

Topology-Preserving Deep Image Segmentation

*Xiaoling Hu¹, Li Fuxin², Dimitris Samaras¹ and Chao Chen¹

¹Stony Brook University

²Oregon State University

Abstract

Segmentation algorithms are prone to make topological errors on fine-scale structures, e.g., broken connections. We propose a novel method that learns to segment with correct topology. In particular, we design a **continuous-valued loss function** that enforces a segmentation to have the same topology as the ground truth, i.e., having the same Betti number. The proposed topology-preserving loss function is differentiable and we incorporate it into end-to-end training of a deep neural network. Our method achieves much better performance on the Betti number error, which directly accounts for the topological correctness. It also performs superiorly on other topology-relevant metrics, e.g., the Adjusted Rand Index and the Variation of Information. We illustrate the effectiveness of the proposed method on a broad spectrum of natural and biomedical datasets.

1 Introduction

Image segmentation, i.e., assigning labels to all pixels of an input image, is crucial in many computer vision tasks. State-of-the-art deep segmentation methods [23, 19, 8–10] learn high quality feature representations through an end-to-end trained deep network and achieve satisfactory per-pixel accuracy. However, these segmentation algorithms are still prone to make errors on fine-scale structures, such as small object instances, instances with multiple connected components, and thin connections. These fine-scale structures may be crucial in analyzing the *functionality* of the objects. For example, accurate extraction of thin parts such as ropes and handles is crucial in planning robot actions, e.g., dragging or grasping. In biomedical images, correct delineation of thin objects such as neuron membranes and vessels is crucial in providing accurate morphological and structural quantification of the underlying system. A broken connection or a missing component may only induce marginal per-pixel error, but can cause catastrophic functional mistakes. See Fig. 1 for an example.

We propose TopoNet, a novel deep segmentation method that *learns to segment with correct topology*. In particular, we propose a *topological loss* that enforces the segmentation results to have the same topology as the ground truth, i.e., having the same *Betti number* (number of connected components and handles). A neural network trained with such loss will achieve high topological fidelity without sacrificing per-pixel accuracy. The main challenge in designing such loss is that topological information, namely, Betti numbers, are discrete values. We need a continuous-valued measurement of the topological similarity between a prediction and the ground truth; and such measurement needs to be differentiable in order to backpropagate through the network.

To this end, we propose to use theory from computational topology [13], which summarizes the topological information from a continuous-valued function (in our case, the likelihood function f is predicted by a neural network). Instead of acquiring the segmentation by thresholding f at 0.5 and inspecting its topology, *persistent homology* [13, 14, 40] captures topological information carried by f over all possible thresholds. This provides a unified, differentiable approach of measuring the

*Correspondence to: Xiaoling Hu <xiaolhu@cs.stonybrook.edu>.

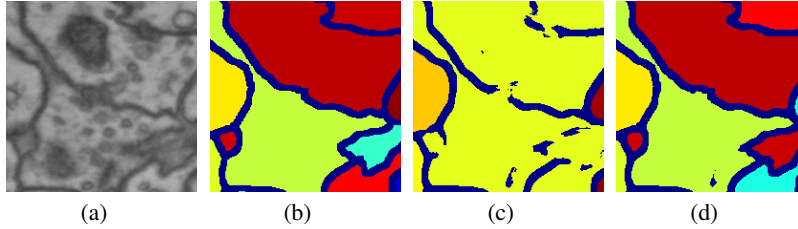


Figure 1: Illustration of the importance of topological correctness in a neuron image segmentation task. The goal of this task is to segment membranes which partition the image into regions corresponding to neurons. (a) an input neuron image. (b) ground truth segmentation of the membranes (dark blue) and the result neuron regions. (c) result of a baseline method without topological guarantee [16]. Small pixel-wise errors lead to broken membranes, resulting in merging of many neurons into one. (d) Our method produces the correct topology and the correct partitioning of neurons.

topological similarity between f and the ground truth, called the topological loss. We derive the gradient of the loss so that the network predicting the likelihood function can be optimized w.r.t. the topological loss.

TopoNet is the first end-to-end deep segmentation network with guaranteed topological correctness. We show that when the topological loss is decreased to zero, the segmentation is guaranteed to be topologically correct, i.e., have identical topology as the ground truth. Our method is empirically validated by comparing with state-of-the-arts on natural and biomedical datasets with fine-scale structures. It achieves superior performance on metrics that encourage structural accuracy. In particular, our method significantly outperforms others on the Betti number error which exactly measures the topological accuracy. Fig. 1 shows a qualitative result.

Our method demonstrates how topological computation and deep learning can be mutually beneficial. While our method empowers deep neural network with topological constraints, it also can be seen as a powerful approach on topological analysis in that the observation function is now learned with a highly nonlinear deep network. This enables topology to be estimated based on a semantically informed and denoised observation function.

Related work. The closest method to ours is by Mosinska *et al.* [25], which also proposes a topology-aware loss. Instead of actually computing and comparing the topology, their approach uses the response of selected filters from a pretrained VGG19 network to construct the loss. These filters prefer elongated shapes and thus alleviate the broken connection issue. But this method is hard to generalize to more complex settings with connections of arbitrary shapes. Furthermore, even if this method achieves zero loss, its segmentation is not guaranteed to be topologically correct.

Different ideas have been proposed to capture fine details of objects, mostly revolving around deconvolution and upsampling [23, 8–10, 27, 32]. However these methods focus on the prediction accuracy of individual pixels and are intrinsically topology-agnostic. Topological constraints, e.g., connectivity and loop-freeness, have been incorporated into variational [18, 22, 36, 33] and MRF/CRF-based segmentation methods [38, 28, 39, 6, 2, 35, 29, 15]. However, these methods focus on enforcing topological constraints in the inference stage, while the trained model is agnostic of the topological prior. In neuron image segmentation, some methods [17, 37] directly find an optimal partition of the image into neurons, and thus avoid segmenting membranes. These methods cannot be generalized to other problems when the object of interest may not be closed loops, e.g., vessels, cracks and roads.

For completeness, we also refer to other existing works on extracting topological features and training kernel classifiers [1, 31, 21, 5, 7]. In graphics, topological similarity was used to simplify and align shapes [30]. As for deep neural networks, Hofer *et al.* [20] proposed a CNN-based topological classifier. This method directly extracts topological information from an input image/shape/graph as input for CNN, hence cannot generate segmentations that preserve topological priors learned from the training set. To the best of our knowledge, no existing work uses topological information as a loss for training an end-to-end deep neural network.

2 TopoNet

Our method, TopoNet, achieves both per-pixel accuracy and topological correctness by training a deep neural network with a new topological loss, $L_{topo}(f, g)$. Here f is the likelihood map predicted

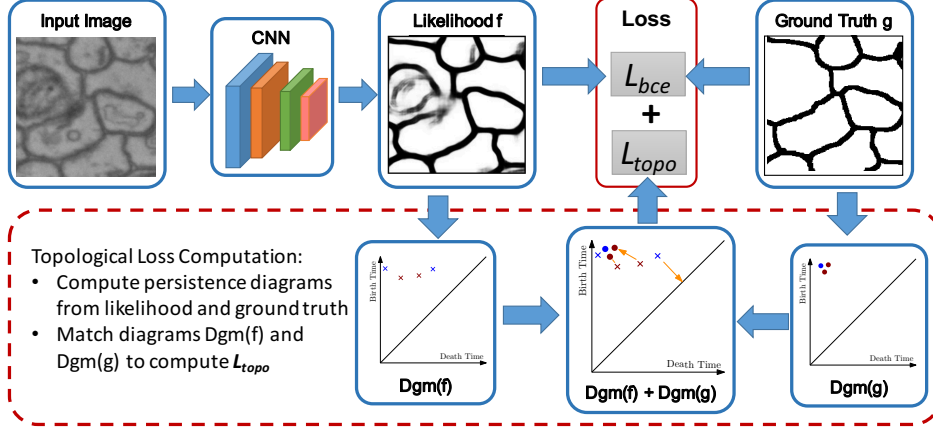


Figure 2: An overview of our method.

by the network and g is the ground truth. The loss function on each training image is a weighted sum of the per-pixel cross-entropy loss, L_{bce} , and the topological loss:

$$L(f, g) = L_{bce}(f, g) + \lambda L_{topo}(f, g), \quad (2.1)$$

in which λ controls the weight of the topological loss. We assume a binary segmentation task. Thus, there is one single likelihood function f , whose value ranges between 0 and 1.

In Sec. 2.1, we introduce the mathematical foundation of topology and how to measure topology of a likelihood map robustly using persistent homology. In Sec. 2.2, we formalize the topological loss as the difference between persistent homology of f and g . We derive the gradient of the loss and prove its correctness. In Sec. 2.3 we explain how to incorporate the loss into the training of a neural network. Although we fix one architecture in experiments, our method is general and can use any neural network that provides pixel-wise prediction. Fig. 2 illustrates the overview of our method.

2.1 Topology and Persistent Homology

Given a continuous image domain, $\Omega \subseteq \mathbb{R}^2$ (e.g., a 2D rectangle), we study a likelihood map $f(x, \omega) : \Omega \rightarrow \mathbb{R}$, which is predicted by a deep neural network (Fig. 3(c)). Note that in practice, we only have samples of f at all pixels. In such case, we extend f to the whole image domain Ω by linear interpolation. Therefore, f is piecewise-linear and is controlled by values at all pixels. A segmentation, $X \subseteq \Omega$ (Fig. 3(a)), is calculated by thresholding f at a given value α (often set to 0.5).

Given X , its d -dimension topological structure, called a *homology class* [13, 26], is an equivalence class of d -manifolds which can be deformed into each other within X .² In particular, 0-dim and 1-dim structures are connected components and handles, respectively. For example, in Fig. 3(a), the segmentation X has two connected components and one handle. Meanwhile, the ground truth (Fig. 3(b)) has one connected component and two handles. Given X , we can compute the number of topological structures, called the *Betti number*, and compare it with the topology of the ground truth.

However, simply comparing Betti numbers of X and g will result in a discrete number. In order to incorporate topological prior into deep neural networks, we need to construct a continuous-valued function that can reveal subtle difference between similar structures. Fig. 3(c) and 3(d) show two likelihood maps f and f' with identical segmentations, both different in topology w.r.t. the ground truth g (Fig. 3(b)). However, f is more preferable as we need much less effort to change it so that the thresholded segmentation X is the same as g . In particular, look closely to Fig. 3(c) and 3(d) near the broken handle and view the magnified view of the function. To restore the broken handle in Fig. 3(d), we need to spend more effort to fill a much deeper gap than Fig. 3(c). The same situation happens near the missing bridge between the two connected components.

To capture such subtle structural difference between different likelihood maps, we need a holistic view. In particular, we use the theory of *persistent homology*. Instead of choosing a fixed threshold, persistent homology theory captures all possible topological structures from all thresholds, and summarize all these information in a concise format, called *persistence diagram*.

²To be exact, a homology class is an equivalent class of cycles whose difference is the boundary of a $(d + 1)$ -dimensional patch.

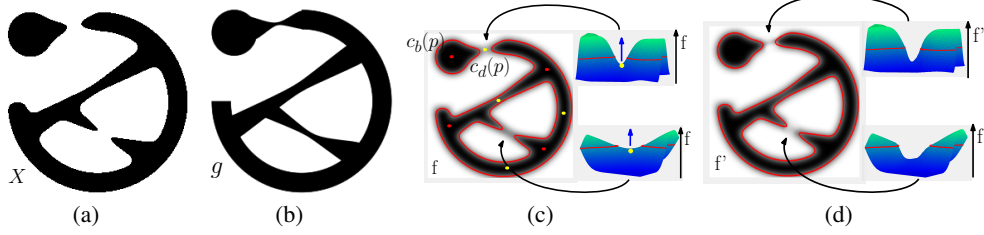


Figure 3: Illustration of topology and topology of a likelihood. For visualization purposes, the higher the function values are, the darker the area is. **(a)** an example segmentation X with two connected components and one handle. **(b)** The ground truth with one connected component and two handles. It can also be viewed as a binary valued function g . **(c)** a likelihood map f whose segmentation (bounded by the red curve) is X . The landscape views near the broken bridge and handle are drawn. Critical points are highlighted in the segmentation. **(d)** another likelihood map f' with the same segmentation as f . But the landscape views reveal that f' is worse than f due to deeper gaps.

All different thresholds α constitute a filtration, i.e., a monotonically growing sequence induced by decreasing the threshold $\alpha : \emptyset \subseteq f^{\alpha_1} \subseteq f^{\alpha_2} \subseteq \dots \subseteq f^{\alpha_n} = \Omega$, where $\alpha_1 \geq \alpha_2 \geq \dots \geq \alpha_n$, and $f^\alpha := \{x \in \Omega | f(x) \geq \alpha\}$. As α decreases, the topology of f^α changes. That means some new topological structures are born while other existing topological structures are killed. When $\alpha < \alpha_n$, only one connected component survives and never gets killed. See Fig. 4(a) and 4(d) for example filtrations induced by the ground truth function g and the likelihood function f .

For a continuous-valued function f , its *persistence diagram*, $\text{Dgm}(f)$, contains a finite number of dots in 2-dimensional plane, called *persistent dots*. Each persistent dot $p = (d, b) \in \text{Dgm}(f)$ corresponds to a topological structure which is born at time/threshold b and killed at time/threshold d . Denote $\text{birth}(p) = b$ and $\text{death}(p) = d$. Fig. 4(b) and 4(e) show the diagrams of g and f , respectively.³ Instead of comparing discrete Betti numbers, we can use the information from persistence diagrams to compare a likelihood f with the ground truth g in terms of topology.

2.2 Topological loss and the Gradient

We are now ready to formalize the topological loss, which measures the topological similarity between the likelihood f and the ground truth g . We abuse the notation and also view g as a binary valued function. We use the dots in the persistence diagram of f as they capture all possible topological structures f potentially has. We slightly modified the *Wasserstein distance* for persistence diagram [12]. For persistence diagrams $\text{Dgm}(f)$ and $\text{Dgm}(g)$, we find a best one-to-one correspondence between the two sets of dots, and measure the total squared distance between them.⁴ Let Γ be the set of all possible bijections between $\text{Dgm}(f)$ and $\text{Dgm}(g)$. The topological loss $L_{\text{topo}}(f, g)$ is:

$$\min_{\gamma \in \Gamma} \sum_{p \in \text{Dgm}(f)} \|p - \gamma(p)\|^2 = \sum_{p \in \text{Dgm}(f)} [\text{birth}(p) - \text{birth}(\gamma^*(p))]^2 + [\text{death}(p) - \text{death}(\gamma^*(p))]^2 \quad (2.2)$$

where γ^* is the optimal matching between two different point sets.

Intuitively, this loss measures the minimal amount of necessary effort to modify the diagram of $\text{Dgm}(f)$ to $\text{Dgm}(g)$ by moving all dots toward their matches. Note there are more dots in $\text{Dgm}(f)$ (Fig. 4(c)) than in $\text{Dgm}(g)$ (Fig. 4(b)); there will usually be some noise in predicted likelihood map. If a dot cannot be matched, we match it to its projection on the diagonal line, $\{(d, b) | d = b\}$. This means we consider it as noise that should be removed. In this example, the extra connected component (a blue cross) in $\text{Dgm}(f)$ will be removed. For comparison, we also show in Fig. 4(f) the matching between diagrams of the worse likelihood f' and g . The cost of the matching is obviously higher, i.e., $L_{\text{topo}}(f', g) > L_{\text{topo}}(f, g)$. As a theoretical reassurance, it has been proven that this metric for diagrams is stable, and the loss function $L_{\text{topo}}(f, g)$ is Lipschitz with regard to the likelihood function f [11].

³Unlike traditional setting, we use the birth time as the y axis and the death time as the x axis. This is because we are using an upperstar filtration, i.e., using the superlevel set, and decreasing α value.

⁴To be exact, the matching needs to be done on separate dimensions. Dots of 0-dim structures (blue markers in Fig. 4(b) and 4(e)) should be matched to the diagram of 0-dim structures. Dots of 1-dim structures (red markers in Fig. 4(b) and 4(e)) should be matched to the diagram of 1-dim structures.

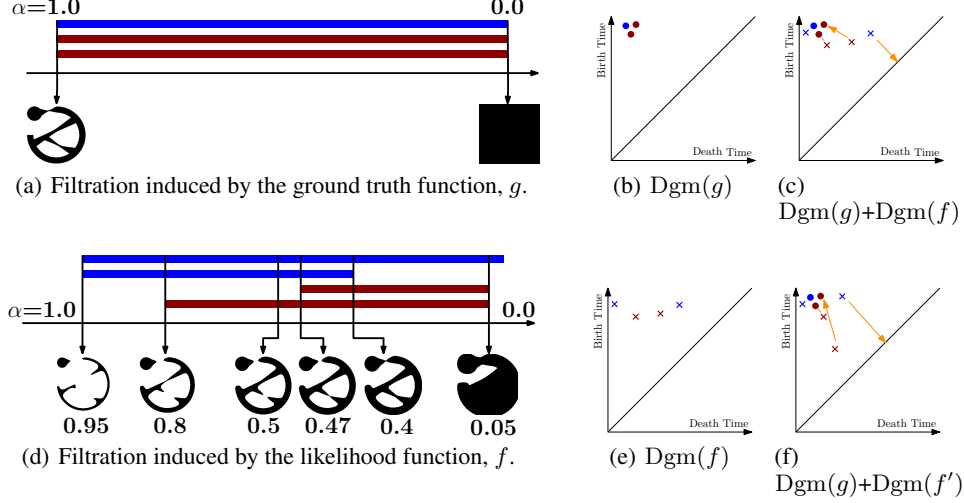


Figure 4: An illustration of persistent homology. **Left** the filtrations on the ground truth function g and the likelihood function f . The bars of blue and burgundy colors are connected components and handles respectively. **(a)** For g , all structures are born at $\alpha = 1.0$ and die at $\alpha = 0$. **(d)** For f , from left to right, birth of two components, birth of the longer handle, segmentation at $\alpha = 0.5$, birth of the shorter handle, death of the extra component, death of both handles. **(b)** and **(e)** the persistence diagrams of g and f . **(c)** the overlay of the two diagrams. Orange arrows denote the matching between the persistent dots. The extra component (a blue cross) from the likelihood is matched to the diagonal line and will be removed if we move $Dgm(f)$ to $Dgm(g)$. **(f)** the overlay of the diagrams of g and the worse likelihood $Dgm(f')$. The matching is obviously more expensive.

We state the following theorem regarding the correctness. The proof is relatively straightforward.

Theorem 1 (Topological Correctness). *When the loss function $L_{topo}(f, g)$ is zero, the segmentation by thresholding f at 0.5 has the same Betti number as g .*

Proof. Assume $L_{topo}(f, g)$ is zero. By Eq. (2.2), $Dgm(f)$ and $Dgm(g)$ are matched perfectly, i.e., $p = \gamma^*(p), \forall p \in Dgm(f)$. The two diagrams are identical and have the same number of dots.

Since g is a binary-valued function, as we decrease the threshold α continuously, all topological structures are created at $\alpha = 1$. All of them except one are killed at $\alpha = 0$. One remaining connected component will survive forever. The number of topological structures (Betti number) of g^α for any $0 < \alpha < 1$ is the same as the number of dots in $Dgm(g)$ plus 1. Note that for any $\alpha \in (0, 1)$, g^α is the ground truth segmentation. Therefore, the Betti number of the ground truth is the number of dots in $Dgm(g)$. Similarly, for any $\alpha \in (0, 1)$, the Betti number of f^α equals to the number of dots in $Dgm(f)$ plus 1. Since the two diagrams $Dgm(f)$ and $Dgm(g)$ are identical, the Betti number of the segmentation $f^{0.5}$ is the same as the ground truth segmentation.⁵ \square

Topological gradient. The loss function (Eq. (2.2)) depends on crucial thresholds at which topological changes happen, e.g., birth and death times of different dots in the diagram. These crucial thresholds are uniquely determined by the locations at which the topological changes happen. When the underlying function f is differentiable, these crucial locations are exactly *critical points*, i.e., points with zero gradients. In the training context, our likelihood function f is a piecewise-linear function controlled by the neural network predictions at pixels. For such f , a critical point is always a pixel, since topological changes always happen at pixels. Denote by ω the neural network parameters. For each dot $p \in Dgm(f)$, we denote by $c_b(p)$ and $c_d(p)$ the birth and death critical points of the corresponding topological structure (See Fig. 3(c) for examples).

Formally, we can show that the gradient of the topological loss $\nabla_\omega L_{topo}(f, g)$ is:

$$\sum_{p \in Dgm(f)} 2[f(c_b(p)) - \text{birth}(\gamma^*(p))] \frac{\partial f(c_b(p))}{\partial \omega} + 2[f(c_d(p)) - \text{death}(\gamma^*(p))] \frac{\partial f(c_d(p))}{\partial \omega} \quad (2.3)$$

To see this, within a sufficiently small neighborhood of f , any other piecewise linear function will have the same super level set filtration as f . The critical points of each persistent dot in $Dgm(f)$

⁵Note that a more careful proof should be done for diagrams of different dimensions separately.

remains constant within such small neighborhood. So does the optimal mapping γ^* . Therefore, the gradient can be straightforwardly computed based on the chain rule, as Eq. (2.3). When function values at different vertices are the same, the gradient does not exist. However, these cases constitute a measure zero subspace in the space of likelihood functions. In summary, $L_{topo}(f, g)$ is a piecewise differentiable loss function over the space of all possible likelihood functions f .

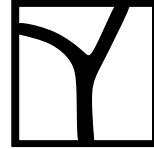
Intuition. During training, we take the negative gradient direction, i.e., $-\nabla_{\omega} L_{topo}(f, g)$. For each topological structure the gradient descent step is pushing the corresponding dot $p \in \text{Dgm}(f)$ toward its match $\gamma^*(p) \in \text{Dgm}(g)$. These coordinates are the function values of the critical points $c_b(p)$ and $c_d(p)$. They are both moved closer to the matched persistent dot in $\text{Dgm}(g)$. We also show the negative gradient force in the landscape view of function f (blue arrow in Fig. 3(c)). Intuitively, force from the topological gradient will push the saddle points up so that the broken bridge gets connected.

2.3 Training a Neural Network

We present some crucial details of our training algorithm. Although our method is architecture-agnostic, we select one architecture inspired by DIVE [16], which was designed for neuron image segmentation tasks. Our network contains six trainable weight layers, four convolutional layers and two fully connected layers. The first, second and fourth convolutional layers are followed by a single max pooling layer of size 2×2 and stride 2 by the end of the layer. Particularly, because of the computational complexity, we use a patch size of 65×65 during all the training process.

We use small patches (65×65) instead of big patches/whole image. The reason is two-folds. First, the computational complexity of topological information is relative high. The computational complexity of persistent homology is $O(n^3)$, where n is the patch size. Second, the matching process between the persistence diagrams of predicted likelihood map and ground truth can be quite difficult. For example, if the patch size is too big, there will be many persistent dots in Fig. 4(b) and even more dots in Fig. 4(e). The matching process in Fig. 4(c) is too complex and prone to errors. *By focusing on smaller patches, we localize topological structures and fix them one by one.*

Topology of small patches and relative homology. The small patches (65×65) often only contain partial branching structures rather than closed loops. To have meaningful topological measure on these small patches, we apply *relative persistent homology* as a more localized approach for the computation of topological structures. Particularly, for each patch, we consider the topological structures relative to the boundary. It is equivalent to padding a black frame to the boundary and compute the topology to avoid trivial topological structures. As shown in the figure on the right, with the additional frame, a Y-shaped branching structure cropped within the patch will create two handles and be captured by persistent homology. Training using these localized topological loss can be very efficient via random patch sampling.



Gradient of the topological loss. Eq. (2.3) shows that the topological loss function is defined on *critical pixels*. In each iteration, the predicted likelihood map from the neural network is different. Therefore, the topological structure and *persistent dots* will also be quite different. The topological loss is computed based on different pixels (*critical points*) at each iteration.

3 Experiments

We evaluate our method on six natural and biomedical datasets: **CREMI**⁶, **ISBI12** [4], **ISBI13** [3], **CrackTree** [41], **Road** [24] and **DRIVE** [34]. The first three are neuron image segmentation datasets. **CREMI** contains 125 images with resolution 1250×1250 with $4 \times 4 \times 40\text{nm}$. We use a three fold cross validation and report the mean performance over the validation set. Same setting applies to the other datasets. **ISBI12** [4] contains 30 images. The resolution is 512×512 with $4 \times 4 \times 50\text{nm}$. **ISBI13** [3] contains 100 images. The resolution is 1024×1024 with $3 \times 3 \times 30\text{nm}$. These three datasets are neuron images (Electron Microscopy images). The task is to segment membranes and eventually partition the image into neuron regions. **CrackTree** [41] contains 206 images of cracks in road. The resolution is 600×800 . **Road** [24] has 1108 images from the Massachusetts Roads Dataset. The resolution is 1500×1500 . **DRIVE** [34] is a retinal vessel segmentation dataset with 20 images. The resolution is 584×565 .

⁶<https://cremi.org/>

Table 1: Quantitative results for different models on several medical datasets

| Dataset | Method | Accuracy | ARI | VOI | Betti Error |
|---------|---------|---------------|---------------|--------------|--------------|
| ISBI12 | DIVE | 0.9640 | 0.9434 | 1.235 | 3.187 |
| | U-Net | 0.9678 | 0.9338 | 1.367 | 2.785 |
| | Mosin. | 0.9532 | 0.9312 | 0.983 | 1.238 |
| | TopoNet | 0.9626 | 0.9444 | 0.782 | 0.429 |
| ISBI13 | DIVE | 0.9642 | 0.6923 | 2.790 | 3.875 |
| | U-Net | 0.9631 | 0.7031 | 2.583 | 3.463 |
| | Mosin. | 0.9578 | 0.7483 | 1.534 | 2.952 |
| | TopoNet | 0.9569 | 0.8064 | 1.436 | 1.253 |
| CREMI | DIVE | 0.9498 | 0.6532 | 2.513 | 4.378 |
| | U-Net | 0.9468 | 0.6723 | 2.346 | 3.016 |
| | Mosin. | 0.9467 | 0.7853 | 1.623 | 1.973 |
| | TopoNet | 0.9456 | 0.8083 | 1.462 | 1.113 |

Table 2: Quantitative results for different models on some other datasets

| Dataset | Method | Accuracy | ARI | VOI | Betti Error |
|-----------|---------|---------------|---------------|--------------|--------------|
| DRIVE | DIVE | 0.9549 | 0.8407 | 1.936 | 3.276 |
| | U-Net | 0.9452 | 0.8343 | 1.975 | 3.643 |
| | Mosin. | 0.9543 | 0.8870 | 1.167 | 2.784 |
| | TopoNet | 0.9521 | 0.9024 | 1.083 | 1.076 |
| CrackTree | DIVE | 0.9854 | 0.8634 | 1.570 | 1.576 |
| | U-Net | 0.9821 | 0.8749 | 1.625 | 1.785 |
| | Mosin. | 0.9833 | 0.8897 | 1.113 | 1.045 |
| | TopoNet | 0.9826 | 0.9291 | 0.997 | 0.672 |
| Road | DIVE | 0.9734 | 0.8201 | 2.368 | 3.598 |
| | U-Net | 0.9786 | 0.8189 | 2.249 | 3.439 |
| | Mosin. | 0.9754 | 0.8456 | 1.457 | 2.781 |
| | TopoNet | 0.9728 | 0.8671 | 1.234 | 1.275 |

Evaluation metrics. We use four different evaluation metrics. **Pixel-wise accuracy** is the percentage of correctly classified pixels. The remaining three metrics are more topology-relevant. The most important one is **Betti number error**, which directly compare the topology (number of handles) between the segmentation and the ground truth⁷. We randomly sample patches over the segmentation and report the average absolute difference between their Betti numbers and the corresponding ground truth patches. Two more metrics are used to indirectly evaluate the topological correctness: **Adapted Rand Index (ARI)** and **Variation of Information (VOI)**. They are often used in neuron image segmentation to compare the partitioning of the image induced by the segmentation. **ARI** is defined as the maximal F-score of the foreground-restricted Rand index, a measure of similarity between two clusters. On this version of the Rand index we exclude the zero component of the original labels (background pixels of the ground truth). **VOI** is a measure of the distance between two clusterings. It is closely related to mutual information; indeed, it is a simple linear expression involving the mutual information.

Baselines. **DIVE** [16] is a state-of-the-art neural network that predicts the probability of every individual pixel in a given image being a membrane (border) pixel or not. **U-Net** [32] is a popular image segmentation method trained with cross-entropy loss. **Mosin.** [25] uses the response of selected filters from a pretrained CNN to construct the topology aware loss.

Quantitative and qualitative results. As stated above, we use pixel accuracy, and three topology-aware metrics: ARI, VOI and Betti Error. Tab. 1 shows the quantitative results for three different neuron image datasets, ISBI12, ISBI13 and CREMI. Tab. 2 shows the quantitative results for DRIVE, CrackTree and Road. We demonstrate that our method significantly outperforms existing methods in topological accuracy (in all three topology-aware metrics), without sacrificing pixel accuracy. Fig. 5 shows qualitative results. TopoNet demonstrates more consistency in terms of structures and topology. It correctly segments fine structures such as membranes, roads and vessels, while all other methods fail to do so.

⁷Note we focus on 1-dimensional topology in evaluation and training as they are more crucial in practice.

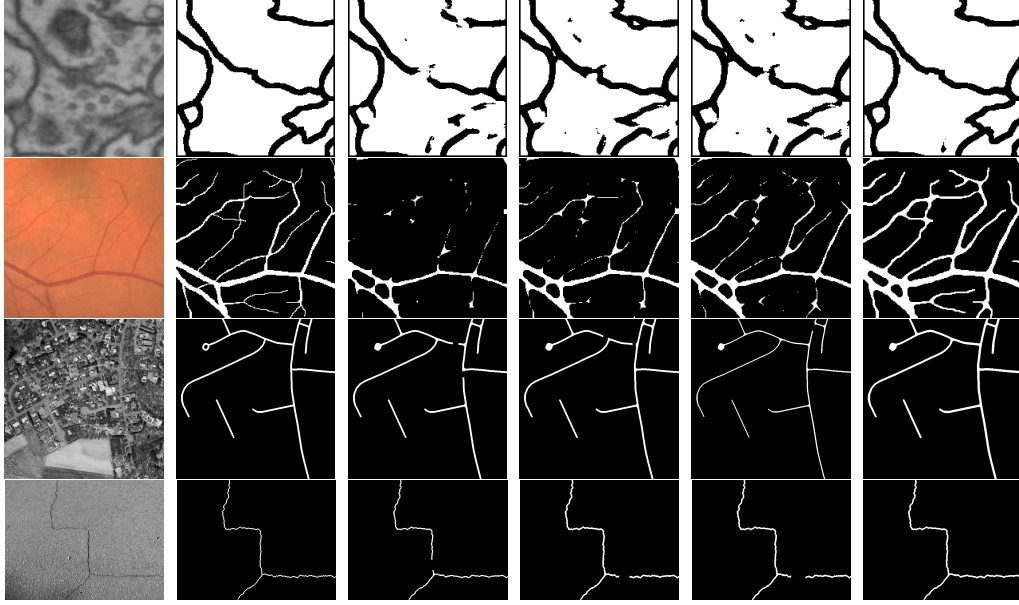


Figure 5: Qualitative results of the proposed method compared to other models. From left to right, sample images, ground truth, results for **DIVE**, **U-Net**, **Mosin**, and **our proposed TopoNet**.

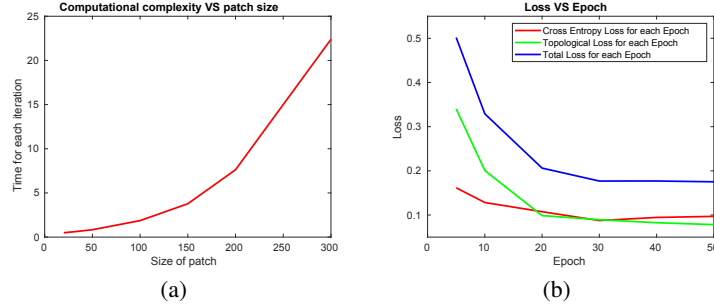


Figure 6: **(a)** Topology computation complexity in terms of patch size. **(b)** Cross Entropy loss, Topological loss and total loss in terms of training epochs.

Computational time vs. patch size. When incorporating our topological loss into deep neural networks, we test our algorithm with different patch sizes. During the training process, for each predicted likelihood map, we are supposed to compute its persistence diagram and the corresponding ground truth persistence diagram. The bigger the patch size is, the longer time it will take for this computation. The matching process will also be more complex and prone to errors (Sec. 2.3). The empirical computation time in terms of patch size is illustrated in Fig. 6(a). To balance the training time and performance of the model, finally we choose 65×65 for all our experiments.

Cross entropy and topological loss vs. training epochs. In this part, we explore the trend of cross entropy loss and topological loss in terms of training epochs. The topological loss, cross entropy loss and total loss are illustrated in Fig. 6(b). From the figure, we can find that after some epochs (in particular, after 30 epochs in Fig. 6(b)), the total loss becomes stable, while the cross entropy loss increases slightly as the topological loss decreases. This is reasonable; incorporating of topological loss may force the network to overtrain on certain locations (near critical pixels), and thus may hurt the overall pixel accuracy slightly. The pixel accuracy performance of our proposed method in Tab. 1 and 2 also confirms this finding.

The effectiveness of the topological loss We plot the thresholded results at different epochs to illustrate the effectiveness of our proposed topological loss (Fig. 7(b) to 7(f)). As the training goes on, the fine structure of the membrane becomes more topologically consistent. This means that our topological term improves the quality of the predictions in terms of topology.

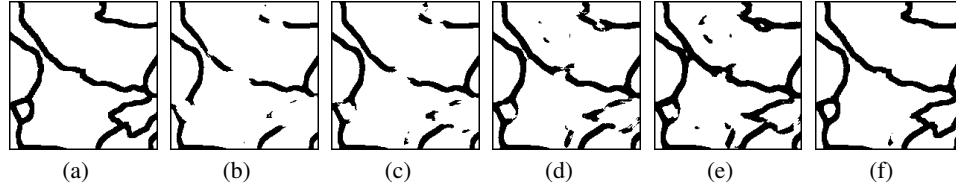


Figure 7: Segmentation at different training epochs. From left to right, ground truth, thresholded results after 10, 20, 30, 40 and 50 epochs.

4 Conclusion

We introduce a new topological loss driven by persistent homology and incorporate it into end-to-end training of deep neural networks. Our method are particularly suitable for fine structure image segmentation. Quantitative and qualitative results show that the proposed topological loss term helps to achieve better performance in terms of topological-relevant metrics. Our proposed topological loss term is generic and can be incorporated into different architectures.

References

- [1] Henry Adams, Tegan Emerson, Michael Kirby, Rachel Neville, Chris Peterson, Patrick Shipman, Sofya Chepushtanova, Eric Hanson, Francis Motta, and Lori Ziegelmeier. Persistence images: A stable vector representation of persistent homology. *The Journal of Machine Learning Research*, 18(1):218–252, 2017.
- [2] Bjoern Andres, Jörg H Kappes, Thorsten Beier, Ullrich Köthe, and Fred A Hamprecht. Probabilistic image segmentation with closedness constraints. In *2011 International Conference on Computer Vision*, pages 2611–2618. IEEE, 2011.
- [3] I Arganda-Carreras, HS Seung, A Vishwanathan, and D Berger. 3d segmentation of neurites in em images challenge-isbi 2013, 2013.
- [4] Ignacio Arganda-Carreras, Srinivas C Turaga, Daniel R Berger, Dan Cireşan, Alessandro Giusti, Luca M Gambardella, Jürgen Schmidhuber, Dmitry Laptev, Sarvesh Dwivedi, Joachim M Buhmann, et al. Crowdsourcing the creation of image segmentation algorithms for connectomics. *Frontiers in neuroanatomy*, 9:142, 2015.
- [5] Mathieu Carriere, Marco Cuturi, and Steve Oudot. Sliced wasserstein kernel for persistence diagrams. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, pages 664–673. JMLR. org, 2017.
- [6] Chao Chen, Daniel Freedman, and Christoph H Lampert. Enforcing topological constraints in random field image segmentation. In *CVPR 2011*, pages 2089–2096. IEEE, 2011.
- [7] Chao Chen, Xiuyan Ni, Qinxun Bai, and Yusu Wang. A topological regularizer for classifiers via persistent homology. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pages 2573–2582, 2019.
- [8] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille. Semantic image segmentation with deep convolutional nets and fully connected crfs. *arXiv preprint arXiv:1412.7062*, 2014.
- [9] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence*, 40(4):834–848, 2018.
- [10] Liang-Chieh Chen, George Papandreou, Florian Schroff, and Hartwig Adam. Rethinking atrous convolution for semantic image segmentation. *arXiv preprint arXiv:1706.05587*, 2017.
- [11] David Cohen-Steiner, Herbert Edelsbrunner, and John Harer. Stability of persistence diagrams. *Discrete & Computational Geometry*, 37(1):103–120, 2007.

- [12] David Cohen-Steiner, Herbert Edelsbrunner, John Harer, and Yuriy Mileyko. Lipschitz functions have l_p -stable persistence. *Foundations of computational mathematics*, 10(2):127–139, 2010.
- [13] Herbert Edelsbrunner and John Harer. *Computational topology: an introduction*. American Mathematical Soc., 2010.
- [14] Herbert Edelsbrunner, David Letscher, and Afra Zomorodian. Topological persistence and simplification. In *Proceedings 41st Annual Symposium on Foundations of Computer Science*, pages 454–463. IEEE, 2000.
- [15] Rolando Estrada, Carlo Tomasi, Scott C Schmidler, and Sina Farsiu. Tree topology estimation. *IEEE transactions on pattern analysis and machine intelligence*, 37(8):1688–1701, 2014.
- [16] Ahmed Fakhry, Hanchuan Peng, and Shuiwang Ji. Deep models for brain em image segmentation: novel insights and improved performance. *Bioinformatics*, 32(15):2352–2358, 2016.
- [17] Jan Funke, Fabian David Tschopp, William Grisaitis, Arlo Sheridan, Chandan Singh, Stephan Saalfeld, and Srinivas C Turaga. A deep structured learning approach towards automating connectome reconstruction from 3d electron micrographs. *arXiv preprint arXiv:1709.02974*, 2017.
- [18] Xiao Han, Chenyang Xu, and Jerry L. Prince. A topology preserving level set method for geometric deformable models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(6):755–768, 2003.
- [19] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 2961–2969, 2017.
- [20] Christoph Hofer, Roland Kwitt, Marc Niethammer, and Andreas Uhl. Deep learning with topological signatures. In *Advances in Neural Information Processing Systems*, pages 1634–1644, 2017.
- [21] Genki Kusano, Yasuaki Hiraoka, and Kenji Fukumizu. Persistence weighted gaussian kernel for topological data analysis. In *International Conference on Machine Learning*, pages 2004–2013, 2016.
- [22] Carole Le Guyader and Luminita A Vese. Self-repelling snakes for topology-preserving segmentation models. *IEEE Transactions on Image Processing*, 17(5):767–779, 2008.
- [23] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015.
- [24] Volodymyr Mnih. *Machine learning for aerial image labeling*. University of Toronto (Canada), 2013.
- [25] Agata Mosinska, Pablo Marquez-Neila, Mateusz Koziński, and Pascal Fua. Beyond the pixel-wise loss for topology-aware delineation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3136–3145, 2018.
- [26] James R Munkres. *Elements of algebraic topology*. CRC Press, 2018.
- [27] Hyeonwoo Noh, Seunghoon Hong, and Bohyung Han. Learning deconvolution network for semantic segmentation. In *Proceedings of the IEEE international conference on computer vision*, pages 1520–1528, 2015.
- [28] Sebastian Nowozin and Christoph H Lampert. Global connectivity potentials for random field models. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 818–825. IEEE, 2009.
- [29] Martin Ralf Oswald, Jan Stühmer, and Daniel Cremers. Generalized connectivity constraints for spatio-temporal 3d reconstruction. In *European Conference on Computer Vision*, pages 32–46. Springer, 2014.

- [30] Adrien Poulenard, Primoz Skraba, and Maks Ovsjanikov. Topological function optimization for continuous shape matching. In *Computer Graphics Forum*, volume 37, pages 13–25. Wiley Online Library, 2018.
- [31] Jan Reininghaus, Stefan Huber, Ulrich Bauer, and Roland Kwitt. A stable multi-scale kernel for topological machine learning. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4741–4748, 2015.
- [32] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.
- [33] Florent Ségonne. Active contours under topology control—genus preserving level sets. *International Journal of Computer Vision*, 79(2):107–117, 2008.
- [34] Joes Staal, Michael D Abràmoff, Meindert Niemeijer, Max A Viergever, and Bram Van Ginneken. Ridge-based vessel segmentation in color images of the retina. *IEEE transactions on medical imaging*, 23(4):501–509, 2004.
- [35] Jan Stuhmer, Peter Schroder, and Daniel Cremers. Tree shape priors with connectivity constraints using convex relaxation on general graphs. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2336–2343, 2013.
- [36] Ganesh Sundaramoorthi and Anthony Yezzi. Global regularizing flows with topology preservation for active contours and polygons. *IEEE Transactions on Image Processing*, 16(3):803–812, 2007.
- [37] Srinivas C Turaga, Kevin L Briggman, Moritz Helmstaedter, Winfried Denk, and H Sebastian Seung. Maximin affinity learning of image segmentation. *arXiv preprint arXiv:0911.5372*, 2009.
- [38] Sara Vicente, Vladimir Kolmogorov, and Carsten Rother. Graph cut based image segmentation with connectivity priors. In *2008 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8. IEEE, 2008.
- [39] Yun Zeng, Dimitris Samaras, Wei Chen, and Qunsheng Peng. Topology cuts: A novel min-cut/max-flow algorithm for topology preserving segmentation in n-d images. *Computer vision and image understanding*, 112(1):81–90, 2008.
- [40] Afra Zomorodian and Gunnar Carlsson. Computing persistent homology. *Discrete & Computational Geometry*, 33(2):249–274, 2005.
- [41] Qin Zou, Yu Cao, Qingquan Li, Qingzhou Mao, and Song Wang. Cracktree: Automatic crack detection from pavement images. *Pattern Recognition Letters*, 33(3):227–238, 2012.