

Structure-Preserving Instance Segmentation via Skeleton-Aware Distance Transform

Zudi Lin^{1†}, Donglai Wei², Aarush Gupta^{3*}, Xingyu Liu^{1*}, Deqing Sun⁴, and Hanspeter Pfister¹

¹Harvard University ²Boston College ³CMU ⁴Google Research

Abstract. Objects with complex structures pose significant challenges to existing instance segmentation methods that rely on boundary or affinity maps, which are vulnerable to small errors around contacting pixels that cause noticeable connectivity change. While the distance transform (DT) makes instance interiors and boundaries more distinguishable, it tends to overlook the intra-object connectivity for instances with varying width and result in over-segmentation. To address these challenges, we propose a *skeleton-aware* distance transform (SDT) that combines the merits of object skeleton in preserving connectivity and DT in modeling geometric arrangement to represent instances with arbitrary structures. Comprehensive experiments on histopathology image segmentation demonstrate that SDT achieves state-of-the-art performance.

1 Introduction

Instances with complex shapes arise in many biomedical domains, and their morphology carries critical information. For example, the structure of gland tissues in microscopy images is essential in accessing the pathological stages for cancer diagnosis and treatment. These instances, however, are usually closely in touch with each other and have non-convex structures with parts of varying widths (Fig. 1a), posing significant challenges for existing segmentation methods.

In the biomedical domain, most methods [3,14,4,22,13] first learn intermediate representations and then convert them into masks with standard segmentation algorithms like connected-component labeling and watershed transform. These representations are not only efficient to predict in one model forward pass but also able to capture object **geometry** (*i.e.*, precise instance boundary), which are hard for top-down methods using low-resolution features for mask generation. However, existing representations have several restrictions. For example, boundary map is usually learned as a pixel-wise binary classification task, which makes the model conduct relatively local predictions and consequently become vulnerable to small errors that break the connectivity between adjacent instances (Fig. 1b). To improve the boundary map, Deep Watershed Transform (DWT) [1] predicts the Euclidean distance transform (DT) of each pixel to the

[†] Currently affiliated with Amazon Alexa. Work was done before joining Amazon.

* Work was done during an internship at Harvard University.

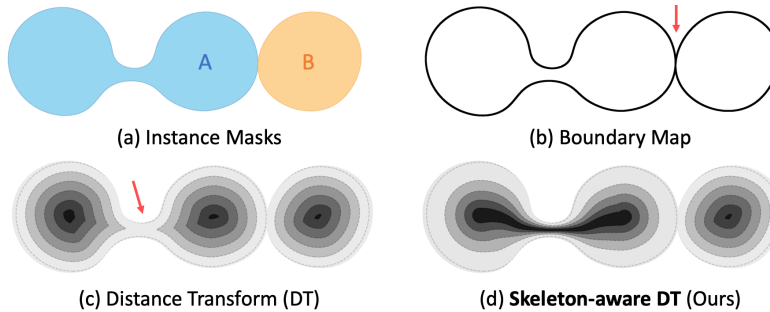


Fig. 1: Skeleton-aware distance transform (SDT). Given (a) instance masks, (b) the boundary map is prone to false merge errors at object contact pixels while (c) the distance transform (DT) struggles to preserve object connectivity. (d) Our SDT can both separate touching instances and enforce object connectivity.

instance boundary. This representation is more aware of the structure for convex objects, as the energy value for centers is significantly different from pixels close to the boundary. However, for objects with non-convex morphology, the boundary-based distance transform produces multiple local optima in the energy landscape (Fig. 1c), which tends to break the intra-instance connectivity when applying thresholding and results in over-segmentation.

To preserve the **connectivity** of instances while keeping the precise instance boundary, in this paper, we propose a novel representation named *skeleton-aware* distance transform (SDT). Our SDT incorporate object skeleton, a concise and connectivity-preserving representation of object structure, into the traditional boundary-based distance transform (DT) (Fig. 1d). In quantitative evaluations, we show that our proposed SDT achieves leading performance on histopathology image segmentation for instances with various sizes and complex structures. Specifically, under the Hausdorff distance for evaluating shape similarity, our approach improves the previous state-of-the-art method by relatively 10.6%.

1.1 Related Work

Instance Segmentation. Bottom-up instance segmentation approaches have become de facto for many biomedical applications due to the advantage in segmenting objects with arbitrary geometry. U-Net [14] and DCAN [4] use fully convolutional models to predict the boundary map of instances. Since the boundary map is not robust to small errors that can significantly change instance structure, shape-preserving loss [22] adds a curve fitting step in the loss function to enforce boundary connectivity. In order to further distinguish closely touching instances, deep watershed transform (DWT) [1] predicts the distance transform (DT) that represents each pixel as its distance to the closest boundary. However, for complex structure with parts of varying width, the boundary-based DT tends to produce relatively low values for thin connections and consequently causes over-

segmentation. Compared to DWT, our SDT incorporates object skeleton (also known as medial axis) [2,24,8] that concisely captures the topological connectivity into standard DT to enforce both the geometry and connectivity.

Object Skeletonization. Object skeleton [15] is a one-pixel wide representation of object masks that can be calculated by topological thinning [24,8,12] or medial axis transform [2]. The vision community has been working on direct object skeletonization from images [16,7,9,19]. Among the works, only Shen *et al.* [16] shows the application of the skeleton on segmenting single-object images. We instead focus on the more challenging instance segmentation task with multiple objects closely touching each other. Object skeletons are also used to correct errors in pre-computed segmentation masks [11]. Our SDT framework instead use the skeleton in the direct segmentation from images.

2 Skeleton-Aware Distance Transform

2.1 SDT Energy Function

Given an image, we aim to design a new representation E for a model to learn, which is later decoded into instances with simple post-processing. Specifically, a good representation for capturing complex-structure masks should have two desired properties: *precise geometric boundary* and *robust topological connectivity*.

Let Ω denote an instance mask, and Γ_b be the boundary of the instance (pixels with other object indices in a small local neighborhood). The *boundary* (or affinity) map is a binary representation where $E|_{\Gamma_b} = 0$ and $E|_{\Omega \setminus \Gamma_b} = 1$. Taking the merits of DT in modeling the geometric arrangement and skeleton in preserving connectivity, we propose a new representation E that satisfies:

$$0 = E|_{\Gamma_b} < E|_{\Omega \setminus (\Gamma_b \cup \Gamma_s)} < E|_{\Gamma_s} = 1 \quad (1)$$

Here $E|_{\Omega \setminus \Gamma_s} < E|_{\Gamma_s} = 1$ indicates that there is only one global maximum for each instance, and the value is assigned to a pixel if and only if the pixel is on the skeleton. This property avoids ambiguity in defining the object interior and preserve connectivity. Besides, $E|_{\Omega \setminus \Gamma_b} > E|_{\Gamma_b} = 0$ ensures that boundary is distinguishable as the standard DT, which produces precise geometric boundary.

For the realization of E , let x be a pixel in the input image, and d be the metric, *e.g.*, Euclidean distance. The energy function for distance transform (DT) is defined as $E_{DT}(x) = d(x, \Gamma_b)$, which starts from 0 at object boundary and increases monotonically when x is away from the boundary. Similarly, we can define an energy function $d(x, \Gamma_s)$ representing the distance from the skeleton. It vanishes to 0 when the pixel approaches the object skeleton. Formally, we define the energy function of the *skeleton-aware* distance transform (Fig. 2) as

$$E_{SDT}(x) = \left(\frac{d(x, \Gamma_b)}{d(x, \Gamma_s) + d(x, \Gamma_b)} \right)^\alpha, \quad \alpha > 0 \quad (2)$$

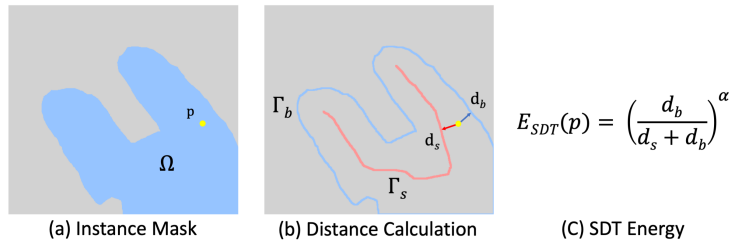


Fig. 2: Illustration of the SDT energy function. (a) Given an instance mask Ω , (b) we calculate the distances of a pixel to both the skeleton and boundary. (c) Our energy function ensures a uniform maximum value of 1 on the skeleton and minimum value of 0 on the boundary, with a smooth interpolation in between.

where α controls the curvature of the energy surface¹. When $0 < \alpha < 1$, the function is concave and decreases faster when being close to the boundary, and vice versa when $\alpha > 1$. In the ablation studies, we demonstrate various patterns of the model predictions given different α .

Besides, since common skeletonization algorithms can be sensitive to small perturbations on the object boundary and produce unwanted branches, we smooth the masks before computing the object skeleton by Gaussian filtering and thresholding to avoid complex branches.

Learning Strategy. Given the ground-truth SDT energy map, there are two ways to learn it using a CNN model. The first way is to *regress* the energy using L_1 or L_2 loss. In the regression mode, the output is a single-channel image. The second way is to quantize the $[0, 1]$ energy space into K bins and rephrase the regression task into a *classification* task [1,18], which makes the model robust to small perturbations in the energy landscape. For the classification mode, the model output has $(K+1)$ channels with one channel representing the background region. We fix the bin size to 0.1 without tweaking, making $K = 10$. Softmax is applied before calculating the cross-entropy loss. We test both learning strategies in the experiments to illustrate the optimal setting for SDT.

2.2 SDT Network

Network Architecture. Directly learning the energy function with a fully convolutional network (FCN) can be challenging. Previous approaches either first regress an easier direction field representation and then use additional layers to predict the desired target [1], or take the multi-task learning approach to predict additional targets at the same time [21,22,16].

Fortunately, with recent progress in FCN architectures, it becomes feasible to learn the target energy map in an end-to-end fashion. Specifically, in all the experiments, we use a DeepLabV3 model [5] with a ResNet [6] backbone

¹ We add $\epsilon = 10^{-6}$ to the denominator to avoid dividing by 0 for the edge case where a pixel is both instance boundary and skeleton (*i.e.*, a one-pixel wide part).

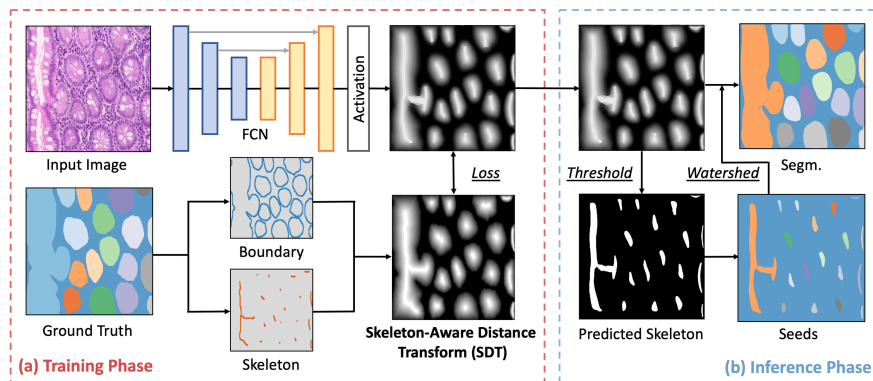


Fig. 3: Overview of the SDT framework. **(a) Training Phase:** target SDT is calculated conditioned on the distance to both the boundary and skeleton. A FCN maps the image into the energy space to minimize the loss. **(b) Inference Phase:** we threshold the SDT to generate skeleton segments, which is processed into seeds with the connected component labeling. Finally, the watershed transform algorithm takes the seeds and the reversed SDT energy to yield the masks.

to directly learn the SDT energy without additional targets (Fig. 3, *Training Phase*). We also add a CoordConv [10] layer before the 3rd stage in the backbone network to introduce spatial information into the segmentation model.

Target SDT Generation. There is an inconsistency problem in object skeleton generation: part of the complete instance skeleton can be different from the skeleton of the instance part (Fig. 4). Some objects may touch the image border due to either a restricted field of view (FoV) of the imaging devices or spatial data augmentation like the random crop. If pre-computing the skeleton, we will get *local skeleton* (Fig. 4c) for objects with missing masks due to imaging restrictions, and *partial skeleton* (Fig. 4b) due to spatial data augmentation, which causes ambiguity. Therefore we calculate the local skeleton for SDT on-the-fly after all spatial transformations instead of pre-computing to prevent the model from hallucinating the structure of parts outside of the currently visible region. In inference, we always run predictions on the whole images to avoid inconsistent predictions. We use the skeletonization algorithm in Lee *et al.* [8], which is less sensitive to small perturbations and produces skeletons with fewer branches.

Instance Extraction from SDT. In the SDT energy map, all boundary pixels share the same energy value and can be processed into segments by direct thresholding and connected component labeling, similar to DWT [1]. However, since the prediction is never perfect, the energy values along closely touching boundaries are usually not sharp and cause split-errors when applying a higher threshold or merge-errors when applying a lower threshold.

Therefore we utilize a skeleton-aware instance extraction (Fig. 3, *Inference Phase*) for SDT. Specifically, we set a threshold $\theta = 0.7$ so that all pixels with the predicted energy bigger than θ are labeled as skeleton pixels. We first perform

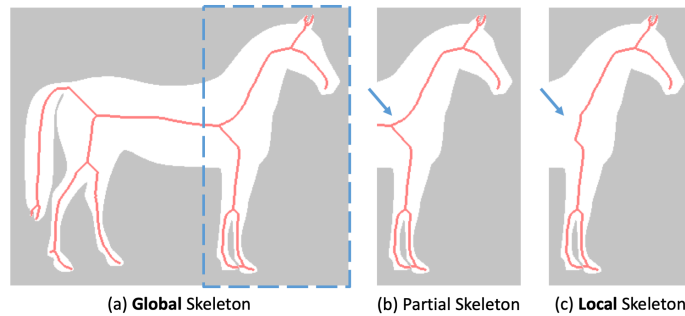


Fig. 4: Skeleton generation rule. (a) Given an instance and the *global* skeleton, (b) the *partial* skeleton cropped from the global skeleton can be different from (c) the *local* skeleton generated from the cropped mask. For SDT, we calculate the local skeleton to prevent the model from extrapolating the unseen parts.

connected component labeling of the skeleton pixels to generate seeds and run the watershed algorithm on the reversed energy map using the seeds as basins (local optima) to generate the final segmentation. We also follow previous works [4,22] and refine the segmentation by hole-filling and removing small spurious objects.

3 Experiments

3.1 Histopathology Instance Segmentation

Accurate instance segmentation of gland tissues in histopathology images is essential for clinical analysis, especially cancer diagnosis. The diversity of object appearance, size, and shape makes the task challenging.

Dataset and Evaluation Metric. We use the gland segmentation challenge dataset [17] that contains colored light microscopy images of tissues with a wide range of histological levels from benign to malignant. There are 85 and 80 images in the training and test set, respectively, with ground truth annotations provided by pathologists. According to the challenge protocol, the test set is further divided into two splits with 60 images of normal and 20 images of abnormal tissues for evaluation. Three evaluation criteria used in the challenge include instance-level F1 score, Dice index, and Hausdorff distance, which measure the performance of object detection, segmentation, and shape similarity, respectively. For the instance-level F1 score, an IoU threshold of 0.5 is used to decide the correctness of a prediction.

Methods in Comparison. We compare SDT with previous state-of-the-art segmentation methods, including DCAN [4], multi-channel network (MCN) [21], shape-preserving loss (SPL) [22] and FullNet [13]. We also compare with suggestive annotation (SA) [23], and SA with model quantization (QSA) [20], which use multiple FCN models to select informative training samples from the dataset.

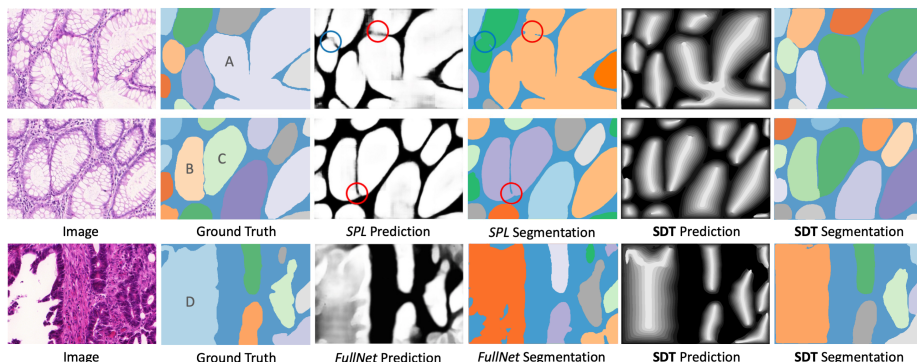


Fig. 5: Visual comparison on histopathology image segmentation. (First 2 rows) Compared with shape-preserving loss (SPL) [22], our SDT unambiguously separates closely touching objects while preserving the structure of complicated masks. (The 3rd row) Compared with FullNet [13], our model infers the SDT energy of instance masks from a global structure perspective instead of boundary that relies on relatively local predictions, which produces high-quality masks.

With the same training settings as our SDT, we also report the performance of skeleton with scales (SS) and traditional distance transform (DT).

Training and Inference. Since the training data is relatively limited due to the challenges in collecting medical images, we apply pixel-level and spatial-level augmentations, including random brightness, contrast, rotation, crop, and elastic transformation, to alleviate overfitting. We set $\alpha = 0.8$ for our SDT in Eqn. 2. We use the classification learning strategy and optimize a model with 11 output channels (10 channels for energy quantized into ten bins and one channel for background). We train the model for 20k iterations with an initial learning rate of 5×10^{-4} and a momentum of 0.9. The same settings are applied to DT. At inference time, we apply argmax to get the corresponding bin index of each pixel and transform the energy value to the original data range. Finally, we apply the watershed-based instance extraction rule described in Sec. 2.2.

Specifically for SS, we set the number of output channels to two, with one channel predicting skeleton probability and the other predicting scales. Since the scales are non-negative, we add a ReLU activation for the second channel and calculate the regression loss. Masks are generated by morphological dilation. We do not quantize the scales as DT and SDT since even ground-truth scales can yield masks unaligned with the instance boundary with quantization.

Results. Our SDT framework achieves state-of-the-art performance on 5 out of 6 evaluation metrics on the gland segmentation dataset (Table 1). With the better distinguishability of object interior and boundary, SDT can unambiguously separate closely touching instances (Fig. 5, first two rows), performs better than previous methods using object boundary representations [4,22]. Besides, under the Hausdorff distance for evaluating shape-similarity between ground-truth and

Method	F1 Score \uparrow		Dice Index \uparrow		Hausdorff \downarrow	
	Part A	Part B	Part A	Part B	Part A	Part B
DCAN [4]	0.912	0.716	0.897	0.781	45.42	160.35
MCN [21]	0.893	0.843	0.908	0.833	44.13	116.82
SPL [22]	0.924	0.844	0.902	0.840	49.88	106.08
SA [23]	0.921	0.855	0.904	0.858	44.74	96.98
FullNet [13]	0.924	0.853	0.914	0.856	37.28	88.75
QSA [20]	0.930	0.862	0.914	0.859	41.78	97.39
SS	0.872	0.765	0.853	0.797	54.86	116.33
DT	0.918	0.846	0.896	0.848	41.84	90.86
SDT	0.931	0.866	0.919	0.851	32.29	82.40

Table 1: Comparison with existing methods on the gland segmentation. Our SDT achieves better or on par F1 score and Dice Index, and significantly better Hausdorff distance for evaluating *shape similarity*. DT and SS represent distance transform and skeleton with scales.

Setting	F1 Score \uparrow		Dice Index \uparrow		Hausdorff \downarrow	
	Part A	Part B	Part A	Part B	Part A	Part B
<i>Loss</i>						
L1	0.916	0.842	0.903	0.850	39.76	94.83
L2	0.896	0.833	0.885	0.837	49.11	110.24
CE	0.931	0.866	0.919	0.851	32.29	82.40
<i>Curvature</i>						
$\alpha = 0.6$	0.912	0.845	0.914	0.855	36.25	91.24
$\alpha = 0.8$	0.931	0.866	0.919	0.851	32.29	82.40
$\alpha = 1.0$	0.926	0.858	0.907	0.849	35.73	86.73
<i>Skeleton</i>						
Partial	0.899	0.831	0.896	0.837	47.50	105.19
Local	0.931	0.866	0.919	0.851	32.29	82.40

Table 2: Ablations studies on the gland dataset. The results suggest that the model trained with cross-entropy loss with $\alpha = 0.8$ and local skeleton generation achieves the best performance.

predicted masks, our SDT reports an average score of 44.82 across two test splits, which improves the previous state-of-the-art approach (*i.e.*, FullNet with an average score of 50.15) by 10.6%. We also notice the different sensitivities of the three evaluation metrics. Taking the instance **D** (Fig. 5, 3rd row) as an example: both SDT and FullNet [13] have 1.0 F1-score (IoU threshold 0.5) for the correct detection; SDT has a slightly higher Dice Index (0.956 vs. 0.931) for better pixel-level classification; and our SDT has significantly lower Hausdorff distance (24.41 vs. 48.81) as SDT yields a mask with much more accurate morphology.

3.2 Ablation Studies

Loss Function. We compare the regression mode using L1 and L2 losses with the classification mode using cross-entropy loss. There is a separate channel for background under the classification mode where the energy values are quantized into bins. However, for regression mode, if the background value is 0, we need to use a threshold $\tau > 0$ to decide the foreground region, which results in shrank masks. To separate the background region from the foreground objects, we assign an energy value of $-b$ to the background pixels ($b \geq 0$). To facilitate the regression, given the predicted value \hat{y}_i for pixel i , we apply a sigmoid function (σ) and affine transformation so that $\hat{y}'_i = (1+b) \cdot \sigma(\hat{y}_i) - b$ has a range of $(-b, 1)$. We set $b = 0.1$ for the experiments. We show that under the same settings, the model trained with quantized energy reports the best results (Table 2). We also notice that the model trained with L_1 loss produces a much sharper energy surface than the model trained with L_2 loss, which is expected.

Curvature. We also compare different α in Eqn. 2 that controls the curvature of the energy landscape. Table 2 shows that $\alpha = 0.8$ achieves the best overall

performance, which is slightly better than $\alpha = 1.0$. Decreasing α to 0.6 introduces more merges and make the results worse.

Global/Local Skeleton. In Sec. 2.2 we show the inconsistency problem of global and local skeletons. In this study, we set $\alpha = 0.8$ and let the model learn the pre-computed SDT energy for the training set. The results show that pre-computed SDT significantly degrades performance (Table 2). We argue this is because pre-computed energy not only introduces inconsistency for instances touching the image border but also restricts the diversity of SDT energy maps.

4 Conclusion

In this paper, we introduce the *skeleton-aware* distance transform (SDT) to capture both the geometry and topological connectivity of instance masks with complex shapes. For multi-class problems, we can use class-aware semantic segmentation to mask the SDT energy trained for all objects that is agnostic to their classes. We hope this work can inspire more research on not only better representations of object masks but also novel models that can better predict those representations with shape encoding. We will also explore the application of SDT in the more challenging 3D instance segmentation setting.

Acknowledgments. This work has been partially supported by NSF awards IIS-2239688 and IIS-2124179. This work has also been partially supported by gift funding and GCP credits from Google.

References

1. Bai, M., Urtasun, R.: Deep watershed transform for instance segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 5221–5229 (2017) 1, 2, 4, 5
2. Blum, H., et al.: A transformation for extracting new descriptors of shape, vol. 4. MIT press Cambridge (1967) 3
3. Briggman, K., Denk, W., Seung, S., Helmstaedter, M.N., Turaga, S.C.: Maximin affinity learning of image segmentation. In: Advances in Neural Information Processing Systems. pp. 1865–1873 (2009) 1
4. Chen, H., Qi, X., Yu, L., Heng, P.A.: Dcan: deep contour-aware networks for accurate gland segmentation. In: Proceedings of the IEEE conference on Computer Vision and Pattern Recognition. pp. 2487–2496 (2016) 1, 2, 6, 7, 8
5. Chen, L.C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L.: 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) 4
6. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016) 4
7. Ke, W., Chen, J., Jiao, J., Zhao, G., Ye, Q.: Srn: side-output residual network for object symmetry detection in the wild. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1068–1076 (2017) 3

8. Lee, T.C., Kashyap, R.L., Chu, C.N.: Building skeleton models via 3-d medial surface axis thinning algorithms. *CVGIP: Graphical Models and Image Processing* **56**(6), 462–478 (1994) [3](#), [5](#)
9. Liu, C., Ke, W., Qin, F., Ye, Q.: Linear span network for object skeleton detection. In: *Proceedings of the European Conference on Computer Vision (ECCV)*. pp. 133–148 (2018) [3](#)
10. Liu, R., Lehman, J., Molino, P., Such, F.P., Frank, E., Sergeev, A., Yosinski, J.: An intriguing failing of convolutional neural networks and the coordconv solution. In: *Advances in Neural Information Processing Systems*. pp. 9605–9616 (2018) [5](#)
11. Matejek, B., Haehn, D., Zhu, H., Wei, D., Parag, T., Pfister, H.: Biologically-constrained graphs for global connectomics reconstruction. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (June 2019) [3](#)
12. Németh, G., Kardos, P., Palágyi, K.: 2d parallel thinning and shrinking based on sufficient conditions for topology preservation. *Acta Cybernetica* **20**(1), 125–144 (Jan 2011). <https://doi.org/10.14232/actacyb.20.1.2011.10> [3](#)
13. Qu, H., Yan, Z., Riedlinger, G.M., De, S., Metaxas, D.N.: Improving nuclei/gland instance segmentation in histopathology images by full resolution neural network and spatial constrained loss. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. pp. 378–386. Springer (2019) [1](#), [6](#), [7](#), [8](#)
14. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: *International Conference on Medical image computing and computer-assisted intervention*. pp. 234–241. Springer (2015) [1](#), [2](#)
15. Saha, P.K., Borgefors, G., di Baja, G.S.: A survey on skeletonization algorithms and their applications. *Pattern recognition letters* **76**, 3–12 (2016) [3](#)
16. Shen, W., Zhao, K., Jiang, Y., Wang, Y., Bai, X., Yuille, A.: Deepskeleton: Learning multi-task scale-associated deep side outputs for object skeleton extraction in natural images. *IEEE Transactions on Image Processing* **26**(11), 5298–5311 (2017) [3](#), [4](#)
17. Sirinukunwattana, K., Pluim, J.P., Chen, H., Qi, X., Heng, P.A., Guo, Y.B., Wang, L.Y., Matuszewski, B.J., Bruni, E., Sanchez, U., et al.: Gland segmentation in colon histology images: The glas challenge contest. *Medical image analysis* **35**, 489–502 (2017) [6](#)
18. Wang, Y., Wei, X., Liu, F., Chen, J., Zhou, Y., Shen, W., Fishman, E.K., Yuille, A.L.: Deep distance transform for tubular structure segmentation in ct scans. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 3833–3842 (2020) [4](#)
19. Wang, Y., Xu, Y., Tsogkas, S., Bai, X., Dickinson, S., Siddiqi, K.: Deepflux for skeletons in the wild. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. pp. 5287–5296 (2019) [3](#)
20. Xu, X., Lu, Q., Yang, L., Hu, S., Chen, D., Hu, Y., Shi, Y.: Quantization of fully convolutional networks for accurate biomedical image segmentation. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 8300–8308 (2018) [6](#), [8](#)
21. Xu, Y., Li, Y., Wang, Y., Liu, M., Fan, Y., Lai, M., Eric, I., Chang, C.: Gland instance segmentation using deep multichannel neural networks. *IEEE Transactions on Biomedical Engineering* **64**(12), 2901–2912 (2017) [4](#), [6](#), [8](#)
22. Yan, Z., Yang, X., Cheng, K.T.T.: A deep model with shape-preserving loss for gland instance segmentation. In: *International Conference on Medical Image Com-*

- puting and Computer-Assisted Intervention. pp. 138–146. Springer (2018) [1](#), [2](#), [4](#), [6](#), [7](#), [8](#)
23. Yang, L., Zhang, Y., Chen, J., Zhang, S., Chen, D.Z.: Suggestive annotation: A deep active learning framework for biomedical image segmentation. In: International conference on medical image computing and computer-assisted intervention. pp. 399–407. Springer (2017) [6](#), [8](#)
 24. Zhang, T., Suen, C.Y.: A fast parallel algorithm for thinning digital patterns. *Communications of the ACM* **27**(3), 236–239 (1984) [3](#)