

分类号: TP391.41

学校代码: 10363

密 级: 公开

学 号: 2220120103



安徽工程大学

Anhui Polytechnic University

硕士学位论文

题 目 基于卷积神经网络的视网膜病变

诊断系统研究

论 文 作 者

解勇正

指 导 教 师

王海(教授)

学 科 (专 业)

机械工程

研 究 方 向

机器视觉

论文提交日期: 2025 年 5 月 12 日

分类号：TP391.41

单位代码：10363

密 级：公开

学号：2220120103

基于卷积神经网络的视网膜病变诊断系统研究

RESEARCH ON A RETINAL LESION DIAGNOSIS SYSTEM
BASED ON CONVOLUTIONAL NEURAL NETWORKS

学生姓名： 解勇正

指导教师： 王海(教授)

专 业： 机械工程

研究方向： 机器视觉

论文答辩日期： 2025 年 5 月 12 日

安徽工程大学学位论文原创性声明

本人郑重声明：我恪守学术道德，崇尚严谨学风。所呈交的学位论文，是本人在导师的指导下，独立进行研究工作所取得的成果。除文中已明确注明和引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写过的作品及成果的内容。论文为本人亲自撰写，我对所写的内容负责，并完全意识到本声明的法律后果由本人承担。

学位论文作者签名：

日期： 年 月 日

安徽工程大学学位论文版权使用授权书

学位论文作者完全了解学校有关保留、使用学位论文的规定，同意学校保留并向国家有关部门或机构送交论文的复印件和电子版，允许论文被查阅或借阅。本人授权安徽工程大学可以将本学位论文的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存和汇编本学位论文。

保密 ☐，在 _____ 年解密后适用本版权书。

本学位论文属于

不保密 ☒。

学位论文作者签名：

指导教师签名：

日期： 2025 年 5 月 12 日

日期： 2025 年 5 月 12 日

基于卷积神经网络的视网膜病变诊断系统研究

摘 要

本文提出一种基于卷积神经网络(CNN)的视网膜病变诊断系统,能够对视网膜病变图像进行高效分割与分类,CNN 作为深度学习领域的强大工具,在图像识别和分类任务中展现出卓越性能,近年随着医疗健康领域的数据量增长,利用 CNN 进行医学影像分析成为研究热点之一,特别是在眼科疾病的早期检测方面,如糖尿病性视网膜病变等,如果不及时发现并治疗可能会导致严重的视力损害甚至失明,因此开发一套基于 CNN 的视网膜病变智能诊断系统显得尤为重要和迫切,本研究旨在探索如何有效地将 CNN 应用于眼底摄影图像上,以实现视网膜特征的自动提取与分析,进而准确判断是否存在病变以及病变程度,针对不同类型的眼底摄影数据构建大规模数据集,再设计适用于视网膜关键区域检测及分割的高效 CNN 模型架构,通过实验验证所提方法在实际应用中的有效性和鲁棒性。

本文的主要研究内容有:

(1)利用数据增强、非线性降维、数据采样、数据变换等方法处理视网膜的扫描 OCT 图像,使得数据集中的图像信息更加丰富,特征更加明显,构建视网膜病变图像的数据集,并应用自适应直方图均衡化与直方图匹配结合的图像处理方法,在提高图像质量的同时,保留图像的细节特征,并在 RMSE 和 PSNR 指标上表现出显著提升, RMSE 降低百分率为 $18.03\% \pm 2.15\%$, PSNR 提升 $0.46\% \pm 0.16\%$, 图像质量较高,噪声和失真较低。

(2)提出 SF-Unet 网络,通过改进 MLP 结构,提出 Shifting-MLP、OS MLP 以及 DOS-MLP 结构,进而提出 SF-Unet 网络以提升图像分割精度和泛化能力,在留一法交叉验证(LOO-CV)中,模型预测结果与真实标签之间误差平方和的平均值为 0.0320,模型的预测结果与真实值接近,在视网膜病变图像分割实验中,SF-Unet 在所有评估指标上均优于传统网络,相较于其他网络,DICE 提升 2.34%~4.68%,PA 提升 5.81%~7.34%,MPA 提升 4.52%~11.51%,MIOU 提升 4.50%~17.12%,SF-Unet 在评估指标上表现出色,通过对不同卷积神经网络在视网膜病变图像分割过程和结果的比较,验证 SF-Unet 在处理复杂场景的优越性。

(3)为解决现有传统分类网络的局限性,研究提出 Sets-ResNet Block 结构,并通过选择合适的优化算法,构建 Sets-ResNet 分类网络,用于视网膜病变图像的分类,采用 ResNet 重复层策略,实现简单可扩展方式,在同样视网膜病变分割任务中,Sets-ResNet50 相较于传统网络,在测试 AUC 上平均提升 18.92%,在验证 AUC 上平均提升 24.71%,表明 Sets-ResNet50 在这些数据集上的表现明显优于传统分类网络模型,相较于其他传统网络具有明显提升。

(4)为进一步提高分类性能,研究对 Sets-ResNet 进行二级并联结构的改进,提出 SetsA-ResNet 分类网络,采用两次并行多分支网络,更多的通道数让每一层能够捕捉到更丰富、更细致的特征,从而在图像分类、目标检测等任务中表现更佳,在处理复杂视觉任务时能够表现出更优异的性能,通过超参数优化和交叉验证,折数为 4 时达到准确度最高点 0.9152, SetsA-ResNet50 在多个类别上的分类效果最佳,进一步改进的 SetsA-ResNet50 在

Sets-ResNet50 的基础上, 测试 AUC 又提升 5.68%, 验证 AUC 提升 8.24%, 数据表明, 通过进一步的优化和调整, SetsA-ResNet50 在性能上有更显著的提升, 显示出其在分类任务中的优越性。

关键词: 卷积神经网络, 图像分类, 图像分割, 视网膜病变诊断系统, 图像识别

RESEARCH ON A RETINAL LESION DIAGNOSIS SYSTEM BASED ON CONVOLUTIONAL NEURAL NETWORKS

ABSTRACT

This paper presents a retinal lesion diagnosis system that utilizes Convolutional Neural Networks (CNN) for efficient segmentation and classification of retinal images. CNNs are powerful tools in deep learning, known for their exceptional performance in image recognition and classification. With the growing volume of healthcare data, CNNs have become a focal point for medical image analysis, particularly for early detection of ocular diseases like diabetic retinopathy, which can cause severe vision impairment or blindness if not addressed promptly. Thus, developing a CNN-based intelligent diagnosis system for retinal lesions is both important and urgent.

The study aims to effectively apply CNNs to fundus photographs for automatic extraction and analysis of retinal features, enabling accurate assessment of lesion presence and severity. A large-scale dataset of various fundus photography types is constructed, and an efficient CNN model architecture is

designed for detecting and segmenting critical retinal areas. The proposed method's effectiveness and robustness are validated through experiments.

The main research contents of this paper are: Image Processing Improvements: The study employs techniques such as data augmentation, nonlinear dimensionality reduction, data sampling, and transformation to enhance scanning OCT images of the retina. This enriches the dataset and highlights features. A new dataset of retinal lesion images is created using adaptive histogram equalization and histogram matching, which improves image quality and preserves detail. Results show that RMSE decreased by $18.03\% \pm 2.15\%$, and PSNR increased by $0.46\% \pm 0.16\%$, resulting in high-quality images with minimal noise and distortion.

Introduction of SF-Unet Network: The paper introduces the SF-Unet network, which improves image segmentation accuracy and generalization by enhancing the MLP structure with Shifting-MLP, OS MLP, and DOS-MLP. In leave-one-out cross-validation (LOO-CV), the average squared error between model predictions and true labels is 0.0320, indicating close alignment with true values. In segmentation experiments, SF-Unet outperforms traditional networks across all metrics, with improvements in DICE (2.34% to 4.68%), PA (5.81% to 7.34%), MPA (4.52% to 11.51%), and MIOU (4.50% to 17.12%). Comparisons confirm SF-Unet's superiority in managing complex scenarios.

Sets-ResNet Block Structure: To overcome limitations in traditional classification networks, the study proposes the Sets-ResNet Block structure and constructs the Sets-ResNet network for retinal disease image classification using

suitable optimization algorithms. Employing a repeated layer strategy from ResNet, this approach is both simple and scalable. In segmentation tasks, Sets-ResNet50 shows an average improvement of 18.92% in testing AUC and 24.71% in validation AUC, demonstrating significant superiority over conventional classification models.

Enhanced SetsA-ResNet Classification Network: To further boost classification performance, the research develops the SetsA-ResNet network, adding a secondary parallel structure. This dual parallel multi-branch network increases channels, allowing richer feature capture, which enhances performance in image classification and object detection tasks. Through hyperparameter optimization and cross-validation, SetsA-ResNet achieves a maximum accuracy of 0.9152 with four folds. It outperforms multiple categories. Furthermore, the improved SetsA-ResNet50 shows a 5.68% increase in testing AUC and an 8.24% increase in validation AUC compared to Sets-ResNet50, highlighting its enhanced classification capabilities.

Overall, this study demonstrates the efficacy of CNNs in retinal lesion diagnosis, showcasing several advancements in image processing, network architecture, and classification performance.

Xie Yongzheng (Mechanical Engineering)

Supervised by Prof. Wang Hai

KEY WORDS: convolutional neural networks, image classification, retinopathy diagnostic system, image recognition

目录

摘 要	I
ABSTRACT	IV
目录	VII
插图清单	XI
表格清单	XIII
第 1 章 绪论	1
1.1 研究背景	1
1.2 国内外研究现状	1
1.2.1 视网膜病变诊断研究现状	1
1.2.2 深度学习模型研究现状	2
1.3 主要研究内容	4
第 2 章 视网膜病变诊断数据集预处理	6
2.1 数据集来源及种类说明	6
2.2 视网膜病变数据集分布	6
2.3 聚类采样	8
2.4 视网膜病变数据集筛选	9
2.5 数据增强	11
2.5.1 随机翻转	11
2.5.2 对比度增强	11
2.6 主成分分析(PCA).....	11
2.7 t-SNE 非线性降维	13
2.8 数据采样	14
2.8.1 分层采样	14
2.8.2 过采样	16
2.9 数据变换	17
2.9.1 对数变换(Log Transformation)	18
2.9.2 平方根变换(Square Root Transformation).....	18

2.9.3 Box-Cox 变换	18
2.9.4 数据变换评价（偏度与峰度）	20
2.9.5 数据变换总结	21
2.10 自适应直方图均衡化	22
2.10.1 累积分布函数(CDF)与直方图(Histogram)	22
2.10.2 自适应直方图均衡化评价指标	22
2.11 直方图匹配	24
2.12 直方图处理后图片评价指标	24
2.12.1 均方根误差(RMSE).....	24
2.12.2 峰值信噪比(PSNR)	25
2.13 直方图均衡化与直方图匹配实验结果	25
第 3 章 SF-Unet 图像分割.....	27
3.1 U-Net 与 CSU-Net 分割	27
3.2 改进 Unet 分割网络：Shifting-Unet	27
3.2.1 总体介绍	27
3.2.2 多层感知机(MLP)	28
3.2.3 激活函数	28
3.2.4 Shifting-MLP 结构.....	29
3.2.5 Orthogonal Shifting MLP 结构.....	32
3.2.6 特征激励块(Excitation-Block)	32
3.2.7 DOS-MLP 结构.....	32
3.2.8 DOSD 和 DOSD-Block 类	33
3.2.9 权重初始化	35
3.2.10 实验平台及超参数配置	36
3.2.11 网络训练	38
3.2.12 参数更新	39
3.2.13 优化算法	40
3.3 留一法交叉验证	42
3.4 SF-Unet 分割评价标准.....	43

3.5 SF-Unet 分割实验结果.....	44
第 4 章 Sets-Resnet 网络	46
4.1 Sets-Resnet 介绍	46
4.1.1 Bottleneck(BTNK)残差块结构	46
4.1.2 优化算法	48
4.1.3 Sets-Resnet50 网络结构配置	48
4.1.4 核心改进机制	49
4.2 Sets-Resnet50 网络分类实验结果	50
第 5 章 SetsA-Resnet 网络	51
5.1 SetsA-Resnet 网络介绍	51
5.2 STAGE0 阶段处理.....	52
5.3 组的划分	52
5.4 SetsA-Resnet Block 结构	52
5.4.1 SetsA-Block.....	54
5.4.2 参数计算	55
5.4.3 可变形卷积操作(Deformable Convolution,DCN)	55
5.4.4 分割与 GAP	55
5.4.5 Dense 低维映射和权重生成	56
5.4.6 SoftmaxA.....	56
5.4.7 加权求和	57
5.5 超参数搜索与参数优化	58
5.5.1 网格搜索	58
5.5.2 随机搜索	58
5.6 SetsA-Resnet 网络分类评估指标	59
5.6.1 混淆矩阵	59
5.6.2 准确率(Accuracy)	60
5.6.3 精确率(Precision).....	61
5.6.4 召回率(Recall)	62
5.6.5 F1 值(F1-score)	63

5.7 K 折交叉验证	63
5.8 SetsA-Resnet 分类网络实验结果及分析	64
总结与展望	68
总结	68
展望	68
参考文献	70
攻读学位期间发表的学术论文情况	74
致谢	75

插图清单

图 2-1 视网膜病变数据集类别分布	6
图 2-2 视网膜病变数据集性别分布	7
图 2-3 视网膜病变数据集箱线图	7
图 2-4 疾病与性别的关系（百分比）	8
图 2-5 疾病与 OS/OD 的关系（百分比）	9
图 2-6 性别与年龄的关系	9
图 2-7 疾病与年龄的关系	9
图 2-8 无效图像	10
图 2-9 正常图片	10
图 2-10 随机翻转	11
图 2-11 对比度增强	11
图 2-12 PCA 降维后的特征图	12
图 2-13 PCA 与 t-SNE 降维结果图	14
图 2-14 视网膜病变数据集类别分布图	15
图 2-15 视网膜病变数据集特征分布	16
图 2-16 数据集类别分布柱状图	16
图 2-17 t-SNE 经过 SMOTE 特征分布	16
图 2-18 对数变换特征分布直方图	18
图 2-19 平方根变换特征分布直方图	18
图 2-20 Box-Cox 变换特征分布直方图	19
图 2-21 原始数据与标准化后数据散点图	19
图 2-22 Quantile-Quantile Plot 概率图	21
图 2-23 自适应直方图匹配处理前后对比	22
图 2-24 直方图与累积分布函数	22
图 2-25 直方图匹配	24
图 2-26 直方图匹配后的直方图与累积分布函数	25
图 3-1 MLP 结构图	30

图 3-2 S-MLP 模块处理特征图示意	31
图 3-3 OS-MLP 结构图	32
图 3-4 对角依赖关系	32
图 3-5 DOS-MLP 结构	33
图 3-6 DOSD 模块结构	34
图 3-7 模型整体结构	34
图 3-8 不同权重初始值的比较	36
图 3-9 动量法	40
图 3-10 RMSprop	41
图 3-11 Adam	42
图 3-12 留一法交叉验证 (LOO-CV) 散点图	43
图 3-13 不同分割模型的训练过程	43
图 3-14 不同分割网络测试结果	43
图 4-1 分类网络训练过程	46
图 4-2 残差块结构	46
图 4-3 残差块的其他不同结构	47
图 4-4 Sets-Resnet50 分类网络训练和验证过程	50
图 5-1 Sets-Resnet 网络结构	51
图 5-2 SetsA-Resnet Block 结构	52
图 5-3 SetsA-Block 结构	54
图 5-4 准确率随阈值变化曲线	60
图 5-5 最佳阈值下的 ROC 曲线	61
图 5-6 真实标签与预测标签效果展示	62
图 5-7 召回率对比	62
图 5-8 Precision-Recall 曲线	63
图 5-9 K 折交叉验证准确率与 AUC	64
图 5-10 不同分割模型的测试过程	65
图 5-11 SetsA-Resnet50 分类混淆矩阵	66
图 5-12 OCT 图像在不同网络下的分类效果	67

表格清单

表 2-1 分层采样前后数据量对比	15
表 2-2 偏度与峰度数据表	20
表 2-3 部分图像 Best Score	23
表 2-4 RMSE 和 PSNR	26
表 3-1 Dice 和 Iou 评分	27
表 3-2 实验平台	36
表 3-3 参数与超参数配置	37
表 3-4 不同分割网络实验结果对比	44
表 4-1 网络结构配置	48
表 5-1 初始参数与超参数配置	58
表 5-2 最终参数与超参数具体数值	59
表 5-3 混淆矩阵参数	60
表 5-4 OCT 图像在改进 ResNet50 网络上的分类结果	66

第1章 绪论

1.1 研究背景

随着全球人口老龄化加剧、生活方式转变及环境变迁，慢性病与退行性疾病患病率持续攀升。在此背景下，早期精准诊断成为疾病防控的核心突破口——医学影像技术凭借其非侵入性、高分辨率特性，已逐步成为病灶检测的"黄金标准"。其中，视网膜影像因其可直观反映全身血管与神经病变的特异性，在糖尿病视网膜病变等致盲性眼病筛查中具有不可替代的价值。

(1)早期识别技术瓶颈：超过 60%的视网膜病变在视力受损前无明显症状（如糖尿病黄斑水肿），传统肉眼阅片易漏诊亚临床期病变；

(2)医疗资源结构性失衡：我国眼科医师与患者比例达 1:3000，基层医疗机构误诊率高达 35%^[1]。

1.2 国内外研究现状

1.2.1 视网膜病变诊断研究现状

视网膜病变是导致全球范围内视力丧失的重要原因，如果这些病变能够在早期被准确诊断，将大大提升患者的治疗成功率，然而传统诊断方法主要依赖于眼科医生主观判断，其效率和准确性受限于医生的经验水平和临床资源的分布不均，近年来随着医学影像技术发展，光学相干断层扫描(Optical Coherence Tomography,OCT)成为诊断视网膜病变的重要手段，其能提供高分辨率的视网膜层次结构图像，由于 OCT 图像数据量巨大，且其判读需要深厚的专业知识，人工解读的效率和准确性都难以满足实际需求，因此基于深度学习技术，开发高效、准确的自动诊断系统，成为该领域研究的热点。

随着医疗技术的不断进步，基于深度学习的自动化诊断系统在医疗领域的应用越来越广泛，为临床医生提供更准确高效的辅助诊断手段，视网膜病变作为一种严重影响视力的眼科疾病，其早期发现与精准诊断尤为重要，OCT 技术因其非侵入性、高分辨率特点，已成为眼科领域的重要成像手段^[2]，在此基础上，结合深度学习技术，可以帮助医生快速识别病变区域，改进传统依赖人工诊断的方式，提高诊断的效率和准确性。

视网膜病变的诊断涵盖多种不同疾病类型，传统诊断通常依赖于专业医生的主观判

断,通过肉眼分析视网膜图像来检测病变区域,这种方法既耗时又可能受到人为因素干扰,为克服这些问题,许多研究人员开发基于视网膜病变图像的自动诊断方法,可以捕捉视网膜组织的微观结构特征,为准确识别病变区域提供丰富信息,例如针对糖尿病视网膜病变,早期研究采用经典的图像处理方法,如边缘检测、特征提取和支持向量机(SVM)分类等^[3],然而传统方法在处理复杂数据、识别细微特征时存在局限性,难以满足实际临床的高精度需求。

1.2.2 深度学习模型研究现状

近年来,深度学习技术在人工智能领域蓬勃发展,以其强大的特征提取和学习能力,广泛应用于图像处理、语音识别和自然语言处理等领域,在医疗图像分析中,深度学习尤其显示出显著的优势,传统医疗图像分析依赖于医生经验,通过手工提取特征并进行分类或判断,容易出现误判,相较之下,基于深度学习的方法能够自动从数据中学习特征并进行分类,大幅提高诊断效率和准确性,为医学领域的自动化发展提供全新的思路。

视网膜是眼球内部的一层薄膜,分布着复杂的微型血管网络,其结构变化通常与多种眼科疾病和系统性疾病密切相关,糖尿病视网膜病变、白内障、高血压和动脉硬化等疾病都会导致视网膜结构(如血管宽度、角度和长度等)的异常^[4],对视网膜图像的分析研究是诊断疾病的重要手段,然而通过肉眼对视网膜图像进行分析和分类不仅依赖于研究者的经验,还容易出现误诊漏诊现象,基于深度学习技术对视网膜图像进行自动化分析成为研究热点,该方法能够实现疾病的高效诊断,通过图像处理和模式识别技术,实现对眼底疾病的快速筛查与分类^[5]。

卷积神经网络(Convolutional Neural Networks,CNN)等深度学习技术的快速发展,为视网膜病变的自动化诊断提供新的思路,在视网膜图像分析领域,深度学习技术已被成功应用于病变程度分类任务,主要集中于优化网络结构和提升分类精度,例如美国 Sudha 等研究人员采用经典 ResNet、DenseNet 等网络进行 OCT 图像分类,并在公共数据集上取得较高的分类准确率^[10]。吴欣等人通过集成学习方法,增强模型对少样本场景的适应能力^[11],许多学者更多关注于结合实际临床需求优化算法,山东中医药大学智能与信息工程学院的孙石磊,李明提出基于 U-Net 的分割模型,用于标注 OCT 图像中的病变区域^[12]。广西师范大学的陆辉辉研究团队还针对不同病变类型进行特征提取与分类,但是面对多个类别的大规模数据集^[13]。Inamullah 等人发表在国际眼科杂志上的对于深度学

习的研究显示其团队在糖尿病视网膜病变诊断中的研究进展,通过 AI 技术分析视网膜图像特征,提高糖尿病视网膜病变的检测准确性和效率,减少人工干预并为早期筛查提供支持,还强调神经网络架构的创新和数据隐私保护的联邦学习方法^[14],为未来的医学影像分析开辟了新路径。Arias-Serrano 等人发表在 AI in Medicine and Healthcare 机构论文研究展示他们利用深度学习设计自动化的糖尿病视网膜图像分类系统^[15],通过在 AlexNet 模型中引入批归一化层,构建一个改进型卷积神经网络(BNnet),结合迁移学习技术,将从大规模数据集中预训练得到的模型迁移到视网膜数据集上,并通过微调获得更好的分类性能,此外还利用全连接层组成的分类器(FCnet),在测试集上分类准确率较高。重庆大学的张思杰团队,在少样本的血管分割中,通过提取血管网络的形态特征来辅助诊断^[16]。云南大学的李琼团队通过限制对比度自适应直方图均衡化和中值滤波等预处理技术,提高视网膜图像的质量^[17]。在特征提取过程中,结合高层卷积层和低层卷积层的特征,可以充分保留局部信息和全局信息,从而提升分割精度;杨培伟团队对于彩色眼底图像的处理与分析是医学图像处理中备受关注的领域之一,这类图像能够反映出眼底结构及其潜在病变区域(如微动脉瘤、眼底出血和渗出物等),为医生的诊断提供重要依据^[18],针对糖尿病视网膜病变的分类问题,提出结合深层监督网络和深度残差网络的深层监督残差网络方法,通过剪切重采样和代价敏感学习等方法,有效解决数据集类别不平衡问题,此外在网络的中间层引入类别监督信号,不仅增强分类性能,还通过集成学习方法提高多尺度分类效果,该方法在糖尿病视网膜病变分类任务中取得较高的准确率和鲁棒性。巴基斯坦管理与技术大学计算机科学系的 Farooq M S, Arooj A 团队在视网膜图像分析的研究中,利用 Caffe 进行灵活的模型搭建框架,针对视网膜图像的特性,引入数据扩增、降噪和迁移学习等关键技术^[19],能够快速实现和验证模型,技术结合为深度学习在医疗图像领域的成功应用提供了有力保障,然而这些方法在实际应用中仍存在模型过于复杂、计算成本高以及对细微病变识别效果有限的问题,并且存在数据集在类别间数据分布不均衡问题,在边界上处理分割效果均存在不足。

深度学习模型具有多层神经网络的自适应学习能力,可以自动提取 OCT 图像中的特征,并在大规模数据集上实现高效训练,基于 CNN 的模型在视网膜病变图像的分割、分类和预测中表现优异,U-Net 模型被广泛用于图像分割,ResNet 等深度网络架构通过改进梯度消失问题,在复杂的分类任务中表现出色,多尺度特征提取技术也被引入到图像分析中,以便更好地捕捉不同分辨率下的图像细节,这些研究证明深度学习技术在提

升诊断效率和准确性方面的潜力。

目前，基于深度学习的视网膜病变诊断系统已经逐渐从实验室研究向实际临床应用转化，除模型架构的优化，研究者还着眼于多模态数据的融合分析，例如将 OCT 图像与眼底彩照、患者病史等数据结合，进一步提高诊断的全面性和精度。一些研究团队开发了能够适应不同分辨率和多种设备生成 OCT 图像的通用化模型，以提高模型的可移植性^[20]，未来随着数据集规模的扩展、模型性能的提升以及计算资源的增强，基于卷积神经网络的视网膜病变诊断系统有望成为眼科诊断的重要辅助工具，在疾病筛查和治疗效果评估中发挥更大作用。

深度学习技术的快速发展为视网膜图像分析带来新的机遇，通过模型的不断优化和方法创新，深度学习必将在医疗图像分析领域发挥更大作用，为疾病的早期诊断和治疗提供更加精准的支持，但是医疗数据的隐私性和可获取性限制大规模数据集的构建，深度学习模型对计算资源的需求较高，可能限制其在资源有限地区的应用，未来的研究可以从以下几个方面进行探索：一是开发更轻量化的模型，提高其在嵌入式设备适用性；二是结合多模态数据（如 OCTA 图像和患者临床信息），提升诊断全面性和准确性。

1.3 主要研究内容

将卷积神经网络应用于眼底图像，实现对视网膜分析，准确判断是否存在病变，为了克服模型过于复杂、计算成本高以及对细微病变识别效果有限的问题，同时实现在图像在类别间数据分布均衡，在边界上处理分割效果较好，本文首先建立视网膜病变诊断数据集，然后对数据集中的图像进行研究，首先进行改进 MLP 结构，构建 SF-Unet 网络实现对图像的分割与提取，再进行 SF-Unet 中的 Shifting-MLP 改进，扩展 SF-Unet-S, SF-Unet-M, SF-Unet-L 网络，通过 Sets-ResNet 与 SetsA-ResNet 实现对视网膜病变图片的分类并结合医院诊断结果实现评测。

章节安排如下：

第 1 章：首先介绍视网膜病变诊断系统研究的可行性及实用性，同时介绍国内外对于视网膜病变的研究方法和成果，再详细介绍卷积神经网络在医学图像分析上的成果。

第 2 章：对数据来源及病变种类进行说明，对模型所要处理的数据集进行构建，通过数据采样、数据变换以及直方图等方法进行处理，同时也进行实验结果的量化评估。

第 3 章：首先介绍 OCT 数据集图片提取使用的 Unet 和 CS-Unet 模型，并基于这两

种卷积神经网络构建视网膜病变数据集分割的网络 Shifting-Unet（简称为 SF-Unet），通过对比实验和交叉验证得出不同分割网络模型对 OCT 图像提取的影响，体现 SF-Unet 对视网膜病变图像分割提取的优越性。

第 4 章：首先介绍分类网络训练过程，再提出 Sets-ResNet Block 结构，选择合适的优化算法，进而提出 Sets-ResNet 分类网络，得到视网膜病变分类模型，结果表明与基础网络相比参数量减少，计算速率加快，验证准确率较高，其分类准确度达到 0.70。

第 5 章：对 Sets-ResNet 的组进行进一步划分形成二级并联结构，提出 SetsA-ResNet 分类网络，同时进行超参数优化与交叉验证，得出分类效果要优于传统的多种网络以及改进后的一级并联结构 Sets-ResNet 分类网络，最后进行网络分类性能评估，得出 OCT 图像在不同网络下的分类效果，其中 SetsA-ResNet50 分类效果表现最好，训练集、验证集、测试集的 AUC 分别为 0.96、0.93、0.92。

总结与展望：视网膜病变诊断系统的智能化与实用化仍需突破数据、效率与解释性瓶颈，未来研究应聚焦多模态深度整合、实时性能优化及临床落地验证，拓展系统至早期筛查与预后预测，结合可穿戴设备构建闭环健康管理生态。

第2章 视网膜病变诊断数据集预处理

2.1 数据集来源及种类说明

视网膜疾病可能损害视网膜的任何部分，每年全世界有数百万人被诊断患有视网膜疾病，并最终导致视力障碍甚至失明，光学相干断层扫描(OCT)是一种高分辨率的非侵入式诊断技术，通过采集视网膜横截面图像信息，帮助医生分析和确定病变视网膜层进而选择适当治疗方案，数据驱动的人工智能技术正在飞速发展，许多国际知名的数据集、样本库都成为开展科学研究不可缺少的重要资源，深度学习模型训练需要海量的数据做支撑，本文采用 OCTA-500 数据集，数据集总大小约 80GB，包含 500 只眼睛的数据近 6 万个 OCT 扫描图像，分辨率为 400×640 像素 BMP 格式图像，每个图像都有与之关联的正常或病变种类，为保证数据充分性，筛选一部分图像来完成前两部分实验，由于后两部分中的分类需要大量数据，使用整个 OCT 数据集，数据集包含内容如下：

糖尿病性视网膜病变(DR)、脉络膜新生血管(CNV)、年龄相关性黄斑变性(AMD)、视网膜静脉阻塞(RVO)、中心性浆液性脉络膜视网膜病变(CSC)、正常(Normal)、其他疾病(Others)，7 个类别共计 6 万张 OCT 的 BMP 格式图片。

2.2 视网膜病变数据集分布

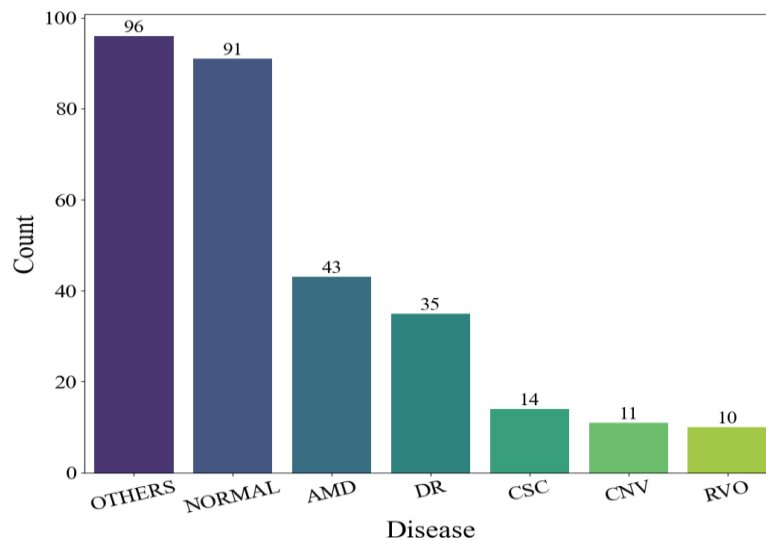


图 2-1 视网膜病变数据集类别分布

图 2-1 便于观察每种疾病流行情况，数据集中共有三百个组，可以看出各个疾病的患病人数排行，其他疾病和正常情况居多，两者占总数 62.33%，剩下五种共占 37.67%，

CSC、CNV、RVO 三者最少，各占总数的 4.67%、3.67%、3.33%。

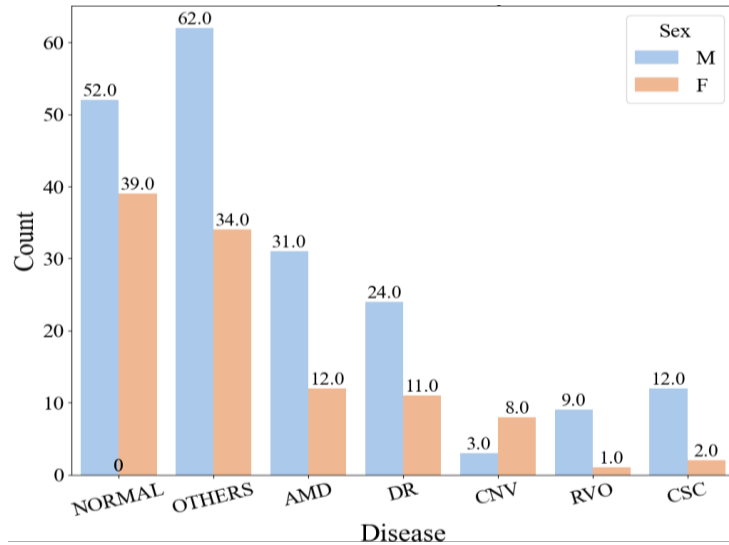


图 2-2 视网膜病变数据集性别分布

从图 2-2 可以看出，男性的平均发病率要高于女性，男性占患病总数的 64.33%，高出女性患者 28.66%，其中在 CNV 类别中，女性患病人数要高于男性，雌激素等性激素对眼部血管有一定影响，雌激素水平下降会增加 CNV 的风险，研究表明，雌激素的保护作用可能在绝经后减弱，从而增加女性患病的概率。

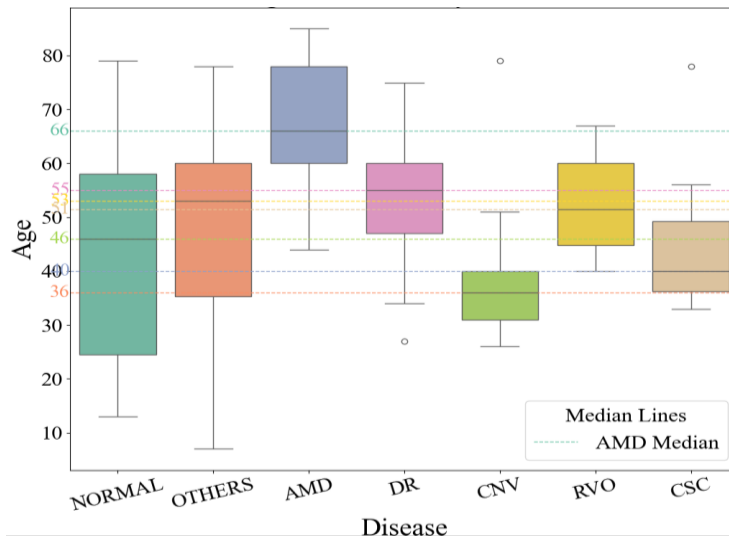


图 2-3 视网膜病变数据集箱线图

图 2-3 的箱体图上下边界分别表示第一四分位数(Q1,25%)和第三四分位数(Q3,75%)，即数据的中间 50%；箱体中间的线表示中位数(Q2,50%)；须的延伸表示数据范围，通常从 Q1 延伸到 Q3，外部点（称为离群点）表示异常值；离群点在箱体外部，表示与其他数据点显著不同年龄值。

观察不同疾病的箱体和中位数，了解到 AMD 患者年龄偏高，年龄中位数显著高于

其他疾病，CNV 患者年龄偏低，年龄与这两种疾病的发生有关，综合分析所有患者年龄大致都分布在二十多岁至七十多岁之间。

2.3 聚类采样

聚类采样适合群组间差异较大的情况，先将数据集分成多个群组，然后从中随机抽取若干个群组中所有数据点，假设数据集分为样本量不同的 M 个群组，每个群组，从这些群组中随机选择 m 个群组($m < M$)，然后在每个群组中选取所有样本进行分析，从群组中选择概率为：

$$P(\text{某个群组被选中}) = \frac{1}{M} \quad (2-1)$$

聚类采样在分组基础上进行，每个群组被选中的概率相同，每个群组中的所有样本被包含在抽样数据中。

为了解患者群体性别、年龄、疾病类型和眼部状况之间关系，为医学研究或临床决策提供数据支持，使用聚类采样对数据进行划分和分析，对不同群体进行特征探索，这里选择群体中个体患疾病相对于的性别、左右眼、年龄三个参数的关系。

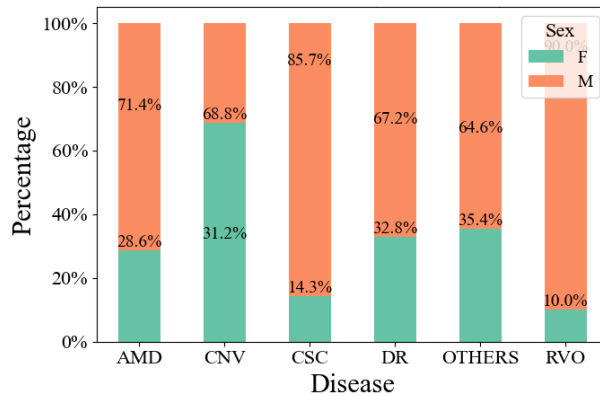


图 2-4 疾病与性别的关系

图 2-4 堆叠柱状图展示不同疾病中男性和女性患者百分比分布，帮助了解性别在疾病中的分布特征，在 CSC 和 RVO 患者中，男性患者人数显著高于女性。

图 2-5 堆叠柱状图展示不同疾病与 OS（左眼）和 OD（右眼）之间关系，与图 2-4 类似，显示每种疾病中患者的眼部状况所占百分比，此图有助于了解不同眼部状况在各个疾病中的分布情况，基本上眼部疾病的左右眼患病几率相同。

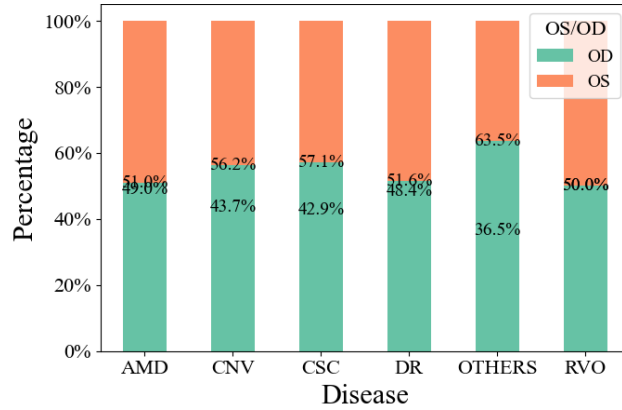


图 2-5 疾病与 OS/OD 的关系（百分比）

图 2-6 展示男性和女性患者年龄分布，还展示某个年龄段有多少患者，直观看到不同性别患者年龄范围和集中程度，对了解疾病在不同年龄和性别群体中的分布有帮助。

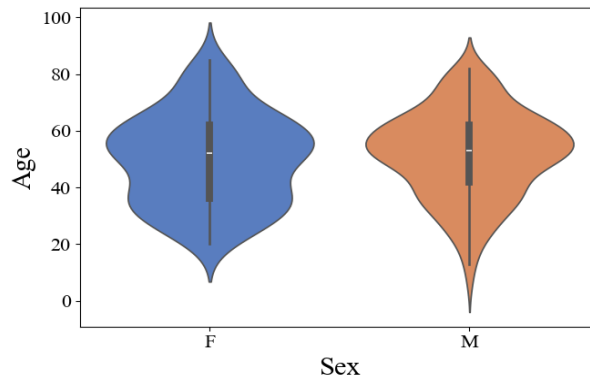


图 2-6 性别与年龄的关系

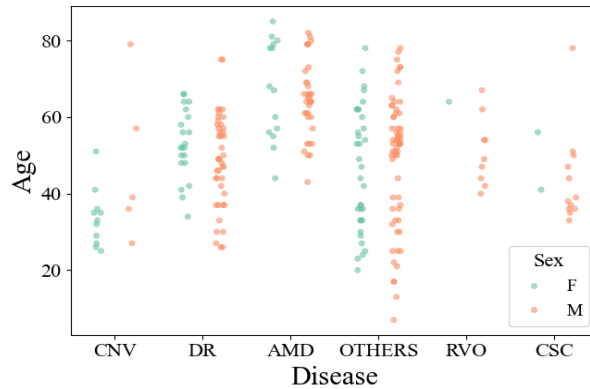


图 2-7 疾病与年龄的关系

图 2-7 展示不同疾病患者的年龄分布，每个点表示一名患者，点的颜色区分患者的性别，通过此图可以观察到不同疾病患者的年龄范围以及各个年龄段内的患者分布情况，帮助了解各疾病在年龄上的特征。

2.4 视网膜病变数据集筛选

通过仔细分析组合数据库以及结合现实情况，OCT 数据集包含很多噪声以及其他造

成干扰的图像，对后续关键位置分割与分类存在影响，对实验有意义仅为清晰图片，所以进行 OCT 筛选，例如图像中具有明显不属于图中的线条等干扰项；图像模糊不清；图像显示的与实验所需无关；以及残缺不全的损坏图像，如图 2-8 所示。

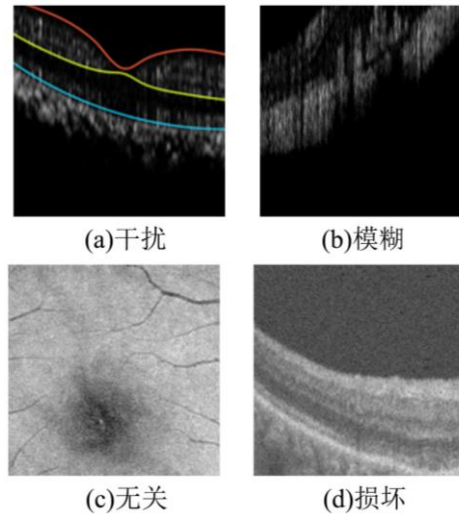


图 2-8 无效图像

筛选后的 OCT 影像数据存在对比度较低、边界不清等问题，通过清理、转换和标准化数据，突出数据对网络特征提取有效性，抑制无用信息，先对图像进行预处理操作，提高数据质量与模型训练效率，确保算法有效性和准确性，良好的数据预处理还可以帮助发掘数据中潜在模式，从而增强模型泛化能力，本文对图像进行不同空间域的变化，最大化扩充数据集，丰富特征信息。

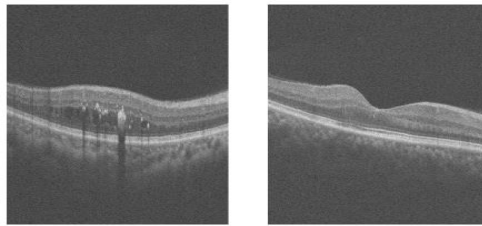


图 2-9 正常图片

数据集中总共有 6 万张 400×640 像素 BMP 格式图片，实验处理的数据集非常大，直接处理全量数据会非常耗时间和资源，同时数据集也存在干扰、无关的图片，这里采用数据采样操作来选择一部分图片来进行实验，采样通过减少数据量来降低计算成本，同时能避免过拟合，提高模型泛化能力，最终在数据集库中筛选 6000 张图像来完成实验^[1]，按照训练集占 60%，验证集占 20%，测试集占 20%的比例，其中训练集有 3600 张，验证集 1200 张，测试集 1200 张。

2.5 数据增强

2.5.1 随机翻转

随机将图像沿水平方向翻转,常用于图像分类任务中增强模型泛化能力,使用 `Resize` 调整到 256×256 , 再进行随机旋转操作, 增强模型对不同角度图像适应能力, 设置图像在 $-30 \sim 30$ 度之间随机旋转。

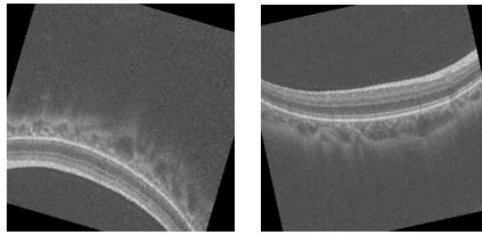


图 2-10 随机翻转

2.5.2 对比度增强

通过对图像进行亮度、对比度、饱和度和色调的调整, 随机改变图像视觉特性, 调整亮度幅度为 $\pm 50\%$, 对比度幅度为 $\pm 50\%$, 饱和度幅度为 $\pm 50\%$, 色调幅度为 ± 0.1 , 对比度增强改善图像对比度和视觉效果, 模拟现实场景中光照和色彩变化, 不仅提升模型对复杂场景的适应能力, 还缓解过拟合, 增强模型泛化性能。

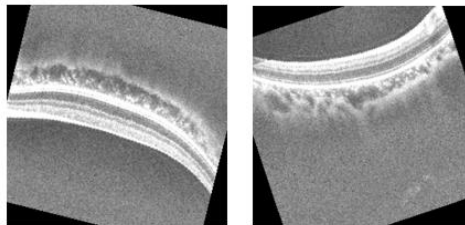


图 2-11 对比度增强

随机选取图片进行随机翻转和对比度调整操作, 为增加鲁棒性, 将处理后的图片和原始数据集保存在一起, 得到一个新的数据集, 大小为 8382 张 `BMP` 格式图片, 通过数据增强手段对训练集进行扩充, 使得训练集在某些特征值区间的样本数量显著增加, 这有助于模型更好地学习到数据的多样性。

2.6 主成分分析(PCA)

数据降维可以减少数据复杂性, 降低数据维度, 提升计算效率, 去除图像噪声和冗余特征, 提高模型性能和泛化能力。

PCA 是广泛使用的线性数据降维技术，保持尽可能多的原始数据变异，通过找到一组新的正交轴（称为主成分），使投影到这些轴上的数据方差最大，具体操作步骤如下：

(1)将数据矩阵 X （大小为 $n \times p$ ，其中 n 是样本数， p 是特征数）中心化，减去每个特征的均值向量 μ ：

$$X_{centered} = X - \mu \quad (2-2)$$

(2)计算中心化数据的协方差矩阵：

$$COV(X) = \frac{1}{n-1} X_{centered}^T X_{centered} \quad (2-3)$$

式中： $X_{centered}^T X_{centered}$ 是一个 $p \times p$ 的矩阵。

(3)对协方差矩阵 $COV(X)$ 进行特征值分解：

$$COV(X) v_i = \lambda_i v_i \quad (2-4)$$

式中： v_i 是特征向量（主成分方向）； λ_i 是对应特征值，是沿着 v_i 方向的方差。

(4)将原始数据投影到前 k 个主成分上：

$$X_{PCA} = X_{centered} V_k \quad (2-5)$$

式中： V_k 是由前 k 个特征向量组成的矩阵。

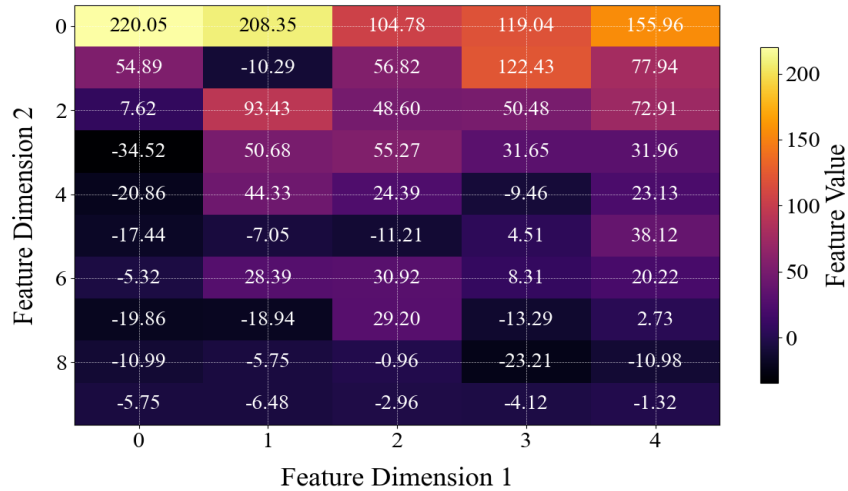


图 2-12 PCA 降维后的特征图

图 2-12 横坐标和纵坐标表示两个主成分数值，颜色表示主成分值在空间中变化，这种可视化能够帮助了解不同样本在新特征空间中分布情况，通过颜色映射来反映每个点在 PCA 降维后主成分值，颜色越亮数字越大代表大的主成分值，通过 PCA 降维后的 50 维样本，清晰展示各维度特征值的分布情况，图中特征值的范围从 -34.52 到 208.35，高值区域如第一行第二列的 208.35 和第一行第四列的 119.04 显著突出，表明这些维度在

降维后的特征空间中具有较高权重，是数据主要特征，低值区域如第三行第一列的-34.52和第四行第四列的-9.46 则相对较弱，反映次要特征或噪声，通过高对比度的颜色映射，为理解数据结构和特征重要性提供全面而直观的支持。

2.7 t-SNE 非线性降维

t-SNE(t-Distributed Stochastic Neighbor Embedding)是非线性数据降维算法，主要用于高维数据映射到二维空间，以便可视化数据结构，针对 64 维图像数据集，采用 t-SNE 将高维特征降维到 2D 空间，揭示高维数据中非线性结构，不依赖数据分布假设，能适应多种复杂数据类型，具体操作步骤如下：

(1)对于每一对数据点 i 和 j ，计算相似性 p_{ij} ，可视为点 j 作为点 i 相邻概率：

$$p_{ji} = \frac{\exp\left(\frac{-\|x_i - x_j\|^2}{2\sigma_i^2}\right)}{\sum_{k \neq i} \exp\left(\frac{-\|x_i - x_k\|^2}{2\sigma_i^2}\right)} \quad (2-6)$$

式中： σ_i 是通过精确寻找使得 p_{ji} 的熵满足预设值。

(2)在低维空间中，为保持高维空间中结构，定义 q_{ij} 来度量点 y_i 和 y_j 相似性：

$$q_{ij} = \frac{(1 + \|y_i - y_j\|^2)^{-1}}{\sum_{k \neq i} (1 + \|y_i - y_k\|^2)^{-1}} \quad (2-7)$$

(3)通过最小化高维和低维空间中相似性分布之间差异，优化目标函数，找到能够保持数据局部结构的低维嵌入，优化过程通过梯度下降法来实现：

$$C = \sum_i \sum_j p_{ij} \log \frac{p_{ij}}{q_{ij}} \quad (2-8)$$

(4)优化过程通过梯度下降不断调整低维嵌入 y_i 来减少目标函数 C ，直到收敛条件。

图 2-13 中左右侧图分别显示 PCA 降维和 t-SNE 后的 2D 可视化，颜色表示不同类别，两幅图横纵坐标都代表降维分量，表示数据点在二维空间中相对位置，两个分量没有直接意义，反映数据在低维空间中分布，图中各个点表示不同类别数据样本，这些点在二维空间中的相对距离反映原始数据在高维空间中的相似度。

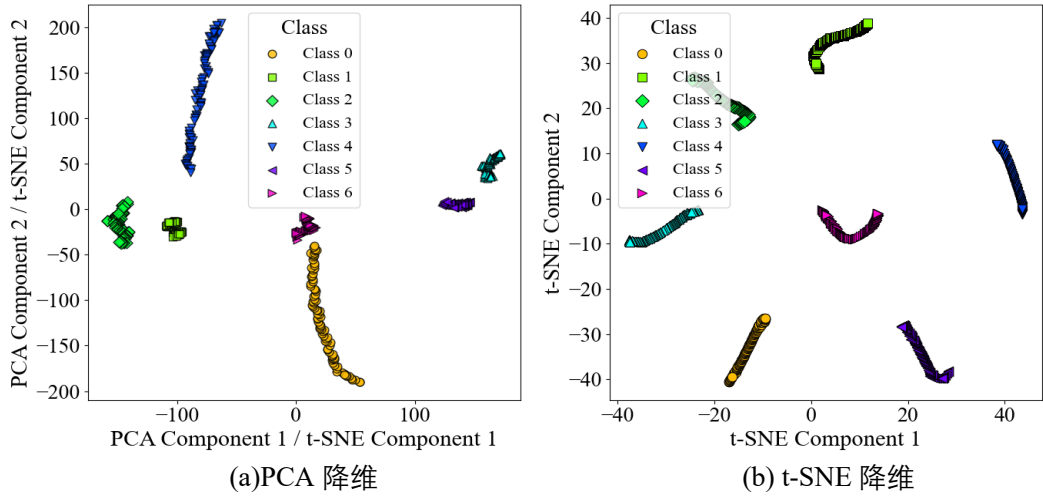


图 2-13 PCA 与 t-SNE 降维结果

由于 PCA 可以有效减少维度、减少数据噪声并降低计算复杂度，通过组合 PCA 和 t-SNE，可以将 PCA 作为预处理，先用 PCA 将数据降到相对较低的维度，再使用 t-SNE 进一步降维到 2D，大大加快 t-SNE 计算速度，利用 PCA 的速度和稳定性的同时具有 t-SNE 对非线性结构的优秀可视化效果。

2.8 数据采样

数据采样决定模型训练时的数据多样性和平衡性，避免因数据偏差导致模型泛化能力下降，数据不平衡时直接随机抽样会导致训练集或测试集中的某些类别样本数量不足或丢失，少数类样本过少会导致模型更倾向于多数类，对少数类预测性能较差，分层采样通过按照原始数据集中类别分布比例抽样，保证训练和测试数据中每个类别代表性，使模型在训练时不会偏向于样本数较多的类别，从而提高模型对所有类别预测性能，过采样通过增加少数类样本数量，使类别分布趋于平衡，减少模型偏差。

2.8.1 分层采样

分层采样适用于数据集中存在多个类别且为非均匀分布，先将数据分为不同的层，然后在每一层中按照比例进行随机采样，确保每一类数据在样本中得到正确标签。假设数据集被分成 k 个层次 $L_1, L_2, L_3, \dots, L_k$ ，每个层次有 $N_1, N_2, N_3, \dots, N_k$ 个样本，总样本量为 N ，抽取总数为 n 的样本，每个层次中样本比例为：

$$P_i = \frac{N_i}{N} \quad (2-9)$$

抽取样本数：

$$n_i = \frac{N_i}{N} \times n \quad (2-10)$$

分层采样目标是确保各层次的数据分布不变，假设 L_i 层的样本比例为 P_i ，抽取样本数量为 n_i ，并且 N_i 保持与总样本量一致，则：

$$n_i = P_i \times n = \frac{N_i}{N} \times n$$

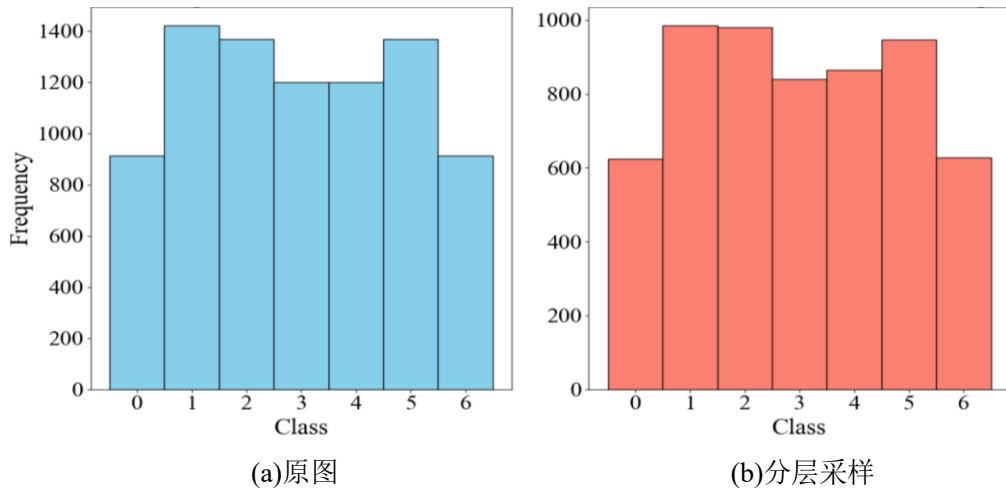


图 2-14 视网膜病变数据集类别分层采样

图 2-14 表示在原始数据集和数据处理后的各个类别图像数量分布，揭示类别间的不平衡，横坐标表示不同类别标签，每个数字对应于数据集中定义的一个特定类别，纵坐标表示每个类别中图像的数量，类别 0 减少不到三百张，而样本数量较多的类别 1 在处理前后减少四百多张，减小类别分布不均匀，并且数据处理后每个类别的样本数量与原始数据相似，确保模型在不同类别上都能学习到足够的信息。

表 2-1 分层采样前后数据量对比

类别	原始数据量	分层采样后
0	912	624
1	1422	985
2	1368	979
3	1200	840
4	1200	865
5	1368	947
6	912	627
总计	8382	5867

表 2-1 展示前面处理过程剔除无关图片等导致每个类别样本数量都不同程度减少，

减少比例在 27.92%到 31.77%之间，即分层采样前后数据量对比，可以更好理解数据处理的影响，为后续模型训练做出相应调整，总体来说分层采样能够有效降低抽样误差。

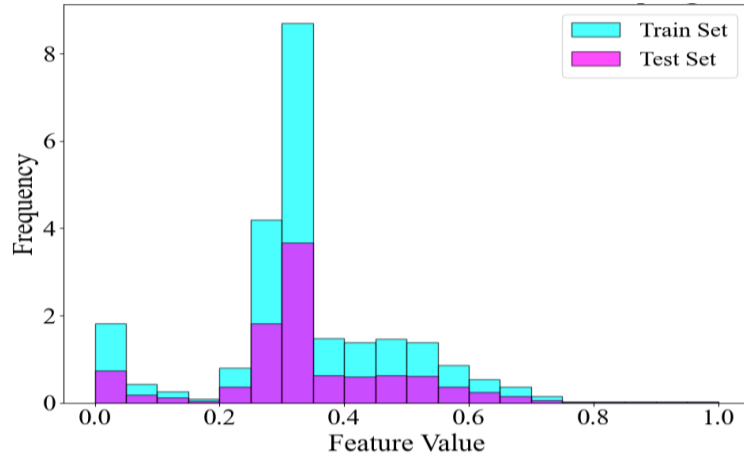


图 2-15 视网膜病变数据集特征分布

图 2-15 显示训练集和测试集中特征值分布情况，特征值是图像经过处理后得到的数值表示，横坐标表示图像特征的数值范围，纵坐标表示特征值出现频率，即有多少个样本的特征值落在某个范围内，训练集的特征值频率在多个区间有明显的高峰，特别是在 0.25 到 0.35 之间，表明这一范围内的数据量非常大，而测试集的特征值分布则相对较为平坦，尤其在高频区间上，频率明显低于训练集，有助于提升模型的泛化能力。

特征分布图为评估模型的泛化能力提供重要信息，理想情况下训练集与测试集的特征分布应当相似，从图 2-15 看出数据表现良好。

2.8.2 过采样

过采样是针对分类问题中类别不平衡问题的处理方法，在训练中少数类样本不足时难以让模型学习到其特征，采用过采样来增加少数类样本数量来平衡类别分布，以提升模型对这些类别的理解和预测能力。

假设类别 C_1 和 C_2 样本量分别为 N_1 和 N_2 （其中 $N_1 > N_2$ ），为平衡类别分布，通过复制或生成少数类 C_2 的样本，使得最终样本量达到 N_1 ，希望少数类样本数与多数类样本数相等，增加的少数类样本数为：

$$n_{\text{new}} = N_1 - N_2 \quad (2-11)$$

利用合成数据生成方法(SMOTE)来进行过采样以生成更多的少数类样本，使类别分布均衡，分别绘制原始和过采样后数据分布柱状图和特征分布散点图：

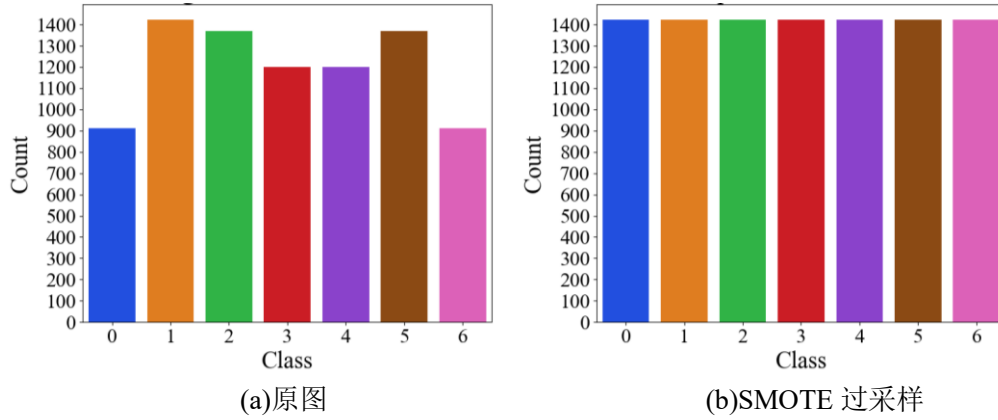


图 2-16 数据集类别分布柱状图

图 2-16(a)图展示原始类别分布柱状图,显示每个类别的样本数量,y 轴的刻度从 200 到 1000,反映样本不平衡分布,(b)图展示经过 SMOTE 过采样后各类别样本数量分布,图中每个类别的样本数量趋于平衡,数据集图像数量保持一致,均为 1429 张图片,共计 10003 张图片,表明数据处理的效果较好。

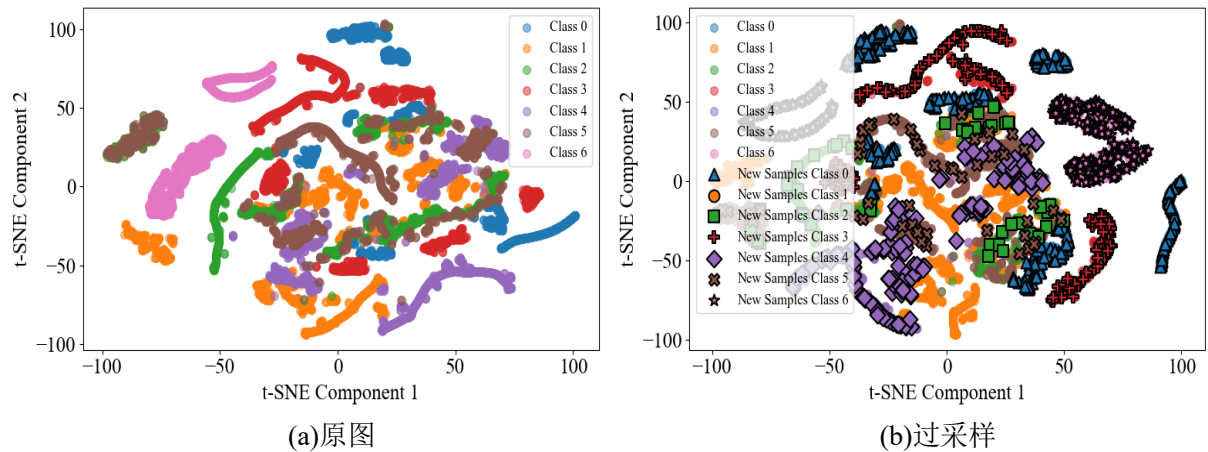


图 2-17 t-SNE 经过 SMOTE 特征分布

图 2-17 特征分布散点图有助于直观观察数据在特征空间中分布,展示少数类样本如何通过过采样补充,(a)图使用 t-SNE 降维技术展示原始数据的特征分布,图中不同类别的样本以散点形式展示,体现类别间的分布情况,(b)图使用 t-SNE 经过 SMOTE 过采样后数据特征分布,并用不同形状标记新生成样本,展示过采样后的数据在特征空间分布情况,图中可以看出新样本与原始样本在特征空间中为合理的分布。

2.9 数据变换

数据变换是对数据进行数学运算以改变其分布或范围,常用于正态化数据或减少偏度,常用方法包括对数变换、平方根变换、Box-Cox 变换等。

2.9.1 对数变换(Log Transformation)

对数变换通过对原始值取对数，用于解决数据方差较大、偏态分布、具有极端差异的情况，对数变换将数据拉近，使数据更加符合正态分布。

$$x' = \log(x) \quad (2-12)$$

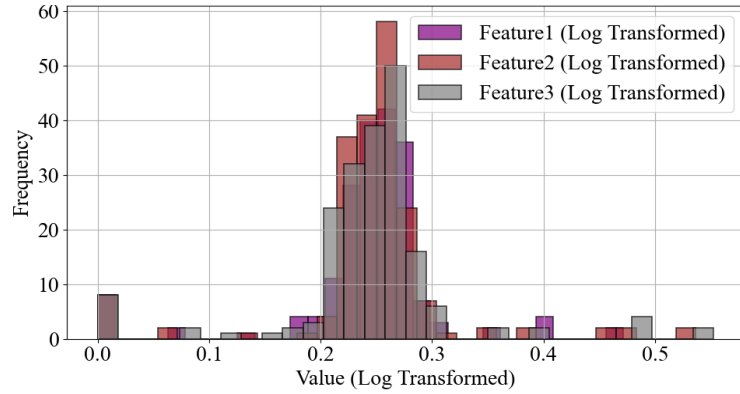


图 2-18 对数变换特征分布直方图

图 2-18 显示经过对数变换后数据分布直方图，每个立柱代表该特征在特定数值范围内数据频率，Feature1、2、3 与图像对比度、颜色分布和形状特征相关，均由图像数据本身决定，经过对数变换后特征分布较平滑，数据均符合正态分布，长尾现象被减弱。

2.9.2 平方根变换(Square Root Transformation)

适用于较小范围数据，能减少数据的偏度，公式如下：

$$x' = \sqrt{x} \quad (2-13)$$

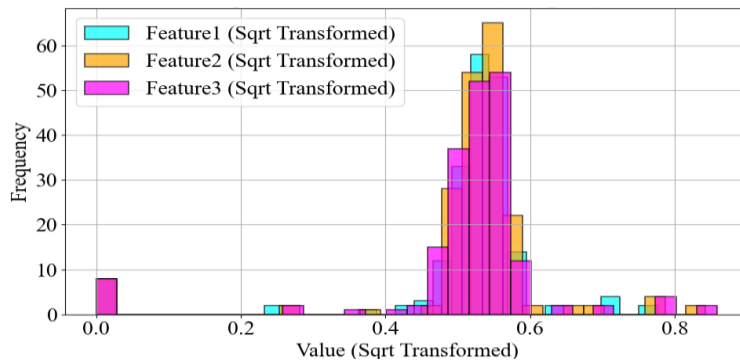


图 2-19 平方根变换特征分布直方图

数据均符合正态分布，经过对数变换后特征分布较平滑，长尾现象被减弱。

2.9.3 Box-Cox 变换

通过调节参数 λ 进行非线性变换，以使数据更接近正态分布，公式如下：

$$x' = \begin{cases} \frac{x^\lambda - 1}{\lambda}, \lambda \neq 0 \\ \log(x), \lambda = 0 \end{cases} \quad (2-14)$$

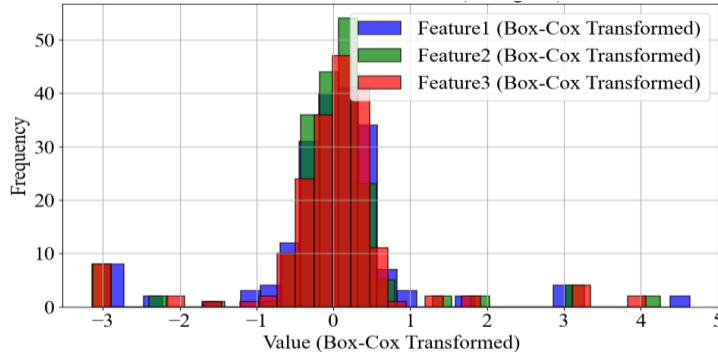
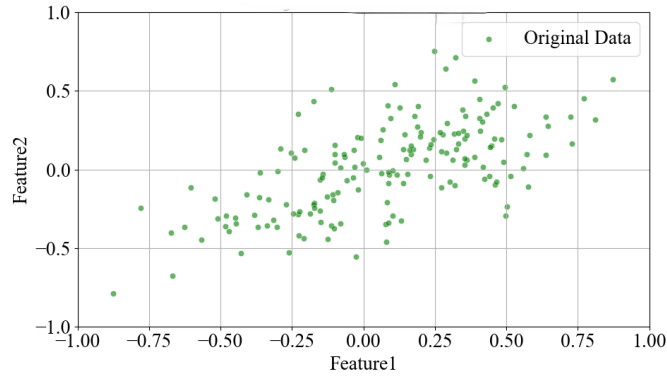
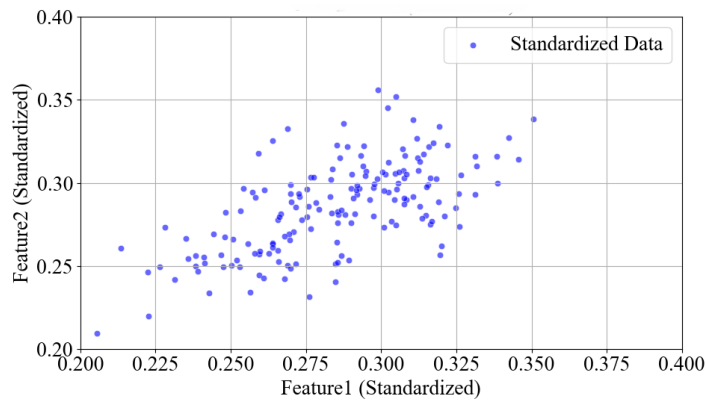


图 2-20 Box-Cox 变换特征分布直方图

数据均符合正态分布，经过对数变换后特征分布较平滑，长尾现象被减弱。



(a)原图



(b)标准化

图 2-21 原始与标准化数据散点图

图 2-21(a)显示特征 Feature1 和 Feature2 原始数据分布，每个点表示一条数据记录，点的位置由 Feature1 的值（X 轴）和 Feature2 的值（Y 轴）决定，是原始数据最直观展示，反映出两个特征之间的相关性与异常点。

图 2-21(b)展示标准化处理后数据分布，将数据会被压缩到更集中统一尺度，标准化

后的散点图与原始散点图的形态相似，目的是消除特征之间尺度不同带来的影响（尤其是在梯度下降算法），以提高模型收敛速度和准确性，通过观察标准化后数据分布形态，存在线性关系且集中在[0.200,0.350]区域。

2.9.4 数据变换评价（偏度与峰度）

(1)Skewness（偏度）：用于衡量数据分布对称性，偏度大于 0 说明数据分布右侧比左侧更长或有更极端的值，分布有“长尾”向右延伸，称为正偏（右偏）；偏度小于 0 反之，称为负偏（左偏）；偏度为 0 表示类似正态分布的对称分布，计算公式如下：

$$Skewness = \frac{n}{(n-1)(n-2)} \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{\sigma} \right)^3 \quad (2-15)$$

式中： n 是样本数量； x_i 是第 i 个样本； $\bar{x}$ 是样本均值； σ 是样本标准差。

(2)Kurtosis（峰度）：用于描述数据分布相比正态分布的陡峭或平坦程度，公式如下：

$$Kurtosis = \frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{\sigma} \right)^4 - \frac{3(n-1)^2}{(n-2)(n-3)} \quad (2-16)$$

偏度展示数据的对称性或倾斜情况，而峰度帮助理解数据集中在中心的程度以及极端值出现的频率。

表 2-2 偏度与峰度数据表

方法	Skewness（偏度）	Kurtosis（峰度）
对数变换	-0.0814	-0.6676
平方根变换	-0.1415	-0.6867
Box-Cox 变换	-0.0032	-0.3583

(3)评价：(a)对数变换：Skewness=-0.0814 表明数据分布略微负偏，接近对称（接近 0）；Kurtosis=-0.6676 表示峰度低于正态分布（峰度为 0），分布相对更平坦，对数变换有效增加数据对称性（Skewness 接近 0），并使得数据峰度较为平缓，尾部较轻。

(b)平方根变换：Skewness= -0.1415 表明数据分布略微向左倾斜（负偏），与对数变换结果相似，接近对称；Kurtosis= -0.6867 表明与对数变换后的峰度接近，分布比较平坦，尾部相对较轻，平方根变换也较好地处理数据对称性和峰度。

(c)Box-Cox 变换：Skewness= -0.0032 表明数据几乎完全对称，很好地消除分布偏斜；Kurtosis= -0.3583 说明数据峰度比前两种变换更低，尾部更轻，分布更加平缓。

综上所述，Box-Cox 变换在消除偏度方面效果最佳（几乎完全对称），但峰度更低，

分布较平坦，尾部极轻，效果在三种变换方式中最佳。

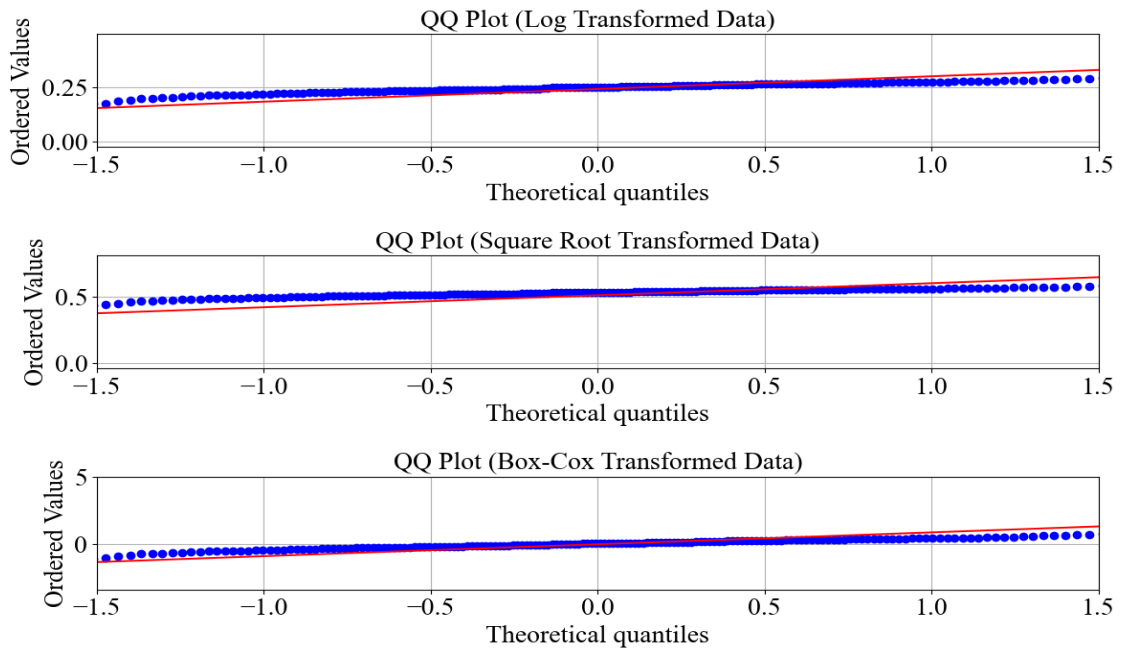


图 2-22 Quantile-Quantile Plot 概率图

图 2-22 的 Quantile-Quantile Plot（分位数-分位数图）是概率图，用于比较两个分布之间相似性或一个数据集是否符合特定的理论分布，采用三种变换来处理数据目标是让数据分布接近正态分布，采用三种变换后的数据理论分布均为正态分布，图中横坐标代表理论分布的分位数，对应于理论上正态分布中的标准分位点，横轴每个点表示正态分布中应该出现的值，纵坐标代表实际样本经过数据变换后的分位数，这些分位数表示经过对数变换后的数据在不同位置上对应的真实分布情况。

图中的点接近红色对角线，说明实际数据的分布与理论分布非常接近，点偏离对角线，说明数据分布与理论分布有差异，偏差越大，数据分布与理论分布差异越明显，图中的数据点大多接近红色对角线，表明经过数据变换后，样本数据的分位数与理论分布的分位数非常接近，数据分布实际形态与理论分布呈现出高度一致性，数据偏度和峰度特征与理论分布的假设相符，证明经过变换后的数据趋近正态性，符合正态分布的假设。

2.9.5 数据变换总结

Box-Cox 变换效果最佳，尤其在处理偏度方面，其结果几乎完全对称，在需要正态分布数据时，其对模型训练而言是最有利的变换，对数变换和平方根变换在处理偏度和峰度方面效果相似，略微向左倾斜，峰度平缓，追求分布对称性，Box-Cox 变换是首选。

2.10 自适应直方图均衡化

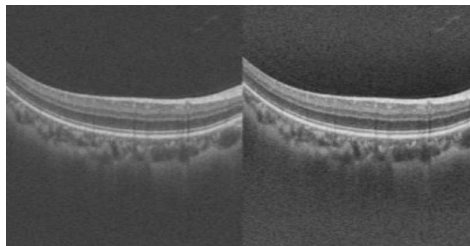
2.10.1 累积分布函数(CDF)与直方图(Histogram)

累积分布函数表示图像像素值的累积分布情况，是直方图的累积形式，显示每个像素值对应的累积像素数量，可以用来描述图像像素值的分布情况。

直方图表示图像灰度级分布情况，显示不同灰度级在图像中出现的频率或数量，从而帮助了解图像对比度、亮度等特征。

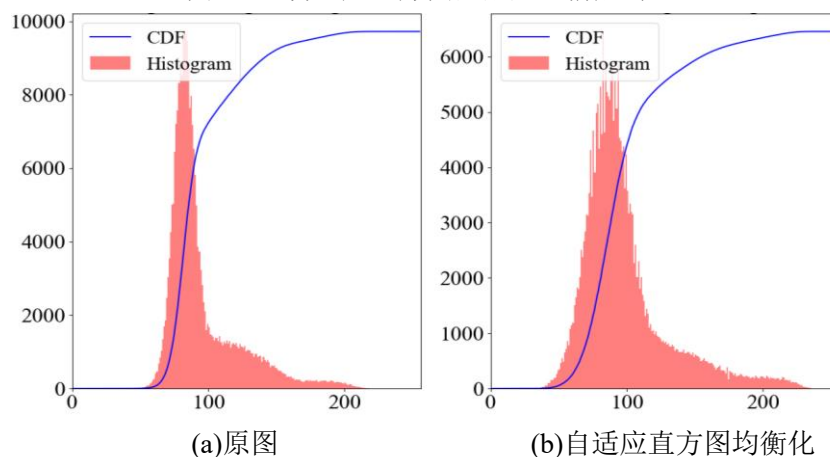
2.10.2 自适应直方图均衡化评价指标

在图像处理任务中，使用的均方误差作为评分标准，**Best Score** 表示所选择的参数组合在均方误差上的表现，较高的 **Best Score** 表示均方误差较低，意味着处理后的图像更接近原始图像，效果更好。



(a)原图 (b)自适应直方图匹配

图 2-23 自适应直方图匹配处理前后对比



(a)原图

(b)自适应直方图均衡化

图 2-24 自适应直方图均衡化的直方图与累积分布函数

图 2-24 中的直方图较为集中并呈现正态分布，意味着图像的亮度值分布相对均匀且对称，图像经过有效的均衡和优化，正态分布的形态显示处理后图像亮度分布比较符合人眼感知习惯，不会出现过暗或过亮的区域，直方图集中表示图像的像素值在合理范围内，避免极端亮度值，改善图像的整体视觉效果，尤其是在细节和层次的呈现上。

通过将直方图波峰处适度降低，减少高亮度区域过度增强，差异更加平衡，避免因对比度过高而导致细节丢失或噪声增强以及过度曝光或曝光不足现象，图像视觉效果更为柔和和自然，有助于减少图像中局部区域的对比度过强而产生的噪声，尤其是在一些高亮或阴影区域中，避免极端的亮度差异对图像质量的负面影响，在提升图像质量方面有显著作用，图像对比度得到均衡优化，细节得到充分保留，而高频噪声和不自然亮度差异得到有效抑制，图像看起来更加自然、舒适且富有层次感，细节也更清晰地展现出来，这种方法不仅增强图像视觉效果，还为后续图像分析提供高质量图像基础。

从图 2-24 中看出自适应直方图均衡化处理方法强调对比度优化，并避免高对比度带来的不自然感，改进后的图像最高值在 7000 左右降低到不足 5000，图像在细节展示上会更加清晰，特别是在低对比度区域，细节层次变得更加明显，提升图像视觉清晰度和质量，第二张图片的图像的纹理、边缘等信息更加突出，而不会损失原有的自然感觉。

在均衡化时用于限制直方图中某些亮度值的累积频率，避免过亮或过暗的像素区域过度增强，如果某个亮度值的频率超过对比度阈值，会被裁剪并重新分配到直方图中较低的频率区域，较低的阈值对比度增强较弱，图像变化较小，效果较好，较高阈值会导致噪声过度放大，这里均采用 1.0。

将输入图像分成小的矩形网格，然后对每个网格单独应用直方图均衡化，如果网格尺寸小，处理细节区域较多，均衡化效果更局部化，但可能导致噪声增多，网格尺寸大，处理的区域更广泛，均衡化效果更平滑，但可能会丢失局部对比度，这里采用自适应的网格裁剪大小，根据图片中的信息来自动划分裁剪大小，得到表 2-3 的 Best Score。

表 2-3 部分图像 Best Score

序号	对比度的阈值	网格大小	Best Score
01	1.0	(12, 1)	78.3906
02	1.0	(3, 5)	78.3417
03	1.0	(2, 3)	78.1950
04	1.0	(2, 3)	78.1180
05	1.0	(3, 3)	76.4262

分数在 78.0 ± 2.0 范围内波动，整体差距较小且处理效果较好，较小的波动意味有较好的效果，从数值来看，最佳分数较高表明 CLAHE 参数选择方法在对比度增强的同时保持相对较低的误差值，效果不错，图像的均方误差较低，表明当前图像处理效果较优，

处理后的图像与原始图像更加相似。

2.11 直方图匹配

对原图和目标图像的直方图都做均衡化，变成相同标准化的均匀直方图，设原图像和输出图像的灰度变量分别为 r 和 z ，则得对应概率密度：

$$T(r) = (L-1) \int_0^r p_r(w) dw \quad (2-17)$$

$$G(z) = (L-1) \int_0^z p_z(w) dw \quad (2-18)$$

上式都为均匀化变换，得 $T(r)=G(z)$ ， z 满足：

$$z = G^{-1}[T(r)] \quad (2-19)$$

在处理离散变量时，仅得到一个近似的直方图，公式如下：

$$s_k = T(r_k) = (L-1) \sum_{i=0}^k P_r(r_i) = \frac{(L-1)}{MN} \sum_{i=0}^k n_i \quad k = 0, 1, \dots, L-1 \quad (2-20)$$

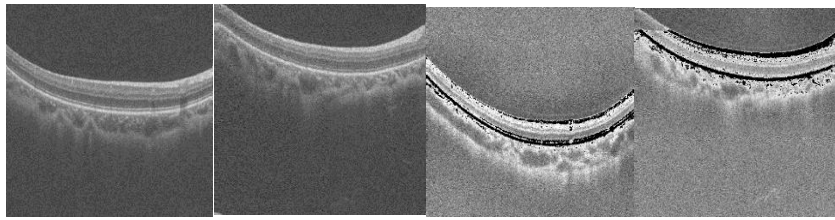
式中： MN 为像素总数； n_j 为灰度值 r_j 像素； L 为灰度级数，计算变换函数：

$$G(z_q) = (L-1) \sum_{i=0}^q p_z(z_i) \quad (2-21)$$

对一个 q 值，有：

$$z_q = G^{-1}(s_k) \quad (2-22)$$

即得到图 2-25 直方图匹配结果：



(a)原图

(b)直方图匹配

图 2-25 直方图匹配结果

2.12 直方图处理后图片评价指标

2.12.1 均方根误差(RMSE)

RMSE 为原始与经过处理后图像的像素差异，值较低表示差异较小，图像质量较高，公式如下：

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \hat{x}_i)^2} \quad (2-23)$$

式中： n 是样本的数量； x_i 是第 i 个样本的实际值； $\hat{x}_i$ 是第 i 个样本的预测值。

2.12.2 峰值信噪比(PSNR)

PSNR 是衡量图像质量的指标，其值越高，表示处理后图像与原始图像之间失真程度越低，图像质量越高，公式如下：

$$PSNR = 20 \log_{10} \left(\frac{MAX_i}{RMSE} \right) \quad (2-24)$$

式中： MAX_i 是图像像素的最大值（对于 8-bit 图像，通常为 255）。

2.13 直方图均衡化与直方图匹配实验结果

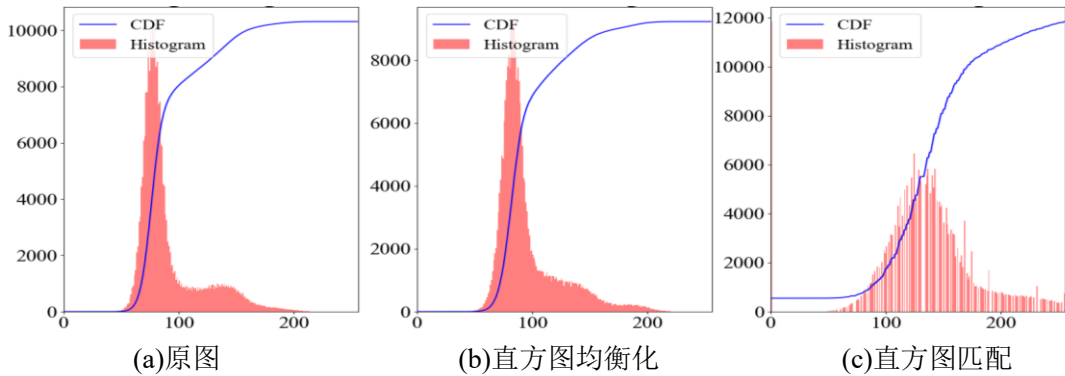


图 2-26 最终的直方图与累积分布函数

图 2-26 中的直方图趋向正态分布并降低波峰，能够提供更为舒适的视觉效果，适中调整提供更加平衡的视觉感受，特别是在图像深色和亮色区域，避免任何过于尖锐或突兀效果，进一步降低波峰使得图像变得更加平缓，图像整体视觉效果看起来更加自然，减少由于过度处理所带来的图像伪影或噪声。

通过直方图的合理调整，将图像的亮度值更好地分布在合适范围内，不仅增强图像细节和对比度，还保证图像整体视觉效果，无论是在保持细节清晰度、减少噪声，还是在提升图像层次感方面，这种处理方法都做到非常好的平衡，采用的图像处理方式是可行的，适合需要准确、真实呈现图像内容应用场景。

直方图更加平滑，高频噪声得到抑制，图像的整体质量得到显著提升，过多的高频成分通常会造成图像变得颗粒感十足，甚至让细节显得不清晰，平滑直方图能够消除这种颗粒感，改善图像质量，尤其是在低对比度区域，细节能够更加清晰地呈现，图像的

亮度值在较低和较高区域得到更均匀的分布，使得图像深色和亮色区域的细节得到保留，并且这些细节不会因为过度增强或压缩而丢失，图像中的每一部分都能展现出更多的纹理和细节，使整体图像看起来更加丰富和自然，并且让图像的亮度过渡更加柔和，增强图像的深度感和层次感，特别是在图像的复杂区域，细节和层次感的增强让图像看起来更加立体和生动。

表 2-4 RMSE 和 PSNR

序号	评价指标	自适应直方图均衡化	直方图匹配
01	RMSE	0.7084	0.5858
	PSNR	47.3755	47.6363
02	RMSE	1.0349	0.8277
	PSNR	47.2753	47.5206
03	RMSE	0.8570	0.7048
	PSNR	47.4165	47.5584

从表 2-4 看出，RMSE 在所有测试图像中都有明显改善，RMSE 降低（图像误差减少） $18.03\% \pm 2.15\%$ ，RMSE 降低说明经过 CLAHE 和直方图匹配处理后，相较于仅进行 CLAHE 的图像差异更小，处理后的图像与原图在视觉上相似度更高，图像失真程度和误差降低，质量得到优化，PSNR 提升 $0.46\% \pm 0.16\%$ ，PSNR 升高表示经过上述流程后值升高，图像质量较高，噪声和失真程度降低。

自适应直方图均衡化先对图像局部对比度进行增强，使得图像的局部细节更加突出，再进行直方图匹配，使得图像亮度分布符合目标图像分布，从而达到图像质量提升和目标一致性的平衡，有效保留原图细节，并在亮度调整方面更加自然，适用于对比度和亮度需要改善的图像，这种图像处理方法（自适应直方图均衡化与直方图匹配结合）对图像质量的提升是有效的，通过优化局部细节和全局亮度分布，提高图像可视效果和质量，表现较好，尤其在 RMSE 指标上，图像误差有明显降低，且 PSNR 指标也有所改善，证明该方法能够优化图像对比度和细节，从而提高视觉质量。

第3章 SF-Unet 图像分割

首先基于 U-Net 和 CS-Unet 两种模型优势，提出更适合视网膜病变数据集分割的改进网络——Shifting-Unet（简称 SF-Unet），在实验部分比较多个网络模型的分割效果，包括 U-Net、CS-U-Net、Attention-Unet、CSP-Unet 以及 SF-Unet，通过对比实验和交叉验证，评估不同分割网络模型在 OCT 图像提取中的表现，旨在提升图像分割的精度和准确度，结果表明 SF-Unet 在视网膜病变图像的分割提取中具有明显的优越性。

3.1 U-Net 与 CSU-Net 分割

首先使用 U-Net 和 CSU-Net 卷积神经网络模型进行 OCT 数据集图像分割提取，得到 Dice 和 Iou 评分如表 3-1：

表 3-1 Dice 和 Iou 评分

网络模型	Mean	Std	Average Dice	Average Iou
U-Net	0.6111	0.0693	0.3375	0.2034
CSU-Net	0.6474	0.0601	0.5337	0.3640

U-Net 和 CSU-Net 在图像分割任务中的性能对比表明，CSU-Net 在各个指标上均优于 U-Net，显示出更高的预测准确性和稳定性。

3.2 改进 Unet 分割网络：Shifting-Unet

3.2.1 总体介绍

遵循 U-Net 设计，由编码器-解码器、瓶颈层和跳跃连接层组成，编解码器有下采样和上采样四个阶段，分别由最大池化和双线性插值实现，编码器-解码器的前三层分别包含两个标准 3×3 卷积块，第四层和瓶颈层采用特征激励块来处理特征通道的压缩和扩展，DOSD 块来捕获图像的局部上下文和远程依赖关系，跳跃连接通过加法融合相同尺度的特征，并且不采用 Transformer，其计算复杂度高，局部特征学习能力差。

CNN 很难捕获长距离依赖关系，因此提出核心 S-MLP 模块，该模块负责学习整个图像在单一方向上的长距离依赖关系，通过对不同方向的 S-MLP 模块进行控制和组合，形成 OS-MLP 和 DOS-MLP 模块，以捕获多方向的远程依赖关系，基于改进的 MLP 模块和 UNet 模型网络，提出名为 Shifting-Unet（简称 SF-Unet）网络，SF-Unet 网络包含

不同变体, 如 SF-Unet S、M、L, 每个变体具有不同的参数配置, 以适应不同任务和计算能力, 主要工作:

- (1)提出一种新的远程依赖捕获方法, 并构建 S-MLP 模块;
- (2)构建 OS-MLP 和 DOS-MLP 模块, 可以获得更多方向上的远程依赖关系;
- (3)提出 DOSD 模块, 同时提取本地上下文信息和远程依赖关系, 而不增加计算负担并且具有与 3×3 卷积相同的参数和计算级别。

3.2.2 多层感知机(MLP)

多层感知机(MLP)拥有提取全局特征能力, 能够学习输入数据与输出目标之间的非线性关系, 是监督学习任务中常用模型, 尤其在分类问题中表现优异。

在运用 CNN 或者 RNN 去解决实际问题时, 长距离记忆问题一直是难题, 就像去浏览一张图片, 想要在较短时间和耗费较少精力的情况下将图片中的所有细节全部记忆下来很困难, 同样在神经网络中, 在不增加网络深度的同时想要将数据集中尽可能多的信息记录下来也很困难, 很好的方法就是添加 MLP, 其基本结构包括输入层(接收输入特征向量, 每个节点代表一个输入特征)、输出层(输出最终预测结果, 输出层的神经元数量取决于具体任务, 如分类任务中的类别数)和至少一层或多层的隐藏层^[22]。

MLP 通过反向传播算法来调节节点权重, 使得模型在训练数据上的预测误差最小化^[26], 反向传播算法步骤如下:

- (1)前向传播: 通过前向传播计算模型输出, 逐层计算输出层, 得到模型预测结果;
- (2)计算损失: 然后计算模型在训练数据上的损失, 使用损失函数来衡量模型预测值与真实值差异;
- (3)反向传播误差: 从输出层开始, 逐层计算每个参数的梯度, 然后将梯度沿着网络反向传播, 直到计算出输入层的参数梯度;
- (4)更新参数: 使用梯度下降来更新参数, 步长由学习率控制, 学习率越大, 参数更新幅度就越大, 学习速度越快, 但是学习率过大可能会导致优化过程不稳定, 甚至发散。

3.2.3 激活函数

激活函数是增强网络表示和学习能力的非线性函数。

- (1)Sigmoid:

$$S(x) = \frac{\exp(x)}{\exp(x) + \exp(0)} = \frac{1}{1 + \exp(-x)} \quad (3-1)$$

图像是 S 型曲线，把实数域的输入 x 压缩到(0,1)之间，当输入值靠近两端时，对输入进行抑制，存在梯度消失问题，大量输入被输入到神经网络，隐藏层数量增加，梯度变得接近于零，从而导致神经网络不准确，并且公式中包括指数运算，使得梯度下降的收敛速度变慢，会降低权重更新效率，一般用于分类网络的输出层。

(2)Softmax:

$$\text{Softmax}(x_i) = \frac{\exp(x_i)}{\sum_j \exp(x_j)} \quad (3-2)$$

以规范化输出，用于网络输出层部分，可以与其他激活函数结合使用，以概率形式产生输出，用于多类分类并生成总和为 1 的概率输出，确保较小值具有较小概率并且不会丢弃，二分类问题的网络输出层使用 Sigmoid，多分类问题的网络输出层使用 Softmax。

(3)ReLU:

$$\text{RELU} = (0, x) \quad (3-3)$$

ReLU 在计算上更高效，相对于 Sigmoid 函数的两端饱和，ReLU 函数为左饱和函数，且在 $x > 0$ 时的导数为 1，Relu 使用的优化器一般是 SGD 算法，收敛速度比 Sigmoid 快。

(4)GELU:

通过门控机制来调整其输出值的激活函数，其表达式为：

$$\text{GELU}(x) = xP(X \leq x) \quad (3-4)$$

式中 $X \leq x$ 是高斯分布 $N(\mu, \sigma^2)$ 的累积分布函数，一般取标准分布 $\mu = 0, \sigma = 1$ ，加入随机正则思想，由于高斯分布累积分布函数为 S 型函数，所以用 Tanh 函数来近似：

$$\text{GELU}(x) \approx 0.5x \left\{ 1 + \tanh \left[\sqrt{\frac{2}{\pi}} (x + 0.04715x^3) \right] \right\} \quad (3-5)$$

GELU 函数在临近原点时梯度不为零，减少训练过程中梯度消失，导函数较光滑。

3.2.4 Shifting-MLP 结构

连接输入层和隐藏层的是权重和偏置，由 X 计算得到 H 的矩阵运算：

$$H_h = X_i w_h + b_h \quad (3-6)$$

加粗的输入、权重及隐藏层神经元组合为一个感知机，也可以表示为：

$$\begin{bmatrix} x_a & \cdots & x_m \\ \vdots & \ddots & \vdots \\ x_j & \cdots & x_{jm} \end{bmatrix} \cdot \begin{bmatrix} w_b & \cdots & w_h \\ \vdots & \ddots & \vdots \\ w_m & \cdots & w_{mh} \end{bmatrix} = \begin{bmatrix} x_a w_b + \dots + x_m w_m & \cdots & x_a w_h + \dots + x_m w_{mh} \\ \vdots & \ddots & \vdots \\ x_j w_b + \dots + x_{jm} w_m & \cdots & x_j w_h + \dots + x_{jm} w_{mh} \end{bmatrix} \quad (3-7)$$

同样可以得到：

$$Y_y = H_h w_k + b_k \quad (3-8)$$

这里将 b 设置为 0，将上面的函数改为：

$$H'_h = X_i w_h^T + b_h \quad (3-9)$$

可以得到：

$$Y_y = H'_h w_k + b_k \quad (3-10)$$

就可以得到：

$$\begin{bmatrix} x_a & \cdots & x_m \\ \vdots & \ddots & \vdots \\ x_j & \cdots & x_{jm} \end{bmatrix} \cdot \begin{bmatrix} w_b & \cdots & w_m \\ \vdots & \ddots & \vdots \\ w_h & \cdots & w_{hm} \end{bmatrix} = \begin{bmatrix} x_a w_b + \dots + x_h w_h & \cdots & x_a w_m + \dots + x_h w_{hm} & \cdots \\ \dots & \ddots & \dots & \dots \\ x_j w_b + \dots + x_{jh} w_h & \cdots & x_j w_m + \dots + x_{jh} w_{hm} & \cdots \\ \dots & \dots & \dots & \dots \end{bmatrix} \quad (3-11)$$

当 $m < h$ 时， $h-m$ 部分不变并裁剪平移到图中；当 $m > h$ 时，对权重进行补充，补充的方法是随机复制 w_i 行成第 $w_{(h+1)}$ ，以此类推。

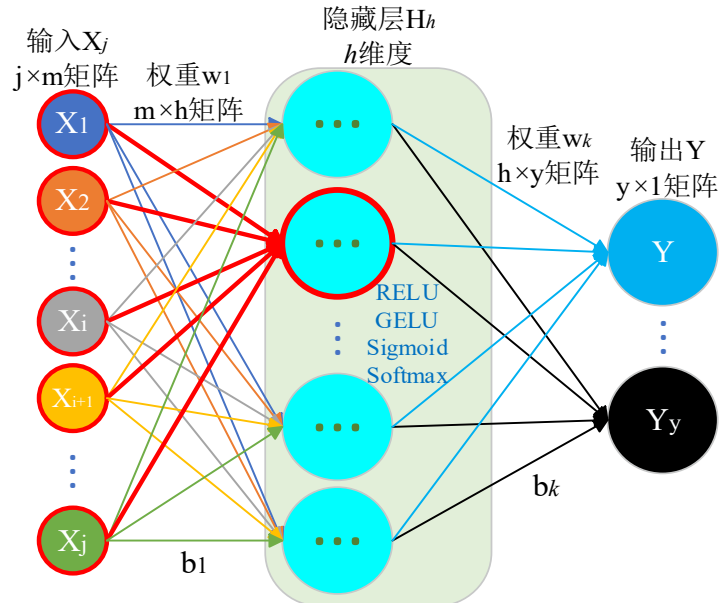


图 3-1 MLP 结构图

对于如图 3-1 结构，假设给定输入图片的尺寸为 $H \times w$ ，则特征矩阵为 $x \in (H \times w \times$

c), Shifting 意味平移, 先不考虑通道数个数, 先根据每个空间位置索引(H_i, w_j), 其中 $i \in (1, H), j \in (1, W)$, 将特征图上的除 $i=j=1$ 区域外所有其他区域进行平移, 水平或者竖直方向根据选择, 即高度或者宽度方向上, 得到 Shifting-MLP 结构 (简称 S-MLP), S-MLP 模块对每个通道层的特征图沿指定方向执行 Shifting 操作, 得到正对角或负对角特征图。

当特征图在某一方向上移动时, 每个通道的特征图都会进行相应移动, 在图形处理中, 一般默认 $H=W$, 但在实际情况中, 这两个参数往往不相等, 移位后的每个通道的特征图可能会超出原始特征图的边界, 此时需要进行相应裁剪和补充, 裁剪步骤会去除多余部分, 并将缺失部分补充回到另外一边, 确保所有通道的特征图在宽度上对齐, 此操作也是根据图像形状而定, 进行类似循环直到输出的尺寸一样为止, S-MLP 就可以在高度方向上捕获长距离依赖, 特别是当 $C < W$ 时, 此时 S-MLP 就对输入信息非常敏感。

在 S-MLP 中, 所有空间位置的通道投影共享相同权重, 无论特征图大小如何, 参数数量保持不变, 有助于减少参数量以提高计算效率, 在通道投影后应用 GELU 激活函数, 完成数据的非线性变换, 解决线性模型的表达、分类能力不足的问题。

Shifting 能计算到特征图上每个点与其他所有点的相关性, 下图中不同颜色深浅代表的是原始位置不同, 在 Shifting 之后取一个坐标点(H_h, H_w)的所有通道, 取的坐标点已经不完全不是原始的(H_h, H_w), 在下图中取任意一宽度坐标($\dots\dots, W_w$)进行观察, 可以看到颜色已经混淆, 代表同一坐标在不同通道已经混合其他位置坐标值, 如图 3-2 所示, 因此通过这一变换, 可以计算不同坐标像素之间相关性, 通过平移操作, S-MLP 能够在不增加计算负担的情况下捕获长距离信息。

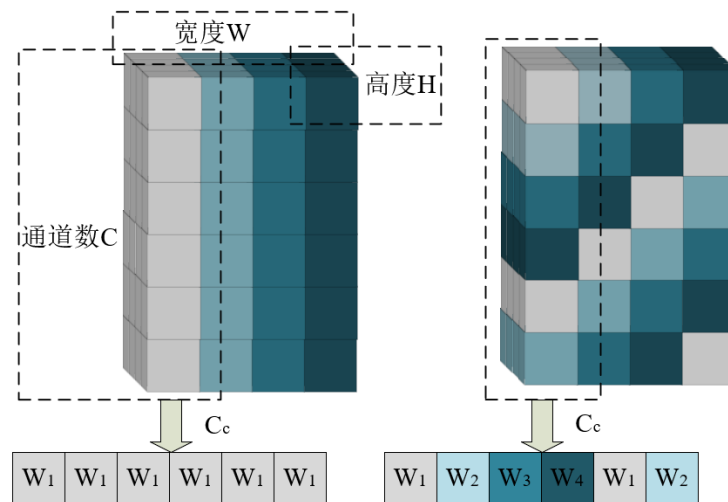


图 3-2 S-MLP 模块处理特征图示意

3.2.5 Orthogonal Shifting MLP 结构

S-MLP 对宽度或者高度方向对长距离依赖进行编码，为得到其他方向上的长距离依赖，假设 X 为输入，先进行宽度方向的 S-MLP，再进行高度的 S-MLP，相当于在两个相互垂直方向上进行平移操作，如图 3-3 所示，得到 Orthogonal Shifting MLP，简称为 OS-MLP，具体操作是先进行一个方向上的 S-MLP 操作，再沿垂直于其的方向上进行第二个 S-MLP 操作，两者之间采用 GELU 激活函数，然后再与输入 X 之间进行残差连接，引入参数 K ，当 K 为正时表示宽度从左往右平移，高度从上到下平移，当 K 为负时反之，能够捕获正对角或者负对角的长距离依赖关系，得到对角依赖关系，如图 3-4 所示。

$$MLP_{Os}(X) = MLP_S^{(2)} \left[GELU \left(MLP_S^{(1)} \right) \right] + X \quad (3-12)$$

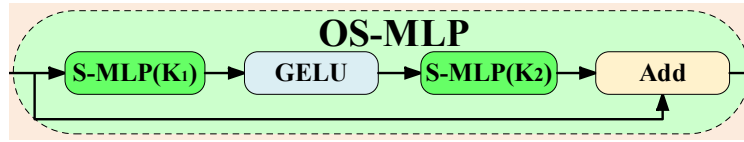


图 3-3 OS-MLP 结构图

3.2.6 特征激励块(Excitation-Block)

采用 MobileNetV3 中的特征激励块(Excitation-Block)，Excitation 通过全连接层对通道进行加权，模块类似于 UNet 的下采样操作^[27]，使用卷积来减小输入特征图空间尺寸，同时增加通道数，本质上是卷积块，主要用于对特征和通道的变化进行编码，其主要由卷积块、GELU 激活函数和 LayerNorm 组成。

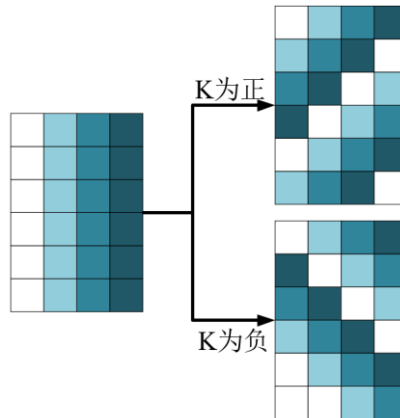


图 3-4 对角依赖关系

3.2.7 DOS-MLP 结构

连续采用两次 OS-MLP，获得宽度、高度、正副对角线，如图 3-5 所示：

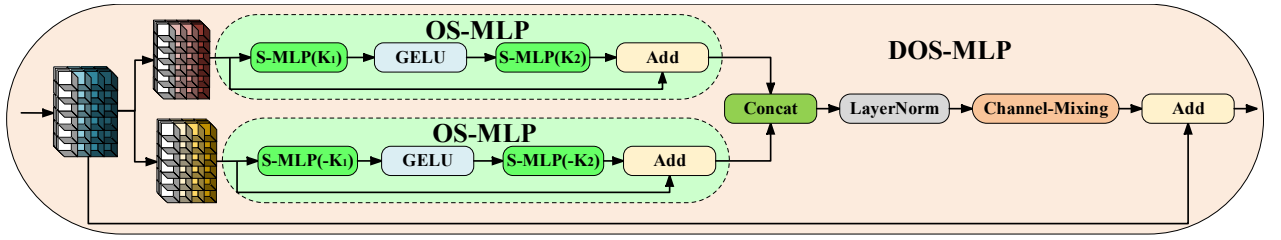


图 3-5 DOS-MLP 结构

$$MLP_{DOS}(X) = CM \left\{ LN \left[Concat \left(MLP_{OS}^{(1)}(X), MLP_{OS}^{(2)}(X) \right) \right] \right\} + X \quad (3-13)$$

DOS-MLP 模块通过并行两个互补的 OS-MLP 模块来捕获四个方向（水平、垂直、对角正向和对角负向）的长距离依赖性。

过程具体如下：

- (1)每个 OS-MLP 都有两个方向正交的 S-MLP 模块，在水平、垂直方向做 Shifting 操作，对应两个参数一正一负（ K_1 与 $-K_1$ ），从而捕获两个长距离依赖；
- (2)两个并行 OS-MLP 中的 S-MLP 顺序不同，第一个 OS-MLP 模块先在水平方向上应用 S-MLP，然后在垂直方向上应用 S-MLP；第二个 OS-MLP 模块先在垂直方向上应用 S-MLP，然后在水平方向上应用 S-MLP，发现第一个 OS-MLP 模块进行第二次 S-MLP 与第二个 OS-MLP 模块进行的第二次 S-MLP 方向正好正交，此时可以表示为一正一负（ K_2 与 $-K_2$ ），这样每个 OS-MLP 模块都能捕获两个正交方向上的依赖性，形成对角线方向的感知能力；
- (3)每个 OS-MLP 中有残差连接，以增强模型学习能力，避免梯度消失；
- (4)两个 OS-MLP 模块输出沿着通道维度进行拼接，以获得不同方向特征表示；
- (5)特征经过 LayerNorm 和 Channel-Mixing，进行残差连接与拼接。

3.2.8 DOSD 和 DOSD-Block 类

为同时提取长距离依赖关系和局部上下文信息，提出 DOSD 模块，该模块由 DOS-MLP 和 DSC^[29]（深度可分离卷积，用来提取局部上下文信息）并行组成，拼接和融合两个子模块输出之后，通过 Channel-Mixing 进一步融合，如图 3-6 所示，这一步骤有助于让不同通道间特征实现交互，增强特征表达能力，在不增加计算负担情况下有效地解决长距离依赖关系；

DOSD 是定制 MLP（多层感知机）模块，其包含多种卷积操作、GELU 非线性激活函数和归一化层，主要负责在特征维度上执行线性变换，DOSD-Block 封装 DOSD 模块

的容器，并添加残差连接，使得网络更深并且更稳定。

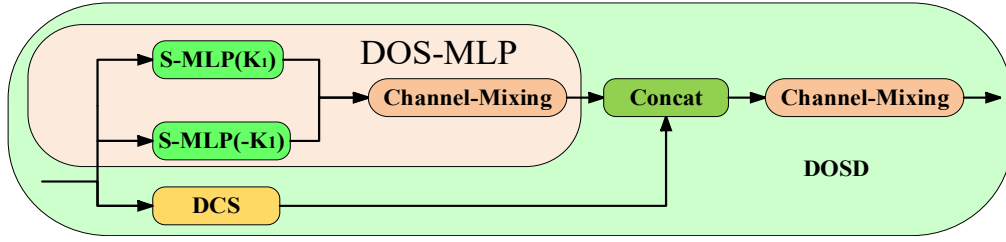


图 3-6 DOSD 模块结构

$$DOSD(X) = CM \left\{ Concat \left[MLP_{DOS}(X), DSC(X) \right] \right\} \quad (3-14)$$

Channel-Mixing 是 MLP-Mixer 架构中提出的技术，用于实现不同通道之间的特征融合，特别是在全连接或混合维度特征场景中，Channel-Mixing 先将特征张量在高度和宽度上展平，使得每个通道特征成为一维向量，然后对每个通道的特征向量应用线性变换，通常是全连接层，并跟随非线性激活函数，Channel-Mixing 的作用类似于 1×1 卷积，允许网络在不增加太多计算负担的情况下学习特征的跨通道交互^[30]。

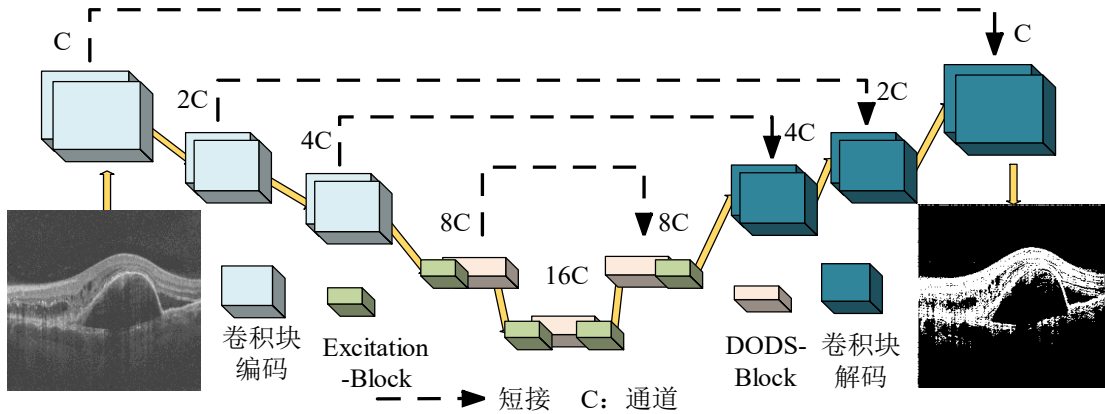


图 3-7 模型整体结构

模型整体结构如图 3-7，是经典 U 型编码器（特征提取网络）-解码器（特征融合网络）结构，但编码器第四层、瓶颈层以及解码器第一层被替换成全新的 DOSD-Block，使网络更好地提取和融合特征，从而提升医学图像分割性能。

总体是一个高分辨率—编码—低分辨率—解码—高分辨率的过程，编码器（特征提取网络）的前三个阶段，首先在图像预处理阶段将输入图像宽度和高度调整为 256 像素，对所有传入图像统一尺寸，确保后续操作能够一致应用于每一张图像，图像是 RGB 图像，输入图像有 3 个通道（红色、绿色、蓝色），输入图片尺寸为 $[3, 256, 256]$ ，3 表示每个图像特征通道数，256 表示空间维度，利用 3×3 卷积对图片进行卷积特征提取，逐步提取图像特征并缩小空间维度，通过 ReLU 激活函数输出特征通道，得到 $[32, 256,$

256], 再在左边每次输出特征通道之后, 对左边下采样过程中图片进行裁剪复制(虚线), 然后通过最大池化对图片进行下采样, 池化核大小为 2×2 (黄色线), 将图片尺寸由 256×256 缩小到 128×128 , 通道数增加到 64, 即为[64, 128, 128], 共经过三次这样操作, 得到图片尺寸为[128, 64, 64], 再将输出的图片送入到特征激励块中进行处理, 裁剪得到[128, 32, 32], 通常在编码器的最后一层(DOSD)获得中间特征, 通常是高维的特征张量, 通过采取展平操作对形状[128, 32, 32]进行处理, 经过展平操作后, 将 32×32 空间维度展平得到 1024, 得到形状 [1, 128, 1024], 批次大小为 1, 通道数为 128, 主要用于将多维输入张量转化为一维张量, 通过转置操作交换输出张量的第 1、2 维, 用于准备数据以适应不同操作或模型要求, 用于后续分割任务, 得到[1, 1024, 128, 128], 再对图像进行上采样, 卷积核大小为 2×2 , 在网络最后使用 1×1 卷积核对图片进行卷积并输出图片。

通过改变类的数量实现 Shifting-Unet 具体变体, 如 Shifting-Unet-S、M、L, 分别代表小型、中型和大型的 Shifting-Unet 网络, Shifting-Unet-S 的整个模型包含编码器、瓶颈和解码器部分, 以及四个 Excitation-Block 和三个 DOSD Block, 网络通过堆叠多层特征激励块和 DOSD Block 来构建。

Shifting-Unet-S 前向传播函数中, 输入通过卷积层再通过特征激励块和 DOSD Block 来进一步处理特征, 接着通过解码路径恢复到原始输入尺寸, 在前向传播中, 网络先通过卷积层提取特征, 再通过池化层降低维度, 通过特征激励块进一步处理特征, 并应用 DOSD-Block 进行多层感知机操作, 网络通过解码器模块恢复空间维度, 并输出分割结果, 前向传播过程通过多个阶段和模块对输入图像进行处理, 从低层次特征到高层次语义特征, 最后生成分割输出, 通过特征激励块和多层感知机块(DOSD)进行多尺度特征提取和增强, 添加跳跃连接来保持特征多样性, 并在网络深度增加时防止梯度消失。

3.2.9 权重初始化

如果全部设成同样的非零值, 在反向传播中所有权重都会进行相同更新为相同值, 并拥有对称值, 不管进行怎样的迭代都不会打破这种对称性, 隐藏层就像只有一个神经元, 无法实现神经网络的表达能力, 为打破权重的对称结构, 必须随机生成初始值。

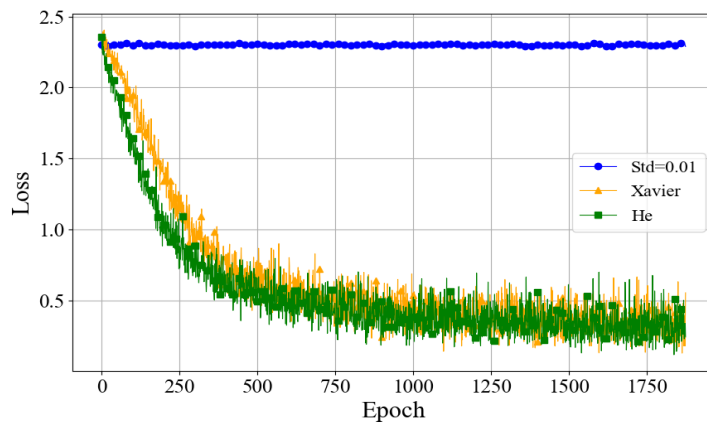


图 3-8 不同权重初始值的比较

从图 3-8 结果可知, Std=0.01 时完全无法进行学习(对应图中 Loss 值无变化), 正向传播中传递值很小(集中在 0 附近数据), 因此逆向传播时得到的梯度也很小, 权重几乎不进行更新, 当权重初始化为 He(权重从均值为 0, 标准差为 $\sqrt{\frac{2}{n_{in}}}$ 的正态分布中采样, 其中 n_{in} 是输入单元的个数)和 Xavier(权重从均值为 0, 标准差为 $\sqrt{\frac{2}{n_{in}+n_{out}}}$ 的正态分布中采样, 其中 n_{out} 是输出单元的个数^[31])初始化时, 学习进行得顺利, 并发现当激活函数使用 ReLU 或 Leaky ReLU 时, 权重初始值使用 He 初始化学习进度更快, 当激活函数为 Sigmoid 或 Tanh 等 S 型曲线函数时, 初始化使用 Xavier 初始值学习进度更快。

3.2.10 实验平台及超参数配置

所有实验均在统一的服务器上面完成, 实验平台配置如表 3-2:

表 3-2 实验平台

实验设备	型号
设备	戴尔工作站
平台	Windows10
CPU	Intel(R) Core(TM) i7-12400F
GPU	NVIDIA A6000
System memory	128GB
SSD	Western Digital 580 3TB

模型训练前设置参数称为超参数, 先设置初始值, 在训练过程中初始值通过优化算法自动调整, 影响模型结构、训练过程从而显著影响模型性能和效率, 进行有效调节可提升模型训练速度、模型泛化能力、减少过拟合, 最终得到超参数配置如表 3-3 所示。

表 3-3 参数与超参数配置

参数与超参数	数值
输入图片尺寸	224×224×3
批量大小	128
训练周期	100
高斯卷积核	19×19
指数衰减因子	(0.9, 0.999)
权重衰减	0.001
稳定常数	1×10^{-8}
学习率	0.01
迭代	100
动量参数	0.9
训练集：验证集	8:2
优化器	Adam
标准化系数Beta1	0.9
标准化系数Beta2	0.999

(1)批量大小(Batch Size =128): 批量过小可能导致训练过程噪声化并导致模型不稳定, 批量过大可能会导致显存不足、计算效率下降, 批量大小决定深度学习训练过程中完成每个周期所需时间和每次迭代(iteration)之间梯度的平滑程度, 批量大小为 b , 样本总数为 n , 每个周期所需的迭代次数(正向+反向)为 n/b ^[32], 完成每个周期所需时间也随着迭代次数增加而增加, 模型收敛快慢取决于梯度方向和更新次数, 批量大小在 6~256(通常在 32、64、128 之间)能够花费最少时间, 并且得到高精度^[33]。

较大批量有助于提高计算效率, 但会增加内存消耗, 选择 128 是在考虑内存和训练稳定性的平衡后作出的决定, 在大多数任务上提供稳定训练结果。

(2)高斯卷积核 (Gaussian Kernel Size=19×19): 用于图像滤波、特征提取和降噪操作, 卷积核大小决定对输入图像局部区域的感知范围, 19×19 是相对较大的卷积核, 能够捕捉到更广泛上下文信息从而帮助模型提取更复杂的特征^[34]。

卷积核大小通常是一个奇数, 如 3×3、5×5、7×7、19×19 等, 这与步长有关, 在特定降噪任务或需要更大感受野的任务中, 需要选择更大的核, 尤其是在图像分辨率较高的

情况下,较大卷积核会增加计算量,但能够提取更大范围的上下文信息^[35],所以在实验条件允许范围内,选择 19×19 能够在较大的图像区域中提取相关特征,这里选择高斯卷积核尺寸为 19×19 。

(3)权重衰减(采用 L2 正则化, $\text{Weight Decay}=0.001$): 是用于控制模型复杂度、避免过拟合的方法,较大权重值可能导致模型过拟合,加入权重衰减能够对大权重进行惩罚来防止过拟合。

对于深度学习任务, $0.0001 \sim 0.01$ 范围的权重衰减值表现良好, 0.001 是常见默认值,能有效地减少过拟合,同时不影响模型学习能力^[36],这里选择 0.001 。

(4)动量参数($\text{Momentum}=0.9$): 优化器中用于加速梯度下降过程并减少震荡,会在更新参数时考虑到之前梯度的加权平均,动量让梯度下降更加平滑和稳定,从而加速收敛。

常见动量值在 0.8 到 0.99 之间,动量值设定为 0.9 能够提供较好平衡,避免震荡并加速收敛,在大量任务中已证明其有效性^[37]。

(5)标准化系数 $\text{Beta1}=0.9$: 是 Adam 优化器中的超参数,控制模型更新时动量衰减速率,在更新时会根据一阶矩估计(均值)调整参数, Beta1 控制历史梯度衰减率。

$\text{Beta1}=0.9$ 表示用 90% 的历史梯度来平滑当前的梯度估计,有助于在优化过程中避免梯度波动过大,通常 Beta1 设置为 0.9 既能够提供良好的训练稳定性^[38],并且不会过度依赖历史梯度,不影响训练过程收敛性,这里选择 Beta1 为 0.9 。

(6)标准化系数 $\text{Beta2}=0.999$: 是 Adam 优化器中第二个超参数, Beta2 控制模型更新时二阶矩估计(梯度的平方)衰减速率, Beta2 决定梯度平方的历史衰减速率^[39]。

$\text{Beta2}=0.999$ 表示梯度平方非常平滑,对较大梯度影响较小,较大的 Beta2 值使得优化过程更加稳定,尤其是在梯度较小或稀疏的情况下, $\text{Beta2}=0.999$ 是 Adam 优化器默认值^[40],在多数任务中都表现良好,较大 Beta2 值会减少梯度更新时的波动,使模型能够稳定地收敛,这里选择 Beta2 为 0.999 。

3.2.11 网络训练

为防止在训练过程中出现梯度爆炸,先使用较小学习率训练模型,然后再由参数更新到最佳的学习率^[41]。

(1)反向传播算法: 简称 BP 算法,是深度学习中神经网络训练重要优化方法,适合于多层神经网络,其包括前向传播、计算误差、反向传播、参数更新等操作^[42],建立在

梯度下降法的基础上，其核心思想是通过计算误差梯度，并利用这些梯度信息来更新网络权重和偏置项，从而最小化预测结果与实际值之间差距。

(2)前向传播：网络学习在权重修改过程中完成，误差达到期望值时网络学习结束，在前向传播过程中，输入数据经过神经网络逐层处理，最终得到输出结果^[43]，公式如下：

$$y = f(x_i \cdot w_i + b) \quad (3-15)$$

式中： $f(x)$ 是激活函数，如 GELU 或 ReLU 等^[44]。

(3)计算误差：采用损失函数，得到值是评估模型预测与真实值偏离程度指标，值越小表示模型性能越好，常见损失函数如均方差函数用于回归任务^[45]，本文的视网膜病变识别是多分类问题，而交叉熵函数和对数似然损失函数用于分类任务，对数似然损失函数用于文本分类，本实验是图片分类，故采用交叉熵损失函数作为训练过程的损失函数，在训练神经网络时，通过计算损失函数并利用梯度下降法更新权重，逐步优化模型，交叉熵损失函数 Loss 公式如下^[46]：

$$Loss = \frac{1}{N} \sum_i L_i = -\frac{1}{N} \sum_i \sum_{c=1}^M y_{ic} \log(p_{ic}) \quad (3-16)$$

式中： M 为类别的数量； y_{ic} 为符号函数； p_{ic} 为类别第 i 层属于类别 c 的预测概率。

(4)反向传播：从后向前传播的方法，反向传播时后一层节点会与前一层多个节点相连，因此需要对所有节点误差求和，计算完误差后需要将误差向不断向前一层传播，因为不同神经元重要性不同，回传时需要考虑权重系数，将误差从输出层逐层向输入层反向传播，并根据链式法则计算每个参数的梯度，具体步骤如下：

(a)误差反向传播：从输出层开始，将误差传递到隐藏层和输入层；

(b)梯度计算：利用链式法则计算各层参数的梯度，对于一个三层神经网络表示为^[47]：

$$\frac{\partial Loss}{\partial w_i} = \sum_{k=1}^M \left(\frac{\partial L}{\partial a_k} \right) \left(\frac{\partial a_k}{\partial z_k} \right) \left(\frac{\partial z_k}{\partial w_i} \right) \quad (3-17)$$

式中： w_i 是第 i 层权重矩阵； a_k 是第 k 个节点激活值； z_k 是第 k 个节点输入值。

3.2.12 参数更新

根据得到的梯度，用梯度下降更新网络中权重和偏置项，更新公式如下：

$$w_i^{new} = w_i^{old} - \alpha \cdot \frac{\partial L}{\partial w_i} \quad (3-18)$$

式中： α 是学习率，用梯度下降法做参数更新时，模型学习速度取决于 α 和偏导值，其中学习率是需要设置的超参数，重点关注偏导值，偏导值大小取决于 x_i 和 $\frac{\partial L}{\partial w_i}$ ，结合交叉熵当损失函数，在模型效果差的时候学习速度变快^[48]。

3.2.13 优化算法

优化算法直接影响模型训练效率和效果，好的优化算法能够加速收敛过程，提高模型性能，并避免陷入局部最优解，显著提升模型准确性和稳定性。

(1)动量法：通过引入动量概念，旨在加快收敛速度并减少震荡，结合当前和之前梯度累积效果，在梯度更新时加速收敛，从而在梯度一致的方向上加快学习速度。

动量的更新公式为：

$$v_{t+1} = \gamma v_t + \alpha \nabla_{\theta} L(\theta_t) \quad (3-19)$$

参数更新公式：

$$\theta_{t+1} = \theta_t - v_{t+1} \quad (3-20)$$

式中： $L(\theta_t)$ 是损失函数； $\nabla_{\theta} L(\theta_t)$ 是损失函数相对于参数的梯度； θ_t 是第 t 次迭代时的参数值； α 是学习率； v_t 是动量项，表示之前梯度的指数加权平均； γ 是动量系数，通常在0.9左右^[49]，动量法通过在梯度更新时保留之前的梯度信息，即每次更新不仅依赖于当前梯度，还依赖于之前的梯度累计。

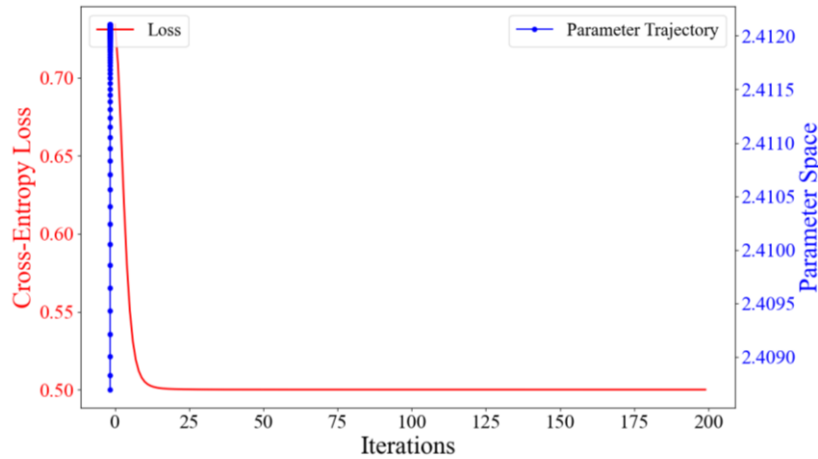


图 3-9 动量法

图 3-9 展示动量法处理结果，红色为损失随迭代次数变化，蓝色为参数更新轨迹。

(2)RMSprop：旨在解决动量法后期学习率过小问题，使用指数加权平均来计算梯度平方和，计算梯度平方和的指数加权平均来防止学习率过快衰减，使得优化算法能够在

训练后期仍然保持有效学习率^[50]，梯度平方和的移动平均公式为：

$$E[g^2]_t = \gamma E[g^2]_{t-1} + (1-\gamma) \nabla_{\theta} L(\theta_t) \odot \nabla_{\theta} L(\theta_t) \quad (3-21)$$

参数更新公式：

$$\theta_{t+1} = \theta_t - \frac{\alpha}{\sqrt{E[g^2]_t + \epsilon}} \odot \nabla_{\theta} L(\theta_t) \quad (3-22)$$

式中： $E[g^2]_t$ 是梯度平方和的指数加权平均； $\odot$ 表示逐元素相乘； ϵ 为稳定常常数，值较小且通常在 1×10^{-8} 左右，防止除零错误^[51]。

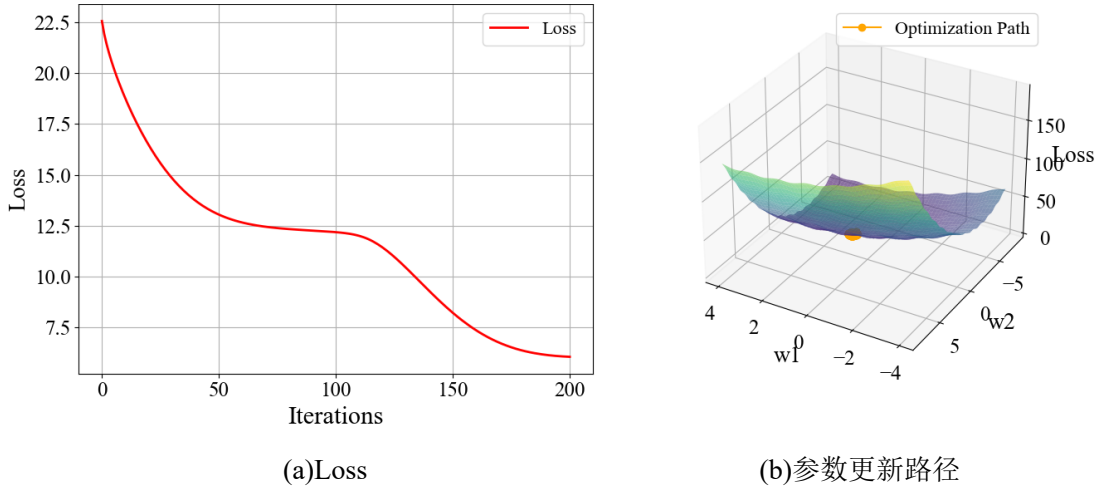


图 3-10 RMSprop

图 3-10(a)展示损失函数收敛曲线，红色线条表示损失值如何随着迭代次数逐渐减小，(b)图展示在三维空间中展示损失函数的曲面以及优化过程中参数更新路径，橙色路径代表优化算法最佳值。

(3)Adam 优化器：结合动量法的加速收敛和 RMSprop 自适应学习率优点，自适应调整参数学习率，通过对梯度的均值和方差进行偏差修正，进一步提高算法稳定性和收敛速度，同时计算梯度的一阶动量（均值）和二阶动量（非中心化方差），并更新参数^[52]。

一阶动量：

$$m_t = \beta_1 m_{t-1} + (1-\beta_1) \nabla_{\theta} L(\theta_t) \quad (3-23)$$

二阶动量：

$$v_t = \beta_2 v_{t-1} + (1-\beta_2) \nabla_{\theta} L(\theta_t) \odot \nabla_{\theta} L(\theta_t) \quad (3-24)$$

偏差校正：

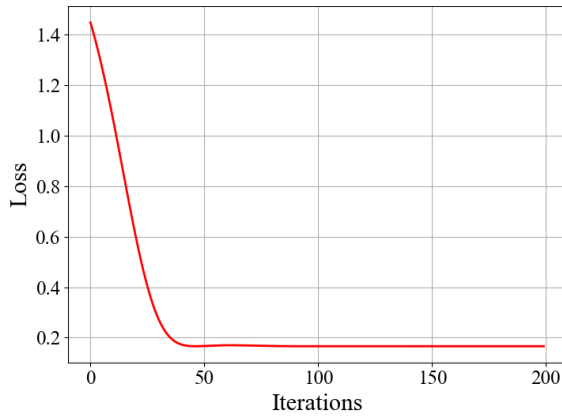
$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t} \quad (3-25)$$

$$\hat{v}_t = \frac{v_t}{1 - \beta_2^t} \quad (3-26)$$

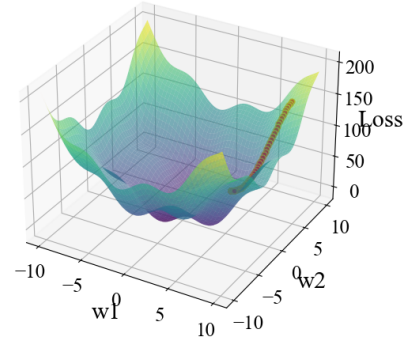
参数更新:

$$\theta_{t+1} = \theta_t - \frac{\alpha}{\sqrt{\hat{v}_t} + \epsilon} \cdot \hat{m}_t \quad (3-27)$$

式中: β 是动量系数^[53]。



(a) Loss



(b) 参数更新路径

图 3-11 Adam

图 3-11(a)展示损失函数随迭代次数的下降过程, 将损失降低到小于 0.2 范围内且较稳定, (b)展示在三维参数空间优化路径, 直观看到优化器在搜索空间中寻找最优解过程。

3.3 留一法交叉验证

使用交叉验证减少单次划分带来的偶然性, 使评估结果更加可靠, 同时也充分利用数据, 所有数据都参与模型训练和测试, 减少过拟合^[54]。

留一法交叉验证将数据集中每个样本都作为一次验证集, 剩余样本作为训练集, 如果有 N 个样本, 模型将训练 N 次, 每次用 $N-1$ 个样本训练, 1 个样本验证, 没有数据丢失, 评估结果更加精确, 每次验证误差为^[55]:

$$E_i = y_{i,true} - y_{i,predicted} \quad (3-28)$$

总误差为:

$$E_{LOO-CV} = \frac{1}{N} \sum_{i=1}^N E_i \quad (3-29)$$

每次训练将数据集分为训练和验证集，验证误差为所有单次误差平均值^[56]。

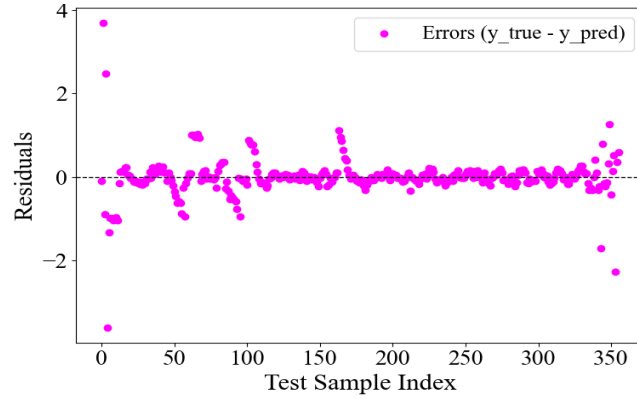


图 3-12 留一法交叉验证（LOO-CV）散点图

图 3-12 中 X 轴表示测试样本索引，每个点对应交叉验证中一个测试样本，Y 轴表示残差值，即每个测试样本实值与预测值之间差异，残差越接近 0 说明模型预测越准确^[57]。从图 3-12 可以看出，残差分布在 0 附近且上下波动较小，说明模型预测较为准确，通过观察残差图直观了解模型预测误差不存在某种明显趋势，说明模型不存在偏差或方差问题，模型效果较好。

在留一法交叉验证（LOO-CV）中，Mean Squared Error of LOO-CV = 0.032，模型预测结果与真实标签之间误差平方和的平均值为 0.032，值越小表示模型预测结果与真实值越接近，本实验 MSE 数值较小，线性回归模型在数据集上表现良好，预测结果与实际分类标签之间误差小。

3.4 SF-Unet 分割评价标准

实验采用 DICE 系数、交并比、像素精度和类别平均像素精度来评估模型性能，假定有 $k+1$ 类， p_{ij} 为属于 i 类预测为 j 类像素点总数， p_{ij} 为 TP， p_{ij} 为 FP， p_{ji} 为 FN，则有：

(1) DICE 系数：用来衡量分割结果和真实标签重叠程度，计算公式如下：

$$L_{dice} = 1 - \frac{2 \sum_{x \in \Omega} p_i(x) g_i(x)}{\sum_{x \in \Omega} p_i^2(x) + \sum_{x \in \Omega} g_i^2(x)} \quad (3-30)$$

式中： $p_i(x)$ 为像素属于背景的概率； $g_i(x)$ 为像素属于前景的概率，DICE 范围从 0 到 1，值越高表示重叠程度越高，分割效果越好。

(2) 交并比(IoU)：计算预测和真实区域重叠程度，帮助了解模型在每个类别上分割效果，高 IoU 值表示预测与实际重叠较多，模型预测较准确，低 IoU 值反之，公式如下：

$$IoU = \frac{Intersection}{Union} \quad (3-31)$$

(3)像素精度：标记正确像素占总像素的百分比，公式如下：

$$PA = \frac{\sum_{i=0}^k p_{ii}}{\sum_{i=0}^k \sum_{j=0}^k p_{ij}} \quad (3-32)$$

(4)类别平均像素精度：每类分类正确和该类所有像素点数比例再求平均，公式如下：

$$MPA = \frac{1}{(k+1)} \sum_{i=0}^k \frac{p_{ii}}{\sum_{j=0}^k p_{ij}} \quad (3-33)$$

3.5 SF-Unet 分割实验结果

表 3-4 不同分割网络实验结果对比

指标	Unet	CS-Unet	CSP-Unet	Attention-Unet	SF-Unet
DICE	0.8910	0.8922	0.8928	0.9114	0.9327
PA	0.8793	0.8824	0.8951	0.8904	0.9439
MPA	0.8618	0.8766	0.9194	0.9151	0.9610
MIOU	0.8756	0.8867	0.9266	0.9322	0.9742

表 3-4 结果展示 SF-Unet 的性能，在同样的视网膜病变图像分割任务中，SF-Unet 模型 DICE 较其他网络提升 2.34%~4.68%，PA 较其他网络提升 5.81%~7.34%，MPA 较其他网络提升 4.52%~11.51%，MIOU 较其他网络提升 4.50%~17.12%，从上述计算可以看出，SF-Unet 在所有评估指标上表现出色，特别是在平均交并比(MIOU)和平均像素准确率(MPA)方面，SF-Unet 相比 Unet 分别提高 17.12%和 11.51%，表明 SF-Unet 在处理图像分割任务具有更高精度和更好泛化能力。

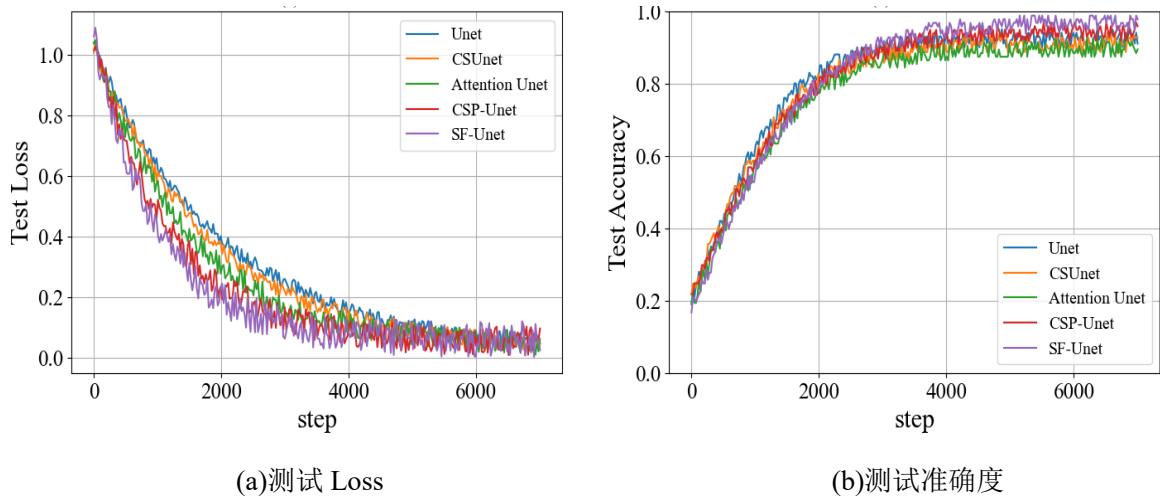


图 3-13 不同分割模型的训练过程

比较不同卷积神经网络在视网膜病变图像的分割效果，得到不同网络测试精度，如

图 3-13(a)和(b)所示，可以看出，SF-Unet 在学习过程中更快更好地到最佳状态。

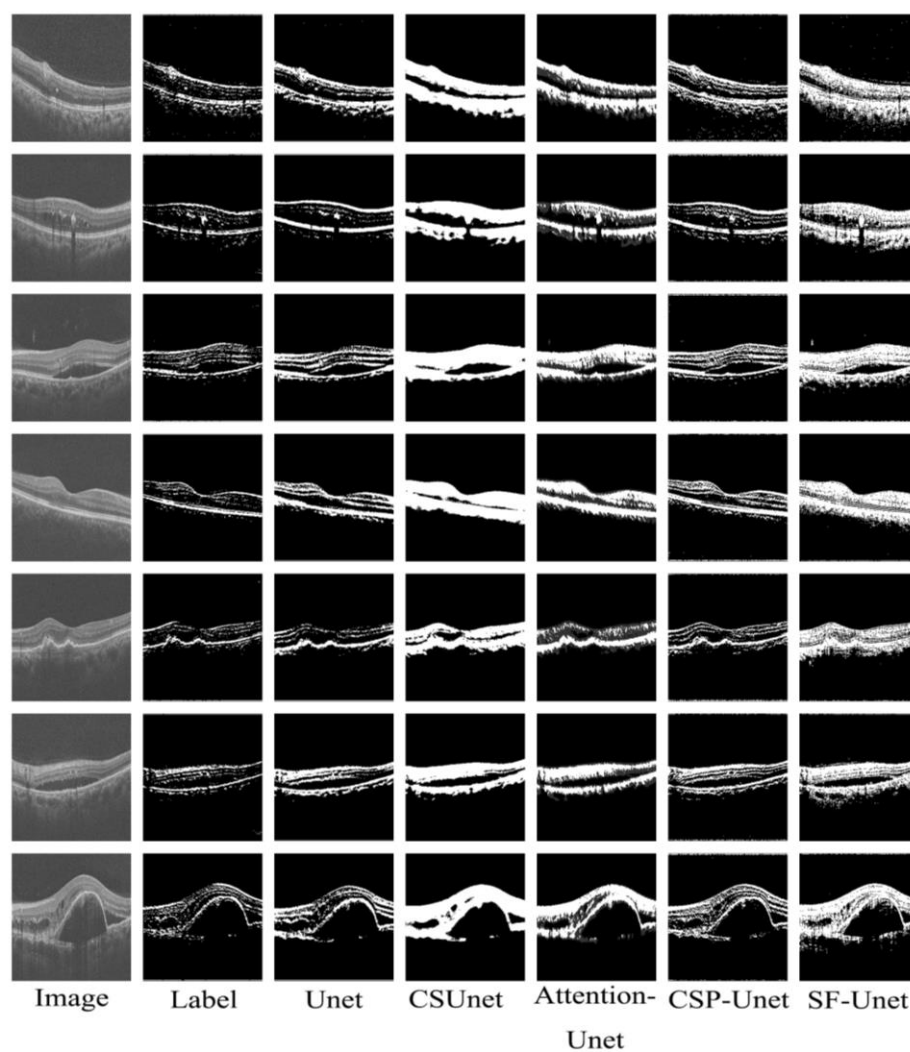


图 3-14 不同分割网络的测试结果

从图 3-14 看出，在复杂分割场景中 CS-Unet 和 Attention-Unet 出现过分割，而 Unet 和 CSP-Unet 出现欠分割，总体而言 SF-Unet 在视网膜病变图像分割任务展现更高性能。

第4章 Sets-Resnet 网络

4.1 Sets-Resnet 介绍

根据图 4-1 网络训练过程，考虑将网络结构中的所有分支设计为相同拓扑结构，采用 ResNet 重复层策略，以简单可扩展方式通过相同的拓扑结构实现求和，可扩展到大量计算，无需专门设计，将 ResNet 中高维特征图分组为多个相同低维特征图，在卷积操作后通过相同拓扑结构将多组结构进行求和，定义网络为 Sets-ResNet，以简单可扩展方式延续 Split-Transform-Merge 结构^[58]，评估不同组时，模型复杂性和参数数量相同情况下，修改某种参数最小化对其他参数修改。

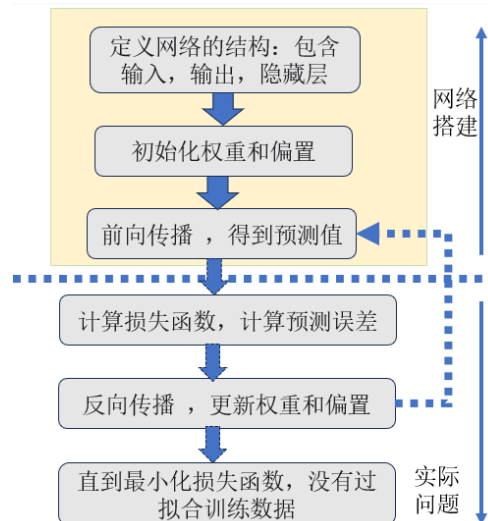


图 4-1 分类网络训练过程

4.1.1 Bottleneck(BTNK)残差块结构

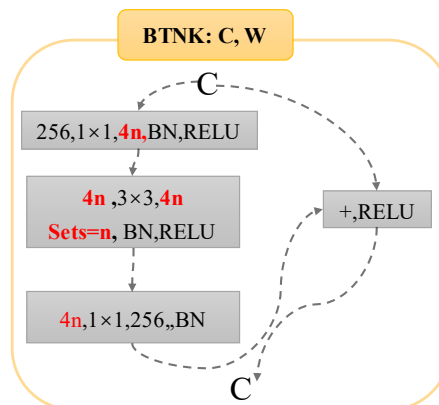


图 4-2 残差块结构

BTNK 先通过第一个 1×1 卷积来降低输入的通道数,使中间卷积通道数减少为 $1/4$, 3×3 卷积核用来扩大感受野并减少特征图尺寸,再次采用 1×1 卷积核来恢复通道数,如图 4-2 所示,第一个 1×1 卷积核使中间 3×3 卷积层输入和输出通道数都减小,效率更高,再次使用 1×1 卷积核来改变维度,两个 1×1 卷积有效减少卷积参数量和计算量,并且两次使用 1×1 卷积核使上一个 BTNK 输出通道数等于下一个 BTNK 输入通道数,使该输出形状仍为 (C,W,W) ,即 BTNK 对应输入与输出通道数相同。

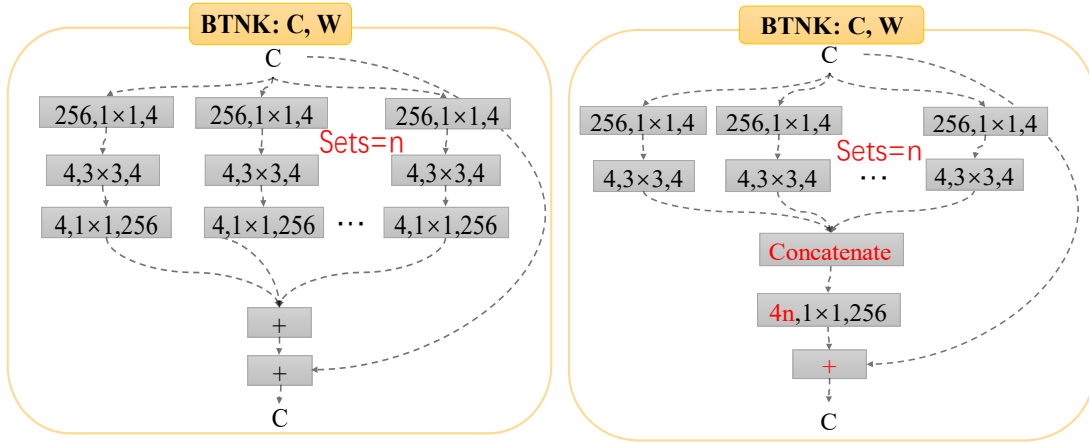


图 4-3 残差块其他不同结构

如图 4-3 所示,把串联结构改成并联结构,是不加深网络并且可以增加模型准确度的有效方法,避免神经网络随着深度增加,导致训练复杂性增加,需要更多计算资源和时间来训练模型,同时也对算法设计和调整提出更高要求,深度神经网络具有丰富参数,网络过深使得容易过拟合训练数据,导致模型在测试集上性能下降,即使在有充足数据情况下,也需要注意防止过拟合,BTNK 有可变参数 C 和 W ,令 (C,W,W) 的输入为 x_i ,BTNK 每组 3 个卷积块为函数 $H(x_i)$,相加可得 $T(x_i)=H(x_i)+x_i$,得到 BTNK 输出:

$$F(x) = \sum_{i=1}^{Sets} T(x_i) \quad (4-1)$$

式中: $Sets=n$ 代表组数目,本质上仍是 Split-Transform-Merge (分割-变换-聚合) 过程: (1)Split: 将向量 x_i 分成低维嵌入(由 1×1 卷积降维),256 个通道变成 4 个通道; (2)Transform: 3×3 卷积对每个低维特征进行变换; (3)Merge: 得到特征图并进行聚合。

对于 RELU 的各种改进方案在不同网络中性能不稳定,没有能够在大部分网络中都能够表现的比 RELU 好且形式简单的函数,除非是专门对激活函数进行研究,一般都选用形式最简单且性能相对稳定的 RELU 作为激活函数,所以在 Sets-Resnet 网络中还是采用 RELU 激活函数。

4.1.2 优化算法

AdaGrad 是早期的自适应学习率优化算法，针对每个参数的学习率进行独立递减，局限性在于学习率会累积过多历史信息，导致学习率过小，难以在后期迭代中继续有效更新参数^[59]；RMSprop 是 AdaGrad 改进版本，主要解决 AdaGrad 学习率递减过快的问题，通过对每个参数的梯度平方进行指数加权平均来调整学习率^[60]；Adam 是 RMSprop 的升级版，这里采用自适应学习率优化器 Adam，用动量来优化随机梯度下降，具有快速收敛、高效稳定等优点。

4.1.3 Sets-Resnet50 网络结构配置

表 4-1 网络结构配置

Stage	Output	Resnet50	Sets-Resnet50	关键改进
Conv1	112×112	7×7, 64, ST=2	7×7, 64, ST=2	初始层无改进，保持输入特征提取能力
		3×3 max pool, ST=2	3×3 max pool, ST=2	池化层未改动，降低特征图尺寸
Conv2	56×56	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 2 \\ 3 \times 3, 2, Sets = 32 \\ 1 \times 1, 8 \end{bmatrix} \times 3$	分组卷积：每组通道数降为 2（原 64），总通道数=32 组 × 2=64，与原中间层一致
Conv3	28×28	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 4 \\ 3 \times 3, 4, Sets = 32 \\ 1 \times 1, 16 \end{bmatrix} \times 4$	每组通道数 4（原 128），总通道数=32×4=128，保持特征维度
Conv4	14×14	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1, 8 \\ 3 \times 3, 8, Sets = 32 \\ 1 \times 1, 32 \end{bmatrix} \times 6$	每组通道数 8（原 256），总通道数=32×8=256，维持信息容量
Conv5	7×7	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 16 \\ 3 \times 3, 16, Sets = 32 \\ 1 \times 1, 64 \end{bmatrix} \times 3$	每组通道数 16（原 512），总通道数=32×16=512，平衡深度与计算效率

注意：50 指卷积层+全连接层的层数，输入为图片维度(64,3,224,224)。

4.1.4 核心改进机制

(1)分组卷积(Sets=32): 每个残差块卷积被分解为 32 个并行组, 单组通道数减少。

(2)总通道数不变: 通过组数(Sets)与单组通道数的乘积保持与原版一致(如 Conv2 中 32 组 $\times$ 2 通道=64 通道)。

(3)参数量(Parameters)减少: 参数量计算公式:

$$\text{参数量} = \text{卷积核尺寸} \times \text{输入通道数} \times \text{输出通道数} \quad (4-2)$$

ResNet50 总参数量(25.6M): 1×1 卷积($64 \rightarrow 64$): $1^2 \times 64 \times 64 = 4,096$; 3×3 卷积($64 \rightarrow 64$): $3^2 \times 64 \times 64 = 36,864$; 1×1 卷积($64 \rightarrow 256$): $1^2 \times 64 \times 256 = 16,384$; 单块参数量: $4,096 + 36,864 + 16,384 = 57,344$; Conv2 阶段 (3 个残差块): $57,344 \times 3 = 172,032$; 其他阶段同理: 累加所有残差块(3+4+6+3)及全连接层 (1M 参数), 总参数量为 25.6M。

Sets-Resnet50 总参数量(7.3M): 1×1 卷积($2 \rightarrow 2$): $1^2 \times 2 \times 2 = 4$ (每组), 32 组总参数量 = $4 \times 32 = 128$; 3×3 卷积($2 \rightarrow 2$): $3^2 \times 2 \times 2 = 36$ (每组), 32 组总参数量 = $36 \times 32 = 1,152$; 1×1 卷积($2 \rightarrow 8$): $1^2 \times 2 \times 8 = 16$ (每组), 32 组总参数量 = $16 \times 32 = 512$; 单块参数量: $128 + 1,152 + 512 = 1,792$; Conv2 阶段 (3 个残差块): $1,792 \times 3 = 5,376$; Conv3 (4 个残差块): 单块参数量 = $32 \times (4 + 36 + 64) = 3,328$; 总参数量 = $3,328 \times 4 = 13,312$; Conv4 (6 个残差块): 每组通道数 = 8; 单块参数量 = $32 \times (64 + 576 + 256) = 28,672$; 总参数量 = $28,672 \times 6 = 172,032$; Conv5 (3 块): 每组通道数 = 16; 单块参数量 = $32 \times (256 + 2,304 + 1,024) = 114,688$; 总参数量 = $114,688 \times 3 = 344,064$

总计: 改进网络各阶段参数量累加 (含初始卷积与池化层) 约为 7.3M, 相比原版减少 71.5%。

4.2 Sets-Resnet50 网络分类实验结果

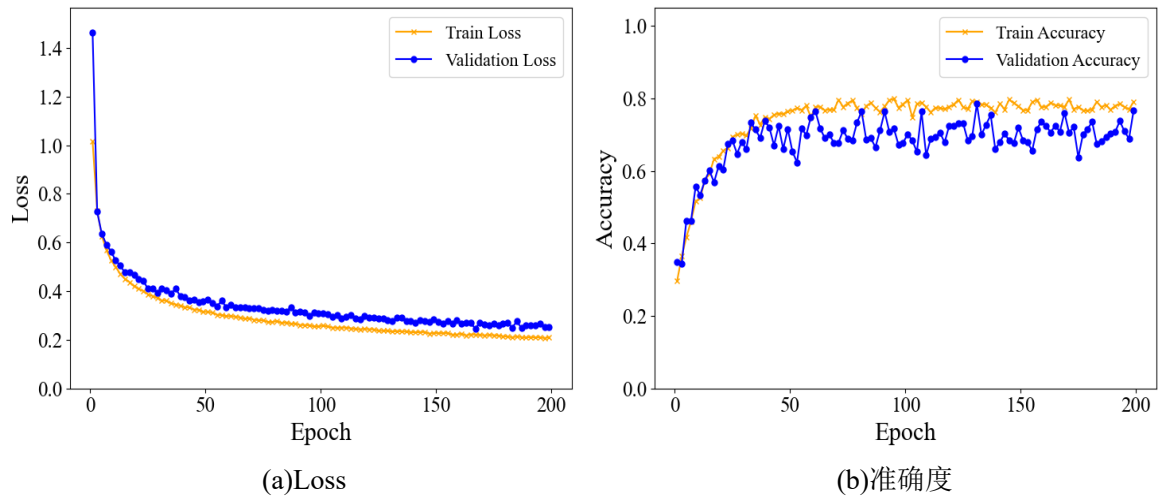


图 4-4 Sets-Resnet50 分类网络训练和验证过程

图 4-4(a)显示训练损失和验证损失随着训练周期(Epoch)的变化情况,训练损失在初始阶段迅速下降,随后趋于平稳,而验证损失在整个过程中也呈现出类似的趋势,但下降速度相对较慢,训练损失在约 20 个轮次后稳定在 0.2 左右,而验证损失在 50 个轮次后稳定在 0.3 左右,训练损失(0.2)与验证损失(0.3)均较低且差距较小,表明模型泛化能力良好,未出现明显过拟合;(b)图展示训练准确率和验证准确率随训练轮数变化,均在开始时迅速上升后增长速率降低后趋于稳定,训练准确率在 20 个轮次后稳定在 0.8 左右,而验证准确率则在 50 个轮次后稳定在 0.7 左右,验证准确率(70%)低于训练准确率(80%),但差距在合理范围内(<10%),说明模型对未见数据的适应能力较强,模型训练损失和验证损失最终都逐渐降低且趋于稳定且数值较低,说明模型性能较好,训练准确率和验证准确率都逐渐提高趋于稳定且数值较高,进一步证明模型的准确性较高。

第5章 SetsA-Resnet 网络

5.1 SetsA-Resnet 网络介绍

SetsA-Resnet Block 通过增加 SetsA-Block 和分组操作(Sets-Again), 提出分组注意力机制, 采用两次并行多分支网络, 每个网络的每个分支、卷积、池化后的尺寸均相同, 一方面增加网络宽度, 另一方面增加网络尺度适应性, 网络卷积层中的网络能提取输入的每一个信息, 3×3 滤波器能够覆盖大部分接受层输入, 进行一个池化操作 (在 SetsA-Block 里进行), 以减少空间大小并降低过拟合, 在每一个卷积层后都进行 ReLU 操作以增加网络非线性, 模型在处理不同特征图时能更关注关键特征, 抑制无关或冗余信息, 提升模型表示能力和性能从而提高网络准确度, 让网络在每层中学习更多特征, 更多通道数让每一层能够捕捉到更丰富、更细致特征从而提高网络表达能力, 使网络能够更多地融合不同通道, 捕捉多尺度和多类型特征以及输入数据中的复杂细节, 更宽的网络能够学习到更多的特征变体和细节, 这种多样性使得网络在面对不同的输入数据时更具鲁棒性, 提升模型的表示和泛化能力, 拥有更高的准确度, 从而, 在处理复杂视觉任务时能够表现出更优异的性能。

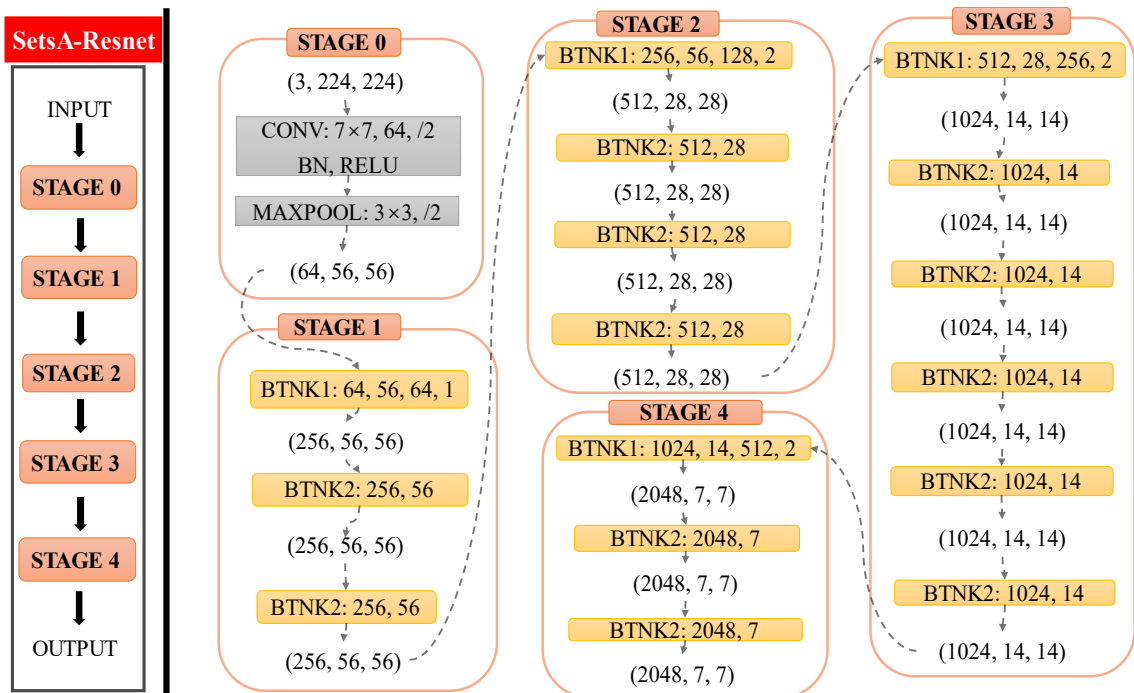


图 5-1 Sets-Resnet 网络结构

5.2 STAGE0 阶段处理

如图 5-1 所示，模型输入先通过 STAGE0 阶段处理，减少输入空间尺寸并增加特征通道数，输入图像(尺寸为 $224 \times 224 \times 3$)经过 7×7 卷积层，输出特征图尺寸为 $112 \times 112 \times 64$ ，再经过 3×3 的最大池化层，特征图尺寸进一步减小到 $56 \times 56 \times 64$ ，STAGE0 阶段处理将输入图像转化为低分辨率、高维度特征图。

5.3 组的划分

特征可按通道维度 C 分为若干组，特征图组数由 Sets 组数 S 给出，为组的一次划分，需另引入基数组 SetsA 超参数 A 表示 SetsA 组数，为基数组内 Sets 的二次划分，特征组总数为 $A \cdot S$ 。

5.4 SetsA-Resnet Block 结构

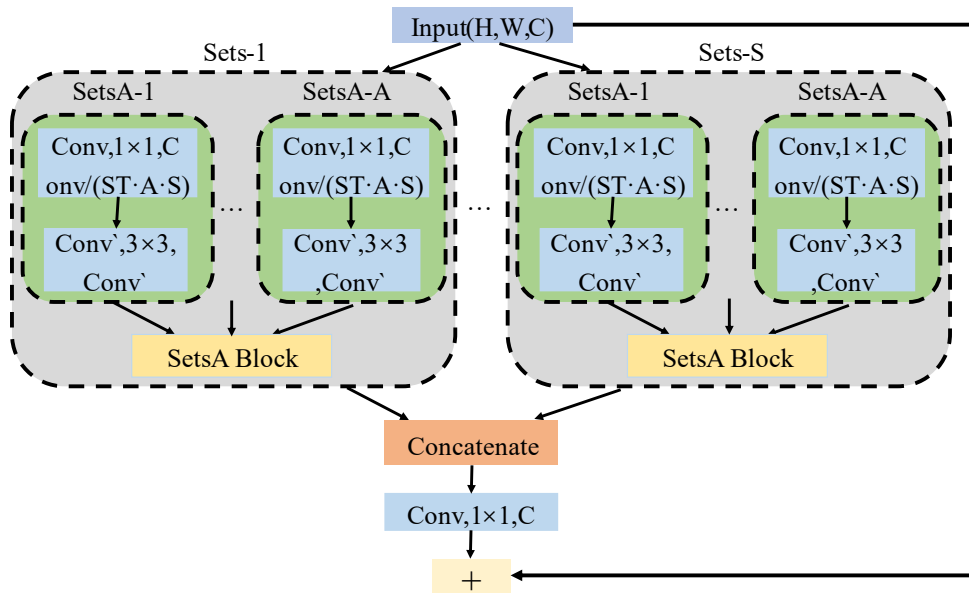


图 5-2 SetsA-Resnet Block 结构

图 5-2 为通过改进两次 Resnet 网络得到的 SetsA-Resnet 网络，创建 Sets-Again 结构以提高网络的特征表示能力，输入特征图尺寸为 (H, W, C) ， H, W, C 分别代表高度、宽度和通道数，输入特征图首先被划分为 S 个 Sets 分支，每个 Sets 分支包含 A 个 SetsA，每个 Sets 分支处理输入一部分特征图，在每个 Sets 分支内，特征图继续被划分为 A 个 SetsA，每个 SetsA 都先经过一个 1×1 卷积层，将通道数压缩到 $ST \cdot A \cdot S$ ，其中 ST 为步

幅,接着通过一个 3×3 卷积层,保持通道数不变,每个 SetsA 之后都经过一个称为 SetsA Block 的模块,用于对各个 SetsA 进行分组, SetsA Block 是 SetsA-ResNet 的关键特点之一,这个模块通过学习每个 SetsA 重要性系数来重新加权各个 SetsA 输出,进而使得网络在处理具有多种特征的场景时更加灵活,加权后的特征图在 SetsA 内进行融合,处理完 S 个 Sets 分支后特征图大小相同,均为 $\text{Conv}/(\text{ST} \cdot A \cdot S)$ 。

深度学习中的特征融合方式主要包括 Concatenate 和 Add 两种操作:(1)Concatenate 操作将两个或多个相同维度张量在某个维度上连接在一起,生成一个更大的张量,虽然图像特征数增加,在每一个特征下的信息都没有发生变化实现特征在纵向空间上叠加融合,用于增加通道数,以便在后续层中更好地捕获不同特征之间关系^[61];(2)Add 操作将两个相同形状的张量逐元素相加在一起,生成一个新的张量,此时特征图的维度并没有发生变化,只在每一个维度下的信息量在不断增加,用于实现跳跃连接或残差连接,以促进梯度流动,提高训练稳定性,以及在深层网络中减少梯度消失问题^[62]。

输入特征图首先经过 1×1 卷积层以压缩通道数,减少后续计算量,通道数减少到输入的 $1/4$,再经过一个 3×3 卷积层,卷积核数量和输出特征图的通道数保持不变,在压缩后的特征图上提取空间特征,再通过一个 1×1 卷积层,通道数恢复到原始输入通道数大小,准备与残差连接分支相加,再计算残差连接分支,如果输入和输出的通道数和步幅相同,则残差连接分支是输入,如果不相同,则需要通过 1×1 卷积匹配通道数和空间维度,添加残差链接有效缓解梯度消失问题,需要将来自不同层次的特征合并到一起,这里选择对所有输出进行拼接(Concat),拼接后的特征图通过一个 1×1 卷积层,用于再次调整通道数,回到原始通道数 C ,经过 1×1 卷积后的输出与输入进行加权,完成残差连接最终输出的特征图尺寸仍为 (H, W, C) ,设主路径得到的函数为 $F(x)$,是包含多个 x 的多项式 $w_1x_1 + w_2x_2 + \dots$,残差连接为 x ,链式求导、反向传输的结果就变成 $w_1 + w_2 + \dots + 1$,有效缓解梯度消失问题,最后将主路径的输出与残差连接分支相加,通过 RELU 激活函数输出, SetsA-Resnet Block 的最终输出 Y 由残差连接生成,即为:

$$Y = F(x) + x \quad (5-1)$$

5.4.1 SetsA-Block

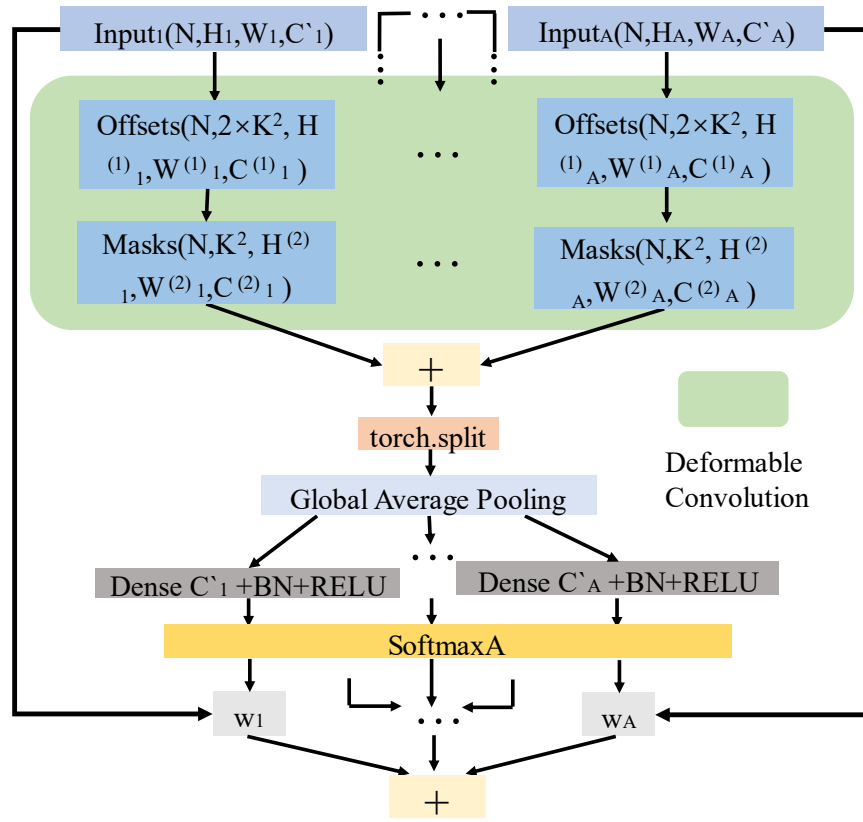


图 5-3 SetsA-Block 结构

图 5-3 展示的 SetsA-Block (Sets Again-Block, 再次分组块) 是整个模块核心, 基于输入通道数和配置参数 (如卷积核大小、步幅、填充等) 来构建注意力机制, 首先引入可变形卷积(DCN)来学习偏移量调整卷积核的采样位置, 以适应空间变形结构, 进一步增强模型对复杂结构的处理能力, 再对不同路径的输入进行加权处理, 并在 $\text{SetsA} > 1$ 时对特征图进行分割, 利用全局平均池化(GAP)提取全局特征信息, 特征经过全连接层 Dense1 进行低维映射, 使用 RELU 激活函数计算每个输入特征权重, 再通过 Dense2 恢复原始维度, 每个输入特征权重通过 SoftmaxA 进行归一化生成注意力权重, 得到基于全局上下文信息权重, 进行加权输入特征, 输出加权后的特征图, 通过这种复杂的注意力机制, 模型能够更好地捕捉多样性特征, 从而提升整体分类与识别效果。

根据不同刺激可动态调节其自身感受野, 感受野计算涉及到卷积层的卷积核、步长 (Stride) 和填充 (Padding) 等参数, 对于第 i 卷积层, 其感受野大小与第 i 层卷积核大小、步长以及第 $(i+1)$ 层感受野大小有关, 最后一个卷积层感受野大小等于该层卷积核大小^[63], 故本文选择在特征图上用于提取空间特征的 3×3 卷积后面添加一个 SetsA Block 结构。

5.4.2 参数计算

第 s 个基数组计算公式如下：

$$Sets^s(i) = \sum_{i=s(A-1)+1}^{sA} Sets_i^s \quad (5-2)$$

将特征图组 $Sets$ 进行二次划分得到基数组 $SetsA$ ：

$$SetsA_c^s = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W Sets^s(i, j) \quad (5-3)$$

式中： $SetsA_c^s$ 为第 s 个基数组第 c 个分量， $s \in 1, 2, \dots$ ，表示 $Sets$ 第 s 个组^[64]。

5.4.3 可变形卷积操作(Deformable Convolution, DCN)

可变形卷积可以克服传统卷积操作局限性，传统卷积操作通过固定卷积核对输入特征图进行采样，比如 3×3 的卷积核会对 3×3 的特征图进行采样，采样点间隔和位置都固定，通过引入可变形卷积，允许卷积核在处理输入时具备更灵活的位置，使得卷积核能够适应输入特征图中变形和非规则结构，而不改变卷积核尺寸，添加偏移量来调整卷积核的采样位置，偏移量是通过学习得到的，并且对于每一个卷积操作，偏移量会根据输入特征图结构自动调整，控制卷积核在每个位置上进行的采样，另外还添加掩码来进一步控制卷积核采样点的权重，掩码同样是通过网络学习得到，可结合偏移量一起使用。

可变形卷积对输入特征图进行处理，将其通道数增加 A 倍，可以处理空间上复杂变形结构，通过可学习偏移量来调整卷积核采样位置，根据输入特征动态变化，卷积核能够灵活地适应不同形变结构，提高对复杂结构适应能力，公式如下：

$$H_i^{(n)} = \frac{H_i^{(n-1)} + 2 \times \text{Padding} - K}{\text{Stride}} + 1 \quad (5-4)$$

$$W_j^{(n)} = \frac{W_j^{(n-1)} + 2 \times \text{Padding} - K}{\text{Stride}} + 1 \quad (5-5)$$

式中： $n=1$ 为经过偏移处理， $n=2$ 为经过掩码处理； $H_i^{(n)}$ 为经过第 i 个图像经过处理后得到的高度； $W_j^{(n)}$ 为经过第 j 个图像经过偏移处理得到的宽度； K 为卷积核的尺寸^[65]。

5.4.4 分割与 GAP

先对不同路径的输入进行加权处理，如果 $SetsA$ 大于 1，则会将特征图沿 $SetsA$ 的通

道维度进行分割，其为张量操作一部分，用于实现特定网络结构，在给定维度上分割数据，不涉及空间通道数总量变化，沿着通道维度(Dim=1)进行分割^[66]，原来的通道数被分配到多个分割张量中，每个分割张量的通道数减少，但所有分割张量的通道数总和等于原来的通道数，计算每个分割平均值，并通过全局平均池化(GAP)计算出特征，将二维特征图压缩成单一的值，从高度和宽度中提取出全局的特征信息，公式如下：

$$y_k = \frac{1}{|\mathcal{R}|} \sum_{(p,q) \in \mathcal{R}} x_{kpq} \quad (5-6)$$

式中： y_k 表示与第 y_k 个特征图的全局平均池化输出值； x_{kpq} 表示第 k 个特征图区域 $\mathcal{R}$ 中位于 (p,q) 处的元素； $|\mathcal{R}|$ 表示第 k 个特征图全部元素的个数^[67]。

5.4.5 Dense 低维映射和权重生成

特征信息先通过第一个全连接卷积层 Dense1，先通过两次线性变换对特征进行处理，调整维度和生成原始注意力特征（也叫中间特征），减少通道数并生成一个中间的特征表示，将输入特征映射到一个低维的表示空间，映射到低维空间后得到中间特征，再经过 BN 以缓解梯度消失和梯度爆炸问题，使输入特征保持在相对稳定的分布范围内，从而缓解网络中层输入变化的影响，中间特征经过激活函数(ReLU)，ReLU 的计算速度相对较快，只比较输入值与 0 的大小，无需进行复杂数学运算^[69]，来计算每个输入特征权重，这些权重反映部分输入特征在特定任务中的相对重要性，再通过第二个全连接卷积层 Dense2，将特征恢复到与输出的 A 倍一样的形状，BN 公式如下：

$$BN(\hat{x}) = \frac{x - \mu_i}{\sqrt{\sigma_i^2 + \epsilon}} \quad (5-7)$$

式中： μ_i 和 σ_i 是第 i 个批次均值和标准差； ϵ 是很小的常数，防止分母为零^[70]。

5.4.6 SoftmaxA

特征信息先通过全连接卷积层生成的特征不是最终的注意力权重，只为后续归一化做准备，经过 Dense 后输出特征仍然是多个通道(Channels×A)，最后进行 SoftmaxA 处理这些通道，将其归一化为注意力权重。

使用 SoftmaxA 是生成实际用于加权特征注意力权重的步骤，将 Dense2 的输出通过 Softmax 或 Sigmoid 进行归一化，输出归一化后的注意力权重，是基于全局上下文信息选择的权重，被用来加权不同路径的特征（即 SetsA 特征）以确保每个分割特征图之间

存在某种竞争性，应用修改的 Softmax 函数，这里“SoftmaxA”中的“A”与将输入特征进行分组有关，通过将输入的特征分为多个组（即“A”个组），可以对每个组分别计算 Softmax，从而在不同组之间引入竞争关系。

SoftmaxA 根据 A 值决定使用 Sigmoid(A=1)还是 Softmax(A>1)进行归一化，当 A 大于 1，SoftmaxA 将不同的子通道（路径）进行 Softmax 归一化，确保这些路径注意力权重相加为 1，将输入张量重新排列为具有 A 和 S 形状，应用 Softmax 函数，最终将其重新调整为原始形状，如果 A=1 将直接对输入张量应用 Sigmoid 激活函数，对权重进行压缩使其在 0 到 1 之间，改进公式：

$$w(x_i, y_j) = \begin{cases} \frac{1}{1 + e^{-x_i}}, A = 1 \\ \frac{e^{x_i}}{\sum_{j=1}^n e^{y_j}}, A > 1 \end{cases} \quad (5-8)$$

式中： x_i 是输入向量第 x 组的分量 i ； y_j 是输入向量第 y 组的分量 j ；分母是所有输入分量指数和，即 $\sum_{j=1}^n e^{y_j}$ ，确保输出为归一化概率分布^[71]。

5.4.7 加权求和

由于 SetsA Block 处理的是神经网络不同层之间生成的特征图，并且处理后的特征图维度均相同，故这里采用相加(Add)操作，可以让模型依旧保持原有特征图大小，逐元素相加可以加上不同层的权重信息，将得到的融合多层信息应用于原始特征图并求和，输出最终注意力加权特征图。

传统深度学习模型需要大量时间和计算资源来训练，结构复杂并且涉及大量的超参数，使得模型训练和优化过程非常耗时且成本高昂，随着数据量增加在处理大规模数据时面临挑战，SetsA-Resnet提供一种替代方案，在不需要深度结构情况下进行有效学习，利用输入数据映射特征作为网络的特征节点，将这些特征通过非线性激活函数扩展为增强节点，将所有特征和增强节点直接连接到输出端，通过SoftmaxA得到输出权重，这种设计使得SetsA-Resnet在结构上不需要深度层次，而是通过增加网络宽度来提升学习能力，相对于深度结构，宽度结构没有层与层之间的耦合而非常简洁，由于没有多层连接所以不需要利用梯度下降来更新权值，在同样计算量下的计算速度优于传统卷积神经网络，在网络精度达不到要求时，可以通过增加网络的宽度来提升精度，SetsA-Resnet适用于数据特征不多但对预测实时性要求较高的情况，通过其增量学习算法，可以高效地处

理新加入的数据,而无需重新训练整个网络,使得网络扩展和更新变得更加容易和高效。

5.5 超参数搜索与参数优化

超参数指在模型训练之前设置的参数,不会在训练过程中自动更新,是通过默认设定的,与之对应的模型参数(如权重和学习率)是可以在训练中通过优化算法自动更新,对模型参数进行优化以提升模型的性能与效果,先根据常见默认值选择参数与超参数进行训练,在训练过程中再进行调整,初始数值表 5-1。

表 5-1 初始参数与超参数配置	
参数与超参数	数值
输入图片尺寸	224×224×3
批量大小	128
训练周期	100
学习率	0.001
迭代次数	50
类别权重	0.9
动量	0.8
训练集: 验证集	8:2
优化器	<i>Adam</i>

5.5.1 网格搜索

网格搜索对参数空间尝试每一种可能的组合,对每个组合进行完整评估,当超参数空间较大时,计算量会非常大,全面覆盖超参数空间,对于小的超参数空间非常有用^[72]。

5.5.2 随机搜索

随机搜索是在超参数空间中随机选择超参数组合进行评估,直到达到最佳效果或者其他停止条件,特别是当超参数空间较大时,随机搜索能够覆盖更广泛的超参数空间,从而找到较好的超参数组合^[73],计算效率高但是计算速度慢,尤其在高维度的超参数空间,需要花费较多的计算资源才能找到好的超参数。

本文将网格搜索和随机搜索结合使用,在超参数选择过程中采用随机搜索,同时,在每个超参数组合中使用网格搜索进行细致调优,先随机定义若干个较大的搜索空间,并

使用随机搜索从中选择一部分超参数组合，对于每个组合再使用网格搜索细化搜索空间，最终得到表 5-2：

表 5-2 最终参数与超参数具体数值	
参数与超参数	数值
输入图片尺寸	224×224×3
批量大小	64
训练周期	100
高斯卷积核	19×19
指数衰减因子	(0.9,0.999)
权重衰减	0.001
稳定常数	1×10^{-8}
学习率	0.000184
迭代次数	100
类别权重	0.896542
动量参数	0.779691
训练集：验证集	8:2
优化器	<i>Adam</i>

5.6 SetsA-Resnet 网络分类评估指标

分类问题是机器学习领域中的常见任务，其核心目标是将输入样本正确地归类到预定义的离散类别中，分类模型通过分析样本特征，学习特征与类别之间的映射关系，在新样本上预测其所属类别，本文研究的视网膜病变诊断系统是分类问题，通过对视网膜病变影像数据集分析，诊断系统需要将患者眼部状况划分到不同的病理类别或者健康类别中，这一任务的复杂性主要体现在影像数据的高维特性以及病理之间存在微小差异，为解决这一分类问题，本文设计一套基于卷积神经网络诊断系统，旨在提高分类准确性和效率，最终为眼科疾病的早期诊断和治疗提供有力的支持，下面是分类评价指标：

5.6.1 混淆矩阵

混淆矩阵是分类器做出正确和错误预测数量的总结，将实际值与模型预测值进行比

较^[74]，对于多分类问题，假设有 K 个类别，混淆矩阵如表 5-3 所示：

表 5-3 混淆矩阵参数

	预测类别 1	预测类别 2	...	预测类别 K
真实类别 1	n_{11}	n_{12}	...	n_{1K}
真实类别 2	n_{21}	n_{22}	...	n_{2K}
...	...	...	...	...
真实类别 K	n_{K1}	n_{K2}	...	n_{KK}

式中： n_{ij} 为实际类别 i 但被预测 j 样本数量^[75]。

5.6.2 准确率(Accuracy)

准确率指模型预测正确占总样本数比例，当各类别样本数量较为均衡时（本实验适用），准确率是一个好的评估指标，公式如下：

$$Accuracy = \frac{\sum_{i=0}^K n_{ii}}{K} \quad (5-9)$$

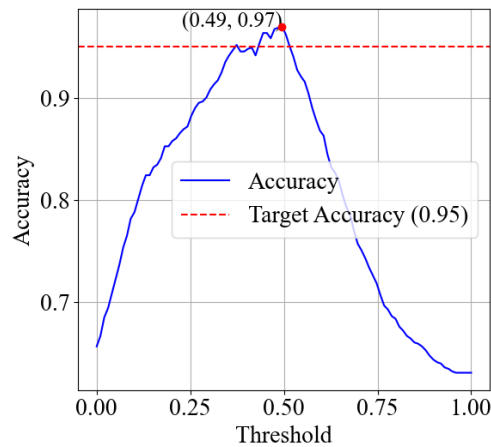


图 5-4 准确率随阈值变化曲线

图 5-4 展示通过遍历多个不同阈值来计算每个阈值下的准确率，横轴是阈值 (Threshold)，纵轴是准确率 (Accuracy)，显示模型准确率随决策阈值变化而变化情况，随阈值增加，准确率先逐渐提高，达到峰值后开始下降，当阈值较高时，模型更倾向于拒绝更多正例，从而降低准确率。

从图 5-4 可以看出最佳效果出现在阈值为 0.49 位置，此时准确率达到最高点 0.97，在这个阈值下，模型能够准确地区分正例和负例，同时避免过多误判，有助于提高模型整体性能，使其在实际应用场景中更加可靠。

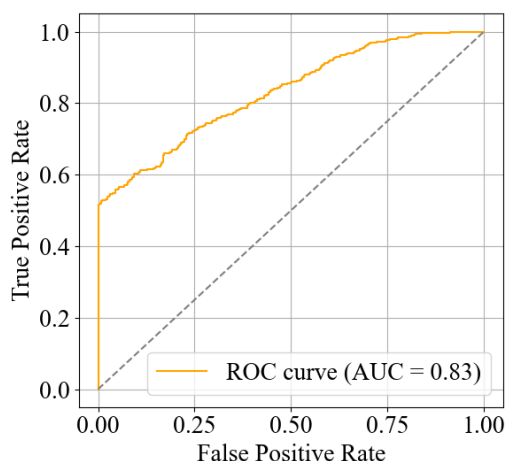


图 5-5 最佳阈值下的 ROC 曲线

图 5-5 显示不同阈值下的真阳性率和假阳性率之间的关系，AUC(Area Under the Curve)表示 ROC 曲线下的面积，AUC 越高模型越好， x 轴预测为患某种病且真正为未患病的概率， y 轴预测为患某种病且真正为这种病的概率，在实际医学检测中，哪怕将未患病的人被误诊后还可以进行下一次检测，但是假如将患病的人识别为未患病，不会再一次进行检查，代价较大，图中看出假阳性率从 0~1 变化表示模型将负样本错误地预测为正样本的概率，真阳性率从 0~1 变化表示模型正确预测正样本的概率，当假阳性率较低时（小于 0.2），真阳性率也较低（接近 0），随着假阳性率的增加，真阳性率也逐渐上升，当假阳性率达到 1 时，真阳性率也达到 1，模型将所有样本都预测为正样本，当阈值为 0.49 时，模型的 AUC 值为 0.83，模型在区分正负例方面表现良好。

5.6.3 精确率(Precision)

真正为类别 i 数量占类别 i 样本数量比例，误将真正为类别 i 样本数量占预测为类别 j 代价较高时（本实验适用），可采用精确率进行评估：

$$Precision = \frac{n_{ii}}{\sum_{j=1}^K n_{ji}} \quad (5-10)$$

精确率用于衡量模型预测的准确性，尤其在不平衡分类问题中，精确率能帮助理解模型在预测正类样本时的表现，高精确率意味着模型在预测为正类时，实际为正类的比例较高，模拟理想情况，预测完全匹配真实标签，可以展示模型在最佳状态下的性能，低精确率场景展示模型在不理想状态下的性能，帮助理解模型可能出现的误判。

图 5-6 前两幅散点图将真实标签和预测标签分布可视化，直观展示模型预测

效果，显示高精确率场景和低精确率场景，颜色鲜艳，易于区分，最后一幅图柱状图，直观看到不同情境下的精确率范围，显示模型在不同情况中的表现差异。

5.6.4 召回率(Recall)

当误将真正为类别 j 的样本数量占预测为类别 i 的代价较高时（本实验适用），召回率侧重于捕获所有正类样本的能力当假阴性的代价很高时（眼睛有疾病但是筛查漏掉病人的情况），召回率是非常重要的指标，在这种场景下更愿意提高召回率，即使会增加假阳性(FP)，可采用召回率进行评估：

$$Recall = \frac{n_{ii}}{\sum_{j=1}^K n_{ij}} \quad (5-11)$$

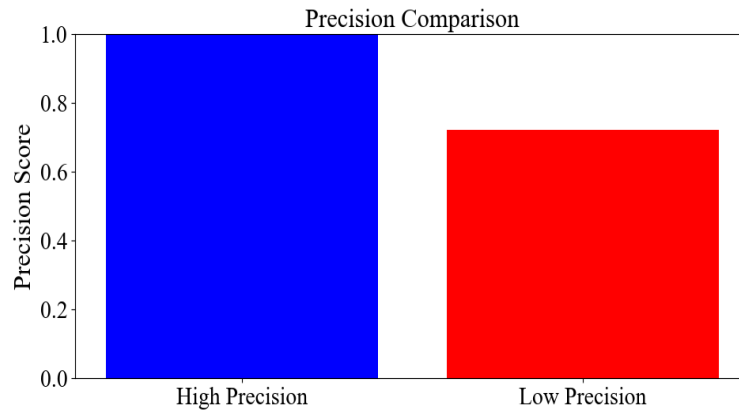


图 5-6 真实标签与预测标签效果展示

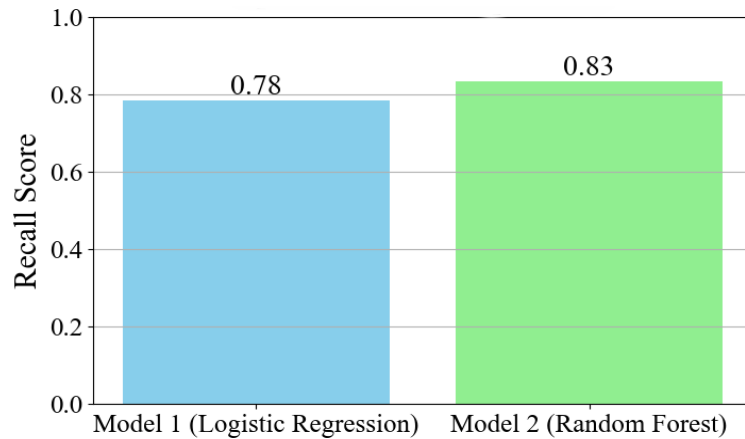


图 5-7 召回率对比

从图 5-7 可以看出，Model1(Sets-ResNet50)召回分数为 0.78，Model2(SetsA-ResNet50)召回分数为 0.83，通过对比可以看出 Model2 召回分数高于 Model1，在识别或检测目标方面表现得更好。

5.6.5 F1 值(F1-score)

F1-Score 是精确率(Precision)和召回率(Recall)的调和平均数,避免单独使用精确率或召回率带来的偏差,提高精确率会降低召回率,F1-Score 提供折中指标来综合评估模型,适用需要在精确率和召回率之间权衡并且类别不均衡的数据集,本文采用数据集有些类别图片较少,经过数据预处理的过采样和 SMOTE 方法进行扩充数据,但是数据集本身存在类别不均衡的情况,F1-Score 公式为:

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (5-12)$$

F1-Score 在精确率和召回率之间找到平衡,这两个值差异较大时会趋向较小那个值。

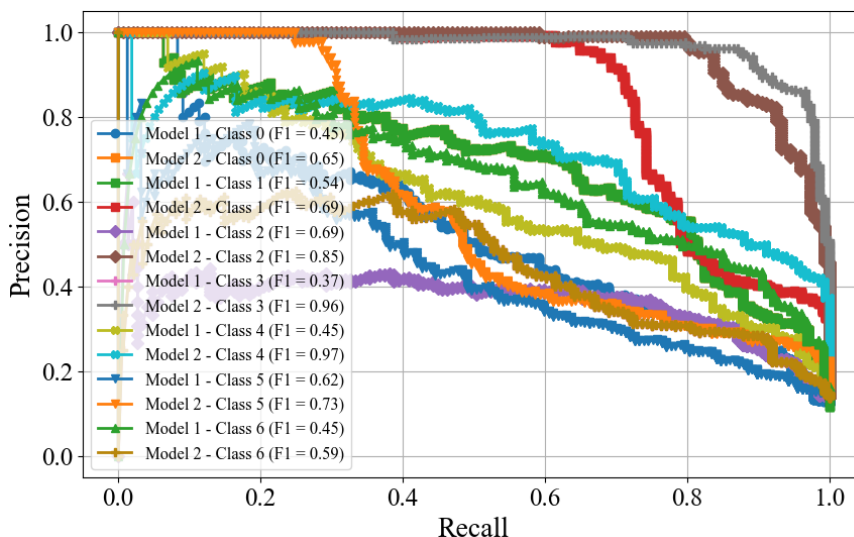


图 5-8 Precision-Recall 曲线

图 5-8 的 Precision-Recall Curve (精确率-召回率曲线) 展示不同阈值下模型精确率和召回率之间的权衡,图像中每条曲线代表一个模型在不同 F1 分数下的表现,显示随着召回率和精确度变化,所围成的面积(F1 分数)的变化,通过对比这些曲线评估各个模型性能表现,所围成面积越大就具有更好性能^[76],从图中 F1 分数可以看出, model1 波动范围为 0.5 ± 0.1 , model2 波动范围为 0.67 ± 0.13 。

还可以观察曲线形状来判断模型稳定性,曲线呈现出平滑且连续的趋势,那么这个模型更加稳定和可靠,相比较于 model1, model2 的曲线更加平滑,模型更加稳定。

5.7 K 折交叉验证

分类网络验证方式常采用 K 折交叉验证,将数据集分成 K 个不重叠子集,每次使用 K-1 个子集作为训练集,剩余 1 个子集作为验证集,重复 K 次,最终得到 K 次验证结果

平均值，数据集被划分为 K 个子集，模型会进行 K 次训练和验证^[77]， K 折交叉验证的优点是能够充分利用数据，并且通过多次评估减少单次划分带来的随机性，使得模型评估更加可靠。

每次验证误差^[79]：

$$E_i = \text{loss}(f_{\text{train on all except fold}}, D_i) \quad (5-13)$$

总误差^[80]：

$$E_{K\text{-Fold}} = \frac{1}{K} \sum_{i=1}^K E_i \quad (5-14)$$

将数据集分为 K 份并训练 K 次，每次使用不同子集作为验证集，再将所有验证结果取平均^[81]。

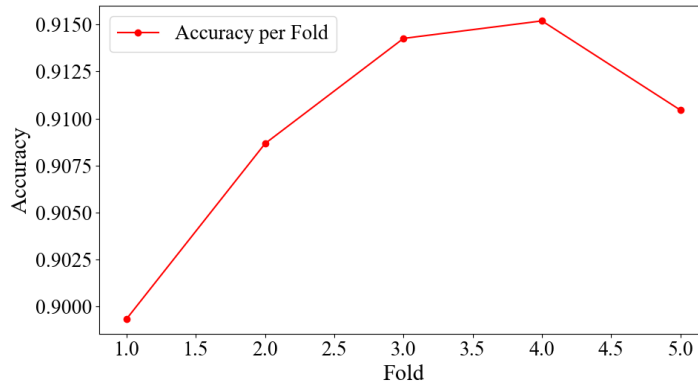


图 5-9 K 折交叉验证准确率与 AUC

图 5-9 显示模型在不同折数(Fold)下的准确度(Accuracy)，清晰地展示模型在不同折数下的准确度变化情况，提供关于模型性能的重要信息，从图中可以看出随着折数增加，模型准确度呈现出先上升后下降的趋势，在折数为 1 时，模型的准确度最低，为 0.896，然后随着折数增加到 2、3 和 4，模型准确度逐渐提高，并在折数为 4 时达到最高点，准确度为 0.9152，然而当折数继续增加到 5 时，模型的准确度开始下降，降到 0.912。

5.8 SetsA-Resnet 分类网络实验结果及分析

在基于卷积神经网络的视网膜病变诊断系统研究中，对传统和改进的网络模型进行测试，得到如图 5-10，横坐标为测试周期数，纵坐标为测试准确度和损失：

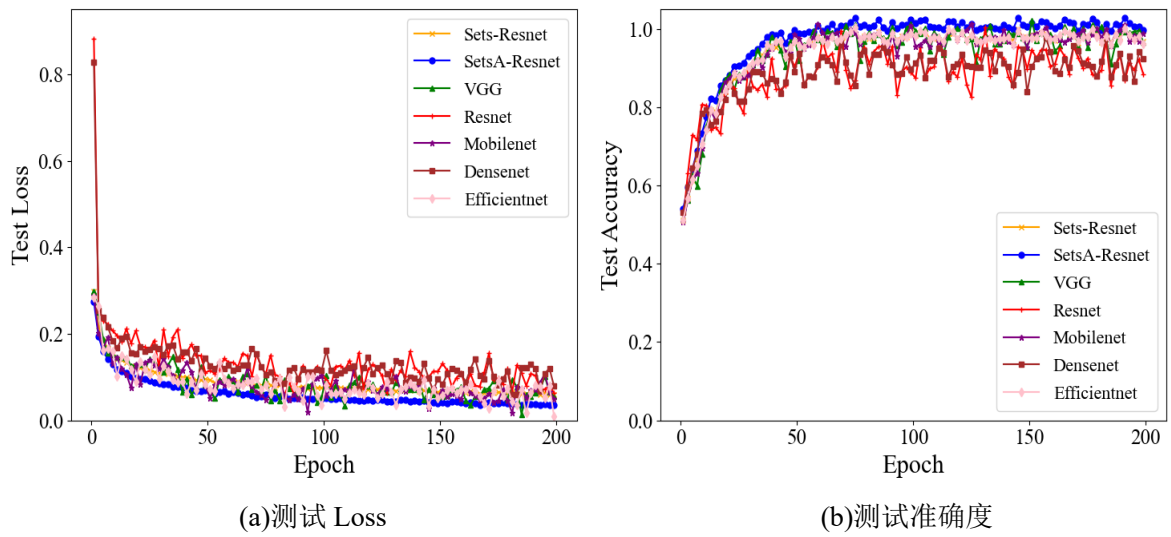


图 5-10 不同分类模型的测试过程

与原始模型相比,改进模型比 Sets-Resnet 更快达到拟合状态,对于测试损失来说,随着迭代次数增加,所有模型损失值都在逐渐降低并伴有小范围起伏波动,在开始阶段,Resnet 和 Densenet 出现明显的不稳定现象,在某些周期上的准确率突然下降,这是由于模型未能很好地泛化到新的数据样本上,整个过程 SetsA-Resnet 测试数据显示更加准确且更加平滑,能够降低损耗值并且提升稳定性,在视网膜病变图像分类任务中可以实现更高性能。

在测试集上, SetsA-Resnet 的初始损失比 Resnet 降低 30%, 最终损失降低 25%, 并且 SetsA-Resnet 最终准确率比 Sets-Resnet 提升 8%, Sets-Resnet 和 SetsA-Resnet 的训练过程都比 Resnet 更平稳, SetsA-Resnet 整体效果更佳。

图 5-11 的混淆矩阵展示多分类模型在不同类别上的预测性能,可以看出该模型在类别 5 上表现最佳,其被正确分类的比例高达 86.1%,而在类别 0 和 1 上的正确分类比例分别为 3.2%和 0.6%,其他类别的分类效果更差,表明模型对类别 5 的识别能力较强,但对类别 0 和 1 以及其他类别的识别能力较弱,混淆矩阵揭示模型在不同类别上预测准确性存在显著差异,并且 SetsA-Resnet 在针对视网膜病变识别分类上有较好的效果。

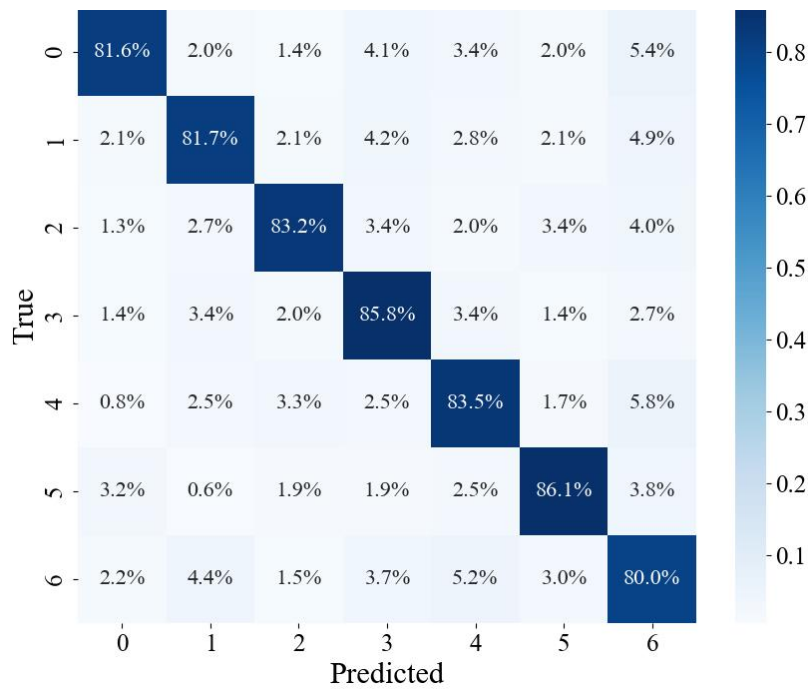


图 5-11 SetsA-Resnet50 分类混淆矩阵

表 5-4 OCT 图像在改进 ResNet50 网络上的分类结果

项目	Sets-ResNet50			SetsA-ResNet50		
	ACC	Precision	F1	ACC	Precision	F1
DR	0.76	0.73	0.80	0.85	0.84	0.90
CNV	0.74	0.75	0.78	0.82	0.82	0.87
AMD	0.78	0.72	0.81	0.87	0.81	0.62
RVO	0.76	0.71	0.77	0.86	0.86	0.83
CSC	0.75	0.74	0.78	0.85	0.84	0.86
normal	0.79	0.76	0.81	0.88	0.85	0.91
others	0.77	0.72	0.79	0.84	0.83	0.89

表 5-4 展示在 OCT 图像分类任务中, Sets-ResNet50 和 SetsA-ResNet50 的表现结果, SetsA-ResNet50 模型相较于 Sets-ResNet50 模型在多个类别上表现出显著性能提升, ACC 提升 9.09%~13.33%, Precision 提升 12.50%~21.13%, F1 分数提升 7.79%~12.50%, 对于糖尿病性视网膜病变(DR)和中心浆液性脉络膜视网膜病变(CSC), SetsA-ResNet50 模型的准确率(ACC)分别提升 11.84%和 13.33%, 精确度(Precision)分别提升 15.07%和 13.51%, 而 F1 分数则分别提升 12.50%和 10.26%, 表明该模型在识别这些疾病方面具有更高的可靠性和准确性, 在脉络膜新生血管(CNV)和视网膜静脉阻塞(RVO)的分类中,

SetsA-ResNet50 模型同样显示出优势, 其中 ACC 分别提升 10.81%和 13.16%, Precision 分别提升 9.33%和 21.13%, F1 分别提升 11.54%和 7.79%, 此外对于正常眼(NORMAL)和其他疾病(OTHERS)的分类, SetsA-ResNet50 模型也表现出较好性能, 其中 ACC 分别提升 11.39%和 9.09%, Precision 分别提升 11.84%和 15.28%, F1 分别提升 12.35%和 12.66%, 总体而言, SetsA-ResNet 模型在多数类别上相较于 Sets-ResNet 模型都有明显提升, 特别是在精确度和 F1 分数上具有更好的分类性能, 能够更准确地区分这些眼部疾病, 有助于医生进行更精确的诊断, 提高医疗影像诊断的准确性和效率。

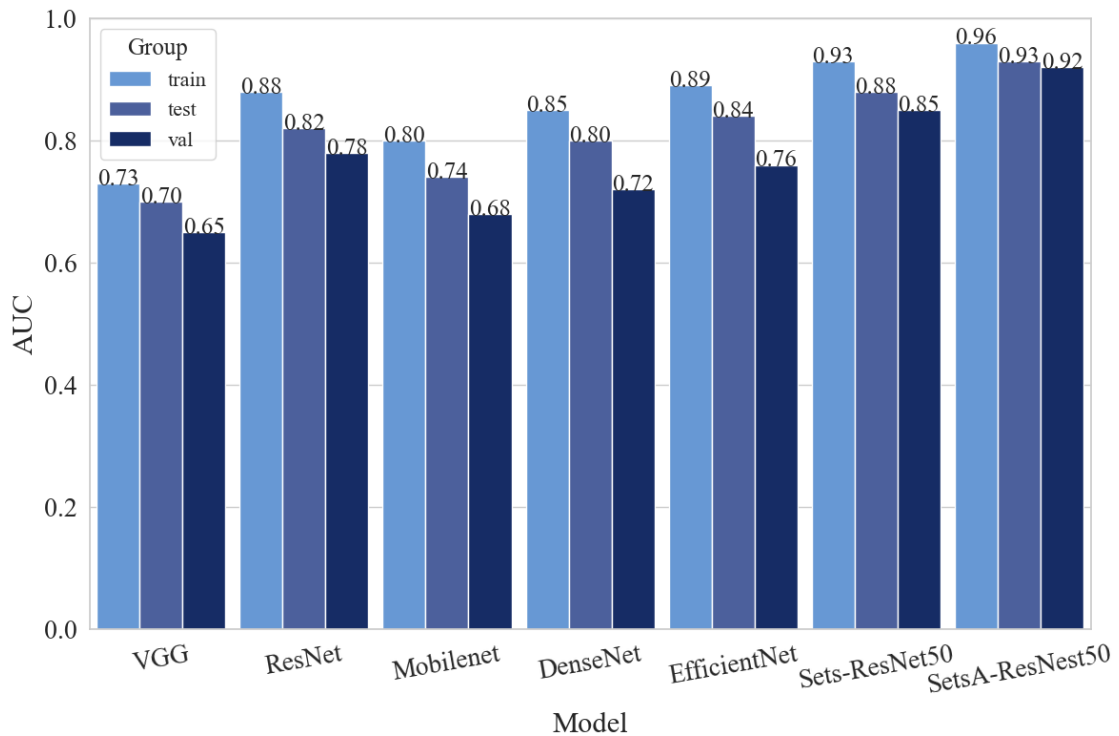


图 5-12 OCT 图像在不同网络下的分类效果

图 5-12 展示在同样视网膜病变分割任务中不同网络分类效果, 改进的 Sets-ResNet50 相较于前五种传统网络, 在测试 AUC 上平均提升 18.92%, 在验证 AUC 上平均提升 24.71%, 表明 Sets-ResNet50 在这些数据集上的表现明显优于传统的网络模型, 此外进一步改进的 SetsA-ResNet50 在 Sets-ResNet50 的基础上, 测试 AUC 又提升 5.68%, 验证 AUC 又提升 8.24%, 数据表明, 通过进一步优化和调整, SetsA-ResNet50 在性能上有更显著提升, 显示出其在分类任务中的优越性。

总结与展望

总结

相比于上述研究，本文在以下方面取得创新和突破，一是提出改进的分割网络 SF-UNet，基于经典 U-Net 网络优化设计，显著提升病变区域分割的边界细节处理能力，SF-UNet 在分割效果上较传统网络如 U-Net、CS-Unet 表现出更好的边界细节提取能力；二是在分类网络设计中，基于 ResNet50 进行优化，提出了 SetsA-ResNet50 网络，通过引入注意力机制和改进的残差模块，提升模型对类别间细微特征的提取能力，特别是在大规模数据集上的泛化性能更为优异，同时减少模型的参数量，提升效率。

在研究中，针对 7 个类别的 6 万张 OCT 图像构建完整的视网膜病变诊断系统，包括 SF-UNet 分割网络和 SetsA-ResNet 分类网络两大模块，与传统方法相比，本文的系统具有以下显著优势：

分割网络的改进：SF-UNet 通过引入空间注意力机制和多尺度特征融合策略，不仅在病变区域的边界处理上精度更高，而且在实际运行中具有较低的计算成本。这对于实际临床环境中的实时需求尤为重要。

分类网络的优化：SetsA-ResNet 网络在传统 ResNet 基础上引入通道注意力机制，提升对复杂病变特征的捕捉能力，此外在网络中通过数据增强和类别平衡策略，有效缓解数据分布不均对分类效果的影响，在大规模数据集的实验中，分类准确率和 F1 分数均显著优于传统网络。

整体系统的集成性：研究将分割与分类模块高效结合并形成完整的视网膜病变诊断系统，与单纯的分割或分类方法相比，该系统在诊断精度和效率上都有明显优势，能够为临床诊断提供全面支持，不仅聚焦于提升单一任务的性能，还注重整体系统的设计和优化，确保其能够在临床环境中具有良好的实用性和鲁棒性，未来将进一步优化模型的可解释性，增强其在真实应用场景中的可信度和可用性。

展望

尽管本文在视网膜病变诊断系统的研究中取得显著的创新和突破，但依然存在一些可以进一步优化和探索的空间：

(1)数据集的多样性与质量：虽然已经使用大规模的 OCT 图像数据集，但对于某些罕见或特殊类型的病变，数据量可能仍然不足。未来可以通过增加更多多样化的数据，特别是对于复杂和罕见病例的深入收集，以进一步提高模型的泛化能力和准确性。

(2)模型可解释性：决策过程缺乏透明度，为提高临床医生对模型预测结果信任度，未来可以研究如何增强模型可解释性，例如通过可视化技术展示模型关注的关键区域，或者开发可解释的 AI 算法来揭示模型的内部工作机制。

(3)实时性能与部署：尽管 SF-UNet 和 SetsA-ResNet 网络在实验中表现出色，但在实际应用中还需要考虑其在临床环境中的实时性能和易用性。未来可以进一步优化算法的计算效率，减少延迟，并开发用户友好的界面和工具，以便医生能够轻松地将该系统集成到现有的临床工作流程中。

(4)多模态数据融合：目前的研究主要基于 OCT 图像，而忽略其他可能有助于诊断的信息源，如眼底照片、荧光素血管造影等，未来可以尝试将多种影像学数据进行融合，以提供更全面和准确的诊断信息。

参考文献

- [1] 曹凯, 胡爱莲.中国 13 个地区 2021 至 2022 年成人盲和视力损伤的横断面调查[J]. 中华眼科杂志, 2025, 61(2): XXX-XXX.
- [2] Jia H, Dai J, He L, et al. EROSION III: a multicenter RCT of OCT-guided reperfusion in STEMI with early infarct artery patency[J]. Cardiovascular Interventions, 2022, 15(8): 846-856.
- [3] 郑雯, 沈琪浩. 基于 Improved DR-Net 算法的糖尿病视网膜病变识别与分级[J]. Acta Optica Sinica, 2023, 41(22): 745-763.
- [4] 吴天柱, 吴学森. 基于深度神经网络原理构建糖尿病视网膜病变辅助诊断模型[J]. 中华疾病控制杂志, 2022, 26(1): 61-77.
- [5] 朱小红, 张云. GhostNet 轻量级网络在糖尿病视网膜病变诊断中的应用价值[J]. 首都医科大学学报, 2024, 45(4): 678-698.
- [6] 李明远, 房丰洲. 基于青光眼影像的人工智能辅助诊断及进展[J]. Optoelectronics Progress, 2024, 61(14): 1404-1426.
- [7] 吴晨玥. 基于卷积神经网络的视网膜图像分割[J]. Acta Optica Sinica, 2023, 32(3): 117-123.
- [8] 吕佳. 边界注意力辅助的动态图卷积视网膜血管分割[J]. 光电工程, 2023, 50(1): 2201-2221.
- [9] 万程. 基于深度学习眼底图像分割[J]. 中华眼底病杂志, 2020, 36(8): 456-468.
- [10] Sudha. A Convolutional Neural Network Classifier VGG-19 Architecture for Lesion Based on Deep Learning[J]. Computers, Materials & Continua, 2023, 66(1): 562-566.
- [11] 吴欣. 深度学习在舌象分类中的研究综述[J]. Journal of Frontiers of Computer Science & Technology, 2023, 17(2): 1247-1266.
- [12] 孙石磊, 李明. 深度学习在糖尿病视网膜病变分类领域的研究进展[J]. Journal of Computer Engineering, 2024, 60(8): 287-302.
- [13] 陆辉辉. 基于卷积神经网络的彩色眼底图像血管分割[D]. 广西师范大学, 2022.
- [14] Inamullah, Hassan S, Alrajeh N A, et al. Data diversity in convolutional neural network based ensemble model for diabetic retinopathy[J]. Biomimetics, 2023, 8(2): 187-198.
- [15] Arias-Serrano. Artificial intelligence based glaucoma and diabetic retinopathy detection—Retrained AlexNet convolutional neural network[J]. F1000Research, 2023, 54(4): 128-134.
- [16] 张思杰. 基于 GAN 少样本视网膜血管分割[J]. 电子测量与仪器学报, 2021, 35(11): 132-142.
- [17] 李琼. 基于深度学习的视网膜病变图像分类研究[D]. 云南大学, 2023.
- [18] 杨培伟. 卷积神经网络在生物医学图像上的应用进展[J]. Journal of Computer Engineering, 2023, 57(7): 435-452.
- [19] Farooq M, Arooj A. Untangling computer-aided diagnostic system for screening diabetic retinopathy based on deep learning techniques[J]. Sensors, 2022, 22(5): 1803-1810.
- [20] Hradeck L. Segmentation and tracking of mammary epithelial organoids in brightfield microscopy[J]. IEEE Transactions on Medical Imaging, 2022, 42(1): 281-290.
- [21] 蹇诗婕. 基于聚类过采样和自动编码器的网络入侵检测方法[J]. Journal of Cyber Security 信息安全学报, 2023, 8(6): 963-975.
- [22] Liu R. A new paradigm shift a survey on visual deep mlp[J]. Patterns, 2022, 3(7): 236-247.

- [23] Naskath.A study on different deep learning algorithms used in deep neural nets:MLP SOM an-d DBN[J].Wireless personal communications,2023,128(4):2913-2936.
- [24] Sachdeva.Resolving autism spectrum disorder through brain topologies using fMRI dataset with multi-layer perceptron (MLP)[J].Psychiatry Research:Neuroimaging,2024,343(5):839-858.
- [25] Naskath.A study on different deep learning algorithms used in deep neural nets:MLP SOM and DBN[J].Wireless personal communications,2023,128(4):2913-2936.
- [26] 王明军,屠思凯.基于反向传播神经网络盲均衡算法的水下涡旋光复用系统性能研究[J].激光与光电子学进展,2023,60(17):1780-1791.
- [27] Zhao L, Wang L.A new lightweight network based on MobileNetV3[J].KSII Transactions on Internet and Information Systems (TIIS), 2022,16(1):11-25.
- [28] Bi C.Identification method of corn leaf disease based on improved Mobilenetv3 model[J].Agronomy,2023,13(2):300-312.
- [29] Vyazovkin.Nonisothermal crystallization kinetics by DSC:practical overview[J].Processes, 2023,11(5):1438-1452.
- [30] Lu Y,Dai L.Near-field channel estimation in mixed environments for extremely large-scale MIMO systems[J].IEEE Transactions on Communications,2023,71(6):3694-3707.
- [31] Li P,Liu J.Early diagnosis and quantitative analysis of stages in retinopathy of prematurity based on deep convolutional neural networks[J].Translational Vision Science & Technology,2022,11(5):17-19.
- [32] Liao Y,Xu Y.Mergesfl:Split federated learning with feature merging and batch size regulation[J].IEEE Transactions on Communications,2024:2054-2067.
- [33] Choudhury N A, Soni B.An adaptive batch size-based-CNN-LSTM framework for human activity recognition in uncontrolled environment[J].IEEE Transactions on Industrial Informatics,2023,19(10):10379-10387.
- [34] Ren L, Heidari A.Gaussian kernel probability-driven slime mould algorithm with new movement mechanism for multi-level image segmentation[J].Measurement,2022,192(5):864-884.
- [35] Ghiasi M A, Shafahi A.Improving robustness with adaptive weight decay[J].Advances in Neural Information Processing Systems,2024,36(2):621-635.
- [36] Wang J, Liu J.Orbital angular momentum and beyond in free-space optical communications[J].Nanophotonics,2022,11(4):645-680.
- [37] Reyad M, Sarhan A M, Arafa M.A modified Adam algorithm for deep neural network optimization[J].Neural Computing and Applications,2023, 35(23):17095-17112.
- [38] Salem H.Predictive modelling for solar power-driven hybrid desalination system using artificial neural network regression with Adam optimization[J].Desalination,2022, 522:1154-1169.
- [39] Abdallah M S,Adam J.Measurements of Proton High-Order Cumulants in Collisions and Implications for the QCD Critical Point[J].Physical review letters, 2022,128(20):2485-2502.
- [40] Ogundokun.Improved CNN based on batch normalization and adam optimizer[J].International Conference on Computational Science,2022:593-604.
- [41] Gill,Sharma.Brain tumor detection using VGG19 and SGD optimizer[J].Communication and Aerospace Technology.2022:85(4)1407-1412.
- [42] Reyad.A modified Adam algorithm for deep neural network optimization[J].Neural Computing and

- Applications,2023,35(23):17095-17112.
- [43] Hassan.The effect of choosing optimizer algorithms to improve computer vision tasks:a comparative study[J].Multimedia Tools and Applications,2023,82(11):16591-16633.
- [44] Singh.Comprehensive overview for image denoising[J].Electronics,2022,11(10):1590-1602.
- [45] Heng,Ridwan.Artificial neural network model with different backpropagation algorithms and meteorological data for solar radiation prediction[J].Scientific reports,2022,12(1):10457-10476.
- [46] 李晓豪, 郑海斌.基于改进交叉熵的模仿学习鲁棒性增强方法[J].2024,62(7):698-712.
- [47] Yang Z.Federated learning with nesterov accelerated gradient[J].IEEE Transactions on Parallel and Distributed Systems,2022,33(12):4863-4873.
- [48] Zhang Z.A model predictive control strategy with switching cost functions for cooperative operation of trains[J].Science China Information Sciences,2023,66(7):1722-1735.
- [49] Prasher S.Analysis of DenseNet201 with SGD optimizer for diagnosis of multiple rice leaf diseases[J].International Conference on Computational Modelling,2022,85(5):182-186.
- [50] Hassan E.The effect of choosing optimizer algorithms to improve computer vision tasks:a comparative study[J].Multimedia Tools and Applications,2023,82(11):16591-16633.
- [51] Elshamy R.Improving the efficiency of RMSProp optimizer by utilizing Nesterov in deep learning[J].Scientific Reports,2023,13(1):8814-8835.
- [52] Ahda.Comparison of Adam Optimization and RMS prop in Minangkabau-Indonesian Bidirectional Translation with Neural Machine Translation[J].International Journal on Informatics Visualization,2024,8(1):231-238.
- [53] Auliya.Green Bean Classification:Fully Convolutional Neural Network with Adam Optimization[J].Mathematical Modelling of Engineering Problems,2024,11(6):857-871.
- [54] Lumumba.Analysis of Cross-Validation Techniques:LOOCV,K-folds Cross-Validation,and Repeated K-folds Cross-Validation in Machine Learning Models[J].2024,32(5):921-945.
- [55] Cavoretto. Adaptive LOOCV-based kernel methods for solving time-dependent BVPs[J].Applied Mathematics and Computation, 2022,429:1272-1298.
- [56] Yates.Cross validation for model selection:a review with examples from ecology[J].Ecological Monographs,2023,93(1):155-178.
- [57] Rani.Numerical solution of one-dimensional nonlinear Sine–Gordon equation using LOOCV with exponential B-spline[J].Computational and Applied Mathematics, 2024, 43(4): 1-19.
- [58] Chen J X.Pattern classification of interstitial lung diseases from computed tomography images using a ResNet-based network with a split-transform-merge strategy and split attention[J].Physical and Engineering Sciences in Medicine,2024:34-45.
- [59] Jagtap.Performance comparison, and future directions[J].Journal of Machine Learning for Modeling and Computing,2023,4(1):896-912.
- [60] Liao L, Shen L, Duan J, et al.Local adagrad-type algorithm for stochastic convex-concave optimization[J].Machine Learning, 2024, 113(4):1819-1838.
- [61] Chan K H.Multiple classifier for concatenate-designed neural network[J].Neural Computing and Applications,2022,34(2):1359-1372.
- [62] Grant J R.Proksee: in-depth characterization and visualization of bacterial genomes[J].Nucleic acids

- p research, 2023,51(1):484-492.
-
- [63] Xu C. RFLA:Gaussian receptive field based label assignment for tiny object detection[J].Springer Nature Switzerland,2022:526-543.
-
- [64] Kim B J.Dead pixel test using effective receptive field[J].Pattern Recognition Letters,2023,167(3):149-156.
-
- [65] Zhang X.Linear deformable convolution for improving convolutional neural networks[J].Image and Vision Computing,2024,149:1051-1090.
-
- [66] Wang W.Internimage:Exploring large-scale vision foundation models with deformable convolutions[J].Proceedings of the IEEE/CVF conference on computer vision and pattern recognition.2023,65(4):14408-14419.
-
- [67] 昆彭.基于可变形卷积和多尺度特征聚焦的 X 线图像非刚性配准[J].Journal of Biomedical Engineering, 2023, 40(3):492-504.
-
- [68] Zhao S.A lightweight dead fish detection method based on deformable convolution and YOLOV4[J].Computers and Electronics in Agriculture,2022,198(2):1070-1098.
-
- [69] 吕佳.条带池化注意力的实时语义分割算法[J].计算机辅助设计与图形学学报,2023,35(9):1395-1404.
-
- [70] Segu.Batch normalization embeddings for deep domain generalization[J].Pattern Recognition,2023,135(3):1091-1115.
-
- [71] Shen K.A study on relu and softmax in transformer[J].Preprint arXiv,2023,65(2):874-897.
-
- [72] Mishra.Optimizing Heart Disease Prediction Model with GridsearchCV for Hyperparameter Tuning[J].Green and Ubiquitous Computing (IC-CGU),2024,54(3):78-95.
-
- [73] Wang H.Quantumnas:Noise-adaptive search for robust quantum circuits[J],International Symposium on High-Performance Computer Architecture (HPCA),2022,68(4): 692-708.
-
- [74] Zhang Y, Zhou K, Liu Z. Neural prompt search[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2024.
-
- [75] 孟元.基于特征类内紧凑性的不平衡医学图像分类方法[J].南京大学学报(自然科学版),2023,59(4):580-589.
-
- [76] Mia.Precision-recall classification trees[J]. Evolutionary intelligence,2022,15(3):1545-1569.
-
- [77] Prusty.SKCV:Stratified K-fold cross-validation on ML classifiers for predicting cervical cancer[J].Frontiers in Nanotechnology,2022,4(3): 972-984.
-
- [78] Allgaier.Cross-validation visualized:a narrative guide to advanced methods[J].Machine Learning and Knowledge Extraction,2024,6(2):1378-1388.
-
- [79] Yan T.Prediction of geological characteristics from shield operational parameters by integrating grid search and K-fold cross validation into stacking classification algorithm[J].Journal of Rock Mechanics and Geotechnical Engineering,2022,14(4):1292-1303.
-
- [80] Dutschmann T.Large-scale evaluation of k-fold cross-validation ensembles for uncertainty estimation[J].Journal of Cheminformatics,2023,15(1):49-68.
-
- [81] Zhang X.Model averaging prediction by K-fold cross-validation[J].Journal of Econometrics,2023,235(1):280-301.

攻读学位期间发表的学术论文情况

(一)参与的科研项目

[1] 基于 3D 视觉的复杂工件缺陷检测机理和实验研究. 重点研发与成果转化项目, 课题编号: 2023yf092

(二)发表的学术论文

[1] An Enhanced YOLOv5-Based Algorithm for Metal Surface Defect Detection[J]. Applied Sciences, 2023, 13(20): 11473. (JCR Q4 已录用)

[2] 基于改进 ResNet50 网络上肢康复动作研究[J].安徽工程大学学报,2025.(已录用)

[3] Research on Retinal Lesion Image Segmentation Based on SF-Unet [J].IEEE Transactions on Pattern Analysis and Machine Intelligence,2025(再审)

致谢

时光如水，笔落终章。当论文画上最后一个句点，恍然惊觉，这场持续三年的学术修行已近尾声。回首求学之路，二十载寒窗恍如一梦，此刻心中涌动的，是难以言表的感恩与敬畏。这篇致谢，不仅是我硕士生涯的终章，更是我漫长求学之路的谢幕。站在人生的这个节点，望着来时路上深深浅浅的足迹，不禁感慨：岁月如歌，韶华易逝。那些在图书馆度过的漫漫长夜，在实验室经历的成败得失，在导师案前聆听的谆谆教诲，此刻都化作生命中最珍贵的馈赠。谨以这些文字，铭刻这段美好的相遇，致敬所有照亮我求知之路的明灯。

“高山仰止，景行行止”，此言用以形容吾师王海教授，再恰当不过。感谢恩师，导我于迷途，示我以明路。其高尚之品行，渊博之学识，严谨之治学态度，为我营造了良好的学术氛围。尤其是其科研成果与坚持创新之精神，对我影响深远，成为我人生路上的明灯。同时，亦感谢杨春来老师在我学业上的悉心指导。饮水思源，学成念师，衷心祝愿各位老师工作顺遂，平安喜乐。

父母之爱子，则为之计深远。特别感恩我的父母，作为我人生路上的导师，给予我无限的支持与包容，鼓励我坚持学业，教导我为人正直、善良、懂礼守分。在我心情低谷之时，从不吝啬对我的夸奖，教会我拥有“自我解嘲”的勇气，教我做“不待扬鞭自奋蹄”的人，不仅为我创造了优质的生活条件，更为我提供了足以饱食一生的精神食粮。

平生感动已，方寸岂悠悠。感谢同门解晓民、卢亮玉、袁义凯，愿大家继续对生活保持热爱，奔赴山海，于更高处再相逢。

以梦为马，不负韶华。感谢三年来自己的坚持，不断学习，充实思想。此虽为我学生生涯的落幕，却只是我精彩人生的开端。路漫漫其修远兮，吾将上下而求索！