InstaSAM: Instance-aware Segment Any Nuclei Model with Point Annotations

Siwoo Nam[1,3], Hyun Namgung[1], Jaehoon Jeong[1], Miguel Luna[1], Soopil Kim[1], Philip Chikontwe[2], and Sang Hyun Park[1,3⋆]

[1] Department of Robotics and Mechatronics Engineering, Daegu Gyeongbuk Institute of Science and Technology (DGIST), Daegu, Korea
[2] Department of Biomedical Informatics, Harvard Medical School, MA, USA
[3] Department of Biostatistics, Epidemiology and Informatics, University of Pennsylvania, PA, USA
{siwoonam,shpark13135}@dgist.ac.kr

Abstract. Weakly supervised nuclei segmentation methods have been proposed to simplify the demanding labeling process by primarily depending on point annotations. These methods generate pseudo labels for training based on given points, but their accuracy is often limited by inaccurate pseudo labels. Even though there have been attempts to improve performance by utilizing power of foundation model *e.g.*, Segment Anything Model (SAM), these approaches require more precise guidance (*e.g.*, box), and lack of ability to distinguish individual nuclei instances. To this end, we propose InstaSAM, a novel weakly supervised nuclei instance segmentation method that utilizes confidence of prediction as a guide while leveraging the powerful representation of SAM. Specifically, we use point prompts to initially generate rough pseudo instance maps and fine-tune the adapter layers in the image encoder. To exclude unreliable instances, we selectively extract segmented cells with high confidence from pseudo instance segmentation and utilize these for the training of binary segmentation and distance maps. Owing to their shared use of the image encoder, the binary map, distance map, and pseudo instance map benefit from complementary updates. Our experimental results demonstrate that our method significantly outperforms state-of-the-art methods and is robust in few-shot, shifted point, and cross-domain settings. The code will be available upon publication.

Keywords: Nuclei instance segmentation · Weakly supervised learning · Segment anything model · Model adaptation · Point annotation

1 Introduction

Separating cells or nuclei from pathology images is a crucial problem, leading to the proposal of numerous deep learning-based techniques [12,1,17]. Annotating a large number of nuclei to create a robust model requires significant effort

⋆ Corresponding author.

and time. Consequently, recent advancements have introduced weakly supervised techniques capable of segmenting entire cells even with point annotations within the cells, aiming to address this challenge [20,21,13,19,22,24]. To train the models using limited point annotations, most existing methods [20,21,13,19] utilize k-means clustering [14] to separate foreground and background, generate pseudo-instance segmentation labels based on Voronoi diagrams using point annotations, and define losses to align predictions with these labels. These methods solely rely on color and distance information from images to generate pseudo labels resulting in inaccuracies, particularly in ambiguous nuclei boundaries, often leading to significant performance degradation.

Recently, several nuclei segmentation models that leverage pre-trained knowledge of the Segment Anything Model (SAM) have been proposed. CellViT [8] was proposed to semantically segment nuclei by fine-tuning SAM with a large amount of nuclei segmentation data, but it requires pixel-level annotations of a large number of nuclei for training. Even if weakly supervised nuclei segmentation models have also been proposed [4,18], their tasks are limited to binary segmentation tasks, leaving instance segmentation out of reach.

In this paper, we propose a novel weakly supervised nuclei instance segmentation method that can accurately segment nuclei and distinguish instances by leveraging powerful representation of SAM. Our model utilizes complementary training of pseudo instance segmentation labels, binary maps and distance maps, and a novel pseudo labeling process. Specifically, we design a nuclei decoder that predicts binary maps and distance maps as labels to utilize SAM in the instance segmentation task. Each point of annotation is used as a SAM prompt to generate pseudo instance nuclei segmentation labels. However, pseudo label may include wrong predictions that can negatively impact model performance. To ensure stable learning, only the predicted foregrounds with high confidence and no overlapping are utilized to define the losses, filtering out inaccurate pseudo-instance segmentation labels during training. Instead of training the entire image encoder, we insert adapters in the image encoder and train only the adapters, and tokens and Multi-Layer Perceptrons (MLPs) in the nuclei decoder. Then, we update the parameters of the adapters and nuclei decoder to generate better pseudo-instance segmentation labels, binary segmentation, and distance maps by minimizing the losses between the predictions and the pseudo labels. Our contributions are as follows:

1. We apply SAM to weakly supervised nuclei instance segmentation task, allowing rapid adaptation to the target domain with minimal parameter changes.
2. We propose a method that simultaneously considers tasks for individually segmenting each cell and segmenting the entire cells from the image by leveraging high-confidence pseudo-labels in training. These significantly improve performance.
3. In various experiments, we achieved state-of-the-art performance in the nuclei instance segmentation task and confirmed the robustness

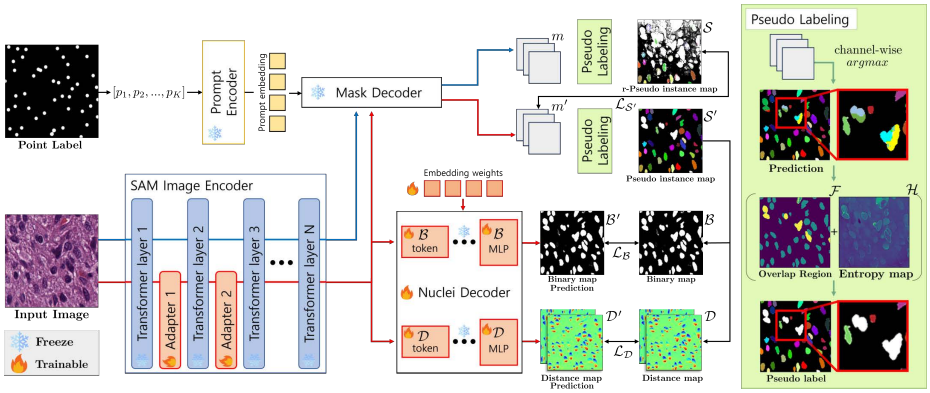

Fig. 1: Overview of the InstaSAM. InstaSAM comprises the adapter for fine-tuning, a mask decoder for prompt-based instance segmentation, and a nuclei decoder for segmentation without prompts. The pseudo instance map used for training is generated from the mask decoder's output, with areas ignored during pseudo labeling shown in white.

of the proposed method in few-shot, shift point, and cross-domain settings.

2 Methodology

Fig. 1 shows an overview of our proposed method of employing a SAM structure for weakly supervised nuclei segmentation. Similar to existing weakly supervised nuclei segmentation methods [20,13,19], we create pseudo labels from point annotations to train our model. Initially, using a vanilla SAM, we created a rough pseudo instance segmentation label $\mathcal{S}$ by prompting each point annotation to obtain a segmentation mask for each nucleus. Pseudo labels are further refined by rejecting mask overlaps and regions of high entropy. Then, we fine-tune the image encoder using adapter layers [2] by inputting each point as a prompt to obtain segmentation results $\mathcal{S}'$, aiming to make $\mathcal{S}'$ similar to $\mathcal{S}$. Simultaneously, predicted instance pseudo labels $\mathcal{S}'$ are refined and used to create additional pseudo labels for binary segmentation $\mathcal{B}$ to detect foreground pixels and horizontal and vertical distance maps $\mathcal{D}$ to distinguish boundaries between adjacent nuclei. An additional fully trainable nuclei decoder is then used to predict binary segmentation B' and distance maps D' from adapted image embeddings. In the early stages of updating, $\mathcal{S}$ guides $\mathcal{S}'$, assisting in achieving favorable results for $\mathcal{B}'$ and $\mathcal{D}'$. As the adapter is updated to better extract nuclei characteristics based on $\mathcal{B}'$ and $\mathcal{D}'$, they mutually complement each other to enhance $\mathcal{S}'$. During inference, only the image is input since point annotations are not provided. Thus, the model generates $\mathcal{B}'$ and $\mathcal{D}'$, from which the final instance segmentation is predicted. The details for each part are described in the subsections below.

SAM Adapter The naive approach to apply the SAM model for nuclei segmentation is to use point coordinates to prompt the model to segment each instance. Specifically, to segment k_{th} nuclei, the image is transformed into an embedded representation via the image encoder while the position coordinates of the target nuclei are encoded by the prompt encoder. Image and prompt embeddings are used by the mask decoder to predict the mask of k_{th} nuclei. However, due to the domain gap with natural images, predicted masks from pathology images tend to include a lot of noise leading to a significant drop in segmentation quality [5]. Therefore, we insert adapter layers [9,3,2] to the image encoder to reduce the discrepancy between domains and facilitate the nuclei segmentation task. In detail, we insert two MLP layers between transformer blocks in the image encoder to fine-tune the model with a minimal number of trainable parameters. Each adapter layer combines the patch embeddings (F^i_{pe}) with the high-frequency components from the Fourier transform of the input image (F_{hfc}) to obtain an adapted embedded representation ($\mathbf{A}^i$):

$$\mathbf{A}^i = \text{MLP}_{up}(\text{GELU}(\text{MLP}^i_{tune}(F_{hfc} + F^i_{pe}))), \tag{1}$$

where MLP^i_{tune} is specific for each adapter layer and MLP_{up} is a projection layer shared across all adapter layers used to adjust the feature dimensions. Following [2], we applied GELU [7] after MLPs.

Pseudo Labeling Previous works on weakly supervised nuclei segmentation methods which are based on point annotations [20,21,13,19] depend on clustering techniques and Voronoi labels to create pseudo labels with the assumption that nuclei are circular. Creating more accurate pseudo labels using Geodesic distance also has the same limitations [19]. In contrast, leveraging the strong learned representation of SAM, we implement instance-level pseudo label generation by using point coordinates as prompts. However, due to the domain gap with natural images, our proposed pseudo labeling process excludes uncertain areas, focusing only on reliable regions for training. Inaccurate areas are identified where there are instance predictions with high entropy values and also where there are nuclei with overlapping masks. Formally, for a set of point annotations $p = p_1, p_2, \ldots, p_K$ there is an equal number of predicted masks $m = m_1, m_2, \ldots, m_K$. For k_{th} nuclei, prompts are constructed as one positive point (p_k) concatenated to four randomly selected negative points from other nuclei. Subsequently, a nuclei binary mask $\hat{m}_k$ is defined by thresholding the probability values of m_k by a factor δ. The total foreground area $\mathcal{F}$ is obtained by adding all the foreground instance masks $\hat{m}_k$. Likewise, the entropy map $\mathcal{H}$ is calculated after applying a sigmoid function to the nuclei mask probability values $\sigma(m_k) = \text{sigmoid}(m_k)$. The refined pseudo label $\hat{y}$ for each pixel position

j is then computed as follows:

$$\mathcal{F} = \sum_{k=1}^{K} \mathbb{1}[\sigma(m_k(j)) > \delta] \tag{2}$$

$$\mathcal{H} = -\sum_{k=1}^{K} \sigma(m_k(j)) \log(\sigma(m_k(j))) \tag{3}$$

$$\hat{y} = \begin{cases} \underset{k}{\operatorname{argmax}} \left(concat[\delta, \sigma(m_k(j))] \right) & \text{if } \mathcal{H}(j) < \epsilon \text{ and } \mathcal{F}(j) < 2 \\ ignore & \text{otherwise,} \end{cases} \tag{4}$$

where ϵ is a thresholding value. $S = \hat{y}$ when the predicted mask m is obtained from image embeddings without adaptation and $S' = \hat{y}$ when the adapter layers are applied. Finally, we generate the pseudo binary label $\mathcal{B}$ and the pseudo distance label $\mathcal{D}$ from $\mathcal{S}'$.

Optimization objective We leverage the strong representation of the SAM model to generate pseudo instance labels $\mathcal{S}$ from point annotations in order to train our nuclei segmentation model. However, the resulting instance map $\mathcal{S}$ contains a large number of pseudo masks m with regions of high entropy leading to big portions of the image to be ignored. Therefore we fine-tune the model via adapter layers to generate a more reliable pseudo instance map $\mathcal{S}'$. Specifically, the predicted mask m'_k, generated from adapted image embeddings and the point prompt p_k, is optimized to match the corresponding mask m_k in the pseudo label map $\mathcal{S}$. For this process, we apply binary cross entropy (BCE) and intersection over union (IOU) losses only to regions of low entropy in $\mathcal{S}$ and ignore others.

$$\mathcal{L}_{\mathcal{S}'} = \text{BCELoss}(m', \mathcal{S}) + \text{IOULoss}(m', \mathcal{S}). \tag{5}$$

As the learning process progresses, the domain gap is gradually closed allowing the adapted image embeddings to generate pseudo instance maps $\mathcal{S}'$ with lower entropy values. Thus, reducing the size of the ignore areas allows the model to use a larger number of pixels for foreground and distance pseudo labels that further improve the adapter layers by optimizing the nuclei decoder outputs.

The nuclei decoder predicts a binary foreground map $\mathcal{B}'$ and distance map $\mathcal{D}'$ (horizontal and vertical distance map) to detect all nuclei pixels in the image and identify boundaries between adjacent nuclei. The nuclei decoder shares most parameters for both tasks, except for input tokens related to each map and output MLP layers. To train the foreground map, we combine a BCE Loss with an IOU Loss while we apply the L1 loss to optimize the regression task of the distance maps.

$$\begin{aligned} \mathcal{L}_{\mathcal{B}} &= \text{BCELoss}(\mathcal{B}', \mathcal{B}) + \text{IOULoss}(\mathcal{B}', \mathcal{B}), \\ \mathcal{L}_{\mathcal{D}} &= \text{L1Loss}(\mathcal{D}', \mathcal{D}). \end{aligned} \tag{6}$$

Finally, our optimization objective is defined as: $\mathcal{L} = \lambda_{\mathcal{S}'}\mathcal{L}_{\mathcal{S}'} + \lambda_{\mathcal{B}}\mathcal{L}_{\mathcal{B}} + \lambda_{\mathcal{D}}\mathcal{L}_{\mathcal{D}}$ where $\lambda_{\mathcal{S}'}$, $\lambda_{\mathcal{B}}$ and $\lambda_{\mathcal{D}}$ are weighting coefficients.

Post-processing During inference, similar to Hover-Net[6], we obtain the instance segmentation result using only $\mathcal{B}'$ and $\mathcal{D}'$ from the nuclei decoder. $\mathcal{D}'$ passes through the Sobel operator [10] to calculate the gradient between pixels in order to identify individual instances, *i.e.*, high gradient values serve to distinguish adjacent instances as well as to locate the center of nuclei. Therefore, combining gradient maps, foreground and instance centers, we perform instance segmentation via the watershed algorithm. It is important to note that point annotations and the original SAM mask decoder are not necessary at inference time.

3 Experiments

Dataset Two independent datasets *i.e.*, CPM17 & MoNuSeg, are utilised for this study. CPM17, originates from digital pathology challenge [23], contains 64 H&E stained histopathology images with a total of 7,570 annotated nuclear boundaries. Their image size varies between 500×500 to 700×700. We split a total 64 images into 32/32 images for training and testing sets. MoNuSeg [11], The multi-organ nuclei segmentation dataset contains 30 H&E stained histopathology images from 7 different organs with a total of 21,623 individually annotated nuclear boundaries. Their image size is equal to 1000×1000. We split 30 images into 16/14 images for training and testing sets.

For the few-shot setting of the ablation study, we reduced the number of training images to four. Specifically, for MoNuSeg, we selected four images ensuring that each one represented one organ type to avoid duplication. Since CPM17 contains only the brain organ, we selected four training images randomly.

Implementation Details For training, we trained the model with a batch size of 4 using the AdamW optimizer [16], and learning rate of 1e-4 which updated through CosineAnnealingLR [15] with a $t_{\max}$ of 20. For pseudo labeling, δ was set to 0.5 and ϵ to 0.3, and for model training, $\lambda_{\mathcal{S}'}$, $\lambda_{\mathcal{B}}$, and $\lambda_{\mathcal{D}}$, were set to 1, 1, and 5, respectively. Finally, for augmentation, we employed techniques similar to conventional methods [13,19] including random resize, random affine transformation, horizontal flip, and random crop. All experiments were conducted with a single NVIDIA RTX A6000 GPU and PyTorch version 1.13.1 environment.

Main Results We compared our method (InstaSAM) with previously proposed various state-of-the-art (SOTA) methods [20,21,13,19,4]. All methods, excluding [4], was experimented under the same conditions as our method. [4] was developed assuming the availability of bounding boxes instead of points, and thus, experiments were conducted with bounding box annotations. Additionally, as this approach only predicted binary segmentation results, we applied the connected component technique to the binary segmentation results following MIDL [20] to obtain the instance segmentation. For evaluation, Dice and AJI are used as metrics. Note that AJI, proposed by Kumar et al. [11], is proved to be the most suitable metric to evaluate the object instance level segmentation performance compared to Dice. As shown in Table 1, InstaSAM surpassed the current SOTA [19] in CPM17 and MoNuSeg only by updating a limited num-

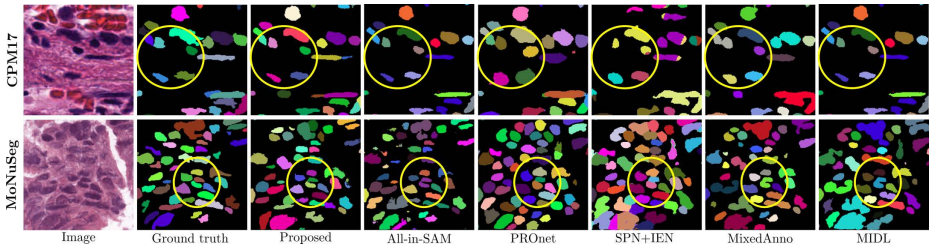

Fig. 2: Nuclei instance segmentation results on CPM17 and MoNuSeg data when nuclei points are centered. Only InstaSAM follows the ground truth shape of each individual nucleus, while other methods have tendency to segment into oval shape.

ber of parameters. In the CPM17 dataset, when the coordinates of point labels are provided at the center, InstaSAM showed a significant performance gain of +5.2% in Dice and +6.8% in AJI compared to PROnet. Despite All-in-SAM utilizing more informative bounding box labels which require stronger guidance, InstaSAM showed superior performance with a +3.2% in Dice and +5.3% in AJI. Also, even when provided with shifted point annotations, InstaSAM gained a performance gap of +3.7% in Dice and +4.5% in AJI compared to PROnet which trained with intact point annotation. Since All-in-SAM only utilizes bounding boxes, we recorded only no-shift condition results for both CPM and MoNuSeg. In MoNuSeg, a more challenging dataset, InstaSAM also surpassed the PROnet by a large margin of performance gap +2.2% in Dice and +1.9% in AJI and the All-in-SAM with +3.4% in Dice and +7.2% in AJI. These achievements are the result of model optimization with high-quality pseudo labels made by our proposed pseudo labeling process. Compared with All-in-SAM, we conclude that InstaSAM fully utilized point label information effectively. Moreover, given only a few-shot dataset, we outperformed the other models trained on whole data.

Fig. 2 shows the instance separation of densely clustered nuclei. Particularly, unlike methods that use cluster or Voronoi labels, which may miss elongated

Table 1: Comparison results of nuclei instance segmentation on two public datasets. Shift indicates how far pixel-wise shifted from each nucleus center.

	CPM17								MoNuSeg							
	shift0		shift2-4		shift4-6		shift6-8		shift0		shift2-4		shift4-6		shift6-8	
	Dice	AJI	Dice	AJI	Dice	AJI	Dice	AJI	Dice	AJI	Dice	AJI	Dice	AJI	Dice	AJI
MIDL[20]	75.0	55.5	75.3	56.9	74.4	53.7	72.2	49.9	70.1	44.9	69.9	45.0	66.3	39.9	61.0	31.5
Mixed Anno[21]	75.3	53.2	75.9	55.5	73.3	52.3	73.1	49.9	73.3	51.6	72.0	49.4	66.0	40.5	66.9	41.8
SPN+IEN[13]	74.3	54.3	72.9	52.1	70.1	47.9	69.4	46.8	74.0	53.4	72.3	50.4	69.1	46.5	65.6	39.4
PROnet[19]	78.7	62.7	78.2	61.8	77.4	60.7	77.0	60.2	75.0	55.5	74.8	54.8	73.3	53.2	72.5	50.9
All-in-SAM[4]	80.7	64.2	-	-	-	-	-	-	73.8	50.2	-	-	-	-	-	-
InstaSAM	**83.9**	**69.5**	**83.9**	**69.5**	**83.1**	**68.7**	**82.4**	**67.2**	**77.2**	**57.4**	**77.3**	**57.1**	**76.9**	**56.0**	**73.3**	**52.6**
InstaSAM$_{fewshot}$	83.0	68.5	83.0	68.4	81.7	65.9	80.5	65.6	75.2	54.1	76.0	54.7	75.5	55.0	70.2	49.8

nuclei shapes or segment them into round shapes, InstaSAM finds their true shape effectively. This qualitative result is the upholder of the superiority of our proposed multi-decoder structure and pseudo labeling process. More results can be found in the supplementary material.

Table 2: Comparison results on cross-domain settings

	MO→CPM17		CPM17→MO	
	Dice	AJI	Dice	AJI
MIDL[20]	71.2	51.7	69.1	45.3
Mixed Anno[21]	75.3	54.1	54.1	26.0
SPN+IEN[13]	72.4	54.0	65.6	41.2
PROnet[19]	76.4	58.7	72.0	51.2
All-in-SAM[4]	77.0	59.7	73.3	49.8
InstaSAM	**79.6**	**63.4**	**75.1**	**53.9**
InstaSAM$_{fewshot}$	79.6	62.0	74.3	52.4

Table 3: Effect of proposed pseudo labeling process

	CPM17		MoNuSeg	
	Dice	AJI	Dice	AJI
$\mathcal{C}+\mathcal{V}$	80.3	63.8	73.2	52.0
$\mathcal{S}$	81.3	66.2	Nan	Nan
${\mathcal{S}'}_{w/o\,\mathcal{H}}$	78.7	63.3	74.8	53.9
${\mathcal{S}'}_{w/o\,\mathcal{F}}$	64.0	45.8	59.5	39.9
$\mathcal{S}'$	83.9	69.5	77.2	57.4
Ground Turth	84.9	71.0	78.6	59.6

Results on Cross-Domain setting In Table 2, we evaluated the robustness of InstaSAM by cross-domain setting. When trained on MoNuSeg and evaluated on CPM17, InstaSAM still surpasses the other methods that were trained and evaluated both on CPM17. Also, the performance of InstaSAM trained on CPM17 and evaluated on MoNuSeg surpassed the other models that were trained and evaluated both on MoNuSeg in Dice. Compared to the cross-domain scenario of PROnet and All-in-SAM, InstaSAM still exhibits a large performance gap even in the few-shot scenario. These results indicate that InstaSAM preserved the generalization ability of SAM while effectively adapting the nuclei instance segmentation task.

Effect of proposed pseudo labeling process We further looked into the benefit of the proposed pseudo labeling process as shown in Table 3. To show the effectiveness of the pseudo instance map, we trained the models respectively with unrefined pseudo instance maps $\mathcal{S}$, and cluster and Voronoi labels $\mathcal{C}+\mathcal{V}$ which are mainly used in previous methods [20,21,13,19]. For the model with $\mathcal{C}+\mathcal{V}$, we employed $\mathcal{B}$ rather than $\mathcal{S}'$ to generate $\mathcal{D}$. We also test our models without $\mathcal{F}$ or $\mathcal{H}$ while maintaining $\mathcal{S}'$ to prove the necessity of each step of our pseudo labeling process. The results indicate that our model surpassed the performance of existing SOTA methods [19], regardless of the pseudo label type for both datasets. For the MoNuSeg dataset, due to the inconsistent shape of each nucleus, inaccurate cluster and Voronoi labels resulted in a serious performance drop. When $\mathcal{S}$ is used, the majority of areas exhibit high entropy values, resulting in a significant portion of the data being unusable for training purposes. In addition, when the instance map is refined without $\mathcal{H}$ or $\mathcal{F}$, an improper segmentation area still exists, which leads to poor performance. On the contrary, our proposed pseudo labeling process helped InstaSAM to achieve a performance close to the

model trained with ground truth annotation. This indicates that our pseudo labeling process is crucial and significantly enhances the learning process.

4 Conclusion

We introduce a novel weakly supervised approach, named InstaSAM, utilizing SAM to address nuclei instance segmentation. We proposed a new pseudo-labeling process into InstaSAM. By incorporating adapter layers into SAM's image encoder and tokens in the decoder, the performance of InstaSAM is well-fitted on the target domain while maintaining SAM's generalization ability. InstaSAM achieved state-of-the-art results on the CPM17 and MoNuSeg datasets. Extensive ablation studies verified the effectiveness of InstaSAM on few-shot, shifted points, and cross-domain settings.

Acknowledgment. This work was supported by IITP grant funded by the Korean government (MSIT) (No.2021-0-02068, Artificial Intelligence Innovation Hub) and (No.RS-2024-00439264, Development of High-Performance Machine Unlearning Technologies for Privacy Protection), Smart Health Care Program funded by the Korean National Police Agency (220222M01), the Digital Innovation Hub project supervised by the Daegu Digital Innovation Promotion Agency (DIP) grant funded by the Korea government (MSIT and Daegu Metropolitan City) in 2023 (DBSD1-01),

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

References

1. Alsubaie, N., Sirinukunwattana, K., Raza, S.E.A., Snead, D., Rajpoot, N.: A bottom-up approach for tumour differentiation in whole slide images of lung adenocarcinoma. In: Medical Imaging 2018: Digital Pathology. vol. 10581, pp. 104–113. SPIE (2018) 1
2. Chen, T., Zhu, L., Deng, C., Cao, R., Wang, Y., Zhang, S., Li, Z., Sun, L., Zang, Y., Mao, P.: Sam-adapter: Adapting segment anything in underperformed scenes. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 3367–3375 (2023) 3, 4
3. Chen, Z., Duan, Y., Wang, W., He, J., Lu, T., Dai, J., Qiao, Y.: Vision transformer adapter for dense predictions. arXiv preprint arXiv:2205.08534 (2022) 4
4. Cui, C., Deng, R., Liu, Q., Yao, T., Bao, S., Remedios, L.W., Tang, Y., Huo, Y.: All-in-sam: from weak annotation to pixel-wise nuclei segmentation with prompt-based finetuning. arXiv preprint arXiv:2307.00290 (2023) 2, 6, 7, 8
5. Deng, R., Cui, C., Liu, Q., Yao, T., Remedios, L.W., Bao, S., Landman, B.A., Wheless, L.E., Coburn, L.A., Wilson, K.T., et al.: Segment anything model (sam) for digital pathology: Assess zero-shot segmentation on whole slide imaging. arXiv preprint arXiv:2304.04155 (2023) 4

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. Medical image analysis **58**, 101563 (2019) 6
7. Hendrycks, D., Gimpel, K.: Gaussian error linear units (gelus). arXiv preprint arXiv:1606.08415 (2016) 4
8. 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. arXiv preprint arXiv:2306.15350 (2023) 2
9. Houlsby, N., Giurgiu, A., Jastrzebski, S., Morrone, B., De Laroussilhe, Q., Gesmundo, A., Attariyan, M., Gelly, S.: Parameter-efficient transfer learning for nlp. In: International Conference on Machine Learning. pp. 2790–2799. PMLR (2019) 4
10. Kanopoulos, N., Vasanthavada, N., Baker, R.L.: Design of an image edge detection filter using the sobel operator. IEEE Journal of solid-state circuits **23**(2), 358–367 (1988) 6
11. Kumar, N., Verma, R., Anand, D., Zhou, Y., Onder, O.F., Tsougenis, E., Chen, H., Heng, P.A., Li, J., Hu, Z., et al.: A multi-organ nucleus segmentation challenge. IEEE transactions on medical imaging **39**(5), 1380–1391 (2019) 6
12. 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) 1
13. Liu, W., He, Q., He, X.: Weakly supervised nuclei segmentation via instance learning. In: 2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI). pp. 1–5. IEEE (2022) 2, 3, 4, 6, 7, 8
14. Lloyd, S.: Least squares quantization in pcm. IEEE transactions on information theory **28**(2), 129–137 (1982) 2
15. Loshchilov, I., Hutter, F.: Sgdr: Stochastic gradient descent with warm restarts. arXiv preprint arXiv:1608.03983 (2016) 6
16. Loshchilov, I., Hutter, F.: Decoupled weight decay regularization. arXiv preprint arXiv:1711.05101 (2017) 6
17. Lu, C., Romo-Bucheli, D., Wang, X., Janowczyk, A., Ganesan, S., Gilmore, H., Rimm, D., Madabhushi, A.: Nuclear shape and orientation features from h&e images predict survival in early-stage estrogen receptor-positive breast cancers. Laboratory investigation **98**(11), 1438–1448 (2018) 1
18. 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) 2
19. Nam, S., Jeong, J., Luna, M., Chikontwe, P., Park, S.H.: Pronet: Point refinement using shape-guided offset map for nuclei instance segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 528–538. Springer (2023) 2, 3, 4, 6, 7, 8
20. Qu, H., Wu, P., Huang, Q., Yi, J., Riedlinger, G.M., De, S., Metaxas, D.N.: Weakly supervised deep nuclei segmentation using points annotation in histopathology images. In: International Conference on Medical Imaging with Deep Learning. pp. 390–400. PMLR (2019) 2, 3, 4, 6, 7, 8
21. Qu, H., Yi, J., Huang, Q., Wu, P., Metaxas, D.: Nuclei segmentation using mixed points and masks selected from uncertainty. In: 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI). pp. 973–976. IEEE (2020) 2, 4, 6, 7, 8
22. Tian, K., Zhang, J., Shen, H., Yan, K., Dong, P., Yao, J., Che, S., Luo, P., Han, X.: Weakly-supervised nucleus segmentation based on point annotations: A coarse-to-fine self-stimulated learning strategy. In: Medical Image Computing and Computer

Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part V 23. pp. 299–308. Springer (2020) 2
23. 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 p. 53 (2019) 6
24. Yoo, I., Yoo, D., Paeng, K.: Pseudoedgenet: Nuclei segmentation only with point annotations. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part I 22. pp. 731–739. Springer (2019) 2