

Efficient Connectivity-Preserving Instance Segmentation with Supervoxel-Based Loss Function

Anna Grim, Jayaram Chandrashekar, Uygur Sümbül

Allen Institute
615 Westlake Avenue
Seattle, WA 98109 USA
anna.grim@alleninstitute.org

Abstract

Reconstructing the intricate local morphology of neurons and their long-range projecting axons can address many connectivity related questions in neuroscience. The main bottleneck in connectomics pipelines is correcting topological errors, as multiple entangled neuronal arbors is a challenging instance segmentation problem. More broadly, segmentation of curvilinear, filamentous structures continues to pose significant challenges. To address this problem, we extend the notion of simple points from digital topology to connected sets of voxels (i.e. supervoxels) and propose a topology-aware neural network segmentation method with minimal computational overhead. We demonstrate its effectiveness on a new public dataset of 3-d light microscopy images of mouse brains, along with the benchmark datasets DRIVE, ISBI12, and CrackTree.

Introduction

High-throughput neuron reconstruction is a challenging 3-d instance segmentation problem and a major bottleneck of many data-driven neuroscience studies (Gouwens et al. 2020; Winnubst et al. 2019). Deep learning-based methods are the leading framework for segmenting individual neurons, which is a critical step in reconstructing neural circuits (Turaga et al. 2010; Januszewski et al. 2018; Lee et al. 2019; Schmidt et al. 2024). While these methods and many others have significantly improved segmentation quality, automated reconstruction of neurons from large 3-d images still suffers from topological errors (i.e. splits and merges) and requires substantial human proofreading. This can be attributed to two basic observations: (i) neuronal branches can be as thin as a single voxel and (ii) branches often touch or even overlap due to the imaging resolution. Consequently, a seemingly innocuous mistake at the single voxel level can produce catastrophically incorrect segmentations.

A natural solution to these challenges is to account for the topology of the underlying objects during training. Several methods use persistent homology to capture the topology of segmented objects in terms of their Betti numbers (Clough et al. 2019; Hu et al. 2021; Stucki et al. 2023). While these methods have achieved state-of-the-art results, they are computationally expensive with complexities that scale nonlinearly. In this paper, we present a more efficient approach

based on an extension of simple voxels from digital topology. Our method aims to preserve the connectivity of segmented objects and has significantly faster training times that scale linearly.

In digital topology, a simple point in a 3-d binary image is defined as a foreground voxel whose deletion does not change the topology of the image (Kong and Rosenfeld 1989). Specifically, deleting a simple voxel does not result in splits or merges, nor does it create or eliminate loops, holes, or objects. This concept has been used to penalize errors involving non-simple voxels when segmenting neurons (Gornet et al. 2019). However, topological errors often involve multiple connected voxels, which is not addressed by a single voxel-based approach.

To overcome this limitation, we extend the concept of simple voxels to supervoxels (i.e., connected sets of voxels). We then propose a differentiable loss function based on this supervoxel characterization, which enables neural networks to be trained to minimize split and merge errors. Finally, we evaluate our approach on 3-d light microscopy images of mouse brains as well as on several 2-d segmentation datasets.

Related Works

Accurately segmenting fine-scale structures such as neurons, vessels, and roads from satellite images is a complex and well-studied problem. There are two general approaches: (1) improve feature representations (e.g., (Wu et al. 2017; Mosinska et al. 2018; Hu et al. 2023; Sheridan et al. 2023)) and (2) incorporate topology-inspired loss functions during training. The latter approach focuses on identifying critical locations where the neural network is prone to topological errors, and then using gradient-based updates to guide improvements.

Bertrand and Malandain (1994) studied topological characterization of simple voxels. The remarkable success of this work is that a complete topological description of a voxel can be derived through basic operations on its immediate neighborhood. Notably, this work does not invoke modern topological concepts such as homology or Morse theory to characterize voxels in low-dimensional digital topology. In many ways, the present study is inspired by that work.

The method most closely related to ours is centerlineDice (cDice), which is also a connectivity-aware loss func-

tion (Shit et al. 2021). In this approach, soft skeletons of both the prediction and ground truth are generated by applying min- and max-pooling k times during training. The loss is then computed by evaluating the overlap between the segmented objects and skeletons. Similar to our method, cDice also preserves topology up to homotopy equivalence – if the hyperparameter k is greater than the maximum observed radius.

Recently, persistent homology has been used in deep learning frameworks to track higher-order topological structures during training (Hofer et al. 2017, 2019; Chen et al. 2019; Stucki et al. 2023). Clough et al. (2019) used the Betti numbers of the ground truth as a topological prior and computed gradients to adjust the persistence of topological features. Hu et al. (2019) introduced a penalty for discrepancies between persistence diagrams of the prediction and ground truth. More recently, Hu et al. (2021) used discrete Morse theory to identify topologically significant structures and employed persistence-based pruning to refine them.

Turaga et al. (2009) introduced MALIS for affinity-based models to improve the predictions at maximin edges. This method involves costly gradient updates requiring a maximin search to identify voxels most likely to cause topological errors. Funke et al. (2018) introduced constrained MALIS to improve computational efficiency by computing gradients in two separate passes: one for affinities within ground truth objects and another for affinities between and outside these objects.

Several studies have leveraged the concept of non-simple voxels to develop topology-aware segmentation methods. Gornet et al. (2019) improved neuron segmentation by penalizing errors at non-simple voxels. Additionally, homotopy warping has been used to prioritize errors at non-simple voxels over minor misalignments at the boundaries (Jain et al. 2010; Hu 2022). However, these approaches are computationally expensive since they require voxel-by-voxel analysis to detect non-simple voxels for each prediction.

Method

Let $G = (V, E)$ be an undirected graph with the vertex set $V = \{1, \dots, n\}$. We assume that G is a graphical representation of an image where the vertices represent voxels and edges are defined with respect to a k -connectivity constraint¹. A ground truth segmentation $y = (y_1, \dots, y_n)$ is a labeling of the vertices with $y_i \in \{0, \dots, m\}$. Let $\hat{y} = (\hat{y}_1, \dots, \hat{y}_n)$ denotes a prediction of the ground truth.

Let $F(y) = \{i \in V : y_i \neq 0\}$ be the foreground of the vertex labeling, which may include multiple and potentially touching objects. Let $\mathcal{S}(y) \subseteq \mathcal{P}(V)$ be the set of connected components induced by the labeling y , where $\mathcal{P}(V)$ is the power set of V .

In a labeled graph, the connected components are determined by the equivalence relation that $i \sim j$ if and only if $y_i = y_j$ with $i, j \in F(y)$ and there exists a path from i to j that is entirely contained within the same segment. An

¹We assume that $k \in \{4, 8\}$ and $k \in \{6, 18, 26\}$ for 2-d and 3-d images, respectively (Kong and Rosenfeld 1989).

equivalence relation induces a partition over a set into equivalence classes that correspond to the connected components in this setting.

We propose a novel connectivity-preserving loss function to train topology-aware neural networks with the goal of avoiding false splits of, and false merges between foreground objects.

Definition 1. Let $\mathcal{L} : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$ be the loss function given by

$$\begin{aligned} \mathcal{L}(y, \hat{y}) = & (1 - \alpha) \mathcal{L}_0(y, \hat{y}) + \alpha \beta \sum_{C \in \mathcal{P}(\hat{y}_+)} \mathcal{L}_0(y_C, \hat{y}_C) \\ & + \alpha (1 - \beta) \sum_{C \in \mathcal{N}(\hat{y}_-)} \mathcal{L}_0(y_C, \hat{y}_C) \end{aligned}$$

such that $\alpha, \beta \in [0, 1]$ and \mathcal{L}_0 is an arbitrary loss function.

We build upon a traditional loss function \mathcal{L}_0 (e.g. cross-entropy or Dice coefficient) by incorporating additional terms that penalize sets of connected voxels (i.e. supervoxels) responsible for connectivity errors. These supervoxels are identified by analyzing connected components in the false negative and false positive masks, which are obtained by comparing the foregrounds of y and \hat{y}

The sets $\mathcal{N}(\hat{y}_-)$ and $\mathcal{P}(\hat{y}_+)$ consist of connected components whose addition or removal changes the number of connected components. A component that changes the underlying topology in this manner is called a *critical* component. Next, we rigorously define this notion, and present an algorithm for detecting critical components.

Critical Components

Critical components generalize the notion of non-simple voxels from digital topology to supervoxels. In digital topology, a voxel is called non-simple if its addition or removal changes the number of connected components, holes, or cavities. Similarly, a supervoxel is called critical if its addition or removal changes the number of connected components. We use the term *critical* as opposed to *non-simple* because the definition is not a direct generalization for computational reasons. This more focused definition enables our supervoxel-based loss function to be computed in linear time, which is a significant improvement over existing topological loss functions.

False Splits Let \hat{y}_- be the false negative mask determined by comparing a prediction to the ground truth. Let $\mathcal{S}_y(\hat{y}_-)$ be the set of connected components of \hat{y}_- with respect to y . The connected components in this set are defined by the following criteria: two voxels $i \sim j$ belong to the same component if:

- (i) $(\hat{y}_-)_i = (\hat{y}_-)_j$ with $i, j \in F(\hat{y}_-)$
- (ii) $y_i = y_j$
- (iii) There exists a path from i to j that lies within the same segments.

This second condition ensures that each component in the false negative mask corresponds precisely to one connected

component in the ground truth².

Negatively critical components are determined by comparing the number of connected components in y and $y \ominus C$ such that $C \in \mathcal{S}_y(\hat{y}_-)$. The notation $y \ominus C$ denotes “removing” a component from the ground truth. The result of this operation is a vertex labeling where the label of node $i \in V$ is given by

$$(y \ominus C)_i = \begin{cases} 0, & \text{if } i \in C \\ y_i, & \text{otherwise} \end{cases} \quad (1)$$

The removal of a component only impacts a specific region within the graph; the component itself and the nodes connected to its boundary. Thus, for topological characterization, it is sufficient to check whether its removal changes the number of connected components in that local region instead of the entire graph. Let $N(C) \subseteq V$ be the neighborhood surrounding a component $C \in \mathcal{S}(y)$ such that $N(C) = \{i \in V : \{i, j\} \in E \text{ and } j \in C\}$. Let $y \cap N(C)$ represent the labeling y within $N(C)$.

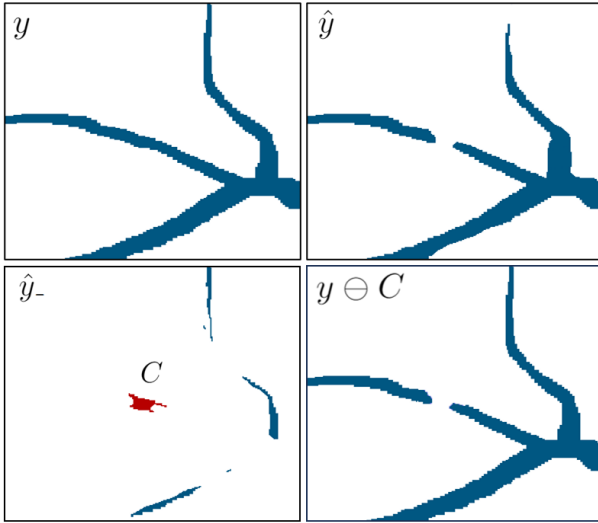


Figure 1: **Top:** Image patches of ground truth and predicted segmentations. **Bottom:** False negative mask with the component C highlighted in red. C is negatively critical since its removal changes the connectivity of the ground truth.

Definition 2. A component $C \in \mathcal{S}_y(\hat{y}_-)$ is said to be negatively critical if $|\mathcal{S}(y \cap N(C))| \neq |\mathcal{S}((y \ominus C) \cap N(C))|$.

Negatively critical components change the local topology by either deleting an entire component or altering the connectivity between vertices in $N(C)$ from the ground truth. In the latter case, the removal of such components locally disconnects some $i, j \in N(C)$ so that it is impossible to find a path (in this neighborhood) that does not pass through C . Based on this intuition, we can establish an equivalent definition of negatively critical components as components that are either (1) identical to a component in the ground truth

²Note that in binary segmentation, $\mathcal{S}(\hat{y}_-)$ is interchangeable with $\mathcal{S}_y(\hat{y}_-)$.

or (2) locally disconnect at least one pair of nodes in $N(C)$ after being removed.

Theorem 1. A component $C \in \mathcal{S}_y(\hat{y}_-)$ is negatively critical if and only if there exists an $A \in \mathcal{S}(y \cap N(C))$ with $A \supseteq C$ such that either (1) $A = C$ or (2) $\exists v_0, v_k \in A \setminus C$ such that there does not exist a path $(v_0, \dots, v_k) \subseteq N(C)$ with $v_i \notin C$ for $i = 1, \dots, k - 1$. (Proof is in the Appendix.)

One computational challenge in both characterizations is the need to recompute connected components within the neighborhood $N(C)$ for every $C \in \mathcal{S}_y(\hat{y}_-)$. In the worst case, the complexity is $\mathcal{O}(n^2)$ with respect to the number of voxels in the image. However, we can develop a more efficient algorithm with $\mathcal{O}(n)$ complexity by leveraging two useful facts: (i) neurons are tree-structured objects, implying that, (ii) negatively critical components change both the local and global topology.

Recall that a negatively critical component $C \in \mathcal{S}_y(\hat{y}_-)$ changes the local topology of $N(C)$ in the sense that $|\mathcal{S}(y \cap N(C))| \neq |\mathcal{S}((y \ominus C) \cap N(C))|$. Analogously, C also changes the global topology if $|\mathcal{S}(y)| \neq |\mathcal{S}(y \ominus C)|$. In this special case, we can establish an equivalent definition, similar to Theorem 1, that utilizes $\mathcal{S}(y)$ and $\mathcal{S}(y \ominus C)$ in place of $\mathcal{S}(y \cap N(C))$ and $\mathcal{S}((y \ominus C) \cap N(C))$.

This characterization can be further simplified by using $\mathcal{S}(y \ominus \hat{y}_-)$ instead of $\mathcal{S}(y \ominus C)$, where $y \ominus \hat{y}_-$ represents the ground truth after removing all components in the false negative mask:

$$(y \ominus \hat{y}_-)_i = \begin{cases} 0, & \text{if } (\hat{y}_-)_i = 1 \\ y_i, & \text{otherwise} \end{cases}$$

Note that this characterization is the key to overcoming nonlinear computational complexity.

Corollary 1. A component $C \in \mathcal{S}_y(\hat{y}_-)$ is negatively critical with $|\mathcal{S}(y)| \neq |\mathcal{S}(y \ominus C)|$ if and only if there exists an $A \in \mathcal{S}(y)$ with $A \supseteq C$ such that either (1) $A = C$ or (2) $\exists B_1, B_2 \in \mathcal{S}(y \ominus \hat{y}_-)$ with $B_1, B_2 \subset A$ such that $B_1 \cup C \cup B_2$ is connected. (Proof is in the Appendix.)

False Merges Let \hat{y}_+ be the false positive mask determined by comparing the prediction to the ground truth. Analogously, a component in the false positive mask is positively critical if its addition to the ground truth changes the number of connected components in the immediate neighborhood. While this notion can be articulated, for consistency and to leverage previous results, we opt for an equivalent formulation. Alternatively, a component in the false positive mask is positively critical if its removal from the predicted segmentation changes the topology.

Definition 3. A component $C \in \mathcal{S}_y(\hat{y}_+)$ is said to be positively critical if $|\mathcal{S}(\hat{y} \cap N(C))| \neq |\mathcal{S}(\hat{y} \ominus C \cap N(C))|$.

Positively critical components change the local topology by either (1) creating a component or (2) altering the connectivity between ground truth objects. In the latter, these components connect pairs of nodes that belong to locally distinct components. Equivalently, their removal causes pairs of nodes to become locally disconnected. Next, we present an equivalent definition that characterizes positively critical components as satisfying one of these conditions.

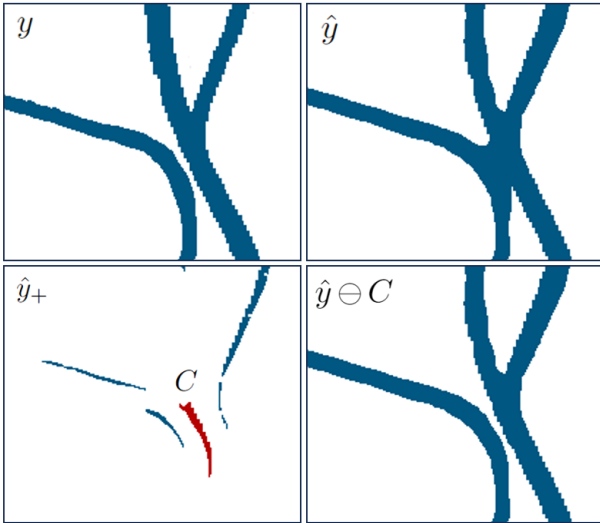


Figure 2: **Top:** Image patches of ground truth and predicted segmentations. **Bottom:** False positive mask with a single component C highlighted. C is positively critical since its removal changes the number of connected components.

Theorem 2. A component $C \in \mathcal{S}_y(\hat{y}_+)$ is positively critical if and only if there exists an $A \in \mathcal{S}(\hat{y})$ with $A \supseteq C$ such that either (1) $A = C$ or (2) $\exists v_0, v_k \in A \setminus C$ such that there does not exist a path $(v_0, \dots, v_k) \subseteq N(C)$ with $v_i \notin C$ for $i = 1, \dots, k - 1$. (Proof is in the Appendix.)

Similarly, positively critical components present the same computational challenge of needing to recompute connected components for every $C \in \mathcal{S}_y(\hat{y}_+)$. However, we can avoid this expensive calculation by utilizing a corollary of Theorem 2 that establishes an equivalent definition of positively critical components that also change the global topology. This characterization uses $\mathcal{S}(\hat{y})$ and $\mathcal{S}(\hat{y} \ominus \hat{y}_+)$, instead of $\mathcal{S}(y \cap N(C))$ and $\mathcal{S}((\hat{y} \ominus C) \cap N(C))$, where $\hat{y} \ominus \hat{y}_+$ denotes removing every component in the false positive mask from the prediction via

$$(\hat{y} \ominus \hat{y}_+)_i = \begin{cases} 0, & \text{if } (\hat{y}_+)_i = 1 \\ \hat{y}_i, & \text{otherwise} \end{cases}$$

Corollary 2. A component $C \in \mathcal{S}_y(\hat{y}_+)$ is positively critical with $|\mathcal{S}(\hat{y})| \neq |\mathcal{S}(\hat{y} \ominus C)|$ if and only if there exists an $A \in \mathcal{S}(\hat{y})$ with $A \supseteq C$ such that either (1) $A = C$ or (2) $\exists B_1, B_2 \in \mathcal{S}(\hat{y} \ominus \hat{y}_+)$ with $B_1, B_2 \subset A$ such that $B_1 \cup C \cup B_2$ is connected. (Proof is in the Appendix.)

Computing Critical Components

Although topological loss functions improve segmentation accuracy, one major drawback is that they are computationally expensive. A key advantage of our proposed method is that the runtime is $\mathcal{O}(n)$ with respect to the number of voxels. In contrast, related methods have either $\mathcal{O}(n \log n)$ or $\mathcal{O}(n^2)$ complexity (e.g. Turaga et al. (2009); Gornet et al. (2019); Hu et al. (2021); Shit et al. (2021); Hu et al. (2023)).

In the case of identifying non-simple voxels, Bertrand and Malandain (1994) proved that it is sufficient to examine the

topology of the neighborhood. Similarly, we can determine whether a component is critical by checking the topology of nodes connected to the boundary. For the remainder of this section, we focus the discussion on computing negatively critical components since the same algorithm can be used to compute positively critical components.

Let $D(C) = N(C) \setminus C$ be the set of nodes connected to the boundary of a component $C \in \mathcal{S}_y(\hat{y}_-)$. Assuming that a negatively critical component also changes the global topology, Corollary 1 can be used to establish analogous conditions on the set $D(C)$ that are useful for fast computation.

Corollary 3. A component $C \in \mathcal{S}_y(\hat{y}_-)$ is negatively critical with $|\mathcal{S}(y)| \neq |\mathcal{S}(y \ominus C)|$ if and only if $\exists A \in \mathcal{S}(y)$ with $A \supseteq C$ such that either (1) $\nexists i \in D(C)$ with $i \in A$ or (2) $\exists B_1, B_2 \in \mathcal{S}(y \ominus \hat{y}_-)$ with $B_1, B_2 \subset A$ such that $i \in B_1$ and $j \in B_2$ for some $i, j \in D(C)$. (Proof is in the Appendix.)

The key to achieving linear complexity is to precompute $\mathcal{S}(y)$ and $\mathcal{S}(y \ominus \hat{y}_-)$, then use a breadth-first search (BFS) to compute $\mathcal{S}_y(\hat{y}_-)$ while simultaneously checking Conditions 1 and 2. Intuitively, the core idea is that once this BFS reaches the boundary of the component, we can visit all nodes in $D(C)$ and efficiently check the conditions.

Let $r \in F(\hat{y}_-)$ be the root of the BFS. Given a node $j \in D(C)$, Conditions 1 and 2 can be efficiently checked using a hash table that stores the connected component label of j in $\mathcal{S}(y)$ and $\mathcal{S}(y \ominus \hat{y}_-)$ as a key-value pair. If we never visit a node $j \in D(C)$ with the same ground truth label as the root, then this label is not a key in the hash table and the component satisfies Condition 1 (Line 19, Algo. 2).

Now consider the case where we visit a node $j \in D(C)$ with the same ground truth label as the root. If the label of j in $\mathcal{S}(y)$ is not already a key in the hash table, a new entry is created (Line 15, Algo. 2). Otherwise, the value corresponding to this key is compared to the label of j in $\mathcal{S}(y \ominus \hat{y}_-)$. If they differ, then the component satisfies Condition 2.

Note that pseudocode for this method is provided in Algo. 1 and 2 and our code is publicly available at <https://github.com/AllenNeuralDynamics/supervoxel-loss>.

Theorem 3. The computational complexity of computing critical components that satisfy either $|\mathcal{S}(y)| \neq |\mathcal{S}(y \ominus C)|$ or $|\mathcal{S}(\hat{y})| \neq |\mathcal{S}(\hat{y} \ominus C)|$ is $\mathcal{O}(n)$ with respect to the number of voxels in the image. (Proof is in the Appendix.)

We emphasize that the statements and algorithms surrounding Theorem 3 are restricted to tree-structured objects (i.e. critical components that satisfy $|\mathcal{S}(y)| \neq |\mathcal{S}(y \ominus C)|$ or $|\mathcal{S}(\hat{y})| \neq |\mathcal{S}(\hat{y} \ominus C)|$). Indeed, a similar algorithm based on the main definitions and deductions can be implemented in a straightforward manner for the general case, except that this algorithm will be super-linear in complexity.

Penalizing Critical Topological Mistakes

Our topological loss function builds upon classical, voxel-based loss functions by adding terms that penalize critical components. The paradigm shift here is to evaluate each mistake at a “structure level” that transcends rectilinear geometry, as opposed to the voxel level, without resorting to expensive search iterations. In standard loss functions, mistakes are detected at the voxel-level by directly comparing

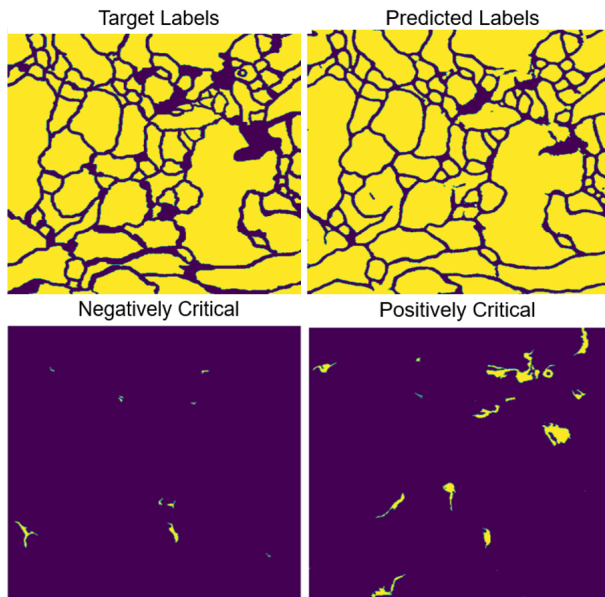


Figure 3: **Segmentation of a 512x512 image from the ISBI12 dataset.** The critical components of this prediction were computed in 1.41 seconds using code at <https://github.com/AllenNeuralDynamics/supervoxel-loss>.

the prediction at each voxel against the ground truth. Instead, we consider the context of each mistake by determining whether a given supervoxel causes a critical topological mistake.

The hyperparameter $\alpha \in [0, 1]$ is a scaling factor that controls the relative importance of voxel-level versus structure-level mistakes. Similarly, $\beta \in [0, 1]$ controls the weight placed on split versus merge mistakes. In situations where false merges are more costly or time-consuming to correct manually, β can be set to a value greater than 0.5 to prioritize preventing merge errors by assigning them higher penalties.

Our topological loss function is architecture agnostic and can be easily integrated into existing deep learning pipelines. We recommend training a baseline model with a standard loss function, then fine-tuning with the topological loss function. This topological function adds little computational overhead since the only additional calculation is to compute the critical components. In Theorem 3, we prove that Algorithms 1 and 2 can be used to compute critical components in linear time. This result can then be used to show that the computational complexity of computing our proposed topological loss function is also $\mathcal{O}(n)$ in the number of voxels in the image.

Experiments

We evaluate our method on the following image segmentation datasets: **DRIVE**, **ISBI12**, **CrackTree** and **EXASPIM**³. DRIVE is a retinal vessel dataset consisting of 20 images with dimensions 584x565 (Staal et al. 2004). ISBI12 is an electron microscopy (EM) dataset consisting of 29

³Download at <s3://aind-msma-morphology-data/EXASPIM25>

images with dimensions 512x512. CrackTree contains 206 images of cracks in roads, where the size of each image is 600x800. EXASPIM is a 3-d light sheet microscopy dataset consisting of 37 images whose dimensions range from 256x256x256 to 1024x1024x1024 and voxel size is $\sim 1 \mu\text{m}^3$ (Glaser et al. 2024).

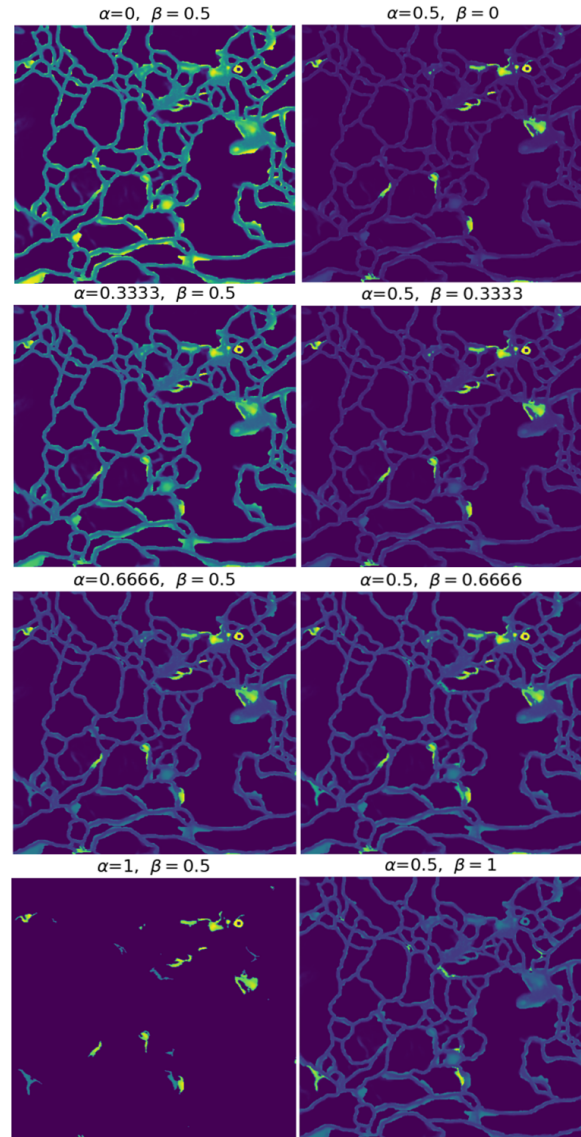


Figure 4: **Visualization of the loss in the prediction in Fig. 3.** **Left:** As α varies from 0 to 1, the loss places higher penalties on critical components. **Right:** As β varies from 0 to 1, the loss shifts from assigning higher penalties to negatively critical to positively critically components.

For the 2-d datasets, we perform 3-fold cross-validation for each method and report the mean and standard deviation across the validation set. For the 3-d dataset, we evaluate the methods on a test set consisting of 4 images. In all experiments, we set $\alpha = 0.5$ and $\beta = 0.5$ in our proposed topological loss function.

Evaluation Metrics. We use four evaluation metrics: pixel-wise accuracy, Dice coefficient, Adjusted Rand Index (ARI), Variation of Information (VOI) and Betti number error. The last three metrics are more topology-relevant in the sense that small topological differences can lead to significant changes in the error. See the Appendix for additional details on each metrics.

Baselines. For the 2-d datasets, we compare our method to **U-Net** (Ronneberger, Fischer, and Brox 2015), **Dive** (Fakhry, Peng, and Ji 2016), **Mosin.** (Mosinska et al. 2018), **TopoLoss** (Hu et al. 2019), and **DMT** (Hu et al. 2023). For the 3-d datasets, we compare our method to **U-Net** (Ronneberger, Fischer, and Brox 2015), **Gornet** (Gornet et al. 2019), **clDice** (Shit et al. 2021), and **MALIS** (Turaga et al. 2009). For the 2-d datasets, the segmentations were generated by applying a threshold of 0.5 to the predicted likelihoods. For the 3-d dataset, where some objects are touching, the segmentations were generated by applying a watershed algorithm to the prediction (Zlateski and Seung 2015).

Results. Table 1 shows quantitative results for the different models on the segmentation datasets. Our proposed method achieves state-of-the-art results; particularly, for the topologically relevant metrics. Figures 5 and 6 (full raw images and segmentations in the Appendix) present qualitative results that visually compare the performance of U-Net models trained with cross-entropy versus our proposed loss function. Although the only difference between these two models is the addition of topological terms, there is a clear difference in topological accuracy.

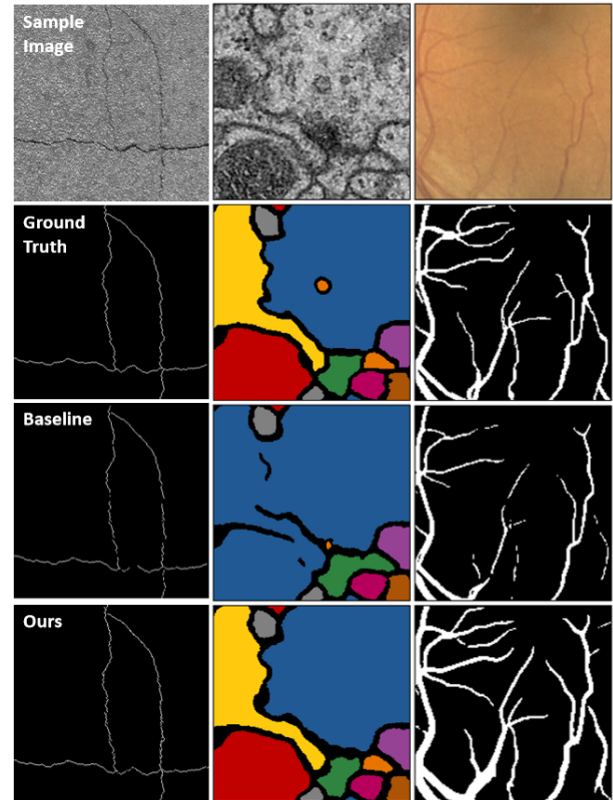


Figure 5: Qualitative results on three 2d datasets.

Method	$\mathcal{O}(\cdot)$	Dice \uparrow	ARI \uparrow	VOI \downarrow	Betti Error \downarrow
DRIVE					
U-Net	n	0.749 \pm 0.003	0.834 \pm 0.041	1.98 \pm 0.05	3.64 \pm 0.54
DIVE	n	0.754 \pm 0.001	0.841 \pm 0.026	1.94 \pm 0.13	3.28 \pm 0.64
Mosin.	n	0.722 \pm 0.001	0.887 \pm 0.039	1.17 \pm 0.03	2.78 \pm 0.29
TopoLoss	$n \log n$	0.762 \pm 0.004	0.902 \pm 0.011	1.08 \pm 0.01	1.08 \pm 0.27
DMT	n^2	0.773 \pm 0.004	0.902 \pm 0.002	0.88 \pm 0.04	0.87\pm0.40
Ours	n	0.809\pm0.012	0.943\pm0.002	0.48\pm0.01	0.94 \pm 0.27
ISBI12					
U-Net	n	0.970 \pm 0.005	0.934 \pm 0.007	1.37 \pm 0.03	2.79 \pm 0.27
DIVE	n	0.971 \pm 0.003	0.943 \pm 0.009	1.24 \pm 0.03	3.19 \pm 0.31
Mosin.	n	0.972 \pm 0.002	0.931 \pm 0.005	0.98 \pm 0.04	1.24 \pm 0.25
TopoLoss	$n \log n$	0.976 \pm 0.004	0.944 \pm 0.008	0.78 \pm 0.02	0.43 \pm 0.10
DMT	n^2	0.980 \pm 0.003	0.953\pm0.005	0.67\pm0.03	0.39\pm0.11
Ours	n	0.983\pm0.001	0.934 \pm 0.001	0.74 \pm 0.03	0.48 \pm 0.02
CrackTree					
U-Net	n	0.649 \pm 0.003	0.875 \pm 0.042	1.63 \pm 0.10	1.79 \pm 0.30
DIVE	n	0.653 \pm 0.002	0.863 \pm 0.0376	1.57 \pm 0.08	1.58 \pm 0.29
Mosin.	n	0.653 \pm 0.001	0.890 \pm 0.020	1.11 \pm 0.06	1.05 \pm 0.21
TopoLoss	$n \log n$	0.673 \pm 0.004	0.929 \pm 0.012	0.99 \pm 0.01	0.67 \pm 0.18
DMT	n^2	0.681\pm0.005	0.931\pm0.017	0.90\pm0.08	0.52 \pm 0.19
Ours	n	0.667 \pm 0.010	0.914 \pm 0.011	0.98 \pm 0.10	0.51\pm0.06
EXASPIIM					
U-Net	n	0.751 \pm 0.047	0.875 \pm 0.082	1.28 \pm 0.46	0.74 \pm 0.03
Gornet	n^2	0.777 \pm 0.083	0.901 \pm 0.049	0.65 \pm 0.17	0.42 \pm 0.07
clDice	kn	0.785 \pm 0.032	0.923 \pm 0.071	0.66 \pm 0.51	0.36 \pm 0.07
MALIS	n^2	0.794\pm0.052	0.927 \pm 0.042	0.64 \pm 0.27	0.34 \pm 0.08
Ours	n	0.770 \pm 0.058	0.953\pm0.038	0.42\pm0.21	0.31\pm0.06

Table 1: **Quantitative results for different models on several datasets.** Results for Dive, Mosin., TopoLoss, and DMT are from (Hu et al. 2023). $\mathcal{O}(\cdot)$: complexity of training iterations, n : number of pixels/voxels, k : number of pooling operations in clDice.

Evaluation with Skeleton-Based Metrics. Although the performance of segmentation models is typically evaluated using voxel-based metrics, such as those in Table 1, skeleton-based metrics are better suited to evaluate neuron segmentations. This is because the main objective of this task is to reconstruct the morphology of individual neurons and their interconnectivity. Thus, skeleton-based metrics are preferable because they quantify to what extent the topological structure of a neuron was accurately reconstructed.

In this evaluation, we compare a set of ground truth skeletons to the predicted segmentations to compute the following metrics: number of splits/neuron (Splits/Neuron), edge accuracy, and normalized expected run length (ERL) (see the Appendix for definitions). The number of merges per neuron was also computed, but the watershed parameters were configured to prevent merges in the segmentations.

Table 2 presents quantitative results for the different models on the EXASPIIM dataset. These results show that our proposed method achieves the highest topological accuracy. In particular, our proposed method has the best normalized ERL, which is regarded as the gold standard for evaluating the accuracy of a neuron’s topological reconstruction (Januszewski et al. 2018). Perhaps as importantly, our method scales linearly and achieves favorable runtimes, consistent with Thm. 3 (see Tables 2 and 4).

Ablation Study of Hyperparameters. The proposed topological loss function has two interpretable hyperparam-

Method	Complexity	Runtime/Epoch ↓	Splits/Neuron ↓	Edge Accuracy ↑	Normalized ERL ↑
U-Net	$\mathcal{O}(n)$	10.03±0.23 sec	9.86±13.30	0.873±0.087	0.596±0.232
Gornet	$\mathcal{O}(n^2)$	71.62±1.83 sec	3.85±2.58	0.937±0.062	0.664±0.106
clDice	$\mathcal{O}(kn)$	48.55±1.60 sec	3.39±1.52	0.911±0.042	0.701±0.091
MALIS	$\mathcal{O}(n^2)$	50.68±1.58 sec	3.33±0.59	0.917±0.053	0.719±0.098
Ours	$\mathcal{O}(n)$	20.12±1.15 sec	2.63±1.36	0.944±0.043	0.784±0.099

Table 2: Skeleton-based metrics for different models for the EXASPIM dataset.

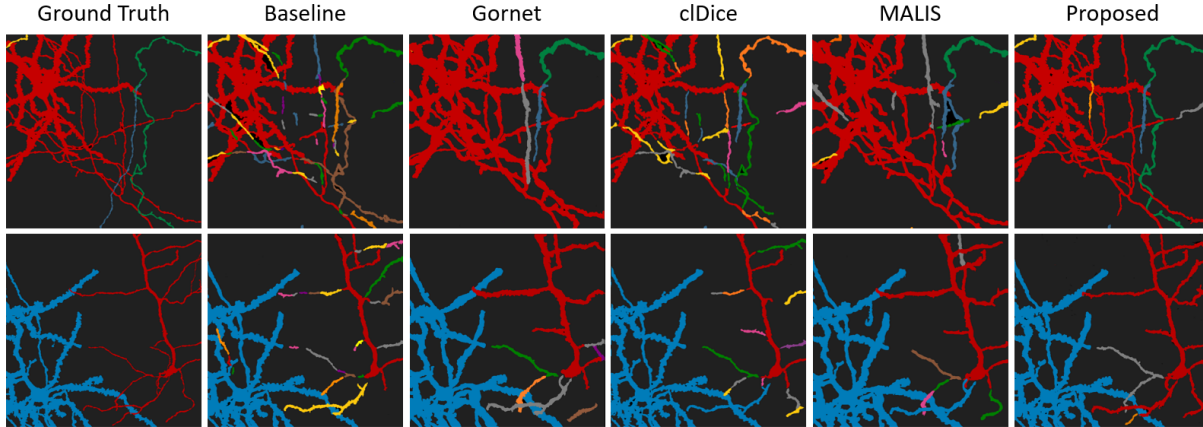


Figure 6: Qualitative results on the 3-d EXASPIM dataset. Raw Image: Fig. 13

eters: α and β . Generally, setting α to a positive value improves the topological accuracy of the segmentation. To better understand their impact on model performance, we performed hyperparameter tuning using Bayesian optimization with the adapted rand index as the objective function (Akiba et al. 2019). We then conducted two separate experiments, where we trained a U-Net using our proposed loss function on both the ISBI12 and EXASPIM datasets.

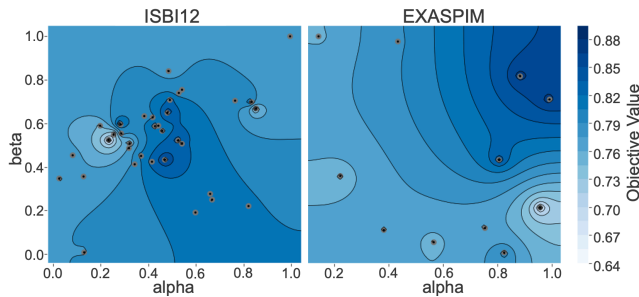


Figure 7: Contour plots of objective functions from hyperparameter optimization, each point is a trial outcome.

Although both datasets consist of neuronal images, they are markedly different: the EXASPIM dataset contains sparse images, while ISBI12 consists of dense ones. Our experiments show that $\alpha \approx 0.5$ and $\beta \approx 0.5$ are optimal for ISBI12, whereas $\alpha \approx 0.9$ and $\beta \approx 0.8$ are optimal for EXASPIM (Figure 7). A rule of thumb is to set $\alpha, \beta \approx 0.5$ for dense images and increase them together as the image becomes sparser. Notably, Figure 7 suggests that careful hy-

perparameter tuning is not critical in practice. In fact, all the results obtained by our method were achieved with a fixed choice for α and β , without using hyperparameter optimization.

Conclusion

Mistakes that change the connectivity of objects (i.e. topological mistakes) are a key problem in instance segmentation. They produce qualitatively different results based on relatively few pixels/voxels, making it a challenge to avoid such mistakes with voxel-level objectives. Existing work on topology-aware segmentation typically requires costly steps to guide the segmentation towards correct decisions, among other problems.

Here, we generalize the concept of simple voxel to connected components of arbitrary shape, and propose a topology-aware method with minimal computational overhead. We demonstrate the effectiveness of our approach on multiple datasets with different resolutions, dimensionality, and image characteristics. Across multiple metrics, our method achieves state-of-the-art results. It is now possible to image not only a local patch of neuronal morphology, but whole arbors, which can reach multiple, distant brain regions. The favorable scalability of our approach will enable efficient analysis of such large-scale datasets.

References

Akiba, T.; Sano, S.; Yanase, T.; Ohta, T.; and Koyama, M. 2019. Optuna: A Next-generation Hyperparameter Optimization Framework. In *Proceedings of the 25th ACM*

SIGKDD International Conference on Knowledge Discovery and Data Mining.

Bertrand, G.; and Malandain, G. 1994. A new characterization of three-dimensional simple points. *Pattern Recognition Letters*, 15(2): 169–175.

Chen, C.; Ni, X.; Bai, Q.; and Wang, Y. 2019. A topological regularizer for classifiers via persistent homology. In *The 22nd International Conference on Artificial Intelligence and Statistics*, 2573–2582. PMLR.

Clough, J. R.; Öksüz, I.; Byrne, N.; Schnabel, J. A.; and King, A. P. 2019. Explicit topological priors for deep-learning based image segmentation using persistent homology. In *Information Processing in Medical Imaging*.

Fakhry, A.; Peng, H.; and Ji, S. 2016. Deep models for brain EM image segmentation: novel insights and improved performance. *Bioinformatics*, 32(15): 2352–2358.

Funke, J.; Tschoop, F.; Grisaitis, W.; Sheridan, A.; Singh, C.; Saalfeld, S.; and 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(7): 1669–1680.

Glaser, A.; Chandrashekar, J.; Vasquez, J.; Arshadi, C.; Ouellette, N.; Jiang, X.; Baka, J.; Kovacs, G.; Woodard, M.; Seshamani, S.; Cao, K.; Clack, N.; Recknagel, A.; Grim, A.; Balaram, P.; Turschak, E.; Liddell, A.; Rohde, J.; Hellevik, A.; Takasaki, K.; Barner, L. E.; Logsdon, M.; Chronopoulos, C.; de Vries, S.; Ting, J.; Perlmutter, S.; Kalmbach, B.; Dembrow, N.; Reid, R. C.; Feng, D.; and Svoboda, K. 2024. Expansion-assisted selective plane illumination microscopy for nanoscale imaging of centimeter-scale tissues. *eLife Sciences Publications, Ltd*.

Gornet, J.; Venkataraju, K. U.; Narasimhan, A.; Turner, N.; Lee, K.; Seung, H. S.; Osten, P.; and Sümbül, U. 2019. Reconstructing neuronal anatomy from whole-brain images. In *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, 218–222. IEEE.

Gouwens, N. W.; Sorensen, S. A.; Baftizadeh, F.; Budzillo, A.; Lee, B. R.; Jarsky, T.; Alfiler, L.; Baker, K.; Barkan, E.; Berry, K.; et al. 2020. Integrated morphoelectric and transcriptomic classification of cortical GABAergic cells. *Cell*, 183(4): 935–953.

Hofer, C.; Kwitt, R.; Niethammer, M.; and Dixit, M. 2019. Connectivity-optimized representation learning via persistent homology. In *International conference on machine learning*, 2751–2760. PMLR.

Hofer, C.; Kwitt, R.; Niethammer, M.; and Uhl, A. 2017. Deep learning with topological signatures. *Advances in neural information processing systems*, 30.

Hu, X. 2022. Structure-aware image segmentation with homotopy warping. In *Advances in Neural Information Processing Systems (NeurIPS)*.

Hu, X.; Li, F.; Samaras, D.; and Chen, C. 2019. Topology-Preserving Deep Image Segmentation. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 32, 5657–5668.

Hu, X.; Li, F.; Samaras, D.; and Chen, C. 2023. Learning probabilistic topological representations using discrete morse theory. In *International Conference on Learned Representations (ICLR)*.

Hu, X.; Wang, Y.; Li, F.; Samaras, D.; and Chen, C. 2021. Topology-Aware Segmentation Using Discrete Morse Theory. In *International Conference on Learned Representations (ICLR)*.

Jain, V.; Bollmann, B.; Richardson, M.; Berger, D. R.; Helmstaedter, M. N.; Briggman, K. L.; Denk, W.; Bowden, J. B.; Mendenhall, J. M.; Abraham, W. C.; Harris, K. M.; Kasthuri, N.; Hayworth, K. J.; Schalek, R.; Tapia, J. C.; Lichtman, J. W.; and Seung, H. S. 2010. Boundary Learning by Optimization with Topological Constraints. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2488–2495.

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

Kong, T.; and Rosenfeld, A. 1989. Digital topology: Introduction and survey. *Computer Vision, Graphics, and Image Processing*, 48(3): 357–393.

Lee, K.; Turner, N.; Macrina, T.; Wu, J.; Lu, R.; and Seung, H. S. 2019. Convolutional nets for reconstructing neural circuits from brain images acquired by serial section electron microscopy. *Current Opinion in Neurobiology*, 55: 188–198.

Mosinska, A.; Marquez-Neila, P.; Kozinski, M.; and Fua, P. 2018. Beyond the Pixel-Wise Loss for Topology-Aware Delineation. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 3136–3145.

Ronneberger, O.; Fischer, P.; and Brox, T. 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, 234–241. Springer International Publishing.

Schmidt, M.; Motta, A.; Sievers, M.; and Helmstaedter, M. 2024. RoboEM: automated 3D flight tracing for synaptic-resolution connectomics. *Nature Methods*, 21: 1–6.

Sheridan, A.; Nguyen, T.; Deb, D.; Lee, W.-C. A.; Saalfeld, S.; Turaga, S.; Manor, U.; and Funke, J. 2023. Local shape descriptors for neuron segmentation. *Nature Methods*, 20: 295–303.

Shit, S.; Paetzold, J.; Sekuboyina, A.; Ezhov, I.; Unger, A.; Zhylyka, A.; Pluim, J.; Bauer, U.; and Menze, B. 2021. clDice - a Novel Topology-Preserving Loss Function for Tubular Structure Segmentation. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 16555–16564. IEEE Computer Society.

Staal, J.; Abramoff, M.; Niemeijer, M.; Viergever, M.; and van Ginneken, B. 2004. Ridge-based vessel segmentation in color images of the retina. *IEEE Transactions on Medical Imaging*, 23(4): 501–509.

Stucki, N.; Paetzold, J. C.; Shit, S.; Menze, B.; and Bauer, U. 2023. Topologically Faithful Image Segmentation via

Induced Matching of Persistence Barcodes. In *Proceedings of the 40th International Conference on Machine Learning*, 32698–32727. PMLR.

Turaga, S.; Briggman, K.; Helmstaedter, M.; Denk, W.; and Seung, S. 2009. Maximin affinity learning of image segmentation. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 22.

Turaga, S.; Murray, J.; Jain, V.; Roth, F.; Helmstaedter, M.; Briggman, K.; Denk, W.; and Seung, S. 2010. Convolutional Networks Can Learn to Generate Affinity Graphs for Image Segmentation. *Neural Computation*, 22(2): 511–538.

Winnubst, J.; Bas, E.; Ferreira, T. A.; Wu, Z.; Economo, M. N.; Edson, P.; Arthur, B. J.; Bruns, C.; Rokicki, K.; Schauder, D.; et al. 2019. Reconstruction of 1,000 projection neurons reveals new cell types and organization of long-range connectivity in the mouse brain. *Cell*, 179(1): 268–281.

Wu, P.; Chen, C.; Wang, Y.; Zhang, S.; Yuan, C.; Qian, Z.; Metaxas, D.; and Axel, L. 2017. Optimal Topological Cycles and Their Application in Cardiac Trabeculae Restoration. In *Information Processing in Medical Imaging*, 80–92. Springer International Publishing.

Zlateski, A.; and Seung, H. S. 2015. Image Segmentation by Size-Dependent Single Linkage Clustering of a Watershed Basin Graph. *ArXiv*.