

---

# Towards Geometry-Aware Cell Segmentation in Microscopy Images

---

**Zhexu Jin**

Division of Natural and Applied Sciences  
Duke Kunshan University  
Jiangsu, China  
zhexu.jin@duke.edu

**Gaoyang Li**

School of Life Sciences and Technology  
Tongji University  
Shanghai, China  
lgyzngc@tongji.edu.cn

**Huanshen Cao**

Division of Natural and Applied Sciences  
Duke Kunshan University  
Jiangsu, China  
hc284@duke.edu

**Dongmian Zou**

Division of Natural and Applied Sciences  
Duke Kunshan University  
Jiangsu, China  
dongmian.zou@duke.edu

## Abstract

We present a new approach to distance based cell instance segmentation. Specifically, we design a new loss that more faithfully matches the shapes in segmentation geometry based on computational topology. This loss takes advantage of regularity of the distance maps that require learning. We test our approach using microscopy images consisting of many tissue types in human cells. The results indicate that the new formulation consistently improves the segmentation performance of commonly used network architectures, and the best result advances state-of-the-art.

## 1 Introduction

Cell segmentation is a long-standing task now taken to a higher level by deep learning. Identifying cell and cell nucleus boundaries is critical to a suite of biological and medical applications such as development, disease diagnosis, and structural and functional relationships [7]. In particular, an accurate description of the geometry of a cell and its nucleus, rather than location of the cell instance, is important for analyzing cell morphology [18, 15, 4].

Recently, many deep learning methods for cell segmentation have been developed. One dominating type of deep learning approach is distance-based, due to the special geometrical construction of cells compared to other instance segmentation tasks [23, 20, 9, 7]. Such approach has two stages, as shown in Figure 1. First, a distance transformed map is created from each cell segmentation instance, which is used for supervised training of a neural network. Then, the trained network is applied to each unlabeled cell microscopy image and predicts the distance transformed map, with which a watershed algorithm is used to obtain an instance segmentation for each cell.

However, the aforementioned approaches have ignored the regularity of the transformed inner-distance since the network may be unaware of the geometry of cells and produce irregular and noisy output during inference. In particular, some perturbations in the distance transformed map may lead to high regression loss, while harmless to watershed procedure. However, some other perturbation may be detrimental to watershed procedure, while producing small regression loss. This is problematic during post-processing, since the watershed segmentation cannot produce stable instance segmentation under a noisy distance map. To address this issue, we leverage the crucial geometric information and propose a geometry-aware loss function for cell segmentation.

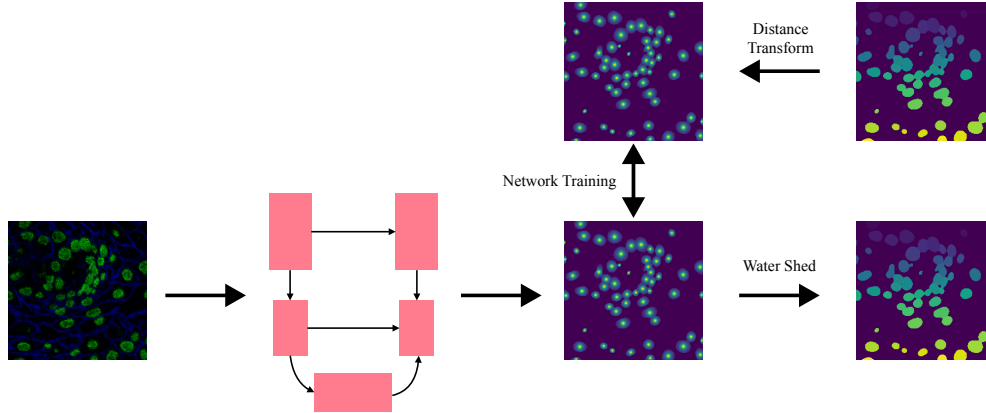


Figure 1: Pipeline for commonly used distance cell segmentation approach.

## 2 Methodology

**Transformer for Medical Image Segmentation** Transformers have outperformed many other architectures on various image tasks, including instance and semantic segmentation [5, 12, 11]. A Fourier decomposition of the learned feature space by multi-headed self-attention layer [16] has shown that multi-headed self-attention facilitates the network to learn low-frequency features. Thus, models equipped with multi-headed self-attention layers are more robust to high-frequency noise. The denoising characteristic of multi-headed self-attention layers has made the transformer architecture variants more suitable for segmentation tasks in medical imaging than traditional convolutional networks. For example, [17] used a transformer for yeast cell instance segmentation, which achieved competitive results. Many other works also used transformers for blood vessel and other image data with tubular structure [19, 14]. In this work, we primarily use a Unet with transformer blocks in each stage. We consider two different backbone ResNext [22] and the Swin transformer [3].

**Geometrically Aware Loss** To alleviate the effect of choosing a distance threshold for a watershed boundary, we enforce the network to learn the geometrical shape of the cell by adding a regularization term to the mean squared error (MSE) loss function of the Unet. The resulting geometry-aware loss is

$$\mathcal{L}(D, \widehat{D}) = \text{MSE}(D, \widehat{D}) + \lambda \sum_{i,j} \mathbf{1}_{(D_{i,j}-\tau)(\widehat{D}_{i,j}-\tau) \leq 0}.$$

Here,  $D$  denotes the ground-truth distance transformed matrix and  $\widehat{D}$  denotes the predicted distance transformed matrix. The indices  $i, j$  indicate pixel locations.  $\mathbf{1}$  is the indicator function which equals 1 if the condition holds and 0 otherwise.  $\tau$  matches the masking threshold used during post-processing the distance transformed mask where we run the watershed algorithm.  $\lambda$  is the regularization parameter. This regularization enforces the net to rely on the mask threshold.

From the viewpoint of persistent homology, the regularization term can be considered as a metric on the 0-dimensional homology group  $H_0$  in the first step of simplex filtration [6]. Here  $H_0$  is characterized by Betti number 0,  $\beta_0$ . Since Betti numbers  $\beta_i$  can be understood as the number of  $i$ -dimensional holes in a simplicial complex,  $\beta_0$  in our case can be intuitively understood as the number of *connected components*. Thus, our loss characterizes the number of connected components at the beginning of simplex filtration. The idea of persistence homology is to track the change of the topological features in the simplicial component evolves over time. The theoretical background in algebraic topology has made persistent homology useful in segmentation applications [1, 10, 8]. However, no extension has been made to instance segmentation or microscopy image data with a cellular structure. In our case, since persistent homology is stable with small perturbation in the simplicial complex, we believe it is particularly helpful for the noisy images. Our work presents the first attempt at incorporating the idea into cell instance segmentation.

We use an example to demonstrate the effectiveness of our proposed loss in Figure 2. Notice that for the two cells highlighted in red bounding boxes, the SwinUnet model trained with geometric

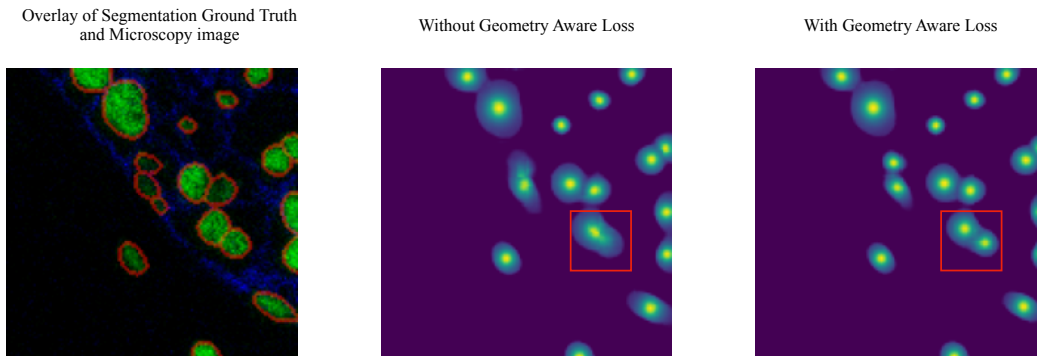


Figure 2: Example of the segmentation results on the same testing image with our proposed geometric loss and without the geometric loss. **Left**: the ground truth segmentation; **Center** and **Right**: the inner-distance produced by SwinUnet trained without and with geometry-aware loss, respectively.

loss successfully differentiates the two cell nuclei that have been merged together. Clearly, with the geometric loss, the network can better differentiate boundaries between different cell instances.

### 3 Results and Discussion

**Dataset** We consider the publicly available annotated human tissue microscopy image dataset TissueNet [7]. The dataset contains 2580 training microscopy images of 512 by 512 resolution, 1324 validation images and 3118 testing images both of 256 by 256 pixel resolution. The dataset represents 6 different human tissue types: immune, pancreas, gi, lung, breast and skin. To generate the training set, a random crop and random flip are performed on the original 512 by 512 resolution image to obtain the 256 by 256 resolution image.

**Implementation Details** We evaluate two segmentation architectures: Unet++[24] with ResNext[22] and the SwinUnet [3] as segmentation backbone. The SwinUnet is composed of 4 encoder blocks and 4 decoder blocks with a window size of 8. The network is trained using the AdamW [13] optimizer on two different loss formulations with a batch size of 32 and a learning rate of 0.001 for 250 epochs. Both architectures include the plain version and the geometric loss version. We also include the original Mesmer [7], which is considered the state-of-the-art. For post-processing, we use the watershed algorithm publicly available in scikit-image [21].

**Results** We present our quantitative results in Table 1, which shows both the mean intersection over union (IoU) and F1-score for the segmentation. The F1-scores are calculated where the true positive is defined by a 0.5 IoU threshold. The results reveal that both network models are improved with our proposed loss. In particular, the geometric loss with the SwinUnet backbone performs the best among all methods, excelling the Mesmer baseline. The results demonstrate that networks trained with our proposed loss more faithfully represents the ground-truth shape of the cell nuclei.

Table 1: Results for Methods with and without Geometry-Aware Losses

	Precision	Recall	F1-score	Average IoU
Mesmer [7]	0.8117	0.6685	0.7332	0.8460
Plain Resnext-Unet++	0.7975	0.6394	0.7098	0.7406
Resnext-Unet++ & Geometric Loss	0.8069	0.6895	0.7259	0.7484
Plain SwinUnet	0.8674	0.8513	0.8592	0.8835
SwinUnet & Geometric Loss	<b>0.8748</b>	<b>0.8594</b>	<b>0.8670</b>	<b>0.9030</b>

**Discussion** Our results show the effectiveness of using geometric information in devising a segmentation algorithm. For future work, we will fully leverage the idea of persistent homology, and provide more theoretical work and experimentation. Diverse datasets will be used to exploit the full potential of prior geometry information for cell segmentation.

One potential problem, when we use the theory of Persistence Homology to design the loss function for segmentation, is the instability of generators [2]. Specifically, distances based on the topological features are agnostic to the exact locations in the image with which the topological features are associated. Given that there are many cell instances in each cell microscopy image, the loss maybe unaware when the detected topological feature is out of place. In future work, we intend to address this problem by fusing the positional patterns into the distances between topological features.

## Acknowledgement

The Authors would like to thank the helpful preliminary work by Eric Qu (Duke Kunshan). The research results of this article are sponsored by the Kunshan Municipal Government research funding.

## Potential Negative Societal Impact

We believe that a segmentation algorithm that correctly characterizes the geometric shape of cell nuclei would support medical research in the long run. One potential negative impact is that supervised learning methods depend on human labeling of the dataset, which may contain mistakes and harmful for medical uses. Nevertheless, this is not a direct impact coming from our work. Robust frameworks or data cleaning will help alleviate this problem.

## References

- [1] K. Arnavaz, O. Krause, J. M. Krivokapic, S. Heilmann, J. A. Bærentzen, P. Nyeng, and A. Feragen. Semi-supervised, topology-aware segmentation of tubular structures from live imaging 3D microscopy. *CoRR*, abs/2105.09737, 2021.
- [2] P. Bendich, P. Bubenik, and A. Wagner. Stabilizing the unstable output of persistent homology computations. *Journal of Applied and Computational Topology*, 4(2):309–338, 2020.
- [3] H. Cao, Y. Wang, J. Chen, D. Jiang, X. Zhang, Q. Tian, and M. Wang. Swin-Unet: Unet-like pure transformer for medical image segmentation. *arXiv preprint arXiv:2105.05537*, 2021.
- [4] M. Dhainaut, S. A. Rose, G. Akturk, A. Wroblewska, S. R. Nielsen, E. S. Park, M. Buckup, V. Roudko, L. Pia, R. Sweeney, et al. Spatial crispr genomics identifies regulators of the tumor microenvironment. *Cell*, 185(7):1223–1239, 2022.
- [5] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. In *9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021*, 2021.
- [6] H. Edelsbrunner and J. L. Harer. *Computational topology: an introduction*. American Mathematical Society, 2022.
- [7] N. F. Greenwald, G. Miller, E. Moen, A. Kong, A. Kagel, T. Dougherty, C. C. Fullaway, B. J. McIntosh, K. X. Leow, M. S. Schwartz, and et al. Whole-cell segmentation of tissue images with human-level performance using large-scale data annotation and deep learning. *Nature Biotechnology*, Mar 2021.
- [8] X. Hu, F. Li, D. Samaras, and C. Chen. Topology-preserving deep image segmentation. In H. M. Wallach, H. Larochelle, A. Beygelzimer, F. d’Alché-Buc, E. B. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, December 8-14, 2019, Vancouver, BC, Canada*, pages 5658–5669, 2019.

- [9] C. F. Koyuncu, G. N. Gunesli, R. Cetin-Atalay, and C. Gunduz-Demir. Deepdistance: a multi-task deep regression model for cell detection in inverted microscopy images. *Medical Image Analysis*, 63:101720, 2020.
- [10] W. Liu, H. Guo, W. Zhang, Y. Zang, C. Wang, and J. Li. Toposeg: Topology-aware segmentation for point clouds. In L. D. Raedt, editor, *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI 2022, Vienna, Austria, 23-29 July 2022*, pages 1201–1208, 2022.
- [11] Z. Liu, H. Hu, Y. Lin, Z. Yao, Z. Xie, Y. Wei, J. Ning, Y. Cao, Z. Zhang, L. Dong, F. Wei, and B. Guo. Swin transformer V2: scaling up capacity and resolution. *CoRR*, abs/2111.09883, 2021.
- [12] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo. Swin transformer: Hierarchical vision transformer using shifted windows. In *2021 IEEE/CVF International Conference on Computer Vision, ICCV 2021, Montreal, QC, Canada, October 10-17, 2021*, pages 9992–10002. IEEE, 2021.
- [13] I. Loshchilov and F. Hutter. Decoupled weight decay regularization. In *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019*, 2019.
- [14] D. H. Pak, A. Caballero, W. Sun, and J. S. Duncan. Efficient aortic valve multilabel segmentation using a spatial transformer network. In *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, pages 1738–1742, 2020.
- [15] G. Palla, D. S. Fischer, A. Regev, and F. J. Theis. Spatial components of molecular tissue biology. *Nature Biotechnology*, 40(3):308–318, 2022.
- [16] N. Park and S. Kim. How do vision transformers work? In *The Tenth International Conference on Learning Representations, ICLR 2022, Virtual Event, April 25-29, 2022*, 2022.
- [17] T. Prangemeier, C. Reich, and H. Koeppl. Attention-based transformers for instance segmentation of cells in microstructures. In *2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 700–707. IEEE, 2020.
- [18] A. Pratapa, M. Doron, and J. C. Caicedo. Image-based cell phenotyping with deep learning. *Current Opinion in Chemical Biology*, 65:9–17, 2021.
- [19] X. Shen, J. Xu, H. Jia, P. Fan, F. Dong, B. Yu, and S. Ren. Self-attentional microvessel segmentation via squeeze-excitation transformer unet. *Comput. Medical Imaging Graph.*, 97:102055, 2022.
- [20] Y. Song, E.-L. Tan, X. Jiang, J.-Z. Cheng, D. Ni, S. Chen, B. Lei, and T. Wang. Accurate cervical cell segmentation from overlapping clumps in pap smear images. *IEEE transactions on medical imaging*, 36(1):288–300, 2016.
- [21] S. Van der Walt, J. L. Schönberger, J. Nunez-Iglesias, F. Boulogne, J. D. Warner, N. Yager, E. Gouillart, and T. Yu. scikit-image: image processing in python. *PeerJ*, 2:e453, 2014.
- [22] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He. Aggregated residual transformations for deep neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1492–1500, 2017.
- [23] Y. Xie, F. Xing, X. Kong, H. Su, and L. Yang. Beyond classification: structured regression for robust cell detection using convolutional neural network. In *International conference on medical image computing and computer-assisted intervention*, pages 358–365. Springer, 2015.
- [24] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang. Unet++: Redesigning skip connections to exploit multiscale features in image segmentation. *IEEE transactions on medical imaging*, 39(6):1856–1867, 2019.