

# 一种结合多尺度特征融合和混合域注意力机制的眼底疾病诊断方法

刘辉<sup>1</sup>, 朱正为<sup>1,2</sup>, 张徐<sup>1</sup>, 仲慧<sup>1</sup>

1. 西南科技大学信息工程学院, 四川 绵阳 621010; 2. 西南科技大学特殊环境机器人技术四川省重点实验室, 四川 绵阳 621010

**【摘要】**针对眼底疾病图像细微特征多、样本数量较少、诊断难度较大的问题,结合深度学习和医学影像技术,提出一种融合多尺度特征融合和混合域注意力机制的眼底疾病诊断网络模型和方法。该方法以Resnet50作为基线网络,通过对其进行改进和优化,利用并行多分支结构提取不同感受野下的眼底疾病特征,有效提高模型的特征提取能力和计算效率,采用混合域注意力机制选择对当前任务目标更关键的信息,有效提高模型的分类性能。最后利用ODIR数据集对该方法进行实验,实验结果表明,提出的方法对正常眼和不同眼底疾病的诊断准确率达到93.2%,相较于基线网络提高5.2%,诊断效果良好。

**【关键词】**眼底疾病;卷积神经网络;多尺度特征融合;混合域注意力机制

**【中图分类号】**R318;TP391.4

**【文献标志码】**A

**【文章编号】**1005-202X(2023)12-1477-09

## A diagnostic method incorporating multi-scale feature fusion and hybrid domain attention mechanism for fundus diseases

LIU Hui<sup>1</sup>, ZHU Zhengwei<sup>1,2</sup>, ZHANG Xu<sup>1</sup>, ZHONG Hui<sup>1</sup>

1. College of Information Engineering, Southwest University of Science and Technology, Mianyang 621010, China;

2. Robot Technology Used for Special Environment Key Laboratory of Sichuan Province, Mianyang 621010, China

**Abstract:** In view of numerous subtle features in fundus disease images, small sample sizes, and difficulties in diagnosis, both deep learning and medical imaging technologies are used to develop a fundus disease diagnosis model that integrates multi-scale features and hybrid domain attention mechanism. Resnet50 network is taken as the baseline network, and it is modified in the study. The method uses parallel multi-branch architecture to extract the features of fundus diseases under different receptive fields for effectively improving the feature extraction ability and computational efficiency, and adopts hybrid domain attention mechanism to select information that is more critical to the current task for effectively enhancing the classification performance. The test on ODIR dataset shows that the proposed method has a diagnostic accuracy of 93.2% for different fundus diseases, which is 5.2% higher than the baseline network, demonstrating a good diagnostic performance.

**Keywords:** fundus disease; convolutional neural network; multi-scale feature fusion; hybrid domain attention mechanism

### 前言

目前,青光眼、白内障、糖尿病视网膜病变等眼底疾病已成为中国致盲的主要原因,严重影响民众的健康,因此需要尽早筛查和治疗。专业眼科医生

一般通过眼底图像对眼科疾病进行诊断,不仅对医生的阅片能力要求高,且耗时耗力,检查周期长,对于一些常见的眼底疾病,例如糖尿病视网膜病变<sup>[1]</sup>和白内障等,早期症状微小,诊断比较困难,可能导致错过最佳治疗时期<sup>[2]</sup>。近年来,深度学习技术在各种医学疾病诊断中取得一些较好的应用成果<sup>[3-5]</sup>。2016年,谷歌研究团队的Gulshan等<sup>[6]</sup>建立一个深度学习网络模型对糖尿病视网膜病变进行检测,实验表明其准确性可与专业眼科医生相媲美。2017年,Karri等<sup>[7]</sup>提出一种利用光学相干断层扫描图像识别不同视网膜病理的方法,该方法对3种不同眼底疾病实现分类。Muhammad等<sup>[8]</sup>提出一种基于AlexNet与随机森林方法混合的深度学习方法,对52名青光眼患者

**【收稿日期】**2023-07-11

**【基金项目】**国家自然科学基金(62071399);特殊环境机器人四川省重点实验室项目(13zxtk08);西南科技大学博士基金(17zx7159)

**【作者简介】**刘辉,硕士,研究方向:深度学习及其在眼底疾病诊断中的应用,E-mail: lh990707@163.com

**【通信作者】**朱正为,博士,副教授,研究生导师,研究方向:图像处理、人工智能和目标识别等,E-mail: zhuzwin@163.com

的OCT眼底图像进行诊断,该方法对OCT图像的识别精度达到87.3%。Grassmann等<sup>[9]</sup>将神经网络与随机神经相结合,提出一种黄斑病变多级分类模型,该模型的分类准确率达到94.3%。2018年,北京致远慧图有限公司联合国内多家眼科医院合作研发了“EyeWisdom”眼科诊断系统,可有效检测糖尿病视网膜病变,受到广泛的关注和认可<sup>[10]</sup>。Google旗下的DeepMind公司、伦敦大学学院和 Moorfields 眼科医院的研究人员,基于深度学习技术研究发布了一套辅助医生诊断眼部疾病的系统<sup>[11]</sup>,该系统对眼底疾病的识别准确率达到94%,并且可以给出相应的诊疗建议<sup>[12]</sup>。尽管深度学习技术在医学疾病诊断领域取得不小的成果,但目前这些研究尚存在一定的局限性,例如在眼底疾病诊断方面,目前的研究大多针对单个类别疾病的诊断,且仅充当一种辅助治疗手段<sup>[13]</sup>。

本文针对上述研究中存在的问题和不足,结合多尺度特征融合和混合域注意力机制对 Resnet50 基线网络模型进行改进<sup>[14]</sup>,提出一种基于多尺度特征融合和混合域注意力机制的眼底疾病诊断新方法。本文提出的改进措施主要有:(1)针对眼底疾病图像的特点,对眼底疾病智能识别 (Ocular Disease Intelligent Recognition, ODIR)数据集进行数据清洗、图像增强和数据增广等预处理操作,以获得高质量的数据集,有利于提高网络模型训练的精度;(2)以 Resnet50 为基线网络,构建一种适合眼底疾病诊断的网络模型,有效提高模型对眼底疾病特征信息的表征能力;(3)利用并行多分支结构提取不同感受野下

的眼底疾病特征,实现多尺度特征融合,提高模型的特征提取能力和计算效率;(4)采用混合域注意力机制,使模型可以获取更多有关疾病的关键信息,提高模型的分类能力;(5)使用更加高效的卷积核替代基线网络中原有的卷积核,在减少网络参数数量的同时保持高效的计算效率。利用 ODIR 数据集对本文方法进行验证,实验验证了多尺度特征融合和混合域注意力机制的有效性,实验结果表明,本文方法对常见眼底疾病具有良好的诊断分类效果,识别准确率可达93.2%。

1 融合多尺度特征和混合域注意力机制的眼底疾病诊断方法

1.1 眼底疾病诊断模型的总体架构

本文以 Resnet50 为基线网络,通过改进和优化,设计出眼底疾病诊断网络模型的总体架构如图1所示。该网络主要包含以下几个子模块:首先是输入图像的卷积部分,如图1中的紫色模块所示,它是一个具有7×7卷积核的卷积层,其输出大小为64×112×112;然后是融合多尺度特征和混合域注意力机制的特征提取模块,这是眼底疾病诊断模型的核心部分,主要作用是提取图像的深层特征,如图1中的4个橙色模块所示,橙色模块的具体实现结构如图2所示,输出大小分别为256×56×56、512×28×28、1 024×14×14和2 048×7×7的特征图;最后经过一个 AvgPooling层及分类器处理,输出眼底疾病诊断分类结果,其中分类器由 Dropout、线性层和 Softmax 层组成。

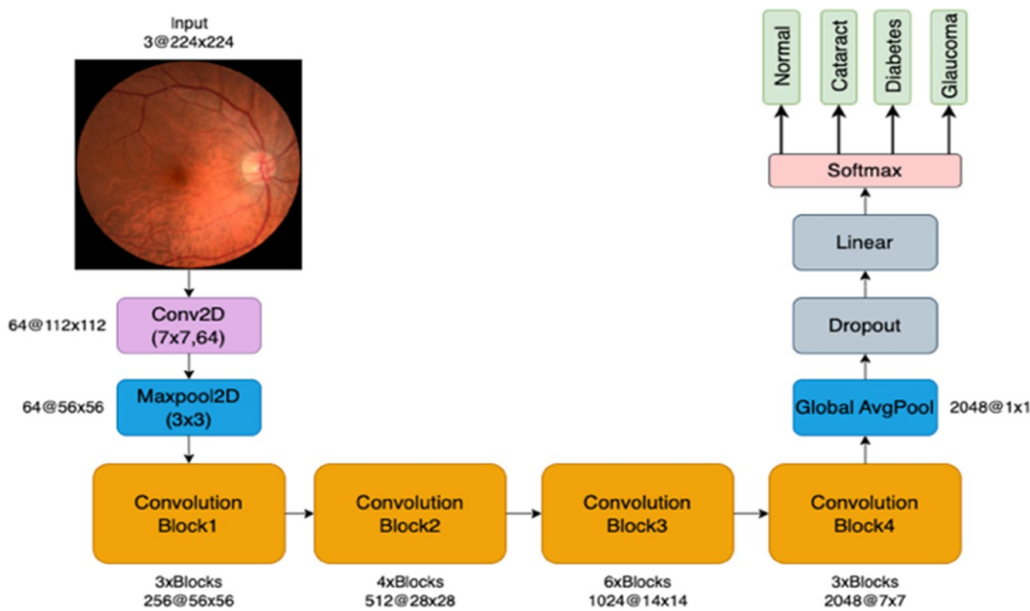


图1 眼底疾病诊断模型的总体架构

Figure 1 Overall structure of fundus disease diagnosis model

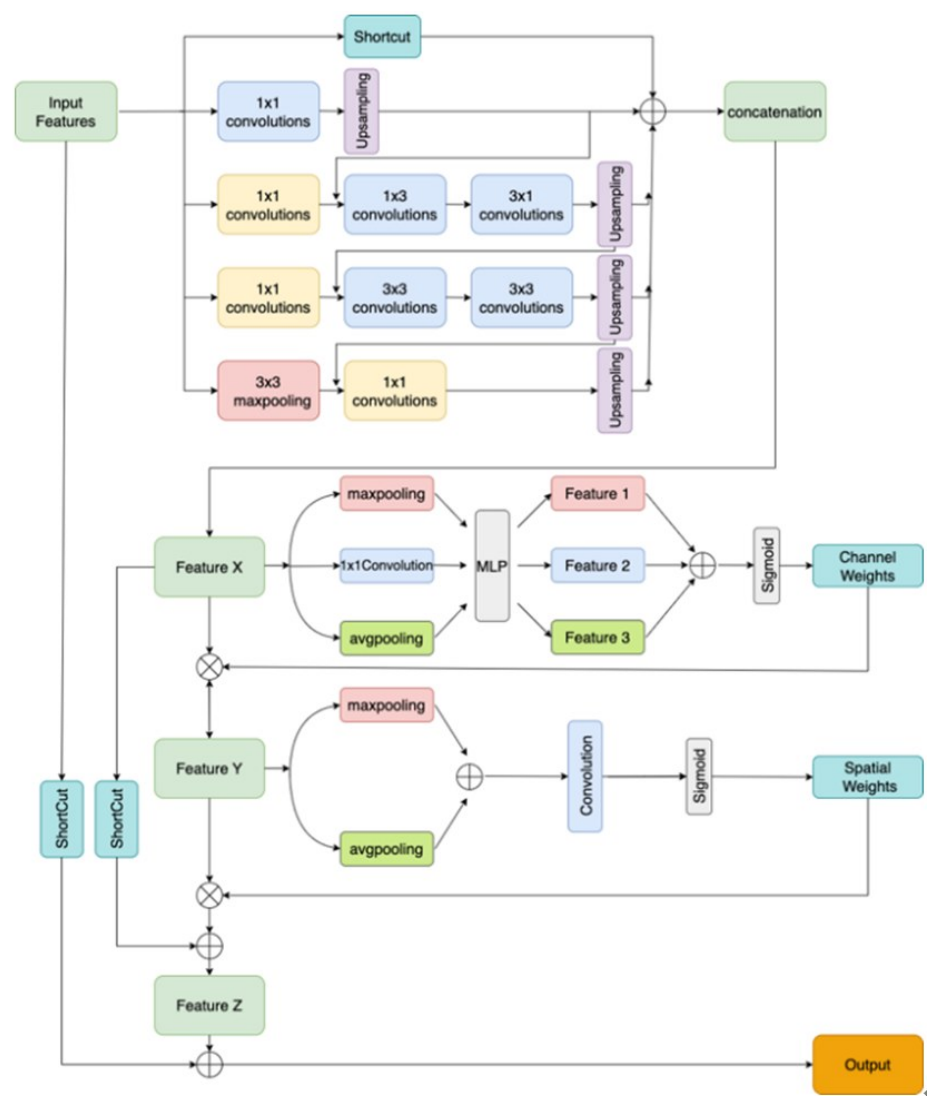


图2 融合多尺度特征和混合域注意力机制的特征提取模块

Figure 2 Feature extraction module incorporating multi-scale feature fusion and hybrid domain attention mechanism

1.2 ODIR 数据集预处理

为了创建一个优秀的眼底疾病诊断网络模型，在进行模型训练前，通常需要对ODIR数据集进行数据清洗、图像增强和数据增广等预处理操作。由于拍摄角度不同，光线明暗不一，以及噪声的存在，ODIR数据集中的眼底图像质量参差不齐，其中不乏存在一些冗余、缺失、异常和低质数据，因此首先需要对数据进行清洗操作。其次，ODIR数据集中的眼底图像还存在光照不均匀，明暗不一致，病理特征不明显的问题，因此需要对眼底图像的灰度值进行预处理操作，本文具体采用限制对比度自适应直方图均衡化（Contrast Limited Adaptive Histogram Equalization, CLAHE）<sup>[15]</sup>方法进行处理。一幅眼底图像经CLAHE处理后的结果图像如图3所示。另外，由于ODIR数据集是由不同人、不同相机采集获得的，其图像尺寸大小不一，因此在利用CLAHE对图

像灰度预处理之前，需要对图像进行裁剪，将所有图像的尺寸统一为224×224，以方便后续操作。最后，构建一个优秀的深度学习模型需要海量的数据资源，由于眼底疾病图像的特殊性，其获取渠道较少，获取方式受限，相对于其它数据集，ODIR数据集中的图像样本数据相对较少，因此要对其进行数据增广，增加其样本数量，以提高模型的泛化性和鲁棒性<sup>[16]</sup>。本文具体采用旋转、缩放、改变对比度、添加噪声等数据增广方法来增加眼底图像样本数据的数量。

1.3 眼底图像的多尺度特征融合

为了增强模型的特征提取能力，获得眼底图像多尺度融合特征<sup>[17]</sup>，本文设计和使用一种并行多分支结构，如图4所示，其采用1×1、1×3、3×1、3×3不同尺寸的卷积核和3×3最大池化层的并行结构，以提取不同感受野下的特征，还引入残差结构，保证良好的



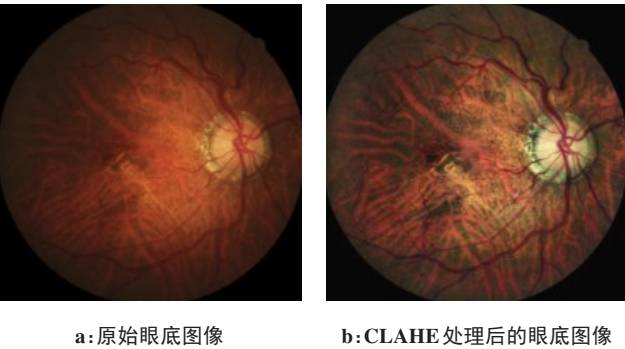


图3 CLAHE处理前后的眼底图像

Figure 3 Fundus images before and after CLAHE processing

信息。然后将上采样完成后的特征在输出端进行多尺度特征融合操作,获得眼底图像不同尺度的特征。采用并行多分支结构的好处是既增加了网络的宽度和深度,可以从图像中提取不同尺度下的眼底疾病特征信息,又保证网络模型的计算效率。浅层网络的几何表征能力强,深层网络的语义信息表征能力强,将浅层信息与深层信息相结合,有利于提高模型的诊断准确性。

1.4 混合域注意力机制的设计与使用

不同眼底疾病具有不同的病理特征,眼底图像

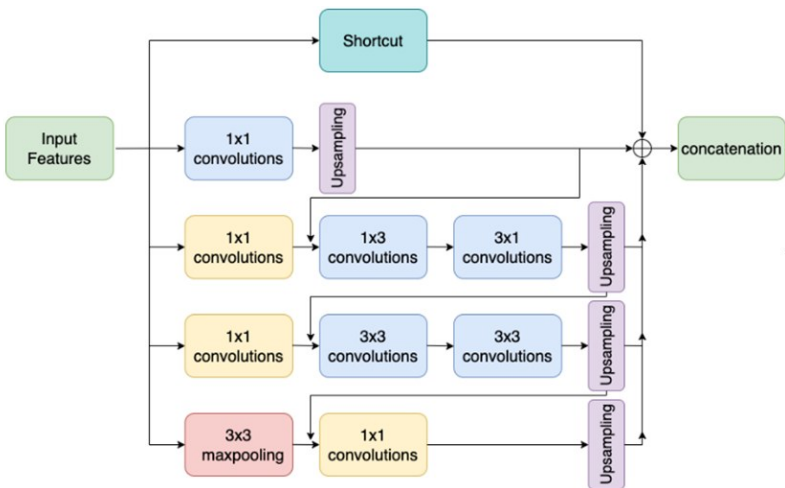


图4 利用并行多分支结构实现多尺度特征融合

Figure 4 Multi-scale feature fusion using parallel multi-branch architecture

具有一定的复杂性,因此获取关键的疾病特征信息进行模型学习就变得十分重要。注意力机制可以从大量信息中选择出对当前任务目标更为关键的特征信息,通过分配更多的权重参数在某个对象身上,提高对此对象特征提取的能力,以提高网络模型的性能<sup>[18]</sup>。

本文设计和采用如图5所示的混合域注意力机制模块,该模块可有效结合通道注意力<sup>[19]</sup>和空间注意力<sup>[20]</sup>工作机制,并在其中引入残差模块<sup>[21]</sup>,对输入的特征图进行通道注意力和空间注意力联合处理。混合域注意力机制具体实现公式如下所示:

$$W_c(F_x) = \sigma \left( \text{MLP}(\text{AvgPool}(F_x)) + \right. \\ \left. \text{MLP}(\text{MaxPool}(F_x)) + \text{MLP}(f^{(1 \times 1)}(F_x)) \right) \quad (1)$$

$$W_s(F_y) = \sigma \left( f^{(7 \times 7)} \left( \left[ \text{AvgPool}(F_y); \text{MaxPool}(F_y) \right] \right) \right) \quad (2)$$

$$F_y = F_x \cdot W_c(F_x) \quad (3)$$

$$F_z = F_y \cdot W_s(F_y) + F_x \quad (4)$$

式中,σ表示 Sigmoid 函数;MLP代表多层感知机;

AvgPool与MaxPool分别代表平均池化操作和最大池化操作; $f^{n \times n}$ 代表 $n \times n$ 的卷积操作; $W_c(F_x)$ 与 $W_s(F_y)$ 分别代表特征图经过通道注意力与空间注意力后的权重; $F_x$ 表示输入特征图, $F_y$ 表示对输入施加通道注意力权重后的特征图, $F_z$ 表示最终输出的特征图。利用本混合域注意力机制,可获取更多有关疾病的重要信息,不仅有利于提取疾病的关键特征,还能保留原有的位置信息,提高网络模型的工作效率和诊断准确性。

2 实验方法

2.1 实验平台和参数设置

本文以 GoogleColaboratory 和 Tensorflow 2.9 深度学习框架为实验平台,实验平台的环境基于 Linux 5.10 和 Python3.6, CPU 和图形处理器分别为 Intel(R) Xeon(R) CPU @ 2.30 GHz 与 Tesla P100-PCIE-16 GB。实验中将经过处理的约 9 500 张 ODIR 图像数据按照 6:2:2 的比例划分为训练集、验证集和测试集。本文使用文献[22]中的交叉熵损失函数作为训练损失函

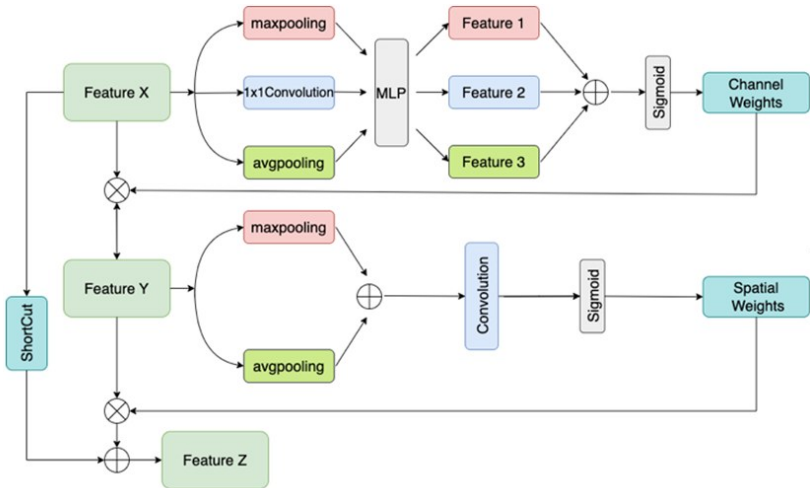


图5 混合域注意力机制的实现

Figure 5 Implementation of hybrid domain attention mechanism

数,使用Adam作为优化器对模型训练过程进行优化<sup>[23]</sup>,初始学习率(LearningRate)参数设置为0.000 1,批量大小(BatchSize)设置为64,训练迭代次数(Epoch)设为300,在训练迭代过程中,模型自动保存准确率最高的Epoch模型权重,同时使用EarlyStop早停策略<sup>[24]</sup>,patience设为10,监视指标为Val\_loss,当Val\_loss经过10个Epoch不再下降时停止训练,加入EarlyStop可以一定程度上避免模型的过拟合,使模型达到良好的收敛效果。

2.2 实验数据

本文采用南开大学与北京上工医信科技有限公司联合实验室推出的ODIR数据集作为实验数据,该数据集是目前世界上最大、公开可用的多类别眼底疾病识别数据集,它包含5 000名患者,涵盖所有年龄段,包括正常眼和7种不同眼底疾病类型的10 000张彩色眼底图像。7类眼底疾病包括糖尿病视网膜病变(D)、白内障(C)、青光眼(G)、年龄性相关病变(A)、近视(M)、高血压(H)和其他疾病(O)。本文选择占比较大的正常眼(N)和糖尿病视网膜病变(D)、白内障(C)、青光眼(G)3类典型眼底疾病作为实验对象。

2.3 性能评价指标

实验中采用准确率(Accuracy)、精确率(Precision)、F1分数(F1-scores)以及工作特征曲线ROC曲线与横坐标轴围成的面积(Area Under Curve, AUC)等医学图像分类中常用的指标<sup>[25]</sup>对本文方法的性能进行评价,各指标的计算公式如式(5)~式(7)所示。AUC取值为0.5~1.0,AUC值越大,检测方法真实性越高,分类效果越好。

$$\text{Accuracy} = \frac{(\text{TP} + \text{TN})}{(\text{TP} + \text{TN} + \text{FP} + \text{FN})}$$

(5)

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

(6)

$$\text{F1-scores} = \frac{2\text{TP}}{(2\text{TP} + \text{FP} + \text{FN})}$$

(7)

其中,TP代表正类判定为正类,FP代表负类判定为正类,TN代表负类判定为负类,FN代表正类判定为负类。

3 结果与分析

3.1 眼底疾病诊断实例展示

为了直观说明本文方法的诊断性能和效果,图6展示了12幅具有不同眼底疾病标签的彩色眼底图像,利用本文方法获得的诊断结果。各图像下方标出了该眼底疾病图像的真实标签、预测标签和置信度。从图6可以看出,绝大多数图像的预测标签与真实标签是一致的,且置信度绝大多数在0.99以上。实例结果直观表明,本文提出的眼底疾病诊断方法是可行和有效的。

3.2 实验结果

采用ODIR数据集对本文方法进行实验,为了检验本文方法的性能优劣,实验同时给出了基线网络方法、基线网络+多尺度特征融合方法、基线网络+混合域注意力机制方法以及本文方法4种眼底疾病诊断方法的实验结果。为了叙述方便,本文将上述4种方法分别命名为Backbone方法、BackboneA方法、BackboneB方法和本文方法。图7给出了上述4种方法的训练(蓝虚线)和验证(红实线)准确率变化曲线。由图7可以看出,4种方法的训练准确率区别不大,但验证集准确率的差异比较明显。(1)Backbone方法迭代30次后验证准确率很难得到更加明显的提高,且波动幅度较大,验证准确率平均为88.0%;(2)

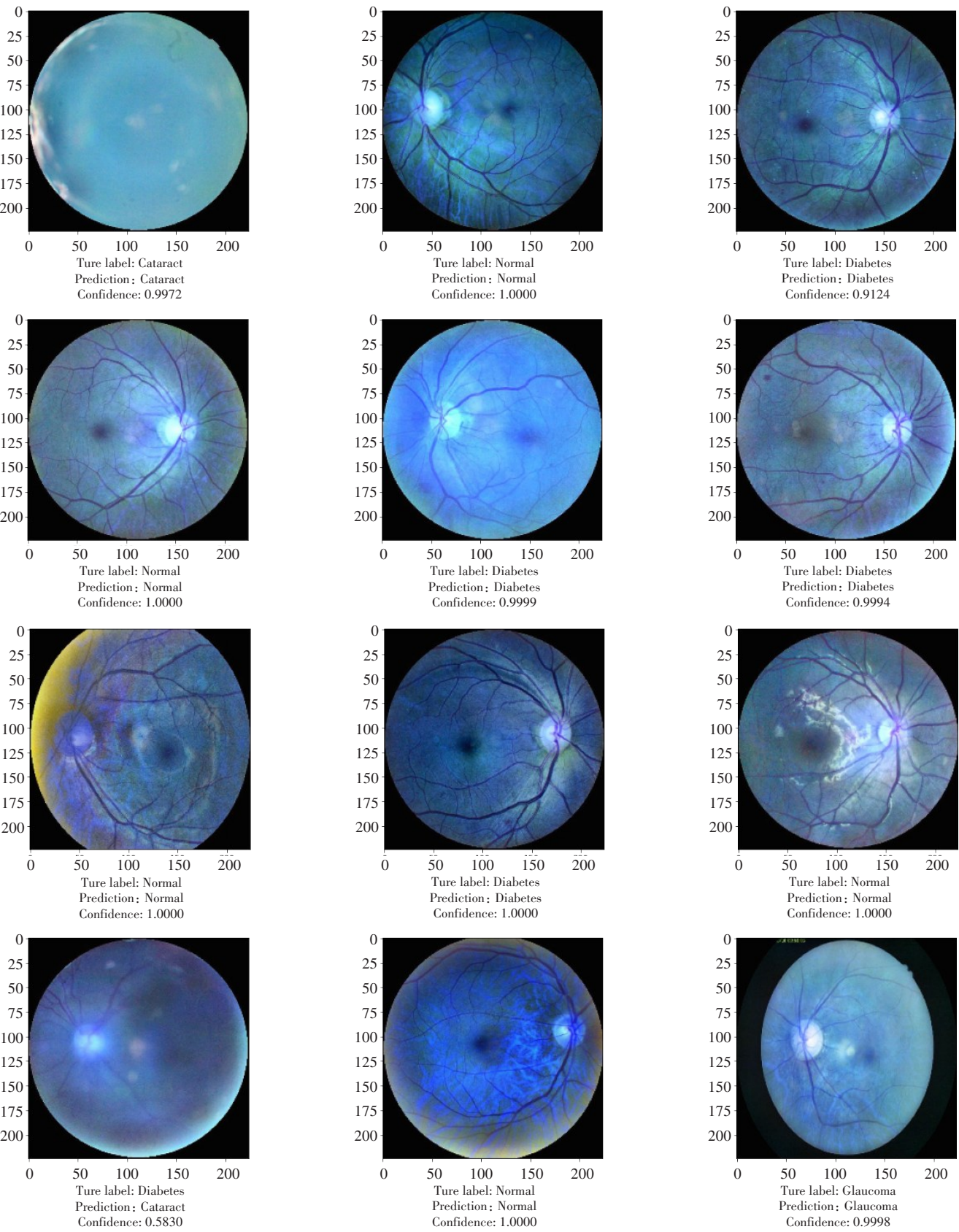


图 6 眼底疾病诊断示例

Figure 6 Examples of fundus disease diagnosis

BackboneA 方法验证准确率曲线波动幅度较小,验证准确率平均为 88.8%,准确率提升比较有限;(3) BackboneB 方法准确率曲线波动幅度比较明显,但验证准确率平均为 89.0%,准确率提升同样较小;(4)本



文方法所得验证准确率曲线波动幅度则相对很小,随着迭代次数的增加,验证准确率逐步上升,最后趋

于稳定,验证准确率达到93.2%。与其它3种方法相比,诊断准确率提升十分显著。

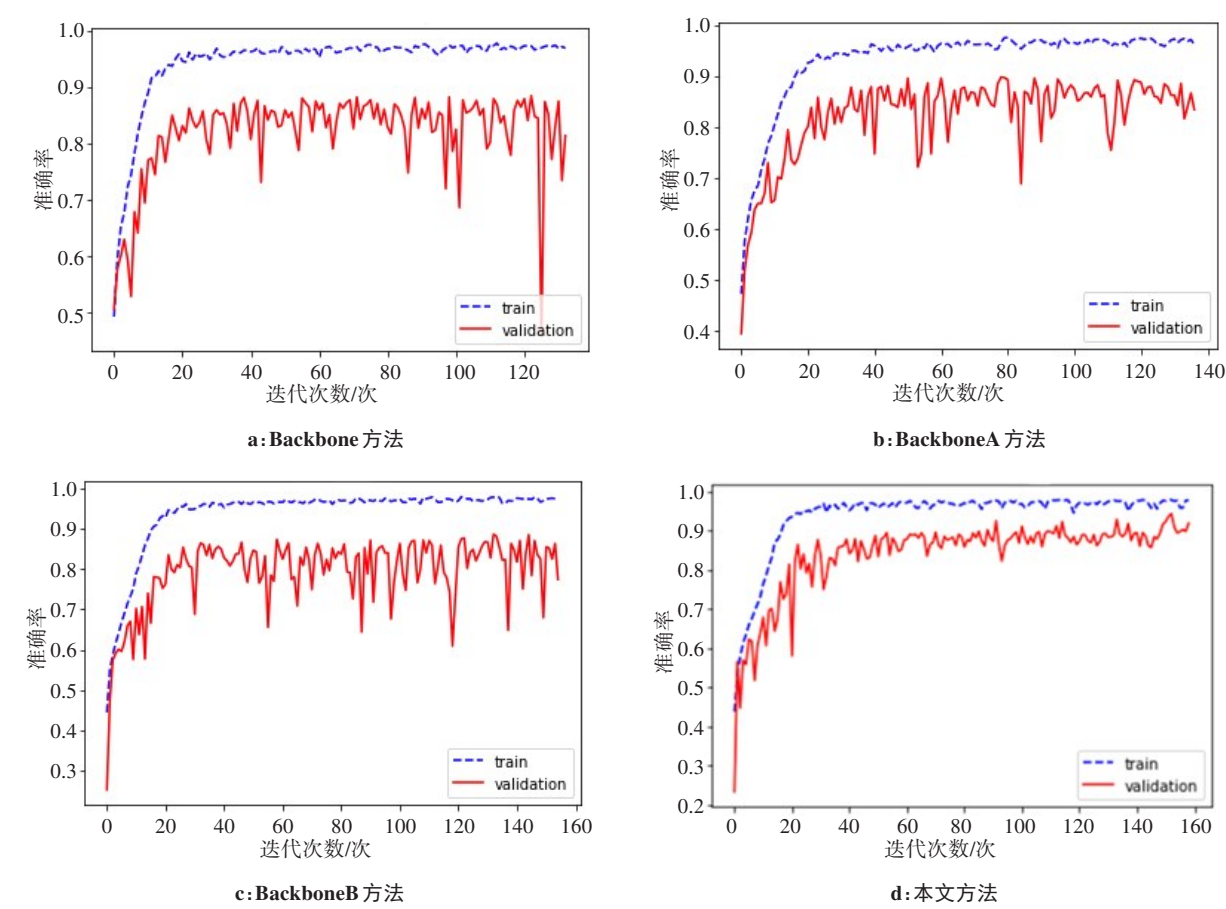


图7 4种不同方法的训练(蓝虚线)和验证(红实线)准确率曲线  
Figure 7 Training (dashed blue line) and validation (solid red line) accuracy curves for 4 different methods

为了更好地观察和分析 Backbone、BackboneA、BackboneB 及本文方法的诊断性能,图 8 给出 4 种方法在测试集上的实验结果混淆矩阵,混淆矩阵的纵坐标表示真实值,横坐标表示预测值,整个坐标系分为 16 个区域,上面的数字为诊断结果数据。对角线上的数值越大,所在区域的颜色越深,代表 TP(正类判定为正类)和 TN(负类判定为负类)的诊断正确率越高,该方法的诊断性能越好。对角线外的数据则代表错分率,例如图 8a 矩阵中第 2 行第 4 列的数值为 0.06,则表示将白内障(C)错误诊断为青光眼(G)的比例为 6%。从图 8d 可以看出,本文方法对正常眼(N)、白内障(C)糖尿病视网膜病变(D)和青光眼(G) 3 类典型眼底疾病的诊断准确率分别为 0.95、0.94、0.92 和 0.90,均达到 90% 以上,本文方法对于 3 类眼底疾病诊断分类性能明显优于其它 3 种方法。

为了更加全面评价上述不同方法的诊断性能,

表 1 给出各种方法的准确率、精确率、F1 分数、AUC 和运行时间等性能数据。从表 1 可以看出,与基线网络 Backbone 方法相比,融合多尺度特征模块的 BackboneA 方法诊断准确率提高 0.8%,增加混合域注意力机制模块的 BackboneB 方法准确率提高 1%,说明多尺度特征融合和混合域注意力机制的引入对提高眼底疾病诊断性能是有效的,可在一定程度上增强模型的特征提取能力,提升模型的诊断分类性能。本文方法由于同时结合了多尺度特征融合和混合域注意力机制,对诊断网络模型进行改进和重构,其性能得到明显的提升,与其它 3 种方法相比,其准确率、精确率、F1 分数、AUC 和运行时间 5 个性能指标均为最优;与基线网络 Backbone 方法相比,其准确率提高 5.2%、精确率提高 5%、F1 分数提高 0.052、AUC 提高 0.021、运行时间减少 25 min,获得了十分优秀的诊断结果。

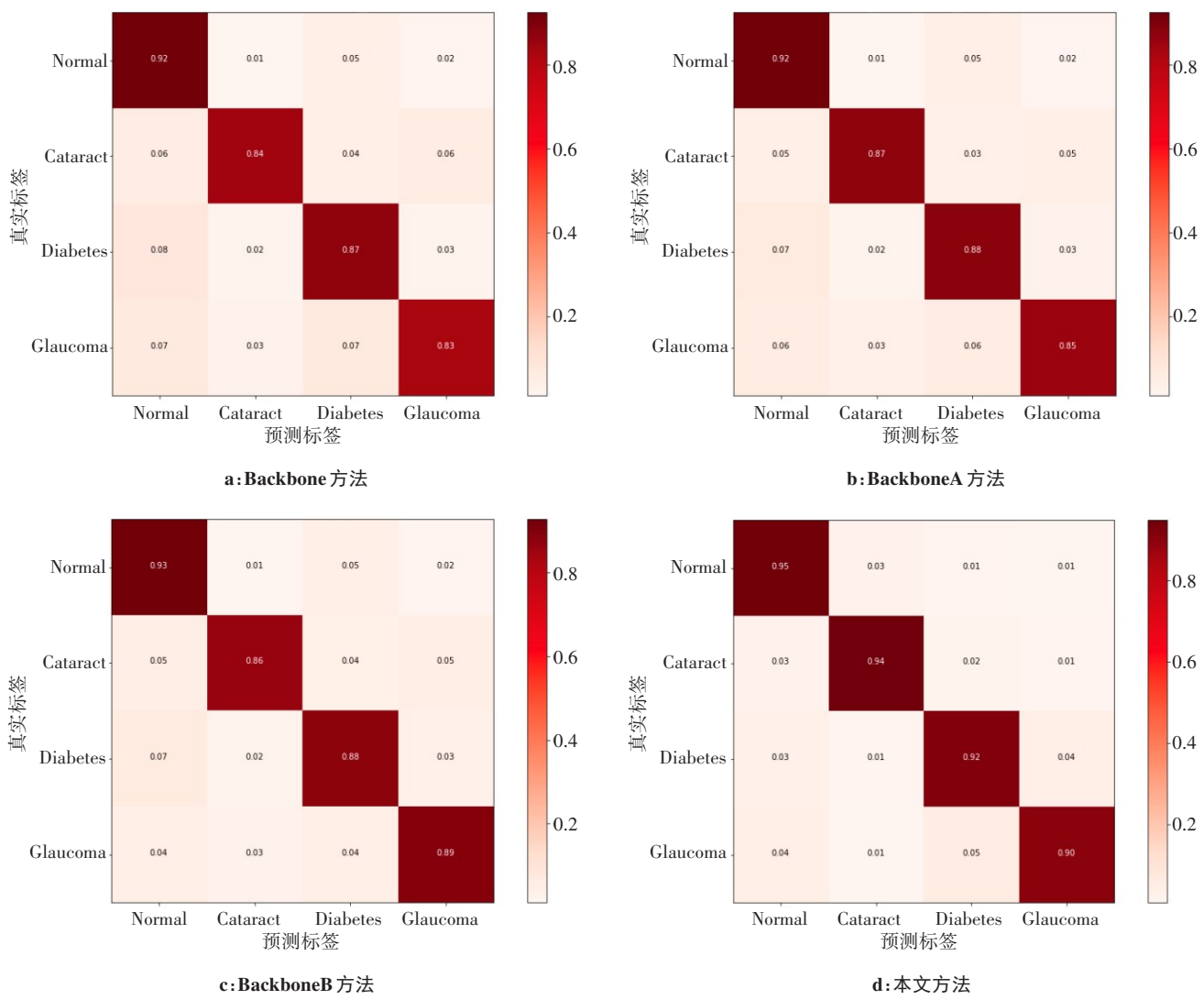


图 8 诊断结果混淆矩阵

Figure 8 Confusion matrices of diagnosis results

表 1 不同方法的性能数据

Table 1 Performance of different methods

方法	准确率/%	精确率/%	F1 分数	AUC	运行时间/min
Backbone 方法	88.0	88.4	0.880	0.966	195
BackboneA 方法	88.8	88.9	0.889	0.969	210
BackboneB 方法	89.0	89.1	0.891	0.973	225
本文方法	93.2	93.4	0.932	0.987	170

**4 结 语**

本文针对眼底疾病图像的特点提出多种改进措施,提出一种融合多尺度特征和混合域注意力机制的眼底疾病诊断网络模型方法,并在 ODIR 眼底疾病数据集上完成了对正常眼和 3 类常见眼底疾病的诊断分类。实验结果表明:(1)在模型训练前,对 ODIR 数据集进行数据清洗、图像增强和数据增广等预处

理操作,可以改善样本数据质量,有助于提高模型的诊断性能;(2)在基线网络中融合多尺度特征模块,有助于增强模型对多尺度特征特别是微小特征的提取能力;(3)增加混合域注意力机制模块,有助于模型获取更多有关疾病的的重要信息,有利于提取疾病的关键特征。本文方法同时结合了多尺度特征融合和混合域注意力机制,并对诊断网络模型进行改进和重构,其性能得到明显的提升,但本文提出的眼



底疾病诊断方法仍然存在一定的误检问题,为了获得更好的诊断性能,需要采集数量更多、质量更好的专业眼底疾病图像数据,同时也需要对本文方法做出进一步的研究和改进。

## 【参考文献】

- [1] Cui Y, Zhang M, Zhang L, et al. Prevalence and risk factors for diabetic retinopathy in a cross-sectional population-based study from rural southern China: Dongguan eye study[J]. *BMJ Open*, 2019, 9(9): e023586.
- [2] Du YF, Liu HR, Zhang Y, et al. Prevalence of cataract and cataract surgery in urban and rural chinese populations over 50 years old: a systematic review and meta-analysis[J]. *Int J Ophthalmol*, 2022, 15(1): 141.
- [3] Jalil NA, Hwang HJ, Dawi NM. Machines learning trends, perspectives and prospects in education sector [C]//Proceedings of the 3rd International Conference on Education and Multimedia Technology. 2019: 201-205.
- [4] Muchuchuti S, Viriri S. Retinal disease detection using deep learning techniques: a comprehensive review[J]. *J Imaging*, 2023, 9(4): 84.
- [5] 王霄, 朱恩照, 艾自胜. 卷积神经网络的原理及其在医学影像诊断中的应用[J]. *中国医学物理学杂志*, 2022, 39(12): 1485-1489.  
Wang X, Zhu EZ, Ai ZS. Principle of convolutional neural network and its applications in medical imaging diagnosis[J]. *Chinese Journal of Medical Physics*, 2022, 39(12): 1485-1489.
- [6] 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.
- [7] Karri SP, Chakraborty D, Chatterjee J. Transfer learning based classification of optical coherence tomography images with diabetic macular edema and dry age-related macular degeneration[J]. *Biomed Opt Express*, 2017, 8(2): 579-592.
- [8] Muhammad H, Fuchs TJ, De Cuir N, et al. Hybrid deep learning on single wide-field optical coherence tomography scans accurately classifies glaucoma suspects[J]. *J Glaucoma*, 2017, 26(12): 1086.
- [9] Grassmann F, Mengelkamp J, Brandl C, et al. A deep learning algorithm for prediction of age-related eye disease study severity scale for age-related macular degeneration from color fundus photography[J]. *Ophthalmology*, 2018, 125(9): 1410-1420.
- [10] Zhang W, Li D, Wei Q, et al. The validation of deep learning-based grading model for diabetic retinopathy[J]. *Front Med (Lausanne)*, 2022, 9: 1-8.
- [11] Hodson H. DeepMind and Google: the battle to control artificial intelligence[J]. *Economist*, 2019: 0013-0613.
- [12] De Fauw J, Ledsam JR, Romera-Paredes B, et al. Clinically applicable deep learning for diagnosis and referral in retinal disease[J]. *Nat Med*, 2018, 24(9): 1342-1350.
- [13] Chelaramani S, Gupta M, Agarwal V, et al. Multi-task knowledge distillation for eye disease prediction[C]//Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 2021: 3983-3993.
- [14] Jidan OJ, Susmoy P, Roy A, et al. A comprehensive study of DCNN algorithms-based transfer learning for human eye cataract detection [J]. *Int J Adv Comput Sci Appl*, 2023, 14(6).
- [15] Sahu S, Singh AK, Ghrera SP, et al. An approach for de-noising and contrast enhancement of retinal fundus image using CLAHE[J]. *Opt Laser Technol*, 2019, 110: 87-98.
- [16] Chlap P, Min H, Vandenberg N, et al. A review of medical image data augmentation techniques for deep learning applications[J]. *J Med Imaging Radiat Oncol*, 2021, 65(5): 545-563.
- [17] Das V, Dandapat S, Bora PK. Multi-scale deep feature fusion for automated classification of macular pathologies from OCT images[J]. *Biomed Signal Process Control*, 2019, 54: 101605.
- [18] Fiebelkorn IC, Kastner S. A rhythmic theory of attention[J]. *Trends Cogn Sci*, 2019, 23(2): 87-101.
- [19] Bastidas AA, Tang H. Channel attention networks[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. 2019: 881-888.
- [20] Zhao H, Zhang Y, Liu S, et al. Psanet: point-wise spatial attention network for scene parsing [C]//Proceedings of the European Conference on Computer Vision (ECCV). 2018: 267-283.
- [21] Gu Z, Li Y, Wang Z, et al. Classification of diabetic retinopathy severity in fundus images using the vision transformer and residual attention[J]. *Comput Intell Neurosci*, 2023: 1-12.
- [22] Ruby U, Yendapalli V. Binary cross entropy with deep learning technique for image classification[J]. *Int J Adv Trends Comput Sci Eng*, 2020, 9(10): 5393-5397.
- [23] Desai C. Comparative analysis of optimizers in deep neural networks [J]. *Int J Innov Sci Res Technol*, 2020, 5(10): 959-962.
- [24] Corneanu C, Madadi M, Escalera S, et al. Explainable early stopping for action unit recognition[C]//International Conference on Automatic Face and Gesture Recognition. 2020: 693-699.
- [25] Wardhani NW, Rochayani MY, Iriany A, et al. Cross-validation metrics for evaluating classification performance on imbalanced data[C]//2019 International Conference on Computer, Control, Informatics and Its Applications. 2019: 14-18.

(编辑:陈丽霞)