

基于深度卷积集成网络的视网膜多种疾病筛查和识别方法

王禾扬, 杨启鸣, 朱旗

(南京航空航天大学计算机科学与技术学院, 江苏 南京 211100)

摘要: 针对视网膜疾病种类繁多、病灶位置不固定等特点, 提出一种基于深度卷积集成网络的视网膜多种疾病筛查和识别方法。首先, 根据视网膜眼底图像裁剪掉两侧黑色边框, 并去除图像中的噪声, 以降低对眼底图像的干扰, 提高图像的清晰度; 之后, 通过对处理完成的视网膜眼底图像使用裁剪、旋转等数据增强方法来扩增数据集; 再建立基于深度卷积神经网络的模型进行特征提取, 并在网络模型微调后完成视网膜疾病筛查和识别任务, 最终将多个模型的结果进行集成。实验结果表明, 该方法针对视网膜疾病的筛查和识别的问题取得了较好的效果, 视网膜疾病筛查的准确率达到 96.05%, 视网膜疾病识别的准确率达到 72.55%。

关键词: 视网膜眼底图像; 疾病筛查; 疾病识别; 深度卷积网络; 集成模型

中图分类号: TP391.4

文献标识码: A

doi: doi:10.11959/j.issn.2096-6652.202127

Retinal multi-disease screening and recognition method based on deep convolution ensemble network

WANG Heyang, YANG Qiming, ZHU Qi

College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 211100, China

Abstract: As for the characteristics of various types of retinal diseases and uncertainty of the location of the lesions, a retinal multi-disease screening and recognition method based on deep convolutional ensemble network was proposed. Firstly, the black borders on both sides of the retinal fundus image were cut off, and the noise in the image was removed to reduce the interference to the retinal image and increase the clarity of the image. After that, data augmentation methods such as cropping and rotating were performed to process retinal fundus image to amplify the dataset. Then, a model based on deep convolutional neural network was built for feature extraction, and the network model was fine-tuned to complete the task of screening and identifying retinal diseases. Finally, the results of multiple models were ensemble. The experimental results show that this method has achieved good results for the screening and recognition of retinal diseases, the accuracy of retinal disease screening is 96.05%, and the accuracy of retinal disease recognition is 72.55%.

Key words: retinal fundus image, disease screening, disease recognition, deep convolutional network, ensemble model

1 引言

视网膜疾病发病率高, 且相关疾病种类繁多。据世界卫生组织报告, 世界视力障碍者总数达 22 亿人, 其中至少 10 亿个病例可以预防或有待解决^[1]。眼部疾病的早期发现和治疗可以很好地预防视力问

题带来的障碍。目前, 人工智能技术在视网膜疾病中最常见的应用是通过彩色眼底数码相机采集患者眼底图像^[2], 将眼部生理结构、病灶特征和颜色变化以清晰的图像展现出来, 医生可以利用这种非侵入式的方法对视网膜疾病做出诊断。然而, 由于病患众多, 仅依靠人工进行眼部疾病的筛查和诊

收稿日期: 2021-06-12; 修回日期: 2021-07-08

通信作者: 朱旗, zhuqinuaa@163.com

基金项目: 中央高校基本科研业务费专项资金项目 (No.NT2020024)

Foundation Item: Fundamental Research Funds for the Central Universities (No. NT2020024)

断会给医生带来极大的工作压力,提高误诊率。因此,计算机视觉技术的发展为眼科医生利用计算机辅助诊断带来了可能。考虑到病人的差异性,眼科医生对患者进行初步诊断的阶段面临诸多困难:

- 视网膜疾病种类繁多,需要医生具备丰富的诊断经验;
- 病灶区域、特征不同,会延长医生的诊断时间;
- 图像采集过程中,光线的影响会对成像质量产生干扰,增加医生工作难度。

近年来,随着深度学习在计算机视觉和模式识别领域的成功发展,基于眼底图像的疾病筛查和识别方法不断涌现。Pratt H 等人^[3]以及 Colas E 等人^[4]提出了使用神经网络的方法对 5 000 位患有糖尿病性视网膜病变(diabetic retinopathy, DR)患者的 10 000 张眼底图像进行特征提取,将 DR 患病阶段分为 4 类:轻度病变、中度病变、重度病变和扩散,为计算机辅助 DR 的诊断做出了贡献;Worrall D E 等人^[5]提出基于深度学习的端到端系统对早产儿视网膜病变(ROP)进行早期诊断,他们对经过预训练的 GoogleNet 模型进行微调后,在私有数据集上对 ROP 疾病的诊断准确率达到 95.4%;Burlina P 等人^[6]提出使用 ImageNet 预训练权重的深度卷积神经网络对年龄相关性黄斑变性(age-related macular degeneration, ARMD)正常、早期、中期、晚期的 4 个阶段进行识别,基于该疾病的特性,深度卷积神经网络从眼底图像中心区域的多个同心矩形中提取特征进行训练,并取得了较好的识别效果。

以上工作为视网膜疾病中的某个单一疾病或某个疾病的不同阶段的诊断做出了突出的贡献。实际上,视网膜疾病的筛查和识别并非针对单一疾病,且同一患者可能同时患有多种疾病,这就涉及对多种疾病进行识别和诊断的问题。区别于对单一疾病进行诊断的模型,针对多种疾病进行筛查和识别的模型更有助于辅助医生做初步判断。其中,参考文献[7-10]使用神经网络的方法在各种疾病图像中提取病灶特征,根据相关疾病的特点设计网络结构,实现对多种疾病的识别。针对视网膜眼底图像,Arunkumar R 等人^[11]提出了一种通过深度学习提取视网膜疾病特征后进行特征降维的方法,使用支持向量机(support vector machine, SVM)分类器对老年性黄斑变性和糖尿病性视网膜病变两

种视网膜疾病进行识别分类;Wang X 等人^[12]使用集成学习的方法对视盘、黄斑和整个视网膜区域分别训练分类模型,并集成最终结果,该方法在糖尿病性视网膜病变和青光眼疾病的识别上有较好的效果。Choi J Y 等人^[13]使用 VGG (visual geometry group)网络和迁移学习的方法对视网膜疾病进行分类,疾病种类少时效果较好,但是在对 10 种视网膜疾病进行分类时,准确率仅为 36.7%。以上这些方法仅适用于对两种或极少种类的疾病进行识别,存在任务难度较小或识别效果不理想等问题。

前人的工作表明,利用深度神经网络对病灶特征进行提取并对疾病进行识别已经成为当前主流的方法,相较于传统机器学习方法,无论是对病灶特征提取的能力,还是对疾病识别的效果,都有极大的提升。但是针对多种疾病的筛查和识别工作,尤其是疾病种类较多时,模型的识别能力仍待改善。基于此,本文做出的贡献有以下几点:

- 只使用一个模型,在仅有 2 500 多个样本的情况下对多达 28 种疾病进行筛查和识别;
- 由于数据集样本数量极不平衡,本文采用分组多任务的方法同时参与网络更新;
- 与以往的工作相比,笔者用更少的成本达到了更好的性能。

2 数据预处理

2.1 黑框裁切

使用眼底相机采集视网膜眼底图像时,因设备的不同和患者间的差异性,采集到的图像存在问题:视网膜眼底图像分辨率不同、眼底在整张图像中的位置不固定、眼底尺寸存在细微差异。另外,眼底图像均存在大量冗余的黑色区域,这些黑色区域并非视网膜区域。因此,为了尽量减少黑框对提取特征过程造成的不必要的干扰,需要去除多余的黑色边框,裁剪出感兴趣区域,并将图像调整为统一尺寸。

为了找到眼底区域,可以通过图像的 RGB 数值定位眼底的边缘位置。眼底边缘位置定位方式如下:

$$\text{sum}(R, G, B) - 3 \times \text{margin} > 0 \quad (1)$$

其中, R 、 G 、 B 为每行(列)的 RGB 色彩空间数值,其取值范围为[0,255]。受光照及其他因素影响,靠近眼底边缘处的一部分区域从 RGB 数值上看并

非完全黑色，若尽可能多地去除黑色区域，则需要减去一个人为设置的阈值（margin），本文设定为5。根据式（1）可从上、下、左、右分别定位眼底边缘位置，再从原图中去除黑色边框，裁剪出感兴趣区域。

2.2 数据增强

深度学习模型的训练需要海量的数据做支撑，而一些罕见的视网膜疾病难以获得足够的训练数据^[14]，深度学习在医学图像处理上的应用使得这一问题尤为凸显。在仅有少量数据集的条件下，训练一个效果良好、泛化性能较强的模型需要根据现有的少量数据进行数据增强，以满足深度神经网络模型对大数据量的需求。本文在训练前和训练过程中使用不同的数据增强方法，有效解决了数据量不足的问题。

首先，在训练前，对输入图像采用图形变换的方式进行数据增强。使用的方法包括随机裁剪、垂直和水平方向翻转、随机角度旋转、增加高斯噪声、限制对比度自适应直方图均衡化（CLAHE）等。其中，Zuiderveld K^[15]提出的 CLAHE 方法是对直方图均衡化的改进，通过将原始图像的灰度直方图划分成若干个子区域，对子区域分别进行受对比度限制的直方图均衡化，再利用双线性差值消除各个子区域间比较明显的截断部分，得到一张对比度增强的图像。这一方法使一些受光照条件影响而过暗或过亮的图像获得更加清晰的成像效果。

另外，在训练过程中引入了 Yun S 等人^[16]提出的 CutMix 数据增强方法，即将每个小批量中的图像进行随机拼接组合，并加权计算总损失。其原理如图1所示。

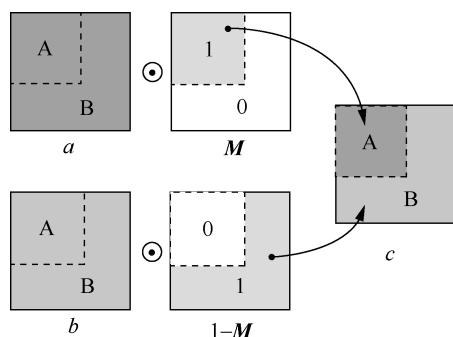


图1 CutMix 数据增强方法原理

CutMix 数据增强方法步骤如下。

步骤1：在一个轮次的每个批量中，采用不放回的方式随机抽取两个图像样本和对应的标签，记为 a 、 b ；

步骤2：随机设置一个概率值 p 。

步骤3：从 a 中截取占比为 p 的局部图像 A ，记录边框对角位置 x_1, y_1, x_2, y_2 ，剩余部分记为 B ；

步骤4：利用 x_1, y_1, x_2, y_2 将 a 中的 B 区域替换为 b 中的 B 区域，组成新的图像 c ，其中 B 区域占比为 $1-p$ ；步骤3、步骤4可用式（2）表示：

$$c = M \odot a + (1 - M) \odot b \quad (2)$$

步骤5：将新图像 c 输入模型，并计算总损失，计算方法如下：

$$\text{Loss} = p \times \text{loss}(f(c), Y_a) + (1 - p) \times \text{loss}(f(c), Y_b) \quad (3)$$

其中，式（2）中 M 为与 a 和 b 同尺寸的掩码矩阵， $\odot$ 表示点乘，即在掩码矩阵中0为被裁剪的区域，1为需要保留的区域。式（3）中 loss 表示损失函数， $f(c)$ 为将新图像 c 输入模型后的输出， Y_a 和 Y_b 分别为图像 a 和图像 b 的标签。

使用该方法，在一次训练过程中拼接两张不同的图像，可以在局部信息中识别出两个不同的目标，从而提高模型的识别能力和泛化能力，达到数据增强的效果。

3 基于深度卷积集成网络的视网膜多种疾病筛查和识别模型

本文将 EfficientNet 作为主干网络进行特征提取，再对整个网络模型进行微调，分别完成视网膜疾病的筛查和识别任务。最后，多个网络模型进行集成，并将最终的结果用于疾病筛查和识别。

3.1 目标函数

在疾病筛查任务中，需要根据被试的眼底图像判断该被试是否有患病风险，因此，可以将疾病筛查问题看作二分类问题。将交叉熵（cross entropy）损失作为多种疾病筛查任务的损失函数：

$$L_{\text{screen}}(Y, p(X)) = -Y \times \log(p(X)) - (1 - Y) \times \log(1 - p(X)) \quad (4)$$

为了给患者的患病种类做出判断，需要根据患者的眼底图像对患病的概率做出预测，而同一患者可能会患有多种疾病，因此，疾病识别任务实际为多标签分类问题。将二分类交叉熵损失作为多种疾病识别任务的损失函数：

$$L_{\text{rec}}(Y, f_2(X)) = -\frac{1}{N} \sum_{i=1}^N (Y \times \log \sigma(f_2(X)) + (1 - Y) \times \log(1 - \sigma(f_2(X)))) \quad (5)$$

综上所述, 最终的损失函数可表示为:

$$\text{Loss} = \alpha \times L_{\text{screen}}(Y, f_1(X)) + \beta \times L_{\text{rec}}(Y, f_2(X)) \quad (6)$$

在式(4)~式(6)中, X 为输入的视网膜眼底图像, Y 为相应图像的标签, $f(*)$ 为主干网络对图像信息编码后预测为正类的概率, N 为疾病类别数量, $\sigma(*)$ 为 Sigmoid 函数, α 和 β 为超参数, 用于平衡两个任务损失函数所占比重。

3.2 特征提取

对视网膜眼底图像进行处理时, 采用卷积的方式提取图像中的信息, 通过叠加多种卷积层构建一个深层的卷积神经网络。相较于传统的机器学习方法, 深层卷积神经网络可以自动完成特征提取, 并达到更好的特征提取效果。经过近几年深度卷积神经网络的发展, 优秀的神经网络往往有以下特点:

- 使用残差结构, 以增加神经网络的深度, 提高特征提取能力;
- 在神经网络的单元结构中增加特征提取层数, 即增加神经网络的宽度;
- 调整神经网络的参数, 以适应更大分辨率图像的输入, 使得网络模型可学习更精确的内容。

本文将谷歌大脑团队提出的 EfficientNet^[17]作为模型的主干网络部分进行特征提取, 该网络可对网络的深度、宽度和输入的分辨率进行适当的调整, 以达到更好的特征提取效果。

在 EfficientNet 中, 最核心的 MBConv 模块结构如图 2 所示。该模块源于 MobileNet-V3^[18]中的倒置残差模块 (inverted residual block), Tan M X 等人^[17]对其进行改进, 采用了 Swish 激活函数^[19], 并引入了 Hu J 等人^[20]在 SENet 中提出的 SE(squeeze-and-excitation) 模块, 将通道注意力机制应用于特征提取过程中。Swish 激活函数如式 (7) 所示:

$$f(x) = \frac{x}{1 + e^{-\beta x}} \quad (7)$$

其中, x 为激活函数的输入数据, β 为常数或可训

练的参数。

MBConv 模块按照以下流程进行计算:

$$f_1 = \text{Swish}(\text{BN}(\text{Conv2d}(X, 1))) \quad (8)$$

$$f_2 = \text{Swish}(\text{BN}(\text{DepthConv2d}(f_1, 3))) \quad (9)$$

$$f_{\text{SE}} = \text{Conv2d}(\text{Conv2d}(\text{AveragePool}(f_2), 1)) \quad (10)$$

$$f_3 = \text{BN}(\text{Conv2d}(f_{\text{SE}} \otimes f_2, 1)) \quad (11)$$

$$\text{output} = X + f_3 \quad (12)$$

首先, MBConv 模块使用 1×1 卷积 (Conv2d) 对输入做升维操作得到特征图 f_1 , 利用 3×3 或 5×5 的通道维度卷积 (DepthConv2d) 得到相同维度的特征图 f_2 , 其中, 卷积操作后含有归一化层 (BN) 和激活层 (Swish); 特征图 f_2 经过 SE 模块中 1×1 卷积和平均池化 (average pool) 后计算出通道注意力权重 f_{SE} , 再将 f_{SE} 与 f_2 相乘, 获得带有注意力的特征图 f_3 。最后, 经过全连接层后, 与原始输入的特征图进行残差连接, 在一定程度上缓解了随着网络深度的增加带来的网络退化和过拟合问题^[21], 最终输出新的特征图 (output)。

从 EfficientNet 的 B0 到 B7, 通过对 MBConv 模块的叠加和网络参数的调整, 构成了不同深度和宽度的网络结构, 输入图像的分辨率随之增大, 特征提取的能力不断增强。本文将采用 EfficientNet-B5、EfficientNet-B6、EfficientNet-B7 所得到的结果进行后续工作。

3.3 网络微调

在使用深度学习对医学图像进行处理的过程中, 常面临数据集的数据量过少的问题, 而且深层的主干网络含有大量的网络参数, 很容易出现过拟合和泛化能力差的情况。为了缓解这一情况, 本文采用了迁移学习的方法, 在模型的训练开始前加载公开的预训练网络权重, 再将训练集数据输入网络进行训练。视网膜眼底图像经过主干网络后, 学习到的特征被提取出来, 因此, 若将提取出的特征用

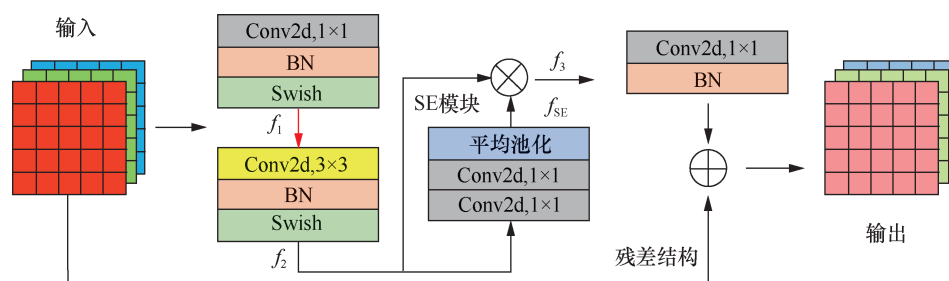


图2 MBConv 模块结构

于多种疾病的筛查任务和识别任务，需要对网络进行微调，微调网络的结构如图3右侧所示。

首先，将主干网络输出的特征图用展平（flatten）操作拉伸成一维向量，经过中间的隐藏层（包括线性映射层 Linear、归一化层 BN 和 ReLU 激活层）增加网络的非线性能力后，输出维度为对应疾病类别数量的向量，用于对多种疾病的患病概率进行预测，并根据该向量计算多种疾病识别损失 L_{screen} ；然后，按照上述类似步骤，根据多种疾病患病概率完成疾病筛查任务，即预测患病风险，并计算疾病筛查损失 L_{rec} ；最后，按照合适的权重对两个损失进行加权，计算出总损失，用于对整个网络模型的反向传播更新网络参数。

3.4 模型集成

在实际的训练过程中，笔者在多种结构的模型上尝试对视网膜疾病进行分类，但并不是所有的模型都能够达到相同的效果，且不同的模型对不同疾病的分类效果也各有差别。因此，本文采用集成学习的方法，综合考虑不同模型的分类结果，使整体的疾病筛查和识别效果在一定程度上有所提升。

集成学习是一种被广泛应用于机器学习和深度学习的算法，即将多个训练器结合使用，其结果往往比单个训练器更优。常用的模型集成策略有以下几种。

- 直接平均法：

$$\text{Result} = \frac{1}{N} \sum_{i=1}^N S_i \quad (13)$$

- 加权平均法：

$$\text{Result} = \frac{1}{N} \sum_{i=1}^N W_i S_i \quad (14)$$

式（13）、式（14）中 N 为集成模型的数量， S_i 为第 i 个模型的预测结果， W_i 为第 i 个模型的权重。

- 投票法：投票法常使用多数表决法，记第 i 个模型对第 j 个类别的预测结果对应的标签为 L_{ij} ，那么所有模型对第 j 个类别的分类结果为 $R_j = \{L_{1j}, L_{2j}, \dots, L_{Nj}\}$ ，将 R_j 中多数投票结果作为最终分类结果。

- 堆叠法：堆叠法实际上通过构建新的学习器，将多个模型预测的结果作为一个新的数据集输入，得到最终预测结果。

如图4所示，本文选用直接平均法模型集成策略，对训练完成的 EfficientNet-B5、EfficientNet-B6、EfficientNet-B7 3 个模型预测结果取平均值，得到最终结果。

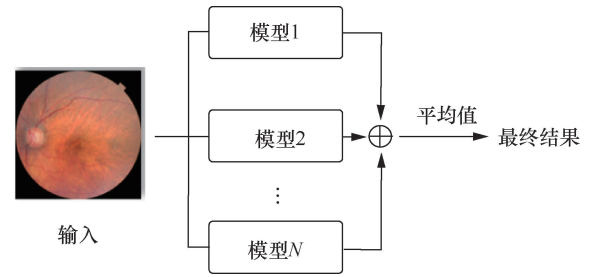


图4 本文使用的模型集成策略

4 实验结果

4.1 数据集

本文将 IEEE ISBI 2021^[22] 中的视网膜眼底图像作为数据集，该数据集分别使用 3 台数码眼底照相机采集，共 2 560 张视网膜眼底图像，其中 1 920 张图像为训练集，640 张图像为测试集。这 2 560 张图像中共包含 3 种分辨率，分别为 4 288×2 848、2 144×1 424 和 2 048×1 536。此外，数据集包含 48 种视网膜疾病，其中较为罕见的疾病将被统一标记为“其他（OTHER）”类别，包括“其他”类别在内共有 28 种疾病，且每个图像样本为正常或至少包含一

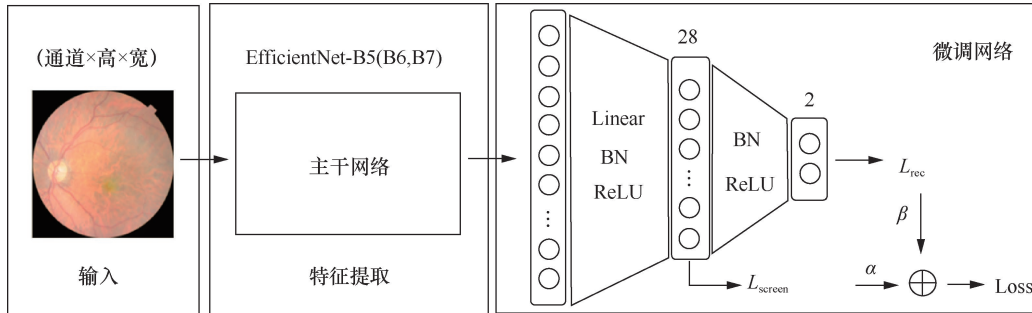


图3 基于深度卷积集成网络的视网膜多种疾病筛查和识别模型

种疾病。其余 27 种疾病如下：糖尿病性视网膜病变 (DR)、年龄相关性黄斑变性 (ARMD)、屈光介质混浊 (media haze, MH)、玻璃疣 (drusen, DN)、近视 (myopia, MYA)、视网膜静脉分枝阻塞 (branch retinal vein occlusion, BRVO)、镶嵌眼底 (tessellation, TSLN)、黄斑视网膜前膜 (macular epiretinal membrane, ERM)、激光瘢痕 (laser scar, LS)、黄斑瘢痕 (macular scar, MS)、中心性浆液性脉络膜视网膜病变 (central serous chorioretinopathy, CSR)、视神经盘凹 (optic disc cupping, ODC)、视网膜中央静脉阻塞 (central retinal vein occlusion, CRVO)、血管迂曲 (vascular circuitry) (原数据集用 TV 表示)、星状玻璃体病变 (asteroid hyalosis, AH)、视盘苍白 (optic disc pallor, ODP)、视盘水肿 (optic disc edema, ODE)、分流 (shunt, ST)、前部缺血性视神经病变 (anterior ischemic optic neuropathy, AION)、中心凹毛细血管扩张 (parafoveal telangiectasia, PT)、视网膜牵引 (retinal traction, RT)、视网膜炎 (retinitis, RS)、脉络膜视网膜炎 (chorioretinitis, CRS)、渗出 (exudation, EDN)、视网膜色素上皮改变 (retinal pigment epithelium change, RPEC)、黄斑裂孔 (macular hole, MHL)、色素性视网膜炎 (retinitis pigmentosa, RP)。

4.2 样本不平衡处理

对样本不平衡的处理是机器学习和深度学习不可避免的一个问题，然而在本文使用的数据集中，样本不平衡的问题尤为突出。在多种疾病筛查和识别问题中，每种疾病往往与其他疾病有一定联系，即各种疾病之间并非互斥关系。训练集中各种疾病的样本数量如图 5 所示。

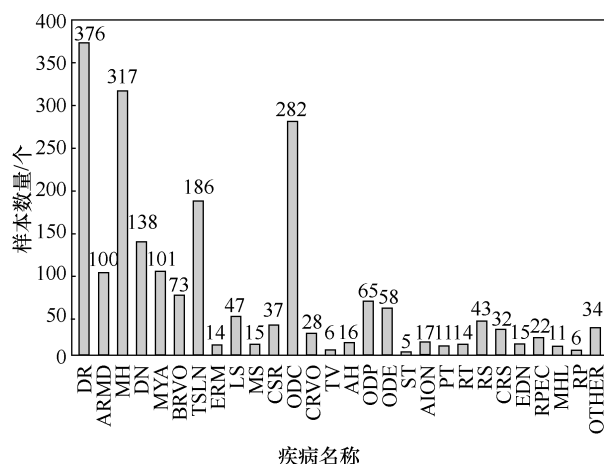


图5 训练集各种疾病样本数量

笔者分析各种疾病之间的关联性和单一疾病样本数量后，根据每种疾病样本数量分为[0,25]、(25,50]、(50,100]、(100,150]、(150,400] 5个区间，处于不同区间内的疾病将采用不同程度的复制、镜像、旋转等方法对样本进行扩充，结果如图 6 所示。

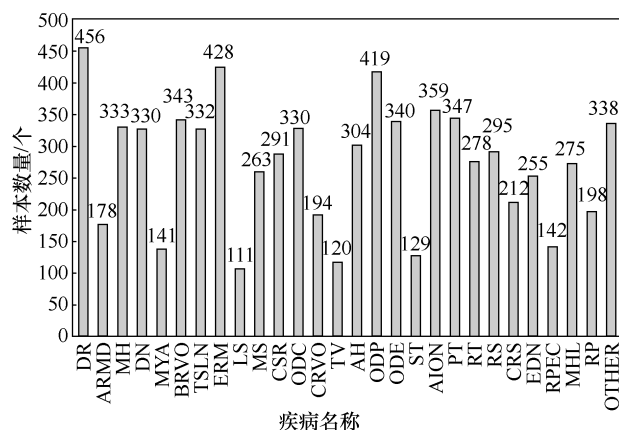


图6 样本不平衡处理结果

从图 6 可以看到，TV、ST 等疾病的样本数量有大幅度的增加，但由于 LS、TV、ST 与 DR、ODP 关联性较大，因此各疾病间样本数量无法实现绝对的平衡。

4.3 数据预处理结果

实验开始前，首先对数据进行预处理，具体方法如第 2 节所述。训练开始前，读入黑框裁切后的 1 920 张训练集图像，再对图像进行尺寸调整、随机裁剪、随机旋转、CLAHE 等数据增强方式扩充训练集，进行归一化处理后输入模型训练。训练过程中随机抽取每个轮次中的两张图像使用 CutMix 数据增强方法再次进行数据增强。黑框裁切的结果如图 7 所示。

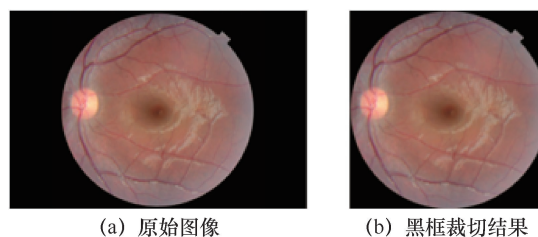


图7 黑框裁切结果

使用 CLAHE 数据增强方法的结果如图 8 所示。图 8 (a) 为原始图像，图 8 (b) 为限制对比的直方图均衡化处理后的效果。可以看到，处理后的图像对病灶区域的显像效果有了明显的增强，这将有助于模型对病灶特点进行学习。

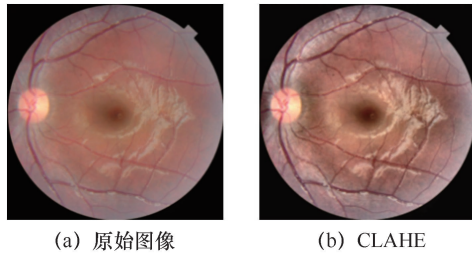


图8 CLAHE数据增强结果

使用 CutMix 数据增强方法的可视化结果如图 9 所示。其中，图 9 (a)、图 9 (b) 为随机抽取的两张图像，图像中的阴影部分为被裁剪出进行拼接的区域，图 9 (c) 为拼接后的图像，拼接后的图像及其标签将被输入模型进行训练。

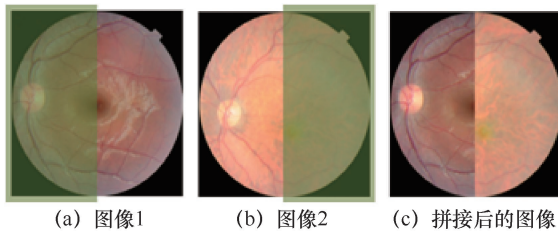


图9 CutMix数据增强结果

实验结果表明，预处理后的数据能减少原始图像中不必要的噪声，降低模型学习难度。此外，数据增强方法能够有效地扩充训练集的数量，有效解决了模型训练过程中数据量少的问题。

4.4 视网膜多种疾病筛查和识别结果

为了避免从零开始对深度卷积神经网络模型进行训练，在训练的初始阶段采用迁移学习的方法，加载经过大量数据预训练后得到的网络模型权重，以降低预训练的成本。

本文使用的目标函数如式 (6) 所示，将式 (4) 多种疾病筛查损失和式 (5) 多种疾病识别损失各

自分配合适的权重后得到的整体损失用于反向传播，更新网络模型的参数。考虑到多种疾病识别任务的难度相对较大，设置 $\alpha=0.4$, $\beta=0.6$ 。另外，实验过程中优化器首先使用 Adam 算法以达到快速收敛的目的，之后使用 SGD 算法继续对模型优化，以达到更好的泛化性能；训练初始学习率设置为 0.001，学习率衰减策略使用固定步长衰减，整个训练过程的轮次设置为 300，受设备影响，批大小设置为 8。

实验采用五折交叉验证方法，训练集中 85% 的样本用于模型训练，剩余 15% 的样本作为验证集用于测试模型的泛化性能，选择训练过程中泛化性能最好的模型进行下一步的测试。在多种疾病筛查任务中，实验评价指标选择准确率 (ACC) 和 ROC 曲线下面积 (AUC)，在多种疾病识别任务中实验评价指标选择准确率 (ACC) 和各类别平均准确率的均值 (mAP)。在测试阶段，采用测试时增强和伪标签的策略，抽取置信度较高的测试样本加入训练集中不断优化网络模型。实验结果表明，该方法对多种疾病的筛查和识别性能有一定的提升。将每个网络模型五折交叉验证的平均结果作为该网络模型的预测结果，再根据式 (13) 中的方法集成各个网络模型，最终的结果为多种疾病筛查和识别的结果。实验结果见表 1。

实验结果表明，本文提出的基于深度卷积集成网络的视网膜多种疾病筛查和识别方法相较于主流的深度学习神经网络有较好的效果提升。如图 10、图 11 所示，本文方法对 EfficientNet 的 3 个模型根据任务进行网络微调后，采用集成、测试增强、打伪标签等策略对训练过程进行优化，在多种疾病筛查任务中准确率达到 96.05%，多种疾病识别任务中准确率达到 72.55%，相

表1 本文方法和对比方法的实验结果

方法	多种疾病筛查结果 ACC	多种疾病筛查结果 AUC	多种疾病识别结果 ACC	多种疾病识别结果 mAP
Three-Stream ^[12]	91.57%	91.36%	63.47%	62.98%
ResNet50 ^[23]	93.77%	93.54%	60.55%	61.09%
VGG16 ^[24]	88.24%	89.12%	57.43%	56.21%
EfficientNet-B5	94.35%	93.89%	67.88%	65.73%
EfficientNet-B6	94.74%	95.33%	68.74%	67.35%
EfficientNet-B7	95.16%	95.32%	70.45%	69.89%
本文方法	96.05%	97.18%	72.55%	73.24%

较于 EfficientNet 的 3 个模型,两个任务分别提升 1%~2%和 2%~5%;相较其他方法,两个任务分别提升 2%~8%和 9%~15%。

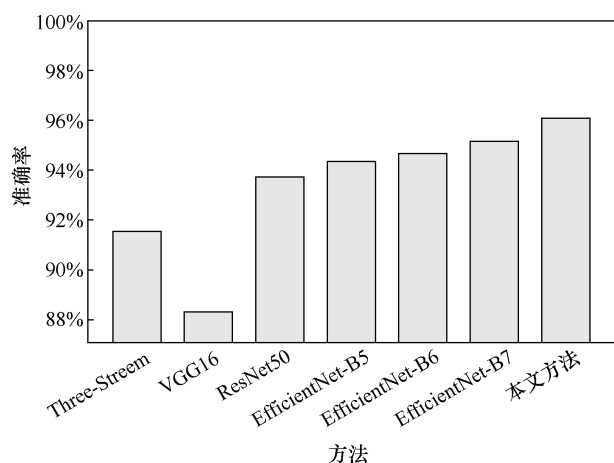


图 10 视网膜多种疾病筛查结果

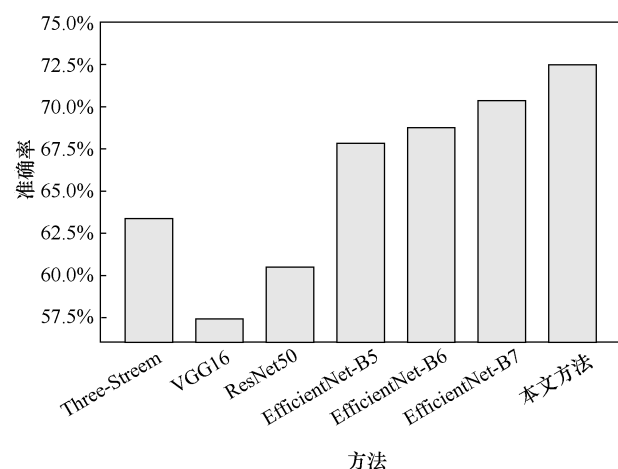


图 11 视网膜多种疾病识别结果

5 结束语

本文在视网膜多种疾病筛查和识别问题中,针对视网膜疾病种类繁多、病灶位置不确定、罕见疾病样本数量少等难题,首先提出了比较有效的图像数据增强和图像去噪方法,去除了视网膜眼底图像中冗余的黑框,提高了图像病灶位置的成像效果;结合使用深度卷积神经网络和集成学习的方法,对视网膜眼底图像的病灶特征进行提取,用于后续对多种疾病的筛查和识别。通过实验表明,在面临近 30 种视网膜疾病时,本文提出的方法能够有效地提升筛查和识别的准确率,可辅助眼科医生对患者进行初步诊断。本文方法虽取得了较好的结果,但在模型训练过程中仍存在耗时长、训练轮次多的问题,单次模型训练时间需 4~6 h,如何对该模型做

进一步的优化以提升模型的性能仍待解决。另外,模型在多种疾病识别任务中的准确率虽能达到 72.55%,但仍有进一步提升的空间,这也是未来研究工作中的一个方向。

参考文献:

- [1] STEINMETZ J D, BOURNE R R A, BRIANT P S, et al. Causes of blindness and vision impairment in 2020 and trends over 30 years, and prevalence of avoidable blindness in relation to VISION 2020: the right to sight: an analysis for the global burden of disease study[J]. The Lancet Global Health, 2021, 9(2): 144-160.
- [2] 王诗惠, 郝晓凤, 谢立科. 人工智能在视网膜疾病中应用的研究现状与展望[J]. 中华眼科医学杂志(电子版), 2020, 10(6): 374-379.
WANG S H, HAO X F, XIE L K. Research status and prospects of application of artificial intelligence in retinal diseases[J]. Chinese Journal of Ophthalmologic Medicine (Electronic Edition), 2020, 10(6): 374-379.
- [3] PRATT H, COENEN F, BROADBENT D M, et al. Convolutional neural networks for diabetic retinopathy[J]. Procedia Computer Science, 2016, 90: 200-205.
- [4] COLAS E, BESSE A, ORGOGOZO A, et al. Deep learning approach for diabetic retinopathy screening[J]. Acta Ophthalmologica, 2016, 94.
- [5] WORRALL D E, WILSON C M, BROSTOW G J. Automated retinopathy of prematurity case detection with convolutional neural networks[C]//Proceedings of the Deep Learning and Data Labeling for Medical Applications. Cham: Springer, 2016: 68-76.
- [6] BURLINA P, PACHECO K D, JOSHI N, et al. Comparing humans and deep learning performance for grading AMD: a study in using universal deep features and transfer learning for automated AMD analysis[J]. Computers in Biology and Medicine, 2017, 82: 80-86.
- [7] ALTAF T, ANWAR S M, GUL N, et al. Multi-class Alzheimer's disease classification using image and clinical features[J]. Biomedical Signal Processing and Control, 2018, 43: 64-74.
- [8] LIAO H F, LI Y C, LUO J B. Skin disease classification versus skin lesion characterization: achieving robust diagnosis using multi-label deep neural networks[C]//Proceedings of the 2016 23rd International Conference on Pattern Recognition. Piscataway: IEEE Press, 2016: 355-360.
- [9] SIDDIQUI M, MUJTABA G, REZA A, et al. Multi-class disease classification in brain MRIs using a computer-aided diagnostic system[J]. Symmetry, 2017, 9(3): 37.
- [10] LEE C S, BAUGHMAN D M, LEE A Y. Deep learning is effective for classifying normal versus age-related macular degeneration OCT images[J]. Ophthalmology Retina, 2017, 1(4): 322-327.
- [11] ARUNKUMAR R, KARTHIGAIAKUMAR P. Multi-retinal disease classification by reduced deep learning features[J]. Neural Computing and Applications, 2017, 28(2): 329-334.
- [12] WANG X, JU L, ZHAO X, et al. Retinal abnormalities recognition using regional multitask learning[C]//Proceedings of the Lecture Notes in Computer Science. Cham: Springer, 2019: 30-38.

- [13] CHOI J Y, YOO T K, SEO J G, et al. Multi-categorical deep learning neural network to classify retinal images: a pilot study employing small database[J]. PLoS One, 2017, 12(11): e0187336.
- [14] 曲毅, 张焕开, 宋先, 等. 人工智能诊断系统在视网膜疾病的研究进展[J]. 山东大学学报(医学版), 2020, 58(11): 39-44.
- QU Y, ZHANG H K, SONG X, et al. Research progress of artificial intelligence diagnosis system in retinal diseases[J]. Journal of Shandong University (Health Sciences), 2020, 58(11): 39-44.
- [15] ZUIDERVELD K. Contrast limited adaptive histogram equalization[M]//Graphics gems. Amsterdam: Elsevier, 1994: 474-485.
- [16] YUN S, HAN D, CHUN S, et al. CutMix: regularization strategy to train strong classifiers with localizable features[C]//Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE Press, 2019: 6022-6031.
- [17] TAN M X, LE Q V. EfficientNet: rethinking model scaling for convolutional neural networks[C]//Proceedings of the International Conference on Machine Learning. [S.l.:s.n.], 2019: 6105-6114.
- [18] HOWARD A, SANDLER M, CHEN B, et al. Searching for MobileNetV3[C]//Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE Press, 2019: 1314-1324.
- [19] RAMACHANDRAN P, ZOPH B, LE Q V. Searching for activation functions[J]. arXiv preprint, 2017, arXiv:1710.05941.
- [20] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]//Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 7132-7141.
- [21] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition[C]//Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2016: 770-778.
- [22] PACHADE S, PORWAL P, THULKAR D, et al. Retinal fundus multi-disease image dataset (RFMiD): a dataset for multi-disease detection research[J]. Data, 2021, 6(2): 14.
- [23] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition[C]//Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2016: 770-778.
- [24] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[J]. arXiv preprint, 2014, arXiv:1409.1556.

[作者简介]



王禾扬(1996-), 男, 南京航空航天大学计算机科学与技术学院硕士生, 主要研究方向为模式识别、医学影像处理。



杨启鸣(1996-), 男, 南京航空航天大学计算机科学与技术学院硕士生, 主要研究方向为模式识别、医学影像处理。



朱旗(1985-), 男, 博士, 南京航空航天大学计算机科学与技术学院副教授, 主要研究方向为模式识别、医学影像分析、生物特征识别。