论著·智能医学与预测模型

# 环形特征融合网络(RFFNet):一种融合CNN与Transformer并校正内镜畸变实现炎症性肠病高精度实时诊断的深度学习模型

李华龙[1],魏艳玲[2],阮广聪[2],孟 薇[2],吴 毅[1],李 颖[1],唐嘉杰[3],刘静静[1],粘永健[1] [1]陆军军医大学(第三军医大学)生物医学工程与影像医学系数字医学教研室,重庆;[2]陆军军医大学(第三军医大学)大坪医院消化内科,重庆;[3]陆军军医大学(第三军医大学)第二附属医院信息科,重庆

[**摘要**] **目的** 借助深度学习技术辅助内镜医师对克罗恩病(Crohn's disease,CD)、溃疡性结肠炎(ulcerative colitis,UC)与正常(Normal)图像肠镜进行准确诊断。**方法** 收集2018年1月至2020年11月陆军军医大学大坪医院消化内科与浙江大学医学院附属邵逸夫医院消化内科1 309例受试者共24 492幅肠镜图像,其中包括4 729张CD图像(424例)、7 074张UC图像(605例)与12 689张正常(Normal)图像(280例)。每类按照7∶1∶2的比例将病例随机划分为训练集915例17 136张图像,内部测试集131例2 626张图像,外部验证集263例4 730张图像,分别用于对模型进行训练、测试及验证。针对卷积神经网络(Convolutional Neural Network,CNN)与Transformer网络特征融合问题及肠镜图像畸变问题,本研究基于ResNeSt50与MViTv2构建环形特征融合网络(Ring-Feature Fusion Network,RFFNet)。该网络通过双向通道-空间注意力实现跨阶段特征融合,有效结合CNN局部特征提取与Transformer全局建模优势;引入环形特征融合机制,减轻肠镜图像桶形畸变与景深现象,提升模型对远端肠腔几何特性的适应能力。与现有5种深度学习模型进行对比,通过混淆矩阵展示各模型在外部验证集上的识别结果,利用准确率、灵敏度、特异度、$F_1$值与曲线下面积(area under the curve,AUC)等各项指标验证RFFNet的性能优势;通过消融实验,在外部验证集上利用准确率与$F_1$值等指标验证RFFNet中各改进点的有效性与必要性;通过梯度加权类激活映射(gradient-weighted class activation mapping,Grad-CAM)算法生成类激活映射图,直观地展示模型关注区域。**结果** RFFNet在外部验证集上总体准确率为95.68%(95%*CI*:94.14%~98.58%),总体AUC为0.987(95%*CI*:0.985~0.990)。CD、UC和正常三类别的AUC分别为0.982(95%*CI*:0.979~0.986)和0.981(95%*CI*:0.977~0.984)和0.997(95%*CI*:0.995~0.998);灵敏度分别为93.35%(95%*CI*:91.67%~94.71%)、94.00%(95%*CI*:92.60%~95.16%)和98.56%(95%*CI*:97.99%~98.97%);特异度分别为94.58%(95%*CI*:93.81%~95.27%)、94.20%(95%*CI*:93.36%~94.94%)和98.61%(95%*CI*:98.05%~99.01%);$F_1$值分别为87.81%(95%*CI*:86.28%~89.25%)、90.05%(95%*CI*:89.00%~91.25%)和98.58%(95%*CI*:98.22%~98.92%),表明RFFNet实现了较高精度诊断,与单独使用CNN、Transformer模型相比,RFFNet的总体准确率提升显著,差异具有统计学意义($P$<0.05)。消融实验证实,模型能够通过动态空间注意力机制,深度融合CNN的局部细粒度特征提取能力与Transformer的全局上下文建模优势,与只使用CNN模型相比,CNN-Transformer架构的总体准确率提升了0.11%;能够通过环形特征校正肠镜光学畸变与景深衰减,增强对远端肠腔几何结构的建模

[**基金项目**] 重庆市技术创新与应用发展专项重点项目(CSTB2022TIAD-KPX0161);重庆市自然科学基金创新发展联合基金项目(CSTB2024NSCQ-LZX0141);重庆英才·创新创业领军人才项目(CQYC20220303576)

[**通信作者**] 粘永健,E-mail:yongjian_nian@163.com

2097-0927/Copyright © 2026,陆军军医大学(第三军医大学)。这篇开放获取文章遵循CC BY许可协议(https://creativecommons.org/licenses/by/4.0/)

**引用格式**:李华龙,魏艳玲,阮广聪,等. 环形特征融合网络(RFFNet):一种融合CNN与Transformer并校正内镜畸变实现炎症性肠病高精度实时诊断的深度学习模型[J]. 陆军军医大学学报,2026,48(04):479-492. DOI:10.16016/j.2097-0927.202510069.

**Citation**:LI H L,WEI Y L,RUAN G C,*et al*. Ring-Feature Fusion Network(RFFNet):a deep learning model integrating CNN and transformer with endoscopic distortion correction for high-accuracy real-time diagnosis of inflammatory bowel disease[J]. J Army Med Univ,2026,48(04):479-492. DOI:10.16016/j.2097-0927.202510069.

能力,使用环形特征后RFFNet总体准确率提升了0.57%。类激活映射图表明改进的模型具有自适应捕获肠镜结构特征的能力。**结论** RFFNet通过双骨干架构深度耦合CNN局部纹理感知与Transformer全局依赖建模,并以环形特征显式校正肠镜畸变,实现了CD、UC与正常黏膜的高精度实时分型,实现了CD、UC和正常图像的高精度实时诊断。

[**关键词**] 炎症性肠病;深度学习;溃疡性结肠炎;克罗恩病
[**中图法分类号**] R319;R445;R574.62 [**文献标志码**] A

# Ring-Feature Fusion Network(RFFNet): a deep learning model integrating CNN and transformer with endoscopic distortion correction for high-accuracy real-time diagnosis of inflammatory bowel disease

LI Hualong[1], WEI Yanling[2], RUAN Guangcong[2], MENG Wei[2], WU Yi[1], LI Ying[1], TANG Jiajie[3], LIU Jingjing[1], NIAN Yongjian[1] [1]Department of Digital Medicine, College of Biomedical Engineering and Imaging Medicine, Army Medical University (Third Military Medical University), Chongqing; [2]Department of Gastroenterology, Daping Hospital, Army Medical University (Third Military Medical University), Chongqing; [3]Department of information, Second Affiliated Hospital, Army Medical University(Third Military Medical University), Chongqing, China

[**Abstract**] **Objective** To leverage deep learning technology to assist endoscopists in accurately diagnosing colonoscopy images of Crohn's disease (CD), ulcerative colitis (UC), and Normal tissue. **Methods** A total of 24 492 colonoscopy images from 1 309 subjects were collected from Department of Gastroenterology, Daping Hospital, Army Medical University, and Sir Run Run Shaw Hospital between January 2018 and November 2020. The dataset included 4 729 CD images from 424 cases, 7 074 UC images from 605 cases, and 12 689 Normal images from 280 cases. The images were randomly split by case in a 7∶1∶2 ratio into a training set (915 cases, 17 136 images), an internal test set (131 cases, 2 626 images), and an external validation set (263 cases, 4 730 images). To address feature fusion between CNN and Transformer architectures and correct endoscopic image distortion, we constructed a Ring-Feature Fusion Network (RFFNet) based on ResNeSt50 and MViTv2. The network employs bidirectional channel-spatial attention for cross-stage feature fusion, effectively combining the local feature extraction strengths of CNN with the global modeling capabilities of Transformer. An innovative ring-feature mechanism was introduced to handle barrel distortion and depth-of-field phenomena in colonoscopy images, enhancing the model's adaptability to the geometric properties of the distal intestinal lumen. RFFNet was compared with five existing deep learning models. Recognition results on the external validation set were visualized using confusion matrices, and performance was evaluated using accuracy, sensitivity, specificity, $F_1$-score, and area under the curve (AUC). Ablation studies were conducted on the external validation set using metrics like accuracy and F1-score to validate the effectiveness and necessity of each improvement in RFFNet. Gradient-weighted class activation mapping (Grad-CAM) was used to generate heatmaps, visually demonstrating the model's focus areas and improving interpretability. **Results** On the external validation set, RFFNet achieved an overall accuracy of 95.68% (95%*CI*: 94.14% to 98.58%) and an overall AUC of 0.987 (95%*CI*: 0.985 to 0.990). For CD, UC, and Normal classes, the AUCs were 0.982 (95%*CI*: 0.979 to 0.986), 0.981 (95%*CI*: 0.977 to 0.984), and 0.997 (95%*CI*: 0.995 to 0.998), respectively; sensitivities were 93.35% (95%*CI*: 91.67% to 94.71%), 94.00% (95%*CI*: 92.60% to 95.16%), and 98.56% (95%*CI*: 97.99% to 98.97%), respectively; specificities were 94.58% (95%*CI*: 93.81% to 95.27%), 94.20% (95%*CI*: 93.36% to 94.94%), and 98.61% (95%*CI*: 98.05% to 99.01%), respectively; F1-scores were 87.81% (95%*CI*: 86.28% to 89.25%), 90.05% (95%*CI*: 89.00% to 91.25%), and 98.58% (95%*CI*: 98.22% to 98.92%), respectively. These results indicate that

RFFNet achieved high diagnostic precision. Compared to using CNN or Transformer models alone, RFFNet's overall accuracy improvement was statistically significant ($P$<0.05). Ablation studies confirmed that the model, through a dynamic spatial attention mechanism, deeply integrates CNN's local fine-grained feature extraction with Transformer's global contextual modeling advantages. The CNN-Transformer architecture improved overall accuracy by 0.11% compared to using CNN alone, respectively. The ring-feature mechanism corrected endoscopic optical distortion and depth-of-field attenuation, enhancing modeling of the distal lumen geometry, and improved RFFNet's overall accuracy by 0.57%. Grad-CAM heatmaps demonstrated the model's adaptive ability to capture structural features in colonoscopy images. **Conclusion** RFFNet, through its dual-backbone architecture that deeply couples CNN's local texture perception with Transformer's global dependency modeling, and explicit correction of endoscopic distortion *via* ring-features, enables high-accuracy real-time classification of CD, UC, and normal mucosa.

[**Key words**] inflammatory bowel disease; deep learning; ulcerative colitis, the Crohn's disease

Supported by the Chongqing Technology Innovation and Application Development Special Key Project (CSTB2022TIAD-KPX0161), the Chongqing Natural Science Foundation Innovation and Development Joint Fund (CSTB2024NSCQ-LZX0141) and the Chongqing Yingcai·Innovation and Entrepreneurship Leading Talent Project (CQYC20220303576). Corresponding author: NIAN Yongjian, E-mail: yongjian_nian@163.com

炎症性肠病(inflammatory bowel disease,IBD)是一组病因未明的慢性非特异性肠道炎症性疾病[1],主要包括溃疡性结肠炎(ulcerative colitis,UC)和克罗恩病(Crohn's disease,CD),二者占总体的85%~90%;另有10%~15%为无法明确分类的"不确定型结肠炎"(indeterminate colitis,IC)[2]。IBD会给患者带来诸多的困扰,除腹泻、出血、腹痛等肠道症状外,还常伴随着如关节炎、皮肤病变、眼部炎症、肝脏疾病等并发症,显著降低患者的生活质量[3]。经研究表明,IBD患者罹患肠道原发恶性肿瘤的风险也显著增高,例如结直肠癌、小肠腺癌、肛管癌等[4]。当前IBD已成为一种全球性疾病,虽然在欧美国家该病的发病率已进入平台期,但在亚洲、中美洲及北欧部分地区呈上升趋势[5-7]。由于IBD涉及多维度病理机制,其治疗需要综合性的手段[8]。在IBD的诊断中,肠镜检查常用于诊断UC和CD[9],但高度依赖操作者经验[10]。基层医院因合格的内镜医师短缺,导致初诊误诊率显著升高,而误诊又进一步增加了不必要的肠段切除风险、术后吻合口瘘发生率及整体手术并发症概率[11]。因此,开发客观、可泛化、易落地的计算机辅助诊断工具成为迫切需求。

人工智能(artificial intelligence,AI)技术在IBD亚型识别领域取得了显著进展。TONG等[12]比较了随机森林和卷积神经网络(convolutional neural network,CNN)两种方法,在拥有6 399例内镜患者的文本描述数据集上实现了疾病的高精度诊断,其中CNN总体表现明显优于随机森林法,在UC、CD和肠结核三分类任务中精确度分别为0.99、0.87和0.52;召回率分别为0.97、0.83和0.81,处理效率显著优于传统人工诊断。此外,戚婧等[13]引入二阶统计方法以增强模型特征表示能力,构建的SABA-ResNet模型在包含1 576例受试者的34 300张内镜图像测试集上,对CD、UC和正常组织的总体分类准确率达到92.67%(95%*CI*:91.91%~93.37%)。然而,以上研究大多采用单一CNN模型架构,难以捕捉肠镜图像特征的长距离空间依赖关系,限制了跨区域特征关联的建模能力。目前,已有研究采用混合架构来实现肠镜疾病的高精度分类。UDRISTOIU等[14]以CNN-LSTM构建模型以融合时空特征,在共焦激光内镜图像上获得了95.30%的准确率。MAURÍCIO等[15]通过视觉变换器(vision transformer,ViT)和CNN共同构建模型以提取肠镜图像特征,并结合知识蒸馏技术压缩模型参数量,显著提升了推理效率。

尽管混合架构颇具成效,但均未考虑肠镜成像的桶形畸变与景深现象对肠镜识别产生的影响。由于广角肠镜的视场角可达170°,肠镜图像会因此产生明显的桶形畸变,其单个方向的畸变偏移量最大可达14像素[16];因肠镜镜头的景深限制,加之肠道弯曲的解剖结构的影响,在操作过程中常使图像呈现出近端失焦、中段清晰、远端模糊的景深分布现象。若忽视肠镜图像特有的桶形畸变以及景深现象,则会使模型在全局特征与局部特征融合时引入畸变噪声,从而限制模型的鲁棒性。如何同时利用CNN局部纹理敏感性与Transformer捕捉长距离依赖建模能力,并针对肠镜畸变进行显式校正,已成为突破性能的关键。

为解决上述问题,本研究从2个角度设计深度学

习模型来提高IBD的诊断分型的性能：其一，设计跨阶段注意力模块，在双向通道-空间注意力基础上，通过动态空间注意力加权机制引导模型聚焦高贡献区域，实现CNN局部特征提取能力与Transformer全局建模优势的深度耦合；其二，基于肠镜几何特性构建环形景深权重图，并融入模型训练过程，提升模型针对远端组织肠腔几何特性的适应性，使模型能够有效校正桶形畸变并适应景深现象，以期实现CD、UC与正常黏膜的高精度实时分型和对肠镜图像的准确诊断。

## 1 材料与方法

### 1.1 总体研究流程

本研究聚焦于CD与UC的计算机辅助鉴别诊断难题，总体研究流程如图1所示。本研究构建了一个包含CD、UC和正常图像（Normal）3种类别的IBD患者的肠镜图像数据集。针对肠镜图像的桶形畸变与景深现象，基于ResNeSt50[17]与MViTv2[18]两种网络，提出环形特征融合网络（Ring-Feature Fusion Network，RFFNet）网络。为评估所提RFFNet的性能，引入5种现有深度学习网络，包括ResNet50[19]、ViT[20]、Swin-T[21]、ResNeSt50和MViTv2。在相同数据集上训练能够识别CD、UC和正常肠镜图像的模型，通过比较不同模型的各分类指标，以验证RFFNet性能的优越性。此外，通过设计消融实验，进一步验证网络中关键模块的有效性与必要性，结合梯度加权类激活映射[22]（gradient-weighted class activation mapping，Grad-CAM）算法来生成类激活映射图。最后，综合对比实验结果，全面评估RFFNet在临床辅助诊断方面的应用潜力与价值。

### 1.2 数据集构建

收集2018年1月至2020年11月陆军军医大学大坪医院消化内科与浙江大学医学院附属邵逸夫医院消化内科共1 576例受试者数据[13]，剔除质量不合格图像后，共1 309例受试者纳入研究，包括CD、UC与Normal 3种类别24 492幅肠镜图像，其中CD患者424例（4 729幅图像），UC患者605例（7 074幅图像），Normal组280例（12 689幅图像）。本研究已经陆军军医大学大坪医院伦理委员会的批准（2018-137）以及浙江大学医学院附属邵逸夫医院伦理委员会批准（2024-1064）。数据集具体情况见表1。

本研究按7∶1∶2的比例将数据集按病例随机划分为训练集、内部测试集和外部验证集，其中训练集共915例17 136张图像，内部测试集共131例2 626张图像，外部验证集共263例4 730张图像。具体数据集划分见表2。

### 1.3 整体网络结构

RFFNet结构如图2所示，CNN部分参考ResNeSt50构建，由一个Stem层和3个阶段组成，Stem层中包括1个7×7卷积层（Conv），1个批归一化层（BatchNorm），1个ReLU激活函数及1个最大池化层（MaxPool），在每个阶段中则包括1个下采样模块（DownBlock）与若干

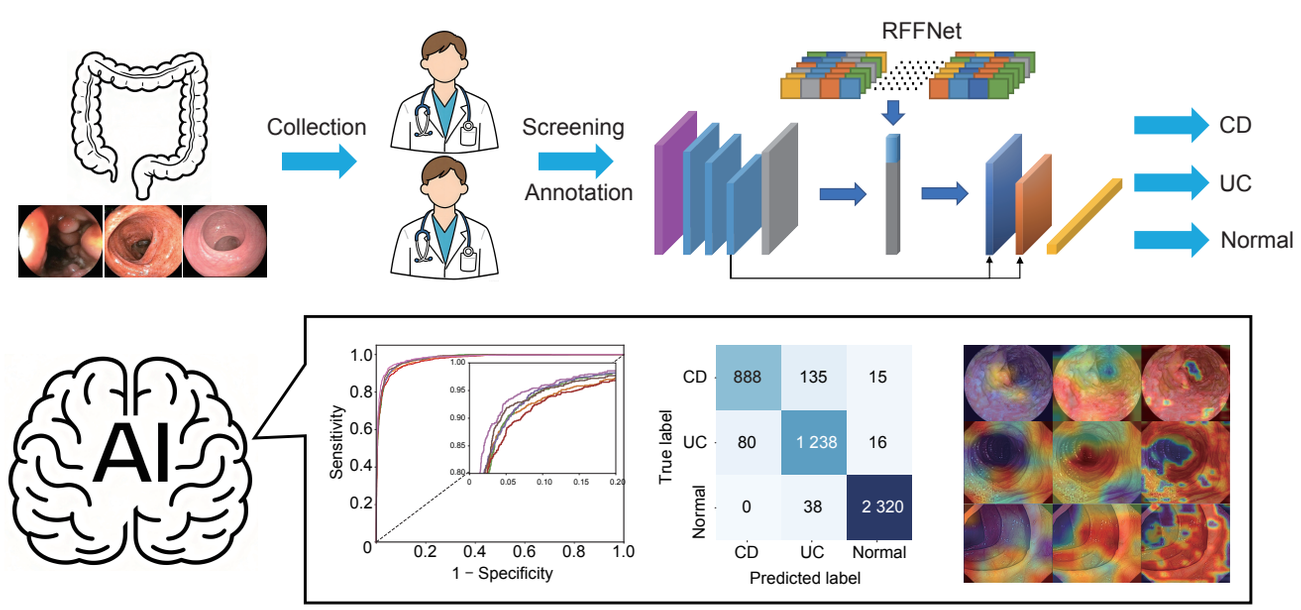



**图1 总体研究流程**

**Figure 1 Overall research process**

表1 大坪医院与邵逸夫医院中CD、UC和Normal的肠镜图像及病例数分布(95%*CI*)

**Table 1 Distribution of colonoscopy images and case counts for Crohn´s disease, ulcerative colitis, and Normal in Daping Hospital and Sir Run Run Shaw Hospital** (95%*CI*)

| Disease subtype | Daping Hospital | | Sir Run Run Shaw Hospital | | Total | |
|---|---|---|---|---|---|---|
| | Number of cases | Number of images | Number of cases | Number of images | Number of cases | Number of images |
| CD | 49(36~65) | 1 093(1 029~1 160) | 375(338~415) | 3 636(3 519~3 756) | 424(385~466) | 4 729(4 595~4 866) |
| UC | 269(238~303) | 3 379(3 266~3 495) | 336(301~374) | 3 695(3 577~3 816) | 605(558~655) | 7 074(6 910~7 241) |
| Normal | 280(248~315) | 12 689(12 469~12 912) | 0(0~3) | 0(0~3) | 280(248~315) | 12 689(12 469~12 912) |
| Total | 598(551~648) | 17 161(16 905~17 420) | 711(660~765) | 7 331(7 164~7 501) | 1 309(1 239~1 382) | 24 492(24 186~24 801) |

表2 CD、UC和Normal肠镜图像在训练集、内部测试集及外部验证集的划分情况(95%*CI*)

**Table 2 Partition of colonoscopy images for Crohn's disease, ulcerative colitis , and Normal across training, internal test, and external validation sets**(95%*CI*)

| Disease subtype | Training set | | Internal test set | | External validation set | |
|---|---|---|---|---|---|---|
| | Number of cases | Number of images | Number of cases | Number of images | Number of cases | Number of images |
| CD | 296(263~332) | 3 224(3 114~3 337) | 43(31~58) | 467(426~511) | 85(68~105) | 1 038(976~1 103) |
| UC | 423(384~465) | 4 972(4 835~5 112) | 60(46~77) | 768(715~824) | 122(101~146) | 1 334(1 263~1 408) |
| Normal | 196(170~225) | 8 940(8 756~9 127) | 28(19~40) | 1 391(1 319~1 466) | 56(42~73) | 2 358(2 264~2 455) |
| Total | 915(857~976) | 17 136(16 880~17 395) | 131(110~155) | 2 626(2 527~2 728) | 263(232~297) | 4 730(4 596~4 867) |

个分散注意力模块(SplAtBlock);Transformer部分参考MViTv2网络,由2个阶段组成,每个阶段包括2个多尺度Transformer模块(MTrans)。在CNN的3个阶段(Stage1~Stage3)输出后将特征图进行2次上采样(UpSample),将上采样后特征与Stem层输出特征进行拼接,通过卷积前馈网络(convolutional feed-forward network,CFFN)[23]增强特征表达,并将特征转换为序列特征。环形图像经过环形特征提取器生成环形特征,与序列特征进行拼接,输入Transformer部分第一阶段(Stage4),使模型能够将环形特征纳入学习范畴,并将Stage4与Stage3的输出特征进行跨阶段注意力融合,实现CNN局部特征与Transformer全局特征的深度融合。将Transformer二阶段(Stage5)输出特征与Stage3的输出特征进行一次跨阶段注意力融合,确保图像浅层特征不会丢失。最后经过平均池化层(Avgpool)与全连接层(fully connected, FC)输出3种类别的概率分数。

### 1.4 跨阶段注意力融合模块

跨阶段注意力融合模块(cross-stage attention fusion module,CSAF)如图3所示,模块需输入不同阶段的CNN特征与Transformer特征。为保证CNN特征与Transformer特征能正常融合,在输入时对CNN特征进行双线性插值上采样,使其空间尺寸与Transformer特征匹配。采样过后的CNN特征与输入的Transformer特征进行拼接得到合并特征,合并特征的通道数为CNN特征与Transformer特征通道数之和,再通过1个平均池化层(AvgPool)与2个1×1卷积层(Conv)对合并特征进行特征提取,并利用残差结构对合并特征进行增强。计算过程如式①所示:

$$F_{Con'}^{CSAF}=\sigma(Conv(Conv(AvgPool(F_{Con}^{CSAF}))))\otimes F_{Con}^{CSAF}\oplus F_{Con}^{CSAF} \quad ①$$

其中$\sigma(\cdot)$为Sigmoid函数,$Conv(\cdot)$为卷积运算,$AvgPool(\cdot)$为平均池化,$\otimes$为逐元素点乘,$\oplus$为逐元素相加。此时CNN特征与Transformer特征已完成初步融合,再通过注意力机制对合并特征进行进一步的提取。首先采用双支池化策略,通过平均池化、最大池化(MaxPool)与多层感知机(multi-layer perception, MLP)生成基于平均的通道权重和基于最大的通道权重,将两分支权重相加后通过Sigmoid函数约束至[0,1]区间与合并特征进行融合得到通道增强特征。过程如式②所示:

$$F_{Con''}^{CSAF}=\sigma(MLP(AvgPool(F_{Con'}^{CSAF}))+MLP(MaxPool(F_{Con'}^{CSAF})))\otimes F_{Con'}^{CSAF} \quad ②$$

其中$MaxPool(\cdot)$为最大池化,$MLP(\cdot)$由2个连接层与1个Rule函数构成。此时通道增强特征已在通道维度上完成增强,但在空间维度上仍有所欠缺,需对空间特征进行进一步的增强。通过拼接通道维的平均、最大化特征图,以生成空间权重矩阵,再与通道

增强特征进行融合得到加权合并特征。最后,通过残差连接将初始合并特征与经过注意力提取融合的加权合并特征进行融合,以保留原始特征分布,缓解梯度消失问题。

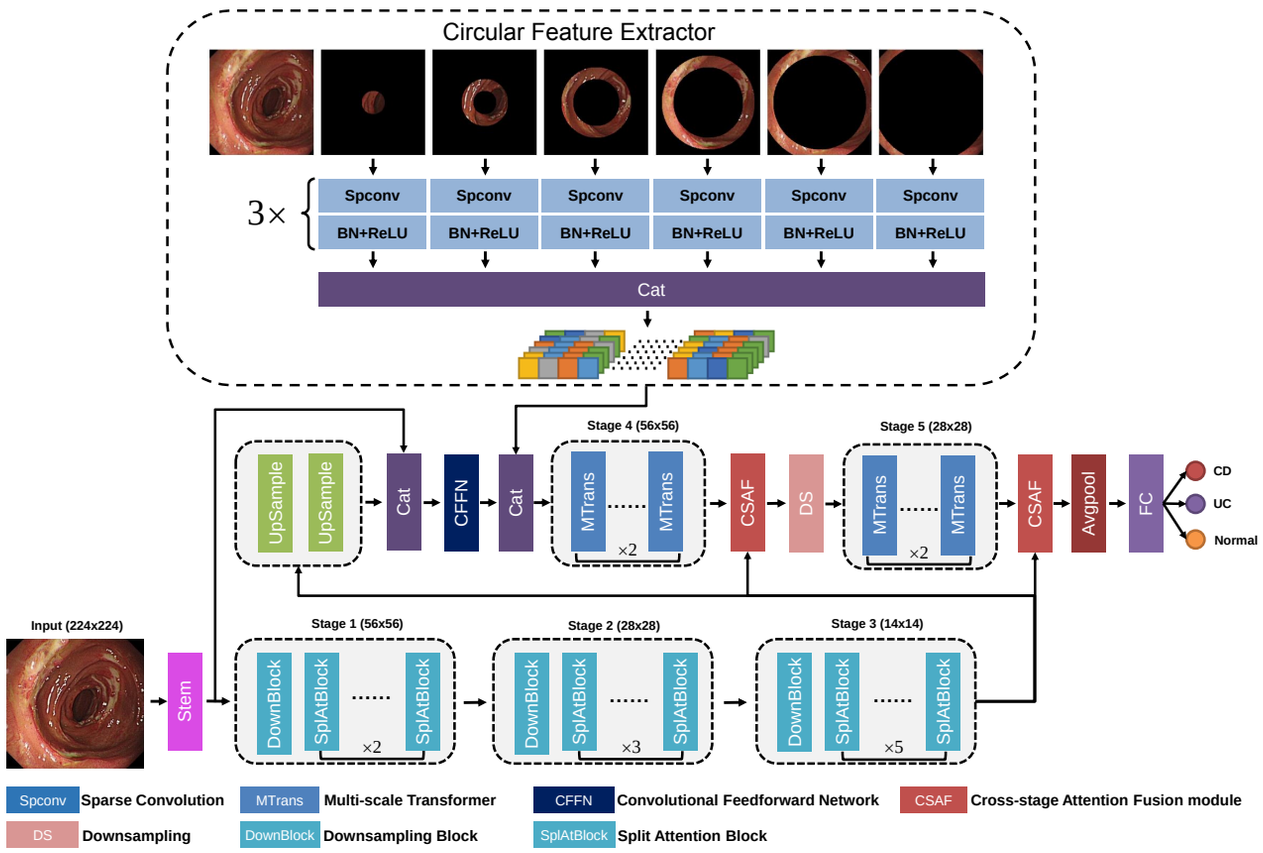



图 2 用于CD、UC及Normal肠镜图像分类的RFFNet整体结构

**Figure 2 Overall architecture of RFFNet for the classification of colonoscopy images of Crohn's disease, ulcerative colitis, and Normal**

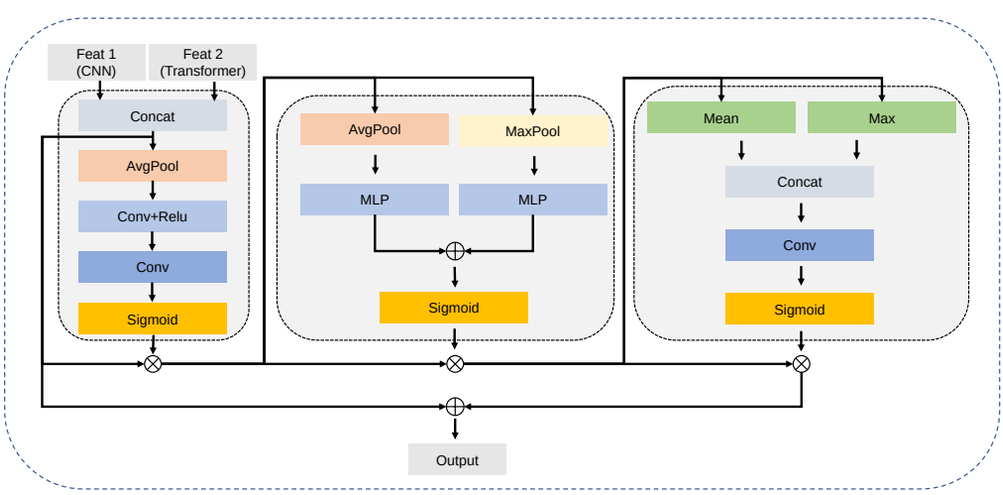



图 3 用于CD、UC和Normal肠镜图像分类的RFFNet中的跨阶段注意力融合模块结构

**Figure 3 Structure of the cross-stage attention fusion module in RFFNet for the classification of colonoscopy images from Crohn's disease, ulcerative colitis, and Normal**

### 1.5 环形特征提取器

环形特征提取器(circular feature extractor,CFE)主要由3个稀疏卷积模块构成,每个稀疏卷积模块包括1个稀疏卷积[24](Spconv)、1个批归一化层和1个ReLU激活函数。由于环形图像中存在大量无效区域,使用传统卷积计算量巨大且效率低下,而稀疏卷积只关注有效数据点,能够显著减少计算量,提高特征提取效率(图2)。该模块通过对图像使用若干个(默认为6个,半径25 px)环形掩码获取相应环形图像,将每个图像通过3个稀疏卷积模块用以提取环形特征,确保提取的特征具备高语义性、空间细节保留能力及计算效率优势。模块第一层计算如式③所示,后续两层计算与第一层一致,输入为上层输出:

$$F_1^{CFE} = ReLU(BN(Spconv(F_{input}^{CFE}))) \quad ③$$

其中$Spconv(\cdot)$为2D稀疏卷积,$BN(\cdot)$为批归一化函数,$ReLU(\cdot)$为ReLU激活函数。环状特征能够建模肠镜景深信息,增强特征空间中的解剖结构感知能力。为使环形特征融入模型中,在经过3层稀疏卷积后,对环形特征进行处理,压缩环形特征的空间维度,只保留通道维度,将每个压缩后的环形特征进行拼接,最终得到的输出特征为所有环形特征拼接后的总特征,通道数与Stage3输出特征一致。

### 1.6 实施细节与评价指标

使用奥林巴斯医疗系统的结肠镜采集肠镜图像,图像采集分辨率为1 920×1 082。采集过程遵循标准化流程:进镜时,内镜直视下从直肠循腔进镜至回盲部,必要时进入回肠末段观察;退镜时,从回盲部、升结肠、横结肠、降结肠、乙状结肠、直肠依次退出,全面观察(尤其皱襞后及转折处),退镜时间不少于6 min;发现可疑病变时,确定具体部位和范围,可综合使用染色内镜、放大内镜、CLE等特殊技术,了解病变大小、范围、浸润深度等,并详细记录;视野受黏液、气泡影响时,用清水或去泡剂及时冲洗。采集图像时,回盲瓣、阑尾隐窝各1张,回盲部、升结肠、肝曲、横结肠、脾曲、降结肠、乙状结肠、直肠各至少1~2张;若发现异常则需额外留图,并保证每张图片清晰度。

为模拟真实世界中肠镜图像的多样性变化,同时增强模型泛化能力,避免模型过拟合,在训练过程中采用随机水平翻转、随机垂直翻转作为数据增强策略,执行概率为0.3,并在增强图像上随机裁剪224×224的区域作为模型输入。在模型测试时,将原始图像缩放至224×224作为输入。训练集作为模型训练的主要数据集,在训练过程中通过内部测试集的损失与准确率选择最优模型。模型初始学习率为0.01,总计经过个300 Epoch迭代,通过余弦退火学习率调整策略在训练时调整每轮学习率。采用交叉熵损失函数与随机梯度下降(stochastic gradient descent,SGD)优化器进行参数优化。SGD优化器中动量(momentum)设置为0.95,L2正则化项(weight_decay)设置为$1\times10^{-3}$。所有模型采用迁移学习,在训练开始前加载在ImageNet数据集上的预训练参数。实验使用Pytorch框架,并在8块NVIDIA GeForce RTX 2080 Ti GPU 11G的显卡上完成所有模型的训练、测试与验证。

### 1.7 统计学分析

计数资料以例数为单位进行记录;基于该类计数资料进一步推导计算得到的总体准确率(accuracy,ACC)、单分类准确率、灵敏度(sensitivity,SEN)、特异度(specificity,SPE)、阳性预测值(positive predictive value,PPV)、阴性预测值(negative predictive value,NPV)和$F_1$值,统一以百分比形式记录。除AUC外各指标计算如式④~⑩所示:

$$\text{总体ACC} = TN_{ALL}/N \quad ④$$

$$ACC = (TP + TN)/N \quad ⑤$$

$$SEN = TP/(TP + FN) \quad ⑥$$

$$SPE = TN/(TN + FP) \quad ⑦$$

$$PPV = TP/(TP + FP) \quad ⑧$$

$$NPV = TN/(TN + FN) \quad ⑨$$

$$F_1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad ⑩$$

其中$TN_{ALL}$为所有类别计算正确的样本数,$N$代表外部验证集所有样本数,$TP$代表真阳性样本数,$FP$代表假阳性样本数,$TN$代表真阴性样本数,$FN$代表假阴性样本数。

所有统计学数据采用Python3.9.19软件进行分析,依托NumPy、SciPy、scikit-learn等核心依赖库开展数值计算、统计检验及基础性能指标与AUC的定量计算,同时借助Bootstrap重采样专用工具包实现置信区间的估计分析。为客观评估各性能指标估计值的统计稳定性与推断可靠性,采用Bootstrap重采样法对外部验证集数据进行1 000次重复抽样,并据此计算得到ACC、单分类准确率、SEN、SPE、PPV、NPV、$F_1$值的95%置信区间(95%*CI*),$P<0.05$为差异具有统计学意义。

## 2 结果

### 2.1 RFFNet各性能较对照模型显著升高

采用Bootstrap重采样法(1 000次重复抽样)计算

95%*CI*,在外部验证集上对RFFNet与现有深度学习模型的ACC、单分类准确率、SEN、SPE及$F_1$值等核心性能指标进行定量比较。RFFNet总体准确率达到95.68%(95%*CI*:94.14%~98.58%),95%*CI*区间宽度为4.44%,显著高于ResNet50($P$<0.01)、ViT($P$<0.001)、Swin-T($P$<0.001)等对照模型(表3);其CD、UC和Normal类别的单分类准确率分别达到94.31%(95%*CI*:93.62%~94.94%)、94.14%(95%*CI*:93.44%~94.78%)和98.58%(95%*CI*:98.21%~98.88%),3个类别的单分类准确率均高于所有对照模型($P$均<0.05,表3);其参数量为48.28×$10^6$,浮点运算性能为22.300 GFLOP/s,单张图像推理耗时8.24 ms(表3),结合分类准确率,初步验证了RFFNet的有效性与实时性。

进一步分析显示,RFFNet在CD、UC和Normal 3类别的灵敏度分别为93.35%(95%*CI*:91.67%~94.71%)、94.00%(95%*CI*:92.60%~95.16%)和98.56%(95%*CI*:97.99%~98.97%);特异度分别为94.58%(95%*CI*:93.81%~95.27%)、94.20%(95%*CI*:93.36%~94.94%)和98.61%(95%*CI*:98.05%~99.01%);$F_1$值分别为87.81%(95%*CI*:86.28%~89.25%)、90.05%(95%*CI*:89.00%~91.25%)和98.58%(95%*CI*:98.22%~98.92%),且RFFNet在3个类别上的$F_1$值均显著高于ResNet50、ViT、Swin-T等对照模型($P$均<0.05,表3)。提示RFFNet在外部验证集上的综合分类性能显著优于现有深度学习模型,在不同疾病亚型及正常对照组中均展现出更优的识别效能,验证了所提出方法的有效性和可靠性。

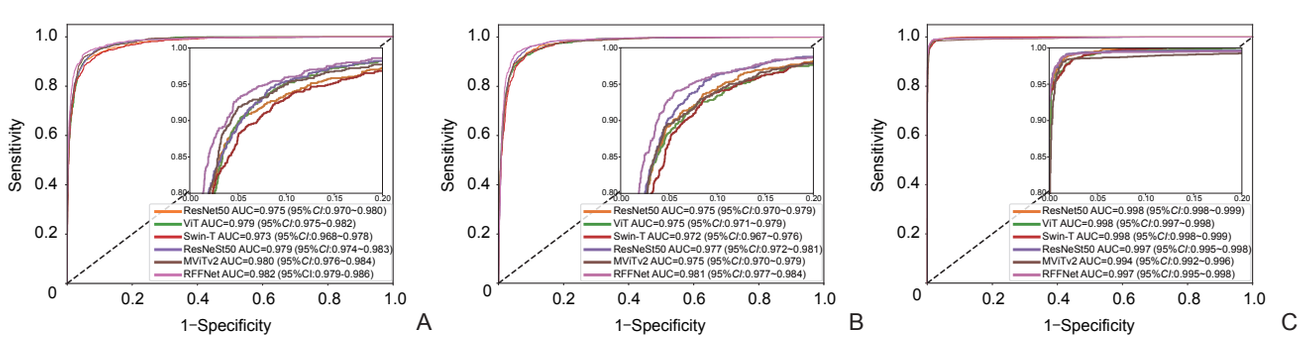
通过绘制ROC曲线直观展示RFFNet与现有模型在各分类类别上的性能对比,并采用Bootstrap重采样法计算95%*CI*,对总体及CD、UC、Normal类别的AUC进行定量分析。结果显示,RFFNet总体AUC为0.987(95%*CI*:0.985~0.990);在CD病变类别上,RFFNet的AUC达到0.982(95%*CI*:0.979~0.986),高于其他对照模型;在UC病变类别上,其AUC为0.981(95%*CI*:0.977~0.984),同样高于其他对照模型;在Normal类别上,各对比模型的AUC差距较小,RFFNet的AUC达到0.997(95%*CI*:0.995~0.998)。提示RFFNet在炎症性肠病相关分类任务中具有优异的整体判别效能,尤其在CD、UC病变类别识别上的效能显著优于现有模型,见图4。

为了解各模型分类结果情况,采用混淆矩阵对各模型的分类性能指标进行测量。所有模型对Normal类别展现出较高的分类准确率,能够有效且稳定地识别正常样本;然而,在CD与UC类别上,各模型普遍存在显著的类别混淆,两类别间出现了明显的相互误判(图5)。各模型在Normal类别识别上具有优势,但在CD与UC类别区分上仍存在不足,需进一步优化模型对这两类别的识别能力。

为探索模型关注区域,采用Grad-CAM方法对模型决策依据相关指标进行测量。与其他模型相比,RFFNet更适应肠镜图像的几何结构,能够更有效地获取肠壁信息,进而提升模型分类性能(图6)。进一步说明RFFNet在肠镜图像特征提取方面具有独特优势,其对肠壁信息的高效获取能力是提升分类性能的重要支撑。

### 2.2 RFFNet消融实验验证各核心模块作用有效

为确定CFE模块的最优参数,采用参数变量法,通过评估不同环形特征数对模型整体准确率的影响,以确定CFE模块的最优参数。具体而言,所评估的环形特征数及对应环半径分别为:3个环(半径40 px)、4个环(半径35 px)、5个环(半径30 px)、6个环(半径25 px)、7个环(半径20 px)、8个环(半径15 px)、9个环




A: ROC curve for CD category; B: ROC curve for UC category; C: ROC curve for Normal category

**图4 各模型ROC曲线对比**

**Figure 4 Comparison of ROC curves of different models**

表3 RFFNet与现有深度学习模型对比[%,(95%*CI*)]

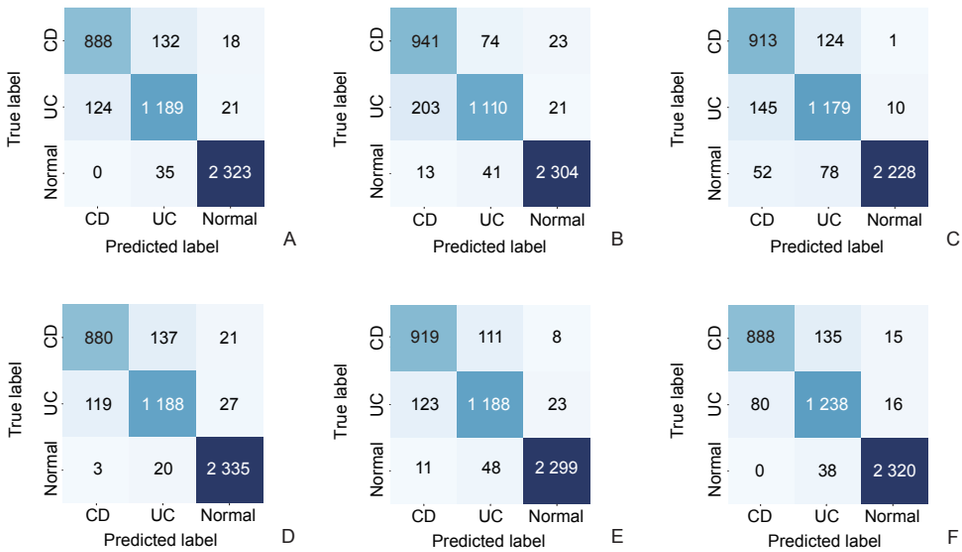
**Table 3 Comparison of RFFNet with state-of-the-art deep learning models [%, (95%*CI*)]**

| Disease subtype and evaluation metric | ResNet50 | ViT | Swin-T | ResNeSt50 | MViTv2 | RFFNet |
|---|---|---|---|---|---|---|
| Total ACC | 94.67$^{b}$(92.05~98.46) | 94.25$^{a}$(92.09~97.95) | 93.67$^{a}$(91.40~97.89) | 94.40$^{b}$(92.18~98.58) | 94.83$^{c}$(92.05~98.25) | 95.68(94.14~98.58) |
| CD | | | | | | |
| ACC | 93.51(92.77~94.18) | 92.71$^{b}$(91.93~93.41) | 91.40$^{b}$(90.56~92.16) | 92.18$^{b}$(91.38~92.91) | 94.19(93.48~94.82) | 94.31(93.62~94.94) |
| SEN | 90.75(88.84~92.37) | 93.06$^{b}$(91.35~94.46) | 92.39$^{c}$(90.62~93.85) | 94.51(92.95~95.74) | 91.91$^{b}$(90.09~93.42) | 93.35(91.67~94.71) |
| SPE | 94.28$^{b}$(93.49~94.99) | 92.61$^{b}$(91.72~93.41) | 91.12$^{b}$(90.15~91.99) | 91.52$^{b}$(90.58~92.38) | 94.83$^{b}$(94.06~95.50) | 94.58(93.81~95.27) |
| PPV | 81.70$^{b}$(79.36~83.83) | 77.97$^{b}$(75.57~80.19) | 74.51$^{b}$(72.06~76.82) | 75.81$^{b}$(73.40~78.07) | 83.32$^{b}$(81.05~85.37) | 82.89(80.63~84.94) |
| NPV | 97.32(96.73~97.80) | 97.94$^{b}$(97.41~98.36) | 97.71$^{c}$(97.15~98.16) | 98.34(97.86~98.72) | 97.66$^{b}$(97.11~98.10) | 98.06(97.56~98.47) |
| $F_1$-score | 85.99$^{b}$(84.38~87.57) | 84.85$^{b}$(83.29~86.45) | 82.49$^{b}$(80.93~84.08) | 84.13$^{b}$(82.44~85.71) | 87.40(85.94~89.01) | 87.81(86.28~89.25) |
| AUC | 97.50$^{a}$(97.03~97.96) | 97.86$^{a}$(97.53~98.23) | 97.29$^{a}$(96.81~97.75) | 97.85$^{a}$(97.39~98.27) | 97.98$^{a}$(97.61~98.35) | 98.24(97.86~98.61) |
| UC | | | | | | |
| ACC | 92.05$^{b}$(91.25~92.79) | 92.09$^{b}$(91.29~92.83) | 91.73$^{b}$(90.91~92.48) | 92.43$^{c}$(91.64~93.15) | 92.05$^{c}$(91.25~92.79) | 94.14(93.44~94.78) |
| SEN | 93.55$^{b}$(92.11~94.75) | 92.43$^{b}$(90.88~93.73) | 93.25$^{b}$(91.78~94.48) | 95.35$^{b}$(94.09~96.36) | 92.95$^{b}$(91.45~94.21) | 94.00(92.60~95.16) |
| SPE | 91.46(90.47~92.35) | 91.96$^{b}$(91.00~92.83) | 91.14$^{c}$(90.13~92.05) | 91.28(90.29~92.19) | 91.70(90.72~92.58) | 94.20(93.36~94.94) |
| PPV | 81.14(79.11~83.02) | 81.87$^{b}$(79.85~83.74) | 80.52$^{b}$(78.47~82.42) | 81.12(79.11~82.98) | 81.47(79.44~83.34) | 86.42(84.56~88.09) |
| NPV | 97.31$^{b}$(96.68~97.81) | 96.87$^{b}$(96.21~97.42) | 97.17$^{b}$(96.54~97.70) | 98.04$^{b}$(97.49~98.47) | 97.07$^{b}$(96.43~97.60) | 97.56(96.97~98.04) |
| $F_1$-score | 86.90$^{b}$(85.61~88.22) | 86.83$^{b}$(85.49~88.06) | 86.42$^{b}$(85.14~87.72) | 87.66$^{b}$(86.46~88.94) | 86.83$^{b}$(85.52~88.13) | 90.05(89.00~91.25) |
| AUC | 97.50$^{a}$(97.07~97.92) | 97.52$^{a}$(97.09~97.92) | 97.17$^{a}$(96.72~97.61) | 97.67$^{a}$(97.26~98.09) | 97.45$^{a}$(97.02~97.86) | 98.06(97.71~98.40) |
| Normal | | | | | | |
| ACC | 98.46(98.06~98.77) | 97.95$^{b}$(97.50~98.32) | 97.89$^{b}$(97.44~98.26) | 98.58(98.21~98.88) | 98.25$^{c}$(97.83~98.58) | 98.58(98.21~98.88) |
| SEN | 98.56(97.99~98.97) | 98.39$^{c}$(97.80~98.82) | 98.30$^{b}$(97.70~98.75) | 98.73$^{b}$(98.19~99.11) | 97.96$^{b}$(97.31~98.46) | 98.56(97.99~98.97) |
| SPE | 98.36(97.76~98.79) | 97.51$^{c}$(96.80~98.07) | 97.47$^{b}$(96.76~98.03) | 98.44$^{b}$(97.86~98.87) | 98.52(97.95~98.94) | 98.61(98.05~99.01) |
| PPV | 98.35(97.75~98.79) | 97.52$^{c}$(96.81~98.07) | 97.48$^{b}$(96.77~98.03) | 98.44$^{b}$(97.85~98.86) | 98.51(97.93~98.92) | 98.60(98.04~99.00) |
| NPV | 98.56(98.00~98.97) | 98.38$^{b}$(97.79~98.82) | 98.30$^{b}$(97.69~98.75) | 98.73$^{b}$(98.19~99.11) | 97.99$^{b}$(97.34~98.48) | 98.57(98.00~98.97) |
| $F_1$-score | 98.45(98.05~98.80) | 97.95$^{b}$(97.50~98.36) | 97.89$^{b}$(97.45~98.28) | 98.58(98.21~98.92) | 98.23$^{c}$(97.85~98.59) | 98.58(98.22~98.92) |
| AUC | 99.86$^{a}$(99.80~99.91) | 99.77$^{a}$(99.69~99.84) | 99.83$^{a}$(99.78~99.88) | 99.65(99.48~99.80) | 99.42$^{a}$(99.23~99.59) | 99.67(99.53~99.80) |
| Number of parameters/($\times10^6$) | 23.51 | 86.39 | 86.75 | 25.44 | 50.71 | 48.28 |
| Floating point operations/(GFLOP/s) | 4.132 | 16.863 | 15.169 | 5.427 | 8.902 | 22.300 |
| Inference time/ms | 1.09 | 0.29 | 0.66 | 0.51 | 4.12 | 8.24 |

a: $P<0.001$, b: $P<0.01$, c: $P<0.05$, compared with RFFNet

(半径10 px)。在6个环时模型总体准确率达到95.68%(95%*CI*: 94.14%~98.58%),此时模型参数达到最优效果(图7)。该参数评估结果可为CFE模块最优环形特征数的确定提供直接依据,进而保障模型整体性能的最优化。

采用消融实验法测量RFFNet中CNN-Transformer架构、CSAF模块及CFE模块对模型总体准确率、各类别单分类准确率及F1值的影响,以验证各模块的有效性。表中基线模型为RFFNet的CNN部分,即从图像输入到Stage3。基线模型本身与原模型ResNeSt50相比在总体ACC上提升了0.59%[基线模型总体ACC为94.99%(95%*CI*: 93.15%~98.37%)],见表4;与基线模型相比,仅引入CNN-Transformer架构时,总体ACC提升了0.11%[总体ACC为95.10%(95%*CI*: 93.07%~98.41%)]。在CNN-Transformer架构基础上分别引入单一模块:仅使用CSAF模块时,总体ACC提升0.31%[总体ACC为95.41%(95%*CI*: 93.81%~98.58%)];仅使用CFE模块时,Normal类别的单分类准确率[98.79%




A: Confusion matrix of ResNet50; B: Confusion matrix of ViT; C: Confusion matrix of Swin-T; D: Confusion matrix of ResNeSt50; E: Confusion matrix of MViTv2; F: Confusion matrix of RFFNet

**图5 各模型混淆矩阵对比**

**Figure 5 Comparison of confusion matrices of different models**

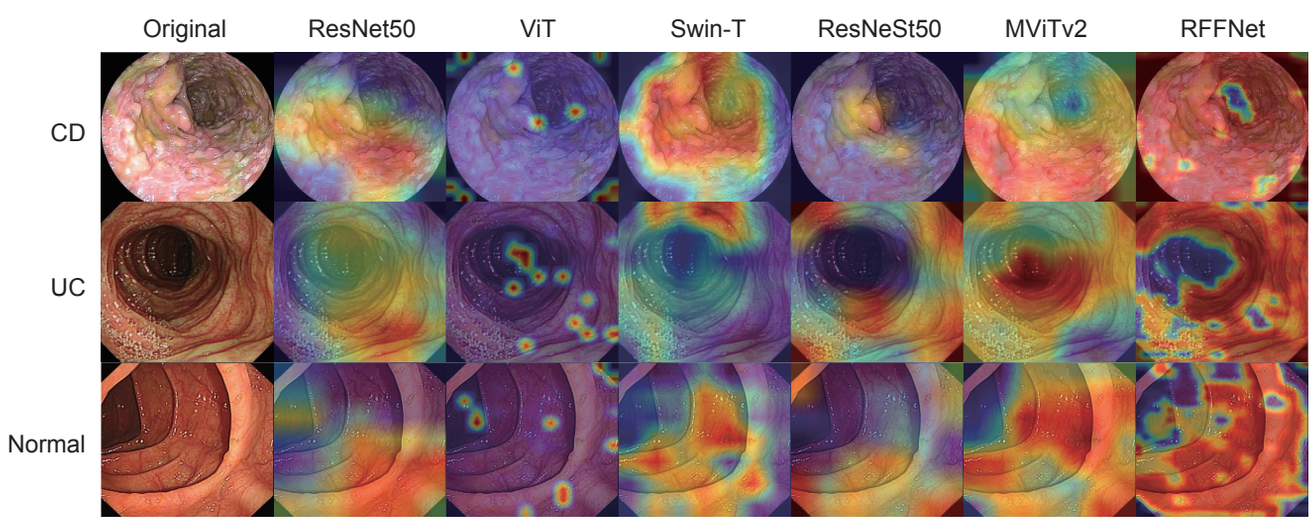



**图6 各模型类激活映射图对比**

**Figure 6 Comparison of class activation maps of different models**

(95%*CI*: 98.44%~99.07%)]与$F_1$值[98.79%(95%*CI*: 98.45%~99.10%)]达到最优,同时总体ACC提升0.57%[总体ACC为95.67%(95%*CI*: 94.02%~98.79%)],见表4。

综合引入CNN-Transformer架构、CSAF模块和CFE模块的完整RFFNet模型,在CD和UC的单分类准确率与$F_1$值上均取得最优结果:与基线模型相比,CD、UC和Normal三类别单类准确率分别提升了0.86%(CD ACC为94.31%, 95%*CI*: 93.62%~94.94%)、0.99%(UC ACC为94.14%, 95%*CI*: 93.44%~94.78%)和0.21%(Normal ACC为98.58%, 95%*CI*: 98.21%~98.88%),总体ACC提升0.69%[总体ACC为95.68%(95%*CI*: 94.14%~98.58%)],见表4;与仅使用CSAF模块相比,CD和UC的单类准确率分别提高了0.50%和0.29%;与仅使用CFE模块相比,CD和UC的单类准确率分别提高了0.10%和0.12%。

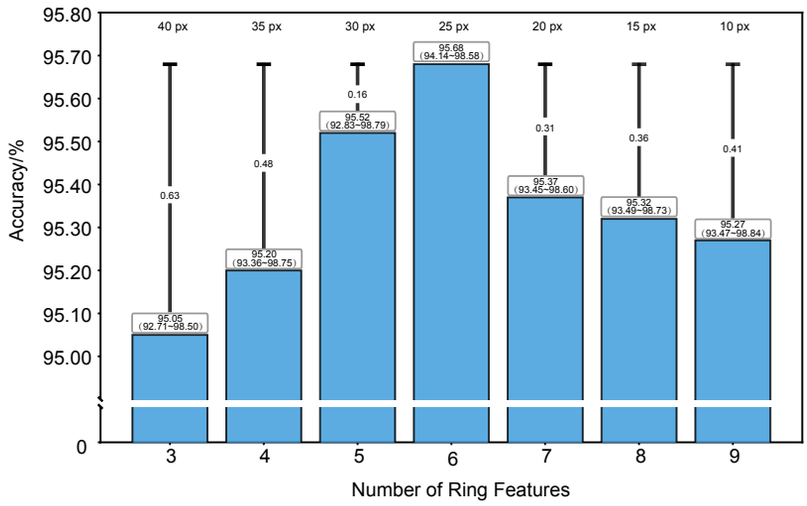



**图7 环形特征数对比**

**Figure 7 Comparison of ring feature numbers**

**表4 消融实验对比结果[%(95%*CI*)]**

**Table 4 Comparison of ablation experiment results[%(95%*CI*)]**

| Model | Total ACC | Disease subtype | ACC | $F_1$-score |
|---|---|---|---|---|
| CNN | 94.99(93.15~98.37) | CD | 93.45(92.70~94.12) | 86.23(84.67~87.77) |
| | | UC | 93.15(92.39~93.84) | 88.50(87.24~89.79) |
| | | Normal | 98.37(97.97~98.70) | 98.36(97.97~98.71) |
| CNN-Transformer | 95.10(93.07~98.41) | CD | 93.07(92.31~93.76) | 85.69(84.00~87.18) |
| | | UC | 93.83(93.10~94.48) | 89.40(88.21~90.68) |
| | | Normal | 98.41(98.02~98.73) | 98.41(98.04~98.76) |
| CNN-Transformer+CSAF | 95.41(93.81~98.58) | CD | 93.81(93.08~94.46) | 87.05(85.44~88.59) |
| | | UC | 93.85(93.13~94.50) | 89.50(88.30~90.67) |
| | | Normal | 98.58(98.21~98.88) | 98.57(98.21~98.91) |
| CNN-Transformer+CFE | 95.67(94.02~98.79) | CD | 94.21(93.50~94.84) | 87.40(85.94~88.84) |
| | | UC | 94.02(93.30~94.66) | 89.83(88.66~90.97) |
| | | Normal | 98.79(98.44~99.07) | 98.79(98.45~99.10) |
| CNN-Transformer+CSAF+CFE | 95.68(94.14~98.58) | CD | 94.31(93.62~94.94) | 87.81(86.28~89.25) |
| | | UC | 94.14(93.44~94.78) | 90.05(89.00~91.25) |
| | | Normal | 98.58(98.21~98.88) | 98.58(98.22~98.92) |

## 3 讨论

### 3.1 深度学习辅助IBD诊断的进展与现存瓶颈

近年,人工智能技术发展迅速,并在多个领域实现了广泛应用。但在医学领域,受数据特殊性、标注难度大以及自然图像训练的深度学习模型难以直接迁移至医学场景等因素限制,模型通常需依赖针对性医学数据集进行迁移学习,因此,AI在医学图像分析中仍面临诸多挑战。对此,AI不再局限于图像数据,通过融合视觉、文本等多源信息,在医学影像分析中展现出显著优势。例如,MedSAM能够有效实现多器官及病变的分割,辅助医生快速标注[25];MedCLIP解决了传统对比学习在医学领域中易将不同患者相同病症被误判为"假阴性"的问题[26];UniVid将理解与生成能力整合于单一模型,为医学影像分析的深层次特征提取与上下文推理提供了借鉴[27]。然而,病变形态复杂多样,单一模型难以准确识别全部病变类型。因此,针对特定病变识别,当前最优策略仍聚焦于提取关键特征或设计适配相似解剖部位的专用模型,以增强识别能力。

本研究构建的端到端分类模型RFFNet,基于ResNeSt50与MViTv2双架构实现特征自适应融合,适

配肠镜图像畸变特性。首先,提出跨阶段注意力模块,通过跨阶段特征融合与注意力加权机制捕获全局上下文与局部细节的互补关系,结合动态空间注意力调控特征贡献度,提升融合效率并加速收敛。其次,引入环形特征适配肠腔几何结构,直接缓解桶形畸变与景深现象的干扰,突破广角镜头成像的物理限制。最后,为规避双架构结合额外卷积带来的计算负担,采用稀疏卷积提取环形特征,聚焦有效区域以减少冗余计算,在保障特征质量的同时大幅降低额外模型参数量。本研究通过利用双架构优势与肠镜特有结构,为提升CD与UC诊断性能提供了新思路。

IBD主要包括CD和UC 2个亚型,患者多表现为消化道系统紊乱及相似的临床症状,如腹痛、腹泻、直肠出血与体质量减轻[28]。鉴于CD和UC的病变主要位于肠壁,且肠镜图像还存在桶形畸变与景深现象等成像特性,深度学习模型提取病灶特征时易受干扰,给CD与UC的精确识别带来挑战。现有研究虽证实了深度学习在IBD诊断任务中兼具高精度与高效性,但仍存在明显不足:多数研究依赖单一架构[12-13, 29],如CHIERICI等[29]通过集成学习堆叠多个ResNet模型作为元模型,但其单一架构本质限制了模型捕获多样化特征的能力;部分研究尝试结合多架构优势提升性能[14-15],如MAURÍCIO等[15]使用Transformer与CNN架构提取肠镜图像特征,但因引入额外卷积层导致参数量激增,需依赖知识蒸馏等方法提高推理速度。此外,现有研究普遍忽视肠镜图像环形结构相似性这一关键先验知识,未能从源头优化特征提取过程。

与现有研究相比,本研究提出的跨阶段注意力模块能够捕获全局上下文与局部细节的互补关系,动态调控2类特征的贡献度,提升特征融合效率,有效弥合CNN与Transformer架构间的特征鸿沟,解决了单一架构模型难以兼顾细粒度病灶识别与整体结构分析的难题。其次,引入环形特征将肠镜特有的环形先验知识嵌入特征提取源头,助力模型学习病灶分布规律,提高病灶特征提取能力。最后,采用稀疏卷积提取环形特征,从根本上减少额外参数量,规避双架构结合额外卷积带来的计算负担,保障模型高效运行,显著提升其临床实用性。本研究通过深度耦合双架构特性与肠镜成像规律,不仅实现特征提取端到端优化,还为CD与UC鉴别诊断构建了更符合医学影像特性的计算框架。

### 3.2 多模块协同赋能RFFNet实现性能最优

本研究结果表明,所有对比模型对正常类别的识别准确率均高于CD与UC类别,提示CD与UC间易出现类别混淆,Normal类别相较于二者更易区分,该现象在图5的混淆矩阵中直观体现。对比模型对CD与UC的相互误判概率显著更高,究其原因,二者临床症状高度相似,溃疡、假性息肉、肠壁炎症等病变特征易使模型提取的特征存在较高相似性,进而引发二者相互误判。模型对比中,ResNeSt50对CD、UC及正常类别的识别灵敏度较高,但三类别的阳性预测值均偏低,提示其对病变敏感但易发生误判。MViTv2在CD类别中特异度与阳性预测值表现突出,对CD的检测可靠性较好,却在UC和正常类别上的性能弱于其他对比模型。RFFNet在所有对比模型中,各类别的ACC与$F_1$值均居首位,且与其他模型存在显著性差异;ROC曲线结果进一步显示,RFFNet在CD与UC类别上的AUC显著高于其他模型,证实其可有效区分这两种易混淆的IBD亚型。在模型参数量方面,RFFNet虽高于ResNeSt50,但仍低于MViTv2;在计算复杂度与推理速度上,RFFNet的FLOPs为对比模型中最高,推理速度亦相对较慢,但其单张图像推理耗时仅8.24 ms,可满足实时性应用需求。

消融实验结果证实,CNN-Transformer架构、特征融合策略与环形特征提取模块均能提升模型分类性能,三者协同作用使RFFNet在特征提取过程中能更精准聚焦病变区域。RFFNet对正常类别的识别性能未达消融实验最优水平,相较而言对CD、UC的识别表现更优,推测系各模块间的相互作用所致——模型被引导更侧重捕获病变特征,一定程度降低了对正常类别特征的关注度,但正常类别仍保持了较高水平的分类准确率,体现出RFFNet在病变识别与正常组织区分上的双重有效性。深度学习模型因多层非线性变换的复杂结构常被视为“黑箱”,其决策逻辑难以直观解释,在医疗领域更易引发信任危机;而Grad-CAM可通过梯度加权类激活映射生成热力图,直观呈现模型预测时的关注区域,有效揭示模型的内部决策依据,提升其可解释性。Grad-CAM类激活映射图的对比结果显示,RFFNet的高响应区域更集中于肠腔溃疡、出血、假性息肉、鹅卵石样改变等关键病变区域,与内镜诊断的核心特征一致[28],提示RFFNet更适配肠镜图像结构,可高效捕获病变区域的关键信息。

本研究仍存在一定的局限性。研究采用2个中心的回顾性数据,但肠镜图像质量受设备、光照、医师操作等多种因素影响,可能制约模型的准确性与稳定性,其泛化能力仍需进一步多中心验证,未来可纳入

多中心的肠镜数据来提升模型的鲁棒性。其次,本研究仅基于静态图像分析,而临床诊断需结合病变动态观察与位置信息综合判断,模型临床部署时仍需医师结合专业知识决策。

综上所述,本研究提出融合环形特征的CNN-Transformer网络RFFNet,有效整合不同架构的特征提取优势,通过环形特征针对性处理肠镜图像的桶形畸变与景深效应,显著提升特征表征能力。RFFNet在CD、UC及正常类别的识别中,均表现出较高的准确率、灵敏度、特异度、$F_1$值与AUC,总体性能优于现有经典模型。作为辅助诊断工具,RFFNet可帮助经验不足的医师快速捕捉病变核心特征,辅助完成鉴别诊断,减轻工作负担的同时提升诊断准确性与效率。

**利益冲突** 本文所有作者声明不存在利益冲突

## 参考文献:


[1] 薛文俊, 杨嫚, 蒋秋焕. 炎症性肠病患者症状群管理的研究进展[J]. 中国护理管理, 2023, 23(2): 298-303. DOI: 10.3969/j.issn.1672-1756.2023.02.028.
XUE W J, YANG M, JIANG Q H. Research progress on symptom cluster management in patients with Inflammatory Bowel Disease[J]. Chin Nurs Manag, 2023, 23(2): 298-303. DOI:10.3969/j.issn.1672-1756.2023.02.028.

[2] ZHANG Y M, CHU X T, WANG L, et al. Global patterns in the epidemiology, cancer risk, and surgical implications of inflammatory bowel disease[J]. Gastroenterol Rep, 2024, 12: goae053. DOI:10.1093/gastro/goae053.

[3] LI L L, CHENG R, WU Y S, *et al*. Diagnosis and management of inflammatory bowel disease[J]. J Evid Based Med, 2024, 17(2): 409-433. DOI:10.1111/jebm.12626.

[4] GORDON H, BIANCONE L, FIORINO G, *et al*. ECCO guidelines on inflammatory bowel disease and malignancies[J]. J Crohns Colitis, 2023, 17(6): 827-854. DOI:10.1093/ecco-jcc/jjac187.

[5] HRACS L, WINDSOR J W, GOROSPE J, *et al*. Global evolution of inflammatory bowel disease across epidemiologic stages[J]. Nature, 2025, 642(8067): 458-466. DOI: 10.1038/s41586-025-08940-0.

[6] HUANG T, ZHAN S Q, CHEN P J, *et al*. Accurate diagnosis of endoscopic mucosal healing in ulcerative colitis using deep learning and machine learning[J]. J Chin Med Assoc, 2021, 84(7): 678-681. DOI:10.1097/JCMA.0000000000000559.

[7] LUO X D, ZHANG J H, LI Z G, *et al*. Diagnosis of ulcerative colitis from endoscopic images based on deep learning[J]. Biomed Signal Process Control, 2022, 73: 103443. DOI:10.1016/j.bspc.2021.103443.

[8] AGRAWAL M, ALLIN K H, PETRALIA F, *et al*. Multiomics to elucidate inflammatory bowel disease risk factors and pathways[J]. Nat Rev Gastroenterol Hepatol, 2022, 19(6): 399-409. DOI:10.1038/s41575-022-00593-y.

[9] KIM K O. Endoscopic activity in inflammatory bowel disease: clinical significance and application in practice[J]. Clin Endosc, 2022, 55(4): 480-488. DOI:10.5946/ce.2022.108.

[10] BUDZINSKAYA A A, BELOUSOVA E A, ORLOVA L P, *et al*. Differential diagnosis of inflammatory bowel diseases by endoscopic ultrasound[J]. Alm Clin Med, 2023, 50(6): 400-407. DOI:10.18786/2072-0505-2022-50-055.

[11] FERNANDEZ C, GAJIC Z, ESEN E, *et al*. Preoperative risk factors for adverse events in adults undergoing bowel resection for inflammatory bowel disease: 15-year assessment of the American college of surgeons national surgical quality improvement program[J]. Am J Gastroenterol, 2023, 118(12): 2230-2241. DOI:10.14309/ajg.0000000000002395.

[12] TONG Y R, LU K M, YANG Y Y, *et al*. Can natural language processing help differentiate inflammatory intestinal diseases in China? Models applying random forest and convolutional neural network approaches[J]. BMC Med Inform Decis Mak, 2020, 20(1): 248. DOI: 10.1186/s12911-020-01277-w.

[13] 戚婧, 阮广聪, 杨毅, 等. 空间双线性注意力网络识别溃疡性结肠炎与克罗恩病[J]. 陆军军医大学学报, 2023, 45(3): 227-234, 242. DOI: 10.16016/j. 2097-0927. 202209187.
QI J, RUAN G C, YANG Y, *et al*. Identification of ulcerative colitis and Crohn's disease based on spatial and bilinear attention network[J]. J Army Med Univ, 2023, 45(3): 227-234, 242. DOI:10.16016/j.2097-0927.202209187.

[14] UDRISTOIU A L, STEFANESCU D, GRUIONU G, *et al*. Deep learning algorithm for the confirmation of mucosal healing in Crohn's disease, based on confocal laser endomicroscopy images[J]. J Gastrointestin Liver Dis, 2021, 30(1): 59-65. DOI:10.15403/jgld-3212.

[15] MAURÍCIO J, DOMINGUES I. Distinguishing between Crohn's disease and ulcerative colitis using deep learning models with interpretability[J]. Pattern Anal Appl, 2024, 27(1): 1. DOI:10.1007/s10044-023-01206-3.

[16] 孙慧贤, 罗飞路, 张玉华. 内窥镜图像非线性畸变数字校正方法[J]. 无损检测, 2009, 31(2): 92-95, 149.
SUN H X, LUO F L, ZHANG Y H. Nonlinear distortion correction of endoscope images with digital technology[J]. Nondestruct Test Technol, 2009, 31(2): 92-95, 149.

[17] ZHANG H, WU C R, ZHANG Z Y, *et al*. ResNeSt: split-attention networks[C]//2022 IEEE/CVF Conference on

Computer Vision and Pattern Recognition Workshops (CVPRW). June 19-20, 2022, New Orleans, LA, USA. IEEE, 2022: 2735-2745. DOI: 10.1109/CVPRW56347.2022.00309.

[18] LI Y H, WU C Y, FAN H Q, *et al*. MViTv2: improved multiscale vision transformers for classification and detection [C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans: IEEE, 2022: 4794-4804. DOI:10.1109/CVPR52688.2022.00476.

[19] 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 (CVPR). Las Vegas: IEEE, 2016: 770-778. DOI:10.1109/CVPR.2016.90.

[20] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, *et al*. An image is worth 16x16 words: transformers for image recognition at scale [EB/OL]. 2020: arXiv: 2010.11929. https://arxiv.org/abs/2010.11929.

[21] LIU Z, LIN Y T, CAO Y, *et al*. Swin transformer: hierarchical vision transformer using shifted windows [C]// 2021 IEEE/CVF International Conference on Computer Vision (ICCV). October 10-17, 2021, Montreal, QC, Canada. IEEE, 2022: 9992-10002. DOI: 10.1109/ICCV48922.2021.00986.

[22] SELVARAJU R R, COGSWELL M, DAS A, *et al*. Grad-CAM: visual explanations from deep networks *via* gradient-based localization [J]. Int J Comput Vis, 2020, 128(2): 336-359. DOI:10.1007/s11263-019-01228-7.

[23] YAO T, LI Y H, PAN Y W, *et al*. HIRI-ViT: scaling vision transformer with high resolution inputs [J]. IEEE Trans Pattern Anal Mach Intell, 2024, 46(9): 6431-6442. DOI: 10.1109/TPAMI.2024.3379457.

[24] GRAHAM B, ENGELCKE M, VAN DER MAATEN L. 3D semantic segmentation with submanifold sparse convolutional networks [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 9224-9232. DOI:10.1109/CVPR.2018.00961.

[25] MA J, HE Y T, LI F F, *et al*. Segment anything in medical images [J]. Nat Commun, 2024, 15: 654. DOI: 10.1038/s41467-024-44824-z.

[26] WANG Z F, WU Z B, AGARWAL D, *et al*. MedCLIP: contrastive learning from unpaired medical images and text [C]//Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing. Abu Dhabi: United Arab Emirates Association for Computational Linguistics, 2022: 3876-3887.

[27] LUO J B, LIN J H, ZHANG Z Y, *et al*. UniVid: the open-source unified video model [EB/OL]. 2025: arXiv: 2509.24200. https://arxiv.org/abs/2509.24200.

[28] MORAN G W, GORDON M, SINOPOLOU V, *et al*. British Society of Gastroenterology guidelines on inflammatory bowel disease in adults: 2025 [J]. Gut, 2025, 74(Suppl 2): s1-s101. DOI:10.1136/gutjnl-2024-334395.

[29] CHIERICI M, PUICA N, POZZI M, *et al*. Automatically detecting Crohn's disease and Ulcerative Colitis from endoscopic imaging [J]. BMC Med Inform Decis Mak, 2022, 22(6): 300. DOI:10.1186/s12911-022-02043-w.

(**收稿**:2025-10-23;**修回**:2025-12-09)

(编辑 栾 嘉)