

# 光电工程

## Opto-Electronic Engineering

中文核心期刊 中国科技核心期刊  
Scopus CSCD

### 融合坐标感知与混合提取的视网膜病变分级算法

梁礼明, 金家新, 冯耀, 卢宝贺

#### 引用本文:

梁礼明, 金家新, 冯耀, 等. 融合坐标感知与混合提取的视网膜病变分级算法[J]. 光电工程, 2024, 51(1): 230276.

Liang L M, Jin J X, Feng Y, et al. Retinal lesions graded algorithm that integrates coordinate perception and hybrid extraction[J]. *Opto-Electron Eng*, 2024, 51(1): 230276.

<https://doi.org/10.12086/oe.2024.230276>

收稿日期: 2023-11-15; 修改日期: 2024-01-04; 录用日期: 2024-01-04

## 相关论文

### 自适应特征融合级联Transformer视网膜血管分割算法

梁礼明, 卢宝贺, 龙鹏威, 阳渊

光电工程 2023, 50(10): 230161 doi: 10.12086/oe.2023.230161

### 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

光电工程 2023, 50(1): 220199 doi: 10.12086/oe.2023.220199

### 引入双编码器模型的OCT视网膜图像分割

陈明惠, 王腾, 袁媛, 柯舒婷

光电工程 2023, 50(10): 230146 doi: 10.12086/oe.2023.230146

### 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

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.2024.230276

## 融合坐标感知与混合提取的视网膜病变分级算法

梁礼明, 金家新\*, 冯耀, 卢宝贺

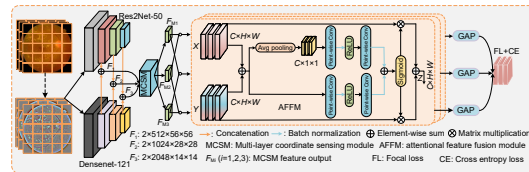
江西理工大学电气工程与自动化学院, 江西 赣州 341000

**摘要:** 针对糖尿病视网膜病变中存在样本分布不平衡和病灶区域特征识别困难等问题, 提出一种融合坐标感知与混合提取的视网膜病变分级算法。该算法首先对视网膜输入图像进行裁剪、高斯滤波等预处理操作, 以增强图像病变前景与噪声背景之间的差异度; 然后由 Res2Net-50 和 Densenet-121 骨干网络组成的混合双模型将增强后的图像进行特征逐层提取, 实现多尺度特征纹理的充分捕捉; 再在混合双模型连接处融入多层坐标感知模块和注意力特征融合模块, 达到剔除聚焦病灶特征干扰的目的, 实现不同病灶语义间的权重重塑; 最后利用组合损失函数缓解样本分布不均匀问题, 进一步监督模型的训练与测试。该文算法在 IDRID 和 APTOS 2019 数据集上进行实验, 二次加权系数分别为 88.76% 和 90.29%; 准确率分别为 81.55% 和 84.42%, 为视网膜病变分级智能辅助诊断提供了新窗口。

**关键词:** 视网膜病变分级; 图像预处理; 混合双模型; 多层坐标感知; 特征融合

**中图分类号:** TP391

**文献标志码:** A



梁礼明, 金家新, 冯耀, 等. 融合坐标感知与混合提取的视网膜病变分级算法 [J]. 光电工程, 2024, 51(1): 230276

Liang L M, Jin J X, Feng Y, et al. Retinal lesions graded algorithm that integrates coordinate perception and hybrid extraction[J]. *Opto-Electron Eng*, 2024, 51(1): 230276

## Retinal lesions graded algorithm that integrates coordinate perception and hybrid extraction

Liang Liming, Jin Jiabin\*, Feng Yao, Lu Baohe

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

**Abstract:** Aiming at the problems of unbalanced sample distribution and difficulty identification of the lesion area in diabetic retinopathy, we propose a retinal lesions grading algorithm that integrates coordinate perception and hybrid extraction. This algorithm first processes the retinal input image and the Gaussian filtering to enhance the difference between the image lesions and the background of the noise, and then the hybrid dual models composed of the backbone network of Res2Net-50 and Densenet-121 will be enhanced. The image is extracted layer by layer to achieve the full capture of the multi-scale feature texture, then the multi-layer coordinate perception module and the attention characteristics fusion module are integrated at the mixed dual model connection to achieve the purpose of eliminating the characteristics of the lesions and the realization of different lesions. The weight of semantics is reshaped, finally uses the combined loss function to relieve the uneven distribution of samples to further supervise the training and test of the model. This article is experimented on the IDRID and Aptos 2019 data

收稿日期: 2023-11-15; 修回日期: 2024-01-04; 录用日期: 2024-01-04

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: 金家新, 1136868701@qq.com。

版权所有©2024 中国科学院光电技术研究所

sets, with the secondary weighted coefficients of 88.76% and 90.29%, respectively. Accuracy rates were 81.55% and 84.42%, which provides a new window for the diagnosis of retinopathy grades and intelligent auxiliary diagnosis.

**Keywords:** retinopathy grading; image preprocessing; hybrid dual model; multi-layer coordinate perception; feature fusion

## 1 引言

糖尿病视网膜病变 (Diabetic retinopathy, DR) 是一种由糖尿病所引发的眼部疾病, 患者长期处在高血糖环境下, 非常容易损伤视网膜, 如未能及时发现疾病并进行治疗则易造成失明, 如及时治疗则有很大概率能保住视力。在临床实践中, 糖尿病视网膜病变诊断, 这一过程非常耗时需要训练有素和经验丰富的眼科医生来进行判别, 诊断的精度需要经济、时间成本和相应的医疗资源。此外, 当前患糖尿病的人群正在逐年递增, 而具有专业眼科经验医生人数严重不足。因此将计算机用于 DR 智能诊断, 提高诊断效率和准确度的需求显得尤为迫切<sup>[1-2]</sup>。

近年来, 深度学习技术逐渐成为主流<sup>[3]</sup>, 使得卷积神经网络 (convolutional neural network, CNN) 在 DR 智能诊断方面得到广泛应用, 但对于当前的视网膜病变分级仍存在不足之处: ①由于本文研究对象视网膜本身情况较为复杂且病灶区域与微血管瘤等周围环境差异较少, 病变特征提取困难。②受糖尿病病变阶段和可治愈性的影响, 导致患者处在不同病变阶段的数量不一致, 使其数据集的样本分布不平衡。针对上述问题, 国内外众多学者进行大量研究, 如 He 等<sup>[4]</sup>构建一种 CABNet 网络, 由新类别的注意力 (category attention block, CAB) 探索 DR 各类别的区域特征以捕捉更细小的病变信息, 从而缓解特征提取不足情况。Ashwini 等<sup>[5]</sup>设计一种基于离散小波变换多分辨率的特征提取方法, 先对图像使用自适应直方图均衡化预处理方式, 然后利用过采样去确保各类别样本数量相等。Zhou 等<sup>[6]</sup>采用一种多任务学习策略, 该策略通过分类和回归损失来预测标签, 同时又通过均方差损失函数去减少预测值与实际值的差异, 但也存在训练时间较长的问题。Shi 等<sup>[7]</sup>使用迁移学习技术学习多个数据集的特征信息, 从而使模型获得充足的特征信息, 强 DR 分级效率, 同时提出一种类平衡损失函数, 缓解数据集样本不平衡问题。Shaik 等<sup>[8]</sup>利用预训练的卷积学习视网膜的图像空间表征, 此外还利用

卷积自编码器和神经支持向量机进行端对端训练, 从而获得基于类别病灶特征的潜在注意力特征。张文轩等<sup>[9]</sup>通过多分支注意力增强机制的卷积神经网络, 利用局部响应图来反映关注目标局部位置, 能够反映存在不同位置的权重, 将会使局部特征较为敏感, 此外注意力机制还可以对所需位置进行针对性强化, 使该网络能够准确定位病灶区域。程小辉等<sup>[10]</sup>构建的一种注意力网络 ME-ANet, 通过模型融合进行特征提取, 应用迁移学习的方式训练模型从而减少训练时间, 但模型的结构较为复杂, 存在可优化空间。针对视网膜病变分级当前面临的技术局限问题, 本文提出一种融合坐标感知与混合提取的视网膜病变分级网络, 其改进工作主要有: ①采用 Res2Net-50 和 Densenet-121 作为骨干网络共同构成混合双模型 (mixed dual model, MDM) 进行逐层特征提取图像的语义信息和空间细节信息, 并利用两个模型不同的特征提取方式进行优势互补, 提升算法的鲁棒性和泛化性。②多层坐标感知模块 (Multi-layer coordinate sensing module, MCSM) 通过全局和局部两种方式获得深层语义信息与细节边缘信息, 同时还利用残差结构捕捉浅层语义信息, 进一步完善病变区域特征信息。③利用焦点损失 (focal loss, FL) 和交叉熵损失函数组成的组合损失函数来抑制样本不平衡而造成的模型性能问题, 弱化因样本引起视网膜病变分级准确率不高, 改善 DR 分级效果。

## 2 网络整体框架

视网膜病变分级相对于其他细粒度图像分类问题较为困难, 因公开的数据集中存在样本分布不均匀、类间差异小和图像噪声干扰较大的问题。本文设计出一种融合坐标感知与混合提取的视网膜病变分级, 其主要由图像预处理、混合双模型 (MDM)、多层坐标感知模块 (MCSM)、注意力特征融合模块 (attentional feature fusion module, AFFM)、组合损失函数共同组成整体网络, 其中 GAP 与 Avg pooling 为全局平均池

化、ReLU 和 Sigmoid 为激活函数、Point-wise Conv 为逐点卷积, 如图 1 所示。在进行特征提取之前对本文所采用的两个数据集进行预处理改变其分辨率的大小, 首先通过混合双模型进行特征提取获得丰富的语义信息, 选择最后三层的尺度特征图大小为  $512 \times 56 \times 56$ 、 $1024 \times 28 \times 28$  和  $2048 \times 14 \times 14$ ; 其次利用 MCSM 对所输出的多种尺度信息进一步挖掘潜在特征, 提高特征识别的效率, 增强图片处理速率; 再次通过卷积操作使通道数一致化, 分别为  $512 \times 56 \times 56$ 、 $512 \times 28 \times 28$  和  $512 \times 14 \times 14$ , 进一步利用 AFFM 将不同尺度特征图  $X$  与  $Y$  特征融合处理, 其结果特征图为  $Z$ , 输出多尺度特征图共同构成完整的图像信息, 增强病灶区域特征的识别效率, 加快模型训练速度; 最后通过组合损失函数进一步缓解类间差异情况。

### 2.1 混合双模型

针对视网膜病变分级中存在病灶区域与背景区域特征相差较小而造成的特征提取困难问题, 本文

提出的混合双模型, 该模型主要由 Res2Net-50<sup>[11]</sup> 和 Densenet-121<sup>[12]</sup> 组成用于提取病灶信息, 相较于单个模型具有较高的分类精度。Res2Net-50 采用跳跃连接方式, 通过残差块有利于输入信息和处理后信息的再次处理, 从而降低需要学习的目标难度且能够捕捉到图像中细节和上下文信息、减少模型的冗余。Densenet-121 能利用密集连接将特征图进行拼接促进信息的流动和减轻梯度消失问题, 同时还能防止图像信息的丢失。残差网络和密集网络共同组成的混合模型能增强特征提取能力并进一步改善算法模型的性能, 将获取更全面的图像全局和局部信息, 其模型的结构如图 2 所示。Res2Net-50 和 Densenet-121 拥有相同的阶段, 但两者的不同阶段所蕴含特征信息不一致, 其每个阶段的输入信息表达式分别为:

$$X_{n(R)} = H_1(X_{n-1}) + X_{n-1}, \quad (1)$$

$$X_{n(D)} = H_1([X_0, X_1, \dots, X_{n-1}]), \quad (2)$$

其中:  $n \in (1, 2, 3, 4)$ ,  $X_{n(R)}$  是 Res2Net-50 输入特征,  $X_{n(D)}$  是 Densenet-121 输入特征,  $H_1(\bullet)$  表示非线性

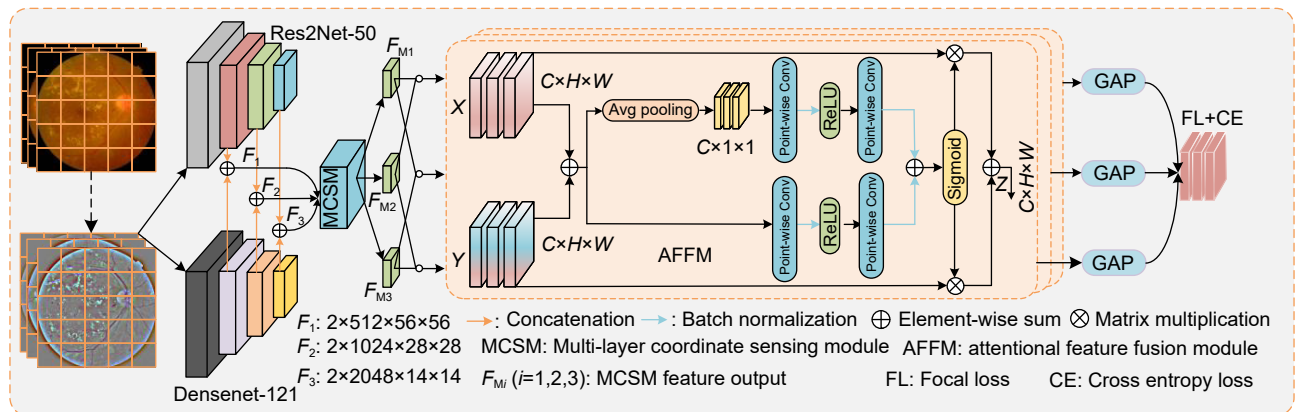


图 1 算法整体框架

Fig. 1 The overall framework of the algorithm

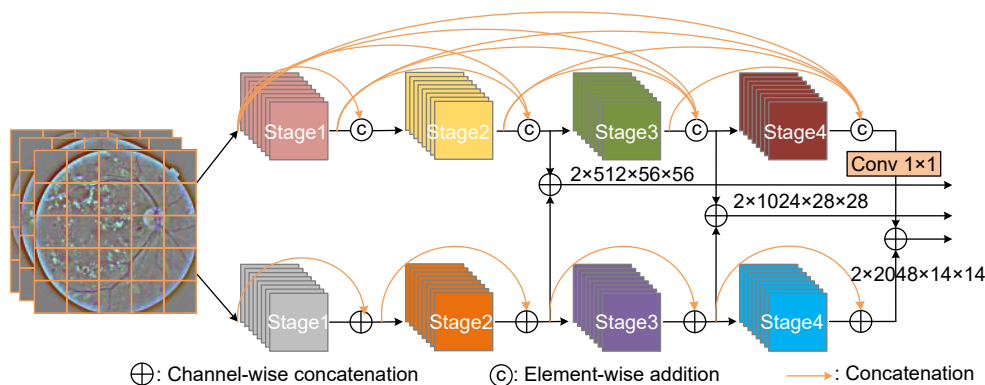


图 2 混合双模型

Fig. 2 The hybrid dual model

转化函数。

本文所需提取最后三个阶段特征图大小一致, 但其通道数还是存在差别, 为使两者的特征信息进行融合处理, 将 Densenet-121 的最后一层的通道数通过 Conv 1×1 从 1024 转为 2048, 最后输出三个不同的特征图大小。

### 2.2 多层坐标感知模块

由于视网膜图像周围存在的微血管瘤、黄白色硬性渗出或有少量的出血斑等细小病变特征, 与正常区域对比差异不明显, 单层的特征学习对信息特征提取可能存在不足之处, 导致视网膜病变分级的效果不佳。为捕获更完整的信息, 得到较高的视网膜分级准确率, 从而设计一种多层坐标感知模块 (MCSM), 使模型更准确地定位并识别病灶区域。MCSM 主要由坐标注意力模块<sup>[13]</sup>(coordinate attention module, CAM)、多层感知机 (multilayer perceptron, MLP) 和残差连接结构三部分组成, 其整体结构如图 3 所示。为加快模型训练的收敛速度而加入 BN (batch normalization), 而后 CAM 是一种嵌入位置信息注意力, 能够获得较大的区域信息且运行参数较少, 同时为避免 2D 全局平均池化造成的信息损失, 设计两个 1D 全局平均池化分别对水平方向和垂直方向的输入特征处理聚合生成两个独立的方向特征图  $f$ , 水平方向的池化核为  $(H, 1)$ , 垂直方向的池化核是  $(1, W)$ ,  $C$  维特征输出分别为:

$$Z_c^h(h) = \frac{1}{W} \sum_{0 \leq i < W} x_c(h, i), \quad (3)$$

$$Z_c^w(w) = \frac{1}{H} \sum_{0 \leq j < H} x_c(j, w), \quad (4)$$

$$f = \delta(F_1([z^h, z^w])), \quad (5)$$

其中:  $F_1$  是变换函数, 特征图  $f \in R^{C/r \times (H \times W)}$ ,  $r$  为缩减比,  $\delta$  是一个非线性激活函数。

同时将存在特定信息特征图编码生成两个注意力

图分别为  $g^h$ 、 $g^w$ , 从而获取两个不同的空间信息, 其中一个空间方向能够捕获与特征图之间的跨通道信息, 则另一个空间方向将获得准确的位置信息, 使注意力图能够同时获取方向感知和位置坐标的深层语义信息, 同时残差连接又能从输入特征获取浅层语义信息, 最终通过融合进一步处理生成具有丰富语义信息的特征注意力图, 即表达式为:

$$g^h = \sigma(F_h(f^h)), \quad (6)$$

$$g^w = \sigma(F_w(f^w)), \quad (7)$$

$$y_c(i, j) = x_c(i, j) \times g_c^h(i) \times g_c^w(j), \quad (8)$$

其中:  $f^h \in R^{C/r \times H}$ ,  $f^w \in R^{C/r \times W}$ ,  $\sigma$  是 sigmoid,  $x_c(i, j)$  输入特征,  $y_c(i, j)$  输出特征,  $F_h$ 、 $F_w$  将  $f^h$ 、 $f^w$  的通道张量转换成与输入值一致。

为排除与病灶不相关的干扰信息, 利用第一个全连接层与激活函数进行线性变换和非线性变换, 其后 dropout 层则丢弃部分隐藏特征信息, 而第二个全连接层将隐藏特征映射回输入特征, 最后通过一个残差连接将弱化非病变特征信息和输入特征信息进行聚合处理, 进一步聚焦病灶区域特征。

### 2.3 注意力特征融合模块

为捕获和聚合不同尺度特征图的局部细节信息和全局信息且能自适应不同大小和位置的病灶区域识别。本文引入了注意力特征融合模块 (AFFM)<sup>[14]</sup> 融合多层坐标感知模块输出的多种尺度信息, 该模块主要由全局和局部两部分组成, 其结构如图 1 中 AFFM 所示。AFFM 中的输入特征  $X$  和  $Y$  来自两个不同的特征图。在全局方向上利用全局平均池化对空间池进行调整, 以实现多个尺度的通道关注。此外, 为控制整体网络的运行参数, 将注意力添加到上下文当中, 随后通过逐点卷积把上下文聚合在一起。在局部方向则只能实现特定通道的关注, 同时利用逐点卷积将上下文信息联会贯通。将经过全局和局部的输入特征通过聚合处

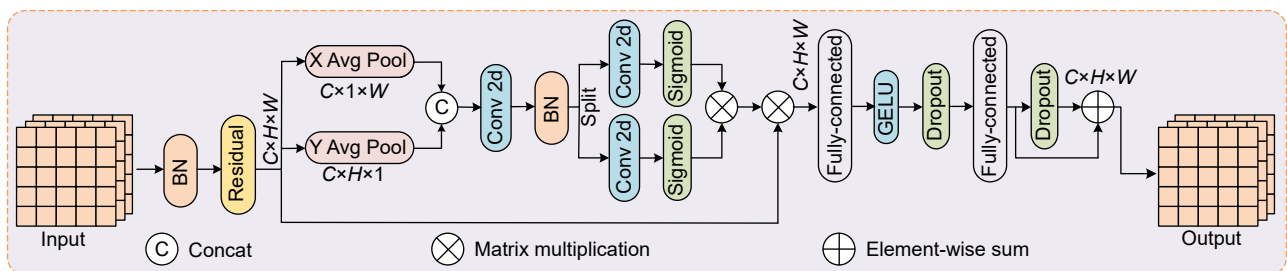


图 3 多层坐标感知模块

Fig. 3 The multi-layer coordinate perception module

理, 最后经过 sigmoid 函数对权重进行再次分配, 并利用跳跃连接与初始输入特征进行融合处理, 其表达式为:

$$Z = M(X \oplus Y) \otimes X + (1 - M(X \oplus Y)) \otimes Y, \quad (9)$$

其中:  $Z \in R^{C \times H \times W}$  是输出特征,  $X, Y \in R^{C \times H \times W}$  是输入特征,  $\oplus$  初始特征积分,  $M$  为 sigmoid 函数,  $X \oplus Y$  是输入 sigmoid 函数的值。

## 2.4 组合损失函数

公开的 DR 数据集是来自糖尿病患者的视网膜图像, 根据患者处在糖尿病的阶段不同可分为五个等级, 而糖尿病阶段不同也造成可被治愈程度不一致, 导致各个类别存在较大的差异。为解决各类别之间样本不均匀情况, 本文提出的组合损失函数由焦点损失函数<sup>[15]</sup>和交叉熵损失函数<sup>[16]</sup>共同构成, 焦点损失函数能够通过一个动态缩放因子, 在易被识别的正确样本时缩放因子就会下降到一个较低值, 相反遇到较难辩别的样本时就会聚焦此样本, 同时影响因子会上升到一个较高值, 两种损失函数的表达式分别为:

$$L_{FL} = -(1 - p^t)^v \ln(p^t), \quad (10)$$

$$L_{CE} = -\frac{1}{N} \sum_{i=1}^{N-1} \sum_{k=0}^{K-1} y_{i,k} \ln p_{i,k}, \quad (11)$$

其中:  $v$  是加权误差调制系数,  $p^t$  是标签预测概率,  $t$  表示样本类别,  $y_{i,k}$  表示第  $i$  个样本的真实标签为  $k$ , 共有  $k$  个标签值的  $N$  个样本,  $p_{i,k}$  表示第  $i$  个样本预测为第  $k$  个标签值的概率。最后, 将所使用的焦点损失和交叉熵损失函数进行加权求和, 其计算式如下:

$$L_{loss} = L_{FL} + \alpha L_{CE}, \quad (12)$$

其中,  $\alpha$  为超参数设置为 0.002。

## 3 实验

### 3.1 实验环境和参数配置

本文所使用设备信息 CPU 为 12th Gen Intel(R) Core(TM) i7-12700H, GPU 为 NVIDIA RTX4060, 16 G 运行内存, 操作系统是 Windows11, 基于 Python3.9 框架建模, Pycharm 的仿真平台。其学习率设为 0.002、batch-size 为 4、epoch 为 150 轮。在 IDRID 数据集实验中平均一轮的训练时间为 45 秒/轮, 测试时间为 11 秒/轮。在 APTOS 2019 数据集实验中平均一轮的训练时间为 2 分 30 秒/轮, 测试时间为 15 秒/轮。

### 3.2 数据的来源和处理

本文使用“印度糖尿病视网膜病变图像 (Indian

DR image dataset, IDRID) 数据集”, 该数据集拥有不同等级的眼底图像一共 516 张, 其图像的分辨率为  $4288 \times 4288$ , 将分辨率进行适当的调整, 改为  $512 \times 512$ 。由于当前数据集所含有的图像数量较少, 其实验结果对于本文的模型解释性不够充分, 因此还采用“亚太远程眼科学会 2019 年失明检测 (Asia Pacific Tele-Ophthalmology Society 2019 Blindness Detection, APTOS 2019 BD)”数据集来增强模型的解释性, 其数据集含有 3662 张眼底图像, 图像分辨率为  $224 \times 224$ , 并通过上采样操作转变成  $256 \times 256$ 。对本文 IDRID 和 APTOS 2019 数据集的眼底图像将分为训练集和测试集, 统一按 8:2 进行划分。根据当前国际医疗领域的评判标准将糖尿病视网膜病变按照不同等级可划分为 5 类<sup>[17]</sup>, 即为无糖尿病 (DR: 0)、轻度非增殖性 (DR: 1)、中度非增殖性 (DR: 2)、重度非增殖性 (DR: 3) 和增殖性 (DR: 4)。在本文所使用的两个数据集中, 存在病变区域和图像背景对比度差异不明显的情况, 因此需要进行预处理操作, 先对原始图像 RGB 通道统一为灰度化操作能在一定程度上加快运算速度, 后利用高斯滤波和加权融合处理去除图像噪声, 进一步增强病变区域和图像背景的对比度, 突出对比的差异性。其表达式分别为:

$$G_{\sigma} = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{x^2+y^2}{2\sigma^2}}, \quad (13)$$

$$I_d = \alpha I + \beta G_{\sigma} * I + \varepsilon, \quad (14)$$

其中:  $\alpha$ 、 $\beta$  和  $\varepsilon$  是加权系数, 本文分别设为 4、-4 和 128;  $G_{\sigma}$  是标准差;  $\sigma$  是二维高斯核;  $*$  是滤波操作;  $I_d$  是加权融合操作后的图像。此外, IDRID 与 APTOS 2019 数据集相比较样本数量较少, 可能会出现过拟合现象, 所以对该数据集进行水平、垂直翻转和镜像翻转、几何变换等操作来数据增强避免出现拟合。其预处理图像前后如图 4 所示。

### 3.3 评价指标

为体现 IDRID 和 APTOS 2019 数据集在算法上的表现效果, 同时能够对其它算法进行比较, 本文在 IDRID 数据集使用灵敏度 (sensitivity, Se)、特异性 (specificity, Sp)、准确率 (accuracy, Acc) 和二次加权 kappa 系数 (quadratic weighted kappa, QWK) 作为评价指标。此外在 APTOS 2019 数据集还利用 ROC 曲线下方的面积 (area under curve, AUC) 和召回率 (recall, Re) 对其进行评估, 其计算式分别为:

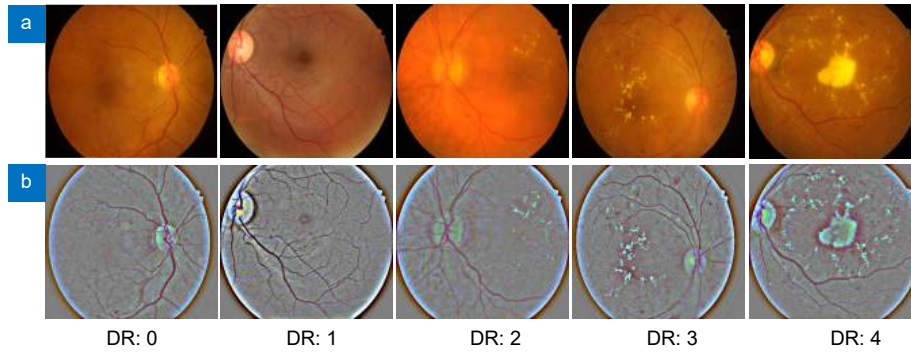


图 4 不同 DR 分级图像预处理对比。(a) 原始图像; (b) 预处理后图像

Fig. 4 Different DR hierarchical images pre-processing comparison. (a) Primitive images; (b) Pre-processing images

$$Acc = \frac{TP+TN}{TP+TN+FP+FN}, \quad (15)$$

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

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

$$W_{i,j} = 1 - \frac{(i-j)^2}{(N-1)^2} i j, \quad (18)$$

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

$$Re = \frac{TP}{TP+FN}, \quad (20)$$

其中:  $TP$  为样本和模型识别结果都为正类,  $TN$  的表示则恰好与之相反,  $FN$  表示样本为负类而模型结果为正类,  $FP$  表示样本为正类而模型结果负类,  $N$  为总类别数,  $W_{i,j}$  表示  $i$  类和  $j$  类的惩罚权重,  $Q_{i,j}$  为第  $i$  类判别为  $j$  类的数量,  $E_{i,j}$  为第  $i$  类的总数  $\times$  第  $j$  类的总数除以总数, 用于平衡分类效果 kappa 系数是检验

一致性的指标, 其系数在  $[-1,1]$  变化通常大于 0, 值越高则表示一致性越高。

### 3.4 实验结果分析

本文算法在 IDRID 和 APTOS 2019 数据集训练过程损失值变化趋势如图 5 所示, IDRID 数据集训练损失值维持在 0.4 左右, APTOS 2019 数据集损失值维持在 0.1 左右, 波动幅度较小, 表明网络已趋于收敛。该算法的两个数据集都经过 150 轮的训练过程, 在 130 轮的时候将达到最高峰, 且通过观察损失函数曲线在 130 轮的时候训练和测试损失曲线将达到平稳, 并且两者之间相差较小达到收敛状态。

#### 3.4.1 热图可视化

为验证本文算法对视网膜图像的学习效果, 利用原始图像通过本文模型生成网络特征热图<sup>[18]</sup>, 通过观察网络特征热图, 发现图像病变区域色彩差异较大效果明显, 如图 6 所示。在图 6(b) 的图像中其蓝色代表低概率病变区域, 绿色代表存在可能病变区域, 红色代表高概率病变区域。其中绿色方框表示原始细微处的病灶区域, 橘色方框为经过模型学习后的效果, 通

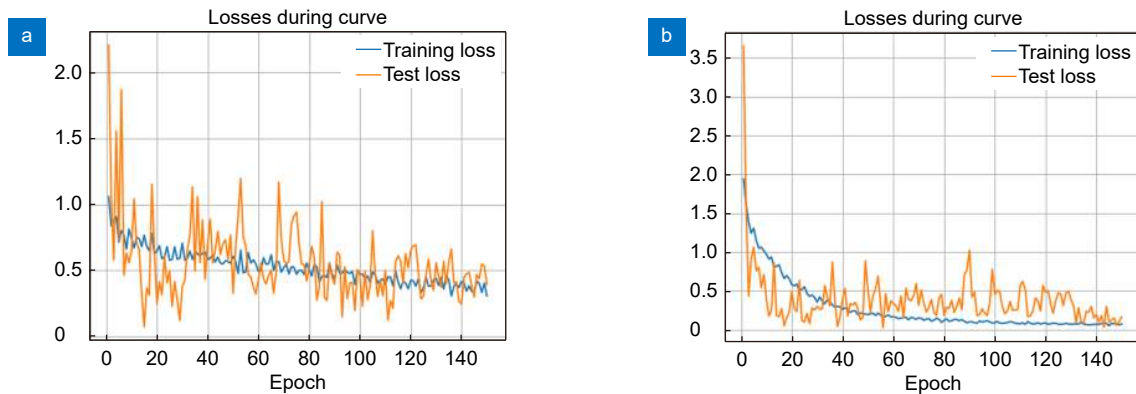


图 5 本文算法在 (a) IDRID 数据集和 (b) APTOS 2019 数据集上的训练损失曲线

Fig. 5 The training loss curves of the proposed algorithm on (a) the IDRID dataset and (b) the APTOS 2019 dataset

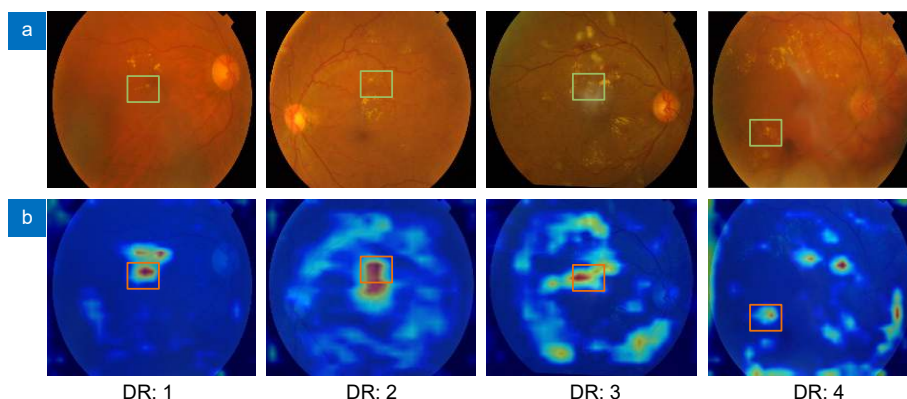


图 6 网络特征热图。(a) 初始图像; (b) 热力图像

Fig. 6 Network feature hot pictures. (a) Initial images; (b) Thermal images

过观察 DR: 1-4 能够发现热力图高亮区域越来越多且分布较广, 既有细微之处的病变也有显著的病灶区域, 图像相较于初始图像对病灶区域的对比差异度明显, 从而表明该算法学习效果较好。

### 3.4.2 消融实验

为探究本文算法中各模块和组合损失函数的有效性, 通过控制变量法在 IDRID 数据集进行消融实验。实验结果如表一所示, 表中 M1: 在整个算法模型中仅去除 MCSM; M2: 在整个算法模型中仅去除 AFFM; M3: 基于 Res2Net-50 的单个模型的网络; M4: 基于 Densenet-121 的单个模型的网络; M5: 基于混合模型的焦点损失函数网络; M6: 完整的算法模型网络。

从表 1 中 M1、M2 和 M6 的数据对比可以得知, 各项的评价指标均有较大的提升, 表明多层坐标感知模块和注意力特征融合模块的加入, 可以进一步增强 DR 分级效果; M3 和 M6 对比结果可知, 混合模型相较于单个 Res2Net-50 模型在准确率、特异性和二次加权 kappa 系数上有较大的提升, 分别为 1.94%、5.88% 和 2.63%; M4 与 M6 对比可知, 对比 Densenet-121 模型在准确率和灵敏度上有较大的提升, 分别为 0.97% 和 7.25%, 说明本文使用的混合双模型结构能够加强模型的一致性和病变区域的敏感性; M5 和 M6 的实验数据分析可知, 各参数指标都有一定程度的增长, 表明组合损失函数能够提升视网膜病变分级的性能。

消融对比实验混淆矩阵如图 7 所示, 从图 7 中比较能够发现 M6 混淆矩阵, 数据多分布在对角线上且数值较大。而错误识别的样本数据较小且分布在对角线周围, 说明本文所提出的改进对视网膜病变区域识别有显著作用, 能够增强 DR 分级能力。

表 1 在 IDRID 数据集的消融结果

Table 1 The ablation results of the IDRID dataset

Model	Acc/%	QWK/%	Se/%	Sp/%
M1	78.64	88.30	<b>94.20</b>	94.11
M2	78.64	87.13	91.30	94.11
M3	79.61	86.13	<b>94.20</b>	91.17
M4	80.58	88.54	86.95	<b>97.05</b>
M5	79.61	88.44	91.30	<b>97.05</b>
M6	<b>81.55</b>	<b>88.76</b>	<b>94.20</b>	<b>97.05</b>

### 3.5 与其他 DR 分级算法的对比

IDRID 和 APTOS 2019 数据集在不同算法中实验对比结果分别为表 2 和表 3。表 2 用 QWK、Acc、Se 和 Sp 作为评价指标, 与之对比该实验的是当前 DR 分级主流算法 (如文献 [7]) 采用多阶段的迁移学习方法, 提取不同数据集中特征表示信息, 能够提高病变分级效果, 与本文结果相比略低。文献 [19] 提出 IFTL DR 预测模型 (CNN+SVM), 先对数据集进行背景消除技术, 再利用 CNN 模型获取图像特征并通过 SVM 进行机器学习分类, 虽然综合了两者的优势, 有较大的进步, 但还是低于本文算法。文献 [20] 构建一种将粗细网络优势互补的方法, 粗网络主要进行二分类作用, 而细网络则进行细分等级, 提高视网膜分级效率, 与本文相比仍有不足之处。文献 [21] 是采用 ResNet-50 为主干网络, 其次还有自适应特征过滤、特征互补融合模块和细粒度分类损失和焦点损失函数共同组成算法模型, 能够在一定程度上缓解样本不平衡, 提高模型 DR 分级效率, Se 和本文一样, 但其他指标略低于本文算法。表 2 中的文献 [22] 和文献 [23] 是关于 IDRID 数据集的复现, 文献 [22] 是细粒度分类, 以 ResNet-50 为骨干网络加入特征增强、抑制模块和信息融合模块共同组成, 有利于挖掘细微病变特



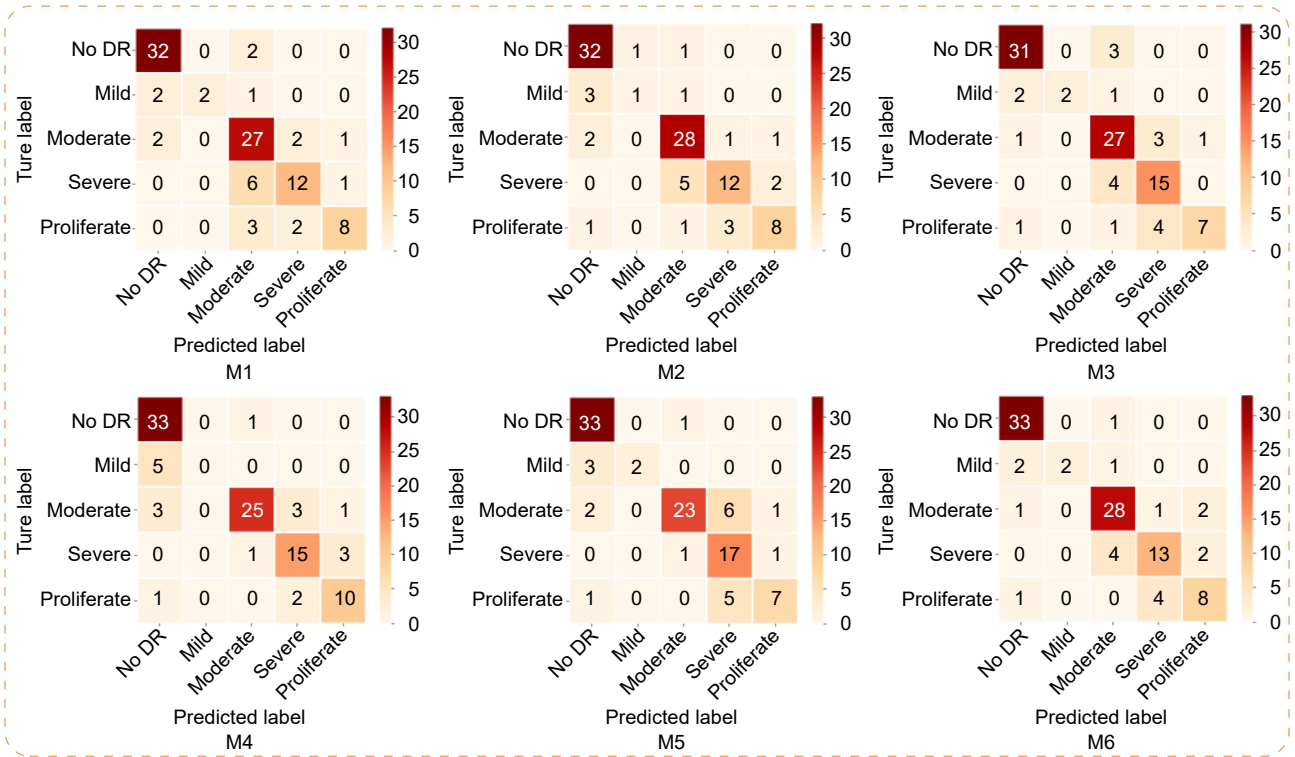


图 7 混淆矩阵

Fig. 7 The confused matrix

表 2 不同算法在 IDRID 数据集的结果表现

Table 2 The results of different algorithms in IDRID data sets

Methods	Model	QWK/%	Acc/%	Se/%	Sp/%
Shi et al. <sup>[7]</sup>	Efficientnet-b5	87.63	79.06	—	—
Bhardwaj et al. <sup>[19]</sup>	ResNet-50	—	79.46	82.85	76.98
Wu et al. <sup>[20]</sup>	ResNet-18	—	56.19	64.21	87.39
Liang et al. <sup>[21]</sup>	ResNet-50+FCFM	88.70	80.58	<b>94.20</b>	94.10
Song et al. <sup>[22]</sup>	ResNet-50+FDM	<b>89.11</b>	80.58	92.75	94.10
Liu et al. <sup>[23]</sup>	ResNet-50	84.42	75.78	92.75	91.17
Ours	Res2Net-50+ Densenet-121	88.76	81.55	<b>94.20</b>	<b>97.05</b>

注: Re和Se计算公式一致, 能够相互替换。

表 3 不同模型在 APTOS 2019 数据集的结果表现

Table 3 The results of different models in the APTOS 2019 data sets

Methods	Model	QWK/%	Acc/%	Re/%	AUC/%
Shaik et al. <sup>[8]</sup>	LA-NSVM	75.64	84.31	66.16	—
Bodapati et al. <sup>[24]</sup>	Xception+VGG16	—	82.54	83.00	79.00
Bodapati et al. <sup>[25]</sup>	VGG16-fc2+Xception	70.90	80.96	—	—
Kobat et al. <sup>[26]</sup>	DenseNet201	78.37	<b>85.93</b>	69.72	—
Song et al. <sup>[22]</sup>	ResNet-50+FDM	86.34	84.28	85.32	92.78
Liu et al. <sup>[23]</sup>	ResNet-50	86.08	81.96	86.43	92.46
Ours	Res2Net-50+ Densenet-121	<b>90.29</b>	84.42	<b>87.40</b>	<b>93.60</b>

注: Re和Se计算公式一致, 能够相互替换。

征, 提高病变分级能力, 只有 QWK 比本文高 0.35%, 其他指标低于本文算法。文献 [23] 是一种跨层相互注意力网络能够循环训练进行特征提取, 但由于参数量较多, 模型过于复杂, 所以指标比本文较低。

如表 3 所示, 在 APTOS 2019 数据集采用 QWK、Acc、Re 和 AUC 作为评价指标, 与最近主流算法作对比。文献 [8] 构建一种多阶段学习, 能够同时获取图像空间表征信息和病灶注意力特征信息, 准确率略高于本文算法, 但其他指标则远低于本文算法, 可能对于潜在的注意力特征病变区域的利用较优于文中算法。文献 [24] 构建了一种具有门控注意力机制的深度神经网络, 能够独立的从不同的通道特征中学习不一致的特征信息, 且有利于提高模型的泛化性, 但与本文算法相比较弱。文献 [25] 使用 VGG16-fc2 和 Xception 结合的混合模型, 从不同模型中提取到多个深度特征, 进行相互补充, 虽然能提高模型性能但其评价指标远低于本文。文献 [26] 应用迁移学习 DenseNet201 算法能在较短时间内生成大量特征, 在对比结果中准确率最高, 超过本文 1.51%, 但其它评价指标低于本文算法, 造成结果相差较大。文献 [26] 基于水平和垂直方向的非固定尺寸分割模型, 对于病灶区域的识别相较于本文算法更为准确。这可能是因实验设备存在差异, 文献 [26] 采用服务器来进行实验而本文则采用电脑的方式进行实验。

此外, 为进一步增强本文算法的说服力, 在相同实验条件下将最近主流算法的两篇文献 [22] 和 [23] 在 APTOS 2019 数据集上进行复现, 其数据结果如表 3 所示。文献 [22] 算法与本文算法在 Acc 评价指标上结果较为接近, 但 QWK、AUC 与 Se 值与本文相比较低。文献 [23] 虽然可以多次循环训练, 但也存

在 (a)、(b)、(c) 模型结构复杂、实验时间较长等问题。

图 8 是 APTOS 2019 数据集的实验结果, 其中 8(a)、8(b) 和 8(c) 分别表示文献 [22] 复现结果、文献 [23] 复现结果和本文算法的 ROC 曲线, 观察图 8 中的三幅图发现 8(c) 中 DR:0-4 的曲线分布较为相近且都靠近左上角 (越靠近左上角表明预测模型的准确率越高), 尤其当 DR 为 4 时对比 8(a) 和 8(b) 曲线图更为明显。ROC 曲线下面积计算 AUC 值分别为 92.78%、92.46% 和 93.60%, 结果表明本文算法在该数据集上较优。通过表 2 和表 3 本文算法和对比算法进行比较其数据较好, 表明模型存在较强的泛化性且具有一定的优越性, 对视网膜病变分级具有良好效果。

## 4 结束语

本文提出一种融合坐标感知与混合提取的糖尿病视网膜病变分级模型。由于存在特征提取不充分问题, 采用混合双模型进行逐级特征提取, 并选取多种尺度信息构成一个完整的图像信息。为了使多尺度特征信息进一步聚焦到病变特征, 通过多层坐标感知模块和注意力特征融合模块筛选干扰噪声后进行两种多尺度信息特征融合; 其次根据自适应病灶特征重新分配权重确保微小病变区域也能够获取足够的权重; 再次利用组合损失函数缓解不同样本之间存在的差异, 从而提升 DR 分级的准确率; 最后进行实验在数据集 IDRID 中的灵敏度与特异性分别为 94.20% 和 97.05%。在数据集 APTOS 2019 上的灵敏度和 ROC 曲线下面积分别为 87.40% 和 93.60%。实验结果表明, 本文算法总体性能优于近年不同算法, 结果表明本文算法具有一定的应用价值。

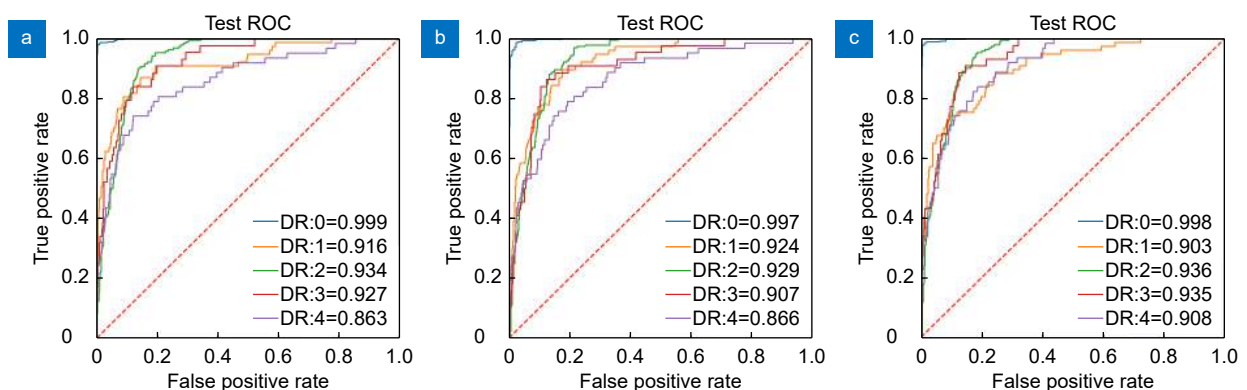


图 8 复现 DR 各类 AUC 值。(a) ResNet-50+FDM<sup>[22]</sup>, (b) ResNet-50<sup>[23]</sup>, (c) 本文方法

Fig. 8 Reapped various types of AUC values. (a) ResNet-50+FDM<sup>[22]</sup>, (b) ResNet-50<sup>[23]</sup>, (c) Method of this article

## 参考文献

- [1] Liang L M, Lu B H, Long P W, et al. Adaptive feature fusion cascade Transformer retinal vessel segmentation algorithm[J]. *Opto-Electron Eng*, 2023, 50(10): 230161.  
梁礼明, 卢宝贺, 龙鹏威, 等. 自适应特征融合级联Transformer视网膜血管分割算法[J]. *光电工程*, 2023, 50(10): 230161.
- [2] Lv J, Wang Z Y, Liang H C. Boundary attention assisted dynamic graph convolution for retinal vascular segmentation[J]. *Opto-Electron Eng*, 2023, 50(1): 220116.  
吕佳, 王泽宇, 梁浩城. 边界注意力辅助的动态图卷积视网膜血管分割[J]. *光电工程*, 2023, 50(1): 220116.
- [3] Chen M H, Wang T, Yuan Y, et al. Study on retinal OCT segmentation with dual-encoder[J]. *Opto-Electron Eng*, 2023, 50(10): 230146.  
陈明惠, 王腾, 袁媛, 等. 引入双编码器模型的OCT视网膜图像分割[J]. *光电工程*, 2023, 50(10): 230146.
- [4] He A L, Li T, Li N, et al. CABNet: category attention block for imbalanced diabetic retinopathy grading[J]. *IEEE Trans Med Imaging*, 2021, 40(1): 143–153.
- [5] Ashwini K, Dash R. Grading diabetic retinopathy using multiresolution based CNN[J]. *Biomed Signal Process Control*, 2023, 86: 105210.
- [6] Zhou K, Gu Z W, Liu W, et al. Multi-cell multi-task convolutional neural networks for diabetic retinopathy grading[C]//*Proceedings of the 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2018: 2724–2727. <https://doi.org/10.1109/EMBC.2018.8512828>.
- [7] 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, 2023. <https://doi.org/10.48550/arXiv.2109.11806>.
- [8] Shaik N S, Cherukuri T K. Lesion-aware attention with neural support vector machine for retinopathy diagnosis[J]. *Mach Vision Appl*, 2021, 32(6): 126.
- [9] Zhang W X, Wu Q. Fine-grained image classification based on multi-branch attention-augmentation[J]. *Comput Sci*, 2022, 49(5): 105–112.  
张文轩, 吴秦. 基于多分支注意力增强的细粒度图像分类[J]. *计算机科学*, 2022, 49(5): 105–112.
- [10] Cheng X H, Li H J, Deng Y, et al. Study on grading of diabetic retinopathy based on Me-ANet model[J]. *Guangxi Sci*, 2022, 29(2): 249–259.  
程小辉, 李贺军, 邓响, 等. 基于ME-ANet模型的糖尿病视网膜病变分级[J]. *广西科学*, 2022, 29(2): 249–259.
- [11] Gao S H, Cheng M M, Zhao K, et al. Res2Net: a new multi-scale backbone architecture[J]. *IEEE Trans Pattern Anal Mach Intell*, 2019, 43(2): 652–662.
- [12] Vellaichamy A S, Swaminathan A, Varun C, et al. Multiple plant leaf disease classification using densenet-121 architecture[J]. *Int J Electr Eng Technol*, 2021, 12(5): 38–57.
- [13] Hou Q B, Zhou D Q, Feng J S. Coordinate attention for efficient mobile network design[C]//*Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021: 13708–13717. <https://doi.org/10.1109/CVPR46437.2021.01350>.
- [14] Dai Y M, Gieseke F, Oehmcke S, et al. Attentional feature fusion[C]//*Proceedings of 2021 IEEE Winter Conference on Applications of Computer Vision*, 2021: 3559–3568. <https://doi.org/10.1109/WACV48630.2021.00360>.
- [15] Mukhoti J, Kulharia V, Sanyal A, et al. Calibrating deep neural networks using focal loss[C]//*Proceedings of the 34th International Conference on Neural Information Processing Systems*, 2020: 1282.
- [16] Rezaei-Dastjerdehei M R, Mijani A, Fatemizadeh E. Addressing imbalance in multi-label classification using weighted cross entropy loss function[C]//*Proceedings of the 27th National and 5th International Iranian Conference on Biomedical Engineering (ICBME)*, 2020: 333–338. <https://doi.org/10.1109/ICBME51989.2020.9319440>.
- [17] 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.
- [18] Adak C, Karkera T, Chattopadhyay S, et al. Detecting severity of diabetic retinopathy from fundus images using ensemble transformers[Z]. arXiv: 2301.00973, 2023. <https://doi.org/10.48550/arXiv.2301.00973>.
- [19] 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.
- [20] 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.
- [21] Liang L M, Lei K, Zhan T, et al. A feature-adaptive filtering algorithm for grading retinopathy[J]. *J Graph*, 2022, 43(5): 815–824.  
梁礼明, 雷坤, 詹涛, 等. 特征自适应过滤的视网膜病变分级算法[J]. *图学学报*, 2022, 43(5): 815–824.
- [22] Song J W, Yang R Y. Feature boosting, suppression, and diversification for fine-grained visual classification[C]//*Proceedings of 2021 International Joint Conference on Neural Networks (IJCNN)*, 2021: 1–8. <https://doi.org/10.1109/IJCNN52387.2021.9534004>.
- [23] Du R Y, Chang D L, Bhunia A K, et al. Fine-grained visual classification via progressive multi-granularity training of jigsaw patches[C]//*Proceedings of the 16th European Conference on Computer Vision*, 2020: 153–168. [https://doi.org/10.1007/978-3-030-58565-5\\_10](https://doi.org/10.1007/978-3-030-58565-5_10).
- [24] Bodapati J D, Shareef N S, Naralasetti V. Composite deep neural network with gated-attention mechanism for diabetic retinopathy severity classification[J]. *J Ambient Intell Human Comput*, 2021, 12(10): 9825–9839.
- [25] Bodapati J D, Naralasetti V, Shareef S N, et al. Blended multi-modal deep ConvNet features for diabetic retinopathy severity prediction[J]. *Electronics*, 2020, 9(6): 914.
- [26] Kobat S G, Baygin N, Yusufoglu E, et al. Automated diabetic retinopathy detection using horizontal and vertical patch division-based pre-trained DenseNET with digital fundus images[J]. *Diagnostics*, 2022, 12(8): 1975.

## 作者简介



梁礼明(1967-), 男, 硕士, 教授, 硕士生导师, 主要研究方向为机器学习、医学影像和系统建模等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。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: lianglm67@163.com



【通信作者】金家新(1999-), 男, 硕士研究生, 目前就读于江西理工大学, 主要研究方向为医学图像病变分级领域。

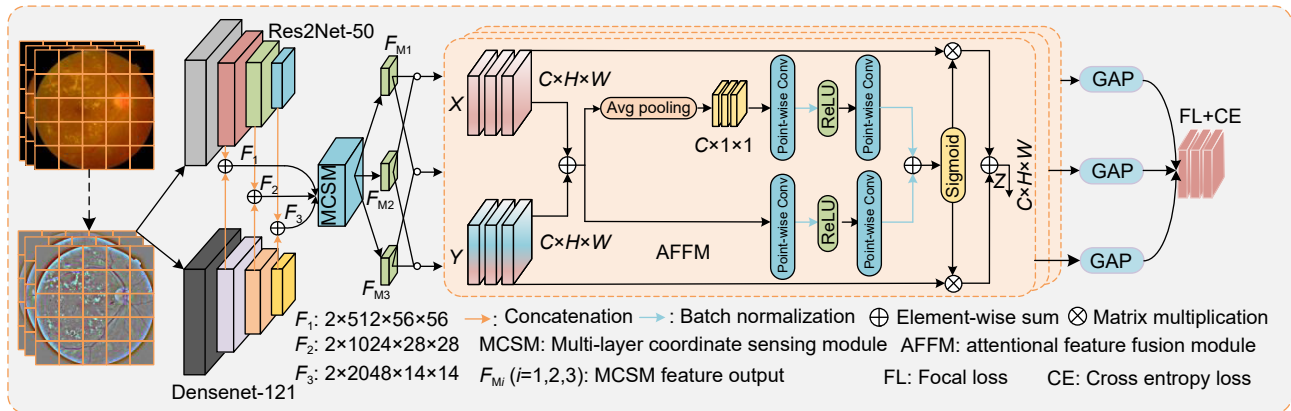
E-mail: 1136868701@qq.com



扫描二维码, 获取PDF全文

# Retinal lesions graded algorithm that integrates coordinate perception and hybrid extraction

Liang Liming, Jin Jiabin\*, Feng Yao, Lu Baohe



**Overview:** Diabetic Retinopathy (DR) is a kind of eye disease caused by diabetes. Patients have been in a high blood sugar environment for a long time. It is very easy to damage the retina. If the disease fails to find the disease in time, it is easy to cause blindness. In recent years, with the development of deep learning technology, it has been widely used in DR intelligent grading diagnosis, but there are still deficiencies: there are small lesions such as micromiel tumors, yellow and white hard exudation or a small number of bleeding spots around the retinal images. The differences between the comparison with normal areas are not obvious. The difference between each child class is difficult to distinguish the difficulty of identification. At the same time, due to the impact of the diabetic disease stage and cure, the number of patients in different stages of different lesions is inconsistent. The sample distribution is unbalanced. In order to solve the above problems, this article proposes a hometown lesion classification algorithm that integrates coordinate perception and mixed extraction. This algorithm first conducts pre-processing operations such as IDRID and APTOS 2019 dataset images, Gaussian filtering and other pre-processing to enhance the differences between image lesions and noise background, then a hybrid dual model composed of the backbone network composed of Res2Net-50 and Densenet-121 backbone network extracts the enhanced images layer by layer, realizes the full capture of multi-scale feature texture, and enhances the robustness and generalization of the algorithm, then integrates the multi-layer coordinate perception module and attention characteristics fusion at the connection between the hybrid dual model connection. The module achieves the purpose of eliminating the characteristics of the lesion, realizing the weight of the semantics of different lesions, ensuring that the small lesion area can also obtain sufficient weight, finally using the focus loss function and cross-entropy. The combination loss function composed of the loss function relieves uneven distribution of sample distribution, weakens the accuracy of the hierarchy of the retinal lesions due to the sample, and further monitors the training and test of the model, thereby improving the accuracy of DR grading.

Experiments on IDRID and APTOS 2019 data sets, the accuracy, secondary weighted coefficients, sensitivity and specificity in IDRID are 81.55%, 88.76%, 94.20%, and 97.05%; The area of the second weighted coefficient, sensitivity and ROC curve was 84.42%, 90.29%, 87.40%, and 93.60%, respectively. The experimental results show that the comparison of the algorithm in this article still has a certain degree of superiority compared with the current mainstream algorithm, which provides a new window for the diagnosis of the hierarchical meta-cooked intelligent auxiliary diagnosis.

Liang L M, Jin J X, Feng Y, et al. Retinal lesions graded algorithm that integrates coordinate perception and hybrid extraction[J]. *Opto-Electron Eng*, 2024, 51(1): 230276; DOI: 10.12086/oe.2024.230276

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: 1136868701@qq.com