

一种基于双流 CNN 模型的糖尿病视网膜病变检测新方法

邓叙言

浙江大学数学科学学院, 杭州 310058, 浙江, 中国

摘要: 视力损害的主要成因——尤其是糖尿病视网膜病变 (DR) 与年龄相关性黄斑变性 (AMD) ——给临床诊断与治疗带来了严峻挑战。早期发现与及时干预有助于避免患者出现严重后果。本研究提出一种利用双流卷积神经网络 (CNN) 模型检测眼部疾病的新方法: 第一条支路处理经预处理的眼底图像, 第二条支路则在空间频率域中分析经高通滤波的眼底图像。为评估模型性能, 本文采用 APTOS 2019 数据集——该数据集最初为 2019 年亚太远程眼科学会失明检测竞赛而整理, 并在 Kaggle 上公开。本文方法的准确率达 0.986, 有望成为 DR 检测的早期筛查工具。

关键词: 糖尿病视网膜病变; 傅里叶变换; 眼底摄影; 松散配对训练; 双流 CNN

A Novel Approach for Detecting Diabetic Retinopathy Using Two-Stream CNNs Model

XuYan Deng

School of Mathematical Sciences, Zhejiang University, Hangzhou 310058, Zhejiang, China

Abstract: Major causes of visual impairment, particularly diabetic retinopathy (DR) and age-related macular degeneration (AMD), have posed significant challenges for clinical diagnosis and treatment. Early detection and prompt intervention can help prevent severe consequences for patients. The study presents a novel approach for detecting eye diseases using a two-stream convolutional neural network (CNN) model. The first stream processes pre-processed fundus images, while the second stream analyzes high-pass filtered fundus images in the spatial frequency domain. To assess the model's performance, we use the APTOS 2019 dataset, which was originally compiled for the Asia Pacific Tele-Ophthalmology Society 2019 Blindness Detection competition and is publicly available on Kaggle. Our method shows promise as an early screening tool for DR detection with an accuracy of 0.986.

Keywords: Diabetic retinopathy; Fourier transform; Fundus photography; Loose pairing training; Two-stream CNN

一、引言

视网膜病变日益成为全球性的健康问题, 需要投入大量关注与资源。在这些疾病中, 视网膜病变是全球范围内导致严重视力损害与失明的首要原因之一[1]。据预测, 到 2040 年年龄相关性黄斑变性 (AMD) 将影响近

3 亿人；与此同时，导致个体视力丧失的主要原因之一是糖尿病视网膜病变（DR），它已被认定为一种全球性流行病。这凸显出针对这些致残性眼病采取预防措施、开展早期诊断与探索创新疗法的迫切需要[2-3]。

医学与生物学正日益依赖自动化系统进行分析与诊断[4]。例如，视网膜图像的计算机辅助诊断（CAD）可帮助医务人员识别疾病、启动早期治疗，并减少图像判读中的不一致[5]。与人工检查相比，自动化分析还具有成本效益高、客观、可靠以及减少对熟练专家依赖等诸多优势[6-7]。在深度学习（DL）技术兴起之前，CAD 系统主要用于视网膜诊断流程中的图像复原与增强等环节；部分 CAD 技术则力图复现医生识别视网膜疾病的步骤，包括图像分割、特征提取以及用机器学习对结果分类[8]。

深度学习技术——尤其是卷积神经网络（CNN）以及近年来的 Transformer 模型——在视网膜疾病自动诊断中取得了显著成效。众多工作聚焦于识别 AMD、DR 与青光眼等常见视网膜疾病。例如，有研究采用两个预训练 CNN（VGG16 与自定义 CNN），通过分析病灶图块的概率来诊断 DR[9-12]；类似地，有系统把 Inception-v3、ResNet152 与 Inception-ResNet-v2 三种 CNN 模型相结合，用于在眼底图像中检测 DR；此外，还有基于深度学习的诊断工具可对多种常见视网膜疾病进行筛查。这些方法仍存在一些不足，例如处理过程中的图像信息损失，或计算需求高、复杂度大。

APTOS 数据集已被广泛用于各类眼病分类方法。文献强调在有针对性的困难样本子集上训练，同时削弱大量易分负样本在训练中可能对检测器造成的淹没效应；文献提出一种神经元内在学习框架，用以识别 CNN 特征空间中的干扰因素，并采用一种新的干扰感知损失函数，在特征空间中把原始图像与其干扰项清晰区分开来；文献利用对比学习进行特征嵌入以应对类别不平衡，取代了传统的交叉熵损失；文献则通过迭代训练过程获得互相关的单位矩阵。

尽管已有显著进展，如何在保持计算效率的同时有效融合多频谱特征信息，仍是一个关键的研究空白。具体而言，本文所针对的主要科学问题是：缺乏一种稳健的模型架构，能够（1）保留原始眼底图像的完整上下文；（2）用专门的通道分离并突出关键的高频病理特征（如微动脉瘤与出血点）；（3）取得优于计算开销沉重的集成方法的性能。要解决这一问题，必须有一种高效的策略来融合上述不同的特征集合。

为填补该空白，本文提出一种基于双流 CNN 模型的 DR 检测新方法。本研究的核心创新与主要贡献在于把所提架构设计与预处理 / 融合方法整合为一体：第一条支路处理经标准预处理（含数据增强）的眼底图像，以捕捉全局的低频特征（如视盘、主要血管等）；第二条支路则处理空间频率域中经高通滤波（由傅里叶变换导出）的眼底图像。这一步至关重要，因为它分离并突出了与 DR 病理相关的关键高频细节，促使网络聚焦于细小而具临床意义的病灶。在全局平均池化（GAP）之后的特征融合策略中，两条支路的输出在 GAP 层之后拼接；这一融合点的选择确保每条支路在合并之前已提取出稳健、精炼的高层语义特征，从而为分类提供更丰富的最终表征。这一双流方法在保持图像完整性的同时，主动引导模型关注特定的病理特征，在 APTOS 2019 数据集上取得了 0.986 的优异准确率。

二、方法

（一）数据集

本研究采用 APTOS 2019 BD 数据集训练并评估所提模型的性能。该数据集包含 3662 个样本，其中正常图像 1805 幅、DR 图像 1857 幅，采集自印度的大规模人群。采集工作由印度 Aravind 眼科医院组织，眼底照片摄于不同条件与不同时期。一组受过训练的医学专家依据国际临床糖尿病视网膜病变严重程度分级标准（ICDRSS）对样本进行了细致复核与标注。据此，APTOS 2019 BD 数据集被划分为五类：增殖性 DR、轻度 DR、中度 DR、

重度 DR 与无 DR。

（二）总体流程

模型的总体流程如图 1 所示。该检测系统使用两个输入通道。如流程图所示，首先施加数据增强以缓解数据量不足与过拟合问题；随后图像被分入两个通道。

第一通道处理原始图像，第二通道则提供额外信息以提升分类准确率。通过在第二通道中把增强后的图像变换到空间频率域，可以分离并去除低频成分。由于这些低频成分编码的是总体信息（如视神经乳头、黄斑与主要血管的形状），去除它们有助于把注意力集中到 DR 的特异性指征上。最后，两个通道被送入分类模块——即双流 CNN 模型——完成疾病检测。

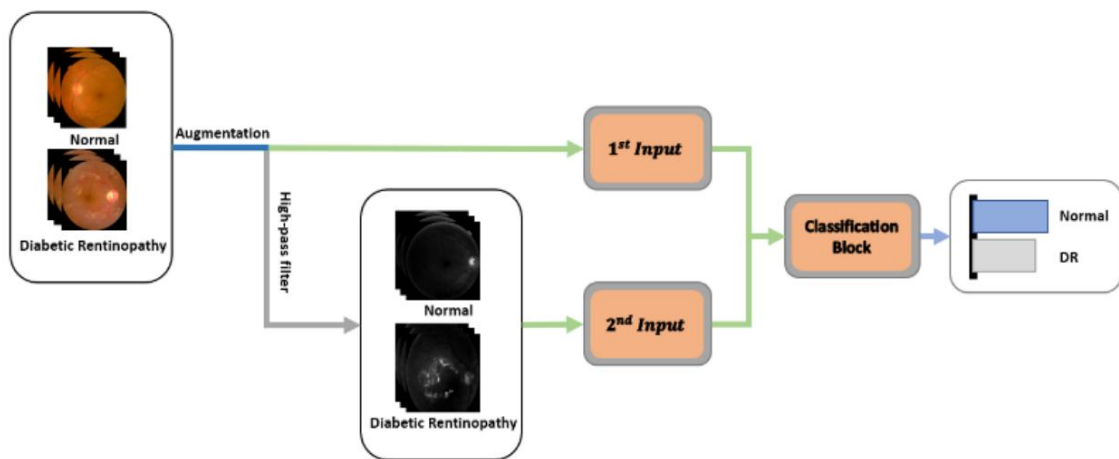


Figure 1 The Workflow of the Proposed Method

图 1 所提方法的总体流程

（三）数据增强

为保证双流模型的训练样本数量足够，本文在训练过程中采用数据增强，对每幅训练图像在输入双流 CNN 模型之前至少施加一种增强。这些增强借助 Albumentations 库实现，包括光学畸变、网格畸变、分段仿射变换、水平与垂直翻转、随机旋转、平移、缩放、红绿蓝（RGB）通道值偏移，以及亮度与对比度的随机调整。经增强后，数据集扩充至 29 296 个样本，其中正常图像 14 440 幅、病变图像 14 856 幅。该增强过程显著提升了数据多样性，降低了过拟合，并改善了模型在 DR 检测与分类上的泛化能力。

（四）高通滤波

在训练与验证阶段，本文使用了原始图像的修改版本，这些照片经过了缩放与裁剪。APTOS 2019 数据集在疾病分期与若干图像元特征（包括亮度、缩放级别、裁剪方式与分辨率）之间存在虚假关联。为防止 CNN 对这些元特征过拟合，并尽量削弱图像内容与此类属性之间的相关性，本文施加了大量增强。随后，对第一条支路的输入施加基于傅里叶变换的高通滤波，以生成第二条支路的输入。

1. 傅里叶变换

傅里叶变换能够通过分析强度（灰度）在空间距离上的变化来分解图像，而非分析随时间变化的信号。该过程把图像的时域转换为空间域。就声音而言，频率表示正弦波重复的（倒数）规律性；类似地，在图像中，空间频率描述强度值波动的（倒数）规律性。空间频率由下式确定：

$$F(u, v) = \iint_{-\infty}^{\infty} I(x, y) \cdot e^{-j2\pi(ux+vy)} dx dy \quad (1)$$

式中， $I(x, y)$ 为空间域中某像素处的灰度值， $F(u, v)$ 为空间频率域中的取值， (u, v) 为空间频率坐标。

2. 频谱

高空间频率对应于强度在短距离内快速变化的图像特征；相反，强度在较长距离上缓慢变化的特征则与低空间频率相关。频谱可通过施加傅里叶变换并对结果取自然对数得到：

$$\text{Spectrum}(u, v) = \log(1 + |F(u, v)|^2) \quad (2)$$

3. 空间频率域分析

空间频率由到离散傅里叶变换（DFT）中心的距离确定。空间频率为零对应图像的平均亮度；距中心两个像素的点表示图像中完成两个正弦周期的正弦分量。因此，图像中的距离 x_i 与其对应的空间频率 f 可表示为：

$$x_i = \Delta x / f \quad (3)$$

式中， Δx 为图像尺寸（以像素计）。

4. 高通滤波

如前所述，在眼底摄影中，低频对应频谱中心附近的区域，高频则与频谱外围区域相关。为施加高通滤波，需定义阈值 D_0 ，表示频谱中心到频谱空间中某点的距离；对频谱中的每一点，若其到中心的距离小于 D_0 ，则将其去除。二维空间中的距离采用欧氏距离计算，点 $P=(p_1, p_2, \dots, p_n)$ 与 $Q=(q_1, q_2, \dots, q_n)$ 之间的欧氏距离公式为：

$$d(P, Q) = \sqrt{\sum_{i=1}^n (p_i - q_i)^2} \quad (4)$$

式中， p_i 与 q_i 分别为点 P 与 Q 的各分量， n 为空间维数。

图 2 与图 3 给出了对眼底图像施加高通滤波的示例。图 2（a）为滤波前图像的频谱，图 2（b）为高通滤波后的频谱；图 3（a）为处理前的眼底图像，图 3（b）为高通滤波后的图像，重要细节更加清晰。经高通滤波后，与 DR 相关的重要信息均得到突出。

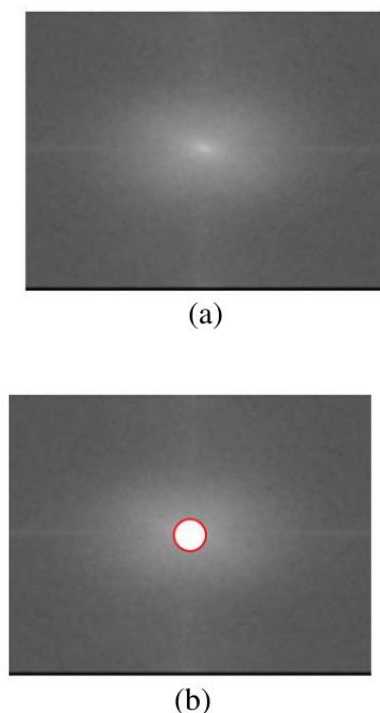


Figure 2 Images of (a) Original and (b) after High-Pass Filtering in Spatial Frequency Domain

图 2 空间频率域中（a）原始图像与（b）高通滤波后的图像

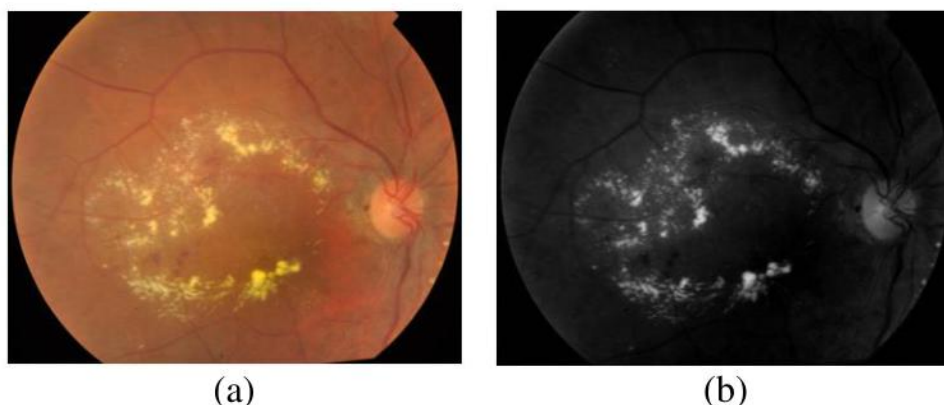


Figure 3 Illustration of (a) before and (b) after High-Pass Filtering

图 3 高通滤波 (a) 之前与 (b) 之后的效果示例

(五) 双流网络架构

本研究设计了一种双流 CNN，用以处理两类不同的特征输入。该模型的架构如图 4 所示，以 ResNet 作为主干网络。这一设计思路与此前在视频人体行为识别领域中探索过的双流网络方法有相似之处。

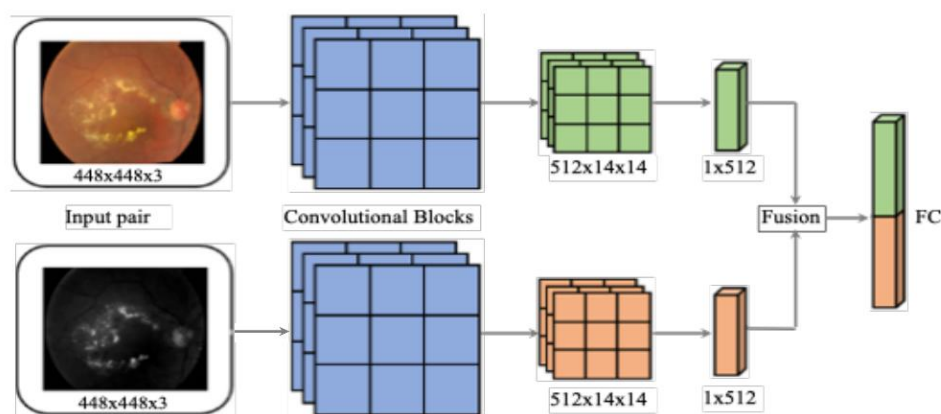


Figure 4 Two-Stream CNNs Model

图 4 双流 CNN 模型

第一条 ResNet 支路专注于从眼底图像中提取特征，第二条支路则提供补充信息以支持第一通道。双流 CNN 的两条分支均基于 ResNet-34。虽然也可采用其他 CNN 架构，但选择 ResNet-34 有若干理由：参数量更少、训练所需数据更少，且已在多种眼底图像识别任务中被证明有效。

单模态 CNN 逐层生成连续的二维特征图以从图像中提取信息。这些特征图尽管尺寸不断缩小，仍保留了原始输入图像的部分空间信息。因此，本文选择在 GAP 层之后实施空间不变的融合：GAP 层把每个特征图约简为单一标量值，从而有效消除空间信息。

在第一条支路中，ResNet-34 的最后一个卷积块产生 512 个特征图，记为 $F_f = \{F_{f,1}, \dots, F_{f,512}\}$ 。每个特征图的尺寸为 $m \times m$ ，其中 m 的大小取决于输入图像尺寸；当输入尺寸为 448×448 时， m 等于 14。若要提取特征图 $F_{f,i}$ 中某位置 (x, y) 处的特征值，记作 $F_{f,i}(x, y)$ 。类似地，第二条支路输出的特征图记为 $F_o = \{F_{o,1}, \dots, F_{o,512}\}$ 。

Ff 与 Fo 同时通过 GAP 层，得到两个 512 维特征向量。

融合模块层把来自两条独立支路的高层输出特征结合起来，有效利用各自的信息以提升整体性能。这种融合层架构具有两项关键优势：第一，它无需像后期融合方法那样训练多个分类器，便能捕捉到对原始数据更为丰富的表征；第二，与在初始层进行融合的早期融合方法不同，各支路中的 CNN 层能够从原始特征中提取出更有用、更精炼的信息。如图 4 所示，第一条支路处理眼底图像，其输出张量的每个元素对应这些图像；第二条支路处理高通滤波图像，其张量元素则表示这些滤波后的数据。

融合后的特征张量记为 $f=\text{concat}(Ff, Fo)$ ，兼具两条支路的优势；随后该张量被送入后续的 CNN 层作进一步处理。最后，由 Softmax 层与全连接（FC）层完成分类。该方法通过自适应调节并平衡不同特征集合的重要性，提升了预测准确率。在训练与测试阶段，采用如下交叉熵损失函数：

$$L = -\sum_{i=0}^{C-1} y_i \cdot \log(p_i) \quad (5)$$

式中， $p=[p_0, \dots, p_{C-1}]$ 表示概率分布， p_i 为样本属于第 i 类的概率； $y=[y_0, \dots, y_{C-1}]$ 为类别标签的独热表示， C 为类别数。

四、结论

本文成功提出了一种用于糖尿病视网膜病变（DR）检测的新型双流 CNN 模型。所提架构以 ResNet-34 为主干，别具特色地融合了来自两个通道的信息：原始眼底图像与频率域中的高通滤波图像。高通通道通过突出与疾病相关的细微细节，增强了特征的区分度。实验结果证明了所提模型的有效性：在 APTOS 2019 数据集上取得 0.986 的准确率与 0.986 的 F1 值，凸显其作为早期筛查工具的可靠性。

尽管性能优异，该模型仍面临若干挑战，包括对最优滤波参数选择的依赖，以及目前仅面向二分类（正常与 DR）而非完整的临床严重程度分级。尽管如此，本研究在改进早期与自动化筛查方面显示出显著的临床潜力，在眼科专家资源匮乏的地区尤其如此。未来工作将聚焦于：按 ICDRSS 严重程度等级把模型扩展至多类 DR 分级；探索注意力机制等自适应特征融合策略，以动态平衡空间信息与频率信息；以及研究自动化滤波器优化技术，在训练过程中学习最优的高通滤波特性。上述改进有望增强模型在多样成像条件下的泛化能力，并进一步提升其临床适用性。

参考文献

- [1] 王树国. 机器人关节驱动器高精度控制. 机器人, 2020, 44(02): 1-8.
- [2] 郝吉明. 工业烟气多污染物协同治理. 中国环境科学, 2022, 42(04): 1501-1508.
- [3] Schmidt-Erfurth U, Sadeghipour A, Gerendas B S, et al. Artificial intelligence in retina. Progress in Retinal and Eye Research, 2018, 67: 1-29.
- [4] 肖立业. 超导输电线路工程化关键问题. 电工电能新技术, 2022, 41(02): 1-10.
- [5] 朱合华. 城市地下管网数字孪生建模. 地下空间与工程学报, 2023, 18(02): 369-378.
- [6] 李元昊. 数控机床热误差补偿技术. 中国机械工程, 2022, 33(08): 905-912.
- [7] 陈发虎. 高寒冻土路基稳定控制技术. 冰川冻土, 2020, 44(02): 478-486.
- [8] 吴一戎. 合成孔径雷达成像优化算法. 雷达学报, 2022, 11(02): 201-212.
- [9] 黄维和. 天然气管网动态调度优化. 油气储运, 2024, 41(04): 361-368.
- [10] 叶列平. 再生混凝土高性能化改性技术. 建筑材料学报, 2022, 25(02): 305-314.
- [11] Keenan T D L, Cukras C A, Chew E Y. Age-related macular degeneration: epidemiology and clinical aspects.

Advances in Experimental Medicine and Biology, 2021, 1256: 1-31.

- [12] Zheng Y, He M, Congdon N. The worldwide epidemic of diabetic retinopathy. Indian Journal of Ophthalmology, 2012, 60(5): 428.