Topology-Preserving Hard Pixel Mining for Tubular Structure Segmentation

Guoqing Zhang[1,2]
zhanggq21@mails.tsinghua.edu.cn
Caixia Dong[3]
caixia_dong@xjtu.edu.cn
Yang Li[1]
yangli@sz.tsinghua.edu.cn

[1] Tsinghua-Berkeley Shenzhen Institute Tsinghua University
[2] Peng Cheng Laboratory
[3] Institute of Medical Artificial Intelligence The Second Affiliated Hospital of Xi'an Jiaotong University

Abstract

Maintaining accurate topology in image segmentation, especially for tubular structures such as borders of neurons, vessels or roads, is essential in many research fields but remains unsolved. Existing neural segmentation networks are prone to topological errors on these fine-scale structures. We propose a skeleton-based hard pixel mining strategy and incorporate it into training end-to-end segmentation networks. In particular, we penalize the most topology-relevant yet mis-segmented pixels in overall cost function. Skeletal hard pixels are mined through several simple logical and morphological operations which are fast and scalable. We evaluate our method on six public datasets of different domains, including neurons, roads and coronary arteries. Our approach demonstrates training speed comparable to the base model, while achieving a significant reduction in the number of erroneous connected components by 17.14% and loops by 53.87%, respectively. Extensive experimental results demonstrate that our method outperforms state-of-the-art computational topology-based method in topological preservation while maintaining competitive volumetric accuracy. Code is available at the public git repository: https://github.com/wlsdzyzl/topohpm.

1 Introduction

As a fundamental task in computer vision, image segmentation has been widely studied and undergone significant developments since the emergence of deep learning. State-of-the-art segmentation models learn multi-scale feature representations through convolutional layers and generate high-quality predictions with satisfactory per-pixel accuracy. Unfortunately, existing methods still suffer from the inaccurate topology of predictions when facing tubular or network-like structures such as blood vessels, nerve cells, road networks and circuits, because topologically incorrect predictions like structures with unintended breaks, may have very few per-pixel errors. On the other hand, topologically accurate segmentation of these structures is crucial in various applications [22, 23, 42]. For example, the correct topology of road networks is necessary for optimal path planning in smart transportation as shown in Fig. 1. In biomedical engineering, predicting the accurate topology of vessels and neurons is helpful for analyzing pathological mechanisms and developing new therapies.

Focusing on hard pixels is a common way to improve the performance of image segmentation models. One of the simplest hard pixel selection methods is to use a threshold to determine the predicted label map [57]. All mis-segmented pixels are considered as hard pixels. Another common solution is to assign a continuous weight map based on the predicted probability and label [7, 24, 29, 41]. These methods perform operations on the whole image and effectively attack class imbalance, but they are not sensitive to topological errors. Hu et al. [20] propose a computational topology-based loss function that enforces a segmentation to have the same Betti number as the ground truth. This strategy has three obvious drawbacks. First, directly aligning the topological features of prediction and ground-truth is insufficient from a geometric perspective. Secondly, the dimensions of topological features need to be carefully chosen according to the specific characteristics of the datasets. Additionally, the expensive computational complexity of topological feature analysis makes it challenging to extend persistent homology-based methods to 3D images.

Figure 1: Significance of topological accuracy in road segmentation and subsequent route planning. The figure contains an aerial image from Massachusetts alongside two segmented road maps predicted by U-Net (top) [40] and ours (bottom), respectively. These two segmentation results achieve similar quality in per-pixel accuracy, but the top one is observed to have more broken roads. Road graphs are extracted from these segmentation [6] and further generate different optimal paths to visit the four locations in the image, which are depicted in red and blue, respectively. Bottom segmentation is evidently preferred, because the blue path is much shorter than the red one.

Our method follows the insight behind hard pixel mining (HPM) and tries to emphasize the importance of the most topology-relevant hard pixels.We adopt the skeleton representation to identify hard pixel due to the work by Choi et al. [10] which proved that the skeleton of an object serves as a strong deformation retract of the original structure and preserves all crucial topological information. In this study, we mine mis-classified skeletal pixels as hard pixels through several simple morphological operations, leading to only a small increase in computational and storage burden. We believe these hard pixels are topologically critical and assign them increased weights for topology preservation. This simple strategy makes our method suitable for training large image patches, allows it to be seamlessly plugged into the training of arbitrary segmentation networks. Additionally, our method achieves competitive

quality in terms of volumetric metrics and demonstrates superiority in topology preservation compared to existing approaches, including persistent homology-based methods.

2 Related Works

Fine-grained segmentation: Various models are proposed to capture fine details of objects in dense image segmentation. Long et al. [30] firstly introduce a fully convolutional network with a skip architecture for semantic segmentation. Chen et al. [8, 9] explore atrous convolution and combine the network with a fully connected conditional random field. Noh et al. [36] learn a deconvolution network with deconvolution and unpooling layers. Li et al. [28] fuses hierarchical features captured by a global transformer and dual local attentions. Deng et al. [13] propose a scale-aware controller to model scale information in renal pathological images. All these methods represent improvements in model architecture and demonstrate notable enhancements in segmentation quality. However, they focus on pixel-wise accuracy and are topology-agnostic.

Class imbalance: The training of tubular structure segmentation model usually faces a large foreground-background imbalance, especially for 3D images [49]. A common approach for addressing class imbalance is to perform hard case mining [17, 45, 48, 53] that selects difficult samples during training. Another solution involves more complex reweighing schemes. Rota Bulo et al. [41] implement loss max pooling to reweigh the distribution of each pixel. Caliva et al. [7] and Kervadec et al. [24] compute weighted loss based on boundary distance. Although the aforementioned literature has achieved great progress in class imbalance, their effectiveness in preserving topology is not particularly evident.

Topology preservation: A range of classical algorithms in binary image processing [26, 27, 31, 59] that employ thinning and medial axis techniques have demonstrated their ability to preserve topology. For topology preservation in non-binary image analysis, while some random field-based methods [3, 19, 37, 38, 39, 43, 50, 52, 55] that utilize shape priors of objects have been proven to be effective. Stuhmer et al. [47] and Kong et al. [25] leverage tree structures in vessel segmentation. Andres et al. [2] investigate the role of closedness constraints in the graph representation of natural images. Vasu et al. [51] jointly train a road segmentation model and a pyramidal discriminator for topological correctness. These approaches require prior assumptions and lack end-to-end topology-preserving solutions.

The closest method to ours is introduced by Shit et al. [44], who implement a differentiable soft skeletonization module through pooling functions and construct a center line-based Dice loss. However, their skeletonization algorithm is constrained by the box filter size and incapable of extracting a fine skeleton. In practice, it is also observed that the result of soft skeletonization does not necessarily maintain the same topological structure as the original object. Mosinska et al. [35] utilize the output of several filters derived from a pre-trained VGG19 neural network to formulate the loss function. These specific filters have been identified as exhibiting a preference for elongated shapes. However, this approach shows fragility on more complex connections of arbitrary shapes. Hu et al. [20, 21], Yang et al. [56] and Clough et al. [12] propose topology-aware cost functions based on persistent homology and Betti numbers [14, 15]. Nevertheless, the high computational complexity of topology analysis limits their applications on real image-size patches.

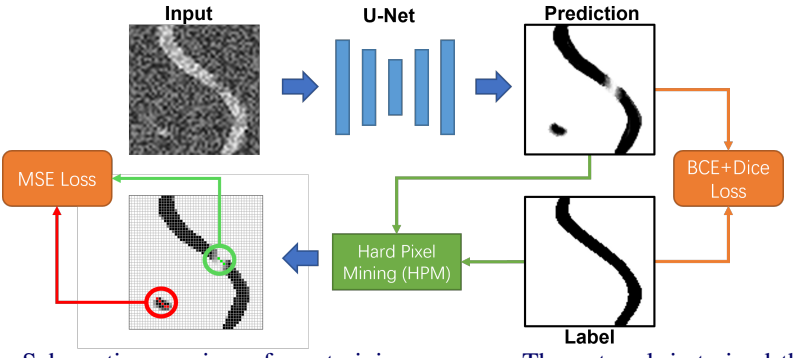

Figure 2: Schematic overview of our training process. The network is trained through a combination of binary cross-entropy and Dice loss. Meanwhile, HPM is performed over prediction and ground truth. An additional loss of hard pixels is calculated as topological constraints.

3 Methods

Our method is designed to address the topological vulnerability of deep segmentation networks in the face of tubular structures. An overview of the training process is presented in Fig 2, in which a U-Net is trained for semantic segmentation. Our HPM module receives both prediction and ground truth as its inputs and utilizes them to determine a set of hard pixels. We employ a hybrid loss integrating both binary cross-entropy (BCE) and Dice loss to train the U-Net. Additionally, we compute the mean squared error (MSE) for hard pixels, so that positive hard pixels marked in green are used to preserve connectivity and negative hard pixels marked in red are used to reduce noise. Assigning a higher penalty to hard pixels within the overall loss function is crucial for training topology-preserving networks.

The remainder of this section is organized as follows: we give the mathematical formulation of our cost functions in Sec. 3.1. Sec. 3.2 introduces our HPM module. Sec.3.3 explains several crucial details of the training algorithms. While we focus on a specific network architecture, it is important to note that our method is generalizable and can be applied to arbitrary segmentation models.

3.1 Cost Function

As mentioned before, the base model is trained with a combination of BCE and Dice loss. We further add a HPM-based cost to enhance the model's ability of topology preservation. The overall loss function can be formulated as follows:

$$\mathcal{L} = \mathcal{L}_{\text{BCE}}(\mathbf{p},\mathbf{y}) + \mathcal{L}_{\text{Dice}}(\mathbf{p},\mathbf{y}) + \lambda \cdot \mathcal{L}_{\text{HPM}}(H,\mathbf{p},\mathbf{y}), \tag{1}$$

where $\mathbf{p}$ is the estimated likelihood map, $\mathbf{y}$ specifies the ground-truth, $\mathcal{L}_{\text{BCE}}$ and $\mathcal{L}_{\text{Dice}}$ denote binary cross-entropy and Dice loss, respectively. BCE directly optimizes probability distributions to achieve high pixel-wise accuracy while Dice loss is more robust to class imbalance and captures fine structures. Combining them leads to an improved performance which is commonly used in medical image segmentation. The topological constraint introduced by

HPM is denoted as $\mathcal{L}_{\text{HPM}}$:

$$\mathcal{L}_{\text{HPM}}(H,\mathbf{p},\mathbf{y}) = \frac{1}{|H|}\sum_{h\in H}(\mathbf{p}[h]-\mathbf{y}[h])^2, \tag{2}$$

where H is the hard pixel set. The hyper-parameter λ controls the strength of topological constraints and is empirically set to 0.5.

3.2 Hard Pixel Mining

In this section, we focus on how to obtain the hard pixel set H. We start from introducing a threshold-based HPM in Sec. 3.2.1. In Sec. 3.2.2, we provide a brief overview of a persistent homology-based approach for comparison. The proposed skeleton-based method is elaborated in Sec. 3.2.3.

3.2.1 Threshold-based HPM

One of the simplest hard pixel mining strategies is to find all mis-segmented pixels that are determined through a probability threshold. The threshold-based hard pixel mining, which is called **THPM** in our experiments, can be formulated as:

$$H_{\text{t}} = \{h\,|\,(\mathbf{p}_t \oplus \mathbf{y})\,[h] = 1\,\}, \tag{3}$$

where $\mathbf{p}_t$ is the predicted label map determined by a probability threshold t which is set to 0.6 empirically, $\oplus$ denotes **XOR** operator. As shown in Fig. 4(a), THPM selects a large number of pixels located at the edge of the foreground as hard samples, which are actually trivial for maintaining the correct topology. Adding penalty on these pixels may lead to a significant decrease in pixel-level accuracy (see Sec. 4.3).

3.2.2 Persistent homology-based HPM

We also provide an approach named **PHPM** to mine hard pixels based on persistent homology, which captures all topological features of given dimensions for the given object and generates its persistent diagram [14, 16]. Each point from the persistent diagram corresponds to the birth and death of a topological feature. Similar to [20], we adopt cubical complex for topology analysis because of its natural adaptability to grid-like data [1, 32, 54], in which every birth or death of the topological feature can be localized to a certain grid called critical point. Instead of directly matching and aligning the critical points like Hu et al. [20], we consider critical points as hard pixels. PHPM can be formulated as follows:

$$H_{\text{p}} = \bigcup_{d\in D} C_d\left(f\left(\mathbf{p},\mathbf{y},t\right)\right), \tag{4}$$

where the set D encompasses the dimensions of topological features that hold our interest, $C_d(\cdot)$ denotes the function to get the d-dim critical points of input image and f is a preprocessing function to reduce the complexity of topology analysis. Specifically, $f(\mathbf{p},\mathbf{y},t)$ returns a matrix $\mathbf{p}_f$ with the same shape as input $\mathbf{p}$ and $\mathbf{y}$. For each quadruplet (p_f, p_t, p, y) that consists of elements taken from the same position of $\mathbf{p}_f$, $\mathbf{p}_t$, $\mathbf{p}$ and $\mathbf{y}$, we have:

$$p_f = \begin{cases} p, & \text{if } p_t \neq y; \\ y, & \text{otherwise.} \end{cases} \tag{5}$$

The topology of a predicted likelihood map could be very complicated and most of the topological features are negligible. We use function f to increase the intensity of all true positive pixels to 1.0 and reduce the intensity of all true negative pixels to 0.0, so that the topology analysis can pay more attention on mis-segmented pixels. Note that PHPM is an improved version of [20]. A detailed introduction about PHPM and the comparison between [20] and PHPM are provided in the supplementary materials.

We visualize the hard pixel set mined through PHPM in Fig. 4(b), where we can see that PHPM only generates very sparse results. Furthermore, analyzing the topology of an image patch is quite expensive, which makes it difficult to be applied on large image patches.

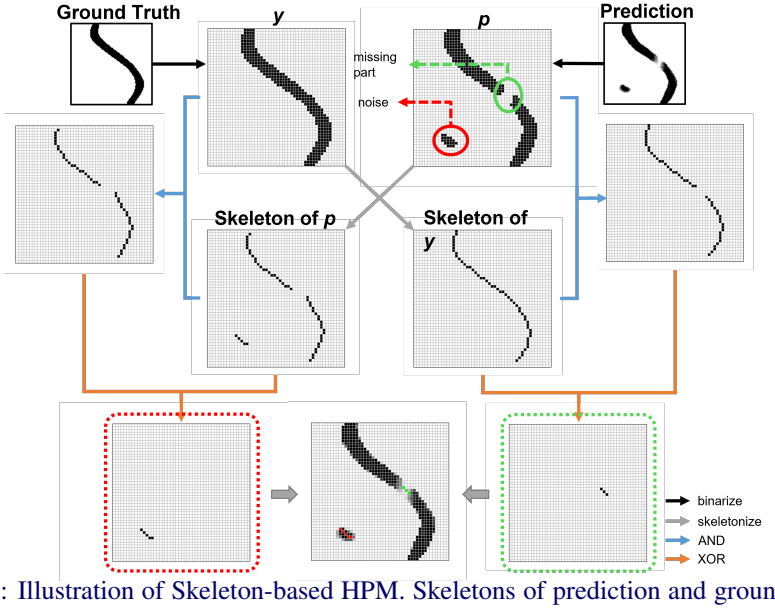

Figure 3: Illustration of Skeleton-based HPM. Skeletons of prediction and ground truth are used to extract negative hard pixels and positive hard pixels, respectively.

3.2.3 Skeleton-based HPM

To overcome the disadvantages of THPM, we further propose **SHPM**, a skeleton-based hard pixel mining strategy. We extract thinner structures of the mis-segmented parts and ignore topologically irrelevant pixels through several simple logical and morphological operations as illustrated in Fig. 3. In specific, the skeletons of ground truth and binarized prediction are extracted by a successive thinning algorithm [59] that iteratively removes pixels on borders of object until the pattern remains unchanged. We use the skeleton of ground-truth to calculate positive hard pixels as defined below:

$$H_{\text{pos}} = \{h | ((\mathbf{p}_t \wedge \text{s}(\mathbf{y})) \oplus \text{s}(\mathbf{y}))[h] = 1\}, \tag{6}$$

where $\wedge$ denotes **AND** operator, $\text{s}(\cdot)$ is the skeletonization function that returns a binary matrix of the same dimension as the input image, where the values of skeletal points are set to 1. Vice-a-versa, we formulate the mining of negative hard pixels as follows:

$$H_{\text{neg}} = \{h | ((\text{s}(\mathbf{p}_t) \wedge \mathbf{y}) \oplus \text{s}(\mathbf{p}_t))[h] = 1\}. \tag{7}$$

The resulting hard pixel set is the union of the above two:

$$H_s = H_{pos} \cup H_{neg}. \tag{8}$$

SHPM emphasizes the importance of skeleton-related mis-segmented pixels as shown in Fig. 4(c). Because skeleton is a thinner structure compared to the original object and remains the same topology, a reasonable prediction should cover the skeleton of the ground truth. Conversely, the ground truth should include the skeleton of its predicted counterpart. Through SHPM we can find the inconsistent parts among these structures. Unlike THPM that penalizes all mis-classified pixels, SHPM disregards the insignificant false pixels located at mask borders, so that we can give a higher penalty for incorrect topology.

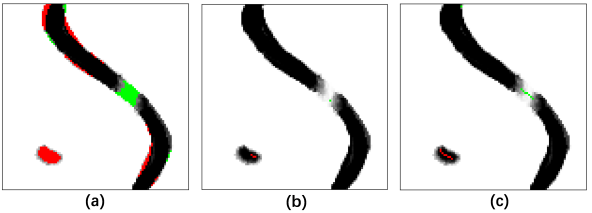

Figure 4: Comparison of mined hard pixel set with different HPM algorithms. Positive and negative hard pixels are marked in green and red, respectively. (a) Threshold-based HPM. (b) Persistent homology-based HPM. (c) Skeleton-based HPM.

3.3 Model Training

Although our proposed method is model-agnostic, we need to select a specific architecture to validate its efficacy. We choose U-Net as our base model due to its wide applications in both 2D and 3D image segmentation [11, 40, 46]. Our network contains 6 double convolutional layers with the first three comprising the encoder, where each layer is followed by a max-pooling layer. Other layers constitute the decoder, with each layer being preceded by an upsampling layer. Due to the computational complexity, the training data is fed into the network in a patch manner. The patch size is 256×256 for 2D images and $128 \times 128 \times 72$ for 3D images. We use stochastic gradient descent to train the base model with a learning rate of 0.01. Topology preservation are performed on pre-trained models. When training with additional loss functions, the learning rates start from 0.001 and follow a plateau schedule.

4 Experiments

4.1 Datasets

We evaluate our method on six public datasets. CREMI-A, CREMI-B, CREMI-C [18] and ISBI12 [8] are 2D neuron image segmentation datasets, in which each of the first three contains 125 images of size 1250×1250 and the last contains 60 images of size 512×512. The Massachusetts Roads dataset [34] contains 1171 images of size 1500×1500 for road segmentation. ICAS-d contains 100 down-sampled 3D images from a large coronary artery segmentation dataset ImageCAS [58]. The images have a size of $128 \times 128 \times 72$ voxels. For all datasets, we perform a five-fold cross validation and report the mean performance over the validation sets.

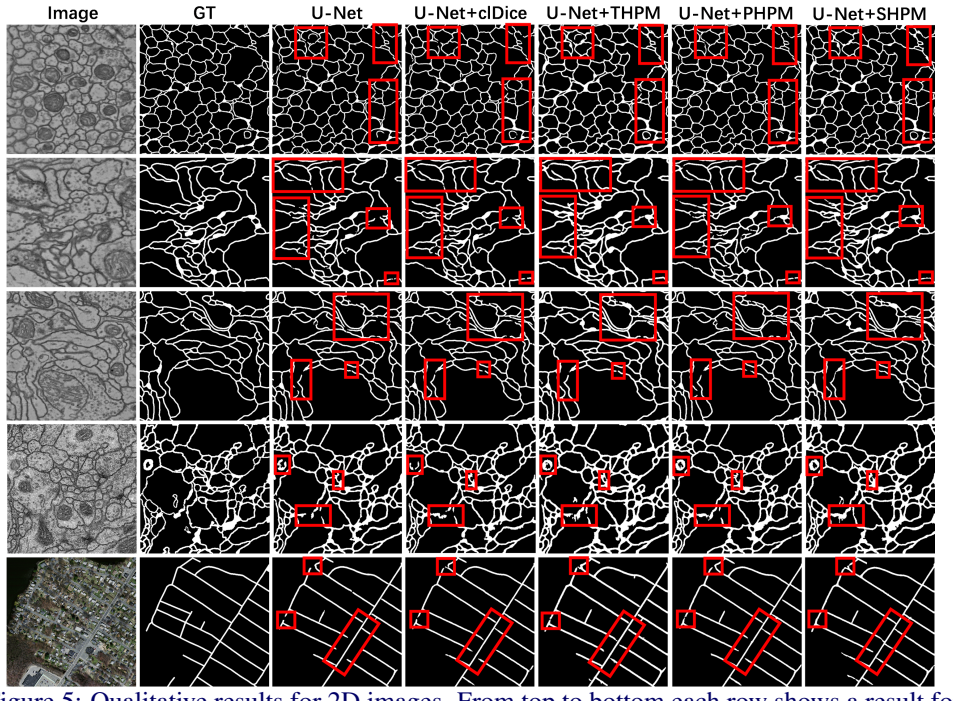

Figure 5: Qualitative results for 2D images. From top to bottom each row shows a result for the CREMI-A, CREMI-B, CREMI-C and ISBI12 neuron datasets, and the Massachusetts road dataset. Columns from left to right are the raw images, the ground truths and the predictions of different approaches. The images indicate that SHPM exhibits superior topological preservation capabilities. Some of the topological discrepancies are highlighted with red boxes.

4.2 Evaluation Metrics

We employ volumetric, clustering-based and topology-based metrics to compare the performance of various experimental settings. For volumetric metrics, we compute **Dice coefficient (Dice)** and the **mean Intersection of Union (mIoU)**. These measurements are topology-agnostic. Two clustering-based metrics are used to compare the partitioning of space by segmented tubular structures: **Variation of Information (VOI)** [33] and **Adapted Rand Error (ARE)** [4]. They are topology-relevant because different topology can result in distinct spatial partitions. ARE is defined as one minus the maximal F-score of the Rand index, with the background zero component excluded. VOI is a measure of the distance between two clusters through mutual information. To directly evaluate the topological disagreements between prediction and ground truth, we calculate the **mean of absolute Betti errors** over Betti numbers β-0 and β-1, that count the number of connected components and holes, respectively. It is worth noting that, unlike Hu et al. [20] and Shit et al. [44] only performing topology-based evaluations on random small patches, all our evaluations are performed on the whole images for both 2D and 3D datasets.

Table 1: Quantitative experimental results.

Dataset	Method	Dice	mIoU	VOI	ARE	β-0 Error	β-1 Error
CREMI-A	U-Net	0.9145	0.8425	0.3969	0.2337	2.176	23.960
	U-Net + clDice	**0.9148**	**0.8431**	0.3958	0.2310	2.200	23.552
	U-Net + THPM	0.9007	0.8195	0.3752	0.1076	2.016	8.960
	U-Net + PHPM	0.9133	0.8405	0.3819	0.2060	2.112	20.352
	U-Net + SHPM	0.9106	0.8326	**0.3426**	**0.1042**	**1.648**	**7.240**
CREMI-B	U-Net	0.8545	0.7506	1.1442	0.6221	18.296	63.864
	U-Net + clDice	**0.8547**	**0.7507**	1.1432	0.6168	17.680	63.584
	U-Net + THPM	0.8461	0.7368	0.7979	0.4074	13.968	29.112
	U-Net + PHPM	0.8540	0.7502	1.0444	0.5724	15.672	56.056
	U-Net + SHPM	0.8509	0.7446	**0.7537**	**0.3806**	**11.968**	**23.832**
CREMI-C	U-Net	0.8937	0.8083	0.6905	0.3693	6.699	46.577
	U-Net + clDice	**0.8951**	**0.8105**	0.6840	0.3607	6.626	45.967
	U-Net + THPM	0.8852	0.7943	0.5366	0.2143	5.569	18.260
	U-Net + PHPM	0.8934	0.8077	0.6383	0.3243	5.772	39.089
	U-Net + SHPM	0.8915	0.8046	**0.4990**	**0.1981**	**4.520**	**12.106**
ISBI12	U-Net	0.8304	0.7108	0.6901	0.1616	3.017	8.300
	U-Net + clDice	0.8300	0.7102	0.6985	0.1706	**2.750**	7.883
	U-Net + THPM	0.8260	0.7058	0.7077	**0.1027**	3.250	7.433
	U-Net + PHPM	0.8313	0.7123	0.6805	0.1400	3.200	7.583
	U-Net + SHPM	**0.8314**	**0.7128**	**0.6759**	0.1160	2.900	**6.750**
Roads	U-Net	0.7317	0.5845	0.7025	0.3191	9.671	28.116
	U-Net + clDice	0.7307	0.5835	0.7520	0.3460	9.759	29.676
	U-Net + THPM	0.7242	0.5744	0.5863	0.2501	9.062	22.096
	U-Net + PHPM	0.7295	0.5823	0.7339	0.3372	8.826	29.232
	U-Net + SHPM	**0.7331**	**0.5858**	**0.5551**	**0.2381**	**7.510**	**20.332**
ICAS-d (3D)	U-Net	0.6129	0.4482	0.0094	0.0011	5.410	0.000
	U-Net + clDice	0.5867	0.4208	0.0098	0.0012	6.440	0.000
	U-Net + THPM	0.6115	0.4453	0.0094	0.0011	5.290	0.000
	U-Net + PHPM	0.6059	0.4407	**0.0091**	0.0012	**4.590**	0.000
	U-Net + SHPM	**0.6316**	**0.4677**	0.0093	**0.0011**	4.740	0.000

4.3 Quantitative and Qualitative Results

We compare the segmentation results of U-Net models with different constraints. CenterlineDice (clDice) [44] is a differentiable skeletonization-based approach. THPM and PHPM refer to constraints with threshold-based and persistent homology-based hard pixel mining, respectively. Our proposed skeleton-based method is denoted by SHPM.

Table 2: Model training time for 100 iterations (s).

	2D Patch (256 × 256)	3D Patch (128 × 128 × 72)
U-Net	1.61	24.21
U-Net + clDice	2.22	27.40
U-Net + THPM	1.65	24.33
U-Net + PHPM	11.83	718.59
U-Net + SHPM	1.78	26.73

Table. 1 shows the quantitative results. Our method demonstrates high superiority in topology-aware metrics, including two clustering-based metrics and two topology-based metrics. The method most closely resembling ours in terms of preserving topology is THPM, but it leads to noticeable decreases in volumetric metrics. On several datasets, clDice outperforms ours in volumetric metrics, but its ability to preserve topology is limited. The improvement of PHPM in topology-preservation is not particularly satisfactory, possibly due to sparse and inaccurate hard set.

Fig. 5 shows qualitative results of 2D datasets. The masks predicted by ours have more

consistent topology. For 3D images, our method generates more complete results comparing to other methods as shown in Fig. 6. We also compare the training time as shown in Table. 2. Training with SHPM is 6.6× and 26.9× faster than PHPM for 2D and 3D image patches. Compared to training of original U-Net and with THPM, our training speed only decreases slightly, which implies that our approach is also applicable to larger-scale training.

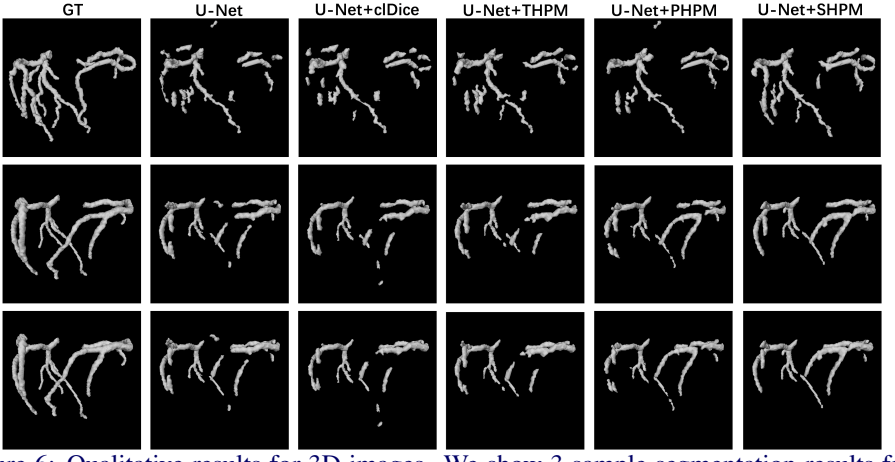

Figure 6: Qualitative results for 3D images. We show 3 sample segmentation results from ICAS-d datasets. Columns from left to right are the visualizations of reconstructed meshes from ground truth and predictions of different approaches. The predictions by ours (U-Net+SHPM) have fewer mis-connections and demonstrate a high level of integrity.

5 Conclusion, Limitation & Future Work

We propose a novel loss function based on hard pixel mining to improve the topology preservation performance of neural networks. We introduce and compare Threshold-based HPM (THPM), Persistent homology-based HPM (PHPM) and Skeleton-based HPM (SHPM), demonstrating the superiority of SHPM in mining topology-relevant hard pixels. We apply SHPM into the training of neural networks and achieve superior quality in topology-aware metrics without a significant compromise on pixel-wise accuracy. Our method is computationally efficient and can be seamlessly plugged into various learning-based segmentation frameworks.

While the experiments have successfully demonstrated the topology-preserving capability of our model, our current implementation is quite simple and can be further improved. For example, the binary thinning algorithm is implemented on CPU. A GPU-accelerated skeletonization can further improve the training efficiency, especially for large batch size. When the tubular structures are extremely thin, such as 1-pixel structures, the skeleton of prediction can hardly be covered by the ground-truth mask, which makes it difficult to mine accurate topology-aware hard pixels. In another case, when two patches with different labels have very similar appearances, our methods might yield the same predictions. The enhancement of model architecture such as introducing local attention mechanisms and hierarchical structures can be considered in future works to address these issues.

Acknowledgement: This work is supported in part by the Natural Science Foundation of China (Grant 62001266) and the Tsinghua SIGS Scientific Research Start-up Fund (Grant QD2021012C).

References

[1] Madjid Allili, Konstantin Mischaikow, and Allen Tannenbaum. Cubical homology and the topological classification of 2d and 3d imagery. In *Proceedings 2001 international conference on image processing (Cat. No. 01CH37205)*, volume 2, pages 173–176. IEEE, 2001.

[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] Ricardo J Araújo, Jaime S Cardoso, and Hélder P Oliveira. A deep learning design for improving topology coherence in blood vessel segmentation. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part I 22*, pages 93–101. Springer, 2019.

[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] 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.

[6] Davide Belli and Thomas Kipf. Image-conditioned graph generation for road network extraction. *arXiv preprint arXiv:1910.14388*, 2019.

[7] Francesco Caliva, Claudia Iriondo, Alejandro Morales Martinez, Sharmila Majumdar, and Valentina Pedoia. Distance map loss penalty term for semantic segmentation. *arXiv preprint arXiv:1908.03679*, 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, 2017.

[10] Hyeong In Choi, Sung Woo Choi, and Hwan Pyo Moon. Mathematical theory of medial axis transform. *pacific journal of mathematics*, 181(1):57–88, 1997.

[11] Özgün Çiçek, Ahmed Abdulkadir, Soeren S Lienkamp, Thomas Brox, and Olaf Ronneberger. 3d u-net: learning dense volumetric segmentation from sparse annotation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part II 19*, pages 424–432. Springer, 2016.

[12] James R Clough, Nicholas Byrne, Ilkay Oksuz, Veronika A Zimmer, Julia A Schnabel, and Andrew P King. A topological loss function for deep-learning based image segmentation using persistent homology. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(12):8766–8778, 2020.

[13] Ruining Deng, Quan Liu, Can Cui, Tianyuan Yao, Jun Long, Zuhayr Asad, R Michael Womick, Zheyu Zhu, Agnes B Fogo, Shilin Zhao, et al. Omni-seg: A scale-aware dynamic network for renal pathological image segmentation. *IEEE Transactions on Biomedical Engineering*, 2023.

[14] Tamal Krishna Dey and Yusu Wang. *Computational topology for data analysis*. Cambridge University Press, 2022.

[15] Herbert Edelsbrunner and John L Harer. *Computational topology: an introduction*. American Mathematical Society, 2022.

[16] 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.

[17] Yanbo Fan, Siwei Lyu, Yiming Ying, and Baogang Hu. Learning with average top-k loss. *Advances in neural information processing systems*, 30, 2017.

[18] Jan Funke, Fabian Tschopp, William Grisaitis, Arlo Sheridan, Chandan Singh, Stephan Saalfeld, and Srinivas C Turaga. Large scale image segmentation with structured loss based deep learning for connectome reconstruction. *IEEE transactions on pattern analysis and machine intelligence*, 41(7):1669–1680, 2018.

[19] 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.

[20] Xiaoling Hu, Fuxin Li, Dimitris Samaras, and Chao Chen. Topology-preserving deep image segmentation. *Advances in neural information processing systems*, 32, 2019.

[21] Xiaoling Hu, Yusu Wang, Li Fuxin, Dimitris Samaras, and Chao Chen. Topology-aware segmentation using discrete morse theory. *arXiv preprint arXiv:2103.09992*, 2021.

[22] Jesse M Hunter, Jason Kwan, Michael Malek-Ahmadi, Chera L Maarouf, Tyler A Kokjohn, Christine Belden, Marwan N Sabbagh, Thomas G Beach, and Alex E Roher. Morphological and pathological evolution of the brain microcirculation in aging and alzheimer's disease. *PloS one*, 7(5):e36893, 2012.

[23] Anne Joutel, Marie Monet-Leprêtre, Claudia Gosele, Céline Baron-Menguy, Annette Hammes, Sabine Schmidt, Barbara Lemaire-Carrette, Valérie Domenga, Andreas Schedl, Pierre Lacombe, et al. Cerebrovascular dysfunction and microcirculation rarefaction precede white matter lesions in a mouse genetic model of cerebral ischemic small vessel disease. *The Journal of clinical investigation*, 120(2):433–445, 2010.

[24] 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*, pages 285–296. PMLR, 2019.

[25] Bin Kong, Xin Wang, Junjie Bai, Yi Lu, Feng Gao, Kunlin Cao, Jun Xia, Qi Song, and Youbing Yin. Learning tree-structured representation for 3d coronary artery segmentation. *Computerized Medical Imaging and Graphics*, 80:101688, 2020.

[26] T. Yung Kong. On topology preservation in 2-d and 3-d thinning. *International journal of pattern recognition and artificial intelligence*, 9(05):813–844, 1995.

[27] Ta-Chih Lee, Rangasami L Kashyap, and Chong-Nam Chu. Building skeleton models via 3-d medial surface axis thinning algorithms. *CVGIP: graphical models and image processing*, 56(6):462–478, 1994.

[28] Yang Li, Yue Zhang, Jing-Yu Liu, Kang Wang, Kai Zhang, Gen-Sheng Zhang, Xiao-Feng Liao, and Guang Yang. Global transformer and dual local attention network via deep-shallow hierarchical feature fusion for retinal vessel segmentation. *IEEE Transactions on Cybernetics*, 2022.

[29] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision*, pages 2980–2988, 2017.

[30] 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.

[31] Cherng Min Ma. On topology preservation in 3d thinning. *CVGIP: Image understanding*, 59(3):328–339, 1994.

[32] Clément Maria, Jean-Daniel Boissonnat, Marc Glisse, and Mariette Yvinec. The gudhi library: Simplicial complexes and persistent homology. In *Mathematical Software–ICMS 2014: 4th International Congress, Seoul, South Korea, August 5-9, 2014. Proceedings 4*, pages 167–174. Springer, 2014.

[33] Marina Meilă. Comparing clusterings: an axiomatic view. In *Proceedings of the 22nd international conference on Machine learning*, pages 577–584, 2005.

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

[35] 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.

[36] 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.

[37] 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.

[38] Martin Ralf Oswald, Jan Stühmer, and Daniel Cremers. Generalized connectivity constraints for spatio-temporal 3d reconstruction. In *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part IV 13*, pages 32–46. Springer, 2014.

[39] Markus Rempfler, Jan-Hendrik Lange, Florian Jug, Corinna Blasse, Eugene W Myers, Bjoern H Menze, and Bjoern Andres. Efficient algorithms for moral lineage tracing. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 4695–4704, 2017.

[40] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, pages 234–241. Springer, 2015.

[41] Samuel Rota Bulo, Gerhard Neuhold, and Peter Kontschieder. Loss max-pooling for semantic image segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017.

[42] Dominik Schultes. *Route planning in road networks*. PhD thesis, Karlsruhe, Univ., Diss., 2008, 2008.

[43] Florent Ségonne. Active contours under topology control—genus preserving level sets. *International Journal of Computer Vision*, 79(2), 2008.

[44] Suprosanna Shit, Johannes C Paetzold, Anjany Sekuboyina, Ivan Ezhov, Alexander Unger, Andrey Zhylka, Josien PW Pluim, Ulrich Bauer, and Bjoern H Menze. cldice-a novel topology-preserving loss function for tubular structure segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16560–16569, 2021.

[45] Abhinav Shrivastava, Abhinav Gupta, and Ross Girshick. Training region-based object detectors with online hard example mining. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 761–769, 2016.

[46] Nahian Siddique, Sidike Paheding, Colin P Elkin, and Vijay Devabhaktuni. U-net and its variants for medical image segmentation: A review of theory and applications. *Ieee Access*, 9:82031–82057, 2021.

[47] 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.

[48] Kah-Kay Sung. Learning and example selection for object and pattern detection. 1996.

[49] Giles Tetteh, Velizar Efremov, Nils D Forkert, Matthias Schneider, Jan Kirschke, Bruno Weber, Claus Zimmer, Marie Piraud, and Bjoern H Menze. Deepvesselnet: Vessel segmentation, centerline prediction, and bifurcation detection in 3-d angiographic volumes. *Frontiers in Neuroscience*, page 1285, 2020.

[50] Engin Türetken, Fethallah Benmansour, Bjoern Andres, Przemysław Głowacki, Hanspeter Pfister, and Pascal Fua. Reconstructing curvilinear networks using path classifiers and integer programming. *IEEE transactions on pattern analysis and machine intelligence*, 38(12):2515–2530, 2016.

[51] Subeesh Vasu, Mateusz Kozinski, Leonardo Citraro, and Pascal Fua. Topoal: An adversarial learning approach for topology-aware road segmentation. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXVII 16*, pages 224–240. Springer, 2020.

[52] 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.

[53] Paul Viola and Michael Jones. Rapid object detection using a boosted cascade of simple features. In *Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001*, volume 1, pages I–I. Ieee, 2001.

[54] Hubert Wagner, Chao Chen, and Erald Vuçini. Efficient computation of persistent homology for cubical data. In *Topological methods in data analysis and visualization II: theory, algorithms, and applications*, pages 91–106. Springer, 2011.

[55] Jan D Wegner, Javier A Montoya-Zegarra, and Konrad Schindler. A higher-order crf model for road network extraction. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1698–1705, 2013.

[56] Jiaqi Yang, Xiaoling Hu, Chao Chen, and Chialing Tsai. 3d topology-preserving segmentation with compound multi-slice representation. In *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*, pages 1297–1301. IEEE, 2021.

[57] Su Yang, Jihoon Kweon, and Young-Hak Kim. Major vessel segmentation on x-ray coronary angiography using deep networks with a novel penalty loss function. 2019.

[58] An Zeng, Chunbiao Wu, Meiping Huang, Jian Zhuang, Shanshan Bi, Dan Pan, Najeeb Ullah, Kaleem Nawaz Khan, Tianchen Wang, Yiyu Shi, et al. Imagecas: A large-scale dataset and benchmark for coronary artery segmentation based on computed tomography angiography images. *arXiv preprint arXiv:2211.01607*, 2022.

[59] Tongjie Y Zhang and Ching Y. Suen. A fast parallel algorithm for thinning digital patterns. *Communications of the ACM*, 27(3):236–239, 1984.