Category Prompt Mamba Network for Nuclei Segmentation and Classification

Ye Zhang[1, 2], **Zijie Fang**[3], **Yifeng Wang**[4], **Lingbo Zhang**[3], **Xianchao Guan**[1, 5], **Yongbing Zhang**[1*]

[1] School of Computer Science and Technology, Harbin Institute of Technology (Shenzhen)
[2] Leibniz-Institut für Analytische Wissenschaften – ISAS – e.V.
[3] Tsinghua Shenzhen International Graduate School, Tsinghua University
[4] School of Science, Harbin Institute of Technology (Shenzhen)
[5] Pengcheng Laboratory
zhangye94@stu.hit.edu.cn, ybzhang08@hit.edu.cn

Abstract

Nuclei segmentation and classification provide an essential basis for tumor immune microenvironment analysis. The previous nuclei segmentation and classification models require splitting large images into smaller patches for training, leading to two significant issues. First, nuclei at the borders of adjacent patches often misalign during inference. Second, this patch-based approach significantly increases the model's training and inference time. Recently, Mamba has garnered attention for its ability to model large-scale images with linear time complexity and low memory consumption. It offers a promising solution for training nuclei segmentation and classification models on full-sized images. However, the Mamba orientation-based scanning method lacks account for category-specific features, resulting in sub-optimal performance in scenarios with imbalanced class distributions. To address these challenges, this paper introduces a novel scanning strategy based on category probability sorting, which independently ranks and scans features for each category according to confidence from high to low. This approach enhances the feature representation of uncertain samples and mitigates the issues caused by imbalanced distributions. Extensive experiments conducted on four public datasets demonstrate that our method outperforms state-of-the-art approaches, delivering superior performance in nuclei segmentation and classification tasks.

Introduction

With the advancements of whole-slide pathology image production and scanning technologies, pathological image analysis tasks represented by nuclei segmentation and classification (Ilyas et al. 2022; Oh and Jeong 2023; Zhang et al. 2024) play a more and more critical role in cancer diagnosis (Kowal and Filipczuk 2014) and patient prognosis analysis (Wang et al. 2022). The morphology and category of nuclei can reflect the cellular differentiation degree and distribution status, which provide valuable information for the tumor micro-environment analysis (Zamanitajeddin et al. 2024).

Recent developments in deep learning have enabled automatic nuclei segmentation and classification (Doan et al. 2022; Pan et al. 2023). However, due to the large size of pathological images, most existing nuclei analysis approaches require splitting these images into smaller, independent patches for training (Ronneberger, Fischer, and Brox 2015; Chen et al. 2016), as illustrated in Fig.1 (a). This patch-based training method presents two significant challenges. **First**, images often suffer from edge effect problems during inference, where nuclei at the borders of adjacent patches are misaligned. **Second**, when the distribution of nuclear categories is highly imbalanced within a single patch, the performance of rare categories tends to deteriorate (Sirinukunwattana et al. 2016; Naylor et al. 2018). This decline is primarily due to the insufficient training of rare categories, stemming from their limited sample sizes.

To address the patch misalignment issue in large-scale images, HoverNet (Graham et al. 2019) introduces a center feature clipping strategy to reduce the uncertainty associated with edge features. Similarly, DoNuSeg (Wang et al. 2024a) employs an overlapped sampling strategy to mitigate edge effects. However, these approaches can merely alleviate misalignment; they do not fundamentally resolve the issue and often lead to increased model training and inference times. In tackling the class imbalance problem, some

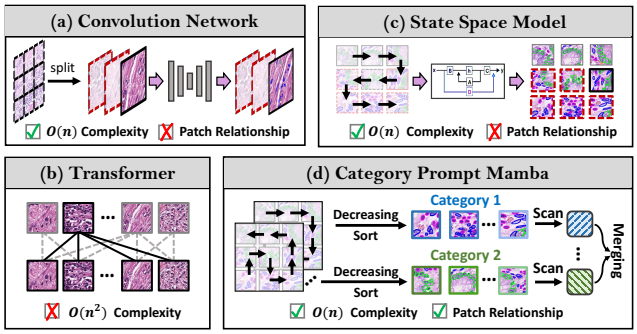

Figure 1: The existing nuclei analysis framework. The images with **black box** represent current training samples, and the images with **red box** do not participate in training. (a) represents convolution network independently trains each image; (b) represents Transformer structure has quadratic computational complexity; (c) represents state space model only considers the preceding samples; (d) is our proposed category prompt network, which utilizes the classification probability as a basis to guide the sequences sorting.

*Corresponding author: Yongbing Zhang

transformer-based methods are proposed (He et al. 2023; Lou et al. 2024a). These approaches enhance the representation of rare categories through global feature interaction. By integrating information from distant regions, transformers can better contextualize rare categories, improving their prediction performance. However, the quadratic computational complexity inherent to the transformer architecture restricts the size of the input image and imposes substantial demands on memory resources, as illustrated in Fig. 1 (b). Additionally, although class-weighted loss functions are proposed to address class imbalance (Schmitz et al. 2021; Hancer et al. 2023), their impact on improving performance remains limited. These limitations underscore the need for more efficient and effective solutions to patch misalignment and class imbalance in nuclei segmentation and classification.

The state space model (SSM) represented by Mamba (Gu and Dao 2023) has recently garnered widespread attention for its ability to perform long-sequence modeling with linear complexity. Mamba's low memory consumption further enables training nuclei segmentation and classification networks directly on large-scale images, thereby eliminating the need for image splitting. However, Mamba's unidirectional scanning inherently limits interaction between patches, as it only considers preceding sequences and lacks awareness of subsequent ones, as illustrated in Fig. 1 (c). To overcome the issue, several multi-directional scanning strategies are proposed. Vim (Zhu et al. 2024) implements forward and backward scanning to enhance interaction across all patches. In contrast, VMamba (Liu et al. 2024) introduces a four-directional cross-scan method that considers both horizontal and vertical directions. Additionally, other scanning strategies (Li, Singh, and Grover 2025; Yang et al. 2024) are developed to more thoroughly address the influence of sequential order. Despite these advancements, the current scanning strategies are orientation-aware rather than class-aware. In class imbalance scenarios, these methods often fail to adequately recognize and enhance class-specific features, leading to suboptimal performance. Hence, more sophisticated approaches are needed to address class imbalance in nuclei segmentation and classification effectively.

This paper addresses the challenges of class imbalance and large-scale image prediction by introducing a probability-guided sorting method built on the Mamba network. Our approach involves learning the feature representation of each category independently and then aggregating them, as depicted in Fig. 1 (d). This method enhances the feature representation of rare categories, thereby improving the accuracy of their predictions. Specifically, we utilize category prompts as supervision to predict multi-category labels, with the predicted probabilities reflecting the confidence level for each category. Based on these probability outputs, the feature sequences are sorted and scanned in descending order, allowing low-confidence features to learn embeddings from high-confidence features. We also provide theoretical proof demonstrating that this probability-guided scanning method offers superior feature learning compared to random scanning. Notably, our method enables direct training on large-scale images without data splitting, eliminating the patch misalignment issue.

In summary, the contributions of this paper are four folds:

(1) We propose a novel category prompt Mamba block in nuclei segmentation and classification network, which enables direct network training on large-scale images without requiring a data splitting process.

(2) We design a patch-level category prompt method, which employs multi-class labels as supervision information to help the network learn class-related features.

(3) We introduce a probability-guided sorting and scanning strategy to enhance the representation of rare categories. We also provide theoretical proof of the sorting method.

(4) We conduct extensive experiments on the four nuclei segmentation and classification datasets, achieving state-of-the-art (SOTA) performance while significantly improving training efficiency.

Related Work

Nuclei Segmentation and Classification

In recent years, deep learning has revolutionized nuclei analysis by eliminating the need for complex threshold selection (Win et al. 2017) and feature extraction processes (Xu et al. 2016; Lou et al. 2024b). Current research in nuclei segmentation focuses on addressing the challenge of boundary overlap. For instance, DCAN (Chen et al. 2016) introduces a separate contour prediction branch to enhance segmentation accuracy. Methods like HoverNet (Graham et al. 2019), and Dist (Naylor et al. 2018) utilize a distance regression branch to improve edge discrimination. Similarly, CDNet (He et al. 2021) and SONNET (Doan et al. 2022) discretize the original distance map to refine segmentation outcomes. Additional methods continue to advance the field of nuclear segmentation (Ahmad et al. 2023; He et al. 2023). In the domain of nuclei classification, several models have been proposed based on graph neural networks (GNNs). For example, MPNet (Hassan et al. 2022) introduces a message-passing network to aggregate global information. While EAGNN (Hasegawa et al. 2023) incorporates edge labels into the GNN architecture. SENC (Lou et al. 2024b) enhances nuclei representation through nuclear structure learning, and CGT (Lou et al. 2024a) combines GNNs with Transformers to establish graph edge relationships. Despite these advancements, imbalanced nuclear category distribution remains a significant hurdle. This paper introduces a novel category prompt network designed to enhance classification accuracy under imbalanced scenarios.

Mamba in Medical Image Analysis

The state space model (SSM) is well-regarded for its ability to capture long sequence dependencies with linear time complexity. Among the SSM models, Mamba (Gu and Dao 2023) gains significant attention due to its state-space selection mechanism, which allows for dynamic adjustment of model parameters based on the input. This adaptability leads to the widespread adoption of Mamba-based models in various domains, including medical image analysis. Several models have successfully integrated Mamba as a core component. For instance, U-Mamba (Ma, Li, and Wang 2024)

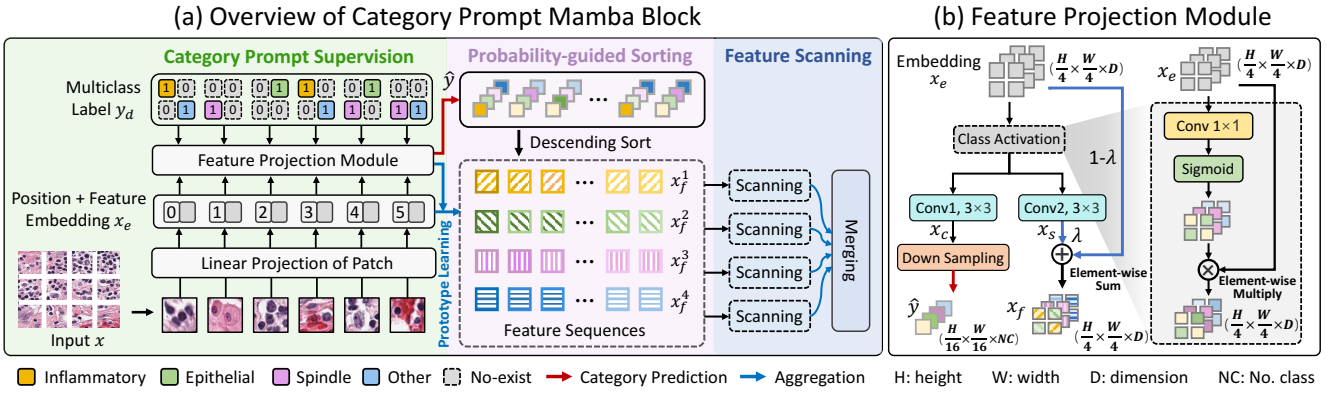

Figure 2: **The framework of our proposed CP-Mamba.** The method employs the category supervision information to learn each type nuclear prototype. Meantime, the probability predictions provide the guides for feature sequence ordering.

and Mamba-UNet (Wang et al. 2024b) utilize Mamba as a feature extractor. Additionally, models like Vim (Zhu et al. 2024), and VMamba (Liu et al. 2024) introduce bidirectional feature scanning and four-direction cross-scanning techniques, respectively, to enhance the feature extraction capabilities of the encoder. These approaches demonstrate effectiveness in various tasks, including semantic and instance segmentation. Further advancements in sequence scanning mechanisms are made with methods such as Zigzag Scan (Hu et al. 2024), Omnidirectional Selective Scan (Shi et al. 2024), and Hierarchical Scan (Zhang et al. 2025). These techniques aim to optimize the scanning process for better feature representation. In the field of multiple instance learning, Mamba-based methods (Fang et al. 2024; Yang, Wang, and Chen 2024) are proposed to leverage Mamba's sequence modeling capabilities, enhancing the interaction between instances. In this paper, we introduce a nuclei segmentation and classification network, which allows for direct training on large-scale images, eliminating the need for patch-based image splitting processing.

Methods

Overview

In this work, we propose a category prompt Mamba (CP-Mamba) as encoder block for feature extraction. Leveraging Mamba's low memory usage and long-sequence modeling capabilities, our approach enables direct training on large-scale images, eliminating the need for image splitting. To tackle the class imbalance challenge, a category phenotype learning module and a probability-guided sorting strategy are designed employing category prompt information to enhance the network's ability to represent class-specific features. The overall framework is illustrated in Fig. 2.

Category Prompt Supervision

Traditional pixel-by-pixel classification methods often struggle with class-imbalanced instance segmentation tasks. These methods treat each pixel equally, overlooking the class distribution. Previous studies (Yue et al. 2024) suggest that category information, as a weakly supervised signal, can help the network learn positive activation features, thereby improving segmentation and classification performance. To enhance the network's perception of nuclei class distribution and improve segmentation performance, we design a category prompt supervision method based on patch-level category information, as illustrated in Fig. 2 (a).

Algorithm 1: Multi-class Labels Generation

Input: Nuclei Ground-truth Classification Mask: y;
Output: Down-sampled Multi-class Label: y_d;
Calculating the Dimension of y: $h, w \leftarrow Shape(y)$;
Down-sampling scale 16
Initializing the Multi-class Label: $y_d \in \{0\}^{\frac{h}{16} \times \frac{w}{16} \times NC}$;
for $i \in \{1, \cdots, NC\}$ **do**
  for $h', w' \in \{0, 16, 32, 48, \cdots\}$ **do**
    if $i \in y_{[h':h'+16, w':w'+16]}$:
      $y_{d[h'/16, w'/16, i]} = 1$
  end for
end for
return y_d # Size(y_d)=($\frac{h}{16} \times \frac{w}{16} \times NC$)

Given an input image $x \in \mathcal{R}^{H \times W \times C}$, we first reshape it into a 2D patch sequence $x_p \in \mathcal{R}^{N \times (P^2 \cdot C)}$, where (H, W) represents the image resolution, C is the number of channels, N is the number of patches, and P is the patch size. For this process, we set the patch size to 4×4. The patch sequences are then fed into the network, where they are concatenated with positional embeddings to generate feature embeddings $x_e = [f(x_p), x_{pos}]$, where f denotes the linear projection layers. Next, we designed a category prompt supervision method. This method uses the multi-class labels to supervise feature learning, with a key component being the feature projection module shown in Fig. 2 (b). In this module, x_e is input into the feature projection module to obtain a new class-related feature representation x_c:

$$x_c = Conv_1(CA(x_e)), \tag{1}$$

where $Conv_1$ denotes a 3×3 convolution layer, and CA represents class activation layers consisting of a 1×1 convolution layer followed by a sigmoid activation function. The

1×1 convolution layer and activation function allow the network to activate features corresponding to specific classes selectively. By multiplying these activated features with the original features, the network can effectively extract and enhance the feature representation for each class.

Predicting category labels directly on a 4×4 patch presents significant challenges. Such a small patch size is insufficient to encompass an entire nucleus, making it challenging to capture the semantic features necessary for accurate classification. Additionally, using smaller patches increases computational complexity due to the more significant number of patches involved in the loss calculation. To address these issues, we apply a down-sampling operation to the output x_c, as illustrated on the left side of Fig. 2 (b):

$$\hat{y} = DS(x_c), \tag{2}$$

where DS denotes the down-sampling operation with a scale factor of 4. This ensures that each pixel in the feature map $\hat{y}$ corresponds to a 16×16 region in the original input x, comparable to the size of a nucleus. After obtaining $\hat{y}$, we compute the multi-label classification loss using the category prompt labels y_d, derived from the ground-truth nuclei classification label y. The generation process of y_d is detailed in Algorithm 1. The loss function is defined as:

$$L_p = MCE(y_d, \hat{y}), \tag{3}$$

where MCE represents the multi-class cross-entropy loss.

Category Phenotype Learning

Following a similar process to the category prompt supervision, the feature x_e is input into the CA module followed by a 3×3 convolution layer, obtaining a new category phenotype representation x_s:

$$x_s = Conv_2(CA(x_e)). \tag{4}$$

In this procedure, since x_s shares the same class activation module as the category prompt supervision, x_s primarily contains class-related semantic features. To better capture feature embeddings relevant to segmentation and classification tasks, we design a feature fusion method as shown in the right branch of Fig. 2 (b). In the design, x_s is fused with feature x_e to consider the class-independent semantic features from the original input. The fused feature will be used for sequence sorting. The fusion method is shown as follows:

$$x_f = \lambda \cdot x_s + (1 - \lambda) \cdot x_e, \tag{5}$$

where λ is a weighting parameter, which combines the original segmentation-related information with the class-related semantic information, enhancing the feature representation x_f. In this paper, we set λ to 0.2, and more parameter ablation experiments are provided in **supplementary materials**.

Probability-guided Sequence Sorting

Previous Mamba network scanning methods are direction-based (Wang et al. 2024b; Liu et al. 2024), which may not effectively handle class-imbalanced samples. When only a few patches contain a specific category, the distance between these patches can hinder the network's ability to extract features for that category. To address this, we propose a probability-guided sequence sorting method that encourages low-confidence features to learn from high-confidence ones. Hence, this sorting mechanism allows preceding patches to provide category-prior knowledge for subsequent patches, reducing classification uncertainty.

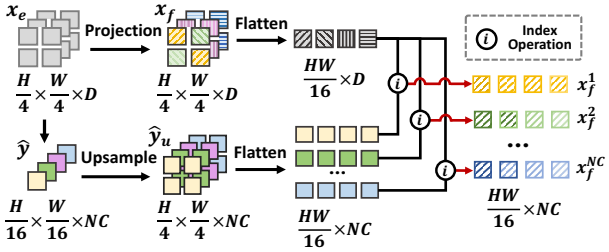

Figure 3: **The category sorting method guided by probability prediction.** "i" represents the sorting by index.

Guided by the probability prediction $\hat{y}$, we implement the probability-guided sorting method shown in Fig. 3. Since $\hat{y}$ is the output of the down-sampling operation and does not align with x_f in the feature dimension, we first up-sample $\hat{y}$ to generate $\hat{y}_u$. Next, we reshape x_f and $\hat{y}_u$ to dimensions $\frac{HW}{16} \times D$ and $\frac{HW}{16} \times NC$, respectively, where D is the feature dimension and NC is the number of categories. We then sort flattened sequences from highest to lowest based on the i^{th} class probability ordering and generate sorted feature sequence $\{x_f^i, i = 1, \cdots, NC\}$ for each category. In the end, these sequences are utilized for feature scanning, which is detailed in the following subsection. To prove the effectiveness of the sorting method, we give the following theorem.

Theorem *Given a random sorting feature sequence* $X_{seq} = \{X_1, X_2, \cdots, X_n\}$*, we use the joint entropy* $H(X_{seq})$ *to represent the classification uncertainty when* X_{seq} *used as classification features. When the sequence is sorted according to task-related confidence level from highest to lowest, then the joint entropy of the sequence will decrease, i.e.* $H(X_1^{'}, X_2^{'}, \cdots, X_n^{'}) \leq H(X_1, X_2, \cdots, X_n)$.

Where $\{X_1^{'}, X_2^{'}, \cdots, X_n^{'}\}$ represents a sorted sequence according to classification confidence level from highest to lowest. This theorem implies that probability-guided sorting can decrease the uncertainty of classification tasks and the proof of the theorem is provided in the **supplementary materials**. In addition, we conducted ablation experiments to analyze the effectiveness of the sorting method.

Network Training

Our training network follows an encoder-decoder architecture. The encoder consists of four CP-Mamba blocks, each consisting of a category phenotype learning process described previously and a feature aggregation process. In the feature aggregation process, we adopt summary operation as VMamba (Liu et al. 2024), which can be detailed as follows:

$$x_{enc} = \sum_{i=1}^{NC} S(x_f^i), \tag{6}$$

where S represents the feature scanning operation as Mamba (Gu and Dao 2023). During the encoding process, category

Datasets	Methods	DICE	AJI	DQ	SQ	PQ	Datasets	Methods	DICE	AJI	DQ	SQ	PQ	Publications
GLySAC	Mask-RCNN†	74.59	60.66	75.04	73.96	56.15	CoNSeP	Mask-RCNN†	74.96	51.88	**66.11**	75.73	49.81	ICCV'2017
	HoverNet‡	80.76	64.21	78.66	**76.54**	61.49		HoverNet†	81.97	53.64	64.14	76.29	50.38	MIA'2019
	Triple-UNet‡	75.27	62.03	76.15	74.10	57.87		Triple-UNet‡	80.39	39.25	49.81	74.66	37.25	MIA'2020
	DoNet†	75.22	61.89	74.34	74.21	56.99		DoNet†	78.23	46.76	57.27	72.89	45.64	CVPR'2023
	Vim‡	79.82	63.11	77.29	75.02	59.49		Vim‡	80.64	49.21	62.88	75.94	49.80	ICML'2024
	Ours‡	**81.43**	**64.87**	**79.02**	75.34	**61.77**		**Ours‡**	**82.23**	**53.90**	65.81	**77.67**	**51.46**	-
Datasets	Methods	DICE	AJI	DQ	SQ	PQ	Datasets	Methods	DICE	AJI	DQ	SQ	PQ	Publications
MoNuSAC	Mask-RCNN†	73.23	60.80	74.18	78.27	59.44	PanNuke	Mask-RCNN†	75.06	61.24	72.34	78.69	60.19	ICCV'2017
	HoverNet‡	74.41	61.27	**75.88**	79.61	60.47		HoverNet‡	80.36	65.64	75.82	80.19	61.22	MIA'2019
	Triple-UNet†	50.84	43.18	62.15	67.67	39.65		Triple-UNet†	74.24	58.83	66.71	73.58	54.06	MIA'2020
	DoNet†	70.03	59.45	70.90	75.30	58.49		DoNet†	78.24	66.86	74.90	79.29	59.91	CVPR'2023
	Vim‡	74.48	62.11	74.97	76.77	58.23		Vim ‡	79.69	64.89	75.59	81.22	60.89	ICML'2024
	Ours†	**75.41**	**62.76**	75.40	**80.08**	**60.68**		**Ours†**	**81.87**	**67.90**	**76.79**	**81.90**	**61.49**	-

Table 1: The nuclei segmentation comparison with the state-of-the-art methods on the GlySAC, ConSeP, MoNuSAC and PanNuke datasets. † represents p-value of AJI < 0.001 and ‡ represents p-value of AJI < 0.05.

prompt supervision and probability-guided sorting facilitate patch interaction and allow for the independent extraction of class-related features. Hence, they strengthen the representation of rare classes. The detailed training architecture is provided in the **supplementary materials**.

The decoder comprises two parallel U-Net decode branches. The first branch is designed to learn the foreground, background, and contour semantic features, while the second branch focuses on the classification task. Instance segmentation is then derived from the post-processing operation. The overall training loss is defined as:

$$L = L_p + \alpha L_{sem} + \beta L_{cls}, \tag{7}$$

where L_{sem} represents the loss of semantic branch, L_{cls} represents the classification loss, and L_p represents the loss of the category prompt supervision. The parameters α and β balance the loss weight. We set α and β to 1 in the paper. In L_{sem} and L_{cls}, we simultaneously employ cross-entropy and dice losses to optimize the objective.

Experiments

Datasets

We evaluate our proposed model on four pathological datasets, including GLySAC (Doan et al. 2022), CoNSeP (Graham et al. 2019), MoNuSAC (Verma et al. 2021), and PanNuke (Gamper et al. 2019) datasets. The **GLySAC** includes 59 H&E stained images of size 1000×1000 pixels from 8 gastric adenocarcinoma WSIs digitized at 40×magnification. The dataset includes 30975 annotated nuclei grouped into three categories: lymphocytes, epithelial, and miscellaneous. The **ConSeP** involves 41 H&E stained images of size 1000×100 pixels. The dataset includes 24319 annotated nuclei grouped into four types: miscellaneous, inflammatory, epithelial, and spindle. The **MoNuSAC** comprises 209 annotated images, ranging from 81×113 pixels to 1422×2162 pixels. The dataset comprises 31411 annotated nuclei of four categories: epithelial nuclei, lymphocytes, macrophages, and neutrophils. The **PanNuke** includes 7899 images of size 256×256 pixels obtained from 19 organs. The dataset contains 189744 annotated nuclei from five categories, including neoplastic, non-neoplastic epithelial, inflammatory, connective, and dead nuclei.

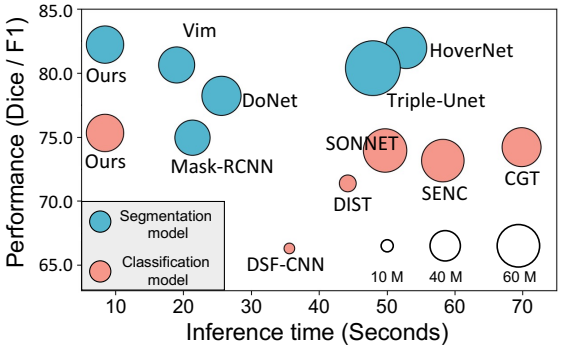

Figure 4: **The model complexity analysis.** The "green" bubbles represent the segmentation model, and the "red" bubbles represent the classification model. The size of the bubble represents the size of the model parameters.

Implementation Details and Evaluation Metrics

Our all experiments are run with PyTorch on two Nvidia RTX 4090 GPUs. We use SGD as an optimizer, and the learning rate, momentum, and weight decay are set at 0.01, 0.9, and 0.0005. Besides, we train the network for 6000 iterations. We evaluate the segmentation performance over metrics of DICE, AJI (Kumar et al. 2017), DQ (Kirillov et al. 2019), SQ, and PQ and evaluate the classification performance with an F1 score. Throughout all the tables in this paper, we bold the **best** and underline the second best.

Computational Complexity

To validate the superiority of our method in inference time, we first perform a complexity analysis as shown in Fig. 4, which shows the inference time, model parameters, and performance comparisons between our proposed method and other segmentation and classification models. We fix the input image size in this experiment as 1000×1000. In detail, segmentation models include Mask-RCNN (He et al. 2017), HoverNet (Graham et al. 2019), Triple-UNet (Zhao et al. 2020), DoNet (Jiang et al. 2023) and Vim (Zhu et al. 2024). The classification models include DIST (Naylor et al. 2018), DSF-CNN (Graham, Epstein, and Rajpoot 2020), SONNET

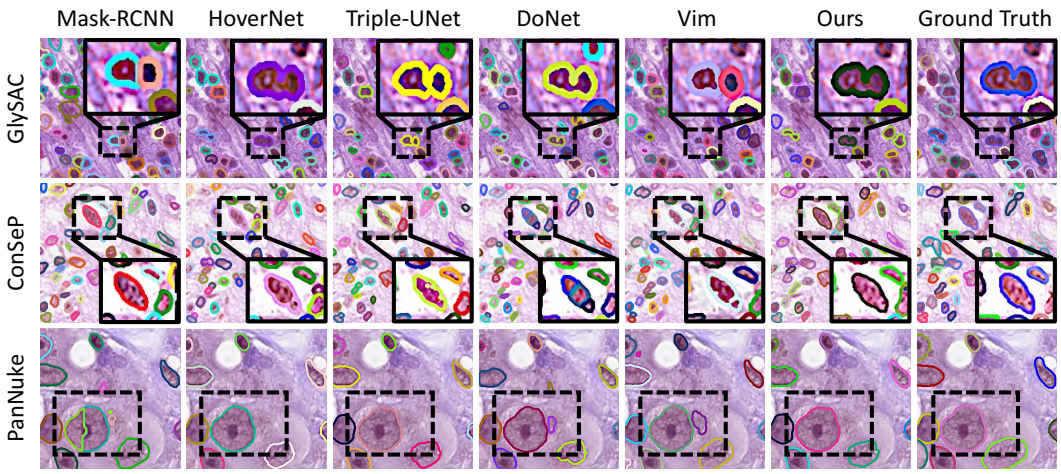

Figure 5: **The visualization comparison on the segmentation task.** The black boxes highlight the segmentation differences.

Datasets	Methods	F_d	F^E	F^L	F^M	-	Datasets	Methods	F_d	F^E	F^I	F^M	F^S	-	Publications
	DIST†	80.11	50.81	50.67	16.92	-		DIST†	71.39	61.65	60.89	17.18	52.67	-	TMI'2018
	DSF-CNN†	82.79	52.64	49.76	25.27	-		DSF-CNN†	66.31	56.79	54.17	11.97	9.32	-	TMI'2020
GLySAC	SONNET‡	83.29	52.57	51.76	32.68	-	ConSeP	SONNET‡	73.96	64.03	61.96	36.78	55.98	-	JBHI'2022
	CGT‡	85.76	55.23	52.61	34.94	-		CGT‡	74.22	64.68	56.23	39.66	54.17	-	AAAI'2024
	SENC‡	84.59	55.08	51.73	35.10	-		SENC‡	73.17	60.60	58.16	39.47	55.73	-	MIA'2024
	Ours‡	**86.94**	**56.15**	**53.49**	**35.92**	-		**Ours**‡	**75.33**	**65.78**	**63.28**	**40.96**	**57.17**	-	-
Datasets	Methods	F_d	F^E	F^L	F^{Ma}	F^N	Datasets	Methods	F_d	F^C	F^D	F^I	F^{Ne}	F^{No}	Publications
	DIST†	60.48	60.82	72.66	14.51	29.18		DIST†	71.80	41.90	2.15	42.63	50.00	34.17	TMI'2018
	DSF-CNN†	82.11	79.18	75.41	40.19	50.80		DSF-CNN†	78.28	44.94	6.89	47.20	59.34	50.09	TMI'2020
MoNuSAC	SONNET‡	**83.63**	**83.48**	78.57	45.73	57.19	PanNuke	SONNET‡	80.02	45.34	29.85	53.17	61.38	55.81	JBHI'2022
	CGT‡	81.59	80.44	79.67	41.64	52.08		CGT‡	78.68	47.27	28.96	52.67	59.80	62.13	AAAI'2024
	SENC‡	75.82	78.19	81.66	39.23	49.22		SENC‡	76.78	48.23	21.64	49.88	60.03	**65.30**	MIA'2024
	Ours‡	82.14	82.90	**82.89**	**46.17**	**58.23**		**Ours**‡	**81.70**	**50.04**	**31.30**	**53.93**	**62.16**	64.08	-

Table 2: The nuclei classification comparison with the state-of-the-art methods on the GLySAC, ConSeP, MoNuSAC and PanNuke datasets. † represents p-value of F_d <0.001 and ‡ represents p-value of F_d < 0.05. F^C, F^D, F^E, F^I, F^L, F^M, F^{Ma}, F^N, F^{Ne}, F^{No}, and F^S denote the F1 score for nuclear types of connective, dead, epithelial, inflammatory, lymphocyte, miscellaneous, macrophages, neutrophils, neoplastic, non-neo plastic, and spindle, respectively.

(Doan et al. 2022), CGT (Lou et al. 2024a), SENC (Lou et al. 2024b). The figure shows that our proposed method is significantly superior in classification and segmentation tasks. On the one hand, our model has less inference time because our method directly inferences on the large images. In contrast, other methods consist of an image splitting step, which enhances the inference time by linear time. On the other hand, our method can achieve better performance under less parameters, and the detailed performance comparison will be shown in the following subsections.

Comparison with the State-of-the-art Methods

Segmentation Performance We validate the effectiveness of our method with five state-of-the-art nuclei segmentation models on four datasets. These comparison methods include Mask-RCNN (He et al. 2017), HoverNet (Graham et al. 2019), Triple-UNet (Zhao et al. 2020), DoNet (Jiang et al. 2023) and Vim (Zhu et al. 2024) and comparison results are shown in Table 1. The results show that our method performs best in Dice and PQ metrics. In addition, compared with the second-best method, HoverNet, our method still significantly improved with an AJI increase of 0.6, 0.3, 0.5, and 0.2 on four datasets, respectively. Combined with the complexity analysis experiment of the previous subsection, the comparison results show that our method does not damage the segmentation performance and dramatically reduces the inference time when training on large-scale images.

Furthermore, the visualization comparisons of nuclei segmentation on GLySAC, ConSeP and PanNuke datasets are shown in Fig. 5. From the results we can see our prediction outputs are closer to the ground-truth labels. In detail, Mask-RCNN and Vim mistakenly divide a single nucleus into multiple ones. Meanwhile, DoNet and Vim incorrectly identify tissue regions as nuclei.

Classification Performance We compare the classification performance of our method with five SOTA classification networks, namely DIST (Naylor et al. 2018), DSF-CNN (Graham, Epstein, and Rajpoot 2020), SONNET (Doan et al. 2022), CGT (Lou et al. 2024a), and SENC (Lou et al. 2024b). The comparison performances are shown in Table 2. In this table, F_d represents the F1 score of the detection, and the others represent the F1 score of classification of

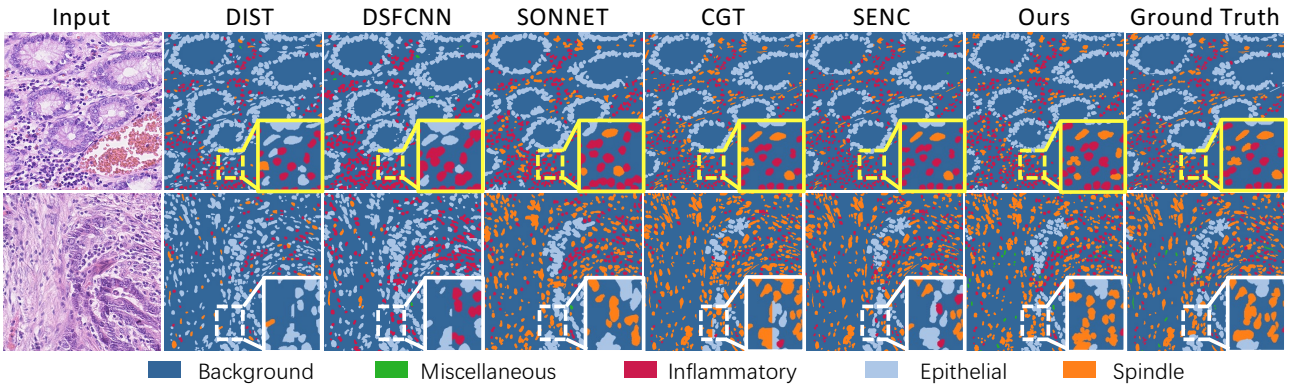

Figure 6: **The visualization comparison on the classification task.** The yellow boxes and white boxes highlight the misclassified and not-aligned nuclei respectively.

Scanning Direction	GlySAC			ConSeP			
	F^E	F^L	F^M	F^E	F^I	F^M	F^S
Bidirectional	53.65	53.12	29.27	64.49	49.14	25.33	56.86
Cross-Scanning	54.34	53.16	31.08	65.19	62.79	34.13	55.73
Probability Sorting	**56.15**	**53.49**	**35.92**	**65.78**	**63.28**	**40.96**	**57.17**

Table 3: The effect of scanning direction for nuclei classification on GlySAC and ConSeP datasets.

Sorting	Phenotype	GlySAC			ConSeP		
		Dice	AJI	F_d	Dice	AJI	F_d
		79.27	61.14	79.42	78.34	51.07	71.12
✓		80.09	62.82	80.20	79.98	52.21	72.22
	✓	79.79	63.34	84.75	80.09	52.23	73.83
✓	✓	**81.43**	**64.87**	**86.94**	**82.23**	**53.90**	**75.33**

Table 4: Ablation experiments of module designing on GlySAC and ConSeP datasets.

each type. Overall, our method's classification performance is better, especially in the face of an unbalanced distribution of categories. In the ConSeP dataset, the F^I scores of DIST and DSF-CNN on miscellaneous nuclei are only 17.18% and 11.97%, and our method has increased by more than 20%. Similar comparison results can be drawn from the MoNuSAC dataset. Such as, when facing rare miscellaneous and neutrophil nuclei, our method still performs best and beyond Dist around 30%.

Furthermore, we provide classification visualizations on the 1000×1000 images from the ConSeP in Fig. 6. In the experiment, we use ground-truth masks as segmentation results for CGT and SENC because they only execute classification tasks and do not have segmentation predictions. From the figure, our method performs better when facing spindle nuclei in yellow boxes. However, DIST and DSFCNN usually misclassified spindles into epithelial or inflammatory. In the meantime, according to the enlarged white areas, we can see methods employing image clipping cannot be aligned between the nucleus, such as DSFCNN and CGT. In contrast, our method has better nuclear integrity.

Ablation Study

To validate the effect of our probability sorting strategy on resolving class-imbalance problems, we compared the different scanning strategies on GLySAC and ConSeP datasets, including bidirectional scanning and cross-scanning. As seen from Table 3, our probability-guided sorting method has a significant improvement, especially in the miscellaneous, the bidirectional scanning method is 6% and 15% lower than our method, and the cross-scanning method is 4% and 6% lower than our method. The above results show that probability sorting vastly improves the classification performance of rare categories. Next, we conduct ablation analysis for the model design, including the probability sorting strategy and phenotype learning, and the results are shown in Table 4. Overall, when either of the two designs is removed, the model's performance is significantly reduced, indicating that the design of both modules is effective. In detail, when phenotype learning is not used, the classification performance decreases by more than 6% and 3% on the two datasets, respectively, indicating that the aggregated class-related features from the category prompt supervision are practical for model classification.

Conclusions

In this paper, we propose a probability-guided sorting method based on Mamba to solve the task of nuclei segmentation and classification in the case of class imbalance. This method uses category prompts to generate confidence probability for each category, and the prediction results are used as reference guide sequence sorting. Then, through independent feature scanning and aggregation of each category sequence, the method boosts the feature representation of rare categories to improve classification accuracy. In addition, our method can be trained directly on large-scale images without an image splitting process, which not only solves the problem of the nuclei not being aligned with the traditional method but also dramatically reduces the training and inference time of the model. We perform extensive comparative experiments on four datasets, and the experiment results also demonstrate the validity of our proposed method.

Acknowledgments

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

References

Ahmad, I.; Xia, Y.; Cui, H.; and Islam, Z. U. 2023. DAN-NucNet: A dual attention based framework for nuclei segmentation in cancer histology images under wild clinical conditions. *Expert Systems with Applications*, 213: 118945.

Chen, H.; Qi, X.; Yu, L.; and Heng, P.-A. 2016. DCAN: deep contour-aware networks for accurate gland segmentation. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, 2487–2496.

Doan, T. N.; Song, B.; Vuong, T. T.; Kim, K.; and Kwak, J. T. 2022. 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*, 26(7): 3218–3228.

Fang, Z.; Wang, Y.; Wang, Z.; Zhang, J.; Ji, X.; and Zhang, Y. 2024. Mammil: Multiple instance learning for whole slide images with state space models. *BIBM*.

Gamper, J.; Alemi Koohbanani, N.; Benet, K.; Khuram, A.; and Rajpoot, N. 2019. Pannuke: an open pan-cancer histology dataset for nuclei instance segmentation and classification. In *Digital Pathology: 15th European Congress, ECDP 2019, Warwick, UK, April 10–13, 2019, Proceedings 15*, 11–19. Springer.

Graham, S.; Epstein, D.; and Rajpoot, N. 2020. Dense steerable filter cnns for exploiting rotational symmetry in histology images. *IEEE Transactions on Medical Imaging*, 39(12): 4124–4136.

Graham, S.; Vu, Q. D.; Raza, S. E. A.; Azam, A.; Tsang, Y. W.; Kwak, J. T.; and Rajpoot, N. 2019. Hover-net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images. *Medical Image Analysis*, 58: 101563.

Gu, A.; and Dao, T. 2023. Mamba: Linear-time sequence modeling with selective state spaces. *ICML*.

Hancer, E.; Traore, M.; Samet, R.; Yıldırım, Z.; and Nemati, N. 2023. An imbalance-aware nuclei segmentation methodology for H&E stained histopathology images. *Biomedical Signal Processing and Control*, 83: 104720.

Hasegawa, T.; Arvidsson, H.; Tudzarovski, N.; Meinke, K.; Sugars, R. V.; and Ashok Nair, A. 2023. Edge-Based Graph Neural Networks for Cell-Graph Modeling and Prediction. In *International Conference on Information Processing in Medical Imaging*, 265–277. Springer.

Hassan, T.; Javed, S.; Mahmood, A.; Qaiser, T.; Werghi, N.; and Rajpoot, N. 2022. Nucleus classification in histology images using message passing network. *Medical Image Analysis*, 79: 102480.

He, H.; Huang, Z.; Ding, Y.; Song, G.; Wang, L.; Ren, Q.; Wei, P.; Gao, Z.; and Chen, J. 2021. Cdnet: Centripetal direction network for nuclear instance segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 4026–4035.

He, K.; Gkioxari, G.; Dollár, P.; and Girshick, R. 2017. Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*, 2961–2969.

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

Hu, V. T.; Baumann, S. A.; Gui, M.; Grebenkova, O.; Ma, P.; Fischer, J.; and Ommer, B. 2024. Zigma: Zigzag mamba diffusion model. *arXiv preprint arXiv:2403.13802*.

Ilyas, T.; Mannan, Z. I.; Khan, A.; Azam, S.; Kim, H.; and De Boer, F. 2022. TSFD-Net: Tissue specific feature distillation network for nuclei segmentation and classification. *Neural Networks*, 151: 1–15.

Jiang, H.; Zhang, R.; Zhou, Y.; Wang, Y.; and Chen, H. 2023. Donet: Deep de-overlapping network for cytology instance segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 15641–15650.

Kirillov, A.; He, K.; Girshick, R.; Rother, C.; and Dollár, P. 2019. Panoptic segmentation. In *CVPR*, 9404–9413.

Kowal, M.; and Filipczuk, P. 2014. Nuclei segmentation for computer-aided diagnosis of breast cancer. *International Journal of Applied Mathematics and Computer Science*, 24(1): 19–31.

Kumar, N.; Verma, R.; Sharma, S.; Bhargava, S.; Vahadane, A.; and Sethi, A. 2017. A dataset and a technique for generalized nuclear segmentation for computational pathology. *IEEE transactions on medical imaging*, 36(7): 1550–1560.

Li, S.; Singh, H.; and Grover, A. 2025. Mamba-nd: Selective state space modeling for multi-dimensional data. In *European Conference on Computer Vision*, 75–92. Springer.

Liu, Y.; Tian, Y.; Zhao, Y.; Yu, H.; Xie, L.; Wang, Y.; Ye, Q.; and Liu, Y. 2024. Vmamba: Visual state space model. *arXiv preprint arXiv:2401.10166*.

Lou, W.; Li, G.; Wan, X.; and Li, H. 2024a. Cell Graph Transformer for Nuclei Classification. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 4, 3873–3881.

Lou, W.; Wan, X.; Li, G.; Lou, X.; Li, C.; Gao, F.; and Li, H. 2024b. Structure embedded nucleus classification for histopathology images. *IEEE Transactions on Medical Imaging*.

Ma, J.; Li, F.; and Wang, B. 2024. U-mamba: Enhancing long-range dependency for biomedical image segmentation. *arXiv preprint arXiv:2401.04722*.

Naylor, P.; Laé, M.; Reyal, F.; and Walter, T. 2018. Segmentation of nuclei in histopathology images by deep regression of the distance map. *IEEE transactions on medical imaging*, 38(2): 448–459.

Oh, H.-J.; and Jeong, W.-K. 2023. Diffmix: Diffusion model-based data synthesis for nuclei segmentation and

classification in imbalanced pathology image datasets. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 337–345. Springer.

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

Ronneberger, O.; Fischer, P.; and Brox, T. 2015. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, 234–241. Springer.

Schmitz, R.; Madesta, F.; Nielsen, M.; Krause, J.; Steurer, S.; Werner, R.; and Rösch, T. 2021. Multi-scale fully convolutional neural networks for histopathology image segmentation: from nuclear aberrations to the global tissue architecture. *Medical image analysis*, 70: 101996.

Shi, Y.; Xia, B.; Jin, X.; Wang, X.; Zhao, T.; Xia, X.; Xiao, X.; and Yang, W. 2024. Vmambair: Visual state space model for image restoration. *arXiv preprint arXiv:2403.11423*.

Sirinukunwattana, K.; Raza, S. E. A.; Tsang, Y.-W.; Snead, D. R.; Cree, I. A.; and Rajpoot, N. M. 2016. Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images. *IEEE transactions on medical imaging*, 35(5): 1196–1206.

Verma, R.; Kumar, N.; Patil, A.; Kurian, N. C.; Rane, S.; Graham, S.; Vu, Q. D.; Zwager, M.; Raza, S. E. A.; Rajpoot, N.; et al. 2021. MoNuSAC2020: A multi-organ nuclei segmentation and classification challenge. *IEEE Transactions on Medical Imaging*, 40(12): 3413–3423.

Wang, Y.; Wang, Y. G.; Hu, C.; Li, M.; Fan, Y.; Otter, N.; Sam, I.; Gou, H.; Hu, Y.; Kwok, T.; et al. 2022. Cell graph neural networks enable the precise prediction of patient survival in gastric cancer. *NPJ precision oncology*, 6(1): 45.

Wang, Z.; Zhang, Y.; Wang, Y.; Cai, L.; and Zhang, Y. 2024a. Dynamic Pseudo Label Optimization in Point-Supervised Nuclei Segmentation. *Medical Image Computing and Computer Assisted Intervention Society*.

Wang, Z.; Zheng, J.-Q.; Zhang, Y.; Cui, G.; and Li, L. 2024b. Mamba-unet: Unet-like pure visual mamba for medical image segmentation. *arXiv preprint arXiv:2402.05079*.

Win, K. Y.; Choomchuay, S.; Choomchuay, S.; and Choomchuay, S. 2017. Automated segmentation of cell nuclei in cytology pleural fluid images using OTSU thresholding. In *2017 International Conference on Digital Arts, Media and Technology (ICDAMT)*, 14–18. IEEE.

Xu, H.; Lu, C.; Berendt, R.; Jha, N.; and Mandal, M. 2016. Automatic nuclei detection based on generalized laplacian of gaussian filters. *IEEE journal of biomedical and health informatics*, 21(3): 826–837.

Yang, C.; Chen, Z.; Espinosa, M.; Ericsson, L.; Wang, Z.; Liu, J.; and Crowley, E. J. 2024. PlainMamba: Improving Non-Hierarchical Mamba in Visual Recognition. *CoRR*.

Yang, S.; Wang, Y.; and Chen, H. 2024. Mambamil: Enhancing long sequence modeling with sequence reordering in computational pathology. *MICCAI*.

Yue, W.; Zhang, J.; Hu, K.; Xia, Y.; Luo, J.; and Wang, Z. 2024. Surgicalsam: Efficient class promptable surgical instrument segmentation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 7, 6890–6898.

Zamanitajeddin, N.; Jahanifar, M.; Bilal, M.; Eastwood, M.; and Rajpoot, N. 2024. Social network analysis of cell networks improves deep learning for prediction of molecular pathways and key mutations in colorectal cancer. *Medical Image Analysis*, 93: 103071.

Zhang, Y.; Wang, Y.; Fang, Z.; Bian, H.; Cai, L.; Wang, Z.; and Zhang, Y. 2024. DAWN: Domain-Adaptive Weakly Supervised Nuclei Segmentation via Cross-Task Interactions. *IEEE Transactions on Circuits and Systems for Video Technology*.

Zhang, Z.; Liu, A.; Reid, I.; Hartley, R.; Zhuang, B.; and Tang, H. 2025. Motion mamba: Efficient and long sequence motion generation. In *European Conference on Computer Vision*, 265–282. Springer.

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

Zhu, L.; Liao, B.; Zhang, Q.; Wang, X.; Liu, W.; and Wang, X. 2024. Vision mamba: Efficient visual representation learning with bidirectional state space model. *ICML*.