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

J. Yang[1 *], X. Hu[2 *], C. Chen[2] and C. Tsai[1]

1. Graduate Center, CUNY 2. Stony Brook University

ABSTRACT

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) and an Iterative Topological-Attention (ITA) module to effectively transfer more stable topologically critical information across slices. Our proposed method outperforms various baselines in terms of topology-aware evaluation metrics.

INTRODUCTION

Motivation: Most existing methods mainly focus on per-pixel accuracy and are prone to structural errors; Differentiable topological methods have their limitation when applied to 3D images.

Contributions:

- **Spatial Topological-Attention**: Spatial topological information across adjacent slices.
- **Iterative Topological-Attention:** Improve the stability of the topologically critical maps
- **Topological-Attention with ConvLSTM:** Better performance on 3D image segmentation

METHOD

- Our method treats a 3D image as 2D slices stacked along the Z-dimension to capture the inter-slice information.
- l consecutive slices ($l = 3$) are fed into ConvLSTM
- Three probabilistic maps are fed into the Topological-Attention module after ConvLSTM
- The segmentation result is returned after the attention module

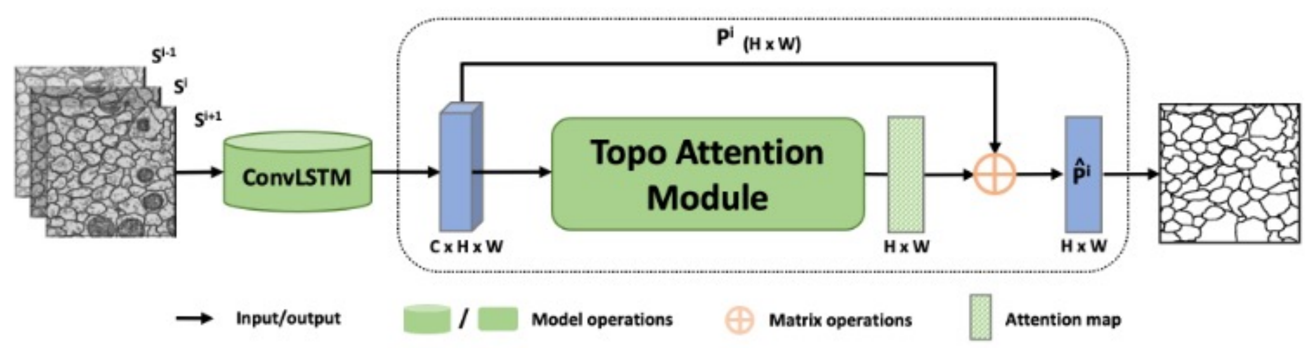

Fig. 2. Overview of the proposed framework

ConvLSTM:

- ConvLSTM is an extension of FC-LSTM, which has the convolutional operators in LSTM gates instead.
- Inputs: $\{S^{i-1}, S^i, S^{i+1}\} \in R^{H \times W}$, and the output $\{P^{i-1}, P^i, P^{i+1}\} \in R^{H \times W}$
- Each is the probabilistic map P^i of the corresponding input slice S^i

Spatial Topological-Attention (STA) Module:

- STA module finds the correlation between the topologically critical information of adjacent slices, without introducing extra parameters.

$$CP^i = Gaussian(PH(P^i)).$$

$$o^i_n = \sum_{m=1}^{N} (P^i{}_m SM_{nm})$$

$$\hat{P}^i_n = \alpha o^i_n + P_n{}^i$$

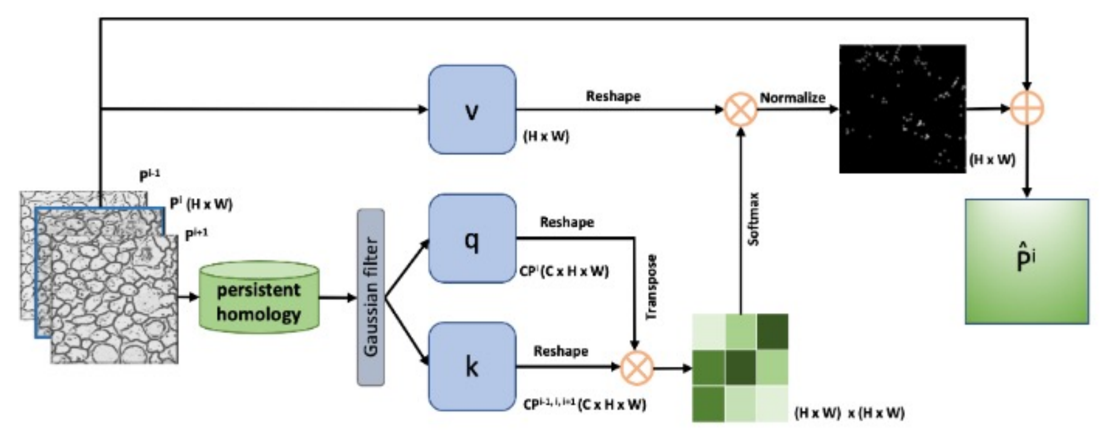

Fig. 3. Illustration of the Spatial Topological-Attention (STA) module

- $PH(\cdot)$ generates isolated critical points and $Gaussian(\cdot)$ is a Gaussian operation.
- Combine the critical maps $CP^{i-1}, CP^i, CP^{i+1} \in R^{H \times W}$ into one single $k \in R^{C \times H \times W}$ $(C = 3)$
- Expand the CP^i into $q \in R^{C \times H \times W}$, the similarity map is obtained by matrix operation and a SoftMax $SM \in R^{N \times N}$
- Obtain normalized topological attention map o^i_n for target slice
- Perform a weighted element-wise sum operation on original probabilistic map P^i

Iterative Topological-Attention (ITA) Module:

- Improve the robustness to obtain stable attention map
- ITA module explores the relationships between the attention maps of different epochs

$$o_t = \beta o_{t-1} + (1 - \beta) o_t.$$

- o is the attention described in STA, β is a parameter and t denotes different training epochs.

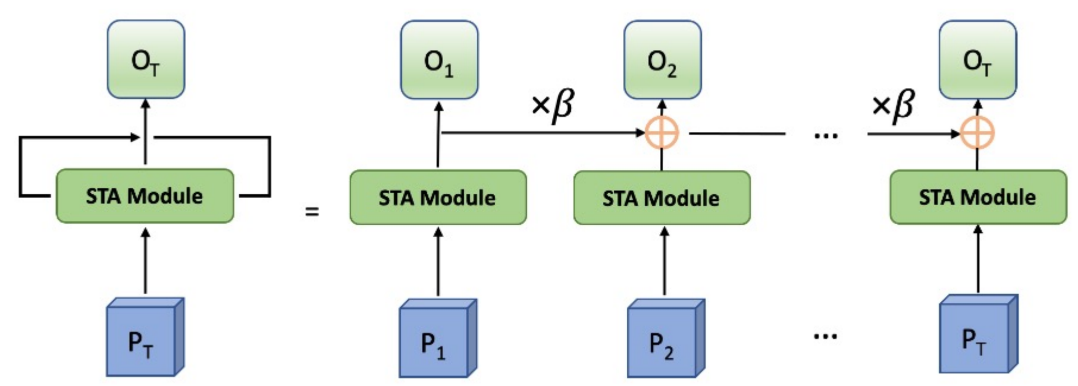

Fig. 1. Illustration of Iterative Topological-Attention (ITA) module. p_T and o_T are the probabilistic map and the output of attention map at T epoch.

RESULTS

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**

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

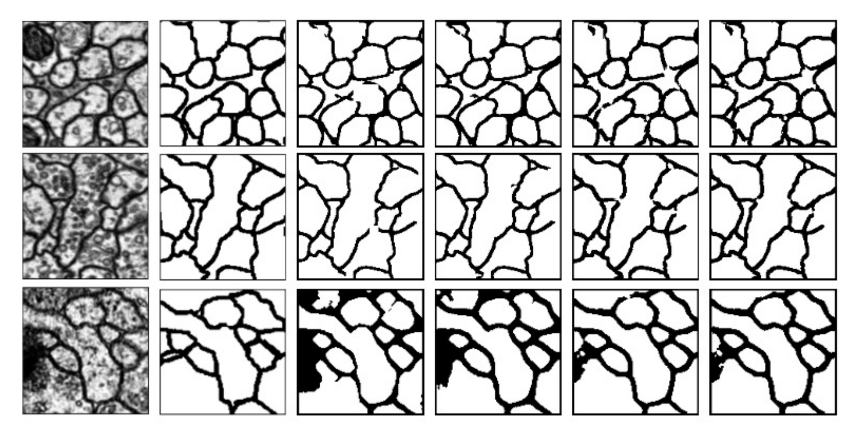

Fig. 4. An illustration of structural accuracy. From left to right: a sample patch, the ground truth, results of UNet, TopoLoss, ConvLSTM and the proposed TACLNet.

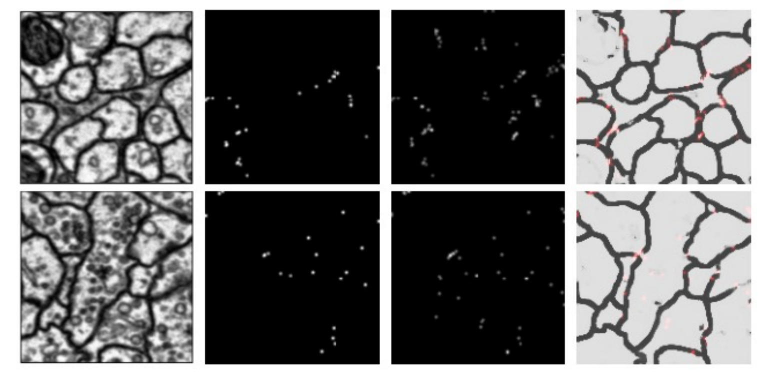

Fig. 5. Illustration of the proposed Topological-Attention. From left to right: original images S^i, the smooth critical points of CP^i, final attention map o^i, and the final probability map $\hat{P}^i$ with o^i superimposed in red (Zoom in and best viewed in color).

- **Datasets:** 3D Electron Microscopic Images ISBI12, ISBI13, and CREMI
- **Ablation Study 1:** ConvLSTM, with Spatial Topological-Attention, and with Iterative Topological-Attention (our TACLNet)
- **Ablation Study 2:** Number of input adjacent slices with their evaluation and running time

CONCLUSIONS

- We proposed a novel Topological-Attention Module with ConvLSTM, named TACLNet, for 3D EM image segmentation.
- Being validated with three EM anisotropic datasets, our method outperforms baselines in terms of topology-aware metrics.
- For the future work, we will apply TACLNet to other medical datasets to prove its efficacy in a broader 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.

CONTACT INFORMATION

J. Yang: jyang2@gradcenter.cuny.edu

X. Hu: xiaolhu@cs.stonybrook.edu

ABSTRACT

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) and an Iterative Topological-Attention (ITA) module to effectively transfer more stable topologically critical information across slices. Our proposed method outperforms various baselines in terms of topology-aware evaluation metrics.

INTRODUCTION

Motivation: Most existing methods mainly focus on per-pixel accuracy and are prone to structural errors; Differentiable topological methods have their limitation when applied to 3D images.

Contributions:

- **Spatial Topological-Attention**: Spatial topological information across adjacent slices.
- **Iterative Topological-Attention:** Improve the stability of the topologically critical maps
- **Topological-Attention with ConvLSTM:** Better performance on 3D image segmentation

METHOD

- Our method treats a 3D image as 2D slices stacked along the Z-dimension to capture the inter-slice information.
- l consecutive slices $(l = 3)$ are fed into ConvLSTM
- Three probabilistic maps are fed into the Topological-Attention module after ConvLSTM
- The segmentation result is returned after the attention module

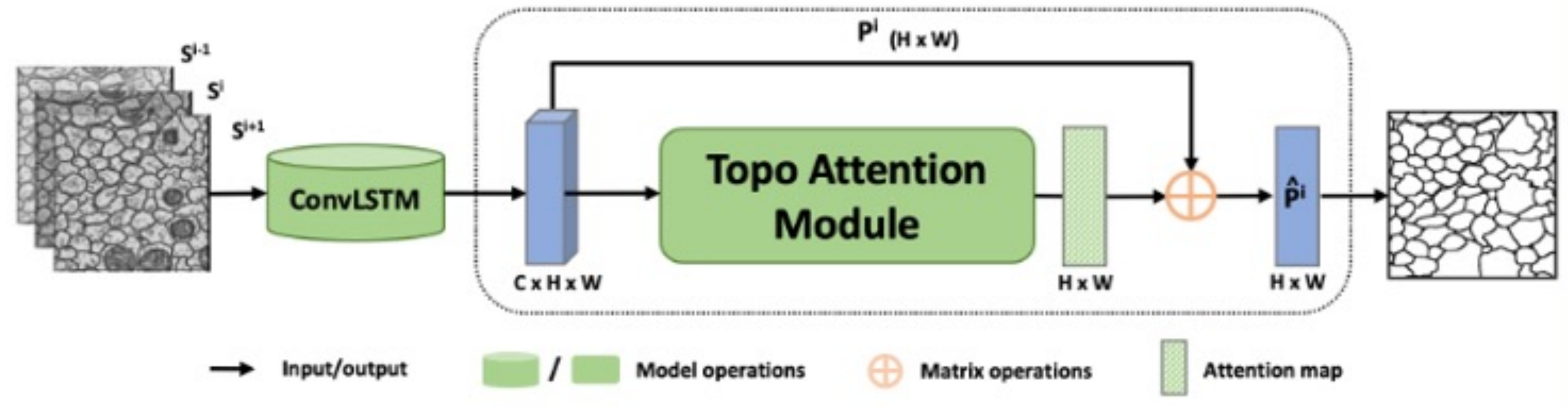

Fig. 2. Overview of the proposed framework

ConvLSTM:

- ConvLSTM is an extension of FC-LSTM, which has the convolutional operators in LSTM gates instead.
- Inputs: $\{S^{i-1}, S^{i}, S^{i+1}\} \in R^{H\times W}$, and the output $\{P^{i-1}, P^{i}, P^{i+1}\} \in R^{H\times W}$
- Each is the probabilistic map P^i of the corresponding input slice S^i

Spatial Topological-Attention (STA) Module:

- STA module finds the correlation between the topologically critical information of adjacent slices, without introducing extra parameters.

$$CP^i = Gaussian(PH(P^i)).$$

$$o_n^i = \sum_{m=1}^{N}(P^{i}_{m} SM_{nm})$$

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

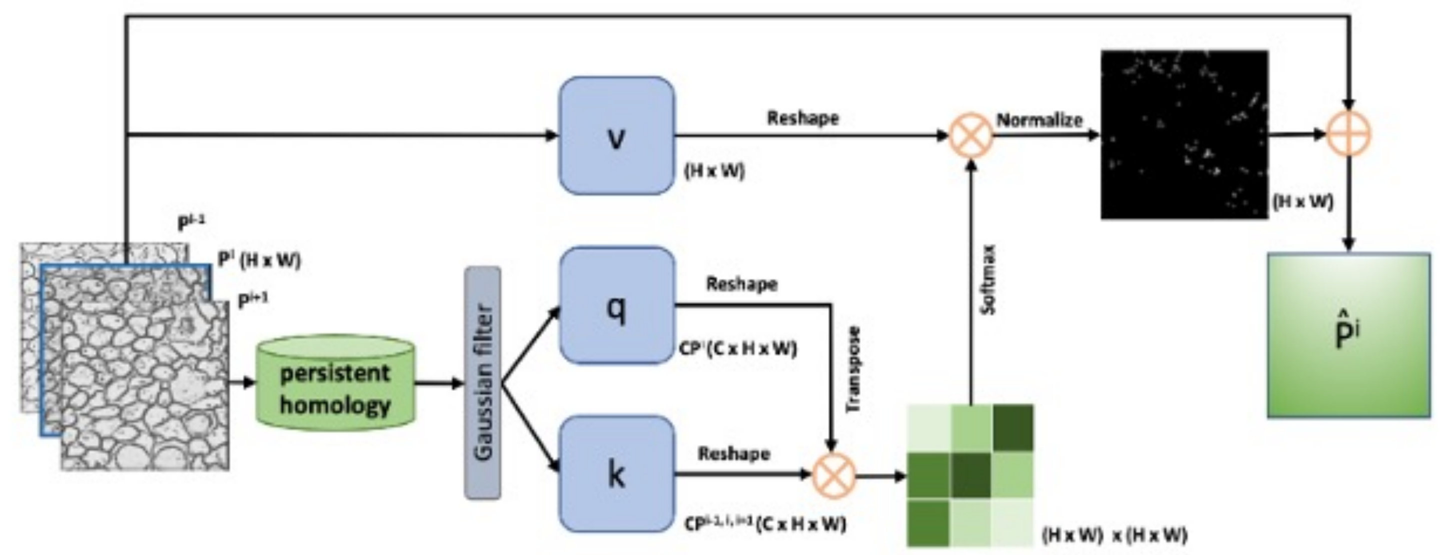

Fig. 3. Illustration of the Spatial Topological-Attention (STA) module

- $PH(\cdot)$ generates isolated critical points and $Gaussian(\cdot)$ is a Gaussian operation.
- Combine the critical maps $CP^{i-1}, CP^{i}, CP^{i+1} \in R^{H\times W}$ into one single $k \in R^{C\times H\times W}$ $(C = 3)$
- Expand the CP^i into $q \in R^{C\times H\times W}$, the similarity map is obtained by matrix operation and a SoftMax $SM \in R^{N\times N}$
- Obtain normalized topological attention map o_n^i for target slice
- Perform a weighted element-wise sum operation on original probabilistic map P^i

Iterative Topological-Attention (ITA) Module:

- Improve the robustness to obtain stable attention map
- ITA module explores the relationships between the attention maps of different epochs

$$o_t = \beta o_{t-1} + (1-\beta)o_t.$$

- o is the attention described in STA, β is a parameter and t denotes different training epochs.

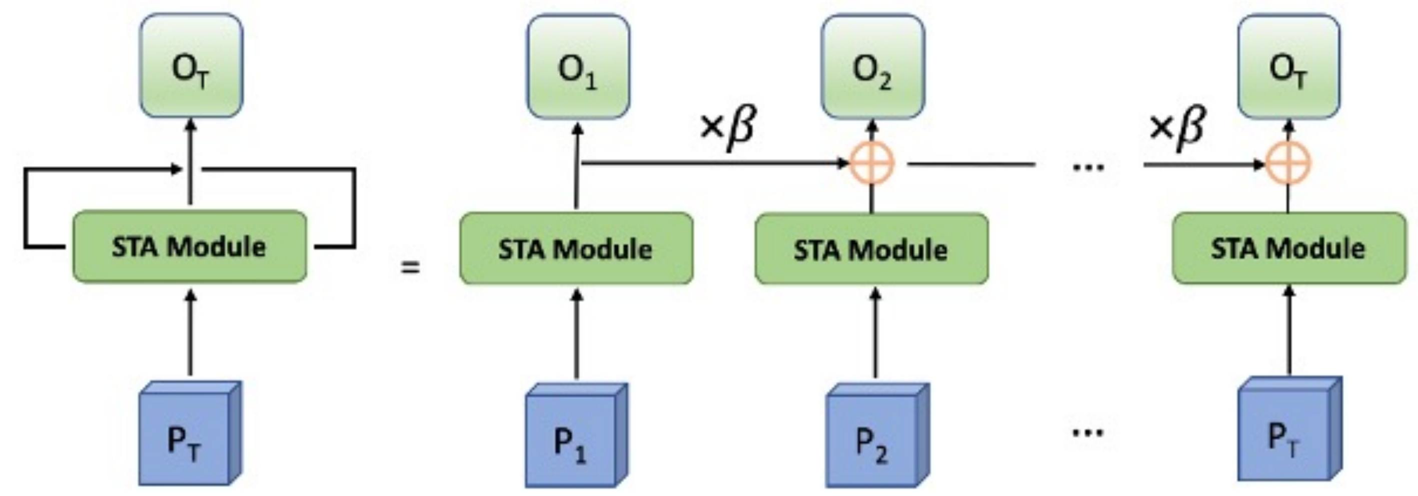

Fig. 1. Illustration of Iterative Topological-Attention (ITA) module. p_T and o_T are the probabilistic map and the output of attention map at T epoch.

RESULTS

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**

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

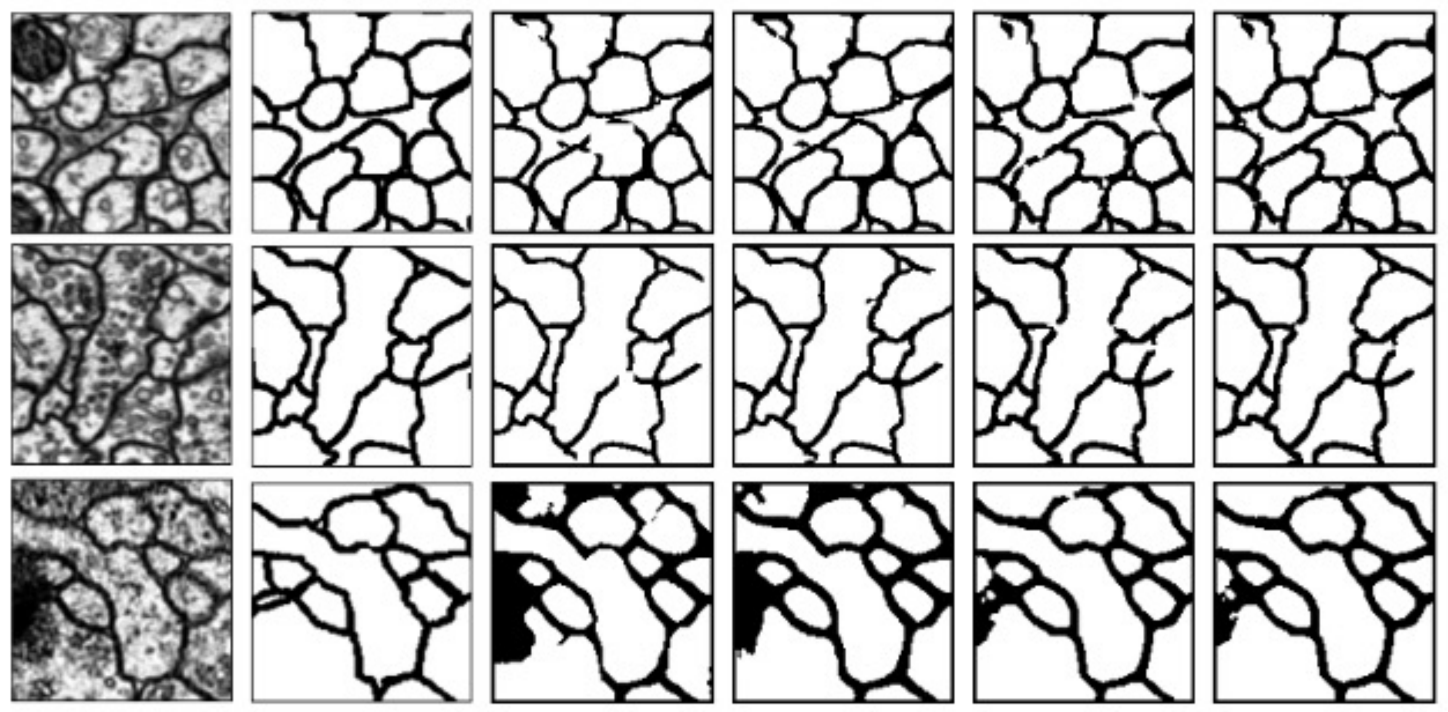

Fig. 4. An illustration of structural accuracy. From left to right: a sample patch, the ground truth, results of UNet, TopoLoss, ConvLSTM and the proposed TACLNet.

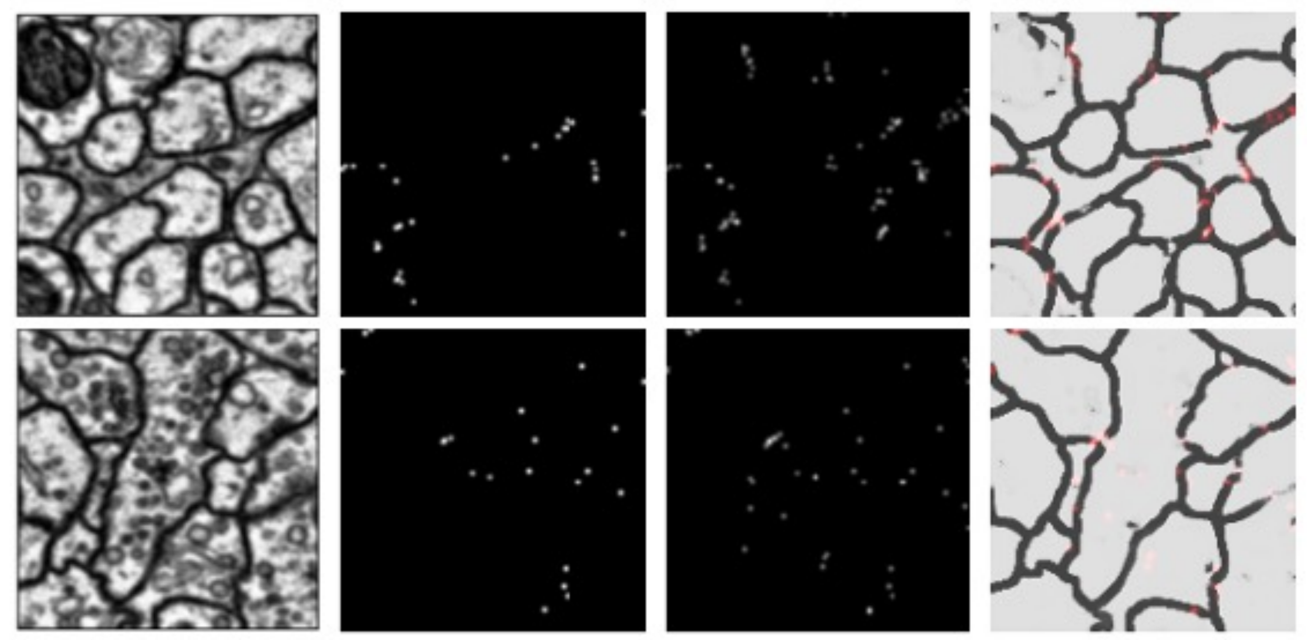

Fig. 5. Illustration of the proposed Topological-Attention. From left to right: original images S^i, the smooth critical points of CP^i, final attention map o^i, and the final probability map $\hat{P}^i$ with o^i superimposed in red (Zoom in and best viewed in color).

- **Datasets:** 3D Electron Microscopic Images ISBI12, ISBI13, and CREMI
- **Ablation Study 1:** ConvLSTM, with Spatial Topological-Attention, and with Iterative Topological-Attention (our TACLNet)
- **Ablation Study 2:** Number of input adjacent slices with their evaluation and running time

CONCLUSIONS

- We proposed a novel Topological-Attention Module with ConvLSTM, named TACLNet, for 3D EM image segmentation.
- Being validated with three EM anisotropic datasets, our method outperforms baselines in terms of topology-aware metrics.
- For the future work, we will apply TACLNet to other medical datasets to prove its efficacy in a broader 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.

CONTACT INFORMATION

J. Yang: jyang2@gradcenter.cuny.edu

X. Hu: xiaolhu@cs.stonybrook.edu