

# 光电工程

## Opto-Electronic Engineering

中文核心期刊 中国科技核心期刊  
Scopus CSCD

### 基于自适应融合和显微成像的乳腺肿瘤分级网络

黄盼, 何鹏, 杨兴, 罗家洋, 肖华亮, 田素坤, 冯鹏

#### 引用本文:

黄盼, 何鹏, 杨兴, 等. 基于自适应融合和显微成像的乳腺肿瘤分级网络[J]. 光电工程, 2023, 50(1): 220158.

Huang P, He P, Yang X, et al. Breast tumor grading network based on adaptive fusion and microscopic imaging[J]. *Opto-Electron Eng*, 2023, 50(1): 220158.

<https://doi.org/10.12086/oe.2023.220158>

收稿日期: 2022-07-08; 修改日期: 2022-09-06; 录用日期: 2022-09-06

## 相关论文

### 采用注意力机制的显微图像智能检测方法

郝如茜, 王祥舟, 张静, 刘娟秀, 杜晓辉, 刘霖

光电工程 2022, 49(3): 210361 doi: 10.12086/oe.2022.210361

### 结合极坐标建模与神经网络的IVUS图像分割

刘靖雨, 蔡怀宇, 郝文月, 左廷涛, 贾忠伟, 汪毅, 陈晓冬

光电工程 2023, 50(1): 220118 doi: 10.12086/oe.2023.220118

### 太赫兹成像技术在肿瘤检测中的应用

施辰君, 吴旭, 彭滢

光电工程 2020, 47(5): 190638 doi: 10.12086/oe.2020.190638

### 基于超声RF信号的乳腺肿瘤分级检测方法

童莹, 严郁

光电工程 2019, 46(1): 180368 doi: 10.12086/oe.2019.180368

更多相关论文见光电期刊集群网站 



<http://cn.ojournal.org/oe>



 OE\_Journal



Website

DOI: 10.12086/oe.2023.220158

# 基于自适应融合和显微成像的 乳腺肿瘤分级网络

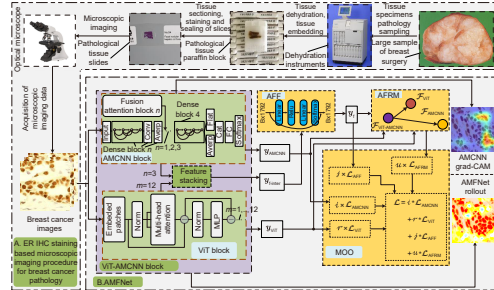
黄盼<sup>1</sup>, 何鹏<sup>1</sup>, 杨兴<sup>2</sup>, 罗家洋<sup>1</sup>,  
肖华亮<sup>3\*</sup>, 田素坤<sup>4\*</sup>, 冯鹏<sup>1\*</sup>

<sup>1</sup>重庆大学光电技术及系统教育部重点实验室,  
光电工程学院, 重庆 400044;

<sup>2</sup>成都理工大学计算机与网络安全学院, 四川 成都 610000;

<sup>3</sup>陆军军医大学大坪医院病理科, 重庆 400037;

<sup>4</sup>山东大学机械工程学院, 山东 济南 250000



**摘要:** 基于显微成像技术的肿瘤分级对于乳腺癌诊断和预后有着重要的意义,且诊断结果需具备高精度和可解释性。目前,集成 Attention 的 CNN 模块深度网络归纳偏差能力较强,但可解释性较差;而基于 ViT 块的深度网络其可解释性较好,但归纳偏差能力较弱。本文通过融合 ViT 块和集成 Attention 的 CNN 块,提出了一种端到端的自适应模型融合的深度网络。由于现有模型融合方法存在负融合现象,无法保证 ViT 块和集成 Attention 的 CNN 块同时具有良好的特征表示能力;另外,两种特征表示之间相似度高且冗余信息多,导致模型融合能力较差。为此,本文提出一种包含多目标优化、自适应特征表示度量和自适应特征融合的自适应模型融合方法,有效地提高了模型的融合能力。实验表明本文模型的准确率达到 95.14%,相比 ViT-B/16 提升了 9.73%,比 FABNet 提升了 7.6%;模型的可视化图更加关注细胞核异型的区域(例如巨型核、多形核、多核和深色核),与病理专家所关注的区域更加吻合。整体而言,本文所提出的模型在精度和可解释性上均优于当前最先进的(state of the art)模型。

**关键词:** 显微镜成像; 可解释性; 深度学习; 自适应融合; 乳腺癌; 肿瘤分级

**中图分类号:** TP391.4

**文献标志码:** A

黄盼, 何鹏, 杨兴, 等. 基于自适应融合和显微成像的乳腺肿瘤分级网络 [J]. 光电工程, 2023, 50(1): 220158

Huang P, He P, Yang X, et al. Breast tumor grading network based on adaptive fusion and microscopic imaging[J]. *Opto-Electron Eng*, 2023, 50(1): 220158

## Breast tumor grading network based on adaptive fusion and microscopic imaging

Huang Pan<sup>1</sup>, He Peng<sup>1</sup>, Yang Xing<sup>2</sup>, Luo Jiayang<sup>1</sup>, Xiao Hualiang<sup>3\*</sup>, Tian Sukun<sup>4\*</sup>, Feng Peng<sup>1\*</sup>

<sup>1</sup>Key Laboratory of Optoelectronic Technology & Systems (Ministry of Education), College of Optoelectronic Engineering, Chongqing University, Chongqing 400044, China;

<sup>2</sup>College of Computer and Network Security, Chengdu University of Technology, Chengdu, Sichuan 610000, China;

<sup>3</sup>Daping Hospital, Department of Pathology, Army Military Medical University, Chongqing 400037, China;

<sup>4</sup>School of Mechanical Engineering, Shandong University, Jinan, Shandong 250000, China

收稿日期: 2022-07-08; 收到修改稿日期: 2022-09-06

基金项目: 国家重点研发计划资助项目(2019YFC0605203); 国家自然科学基金资助项目(52105265); 重庆市基础研究与前沿探索专项资助项目(cstc2020jcyj-msxmX0553)

\*通信作者: 肖华亮, dpbl\_xhl@163.com; 田素坤, sukhum169@hotmail.com; 冯鹏, coe-fp@cqu.edu.cn.

版权所有©2023 中国科学院光电技术研究所

**Abstract:** Tumor grading based on microscopic imaging is critical for the diagnosis and prognosis of breast cancer, which demands excellent accuracy and interpretability. Deep networks with CNN blocks combined with attention currently offer better induction bias capabilities but low interpretability. In comparison, the deep network based on ViT blocks has stronger interpretability but less induction bias capabilities. To that end, we present an end-to-end adaptive model fusion for deep networks that combine ViT and CNN blocks with integrated attention. However, the existing model fusion methods suffer from negative fusion. Because there is no guarantee that both the ViT blocks and the CNN blocks with integrated attention have acceptable feature representation capabilities, and secondly, the great similarity between the two feature representations results in a lot of redundant information, resulting in a poor model fusion capability. For that purpose, the adaptive model fusion approach suggested in this study consists of multi-objective optimization, an adaptive feature representation metric, and adaptive feature fusion, thereby significantly boosting the model's fusion capabilities. The accuracy of this model is 95.14%, which is 9.73% better than that of ViT-B/16, and 7.6% better than that of FABNet; secondly, the visualization map of our model is more focused on the regions of nuclear heterogeneity (e.g., mega nuclei, polymorphic nuclei, multi-nuclei, and dark nuclei), which is more consistent with the regions of interest to pathologists. Overall, the proposed model outperforms other state-of-the-art models in terms of accuracy and interpretability.

**Keywords:** microscopic imaging; interpretability; deep learning; adaptive fusion; breast cancer; tumor grading

## 1 引言

乳腺癌是女性中最常见的癌症<sup>[1]</sup>。乳腺癌肿瘤的组织学分级已被证明是一个强有力的诊断和预后指标,其主要依靠病理专家评估组织切片的显微镜成像中组织和细胞的形态学状况,例如腺体分化、细胞核异型和有丝分裂计数等<sup>[2]</sup>。已有研究文献表明苏木精-伊红 (Hematoxylin-Eosin, HE) 染色的显微成像样本和孕激素受体 (Estrogen Receptor, ER) 的免疫组化 (Immunohistochemically, IHC) 染色的显微成像样本存在较强统计相关性,即 ER 状态与肿瘤组织分级有着较强的相关性<sup>[3-9]</sup>。其中,Desai 等人<sup>[3]</sup>的研究表明 ER 状态值的下降与肿瘤恶性程度有一定的相关性。Zafrani 等人<sup>[4]</sup>发现 ER 状态值和肿瘤分级有着很强的统计相关性 ( $P < 10^{-4}$ )。Fuqua 等人<sup>[5]</sup>发现 ER 状态值和肿瘤组织分级有着明显的相关性。Vagunda 等人<sup>[6]</sup>的研究也发现 ER 值与肿瘤分级有统计学相关性。Baqai 和 Shousha 等人<sup>[7]</sup>的工作表明 ER 阴性与肿瘤高级别有着明显相关性 ( $P < 0.001$ )。Sofi 等人<sup>[8]</sup>发现 ER 在 66.3% 的病例中呈阳性,随着年龄的增长而增加,高级别病变和较大尺寸的肿瘤更可能是 ER 阴性。Azizun-Nisa 等人<sup>[9]</sup>研究表明高级肿瘤中的 ER 表达显著降低 (ER 5.6% vs 10.5%)。因此,在基于 ER IHC 染色的病理组织显微镜成像中进行肿瘤分级有一定的统计意义,可以作为临床诊断和预后的一个参考指标。

虽然数字病理全玻片图像技术发展迅速,但基于显微镜成像的方式仍然是病理科诊断的主要手段,这是因为全玻片机昂贵的价格和生成数字全玻片图像的速度较慢,无法满足病理科大量的病例诊断需求。现有肿瘤分级仍十分依赖病理专家的经验 and 精力,因为病理图像的肿瘤级别之间差异小,分级指标复杂且带有很强的主观性<sup>[10]</sup>。因此,研究计算机肿瘤分级算法来辅助病理专家进行肿瘤分级的工作显得十分重要。

在癌症组织病理显微镜成像分析中,基于集成注意力机制的卷积神经网络 (attention mechanism based on convolutional neural networks, AMCNN) 块的深度学习模型辅助病理专家进行诊断已取得快速发展<sup>[11-17]</sup>。Yang 等人<sup>[11]</sup>采用一种引导注意力机制来提高 CNN 模型识别乳腺癌组织病理图像的能力。BenTaieb 等人<sup>[12]</sup>提出一种基于循环注意力机制的检测模型,实现癌症数字全玻片图像中对有无肿瘤病变进行预测。Tomita 等人<sup>[13]</sup>采用注意力机制改进 ResNet18 来提高食道癌病理图像分类性能。Yao 等人<sup>[14]</sup>提出孪生注意力机制和全连接层从数字全玻片中识别癌变。Huang<sup>[15]</sup>和 Zhou<sup>[16]</sup>等人构建了一种融合注意力机制的改进 CNN 模型,提高了喉癌病理图像肿瘤分级的性能。虽然,以上基于 AMCNN 块的深度学习模型的识别性能和归纳偏差能力较好,但是其可解释性较差,导致模型结果的临床可信度不高<sup>[18-21]</sup>。

基于 Vision Transformer (ViT) 块的深度学习模型

流行于计算机视觉领域, 在癌症组织病理图像处理方向也有一些应用<sup>[22-27]</sup>。Gao 等人<sup>[22]</sup>提出一种基于示例学习的 ViT 模型从病理图像中学习更加鲁棒的特征表示。Wang 等人<sup>[23]</sup>提出一种混合 ViT 模型来探索图像固有特性和在特定域捕获新的知识。Li 等人<sup>[24]</sup>采用一种可变型的 ViT 模块, 来提高模型的归纳偏差能力。Chen 等人<sup>[25]</sup>提出一种多尺度可视化 ViT 模型, 用于胃部组织病理图像的分类。Zou 等人<sup>[26]</sup>提出一种双 ViT 网络模型, 可通过同时捕获图像局部和全局信息识别癌症图像。尽管 ViT 块采用多 Attention 机制的串、并联结构, 使其具有较好的可解释性。但是其缺乏旋转不变性、共享权重和尺度不变性等优势, 导致基于 ViT 块的深度学习模型的归纳偏差能力较差<sup>[27]</sup>。

为使乳腺癌肿瘤分级结果同时具有良好的可解释性和高精度, 本文提出一种基于自适应模型融合的端到端深度网络 (deep network with adaptive model fusion, AMFNet)。由于现有的模型融合方法 bagging<sup>[28]</sup>, boosting<sup>[29]</sup> 和 stacking<sup>[30]</sup> 存在负融合现象<sup>[31]</sup>, 原因: 1) ViT 块参数量是 AMCNN 块的十倍, 学习率

过大, ViT 块容易错过局部最优解, 学习率过小, AMCNN 块会陷入“鞍部”。因此, ViT 块和 AMCNN 块无法同时具备良好的特征表示能力; 2) ViT 块和 AMCNN 块所学的特征表示相似度高, 导致融合特征矩阵的冗余度高, 并且现有方法去冗余能力较弱, 故 AMFNet 融合 ViT 块和 AMCNN 块的能力不足<sup>[32]</sup>。因此, 为了解决 AMFNet 在融合 ViT 块和 AMCNN 块时出现负融合现象, 本文提出一种自适应模型融合 (adaptive model fusion, AMF) 方法, 包括多目标优化 (multi-objective optimization, MOO)、自适应特征表示度量 (adaptive feature representation metric, AFRM) 和自适应特征融合 (adaptive feature fusion, AFF) 三部分。其中, 多目标优化是保证 ViT 块和 AMCNN 块能够同时具备良好的特征表示能力; 而自适应特征表示度量和自适应特征融合是提高 AMFNet 的去冗余能力, 进而提升模型的融合能力。

本文的主要贡献如下:

1) 提出一种多目标优化方法, 能同时保证 ViT 块和 AMCNN 块具有良好的特征表示能力, 从而提

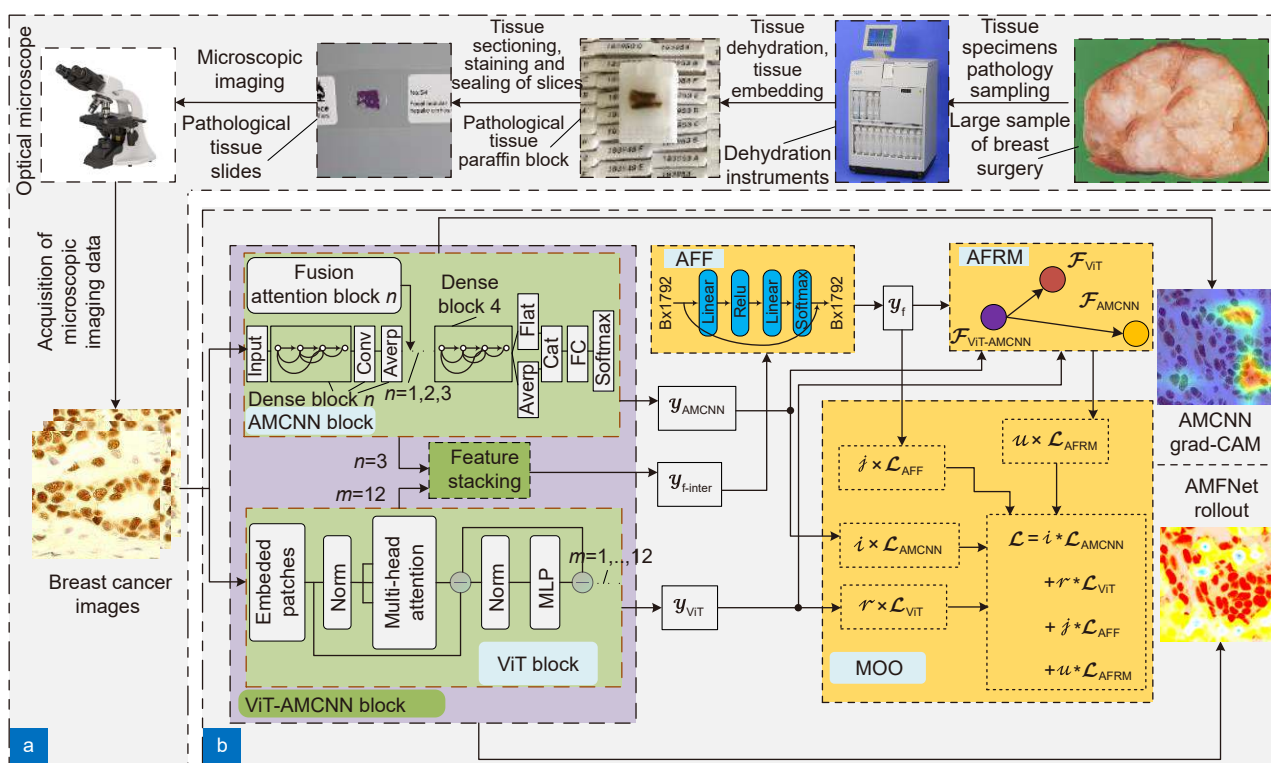


图 1 本文算法框图。

(a) 介绍基于 ER IHC 染色的乳腺癌病理组织显微镜成像的基本流程; (b) 描述 AMFNet 整体流程。

Fig. 1 Block diagram of the algorithm in this paper.

(a) ER IHC staining based microscopic imaging procedure for breast cancer pathology; (b) AMFNet

高 AMFNet 的模型融合能力;

2) 提出一种自适应特征表示度量方法, 能降低 ViT 块和 AMCNN 块所学特征表示之间的相似性和冗余度, 进而提高 AMFNet 的模型融合能力;

3) 提出一种自适应特征融合方法, 能够自适应和更加稀疏地去除融合特征矩阵中的冗余信息, 从而提高 AMFNet 的模型融合能力。

## 2 方法

为实现对 ER IHC 染色的乳腺癌病理组织显微镜成像进行高精度和可解释性的肿瘤分级, 本文通过有效地融合 ViT 块和 AMCNN 块, 提出了一种基于自适应模型融合的深度网络, 如图 1 所示。该方法主要包括五部分: 基于 ER IHC 染色的乳腺癌病理组织显微镜成像部分、ViT 块、AMCNN 块、ViT-AMCNN 块和 AMF 方法(包括 MOO、AFRM 和 AFF)。其中, ViT 块采用 ViT-B/16<sup>[27]</sup>, AMCNN 块使用 FABNet 模型<sup>[15]</sup>。依据图 1, 对本文算法流程进行简要的介绍。首先从图 1(a) 中采集显微镜成像数据分别输入到图 1(b) AMFNet 的 AMCNN 块、ViT 块、ViT-AMCNN 块中得到对应的特征矩阵 ( $\mathcal{Y}_{AMCNN}$ ,  $\mathcal{Y}_{ViT}$  和  $\mathcal{Y}_{F-inter}$ )。第二, 将  $\mathcal{Y}_{F-inter}$  输入到 AFF 方法中去, 计算每个特征通道的权重分数, 然后乘以权重分数进行自适应地去除特征矩阵的冗余信息, 得到  $\mathcal{Y}_r$ 。第三, 将每个块所学到的特征表示 ( $\mathcal{F}_{AMCNN} = \{\mathcal{Y}_{AMCNN}, \ell\}$ ,  $\mathcal{F}_{ViT} = \{\mathcal{Y}_{ViT}, \ell\}$  和  $\mathcal{F}_{ViT-AMCNN} = \{\mathcal{Y}_r, \ell\}$ ,  $\ell$  表示标签矩阵) 一起输入到 AFRM 方法中去, 采用深度度量学习来度量特征表示之间的联合分布距离, 拉开  $\mathcal{F}_{ViT}$  和  $\mathcal{F}_{AMCNN}$  的距离, 从而降低  $\mathcal{F}_{ViT-AMCNN}$  的冗余性。第四, 采用 MOO 方法对 ViT、AMCNN、AFRM 和 AFF 的输出目标同时进行多目标和多学习率优化。最后采用 Grad-CAM<sup>[33]</sup> 和 Rollout<sup>[34]</sup> 技术可视化模型所学知识。

### 2.1 基于 ER IHC 染色的乳腺癌病理组织显微镜成像流程

基于 ER IHC 染色的乳腺癌病理组织显微镜成像流程主要包括六个步骤: 组织标本病理取材、组织脱水、组织包埋、组织切片、免疫组化染色封片和显微镜成像, 如图 1(a) 所示。

1) 组织标本病理取材: 对送检的手术大样本按标准进行逐一剖开, 并充分固定 (>12 h);

2) 组织脱水: 采用脱水仪器 Leica ASP300S 对组织病理标本进行脱水处理, 整个脱水过程有 13 个步

骤, 每个步骤都需要将脱水仪器设置真空、加压或两者交替进行的脱水模式。而脱水过程主要用到的试剂有: AF 液、75% 酒精、85% 酒精、95% 酒精、无水酒精、二甲苯和石蜡;

3) 组织包埋: 采用 Leica EG1150 H 石蜡包埋机进行组织包埋, 其蜡缸的容量为 3L, 在工作日和工作时间设置仪器自动运行。包埋盒和模具的加热槽可互换, 以适应变化的包埋流程。Leica EG1150 H 的冷台维持 -5 °C 恒温, 实现包埋蜡块和模具的快速冷却。该过程主要耗材是包埋盒和石蜡等;

4) 组织切片: 采用 Leica RM2235 石蜡切片机进行组织切片, 其选配的 8° X/Y 轴精确定位系统带校准调节功能, 有助于快速定位切过的样品(重切)。定位系统随后快速返回到有 2 个红色指示标记的基准 0 刻度位。垂直行程最长达 70 mm, 不仅能使用超大包埋盒, 并且由于加大了距离刀刃的工作距离;

5) 免疫组化染色封片: 采用 Autostainer XL Leica ST5010 染色封片系统、医嘱所需的免疫组化试剂、酒精、二甲苯、玻片、封片胶带等耗材对组织切片进行免疫组化染色并封装。该过程分为 23 个步骤, 需要精准地设置每个步骤的时间和温度;

6) 显微镜成像: 采用 Olympus BX41 多头显微镜对病理组织玻片进行成像, 其中物镜分别设置为 20X (N.A. 0.4, W.D. 1.2 Spring) 和 40X (N.A. 0.65, W.D. 0.6 Spring)。齐焦距为 45 mm, 观察镜筒的倾角为 30°, 使用柯勒照明器作为内置透射光源头。

### 2.2 多目标优化

为有效地融合 ViT 块和 AMCNN 块, 来提高 AMFNet 的模型融合能力, 使模型的肿瘤分级具有高精度和可解释性的特性。本文提出一个多目标优化的损失函数  $\mathcal{L}$  来实现上述目的, 同时也优化 AMF 方法中的各个部分。

$$\mathcal{L} = i * \mathcal{L}_{AMCNN} + r * \mathcal{L}_{ViT} + j * \mathcal{L}_{AFF} + u * \mathcal{L}_{AFRM}, \quad (1)$$

式中:  $\mathcal{L}_{AMCNN}$  和  $\mathcal{L}_{ViT}$  分别优化 AMCNN 块和 ViT 块, 保证它们同时具有良好特征表示的能力, 且保证 AMFNet 模型的肿瘤分级结果具有高精度和可解释性。 $\mathcal{L}_{AFF}$  用于优化 AFF 方法部分, 其输出为模型最终的肿瘤分级结果。 $\mathcal{L}_{AFRM}$  用于优化 AFRM 方法部分。其中  $\mathcal{L}_{AMCNN}$ ,  $\mathcal{L}_{ViT}$ ,  $\mathcal{L}_{AFF}$  使用了交叉熵损失函数 ( $\mathcal{L}_{CE}$ ), 如式 (2) 所示。 $i, r, j, u$  是对应于损失函数的权重系数。由于 AMCNN 块和 ViT 块参数量差异较大, 本文设置了多学习率优化策略:  $\mathcal{L}_{AMCNN}$  和  $\mathcal{L}_{AFF}$  的学习率设置

为 1~50 epochs:  $1 \times 10^{-4}$ , 51~75 epochs:  $2 \times 10^{-5}$ , 76~100 epochs:  $1 \times 10^{-6}$ ;  $\mathcal{L}_{ViT}$  为 1~50 epochs:  $1 \times 10^{-5}$ , 51~75 epochs:  $5 \times 10^{-6}$ , 76~100 epochs:  $1 \times 10^{-6}$ ;  $\mathcal{L}_{AFRM}$  没有设置专门对应的可学习参数, 而是在优化过程中  $\mathcal{L}_{AFRM}$  会同时影响 ViT 块, AMCNN 块和 AFF 方法所对应的可学习参数。

$$\mathcal{L}_{CE} = -\ell * \sum_{i=1}^M \ell^i * \log(\delta(\mathcal{Y}^i)), \quad (2)$$

式中:  $\ell$  表示损失函数系数,  $M$  表示模型输入张量的 batch size,  $\ell$  表示真实的标签矩阵,  $\mathcal{Y}$  表示模型输出的特征矩阵,  $\delta$  表示 Softmax 激活函数。

### 2.3 自适应特征融合

为去除 ViT 块和 AMCNN 块采用特征拼接 (feature stacking, 按特征矩阵的第二维度进行拼接) 方法融合后的特征矩阵中存在较多冗余信息, 而传统特征选择方法无法实现自适应和稀疏性的去冗余。因此, 本文提出一种 AFF 方法, 采用编解码的神经网络和 Softmax 概率权重的思路, 分别实现自适应和稀疏性求解各个特征的权重系数。从而有效地提高了 AMFNet 的模型融合能力, 让模型的肿瘤分级结果能够具有更高的准确率和更好的可解释性。实现细节如图 2 和式 (3) 所示。给出 AMFNet 的 AMCNN 块、ViT 块、ViT-AMCNN 块中得到对应的特征矩阵 ( $\mathcal{Y}_{AMCNN}, \mathcal{Y}_{ViT}, \mathcal{Y}_{f-inter}$ ), 自适应融合后的最终特征矩阵为  $\mathcal{Y}_f$ 。

$$\mathcal{Y}_f^{(B,1792)} = \delta \left[ \begin{array}{c} \theta \left( \mathcal{Y}_{f-inter}^{(B,1792)} \times \mathcal{W}_{l1}^{(1792,112)} + \mathcal{B}_{l1}^{(1,112)} \right) \times \mathcal{W}_{l2}^{(112,1792)} \\ + \mathcal{B}_{l2}^{(1,1792)} \end{array} \right], \quad (3)$$

式中:  $\delta$  表示 Softmax 激活函数,  $\theta$  表示 ReLU 激活函数,  $\mathcal{W}_{l1}, \mathcal{B}_{l1}$  表示 Linear 1 层 (编码层) 的权重和偏置项,  $\mathcal{W}_{l2}, \mathcal{B}_{l2}$  表示 Linear 2 层 (解码层) 的权重和偏

置项。

### 2.4 自适应特征表示度量

解决 ViT 和 AMCNN 块融合后冗余信息过多的另外一个方法是降低两个模块所学特征表示之间的相似度。为此, 本文提出一种 AFRM 的方法, 通过深度度量学习, 来拉开 ViT 和 AMCNN 块的特征表示之间的距离 (特征表示: 由特征矩阵和所对应的标签矩阵构成联合概率分布, 其中图 3 中的  $\mathcal{F}_{ViT}, \mathcal{F}_{AMCNN}$  和  $\mathcal{F}_{ViT-AMCNN}$  分别表示 ViT 块、AMCNN 块和 ViT-AMCNN 块所学的特征表示), 从而降低融合特征矩阵的冗余度, 提高 AMFNet 的模型融合能力, 进而提升模型肿瘤分级结果的准确率和可解释性。为了实现这个目标, 必须要设置一个目标函数 (如式 (4) 所示), 采用梯度反向传播的方式进行多次优化, 从而得到一个最优的 AMFNet 权重参数, 实现细节如式 (4) 和图 3 所示。

$$\mathcal{H}^* = \arg \min_{\mathcal{H}} \left\{ \frac{1}{N} \sum_i \mathcal{L}_{CE}(\mathcal{H}(\mathcal{Y}_f), \ell) + \varphi * \mathcal{D}(\mathcal{P}(\mathcal{Y}_f), \mathcal{P}(\mathcal{Y}_{AMCNN})) - \varphi * \mathcal{D}(\mathcal{P}(\mathcal{Y}_f), \mathcal{P}(\mathcal{Y}_{ViT})) + (1 - \varphi) * \mathcal{D}(\mathcal{P}(\ell|\mathcal{Y}_f), \mathcal{P}(\ell|\mathcal{Y}_{AMCNN})) - (1 - \varphi) * \mathcal{D}(\mathcal{P}(\ell|\mathcal{Y}_f), \mathcal{P}(\ell|\mathcal{Y}_{ViT})) \right\}, \quad (4)$$

式中:  $N$  表示训练的次数,  $\mathcal{H}$  表示 AMFNet 的张量函数,  $\mathcal{H}^*$  表示最优的张量函数,  $\mathcal{P}(\cdot)$  表示边缘概率分布,  $\mathcal{P}(\cdot|\cdot)$  表示条件概率分布,  $\varphi$  表示分布的权重系数,  $\mathcal{D}$  表示在高斯空间的欧式距离。

为了上述目标实现自适应的优化, 本文为自适应特征表示度量设计了专门的损失函数 ( $\mathcal{L}_{AFRM}$ ), 如式 (5) 所示。

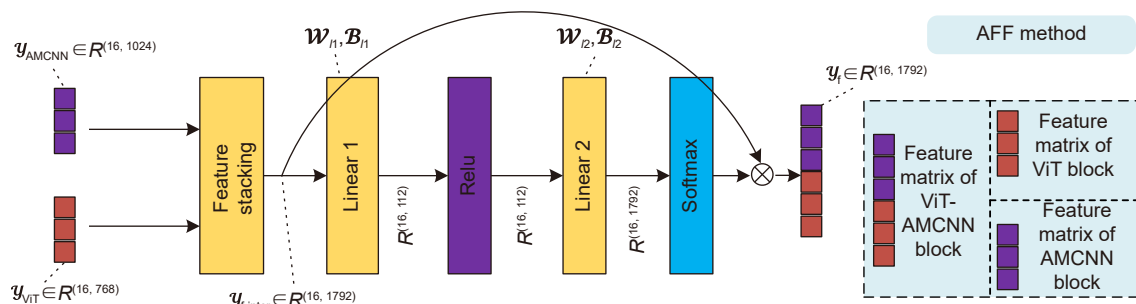


图 2 AFF 方法实现细节图

Fig. 2 Detail diagram of the AFF method implementation

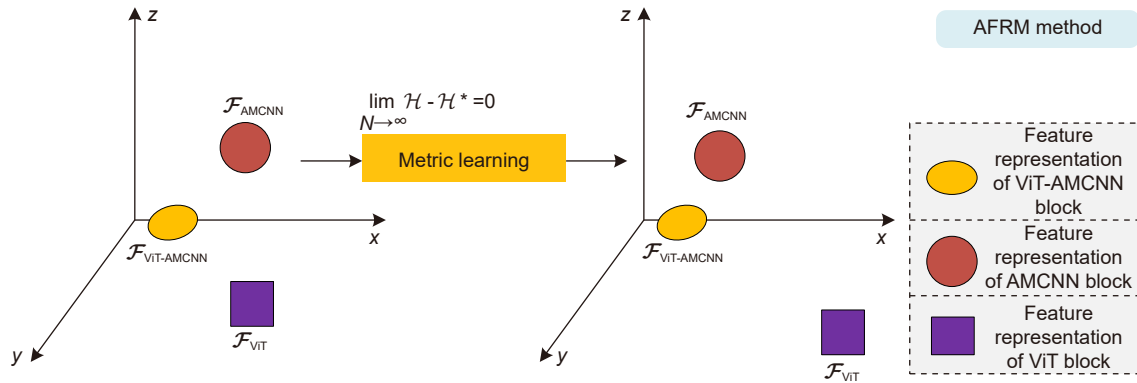


图 3 AFRM 方法实现原理图  
Fig. 3 AFRM method implementation schematic

$$\begin{aligned} \mathcal{L}_{AFRM} = & \mu * \left\{ \left\| \frac{1}{B} * \sum_i^B \phi(\mathcal{Y}_f^{(i,1792)}) - \frac{1}{B} * \sum_i^B \phi(\mathcal{Y}_{AMCNN}^{(i,1024)}) \right\|_2^2 \right. \\ & - \left. \left\| \frac{1}{B} * \sum_i^B \phi(\mathcal{Y}_f^{(i,1792)}) - \frac{1}{B} * \sum_i^B \phi(\mathcal{Y}_{ViT}^{(i,768)}) \right\|_2^2 \right\} + \frac{1-\mu}{G} \\ & * \left\{ \sum_j^G \left\| \frac{1}{B^{(G)}} * \sum_i^{B^{(G)}} \phi(\mathcal{Y}_f^{(i,1792)}) - \frac{1}{B^{(G)}} * \sum_i^{B^{(G)}} \phi(\mathcal{Y}_{AMCNN}^{(i,1024)}) \right\|_2^2 \right. \\ & \left. - \sum_j^G \left\| \frac{1}{B^{(G)}} * \sum_i^{B^{(G)}} \phi(\mathcal{Y}_f^{(i,1792)}) - \frac{1}{B^{(G)}} * \sum_i^{B^{(G)}} \phi(\mathcal{Y}_{ViT}^{(i,768)}) \right\|_2^2 \right\} \quad (5) \end{aligned}$$

式中: AMFNet的AMCNN块、ViT块和ViT-AMCNN块所得到的特征矩阵分别是 $\mathcal{Y}_{AMCNN}$ ,  $\mathcal{Y}_{ViT}$ 和 $\mathcal{Y}_{f-inter}$ ,  $B^{(G)}$ 表示每个肿瘤级别的对应的样本数,  $G$ 表示肿瘤的分级数,  $\phi$ 表示高斯核映射。

### 3 实验

#### 3.1 数据集和实验平台

基于ER IHC染色的乳腺癌病理组织显微镜成像流程数据集<sup>[35]</sup>包含414张分辨率为1300 pixels×1030 pixels的图像。该数据集包括I级、II级和III级的肿瘤病理显微镜图像。本文依此顺序裁剪出1640张patch图像,并将乳腺癌patch图像的分辨率调整为224 pixels×224 pixels。此外,1640张patch图像被分

为983、328和329张图像,分别用于训练、验证和测试,具体如表1所示。所提出的AMFNet是在PyTorch框架中使用Rmsprop进行优化。该网络模型中使用的超参数具体如下:

$i=1.0$ ,  $r=0.5$ ,  $j=0.2$ ,  $u=0.2$ , Epoch=100, Batch size=16。AMFNet的输入尺寸为224×224×3。实验的GPU是NVIDIA GeForce RTX3060 12 GB。并且本文采用类别准确度(Grade I Acc、Grade II Acc、Grade III Acc)、平均准确度(Average Acc)、特异性(Precision, P)、灵敏度(Recall, R)、F1得分(F1)和ROC曲线下面积(AUC)对模型性能进行定量分析。

#### 3.1.1 肿瘤分级准确率

本文的乳腺癌ER IHC染色的组织病理图像分为三个等级(I级、II级和III级),对于类别准确度和平均准确度的计算准则是:取置信度为0.5,大于0.5的算判对,最终将判对的除以总数得到准确度百分比。本文对应的是多分类,采用了类别数量权重来计算三种类别的平均肿瘤分级准确率,其次本文采用了Marco方式来计算三种类别的平均P、R、AUC和F1等指标。

#### 3.1.2 灵敏性(R)

当对乳腺癌ER IHC染色的组织病理图像进行自

表 1 乳腺癌 ER IHC 数据集数量分布表

Table 1 Distribution of the number of ER IHC datasets for breast cancer

Datasets	Parameter			Image size	Total
	Grade I	Grade II	Grade III		
Training set	268	355	360	224×224	983
Validation set	90	118	120	224×224	328
Testing set	90	119	120	224×224	329
Total	448	592	600	224×224	1640

动分类时, 会出现分类混淆矩阵 (如表 1 所示) 中的各种情况。为了分析真正的阳性样本中有多大比例被预测正确, 本文用召回率来分析, 见式 (6)。

$$R = \frac{TP}{TP + FN} \quad (6)$$

### 3.1.3 特异性 (P)

如表 2 所示, 为了确定预测为阳性的样本中有多少比例是真正的阳性样本, 本文使用特异性指标来进行分析, 其计算如式 (7) 所示。

$$P = \frac{TP}{TP + FP} \quad (7)$$

### 3.1.4 F1 分数 (F1)

F1 分数, 也被称为平衡 F 分数 (balanced F score), 被定义为精度和召回率的平均值。其计算如式 (8) 所示。

$$F1 = 2 \times \frac{P * R}{P + R} \quad (8)$$

### 3.1.5 ROC 曲线下面积 (AUC)

本文使用 ROC 曲线和 AUC (area under ROC curve, AUC) 来评价模型性能。根据分类器的预测结果的置信度对样本进行排序, 将样本作为阳性样本按此顺序进行预测, 每次计算真阳性率 (true positive rate, TPR) 如式 (9), 假阳性率 (false positive rate, FPR) 如式 (10), TPR 作为 ROC 曲线的纵轴, FPR 作为 ROC 曲线的横轴。

$$TPR = \frac{TP}{TP + FN} \quad (9)$$

$$FPR = \frac{FP}{TN + FP} \quad (10)$$

## 3.2 AMF 方法的消融实验

为进一步研究 AMF 方法对网络模型的影响, 本文分别对其进行消融实验验证。从表 3 可知, 在 AMF 方法中同时使用 MOO、AFRM 和 AFF 方法的平均准确度达到了 95.14%, 优于没有使用 AMF 方法的 69.00%, 优于只使用了 MOO 方法的 92.40%, 也优于使用了 MOO 和 AFRM 方法的 93.92%。

## 3.3 肿瘤分级准确率对比实验

从表 4 可知, 本文提出的 AMFNet 模型的平均分级精度为 95.14%, 明显高于 FABNet 的 87.54%、ViT-L/16 的 79.94%、ViT-B/16 的 85.41%、DenseNet+HIENet 的 85.71%、DenseNet+Nonlocal 的 83.89%、DenseNet+CBAM 的 82.98% 和 DenseNet+SENet 的 82.67%。因此, 本文提出的 AMFNet 比其他先进方法 (State Of the Art, SOTA) 有更好的肿瘤分级性能。此外, 由表 3 可知, 本文的 AMF 方法可以有效地提高 AMFNet 的模型融合能力。

## 3.4 分级结果可视化实验

从图 4 观察可知, 本文提出的 AMFNet 模型所学知识的可视化图更关注肿瘤细胞的异型核, 如多核、巨核、多形核和深色核。而病理肿瘤分级主要依赖于细胞核的异型程度。而其他 SOTA 模型 (如 FABNet) 只是关注部分异型核区和许多非典型异型细胞核区, 这导致了整体可解释性差。因此, 与 SOTA 模型相比, 本文提出的 AMFNet 得到的可解释性的肿瘤分级结果与病理理论更加一致。

此外, 本文的目的是融合 ViT 和 AMCNN 模块

表 2 分类混淆矩阵

Table 2 Classification confusion matrix

Reality	Forecast result	
	Positive	Negative
Positive	True positive (TP)	False negative (FN)
Negative	False positive (FP)	True negative (TN)

表 3 在 ER IHC 染色的乳腺癌病理组织显微镜成像中对 AMF 方法进行消融 (X 表示无, ✓ 表示有)

Table 3 Ablation of AMF method in ER IHC pathological microimaging of breast cancer

Model	AMF method			Average acc/%	P	R	F1	AUC
	MOO	AFRM	AFF					
AMFNet (ViT-AMCNN blocks)	X	X	✓	69.00	0.6934	0.6900	0.6903	0.7643
	✓	X	X	92.40	0.9240	0.9240	0.9240	0.9423
	✓	✓	X	93.92	0.9403	0.9392	0.9394	0.9532
	✓	✓	✓	<b>95.14</b>	0.9520	0.9514	0.9513	0.9617



表 4 乳腺癌 ER IHC 染色的病理组织显微成像的肿瘤分级准确率

Table 4 Tumor grading accuracy of breast cancer ER IHC histopathology microscopic imaging

Model	Grade I acc/%	Grade II acc/%	Grade III acc/%	Average acc/%	P	R	F1	AUC
Inception V3 <sup>[36]</sup>	74.71	78.19	78.01	77.20	0.7721	0.7720	0.7717	0.8290
Xception V3 <sup>[37]</sup>	71.82	68.40	76.42	72.34	0.7226	0.7234	0.7226	0.7925
ResNet50 <sup>[38]</sup>	72.96	73.77	76.08	74.47	0.7524	0.7447	0.7439	0.8085
DenseNet121 <sup>[39]</sup>	82.80	77.53	81.63	80.55	0.8059	0.8055	0.8047	0.8541
DenseNet121+Nonlocal <sup>[40]</sup>	85.88	84.39	81.97	83.89	0.8396	0.8389	0.8391	0.8791
DenseNet121+SENet <sup>[41]</sup>	79.55	83.27	84.39	82.67	0.8272	0.8267	0.8266	0.8700
DenseNet121+CBAM <sup>[42]</sup>	82.76	81.20	84.80	82.98	0.8307	0.8298	0.8294	0.8723
DenseNet121+HIENet <sup>[43]</sup>	86.21	85.59	85.48	85.71	0.8585	0.8571	0.8572	0.8928
FABNet <sup>[45]</sup>	86.05	91.29	84.90	<b>87.54</b>	0.8765	0.8754	0.8752	0.9036
ViT-S/16 <sup>[27]</sup>	54.02	55.87	69.20	60.18	60.37	60.18	0.6023	0.6970
ViT-B/16 <sup>[27]</sup>	85.87	86.32	84.17	<b>85.41</b>	0.8546	0.8541	0.8541	0.8913
ViT-B/32 <sup>[27]</sup>	68.60	70.00	73.98	71.12	0.7114	0.7112	0.7107	0.7799
ViT-L/16 <sup>[27]</sup>	78.98	79.17	81.23	79.94	0.8113	0.7994	0.7987	0.8430
ViT-L/32 <sup>[27]</sup>	50.35	61.21	59.83	58.36	0.6018	0.5836	0.5774	0.6770
AMFNet (ours)	92.66	97.02	95.12	<b>95.14</b> ↑7.6	0.9520	0.9514	0.9513	0.9617

的优点, 在优化过程中融合特征表示较为倾向 ViT, 所以 ViT-AMCNN 所学的融合特征表示的归纳偏差能力较强, 同时具备较好的可解释性。显然这样的学习策略导致 AMFNet 的可解释性优于 AMCNN 模型而弱优于 ViT 模型。这里可以从图 4 中看出, 我们发现在四幅可视化图中, AMFNet 的可解释性均优于传统的 AMCNN 模型, 而只有第一幅图优于 ViT-B/16 模型, 其余三幅是持平的。但是, 从可解释性和肿瘤分

级的总体性能来说, 本文模型都远优于 ViT 和 AMCNN 模型。

## 4 讨论与分析

### 4.1 模型的肿瘤分级精准性

由 2.2 和 2.3 节实验分析可知: 1) MOO 方法能够同时保证 ViT 块和 AMCNN 块具有良好特征表示的能力, 从而提高 AMFNet 的模型融合能力; 2) AFF

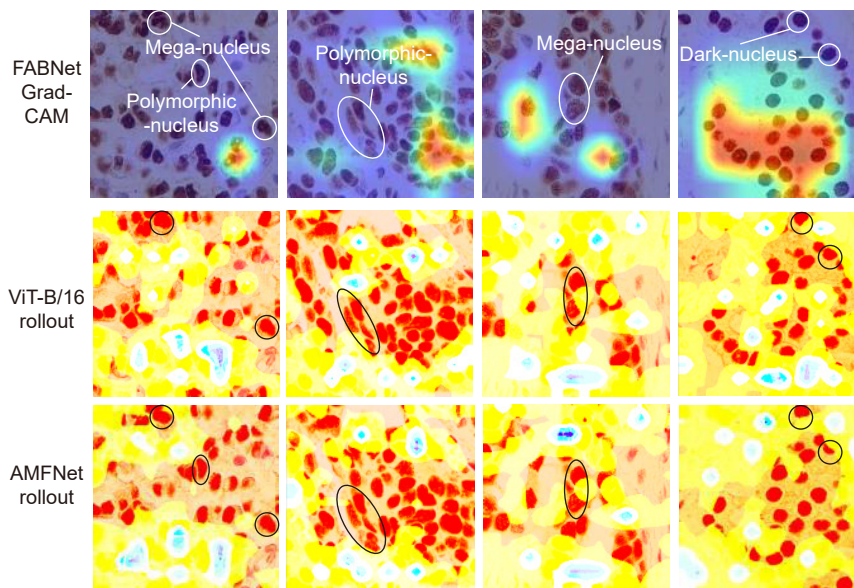


图 4 在乳腺癌 ER IHC 染色的病理组织显微成像上与 SOTA 模型进行肿瘤分级结果的可解释性对比

Fig. 4 Visually interpretative results comparisons with SOTA and our method on the breast cancer IHC microscopic imaging

方法能够实现自适应和稀疏性地去除 ViT 块和 AMCNN 块融合后的特征矩阵的冗余信息, 从而提高 AMFNet 的模型融合能力; 3) AFRM 方法通过深度度量学习, 能够有效地拉开 ViT 和 AMCNN 块所学特征表示的联合概率分布之间的距离, 从而降低融合特征矩阵的冗余度, 进而提高 AMFNet 模型的融合能力。

此外, MOO 方法充分地保证了 ViT 块和 AMC 块具有良好的特征表示能力, 是解决模型负融合现象的根本, 而 AFF 方法在 MOO 方法的基础上, 自适应和稀疏性地去除模型内部融合特征矩阵中的冗余信息, 进一步提高了模型融合能力和缓解负融合现象, 最后进行的 AFRM 方法有效地拉开了 ViT 块和 AMC 块所学特征表示之间的相似度, 进一步提升了模型融合能力。总之, 由 MOO、AFF 和 AFRM 所组成的 AMF 方法有效地解决了 AMFNet 的负融合问题, 从而提升 AMFNet 模型融合能力, 进而提升了 AMFNet 的肿瘤分级准确率。

#### 4.2 模型的可解释性

由 2.4 节的实验分析可知, 与 SOTA 模型相比, 本文提出的端到端 AMFNet 模型更加关注细胞核异型区域, 例如多核、巨核、多形核和深色核等。这与病理医生所关注的区域更加吻合, 且更加符合病理的相关分析。因此, 本文所提出的 AMFNet 模型能同时具备较好的可解释性和肿瘤分级准确性。故本文的 AMFNet 模型一定程度上解决了现有病理图像的计算机辅助诊断算法的可解释性不足的问题。

#### 4.3 模型的泛化能力分析

本文为了验证 AMFNet 方法的泛化能力和稳定性, 在更大的脑癌组织病理图像数据集<sup>[44]</sup>(4214 张病理图像, 具体如表 5 所示) 上进行了实验。其中训练集、测试集和验证集仍是按照 6:2:2 的比例进行划分, 其中训练集 2532 张, 验证集 841 张, 测试集 841 张, 其余的实验条件与乳腺癌组织病理数据集实验保持一致。脑癌组织病理图像的实验结果如表 6 和图 5 所示,

表 5 脑癌组织病理图像数据集数量分布表

Table 5 Distribution of the number of brain cancer histopathology image datasets

Datasets	Grade I	Grade II	Grade III	Grade IV	Total
Training set	276	492	873	891	2532
Validation set	91	164	290	296	841
Testing set	91	164	290	296	841
Total	458	820	1453	1483	4214

表 6 脑癌组织病理图像的肿瘤分级准确率的对比表

Table 6 Comparison table of tumor grading accuracy of histopathological images of brain cancer

Model	Metrics								
	Grade I acc/%	Grade II acc/%	Grade III acc/%	Grade IV acc/%	Average acc/%	P	R	F1	AUC
Inception V3 <sup>[36]</sup>	84.16	68.11	80.62	78.05	77.76	0.7779	0.7776	0.7766	0.8575
Xception V3 <sup>[37]</sup>	84.95	72.34	78.31	80.80	78.83	0.7976	0.7883	0.7874	0.8588
ResNet50 <sup>[38]</sup>	82.00	65.62	66.43	71.12	69.80	0.6963	0.6980	0.6961	0.8091
DenseNet121 <sup>[39]</sup>	87.05	74.75	74.86	81.75	78.95	0.7978	0.7895	0.7858	0.8625
DenseNet121+Nonlocal <sup>[40]</sup>	91.84	81.90	86.70	89.39	87.40	0.8763	0.8740	0.8727	0.9195
DenseNet121+SENet <sup>[41]</sup>	96.22	84.97	88.05	90.25	89.18	0.8925	0.8918	0.8911	0.9279
DenseNet121+CBAM <sup>[42]</sup>	92.71	82.00	89.04	86.77	87.40	0.8750	0.8740	0.8726	0.9162
DenseNet121+HIENet <sup>[43]</sup>	95.70	85.99	87.81	87.50	88.23	0.8851	0.8823	0.8820	0.9244
FABNet <sup>[15]</sup>	93.68	87.70	90.97	90.82	<b>90.61</b>	0.9072	0.9061	0.9057	0.9391
ViT-S/16 <sup>[27]</sup>	65.98	46.98	65.97	69.73	64.21	0.6375	0.6421	0.6359	0.7489
ViT-B/16 <sup>[27]</sup>	83.90	80.00	87.83	86.83	<b>85.61</b>	0.8618	0.8561	0.8553	0.9028
ViT-B/32 <sup>[27]</sup>	81.48	62.59	78.93	77.20	75.74	0.7550	0.7574	0.7541	0.8319
ViT-L/16 <sup>[27]</sup>	70.79	54.49	72.05	73.56	69.32	0.6897	0.6932	0.6902	0.7808
ViT-L/32 <sup>[27]</sup>	75.49	58.02	73.76	75.08	71.70	0.7152	0.7170	0.7134	0.8084
AMFNet (ours)	98.32	93.38	94.30	93.90	<b>94.41↑3.8</b>	0.9451	0.9441	0.9442	0.9611

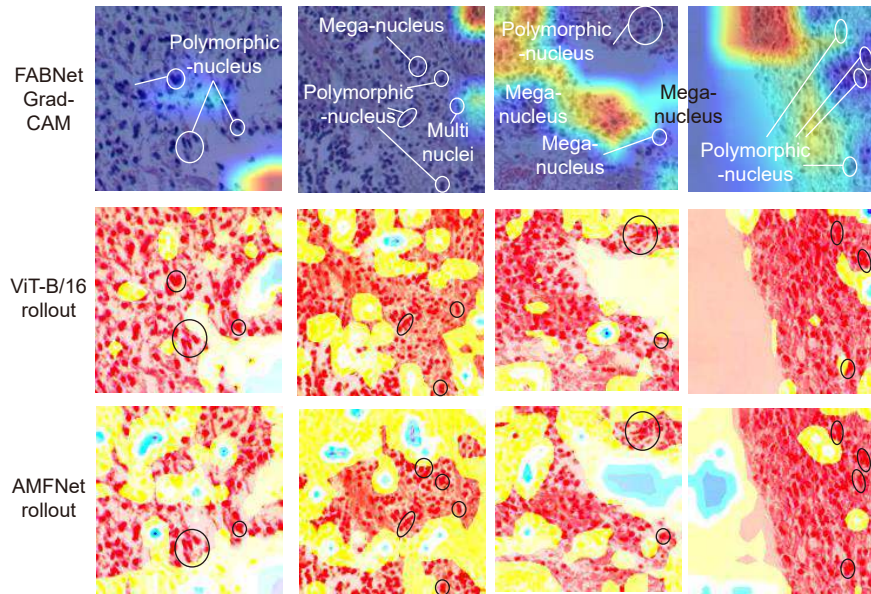


图 5 脑癌组织病理图像的可视化结果对比图

Fig. 5 Comparison of visualization results of histopathological images of brain cancer

可以看出, 本文所提出的 AMFNet 在更大的脑癌数据集上, 相比其他 13 种前沿先进模型, 仍有最佳的肿瘤分级准确率和可解释性。

## 5 结论

基于 ER IHC 染色的病理组织显微成像对乳腺癌肿瘤进行分级是一项重要而具有挑战性的工作, 它依赖于病理学家的经验和时间。特别是, 现有的辅助诊断算法的肿瘤分级结果不能同时具有良好的准确性和可解释性。因此, 本文提出了一种基于自适应模型融合的端到端深度神经网络, 其保证了 ViT 和 AMCNN 块同时具有良好的特征表示能力, 可以提高模型融合和冗余去除能力。大量的实验结果表明, 本文提出的 AMFNet 不仅比其他 SOTA 模型具有更高的肿瘤分级准确性, 而且能够对肿瘤分级结果进行更好的可视化解释, 特别是模型对异型核的关注与病理分析更加一致。通过大量的实验证实本文提出的 AMFNet 在一定程度上提高了现有乳腺癌辅助诊断算法的临床可信度。

## 参考文献

- [1] Parkin D M, Bray F, Ferlay J, et al. Estimating the world cancer burden: globocan 2000[J]. *Int J Cancer*, 2001, **94**(2): 153-156.
- [2] Robertson S, Azizpour H, Smith K, et al. Digital image analysis in breast pathology—from image processing techniques to artificial intelligence[J]. *Transl Res*, 2018, **194**: 19-35.
- [3] Desai S B, Moonim M T, Gill A K, et al. Hormone receptor status of breast cancer in India: a study of 798 tumours[J]. *Breast*, 2000, **9**(5): 267-270.
- [4] Zafrani B, Aubriot M H, Mouret E, et al. High sensitivity and specificity of immunohistochemistry for the detection of hormone receptors in breast carcinoma: comparison with biochemical determination in a prospective study of 793 cases[J]. *Histopathology*, 2000, **37**(6): 536-545.
- [5] Fuqua S A W, Schiff R, Parra I, et al. Estrogen receptor  $\beta$  protein in human breast cancer: correlation with clinical tumor parameters[J]. *Cancer Res*, 2003, **63**(10): 2434-2439.
- [6] Vagunda V, Šmardová J, Vagundová M, et al. Correlations of breast carcinoma biomarkers and p53 tested by FASAY and immunohistochemistry[J]. *Pathol Res Pract*, 2003, **199**(12): 795-801.
- [7] Baqai T, Shousha S. Oestrogen receptor negativity as a marker for high-grade ductal carcinoma in situ of the breast[J]. *Histopathology*, 2003, **42**(5): 440-447.
- [8] Sofi G N, Sofi J N, Nadeem R, et al. Estrogen receptor and progesterone receptor status in breast cancer in relation to age, histological grade, size of lesion and lymph node involvement[J]. *Asian Pac J Cancer Prev*, 2012, **13**(10): 5047-5052.
- [9] Azizun-Nisa, Bhurgri Y, Raza F, et al. Comparison of ER, PR and HER-2/neu (C-erb B 2) reactivity pattern with histologic grade, tumor size and lymph node status in breast cancer[J]. *Asian Pac J Cancer Prev*, 2008, **9**(4): 553-556.
- [10] Gurcan M N, Boucheron L E, Can A, et al. Histopathological image analysis: a review[J]. *IEEE Rev Biomed Eng*, 2009, **2**: 147-171.
- [11] Yang H, Kim J Y, Kim H, et al. Guided soft attention network for classification of breast cancer histopathology images[J]. *IEEE Trans Med Imaging*, 2020, **39**(5): 1306-1315.
- [12] BenTaieb A, Hamarneh G. Predicting cancer with a recurrent

- visual attention model for histopathology images[C]//21st International Conference on Medical Image Computing and Computer Assisted Intervention, 2018: 129–137. [https://doi.org/10.1007/978-3-030-00934-2\\_15](https://doi.org/10.1007/978-3-030-00934-2_15).
- [13] Tomita N, Abdollahi B, Wei J, et al. Attention-based deep neural networks for detection of cancerous and precancerous esophagus tissue on histopathological slides[J]. *JAMA Netw Open*, 2019, 2(11): e1914645.
- [14] Yao J W, Zhu X L, Jonnagaddala J, et al. Whole slide images based cancer survival prediction using attention guided deep multiple instance learning networks[J]. *Med Image Anal*, 2020, 65: 101789.
- [15] Huang P, Tan X H, Zhou X L, et al. FABNet: Fusion attention block and transfer learning for laryngeal cancer tumor grading in p63 IHC histopathology images[J]. *IEEE J Biomed Health Inform*, 2022, 26(4): 1696–1707.
- [16] Zhou X L, Tang C W, Huang P, et al. LPCANet: classification of laryngeal cancer histopathological images using a CNN with position attention and channel attention mechanisms[J]. *Interdiscip Sci*, 2021, 13(4): 666–682.
- [17] Liu J X, Fan J Y, Wang Q, et al. Imaging of skin structure and vessels in melanoma by swept source optical coherence tomography angiography[J]. *Opto-Electron Eng*, 2020, 47(2): 190239.  
刘敬璇, 樊金宇, 汪权, 等. SS-OCTA对黑色素瘤皮肤结构和血管的成像实验[J]. *光电工程*, 2020, 47(2): 190239.
- [18] Abnar S, Zuidema W. Quantifying attention flow in transformers[C]//Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 2020: 4190–4197. <https://doi.org/10.18653/v1/2020.acl-main.385>.
- [19] Chefer H, Gur S, Wolf L. Transformer interpretability beyond attention visualization[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 782–791. <https://doi.org/10.1109/CVPR46437.2021.00084>.
- [20] Liang L M, Zhou L S, Chen X, et al. Ghost convolution adaptive retinal vessel segmentation algorithm[J]. *Opto-Electron Eng*, 2021, 48(10): 210291.  
梁礼明, 周珑颂, 陈鑫, 等. 鬼影卷积自适应视网膜血管分割算法[J]. *光电工程*, 2021, 48(10): 210291.
- [21] Liu X, Gan Q, Liu X, et al. Joint energy active contour CT image segmentation method based on super-pixel[J]. *Opto-Electron Eng*, 2020, 47(1): 190104.  
刘侠, 甘权, 刘晓, 等. 基于超像素的联合能量主动轮廓CT图像分割方法[J]. *光电工程*, 2020, 47(1): 190104.
- [22] Gao Z Y, Hong B Y, Zhang X L, et al. Instance-based vision transformer for subtyping of papillary renal cell carcinoma in histopathological image[C]//24th International Conference on Medical Image Computing and Computer Assisted Intervention, 2021: 299–308. [https://doi.org/10.1007/978-3-030-87237-3\\_29](https://doi.org/10.1007/978-3-030-87237-3_29).
- [23] Wang X Y, Yang S, Zhang J, et al. TransPath: Transformer-based self-supervised learning for histopathological image classification[C]//24th International Conference on Medical Image Computing and Computer Assisted Intervention, 2021: 186–195. [https://doi.org/10.1007/978-3-030-87237-3\\_18](https://doi.org/10.1007/978-3-030-87237-3_18).
- [24] Li H, Yang F, Zhao Y, et al. DT-MIL: deformable transformer for multi-instance learning on histopathological image[C]//24th International Conference on Medical Image Computing and Computer Assisted Intervention, 2021: 206–216. [https://doi.org/10.1007/978-3-030-87237-3\\_20](https://doi.org/10.1007/978-3-030-87237-3_20).
- [25] Chen H Y, Li C, Wang G, et al. GasHis-transformer: a multi-scale visual transformer approach for gastric histopathological image detection[J]. *Pattern Recognit*, 2022, 130: 108827.
- [26] 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), 2021: 1235–1240. <https://doi.org/10.1109/BIBM52615.2021.9669903>.
- [27] Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16x16 words: transformers for image recognition at scale[C]//9th International Conference on Learning Representations, 2021.
- [28] Breiman L. Bagging predictors[J]. *Mach Learn*, 1996, 24(2): 123–140.
- [29] Schapire R E. The boosting approach to machine learning: an overview[M]//Denison D D, Hansen M H, Holmes C C, et al. *Nonlinear Estimation and Classification*. New York: Springer, 2003: 149–171. [https://doi.org/10.1007/978-0-387-21579-2\\_9](https://doi.org/10.1007/978-0-387-21579-2_9).
- [30] Ting K M, Witten I H. Stacking bagged and Daggged models[C]//Proceedings of the Fourteenth International Conference on Machine Learning, 1997: 367–375. <https://doi.org/10.5555/645526.657147>.
- [31] Ke W C, Fan F, Liao P X, et al. Biological gender estimation from panoramic dental X-ray images based on multiple feature fusion model[J]. *Sens Imaging*, 2020, 21(1): 54.
- [32] Zhou Z H. *Machine Learning*[M]. Beijing: Tsinghua University Press, 2016: 171–196.  
周志华. *机器学习*[M]. 北京: 清华大学出版社, 2016: 171–196.
- [33] Selvaraju R R, Cogswell M, Das A, et al. Grad-CAM: visual explanations from deep networks via gradient-based localization[C]//Proceedings of the 2017 IEEE International Conference on Computer Vision, 2017: 618–626.
- [34] Selvaraju R R, Cogswell M, Das A, et al. Grad-CAM: visual explanations from deep networks via gradient-based localization[C]//Proceedings of the IEEE International Conference on Computer Vision, 2017: 618–626. <https://doi.org/10.1109/ICCV.2017.74>.
- [35] Kostopoulos S, Glotsos D, Cavouras D, et al. Computer-based association of the texture of expressed estrogen receptor nuclei with histologic grade using immunohistochemically-stained breast carcinomas[J]. *Anal Quant Cytol Histol*, 2009, 31(4): 187–196.
- [36] Szegedy C, Vanhoucke V, Ioffe S, et al. Rethinking the inception architecture for computer vision[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016: 2818–2826. <https://doi.org/10.1109/CVPR.2016.308>.
- [37] Chollet F. Xception: deep learning with depthwise separable convolutions[C]//30th IEEE Conference on Computer Vision and Pattern Recognition, 2017: 1800–1807. <https://doi.org/10.1109/CVPR.2017.195>.
- [38] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770–778. <https://doi.org/10.1109/CVPR.2016.90>.
- [39] Huang G, Liu Z, Van Der Maaten L, et al. Densely connected

- convolutional networks[C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017: 2261–2269. <https://doi.org/10.1109/CVPR.2017.243>.
- [40] Wang X L, Girshick R, Gupta A, et al. Non-local neural networks[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018: 7794–7803. <https://doi.org/10.1109/CVPR.2018.00813>.
- [41] Hu J, Shen L, Albanie S, et al. Squeeze-and-excitation networks[J]. *IEEE Trans Pattern Anal Mach Intell*, 2020, 42(8): 2011–2023.
- [42] Woo S, Park J, Lee J Y, et al. CBAM: convolutional block attention module[C]//*Proceedings of the 15th European Conference on Computer Vision*, 2018: 3–19. [https://doi.org/10.1007/978-3-030-01234-2\\_1](https://doi.org/10.1007/978-3-030-01234-2_1).
- [43] Sun H, Zeng X X, Xu T, et al. Computer-aided diagnosis in histopathological images of the endometrium using a convolutional neural network and attention mechanisms[J]. *IEEE J Biomed Health Inform*, 2020, 24(6): 1664–1676.
- [44] Glotsos D, Kalatzis I, Spyridonos P, Kostopoulos S, Daskalakis A, Athanasiadis E, Ravazoula P, Nikiforidis G, and Cavouras D. Improving accuracy in astrocytomas grading by integrating a robust least squares mapping driven support vector machine classifier into a two level grade classification scheme[J]. *Comput Methods Programs Biomed*, 2008, 90(3): 251–261.

## 作者简介



黄盼 (1992-), 男, 重庆大学光电工程学院, 博士研究生, 研究方向为病理图像处理、可解释性学习、域自适应及多模态融合。

E-mail: panhuang@cqu.edu.cn



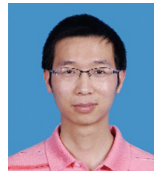
【通信作者】肖华亮 (1973-), 男, 副主任医师, 副教授, 硕士研究生导师, 病理学博士。现任陆军特色医学中心 (陆军军医大学大坪医院) 病理科主任、肿瘤分子病理实验室负责人。研究方向为肿瘤微环境调控与侵袭、分子病理学。

E-mail: dpbl\_xhl@163.com



【通信作者】田素坤 (1990-), 男, 山东大学机械工程学院, 博士后, 研究方向为牙龈图像重建、医学图像处理。

E-mail: sukhum169@hotmail.com

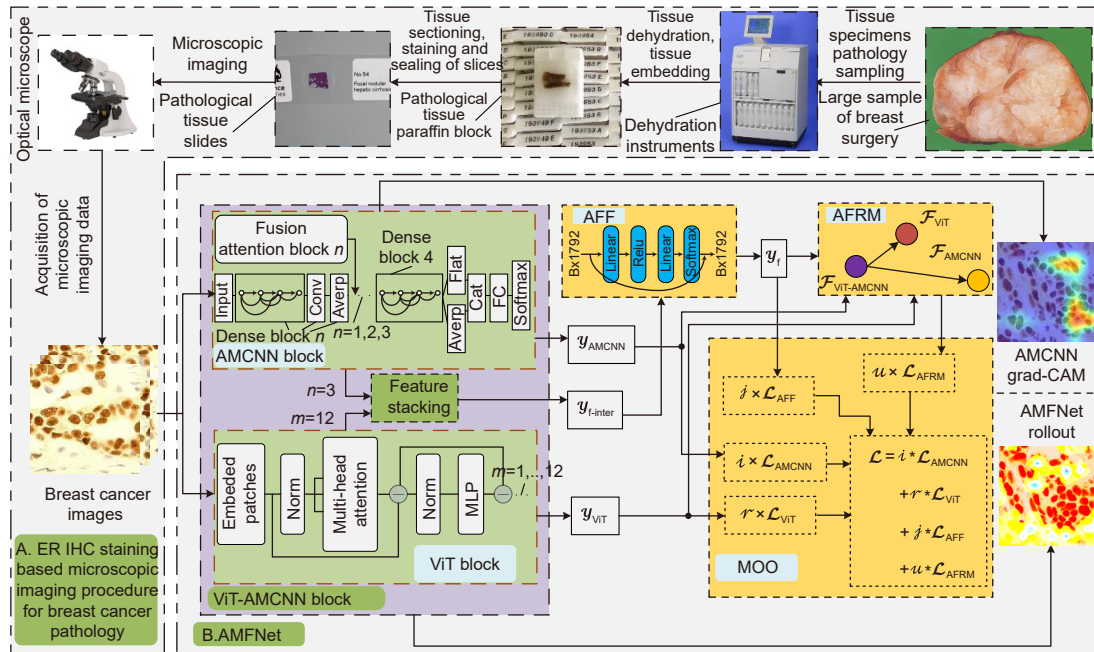


【通信作者】冯鹏 (1981-), 男, 重庆大学光电工程学院, 副教授, 研究方向为基于人工智能的多模医学图像处理与分析研究。

E-mail: coe-fp@cqu.edu.cn

# Breast tumor grading network based on adaptive fusion and microscopic imaging

Huang Pan<sup>1</sup>, He Peng<sup>1</sup>, Yang Xing<sup>2</sup>, Luo Jiayang<sup>1</sup>, Xiao Hualiang<sup>3\*</sup>, Tian Sukun<sup>4\*</sup>, Feng Peng<sup>1\*</sup>



**Overview:** Breast cancer is the most common cancer among women. Tumor grading based on microscopic imaging is important for the diagnosis and prognosis of breast cancer, and the results need to be highly accurate and interpretable. Breast cancer tumor grading relies on pathologists to assess the morphological status of tissues and cells in the microscopic images of tissue sections, such as tissue differentiation, nuclear isotypes, and mitotic counts. A strong statistical correlation between the hematoxylin-Eosin (HE) stained microscopic imaging samples and the progesterone Receptor (ER) immunohistochemically (IHC) stained microscopic imaging samples has been documented, i.e., the ER status is strongly correlated with the tumor tissue grading. Therefore, it is a meaningful task to use deep learning models to research the breast tumor grading in ER IHC pathology images exploratively. At present, the CNN module integrating attention has a strong ability of induction bias but poor interpretability, while the Vision Transformer (ViT) block-based deep network has better interpretability but weaker ability of induction bias. In this paper, we propose an end-to-end deep network with adaptive model fusion by fusing ViT blocks and CNN blocks with integrated attention. Due to the negative fusion phenomenon of the existing model fusion methods, while it is impossible to guarantee that ViT blocks and CNN blocks with integrated attention have good feature representation capability at the same time; in addition, the high similarity and redundant information between the two feature representations lead to a poor model fusion capability. To this end, this paper proposes an adaptive model fusion method that includes multi-objective optimization, adaptive feature representation metric, and adaptive feature fusion, which effectively improves the fusion ability of the model. The accuracy of the model is 95.14%, which is 9.73% better than that of ViT-B/16 and 7.6% better than that of FABNet. The visualization of the model focuses more on the regions of nuclear heterogeneity (e.g., giant nuclei, polymorphic nuclei, multinuclei and dark nuclei), which is more consistent with the regions of interest to pathologists. Overall, the proposed model outperforms the current state-of-the-art model in terms of accuracy and interpretability.

Huang P, He P, Yang X, et al. Breast tumor grading network based on adaptive fusion and microscopic imaging[J]. *Opto-Electron Eng*, 2023, 50(1): 220158; DOI: 10.12086/oe.2023.220158

<sup>1</sup>Key Laboratory of Optoelectronic Technology & Systems (Ministry of Education), College of Optoelectronic Engineering, Chongqing University, Chongqing 400044, China; <sup>2</sup>College of Computer and Network Security, Chengdu University of Technology, Chengdu, Sichuan 610000, China; <sup>3</sup>Daping Hospital, Department of Pathology, Army Military Medical University, Chongqing 400037, China; <sup>4</sup>School of Mechanical Engineering, Shandong University, Jinan, Shandong 250000, China

\* E-mail: dpbl\_xhl@163.com; sukhum169@hotmail.com; coe-fp@cqu.edu.cn