

基于迁移学习的糖尿病视网膜病变眼底图像分类研究

粘伟杰, 金海龙

燕山大学电气工程学院生物医学工程系, 河北 秦皇岛

收稿日期: 2024年10月25日; 录用日期: 2025年1月16日; 发布日期: 2025年1月24日

摘要

全世界有4亿多人患有糖尿病, 中国的患病人数超过1个亿。而其中1/3的糖尿病患者会出现可能导致视力损伤甚至失明的并发症——糖尿病视网膜病变。因此, 检查眼底是每个患有糖尿病的人都需要做的检查, 病情越重检查频率越高。而如此多的检查给医疗人员造成很大的工作负担, 同时也增加了国家的医疗支出。为了更经济、高效和准确地检查, 需要建立计算机自动诊断系统。因此, 这项工作成为科研人员重要的研究课题。随着深度学习在图像分类任务中取得的重大进步, 利用这一技术建立高效的糖尿病自动诊断系统成为研究热点。然而, 由于医疗数据自身的特点(属于小数据集并且数据不平衡)以及神经网络的“黑箱”问题, 使得利用深度学习进行疾病分类效果不理想。针对以上问题, 本文从以下方面进行研究: 提出了一个基于迁移学习和DenseNet121的可解释性分类器。在预先训练的DenseNet121网络上, 利用微调迁移学习技术, 更改部分网络结构, 实现DR自动诊断任务, 实验结果为敏感性0.77、特异性0.91、准确率88%。同时对神经网络解释性可视化, 采用梯度加权类别激活映射对分类器的分类结果可视化。通过这种方式, 可显示输入图像的哪个部分对分类器的分类结果影响更大。通过这种可视化的方式发现, 虽然分类器达到了较高的准确率, 但分类器并非总是关注那些病灶部位, 并且存在对类别少的眼底图像样本的偏见, 反映出不平衡样本所训练的模型存在的问题, 这让我们更好地了解我们所训练的模型所存在的不足之处。

关键词

计算机辅助诊断, 深度学习, 迁移学习, 糖尿病视网膜病变

A Study on Image Classification of Diabetic Retinopathy Based on Transfer Learning

Weijie Nian, Hailong Jin

Department of Biomedical Engineering, School of Electrical Engineering, Yanshan University, Qinhuangdao Hebei

Received: Oct. 25th, 2024; accepted: Jan. 16th, 2025; published: Jan. 24th, 2025

文章引用: 粘伟杰, 金海龙. 基于迁移学习的糖尿病视网膜病变眼底图像分类研究[J]. 生物医学, 2025, 15(1): 232-243.
DOI: 10.12677/hjbm.2025.151027

Abstract

More than 400 million people worldwide suffer from diabetes and more than 100 million in China. A third of diabetics have complications that can lead to visual impairment or even blindness, namely, diabetic retinopathy. Therefore, fundus examination is a necessary check for everyone with diabetes, and the more severe the condition, the higher the frequency of the check. Such a large number of examinations places a heavy workload on medical staff and also increases the country's medical expenditure. To conduct the examination more economically, efficiently, and accurately, it is necessary to establish a computer-aided diagnostic system. Hence, this work has become an important research topic for scientific researchers. However, due to the characteristics of medical data itself (belonging to small datasets and data imbalance) as well as the "black box" issue of neural networks, the effect of using deep learning for disease classification is not satisfactory. In response to the above problems, this paper conducts research in the following aspects: an interpretable classifier based on transfer learning and DenseNet121 is proposed. By utilizing fine-tuning transfer learning techniques on the pre-trained DenseNet121 network and modifying some of the network structures, the automatic diagnosis task of DR is realized. The experimental results show a sensitivity of 0.77, specificity of 0.91, and accuracy of 88%. At the same time, the interpretability of the neural network is visualized by using gradient-weighted class activation mapping to visualize the classification results of the classifier. Through this method, it can be shown which part of the input image has a greater impact on the classification results of the classifier. By this visualization approach, it is found that although the classifier has achieved a high accuracy rate, it does not always focus on the lesion areas, and there is a bias towards fundus image samples with fewer categories. This reflects the problems existing in models trained with imbalanced samples, allowing us to better understand the shortcomings of the models we have trained.

Keywords

Computer-Aided Diagnosis, Deep Learning, Transfer Learning, Diabetic Retinopathy

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

1.1. 研究背景

DR 临床诊断过程: 临床医生通过分析微血管动脉瘤(MA)、渗出、出血、新生毛细血管和棉絮状点的数量、位置、大小来判断 DR 病情[1]。DR 自动诊断系统的发展经历了几个阶段。针对不同症状有几类糖尿病视网膜病变计算机辅助诊断系统: 主要有基于微血管动脉瘤、基于渗出、基于出血和基于所有特征的糖尿病视网膜病变计算机辅助诊断系统。本文主要针对基于所有特征(整个图像)的糖尿病视网膜病变开展研究和一维信号相比, 图像占用的内存空间较大, 并且图像特征更复杂, 要处理这么多特征, 算法就需要更加复杂。但因为算力不足, 在早期复杂的算法计算机无法实现。因此, 早期研究人员难以开发出诊断效果很好的自动诊断系统。另外, 眼底筛查设备很少和图像分辨率较低, 因此数据量少并且数据质量偏低, 也是限制科研人员开发效果更好的诊断系统的原因。然而, 随着数字时代的到来, 各种眼底成像设备, 更加完善的筛查以及算力更强的计算机的出现, 使得科研人员可以在大量高清晰的眼底图

像上采用更加复杂的算法进行研究。相比之下, 传统机器学习通常采用人工标注特征的方法进行 DR 诊断[2]。因此模型效果受数据标注的质量和数量影响较大, 而且这种方法费时费力。在互联网、移动设备和各种电子成像设备产生海量数据, 加上 GPU 产生的大算力以及深度学习算法的发展, 深度学习以压倒性的优势完胜基于传统手工标注的方法[2]。受深度神经网络在图像分类任务中取得的惊人成功影响, 国内外大量研究开始利用深度神经网络建立糖尿病视网膜病变自动分类系统。

1.2. 文献综述

国内, 2015 年张庆鹏开发了一个自动判断眼底是否出现微血管瘤的软件[3]。李博开发了自动诊断新生血管的算法, 该算法主要针对出现威胁视力的 DR 及时做出诊断[4]。2017 年, 丁蓬莉在 CompactNet 网络上分类 DR 达到了 0.69 的准确率[5]。罗院生使用 SegVessel 网络在 Kaggle 的 diabetic-retinopathy-detection 数据集上达到了 0.93 的准确率[6]。马文俊采用 DrNet 网络, 在上述数据集上达到了 0.81 的准确率[7]。在 California Healthcare Foundation 和 EyePACS 发布的眼底图像集(35,126 张彩色眼底图像)上达到了 0.81 的分类准确率。2018 年, Zhixi Li 等人在 10 万多张包括多种族 DR 眼底图像上, 建立了基于卷积神经网络的模型(DLA), 达到了较高的准确率[8]。李郭利用卷积神经网络特征提取, 用多层感知机做分类器, 在 Kaggle 的一个 6000 张 DR 眼底图像上达到了, 55.13%的准确率[9]。方全在 Kaggle 公开的 2015 年竞赛数据集(29,874 张)上, 利用改进的 AlexNet 和 VGG-16 网络进行分类, 达到了 0.77 的准确率[10]。蔡石林在 ResNet 和 DenseNet 网络基础上建立神经网络, 在 8 万张 Kaggle 数据集和 Messidor-2 数据集上达到了 0.9 的准确率。2019 年的商逸凡利用集成学习, 即将数据预处理、数据增强和卷积神经网络集成在一起对眼底图片进行分类[11], 以及 2020 年的李芳君利用从浅层神经网络到深层逐步训练的方式[12], 在各自的数据集取得了较好的分类效果。

国外, 2019 年, Dekhil 等人利用迁移学习和卷积神经网络在 Kaggle APTOS 2019 数据集上达到了 77% 的诊断准确率[13]。Yuchen、Wu 等人在预训练的 VGG19、ResNet50 网络上对 DR 疾病进行分类, 结果显示迁移学习比无迁移学习达到更高的准确率[14], 这些方法表明迁移学习提高了模型效果。另外, 在工作[15]-[18]中采用不同的算法实现 DR 的诊断, 并且准确率均已很高: 在各自不同的数据集上达到超 97% 的诊断准确率或者与眼科专家相当或者超过专家。其中, 2019 年, Jordi 等人在基于卷积神经网络的基础上, 提出了多次迭代的方法达到了专家级别的准确率。2020 年, Wang、Xiangning 等人, 在 500 多张 DR 眼底图片损伤和特征信息已经手工标注的图片数据集以及多个公开数据集上, 利用强化学习和多级迭代卷积神经网络训练, 进行 DR 分级, 达到了超 98% 的准确率。2021 年, Phong 等人利用粒子群算法分类 DR, Shankar 等人用的是 Inception 模型。

以上这些研究, 效果越来越好, 因为采用的深度神经网络越来越深, 算力也更强, 再加上迁移学习等技术的应用。从以上论文中的实验结果上看, 已经使得 DR 分类达到了很好的效果, 但由于以下问题还没有得到解决, 大部分神经网络还未能医疗领域得到实际应用: 神经网络的黑箱问题; 数据不平衡和数据量小等。虽然神经网络的分类准确率很高, 但考虑到大部分样本是不需要治疗的 DR, 而真正分类那些需要治疗的样本很少。这就是说, 如果神经网络判断所有样本都为正常眼底图像, 则其准确率也很高。而这样的分类器显然是毫无意义的, 因此数据不平衡对分类器影响很大, 而且采用的神经网络不具有可解释性, 导致无法对输出分类结果做出解释分析, 导致即使模型犯了错误也难以发现。2019 年, Li T.等人比较各个常用模型在 DR 分类和特征识别中的表现, 发现虽然大多数算法准确率很高, 但与医生判断 DR 病情(通过观察和计量眼底损失程度)并不相同, 而且这些方法在损失分割和检测上表现很不理想[19], 这也说明了高准确率分类模型仍然没有得到实际应用的原因, 因此我们除了准确率等相关标准外, 需要在可解释性上进行研究。

关于神经网络的可解释性, 深度学习要在医疗领域得到应用, 除了很高准确率外, 神经网络的诊断

过程和原理也同样重要, 否则, 医疗机构通常不会选择这种具有不确定性设备, 因为, 医疗机构要为自己所用的设备造成的后果负责。这导致深度学习难以在医疗领域大范围应用[20]。因此, 研究具有可解释性的神经网络十分重要。

关于分类模型可解释性主要分为两大类: 对分类器的可解释性和对抗生样本的可解释性。本文主要介绍对分类器的可解释性。对分类器的可解释性包括: 对预测结果的解释和对内部表征的可视化和分析。

对预测结果的可解释性, 最早是 2016 年, Zhou 等人在[21]中, 建立了一种对分类器分类结果做出可视化热图的方法: 类别激活映射(CAM) [21], 主要是根据模型预测结果计算模型最后一层卷积层后的全局池化层的梯度值, 然后将该值乘到其上一层卷积层的输出上, 形成一个热力图, 通过这种方法可以显示输入图片中各个部分在预测结果中的影响大小。在 CAM 基础上, 相继发展出梯度加权类别激活映射(Grad-CAM) [22]和 GBP (Guided-Backpropagation) [23]等解释性可视化方法。但对效果评价方法一直没有。直到 2020 年, Junsuk Choe 等人设计了一个评估模型弱监督定位效果的标准(CAM 原本用于模型分类结果解释, 也可用于弱监督定位), 在此标准上评估了过去几年弱监督定位方面的工作, 根据这个标准显示, 效果最好的依然是 CAM [24], 这也反映出 CAM 的效果很好。但由于 CAM 采用了全局平均池化层, 因此会一定程度地降低分类结果的准确性[22], 因此综合考虑分类器的准确性和可解释效果, 采用 Grad-CAM 对神经网络输出结果进行解释是相对较好的方法。在 2020 年, Hongyang Jiang 等人在 25 名医生手工标注病灶点的 3228 张 DR 眼底图像上, 采用了基于 Resnet 和 Grad-CAM 的模型, 建立了具有可解释性的眼底病灶识别的神经网络, 达到了很好的识别效果和可解释性热图[25]。

内部表征的可解释性工作: Mu 等人通过给模型内部神经元打上标签, 利用神经元的语义概念给出可解释性[26]。McDuff 等人利用生成模型生成的样本图片来测试分类模型的, 结果发现分类器识别错误的多是用于训练分类模型时, 数据较少类别[27]。由于内部表征的可解释性通常需要设计网络结构, 训练好的模型较少, 而且我们的目标任务属于小数据集上的分类任务, 重新训练这样的网络不太适合。所以没有选择这种可解释性方法。

除此之外, 越复杂的数据需要更深的网络才能取得较好的效果, 而越深的网络需要更多数据才能充分训练模型, 进而取得好的效果。然而糖尿病视网膜病变数据集通常为小数据, 但 DR 图像却十分复杂, 因此需要大型网络来拟合数据, 所以 DR 分类任务面临数据小的问题。

为了利用小数据集训练大型网络, 但小数据难以训练整个大型网络。那能否利用那些花费了大量人力物力训练好的大型网络中的大部分网络, 而用我们目标数据集训练其中的小部分网络参数, 使得训练后的大型网络能够适用于数据不足的任务, 迁移学习就是采用了这个想法。对于小数据集的分类问题, 用迁移学习是很好的解决办法, 很多工作[14][28]都利用了这一技术达到了很好的效果。由于本课题属于小数据集, 因此本文也研究利用迁移学习这项技术。

1.3. 研究贡献

基于以上分析, 本研究采用迁移学习, 通过在小数据集上使用大型预训练模型来训练 DR 分类模型。随后, 利用 Grad-CAM 对模型的分类结果进行可视化解, 从而构建一个具有可解释性的 DR 分类深度学习模型。

2. 方法

2.1. DR 分类模型

2.1.1. 数据来源

数据来源于 Kaggle 公开数据集 Diabetic Retinopathy Detection 公开竞赛数据集, 数据集是从多家医院

收集的, 其中的图像和标签都含有噪声, 图像是很多成像设备拍摄的, 有的图像包含伪影、曝光不够或曝光过度、不聚焦和尺寸不同等缺陷, 而且部分图像无法判断属于哪个类别。数据集中包括 35126 张眼底图像, 其中 33545 张是不影响视力的眼底图像, 其它 1581 张是对视力造成伤害的 DR 眼底图像, 显然数据存在分布不均的情况。

2.1.2. 数据预处理和数据增强

数据下载后, 使用 torch 库中 `transposes` 方法进行处理: 由于数据本身特点(眼睛是圆形的), 因此数据处理采用随机旋转、水平翻转、垂直翻转、平移、随机灰度化和随机颜色变化。并同时把所有图像数据统一更改为统一尺寸并转换为 `tensor` 格式以便输入模型。

除传统方法外, 并采用过采样(即喂入模型中的数据按照各个类别的数据数量基本相同使得网络学习的样本数量各个类别一样, 有些类别样本偏少, 那么该类样本就会反复喂入网络)。按照放回抽样的方式, 从各个类别中均匀抽取样本, 根据样本分布 d 设计取样器权重 w 。

$$d = [25810, 2443, 5292, 873, 708] \quad w = [0.1, 1, 0.5, 3, 4] \quad (2-1)$$

2.2. DR 分类模型搭建

2.2.1. 框架选择

考虑到本研究是针对大量图像数据, 选择 GPU 会节省模型训练时间。考虑到易用性和灵活性, 采用 `pytorch` 框架。运行环境如下:

操作系统: Windows 10; 编程语言: Python 3.8.4; 深度学习框架: Pytorch 1.71。

CPU 型号: IntelCorei3-9100F @ 3.60 GHz; 31 GB 内存。

GPU: GeForce RTX 2080 Ti。

2.2.2. 模型设置

为了使 DR 图片分类任务效果更好, 考虑到糖尿病视网膜病变图像比较复杂, 为了达到更好的效果, 采用相对较深的网络, 因为深层网络被证明能更好地处理复杂的分类任务。由于采用梯度下降法来优化神经网络, 使得深层网络容易出现梯度消失的问题, 处理该问题常用方法有采用长短时记忆(LSTM)、ReLU、Residual neural networks (ResNets)、Batch Normalization (BN)和 DenseNet。ReLU 和 BN 对梯度消失问题帮助有限, 直到 ResNets 的提出, 才使得深层网络模型开始大量使用。在 ResNets 基础上, DenseNet 在更少的模型参数和更短的训练时间达到了更好效果[28]。因此, 本研究选择以 DenseNet 网络为基础进行 DR 分类研究。

模型采用 DenseNet121 预训练的神经网络, 采用微调迁移学习, 即更改网络最后一层(FC 层), 使得输出通道为 5, 训练该 FC 层和最后一个 dense 块中的最后三个 dense 层, 固定其它层的网络参数。本研究采用的更改后的网络如表 1 所示:

模型搭建(Pytorch 代码如下):

```
model = models.densenet121(pretrained=True)
```

其中最后一层全连接层网络, 即 FC 网络(只显示全连接层)如下:

```
model.classifier = nn.Sequential(
    nn.Linear(in_features=1024, out_features=512, bias=True),
    nn.Linear(in_features=512, out_features=256, bias=True),
    nn.Linear(in_features=256, out_features=128, bias=True),
    nn.Linear(in_features=128, out_features=64, bias=True),
```

```

nn.Linear(in_features=64, out_features=32, bias=True),
nn.Linear(in_features=32, out_features=16, bias=True),
nn.Linear(in_features=16, out_features=5, bias=True),
)

```

Table 1. Structure of model**表 1.** 模型结构和参数

层的名称	输出尺寸	参数信息
Convolution	112, 112	7×7 conv, stride 2
Pooling	56, 56	3×3 max pool, stride 2
Dense Block (1)	56, 56	$[1 \times 1 \text{ conv}, 3 \times 3 \text{ conv}] \times 6$
Transition Layer (1)	28, 28	1×1 conv, 2×2 average pool, stride 2
Dense Block (2)	28, 28	$[1 \times 1 \text{ conv}, 3 \times 3 \text{ conv}] \times 12$
Transition Layer (2)	14, 14	1×1 conv, 2×2 average pool, stride 2
Dense Block (3)	14, 14	$[1 \times 1 \text{ conv}, 3 \times 3 \text{ conv}] \times 24$
Transition Layer (3)	7, 7	1×1 conv, 2×2 average pool, stride 2
Dense Block (4)	7, 7	$[1 \times 1 \text{ conv}, 3 \times 3 \text{ conv}] \times 16$
Pooling	1, 1	7×7 global average pool
Classification Layer	1, 1	FC: 1024...5

超参数设置:

学习率: 0.0001, batch size: 256 (设置大些以提高训练速度和训练效果)。

数据集: 将数据集 80%用于训练集, 20%用于测试集, 由于数据量小, 训练集比重设置大些, 提高模型训练效果。

模型: 预训练 DenseNet121。

优化器: Adam; 损失函数: 交叉熵。

主要程序如下:

#设置可训练层

```
for name, param in critic0.features.denseblock4.denselayer14.named_parameters():
```

```
    param.requires_grad=True
```

```
for name, param in critic0.features.denseblock4.denselayer15.named_parameters():
```

```
    param.requires_grad=True
```

```
for name, param in critic0.features.denseblock4.denselayer16.named_parameters():
```

```
    param.requires_grad=True
```

```
critic0 = critic0.to(device)
```

#选择交叉熵损失函数

```
criterion = nn.CrossEntropyLoss()
```

#优化器选择 Adam

```
optim = optim.Adam(critic0.parameters(), lr=LEARNING_RATE, betas=(0., 0.99), eps=1e-8)
```

```
d_scheduler = build_lr_scheduler(optim, -1)
```

```
#采用 tensorboard 可视化训练过程
writer_real = SummaryWriter(f"logs/GAN_MNIST/real")
writer_fake = SummaryWriter(f"logs/GAN_MNIST/fake")
step = 0
```

2.2.3. 模型输出结果

为了验证 DenseNet121 网络的优势, 与常用图片分类网络 Resnet101 和 VGG16 进行比较, 网络都是在相同的图像大数据集 ImageNet 上训练好的网络, 并更改了网络最后一层全连接层, 全连接层的设置相同, 并且都相应选择最后一个包含卷积的网络块设为可训练, 其它浅层网络设置为不可训练, 超参数选择都相同。

采用 tensorboard 可视化模型的损失曲线(DenseNet、VGG 和 ResNet)如图 1~3 所示, 图中深色线为损失平滑后的曲线, 浅蓝色线为实际损失曲线。

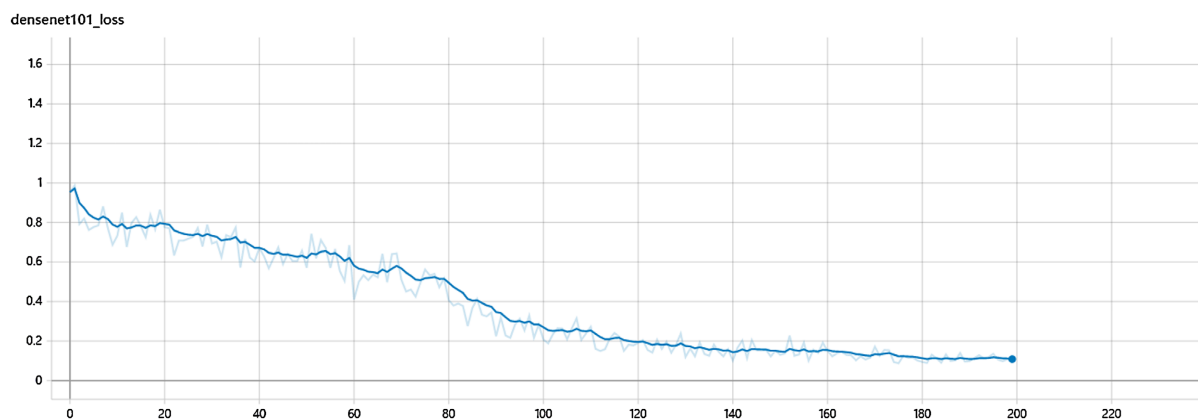


Figure 1. Loss curve of DenseNet121 network training process

图 1. DenseNet121 网络训练过程损失曲线

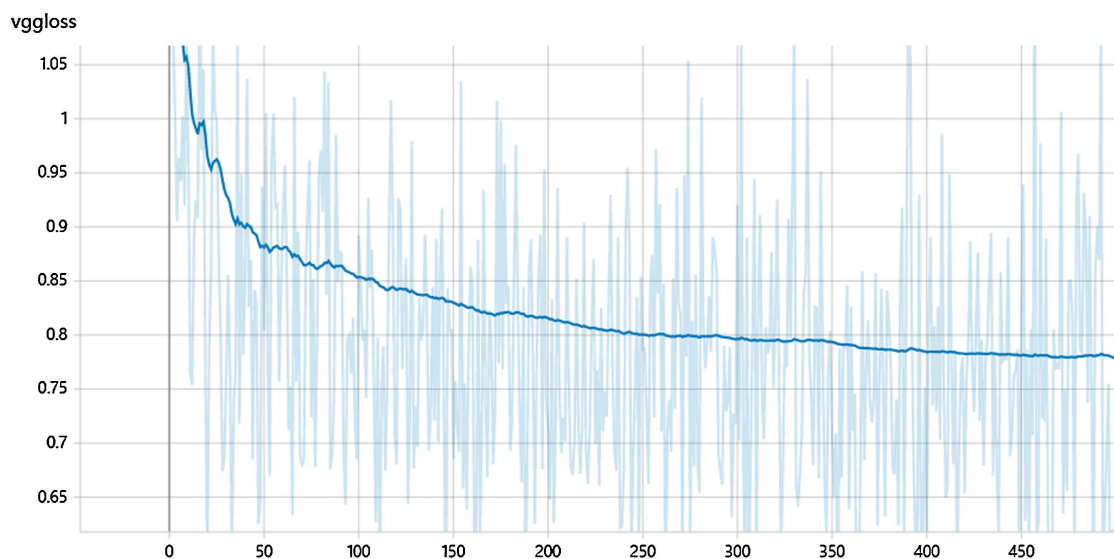


Figure 2. Loss curve of VGG16 network training process

图 2. VGG16 网络训练过程损失曲线

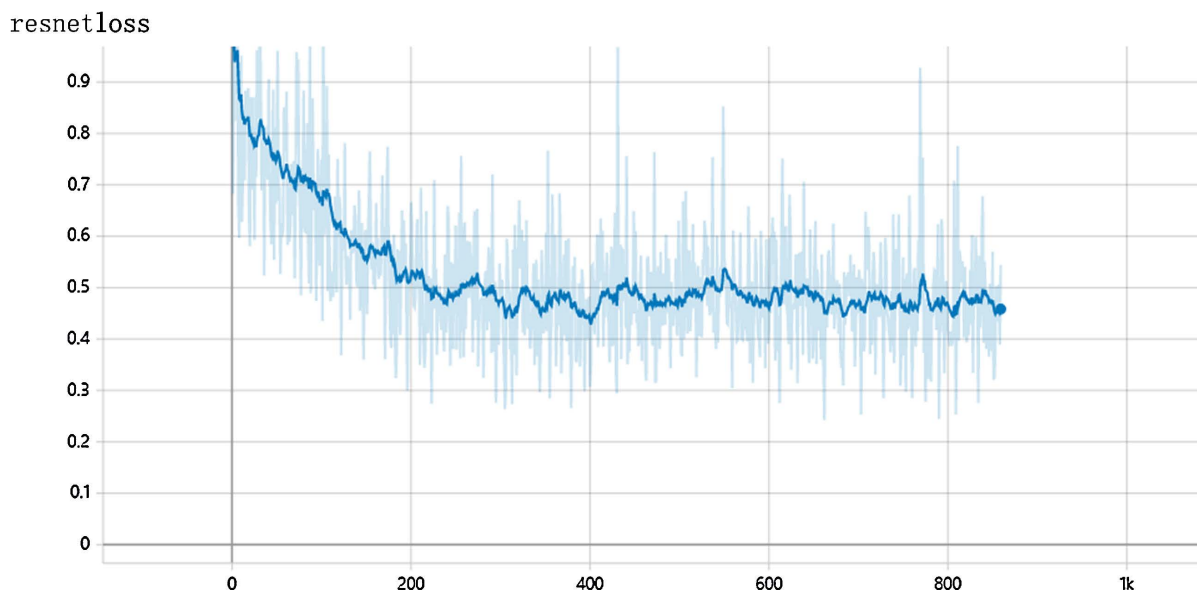


Figure 3. Loss curve of Resnet101 network training process

图 3. Resnet101 网络训练过程损失曲线

由图 1 可见, Densenet121 网络更易经过训练损失达到稳定后, 模型损失为 0.15 左右, 而且损失下降较平稳。

对比图 1~3 可见, Densenet121 网络经过 100 轮训练后收敛, 比 VGG 网络和 Resnet 网络收敛速度更快, 训练过程也更加稳定。

进一步比较基于迁移学习 DenseNet121 网络与 ReseNet101 网络和 VGG16 网络之间的敏感性 (Sensitivity)、特异性 (Specificity) 和准确率 (Accuracy), 计算公式如 2-2, 2-3, 2-4 所示:

$$\text{Sensitivity} = \frac{TP}{TP + TF} \quad (2-2)$$

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (2-3)$$

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (2-4)$$

其中: TP 为真正样本数、 TN 为真负样本数、 FP 为假正样本数、 FN 为假负样本数。

通过表 2 可见, 基于迁移学习的 Densenet121 网络性能更好, 仅特异性相较于 VGG 网络略差, 敏感性和准确率都为三者最高。而且, 模型收敛很快, 100 多个 batch 损失就降到 0.15 左右, 显示了 Densenet 网络在性能上的优越性, 也显示了迁移学习在提高模型训练效果上的明显作用。

Table 2. Comparison of model classification results

表 2. 模型分类结果对比

模型	敏感性	特异性	准确率
Resnet101_based network	29%	89%	76%
VGG16_based network	6%	96%	74%
Densenet121_based network	77%	91.3%	88.3%

2.3. 分类结果解释性的可视化

2020 年, Junsuk Choe 等人在中提出了一种评估弱监督定位的方法, 基于此方法, 评估了最近的几个模型弱监督定位的方法, 包括 CAM、HaS、SPG、ADL 和 CutMix, 发现在 2016 年的 CAM 方法之后提出的其他方法都没有比 CAM 效果更好[24]。这也一定程度上反映出 CAM 寻找模型关注点方面的优越性。但考虑到 CAM 存在的不足(需要改变网络结构以及降低分类准确性)。因此, 本文选用 CAM 的进阶版——梯度加权类别激活映射(Grad-CAM)作为 DR 分类器模型可解释性方法。

具体步骤为, 在模型训练前, 将 Grad-CAM 加入到网络中, 选取要加入的位置以及输出热力图保存的路径, 本研究采用 Densenet 网络, Grad-CAM 加到 DensNet121 的最后一个 Dense Block 的输出端, 即 Dense Block 4 为最后一个 Dense Block, 其输出为(7, 7, 1024)。然后, 计算模型分类结果中概率最大的值, 然后计算损失函数再反向传播, 计算那 1024 个神经元的梯度并求平均, 用此梯度乘上一层的网络输出再求平均(在最后一个轴)得到(7, 7)大小的图, 再上采样就得到了和图片大小相同的热力图。测试时, 通过任选一个样本输入网络, 即可得到分类结果和解释性热图。主要程序如下:

```
#选取样本
sample_x, sample_y = next(iter(train_loader))
critic0 = critic0.to('cpu')
#输出热图保存至文件夹: attention_maps
critic0=medcam.inject(critic0,output_dir=f"attention_maps", lay-er='features.denseblock4.denselayer16',
backend='gcam', save_maps=True)
sample_x = sample_x.cpu()
s = s_x.unsqueeze(0)
prep = critic0(s)
#将图片通道移至最后一个轴上
sample_x1 = s.permute(0,2,3,1)
sample_np = sample_x1.cpu().squeeze(0).numpy()
#标准化图片
a = sample_np - np.min(sample_np)
b = np.max(a)
sample_np_nml = (a)/(b)
#读取热图
interpre_img_dir=f'/home/featurize/work/CWGAN-GP/attention_maps
/features.denseblock4.denselayer16/attention_map_0_0_0.png'
image_fm_PIL = PIL.Image.open(interpre_img_dir)
img_inter = transforms(image_fm_PIL).permute(1,2,0)
#将热图和输入图片进行叠加:
mix_img = img_inter * 0.1 + sample_np_nml*0.8
```

其中选择热图 `img_inter` 占比 0.1, 是因为热图比重太大就会看不出眼底图像的细节信息, 同理如果样本图像 `sample_np_nml` 比重太小也看不清眼底图像了。

如图 4 所示, 图 4(a)为 DR 眼底图片, 图 4(b)为对应热图, 图 4(c)为前两者叠加后的效果图。

从图 4 中可见, 红色区域为 DR 图片中对模型分类结果影响大的部位, 黄色其次。通过热力图可见,

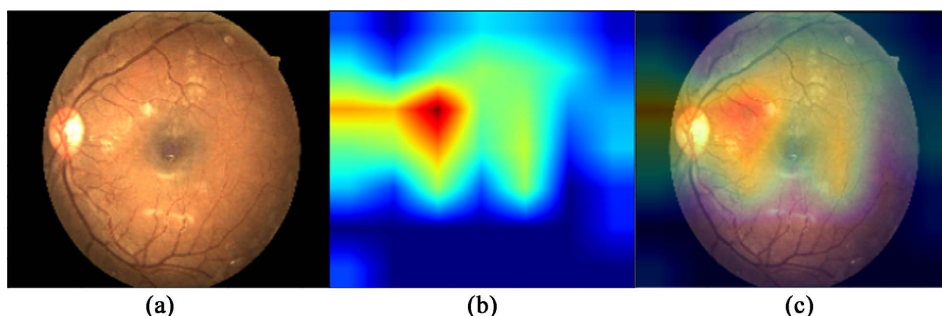


Figure 4. Results with heat map of DR classification

图 4. 糖尿病视网膜病变分类结果热图显示

网络判断是根据图片中的一些特征包括病灶部位和其它信息, 通过这种方式, 我们能够更好地了解训练的模型, 是否是按照人类判断的方法, 同时, 也许会为医疗人员带来一些新的想法, 比如可能人类并没有发现的一些特征, 而神经网络发现了, 然后网络通过热力图告诉我们, 它们关注的点。

热力图显示比较直观, 但也存在不足, 从图 4(c)中可见, 有些病灶部位(如中央凹下方的病灶)模型并未给予足够关注, 有些非病灶部位(视盘左侧)却得到了一定关注, 这与医疗人员通过眼底损失来确诊有一定不同, 但通过这种方法, 我们可以在训练模型时, 得到更多信息, 而不仅仅是看准确率。在医疗影像自动诊断中, 我们更需要多看过程, 在过程正确的基础上再提高准确率。下面进一步分析模型各种判断情况下热力图给出的解释。

如图 5 所示, 图 5(a)为模型判断正确(严重非增殖性 DR)的热力图, 从图中可见模型关注图中央病灶部位, 显然模型判断过程是合理的。图 5(b)为人类也难以判断类别的眼底图像图片(严重非增殖性 DR 判断为正常 DR), 模型选择了一些血管纹理进行判断, 判断过程并不正确。这反映出模型可能会找一些医生不会关注的地方来进行识别。图 5(c)为模型判断错误的图片(严重非增殖性 DR 判断为正常 DR)。从图中可见, 模型并未关注图中央病灶部位, 而是关注于其它部位。这反映出模型出错的情况, 而这种情况是

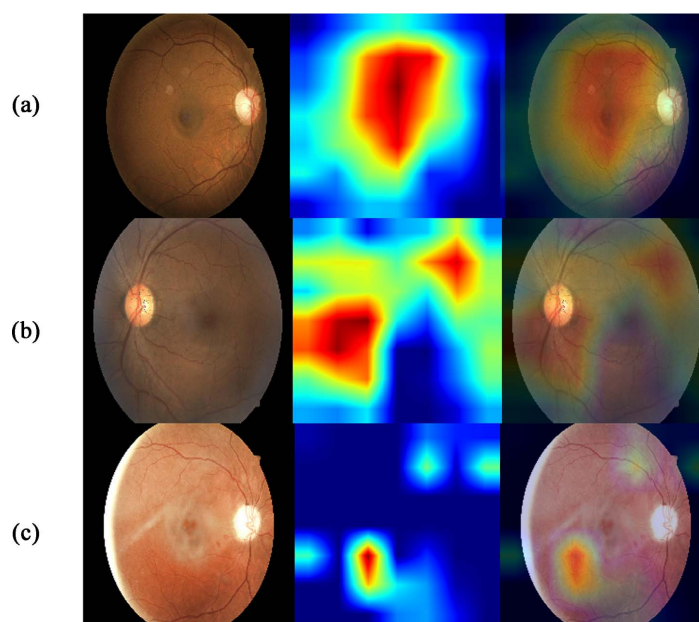


Figure 5. Comparison of heat maps

图 5. 热力图对比

可能存在的, 因为数据不平衡, 导致模型倾向于寻找无病的 DR 特征, 而无病的 DR 眼底图像是不存在病灶特征的, 更多的是血管特征。从图 5(b)和图 5(c)可见, 模型倾向于寻找血管特征, 这反映出数据不平衡带来的问题。从热力图分析中我们可以看出我们的模型, 存在对数据较少类别的 DR 眼底图像的偏见, 在判断错误时更倾向于把更严重的病例样本判断为无病。

根据对热力图的分析发现, 我们的模型还存在不足, 如果单从准确率和敏感性特异性上看, 只能看出模型效果并非十分理想, 敏感性稍差 0.77, 同时准确率也只有 88%, 但除此之外并没有对模型判断过程或者关注点有更直观的信息。但通过热力图, 我们可以明显发现模型出错的原因(数据少造成模型产生偏见), 这给我们如何进一步提高模型, 带来了更直观和重要的信息。也为我们更加直观地显示出模型判断过程是否正确。显然, 通过这种方法让我们更好地了解模型, 也给“黑箱”带进来一缕阳光。

3. 结论

糖尿病视网膜病变(DR)是糖尿病的一种并发症, 如不及时检查和治疗, 可能导致病人视力受损甚至失明, 将给病人、病人家属和社会带来极大的负担。这种疾病是由于病人血糖控制问题所导致的, 一般病人血糖控制都存在一定问题, 因此这种疾病发病率高, 检测次数也就因此非常大。为了缓解眼科医生的工作负担, 同时也为了降低就医费用。利用深度学习来完成这项检测分类任务十分必要。

然而, 深度学习在医学领域的应用还面临一些困难: 医学数据通常是大数据集, 而且数据存在不平衡现象。因此, 本研究利用生成对抗网络、迁移学习、可解释性工具对这项任务进行研究, 主要研究: (1) 研究利用迁移学习搭建生成对抗网络(GAN)生成 DR 眼底图像; (2) 研究利用预训练的生成对抗网络构造条件生成对抗网络, 增加对生成 DR 眼底图像的控制; (3) 研究搭建有可解释性的 DR 病情分类神经网络。

主要创新点: 建立了带可解释性的基于迁移学习的 DensNet121 分类神经网络, 用小数据训练大型分类神经网络的目的, 并且达到了较好分类的效果: 准确率: 88%, 敏感性: 77%, 特异性 91%。并且, 实现了网络对每个样本的预测, 做出了可视化的解释。

本研究还需进一步的改善如下: 对神经网络可解释性, 还需进一步研究, 如何把热图做到更精细化, 使得 Grad-CAM 的解释能力更强, 还需要继续研究。

参考文献

- [1] Biyani, R.S. and Patre, B.M. (2018) Algorithms for Red Lesion Detection in Diabetic Retinopathy: A Review. *Biomedicine & Pharmacotherapy*, **107**, 681-688. <https://doi.org/10.1016/j.biopha.2018.07.175>
- [2] Asiri, N., Hussain, M., Al Adel, F. and Alzaidei, N. (2019) Deep Learning Based Computer-Aided Diagnosis Systems for Diabetic Retinopathy: A Survey. *Artificial Intelligence in Medicine*, **99**, Article 101701. <https://doi.org/10.1016/j.artmed.2019.07.009>
- [3] 张欣鹏. 彩色眼底图像微动脉瘤检测方法研究[D]: [博士学位论文]. 天津: 天津工业大学, 2017.
- [4] 李博. 基于彩色眼底图像的新生血管自动检测方法[D]: [硕士学位论文]. 沈阳: 东北大学, 2015.
- [5] 丁蓬莉. 基于深度学习的糖尿病性视网膜图像分析算法研究[D]: [硕士学位论文]. 北京: 北京交通大学, 2017.
- [6] 罗院生. 从视网膜图像到糖尿病视网膜病变诊断[D]: [硕士学位论文]. 成都: 电子科技大学, 2017.
- [7] 马文俊. 基于机器学习的糖尿病视网膜病变分级研究[D]: [硕士学位论文]. 哈尔滨: 哈尔滨工程大学, 2017.
- [8] Li, Z., Keel, S., Liu, C., He, Y., Meng, W., Scheetz, J., et al. (2018) An Automated Grading System for Detection of Vision-Threatening Referable Diabetic Retinopathy on the Basis of Color Fundus Photographs. *Diabetes Care*, **41**, 2509-2516. <https://doi.org/10.2337/dc18-0147>
- [9] 李郭. 基于深度学习的糖尿病视网膜眼底图像病变分析算法研究[D]: [硕士学位论文]. 上海: 上海交通大学, 2018.
- [10] 方全. 融合双眼特征的糖网病图像识别方法[D]: [硕士学位论文]. 武汉: 中南民族大学, 2019.

- [11] 商逸凡. 基于数据增强的糖尿病性视网膜病变检测[D]: [硕士学位论文]. 北京: 北京交通大学, 2019.
- [12] 李芳君. 基于机器学习的医学数据分类算法研究[D]: [硕士学位论文]. 济南: 山东大学, 2020.
- [13] Dekhil, O., Naglah, A., Shaban, M., Ghazal, M., Taher, F. and Elbaz, A. (2019) Deep Learning Based Method for Computer Aided Diagnosis of Diabetic Retinopathy. 2019 *IEEE International Conference on Imaging Systems and Techniques (IST)*, Abu Dhabi, 9-10 December 2019, 1-4. <https://doi.org/10.1109/ist48021.2019.9010333>
- [14] Wu, Y. and Hu, Z. (2019) Recognition of Diabetic Retinopathy Based on Transfer Learning. 2019 *IEEE 4th International Conference on Cloud Computing and Big Data Analysis (ICCCBDA)*, Chengdu, 12-15 April 2019, 398-401. <https://doi.org/10.1109/icccbda.2019.8725801>
- [15] Shankar, K., Perumal, E., Elhoseny, M. and Thanh Nguyen, P. (2021) An IoT-Cloud Based Intelligent Computer-Aided Diagnosis of Diabetic Retinopathy Stage Classification Using Deep Learning Approach. *Computers, Materials & Continua*, **66**, 1665-1680. <https://doi.org/10.32604/cmc.2020.013251>
- [16] Thanh Nguyen, P., Dang Bich Huynh, V., Dang Vo, K., Thanh Phan, P., Yang, E. and Prasad Joshi, G. (2021) An Optimal Deep Learning Based Computer-Aided Diagnosis System for Diabetic Retinopathy. *Computers, Materials & Continua*, **66**, 2815-2830. <https://doi.org/10.32604/cmc.2021.012315>
- [17] Wang, X., Dai, L., Li, S., Kong, H., Sheng, B. and Wu, Q. (2020) Automatic Grading System for Diabetic Retinopathy Diagnosis Using Deep Learning Artificial Intelligence Software. *Current Eye Research*, **45**, 1550-1555. <https://doi.org/10.1080/02713683.2020.1764975>
- [18] Wang, J., Bai, Y. and Xia, B. (2020) Simultaneous Diagnosis of Severity and Features of Diabetic Retinopathy in Fundus Photography Using Deep Learning. *IEEE Journal of Biomedical and Health Informatics*, **24**, 3397-3407. <https://doi.org/10.1109/jbhi.2020.3012547>
- [19] Li, T., Gao, Y., Wang, K., Guo, S., Liu, H. and Kang, H. (2019) Diagnostic Assessment of Deep Learning Algorithms for Diabetic Retinopathy Screening. *Information Sciences*, **501**, 511-522. <https://doi.org/10.1016/j.ins.2019.06.011>
- [20] de la Torre, J., Valls, A. and Puig, D. (2020) A Deep Learning Interpretable Classifier for Diabetic Retinopathy Disease Grading. *Neurocomputing*, **396**, 465-476. <https://doi.org/10.1016/j.neucom.2018.07.102>
- [21] Zhou, B., Khosla, A., Lapedriza, A., Oliva, A. and Torralba, A. (2016) Learning Deep Features for Discriminative Localization. 2016 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, 27-30 June 2016, 2921-2929. <https://doi.org/10.1109/cvpr.2016.319>
- [22] Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D. and Batra, D. (2017) Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. 2017 *IEEE International Conference on Computer Vision (ICCV)*, Venice, 22-29 October 2017, 618-626. <https://doi.org/10.1109/iccv.2017.74>
- [23] Simonyan, K., Vedaldi, A. and Zisserman, A. (2013) Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps. arXiv: 1312.6034. <https://ui.adsabs.harvard.edu/abs/2013arXiv1312.6034S>
- [24] Choe, J., Oh, S.J., Lee, S., Chun, S., Akata, Z. and Shim, H. (2020) Evaluating Weakly Supervised Object Localization Methods Right. 2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, 13-19 June 2020, 3130-3139. <https://doi.org/10.1109/cvpr42600.2020.00320>
- [25] Jiang, H., Xu, J., Shi, R., Yang, K., Zhang, D., Gao, M., et al. (2020) A Multi-Label Deep Learning Model with Interpretable Grad-CAM for Diabetic Retinopathy Classification. 2020 *42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, Montreal, 20-24 July 2020, 1560-1563. <https://doi.org/10.1109/embc44109.2020.9175884>
- [26] Mu, J. and Andreas, J. (2020) Compositional Explanations of Neurons. arXiv: 2006.14032. <https://ui.adsabs.harvard.edu/abs/2020arXiv200614032M>
- [27] McDuff, D., Ma, S., Song, Y., et al. (2019) Characterizing Bias in Classifiers Using Generative Models. arXiv: 1906.11891. <https://ui.adsabs.harvard.edu/abs/2019arXiv190611891M>
- [28] 柴源. 基于多重迁移的糖尿病性视网膜病变检测[D]: [硕士学位论文]. 青海: 青海大学, 2020.