

3D TOPOLOGY-PRESERVING SEGMENTATION WITH COMPOUND MULTI-SLICE REPRESENTATION

Jiaqi Yang^{1,*}

Xiaoling Hu^{2,*}

Chao Chen²

Chialing Tsai¹

¹ Graduate Center, CUNY

² Stony Brook University

ABSTRACT

We propose a new topology-preserving method for 3D image segmentation. We treat the image as a stack of 2D images so that the topological computation can be carried only within 2D in order to achieve computational efficiency. To enforce the continuity between slices, we propose a compound multi-slice representation and a compound multi-slice topological loss that incorporates rich topological information from adjacent slices. The quantitative and qualitative results show that our proposed method outperforms various strong baselines, especially for structure-related evaluation metrics.

Index Terms— Topology, 3D Image Segmentation, Compound Multi-Slice

1. INTRODUCTION

In biomedical image analysis, accurate segmentation of biological entities or anatomical structures remains a very crucial step for a wide range of clinical applications, such as computer-aided diagnosis [1], screening of medical conditions [2] and visual augmentation of medical devices [3]. In recent years, deep-learning-based methods have achieved close-to-human-level performance [4, 5]. However, existing algorithms, mostly focusing on per-pixel accuracy, are prone to structural errors, e.g., missing connected components and broken connections. Such structural errors can be fatal for downstream analysis of fine-scale structures such as neuron membranes or processes, vessels, cells, etc. The topology of these structures carries significant semantic / functional information. For example, correct delineation of thin objects, such as neuron membranes and processes, is vital in providing accurate morphological and structural quantification of the underlying system; small pixel-wise errors can lead to broken membranes, merging different neurons into one connected region. See Fig. 1 for an illustration.

Recently, a novel method has been proposed to train deep nets to segment objects with correct topology [6]. The idea is to introduce a topological loss that complements existing cross-entropy loss and enforces the network to learn correct topology. However, the method is restricted to 2D images.



Fig. 1. An illustration of structural accuracy. From left to right: a sample patch, the ground truth, results of UNet, TopoLoss, and our compound multi-slice Topo-UNet.

An extension to 3D images, although feasible in theory, can be computationally prohibitive.

Generating 3D results by stacking 2D predictions is a natural idea. The major shortcoming of such approach is that slices are treated independently, whereas their spatial continuity could have been leveraged. When segmenting one slice, observations from adjacent slices should be used to clarify challenging cases as consecutive slices likely have similar segmentations.

To fully leverage the continuity information, we propose to learn a *compound multi-slice representation*. For each slice, we learn a feature representation using not only itself, but also its adjacent slices. The feature representation is a compound of multiple slices along Z-dimension; it has compound channels, representing the current slice (s_i), itself plus 2 immediately adjacent slices ($[s_{i-1}, s_i, s_{i+1}]$), and itself plus 4 nearby slices ($[s_{i-2}, \dots, s_{i+2}]$). The compound multi-slice representation carries contextual information from neighboring slices. It is more robust to noise and leads to more accurate segmentation for the current slice.

We cannot simply use the topology of adjacent slices as additional supervision, as they may be incorrect for the current slice. Therefore, we propose a novel *compound multi-slice topological loss* that leverages the multi-slice representation and enforces better topological accuracy of the current slice. The core idea is that the critical locations for topological accuracy can be shared among adjacent slices.

In summary, we propose a compound multi-slice representation that incorporate information from adjacent slices, and a compound multi-slice topological loss for training the network to incorporate topological information. Our proposed topology-preserving approach achieves better performance in both per-pixel accuracy and topological accuracy. See Fig. 1 for an illustration.

* indicates equal contributions.

2. RELATED WORKS

In pre-deep learning era, heuristic methods such as random forest [7] are always used to obtain probability maps. With the advent of deep learning, convolutional neural networks (DCNN) [8, 9] have been used in more and more scenarios for medical image segmentation.

In terms of 3D medical image data, we usually segment the images slice by slice in a 2D way, and then linking methods are adopted to generate 3D neuronal objects based on previous 2D segmentation results [10]. Researchers have already developed efficient deep learning methods for segmenting 2D EM images [11]. Good 3D segmentation performance depends on both robust 2D segmentation for slices and reasonable linking methods to aggregate 2D segmentation results. But the problem is that the probability maps of different slices are generated independently, and the contextual information is ignored. As a result, errors are introduced due to inconsistency in the linking stage.

Some other researchers have tried to solve 3D medical segmentation in a direct 3D way [12, 13, 14]. Though these approaches could take advantage of contextual information between different slices, they demand much higher computation cost, which can be unacceptable in practice. In addition, 3D networks are lacking the resources of pre-trained models using 2D natural images that many 2D methods adopt. As a result, it takes much longer training time and can possibly suffer from unstable convergence [15].

However almost all these methods focus on pixel accuracy and treat every pixel equally. Beside pixel accuracy, structure should also be critical for medical scenarios. To preserve structural accuracy, topological loss functions based on persistent homology was proposed for various tasks such as image segmentation [6, 16], crowd localization [17] and classifier regularization [18]. Also, another topological loss function based on discrete Morse theory has been proposed for image segmentation [19]. We note in the EM image context, one may directly learn to better partition the image rather than segmenting membranes [14]. But this method cannot handle other types of topology as our method can.

3. METHOD

Our method treats a 3D image as 2D slices stacked along the Z-dimension. It segments these 2D slices one by one. When segmenting each slice, our method leverages adjacent slices' information using a compound representation. Topological information of adjacent slices are also used through a compound topological loss, so that the trained network preserves the correct topology. Figure 2 illustrates the whole pipeline of our method.

Figure 2(a) illustrates how the compound multi-slice representation is learnt. We learn feature representation of the same slice with compound networks, namely, using different

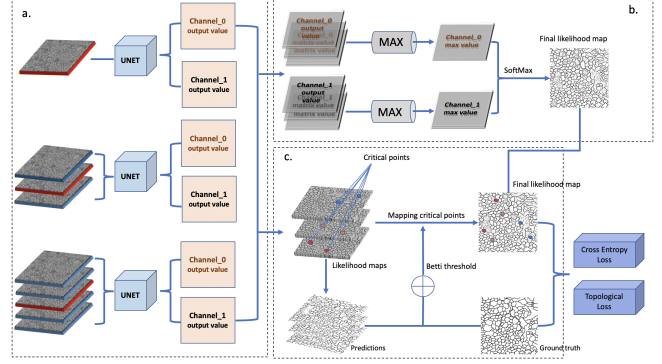


Fig. 2. The segmentation pipeline of the proposed method, and the architecture contains three modules: (a) compound multi-slice representation module; (b) computation process of final feature map from compound models; (c) computation of compound multi-slice topological loss.

collections of nearby slices as input for each. Assume the slice of interest is the i -th one, denoted as s_i . We use three different multi-slice representations (hence the word “compound”), learnt with the slice itself ($[s_i]$), the slice and its two adjacent ones ($[s_{i-1}, s_i, s_{i+1}]$), the slice and its 4 neighboring slices ($[s_{i-2}, \dots, s_{i+2}]$). Take the 3 slices representation for example. We use a UNet-based sub-network, which takes the three slices (s_{i-1} , s_i and s_{i+1}) as input and predicts the likelihood maps of foreground and background for slice s_i . These likelihood maps are considered the representation of the 3-slices (3s) sub-network. Similarly, we can have representations of 1s sub-network and 5s sub-network. For convenience, we call these sub-networks UNet[1s], UNet[3s] and UNet[5s], respectively. These sub-networks are pre-trained as standard segmentation UNets, except that they take 1, 3 or 5 input channels. During training, these sub-networks are combined with the prediction and the topological loss evaluation and are trained end-to-end.

3.1. Compound Multi-Slice Representations

These representations of 1s, 3s and 5s are used to predict a final likelihood map. We use a max-projection over the foreground likelihood maps of UNet[1s], UNet[3s] and UNet[5s] to generate the foreground likelihood map. Similarly, we use a max-projection over the background likelihood maps of the three UNets to generate the background likelihood map. After a re-normalization via softmax, we obtain the final likelihood map. This is illustrated in Figure 2(b). The generated likelihood map is used for final prediction during testing. During training, it will be compared with ground truth labeling through cross-entropy loss. Here we use max-project over both the foreground likelihood maps and the background likelihood maps in order to avoid false negative detection in both foreground and background. This way, we mitigate the error

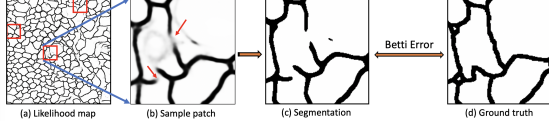


Fig. 3. Illustration of the topological loss in [6].

rate in prediction by aggregating over the compound of multi-slice models (1s, 3s and 5s). The network avoids being either over-confident or under-confident.

The final likelihood map prediction, denoted as $\hat{y}_i$, will also be used to evaluate the topological loss, as will be explained below.

3.2. Compound Multi-Slice Topological Loss

It remains to introduce the topological loss, which enforces the final prediction $\hat{y}_i$ to have the correct topology. The key difference between our new topological loss and the one by Hu et al. [6] is the usage of the compound of multi-slice. Our compound multi-slice topological loss incorporates topological information from nearby slices to better segment the current slice.

In [6], the topological loss compares a predicted likelihood map and the ground truth segmentation in terms of their topology. Essentially, the loss identifies critical points of the likelihood map, e.g., saddles and extrema, which correspond to topological error. See Figure 3 for an illustration. The algorithm first samples patches over the image domain (Figure 3(a)). For the sample patch in (Figure 3(b)), the topological loss identifies saddle points (highlighted with red arrows) corresponding to broken handles in the prediction (Figure 3(c)). These critical points are used to define the topological loss (for the i -th slice) as $L_{topo}(\hat{y}_i, y_i) = \sum_{c \in \mathcal{C}(\hat{y}_i)} (\hat{y}_i(c) - y_i(c))^2$, in which $\mathcal{C}(\hat{y}_i)$ is the set of identified critical points (corresponding to critical pixels) of the likelihood map $\hat{y}_i$. y_i is the ground truth mask. In other words, the topological loss identifies these topology-error-relevant critical points that correspond to difficult locations (e.g., blurred boundary region). Then it pushes the prediction at these locations closer to ground truth.

To incorporate topological information from nearby slices, we need to use the compound representations. We observe in the original topological loss, and noticed that the key is to identify critical points and to force the neural nets to be correct at these critical locations. This inspires us to identify multiple sets of critical points, each from a different multi-slice model. These critical points are complementary to each other and should all be taken into consideration for computing the topological loss. In particular, recall the representations are likelihood maps predicted by different UNets (UNet[1s], UNet[3s] and UNet[5s]). We can compute the critical point sets of these likelihood maps corresponding to incorrect topology. Denote these critical point sets as $\mathcal{C}_{1s}$, $\mathcal{C}_{3s}$, $\mathcal{C}_{5s}$. We can then rewrite the compound multi-slice

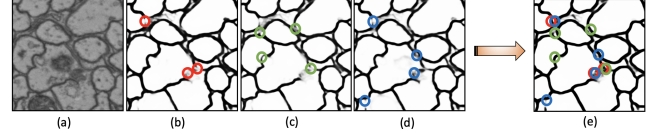


Fig. 4. The motivation of combining critical points from multiple predicted likelihood maps. (a) the original image, (b), (c) and (d) the middle three images are likelihood maps predicted by different sub networks and their corresponding critical points identified by TopoLoss, (e) the final likelihood map with combining critical points. Looking at the original image, actually all these critical points are *difficult locations* for prediction, which means that these critical points are complementary to each other and can be combined to achieve better results.

topological loss as

$$L_{mr-topo}(\hat{y}_i, y_i) = \sum_{c \in \mathcal{C}_{1s} \cup \mathcal{C}_{3s} \cup \mathcal{C}_{5s}} (\hat{y}_i(c) - y_i(c))^2.$$

Note we are still forcing the final prediction $\hat{y}_i$ to be correct at critical points. The only difference is that the critical points are identified by inspecting predictions of different models. See Figure 4 for an example. The critical points identified by representations of different models ((b)-(d)) all correspond to challenging locations. Using all of them in the final prediction will efficiently train the neural network to achieve high topological accuracy.

Note that we also need to sample patches over the whole image. To further improve efficiency, we only select patches on which topological error happens. In particular, for each patch we calculate the *Betti error* by comparing the Betti numbers of the prediction and the ground truth mask (Figure 3 (c)(d)). Betti number counts the number of topological structures, e.g., connected components or handles. The incentive is that computing and comparing Betti numbers is much cheaper than finding critical points corresponding to incorrect topology (which requires calling the algorithm for persistent homology [20]). Figure 2(c) illustrates how the topological loss is computed. Finally, our overall loss is $L = L_{CE} + \lambda L_{mr-topo}$, where λ is the weight of topological loss.

4. EXPERIMENTS

Datasets. We demonstrate the effectiveness of our proposed method with two different 3D Electron Microscopic Images datasets: **ISBI13** [21] and **CREMI**¹. The size of **ISBI13** is 100x1024x1024 and the size of **CREMI** is 125x1250x1250. The goal of this segmentation task is to recognize the membranes and then partition the original image into different regions/neurons. For all the experiments, we use 80% of the training samples as training and the remaining 20% as validation, and report the performance over the validation set.

¹<https://cremi.org/>

Evaluation metrics. In this paper we adopt similar evaluation metrics as of [6], which are pixel accuracy and Betti number error. Details of the evaluation metrics can be found in the Sec.3 of [6]. Generally speaking, pixel accuracy is the most traditional metric for segmentation task and could roughly evaluate the performance of different segmentation models. Betti number error is directly related to image structure, which is more applicable to biomedical domain tasks.

4.1. Implementation details

We adopt a classical UNet with depth five as our backbone. Also, we use data augmentation techniques, random noise and brightness changes, to improve the performance.

Baselines. To demonstrate the effectiveness of compound multi-slice topological loss, we compare the results of the proposed method with five baselines: 1) Classical UNet with only one slice as input (traditional 2D segmentation, UNet[1s]); 2) UNet with 3 sequential slices as input (UNet[3s]); 3) UNet with 5 sequential slices as input (UNet[5s]); 4) UNet with compound multi-slice representation (UNet[1s3s5s]); 5) TopoLoss [6]. Because of the computational cost and fairness for comparison, we re-implement the TopoLoss with our UNet backbone instead of the original version.

Train setting. For UNet training parameters, we set 0.0001 as the learning rate. The number of training epochs are 30 and 100 for CREMI and ISBI13 respectively (without topological training), followed by another 3 and 10 epochs to train models with topological loss. So our training contains two parts: for the first part we only use cross entropy as loss function; for the second part, weighted topological loss is incorporated into the networks to fine-tune the pre-trained model.

Topo-incorporating details. As described above, topological training is added to the last “UNet with multi deep models” setting to improve the structure accuracy of the segmentation. Specifically, the patch size is also $65 * 65$ for topological loss function, and the balanced weight for topological loss is $\lambda = 0.01$. To reduce the computation cost, we use pre-trained model before incorporating the topological loss. To be noted that for the training of CREMI dataset, the topological loss is fed into training after 30 epochs. The total loss is somehow stable before over-training (34 epochs in this case). Before over-training, the topological loss decreases while the cross entropy loss is oscillating. This reason is that topological loss forces the network to identify critical pixels and fix them and may fool the cross entropy loss.

4.2. Result Analysis

We compare the proposed method (Topo-UNet) with five different networks configurations using two different evaluation metrics mentioned above. The quantitative results are illustrated in Table. 1. The best results are highlighted, most of which generated by our proposed method. For accuracy

Table 1. Experiment results for different models on two popular 3D medical datasets

Dataset	MODELS	Accuracy	Betti Error
CREMI	UNet[1s]	0.9620	2.59
	UNet[3s]	0.9601	1.98
	UNet[5s]	0.9690	1.85
	UNet[1s3s5s]	0.9700	1.67
	TopoLoss [6]	0.9664	1.90
	Topo-UNet[1s3s5s]	0.9713	1.32
ISBI13	UNet[1s]	0.9123	2.54
	UNet[3s]	0.9120	1.88
	UNet[5s]	0.9099	1.41
	UNet[1s3s5s]	0.9168	1.56
	TopoLoss	0.9138	1.67
	Topo-UNet[1s3s5s]	0.9156	1.25

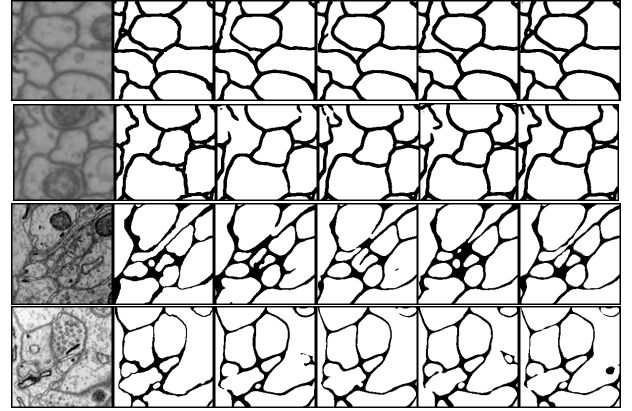


Fig. 5. Qualitative results of the proposed method compared to other models. From left to right, sample images, ground truth, results for UNet[5s], UNet[1s3s5s], TopoLoss and our proposed Topo-UNet[1s3s5s].

of ISBI13 dataset, our method is second only to our compound multi-slice network without TopoLoss integrated. It can also be observed that results of compound multi-slice (UNet[1s3s5s]) are in general better than the results of single-slice (UNet[1s]) or multi-slices (UNet[3s] and UNet[5s]) configurations.

Fig. 5 shows qualitative results. Results in column 4 (UNet[1s3s5s]) are less fragmented, comparing to one multi-slice network (column 3). The last column corresponds to the results of our proposed system, Topo-UNet, which demonstrates better connectivity in terms of structures and topology, comparing to other network configurations.

5. CONCLUSION

In this paper, a novel method is proposed to extend the topology-preserving training to 3D EM images without much additional computational cost. When segmenting each slice, our proposed method leverages adjacent slices’ information using a compound multi-slice representation. Topological information of adjacent slices are also used through a compound multi-slice topological loss to segment with correct topology.

6. COMPLIANCE WITH ETHICAL STANDARDS

The authors of this work have undergone CITI training in human subjects and medical ethics. All datasets are publicly available. This research study does not require the institutional IRB.

7. ACKNOWLEDGMENTS

This work was partially supported by grants NSF IIS-1909038, CCF-1855760, and NCI 1R01CA253368-01.

8. REFERENCES

- [1] A. Esteva, B. Kuprel, and R. et al. Novoa, “Dermatologist-level classification of skin cancer with deep neural networks,” *Nature*, 2017.
- [2] R. Poplin, A.V. Varadarajan, and K. et al. Blumer, “Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning,” *Nature Biomedical Engineering*, 2018.
- [3] P.C. Chen, K. Gadepalli, and R. et al. MacDonald, “An augmented reality microscope with real-time artificial intelligence integration for cancer diagnosis,” *Nature Medicine*, 2019.
- [4] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick, “Mask r-cnn,” in *ICCV*, 2017.
- [5] 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,” *TPAMI*, 2018.
- [6] Xiaoling Hu, Fuxin Li, Dimitris Samaras, and Chao Chen, “Topology-preserving deep image segmentation,” in *NeurIPS*, 2019.
- [7] Verena Kaynig, Amelio Vazquez-Reina, Seymour Knowles-Barley, Mike Roberts, Thouis R Jones, Narayanan Kasthuri, Eric Miller, Jeff Lichtman, and Hanspeter Pfister, “Large-scale automatic reconstruction of neuronal processes from electron microscopy images,” *Medical image analysis*, 2015.
- [8] Manuel Berning, Kevin M Boergens, and Moritz Helmstaedter, “Segem: efficient image analysis for high-resolution connectomics,” *Neuron*, 2015.
- [9] Dan Ciresan, Alessandro Giusti, Luca M Gambardella, and Jürgen Schmidhuber, “Deep neural networks segment neuronal membranes in electron microscopy images,” in *NeurIPS*, 2012.
- [10] Juan Nunez-Iglesias, Toufiq Parag Ryan Kennedy, Jianbo Shi, and Dmitri B Chklovskii, “Machine learning of hierarchical clustering to segment 2d and 3d images,” *PloS one*, 2013.
- [11] Ahmed Fakhry, Hanchuan Peng, and Shuiwang Ji, “Deep models for brain em image segmentation: novel insights and improved performance,” *Bioinformatics*, 2016.
- [12] Özgün Çiçek, Ahmed Abdulkadir, Soeren S Lienkamp, Thomas Brox, and Olaf Ronneberger, “3d u-net: learning dense volumetric segmentation from sparse annotation,” in *MICCAI*. Springer, 2016.
- [13] Yaron Meirovitch, Lu Mi, Hayk Saribekyan, Alexander Matveev, David Rolnick, and Nir Shavit, “Cross-classification clustering: An efficient multi-object tracking technique for 3-d instance segmentation in connectomics,” in *CVPR*, 2019.
- [14] 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.
- [15] Nima Tajbakhsh, Jae Y Shin, Suryakanth R Gurudu, R Todd Hurst, Christopher B Kendall, Michael B Gotway, and Jianming Liang, “Convolutional neural networks for medical image analysis: Full training or fine tuning?,” *TMI*, 2016.
- [16] James R Clough, Ilkay Oksuz, Nicholas Byrne, Julia A Schnabel, and Andrew P King, “Explicit topological priors for deep-learning based image segmentation using persistent homology,” in *IPMI*. Springer, 2019.
- [17] Shahira Abousamra, Minh Hoai, Dimitris Samaras, and Chao Chen, “Localization in the crowd with topological constraints,” in *AAAI*, 2021.
- [18] Chao Chen, Xiuyan Ni, Qinxun Bai, and Yusu Wang, “A topological regularizer for classifiers via persistent homology,” in *AISTATS*. PMLR, 2019, pp. 2573–2582.
- [19] Xiaoling Hu, Yusu Wang, Li Fuxin, Dimitris Samaras, and Chao Chen, “Topology-aware segmentation using discrete morse theory,” in *ICLR*, 2021.
- [20] Herbert Edelsbrunner and John Harer, *Computational topology: an introduction*, American Mathematical Soc., 2010.
- [21] I Arganda-Carreras, HS Seung, A Vishwanathan, and D Berger, “3d segmentation of neurites in em images challenge-isbi 2013,” 2013.