CellMamba: Adaptive Mamba for Accurate and Efficient Cell Detection

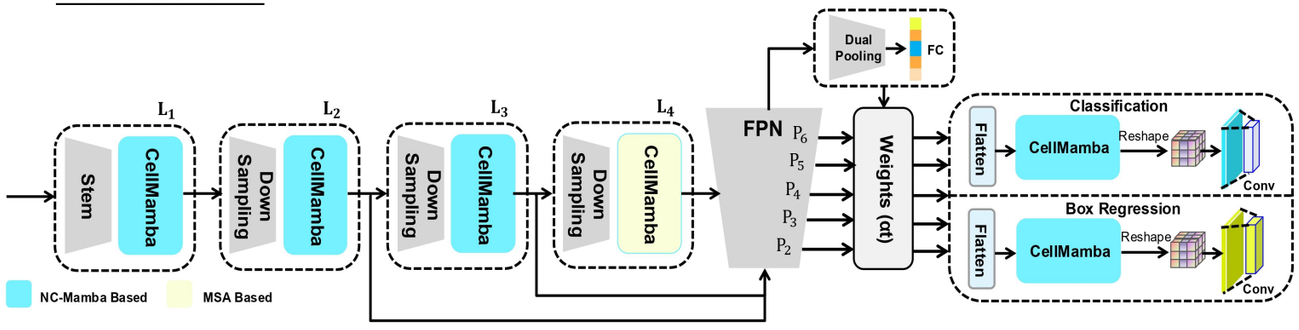
Ruochen Liu, Yi Tian, Jiahao Wang, Hongbin Liu, Xianxu Hou, Jingxin Liu

Faculty of Science and Engineering, University of Liverpool
Yong Loo Lin School of Medicine, National University of Singapore
School of AI and Advanced Computing, Xi'an Jiaotong-Liverpool University

Introduction

Cell detection in pathological images presents unique challenges due to densely packed objects, subtle inter-class differences, and severe background clutter. In this paper, we propose CellMamba, a lightweight and accurate one-stage detector tailored for fine-grained biomedical instance detection. Built upon a VSSD backbone, Cell- Mamba integrates CellMamba Blocks, which couple either NC-Mamba or Multi-Head Self-Attention (MSA) with a novel Triple-Mapping Adaptive Coupling (TMAC) module. TMAC enhances spatial discriminability by splitting channels into two parallel branches, equipped with dual idiosyncratic and one consensus attention map, adaptively fused to preserve local sensitivity and global consistency. Furthermore, we design an Adaptive Mamba Head that fuses multi-scale features via learnable weights for robust detection under varying object sizes. Extensive experiments on two public datasets—CoNSeP and CytoDArk0—demonstrate that CellMamba outperforms both CNN-based, Transformer- based, and Mamba-based baselines in accuracy, while significantly reducing model size and inference latency. Our results validate CellMamba as an efficient and effective solu- tion for high-resolution cell detection. Subsequent updates will be made available on arXiv.

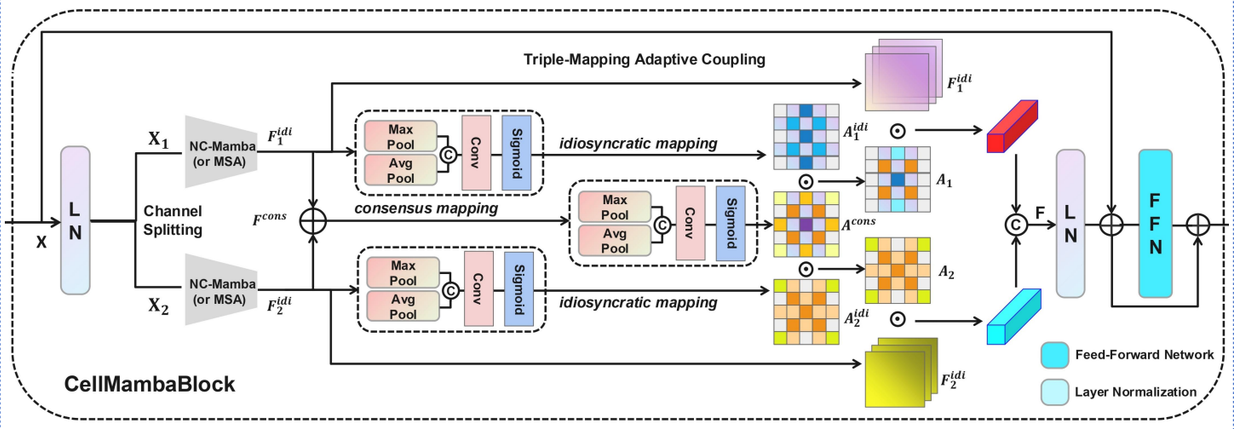
Framework:



Main contributions:

- *Efficient and Accurate One Stage Detector.* We propose CellMamba, a Mamba-based one-stage object detector for efficient and accurate cell detection in pathology images.
- *Triple-Mapping Adaptive Coupling.* We introduce the Triple-Mapping Adaptive Coupling (TMAC) module to reduce fea- ture interference and enhance spatial focus, improving detection in dense, low-contrast regions.
- *Bounding-Box-Based Detection and SOTA Model.* We bridge the gap in bounding-box-based cell detection with Mamba models and validate our method on two public datasets, achieving state-of-the-art performance.

Method



Backbone:

The first three stages of the backbone use NC-Mamba blocks to capture long-range spatial dependencies, which are essential for separating visually similar nuclei and suppressing background noise. The final stage adopts Multi-Head Self-Attention (MSA) to enhance global contextual modelling and complex spatial reasoning for low-resolution abstract cellular features. We refer to each stage's combined structure of sequence modeling (NC-Mamba or MSA) followed by TMAC as a CellMamba block, which serves as the core unit of our backbone. The four stages contain 2, 2, 8, and 4 such blocks respectively, each equipped with our proposed TMAC module.

Triple-Mapping Adaptive Coupling (TMAC):

While sequence-based models like NC-Mamba and MSA are effective at modeling long-range dependencies, they often overlook localized spatial cues. To address this limitation, we introduce the TMAC module, which augments each attention block with a lightweight yet expressive spatial refinement mechanism. By incorporating channel splitting and triple mapping, it enables fine-grained spatial focus and robust feature alignment.

Channel Splitting: To enhance spatial specialisation and reduce interference among heterogeneous features, we adopt a Dual-Channel structure by splitting the input feature map along the channel dimension:

$$\{\mathbf{X}_1, \mathbf{X}_2\} = \text{Split}(\text{Flatten}(\mathbf{X})), \quad \mathbf{X}_m \in \mathbb{R}^{B\times L\times \frac{C}{2}}, \quad m \in \{1,2\}$$

Triple Mapping: TMAC generates two idiosyncratic attention maps (one for each sub-path) and one consensus attention map.

$$\mathbf{A}_m^{idi} = \sigma\left(\text{Conv}\left(\text{concat}\left[\underset{\frac{C}{2}}{\text{mean}}(\mathbf{F}_m^{idi}), \underset{\frac{C}{2}}{\max}(\mathbf{F}_m^{idi})\right]\right)\right), \quad m \in \{1,2\}$$

To capture common regions of interest, we compute the consensus feature map:

$$\mathbf{F}^{cons} = \sum_{m=1}^{2} \mathbf{F}_m^{idi}, \quad \mathbf{F}^{cons} \in \mathbb{R}^{H\times W\times \frac{C}{2}}$$

The consensus attention map: (All three attention branches share weights to ensure consistency and reduce complexity.)

$$\mathbf{A}^{cons} = \sigma\left(\text{Conv}\left(\text{concat}\left[\underset{\frac{C}{2}}{\text{mean}}(\mathbf{F}^{cons}), \underset{\frac{C}{2}}{\max}(\mathbf{F}^{cons})\right]\right)\right)$$

Adaptive Mamba Head:

Due to Mamba's continuous state updates, it possesses stronger global modeling capability compared to convolutions. Leveraging this, we design a lightweight detection head by employing a single CellMamba block in both the classification and box regression branches.

$$\text{Class\_sub} = \text{Conv}_1(\text{Reshape}(\text{CellMamba}(\text{Flatten}(\alpha_t \mathrm{P}_i)))) \in \mathbb{R}^{K\times H'\times W'}, \quad i \in [2,6]$$

$$\text{Box\_sub} = \text{Conv}_2(\text{Reshape}(\text{CellMamba}(\text{Flatten}(\alpha_t \mathrm{P}_i)))) \in \mathbb{R}^{4\times H'\times W'}, \quad i \in [2,6]$$

Experiment

Datasets:

We evaluate our method on two publicly available histopathological datasets, CoNSeP and CytoDArk0 dataset, adapting them for bounding-box-based cell detection. The CoNSeP dataset consists of H&E-stained colorectal adenocarcinoma images with instance segmentation annotations. The CytoDArk0 comprises Nissl-stained brain tissue images at 40× magnification. While CoNSeP is designed for nucleus detection and CytoDArk0 for cell detection, evaluating on these datasets fully validates its capabilities across multiple pathological detection tasks.

Experiment Results:

We compare our proposed model with state-of-the-art (SOTA) methods across three architecture categories: CNN, Transformer, and Mamba

Category	Model	CoNSeP (mAP %)			CytoDArk0 (mAP %)		
		mAP	@50	@75	mAP	@50	@75
CNN	RetinaNet (ICCV'17)	19.1	39.3	17.1	49.1	77.9	56.1
	Mask R-CNN (ICCV'17)	17.1	35.9	16.3	48.8	78.5	55.0
Transformer	Deformable-DETR (ICLR'21)	23.8	44.2	23.5	47.9	79.3	51.7
	DINO (ICLR'23)	24.2	45.1	**24.2**	53.0	81.5	**61.2**
Mamba	VSSD-Micro-R (ICCV'25)	24.5	48.2	23.3	50.5	80.8	55.5
	VSSD-Micro-M (ICCV'25)	23.1	44.8	22.5	47.6	78.3	50.9
	Mamba-YOLO-Base (AAAI'25)	25.2	50.7	23.6	52.3	81.2	56.6
	MobileMamba-B1-R (CVPR'25)	22.8	43.9	22.6	41.7	72.3	43.9
	MobileMamba-B1-M (CVPR'25)	23.2	45.3	22.9	44.1	74.9	48.2
	2D-Mamba-R (CVPR'25)	16.8	34.2	15.5	40.1	68.3	41.5
	2D-Mamba-M (CVPR'25)	20.4	40.2	19.7	42.8	74.8	45.6
	Spatial-Mamba-Tiny-R (ICLR'25)	21.1	40.8	20.8	47.7	77.8	51.2
	Spatial-Mamba-Tiny-M (ICLR'25)	22.2	42.5	20.7	48.3	78.4	53.9
	Ours	**25.7**	**51.1**	23.8	**53.3**	**83.5**	59.8

Efficiency Analysis:

Model	Params (M)	Time (ms)	mAP@50
CNN			
RetinaNet (ICCV'17)	36.0	3.4	77.9
Mask R-CNN (ICCV'17)	44.0	4.3	78.5
Transformer			
Deformable-DETR (ICLR'21)	40.0	1.8	79.3
DINO (ICLR'23)	48.0	4.5	81.5
Mamba			
VSSD-Micro-R (ICCV'25)	21.4	2.1	80.8
VSSD-Micro-M (ICCV'25)	33.0	5.3	78.3
Mamba YOLO-Base (AAAI'25)	22.0	2.2	81.2
MobileMamba-B1-R (CVPR'25)	27.1	6.6*	72.3
MobileMamba-B1-M (CVPR'25)	38.0	6.5	74.9
2D-Mamba-R (CVPR'25)	38.8	2.4	68.3
2D-Mamba-M (CVPR'25)	49.0	6.9	74.8
Spatial-Mamba-Tiny-R (ICLR'25)	36.3	5.8	77.8
Spatial-Mamba-Tiny-M (ICLR'25)	46.0	6.2	78.4
Ours	**14.7**	**1.6**	**83.5**

Our model achieves the best overall trade-off—only 14.7M parameters and 1.6ms latency per 256×256 patch. Overall, we adopt Mamba as a powerful yet lightweight feature extractor. Instead of introducing complex attention modules, we enhance the model's focus on fine-grained features via targeted structural designs, and parameter sharing among parallel modules further reinforces its lightweight property.

Conclusion

In this paper, we propose CellMamba, a lightweight and accurate one-stage cell detector designed for high-resolution pathological images. The architecture combines a hierarchical backbone with the proposed TMAC module, which enhances spatial focus through complementary attention maps and adaptive fusion. In addition, an adaptive Mamba head improves detection across varied cell sizes and densities. These findings underscore the effectiveness of structured state-space models in dense and fine-grained biomedical detection and offer a scalable framework for future research in computational pathology.