Knowledge-guided Multi-scale Graph Mamba for Whole Slide Image Classification

Minghong Duan[1,2], Zhiwei Yang[2,3], Yingfan Ma[1,2], Manning Wang[1,2,(✉)], and Zhijian Song[1,2,(✉)]

[1] Digital Medical Research Center, School of Basic Medical Science, Fudan University, Shanghai 200032, China
[2] Shanghai Key Lab of Medical Image Computing and Computer Assisted Intervention, Shanghai 200032, China
[3] Academy for Engineering & Technology, Fudan University, Shanghai 200433, China.
{mnwang, zjsong}@fudan.edu.cn

Abstract. Whole Slide Images (WSIs) are crucial for cancer diagnosis in digital pathology. WSI classification typically relys on Multiple Instance Learning (MIL). Existing MIL methods use attention mechanisms to highlight key instances but struggle to capture instance interactions. Although Transformers, State Space Models (SSMs), and Graph Neural Networks (GNNs) have made progress in solving this problem, they still face two main issues: (1) insufficient guidance from class-related information in modeling instance relationships, and (2) inadequate interaction between slides at different magnifications. To address these issues, we propose Knowledge-guided Multi-scale Graph Mamba (KMG-Mamba), which incorporates a Knowledge-guided Graph Representation (KGR) method for class-related guidance and Cross-scale Knowledge Interaction Mamba (CKIM) to facilitate effective cross-magnification information exchange. Experimental results on three public datasets show KMG-Mamba outperforms current MIL methods in WSI classification.

Keywords: Whole slide images · Multiple instance learning · Mamba.

1 Introduction

Pathological slide examination is the gold standard for cancer diagnosis. Deep learning-based automated classification of Whole Slide Images (WSIs) can significantly improve diagnostic efficiency [2], but the enormous size of WSIs and the requirement for detailed pixel-level annotations present substantial challenges [12, 22]. To address the challenges, weakly supervised learning methods that require only slide-level labels have been developed [18, 14, 15, 17]. Many of these methods are based on Multiple Instance Learning (MIL), where a WSI is treated as a bag and patches cropped from the slide are considered its instances within the bag. The final prediction is obtained by aggregating information from these instances, with attention mechanisms often used to highlight important ones. However, these methods typically overlook the contextual relationships between instances, leading to suboptimal WSI representations. Recent advancements have leveraged Transformers, State Space Models (SSMs), and Graph Neural Networks (GNNs) to model inter-instance relationships [25, 3, 27]. For example, Transformer-based methods such as TransMIL [20] explore instance relationships via self-attention mechanisms, while SSM-based approaches like MambaMIL [27] employ selective state space models to capture long-range dependencies. Meanwhile, GNNs have become effective tools for WSI analysis due to their ability to capture local similarities in the topological structure of entities, improving local relationship modeling [3]. As shown in Fig. 1A (a,b), GNN-based methods treat instance representations as nodes, constructing graphs by assigning edges based on feature similarity or spatial proximity [11, 4, 16, 21]. This enables efficient local information propagation through graph convolution or attention mechanisms. However, these methods lack category-specific knowledge to guide interactions between instances [11, 4, 16, 21], limiting the flow of useful information and reducing the model's ability to identify discriminative regions in WSIs. On the other hand, pathologists diagnose tumors by zooming in and out of magnification levels, highlighting the need for multi-scale analysis. As shown in Fig. 1B (a-c), although some methods have proposed feature concatenation or weighted fusion to integrate information from different magnifications [15, 9], these approaches lack effective interaction between features at different scales. Recent work using hierarchical Transformers for cross-scale feature interaction can capture long-range dependencies [8, 19], but suffer from quadratic computational complexity of the self-attention mechanisms, leading to high overhead when processing instances at different magnifications.

To address these challenges, this paper presents the Knowledge-guided Multiscale Graph Mamba (KMG-Mamba). Specifically, to guide the modeling of inter-instance relationships using category knowledge, we introduce a knowledge-guided graph representation (KGR) method. As shown in Fig. 1A(c), each instance feature is parameterized as a node containing base, head, and tail embeddings. A graph is constructed based on the directed connections between the head and tail of the nodes. The base embedding of the node with the highest degree of association is selected as the prototype, serving as the communication hub for instance interactions. Then, a prototype-guided graph aggregation module is employed to aggregate each node and its neighbors. To this end, we use an instance-level classifier to predict the prototype's logits and create a loss function with class labels for optimization, thereby guiding the graph representation with class knowledge. This ensures that, under class-related constraints, the graph focuses on learning discriminative information for WSI classification. To effectively utilize multi-scale information, we introduce the Cross-scale Knowledge Interaction Mamba (CKIM). As shown in Fig. 1B(d), inspired by the Selective Scan Space State Sequential model (Mamba) in computer vision [28, 26], we develop a cross-scale selective scanning module that models long sequences and promotes instance interactions across magnifications through compressed hidden states, facilitating information exchange at varying scales while maintaining linear complexity. Our contributions are as follows: (1) We propose the Knowledge-guided

Multi-scale Graph Mamba (KMG-Mamba) framework, which integrates class-specific knowledge to enhance graph representations for WSI classification; (2) We introduce the Cross-scale Knowledge Interaction Mamba to enable efficient instance interaction across magnifications; (3) Extensive comparison and ablation experiments on three publicly available datasets demonstrate the effectiveness of KMG-Mamba.

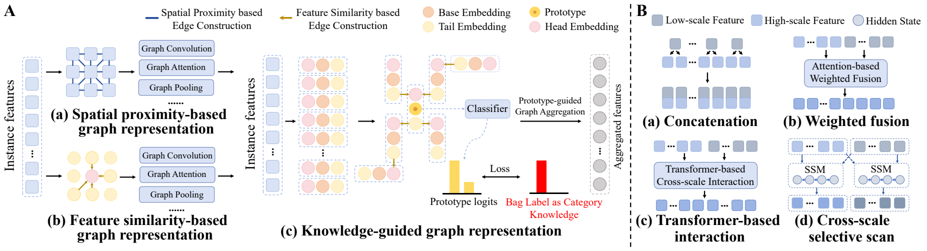
A
Spatial Proximity based Edge Construction
Feature Similarity based Edge Construction
Base Embedding
Tail Embedding
Prototype
Head Embedding
Instance features
Graph Convolution
Graph Attention
Graph Pooling
......
(a) Spatial proximity-based graph representation
Graph Convolution
Graph Attention
Graph Pooling
......
(b) Feature similarity-based graph representation
Instance features
Classifier
Prototype-guided Graph Aggregation
Aggregated features
Loss
Prototype logits
Bag Label as Category Knowledge
(c) Knowledge-guided graph representation
B
Low-scale Feature
High-scale Feature
Hidden State
(a) Concatenation
Attention-based Weighted Fusion
(b) Weighted fusion
Transformer-based Cross-scale Interaction
(c) Transformer-based interaction
SSM
SSM
(d) Cross-scale selective scan

Fig. 1. Motivation of our method. (A) Existing methods and our knowledge-guided graph representation for modeling the relationship between instances. (B) Existing methods and our cross-scale selective scan module for multi-scale processing.

2 Method

The pipeline of the proposed Knowledge-guided Multi-scale Graph Mamba (KMG-Mamba) is shown in Fig. 2(a). The low-scale and high-scale WSIs are cropped into patches and extracted as instance features, which are input into the Knowledge-guided Graph Representation module to construct the graph and generate prototypes. The prototype-guided graph aggregation module enhances cross-scale interactions and WSI representation, obtaining aggregated features, F_{LR} and F_{HR}. Prototypes are predicted by a linear classifier to generate prototype logits, which, along with class labels, form a loss function for optimization. F_{LR} and F_{HR} are then fed into the Cross-scale Knowledge Interaction Mamba, where a cross-scale selective scanning module facilitates interactions between instances across different scales. Finally, an attention-based aggregator and classifier produce the bag logits.

2.1 Preliminaries

State-space models (SSMs) maps a one-dimensional function or sequence $x(t) \in \mathbb{R}$ to the output $y(t) \in \mathbb{R}$ through the hidden state $h(t) \in \mathbb{R}^N$, as calculated as:

$$h'(t) = Ah(t) + Bx(t), y(t) = Ch(t) \tag{1}$$

where $h'(t)$ represents the time derivative of $h(t)$. Additionally, $A \in \mathbb{R}^{N\times N}$, $B \in \mathbb{R}^{N\times 1}$ and $C \in \mathbb{R}^{1\times N}$ are system matrices. To handle discrete sequences,

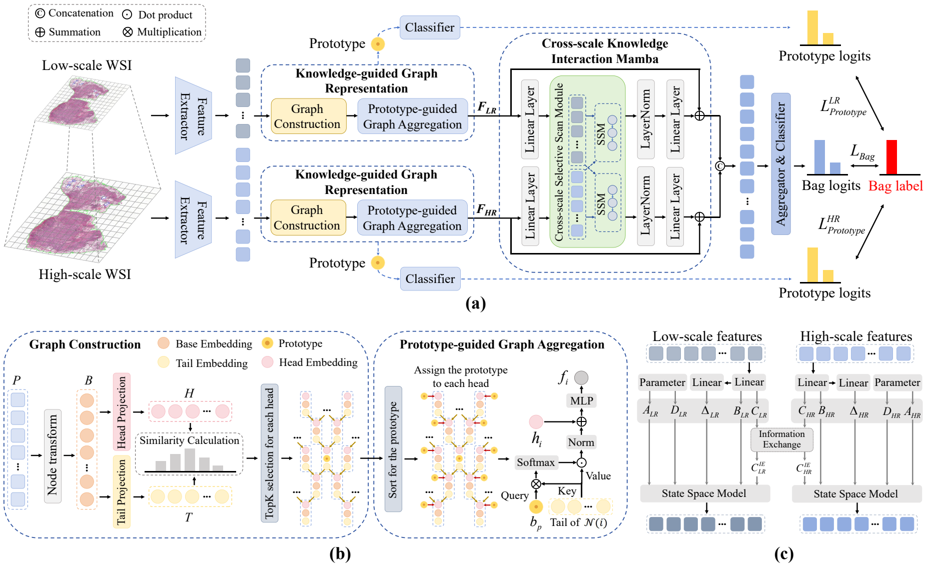

Fig. 2. (a) Pipeline of the proposed KMG-Mamba; (b) Knowledge-guided Graph Representation; (c) Cross-scale selective scanning module.

the SSM employs a zero-order hold (ZOH) discretization method, mapping the input sequence x_t to the output sequence y_t through the hidden state h_t. The specific discretization process can be expressed as:

$$\bar{A} = \exp(\Delta A), \bar{B} = (\Delta A)^{-1}(\exp(\Delta A) - I)\Delta B \approx \Delta B, \bar{C} = C \tag{2}$$

$$h_t = \bar{A}h_{t-1} + \bar{B}x_t, y_t = Ch_t \tag{3}$$

where Δ is used to convert the continuous parameters A and B into their discrete counterparts $\bar{A}$ and $\bar{B}$. Mamba [7] further introduces the selective scan mechanisms to ensure parameter dependency on the input, enabling the model to possess contextual awareness. This mechanism allows Mamba to effectively model complex interactions within long sequences.

2.2 Knowledge-guided Graph Representation

The proposed knowledge-guided graph representation is shown in Figure 2(b). It consists of two main components: graph construction and prototype-guided graph aggregation. Prototypes are generated during the graph construction process, and are predicted by a linear classifier as prototype logits. The prototype logits along with the category labels are used to construct a loss function that optimizes the training process. This approach effectively injects category knowledge into the graph representation and enables class-aware message passing, guiding both the graph construction and the information interaction between nodes.

Graph Construction The instance features $P = \{p_1, p_2, \ldots, p_L\}$ are mapped to the base embeddings $B = \{b_1, b_2, \ldots, b_L\}$ in the graph through a node transformation layer based on linear mapping. These are then further transformed into head embeddings $H = \{h_1, h_2, \ldots, h_L\}$ and tail embeddings $T = \{t_1, t_2, \ldots t_L\}$ using two additional linear mapping layers, thereby constructing a node for each instance that simultaneously contains base, head, and tail embeddings. The process of constructing the neighboring nodes $\mathcal{N}(i)$ for each node is as follows:

$$\mathcal{N}(i) = \left\{ t_j \in T : \omega_{i,j} \in Topk \left\{\omega_{i,j}\right\}_{j=1}^{L}, \omega_{i,j} = h_i^{\mathrm{T}} t_j \right\} \tag{4}$$

where $\omega_{i,j}$ denotes the similarity between the head embedding h_i and the tail embedding t_j. For each node, the top K nodes corresponding to the tail embeddings with the highest similarity scores to its head embedding are selected as its neighboring nodes. Each instance corresponding to a node can be associated with other instances through its head, or be linked to other instances via its tail, ensuring directed relationships and contributions between adjacent entities.

Prototype-guided Graph Aggregation We select the base embedding of the node with the highest degree in the graph as the prototype b_{pt}, and assign the prototype to the head embedding h_i of each node. Then the prototype-guided graph aggregation is achieved through cross-attention. In this process, b_{pt} is mapped as the Query q_{pt}, and the tail embeddings of $\mathcal{N}(i)$ linked to the head of each node are mapped as the Key $\{k_1, k_2, \ldots, k_K\}$ and Value $\{v_1, v_2, \ldots, v_K\}$. The detailed procedure is as follows:

$$f_i = W_b \left(Norm \left(\sum_{i=1}^{K} \left(\frac{\exp\left(q_{pt} k_i\right)}{\sqrt{d} \sum_{i=1}^{K} \exp\left(q_{pt} k_i\right)} \right) v_i \right) + h_i \right) \tag{5}$$

where $Norm(\cdot)$ denotes the layer normalization operation, d represents the feature dimension, and f_i signifies the aggregated features. W_b is a trainable weight matrix of dimension $d \times d$, which implements the MLP-based linear transformation. The knowledge-guided graph representation ultimately yields the instance feature sequence $F = \{f_1, f_2, \ldots f_L\}$.

2.3 Cross-scale Knowledge Interaction Mamba

The proposed Cross-scale Knowledge Interaction Mamba is shown in Fig. 2(a), the low-scale feature F_{LR} and high-scale feature F_{HR} are processed by linear layers and then fed into the cross-scale selective scanning module. This module's output is subsequently processed through layer normalization and linear layers, with residual connections further enhancing its ability to model long-range spatial dependencies. The proposed cross-scale scanning module is shown in Fig. 2(c). Based on the selective scanning mechanism discussed in Section 2.1, we use linear mapping layers to generate system matrices $B_{\mathrm{LR}} \in \mathbb{R}^{L_{lr} \times N}$, $C_{\mathrm{LR}} \in \mathbb{R}^{L_{lr} \times N}$, and $\mathit{\Delta}_{\mathrm{LR}} \in \mathbb{R}^{L_{lr} \times D}$ for processing low-scale feature F_{LR}, as well as system matrices $B_{\mathrm{HR}} \in \mathbb{R}^{L_{hr} \times N}$, $C_{\mathrm{HR}} \in \mathbb{R}^{L_{hr} \times N}$, and $\Delta_{\mathrm{HR}} \in \mathbb{R}^{L_{hr} \times D}$ for processing high-scale feature F_{HR}, thereby enabling the the module parameters to be context-aware of the input. According to Equation (3), the system matrix C is used to decode information from the hidden state h_t to obtain the output y_t. Inspired by cross-attention mechanisms, we design the cross-scale selective scanning module to facilitate information exchange, which is represented as:

$$\bar{A}_{\mathrm{HR}} = \exp\left(\Delta_{\mathrm{HR}} A_{\mathrm{HR}}\right), \bar{A}_{\mathrm{LR}} = \exp\left(\Delta_{\mathrm{LR}} A_{\mathrm{LR}}\right) \tag{6}$$

$$\bar{B}_{\mathrm{HR}} = \Delta_{\mathrm{HR}} B_{\mathrm{HR}}, \bar{B}_{\mathrm{LR}} = \Delta_{\mathrm{LR}} B_{\mathrm{LR}} \tag{7}$$

$$h_{\mathrm{HR}}^{t} = \bar{A}_{\mathrm{HR}} h_{\mathrm{HR}}^{t-1} + \bar{B}_{\mathrm{HR}} x_{\mathrm{HR}}^{t}, h_{\mathrm{LR}}^{t} = \bar{A}_{\mathrm{LR}} h_{\mathrm{LR}}^{t-1} + \bar{B}_{\mathrm{LR}} x_{\mathrm{LR}}^{t} \tag{8}$$

$$C_{\mathrm{HR}}^{\mathrm{IE}} = C_{\mathrm{HR}} + Norm\left(C_{\mathrm{HR}} C_{\mathrm{LR}}^{\mathrm{T}} C_{\mathrm{LR}}\right), C_{\mathrm{LR}}^{\mathrm{IE}} = C_{\mathrm{LR}} + Norm\left(C_{\mathrm{LR}} C_{\mathrm{HR}}^{\mathrm{T}} C_{\mathrm{HR}}\right) \tag{9}$$

$$y_{\mathrm{HR}}^{t} = C_{\mathrm{HR}}^{\mathrm{IE}} h_{\mathrm{HR}}^{t}, y_{\mathrm{LR}}^{t} = C_{\mathrm{LR}}^{\mathrm{IE}} h_{\mathrm{LR}}^{t} \tag{10}$$

where x_{HR}^{t} and x_{LR}^{t} represent the inputs at time step t, while y_{HR}^{t} and y_{LR}^{t} denote the outputs of the module. $C_{\mathrm{HR}}^{\mathrm{IE}}$ and $C_{\mathrm{LR}}^{\mathrm{IE}}$ are cross-scale matrices, and their generation process leverages the interaction between C_{HR} and C_{LR}, thereby capturing information from different scales and ultimately recovering the output at each time step from the hidden states.

2.4 Training Strategy

After obtaining the prototypes from KMG-Mamba, we use a linear classifier to convert them into prototype logits and apply a smoothed support vector machine loss function with the bag labels to obtain $L_{Prototype}^{HR}$ and $L_{Prototype}^{LR}$. Meanwhile, we apply the cross-entropy loss function between the predicted bag logits and the labels to obtain L_{Bag}. The total loss function L for training is given as:

$$L = 0.5\left(L_{Prototype}^{HR} + L_{Prototype}^{LR}\right) + 0.5 L_{Bag} \tag{11}$$

3 Experiments

3.1 Datasets and Implementation Details

We evaluated the proposed KMG-Mamba on three public datasets: (1) CAMELYON16 [1], a dataset for detecting lymph node metastasis in breast cancer patients, which consists of 399 WSIs, with 159 slides containing lymph node metastasis and 240 slides without metastases; (2) TCGA-NSCLC [24], which comprises 1040 WSIs for two lung cancer subtypes: adenocarcinoma (LUAD, 529 slides) and squamous cell carcinoma (LUSC, 511 slides); (3) TCGA-BRCA [24], containing 977 WSIs for two breast cancer subtypes: invasive ductal carcinoma (IDC, 779 slides) and invasive lobular carcinoma (ILC, 198 slides). We extracted features from non-overlapping patches (256×256 pixels) at different magnifications (e.g., ×10, ×20), using ResNet-50 [10] pre-trained on ImageNet [5] and PLIP [13] pre-trained on OpenPath. We compared the proposed method with the current state-of-the-art (SOTA) approaches, including: (1) graph-based methods (Patch-GCN [4], WiKG [16]); (2) state-space model-based methods (S4MIL [6], MambaMIL [27]); (3) multi-scale methods (DSMIL [15], MG-Trans [21]); and (4) Attention-based methods (ABMIL [14], CLAM-MB [17], TransMIL [20], and RRT-MIL [23]). To ensure robust evaluation, we employed 10-fold Monte Carlo cross-validation and split datasets into training, validation, and test sets at an 8:1:1 ratio. A learning rate of 2×10^{-4} was used, and performance was assessed using the area under the curve (AUC), accuracy (ACC), and their standard deviations.

3.2 Comparison Results

Table 1 presents the experimental results on three datasets. Compared to the current SOTA methods, the proposed KMG-Mamba outperforms existing approaches in both feature settings. Additionally, we compared the computational efficiency of our method with several existing advanced methods. As shown in Fig. 3(a), we visualized the relationship between GPU memory allocation (GB) and performance (AUC) on the TCGA-NSCLC dataset. The results demonstrate that our model achieves excellent predictive performance while reducing computational costs.

Table 1. Comparisons on TCGA-BRCA, TCGA-NSCLC and CAMELYON16.

Dataset		TCGA-BRCA		TCGA-NSCLC		CAMELYON16	
Metric		AUC	ACC	AUC	ACC	AUC	ACC
ResNet-50	ABMIL [14]	0.881±0.031	0.873±0.027	0.939±0.028	0.837±0.050	0.855±0.048	0.805±0.048
	CLAM [17]	0.896±0.036	0.871±0.027	0.946±0.019	0.873±0.030	0.855±0.032	0.792±0.046
	TransMIL [20]	0.875±0.036	0.851±0.017	0.946±0.023	0.871±0.030	0.839±0.054	0.783±0.073
	RRT-MIL [23]	0.900±0.034	0.862±0.029	0.949±0.020	0.865±0.040	0.795±0.097	0.738±0.100
	Patch-GCN [4]	0.896±0.040	0.871±0.037	0.955±0.016	0.874±0.032	0.838±0.036	0.773±0.085
	WiKG [16]	0.895±0.043	0.849±0.048	0.942±0.030	0.847±0.044	0.843±0.069	0.805±0.040
	S4MIL [6]	0.892±0.032	0.862±0.028	0.939±0.022	0.864±0.038	0.875±0.036	0.803±0.036
	MambaMIL [27]	0.897±0.037	0.861±0.039	0.952±0.012	0.849±0.030	0.802±0.082	0.713±0.161
	DSMIL [15]	0.877±0.038	0.836±0.035	0.932±0.029	0.860±0.044	0.848±0.045	0.810±0.037
	MG-Trans [21]	0.889±0.036	0.871±0.031	0.954±0.013	0.873±0.024	0.815±0.054	0.760±0.058
	KMG-Mamba	**0.904±0.036**	**0.874±0.034**	**0.959±0.017**	**0.878±0.032**	**0.878±0.033**	**0.813±0.041**
PLIP	ABMIL [14]	0.900±0.040	0.877±0.037	0.964±0.012	0.890±0.035	0.900±0.051	0.840±0.070
	CLAM [17]	0.885±0.035	0.870±0.032	0.965±0.014	0.894±0.029	0.911±0.051	0.863±0.036
	TransMIL [20]	0.884±0.052	0.882±0.030	0.957±0.015	0.867±0.036	0.899±0.032	0.820±0.082
	RRT-MIL [23]	0.887±0.034	0.868±0.038	0.962±0.017	0.890±0.036	0.913±0.053	0.847±0.060
	Patch-GCN [4]	0.894±0.034	0.862±0.028	0.964±0.015	0.889±0.028	0.908±0.040	0.850±0.046
	WiKG [16]	0.887±0.034	0.874±0.026	0.958±0.018	0.876±0.036	0.887±0.046	0.863±0.032
	S4MIL [6]	0.899±0.036	0.865±0.021	0.959±0.017	0.886±0.031	0.912±0.041	0.863±0.056
	MambaMIL [27]	0.891±0.045	0.885±0.037	0.963±0.011	0.888±0.034	0.885±0.043	0.813±0.042
	DSMIL [15]	0.903±0.047	0.868±0.020	0.956±0.018	0.892±0.026	0.909±0.047	0.847±0.038
	MG-Trans [21]	0.898±0.048	0.891±0.021	0.962±0.012	0.881±0.028	0.877±0.037	0.813±0.068
	KMG-Mamba	**0.903±0.046**	**0.893±0.036**	**0.967±0.013**	**0.899±0.030**	**0.919±0.038**	**0.878±0.028**

3.3 Ablation Study

We conducted ablation experiments to evaluate the contribution of each module in our framework: (1) w/o KGR: remove the knowledge-guided graph represen-

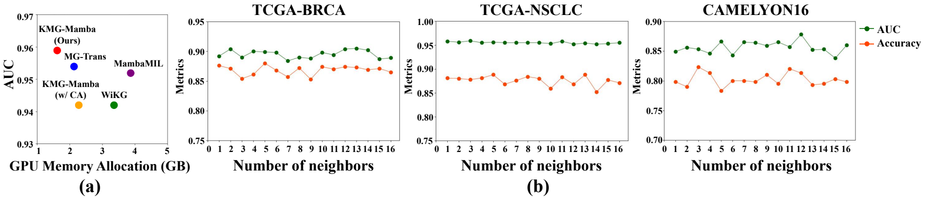

Fig. 3. (a) Computational analysis of our method and some selected SOTA methods; (b) Results of AUC and ACC with different numbers of neighbor nodes on three datasets.

Table 2. Ablation analysis on three datasets.

Method	TCGA-BRCA		TCGA-NSCLC		CAMELYON16	
	AUC	ACC	AUC	ACC	AUC	ACC
w/o KGR	0.900±0.042	0.866±0.032	0.951±0.022	0.864±0.027	0.839±0.048	0.795±0.038
w/o PL	0.893±0.036	0.862±0.032	0.954±0.019	0.877±0.037	0.821±0.056	0.780±0.080
w/o CKIM	0.884±0.054	0.870±0.032	0.949±0.021	0.870±0.035	0.844±0.052	0.807±0.039
w/ CA	0.893±0.032	0.860±0.023	0.942±0.025	0.848±0.038	0.830±0.051	0.800±0.030
Ours	**0.904±0.036**	**0.874±0.034**	**0.959±0.017**	**0.878±0.032**	**0.878±0.033**	**0.813±0.041**

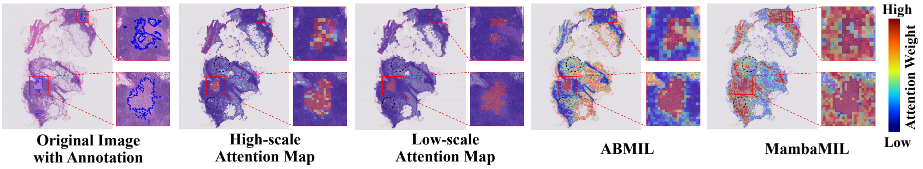

Fig. 4. Attention map visualization on CAMELYON16.

tation (KGR); (2) w/o PL: remove the loss functions $L_{Prototype}^{HR}$ and $L_{Prototype}^{LR}$ supporting KGR; (3) w/o CKIM: remove the cross-scale knowledge interaction Mamba, and F_{LR} and F_{HR} are directly fed into the aggregator and classifier; (4) w/ CA: replace the cross-scale scanning module with a cross-attention-based Transformer. As shown in Table 2, our method surpasses the performance of variant models across three datasets, confirming the importance of each module. Fig. 3(a) shows that the cross-scale scanning module uses less GPU memory than the cross-attention module, improving both computational efficiency and performance. Additionally, we investigated the effect of the number of neighbor nodes during graph construction in KGR as shown in Fig. 3(b), and found the number of neighbor nodes has minimal impact. Finally, we visualized the attention maps from the aggregator of KMG-Mamba on CAMELYON16, as shown in Fig. 4, and the attention maps of our method on the slides of both magnifications demonstrate superior localization by focusing on cancerous regions while minimizing normal tissue highlights. Although the attention maps from ABMIL and MambaMIL also focus on the positive regions, they highlight a large amount of normal tissue, confirming the effectiveness of KMG-Mamba in identifying discriminative regions within WSIs.

4 Conclusion

In this paper, we propose a novel knowledge-guided multi-scale graph Mamba for WSI classification, incorporating class knowledge through prototypes. Additionally, we propose the cross-scale knowledge interaction Mamba efficiently utilizes slide information across different magnifications. Extensive experiments on three public datasets demonstrate its effectiveness in WSI classification.

Acknowledgments. This work was supported by National Natural Science Foundation of China under Grant 82072021.

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

References

1. Bejnordi, B.E., Veta, M., Van Diest, P.J., Van Ginneken, B., Karssemeijer, N., Litjens, G., Van Der Laak, J.A., Hermsen, M., Manson, Q.F., Balkenhol, M., et al.: Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. Jama **318**(22), 2199–2210 (2017)
2. Bera, K., Schalper, K.A., Rimm, D.L., Velcheti, V., Madabhushi, A.: Artificial intelligence in digital pathology—new tools for diagnosis and precision oncology. Nature reviews Clinical oncology **16**(11), 703–715 (2019)
3. Brussee, S., Buzzanca, G., Schrader, A.M., Kers, J.: Graph neural networks in histopathology: Emerging trends and future directions. Medical Image Analysis p. 103444 (2025)
4. Chen, R.J., Lu, M.Y., Shaban, M., Chen, C., Chen, T.Y., Williamson, D.F., Mahmood, F.: Whole slide images are 2d point clouds: Context-aware survival prediction using patch-based graph convolutional networks. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part VIII 24. pp. 339–349. Springer (2021)
5. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: A large-scale hierarchical image database. In: 2009 IEEE conference on computer vision and pattern recognition. pp. 248–255. Ieee (2009)
6. Fillioux, L., Boyd, J., Vakalopoulou, M., Cournède, P.H., Christodoulidis, S.: Structured state space models for multiple instance learning in digital pathology. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 594–604. Springer (2023)
7. Gu, A., Dao, T.: Mamba: Linear-time sequence modeling with selective state spaces. arXiv preprint arXiv:2312.00752 (2023)
8. Guo, Z., Zhao, W., Wang, S., Yu, L.: Higt: Hierarchical interaction graph-transformer for whole slide image analysis. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 755–764. Springer (2023)

9. Hashimoto, N., Fukushima, D., Koga, R., Takagi, Y., Ko, K., Kohno, K., Nakaguro, M., Nakamura, S., Hontani, H., Takeuchi, I.: Multi-scale domain-adversarial multiple-instance cnn for cancer subtype classification with unannotated histopathological images. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 3852–3861 (2020)
10. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016)
11. Hou, W., Huang, H., Peng, Q., Yu, R., Yu, L., Wang, L.: Spatial-hierarchical graph neural network with dynamic structure learning for histological image classification. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 181–191. Springer (2022)
12. Huang, S.C., Chen, C.C., Lan, J., Hsieh, T.Y., Chuang, H.C., Chien, M.Y., Ou, T.S., Chen, K.H., Wu, R.C., Liu, Y.J., et al.: Deep neural network trained on gigapixel images improves lymph node metastasis detection in clinical settings. Nature communications **13**(1), 3347 (2022)
13. Huang, Z., Bianchi, F., Yuksekgonul, M., Montine, T.J., Zou, J.: A visual–language foundation model for pathology image analysis using medical twitter. Nature medicine **29**(9), 2307–2316 (2023)
14. Ilse, M., Tomczak, J., Welling, M.: Attention-based deep multiple instance learning. In: International conference on machine learning. pp. 2127–2136. PMLR (2018)
15. Li, B., Li, Y., Eliceiri, K.W.: Dual-stream multiple instance learning network for whole slide image classification with self-supervised contrastive learning. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 14318–14328 (2021)
16. Li, J., Chen, Y., Chu, H., Sun, Q., Guan, T., Han, A., He, Y.: Dynamic graph representation with knowledge-aware attention for histopathology whole slide image analysis. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11323–11332 (2024)
17. Lu, M.Y., Williamson, D.F., Chen, T.Y., Chen, R.J., Barbieri, M., Mahmood, F.: Data-efficient and weakly supervised computational pathology on whole-slide images. Nature biomedical engineering **5**(6), 555–570 (2021)
18. Qu, L., Liu, S., Liu, X., Wang, M., Song, Z.: Towards label-efficient automatic diagnosis and analysis: a comprehensive survey of advanced deep learning-based weakly-supervised, semi-supervised and self-supervised techniques in histopathological image analysis. Physics in Medicine & Biology **67**(20), 20TR01 (2022)
19. Qu, L., Yang, Z., Duan, M., Ma, Y., Wang, S., Wang, M., Song, Z.: Boosting whole slide image classification from the perspectives of distribution, correlation and magnification. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 21463–21473 (2023)
20. Shao, Z., Bian, H., Chen, Y., Wang, Y., Zhang, J., Ji, X., et al.: Transmil: Transformer based correlated multiple instance learning for whole slide image classification. Advances in neural information processing systems **34**, 2136–2147 (2021)
21. Shi, J., Tang, L., Gao, Z., Li, Y., Wang, C., Gong, T., Li, C., Fu, H.: Mg-trans: Multi-scale graph transformer with information bottleneck for whole slide image classification. IEEE Transactions on Medical Imaging (2023)
22. Srinidhi, C.L., Ciga, O., Martel, A.L.: Deep neural network models for computational histopathology: A survey. Medical image analysis **67**, 101813 (2021)
23. Tang, W., Zhou, F., Huang, S., Zhu, X., Zhang, Y., Liu, B.: Feature re-embedding: Towards foundation model-level performance in computational pathology. In: Pro-

ceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11343–11352 (2024)

24. Weinstein, J.N., Collisson, E.A., Mills, G.B., Shaw, K.R., Ozenberger, B.A., Ellrott, K., Shmulevich, I., Sander, C., Stuart, J.M.: The cancer genome atlas pan-cancer analysis project. Nature genetics **45**(10), 1113–1120 (2013)
25. Xu, H., Xu, Q., Cong, F., Kang, J., Han, C., Liu, Z., Madabhushi, A., Lu, C.: Vision transformers for computational histopathology. IEEE Reviews in Biomedical Engineering **17**, 63–79 (2023)
26. Xu, R., Yang, S., Wang, Y., Du, B., Chen, H.: A survey on vision mamba: Models, applications and challenges. arXiv e-prints pp. arXiv–2404 (2024)
27. Yang, S., Wang, Y., Chen, H.: Mambamil: Enhancing long sequence modeling with sequence reordering in computational pathology. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 296–306. Springer (2024)
28. Zhu, L., Liao, B., Zhang, Q., Wang, X., Liu, W., Wang, X.: Vision mamba: Efficient visual representation learning with bidirectional state space model. In: Forty-first International Conference on Machine Learning