

用于糖尿病视网膜病变检测的深度学习模型*

庞浩^{1,2}, 王枏^{1,2}



¹(北京邮电大学 软件学院, 北京 100876)

²(可信分布式计算与服务教育部重点实验室(北京邮电大学), 北京 100876)

通讯作者: 庞浩, E-mail: panghao@bupt.edu.cn

摘要: 近年来,深度学习在计算机视觉方面取得了巨大的进步,并在利用计算机视觉完成医学影像的阅片工作方面展现出了良好的应用前景.针对糖尿病眼底病变筛查工作,通过构建两级深度卷积神经网络,完成了原始照片的特征提取、特征组合和结果分类,最终得出筛查结果.通过与医生的诊断结果进行比较,证明了模型的输出结果与医生诊断结果之间具有高度的一致性.同时,提出了利用弱监督学习进行细粒度图像分类的改进方法.最后,对未来研究的方向进行了展望.

关键词: 计算机视觉;卷积神经网络;深度学习;弱监督学习;糖尿病视网膜病变
中图分类号: TP182

中文引用格式: 庞浩,王枏.用于糖尿病视网膜病变检测的深度学习模型.软件学报,2017,28(11):3018-3029. <http://www.jos.org.cn/1000-9825/5332.htm>

英文引用格式: Pang H, Wang C. Deep learning model for diabetic retinopathy detection. Ruan Jian Xue Bao/Journal of Software, 2017, 28(11): 3018-3029 (in Chinese). <http://www.jos.org.cn/1000-9825/5332.htm>

Deep Learning Model for Diabetic Retinopathy Detection

PANG Hao^{1,2}, WANG Cong^{1,2}

¹(School of Software, Beijing University of Posts and Telecommunications, Beijing 100876, China)

²(Key Laboratory of Trustworthy Distributed Computing and Service (Beijing University of Posts and Telecommunications), Ministry of Education, Beijing 100876, China)

Abstract: In recent years, deep learning in the computer vision has made great progress, showing good application prospects in medical image reading. In this paper, a model with construction of two-level deep convolution neural network is designed to achieve feature extraction, feature blend, and classification of the fundus photo. By comparing with doctor's diagnosis, it is shown that the output of the model is highly consistent with the doctor's diagnosis. In addition, an improved method of fine-grained image classification using weak supervised learning is proposed. Finally, future research direction is discussed.

Key words: computer vision; convolutional neural network; deep learning; weak supervised learning; diabetic retinopathy

糖尿病视网膜病变(diabetic retinopathy,简称 DR)是糖尿病最常见的眼部并发症,是主要的致盲眼病之一.长期的高血糖环境会损伤视网膜血管的内皮,引起一系列的眼底病变,如微血管瘤、硬性渗出、棉絮斑、新生血管、玻璃体增殖甚至视网膜脱离.一般糖尿病出现 10 年以上的病人开始出现眼底病变,但如果血糖控制差,或者是胰岛素依赖型糖尿病的患者则可能更早出现眼底病变,故糖尿病患者需要定期到眼科检查眼底.

目前,检测 DR 是一个耗时的人工过程,需要临床医生检查和评估视网膜的数字彩色眼底照片.医生通常一

* 基金项目: 国家重点研发计划(2016YFF0201003)

Foundation item: National Key Research and Development Program of China (2016YFF0201003)

本文由复杂环境下的机器学习研究专刊特约编辑胡清华教授推荐.

收稿时间: 2017-01-03; 修改时间: 2017-04-11, 2017-06-16, 2017-08-23; 采用时间: 2017-09-06

两天后才能提交检查结果,这样的延迟往往增加医患沟通的成本,有可能错过治疗的时机.另外,即使临床医生可以通过糖尿病引起的眼底血管异常类损伤来有效地鉴别 DR,糖尿病高发地区也未必有足够的眼科专业医生.随着糖尿病个体的数量持续增长,预防由于 DR 引起的失明所需的基础设施将变得更加不足.

人们已经认识到,需要一种全面且自动化的 DR 筛选方法.已有的一些工作已经利用图像分类、模式识别和机器学习技术取得了一些进展.随着深度学习技术全面应用于实践,使用深度学习模型处理医学影像成为未来一个重要的发展方向,未来会有更加有效的模型能够自动从眼底照片中识别出糖尿病视网膜病变.

1 相关工作

当前,在计算机视觉领域,最有效的深度模型是卷积神经网络(convolutional neural network,简称 CNN)模型. ImageNet 比赛(ImageNet Large Scale Visual Recognition Challenge,简称 ILSVRC)见证了 CNN 模型在计算机视觉领域的发展.从 2012 年的 AlexNet^[1]、2014 年的 VGGNet^[2]和 GoogLeNet^[3]到 2015 年的 Residual Network^[4],CNN 的规模和深度在急剧地增加;同时,对物体的识别错误率也在快速地降低.目前,针对竞赛的 1 000 个类别物体的识别错误率,CNN 模型已经低于人类.此外,CNN 已成功应用于大量的一般性识别任务,如对象检测^[5]、语义分割^[6]和轮廓检测^[7].

在自动分析眼底照片方面,现有的工作主要集中在对血管和视盘的分割上.有一种名为 Deep Vessel^[8]的方法使用深度学习和条件随机场来做视网膜血管分割.DRIU^[9]方法提出了一个统一的框架进行视网膜血管和视盘的分割.由于微血管瘤通常作为 DR 判别的标志性病变,有的方法^[10]通过检测微血管瘤间接进行 DR 检测.

在本文中,我们利用深度学习的最新进展,构建了两级 CNN 进行自动化 DR 检测:第 1 级 CNN 用于原始眼底照片的特征提取,第 2 级 CNN 用于特征变换与特征混合.通过两级 CNN 的处理,网络得出 DR 检测的结果.

2 第 1 级 CNN 的设计

第 1 级 CNN 共包含 3 个组件:输入部分、输出部分和多分支特征变换部分.其中,输入部分类似于 VGGNet 的结构,用于接受原始图像的输入,并进行初步变换以降低数据尺寸;输出部分支持多种输出模式,得出 DR 的检测结果,用于对 CNN 模型的特征提取能力进行评估;多分支特征变换部分整合了 GoogLeNet 和 Residual Network 的基础结构,兼顾了网络的深度、宽度和子网个数,能够提取更深层次的特征.具体架构如图 1 所示.

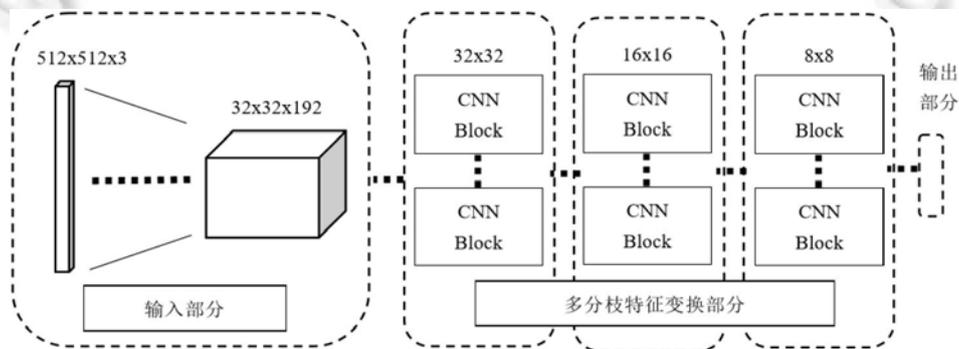


Fig.1 Level 1 CNN framework of the proposed DR detection

图 1 用于 DR 检测的 CNN 模型第 1 级框架

2.1 输入和输出设计

由于原始眼底照片的分辨率很高,早期 DR 病变的特征又很微小,高分辨率图像能够包含更多的原始特征细节,对于 DR 检测是有意义的.但是,过高的分辨率输入会占用大量的显存,而且模型的特征空间本身也需要占用大量的显存,为了平衡原始图像的大小和模型的复杂度,经过多次尝试,我们使用 512×512 的彩色图像作为输

入,这样的分辨率能够在主流的 8G 显存的显卡上完成模型的训练.

原始图像具有最大的尺寸以及最低级别的特征,为了后续的处理,需要缩小图像的尺寸,同时进行特征提取.输入部分的设计就是用来完成这一工作的.由于 CIFAR-10 是多个 CNN 模型的基准测试数据集,其数据为 32×32 的彩色图像.因此,输出部分使用 32×32 作为输出尺寸满足特征提取需求.布局设计见表 1.

Table 1 Layout of the input part

表 1 输入部分的布局设计

类型	卷积核数/窗口尺寸	步长	输入尺寸/输出尺寸
卷积	32/3×3	2	512×512×3/256×256×32
卷积	32/3×3	1	256×256×32/256×256×32
卷积	64/3×3	1	256×256×32/256×256×64
卷积	64/3×3	2	256×256×64/128×128×64
卷积	64/3×3	1	128×128×64/128×128×64
卷积	128/3×3	1	128×128×64/128×128×128
最大池化	3×3	2	128×128×128/64×64×128
卷积	192/3×3	1	64×64×128/64×64×192
卷积	192/3×3	1	64×64×192/64×64×192
最大池化	3×3	2	64×64×192/32×32×192

根据最新的《美国眼科协会临床指南:糖尿病视网膜病变》(2016 版)^[11],糖尿病视网膜病变早期治疗研究(ETDRS)分类标准见表 2.

Table 2 Classification of DR in the early treatment of DR study

表 2 糖尿病视网膜病变早期治疗研究分类

序号	疾病严重程度级别
1	轻度非增生型
2	中度非增生型
3	重度非增生型
4	早期增生型
5	高危增生型

根据上述分类,可以将 DR 检测看作是一个分类问题.然而,大多数分类问题,各类别之间是弱相关的,可称其为正交分类.例如猫、桌子、汽车、树.如果分类产生错误,那么将猫识别成汽车和将猫识别成树是等价的.由于 DR 判断的主要依据是微血管瘤、硬性渗出、棉絮斑、新生血管的多少等,为方便起见,在实际诊疗中得出病人 DR 的严重程度分级,可视为将回归问题转化为分类问题.这时,各类别之间是强相关的,可称其为非正交分类.如果误诊无法避免,则将中度 DR 识别成轻度 DR 与将中度 DR 识别成没有 DR 是完全不同的.因此,医学问题对分类的要求更加严格,一般不使用正确率而是使用一致性来评估分类的结果.因此,输出部分需要同时支持分类和回归.输出部分的输入尺寸为 $8 \times 8 \times 2048$,其布局见表 3.

Table 3 Layout of the output part

表 3 输出部分的布局设计

分类	回归
平均池化,窗口尺寸为 8×8 ,步长为 1	
卷积层,卷积核数为 5,窗口大小为 1×1	Dropout 层,丢弃概率为 0.5 全连接层,单元数为 1 024
扁平化	特征池化,窗口大小为 2
Softmax 输出	全连接层,单元数为 1

2.2 多分支设计

2014 年,ImageNet 比赛的冠军模型是 GoogLeNet.这个模型使用了平行扩展的架构,并通过合并多个分支的权重来进行特征的归集.2015 年,ImageNet 比赛的冠军模型是 Residual Network.这个模型使用了元素级权重累

加的方式来进行跨层权重计算,以便支持超过 100 层的超深网络结构.为了在深度和宽度上进行拓展,这里参考 Inception v3 和 ResNet v2 的结构,设计了一个多分支的网络,用于进行细节特征的提取.

根据人们所做的大量尝试以及 Inception 结构和 ResNet 结构的特点,为了得到一个特征抽取能力较强的神经网络,网络结构的设计原则可总结为如下 4 点.

- 更大的有效深度;
- 包含尽可能多的子网;
- 管理好梯度传递,防止梯度爆炸和梯度消失;
- 控制网络中的参数总量.

相关研究所做的模型改进也遵循了这样的原则.Zagoruyko 等人的研究^[12]指出:原始 ResNet 网络的深度过深,为了追求深度而牺牲了网络宽度,而网络宽度的增加能够提高性能.在参数规模相同时,更宽的网络比更深的网络好训练.Lee 等人提出的 Wide-Residual-Inception 结构^[13],在基本 Residual Block 中,用 Inception 模块代替 3×3 的卷积操作,取得了良好的效果.Wu 等人的研究^[14]进一步指出:ResNet 结构的主要优势是通过快捷方式传递梯度,在改善梯度消失问题的同时,增加了网络中的子网个数.神经网络能够通过子在子网之间共享权重获得良好的正则化.

与参考文献[13]中以 ResNet 结构为基础,将 Inception 模块嵌入 ResNet 不同,本文以 Inception 结构为基础,通过在 Inception 模块间及模块内添加 Residual 连接,构建多分支连接来实现相应的设计目标.

整体上,从 32×32 到 16×16 再到 8×8 ,整个多分支部分的架构分为 3 个阶段.

- 第 1 阶段的输入为 $32 \times 32 \times 192$,输出为 $32 \times 32 \times 288$.

使用 3 个 4 分支的 Inception 单元,并将后面两个输出进行 Residual 连接,作为第 1 阶段的输出.如图 2 所示.

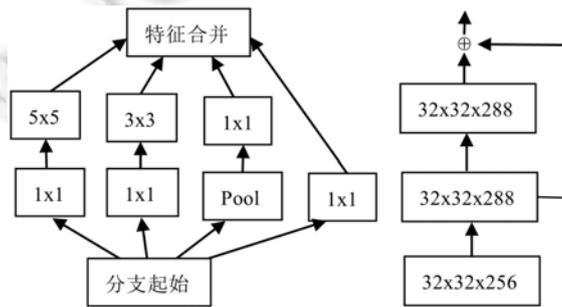


Fig.2 Standard inception module
图 2 标准 Inception 模块

第 1 阶段的网络规模较小,而且输入和输出宽度不同.由于 Residual 连接的效果在 Identity 条件下最好,因此仅对两个宽度相同的 Block 添加了 Residual 连接.

之后,使用另外一个 3 分支的 inception 模块来降低数据的尺寸,从 32×32 降低到 16×16 ,参见文献[15].

- 第 2 阶段处理 16×16 的数据,输入为 $16 \times 16 \times 768$,输出为 $16 \times 16 \times 768$.

使用 4 个 4 分支的 Inception 单元,如图 3 所示,整体上对分支进行了混合连接,并将输入和最后一个 Inception 模块的输出进行了 Residual 连接,作为第 2 阶段的输出.如图 4 所示.

第 2 阶段的网络规模较大,而且输入输出宽度相同,符合 Identity 条件,因此在输入和输出之间添加了 Residual 连接.此外,还在模块间添加了多条分支连接,用于构建尽可能多的子网.

传统构建子网的方法大多是复制相同的网络结构进行堆叠,因此子网结构并不丰富.本文在这里虽然也进行了相同结构网络的堆叠,但通过对不同模块引入不同的输入连接点,使得子网结构尽可能地丰富.更丰富的网络结构有利于学习到更多样化的特征.

之后,使用另外一个 3 分支的 inception 模块来降低数据的尺寸,从 16×16 降低到 8×8 ,参见文献[15].

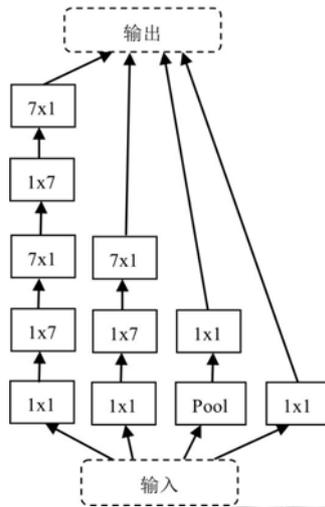


Fig.3 Inception module used in Stage 2

图3 第2阶段 Inception 模块

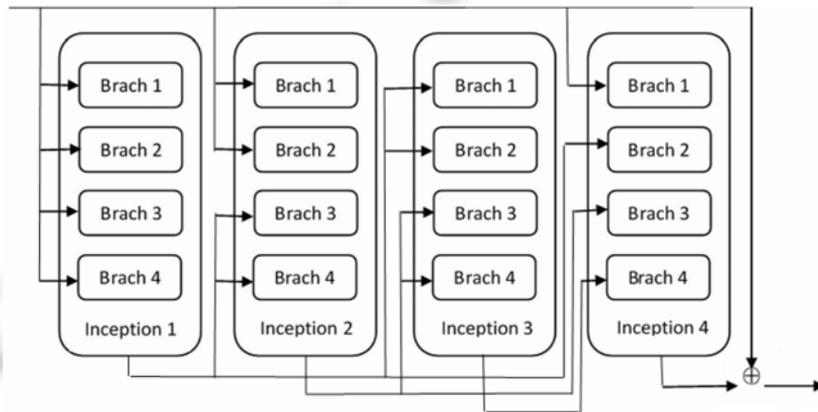


Fig.4 Multi-Branch design in Stage 2

图4 第2阶段多分支设计

- 第3阶处理 8×8 的数据,输入为 $8 \times 8 \times 1280$,输出为 $8 \times 8 \times 2048$.

使用2个6分支的 Inception 单元,并将这两个 Inception 模块的输出进行了 Residual 连接,作为第3阶段的输出.如图5所示.

网络的训练依靠的是误差的反向传播,并通过链式法则进行计算,这意味着梯度计算需要根据网络深度进行多次连乘.由于神经元输出信号尺度问题,传递到浅层的信号会产生剧烈的衰减.

因为第3阶段的网络也不符合 Identity 的条件,而且已经非常接近网络底部,不会有严重的梯度消失的问题,所以并没有建立输入到输出的 Residual 连接.与第1阶段不同,在这一阶段,输入通过两个相同的路径调整了宽度,并对两个宽度相同的输出进行了 Residual 连接,作为最终的输出.这种结构的网络子网等价深度最大,有利于学习到更为抽象的特征.

综上,本文设计的第1级 CNN 网络的有效深度与 InceptionV3 保持一致,通过添加 Residual 连接和多分支连接,在参数个数可控的情况下,既增加了子网个数和结构多样性,又改善了梯度消失的情况.

至此,模型的1级 CNN 设计完成.通过1级 CNN 模型,已经可以进行 DR 的检测工作了.

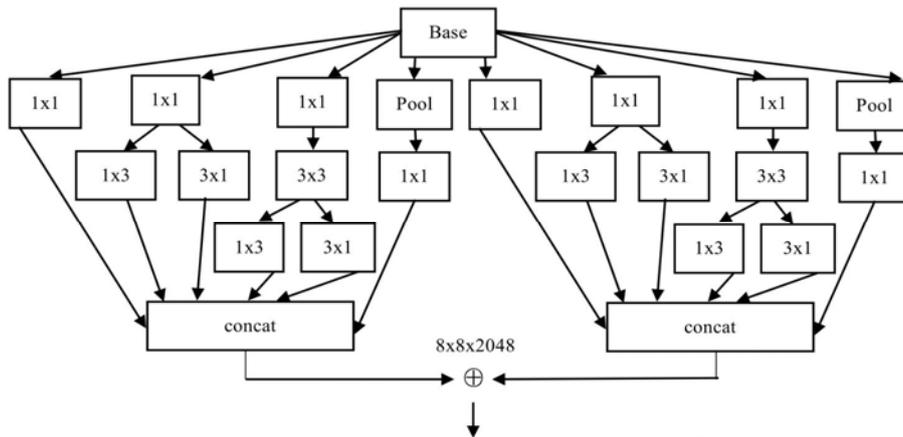


Fig.5 Two inception modules in Stage 3
 图 5 第 3 阶段的两个 inception 模块

3 第 2 级 CNN 的设计

通常,在进行 DR 检测时,医生都会获取被筛查对象双眼的眼底照片进行判断.对于用于辅助诊断的 CNN 模型来说,同时处理双眼数据也是具有重要意义的.因为虽然病灶可能出现的具体位置是随机的,但由同一病因引起的病灶信息在双眼的分布具有强相关性.

通过对第 1 级 CNN 模型的训练,模型已经能够较好地提取单眼眼底照片中的特征.需要构建另外一个 CNN 模型,通过同时处理双眼特征来提高检测的效果.

按照迁移学习的模式,将第 1 级 CNN 模型作为双眼特征联合 DR 检测任务的固定特征提取器,再将提取的特征做进一步的变换与融合,充分利用深度学习可迁移的特性,能够大幅度提高 DR 的检测能力.

首先需要进行特征设计,这里参考了 Kaggle 机器学习竞赛平台上糖尿病视网膜膜病变检测比赛第 2 名的方案.由于多分支特征变换部分的输出为 $8 \times 8 \times 2048$,根据输出部分的设计,在输入分类器前统一进行了平均池化操作,变换成了 $1 \times 1 \times 2048$.由此,一张眼底照片图像最终得到了 2 048 个特征.以一副原始图像为基础,进行几何色彩等变换生成若干图像,进而得到更多特征的方法称为数据增强,广泛应用于 CNN 的训练.考虑到数据增强,围绕一张眼底照片能够得到多组 2 048 个特征,这些特征具有很高的相似性.以这一系列特征的均值和方差为特征来描述原始图,就将原始 2 048 个特征扩展为 4 096 个特征.另外,考虑到糖尿病一般同时侵害双眼,左、右眼往往同时产生病变,相关特征之间具有一定的内在联系,因此,将双眼特征进行整合也能增加有效的特征.最终的特征数为 8 192.由于所有的操作均没有损失相对位置信息,因此将特征形状重新格式化为 $32 \times 32 \times 8$,后续仍然可以使用 CNN 进行计算.具体特征设计的流程如图 6 所示.

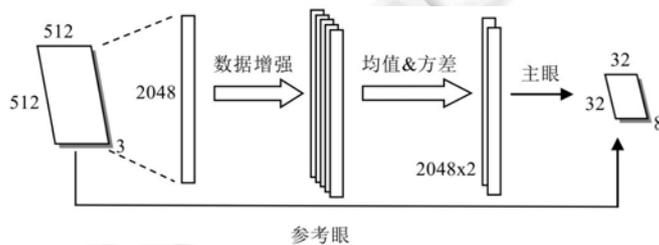


Fig.6 Steps of feature extraction
 图 6 特征提取的步骤

由于特征不再稀疏,这里仅设计了一个类似于 AlexNet 的简单 CNN 来进行分类.另外,为了更有力地支持医生的诊断,第 2 级 CNN 采用回归目标函数.拓扑结构见表 4.

Table 4 Design of level 2 CNN

表 4 第 2 级 CNN 的设计

类型	参数	步长
卷积	32/3×3	1
最大池化	3×3	2
卷积	64/3×3	1
最大池化	3×3	2
卷积	128/3×3	1
最大池化	3×3	2
全连接	500	-
Dropout	0.5	-
全连接	500	-
特征池化	2	-
全连接	1	-

4 实验验证

实验环境如下:

- 硬件: Intel i5-6600K, 32G, Nvidia GTX 1080;
- 软件: Ubuntu 16.04, CUDA8.0, cudnn5.1, Theano0.8, Lasagne, Keras1.x.

目前,针对 DR 检测最大的开放数据源就是 Kaggle 的 DR 检测竞赛(Kaggle diabetic retinopathy detection competition)数据.模型的优化目标为最小的平方加权 Kappa 值,即模型诊断结果与专家诊断结果的一致性.本次实验使用了 35 124 张图片进行训练,数据具体信息见表 5.

Table 5 Details of data

表 5 数据信息

类别	名称	图片数	占比(%)
0	Normal	25 808	73.48
1	Mild NPDR	2 443	6.96
2	Moderate NPDR	5 292	15.07
3	Severe NPDR	873	2.48
4	PDR	708	2.01

为了进行交叉验证,实验对数据进行了分割,90%的数据用于训练,10%的数据用于验证.另外,由于原始数据严重有偏,在训练中参考 Gary King 的方法^[16]进行了数据的重采样.网络的训练使用 Nesterov momentum^[17]作为优化方法,权重初始化的方法参见文献^[18].

针对第 1 级 CNN,当使用分类方法进行训练时,得到的最好分类准确率为 0.838 361,最好的分类 Kappa 一致性为 0.754 13.具体训练曲线如图 7 所示.当使用分类目标函数(交叉熵损失函数)时,模型能够得到较高的分类准确率,但 Kappa 一致性和准确率始终有较大差距.

当使用回归方法进行训练时,得到的最好分类准确率为 0.783 153,最好的分类 Kappa 一致性为 0.765 98.具体训练曲线如图 8 所示.当使用回归目标函数(均方差损失函数)时,模型能够得到较高的 Kappa 值,但分类准确率会有较大的下降.

通过图 7、图 8 可以看出,使用不同的损失函数可使模型获得不同的优化结果,这与输出层神经元的敏感度直接相关.事实上,以均方差或交叉熵作为损失函数的 CNN,其输出层神经元的敏感度均为它的激活值与目标值的差值.

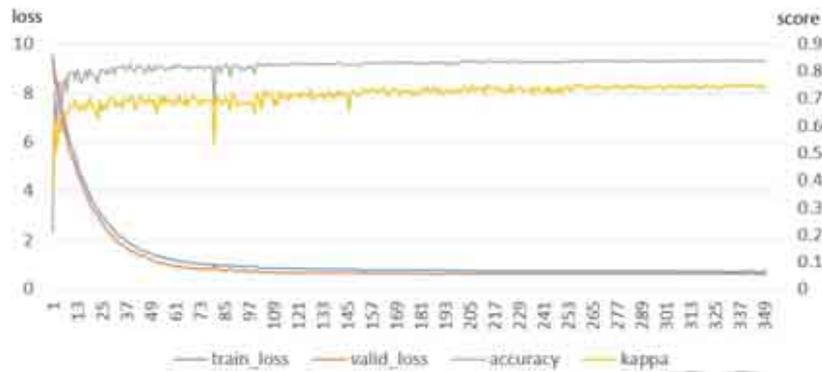


Fig.7 Training curves of level 1 CNN (classification) on the Kaggle dataset
图 7 第 1 级 CNN 在 Kaggle 数据集上进行分类训练的训练曲线

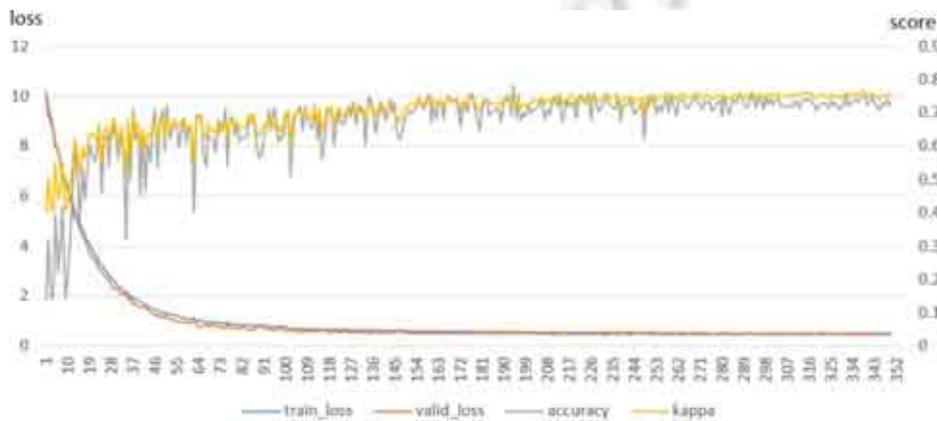


Fig.8 Training curves of level 1 CNN (regression) on the Kaggle dataset
图 8 第 1 级 CNN 在 Kaggle 数据集上进行回归训练的训练曲线

在使用交叉熵损失函数时,模型的输出神经元使用的是 Softmax 激活函数.在计算梯度时,只考虑真值所在的类别得到的概率值与 1 之间的差,剩余概率的分配则不予考虑,即网络只关心如何更正确,而不关心是怎么错的.从图 7 可以明显看到,网络只是在优化准确率的同时,顺带提高了 Kappa 一致性而已.当准确率不能进一步提高时,Kappa 也无法得到提高了.

在使用均方差损失函数时,模型的输出神经元使用的是线性激活函数,输出反映的是对 DR 的严重程度的判断.计算梯度时,输出与真值的差反映的是判断错误的程度,即网络不考虑分类结果,而只关注如何得到正确的 DR 严重程度值.这样的优化方向更符合 Kappa 的需要.从图 8 可以明显看到,训练曲线波动很大,分类准确率不佳,但得到了更好的 Kappa 结果.因此,若要得到较好的 Kappa 值,应使用回归的方式进行分类,这也是在第 2 级 CNN 的设计中直接采用回归目标函数的原因.

如果将第 1 级 CNN 视为特征提取器,而不是分类器,引入第 2 级 CNN 作为分类器,则需要评估哪一种目标函数对特征提取更为有利.实验中,针对第 1 级 CNN 做特征提取,分别使用回归方式和分类方式进行特征提取,对于回归特征和分类特征的训练结果如图 9 所示.

使用回归特征得到的结果是:准确率为 0.812,Kappa 值为 0.833;使用分类特征得到的结果是:准确率为 0.816,Kappa 值为 0.838.使用分类特征的结果略优于使用回归特征的结果.

为了避免偶然性,实验在使用分类目标函数时提取了 3 组,共 150 个批次的特征;在使用回归分类目标函数时,提取了 3 组,共 150 个批次的特征.实验分别利用第 2 级 CNN 对分类获取的 3 组特征,回归获取的 3 组特征,

以及全部 6 组特征进行了分类,分别得到了 3 个实验结果.其中,使用 3 组回归特征时的综合 Kappa 值为 0.842,使用 3 组分类特征时的综合 Kappa 值为 0.85,使用全部 6 组特征时的综合 Kappa 值为 0.848.由此得出,使用分类目标函数提取的特征更有效.

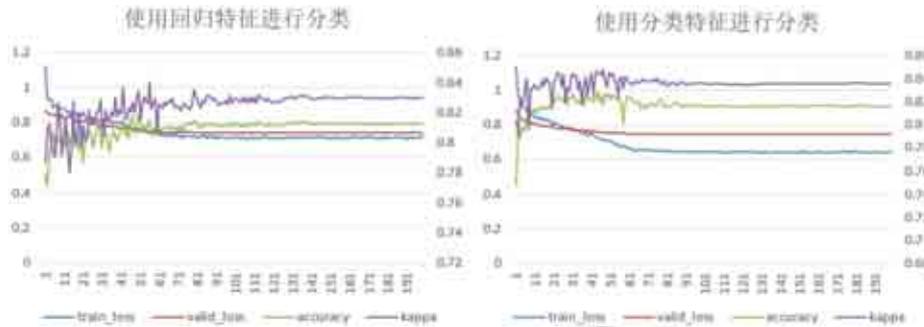


Fig.9 Comparison of training curves for the level 2 CNN training using two different feature sources

图 9 使用两个不同特征来源进行第 2 级 CNN 训练的训练曲线比较

使用分类目标函数时,Softmax 激活函数输出 5 个值,使得每一个值都可以建立与图像特征之间的关系.而使用回归目标函数时,线性激活函数只输出 1 个值,只能通过这 1 个值建立与图像特征之间的关系.与图像特征之间建立的关系越多越复杂,就更能挖掘到有意义的图像特征.这可能是使用分类特征能够得到更好的 Kappa 值的原因.

针对验证集的 3 514 张图片,综合 Kappa 取 0.85 时的混淆矩阵见表 6.

Table 6 Confution matrix

表 6 混淆矩阵

医生	模型				
	无 DR	轻度	中度	重度	增生
无 DR	2 408	153	16	0	0
轻度	109	114	32	0	0
中度	68	95	325	30	0
重度	0	3	40	50	0
增生	0	2	13	36	20

有无 DR 漏诊/误诊量为 329,误诊率为 0.093 6.

另外,二级 CNN 使用回归目标函数时,能够更好地指导医生得出诊断结果.在实践中,某 DR 检测样本结论见表 7.

Table 7 Conclusion of computer-aided diagnosis

表 7 计算机辅助诊断结论

仅供参考	DR 指标(0~4)	参考概率(0~1)	预测结果
左眼	0.700 1	0.044 7	轻度非增生型 DR
右眼	0.471 4	0.776 3	无 DR

由表 7 可知,左眼的 DR 指标值,即回归值为 0.7,是由第 2 级 CNN 给出的.由于回归转分类的阈值分别为 0.5,1.5,2.5 和 3.5,因此预测结果为轻度非增生型 DR.由于参考概率为 0.044 7 来源于第 1 级分类 CNN 模型,可以看出,最终的预测结果分类模型与回归模型不一致,以回归模型为准.右眼虽然预测结果为无 DR,且通过参考概率了解到分类模型与回归模型的结论一致,然而 DR 指标非常接近于 0.5 的阈值,因此,右眼是一个疑似的轻度非增生型 DR.

5 细粒度图像分类改进

在进行 CNN 模型设计时,通常着重考虑以下两点,以便根据任务特点来调整网络结构.

- 像素级别的精度,即怎样处理局部信息;
- 目标级别的抽象,即怎样整合全局信息.

这两点是相辅相成、不可分割的.在平衡这两点时,最主要的约束条件就是模型的复杂度.

CNN 的复杂度主要是由层数、各层卷积核大小以及各层卷积核个数决定.

$$cnn_param = \sum_l [k_num_l \times (k_size_l)^2 + k_num_l],$$

其中, cnn_Param 是模型中所有卷积层的参数, k_num_l 是第 l 层的卷积核数量, k_size_l 是第 l 层的卷积核大小.

DR 检测任务属于细粒度图像分类(fine-grained image classification)任务,因为相邻 DR 级别的眼底图像差异很小,需要经过专业培训的医生才能进行诊断.使用解决通用图像任务的方法,模型的能力会受到很大的限制.在这方面,本文也做了初步的尝试.

针对细粒度图像分类,有一些算法使用了部位标注点等额外的人工标注信息,以便定位关键特征来源.而在 DR 检测场景下,医生只会给出诊断结果,不会在眼底照片中标记特征的区域和特征的性质.因此,仅使用图像级别标注信息进行 DR 检测,需要综合使用若监督学习方法定位特征区域.已有的算法通常使用反卷积或梯度计算来获得特征区域.

在 CNN 中,原始图像经多次卷积计算,通过根据梯度不断修正的权重,连接至最终的输出结果.因此,卷积输出信号和梯度输出信号直接决定了原始图像到最终结果的映射关系.由于卷积计算并不损失空间相对位置信息,且卷积信号和梯度信号维度数与原始图像一致,因此可以通过计算这些信号来进行特征定位.

如图 10 所示,在原始图像中标记出与结果的得出关系最密切的特征来源区域,这一区域的标定与最后一个卷积层的输出及其与最终分类结果的梯度相关.只有输出特征明显且梯度明显的区域才会被标记.而相关区域被视为特征来源区域.

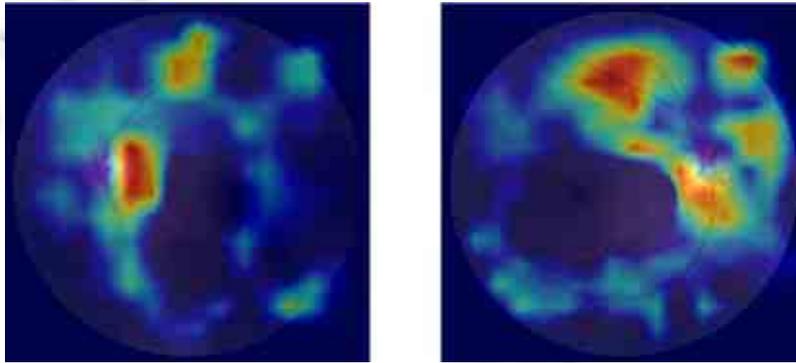


Fig.10 Weak supervisory characterization

图 10 弱监督特征定位

原始眼底图像的分辨率很高,如果使用原始分辨率进行建模,则需要大量的下采样操作.虽然模型复杂度与网络输入图像的分辨率并不直接相关,但为了获得高级别的抽象(从局部到全局),每一次下采样都会大幅度提高卷积核的个数.过高的分辨率会显著增加下采样的次数,从而导致模型的参数过多,超出了计算机求解的能力.由于医学影像的特征,相对于一般场景下的图像,特征的特异性较差,特征的分布较为稀疏,因此,获得特征的大致区域是一个可行的改进方向.

在本次实验中,图像的分辨率被降低到了 512×512 ,这样会损失大量的图像细节.若能首先获取特征区域,再针对具体区域使用更大的分辨率进行分析,就能获得更加准确的预测结果.本文已在分类和弱监督特征定位方

面进行了有益的探索.将两者相结合做进一步的改进,还需要更加深入的研究与实验.

6 结论与展望

本文首先提出了一个包含多分支结构的 CNN,用于 DR 检测,并提出将这一多分支 CNN 视为特征提取器,构建第 2 级 CNN 进行二次特征分类的方式进行 DR 检测.在实验环节,使用 Kaggle 竞赛数据对模型进行了训练,得出使用回归目标函数更有利于一级 CNN 获得较优结果,而用分类目标函数更有利于一级 CNN 获得较优特征的结论.最终,通过两级 CNN 的建模和训练,在验证集上获得了 0.85 的 Kappa 结果.这一结果已达到了参考 Kaggle 竞赛的最高水平.另外,本文给出了利用弱监督学习获得特征定位的方法,并指出了进行细粒度图像分析的实现方向.

在网络架构设计方面,进行多分支设计、平衡网络的深度和宽度是发展的大方向.根据 Inception 架构衍生的 Xception^[19]和根据 ResNet 架构的衍生的 ResNeXt^[20],都在这方面进行了有益的尝试.其中,ResNeXt 获得了 2016 年度 ImageNet 比赛的亚军.

针对 DR 检测,还有很多工作有待于完成,例如,目前还没有一个专门针对 Kappa 一致性进行优化的目标函数,使用交叉熵函数或均方差函数均不是针对 Kappa 进行直接优化.针对这一点,已有文献进行了一些探索^[21],竞赛第 5 名的解决方案也进行了有益的尝试,但效果并不显著.这将是未来研究的一个方向.

另外,数据集偏小是目前围绕医学影像数据集的一个普遍问题,由于模型并没有显著的过拟合问题,一旦能够获得更多的数据,必定能够得到更好的结果.事实上,受制于成本、时间等因素,完全依赖有标注数据是不现实的.由于 Kaggle 竞赛还有 5 万多张未标记数据,如何构建一个半监督学习模型、充分利用未标记数据,也是未来研究的一个重要方向.

利用人工智能技术进行医学影像辅助诊断是一个快速发展的领域,最近获得了多个令人瞩目的成果.2016 年底,Google 在医学顶级期刊《JAMA》上发表的论文显示,其 DR 检测算法的表现与眼科医生旗鼓相当^[22].2017 年初,斯坦福大学在《Nature》上发表的论文显示,其构建的皮肤癌检测模型的正确率及灵敏度均可达到甚至超过专业医师水平^[23].随着深度学习技术的快速发展以及医学影像数据的不断开放,人工智能在早期疾病筛查和医生辅助诊断方面正在逐步走向成熟.

References:

- [1] Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, 2017,60(8):84–90. [doi: 10.1145/3065386]
- [2] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. In: *Proc. of the Int'l Conf. on Learning Representations*. 2015.
- [3] Szegedy C, Liu W, Jia YQ, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A. Going deeper with convolutions. In: *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*. 2015. [doi: 10.1109/CVPR.2015.7298594]
- [4] He KM, Zhang XY, Ren SQ, Sun J. Deep residual learning for image recognition. *arXiv: 1512.03385 [cs.CV]*, 2015.
- [5] Girshick R, Donahue J, Darrell T, Malik J. Region-Based convolutional networks for accurate object detection and segmentation. *IEEE T-PAMI*, 2016,38(1):142–158. [doi: 10.1109/TPAMI.2015.2437384]
- [6] Hariharan B, Arbelaez P, Girshick R, Malik J. Hypercolumns for object segmentation and fine-grained localization. In: *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*. 2015. [doi: 10.1109/CVPR.2015.7298642]
- [7] Xie SN, Tu ZW. Holistically-Nested edge detection. In: *Proc. of the Int'l Conf. on Computer Vision*. 2015. [doi: 10.1109/ICCV.2015.164]
- [8] Fu HZ, Xu YW, Lin S, Wong DWK, Liu J. Deep vessel: Retinal vessel segmentation via deep learning and conditional random field. In: *Proc. of the Medical Image Computing and Computer Assisted Intervention*. 2016. [doi: 10.1007/978-3-319-46723-8_16]
- [9] Maninis KK, Pont-Tuset J, Arbelaez P, Van Gool L. Deep retinal image understanding. In: *Proc. of the Medical Image Computing and Computer Assisted Intervention*. 2016. [doi: 10.1007/978-3-319-46723-8_17]
- [10] Haloi M. Improved microaneurysm detection using deep neural networks. *arXiv: 1505.04424 [cs.CV]*, 2015.

- [11] American Academy of Ophthalmology Retina/Vitreous Panel. Preferred practice pattern® guidelines. In: Proc. of the Diabetic Retinopathy. San Francisco: American Academy of Ophthalmology, 2016. <http://www.aao.org/ppp>
- [12] Zagoruyko S, Komodakis N. Wide residual networks. arXiv:1605.07146 [cs.CV], 2016.
- [13] Lee YW, Kim HE, Park ES, Cui XN, Kim HK. Wide-Residual-Inception networks for real-time object detection. In: Proc. of the IEEE Intelligent Vehicles Symp. 2017. [doi: 10.1109/IVS.2017.7995808]
- [14] Wu ZF, Shen CH, van den Hengel A. Wider or deeper: Revisiting the ResNet model for visual recognition. arXiv: 1611.10080 [cs.CV], 2016.
- [15] Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the inception architecture for computer vision. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. 2015. [doi: 10.1109/CVPR.2016.308]
- [16] King G, Zeng LC. Logistic regression in rare events data. Political Analysis, 2001,9:137–163. [doi: 10.1093/oxfordjournals.pan.a004868]
- [17] Sutskever I, Martens J, Dahl G, Hinton G. On the importance of initialization and momentum in deep learning. In: Proc. of the 30th Int'l Conf. on Machine Learning. 2013. 1139–1147.
- [18] He KM, Zhang XY, Ren SQ, Sun J. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In: Proc. of the Int'l Conf. on Computer Vision. 2015. [doi: 10.1109/ICCV.2015.123]
- [19] Chollet F. Xception: Deep learning with depthwise separable convolutions. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. 2017.
- [20] Xie SN, Girshick R, Dollár P, Tu ZW, He KM. Aggregated residual transformations for deep neural networks. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. 2017.
- [21] Vaughn D, Justice D. On the direct maximization of quadratic weighted Kappa. arXiv: 1509.07107 [cs.LG], 2015.
- [22] Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, Venugopalan S, Widner K, Madams T, Cuadros J, Kim R, Raman R, Nelson PC, Mega JL, Webster DR. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA Online, 2016. <https://jamanetwork.com/journals/jama/article-abstract/2588763>
- [23] Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, Thrun S. Dermatologist-Level classification of skin cancer with deep neural networks. Nature, 2017,542:115–118. [doi: 10.1038/nature21056]



庞浩(1979 -),男,陕西岐山人,博士生,主要研究领域为机器学习,计算机视觉,医学影像计算机辅助诊断.



王枞(1958 -),女,博士,教授,博士生导师,CCF 高级会员,主要研究领域为智能信息处理,网络信息安全,可信计算与服务,医疗健康数据分析挖掘.