

联合多视角Transformer编码与在线融合互学习的 乳腺癌病理图像分类模型

李广丽¹, 叶艺源¹, 吴光庭¹, 李传秀¹, 吕敬钦², 张红斌²

(1. 华东交通大学信息工程学院, 江西南昌 330013; 2. 华东交通大学软件学院, 江西南昌 330013)

摘要: 乳腺癌是女性最常见的癌症. 单一网络在乳腺癌病理图像分类中存在缺陷, 卷积神经网络无法提取全局上下文, 而Transformer不能准确描述局部细节. 本文提出联合多视角Transformer编码与在线融合互学习的乳腺癌病理图像分类模型(Multi-View Transformer Online Fusion Mutual Learning, MVT-OFML). 采用ResNet-50(Residual Network-50)提取图像局部特征, 设计多视角Transformer编码模块, 捕获图像中全局上下文; 联合Logits和中间特征层构建OFML框架, 实现ResNet-50与多视角Transformer编码模块间双向传递知识, 使2个网络优势互补以完成乳腺癌病理图像分类. 实验表明, 在BreakHis和BACH数据集上, MVT-OFML的准确率比最强基线分别提升0.90%和2.26%, F_1 均值比最强基线分别提升4.75%和3.21%.

关键词: 乳腺癌; 病理图像分类; 多视角Transformer; 卷积神经网络; 在线融合互学习

基金项目: 国家自然科学基金(No.62161011); 江西省重点研发计划重点项目(揭榜挂帅)(No.20223BBE51036); 江西省自然科学基金(No.20212BAB202006); 江西省社科规划项目(No.22TQ01); 江西省教育厅科技项目(No.GJJ200628, No.GJJ2200639); 江西省研究生创新基金(No.YC2022-s546)

中图分类号: TP391 **文献标识码:** A **文章编号:** 0372-2112(2024)07-2369-13

电子学报URL: <http://www.ejournal.org.cn>

DOI: 10.12263/DZXB.20230305

Breast Cancer Pathological Image Classification Model via Combining Multi-View Transformer Coding and Online Fusion Mutual Learning

LI Guang-li¹, YE Yi-yuan¹, WU Guang-ting¹, LI Chuan-xiu¹, LÜ Jing-qin², ZHANG Hong-bin²

(1. School of Information Engineering, East China Jiaotong University, Nanchang, Jiangxi 330013, China;

2. School of Software, East China Jiaotong University, Nanchang, Jiangxi 330013, China)

Abstract: Breast cancer is the most common cancer in women. The single neural network used in breast cancer pathological image classification has the following defects: the convolutional neural network (CNN) lacks the ability to extract global context information while the Transformer lacks the ability to depict local lesion details. To alleviate the problem, a novel model, named multi-view Transformer coding and online fusion mutual learning (MVT-OFML), is proposed for breast cancer pathological image classification. First, ResNet-50 is employed to extract local features in images. Then, a new multi-view Transformer (MVT) coding module is designed to capture the global context information. Finally, a novel online fusion mutual learning (OFML) framework based on the Logits and middle feature layers is designed to implement the bi-directional knowledge transfer between ResNet-50 and the MVT coding module. This makes the two networks complement each other to complete breast cancer pathological image classification. Experiments validated on BreakHis and BACH show that compared to the best baseline, the performance improvements of accuracy are 0.90% and 2.26%, respectively, whereas the corresponding improvements of average F_1 score are 4.75% and 3.21%, respectively.

Key words: breast cancer; pathological image classification; multi-view Transformer; convolution neural network; online fusion mutual learning

Foundation Item(s): National Natural Science Foundation of China (No.62161011); Key Research and Development Plan of Jiangxi Province (Key Project) (No.20223BBE51036); Natural Science Foundation of Jiangxi Province (No.20212BAB202006); Humanity and Social Science Foundation of Jiangxi Province (No.22TQ01); Science and Technol-

ogy Projects of Jiangxi Provincial Department of Education (No.GJJ200628, No.GJJ2200639); Graduate Innovation Foundation Project of Jiangxi Province (No.YC2022-s546)

1 引言

乳腺癌已成为全球女性第一大癌症^[1]. 计算病理学为癌症诊断和精准医学提供了光明的未来. 病理学家在临床实践中手动评估 H&E (Hematoxylin-Eosin) 染色组织切片完成诊断过程^[2]. 随着全球癌症病例数量激增, 病理学家的工作量也迅速增加, 人工分析已成为诊断效率的瓶颈. 因此, 计算机辅助病理学分析是应对这一问题的有力手段.

病理图像分类依赖细胞核形状和排列, 细胞核的形态变化是病理学家判断患者是否患癌的重要标准^[3]. 在临床上, 病理学家须综合考虑图像中局部特征和上下文全局信息, 方可做出诊断决定. 例如, 从核的多形性和不规则排列中提取局部形态信息, 从腺体组织结构中提取全局上下文信息. 近年来, 研究人员尝试构建各种病理图像分类模型. 早期, 研究者基于手工特征和传统分类器完成分类, 这些方法非常依赖病理学家的标注. 接着, 卷积神经网络 (Convolutional Neural Network, CNN)^[4] 被广泛用于乳腺癌病理学图像分类. CNN 模型具备出色的表征能力, 但它不擅于提取全局上下文信息, 会丢失病理图像中的部分组织结构信息. 相反, 最近的 Transformer^[5] 基于多头自注意力建模长距离依赖, 捕获图像中全局上下文信息. 然而, Transformer 缺乏 CNN 的归纳偏置, 更依赖大数据集来完成训练. 较少工作考虑联合 Logits 层、中间特征层来挖掘异构的 CNN 和 Transformer 模型间的互补信息. 为此, 本文提出联合多视角 Transformer 编码与在线融合互学习的乳腺癌病理图像分类模型 (Multi-View Transformer Online Fusion Mutual Learning, MVT-OFML), 采用 CNN 与 Transformer 设计双路网络结构, 充分发挥异构 CNN、Transformer 的优势, 提高乳腺癌病理图像分类精度. 本文贡献如下:

(1) 设计 1 种联合 CNN 和 Transformer 的双路网络结构, 提取病理图像中互补的局部特征和全局特征, 以改善乳腺癌病理图像分类精度.

(2) 设计 1 个新的多视角 Transformer 编码模块, 通过融合不同视角的编码信息捕获乳腺癌病理图像中全局上下文特征.

(3) 构建在线融合互学习方法, 联合 Logits 输出层和中间特征层, 深入挖掘异构的 CNN 和 Transformer 模型之间的互补性.

2 相关工作

2.1 传统方法

早期研究先对病理图像进行预处理和分割, 提取特征完成分类. Sansone 等人^[6] 采用灰度共生矩阵训练分类器, 以判别良恶性病理图像. Wang 等人^[7] 结合多尺度区域特征和小波变换算法对细胞核进行检测和分割, 采用支持向量机 (Support Vector Machine, SVM)^[8] 完成乳腺癌病理图像分类. 范虹等人^[9] 利用复小波域双变量模型完成乳腺肿瘤图像的去噪和分割. 此类方法虽然具备良好的生物学解释, 且在特定任务中效果良好, 但它需要手动设计特征, 模型泛化能力有限.

2.2 基于 CNN 的方法

CNN 利用卷积和池化操作提取图像特征, 无须任何先验知识. 诸多 CNN 架构, 如 ResNet-50^[10]、融合类模型^[11] 都被用于乳腺病理图像分析. Cruz-Roa 等人^[12] 设计 1 个 3 层 CNN 来识别乳腺浸润性导管癌, 结果表明其性能优于传统方法. Zhou 等人^[13] 提出基于分辨率自适应网络 (Resolution Adaptive Network, RANet) 的方法, 采用 ADSVM (Anomaly Detection with a Support Vector Machine) 方法进行异常检测. 为减少对医学专家标注的依赖, Wang 等人^[14] 试图从弱标记数据中学习出可迁移特征, 并在乳腺癌病理图像分类中获得有竞争力的性能. 上述研究推动了乳腺癌病理图像分类的研究, 但由于采用卷积和池化操作, CNN 会丢失病理图像中的全局上下文信息.

2.3 基于 Transformer 方法

Transformer 基于多头自注意力机制捕获图像中全局上下文信息已广泛应用于视觉图像领域^[15]. Parmar 等人^[16] 用自注意力机制关注局部区域, 增加每层感受野. Alotaibi 等人^[17] 提出 ViT-DeiT (Vision Transformer and Data-efficient image Transformer) 模型, 完成乳腺癌病理图像分类. Zou 等人^[18] 设计双路 DCET-Net (Dual-stream Convolution Expanded Transformer Network) 模型, 同时提取局部特征和全局特征, 完成乳腺癌病理图像分类. 由于 Transformer 缺乏归纳偏置, 其局部特征提取能力不强. Tummala 等人^[19] 使用改进的 Swin Transformer 实现乳腺癌多分类, 其性能优于标准 ViT. 但这些研究都很少考虑联合 Logits 层、中间特征层来挖掘异构的 CNN 和 Transformer 模型间的互补信息.

2.4 多实例学习方法

多实例学习 (Multiple Instance Learning, MIL) 通过

将病理图像划分成实例并构建成包,将病理图像分类问题转化为弱监督的MIL问题,以降低模型对标注的依赖.MIL在处理高分辨率组织病理图像上表现良好,常被用于全切片图像(Whole Slide Image, WSI)的病理诊断.Sudharshan 等人^[20]提出了弱监督学习框架,利用多实例神经网络(Multiple Instance Neural Network, MINN)完成乳腺癌组织病理图像分类. Ilse 等人^[21]提出基于注意力的聚合操作,在MINN中加入单一注意力模块来学习每个实例额外的贡献信息,完成组织病理图像分类. Shao 等人^[22]设计基于Transformer的多实例学习(Multiple Instance Learning, TransMIL)框架,学习不同实例之间的相关性,完成高分辨率组织病理图像分类. Li 等人^[23]设计多视角注意力引导的多实例检测网络(Multi-View Attention-guided Multiple Instance Detection Network, MVA-MIDN),完成高分辨率乳腺癌组织病理图像分类. 然而,目前的MIL方法无法对实例进行精准筛选,制约了模型性能提升.

综上, CNN、Transformer 等深度学习模型在乳腺癌病理图像分类中发挥了重要作用. 然而,乳腺癌病理图像分类既依赖局部细胞核特征,也需要乳腺组织结构的全局上下文信息. 现有工作未充分利用 CNN 和 Transformer 各自优势. 为此,设计1种联合 CNN 和 Transformer 的双路网络结构,基于 ResNet-50 提取乳腺癌病理图像中的局部特征;设计多视角 Transformer 编码模块,更好地提取乳腺癌病理图像中全局上下文信息;构建在线融合互学习方法,搭建 ResNet-50 与 Transformer 之间的互学习通道,即设计集成分类器和融合分类器,联合 Logits 输出层和中间特征层完成模型间的互学习,深入挖掘异构 CNN 和 Transformer 之间的相关性,完成乳腺癌病理图像分类.

3 模型实现

MVT-OFML 模型如图 1 所示. 该模型是 1 个端到端的双路网络,包括 1 个 CNN 主干网络和 1 个 Transformer 主干网络.

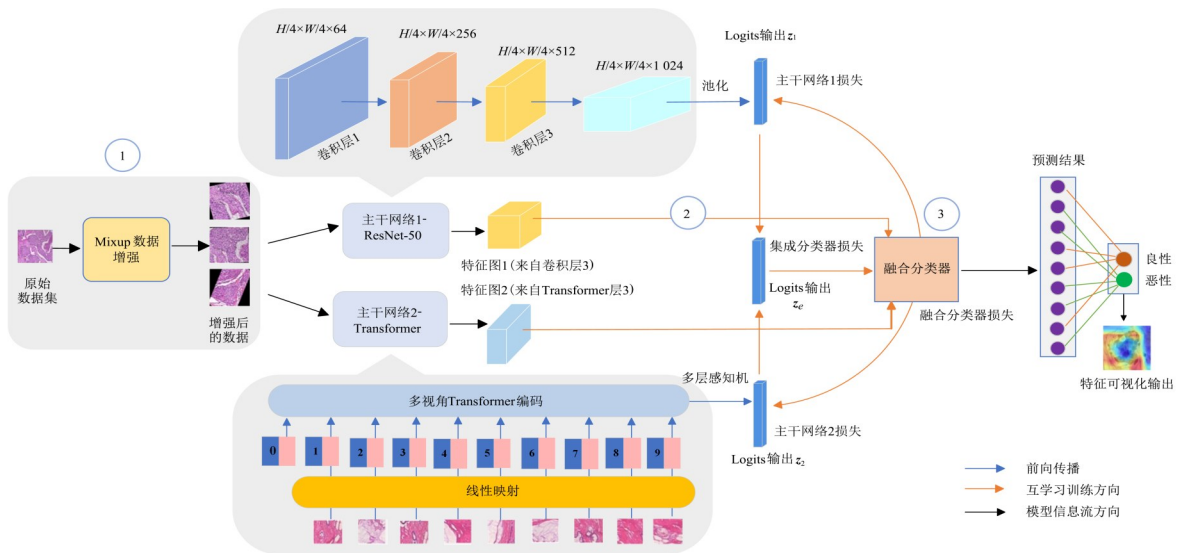


图 1 MVT-OFML 模型总体图

模型执行步骤如下:

第 1 阶段,采用 Mixup^[24] 获得足够多高质量训练图像. Mixup 在增强中利用相关标签线性插值来扩展训练分布,得到高质量图像样本,最大限度保留原病理图像中的重要信息.

第 2 阶段,构建包含 CNN 和 Transformer 的双路网络结构. 设计多视角 Transformer 编码器,更好地提取并融合图像中全局上下文特征;选取 ResNet-50 以捕捉图像中局部特征. 设计集成分类器将来自 2 个异构网络的 Logits 进行集成,为后续激励融合分类器做好准备.

第 3 阶段,设计在线融合互学习方法. 选取 ResNet-50

的卷积层 3 作为特征图 1,选取多视角 Transformer 的层 3 作为特征图 2. 联合特征图 1 和特征图 2 为中间特征,将其作为融合分类器的输入. 构建融合分类器,对 2 个主干网络的中间层特征进行自适应融合,充分挖掘异构网络间的隐式互补性. 如图 2 所示,融合分类器首先使用 1 个自适应平均池化对输入的特征图 1 (512, 28, 28) 和特征图 2 (512, 14, 14) 进行池化. 中间特征经过自适应池化后,每个特征图的宽度和高度均变为 1;然后,采用拼接操作完成特征融合,该融合特征图的通道数是参与融合的特征图通道数之和,即 1 024;接着,通过逐点卷积改变融合特征图的通道数为 C (C 设置为数据

集 的 类 别 数 量), 提 升 局 部 模 块 的 抽 象 表 示 能 力. 自 适 应 平 均 池 化 和 逐 点 卷 积 不 仅 保 证 了 融 合 分 类 器 可 自 适 应 匹 配 任 何 2 个 特 征 图, 还 避 免 了 传 统 融 合 方 法 需 调 制 大 量 参 数 的 缺 陷. 同 时, 将 集 成 分 类 器 的 输 出 传 递 到 融 合 分 类 器 中, 即 通 过 来 自 Logits 层 的 信 息 激 励 融 合 分 类

器. 接 着, 将 融 合 分 类 器 中 的 信 息 回 传 给 2 个 主 干 网 络, 即 联 合 Logits 输 出 层 和 中 间 特 征 层 进 一 步 增 强 主 干 网 络 之 间 的 病 理 知 识 交 流, 通 过 相 互 学 习 达 到 提 高 各 主 干 网 络 性 能 的 目 的. 最 后, 计 算 融 合 分 类 器 的 类 概 率 输 出, 得 到 乳 腺 癌 病 理 图 像 分 类 结 果.

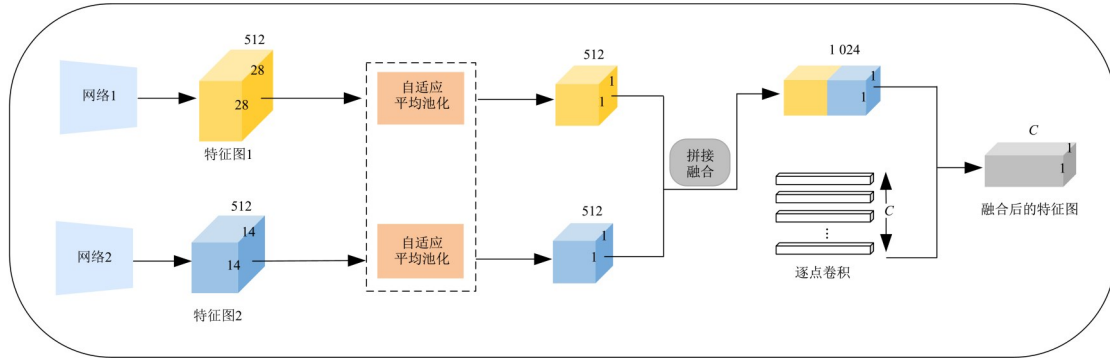


图 2 融合分类器

3.1 多视角 Transformer 编码

传统 Transformer 网络对乳腺癌病理图像中复杂的纹理、结构和空间信息捕捉不充分, 故设计多视角

Transformer 编码模块, 更好地刻画乳腺癌组织结构特征和全局上下文信息. 多视角 Transformer 编码模块如图 3 所示.

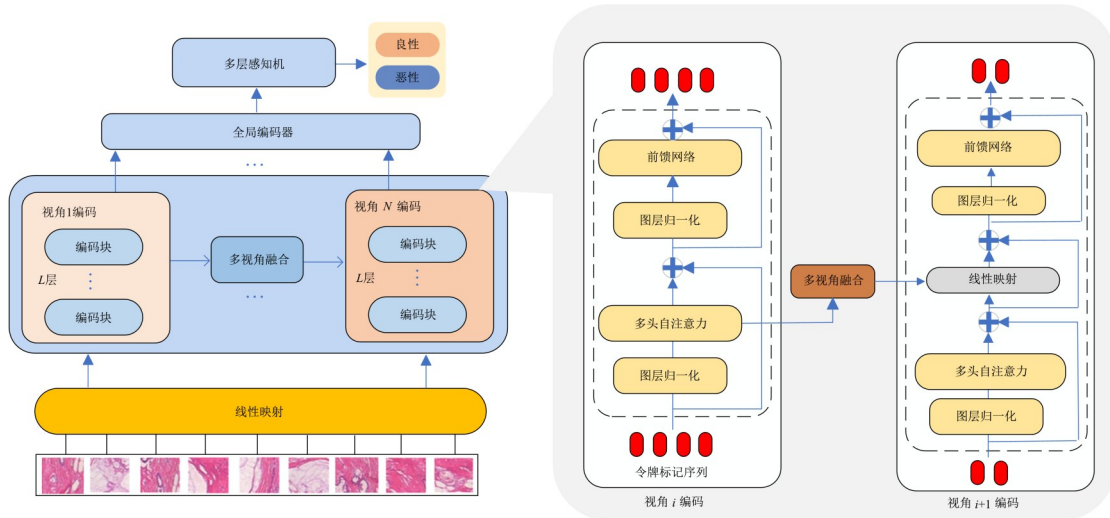


图 3 多视角 Transformer 编码模块实现细节

首先, 将输入的乳腺癌病理图像表示为 $x \in \mathbb{R}^{H \times W \times C}$, 其中, H 表示图像高度, W 表示图像宽度, C 表示图像维度. 图像经过剪裁后变成 $N = (H \times W) / P$ 个尺寸为 $P \times P \times C$ 的图像块, P 为剪裁后图像块的高度和宽度. 将这些图像块展开操作成一维向量 $x_{\text{patch}} \in \mathbb{R}^{N \times (P \times P \times C)}$, 这些一维向量经过线性映射后得到多个带有病理特征信息的令牌标记序列 z^0 , 这个序列前面带 1 个可学习的类标记 $z_{\text{class}} \in \mathbb{R}^d$ 和 1 个位置嵌入 $p \in \mathbb{R}^{(N+1) \times d}$, 其中, d 是输入嵌入向量的维度, E 表示实现线性映射的矩阵, 故令牌标记序列如式 (1) 所示:

$$z^0 = [z_{\text{class}} \quad x_{\text{patch}}^1 E \quad x_{\text{patch}}^2 E \quad \cdots \quad x_{\text{patch}}^N E] + p \quad (1)$$

令牌标记序列 z^0 被 N 个视角 Transformer 编码器处理, 每个编码器包含 L 层, 每 1 层 l 是按顺序排列. 在多视角 Transformer 编码器中包括多视角融合模块和全局编码器模块. 如图 3 所示, 多视角融合模块融合其他视角编码信息, 并通过交叉视角注意力 (Cross View Attention, CVA) 多维度交互处理标记序列, 提取交互后的病理信息. CVA 如式 (2) 所示:

$$\text{CVA}(x, y) = \text{Softmax}\left(\frac{W^Q x W^K y^T}{\sqrt{d_k}}\right) W^V y \quad (2)$$

其中, W^Q, W^K 和 W^V 是自注意力层通过映射矩阵对序列投影得到的 3 个向量矩阵; y 表示可学习的参数矩阵; 基

于CVA得到 $\mathbf{z}^{(i)}$,它表示第*i*个视觉编码器序列的特征, $\mathbf{z}^{(i+1)}$ 表示第*i+1*个视觉编码器序列的特征信息, \mathbf{W}^{proj} 表示投影向量.

$$\mathbf{z}^i = \text{CVA}(\mathbf{z}^{(i)}, \mathbf{W}^{\text{proj}} \mathbf{z}^{(i+1)}) \quad (3)$$

CVA会根据病理特征的重要性排序,并融合2个相邻视角输出的特征,获得更为丰富的病理特征.由于不同视角之间的隐藏维度不同,需将不同维度的视角特征投影到相同维度,投影后再对2者进行融合.最后,经过多视角编码交互融合后的标记信息被输入到全局编码器,全局编码器、多层感知器的计算如式(4)和式(5)所示:

$$\mathbf{y}^i = \text{MSA}(\text{LN}(\mathbf{z}^{i-1})) + \mathbf{z}^{i-1} \quad (4)$$

$$\mathbf{z}^i = \text{MLP}(\text{LN}(\mathbf{y}^i)) + \mathbf{y}^i \quad (5)$$

MSA (Multi-head Self Attention) 是多头注意力机制^[6], LN 表示图层归一化^[6], MLP (MultiLayer Perceptron) 是多层感知机制,它包括2个线性投影,由非线性激活函数 GELU (Gaussian Error Linear Units) 隔开排列.从多层感知机中获得最终分类信息.

3.2 在线融合互学习

在离线KD中,当教师-学生网络容量差距增加时,学生网络的性能下降.与离线KD相比,在线KD具有明显优势,教师-学生网络能同时实施蒸馏操作.较少工作联合 Logits 层、中间特征层来挖掘异构 CNN 和 Transformer 模型间的互补信息, MVT-OFML 结合2个网络的最后1层 Logits 输出和异构特征图的融合信息,充分挖掘异构网络之间的病理知识,共同监督网络训练,最终建立2个异构网络之间的互学习关系.

在线融合学习中,使用2个异构主干网络完成相互学习,本文将 ResNet-50 作为主干网络1,将3.1节多视角编码Transformer作为主干网络2.令*m*类中的*N*张病理图像样本为 $X = \{\mathbf{x}_i\}$,其样本标签表示为 $Y = \{y_i\}$,其中, $y_i \in \{1, 2, \dots, m\}$, $i = 1, 2, 3, \dots, N$. ResNet-50 和多视角编码Transformer输出样本 \mathbf{x}_i 的*m*类的软化概率分别定义为式(6)和式(7):

$$\mathbf{p}_1^m(\mathbf{x}_i, T) = \frac{\exp(\mathbf{z}_1^m/T)}{\sum_{m=1}^M \exp(\mathbf{z}_1^m/T)} \quad (6)$$

$$\mathbf{p}_2^m(\mathbf{x}_i, T) = \frac{\exp(\mathbf{z}_2^m/T)}{\sum_{m=1}^M \exp(\mathbf{z}_2^m/T)} \quad (7)$$

其中, \mathbf{z}_1^m 是 ResNet-50 的 Softmax 层特征, \mathbf{z}_2^m 是多视角Transformer的 MLP 层特征,*T*为蒸馏温度,*T*越大,对应的病理类概率分布越平滑.因此,蒸馏温度使概率比以前更软化,有助于从其他相关病理类别中挖掘出更多有价值的病理知识,提升乳腺癌病理图像分类精度.由于病理图像的复杂性,在嵌入融合互学习框架中还包

含1个集成分类器,用于从2个主干子网络中学习更有价值的病理知识.如图1所示,集成分类器的 Logits 输出计算如下:

$$\mathbf{z}_e = \frac{\mathbf{z}_1 + \mathbf{z}_2}{2} \quad (8)$$

集成分类器的交叉熵损失计算如下:

$$L_{\text{ensemble}}^f = \sum_{i=1}^N \sum_{m=1}^M I(y_i, m) \log(\mathbf{p}_e^m(\mathbf{x}_i, 1)) \quad (9)$$

在每次训练中,集成分类器将 ResNet-50 和多视角Transformer编码器中的病理知识迁移到融合分类器中,此过程称为集成病理学知识转移 (Ensemble Pathological Knowledge Transfer, EPKT),其损失使用KL散度计算如下:

$$L_{\text{EPKT}} = D_{\text{KL}}(\mathbf{p}_e \| \mathbf{p}_f) \quad (10)$$

对于多分类任务, ResNet-50 的目标损失函数定义为预测标签与真实标签之间的交叉熵损失 L_{net1}^1 .在线融合互学习框架中,2个主干网络的病理知识与融合分类器的病理知识互相传递,促使主干网络及融合分类器的性能均得以提升.使用融合分类器的后验概率 p_f 提供训练经验,采用KL散度量 \mathbf{p}_1 和 \mathbf{p}_f 之间的差异(即 $D_{\text{KL}}(\mathbf{p}_1 \| \mathbf{p}_f)$),具体公式如下:

$$L_{\text{net1}}^1 = \sum_{i=1}^N \sum_{m=1}^M I(y_i, m) \log(\mathbf{p}_1^m(\mathbf{x}_i, 1))$$

$$L_{\text{FPKT}}^1 = D_{\text{KL}}(\mathbf{p}_1 \| \mathbf{p}_f) = \sum_{i=1}^N \sum_{m=1}^M \mathbf{p}_e^m(\mathbf{x}_i) \log \frac{\mathbf{p}_f^m(\mathbf{x}_i)}{\mathbf{p}_1^m(\mathbf{x}_i)}$$

$$L_1 = L_{\text{net1}}^1 + T^2 \times L_{\text{FPKT}}^1 \quad (11)$$

这里,*I*表示一个指标函数,其定义为

$$I(y_i, m) = \begin{cases} 1, & y_i = m \\ 0, & y_i \neq m \end{cases} \quad (12)$$

L_1 为 ResNet-50 的总损失函数,由 ResNet-50 的 L_{net1}^1 损失和融合分类器 L_{FPKT}^1 损失共同组成. L_{EPKT} 损失鼓励网络生成更有效的特征图,并提高融合性能.融合分支将中间特征层隐含的病理知识回传给每个主干网络,为更好地训练每个主干网络,此过程称为融合病理知识转移 (Fusion Pathological Knowledge Transfer, FPKT),将融合分类器的软化概率分布传递到每个主干网络.联合主干网络 Logits 输出和来自融合分类器的病理知识提高最终性能.

同理,多视角Transformer编码器的总损失函数表示为式(13):

$$L_{\text{net2}}^2 = \sum_{i=1}^N \sum_{m=1}^M I(y_i, m) \log(\mathbf{p}_2^m(\mathbf{x}_i, 1))$$

$$L_{\text{FPKT}}^2 = D_{\text{KL}}(\mathbf{p}_2 \| \mathbf{p}_f) = \sum_{i=1}^N \sum_{m=1}^M \mathbf{p}_e^m(\mathbf{x}_i) \log \frac{\mathbf{p}_f^m(\mathbf{x}_i)}{\mathbf{p}_2^m(\mathbf{x}_i)}$$

$$L_2 = L_{\text{net2}}^2 + T^2 \times L_{\text{FPKT}}^2 \quad (13)$$

根据式(13)将 T^2 乘以相应的融合病理知识迁移损失,因为软化的概率分布被缩放 $1/T^2$.最后,整个模型的损失为

$$L_f = L_1 + L_2 + T^2 \times L_{ensemble}^f \quad (14)$$

2个主干网络和融合分支都是同时训练的,互学习策略嵌入在每个小批量的网络更新步骤中并贯穿整个训练过程,该过程从标签中挖掘监督信息.同时,融合分支将融合后的病理知识回传到每个主干网络,这将激励ResNet-50和多视角Transformer编码器从异构的中间特征层学习更多有价值的病理知识.而集成分类器将各主干网络产生的深层病理知识回传到融合分支,激励融合分类器的训练.故2个主干网络的Logits输出和特征蒸馏知识都被用来提升融合分类器的判别能力,2个主干网络和融合分类器相互鼓励、互为补充,为病理图像分类挖掘足够知识.

4 实验结果和详细分析

4.1 数据集

本文使用2个公共数据集来验证模型的有效性和鲁棒性.数据集详细如下所示:

(1)BreakHis:它拥有4种不同的放大倍数,包括40倍、100倍、200倍和400倍,每张乳腺组织病理学图像有3个RGB通道,每个RGB通道是8位颜色深度.每张病理图像为700×460像素.BreakHis数据集包括2408张良性肿瘤图像和5429张恶性肿瘤图像.每张图像还被标记为以下8种亚型之一:乳腺病(A)、纤维腺瘤(F)、叶状瘤(PT)、管状腺瘤(TA)、导管癌(DC)、小叶癌(LC)、黏液癌(MC)和乳头状癌(PC).本文分别执行BreakHis数据集2分类和8分类实验.使用Mixup算法对原始数据集进行4倍增强,降低模型过拟合风险.表1是Mixup增强后的BreakHis数据集分布情况.

(2)BACH:BACH根据每个图像中的主要癌症类型,将显微图像标记为正常(Normal)、良性(Benign)、原位癌(Insitu)和浸润性癌(Invasive)4个类别.病理学家将正常和良性标签视为健康组织,将原位和浸润性标签视为癌组织.该数据集包含400张RGB3通道图像,但它未对图像的放大倍数进行说明.每张图像为2048×1536像素,像素尺度为 $0.42 \mu\text{m} \times 0.42 \mu\text{m}$.表2给出了Mixup增强后的BACH数据集分布情况.

表1 经过Mixup数据增强之后的BreakHis数据集分布情况

数据集	训练		测试	
	良性图像	恶性图像	良性图像	恶性图像
原始数据集	1 736	3 800	744	1 629
增强数据集	6 944	15 200	744	1 629

表2 经过Mixup数据增强之后的BACH数据集分布情况

数据集	训练				测试			
	正常	良性	原位癌	浸润性癌	正常	良性	原位癌	浸润性癌
原始数据集	70	70	70	70	30	30	30	30
增强数据集	280	280	280	280	30	30	30	30

综上,2个数据集在数量、分辨率等方面具有较大差异,围绕其开展实验有助于更好地验证MVT-OFML模型的鲁棒性.

4.2 实现细节

本文模型基于PyTorch中实现,并在包含2块GTx3080Ti的高性能服务器上完成全部实验.为了获得最佳性能,使用Adam优化器,初始学习率和动量分别设置为0.01、0.8,权重衰减为 5×10^{-4} .损失函数中的蒸馏温度 T 设置为4.实验中设置可学习的位置嵌入,图像经过Mixup增强后直接变换尺寸为224×224输入模型,并以64的批次大小训练300轮次.训练集和测试集设置为7:3.为了全面评估模型,本文使用准确率、AUC、精确率、召回率和 F_1 分数5个指标.

4.3 实验结果

4.3.1 在BreakHis数据集上的实验结果

首先,将MVT-OFML模型在BreakHis数据集上进行2分类和8分类实验,与主流乳腺癌病理图像分类模型作为基线进行比较.使用准确率、精确率、召回率和 F_1 分数这4个指标来衡量不同模型或方法的性能,在2分类上性能对比的详细结果如表3(加粗数据为最佳结果)所示.

从表3中可看出:MVT-OFML模型在40倍2分类任务中达到99.77%的最佳准确率,与最强模型ViT-DeiT相比,MVT-OFML模型在4个指标中均取得最优,从提升情况可看出,MVT-OFML模型与最优基线相比高出0.25%~1.53%,这表明MVT-OFML模型可有效改善乳腺癌病理图像分类性能.更重要的是,在不同放大倍数下,MVT-OFML模型获取了更平衡性能,这得益于多视角Transformer编码器能独立于放大倍数,提取更鲁棒、有效的特征,最终提高分类性能.从4种放大倍数的平均准确率来看,IDSNet(Intrusion Detection System Network)、DCET-Net、RANet-ADSVM、ViT-DeiT模型分别取得的平均准确率为:90.98%、98.79%、94.29%和98.72%,MVT-OFML模型则取得了99.64%的平均准确率,MVT-OFML模型其他指标(精确率、召回率和 F_1 分数)的平均性能都最佳,有力证明了模型的稳健性和鲁棒性.

与单路模型IDSNet比较,MVT-OFML模型在40倍下的准确率、精确率、召回率和 F_1 分数上分别提升了8.27%、9.28%、8.71%和9.23%,这得益于MVT-OFML模

表3 BreakHis数据集上MVT-OFML模型与主流方法的2分类性能对比

主干网络或方法	发表年份	放大倍数/倍	准确率/%	精确率/%	召回率/%	F_1 分数/%
IDSNet ^[25]	2020	40	91.50	90.55	91.00	90.54
		100	90.40	91.23	90.56	90.78
		200	95.30	95.33	95.66	95.39
		400	86.70	89.35	88.42	89.47
DCET-Net ^[18]	2021	40	99.00	99.47	97.38	98.41
		100	98.08	94.79	98.91	96.81
		200	99.34	97.66	97.82	98.82
		400	98.72	98.22	97.65	97.93
RANet-ADSV ^[13]	2022	40	91.96	93.83	94.91	94.36
		100	96.83	98.52	98.30	98.32
		200	98.05	98.92	99.15	99.13
		400	90.30	93.17	93.56	93.35
ViT-DeiT ^[17]	2022	40	99.43	99.38	99.46	99.40
		100	98.34	98.31	98.51	98.35
		200	98.27	98.32	98.27	98.23
		400	98.82	98.57	98.78	98.65
MVT-OFML	2023	40	99.77	99.75	99.71	99.77
		100	99.56	99.74	99.54	99.44
		200	99.76	99.65	99.43	99.62
		400	99.45	99.30	99.69	99.33
提升情况		40	0.34 ↑	0.37 ↑	0.25 ↑	0.37 ↑
		100	1.22 ↑	1.43 ↑	1.03 ↑	1.09 ↑
		200	1.40 ↑	1.33 ↑	1.16 ↑	1.39 ↑
		400	0.63 ↑	0.73 ↑	0.96 ↑	0.91 ↑

型在双路网络结构下充分利用异构网络之间的互补性病理知识. 与双路模型 ViT-DeiT 相比, MVT-OFML 在 400 倍下的准确率、精确率、召回率和 F_1 分数上分别提升了 0.63%、0.73%、0.91% 和 0.68%. ViT-DeiT 仅利用最后的 Logits 输出, 而中间层隐含的病理特征未充分利用. 相反, MVT-OFML 打通了异构 CNN 和 Transformer 模型间的“鸿沟”, 它联合 Logits 层、中间特征层深入挖掘异构 CNN 和 Transformer 模型间的互补信息, 充分发挥其优势, 提取更具判别性的病理图像特征. 综上, 与各类基线方法相比, 本文方法 MVT-OFML 是最优的.

此外, 图 4(a) 是 BreakHis 数据集上的 2 分类 ROC 曲线. 如图 4(a) 所示, MVT-OFML 模型的 ROC 曲线非常接近 (0, 1) 点, 并远离坐标轴的 45° 对角线, 模型的 AUC 值达到 0.998 7. 优异的 ROC 曲线图清楚表明了 MVT-OFML 模型充分发挥 CNN 和 Transformer 模型在各自特征学习中的优势, 异构模型之间的互补病理信息知识, 提取出更具判别性的病理图像特征来改善模型的各项性能指标, 并对乳腺癌病理图像的 2 分类任务是高效、实用的. 总体而言, MVT-OFML 模型在 4 种不同放大倍率上和不同性能评价指标中获得了较好性能, 进一步证明了 MVT-OFML

模型具有良好的泛化能力和强大的鲁棒性. 同时, 也说明 MVT-OFML 模型对乳腺癌病理图像分类是有效的.

相比 2 分类任务, 乳腺癌病理图像多分类任务更具挑战性. 将 MVT-OFML 模型在 BreakHis 数据集上进行 8 分类实验, 定量结果如表 4 (加粗数据为最佳结果) 所示. 同样, 使用准确率、精确率、召回率和 F_1 分数这 4 个指标来衡量不同方法的性能. 图 4(b) 是 BreakHis 数据集上得到的 8 分类 ROC 曲线.

从表 4 中可看出: 相比其他文献的最优基线 (BreaST-Net), MVT-OFML 模型在任何放大比例下都获得最佳分类性能, 各类指标均接近饱和. 例如, MVT-OFML 模型在 400 倍放大倍率下的 8 分类任务达到 99.63% 的最佳准确率. 同样, 从 4 种放大倍率的平均准确率来看, Deep-Net^[11]、AnoGAN^[26] (Anomaly detection with Generative Adversarial Networks)、BHC-Net^[27]、BreaST-Net^[28] 模型分别取得的平均精准度为: 93.08%、92.78%、93.29% 和 93.40%, MVT-OFML 模型取得了 99.36% 的最佳平均准确率, 且其他指标 (精确率、召回率和 F_1 分数) 的平均性能也都是 MVT-OFML 最佳. 这表明: 无论从最佳结果还是从平均性能来看, MVT-OFML 模型都是最优的, 有力证明了模型整体的稳健性和鲁棒性.

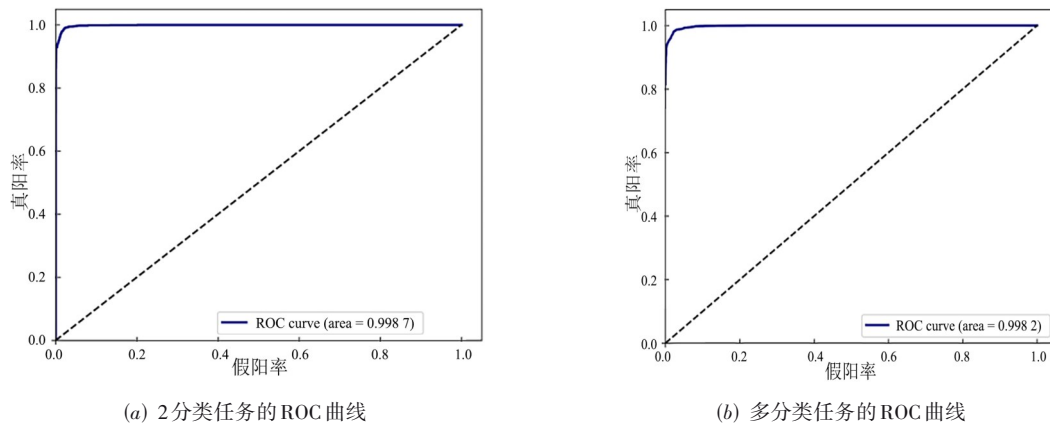


图4 MVT-OFML模型在BreakHis数据集上的ROC曲线

表4 MVT-OFML模型在BreakHis数据集上与其他主流方法的分类性能对比

主干网络或方法	发表年份	放大倍数/倍	准确率/%	精确率/%	召回率/%	F_1 分数/%
Deep-Net ^[11]	2020	40	94.43	95.25	95.55	95.39
		100	94.45	94.64	94.64	94.42
		200	92.27	90.71	92.24	91.42
		400	91.15	90.74	91.09	90.75
AnoGAN ^[26]	2021	40	99.15	99.64	99.46	99.78
		100	97.09	98.07	98.49	98.22
		200	87.58	88.19	92.82	90.62
		400	87.30	82.77	92.50	88.23
BHC-Net ^[27]	2022	40	94.71	95.25	95.55	95.39
		100	94.60	94.51	94.64	94.42
		200	92.35	90.71	92.24	91.42
		400	91.50	90.74	91.09	90.75
BreaST-Net ^[28]	2022	40	96.00	—	—	95.80
		100	92.60	—	—	92.40
		200	93.50	—	—	93.60
		400	91.50	—	—	93.20
MVT-OFML	2023	40	99.19	98.93	98.88	98.46
		100	99.05	97.44	98.90	97.77
		200	99.60	97.88	99.30	99.33
		400	99.63	96.19	99.54	98.45
提升情况		40	3.19 ↑	—	—	2.66 ↑
		100	6.45 ↑	—	—	5.37 ↑
		200	6.10 ↑	—	—	5.73 ↑
		400	8.31 ↑	—	—	5.25 ↑

注：“—”表示相应工作未提供实验结果。

与最佳基线 BreaST-Net 模型相比,在任何放大倍率下, MVT-OFML 都可获得更明显的性能改进. 例如,在 400 倍放大倍率下可以观察到 8.31% 的准确率改善,有助于提高分类模型的实用性. 更重要的是, MVT-OFML 模型在 200 倍放大倍率下实现了 5.73% 的 F_1 分数改进. 值得注意的是,与其他放大倍率相比,在 400 倍放大倍率下可获得更大性能改进, MVT-OFML 模型能很好处理大

放大倍率有 2 个关键因素: 首先,多视角 Transformer 编码器能够独立于放大倍率提取更鲁棒、有效的特征,从其他病理类别中学习有价值的信息,以应对复杂的多分类任务,取得更好性能; 其次, MVT-OFML 模型联合 Logits 层、中间特征层深入挖掘异构的 CNN 和 Transformer 模型间的互补信息,模拟真实病理诊断场景,充分发挥 CNN 和 Transformer 模型在各自特征学习中的优势,提取

乳腺癌病理图像中更具判别性的特征.

相比2分类任务,乳腺癌8分类更具挑战性,因此,在挑战性更高的任务上表现优异能更好说明模型的有

效性和健壮性. 联合 CNN 和 Transformer 的双路网络可捕获更多有价值的深层病理特征. 图 5 展示的是本文方法在 BreakHis 数据集上 8 分类的混淆矩阵.

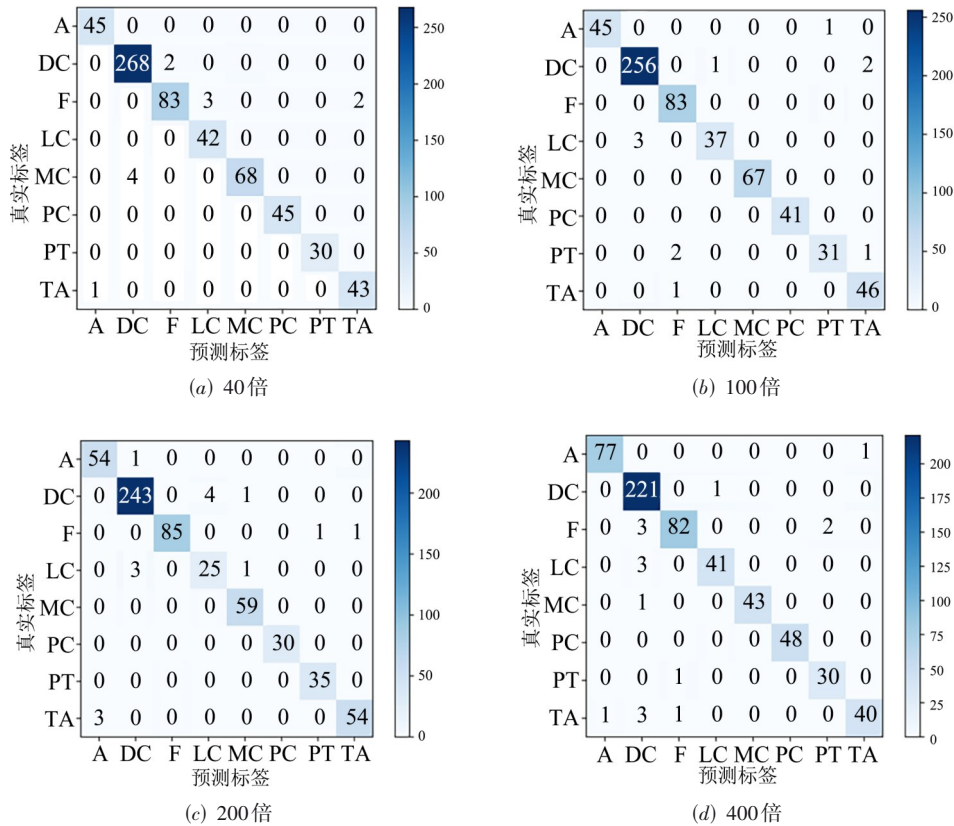


图 5 MVT-OFML 模型在 BreakHis 数据集上 8 分类任务不同放大倍数的混淆矩阵

4.3.2 在 BACH 数据集上的实验结果

与 BreakHis 数据集相比, BACH 数据集中的图像分辨率更高,且内容复杂,包含更多、更强的黏附性和来自背景的噪声,针对该数据集的分类也更具挑战. 故本节在 BACH 数据集上进行分类性能比较,详细的定量结果如表 5 所示. 将 MVT-OFML 模型与下列主流基线进行比较: Patch+Vote^[29]、HybridDNN (Hybrid Deep Neural Networks)^[30]、3E-Net^[31]、TransMIL^[22]、MA-MIDN^[23] (Multi-view Attention guided Multiple Instance Detection Network) 和 MSMV-PFENet (Multi-Scale Multi-View Progressive Feature Encoding Network)^[32].

实验结果表明: MVT-OFML 模型在各类指标上均优于基线,与最优基线 MSMV-PFENet 相比,其准确率、精准率、召回率和 F_1 分数分别提升 2.26%、3.10%、3.03% 和 3.21%, MVT-OFML 模型对于分类难度更大的病理图像数据集同样有效. 尽管大尺寸的 BACH 乳腺癌病理图像中包含大量噪声,但是蕴含交叉视角注意力机制的多视角 Transformer 编码模块能准确捕捉图像中的全局上下文信息,并与 CNN 模型捕获的细胞形态、

纹理、颜色等局部信息形成良性互补,进而抑制噪声干扰并降低黏附性的负面影响,实现更稳健的和优越的分类性能,这表明 MVT-OFML 模型是高效的、鲁棒的. 此外,在线融合互学习机制也扮演了重要角色. ResNet-50 和多视角 Transformer 编码器能从异构的中间特征层学习更多有价值的病理知识,而集成分类器将 2 个主干网络产生的深层病理知识回传到融合分支,故 2 个主干网络的 Logits 输出和特征蒸馏知识都被用来提升融合分类器判别能力. 2 个主干网络和融合分类器互为补充,为乳腺癌病理图像分类挖掘足够知识,改善最终分类性能.

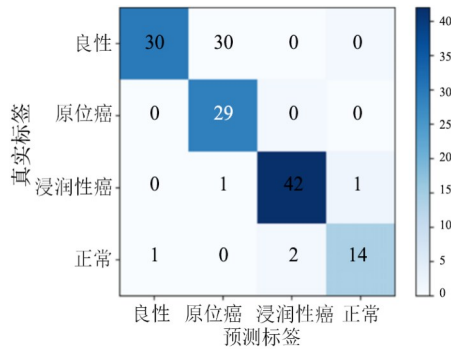
图 6(a) 是 BACH 上 MVT-OFML 模型的分混淆矩阵. 在图 6(a) 中, BACH 上 4 个类型的图像均取得较好分类性能. 正常类图像由于样本数少,其分类挑战较大,有部分样本被错分为浸润性癌,未来可通过对比学习方法进一步增强特征的判别性,减少错分现象出现. 图 6(b) 是 MVT-OFML 模型的 ROC 曲线. 在图 6(b) 中, MVT-OFML 模型取得的 AUC 为 0.997 6, 这表明 MVT-OFML 模型综合分类性能优秀.

综合各类指标可知: MVT-OFML 模型能学习到

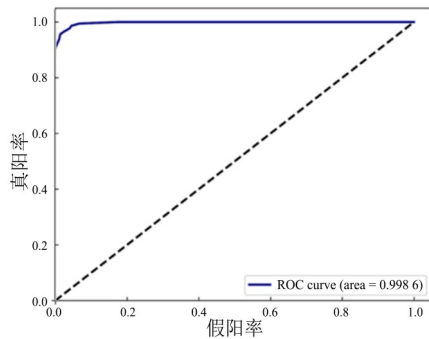
尺寸乳腺癌病理图像中的关键判别信息,为提升分类性能奠定重要的基础.

表 5 MVT-OFML模型在BACH数据集上与其他主流方法的性能对比
单位:%

主干网络或方法	发表年份	准确率	精确率	召回率	F_1 分数
Patch+Vote ^[29]	2019	85.00	86.77	81.91	84.23
Hybrid DNN ^[30]	2020	95.29	94.46	94.43	94.31
3E-Net ^[31]	2021	96.68	95.46	95.45	95.46
TransMIL ^[22]	2021	85.83	86.90	84.69	85.78
MA-MIDN ^[23]	2021	93.57	96.18	94.26	95.18
MSMV-PFENet ^[32]	2022	94.80	95.20	94.89	94.79
MVT-OFML	2023	98.94	98.56	98.48	98.67
提升情况		2.26 ↑	3.10 ↑	3.03 ↑	3.21 ↑



(a) 4分类任务混淆矩阵



(b) 4分类任务的ROC曲线

图6 MVT-OFML模型在BACH数据集上的结果

4.3.3 消融实验

为评估MVT-OFML模型各部分实际贡献,进行消融分析实验.消融分析实验在BreakHis数据集的8分类任务中完成.设计若干变种模型,包括:ResNet-50(模型A)、Transformer(模型B)、ResNet-50+Transformer(模型C)、ResNet-50+Transformer+多视角编码(模型D)、ResNet-50+Transformer+多视角编码+EC(模型E)、ResNet-50+Transformer+多视角编码+AFC(模型F)、

ResNet-50+Transformer+多视角编码+EC+AFC+Mixup数据增强(模型G)、MVT-OFML模型,其中,“EC”表示集成分类器,“AFC”表示融合分类器.实验结果如表6所示.

如表6所示,与ResNet-50+Transformer(模型C)相比,添加多视角Transformer编码模块后,模型准确率大幅提高4.81%,这表明多视角Transformer编码对乳腺癌病理图像分类是有效的.在模型D的基础上分别增加集成分类器EC(模型E)和增加融合分类器AFC(模型F),分类性能分别提升0.77%和3.08%.这表明,融合分类器作用更明显.这是因为集成分类器基于集成学习思想,它是对最终结果进行融合,并未处理异构特征间的冗余数据,所以模型性能提升相对微弱.相反,引入融合分类器AFC后,一方面,它会自适应融合来自异构网络的特征;另一方面,通过融合病理知识转移FPKT,充分利用KL散度将融合分类器的软化概率分布回传到每个子网络中.异构网络与融合分类器的冗余数据越少,KL散度就越大,融合分类器可以给异构网络更强的监督信号;相反,异构网络上融合分类器的冗余数据越多,KL散度就会越小,融合分类器给异构网络的监督信号更弱,此时,异构网络的Logits输出和真实标签的交叉熵损失会给异构网络更强的监督信号.通过联合融合病理知识转移FPKT和异构网络的Logits输出动态调整网络参数,促进网络优化及降低冗余数据对模型训练的影响.

融合分类器中蕴含来自异构网络中的深层病理信息.相比AFC,多视角Transformer编码模块更重要,因为它能准确捕捉乳腺癌病理图像的全局上下文信息,并从异构中间特征层学习更多有价值的病理知识.最后,Mixup数据增强对模型也起到了增益效果,最大的性能提升来自MVT-OFML模型,它联合多视角Transformer编码、集成分类器和融合分类器来提升分类性能.因此,在MVT-OFML模型中多视角Transformer编码模块作用最大,其次是融合分类器、集成分类器,最后是Mixup数据增强操作.针对输入AFC的特定中间特征进行消融分析,确定最佳中间特征组合.消融实验在BreakHis数据集的8分类任务中完成,具体实现:在模型F(ResNet-50+Transformer+多视角编码+AFC)的基础上分别选取ResNet-50卷积层1~3和Transformer的1~3层构建出输入AFC的中间特征组合,包括:AFC1(ResNet-50卷积层1+Transformer层1)、AFC2(ResNet-50卷积层2+Transformer层2)、AFC3(ResNet-50卷积层3+Transformer层3).实验结果如表7所示.在表7中,当联合ResNet-50的卷积层3和Transformer层3作为中间特征输入融合分类器,MVT-OFML模型可以获得最佳分类性能.故MVT-OFML联合ResNet-50的卷积层3和Transformer的层3特征执行在线融合互学习,实现乳腺癌病理图像分类.

表 6 在 BreakHis 数据集上的消融分析实验结果

单位:%

模型	Mixup	ResNet-50	Transformer	多视角编码	EC	AFC	准确率	F_1 分数
A	×	√	×	×	×	×	86.58	87.26
B	×	×	√	×	×	×	84.33	84.95
C	×	√	√	×	×	×	88.75	88.69
D	×	√	√	√	×	×	93.56	94.07
E	×	√	√	√	√	×	94.33	94.34
F	×	√	√	√	×	√	96.64	96.38
G	×	√	√	√	√	√	97.12	97.00
MVT-OFML	√	√	√	√	√	√	98.65	98.83

表 7 特定中间层特征的消融分析实验结果

单位:%

模型	Mixup	ResNet-50	Transformer	多视角编码	EC	AFC	准确率	F_1 分数
F	×	√	√	√	×	AFC1	95.78	95.01
F	×	√	√	√	×	AFC2	96.31	96.17
F	×	√	√	√	×	AFC3	96.64	96.38

4.3.4 特征可视化结果

本节使用 t-SNE^[33]对模型所产生的深层特征进行可视化展示,选取 BreakHis 上 8 分类任务完成可视化实验,结果如图 7 所示.

图 7(a)显示 ResNet-50 在最后 1 层池化层直接拼接融合特征的可视化结果. 图 7(b)是本文方法 MVT-OFML 的在线融合可视化结果. 与直接融合相比,本文模型将同类别乳腺癌病理图像聚合在一起,将不同类别样本推得很远,有助于构建明确的决策边界,提升模型分类准确率. 因此,基于可视化结果可知, MVT-OFML 模型产生的融合特征具有更强判别力,有助于提升乳腺癌病理图像分类性能.

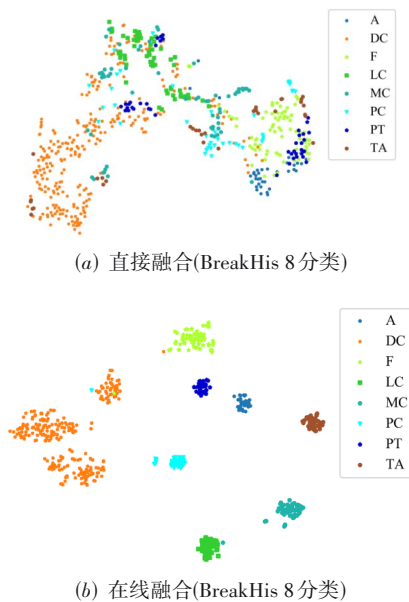


图 7 t-SNE 可视化结果

本节使用 Grad-CAM^[34]来可视化网络关注的区域. 图 8 比较了分别由 ResNet-50、Transformer 和 MVT-OFML 在 BreakHis 和 BACH 数据集上的 CAM 可视化结果,如图 8(a)~(e)所示.

如图 8 所示,在 ResNet-50 的可视化图中,模型只关注到零碎的局部特征信息,比如图 8(b)和图 8(c), ResNet-50 对病理图像的定位能力有限,故无法提取全局信息. 在 Transformer 的可视化图中,模型虽然关注到全局上下文信息,但对局部细节无法准确定位,某些无须关注的区域也被关注了,比如图 8(c)和图 8(d). MVT-OFML 模型利用融合特征中的丰富信息,实现对乳腺癌病理图像关键病灶区域定位. 此外,图 8 的结果具备一定可解释性,能更好辅助医生的临床诊断活动.

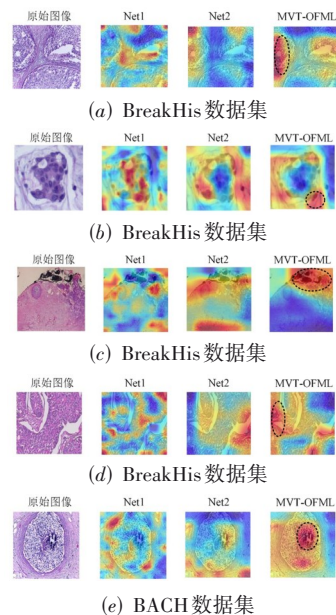


图 8 Grad-CAM 可视化结果

5 结论

乳腺癌病理图像分类不仅依赖局部的细胞核特征,也需要乳腺组织结构的全局上下文信息. 提出 MVT-OFML 模型,它将 CNN 与 Transformer 相结合,其中 CNN 聚焦于提取局部特征,Transformer 则捕获组织结构蕴含的全局上下文特征,全局特征和局部特征形成良性互补,更好地刻画图像中病理语义. 此外,设计在线融合互学习方法,联合集成分类器和融合分类器进一步改善乳腺病理图像分类性能. 实验表明,MVT-OFML 模型在 2 个公共数据集上均优于主流方法. 未来,计划探索多视角双向融合机制,以获取更为丰富、判别性更强的病理特征.

参考文献

- [1] SUNG H, FERLAY J, SIEGEL R L, et al. Global cancer statistics 2020: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries[J]. *A Cancer Journal for Clinicians*, 2021, 71(3): 209-249.
- [2] 宋杰, 肖亮, 练智超, 等. 基于深度学习的数字病理图像分割综述与展望[J]. *软件学报*, 2021, 32(5): 1427-1460. SONG J, XIAO L, LIAN Z C, et al. Overview and prospect of deep learning for image segmentation in digital pathology[J]. *Journal of Software*, 2021, 32(5): 1427-1460. (in Chinese)
- [3] KRITHIGA R, GEETHA P. Breast cancer detection, segmentation and classification on histopathology images analysis: A systematic review[J]. *Archives of Computational Methods in Engineering*, 2021, 28(4): 2607-2619.
- [4] SRINIDHI C L, CIGA O, MARTEL A L. Deep neural network models for computational histopathology: A survey[J]. *Medical Image Analysis*, 2021, 67: 101813.
- [5] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16×16 words: Transformers for image recognition at scale[EB/OL]. (2021-06-03) [2023-03-02]. <http://arxiv.org/pdf/2010.11929>.
- [6] SANSONE M, FUSCO R, GRASSI F, et al. Machine learning approaches with textural features to calculate breast density on mammography[J]. *Current Oncology*, 2023, 30(1): 839-853.
- [7] WANG P, HU X L, LI Y M, et al. Automatic cell nuclei segmentation and classification of breast cancer histopathology images[J]. *Signal Processing*, 2016, 122: 1-13.
- [8] XUE H, YANG Q, CHEN S. SVM: Support vector machines[M]//*The Top Ten Algorithms in Data Mining*. New York: Chapman and Hall/CRC, 2009: 51-74.
- [9] 范虹, 张程程, 侯存存, 等. 结合双树复小波变换和改进密度峰值快速搜索聚类的乳腺 MR 图像分割[J]. *电子学报*, 2019, 47(10): 2149-2157. FAN H, ZHANG C C, HOU C C, et al. Dual-tree complex wavelet transform and improved density peak fast search and clustering method for breast MR image segmentation[J]. *Acta Electronica Sinica*, 2019, 47(10): 2149-2157. (in Chinese)
- [10] 蒲秀娟, 刘浩伟, 韩亮, 等. 使用深度学习与海马体异构特征融合的阿尔茨海默病分类方法[J]. *电子学报*, 2023, 51(11): 3305-3319. PU X J, LIU H W, HAN L, et al. Method on Alzheimer's disease classification utilizing deep learning and hippocampus heterogeneous feature fusion[J]. *Acta Electronica Sinica*, 2023, 51(11): 3305-3319. (in Chinese)
- [11] ZOU Y, ZHANG J X, HUANG S, et al. Breast cancer histopathological image classification using attention high-order deep network[J]. *International Journal of Imaging Systems and Technology*, 2022, 32(1): 266-279.
- [12] CRUZ-ROA A, BASAVANHALLY A, GONZÁLEZ F, et al. Automatic detection of invasive ductal carcinoma in whole slide images with convolutional neural networks[C]//*SPIE Proceedings of Medical Imaging 2014: Digital Pathology*. San Diego: SPIE, 2014: 904103.
- [13] ZHOU Y P, ZHANG C, GAO S S. Breast cancer classification from histopathological images using resolution adaptive network[J]. *IEEE Access*, 2019, 10: 35977-35991.
- [14] WANG Z N, MA Y T, ZHENG Y L, et al. Weakly-supervised learning using pretraining for classification in HER2 immunohistochemistry image of breast cancer[C]//*Proceedings of the 2021 6th International Conference on Mathematics and Artificial Intelligence*. New York: ACM, 2021: 66-71.
- [15] 田永林, 王雨桐, 王建功, 等. 视觉 Transformer 研究的关键问题: 现状及展望[J]. *自动化学报*, 2022, 48(4): 957-979. TIAN Y L, WANG Y T, WANG J G, et al. Key problems and progress of vision transformers: The state of the art and prospects[J]. *Acta Automatica Sinica*, 2022, 48(4): 957-979. (in Chinese)
- [16] PARMAR N, VASWANI A, USZKOREIT J, et al. Image transformer[EB/OL]. (2018-06-15) [2023-03-02]. <http://arxiv.org/pdf/1802.05751>.
- [17] ALOTAIBI A, ALAFIF T, ALKHILAIWI F, et al. ViT-DeiT: An ensemble model for breast cancer histopathological images classification[EB/OL]. (2022-11-01) [2023-03-02]. <http://arxiv.org/abs/2211.00749>.
- [18] ZOU Y, CHEN S N, SUN Q L, et al. DCET-net: Dual-stream convolution expanded transformer for breast cancer histopathological image classification[C]//*2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. Piscataway: IEEE, 2021: 1235-1240.
- [19] TUMMALA S, KADRY S, BUKHARI S A C, et al. Classification of brain tumor from magnetic resonance imaging using vision transformers ensembling[J]. *Current Oncology*, 2022, 29(10): 7498-7511.
- [20] SUDHARSHAN P J, PETITJEAN C, SPANHOL F, et al. Multiple instance learning for histopathological breast cancer image classification[J]. *Expert Systems with Applications*, 2019, 117: 103-111.

- [21] ILSE M, TOMCZAK J M, WELLING M. Attention-based deep multiple instance learning[EB/OL]. (2018-06-28)[2023-03-02]. <http://arxiv.org/abs/1802.04712>.
- [22] SHAO Z C, BIAN H, CHEN Y, et al. TransMIL: Transformer based correlated multiple instance learning for whole slide image classification[EB/OL]. (2021-06-02)[2023-03-02]. <http://arxiv.org/abs/2106.00908>.
- [23] LI G L, LI C X, WU G T, et al. Multi-view attention-guided multiple instance detection network for interpretable breast cancer histopathological image diagnosis[J]. IEEE Access, 2021, 9: 79671-79684.
- [24] ZHANG H Y, CISSE M, DAUPHIN Y N, et al. Mixup: Beyond empirical risk minimization[EB/OL]. (2018-04-27)[2023-03-02]. <http://arxiv.org/abs/1710.09412>.
- [25] LI X, SHEN X, ZHOU Y X, et al. Classification of breast cancer histopathological images using interleaved DenseNet with SENet (IDSNet)[J]. PLoS One, 2020, 15(5): e0232127.
- [26] MAN R, YANG P, XU B W. Classification of breast cancer histopathological images using discriminative patches screened by generative adversarial networks[J]. IEEE Access, 2020, 8: 155362-155377.
- [27] JIANG Y, CHEN L, ZHANG H, et al. Breast cancer histopathological image classification using convolutional neural networks with small SE-ResNet module[J]. PLoS One, 2019, 14(3): e0214587.
- [28] TUMMALA S, KIM J, KADRY S. BreaST-net: Multi-class classification of breast cancer from histopathological images using ensemble of swin transformers[J]. Mathematics, 2022, 10(21): 4109.
- [29] ZEROUAOUI H, IDRI A. Deep hybrid architectures for binary classification of medical breast cancer images[J]. Biomedical Signal Processing and Control, 2022, 71: 103226.
- [30] YE H C, ZHANG X F, HUANG Z Z, et al. HybridDNN: A framework for high-performance hybrid DNN accelerator design and implementation[C]//2020 57th ACM/IEEE Design Automation Conference (DAC). Piscataway: IEEE, 2020: 1-6.
- [31] SENOUSY Z, ABDELSAMEA M M, MOHAMED M M, et al. 3E-net: Entropy-based elastic ensemble of deep convolutional neural networks for grading of invasive breast carcinoma histopathological microscopic images[J]. Entropy, 2021, 23(5): 620.
- [32] LIU L X, FENG W X, CHEN C, et al. Classification of breast cancer histology images using MSMV-PFENet[J]. Scientific Reports, 2022, 12(1): 17447.
- [33] VAN DER MAATEN L, HINTON G. Visualizing data using t-SNE[J]. Journal of Machine Learning Research, 2008, 9: 2579-2625.
- [34] SELVARAJU R R, DAS A, VEDANTAM R, et al. Grad-CAM: Why did you say that? [EB/OL]. (2017-01-25)[2023-03-02]. <http://arxiv.org/abs/1611.07450>.

作者简介



李广丽 女,1977年11月出生,广西壮族自治区玉林人.华东交通大学信息工程学院教授,硕士生导师.主要研究方向为医学图像处理、数据挖掘、跨模态分析、推荐系统等.
E-mail: 1333@ecjtu.edu.cn



叶艺源 女,1998年12月出生,江西上饶人.华东交通大学信息工程学院硕士研究生.主要研究方向为医学图像处理、肿瘤图像识别、深度学习等.
E-mail: 1043549290@qq.com



吴光庭 男,1996年3月出生,江西抚州人.华东交通大学信息工程学院硕士.主要研究方向为医学图像处理、肿瘤图像识别、深度学习等.
E-mail: 1394692438@qq.com



李传秀 男,1995年9月出生,山东青岛人.华东交通大学信息工程学院硕士.主要研究方向为医学图像处理、肿瘤图像识别、深度学习等.
E-mail: 523536220@qq.com



吕敬钦 男,1984年11月出生,江西九江人.华东交通大学软件学院副教授.主要研究方向为行人检测、机器视觉等.
E-mail: jingqinlv@ecjtu.edu.cn



张红斌 男,1979年10月出生,江苏如皋人.华东交通大学软件学院教授,硕士生导师.主要研究方向为医学影像分析、自然语言处理、跨模态分析等.
E-mail: zhanghongbin@whu.edu.cn