Bridging RGB and Hematoxylin Components: An Interleaved Guidance and Fusion Framework for Point Supervised Nuclei Segmentation

Supplementary Material

1. Discussion and Visualization on IGA

To further examine the contribution of the proposed Interleaved Guidance Attention (IGA), we visualize four components: the RGB input, the ground-truth instance mask (GT), the raw prediction maps with IGA (IGA), and those without IGA (NIGA). As shown in Figure 1, incorporating IGA substantially enhances the model's ability to focus on nuclei regions under weak supervision.

Compared with the RGB input, the IGA responses suppress irrelevant background textures and staining noise, enabling the network to concentrate on coherent morphological structures. When aligned with GT, the IGA prediction maps exhibit stronger spatial compactness and clearer boundaries, reflecting a more faithful reconstruction of nuclear shapes and a better separation between adjacent instances. In contrast, NIGA predictions often suffer from dispersed activations, spurious responses on background regions, and incomplete emphasis on nuclei with irregular morphology.

These observations highlight the essential role of IGA in establishing robust cross-representation guidance. By enforcing complementary semantic cues and spatial consistency, IGA encourages the network to refine activations at nuclei centers while imposing additional structural constraints along boundaries. This mechanism effectively mitigates the instability caused by sparse point supervision and label noise, yielding more focused, consistent, and separable nuclei representations. Overall, IGA significantly improves the reliability of subsequent instance segmentation.

2. Efficiency and Performance Analysis

As shown in Figure 2, our method (Final) achieves the best overall performance across multiple metrics while maintaining moderate computational complexity. On the MO dataset [2], it records the highest Accuracy (91.38%), F1-score (79.03%), Dice_{obj} (74.89%), and AJI (54.85%), demonstrating the effectiveness of our dual-representation fusion mechanism in enhancing instance-level nuclei segmentation quality.

Compared with the lightweight D2E2 [7], our method yields substantial improvements, including a +3.9% gain in F1-score and a +4.9% increase in AJI. Although it requires considerably fewer FLOPs and parameters than the more complex SCNet [3] (46.18 G rather than 73.50 G FLOPs, and 65.1 M rather than 80.1 M parameters), it still surpasses SCNet across all evaluation metrics, representing a 37% reduction in computational cost while delivering superior accuracy.

Train→Test	Method	Acc	F1-score	Dice_{obj}	AJI
M→S	Ours	$90.23_{0.21}$	$65.59_{0.18}$	$60.97_{0.22}$	$36.30_{0.19}$
	MA	$88.95_{0.37}$	$62.42_{0.41}$	$58.15_{0.33}$	$34.21_{0.48}$
S→M	Ours	$88.57_{0.25}$	$67.62_{0.20}$	$56.34_{0.28}$	$38.35_{0.24}$
	MA	$86.70_{0.46}$	$64.75_{0.39}$	$51.91_{0.44}$	$32.80_{0.40}$
M→C	Ours	$87.96_{0.17}$	$69.83_{0.23}$	$67.62_{0.19}$	$48.24_{0.22}$
	MA	$89.45_{0.35}$	$71.02_{0.44}$	$69.85_{0.38}$	$49.13_{0.41}$
C→M	Ours	$88.65_{0.26}$	$69.98_{0.22}$	$66.89_{0.25}$	$46.37_{0.20}$
	MA	$84.36_{0.49}$	$66.15_{0.34}$	$64.78_{0.42}$	$43.92_{0.37}$
C→S	Ours	$86.28_{0.19}$	$48.87_{0.27}$	$47.82_{0.24}$	$25.86_{0.23}$
	MA	$83.96_{0.45}$	$40.65_{0.38}$	$44.76_{0.41}$	$22.51_{0.36}$
S→C	Ours	$86.38_{0.23}$	$63.15_{0.21}$	$60.88_{0.28}$	$39.25_{0.26}$
	MA	$85.18_{0.33}$	$60.21_{0.40}$	$54.50_{0.43}$	$35.50_{0.39}$

Table 1. Adaptive domain generalization results across three datasets. Each training-testing pair is evaluated using two methods: our adaptive fusion approach (Ours) and MixAnno (MA). Evaluation metrics include Accuracy (Acc), F1-score (F1), Dice_{obj} , and Aggregated Jaccard Index (AJI). M, S, and C denote the MO, CoNSeP, and CPM-17 datasets, respectively.

Relative to the mixed-annotation method MA [5], our approach achieves slightly higher accuracy (91.38% compared to 91.21%) and produces notable improvements in F1-score (+2.1%) and AJI (+3.2%), reflecting stronger robustness and generalization under point level supervision.

Overall, the proposed framework provides a well-balanced trade-off between segmentation precision and computational efficiency, which can be attributed to its cross-representation complementarity and adaptive fusion strategy.

3. Domain Generalization Performance

Table 1 presents the cross-domain generalization results among MO (M) [2], CoNSeP (S) [1], and CPM-17 (C) [6]. Our adaptive fusion framework consistently demonstrates superior or comparable performance across all training–testing configurations, highlighting its robustness to domain shifts in histopathological images. Specifically, when transferring from MO to CoNSeP (**M→S**), our method surpasses MixAnno [5] by **+2.8%** in F1-score and **+2.8%** in Dice, reflecting improved structural alignment under staining and morphological variations. Similarly, under the reverse setting (**S→M**), we observe gains of **+2.9 F1%** and **+4.4 Dice%**, confirming enhanced cross-

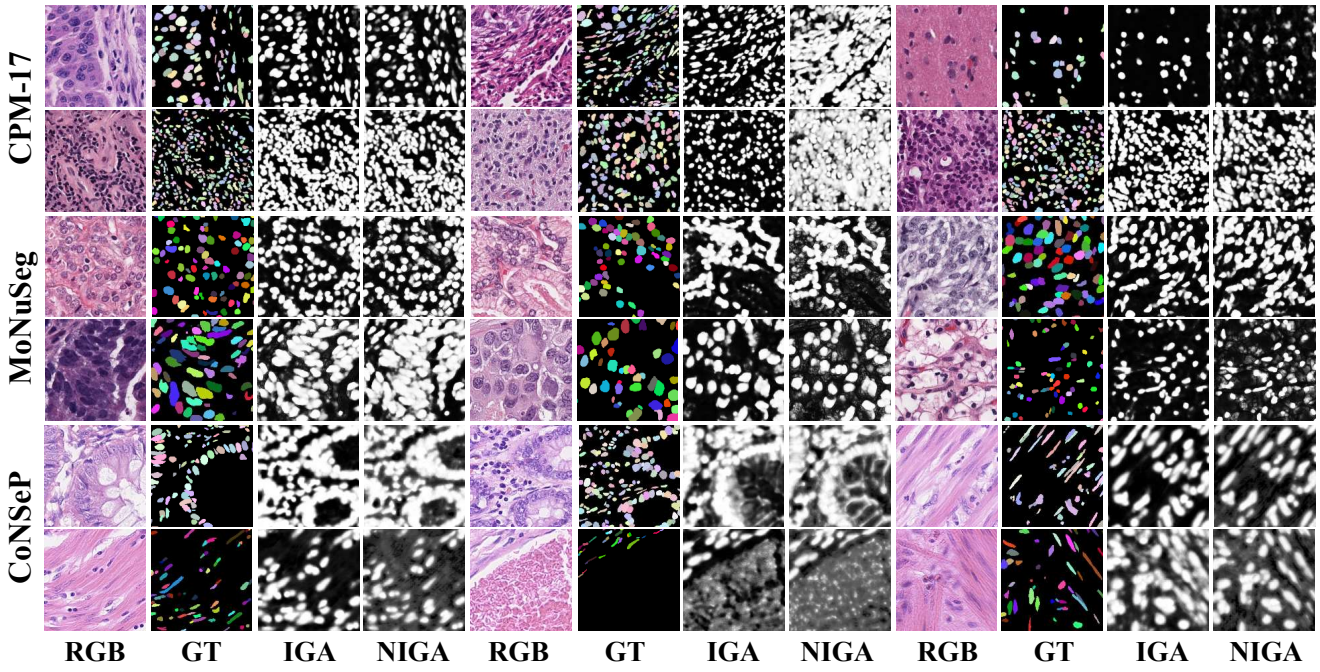


Figure 1. Visualization of IGA across the MoNuSeg (MO), CPM-17, and CoNSeP datasets. For each dataset, the four columns show: (1) RGB input image; (2) ground-truth (GT) instance annotations; (3) IGA prediction logits generated with the proposed IGA module; and (4) NIGA prediction logits obtained without IGA.

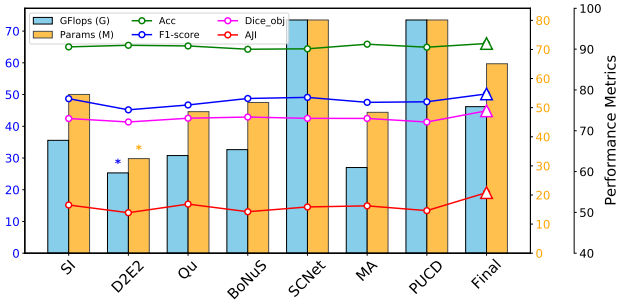


Figure 2. Comparison of different methods in terms of segmentation metrics, parameter counts, and computational complexity on the MO dataset. The left y-axis represents **GFLOPs** (blue), while the right y-axis indicates the number of **parameters** (orange). The best metric results are marked with a **triangle**, and the most efficient methods in terms of GFLOPs and parameters are highlighted with a **star**.

domain representation learning. Even in the most challenging transfer (**C→S**), our approach outperforms MixAnno by a significant margin (**+8.2% in F1-score** and **+3.1% in Dice**$_{obj}$), indicating effective adaptation to unseen data distributions. Although MixAnno slightly excels in **M→C** due to dataset-scale bias, our model maintains competitive results and exhibits more stable generalization, as evidenced by the smaller standard deviations. Overall, these results demonstrate that the proposed interleaved adaptive fusion strategy effectively leverages complementary cues between RGB and Hematoxylin representations, enabling strong domain generalization across diverse pathological datasets without retraining.

4. Complementarity Analysis

As discussed in the introduction, there are four key issues we will analyze: excessive staining, staining artifacts, " Two or one nuclei? " , and the challenge of separating nuclei from the background. These issues highlight the complementarity between RGB and H-channel methods, as each method has its own strengths and limitations. In this section, we analyze these issues in detail through the use of RGB and H-channel images.

4.1. Excessive Staining

Figure 3(B) illustrates the issue of excessive staining. In the original RGB images, over-staining leads to blurred cell boundaries or background regions being dyed, which causes incorrect segmentation of nuclei. Grayscale images, derived from RGB images, do not significantly improve upon this problem as they still suffer from over-staining effects, making it difficult to differentiate between nuclei and background. In contrast, the H-channel extraction method filters out background regions and enhances the contrast between the nuclei and background, improving nuclei separa-

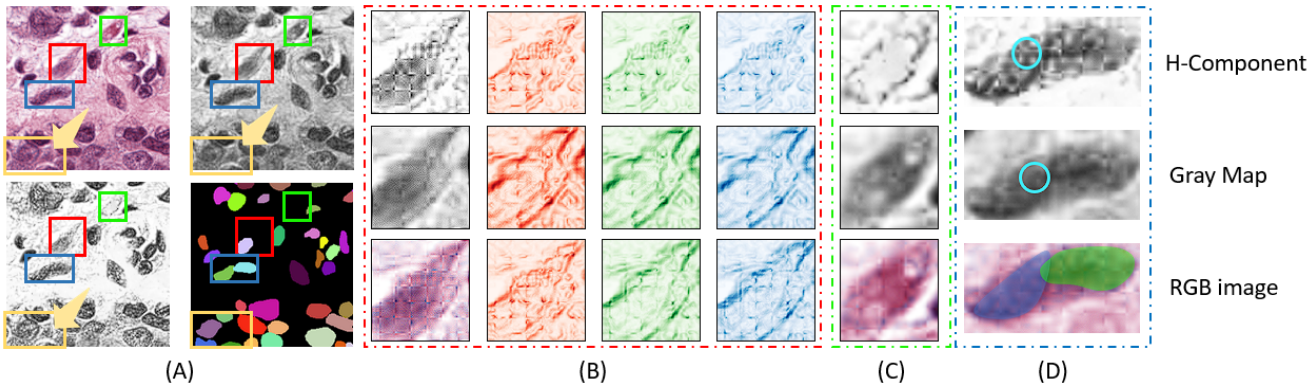


Figure 3. Analysis of image segmentation challenges and H-channel complementarity. (A) Visualizations of different image issues: top-left shows the original RGB image, top-right shows the corresponding grayscale image, bottom-left shows the H-channel image, and bottom-right shows the ground truth labels. The complementarity between RGB and H-channel is highlighted in the yellow box and arrow, particularly in regions where H-channel extraction struggles with background separation in dense nuclear areas. (B) Gradient analysis of excessive staining in RGB images. (C) Staining artifact issue, demonstrating residual staining in non-nuclear regions. (D) The "Two or One Nuclei?" issue, where weak boundaries or insufficient staining lead to difficulties in differentiating overlapping nuclei.

tion. However, in areas with severe staining artifacts, the H-channel may still incorrectly classify stained regions as nuclei, leading to segmentation errors.

4.2. Staining Artifact

Figure 3(C) shows the staining artifact problem. This issue originates from the uneven staining in the original RGB image, where non-nuclear areas still exhibit artificial staining. This residual staining impacts nuclei segmentation, misclassifying background regions as part of the nuclei. The H-channel method, by suppressing areas dominated by Eosin, greatly reduces such artifacts, correctly identifying them as background. Thus, H-channel extraction helps significantly in reducing staining artifacts and improving nuclei segmentation accuracy.

4.3. Two or One Nuclei?

In Figure 3(D), we show the "Two or one nuclei?" problem. In both RGB and grayscale images, weakly stained cell boundaries or insufficient staining may cause nuclei to merge or be hard to distinguish. This results in misclassification, such as a single nucleus being misidentified as two or vice versa. Through gradient analysis, we show that the H-channel method enhances the contrast between nuclei and background, assisting in separating overlapping nuclei. However, in cases of severe overlap or unclear boundaries, the H-channel method may still misclassify multiple nuclei as a single instance, leading to blurred nuclei boundaries and inaccurate segmentation.

4.4. Background Separation in Dense Nuclei Regions

Finally, as shown in the yellow box and arrow in Figure 3(A), the H-channel method struggles to separate nuclei from the background in densely packed nuclear regions. The staining in these areas may spill over into the background, leading to staining accumulation. In the H-channel image, this accumulation appears as falsely segmented nuclei, where surrounding regions form what seems like a mound of staining. Therefore, to accurately identify such areas, it is necessary to analyze the original RGB image alongside the H-channel image. This highlights the complementarity between RGB and H-channel methods, as each method excels in different areas and compensates for the other's shortcomings.

4.5. Conclusion

The RGB and H-channel methods each have their own strengths and weaknesses. RGB images provide rich color information and retain structural context, but are prone to over-staining and background interference, which can cause boundary blurring and staining artifacts. H-channel extraction excels at enhancing contrast and suppressing background interference, but it struggles in densely stained regions and may misinterpret overlapping nuclei as a single instance. Therefore, the complementarity between RGB and H-channel methods is critical for improving nuclei segmentation accuracy. By combining both methods, we can leverage their respective strengths and mitigate their weaknesses, leading to more robust and accurate segmentation results.

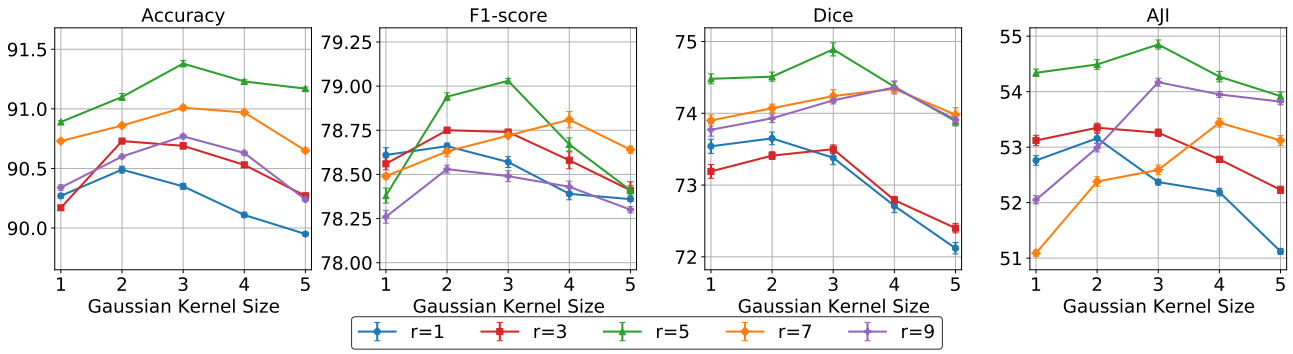


Figure 4. Line chart illustrating the effect of Gaussian kernel radius on four key metrics. Here, r denotes the radius setting, and different colors indicate metric performance under different radius settings.

M	L	S	R	Dice_{obj}	AJI	DQ	SQ	PQ
✓				$74.29_{0.22}$	$54.41_{0.19}$	$69.31_{0.09}$	$71.39_{0.17}$	$51.19_{0.24}$
	✓			$74.38_{0.18}$	$54.36_{0.21}$	$69.43_{0.20}$	$71.25_{0.15}$	$51.18_{0.19}$
		✓		$74.56_{0.25}$	$54.56_{0.13}$	$69.42_{0.17}$	$71.72_{0.18}$	$51.21_{0.25}$
			✓	$74.55_{0.20}$	$54.58_{0.16}$	$69.48_{0.22}$	$71.94_{0.19}$	$51.46_{0.26}$
✓			✓	$74.69_{0.19}$	$54.57_{0.24}$	$70.33_{0.15}$	$72.26_{0.37}$	$51.39_{0.27}$
	✓	✓	✓	$74.72_{0.13}$	$54.69_{0.10}$	$70.34_{0.16}$	$72.38_{0.18}$	$51.38_{0.25}$
✓	✓	✓	✓	$\mathbf{74.89}_{0.11}$	$\mathbf{54.85}_{0.34}$	$\mathbf{70.65}_{0.17}$	$\mathbf{72.54}_{0.29}$	$\mathbf{51.45}_{0.13}$

Table 2. Ablation study of each component on the MO dataset. M, L, S, and R denote different modules. Metrics include Dice_{obj}, AJI, DQ, SQ, and PQ (%).

5. Impact of Kernel Size and Radius

As shown in Figure 4, the *Gaussian kernel size* denotes the spatial extent of the kernel applied during Gaussian blurring, while the *point radius* (denoted as r in the legend) specifies the radius assigned to each annotated point prior to blurring.

Empirically, we observe that moderate kernel sizes yield more stable and accurate instance segmentation results. Smaller kernels are insufficient to suppress annotation noise, whereas excessively large kernels tend to oversmooth structural boundaries and degrade fine-grained localization. A kernel size within the range of 3–5 offers an effective balance between noise attenuation and structural fidelity.

Regarding the point radius, increasing r from 1 to 5 consistently enhances segmentation performance by providing richer contextual cues around annotated centers. However, an overly large radius (e.g., $r = 9$) introduces spatial ambiguity, and the localization accuracy slightly deteriorates.

Overall, combining a moderate Gaussian kernel with a relatively large point radius achieves an optimal trade-off between contextual aggregation and spatial precision in weakly supervised nuclei instance segmentation.

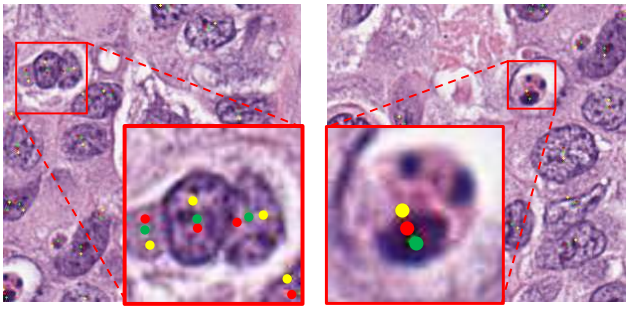
6. Ablation Analysis of RCDF Components

Table 2 presents the ablation results of different components within the RCDF module, including the multi-scale feature-aware units (M), the lightweight global context-aware scale modulator (L), the sequential attention (S), and the residual guided fusion (R). Individually adding each component leads to consistent improvements across all metrics, demonstrating their effectiveness in enhancing feature representation and fusion. In particular, introducing the residual guided fusion (R) yields a notable gain in Dice_{obj} (from 74.29 to 74.55) and AJI (from 54.41 to 54.58), indicating that the residual pathway helps preserve fine-grained structural information. Combining M and R further enhances the discriminative ability of the feature space. When all four components (M, L, S, and R) are integrated, our complete RCDF achieves the highest performance, with Dice_{obj} = 74.89, AJI = 54.85, DQ = 70.65, SQ = 72.54, and PQ = 51.45, confirming the complementary roles of each sub-module in achieving robust and accurate feature alignment.

7. Metrics Definition

Given that nuclei segmentation inherently targets instance-level predictions, the evaluation must assess how well individual nuclei are delineated rather than focusing solely on pixel-level accuracy. Therefore, two complementary object-centric metrics are employed: the *instance-level Dice coefficient* (D_{inst}) and the *Aggregated Jaccard Index* (AJI) [4]. The first metric, $Dice_{obj}$, quantifies the agreement between matched nuclei pairs and is defined as:

$$Dice_{obj}(Y,P) = \frac{1}{2}\sum_{i=1}^{N_Y} \alpha_i \, D(Y_i, \tilde{P}(Y_i)) + \frac{1}{2}\sum_{j=1}^{N_P} \beta_j \, D(\tilde{Y}(P_j), P_j), \quad (1)$$



(a) Inside nuclei (b) Outside nuclei

Figure 5. The figure illustrates simulated clinical annotation scenarios with different pixel offset distances. (a) shows all offset points located inside the nuclei, while (b) shows cases where some offset points fall outside the nuclei.

where $Y = \{Y_1, \ldots, Y_{N_Y}\}$ and $P = \{P_1, \ldots, P_{N_P}\}$ denote the sets of ground-truth and predicted nuclei, respectively. For each ground-truth nucleus Y_i, $\tilde{P}(Y_i)$ represents the predicted region that maximizes the overlap with Y_i, while $\tilde{Y}(P_j)$ indicates the ground-truth region having the greatest overlap with P_j. The weights α_i and β_j are proportional to the object areas. A one-to-one correspondence is established only if the overlapping region exceeds 50%. This formulation jointly evaluates prediction completeness and precision by averaging the Dice similarity across both annotation and prediction domains.

The *Aggregated Jaccard Index* (AJI) provides a complementary, global view of segmentation quality and is formulated as:

$$AJI = \frac{\sum_{i=1}^{N_Y} |Y_i \cap P^*(Y_i)|}{\sum_{i=1}^{N_Y} |Y_i \cup P^*(Y_i)| + \sum_{r \in \mathcal{R}} |P_r|}, \qquad (2)$$

where $P^*(Y_i)$ is the predicted nucleus exhibiting the highest Jaccard similarity with Y_i, and $\mathcal{R}$ is the set of unmatched predictions that do not correspond to any labeled instance. The numerator aggregates all intersection areas of valid matches, whereas the denominator accounts for both the union regions of matched pairs and the areas of remaining unmatched predictions. Consequently, J_{agg} serves as a holistic indicator that captures the overall consistency between predicted and ground-truth nuclei across the entire image.

8. Point Shift Visualization

Figure 5 illustrates the visualization of the simulated point-shift experiment designed to mimic real-world annotation deviations. Red dots represent randomly sampled points located 4 pixels away from the true nuclear centers (green dots), while yellow dots denote those 8 pixels away. In Figure 5 (a), all simulated points remain within the nuclear boundaries, which corresponds to the ideal annotation scenario. In contrast, Figure 5 (b) includes yellow points falling outside nuclei, simulating mislabeled cases commonly encountered in practice and testing the robustness of our method. As shown in the main paper (Figure 6), even when roughly one-quarter of the annotations are perturbed, our method still achieves strong performance. This robustness mainly stems from the complementary dual-representation design, which mitigates noise introduced by inaccurate annotations and reinforces nuclear features under both structural and semantic representations, leading to more stable instance delineation.

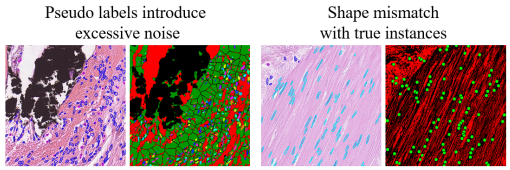


Figure 6. Visualization of the limitations of clustering-based pseudo labels on the CoNSeP dataset.

9. Limitation on the CoNSeP Dataset

As illustrated in Fig. 6, the CoNSeP dataset has relatively low image resolution, which makes the generation of clustering-based pseudo labels more prone to over-clustering and noise. In addition, pseudo labels produced by clustering methods tend to form compact and near-circular regions. However, many nuclei in the CoNSeP dataset exhibit elongated morphologies. This discrepancy between the pseudo-label shapes and the true nuclear structures can degrade the quality of the generated supervision and introduce additional noise during training. Such limitations are commonly encountered in weakly supervised nuclei instance segmentation when pseudo labels are derived from clustering-based strategies.

References

[1] Simon Graham, Quoc Dang Vu, Shan E Ahmed Raza, Ayesha Azam, Yee Wah Tsang, Jin Tae Kwak, and Nasir Rajpoot. Hover-net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images. *Medical image analysis*, 58:101563, 2019. 1

[2] Neeraj Kumar, Ruchika Verma, Sanuj Sharma, Surabhi Bhargava, Abhishek Vahadane, and Amit Sethi. A dataset and a technique for generalized nuclear segmentation for computational pathology. *IEEE transactions on medical imaging*, 36 (7):1550–1560, 2017. 1

[3] Yi Lin, Zhiyong Qu, Hao Chen, Zhongke Gao, Yuexiang Li, Lili Xia, Kai Ma, Yefeng Zheng, and Kwang-Ting Cheng. Nuclei segmentation with point annotations from pathology images via self-supervised learning and co-training. *Medical Image Analysis*, 89:102933, 2023. 1

[4] Hui Qu, Pengxiang Wu, Qiaoying Huang, Jingru Yi, Zhen-

nan Yan, Kang Li, Gregory M Riedlinger, Subhajyoti De, Shaoting Zhang, and Dimitris N Metaxas. Weakly supervised deep nuclei segmentation using partial points annotation in histopathology images. *IEEE transactions on medical imaging*, 39(11):3655–3666, 2020. 4

[5] Hui Qu, Jingru Yi, Qiaoying Huang, Pengxiang Wu, and Dimitris Metaxas. Nuclei segmentation using mixed points and masks selected from uncertainty. In *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, pages 973–976. IEEE, 2020. 1

[6] Quoc Dang Vu, Simon Graham, Tahsin Kurc, Minh Nguyen Nhat To, Muhammad Shaban, Talha Qaiser, Navid Alemi Koohbanani, Syed Ali Khurram, Jayashree Kalpathy-Cramer, Tianhao Zhao, et al. Methods for segmentation and classification of digital microscopy tissue images. *Frontiers in bioengineering and biotechnology*, 7:53, 2019. 1

[7] Kunzi Xie, Haonan Zhong, Jiamin Chang, Maurice Pagnucco, and Yang Song. D2e2-net: Double deep edge enhancement for weakly-supervised cell nuclei segmentation with incomplete point annotations. In *2022 International Conference on Digital Image Computing: Techniques and Applications (DICTA)*, pages 1–8. IEEE, 2022. 1