ISSN 1004-9037,CODEN SCYCE4
Journal of Data Acquisition and Processing Vol. 41,No. 1,Jan. 2026,pp. 231－243
DOI:10. 16337/j. 1004-9037. 2026. 01. 016

http://sjcj. nuaa. edu. cn
E-mail:sjcj@nuaa. edu. cn
Tel/Fax:＋86-025-84892742

基于可变形注意力和多尺度多实例学习的全切片病理图像分类方法

薛　保，周俊杰，邵　伟

（南京航空航天大学人工智能学院脑机智能技术教育部重点实验室，南京 211106）

摘　要：全切片图像（Whole slide images，WSIs）是病理学诊断的金标准。准确的组织病理图像分类为肿瘤的类型、分级和分期提供了详细信息，对癌症预后和治疗策略选择具有重要意义。目前，在计算病理学领域中，基于多实例学习（Multi-instance learning，MIL）的分析方法正成为针对WSIs分类问题的主流方法，但该方法大多针对单一尺度病理图像展开，无法在不同层次上理解癌症的产生与发展机制。此外，病理图像的高分辨率特性以及不同尺度病理图像蕴含信息的差异性，也给高效分析单一尺度内的病理图像块以及融合不同尺度下的病理信息带来挑战。为此，本文提出了一种基于可变形注意力和多尺度多实例学习（Deformable attention and multi-scale multi-instance learning，DMSMIL）的全切片病理图像分类方法。具体而言，该方法通过设计可变形注意力分支学习尺度内不同图像块的关联，提升了注意力计算的效率。同时，设计了基于最优传输（Optimal transport，OT）的关联算法融合不同尺度的病理图像，实现了对多尺度病理信息的高效对齐。在乳腺癌亚型分类和肺癌亚型分类任务上的实验结果表明，所提方法分别取得了85.39％和92.00％的分类准确率，相较于主流的WSIs分类方法，性能得到了显著提升。

关键词：多实例学习；可变形注意力；多尺度学习；全切片病理图像分类；最优传输

中图分类号：TP18;R446　　**文献标志码：**A

引用格式：薛保，周俊杰，邵伟. 基于可变形注意力和多尺度多实例学习的全切片病理图像分类方法[J]. 数据采集与处理，2026，41(1)：231-243. XUE Bao，ZHOU Junjie，SHAO Wei. Whole slide pathology image classification method based on deformable attention and multi-scale multi-instance learning[J]. Journal of Data Acquisition and Processing，2026，41(1)：231-243.

引　　言

癌症已成为全球主要的死亡原因之一[1]。从原位癌到侵袭性转移性癌，癌症的临床表现差异很大。因此，对癌症进行有效准确的诊断，并将癌症患者进行亚组分类，进行相应的个性化癌症治疗越来越受到人们的重视。在过去的几十年里，癌症研究取得了重大进展。迄今为止，大量的生物标志物已经被发现并应用于癌症的亚型分类以及诊断，包括影像组学数据[2]、组织病理图像、基因突变、基因表达特征和蛋白质标记。在所有生物标志物中，全切片图像（Whole slide images，WSIs）可以从不同水平反映癌症的潜在分子过程和疾病进展，因此被广泛视为癌症诊断的金标准。

对病理图像进行分类的传统方法主要通过病理医生肉眼对全切片病理图像进行观察，但由于全切片病理图像的尺寸巨大（通常10 000像素×100 000像素），因此整个过程极为繁琐、耗时费力。且医生的判断可能因为病理切片的质量、染色等客观因素受到干扰，导致出现偏差，这大大增加了对全切片病

收稿日期：2024-09-12；**修订日期：**2024-12-02

理图像进行精准分类的挑战性[3]。近年来,数字扫描技术日益成熟,它可以将病理切片快速高质量地转换为数字图像,进而方便病理医生能够更有效地观测和管理病理数据。此外,随着深度学习在计算机视觉任务中的迅速发展,许多新技术不断涌现,并被应用到零售、安防、医疗和自动驾驶等各个行业,其中深度学习技术应用于全切片病理图像分类任务受到了很多计算机科学家以及病理学家的关注,并被认为可以更高效地对病理图像进行分类。

一方面,鉴于全切片病理图像尺寸巨大,从而造成对其所含图像块进行标注耗时且成本高,因此,弱监督多实例学习(Multi-instance learning, MIL)正成为目前对全切片病理图像进行分类的主流方法。在MIL框架下,需要首先从全切片病理图像中提取感兴趣区域图像块,之后汇总图像块分析结果以生成代表该全切片图像的包级表示,并用于下游分类任务。具体而言,Ilse等[4]首次提出了基于注意力机制的MIL框架,通过计算每个实例的权重来调整其重要性,从而提升包级分类性能。Lu等[5]提出聚类约束注意多实例学习(Clustering-constrained-attention multi-instance learning, CLAM)方法,利用聚类信息捕捉实例之间的关系,增强多实例学习的包级分类效果。然而,这些方法都依赖于传统的全局注意力机制,在处理高分辨率图像时通常面临高昂的开销。此外,这些方法在特征加权过程中受到固定注意力范围的限制,缺乏足够的灵活性,未能充分适应局部特征的变化。

另一方面,目前多实例学习方法大多针对单一尺度病理图像进行分析[4-5],而忽略了病理数据在不同层面展现出的多尺度特性。为了进一步提升多实例学习在全切片病理图像上的分类效果,研究者们开始探讨如何高效融合多尺度病理特征。例如,MS-DA-MIL[6]结合卷积神经网络(Convolutional neural network, CNN)和领域对抗学习(Domain adversarial, DA)对多尺度特征进行融合学习。MultiAttentionMIL[7]利用带有残差连接(Residual connection)的注意机制在聚合层可靠地融合多尺度特征。以上方法虽然取得了一定进展,但大多采用注意力机制线性融合多尺度信息,难以捕捉不同尺度病理信息的复杂关联。

基于此,本文提出了一种基于可变形注意力和多尺度多实例学习(Deformable attention and multi-scale multi-instance learning, DMSMIL)的全切片病理图像分类方法。一方面,设计了可变形注意力分支,通过动态调整注意力区域,灵活地选择少量关键点进行计算,优化了注意力计算效率;另一方面,提出了基于最优传输(Optimal transport, OT)的关联算法,通过最小化不同尺度图像之间的分布差异,以实现不同尺度病理图像的高效对齐,从而有效融合多尺度病理信息。实验结果表明,相较于主流的全切片病理图像分类方法,该方法在性能上得到了提升。

本文的主要贡献如下:

(1)提出了一种基于可变形注意力和多尺度多实例学习的全切片病理图像分类方法,利用可变形注意力机制学习同一尺度内不同图像块的关联,提升了注意力计算的效率。

(2)提出了一种基于最优传输的关联算法,通过最小化不同尺度图像块之间的分布差异,实现多尺度病理信息的高效对齐。

1 相关工作

1.1 基于注意力机制的多实例学习分析方法

近年来,基于注意力机制的多实例学习[4-5,8-10]分析方法在计算病理学中得到了广泛应用,尤其在全切片病理图像分析中。注意力机制能够帮助模型有效聚焦于关键区域,从而提升分析的精度和效率。例如,ABMIL方法[4]通过为每个实例分配不同的权重,优先考虑重要实例,以便在包级别上获得更准确的分类效果;TransMIL方法[8]利用Transformer的全局自注意力机制建模病理图像中多个实例间的复杂关系,以提升对整体图像的理解和分类性能;LA_MIL[9]采用基于图结构的注意力机制,结合局部和

全局关系,使得特征聚合和决策过程更加精确;Xiong等[10]等则结合层次注意力机制和图结构,优化了实例间关系,从而提升了学习效果。然而,这些方法通常依赖传统的全局注意力机制,因此在处理高分辨率图像时会面临较大的计算开销。同时,特征加权过程中固定的注意力范围限制了方法的灵活性,无法有效适应局部特征的变化。因此,本文提出了一种基于可变形注意力的分析方法,通过动态调整注意力区域,灵活地选择少量关键点进行计算以提升注意力计算的效率。

1.2　基于多尺度学习的分析方法

近年来,针对多尺度的组织病理图像分析[6-7,11-13]日益受到关注。不同尺度的病理图像能够揭示不同层次的特征信息:低尺度下可观察到全切片图像中的组织整体结构,而高尺度下则能细致地观察细胞核的形状。MultiAttentionMIL[7]结合注意力机制和多实例学习方法,将带有残差连接的注意力机制应用于模型,从而在特征聚合层有效地融合多尺度特征;ScaleFormer[11]以分层的形式提取CNN网络中的多尺度信息,并聚合多尺度特征,提升了全切片病理图像的分类性能;MDF-Net[12]采用一个两阶段端到端架构,能够较好地进行特征表示和融合;Ding等[13]提出了一种基于多尺度Transformer的全切片图像分类方法,分类精度得到一定提升。尽管这些方法在多尺度分析上有所进展,但它们通常依赖于注意力机制进行简单的线性融合,难以有效捕捉不同尺度间复杂的病理信息关系。本文设计了基于最优传输的关联算法,通过最小化不同尺度图像之间的分布差异来实现不同尺度病理图像的高效对齐,从而有效融合多尺度病理信息。

2　本文方法

如图1所示,本文所提出的DMSMIL模型由3个主要模块组成,分别为尺度间最优传输模块、尺度内可变形注意力模块和注意力融合分类模块。最优传输模块负责计算不同尺度图像块(10×和20×)

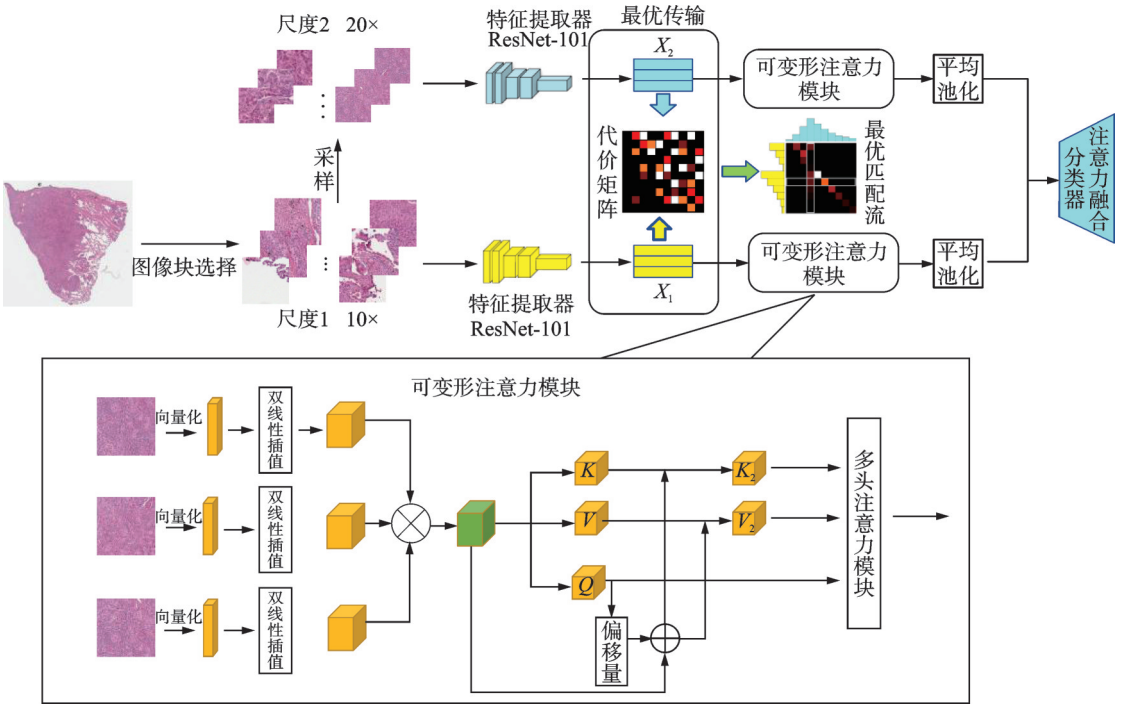

图1　DMSMIL的总体框架

Fig.1　Overall framework of DMSMIL

间的最优传输损失,以实现不同尺度下病理图像的高效对齐;可变形注意力模块通过动态调整注意力区域,灵活地选择少量关键点进行计算,以捕捉实例之间的复杂关联;注意力融合分类模块则用于融合不同尺度的病理信息,最终实现全切片病理图像的分类。

2.1 基本定义

本文定义一个包含N个全切片病理图像$W=\{W_1, W_2, \cdots, W_N\}$的分类任务,其中$Y_i$为$W_i$对应的类别标签。对于$W_i$,分别在10×和20×尺度上挑选$K$个图像块,记为$P_i^1=\{p_i^{1,1}, p_i^{1,2}, \cdots, p_i^{1,K}\}$和$P_i^2=\{p_i^{2,1}, p_i^{2,2}, \cdots, p_i^{2,K}\}$。随后,利用ResNet-101网络提取对应图像块的特征,记为$x_i^1=\{\boldsymbol{x}_i^{1,1}, \boldsymbol{x}_i^{1,2}, \cdots, \boldsymbol{x}_i^{1,K}\}$和$x_i^2=\{\boldsymbol{x}_i^{2,1}, \boldsymbol{x}_i^{2,2}, \cdots, \boldsymbol{x}_i^{2,K}\}$

2.2 数据预处理

如图1所示,DMSMIL首先在10×尺度下将全切片病理图像划分为不重叠图像块。考虑到生成的图像块中可能存在非组织区域,本文选择图像密度最高的200个图像块进一步分析。这里,图像密度定义为图像块中非白色像素(即在24位RGB色彩空间中,红、绿、蓝三色中至少有一个通道的值低于200)的百分比[14]。针对每一个从10×尺度下采集的图像块,本文在20×尺度下随机采样一个相同大小的图像块,作为另一个尺度的病理数据。

2.3 尺度间最优传输模块

考虑到不同尺度的病理图像所提取的信息差异性较大,为了充分获取不同尺度病理图像间的复杂关联,本文基于最优传输理论,利用Wasserstein距离最小化不同尺度图像之间的分布差异,以实现不同尺度病理图像的高效对齐。具体而言,参考文献[15-16]中的方法,本文定义第i个全切片病理图像在两个尺度下对应图像块特征间的Wasserstein距离为

$$D_i(x_i^1, x_i^2)=\inf_{\gamma\in\Gamma(Q_1, Q_2)} E_{(\boldsymbol{x}_i^{1,j}, \boldsymbol{x}_i^{2,j})\in\Gamma}\left[\| \boldsymbol{x}_i^{1,j}-\boldsymbol{x}_i^{2,j} \|\right] \tag{1}$$

式中:x_i^1和x_i^2分别表示10×和20×尺度下所有图像块的特征;Q_1和Q_2表示不同尺度下特征对应的边缘分布,$\Gamma(Q_1, Q_2)$表示其联合分布;$\| \boldsymbol{x}_i^{1,j}-\boldsymbol{x}_i^{2,j} \|$表示在每个可能的联合分布$\Gamma$中,计算出不同尺度图像块特征间的距离。针对$N$张全切片病理图像特征,最优传输损失可以定义为

$$L_{\text{ot}}=\frac{1}{N}\sum_{i=1}^{N} D_i(x_i^1, x_i^2) \tag{2}$$

2.4 尺度内可变形注意力模块

针对传统注意力计算方式需要考虑所有的空间位置,导致计算效率低下的问题,本文采用可变形注意力模块[17-18]计算同一尺度下不同图像块之间的关联。可变形注意力模块通过动态调整注意力区域,灵活地选择少量关键点进行计算,进而显著提升注意力计算的效率。

具体的可变形注意力模块如图2所示。对于尺度s下病理图像W_i的第j个图像块的特征$\boldsymbol{x}_i^{s,j}\in\mathbf{R}^C$,经过重塑和插值操作后得到对应的特征图$\tilde{\boldsymbol{x}}_i^{s,j}\in\mathbf{R}^{H\times W\times C}$,其中$H$和$W$分别表示特征图的高度和宽度。基于$\tilde{\boldsymbol{x}}_i^{s,j}$,生成一个统一网格$\boldsymbol{O}\in\mathbf{R}^{H\times W\times 2}$作为参考,在本文中,设置网格尺寸与特征图尺寸一致。接着,本文从网格中选取参考点r_q,并将其坐标归一化到[−1, +1]的范围中。为了获得每个参考点的偏移量和注意力权重,本文首先计算每个特征图的查询向量,即$\boldsymbol{z}_q=W_{\text{query}}\tilde{\boldsymbol{x}}_i^{s,j}$,然后将其输入到$T=3\times M\times U$通道的线性层中,其中$M$表示注意力头数,$U$表示每一个注意力头中偏移的方向数。本文应用前$2\times M\times U$通道计算编码采样偏移量$\Delta r_{mqu}$,剩余的$M\times U$通道用于计算输入到Softmax算子后的注意力权重A_{mqu},其可以表示为

$$\Delta r_{mqu}=\text{Linear}_{2MU}(\boldsymbol{z}_q) \tag{3}$$

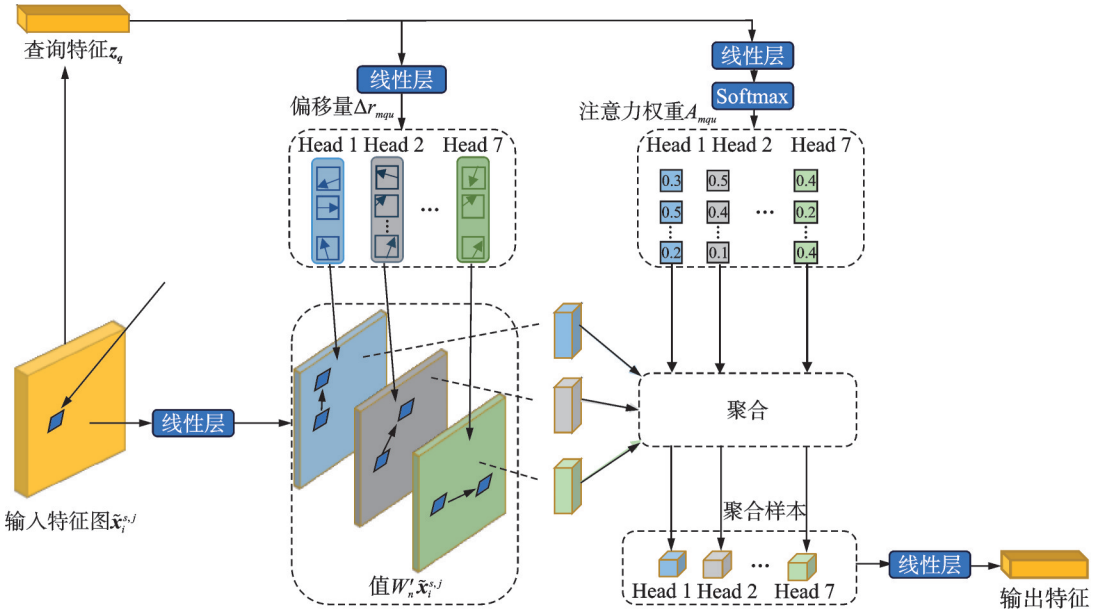

图2　可变形注意力模块

Fig.2　Deformable attention module

$$A_{mqu}=\mathrm{Softmax}(\mathrm{Linear}_{MU}(z_q)) \tag{4}$$

为了稳定训练过程,本文引入预定义因子c来缩放Δr_{mqu},以防止偏移量太大,具体计算为$\Delta r_{mqu}=c\cdot\tanh(\Delta r_{mqu})$。随后,在变形点的位置$(r_q+\Delta r_{mqu})$对偏移后的特征$\bar{x}_i^{s,j}$进行采样,以获得键和值。考虑到$(r_q+\Delta r_{mqu})$不一定为整数,进一步利用文献[19]中提到的双线性插值方法来确定其所在网格点的位置。具体表示为

$$\bar{x}_i^{s,j}=\tilde{x}_i^{s,j}(r_q+\Delta r_{mqu}) \tag{5}$$

$$\bar{z}_{\mathrm{k}}=W_{\mathrm{k}}\bar{x}_i^{s,j},\bar{z}_{\mathrm{v}}=W_{\mathrm{v}}\bar{x}_i^{s,j} \tag{6}$$

式中$\bar{z}_{\mathrm{k}}$和$\bar{z}_{\mathrm{v}}$分别表示经过变形注意力计算后得到的键和值的编码。

在获取偏移量Δr_{mqu}和注意力权重A_{mqu}后,进一步计算加权特征。首先,偏移后的特征$\bar{x}_i^{s,j}$经过线性映射W_m'后得到$\vec{x}_i^{s,j}$;接着,利用注意力权重A_{mqu}对$\vec{x}_i^{s,j}$进行加权求和,聚合得到样本特征;最后,经过线性层W_m,得到最终的输出特征$\hat{x}_i^{s,j}$。具体可表示为

$$\vec{x}_i^{s,j}=W_m'\bar{x}_i^{s,j}=W_m'\tilde{x}_i^{s,j}(r_q+\Delta r_{mqu}) \tag{7}$$

$$\hat{x}_i^{s,j}=\mathrm{DeformAttn}(z_q,r_q,\tilde{x}_i^{s,j})=\sum_{m=1}^{M}W_m\left[\sum_{u=1}^{U}A_{mqu}W_m'\tilde{x}_i^{s,j}(r_q+\Delta r_{mqu})\right] \tag{8}$$

式中:M为总的注意力头数;U为总的采样键数,即每个注意力头中偏移方向的数量;W_m为第m个注意力头的权重矩阵。

相较于传统注意力机制需要对特征图中所有位置间的关系进行计算,其计算复杂度为$O(H^2\cdot W^2\cdot C)$,其中H和W分别为特征图的高度和宽度,C为通道数。可变形注意力机制通过仅在偏移后的局部

区域内进行计算，从而将计算复杂度降低至$O(M\cdot U\cdot C\cdot L^2)$，其中$M$为注意力头数，$U$为偏移方向数，$L$为局部区域的边长。由于$U$、$M$、$L$均远小于$H$和$W$，故可变形注意力机制显著减少了计算量，提升了计算效率。

2.5 注意力融合分类模块

针对从全切片病理图像W_i提取的不同尺度的特征，本文首先通过均值池化操作对相同尺度的特征进行聚合，得到$\hat{\boldsymbol{x}}_i^1$和$\hat{\boldsymbol{x}}_i^2$，表示为

$$\hat{\boldsymbol{x}}_i^1=\frac{1}{K}\sum_{j=1}^{K}\hat{\boldsymbol{x}}_i^{1,j},\ \hat{\boldsymbol{x}}_i^2=\frac{1}{K}\sum_{j=1}^{K}\hat{\boldsymbol{x}}_i^{2,j} \tag{9}$$

接着，针对聚合后的特征，采用基于注意力机制的多尺度特征融合方法[20]进行癌症亚型分类。该方法对不同尺度下图像特征分配不同的权重，相较于传统的最大池化或平均池化策略更为灵活。具体而言，本文对两个尺度下的图像特征$\hat{\boldsymbol{x}}_i^1$和$\hat{\boldsymbol{x}}_i^2$进行加权求和，获得其联合表示为

$$g(\hat{\boldsymbol{x}}_i)=\sum_{s=1}^{S}a_i^s\hat{\boldsymbol{x}}_i^s \tag{10}$$

其中

$$a_i^s=\frac{\exp\left\{\boldsymbol{\omega}^{\mathrm{T}}\tanh(\hat{\boldsymbol{x}}_i^s)\right\}}{\sum_{s=1}^{S}\exp\left\{\boldsymbol{\omega}^{\mathrm{T}}\tanh(\hat{\boldsymbol{x}}_i^s)\right\}} \tag{11}$$

式中$\boldsymbol{\omega}$为可学习的参数。

对于融合后的特征$g(\hat{\boldsymbol{x}}_i)$，将其送入分类器进行亚型分类，并获得交叉熵损失$L_{\mathrm{ce}}$，表示为

$$y_i=\mathrm{Classifier}(g(\hat{\boldsymbol{x}}_i)) \tag{12}$$

$$L_{\mathrm{ce}}=-\frac{1}{N}\sum_{i=1}^{N}[y_i\log\alpha_i+(1-y_i)\log(1-\alpha_i)] \tag{13}$$

2.6 总损失

结合最优传输损失（式(2)）以及交叉熵分类损失（式(13)），本文提出的DMSMIL方法的目标函数为

$$L_{\mathrm{train}}=L_{\mathrm{ot}}+\lambda L_{\mathrm{ce}} \tag{14}$$

式中λ为正则化参数。通过引入λ，本文能够在优化过程中平衡最优传输损失与交叉熵分类损失的贡献，从而实现分类精度与样本间关系建模的统一优化。

为了验证λ的作用，本文通过实验研究了不同λ值对模型性能的影响。最终，选择$\lambda=1$作为最优正则化参数，以确保模型在分类精度和鲁棒性方面均能达到最佳表现。不同λ值在两个数据集上的分类结果（包括ACC、F_1评分、AUC这3个评价指标）如图3所示。从图3中可以看出，不同的λ对模型在TCGA-BRCA和TCGA-NSCLC数据集上的分类性能具有显著影响。当$\lambda=1$时，3个评价指标均达到了最高值。这表明，设置$\lambda=1$能够有效提升模型泛化能力与鲁棒性，为后续实验提供了优化依据。

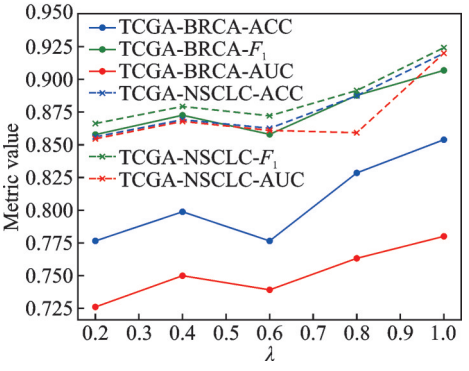

图3 不同λ在TCGA-BRCA和TCGA-NSCLC数据集上的结果图

Fig.3 Results of different λ on TCGA-BRCA and TCGA-NSCLC datasets

3　实验分析与结果

3.1　数据集和实验设置

本文在全切片病理图像亚型分类任务上评估所提出模型的性能,主要使用了TCGA公共数据库中的两个数据集(TCGA-BRCA和TCGA-NSCLC)。TCGA-BRCA为乳腺癌数据集,共包含546张全切片病理图像,主要进行对乳腺浸润性癌(426张)和乳腺小叶癌(120张)两种亚型的分类任务;TCGA-NSCLC为肺癌数据集,共包含790张全切片病理图像,主要用于肺腺癌(385张)和肺鳞状细胞癌(405张)两种亚型的分类。对于每一张全切片病理图像,本文在10×和20×这两个尺度下分别提取尺寸为512×512的200个图像块,并利用ResNet-101网络对其进行特征提取。对于本文提出的方法和所有对比实验,本文将输入图像统一调整为与512×512图像块相同的尺寸,设置每个实验的输入图像块训练轮数为120,并利用Adam优化器更新模型参数,学习率固定为1×10^{-4},以确保实验的一致性。为了评估所提方法的有效性,本文基于准确率(ACC)、F_1分数和AUC值对不同方法进行比较。所有实验均在Pytorch2.1环境下完成,并在Nvidia RTX3060GPU上进行训练。

为了更全面地评估本文提出的方法,选择了近年来在病理图像分类任务中取得较好效果的单尺度和多尺度病理图像分类方法进行比较,其中单尺度方法包括:(1)ABMIL[4],基于注意力机制的多实例学习方法;(2) TransMIL[8],基于Transformer结构的多实例学习方法;(3) CLAM[5],基于聚类约束注意力的多实例学习方法;(4)DTFD-MIL[21],结合注意力机制和双层特征蒸馏机制的多实例学习方法;(5) LA-MIL[9],基于局部和全局注意力的多实例学习方法;(6) HAG-MIL[10],基于分层注意力引导的多实例学习方法。多尺度方法包括:(1) DSMIL[22],基于金字塔融合机制的双流多实例学习方法;(2) MultiAttnMIL[7],结合注意力机制与残差结构,融合多尺度信息的多实例学习方法;(3) HiFuse[23],融合空间注意力和通道注意力的多尺度特征融合学习方法;(4) AMGCMN[24],融合注意力多跳图和多尺度卷积的融合学习方法。

3.2　实验结果

3.2.1　单尺度学习方法比较

首先,在单一尺度(10×、20×)下,将所提出的基于可变形注意力的方法(DAMIL_10×、DAMIL_20×)与现有方法进行比较,实验结果如图4和图5所示。从图4和图5中可以看出,本文提出的基于可变形注意力和多实例学习的全切片病理图像分类方法在两个数据集上均有一定优势。与性能次优的

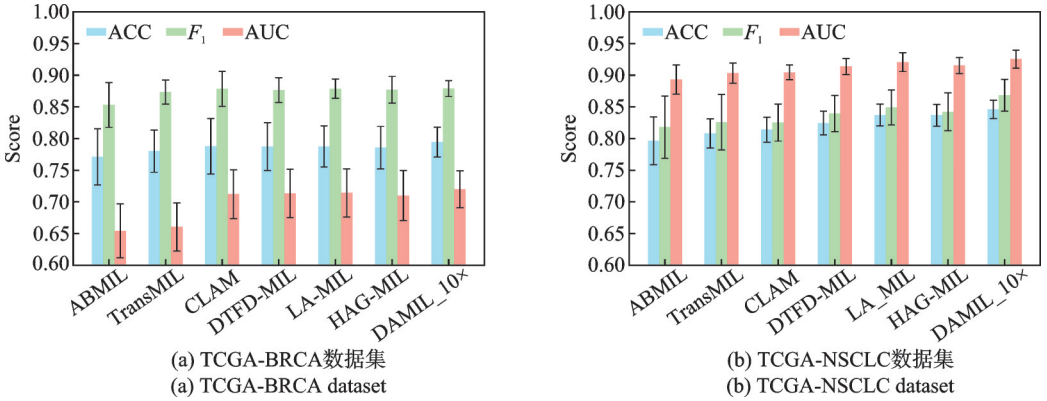

图4　单尺度方法在TCGA-BRCA和TCGA-NSCLC数据集10×尺度上的对比实验结果

Fig.4　Comparison of experimental results for single-scale (10×) methods on TCGA-BRCA and TCGA-NSCLC datasets

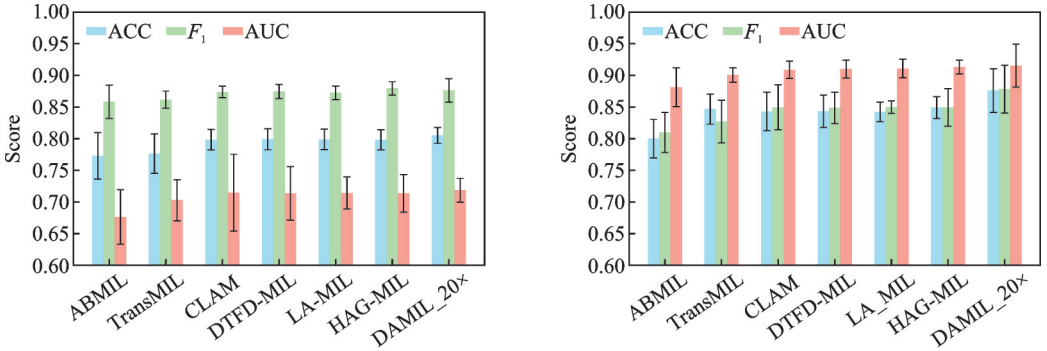

图5 单尺度方法在TCGA-BRCA和TCGA-NSCLC数据集20×尺度上对比实验结果

Fig.5 Comparison of experimental results for single-scale (20×) methods on TCGA-BRCA and TCGA-NSCLC datasets

模型相比,所提出的方法在10×尺度上的准确率提高了约3%,在20×尺度上的准确率提高了约4%。这一提升主要归因于可变形注意力模块能够精确地选择关键点进行学习,并有效地融合分类信息。

3.2.2 多尺度学习方法比较

为了更全面地评估所提出的DMSMIL方法的性能,本文将其与现有的多尺度学习方法进行了对比,实验结果如表1和表2所示。从表1和表2中可以看出,DMSMIL在ACC、F_1和AUC这3项指标上均显著优于现有方法。具体而言,与性能次优的模型相比,DMSMIL在TAGC-BRCA数据集上的分类准确率提高了超过4%,在TCGA-NSCLC数据集上的分类准确率提高了超过6%。这一实验结果进一步验证了所提出的可变形注意力机制和尺度间最优传输模块的有效性。此外,从表中还观察到,DMSMIL总体在各项指标上都体现出了较低的标准差,表明该模型相较于对比方法具有更强的泛化能力。

表1 基于多尺度方法的TCGA-BRCA数据集亚型分类结果

Table 1 Subtype classification results of TCGA-BRCA dataset based on multi-scale methods

方法	ACC	F_1	AUC
DSMIL	0.796 7±0.026 9	0.867 9±0.031 9	0.701 4±0.025 4
MultiAttnMIL	0.805 9±0.038 2	0.876 3±0.026 1	0.701 1±0.055 5
HiFuse	0.809 8±0.027 2	0.877 4±0.022 9	0.710 7±0.022 5
AMGCFN	0.810 8±0.015 9	0.880 8±0.010 8	0.703 5±0.016 9
DMSMIL	**0.853 9±0.011 2**	**0.906 9±0.006 0**	**0.780 0±0.018 6**

表2 基于多尺度方法的TCGA-NSCLC数据集亚型分类结果

Table 2 Subtype classification results of TCGA-NSCLC dataset based on multi-scale methods

方法	ACC	F_1	AUC
DSMIL	0.856 8±0.034 6	0.852 8±0.046 4	0.889 3±0.035 4
MultiAttnMIL	0.858 1±0.042 9	0.854 6±0.045 0	0.847 2±0.042 1
HiFuse	0.859 9±0.032 1	0.850 5±0.044 2	0.889 1±0.024 3
AMGCFN	0.857 2±0.035 8	0.858 2±0.048 9	0.897 0±0017 9
DMSMIL	**0.920 0±0.015 9**	**0.924 3±0.013 6**	**0.920 0±0.016 5**

3.2.3　可解释性分析

本文还对所提出的方法进行了可解释性分析。具体而言,针对TCGA-BRCA和TCGA-NSCLC数据集中不同类型的全切片病理图像,将其裁剪为图像块并输入到所提出的深度神经网络中,计算每个图像块的归一化注意力得分,结果分别如图6和图7所示。从图6和图7可以看出,不同类别的全景病理切片中各个图像块的注意力得分存在较大差异。具体而言,在乳腺小叶癌切片中,各个图像块对应的注意力得分较低,通过分析得分最低的几个图像块,发现其对应浸润淋巴细胞区域,这与乳腺浸性小叶癌淋巴浸润程度较高的临床病理特征一致[25]。与之相反,乳腺浸润性癌切片中的图像块注意力得分较高,而这些高分图像块通常对应上皮细胞区域,这正是乳腺浸润性细胞癌的多发区域[26]。此外,通过对肺腺癌和肺鳞状细胞癌全景病理切片的注意力得分结果进行可视化,发现肺鳞状细胞癌的图像块具有较高的注意力得分,且这些高得分图像块的细胞核面积大于肺腺癌图像块细胞核面积,这与肺鳞状细胞癌细胞核相较于其他类型癌症细胞核大的特点一致[27]。综上所述,本文所提出的方法不仅能够有效区分不同类型的全切片病理图像,还具有一定的医学可解释性。

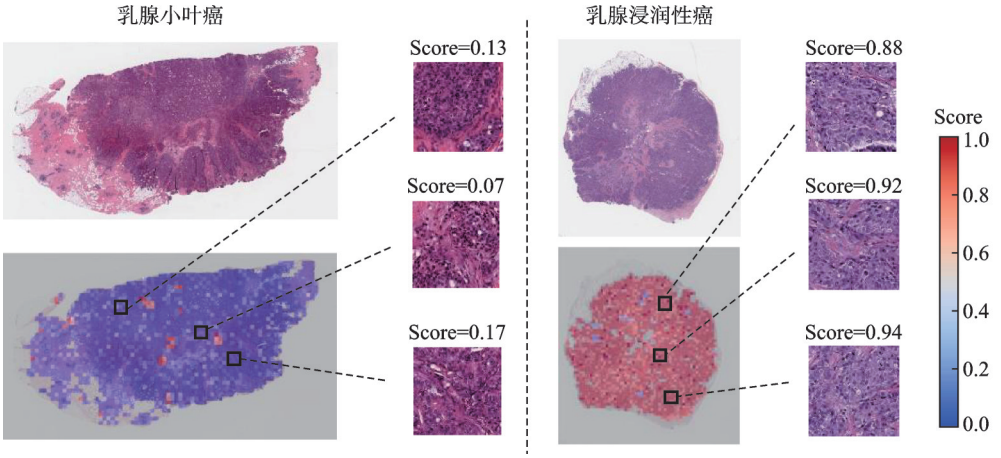

图6　TCGA-BRCA数据集上基于DMSMIL分类方法的注意力得分可视化结果

Fig.6　Attention scores based visualization results of the DMSMIL method on TCGA-BRCA dataset

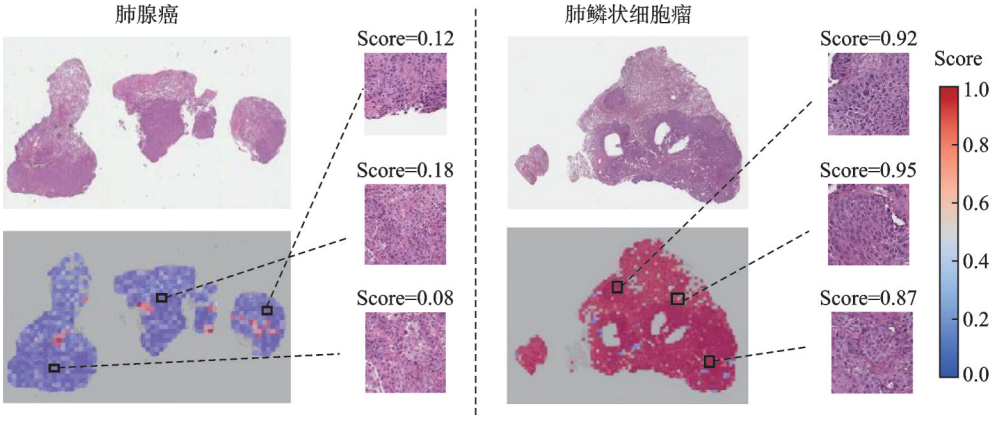

图7　TCGA-NSCLC数据集上基于DMSMIL分类方法的注意力得分可视化结果

Fig.7　Attention scores based visualization results of the DMSMIL method on TCGA-NSCLC dataset

3.2.4　方法效率分析

为了评估DMSMIL的计算效率,本文从参数数量、训练时间以及推理时间等多个维度上将其与对

比方法进行比较,具体结果如表3所示。从表3可以看出,所提出的DMSMIL方法参数量较少,因此在模型训练和推理效率上相较于现有方法有一定的提升。具体而言,本文所提方法每一轮的训练时间为4.05 s,推理时间为0.007 2 s,相较于对比方法的计算效率提升了至少2倍。这主要归因于可变形注意力模块能够通过动态调整注意力区域,灵活地选择少量关键点进行计算,从而提高了计算效率。结合表2的实验结果可以看出,不仅所提方法的分类精度有所提升,而且其计算时间也显著降低,因此它在临床任务中具有更好的适用性。

表3 各多尺度融合方法中模型的参数量和时间成本比较

Table 3 Comparison of model parameters and time cost in various multi-scale fusion methods

方法	参数量	训练时间/s	推理时间/s
DSMIL	1 894 072	13.72	0.014 1
MultiAttnMIL	854 789	9.68	0.009 6
HiFuse	3 596 964	27.16	0.029 8
AMGCFN	1 065 247	19.83	0.017 9
DMSMIL	**612 096**	**4.05**	**0.007 2**

3.3 消融实验

本文还进行了系统的消融实验,以验证方法中各个关键模块的有效性。实验内容涵盖了对多尺度融合模块、尺度间最优传输模块以及尺度内可变形注意力模块的逐一评估。在TCGA-BRCA和TCGA-NSCLC数据集上的消融实验结果分别如表4和表5所示。实验结果表明,随着多尺度融合策略、尺度间最优传输模块以及尺度内可变形注意力模块的逐步引入,模型整体的分类准确率也随之提升。这些结果充分证明了所引入模块在多尺度信息融合和分类中的重要作用。

表4 在TCGA-BRCA数据集上的消融实验结果

Table 4 Ablation experimental results on TCGA-BRCA dataset

多尺度融合策略	尺度间最优传输模块	尺度内可变形注意力模块	准确率±标准差
—	—	—	0.798 5±0.031 6
—	—	√	0.805 1±0.012 5
√	—	—	0.805 6±0.032 4
√	√	—	0.815 3±0.027 5
√	—	√	0.845 1±0.025 1
√	√	√	**0.853 9±0.011 2**

注:“√”表示采用该模块,“—”表示不使用该模块。

表5 在TCGA-NSCLC数据集上的消融实验结果

Table 5 Ablation experimental results on TCGA-NSCLC dataset

多尺度融合策略	尺度间最优传输模块	尺度内可变形注意力模块	准确率±标准差
—	—	—	0.848 1±0.020 8
—	—	√	0.876 0±0.034 6
√	—	—	0.885 7±0.021 0
√	√	—	0.889 7±0.019 4
√	—	√	0.902 0±0.018 5
√	√	√	**0.920 0±0.015 9**

注:“√”表示采用该模块,“—”表示不使用该模块。

4 结束语

本文提出了一种基于可变形注意力和多尺度多实例学习的全切片病理图像分类方法(DMSMIL),旨在解决全切片病理图像的亚型分类问题。在相同尺度内,DMSMIL通过可变形注意力模块动态调整注意力区域,灵活地选择少量关键点进行计算,从而优化了注意力计算效率。此外,DMSMIL设计了一种基于最优传输的关联算法,以最小化不同尺度图像间的分布差异,实现了高效的尺度间对齐。最终,DMSMIL引入注意力融合分类模块,显著提升了全切片病理图像的分类性能。通过在两个公共癌症数据集上的对比实验和消融实验,验证了所提方法的有效性。

参考文献:

[1] BRAY F, LAVERSANNE M, SUNG H, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries[J]. CA: A Cancer Journal for Clinicians, 2024, 74(3): 229-263.

[2] 杨印凯,万鹏,石航,等.基于多模态超声对比学习的肝癌诊断方法[J].数据采集与处理,2024,39(4): 874-885.
YANG Yinkai, WAN Peng, SHI Hang, et al. Liver cancer diagnosis method based on multi-modal ultrasound contrast learning [J]. Journal of Data Acquisition and Processing, 2024, 39(4): 874-885.

[3] CAI H, FENG X, YIN R, et al. MIST: Multiple instance learning network based on Swin Transformer for whole slide image classification of colorectal adenomas[J]. The Journal of Pathology, 2023, 259(2): 125-135.

[4] ILSE M, TOMCZAK J, WELLING M. Attention-based deep multiple instance learning[C]//Proceedings of International Conference on Machine Learning. [S.l.]: PMLR, 2018: 2127-2136.

[5] LU M Y, WILLIAMSON D F K, CHEN T Y, et al. Data-efficient and weakly supervised computational pathology on whole-slide images[J]. Nature Biomedical Engineering, 2021, 5(6): 555-570.

[6] HASHIMOTO N, FUKUSHIMA D, KOGA R, et al. Multi-scale domain-adversarial multiple-instance CNN for cancer subtype classification with unannotated histopathological images[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2020: 3852-3861.

[7] WIBAWA M S, LO K W, YOUNG L S, et al. Multi-scale attention-based multiple instance learning for classification of multi-gigapixel histology images[C]//Proceedings of European Conference on Computer Vision. Cham: Springer Nature Switzerland, 2022: 635-647.

[8] SHAO Z, BIAN H, CHEN Y, et al. TransMIL: Transformer based correlated multiple instance learning for whole slide image classification[J]. Advances in Neural Information Processing Systems, 2021, 34: 2136-2147.

[9] WAGNER S J, REISENBÜCHLER D, WEST N P, et al. Transformer-based biomarker prediction from colorectal cancer histology: A large-scale multicentric study[J]. Cancer Cell, 2023, 41(9): 1650-1661.

[10] XIONG C, CHEN H, SUNG J J Y, et al. Diagnose like a pathologist: Transformer-enabled hierarchical attention-guided multiple instance learning for whole slide image classification[EB/OL]. (2023-01-19). https://doi. org/10.48550/arXiv. 2301.08125.

[11] HUANG H, XIE S, LIN L, et al. ScaleFormer: Revisiting the Transformer-based backbones from a scale-wise perspective for medical image segmentation[EB/OL]. (2022-07-29). https://doi.org/10.48550/arXiv.2207.14552.

[12] QI W, WU H C, CHAN S C. MDF-Net: A multi-scale dynamic fusion network for breast tumor segmentation of ultrasound images[J]. IEEE Transactions on Image Processing, 2023, 32: 4842-4855.

[13] DING S, WANG J, LI J, et al. Multi-scale prototypical Transformer for whole slide image classification[C]//Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham: Springer Nature Switzerland, 2023: 602-611.

[14] YU K H, ZHANG C, BERRY G J, et al. Predicting non-small cell lung cancer prognosis by fully automated microscopic pathology image features[J]. Nature Communications, 2016, 7(1): 12474.

[15] YANG Y, GU X, SUN J. Prototypical partial optimal transport for universal domain adaptation[C]//Proceedings of the AAAI

Conference on Artificial Intelligence. [S.l.]: AAAI, 2023, 37(9): 10852-10860.

[16] LUO D, XU H, CARIN L. Differentiable hierarchical optimal transport for robust multi-view learning[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 45(6): 7293-7307.

[17] KIM S, AHN D, KO B C. Cross-modal learning with 3D deformable attention for action recognition[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. [S.l.]: IEEE, 2023: 10265-10275.

[18] LUO J, REN W, GAO X, et al. Multi-exposure image fusion via deformable self-attention[J]. IEEE Transactions on Image Processing, 2023, 32: 1529-1540.

[19] KIRKLAND E J. Advanced computing in electron microscopy[M]. New York: Plenum Press, 1998.

[20] SHAO W, ZUO Y, SHI Y, et al. Characterizing the survival-associated interactions between tumor-infiltrating lymphocytes and tumors from pathological images and multi-omics data[J]. IEEE Transactions on Medical Imaging, 2023, 42(10): 3025-3035.

[21] ZHANG H, MENG Y, ZHAO Y, et al. DTFD-MIL: Double-tier feature distillation multiple instance learning for histopathology whole slide image classification[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2022: 18802-18812.

[22] LI B, LI Y, ELICEIRI K W. Dual-stream multiple instance learning network for whole slide image classification with self-supervised contrastive learning[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2021: 14318-14328.

[23] HUO X, SUN G, TIAN S, et al. HiFuse: Hierarchical multi-scale feature fusion network for medical image classification[J]. Biomedical Signal Processing and Control, 2024, 87: 105534.

[24] ZHOU H, LUO F, ZHUANG H, et al. Attention multihop graph and multiscale convolutional fusion network for hyperspectral image classification[J]. IEEE Transactions on Geoscience and Remote Sensing, 2023, 61: 1-14.

[25] 樊紫瑜,房煊,张晟.乳腺浸润性小叶癌的临床病理特征、诊疗现状及展望[J].中国全科医学,2021,24(30): 3806-3813,3820.
FAN Ziyu, FANG Xuan, ZHANG Sheng. Clinicopathological features, diagnosis and treatment status, and prospects of invasive lobular carcinoma of the breast[J]. Chinese General Practice, 2021, 24(30): 3806-3813, 3820.

[26] 张世超,陆苏,赵紫薇,等.乳腺浸润性微乳头状癌的临床病理特征及预后分析[J].中国肿瘤临床,2020,47(2): 77-81.
ZHANG Shichao, LU Su, ZHAO Ziwei, et al. Clinicopathological features and prognostic analysis of invasive micropapillary carcinoma of the breast[J]. Chinese Journal of Clinical Oncology, 2020, 47(2): 77-81.

[27] FERRÍS L B, PÜTTMANN S, MARINI N, et al. A full pipeline to analyze lung histopathology images[C]//Proceedings of Medical Imaging 2024: Digital and Computational Pathology. [S.l.]: SPIE, 2024, 12933: 1293303.

作者简介:

薛保(2001-),男,硕士研究生,研究方向:病理图像分析,E-mail: 1359647020@qq.com。

周俊杰(2000-),男,博士研究生,研究方向:病理染色生成、计算病理学。

邵伟(1986-),**通信作者**,男,副教授,研究方向:机器学习、医学图像处理,E-mail: shaowei20022005@nuaa.edu.cn。

(编辑:王静)

Whole Slide Pathology Image Classification Method Based on Deformable Attention and Multi-scale Multi-instance Learning

XUE Bao, ZHOU Junjie, SHAO Wei*

(Key Laboratory of Brain-Machine Intelligence Technology, Ministry of Education, College of Artificial Intelligence, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China)

Abstract: Whole slide images (WSIs) serve as the golden standard for pathological diagnosis, and their accurate classification provides critical information on tumor type, grade, and stage, which is essential for cancer prognosis and treatment strategy selection. In computational pathology, multi-instance learning (MIL) has become the mainstream approach for WSI classification. However, most existing MIL methods focus on single-scale pathological images, limiting the understanding of cancer development and progression mechanisms across different levels. Additionally, the high resolution of WSIs and information discrepancies across scales pose challenges to efficiently integrating and analyzing patches both within a single scale and across multiple scales. To address these issues, this paper proposes a WSI classification method based on deformable attention and multi-scale multi-instance learning (DMSMIL). Specifically, a deformable attention branch is designed to learn associations among patches within the same scale, enhancing attention computation efficiency. Meanwhile, an optimal transport (OT)-based association algorithm is developed to integrate pathological information across different scales, enabling efficient multi-scale information alignment. Experimental results on breast cancer and lung cancer subtype classification tasks demonstrate that the proposed method achieves classification accuracies of 85.39% and 92.00%, respectively, outperforming mainstream WSI classification methods. The proposed DMSMIL effectively integrates multi-scale pathological features and improves the accuracy of WSI-based cancer subtype classification, providing a promising approach for computational pathological diagnosis.

Highlights:

1. Propose a novel DMSMIL framework for WSI classification, integrating deformable attention and multi-scale MIL to address single-scale limitations.
2. Design a deformable attention branch to enhance intra-scale patch association learning and attention computation efficiency.
3. Develop an OT-based association algorithm for efficient multi-scale pathological information integration and alignment.
4. Achieve 85.39% and 92.00% accuracy on breast and lung cancer subtype classification, outperforming mainstream methods.

Key words: multi-instance learning (MIL); deformable attention; multi-scale learning; whole slide pathology image classification; optimal transport (OT)

Received: 2024-09-12; **Revised:** 2024-12-02

***Corresponding author**, **E-mail:** shaowei2002005@nuaa.edu.cn.