E-mail: jig@aircas.ac.cn
Website: www.cjig.cn
Tel: 010-58887035
中国图象图形学报
JOURNAL OF IMAGE AND GRAPHICS
©中国图象图形学报版权所有

中图法分类号:TP391;TP18　　**文献标识码**: A　　**文章编号**: 1006-8961(2026)10-3802-17

论文引用格式:Teng X F, Liu H, Mao N, Gao J and Chen G G. 2026. Multiscale histopathological image classification with graph edge-attribute attention. Journal of Image and Graphics, 31(10):3802-3818(滕晓菲, 刘慧, 毛宁, 高婧, 陈公冠. 2026. 图边属性注意力的多尺度组织病理图像分类. 中国图象图形学报, 31(10):3802-3818)[DOI:10.11834/jig.260009]

图边属性注意力的多尺度组织病理图像分类

滕晓菲[1,2],刘慧[1,2*],毛宁[3],高婧[3],陈公冠[1,2]

1. 山东财经大学计算机与人工智能学院,济南　250014; 2. 山东省数字经济轻量智算与可视化重点实验室(山东财经大学),济南　250014; 3. 青岛大学附属医院烟台毓璜顶医院,烟台　264099

摘　要: **目的**　微卫星不稳定性(microsatellite instability,MSI)是癌症预后与治疗决策的重要指标,而组织病理图像作为诊断的重要依据,其数字化形式——全玻片图像(whole slide image,WSI)已广泛应用于病理分析。受限于WSI的尺寸,现有方法常采用多实例学习(multiple instance learning,MIL)框架,将WSI划分为图像块后通过平均池化等方式获得整体预测结果,但其忽略了块间关联性,限制了模型的建模能力与检测性能。此外,单尺度建模难以同时兼顾组织的宏观结构与局部细节,而简单的多尺度融合策略缺乏对不同尺度语义差异的刻画,容易引入尺度偏置。**方法**　为解决上述问题,提出了一种图边属性注意力的多尺度融合网络(graph edge-attribute attentional multi-scale fusion network, GEAMF)。该模型首先使用图构建模块(graph construct module,GC-M)将WSI图像块构建为图结构,然后基于Transformer的注意力机制设计了图边属性注意力模块(graph edge-attribute attention module,GEA-M)来显式编码图节点间的关系,以实现节点特征的动态加权聚合。此外,进一步设计了一种双流分支来学习不同尺度下的WSI特征,最终通过可学习的尺度感知门控残差模块(scale-aware gated ResMLP module,SGRes-M)实现多尺度特征的互补优化。**结果**　实验表明,GEAMF在多种公开WSI数据集上均显著优于对比方法,验证了其在MSI分类任务中对多尺度组织表征和结构依赖关系建模的有效性。此外,在Camelyon16淋巴结转移检测这一不同病理任务上,所提方法同样取得了稳定性能提升,验证了模型在跨病理任务场景下的泛化能力。**结论**　本文提出的GEAMF模型通过显式建模图像块之间的结构依赖关系,并联合多尺度组织表征,有效弥补了传统弱监督方法忽视结构与尺度信息的不足。

关键词: 全玻片图像(WSI);多实例学习(MIL);图神经网络(GNN);多尺度特征融合;残差多层感知机(RasMLP)

Multiscale histopathological image classification with graph edge-attribute attention

Teng Xiaofei[1,2], Liu Hui[1,2*], Mao Ning[3], Gao Jing[3], Chen Gongguan[1,2]

1. School of Computing and Artificial Intelligence, Shandong University of Finance and Economics, Jinan 250014, China; 2. Shandong Key Laboratory of Lightweight Intelligent Computing and Visualization for Digital Economy (Shandong University of Finance and Economics), Jinan 250014, China; 3. Yantai Yuhuangding Hospital, Qingdao University Affiliated Hospital, Yantai 264099, China

Abstract: **Objective** Microsatellite instability (MSI) is a key biomarker for cancer prognosis and treatment decisions, par-

收稿日期:2026-01-12;**修回日期**:2026-02-13;**预印本日期**:2026-02-20

*通信作者:刘慧　liuh_lh@sdufe.edu.cn

基金项目:山东省泰山学者特聘专家计划项目(tstp20221137);山东省重点研发计划资助(2025CXPT100);济南市人才发展专项资金项目(202333037)

Supported by: Taishan Scholars Program of Shandong Province(tstp20221137); Key R&D Program of Shandong Province, China(2025CXPT100); Jinan Municipal Talent Development Special Fund(202333037)

ticularly in colorectal and gastric cancers. With the widespread application of digital pathology, whole slide image (WSI) have become an important data source for MSI prediction. However, automatically predicting MSI from WSIs remains challenging due to their gigapixel scale, insufficient supervised information at the slice level, and considerable heterogeneity in histopathological tissue structures. Most existing methods rely on multiple instance learning strategies, processing WSIs by segmenting them into isolated image patches and aggregating block-level predictions using pooling strategies. While this approach is effective to some extent, it inherently ignores the spatial and semantic dependencies between tissue regions, leading to poor representation learning performance. Furthermore, pathological diagnosis is inherently a multiscale process, requiring joint reasoning about global tissue structure and local cell morphology. Many current methods either operate at only a single magnification or employ coarse and nonadaptive multiscale fusion strategies, limiting their ability to capture complementary information across scales. Therefore, this work aims to develop a weakly supervised WSI modeling framework that explicitly captures structural relationships between patches while effectively integrating complementary information across multiple magnification scales to achieve reliable MSI classification. **Method** To achieve this goal, this research proposes a graph edge-attribute attentional multiscale fusion network (GEAMF). In GEAMF, each WSI is represented as a graph, where image patches correspond to nodes, and interpatch relationships are modeled as edges. First, this study designs a graph construct module(GC-M) that combines constraints on semantic similarity and spatial proximity to establish reliable structural connections between image patches. Specifically, cosine similarity can effectively measure semantic alignment in a high-dimensional feature space, but it can introduce nonphysical long-range edges in heterogeneous WSI. These connections propagate irrelevant contextual information to distant tissue regions. To address this limitation, we introduce Euclidean distance as a spatial constraint to reweight semantic similarity, thereby achieving locally perceptive edge modeling. This design explicitly balances semantic consistency and spatial coherence, achieving graph modeling that is locally perceptive and structurally meaningful. On the basis of this graph representation, this study proposes a graph edge-attribute attention module (GEA-M) to perform structure-aware message passing. In contrast to traditional graph attention mechanisms that only utilize node features, GEA-M explicitly incorporates edge attributes into the attention computation. This design allows the model to adjust the information propagation intensity dynamically on the basis of the discriminative importance of different structural relationships, suppressing noisy or irrelevant connections while emphasizing diagnostically remarkable interactions between tissue regions. To address the inherent multiscale nature of pathological diagnosis further, GEAMF employs a dual architecture, independently processing WSI at two magnification levels. Each branch learns a graph-level representation for its specific scale, capturing either global organizational structure or fine-grained cell morphology. To integrate these heterogeneous representations adaptively, this work further designs a scale-aware gated ResMLP module (SGRes-M). This module introduces learnable scale-aware gating, dynamically reweighting the magnification contribution of each sample, and then uses a residual multilayer perceptron to model higher-order interactions between scales while preserving scale-specific discriminative cues. **Result** This study systematically evaluates the proposed GEAMF model on several publicly available WSI benchmark datasets. The TCGA_CRC and TCGA_STAD datasets were used to evaluate the model's performance on the MSI classification task, while the Camelyon16 dataset was used to test the model's generalization ability in different pathological task scenarios. On the TCGA_CRC dataset, GEAMF achieved Acc, AUC, and F1 of 0.945, 0.835, and 0.887. On the TCGA_STAD dataset, the model achieved Acc, AUC, and F1 of 0.924, 0.829, and 0.897, respectively. Furthermore, GEAMF also achieved good performance on the Camelyon16 lymph node metastasis detection dataset, with Acc, AUC and F1 of 0.907, 0.877 and 0.848, validating broad generalization on different tasks. Compared with various existing WSI classification methods, the proposed model demonstrates more stable and superior overall performance across different datasets and task settings. To validate the rationality of each module design further, this study systematically analyzes the impact of the proposed modules on model performance through ablation experiments. Furthermore, this study improves the model's interpretability from two levels: t-SNE and attention heatmaps. **Conclusion** This study proposes a novel weakly supervised framework, GEAMF for WSI classification, which explicitly integrates graph-based structural modeling and adaptive multiscale fusion. By introducing an edge-attribute-aware attention mechanism and scale-aware gated residual fusion, GEAMF effectively captures intrascale structural dependencies and cross-scale complementary information. Experimental results on the MSI benchmark datasets demonstrate that

explicitly modeling the structural relationships between image patches and combining them with multiscale representations remarkably improves classification performance, validating the effectiveness of introducing structure-aware and scale-aware inductive bias into WSI analysis. Furthermore, GEAMF exhibits stable performance in the Camelyon16 lymph node metastasis detection task, further illustrating the framework's generalization ability across different pathological tasks. This framework provides a general and scalable paradigm for graph-based multiscale modeling in computational pathology and has potential applications in various histopathological prediction tasks.

Key words: whole slide image (WSI); multiple instance learning (MIL); graph neural network (GNN); multi-scale feature fusion; residual multilayer perceptron (ResMLP)

0 引 言

在临床诊疗中,微卫星不稳定性(microsatellite instability,MSI)被列为多个癌症的重要生物标志物(Willis等,2019)。传统的MSI检测主要依赖聚合酶链式反应和免疫组化等实验方法,这些方法在实际应用中存在组织样本消耗大、检测周期长以及人工判读偏差等问题(Saillard等,2023)。为了减轻病理学家的工作负担并提升诊断准确率,组织病理图像逐步以全玻片图像(whole slide image,WSI)的形式被数字化,为MSI检测方法向数字化、自动化方向的发展奠定了基础。

在此背景下,研究者开始引入深度学习方法(冯龙锋 等,2022)对WSI进行自动化分析。尽管卷积神经网络(convolutional neural network,CNN)在自然图像领域取得了显著成果(Xie等,2020;Gregor等,2015;Bayramoglu等,2016),但受限于WSI的超高分辨率特性,其难以直接对整幅图像进行端到端建模。早期研究通常采用图像分块的策略(Wang等,2021),从WSI中提取局部图像块作为训练单元,学习图像块与标签之间的对应关系,如图1(b)所示。然而该方法依赖于块级标注,在实际病理场景中,精细标注代价高昂且难以获得。因此,研究者进一步引入多实例学习(multiple instance learning,MIL)框架,以实现弱监督下的有效建模。

在基于MIL(Hou等,2016;Chen等,2022;Alon和Zhou,2022;Zhao等,2020)的方法中,研究者仅需依赖WSI级标签,通过对图像块特征进行聚合实现整体预测,在一定程度上缓解了精细标注缺失的问题,如图1(c)所示。近年来,部分研究通过引入视觉解释机制(He等,2022)和特征蒸馏(Zhang等,2022)等进一步提升了特征表达能力和全局建模能力。然而,现有MIL方法通常假设图像块之间相互独立,未能显式建模实例间潜在的空间结构关系与语义依赖。在高度异质的病理组织中,这种独立性假设容易使大量与诊断无关的背景区域被错误地赋予相同的图像级监督信号,从而引入显著噪声并削弱模型对关键病灶区域的判别能力(Lu等,2021;Chan等,2023)。因此,如何在弱监督条件下显式建模图像块之间的结构关系,成为提升MSI预测性能的关键问题之一。

为弥补MIL中结构信息缺失的问题,部分研究(Lu等,2022;Bontempo等,2024;沈熠婷 等,2024;马帅 等,2022)进一步引入了图神经网络(graph neural network,GNN),如图1(d)所示。Shen等人(2023)通过结合CNN与GNN实现局部空间信息与全局语义特征的联合建模。随后,Chan等人(2023)开始关注图结构内部的异质性建模,设计异构图边属性变换器,提升结构表达能力。进一步地,Shi等人(2024)基于超图提出MaskHGL框架,扩展了传统图的表示范围,从而增强模型在复杂组织结构下的鲁棒性。然而,现有GNN方法通常采用统一的消息传递策略,对不同邻接关系进行无差别聚合,隐含假设所有邻居在语义上同等重要(Ozen等,2021;Zheng等,2022)。这种"一视同仁"的信息融合方式在组织病理图像中尤为受限:相邻但组织类型不同的图像块可能引入噪声干扰,而空间上并不紧邻但语义高度一致的区域却可能承载关键的病灶信息。缺乏对结构关系判别性的建模,使得GNN难以在复杂组织异质性背景下有效聚焦于具有诊断价值的区域关联。因此,WSI图建模的关键不在于简单引入图结构,而在于如何在消息传递过程中显式区分不同结构关系的重要性,使模型能够自适应地选择"重要关系"并抑制冗余或噪声关联。这一观察为在GNN中引入结构感知的注意力机制提供了新的建模视角。

此外，在实际应用中，WSI的分析不仅需要识别肿瘤的宏观特征，还需要关注其微观特征。低倍率图像提供了肿瘤边界、组织分布等宏观结构，而高倍率图像揭示了细微的细胞形态与微环境特征，如何实现不同尺度的信息融合至关重要（丁维龙 等，2024；Ding等，2023；Xiang等，2023）。现有方法大多局限于单尺度建模（Lu 等，2022；Bontempo 等，2024），这导致必须在两者之间做出取舍：低倍率缺失细节、高倍率丢失上下文。局限于单一视角也会造成全局或局部信息的丢失，影响模型对病灶区域的精细识别（Bhojanapalli 等，2020）。此外，由于不同的WSI样本在诊断过程中对不同尺度信息的依赖程度不一致，固定或简单的多尺度融合策略容易引入尺度偏置，从而影响模型的稳定性与泛化能力。

针对上述问题，本文首先从单尺度图结构建模出发，通过图构建模块（graph construct module，GC-M），建立联合空间与语义约束的结构连接，以缓解单一构图方式带来的逻辑偏差。在此基础上，提出图边属性注意力模块（graph edge-attribute attention module，GEA-M），将边属性显式引入注意力计算过程中，对不同邻接关系进行结构判别性加权，以抑制组织异质性背景下由弱相关邻接引发的噪声传播，形成结构感知的单尺度图建模框架。进一步地，考虑到不同放大倍率间的病理信息互补性，本文在上述单尺度图建模框架的基础上，扩展构建了一个双分支的图边属性注意力的多尺度融合网络（graph edge-attribute attentional multi-scale fusion network，GEAMF），并设计尺度感知门控残差模块（scale-aware gated resmlp module，SGRes-M）自适应地平衡多尺度贡献，从而降低固定融合策略引入的尺度偏置，实现更稳定的跨尺度特征互补融合。

本文的主要贡献如下：1）从单尺度图建模出发，提出了一种联合约束语义一致性与空间邻近性的图构建方法，为弱监督病理图像分析提供更具表达力的结构表示。2）为缓解组织异质性背景下消息传递缺乏判别性的问题，设计了图边属性注意力模块（GEA-M），将边属性显式引入注意力计算，实现了节点间信息流动的动态引导，有效提升了节点特征的结构感知能力与判别性。3）为解决多尺度信息融合中的尺度偏置与样本特异性依赖问题，提出了图边属性注意力的多尺度融合网络GEAMF。该网络通过尺度感知门控残差机制实现了样本级的自适应互补建模，在保证各尺度判别性的同时，显著提升了多尺度表征的稳定性与泛化性能。

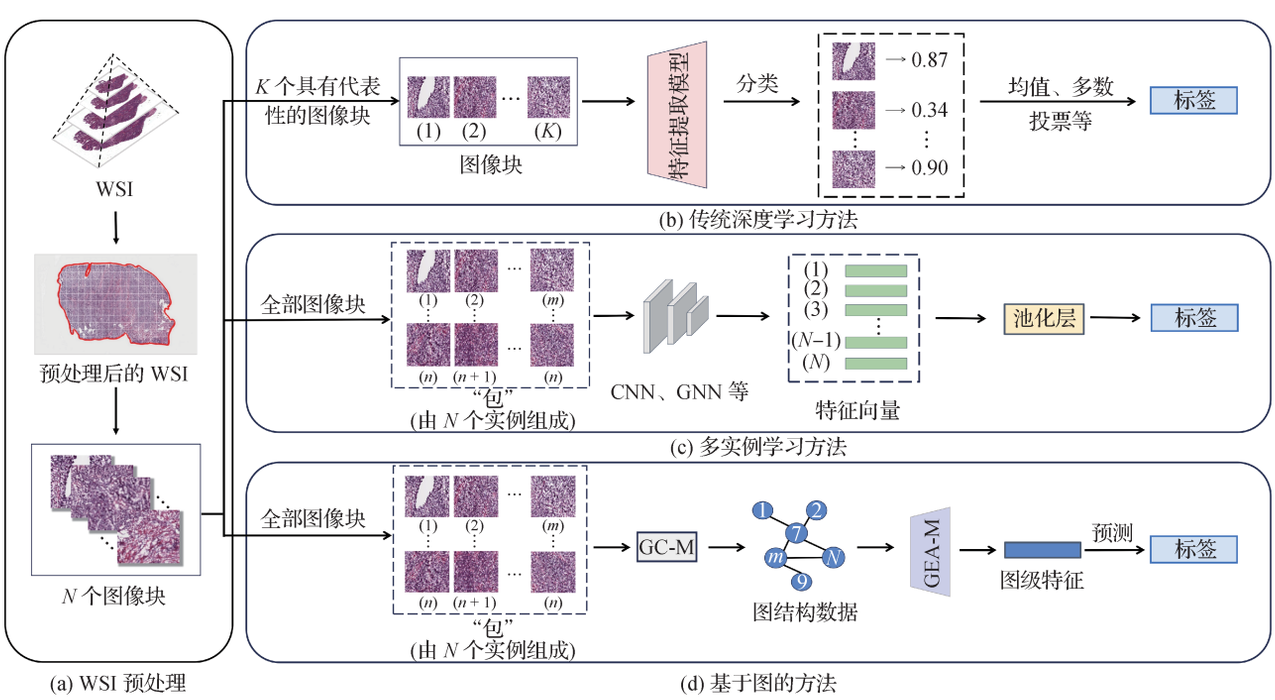


图 1　WSI分析的不同方法比较

Fig. 1　Comparison of different methods for WSI analysis ((a) WSI preprocessing; (b) traditional deep learning methods; (c) multiple instance learning methods; (d) graph-based modeling methods)

1 本文方法

本文提出了一种基于多实例学习框架的图边属性注意力的多尺度融合网络(GEAMF),用于实现对WSI的MSI检测。模型的整体架构如图2所示,图中$\boldsymbol{X}_H$与$\boldsymbol{X}_L$分别表示高低倍率节点特征矩阵,$\boldsymbol{G}_H$与$\boldsymbol{G}_L$分别表示高低倍率图结构。整体框架包括图构建模块(GC-M)、图边属性注意力模块(GEA-M)以及多尺度融合模块(SGRes-M)。本节对这些模块以及最终的模型设计进行详细说明:首先介绍数据预处理步骤,然后介绍GC-M和GEA-M,最后提出SGRes-M以融合双分支的多尺度信息。

1.1 数据预处理

WSI通常具有金字塔结构,每层代表不同的放大倍率,其存储格式有SVS(scanscope virtual slide)、NDPI(nanozoomer digital pathology image)等。本文选用20×和40×两种尺度作为多尺度输入,主要基于如下考虑:20×下的图像块具有较大的感受视野,有利于捕捉组织形态的上下文信息;40×下的图像块保留了更多细粒度的形态特征,有助于肿瘤区域的边界识别。两种尺度在语义与纹理层面互补,有助于提升模型的判别能力。本文借助OpenSlide库,读取WSI中40×下的图像,采用与Zhu等人(2025)相同的处理方式得到20×的WSI图像。

针对提取出来的不同放大倍率的图像,均采用Lerousseau等人(2021)提出的WSI预处理方法,将WSI图像划分为512×512像素的图像块。为减少无组织区域对后续分析的干扰,对图像块进行背景过滤,移除大部分为背景的图像块,只保留包含实际组织信息的图像块。

1.2 图构建模块GC-M

在弱监督WSI场景下,图结构的构建方式直接决定了后续消息传递的有效性。仅基于特征相似度构建图结构,容易在组织高度异质的病理图像中引入与空间结构不一致的"长程连接";而仅依赖空间邻近性又可能忽略跨区域但语义一致的关键病灶关系。为了缓解这种构图偏差,本文设计了GC-M,通过联合约束语义相似性与空间邻近性,建立更具病理可信度的结构连接。

以20×为例,将一幅WSI表示为图结构$\boldsymbol{G}=(\boldsymbol{V},\boldsymbol{E})$,其中,$\boldsymbol{V}$是节点的集合,$\boldsymbol{E}$是边的集合。每个节点$\boldsymbol{v}_i\in\boldsymbol{V}$对应一个图像块,其特征表示为$\boldsymbol{x}_i\in\mathbf{R}^D$,$D=512$表示特征维度,边$e_{ij}$关联一对节点$(\boldsymbol{v}_i,\boldsymbol{v}_j)$,

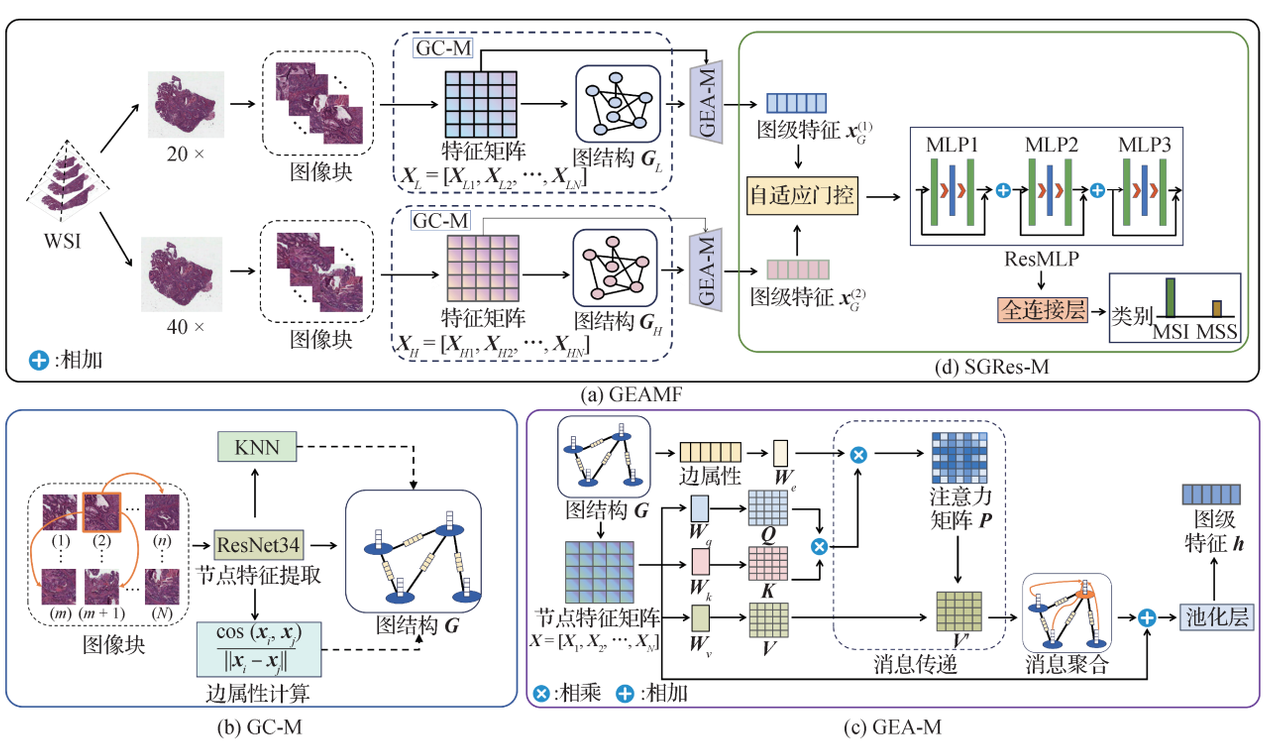


图2 本文方法架构

Fig. 2 Architecture of the proposed method ((a) GEAMF network; (b) GC-M module; (c) GEA-M module; (d) SGRes-M module)

并具有对应的属性$\boldsymbol{w}_{ij} \in \mathbf{R}^k$，其中，$k$表示边属性向量的维度。在特征提取阶段，本文采用在ImageNet上预训练的ResNet34（34-layer residual network）对每个图像块进行编码，提取其512维特征向量，构成节点特征矩阵$\boldsymbol{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_{n_p}]^{\mathrm{T}} \in \mathbf{R}^{n_p \times D}$，其中，$n_p$表示WSI被分成的图像块数量。

随后，利用图像块在WSI中的二维空间坐标，对所有节点执行K-近邻（K-nearest neighbor，KNN）搜索，为每个节点$\boldsymbol{v}_i$建立与其最近的K个邻居节点之间的边连接，得到邻接矩阵$\boldsymbol{A}$。在构建图结构时，如何合理刻画节点间的关联强度是影响后续消息传递效果的关键因素。已有研究表明（Zhang等，2022），余弦相似度因其对向量尺度不敏感、适于刻画高维表征的方向一致性，常用做基于特征的KNN图构建或边权初始化策略。基于这一动机，本文首先考虑利用节点特征之间的余弦相似度来刻画图像块间的语义关联程度，具体为

$$\cos(\boldsymbol{v}_i, \boldsymbol{v}_j) = \frac{\boldsymbol{x}_i \boldsymbol{x}_j}{\|\boldsymbol{x}_i\| \|\boldsymbol{x}_j\|} \tag{1}$$

式中，$\|\cdot\|$表示L2范数。由于病理组织具有高度异质性，仅基于特征相似度可能产生与组织空间结构不一致的“长程连接”，从而在消息传递中引入跨区域干扰（Alzoubi等，2024）。为缓解上述问题，本文进一步引入欧氏距离作为空间约束项（林忠辉 等，2002），对基于语义相似度构建的结构关联进行调制。具体而言，欧氏距离用于刻画节点在空间域中的几何邻近性，具体为

$$d(\boldsymbol{v}_i, \boldsymbol{v}_j) = \|\boldsymbol{x}_i - \boldsymbol{x}_j\| \tag{2}$$

当两个图像块在特征空间中高度相似但空间距离较远时，其结构关联应被适当抑制；仅当语义一致性与空间邻近性同时满足时，该关联才具有较高的病理可信度。基于这一考虑，本文采用“语义相似度—空间距离调制”的方式定义边属性，具体为

$$\boldsymbol{w}_{ij} = \begin{cases} \dfrac{\cos(\boldsymbol{v}_i, \boldsymbol{v}_j)}{d(\boldsymbol{v}_i, \boldsymbol{v}_j)} & d(\boldsymbol{v}_i, \boldsymbol{v}_j) \leqslant \theta \\ 0 & d(\boldsymbol{v}_i, \boldsymbol{v}_j) > \theta \end{cases} \tag{3}$$

式中，$\boldsymbol{v}_j \in \boldsymbol{N}_K(\boldsymbol{v}_i)$，$\boldsymbol{N}_K(\boldsymbol{v}_i)$表示节点$\boldsymbol{v}_i$的邻居节点集合，$\theta$为阈值，用于抑制结构关联度较低的边，以减少噪声连接对消息传递的干扰。最终生成的边属性矩阵$\boldsymbol{W} \in \mathbf{R}^{n_p \times n_p}$作为GNN模型中的结构先验信息。40×的图结构构建过程与20×相同。

1.3　图边属性注意力模块GEA-M

为缓解GNN在弱监督场景下邻接关系缺乏判别性且易受组织异质性噪声干扰的问题，本文设计了GEA-M。不同于仅基于节点特征相似性的传统机制，如图注意力网络（graph attention network，GAT），GEA-M在计算Query-Key相似性时显式引入了边属性$\boldsymbol{w}_{ij}$作为调制因子，对每一对节点间的注意力权重进行引导修正。该模块能够动态调节邻居节点间的信息传递强度，从而实现更精准的消息传递控制。其具体计算过程如下：

首先通过线性变换获得查询（Query）、键（Key）、值（Value）向量。具体为

$$\begin{aligned} \boldsymbol{Q}_i^h &= \boldsymbol{W}_Q^h \boldsymbol{x}_i \\ \boldsymbol{K}_i^h &= \boldsymbol{W}_K^h \boldsymbol{x}_i \\ \boldsymbol{V}_i^h &= \boldsymbol{W}_V^h \boldsymbol{x}_i \end{aligned} \tag{4}$$

式中，h表示第h个注意力头，$\boldsymbol{W}_Q^h$，$\boldsymbol{W}_K^h$，$\boldsymbol{W}_V^h$为第h个注意力头的可学习参数。由于在WSI中，相邻但纹理差异大的图像块可能位于不同组织区域，应降低信息传递强度；而空间上较远但纹理相似的图像块可能属于同一病灶区域，应强化其信息聚合。因此引入边属性调制注意力，得到边感知注意力权重，具体为

$$\boldsymbol{\alpha}_{ij}^h = f_{\mathrm{softmax}}\left(\frac{(\boldsymbol{Q}_i^h \boldsymbol{K}_j^{h\mathrm{T}}) \cdot \boldsymbol{w}_{ij}}{\sqrt{d_k}}\right) \tag{5}$$

式中，注意力权重$\boldsymbol{\alpha}_{ij}^h$控制从节点$j$到节点$i$的消息传递强度，$d_k$表示每个注意力头中Key向量的维度，最终消息向量为

$$\boldsymbol{m}_{ij}^h = \boldsymbol{\alpha}_{ij}^h \boldsymbol{V}_j^h \tag{6}$$

多头自注意力机制通过并行的多个注意力头分别建模不同的上下文依赖关系，以增强模型的表达能力。各个注意力头的输出经拼接后，通过线性映射生成最终多头注意力特征，具体为

$$\boldsymbol{m}_{ij}^{(l)} = f_{\mathrm{Concat}}(\boldsymbol{m}_{ij}^1, \boldsymbol{m}_{ij}^2, \cdots, \boldsymbol{m}_{ij}^H)\boldsymbol{W}_o \tag{7}$$

式中，H为注意力头数量，$\boldsymbol{W}_o$为可学习的线性投影矩阵。最终，节点v_i的特征通过聚合其邻居节点的多头消息向量进行更新，具体为

$$\boldsymbol{x}_i' = f_{\mathrm{Aggregate}}(\{\boldsymbol{m}_{ij}^{(l)} \mid j \in \boldsymbol{N}(i)\}) \tag{8}$$

式中，$\boldsymbol{N}(i)$表示节点$\boldsymbol{v}_i$的邻居节点的集合。为增强节点表示的深层表达能力，本文堆叠了多层GEA-M，节点表示的第$l+1$层通过如下方式迭代更新，具体为

$$\boldsymbol{x}^{(l+1)\prime} = f_{\mathrm{GEA\text{-}M}}(\boldsymbol{x}^{(l)\prime}, \boldsymbol{W}_{\mathrm{edge}}) \tag{9}$$

式中，$\boldsymbol{W}_{\text{edge}}$表示边属性矩阵，$x^{(l)\prime}$为第$l$层的节点表示。为缓解深层网络训练中的梯度消失问题，引入可学习的残差连接，具体为

$$\boldsymbol{x}^{(l+1)\prime}=\beta\cdot\boldsymbol{x}^{(l+1)\prime}+(1-\beta)\cdot\boldsymbol{x}^{(l)\prime} \tag{10}$$

式中，β为可学习的参数。最后通过均值—最大值双通道池化聚合图级特征，具体为

$$\boldsymbol{x}_G=f_{\text{Concat}}[GMP(\boldsymbol{x}'),\ GAP(\boldsymbol{x}')] \tag{11}$$

1.4 尺度感知门控残差模块SGRes-M

在多尺度建模中，低倍率分支与高倍率分支分别侧重于捕获组织的宏观结构与局部微观细节。然而，由于不同样本在诊断过程中对尺度信息的依赖程度存在显著差异，固定的特征融合策略极易引入尺度偏置，导致模型过度依赖单一尺度，进而影响融合表示的稳定性。为解决上述问题，本文设计了SGRes-M，如图3所示。该模块通过引入自适应门控机制，动态调节不同尺度表征的贡献权重，以实现多尺度信息的判别性对齐与稳定互补。

设经过GEA-M模块后得到的低倍率下的图级特征表示为$\boldsymbol{x}_G^{(1)}\in\mathbf{R}^d$，高倍率图级特征表示为$\boldsymbol{x}_G^{(2)}\in\mathbf{R}^d$。考虑到不同样本在多尺度信息上的依赖具有显著差异性，SGRes-M首先引入尺度感知门控机制，对每个尺度的图级特征进行自适应重标定。具体地，对第s个尺度$\left(s\in\{1,2\}\right)$，通过一个可学习的门控函数生成对应的尺度权重，具体为

$$\boldsymbol{\alpha}_s=\sigma(\boldsymbol{W}_g\boldsymbol{x}_G^{(s)}) \tag{12}$$

式中，$\boldsymbol{W}_g\in\mathbf{R}^{1\times d}$为可学习参数，$\sigma(\cdot)$表示sigmoid激活函数，$\boldsymbol{\alpha}_s\in(0,1)$表示当前样本在该尺度下的结构重要性权重。该门控机制使模型能够根据样本特征自适应地调整不同尺度图表征在融合过程中的贡献程度。随后，对各尺度图级特征进行门控调制，具体为

$$\tilde{\boldsymbol{x}}_G^{(s)}=\boldsymbol{\alpha}_s\cdot\boldsymbol{x}_G^{(s)} \tag{13}$$

在完成尺度感知重标定后，将不同尺度下的门控图级特征进行拼接，并输入至残差多层感知机（residual multi-layer perceptron，ResMLP）进行融合，具体为

$$z=f_{\text{ResMLP}}\left(f_{\text{Concat}}\left(\tilde{\boldsymbol{x}}_G^{(1)},\tilde{\boldsymbol{x}}_G^{(2)}\right)\right) \tag{14}$$

式中，f_{ResMLP}表示残差多层感知机，由多层非线性映射与残差连接组成，用于建模多尺度结构表征之间的高阶交互关系（Touvron等，2023）。残差结构的引入有助于在融合过程中保留各尺度原始的判别性结构信息，同时缓解深层非线性映射可能带来的信息退化问题。ResMLP每一层包含一个降维和一个升维映射，配合残差连接完成特征变换，具体为

$$\boldsymbol{x}_{\text{temp}}^{(l)}=g(\boldsymbol{W}_{\text{down}}^{(l)}\boldsymbol{x}^{(l-1)}+\boldsymbol{b}_{\text{down}}^{(l)}) \tag{15}$$

$$\boldsymbol{x}^{(l)}=g(\boldsymbol{W}_{\text{up}}^{(l)}\boldsymbol{x}_{\text{temp}}^{(l)}+\boldsymbol{b}_{\text{up}}^{(l)})+\boldsymbol{x}^{(l-1)} \tag{16}$$

式中，$l=1,2,3$表示堆叠层数，$\boldsymbol{W}_{\text{down}}^{(l)}\in\mathbf{R}^{d_{\text{hidden}}\times d_{\text{in}}}$和$\boldsymbol{W}_{\text{up}}^{(l)}\in\mathbf{R}^{d_{\text{in}}\times d_{\text{hidden}}}$分别表示第$l$层多层感知机（multi-layer perceptron，MLP）中降维和升维的可学习权重矩阵，$\boldsymbol{b}_{\text{down}}^{(l)}$与$\boldsymbol{b}_{\text{up}}^{(l)}$为对应的偏置项，$g(\cdot)$表示非线性激活函数（如ReLU（rectified linear unit）或GELU（Gaussian error linear unit））。在本文模型中，输入层特征维度为$d_{\text{in}}=2d$，中间隐藏层维度$d_{\text{hidden}}<d_{\text{in}}$。

经过3层ResMLP处理后，获得融合后的图级特征表示，具体为

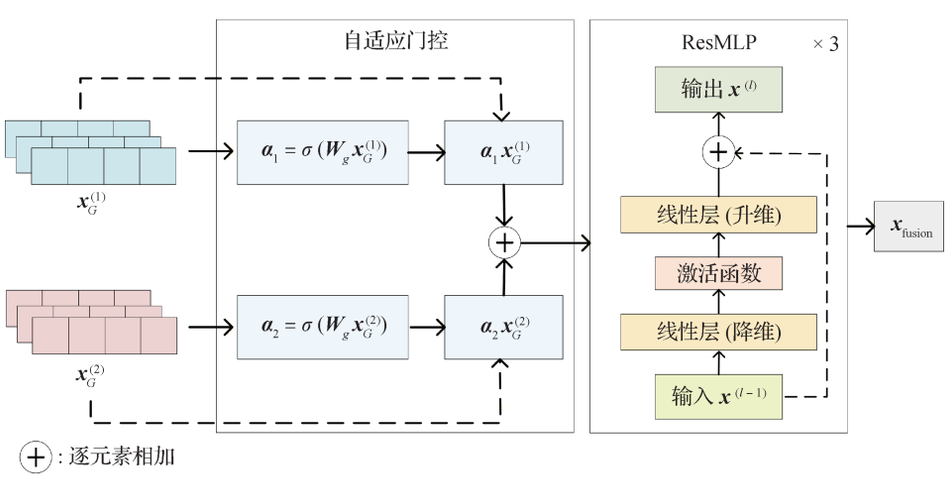


图3 SGRes-M框架示意图

Fig. 3 Overview of SGRes-M

$$\boldsymbol{x}_{\text{fusion}} = \boldsymbol{x}^{(3)} \tag{17}$$

最后，融合特征 $\boldsymbol{x}_{\text{fusion}}$ 经过全连接层映射到类别空间，实现微卫星不稳定性状态的预测，具体为

$$\boldsymbol{y} = \boldsymbol{W}_{\text{fc}}\boldsymbol{x}_{\text{fusion}} + \boldsymbol{b}_{\text{fc}} \tag{18}$$

式中，$\boldsymbol{W}_{\text{fc}} \in \mathbf{R}^{C \times d_{\text{in}}}$ 和 $\boldsymbol{b}_{\text{fc}} \in \mathbf{R}^{C}$ 分别为分类器的权重矩阵和偏置项，C 为预测类别的数量。

2　实验结果与分析

2.1　实验环境与数据集

本文实验在 NVIDIA RTX4090 D GPU 上使用 PyTorch 框架进行训练和测试，PyTorch 版本为 2.4.1+cu118，图神经网络所使用的 PyTorch Geometric 版本为 2.6.1。在感知模块中，GEA-M 模块的堆叠层数为 2，激活函数使用 GELU，其中 Transformer 的注意力头数量为 2。在融合模块中，ResMLP 模块使用 LeakyReLU 作为激活函数。在训练过程中，使用 AdamW 优化器，设定初始学习率为 5.0×10^{-4}，权重衰减系数为 1.0×10^{-5}，批量大小设置为 8。实施早停策略，在连续 10 次的训练中验证集的性能没有任何改进时终止训练。

本文采用两个公共的 TCGA（the cancer genome atlas）数据集（Kather 等，2019）：结直肠癌（TCGA_CRC）和胃腺癌（TCGA_STAD）。TCGA_CRC 数据集由结直肠癌患者的 H&E（hematoxylin and eosin）染色切片图像构成，共包括 430 个样本，根据已标注的 MSI 状态，样本分为微卫星不稳定（MSI）和微卫星稳定（microsatellite stable，MSS）两类，其中，78 例为 MSI 样本，352 例为 MSS 样本。TCGA_STAD 数据集为胃腺癌患者的 H&E 切片图像，共包含 326 个样本，其中，64 例为 MSI 样本，262 例为 MSS 样本。

此外，为验证所提出多尺度图神经网络模型在不同病理任务场景下的泛化能力与结构建模稳健性，本文进一步在 Camelyon16 数据集（Ehteshami Bejnordi 等，2017）上进行了跨任务的补充实验。Camelyon16 是一个广泛使用的 WSI 分析基准数据集，其任务目标为乳腺癌前哨淋巴结转移检测，即对 WSI 进行正常与肿瘤转移二分类判别。该数据集在训练集中提供了转移灶区域的像素级标注，在测试阶段仅提供切片级标签，常用于评估 WSI 分类模型在复杂病理场景下的判别能力。该数据集共包含 400 幅全切片病理图像，官方提供了训练集与测试集的划分。其中，训练集含有 111 张阳性切片与 159 张正常切片，测试集共 130 幅，包含另外的独立阴性和阳性图像。为了均衡正负样本比例，在训练集中使用 Weighted-RandomSampler 进行加权随机采样，提高模型对少数类的识别能力。数据集划分遵循 Kather 等人（2019）提出的划分方式。数据集类别统计如表 1 所示。

表 1　数据集类别统计

Table 1　Dataset category statistics

数据集	类别及数量/幅
TCGA_CRC	MSI: 78 / MSS: 352
TCGA_STAD	MSI: 64 / MSS: 262
Camelyon16	Normal: 240 / Tumor: 160

2.2　评估指标

本文采用准确率（accuracy，Acc）、AUC（area under curve）和 F1 分数作为模型性能评估指标。Acc 用于衡量整体分类准确性，AUC 用于评估模型在不同判别阈值下的整体判别能力，F1 分数综合反映模型在类别不平衡场景下对少数类的识别性能。

2.3　对比实验

本文展示了在 TCGA_CRC 和 TCGA_STAD 两个数据集上的实验结果，并与以下方法进行了比较：两个传统方法 Mean-Pooling 和 Max-Pooling，以及 Shen 等人（2023）提出的 GNN-CNN 融合框架、DTFD-MIL（double-tier feature distillation multiple instance learning）（Zhang 等，2022）、CLAM_SB（clustering-constrained attention multiple instance learning）（Lu 等，2021）、CLAM_MB（Lu 等，2021）、ILRA-MIL（iterative low-rank attention multiple instance learning）（Xiang 和 Zhang，2022）、CAMIL（channel attention-based multiple instance learning）（Mao 等，2025）、GAT-Mamba（graph attention network-Mamba）（Ding 等，2025）、WiKG（whole slide image is a knowledge graph）（Li 等，2024）和 DGR-MIL（diverse global representation multiple instance learning）（Zhu 等，2025）。所有方法的结果均来自于在相同设置下使用其官方代码进行的实验，结果如表 2 所示。

在 TCGA_CRC 数据集上，GEAMF 在 3 项指标上均取得当前最优性能，其中，Acc 相较最优对比方法提升 6.3%，AUC 提升 1%，F1 提升 0.3%。在

表2　不同模型在TCGA_CRC、TCGA_STAD和Camelyon16数据集上的性能比较

Table 2　Comparisons of model performance on the TCGA_CRC, TCGA_STAD and Camelyon16 datasets

方法	TCGA_CRC			TCGA_STAD			Camelyon16		
	Acc	AUC	F1	Acc	AUC	F1	Acc	AUC	F1
Mean-Pooling	0.829	0.711	0.777	0.847	0.790	0.733	0.811	0.779	0.784
Max-Pooling	0.837	0.783	0.804	0.865	0.755	0.707	0.834	0.746	0.754
GNN-CNN fusion(Shen等,2023)	0.829	0.770	0.863	0.825	0.818	0.852	0.812	0.769	0.705
DTFD-MIL(Zhang等,2022)	0.806	0.823	0.853	0.816	0.805	0.878	0.821	0.829	0.814
CLAM_SB(Lu等,2021)	0.861	0.818	0.762	0.806	0.803	0.707	0.820	0.825	0.775
CLAM_MB(Lu等,2021)	0.845	0.822	0.744	0.816	0.810	0.664	0.821	0.831	0.774
ILRA-MIL(Xiang和Zhang,2022)	0.830	0.760	0.876	0.837	0.732	**0.905**	0.848	0.826	0.868
CAMIL(Mao等,2025)	0.826	0.825	0.884	0.677	0.768	0.747	0.869	0.861	**0.914**
GAT-mamba(Ding等,2025)	0.889	0.823	0.862	0.830	0.812	0.847	0.834	0.865	0.832
WiKG(Li等,2024)	0.870	0.807	0.864	0.871	0.808	0.851	0.800	0.847	0.801
DGR-MIL(Zhu等,2025)	0.852	0.826	0.855	0.841	0.818	0.862	0.845	0.858	0.829
GEAMF(本文)	**0.945**	**0.835**	**0.887**	**0.924**	**0.829**	0.897	**0.907**	**0.877**	0.848

注:加粗字体表示各列最优结果。

TCGA_STAD数据集上,GEAMF在Acc和AUC指标上分别提升6%和1.3%,F1略低于ILRA-MIL,这可能与ILRA-MIL针对胃腺癌组织复杂性引入低秩预训练和迭代全局注意力机制,因此更贴合该数据分布有关。为进一步评估模型在不同病理任务场景下的稳定性,本文在Camelyon16淋巴结转移检测数据集上进行了跨病理任务对比实验。实验结果表明,GEAMF在Acc(0.907)和AUC(0.877)上取得最优性能,显示出在病灶分布高度不均衡、尺度差异显著的转移检测任务中仍具备良好的判别能力与泛化性。该数据集上F1指标略低于CAMIL,可能与其专门设计的通道注意力模块在少数类样本细粒度表征与召回方面具有优势有关。此外,GEAMF在Acc指标上取得了尤为显著的提升,其核心原因在于模型在图结构建模与多尺度融合阶段有效降低了多数类样本的误判率。具体而言,GC-M模块通过联合约束语义一致性与空间邻近性,减少了背景区域之间的无效连接;GEA-M进一步在消息传递过程中显式抑制低判别性边的传播强度,从而避免背景噪声在图结构中的累积放大。上述机制使模型在高度类别不平衡的MSI任务中能够更稳定地区分真实病灶与非病灶区域,形成更可靠的决策边界。与此同时,多尺度图结构的协同建模及尺度感知门控融合增强了切片级表征的一致性,使该优势在跨病理任务实验中同样得以保持。

2.4　消融实验

2.4.1　GC-M模块消融实验

为了验证GC-M模块中不同边属性构建方式对模型性能的影响,本文在TCGA_CRC、TCGA_STAD以及Camelyon16这3个数据集上进行了消融实验,比较了基于余弦相似度(cos)、欧氏距离(eucl)以及两者融合(cos + eucl)3种边属性定义方式的分类效果,实验结果如表3所示。具体来看,cos在捕捉语义相似性方面表现出一定优势,在各类数据集上其AUC均高于eucl;而eucl在反映空间邻近性上效果更好,例如在TCGA_CRC上的F1分数达到0.846,优于cos。但两者单独使用时均存在明显短板:cos可能引入“语义相近但空间离散”的弱关联边,导致局部结构约束不足;eucl则可能将“空间相邻但语义无关”的噪声patch连接起来,造成信息误聚合。融合策略(cos + eucl)在3项任务上均取得近最优结果(例如TCGA_CRC上达到Acc = 0.945、AUC = 0.835、F1 = 0.887),并在TCGA_STAD上保持一致增益,说明联合刻画“语义一致性 + 空间邻近性”能够有效减少单一度量带来的结构偏置,提升图结构的鲁棒性。

表 3　GC-M 模块消融实验

Table 3　Ablation study on GC-M module

模块	TCGA_CRC			TCGA_STAD			Camelyon16		
	Acc	AUC	F1	Acc	AUC	F1	Acc	AUC	F1
cos	0.927	0.778	0.818	0.919	0.739	0.832	0.903	0.813	0.816
eucl	0.925	0.771	0.846	0.910	0.713	0.779	0.880	0.788	**0.854**
cos+eucl	**0.945**	**0.835**	**0.887**	**0.924**	**0.829**	**0.897**	**0.907**	**0.877**	0.848

注：加粗字体表示各列最优结果。

在 Camelyon16 这一不同病理任务场景下，融合策略（cos + eucl）在 F1 指标上略低于使用 eucl 的设置。这与该任务高度类别不平衡且病灶具有显著空间局部性的特点有关：仅依赖空间距离的构图方式更易形成连续响应，从而在固定阈值下获得较高召回率。然而，融合策略在 Acc 与 AUC 等整体判别指标上仍保持优势，说明其在预测排序稳定性与噪声抑制方面更具鲁棒性。该结果表明，所提出的 GC-M 模块在不同病理任务中均表现出较稳定而有效的建模能力，从而验证了该模块在跨病理任务场景下的泛化稳定性。

2.4.2　GEA-M 模块有效性对比

本文进一步对比了所提出的 GEA-M 模块与经典的 GNN 模型 GCN（graph convolutional network）（Kipf 和 Welling，2017）、GIN（graph isomorphism network）（Xu 等，2019）、GAT（Veličković 等，2018）在多尺度输入下的性能，结果如表 4 所示。具体而言，在 TCGA_CRC 数据集上，GEA-M 在 Acc、AUC 以及 F1 分数上分别达到 0.945、0.835 与 0.887，显著优于其他 3 种模块，表明 GEA-M 能够更准确地区分 MSI 与 MSS 两类样本，提升模型的判别能力。在 TCGA_STAD 数据集上，GEA-M 同样展现出稳健的性能，Acc、AUC 与 F1 分数分别达到 0.924、0.829 与 0.897，均优于其他方法。在 Camelyon16 淋巴结转移检测任务中，GEA-M 模块同样展现出显著优势。具体而言，GEA-M 在该数据集上取得了 Acc = 0.907、AUC = 0.877、F1 = 0.848，分别较 3 种经典图神经网络最高的 Acc（GCN 的 0.887）、AUC（GIN 的 0.790）和 F1（GCN 的 0.818）提升了 2.2%、11% 和 3.7%。该结果表明，通过在多尺度图上融合边属性与注意力机制，GEA-M 不仅在以分子状态区分为目标的 MSI 检测任务中取得了稳定提升，在以空间局部病灶识别为核心的淋巴结转移检测任务中同样保持了良好性能，从而验证了该模块在不同病理任务场景下对多尺度结构关系建模的有效性。

2.4.3　融合模块消融

为验证所提出尺度感知门控残差融合模块（SGRes-M）的有效性，本文在 3 个数据集上进行了消融实验，将 SGRes-M 分别替换为简单多层感知机（MLP）和 Transformer 融合模块，实验结果如表 5 所示。在 TCGA_CRC 数据集上，SGRes-M 在 AUC 和 F1 指标上分别达到 0.835 和 0.887，显著优于 MLP 与 Transformer；在 TCGA_STAD 数据集上，SGRes-M 同样在 AUC 和 F1 指标上领先；在 Camelyon16 淋巴结转移检测任务中，SGRes-M 在 AUC 指标上继续保持领先，而在 Acc 上略低于 Transformer，F1 略低于

表 4　GEA-M 模块消融实验

Table 4　Ablation study results of the GEA-M module

模块	TCGA_CRC			TCGA_STAD			Camelyon16		
	Acc	AUC	F1	Acc	AUC	F1	Acc	AUC	F1
GCN	0.773	0.687	0.765	0.756	0.619	0.840	0.887	0.741	0.818
GIN	0.925	0.674	0.806	0.830	0.790	0.805	0.854	0.790	0.805
GAT	0.940	0.659	0.831	0.842	0.774	0.813	0.878	0.743	0.771
GEA-M	**0.945**	**0.835**	**0.887**	**0.924**	**0.829**	**0.897**	**0.907**	**0.877**	**0.848**

注：加粗字体表示各列最优结果。

MLP。从融合策略对比来看，MLP融合仅通过线性或浅层非线性映射对多尺度特征进行压缩，缺乏对尺度间结构差异与互补关系的显式建模能力，因此在3个数据集上的综合判别性能未表现出明显优势。然而，该融合方式往往产生较为激进的阳性预测，在固定判别阈值下能够获得略高的召回率，从而在特定设置下表现出相对较高的F1分数。Transformer融合模块在参数量显著增加（245 K）的情况下，其性能提升却较为有限，甚至在部分数据集（如TCGA_CRC）上不及MLP，表明直接引入全局自注意力并不能有效适配WSI多尺度图级特征的融合需求，反而可能因过度建模全局依赖而引入尺度噪声。值得注意的是，Transformer在3个数据集上Acc更高，主要因为其容量更大，更容易在默认阈值下拟合“多数类占优”的决策边界；在类别不平衡的WSI任务中，这种偏向多数类的输出会抬高总体正确率。

表5　SGRes-M模块的消融实验

Table 5　Ablation study on SGRes-M module

模块	TCGA_CRC			TCGA_STAD			Camelyon16			
	Acc	AUC	F1	Acc	AUC	F1	Acc	AUC	F1	参数量/k
MLP	0.944	0.805	0.873	0.918	0.817	0.889	0.897	0.800	**0.852**	228
Transformer	**0.948**	0.754	0.815	**0.960**	0.712	0.792	**0.914**	0.835	0.847	245
SGRes-M	0.945	**0.835**	**0.887**	0.924	**0.829**	**0.897**	0.907	**0.877**	0.848	236

注：加粗字体表示各列最优结果。

2.4.4　多尺度消融

为了系统验证所选放大倍率组合的合理性与泛化能力，本文在多个数据集上对单倍率与多倍率输入配置进行了消融实验，结果如表6所示。

表6　不同输入倍率的消融实验

Table 6　Ablation study on different input magnifications

倍率	TCGA_CRC			TCGA_STAD		
	Acc	AUC	F1	Acc	AUC	F1
20×	0.837	0.703	0.659	0.837	0.739	0.830
40×	0.814	0.730	0.727	0.831	0.775	0.848
20× + 40×	**0.945**	**0.835**	**0.887**	**0.924**	**0.829**	**0.897**

注：加粗字体表示各列最优结果。

首先，在TCGA_CRC与TCGA_STAD数据集上，多倍率融合显著优于单倍率输入。在TCGA_CRC数据集上，20× + 40×组合分别取得了0.945的Acc、0.835的AUC和0.887的F1分数；在TCGA_STAD数据集上亦达到0.924的Acc、0.829的AUC和0.897的F1分数，均明显优于单一倍率模型。该结果表明，仅依赖单倍率输入难以充分表达WSI在不同尺度下所蕴含的多层次结构信息，而融合多尺度特征有助于同时捕捉组织的全局结构与局部细节，从而提升模型对MSI与MSS样本差异的建模能力。

在此基础上，本文进一步在Camelyon16淋巴结转移检测这一不同病理任务场景下开展多尺度消融，用于检验倍率组合选择是否具有跨任务的稳健性，如表7所示。单尺度结果显示，随着放大倍率的提升，模型对病灶细节纹理的捕捉能力逐步增强。然而，当将10×与20×或40×进行融合时，尽管AUC相较于10×单独输入有所提高，但F1分数仍低于20×或40×的单尺度结果。相比之下，20× + 40×组合在Acc、AUC和F1这3项指标上均取得最优表现，较最佳单尺度40×分别提升了2.7%、10.9%和1.4%。这一结果表明，20× + 40×的倍率组合不仅在MSI分类任务上带来稳定增益，同时在不同病理类型与判别目标的Camelyon16场景下仍保持优势，

表7　Camelyon16数据集不同输入倍率的消融实验结果

Table 7　Ablation study on different input magnifications on the Camelyon16 dataset

倍率	Acc	AUC	F1
10×	0.837	0.664	0.697
20×	0.870	0.763	0.846
40×	0.883	0.791	0.836
10× + 20×	0.871	0.739	0.787
10× + 40×	0.840	0.801	0.787
20× + 40×	**0.907**	**0.877**	**0.848**

注：加粗字体表示各列最优结果。

从而为本文所采用的多尺度配置提供了补充性的跨任务验证。

2.5　可视化分析

2.5.1　t-SNE 分布

本文首先在 TCGA_CRC 与 TCGA_STAD 两个 MSI 分类数据集上，采用 t 分布随机邻域嵌入（t-distributed stochastic neighbor embedding，t-SNE）对不同输入倍率下模型提取的图级特征进行可视化分析，以验证多尺度融合对特征判别性的直接影响，如图 4 所示。图中“维度 1”和“维度 2”分别表示 t-SNE 降维后得到的两个嵌入坐标轴。第 1、2 行分别对应 TCGA_CRC 和 TCGA_STAD 两个数据集，在单一倍率（20×、40×）和多倍率（20× + 40×）输入下，模型学习到的样本分布存在显著差异。在仅使用单尺度时，TCGA_CRC 中 MSI（蓝点）与 MSS（红点）样本高度混杂，难以划出清晰的分界；而融合 20× + 40× 后，两类样本在二维投影中被明显拉开，MSI 与 MSS 各自形成紧凑簇群。在 TCGA_STAD 上，虽然单尺度输入已经能将两类大致分离，但簇内点分布松散；多尺度融合后，簇内紧致度大幅提高，说明模型更好地捕获了同类样本的共同语义特征。

在此基础上，为进一步验证多尺度融合策略在不同任务与数据分布下的泛化表现，本文在全新的 Camelyon16 淋巴结转移检测数据集上对不同倍率及其组合进行了 t-SNE 可视化分析，结果如图 5 所示。单一倍率（10×、20×、40×）下，样本在低维嵌入空间中均存在不同程度的类间混叠，其中 40× 的分离趋势相对较好，但仍难以形成清晰边界。多倍率融合后，10× + 20× 与 10× + 40× 的类边界清晰，但类内较松散；相比之下，20× + 40× 的特征嵌入呈现出最清晰、稳定的聚类结构，Normal 与 Tumor 样本形成两个几乎完全分离的簇，表明融合后的特征具有更强的判别一致性。值得注意的是，尽管 10× + 20× 在部分指标上未优于单尺度输入，但其 t-SNE 分布相较单倍率呈现出更清晰的局部结构。这一现象说明，多倍率融合能够在特征空间中构建更合理的邻域关系，但其最终分类性能仍高度依赖于各尺度所提供信息的判别有效性。在 Camelyon16 这一以局部病灶为核心的高度不平衡任务中，中高倍率的协同建模显然更具优势。

2.5.2　注意力热力图对比

在 Camelyon16 数据集上对模型的关注区域进行了可视化分析，结果如图 6 所示。图 6（b）为通过预处理算法得到的组织区域分割结果，图 6（c）为 Camelyon16 数据集提供的阳性病灶标注，图 6（h）为本文提出的 GEAMF 的分割结果。从图 6 可以观察

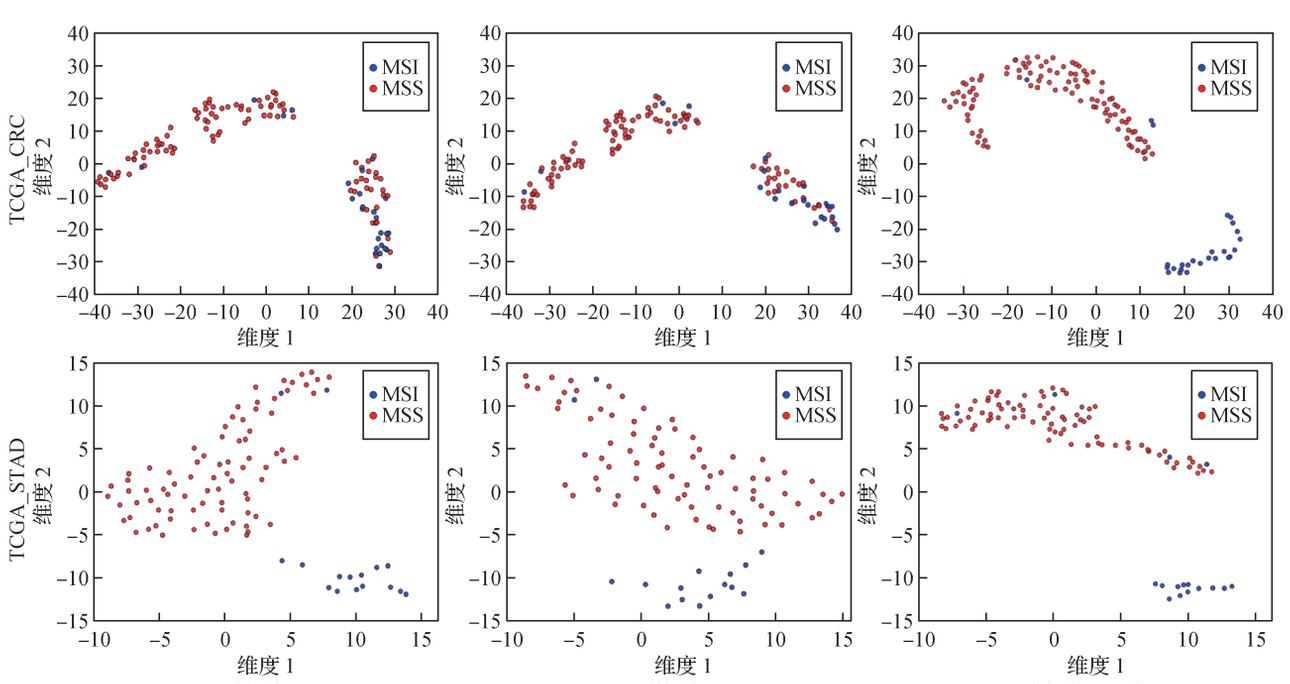


图 4　TCGA：t-SNE 分布对比

Fig. 4　TCGA: t-SNE distribution comparison ((a) 20×; (b) 40×; (c) 20× + 40×)

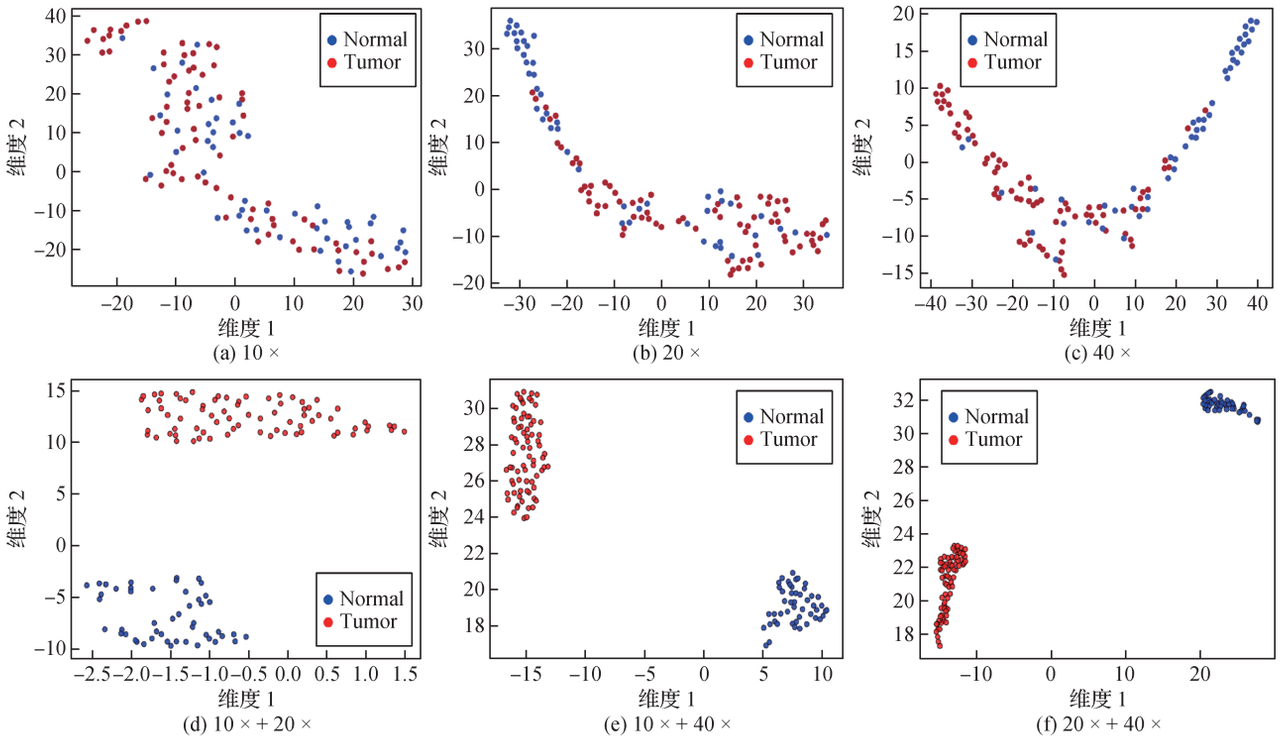


图5　Camelyon16:t-SNE分布对比

Fig. 5　Camelyon16: t-SNE distribution comparison

((a) 10 ×; (b) 20 ×; (c) 40 ×; (d) 10 × + 20 ×; (e) 10 × + 40 ×; (f) 20 × + 40 ×)

到,CLAM所生成的注意力图呈现出高度稀疏的响应模式,其高响应区域通常集中于少量局部区域,体现了基于注意力池化机制对少数判别性实例的强选择能力。然而,该特性也使其对局部噪声较为敏感,且对病灶整体空间结构的刻画相对有限。相比之下,DTFD-MIL的注意力分布更为弥散,高响应区域在组织范围内呈现出较为连续的分布趋势,这反映了其通过知识蒸馏机制聚合多实例弱判别证据的建

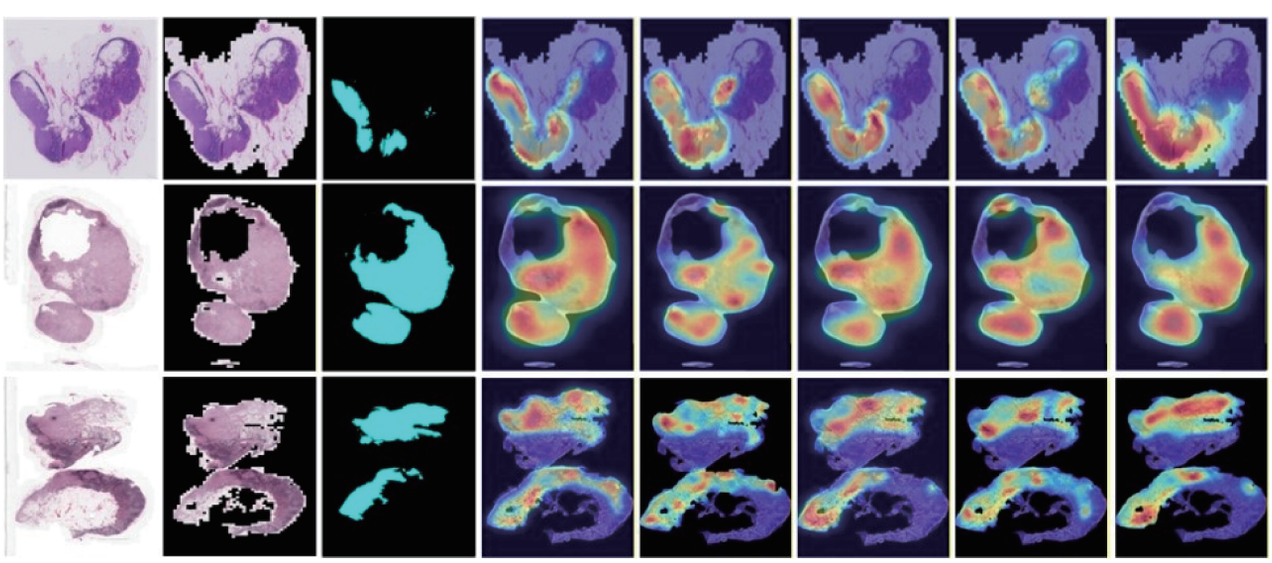


图6　注意力热力图

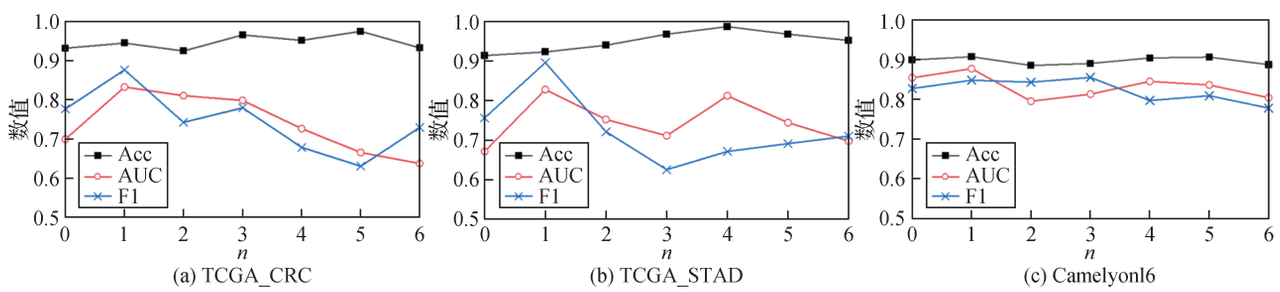
Fig. 6　Attention heatmap ((a) WSI; (b) the tissue segmented with a pre-processing algorithm; (c) ground truth provided by Camelyon16; (d) GNN-CNN fusion; (e) CLAM; (f) CAMIL; (g) DTFD-MIL; (h) GEAMF(proposed))

模特点，但在病灶边界定位方面判别性相对不足。与上述方法不同，GEAMF 在保持注意力分布稀疏性的同时，能够在病灶相关区域形成更具结构一致性的响应模式，其高响应区域在空间上与 Camelyon16 提供的阳性病灶标注具有更高的一致性。这表明，通过显式建模多尺度结构关系并引入跨尺度语义对齐机制，GEAMF 能够在细粒度判别能力与全局结构感知之间取得更为平衡的注意力表达，从而提升模型的可解释性与鲁棒性。

2.6　参数敏感性分析

2.6.1　多头注意力参数敏感性分析

为了探究多头注意力机制中注意力头数量对性能的影响，在 TCGA_CRC、TCGA_STAD 和 Camelyon16 这 3 个数据集上分别进行了消融实验，如图 7 所示，在该实验中，由于隐藏层维度固定为 128，本文将注意力头的数量取值为 2^n，并在图 7 中以指数 n 作为横坐标，以表示注意力头的数量的指数增长关系。



(a) TCGA_CRC　(b) TCGA_STAD　(c) Camelyonl6

图 7　不同注意力头数量（2^n）下的模型性能对比

Fig. 7　Model performance under varying numbers of attention heads (2^n) ((a) TCGA_CRC; (b) TCGA_STAD; (c) Camelyon16)

实验结果表明，随着注意力头数量的增加，部分数据集上的 Acc 指标呈上升趋势（例如在 TCGA_CRC 中达到最高值），但 AUC 与 F1 并未同步提升，且在头数过多时明显下降。这表明，在特征维度固定的条件下，过多注意力头会压缩单头子空间维度，导致表征能力分散，并可能引入冗余或不稳定的注意力模式，从而削弱对少数类样本的判别能力。在类别分布高度不均衡的 WSI 分析任务中，Acc 易受到多数类主导，而 AUC 与 F1 更能反映模型在整体排序能力和少数类识别上的稳定性。综合 3 项指标可见，当注意力头数量设置为 2^1 时，3 个数据集在 AUC 与 F1 上均达到或接近最优（如 TCGA_CRC：AUC = 0.835，F1 = 0.877；TCGA_STAD：AUC = 0.829，F1 = 0.897；Camelyon16：AUC = 0.877，F1 = 0.848），表现出更稳定的判别性能与更好的泛化能力。因此，在主实验中采用了注意力头数量为 2^1 作为默认设置，以实现性能的兼顾。

2.6.2　图构建超参数敏感性分析

在图构建过程中，选用 $\theta = 10$，$k = 4$ 的超参数设定，为分析构图超参数对模型性能的影响，在 Camelyon16 数据集上分别对阈值参数 θ 和 KNN 邻居数 k 进行了系统消融实验。

如表 8 所示，在固定 $k = 4$ 时，随着 θ 从 5 增加至 10，模型在 Acc、AUC 与 F1 指标上均达到最优，表明适当放宽空间约束有助于增强病灶区域内的局部连通性，促进弱阳性证据的有效聚合；而当 θ 进一步增大时，引入的空间邻近但语义不一致的背景节点增多，导致邻域纯度下降并削弱模型的判别稳定性。如表 9 所示，在 k 的消融实验（固定 $\theta = 10$）中，模型在 $k = 4$ 时取得最佳性能：k 过小邻域信息不足，k 过大则会引入冗余背景节点，稀释病灶判别信息，造成 AUC 与 F1 的同步下降。该结果表明，在 Camelyon16 这类高度不平衡的转移检测任务中，适度的邻域规模更有利于兼顾信息充分性与判别鲁棒性。

表 8　Camelyon16 数据集图构建超参数 θ 的消融实验（$k = 4$）

Table 8　Ablation results of graph construction by parameter θ on Camelyon16 dataset ($k = 4$)

θ ($k = 4$)	Acc	AUC	F1
5	0.892	0.859	0.828
10	**0.907**	**0.877**	**0.848**
15	0.899	0.868	0.837
20	0.885	0.852	0.820

注：加粗字体表示各列最优结果。

表9 Camelyon16数据集图构建超参数k的消融实验（$\theta = 10$）

Table 9 Ablation results of graph construction by parameter k on Camelyon16 dataset ($\theta = 10$)

k ($\theta = 10$)	Acc	AUC	F1
2	0.884	0.851	0.822
4	**0.907**	**0.877**	**0.848**
6	0.902	0.872	0.843
8	0.895	0.865	0.835

注：加粗字体表示各列最优结果。

3 结 论

本文提出了一种基于多实例学习框架的图边属性注意力的多尺度融合网络（GEAMF），用于实现对组织病理图像中微卫星不稳定性（MSI）状态的预测。该模型以WSI的图像块作为实例，构建图结构以建模图像块之间的空间与语义关系，并融合多尺度病理信息以增强整体表征能力。对比实验结果表明，本文方法在TCGA_CRC和TCGA_STAD数据集上均取得了优于现有方法的性能。同时，通过在Camelyon16淋巴结转移检测这一不同病理任务上的实验，进一步验证了模型在跨任务场景下的强泛化能力与结构建模的稳健性。在消融实验中，验证了所提出的GEA-M模块相较于GCN、GIN和GAT模型的优越性。进一步地，验证了多尺度融合在WSI分析中的优势。

本文提出的图边属性注意力机制为结构关系的可学习建模提供了一种通用范式。未来工作可将该机制拓展至跨模态或跨任务场景，例如联合建模组织病理图像与分子组学、临床变量等异构信息，以探索结构感知图建模在多模态医学分析中的潜力。

参考文献（References）

Alon Y and Zhou H Y. 2022. Neuroplastic graph attention networks for nuclei segmentation in histopathology images [EB/OL]. [2026-01-10]. https://arxiv.org/pdf/2201.03669.pdf

Alzoubi I, Zhang L, Zheng Y Q, Loh C, Wang X Y and Graeber M B. 2024. PathoGraph: an attention-based graph neural network capable of prognostication based on CD276 labelling of malignant glioma cells. Cancers, 16 (4): #750 [DOI: 10.3390/cancers16040750]

Bayramoglu N, Kannala J and Heikkilä J. 2016. Deep learning for magnification independent breast cancer histopathology image classification//Proceedings of the 23rd International Conference on Pattern Recognition (ICPR). Cancun, Mexico: IEEE, 2440-2445 [DOI: 10.1109/ICPR.2016.7900002]

Bhojanapalli S, Yun C, Rawat A S, Reddi S and Kumar S. 2020. Low-rank bottleneck in multi-head attention models//Proceedings of the 37th International Conference on Machine Learning. JMLR.org: #81 [DOI: 10.5555/3524938.3525019]

Bontempo G, Bolelli F, Porrello A, Calderara S and Ficarra E. 2024. A graph-based multi-scale approach with knowledge distillation for WSI classification. IEEE Transactions on Medical Imaging, 43(4): 1412-1421 [DOI: 10.1109/TMI.2023.3337549]

Chan T H, Cendra F J, Ma L, Yin G S and Yu L Q. 2023. Histopathology whole slide image analysis with heterogeneous graph representation learning//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, Canada: IEEE: 15661-15670 [DOI: 10.1109/CVPR52729.2023.01503]

Chen R J, Chen C K, Li Y C, Chen T Y, Trister A D, Krishnan R G, et al. 2022. Scaling vision transformers to gigapixel images via hierarchical self-supervised learning//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 16123-16134 [DOI: 10.1109/CVPR52688.2022.01567]

Ding R W, Luong K D, Rodriguez E, da Silva A C A L and Hsu W. 2025. Combining graph neural network and Mamba to capture local and global tissue spatial relationships in whole slide images. Scientific Reports, 15 (1): #18261 [DOI: 10.1038/s41598-025-99042-4]

Ding S S, Li J C, Wang J, Ying S H and Shi J. 2023. Multi-scale efficient graph-transformer for whole slide image classification. IEEE Journal of Biomedical and Health Informatics, 27(12): 5926-5936 [DOI: 10.1109/JBHI.2023.3317067]

Ding W L, Zhu F L, Zheng K and Jia X P. 2024. Classification of breast pathological images based on multiscale information interaction and fusion. Journal of Image and Graphics, 29(4): 1085-1099 (丁维龙, 朱峰龙, 郑魁, 贾秀鹏. 2024. 多尺度信息交互与融合的乳腺病理图像分类. 中国图象图形学报, 29(4): 1085-1099) [DOI: 10.11834/jig.221178]

Ehteshami Bejnordi B, Veta M, Van Diest P J, Van Ginneken B, Karssemeijer N, Litjens G, et al. 2017. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. JAMA, 318(22): 2199-2210 [DOI: 10.1001/jama.2017.14585]

Feng L F, Chen Y, Zhou T H, Hu F and Yi Z. 2022. Review of human lung and lung lesion regions segmentation methods based on CT images. Journal of Image and Graphics, 27(3): 722-749 (冯龙锋, 陈英, 周滔辉, 胡菲, 易珍. 2022. CT图像肺及肺病变区域分割

方法综述. 中国图象图形学报, 27(3): 722-749)[DOI: 10.11834/jig.210769]

Gregor K, Danihelka I, Graves A, Rezende D and Wierstra D. 2015. DRAW: a recurrent neural network for image generation//Proceedings of the 32nd International Conference on Machine Learning. Lille, France: PMLR: 1462-1471

He F, Zhu Y M, Wang W B, Nanding A, Kuai Z X, Li X M, et al. 2022. Multi-instance classification of histopathological breast cancer images with visual explanation//Proceedings of the 16th IEEE International Conference on Signal Processing. Beijing, China: PMLR: 431-436[DOI: 10.1109/ICSP56322.2022.9965331]

Hou L, Samaras D, Kurc T M, Gao Y, Kong J E, Saltz J H, et al. 2016. Patch-based convolutional neural network for whole slide tissue image classification//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE: 2424-2433[DOI: 10.1109/CVPR.2016.266]

Kather J N, Pearson A T, Halama N, Jäger D, Krause J, Loosen S H, et al. 2019. Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer. Nature Medicine, 25(7): 1054-1056[DOI: 10.1038/s41591-019-0462-y]

Kipf T N and Welling M. 2017. Semi-supervised classification with graph convolutional networks//Proceedings of 2017 International Conference on Learning Representations. Toulon, France: ICLR: 1-14.

Lerousseau M, Vakalopoulou M, Classe M, Adam J, Battistella E, Carré A, et al. 2021. Weakly supervised multiple instance learning histopathological tumor segmentation//Proceedings of the 23rd International Conference on Medical Image Computing and Computer Assisted Intervention(MICCAI). Lima, Peru: Springer: 470-479[DOI: 10.1007/978-3-030-59722-1_45]

Li J W, Chen Y X, Chu H B, Sun Q H, Guan T, Han A J, et al. 2024. Dynamic graph representation with knowledge-aware attention for histopathology whole slide image analysis//Proceedings of 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 11323-11332[DOI: 10.1109/CVPR52733.2024.01076]

Lin Z H, Mo X G, Li H X and Li H B. 2002. Comparison of three spatial interpolation methods for climate variables in China. Acta Geographica Sinica, 57(1): 47-56(林忠辉, 莫兴国, 李宏轩, 李海滨. 2002. 中国陆地区域气象要素的空间插值. 地理学报, 57(1): 47-56)[DOI: 10.3321/j.issn:0375-5444.2002.01.006]

Lu M Y, Williamson D F K, Chen T Y, Chen R J, Barbieri M and Mahmood F. 2021. Data-efficient and weakly supervised computational pathology on whole-slide images. Nature Biomedical Engineering, 5(6): 555-570[DOI: 10.1038/s41551-020-00682-w]

Lu W Q, Toss M, Dawood M, Rakha E, Rajpoot N and Minhas F. 2022. SlideGraph$^+$: whole slide image level graphs to predict HER2 status in breast cancer. Medical Image Analysis, 80: #102486[DOI: 10.1016/j.media.2022.102486]

Ma S, Liu J W and Zuo X. 2022. Survey on graph neural networks. Journal of Computer Research and Development, 59(1): 47-80(马帅, 刘建伟, 左信. 2022. 图神经网络综述. 计算机研究与发展, 59(1): 47-80)[DOI: 10.7544/issn1000-1239.20201055]

Mao J Y, Xu J L, Tang X F, Liu Y J, Zhao H, Tian G, et al. 2025. CAMIL: channel attention-based multiple instance learning for whole slide image classification. Bioinformatics, 41(2): #btaf024[DOI: 10.1093/bioinformatics/btaf024]

Ozen Y, Aksoy S, Kösemehmetoğlu K, Önder S and Üner A. 2021. Self-supervised learning with graph neural networks for region of interest retrieval in histopathology//Proceedings of the 25th International Conference on Pattern Recognition. Milan, Italy: IEEE: 6329-6334[DOI: 10.1109/ICPR48806.2021.9412903].

Saillard C, Dubois R, Tchita O, Loiseau N, Garcia T, Adriansen A, et al. 2023. Validation of MSIntuit as an AI-based pre-screening tool for MSI detection from colorectal cancer histology slides. Nature Communications, 14(1): #6695[DOI: 10.1038/s41467-023-42453-6]

Shen Y Q, Zhou B X, Xiong X Y, Gao R T and Wang Y G. 2023. How GNNs facilitate CNNs in mining geometric information from large-scale medical images//Proceedings of 2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). Istanbul, Turkiye: IEEE: 2227-2230[DOI: 10.1109/BIBM58861.2023.10385379]

Shen Y T, Chen Z, Zhang Q H, Chen J H and Wang Q G. 2024. Fully and weakly supervised graph networks for histopathology image segmentation. Journal of Image and Graphics, 29(3): 697-712(沈熠婷, 陈昭, 张清华, 陈锦豪, 王庆国. 2024. 全监督和弱监督图网络的病理图像分割. 中国图象图形学报, 29(3): 697-712)[DOI: 10.11834/jig.230477]

Shi J, Shu T, Wu K, Jiang Z G, Zheng L P, Wang W, et al. 2024. Masked hypergraph learning for weakly supervised histopathology whole slide image classification. Computer Methods and Programs in Biomedicine, 253: #108237[DOI: 10.1016/j. cmpb. 2024.108237]

Touvron H, Bojanowski P, Caron M, Cord M, El-Nouby A, Grave E, et al. 2023. ResMLP: feedforward networks for image classification with data-efficient training. IEEE Transactions on Pattern Analysis and Machine Intelligence, 45(4): 5314-5321[DOI: 10.1109/TPAMI.2022.3206148]

Veličković P, Cucurull G, Casanova A, Romero A, Liò P and Bengio Y. 2018. Graph attention networks//Proceedings of the 6th International Conference on Learning Representations (ICLR). Vancouver, Canada: ICLR

Wang X Y, Yang S, Zhang J, Wang M H, Zhang J, Huang J Z, et al. 2021. TransPath: transformer-based self-supervised learning for histopathological image classification//Proceedings of the 24th International Conference on Medical Image Computing and Computer Assisted Intervention——MICCAI 2021. Strasbourg, France:

Springer: 186-195 [DOI: 10.1007/978-3-030-87237-3_18]

Willis J, Lefterova M I, Artyomenko A, Kasi P M, Nakamura Y, Mody K, et al. 2019. Validation of microsatellite instability detection using a comprehensive plasma-based genotyping panel. Clinical Cancer Research, 25(23): 7035-7045 [DOI: 10.1158/1078-0432.CCR-19-1324]

Xiang H C, Shen J Y, Yan Q G, Xu M L, Shi X S and Zhu X F. 2023. Multi-scale representation attention based deep multiple instance learning for gigapixel whole slide image analysis. Medical Image Analysis, 89: #102890 [DOI: 10.1016/j.media.2023.102890]

Xiang J X and Zhang J. 2022. Exploring low-rank property in multiple instance learning for whole slide image classification//Proceedings of the 11th International Conference on Learning Representations (ICLR). Kigali, Rwanda: ICLR

Xie Q Z, Luong M T, Hovy E and Le Q V. 2020. Self-training with noisy student improves ImageNet classification//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 10684-10695 [DOI: 10.1109/CVPR42600.2020.01070]

Xu K, Hu W H, Leskovec J and Jegelka S. 2019. How powerful are graph neural networks?//Proceedings of the 7th International Conference on Learning Representations (ICLR). New Orleans, USA: ICLR

Zhang H R, Meng Y D, Zhao Y T, Qiao Y H, Yang X Y, Coupland S E, et al. 2022. DTFD-MIL: double-tier feature distillation multiple instance learning for histopathology whole slide image classification//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 18780-18790 [DOI: 10.1109/CVPR52688.2022.01824]

Zhao Y, Yang F, Fang Y Q, Liu H L, Zhou N Y, Zhang J, et al. 2020. Predicting lymph node metastasis using histopathological images based on multiple instance learning with deep graph convolution//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 4836-4845 [DOI: 10.1109/CVPR42600.2020.00489]

Zheng Y, Gindra R H, Green E J, Burks E J, Betke M, Beane J E, et al. 2022. A graph-transformer for whole slide image classification. IEEE Transactions on Medical Imaging, 41(11): 3003-3015 [DOI: 10.1109/TMI.2022.3176598]

Zhu W H, Chen X W, Qiu P J, Sotiras A, Razi A and Wang Y L. 2025. DGR-MIL: exploring diverse global representation in multiple instance learning for whole slide image classification//Proceedings of the 18th European Conference on Computer Vision——ECCV 2024. Milan, Italy: Springer: 333-351 [DOI: 10.1007/978-3-031-72920-1_19]

作者简介

滕晓菲，女，硕士研究生，主要研究方向为计算机视觉。
E-mail：t_17nl@163.com
刘慧，通信作者，女，教授，主要研究方向为医学图像处理、数据挖掘与机器学习。E-mail：liuh_lh@sdufe.edu.cn
毛宁，男，研究员，主要研究方向为医学影像与核医学、大数据与人工智能在医学影像中的应用、脑结构与功能成像。
E-mail：maoning@pku.edu.cn
高婧，女，医师，主要研究方向为乳腺癌影像人工智能。
E-mail：gaojingstrive@163.com
陈公冠，男，博士研究生，主要研究方向为图像处理和机器学习。E-mail：cgg970411@163.com