NeuroRefiner: Morphology-Aware Multi-Agent Refinement for 3D Fluorescence Microscopy Neuron Segmentation

Haiyang Yan[1,2], Jinyue Guo[1,3], Yanchao Zhang[1,2], Bingqing Wang[1,3], Zhenchen Li[4], Jing Liu[1], Jiazheng Liu[1], Linlin Li[1], and Hua Han[1,2(✉)]

[1] State Key Laboratory of Brain Cognition and Brain-inspired Intelligence Technology, Institute of Automation, Chinese Academy of Sciences
{yanhaiyang2022, hua.han}@ia.ac.cn
[2] School of Future Technology, University of Chinese Academy of Sciences
[3] School of Artificial Intelligence, University of Chinese Academy of Sciences
[4] The State Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University, Beijing, China

Abstract. Accurate 3D neuron segmentation in fluorescence microscopy is critical for neuroscience. However, the sparse and elongated morphology of neurons poses significant challenges to existing segmentation methods. These methods struggle to preserve both local details and global topology, leading to fragmented results. To address this, we propose NeuroRefiner, a multi-agent system that formalizes the human expert workflow involving iterative global observation and local editing. Specifically, NeuroRefiner comprises three collaborative agents dedicated to diagnosing topological errors, generating correction instructions, and validating refinement quality. To facilitate agent instruction-guided segmentation refinement, we propose TopoRefineNet, a dedicated 3D U-Net-based tool that leverages cross-modality feature fusion to generate refined masks. Through multi-round agent reasoning and voxel-level editing, NeuroRefiner produces topologically more accurate segmentations with enhanced interpretability. Experiments on the BigNeuron, CWMBS, and ZBFWB datasets demonstrate that NeuroRefiner outperforms state-of-the-art methods, notably achieving a 3.02% improvement in F1 score on the challenging ZBFWB dataset.

Keywords: Neuron reconstruction · Multi-agent system · Segmentation refinement · Fluorescence Microscopy

1 Introduction

Neuronal reconstruction aims to derive neuronal skeletons from 3D fluorescence microscopy volumes, thereby enabling quantitative analysis in neuroscience [10, 47]. Accurate reconstruction requires not only recovering fine-grained structural details but also preserving topological correctness. Unlike electron microscopy (EM) neuron reconstruction [14,34,49], which typically deals with dense

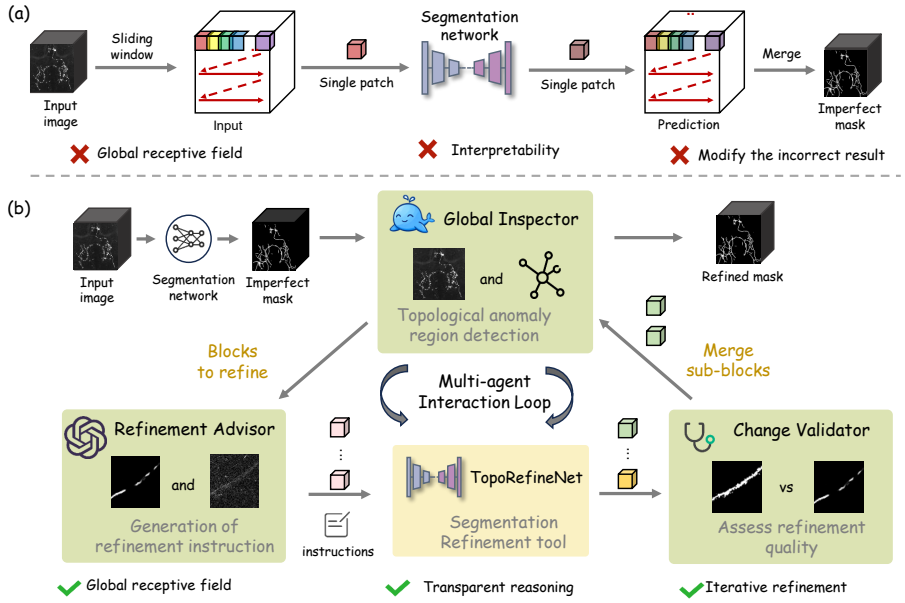

Fig. 1: (a) Mainstream neuron segmentation paradigms. Sliding window-based methods fail to preserve holistic neuronal morphology. Moreover, a single forward pass cannot rectify segmentation errors and lacks interpretability. (b) Overview of NeuroRefiner. By combining global topological observation with local fine-grained refinement, it generates continuous segmentation results consistent with topological priors. The explicit and transparent reasoning process of the agents offers enhanced interpretability.

cellular structures, fluorescence microscopy poses distinct challenges. These challenges stem from extreme signal sparsity and filamentous morphology within high-resolution 3D volumes that are frequently corrupted by substantial noise [4, 23]. Consequently, robust automated reconstruction requires multi-scale features to preserve both structural details and global connectivity.

Neuron segmentation plays a pivotal role in neuronal reconstruction by enhancing weak signals and suppressing noise [20,22]. As shown in Fig. 1 (a), given the prohibitive size of raw image volumes, existing segmentation approaches [37, 38, 41, 44] typically rely on sliding-window strategies for inference. Nevertheless, the limited receptive field in each local window precludes the modeling of long-range dependencies essential for preserving global neuronal morphology [42], resulting in fragmented segmentation. Furthermore, these end-to-end approaches typically operate as black boxes, lacking interpretability and offering no mechanism for the correction of significant topological errors in segmentation. Such errors are challenging to localize and rectify, ultimately constituting a critical bottleneck for the automation of neuronal reconstruction.

Agent-based biomedical image segmentation methods [15,16,45] have demonstrated enhanced performance and interpretability via the reasoning and tool in-

vocation capabilities of large language models (LLMs). However, existing agents and their associated tool designs remain ill-suited for neuron segmentation. First, they depend on foundation models (e.g., MedSAM2 [25]) as tools, which process 3D volumes slice-by-slice using 2D encoders. Yet, this proves inadequate for neurons, as their structural sparsity and ambiguous boundaries hinder effective feature extraction within isolated 2D planes [51]. Furthermore, these agents confine the workflow to single tool invocations, which lack the flexibility to iteratively refine the segmentation via agent instructions. These limitations underscore the critical need for a dedicated agent framework equipped with domain-specific tools tailored for neuron segmentation.

Motivated by the neuroscientist's cognitive paradigm of global observation and local refinement, we propose NeuroRefiner, a novel multi-agent framework designed for iterative, topology-aware neuron segmentation. As depicted in Fig. 1 (b), our approach comprises three specialized agents: the Global Inspector, which leverages initial segmentation masks and their corresponding Betti numbers to pinpoint sub-regions requiring correction; the Refinement Advisor, which generates detailed refinement instructions for each sub-region; and the Change Validator, which assesses refinement validity and determines whether to accept or regenerate the refinement instructions. Complementing these reasoning agents is a dedicated execution tool, TopoRefineNet, which performs instruction-guided editing of the segmentation mask. This closed-loop pipeline effectively leverages multi-scale features and operates iteratively until the segmentation satisfies the structural completeness of neurons. In summary, our contributions are threefold:

- To overcome the limitations of existing neuron segmentation approaches, we present NeuroRefiner—the first LLM-based multi-agent system that achieves topology-aware iterative optimization of segmentation results.
- To enable efficient collaboration between reasoning agents and specialized segmentation models, we design TopoRefineNet, a dedicated segmentation editing tool, together with a tailored two-stage training strategy that aligns model behavior with structural refinement instructions.
- Experiments on three benchmarks demonstrate that our method outperforms existing neuron segmentation and segmentation refinement approaches.

2 Related Work

Neuron Segmentation. Recent advancements have prioritized the design of specialized architectures to extract multi-scale features, aiming to simultaneously model global neuronal morphology and fine-grained structural details. Architecturally, this is typically realized through the integration of dedicated convolution [41, 43] or graph-based reasoning modules [35]. In terms of morphological constraints, methods such as SGSNet [44] and Tubular [37] augment feature representation via multi-task learning that predicts auxiliary geometric cues. To capture intrinsic morphological priors, MP-NRGAN [5] employs adversarial training utilizing weak supervision signals derived from reconstructed neurons. Furthermore, to mitigate the limitations of restricted receptive fields inherent in

patch-based inference, GBP-Net [42] encodes long-range contextual information and fuses it with intra-patch features.

Despite these advancements, prevailing models remain hindered by single-pass inference paradigms, which preclude iterative error correction. While certain approaches [3, 52] utilize point cloud networks to refine segmentation outputs, they fundamentally fail to recall omitted structural segments (false negatives) missed during the initial prediction. Conversely, our framework introduces a multi-round refinement mechanism guided by multi-scale features, effectively mitigating the errors in single-pass models.

Agent for Biomedical Image Segmentation. Existing agent-based approaches generally fall into two distinct paradigms. The first paradigm relies on conventional multi-agent reinforcement learning (MARL) for low-level, pixel-wise decision-making. Methods such as IteR-MRL [19] and BS-IRIS [24] employ multi-agent reinforcement learning for interactive 3D segmentation, treating voxels as collaborative agents to iteratively refine results. Similarly, dbMiM [6] employs MARL to optimize masking strategies for self-supervised neuron segmentation. However, these approaches are predominantly tailored for electron microscopy or general 3D modalities (e.g., MRI/CT), rendering them ineffective for handling neurons in fluorescence microscopy.

The second paradigm represents a more advanced, LLM-driven approach. These agents are LLM-centric and leverage planning-execution loops to invoke external segmentation tools. For instance, Ophiuchus [16] adopts a three-stage training strategy to coordinate vision tools like SAM2 [31] and BiomedParse [53] for enhanced lesion segmentation. IBISAgent [15] generates textual click instructions to invoke interactive segmentation tools for progressive mask refinement, while GenCellAgent [45] introduces a multi-agent framework that achieves intelligent tool routing via a planner-executor-evaluator loop. Although these methods have improved accuracy, they lack topology-aware agent design and tools specifically tailored for sparse neurons.

Segmentation Refinement. High-resolution image segmentation conventionally necessitates downsampling operations or patch-based inference strategies, which incur boundary artifacts or information loss. To mitigate these issues, a series of refinement methods has been proposed. Mask Transfiner [17] identifies incoherent regions to perform fine-grained label correction. SegRefiner [36] introduced a generic, iterative framework leveraging discrete diffusion processes to correct diverse segmentation errors. To achieve accurate segmentation of EM neurons, FGNet [18] refines the semantic features extracted by SAM2 [30] by training a lightweight fine-grained encoder. While effective for boundary regularization, existing methods neglect global topology, which is paramount for preserving filamentous neuronal structures. Hence, topology-aware refinement for neuronal structures remains a critical, underexplored frontier.

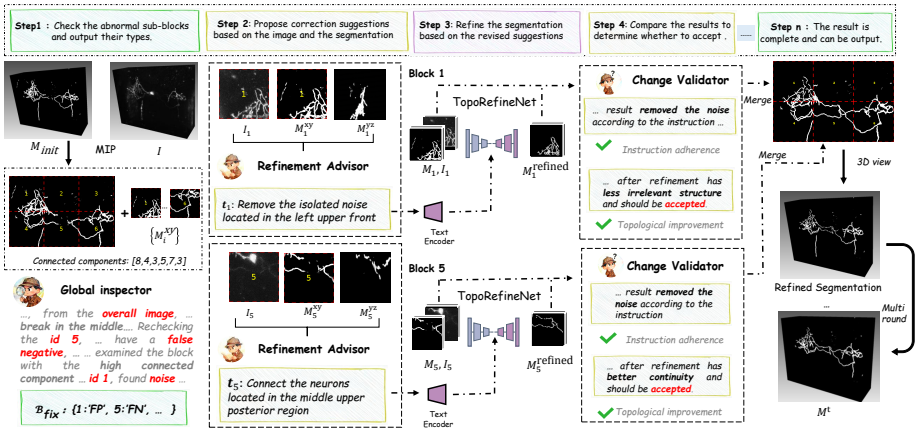

Fig. 2: Collaborative NeuroRefiner workflow. The system performs multi-round refinement via the interaction of agents and TopoRefineNet to obtain topologically complete segmentation results.

3 Method

Sec. 3.1 describes the collaborative mechanism among agents in NeuroRefiner. Sec. 3.2 details the design of each individual agent and Sec. 3.3 presents the architecture and training strategy of our proposed tool, TopoRefineNet.

3.1 NeuroRefiner

Manual refinement of neuron segmentation follows a hierarchical workflow in which experts iteratively alternate between global morphological assessment and fine-grained local editing. Motivated by this paradigm, we propose a multi-agent collaborative framework, as outlined in Algorithm 1. The segmentation mask $M^{(t)}$ is initialized using an off-the-shelf segmentation model. At each iteration, the Global Inspector conducts a holistic assessment of $M^{(t)}$ to detect sub-regions that violate topological priors. Subsequently, the Refinement Advisor generates a set of structure-aware refinement instructions $\{\mathbf{t}_i\}$, each tailored to a specific anomalous sub-region. These instructions guide TopoRefineNet to produce refined sub-region masks, denoted as $\{M_i^{\text{refined}}\}$. As a validation mechanism, the Change Validator acccepts only those refinements that yield an improvement over the original segmentation, as

$$M^{(t+1)} = \text{Merge}\left(M^{(t)}, \left\{M_i^{\text{refined}} \mid \text{Validator}(M_i^{\text{refined}}) = \text{True}\right\}\right). \quad (1)$$

As illustrated in Fig. 2, the process terminates when the Global Inspector determines that the neuronal morphology is complete or upon reaching $T_{\max}$ iterations. This iterative protocol ensures transparency, as the explicit diagnostic decisions and instructions generated at each step are auditable.

Algorithm 1 NeuroRefiner: Morphology-Aware Multi-Agent Refinement

Require: 3D image I, initial mask $M^{(0)}$, max iterations T_{max}, iteration $t \leftarrow 0$
while $t < T_{max}$ **do**
  Compute z-axis MIP M^{xy}, sub-blocks $\{b_i\}$, and Betti numbers $\{\beta_0(b_i)\}$
  $B_{fix}, \{\tau_i\} \leftarrow \pi_{ins}(M^{xy}, \{M_i^{xy}\}, \{\beta_0(b_i)\})$ // Identify blocks to fix
  if $B_{fix} = \varnothing$ then **break** //morphologically converged
  for each block $b_i \in B_{fix}$ **do**
    Initialize retry counter $k \leftarrow 0$
    while $k < T_{max}$ **do**
      $t_i \leftarrow \pi_{adv}(I_i^{xy}, I_i^{yz}, M_i^{xy}, M_i^{yz}, \tau_i)$ // Generate refinement instruction
      $M_i^{refined} \leftarrow$ TopoRefineNet(I_i, M_i, t_i) // Instruction-guided editing
      $r_i \leftarrow \pi_{val}(M_i^{xy}, M_i^{\text{refined}}, t_i, \tau_i)$ // Validate refinement quality
      if r_i is Accept **then**
        $M[b_i] \leftarrow M_i^{\text{refined}}$, break //Update local mask
      else
        $k \leftarrow k + 1$ //Retry
      end if
    end while
  end for
  $t \leftarrow t + 1$
end while
return $M^{\text{final}} \leftarrow M^{(t)}$

3.2 Design of Agents in NeuroRefiner

Global Inspector leverages the high discriminability of global morphological anomalies to identify sub-regions violating topological priors. Given an initial 3D segmentation mask M_{init} and the corresponding image I, we first apply Maximum Intensity Projection (MIP) along the z-axis to obtain projection M^{xy} and I^{xy}. Performing MIP on sparse volumes facilitates their input into VLMs while inducing minimal signal overlap. The projected volume is then partitioned into a set of non-overlapping 2D sub-blocks $\mathcal{B} = \{b_i\}_{i=1}^{N}$, yielding corresponding image and mask patches $\{I_i^{xy}\}$ and $\{M_i^{xy}\}$.

As a crucial metric for neuron segmentation, topological correctness serves as an effective guide for identifying erroneous regions. For each sub-block, the 0th-order Betti number is computed to quantify topological irregularities as $\beta_0(b_i) = CC(M_i)$, where $CC(\cdot)$ denotes the connected component counting function in the corresponding 3D region. An elevated β_0 value, which indicates a greater number of isolated fragments within a sub-block, suggests a higher likelihood of topological errors. Based on this topological metric and the mask, the inspection policy π_{ins} identifies abnormal regions and categorizes their error types:

$$\mathrm{B}_{\text{fix}}, \tau_i = \pi_{\text{ins}}(M^{xy}, \{M_i^{\text{xy}}\}, \beta_0(b_i)), \tag{2}$$

where $\mathrm{B}_{fix} \subseteq \mathcal{B}$ denotes the subset of blocks requiring refinement and τ_i labels each block as either false-positive or false-negative. They provide guidance for the Refinement Advisor to generate detailed instructions.

Refinement Advisor generates structured, morphology-aware instructions to guide the correction of identified sub-blocks. To precisely delineate the abnormal region along the z-axis, the agent first applies MIP to the 3D region corresponding to b_i along the x-axis and then localizes the defect in the yz-plane. Upon obtaining the xy and yz plane projections of the target region, it jointly analyzes the image and mask from orthogonal views to uncover connectivity cues and formulate a refinement instruction, as

$$\mathbf{t}_i = \pi_{\text{adv}}\big(I_i^{xy}, I_i^{yz}, M_i^{xy}, M_i^{yz}, \tau_i\big). \tag{3}$$

Each instruction $\mathbf{t}_i$ explicitly encodes both a spatial location (e.g., "upper-left-posterior") and an editing operation (e.g., "connect" or "remove"). The editing module TopoRefineNet subsequently performs refinement based on the instruction and outputs M_i^{refined}. This design decouples reasoning from voxel-level operations by using natural language as an intermediary, thereby fully unleashing the agent's reasoning and expressive capabilities.

Change Validator ensures the reliability of iterative refinement by evaluating the output of TopoRefineNet and preventing erroneous edits from being merged. Specifically, it compares the initial mask M_i against the refined result M_i^{refined}, conditioned on the instruction $\mathbf{t}_i$ and the error type τ_i to assess whether the edit is beneficial to morphology. This evaluation is formalized as

$$r_i = \pi_{\text{val}}\big(M_i^{xy}, M_i^{\text{refined}}, \mathbf{t}_i, \tau_i\big). \tag{4}$$

The Change Validator only approves an edit when two criteria are satisfied: (1) instruction adherence—the modification correctly implements the spatial and operational intent of $\mathbf{t}_i$, and (2) topological improvement—the refined mask exhibits reduced fragmentation or noise without introducing new artifacts. If either condition fails, it triggers the refinement advisor to generate a new instruction, thereby closing the loop with corrective feedback. For cases that remain unresolved after multiple refinement cycles, the system escalates the sub-block to human experts for manual intervention.

3.3 TopoRefineNet

Architecture. Serving as the critical bridge between high-level agent reasoning and low-level pixel operations, TopoRefineNet executes the conversion of topological instructions into voxel-level mask modifications. Inspired by recent advances in multimodal conditional generation [8,33,48], TopoRefineNet formulates segmentation refinement as an instruction-guided image editing task. As illustrated in Fig. 3 (a), it takes the image I_i and the noisy mask $M_i \in \mathrm{R}^{H\times W\times D}$ as input, which are concatenated and encoded by a visual encoder $\mathcal{E}_\theta$ to produce a hierarchical feature pyramid, as $\{\mathbf{F}_i^v\}_{i=1}^L = \mathcal{E}_\theta([I_i; M_i])$. The instruction $\mathbf{t_i}$ is embedded via a frozen pre-trained text encoder, as $\mathbf{F}^t = \mathcal{T}_\phi(\mathbf{t}) \in \mathrm{R}^d$. The interaction is achieved via cross-attention between the deepest visual features $\mathbf{F}_L^v$ and $\mathbf{F}^t$, formulated as

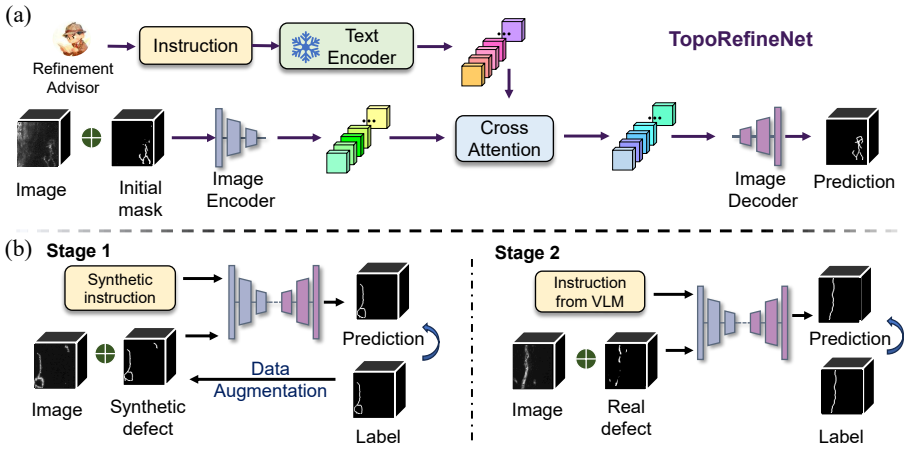

Fig. 3: (a) Architecture of TopoRefineNet. The image and instruction are encoded by their respective encoders, fused via cross-attention, and decoded to generate the corrected mask. (b) The proposed two-stage training paradigm. The first stage is trained on synthetic defects, while the second stage utilizes real masks with defects.

$$\mathbf{F}_L'^v = \text{CrossAttn}(\mathbf{F}_L^v, \mathbf{F}^t) = \text{softmax}\left(\frac{\mathbf{W}_q\mathbf{F}_L^v(\mathbf{W}_k\mathbf{F}^t)^\top}{\sqrt{d_k}}\right)\mathbf{W}_v\mathbf{F}^t, \tag{5}$$

where $\mathbf{W}_q, \mathbf{W}_k, \mathbf{W}_v$ are learnable projection matrices and d_k denotes the feature dimension. This mechanism dynamically recalibrates visual features by attending to regions and editing operations specified in the instruction. The fused feature $\mathbf{F}_L'^v$ is passed to the decoder $\mathcal{D}$, which integrates it with shallow features via skip connections and upsamples to yield the refined mask M_{refined}. The entire process is concisely expressed as a conditional generation function, as

$$M_{\text{refined}} = \mathcal{F}_\theta(I_i, M_{\text{i}}, \mathbf{t_i}) = \mathcal{D} \circ \text{CrossAttn}\big(\mathcal{E}_\theta([I_i; M_{\text{i}}]), \mathcal{T}_\phi(\mathbf{t_i})\big). \tag{6}$$

This design addresses the limitations of general tools in editing sparse neuronal structures, empowering the system with essential capabilities for fine-grained morphological refinement.

Training Strategy. To endow TopoRefineNet with precise instruction-to-voxel editing capability, we devise a two-stage training strategy as shown in Fig. 3 (b). The initial stage prioritizes cross-modal alignment using synthetic data with morphological defects. Specifically, ground-truth masks M_{gt} are perturbed via data augmentation to yield corrupted variants (Sec. 4.1), with corresponding refinement instructions generated based on the location and type of perturbation. The second stage subsequently shifts focus toward real-world generalization. We curate segmentation outputs from multiple networks trained on diverse neuronal datasets and specifically select instances that exhibit representative topological

errors. For each identified error, a refinement instruction is synthesized using Qwen3-VL [1]. By progressively increasing defect complexity from synthetic scenarios to complex real-world cases, this scheme stabilizes optimization and enhances robustness against the diverse instructions encountered during inference.

4 Experiments

4.1 Experiments Setup

Dataset. To evaluate NeuroRefiner across diverse neuronal morphologies and imaging conditions, we conducted experiments on three representative benchmarks, adhering to the training–testing splits established in prior work [22,42,44]. The BigNeuron [28] dataset comprises expert-annotated neurons spanning multiple species. Images were acquired across different laboratories and imaging platforms, exhibiting substantial variations in appearance, resolution, and scale. The CWMBS [22] dataset is derived from whole-brain mouse imaging and contains volumes of size 256 × 256 × 256 voxels (physical resolution: 0.2µm × 0.2µm × 1µm), of which 83 exhibit strong background noise and 162 feature thin filamentary structures with weak signals. The ZBFWB dataset [9] consists of confocal microscopy images from 6-day-old zebrafish larvae (1000 × 2000 × 250 voxels; 0.5µm × 0.5µm × 1µm), presenting challenges due to its low contrast, long-range axonal projections, intricate network topologies, and low contrast.

Metrics. To ensure consistency with prior work [22, 42, 44], we use reconstructions generated by APP2 [39] to assess segmentation quality from point-level accuracy to topological completeness. We compute the precision, recall, and F1-score to evaluate voxel-level accuracy and further employ the distance-based metric Spatial Distance (SD) to measure the average bidirectional distance between prediction and ground-truth structures. Its variant SSD [29], excludes matched points within 2 voxels to focus on significant topological errors. Finally, MES [40] evaluates the overall morphological fidelity of topology by quantifying the lengths of missing and extra structures. Lower SD and SSD values, along with higher values for all other metrics, indicate superior reconstruction.

Details. In our experiments, all agents use Qwen3-VL-8B [1] as the foundation model, and text embeddings are obtained from Qwen3 Embedding 0.6B [50]. The Global Inspector performs error detection using a region of 128 × 128 pixels as its fundamental unit. The maximum number of iterations $T_{\max}$ is set to 5. The prompts for all agents and the template of instructions for TopoRefineNet are provided in the supplementary material.

Both training stages of TopoRefineNet employ cross-entropy loss and dice loss [26]. In the first stage, to synthesize FN segmentations, we randomly select a positive sample as the center and perform an erosion operation with a random kernel size $k \in [5, 20]$ to generate breaks of varying lengths. For FP samples,

Table 1: Quantitative results (%) on the CWMBS dataset. The best results are **bolded** and the second-best results are underlined in the following table.

Models	Weak Signal						Strong Noise					
	SD↓	SSD↓	PRE↑	REC↑	F1↑	MES↑	SD↓	SSD↓	PRE↑	REC↑	F1↑	MES↑
3D Segmentation Baselines												
3D U-Net	28.35	33.71	93.85	62.37	74.94	46.21	24.98	31.61	94.16	69.57	79.23	54.30
SwinU	24.94	31.14	94.07	65.07	76.93	56.45	19.43	27.17	93.71	74.40	80.98	56.30
nnUNet	25.97	31.66	95.03	68.82	79.83	54.76	33.65	39.89	90.38	61.43	73.14	52.19
Neuron-tailored architectures												
V-Net	23.80	29.97	93.58	71.12	80.82	55.47	34.96	40.17	88.16	56.34	68.75	46.95
SGSNet	35.70	40.82	92.44	55.63	69.46	50.29	19.59	26.79	94.79	73.13	82.56	58.25
GBP-Net	11.85	19.27	95.06	83.50	88.91	67.09	17.15	24.94	93.41	75.93	83.77	59.41
ADTL-Net	25.96	31.65	93.49	62.72	75.07	53.51	28.11	35.18	95.24	60.68	74.13	52.79
Refinement approaches on 3D UNet												
Transfiner	27.33	32.14	94.10	62.28	74.95	46.22	24.53	31.52	94.33	70.22	80.51	56.28
SegFix	26.74	32.09	93.00	64.72	76.32	48.93	23.22	30.18	**95.39**	70.11	80.82	57.52
SegRefiner	17.61	23.79	94.92	77.84	85.54	62.31	21.64	27.80	92.19	72.84	81.38	57.91
Ours	9.17	15.21	95.72	85.05	90.07	70.32	**13.29**	**19.52**	94.31	**79.66**	**86.37**	**64.39**
Refinement approaches on nnUNet												
Transfiner	25.34	31.73	95.63	66.29	78.30	53.16	33.24	39.70	91.69	63.79	75.24	53.27
SegFix	24.82	30.57	95.47	69.69	80.57	56.79	30.09	35.35	92.74	61.02	73.61	55.38
SegRefiner	15.98	22.60	96.71	73.46	83.50	61.41	22.53	27.83	94.35	66.37	77.92	54.93
Ours	**7.26**	**11.86**	**96.79**	**87.39**	**91.85**	**72.60**	15.96	20.17	93.76	77.47	84.84	61.82

we either randomly copy-paste annotations of other neurons or introduce Gaussian noise within local regions. In the second stage, we trained several common architectures—namely, 3D U-Net [7], V-Net [21], UNETR [12], nnFormer [54], and SwinUNETR [11]—on a subset of the training set, followed by inference on the complete training set. Based on connected component and F1 scores, we selected a total of 5,310 volumes of $128 \times 128 \times 64$ voxels exhibiting significant segmentation errors for training.

4.2 Quantitative Results

We comprehensively evaluated NeuroRefiner against a diverse set of state-of-the-art (SOTA) methods for fluorescence microscopy neuron segmentation. The baselines include 3D U-Net [7], SwinUNETR [11], and nnUNet [13]; neuron-tailored architectures SGSNet [44], GBP-Net [42], Improved V-Net [21], and ADTL-Net [22]; refinement-based methods Mask Transfiner [17], SegFix [46], and SegRefiner [36]. We reproduced the corresponding 3D versions of the 2D-based refinement-based methods using the same training data as TopoRefineNet.

Table 2: Quantitative results (%) on the BigNeuron and ZBFWB datasets.

	Method	SD↓	SSD↓	PRE↑	REC↑	F1↑	MES↑
BigNeuron	3D U-Net [7]	7.23	11.27	83.15	89.76	86.33	66.82
	nnUNet [13]	5.16	8.68	87.34	86.87	87.10	68.72
	Improved V-Net [21]	7.40	11.16	82.50	**90.41**	86.27	67.16
	SGSNet [44]	5.90	9.34	86.22	90.28	88.20	69.90
	GBP-Net [42]	4.89	8.41	88.83	89.23	89.03	70.77
	ADTL-Net [22]	5.12	8.39	88.25	87.14	87.69	69.32
	Mask Transfiner [17]	6.82	10.57	86.14	88.89	87.49	69.01
	SegFix [46]	7.18	11.35	83.92	89.19	86.47	66.32
	SegRefiner [36]	5.79	9.94	87.18	89.23	88.19	69.27
	Ours	**3.39**	**6.94**	**89.74**	90.37	**90.05**	**72.02**
ZBFWB	3D U-Net [7]	38.65	45.41	86.33	63.15	72.94	53.63
	nnUNet [13]	27.94	34.40	86.40	67.60	75.85	61.78
	Improved V-Net [21]	42.43	49.06	84.27	58.16	68.82	51.45
	SGSNet [44]	44.34	50.29	82.54	55.36	66.27	48.14
	GBP-Net [42]	20.35	28.44	91.13	76.39	83.11	71.41
	ADTL-Net [22]	46.74	54.87	82.53	52.77	64.38	48.79
	Mask Transfiner [17]	36.23	43.12	88.26	63.92	74.14	54.87
	SegFix [46]	36.42	42.97	87.21	62.94	73.11	55.96
	SegRefiner [36]	30.18	36.25	88.25	65.19	74.99	60.32
	Ours	**16.31**	**23.66**	**92.77**	**80.38**	**86.13**	**74.25**

Tab. 1 reports quantitative results on the CWMBS subsets. CNN-based models (3D U-Net, SGSNet) are hampered by local receptive fields, limiting generalization under distribution shifts, whereas multiscale methods (GBP-Net, SwinUNETR) demonstrate superior accuracy. Due to a lack of topology-centric refinement, Transfiner and SegFix yield only marginal gains on the masks of 3D U-Net and nnUNet. In contrast, our approach leverages agent-based reasoning and topology-centric refinement and improves F1 scores of various segmentation masks by over 10% on the weak-signal subset, establishing SOTA results across both subsets.

Tab. 2 summarizes results on the BigNeuron and ZBFWB datasets, where all refinement approaches utilize 3D U-Net segmentation outputs as initialization. Benefiting from multi-step refinement, both SegRefiner and our approach boost F1 scores on the BigNeuron dataset. Specifically, our method improves MES by 5.20% and reduces SSD by 4.33 relative to the initial segmentation. On the more challenging ZBFWB dataset, conventional CNN-based methods struggle to handle low-contrast regions, resulting in fragmented segmentation and suppressed recall. Although GBP-Net leverages enhanced multi-scale features to surpass other methods, it remains unable to correct topological defects. Furthermore, we observe that all segmentation refinement approaches yield consistent improvements. Notably, our method elevates the F1 score of 3D U-Net by 13.19%, achieving an F1 score 3.02% higher than GBP-Net. These results validate the effectiveness of our multi-agent system in leveraging topological priors and global features for iterative refinement.

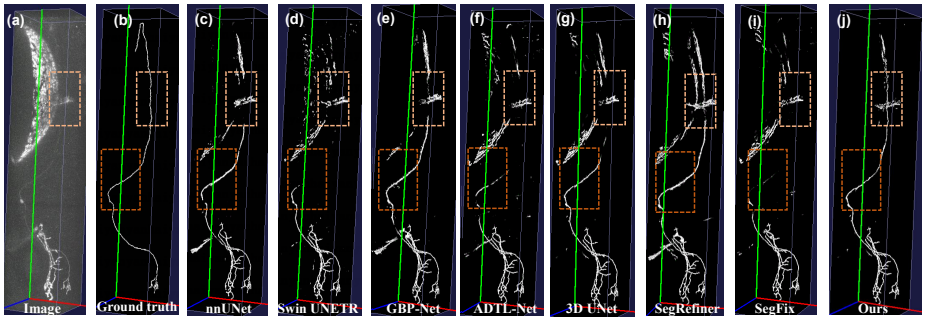

Fig. 4: Qualitative comparison of original images, ground truth, and segmentation results on the ZBFWB dataset. The orange boxes highlight regions with weak signals. Our method preserves the complete neuronal structure and yields superior performance.

Table 3: Ablation Study on Agents in NeuroRefiner.

Global Inspector	Refinement Advisor	Change Validator	ZBFWB SSD	ZBFWB F1 (%)	CWMBS SSD	CWMBS F1 (%)
			34.52	76.39	30.23	79.96
✓			32.05	77.29	28.47	80.34
✓	✓		25.24	83.47	18.93	87.14
✓		✓	30.69	81.43	24.53	83.12
✓	✓	✓	**23.66**	**86.13**	**16.65**	**88.83**

Fig. 4 depicts a neuron with long-range projections from the ZBFWB dataset, where irrelevant tissue signals significantly obscure the target structure. All single-pass segmentation methods yield results plagued by topological violations, characterized by extensive fragmentation and artifacts. Although refinement-based paradigms correct local errors in segmentation, they lack a neuron-specific error identification and correction protocol. In contrast, our method leverages agent-based reasoning and neuronal topological priors to guide the refinement with TopoRefineNet, achieving superior reconstructions.

4.3 Ablation Study

Ablation Study on Agents in NeuroRefiner. To validate the efficacy of the proposed multi-agent system, we conducted ablation studies on the ZBFWB and CWMBS datasets, summarized in Table 3. In the absence of agents, TopoRefineNet refines all blocks indiscriminately for T_{max} rounds. While this improves initial segmentation, the lack of global context and targeted guidance yields suboptimal results. Employing solely the Global Inspector for both error localization and instruction generation provides negligible gains, as a single agent struggles to manage these distinct tasks and lacks quality control mechanisms. The Refinement Advisor decouples instruction generation from inspection. This functional specialization yields F1 gains of 6.18% and 6.80%, respectively. Further incorporating the Change Validator to filter unreliable refinements yields

Table 4: Ablation Study on Foundation Model.

Method	CWMBS		ZBFWB	
	SSD	F1	SSD	F1
Intern-VL3-8B	17.70	87.94	28.11	83.81
Qwen2.5-VL-7B	18.33	87.14	27.31	84.77
Qwen3-VL-8B	**16.65**	**88.83**	**23.66**	**86.13**

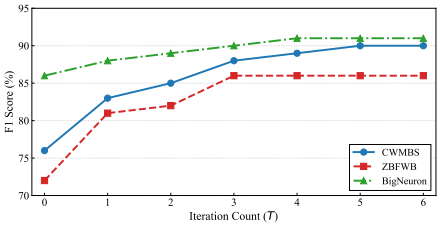

Fig. 5: Ablation Study on Iteration Numbers $T_{\max}$.

additional gains of 2.66% and 1.68%. Ultimately, the full system achieves optimal performance, confirming that synergistic collaboration is essential for robust, high-fidelity neuron segmentation refinement.

Fig. 5 illustrates the impact of the number of iterations. The results indicate that across all three datasets, the initial iteration yields substantial gains, whereas improvements diminish significantly after the fifth iteration. Consequently, we set the maximum iteration count $T_{max} = 5$. Notably, the more challenging ZBFWB and CWMBS datasets require more steps to converge and demonstrate significant improvement over the baseline segmentation.

Ablation Study on Foundation models. We evaluated NeuroRefiner using alternative open-source VLMs. As shown in Tab. 4, while Qwen3-VL-8B achieves optimal performance due to its advanced reasoning capabilities, substituting it with weaker models like Qwen2.5-VL-7B [2] and Intern-VL3-8B [55] still yields significant improvements over the baseline (+4.82% F1 with Intern-VL3 on CWMBS). This confirms that the performance gain stems primarily from the multi-agent refinement framework and TopoRefineNet, rather than the specific foundation model.

Ablation Study on Global Inspector. We conduct ablation studies on the input configuration, as summarized in Tab. 5. When relying solely on M^{xy}, the inspector tends to overlook fragmentation artifacts, leading to marginal refinement gains. In contrast, providing the full sequence of patch-level projections alongside the global view enables it to jointly leverage holistic context and local cues. This significantly enhances the ability to pinpoint topological defects and improves the F1 score by 3.43% and 3.20%. Although the MIP of the segmentation results is sufficient to identify most topological anomalies in sparsely distributed neurons, a small number of complex volumes may still suffer from interference due to neuron overlap. Therefore, incorporating the 0-th Betti number β_0 as an auxiliary topological indicator effectively directs attention toward regions with high connected component counts. This configuration achieved the best performance across multiple datasets.

Table 5: Ablation Study on Global Inspector.

Inputs			ZBFWB		CWMBS	
M^{xy}	$\{M_i^{xy}\}$	$\beta_0(b_i)$	SSD	F1 (%)	SSD	F1 (%)
✓			32.86	77.36	27.11	81.89
✓	✓		26.27	80.79	21.47	85.09
✓		✓	29.16	82.41	25.93	82.64
✓	✓	✓	**23.66**	**86.13**	**16.65**	**88.83**

Table 6: Ablating TopoRefineNet.

Method	CWMBS		ZBFWB	
	SSD	F1	SSD	F1
AdaLN	19.37	86.46	26.78	84.38
CDP	17.89	88.07	**22.14**	86.07
CAB	**16.65**	**88.83**	23.66	**86.13**

Ablation Study on TopoRefineNet. We carry out ablation studies on the architecture and training strategies of TopoRefineNet. First, we compared three multimodal fusion strategies: channel-wise dot product (CDP) [32], cross-attention in the bottleneck (CAB), and AdaLN [27]. As shown in Tab. 6, since CDP and CAB yield comparable accuracy, we adopt the more efficient CAB. Second, we replace the text encoder with the larger Qwen3-4B. It yields no significant improvement, indicating that the 0.6B variant suffices for precise encoding of the structured instructions. To validate the two-stage curriculum learning strategy, we trained on real defects from scratch as a comparison. This single-stage variant exhibited a 23.57% increase in the rejection rate of Change Validator, alongside F1 score degradations of 1.17% and 1.89% in ZBFWB and CWMBS, respectively. The decline confirms the necessity of our training strategy.

5 Conclusion

In this paper, we present NeuroRefiner, the first LLM-based multi-agent system for neuron segmentation refinement. By coordinating morphology-aware global inspection, language-guided editing, and validation, our approach achieves progressive refinement without agent training. Consequently, it overcomes the limited topological fidelity and poor interpretability inherent in end-to-end segmentation models. The proposed TopoRefineNet bridges the gap between instructions and segmentation models, offering novel insights into agent–tool interaction design. Notably, our framework is agnostic to the foundation model, ensuring it evolves in tandem with advancements in VLMs.

Limitation While NeuroRefiner's 2D MIP detects topological anomalies in sparse neurons, it loses depth information in dense regions, hindering precise 3D error localization. Future work can introduce depth-aware encoding, which maps Z-axis depth to pseudo-color channels or generates depth-weighted projections, allowing the Global Inspector to resolve occlusions and accurately infer the Z-axis depth of topological defects.

Acknowledgement This work was supported by the Brain Science and Brain-like Intelligence Technology - National Science and Technology Major Project (2021ZD0204500, 2021ZD0204503 to L.L.), National Key Research and Development Program of China (2025YFA1614600 to J.L.).

References

1. Bai, S., Cai, Y., Chen, R., Chen, K., Chen, X., Cheng, Z., Deng, L., Ding, W., Gao, C., Ge, C., et al.: Qwen3-vl technical report. arXiv preprint arXiv:2511.21631 (2025)
2. Bai, S., Chen, K., Liu, X., Wang, J., Ge, W., Song, S., Dang, K., Wang, P., Wang, S., Tang, J., et al.: Qwen2. 5-vl technical report. arXiv preprint arXiv:2502.13923 (2025)
3. Balsiger, F., Soom, Y., Scheidegger, O., Reyes, M.: Learning shape representation on sparse point clouds for volumetric image segmentation (2019)
4. Chen, R., Liu, M., Chen, W., Wang, Y., Meijering, E.: Deep learning in mesoscale brain microscopy image analysis: A review. Computers in Biology and Medicine p. 107617 (2023)
5. Chen, X., Zhang, C., Zhao, J., Xiong, Z., Zha, Z.J., Wu, F.: Weakly supervised neuron reconstruction from optical microscopy images with morphological priors. IEEE Trans. Med. Imag. **40**(11), 3205–3216 (2021)
6. Chen, Y., Huang, W., Zhou, S., Chen, Q., Xiong, Z.: Self-supervised neuron segmentation with multi-agent reinforcement learning. arXiv preprint arXiv:2310.04148 (2023)
7. Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O.: 3d u-net: learning dense volumetric segmentation from sparse annotation. In: International conference on medical image computing and computer-assisted intervention. pp. 424–432. Springer (2016)
8. Couairon, G., Verbeek, J., Schwenk, H., Cord, M.: Diffedit: Diffusion-based semantic image editing with mask guidance. arXiv preprint arXiv:2210.11427 (2022)
9. Du, X., Yue, Z., Wei, J., Li, W., Chen, M., Chen, T., Hu, H., Ren, H., Jia, Z., Ning, X., et al.: Central nervous system atlas of larval zebrafish constructed using the morphology of single excitatory and inhibitory neurons. bioRxiv pp. 2025–06 (2025)
10. Gou, L., Wang, Y., Gao, L., Zhong, Y., Xie, L., Wang, H., Zha, X., Shao, Y., Xu, H., Xu, X., et al.: Gapr for large-scale collaborative single-neuron reconstruction. Nature Methods **21**(10), 1926–1935 (2024)
11. 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. Springer (2021)
12. 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: Proceedings of the IEEE/CVF winter conference on applications of computer vision. pp. 574–584 (2022)
13. 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)
14. Jiang, L., Lu, Y., Zhang, Y., Liu, J., Han, H.: Neuromamba: Multi-perspective feature interaction with visual mamba for neuron segmentation. arXiv preprint arXiv:2601.15929 (2026)
15. Jiang, Y., Li, Q., Xu, B., Sun, H., Ding, C., Dong, J., Cai, Y., Zhang, X., Yin, J.: Ibisagent: Reinforcing pixel-level visual reasoning in mllms for universal biomedical object referring and segmentation. arXiv preprint arXiv:2601.03054 (2026)
16. Jiang, Y., Zhang, Y., Zhang, P., Li, Y., Chen, J., Shi, X., Zhen, S.: Incentivizing tool-augmented thinking with images for medical image analysis. arXiv preprint arXiv:2512.14157 (2025)

17. Ke, L., Danelljan, M., Li, X., Tai, Y.W., Tang, C.K., Yu, F.: Mask transfiner for high-quality instance segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 4412–4421 (2022)
18. Li, Z., Chen, H., Sun, Z., Li, K., Hu, X.: Fgnet: Leveraging feature-guided attention to refine sam2 for 3d em neuron segmentation. arXiv preprint arXiv:2511.13063 (2025)
19. Liao, X., Li, W., Xu, Q., Wang, X., Jin, B., Zhang, X., Wang, Y., Zhang, Y.: Iteratively-refined interactive 3d medical image segmentation with multi-agent reinforcement learning. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 9394–9402 (2020)
20. Liu, C., Jiang, Y., Zheng, N.: Netracer: A topology-aware iterative tracing approach for tubular structure extraction. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 20593–20602 (2025)
21. Liu, M., Luo, H., Tan, Y., Wang, X., Chen, W.: Improved v-net based image segmentation for 3D neuron reconstruction. In: Proc. 2018 IEEE Int. Conf. Bioinf. Biomed. (BIBM). pp. 443–448. IEEE (2018)
22. Liu, M., Wu, S., Chen, R., Lin, Z., Wang, Y., Meijering, E.: Brain image segmentation for ultrascale neuron reconstruction via an adaptive dual-task learning network. IEEE Trans. Med. Imag. (2024)
23. Liu, Y., Wang, G., Ascoli, G.A., Zhou, J., Liu, L.: Neuron tracing from light microscopy images: Automation, deep learning and bench testing. Bioinformatics **38**(24), 5329–5339 (2022)
24. Ma, C., Xu, Q., Wang, X., Jin, B., Zhang, X., Wang, Y., Zhang, Y.: Boundary-aware supervoxel-level iteratively refined interactive 3d image segmentation with multi-agent reinforcement learning. IEEE Transactions on Medical Imaging **40**(10), 2563–2574 (2020)
25. Ma, J., Yang, Z., Kim, S., Chen, B., Baharoon, M., Fallahpour, A., Asakereh, R., Lyu, H., Wang, B.: Medsam2: Segment anything in 3d medical images and videos. arXiv preprint arXiv:2504.03600 (2025)
26. Milletari, F., Navab, N., Ahmadi, S.A.: V-net: Fully convolutional neural networks for volumetric medical image segmentation. In: Proc. 2016 4th Int. Conf. 3D Vis. (3DV). pp. 565–571. IEEE (2016)
27. Peebles, W., Xie, S.: Scalable diffusion models with transformers. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 4195–4205 (2023)
28. Peng, H., Hawrylycz, M., Roskams, J., Hill, S., Spruston, N., Meijering, E., Ascoli, G.A.: Bigneuron: Large-scale 3D neuron reconstruction from optical microscopy images. Neuron **87**(2), 252–256 (2015)
29. Peng, H., Ruan, Z., Long, F., Simpson, J.H., Myers, E.W.: V3d enables real-time 3D visualization and quantitative analysis of large-scale biological image data sets. Nature Biotechnol. **28**(4), 348–353 (2010)
30. Ravi, N., Gabeur, V., Hu, Y.T., Hu, R., Ryali, C., Ma, T., Khedr, H., Rädle, R., Rolland, C., Gustafson, L., et al.: Sam 2: Segment anything in images and videos. arXiv preprint arXiv:2408.00714 (2024)
31. Ravi, N., Gabeur, V., Hu, Y.T., Hu, R., Ryali, C., Ma, T., Khedr, H., Rdle, R., Rolland, C., Gustafson, L.: Sam 2: Segment anything in images and videos (2024)
32. Rokuss, M., Langenberg, M., Kirchhoff, Y., Isensee, F., Hamm, B., Ulrich, C., Regnery, S., Bauer, L., Katsigiannopulos, E., Norajitra, T., et al.: Voxtell: Free-text promptable universal 3d medical image segmentation. arXiv preprint arXiv:2511.11450 (2025)

33. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B.: High-resolution image synthesis with latent diffusion models. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 10684–10695 (2022)
34. 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)
35. Wang, H., Song, Y., Zhang, C., Yu, J., Liu, S., Pengy, H., Cai, W.: Single neuron segmentation using graph-based global reasoning with auxiliary skeleton loss from 3D optical microscope images. In: IEEE 18th Int. Symp. Biomed. Imaging (ISBI). pp. 934–938. IEEE (2021)
36. Wang, M., Ding, H., Liew, J.H., Liu, J., Zhao, Y., Wei, Y.: Segrefiner: Towards model-agnostic segmentation refinement with discrete diffusion process. arXiv preprint arXiv:2312.12425 (2023)
37. Wang, X., Liu, M., Wang, Y., Fan, J., Meijering, E.: A 3D tubular flux model for centerline extraction in neuron volumetric images. IEEE Trans. Med. Imag. **41**(5), 1069–1079 (2021)
38. Wang, Y., Lang, R., Li, R., Zhang, J.: Nrtr: Neuron reconstruction with transformer from 3D optical microscopy images. IEEE Trans. Med. Imag. (2023)
39. Xiao, H., Peng, H.: App2: Automatic tracing of 3D neuron morphology based on hierarchical pruning of a gray-weighted image distance-tree. Bioinformatics **29**(11), 1448–1454 (2013)
40. Xie, J., Zhao, T., Lee, T., Myers, E., Peng, H.: Anisotropic path searching for automatic neuron reconstruction. Med. Image Anal. **15**(5), 680–689 (2011)
41. Yan, H., Zhai, H., Guo, J., Li, L., Han, H.: Neurolink: Bridging weak signals in neuronal imaging with morphology learning. In: Int. Conf. Med. Image Comput. Comput.-Assist. Interv. (MICCAI). pp. 467–477. Springer (2024)
42. Yan, H., Zhang, Y., Li, Z., Guo, J., Zhai, H., Liu, J., Zhong, Y., Yuan, J., Shen, L., Li, L., et al.: Glancing beyond patch: Spatial contextual cues for 3d neuron segmentation. IEEE Transactions on Medical Imaging (2025)
43. Yang, B., Chen, W., Luo, H., Tan, Y., Liu, M., Wang, Y.: Neuron image segmentation via learning deep features and enhancing weak neuronal structures. IEEE J. Biomed. Health Informat. **25**(5), 1634–1645 (2020)
44. Yang, B., Liu, M., Wang, Y., Zhang, K., Meijering, E.: Structure-guided segmentation for 3D neuron reconstruction. IEEE Trans. Med. Imag. **41**(4), 903–914 (2021)
45. Yu, X., Yang, Y., Liu, Q., Du, Y., McSweeney, S., Lin, Y.: Gencellagent: Generalizable, training-free cellular image segmentation via large language model agents. arXiv preprint arXiv:2510.13896 (2025)
46. Yuan, Y., Xie, J., Chen, X., Wang, J.: Segfix: Model-agnostic boundary refinement for segmentation. In: European conference on computer vision. pp. 489–506. Springer (2020)
47. Zhang, L., Huang, L., Yuan, Z., Hang, Y., Zeng, Y., Li, K., Wang, L., Zeng, H., Chen, X., Zhang, H., et al.: Collaborative augmented reconstruction for scaled production of 3d neuron morphology in mouse and human brains. bioRxiv pp. 2023–10 (2023)
48. Zhang, L., Rao, A., Agrawala, M.: Adding conditional control to text-to-image diffusion models. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 3836–3847 (2023)
49. Zhang, Y., Guo, J., Zhai, H., Liu, J., Han, H.: Segneuron: 3d neuron instance segmentation in any em volume with a generalist model. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 589–600. Springer (2024)

50. Zhang, Y., Li, M., Long, D., Zhang, X., Lin, H., Yang, B., Xie, P., Yang, A., Liu, D., Lin, J., et al.: Qwen3 embedding: Advancing text embedding and reranking through foundation models. arXiv preprint arXiv:2506.05176 (2025)
51. Zhang, Y., Jiao, R.: Towards segment anything model (sam) for medical image segmentation: a survey. arXiv preprint arXiv:2305.03678 (2023)
52. Zhao, R., Wang, H., Zhang, C., Cai, W.: Pointneuron: 3d neuron reconstruction via geometry and topology learning of point clouds. In: Proceedings of the IEEE/CVF winter conference on applications of computer vision. pp. 5787–5797 (2023)
53. Zhao, T., Gu, Y., Yang, J., Usuyama, N., Lee, H.H., Naumann, T., Gao, J., Crabtree, A., Abel, J., Moung-Wen, C.: Biomedparse: a biomedical foundation model for image parsing of everything everywhere all at once (2024)
54. Zhou, H.Y., Guo, J., Zhang, Y., Han, X., Yu, L., Wang, L., Yu, Y.: nnformer: Volumetric medical image segmentation via a 3d transformer. IEEE transactions on image processing **32**, 4036–4045 (2023)
55. Zhu, J., Wang, W., Chen, Z., Liu, Z., Ye, S., Gu, L., Tian, H., Duan, Y., Su, W., Shao, J., et al.: Internvl3: Exploring advanced training and test-time recipes for open-source multimodal models. arXiv preprint arXiv:2504.10479 (2025)