# Dynamic Pseudo Label Optimization in Point-Supervised Nuclei Segmentation

Ziyue Wang[1,2]*, Ye Zhang[1]*, Yifeng Wang[3]*, Linghan Cai[1], and Yongbing Zhang[1](✉)
{200111326, zhangye94, wangyifeng,cailh}@stu.hit.edu.cn,
ybzhang08@hit.edu.cn

[1] School of Computer Science and Technology, Harbin Institute of Technology (Shenzhen),
[2] Department of Electrical and Computer Engineering, National University of Singapore,
[3] School of Science, Harbin Institute of Technology (Shenzhen).

**Abstract.** Deep learning has achieved impressive results in nuclei segmentation, but the massive requirement for pixel-wise labels remains a significant challenge. To alleviate the annotation burden, existing methods generate pseudo masks for model training using point labels. However, the generated masks are inevitably different from the ground truth, and these dissimilarities are not handled reasonably during the network training, resulting in the subpar performance of the segmentation model. To tackle this issue, we propose a framework named DoNuSeg, enabling **D**ynamic pseudo label **O**ptimization in point-supervised **Nu**clei **Seg**mentation. Specifically, DoNuSeg takes advantage of class activation maps (CAMs) to adaptively capture regions with semantics similar to annotated points. To leverage semantic diversity in the hierarchical feature levels, we design a dynamic selection module to choose the optimal one among CAMs from different encoder blocks as pseudo masks. Meanwhile, a CAM-guided contrastive module is proposed to enhance the accuracy of pseudo masks further. In addition to exploiting the semantic information provided by CAMs, we consider location priors inherent to point labels, developing a task-decoupled structure for effectively differentiating nuclei. Extensive experiments demonstrate that DoNuSeg outperforms state-of-the-art point-supervised methods. The code is available at https://github.com/shinning0821/MICCAI24-DoNuSeg.

**Keywords:** Nuclei Instance Segmentation · Point-supervised Segmentation · Pseudo Label Optimization · Class Activation Map.

## 1 Introduction

Nuclei segmentation in whole-slide images (WSIs) is crucial for uncovering tumor microenvironment and thus informing relevant decisions in disease treatment

---

* Ziyue Wang, Ye Zhang, and Yifeng Wang contributed equally.

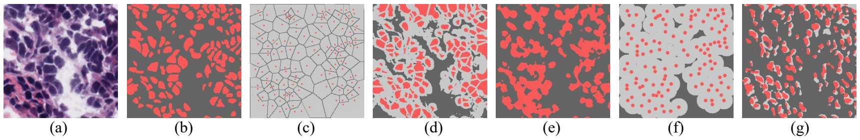

Fig. 1: (a) image input; (b) ground truth; (c) Voronoi label; (d) cluster label; (e) LSM label; (f) our initial label $M$; (g) our optimized label $P$. In (b)-(g), red, dark gray and light gray pixels denote nuclei, background and ignored areas, respectively.

[13,16,27,7]. Recently, deep learning techniques [19,15,21,3,8,2] have promoted nuclei segmentation. However, the success of the segmentation algorithms is contingent on the availability of high-quality imaging data with corresponding pixel-wise labels provided by experts. The annotating process is time-consuming and labor-intensive, limiting the development of models. Meanwhile, point labels that annotate nuclei with single points effectively reduce the annotation cost, making it essential to develop point-supervised segmentation methods.

Existing point-supervised methods [12,28] generally adopt a two-stage framework, first utilizing the biological morphology of nuclei to generate pixel-wise pseudo masks, then training the segmentation model. For precise nuclei segmentation, current research investigates various approaches to improve the quality of the pseudo masks. As shown in Figure 1(c)(d), [18,22] integrates the Voronoi diagram for mask generation, which considers the distance between points to distinguish overlapping instances and then generates cluster labels in separate regions. [26] develops a level-set method (LSM) to consider the nuclei's topology further as shown in Figure 1(e). However, these algorithms inevitably bring noisy labels due to the enormous variation in nuclei shape, color, and distribution. Meanwhile, these methods lack effective solutions for handling inaccurate labels. This oversight may impair the model training, leading to insufficient nucleus feature representation. To this end, devising reliable pseudo label enhancement in the training phase is critical for point-supervised nuclei segmentation.

Out of the advantages of targeting class-related areas, class activation maps (CAMs) are widely adopted for weakly supervised segmentation methods under natural scenes [20,29]. During training, CAMs are gradually optimized and can increasingly localize foreground regions. Therefore, we argue that CAMs have the potential to serve as pseudo-labels. However, the direct application of CAMs for labels encounters great challenges. Firstly, nuclei are densely distributed in pathological images, while CAMs tend to capture the most salient regions, resulting in frequently missed detection. Secondly, nuclei present low contrast with the surrounding tissue, posing difficulty in determining instance boundaries using low-resolution CAMs. Therefore, it is imperative to enhance the quality of CAM generation and address the limitations of CAMs in distinguishing instances.

Motivated by the above discussions, this paper presents a **D**ynamic pseudo label **O**ptimization method in point-supervised **Nu**clei **Seg**mentation (DoNuSeg). DoNuSeg takes advantage of the location priors provided by point labels and the semantic-level representation of CAMs, decoupling nuclei instance segmentation into object detection and semantic segmentation. To alleviate the miss-detection problem, we develop a novel Dynamic CAM Selection (DCS) module that incorporates the hierarchical features of the encoder for CAM generation, enriching the nucleus-related features and obtaining more activated foreground areas. To suppress the CAMs' uncertainty, DoNuSeg adopts a CAM-guided contrastive learning (CCL) module highlighting the representation differences between nuclei and the surrounding tissues, thereby accurately distinguishing nuclei boundaries. Overall, our contributions can be summarized as following aspects:

- We propose a novel weakly supervised nuclei instance segmentation framework termed DoNuSeg, which effectively leverages CAMs to achieve dynamic optimization of the pseudo label.
- We develop a pseudo label optimizing method that measures the accuracy of CAMs generated by different feature levels and adaptively selects the optimal CAM for label generation.
- We integrate a contrastive learning approach that utilizes the location information provided by points to widen the gap between nuclei and tissues, refining the feature representation and improving CAMs' location accuracy.
- Extensive experiments demonstrate the superiority of our method, outperforming state-of-the-art methods on three public datasets.

## 2 Methods

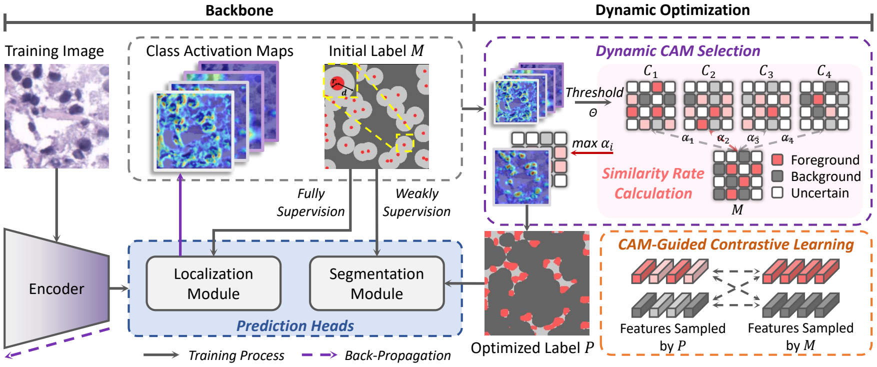

Fig. 2: Overview of our DoNuSeg method, which utilizes a Dynamic CAM Selection (DCS) module and a CAM-guided Contrastive Learning (CCL) module to dynamically select and optimize pseudo labels.

Our DoNuSeg develops a dynamic pseudo label optimization method to solve the challenges in point-supervised nuclei segmentation by the proposed DCS and CCL module as shown in Figure 2. To further utilize the location prior, we also take a task-decoupled structure as shown in Figure 3, which combines the detection task and semantic segmentation task to achieve instance segmentation.

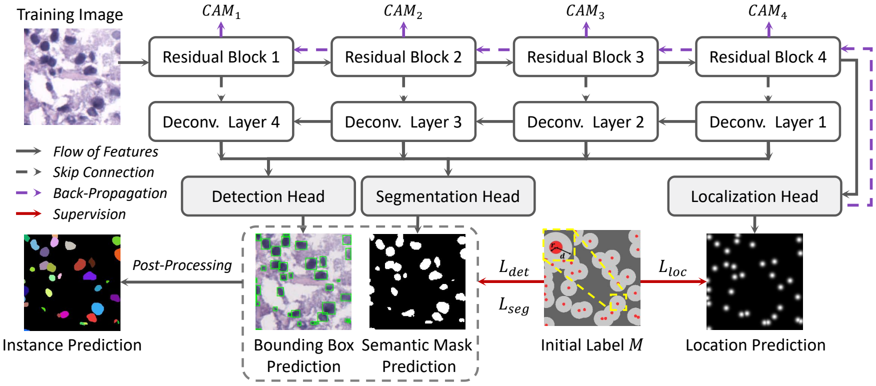

Fig. 3: Backbone structure of DoNuSeg. The detection and segmentation head takes hierarchical feature levels in the decoder as their input.

### 2.1 Backbone

Point annotations are challenging for pixel-wise segmentation but can be utilized to train fully supervised agent tasks to generate CAMs. CAMs offer valuable insights into the model's focus on crucial foreground regions, providing valuable guidance for training segmentation networks. However, CAMs only capture semantic information and require additional assistance differentiating individual instances. To address this limitation, we propose a decoupled instance segmentation method, as illustrated in Figure 3. Our approach leverages the positional priors obtained from point annotations to accurately predict bounding boxes, facilitating the distinction of instances within the semantic masks.

We take FPN [11] as the backbone with a ResNet50 [6] encoder. The decoder has a shared detection and segmentation head for each feature level. For the detection head, our design is based on the efficient detector FCOS [23] while the segmentation head is composed of four convolutional layers. The predicted bounding boxes are used to distinguish instances from the semantic masks.

We first calculate the pseudo bounding box for point labels following dense object detection in natural scenarios [24], which is used to compute the loss of detection heads $L_{det}$ following FCOS [23] (details seen in the supplementary materials). As shown in Figure 1(f), we obtain the initial label $M$ for segmentation by assuming that pixels within $r$ units around the point labels are foreground and pixels more than $d$ units away from the point labels are background. Other pixels are ignored while training to avoid introducing noise. The segmentation loss is computed as follows:

$$\mathcal{L}_{seg} = -\frac{1}{|\Omega_M|}\sum_{i\in\Omega_M}[(1-M_i)log(1-Y_i)+M_ilog(Y_i)], \quad (1)$$

where $Y_i$ and $M_i$ denote segmentation prediction $Y$ and initial label $M$ at the $i$th pixel, and $\Omega_M$ is the set of non-ignored pixels in $M$. For the generation of

CAMs, a localization head following the encoder is employed to conduct fully supervised point localization, which consists of three fully connected layers and is trained by an MSE loss $\mathcal{L}_{loc}$. The total loss function is computed as follows:

$$\mathcal{L} = \mathcal{L}_{det} + \mathcal{L}_{seg} + \mathcal{L}_{loc} + \omega_1 \mathcal{L}_{dcs} + \omega_2 \mathcal{L}_{ccl}, \tag{2}$$

where $\mathcal{L}_{dcs}$ and $\mathcal{L}_{ccl}$ are the loss of the proposed DCS and CCL module and will be introduced in the following subsections. $\omega_1$ and $\omega_2$ are hyperparameters.

### 2.2 Dynamic CAM Selection Module

As shown in Figure 2, CAMs can reflect the attention area of the encoder and exhibit different semantic information depending on specific layers. Previous studies [20,29] primarily utilize CAMs generated by the encoder's last layer, which has limited coverage of the foreground regions and coarse-grained boundaries resulting from upsampling from the small-size deep feature map. However, the encoder's intermediate layers' CAMs, which capture more nuclei and can provide fine-grained information, are often ignored. Therefore, the DCS module is utilized to choose the proper CAM dynamically. As shown in Figure 2, we first filter the generated CAM by a threshold $\theta$ to get a binary map $C$:

$$C_i = \begin{cases} 1, & CAM_i > \theta, \\ 0, & CAM_i \leq 1 - \theta, \\ \text{ignored}, & 1 - \theta < CAM_i \leq \theta, \end{cases} \tag{3}$$

where $CAM_i$ and $C_i$ denote CAM and $C$ at the $i$th pixel. The similarity rate $\alpha$ of $C$ is defined as follows:

$$\alpha = \frac{1}{|\Omega_M|} \sum_{i \in \Omega_M} M_i C_i + \bar{M}_i \bar{C}_i, \tag{4}$$

where $M_i$ denote the initial label $M$ at the $i$th pixel, and $\Omega_M$ is the set of non-ignored pixels in $M$. We choose $C$ with the maximum $\alpha$ as the optimized pseudo label $P$ to dynamically supervise segmentation network training by:

$$\mathcal{L}_{dcs} = -\frac{\alpha_P}{|\Omega_P|} \sum_{i \in \Omega_P} [(1 - P_i) log(1 - Y_i) + P_i log(Y_i)], \tag{5}$$

where $P_i$ and $Y_i$ denotes $P$ and the segentation prediction $Y$ at the $i$th pixel, $\alpha_P$ is the similarity rate of $P$, and $\Omega_P$ is the set of non-ignored pixels in $P$.

### 2.3 CAM-guided Contrastive Learning Module

To better segment nuclei boundaries, we propose a CAM-guided Contrastive Learning module to enhance intra-class coherence and inter-class discrimination of nuclei and background features. As described in Section 2.1, the initial pseudo label $M$ hardly contains noise, thus by aligning the CAM's attention regions with the initial label $M$, CAM concentrates more on nuclei and less on the background, which enhances the accuracy of CAM. Details are described below:

We use a projector $g$ which consists of four $3 \times 3$ convolutional layers and a three-layer MLP to preserve critical contextual information following [1]. The features outputted by the encoder are enhanced by $g$ and then upsampled to the size of the original image as the enhanced feature map $Z$. Let $Z_i$, $M_i$, $P_i$ denote $Z$, the initial Label $M$ and the optimized label $P$ at the $i$th pixel respectively, nuclei and background feature sets are defined as $\mathcal{F}^+ = \{Z_i \in Z | M_i = 1\}$ and $\mathcal{F}^- = \{Z_i \in Z | M_i = 0\}$. The anchor features $a^+$ and $a^-$ are computed by the average of $\mathcal{F}^+$ and $\mathcal{F}^-$, which can be regarded as the ground truth for nuclei and background features.Positive and negative feature sets can be sampled by $P$: $\mathcal{S}^+ = \{Z_i \in Z | P_i = 1\}$ and $\mathcal{S}^- = \{Z_i \in Z | P_i = 0\}$. The pixel-wise contrastive learning loss is computed by:

$$\mathcal{L}_{ccl} = \alpha_P(\mathcal{L}_{con}(a^+, S^+, S^-) + \mathcal{L}_{con}(a^-, S^-, S^+)), \tag{6}$$

where $\alpha_P$ is the similarity rate of $P$. $\mathcal{L}_{con}$ is defined as follows:

$$\mathcal{L}_{con}(q, U, V) = -\frac{1}{|U|} \sum_{u \in U} [\phi(q, u)/\tau - \log(e^{\phi(q,u)/\tau} + \sum_{v \in V} e^{\phi(q,v)/\tau})], \tag{7}$$

where $\tau$ is a hyperparameter and $\phi$ denotes the consine similarity, $q$ is an anchor feature and $U, V$ are the similar and dissimilar feature sets. The proposed module can make the features of the foreground and background pixels in $P$ closer to the corresponding anchor features, thus enhancing the feature representation.

## 3 Experiments and Results

Table 1: Performance comparison (%) with SOTA point-supervised methods. The best performance is shown in **bold**, and the second is underlined.

| Datasets | Methods | DICE | AJI | DQ | SQ | PQ |
|---|---|---|---|---|---|---|
| CryoNuSeg | WeakSeg [18] | 59.08±2.27 | 29.07±1.81 | 34.97±1.15 | 62.91±1.42 | 23.78±1.51 |
| | PseudoEdgeNet [25] | 60.42±1.71 | 36.78±2.17 | 35.03±3.32 | 63.10±1.53 | 22.37±1.29 |
| | MaskGA-Net [5] | 65.94±2.85 | 40.13±0.73 | 41.18±1.81 | **67.47**±0.89 | 28.19±1.13 |
| | DDTNet [26] | **68.32**±1.97 | 34.05±0.93 | 41.83±0.81 | 66.86±1.40 | 27.91±2.17 |
| | SC-Net [12] | 63.17±0.74 | 38.63±1.28 | 38.82±0.44 | 65.32±1.34 | 25.43±2.03 |
| | **DoNuSeg** (Ours) | 67.22±1.30 | **44.08**±0.77 | **46.41**±1.67 | 65.74±1.31 | **30.58**±1.53 |
| CoNSeP | WeakSeg [18] | **63.32**±1.16 | 33.41±1.34 | 34.97±2.10 | 64.91±1.60 | 23.17±0.58 |
| | PseudoEdgeNet [25] | 33.07±2.24 | 22.07±1.10 | 14.38±1.82 | 52.54±0.67 | 15.26±1.11 |
| | MaskGA-Net [5] | 28.69±1.61 | 20.70±1.55 | 19.97±0.36 | 52.37±2.03 | 18.81±1.29 |
| | DDTNet [26] | 58.82±1.29 | 29.58±2.18 | 28.33±0.25 | **65.20**±1.77 | 20.67±0.43 |
| | SC-Net [12] | 49.95±1.55 | 31.70±0.43 | 27.77±0.17 | 57.12±0.76 | 21.58±2.05 |
| | **DoNuSeg** (Ours) | 60.20±2.10 | **36.39**±1.77 | **36.68**±0.74 | 62.33±1.83 | **23.95**±0.99 |
| TNBC | WeakSeg [18] | 66.22±0.33 | 44.82±1.91 | 50.45±0.37 | 67.91±1.58 | 33.05±2.34 |
| | PseudoEdgeNet [25] | 48.27±0.40 | 32.83±1.38 | 30.31±2.74 | 54.76±0.72 | 20.26±1.92 |
| | MaskGA-Net [5] | 55.01±2.07 | 36.16±1.09 | 40.26±1.40 | 56.24±2.45 | 27.13±1.98 |
| | DDTNet [26] | **67.88**±1.59 | 47.83±2.01 | 58.56±0.57 | 66.79±2.26 | 40.08±0.76 |
| | SC-Net [12] | 63.24±1.91 | 44.75±1.92 | 53.62±1.21 | 58.38±0.39 | 35.29±0.67 |
| | **DoNuSeg** (Ours) | 63.99±1.18 | **50.06**±0.94 | **58.77**±0.69 | **68.75**±1.34 | **40.64**±0.54 |

### 3.1 Datasets and Metrics

We evaluate the proposed method on three public datasets, namely CryoNuSeg [14], ConSeP [4], and TNBC [17]. CryoNuSeg contains 30 images sampled from 10 organ tissues with the size of $512 \times 512$. ConSeP includes 41 images sampled from colon patients with the size of $1000 \times 1000$. TNBC consists of 50 images from 11 breast cancer patients with a size of $512 \times 512$. Datasets are divided into training, validation, and test sets in a ratio of 3:1:1. All images are then cropped into $256 \times 256$ sized patches with an overlapping of 128 pixels.

We adopt five widely used metrics for quantitative evaluation: DICE, Aggregated Jaccard Index (AJI) [10], Detection Quality (DQ), Segmentation Quality (SQ), and Panoptic Quality (PQ) [9]. The higher value is better for these metrics. To avoid randomness, we adopt 5-fold cross-validation and report the average values and the standard deviation in the testing set.

### 3.2 Implementation Details

Our experiments are implemented on PyTorch 1.10.0 using an Nvidia RTX 3090 GPU. We adopt an SGD optimizer for model training with a learning rate of 0.01, a momentum of 0.9, and a weight decay of 0.0005. Each model is trained for up to 40 epochs with a mini-batch size of 8. We set hyperparameters $r = 4$, $d = 20$, $\tau = 1$, $\omega_1 = 0.5$, $\omega_2 = 2$, and $\theta = 0.8$. Online data augmentation is employed including random flipping, random rotation, and random cropping.

### 3.3 Comparisons with State-of-the-art Methods

We compare DoNuSeg with SOTA point-supervised segmentation methods: WeakSeg [18], PseudoEdgeNet [25], MaskGA-Net [5], DDTNet [26], and SC-Net [12]. Since MaskGA-Net and PseudoEdgeNet only conduct semantic segmentation, we obtain the instance mask by applying post-processing following [4].

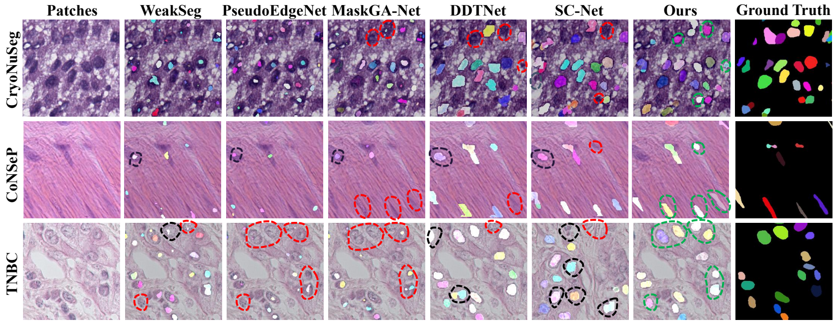

Fig. 4: Visualization comparison of segmentation results on three datasets. Red and black circles indicate the false negative (FN) and false positive (FP) errors. Green circles denote how DoNuSeg corrects these errors.

Table 2: Effects (%) of $\mathcal{L}_{dcs}$ and $\mathcal{L}_{ccl}$ on CryoNuSeg and CoNSeP.

| $\mathcal{L}_{dcs}$ | $\mathcal{L}_{ccl}$ | CryoNuSeg | | | | | CoNSeP | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | DICE | AJI | DQ | SQ | PQ | DICE | AJI | DQ | SQ | PQ |
| | | 60.93 | 41.78 | 42.68 | 63.93 | 27.32 | 53.69 | 35.51 | 35.77 | 55.18 | 22.85 |
| ✓ | | 66.48 | 43.82 | **46.41** | 64.99 | 30.20 | 57.94 | 35.75 | 35.92 | 60.14 | 23.36 |
| | ✓ | 64.02 | 43.71 | 45.66 | 65.60 | 30.06 | 52.98 | **36.53** | 36.55 | 60.43 | 23.12 |
| ✓ | ✓ | **67.22** | **44.08** | **46.41** | **65.74** | **30.58** | **60.20** | 36.39 | **36.68** | **62.33** | **23.95** |

**Quantitative Evaluation.** Table 1 presents performance comparisons in terms of five metrics. It can be seen that previous methods present poor performance due to the absence of correction for noisy pseudo labels. In contrast, our method outperforms state-of-the-art methods on AJI, DQ, and PQ across all the datasets. Notably, DoNuSeg achieves significant improvements in terms of AJI, surpassing the second-best by 3.9%, 2.9%, and 2.2% on the three datasets, respectively.

**Qualitative Evaluation.** Figure 4 displays the visual comparison results. As challenging datasets, CryoNuSeg and CoNSeP have a low distinction between nuclei and background tissue. Thus, the generated pseudo-labels based on morphology measure often involve much noise and lead to numerous FN and FP errors. Surprisingly, DoNuSeg can dynamically select and optimize pseudo labels, thus performing well on these challenging datasets.

## 3.4 Ablation Study

We conduct ablation experiments on CryoNuSeg and CoNSeP to prove the effectiveness of the proposed method. As shown in Table 2, the method significantly improves when adding $\mathcal{L}_{dcs}$ and $\mathcal{L}_{ccl}$ and achieves the best performance when both are added. This shows that DCS and CCL improve the training performance of the segmentation network by improving the quality of the pseudo-label.

Table 3: Effects (%) of the CAM selection strategy on CryoNuSeg.

| block | DICE | AJI | DQ | SQ | PQ |
|---|---|---|---|---|---|
| 1 | 61.95 | 40.14 | 39.88 | 64.81 | 27.14 |
| 2 | 64.53 | 40.91 | 41.16 | 65.35 | 29.06 |
| 3 | 67.10 | 41.44 | 40.97 | **66.02** | 27.37 |
| 4 | 62.84 | 41.73 | 45.60 | 64.38 | 28.35 |
| ours | **67.22** | **44.08** | **46.41** | 65.74 | **30.58** |

Fig. 5: Effects (%) of different $r$ and $d$ when generating $M$ on CryoNuSeg.

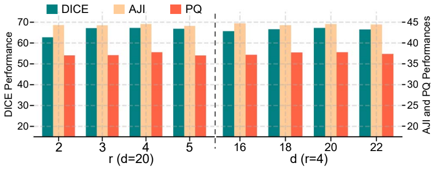

Table 3 shows the effect of the CAM selection strategy in the DCS module. It can be seen that the proposed method can combine semantic information in hierarchical features and achieve the best performance. Notably, compared to merely using CAM generated by the fourth block as in previous methods, the performance improves by 4.4%, 2.4%, and 2.1% on DICE, AJI, and PQ, respectively. Furthermore, as shown in Figure 5, when changing the hyperparameters $r$ and $d$

in generating the initial label $M$, the model's performance is almost unaffected, which shows that our method is robust to the selection of parameters.

## 4 Conclusion

This paper proposes a point-supervised nuclei segmentation framework, DoNuSeg, to reduce the cost of pixel-level annotations. DoNuSeg utilizes CAMs to achieve a dynamic optimization mechanism of the noisy pseudo labels. A DCS module and a CCL module are proposed to dynamically select and optimize CAMs and gradually correct the pseudo label. To better distinguish nuclei, we develop a task-decouple structure to leverage location priors in point labels. Experiments show that our method achieves SOTA performance, and the ablation study shows the effectiveness of the proposed method. In conclusion, DoNuSeg provides fresh insights for point-supervised nuclei instance segmentation.

**Acknowledgements.** This work was supported in part by the National Natural Science Foundation of China (62031023&62331011), in part by the Shenzhen Science and Technology Project (GXWD20220818170353009), and in part by the Fundamental Research Funds for the Central Universities (Grant No. HIT.OCEF.2023050)

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

## References

1. Chen, T., Kornblith, S., Norouzi, M., Hinton, G.: A simple framework for contrastive learning of visual representations. In: International conference on machine learning. pp. 1597–1607. PMLR (2020)
2. Choudhuri, R., Halder, A.: Histopathological nuclei segmentation using spatial kernelized fuzzy clustering approach. In: Soft Computing for Problem Solving: Proceedings of the SocProS 2022, pp. 225–238. Springer (2023)
3. Feng, Z., Wang, Z., Wang, X., Mao, Y., Li, T., Lei, J., Wang, Y., Song, M.: Mutual-complementing framework for nuclei detection and segmentation in pathology image. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 4036–4045 (2021)
4. 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). https://doi.org/10.1016/j.media.2019.101563
5. Guo, R., Pagnucco, M., Song, Y.: Learning with noise: Mask-guided attention model for weakly supervised nuclei segmentation. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part II 24. pp. 461–470. Springer (2021)

6. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016)
7. Hou, L., Agarwal, A., Samaras, D., Kurc, T.M., Gupta, R.R., Saltz, J.H.: Robust histopathology image analysis: To label or to synthesize? In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 8533–8542 (2019)
8. Huang, J., Li, H., Wan, X., Li, G.: Affine-consistent transformer for multi-class cell nuclei detection. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 21384–21393 (2023)
9. Kirillov, A., He, K., Girshick, R., Rother, C., Dollár, P.: Panoptic segmentation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 9404–9413 (2019)
10. 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)
11. Lin, T.Y., Dollár, P., Girshick, R., He, K., Hariharan, B., Belongie, S.: Feature pyramid networks for object detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 2117–2125 (2017)
12. Lin, Y., Qu, Z., Chen, H., Gao, Z., Li, Y., Xia, L., Ma, K., Zheng, Y., Cheng, K.T.: Nuclei segmentation with point annotations from pathology images via self-supervised learning and co-training. Medical Image Analysis **89**, 102933 (2023)
13. 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)
14. Mahbod, A., Schaefer, G., Bancher, B., Löw, C., Dorffner, G., Ecker, R., Ellinger, I.: Cryonuseg: A dataset for nuclei instance segmentation of cryosectioned h&e-stained histological images. Computers in biology and medicine **132**, 104349 (2021)
15. Mahmood, F., Borders, D., Chen, R.J., McKay, G.N., Salimian, K.J., Baras, A., Durr, N.J.: Deep adversarial training for multi-organ nuclei segmentation in histopathology images. IEEE transactions on medical imaging **39**(11), 3257–3267 (2019)
16. Natarajan, V.A., Kumar, M.S., Patan, R., Kallam, S., Mohamed, M.Y.N.: Segmentation of nuclei in histopathology images using fully convolutional deep neural architecture. In: 2020 International Conference on computing and information technology (ICCIT-1441). pp. 1–7. IEEE (2020)
17. Naylor, P., Laé, M., Reyal, F., Walter, T.: Segmentation of nuclei in histopathology images by deep regression of the distance map. IEEE transactions on medical imaging **38**(2), 448–459 (2018)
18. Qu, H., Wu, P., Huang, Q., Yi, J., Yan, Z., Li, K., Riedlinger, G.M., De, S., Zhang, S., Metaxas, D.N.: Weakly supervised deep nuclei segmentation using partial points annotation in histopathology images. IEEE transactions on medical imaging **39**(11), 3655–3666 (2020)
19. 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)
20. Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D.: Grad-cam: Visual explanations from deep networks via gradient-based localization. In:

Proceedings of the IEEE international conference on computer vision. pp. 618–626 (2017)

21. Stringer, C., Wang, T., Michaelos, M., Pachitariu, M.: Cellpose: a generalist algorithm for cellular segmentation. Nature methods **18**(1), 100–106 (2021)
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)
23. Tian, Z., Shen, C., Chen, H., He, T.: Fcos: Fully convolutional one-stage object detection. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 9627–9636 (2019)
24. Wang, Y., Hou, J., Hou, X., Chau, L.P.: A self-training approach for point-supervised object detection and counting in crowds. IEEE Transactions on Image Processing **30**, 2876–2887 (2021)
25. 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)
26. Zhang, X., Zhu, X., Tang, K., Zhao, Y., Lu, Z., Feng, Q.: Ddtnet: A dense dual-task network for tumor-infiltrating lymphocyte detection and segmentation in histopathological images of breast cancer. Medical image analysis **78**, 102415 (2022)
27. Zhang, Y., Cai, L., Wang, Z., Zhang, Y.: Seine: Structure encoding and interaction network for nuclei instance segmentation. arXiv preprint arXiv:2401.09773 (2024)
28. Zhao, T., Yin, Z.: Weakly supervised cell segmentation by point annotation. IEEE Transactions on Medical Imaging **40**(10), 2736–2747 (2020)
29. Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., Torralba, A.: Learning deep features for discriminative localization. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 2921–2929 (2016)