TOPOLOGY-AWARE SEGMENTATION USING DISCRETE MORSE THEORY

Anonymous authors
Paper under double-blind review

ABSTRACT

In the segmentation of fine-scale structures from natural and biomedical images, per-pixel accuracy is not the only metric of concern. Topological correctness, such as vessel connectivity and membrane closure, is crucial for downstream analysis tasks. In this paper, we propose a new approach to train deep image segmentation networks for better topological accuracy. In particular, leveraging the power of discrete Morse theory (DMT), we identify global structures, including 1D skeletons and 2D patches, which are important for topological accuracy. Trained with a novel loss based on these global structures, the network performance is significantly improved especially near topologically challenging locations (such as weak spots of connections and membranes). On diverse datasets, our method achieves superior performance on both the DICE score and topological metrics.

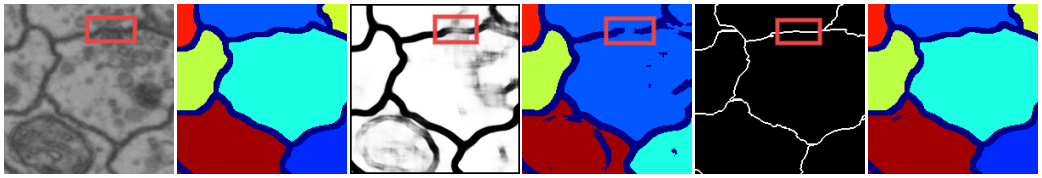
1 INTRODUCTION

Segmenting objects while preserving their global structure is a challenging yet important problem. Various methods have been proposed to encourage neural networks to preserve fine details of objects (Long et al., 2015; He et al., 2017; Chen et al., 2014; 2018; 2017). Despite their high per-pixel accuracy, most of them are still prone to structural errors, such as missing small object instances, breaking thin connections, and leaving holes in membranes. These structural errors can significantly damage downstream analysis. For example, in the segmentation of biomedical structures such as membranes and vessels, small pixel errors at junction will induce significant structure error, leading to catastrophic functional mistakes. See Fig. 1 for an illustration.

Topology is a very global characterization that needs a lot of observations to learn. Any training set is insufficient in teaching the network to correctly reason about topology, especially near challenging spots, e.g., blurred membrane locations or weak vessel connections. A neural network tends to learn from the clean-cut cases and converge quickly. Meanwhile, topologically-challenging locations remain mis-classified, causing structural/topological errors. We note that this issue cannot be alleviated even with more annotated (yet still unbalanced) images.

We propose a novel approach that identifies critical topological structures during training and teaches a neural network to learn from these structures. Our method can produce segmentations with correct topology, i.e., having the same *Betti number* (i.e., number of connected components and handles/tunnels) as the ground truth. Underlying our method is the classic Morse theory (Milnor, 1963), which captures singularities of the gradient vector field of the likelihood function. Intuitively speaking, we treat the likelihood as a terrain function and Morse theory helps us capture terrain structures such as ridges and valleys. See Fig. 1 for illustrations. These structures, constituted of 1D and 2D manifold pieces, reveal the topological information captured by the (potentially noisy) likelihood function.

We consider these Morse structures as *topologically critical*; they encompass all potential skeletons of the object. We propose a new loss that utilizes Morse theory to identify these topologically important structures during training and automatically adjust the loss function based on their importance. By identifying these topologically critical structures, we effectively address the sampling bias issue and ensure that the networks predict correctly near these topologically difficult locations. Since the Morse structures are identified based on the (potentially noisy) likelihood function, they can be both false negatives (a structure can be a true structure but was missed in the segmentation) and false positives (a hallucination of the model and should be removed). Our loss ensures that both kinds of structural mistakes are corrected.



(a) (b) (c) (d) (e) (f)

Figure 1: Illustration of the importance of topological correctness in a neuron image segmentation task and the effectiveness of the proposed DMT-loss. The goal of this task is to segment membranes which partition the image into regions corresponding to neurons. **(a)** an input neuron image with challenging locations (blur regions) highlighted. **(b)** ground truth segmentation of the membranes (dark blue) and the result neuron regions. **(c)** likelihood map of a baseline method without topological guarantee (Ronneberger et al., 2015). **(d)** segmentation results of the baseline method. Small pixel-wise errors lead to broken membranes, resulting in the merging of many neurons into one. **(e)** The topologically critical structure captured by the proposed DMT-loss (based on the likelihood in (c)). **(f)** Our method produces the correct topology and the correct partitioning of neurons.

Several technical challenges need to be tackled. First, the classical Morse theory was defined for smooth functions on smooth domains. Computing the Morse structures can be expensive and numerically unstable. Furthermore, the entire set of Morse structures may include an excessive amount of structures, a large portion of which can be noisy, irrelevant ones. To address these challenges, we use the discrete version of Morse theory by Forman (1998; 2002). For efficiency purposes, we also use an approximation algorithm to compute 2D Morse structures with almost linear time. The idea is to compute zero dimensional Morse structures of the dual image, which boils down to a minimum spanning tree computation. Finally, we use the theory of persistent homology (Edelsbrunner et al., 2000; Edelsbrunner & Harer, 2010) to prune spurious Morse structures that are not topologically critical.

Our discrete-Morse-theory based loss, called the *DMT-loss*, can be evaluated efficiently and can effectively train the neural network to achieve high performance in both topological accuracy and per-pixel accuracy. Our method outperforms state-of-the-art methods in multiple topology-relevant metrics (e.g., ARI and VOI), on various 2D and 3D benchmarks. It has superior performance in the Betti number error, which is an exact measurement of the topological fidelity of the segmentation.

Related work. Closely related to our method are recent works on persistent-homology-based losses (Hu et al., 2019; Clough et al., 2019). These methods identify a set of critical points of the likelihood function, e.g., saddles and extrema, as topologically critical locations for the neural network to memorize. However, only identifying a *sparse set* of critical points at every epoch is inefficient in terms of training. Instead, our method identifies a much bigger set of critical locations at each epoch, i.e., 1D or 2D Morse skeletons (curves and patches). This brings benefits in both training efficiency and model performance. Extending the critical location sets from points to 1D curves and 2D patches makes it much more efficient in training. Compared with TopoLoss, we observe a 3-time speedup in practice. Furthermore, by focusing on more critical locations early, our method is more likely to escape poor local minima of the loss landscape. Thus it achieves better topological accuracy than TopoLoss. The shorter training time may also contribute to better stability of the SGD algorithm, and thus better test accuracy (Hardt et al., 2016).

Another topology-aware loss (Mosinska et al., 2018) uses pretrained filters to detect broken connections. However, this method cannot be generalized to unobserved geometry and higher dimensional topology (loops and voids). We also refer to other existing work on topological features and their applications (Adams et al., 2017; Reininghaus et al., 2015; Kusano et al., 2016; Carriere et al., 2017; Ni et al., 2017; Chen & Quadrianto, 2016; Wu et al., 2017). Deep neural networks have also been proposed to learn from topological features directly extracted from data (Hofer et al., 2017; Carrière et al., 2019). Persistent-homology-inspired objective functions have been proposed for graphics (Poulenard et al., 2018) machine learning (Chen et al., 2019; Hofer et al., 2019). Discrete Morse theory has been used for image analysis (Delgado-Friedrichs et al., 2015; Robins et al., 2011; Wang et al., 2015; Dey et al., 2019), but only as a preprocessing step. To the best of our knowledge, **our method is the first to use the Morse structures for the end-to-end training of a network.**

Many methods have leveraged the power of deep neural networks (Ronneberger et al., 2015; Long et al., 2015; Badrinarayanan et al., 2017; Ding et al., 2019; Kervadec et al., 2019; Karimi & Salcudean, 2019; Mosinska et al., 2018) for fine-scale structure segmentation. One may also enforce connectivity

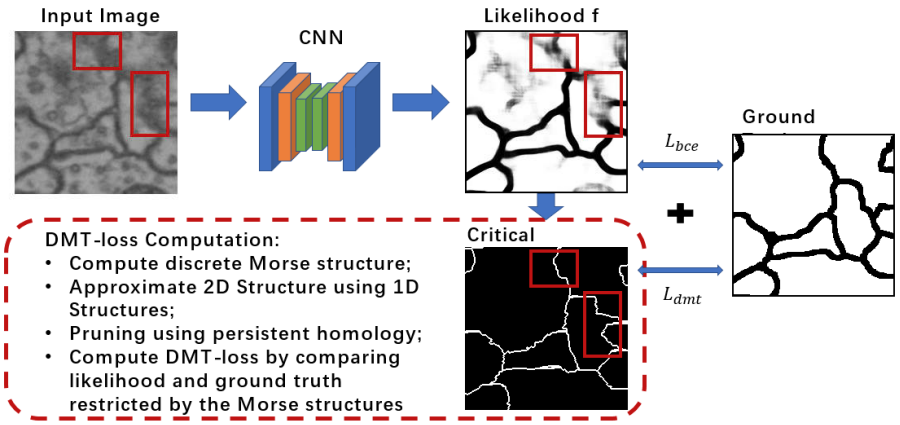


Figure 2: Overview of our method. Topologically critical and error-prune structures are highlighted.

constraints when postprocessing the likelihood map (Han et al., 2003; Le Guyader & Vese, 2008; Sundaramoorthi & Yezzi, 2007; Ségonne, 2008; Wu et al., 2017; Gao et al., 2013; Vicente et al., 2008; Nowozin & Lampert, 2009; Zeng et al., 2008; Chen et al., 2011; Andres et al., 2011; Stuhmer et al., 2013; Oswald et al., 2014; Estrada et al., 2014). However, when the deep neural network itself is topology-agnostic, the likelihood map may be fundamentally flawed and cannot be salvaged topologically. Specific to neuron image segmentation, some methods (Funke et al., 2017; Turaga et al., 2009; Januszewski et al., 2018) directly find neuron regions instead of their boundary/membranes. These methods cannot be generalized to other types of data such as satellite images, retinal images, vessel images, etc.

2 METHOD

We propose a novel loss to train a topology-aware network end-to-end. It uses global structures captured by discrete Morse theory (DMT) to discover critical topological structures. In particular, through the language of 1- and 2-stable manifolds, DMT helps identify 1D skeletons or 2D sheets (separating 3D regions) that may be critical for structural accuracy. These Morse structures are used to define a DMT-loss that is essentially the cross-entropy loss constrained to these topologically critical structures. As the training continues, the neural network learns to better predict around these critical structures, and eventually achieves a better topological accuracy. Please refer to Fig. 2 for an overview of our method. In what follows, we provide relevant concepts in (discrete) Morse theory (Sec. 2.1), describe how to compute and prune the Morse structures efficiently (Sec. 2.2), and introduce our proposed DMT-loss function (Sec. 2.3). Fig. 2 shows an overview of the proposed method.

2.1 MORSE THEORY

Morse theory (Milnor, 1963) identifies topologically critical structures from a likelihood map (Fig. 3(a)). In particular, it views the likelihood as a terrain function (Fig. 3(b)) and extracts its landscape features such as mountain ridges and their high-dimensional counterparts. The broken connection in the likelihood map corresponds to a local dip in the mountain ridge of the terrain in Fig. 3(b) and Fig. 3(c). The bottom of this dip is captured by a so-called saddle point (S in Fig. 3(c)) of the likelihood map. The mountain ridge connected to this bottom point captures the main part of missing pixels. Such "mountain ridges" can be captured by the so-called stable manifold w.r.t. the saddle point using the language of Morse theory. By finding the saddle points and the stable manifold of the saddle points on the likelihood map, we can ensure the model learns to "correctly" handle pixels near these structures. We note that an analogous scenario can also happen with such 1D signals (such as blood vessels) as well as 2D signals (such as membranes of cells) in 3D images – they can also be captured by saddles (of different indices) and their stable manifolds.

In this paper, we focus on the application of segmenting 2D and 3D images. Specifically, suppose we have a smooth function $f : \mathbb{R}^d \to \mathbb{R}$ to model the likelihood (density) map. Given any point $x \in \mathbb{R}^d$, the negative gradient $-\nabla f(x) = -[\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \ldots, \frac{\partial f}{\partial x_d}]^T$ indicates the steepest descending direction of f. A point $x = (x_1, x_2, \ldots, x_k)$ is *critical* if the gradient of function at this point vanishes (i.e., $\nabla f(x) = 0$). For a well-behaved function (more formally, called Morse function) defined on $\mathbb{R}^d$, a critical point could be a minimum, a maximum, or $d-1$ types of saddle points. See Fig. 3(c) as an example. For $d = 2$, there is only one type of saddle point. For $d = 3$, there are two types of saddle

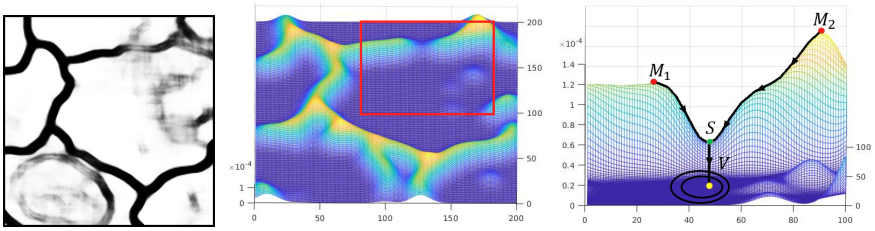


Figure 3: From left to right: **(a)** Likelihood map. **(b)** Density map: the z-axis value is the probability of the likelihood map in left figure. **(c)** Density map for the highlighted region in middle figure. M_1 and M_2 are maxima (red dots), V is a minimum (yellow), S is a saddle (green) with its stable manifolds flowing to it from M_1 and M_2.

points, referred to as index-1 and index-2 saddles. Formally, when taking the eigen-values of the Hessian matrix at a critical point, its index is equal to the number of negative eigenvalues.

Intuitively, imagine we put a drop of water on the graph of f (i.e, the terrain in Fig. 3(b)) at the lift of x onto this terrain, then $-\nabla f(x)$ indicates the direction along which the water will flow down. If we track the trajectory of this water drop as it flows down, this gives rise to so-called *integral line* (a flow line). Such flow lines can only start and end at critical points[1], where the gradient vanishes.

The stable manifold $\mathsf{S}(p)$ of a critical point p is defined as the collection of points whose flow line ends at p. For a 2D function $f : \mathbb{R}^2 \to \mathbb{R}$, for a saddle q, its stable manifold $\mathsf{S}(q)$ starts from local maxima (mountain peaks in the terrain) and ends at q, tracing out the mountain ridges separating different valleys (Fig. 3(c)). The stable manifold $\mathsf{S}(p)$ of a minimum p, on the other hand, corresponds to the entire valley around this minimum p. See the valley point V and its corresponding stable field in Fig. 3(c). For a 3D function $f : \mathbb{R}^3 \to \mathbb{R}$, the stable manifold w.r.t. an index-2 saddle connects mountain peaks to saddles, tracing 1D mountain ridges as in the case of a 2D function. The stable manifold w.r.t. an index-1 saddle q consists of flow lines starting at index-2 saddles and ending at q. Their union, called the 2-stable manifold of f, consists of a collection of 2-manifold pieces.

These stable manifolds indicate important topological structures (graph-like or sheet-like) based on the likelihood of the current neural network. Using these structures, we will propose a novel loss (Sec. 2.3) to improve the topological awareness of the model. In practice, for images, we will leverage the discrete version of Morse theory for both numerical stability and easier simplification.

Discrete Morse theory. Due to space limitations, we briefly explain discrete Morse theory, leaving technical details to the Appendix A.2.1. We view a d-dimensional image, $d = 2$ or 3, as a d-dimensional cubical complex, meaning it consists of 0-, 1-, 2- and 3-dimensional cells corresponding to vertices (pixels/voxels), edges, squares, and cubes as its building blocks. Discrete Morse theory (DMT), originally introduced in (Forman, 1998; 2002), is a combinatorial version of Morse theory for general cell complexes. In this setting, the analog of a gradient vector is a pair of adjacent cells, called discrete gradient vectors. The analog of an integral line (flow line) is a sequence of such cell-pairs (discrete gradient vectors), forming a so-called *V-path*. Critical points correspond to critical cells which do not participate in any discrete gradient vectors. A minimum, index-1 saddle, index-2 saddle and a maximum for 3D domain intuitively correspond to a critical vertex, critical edge, critical square and critical cube respectively. A 1-stable manifold in 2D will correspond to a V-path, i.e., a sequence of cells, connecting a critical square (a maximum) and a critical edge (a saddle). See Fig. 3(c) for an illustration. In 3D, it will be a V-path connecting a critical cube and a critical square.

2.2 Simplification and Computation

In this section, we describe how we extract discrete Morse structures corresponding to the 1-stable manifold and 2-stable manifold in the continuous analog. First, we prune unnecessary Morse structures, based on the theory of persistent homology (Edelsbrunner et al., 2000; Edelsbrunner & Harer, 2010). Second, the 2-stable manifold is rather involved to compute based on the original definition. We approximate the 2-stable manifold structures using 0-stable manifolds of the dual to achieve high efficiency in practice.

Persistence-based structure pruning. While Morse structures reveal important structural information, they can be sensitive to noise. Without proper pruning, there can be an excessive amount of

[1]More precisely, flow lines only tend to critical points in the limit and never reach it.

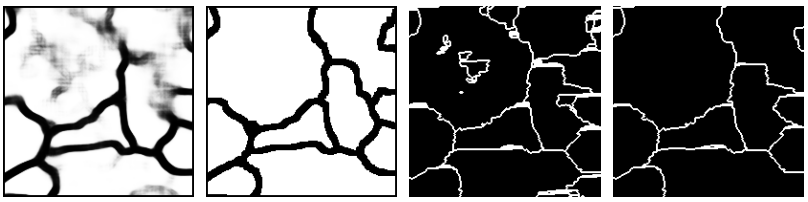


Figure 4: From left to right: **(a)** Sample likelihood map, **(b)** Ground truth, **(c)** improperly pruned structures and **(d)** properly pruned structures.

Morse structures, many of which are spurious and not relevant to the true signal. See Fig. 4(c) for an example. We propose to prune these structures using persistent homology.

Persistent homology is one of the most important developments in the field of topological data analysis in the past two decades (Edelsbrunner & Harer, 2010; Edelsbrunner et al., 2000). Intuitively speaking, we grow the complex by starting from the empty set and gradually include more and more cells using a decreasing threshold. Through this course, new topological features can be created upon adding a critical cell, and sometimes a feature can be destroyed upon adding another critical cell. The persistence algorithm (Edelsbrunner et al., 2000) will pair up these critical cells; that is, its output is a set of pairs of critical cells, where each pair captures the birth and death of topological features during this evolution. The persistence of a pair is defined as the difference of function values of the two critical cells, intuitively measuring how long the topological feature lives in term of f.

Using persistence, we can prune critical cells that are less topologically salient, and thus their corresponding Morse structures. Recall each 1- and 2-stable Morse structure is constituted by V-paths flowing into a critical cell (corresponding to a saddle in continuous setting). We then use the persistence associated with this critical cell to determine the saliency of the corresponding Morse structure. If the persistence is below a certain threshold ϵ, we prune the corresponding Morse structure via an operation called *Morse cancellation* (more details are in the Appendix A.2.2).[2] See Fig. 4(d) for example Morse structures after pruning. We denote by $\mathcal{S}_1(\epsilon)$ and $\mathcal{S}_2(\epsilon)$ the remaining sets of 1- and 2-stable manifolds after pruning. We'll use these Morse structures to define the loss (Sec. 2.3).

Computation. We need an efficient algorithm to compute $\mathcal{S}_1(\epsilon)$ and $\mathcal{S}_2(\epsilon)$ from a given likelihood f, because this computation needs to be carried out at each epoch. It is especially more involved to define and compute $\mathcal{S}_2(\epsilon)$ in the discrete Morse setting (Delgado-Friedrichs et al., 2015). Furthermore, this also requires the computation of persistent homology up to 2-dimensions, which takes time $T = O(n^\omega)$ (where $\omega \approx 2.37$ is the exponent in the matrix multiplication time, i.e., time to multiply two $n \times n$ matrices). To this end, we propose to approximate $\mathcal{S}_2(\epsilon)$ by $\widehat{\mathcal{S}}_2(\epsilon)$ (more details could be found in Appendix A.2.3) which intuitively comes from the "boundary" of the stable manifold for minima. Note that in the smooth case for a function $f : \mathbb{R}^3 \to \mathbb{R}$, the closure of 2-stable manifolds exactly corresponds to the 2D sheets on the boundary of stable manifolds of minima. [3] This is both conceptually clear and also avoids the costly persistence computation. In particular, given a minimum q with persistence greater than the pruning threshold ϵ, the collections of V-paths ending at the minimum q form a spanning tree T_q. In fact, consider all minima $\{q_1, \ldots, q_\ell\}$ with persistence at least ϵ. $\{T_{q_i}\}$ form a maximum spanning forest among all edges with persistence value smaller than ϵ (Bauer et al., 2012; Dey et al., 2018). Hence it can be computed easily in $O(n \log n)$ time, in which n is the image size.

We then take all the edges incident to nodes from different trees. The dual of these edges, denoted as $\widehat{\mathcal{S}}_2(\epsilon)$, serve as the "boundaries" separating different spanning trees (representing stable manifolds to different minima with persistence $\geq \epsilon$). See Fig. 5. Overall, the computation of $\widehat{\mathcal{S}}_2(\epsilon)$ takes only $O(n \log n)$ by a maximum spanning tree algorithm.

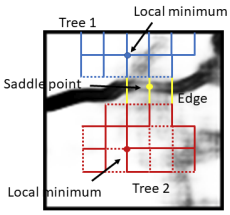


Figure 5: Spanning tree illustration.

As for $\mathcal{S}_1(\epsilon)$, we use a simplified algorithm of (Dey et al., 2018), which can compute $\mathcal{S}_1(\epsilon)$ in $O(n \log n)$ time for a 2D image, in which n is the image size. For a 3D image, the time is $O(n \log n + T)$, where $T = O(n^\omega)$ is the time to compute persistent homology, where $\omega \approx 2.37$ is the exponent in matrix multiplication time.

[2] Technically, not all spurious structures can be pruned/cancelled. But in practice, most of them can.

[3] This however is not always true for the discrete Morse setting.

2.3 THE DMT-BASED LOSS FUNCTION AND TRAINING DETAILS

Our loss has two terms, the cross-entropy term, L_{bce} and the DMT-loss, L_{dmt}: $L(f, g) = L_{bce}(f, g) + \beta L_{dmt}(f, g)$, in which f is the likelihood, g is the ground truth, and β is the weight of L_{dmt}. Here we focus on one single image, while the actual loss is aggregated over the whole training set.

The DMT-loss enforce the correctness of the topologically challenging locations discovered by our algorithm. These locations are pixels of the (approximation of) 1- and 2-stable manifolds $\mathcal{S}_1(\epsilon)$ and $\widehat{\mathcal{S}}_2(\epsilon)$ of the likelihood, f (Fig. 1(e)). Denote by $\mathcal{M}_f$ a binary mask of the union of pixels of all Morse structures in $\mathcal{S}_1(\epsilon) \cup \widehat{\mathcal{S}}_2(\epsilon)$. We want to enforce these locations to be correctly segmented. We use the cross-entropy between the likelihood map f and ground truth g restricted to the Morse structures, formally, $L_{dmt}(f, g) = L_{bce}(f \circ \mathcal{M}_f, g \circ \mathcal{M}_f)$, in which $\circ$ is the Hadamard product.

Different types of topological errors. Recall the Morse structures are computed over the potentially noisy likelihood function of a neural network, which can help identify two types of structural errors: (1) **false negative**: a true structure that is incomplete in the segmentation, but can be visible in Morse structures. This types of false negatives (broken connection, holes in membrane) can be restored as the additional cross-entropy loss near the Morse structures will force the network to increase its likelihood value on these structures. (2) **false positive**: a phantom structure hallucinated by the network when they do not exist (spurious branches, membrane pieces). These errors can be eliminated as the extra cross entropy loss on these structures will force the network to decrease the likelihood values along these structures. We will provide a more thorough discussion in the Appendix A.1.2.

Differentiability. We note that the Morse structures are recomputed at every epoch. The structures, as well as their mask $\mathcal{M}_f$, may change with f. However, the change is not continuous; the output of the discrete Morse algorithm is a combinatorial solution that does not change continuously with f. Instead, it only changes at singularities, i.e., when function values of f at different pixels/voxels are the same. In other words, for a general f, the likelihood function is real-valued, so it is unlikely two pixels share the exact same value. In case if they are, the persistence homology algorithm by default will break the tie and choose one as critical. The mask $\mathcal{M}_f$ remains a constant within a small neighborhood of current f. Therefore, the gradient of L_{dmt} exists and can be computed naturally.

Training details. Although our method is architecture-agnostic, for 2D datasets, we select an architecture driven by 2D U-net (Ronneberger et al., 2015); for 3D datasets, we select one architecture inspired by 3D U-Net (Çiçek et al., 2016). Both U-Net and 3D U-Net were originally designed for neuron image segmentation tasks, capturing the fine-structures of images. In practice, we first pretrain the network with only the cross-entropy loss, and then train the network with the combined loss.

3 EXPERIMENTS ON 2D AND 3D DATASETS

Experiments on 2D dataset. Six natural and biomedical 2D datasets are used: **ISBI12** (Arganda-Carreras et al., 2015), **ISBI13** (Arganda-Carreras et al., 2013), **CREMI**, **CrackTree** (Zou et al., 2012), **Road** (Mnih, 2013) and **DRIVE** (Staal et al., 2004). More details about the datasets are included in Appendix A.3.1. For all the experiments, we use a 3-fold cross-validation to tune hyperparameters for both the proposed method and other baselines, and report the mean performance over the validation set. This also holds for 3D experiments.

Evaluation metrics. We use five different evaluation metrics: **Pixel-wise accuracy**, **DICE score**, **ARI**, **VOI** and the most important one is **Betti number error**, which directly compares the topology (number of handles/voids) between the segmentation and the ground truth. More details about the evaluation metrics are provided in Appendix A.3.2. The last three metrics are topology-aware.

Baselines. DIVE (Fakhry et al., 2016) is a popular neural network that predicts the probability of a pixel being a membrane (border) pixel or not. **U-Net** (Ronneberger et al., 2015) is a one of the most powerful image segmentation method trained with cross-entropy loss. **Mosin.** (Mosinska et al., 2018) constructs a topology-aware loss based on the response of selected filters from a pretrained CNN. **TopoLoss** (Hu et al., 2019) proposes a novel topological loss to learn to segment with correct topology. For all methods, we generate segmentations by thresholding the predicted likelihood maps at 0.5, and this also holds for 3D experiments.

Quantitative and qualitative results. Table 1 shows quantitative results for four 2D image datasets, ISBI13, CREMI, CrackTree and Road. The results are highlighted when they are significantly better,

Table 1: Quantitative results for different models on several 2D medical datasets

Method	Accuracy	DICE	ARI	VOI	Betti Error
		ISBI13			
DIVE	0.9642 ± 0.0018	0.9658 ± 0.0020	0.6923 ± 0.0134	2.790 ± 0.025	3.875 ± 0.326
U-Net	0.9631 ± 0.0024	0.9649 ± 0.0057	0.7031 ± 0.0256	2.583 ± 0.078	3.463 ± 0.435
Mosin.	0.9578 ± 0.0029	0.9623 ± 0.0047	0.7483 ± 0.0367	1.534 ± 0.063	2.952 ± 0.379
TopoLoss	0.9569 ± 0.0031	$\mathbf{0.9689 \pm 0.0026}$	0.8064 ± 0.0112	1.436 ± 0.008	$\mathbf{1.253 \pm 0.172}$
DMT	0.9625 ± 0.0027	$\mathbf{0.9712 \pm 0.0047}$	$\mathbf{0.8289 \pm 0.0189}$	$\mathbf{1.176 \pm 0.052}$	$\mathbf{1.102 \pm 0.203}$
		CREMI			
DIVE	0.9498 ± 0.0029	0.9542 ± 0.0037	0.6532 ± 0.0247	2.513 ± 0.047	4.378 ± 0.152
U-Net	0.9468 ± 0.0048	0.9523 ± 0.0049	0.6723 ± 0.0312	2.346 ± 0.105	3.016 ± 0.253
Mosin.	0.9467 ± 0.0058	0.9489 ± 0.0053	0.7853 ± 0.0281	1.623 ± 0.083	1.973 ± 0.310
TopoLoss	0.9456 ± 0.0053	0.9596 ± 0.0029	0.8083 ± 0.0104	1.462 ± 0.028	$\mathbf{1.113 \pm 0.224}$
DMT	0.9475 ± 0.0031	$\mathbf{0.9653 \pm 0.0019}$	$\mathbf{0.8203 \pm 0.0147}$	$\mathbf{1.089 \pm 0.061}$	$\mathbf{0.982 \pm 0.179}$
		CrackTree			
DIVE	0.9854 ± 0.0052	0.6530 ± 0.0017	0.8634 ± 0.0376	1.570 ± 0.078	1.576 ± 0.287
U-Net	0.9821 ± 0.0097	0.6491 ± 0.0029	0.8749 ± 0.0421	1.625 ± 0.104	1.785 ± 0.303
Mosin.	0.9833 ± 0.0067	0.6527 ± 0.0010	0.8897 ± 0.0201	1.113 ± 0.057	1.045 ± 0.214
TopoLoss	0.9826 ± 0.0084	0.6732 ± 0.0041	$\mathbf{0.9291 \pm 0.0123}$	0.997 ± 0.011	$\mathbf{0.672 \pm 0.176}$
DMT	0.9842 ± 0.0041	$\mathbf{0.6811 \pm 0.0047}$	$\mathbf{0.9307 \pm 0.0172}$	$\mathbf{0.901 \pm 0.081}$	$\mathbf{0.518 \pm 0.189}$
		Road			
DIVE	0.9734 ± 0.0077	0.6743 ± 0.0051	0.8201 ± 0.0128	2.368 ± 0.203	3.598 ± 0.783
U-Net	0.9786 ± 0.0052	0.6612 ± 0.0016	0.8189 ± 0.0097	2.249 ± 0.175	3.439 ± 0.621
Mosin.	0.9754 ± 0.0043	0.6673 ± 0.0044	0.8456 ± 0.0174	1.457 ± 0.096	2.781 ± 0.237
TopoLoss	0.9728 ± 0.0063	0.6903 ± 0.0038	0.8671 ± 0.0068	1.234 ± 0.037	$\mathbf{1.275 \pm 0.192}$
DMT	0.9744 ± 0.0049	$\mathbf{0.7056 \pm 0.0022}$	$\mathbf{0.8819 \pm 0.0104}$	$\mathbf{1.092 \pm 0.129}$	$\mathbf{0.995 \pm 0.301}$

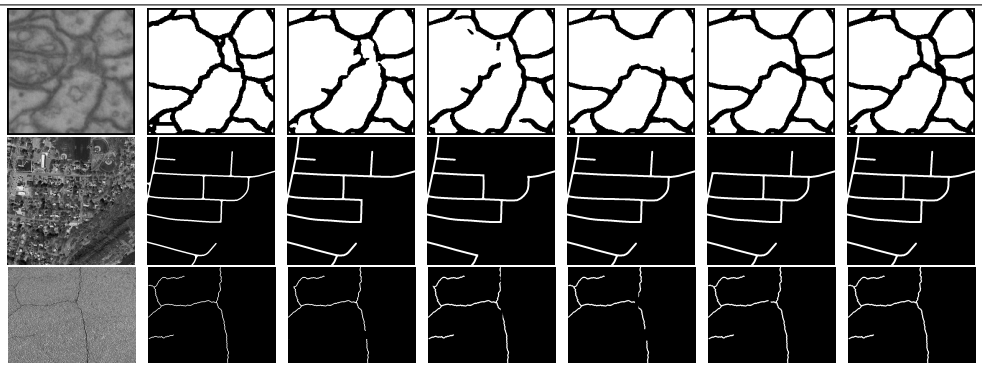


Figure 6: Qualitative results of the proposed method compared to other models. From left to right: sample images, ground truth, results for **DIVE**, **U-Net**, **Mosin.**, **TopoLoss** and our proposed **DMT**.

and the statistical significance is determined by t-test. Results for ISBI12 and DRIVE are included in Appendix A.3.3. DMT-loss outperforms others in both DICE score and topological accuracy (ARI, VOI and Betti Error). Please note that here the backbone of TopoLoss is the same as in (Hu et al., 2019), a heavily designed network. The performance of TopoLoss will be worse if we use the U-Net backbone as DMT-Loss. Comparisons with the same backbones can be found in Appendix A.3.4.

Fig. 6 shows qualitative results. Our method correctly segments fine structures such as membranes, roads and vessels. Our loss is a weighted combination of cross entropy loss and DMT-loss. When $\beta = 0$, the proposed method degrades to standard U-Net. The improvement of the performance over all datasets (Line U-Net and DMT in Table 1) demonstrates that our DMT-loss is helping the deep neural nets to learn a better structural segmentation.

Experiments on 3D datasets. We use three different biomedical 3D datasets: **ISBI13**, **CREMI** and **3Dircadb** (Soler et al., 2010). **ISBI13** and **CREMI**, which have been mentioned above, are originally 3D datasets, can be used in both 2D and 3D settings. We also use the open dataset **3Dircadb** to evaluate our proposed method, which contains 20 enhanced CT volumes with artery annotations.

Evaluation metrics. We use similar evaluation metrics as the 2D part. To be noted that, in terms of 2D cases, for ARI and VOI, we compute the numbers for each slice and then average the numbers for

Table 2: Quantitative results for different models on several 3D datasets

Method	Accuracy	DICE	ARI	VOI	Betti Error
ISBI13					
3D DIVE	0.9723 ± 0.0021	0.9681 ± 0.0043	0.8719 ± 0.0189	1.208 ± 0.149	2.375 ± 0.419
3D U-Net	0.9746 ± 0.0025	0.9701 ± 0.0012	$\mathbf{0.8956 \pm 0.0391}$	1.123 ± 0.091	1.954 ± 0.585
MALA	0.9701 ± 0.0018	0.9699 ± 0.0013	$\mathbf{0.8945 \pm 0.0481}$	0.901 ± 0.106	1.103 ± 0.207
3D TopoLoss	0.9689 ± 0.0031	0.9752 ± 0.0045	$\mathbf{0.9043 \pm 0.0283}$	0.792 ± 0.086	0.972 ± 0.245
DMT	0.9701 ± 0.0026	$\mathbf{0.9803 \pm 0.0019}$	$\mathbf{0.9149 \pm 0.0217}$	$\mathbf{0.634 \pm 0.086}$	$\mathbf{0.812 \pm 0.134}$
CREMI					
3D DIVE	0.9503 ± 0.0061	0.9641 ± 0.0011	0.8514 ± 0.0387	1.219 ± 0.103	2.674 ± 0.473
3D U-Net	0.9547 ± 0.0038	0.9618 ± 0.0026	0.8322 ± 0.0315	1.416 ± 0.097	2.313 ± 0.501
MALA	0.9472 ± 0.0027	0.9583 ± 0.0023	0.8713 ± 0.0286	1.109 ± 0.093	1.114 ± 0.309
3D TopoLoss	0.9523 ± 0.0043	0.9672 ± 0.0010	0.8726 ± 0.0194	1.044 ± 0.128	1.076 ± 0.206
DMT	0.9529 ± 0.0031	$\mathbf{0.9731 \pm 0.0045}$	$\mathbf{0.9013 \pm 0.0202}$	$\mathbf{0.891 \pm 0.099}$	$\mathbf{0.726 \pm 0.187}$
3Dircadb					
3D DIVE	0.9618 ± 0.0054	0.6097 ± 0.0034	/	/	4.571 ± 0.505
3D U-Net	0.9632 ± 0.0009	0.5898 ± 0.0025	/	/	4.131 ± 0.483
MALA	0.9546 ± 0.0033	0.5719 ± 0.0043	/	/	2.982 ± 0.105
3D TopoLoss	0.9561 ± 0.0019	0.6138 ± 0.0029	/	/	2.245 ± 0.255
DMT	0.9587 ± 0.0023	$\mathbf{0.6257 \pm 0.0021}$	/	/	$\mathbf{1.415 \pm 0.305}$

different slices as the final performance; for 3D images, we compute the performance for the whole volume. For 2D images, we compute 1D Betti number (number of holes) to obtain the Betti Error; while for 3D images, we compute 2D Betti number (number of voids) to obtain the Betti Error.

Baselines. **3D DIVE** (Zeng et al., 2017), **3D U-Net** (Çiçek et al., 2016), **3D TopoLoss** (Hu et al., 2019) are the 3D versions for **DIVE**, **U-Net** and **TopoLoss**. **MALA** (Funke et al., 2017) trains the U-Net using a new structured loss function.

Quantitative and qualitative results. Table 2 shows the quantitative results for three different 3D image datasets, ISBI13, CREMI and 3Dircadb. Our method outperforms existing methods in topological accuracy (in all three topology-aware metrics), which demonstrates the effectiveness of the proposed method. More qualitative results for 3D cases are included in Appendix A.3.5.

The benefit of the proposed DMT-loss. Instead of capturing isolated critical points in TopoLoss (Hu et al., 2019), the proposed DMT-loss captures the whole V-path as critical structures. Taking the patch in Fig. 4(a) as an example, TopoLoss identifies ≈ 80 isolated critical pixels for further training, whereas the critical structures captured by DMT-loss contain ≈ 1000 critical pixels (Fig. 4(d)). We compare efficiency of DMT-loss and TopoLoss using a same backbone network, evaluated on CREMI 2D dataset. Both methods start from a pre-trained reasonable likelihood map. TopoLoss achieves 1.113 (Betti Error), taking $\approx$3h to converge; while DMT-loss achieves 0.956 (Betti Error), taking $\approx$1.2h to converge (the standard U-Net takes $\approx$0.5h instead). Aside from converging faster, DMT-loss is also less likely to converge to low-quality local minima. We hypothesize that the loss landscape of the topological loss will have more local minima than that of the DMT-loss, even though the global minima of both landscapes may have the same quality.

Ablation study for persistence threshold ϵ. As illustrated in Fig. 4, different persistence thresholds ϵ will identify different critical structures. The ablation experiment is also conducted on CREMI 2D dataset. When $\epsilon = 0.2$ (See Fig. 4(d)), the proposed DMT-loss achieves best performance 0.982 (Betti Error). When $\epsilon = 0.1$ and $\epsilon = 0.3$, the performance drops to 1.206 and 2.105 (both in Betti Error), respectively. This makes sense for following reasons: 1) for $\epsilon = 0.1$, the DMT-loss captures lots of unnecessary structures which mislead the neural networks; 2) for $\epsilon = 0.3$, the DMT-loss misses lots of important critical structures, making the performance drop significantly. The ϵ is chosen via cross-validation. The ablation study for balanced term β is provided in Appendix A.3.6.

4 Conclusion

In this paper, a new DMT-loss is proposed to train deep image segmentation neural networks for better topological accuracy. With the power of discrete Morse theory (DMT), we could identify 1D skeletons and 2D patches which are important for topological accuracy. Trained with the new loss based on these global structures, the networks are significantly improved especially near topologically challenging locations (such as weak spots of connections and membranes).

REFERENCES

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.

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*, pp. 2611–2618. IEEE, 2011.

I Arganda-Carreras, HS Seung, A Vishwanathan, and D Berger. 3d segmentation of neurites in em images challenge-isbi 2013, 2013.

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.

Vijay Badrinarayanan, Alex Kendall, and Roberto Cipolla. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE transactions on pattern analysis and machine intelligence*, 39(12):2481–2495, 2017.

U. Bauer, C. Lange, and M. Wardetzky. Optimal topological simplification of discrete functions on surfaces. *Discr. Comput. Geom.*, 47(2):347–377, 2012. ISSN 1432-0444.

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*, pp. 664–673. JMLR. org, 2017.

Mathieu Carrière, Frédéric Chazal, Yuichi Ike, Théo Lacombe, Martin Royer, and Yuhei Umeda. A general neural network architecture for persistence diagrams and graph classification. *arXiv preprint arXiv:1904.09378*, 2019.

Chao Chen and Novi Quadrianto. Clustering high dimensional categorical data via topographical features. In *Proceedings of the 33rd International Conference on Machine Learning; New York; 19-24 June 2016*, volume 48, pp. 2732–2740. JMLR, 2016.

Chao Chen, Daniel Freedman, and Christoph H Lampert. Enforcing topological constraints in random field image segmentation. In *CVPR 2011*, pp. 2089–2096. IEEE, 2011.

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*, pp. 2573–2582, 2019.

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.

Liang-Chieh Chen, George Papandreou, Florian Schroff, and Hartwig Adam. Rethinking atrous convolution for semantic image segmentation. *arXiv preprint arXiv:1706.05587*, 2017.

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.

Özgün Çiçek, Ahmed Abdulkadir, Soeren S Lienkamp, Thomas Brox, and Olaf Ronneberger. 3d u-net: learning dense volumetric segmentation from sparse annotation. In *International conference on medical image computing and computer-assisted intervention*, pp. 424–432. Springer, 2016.

James R. Clough, Ilkay Oksuz, Nicholas Byrne, Veronika A. Zimmer, Julia A. Schnabel, and Andrew P. King. A topological loss function for deep-learning based image segmentation using persistent homology, 2019. arXiv preprint at arXiv:1910.01877.

O. Delgado-Friedrichs, V. Robins, and A. Sheppard. Skeletonization and partitioning of digital images using discrete morse theory. *IEEE Trans. Pattern Anal. Machine Intelligence*, 37(3):654–666, March 2015. ISSN 0162-8828.

T. Dey, J. Wang, and Y. Wang. Road network reconstruction from satellite images with machine learning supported by topological methods. In *Proc. 27th ACM SIGSPATIAL Intl. Conf. Adv. Geographic Information Systems (GIS)*, pp. 520–523, 2019.

Tamal K. Dey, Jiayuan Wang, and Yusu Wang. Graph reconstruction by discrete morse theory. In *Proc. 34th Intl. Sympos. Comput. Geom. (SoCG)*, pp. 31:1–31:15, 2018.

Henghui Ding, Xudong Jiang, Ai Qun Liu, Nadia Magnenat Thalmann, and Gang Wang. Boundary-aware feature propagation for scene segmentation. In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 6819–6829, 2019.

Herbert Edelsbrunner and John Harer. *Computational topology: an introduction*. American Mathematical Soc., 2010.

Herbert Edelsbrunner, David Letscher, and Afra Zomorodian. Topological persistence and simplification. In *Proceedings 41st Annual Symposium on Foundations of Computer Science*, pp. 454–463. IEEE, 2000.

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.

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.

R. Forman. Morse theory for cell complexes. *Advances in mathematics*, 134(1):90–145, 1998.

Robin Forman. A user's guide to discrete morse theory. *Sém. Lothar. Combin*, 48:35pp, 2002.

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.

Mingchen Gao, Chao Chen, Shaoting Zhang, Zhen Qian, Dimitris Metaxas, and Leon Axel. Segmenting the papillary muscles and the trabeculae from high resolution cardiac ct through restoration of topological handles. In *International Conference on Information Processing in Medical Imaging*, pp. 184–195. Springer, 2013.

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.

Moritz Hardt, Ben Recht, and Yoram Singer. Train faster, generalize better: Stability of stochastic gradient descent. In *International Conference on Machine Learning*, pp. 1225–1234, 2016.

Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pp. 2961–2969, 2017.

Christoph Hofer, Roland Kwitt, Marc Niethammer, and Andreas Uhl. Deep learning with topological signatures. In *Advances in Neural Information Processing Systems*, pp. 1634–1644, 2017.

Christoph Hofer, Roland Kwitt, Mandar Dixit, and Marc Niethammer. Connectivity-optimized representation learning via persistent homology. *arXiv preprint arXiv:1906.09003*, 2019.

Xiaoling Hu, Fuxin Li, Dimitris Samaras, and Chao Chen. Topology-preserving deep image segmentation. In *Advances in Neural Information Processing Systems*, pp. 5658–5669, 2019.

Michał Januszewski, Jörgen Kornfeld, Peter H Li, Art Pope, Tim Blakely, Larry Lindsey, Jeremy Maitin-Shepard, Mike Tyka, Winfried Denk, and Viren Jain. High-precision automated reconstruction of neurons with flood-filling networks. *Nature methods*, 15(8):605–610, 2018.

Davood Karimi and Septimiu E Salcudean. Reducing the hausdorff distance in medical image segmentation with convolutional neural networks. *arXiv preprint arXiv:1904.10030*, 2019.

Hoel Kervadec, Jihene Bouchtiba, Christian Desrosiers, Eric Granger, Jose Dolz, and Ismail Ben Ayed. Boundary loss for highly unbalanced segmentation. In *International Conference on Medical Imaging with Deep Learning*, pp. 285–296, 2019.

Genki Kusano, Yasuaki Hiraoka, and Kenji Fukumizu. Persistence weighted gaussian kernel for topological data analysis. In *International Conference on Machine Learning*, pp. 2004–2013, 2016.

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.

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*, pp. 3431–3440, 2015.

John Milnor. *Morse theory.(AM-51)*, volume 51. Princeton university press, 1963.

Volodymyr Mnih. *Machine learning for aerial image labeling*. University of Toronto (Canada), 2013.

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*, pp. 3136–3145, 2018.

Xiuyan Ni, Novi Quadrianto, Yusu Wang, and Chao Chen. Composing tree graphical models with persistent homology features for clustering mixed-type data. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, pp. 2622–2631. JMLR. org, 2017.

Sebastian Nowozin and Christoph H Lampert. Global connectivity potentials for random field models. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 818–825. IEEE, 2009.

Martin Ralf Oswald, Jan Stühmer, and Daniel Cremers. Generalized connectivity constraints for spatio-temporal 3d reconstruction. In *European Conference on Computer Vision*, pp. 32–46. Springer, 2014.

Adrien Poulenard, Primoz Skraba, and Maks Ovsjanikov. Topological function optimization for continuous shape matching. In *Computer Graphics Forum*, volume 37, pp. 13–25. Wiley Online Library, 2018.

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*, pp. 4741–4748, 2015.

V. Robins, P. J. Wood, and A. P. Sheppard. Theory and algorithms for constructing discrete morse complexes from grayscale digital images. *IEEE Trans. Pattern Anal. Machine Intelligence*, 33(8): 1646–1658, Aug 2011. ISSN 0162-8828.

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*, pp. 234–241. Springer, 2015.

Florent Ségonne. Active contours under topology control—genus preserving level sets. *International Journal of Computer Vision*, 79(2):107–117, 2008.

L Soler, A Hostettler, V Agnus, A Charnoz, J Fasquel, J Moreau, A Osswald, M Bouhadjar, and J Marescaux. 3d image reconstruction for comparison of algorithm database: a patient-specific anatomical and medical image database. *IRCAD, Strasbourg, France, Tech. Rep*, 2010.

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.

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*, pp. 2336–2343, 2013.

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.

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.

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*, pp. 1–8. IEEE, 2008.

S. Wang, Y. Wang, and Y. Li. Efficient map reconstruction and augmentation via topological methods. In *Proc. 23rd ACM SIGSPATIAL*, pp. 25. ACM, 2015.

Pengxiang Wu, Chao Chen, Yusu Wang, Shaoting Zhang, Changhe Yuan, Zhen Qian, Dimitris Metaxas, and Leon Axel. Optimal topological cycles and their application in cardiac trabeculae restoration. In *International Conference on Information Processing in Medical Imaging*, pp. 80–92. Springer, 2017.

Tao Zeng, Bian Wu, and Shuiwang Ji. Deepem3d: approaching human-level performance on 3d anisotropic em image segmentation. *Bioinformatics*, 33(16):2555–2562, 2017.

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.

Afra Zomorodian and Gunnar Carlsson. Computing persistent homology. *Discrete & Computational Geometry*, 33(2):249–274, 2005.

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.

A APPENDIX

In Sec. A.1, we will discuss different types of topological errors. In Sec. A.2, we will provide more details for the method part. In Sec. A.3, more details and results for the experiments will be provided.

A.1 DIFFERENT TYPES OF TOPOLOGICAL ERRORS

A.1.1 ILLUSTRATION OF 3D TOPOLOGICAL ERRORS

In Sec.2.1 of the main paper, we have already introduced discrete Morse theory with a 2D example. Here, we would like to illustrate 3D topological errors with 3D examples.

Fig. 7 and Fig. 8 illustrate two different types of topological errors for 3D data. Fig.7 illustrates an index-1 topological error for 3D synthetic data. 3D EM/neuron has the same type of topological error as the synthetic data. Fig.8 illustrates index-2 topological error for 3D vessel data.

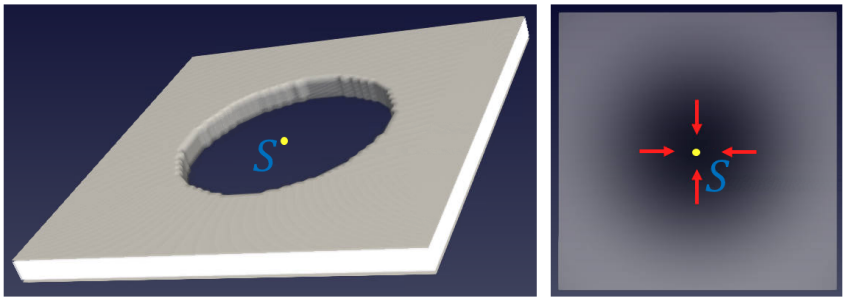


Figure 7: Illustration of an index-1 topological error (3D hole in the middle) for a 3D synthetic data. The ground truth is a complete sheet without the hole. We intentionally weaken the likelihood function in the middle. So the segmentation has a hole in the middle. **Left**: 3D segmentation result. S is a saddle point of the likelihood function. Its Hessian has 1 negative and 2 positive eigenvalues. The stable manifold of the saddle point S is a 2D plane going through the saddle point and cutting the segmentation into two thin slices. **Right**: the likelihood function visualized on the 2D stable manifold of S. Red arrows illustrate how different V-paths (streamlines of negative gradient) flow to the saddle S.

A.1.2 FALSE NEGATIVE AND FALSE POSITIVE ERRORS

In the main paper, we have mentioned that the proposed DMT-loss can capture and fix two different types of topological errors: false negative and false positive. We illustrate these two types in Fig. 9. The two highlighted red rectangles represent the two types of topological errors: 1) The red rectangle on the right represents a sample of false negative error; part of the membrane structure is missing, due to a blurred region near the membrane. 2) The red rectangle on the left represents a sample of false positive error. In this specific case, it is caused by mitochondria which are not the boundary of neurons.

In summary, with the help of the proposed DMT-loss, we can identify both these two types of topological errors, and then force the network to increase/decrease its likelihood value on these structures to correctly segment the images with with correct topology.

A.2 ADDITIONAL DETAILS ON THE METHOD

A.2.1 DISCRETE MORSE THEORY

We view a dD image, $d = 2$ or 3, as a d-dimensional cubical complex, meaning it consists of 0-, 1-, 2- and 3-dimensional cells corresponding to vertices, edges, squares, and voxels (cubes) as its building blocks.

Discrete Morse theory (DMT), originally introduced in (Forman, 1998; 2002), is a combinatorial version of Morse theory for general cell complexes. There are many beautiful results established for

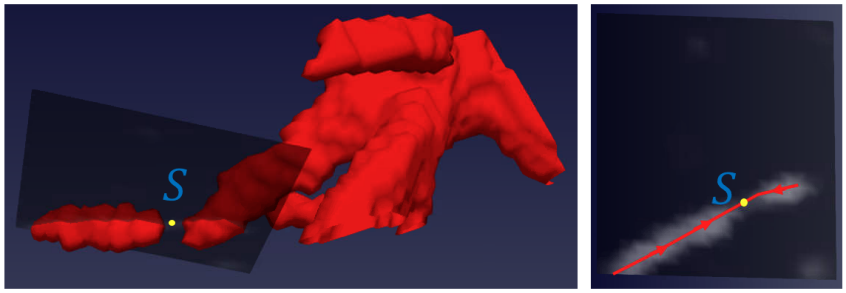
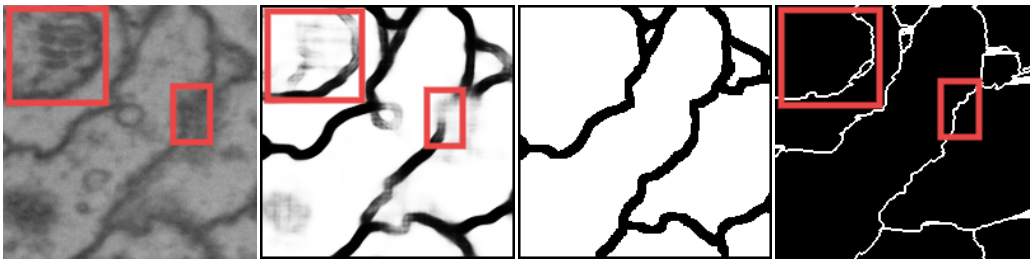


Figure 8: Illustration of an index-2 topological error from a 3D vessel image. The vessel segmentation has a broken connection near bottom-left of the Left image. **Left**: part of the 3D segmentation result. A The saddle point S corresponds to a broken connection. The Hessian of the likelihood at the saddle point has 2 negative and 1 positive eigenvalues. **Right**: One slice of the 3D likelihood map passing the saddle point. The saddle point (yellow) and its 1-stable manifold (red) are also drawn.



(a) (b) (c) (d)

Figure 9: Illustration of two different types of topological errors captured by DMT-loss. **(a)** an input neuron image with challenging locations highlighted. **(b)** likelihood map of a baseline method without topological guarantee (Ronneberger et al., 2015). **(c)** ground truth. **(d)** The topologically critical structure from the likelihood, captured by the proposed discrete Morse algorithm. These structures will be used in the DMT-Loss.

DMT, analogous to classical Morse theory. We will however only briefly introduce some relevant concepts for the present paper, and we will describe it in the setting of cubical complexes (instead of simplicial complexes) as it is more suitable for images.

Let K be a cubical complex. Given a p-cell τ, we denote by $\sigma < \tau$ if σ is a $(p-1)$-dimensional face for τ. A *discrete gradient vector* (also called *a V-pair* for simplicity) is a pair (τ, σ) where $\sigma < \tau$. Now suppose we are given a collection of V-pairs $\mathsf{M}(K)$ over the cubical complex K. A sequence of cells $\pi : \tau_0^{p+1}, \sigma_1^p, \tau_1^{p+1}, \sigma_2^p, \cdots, \sigma_k^p, \tau_k^{p+1}, \sigma_{k+1}^p$, where the superscript p in α^p stands for the dimension of this cell α, form a *V-path* if $(\tau_i, \sigma_i) \in \mathsf{M}(K)$ for for any $i \in [1, k]$ and $\sigma_i < \tau_{i-1}$ for any $i \in [1, k+1]$. A V-path π is *acyclic* if $(\tau_0, \sigma_{k+1}) \notin \mathsf{M}(K)$. This collection of V-pairs $\mathsf{M}(K)$ form a *discrete gradient vector field*[4] if (cond-i) each cell in $\mathsf{M}(K)$ can only appear in at most one pair in $\mathsf{M}(K)$; and (cond-ii) all V-paths in $\mathsf{M}(K)$ are acyclic. Given a discrete gradient vector field $\mathsf{M}(K)$, a simplex $\sigma \in K$ is *critical* if it is not in any V-pair in $\mathsf{M}(K)$.

Even though a discrete gradient vector (a V-pair), say (τ, σ) is a combinatorial pair instead of a real vector, it still indicates a "flow" from τ to its face σ. A V-path thus corresponds to a flow path (integral line) in the smooth setting. However, to make a collection of V-pairs a valid analog of gradient field, (cond-i) says that at each simplex there should only be one "flow" direction; while (cond-ii) is necessarily as flow lines traced by gradient can only go down in function values and thus never come back (thus acyclic).

[4]We will not introduce the concept of discrete Morse function, as the discrete gradient vector field is sufficient to define all relevation notations.

A critical simplex has "vanishing gradient" as it is not involved in any V-pair in $\mathsf{M}(K)$ (i.e, there is no flow at this simplex). Given a 2D cubical complex K a a discrete gradient vector field $\mathsf{M}(K)$, we can view critical 0-, 1-, 2- and 3-cells as minima, saddle points, and maxima, respectively. If K is 3D, then we can view critical 0-, 1-, 2- and 3-cells as minima, index-1 saddle, index-2 saddle and maxima, respectively.

Hence, a 1-stable manifold in 2D will correspond to a V-path connecting a critical square (a maximum) and a critical edge (a saddle), while in 3D, it will be a V-path connecting a critical cube and a critical square.

Morse cancellation. A given discrete gradient field $\mathsf{M}(K)$ could be noisy, e.g, there are shallow valleys where the mountain ridge around it should be ignored. Fortunately, the discrete Morse theory provides an elegant and purely combinatorial way to cancel pairs of critical simplices (and thus reduce their stable manifolds). In particular, given $\mathsf{M}(K)$, a pair of critical simplices $\langle \delta^{(p+1),\gamma^p} \rangle$ is *cancellable* if there is a unique V-path $\pi = \delta = \delta_0, \gamma_1, \delta_1, \ldots, \delta_s, \gamma_{s+1} = \gamma$ from δ to γ. The *Morse cancellation operation* simple reverse all V-pairs along this path by removing all V-pairs along these path, and adding (δ_{i-1}, γ_i) to $\mathsf{M}(K)$ for any $i \in [1, s+1]$. It is easy to check that after the cancellation neither δ nor γ is critical.

A.2.2 Persistence pruning

We can extend this vertex-valued function ρ to a function $\rho : K \to \mathbb{R}$, by setting $\rho(\sigma)$ for each cell to be the maximum ρ-value of each vertex in σ. How to obtain a discrete gradient vector field from such function $\rho : K \to \mathbb{R}$? Following the approach developed in (Wang et al., 2015; Dey et al., 2018), we initialize a trivial discrete gradient vector field where all cells are initially critical. Let $\epsilon > 0$ be a threshold for simplification. We then perform persistence algorithm (Edelsbrunner et al., 2000) induced by the super-level set filtration of ρ and pair up all cells in K, denoted by $\mathcal{P}_\rho(K)$.

Persistent homology is one of the most important development in the field of topological data analysis in the past two decades (Edelsbrunner & Harer, 2010; Edelsbrunner et al., 2000; Zomorodian & Carlsson, 2005). We will not introduce it formally here. Imagine we grow the complex K by starting from the empty set and gradually include more and more cells in decreasing ρ values. (More formally, this is the so-called super-level set filtration of K induced by ρ.) Through this course, new topological feature can be created upon adding a simplex σ, and sometiems a feature can be destroyed upon adding a simplex τ. Persistence algorithm (Edelsbrunner et al., 2000) will pair up simplices; that is, its output is a set of pairs of simplices $\mathcal{P}_\rho(K) = \{(\sigma, \tau)\}$, where each pair captures the birth and death of topological features during this evolution. The persistence of a pair, say $\mathsf{p} = (\sigma, \tau)$, is defined as $\mathrm{pers}(\mathsf{p}) = \rho(\sigma) - \rho(\tau)$, measuring how long the topological feature captured by p lives in term of ρ. In this case, we also write $\mathrm{pers}(\sigma) = \mathrm{pers}(\tau) = \mathrm{pers}(\mathsf{p})$ – the persistence of a simplex (say σ or τ) can be viewed as the importance of this simplex.

With this intuition of the persistence pairings, we next perform Morse-cancellation operation to all pairs of these cells $(\sigma, \tau) \in \mathcal{P}_\rho(K)$ in increasing order their persistence if (i) its persistence $\mathrm{pers}(\delta, \gamma) < \epsilon$ (i.e, this pair has low persistence and thus not important); and (ii) this pair (δ, γ) is cancellable.

Let $\mathsf{M}_\epsilon(K)$ be the resulting discrete gradient field after simplifying all low-persistence critical simplices. We then construct the 1-stable and 2-stable manifolds for the remaining (high persistence, and thus important) saddles (critical 1-cell and 2-cells) from $\mathsf{M}_\epsilon(K)$. Let $\mathcal{S}_1(\epsilon)$ and $\mathcal{S}_2(\epsilon)$ be the resulting collection of 1- and 2-stable manifolds respectively. In particular, see an illustration of a V-path (highlighted in black) corresponding to a 1-stable manifold of the green saddle in Fig. 3(c).

A.2.3 More details on the approximation of $\mathcal{S}_2$ via $\hat{S}_2$.

We approximate S_2 by taking the boundary of the stable manifold of the minima (basins/valleys in the terrain). This is like a watershed algorithm: growing the basins from all minima until they meet. The stable manifolds of the minima is approximated using spanning trees. This algorithm is inspired by the continuous analog for Morse functions.

A.3 EXPERIMENTS

A.3.1 DATASETS

We conduct experiments on six different 2D datasets. The first three datasets are neuron images (Electron Microscopy images). The task is to segment membranes and eventually partition the image into neuron regions.

CREMI contains 125 images. The resolution for each image is 1250x1250.

ISBI12 (Arganda-Carreras et al., 2015) contains 30 images. The resolution for each image is 512x512.

ISBI13 (Arganda-Carreras et al., 2013) contains 100 images. The resolution for each image is 1024x1024.

The next three are natural image datasets, and their structures are vital for their functionality.

CrackTree (Zou et al., 2012) contains 206 images of cracks in road. The resolution for each image is 600x800.

Road (Mnih, 2013) has 1108 images from the Massachusetts Roads Dataset. The resolution for each image is 1500x1500.

DRIVE (Staal et al., 2004) is a retinal vessel segmentation dataset with 20 images. The resolution for each image is 584x565.

A.3.2 EVALUATION METRICS

We use five different evaluation metrics to evaluate the proposed DMT-loss.

Pixel-wise accuracy: Pixel-wise accuracy is one of the most common metrics which measures the percentage of correctly classified pixels.

DICE score: DICE score (also known as DICE coefficient, DICE similarity index) is the same as the F1 score.

Adapted Rand Index (ARI): ARI is 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).

Variation of Information (VOI): 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.

Betti number error: which directly compares the topology (number of handles) between the segmentation and the ground truth. We randomly sample patches over the segmentation and report the average absolute difference between their Betti numbers and the corresponding ground truth patches.

A.3.3 QUANTITATIVE RESULTS FOR MORE 2D DATASETS

Table. 3 shows quantitative results for ISBI12 and DRIVE.

A.3.4 FAIRNESS COMPARISONS WITH SAME BACKBONE NETWORKS

We copy the numbers of TopoLoss from (Hu et al., 2019), which is TopoLoss+DIVE. And in this paper, we use U-Net as the backbone. Indeed, with U-Net, TopoLoss will be worse and the gap will be even bigger. The DIVE used in (Hu et al., 2019) is more expensive and better designed specifically for EM images. We choose U-Net in this manuscript as it is lightweight and easy to generalize to many datasets. We also apply our backbone-agnostic DMT-loss to the DIVE network (Fakhry et al., 2016). All the experiments are conducted on CREMI 2D dataset. The quantitative results (Betti Error) are shown in the Table 4.

A.3.5 QUALITATIVE RESULTS FOR 3D DATASETS

Fig. 10 shows qualitative results for ISBI13 dataset.

Table 3: Quantitative results for different models on several 2D medical datasets

Method	Accuracy	DICE	ARI	VOI	Betti Error
ISBI12					
DIVE	0.9640 ± 0.0042	0.9709 ± 0.0029	0.9434 ± 0.0087	1.235 ± 0.025	3.187 ± 0.307
U-Net	0.9678 ± 0.0021	0.9699 ± 0.0048	0.9338 ± 0.0072	1.367 ± 0.031	2.785 ± 0.269
Mosin.	0.9532 ± 0.0063	0.9716 ± 0.0022	0.9312 ± 0.0052	0.983 ± 0.035	1.238 ± 0.251
TopoLoss	0.9626 ± 0.0038	0.9755 ± 0.0041	0.9444 ± 0.0076	0.782 ± 0.019	$\mathbf{0.429 \pm 0.104}$
DMT	0.9593 ± 0.0035	$\mathbf{0.9796 \pm 0.0033}$	$\mathbf{0.9527 \pm 0.0052}$	$\mathbf{0.671 \pm 0.027}$	$\mathbf{0.391 \pm 0.114}$
DRIVE					
DIVE	0.9549 ± 0.0023	0.7543 ± 0.0008	0.8407 ± 0.0257	1.936 ± 0.127	3.276 ± 0.642
U-Net	0.9452 ± 0.0058	0.7491 ± 0.0027	0.8343 ± 0.0413	1.975 ± 0.046	3.643 ± 0.536
Mosin.	0.9543 ± 0.0047	0.7218 ± 0.0013	0.8870 ± 0.0386	1.167 ± 0.026	2.784 ± 0.293
TopoLoss	0.9521 ± 0.0042	0.7621 ± 0.0036	0.9024 ± 0.0113	1.083 ± 0.006	$\mathbf{1.076 \pm 0.265}$
DMT	0.9495 ± 0.0036	$\mathbf{0.7733 \pm 0.0039}$	$\mathbf{0.9077 \pm 0.0021}$	$\mathbf{0.876 \pm 0.038}$	$\mathbf{0.873 \pm 0.402}$

Table 4: Comparison with same backbones

	U-Net	DIVE
TopoLoss	1.451 ± 0.216	1.113 ± 0.224
DMT-Loss	$\mathbf{0.982 \pm 0.179}$	$\mathbf{0.956 \pm 0.142}$

A.3.6 THE ABLATION STUDY FOR BALANCED TERM β.

We conduct another ablation study for the balanced weight of parameter β. To be noted that, the parameter β is dataset dependent. We conduct the ablation experiment on CREMI 2D dataset. When $\beta = 3$, the proposed DMT-loss achieves best performance 0.982 (Betti Error). When $\beta = 2$ and $\beta = 4$, the performance drops to 1.074 and 1.181 (both in Betti Error), respectively. The parameter β is also chosen via cross-validation.

A.3.7 COMPARED WITH RE-WEIGHTING CROSS ENTROPY LOSS

We run an additional ablation study to compare with a baseline of reweighting the FP and FN pixels in cross-entropy loss (Reweighting CE). The weights of the FP/FN pixels are hyperparameters tuned via cross-validation. DMT-loss is better than this baseline. The reason is that the DMT-loss specifically penalizes FP and FN pixels which are topology-critical. Meanwhile, reweighting CE adds weights to FP/FN pixels without discrimination. A majority of these misclassified pixels are not topology-critical. They are near the boundary of the foreground. Please refer to Fig. 11 for an illustration.

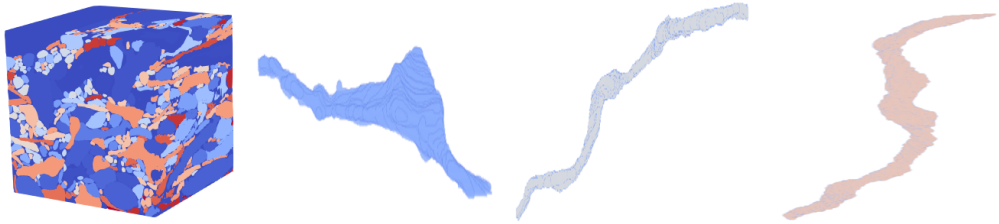


Figure 10: Segmentation results for ISBI13 dataset and 3 randomly selected neurons.

(a) (b) (c) (d)

Figure 11: Illustration of the comparison between our proposed DMT-loss and simple re-weighting cross entropy loss. Please refer to Fig.9(b) for the likelihood map of a baseline method without topological guarantee. **(a)** an input neuron image. **(b)** The topologically critical structures from the likelihood, captured by the proposed discrete Morse algorithm. These structures will be used in the DMT-Loss. **(c)** ground truth. **(d)** The FP/FN pixels identified by simple re-weighting cross entropy loss.