

光电工程

Opto-Electronic Engineering

中文核心期刊 中国科技核心期刊
Scopus CSCD

基于注意力机制多特征融合的视网膜病变分级算法

梁礼明, 董信, 李仁杰, 何安军

引用本文:

梁礼明, 董信, 李仁杰, 等. 基于注意力机制多特征融合的视网膜病变分级算法[J]. *光电工程*, 2023, 50(1): 220199.

Liang L M, Dong X, Li R J, et al. Classification algorithm of retinopathy based on attention mechanism and multi feature fusion[J]. *Opto-Electron Eng*, 2023, 50(1): 220199.

<https://doi.org/10.12086/oe.2023.220199>

收稿日期: 2022-08-17; 修改日期: 2022-10-25; 录用日期: 2022-11-01

相关论文

鬼影卷积自适应视网膜血管分割算法

梁礼明, 周珑颂, 陈鑫, 余洁, 冯新刚

光电工程 2021, 48(10): 210291 doi: 10.12086/oe.2021.210291

边界注意力辅助的动态图卷积视网膜血管分割

吕佳, 王泽宇, 梁浩城

光电工程 2023, 50(1): 220116 doi: 10.12086/oe.2023.220116

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

Customized anterior segment photoacoustic imaging for ophthalmic burn evaluation *in vivo*

Huangxuan Zhao, Ke Li, Fan Yang, Wenhui Zhou, Ningbo Chen, Liang Song, Chuansheng Zheng, Zhicheng Liu, Chengbo Liu

Opto-Electronic Advances 2021, 4(6): 200017 doi: 10.29026/oea.2021.200017

更多相关论文见光电期刊集群网站 



光电工程
Opto-Electronic Engineering

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



OE_Journal



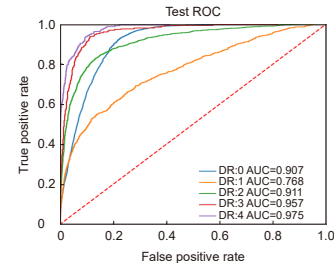
Website

DOI: 10.12086/oe.2023.220199

基于注意力机制多特征融合的视网膜病变分级算法

梁礼明*, 董信, 李仁杰, 何安军

江西理工大学电气工程与自动化学院, 江西赣州 341000



摘要: 糖尿病视网膜病变 (Diabetic Retinopathy, DR) 是糖尿病的一个常见的急性阶段, 可引起视网膜的视功能异常。针对视网膜眼底图像病灶区域识别困难以及分级效率不高等问题, 本文提出一种基于注意力机制多特征融合的算法来对 DR 进行诊断分级。首先对输入的图像采用高斯滤波等形态学预处理来提升眼底图像特征对比度; 然后用 ResNeSt50 残差网络作为模型的骨干, 引入多尺度特征增强模块对视网膜病变图像病变区域进行特征增强, 提高分级准确率; 再后利用图形特征融合模块对主干输出的特征增强后的局部特征进行信息融合; 最后采用中心损失和焦点损失组合的加权损失函数进一步提升分类效果。在印度糖尿病视网膜病变 (IDRID) 数据集中灵敏度和特异性分别为 95.65% 和 91.17%, 二次加权一致性检验系数为 90.38%。在 Kaggle 比赛数据集中准确率为 84.41%, 受试者工作特征曲线下的面积为 90.36%。仿真实验表明, 本文算法在糖尿病视网膜病变分级中具有一定的应用价值。

关键词: 视网膜病变分级; 多尺度特征; 注意力机制; 图形特征融合

中图分类号: TP391

文献标志码: A

梁礼明, 董信, 李仁杰, 等. 基于注意力机制多特征融合的视网膜病变分级算法 [J]. 光电工程, 2023, 50(1): 220199

Liang L M, Dong X, Li R J, et al. Classification algorithm of retinopathy based on attention mechanism and multi feature fusion[J]. *Opto-Electron Eng*, 2023, 50(1): 220199

Classification algorithm of retinopathy based on attention mechanism and multi feature fusion

Liang Liming*, Dong Xin, Li Renjie, He Anjun

School of Electrical Engineering and Automation, Jiangxi University of Science and Technology, Ganzhou, Jiangxi 341000, China

Abstract: Diabetic Retinopathy (DR) is a prevalent acute stage of diabetes mellitus that causes vision-affecting abnormalities on the retina. In view of the difficulty in identifying the lesion area in retinal fundus images and the low grading efficiency, this paper proposes an algorithm based on multi-feature fusion of attention mechanism to diagnose and grade DR. Firstly, morphological preprocessing such as Gaussian filtering is applied to the input image to improve the feature contrast of the fundus image. Secondly, the ResNeSt50 residual network is used as the backbone of the model, and a multi-scale feature enhancement module is introduced to enhance the feature of the lesion area of the retinopathy image to improve the classification accuracy. Then, the graphic feature fusion module is used to fuse the enhanced local features of the main output. Finally, a weighted loss function combining center loss and focal loss is used to further improve the classification effect. In the Indian Diabetic Retinopathy

收稿日期: 2022-08-17; 修回日期: 2022-10-25; 录用日期: 2022-11-01

基金项目: 国家自然科学基金资助项目 (51365017, 6146301); 江西省自然科学基金资助项目 (20192BAB205084)

*通信作者: 梁礼明, 9119890012@jxust.edu.cn。

版权所有©2023 中国科学院光电技术研究所

(IDRID) dataset, the sensitivity and specificity were 95.65% and 91.17%, respectively, and the quadratic weighted agreement test coefficient was 90.38%. In the Kaggle competition dataset, the accuracy rate is 84.41%, and the area under the receiver operating characteristic curve was 90.36%. Simulation experiments show that the proposed algorithm has certain application value in the grading of diabetic retinopathy.

Keywords: retinopathy grade; multiscale features; attention mechanism; graphic feature fusion

1 引言

糖尿病视网膜病变 (Diabetic retinopathy, DR) 是糖尿病一个常见的急性阶段, 可引起视网膜的视功能异常, 如果不及早发现治疗进而会导致失明。因此, 尽早诊断和治疗 DR 将极大地减少失明的机率^[1]。在医学临床实践中, 对 DR 诊断主要通过眼科医生分辨眼底照相机拍摄的二维 (2D) 彩色眼底图片的方式进行, 但需要经验丰富的眼科医生才能较为准确地判断。而我国具有这样高水准的眼科医生群体较小, 满足不了国内庞大的 DR 患者群体需求。因此借助计算机辅助诊断技术实现 DR 自动分级极具研究价值^[2]。

近年来, DR 分级智能诊断研究一直是医学图像处理领域中热门话题。随着深度学习的高质量发展, 利用深度神经网络算法对 DR 图像处理进行智能诊断的技术受到广泛应用。但目前的 DR 分级智能诊断过程中还存在以下两点局限: ① DR 图像中微血管瘤、硬性渗出物以及出血点等微小病变特征与视网膜周围环境差异小, 特征提取不充分; ② 医学领域公开的数据集各类样本分布不均衡。针对这些问题, 国内外学者开展了相关研究, 如 Zhou 等人^[3]提出一种多单元和多任务结构, 该结构采用多个网格结构对不同眼底输入增加网络的深度来进一步对眼底图片特征提取, 同时引入均方差损失函数最小化标签值和预测值的差异, 提高预测精度。Du 等人^[4]提出一种渐进式训练策略和随机拼图补丁生成器来有效融合不同粒度的特征和学习潜在特定粒度特征, 解决微小病变差异小提取不充分问题, 但网络推理时间长而效率不高。郑雯等人^[5]采用多种数据增强策略并对数据集进行无偏修正来解决样本分布不均衡问题, 在 ResNext50 残差网络下引入空洞卷积解决图片分辨率降低、信息丢失等问题, 再融合空间和通道注意力机制淡化图像中无用信息, 加强对病灶区域的关注, 最后采用迁移学习模式进一步学习病灶的细节特征。顾婷菲等^[6]提出一种基于多通道注意力选择机制的细粒度分级算法, 该算法结合细粒度分类方法和多通道注意力选择机制来获

取局部特征, 并引入排序损失以及优化每层通道信息量, 得到更多具有信息量的局部区域。

但以上方法在准确定位病灶区域等方面仍然具有一定的局限性。针对上述问题, 本文提出了一种基于注意力机制多特征融合算法来对 DR 进行有效地诊断分级。该算法主要改进内容如下:

1) 通过多尺度特征增强模块 (multi-scale features boosting module, MSFB) 来获取更大的感受野, 解决训练过程中下采样引起的图像分辨率降低和特征提取不充分的问题, 并且模块中空间和通道注意力机制可以加强对病变区域的关注, 优化模型性能;

2) 再通过图形特征融合模块 (graphic feature fusion module, GFFM) 进行信息融合, 不同阶段的局部特征图两两融合加强病变区域的权重, 提高模型对病变区域识别准确率;

3) 最后利用中心损失和焦点损失的加权组合缓解数据集类间差异小难以区分和数据集样本分布不平衡等问题, 进一步提升 DR 分级效果。

2 网络体系结构

2.1 主干网络原理

本文使用基于 ResNet 改进的残差网络 ResNeSt (Split-Attention Networks) 作为主干网络。ResNeSt 网络设计了一个分割的注意力模块, 可以实现跨通道注意力。ResNeSt 网络保留了完整的 ResNet 结构, 以 ResNet 样式堆叠分割的注意力块, 其基数组沿着通道的维度进行连接: $V = \text{Concat}\{V^1, V^2, \dots, V^K\}$ 。与标准残差块相同, 如果输入和输出的特征图形状相同, 则其他分割的注意力块产生的最终输出 Y 通过快捷的连接方式: $Y = V + X$ 。对于跨步残差块, 将通过适当地转换 $\tau(\cdot)$ 后再进行快捷的连接方式: $Y = V + \tau(X)$ 。例如, $\tau(\cdot)$ 可以是跨步卷积或者卷积与池化的共同操作^[7]。ResNeSt 块的结构如图 1 所示。

2.2 多尺度特征增强模块

引入空洞卷积策略通常是在不损失原始分辨率的

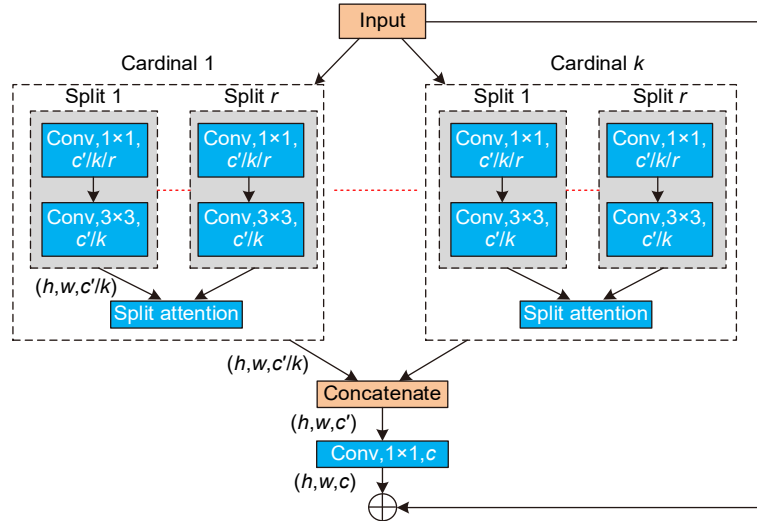


图 1 ResNeSt 块结构

Fig. 1 Structure of the ResNeSt block

情况下来扩展接收域, 该策略可以在不增加算法参数复杂度的情况下增大网络的感受野, 从而提高网络捕捉特征信息的能力^[8]。相比原来的标准卷积, 空洞卷积多了一个超参数称之为扩张率, 指的是卷积核各点之前的间隔数量。例如扩张率为 1 时的卷积核, 相当于网络进行 3×3 卷积操作; 扩张率为 2 时, 相当于进行正常的 5×5 卷积操作; 扩张率为 3 时, 进行 7×7 的卷积操作。给定输入特征图 C_i^D , 空洞卷积数学定义为

$$y(x) = \sum_k C_i^D(x + c \cdot k) \cdot W(x), \quad (1)$$

其中, x 为当前正在处理的像素位置, k 为内核大小, c 为扩张率, W 为滤波器, y 为空洞卷积的输出。本文使用三个空洞卷积 (即 d_1, d_2, d_3), 并保持滤波器权重不变, 输入的特征图 C_i^D 经过空洞卷积的输出如下:

$$H(C_i^D) = \{D_{r1}(C_i^D); D_{r2}(C_i^D); D_{r3}(C_i^D); M(C_i^D)\}, \quad (2)$$

其中, H 为扩张的卷积结果 $D(C_i^D)$ 和维度映射结果 $M(C_i^D)$ 的串联。我们将串联结果表示为 H_i^D 。维度映射操作 M 是对输入特征图进行 1×1 卷积操作, 以保证通道对空洞卷积结果的一致性, 同时保持输入的原始特征信息。

为了对空洞卷积的结果进行细化, 找出更为关键的特征信息, 在卷积结果后添加了通道和空间两个维度的注意力机制计算特征图的注意力权重。它首先通过一个全局池化操作挤压 H_i^D , 然后通过一个 1×1 卷积层激活减少特征映射^[9]。通道注意力激活数学表达式为

$$F_{ch}(H_i^D) = P_{avg}(H_i^D) \cdot W_{1 \times 1}, \quad (3)$$

其中, P_{avg} 和 $W_{1 \times 1}$ 分别表示全局池化和卷积操作。通道注意力输出 $F_{ch}(H_i^D)$ 是对 C_i^D 重新加权的一维向量。该网络更加关注了 H_i^D 子通道中的病变区域精确尺度响应。由通道注意力得到的重新加权特征图可以写成:

$$H_i^{Dch} = F_{ch}(H_i^D) \otimes H_i^D, \quad (4)$$

其中, \otimes 是对矩阵中元素进行乘法运算。本文再对重新加权的特征图进行空间维度上的计算来增加网络对 H_i^{Dch} 中病变区域反应的注意, 先使用最大池化操作挤压 H_i^{Dch} , 然后进行 3×3 卷积操作。空间注意力激活过程为

$$F_{sp}(H_i^{Dch}) = P_{max}(H_i^{Dch}) \cdot W_{3 \times 3}, \quad (5)$$

其中, P_{max} 和 $W_{3 \times 3}$ 分别表示最大池化和卷积操作。空间注意力 $F_{sp}(H_i^{Dch})$ 是一个单通道特征图, 用于过滤掉 H_i^{Dch} 的无关信息。因此, 网络会将注意力集中在病变区周围。由空间注意力得到的加强特征图映射可以表示为

$$\varphi_i = F_{sp}(H_i^{Dch}) \otimes H_i^{Dch}, \quad (6)$$

其中, \otimes 表示矩阵元素乘法运算。MSFB 结构如图 2 所示。

2.3 图形特征融合模块

2.3.1 图形轮廓投影

给定区域特征图 $\psi \in R^{N \times C}$ 和等值线图 $\varphi \in R^{N \times C}$, 空间分辨率为 $N = H \times W$, C 为通道数。本文通过将 ψ 映

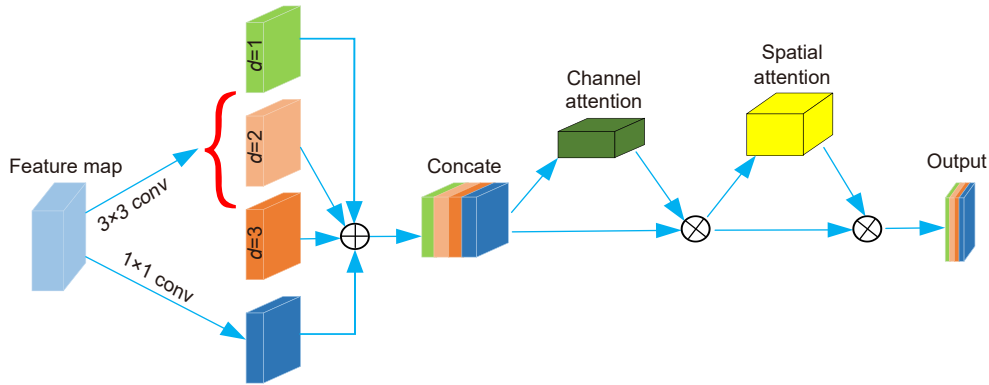


图 2 多尺度特征增强模块
Fig. 2 Multi-scale feature boosting module

射到一个由 φ 引导的图的顶点来建立一个投影矩阵 \mathbf{P} 。具体来说, 首先在特征空间中对 ψ 进行 1×1 卷积操作降低维数, 结果为 $\tau(\psi)$; 其次为了便于计算, 将等值线图 φ 置于同一维度中, 并将 $\tau(\psi)$ 与 φ 进行哈达玛积 (Hadamard product) 运算, 使得轮廓信息融入到投影矩阵中, 通过哈达玛积运算为轮廓像素的特征赋予权重, 这样轮廓像素的特征就更大; 最后采用平均池化操作获取图像中顶点的锚点, 这些锚点表示每个像素区域的中心, 本文使用 $\tau(\psi)$ 和锚点的乘积来学习每个像素和锚点之间的相似性^[10]。投影矩阵 \mathbf{P} 可定义为

$$\mathbf{P} = \text{softmax}(\psi(\tau(\psi) \odot \varphi) \cdot \tau(\psi)^T), \quad (7)$$

其中, \odot 表示哈达玛积运算。得到投影矩阵之后, 将区域像素特征图 ψ 投影到图域上可表示为

$$\psi_G = \mathbf{P}\vartheta(\psi), \quad (8)$$

其中, $\vartheta(\cdot)$ 是一个 1×1 的卷积层。这种投影操作将具有相似特征的像素聚集到相同的节点上, 每个顶点本质上表示目标图像的一个子区域^[11]。因此, 本文通过引入轮廓引导的图投影将相似的像素和每个区域联系起来。

2.3.2 图形推理

本文从 ψ_G 中提取顶点之间的连通性, 即轮廓特征和区域特征之间的关系。同时, 通过在顶点之间传播信息来学习更高层次的语义信息, 从而推断出这些潜在的关系。该策略可以通过单层图卷积网络 (GCN) 来实现^[12-13]。具体来说, 将顶点特征 ψ_G 输入到频谱图卷积的一阶近似中。 γ_G 为输出特征图, 其数学表达式如下:

$$\gamma_G = \delta[(\mathbf{I} - \mathbf{A})\mathbf{P}\vartheta(\psi)W_G], \quad (9)$$

其中, \mathbf{A} 表示对目标图进行编码学习连通性的邻接矩阵, W_G 为 GCN 的权重, \mathbf{I} 为单位矩阵, $\delta[\cdot]$ 表示

ReLU 激活函数。

2.3.3 特征投影融合

为了将改进的图表示映射回原来的 2D 像素坐标空间, 本文将提取的顶点特征重新投影到图域的原始像素网格中。考虑到学习图表示的目的是为了计算一个将 γ_G 映射到像素空间的矩阵 \mathbf{Z} , 理论上 \mathbf{Z} 可以认为是投影矩阵 \mathbf{P} 的逆, 由于 \mathbf{P} 不是方阵, 计算比较复杂。为了解决这个问题, 本文将重投影矩阵^[14]认为是一个转置矩阵 \mathbf{P}^T , 重投影后再经过 1×1 卷积运算将特征通道添加到输入区域特征 ψ 中, 最后将输入的区域特征与重投影精细化后的特征融合作为最终输出。对应的像素级特征图 M 可以定义为

$$M = \psi + \theta(\mathbf{P}^T \cdot \gamma_G), \quad (10)$$

其中, $\theta(\cdot)$ 表示 1×1 卷积层。在 GFFM 中, 输入的区域特征图是由 MSFB 得到的三个尺度一致的局部特征输出, 三个局部特征在 GFFM 中分别作为原始区域特征和等值线特征图两两融合。GFFM 结构如图 3 所示。

2.4 损失函数

由于 DR 数据集样本存在分布不均的问题会对分级的结果造成影响, 并且 DR 各级别样本之间特征差异比其他图像分类特征差异更加的精细, 这使得传统的交叉熵损失函数很难区分 DR 各个类别。为了解决以上问题, 本文提出一种焦点损失 (focal loss, FL) 函数和中心损失 (center loss, CL) 函数组合的损失函数。FL 损失函数是为了缓解样本分布不均对 DR 分级造成的影响, 且 FL 损失函数动态缩放交叉熵损失函数, 在正确类别增加时, 缩放因子下降为 0, 直观地, 这个缩放因子能自动地降低在训练期间简单样本贡献的权重和迅速关注模型中的困难样本^[15]。FL 损失函数

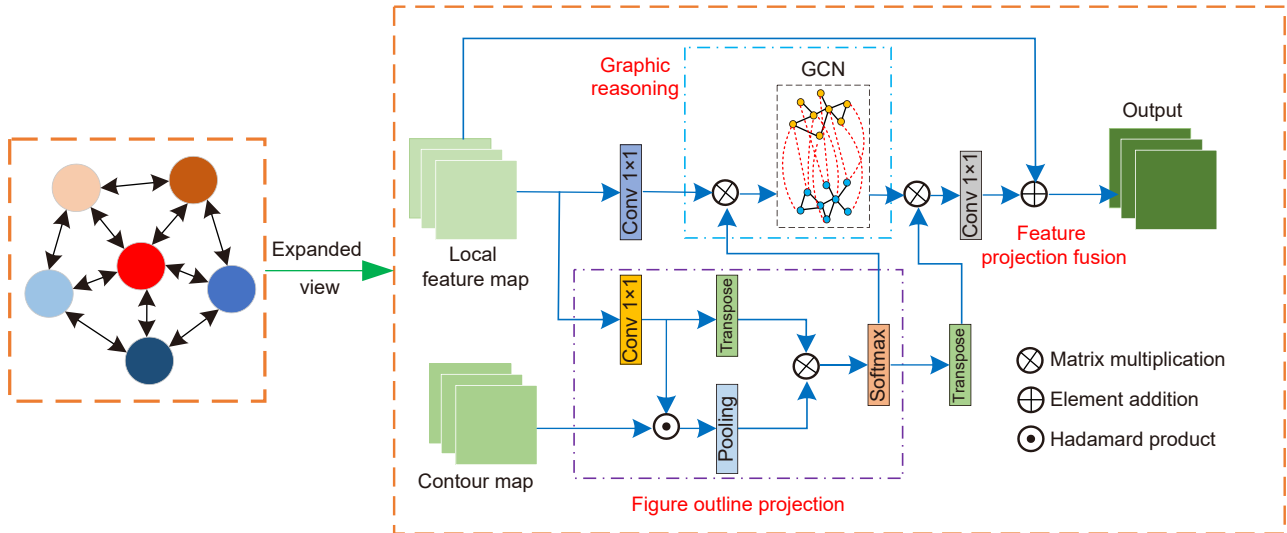


图 3 图形特征融合模块
Fig. 3 Graphic feature fusion module

计算式为

$$L_{FL} = -(1 - p_i)^r \ln(p_i), \quad (11)$$

其中, r 为加权误差调制系数(缩放因子); p_i 为标签预测概率。

CL 损失函数用于扩大类间差异来解决糖尿病视网膜病变分级中不同类别差异小难以区分的问题。CL 损失函数主要是在 softmax loss 的基础上, 通过对训练集的每个类别在特征空间分别维护一个类中心, 在训练过程, 增加样本经过网络映射后在特征空间与类中心的距离约束, 从而兼顾了类内聚合与类间分离, 即:

$$L_{CL} = \frac{1}{2N} \sum_{i=1}^N \|x_i - c_i\|_2^2, \quad (12)$$

其中, N 表示分类的类别, x_i 表示全连接层之前的特征, c_i 表示第 i 个类别的特征中心。

最终通过将 CL 损失与 FL 损失进行加权求和, 实现整体的分类任务学习。其形式如下:

$$L_{loss} = L_{FL} + \mu \cdot L_{CL}, \quad (13)$$

其中, μ 为超参数, 设为 0.001。

2.5 注意力机制多特征融合的视网膜病变分级

相比较其他的细粒度图像分类任务, DR 分级由于各等级样本的特征差异小、样本数量分布不均衡现象严重和存在不完整或噪声严重的无效图片等问题变得更具有挑战性。本文提出的注意力机制多特征融合 (attention mechanism multi-feature fusion, AMMF) 网络由 ResNeSt 主干网络、多尺度特征增强模块

(MSFB) 和图形特征融合模块 (GFFM) 三部分组成。其中 ResNeSt 主干网络采用 ImageNet 数据集训练的权重作为模型的初始参数, 这样有效地避免由于训练数据少而难以提高模型精度和过拟合的问题。首先取 ResNeSt 骨干网络中最后三层作为输出, 不同尺度输出经过 MSFB 模块, 利用扩张率不同的空洞卷积增大感受野, 通道和空间融合的注意力机制加强区域特征的权重, 过滤掉无关信息来提高模型对任务处理的效率和准确性^[16]。其次将不同分支输出的增强局部特征图在 GFFM 中两两融合进行信息互补为最终输出。最后采用 FL 和 CL 组合损失函数对梯度优化。本文算法整体框架如图 4 所示。

3 数据来源及预处理

本文采用的数据集来自国际医学影像顶级会议 ISBI 举办的 IDRID (印度糖尿病视网膜病变图像数据集) 眼底图分析竞赛中开源数据集和 Eye-PACS 数据集^[17]进行实验。IDRID 数据集包含 516 张视网膜眼底图像, 图像分辨率为 4288×4288。Eye-PACS 数据集中包含 35126 张眼底图片, 图像分辨率为 256×256。两个数据集分辨率差异较大, 对 IDRID 数据集图像进行下采样处理, 对 Eye-PACS 数据集图像进行上采样处理, 输入时统一分辨率为 512×512。专家根据医疗领域的分级标准将每幅图像都被诊断为从 0 到 4 共 5 个 DR 阶段, 标记为 {0, 1, 2, 3, 4}, 分别表示为健康、轻度、中度、重度和增殖性 DR。

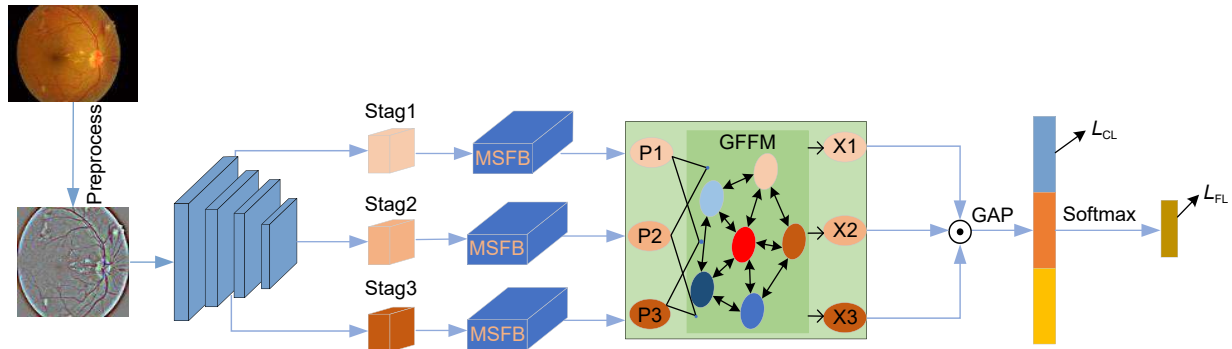


图 4 算法整体框架

Fig. 4 Overall framework of the algorithm

用于实验的两个数据集中的病变图像的背景与病变区域对比不明显, 因此训练前需要对眼底图像进行高斯滤波等形态学预处理, 具体过程如下:

1) 原始图像存在有大量黑色背景, 对其先进行灰度化, 再设定阈值进行掩膜处理, 为了突出眼底图像中小病变区域特征, 最后将原始图像与进行高斯滤波操作后的图像加权融合。其计算公式为

$$G_{\sigma} = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{x^2+y^2}{2\sigma^2}}, \quad (14)$$

$$I_d = \alpha I + \beta G_{\sigma} * I + \varepsilon, \quad (15)$$

其中, α 、 β 、 ε 为加权系数, 其值分别设定为 4、-4、128; G_{σ} 是标准差为 σ 的二维高斯核, $*$ 表示为滤波操作, I_d 为加权融合后的图像。IDRID 数据集不同 DR 阶段眼底图像预处理前后结果对比如图 5 所示。

2) 本文对 IDRID 小样本数据集训练时, 为了避免过拟合, 采取了翻转、旋转、像素值标准化和归一化的在线数据增强操作。相比较 IDRID 数据集, Eye-PACS 中样本数量大幅度提升, 避免了小数据集实验结果的随机性, 且 Eye-PACS 数据集眼底图像分辨率较低, 也可验证模型在不同数据集中的泛化性能。

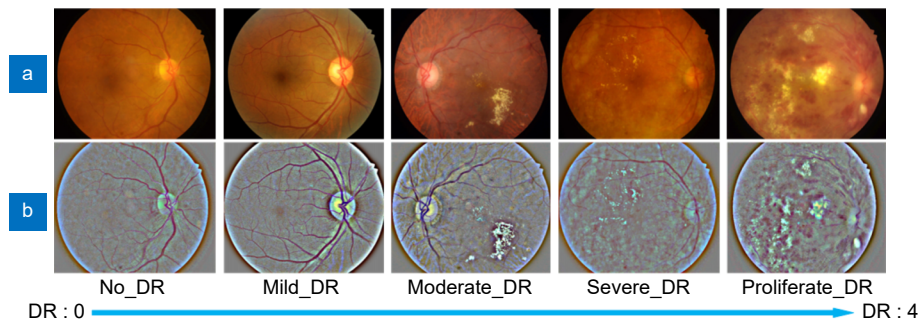


图 5 不同 DR 阶段眼底图像预处理前后对比。(a) 原图; (b) 预处理后图像

Fig. 5 Comparison of fundus images before and after preprocessing in different DR stages. (a) Original image; (b) Preprocessed image

4 实验结果

本文算法实验设备配置采用 Intel® Core™ i7-6700H CPU, 16 GB 内存, NVIDIA GeForce GTX 2070 GPU, 64 位 Win10 操作系统。基于 Pytorch1.5 框架建模, 仿真平台为 Pycharm。在训练中, 设定 batch size 为 4, 学习率为 0.002。本文算法在 IDRID 数据集上网络的一轮更新平均训练时间为 44 s/轮, 测试时间为 2 s/轮; 在 Eye-PACS 数据集上网络的一轮更新平均训练时间为 43 min/轮, 测试时间为 5 min/轮。

4.1 性能指标

为了准确评估本文所提算法在 IDRID 和 Eye-PACS 数据集上的表现, 对于不同数据集评判指标的差异性, IDRID 数据集使用了灵敏度 (sensitivity, Se)、特异性 (specificity, Sp)、准确率 (accuracy, Acc) 以及二次加权一致性检验系数 (quadratic weighted kappa, QWK, 在公式中用 η_{QWK} 表示) 等指标来评估本文算法分级的性能。其中 η_{QWK} 范围在 -1 到 1 之间, 值越大, 一致性就越高。这些指标计算公式分别如下:

$$Se = \frac{TP}{TP + FN}, \quad (16)$$

$$Sp = \frac{TN}{TN + FP}, \quad (17)$$

$$Acc = \frac{TP + TN}{TP + TN + FP + FN}, \quad (18)$$

$$\eta_{QWK} = 1 - \frac{\sum_{i,j} W_{i,j} Q_{i,j}}{\sum_{i,j} W_{i,j} E_{i,j}}, \quad (19)$$

其中, TP 和 FP 分别表示模型正确识别样本为正样本数和模型错误识别样本为正样本数。 TN 表示模型正确识别样本为负样本数。 FN 表示模型错误识别样本为负样本数。 i 和 j 代表预测类别和标签类别, $W_{i,j}$ 表示每一对 (i, j) 的惩罚权重, $Q_{i,j}$ 表示第 i 类识别为 j 类的数量, $E_{i,j}$ 表示第 i 类总数 \times 第 j 类总数/总样本数。同时在 Eye-PACS 数据集上验证评估采用 Acc、Se、Sp 以及 ROC 曲率下方的面积 (area under curve, AUC) 等指标对模型进行评判。

4.2 实验结果分析

观察 AMMF 整体网络模型的特征热图, 从而可知提出的网络模型是否对 DR 各类间的微小病理特征

和关键的病理特征进行了更多关注。从图 6 可知, AMMF 整体网络模型确实对 DR 的微小病理和关键病理特征进行了学习关注, 图 6 中蓝色方框为标注的部分病理特征, 红色方框为网络学习关注效果。这表明了 AMMF 网络模型能让微小的病理特征分配到足够的权重, 让网络自动学习识别这些病理特征, 验证了网络整体的有效性。

本文算法在 IDRID 和 Eye-PACS 数据集上训练过程损失趋势如图 7 所示。训练损失值下降迅速, 两数据集上损失基本上维持在 0.1 左右, 且波动幅度小, 则表明网络已趋于收敛。

为了更加直观地表现 AMMF 网络模型中各个模块的效果, 对模型进行消融实验。本文设计了五组对照实验: ① M1: 只保留模型算法的骨干网络 ResNeSt; ② M2: 将模型 AMMF 的组合损失函数改为交叉熵损失函数; ③ M3: 仅去掉 AMMF 算法中的 MSFB 模块; ④ M4: 只去掉 AMMF 算法中的 GFFM 模块; ⑤ 本文所提完整的 AMMF 模型算法。结果如表 1 所示。

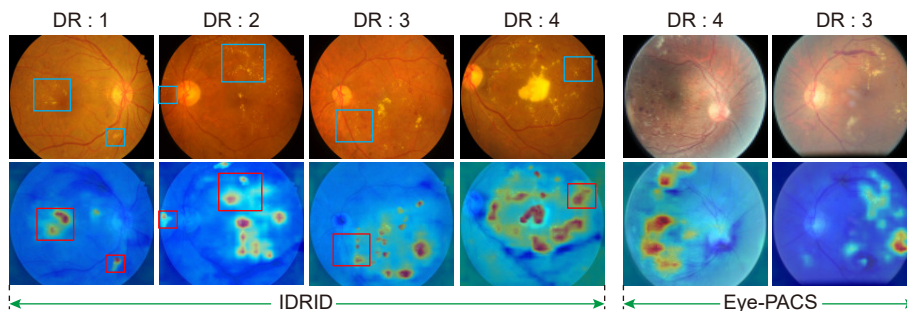


图 6 AMMF 网络特征热图
Fig. 6 AMMF network feature heatmap

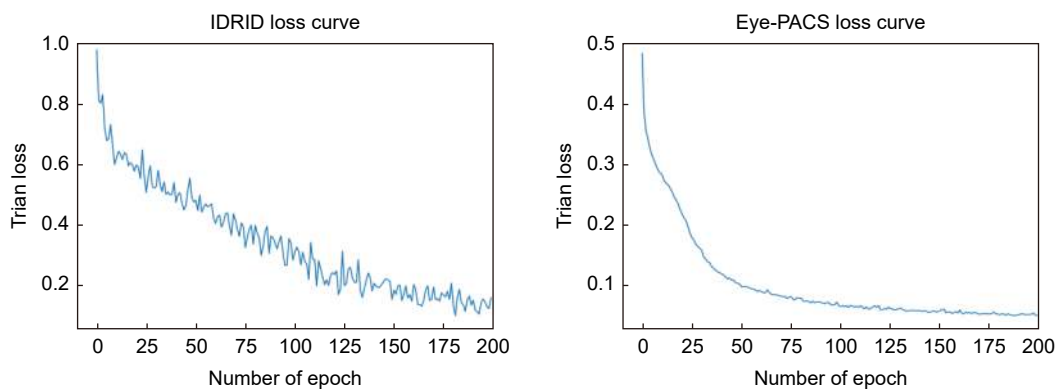


图 7 在 IDRID 和 Eye-PACS 数据集上的训练损失曲线
Fig. 7 Training loss curves on IDRID and Eye-PACS datasets

表 1 消融实验在 IDRID 数据集中的表现

Table 1 Performance of ablation experiments in the IDRID dataset

Model	QWK/%	Acc/%	Se/%	Sp/%
M1	84.83	71.84	88.48	88.23
M2	85.08	74.75	89.80	94.11
M3	87.70	77.66	92.75	89.80
M4	89.59	76.69	96.10	88.23
AMMF	90.38	78.64	95.65	91.17

由表 1 中 M1 和 AMMF 两组实验可知, 多尺度特征增强模块、图形特征融合模块和组合的损失函数能够使二次加权一致性系数、准确率、灵敏度和特异性这些指标分别提升 5.55%、6.8%、7.17%、2.94%, 说明本文网络中相应模块对视网膜病变具有较高的捕获能力; M2 在 AMMF 模型中使用传统的交叉熵损失函数, 可以看出特异性比 AMMF 模型高出 2.94%, 但其他指标都远低于 AMMF 网络模型, 说明了组合的加权损失函数对视网膜病变分级精度上的有效性; 由 M3 和 AMMF 实验对比可知各项指标均较大幅度的提升, 说明多尺度增强模块可以有效地指导不同层次的特征, 以增大网络感受野提高网络捕捉特征信息的能力; 而 M4 和 AMMF 实验可以看出图形特征融合模块使得灵敏度降低了 0.45%, 但其他指标都有一定程度上的提升, 说明了图形特征融合模块能够有效地提升 DR 性能。上述实验整体表明, 本文算法具有

一定的有效性和合理性。

4.3 其他 DR 分级算法对比分析

为了充分说明本文算法对糖尿病视网膜病变分级的性能, 表 2 和表 3 分别给出了本文算法模型与其他 DR 分级算法模型在 IDRID 和 Eye-PACS 数据集上实验结果对比。

如表 2 所示, 本文在 IDRID 数据集上对比主流细粒度分类和近几年其他的糖尿病视网膜分级算法的分类结果, 其 QWK、Se 和 Sp 均低于本文所提算法。文献 [18] 采用深度神经网络提取眼底图像的特征, 并与机器学习方法结合而有效地提高了 DR 分级性能, 比本文算法准确率高 0.82%, 但 Se 和 Sp 指标都远低于本文算法。文献 [19] 先采用粗粒度网络去判定眼底图像是否发生病变, 再利用细粒度网络对粗粒度网络的预测结果进一步分级预测, 增强了病变特征并在

表 2 不同模型在 IDRID 数据集中的表现

Table 2 Performance of different models in the IDRID dataset

Methods	Backbone	QWK/%	Acc/%	Se/%	Sp/%
Bhardwaj et al. ^[18]	ResNet-50	—	79.46	82.85	76.98
Wu et al. ^[19]	ResNet-18	—	56.19	64.21	87.39
Song et al. ^[20]	ResNet-50	86.54	76.69	89.85	88.00
Shi et al. ^[21]	Efficientnet-b5	87.63	79.06	—	—
Ours	ResNeSt-50	90.38	78.64	95.65	91.17

表 3 不同模型在 Eye-PACS 数据集中的表现

Table 3 Performance of different models in the Eye-PACS dataset

Methods	Backbone	Acc/%	AUC/%	Se/%	Sp/%
Wu et al. ^[19]	ResNet-18	83.10	89.00	53.99	91.22
Song et al. ^[20]	ResNet-50	82.25	87.08	61.27	90.21
Kamothi et al. ^[22]	Modified VGG-16	77.60	—	77.00	—
Thota et al. ^[23]	VGGNet	74.00	80.00	80.00	65.00
Saxena et al. ^[24]	InceptionResNetV2	76.68	79.51	50.77	91.25
Ours	ResNeSt-50	84.41	90.36	77.85	97.05

一定程度上解决了患病样本与正常样本之间的分布不平衡问题, 但分级效果与本文算法仍有一定差距。文献 [20] 提出了一种特征增强、抑制和多样化信息融合的细粒度图像分类模型, 对其进行复现, 但分级结果还是弱于本文算法。文献 [21] 利用提出一种多阶段迁移的方法对特征学习和分类过程中进行独立训练, 该方法有效缓解了样本分布不均的影响, 进而提高分级的性能, 准确率高出本文算法 0.42%, 但最终结果整体上依然弱于本文所提算法。由分类结果可知, 本文所提算法在 IDRID 数据集上达到了不错的效果。为进一步清晰展示本文实验结果, 图 8 给出部分混淆矩阵对比结果, 本文算法混淆矩阵对角线样本分布更为集中, 且大多数误分类样本都被分成相邻类别, 而在文献 [20] 的混淆矩阵中样本分布效果没有本文算法好。

如表 3 所示, 近几年糖尿病视网膜病变分级算法在 Eye-PACS 数据集中与本文算法结果对比。可以看

出在分类的准确率上, 本文所提算法比其他算法具有更好的结果, 达到了 84.41%, 比其中效果最好的 Wu 等^[19] 高出了 1.31%。AUC 和 Sp 的结果也优于其他算法。本文算法的五分类平均 AUC 值高达 90.36%, 这表明了本文算法的分类效果最好。但是灵敏度较文献 [23] 低 2.15%, 而 Acc、AUC 和 Sp 比其分别高出 10.41%、10.36% 和 32.05%。这也表明本文算法在 Eye-PACS 数据集上依然具有一定的泛化能力。本文所提算法的受试者工作特征曲线下的面积 (AUC), 如图 9(a) 所示。

从图 9(a) 中可知, 轻度病变 (DR: 0) 情况最难分类, 其他病变等级分类效果都较为理想。根据 DR 各类别的受试者工作特征曲线下的面积可以计算出整体分类的平均 AUC 值为 90.36%。由图 9(b) 绘制的各指标对比的柱状图可以更加清晰地表明本文算法优于其他算法。

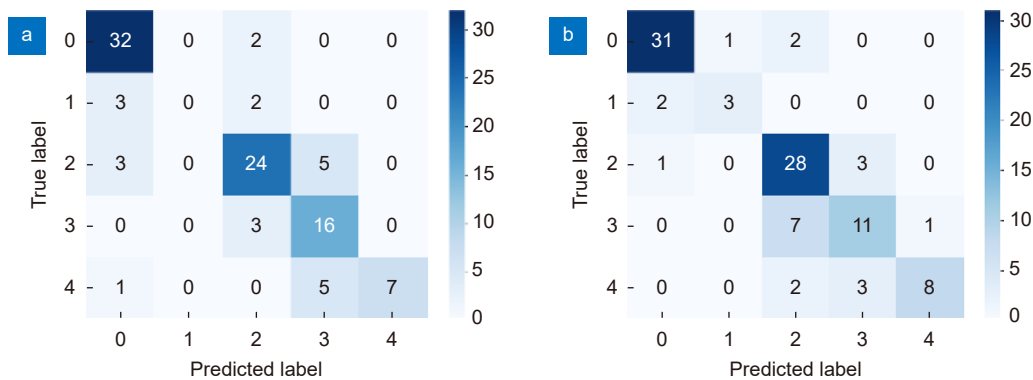


图 8 混淆矩阵。(a) 文献 [20]; (b) 本文方法
Fig. 8 Confusion matrix. (a) Ref. [20]; (b) Ours

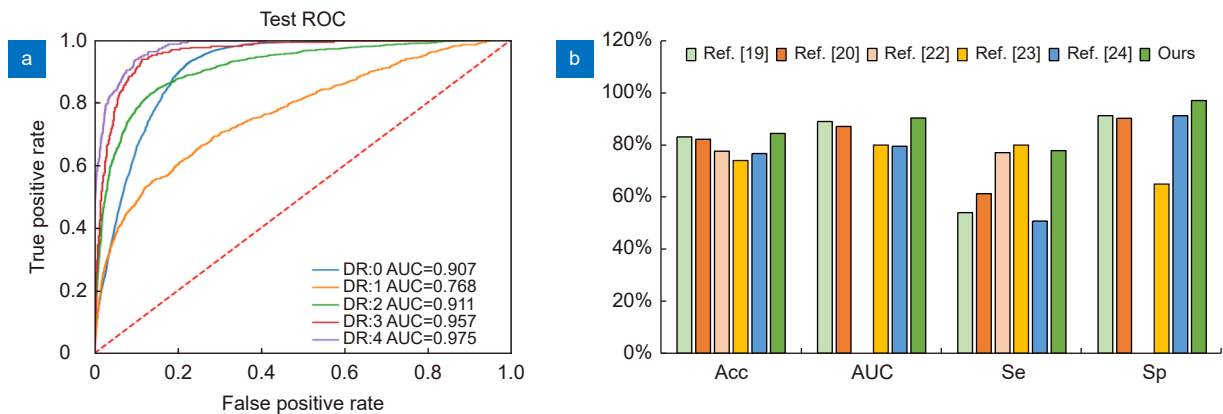


图 9 对比趋势图。(a) DR 各类别 AUC 值; (b) 各指标对比柱状图
Fig. 9 Comparison trend graph. (a) AUC value of each category of DR; (b) Comparison of each indicator bar chart

5 结 语

本文提出了一个新的注意力机制多特征融合的糖尿病视网膜病变分级模型, 其整体结构由 ResNeSt 主干网络、多尺度特征增强模块和图形特征融合模块组成, 并引入组合的加权损失函数来缓解样本分布不均衡和类间差异小难以区分的问题。实验结果证明, 本文所提算法能够有效地对 DR 各类间微小特征差异进行分级判断, 且与同类型病变分级网络进行了比较, 验证了本文所提算法模型优于其他网络模型, 在 DR 领域具有较大的临床应用价值。由于本文网络整体参数量较大, 部分网络还有精简的空间, 网络的参数大和复杂性增加了训练时间, 因此下一步研究重心将去对网络模型进行轻量化, 减少参数精简化来缩短训练时间, 进一步提升糖尿病视网膜病变分级的精度。

参考文献

- [1] 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.
- [2] Wei C M, Xu Y, Jiang X H, et al. Study on dermoscope image classification method based on PiT[J]. *J Optoelectron-Laser*, 2022, **33**(5): 505–512.
韦春苗, 徐岩, 蒋新辉, 等. 基于PiT的皮肤镜图像分类方法研究[J]. *光电子·激光*, 2022, **33**(5): 505–512.
- [3] Zhou K, Gu Z W, Liu W, et al. Multi-cell multi-task convolutional neural networks for diabetic retinopathy grading[C]//*Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2018: 2724–2727. <https://doi.org/10.1109/EMBC.2018.8512828>.
- [4] Du R Y, Chang D L, Bhunia A K, et al. Fine-grained visual classification via progressive multi-granularity training of jigsaw patches[C]//*16th European Conference on Computer Vision*, 2020: 153–168. https://doi.org/10.1007/978-3-030-58565-5_10.
- [5] Zheng W, Shen Q H, Ren J. Recognition and classification of diabetic retinopathy based on improved DR-Net algorithm[J]. *Acta Opt Sin*, 2021, **41**(22): 2210002.
郑雯, 沈琪浩, 任佳. 基于Improved DR-Net算法的糖尿病视网膜病变识别与分级[J]. *光学学报*, 2021, **41**(22): 2210002.
- [6] Gu T F, Hao P Y, Bai Z, et al. Diabetic retinopathy grading based on multi-channel attention[J]. *J Image Graphics*, 2021, **26**(7): 1726–1736.
顾婷菲, 郝鹏翼, 白琼, 等. 结合多通道注意力的糖尿病性视网膜病变分级[J]. *中国图象图形学报*, 2021, **26**(7): 1726–1736.
- [7] Zhang H, Wu C R, Zhang Z Y, et al. ResNeSt: split-attention networks[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2022: 2735–2745. <https://doi.org/10.1109/CVPRW56347.2022.00309>.
- [8] Tang X L, Du Y M, Liu Y W, et al. Image recognition with conditional deep convolutional generative adversarial networks[J]. *Acta Autom Sin*, 2018, **44**(5): 855–864.
唐贤伦, 杜一铭, 刘雨微, 等. 基于条件深度卷积生成对抗网络的图像识别方法[J]. *自动化学报*, 2018, **44**(5): 855–864.
- [9] Shao Q B, Gong L J, Ma K, et al. Attentive CT lesion detection using deep pyramid inference with multi-scale booster[C]//*22nd International Conference on Medical Image Computing and Computer Assisted Intervention*, 2019: 301–309. https://doi.org/10.1007/978-3-030-32226-7_34.
- [10] Chen L C, Papandreou G, Kokkinos I, et al. DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs[J]. *IEEE Trans Pattern Anal Mach Intell*, 2018, **40**(4): 834–848.
- [11] Wang K, Zhang X H, Lu Y T, et al. CGRNet: contour-guided graph reasoning network for ambiguous biomedical image segmentation[J]. *Biomed Signal Process Control*, 2022, **75**: 103621.
- [12] Kipf T N, Welling M. Semi-supervised classification with graph convolutional networks[Z]. arXiv: 1609.02907, 2016. <https://arxiv.org/abs/1609.02907v3>.
- [13] Li Q M, Han Z C, Wu X M. Deeper insights into graph convolutional networks for semi-supervised learning[C]//*32nd AAAI Conference on Artificial Intelligence*, 2018: 3538–3545.
- [14] Li Y, Gupta A. Beyond grids: learning graph representations for visual recognition[C]//*Proceedings of the 32nd International Conference on Neural Information Processing Systems*, 2018: 9245–9255. <https://doi.org/10.5555/3327546.3327596>.
- [15] Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection[C]//*Proceedings of the IEEE International Conference on Computer Vision*, 2017: 2999–3007. <https://doi.org/10.1109/ICCV.2017.324>.
- [16] Shang Q F, Shen W, Shuai S Y. Semantic segmentation of high resolution remote sensing image based on deep learning[J]. *Comput Syst Appl*, 2020, **29**(7): 180–185.
尚群峰, 沈炜, 帅世渊. 基于深度学习高分辨率遥感影像语义分割[J]. *计算机系统应用*, 2020, **29**(7): 180–185.
- [17] Cuadros J, Bresnick G. EyePACS: an adaptable telemedicine system for diabetic retinopathy screening[J]. *J Diabetes Sci Technol*, 2009, **3**(3): 509–516.
- [18] Bhardwaj C, Jain S, Sood M. Transfer learning based robust automatic detection system for diabetic retinopathy grading[J]. *Neural Comput Appl*, 2021, **33**(20): 13999–14019.
- [19] Wu Z, Shi G L, Chen Y, et al. Coarse-to-fine classification for diabetic retinopathy grading using convolutional neural network[J]. *Artif Intell Med*, 2020, **108**: 101936.
- [20] Song J W, Yang R Y. Feature boosting, suppression, and diversification for fine-grained visual classification[C]//*2021 International Joint Conference on Neural Networks (IJCNN)*, 2021: 1–8. <https://doi.org/10.1109/IJCNN52387.2021.9534004>.
- [21] Shi L, Zhang J X. Few-shot learning based on multi-stage transfer and class-balanced loss for diabetic retinopathy grading[Z]. arXiv: 2109.11806, 2021. <https://arxiv.org/abs/2109.11806>.
- [22] Kamothi N, Thakur R. Detection of diabetic retinopathy using transfer learning approach[J]. *BIOINFOLET*, 2021, **18**(1a): 51–54.
- [23] Thota N B, Reddy D U. Improving the accuracy of diabetic retinopathy severity classification with transfer learning[C]//*2020 IEEE 63rd International Midwest Symposium*

on *Circuits and Systems (MWSCAS)*, 2020: 1003–1006.
<https://doi.org/10.1109/MWSCAS48704.2020.9184473>.

[24] Saxena G, Verma D K, Paraye A, et al. Improved and robust

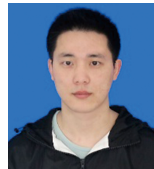
deep learning agent for preliminary detection of diabetic retinopathy using public datasets[J]. *Intell Based Med*, 2020, 3–4: 100022.

作者简介



【通信作者】梁礼明 (1967-), 男, 江西吉安人, 教授, 2005 年于江西理工大学获得硕士学位, 主要从事机器学习和医学影像方面的研究。获得国家教学成果二等奖一项, 江西省教学成果一等奖一项、二等奖一项、三等奖二项。公开发表学术论文百余篇, 其中被 SCI、EI、ISTP 收录论文二十余篇, 全国中文核心期刊四十余篇; 获得中国发明专利六项 (排名第一)、出版研究生教材一部。

E-mail: 9119890012@jxust.edu.cn

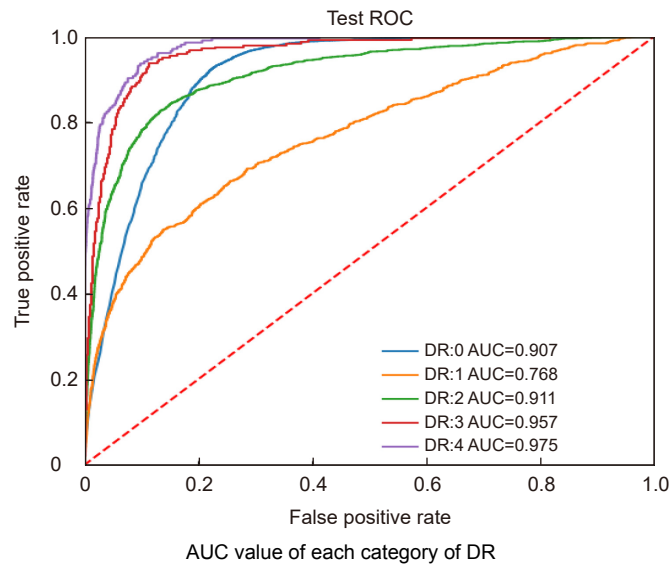


董信 (1997-), 男, 硕士研究生, 目前就读于江西理工大学, 主要研究方向是医学图像病变分级领域。

E-mail: 2228317846@qq.com

Classification algorithm of retinopathy based on attention mechanism and multi feature fusion

Liang Liming*, Dong Xin, Li Renjie, He Anjun



Overview: Diabetic Retinopathy (DR) is a common acute stage of diabetes, which can cause abnormal retinal visual function; if not detected early and treated, it can lead to blindness. In recent years, the research on the intelligent diagnosis of DR classification has been a hot topic in the field of medical image processing. With the high-quality development of deep learning, the technology of intelligent diagnosis using deep neural network algorithms for DR image processing has been widely used. However, there are still two limitations in the current DR grading intelligent diagnosis process: ① In DR images, the features of microscopic lesions such as microaneurysms, hard exudates, and hemorrhages have little difference with the surrounding environment of the retina, and the feature extraction is insufficient; ② The distribution of various samples in the public datasets in the medical field is uneven.

In view of the difficulty in identifying the lesion area in retinal fundus images and the low grading efficiency, this paper proposes an algorithm based on multi-feature fusion of attention mechanism to diagnose and grade DR. The overall structure of the algorithm is mainly composed of ResNeSt backbone network, multi-scale feature enhancement module (MSFB), and graph feature fusion module (GFFM), and a combined weighted loss function is introduced to alleviate the problem of unbalanced sample distribution and indistinguishable differences between classes. Firstly, feature enhancement is performed on the pathological area of retinopathy image through MSFB, improving classification accuracy, and optimizing model performance. Then, the graphic feature fusion module is used to perform information fusion on the local features after the feature enhancement of the backbone output. Finally, a weighted loss function combining center loss and focal loss is used to further improve the classification effect. Although it shows good performance on two datasets, this paper also has some shortcomings. For example, the overall number of parameters is slightly larger, which makes the network more complicated and increases the training and testing time. The average time per round is 46 seconds on the IDRID dataset and 48 minutes per round on the Eye-PACS dataset.

In IDRID dataset, the sensitivity and specificity were 95.65% and 91.17%, respectively, and the quadratic weighted agreement test coefficient was 90.38%. In the Kaggle competition dataset, the accuracy rate is 84.41%, and the area under the receiver operating characteristic curve was 90.36%. The experimental results show that the algorithm in this paper has certain application value in the field of DR. In view of the shortcomings of the above model, the next key task is to streamline the network model and further improve the model performance as much as possible.

Liang L M, Dong X, Li R J, et al. Classification algorithm of retinopathy based on attention mechanism and multi feature fusion[J]. *Opto-Electron Eng*, 2023, 50(1): 220199; DOI: 10.12086/oe.2023.220199

Foundation item: National Natural Science Foundation of China (51365017,6146301), and Natural Science Foundation of Jiangxi Province (20192BAB205084)

School of Electrical Engineering and Automation, Jiangxi University of Science and Technology, Ganzhou, Jiangxi 341000, China

* E-mail: 9119890012@jxust.edu.cn