Addressing Distribution Mismatch for Effective Semi-Supervised Neuron Segmentation

Yanchao Zhang[a,b,1], Hao Zhai[a,b,1], Jinyue Guo[a,c], Jing Liu[a,*], Qiwei Xie[d] and Hua Han[a,b,*]

[a]*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, 100190, China*

[b]*School of Future Technology, University of Chinese Academy of Sciences, Beijing, 101408, China*

[c]*School of Artificial Intellengace, University of Chinese Academy of Sciences, Beijing, 101408, China*

[d]*Research Base of Beijing Modern Manufacturing Development, Beijing University of Technology, Beijing China*

ARTICLE INFO

Keywords:
Neuron segmentation
Semi-supervised learning
Connectomics
Volume electron microscopy
Deep learning

ABSTRACT

Semi-supervised learning offers a cost-effective approach for neuron segmentation in electron microscopy (EM) volumes. This technique leverages extensive unlabeled data to regularize supervised training for more robust predictions of neuron affinities. However, the distribution mismatch between labeled and unlabeled datasets, arising from limited annotations and the diversity of neuronal patterns, impedes the generalization of semi-supervised models. In this paper, we develop a dual-level pipeline to address the inherent mismatch issue and enhance semi-supervised neuron segmentation. At the data level, we propose an unsupervised heuristic to select valuable sub-volumes as labeled data based on distribution similarity in a pretrained feature space, ensuring a representative coverage of neuronal structures. At the model level, we introduce an axial-through mixing strategy into anisotropic neuron segmentation and integrate it into a semi-supervised framework. Building on this, we establish cross-view consistency constraints through intra- and inter-mixing of labeled and unlabeled datasets, which facilitates the learning of shared semantics across distributions while avoiding ambiguity in neuron affinities. Extensive comparative experiments and ablation studies on publicly available datasets demonstrate the effectiveness of the proposed method across different EM modalities and spatial resolutions.

1. Introduction

Connectomics aims to gain a comprehensive understanding of brain mechanisms and intelligence by reconstructing and interpreting neural circuits at the synaptic level (Ngai, 2022). Currently, only volume electron microscopy (EM), with its nanoscale spatial resolution, supports imaging the intricate morphology of neurons and individual synapses in brain tissue (Peddie et al., 2022; Peddie and Collinson, 2014). However, this technique generates massive volumetric data ranging from several hundred terabytes to petabytes (Abbott et al., 2020; Shapson-Coe et al., 2024; Zheng et al., 2018; Dorkenwald et al., 2023; Scheffer et al., 2020), which calls for automated methods to accelerate data processing and analysis. Therefore, efficient and accurate neuron reconstruction from EM volumes has become critical to advance the digitalization of the complete atlas of the nervous system.

Recently, learning-based methods have been developed for neuron instance segmentation across various EM modalities and voxel resolutions (Lee et al., 2017; Januszewski et al., 2018; Lee et al., 2021; Funke et al., 2018; Sheridan et al., 2022). State-of-the-art methods typically start with convolutional neural networks (CNNs) to predict descriptors of neuronal membranes, followed by supervoxel extraction and agglomeration to achieve instance segmentation (Fig. 1(a)). Nevertheless, optimizing such networks faces two major challenges. Firstly, the massive EM data spanning different neural structures leads to diverse textures and patterns. Secondly, labeled data typically takes up a small fraction of the overall volume due to the high cost of expert annotations. For instance, in the reconstruction of a fly hemi-brain, the segmentation model was trained using 8 manually selected sub-volumes, which account for a mere 0.003% of the entire dataset (0.8GB out of 26TB) (Scheffer et al., 2020).

Semi-supervised learning holds immense potential for tackling these issues by leveraging large-scale unlabeled data to enhance model generalization. Related works in biomedical image segmentation have developed techniques,

*Corresponding author
liujing2016@ia.ac.cn (J. Liu); hua.han@ia.ac.cn (H. Han)
ORCID(s): 0000-0003-4713-4631 (H. Han)
[1]Equal contributation

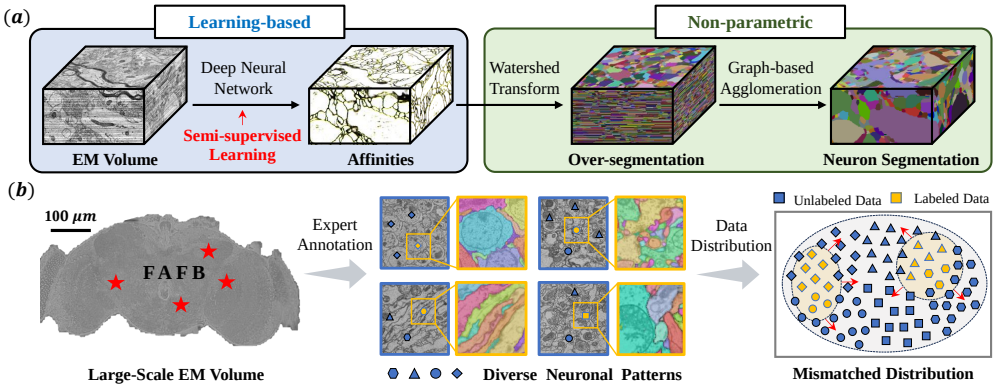

Figure 1: Motivation.(a) 3D neuron segmentation assigns unique labels to voxels within the same neurite. We aim to improve affinity predictions for enhanced instance segmentation by leveraging limited labeled data and extensive unlabeled data in a semi-supervised manner. (b) The vast volume of EM data across diverse brain regions exhibits complex textures and patterns. Limited by annotation budgets, the labeled data may inadequately reflect the distribution of unlabeled data. Such distribution mismatch results in sub-optimal semi-supervised learning and degradation in segmentation performance.

such as consistency learning (Huang et al., 2022; Luo et al., 2021; Yu et al., 2019) and self-training (Wang et al., 2022b; Shi et al., 2021), to regularize supervised training on insufficient labeled data, aiming to reduce the performance gap between fully supervised and semi-supervised methods. Note that these studies typically assume that the labeled and unlabeled data are drawn from the same distribution, thereby enabling network optimization with the aid of unlabeled data. However, ensuring matched distribution can be challenging in practical scenarios, and it is particularly prominent in dense neuron segmentation because of the intricate nature of neural structures, the large-scale database, and limited annotation budgets. The mismatch problem, as shown in Fig. 1(b), may lead to the discarding of knowledge learned from limited labeled data, thereby hindering the generalization of semi-supervised models.

From the data level, this obstacle can be largely overcome if limited labeled data is carefully selected to reflect the overall data distribution. Unfortunately, there is a lack of research on the targeted annotation of volumetric sub-regions within unlabeled EM data. The challenges lie in two aspects: (1) Volumetric data. Extracting representative sub-regions with flexible sizes from continuously distributed EM volume requires data discretization and customized evaluation criteria. In contrast, selective annotation of discrete instances (*e.g.*, ImageNet (Russakovsky et al., 2015)) can be achieved directly through diversity sampling (Wang et al., 2022a; Mannix and Bondell, 2023). (2) Unsupervised scene. Selecting the most representative data for labeling becomes challenging in the absence of initial labeled data, as it is difficult to assess the potential value of raw EM data (Zheng et al., 2019; Jin et al., 2022).

From the model level, promoting the mutual transfer of information between labeled and unlabeled data in semi-supervised training can further narrow the distribution gap by encouraging the network to learn shared semantics (Bai et al., 2023; Wang et al., 2019). However, existing efforts in semi-supervised neuron segmentation treat labeled and unlabeled data separately, leading to asymmetry in knowledge extraction and feature learning. Establishing a connection between labeled and unlabeled EM data within a semi-supervised framework poses another challenge in addressing the mismatch problem in neuron segmentation.

This paper introduces a novel pipeline for semi-supervised neuron segmentation. The core concept is that alleviating the distribution mismatch between labeled and unlabeled EM data facilitates semi-supervised learning. We begin by conducting self-supervised learning on the unlabeled EM dataset to develop an embedding network that provides meaningful representations for local patches. Based on this, we present a quantitative heuristic to identify representative sub-volumes comprising spatially associated patches with a coverage-based criterion. These sub-volumes can effectively reflect the unlabeled distribution even within a limited annotation budget, thus establishing a robust data foundation for semi-supervised learning. Afterward, we train a segmentation model using selectively labeled sub-volumes and the unlabeled data in a semi-supervised manner. By extending CutMix (Yun et al., 2019; French et al., 2019) to anisotropic neuron segmentation, we conduct inter- and intra-dataset mixing to connect labeled and unlabeled data by enforcing cross-view consistency while avoiding ambiguity in neuron affinities. This two-stage approach effectively reduces the distribution gap from both the data and model levels, demonstrating practical value in large-scale neuron reconstruction. Our contributions are summarized as follows:

1. We present an improved pipeline for semi-supervised neuron segmentation that effectively addresses the mismatch between labeled and unlabeled distributions at both data and model levels.
2. We introduce an unsupervised heuristic to identify valuable sub-volumes within EM volumes, ensuring a fine representation of the overall data distribution and various neuronal structures under a limited annotation budget.
3. We make the first attempt to integrate spatial mixing into semi-supervised neuron segmentation, which establishes a connection between labeled and unlabeled datasets and enables cross-view consistency regularization through inter- and intra-dataset mixing.
4. We validate the effectiveness of the proposed pipeline on representative EM datasets and large-scale scenarios, offering novel insights for cost-effective neuron segmentation.

2. Related work

2.1. Neuron Segmentation

As an instance segmentation problem, neuron reconstruction aims to assign a unique label to every voxel in the EM volumes that belongs to the same neuron. This task is challenging due to the unprecedented scale of EM data, the varied textures of neuronal structures, and the scarcity of expert annotations. State-of-the-art methods can be sketchily categorized as boundary-based and object-based. Specifically, boundary-based approaches first use 3D U-Net (Çiçek et al., 2016) variants to predict descriptors for neuron boundaries (*e.g.*, voxel affinity graphs, dense voxel embeddings (Lee et al., 2021), and local shape descriptors (Sheridan et al., 2022)). After over-segmentation via watershed transform, graph-based agglomeration is adopted to group supervoxels for instance segmentation (Beier et al., 2017; Funke et al., 2018; Li et al., 2024). In contrast, object-based approaches, as seen in (Januszewski et al., 2018; Meirovitch et al., 2019; Schmidt et al., 2024), extend the segmented area from the seed points to complete neurites iteratively. While effective, large-scale neuron reconstructions still require years of manual proofreading before they can be used for connectivity analysis (Zheng et al., 2018). Pursuing higher accuracy in the automatic segmentation stage plays a vital role in cost-effective connectomics.

In this study, we employ a boundary-based pipeline that starts with a 3D network to predict voxel affinities along the x, y, and z axes. As in Fig. 1(a), we aim to improve affinity predictions by leveraging both limited labeled data and extensive unlabeled data in a semi-supervised manner, ultimately achieving better instance segmentation results and reducing proofreading costs.

2.2. Semi-supervised Segmentation

Labeled data often constitutes a tiny fraction of the total volume in practical connectomics reconstruction. Semi-supervised learning (SSL) is a promising method to address limited annotation by leveraging unlabeled data to improve model generalization. Self-training and consistency regularization are notable techniques for generating supervision signals from unlabeled data. Self-training typically involves assigning pseudo labels to unlabeled data and combining them with the labeled for iterative model retraining (Wang et al., 2022b; Shi et al., 2021; Yang et al., 2022). In contrast, the premise of consistency regularization is that the prediction of unlabeled data should remain consistent under various perturbations (Tarvainen and Valpola, 2017; Yu et al., 2019; Luo et al., 2021; Wu et al., 2021). SSNS-Net (Huang et al., 2022) makes the first effort in semi-supervised neuron segmentation using consistency learning and validates the effectiveness of specific perturbations for EM data, highlighting the potential of semi-supervised techniques in this field. Notably, SSL is based on the assumption that labeled and unlabeled data are sampled from the identical distribution (Kurakin et al., 2019; Berthelot et al., 2019). However, distribution mismatch between labeled and unlabeled data is prominent in semi-supervised neuron segmentation due to the complex neural structures and limited budgets. Still, it has received little attention in the research community.

In this paper, we aim to address the distribution mismatch issue inherent in semi-supervised neuron segmentation. At the data level, we perform unsupervised selective labeling on unlabeled EM volumes for representative sub-volumes as labeled data. At the model level, we introduce intra- and inter-dataset mixing to promote shared feature learning within a semi-supervised framework. Both strategies reduce the distribution gap between labeled and unlabeled datasets, paving the way for more effective semi-supervised learning.

2.3. Selective Labeling

Selective Labeling (SL), the focus of active learning, seeks to annotate a valuable subset of an unlabeled dataset to optimize supervised learning under limited budgets. Active learning can be divided into two categories based on

whether or not the data are selected iteratively: ITerative Active Learning (ITAL) and One-Shot Active Learning (OSAL).

ITAL has received significant attention in prior research. It typically involves the iterative selection and labeling of valuable samples under uncertainty- (Beluch et al., 2018; Konyushkova et al., 2019; Wen et al., 2018) or diversity- (Yang et al., 2017; Lin et al., 2020; Sener and Savarese, 2018) based strategies, accompanied by continual network fine-tuning as more labeled data become available. However, ITAL rarely discusses which samples to choose initially when no trained model is available. Worse still, the iterative training, inference, and annotation procedures can be computationally expensive and time-consuming, particularly in large-scale neuron segmentation. In contrast, OSAL selects all valuable samples simultaneously before network training. Specifically, OSAL algorithms first measure unlabeled distribution via unsupervised representation learning, which projects raw data onto a low-dimensional embedding space. Then, customized selection strategies (Zheng et al., 2019; Jin et al., 2022; Wang et al., 2022a; Kolluru et al., 2021; Moenning and Dodgson, 2003; Lou et al., 2022) are proposed to identify representative and informative instances as labeled data. Unfortunately, limited attention has been given to the selective labeling of volumetric data including EM volumes. Due to the continuity and significant variance in the size of neuronal structures, labeled data in practical neuron reconstruction are typically sub-volumes with flexible sizes, from which local patches are randomly cropped for model training. Nevertheless, existing OSAL algorithms operate on patch-level discrete instances for selective labeling. Those annotations lack flexibility in size and fail to reflect the continuous changes in neuron boundaries, resulting in performance degradation of the segmentation network.

In this work, we propose a novel OSAL approach for EM volumes. This unsupervised algorithm allows for selecting valuable sub-regions with flexible sizes from the continuous unlabeled volume, which ensures a better representation of the overall data distribution even within a limited annotation budget.

3. Method

We seek to address the distribution mismatch between labeled and unlabeled datasets to improve semi-supervised neuron segmentation. The proposed approach highlights two primary components: selective labeling of valuable sub-volumes (Sec. 3.1) and semi-supervised training via mixed-view consistency regularization (Sec. 3.2), trying to tackle the mismatch issue from data and model levels, respectively.

3.1. Unsupervised Selective Labeling for EM Sub-volumes

Given an unlabeled dataset D_u and a labeling budget, the task is to identify sub-volumes that effectively represent the entire dataset as labeled data for semi-supervised training. As shown in Fig. 2(a), we devise a two-step approach involving patch representations and coverage-based sub-volume selection.

3.1.1. Unsupervised Representation Learning for EM Patches

To begin with, we perform self-supervised learning to develop an embedding network that can generate low-dimensional and semantic-rich representations for 3D EM patches. Specifically, the framework, constructed with Siamese networks (Chen and He, 2021), is optimized by maximizing the similarity between two variations of one volumetric image in the learned feature space, as depicted in Fig. 2(b). We cautiously employ perturbation strategies to augment EM inputs without substantially altering the semantic content, including reflections, rotations, Gaussian blurring, Gaussian noise, simulated misalignment, photometric perturbation, and randomly masked pixels.

During training, each input $u \in \mathbb{R}^{D\times H\times W}$ is sampled from unlabeled dataset D_u and randomly augmented to produce two views $\widetilde{u}_1$ and $\widetilde{u}_2$. These views are then fed into Siamese networks, which comprise two weight-sharing encoders (denoted as f), two weight-sharing projectors (denoted as g), and a predictor (denoted as h). The networks' parameters are updated by minimizing a symmetrized loss function given by

$$\mathcal{L}_{Sim} = \mathcal{N}(h(g(f(\widetilde{u}_1))), \underline{g(f(\widetilde{u}_2))}), \tag{1}$$

where $\mathcal{N}(A, B)$ represent the negative cosine similarity between A and B. The <u>underlined</u> items are treated as constants and not involved in gradient calculation. The trained encoder, referred to as $f(\cdot)$, then serves as an embedding network that projects EM patches onto an L-dimensional hyperspace.

3.1.2. Coverage-based Sub-volume Selection

With the embedding network for local patches, this subsection proposed a coverage-based heuristic to select representative and informative sub-volumes within the allocated budget.

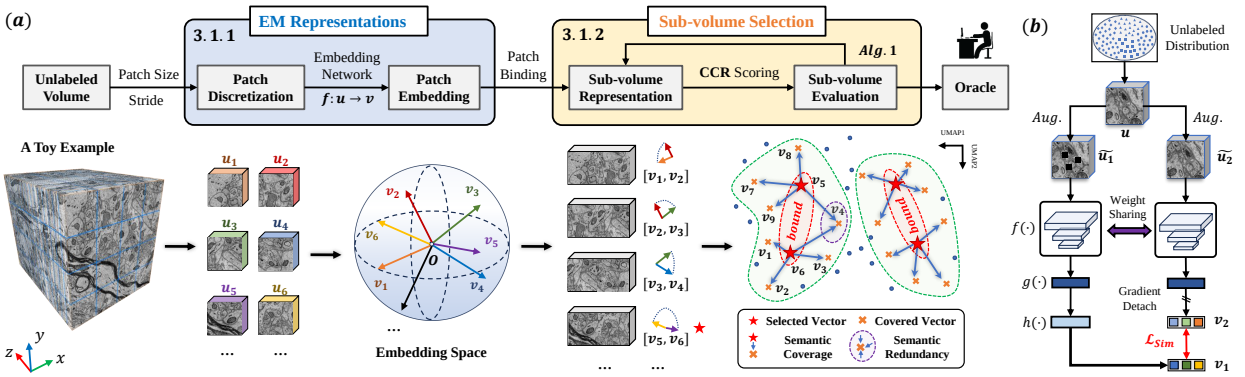

Figure 2: Selective labeling. (a) After patch discretization (presented as non-overlapping for clarity), we employ an unsupervised trained network $f(\cdot)$ to map unlabeled data to the embedding space for EM representation. Then a quantitative heuristic (*i.e.*, CGS) is designed to identify representative sub-volumes as labeled data iteratively. Well-selected sub-volumes, represented by bound embedding vectors, can effectively reflect the overall data distribution with typical and diverse patches. (b) The embedding network $f(\cdot)$ for 3D EM patches is optimized via contrastive learning.

For clarity, we highlight the difference between a patch and a sub-volume. A patch serves as the smallest unit for representation learning, whereas a sub-volume represents an arbitrarily sized sub-region cropped from the EM volume. The sub-volume annotation technique is customized for neuron segmentation. It addresses the limitations of patch-level annotation, such as the lack of size flexibility and the inability to capture the morphology of continuously distributed neural structures. Specifically, We employ spatially neighboring patches to jointly represent each candidate sub-volume. This enables us to utilize patch embeddings to represent sub-volumes directly in feature space. Well-selected sub-volumes covering typical and diverse patches can provide a comprehensive reflection of the overall data distribution.

By discretizing the unlabeled volume(s) into local patches via sliding window inference, we can leverage the embedding network to measure the unlabeled distribution. Specifically, the unlabeled data can be embedded as the *universal set* $U : \{v_n = f(u_n)\}_{n=1}^{N}$, where $u_n \in \mathbb{R}^L$ is the output of the embedding network $f(\cdot)$ applied to a 3D input patch u_n. Depending on the stride of patch discretization, this measurement approach can be readily extended to datasets of different scales. The above process maps the continuously distributed data to discrete vectors on a learned feature space. On this basis, candidate sub-volumes that are composed of patches with spatial constraints can be mapped to the *selected set* $S : \{v_m | v_m \in U\}_{m=1}^{M}$ by extracting corresponding vectors from U, as depicted in Fig. 2(a) sub-volume representation.

Intuitively, the distance between embedding vectors can reflect the semantic similarity between their corresponding patches. We postulate that each selected vector in S can represent its K-nearest neighbors in U (denoted as covered vectors). We call it semantic coverage in the embedding space. In this case, a fixed-distance coverage strategy is not suitable due to variations in the distribution density of embedding vectors across different semantic distinctions, which may result in over- or under-representation of particular semantics. We have validated this statement in the experiments. We notice that some selected vectors may cover the same vector in U (see Fig. 2(a)), leading to local semantic redundancy and insufficient data diversity. Therefore, well-selected sub-volumes can ensure a fine representation of the overall data distribution and various neuronal structures with diverse and representative local patches.

With this in mind, we establish a criterion, **C**onstrained **C**overage **R**ate (CCR), to measure the semantic coverage ability of the *selected set* S. Formally,

$$V(v, U) = \{v_n \in U \mid ||v_n - v||_2 \leq ||u - v||_2,\ u \in U,\ |V| = K\}, \tag{2}$$

$$C(S, U) = \bigcup_{v_m \in S} V(v_m, U), \tag{3}$$

$$\text{CCR}(S, U) = \frac{|\ C(S, U)\ |}{|\ U\ |}, \tag{4}$$

where $||\cdot||_2$ represent l_2-norm, $V(v, U)$ returns the K-nearest neighbors of v in U, and thus *covered set* C contains all the vectors in U that are within the K-nearest neighbors of at least one vector in S. CCR defines the coverage rate as the percentage of covered vectors in U. A higher CCR indicates that the selected sub-volumes cover a larger portion of the neural structures in the entire volume and thus better represent the overall data distribution.

Given a labeling budget, we aim to select a subset S in U to maximize the $\mathrm{CCR}(S, U)$, where the vectors in S have spatial binding constraints. The optimal subset can be obtained by calculating the coverage rate across all sub-volume combinations. However, as the data scale grows and the annotation budget increases, the computational complexity of identifying the optimal sub-volumes becomes infeasible. As a compromise, we propose a **C**overage-based **G**reedy **S**election (CGS) algorithm to tackle the issue with linear time complexity, providing a promising solution for selection in large-scale EM volumes. Specifically, each iteration of CGS involves traversing the entire volume for the candidate sub-volume and extracting the corresponding vectors from U. These vectors are then added to the S to evaluate the selected sub-volume(s). At the end of each iteration, the sub-volume that garners the highest coverage rate is chosen for labeling. This process is repeated until the desired number of sub-volumes has been selected. CGS effectively mitigates the bias between labeled and unlabeled distributions. Please refer to Algorithm 1 for more details.

Algorithm 1: Coverage-based Greedy Selection for Sub-volumes (CGS)

Input: unlabeled volume D_u, embedding network f_θ, sub-volume sizes $\{s_i\}_{i=1}^{t}$
Output: selected sub-volumes D_l

```
selected set : S ← ∅, universal set : U ← patch-level sliding window inference on D_u using f(·)
for each sub-volume of size s_i ∈ s_1, s_2, ..., s_t do
    best_subset ← ∅, best_rate ← −∞
    for each position p in raw volume using sliding window of size s_i do
        subset ← extract vectors from U with {p, s_i}
        coverage_rate ← CCR(subset ∪ S, U)
        if coverage_rate>best_rate then
            best_subset ← subset, best_rate ← coverage_rate
        end
    end
    selected_set ← selected_set ∪ best_subset
end
D_l ← extract sub-volumes from D_u according to S
```

3.2. Semi-supervised Neuron Segmentation via Mixed-view Consistency Learning

To integrate information from selected sub-volumes D_l and unlabeled data D_u, we present IIC-Net, an innovative SSL framework for neuron segmentation. As illustrated in Fig. 3(a), axial-through mixing for volumetric EM data creates a connection between labeled and unlabeled datasets. On this basis, enforcing cross-view consistency on mixed data not only fosters the network's learning of common semantics but also enhances the discriminability of neuron boundaries.

3.2.1. Axial-through Spatial Mixing for Neuron Segmentation

The affinity graph used to characterize neuron boundaries is generated from dense instance annotations. Each affinity value ranges between 0 and 1, which indicates whether two associated voxels belong to the same instance (1) or not (0) along the $x/y/z$ axis. Let two patches with affinity map be $\{(x_i, y_i), (x_j, y_j)\}$, where $x_i, x_j \in \mathbb{R}^{D\times H\times W}$ and $y_i, y_j \in \{0, 1\}^{C\times D\times H\times W}$. Spatical mixing for raw patches and affinities are denoted as $\mathrm{M}^R(\cdot)$ and $\mathrm{M}^A(\cdot)$, respectively.

Currently, large-scale volume EM data typically exhibits strong anisotropy (Zheng et al., 2018; Shapson-Coe et al., 2024), *i.e.* the z-axis resolution is much lower than that of the x- and y-axes. To prevent potential semantic ambiguity caused by asymmetric axial and tangential resolution, we employ axial-through masking strategies in spatial mixing, as shown in Fig. 3(b). The mixed result for x_i and x_j is

$$\mathrm{M}^R(x_i, x_j) = \Pi^R \odot x_i + (\mathbf{1} - \Pi^R) \odot x_j, \tag{5}$$

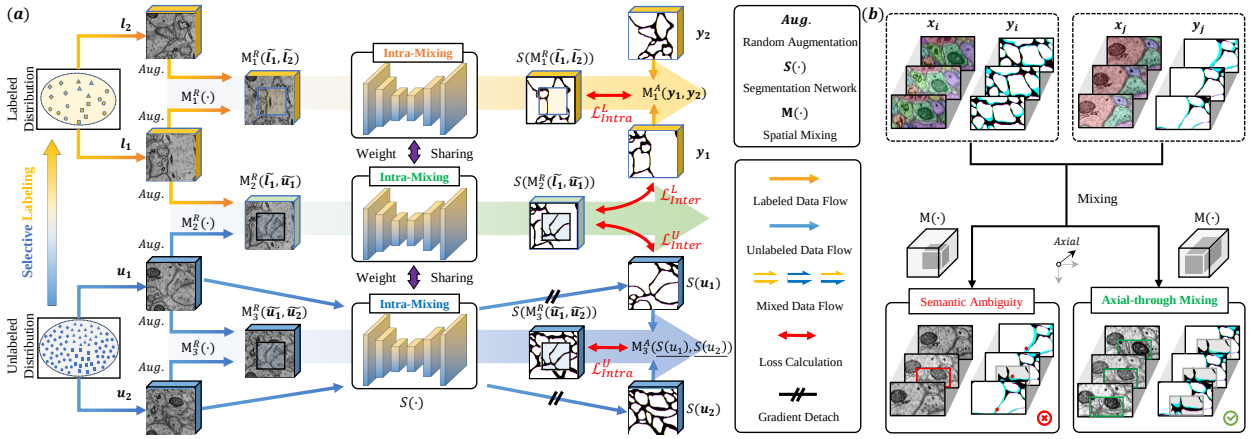

Figure 3: Semi-supervised neuron segmentation. (a) Training framework. Initially, representative and diverse sub-volumes are extracted and annotated from the unlabeled EM volume. All data resources are then fully exploited through intra- and inter-dataset mixing. On this basis, we reinforce mixed-view consistency to encourage shared semantic learning across datasets. (b) Axial-through mixing. The much lower resolution of anisotropic EM volume in the axial direction brings semantic confusion in axial affinities when using regular spatial mixing. We mitigate this issue by using an axial-through strategy to avoid drastic changes in affinity values.

where $\odot$ denote Hadamard product, $\Pi^R \in \{0,1\}^{D\times H\times W}$ is a binary matrix determined by mixing strategy, indicating whether the voxel comes from x_i (1) or x_j (0). The mixed result for c-channel affinities y_i and y_j is computed by

$$\mathrm{M}^A(y_i, y_j) = \Pi^A \odot y_i + (\mathbf{1} - \Pi^A) \odot y_j, \tag{6}$$

where $\Pi^A_{c,d,h,w} \equiv \Pi^R_{d,h,w}$, indicating that the same mixing operation is performed on each channel of the affinity graph.

3.2.2. Mixed-view Consistency Learning

Given that backbone design is not the primary focus, we utilize a modified 3D U-Net as the segmentation network $S(\cdot)$, which is widely used in neuron segmentation in anisotropic EM data (Lee et al., 2017, 2021; Huang et al., 2022). During training, two labeled pairs $\{(l_1, y_1), (l_2, y_2)\}$ and two unlabeled inputs $\{u_1, u_2\}$ are randomly cropped from D_l and D_u. M_i denotes the mixing operation with the masking matrix Π_i. Random data augmentation (including Gaussian blurring, simulated misalignment, Gaussian noise, randomly masked pixels, and photometric perturbation) is performed before mixing, providing diverse combinations of augmented views. As illustrated in Fig. 3, the supervision signal is built on intra- and inter-dataset mixing. Specifically,

Intra-dataset mixing. To maximize the utility of limited annotations and facilitate network adaptation to mixed inputs, the mixing strategy is employed within both labeled and unlabeled datasets. For labeled data, the supervised loss $\mathcal{L}^L_{Intra}$ is defined as follows,

$$\mathcal{L}^L_{Intra} = \mathcal{L}_{CE}(S(\mathrm{M}^R_1(\widetilde{l}_1, \widetilde{l}_2)), \mathrm{M}^A_1(y_1, y_2)), \tag{7}$$

where $\mathcal{L}_{CE}$ represents cross entropy loss, $\widetilde{l}_1$ and $\widetilde{l}_2$ are randomly augmented views of l_1 and l_2, respectively.

We leverage unlabeled data by imposing consistency constraints between predictions of perturbated and original inputs. Specifically, the pseudo labels for randomly augmented inputs $\widetilde{u}_1$ and $\widetilde{u}_2$ are computed by $S(u_1)$ and $S(u_2)$, and thus the mixed-view consistency loss for unlabeled data is given by

$$\mathcal{L}^U_{Intra} = \mathcal{L}_{SE}(S(\mathrm{M}^R_2(\widetilde{u}_1, \widetilde{u}_2)), \mathrm{M}^A_2(\underline{S(u_1)}, \underline{S(u_2)})), \tag{8}$$

where $\mathcal{L}_{SE}$ represents square error loss.

Inter-dataset mixing. Spatial mixing is also performed across datasets to connect labeled and unlabeled data. In detail, the annotated area in the mixed image $\mathrm{M}^R_3(\widetilde{l}_1, \widetilde{u}_1)$ is supervised under label map y_1, while the unlabeled area with the pseudo label $S(u_1)$. $\Pi^A_3 \in \{0,1\}^{C\times D\times H\times W}$ is a binary mask for the affinity graph, which indicates whether

the voxel comes from labeled part (1) or unlabeled part (0) under M_3^R. Therefore, the supervision signal of the mixed image can be divided into two parts, denoted as $\mathcal{L}_{Inter}^L$ and $\mathcal{L}_{Inter}^U$, respectively. Formally,

$$\mathcal{L}_{Inter}^L = \Pi_3^A \odot \mathcal{L}_{CE}(S(\mathrm{M}_3^R(\widetilde{l}_1, \widetilde{u}_1)), y_1), \tag{9}$$

$$\mathcal{L}_{Inter}^U = (\mathbf{1} - \Pi_3^A) \odot \mathcal{L}_{SE}(S(\mathrm{M}_3^R(\widetilde{l}_1, \widetilde{u}_1)), \underline{S(u_1)}). \tag{10}$$

Given that expert-annotated masks for labeled patches are typically more accurate than soft pseudo labels assigned to unlabeled patches, we apply λ as a weighting coefficient to balance the contribution of labeled and unlabeled data. The full loss function is formally given by,

$$\mathcal{L}_{Semi} = \mathcal{L}_{Intra}^L + \mathcal{L}_{Inter}^L + \lambda(\mathcal{L}_{Intra}^U + \mathcal{L}_{Inter}^U). \tag{11}$$

Warming-up. Consistency regularization operates effectively when applied to meaningful predictions of unlabeled data. However, achieving this condition can be challenging when the weights of the segmentation network are randomly initialized. To prevent convergence to sub-optimal solutions, we first train the network with $\mathcal{L}_{Intra}^L$, and then fine-tune it with the complete loss function.

4. Experiments setup

4.1. Dataset

The experimental evaluations are conducted on two representative and competitive public datasets with different visual appearances and spatial resolutions, *i.e.*, AC3/AC4, CREMI. Moreover, Kasthuri is utilized to evaluate the practical application value of the pipeline on large-scale datasets.

AC3/AC4[1]. This dataset consists of two densely labeled subsets of the Kasthuri dataset (Kasthuri et al., 2015), which was imaged from the mouse somatosensory cortex by scanning EM (SEM). AC3 and AC4 consist of 256 and 100 sequential images of size 1024×1024, respectively, at $6 \times 6 \times 29$ nm^3 voxel resolution. To ensure the reliability of the experimental results, we carefully divide them into two parts: the bottom 100 layers of AC3 and the complete 100 layers of AC4 are used to simulate unlabeled data D_u; the top 100 layers of AC3 are used for testing. Notably, the myelin regions in AC3/AC4 annotations are ignored, leading to confusion in boundary learning. Therefore, we adopt a revised version of annotation for AC4 with myelin annotation (*i.e.*, the training set of SNEMI[2]) and restrict all algorithms to select sub-volumes (D_l) from AC4 only. The sub-volume size for AC3/AC4 datasets is set as $18 \times 380 \times 380$.

CREMI[3]. The CREMI dataset comes from an adult Drosophila brain, imaged using serial section transmission EM (TEM) at a resolution of $4 \times 4 \times 40$ nm^3. The CREMI dataset comprises three labeled subsets (CREMI-A/B/C), each containing 125 sequential images associated with different regions. Since CREMI-A features a homogenous neuron pattern, our experiments employ the CREMI-B/C dataset. Similarly, we divide them into two parts: the top 60 layers of CREMI-B and CREMI-C are considered unlabeled data, denoted as D_u; the bottom 60 layers are used for testing. We select and label sub-volumes (D_l) of size $18 \times 340 \times 340$ from D_u. Moreover, the unlabeled set of CREMI, *i.e.* CREMI-B+/C+, is used for the investigation of distribution mismatch in Sec. 5.4.

Kasthuri[4]. The size of the Kasthuri dataset is $1850 \times 10747 \times 12895$, with a voxel resolution of $6 \times 6 \times 29$ nm^3. Our method is evaluated on a sparsely annotated subset cropped from the Kasthuri15 dataset, measuring $300 \times 4096 \times 4096$. The bounding box of this volume ranges from (1050, 6500, 3200) to (1350, 10596, 7296). In the experiments, the top 200 layers serve as unlabeled data (D_u), while the bottom 100 layers are designated as the test data. Since the myelin regions are also ignored in official annotations, we select labeled sub-volumes (D_l) sized $18 \times 380 \times 380$ from the training set of SNEMI based on embeddings of the unlabeled volume D_u.

AxonEM-H.[5] AxonEM-H dataset is imaged using ATUM-SEM from layer 2 in the temporal lobe of an adult human, containing nine $50 \times 512 \times 512$ sub-volumes at $8 \times 8 \times 30$ nm^3 resolution. We use the first six as unlabeled data (D_u) and the rest for evaluation. We select labeled sub-volumes (D_l) sized $18 \times 340 \times 340$ from the unlabeled data.

[1] https://lichtman.rc.fas.harvard.edu/vast/AC3AC4Package.zip
[2] https://snemi3d.grand-challenge.org
[3] https://cremi.org
[4] https://neurodata.io/data/kasthuri15
[5] https://axonem.grand-challenge.org

Hemi-Brain.[1] Hemi-Brain is a FIB-SEM volume of the Drosophila melanogaster central brain, imaged at 8 *nm* isotropic resolution. The training set for reconstruction comprises 8 densely labeled volumes. We utilize 6 volumes for subsequent experiments, each with a size of $520 \times 520 \times 520$. We partition these volumes into two groups based on their location in the brain. One is designated as unlabeled data (D_u), and the other serves as the test data. We select labeled sub-volumes (D_l) sized $18 \times 300 \times 300$ from the unlabeled data.

4.2. Evaluation Metrics

Two widely-used metrics for neuron segmentation (Funke et al., 2018; Sheridan et al., 2022; Lee et al., 2021; Huang et al., 2022), the variation of information VI (Nunez-Iglesias et al., 2013) and adapted Rand error ARAND (Arganda-Carreras et al., 2015), are adopted to evaluate the results.

VI is defined as a sum of the conditional entropies $\mathrm{H}(\cdot|\cdot)$ between the segmentation proposal A and the ground truth T, and can be broken down into two distinct components as

$$\mathrm{VI} = \mathrm{VI_s} + \mathrm{VI_m} = \mathrm{H}(S|T) + \mathrm{H}(T|S) \tag{12}$$

where VI_s and VI_m, corresponding to split and merge errors, respectively. Define p_{ij} as the probability that a randomly selected voxel belongs to the segment i in A and the segment j in T, the ARAND is formally given by

$$\mathrm{ARAND} = 1 - \frac{2\sum_{ij} p_{ij}^2}{\sum_i(\sum_j p_{ij}^2)^2 + \sum_j(\sum_i p_{ij}^2)^2} \tag{13}$$

Lower values correspond to higher segmentation quality in both metrics. The background is ignored when evaluating.

4.3. Implementation Details

All models are trained on 2 NVIDIA RTX 3090 GPUs using Adam optimizer. The input patch size is fixed at $18 \times 160 \times 160$, the batch size is 4, and the learning rate is set as 1e-4 during unsupervised and semi-supervised training. For selective labeling, the discretization of unlabeled volume is implemented by sliding window cropping with a stride of $8 \times 40 \times 40$. The dimension of the embedding space L is 80, and the number of neighbors K in CCR is set as 30. For semi-supervised learning, the segmentation network is the residual symmetric U-Net (RSU-Net) proposed in (Lee et al., 2017) that is widely adopted in neuron segmentation. We employ the rectangle and quarter masking strategies in axial-through spatial mixing. The weighting coefficient λ in $\mathcal{L}_{semi}$ is set as 0.2. The affinity graph has three channels (*i.e.*, $C = 3$), each representing the likelihood between a pixel and its nearest neighbors along the z, x, and y directions. Once the affinity map is predicted, we obtain the instance segmentation with two commonly adopted approaches, Multicut (Beier et al., 2017) and Waterz (Funke et al., 2018). All codes are available at https://github.com/yanchaoz/SL-SSNS.

5. Results

In this section, we first verify the effectiveness of CGS and IIC-Net separately, and then explore their cooperation for more effective semi-supervised segmentation. Additionally, we investigate the impact of distribution mismatch on the performance of semi-supervised models. Finally, we validate the improved pipeline in a large-scale scenario.

5.1. Effectiveness of Selective Labeling for Sub-volumes

In the following, we perform a comparative study on AC3/AC4 and CREMI datasets to confirm the superiority of the proposed sub-volume selection algorithm.

5.1.1. Quantitative Comparisons

All strategies, including patch-level and sub-volume selection schemes, are quantitatively evaluated through supervised training on the selected data, and their ability to reflect overall data distribution is assessed through segmentation performance.

In this subsection, the labeling budget for all baseline methods applied to the AC3/AC4 and CREMI datasets is set at 7.5% and 6%, respectively. For patch-level selective labeling, we consider the following competitors: (a) Random

[1] https://github.com/funkelab/lsd#notebooks

Table 1
The quantitative comparisons of selective labeling strategies on the AC3/AC4 and CREMI datasets.

		Methods	Multicut (Beier et al., 2017)				Waterz (Funke et al., 2018)			
			$VI_s \downarrow$	$VI_m \downarrow$	VI ↓	ARAND ↓	$VI_s \downarrow$	$VI_m \downarrow$	VI ↓	ARAND ↓
AC3/AC4	Patch-level	*Random*	0.9042	0.3883	1.2925	0.1477	0.9822	0.3837	1.3659	0.1637
		Equispaced	0.9459	0.3270	1.2729	0.1224	0.9441	0.3249	1.2690	0.1220
		K-means	0.9818	0.5333	1.5152	0.2147	0.9859	0.4220	1.4079	0.1401
		K-media	0.9826	0.3037	1.2863	0.1245	0.9467	0.3692	1.3159	0.1602
		FPS (Jin et al., 2022)	0.9594	0.3583	1.3177	0.1225	0.9314	0.3769	1.3083	<u>0.1177</u>
		RA(Zheng et al., 2019)	0.9424	0.3091	1.2515	0.1158	0.9212	0.2999	1.2211	0.1193
		USL (Wang et al., 2022a)	0.8997	0.3507	1.2504	0.1201	0.8974	0.3678	1.2651	0.1288
	Sub-volume	*Random*	0.8774	0.3256	1.2030	0.1182	0.8915	0.3417	1.2307	0.1432
		Equlispaced	0.8407	0.3430	1.1837	0.1189	0.8520	0.3986	1.2506	0.1499
		CGS (FD)	0.8733	0.2888	<u>1.1622</u>	<u>0.1150</u>	0.9000	0.2958	<u>1.1958</u>	0.1239
		CGS (Ours)	0.8546	0.2717	**1.1263**	**0.1103**	0.7705	0.3196	**1.0901**	**0.1144**
		Δ ↑	-	-	+6.4%	+6.7%	-	-	+11.4%	+20.1%
CREMI	Patch-level	*Random*	1.0034	0.9599	1.9633	0.2280	1.2531	1.4793	2.7324	0.3742
		Equispaced	1.0014	0.8144	1.8158	0.2026	1.0643	1.1739	2.2382	0.3161
		K-means	1.0656	0.8106	1.8761	0.2240	1.1551	1.2520	2.4070	0.3061
		K-media	0.9511	0.7897	1.7407	0.2034	1.1473	1.3617	2.5090	0.3919
		FPS (Jin et al., 2022)	1.0391	0.8372	1.8763	0.1817	1.2892	1.3643	2.6706	0.3406
		RA (Zheng et al., 2019)	0.9810	0.7769	1.7579	0.1790	1.2552	1.5709	2.8261	0.4077
		USL (Wang et al., 2022a)	1.0014	0.8144	1.8158	0.2026	1.0643	1.1739	2.2381	0.3161
	Sub-volume	*Random*	0.9775	0.7039	1.6814	0.1843	1.0961	1.2460	2.3421	0.3360
		Equispaced	0.8838	0.7897	1.6735	0.1626	1.0758	1.0967	2.1725	0.3219
		CGS (FD)	0.9463	0.5881	<u>1.5326</u>	<u>0.1503</u>	1.0152	1.0728	<u>2.0880</u>	<u>0.2595</u>
		CGS (Ours)	0.9169	0.5675	**1.4843**	**0.1281**	0.9693	0.8520	**1.8208**	**0.1944**
		Δ ↑	-	-	+11.7%	+30.5%	-	-	+22.3%	+14.2%

Δ indicates the gain of CGS over random sub-volume selection. **Bold** and <u>underlined</u> items indicate the first and second highest scoring results, respectively.

sampling, we repeat the random selection four times under different seeds and report the average value *Random*. (b) Equispaced sampling denotes as *Equispaced*. (c) One-shot active learning methods include K-median (Kolluru et al., 2021), K-means (Arthur and Vassilvitskii, 2007), Farthest Point Sampling (FPS) (Moenning and Dodgson, 2003; Jin et al., 2022), Representative Annotation (RA) (Zheng et al., 2019), and Unsupervised Selective Labeling (USL) (Wang et al., 2022a). For sub-volume selection methods, we only compare CGS with random and equispaced selection due to the lack of related research. Similarly, we repeat the random selection process four times and calculate the average value. Note that we also compare with CGS's selection under the fixed-distance coverage function (*i.e.*, each patch can represent patches within a fixed radius in the embedding space), denoted as CGS (FD). To ensure a fair comparison, we use the same unlabeled volumes and embedding network for all the competing methods.

Quantitative results in TABLE 1 support the following conclusions: (1) The sub-volume annotation technique achieves a significant improvement compared to patch-level annotation in neuron segmentation. Notably, even random sub-volume annotation outperforms most patch-level selection baselines. This highlights the limitations of existing OSAL methods when directly applied to volumetric EM segmentation tasks. (2) As a greedy strategy, CGS provides data selection with linear time complexity, which can effectively reflect the unlabeled distribution and produce excellent segmentation results. (3) The superior performance on both SEM and TEM datasets shows good adaptability and stability of the proposed heuristic. (4) The fixed-distance coverage strategy, *i.e.*, CGS (FD), fails to account for variations in the distribution density of patch embeddings, leading to a degradation in segmentation performance compared to K-nearest coverage function. Qualitatively, the examples of segmentation results from both datasets are provided in Fig 4.

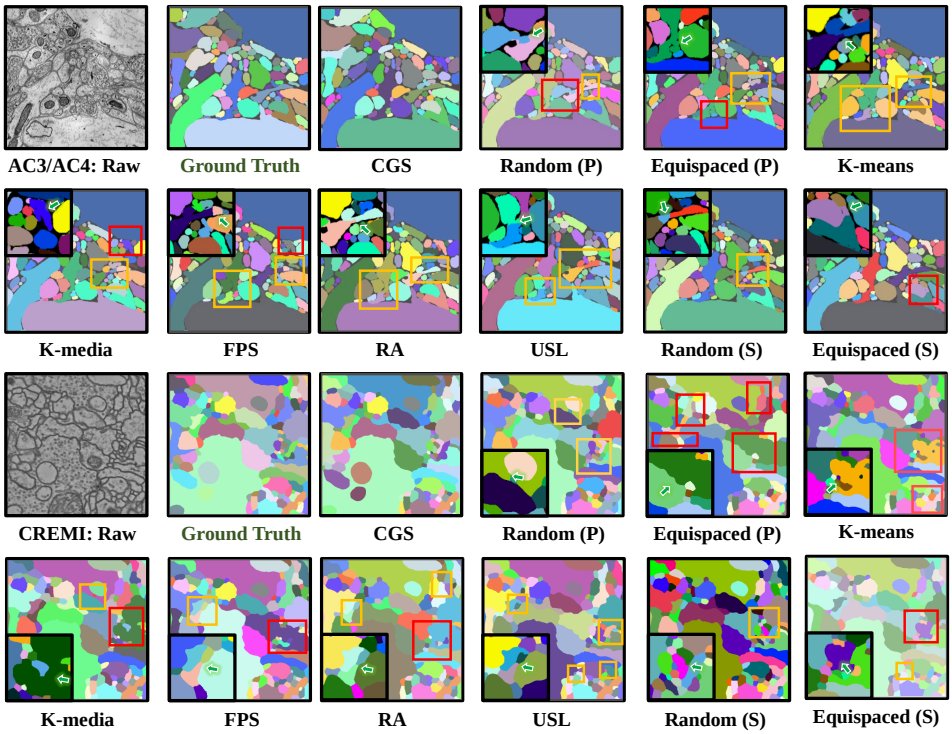

Figure 4: Qualitative comparison of selective labeling strategies on the AC3/AC4 (top) and CREMI (bottom) dataset. The red and yellow boxes represent merge and split errors, respectively. We also provide a zoomed-in view of typical errors within the black box. Multicut (Beier et al., 2017) algorithm is adopted for graph-based agglomeration. (P) and (S) denote patch-level and sub-volume annotations, respectively.

Table 2
Quantitative Comparisons on AC3/AC4 dataset with different sub-volume sizes with a budget of 7.5%.

Size (pixels)	Method	Multicut		Waterz	
		VI ↓	ARAND ↓	VI ↓	ARAND ↓
160 (patch)	*Random*	1.2925	0.1477	1.3659	0.1637
220 (sub-volume)	*Random*	1.2254	0.1432	1.2838	0.1538
	CGS	**1.1524**	**0.1138**	**1.1870**	**0.1170**
300 (sub-volume)	*Random*	1.1958	0.1214	1.2127	0.1406
	CGS	**1.1484**	**0.1111**	**1.1241**	**0.1121**
380 (sub-volume)	*Random*	1.2030	0.1182	1.2307	0.1432
	CGS	**1.1263**	**0.1103**	**1.0901**	**0.1144**
460 (sub-volume)	*Random*	1.1620	0.1348	1.2445	0.1744
	CGS	**1.1024**	**0.1163**	**1.1210**	**0.1130**

Bold item indicates the first scoring result.

Furthermore, we explore the influence of sub-volume size on segmentation outcomes through a comparison between CGS and random selections. We repeat the random selection four times and show the average value *Random*. The quantitative results presented in TABLE 2 demonstrate several key findings. Firstly, CGS exhibits stable performance and consistent improvements over random selections across different sizes. Secondly, when the size is too small, such as Patch and 220, labeled data may fail to capture the continuous changes in neuronal structures, leading to suboptimal performance for both random selections and CGS. Conversely, in the case of a large size, such as 460,

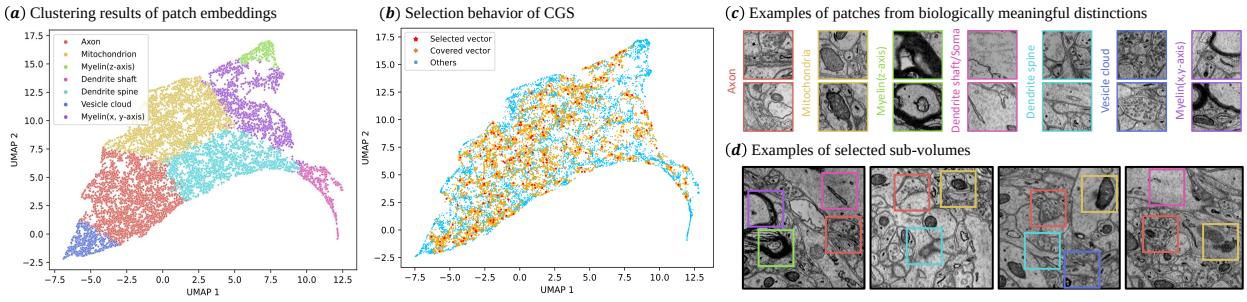

Figure 5: UMAP projection of patch embeddings extracted by $f(\cdot)$ on AC3/AC4 dataset. (a) Spectral clustering results exhibit biologically meaningful distinctions (legend writes dominant structures). (b) The sub-volumes selected via CGS are represented by selected and covered vectors in the embedding space. (c) Patch examples from each distinction. (d) CGS-selected sub-volumes cover diverse and representative patches.

Table 3
The quantitative comparisons of semi-supervised methods on the AC3/AC4 and CREMI datasets.

		Methods	Multicut (Beier et al., 2017)				Waterz (Funke et al., 2018)			
			$VI_s \downarrow$	$VI_m \downarrow$	VI ↓	ARAND ↓	$VI_s \downarrow$	$VI_m \downarrow$	VI ↓	ARAND ↓
AC3/AC4	Sup.	RSU-Net (Lee et al., 2017)	0.8546	0.2717	1.1263	0.1103	0.7705	0.3196	1.0901	0.1144
		RSU-Net *w/* Mix	0.7877	0.3482	1.1059	0.1061	0.7983	0.3144	1.1127	0.1031
	Semi.	PseudoSeg (Zou et al., 2020)	0.8190	0.2687	1.0876	0.1073	0.8489	0.2934	1.1424	0.1270
		UA-MT (Yu et al., 2019)	0.8018	0.2645	1.0662	0.0998	0.8588	0.2886	1.1473	0.1172
		SASSNet (Li et al., 2020)	0.7784	0.2914	1.0698	0.1143	0.8434	0.2467	1.0902	0.1086
		MC-Net (Wu et al., 2021)	0.7945	0.2592	1.0538	0.1082	0.8329	0.2390	1.0719	0.1049
		SSNS-Net (Huang et al., 2022)	0.7308	0.3166	1.0473	0.1025	0.7781	0.3113	1.0894	0.1165
		IIC-Net *w/o* Inter-mix	0.7395	0.2899	1.0293	0.1008	0.7732	0.2882	<u>1.0614</u>	**0.0960**
		IIC-Net *w/o* Intra-mix	0.7335	0.2792	<u>1.0127</u>	<u>0.0937</u>	0.8136	0.2559	1.0695	0.1072
		IIC-Net (Ours)	0.6895	0.2855	**0.9751**	**0.0922**	0.7813	0.2693	**1.0506**	<u>0.1002</u>
		Δ ↑	-	-	+13.4%	+16.4%	-	-	+3.6%	+12.4%
CREMI	Sup.	RSU-Net (Lee et al., 2017)	0.9169	0.5675	1.4843	0.1281	0.9693	0.8520	1.8208	0.1944
		RSU-Net *w/* Mix	0.8462	0.6004	1.4466	0.1336	0.9451	0.8743	1.8193	0.2034
	Semi.	PseudoSeg (Zou et al., 2020)	0.9068	0.5406	1.4474	0.1362	0.9819	0.8659	1.8478	0.2289
		UA-MT (Yu et al., 2019)	0.8966	0.5420	1.4386	<u>0.1216</u>	0.9718	0.7593	1.7311	0.1730
		SASSNet (Li et al., 2020)	0.8917	0.5552	1.4469	0.1265	0.9804	0.7711	1.7514	0.1815
		MC-Net (Wu et al., 2021)	0.8970	0.5409	1.4379	0.1235	0.9678	0.7537	1.7215	0.1942
		SSNS-Net (Huang et al., 2022)	0.8922	0.5329	1.4251	0.1217	0.9451	0.7229	1.6679	0.1633
		IIC-Net *w/o* Inter-mix	0.8472	0.5367	1.3838	0.1229	0.9323	0.7099	1.6400	0.1831
		IIC-Net *w/o* Intra-mix	0.8767	0.4986	<u>1.3753</u>	0.1258	0.8886	0.6707	<u>1.5593</u>	<u>0.1576</u>
		IIC-Net (Ours)	0.8534	0.4992	**1.3526**	**0.1122**	0.8669	0.6830	**1.5499**	**0.1416**
		Δ ↑	-	-	+8.9%	+12.4%	-	-	+14.9%	+27.2%

Sup. and Semi. are abbreviations for supervised and semi-supervised learning. Δ shows improvements of our approach over the supervised baseline (RSU-Net). **Bold** and <u>underlined</u> items indicate the first and second scoring results, respectively.

randomly selected data might struggle to cover diverse and informative neuronal patterns, whereas CGS effectively mitigates this issue and demonstrates superior robustness.

5.1.2. Visualization of Selection Strategy in Feature Space

We employ the UMAP to project the patch embeddings for the AC3/AC4 dataset into a two-dimensional feature space. Clustering those vectors allows for identifying biologically meaningful distinctions. We visualize the selected sub-volumes with associated patch embeddings, *i.e.*, selected and covered vectors. As in Fig. 5, representative neural structures in the raw volume are semantically covered by the selected sub-volumes, including myelin, dendrite shaft

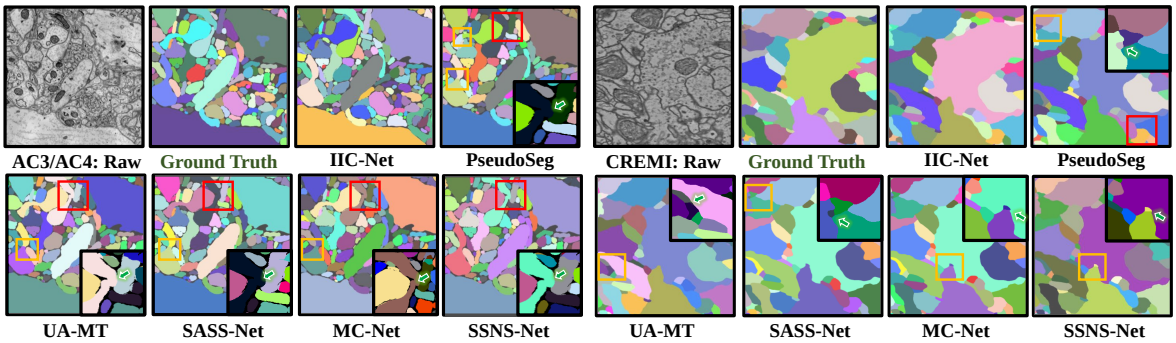

Figure 6: Qualitative comparison of semi-supervised methods on the AC3/AC4 (left) and CREMI (right) dataset. The red and yellow boxes represent merge and split errors, respectively. We also provide a zoomed-in view of typical errors within the black box. Multicut (Beier et al., 2017) algorithm is adopted for graph-based agglomeration.

(or soma), dendrite spine, mitochondria, vesicle cloud, and axon. This visualization validates the effectiveness of CGS in extracting informative and diverse sub-regions from EM volumes.

5.2. Effectiveness of Mixed-view Consistency Learning

In this subsection, we verify the state-of-the-art performance of the proposed semi-supervised paradigm, *i.e.*, IIC-Net. Experiments on AC3/AC4 and CREMI datasets are conducted on CGS-selected sub-volumes D_l and unlabeled data D_u. The labeling budget is consistent with the setting in Sec. 5.1.

Two supervised learning methods and five semi-supervised learning methods are adopted for comparison, including (1) RSU-Net (Lee et al., 2017), using only labeled data in a supervised manner, (2) RSU-Net w/ Mixing, incorporating axial-through spatial mixing for labeled data, (3) PseudoSeg (Zou et al., 2020), a classical semi-supervised approach that enforces consistency between unlabeled data and hard pseudo labels, (4) UA-MT (Yu et al., 2019), utilizing plausible predictions to guide consistency learning, (5) SASSNet (Li et al., 2020), employs an adversarial learning strategy to enforce a geometric shape constraint between the affinity maps of labeled and unlabeled data, (6) MC-Net (Wu et al., 2021), proposes mutual consistency regulation to encourage low-entropy predictions, and (7) SSNS-Net (Huang et al., 2022), the first semi-supervised neuron segmentation framework introducing innovative network initialization and modeling consistency between soft pseudo labels and perturbed predictions of unlabeled data. Additionally, we perform an ablation study to demonstrate the benefits of intra- and inter-dataset mixing, respectively. We consider the following variants: IIC-Net w/o Inter-mixing, where we disable $\mathcal{L}^{L}_{Inter}$ and $\mathcal{L}^{U}_{Inter}$, and IIC-Net w/o Intra-mixing, where Π in M_1 and M_3 are set as $\mathbf{1}$. For a fair comparison, we use consistent network architecture and augmentation strategies across all baseline methods.

The quantitative results under two post-processing algorithms present in TABLE 3 support the following conclusions: (1) Performing spatial mixing on labeled data (RSU-Net w/ Mix) can directly improve the segmentation performance in the supervised setting. This reveals the effectiveness of mixing strategies in promoting a better understanding of neuron boundaries. (2) The superior performance of soft pseudo-labeling strategies (UA-MT, SASSNet, MC-Net, SSNS-Net) substantiates its greater efficacy compared to hard pseudo-labeling (PseudoSeg) in semi-supervised neuron segmentation. (3) The proposed IIC-Net achieves consistent and considerable improvements over all baselines, which confirms the effectiveness of mixed-view consistency regularization. (4) The performance of IIC-Net decreases when intra-dataset mixing is turned off (w/o Intra-mix), emphasizing its critical role in facilitating adaptation to mixed inputs. Notably, when inter-dataset mixing (w/o Inter-mix) is disabled, the performance drops even more. Such results suggest that inter-mixing serves as a key factor. It connects labeled and unlabeled datasets and encourages the network to learn shared semantics. We also provide a visual comparison of different semi-supervised methods in Fig. 6.

5.3. Collaboration between Selective Labeling and Semi-supervised Training

After evaluating the effectiveness of selective labeling and mixed-view consistency learning individually, we explore how these work together for cost-effective neuron segmentation in different labeling budgets. This subsection can be regarded as an ablation study of the proposed pipeline.

Table 4
Quantitative Comparisons on AC3/AC4 dataset using different label budgets.

Budget	SL	SSL	VI ↓	ARAND ↓
5%	*Random**	*w/o*	1.4155	0.2082
	*Random**	IIC-Net	1.0745	0.1422
	CGS	*w/o*	1.2458	0.1468
	CGS	SSNS-Net	<u>1.0680</u>	<u>0.1220</u>
	CGS	IIC-Net	**1.0436**	**0.1107**
7.5%	*Random**	*w/o*	1.1608	0.1135
	*Random**	IIC-Net	<u>1.0225</u>	0.1057
	CGS	*w/o*	1.1263	0.1103
	CGS	SSNS-Net	1.0473	<u>0.1025</u>
	CGS	IIC-Net	**0.9751**	**0.0922**
10%	*Random**	*w/o*	1.0649	0.0998
	*Random**	IIC-Net	<u>0.9753</u>	0.1053
	CGS	*w/o*	1.0343	0.0953
	CGS	SSNS-Net	1.0110	<u>0.0936</u>
	CGS	IIC-Net	**0.9688**	**0.0899**
100%	*w/o*	*w/o*	**0.9583**	**0.0899**

SL and SSL denote selective labeling and semi-supervised learning. Multicut (Beier et al., 2017) algorithm is used for post-processing. **Bold** and <u>underlined</u> items indicate the first and second scoring results, respectively.

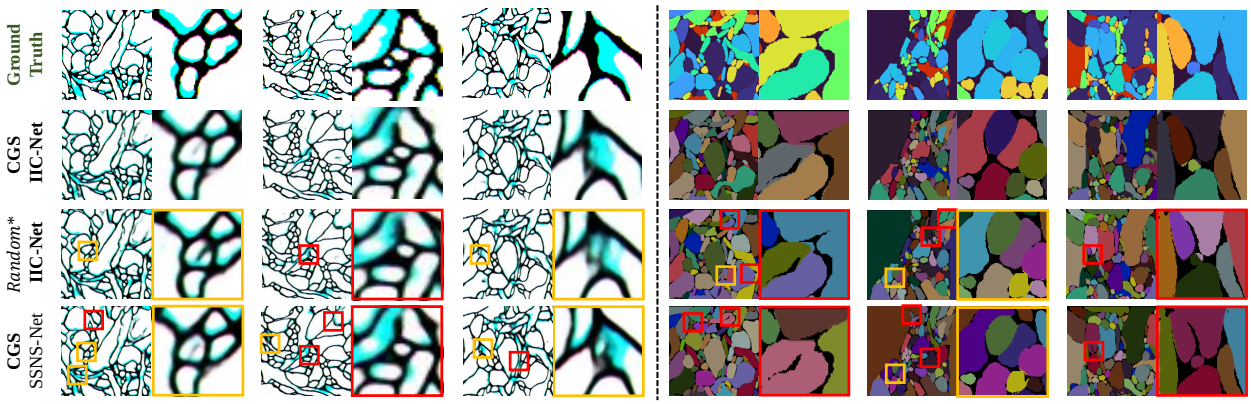

Figure 7: 2D visual comparison of affinity (left) and segmentation (right) results on the AC3/AC4 dataset with labeling cost of 7.5%. The red and yellow boxes represent merge and split errors, respectively. Multicut (Beier et al., 2017) algorithm is adopted for graph-based agglomeration.

We gradually increase the labeling cost on AC3/AC4 and evaluate the performance under supervised and semi-supervised settings. For selective labeling, we compare CGS with random sub-volume selection and report the optimal result in four experiments, denoted as *Random**. For semi-supervised learning, we adopt the most representative baseline, SSNS-Net (Huang et al., 2022), for comparison. The results under Multicut post-processing are presented in TABLE 4. The proposed framework, utilizing labeled data selected via CGS and a segmentation network trained by IIC-Net, demonstrates consistent improvements over other combinations across all annotation budgets. It achieves comparable performance with fully supervised methods, even with only 10% labeled data on the AC3/AC4 dataset. Also, we provide both 2D and 3D visual comparisons in Fig. 7 and Fig. 8, respectively. The proposed method shows significant advantages in maintaining the precise connectivity of reconstructed structures.

In addition, we perform experiments on the AxonEM-H dataset and the Hemi-Brain dataset to verify the generalizability of the proposed method. The same conclusions can be obtained from quantitative results in Fig. 9, which demonstrates the adaptability of the proposed framework to diverse data distributions and voxel resolutions.

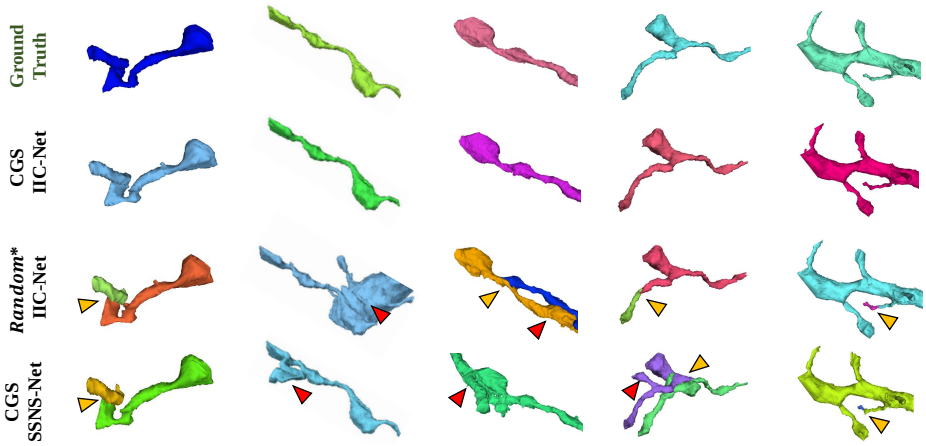

Figure 8: 3D visual comparison of segmentation results on the AC3/AC4 dataset with labeling cost of 7.5%. The red and yellow arrows represent merge and split errors, respectively. Multicut (Beier et al., 2017) algorithm is adopted for graph-based agglomeration.

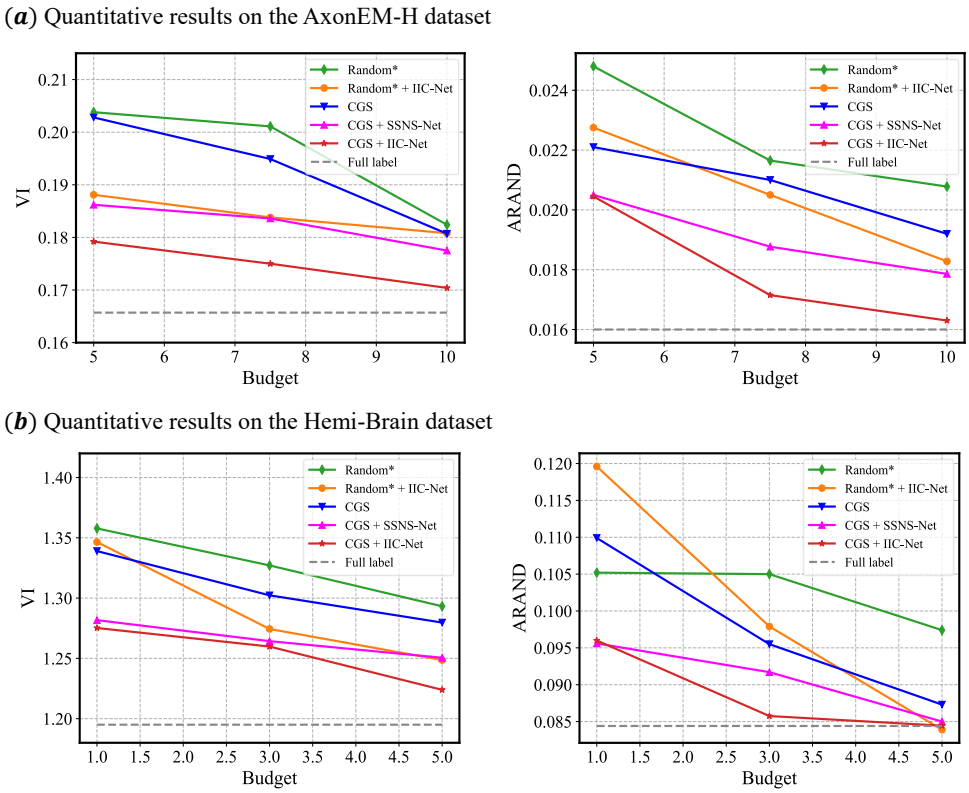

Figure 9: Quantitave Comparisons on the AxonEM-H and HemiBrain dataset with different annotation budgets. Multicut (Beier et al., 2017) algorithm is adopted for graph-based agglomeration.

5.4. Investigation of Distribution Mismatch

In the following, we conducted semi-supervised training with IIC-Net on the AC3/AC4 and CREMI datasets under different data settings to investigate the impact of the mismatch between labeled and unlabeled distributions.

Building on the conclusions from Sec. 5.1, the labeled data (D_l) selected by CGS can ensure a matching distribution with the original unlabeled dataset (D_u), establishing a baseline (Match) for other data combinations. For comparison,

Table 5
Investigation of distribution mismatch by changing labeled and unlabeled data on AC3/AC4 and CREMI dataset.

Methods			VI ↓	ARAND ↓
AC3/AC4	Match	D_l (CGS), D_u	**0.9751**	**0.0922**
	Mismatch	Changing Labeled Data	1.0694	0.1036
		Changing Unlabeled Data	1.0629	0.1107
CREMI	Match	D_l (CGS), D_u	**1.3526**	**0.1122**
	Mismatch	Changing Labeled Data	1.3767	0.1264
		Changing Unlabeled Data	1.4313	0.1362

Mismatch shows the average performance of four random selection experiments. Multicut (Beier et al., 2017) algorithm is used for post-processing. **Bold** item indicates the first scoring result.

Table 6
Evaluation on Kasthuri15 with the proposed pipeline.

Budget	SL	SSL	VI ↓	ARAND ↓
10%	CGS	w/o	0.8540	0.1810
	CGS	IIC-Net	0.8100	0.1741
	-	-	+5.15%	+3.81%
20%	CGS	w/o	0.8236	0.1795
	CGS	IIC-Net	0.7842	0.1710
	-	-	+4.78%	+4.74%
100%	w/o	w/o	0.7751	0.1731
	w/o	IIC-Net	0.7362	0.1632
	-	-	+5.02%	+5.72%

SL and SSL denote selective labeling and semi-supervised learning. Multicut (Beier et al., 2017) algorithm is adopted for graph-based agglomeration.

we introduce mismatch by changing labeled and unlabeled distributions respectively: (1) Fix the unlabeled data and randomly change the labeled data. These labeled sub-volumes are randomly cropped from the original unlabeled datasets four times under the same budget. (2) Fix the labeled data and randomly change the unlabeled data. For AC3/AC4, we took eight unlabeled volumes of size 100×1024×1024 from the Kasthuri dataset; for CREMI, we randomly cropped eight unlabeled volumes from CREMI-B+/C+, each containing 60 sequential images. Both are grouped equally into four sets. The quantitative results, as depicted in TABLE 5, indicate that: (1) Mismatched distribution hinders the generalization of segmentation models and leads to significant degradation in performance. (2) The CGS algorithm provides a practical and effective solution for the mismatch problem in semi-supervised neuron segmentation from the data level.

5.5. Evaluation on Large-scale Data

In the following, we conduct the proposed pipeline, *i.e.* selective labeling via CGS, and semi-supervised training via IIC-Net, on the Kasthuri dataset to study the practical value of the proposed method in large-scale neuron reconstruction.

Quantitative results are shown in TABLE 6. Under the same labeling budget including 100%, IIC-Net achieves consistent improvements over the supervised baseline on the segmentation results. Since expert proofreading is extremely time-consuming and expensive, semi-supervised training provides more robust predictions with the same labeling budget and inference overhead, which holds great potential in large-scale connectomics. In addition, the performance of the fully supervised method (VI: 0.7751, ARAND: 0.1731) is matched by the proposed pipeline with only 20% of labeled data (VI: 0.7842, ARAND: 0.1710), demonstrating the effectiveness of cooperation between coverage-based sub-volume selection and mixed-view consistency regularization.

6. Summary and Discussion

Automatic neuron reconstruction in large-scale EM volume produces massive merge and split errors, which require years of manual proofreading before they can be used for connectivity analysis. Therefore, achieving higher accuracy in the automatic stage plays a vital role in cost-effective connectomics. Semi-supervised learning offers a pathway to enhance model performance without additional expert annotations and inference overhead, by leveraging readily available unlabeled data.

This study introduces a novel semi-supervised pipeline for 3D neuron segmentation in electron microscopy volumes. By integrating selective labeling and mixed-view consistency regularization, we mitigate the mismatch between labeled and unlabeled distributions at both the data and model levels, paving the way for more effective semi-supervised learning. Extensive experiments validate the method's superiority and generalizability, demonstrating its potential to reduce manual workload and expedite scientific analysis in connectomics.

Despite the effectiveness of this pipeline, there are opportunities for improvement in three aspects:

1. Given the added expense of retraining the embedding network on unseen datasets, we aim to design a generalist embedding model for volumetric EM data. This model will not only support coverage-based selective labeling but also facilitate the discovery and analysis of local patterns from unlabeled EM data.
2. Noisy regions in pseudo labels can substantially degrade the performance of semi-supervised models. Hence, we will investigate a denoising technique for pseudo-labels in affinity graphs for neuron instance segmentation to enhance consistency learning.
3. While OSAL demonstrates computational efficiency, integrating ITAL technology with semi-supervised training for neuron segmentation can leverage both model performance feedback and data distribution to improve labeled data selection, potentially enhancing segmentation accuracy for large-scale reconstructions.

Acknowledgments

This work was supported in part by the STI 2030–Major Projects under Grant 2021ZD0204500 and Grant 2021ZD0204503, in part by the National Natural Science Foundation of China under Grant 32171461, and in part by the Scientific Research Instrument and Equipment Development Project of Chinese Academy of Sciences under Grant YJKYYQ20210022. We thank the Transdisciplinary Platform of Functional Connectome and Brain-inspired Intelligence in Huairou Science City in Beijing for providing technical support and device resources.

References

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., 2020. The mind of a mouse. Cell 182, 1372–1376.

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., 2015. Crowdsourcing the creation of image segmentation algorithms for connectomics. Frontiers in Neuroanatomy 9, 142.

Arthur, D., Vassilvitskii, S., 2007. K-means++ the advantages of careful seeding, in: Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms, pp. 1027–1035.

Bai, Y., Chen, D., Li, Q., Shen, W., Wang, Y., 2023. Bidirectional copy-paste for semi-supervised medical image segmentation, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 11514–11524.

Beier, T., Pape, C., Rahaman, N., Prange, T., Berg, S., Bock, D.D., Cardona, A., Knott, G.W., Plaza, S.M., Scheffer, L.K., et al., 2017. Multicut brings automated neurite segmentation closer to human performance. Nature Methods 14, 101–102.

Beluch, W.H., Genewein, T., Nürnberger, A., Köhler, J.M., 2018. The power of ensembles for active learning in image classification, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 9368–9377.

Berthelot, D., Carlini, N., Goodfellow, I., Papernot, N., Oliver, A., Raffel, C.A., 2019. Mixmatch: A holistic approach to semi-supervised learning. Advances in Neural Information Processing Systems 32.

Chen, X., He, K., 2021. Exploring simple siamese representation learning, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 15750–15758.

Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O., 2016. 3d u-net: learning dense volumetric segmentation from sparse annotation, in: International Conference on Medical Image Computing and Computer Assisted Intervention, Springer. pp. 424–432.

Dorkenwald, S., Matsliah, A., Sterling, A.R., Schlegel, P., Yu, S.C., McKellar, C.E., Lin, A., Costa, M., Eichler, K., Yin, Y., et al., 2023. Neuronal wiring diagram of an adult brain. bioRxiv .

French, G., Laine, S., Aila, T., Mackiewicz, M., Finlayson, G.D., 2019. Semi-supervised semantic segmentation needs strong, varied perturbations. British Machine Vision Conference .

Funke, J., Tschopp, F., Grisaitis, W., Sheridan, A., Singh, C., Saalfeld, S., Turaga, S.C., 2018. Large scale image segmentation with structured loss based deep learning for connectome reconstruction. IEEE Transactions on Pattern Analysis and Machine Intelligence 41, 1669–1680.

Huang, W., Chen, C., Xiong, Z., Zhang, Y., Chen, X., Sun, X., Wu, F., 2022. Semi-supervised neuron segmentation via reinforced consistency learning. IEEE Transactions on Medical Imaging 41, 3016–3028.

Januszewski, M., Kornfeld, J., Li, P.H., Pope, A., Blakely, T., Lindsey, L., Maitin-Shepard, J., Tyka, M., Denk, W., Jain, V., 2018. High-precision automated reconstruction of neurons with flood-filling networks. Nature Methods 15, 605–610.

Jin, Q., Yuan, M., Qiao, Q., Song, Z., 2022. One-shot active learning for image segmentation via contrastive learning and diversity-based sampling. Knowledge-Based Systems 241, 108278.

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., 2015. Saturated reconstruction of a volume of neocortex. Cell 162, 648–661.

Kolluru, C., Lee, J., Gharaibeh, Y., Bezerra, H.G., Wilson, D.L., 2021. Learning with fewer images via image clustering: Application to intravascular oct image segmentation. IEEE Access 9, 37273–37280.

Konyushkova, K., Sznitman, R., Fua, P., 2019. Geometry in active learning for binary and multi-class image segmentation. Computer vision and image understanding 182, 1–16.

Kurakin, A., Raffel, C., Berthelot, D., Cubuk, E.D., Zhang, H., Sohn, K., Carlini, N., 2019. Remixmatch: Semi-supervised learning with distribution matching and augmentation anchoring, in: International Conference on Learning Representations.

Lee, K., Lu, R., Luther, K., Seung, H.S., 2021. Learning and segmenting dense voxel embeddings for 3d neuron reconstruction. IEEE Transactions on Medical Imaging 40, 3801–3811.

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

Li, S., Zhang, C., He, X., 2020. Shape-aware semi-supervised 3d semantic segmentation for medical images, in: Medical Image Computing and Computer Assisted Intervention, Springer. pp. 552–561.

Li, Z., Yang, X., Liu, J., Hong, B., Zhang, Y., Zhai, H., Shen, L., Chen, X., Liu, Z., Han, H., 2024. Deepmulticut: Deep learning of multicut problem for neuron segmentation from electron microscopy volume. IEEE Transactions on Pattern Analysis and Machine Intelligence doi:10.1109/TPAMI.2024.3409634.

Lin, Z., Wei, D., Jang, W.D., Zhou, S., Chen, X., Wang, X., Schalek, R., Berger, D., Matejek, B., Kamentsky, L., et al., 2020. Two stream active query suggestion for active learning in connectomics, in: European Conference on Computer Vision, Springer. pp. 103–120.

Lou, W., Li, H., Li, G., Han, X., Wan, X., 2022. Which pixel to annotate: a label-efficient nuclei segmentation framework. IEEE Transactions on Medical Imaging 42, 947–958.

Luo, X., Chen, J., Song, T., Wang, G., 2021. Semi-supervised medical image segmentation through dual-task consistency, in: Proceedings of the AAAI Conference on Artificial Intelligence, pp. 8801–8809.

Mannix, E.J., Bondell, H.D., 2023. Cold paws: Unsupervised class discovery and the cold-start problem. arXiv:2305.10071 .

Meirovitch, Y., Mi, L., Saribekyan, H., Matveev, A., Rolnick, D., Shavit, N., 2019. Cross-classification clustering: An efficient multi-object tracking technique for 3-d instance segmentation in connectomics, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 8425–8435.

Moenning, C., Dodgson, N.A., 2003. Fast marching farthest point sampling for implicit surfaces and point clouds. Computer Laboratory Technical Report 565, 1–12.

Ngai, J., 2022. Brain 2.0: Transforming neuroscience. Cell , 4–8.

Nunez-Iglesias, J., Kennedy, R., Parag, T., Shi, J., Chklovskii, D.B., 2013. Machine learning of hierarchical clustering to segment 2d and 3d images. PloS one 8, e71715.

Peddie, C.J., Collinson, L.M., 2014. Exploring the third dimension: volume electron microscopy comes of age. Micron 61, 9–19.

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

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., et al., 2015. Imagenet large scale visual recognition challenge. International Journal of Computer Vision 115, 211–252.

Scheffer, L.K., Xu, C.S., Januszewski, M., Lu, Z., Takemura, S.y., Hayworth, K.J., Huang, G.B., Shinomiya, K., Maitlin-Shepard, J., Berg, S., et al., 2020. A connectome and analysis of the adult drosophila central brain. Elife 9, e57443.

Schmidt, M., Motta, A., Sievers, M., Helmstaedter, M., 2024. Roboem: automated 3d flight tracing for synaptic-resolution connectomics. Nature Methods , 1–6.

Sener, O., Savarese, S., 2018. Active learning for convolutional neural networks: A core-set approach, in: International Conference on Learning Representations.

Shapson-Coe, A., Januszewski, M., Berger, D.R., Pope, A., Wu, Y., Blakely, T., Schalek, R.L., Li, P.H., Wang, S., Maitin-Shepard, J., et al., 2024. A petavoxel fragment of human cerebral cortex reconstructed at nanoscale resolution. Science 384, eadk4858.

Sheridan, A., Nguyen, T.M., Deb, D., Lee, W.C.A., Saalfeld, S., Turaga, S.C., Manor, U., Funke, J., 2022. Local shape descriptors for neuron segmentation. Nature Methods , 1–9.

Shi, Y., Zhang, J., Ling, T., Lu, J., Zheng, Y., Yu, Q., Qi, L., Gao, Y., 2021. Inconsistency-aware uncertainty estimation for semi-supervised medical image segmentation. IEEE Transactions on Medical Imaging 41, 608–620.

Tarvainen, A., Valpola, H., 2017. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. Advances in neural information processing systems 30.

Wang, Q., Li, W., Gool, L.V., 2019. Semi-supervised learning by augmented distribution alignment, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 1466–1475.

Wang, X., Lian, L., Yu, S.X., 2022a. Unsupervised selective labeling for more effective semi-supervised learning, in: European Conference on Computer Vision, Springer. pp. 427–445.

Wang, X., Yuan, Y., Guo, D., Huang, X., Cui, Y., Xia, M., Wang, Z., Bai, C., Chen, S., 2022b. Ssa-net: Spatial self-attention network for covid-19 pneumonia infection segmentation with semi-supervised few-shot learning. Medical Image Analysis 79, 102459.

Wen, S., Kurc, T.M., Hou, L., Saltz, J.H., Gupta, R.R., Batiste, R., Zhao, T., Nguyen, V., Samaras, D., Zhu, W., 2018. Comparison of different classifiers with active learning to support quality control in nucleus segmentation in pathology images. AMIA Summits on Translational Science Proceedings 2018, 227.

Wu, Y., Xu, M., Ge, Z., Cai, J., Zhang, L., 2021. Semi-supervised left atrium segmentation with mutual consistency training, in: Medical image computing and computer assisted intervention, Springer. pp. 297–306.

Yang, L., Zhang, Y., Chen, J., Zhang, S., Chen, D.Z., 2017. Suggestive annotation: A deep active learning framework for biomedical image segmentation, in: International Conference on Medical Image Computing and Computer Assisted Intervention, Springer. pp. 399–407.

Yang, L., Zhuo, W., Qi, L., Shi, Y., Gao, Y., 2022. St++: Make self-training work better for semi-supervised semantic segmentation, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 4268–4277.

Yu, L., Wang, S., Li, X., Fu, C.W., Heng, P.A., 2019. Uncertainty-aware self-ensembling model for semi-supervised 3d left atrium segmentation, in: International Conference on Medical Image Computing and Computer Assisted Intervention, Springer. pp. 605–613.

Yun, S., Han, D., Oh, S.J., Chun, S., Choe, J., Yoo, Y., 2019. Cutmix: Regularization strategy to train strong classifiers with localizable features, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 6023–6032.

Zheng, H., Yang, L., Chen, J., Han, J., Zhang, Y., Liang, P., Zhao, Z., Wang, C., Chen, D.Z., 2019. Biomedical image segmentation via representative annotation, in: Proceedings of the AAAI Conference on Artificial Intelligence, pp. 5901–5908.

Zheng, Z., Lauritzen, J.S., Perlman, E., Robinson, C.G., Nichols, M., Milkie, D., Torrens, O., Price, J., Fisher, C.B., Sharifi, N., et al., 2018. A complete electron microscopy volume of the brain of adult drosophila melanogaster. Cell 174, 730–743.

Zou, Y., Zhang, Z., Zhang, H., Li, C.L., Bian, X., Huang, J.B., Pfister, T., 2020. Pseudoseg: Designing pseudo labels for semantic segmentation, in: International Conference on Learning Representations.