

糖尿病视网膜病变眼底图像筛查研究

李家昱¹, 陈明惠^{1*}, 杨瑞君², 马文飞¹, 赖湘玲¹, 黄鐸文¹, 刘渡新¹, 马昕宏¹, 沈越¹¹上海理工大学医疗器械与食品学院生物医学工程系, 上海介入医疗器械工程技术研究中心,
教育部医学光学工程中心, 上海 200093;²四川省绵阳市第三人民医院, 四川 绵阳 621000

摘要 眼底照相是获取眼部图像的主要技术之一。利用眼底相机对视网膜病变区域进行拍摄可以获得清晰的图像,从获取的图像中能够直接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、数量和位置等信息可进行糖尿病视网膜病变分类。基于此,本文利用深度神经网络对糖尿病视网膜病变进行自动分类识别,提出了一种体系结构简单、在通用设备上运行速度快的卷积神经网络 CA-RepVGG(CA 代表 Channel Attention, RepVGG 为现有模块)。利用单路极简结构的 RepVGG 模块替代复杂的可使用性较差的模块作为分类模型的主体部位,并选用高效通道注意力机制 ECA 替代压缩注意力机制 SE,以此来提升模型对病变分级的能力。此外,本文还将 CA-RepVGG 模型与传统的分类模型 VGG-16、Inception-V3、ResNet-50 和 ResNext-50 模型进行了比较。从比较结果可以看出,虽然 CA-RepVGG 模型的参数量最大,但由于其是单分支结构,且只有 3×3 卷积块,因此它的模型复杂度并不高,分类速度很快,比另外 4 个模型中分类速度最快的 ResNet-50 还高出 15.3%。另外,利用两个混淆矩阵展示了所提模型的分类结果,其在两个数据集上的准确度都超过了 92.4%,精确度不低于 91.6%,灵敏度在 93.8% 以上。从实验结果可知,所提模型不仅可对糖尿病视网膜病变进行分类,而且相比其他现有模型具有一定的优越性。若将该模型应用在临床上,可以提高专业眼科医生在眼科疾病上的诊断效率。

关键词 医用光学; 眼科; 糖尿病视网膜病变分级; 眼底照相机; 深度学习; 眼底图像; 自动检测

中图分类号 TP391

文献标志码 A

DOI: 10.3788/CJL202249.1107001

1 引言

糖尿病是一种世界性的慢性疾病,会引起血管功能改变,最终诱发诸如糖尿病视网膜病变(DR)之类的并发症。据统计,几乎一半的糖尿病患者会患上糖尿病视网膜病变,并最终导致失明。所以,糖尿病视网膜病变的早发现 and 早治疗已成为世界各国降低失明率的重要措施。国际上将临床糖尿病视网膜病变的严重程度分为 5 个级别,包括无糖尿病视网膜病变、轻度糖尿病视网膜病变、中度糖尿病视网膜病变、重度糖尿病视网膜病变以及增殖性糖尿病视网膜病变。当前获取眼底图像主要依靠免散瞳彩色眼底照相技术,它能清晰地拍摄到视网膜上的软硬渗出物、出血点、微血管瘤等^[1-3]。眼科专家根据病

灶的类型、数量以及位置对糖尿病视网膜病变的程度进行分级,但目前的专家资源不足,给医生带来了沉重的工作负担。因此,采用计算机辅助诊断替医生分担工作就非常有必要。计算机辅助诊断不仅会大大减轻眼科医生的工作量,而且具有更精确的诊断结果,因此在临床上得到了越来越多的应用^[4-5]。

研究人员利用机器学习的经典算法以及它的重要分支——深度学习,对医学图像的分类进行了广泛研究^[6],并逐渐开辟了糖尿病视网膜病变眼底图像分类这一研究领域。庞浩等^[7]通过将 Inception 模块融入 ResNet 模型的方法搭建了一个多分支神经网络,但该网络的处理速度太慢,需要数周时间才能完成。Li 等^[8]提出了一种可以自动提取眼底图

收稿日期: 2021-09-16; 修回日期: 2021-10-19; 录用日期: 2021-11-08

基金项目: 国家自然科学基金青年科学基金(61308115)、上海市科委产学研医项目(15DZ1940400)

通信作者: *cmhui.43@163.com

像主要特征的方法,但该方法有一个很明显的缺陷,即:病变区域的大小不均衡会对图像分类的准确性产生严重影响。Sopharak 等^[9]采用经典的模糊 C 均值聚类算法对原始眼底视网膜图像进行处理,提取局部方差、灰度、色调以及边缘像素点这 4 个特征,然后根据这 4 个特征对图像数据集进行聚类,但他们并未给出聚类的图片类别总数。Pan 等^[10]提出了一种可以自动提取眼底图像中主要特征的方法,但该方法有一个缺陷,即:浅层网络无法提取完备的特征表示,最终导致模型易将病变样本误诊。Haloi 等^[11]利用深度学习的像素智能微动脉瘤方法进行糖尿病视网膜病变的分类,但该方法会将其他类型的小病变区域错误地识别为糖尿病视网膜病变,严重影响了分类的准确度。Antal 等^[12]提出了一种集成的糖尿病视网膜病变检测算法,该算法首先根据眼底图像中是否存在病灶点分辨出是否有糖尿病视网膜病变,然后对存在病变的图像进行分级。Lin 等^[13]提出了一种基于病变的离线加权方法,以此来提高糖尿病视网膜病变分级的准确性。Zhou 等^[14]提出了一种双阶段框架,试图将斑块级病态特征、图像级全局特征与一个注意力机制结合在一起,以对抗式网络协助优化糖尿病视网膜病变的分级。

但是,目前多数研究人员开发出的针对眼底图像进行分类的机器学习算法和深度学习算法存在一个普遍问题,即:为了追求分类的高准确率而不断增大模型的复杂度,导致计算复杂度增大以及浮点运算量和网络模型的参数量逐渐增大,从而降低了推理速度和内存利用率,使得开发出来的分类模型在临床上使用的可能性较低。鉴于此,本课题组提出了一种简单的网络模型,它不仅具有与最先进模型可比拟的高准确度、精确度及灵敏度,还具有较快的运算速度,更重要的是其具有帮助医生进行临床诊断的潜力,在临床上具有一定的应用价值。

本课题组对旷视实验室提出的 RepVGG 框架^[15]进行改进,提出了一个新颖的模型 CA-RepVGG,用于对糖尿病视网膜病变进行分级,并在新的糖尿病视网膜病变数据集上进行了模型性能的验证。本文的主要内容为:首先,利用结构重参数化方法将训练时的多分支体系结构和推理时的单分支体系结构解耦,大大降低模型的复杂度,满足结构简单和功能良好的要求;其次,利用一种新的极轻量级的注意力模块 ECA 来提高网络模型对视网膜眼底图像的特征提取能力;最后,在运算速度、参数量、精

确度、准确度以及灵敏度等方面,将所提网络与几个经典的网络模型进行比较。

2 实验方法与原理

本研究包括以下三个步骤:1)数据集的划分,即,把数据集图像划分为训练集、验证集和测试集。2)数据增强,即,对训练集中不同病变程度的图像进行数据扩充,使不同病变程度的图像在数量上保持相对平衡;采用的扩充方法为对比度受限的自适应直方图均衡化、旋转、镜像、增加亮度、伽马变换等。3)分类,即,利用训练好的模型在验证集和测试集上进行分类,测试模型的性能。

为了提高临床上使用的可能性,本文提出了 CA-RepVGG 网络模型。首先利用改进后的 RepVGG 模块替代传统的、复杂的网络模块,然后在网络模型中使用极轻量级注意力机制来提高整个网络对图像分级的效果,利用上述两个模块建立 CA-RepVGG 网络模型。这是一种简单的、性能优异的网络架构,如图 1 所示。蓝色层表示的是在 RepVGG 模块中添加了极轻量级注意力机制的层。

2.1 极轻量级注意力模块

最近几年,将注意力机制引入卷积块引起了广大研究者的关注,这是因为注意力机制在改善深层卷积神经网络性能方面具有巨大潜力,它的作用就是关注图像的重点区域,使分类效果更佳。注意力机制中最经典的是 SE(squeeze-and-excitation),它的原理是在输入特征后,首先采用全局池化操作对每个通道进行特征压缩,然后通过两个非线性全连接层对通道间的相关性进行建模,使输出的权重值数目和输入特征的权重值数目相同,最后使用归一化操作,把输出得到的权重加权到每个通道的特征上。上述步骤能使图像的重点区域得到关注。研究人员设计两个非线性全连接层的初衷是要利用它生成非线性的跨通道交互作用,这会涉及因避免过高的模型复杂度而进行的降维,但是降维会给权重的分布带来负面影响。除此之外,越来越多的研究者为了获得更好的分类效果,都致力于开发更复杂的注意力模块,但他们却忽略了一个事实,那就是这样做会不可避免地增加计算负担。因此,为了克服性能和复杂度之间的矛盾,本文选择利用极轻量级的注意力机制模块 ECA (efficient channel attention)^[16]。使用 ECA 的深度卷积神经网络引入的附加参数很少,计算量几乎可以忽略不计,同时会带来显著的性能提升。

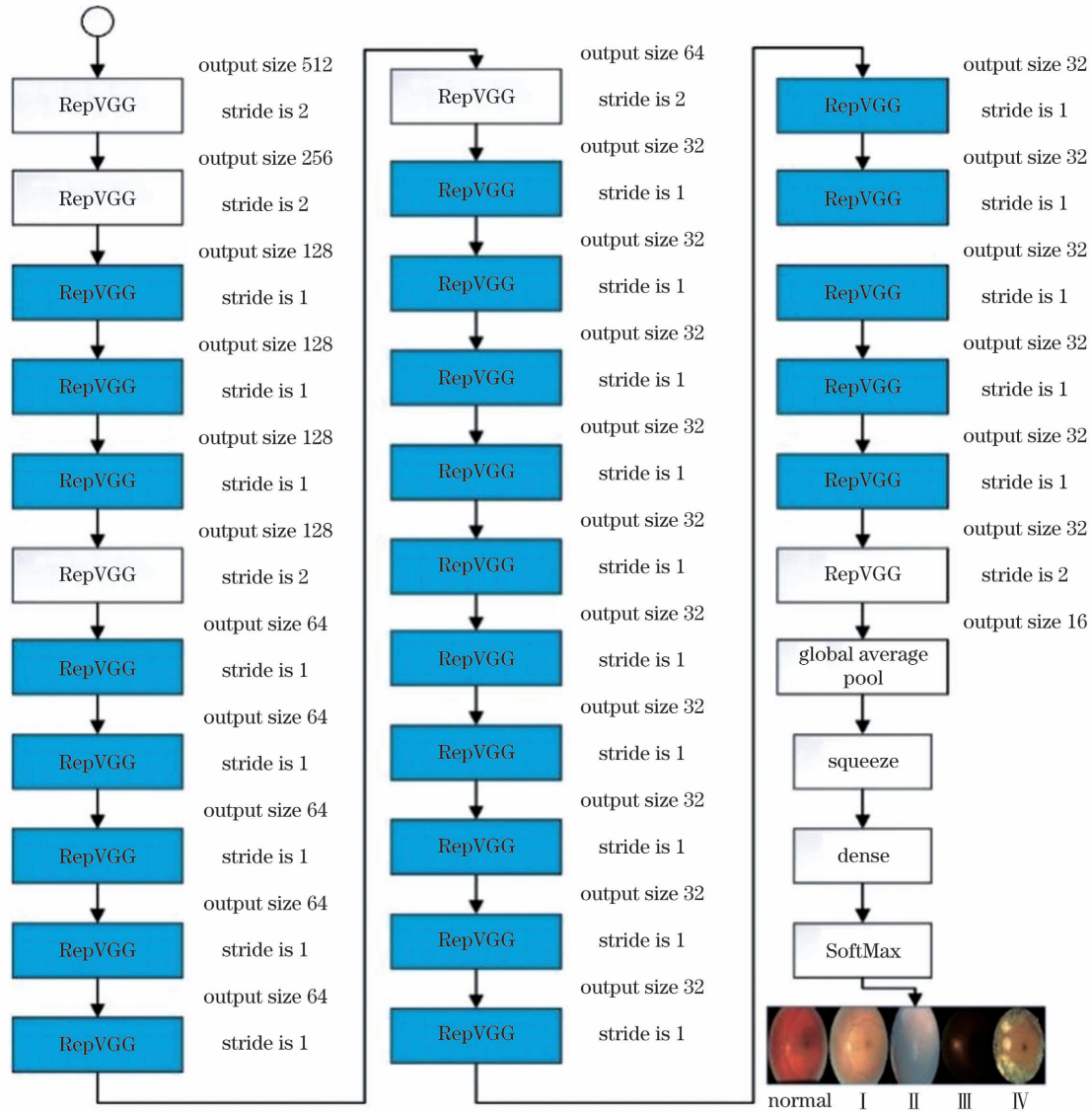


图 1 CA-RepVGG 网络的架构图

Fig. 1 Architecture of CA-RepVGG

基于上述分析可知,避免降维以及适当的跨通道交互作用在通道注意力机制的形成中起到了关键作用。而本文所使用的极轻量级注意力模块 ECA 就包含上述两个性质。具体地说,ECA 在没有降维的情况下进行通道式全局池化后,通过考虑每个通道及其 $k(k < 10)$ 个邻近通道来捕获局部跨通道交互信息。因此,ECA 通过大小为 k 的快速一维卷积生成通道权重。核大小 k 表示局部跨通道交互的覆盖,即有多少个邻近通道参与了一个通道的注意预测。显然,这将影响 ECA 的效率和性能。交互的覆盖与通道维度有关,因此通过一种与通道维度相关的函数来自适应地确定一维卷积核的大小是很有必要的。如式(1)所示,用 \mathbf{W}_k 表示学习到的注意力

$$\begin{bmatrix} \omega^{1,1} & \dots & \omega^{1,k} & 0 & 0 & \dots & \dots & 0 \\ 0 & \omega^{2,2} & \dots & \omega^{2,k+1} & 0 & \dots & \dots & 0 \\ \vdots & \vdots & & \vdots & & & & \vdots \\ 0 & \dots & 0 & 0 & \dots & \omega^{C,C-k+1} & \dots & \omega^{C,C} \end{bmatrix} \circ \quad (1)$$

\mathbf{W}_k 涉及 $K \times C$ 个参数,并且 \mathbf{W}_k 避免了不同组的完全独立。权重 y_i 与它的 k 个邻居之间的信息交互如图 2 所示,计算公式为

$$\omega_i = \sigma \left(\sum_{j=1}^k \omega_i^j y_i^j \right), y_i^j \in \Omega_i^k, \quad (2)$$

式中: ω_i^j 代表当前通道 i 的通道注意力; y_i^j 代表当前通道 i 的权重; Ω 表示全集。

为了提高模型的性能,还可以共享权重信息,即: $\omega_i = \sigma \left(\sum_{j=1}^k \omega^j y_i^j \right), y_i^j \in \Omega_i^k$,它利用通道数为 k 的

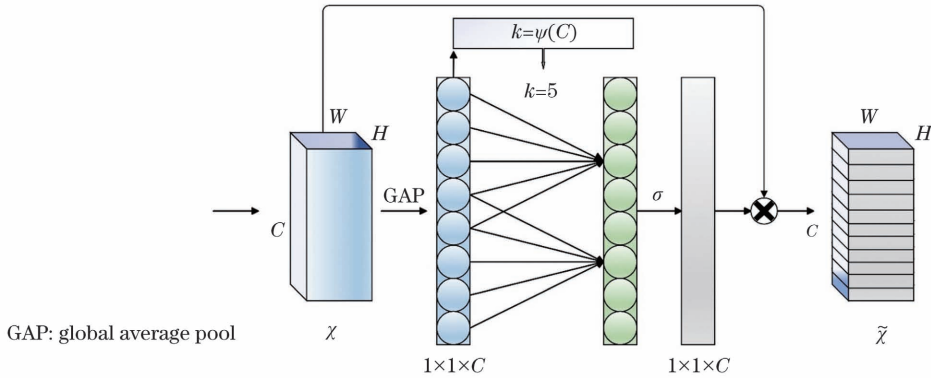


图 2 极轻量级注意力模块 ECA-Net

Fig. 2 Ultra lightweight attention module ECA-Net

一维卷积实现通道之间的信息交互,即

$$\omega = \sigma [C1D_k(y)], \quad (3)$$

式中: $C1D_k$ 代表一维卷积。这种方法就是 ECA 模块的原理,它只涉及 k 个参数信息。虽然目前可以利用手动优化来调整信息交互的最佳范围,但是会在计算资源上提出更高要求。针对此,在固定组数量不变的条件下,可以考虑利用分组卷积对网络架构进行改进,即,跨通道信息交互作用的覆盖范围(即 k 的大小)与通道的维数 C 之间存在一种映射关系

$$C = \varphi(k). \quad (4)$$

由于线性映射在某些相关特征上表现出了局限性,在此选择非线性映射。此外,通道维数一般选择为 2 的整数倍,所以 C 与 k 之间的关系采用以 2 为底的指数函数来表示,即

$$C = \varphi(k) = 2^{(\gamma \times k - d)}. \quad (5)$$

在本文中, γ 取值为 2, d 取值为 1。所以如果给出了通道维数 C , 那么 k 的大小便可通过转换得到,转换公式为

$$k = \psi(C) = \left\lfloor \frac{\log_2(C)}{\gamma} + \frac{d}{\gamma} \right\rfloor_{\text{odd}}. \quad (6)$$

实验证明,当 $k=3$ 时, ECA 模块不仅具有最好的权重重分配效果和最低的模型复杂度,而且非常高效。

2.2 RepVGG 模块

RepVGG 模块用结构重参数化方法实现了 VGG 式单路极简结构,在整个网络结构中,从浅层卷积层到深层卷积层只使用了 3×3 卷积和 ReLU,不使用各种新颖的激活函数,也不使用多分支结构,更没有自动搜索和复合缩放等繁重设计,因此能在速度和性能上有很大提升。网络结构只有 28 层,分为 5 个阶段,分别包含 1、4、6、16、1 个 RepVGG 模

块,每一阶段的第一层都用 $\text{stride}=2$ 进行降采样。图 3 展示了第 2~4 阶段的前两层,图 3(a)是训练阶段的 RepVGG 模块,分类效果比较好,但速度较慢且耗内存。图 3(b)是推理阶段的 RepVGG 模块,其优点是速度快和省内存。

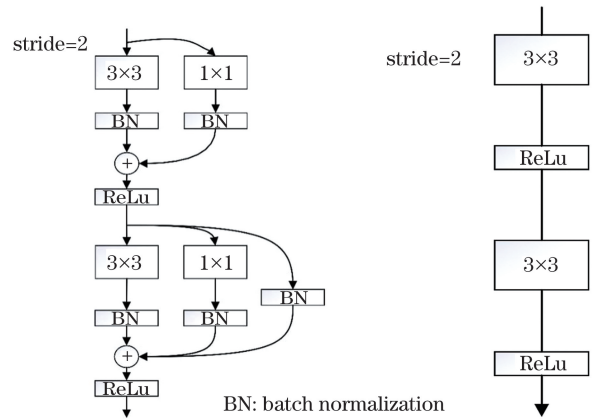


图 3 RepVGG 模块图。(a)训练阶段; (b)推理阶段
Fig. 3 RepVGG module. (a) Training; (b) inference

借鉴 ResNet 的思想,在训练时采用多分支结构,把 1×1 卷积核分支、 3×3 卷积核分支和残差路径分支称为一个块,相当于一个块内所做的计算为

$$y = F(X) + G(X) + X, \quad (7)$$

式中: $F(X)$ 为 3×3 的卷积; $G(X)$ 为 1×1 的卷积; X 是残差路径分支。为了将多分支结构变为单分支结构,首先需要将卷积层和 BN 层进行融合。假设卷积层的计算公式为

$$\text{Conv}(x) = W(x) + b, \quad (8)$$

BN 层的计算公式为

$$\text{BN}(x) = y \times \frac{(x - V_{\text{mean}})}{\sqrt{d_{\text{var}}}} + \beta, \quad (9)$$

然后将卷积层的结果代入到 BN 层公式中,即

$$\text{BN}[\text{Conv}(x)] = y \frac{W(x) + b - V_{\text{mean}}}{\sqrt{d_{\text{var}}}} + \beta, \quad (10)$$

进一步化简为

$$\text{BN}[\text{Conv}(x)] = \frac{y \times W(x)}{\sqrt{d_{\text{var}}}} + \frac{y \times (b - V_{\text{mean}})}{\sqrt{d_{\text{var}}}} + \beta, \quad (11)$$

式中: $W(x)$ 为卷积层; b 为偏置参数; V_{mean} 为累计平均值; $\sqrt{d_{\text{var}}}$ 为标准差; y 为学习比例因子; β 为偏置。这其实就是一个卷积层, 只不过权重考虑了 BN 的参数。令 $W_{\text{fused}}(x) = [y \times W(x)] / \sqrt{d_{\text{var}}}$,

$$B_{\text{fused}} = \frac{y \times (b - V_{\text{mean}})}{\sqrt{d_{\text{var}}}} + \beta, \text{ 最终融合的结果是}$$

$$\text{BN}[\text{Conv}(x)] = W_{\text{fused}}(x) + B_{\text{fused}}. \quad (12)$$

接下来考虑将 1×1 卷积变成 3×3 卷积。如图 4 所示, 1×1 卷积核在图片特征图上滑动, 这里 1×1 卷积不需要在边界填充 0 就能保证特征图的大小前后一致。为了保证 3×3 卷积核在特征图上获取的信息一致, 需要在原特征图边界上填充一圈 (图中浅蓝色部分表示边界填充, 值全为 0), 这样就能保证特征图的信息不变。

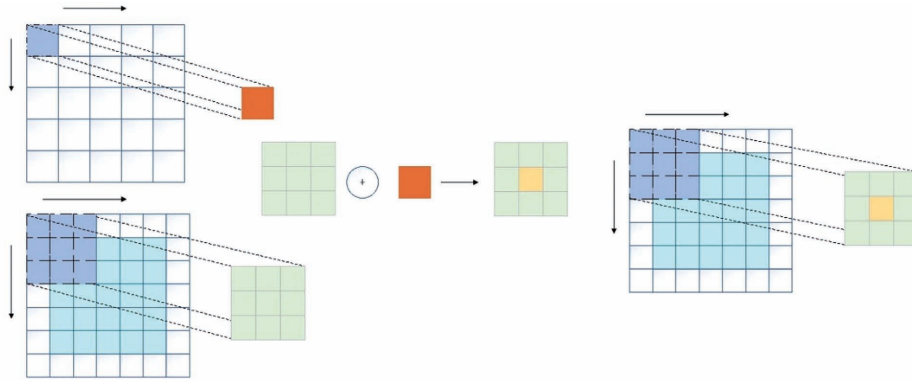


图 4 卷积变换示意图

Fig. 4 Schematic of convolution transformation

通过观察 3×3 卷积核中间的那个核会发现它卷积的路径就是前面 1×1 卷积的路径。一个很自然的想法就是把 1×1 卷积核加在 3×3 卷积核中间, 这样就能完成卷积分支的融合。

1×1 卷积核经边界填充后成为 3×3 卷积, 然后将其与另外的 3×3 卷积核相加, 得到新的卷积核, 这个新卷积核在原始特征图上作卷积操作得到的新特征图与 1×1 卷积核和 3×3 卷积核在原始特征图上分别操作然后相加得到的特征图一样。

最后, 将残差路径分支与 3×3 卷积融合。残差路径分支其实就是每个通道中的每个元素对应相加。现在考虑将残差路径分支用一个 3×3 卷积层表示, 为保证残差路径分支前后值不变, 利用权重等于 1 的卷积核并分通道进行卷积, 相当于单独对每个通道的每个元素乘以 1, 然后再输出来, 这样就完成了残差路径分支的融合。

3 眼底图像数据集与实验设置

本文的数据集 1 来自 2015 年糖尿病视网膜病变检测比赛^[17-20]。为进一步验证本文模型的性能,

使用温州医科大学和四川省绵阳市第三人民医院的临床数据集进行验证。每个图像都经过调整和裁剪, 最大尺寸为 $1024 \text{ pixel} \times 1024 \text{ pixel}$ 。2015 年糖尿病视网膜病变检测数据集的具体情况如表 1 所示, 其中包含了多名患者的左右眼底图像, 而且临床医生根据 0~4 等级对每张图像的糖尿病视网膜病变严重程度进行了评级 (0 级为正常, 1 级为轻度糖尿病视网膜病变, 2 级为中度糖尿病视网膜病变, 3 级为重度糖尿病视网膜病变, 4 级为增殖性糖尿病视网膜病变)。将 2015 年糖尿病视网膜病变检测训练集中的原始训练集作为模型的训练集, 然后从测试集图像中取出部分图像, 一部分作为测试集, 剩下的作为模型的验证集。由表 1 可知, 不同等级的图像分类极不均匀, 采用神经网络进行训练时, 会导致网络模型出现严重的过拟合现象。因此, 本文采用数据增强技术对标签为 1~4 的图像进行扩充, 使各个标签的数据保持相对平衡。扩充的方法包括对比度受限的自适应直方图均衡化、旋转、镜像、增加亮度、伽马变换操作, 数据增强的具体细节如图 5 所示。

表 1 2015 年糖尿病视网膜病变检测数据集
Table 1 2015 diabetic retinopathy detection dataset

Grade	Degree of illness	Number of training images	Number of testing images	Total
0	Healthy	25810	39533	65343
1	Light	2443	3762	6205
2	Moderate	5292	7861	13153
3	Severe	873	1214	2087
4	Proliferative	708	1206	1914

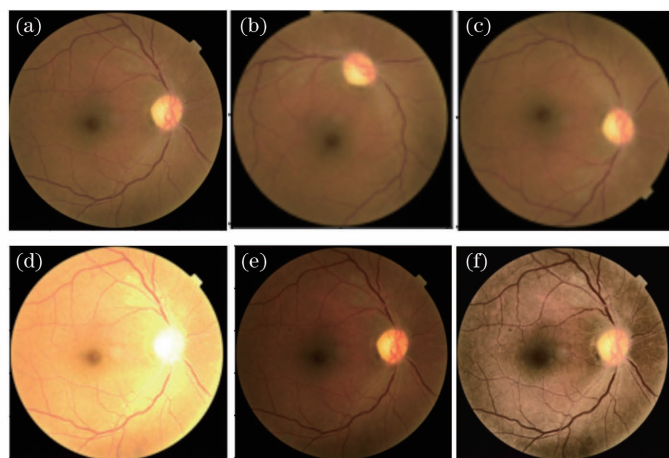


图 5 数据增强。(a)原始图像;(b)旋转 60°;(c)水平镜像;(d)亮度增强;(e)伽马变换;(f)自适应直方图均衡化
Fig. 5 Data enhancement. (a) Original image; (b) rotation of 60°; (c) mirror horizontally; (d) enhancement of brightness; (e) Gamma transformation; (f) adaptive histogram equalization

对数据集 2 使用与数据集 1 一样的数据增强方式,提取原始测试集中每一类标签中的 100 张图像作为测试集,剩下的作为验证集,并且在训练好的模型上进行微调,证明所提网络在不同数据集上的良好的分类性能。微调步骤如下:首先在数据集 1 训练好的模型上添加自定义网络,然后冻结训练好的网络并训练所添加的部分,最后解冻网络部分层,并联合训练解冻的这些层和添加的部分。

在具有 32 GB 显存的 Nvidia Tesla V100 GPU HPC^[21]集群上进行训练,使用 PyTorch^[22]作为深度学习开发框架,在神经网络模型训练过程中,输入的眼底图像的大小为 512 pixel×512 pixel,使用 Adam 优化器对糖尿病视网膜病变图像进行分类。在网络模型训练过程中,初始学习率设置为 1×10^{-5} ,在每 10 个周期后将学习率乘以 0.5,训练的最大周期为 50。其他模型的对比实验都使用相同的训练参数。

由表 1 可知数据集存在正负样本比例严重失衡的问题,因此模型训练过程中的损失函数选用 focal loss。该损失函数可以降低正样本在训练时的权重,使神经网络能够更容易地进行训练。focal loss 多分类的具体实现公式为

$$F_L = -(1 - p_{\text{prediction}} \times p_{\text{groundtruth}})^{\gamma} \lg(p_{\text{prediction}}), \quad (13)$$

式中: $p_{\text{prediction}}$ 表示的是模型预测分类的概率值; $p_{\text{groundtruth}}$ 表示的是模型正确分类的概率值; γ 代表不同难易程度的样本对模型的贡献程度,根据实验,当 $\gamma=0.25$ 时效果最佳。

在样本数据集很少的情况下,可以使用 4 折交叉验证方法进行验证。该方法的原理是将数据划分为 4 个分区,得到 4 个模型,然后用 3 个分区训练每一个模型,并在留下的分区上进行评估,将 4 个预测值的平均值作为最终的预测值进行输出。

在正负样本不平均的前提下,仅用准确率衡量算法有很大缺陷,甚至不具有参考意义。因此,本文使用 A(准确度)、P(精确度)及 S(灵敏度)来评价算法的优劣。在临床上,准确度能反映平均分类的准确性,灵敏度能反映糖尿病视网膜病变图像没有被漏判的概率。具体计算公式为

$$A = \frac{f_{\text{TP}} + f_{\text{TN}}}{f_{\text{TP}} + f_{\text{TN}} + f_{\text{FP}} + f_{\text{FN}}}, \quad (14)$$

$$P = \frac{f_{\text{TP}}}{f_{\text{TP}} + f_{\text{FP}}}, \quad (15)$$

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

式中： f_{TP} 表示五分类中实际患病情况和预测结果相同的样本数量； f_{TN} 表示实际结果和预测结果都是无病症的样本数量； f_{FP} 表示实际患病情况比预测患病情况要轻的样本数量； f_{FN} 表示实际患病情况比预测患病情况要重的样本数量。

4 结果与分析

4.1 分类结果

本文使用不同的网络模型对眼底图像进行分类验证。使用 VGG-16^[23]、Inception-V3^[24]、ResNet-50^[25] 和 ResNext-50^[26] 网络模型进行对比实验，验

证所提模型的先进性，并使用上述三个评价指标进行性能评估，同时用速度(speed)表示模型分类的快慢。速度是在 1080Ti 上测试的，批量大小为 128，全精度(fp32)，以每秒处理的图片数量(photo/s)为单位进行比较，图片尺寸为 512 pixel×512 pixel。

模型性能验证在数据集 1 的 1096 张图片和数据集 2 的 500 张图片上进行。表 2、3 给出了各模型的分类速度，其中 CA-RepVGG 的分类速度最快，比 ResNet-50 的分类速度还高出 15.3%。同时，CA-RepVGG 在数据集 1 上的准确度达到 92.4%，精确度达到 91.6%，灵敏度达到 96.5%；在数据集 2 上的准确度达到 93.9%，精确度达到 96.3%，灵敏度达到 93.8%。

表 2 数据集 1 上不同模型的性能指标对比

Table 2 Comparison of performance indicators of different models for dataset 1

Model	Parameter number /10 ⁶	Speed /(photo·s ⁻¹)	Accuracy /%	Precision /%	Sensitivity /%
VGG-16	13.84	208	82.0	83.9	85.0
Inception-V3	27.16	295	87.3	89.4	93.2
ResNet-50	25.56	360	89.5	90.7	97.0
ResNext-50	25.03	242	91.0	91.3	96.5
CA-RepVGG (ours)	51.82	415	92.4	91.6	96.5

表 3 数据集 2 上不同模型的性能指标对比

Table 3 Comparison of performance indicators of different models for dataset 2

Model	Parameter number /10 ⁶	Speed /(photo·s ⁻¹)	Accuracy /%	Precision /%	Sensitivity /%
VGG-16	13.84	208	83.0	82.7	84.0
Inception-V3	27.16	295	85.3	88.4	86.2
ResNet-50	25.56	360	89.4	89.7	91.0
ResNext-50	25.03	242	91.5	90.5	90.4
CA-RepVGG (ours)	51.82	415	93.9	96.3	93.8

根据验证集的结果绘制出混淆矩阵，如图 6 和图 7 所示，图 6 是数据集 1 的混淆矩阵，图 7 是数据

集 2 的混淆矩阵。一个分类器是否具有良好的分类效果可在混淆矩阵中体现出来。图 6、7 显示，图片

confusion matrix (accuracy is 0.9388686131386861)

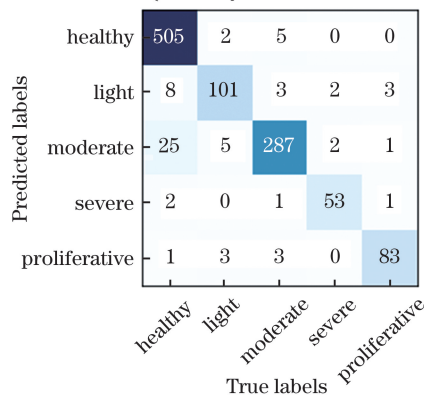


图 6 数据集 1 的混淆矩阵

Fig. 6 Confusion matrix of dataset 1

confusion matrix (accuracy is 0.924)

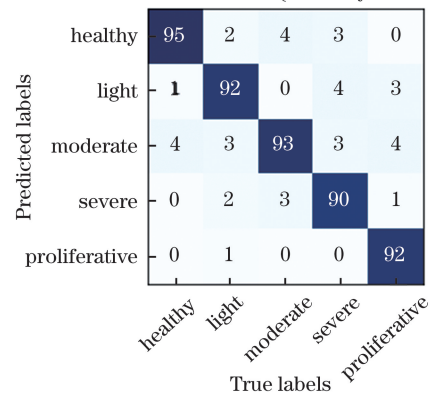


图 7 数据集 2 的混淆矩阵

Fig. 7 Confusion matrix of dataset 2

的分类结果在对角线上集中分布,而对角线上的数值是模型正确分类的图片数量,只有极少量图片分类错误,其中过度诊断分布在对角线左下方,漏诊分布在对角线右上方。混淆矩阵表明所提网络模型分类效果较好。

4.2 分析与讨论

CA-RepVGG 的优点很显著,它是为在 GPU 和专用硬件上实现最大速度设计的,具有模型复杂度低、运行速度快、准确率和精确度高、灵敏度好等特点。具体原因如下:1)本文所使用的模型虽然参数量较大,但模型参数并不是决定模型复杂度的重要因素。本文提出的 CA-RepVGG 模型利用结构重参数化方法,使得其在推理阶段时所有的卷积层都是 3×3 卷积,而且是单分支结构,这就在在很大程度上降低模型的复杂度,而其他现有模型,比如 Inception-V3、ResNet-50、ResNext-50 等都包含有多分支结构,多分支结构必然会大大增加模型的复杂度。2)其他模型除了含有 3×3 卷积块以外,还混杂了 1×1 卷积块、 5×5 卷积块、 7×7 卷积块,甚至在 ResNext-50 上还混杂了 9×9 、 11×11 这样大的卷积块,多种尺度的卷积层混杂在一起必然会大大增加模型的复杂度。综合以上原因,即使 CA-RepVGG 与对比模型相比参数量较多,也不会影响它的运行速度。VGG-16 的运行速度是 208 photo/s, Inception-V3 的速度是 295 photo/s, ResNet-50 的速度是 360 photo/s, ResNext-50 的速度是 242 photo/s,而 CA-RepVGG 的速度是 415 photo/s,比另外 4 个模型中处理速度最快的 ResNet-50 还要高出 15.3%。但是,在低功耗设备的移动机制方面,本文模型可能不如 MobileNets 和 ShuffleNets 等受欢迎。

在接下来的实验中拟考虑使用 5×5 卷积块,虽然它的参数量比 3×3 卷积块还要大,但它可能比 3×3 卷积块更适合处理更大尺寸的图片。

5 结 论

本课题组提出的 CA-RepVGG 用 RepVGG 模块替代复杂的网络模型,用 ECA 这样简单高效的注意力机制替换计算量大的注意力机制,在功能上很好地实现了对眼底图像的分类识别,不仅保证了模型的结构简单、计算量小,还提升了其在临床上应用的可行性。

利用 CA-RepVGG 模型在数据集上测试和评估了糖尿病视网膜病变图像的分类效果,同时将该

模型与传统的 VGG-16、Inception-V3、ResNet-50 和 ResNext-50 进行了对比实验,并使用运行速度、准确度、精确度和灵敏度指标验证了 CA-RepVGG 网络模型的先进性。由实验结果可知,该算法在糖尿病视网膜病变分类上不仅可行,而且具有优越性。若应用在临床上,可以提高专业眼科医生在眼科疾病上的诊断效率,保证更多的患者能及时就医,避免失明。如果未来能采用更多的数据集对所提模型进行训练,就能进一步提高其自动分类的准确度、精确度和灵敏度,在临床上达到更好的效果。

参 考 文 献

- [1] 张鹏飞, 张廷玮, 宋维业, 等. 从小鼠视网膜多种成像方式探讨眼科光学成像技术进展[J]. 中国激光, 2020, 47(2): 0207003.
Zhang P F, Zhang T W, Song W Y, et al. Review of advances in ophthalmic optical imaging technologies from several mouse retinal imaging methods [J]. Chinese Journal of Lasers, 2020, 47(2): 0207003.
- [2] 张洋, 倪敬书, 张元志, 等. 组织固有荧光光谱复原算法及其在糖尿病筛查中的应用研究[J]. 中国激光, 2018, 45(7): 0707001.
Zhang Y, Ni J S, Zhang Y Z, et al. Tissue intrinsic fluorescence spectrum recovery algorithm and its application in diabetes screening [J]. Chinese Journal of Lasers, 2018, 45(7): 0707001.
- [3] Dutta S, Manideep B C, Basha S M, et al. Classification of diabetic retinopathy images by using deep learning models [J]. International Journal of Grid and Distributed Computing, 2018, 11(1): 99-106.
- [4] Li Z X, Keel S, Liu C, et al. An automated grading system for detection of vision-threatening referable diabetic retinopathy on the basis of color fundus photographs [J]. Diabetes Care, 2018, 41(12): 2509-2516.
- [5] 王亮, 陈春晓, 傅雪, 等. 基于 FDMU-net 的早产儿视网膜图像血管分割 [J]. 激光与光电子学进展, 2021, 58(14): 1417003.
Wang L, Chen C X, Fu X, et al. Retinal vessel segmentation of prematurity infants based on FDMU-net [J]. Laser & Optoelectronics Progress, 2021, 58(14): 1417003.
- [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] 庞浩, 王枫. 用于糖尿病视网膜病变检测的深度学习

- 模型[J]. 软件学报, 2017, 28(11): 3018-3029.
- Pang H, Wang Z, Deep learning model for diabetic retinopathy detection[J]. Journal of Software, 2017, 28(11): 3018-3029.
- [8] Li Z, Keel S, Liu C, et al. An automated grading system for detection of vision-threatening referable diabetic retinopathy on the basis of color fundus photographs[J]. Diabetes Care, 2018, 41(12): 2509-2516.
- [9] Sopharak A, Uyyanonvara B, Barman S. Automatic exudate detection from non-dilated diabetic retinopathy retinal images using fuzzy C-means clustering[J]. Sensors, 2009, 9(3): 2148-2161.
- [10] Pan C, Li J, Wang Y, et al. Collaborative learning for hyperspectral image classification [J]. Neurocomputing, 2018, 275: 2512-2524.
- [11] Haloi M. Improved microaneurysm detection using deep neural networks[EB/OL]. (2015-05-17)[2021-09-16]. <https://doi.org/10.48550/arXiv.1505.04424>.
- [12] Antal B, Hajdu A. An ensemble-based system for microaneurysm detection and diabetic retinopathy grading [J]. IEEE Transactions on Biomedical Engineering, 2012, 59(6): 1720-1726.
- [13] Lin Z, Guo R, Wang Y, et al. A framework for identifying diabetic retinopathy based on anti-noise detection and attention-based fusion[C]//International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham: Springer, 2018: 74-82.
- [14] Zhou Y, He X, Huang L, et al. Collaborative learning of semi-supervised segmentation and classification for medical images[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 2079-2088.
- [15] Ding X H, Zhang X Y, Ma N N, et al. RepVGG: making VGG-style ConvNets great again[EB/OL]. (2021-01-11)[2021-06-02]. <https://arxiv.org/abs/2101.03697>.
- [16] Wang Q L, Wu B G, Zhu P F, et al. ECA-net: efficient channel attention for deep convolutional neural networks[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 11531-11539.
- [17] Ming S, Xie K P, Lei X, et al. Evaluation of a novel artificial intelligence-based screening system for diabetic retinopathy in community of China: a real-world study[J]. International Ophthalmology, 2021, 41(4): 1291-1299.
- [18] Ruamviboonsuk P, Krause J, Chotcomwongse P, et al. Deep learning versus human graders for classifying diabetic retinopathy severity in a nationwide screening program[J]. Npj Digital Medicine, 2019, 2: 25.
- [19] Natarajan S, Jain A, Krishnan R, et al. Diagnostic accuracy of community-based diabetic retinopathy screening with an offline artificial intelligence system on a smartphone[J]. JAMA Ophthalmology, 2019, 137(10): 1182-1188.
- [20] Abràmoff M D, Lou Y Y, Erginay A, et al. Improved automated detection of diabetic retinopathy on a publicly available dataset through integration of deep learning [J]. Investigative Ophthalmology & Visual Science, 2016, 57(13): 5200-5206.
- [21] Knight J C, Nowotny T. GPUs outperform current HPC and neuromorphic solutions in terms of speed and energy when simulating a highly-connected cortical model[J]. Frontiers in Neuroscience, 2018, 12: 941.
- [22] 袁坤, 霍力. 用于视网膜 OCT 图像分层的多尺度补全卷积神经网络 [J]. 中国激光, 2021, 48(15): 1507004.
- Yuan K, Huo L. Multiple-scale inpainting convolutional neural network for retinal OCT image segmentation[J]. Chinese Journal of Lasers, 2021, 48(15): 1507004.
- [23] Aswathy A L, Anand H S, Vinod C S S. COVID-19 diagnosis and severity detection from CT-images using transfer learning and back propagation neural network[J]. Journal of Infection and Public Health, 2021, 14(10): 1435-1445.
- [24] Cao J F, Yan M M, Jia Y M, et al. Application of a modified Inception-v3 model in the dynasty-based classification of ancient murals[J]. EURASIP Journal on Advances in Signal Processing, 2021, 2021: 49.
- [25] Yang Y J, Cho B J, Lee M J, et al. Automated classification of colorectal neoplasms in white-light colonoscopy images via deep learning[J]. Journal of Clinical Medicine, 2020, 9(5): 1593.
- [26] Vidyarthi A, Malik A. A hybridized modified densenet deep architecture with CLAHE algorithm for humpback whale identification and recognition [J]. Multimedia Tools and Applications, 2021: 1-15.

Fundus Image Screening for Diabetic Retinopathy

Li Jiayu¹, Chen Minghui^{1*}, Yang Ruijun², Ma Wenfei¹, Lai Xiangling¹, Huang Duowen¹,
Liu Duxin¹, Ma Xinhong¹, Shen Yue¹

¹ Shanghai Engineering Research Center of Interventional Medical, the Ministry of Education of Medical Optical Engineering Center, Department of Biomedical Engineering, School of Medical Instrument and Food Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China;

² The Third People's Hospital of Mianyang City, Mianyang 621000, Sichuan, China

Abstract

Objective Diabetes is a worldwide chronic disease, which can cause changes in vascular performance and result in complications such as diabetic retinopathy (DR). If a patient has DR and it goes undetected, the patient's eyesight may be lost. Therefore, early detection and early treatment of DR is important for reducing the blindness rate in various countries. At present, the acquisition of fundus images mainly depends on nonmydriasis color fundus photography, which can clearly capture the soft and hard exudates, bleeding points, microvessels, and so on. Currently, the ophthalmologists who observe fundus images have to endure heavy workloads and massive job-related stress owing to problems such as insufficient expert resources. To reduce their workload, it is necessary to use computer-aided diagnosis. Moreover, computer-aided diagnosis is more accurate. In the field of medical imaging, using classical machine learning and deep learning to classify medical images has become key research areas and gradually the subfield of retinal fundus image classification has developed. However, there is a serious problem with machine learning and deep learning algorithms for fundus image classification developed by many researchers, that is, they continuously increase the complexity of the model to pursue high accuracy. This results in a corresponding increase in computational complexity in terms of the number of floating-point operations and the size of the parameters of the network model, thus reducing the speed and increasing the memory utilization. An inefficient classification model is less likely to be used in clinical practice. The purpose of this paper is to propose a simple network model. Compared with the most advanced model, our model has not only high accuracy, precision, and sensitivity but also high speed. More importantly, it has the potential to be used in clinical practice.

Methods We improved the network architecture of the RepVGG model proposed by the Kuangshi's group and proposed a novel model Channel Attention-RepVGG (CA-RepVGG). The RepVGG with simple structure was used to replace the complex module as the main part of the classification model, and efficient channel attention was selected to replace the squeeze-and-excitation (SE) for a good classification of images related to DR. Then, CA-RepVGG was tested on the new dataset of DR images. The main research includes the following. First, the multibranch architecture employed during training and the single-branch architecture employed during inference were decoupled using the structural reparameterization method, which greatly reduced the complexity of the model and met the requirements of simplicity in structure. Second, a new lightweight attention module was used to improve the performance of the convolutional neural network and enhance the ability of feature extraction from retinal fundus images. Finally, the parameters, speed, precision, accuracy, and sensitivity of several classical networks in image classification were compared.

Results and Discussions The proposed model is tested on 1096 pictures of dataset 1 and 500 pictures of dataset 2. The accuracy of dataset 1 is 92.4%, the precision is 91.6%, and the sensitivity is 96.5%. For dataset 2, the accuracy is 93.9%, the precision is 96.3%, and the sensitivity is 93.8%. The confusion matrix of dataset 1 is shown in Fig. 6 and that of dataset 2 is shown in Fig. 7. The two figures show the classification results of the pictures; the values on the diagonal present the number of patients correctly classified by the model. Only a few images have errors, the overdiagnosis distributes on the lower left of the diagonal in the image and the missed diagnosis distributes on the upper right of the diagonal. The confusion matrices show the superiority of the proposed model. Furthermore, CA-RepVGG has the fastest speed, although it has the largest parameters in the compared models. Owing to the single-branch architecture and 3×3 convolution, our model is not complex. The CA-RepVGG can process 415 pictures per second, which is 15.3% higher than those by ResNet-50.

Conclusions CA-RepVGG can be used in clinical practice. The simplicity of the model and the small amount of

calculation ensure the feasibility and reliability of CA-RepVGG. In this paper, CA-RepVGG is used to test and evaluate the classification effect of DR images in two datasets. At the same time, VGG-16, Inception-V3, ResNet-50, and ResNext-50 are compared with our model, and the accuracy, precision, and sensitivity of the network demonstrate the advanced nature of our model. The experimental results show that the model is not only feasible but also superior in classification. In the future, if our proposed model is applied to clinical practice, it can enhance the diagnostic efficiency of professional ophthalmologists regarding ophthalmic diseases, especially in remote and poor areas, ensuring that more patients can be treated in time and avoid losing their eyesight. If more datasets can be used to train the model in the future, the accuracy of automatic classification can be further enhanced and better results can be achieved in clinical practice.

Key words medical optics; ophthalmology; grading of diabetic retinopathy; fundus camera; deep learning; fundus image; automatic detection