

中图法分类号: TP391.4 文献标识码: A 文章编号: 1006-8961(2025)11-3680-14

论文引用格式: Ding W L, Deng R R, Xu L F, Wang C N, Zhu X Q and Zheng K. 2025. MoLiNet: a local and global information interactive fusion network for improving multi-classification of pathological image artifacts. Journal of Image and Graphics, 30(11):3680-3693(丁维龙, 邓戎戎, 徐利锋, 汪春年, 祝行琴, 郑魁. 2025. 改善病理图像伪影多分类的局部与全局信息交互融合网络 MoLiNet. 中国图象图形学报, 30(11):3680-3693)[DOI:10.11834/jig.240691]

改善病理图像伪影多分类的局部与全局信息交互融合网络 MoLiNet

丁维龙¹, 邓戎戎¹, 徐利锋¹, 汪春年^{2*}, 祝行琴², 郑魁³

1. 浙江工业大学计算机科学技术学院, 杭州 310023; 2. 宁波市临床病理诊断中心, 宁波 315021;

3. 上海派影医疗科技有限公司, 上海 201306

摘要: 目的 在组织病理切片制备过程中, 会不可避免地产生伪影, 可能严重影响病理医生诊断准确性及计算机辅助诊断系统效能。现有伪影分类算法在处理相似伪影时, 因缺乏全局分布规律学习能力, 导致分类准确度并不理想。为此, 提出一种改善病理图像伪影多分类的局部与全局信息交互融合网络 MoLiNet (mobile linear net)。方法 首先, 构建了一个基于 HSV 颜色域的边缘检测模块以增强对裂缝的关注。然后, 在双分支架构的基础上, 结合门控线性单元 (gated linear unit, GLU) 与移动倒残差瓶颈卷积模块 (mobile inverted bottleneck convolution, MBConv) 构建出一种新型前馈层, 以自适应地筛选出关键特征, 同时提高整体计算效率。此外, 还提出线性聚焦交叉注意力, 以对全局特征和局部特征进行充分交互, 从而增强特征表达能力。**结果** 本文提出的网络在分类精度和计算效率方面均超越同类先进方法。在 NCPDCSAD (ningbo clinical pathology diagnosis center similar artifact dataset) 数据集测试中, 本文模型取得了 94.24% 准确率 (accuracy), 超越了专门用于伪影分类的 DKL 网络。另外, 本文构建的网络在参数量 (Params) 和计算量 (Flops) 上相较于 DKL (deep kernel learning) 分别减少了 80.91% 和 96.94%。通过热力图可视化分析, 进一步验证了本文提出的伪影分类网络对伪影特征具有敏锐捕捉能力。**结论** 本文方法不仅能够有效地区分组织病理切片中相似伪影, 而且在计算资源消耗上实现了显著降低。这一研究成果为病理图像质量评估领域带来了一种有效解决方案。

关键词: 数字病理切片; 伪影分类; 注意力机制; 特征融合; 轻量级

MoLiNet: a local and global information interactive fusion network for improving multi-classification of pathological image artifacts

Ding Weilong¹, Deng Rongrong¹, Xu Lifeng¹, Wang Chunlian^{2*}, Zhu Xingqin², Zheng Kui³

1. College of Computer Science and Technology, Zhejiang University of Technology, Hangzhou 310023, China;

2. Ningbo Diagnostic Pathology Center, Ningbo 315021, China; 3. Shanghai Paiying Medical Science and Technology Limited Company, Shanghai 201306, China

Abstract: Objective The analysis of histopathological images provides essential pathological information for tumor diagnosis and treatment, and it is widely regarded by experts as the gold standard for tumor diagnosis. Prior to digitization, histopathological images undergo a complex, multi-step process that includes sampling, fixation, dehydration, embedding, sec-

收稿日期: 2024-11-21; 修回日期: 2025-03-14; 预印本日期: 2025-03-21

* 通信作者: 汪春年 niana217@126.com

基金项目: 浙江省基础公益研究计划项目 (LTGY24F020001, TGY23F020005)

Supported by: Zhejiang Public Welfare Technology Research Plan Project (LTGY24F020001, TGY23F020005)

tioning, staining, and mounting. This process is labor intensive, has limited automation, and is subject to the subjective actions of pathologists, which can introduce artifacts. These artifacts may obscure normal tissue and complicate the assessment of true lesions, which leads to potential misdiagnoses or missed diagnoses that can impact patient treatment and prognosis. Furthermore, artifacts in histopathological images can hinder the performance of computer-aided diagnostic systems, which underscores the importance of effective artifact detection in finalized images. Deep learning methods for identifying and differentiating artifacts with prominent morphological and color distinctions, such as ink marks, tissue folds, and blurriness, have shown promising results. However, certain artifact types, such as looseness and tissue bubbles, are visually similar, which makes them particularly challenging to differentiate. Effective differentiation requires models that can maintain sensitivity to local features while also achieving a comprehensive understanding of global patterns and structures, which ensures the robust integration of local and global information. However, current methods for artifact detection have not fully addressed the challenges posed by these visually similar artifacts, which indicates an urgent need for further investigation in this area. This study aims to address the limitations of existing methods in integrating and interacting with local and global information for detecting highly similar artifacts. Thus, this study proposes a local and global information interactive fusion network MoLiNet (mobile linear net) to improve the multi-classification of pathological image artifacts.

Method In pathology, histopathological images are generally stored as whole-slide images (WSIs) organized in a pyramid structure. Researchers typically preprocess WSIs by segmenting them into smaller image patches. After patch segmentation, we further process these patches using an HSV color space-based edge detection module, which is called HSVEF. HSVEF is designed to improve the attention of the network to fissures. In this work, we extend the concept of a parallel framework by employing a dual-branch network. The CNN branch utilizes the lightweight CNN MobileNetV3, which enables efficient feature extraction while enhancing the capacity of the network to capture local details. As image patches pass through this branch, features are progressively refined from low- to higher-level representations. The Transformer branch is constructed using MGL-Trans (mobile gated linear Transformer), which takes dimensionally compressed image features as input. Within the MGL-Trans module, the conventional feedforward layer is replaced by the mobile gated linear unit (MobileGLU), which leverages a gated linear unit (GLU). The MobileGLU design enables the model to selectively amplify or suppress features at different stages, which enhances its capability to recognize complex image patterns. A focused linear cross-attention (FLCA) mechanism is incorporated between the two branches, which enables effective interaction between local and global features. In our model, this attention mechanism is concretely implemented as the feature gather module and the feature diffusion module. Finally, the features extracted from both branches are concatenated and fed into a classification head for categorization. This architecture not only enriches the overall understanding of image features by the model but also enhances its performance in complex image analysis tasks. **Result** The proposed network surpasses comparable state-of-the-art methods in classification accuracy and computational efficiency. On the Ningbo Clinical Pathology Diagnosis Center Similar Artifact Dataset, our model achieves accuracy, precision, recall, and F1-score values of 0.942 4, 0.947 7, 0.949 8, and 0.948 6, respectively. These values outperform the dual-branch architecture Dyn-Perceiver and artifact classification methods AR-Classifier and DKL. Notably, our network reduces parameters and computational load by 80.91% and 96.94%, respectively, compared with DKL. In addition, heatmap visualization analysis confirms that our method accurately identifies critical regions of pathological artifacts and assigns higher weights to these areas, which greatly enhance the accuracy of multi-class artifact detection tasks. **Conclusion** Compared with similar methods, our approach effectively extracts artifact information from images. In this network, we designed the MGL-Trans based on Transformer principles. It allows the model to selectively enhance or suppress features at various stages, which improves its capability to recognize complex image patterns. We also developed a focused linear cross-attention mechanism based on cross-attention. FLCA utilizes an efficient focusing function to better capture dependencies between two input sequences while maintaining computational efficiency and high performance. These innovations enable the model to effectively distinguish similar artifacts in histopathological slides while significantly reducing computational resource consumption. This achievement offers an innovative and efficient solution for the quality assessment of pathological images.

Key words: whole slide images; artifact classification; attention mechanism; feature fusion; lightweight

0 引言

当前,癌症依然是中国乃至众多发达国家面临的主要致死疾病之一(Xia等,2022)。组织病理图像分析为肿瘤诊断和治疗提供了精确的病理信息,被该领域专家公认为肿瘤诊断的金标准(Liang等,2021)。组织病理切片在数字化前,需经过取材、固定、脱水、包埋、切片、染色和封片等一系列复杂步骤(丁维龙等,2024)。这一过程烦琐、自动化程度低,并受技术员手工操作影响,容易产生意外的缺陷(本文称其为“伪影”)。伪影主要分为3级:1)组织级,如刀痕、松散;2)切片级,如气泡、墨水;3)扫描级,如失焦(Kanwal等,2022)。这些伪影可能会使病理医生在观察和判断真实病变时无法观察到正常组织,因而导致误诊或漏诊,进而影响病人的治疗和预后。此外,存在这些问题的组织病理图像对计算机辅助诊断系统也造成负面影响。因此,对制作完成的组织病理图像进行相关检测就显得尤为重要。

现有对墨水、组织折叠、模糊等形态及颜色差异较大的伪影进行识别与区分的深度学习方法,已取得较好的效果(Babaie和Tizhoosh,2019;Hemmatirad等,2022;Shakarami等,2023)。但伪影种类繁多,某些如松散和组织气泡等种类的伪影在外观上极为相似,如图1所示,这使得区分它们变得极为困难。从局部来看(图1中4个小框区域),这些伪影特征都表现为在组织上分布着不规则的各种裂缝;从整体视角来看(图1第2列方框中图像),这4种伪影同样显示出较高的相似度。这表明,如果模型仅聚焦于细节或过度依赖全局信息,将难以有效区分这些伪影。在相同种类的伪影中,不同图像间差异也很明显,以伪影刀痕为例(图1中最后一行),最左侧的图像与右侧两幅图像在纹理和颜色上存在显著差异,它们的共同特征是存在线状分布的裂缝(如图1中最后一行图中线条所示)。因此,为了有效区分这些伪影,模型必须在保持对局部特征的敏感性的同时,理解整体模式和结构,并实现局部信息与全局信息的充分融合。然而,当前的伪影检测方法尚未针对这些相似伪影展开相关研究。

本文针对现有方法在识别高度相似的伪影时所面临的局部与全局信息融合与交互不足的问题,提出一种面向病理图像伪影多分类的局部与全局信息

交互融合网络。这一双分支架构由两个关键部分组成:卷积神经网络(convolutional neural network, CNN)分支和变换器(Transformer)分支。前者用以充分提取纹理和颜色的局部信息,而后者则用以捕捉裂缝的整体分布规律等全局信息。

为增强Transformer分支在并行框架中的特征提取能力,设计了基于Transformer的MGL-Trans(mobile gated linear Transformer),该模块中使用了基于门控线性单元(gated linear unit, GLU)(Dauphin等,2017)和基于MBConv的MobileGLU(mobile convolution gated linear unit)模块来替代传统的前馈层。MobileGLU的设计使得模型能够在不同阶段有选择性地强化或抑制特征,进而增强了对复杂图像模式的识别能力。为了进一步提升局部与全局信息的交互和融合,设计了一种基于cross-attention(Chen等,2021)的线性聚焦交叉注意力机制(focused linear cross-attention, FLCA)。FLCA采用高效的聚焦函数使模型在更好地捕捉两个输入序列之间的依赖关系的同时,保证了计算效率和性能。针对伪影中常见的大量裂缝特征,还设计了一个基于HSV(hue, saturation, value)颜色空间的边缘检测模块。该模块通过整合HSV 3个通道的边缘信息,显著提升了网络对裂缝特征的识别与关注。通过这些设计,本文提出的模型能够更精准地对病理图像中的相似伪影进行分类,为病理图像伪影检测提供了一种新的解决方案。

本文主要贡献如下:1)提出基于Transformer的

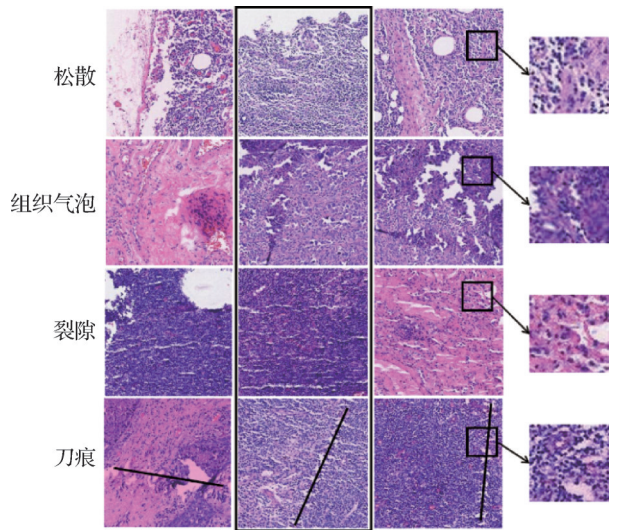


图1 伪影示例

Fig. 1 Examples of artifacts

全局特征提取模块 MGL-Trans。该模块采用的前馈层 MobileGLU 通过门控机制有效减少不必要的特征流动,其中倒残差结构通过压缩和扩展特征,能够在特征变换中保持信息的传递与梯度流动的稳定性。2)提出基于 cross-attention 的线性聚焦交叉注意力机制,该机制能够促进全局特征与局部特征之间的高效交互,通过线性聚焦策略确保信息的有效传递和整合。3)提出基于 HSV 颜色域的边缘检测模块,该模块能够同时保留组织病理图像在 HSV 颜色空间的 3 个通道上提取的边缘信息,以增强对伪影中存在的空白裂缝的关注。

1 相关工作

当前,病理图像的伪影检测方法主要分为 3 种:人工质检、机器学习和深度学习。人工质检是医疗机构目前普遍采用的方法。由于每天都会制备大量的组织病理切片,临床上大的病理科室只能按一定比例进行人工抽检,一般无法覆盖所有新制备的切片。此外,由于缺乏统一的评判标准,不同人员的判断结果可能存在差异(Hosseini 等, 2019)。相比之下,机器学习和深度学习方法则提供了一种更为高效和客观的解决方案。

1.1 基于机器学习的方法

在早期的研究中,伪影在组织病理图像上的检测与识别主要通过传统的机器学习技术实现。其中许多研究专注于利用图像的颜色特征,如 Palokangas 等人(2007)通过将切片图像转换到 HSV 颜色空间进行处理来提高目标像素的识别度,然后用 k -means 聚类算法从图像中准确提取褶皱像素簇。Bautista 和 Yagi(2010)提出一种基于颜色的方法,采用具有固定阈值的策略来检测组织折叠。Kothari 等人(2013)提出一种基于颜色特性和自适应连通性阈值的组织折叠伪影检测方法,用于检测低分辨率全切片图像中的组织折叠伪影。然而,由于医疗机构间染色技术和数字扫描仪不同,甚至同一机构不同批次的病理切片在染色过程中也存在变化,导致图像的颜色特征可能呈现不一致性。这使得上述过于依赖颜色的方法难以适应这些变化,影响识别效果的稳定性。

此外,一些研究尝试采用人为设计的特征,如 Hashimoto 等人(2012)采用线性回归分析方法,选取

图像的锐度和噪声特征来进行计算,以评估全切片成像的图像质量。Shakhawat 等人(2020)则使用支持向量机(support vector machine, SVM)学习基于灰度共生矩阵(gray-level co-occurrence matrix, GLCM)的纹理特征,以区分气泡和组织折叠。然而,人工设计的特征可能无法全面捕捉图像中的纹理和颜色信息,所设计的特征的有效性难以保证,从而限制了这些方法对图像信息的准确映射。

1.2 基于深度学习的方法

随着深度学习技术在计算机视觉领域的迅速推进, CNN 在病理图像伪影检测的应用上已经取得了显著的进展。Foucart 等人(2018)利用深度残差网络从粗糙的标签中提取信息,用于苏木精-伊红(hematoxylin-eosin, H&E)和免疫组化(immunohistochemistry, IHC)全切片图像(whole slide image, WSI)中的伪影检测和分割。Haghighat 等人(2022)使用基于 ResNet18 的多任务深度神经网络在 5 倍放大率下,对 WSI 的质量进行逐块评估。然而,这些研究中使用的网络模型,如深度残差网络和 ResNet18 在处理全局信息方面存在一定的局限性。为了克服这一挑战,Shakarami 等人(2023)提出一种用于识别 WSI 上伪影的新 CNN 架构 TCNN(Transformer convolutional neural network),通过在卷积层的最后几层中应用扩张卷积,以增强模型捕获图像全局上下文信息的能力。Pizarro 等人(2023)通过堆叠多个卷积层的方式,逐层增加其感受野。上述基于 CNN 的改进工作在一定程度上拓宽了网络的感受野,使其能够更有效地捕捉全局信息。然而, CNN 的感受野仍然受限于卷积核的尺寸和网络架构的深度(李金星等, 2022)。特别是随着网络深度的增加,细节信息的丢失风险也随之增加。

随着 ViT(vision Transformer)的提出(Dosovitskiy 等, 2021), Transformer 架构在计算机视觉领域的应用引起了广泛关注。得益于其自注意力机制, Transformer 的感受野可以立即覆盖整个输入图像(Sun 等, 2021; Khan 等, 2022)。Guo 等人(2023)提出 BIQA(blind image quality assessment)模型。该模型的创新之处在于融合了基于 DInet(dilated inception network)(Yang 等, 2020)的显著性模块,以及一个基于多头自注意力和金字塔位置编码生成器的图像块模块。该模型结合 CNN 和基于 Transformer 的网络模块,分别针对局部和全局伪影特征进行处理。然

而,模型在最终分类头之前仅通过简单地拼接来整合局部信息与全局信息,这种方法未能充分发挥两者结合的优势,导致性能提升受限。

对此,Chen 等人(2022)和 Han 等人(2023b)相继提出 mobile-former 和 Dyn-Perceiver。这些方法均采用了 CNN 与 Transformer 并行的设计,并通过双向桥接结构来实现两者之间的特征融合。然而, mobile-former 中 Transformer 分支使用的前馈层(feed-forward network, FFN)在并行框架中的表现并不理想,且其中 MobileNet 子块的输入依赖于 Transformer 子块的输出,这限制了两个分支的独立并行性能。Dyn-Perceiver 则更侧重于动态提前退出。本文延续了并行框架的思想,并对 CNN 分支与 Transformer 分支的交互时机进行了优化调整。

在当前基于深度学习的伪影检测方法中,一些方法借鉴了传统机器学习领域中对颜色处理的策略,但与传统方法相比,它们对颜色的依赖显著降低。如 TCNN 在初始层中使用了一个色彩映射层,该层将图像的色彩空间从 RGB 域转换为 HSV 域(Shakarami 等,2023)。但该方法仅进行了简单的颜色域转换。Ke 等人(2023)提出的 AR-Classifier(artifact recognition classifier)则进一步利用来自不同颜色空间(如 LAB(CIELAB)、HED(hematoxylin-eosin-DAB)和 HSV)的互补信息,但同时也导致计算量的成倍增加。Tellez 等人(2019)指出采用 HSV 颜色空间转换技术,可以显著提高 CNN 在病理学图像分类任务中的性能。而本文要区分的这些伪影中存在大量的裂缝,这是区分的关键点。由此,本文提出基于 HSV 的边缘检测模块(HSV edge fusion, HSVEF),该模块在将颜色域转换为 HSV 域的基础上,引入了 Canny 边缘检测器,以增强图像中裂缝与正常组织之间的边界信息。

当前,所讨论的伪影检测技术主要集中于识别特征差异较大的伪影。但是,对于本文探讨的那些特征相似度较高的伪影,现有的检测方法尚未涉及,且在区分这些相似伪影的能力上表现出明显不足。

2 方 法

2.1 网络整体结构

在病理学领域,组织病理图像通常以全切片图像(WSI)的形式存储,这种格式中图像以金字塔结

构进行组织。当使用 40 倍的放大倍率扫描时,WSI 的分辨率可达到 $50\,000 \times 50\,000$ 像素(金旭 等,2020)。因此,研究人员通常会对 WSI 进行图像预处理以分割成小图像块。具体步骤如图 2(a)所示。首先,利用 OTSU 阈值分割算法将图像中的组织区域与背景区域分离,生成一个仅包含黑白两色的掩码图(赵樱莉 等,2023)。接着,通过对掩码图与原始 WSI 进行按位与操作,提取出图像中的组织区域。最后,采用滑动窗口技术将选定的组织区域划分成较小的图像块。在得到图像块后,进一步利用基于 HSV 颜色域的边缘检测模块 HSVEF 对图像进行处理。这一步骤将图像转换为更符合人眼颜色感知的 HSV 颜色域,并增强图像中组织区域与裂缝区域的边缘,从而为后续的分析 and 识别提供更为清晰的图像特征。

本文方法延续了并行框架的思想(Chen 等,2022; Han 等,2023b),并对其 CNN 分支与 Transformer 分支的交互时机进行了优化调整。进一步地,为了深化特征的交互与融合,设计了一种新型的交叉注意力模块 FLCA,旨在提升特征交互机制的效率与效果。此外,为了增强 Transformer 分支在并行框架中的特征提取能力,本文基于 MobileNet2 中倒残差结构(Sandler 等,2018)和门控线性单元 GLU 设计了 MobileGLU,用以替代传统的自注意力机制中的前馈层,以期实现更优的性能表现。

具体实现如图 2(b)所示,将图像特征编码为一个 4 维数组 $F \in \mathbf{R}^{(B \times C \times H \times W)}$,其中 B 表示处理的图像批次的数量, C 为特征通道数, H 和 W 分别表示图像的高和宽。对于双分支网络中 CNN 分支,本文采用轻量化卷积神经网络 MobileNetV3(Howard 等,2019),旨在保证特征提取效率的同时,增强网络对局部细节的提取。将图像输入该分支后,其特征被逐步从低级特征提炼至更抽象的高级特征。Transformer 分支则由 MGL-Trans 实现,它接收经过维度压缩后的图像特征作为输入。基于 Transformer 的模型(Zhu 等,2023; Shi, 2024)的常见处理方法,会在两个 Transformer 子模块之间使用下采样模块(patch merging)来缩小图像的分辨率。参考这种做法,也在每两个 MGL-Trans 模块之间引入了令牌混合器(token mixer, TM)(Han 等,2023b),以逐步减少令牌(token)的数量和增加潜在编码(latent code)的维度。

在双分支之间,本文构建了线性聚焦交叉注意力机制 FLCA,实现了局部特征与全局特征的有效交互。该注意力机制在本模型中具体实现为特征聚焦模块 (feature gather module, FGM) 和特征扩散模块

(feature diffusion module, FDM)。最终,两个分支提取的特征被拼接起来,输入到分类头中进行分类。这种设计不仅提高模型对图像特征的全面理解,还增强了其在复杂图像分析任务中的性能。

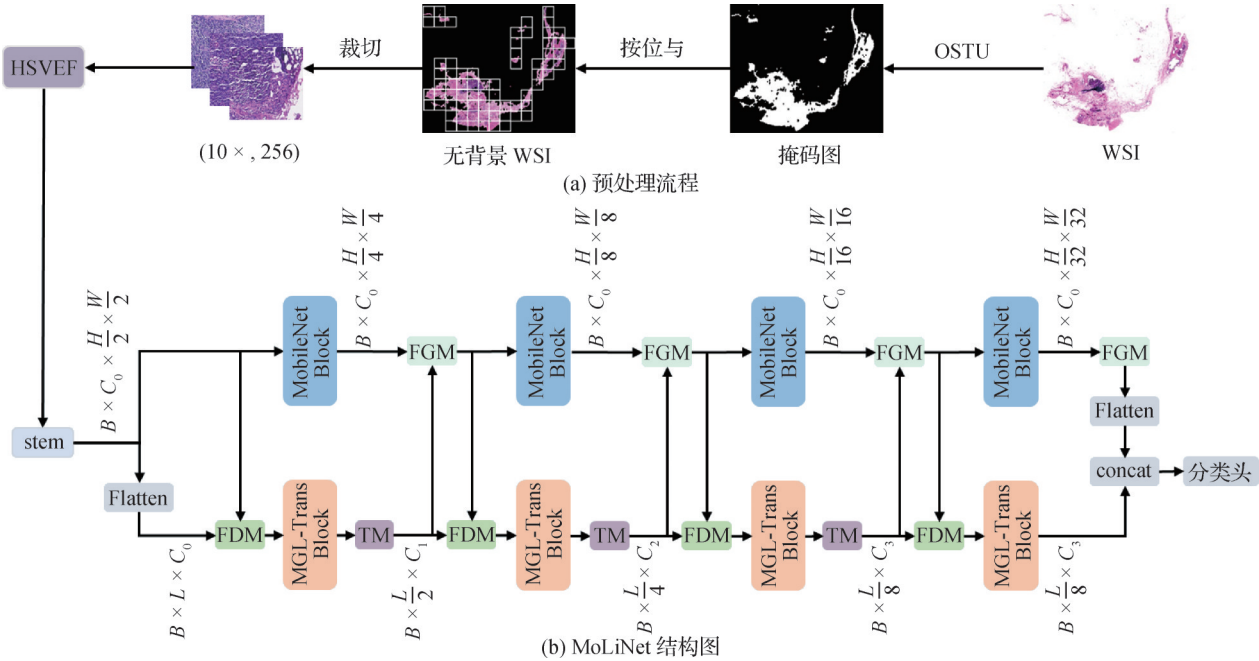


图 2 局部与全局信息交互融合网络的整体架构

Fig. 2 The overall architecture of the local and global information interaction fusion network
((a) preprocessing process; (b) MoLiNet structure diagram)

2.2 基于HSV颜色域的边缘检测模块

组织病理图像上有着丰富的细胞结构,而正常细胞与受伪影影响的细胞在其形状、完整性以及颜色上都有一定的差异,这使得模型对细胞特征的准确学习尤为关键。而在HSV颜色域中,模型可以类似于人类视觉感知系统的工作方式对细胞进行分析学习(Najjar等,2023)。基于此,选择将图像从RGB颜色域转化到HSV颜色域,并针对病理图像伪影中

具有大量空白裂缝的特点,设计了一种在HSV颜色空间中应用Canny边缘检测器的模块,以强化裂缝特征,具体结构如图3所示。

在该模块中,首先将图像转至HSV颜色域,然后分别提取色调(hue)、饱和度(saturation)和明度(value)3个通道,得到3个独立的图像。接下来,对这3个通道的图像分别进行高斯模糊处理,以减少噪声并保留边缘信息。随后,利用Canny边缘检测

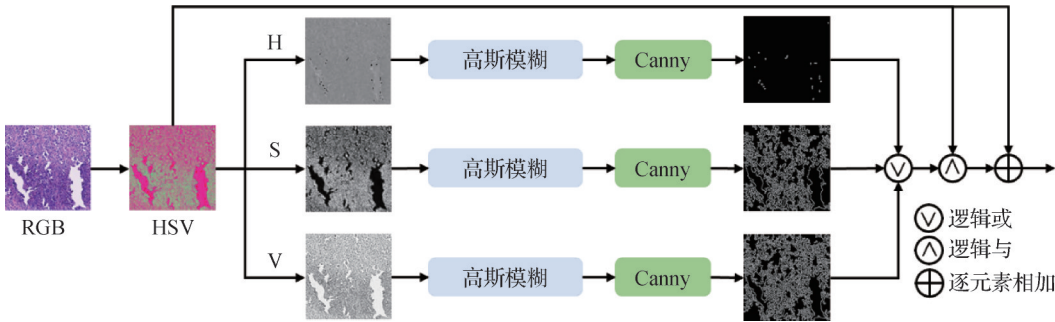


图 3 基于HSV颜色域的边缘检测模块 HSVEF

Fig. 3 HSVEF edge detection module based on HSV color domain

器在每个通道上识别边缘。识别出的边缘掩码信息随后通过按位或运算合并,形成边缘掩码的并集,这一操作确保了所有通道检测到的边缘都被纳入考虑。然后得到的边缘掩码的并集与原始的 HSV 图像进行按位与运算,这一操作定位到原图中的边缘部分,以得到边缘图。最后将边缘图与原图逐位相加,实现边缘增强的效果。该过程具体为

$$\begin{aligned} F &= HSVEF(P) = \\ &(C(P^H) \vee C(P^S) \vee C(P^V)) \wedge \cap P^{HSV} \oplus P^{HSV} \\ &C(x) = Canny(GB(x)) \end{aligned} \quad (1)$$

式中, F 是经过 HSVEF 模块后得到的图像特征, P 表示输入的图像, P^{HSV} 表示 HSV 颜色域的图像, P^H 、 P^S 和 P^V 分别是 HSV 颜色域中色调(hue)、饱和度(saturation)和色明度(value)通道的图像, GB 表示高斯模糊(Gaussian blur), $\oplus$ 为按元素相加、 $\wedge$ 是逻辑与、 $\vee$ 是逻辑或。

2.3 MGL-Trans 模块

组织病理图像与自然图像不同,组织病理图像在纹理特征上更加复杂,尤其是其中的各种组织间的大片空白以及大量不规则裂缝为图像的识别带来挑战,使 Transformer 分支更难以学习到图像中伪影的分布规律。而在并行框架中的 Transformer 模块所带来的增益非常有限,尤其是其中的前馈层 MLP (multi-layer perceptron) 相比在原始的 Transformer 架构中带来的增益也很有限(Chen 等, 2022)。为了使 Transformer 模块在并行框架中能有更强的特征提取效果,本文引入了 MobileNet 中的倒置残差块(Sandler 等, 2018),考虑到 MLP 和倒置残差块两者都采用反向瓶颈设计,而倒置残差块则有更强的特征提取能力,故将倒置残差块与门控线性单元 GLU 进行联合以构建成新的前馈层 MobileGLU。然后将 MobileGLU 与线性聚焦注意力(focused linear attention, FLA)结合形成 MGL-Trans 模块(Han 等, 2023a),其网络结构如图 4 所示,图中 c 表示通道(channel)。

特征在该模块的提取过程的表达式为

$$Y_{i+1} = MobileGLU(Y_i \oplus FLA(LN(Y_i))) \quad (2)$$

式中, Y_i 表示输入 MGL-Trans 模块的图像特征,经过 MGL-Trans 模块后的输出表示为 Y_{i+1} , LN 为层归一化(layer normalization), $\oplus$ 为按元素相加操作。其中 MobileGLU 模块对应图 4 中虚线框内部分,计算为

$$MobileGLU(x) =$$

$$x \oplus Linear(MBConv(Linear(x)) \odot Linear(x)) \quad (3)$$

式中, $Linear$ 为线性层, $\odot$ 表示按元素相乘操作,其中, MBConv 模块如图 4 所示,计算为

$$\begin{cases} MBConv(x) = x + (C_1 \circ C_2 \circ C_3)(BN(x)) \\ C_1(x) = GeLU(BN(Conv^{1 \times 1}(x))) \\ C_2(x) = GeLU(BN(DWC(x))) \\ C_3(x) = Conv^{1 \times 1}(x) \end{cases} \quad (4)$$

式中,“ $\circ$ ”表示函数复合符号, $Conv^{1 \times 1}$ 表示 1×1 卷积, DWC 为深度卷积(depthwise convolution), BN 表示批归一化(batch normalization), $GeLU$ 为激活函数。

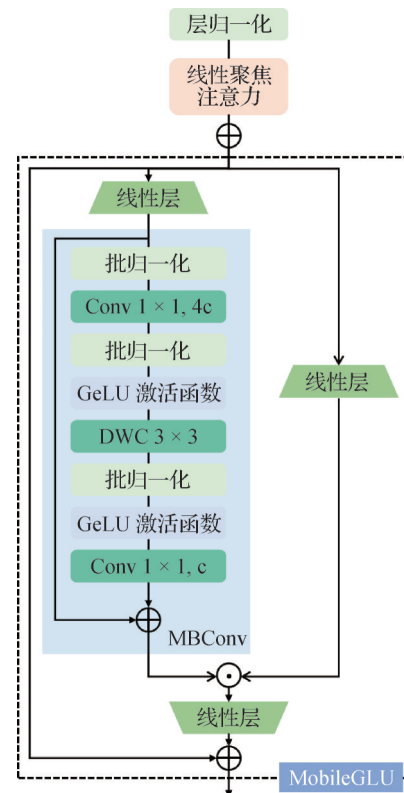


图4 MGL-Trans 模块

Fig. 4 MGL-Trans block

2.4 线性聚焦交叉注意力模块

一些伪影不仅在局部特征上相似,在整体结构上也有高度的相似性。因此,模型不仅需要捕捉到局部的纹理和色彩细节,还必须理解伪影的整体分布模式,并将这些局部与全局信息有效地结合。为了实现这一目标,本文采用了交叉注意力机制,并进一步引入了线性聚焦注意力,以构建一种新型的线性聚焦交叉注意力机制。这一机制在本文中主要通

过特征扩散模块和特征聚焦模块来实现,其结构详见图5。

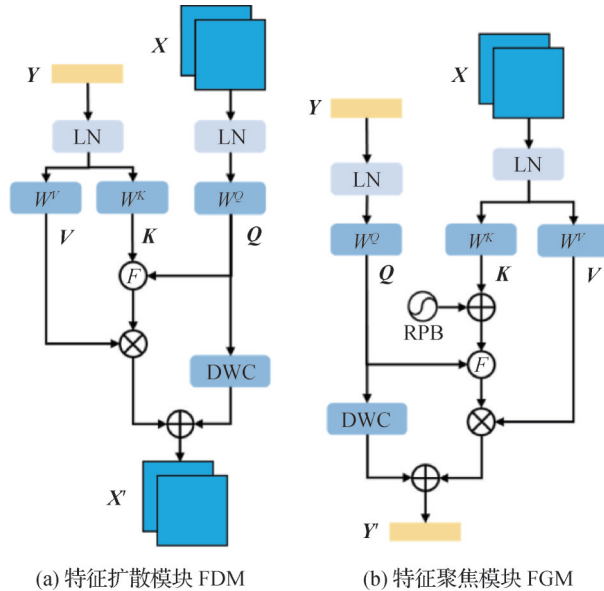


图5 线性聚焦交叉注意力机制 FLCA

Fig. 5 Linear focused cross attention mechanism FLCA

((a) feature diffusion module; (b) feature gather module)

由于 MobileNet 分支仅具有较小的感受野,故其主要致力于捕捉图像的局部细节。为了弥补这一点,本文设计了 FGM 模块。该模块能为该分支补充伪影整体分布特征,以增强该分支对关键区域的聚焦能力。相对地, MGL-Trans 分支则更侧重于全局信息的捕捉。为了进一步丰富这一分支的局部特征理解, FDM 模块被设计用于将丰富的局部细节信息扩散到全局特征中。通过 FDM 和 FGM 这两个模块的协同工作,实现了局部与全局特征的互补和交互。

这两个模块的输入都是 X 和 Y , 其中 X 表示 MobileNet 分支获取的局部特征, 而 Y 表示 MGL-Trans 分支获取的全局特征。特征在 FDM 模块的提取过程计算为

$$X' = FDM(X, Y) = (F(K, Q) \otimes V) \oplus DWC(Q) \quad (5)$$

$$Q = W^Q(LN(X))$$

式中, X' 表示输出的新局部特征。查询向量 Q 由局部特征 X 经过归一化再经过线性变化得到, 键向量 K 和值向量 V 则由全局向量 Y 经过类似的处理得到。 F 表示聚焦相似度, 具体为

$$F(Q, K) = \psi_p(Q) \psi_p(K)^T$$

$$\psi_p(x) = f_p(ReLU(x))$$

$$f_p(x) = \frac{\|x\|}{\|x^{**p}\|} x^{**p} \quad (6)$$

式中, x^{**p} 是对 x 中每个元素进行 p 次幂操作, 聚焦函数 f_p 可以帮助注意力集中在相似的特征上。特征在映射前后的范数相同, 即 $\|x\| = \|f_p(x)\|$, 表示该函数仅调整了特征方向。在 FGM 模块中, 对两个分支的输入特征进行了与 FDM 模块核心操作相反的处理, 区别在于计算键向量时额外引入了相对位置偏差 (relative position bias, RPB), 以编码位置信息并增强模型对空间关系的感知。

3 实验

3.1 数据集介绍

深度学习技术的快速发展极大地推动了病理图像伪影研究的进展。尽管如此, 目前的研究大多依赖于私有数据集, 而公开可用的数据集相对稀缺。现有的公开数据集在图像数量上也显得捉襟见肘, 例如 HistoQCRRepo 仅包含 171 幅图像 (Janowczyk 等, 2019)。此外, 现有的数据集尚未涵盖能够支持相似伪影细粒度分类任务的图像。

为了克服这一挑战, 本文与宁波市临床病理诊断中心携手合作, 共同采集了 2 552 幅病理全切片图像 (朱伟, 2024)。这些图像均经过该中心 11 位经验丰富的病理专家的精心标注与严格审核。这批全切片图像中所包含的伪影种类众多, 本文基于病理专家的精确标注, 从全切片图像中裁剪出了 256×256 像素的图像块。然后收集整理了其中 4 种相似伪影: 松散、组织气泡、裂隙和刀痕, 同时还收集了部分正常组织区域的图像构建出数据集 NCPDCSAD, 各类别的具体数量列于表 1。

表1 NCPDCSAD数据集数量统计
Table 1 Statistics of the NCPDCSAD dataset

图像	数量/幅
松散	84 201
组织气泡	34 277
裂隙	96 599
刀痕	108 292
正常	82 754
总计	406 123

3.2 实验设置

本文的实验基于一台配备有 Intel i9-10940X @ 3.30 GHz 处理器和 NVIDIA GTX3080Ti 显卡的计算机。实验所采用的深度学习框架是 PyTorch, 版本为 1.12.1。在模型训练阶段, 采用了余弦退火策略来调整学习率, 初始学习率设定为 $4E-3$, 同时设定了一个最小学习率 $1E-6$ 以保证训练过程的稳定性。此外, 为了加快模型的收敛速度, 设置了 20 个周期的学习率预热阶段。优化算法方面, 选择了结合权重衰减技术的 AdamW 优化器, 其 `weight_decay` 参数设置为 0.05。实验设置的 `batch_size` 为 128, 整个训练过程计划执行 200 个轮次。为了全面评估模型性能, 选用了准确率、精确率、召回率和 F1 分数等关键指标, 这些指标均为多分类任务中的宏观平均结果, 能够从不同角度反映模型的分类效果。在实验设计上, 将数据集以 60%、20% 和 20% 的比例随机划分为训练集、验证集和测试集。为了防止模型过拟合, 在训练中引入了 Dropout 正则化技术。

3.3 对比实验

为验证本文提出的网络模型的有效性, 在 NCPDCSAD 数据集上将本文方法与 MobileViT-XS

(Mehta 和 Rastegari, 2022)、VanillaNet-5 (Chen 等, 2023)、AR-Classifier、Dyn-Perceiver 和 DKL (deep kernel learning) (Kanwal 等, 2024) 等网络模型进行对比, 结果如表 2 所示。在这些模型中, MobileViT-XS 和 VanillaNet-5 是用于分类任务的主流分类网络模型, 其中 MobileViT-XS 串行结合了 CNN 和 ViT。Dyn-Perceiver 是一个并行结合了 CNN 与 Transformer 的双分支网络模型。AR-Classifier 和 DKL 是专门针对组织病理图像伪影分类的先进方法。为了确保比较的公平性, 所有模型的输入图像都经过了相同的预处理, 并将输入模型的图像尺寸都统一转换为 224×224 像素。

与其他算法相比, 本文方法展现出了最好的性能, 准确率 (accuracy) 达到了 94.24%, 超越了同为双分支架构的 Dyn-Perceiver 以及专注于伪影分类的 AR-Classifier 和 DKL。此外, 本文方法在计算量 (Flops) 和参数量 (Params) 方面也显示出了优势。在所有对比方法中, 本文方法在参数量上仅次于 MobileViT-XS, 而在计算量上则是最低的。这表明, 本文方法在保持高检测性能的同时, 也成功地保持了较低的计算成本。

表 2 不同方法的性能对比结果
Table 2 Performance evaluation results of different methods

模型	准确率	精确率	召回率	F1 分数	Params/M	Flops/M
MobileViT-XS	0.931 3	0.937 4	0.935 9	0.936 5	1.94	909
VanillaNet-5	0.928 2	0.935 5	0.936 1	0.935 7	11.80	3 070
AR-Classifier	0.936 9	0.943 1	0.944 3	0.943 7	11.17	27 336
Dyn-Perceiver	0.938 9	0.944 2	0.946 4	0.945 2	5.78	256
DKL	0.937 1	0.943 0	0.944 8	0.943 8	27.67	8 038
本文方法	0.942 4	0.947 7	0.949 8	0.948 6	5.28	246

注: 加粗字体表示各列最优结果。

3.4 消融实验

为了验证本文设计的各子模块有效性, 在 NCPDCSAD 数据集上进行了一系列的消融实验。

1) HSV 颜色边缘融合模块。本实验深入探讨了不同颜色域处理方法对模型性能的影响。实验结果如表 3 所示, 首先分别在 RGB 颜色域上和 HSV 颜色域上进行了实验, 结果显示在 HSV 颜色域中的实验效果显著优于 RGB 颜色域。这表明, 在 HSV 颜色空间中处理组织病理图像分类任务时, 能够更好地提

取颜色特征并提供更稳定的性能表现。随后, 为了增强模型对图像裂缝特征的關注, 本文基于 HSV 颜色空间使用 Canny 边缘检测器提取裂缝边缘。由于 HSV 颜色空间由色调 (hue)、饱和度 (saturation) 和亮度 (value) 3 个通道组成, 本文分别在每个通道上独立应用了 Canny 检测进行实验。实验结果表明, 单独在每个通道上应用 Canny 的效果并不如同时在 3 个通道上应用 Canny 的效果好, 这进一步凸显了综合利用 HSV 颜色域 3 个通道信息的重要性。

表 3 HSV 颜色边缘融合模块消融结果
Table 3 Ablation results of HSV color edge fusion module

预处理	准确率	精确率	召回率	F1 分数
RGB	0.919 2	0.925 6	0.929 6	0.927 0
HSV	0.941 3	0.946 6	0.948 4	0.947 5
H+Canny	0.940 0	0.945 4	0.947 2	0.946 2
S+Canny	0.939 0	0.944 4	0.946 6	0.945 4
V+Canny	0.940 7	0.946 3	0.948 2	0.947 2
HSVEF	0.942 4	0.947 7	0.949 8	0.948 6

注:加粗字体表示各列最优结果。

2) MGL-Trans 模块。为了评估 FLA 模块和 MobileGLU 模块对模型性能的具体影响,本研究实施了一系列消融实验,结果如表 4 所示。先采用多头自注意力模块(multi head attention, MHA)和两层组成的前馈网络(MLP)作为基准模型。在此基础上,探索了用 FLA 模块替代 MHA 的潜在优势,实验数据表明 FLA 模块在性能上取得了更优的结果。进一步对比了多种不同的前馈层配置,包括 GLU(Dauphin 等, 2017)、ConvGLU (Shi, 2024) 和 MBCConv (Sandler 等, 2018)等变体。然而,这些替代方案的实验表现并不理想。本文提出的 MobileGLU 模块,凭借其 GLU 的门控机制显著提升了非线性处理能力,使得网络能够从复杂的组织结构中捕捉到更为细微的特征模式。此外,通过引入倒置残差块, MobileGLU 模块显著提升了前馈层的特征提取能力,解决了传统设计中特征提取能力不足的问题。这些设计使得 MobileGLU 模块在提升模型整体性能方面发挥了关键作用。

表 4 MGL-Trans 模块消融结果
Table 4 Ablation results of MGL-Trans module

注意力模块	前馈层	准确率	精确率	召回率	F1 分数
MHA	MLP	0.940 3	0.946 3	0.947 4	0.946 8
FLA	MLP	0.941 1	0.946 8	0.948 0	0.947 4
FLA	GLU	0.940 1	0.945 7	0.947 2	0.946 4
FLA	ConvGLU	0.939 1	0.945 1	0.946 4	0.945 7
FLA	MBCConv	0.940 4	0.945 9	0.947 6	0.946 6
MHA	MobileGLU	0.938 8	0.944 8	0.945 7	0.045 2
FLA	MobileGLU	0.942 4	0.947 7	0.949 8	0.948 6

注:加粗字体表示各列最优结果。

3)线性聚焦交叉注意力模块。本文研究了不同交叉注意力模块对模型性能的影响。如表 5 所示,首先进行了基于自注意力机制的 cross-attention 模块的实验。随后,对比分析了在模型中加入深度卷积模块 DWC 与否的效果。结果表明,与 cross-attention 相比,FLCA 模块显著提升了模型性能。此外,通过引入 DWC 模块,增强了特征的多样性,从而增强了线性注意力的特征表达能力,这也为模型性能带来了一定的提升。

表 5 线性聚焦交叉注意力模块消融结果
Table 5 Ablation results of linear focused cross attention module

交叉模块	准确率	精确率	召回率	F1 分数
cross-attention	0.940 8	0.946 2	0.948 4	0.947 2
FLCA(无 DWC)	0.941 2	0.947 0	0.947 9	0.947 4
FLCA	0.942 4	0.947 7	0.949 8	0.948 6

注:加粗字体表示各列最优结果。

4) MGL-Trans 分支输入研究。本文还研究了 MGL-Trans 分支中不同输入对模型效果的影响,在 mobile-former 和 Dyn-Perceiver 的双分支框架设计中,Transformer 分支的输入使用了可学习的随机参数。本文进一步探索了直接使用图像信息作为该分支输入的效果,如表 6 所示,实验结果表明,在相似伪影细粒度分类场景中,使用图像信息作为输入能使模型更直接和有效地学习伪影的内在特征,提升了模型在处理精细视觉差异时的性能。

表 6 MGL-Trans 分支的输入消融结果
Table 6 Input ablation results of MGL-Trans branch

输入	准确率	精确率	召回率	F1 分数
随机可学习	0.939 6	0.945 4	0.946 8	0.946 1
图像信息	0.942 4	0.947 7	0.949 8	0.948 6

注:加粗字体表示各列最优结果。

5)局部特征提取模块选取。为了选择合适的局部特征提取器,本研究对比实验了卷积神经网络 ResNet (He 等, 2016)、RegNet (Xu 等, 2023) 以及 MobileNet-V3,实验结果如表 7 所示。在所有评估指标上, MobileNet-V3 均展现出超越其他模型的性能。作为 MobileNet 家族的最新成员,不仅继承了前代版本的深度可分离卷积和倒残差结构设计,还引入了

挤压—激励模块(squeeze-and-excitation block)来进一步提升网络的特征表达能力。MobileNet-V3在确保高准确率的同时,也维持了较低的计算复杂度。因此,本文选择 MobileNet-V3 作为局部特征提取器。

表 7 局部特征提取模块的消融结果
Table 7 Ablation results of local feature extraction module

模型	准确率	精确度	召回率	F1 分数
ResNet50	0.934 3	0.941 2	0.941 9	0.941 5
RegNetY	0.941 3	0.946 9	0.948 3	0.947 6
MobileNet-V3	0.942 4	0.947 7	0.949 8	0.948 6

注:加粗字体表示各列最优结果。

3.5 注意力热图可视化分析

为展示各模型对组织病理图像中伪影区域的关注程度,本文通过 Grad-CAM(Selvaraju 等,2017),分析模型中最后一个卷积层的梯度生成热力图,以显示出模型对图像的关注区域。图6展示了生成的热力图,颜色越红表明模型的关注度越高。可以看出,本文方法的注意力区域及其权重分布更接近伪影的真实分布区域。基于双分支框架的 Dyn-Perceiver 在区域关注上与本文方法接近,但在定位精准度上略有不足。而 DKL 虽然在注意力定位上表现尚可,但其在注意力权重分配上不如本文方法精确。综合分

析各模型的注意力热力图,本文方法成功定位了病理伪影的关键区域,并对这些区域赋予了较高权重,有效提升了伪影多分类任务的准确性。

4 讨 论

结合表2和图6可知,本文所提出的网络相比其他模型在准确率等分类性能指标上以及对伪影区域的关注上都达到了最佳。与轻量化模型 MoblieVit-XS 和 VanillaNet-5 相比,本文方法在分类性能上有显著的提升。而且本文的模型计算量低于轻量化模型 MoblieVit-XS。 MoblieVit-XS 与本文方法一样融合了 CNN 的局部感受野和 ViT 的全局感受野,但其 CNN 和 ViT 模块是串行的,而本文的模型采用了并行结构。这种并行结构使得 CNN 和 ViT 分支更加独立,并通过线性交叉注意力机制进行有效交互,从而在保持较低计算量的同时实现两者的优势互补,生成更丰富的特征表示。相比之下, VanillaNet-5 由于是基于 CNN 的简单层次结构实现了较低的参数量,这使得其在处理较为复杂的计算机视觉任务时能力受限。

与同样是伪影分类模型的 AR-Classifier 和 DKL 相比,本文的模型在分类性能上略有提升,但在计算量和参数量上的优势更为明显。AR-Classifier 通过利用多个颜色域中的特征来增强对不同染色风格的

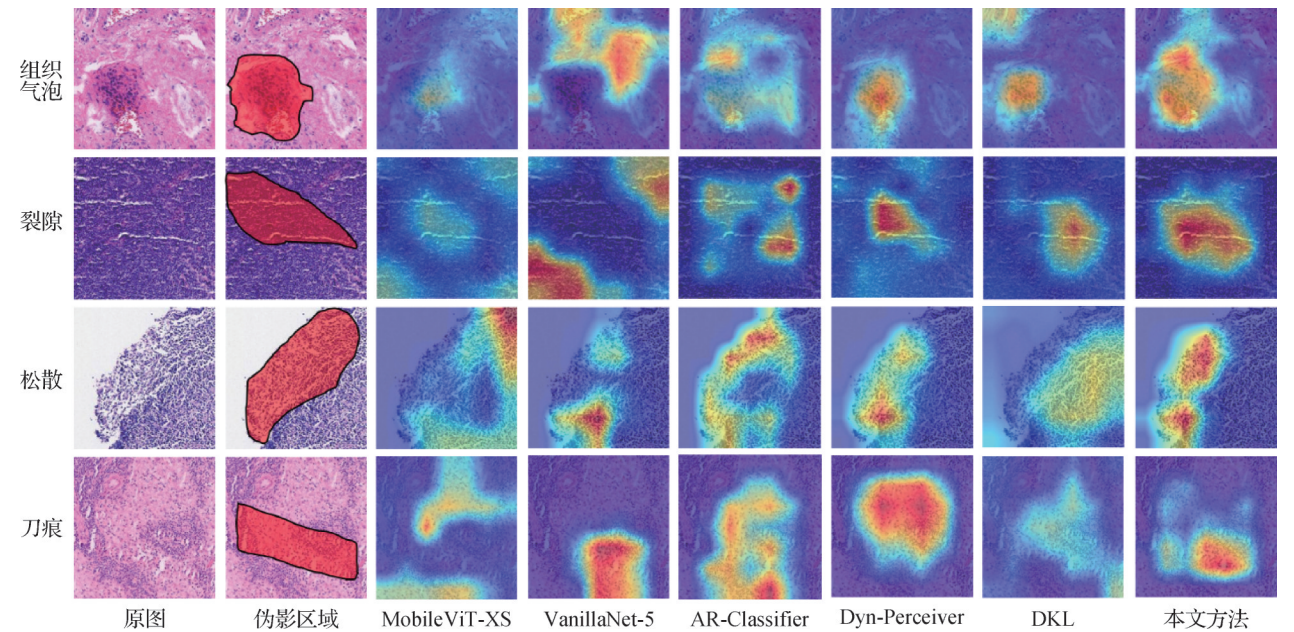


图6 所有对比方法的注意力图
Fig. 6 Attention maps of all comparison methods

组织切片的鲁棒性,但其计算量随着颜色空间数量的增加而显著增加。DKL同样面临计算量大的问题,它使用DenseNet作为特征提取器,并在此基础上应用高斯过程(Gaussian process, GP)来提供预测的不确定性估计,以增强模型的可解释性。然而, DenseNet的密集连接导致了模型参数数量的显著增加。

与同为双分支架构的模型 Dyn-Perceiver 相比,本文方法在各项性能指标上都有一定提升。主要原因是在双分支框架的基础上提出新的前馈层 MobileGLU。该模块结合了门控线性单元 GLU 和 MBConv。MBConv 的倒残差结构能够让模型在不同尺度上处理和压缩信息,结合 GLU 的门控机制,模型可以根据任务需求动态地调整特征流的通过量,并且基于线性聚焦注意力提出新的交叉注意力机制 FLCA。该机制能够让全局和局部特征之间进行更加有效的交互,通过线性聚焦来确保信息的有效传递与融合,使得模型在捕捉细节与整体结构之间达到平衡,从而提升模型的识别能力。

尽管本文模型中已引入了门控线性单元 GLU,并采用了 MobileNet 来降低参数量和计算量,但模型的参数量依然接近 MobileViT-XS 的 3 倍。这可能是由于 MGL-Trans 模块中的线性聚焦注意力相较于自注意力在参数压缩上效果不够显著,且分支间进行了较频繁的交叉注意力操作。该问题在计算资源有限的情况下尤其突出,未来需要进一步优化以实现更高的效率。

本文方法展现了较强的伪影检测能力,但其临床实用性仍需进一步验证。未来,该算法可集成至病理图像质量控制系统,对获取的 WSI 进行切割,提取图像块并输入算法进行伪影识别。根据不同类型伪影对诊断的影响程度设定惩罚系数,并通过加权计算生成质量评分,从而实现对病理图像的质量评估。此外,该算法还可嵌入肿瘤诊断等基于组织病理图像的智能诊断算法流程中,对待输入的病理图像块进行伪影筛选,剔除可能影响后续算法准确性的低质量图像,以提升整体诊断的可靠性和精度。

5 结 论

为了克服现有伪影分类算法在全局信息捕获能力上的局限,以及在相似伪影分类方面的不足,本文提出一种新的改善病理图像伪影多分类的局部与全

局信息交互融合网络 MoLiNet。首先,针对组织病理图像伪影中常见的大量缝隙特征,开发了一种新的预处理模块 HSVEF,该模块在 HSV 颜色空间中结合 Canny 边缘检测技术,以增强对裂缝特征的关注。接着,提出基于 Transformer 的 MGL-Trans 模块,融合了线性聚焦注意力和创新的前馈层 MobileGLU。MobileGLU 结合了 MbConv 的倒残差结构和 GLU 的门控机制,实现了计算效率和特征选择的双重优化。进一步地,引入了线性聚焦注意力机制,提出新的交叉注意力机制 FLCA。在本文中,FLCA 具体实现为特征扩散模块和特征聚焦模块,以促进全局特征和局部特征之间的双向融合。通过在 NCPDCSAD 数据集上的实验,验证了提出的网络在分类精度和计算效率方面均超越了同类先进方法。未来将致力于构建包含更多伪影类型的数据集,以增强模型的泛化能力。同时,也将探索更轻量化的算法,以提高组织病理图像伪影检测的性能,使其能够满足嵌入切片扫描设备的需要。

参考文献(References)

- Babaie M and Tizhoosh H R. 2019. Deep features for tissue-fold detection in histopathology images//Proceedings of the 15th European Congress on Digital Pathology. Warwick, UK: Springer: 125-132 [DOI: 10.1007/978-3-030-23937-4_15]
- Bautista P A and Yagi Y. 2010. Improving the visualization and detection of tissue folds in whole slide images through color enhancement. Journal of Pathology Informatics, 1(1): #25 [DOI: 10.4103/2153-3539.73320]
- Chen C F R, Fan Q F and Panda R. 2021. CrossViT: cross-attention multi-scale vision transformer for image classification//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE: 347-356 [DOI: 10.1109/ICCV48922.2021.00041]
- Chen H T, Wang Y H, Guo J Y and Tao D C. 2023. VanillaNet: the power of minimalism in deep learning//Proceedings of the 37th International Conference on Neural Information Processing Systems. New Orleans, USA: Curran Associates Inc.: #308
- Chen Y P, Dai X Y, Chen D D, Liu M C, Dong X Y, Yuan L and Liu Z C. 2022. Mobile-Former: bridging MobileNet and transformer//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans, USA: IEEE: 5260-5269 [DOI: 10.1109/CVPR52688.2022.00520]
- Dauphin Y N, Fan A, Auli M and Grangier D. 2017. Language modeling with gated convolutional networks//Proceedings of the 34th International Conference on Machine Learning. Sydney, Australia:

- PMLR; 933-941
- Ding W L, Liu J L, Zhu W and Liao W Y. 2024. Review of quality control algorithms for pathological slides based on deep learning. *Computer Science*, 51(10): 276-286 (丁维龙, 刘津龙, 朱伟, 廖婉茵. 2024. 基于深度学习的病理切片质量控制算法综述. *计算机科学*, 51(10): 276-286) [DOI: 10.11896/jsjx.231000167]
- Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X H, Unterthiner T, Dehghani M, Minderer M, Heigold G, Gelly S, Uszkoreit J and Housley N. 2021. An image is worth 16×16 words: transformers for image recognition at scale//*Proceedings of the 9th International Conference on Learning Representations*. [s. l.]: OpenReview.net
- Foucart A, Debeir O and Decaestecker C. 2018. Artifact identification in digital pathology from weak and noisy supervision with deep residual networks//*Proceedings of the 4th International Conference on Cloud Computing Technologies and Applications (CloudTech)*. Brussels, Belgium: IEEE: 1-6 [DOI: 10.1109/CloudTech.2018.8713350]
- Guo Y F, Hu M H, Min X K, Wang Y, Dai M, Zhai G T, Zhang X P and Yang X K. 2023. Blind image quality assessment for pathological microscopic image under screen and immersion scenarios. *IEEE Transactions on Medical Imaging*, 42(11): 3295-3306 [DOI: 10.1109/TMI.2023.3282387]
- Haghighat M, Browning L, Sirinukunwattana K, Malacrino S, Alham N K, Colling R, Cui Y, Rakha E, Hamdy F C, Verrill C and Rittscher J. 2022. Automated quality assessment of large digitised histology cohorts by artificial intelligence. *Scientific Reports*, 12(1): #5002 [DOI: 10.1038/s41598-022-08351-5]
- Han D C, Pan X R, Han Y Z, Song S J and Huang G. 2023a. FLatten transformer: vision transformer using focused linear attention//*Proceedings of 2023 IEEE/CVF International Conference on Computer Vision*. Paris, France: IEEE: 5938-5948 [DOI: 10.1109/ICCV51070.2023.00548]
- Han Y Z, Han D C, Liu Z Y, Wang Y L, Pan X R, Pu Y F, Deng C, Feng J L Song S J and Huang G. 2023b. Dynamic perceiver for efficient visual recognition//*Proceedings of 2023 IEEE/CVF International Conference on Computer Vision*. Paris, France: IEEE: 5969-5979 [DOI: 10.1109/ICCV51070.2023.00551]
- Hashimoto N, Bautista P A, Yamaguchi M, Ohyama N and Yagi Y. 2012. Referenceless image quality evaluation for whole slide imaging. *Journal of Pathology Informatics*, 3(1): #9 [DOI: 10.4103/2153-3539.93891]
- He K M, Zhang X Y, Ren S Q and Sun J. 2016. Deep residual learning for image recognition//*Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas, USA: IEEE: 770-778 [DOI: 10.1109/CVPR.2016.90]
- Hemmatirad K, Babaie M, Afshari M, Maleki D, Saiadi M and Tizhoosh H R. 2022. Quality control of whole slide images using the YOLO concept//*Proceedings of the 10th IEEE International Conference on Healthcare Informatics (ICHI)*. Rochester, USA: IEEE: 282-287 [DOI: 10.1109/ICHI54592.2022.00049]
- Hosseini M S, Zhang Y Y and Plataniotis K N. 2019. Encoding visual sensitivity by MaxPol convolution filters for image sharpness assessment. *IEEE Transactions on Image Processing*, 28(9): 4510-4525 [DOI: 10.1109/TIP.2019.2906582]
- Howard A, Sandler M, Chen B, Wang W J, Chen L C, Tan M X, Chu G, Vasudevan V, Zhu Y K, Pang R M, Adam H and Le Q. 2019. Searching for MobileNetV3//*Proceedings of 2019 IEEE/CVF International Conference on Computer Vision*. Seoul, Korea (South): IEEE: 1314-1324 [DOI: 10.1109/ICCV.2019.00140]
- Janowczyk A, Zuo R, Gilmore H, Feldman M and Madabhushi A. 2019. HistoQC: an open-source quality control tool for digital pathology slides. *JCO Clinical Cancer Informatics*, 3: 1-7 [DOI: 10.1200/CCI.18.00157]
- Jin X, Wen K, Lyu G F, Shi J, Chi M X, Wu Z and An H. 2020. Survey on the applications of deep learning to histopathology. *Journal of Image and Graphics*, 25(10): 1982-1993 (金旭, 文可, 吕国锋, 石军, 迟孟贤, 武铮, 安虹. 2020. 深度学习在组织病理学中的应用综述. *中国图象图形学报*, 25(10): 1982-1993) [DOI: 10.11834/jig.200460]
- Kanwal N, López-Pérez M, Kiraz U, Zuiverloon T C M, Molina R and Engan K. 2024. Are you sure it's an artifact? Artifact detection and uncertainty quantification in histological images. *Computerized Medical Imaging and Graphics*, 112: #102321 [DOI: 10.1016/j.compmedimag.2023.102321]
- Kanwal N, Pérez-Bueno F, Schmidt A, Engan K and Molina R. 2022. The devil is in the details: whole slide image acquisition and processing for artifacts detection, color variation, and data augmentation: a review. *IEEE Access*, 10: 58821-58844 [DOI: 10.1109/ACCESS.2022.3176091]
- Ke J, Liu K, Sun Y X, Xue Y Y, Huang J X, Lu Y Z, Dai J, Chen Y B, Han X D, Shen Y Q and Shen D G. 2023. Artifact detection and restoration in histology images with stain-style and structural preservation. *IEEE Transactions on Medical Imaging*, 42(12): 3487-3500 [DOI: 10.1109/TMI.2023.3288940]
- Khan S, Naseer M, Hayat M, Zamir S W, Khan F S and Shah M. 2022. Transformers in vision: a survey. *ACM Computing Surveys (CSUR)*, 54(10s): #200 [DOI: 10.1145/3505244]
- Kothari S, Phan J H and Wang M D. 2013. Eliminating tissue-fold artifacts in histopathological whole-slide images for improved image-based prediction of cancer grade. *Journal of Pathology Informatics*, 4(1): #22 [DOI: 10.4103/2153-3539.117448]
- Li J X, Sun J, Li C and Ahmad B. 2022. A MHA-based integrated diagnosis and segmentation method for COVID-19 pandemic. *Journal of Image and Graphics*, 27(12): 3651-3662 (李金星, 孙俊, 李超, Ahmad B. 2022. 融合多头注意力机制的新冠肺炎联合诊断与分割. *中国图象图形学报*, 27(12): 3651-3662) [DOI: 10.11834/jig.211015]
- Liang Z Y, Wang Q Y, Liao H W, Zhao M, Lee J, Yang C, Li F Y and Ling D S. 2021. Artificially engineered antiferromagnetic nano-

- probes for ultra-sensitive histopathological level magnetic resonance imaging. *Nature Communications*, 12(1): #3840 [DOI: 10.1038/s41467-021-24055-2]
- Mehta S and Rastegari M. 2022. MobileViT: light-weight, general-purpose, and mobile-friendly vision transformer//*Proceedings of the 10th International Conference on Learning Representations*. [s.l.]: OpenReview.net
- Najjar F H, Khudhair K T, Khudhair Z N, Alwan H H and Al-Khaykan A. 2023. Acute lymphoblastic leukemia image segmentation based on modified HSV model. *Journal of Physics: Conference Series*, 2432(1): #012020 [DOI: 10.1088/1742-6596/2432/1/012020]
- Palokangas S, Selinummi J and Yli-Harja O. 2007. Segmentation of folds in tissue section images//*Proceedings of the 29th Annual International Conference of 2027 IEEE Engineering in Medicine and Biology Society*. Lyon, France: IEEE: 5641-5644 [DOI: 10.1109/IEMBS.2007.4353626]
- Pizarro R, Assemblal H E, Jegathambal S K B, Jubault T, Antel S, Arnold D and Shmuel A. 2023. Deep learning, data ramping, and uncertainty estimation for detecting artifacts in large, imbalanced databases of MRI images. *Medical Image Analysis*, 90: #102942 [DOI: 10.1016/j.media.2023.102942]
- Sandler M, Howard A, Zhu M L, Zhmoginov A and Chen L C. 2018. MobileNetV2: inverted residuals and linear bottlenecks//*Proceedings of 2018 IEEE Conference on Computer Vision and Pattern Recognition*. Salt Lake City, USA: IEEE: 4510-4520 [DOI: 10.1109/CVPR.2018.00474]
- Selvaraju R R, Cogswell M, Das A, Vedantam R, Parikh D and Batra D. 2017. Grad-CAM: visual explanations from deep networks via gradient-based localization//*Proceedings of 2017 IEEE International Conference on Computer Vision*. Venice, Italy: IEEE: 618-626 [DOI: 10.1109/ICCV.2017.74]
- Shakarami A, Nicolè L, Terreran M, Tos A P D and Ghidoni S. 2023. TCNN: a transformer convolutional neural network for artifact classification in whole slide images. *Biomedical Signal Processing and Control*, 84: #104812 [DOI: 10.1016/j.bspc.2023.104812]
- Shakhawat H M, Nakamura T, Kimura F, Yagi Y and Yamaguchi M. 2020. Automatic quality evaluation of whole slide images for the practical use of whole slide imaging scanner. *ITE Transactions on Media Technology and Applications*, 8(4): 252-268 [DOI: 10.3169/mta.8.252]
- Shi D. 2024. TransNeXt: robust foveal visual perception for vision transformers//*Proceedings of 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Seattle, USA: IEEE: 17773-17783 [DOI: 10.1109/CVPR52733.2024.01683]
- Sun J M, Shen Z H, Wang Y A, Bao H J and Zhou X W. 2021. LoFTR: detector-free local feature matching with transformers//*Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Nashville, USA: IEEE: 8918-8927 [DOI: 10.1109/CVPR46437.2021.00881]
- Tellez D, Litjens G, Bándi P, Bulten W, Bokhorst J M, Ciompi F and Van Der Laak J. 2019. Quantifying the effects of data augmentation and stain color normalization in convolutional neural networks for computational pathology. *Medical Image Analysis*, 58: #101544 [DOI: 10.1016/j.media.2019.101544]
- Xia C F, Dong X S, Li H, Cao M M, Sun D Q, He S Y, Yang F, Yan X X, Zhang S L, Li N and Chen W Q. 2022. Cancer statistics in China and United States, 2022: profiles, trends, and determinants. *Chinese Medical Journal*, 135(5): 584-590 [DOI: 10.1097/CM9.0000000000002108]
- Xu J, Pan Y, Pan X L, Hoi S, Yi Z and Xu Z L. 2023. RegNet: self-regulated network for image classification. *IEEE Transactions on Neural Networks and Learning Systems*, 34(11): 9562-9567 [DOI: 10.1109/TNNLS.2022.3158966]
- Yang S, Lin G S, Jiang Q P and Lin W S. 2020. A dilated inception network for visual saliency prediction. *IEEE Transactions on Multimedia*, 22(8): 2163-2176 [DOI: 10.1109/TMM.2019.2947352]
- Zhao Y L, Ding W L, You Q H, Zhu F L, Zhu X J, Zheng K and Liu D D. 2023. Classification of whole slide images of breast histopathology based on spatial correlation characteristics. *Journal of Image and Graphics*, 28(4): 1134-1145 (赵樱莉, 丁维龙, 游庆华, 朱峰龙, 朱筱婕, 郑魁, 刘丹丹. 2023. 融合空间相关性特征的乳腺组织病理全切片分类. *中国图象图形学报*, 28(4): 1134-1145) [DOI: 10.11834/jig.211133]
- Zhu L, Wang X J, Ke Z H, Zhang W and Lau R. 2023. BiFormer: vision transformer with bi-level routing attention//*Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Vancouver, Canada: IEEE: 10323-10333 [DOI: 10.1109/CVPR52729.2023.00995]
- Zhu W. 2024. Research and Development of Key Technologies of Detection System for Pathological Slide Artifacts in Scanning and Preparation Based on Deep Learning. Hangzhou: Zhejiang University of Technology (朱伟. 2024. 基于深度学习的病理扫描与制片伪影检测系统关键技术研发. 杭州: 浙江工业大学)

作者简介

丁维龙,男,教授,博士生导师,主要研究方向为医疗人工智能与虚拟仿真。E-mail: wlding@zjut.edu.cn

汪春年,通信作者,女,主任医师,硕士生导师,主要研究方向为人工智能技术在肿瘤病理诊断中的应用。

E-mail: niana217@126.com

邓戎戎,男,硕士研究生,主要研究方向为数字病理图像处理。E-mail: isdengrr@163.com

徐利锋,男,讲师,硕士生导师,主要研究方向为虚拟仿真与医疗人工智能。E-mail: lfxu@foxmail.com

祝行琴,女,主管技师,主要研究方向为病理技术 HE 染色制片质量控制。E-mail: zhu20061050223@163.com

郑魁,男,工程师,主要研究方向为医疗智能系统管理。

E-mail: zhengkuimail@126.com