

DeNuC: Decoupling Nuclei Detection and Classification in Histopathology

Zijiang Yang¹, Chen Kuang¹, and Dongmei Fu^{1,2*}

¹ School of Automation and Electrical Engineering, University of Science and Technology Beijing, Beijing 100083, China

² Beijing Engineering Research Center of Industrial Spectrum Imaging
{zijiangyang, ckuang}@xs.ustb.edu.cn, fdm_ustb@ustb.edu.cn

Abstract. Pathology Foundation Models (FMs) have shown strong performance across a wide range of pathology image representation and diagnostic tasks. However, FMs do not exhibit the expected performance advantage over traditional specialized models in Nuclei Detection and Classification (NDC). In this work, we reveal that jointly optimizing nuclei detection and classification leads to severe representation degradation in FMs. Moreover, we identify that the substantial intrinsic disparity in task difficulty between nuclei detection and nuclei classification renders joint NDC optimization unnecessarily computationally burdensome for the detection stage. To address these challenges, we propose **DeNuC**, a simple yet effective method designed to break through existing bottlenecks by **Decoupling Nuclei** detection and **Classification**. DeNuC employs a lightweight model to detect nuclei, subsequently leveraging a pathology FM to encode input images and query nucleus-specific features based on the detected coordinates for classification. Extensive experiments on three widely used benchmarks demonstrate that DeNuC effectively unlocks the representational potential of FMs for NDC and significantly outperforms state-of-the-art methods. Notably, DeNuC improves F1 scores by 4.2% and 3.6% (or higher) on the BRCAM2C and PUMA datasets, respectively, while using only 16% (or fewer) trainable parameters compared to other methods. Code is available at <https://github.com/ZijiangY1116/DeNuC>.

Keywords: Computational pathology · Nuclei detection and classification · Foundation model.

1 Introduction

Nuclei Detection and Classification (NDC) is a cornerstone of quantitative analysis and diagnosis in computational pathology [3, 24]. Recent efforts have shown promising results by exploring spatial context representation [1], nuclei graph [10], or leveraging tissue-level context features [16]. Despite these approaches generally validating that enhancing nuclear morphology and tissue context representation is key to improving performance, they often come at the cost of introducing

* Corresponding author: Dongmei Fu

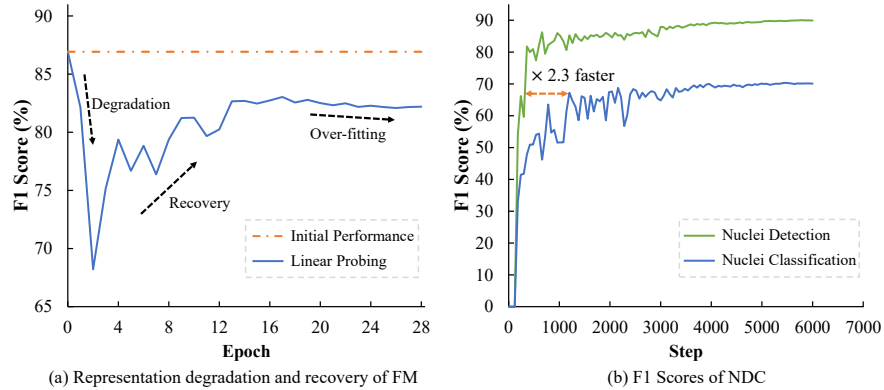


Fig. 1. Analysis of model performance during the NDC training. (a) We evaluate the representation capability of UNI2-H [2] during joint optimization for NDC on the OCELOT [14] dataset via linear probing. UNI2-H undergoes a severe representation degradation in the early training phase. Although the performance subsequently recovers, it fails to regain its initial optimal level and rapidly deteriorates into over-fitting. (b) We evaluate the detection and classification F1 scores of a ConvNeXt-S [8] model pre-trained solely on ImageNet-1K throughout the training process on PUMA [15]. Despite lacking domain-specific pathology pre-training, nuclei detection converges approximately $2.3 \times$ faster than classification, highlighting the inherent difficulty disparity between the two tasks.

highly complex module designs, which not only necessitate excessive design for specific datasets but also increase computational redundancy, thereby limiting the generalization of models across diverse clinical scenarios.

Recently, benefiting from a generic design and pre-training on large-scale unlabeled data, pathology Foundation Models (FMs) [2, 21] exhibit strong general visual representation capabilities for pathology images, significantly improving performance on numerous downstream tasks [20]. This has fostered the expectation that high-performance NDC can be achieved with simple architectures by directly leveraging the robust feature extraction of FMs. However, compared to traditional specialized models, FMs have not demonstrated a significant performance advantage in nucleus-level tasks [5].

In this work, we reveal that FMs suffer from severe representation degradation when jointly optimizing nuclei detection and classification. Although pathology pre-trained models effectively encode semantic representations of images, they are not inherently designed for coordinate regression. Forcing simultaneous joint optimization of classification and regression disrupts the pre-trained feature space, preventing the effective utilization of the original robust representations of FMs for NDC. As illustrated in Fig.1 (a), although the FM initially possesses strong nuclei representation capabilities, the model parameters undergo updates during joint training to rapidly adapt to nuclei detection requirements. While this improves detection performance, it causes a severe decline in the quality of

pure nuclei representations. In subsequent training stages, although classification representation capability recovers, it fails to return to its initial optimal level.

Furthermore, we identify a significant disparity in optimization difficulty between nuclei detection and nuclei classification, which exacerbates the inefficiency of joint optimization. Nuclei typically exhibit a relatively fixed size range and distinct contrast against the stromal background, resulting in lower optimization difficulty for detection. As shown in Fig.1 (b), even a model without pathology pre-training can achieve high-performance nuclei detection within a very short training period, whereas nuclei detection converges approximately $2.3 \times$ faster than classification. This implies that formulating NDC as a multi-task joint optimization problem not only fails to achieve mutual complementarity between tasks but also significantly inflates the computational cost of detection.

To address these issues, we propose a simple yet effective method named **DeNuC**, designed to break through existing bottlenecks by **Decoupling Nuclei** detection and **Classification**. Specifically, DeNuC first employs a lightweight detection model to localize all nuclei in the input pathology image. Subsequently, we leverage a pathology foundation model to encode the input image and utilize the detected coordinates to query nucleus-specific features from the feature maps for classification. For nuclei detection, DeNuC not only compresses model parameters to an extreme minimum but also unlocks the capability for cross-dataset joint detection learning. For nuclei classification, this decoupled design allows the foundation model to focus on fine-grained nuclei representation without being disturbed by detection task gradients, thereby fully unleashing its pre-trained representation potential. As shown in Fig.2, extensive experiments on three widely used benchmarks indicate that DeNuC not only significantly outperforms existing State-Of-The-Art (SOTA) methods in performance but also requires only 16% or less of the training parameters compared to other models.

2 Methodology

2.1 Problem Formulation

Given an input pathology image $X \in \mathbb{R}^{H \times W \times 3}$, the objective of NDC is to identify both the spatial localization and category of nuclei:

$$P = \mathcal{F}(\mathbf{X}) = \{(\mathbf{p}_k, c_k)\}_{k=1}^K, \quad (1)$$

where P is the set of predictions, $\mathcal{F}$ is the model, and K denotes the total number of detected nuclei. Each prediction consists of a centroid coordinate vector $\mathbf{p} = (x_k, y_k) \in \mathbb{R}^2$ locating the nucleus in X , and a scalar $c_k \in \{1, \dots, C\}$ representing the predicted class category, with C being the total number of defined nuclei types. The set of nuclei coordinates $\{\mathbf{p}_k\}_{k=1}^K$ and the set of nuclei types $\{c_k\}_{k=1}^K$ are denoted as P_{det} and P_{cls} , respectively.

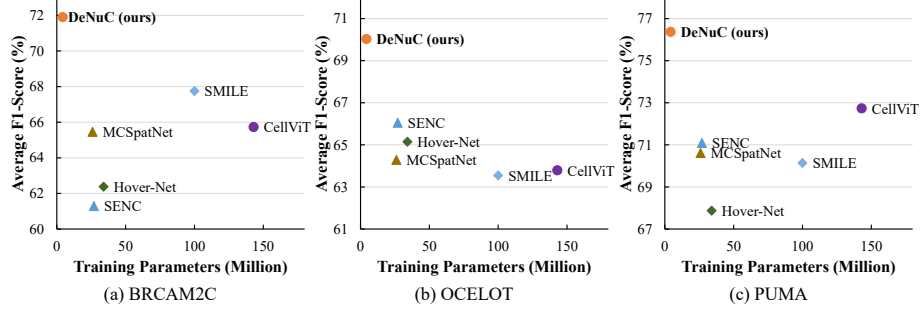


Fig. 2. Comparison of DeNuC and SOTA methods. DeNuC achieves significantly superior performance across three benchmark datasets.

2.2 DeNuC

Decoupling NDC. Existing methods typically formulate Equation (1) as a multi-task problem that share the backbone:

$$P = \underbrace{\{\mathcal{H}_{det}(\mathcal{B}(\mathbf{X}))\}}_{\text{Detection}}, \underbrace{\{\mathcal{H}_{cls}(\mathcal{B}(\mathbf{X}))\}}_{\text{Classification}}, \quad (2)$$

where $\mathcal{B}$ denotes the backbone encoder, while $\mathcal{H}_{cls}$ and $\mathcal{H}_{det}$ represent the detection and classification modules, respectively. In map-based methods [4], an additional decoder is required for nuclei segmentation [17]. Moreover, $\mathcal{H}_{cls}$ and $\mathcal{H}_{det}$ are implemented as parameter-free post-processing algorithms. Conversely, anchor-based methods [18, 16, 22] implement them as lightweight Multi-Layer Perceptrons (MLPs) for regression and prediction. Given that map-based methods are susceptible to morphological variations and necessitate heuristic tuning—thereby limiting performance, we focus our analysis primarily on the anchor-based methods to streamline the discussion.

In anchor-based frameworks, the backbone is required to extract feature maps that not only effectively represent nuclei appearance but can also be exploited by $\mathcal{H}_{cls}$ to regress the relative offsets between anchor points and potential ground-truth nuclei locations. Although pathology pre-trained models excel at representing histopathological images, they are inherently unsuitable for coordinate regression. Jointly optimizing NDC disrupts the pre-trained feature space, hindering the effective utilization of the robust representations derived from foundation models.

As illustrated in Fig.3, we address this bottleneck by decoupling detection and classification:

$$P = \underbrace{\{\mathcal{D}(\mathbf{X})\}}_{\text{Detection}}, \underbrace{\{\mathcal{H}_{cls}(\mathcal{B}(\mathbf{X}, \mathcal{D}(\mathbf{X})))\}}_{\text{Classification}}, \quad (3)$$

where $\mathcal{D}(X)$ represents an lightweight detection model. This formulation separates NDC into two distinct optimization objectives: (i) $\mathcal{D}(X)$ is dedicated solely

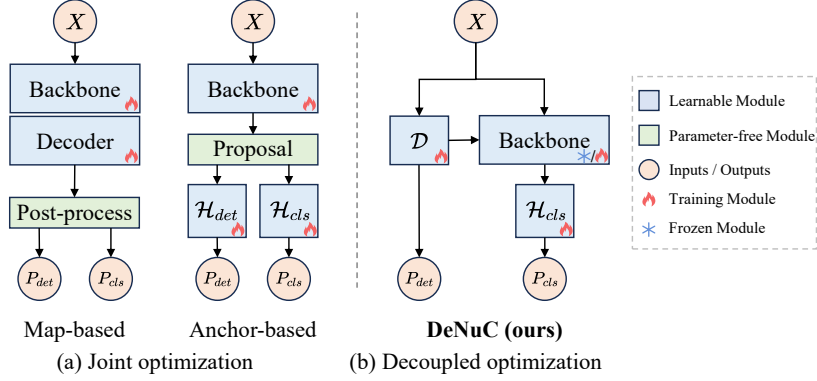


Fig. 3. Illustration of (a) Joint optimization and (b) Decoupled optimization for NDC. Existing methods require the backbone to accommodate additional optimization objectives beyond nuclei representation, leading to representation degradation. In contrast, DeNuC employs an independent lightweight detection network $\mathcal{D}$ for nuclei detection, thereby allowing the backbone to focus exclusively on representation learning.

to nuclei detection, and (ii) $\mathcal{B}$ focus on extracting nucleus-specific features for classification, conditioned on the input image and the detected coordinates. As demonstrated in our experiments, this decoupled framework not only fully exploits the powerful representation capabilities of pathology foundation models but also significantly alleviates the computational burden of detection.

Nuclei Detection. To achieve efficient and robust nuclei detection, we adopt a single-stage point detection architecture inspired by P2PNet [19]. Given an input pathology image X , the detection model $\mathcal{D}$ extracts feature maps and constructs a corresponding reference grid $\mathcal{G}$. For each spatial location (i, j) on the grid, the network concurrently predicts a nuclei confidence score $s_{i,j}$ and a spatial coordinate offset $\delta_{i,j} \in \mathbb{R}^2$ relative to the grid point. The final set of detected nuclei P_{det} is then directly formulated as:

$$P_{det} = \{\mathbf{g}_{i,j} + \delta_{i,j} \mid s_{i,j} > \tau, \forall (i, j) \in \mathcal{G}\}, \quad (4)$$

where $\mathbf{g}_{i,j}$ denotes the original coordinates of the grid point, and $\tau = 0.5$ is the confidence threshold used to filter out background noise. Crucially, Equation (4) only necessitates binary classification between nuclei and background, thereby enabling the use of an extremely lightweight model for detection and facilitating joint training across multiple datasets.

Nuclei Classification. Based on the detected nuclei locations P_{det} , the classification module further predicts the specific category for each nucleus. First, a pre-trained foundation model $\mathcal{B}$ extracts high-dimensional semantic feature maps $F \in \mathbb{R}^{C \times H' \times W'}$ of X . Subsequently, to precisely capture the local context of each nucleus, we employ a bilinear interpolation sampling operation $\mathcal{S}$ to directly query the corresponding feature vectors from F using the coordinate set

Table 1. Comparison of nucleus detection and classification in F1-score % ($\uparrow$). $F^{Tum.}$, $F^{Lym.}$, $F^{Oth.}$, and $F^{Avg.}$ denote the F1 score of tumor nucleus, lymphocytes, other nucleus, and average, respectively. The best results are highlighted in **bold**, and the second-best results are in underlined. DeNuC significantly outperforms other methods.

Method	Training Params.	BRCAM2C				OCELOT			PUMA			
		$F^{Lym.}$	$F^{Tum.}$	$F^{Oth.}$	$F^{Avg.}$	$F^{Tum.}$	$F^{Oth.}$	$F^{Avg.}$	$F^{Lym.}$	$F^{Tum.}$	$F^{Oth.}$	$F^{Avg.}$
CGT [10]	37M	56.42	75.98	50.44	60.95	68.77	61.30	65.03	76.55	79.66	54.20	70.14
SENC [11]	27M	57.94	76.50	49.42	61.29	70.02	62.08	66.05	77.38	81.51	54.38	71.09
CellViT [6]	143M	67.20	78.20	51.81	65.73	67.36	60.22	63.79	79.07	81.16	<u>57.96</u>	<u>72.73</u>
Hover-Net [4]	34M	62.31	75.25	49.58	62.38	69.27	61.03	65.15	75.72	78.36	49.53	67.87
DPA-P2PNet [18]	32M	59.65	77.26	<u>55.26</u>	64.06	70.07	59.92	64.99	76.80	81.87	54.04	70.90
MCSpatNet [1]	26M	63.15	78.56	54.66	65.46	68.60	59.99	64.29	78.25	<u>82.05</u>	51.54	70.61
PointNu-Net [23]	160M	<u>71.51</u>	76.02	51.95	66.50	66.72	56.96	61.84	76.31	79.57	52.68	69.52
SMILE [13]	100M	72.59	<u>79.61</u>	51.06	<u>67.75</u>	66.99	60.10	63.55	<u>80.35</u>	77.54	52.52	70.14
MUSE [22]	123M	-	-	-	-	<u>73.48</u>	<u>64.52</u>	<u>69.00</u>	-	-	-	-
DeNuC (ours)	4.3M	69.73	85.10	61.08	71.97	73.83	66.04	69.94	81.00	85.25	62.85	76.37

P_{det} . The final classification prediction set P_{cls} is obtained via $\mathcal{H}_{cls}$:

$$P_{cls} = \{\mathcal{H}_{cls}(\mathbf{f}_k) \mid \mathbf{f}_k = \mathcal{S}(F, \mathbf{p}_k), \forall \mathbf{p}_k \in P_{det}\}, \quad (5)$$

where $\mathbf{p}_k$ denotes the k -th detected nuclei, and $\mathbf{f}_k$ represents the sampled feature vector specific to that nucleus. This coordinate-guided feature querying mechanism not only eliminates redundant computations but also ensures a high alignment between classification features and the spatial locations of nuclei.

Learning. We optimize DeNuC via a two-stage training paradigm. First, $\mathcal{D}$ is trained using a combination of L2 regression loss and binary cross-entropy loss. Unless otherwise specified, $\mathcal{D}$ is jointly trained across multiple datasets to maximize the utility of all available annotations. Then, $\mathcal{D}$ serves as an auxiliary network to facilitate classifier training. The classifier is optimized with standard cross-entropy loss. To maximize training efficiency, we freeze the backbone $\mathcal{B}$ and only optimize the classification head $\mathcal{H}_{cls}$, unless stated otherwise. As demonstrated in our experiments, optimizing solely $\mathcal{D}$ and $\mathcal{H}_{cls}$ is sufficient to achieve SOTA performance.

3 Experiments

3.1 Experiment Settings

Datasets and evaluation metrics. To comprehensively evaluate the performance of DeNuC, we conduct extensive experiments on three widely used public datasets, including BRCAM2C [1], OCELOT [14], and PUMA [15]. Following the common practice of NDC [22, 16], we employ the distance-based F1-score for evaluation. Specifically, we perform a one-to-one matching between the predicted and ground-truth (GT) nuclei within the same category. A prediction is identified as a True Positive (TP) if it successfully matches a GT nucleus within

Table 2. Ablation study of $\mathcal{D}$ in F1-score % ($\uparrow$). **(Left)** Comparison of different backbones. **(Right)** Ablation of cross-dataset training strategy. N_{nu} denotes the number of nuclei in the training set. $F^{Det.}$ denotes the F1 score of detection. "SN" denotes ShuffleNetV2. "Separated" and "Joint" indicate that models are trained individually on each dataset and jointly across all datasets, respectively.

Backbone	Params.	BRCAM2C		OCELOT		PUMA	
		$F^{Det.}$	$F^{Avg.}$	$F^{Det.}$	$F^{Avg.}$	$F^{Det.}$	$F^{Avg.}$
SN (0.5 $\times$)	0.3M	86.61	71.43	79.89	68.85	92.23	75.45
SN (1.0 $\times$)	1.0M	86.98	71.58	80.97	69.74	93.13	75.98
SN (1.5 $\times$)	2.7M	87.00	71.90	81.07	69.76	93.28	76.19
SN (2.0 $\times$)	4.3M	87.52	71.97	81.33	69.94	93.57	76.37
ResNet-50	26M	87.29	71.90	81.48	70.03	93.22	76.07

Dataset	N_{nu}	Separated		Joint	
		$F^{Det.}$	$F^{Avg.}$	$F^{Det.}$	$F^{Avg.}$
BRCAM2C	18.6k	85.15	70.26	87.52	71.97
OCELOT	65.8k	81.22	69.90	81.33	69.94
PUMA	56.9k	93.42	76.25	93.57	76.37

a predefined distance threshold. Unmatched predictions, or those exceeding the distance threshold, are counted as False Positives (FP), while unmatched GT nuclei are designated as False Negatives (FN). Furthermore, we report the average F1-score across all classes to assess the overall performance on the dataset.

Baselines. We compare DeNuC with SOTA methods, including SENC [11], CellViT [6], CGT [10], Hover-Net [4], DPA-P2PNet [18], MCSpatNet [1], PointNuNet [23], SMILE [13], and MUSE [22]. For a fair comparison, we cite the results of MUSE evaluated at the same input size as the other models, instead of using LFoV samples.

Implementation. $\mathcal{D}$ employs a pre-trained ShuffleNetV2 [12] as the backbone and a Path Aggregation Network (PAN) [7] as the FPN. The detector is trained for 100 epochs, utilizing a batch size of 32 and an initial learning rate of 0.001. UNI2-H [2] is employed as the feature extractor. $\mathcal{H}_{cls}$ is implemented as a single fully connected layer. The classifier is trained for 100 epochs with a batch size of 256 and an initial learning rate of 0.01. Across all training phases, we employ the AdamW [9] optimizer to train models, coupled with the cosine annealing scheduler for learning rate decay.

3.2 Main Results

Table 1 reports the comparative results for nuclei detection and classification. DeNuC demonstrates significant performance advantages over both existing map-based and anchor-based approaches. Specifically, DeNuC outperforms current SOTA methods on BRCAM2C, OCELOT, and PUMA by substantial margins of 4.2%, 0.9%, and 3.6% in $F^{Avg.}$, respectively. Notably, DeNuC achieves these superior results with only 4.3M trainable parameters, which constitutes 16% or less of the parameters required by other methods. The results clearly demonstrate that DeNuC not only substantially improves the performance of nuclear analysis but also achieves remarkably high parameter efficiency.

3.3 Ablation Studies

Detection model. Table 2 reports the ablation results for the detection module $\mathcal{D}$. Table 2 (Left) shows that ShuffleNetV2 (0.5 $\times$) with only 0.3M parameters

Table 3. Ablation study on the optimization strategy of the classification network in F1-score % ($\uparrow$). Providing nuclei coordinates highly improves performance.

Method	P_{det}	Training Params.	BRCAM2C				OCELOT			PUMA			
			$F^{Lym.}$	$F^{Tum.}$	$F^{Oth.}$	$F^{Avg.}$	$F^{Tum.}$	$F^{Oth.}$	$F^{Avg.}$	$F^{Lym.}$	$F^{Tum.}$	$F^{Oth.}$	$F^{Avg.}$
Linear	✓	4.6K	69.73	85.10	61.08	71.97	73.83	66.04	69.94	81.00	85.25	62.85	76.37
Lora	✓	2.4M	71.79	85.33	61.40	72.84	74.19	66.93	70.56	81.23	85.70	63.38	76.77
Fully	✓	681M	71.03	85.50	61.11	72.55	73.87	65.52	69.69	81.45	85.37	63.13	76.65
End-to-end	✗	686M	68.21	79.42	56.67	68.10	67.37	58.15	62.76	79.50	83.19	62.20	74.97

achieves a detection F1-score highly comparable to that of the 26M ResNet-50. This suggests that model capacity has a negligible impact on nuclei detection, further verifying the low optimization difficulty of this task. In addition, Table 2 (Rights) shows that joint multi-dataset training effectively boosts performance on datasets with fewer samples. Specifically, on BRCAM2C, joint training increases the detection F1-score by 2.4%, which subsequently improves classification performance by 1.7%.

Optimization strategy of the classification network. Table 3 reports the ablation study on the optimization strategy for the classification network. The results indicate that providing P_{det} to allow the classifier to focus solely on nuclei representation significantly improves F1-scores. Notably, compared to end-to-end training, simply freezing the backbone $\mathcal{B}$ and optimizing only the linear classification head yields substantial gains of 3.9%, 7.2%, and 1.4% on BRCAM2C, OCELOT, and PUMA, respectively. Furthermore, optimizing $\mathcal{B}$ via LoRA or full fine-tuning leads to additional performance improvements. These results strongly suggest that simultaneous end-to-end detection and classification induces representation degradation, whereas DeNuC effectively capitalizes on the robust visual representation capabilities of FMs for pathology images.

4 Conclusion

In this work, we reveal that FMs suffer from severe representation degradation when jointly optimizing nuclei detection and classification. Furthermore, we identify a significant disparity in task difficulty between nuclei detection and classification, implying that formulating NDC as a multi-task joint optimization problem not only fails to achieve mutual complementarity but also unnecessarily inflates the computational cost of detection. To address these challenges, we propose **DeNuC**, a simple yet effective method designed to break through existing bottlenecks by **Decoupling Nuclei** detection and **Classification**. Extensive experiments demonstrate that DeNuC significantly outperforms existing methods while requiring only 16% or fewer training parameters compared to existing methods. These results validate DeNuC as a new paradigm that combines high performance with high training efficiency, offering a novel perspective for developing high-precision and resource-efficient algorithms in NDC.

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

References

1. Abousamra, S., Belinsky, D., Van Arnem, J., Allard, F., Yee, E., Gupta, R., Kurc, T., Samaras, D., Saltz, J., Chen, C.: Multi-class cell detection using spatial context representation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 4005–4014 (2021)
2. Chen, R.J., Ding, T., Lu, M.Y., Williamson, D.F., Jaume, G., Song, A.H., Chen, B., Zhang, A., Shao, D., Shaban, M., et al.: Towards a general-purpose foundation model for computational pathology. *Nature medicine* **30**(3), 850–862 (2024)
3. Diao, J.A., Wang, J.K., Chui, W.F., Mountain, V., Gullapally, S.C., Srinivasan, R., Mitchell, R.N., Glass, B., Hoffman, S., Rao, S.K., et al.: Human-interpretable image features derived from densely mapped cancer pathology slides predict diverse molecular phenotypes. *Nature communications* **12**(1), 1613 (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)
5. Hörst, F., Rempe, M., Becker, H., Heine, L., Keyl, J., Kleesiek, J.: Cellvit++: Energy-efficient and adaptive cell segmentation and classification using foundation models. *arXiv preprint arXiv:2501.05269* (2025)
6. 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)
7. Liu, S., Qi, L., Qin, H., Shi, J., Jia, J.: Path aggregation network for instance segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 8759–8768 (2018)
8. 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)
9. Loshchilov, I., Hutter, F.: Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101* (2017)
10. Lou, W., Li, G., Wan, X., Li, H.: Cell graph transformer for nuclei classification. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 38, pp. 3873–3881 (2024)
11. Lou, W., Wan, X., Li, G., Lou, X., Li, C., Gao, F., Li, H.: Structure embedded nucleus classification for histopathology images. *IEEE Transactions on Medical Imaging* **43**(9), 3149–3160 (2024)
12. Ma, N., Zhang, X., Zheng, H.T., Sun, J.: Shufflenet v2: Practical guidelines for efficient cnn architecture design. In: Proceedings of the European conference on computer vision (ECCV). pp. 116–131 (2018)
13. 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)
14. Ryu, J., Puche, A.V., Shin, J., Park, S., Brattoli, B., Lee, J., Jung, W., Cho, S.I., Paeng, K., Ock, C.Y., et al.: Ocelot: Overlapped cell on tissue dataset for histopathology. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 23902–23912 (2023)

15. Schuiveling, M., Liu, H., Eek, D., Breimer, G.E., Suijkerbuijk, K.P., Blokk, W.A., Veta, M.: A novel dataset for nuclei and tissue segmentation in melanoma with baseline nuclei segmentation and tissue segmentation benchmarks. *GigaScience* **14**, giaf011 (2025)
16. Shui, Z., Li, H., Zhang, Y., Sun, Y., Ye, Y., Chen, P., Guo, R., Cui, L., Zhu, C., Yang, L.: Towards effective and efficient context-aware nucleus detection in histopathology whole slide images. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. vol. 40, pp. 9042–9050 (2026)
17. 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)
18. 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)
19. Song, Q., Wang, C., Jiang, Z., Wang, Y., Tai, Y., Wang, C., Li, J., Huang, F., Wu, Y.: Rethinking counting and localization in crowds: A purely point-based framework. In: *Proceedings of the IEEE/CVF international conference on computer vision*. pp. 3365–3374 (2021)
20. Wang, X., Yang, S., Zhang, J., Wang, M., Zhang, J., Yang, W., Huang, J., Han, X.: Transformer-based unsupervised contrastive learning for histopathological image classification. *Medical image analysis* **81**, 102559 (2022)
21. Xu, H., Usuyama, N., Bagga, J., Zhang, S., Rao, R., Naumann, T., Wong, C., Gero, Z., González, J., Gu, Y., et al.: A whole-slide foundation model for digital pathology from real-world data. *Nature* **630**(8015), 181–188 (2024)
22. Yang, Z., Chao, H., Zhao, B., Yang, Y., Zhang, Y., Fu, D., Zhang, J., Lu, L., Yan, K., Jin, D., et al.: Muse: Multi-scale dense self-distillation for nucleus detection and classification. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. vol. 40, pp. 11838–11847 (2026)
23. 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)
24. Zhang, P., Gao, C., Zhang, Z., Yuan, Z., Zhang, Q., Zhang, P., Du, S., Zhou, W., Li, Y., Li, S.: Systematic inference of super-resolution cell spatial profiles from histology images. *Nature Communications* **16**(1), 1838 (2025)