A Morphology-Aware Deformable Hybrid CNN-Mamba Network for Nuclei Instance Segmentation and Classification

Saiyi Ma[1], Shaoguo Cui[1*], Guofen Wang[1], Yanhe Gong[1]
[1]College of Computer and Information Science, Chongqing Normal University, Chongqing, China
455116275@qq.com, csg@cqnu.edu.cn, wanggf@cqnu.edu.cn, 2645968486@qq.com

***Abstract*—Accurate nuclei instance segmentation and classification in pathological images are crucial for quantitative cancer diagnosis. However, nuclei in histopathological images are often densely distributed and exhibit irregular morphologies, which makes precise nuclei instance segmentation and classification highly challenging. CNN-based methods are limited by local receptive fields and fail to global contextual dependencies, while Transformer-based approaches incur high computational costs on high-resolution images. Although Mamba state-space models can efficiently model global contextual information, they remain insufficient in capturing fine-grained local nuclei morphological variations. To address this, this paper proposes a morphology-aware deformable hybrid CNN-Mamba network, MDH-CMambaNet, for nuclei instance segmentation and classification. The network employs a CNN encoder to extract multi-scale features and introduces DeformHybridMamba (DHMamba) block during decoding as the core modeling unit, which adopts a dual-branch parallel structure to jointly model local morphological variations of nuclei and global contextual dependencies. Moreover, an adaptive feature fusion mechanism (AFM) is incorporated at skip connections to dynamically fuse encoder and decoder features via channel and spatial weighting, enhancing critical nuclei regions while suppressing background tissue interference. Experimental results demonstrate that the proposed method outperforms existing approaches in overall performance, particularly in scenarios with irregular nuclei morphology and dense distribution, validating the effectiveness and superiority of the proposed framework. Code is available at https://github.com/pppumpk/MDH-CMambaNet.**

***Index Terms*—Digital pathology, Nuclei instance segmentation, Nuclei classification, Hybrid CNN–Mamba network, Morphology-aware feature modeling**

I. Introduction

Accurate nuclei instance segmentation and classification are essential for digital histopathology image analysis [1]. Nuclei morphological features and class distributions reflect cell differentiation and tissue microenvironment changes, aiding tumor quantification and prognosis [2]. However, in pathological images, nuclei are often densely packed, irregular morphology, and have blurred boundaries. Therefore, constructing an automated framework for nuclei instance segmentation and classification is especially important.

Early methods mainly relied on handcrafted features and traditional image processing techniques [3]–[5], but performed poorly in complex scenarios. With the development of deep learning, convolutional neural network (CNN)-based models have gradually become mainstream. U-Net and its variants [2], [6], [7] significantly improve nuclei instance segmentation and classification by capturing local features, but are inherently limited in modeling global contextual dependencies due to their local receptive fields.

To address this limitation, Transformer-based models have been introduced into nuclei analysis tasks, enhancing performance through global feature interactions [8], [9]. However, when processing high-resolution pathological images, the quadratic complexity of self-attention in Transformers results in high computational costs. Recently, state space models (SSMs), represented by Mamba [10], have attracted attention for their linear complexity in modeling long sequences. By employing multi-directional scanning mechanisms [11], [12], Mamba has been extended to the computer vision domain, achieving efficient global context modeling. Meanwhile, CP-Mamba [13] further improves classification performance on rare classes by incorporating a class prompt mechanism. Nevertheless, existing Mamba-based methods for nuclei instance segmentation and classification remain insufficient in capturing fine-grained local nuclei morphological variations, limiting their ability to characterize irregular nuclei morphology and fine structural details.

In pathological images, different cell types exhibit substantial variations in nuclei morphology, often accompanied by irregular boundaries and adhesion. Standard convolution, constrained by fixed sampling locations, struggles to flexibly model such complex structures. Deformable convolution [14] improves modeling of irregular objects via adaptive sampling, but suffers from high computational overhead. Linear Deformable Convolution (LDConv) [15] retains morphological adaptability while significantly reducing computational complexity. Although existing methods have made progressive advances, there remain clear deficiencies in simultaneously accurately characterizing the local irregular nuclei morphology and efficiently modeling the global dynamic relationships among nuclei instances.

To address this, we propose MDH-CMambaNet, a CNN-Mamba hybrid network for nuclei instance segmentation and classification. A CNN encoder is employed to extract multi-

* Shaoguo Cui is the corresponding author.

scale features, while DeformHybridMamba (DHMamba) block is introduced during decoding as the core modeling unit. DHMamba combines the LDConv's ability to model local irregular morphology with Mamba's capacity for long-range contextual modeling, achieving collaborative modeling of complex nuclei morphology and inter-nuclei relationships. In addition, an Adaptive Feature Fusion Mechanism (AFM) is designed at skip connections to enable adaptive cross-scale feature fusion via channel and spatial weighting, emphasizing critical nuclei regions while suppressing background tissue interference.

The main contributions of this paper are as follows:

- We propose MDH-CMambaNet, a CNN-Mamba hybrid network that effectively integrates multi-scale CNN features with global context modeling for accurate nuclei instance segmentation and classification.
- We design the DHMamba module with a dual-branch parallel structure to jointly model local morphological variations and global contextual dependencies, enhancing the model's overall perception and discrimination ability for irregular morphology and densely distributed nuclei.
- An Adaptive Feature Fusion Mechanism (AFM) is developed to achieve adaptive fusion of cross-scale features between the encoder and decoder, precisely focusing on key regions of the nuclei while suppressing background tissue interference.
- Results demonstrate that the proposed MDH-CMambaNet achieves significant improvements on two public datasets, showing clear advantages especially in scenarios with irregular nuclei morphology and dense distributions.

II. Related works

Research on nuclei instance segmentation and classification has gradually evolved from traditional image processing to deep learning approaches. Early studies relied on handcrafted features and heuristic rules. For example, Yang et al. (2006) [3] combined watershed and mean shift algorithms to segment clustered nuclei. Liao et al. (2016) [5] employed ellipse fitting to separate touching nuclei. However, these methods rely on manual rules and thresholds, limiting generalizability.

With the development of deep learning, CNN-based methods quickly became mainstream. Graham et al. (2019) [6] proposed HoVer-Net, which mitigates nuclei adhesion by regressing horizontal and vertical displacement maps. Subsequent works, such as Zhao et al. (2020) [7]and Pan et al. (2023) [2], further improved nuclei instance segmentation and classification through dense feature aggregation and cost-sensitive multi-task learning. Additionally, Transformer-based methods such as Chen et al. (2023) [16] incorporated Transformer self-attention into U-Net. Hörst et al. (2024) [9] proposed CellViT using large-scale pretrained Vision Transformers to enhance global feature interaction. However, the high computational complexity of Transformers limits their application to high-resolution pathology images. Recently, state space models (SSMs), exemplified by Mamba [10], have attracted attention for their linear complexity in long-sequence modeling. Zhu et al. (2024) [11] proposed Vision Mamba (vim) to capture global visual context with bidirectional SSMs. Ma et al. (2024) [17] integrated Mamba into U-Net for medical image segmentation. Zhang et al. (2025) [13] further introduced a class prompt mechanism into Mamba to improve segmentation and classification performance for rare categories. However, these methods remain insufficient in capturing fine-grained local nuclei morphological variations, such as irregular nuclei morphology and fine boundaries.

To overcome the limitations of existing methods in modeling both global context and fine-grained local nuclei morphology, we propose MDH-CMambaNet, a hybrid CNN-Mamba architecture that jointly models global contextual dependencies and local morphological variations of nuclei. Moreover, AFM is introduced to achieve cross-scale feature integration, facilitating accurate nuclei instance segmentation and classification in complex pathological images.

III. Methods

The overall architecture of MDH-CMambaNet is illustrated in Fig. 1 and comprises three core components: a CNN-based encoder, AFM (in Sec. III-A), and a DHMamba-based decoder (in Sec. III-B). The network follows a U-shaped design, with a ResNet-34 encoder extracting multi-scale features. During decoding, the deepest features are first processed by two stacked DHMamba blocks to model global contextual dependencies and local nuclei morphology. Then, at each upsampling stage, encoder and decoder features are adaptively fused via AFM and further refined by DHMamba blocks, enabling the decoder to progressively restore spatial details while maintaining awareness of inter-nuclei relationships and morphological variations. Finally, we employ a multi-task output branch that simultaneously generates the nuclei pixel (NP) branch, the horizontal-vertical distance (HV) branch, and the nuclei classification (NC) branch.

- **NP branch:** Predicts whether each pixel belongs to a nuclei or the background.
- **HV branch:** Predicts the horizontal and vertical distances from nuclei pixels to the nuclei centroid.
- **NC branch:** Predicts the nuclei type for each pixel and further determines the category of each nuclei instance.

A. Adaptive Feature Fusion Mechanism

Adaptive Feature Fusion Mechanism (AFM) adaptively fuses encoder and decoder features at skip connections to enhance nuclei regions while suppressing background interference. As shown in Fig. 2. Specifically, the upsampled decoder feature F_1 and the corresponding encoder feature F_2 are first concatenated along the channel dimension:

$$F = \text{Concat}([F_1, F_2]) \tag{1}$$

Subsequently, global average pooling followed by a 1×1 convolution is applied to generate channel-wise attention weights

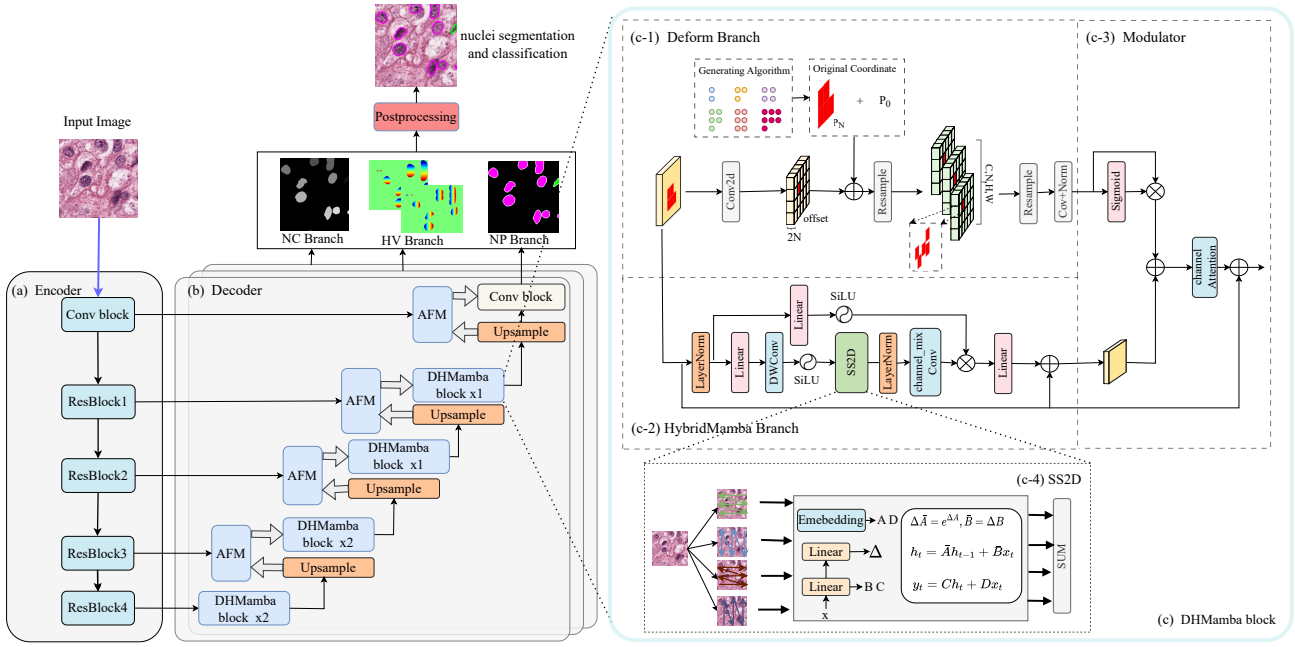

Fig. 1. An overview of our proposed MDH-CMambaNet

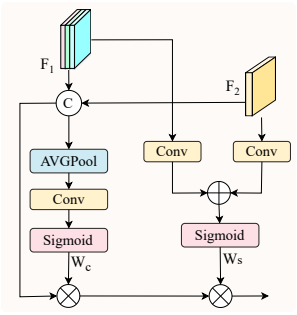

Fig. 2. Architecture of the proposed Adaptive Feature Fusion Mechanism

W_c, which are used to reweight F and enhance discriminative channel features:

$$W_c = \sigma\big(\text{Conv}(\text{AVGPool}(F))\big)$$
$$F_c = F \cdot W_c. \tag{2}$$

On this basis, spatial dependencies are further modeled. Specifically, F_1 and F_2 are first compressed via 1×1 convolutions to aggregate spatial information. The resulting features are then combined and passed through a Sigmoid activation function to obtain the spatial attention weights W_s, which are applied to F_c to produce the final fused feature:

$$W_s = \sigma(\text{Conv}(F_1) + \text{Conv}(F_2)),$$
$$F = F_c \cdot W_s. \tag{3}$$

Overall, AFM jointly models channel importance and spatial consistency, enabling adaptive feature fusion that highlights key nuclei regions and provides more discriminative feature representations for subsequent modeling.

B. DeformHybridMamba block

To address Mamba's limitations in capturing local structures such as irregular nuclei morphology and fine boundaries, we designed the DHMamba block. This block employs a dual-branch parallel structure to collaboratively model the fused features, thereby simultaneously capturing local morphological variations of nuclei and global contextual dependencies. Its architecture comprises a Deform branch and a HybridMamba branch, fused via a Modulator, as shown in Fig. 1(c).

1) Deform Branch: The Deform branch employs LDConv to adaptively model irregular nuclei morphology, enabling boundary alignment and enhanced perception of local variations. Unlike conventional deformable convolutions that rely on regular sampling, LDConv leverages irregular sampling to more effectively capture local morphological features, as shown in Fig. 1(c-1).

Specifically, for an irregular convolution kernel with N sampling points, the initial sampling coordinates P_n are first generated, corresponding to the original positions $(P_0 + P_n)$ on the feature map. The input feature X is then passed through a 3×3 convolution to predict the sampling offsets $\Delta P \in R^{2N \times H \times W}$, which are added to the original coordinates to obtain the final sampling locations P, enabling adaptive sampling of irregular nuclei morphology:

$$\Delta P = \text{Conv}(X)$$
$$P = P_0 + P_n + \Delta P \cdot s. \tag{4}$$

where s is a learnable scaling factor. Bilinear interpolation resamples the features to obtain offset features X_{offset}, which are aggregated along the column dimension to produce the output:

$$X_{def} = Conv_{(N,1)}(X_{offset}) \tag{5}$$

2) HybridMamba Branch: The HybridMamba Branch is designed to model long-range dependencies across different spatial positions in the feature map and enhance cross-channel information interaction through channel mixing, as illustrated in Fig. 1(c-2). The input feature X is first layer-normalized and split into two parallel paths.

The first path performs linear projection on the input, followed by SiLU activation. The second path sequentially applies a linear projection, depthwise separable convolution (DWConv), and SiLU, followed by 2D-Selective-Scan (SS2D) for long-range dependency modeling. After SS2D, layer normalization and a 1×1 channel-mixing convolution are applied. The two paths are fused by element-wise multiplication and combined with the input via a residual connection to generate the HybridMamba output:

$$\begin{aligned} X_1 &= \text{Conv}(\text{LN}(\text{SS2D}\,(\text{SiLU}\,(\text{DWConv}(\text{Linear}(X)))))) \\ X_2 &= \text{SiLU}(\text{Linear}(X)) \\ X_{\text{global}} &= X + \text{Linear}(X_1 \cdot X_2) \end{aligned} \tag{6}$$

Fig. 1(c-4) shows the specific implementation of SS2D. The scan expansion operation first unfolds the image into sequences in four different directions, fully extracting useful information. Each sequence is further processed by a Selective State Space model (S6) to produce enhanced features. Finally, the scan merging operation combines the four features from the different sequences to restore the original size.

3) Modulator: As shown in Fig. 1(c-3), the Modulator generates a soft attention mask M to modulate the Deform branch output feature, highlighting critical regions while suppressing irrelevant responses:

$$\begin{aligned} M &= \sigma(\text{BN}(\text{Conv}(X_{\text{def}}))) \\ X_{\text{local}} &= X_{\text{def}} \cdot M \end{aligned} \tag{7}$$

The outputs of the two branches are fused via residual connections and channel attention, and further combined with the input feature through a residual path to obtain the final feature representation.

In summary, DHMamba jointly models local nuclei morphology and global contextual dependencies through a dual-branch design, enhancing the model's perception and discrimination capability for irregular morphology and densely distributed nuclei.

C. Loss Function

The overall loss $\mathcal{L}$ is calculated as the weighted sum of the NP branch loss ($\mathcal{L}_{np}$), HV branch loss ($\mathcal{L}_{hv}$), and NC branch loss ($\mathcal{L}_{nc}$):

$$\mathcal{L} = \mathcal{L}_{np} + \mathcal{L}_{hv} + \mathcal{L}_{nc} \tag{8}$$

IV. Experiments

A. Dataset

CoNSeP. CoNSeP dataset contains 41 H&E-stained images. Following the protocol in Graham et al. [6], Normal and Malignant/Dysplastic epithelial nuclei are merged into Epithelial, while Fibroblast, Muscle, and Endothelial are grouped as Spindle-shaped. Nuclei are finally categorized into four types: Epithelial, Spindle-shaped, Inflammatory and Miscellaneous.

PanNuke. PanNuke is a large pancancer cell nuclei segmentation and classification dataset [30], covering 19 tissue types and five nuclei classes with 189,744 annotated nuclei. We follow the official three-fold cross-validation protocol and took the average results.

B. Implementation Details and Evaluation Metrics

All experiments were conducted on an NVIDIA RTX A6000 GPU with data augmentation including random scaling, cropping, and Gaussian blur. A two-stage training strategy was adopted. In the first stage, ImageNet pre-trained weights [6] were loaded, and the model was trained for 50 epochs using Adam with an initial learning rate of 10^{-4}, decayed by a factor of 10 every 25 epochs, with momentum 0.9 and weight decay 10^{-4}. In the second stage, all network parameters were unfrozen and training continued for 120 epochs using the same optimization and loss settings.

Segmentation performance was evaluated using Dice, AJI, DQ, SQ, and PQ metrics, while classification performance was assessed with the F1 score.

C. Comparative Test

1) Segmentation Performance: We compared our method with nine state-of-the-art nuclei instance segmentation models as shown in Table I. On CoNSeP, our method improves Dice, AJI, and PQ by 1.4%, 0.4%, and 2.3% over HoVer-Net, and achieves a 4.0% improvement in PQ compared with CP-Mamba. On PanNuke, Dice, SQ, and PQ are improved by 2.8%, 2.4%, and 1.7%, respectively, over HoverNet.

In addition, a visual comparison of nuclei instance segmentation on CoNSeP and PanNuke is shown in Fig. 3. The results indicate that, in regions with irregular nuclei morphology and overlapping nuclei, our method outperforms the baseline models, accurately delineating irregular nuclei boundaries and effectively separating clustered nuclei. Even in the presence of small nuclei or partially overlapping instances, our model can reliably segment and identify individual nuclei.

2) Classification Performance: We compare the classification performance with twelve SOTA classification networks, with results summarized in Table II, where F_d denotes the detection F1 score, and the remaining columns correspond to the F1 scores of each category. On CoNSeP, our method improves F_d by 2.1% over the second-best HistoNetX. For the miscellaneous category, which accounts for only 2%, our approach outperforms the second-best method SMILE by 2.3%. In addition, our method achieves the best performance on epithelial and spindle-shaped nuclei, surpassing CP-Mamba by 2.3% and 2.7%, respectively. On PanNuke, our method maintains a high F_d and achieves the highest F1 scores on all nuclei types except inflammatory and dead categories, indicating strong overall classification capability.

TABLE I
NUCLEI INSTANCE SEGMENTATION COMPARISON ON CONSEP AND PANNUKE.

Methods	Year	CoNSeP					PanNuke				
		Dice	AJI	DQ	SQ	PQ	Dice	AJI	DQ	SQ	PQ
Mask-RCNN [18]	2017	0.740	0.518	0.661	0.757	0.498	0.750	0.612	0.723	0.786	0.568
HoverNet [6]	2019	0.833	0.559	0.691	0.768	0.531	0.803	0.656	0.758	0.801	0.607
Triple-UNet [7]	2020	0.823	0.551	0.498	0.746	0.372	0.742	0.588	0.667	0.735	0.500
TSHVNet [19]	2022	0.825	0.501	0.660	0.752	0.500	0.801	0.642	0.740	0.801	0.590
SMILE [11]	2023	0.830	0.549	0.678	0.768	0.520	0.811	0.618	0.736	0.817	0.602
DoNet [20]	2023	0.782	0.467	0.572	0.728	0.416	0.782	0.668	0.749	0.792	0.599
Vim [11]	2024	0.816	0.492	0.628	0.759	0.478	0.790	0.641	0.746	0.808	0.603
CP-Mamba [13]	2025	0.822	0.539	0.658	**0.776**	0.514	0.818	0.670	**0.767**	0.819	0.614
SEINE [21]	2025	0.829	0.560	–	–	0.510	0.820	**0.682**	–	–	0.609
Ours	∼	**0.847**	**0.563**	**0.714**	0.775	**0.554**	**0.831**	0.671	0.754	**0.825**	**0.624**

TABLE II
NUCLEI CLASSIFICATION COMPARISON ON CONSEP AND PANNUKE. F^e, F^i, F^s, F^m, F^c, F^d, F^{ne}, F^{no} DENOTE THE F1 SCORE FOR NUCLEI TYPES OF EPITHELIAL, INFLAMMATORY, SPINDLE-SHAPED, MISCELLANEOUS, CONNECTIVE, DEAD, NEOPLASTIC, AND NON-NEOPLASTIC, RESPECTIVELY.

Methods	Year	CoNSeP					PanNuke					
		F_d	F^e	F^i	F^s	F^m	F_d	F^c	F^d	F^i	F^{ne}	F^{no}
DIST [22]	2018	0.730	0.616	0.608	0.526	0.171	0.718	0.419	0.000	0.426	0.500	0.341
HoverNet [6]	2019	0.748	0.635	0.631	0.566	0.426	0.800	0.491	0.310	0.514	0.565	0.552
DST-CNN [23]	2020	0.663	0.567	0.541	0.532	0.119	0.782	0.449	0.068	0.472	0.593	0.500
SONNET [24]	2022	0.739	0.640	0.619	0.559	0.367	0.800	0.453	0.298	0.531	0.613	0.558
SMILE [2]	2023	0.763	0.668	0.624	0.578	0.500	0.750	0.485	0.312	0.524	0.560	0.543
DRCA-Net [25]	2023	0.735	0.648	0.622	0.538	0.372	0.807	0.496	0.339	0.544	0.626	0.587
CGT [13]	2024	0.742	0.646	0.562	0.541	0.396	0.786	0.472	0.289	0.526	0.598	0.621
SENC [26]	2024	0.731	0.606	0.581	0.557	0.394	0.767	0.482	0.216	0.498	0.600	0.653
ANet [27]	2024	0.743	0.625	0.613	0.579	0.351	–	–	–	–	–	–
HistoNetXt [28]	2025	0.767	0.653	**0.653**	0.580	0.394	0.811	0.520	**0.353**	**0.573**	0.701	0.700
NuHTC [29]	2025	–	–	–	–	–	0.815	0.565	0.309	0.555	0.703	0.715
CP-Mamba [13]	2025	0.753	0.657	0.632	0.571	0.409	0.817	0.500	0.313	0.539	0.621	0.640
Ours	∼	**0.788**	**0.680**	0.651	**0.598**	**0.523**	**0.826**	**0.578**	0.344	0.562	**0.714**	**0.728**

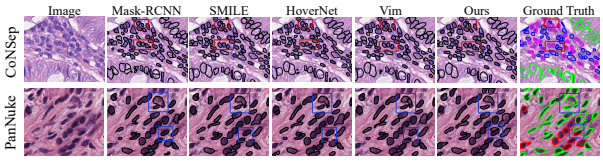

Fig. 3. Visual comparison of instance segmentation results. Nuclei are shown in one color, with differences marked by boxes.

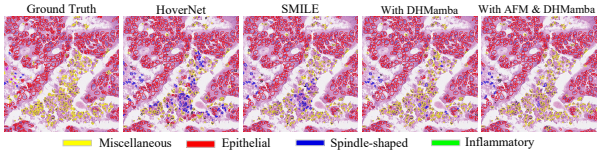

Fig. 4. Visual ablation results of DHMamba and AFM.

D. Ablation Study

We conducted ablation studies on the CoNSeP and PanNuke datasets to evaluate the effectiveness of each key component in MDH-CMambaNet, as shown in Table III and Table IV . Adding AFM alone improves PQ by 1.8% and 1.4% and increases Fd by 1.1% and 0.7% on CoNSeP and PanNuke, respectively. These results indicate that AFM effectively suppresses background interference and focuses on critical nuclei regions, thereby improving both instance segmentation and classification accuracy.Adding DHMamba alone improves PQ by 3.8% and 2.2% on CoNSeP and PanNuke, respectively, while F_d increases by 2.6% and 1.2%. Notably, for the miscellaneous category with highly irregular nuclei morphology and large scale variations, the F1 score increases from 0.420 to 0.523, highlighting the benefit of the dual-branch design: the Deform branch enhances the model's ability to perceive local structural variations of irregular nuclei, while the HybridMamba branch captures global contextual dependencies, thereby improving sensitivity to irregular morphology changes and category discrimination. When AFM and DHMamba were jointly integrated to form the complete MDH-CMambaNet, the model achieves the best overall performance, with PQ improvements of 6.1% and 2.9% and Fd gains of 4.4% and 2.3% on CoNSeP and PanNuke, respectively. The complete model consistently achieves the best results across all evaluation metrics, fully validating the complementary effects of AFM and DHMamba. Visualization results are shown in Fig. 4.

TABLE III
ABLATION STUDY ON CONSEP DATASET.

AFM	DHMamba	Dice	PQ	$\mathbf{F_d}$	$\mathbf{F^e}$	$\mathbf{F^i}$	$\mathbf{F^s}$	$\mathbf{F^m}$
		0.816	0.493	0.744	0.623	0.548	0.536	0.420
✓		0.825	0.511	0.755	0.662	0.589	0.568	0.490
	✓	0.831	0.531	0.770	0.660	0.642	0.593	0.510
✓	✓	**0.847**	**0.554**	**0.788**	**0.680**	**0.651**	**0.598**	**0.523**

V. CONCLUSION

This paper proposes a hybrid CNN-Mamba architecture called MDH-CMambaNet for nuclei instance segmentation

TABLE IV
ABLATION STUDY ON PANNUKE DATASET.

AFM	DHMamba	Dice	PQ	F_d	F^c	F^d	F^i	F^{ne}	F^{no}
		0.810	0.595	0.803	0.512	0.301	0.519	0.620	0.683
✓		0.816	0.609	0.810	0.537	0.316	0.523	0.663	0.716
	✓	0.823	0.617	0.815	0.554	0.327	0.546	0.668	0.714
✓	✓	**0.831**	**0.624**	**0.826**	**0.578**	**0.344**	**0.562**	**0.714**	**0.728**

and classification. By jointly modeling local morphological features of nuclei and global contextual dependencies, the proposed method effectively addresses the challenges posed by irregular nuclei morphology and dense distributions. Specifically, the DHMamba integrates deformable local modeling with Mamba-based long-range dependency modeling, while AFM enables adaptive cross-scale feature fusion to enhance nuclei regions and suppress background interference. Experimental results demonstrate the effectiveness of MDH-CMambaNet in nuclei instance segmentation and classification tasks, especially under irregular morphology and dense nuclei scenarios. Future work will further explore extending MDH-CMambaNet to larger and more diverse pathological datasets, continuously improving its accuracy in high-resolution whole-slide image analysis.

VI. ACKNOWLEDGMENTS

Key Project of Chongqing Natural Science Foundation Innovation Development Joint Fund (CSTB2025NSCQ-LZX0073)

REFERENCES

[1] Z. Lu, S. Xu, W. Shao, Y. Wu, J. Zhang, Z. Han, Q. Feng, and K. Huang, "Deep-learning–based characterization of tumor-infiltrating lymphocytes in breast cancers from histopathology images and multiomics data," *JCO clinical cancer informatics*, vol. 4, pp. 480–490, 2020.

[2] X. Pan, J. Cheng, F. Hou, R. Lan, C. Lu, L. Li, Z. Feng, H. Wang, C. Liang, Z. Liu *et al.*, "Smile: Cost-sensitive multi-task learning for nuclear segmentation and classification with imbalanced annotations," *Medical Image Analysis*, vol. 88, p. 102867, 2023.

[3] X. Yang, H. Li, and X. Zhou, "Nuclei segmentation using marker-controlled watershed, tracking using mean-shift, and kalman filter in time-lapse microscopy," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 53, no. 11, pp. 2405–2414, 2006.

[4] M. Veta, P. J. Van Diest, R. Kornegoor, A. Huisman, M. A. Viergever, and J. P. Pluim, "Automatic nuclei segmentation in h&e stained breast cancer histopathology images," *PloS one*, vol. 8, no. 7, p. e70221, 2013.

[5] M. Liao, Y.-q. Zhao, X.-h. Li, P.-s. Dai, X.-w. Xu, J.-k. Zhang, and B.-j. Zou, "Automatic segmentation for cell images based on bottleneck detection and ellipse fitting," *Neurocomputing*, vol. 173, pp. 615–622, 2016.

[6] S. Graham, Q. D. Vu, S. E. A. Raza, A. Azam, Y. W. Tsang, J. T. Kwak, and N. Rajpoot, "Hover-net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images," *Medical image analysis*, vol. 58, p. 101563, 2019.

[7] B. Zhao, X. Chen, Z. Li, Z. Yu, S. Yao, L. Yan, Y. Wang, Z. Liu, C. Liang, and C. Han, "Triple u-net: Hematoxylin-aware nuclei segmentation with progressive dense feature aggregation," *Medical Image Analysis*, vol. 65, p. 101786, 2020.

[8] Z. He, M. Unberath, J. Ke, and Y. Shen, "Transnuseg: A lightweight multi-task transformer for nuclei segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2023, pp. 206–215.

[9] F. Hörst, M. Rempe, L. Heine, C. Seibold, J. Keyl, G. Baldini, S. Ugurel, J. Siveke, B. Grünwald, J. Egger *et al.*, "Cellvit: Vision transformers for precise cell segmentation and classification," *Medical Image Analysis*, vol. 94, p. 103143, 2024.

[10] A. Gu and T. Dao, "Mamba: Linear-time sequence modeling with selective state spaces," in *First conference on language modeling*, 2024.

[11] L. Zhu, B. Liao, Q. Zhang, X. Wang, W. Liu, and X. Wang, "Vision mamba: Efficient visual representation learning with bidirectional state space model," *arXiv preprint arXiv:2401.09417*, 2024.

[12] Y. Liu, Y. Tian, Y. Zhao, H. Yu, L. Xie, Y. Wang, Q. Ye, J. Jiao, and Y. Liu, "Vmamba: Visual state space model," *Advances in neural information processing systems*, vol. 37, pp. 103 031–103 063, 2024.

[13] Y. Zhang, Z. Fang, Y. Wang, L. Zhang, X. Guan, and Y. Zhang, "Category prompt mamba network for nuclei segmentation and classification," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 39, no. 10, 2025, pp. 10 284–10 292.

[14] X. Zhu, H. Hu, S. Lin, and J. Dai, "Deformable convnets v2: More deformable, better results," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 9308–9316.

[15] X. Zhang, Y. Song, T. Song, D. Yang, Y. Ye, J. Zhou, and L. Zhang, "Ldconv: Linear deformable convolution for improving convolutional neural networks," *Image and Vision Computing*, vol. 149, p. 105190, 2024.

[16] B. Chen, Y. Liu, Z. Zhang, G. Lu, and A. W. K. Kong, "Transattunet: Multi-level attention-guided u-net with transformer for medical image segmentation," *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 8, no. 1, pp. 55–68, 2023.

[17] C. Ma and Z. Wang, "Semi-mamba-unet: Pixel-level contrastive and pixel-level cross-supervised visual mamba-based unet for semi-supervised medical image segmentation," *arXiv preprint arXiv:2402.07245*, 2024.

[18] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2961–2969.

[19] Y. Chen, Y. Jia, X. Zhang, J. Bai, X. Li, M. Ma, Z. Sun, and Z. Pei, "Tshvnet: Simultaneous nuclear instance segmentation and classification in histopathological images based on multiattention mechanisms," *BioMed Research International*, vol. 2022, no. 1, p. 7921922, 2022.

[20] H. Jiang, R. Zhang, Y. Zhou, Y. Wang, and H. Chen, "Donet: Deep de-overlapping network for cytology instance segmentation," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2023, pp. 15 641–15 650.

[21] Y. Zhang, L. Cai, Z. Wang, and Y. Zhang, "Seine: Structure encoding and interaction network for nuclei instance segmentation," *IEEE Journal of Biomedical and Health Informatics*, 2025.

[22] P. Naylor, M. Laé, F. Reyal, and T. Walter, "Segmentation of nuclei in histopathology images by deep regression of the distance map," *IEEE transactions on medical imaging*, vol. 38, no. 2, pp. 448–459, 2018.

[23] S. Graham, D. Epstein, and N. Rajpoot, "Dense steerable filter cnns for exploiting rotational symmetry in histology images," *IEEE Transactions on Medical Imaging*, vol. 39, no. 12, pp. 4124–4136, 2020.

[24] T. N. Doan, B. Song, T. T. Vuong, K. Kim, and J. T. Kwak, "Sonnet: A self-guided ordinal regression neural network for segmentation and classification of nuclei in large-scale multi-tissue histology images," *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 7, pp. 3218–3228, 2022.

[25] G. M. Dogar, M. Shahzad, and M. M. Fraz, "Attention augmented distance regression and classification network for nuclei instance segmentation and type classification in histology images," *Biomedical Signal Processing and Control*, vol. 79, p. 104199, 2023.

[26] W. Lou, X. Wan, G. Li, X. Lou, C. Li, F. Gao, and H. Li, "Structure embedded nucleus classification for histopathology images," *IEEE Transactions on Medical Imaging*, vol. 43, no. 9, pp. 3149–3160, 2024.

[27] M. Kadaskar and N. Patil, "Anet: Nuclei instance segmentation and classification with attention-based network," *SN Computer Science*, vol. 5, no. 4, p. 348, 2024.

[28] J. Chen, R. Wang, W. Dong, H. He, and S. Wang, "Histonext: dual-mechanism feature pyramid network for cell nuclear segmentation and classification," *BMC Medical Imaging*, vol. 25, no. 1, p. 9, 2025.

[29] B. Li, Z. Liu, S. Zhang, X. Liu, C. Sun, J. Liu, B. Qiu, and J. Tian, "Nuhtc: A hybrid task cascade for nuclei instance segmentation and classification," *Medical Image Analysis*, p. 103595, 2025.

[30] J. Gamper, N. A. Koohbanani, K. Benes, S. Graham, M. Jahanifar, S. A. Khurram, A. Azam, K. Hewitt, and N. Rajpoot, "Pannuke dataset extension, insights and baselines," *arXiv preprint arXiv:2003.10778*, 2020.