

基于特征融合网络的糖尿病视网膜病变分类

赵爽¹, 穆鹤¹, 赵文华^{2*}, 马志庆²¹山东中医药大学实验室管理处, 山东 济南 250355;²山东中医药大学智能与信息工程学院, 山东 济南 250355

摘要 糖尿病视网膜病变是糖尿病最常见也是最严重的并发症之一。为提高对糖尿病视网膜病变严重程度的诊断准确率,进一步为糖尿病视网膜病变治疗的精准用药提供依据,提出一种新型的特征融合网络模型。首先利用轻量化网络 EfficientNet-B0 提取眼底图像的不同层特征,使用高层特征结合三个不同空洞率的空洞卷积形成多尺度特征。然后引入多尺度通道注意力模块(MS-CAM),赋予高层特征和低层特征新的权重,对高低层特征进行融合,形成最终的特征表征,从而完成对糖尿病视网膜病变严重程度的分类。实验结果表明,所提模型的分类准确率达 85.25%,表明其具有较好的可行性,在临床上给医生诊断起到了辅助作用,能更有效地预防糖尿病视网膜病变的进一步恶化。

关键词 自动分类; 糖尿病视网膜病变; 特征融合网络; 空洞卷积; 注意力机制

中图分类号 R445

文献标志码 A

DOI: 10.3788/LOP222415

Classification of Diabetic Retinopathy with Feature Fusion Network

Zhao Shuang¹, Mu Ge¹, Zhao Wenhua^{2*}, Ma Zhiqing²¹Laboratory Management Office, Shandong University of Traditional Chinese Medicine, Jinan 250355, Shandong, China;²College of Intelligence and Information Engineering, Shandong University of Traditional Chinese Medicine, Jinan 250355, Shandong, China

Abstract Diabetic retinopathy is a serious and common complication of diabetes. Herein, we propose a new feature fusion network model to improve the accuracy of the diagnosis for the severity of diabetic retinopathy and provide a basis for its precise drug treatment. A lightweight network, EfficientNet-B0, was used to extract layer information from fundus images, and high-level elements were combined with three dilated convolutions with various dilation rates to obtain multiscale features. The multiscale channel attention module (MS-CAM) was introduced to weigh high- and bottom-level features, which were then fused to form final feature representations and thereby complete the classification of the diabetic retinopathy. Experimental results show the classification accuracy of the proposed model is 85.25%; hence, the network is appropriate for practical applications. Furthermore, the model can play an auxiliary role for clinical diagnosis and can effectively prevent further deterioration in diabetic retinopathy.

Key words automatic classification; diabetic retinopathy; feature fusion network; dilated convolution; attention mechanism

1 引言

糖尿病性视网膜病变(DR)是常见的导致视力严重损害的视网膜血管病变,是糖尿病的三大严重并发症之一^[1]。对于已经被确诊为DR的患者,随着DR病程的延长,若不进行及时有效的治疗,严重则会导致失明且不可逆^[2]。另外根据DR分类结果可进行精准治

疗,因此对DR严重程度进行精确分类是当前研究的重点,同时也是一项极具挑战的任务。这需要眼科医生从眼底图像中观察到眼球中的渗出物、出血点和微血管瘤,根据检测出的病灶类型、数量和位置等信息进行糖尿病视网膜病变分析^[3],医生容易疲乏、劳累而出现错误诊断,所以借助计算机辅助诊断(CAD)方法实现对DR严重程度的自动检测尤为重要。

收稿日期: 2022-08-29; 修回日期: 2022-08-31; 录用日期: 2022-09-23; 网络首发日期: 2022-09-30

基金项目: 山东省研究生教育质量提升计划课题(SDYJG1943)

通信作者: *zhaowh0621@163.com

现阶段深度学习得到快速发展,尤其是卷积神经网络(CNN),已成为对DR严重程度进行分类的首推举措^[4]。例如,Li等^[5]采用迁移学习的方法^[6]对DR严重程度进行预测,使用经过预训练的CNN模型对眼底图像进行特征提取,再训练支持向量机(SVM)^[7]实现分类,并在DR1^[8]和MESSIDOR^[9]数据库进行评估,方法的准确率分别达94.12%和92.01%。Gargeya等^[10]提出了一种新的方法来自动区分健康眼底和DR,首先通过CNN进行特征学习,对于学习到的特征可以通过生成的异常热图进行可视化,突出显示每张眼底图像中的子区域以供进一步临床审查,再训练分类器实现分类,设计的方法的有效性在MESSIDOR2^[11]和E-Ophtha^[12]公共数据集上得到验证,曲线下的面积(AUC)达0.97。Krishnan等^[13]采用迁移学习的方法自动识别DR严重程度,使用二次加权kappa系数评估不同的CNN架构,如ResNet^[14]、Inception-ResNet-v2^[15]等,最后在DR数据集^[16]上进行验证,结果表示所提出的模型的kappa值达0.76。Li等^[17]利用与文献^[12]相同的数据集,提出了一种新的基于CNN的算法,将常用的最大池化层替换为分数阶最大值池化层,然后通过深度神经网络获得更有区别性的分类特征,再结合图像的原数据特征训练SVM分类器,该方法的识别率达86.17%。Amalia等^[18]将CNN和长短期记忆(LSTM)网络^[19]两种深度学习体系结合,根据病变使用LSTM生成描述语句,使用CNN检测DR的严重程度,在MESSIDOR数据集上进行实验,方法的准确率在90%左右。Yang等^[20]提出了一种新的网络模型TransEye对眼底图像进行分类。TransEye结合了CNN和Transformer模型^[21]的优势,并在最后一个注意力层的输入之前增加了有效区域选择模块,不仅可以有效地提取图像的底层特征,还可以建立图像的远程依赖关系,最后在OIA数据集^[22]上进行了评估,达到了84.1%的准确率。Sun等^[23]提出了一种病变感知的Transformer模型(LAT),LAT通过编解码结构同时实现对DR严重程度的分类和对病变的定位识别。其中编码器通过学习像素之间的关联可以有效适应图像外观的变化,而基于病变滤波器的解码器负责识别病变区域。在MESSIDOR、MESSIDOR2和E-Ophtha这三个基准数据集上的实验结果表明LAT优于最先进的模型。Bodapati等^[24]从多个预训练的CNN模型中提取特征,然后利用池化方法进行特征融合用于对DR严重程度的预测。模型在APTOS 2019数据集^[25]上进行了评估,准确率达80.96%和kappa值达0.71。Kassani等^[26]提出了一种新的分类方法,该方法拼接Xception网络^[27]不同卷积层的特征,然后通过训练多层感知器对DR严重程度进行预测,所提出模型的性能在APTOS 2019数据集^[25]上得到评估,准确率为83.09%。

这些CAD方案取得了一定成果,由于临床上对DR严重程度识别速度和精度的需求进一步提高,仍然需要进一步研究以提高对DR严重程度的诊断准确率。又因眼底图像病变的形状、大小特征等病变区域信息通常蕴藏在浅层网络中,但是由于病变区域过小,在特征提取过程中会损失大量病变区域的细节信息。本文提出一种新型的特征融合网络模型对眼底图像进行自动分类,利用轻量化网络EfficientNet-BO^[28]提取眼底图像不同层的特征,结合空洞卷积^[29]增大图像的感受野。引入多尺度通道注意力模块(MS-CAM)^[30]赋予高低层特征新的权重,对高低层特征进行融合,形成最终的特征,旨在提高对DR严重程度的分类准确率,及早对DR进行精准治疗。

2 数据和方法

2.1 数据集

采用的APTOS 2019数据集是印度亚拉文眼科医院为了筛查农村人口的病情得到的^[25]。数据集包含了3662张眼底图像,且大多数图像都有较高的分辨率。将糖尿病造成的视网膜病变的严重程度分为5级,分别是无DR(No DR)、轻度DR(Mild DR)、中度DR(Moderate DR)、重度DR(Severe DR)和增殖期DR(Proliferative DR)。为保证实验过程中训练集与测试集图像的独立性,提高模型的泛化能力,按比例8:1:1将数据集随机分为训练集、验证集和测试集。APTOS 2019数据集的分布如表1所示,具体的5类图像如图1所示。

表1 APTOS 2019数据集的分布
Table 1 Distribution of APTOS 2019 dataset

Label	Number of images
No DR	1805
Mild DR	370
Moderate DR	999
Severe DR	193
Proliferative DR	295

2.2 数据预处理

APTOS 2019数据集是用不同的摄像工具从各个诊所收集得到的,获取的图像有很大差异。其中主要有4种不同类型的图像:矩形图像,无裁剪;矩形图像,有损垂直裁剪;方形图像,紧密裁剪;矩形图像,有损垂直和水平裁剪。4种不同类型的图像如图2所示。

因此,为优化训练过程,首先对输入的眼底图像进行预处理,预处理过程如下。

1) 眼底图像多为矩形,但感兴趣区域为眼球中的渗出物、出血点和微血管瘤等所在区域,图像有大量黑色区域,首先将图中无信息黑色区域裁剪掉;

2) 以各个眼底图像内感兴趣区域的中心为图片中心,对图片进行调整;

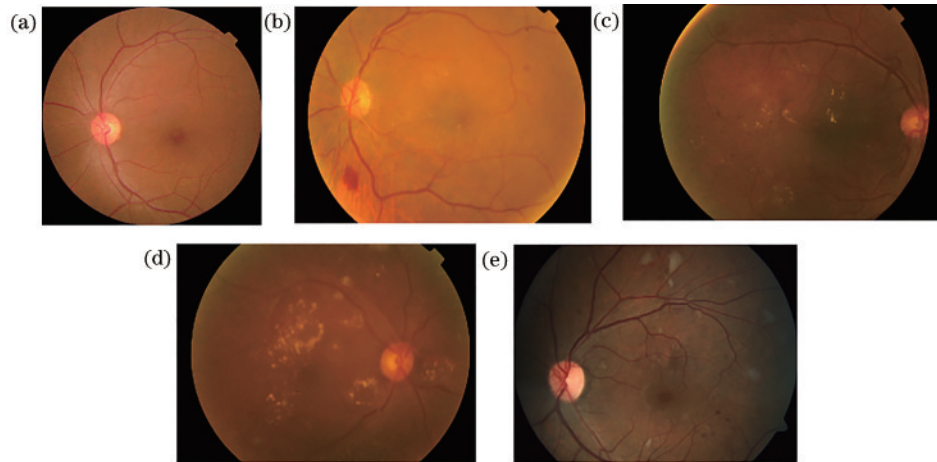


图1 APTOS 2019数据集的5类图像示例。(a) 无DR; (b) 轻度DR; (c) 中度DR; (d) 重度DR; (e) 增殖期DR

Fig. 1 Five categories in the APTOS 2019 dataset. (a) No DR; (b) Mild DR; (c) Moderate DR; (d) Severe DR; (e) Proliferative DR

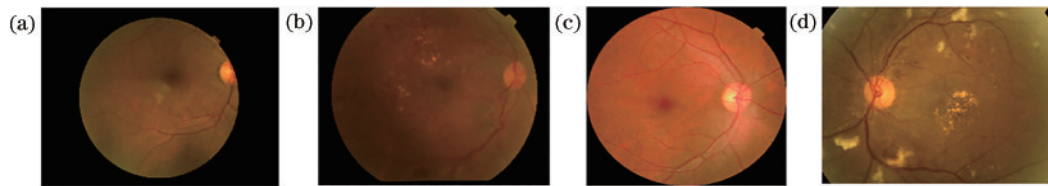


图2 APTOS数据集中4种不同类型的图像。(a)矩形图像,无裁剪;(b)矩形图像,有损垂直裁剪;(c)方形图像,紧密裁剪;(d)矩形图像,有损垂直和水平裁剪

Fig. 2 Four types of images in the APTOS dataset. (a) Rectangular image, no cropping; (b) rectangular image, lossy vertical cropping; (c) square image, tight cropping; (d) rectangular image, lossy vertical and horizontal cropping

3) 统一将图像尺寸调整为 1024×1024 ,在训练过程中,通过对图像进行水平、垂直翻转和随机角度的旋转变换等来增强训练数据,以提高CNN对眼底图像特

征的鲁棒性。

预处理的结果如图3所示。

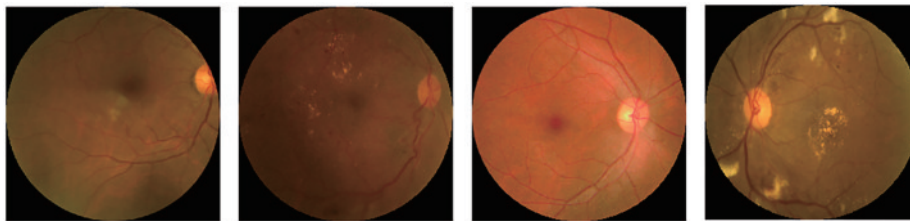


图3 预处理结果图

Fig. 3 Preprocessed images

2.3 网络结构设计

将 APTOS 2019 数据集中图像的尺寸调整为 1024×1024 ,因图像像素密度大、纹理丰富且有更高的可信度,在训练数据过程中使用高计算性能的处理器和过深的网络容易导致过拟合等问题。因此采用轻量化网络 EfficientNet 作为基础网络进行特征提取。传统方法通常只在深度、宽度、输入图像大小中的单一维度上进行放大,模型的参数也会随之增多。通过分析卷积网络的深度、宽度、输入图像大小对卷积网络性能表现的影响,提出了一种混合模型尺度的方法,通过设置一定的参数值平衡调节卷积网络的深度、宽度和输入图像大小,来达到高效高性能的目的^[28]。整体网

络结构设计如图4所示。整体结构包括三个模块:特征提取模块、MS-CAM和分类模块。

特征提取模块首先利用 EfficientNet-B0 网络提取眼底图像的深度特征,将网络第8阶段输出的特征融合多个感受野特征,采用了并行的三个渐进空洞率(空洞率分别为3、6、9)的设置,捕获不同尺度下的感受野特征;再将网络第8阶段输出的特征通过全局平均池化操作,增强网络的全局特征,并使用 1×1 卷积核降低特征的维度;将4个不同尺度的特征通过插值上采样到相同大小,对它们进行拼接,再经过 1×1 卷积层处理,连接非线性激活函数 ReLU,增强网络的非线性特性,得到多尺度特征;同时对第7阶段输出的特征进

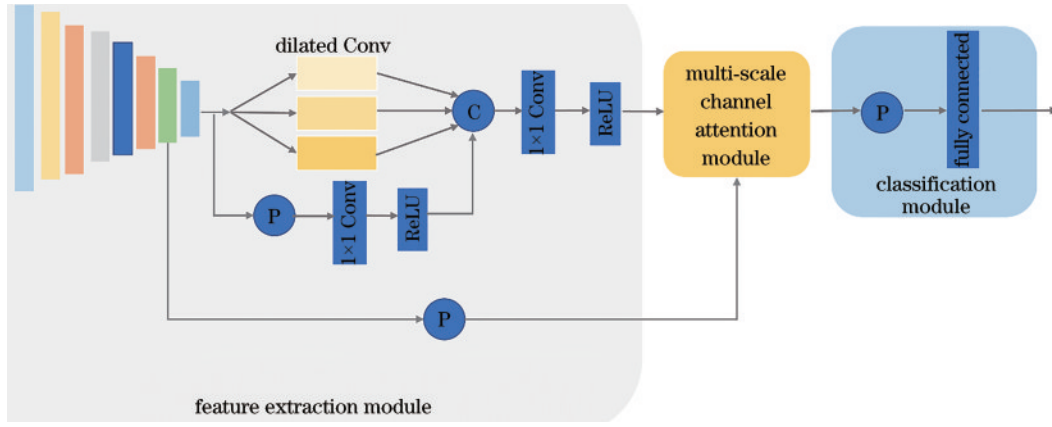


图 4 整体网络结构

Fig. 4 Overall network architecture

行全局平均池化操作,利用MS-CAM赋予其与第8阶段最后输出的多尺度特征新的权重,再进行融合,形成最终的特征表征。其中EfficientNet-B0(阶段1到阶段8)的网络结构如表2所示,其中 k 为所采用的卷积核大小。

表 2 EfficientNet-B0(阶段1~8)网络结构

Table 2 Network architecture of EfficientNet-B0 (stage1~8)

Stage	Operator	$k \times k$	Output size	Number of layers
1	Conv	3×3	$256 \times 256 \times 32$	$\times 1$
2	MBCConvBlock1	3×3	$128 \times 128 \times 16$	$\times 1$
3	MBCConvBlock6	3×3	$128 \times 128 \times 24$	$\times 2$
4	MBCConvBlock6	5×5	$64 \times 64 \times 40$	$\times 2$
5	MBCConvBlock6	3×3	$32 \times 32 \times 80$	$\times 3$
6	MBCConvBlock6	5×5	$16 \times 16 \times 112$	$\times 3$
7	MBCConvBlock6	5×5	$16 \times 16 \times 192$	$\times 4$
8	MBCConvBlock6	3×3	$8 \times 8 \times 320$	$\times 1$

如图5所示,MS-CAM不在主干网络中,通过两个不同尺度的分支来提取通道注意力,能更加地集中和突出重点,更好地融合尺度不一致的特征^[30]。其中两个不同尺度的分支分别是局部通道特征分支和全局通道特征分支。为了使MS-CAM尽可能轻量化,MS-CAM通过逐点卷积(PWConv)^[31]替换不同大小的卷积核。局部通道特征 $F(\mathbf{X}) \in \mathbf{R}^{C \times H \times W}$ 通过类似于瓶颈结构的结构计算得到:

$$F(\mathbf{X}) = \text{B} \left\{ \text{PWConv}_2 \left\{ \delta \left\{ \text{B} \left[\text{PWConv}_1(\mathbf{X}) \right] \right\} \right\} \right\}, \quad (1)$$

式中:PWConv₁将输入特征 \mathbf{X} 的通道数减少到原来的 $1/r$;B代表BatchNorm; δ 代表ReLU;PWConv₂将通道数恢复到原始输入通道数。与 $F(\mathbf{X})$ 不同的是,全局通道特征 $G(\mathbf{X})$ 对输入特征 \mathbf{X} 首先进行1次全局平均池化。

对于输入特征 \mathbf{X} ,利用全局通道特征 $G(\mathbf{X})$ 和局部通道特征 $F(\mathbf{X})$,所对应的输出特征 $\mathbf{X}' \in \mathbf{R}^{C \times H \times W}$ 为

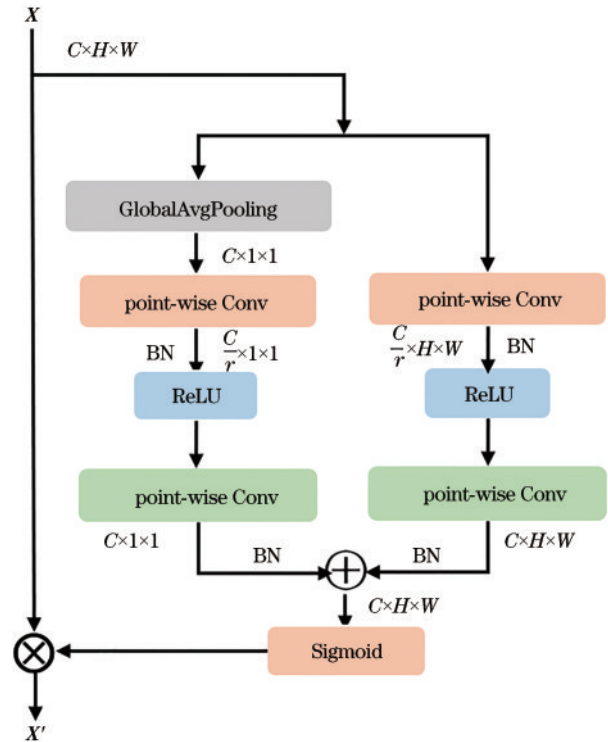


图 5 多尺度通道注意力模块

Fig. 5 Multi-scale channel attention module

$$\mathbf{X}' = \mathbf{X} \otimes \sigma [F(\mathbf{X}) \oplus G(\mathbf{X})], \quad (2)$$

式中: $\sigma [F(\mathbf{X}) \oplus G(\mathbf{X})]$ 表示生成的注意力权重; σ 指Sigmoid激活函数; \oplus 代表广播加法运算; \otimes 代表两个对应元素相乘。

对EfficientNet-B0网络第7阶段最后输出的特征 \mathbf{X}_L 和第8阶段最后输出的特征 \mathbf{X}_H 按对应元素进行融合。融合后的特征包含了更多的特征信息,但是存在着较多的冗余信息。因此,引入多尺度通道注意力模块进行权重的再分配,得到最后的融合特征,计算公式为

$$\mathbf{Z} = \mathbf{X}_L \otimes \sigma [F(\mathbf{X}_L) \oplus G(\mathbf{X}_L)] + \mathbf{X}_H \otimes \left\{ \mathbf{1} - \sigma [F(\mathbf{X}_H) \oplus G(\mathbf{X}_H)] \right\}. \quad (3)$$

分类模块包含 1 个全局平均池化层和 1 个全连接层。

2.4 实验环境

使用 Python 语言, 基于 PyTorch 框架进行编程。实验平台为基于 GeForce GTX1080Ti 的 Ubuntu 系统。在 GPU 上训练模型, 以加快对数据的计算, 提高实验效率。

2.5 实验设置

使用 ImageNet 上预训练的权值初始化所提模型的部分权重, 优化方法采用自适应矩估计算法 (Adam)^[32], 损失函数使用双稳态逻辑损失函数^[33]。将初始学习率设置为 0.0001, 采用小批量数据 (batch size 为 4) 训练模型, epoch 设置为 50。

双稳态逻辑损失函数是 Amid 等^[33]提出的, 通过引入温度 t_1 和尾部重量 t_2 两个参数来解决逻辑损失函数在训练有噪声数据集时的缺陷, 具体公式为

$$\hat{y}_i = \exp_{t_2}[\hat{a}_i - \lambda_{t_2}(\hat{a})], \quad (4)$$

式中: $\sum_i^C \exp_{t_2}[\hat{a}_i - \lambda_{t_2}(\hat{a})] = 1$; \hat{a} 为激活函数; $\lambda_{t_2}(\hat{a})$ 为归一化的值。

$$L_{\text{loss}} = \sum_i^C \left\{ y_i \left[\log(t_1 y_i) - \log(t_1 \hat{y}_i) \right] - \frac{1}{2 - t_1} \left[y_i^{2-t_1} - \hat{y}_i^{2-t_1} \right] \right\}, \quad (5)$$

当 $t_1 = t_2 = 1$ 时, 双稳态函数就退化为普通的逻辑损失函数。 $0 < t_1 < 1$ 时, 数值越小, 对逻辑损失函数界限的约束就越厉害。 $1 \leq t_2$ 时, 数值越大, 尾部就越“厚”, 对指数函数来说衰减也就越慢。本文设置 $t_1 = 0.8$ 和 $t_2 = 1.2$ 。

2.6 评价标准

选择准确率 (accuracy, A)、召回率 (recall, R) 与精确率 (precision, P) 三个指标评价模型的性能。在二分类任务中, 计算准确率、召回率与精确率的公式分别为

$$A = \frac{N_{\text{TP}} + N_{\text{TN}}}{N_{\text{TP}} + N_{\text{TN}} + N_{\text{FN}} + N_{\text{FP}}}, \quad (6)$$

$$R = \frac{N_{\text{TP}}}{N_{\text{TP}} + N_{\text{FN}}}, \quad (7)$$

$$P = \frac{N_{\text{TP}}}{N_{\text{TP}} + N_{\text{FP}}}, \quad (8)$$

式中: TP、TN、FP、FN 分别代表阳性、阴性、假阳性、假阴性。在本文的多分类任务中, 使用宏平均的评价方法, 首先对每个类别分别统计准确率、召回率与精确率三个评价指标, 然后求取算术平均值作为整体的准确率、召回率和精确率。

3 结果与分析

使用 2930 张图像进行模型训练, 366 张图像进行验证, 实验结果来自 366 张未在训练和验证阶段使用的测试图像。50 个 epoch 总用时为 22 h 10 min 10 s, 平

均每个 epoch 用时约为 26 min。图 6 展示了在训练阶段损失函数在训练集、验证集上随着迭代次数变化而变化的情况。随着迭代次数的增加, 训练集的损失值不断下降, 验证集的损失值在开始不断振荡, epoch 为 30 时趋于平稳。

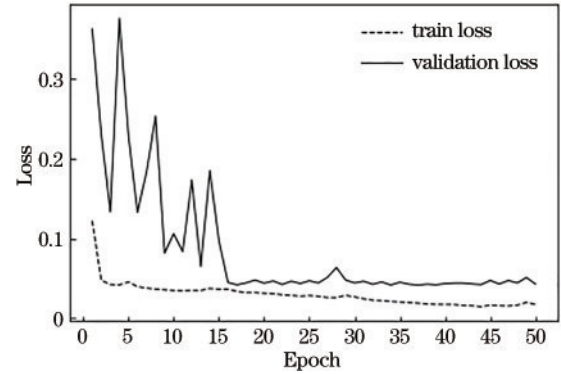


图 6 损失曲线

Fig. 6 Loss curve

图 7 是测试集的混淆矩阵, 其中竖轴的标签表示真实属性, 而横轴的标签表示预测结果。矩阵中的第 1 行第 1 列表示 No DR 被分类成 No DR 的样本数目, 为 176 张; 第 1 行第 2 列表示 No DR 被分类成 Mild DR 的样本数目, 为 4 张; 第 1 行第 3 列表示 No DR 被分类成 Moderate DR 的样本数目, 为 1 张; ……。对角线上的数值 (从左上角到右下角) 代表被正确预测的样本数目, 共为 312 张。

True label	Predicted label				
	No DR	Mild DR	Moderate DR	Severe DR	Proliferative DR
No DR	176	4	1	0	0
Mild DR	0	28	9	0	0
Moderate DR	0	9	86	0	5
Severe DR	1	0	12	4	2
Proliferative DR	0	1	10	0	18

图 7 混淆矩阵

Fig. 7 Confusion matrix

所提模型可较好地地区分正常眼底图像和病变的眼底图像, 但因病变初期眼底图像的症状十分轻微, 故 4 张 Mild DR 样本和 1 张 Moderate DR 样本被划分为 No DR 样本; 同时, 有的正常眼底存在一些伪影等噪声, 这些噪声容易被误判为病灶点, 故极小部分 No DR 样本被划分为 Severe DR 样本; 又因数据集中样本

在病变眼底图像各个级别分布不均衡且相近级别的图像相差不大, Moderate DR 的样本数目明显高于其他病变图像, 故被错误分类成 Moderate DR 的样本数目明显高于被错误分类成其他的样本数目。

设计了 2 组消融实验来验证网络中不同模块的有效性, 分别使用 EfficientNet-B0 和 EfficientNet-B0+Dilated convolution 两个模型与所提模型进行对比, 结果显示在表 3 中。其中 EfficientNet-B0 是使用 EfficientNet-B0 网络的分类结果, EfficientNet-B0+Dilated convolution 是使用 EfficientNet-B0(阶段 1 到阶段 8)网络结合空洞卷积的分类结果。从表 3 可以看到, 所提方法的准确率达 85.25%, 明显优于其他模型。结果说明: 结合不同空洞率的空洞卷积后, 网络具有不同大小的感受野, 能获取多尺度感受野特征, 从而更好地适应不同类型的眼底图像, 使用全局平均池化操作后能够增强全局特征, 有助于增强对糖尿病视网膜病变程度的识别能力; 引入了 MS-CAM, 采用注意力融合的机制融合 CNN 中的高低层特征, 赋予高低层特征不同的权重, 更加地集中和突出眼底图像中细粒度的病变特征。

表 3 消融实验
Table 3 Ablation study unit: %

Model	Accuracy	Recall	Precision
EfficientNet-B0	83.92	62.33	78.94
EfficientNet-B0+Dilated convolution	84.74	64.68	77.08
Proposed model	85.25	68.41	82.20

为验证所提模型的有效性, 选择了现有的网络和研究成果进行对比, 且研究成果中均利用了特征融合的方法, 结果如表 4 所示。从表 4 可以看出, 无论是和文献[34]中现有的网络 ResNet-50、SE-Net 和 SE-ResNet-50 进行对比, 还是和文献[23]和文献[25]中的模型进行对比, 所提模型具有比其他模型更高的准确率, 召回率低于文献[23]中的模型。结果表明, 所提模型通过对眼底图像局部特征和全局特征的有效融合, 可获得最优的特征表征, 分类准确率达到最高, 但因不同亚型之间的数据不平衡问题, 导致召回率较低。

表 4 与现有的研究对比结果

Model	Accuracy	Recall	Precision
ResNet-50 ^[34]	73.778		
SE-Net ^[34]	72.873		
SE-ResNet-50 ^[34]	76.144		
Model in Ref. [23]	80.06	80.00	81.00
Model in Ref. [25]	83.09		
Proposed model	85.25	68.41	82.20

4 结 论

提出一种新型的特征融合网络, 用于对 DR 严重程度的预测, 在 APTOS 2019 数据集上进行了验证。针对 DR 原始数据集图像差异性较大的问题, 在原始的图像上进行预处理, 来统一 DR 数据集的格式, 以便优化训练过程; 利用 EfficientNet-B0 提取眼底图像的不同层特征, 使用高层特征结合空洞卷积, 增大图像的感受野, 引入 MS-CAM 对高低层特征进行融合, 使 DR 严重程度的分类精度得以提升。与现有模型相比, 所提模型表现出一定的优越性, 表明所提模型具有一定的可行性和有效性, 可使糖尿病视网膜病变的临床诊断准确率得到很大提升, 极大程度上阻止了糖尿病视网膜病变的进一步恶化。然而, 不同亚型之间的数据不平衡问题仍然难以解决, 下一步应该收集更多的糖尿病视网膜病变数据集以及探索更有效的生成高分辨率眼底图像的方法, 以增加数据的多样性。

参 考 文 献

- [1] Teo Z L, Tham Y C, Yu M, et al. Global prevalence of diabetic retinopathy and projection of burden through 2045: systematic review and meta-analysis[J]. *Ophthalmology*, 2021, 128(11): 1580-1591.
- [2] Prawej A, Noushin T, Nuren S N, et al. Diabetic retinopathy: an overview on mechanisms, pathophysiology and pharmacotherapy[J]. *Diabetology*, 2022, 3(1): 159-175.
- [3] 李家昱, 陈明惠, 杨瑞君, 等. 糖尿病视网膜病变眼底图像筛查研究[J]. *中国激光*, 2022, 49(11): 1107001.
Li J Y, Chen M H, Yang R J, et al. Fundus image screening for diabetic retinopathy[J]. *Chinese Journal of Lasers*, 2022, 49(11): 1107001.
- [4] Vujosevic S, Aldington S J, Silva P, et al. Screening for diabetic retinopathy: new perspectives and challenges[J]. *The Lancet Diabetes & Endocrinology*, 2020, 8(4): 337-347.
- [5] Li X G, Pang T T, Xiong B, et al. Convolutional neural networks based transfer learning for diabetic retinopathy fundus image classification[C]//2017 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), October 14-16, 2017, Shanghai, China. New York: IEEE Press, 2017.
- [6] Pan S J, Yang Q. A survey on transfer learning[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2010, 22(10): 1345-1359.
- [7] Cortes C, Vapnik V. Support-vector networks[J]. *Machine Learning*, 1995, 20(3): 273-297.
- [8] Rocha A, Carvalho T, Jelinek H F, et al. Points of interest and visual dictionaries for automatic retinal lesion detection[J]. *IEEE Transactions on Biomedical Engineering*, 2012, 59(8): 2244-2253.
- [9] Decencièrre E, Zhang X W, Cazuguel G, et al. Feedback

- on a publicly distributed image database: the messidor database[J]. *Image Analysis & Stereology*, 2014, 33(3): 231-234.
- [10] Gargeya R, Leng T. Automated identification of diabetic retinopathy using deep learning[J]. *Ophthalmology*, 2017, 124(7): 962-969.
- [11] Gulshan V, Peng L, Coram M, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs[J]. *JAMA*, 2016, 316(22): 2402-2410.
- [12] Cuadros J, Bresnick G. EyePACS: an adaptable telemedicine system for diabetic retinopathy screening[J]. *Journal of Diabetes Science and Technology*, 2009, 3(3): 509-516.
- [13] Krishnan A S, Clive R D, Bhat V, et al. A transfer learning approach for diabetic retinopathy classification using deep convolutional neural networks[C]//2018 15th IEEE India Council International Conference, December 16-18, 2018, Coimbatore, India. New York: IEEE Press, 2018.
- [14] 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, June 26-July 1, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 770-778.
- [15] Szegedy C, Liu W, Jia Y Q, et al. Going deeper with convolutions[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition, June 8-14, 2015, Boston, MA. New York: IEEE Press, 2015.
- [16] Kaggle. Diabetic retinopathy detection[EB/OL]. [2022-03-06]. <https://www.kaggle.com/c/diabetic-retinopathy-detection>.
- [17] Li Y H, Yeh N N, Chen S J, et al. Computer-assisted diagnosis for diabetic retinopathy based on fundus images using deep convolutional neural network[J]. *Mobile Information Systems*, 2019, 2019: 6142839.
- [18] Amalia R, Bustamam A, Sarwinda D. Detection and description generation of diabetic retinopathy using convolutional neural network and long short-term memory [J]. *Journal of Physics: Conference Series*, 2021, 1722 (1): 012010.
- [19] Greff K, Srivastava R K, Koutnik J, et al. LSTM: a search space odyssey[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2017, 28(10): 2222-2232.
- [20] Yang H G, Chen J J, Xu M F. Fundus disease image classification based on improved transformer[C]//2021 International Conference on Neuromorphic Computing (ICNC), October 11-17, 2021, Wuhan, China. New York: IEEE Press, 2021: 207-214.
- [21] Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16×16 words: transformers for image recognition at scale[EB/OL]. (2020-10-22) [2022-03-01]. <https://arxiv.org/abs/2010.11929>.
- [22] Nankai University. Nankai University participated in the research of OIA series datasets, which were released to the public[EB/OL]. [2022-03-06]. <https://cyber.nankai.edu.cn/2019/1209/c13342a255556/page.htm>.
- [23] Sun R, Li Y H, Zhang T Z, et al. Lesion-aware transformers for diabetic retinopathy grading[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), December 18-20, 2021, Nashville, TN, USA. New York: IEEE Press, 2021: 10933-10942.
- [24] 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.
- [25] Kaggle. APTOS 2019 blindness detection[EB/OL]. [2022-03-06]. <https://www.kaggle.com/c/aptos2019-blindness-detection>.
- [26] Kassani S H, Kassani P H, Khazaeinezhad R, et al. Diabetic retinopathy classification using a modified xception architecture[C]//2019 IEEE International Symposium on Signal Processing and Information Technology, December 10-12, 2019, Ajman, United Arab Emirates. New York: IEEE Press, 2019.
- [27] Chollet F. Xception: deep learning with depthwise separable convolutions[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 1800-1807.
- [28] Tan M, Le Q V. Efficientnet: rethinking model scaling for convolutional neural networks[C]//36th International Conference on Machine Learning, June 9-15, 2019, Long Beach, California, USA. Copenhagen: MLR Press, 2019: 6105-6114.
- [29] 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 Transactions on Pattern Analysis and Machine Intelligence*, 2018, 40(4): 834-848.
- [30] Dai Y M, Gieseke F, Oehmcke S, et al. Attentional feature fusion[C]//2021 IEEE Winter Conference on Applications of Computer Vision, January 6-8, 2021, Waikoloa, HI, USA. New York: IEEE Press, 2021: 3559-3568.
- [31] Zhang J, Cao Y, Wang Y, et al. Fully point-wise convolutional neural network for modeling statistical regularities in natural images[C]//Proceedings of the 26th ACM International Conference on Multimedia, October 22-26, 2018, Seoul, Republic of Korea. New York: ACM Press, 2018: 984-992.
- [32] Newey W K. Adaptive estimation of regression models via moment restrictions[J]. *Journal of Econometrics*, 1988, 38(3): 301-339.
- [33] Amid E, Warmuth M K, Anil R, et al. Robust Bi-tempered logistic loss based on Bregman divergences [EB/OL]. (2019-06-08) [2022-03-06]. <https://arxiv.org/abs/1906.03361>.
- [34] 张慧婷. 用于糖尿病视网膜病变分级的深度学习模型研究[D]. 太原: 太原理工大学, 2021.
- Zhang H T. The research of deep learning model for diabetic retinopathy grading[D]. Taiyuan: Taiyuan University of Technology, 2021.