Co-Seg: Mutual Prompt-Guided Collaborative Learning for Tissue and Nuclei Segmentation

Qing Xu[1,2], Wenting Duan[1], and Zhen Chen[3]([✉])

[1] University of Lincoln, Brayford Pool, UK
[2] University of Nottingham, Nottinghamshire, UK
[3] Yale University, New Haven CT 06510, USA
zchen.francis@gmail.com

Abstract. Histopathology image analysis is critical yet challenged by the demand of segmenting tissue regions and nuclei instances for tumor microenvironment and cellular morphology analysis. Existing studies focused on tissue semantic segmentation or nuclei instance segmentation separately, but ignored the inherent relationship between these two tasks, resulting in insufficient histopathology understanding. To address this issue, we propose a Co-Seg framework for collaborative tissue and nuclei segmentation. Specifically, we introduce a novel co-segmentation paradigm, allowing tissue and nuclei segmentation tasks to mutually enhance each other. To this end, we first devise a region-aware prompt encoder (RP-Encoder) to provide high-quality semantic and instance region prompts as prior constraints. Moreover, we design a mutual prompt mask decoder (MP-Decoder) that leverages cross-guidance to strengthen the contextual consistency of both tasks, collaboratively computing semantic and instance segmentation masks. Extensive experiments on the PUMA dataset demonstrate that the proposed Co-Seg surpasses state-of-the-arts in the semantic, instance and panoptic segmentation of tumor tissues and nuclei instances. The source code is available at `https://github.com/xq141839/Co-Seg`.

Keywords: Collaborative Learning · Tissue Segmentation · Nuclei Segmentation · Mutual Prompt.

1 Introduction

Medical image segmentation plays a crucial role in clinical applications and has received extensive attention in the research [7, 29, 10, 5]. Particularly, histopathology image analysis is challenged by the delineation of tissue regions and further separation of individual nuclei within the regions to assess tissue subtypes and tumor grading [36]. These requirements have led to two significant segmentation tasks in the field, *i.e.*, tissue semantic segmentation and nuclei instance segmentation.

For medical semantic segmentation, existing methods [26, 35, 24, 34] rely on hierarchal encoding and decoding to enhance multi-scale awareness for precise
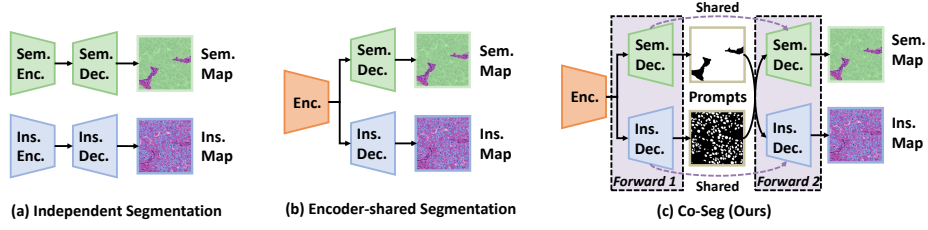
Fig. 1. Comparison of our Co-Seg and existing works in tissue and nuclei segmentation. (a) Two independent networks for tissue and nuclei segmentation. (b) A shared image encoder but separated task decoders for tissue and nuclei segmentation. (c) Our Co-Seg leverages mutual prompts for collaborative tissue and nuclei segmentation.

mask generation. In particular, ViT-based [1] and Mamba-based [17] architectures model long-range dependencies to gain global contexts of target regions. For instance segmentation, a series of works [6, 30, 2, 10] adopted different distance proxy maps to improve the boundary understanding of instances. Recent SAM-based methods [3, 21, 29, 5, 29, 4, 33, 16, 19] have revealed impressive medical semantic and instance segmentation performance by manually providing corresponding task prompts.

Despite the advancements, existing methods [1, 31, 18, 32, 8] usually focus on the optimization of tissue semantic or nuclei instance segmentation tasks, which can only be supervised by the isolated knowledge. In fact, both tasks are highly correlated as they aim to achieve the adequate perception and understanding of histopathology images. For example, accurately identifying nuclei provides valuable cues for understanding the underlying tissue structures, while tissue segmentation can aid in localizing nuclei. Therefore, this strong interdependence motivates us to develop a collaborative approach that integrates tissue and nuclei segmentation to advance state-of-the-art histopathology image analysis.

To overcome this bottleneck, we propose **Co-Seg**, a collaborative tissue and nuclei segmentation framework that allows semantic and instance segmentation tasks to mutually enhance each other. As illustrated in Fig. 1(c), the Co-Seg, based on a co-segmentation paradigm, improves segmentation mask quality by capturing contextual dependencies between the two tasks. Specifically, we first devise the region-aware prompt encoder (RP-Encoder) to extract high-quality semantic and instance prompts from the target region of both tasks, guiding segmentation decoding. We further introduce a mutual prompt mask decoder (MP-Decoder), which jointly computes semantic and instance maps by leveraging cross-guidance. This collaborative learning approach achieves contextual consistency, reducing prediction errors in both tasks. Experimental results on melanoma tissue and nuclei segmentation tasks demonstrate that our Co-Seg achieves remarkable performance over state-of-the-art segmentation methods.
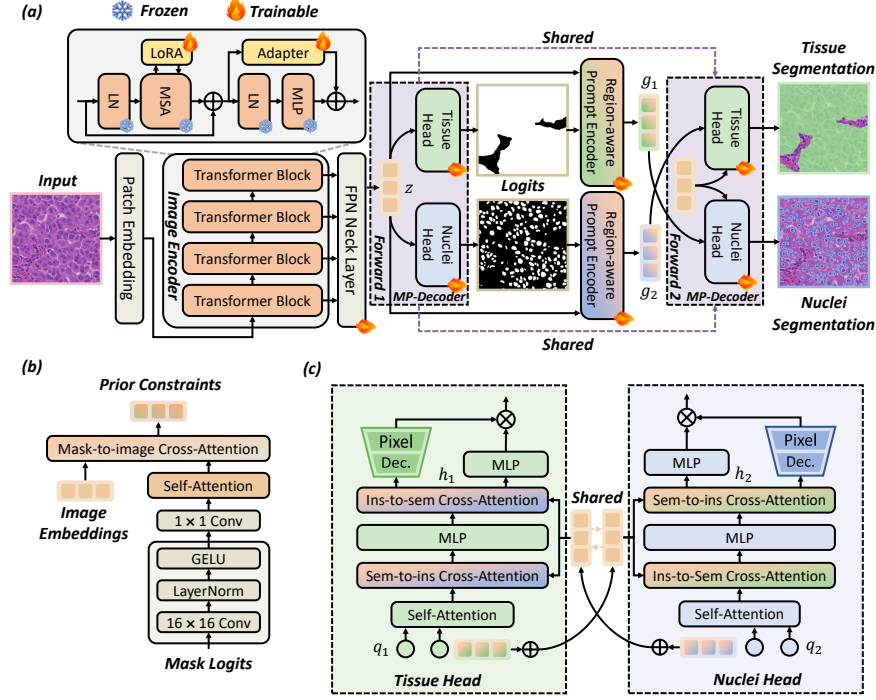
Fig. 2. (a) The overview of the proposed Co-Seg framework for collaborative tissue and nuclei segmentation, consisting of (b) RP-Encoder and (c) MP-Decoder. Co-Seg fully exploits complementary information by leveraging mutual prompts.

2 Methodology

As elaborated in Fig. 2, we devise the Co-Seg framework with the co-segmentation paradigm, achieving the mutual optimization between tissue and nuclei segmentation in medical imaging. To accomplish this, we design the RP-Encoder to provide prior constraints to guide both tasks, and the MP-Decoder to cooperatively generate semantic and instance maps through bidirectional information interactions. This integrated approach enables the two tasks to enhance each other, leveraging mutual prompts for improved segmentation accuracy.

2.1 Co-Segmentation Paradigm

Existing tissue [9, 13, 22] and nuclei [23, 20, 3] segmentation methods decouple the parameter space of semantic and instance segmentation, disrupting their interdependencies, degrading the performance. To address this issue, we propose the co-segmentation paradigm that leverages closed-loop bidirectional interaction to realize dual-task collaborative optimization. Specifically, our goal is to

train a model $f_\Theta : x \to \{y_1, y_2\}$, where $\Theta = \{\theta_1, \theta_2\}$ represents the parameters for semantic segmentation θ_1 and instance segmentation θ_2, with y_1 and y_2 as the corresponding segmentation masks. To model the interdependence between the two tasks, we define their joint prediction of y_1 and y_2 as:

$$p(y_1, y_2|x, \theta_1, \theta_2) = p(y_1|x, \theta_1) \cdot p(y_2|y_1, x, \theta_2) = p(y_2|x, \theta_2) \cdot p(y_1|y_2, x, \theta_1). \quad (1)$$

Note that Eq. (1) reveals the mutual dependency between semantic and instance segmentation through intertwined conditional probabilities. It states that $p(y_1, y_2|x, \theta_1, \theta_2)$, the joint probability of obtaining both segmentations given the image and model parameters, can be decomposed in two symmetrical ways:

- $p(y_1|x, \theta_1) \cdot p(y_2|y_1, x, \theta_2)$ suggests that once the semantic segmentation y_1 is known, it directly influences the prediction of the instance segmentation y_2.
- Similarly, $p(y_2|x, \theta_2) \cdot p(y_1|y_2, x, \theta_1)$ implies that knowing the instance segmentation y_2 affects the outcome of the semantic segmentation y_1.

Both expressions are mathematically equivalent because they describe the same joint probability from different perspectives, highlighting that each segmentation task provides crucial context that enhances the accuracy of the other. This reciprocal relationship underscores the need for a collaborative approach, where both segmentation tasks are optimized together rather than in isolation, leveraging the full context available from each task to improve overall segmentation results. Translating this relationship to the gradient calculations for θ_1 and θ_2, we have:

$$\nabla_{\theta_i} \ell_i = -\nabla_{\theta_i} \mathbb{E}\left[\log p(y_i|x, \theta_i)\right] - \nabla_{\theta_i} \mathbb{E}\left[\log p(y_j|y_i, x, \theta_j)\right] + \nabla_{\theta_i} I(y_1; y_2|x), \quad (2)$$

for $i, j \in \{1, 2\}, i \neq j$ where $\nabla_\theta I(y_1; y_2|x) \approx \mathbb{E}\left[\nabla_\theta \log \frac{p(y_1, y_2|x)}{p(y_1|x)p(y_2|x)}\right]$. This term captures the mutual information between y_1 and y_2, ensuring that both segmentation tasks benefit from shared feature learning. The first term of Eq. (2) represents the main segmentation loss gradient, while the second term incorporates the implicit gradient of task interdependencies, modeling the complementarity between θ_1 and θ_2. This leads to the following optimization update rule:

$$\theta_1 \leftarrow \theta_1 - \eta \nabla_{\theta_1} \ell_1, \quad \theta_2 \leftarrow \theta_2 - \eta \nabla_{\theta_2} \ell_2, \quad (3)$$

where η is the learning rate. In this way, the co-segmentation paradigm breaks the barrier of isolating gradient flows from each other in multi-task learning, enabling our Co-Seg framework to achieve the bidirectional optimization of tissue and nuclei segmentation.

2.2   Region-aware Prompt Encoder

The proposed co-segmentation paradigm requires establishing implicit interdependencies between tissue and nuclei segmentation tasks. To achieve this, we introduce the RP-Encoder that leverages mask prompts to capture task-specific

target regions as prior constraints, as presented in Fig. 2(b). Specifically, we first employ several convolutional blocks to extract features m_i from mask logits μ_i. After that, the RP-Encoder performs self-attention, followed by cross-attention with shared image embeddings z to generate a set of dense prompts g_i, as follows:

$$m_i = \text{Conv}_2(\text{GELU}(\text{LN}(\text{Conv}_1(\mu_i)))), \tag{4}$$

$$g_i = \text{CrossAttention}(\text{SelfAttention}(m_i), z, z), \tag{5}$$

where Conv_1 denotes a 16×16 convolutions for downsampling, Conv_2 is a 1×1 convolution for channel expansion, $\text{LN}(\cdot)$ is LayerNorm. These operations encode the structural context of mask logits. On this basis, we apply RP-Encoder to semantic and instance mask logits, effectively obtaining region-aware prompts: g_1 and g_2 as prior constraints of Co-Seg for tissue and nuclei segmentation tasks.

2.3   Mutual Prompt Mask Decoder

Following the co-segmentation paradigm, we propose the MP-Decoder to enforce bidirectional interactions between tissue and nuclei segmentation, ensuring mutual refinement rather than treating them as isolated tasks, as illustrated in Fig. 2(c). Specifically, the MP-Decoder contains tissue and nuclei heads that leverage two sets of query embeddings: q_1 and q_2 to save the decoding information of both tasks. To leverage complementary effects between tasks, we first perform self-attention on each query, followed by cross-attention with the other task's prompts, as follows:

$$q_i' = \text{CrossAttention}(\text{SelfAttention}(q_i), z \oplus g_j, z \oplus g_j), \tag{6}$$

where $\oplus$ represents the element-wise addition. This is followed by applying MLP to ensure the refined query embeddings are expressive. The MP-Decoder further applies the reverse cross-attention to generate task-specific image embeddings h_1 and h_2, as follows:

$$h_i = \text{CrossAttention}(z \oplus g_j, q_i', q_i'). \tag{7}$$

On this basis, nuclei context prompts can enhance tissue segmentation consistency, while tissue boundary feedback provides prior constraints for nuclei discrimination. Following the standard SAM [15], each segmentation head adopts a pixel decoder to upsample the refined task-specific image embedding. Finally, our MP-Decoder generates tissue and nuclei segmentation predictions by performing the dot product between their upscaled image embeddings and corresponding task queries. Overall, the proposed MP-Decoder utilizes the bidirectional interactions of tissue and nuclei segmentation information to mutually improve the quality of segmentation masks of our Co-Seg framework.

Table 1. Comparison with state-of-the-arts on tissue semantic segmentation.

Methods	Dice	mIoU	HD
nnUnet [14]	90.90	84.48	236.03
Swin-Umamba [17]	89.67	82.75	265.43
EMCAD [24]	90.96	84.61	306.36
SAC [21]	91.19	85.01	295.57
H-SAM [5]	91.47	85.47	246.16
Co-Seg	**92.51**	**87.18**	**206.72**

Table 2. Comparison with state-of-the-arts on nuclei instance segmentation

Methods	F1	Prec.	Rec.	AJI
HoverNet [6]	75.53	70.53	82.80	64.40
CellPose [30]	75.30	71.12	81.79	65.15
CPPNet [2]	74.96	69.58	83.06	64.63
CellViT [10]	76.04	70.21	83.51	66.13
PromptNucSeg [29]	76.47	71.81	83.02	66.71
Co-Seg	**79.70**	**77.05**	**83.61**	**69.14**

2.4   Optimization Pipeline

To construct our Co-Seg framework, we first adopt Hirea ViT [27] as the shared image encoder for tissue and nuclei segmentation, ensuring the consistency of feature learning. In particular, we load the pre-trained SAM2-L [25] to initialize corresponding parameters and freeze these weights to preserve pre-trained knowledge. Additionally, we insert LoRA [12] and Adapter [11] into attention and FFN layers to achieve parameter-efficient fine-tuning from natural to histopathology domains. Based on the co-segmentation paradigm, Co-Seg includes two decoding forwards, where the tissue and nuclei heads of the MP-Decoder share parameters in these two stages. Specifically, the MP-Decoder leverages the image embedding generated from the image encoder to calculate binary segmentation masks y_1' and y_2' of both tasks without any prompts. Then, we deliver them to the RP-Encoder to provide tissue and nuclei prompts. In the second forward, they will be directly transferred back to the MP-Decoder as prior constraints. Finally, the MP-Decoder utilizes cross-guidance to co-generate tissue and nuclei segmentation masks. The training of Co-Seg consists of two parts: (1) the optimization of tissue g_1 and nuclei g_2 prior constraints and (2) the joint optimization of tissue and nuclei segmentation. The overall loss is formularized as:

$$\mathcal{L}_{\text{CoSeg}} = \underbrace{\lambda_1 \mathcal{L}_{\text{sem}}^{g_1}(\hat{y}_1', y_1') + \lambda_2 \mathcal{L}_{\text{ins}}^{g_2}(\hat{y}_2', y_2')}_{\text{Prior Constraint Loss}} + \underbrace{\mathcal{L}_{\text{sem}}(\hat{y}_1, y_1) + \mathcal{L}_{\text{ins}}(\hat{y}_2, y_2)}_{\text{Segmentation Loss}}, \quad (8)$$

where λ_1 and λ_2 are factors to adjust the contribution of each term. By optimizing $\mathcal{L}_{\text{CoSeg}}$, our Co-Seg achieves accurate tissue and nuclei segmentation with superior performance across different histopathology images.

3   Experiments

3.1   Experimental Setup

To validate the effectiveness of the proposed Co-Seg, we adopt the melanoma-specific histopathology dataset: PUMA [28] for tissue semantic and nuclei instance segmentation tasks. It contains 206 histopathology images of 1024×1024 resolution and adopts a common split of training, validation, and test sets as

Table 3. Comparison with state-of-the-arts on histopathology panoptic segmentation.

Methods	PQ_{tissue}	DQ_{tissue}	SQ_{tissue}	PQ_{nuclei}	DQ_{nuclei}	SQ_{nuclei}
nnUnet [14]	57.34	68.27	70.13	58.60	72.83	80.24
HoverNet [6]	58.95	67.25	69.92	60.91	74.96	81.04
CellViT [10]	60.55	71.32	69.17	62.91	76.59	81.95
H-SAM [5]	61.62	71.17	70.81	62.68	76.39	81.86
PromptNucSeg [29]	61.58	70.84	70.92	63.37	77.41	81.79
Co-Seg	**63.09**	**72.82**	**70.99**	**66.11**	**79.70**	**82.80**

Table 4. Ablation study of Co-Seg on PUMA tissue and nuclei Segmentation.

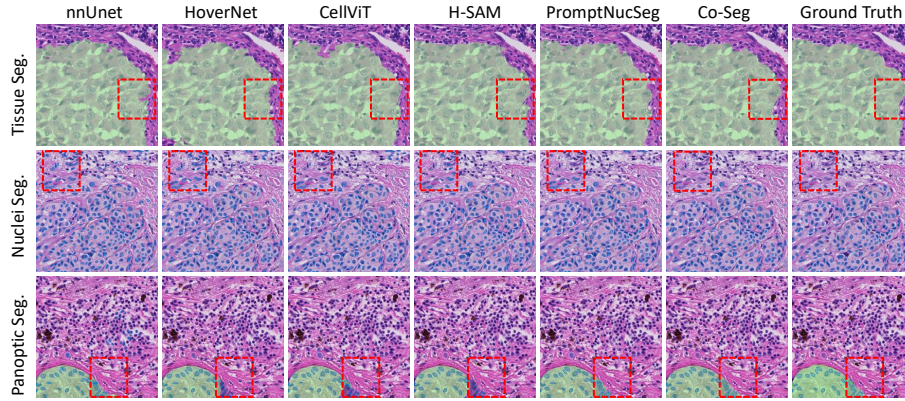| P	D	C	Tissue Semantic Seg.			Nuclei Instance Seg.			
			Dice	mIoU	HD	F1	Prec.	Rec.	AJI
			90.98	84.71	287.19	76.35	73.18	82.27	66.59
✓			91.23	85.18	261.37	77.01	73.25	83.01	67.08
✓	✓		91.76	86.13	246.56	78.19	74.91	82.54	67.93
✓	✓	✓	**92.51**	**87.18**	**206.72**	**79.70**	**77.05**	**83.61**	**69.14**

7:1:2. We perform all experiments on a NVIDIA A5000 GPU using PyTorch. For fair comparisons, we implement all tissue and nuclei segmentation methods with the same training settings and configurations. We utilize the pre-trained SAM ViT-H [15] structure as the image encoder of medical SAM baselines. We apply the optimizer using Adam with an initial learning rate of 1×10^{-4} and use the exponential decay strategy to adjust the learning rate with a factor of 0.98. The batch size and epochs are set to 16 and 300. We leverage the combination of cross-entropy loss and Dice loss to supervise the tissue segmentation. For nuclei segmentation, we follow the standard combination loss of Focal loss, Dice loss, MSE loss and MSGE loss [6, 10]. The loss coefficient λ_1 and λ_2 are set to 2 and 1. Our Co-Seg introduces additional 9.1% learnable parameters to the baseline (in Table 4) during fine-tuning.

3.2 Comparison with State-of-the-art Methods

Tissue Semantic Segmentation. We first evaluate the performance of all models in the tissue semantic segmentation. As illustrated in Table 1, we observe that classical semantic segmentation methods are inferior to medical SAMs, *e.g.,* H-SAM [5] surpasses EMCAD [24] with a 0.51% Dice increase. Remarkably, our Co-Seg achieves the best performance with a P-value < 0.005, Dice of 92.51%, and the lowest HD of 206.72, indicating precise tissue boundary localization.
Nuclei Instance Segmentation. The comparison of nuclei instance segmentation is shown in Table 2. Our Co-Seg framework reveals the overwhelming performance of four metrics. In particular, Co-Seg achieves a significant advantage over the second-best PromptNucSeg [29], *e.g.,* a P-value < 0.001, a F1-score increase of 3.23%, and an AJI increase of 2.43%.
Histopathology Panoptic Segmentation. We further comprehensively evaluate the performance of our Co-Seg with the panoptic segmentation protocol.

Fig. 3. Qualitative comparison of three evaluation protocols on tissue semantic and nuclei instance segmentation. Benefiting from the mutual reinforcement of semantic and instance segmentation tasks, our Co-Seg can delineate precise tissue regions and segment accurate nuclei, containing fewer false positives.

As presented in Table 3, our Co-Seg benefits from the novel co-segmentation paradigm, achieving the best performance of all six metrics, with a remarkable PQ increase of 1.51% and 2.74% in tissue and nuclei segmentation tasks.

3.3   Ablation Study

To investigate the effectiveness of our proposed co-segmentation paradigm C, RP-Encoder P and MP-Decoder D, we further conduct the comprehensive ablation study on the tissue semantic and nuclei instance segmentation of the PUMA dataset, as illustrated in Table 4. By removing the tailored modules from Co-Seg, we construct two independent Hiera ViT-based semantic and instance segmentation networks as the ablation baseline. By introducing the RP-Encoder, the performance achieves Dice and F1-score increases of 0.25% and 0.66% on tissue semantic and nuclei instance segmentation, respectively. Moreover, we investigate the effect of combined RP-Encoder and MP-Decoder, resulting in superior performance, with the Dice of 91.76% and the F1-score of 78.19%. Finally, we establish the framework using the proposed co-segmentation paradigm. The result proves that this design can significantly prompt the semantic and instance segmentation capabilities. In this way, these ablation experiments demonstrate the effectiveness of the RP-Encoder, MP-Decoder and co-segmentation paradigm in our Co-Seg framework.

4   Conclusion

In this work, we identify the potential relationship between tissue and nuclei segmentation in histopathology and propose a novel co-segmentation paradigm

to establish the Co-Seg framework for promoting nuclei and tissue segmentation mutually. It comprises two modules: the RP-Encoder aims to provide task-specific region prompts by perceiving the target area, and the MP-Decoder adopts cross-guidance to cooperatively generate tissue and nuclei segmentation masks. Extensive experiments on the melanoma tissue and nuclei instance segmentation dataset demonstrate that Co-Seg outperforms existing methods by remarkable margins.

Disclosure of Interests. The authors declare no competing interests.

References

1. Chen, J., Mei, J., Li, X., Lu, Y., Yu, Q., Wei, Q., Luo, X., Xie, Y., Adeli, E., Wang, Y., et al.: Transunet: Rethinking the u-net architecture design for medical image segmentation through the lens of transformers. Med. Image Anal. **97**, 103280 (2024)
2. Chen, S., Ding, C., Liu, M., Cheng, J., Tao, D.: Cpp-net: Context-aware polygon proposal network for nucleus segmentation. IEEE Trans. Image Process. **32**, 980–994 (2023)
3. Chen, Z., Xu, Q., Liu, X., Yuan, Y.: Un-sam: Domain-adaptive self-prompt segmentation for universal nuclei images. Med. Image Anal. p. 103607 (2025)
4. Chen, Z., Zhang, Z., Guo, W., Luo, X., Bai, L., Wu, J., Ren, H., Liu, H.: Asi-seg: Audio-driven surgical instrument segmentation with surgeon intention understanding. In: IROS. pp. 13773–13779. IEEE (2024)
5. Cheng, Z., Wei, Q., Zhu, H., Wang, Y., Qu, L., Shao, W., Zhou, Y.: Unleashing the potential of sam for medical adaptation via hierarchical decoding. In: CVPR. pp. 3511–3522 (2024)
6. 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. Med. Image Anal. **58**, 101563 (2019)
7. Guo, X., Chen, Z., Yuan, Y.: Complementary network with adaptive receptive fields for melanoma segmentation. In: ISBI. pp. 2010–2013. IEEE (2020)
8. Hao, Z., Quan, H., Lu, Y.: Emf-former: An efficient and memory-friendly transformer for medical image segmentation. In: MICCAI. pp. 231–241. Springer (2024)
9. He, A., Wang, K., Li, T., Du, C., Xia, S., Fu, H.: H2former: An efficient hierarchical hybrid transformer for medical image segmentation. IEEE Trans. Med. Imaging **42**(9), 2763–2775 (2023)
10. Hörst, F., Rempe, M., Heine, L., Seibold, C., Keyl, J., Baldini, G., Ugurel, S., Siveke, J., Grünwald, B., Egger, J., et al.: Cellvit: Vision transformers for precise cell segmentation and classification. Med. Image Anal. **94**, 103143 (2024)
11. Houlsby, N., Giurgiu, A., Jastrzebski, S., Morrone, B., De Laroussilhe, Q., Gesmundo, A., Attariyan, M., Gelly, S.: Parameter-efficient transfer learning for nlp. In: ICML. pp. 2790–2799. PMLR (2019)
12. Hu, E.J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., Chen, W.: LoRA: Low-rank adaptation of large language models. In: ICLR (2022)
13. Ibtehaz, N., Kihara, D.: Acc-unet: A completely convolutional unet model for the 2020s. In: MICCAI. pp. 692–702. Springer (2023)

14. Isensee, F., Jaeger, P.F., Kohl, S.A., Petersen, J., Maier-Hein, K.H.: nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. Nature methods **18**(2), 203–211 (2021)

15. Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.Y., Dollar, P., Girshick, R.: Segment anything. In: ICCV. pp. 4015–4026 (October 2023)

16. Li, C., Huang, Y., Li, W., Liu, H., Liu, X., Xu, Q., Chen, Z., Huang, Y., Yuan, Y.: Flaws can be applause: Unleashing potential of segmenting ambiguous objects in sam. In: NeurIPS. vol. 37, pp. 45578–45599 (2024)

17. Liu, J., Yang, H., Zhou, H.Y., Xi, Y., Yu, L., Li, C., Liang, Y., Shi, G., Yu, Y., Zhang, S., et al.: Swin-umamba: Mamba-based unet with imagenet-based pretraining. In: MICCAI. pp. 615–625. Springer (2024)

18. Liu, Y., Zhu, H., Liu, M., Yu, H., Chen, Z., Gao, J.: Rolling-unet: Revitalizing mlp's ability to efficiently extract long-distance dependencies for medical image segmentation. In: AAAI. pp. 3819–3827 (2024)

19. Lou, Z., Xu, Q., Jiang, Z., He, X., Li, C., Chen, Z., Wang, Y., He, M.M., Duan, W.: Nusegdg: Integration of heterogeneous space and gaussian kernel for domain-generalized nuclei segmentation. Knowledge-Based Systems p. 113641 (2025)

20. Meng, Z., Dong, J., Zhang, B., Li, S., Wu, R., Su, F., Wang, G., Guo, L., Zhao, Z.: Nusea: Nuclei segmentation with ellipse annotations. IEEE J. Biomed. Health Inform. (2024)

21. Na, S., Guo, Y., Jiang, F., Ma, H., Huang, J.: Segment any cell: A sam-based auto-prompting fine-tuning framework for nuclei segmentation. arXiv preprint arXiv:2401.13220 (2024)

22. Nam, J.H., Syazwany, N.S., Kim, S.J., Lee, S.C.: Modality-agnostic domain generalizable medical image segmentation by multi-frequency in multi-scale attention. In: CVPR. pp. 11480–11491 (2024)

23. Nam, S., Jeong, J., Luna, M., Chikontwe, P., Park, S.H.: Pronet: Point refinement using shape-guided offset map for nuclei instance segmentation. In: MICCAI. pp. 528–538. Springer (2023)

24. Rahman, M.M., Munir, M., Marculescu, R.: Emcad: Efficient multi-scale convolutional attention decoding for medical image segmentation. In: CVPR. pp. 11769–11779 (2024)

25. Ravi, N., Gabeur, V., Hu, Y.T., Hu, R., Ryali, C., Ma, T., Khedr, H., Rädle, R., Rolland, C., Gustafson, L., Mintun, E., Pan, J., Alwala, K.V., Carion, N., Wu, C.Y., Girshick, R., Dollar, P., Feichtenhofer, C.: SAM 2: Segment anything in images and videos. In: ICLR (2025)

26. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18. pp. 234–241. Springer (2015)

27. Ryali, C., Hu, Y.T., Bolya, D., Wei, C., Fan, H., Huang, P.Y., Aggarwal, V., Chowdhury, A., Poursaeed, O., Hoffman, J., et al.: Hiera: A hierarchical vision transformer without the bells-and-whistles. In: ICML. pp. 29441–29454. PMLR (2023)

28. Schuiveling, M., Liu, H., Eek, D., Breimer, G.E., Suijkerbuijk, K.P.M., Blokx, W.A.M., Veta, M.: A novel dataset for nuclei and tissue segmentation in melanoma with baseline nuclei segmentation and tissue segmentation benchmarks. GigaScience (2025)

29. Shui, Z., Zhang, Y., Yao, K., Zhu, C., Zheng, S., Li, J., Li, H., Sun, Y., Guo, R., Yang, L.: Unleashing the power of prompt-driven nucleus instance segmentation. In: ECCV. pp. 288–304. Springer (2024)
30. Stringer, C., Wang, T., Michaelos, M., Pachitariu, M.: Cellpose: a generalist algorithm for cellular segmentation. Nature methods **18**(1), 100–106 (2021)
31. Valanarasu, J.M.J., Patel, V.M.: Unext: Mlp-based rapid medical image segmentation network. In: MICCAI. pp. 23–33. Springer (2022)
32. Xing, Z., Ye, T., Yang, Y., Liu, G., Zhu, L.: Segmamba: Long-range sequential modeling mamba for 3d medical image segmentation. In: MICCAI. pp. 578–588. Springer (2024)
33. Xu, Q., Li, J., He, X., Liu, Z., Chen, Z., Duan, W., Li, C., He, M.M., Tesema, F.B., Cheah, W.P., et al.: Esp-medsam: Efficient self-prompting sam for universal domain-generalized medical image segmentation. arXiv preprint arXiv:2407.14153 (2024)
34. Xu, Q., Lou, Z., Li, C., He, X., Qu, R., Berhanu, T.F., Wang, Y., Duan, W., Chen, Z.: Hrmedseg: Unlocking high-resolution medical image segmentation via memory-efficient attention modeling. arXiv preprint arXiv:2504.06205 (2025)
35. Xu, Q., Ma, Z., Na, H., Duan, W.: Dcsau-net: A deeper and more compact split-attention u-net for medical image segmentation. Comput. Biol. Med. **154**, 106626 (2023)
36. Zhao, T., Gu, Y., Yang, J., Usuyama, N., Lee, H.H., Kiblawi, S., Naumann, T., Gao, J., Crabtree, A., Abel, J., et al.: A foundation model for joint segmentation, detection and recognition of biomedical objects across nine modalities. Nature methods pp. 1–11 (2024)