DOI:10.16016/j.2097-0927.202209187

空间双线性注意力网络识别溃疡性结肠炎与克罗恩病

戚 婧[1],阮广聪[2],杨 毅[1],吴 毅[1],曹 倩[3],魏艳玲[2],粘永健[1] 400038 重庆,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[1];400042 重庆,陆军特色医学中心消化内科[2];310016 杭州,浙江大学医学院附属邵逸夫医院消化内科[3]

[摘要] **目的** 利用深度学习技术辅助内镜医师识别溃疡性结肠炎(ulcerative colitis, UC)与克罗恩病(Crohn's disease, CD)。**方法** 收集2018年1月至2020年11月陆军特色医学中心消化内科与邵逸夫医院消化内科共1 576例受试者的内镜图像,包括CD、UC与正常三类共计34 300幅,并按照9∶1的比例随机划分训练集与测试集,用于对网络进行训练与测试。在ResNet50基础上构建新颖的空间双线性深度网络(SABA-ResNet),引入空间注意机制,通过膨胀卷积扩大感受野以联系上下文信息,并与普通卷积局部归纳特性相配合,自适应聚焦病变区域。利用双线性注意提高网络的特征表示能力,以二阶信息加权特征映射的通道信息,从而提高模型的分类性能。**结果** SABA-ResNet在测试集上对CD、UC和正常识别的总体准确率为92.67%(95%*CI*:91.91~93.37),AUC分别为0.978(95%*CI*:0.972~0.983)、0.977(95%*CI*:0.971~0.982)和0.999(95%*CI*:0.998~1.000),灵敏度分别为88.40%、89.07%、98.89%,特异性分别为95.49%、94.88%、98.93%,F1值分别为88.80%、89.01%和98.60%。消融实验与类激活映射图表明空间注意与双线性注意可帮助模型捕获更多病变区域的特征。**结论** 所构建的网络将空间注意与双线性注意相结合,在CD、UC与正常的识别中取得了良好的性能,可以有效辅助内镜医师对UC与CD进行识别。

[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; R447; R574.62 [文献标志码] A

Identification of ulcerative colitis and Crohn's disease based on spatial and bilinear attention network

QI Jing[1], RUAN Guangcong[2], YANG Yi[1], WU Yi[1], CAO Qian[3], WEI Yanling[2], NIAN Yongjian[1] [1]Department of Digital Medicine, School of Biomedical Engineering and Imaging Medicine, Army Medical University (Third Military Medical University), Chongqing, 400038; [2]Department of Gastroenterology, Army Medical Center of PLA, Chongqing, 400042; [3]Department of Gastroenterology, Sir Run Run Shaw Hospital, Zhejiang University School of Medicine, Hangzhou, Zhejiang Province, 310016, China

[**Abstract**] **Objective** To identify ulcerative colitis (UC) and Crohn's disease (CD) with aid of deep learning technology for endoscopists. **Methods** From January 2018 to November 2020, the endoscopic images of 1 576 subjects (including 34 300 CD, UC and normal images) were collected from the Department of Gastroenterology of Army Medical Center of PLA and Sir Run Run Shaw Hospital. The training set and test set were randomly divided according to the ratio of 9∶1 to train and test the neural network. A novel spatial and bilinear deep network (SABA-ResNet) was constructed on the basis of ResNet50. The spatial attention mechanism was introduced, and the receptive field was expanded by dilated convolution to leverage contextual information, which was combined with the local induction of standard convolution to adaptively focus the lesion region. Bilinear attention was applied to improve the feature representation ability of the network, and the second-order information was used to weight the channel information of the feature map, so as to improve the

[基金项目] 重庆市研究生科研创新项目(CYS22746)

[通信作者] 粘永健,E-mail:yjnian@tmmu.edu.cn

2097-0927/Copyright © 2023, 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/)

classification performance of the model. **Results** The overall accuracy of SABA-ResNet for the recognition of CD, UC and normal tissues on the test set was 92.67% (95%*CI*: 91.91 ~ 93.37), the AUC value was 0.978 (95%*CI*: 0.972~0.983), 0.977 (95%*CI*: 0.971~0.982) and 0.999 (95%*CI*: 0.998~1.000), the sensitivity was 88.40%, 89.07% and 98.89%, the specificity was 95.49%, 94.88% and 98.93%, and the F1 value was 88.80%, 89.01% and 98.60%, respectively. The ablation experiment and the class activation map suggested that spatial attention and bilinear attention could help the model capture more features of the lesion region. **Conclusion** Our constructed network combines spatial attention and bilinear attention, achieves excellent performance in the recognition of CD, UC and normal tissue, and effectively assist endoscopists in the diagnosis of UC and CD.

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

Supported by the Project of Postgraduate Scientific Research Innovation of Chongqing (CYS22746). Corresponding author: NIAN Yongjian, E-mail: yjnian@tmmu.edu.cn

炎症性肠病(inflammatory bowel disease, IBD)包括溃疡性结肠炎(ulcerative colitis, UC)和克罗恩病(Crohn's disease, CD),是由遗传、环境、免疫和微生物因素相互作用引起的慢性胃肠道炎症性疾病。IBD具有遗传倾向且不可治愈,还将增加患结直肠癌的风险[1]。作为一种多系统疾病,除胃肠道外,IBD还会影响肌肉骨骼、眼部和皮肤系统[2],严重影响了患者的生活质量。21世纪以来,受环境因素影响,亚洲国家IBD的发病率和流行率逐渐增高[3]。由于UC和CD在治疗策略、预后、综合评估和临床护理等方面存在较大差异,因此,针对两者的精准识别对于临床治疗决策的制定以及提高预后具有重要意义。

内镜检查在IBD的诊断、治疗和随访监测中起着至关重要的作用,可用于鉴别UC和CD,识别病变的严重程度及范围[4],但其鉴别过程主要依靠内镜医师的经验与水平,当CD病变仅累及结肠时,易混淆为UC。此外,内镜医师长时间的工作疲劳也在一定程度上影响了UC与CD的识别准确性。

人工智能技术的发展改变了临床医师和研究人员处理与分析大数据集的方式。已有一些研究探索了人工智能在IBD亚型分类中的应用,在一项对59例CD患者、26例UC患者和42例健康对照者的横断面研究中,SMOLANDER等[5]将深度信念网络和支持向量机应用于基因表达数据集,对CD与UC的总准确率可达(95.71±0.21)%。TONG等[6]在一项回顾性队列研究中,收集了875例CD患者、5 128例UC患者以及396例肠结核患者的内镜图像的文本描述,通过随机森林和卷积神经网络(CNNs)进行识别,RF和CNNs诊断UC/CD的精确度分别为0.97/0.65和0.99/0.87。相比于基因表达与文本信息,图像可以更加直观地展现病变特点并具有无限细粒度的特性,包含了更加丰富的语义信息。WANG等[7]开发了ResNeXt101网络用于识别CD、UC和正常3种内镜图像,总体准确率为92.04%,其性能优于大多数临床医师,展现了深度学习在IBD亚型辅助诊断中的应用潜力。

CD和UC的病灶区域在内镜图像上呈现弥散状,在很多情况下具有较强的相似性,如何使深度学习模型更加聚焦于任务相关的高贡献区域,同时避免背景噪声的影响,是提高识别性能的关键因素。本研究基于ResNet50网络基础,提出SABA-ResNet网络,通过引入空间注意机制,加强对病灶区域的特征提取,引入双线性注意模块,在捕获二阶统计信息的同时,建立了二阶局部与全局特征之间的相互依赖性,加强了网络对关键特征的识别能力,有利于提升模型的分类性能。

1 材料与方法

1.1 数据集构建

收集2018年1月至2020年11月期间,陆军特色医学中心消化内科与浙江大学邵逸夫医院消化内科共1 576例受试者的34 300张内镜图像,其中CD患者556例,UC患者658例,正常362例。本研究经过陆军特色医学中心伦理委员会的批准(2021-285)。

本研究按9∶1的比例将数据集随机划分为训练集与测试集,其中训练集和测试集分别有29 414张和4 886张内镜图像。具体的数据来源分布与数据集的划分见表1。

表1 数据来源与数据集划分(张)

数据来源	CD	UC	正常	总计
陆军特色医学中心	1 093	3 379	12 689	17 161
邵逸夫医院	9 670	7 469	0	17 139
训练集	9 315	9 302	10 797	29 414
测试集	1 448	1 546	1 892	4 886
总计	10 763	10 848	12 689	34 300

1.2　整体网络结构

SABA-ResNet 结构如图 1 所示,主干网络为 ResNet50,其主体有一个茎(Stem)和 4 个阶段,茎中包括一个 7×7 卷积层(Convolution),一个批归一化层(BatchNorm),一个 ReLU 激活函数和一个最大池化层(Maxpool),而每个阶段中包括一个下采样模块(Bottleneck Downsample)和若干个残差模块(Bottleneck)。在前 3 个阶段输出后,引入 SA 模块,而在最后一个阶段输出后,引入 BA 模块,最后经过平均池化层(Avgpool)与全连接层(FC)后输出 3 种类别的概率分数。

1.3　空间注意模块

空间注意模块的结构如图 2 所示,若空间注意模块的输入特征为 $F_{input}^{SA}=\mathbb{R}^{C_i\times H_i\times W_i}$,其中 C_i、H_i 和 W_i 分别为第 i 阶段的通道、高度和宽度。F_{input}^{SA} 首先由 1×1 卷积进行通道降维以减少计算量,得到大小为 $\mathbb{R}^{C_i\times H_i\times W_i}$ 的中间特征,其中 r 为降维因子,实验中设置为 4。随后利用 3×3 的普通卷积和膨胀卷积共同编码空间信息,膨胀卷积[8]具有更大的感受野,可有效利用上下文信息,对弥漫状病灶有更好的辨别能力,而普通卷积结构紧凑,用以捕获局部细节信息,对散发状病灶适应性更强,随后两个分支均由 3×3 卷积将通道维度压缩至 1,得到的空间映射分别为 $M_1\in\mathbb{R}^{1\times H_i\times W_i}$ 和 $M_2\in\mathbb{R}^{1\times H_i\times W_i}$,其计算如下:

$$M_1=\mathrm{Conv}_{3\times3}(\mathrm{Conv}_{3\times3}^{d=1}(\mathrm{Conv}_{1\times1}(F_{input}^{SA})))\quad(1)$$

$$M_2=\mathrm{Conv}_{3\times3}(\mathrm{Conv}_{3\times3}^{d=2}(\mathrm{Conv}_{1\times1}(F_{input}^{SA})))\quad(2)$$

其中 Conv(·)为卷积运算,d 为膨胀因子,$d=1$ 时为普通卷积。通过平均池化加入原始输入 F_{input}^{SA} 的空间信息,并与 M_1、M_2 沿通道维度拼接得到的特征映射为 $M_S\in\mathbb{R}^{3\times H_i\times W_i}$。最终,采用 5×5 卷积与激活函数生成空间注意映射 $A\in\mathbb{R}^{1\times H_i\times W_i}$,计算过程如式(3)和(4)所示:

$$M_S=\mathrm{Concat}(M_1,M_2,\mathrm{Avg}(F_{input}^{SA}))\quad(3)$$

$$A=\mathrm{Sig}(\mathrm{Conv}_{5\times5}(M_S))\quad(4)$$

其中 Avg(·)为平均池化,Sig(·)为 Sigmoid 激活函数,Concat(·)为拼接操作。生成的空间注意映射 A 与 F_{input}^{SA} 进行元素级相乘后再与 F_{input}^{SA} 相加得到 $F_{output}^{SA}\in\mathbb{R}^{C_i\times H_i\times W_i}$,以此作为下一个阶段的输入,计算过程如式(5)所示:

$$F_{output}^{SA}=(A\otimes F_{input}^{SA})\oplus F_{input}^{SA}\quad(5)$$

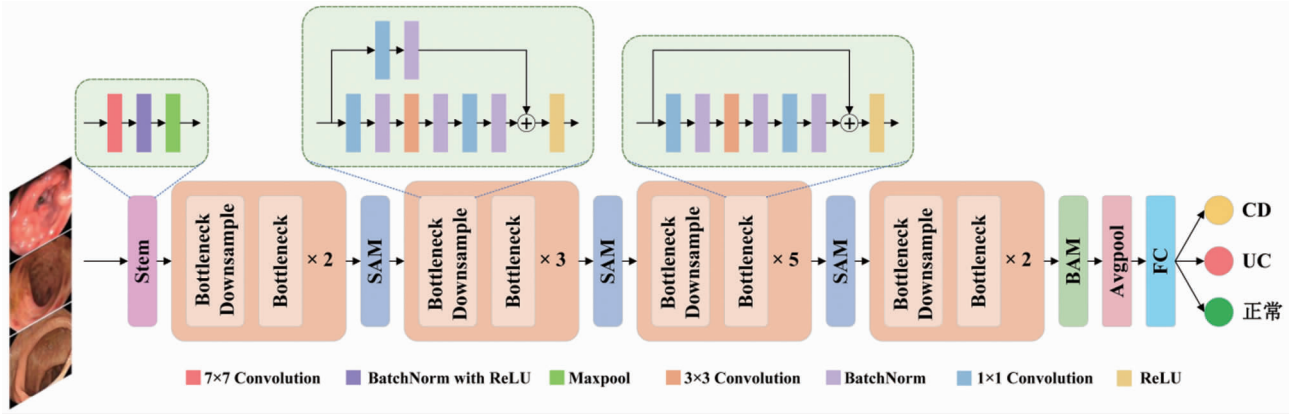


图 1　SABA-ResNet 整体网络结构

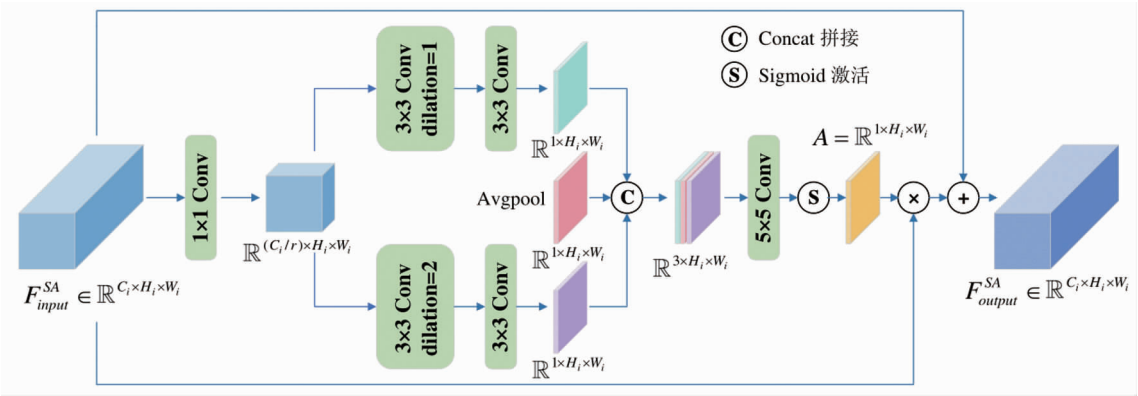


图 2　空间注意模块详细结构与特征映射变化

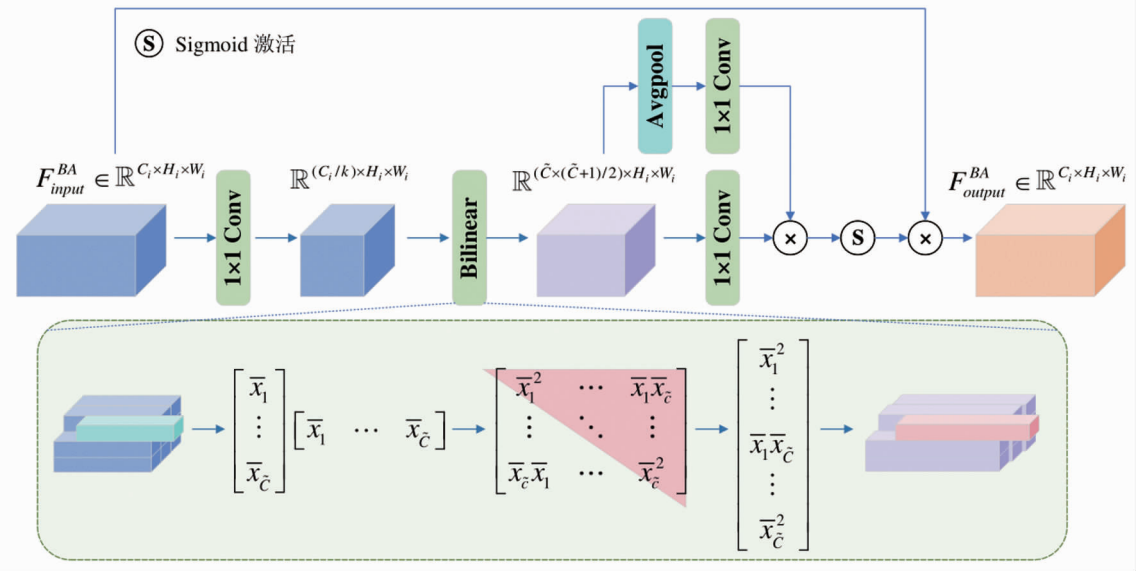


图 3 双线性注意模块详细结构与特征映射变化

1.4 双线性注意模块

双线性注意模块结构如图 3 所示,设双线性注意模块的输入特征图为 $F_{input}^{BA} \in \mathbb{R}^{C_i \times H_i \times W_i}$,设在空间($m$,$n$)位置的向量为 $\boldsymbol{x} \in \mathbb{R}^{C_i}$,其中 $m \in \{1,2\cdots\cdots,H_i\}$,$n \in \{1,2\cdots\cdots,W_i\}$,$F_{input}^{BA}$ 首先由 1×1 卷积进行通道降维得到$\mathbb{R}^{C_i \times H_i \times W_i}$ 大小的中间特征,双线性算子可实现为:

$$Y = \theta(\boldsymbol{x})\theta(\boldsymbol{x})^{\top} = \bar{\boldsymbol{x}}\,\bar{\boldsymbol{x}}^{\top} = \begin{bmatrix} \bar{x}_1^2 & \cdots & \bar{x}_1\bar{x}_{\tilde{c}} \\ \vdots & \ddots & \vdots \\ \bar{x}_{\tilde{c}}\bar{x}_1 & \cdots & \bar{x}_{\tilde{c}}^2 \end{bmatrix} \tag{6}$$

其中 $\theta(\boldsymbol{x}) \in \mathbb{R}^{\tilde{C}}$,$\tilde{C} = C_i/k$,$k$ 为通道降维因子。此时 Y 是一个对称矩阵,为了减少特征冗余,取 Y 的上三角区域再向量化可将通道维度从 $\tilde{C}^2$ 减少到 $\tilde{C}\times(\tilde{C}+1)/2$,如式(7)所示:

$$\mathrm{Bu}(\bar{\boldsymbol{x}}) = \mathrm{Vec}(\mathrm{Triu}(\bar{\boldsymbol{x}}\bar{\boldsymbol{x}}^{\mathrm{T}})) \tag{7}$$

其中 Triu(·)和 Vec(·)分别代表上矩阵三角区域提取与向量化操作。Bi(·)运算后得到的中间特征为 $U \in \mathbb{R}^{(\tilde{C}\times(\tilde{C}+1)/2)\times H_i \times W_i}$,$U$ 中的向量 $\hat{\boldsymbol{x}}$ 包含了 $\bar{\boldsymbol{x}}$ 的二阶信息。与 SE 模块中的挤压激励相似,U 经全局平均池化后由 1×1 卷积将通道维度降低至与输入特征的通道维度一致,全局信息与局部特征建立相互依赖联系之后,可自适应地重新加权二阶特征响应,最终由 Sigmoid 函数激活,通道注意映射与 F_{input}^{BA} 进行元素级相乘得到 $F_{output}^{BA} \in \mathbb{R}^{C_i \times H_i \times W_i}$,计算过程如式(8)所示。

$$F_{output}^{BA} = \mathrm{Sig}(\mathrm{Conv}_{1\times1}(\mathrm{Avg}(U)) \otimes \mathrm{Conv}_{1\times1}(U)) \otimes F_{input}^{BA} \tag{8}$$

1.5 实施细节与评价指标

在训练过程中,为了增强图像的多样性,在原始图像上随机裁剪 224×224 大小的区域作为模型的输入,而在测试时,将原始图像缩放到 256×256 大小,然后中心裁剪 224×224 大小的区域作为输入。对于对比模型中的 InceptionV4 网络,图像缩放到 320×320 大小,网络输入尺寸为 300×300。为了增强模型的泛化性同时避免过拟合,采用了随机水平和随机垂直翻转的数据增强策略,执行概率设置为 0.3。模型的初始学习率设置为 0.01,一共经过 50 个 epoch 迭代,每 10 个 epoch 学习率降低 0.1 倍。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 中的 momentum 和 weight_decay 分别设置为 0.9 和 1×10^{-4},采用迁移学习策略,所有模型均使用了在 ImageNet 数据集上的预训练参数。实验使用 Pytorch 框架,并在两块 NVIDIA Quadro GV100 32G 显卡上完成模型的训练与测试。

$$\text{总体 } ACC = TN_{all}/N \tag{9}$$

$$ACC = (TP+TN)/N \tag{10}$$

$$SEN = TP/(TP+FN) \tag{11}$$

$$SPE = TN/(TN+FP) \tag{12}$$

$$PPV = TP/(TP+FP) \tag{13}$$

$$NPV = TN/(TN+FN) \tag{14}$$

$$F1 = 2\times\frac{Recall\times Precision}{Recall+Precision} \tag{15}$$

使用总体准确率(accuracy, ACC)、单分类 ACC、灵敏度(sensitivity, SEN)、特异性(specificity, SPE)、阳性预测值(positive predictive value, PPV)、阴性预测值(negative predictive value, NPV)与 F1 值全方位评估模型的性能。所有指标的计算如式(9)~(15)所示,其中 N 代表所有测试集所有样本,TN_{all} 为所有计

算正确的样本,*TP*、*FP*、*TN* 和 *FN* 分别表示真阳性、假阳性、真阴性和假阴性的样本数。

2 结果

2.1 通道降维因子 *k* 的影响

首先探索 BA 模块中通道降维因子 *k* 的影响,由于 ResNet50 网络最后一个阶段输出特征的通道数为 2 048,因而设置 *k* 分别取 4、8、16 和 32 时,降维后的通道分别为 512、256、128 和 64,降维后的通道数目越多,执行双线性运算时涵盖的二阶信息越多,但计算成本开销也更大。结果显示,*k* 取 16 时模型取得最高的识别性能(图 4)。因此,本研究将 *k* 设置为 16。

2.2 模型对比

在训练集上进行 10 折交叉验证用以确定模型训练过程中所有的超参数,交叉验证的总体准确率为 92.46%(95%*CI*:92.16~92.76),CD、UC 和正常的 AUC 分别为 0.977(95%*CI*:0.975~0.979)、0.972(95%*CI*:0.969~0.974)与0.998(95%*CI*:0.998~0.999)。为了验证所提出方法的有效性,将 SABA-ResNet 与 VGG19、ResNet34、InceptionV4、SE-ResNet50 和 DenseNet121 等模型在测试集上进行性能比较,结果见表 2。显然 SABA-ResNet 具有明显的优势,总体准确率达到 92.67%(95%*CI*:91.91~93.37),CD、UC 和正常的单分类准确率分别为 93.39%、93.04%和 98.92%。

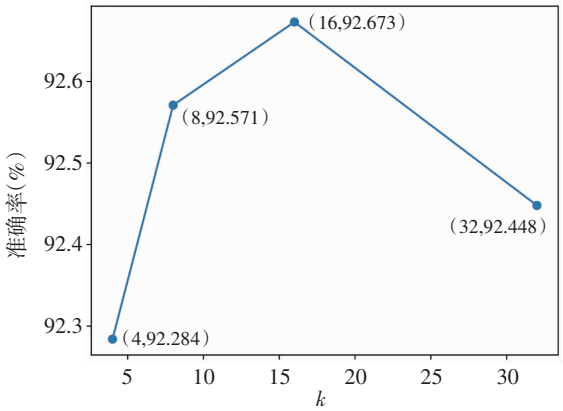
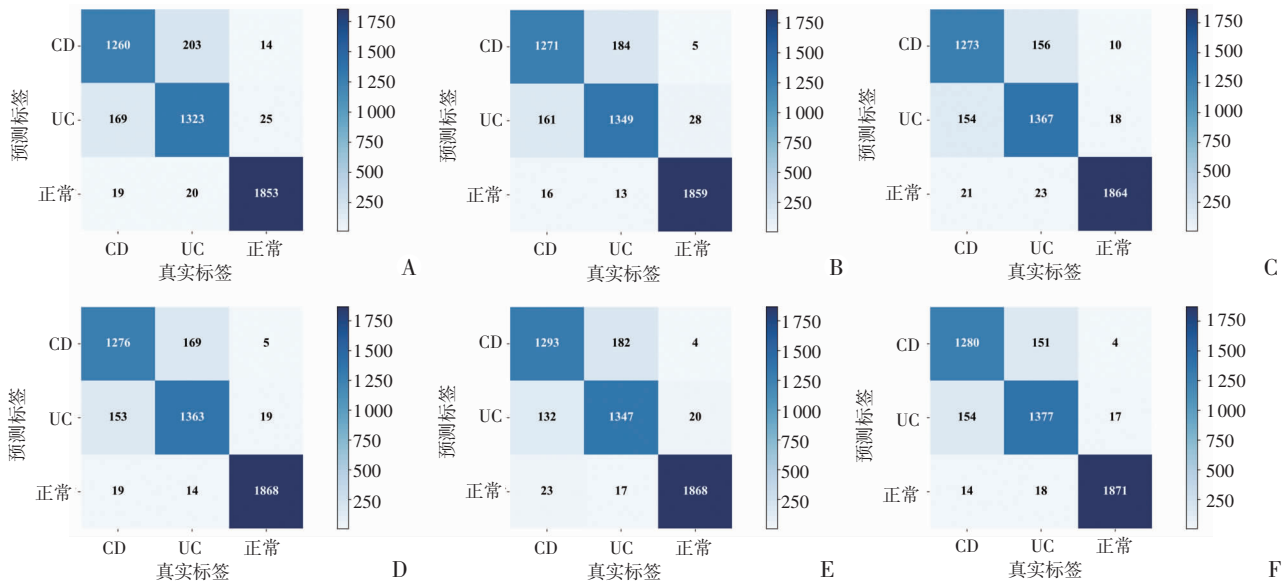


图 4 通道降维因子 *k* 的影响

表 2 SABA-ResNet 与经典卷积神经网络模型的指标对比(%)

指标	VGG19	ResNet34	InceptionV4	SE-ResNet50	DenseNet121	SABA-ResNet
总体 ACC	90.79	91.67	92.18	92.24	92.26	92.67
CD						
ACC	91.71	92.51	93.02	92.92	93.02	93.39
SEN	87.02	87.78	87.91	88.12	89.30	88.40
SPE	93.69	94.50	95.17	94.94	94.59	95.49
PPV	85.31	87.05	88.46	88.00	87.42	89.20
NPV	94.49	94.83	94.92	94.99	95.45	95.13
F1	86.15	87.41	88.19	88.06	88.35	88.80
UC						
ACC	91.47	92.10	92.82	92.73	92.82	93.04
SEN	85.58	87.26	88.42	88.16	87.13	89.07
SPE	94.19	94.34	94.85	94.85	95.45	94.88
PPV	87.21	87.71	88.82	88.79	89.86	88.95
NPV	93.38	94.12	94.65	94.54	94.12	94.94
F1	86.39	87.48	88.62	88.48	88.47	89.01
正常						
ACC	98.40	98.73	98.53	98.83	98.69	98.92
SEN	97.94	98.26	98.52	98.73	98.73	98.89
SPE	98.70	99.03	98.53	98.90	98.66	98.93
PPV	97.94	98.46	97.69	98.26	97.90	98.32
NPV	98.70	98.90	99.06	99.20	99.19	99.30
F1	97.94	98.36	98.11	98.50	98.32	98.60



A:VGG19;B:ResNet34;C:InceptionV4;D:SE-ResNet50;E:DenseNet121;F:SABA-ResNet

图 5 各模型的混淆矩阵对比

3 个类别的灵敏度分别为 88.40%、89.07% 和 98.89%,特异性分别为 95.49%、94.88% 和 98.93%,F1 值分别为 88.80%、89.01% 和 98.60%。各模型的混淆矩阵见图 5,直观展示了不同模型的分类性能。绘制 SABA-ResNet 在测试集上的 ROC 曲线,CD、UC 和正常的 AUC 分别为 0.978(95%*CI*:0.972~0.983)、0.977(95%*CI*:0.971~0.982)和 0.999(95%*CI*:0.998~1.000),见图 6。

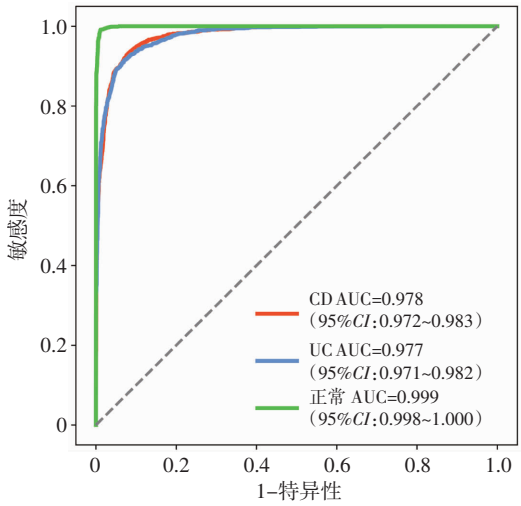


图 6 测试集 ROC 曲线

2.3 消融实验

通过消融实验可以验证 BA 与 SA 模块的有效性,实验结果见表 3。与基线模型 ResNet50 相比,BA 与 SA 模块分别在总体准确率上提升了 0.36%和 0.45%,而综合 BA 与 SA 的模型 SABA-ResNet 总体准确率提升了 0.73%,单分类准确率与 F1 值均取得最优结果。

借助梯度加权类激活映射(gradient-weighted class activation mapping, Grad-CAM)算法[9]可以从视觉角度解释模型做出决策的依据,展示模型所关注的区域。利用 Grad-CAM 生成的类激活映射见图 7。

3 讨论

受人类视觉系统感知过程的启发,注意力机制在自然语言处理[10]和计算机视觉[11]领域得到了广泛应用。注意力机制主要分为通道注意、空间注意以及混合注意,经典代表分别为 SENet 网络[12]、STN 模型[13]和 CBAM 模块[14],其中通道注意自适应校准每一个通道的权重,决定需要注意什么,而空间注意为一种适应性空间区域选择,聚焦哪里需要注意,混合注意耦合了空间注意与通道注意的优势,自适应选择重要的目标与区域。然而,SE 模块中过于简单的平均池化难以捕获复杂的全局信息,因此,不少研究通过改进挤压激励过程来提高网络的建模能力,如采用二阶池化[15]、离散余弦变换[16]等。双线性池化最早在 LIN 等[17]的研究中用于特征融合,随后,双线性池化不断改进,双线性注意应运而生。KIM 等[18]构建了双线性注意网络用于多模态学习,以有效地利用给定的视觉和语言信息。FANG 等[19]设计了一种新颖的双线性注意模块,在人的检索中表现优异。

表 3 消融实验指标对比结果(%)

模型	总体 ACC		ACC	SEN	SPE	PPV	NPV	F1
ResNet50	91.94	CD	92.90	87.78	95.06	88.20	94.86	87.99
		UC	92.55	87.84	94.73	88.53	94.39	88.18
		正常	98.42	98.47	98.40	97.49	99.03	97.98
ResNet50+BA	92.30	CD	93.12	88.74	94.97	88.13	95.25	88.44
		UC	92.84	87.84	95.15	89.34	94.41	88.58
		正常	98.65	98.68	98.63	97.85	99.16	98.26
ResNet50+SA	92.39	CD	93.21	88.88	95.03	88.27	95.30	88.58
		UC	92.98	88.16	95.21	89.49	94.56	88.82
		正常	98.59	98.52	98.63	97.85	99.06	98.18
SABA-ResNet	92.67	CD	93.39	88.40	95.49	89.20	95.13	88.80
		UC	93.04	89.07	94.88	88.95	94.94	89.01
		正常	98.92	98.89	98.93	98.32	99.30	98.60

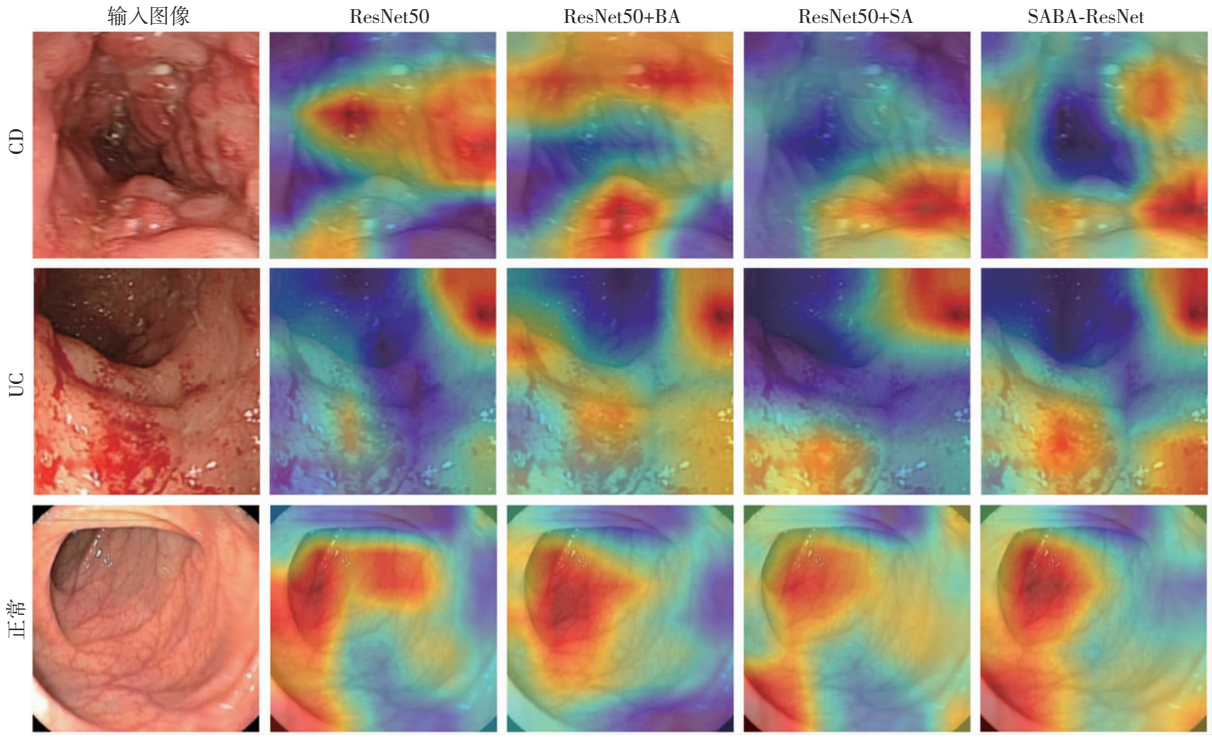


图 7 不同数据集内镜图像类激活映射图对比

在利用各类医学图像构建计算机辅助诊断模型时,注意力机制常被用来聚焦病变区域,增强关键特征的提取能力。XIAO 等[20]设计了由通道注意与空间注意并联组成的并行注意模块,在利用 CT 图像诊断新冠肺炎中自适应校准特征。ZHANG 等[21]提出了一种注意力残差学习模型,利用高层学习到的特征映射生成低层的注意力映射,在皮肤病分类中表现优异。XING 等[22]在胶囊内镜图像分类任务中,通过注意力引导进行病变放大,同时采用三阶长距离特征聚合,取得了先进的实验结果。

CD 和 UC 是 IBD 的两个主要亚型,均会导致消化道系统紊乱与胃肠道炎症,且具有相似的临床症状,如腹痛腹泻、直肠出血和体质量减轻[23]。消化内镜的普及使得内镜检查在 CD 与 UC 的识别中占主导地位,然而其严重依赖于内镜医师的经验与水平,因经验缺乏或审查疲劳对内镜图像产生的误判将会导致漏诊或者治疗延误[24]。本研究通过构建深度学习网络模型并引入注意力机制,为提高 CD 与 UC 的诊断性能提供了一种新的解决方案。总体来看,正常类别相对容易区分,而 CD 与 UC 之间更容易发生混淆。对比实验中,DenseNet121 模型在识别 CD 时具有更低的假阴性,而识别 UC 时具有更低的假阳性,这将会导致模型更偏向于将 UC 识别为 CD,兼顾假阳性与假阴性,SABA-ResNet 在所有被比较的模型中表现最好。虽然 SE-ResNet50 在 ResNet50 网络结构基础上加入了通道注意 SE 模块,但其相对于 ResNet50 并没有展现出强大

的优势,这表明仅有通道注意是不够的,空间注意在聚焦病灶区域方面仍然发挥着重要作用。Grad-CAM 类激活映射图中高亮红色区域集中在肠腔中鹅卵石样病变、假息肉、溃疡与出血等位置,这与临床中 CD 与 UC 内镜下诊断的主要特征一致[4]。对比类激活映射可以发现,引入 BA 和 SA 模块,有助于模型捕获更多病变区域的信息。鉴于 CD 与 UC 在内镜下特征存在一定的相似性,仅靠局部位置的信息很容易造成混淆,多病变区域信息综合分析更有利于模型做出准确的判断。对于正常图像来说,高亮部分是相对分散的,但是也是作为模型的判别依据。

本研究基于 ResNet50 构建了端到端训练的分类模型 SABA-ResNet,针对静态肠镜图像复杂多变的弥散状病变特征,在网络早期引入了空间注意机制,采用感受野更大的膨胀卷积联系上下文信息,与普通卷积的局部归纳特性相配合,促使网络自适应提取重要病灶区域的信息,减少背景噪声的干扰。在网络输出之前引入双线性注意机制,作为一种带有二阶信息的通道注意,加权了用于全局平均池化特征的不同通道,对于分类任务来说,提升了不同类别特征的可鉴别性,有效提升模型的分类性能。虽然二阶信息相对于一阶信息具有更好的建模能力,但目前对二阶信息的使用存在计算开销大的问题。本研究在实现过程中,通过通道降维,减少了计算开销,并讨论了通道降维因子的影响,实验结果表明,用于二阶信息构建的通道并不是越多越好,这与文献[19]中的实验讨论一致。过多的二阶信息可能会扰乱特征分布,给分类带来不利的影响,由此可以提出一个新的问题,二阶信息甚至更高阶的信息提取会有利于哪些非抽象特征的进一步建模,这将是一个值得探索的研究方向。此外,二阶统计量带来的计算量给模型训练带来了较大的负担,图形处理器(graphics processing unit, GPU)对矩阵特征值分解与奇异值分解的不友好性,也成为二阶信息在神经网络构建中应用的阻碍。LI 等[25]提出了一种矩阵协方差迭代归一化方法,在可接受的计算成本中实现了模型的端到端训练,为高阶统计量的应用提供了新的思路。

本研究仍然存在一些局限性。虽然回顾性收集的内镜图像数据来自两个医学中心,但对于模型泛化性的评估仍然不足。内镜图像拍摄时的设备、光照、角度甚至医师的操作差异,会给模型使用时的稳定性与准确性带来挑战。其次,使用的是静态三通道彩色内镜图像,在临床的内镜检查中,医师会动态地观察患者的肠道情况,结合病变的位置信息做出综合判断。因此,本研究是计算机辅助诊断在炎症性肠病应用中的一个初步探索。

本研究设计了一种带有空间注意与双线性注意的卷积网络模型 SABA-ResNet,以空间注意自适应聚焦病变区域的同时,以双线性注意提高特征表示能力,在识别 CD、UC 与正常内镜图像中具有较高的准确率、灵敏度、特异性与 F1 值,其性能优于目前经典的卷积模型。作为一种辅助诊断工具,可帮助经验不足的内镜医师快速了解 CD 与 UC 的病灶特点,同时减轻内镜医师的工作负担,提高 IBD 亚型的诊断准确率。

参考文献:

[1] MCDOWELL C, FAROOQ U, HASEEB M. Inflammatory bowel disease[M]. Treasure Island: StatPearls Publishing, 2022.

[2] MALIK T F, AURELIO D M. Extraintestinal manifestations of inflammatory bowel disease [M]. Treasure Island: StatPearls Publishing, 2022.

[3] PARK J, CHEON J H. Incidence and prevalence of inflammatory bowel disease across Asia[J]. Yonsei Med J, 2021, 62(2): 99-108. DOI:10.3349/ymj.2021.62.2.99.

[4] FLYNN S, EISENSTEIN S. Inflammatory bowel disease presentation and diagnosis[J]. Surg Clin North Am, 2019, 99(6): 1051-1062. DOI:10.1016/j.suc.2019.08.001.

[5] SMOLANDER J, DEHMER M, EMMERT-STREIB F. Comparing deep belief networks with support vector machines for classifying gene expression data from complex disorders [J]. FEBS Open Bio, 2019, 9(7): 1232-1248. DOI:10.1002/2211-5463.12652.

[6] 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.

[7] WANG L J, CHEN L P, WANG X Y, *et al*. Development of a convolutional neural network-based colonoscopy image assessment model for differentiating Crohn's disease and ulcerative colitis[J]. Front Med, 2022, 9: 789862. DOI: 10.3389/fmed.2022.789862.

[8] YU F, KOLTUN V. Multi-scale context aggregation by dilated convolutions[EB/OL]. 2015, arXiv: 1511.07122. https://arxiv.org/abs/1511.07122.

[9] SELVARAJU R R, COGSWELL M, DAS A, *et al*. Grad-CAM: visual explanations from deep networks via gradient-based localization [C]. IEEE International Conference on Computer Vision, 2017: 618-626.

[10] VASWANI A, SHAZEER N, PARMAR N, *et al*. Attention is all You need[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. New York: ACM, 2017: 6000-6010. DOI: 10.5555/3295222.3295349.

[11] GUO M H, XU T X, LIU J J, *et al*. Attention mechanisms in computer vision: a survey[J]. Comput Vis Media, 2022, 8(3): 331-368. DOI:10.1007/s41095-022-0271-y.

(下转 242 页)

intra-observer reproducibility of cardiovascular magnetic resonance myocardial feature tracking strain assessment is independent of field strength[J]. Eur J Radiol, 2013, 82(2): 296-301. DOI:10.1016/j.ejrad.2012.11.012.

[16] 郑娇凤, 王磊, 胡英, 等. 磁共振组织追踪成像对早期左心室心肌肥厚大鼠模型的评估[J]. 中国医学影像学杂志, 2021, 29(8): 763-769. DOI:10.3969/j.issn.1005-5185.2021.08.002.
ZHENG J F, WANG L, HU Y, *et al*. Evaluation of magnetic resonance tissue tracking in rat model of early left ventricular hypertrophy[J]. Chin J Med Imaging, 2021, 29(8): 763-769. DOI:10.3969/j.issn.1005-5185.2021.08.002.

[17] WANG X X, CHEN T, ZHANG S T, *et al*. Meshless deformable models for 3D cardiac motion and strain analysis from tagged MRI[J]. Magn Reson Imaging, 2015, 33(1): 146-160. DOI:10.1016/j.mri.2014.08.007.

[18] 干浩庆, 张海光. 老年脓毒症休克患者血清氨基末端脑钠肽前体、肌钙蛋白I水平及其与预后的相关性[J]. 中国老年学杂志, 2019, 39(23): 5760-5762. DOI:10.3969/j.issn.1005-9202.2019.23.037.
GAN H Q, ZHANG H G. Serum amino-terminal brain natriuretic peptide precursor and troponin I levels in elderly patients with septic shock and their correlation with prognosis[J]. Chin J Gerontol, 2019, 39(23): 5760-5762. DOI:10.3969/j.issn.1005-9202.2019.23.037.

[19] 袁清茹, 刘莉娟, 王晋华, 等. 慢性心力衰竭患者血清心肌肌钙蛋白与心脏结构功能的关系及对预后的影响[J]. 中国老年学杂志, 2016, 36(14): 3421-3423. DOI:10.3969/j.issn.1005-9202.2016.14.031.
YUAN Q R, LIU L J, WANG J H, *et al*. Relationship between serum cardiac troponin and cardiac structure and function in patients with chronic heart failure and its effect on prognosis[J]. Chin J Gerontol, 2016, 36(14): 3421-3423. DOI:10.3969/j.issn.1005-9202.2016.14.031.

[20] REINDL M, TILLER C, HOLZKNECHT M, *et al*. Global longitudinal strain by feature tracking for optimized prediction of adverse remodeling after ST-elevation myocardial infarction[J]. Clin Res Cardiol, 2021, 110(1): 61-71. DOI:10.1007/s00392-020-01649-2.

(收稿:2022-07-10;修回:2022-09-15)
(编辑 汪勤俭)

(上接234页)

[12] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE, 2018: 7132-7141. DOI: 10.1109/CVPR.2018.00745.

[13] JADERBERG M, SIMONYAN K, ZISSERMAN A, *et al*. Spatial transformer networks[C]//Proceedings of the 28th International Conference on Neural Information Processing Systems. New York: ACM, 2015: 2017-2025. DOI: 10.5555/2969442.2969465.

[14] WOO S, PARK J, LEE J Y, *et al*. CBAM: convolutional block attention module[EB/OL]. arXiv, 2018:1807.06521. https://arxiv.org/abs/1807.06521.

[15] GAO Z L, XIE J T, WANG Q L, *et al*. Global second-order pooling convolutional networks[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2020: 3019-3028. DOI: 10.1109/CVPR.2019.00314.

[16] QIN Z Q, ZHANG P Y, WU F, *et al*. FcaNet: frequency channel attention networks[C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV). IEEE, 2022: 763-772. DOI: 10.1109/ICCV48922.2021.00082.

[17] LIN T Y, ROYCHOWDHURY A, MAJI S. Bilinear CNN models for fine-grained visual recognition[C]//2015 IEEE International Conference on Computer Vision (ICCV). IEEE, 2016: 1449-1457. DOI: 10.1109/ICCV.2015.170.

[18] KIM J H, JUN J, ZHANG B T. Bilinear attention networks[C]//Proceedings of the 32nd International Conference on Neural Information Processing Systems. New York: ACM, 2018: 1571-1581. DOI: 10.5555/3326943.3327087.

[19] FANG P F, ZHOU J M, ROY S, *et al*. Bilinear attention networks for person retrieval[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV). IEEE, 2020: 8029-8038. DOI: 10.1109/ICCV.2019.00812.

[20] XIAO B, YANG Z Y, QIU X M, *et al*. PAM-DenseNet: a deep convolutional neural network for computer-aided COVID-19 diagnosis[J]. IEEE Trans Cybern, 2022, 52(11): 12163-12174. DOI:10.1109/TCYB.2020.3042837.

[21] ZHANG J P, XIE Y T, XIA Y, *et al*. Attention residual learning for skin lesion classification[J]. IEEE Trans Med Imaging, 2019, 38(9): 2092-2103. DOI: 10.1109/TMI.2019.2893944.

[22] XING X H, YUAN Y X, MENG M Q H. Zoom in lesions for better diagnosis: attention guided deformation network for WCE image classification[J]. IEEE Trans Med Imaging, 2020, 39(12): 4047-4059. DOI:10.1109/TMI.2020.3010102.

[23] SEYEDIAN S S, NOKHOSTIN F, MALAMIR M D. A review of the diagnosis, prevention, and treatment methods of inflammatory bowel disease[J]. J Med Life, 2019, 12(2): 113-122. DOI:10.25122/jml-2018-0075.

[24] RUAN G C, QI J, CHENG Y, *et al*. Development and validation of a deep neural network for accurate identification of endoscopic images from patients with ulcerative colitis and Crohn's disease[J]. Front Med (Lausanne), 2022, 9(8): 854677. DOI:10.3389/fmed.2022.854677.

[25] LI P H, XIE J T, WANG Q L, *et al*. Towards faster training of global covariance pooling networks by iterative matrix square root normalization[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE, 2018: 947-955. DOI: 10.1109/CVPR.2018.00105.

(收稿:2022-09-21;修回:2022-10-18)
(编辑 张 维)