A Topological-Attention ConvLSTM Network and Its Application to EM Images

Jiaqi Yang[1⋆], Xiaoling Hu[2⋆], Chao Chen[2], and Chialing Tsai[1]

[1] Graduate Center, CUNY
[2] Stony Brook University

Abstract. Structural accuracy of segmentation is important for fine-scale structures in biomedical images. We propose a novel Topological-Attention ConvLSTM Network (TACLNet) for 3D anisotropic image segmentation with high structural accuracy. We adopt ConvLSTM to leverage contextual information from adjacent slices while achieving high efficiency. We propose a Spatial Topological-Attention (STA) module to effectively transfer topologically critical information across slices. Furthermore, we propose an Iterative Topological-Attention (ITA) module that provides a more stable topologically critical map for segmentation. Quantitative and qualitative results show that our proposed method outperforms various baselines in terms of topology-aware evaluation metrics.

Keywords: Topological-Attention, Spatial, Iterative, ConvLSTM

1 Introduction

Deep learning methods have achieved state-of-the-art performance for image segmentation. However, most existing methods focus on per-pixel accuracy (e.g., minimizing the cross-entropy loss) and are prone to structural errors, e.g., missing connected components and broken connections. These structural errors can be fatal in downstream analysis, affecting the functionality of the extracted fine-scale structures such as neuron membranes, vessels and cells.

To address this issue, differentiable topological losses [12,6,13,27] have been proposed to enforce the network to learn to segment with correct topology. However, these methods have their limitations when applied to 3D images, due to the high computational cost of topological information. Furthermore, we often encounter anisotropic images, i.e., images with low resolution in z-dimension. The topological loss cannot be directly applied to 3D anisotropic images. For example, a tube in 3D may manifest as a series of rings across different slices rather than a seamless tube. Directly enforcing a 3D tube topology cannot work.

In this paper, a novel 3D topology-preserving segmentation method is proposed to address the aforementioned issues. Inspired by existing approaches for anisotropic images [19,28], we propose to first segment individual slices, and then stack the results together as the 3D output. We use *convolutional LSTM*

⋆ The two authors contributed equally to this paper.

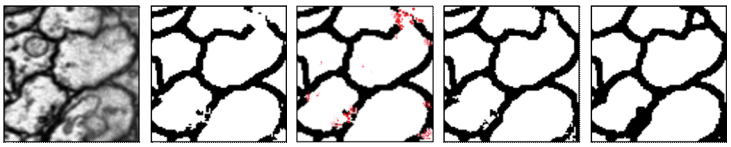

Fig. 1. An illustration of our method. From left to right: original image, ConvLSTM segmentation result, topological attention map (overlaid with segmentation), result of our method, and ground truth.

.

(ConvLSTM) [4] as our backbone. Specifically designed for 3D anisotropic images, ConvLSTM uses 2D convolution and exploits inter-slice correlation to achieve high quality results while being more efficient than 3D CNNs. We incorporate topological loss into each of the 2D slices. This way, the topological computation is restricted within each 2D slice, and thus is very efficient.

However, simply enforcing topological loss at each slice is insufficient. A successful method should account for the fact that the topology of consecutive slices share some similarity, but are not the same. When segmenting one slice, the topology of other slices should help recalibrate the prediction, but in a soft manner. To effectively propagate topological information across slices, we propose a *Spatial Topological-Attention module*, which redirects the convolutional network's attention toward topologically critical locations of each slice, based on the topology of itself and its adjacent slices. These critical locations are locations at which the model is prone to topological mistakes. Redirecting the attention to these locations will enforce the model to make topologically correct predictions. See Fig. 1 for an illustration of our method.

Another challenge is that the topologically critical map can be inconsistent across different slices and unstable through training epochs. During the training process, the predicted probability maps will change slightly, while the corresponding Topological-Attention maps can be quite different, leading to instability of the training process. To this end, we propose an *Iterative Topological-Attention module* that iteratively refines the topologically critical map through epoches.

Our method, called *Topological-Attention ConvLSTM Network (TACLNet)*, fully utilizes topological information from adjacent slices for 3D images without much additional computational cost. Empirically, our method outperforms baselines in terms of topology-aware metrics. In summary, our main contribution is threefold:

1. A novel Spatial Topological-Attention (STA) module to propagate spatial contextual topological information across adjacent slices.
2. An Iterative Topological-Attention (ITA) module to improve the stability of the topologically critical maps, and consequently the quality of the final results.
3. Combining Topological-Attention with ConvLSTM to achieve high performance on 3D image segmentation benchmarks.

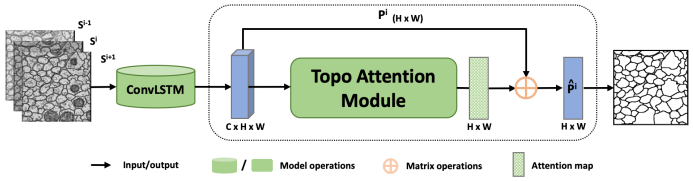

Fig. 2. Overview of the proposed framework

.

2 Related Works

Standard 3D medical image segmentation methods directly apply the networks to 3D images [5,17,16,11]. These methods could be computationally expensive. Alternatively, one may first segment each 2D slice, and then link the 2D segmentation results to generate 3D results [19,27]. Note that this segment-and-link approach ignores the contextual information shared among adjacent slices at the segmentation step. To address this, one may introduce pooling techniques across adjacent slices [9]. But these methods are not explicitly modeling the topology as our method does.

Persistent homology. Our topological approach is based on the theory of persistent homology [7,8], which has attracted a great amount of attention both from theory [3,10] and from applications [25,24]. In image segmentation, persistent-homology-based topological loss functions [12,6] have been proposed to train a neural network to preserve the topology of the segmentation. The key insight of these methods is to identify critical locations for topological correctness, and improve the neural network's prediction at these locations. These critical locations are computed using the theory of persistent homology, and correspond to critical points (local maxima/minima and saddles) of the likelihood function.

Attention mechanism. Attention modules model relationships between pixels/channels/feature maps and have been widely applied in both vision and natural language processing tasks [14,15,21]. Specifically, self-attention mechanism [22] is proposed to draw global dependencies of inputs and has been used in machine translation tasks. [29] tries to learn a better image generator via self-attention mechanism. [23] mainly explores effectiveness of non-local operation, which is similar to self-attention mechanism. [30] learns an attention map to aggregate contextual information for each individual point for scene parsing.

3 Method

The overview of the proposed architecture is illustrated in Fig. 2. To capture the inter-slice information, l consecutive slices along Z-dimension are fed into a ConvLSTM. For ease of exposition, we set $l = 3$ when describing our method. But our method can be easily generalized to arbitrary l. ConvLSTM is an extension

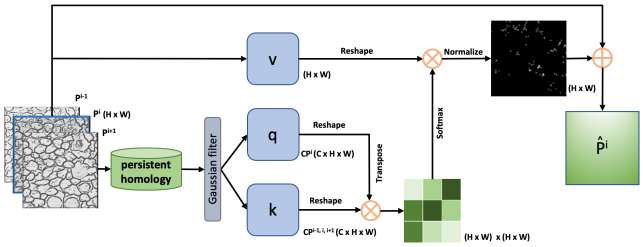

Fig. 3. Illustration of the Spatial Topological-Attention (STA) module

of FC-LSTM [26], which has the convolutional operators in LSTM gates and is particularly efficient in exploiting image sequences. Note that the inputs to ConvLSTM are three adjacent slices, $\{S^{i-1}, S^i, S^{i+1}\} \in R^{H\times W}$, and the output also has three channels, $\{P^{i-1}, P^i, P^{i+1}\} \in R^{H\times W}$, each being the probabilistic map P^i of the corresponding input slice S^i. We use $i-1, i, i+1$ to represent input slice indices in this paper.

The three probabilistic maps $\{P^{i-1}, P^i, P^{i+1}\}$ are then fed into the Topological-Attention module. In this module, each pixel in the feature maps gathers rich structural information from both the current and adjacent slices, without introducing extra parameters. We propose a Spatial Topological-Attention module to model the correlation between the topologically critical information of adjacent slices. A Topological-Attention map is generated to highlight the locations which are structurally critical. See Sec. 3.1 for details. In Sec. 3.2, we introduce the Iterative Topological-Attention module to stabilize the critical map.

3.1 Spatial Topological-Attention (STA) Module

Continuation in contextual information across slices is essential for 3D image understanding, which can be obtained by taking adjacent slices into consideration. In order to collect contextual information in the Z-dimension to enhance the prediction quality, we introduce a STA module which encodes the inter-slices contextual information into the focused slice.

As illustrated in Fig. 2, we can obtain three predicted probabilistic maps, $\{P^{i-1}, P^i, P^{i+1}\} \in R^{H\times W}$ which corresponds to the input slices after ConvLSTM. Fig. 3 shows the complete process that passes the probabilistic maps to the STA module, and yields the final probabilistic map $\hat{P}^i$ in the end. Next, we elaborate the process of aggregating the topological context of adjacent slices.

Persistent homology and critical points. Given a 2D image likelihood map, we obtain the binary segmentation by thresholding at $\alpha = 0.5$. The 2D likelihood map can be represented as a 2D continuous-valued function f. We consider thresholding the continuous function f with all possible thresholds. Denote by Ω the image domain. For a specific threshold α, we define the thresholded results $f^\alpha := \{x \in \Omega | f(x) \geq \alpha\}$. By decreasing α, we obtain a monotonically growing sequence $\varnothing \subseteq f^{\alpha_1} \subseteq f^{\alpha_2} \subseteq ... \subseteq f^{\alpha_n} = \Omega$, where $\alpha_1 \geq \alpha_2 \geq ... \geq \alpha_n$. As α

changes, the topology of f^α changes. New topological structures are born while existing ones are killed. The theory of persistent homology captures all the birth time and death time of these topological structures and summarize them as a *persistence diagram*. One can define a topological loss as the matching distance between the persistence diagrams of the likelihood function and the ground truth. When the loss is minimized, the two diagrams are the same and the likelihood map will generate a segmentation with the correct topology.

As shown in [12], the topological loss can be written as a polynomial function of the likelihood function at different critical pixels. These critical pixels correspond to critical points of the likelihood function (e.g., saddles, local minima and local maxima), and these critical points are crucial locations at which the current model is prone to make topological mistakes. The loss essentially forces the network to improve its prediction at these *topologically critical* locations.

Aggregating topologically critical maps via topological attention. The likelihood maps at different slices generate different critical point maps. Here we propose an attention mechanism to aggregate these critical point maps across different slices to generate topological attention map for the current slice (third column in Fig. 1). For $\{P^{i-1}, P^i, P^{i+1}\}$, using persistent homology algorithm, we identify the critical points. Here we use a Gaussian operation to expand the isolated critical points to blobs because the surrounding regions will likely also be vital for structures. This way we obtain a soft version of critical point map: $CP^i = Gaussian(PH(P^i))$. Here, $PH(\cdot)$ is the operation to generate isolated critical points and $Gaussian(\cdot)$ is a Gaussian operation. CP^i has the same dimension as $P^i \in R^{H\times W}$.

As shown in Fig. 3, these critical maps are used as the query and the key for the attention mechanism. To improve the computational efficiency, we combine the critical maps CP^{i-1}, CP^i, CP^{i+1} into one single $k \in R^{C\times H\times W}$ $(C = 3)$ and expand the CP^i into same size as $q \in R^{C\times H\times W}$. To obtain the correlation between target map (q) and consecutive slices (k), we reshape them to $R^{C\times N}$, where $N = H \times W$, and perform a matrix multiplication between the transpose of q and k. The similarity map $SM \in R^{N\times N}$ is generated after a softmax. SM_{nm} measures the correlation between two pixels, m and n.

Next, we reshape probabilistic map P^i and perform a matrix multiplication between P^i and SM. For pixel n, we obtain normalized $o_n^i = \sum_{m=1}^{N}({P^i}_m SM_{nm})$ and perform an element-wise sum operation with the probabilistic map P^i to get the final output:

$$\hat{P}_n^i = \alpha o_n^i + {P_n}^i \tag{1}$$

where α is initialized as 0 to capture stable probabilistic maps first. As training continues, we assign more weight on attention map so that the $\hat{P}^i$ at each position is a weighted sum across all positions and original probabilistic map P^i.

3.2 Iterative Topological-Attention (ITA) Module

As mentioned above, the critical points generated by persistent homology are sensitive, and consequently the attention map is also relatively unstable.

By exploiting the correlation of probabilistic maps between different epochs, we can further improve the robustness of the obtained attention maps, which can lead to a better representation of the predicted probabilistic maps. Therefore, we introduce an ITA module to explore the relationships between the attention maps of different epochs. The iterative method not only helps with stability, but also enforces faster convergence. Formally, the ITA is calculated as $o_t = \beta o_{t-1} + (1-\beta)o_t$. Here, β is a parameter to deal with the sensitiveness of the critical points, and t denotes different training epochs. o is the output of attention map which was described in Eq. (1). More details of ITA module is illustrated in Supplementary Fig. 1. During the training process, the final output $\hat{P}^i$ is generated by the sum of iterative attention map o_T and the original probabilistic map P^i. Therefore, it has a global contextual view and selectively aggregates contexts according to the spatial attention map.

4 Experiments

We use three EM datasets with rich structure information to demonstrate the effectiveness of the proposed method. In this section, we will introduce the implementation details, datasets, and the experiment results on both datasets.

Datasets. We demonstrate the effectiveness of our proposed method with three different 3D Electron Microscopic Images datasets: **ISBI12** [2], **ISBI13** [1] and **CREMI**. The size of **ISBI12**, **ISBI13** and **CREMI** are $30 \times 512 \times 512$, $100 \times 1024 \times 1024$ and $125 \times 1250 \times 1250$, respectively.

Train settings. We adopt ConvLSTM as our backbone architecture. Also, we apply simple data augmentation, Contrast-Limited Adaptive Histogram Equalization (CLAHE) and random flipping (for ISBI12 only to enlarge training size). For the training parameters, we initialize learning rate (lr) as 0.001 and multiply by 0.5 every 50 epochs. We train our model with batch size of 15 for CREMI and ISBI13, and 30 for ISBI12. The number of training epochs are 35, 900 and 150 for CREMI, ISBI13 and ISBI12, respectively (without attention module). We use cross entropy loss as the optimization metric.

Attention module details. As described in Sec 3.1, Topological-Attention module comes after ConvLSTM. We train the TACLNet for another 15 epochs, with $lr = 0.00001$. Specifically, the patch size is 39×39 for critical points extraction. The iterative rate β is set as 0.5 for ITA module.

Quantitative and qualitative results. In this paper we use similar topology-aware metrics as of [12] for structural accuracy, Adapted Rand Index (ARI), Variation of Information (VOI) and Betti number error. We also report dice scores for pixel accuracy for all the baselines and the proposed method for completeness. The details of the evaluation metrics can be found in the Sec.3 of [12]. For all the experiments, we use three-fold cross-validation to report the average performance and standard deviation over the validation set. Tab. 1 shows the quantitative results for the three different datasets. Note that we remove small connected components as a post-processing step to obtain final segmentation

Table 1. Experiment results for different models on CREMI dataset

Datasets	Models	DICE	ARI	VOI	Betti Error
CREMI	DIVE	0.9542 ± 0.0037	0.6532 ± 0.0247	2.513 ± 0.047	4.378 ± 0.152
	U-Net	0.9523 ± 0.0049	0.6723 ± 0.0312	2.346 ± 0.105	3.016 ± 0.253
	Mosin.	0.9489 ± 0.0053	0.7853 ± 0.0281	1.623 ± 0.083	1.973 ± 0.310
	TopoLoss	0.9596 ± 0.0029	0.8083 ± 0.0104	1.462 ± 0.028	1.113 ± 0.224
	TACLNet	**0.9665 ± 0.0008**	**0.8126 ± 0.0153**	**1.317 ± 0.165**	**0.853 ± 0.183**
ISBI12	DIVE	0.9709 ± 0.0029	0.9434 ± 0.0087	1.235 ± 0.025	3.187 ± 0.307
	U-Net	0.9699 ± 0.0048	0.9338 ± 0.0072	1.367 ± 0.031	2.785 ± 0.269
	Mosin.	0.9716 ± 0.0022	0.9312 ± 0.0052	0.983 ± 0.035	1.238 ± 0.251
	TopoLoss	**0.9755 ± 0.0041**	**0.9444 ± 0.0076**	0.782 ± 0.019	0.429 ± 0.104
	TACLNet	0.9576 ± 0.0047	0.9417 ± 0.0045	**0.771 ± 0.027**	**0.417 ± 0.117**
ISBI13	DIVE	0.9658 ± 0.0020	0.6923 ± 0.0134	2.790 ± 0.025	3.875 ± 0.326
	U-Net	0.9649 ± 0.0057	0.7031 ± 0.0256	2.583 ± 0.078	3.463 ± 0.435
	Mosin.	0.9623 ± 0.0047	0.7483 ± 0.0367	1.534 ± 0.063	2.952 ± 0.379
	TopoLoss	**0.9689 ± 0.0026**	**0.8064 ± 0.0112**	1.436 ± 0.008	1.253 ± 0.172
	TACLNet	0.9510 ± 0.0022	0.7943 ± 0.0127	**1.305 ± 0.016**	**1.175 ± 0.108**

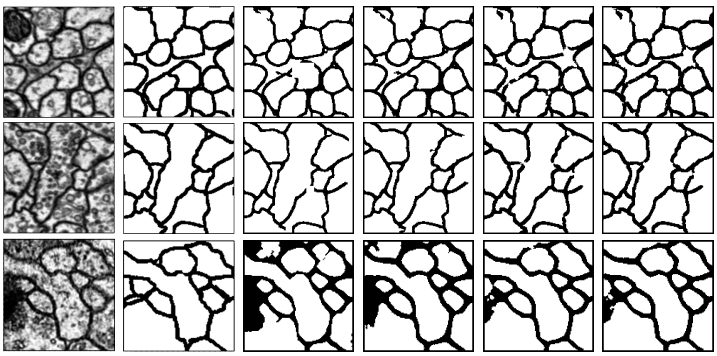

Fig. 4. An illustration of structural accuracy. From left to right: a sample patch, the ground truth, results of UNet, TopoLoss, ConvLSTM and the proposed TACLNet.

results. Our method generally outperforms existing methods [9,20,18] in terms of topology-aware metrics. Fig. 4 shows qualitative results. Our method achieves better consistency/connection compared with other baselines.

In Tab. 1, the Betti Error of our TACLNet brings 23.3% improvement compared to the best baseline, TopoLoss [12], for CREMI dataset. Meanwhile, TACLNet achieves best performances in terms of Betti Error and VOI on both ISBI12 and ISBI13 datasets. In Fig. 4, the results from TACLNet possess better structures with less broken boundaries comparing to other baseline methods. Results show that our attention module strengthens the structural performance overall. Also, the proposed method computes topological information on a stack of 2D images rather than directly on a 3D image, and this significantly reduces the computational expense. Specifically, for CREMI dataset, our method takes

Table 2. Ablation study results for TACLNet on CREMI dataset

Models	DICE	ARI	VOI	Betti Error
ConvLSTM	0.9667 ± 0.0007	0.7627 ± 0.0132	1.753 ± 0.212	1.785 ± 0.254
ConvLSTM + STA	0.9663 ± 0.0004	0.7957 ± 0.0144	1.496 ± 0.156	0.873 ± 0.212
Our TACLNet	0.9665 ± 0.0008	**0.8126 ± 0.0153**	**1.317 ± 0.165**	**0.853 ± 0.183**

Table 3. Ablation study results for number of input slices on CREMI

Number	ARI	VOI	Betti Error	Time
1s	0.7813 ± 0.0141	1.672 ± 0.191	1.386 ± 0.117	0.99h/epoch
3s	**0.8126 ± 0.0153**	**1.317 ± 0.165**	**0.853 ± 0.183**	1.20h/epoch
5s	0.8076 ± 0.0107	1.461 ± 0.125	0.967 ± 0.098	2.78h/epoch

≈1.2 hours per epoch to train, whereas topoloss (3D version) takes ≈2.8h per epoch.

Ablation study for TACLNet. Table. 2 shows the ablation study of the proposed method, which demonstrates the individual contributions of the two proposed modules, Spatial Topological-Attention and Iterative Topological-Attention. As shown in Table. 2, compared with the backbone ConvLSTM model (Betti Error = 1.785), the STA improves the performance remarkably to 0.873. After applying the ITA module, the network further improves the performance by $2.3\%, 11.9\%, 2.1\%$ in Betti Error, VOI, and ARI, respectively. In addition, the combination of STA and ITA also improves the speed of convergence. For our ablation study, STA was trained with 50 epochs, but STA + ITA (our TACLNet) was trained with fewer than 15 epochs for a better performance, which demonstrates that the ITA module can stabilize the training procedure.

Ablation study for number of input slices. Table. 3 is an illustration for the number of input slices. As shown in Table. 3, compared with 1 slice (Betti Error = 1.386) or 5 slices (Betti Error = 0.967), the adopted setting of 3 slices achieves the best results (Betti Error = 0.853). It's not surprised that the 3 slices achieves better performance than 1 slice, as it makes use of inter-slice information. On the other hand, the dataset is anisotropic, and the slices further away are increasingly different from the center slice, which degrades the performance for 5 slices setting.

Illustration of the attention module. We select two images to show the effectiveness of the attention module in Fig. 5. Compared with the second column showing only the critical points detected in the current slice, the attention map in the third column captures more information with the structural similarity from adjacent slices. The attention areas on the final probabilistic map (last column) are highlighted with red color. We observe that the responses of most broken connections are high with attention module enhancement. In summary, Fig. 5 demonstrates that our TACLNet successfully captures the structure information and further improves responses on those essential areas.

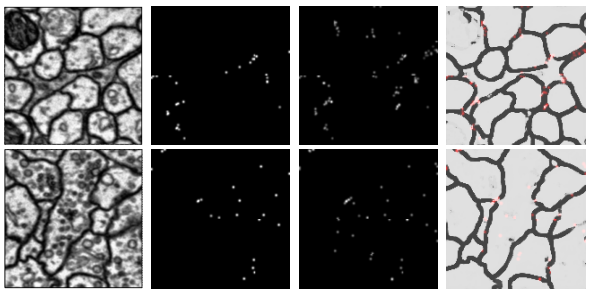

Fig. 5. Illustration of the proposed Topological-Attention. From left to right: original images S^i, the smooth critical points of CP^i, final attention map o^i, and the final probability map $\hat{P}^i$ with o^i superimposed in red (Zoom in and best viewed in color).

5 Conclusion

In this paper, we proposed a novel Topological-Attention Module with ConvLSTM, named TACLNet, for 3D EM image segmentation. Validated with three EM anisotropic datasets, our method outperforms baselines in terms of topology-aware metrics. We expect the performance to further improve on isotropic datasets, because slices are closer (due to higher sampling rate in the z-dimension) with more consistent topologies across slices. For the future work, we will apply TACLNet to datasets of other medical structures, such as cardiac and vascular images, to prove its efficacy in a broader medical domain.

Acknowledgements

This work was partially supported by grants NSF IIS-1909038, CCF-1855760, NCI 1R01CA253368-01 and PSC-CUNY Research Award 64450-00-52.

References

1. Arganda-Carreras, I., Seung, H., Vishwanathan, A., Berger, D.: 3d segmentation of neurites in em images challenge-isbi 2013 (2013)
2. Arganda-Carreras, I., Turaga, S.C., Berger, D.R., Cireşan, D., Giusti, A., Gambardella, L.M., Schmidhuber, J., Laptev, D., Dwivedi, S., Buhmann, J.M., et al.: Crowdsourcing the creation of image segmentation algorithms for connectomics. Frontiers in neuroanatomy **9**, 142 (2015)
3. Bubenik, P.: Statistical topological data analysis using persistence landscapes. JMLR **16**(1), 77–102 (2015)
4. Chen, J., Yang, L., Zhang, Y., Alber, M., Chen, D.Z.: Combining fully convolutional and recurrent neural networks for 3d biomedical image segmentation. In: NeurIPS. pp. 3036–3044 (2016)
5. Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O.: 3d u-net: learning dense volumetric segmentation from sparse annotation. In: MICCAI. pp. 424–432. Springer (2016)

6. Clough, J.R., Oksuz, I., Byrne, N., Schnabel, J.A., King, A.P.: Explicit topological priors for deep-learning based image segmentation using persistent homology. In: IPMI. pp. 16–28. Springer (2019)
7. Edelsbrunner, H., Harer, J.: Computational topology: an introduction. American Mathematical Soc. (2010)
8. Edelsbrunner, H., Letscher, D., Zomorodian, A.: Topological persistence and simplification. In: Proceedings 41st Annual Symposium on Foundations of Computer Science. pp. 454–463. IEEE (2000)
9. Fakhry, A., Peng, H., Ji, S.: Deep models for brain em image segmentation: novel insights and improved performance. Bioinformatics **32**(15), 2352–2358 (2016)
10. Fasy, B.T., Lecci, F., Rinaldo, A., Wasserman, L., Balakrishnan, S., Singh, A., et al.: Confidence sets for persistence diagrams. The Annals of Statistics **42**(6), 2301–2339 (2014)
11. Funke, J., Tschopp, F., Grisaitis, W., Sheridan, A., Singh, C., Saalfeld, S., Turaga, S.C.: Large scale image segmentation with structured loss based deep learning for connectome reconstruction. TPAMI **41**(7), 1669–1680 (2018)
12. Hu, X., Li, F., Samaras, D., Chen, C.: Topology-preserving deep image segmentation. In: NeurIPS. pp. 5658–5669 (2019)
13. Hu, X., Wang, Y., Fuxin, L., Samaras, D., Chen, C.: Topology-aware segmentation using discrete morse theory. In: ICLR (2021), https://openreview.net/forum?id=LGgdb4TS4Z
14. Lin, G., Shen, C., Van Den Hengel, A., Reid, I.: Efficient piecewise training of deep structured models for semantic segmentation. In: CVPR. pp. 3194–3203 (2016)
15. Lin, Z., Feng, M., Santos, C.N.d., Yu, M., Xiang, B., Zhou, B., Bengio, Y.: A structured self-attentive sentence embedding. arXiv preprint arXiv:1703.03130 (2017)
16. Meirovitch, Y., Mi, L., Saribekyan, H., Matveev, A., Rolnick, D., Shavit, N.: Cross-classification clustering: An efficient multi-object tracking technique for 3-d instance segmentation in connectomics. In: CVPR. pp. 8425–8435 (2019)
17. Milletari, F., Navab, N., Ahmadi, S.A.: V-net: Fully convolutional neural networks for volumetric medical image segmentation. In: 3DV. pp. 565–571. IEEE (2016)
18. Mosinska, A., Marquez-Neila, P., Koziński, M., Fua, P.: Beyond the pixel-wise loss for topology-aware delineation. In: CVPR. pp. 3136–3145 (2018)
19. Nunez-Iglesias, J., Ryan Kennedy, T.P., Shi, J., Chklovskii, D.B.: Machine learning of hierarchical clustering to segment 2d and 3d images. PloS one **8**(8) (2013)
20. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: MICCAI. pp. 234–241. Springer (2015)
21. Shen, T., Zhou, T., Long, G., Jiang, J., Pan, S., Zhang, C.: Disan: Directional self-attention network for rnn/cnn-free language understanding. In: AAAI (2018)
22. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. In: NeurIPS. pp. 5998–6008 (2017)
23. Wang, X., Girshick, R., Gupta, A., He, K.: Non-local neural networks. In: CVPR. pp. 7794–7803 (2018)
24. Wong, E., Palande, S., Wang, B., Zielinski, B., Anderson, J., Fletcher, P.T.: Kernel partial least squares regression for relating functional brain network topology to clinical measures of behavior. In: ISBI. pp. 1303–1306. IEEE (2016)
25. Wu, P., Chen, C., Wang, Y., Zhang, S., Yuan, C., Qian, Z., Metaxas, D., Axel, L.: Optimal topological cycles and their application in cardiac trabeculae restoration. In: International Conference on Information Processing in Medical Imaging. pp. 80–92. Springer (2017)

26. Xingjian, S., Chen, Z., Wang, H., Yeung, D.Y., Wong, W.K., Woo, W.c.: Convolutional lstm network: A machine learning approach for precipitation nowcasting. In: NeurIPS. pp. 802–810 (2015)
27. Yang, J., Hu, X., Chen, C., Tsai, C.: 3d topology-preserving segmentation with compound multi-slice representation. In: ISBI. pp. 1297–1301. IEEE (2021)
28. Ye, Z., Chen, C., Yuan, C., Chen, C.: Diverse multiple prediction on neuron image reconstruction. In: MICCAI. pp. 460–468. Springer (2019)
29. Zhang, H., Goodfellow, I., Metaxas, D., Odena, A.: Self-attention generative adversarial networks. In: ICML. pp. 7354–7363. PMLR (2019)
30. Zhao, H., Zhang, Y., Liu, S., Shi, J., Change Loy, C., Lin, D., Jia, J.: Psanet: Point-wise spatial attention network for scene parsing. In: ECCV. pp. 267–283 (2018)