Rethinking the Adaptation of Vision Foundation Models for Efficient Cell Segmentation

Qing Xu[1,2], Xiangjian He[2(✉)], Wenting Duan[3], Jiebo Luo[4], and Zhen Chen[5(✉)]

[1] School of Computer Science, University of Nottingham, UK
[2] School of Computer Science, University of Nottingham Ningbo China, China
[3] School of Engineering and Physical Science, University of Lincoln, UK
[4] Department of Computer Science, University of Rochester, USA
[5] Department of Data Science and Artificial Intelligence, The Hong Kong Polytechnic University, Hong Kong SAR
sean.he@nottingham.edu.cn, z.chen@polyu.edu.hk

Abstract. Cell segmentation is critical for computational pathology and biomedical discovery. While recent Vision Foundation Models (VFMs) have demonstrated remarkable universal feature representations, unlocking their full potential for cellular imaging is currently bottlenecked by resource-intensive adaptation paradigms. Existing methods typically rely on fine-tuning heavy visual encoders, leading to extensive computational overhead and a dependency on large-scale annotations. To address this, we propose the EffiCell-Seg framework for highly efficient cell segmentation without re-training the visual encoder. Our core insight is that pretrained VFMs intrinsically encode complementary structural priors: global saliency for localizing potential cells, and local morphological patterns for delineating cellular structures. To harness these priors, we devise a Cell Structure Prompt Encoder (CSP-Encoder) that synthesizes semantic-aware saliency and principal morphological features from frozen VFM representations into explicit structural prior maps. Moreover, we propose a Synergistic Mask Decoder (SM-Decoder) that enforces contextual consistency by jointly predicting geometric distance fields and semantic maps via mutual cross-guidance. Extensive experiments demonstrate that EffiCell-Seg outperforms state-of-the-art methods across diverse cell imaging modalities while requiring only ∼5M trainable parameters, over 130× fewer than fully fine-tuned VFM counterparts. The code is available at https://github.com/xq141839/EffiCell-Seg.

Keywords: Cellular Imaging · Cell Segmentation · Vision Foundation Models · Efficient Adaptation

1 Introduction

Cell segmentation plays a crucial role in computational pathology and high-throughput biomedical discovery [25, 12, 26, 27]. However, this task is inherently challenging due to the diversity of imaging modalities (*e.g.*, H&E staining, fluorescence microscopy), which exhibit varying contrast ratios and heterogeneous

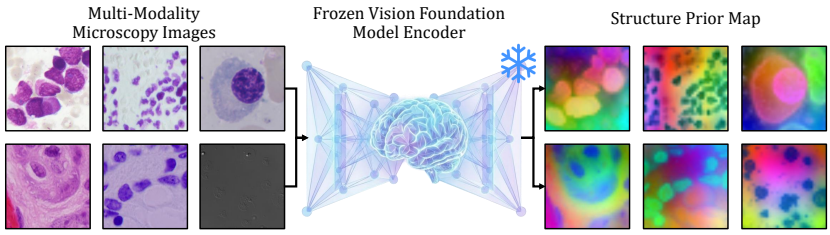

Fig. 1. Visualization of structural prior maps derived from the pretrained VFM. Given diverse multi-modality microscopy images, the frozen VFM encoder produces structural prior maps that effectively highlight potential cell regions and capture cellular structures.

cellular morphologies [15, 8, 16]. Recently, Vision Foundation Models (VFMs) such as DINOv3 [24], pre-trained on massive natural images, have shown remarkable capabilities in capturing universal feature representations, offering a promising avenue to address these issues. While VFMs possess immense potential, unlocking their full capability for cellular imaging is currently bottlenecked by resource-intensive adaptation paradigms, rather than the intrinsic representational limits of the models themselves.

The classical approaches [9, 18, 1] fully fine-tuned the entire VFM encoder, incurring substantial computational overhead. To further enhance region-level guidance, [20, 23] trained auxiliary segmentation or detection models to provide bounding box or point prompts, assisting the VFM in localizing potential cell regions, but at the cost of additional model complexity. Moreover, parameter-efficient fine-tuning techniques [13, 28] such as Adapter [4] and LoRA [10] have emerged as promising alternatives that enable effective adaptation to cell instance segmentation while reducing the number of trainable parameters [21, 5].

Despite these advancements, a fundamental limitation of existing methods [18, 1, 6] is their reliance on re-training the feature representations of the pretrained visual encoder, which typically requires large-scale annotated data. However, acquiring pixel-level cell annotations is notoriously labor-intensive, making it often prohibitive in practice. In fact, VFMs pre-trained on massive natural datasets already possess the capability to perceive universal structural patterns (*e.g.*, textures, boundaries, and spatial continuity), as illustrated in Fig. 1. These structural priors are largely shared across natural and cellular images, as cell instances inherently exhibit closed boundaries and locally coherent regions that closely resemble general visual structures. This observation motivates us to rethink the adaptation of VFMs and selectively re-align the existing structural knowledge of pretrained encoders to cellular instance segmentation.

To overcome this bottleneck, we propose the EffiCell-Seg framework for highly efficient cell segmentation that completely circumvents the need to re-train the heavy visual encoder. Our core insight is that the pretrained VFM intrinsically encodes two complementary structural priors: global saliency that highlights potential cell regions, and local morphological patterns that delineate cellular boundaries. Based on this observation, we make the following contributions:

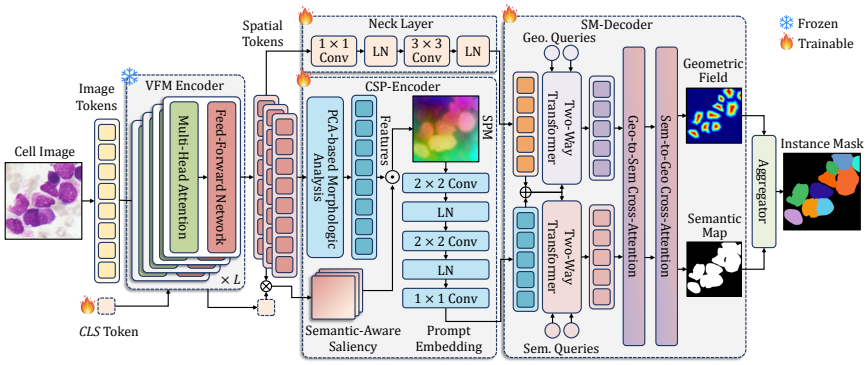

Fig. 2. Overview of the proposed EffiCell-Seg framework. The CSP-Encoder fuses semantic-aware saliency and PCA-based morphology from pretrained VFM features into structural prior maps, generating high-quality prompts for the SM-Decoder to jointly infer cell geometric structure and semantic map through cross-guided decoding.

(1) We devise a Cell Structure Prompt Encoder (CSP-Encoder) that synthesizes semantic-aware saliency and principal morphological features from frozen VFM representations into explicit structural prior maps, generating high-quality prompts that bridge pretrained knowledge and cellular structures. (2) We further devise a Synergistic Mask Decoder (SM-Decoder) that enforces contextual consistency by jointly predicting geometric distance fields and semantic maps via mutual cross-guidance. To the best of our knowledge, EffiCell-Seg is the first to achieve cell segmentation by exclusively re-aligning the inherent structural priors of a fully frozen VFM. (3) Extensive experiments show that EffiCell-Seg achieves state-of-the-art performance across diverse imaging modalities. Remarkably, it achieves such superior accuracy while requiring only ∼5M trainable parameters, over 130× fewer than fully fine-tuned VFM counterparts.

2 Methodology

As illustrated in Fig. 2, we propose the EffiCell-Seg framework that unlocks the structural knowledge of pretrained VFMs for efficient cell segmentation without re-training the heavy visual encoder. Specifically, we devise a CSP-Encoder that synthesizes semantic-aware saliency and morphological features from frozen VFM representations into structural prior maps, bridging pretrained knowledge and cellular structures. We further design an SM-Decoder that jointly predicts geometric distance fields and semantic maps via mutual cross-guidance, ensuring contextual consistency for accurate segmentation.

2.1 Preliminary and Insights

Current VFMs such as DINOv3 [24] typically adopt Vision Transformer (ViT) as the visual encoder, pre-trained on large-scale natural image datasets. Given

an input image x, we partition it into N non-overlapping patches and feed them into the pretrained VFM encoder. The encoder consists of L stacked transformer blocks, where each block applies multi-head self-attention (MHSA) and a feed-forward network (FFN) with layer normalization (LN) and residual connections:

$$h'_l = \mathrm{MHSA}(\mathrm{LN}(h_{l-1})) + h_{l-1}, \quad h_l = \mathrm{FFN}(\mathrm{LN}(h'_l)) + h'_l, \tag{1}$$

where $h_0 = [t_{\mathrm{cls}}; t_1, t_2, \ldots, t_N]$ denotes the initial token sequence. After L layers, we obtain the output representations $h_L = \{t_{\mathrm{cls}}, t_1, t_2, \ldots, t_N\}$, where $t_{\mathrm{cls}} \in \mathbb{R}^D$ is the CLS token and $\{t_i\}_{i=1}^N \in \mathbb{R}^{N \times D}$ are the spatial tokens.

Through self-attention across L layers, t_{cls} progressively aggregates information from all spatial tokens, forming a holistic summary of the entire image. Its affinity with individual spatial tokens thus naturally reflects the structural saliency of each local region, indicating *where* potential cell regions reside. Meanwhile, the spatial tokens $\{t_i\}_{i=1}^N$ preserve rich patch-level information encoding morphological patterns, characterizing *what* cellular structures look like. These two structural priors are inherently complementary: global saliency offers region-level localization while local patterns provide fine-grained structural details. As illustrated in Fig. 1, this insight motivates our EffiCell-Seg, which extracts and fuses these priors into structural prior maps for guiding cell instance segmentation.

2.2 Cell Structure Prompt Encoder

Following our insight that the CLS token and spatial tokens of pretrained VFMs encode complementary structural priors, we devise the CSP-Encoder to extract and fuse these priors into structure-aware prompts without any encoder fine-tuning. Specifically, to capture *where* potential cells reside, we compute the cosine similarity between t_{cls} and each spatial token t_i as a saliency map:

$$a_i = \frac{t_i \cdot t_{\mathrm{cls}}^\top}{\|t_i\| \|t_{\mathrm{cls}}\|}, \tag{2}$$

which effectively highlights foreground cell regions while suppressing background noise. To characterize *what* cellular structures look like, we apply PCA to $\{t_i\}_{i=1}^N$ and retain the top-K components as follows:

$$m_i = \mathrm{PCA}_K(\{t_i\}_{i=1}^N) \in \mathbb{R}^K. \tag{3}$$

The principal components capture dominant *morphological* and *textural* variations, encoding fine-grained structural details such as cell boundaries and internal textures. We then fuse these two priors through element-wise multiplication to obtain the structural prior map (SPM) as $z_i = a_i \cdot m_i$, where the saliency score acts as a spatial gate that selectively activates morphological features in structurally relevant regions. Finally, we employ a lightweight convolutional block to transform the structural prior map into the prompt embedding z, enforcing local spatial consistency for downstream dense prediction. In this way, the CSP-Encoder bridges frozen VFM representations and cellular morphology, providing reliable spatial guidance for subsequent geometric and semantic decoding.

2.3 Synergistic Mask Decoder

The CSP-Encoder provides structure-aware prompt embedding z from frozen VFM features. To achieve accurate cell instance segmentation, we further need to coherently decode this structural prior into dense predictions. To this end, we devise the SM-Decoder that jointly predicts geometric distance fields and semantic maps through cross-guided decoding. The core motivation is that geometric cues (*i.e.*, distance fields) provide precise boundary localization and instance separation but are sensitive to noise in low-contrast regions, while semantic confidence suppresses spurious responses but lacks the spatial precision to separate touching cells. Treating them independently often leads to inconsistent or fragmented predictions. The SM-Decoder therefore enforces mutual structural consistency between them through bidirectional cross-guidance. Specifically, the SM-Decoder maintains two sets of learnable query embeddings: geometric queries q_g for capturing spatial organization and instance centrality, and semantic queries q_s for encoding region-level category-aware features. We first adopt a projection layer $\mathrm{Prj}(\cdot)$ to align the dimensions of frozen VFM features with the prompt embedding, and then employ a two-way transformer [11], denoted as $\mathrm{TWA}(\cdot)$, to initialize geometric and semantic-aware representations as follows:

$$h_g = \mathrm{TWA}(q_g, q_g, \mathrm{Prj}(h) \oplus z), \quad h_s = \mathrm{TWA}(q_s, q_s, \mathrm{Prj}(h) \oplus z), \tag{4}$$

where $\oplus$ denotes element-wise addition. This grounds both query sets in the same structure-aligned feature space. We further leverage cross-guidance attention to facilitate bidirectional interaction between geometric and semantic embeddings:

$$h_g \leftarrow \mathrm{CrossAttn}(q_g, q_g, h_g \oplus h_s), \quad h_s \leftarrow \mathrm{CrossAttn}(q_s, q_s, h_g \oplus h_s), \tag{5}$$

which ensures that geometric precision and semantic confidence mutually reinforce each other, preventing degenerate solutions where either dominates in isolation. Finally, the SM-Decoder generates the distance fields and semantic maps through lightweight transposed convolutions. Through this design, the SM-Decoder coherently decodes structure-aware prompts into geometrically precise and semantically consistent predictions, enabling accurate cell instance segmentation.

2.4 Optimization Pipeline

In the design of our EffiCell-Seg framework, we adopt a pretrained DINOv3 [24] as the VFM encoder and freeze its weights throughout training to preserve the universal structural knowledge learned from large-scale natural datasets. This design not only reduces computational overhead but also prevents overfitting to domain-specific appearance variations. Adaptation to cellular scenarios is achieved exclusively through optimizing the lightweight CSP-Encoder and SM-Decoder. During the forward pass, frozen VFM features are first processed by the CSP-Encoder to generate the structure-aware prompt embedding, which then guides the SM-Decoder to jointly infer the horizontal and vertical distance fields and the semantic map through cross-guided decoding.

The training is supervised by a joint loss combining cross-entropy loss $\mathcal{L}_{\text{CE}}$ with dice loss $\mathcal{L}_{\text{Dice}}$ for semantic prediction, and MSE loss $\mathcal{L}_{\text{MSE}}$ for accurate distance estimation. The overall objective is formulated as follows:

$$\mathcal{L}_{\text{Cell}} = \mathcal{L}_{\text{CE}} + \mathcal{L}_{\text{Dice}} + \lambda_1 \mathcal{L}_{\text{MSE}}, \tag{6}$$

where λ_1 is the balancing coefficient that compensates for the magnitude discrepancy between regression and classification losses [7]. During inference, we apply a watershed algorithm [22] on the predicted distance field, guided by the semantic map, to separate individual cell instances. By optimizing $\mathcal{L}_{\text{Cell}}$, our EffiCell-Seg effectively activates and re-aligns the universal structural knowledge of the frozen VFM for accurate cell segmentation with superior computational efficiency.

3 Experiments

3.1 Experimental Setup

To validate the effectiveness of the proposed EffiCell-Seg, we conduct experiments on two benchmarks: the CellSeg dataset [17], containing 1,101 multimodal microscopy images, and the DSB dataset [2], comprising 670 images. These two datasets collectively cover diverse cellular imaging modalities and cell types, including plasma cells for Multiple Myeloma and white blood cells for Leukaemia. Both datasets adopt a common split of training, validation, and test sets as 7:1:2. We perform all experiments on a NVIDIA H20 GPU using PyTorch. We adopt DINOv3 [24] as the VFM encoder, whose weights remain frozen throughout training. All input images are resized to 512×512. We apply the AdamW optimizer with a learning rate of 1×10^{-4} and use CosineAnnealingLR for the scheduling strategy. The batch size and training epochs are set to 16 and 100, respectively. The balancing coefficients λ_1 are set to 10. For fair comparisons, all baselines are fine-tuned on the same training set and evaluated under the same input resolution. All SAM-based baselines [18, 23, 9, 19] adopt their automatic prompting mode. For quantitative evaluation, we adopt Dice and mIoU to measure semantic segmentation overlap, and F1 and Panoptic Quality (PQ) to assess instance-level detection and segmentation quality.

3.2 Comparison with State-of-the-art Methods

Cell Semantic Segmentation. We first evaluate all methods on cell semantic segmentation. As shown in Table 1, classical approaches [3, 14] show limited performance due to the lack of large-scale pretrained knowledge. Among VFM-based methods, CellSAM [18] and PromptNucSeg [23] achieve moderate results but require over 660M trainable parameters owing to fine-tuning the heavy visual encoder or training auxiliary models. CellViT [9] and SAC [19] reduce the parameter cost while improving performance, yet still depend on encoder fine-tuning. Our EffiCell-Seg achieves the best results on both datasets, outperforming the second-best SAC by 0.78% Dice and 0.97% mIoU on CellSeg, and 0.83%

Table 1. Comparison with state-of-the-art methods on cell semantic segmentation.

Methods	Trainable #Params	CellSeg		DSB	
		Dice	mIoU	Dice	mIoU
CPPNet [3]	32.67M	84.13	74.43	90.57	84.19
Swin-UMamba [14]	27.43M	85.24	75.81	91.23	85.02
CellSAM [18]	682.92M	85.87	76.52	91.58	85.46
PromptNucSeg [23]	661.74M	86.14	76.93	91.85	85.73
CellViT [9]	68.56M	86.48	77.12	92.04	85.96
SAC [19]	51.83M	86.92	77.51	92.31	86.34
EffiCell-Seg (Ours)	**4.99M**	**87.70**	**78.48**	**93.14**	**87.37**

Table 2. Comparison with state-of-the-art methods on cell instance segmentation.

Methods	Trainable #Params	CellSeg		DSB	
		PQ	F1	PQ	F1
CPPNet [3]	32.67M	55.83	54.96	70.12	68.75
Swin-UMamba [14]	27.43M	56.74	55.81	71.28	69.83
CellSAM [18]	682.92M	57.92	57.13	72.17	70.92
PromptNucSeg [23]	661.74M	58.31	57.52	72.85	71.43
CellViT [9]	68.56M	58.65	57.94	73.24	71.86
SAC [19]	51.83M	59.78	58.96	74.05	72.54
EffiCell-Seg (Ours)	**4.99M**	**61.16**	**60.34**	**75.32**	**73.89**

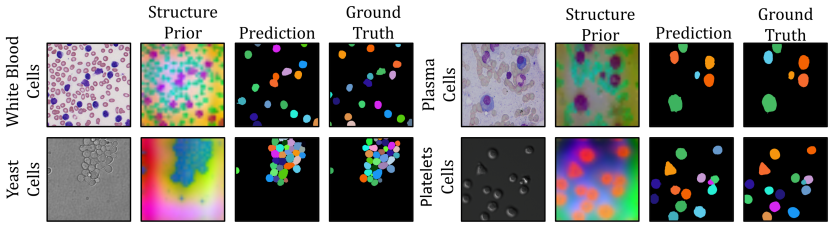

Fig. 3. Qualitative results of EffiCell-Seg on diverse cell types. By effectively leveraging structural priors from the pretrained VFM, our model achieves precise boundary delineation and robust instance separation across diverse imaging modalities.

Dice and 1.03% mIoU on DSB, while requiring only **4.99M** trainable parameters, approximately **10× fewer** than SAC and over **130× fewer** than CellSAM. These results demonstrate that re-aligning structural knowledge of pretrained VFMs is more effective than re-training the heavy visual encoder for cell segmentation.

Cell Instance Segmentation. We further evaluate all methods on cell instance segmentation. As shown in Table 2, the classical methods [3, 14] struggle in instance-level prediction, while heavily parameterized VFM-based methods such as CellSAM [18] and PromptNucSeg [23] do not necessarily translate to better instance discrimination. Our EffiCell-Seg achieves the best performance across both datasets, surpassing SAC by 1.27% PQ and 1.35% F1 on DSB, demonstrating that the cross-guided decoding in our SM-Decoder effectively leverages geometric

Table 3. Ablation study of our EffiCell-Seg framework on cell segmentation.

$\mathcal{P}$		$\mathcal{D}$	CellSeg				DSB			
α	m		Dice	mIoU	PQ	F1	Dice	mIoU	PQ	F1
			83.52	73.86	56.42	55.63	89.74	83.15	71.05	69.48
✓			85.41	75.92	58.15	57.26	91.26	84.83	72.74	71.23
	✓		84.93	75.48	57.68	56.84	90.87	84.41	72.31	70.85
✓	✓		86.28	76.85	59.53	58.71	91.94	85.72	73.86	72.34
✓	✓	✓	**87.70**	**78.48**	**61.16**	**60.34**	**93.14**	**87.37**	**75.32**	**73.89**

and semantic complementarity for coherent instance predictions. The qualitative results in Fig. 3 further illustrate that EffiCell-Seg produces highly precise boundary delineation and clear instance separation in challenging regions with overlapping and morphologically diverse cells. These results explicitly validate our core insight: rather than expensively re-training the entire vision foundation model, strategically re-aligning its inherent structural priors via our EffiCell-Seg is a substantially more efficient and robust paradigm for cellular analysis.

3.3 Ablation Study

To validate the effectiveness of the proposed CSP-Encoder $\mathcal{P}$ (*i.e.*, including semantic-aware saliency α and PCA Map m), and SM-Decoder $\mathcal{D}$, we conduct ablation studies on both CellSeg and DSB datasets for cell segmentation, as shown in Table 3. Starting from the baseline that directly decodes frozen VFM features without any proposed component, introducing the Similarity Map alone yields 1.89% Dice and 1.73% PQ improvements on CellSeg. Adding the PCA Map alone brings comparable gains of 1.41% Dice and 1.26% PQ. Incorporating a complete CSP-Encoder achieves further improvements across all metrics, demonstrating that this prior provides reliable region localization while local morphological patterns offer fine-grained structural details. Finally, incorporating the SM-Decoder with cross-guided decoding brings the most significant boost of 1.42% Dice and 1.63% PQ on CellSeg and 1.20% Dice and 1.46% PQ on DSB, highlighting the importance of enforcing geometric-semantic consistency for both semantic and instance-level predictions. These comprehensive results validate that the tailored CSP-Encoder and SM-Decoder collectively contribute to the superior performance of our framework across diverse cell segmentation scenes.

4 Conclusion

We present EffiCell-Seg, a highly *efficient* and *effective* cell segmentation framework that circumvents heavy visual encoder fine-tuning. Our core insight is that pretrained VFMs intrinsically encode complementary structural priors for cellular analysis. To rethink the VFM adaptation, we devise the CSP-Encoder to synthesize semantic-aware saliency and principal morphological features into structural prior maps. Moreover, we propose the SM-Decoder to jointly infer

geometric distance fields and semantic maps via cross-guided decoding, ensuring robust contextual consistency. Extensive evaluations demonstrate that the proposed EffiCell-Seg framework outperforms state-of-the-art methods across diverse imaging modalities, delivering superior segmentation accuracy while requiring much fewer trainable parameters than existing VFM adaptation paradigms.

Acknowledgments. This work is partially supported by the Yongjiang Technology Innovation Project (2022A-097-G), Zhejiang Department of Transportation General Research and Development Project (2024039), National Natural Science Foundation of China grant (62476037), Ningbo Science and Technology Bureau under NBNSF General Programme (2025J114), and PolyU Start-up Fund (P0060371).

Disclosure of Interests. The authors declare no competing interests.

References

1. Archit, A., Freckmann, L., Nair, S., Khalid, N., Hilt, P., Rajashekar, V., Freitag, M., Teuber, C., Spitzner, M., Tapia Contreras, C., et al.: Segment anything for microscopy. Nat. Methods **22**(3), 579–591 (2025)
2. Caicedo, J.C., Goodman, A., Karhohs, K.W., Cimini, B.A., Ackerman, J., Haghighi, M., Heng, C., Becker, T., Doan, M., McQuin, C., et al.: Nucleus segmentation across imaging experiments: the 2018 data science bowl. Nat. Methods **16**(12), 1247–1253 (2019)
3. 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)
4. Chen, Z., Duan, Y., Wang, W., He, J., Lu, T., Dai, J., Qiao, Y.: Vision transformer adapter for dense predictions. In: ICLR
5. 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)
6. Gao, Y., Li, H., Yuan, F., Wang, X., Gao, X.: Dino u-net: Exploiting high-fidelity dense features from foundation models for medical image segmentation. arXiv preprint arXiv:2508.20909 (2025)
7. 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)
8. Griebel, T., Archit, A., Pape, C.: Segment anything for histopathology. In: MIDL (2025)
9. 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)
10. 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)
11. 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)

12. Li, B., Liu, Z., Zhang, S., Liu, X., Sun, C., Liu, J., Qiu, B., Tian, J.: Nuhtc: A hybrid task cascade for nuclei instance segmentation and classification. Med. Image Anal. p. 103595 (2025)
13. Li, Y., Xu, Q., Zhang, Y., He, X., Zhang, Q., Yao, Y., Tesem, F.B., Chen, X., Wang, R., Chen, Z., et al.: Uniultra: Interactive parameter-efficient sam2 for universal ultrasound segmentation. IEEE Transactions on Multimedia (2026)
14. 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)
15. Lou, W., Li, H., Li, G., Lou, X., Xiong, Y., Wan, X., Wu, X.: Instance-aware multi-task learning for nuclei segmentation. IEEE Trans. Med. Imaging (2025)
16. 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. Knowl.-Based Syst. p. 113641 (2025)
17. Ma, J., Xie, R., Ayyadhury, S., Ge, C., Gupta, A., Gupta, R., Gu, S., Zhang, Y., Lee, G., Kim, J., et al.: The multimodality cell segmentation challenge: toward universal solutions. Nat. Methods **21**(6), 1103–1113 (2024)
18. Marks, M., Israel, U., Dilip, R., Li, Q., Yu, C., Laubscher, E., Iqbal, A., Pradhan, E., Ates, A., Abt, M., et al.: Cellsam: a foundation model for cell segmentation. Nat. Methods pp. 1–9 (2025)
19. Na, S., Guo, Y., Jiang, F., Ma, H., Gao, J., Huang, J.: Segment any cell: A sam-based auto-prompting fine-tuning framework for nuclei segmentation. IEEE Trans. Neural Netw. Learn. Syst. (2025)
20. 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)
21. Nam, S., Namgung, H., Jeong, J., Luna, M., Kim, S., Chikontwe, P., Park, S.H.: Instasam: Instance-aware segment any nuclei model with point annotations. In: MICCAI. pp. 232–242. Springer (2024)
22. Roerdink, J.B., Meijster, A.: The watershed transform: Definitions, algorithms and parallelization strategies. Fundam. Inform. **41**, 187–228 (2000)
23. 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)
24. Siméoni, O., Vo, H.V., Seitzer, M., Baldassarre, F., Oquab, M., Jose, C., Khalidov, V., Szafraniec, M., Yi, S., Ramamonjisoa, M., et al.: Dinov3. arXiv preprint arXiv:2508.10104 (2025)
25. Stringer, C., Wang, T., Michaelos, M., Pachitariu, M.: Cellpose: a generalist algorithm for cellular segmentation. Nat. Methods **18**(1), 100–106 (2021)
26. Xu, Q., Duan, W., Chen, Z.: Co-seg: Mutual prompt-guided collaborative learning for tissue and nuclei segmentation. In: MICCAI. pp. 130–140. Springer (2025)
27. Xu, Q., Luo, Y., Duan, W., Chen, Z.: Co-seg++: Mutual prompt-guided collaborative learning for versatile medical segmentation. IEEE Transactions on Medical Imaging (2025)
28. Zhang, Y., Xu, Q., Li, Y., He, X., Zhang, Q., Haque, M., Qu, R., Duan, W., Chen, Z.: Freqdino: Frequency-guided adaptation for generalized boundary-aware ultrasound image segmentation. In: ISBI. pp. 1–5. IEEE (2026)