




This MICCAI paper is the Open Access version, provided by the MICCAI Society. It is identical to the accepted version, except for the format and this watermark; the final published version is available on SpringerLink.

DES-SAM: Distillation-Enhanced Semantic SAM for Cervical Nuclear Segmentation with Box Annotation

Lina Huang¹, Yixiong Liang¹ ^[0000-0003-0407-5838], and Jianfeng Liu²

¹ School of Computer Science and Engineering,
Central South University, Changsha, China
yxliang@csu.edu.cn

² School of Automation, Central South University, Changsha, China

Abstract. Nuclei segmentation in cervical cell images is a crucial technique for the automatic diagnosis of cervical cell pathology. The current state-of-the-art (SOTA) nuclei segmentation methods often require significant time and resources to provide pixel-level annotations for training. To reduce the labor-intensive annotation costs, we propose DES-SAM, a box-supervised cervical nucleus segmentation network with strong generalization ability based on self-distillation prompting. We utilize Segment Anything Model (SAM) to generate high-quality pseudo-labels by integrating a lightweight detector. The main challenges lie in the poor generalization ability brought by small-scale training datasets and the large-scale training parameters of traditional knowledge distillation frameworks. To address these challenges, we propose leveraging the strong feature extraction ability of SAM and a self-distillation prompting strategy to maximize the performance of the downstream nuclear semantic segmentation task without compromising SAM’s generalization. Additionally, we propose an Edge-aware Enhanced Loss to improve the segmentation capability of DES-SAM. Various comparative and generalization experiments on public cervical cell nuclei datasets demonstrate the effectiveness of the proposed method. The code is available at <https://github.com/CVIU-CSU/DES-SAM>.

Keywords: Weakly-supervised Learning · Knowledge Distillation · Segment Anything Model (SAM) · Nuclei Segmentation.

1 Introduction

The morphological and visual characteristics of cervical cell nuclei, such as comprehensive optical density, average size, and heterogeneity, play a significant role in determining the malignancy degree of tumors. These features can be calculated after segmenting individual nuclei. Therefore, nuclear image segmentation is a crucial task for analyzing cervical images. Most of the previous cell nucleus segmentation methods [2,3,27] are fully supervised, which typically require large-scale datasets with well-annotated pixel-level labels, making it expensive and

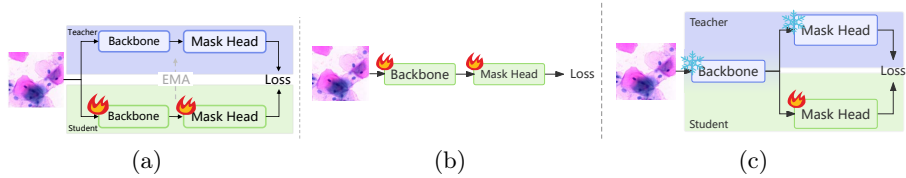


Fig. 1. A Comparison of Three Box-Supervised Image Segmentation Paradigms.

time-consuming. Various weakly-supervised nuclear segmentation methods have been proposed to reduce annotation costs, learning from point [20], scribble [12], or bounding box [16,26] annotations. Among these methods, weakly-supervised approaches leveraging bounding box annotations often achieve a favorable trade-off between performance and annotation costs.

In order to learn from box annotations, previous box-supervised image segmentation methods [11,13,24,26] have used pseudo-label or multiple instance learning from box annotations to achieve performance similar to fully supervised methods. Most of them adhere to two paradigms, as depicted in Fig. 1. One prevalent paradigm following the knowledge distillation framework is exemplified in works such as [11,16,26], as illustrated in Fig. 1(a). These methods often use traditional approaches to first generate noisy pseudo-labels to train the teacher, and then use the teacher’s outputs as pseudo-labels to train the student. However, the generated pseudo-labels often exhibit low quality and noise, and training the entire network with such noisy labels can significantly degrade performance. Another paradigm is shown in Fig. 1(b) which is often based on affinity pairwise loss, exemplified by BoxSnake [25] and BoxLevelSet [13]. However, this method is prone to misinterpreting intricate textures found in images and requires training the entire network, also leading to suboptimal segmentation performance.

To address shortcomings in existing paradigms, we propose DES-SAM, an efficient cervical cell nuclear segmentation network illustrated in Fig. 1(c). We follow the teacher-student architecture but utilize a frozen, powerful vision foundation model to extract features effectively. Based on a self-distillation strategy, we then employ Parameter-Efficient Fine-Tuning (PEFT) to adapt the foundation model to the box-supervised cell nucleus segmentation task. Our approach builds upon the Segment Anything Model (SAM) [8], pretrained on 11 million images and demonstrating strong performance on various downstream tasks. Specifically, we use SAM’s image encoder as the feature extractor and extend SAM by introducing an additional detection branch to automatically produce box prompts, which are then fed into SAM’s mask decoder to generate high-quality segmentation pseudo-labels. Furthermore, due to the limited data for cervical cell nucleus segmentation images, we use the original SAM’s mask head as a teacher to fine-tune the student mask head, which mirrors the teacher’s architecture but appends a few learnable prompts to the `output` tokens. This self-distillation strategy effectively transfers accumulated knowledge from the

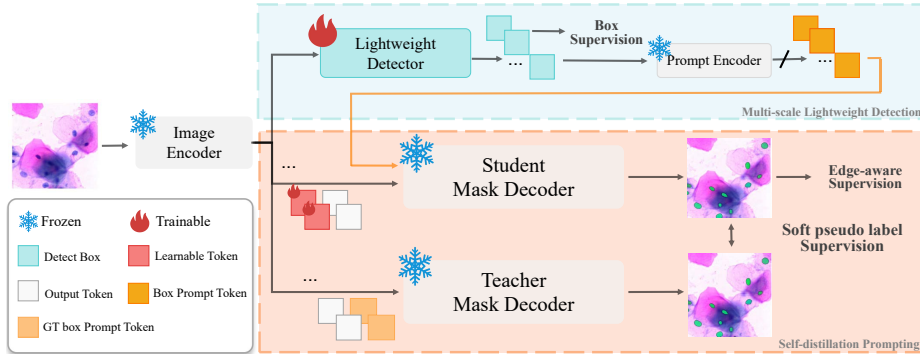


Fig. 2. The overview of our proposed method, including Multi-scale Light-weight Detection module and Self-distillation Prompting strategy.

teacher network to the cervical cell nucleus segmentation task. Additionally, we introduce an Edge-aware Enhanced Loss based on contrastive learning to refine the segmentation boundary of the cell nucleus. We conducted comparative and generalization experiments on the publicly available CNSeg dataset [28] and the 2014 ISBI Challenge dataset [17] to validate our method’s performance. The results showcase the effectiveness and robustness of DES-SAM.

2 Method

An overview of our DES-SAM framework is shown in Fig. 2. We extend SAM [8] to the box-supervised nuclear segmentation task by first introducing an additional Multi-scale Lightweight Detection module to automatically generate the box prompts, thereby eliminating SAM’s reliance on additional input prompts. We then propose a self-distillation prompting strategy and an Edge-aware Enhanced Loss to perform PEFT of SAM, enhancing its capability for nucleus segmentation in cervical cells. We will elaborate on each component in the subsequent subsections.

2.1 Multi-scale Lightweight Detection

As an interactive segmentation method, SAM [8] often requires additional inputs such as points or box prompts in addition to the input image. Inspired by the Regional Proposal Network in Faster R-CNN [21] to automatically generate proposals, we extend SAM to our box-supervised nuclear segmentation task by integrating an additional Multi-scale Lightweight Detection module into SAM to automatically generate the box prompts. Our lightweight detection module is borrowed from ViTDet [15]. Specifically, we use the image encoder from SAM as the backbone network, with parameters frozen, obtaining richer image features while avoiding a large number of learnable parameters. Based on the single-scale

extracted features, we build a simple multi-scale feature pyramid using strided convolutions and deconvolutions to address the multiscale variations of the target cell nucleus, replacing the feature pyramid network in traditional object detection frameworks. The frozen backbone helps our DES-SAM transfer the strong segmentation capability of SAM [8] to downstream detection tasks using simple biases. We then simply attach a detection head based on Faster R-CNN [21] and the entire detection module is supervised by the detection loss

$$\mathcal{L}_{dect} = \mathcal{L}_1(Y_B^*, Y_B) + \mathcal{L}_{BCE}(Y_C^*, Y_C), \quad (1)$$

where $\mathcal{L}_1(\cdot, \cdot)$ is the mean absolute error loss used to supervise the prediction boxes Y_B^* with the ground truth bounding boxes Y_B , while $\mathcal{L}_{BCE}(\cdot, \cdot)$ denotes the binary cross-entropy loss used to supervise the class scores Y_C^* with labels Y_C .

2.2 Self-distillation Prompting Strategy

We propose further fine-tuning SAM’s mask head to enhance its performance in our cervical cell nuclear segmentation task. However, this task often encounters limitations of insufficient data and a single data source, with excessive fine-tuning resulting in suboptimal performance, particularly when tested on datasets with domain shifts. To address this issue, we employ a simple yet effective PEFT method, namely prompt tuning [6,14], and propose a self-distillation prompting strategy that retains as much of the knowledge accumulated in SAM as possible, to finetune the SAM’s mask head for the cervical cell nucleus segmentation task.

Concretely, unlike traditional knowledge distillation, as shown in Fig. 2, our teacher-student network shares the frozen backbone and even the frozen mask decoder that takes both image features and box prompt tokens concatenated with `output` tokens as inputs. The box prompt tokens are obtained by feeding the box into SAM’s prompt encoder. Unlike the teacher network, which directly uses the original SAM’s `output` tokens and ground-truth box prompt tokens, the student network takes the predicted box prompt tokens obtained by feeding the output box of our lightweight detection module into SAM’s prompt encoder and appends a few learnable prompts to the `output` tokens. The learnable prompts are highlighted as visual prompts [6], which are supervised by the following distillation loss

$$\mathcal{L}_{distill} = \mathcal{L}_{BCE}(Y_{np}^t, Y_{np}^s) + \mathcal{L}_{DICE}(Y_{np}^t, Y_{np}^s), \quad (2)$$

where $\mathcal{L}_{DICE}(\cdot, \cdot)$ denotes the dice loss and Y_{np}^t and Y_{np}^s are the predicted results of teacher and student respectively, which are obtained by

$$\begin{aligned} Y_{np}^t &= D(F_I, P(Y_B), T), \\ Y_{np}^s &= D(F_I, P(Y_B^*), T \oplus \tilde{T}), \end{aligned} \quad (3)$$

where $D(\cdot, \cdot, \cdot)$ is SAM’s mask decoder, which takes three kinds of inputs including the image feature map F_I , prompt tokens output by SAM’s prompt encoder $P(\cdot)$, SAM’s `output` tokens T along with our additional learnable visual prompt \tilde{T} . The symbol \oplus denotes the concatenation operation.

2.3 Edge-aware Enhanced Loss

We further improve the segmentation performance of our DES-SAM by considering that object boundaries typically exist within regions of local color variations in images. Inspired by [13,25], we propose the following Edge-aware Enhanced Loss, which is a weighted combination of local pairwise loss \mathcal{L}_{lp} [25] and global pairwise loss \mathcal{L}_{gp} [13]

$$\mathcal{L}_{edge} = \alpha\mathcal{L}_{lp} + \beta\mathcal{L}_{gp}, \quad (4)$$

where α and β are the weight hyperparameters. The local pairwise loss \mathcal{L}_{lp} is used to facilitate the alignment between the predicted mask and the nucleus edge. Unlike BoxSnake [25], which rasterizes polygons predicted from the model to obtain masks, our method associates proposals and masks through box prompts. Furthermore, as there may be color variations in local regions of the image, training with only the local pairwise loss \mathcal{L}_{lp} may produce unexpected segmentation boundaries. Therefore, we also utilize the global pairwise loss \mathcal{L}_{gp} [13] to reduce the effect of local noise.

In summary, the overall loss can be represented by

$$\mathcal{L}_{total} = \lambda_{dect}\mathcal{L}_{dect} + \lambda_{distill}\mathcal{L}_{distill} + \lambda_{pairwise}\mathcal{L}_{edge}, \quad (5)$$

where $\lambda_{distill}$, λ_{dect} and $\lambda_{pairwise}$ are modulation weights for each loss term.

3 Experiments

3.1 Experimental Setting

Datasets. We conduct experiments on the CNSeg dataset [28], which contains 124,353 annotated cell nuclei collected from 1,530 patients, making it the largest publicly available dataset in the field of cell nucleus segmentation to our best knowledge. The CNSeg dataset [28] is divided into three subsets: PatchSeg, ClusterSeg, and DomainSeg. The PatchSeg subset includes small patch images cropped from whole-slide images under complex conditions. The ClusterSeg subset is further divided into three subsets: Sample, Normal, and Difficult, consisting of nuclear images with overlapping clusters. The DomainSeg subset includes new domain images, which are divided into the TargetA and TargetB subsets. Since the PatchSeg and ClusterSeg subsets provide training and testing splits, following [28], we use these subsets for training and comparative testing. Furthermore, we use the model trained on ClusterSeg to test generalization on DomainSeg.

In addition to the CNSeg dataset, we also conduct generalization experiments on the 2014 ISBI Challenge dataset [17], which consists of 16 EDF real cervical cytology images and 945 synthetic images. Following the public challenge settings, we perform the generalization evaluation on the 2014ISBI test set consisting of 900 synthetic images.

Table 1. Comparison with the weakly-supervised methods on the PatchSeg Test, ClusterSeg Test. The ‘Sup.’ column indicates the supervision type: ‘P’ for point, ‘S’ for scribble, and ‘B’ for bounding box.

Model	Sup.	PatchSeg			ClusterSeg			Params
		DICE	AJI	PQ	DICE	AJI	PQ	
WSPP [20]	P	62.99	42.47	48.64	68.37	49.51	48.11	70.59M
Scribble2Label [12]	S	77.62	59.26	53.71	76.78	48.21	47.13	32.25M
BoxInst [24]	B	82.23	67.73	66.43	80.61	65.74	60.92	32.28M
DiscoBox [11]	B	82.35	67.82	65.29	81.07	67.01	61.10	43.87M
WNS [16]	B	83.60	70.23	68.05	75.38	59.92	61.19	41.72M
BoxLevelSet [13]	B	76.92	60.35	53.82	83.80	71.04	67.57	33.98M
MAL [10]	B	78.37	61.65	57.16	80.39	65.72	61.93	43.87M
BoxSnake [25]	B	84.65	71.40	70.16	84.52	71.90	70.59	45.90M
SAM-Base [8]	-	41.09	12.99	30.47	57.92	39.04	42.32	-
DES-SAM (ours)	B	83.37	69.99	69.85	83.58	70.02	70.16	18.34M

Evaluation Metrics. To comprehensively evaluate the performance of our DES-SAM for nuclei segmentation, we use three widely-used evaluation metrics [3,28], including the Dice Coefficient (DICE), Aggregated Jaccard Index (AJI), and Panoptic Quality (PQ).

Implementation Details. We use SAM-Base [8] to initialize the image encoder and mask head of the network, and following VitDet [15], we adopt the Faster R-CNN detection head [21], performing detection on feature maps at four different scales. For loss weights, we set α , β , $\lambda_{distill}$, λ_{dect} and $\lambda_{pairwise}$ to 1.0, 0.001, 1.0, 1.0, 1.0 respectively. Our experiments start with an initial learning rate of 2.0×10^{-3} , using SGD optimizer for 50 epochs. Unless otherwise specified, the reported results are averaged over three trials.

3.2 Comparison to SOTA Methods

Comparison to Weakly-supervised Methods. We first conduct comparative training and testing experiments on the PatchSeg and ClusterSeg subsets with eight weakly-supervised methods, including point-supervised WSPP [20], scribble-supervised Scribble2Label [12], and six box-supervised methods: BoxInst [24], DiscoBox [11], WNS [16], BoxLevelSet [13], MAL [10], and BoxSnake [25], along with the unsupervised SAM-Base method [8]. Table 1 reports the comparative segmentation performance along with the trainable parameters (Params). As we can see, our DES-SAM exceeds MAL [10] and DiscoBox [11], which excel in natural image settings, and achieves performance on par with the SOTA BoxSnake [25]. It is noteworthy that our approach significantly reduces the number of trainable parameters, accounting for only 40% of those used in the BoxSnake [25]. This indicates that our model achieves an excellent trade-off between performance and parameter efficiency for cervical cell nuclear segmentation.

We further conduct experiments to compare the generalization performance of each method. Following [28], we directly perform comparative generalization testing of the model trained on ClusterSeg on the TargetA, TargetB, and

Table 2. Generalization test on TargetA Test, TargetB Test and 2014ISBI Test.

Model	Sup.	TargetA			TargetB			2014ISBI		
		DICE	AJI	PQ	DICE	AJI	PQ	DICE	AJI	PQ
WSPP [20]	P	52.80	32.15	31.76	62.24	39.24	42.52	79.88	66.48	64.91
Scribble2Label [12]	S	54.43	23.80	23.24	68.13	41.47	42.83	68.89	49.27	43.79
BoxInst [24]	B	72.23	52.73	47.68	72.36	54.33	52.82	34.45	18.10	26.02
DiscoBox [11]	B	72.40	53.84	49.30	74.97	57.84	52.21	55.25	35.81	43.49
WNS [16]	B	74.68	57.66	58.07	79.14	63.12	65.39	61.23	40.07	42.60
MAL [10]	B	71.29	51.94	49.37	73.67	55.07	50.05	42.69	20.31	32.80
BoxSnake [25]	B	74.60	56.82	56.92	77.37	63.09	64.09	43.01	22.09	25.35
SAM-Base [8]	-	50.66	32.67	33.46	54.88	33.41	42.61	30.67	8.43	25.86
DES-SAM (ours)	B	76.15	59.07	60.45	76.58	59.23	63.63	83.98	71.37	80.81

2014ISBI Test datasets. Table 2 lists the detailed test results, demonstrating that our method achieves overall optimal performance on various metrics across the three subsets. Specifically, on the TargetA subset, our method improved overall performance by 1.47% in DICE, 1.41% in AJI, and 2.38% in PQ compared to the best existing methods, while on the TargetB subset, it achieved results nearly equivalent to the current best performance. Similarly, on the 2014ISBI dataset, our method improved performance by 4.10% in DICE, 4.89% in AJI, and 15.90% in PQ. This shows the strong generalization ability of our DES-SAM and demonstrates the effectiveness of the self-distillation prompting strategy in improving nucleus segmentation performance on downstream tasks while inheriting the powerful generalization ability of SAM.

Table 3. Comparison of model performance with fully supervised methods on the CNSeg dataset. ‘S’ and ‘I’ stand for semantic segmentation and instance segmentation methods, respectively.

Model	Task	ClusterSeg		Difficult		Normal	
		AJI	PQ	AJI	PQ	AJI	PQ
Blend Mask [1]	I	68.34	69.82	62.67	65.53	70.00	71.08
CondInst [23]	I	67.72	68.07	62.52	64.37	68.62	69.15
BC-Net [7]	I	67.65	69.39	60.63	64.29	69.71	70.88
Mask RCNN [5]	I	69.47	70.57	64.29	66.31	71.00	71.82
U-Net [22]	S	59.53	58.62	51.15	50.24	61.99	61.08
U-Net++ [29]	S	61.91	61.39	54.64	53.91	64.04	63.58
Attention U-Net [18]	S	61.22	61.27	53.74	53.16	63.41	63.65
CE-Net [4]	S	60.48	60.39	51.46	51.72	63.13	62.94
Joint segmentation [19]	S	56.16	56.03	45.83	45.95	59.19	58.98
NucleiSegNet [9]	S	60.93	60.24	54.03	52.15	62.95	62.62
AL-Net [27]	S	69.58	69.09	63.66	63.09	71.32	70.85
DES-SAM (ours)	S	70.02	70.16	65.23	66.39	68.47	68.52

Comparison to Fully-supervised Methods. We also conduct comparative experiments between our method and various pixel-level fully-supervised methods for cervical cell nucleus segmentation. We train and test on the ClusterSeg dataset, comparing the performance on Difficult and Normal datasets in detail, as shown in Table 3. Surprisingly, the overall performance of our DES-SAM surpasses most classic fully-supervised segmentation networks like U-Net [22], U-

Table 4. Ablation study results on the CNSeg dataset.

prompt	mask_loss	pairwise_loss	PatchSeg			ClusterSeg		
			DICE	AJI	PQ	DICE	AJI	PQ
			83.13	69.75	69.71	83.55	69.98	70.16
✓	✓		83.18	69.88	69.78	83.48	69.98	69.55
✓	✓	✓	83.37	69.99	69.85	83.58	70.02	70.16

Table 5. Generalization test of ablation study on the CNSeg dataset.

prompt	mask_loss	pairwise_loss	TargetA			TargetB		
			DICE	AJI	PQ	DICE	AJI	PQ
			76.30	59.31	60.98	76.31	58.62	63.93
✓	✓		76.39	59.50	60.17	76.57	59.18	63.43
✓	✓	✓	76.15	59.07	60.45	76.58	59.23	63.63

Net++ [29] and CondInst [23]. The experimental results confirm that our model can achieve performance comparable to fully-supervised segmentation methods.

3.3 Ablation studies

We then conduct several ablations on the CNSeg dataset to evaluate the effectiveness of key settings in our DES-SAM model. The model is trained and tested on the PatchSeg and ClusterSeg subsets, with additional generalization testing on the DomainSeg subset.

Self-distillation Prompting. We conduct ablation experiments to evaluate the effectiveness of the self-distillation prompting strategy. As shown in Tables 4 and 5, we observed performance improvements in PatchSeg and ClusterSeg datasets when the model incorporated this strategy, along with enhanced generalization capability. Additionally, the supplementary materials list the model’s generalization performance when trained on PatchSeg, which also showed significant improvement. These results indicate that the self-distillation prompting strategy enables the model to achieve performance gains with minimal parameter training while maintaining the generalization ability of SAM.

Edge-aware Enhanced Loss. We also conduct ablation experiments to evaluate the effectiveness of the Edge-aware Enhanced Loss. Our model showed improvements in both performance and generalization on PatchSeg, as indicated in the second and third rows of Tables 4 and 5. Additionally, generalization performance testing on the model trained on PatchSeg, presented in the supplementary materials, demonstrated significant enhancement with the inclusion of edge-aware enhancement loss. These results collectively suggest that DES-SAM learned more about nuclear boundaries and universally applicable nuclear features through edge-aware enhancement loss supervision.

4 Conclusion

In this paper, we propose DES-SAM, a detection-based, box-supervised method for cervical cell nuclei segmentation utilizing knowledge distillation. The detection module of DES-SAM leverages the powerful feature extraction capability of

SAM to automatically generate cervical nucleus region proposals, thereby eliminating the need for additional input prompts for SAM. Additionally, by introducing a self-distillation prompting strategy, DES-SAM extends SAM’s generalization ability to the box-supervised nucleus segmentation task, addressing the challenges of insufficiently annotated and limited labeled cell nuclei in cervical datasets. Furthermore, we introduce an Edge-aware Enhancement Loss to improve segmentation performance at nucleus boundaries. Our DES-SAM achieves an excellent balance between performance and generalization ability with very few training parameters. While our study shows promising results, future research is needed to explore more effective prompting strategies and extend them to other types of cell nucleus segmentation.

Acknowledgments. This manuscript was supported in part by the National Key Research and Development Program of China under Grant 2021YFF1201202, the Natural Science Foundation of Hunan Province under Grant 2024JJ5444 and 2023JJ30699, and the Key Research and Development Program of Hunan Province under Grant 2023SK2029. The authors wish to acknowledge High Performance Computing Center of Central South University for computational resources.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Chen, H., Sun, K., Tian, Z., Shen, C., Huang, Y., Yan, Y.: BlendMask: Top-down meets bottom-up for instance segmentation. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 8570–8578. IEEE (2020)
2. Doan, T.N., Song, B., Vuong, T.T., Kim, K., Kwak, J.T.: SONNET: A self-guided ordinal regression neural network for segmentation and classification of nuclei in large-scale multi-tissue histology images. *IEEE Journal of Biomedical and Health Informatics* **26**(7), 3218–3228 (2022)
3. Graham, S., Vu, Q.D., Raza, S.E.A., Azam, A., Tsang, Y.W., Kwak, J.T., Rajpoot, N.: Hover-Net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images. *Medical Image Analysis* **58**, 101563 (2019)
4. Gu, Z., Cheng, J., Fu, H., Zhou, K., Hao, H., Zhao, Y., Zhang, T., Gao, S., Liu, J.: CE-Net: Context encoder network for 2D medical image segmentation. *IEEE Transactions on Medical Imaging* **38**(10), 2281–2292 (2019)
5. He, K., Gkioxari, G., Dollár, P., Girshick, R.B.: Mask R-CNN. In: IEEE International Conference on Computer Vision (ICCV). pp. 2980–2988. IEEE (2017)
6. Jia, M., Tang, L., Chen, B.C., Cardie, C., Belongie, S., Hariharan, B., Lim, S.N.: Visual prompt tuning. In: European Conference on Computer Vision (ECCV). pp. 709–727. Springer (2022)
7. Ke, L., Tai, Y., Tang, C.: Deep occlusion-aware instance segmentation with overlapping bilayers. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 4019–4028 (2021)
8. Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.Y., et al.: Segment anything. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 4015–4026 (2023)

9. Lal, S., Das, D., Alabhya, K., Kanfade, A., Kumar, A., Kini, J.: NucleiSegNet: Robust deep learning architecture for the nuclei segmentation of liver cancer histopathology images. *Computers in Biology and Medicine* **128**, 104075 (2021)
10. Lan, S., Yang, X., Yu, Z., Wu, Z., Alvarez, J.M., Anandkumar, A.: Vision transformers are good mask auto-labelers. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. pp. 23745–23755 (2023)
11. Lan, S., Yu, Z., Choy, C.B., Radhakrishnan, S., Liu, G., Zhu, Y., Davis, L.S., Anandkumar, A.: DiscoBox: Weakly supervised instance segmentation and semantic correspondence from box supervision. In: *IEEE/CVF International Conference on Computer Vision (ICCV)*. pp. 3386–3396. IEEE (2021)
12. Lee, H., Jeong, W.K.: Scribble2Label: Scribble-supervised cell segmentation via self-generating pseudo-labels with consistency. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. pp. 14–23. Springer (2020)
13. Li, W., Liu, W., Zhu, J., Cui, M., Hua, X.S., Zhang, L.: Box-supervised instance segmentation with level set evolution. In: *European Conference on Computer Vision (ECCV)*. pp. 1–18. Springer (2022)
14. Li, X.L., Liang, P.: Prefix-Tuning: Optimizing continuous prompts for generation. In: *Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL)*. pp. 4582–4597. Association for Computational Linguistics (2021)
15. Li, Y., Mao, H., Girshick, R., He, K.: Exploring plain vision transformer backbones for object detection. In: *European Conference on Computer Vision (ECCV)*. pp. 280–296. Springer (2022)
16. Liang, Y., Yin, Z., Liu, H., Zeng, H., Wang, J., Liu, J., Che, N.: Weakly supervised deep nuclei segmentation with sparsely annotated bounding boxes for DNA image cytometry. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* **20**(1), 785–795 (2023)
17. Lu, Z., Carneiro, G., Bradley, A.P.: An improved joint optimization of multiple level set functions for the segmentation of overlapping cervical cells. *IEEE Transactions on Image Processing* **24**(4), 1261–1272 (2015)
18. Oktay, O., Schlemper, J., Folgoc, L.L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N.Y., Kainz, B., Glocker, B., Rueckert, D.: Attention U-Net: Learning where to look for the pancreas. In: *Medical Imaging with Deep Learning (MIDL)* (2018)
19. Qu, H., Riedlinger, G., Wu, P., Huang, Q., Yi, J., De, S., Metaxas, D.: Joint segmentation and fine-grained classification of nuclei in histopathology images. In: *IEEE International Symposium on Biomedical Imaging (ISBI)*. pp. 900–904. IEEE (2019)
20. Qu, H., Wu, P., Huang, Q., Yi, J., Yan, Z., Li, K., Riedlinger, G.M., De, S., Zhang, S., Metaxas, D.N.: Weakly supervised deep nuclei segmentation using partial points annotation in histopathology images. *IEEE Transactions on Medical Imaging* **39**(11), 3655–3666 (2020)
21. Ren, S., He, K., Girshick, R., Sun, J.: Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **39**(6), 1137–1149 (2016)
22. 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 (MICCAI)*. pp. 234–241. Springer (2015)
23. Tian, Z., Shen, C., Chen, H.: Conditional convolutions for instance segmentation. In: *European Conference on Computer Vision (ECCV)*. pp. 282–298. Springer (2020)

24. Tian, Z., Shen, C., Wang, X., Chen, H.: BoxInst: High-performance instance segmentation with box annotations. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 5443–5452 (2021)
25. Yang, R., Song, L., Ge, Y., Li, X.: BoxSnake: Polygonal instance segmentation with box supervision. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 766–776 (2023)
26. Zhang, C., Liang, Y., Liu, Q.: CCBox: Improving box-supervised nuclei segmentation with consistency constraint. In: IEEE International Conference on Bioinformatics and Biomedicine (BIBM). pp. 2412–2415. IEEE (2023)
27. Zhao, J., He, Y.J., Zhao, S.Q., Huang, J.J., Zuo, W.M.: AL-Net: Attention learning network based on multi-task learning for cervical nucleus segmentation. IEEE Journal of Biomedical and Health Informatics **26**(6), 2693–2702 (2021)
28. Zhao, J., He, Y.j., Zhou, S.H., Qin, J., Xie, Y.n.: CNSeg: A dataset for cervical nuclear segmentation. Computer Methods and Programs in Biomedicine **241**, 107732 (2023)
29. Zhou, Z., Siddiquee, M.M.R., Tajbakhsh, N., Liang, J.: UNet++: Redesigning skip connections to exploit multiscale features in image segmentation. IEEE Transactions on Medical Imaging **39**(6), 1856–1867 (2019)