1423–1431 (2023)

26. Winding, M., Pedigo, B.D., Barnes, C.L., Patsolic, H.G., Park, Y., Kazimiers, T., Fushiki, A., Andrade, I.V., Khandelwal, A., Valdes-Aleman, J., et al.: The connectome of an insect brain. Science **379**(6636), eadd9330 (2023)
27. Yang, J., Gonzalez-Bellido, P.T., Peng, H.: A distance-field based automatic neuron tracing method. BMC bioinformatics **14**(1), 93 (2013)