CA-SAM2: SAM2-based Context-Aware Network with Auto-Prompting for Nuclei Instance Segmentation

Hanbin Huang, Hongliang He(✉), Liying Xu, Xudong Zhu, Siwei Feng, and Guohong Fu

School of Computer Science and Technology, Soochow University, Suzhou, China
hlhe2023@suda.edu.cn

Abstract. Nuclei instance segmentation is crucial for biomedical research and disease diagnosis. Pathologists utilize information such as color, shape, and the surrounding tissue microenvironment to distinguish nuclei. However, existing models are limited as they rely solely on features from the current patch, neglecting contextual information from neighboring patches. This limitation impedes the model's ability to accurately identify nuclei. To address this issue, we propose CA-SAM2, a novel framework that enhances the prompt propagation capability of the Segment Anything Model 2 (SAM2) through a Context Injection Module(CIM), integrating surrounding contextual information during segmentation. Additionally, to adapt SAM2 to the pathology image domain, we introduce a convolutional branch to extract domain-specific features from pathological images. We further design a Multi-Level Feature Refinement Block (MFRB) to refine the prior features extracted by SAM2 and integrate domain features. Finally, we incorporate a regression head and a classification head after the convolutional branch to automatically generate point prompts, eliminating the need for manual annotation. Extensive evaluations of CA-SAM2 on the MoNuSeg and CPM-17 datasets demonstrate its effectiveness and practicality in enhancing nuclei segmentation. The code is available at https://github.com/HanbinHuang123/CA-SAM2.

Keywords: Nuclei instance segmentation · Context-Aware · SAM2.

1 Introduction

Nuclear instance segmentation in pathological images is a fundamental step in extracting meaningful biological information. The relative topological structure, size, and shape of nuclei are crucial for downstream tasks such as cancer diagnosis and tumor grading [11]. However, digitized Whole-Slide Images (WSIs) possess extremely high resolution and contain numerous nuclei. Even when divided into smaller patches, pathologists still require significant time to review and annotate these patches [9], making label acquisition challenging. With the advancement of deep learning (DL) techniques, several automated nuclear segmentation models,

such as HoVer-Net [2], U-Net [14], and nn-UNet [5], have achieved impressive performance. These models employ an encoder to extract features from patches and a decoder to process these features into final segmentation masks. However, relying solely on local patch features may not be sufficient to achieve optimal results. In practice, pathologists also consider contextual information from surrounding patches when annotating, including nuclear spatial distribution and tissue structure. Therefore, we propose a novel pipeline that incorporates contextual information from surrounding patches during model prediction, leading to more accurate mask predictions.

Recently, Meta company introduced SAM2 [13], a foundational model for video segmentation, which incorporates a stream memory for real-time video processing and prompt propagation. Trained on the large-scale Segment Anything Video (SA-V) dataset, SAM2 demonstrates exceptional performance and generalization capabilities. Although originally designed for video segmentation, its unique memory mechanism holds potential for application in 2D pathological images. As shown in Fig. 1, when patches are treated as video frames, the scanning process of a WSI resembles a complete video. In this scenario, by applying SAM2 to process this "video", the prompt propagation mechanism can be used for context awareness. Based on this thinking, we apply SAM2 to nuclear instance segmentation in pathological images and propose a Context Injection Module(CIM), that extends SAM2's memory module to better capture surrounding contextual information.

However, the application of SAM2 in pathological scenarios presents two primary challenges. First, the SAM2 training dataset does not include pathological images. Directly applying the model to these images would degrade segmentation performance, as SAM2 is unable to extract domain-specific features present in pathological images. One possible approach would be to retrain the entire SAM2 model on pathological images. However, this would require a large-scale pathological dataset and substantial computational resources, which is impractical given the difficulty of annotating pathological images. Second, although SAM2 supports prompt propagation and is capable of segmenting an entire video using only the prompt from the first frame, the nuclear distribution in pathological images is relatively dense, making it difficult to capture all nuclear instances through prompt propagation alone. This often leads to missed detections that impair segmentation performance. To address these challenges, we freeze SAM2's image encoder and introduce a convolutional branch to assist in extracting domain-specific features. Additionally, we design a Multi-level Feature Refinement Block (MFRB), which facilitates the fine-tuning of features and enables efficient information flow between the ViT and convolutional branches. This design helps SAM2 adapt to the pathology image domain. Finally, we incorporate a regression head and a classification head into the convolutional branch, which automatically generates point prompts for each nuclei, reducing the number of missed detections.

In summary, our main contributions are as follows: (1) We propose a novel context-aware network based on SAM2, which incorporates surrounding context

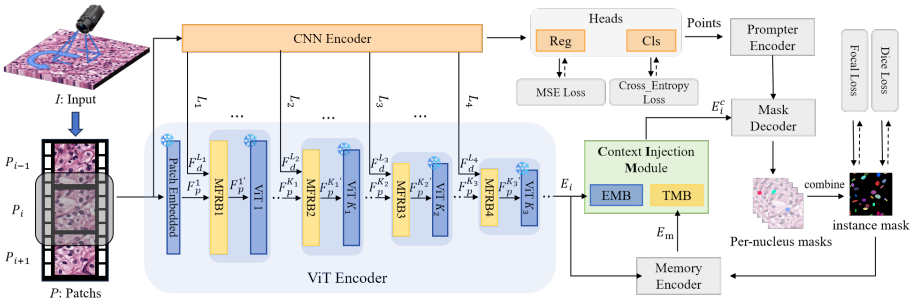

Fig. 1. Overall architecture of the proposed CA-SAM2. Consisting of two parallel branches: the ViT branch and the convolutional branch. In the ViT branch, 1, K_1, K_2, and K_3 represent the starting indices for the four stages of the SAM2 image encoder, with specific values depending on the backbone. The ViT branch captures prior features, while the convolutional branch extracts domain-specific features. These features are refined and fused through the MFRB to adapt to the pathology image domain. Subsequently, the CIM supplements the image embedding with contextual information to better decode the mask. Finally, both the image embedding and memory embedding are input into the CIM to provide contextual information for the next patch.

information during model segmentation, thereby improving segmentation performance.(2) We introduce the Multi-level Feature Refinement Block (MFRB), which fine-tunes the prior features extracted by the ViT branch and integrates the domain features extracted by the convolutional branch, bridging the domain gap between natural images and pathological images.(3) Extensive experiments on the MoNuSeg and CPM-17 datasets demonstrate the effectiveness and superiority of the proposed method.

2 Method

2.1 Overview of CA-SAM2

The proposed network is illustrated in Fig 1. For the input pathological image $I \in \mathbb{R}^{H\times W\times 3}$, we employ a sliding window to scan I in a counterclockwise direction, starting from the center, generating a set of patches $P = \{p\}_{i=1}^{n} \in \mathbb{R}^{256\times 256\times 3}$, which correspond to video frames. For the i-th input patch P_i, it is first passed through the encoders of both branches to extract prior features F_p and domain features F_d, respectively. Meanwhile, by incorporating the Multi-level Feature Refinement Block (MFRB) before each ViT block, the prior features F_p are refined, and missing domain features are extracted and fused from F_d , facilitating adaptation to the pathology image domain. The enhanced image embedding E_i is then input into the Context Injection Module (CIM) to integrate contextual information from the surrounding segmented patches, generating the final image embedding E_i^c. Additionally, a regression(Reg) head and a classification(Cls)

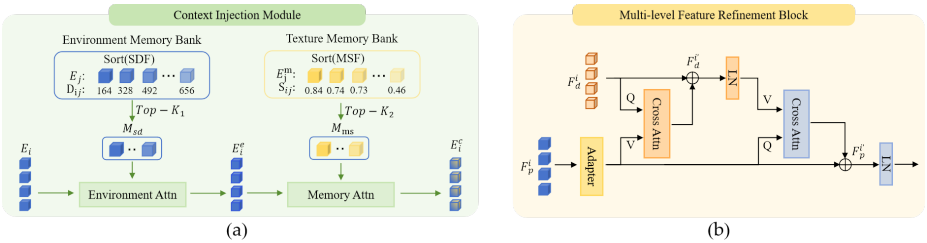

Fig. 2. Detailed architecture of CIM and MFRB. The blue, orange, and yellow embeddings represent the Prior embedding, Domain embedding, and Memory embedding, respectively.

head are added after the convolutional branch to predict the points corresponding to each nuclei[16], which are treated as point prompts and passed into the prompt encoder. Inspired by PromptNucSeg[15] we adopt a "one-prompt-one-nucleus" approach for decoding. After obtaining the mask for each nucleus from the mask decoder, these masks are combined to form the instance mask. Finally, the instance mask, along with the image embedding E_i, is input into the memory encoder to generate the memory embedding E_m. Specifically, the E_m and the E_i are separately stored in different memory banks within the CIM, where they continue to provide contextual information for the next patch.

2.2 Context Injection Module

Although SAM2's memory module utilizes foreground information from segmented patches to assist in segmenting the current patch, the first-in, first-out management strategy does not effectively provide the necessary foreground information. Additionally, background information, such as the tissue microenvironment plays an essential role in identifying nuclear instances, which SAM2 currently lacks. To address these issues, we have designed the CIM to complement the image embedding with both foreground information (e.g., nucleus texture and size) and background information (e.g., tissue microenvironment) from surrounding segmented patches. The CIM consists primarily of two Memory Banks: Environment Memory Bank(EMB) and Texture Memory Bank(TMB). Environment Memory Bank is designed to supplement the surrounding tissue microenvironment information, while Texture Memory Bank provides additional foreground information, such as the texture and size of similar nuclei. Consequently, our CIM effectively extracts and utilizes surrounding contextual information.

Environment Memory Bank: To better capture the characteristics of the nuclei and their surrounding tissue environment, Environment Memory Bank(EMB) stores the image embeddings E_j (where $j \in 1 \sim i-1$) of all previously segmented patches from the same image. As shown in Fig 2(a), when an E_i is input,

the EMB reorders itself using a Shortest Distance First (SDF) strategy, and select the K embeddings closest to E_i to create the shortest distance memory M_{sd}. Then apply M_{sd} to inject surrounding environmental information into E_i:

$$E_i^e = \mathcal{A}_e(E_i, M_{sd}) \tag{1}$$

$$M_{sd} = \{E_k \,|\, \forall k \in \{\text{Top-K1}(D_{ij})\},\, k = 1, \ldots, K1\} \tag{2}$$

$$D_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \tag{3}$$

where D_{ij} represents the distance between E_i and E_j, in this paper, we calculate this distance using Euclidean distance. Where (x_i, y_i) and (x_j, y_j) are the coordinates of E_i and E_j, and $\mathcal{A}_e$ uses one layer of the memory attention network from SAM2, trained independently.

Finally, the resulting E_i^e is input into Texture Memory Bank for integrating foreground information. Notably, to maintain the EMB, it is cleared whenever the image is changed.

Texture Memory Bank: To better provide nucleus features similar to those in the input E_i^e, along with their corresponding masks. Texture Memory Bank(TMB) stores the memory embeddings of all previously segmented patches with the maximum similarity differences. Inspired by MedSAM2[20], we use the similarity of image embeddings to measure the similarity of nucleus features, with confidence incorporated to update the TMB. When a memory embedding E_m is to be enqueued and the TMB is full, first enqueue E_m and calculate the similarity between every pair of embeddings E_i and E_j in the current TMB, resulting in the similarity matrix S_a:

$$S_a = \{S_{ij} | i \in [1, M_t + 1], j \in [1, M_t + 1]\} \tag{4}$$

$$S_{ij} = \text{CosSim}(E_i, E_j) \tag{5}$$

where S_{ij} represents the cosine similarity between E_i and E_j, and M_t is the size of the TMB.

Then select the embedding $E_{\max}$ whose sum of similarities with other embeddings is the largest. If its Intersection over Union(IoU) value minus 0.1 is smaller than that of E_m, $E_{\max}$ will be dequeued. Otherwise, E_m will be dequeued. This strategy enables real-time updates to the TMB, storing reliable and diverse memory embeddings within it.

Furthermore, as shown in Fig 2(a), when a embedding E_i^e is input, Texture Memory Bank reorders itsel using a Maximum Similarity First (MSF) strategy, and selecting the top K embeddings with the highest similarity to E_i^e, forming the maximum similarity memory M_{ms}. Then apply M_{ms} to inject foreground information into E_i^e. This process can be expressed as:

$$E_i^c = \mathcal{A}_m(E_i^e, M_{\text{ms}}) \tag{6}$$

$$M_{\text{ms}} = \{E_k^m | \forall k \in \{\text{Top-K2}(S_{ij})\}, k = 1, \ldots, K_2\} \tag{7}$$

where E_j^m is the memory embedding stored in the TMB, S_{ij} represents the cosine similarity between E_i^e and E_j^m, and $\mathcal{A}_m$ is the memory attention module used in SAM2.

Finally, the resulting E_i^c is passed to the mask decoder to decode the per-nucleus masks.

2.3 Multilevel Feature Refinement Block

Although Adapter[1] achieves good performance with relatively few parameters, this fine-tuning strategy may not fully adapt the model to specific domains, especially in complex scenarios such as pathological images. To address this limination, we introduce a convolutional branch based on ConvNeXt[10] to extract domain-specific features from pathological images. Additionally, we designed a MFRB to facilitate information flow between the two branches. As shown in Fig 2(b), the features from the ViT branch and the convolutional branch at the i-th layer are denoted as F_p^i and F_d^i, respectively. First, F_p^i undergoes initial fine-tuning through the Adapter to incorporate domain knowledge. Then, it is used to refine the domain feature representation in F_d^i, generating $F_d^{i'}$. Finally, $F_d^{i'}$ is used to update F_p^i, complementing it with missing domain features. This process can be formulated as follows:

$$F_d^{i'} = F_d^i + \mathcal{A}_c(F_d^i, Adapter(F_p^i)) \tag{8}$$

$$F_p^{i'} = F_p^i + \mathcal{A}_c(F_p^i, LN(F_d^{i'})) \tag{9}$$

where LN represents layer normalization, and $\mathcal{A}_c$ is the cross-attention.

The resulting $F_p^{i'}$ is normalized and then input to the next ViT block for further refinement, ultimately generating the enhanced image embedding E_i.

3 Experiments

3.1 Datasets and Implementation

To validate the effectiveness of CA-SAM2, we evaluated the model on the MoNuSeg[8] dataset from the 2018 MICCAI challenge and the public brain glioma CPM-17[17] dataset. The MoNuSeg dataset consists of 30 histopathological images with a resolution of 1000×1000, including both benign and malignant tissue samples from seven different organs: breast, liver, kidney, bladder, prostate, colon, and stomach. The CPM-17 dataset contains 64 H&E stained images (with size: 500×500 or 600×600) and includes annotations for 7,570 annotated nuclei. Both the training and test sets consist of 32 images each.

The experiments were conducted on an NVIDIA V100 GPU, with a patch size of 256×256, a stride of 164, and the size of the TMB was set to 64. Considering the influence of input order on contextual representation, we evaluated four strategies: Row-wise, Zigzag, Spiral inward, and Spiral outward, and used the best-performing Spiral outward. We also stored the training-time TMB in checkpoints to provide foreground information from the train set for the test set.

Table 1. Performance Comparison of SAM-Based Models, SAM2-Based Models and state-of-the-art Expert Models on the CPM-17 and MoNuSeg Datasets.

Dataset	Model	Prompt	Dice	AJI	DQ	SQ	PQ
CPM-17	U-Net [14] (2015)	✗	0.813	0.643	0.778	0.734	0.578
	Mask R-CNN [3] (2017)	✗	0.850	0.684	0.848	0.792	0.674
	HoVer-Net [2] (2019)	✗	0.878	0.712	0.851	0.808	0.690
	SMILE [12] (2023)	✗	0.881	0.723	0.779	0.759	0.705
	PointNu-Net [19] (2024)	✗	0.859	0.705	0.872	0.803	0.701
	SAM(Zero-Shot) [7] (2023)	✗	0.673	0.458	0.450	0.721	0.340
	MedSA [18] (2023)	✗	0.823	0.577	0.740	0.743	0.551
	HQ-SAM [6] (2023)	✗	0.742	0.640	0.804	0.795	0.641
	CellViT [4] (2024)	✗	0.874	0.711	0.858	0.806	0.693
	PromptNucSeg-L [15] (2024)	Auto-point	0.870	0.718	0.879	0.812	0.715
	SAM2 [13] (2024)	One-point	0.826	0.517	0.626	0.738	0.466
	MedSAM2 [20] (2024)	One-point	0.825	0.519	0.634	0.739	0.472
	CA-SAM2	Auto-point	**0.881**	**0.746**	**0.881**	**0.819**	**0.723**
MoNuSeg	U-Net [14] (2015)	✗	0.801	0.573	0.666	0.739	0.495
	Mask R-CNN [3] (2017)	✗	0.760	0.546	0.521	0.700	0.374
	HoVer-Net [2] (2019)	✗	0.803	0.595	0.743	0.754	0.563
	SMILE [12] (2023)	✗	0.782	0.576	0.709	0.749	0.533
	PointNu-Net [19] (2024)	✗	0.792	0.582	0.754	0.750	0.568
	SAM(Zero-Shot) [7] (2023)	✗	0.448	0.188	0.106	0.562	0.069
	MedSA [18] (2023)	✗	0.669	0.361	0.359	0.671	0.242
	HQ-SAM [6] (2023)	✗	0.663	0.526	0.685	**0.759**	0.519
	CellViT [4] (2024)	✗	0.797	0.584	0.722	0.749	0.543
	PromptNucSeg-L [15] (2024)	Auto-point	0.804	0.607	**0.767**	0.758	**0.584**
	SAM2 [13] (2024)	One-point	0.732	0.382	0.377	0.673	0.257
	MedSAM2 [20] (2024)	One-point	0.735	0.394	0.403	0.675	0.276
	CA-SAM2	Auto-point	**0.810**	**0.619**	0.760	**0.759**	0.579

3.2 Results

Table 1 compares our method with several state-of-the-art models, including the Expert model, SAM-based model, and SAM2-based model. In the SAM-based model, our method exhibits a 0.5% reduction in PQ compared to PromptNucSeg on MoNuSeg. This is due to the relatively dense distribution of nuclei in MoNuSeg and the fact that the Prompter in PromptNucSeg is independently trained to generate point prompts, endowing it with more specialized capabilities compared to the convolutional branch in our method. However, in terms of AJI, our CA-SAM2 outperforms PromptNucSeg by 2.8% and 1.2%, demonstrating that our method achieves better segmentation results even with suboptimal point prompts, thereby proving that our module significantly enhances segmentation performance. Compared to the Expert model and the SAM2-based model, our method performs better on both datasets. Specifically, our AJI improves by 4.1% and 3.7% when compared to the PointNu-Net. Notably, for our CA-SAM2, SAM2, MedSAM2, and PromptNucSeg architectures, we use Hiera-L and ViT-L as backbones, while all SAM-based models utilize the ViT-B configuration to

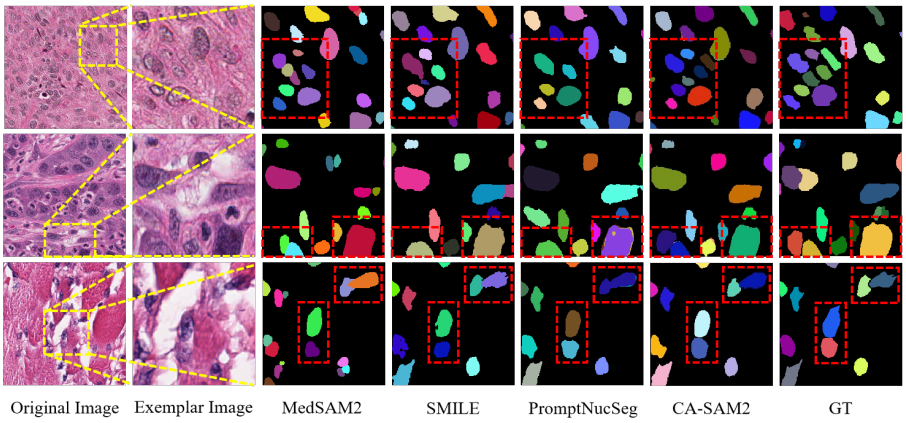

Fig. 3. Qualitative examples of the MedSAM2, SMILE, PromptNucSeg, our model and GT.

Table 2. Ablation study of the proposed modules on the CPM-17 dataset, where TTA stands for test-time augmentation.

MFRB	EMB	TMB	TTA	AJI	PQ
				0.711	0.687
✔				0.732	0.706
✔		✔		0.736	0.708
✔	✔			0.739	0.712
✔	✔	✔		0.742	0.719
✔	✔	✔	✔	0.746	0.723

Table 3. Ablation experiments on four backbones of SAM2.

Dataset	Model	AJI	PQ
CPM-17	Hiera-T	0.724	0.702
	Hiera-S	0.730	0.707
	Hiera-B+	0.734	0.712
	Hiera-L	0.746	0.723
MoNuSeg	Hiera-T	0.597	0.558
	Hiera-S	0.596	0.547
	Hiera-B+	0.599	0.549
	Hiera-L	0.619	0.579

ensure fairness in the experiments and the accuracy of the performance evaluation.

Fig 3 presents a qualitative comparison of SMILE, PromptNucSeg, MedSAM2, and our CA-SAM2 on the CPM-17 dataset, demonstrating that CA-SAM2 excels in accurately segmenting nuclear instances. Specifically, compared to SMILE, PromptNucSeg, and MedSAM2, our proposed method shows a accuracy detection region on nuclear edges, provides more precise boundary differentiation.

3.3 Ablation Study

We conducted ablation experiments to verify the effectiveness of the proposed MFRB, EMB, and TMB. For the baseline, we use Hiera-L as the backbone of SAM2 and remove the memory encoder. As shown in Table 2, each module significantly enhances the overall performance, and their combination yields the best overall results.

Although we freeze the image encoder, in order to compare the performance differences caused by the pre-trained scale of SAM2, we conducted ablation experiments on different backbones. As shown in Table 3, the segmentation performance of the model gradually decreases as the backbone decreases. However, even with the use of Hiera-B+, our model remains highly competitive.

4 Conclusion

We propose a SAM2-based context-aware network, CA-SAM2, which enables the model to incorporate surrounding contextual information during segmentation. The CIM enhances the context from the background and foreground. Additionally, the MFRB promotes the model adaptation to the pathology image domain by fine-tuning prior features and integrating domain features. Extensive experiments on the MoNuSeg and CPM-17 datasets demonstrate the effectiveness of CA-SAM2, achieving promising segmentation results.

Acknowledgments. This work is supported by the National Nature Science Foundation of China (No. 62406214, 62476187) and the Natural Science Foundation of Jiangsu Province, China (No. BK20240781).

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

References

1. Chen, T., Lu, A., Zhu, L., Ding, C., Yu, C., Ji, D., Li, Z., Sun, L., Mao, P., Zang, Y.: Sam2-adapter: Evaluating & adapting segment anything 2 in downstream tasks: Camouflage, shadow, medical image segmentation, and more. arXiv preprint arXiv:2408.04579 (2024)
2. 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)
3. He, K., Gkioxari, G., Dollár, P., Girshick, R.: Mask r-cnn. In: Proceedings of the IEEE international conference on computer vision. pp. 2961–2969 (2017)
4. 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. Medical Image Analysis **94**, 103143 (2024)
5. Isensee, F., Petersen, J., Klein, A., Zimmerer, D., Jaeger, P.F., Kohl, S., Wasserthal, J., Koehler, G., Norajitra, T., Wirkert, S., et al.: nnu-net: Self-adapting framework for u-net-based medical image segmentation. arXiv preprint arXiv:1809.10486 (2018)
6. Ke, L., Ye, M., Danelljan, M., Tai, Y.W., Tang, C.K., Yu, F., et al.: Segment anything in high quality. Advances in Neural Information Processing Systems **36** (2024)
7. 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. pp. 4015–4026 (2023)

8. Kumar, N., Verma, R., Sharma, S., Bhargava, S., Vahadane, A., Sethi, A.: A dataset and a technique for generalized nuclear segmentation for computational pathology. IEEE transactions on medical imaging **36**(7), 1550–1560 (2017)
9. Li, Y., Ren, H., Deng, J., Ma, X., Xie, X.: Centersam: Fully automatic prompt for dense nucleus segmentation. In: 2024 IEEE International Symposium on Biomedical Imaging (ISBI). pp. 1–5. IEEE (2024)
10. Liu, Z., Mao, H., Wu, C.Y., Feichtenhofer, C., Darrell, T., Xie, S.: A convnet for the 2020s. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 11976–11986 (2022)
11. Nasir, E.S., Parvaiz, A., Fraz, M.M.: Nuclei and glands instance segmentation in histology images: a narrative review. Artificial Intelligence Review **56**(8), 7909–7964 (2023)
12. Pan, X., Cheng, J., Hou, F., Lan, R., Lu, C., Li, L., Feng, Z., Wang, H., Liang, C., Liu, Z., et al.: Smile: Cost-sensitive multi-task learning for nuclear segmentation and classification with imbalanced annotations. Medical Image Analysis **88**, 102867 (2023)
13. Ravi, N., Gabeur, V., Hu, Y.T., Hu, R., Ryali, C., Ma, T., Khedr, H., Rädle, R., Rolland, C., Gustafson, L., et al.: Sam 2: Segment anything in images and videos. arXiv preprint arXiv:2408.00714 (2024)
14. 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)
15. 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: European Conference on Computer Vision. pp. 288–304. Springer (2024)
16. Shui, Z., Zheng, S., Zhu, C., Zhang, S., Yu, X., Li, H., Li, J., Chen, P., Yang, L.: Dpa-p2pnet: deformable proposal-aware p2pnet for accurate point-based cell detection. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 38, pp. 4864–4872 (2024)
17. Vu, Q.D., Graham, S., Kurc, T., To, M.N.N., Shaban, M., Qaiser, T., Koohbanani, N.A., Khurram, S.A., Kalpathy-Cramer, J., Zhao, T., et al.: Methods for segmentation and classification of digital microscopy tissue images. Frontiers in bioengineering and biotechnology **7**, 433738 (2019)
18. Wu, J., Ji, W., Liu, Y., Fu, H., Xu, M., Xu, Y., Jin, Y.: Medical sam adapter: Adapting segment anything model for medical image segmentation. arXiv preprint arXiv:2304.12620 (2023)
19. Yao, K., Huang, K., Sun, J., Hussain, A.: Pointnu-net: Keypoint-assisted convolutional neural network for simultaneous multi-tissue histology nuclei segmentation and classification. IEEE Transactions on Emerging Topics in Computational Intelligence **8**(1), 802–813 (2023)
20. Zhu, J., Qi, Y., Wu, J.: Medical sam 2: Segment medical images as video via segment anything model 2. arXiv preprint arXiv:2408.00874 (2024)