

分类号: TP39

密 级: 公开

学校代码: 10596

学 号: 1020190592



桂林理工大学
GUILIN UNIVERSITY OF TECHNOLOGY

硕 士 学 位 论 文

(全日制学术型硕士)

基于深度学习的糖尿病视网膜病变 分析与研究

研 究 生 姓 名: 李贺军

导 师: 程小辉 教授

学 科 专 业: 计算机科学与技术

研 究 方 向: 医学图像处理

所 在 单 位: 信息科学与工程学院

二〇二二年六月



桂林理工大学
GUILIN UNIVERSITY OF TECHNOLOGY

Thesis for Master Degree

Analysis and research of diabetic retinopathy based on deep learning

Graduate Student:	Li Hejun
Supervisor:	Prof. Cheng Xiaohui
Major:	Computer Science and Technology
Direction of Study:	Medical Image Processing
Affiliation:	College of Information Science and Engineering Guilin University of Technology

June, 2022

桂林理工大学研究生学位论文独创性声明

本人声明：所呈交的论文是本人在导师指导下进行的研究工作及取得的研究成果。据我所知，除了文中特别加以标注和致谢的地方外，论文中不包含他人已经发表或撰写过的研究成果，也不包含为获得桂林理工大学或其它教育机构的学位或证书而使用过的材料。对论文的完成提供过帮助的有关人员已在论文中作了明确的说明并表示谢意。

学位论文作者（签名）：李贺军

2022年6月16日

学位论文版权使用授权书

本学位论文作者完全了解桂林理工大学有关保留、使用学位论文的规定，有权保留并向国家有关部门或机构及学校图书馆送交论文的印刷本和电子版本，允许论文被查阅和借阅。本人授权桂林理工大学可以将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、汇编学位论文，并通过网络向社会公众提供信息服务。（保密的学位论文在解密后适用本授权书）

本论文是否保密： 是 ☒ 否 ☐ 如需保密，保密期限为： 年

学位论文作者（签名）：李贺军

2022年6月16日

导师（签名）：程心远

2022年6月16日

摘 要

糖尿病视网膜病变 (Diabetic retinopathy, DR) 是糖尿病最常见的并发症之一, 是世界上致盲率最高的眼科疾病。在 DR 早期, 患者通常无法察觉到它所带来的影响, 容易错过最佳的治疗时间。在临床诊断中, 眼科医生通过观察患者的视网膜图像进行诊断, 该方式非常依赖于医生的诊断经验, 不仅诊断时间较长还非常耗费精力, 易发生漏诊、误诊等情况。另外, 我国患者人数众多而医疗资源又相对紧张, 导致有的患者无法及时接受诊断和治疗。而定期筛查可以实现尽早发现病情, 避免患者错过最佳的治疗时期。近年来, 深度学习在医学图像领域应用广泛, 并取得了显著的成果, 利用计算机辅助诊断技术使得眼科医生能够根据 DR 图像对患者进行诊断已成为迫切需要。基于此, 本文针对糖尿病视网膜病变图像, 利用现有的深度学习框架和学习模型, 构建 DR 分类和病灶分割模型, 实现对 DR 图像的准确分类和关键病灶点渗出物的精确分割, 这对于实现计算机辅助医生高效诊断具有非常重要的作用, 主要研究工作如下:

(1) 针对现有深度学习模型由于 DR 不同病变图像之间差异过小而导致分类准确率较低的问题, 本文设计了一种基于 MobileNetV2 和 EfficientNetB0 模型集成及注意力机制 (Attention mechanisms) 的分类网络: ME-ANet。设计网络时采用集成的思想, 将深度学习模型 MobileNetV2 和 EfficientNetB0 的核心结构分阶段进行集成, 构成模型的特征提取器, 实现了对视网膜图像多尺度特征信息的融合。同时, 设计全局注意力机制 (Global attention mechanisms, GAM), 并将其嵌入至不同阶段的集成模型中, 来获取病变特征的权重信息, 抑制非病变特征噪声。该网络实现了对图像浅层特征信息的充分提取, 减少了微病灶特征信息在训练过程中的卷积丢失问题, 有效提升了模型的分类性能。

(2) 针对现有深度模型对 DR 图像中渗出物病灶点分割精度不高而影响后期诊断的问题, 本文构建了一种基于深度学习的糖尿病视网膜渗出物分割模型, 该模型以 U-Net 编码器-解码器结构为基本网络架构, 同时结合 ResNet50 作为编码端的特征提取网络, 这不仅实现对特征提取网络的进一步加深, 还避免了模型训练中梯度消失的问题; 在解码端, 设计注意力层结构, 并嵌入到解码单元中, 来提升模型对小病灶点的识别能力; 最后在模型解码端构建多特征尺度融合层, 实现模型多尺度语义特征的学习。实验过程中采用 Focal loss 焦点损失函数, 进一步强化模型对样本中小目标物体特征的学习优化。与其他渗出物分割模型相比, 该模型分割性能较优, 且具有普遍适用性。

综上所述, 本文构建深度学习分类模型, 完成对糖尿病视网膜的病变分类任务, 然后构建深度分割模型完成对糖尿病视网膜渗出物的分割任务, 从而更加有针对性的实现计算机辅助医疗诊断的目的, 进一步提高糖尿病视网膜病变的筛查和诊断效率。

关键词: 糖尿病视网膜病变; 深度学习; 注意力机制; 病变分类; 渗出物分割

Abstract

Diabetic retinopathy (DR) is one of the most common complications of diabetes and the eye disease with the highest blindness rate in the world. In the early stages of DR, patients are often unaware of its effects and easily miss the best time for treatment. In clinical diagnosis, the ophthalmologist makes a diagnosis by observing the retinal images of the patient. This method is very dependent on the doctor's diagnostic experience, which not only takes a long time to diagnose, but also consumes a lot of energy, and is prone to missed diagnosis and misdiagnosis. In addition, due to the large number of patients in my country and the relatively tight medical resources, some patients cannot receive timely diagnosis and treatment. Regular screening can achieve early detection of the disease and prevent patients from missing the best treatment period. In recent years, deep learning has been widely used in the field of medical images and has achieved remarkable results. It has become an urgent need for ophthalmologists to diagnose patients based on DR images using computer-aided diagnosis technology. Based on this, this thesis uses the existing deep learning framework and learning model to build a DR classification and lesion segmentation model for diabetic retinopathy images, so as to achieve accurate classification of DR images and accurate segmentation of exudates at key lesions. Computer-aided doctor's efficient diagnosis plays a very important role. The main research work is as follows:

(1) Aiming at the problem that the existing deep learning model has low classification accuracy due to the small difference between different DR lesions images, this thesis designs a classification network based on MobileNetV2 and EfficientNetB0 model integration and attention mechanisms: ME-ANet. When designing the network, the idea of integration is adopted, and the core structures of the deep learning model MobileNetV2 and EfficientNetB0 are integrated in stages to form the feature extractor of the model, which realizes the fusion of multi-scale feature information of retinal images. At the same time, a global attention mechanism (GAM) is designed and embedded into the ensemble model at different stages to obtain the weight information of lesion features and suppress the noise of non-lesion features. The network realizes the full extraction of image shallow feature information, reduces the convolution loss problem of micro-lesion feature information in the training process, and effectively improves the classification performance of the model.

(2) Aiming at the problem that the existing deep model segmentation accuracy of exudate lesions in DR images is not high and affects the later diagnosis, this paper constructs a deep learning-based model for the segmentation of diabetic retinal exudates. The model uses U-Net

encoder-decoder structure is the basic network architecture, and ResNet50 is used as the feature extraction network at the encoding end, which not only realizes the further deepening of the feature extraction network, but also avoids the problem of gradient disappearance during model training; at the decoding end, design attention. Finally, a multi-feature scale fusion layer is constructed at the decoding end of the model to realize the learning of multi-scale semantic features of the model. During the experiment, the Focal loss focal loss function is used to further strengthen the model's learning and optimization of the characteristics of small target objects in the sample. Compared with other exudate segmentation models, this model has better segmentation performance and has universal applicability.

In summary, this paper builds a deep learning classification model to complete the task of classifying diabetic retinal lesions, and then builds a deep segmentation model to complete the task of segmenting diabetic retinal exudates, so as to achieve the purpose of computer-aided medical diagnosis more targeted, to further improve the screening and diagnosis efficiency of diabetic retinopathy.

Key words: Diabetic retinopathy; Deep learning; Attention mechanism; Lesion classification; Exudate segmentation

目 录

摘 要	I
Abstract	II
目 录	IV
第 1 章 绪论	1
1.1 课题研究背景及意义	1
1.1.1 研究背景	1
1.1.2 研究意义	2
1.2 糖尿病视网膜病变	2
1.3 国内外研究现状	4
1.3.1 糖尿病视网膜病变分类研究现状	4
1.3.2 糖尿病视网膜渗出物分割研究现状	6
1.4 主要研究内容	8
1.5 本文组织结构	9
第 2 章 相关工作	10
2.1 神经网络相关知识	10
2.1.1 神经网络的发展	10
2.1.2 人工神经元	11
2.1.3 常用的激活函数	12
2.2 卷积神经网络	13
2.2.1 卷积层	14
2.2.2 池化层	16
2.2.3 全连接层	16
2.3 注意力机制	17
2.3.1 SE 注意力机制	18
2.3.2 CBAM 注意力机制	19
2.4 深度学习主流开源框架	21
2.5 本章小结	22
第 3 章 基于 ME-ANet 的糖尿病视网膜病变分类	23
3.1 引言	23
3.2 糖尿病视网膜病变分类算法流程	23
3.3 糖尿病视网膜眼底图像预处理	24
3.3.1 眼底图像数据集	24

3.3.2 数据预处理	26
3.3.3 不平衡数据处理	29
3.4 糖尿病视网膜眼底图像分类模型架构设计	31
3.4.1 MobileNetV2 模型结构	31
3.4.2 EfficientNetB0 模型结构	33
3.4.3 GAM 注意力模块设计	36
3.4.4 模型集成设计	37
3.5 实验结果与分析	40
3.5.1 实验环境	40
3.5.2 实验评估指标	40
3.5.3 实验设计与分析	42
3.6 本章小结	47
第4章 基于深度学习的糖尿病视网膜渗出物分割	48
4.1 引言	48
4.2 视网膜渗出物分割模型	49
4.2.1 模型总体结构	49
4.2.2 具体模型设计	50
4.3 糖尿病视网膜眼底图像处理	51
4.3.1 实验数据集	51
4.3.2 数据预处理	51
4.4 实验结果与分析	53
4.4.1 实验环境	53
4.4.2 损失函数	53
4.4.3 实验评估指标	54
4.4.4 实验设计与分析	55
4.5 本章小结	59
第5章 总结与展望	60
5.1 全文总结	60
5.2 未来工作展望	61
参考文献	62
攻读学位期间的成果及发表的学术论文	67
致 谢	68

第 1 章 绪论

1.1 课题研究背景及意义

1.1.1 研究背景

随着物质文明的极大丰富，人们的生活方式已经彻底改变，多种慢性疾病如糖尿病也随之而来，严重影响人们的健康。糖尿病是一种由高血糖引起的代谢疾病，是现代社

会的一种高发病，正逐渐成为全世界范围内的一种慢性流行性疾病，据国际权威机构预测，到 2035 年全世界范围内糖尿病患者的人数将会增至 5.93 亿，而我国大概有 1.164 亿糖尿病患者^[1]，成为世界上糖尿病患者人数最多的国家。目前，我国的糖尿病防治工作面临着严峻的挑战。

糖尿病视网膜病变（Diabetic retinopathy, DR）是糖尿病众多并发症中最易爆发的一种典型疾病之一，特别是在工作人群中，它是一种最为常见、不可逆的眼部疾病^[2]。人类眼部视网膜血管分布比较密集且非常的脆弱，而在糖尿病患者中，高血糖的情况下会造成眼部血压长时间维持在较高水平，非常容易导致眼部视网膜血管发生病变^[3]，若不及时诊断治疗，将会影响人们的视力正常，严重的会导致失明。DR 最有效的治疗时期是在其发病早期，然而在 DR 早期，患者眼部状况往往不会发生明显变化，通常无法感觉到它所带来的影响，容易使患者忽略病情，贻误最佳的治疗时期^[4]。眼科医生建议糖尿病患者最好要做一年两至三次的 DR 诊断检查，这是由于糖尿病患者的病期越长，DR 的发病率也会不断增高，因此糖尿病患者的定期检查是非常必要的。定期检查不仅可以及时发现病情，还可以尽早抑制病情恶化，降低致盲率^[5]。一项包含了全球范围内 35 项 22896 名糖尿病患者的数据分析表明，DR 患病率为 34.6%，而我国 DR 患病率为 24.7%~37.5%^[6]，我国不同病程时间段 DR 患病率如下图 1.1 所示。

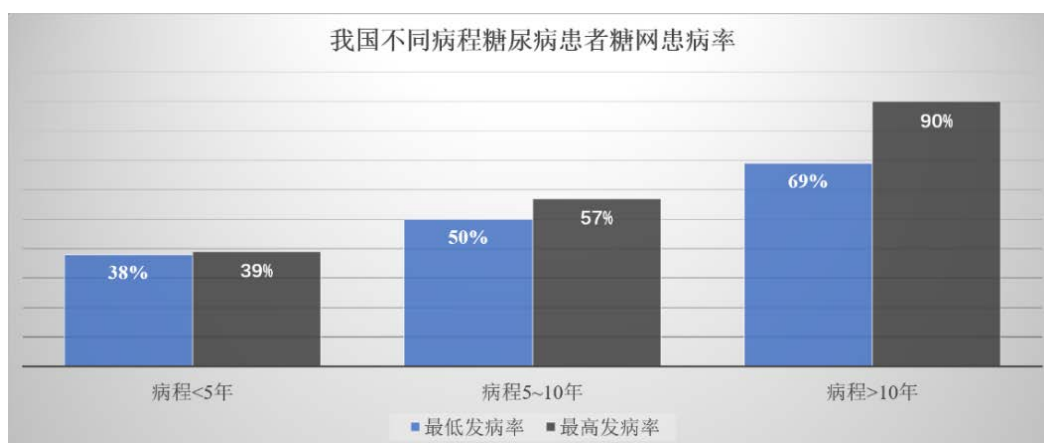


图 1.1 我国不同病程糖尿病患者糖网患病概率图

目前对于 DR 的检测诊断，分为检眼镜直接观察和眼底图像诊断两种方式，前者由

专业医生使用检眼镜直接观察患者眼底图片进行诊断,这种方式效率较低,而且诊断结果依赖于观察者的专业素养和经验,不同的专业医生由于其临床经验和主观认知不同,就会产生不同的诊断结果,有可能发生误诊或漏诊的情况;后者首先利用专业医学图像采集设备对患者眼底进行拍照来获取大量的眼底图像,然后专业医生对这些图像进行观察分析,根据图像是否含有早期病变症状来判断患者是否需要早期进行饮食控制和药物治疗^[7]。在我国十几亿的人群中,进行全面有效的 DR 筛查是一项十分必要且艰巨的任务。而 DR 筛查是及时发现早期病情的主要方式,在早期进行相应的治疗会极大抑制病情进一步恶化、大大降低致盲率^[9],使得 DR 的定期筛查成为预防 DR 视力损伤的关键环节。

1.1.2 研究意义

筛查是根据目标群体进行测试,以便通过表现鉴定出可能与特定病症相关的早期症状或符合某种倾向指标的目标个体^[10]。DR 筛查的目的是为了确定被筛查人员是否需要随访或转诊进行进一步治疗,定期的筛查预防工作通常包括:在糖尿病患者早期,诊断其是否存在相关疾病;检查患者是否需要血压及血糖方面的控制和治疗需求;宣传普及糖尿病视网膜病变的发病原因以及有效降低视网膜病变风险的方法^[11]。另外,早期的 DR 可以通过调节体内血压和血糖水平来缓解疾病的恶化^[12]。目前,利用计算机辅助技术对采集到的视网膜图像进行一系列的处理,能够实现快速高效的判断出患者的眼部疾病的类型及病变程度^[13],这对于推进国内 DR 定期筛查任务的顺利进行有着积极的作用。

在大数据时代,计算机硬件如内存、CPU、GPU 的快速升级,推动了深度学习的蓬勃发展,特别是在医学图像领域得到了广泛的应用。深度学习在医学图像分类、病灶检测和分割任务中表现出色,凭借其出色的特征提取和学习能力,成为当代学者研究的热点。为进一步使计算机技术和辅助医学相融合,有效地提高临床医师的诊断精度和效率,为了更好的解决我国目前大规模的糖尿病视网膜病变筛查和诊断问题,将深度学习技术应用在 DR 筛查任务中有巨大潜力的发展空间。深度学习可以很好地克服人工设计复杂度高、泛化能力低等问题,从而提高了识别的可信度,为眼底疾病的早期筛查与诊断提供了新的思路。

1.2 糖尿病视网膜病变

糖尿病视网膜的发病机制较为复杂,主要原因是糖尿病患者中血液成分发生变化导致视网膜微血管的内皮细胞发生功能异常,最终导致视网膜组织被损坏。在 DR 视网膜图像中非增生阶段主要有四种典型的病灶:微动脉瘤(MA)、硬性渗出物(HE)、软性渗出物(SE)及出血斑点(HA),在早期诊断中,以上四种病灶成为诊断的关键依据,图 1.2 展示了正常眼底图像和 DR 视网膜图像。MA 是早期眼底病变图像中最先病变的

组织，由毛细血管囊样扩张引发，在视网膜中一般呈现暗红色点状，边界较为清晰，该病变是诱发视网膜出血和组织渗出的最主要原因；HE 呈黄白色的不规则点状或者片状，主要是视网膜静脉中的血液发生渗漏导致，视网膜血管的长期渗漏造成血液中的脂质堆积在血管附近，HE 形成初期，边界较为清晰，随着病情的恶化，堆积点相互连接，可呈片状出现；SE 是一种灰白色块状物，边界较为模糊，主要是由于动脉血管发生闭塞，导致视网膜供血不足，影响轴浆流动，导致轴浆积聚在视网膜神经节轴突中；HA 是一种不规则的红色斑块，边缘模糊，主要是由于毛细血管破裂进而引起组织渗漏造成的，具体如下图 1.2 所示。

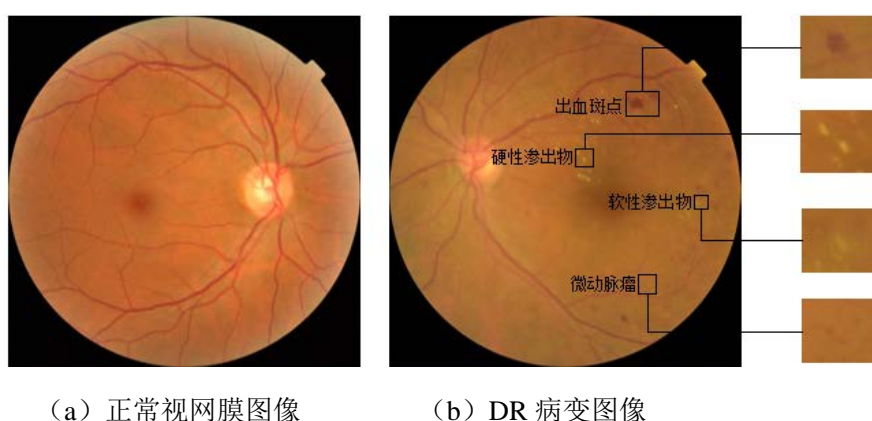


图 1.2 视网膜彩色眼底图像

图 1.2 中的图 (a) 为正常眼底图像，视网膜表面光滑且血管分布脉络清晰，无病灶点和新生血管的生成。图 (b) 为糖尿病视网膜病变患者的眼底图像，从图像中可以清楚的看出上述 4 种典型病灶。

对于糖尿病视网膜病变分期，专家往往按照三个阶段来进行划分。在第一个阶段是视网膜无异常阶段（NDR），糖尿病患者的眼部组织结构无病变状况，视力未发生异常变化；第二阶段被称为非增殖性 DR（NPDR），该阶段包括轻度 NPDR、中度 NPDR、重度 NPDR 三个不同程度的病变期，属于 DR 病变的早期阶段，在该阶段的诊断更有助于控制病情，大大降低视力丧失的风险；第三个阶段为增殖性 DR（PDR），这一时期的典型病变是新生的视网膜血管，它破坏了视网膜的组织结构，导致视网膜发生不同程度的脱离，使人的视力受到不可逆转的损害。在轻度 NPDR 阶段中，DR 患者的主要症状是微动脉瘤；在中度 NPDR 阶段，微动脉瘤和出血点病变在该阶段是最为常见的；而在重度的 NPDR 阶段主要表现为眼底视网膜内发生出血现象、部分静脉血管扩张并伴有明显的视网膜内微血管病变^[14]。糖尿病患者视网膜如果有出血、渗出物和新生血管的症状，说明患者的视网膜血管已经存在较为严重的问题，需要患者及时就医治疗。在医学眼科领域中，专家约定中度 DR 以及比其更严重的病变期被称为就诊病变，糖尿病视网膜病变分级标准^[1]如下表 1.1 所示，它是专家讨论制定的国际标准，成为目前诊断病变诊断的关键依据。

表 1.1 糖尿病视网膜病变国际五级诊断标准^[15]

糖尿病视网膜病变	具体临床表现
无视网膜病变 (无 DR)	无异常
轻度非增殖性糖尿病视网膜病变 (轻度 NPDR)	仅有微动脉瘤
中度非增殖性糖尿病视网膜病变 (中度 NPDR)	微动脉瘤, 存在轻于重度非增殖性 DR 病变表现
重度非增殖性糖尿病视网膜病变 (重度 NPDR)	出现下列任何 1 种改变, 但无增殖性 DR 病变表现: 任意 1 个象限内有超过 20 处视网膜出血点; 在 2 个以上象限内存在静脉串珠样病变现象; 在 1 个以上象限内有显著的微血管病变;
增殖性糖尿病视网膜病变 (PDR)	出现以下 1 种或多种改变: 新生毛细血管生成、玻璃体积血、视网膜前出血

1.3 国内外研究现状

在全球范围内, 糖尿病作为一种慢性疾病长期影响着人们的健康, 而由糖尿病引发的一系列的其他疾病更是困扰着人们的健康。糖尿病视网膜病变会损害眼部组织, 对眼睛造成不可恢复的损伤, 对人们的正常生活带来了极大的困扰。随着当代物质文明的极大丰富, 促使人们越来越重视自己的身体健康, 注重提高自己的生活品质。显然, 这对当代医疗技术的发展提出了更高的要求, 科技创新推动医疗领域的大发展已是大势所趋。目前, 研究者对于医学图像的研究已经从传统的图像处理算法转向深度学习方法中, 并取得了显著的成果。在糖尿病视网膜病变研究当中, 图像处理算法一般围绕分类、检测和分割三大任务, 实现从整体评估到具体病变细节的诊断, 这一成就大幅度提升了医生的诊断效率。根据本文的内容, 接下来从糖尿病视网膜病变分类和糖尿病视网膜病变渗出物分割两个方面进行讨论国内外的研究现状。

1.3.1 糖尿病视网膜病变分类研究现状

近年来, 研究人员根据糖尿病视网膜的病变程度通常将其分为五个类别: 无异常、轻度非增殖性 DR、中度非增殖性 DR、重度非增殖性 DR 和增殖性 DR^[16]。五分类是最为常见的分类方式, 能够实现对眼底病变图像更为细腻的分类和研究, 也符合专家讨论指定的关于糖尿病视网膜病变分类的国际标准, 更加有针对性实现计算机辅助医生诊断的目的。近年来, 国内外针对糖尿病视网膜病变分类任务有较多的研究, 主要分为基于

传统图像处理和基于深度学习模型两种方法。

在传统医学图像处理中主要依赖人工特征提取和传统机器学习分类。如 Adarsh 等人^[17]应用形态学方法对糖尿病视网膜图像进行特征提取,同时结合 SVM 多分类器进行训练。Shahin 等人^[18]利用形态学方式对病变图像进行处理,进而得到视网膜中渗出物区域和病变微血管面积等特征,然后计算获得熵和同质性两个评估值,并将其相应的病理特征传输到浅层深度模型中,实现视网膜病变图像中非增殖性 DR 和增殖性 DR 的分类,最后模型训练得到 88% 的灵敏度和 100% 的特异性。Nayak 等人^[19]通过将计算出的眼底图像的渗出物面积、血管面积和组织纹理等特征作为训练模型的输入信息,利用 140 张图片进行训练,该模型在准确率上为 93%,灵敏度为 90%,特异性达到 100%。以上方法都需要依靠专业医生进行人工标注,而手工特征提取不仅需要耗费大量的时间和精力,在分类训练过程中也带来了一些复杂性和不稳定性。

深度卷积神经网络对于图像的特征提取和学习能力突出,尤其在医学图像处理领域中应用广泛^[20],深度学习与医学图像领域的结合也极大的推动了智慧医疗的快速发展。深度学习辅助诊断已逐渐成为主流趋势。2015 年 Kaggle 发起了关于糖尿病视网膜病变分类竞赛,该竞赛平台提供了 Kaggle 大型数据集,该数据集含有 10 万张视网膜图像,且每张图像都具有相应的等级标签,它的出现极大的促进了深度学习在糖尿病视网膜研究领域中的应用发展。近年来基于深度学习的糖尿病视网膜分类研究成果显著。在糖尿病视网膜五分类任务中,Gulshan 等人^[22]以 Inception-v3 为基本模型架构,将约 12 万张糖尿病视网膜眼底图像输入到模型中进行训练,由于数据量比较充分以及眼科专家对眼底图像的筛选,模型的灵敏度为 97.5%,特异性达到 93.4%。李琼等人^[23]首先对视网膜图像做了一系列预处理,然后对 AlexNet 模型进行改进,主要是在卷积操作和全连接操作之后分别再进行一次批量归一化层,得到一个更为复杂的模型作为特征提取器,最后训练过程中采用迁移学习的策略,模型准确率达到 93%,具有较好的鲁棒性。丁蓬莉等人^[24]设计了一种紧凑的 CompactNet 网络模型,该模型加载了 AlexNet 的浅层模型参数,但由于缺少标记的视网膜眼底图像,模型在训练过程中提取特征信息不足,导致分类准确率不高。Zhou 等人^[25]利用多模块结构对高分辨率的眼底图片进行多目标学习,并采用分类和回归的形式进行标签预测,利用 Kaggle 数据集训练模型,在测试集上的 Kappa 值达到 84.1%,但该模型适合分辨率较高的眼底图像,对于分辨率较低的图像无法保证其精度。Pratt 等人^[26]通过利用 Kaggle 数据集训练模型,并采用数据扩增来平衡各类数据的规模,实现了 DR 五个病变等级的分类,得到的特异性为 95%,准确率为 75%。李轩屹等人^[27]设计了一种多尺度 ResNext 网络模型,同时利用多分辨率图像作为模型的输入,采用迁移学习的策略训练模型,加速模型收敛,该方法分类准确率达到 97%。闫育铭等人^[28]提出了 Inception-v3 模型,通过使用不同扩张率的膨胀卷积,进而提高了模型的感受野,在训练中使用 ImageNet 大数据集预先训练过的 Inception-v3 模型,并采用微调的方式进行模型训练,该方法取得了较高的准确度。

综上所述,深度学习技术与传统医学图像处理的有效结合,极大的提高了现代医学诊断的效率,成为辅助医疗的强有力的工具。目前,利用深度学习对糖尿病视网膜病变的诊断方式主要是基于全局图像进行分类,这种处理方式容易受到相似样本的干扰,从而影响模型分类效果。在糖尿病视网膜的不同病变时期,病灶表现各不相同,且相邻类别之间的病变图像相似度较大,这种小差异的样本严重制约着模型的性能提升,使得模型难以达到一个理想的鲁棒状态。近年来国内外虽然取得了不错的研究成果,但为了真正客观的实现辅助医疗诊断的目的,利用当代科技为人们健康事业做出一定的贡献,仍需要研究学者不断深入研究,根据实际需求,优化算法,保证模型的准确性、客观性。

1.3.2 糖尿病视网膜渗出物分割研究现状

目前,以深度学习方法为基础的糖尿病视网膜病变分级任务日趋成熟,而糖尿病视网膜病变的不同阶段会出现不同类别的病理特征,如微动脉瘤、出血病变、硬性渗出和软性渗出等病灶点。由于眼底视网膜环境的复杂性、病灶形状的多样性和图像噪声干扰等问题,实现自动精准的病灶检测仍具有较大的挑战。在 DR 筛查任务中,视网膜图像简单的分类任务远不能达到有效辅助医疗的目的,在准确分类的基础上,实现对关键病灶病变区域的有效分割更有利于专家诊断。在中国繁重的 DR 筛查任务中,这种分类和分割任务的实现无疑推动了筛查任务的高效进行,具有非常重要的现实意义^[29]。专业医疗机构和医院对大规模医学图像数据集的开放也进一步推动了深度医疗产业的研究和发展。深度分割模型能够从特定尺度的眼底图像中渗出物病灶特征,与传统图像处理方式相比较,具有明显的优势^[30]。在深度学习中,根据算法是否需要标记特征信息,一般将其划分为非监督 and 全监督两种学习方式。接下来,本文将具体介绍近年来基于这两种学习方式而设计的渗出物分割算法。

在糖尿病视网膜图像研究的早期阶段,对于渗出物的处理主要使用非监督的学习方式,具体包括形态学滤波处理、聚类和区域生长的方法。在形态学方法中,主要利用不同的形态学滤波算子进行处理。Zhang 等人^[31]提出了一种基于形态学的候选区域分割方法,这些候选区域包括典型的病灶特征,也存在新的背景信息,最后使用随机森林算法来检测候选区域的渗出物。Harangi 等人^[32]设计了一种硬性渗出物的分割模型,利用形态学方式来提取图片中亮度较大的区域,由于活动轮廓方式可以对区域边界进行标记,作者使用该方式对提取到的区域进行标记,最后去除了标记区域中的非硬性渗出物,获得具体的渗出物分割图;在基于聚类方法中,将视网膜病变图像中不同纹理、组织、颜色等特征进行分类,重新计算各类别之间的距离,不断调整中心点的位置,直到收敛状态。高玮玮等人^[33]对模糊 C-均值聚类算法进行了改进,并结合支持向量机分类器对糖尿病视网膜眼底图像进行渗出物的检测,并得到了不错的结果。基于区域生长则是利用病灶点边界的清晰程度来识别病变区域。Pereira 等人^[34]提出将蚂蚁优化算法应用到候选边

界的处理中,该算法具有增强渗出物边界的效果,有利于区域生长方法对边界的分割。

基于监督学习的病灶识别分割算法包含传统机器学习和深度学习两种方法。在机器学习方法中,通常采用支持向量机、贝叶斯等分类器实现特征分类的目的,而特征的提取如视网膜渗出物的颜色、亮度、血管面积等,大多需要利用形态学方式处理图像后获得。**Chen** 等人^[35]使用直方图分割方法提取病变候选区域,然后设计 **SVM** 分类器并利用候选区域的相关特征进行分类。**Kang** 等人^[36]提出一种基于统计方法和贝叶斯分类器的眼底视网膜渗出物分割模型。首先对眼底图像做通道分离、对比度增强等预处理操作,通过对 3 组纹理特征进行处理,将提取到的多组纹理特征作为贝叶斯分类器的输入,实现病变的自动检测。相比于早期的非监督图像处理方式,机器学习的方式很大程度上提升了模型的分割性能,但不足的是,该方式不能完全自动的利用图像样本进行特征学习,需要人工提取渗出物的相关特征,然后输入到分类器或者神经网络中进行训练分割。

近年来,在计算机硬件的全面升级、大数据智能化背景下,许多优秀的深度学习模型应运而生,且被广泛应用于多领域、多任务当中。特别是在医学图像领域,一定程度上革新了传统的医学图像诊断方式。**Jiang** 等人^[37]利用多个不同扩展率的空洞卷积来替换传统卷积,在保证模型感受野大小和原始图像分辨率的前提下,进一步提高了模型的特征表达能力,所提出的空洞卷积方法巧妙地权衡图像的分辨率与感受野之间的相互制约关系,又能使模型学习到更加丰富的图像细粒度特征,该模型训练后最终得到的准确率为 97.81%,特异性达到 99.04%。**叶显一**等人^[38]设计了一种提取渗出物候选区的方法,然后采用主成分分析方法对渗出物像素点处理,实现渗出物分割,同时提出了利用 **U-Net** 网络的渗出物分割方法,最终获得 98.1%的准确率和 96.7%的灵敏度。在眼底病变图像中,渗出物像素点所占的比例相比于整个背景区域来说非常小,模型训练中,造成样本比例极不均衡,为了更好的处理这一问题,**Guo** 等人^[39]提出 **bin loss** 损失训练模型,解决了类平衡交叉熵损失 (**CBCE loss**) 所带来的损失不平衡问题,并取得了不错的效果。**Guo** 等人^[40]提出一种轻量级的病灶分割网络,该网络以编码-解码结构为主体框架,在编码端以多尺度的方式进行特征提取同时采用上下文模块和跳跃连接来保证梯度的正常传播。**刘梦颖**等人^[41]在 **U-Net** 模型的基础上,结合 **ResNet** 核心结构,构成 **ResU-net** 病灶分割模型,利用 **DIARETDB1** 数据集训练并测试模型,实验结果表明,实现了对糖尿病视网膜病变区域进行较好的识别分割。**Costa** 等人^[42]将多实例学习引入模型中,将优化实例编码与糖尿病视网膜病变等级分级任务相结合,提升模型对 **DR** 图像的中级特征的学习能力,利用大量的中级特征实现对病灶的精确检测,精准率达到 96%。**Wang** 等人^[43]提出了一种基于深度学习的数学形态学优化方法,用来识别渗出物的候选区域,然后将图像的几何、纹理和灰度特征进行融合,最后利用随机森林算法从候选区域中分割出渗出物。

综上所述,糖尿病视网膜渗出物的分割已由传统的图像处理方式发展到近期的基于机器学习的分割,再发展到目前的基于深度学习模型对眼底图像区域检测分割,国内外

学者已经取得了不错的成绩。但该方法仍存在一些问题亟需解决，基于深度学习的全监督训练模型虽然获得了显著效果，但若将其转移到非完全监督模型后，并不能得到较好地效果；深度模型的训练参数应该被压缩，轻量级的模型更适合移动端的嵌入式设备，充分发挥其诊断功能；另外，糖尿病视网膜眼底图像环境复杂，多种病灶大小不一、形状不同且分布无规律，部分病灶还与眼底组织属性相似，导致现有的技术对一些特殊病灶的区域识别分割精度不高。因此，本文基于深度模型构建高特异性和灵敏度的分割模型，以提高模型对病变区域分割性能。

1.4 主要研究内容

本课题基于深度学习的方式对糖尿病视网膜病变图像进行了一系列的分析研究，设计算法模型自动识别病变图像对应的类别，自动实现对病变图像渗出物的分割。算法训练中无需人工干预，实现端到端的图像分析，从而起到辅助医疗的作用。本文针对分类和分割任务分别设计了两种完全不同的网络模型，经过实验测试证明，两种模型均具有较强的泛化能力，主要研究内容如下几点：

(1) 主要研究了基于深度学习的图像分类与分割方法，尤其关注学习医疗图像领域中的相关深度模型。探究深度模型的运行机制，深入学习卷积神经网络相关知识和图像特征提取的原理。在实验室服务器环境下搭建深度学习运行环境，熟悉深度学习框架，搭建用于图像分类和分割的基本模型框架，同时训练基本模型并学习模型优化等工作。

(2) 由于糖尿病视网膜病变图像的分类与渗出物分割公开数据集数量有限，而深度模型训练任务需要大量的数据集支撑。为此，针对本课题使用的分类、分割数据集进行一系列的预处理，如必要的质量筛选，对比度调整、图像平滑去噪、数据增强等方法。同时采用图像扩增技术对数据进行平衡扩增，克服数据量不足和分布不均衡问题。

(3) 对于 DR 分类问题，本文基于深度学习设计糖尿病视网膜病变分类模型，提出了一种集成 MobileNetV2 和 EfficientNetB0 深度模型的注意力网络：ME-ANet。将集成模型作为特征提取器，训练中利用迁移学习的策略对模型参数进行初始化，既实现了模型的加速训练，又有效避免了模型训练中的过拟合问题。同时改进设计注意力机制并分别嵌入到集成模块中，利用该模型实现对糖尿病视网膜病变的五分类任务。模型实现了对图像浅层信息的特征融合提取，减少了微病灶特征信息在训练过程中的卷积丢失问题。注意力机制进一步强化了模型对目标特征的学习，最后采用 Kaggle 数据集对模型进行训练，实验证明本文提出的网络结构具有较好的分类性能。

(4) 在糖尿病视网膜病变分类的基础上，实现对关键病灶点渗出物的精确分割，将有效提高计算机辅助医疗诊断的效率。为此，本文提出一种基于深度学习的糖尿病视网膜渗出物的分割算法，以 U-Net 编码器-解码器结构为基本网络架构，并使用 ResNet50 作为编码端的特征提取器，实现了特征提取网络的加深，避免了模型训练中梯度消失的

问题；在解码端设计注意力层结构，并嵌入到解码单元中，提升模型对小病灶点的识别能力；最后在模型解码端构建多特征尺度融合层，实现模型多尺度语义特征的学习。实验过程中利用 IDRID 数据集进行训练，训练时采用 Focal loss 焦点损失函数，强化模型对样本中小目标物体特征的学习能力。与其他渗出物分割算法相比，该算法分割性能较优，且具有普遍适用性。

1.5 本文组织结构

本文共分为五个章节：

第一章 绪论。首先介绍了本课题的研究背景和意义，然后对糖尿病视网膜的病变机制和病变类别进行描述。在国内外研究现状中，主要针对糖尿病视网膜病变分类和渗出物分割两个研究点，介绍近年来国内外的相关研究成果。最后对本文的研究内容和组织结构进行概述。

第二章 相关工作。首先从神经网络的发展历史、人工神经元和激活函数三个方面对神经网络进行介绍，然后详细描述了卷积神经网络中的三个核心层：卷积层、池化层和全连接层的工作机制，同时介绍了注意力机制的相关概念以及典型注意力模型。最后，对目前主流的深度学习框架进行简要概述，并简要分析框架的优缺点。

第三章 基于深度学习设计构建糖尿病视网膜病变分类模型：ME-ANet。在数据准备过程中，利用 Kaggle 竞赛平台公开的 DR 数据集，对该数据集进行一系列的预处理，获取最终用于模型训练的数据。利用处理过的图像训练本章构建的 DR 分类模型，训练过程中采用迁移学习的策略，最后通过实验获得的评价指标对模型进行评估。

第四章 实现 DR 图像渗出物区域的精确分割。本章改进基于 U-Net 编码器-解码器框架的深度分割模型。采用 IDRID 公开数据集训练模型，同样需要先对该数据集进行相应的预处理和扩增处理。训练时采用 Focal loss 焦点损失函数，强化模型对样本中小目标物体特征的学习能力。最后分析模型的渗出物分割效果并与专家的渗出物标注作对比，同时利用相关评价指标分析模型的分割性能。

第五章 总结与展望。对本文的研究工作进行总结，具体针对分类和分割两个研究点，分析总结模型在具体任务中的完成效果，同时指出本课题研究工作中存在的不足及一些对未来研究工作的展望。

第 2 章 相关工作

2.1 神经网络相关知识

2.1.1 神经网络的发展

人工神经网络（Artificial Neural Network, ANN）是指一系列受到生物科学和神经科学启迪而设计的数学模型，这些模型主要通过对人脑的生物神经元网络进行抽象，构造人工神经元模型，并根据特定的连接模式来创建神经元之间的相互连接，多个神经单元连接组成的网络用来实现数据的传递，达到一种模拟生物神经网络的效果^[44]。

人们对于神经网络的研究从上世纪 40 年代就已经开始，Warren McCulloch 和 Walter Pitts 两人在 1943 年最早提出了具有单一运算功能的人工神经网络，将具有二进制阈值的神经元与布尔逻辑进行比较，被称为 MP 模型，从那以后，研究者对神经网络的研究不断深入创新。1958 年，Rosenblatt^[45]后来在对 MP 模型进行改进，设计了具有模拟人类感知能力的感知器（Perceptron）模型，能够让计算机学习如何区分左侧或右侧标记的卡片。1962 年 Hubel 和 Wiesel^[46]设计并验证多种视觉感知实验，并结合生物神经元的研究，首次提出了感受野（Receptive field）的概念，这一概念的提出进一步推动了神经网络的发展。1974 年，Paul Werbos^[47]发明反向传播算法（BackPropagation, BP），并在他的博士论文中指出其在神经网络中的应用，遗憾的是，这一算法在当时并没有得到广泛的应用。1980 年，福岛邦彦^[48]受到生物细胞感受野的启发，提出了新知机（Neocognitron）的结构，该网络使用分层、多层设计，这种设计允许计算机学习识别视觉模式，并采用无监督学习的方式训练，这一发明也没有得到足够的重视。20 世纪 80 年代中后期，反向传播算法重新进入研究者的研究视野当中，LeCun 等人^[49]将反向传播算法引入卷积神经网络中，指出如何在反向传播算法中使用约束并将其集成到神经网络架构中来训练算法，后来这项研究成功地利用神经网络识别美国邮政服务提供的手写邮政编码数字。在 2006 年，Hinton 等人^[50]第一次提出了深度学习的概念，后来构造深度信念网络模型（Deep Belief Networks, DBN），并采用逐层预训练和反向传播算法对模型训练，成功解决了梯度传播过程中的弥散现象。至此，深层神经网络的训练方式逐渐成熟，极大推动了神经网络的发展。近年来，随着计算机科学的快速发展，各类大数据集也被公开，再加上大型科技互联网的大量投资，使得深度神经网络成为当代计算机前沿科技的重要组成部分，也成为各计算机领域与其他领域交叉的重要技术媒介。图 2.1 展示了本文描述神经网络发展中的时间点和成果。

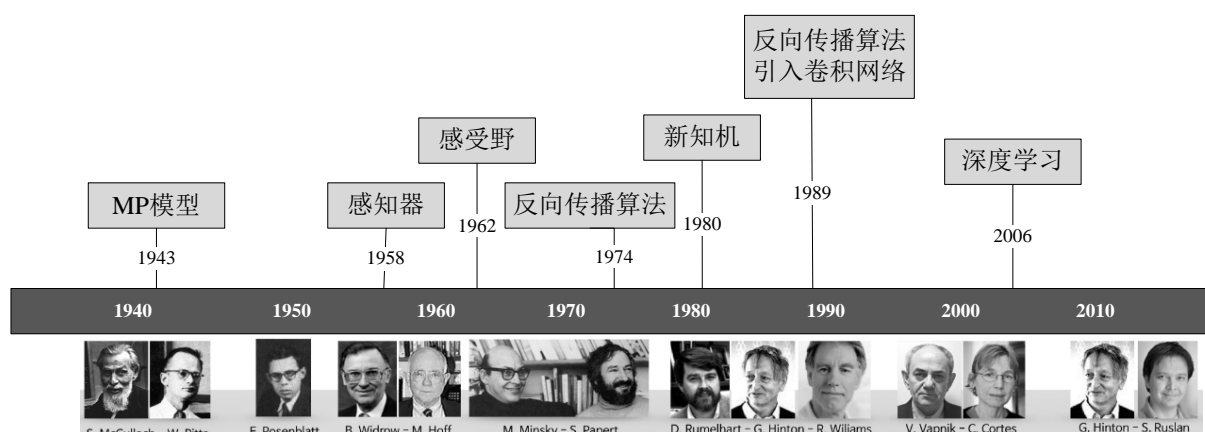


图 2.1 神经网络发展历程图

2.1.2 人工神经元

神经网络是由多个神经元相互连接组成，其结构与生物组织中的神经元构成形式类似。将不同的权重分配给不同的神经元结点，这些权重决定上一层神经元对下一层神经元的干预程度。单个神经单元中有一个特定的函数，即激活函数，来自其他神经单元中的信息经过相应的权重求和后，输入到激活函数中，生成一个新的激活值。神经元是否转变为激活状态是由激活值与该函数阈值比较决定的，当激活值大于提前设置的阈值时，神经元就转变为激活状态，否则神经元不会改变状态，图 2.2 为神经元结构示意图。

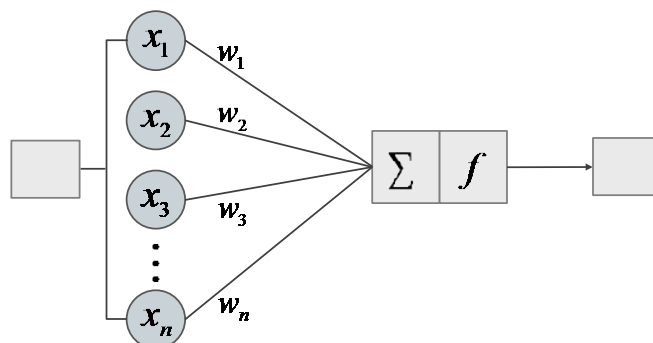


图 2.2 神经元结构示意图

上图中，神经元的输入端为 x_1, x_2, \dots, x_n ，对应上一层的连接权重为 w_1, w_2, \dots, w_n ，所有输入通过相应加权计算后求和，将值传递给激活函数，如果输入值高于某个阈值，则会触发激活条件得出激活函数对应的输出值。若激活函数用 f 表示，偏置用 b_i 表示，最后产生神经元的输出用 y 表示，则神经元加权求和公式如式 (2.1) 所示。

$$y = f\left(\sum_{i=1}^n w_i x_i + b_i\right) \quad i = 1, 2, 3, \dots, n \quad \text{公式 (2.1)}$$

2.1.3 常用的激活函数

线性激活函数的激活与输入呈正比，数学表达如下式 (2.2) 所示。线性激活函数主要有两个问题：在反向传播过程中无法使用，这是因为线性激活函数的导数是一个常量，其值与输入变量 x 是无关的；若在神经网络中均使用线性激活函数，仅仅相当于将网络中所有的层结构进行简单的线性叠加，无法实现数据的复杂映射。

$$f(x) = x \quad \text{公式 (2.2)}$$

为了让神经网络具有非线性建模能力，卷积操作之后添加非线性激活层是非常必要的。神经元中的激活函数只有是非线性激活函数，才能得到最后的非线性输出。具体来讲，非线性激活函数允许反向传播，其导函数的值与输入是相关联的，并且可以反馈给网络哪些神经元节点中的权重值可以提供更佳的预测。另外，由于输出是通过多层输入的非线性组合，故可以允许神经元的多层堆叠，组建更加深层次的网络。下面介绍 4 种常见的激活函数，函数曲线如下图 2.3 所示。

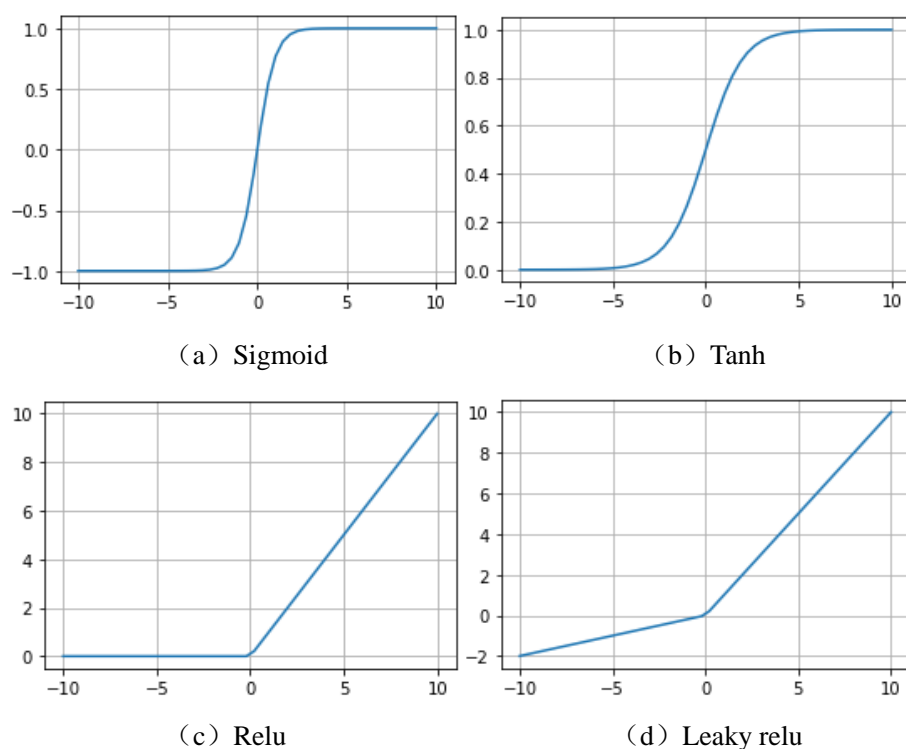


图 2.3 常用的激活函数

Sigmoid 函数将任何实数值作为输入并输出 0 到 1 范围内，且函数具有对称性如图 2.3 (a) 所示，输入值越大，输出值约接近于 1，而输入值越小，输出值越接近于 0。使用该激活函数的神经单元的输出值可以看作概率分布，使神经网络更好地和统计模型进行结合。该函数在神经网络中的最大缺陷是易饱和，输入值在 $[-5, 5]$ 区间外的值，其导数接近于 0，导致模型权重无法更新。函数数学表达式如式 (2.3) 所示：

$$f(x) = \frac{1}{1 + e^{-x}} \quad \text{公式 (2.3)}$$

Tanh 函数如图 2.3 (b) 所示, 与 sigmoid 函数相似, 输出函数值在-1 到 1 之间, 以零为中心, 可以将其输出值映射为正、中、负三种状态。另外, tanh 激活函数虽然克服了 sigmoid 函数只能得到正值的问题, 但是无法彻底解决模型训练中的梯度消失现象。由上图观察可知, tanh 函数的梯度比 sigmoid 要陡峭的多, 也进一步表明 tanh 的非线性要比 sigmoid 更好。其数学表达式如式 (2.4) 所示:

$$f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad \text{公式 (2.4)}$$

Relu 函数如图 2.3 (c) 所示, 是目前常用的激活函数之一。尽管它给人一种类似于线性函数的印象, 但 relu 具有导数函数, 并允许反向传播, 同时使其计算效率更高。当函数的激活输出值为 0 时, 该神经元不会转变为激活状态。也就是说, 在使用 relu 激活函数的神经网络区域中, 神经元不会同时被激活, 在反向传播过程中大大减少了梯度的计算更新, 计算效率比 sigmoid 和 tanh 激活函数要高得多。另外, 由于其线性和非饱和特性, 该函数加速了梯度下降过程, 实现向损失函数的全局最小值的快速收敛。Relu 函数的表达式如式 (2.5) 所示:

$$f(x) = \max(0, x) \quad \text{公式 (2.5)}$$

Leaky relu 函数如图 2.3 (d) 所示, 它是 relu 函数的改良版本, 用一个非水平线简单地代替了水平线, 当输入为负值时, 添加一个非零斜率 ε , ε 为极小负值, 使左侧的梯度变为非 0 值, 则该区域的神经元不会成为死神经元。虽然改善了死神经元的特性, 但是损失了稀疏性。其数学表达式如式 (2.6) 所示:

$$f(x) = \max(\varepsilon x, x) \quad \text{公式 (2.6)}$$

2.2 卷积神经网络

在过去几十年中, 深度学习凭借计算机存储、计算性能的不断提升, 已能够轻松地处理海量数据。目前, 卷积神经网络 (CNN) 已经成为最受欢迎的神经网络之一, 常用于处理分类和计算机视觉任务, 下图 2.4 展示了卷积神经网络典型的分类架构。在 CNN 之前, 一般使用手工特征提取方法来对图像中的多种目标进行识别。然而, 卷积神经网络现在为图像分类和对象识别任务提供了一种更具可扩展性的方法, 将线性代数中的矩阵研究应用到图像像素点的相关计算中来, 完成图像中的具体特征的识别和提取工作。当然, 这种方式对计算机性能的要求是非常高的。

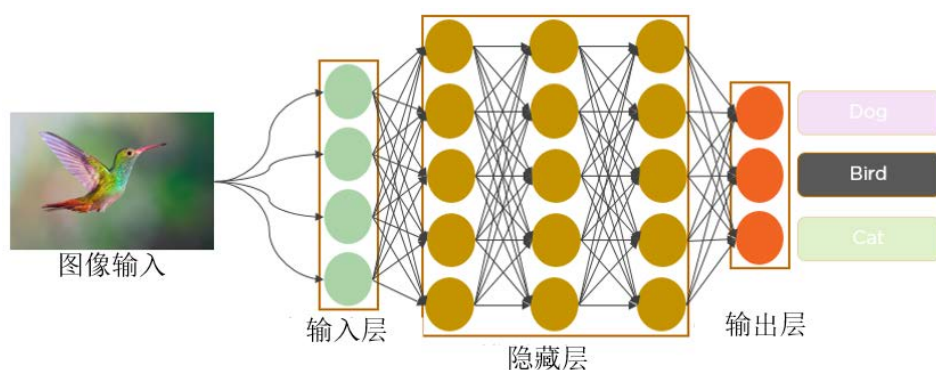


图 2.4 典型的卷积神经网络分类模型

自 1950 年代人工智能早期以来，研究人员一直在思考如何构建一个智能系统，实现计算机与现实视觉数据的交互，这个领域后来被称为计算机视觉。2012 年，多伦多大学的研究人员设计出 AlexNet 深度模型，该模型超越了当时性能最佳的图像分类方法，在 2012 年 ImageNet 计算机视觉竞赛上，该模型达到 85% 的准确率，斩获第一名，而当年的第二名仅仅得到 74% 的准确率。AlexNet 的核心是卷积神经网络，该网络与其他深度网络不同的是，其在图片、语音等方面具有无可替代的优势。卷积神经网络是一类具有局部连接和权重共享的深度前向神经网络，其最核心的三种层结构分别是卷积层、池化层和全连接层，具体结构如图 2.5 所示。

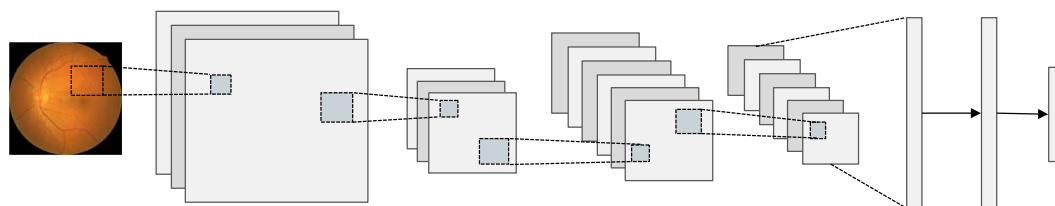


图 2.5 卷积神经网络结构示意图

2.2.1 卷积层

卷积层，顾名思义，它实现的是对图片的卷积操作。随着网络层数的加深，CNN 的复杂性也在不断增加。在卷积神经网络中，前面的浅层卷积主要学习简单的特征，如物体的颜色和边缘等，在网络深层卷积中，则主要学习更加抽象的、全局的特征，如物体的部分结构和形状，直到最终识别出预期的物体。如下图 2.6 所示，由下至上可以理解为模型通过卷积层捕获到的浅层特征、中层特征和高级特征，低层零件部分在神经网络中构成了一个较低级别的特征，其余不同程度的零件组合表示不同高级程度的特征。总之，卷积神经网络中处于不同深度的卷积层会学习到不同层次的特征，它的厉害之处就在于其强大的特征提取能力。

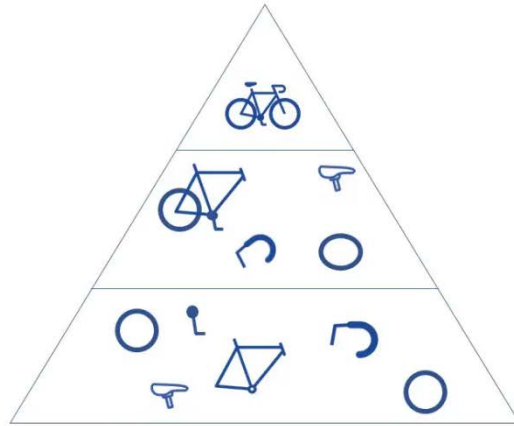


图 2.6 不同卷积所对应的特征信息举例图

在提取特征时，大多数的计算都在卷积层进行，而卷积核是它的主要特征提取结构。在卷积神经网络中，利用卷积核与输入的特征图进行运算，得到对应的特征矩阵。一般卷积核是 3×3 矩阵，它的尺度决定了模型感受野的大小。在卷积操作中，卷积核应用于图像的某个区域，通过计算输入像素值与卷积核之间的点积，并将计算结果并记录在输出矩阵中。之后，卷积核按照预设的步长移动，重复上述计算过程，直到扫过整个图像，最后的输出矩阵又被称为特征图。卷积计算的公式如式（2.7）所示。

$$\text{conv}(x,y)=\sum_{i=0}^m \sum_{j=0}^n f(x+i,y+j)g(i,j) \quad \text{公式 (2.7)}$$

其中， $f(i,j)$ 表示特征图的输入， $g(i,j)$ 表示卷积核。为了方便表示，本文采用单通道灰度图来演示计算过程，具体过程如下图 2.7 所示，输入特征图为 5×5 矩阵，卷积核大小为 3×3 ，输出特征图的大小为 3×3 ，对于输入特征数据，滤波器按照一定的移动步长滑动，并依次与滑过的特征区域进行卷积计算。

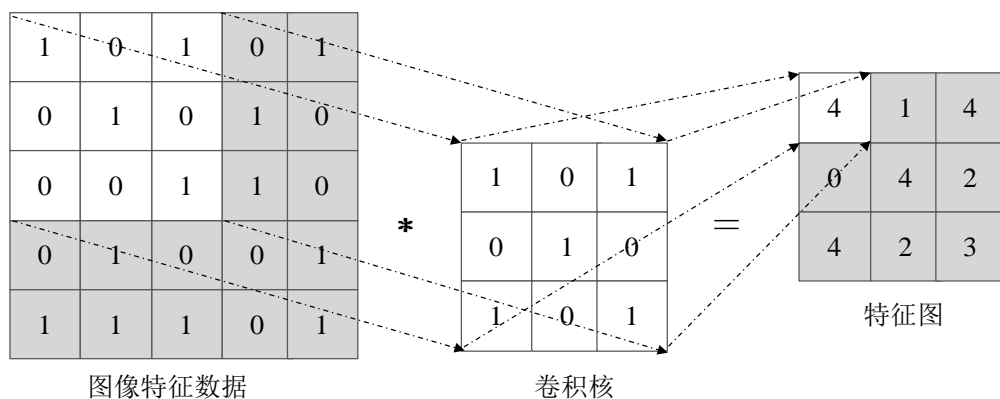


图 2.7 卷积计算过程示意图

如上图所示，特征图中的每个输出值对应输入图像的局部区域，而不必连接到输入图像中的每个像素值。另外，卷积核的矩阵参数在图像上移动卷积计算时保持固定，该

特性被称为参数共享。输出特征的尺度由卷积核数量、步幅和填充方式决定，卷积核的数量影响输出的深度，例如三个不同的卷积核将会产生三个不同的特征图；步幅是卷积核在输入矩阵上移动的距离，较大步幅会产生较小的输出；填充方式分为有效填充、相同填充和全填充三种方式，有效填充又叫无填充，在这种方式下，如果卷积核在图像边缘无法对齐，将不会执行卷积操作，相同填充能够实现输出特征与输入特征具有相同的维度，完全填充是通过对输入特征的边缘进行零点填充，从而实现对输出特征图尺寸的控制。

2.2.2 池化层

池化层，也称为下采样，一般处于卷积层之后，它能够减小卷积后输出特征的尺度大小，以此来降低模型训练中所需要的算力，这不仅节省了计算机的资源，还有能够加速模型的训练。同时，池化层也能降低图像的噪音，使各个特征表示更加稳健。池化操作就是选择一个类似于卷积核的滤波器，但是在该滤波器中不分配任何权重，而是将聚合函数应用于感受域内的值。池化操作分为最大池化和平均池化，如果在指定窗口大小内取最大值，则为最大池化，在指定窗口大小内取平均值，则为平均池化。在平均池化中，只是简单地进行区域平均值来进行降维，图像去噪效果较差；最大池化在区域中取最大值进行降维，具有较好的图像去噪效果，模型训练中最大池化的使用更为常见。两种池化过程如图 2.8 所示，其中，池化窗口大小设置为 2×2 ，步长设置为 2。

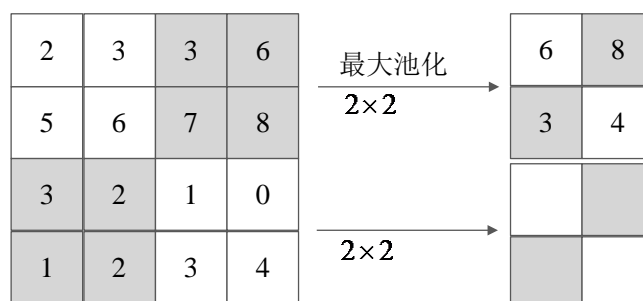


图 2.8 池化操作示意图

2.2.3 全连接层

卷积层和池化层均为多维矩阵，而全连通层则更为直观、易于理解。在全连通层中，各结点与上一层的结点直接相连，它将上一层的输出维度拉伸成一维向量。全连接层本质上就是传统的前馈神经网络结构，一般出现在卷积神经网络的末端，该层通过前几层提取的特征进行组合分类，一般使用 softmax 激活函数对输入生成相应的类别概率值。具体全连接层结构如下图 2.9 所示。

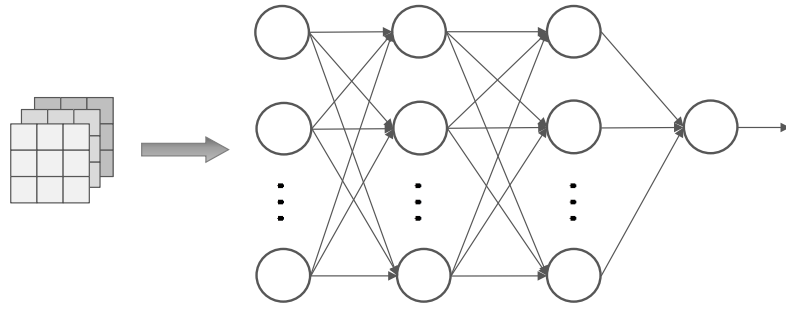


图 2.9 全连接层结构示意图

上图中，来自前一层卷积层或池化层的数据特征被展开成一维向量，由 16 个神经元构成，接着是两层前馈神经网络层，各包含 12 个和 8 个神经元，然后就是整个神经网络的输出层，输出层可以是单个神经元用于二分类任务，也可以是 softmax 层，含有多个神经元实现多分类任务，最后输出的是预测的概率值。

全连接层将学到的局部特征整合到一维特征空间，根据高层抽象特征的学习组合，假设全连接层的上一层有 n 个神经元作为输入， (x_1, x_2, \dots, x_n) 为神经元的输入，且下一层的输出有 m 个神经元，且 (y_1, y_2, \dots, y_m) 为其对应的神经元节点^[41]，则计算过程如式(2.8)所示。

$$\begin{pmatrix} w_{11} & \cdots & w_{1n} \\ \vdots & \ddots & \vdots \\ w_{m1} & \cdots & w_{mn} \end{pmatrix} \times \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} + \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix} \quad \text{公式 (2.8)}$$

直观地说，卷积层提取的是局部视野，而全连接层能够利用卷积层提取的特征进行特定任务的判断，全连接层的功能是把所学到的分布特性的表现映射到样本空间中。

2.3 注意力机制

注意力是一种非常复杂的人类认知功能，即人们可以在对某些信息进行关注的时候，忽视其他信息，而专注于对个人有帮助的关键信息。在日常生活中，人们通过视觉、听觉、触觉等多种形式接收大量的信息^[51]，而大脑能够轻松的对这些日常海量的信息进行有效的处理，使我们能够有节奏的处理问题，这是因为人脑可以有针对性地选择小部分消息，而忽略大部分无用的消息^[52]，这种能力被称为注意力。在有限的计算资源中，利用注意力机制对计算资源进行高效的配置^[53]，实现将有限的计算资源应用在处理关键信息的过程中。特别是在计算机视觉中，为了提高模型对特定任务中关键特征的学习，减少对无关特征的关注，引入注意力机制效果显著。

在计算机领域中，注意力机制是由 Bahdanau 等人^[54]在 2014 年首次被提出，在那之

后注意力机制不断升级，并逐渐被应用在各种模型中用于提升性能。由于注意力机制相比于 CNN 和 RNN 复杂度较低且训练参数少，而且还能够学习全局信息和局部信息，近年来被经常应用在分类和分割模型中。在自然语言处理任务中，注意力机制表现出独特的优势，实现了高效准确的完成对关键文本词汇识别分析任务。后来随着计算机视觉地不断发展，注意力机制被应用到该领域进行视觉信息优化处理任务，之后便得到广泛关注。在深度模型中，注意力机制一般是作为一个独立的模块被嵌入到模型中，通过控制输入网络的特征信息或分配特定的权重来实现对大量信息的筛减。

2.3.1 SE 注意力机制

2018 年，Hu 等人^[55]在 CVPR 上发表了一篇名为 Squeeze-and-Excitation Networks (SENet) 的论文，这篇论文成为注意力机制领域最具代表性的论文之一。它可以作为 CNN 结构中一个全新的独立模块，它提高了通道间之间的相关性，而不需要耗费太多的计算资源。SE 注意力的核心设计理念是在卷积层的输出特性图中加入一些参数，以便网络可以自适应地调整每个特征图的权重，使其对重要特征变得更加敏感，同时忽略不相关的特征。CNN 通过卷积操作从输入图像中捕获不同程度的特征信息，在创建特征图时，网络对每个通道的权重相等，这意味着每个通道在训练过程中，都被无差别对待的保留下来，这不利于提高模型的特征表示能力。

SENet 注意力机制由挤压模块 (Squeeze) 和激励模块 (Excitation) 两部分组成^[56]，如图 2.10 所示。在挤压模块中，挤压操作用于从特征图的每个通道中提取全局信息。为了使通道注意力适应每个通道本身，同时又为了避免大量像素值的计算，需要使用特征描述符将每个特征图的信息分解为奇异值，以此来降低整个操作的计算复杂度，这也是进行挤压操作最根本的目的。假设给定一个输入张量 X ，第 c 个通道的挤压过程公式(2.9)所示：

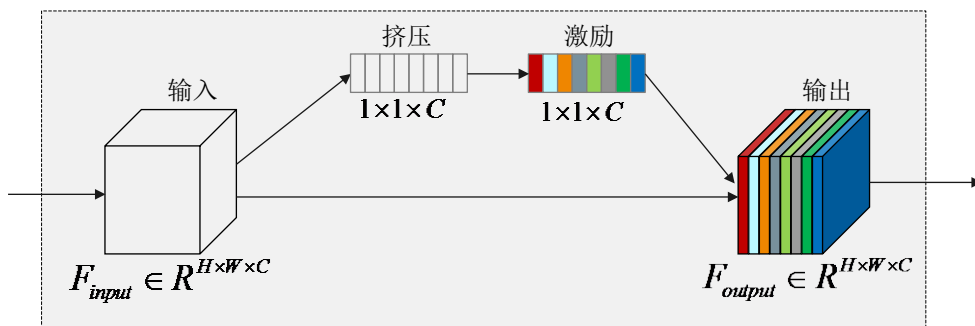


图 2.10 SENet 注意力结构

$$z_c = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W x_c(i, j) \quad \text{公式 (2.9)}$$

其中 Z_c 是与第 c 通道相关的输出, 由于输入 x 来自一个固定卷积核大小的卷积层输出, 即可将其视为局部特征描述的集合, 故挤压操作能够收集全局特征信息。

卷积神经网络中的池化一般用于减少空间大小, 通过消融实验, 在激励模块中使用全局平均池化 (GAP) 操作效果较优, 通过取该特征图中所有像素的平均值, 基本上将整个特征图减少到一个奇异值。就维度而言, 若输入张量维度为 $B \times H \times W \times C$, 则处理后输出张量的维度变为 $B \times 1 \times 1 \times C$, 在特征图中 B 表示批量大小, H 表示宽度, W 表示高度, C 表示通道数。在激励模块中, 使用具有瓶颈结构的全连接多层感知器 (MLP), 如图 2.11 所示, MLP 用于生成权重以自适应地缩放特征图的每个通道。另外, 采用两层完全连接的方法, 不但可以提高系统的非线性, 而且能够很好地适应通道的复杂性。

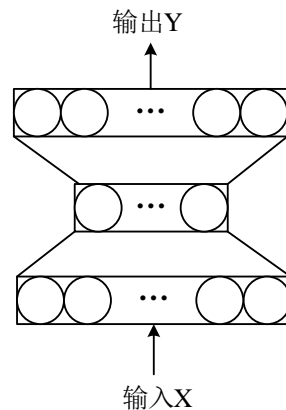


图 2.11 MLP 结构示意图

MLP 由三层组成, 输入的特征为挤压模块的输出, 大小为 $B \times 1 \times 1 \times C$, 输入层具有 C 个神经元, 在隐藏层中使用缩减率 r (默认值为 16) 来减少特征数量, 此时的尺度为 $B \times 1 \times 1 \times C/r$, 隐藏层有 C/r 个神经元, 最后在输出层, 进行尺度恢复, 神经元个数又恢复为 C 个。

通过挤压和激励两个操作后, 得到被 sigmoid 激活过生成的矩阵, 矩阵值被归一化到 0-1 之间, 最后将该矩阵与输入特征图之间进行逐元素相乘, 即得到最后的注意力特征图。

2.3.2 CBAM 注意力机制

在 SE 注意力机制中, 为了降低计算复杂度, 在激励层添加 MLP 瓶颈结构, 按照一定的比例缩减通道数, 然而, 这一操作容易导致特征信息的丢失, 不是最好的处理方式。其次, SE 注意力使用全局平均池化进行空间压缩, 意味着所有的像素都均匀地映射在特征图的空间域中, 无法区分像素值的类别和重要程度, 进一步造成特征信息的丢失。为解决上述问题, Woo 等人^[57]提出基于卷积块的注意力机制 (Convolutional Block Attention Module, CBAM), 从结构上看, 它将空间注意力机制和通道注意力机制进行

了结合；从效果上看，该注意力机制可以提高分类检测的准确率。**CBAM** 注意力机制中的通道注意力结构如图 2.12 所示，它缓解了全局平均池化操作所带来的一些影响。其核心思想如下：给定一个特征输入 $F \in R^{H \times W \times C}$ ，为了提高通道注意力的效率，必须先压缩输入特征图的空间维度，一般使用平均池化来聚集特性空间的信息。在 **CBAM** 中，同时也使用了最大池化操作，有利于捕获特定对象特征的信息，通过以上两步操作最终实现空间特征的压缩，得到两个不同的空间语义描述符： F_{\max}^c 和 F_{avg}^c 。随后将这两个描述符转发到共享网络以生成通道注意力图，共享网络是由一个具有隐藏层的多层感知器组成，且隐藏层缩减率为 r ，将共享网络的输出逐元素求和得到最后输出特征向量，即通道注意力向量 $M_c \in R^{1 \times 1 \times C}$ ，计算过程如式 (2.10) 所示：

$$\begin{aligned} M_c(F) &= \sigma(MLP(AvgPool(F)) + MLP(MaxPool(F))) \\ &= \sigma(W_1(W_0(F_{\text{avg}}^c)) + W_1(W_0(F_{\max}^c))) \end{aligned} \quad \text{公式 (2.10)}$$

其中 σ 表示 *sigmoid* 激活函数， $W_0 \in R^{C/r \times C}$ ， $W_1 \in R^{C \times C/r}$ ， W_0 ， W_1 均表示多层感知机的权重，在 W_0 后使用了 **ReLU** 作为激活函数。

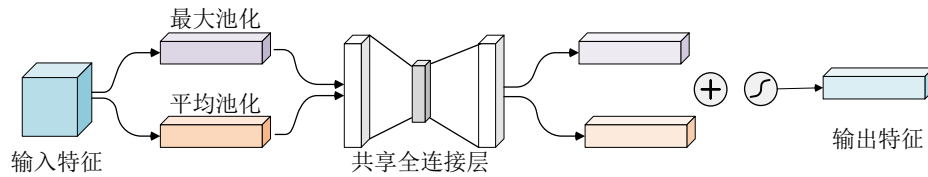


图 2.12 CBAM 通道注意力结构图

对于空间注意力分支结构如下图 2.13 所示，空间注意力是对通道注意力的补充，它能够识别到哪些是信息丰富的部分。其原理如下：假设给定一个输入 $F \in R^{H \times W \times C}$ ，首先沿着通道轴应用平均池化和最大池化操作聚合特征图的通道信息，即得到两种不同的通道特征描述符： F_{avg}^s 和 F_{\max}^s ，再将两者进行拼接，然后使用一个卷积核大小为 7×7 的卷积操作，再经过激活函数层激活，最后得到空间注意力向量 $M_s \in R^{H \times W \times 1}$ ，它编码了要强调或抑制的位置，具体描述如式 (2.11) 所示。

$$\begin{aligned} M_s(F) &= \sigma(f^{7 \times 7}([AvgPool(F); MaxPool(F)])) \\ &= \sigma(f^{7 \times 7}([F_{\text{avg}}^s; F_{\max}^s])) \end{aligned} \quad \text{公式 (2.11)}$$

其中， σ 表示 *sigmoid* 操作， 7×7 表示卷积核的大小。

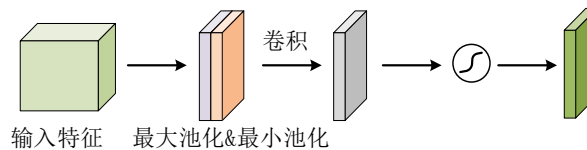


图 2.13 CBAM 空间注意力结构图

CBAM 注意力机制的整体结构如图 2.14 所示。从上文可知，输入特征图 F 分别经过通道注意力模块和空间注意力模块得到 $M_c(F) \in R^{1 \times 1 \times C}$ ， $M_s(F) \in R^{H \times W \times 1}$ 两个特征图，最后 CBAM 注意力机制的输出用 F_{CBAM} 表示，计算公式如式 (2.12) 所示，式中 \otimes 表示逐元素相乘。

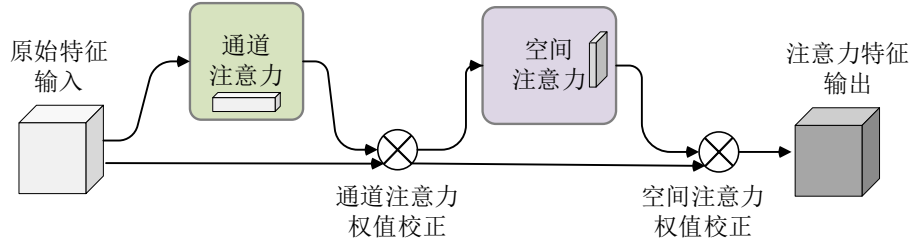


图 2.14 CBAM 注意力整体结构图

$$\begin{aligned} F' &= M_c(F) \otimes F \\ F_{CBAM} &= M_s(F') \otimes F' \end{aligned} \quad \text{公式 (2.12)}$$

综上所述，CBAM 的注意力机制既考虑到了通道的重要性，也考虑到了相同的位置对其重要性的影响，经过在两种不同层面上的特征加强效果进行融合，使网络能够达到对目标特征信息的学习，同时抑制不相关特征信息的效果^[58]。

2.4 深度学习主流开源框架

深度学习框架用于创建深度和机器学习模型，该框架通过简化机器学习算法为模型的设计构建和训练提供了一系列的工具和方法。这些框架包括接口、库和工具，通常是开源的，可以实现快速构建模型并训练模型。总之，深度学习框架使研究人员和开发人员能够快速稳健地实现最先进技术的研究设计^[59]，有利于推动深度学习的发展和推广。迄今为止，已经出现了多种深度学习框架，在不同的研究领域被广泛应用。接下来，本文将重点介绍几种比较流行的深度学习框架。

Theano 是蒙特利尔大学 LISA 算法学院开发的第一个深度学习框架^[59]，主要用于多维数组的计算和优化，它是基于 Python 和 CUDA 语言开发，且支持多种操作系统，如 Linux、Windows 和 macOS 等。Theano 框架整合了 Numpy，可以直接使用 Numpy 等功能，用于多维数组的优化，可以提供更高效的计算，非常适合大规模神经网络的训练。但是由于该框架在底层缺少 C++ 的接口，导致在部署时需要依赖相关的 Python 工具库，也正是因为这一缺陷，使得该框架没有在工业生产上使用。

Caffe 由加州大学伯克利分校研发，基于 C++ 和 CUDA 开发，支持 Linux、Windows 和 macOS 操作系统^[60]。可直接在 CPU 和 GPU 中自由切换，可保证模型的训练效率。Caffe 是一个兼顾速度和层次化的现代深度学习框架，该框架以其速度和效率而闻名，它每天处

理超过6000万张图像，拥有完善的技术文档和活跃的开发人员社区。Caffe代码移动好理解，实现基础算法比较方便，适合工业快速部署与应用。

TensorFlow由Google Brain公开的开源机器学习库，安装十分简单方便，仅需要pip命令执行安装，不需要手动添加各种依赖库，属于最流行的深度学习框架之一。该框架由Python、C++和CUDA编写的，几乎可以在所有的平台上运行并且可以与CPU、GPU和TPU一起使用。TensorFlow最大的特点就是计算图机制，利用计算图，表示计算流，着重强调了数据的流向。利用Python接口，很容易将设计思路体现在TensorFlow中，同时该框架支持TensorBoard模块，实现模型训练曲线和数据流图等可视化效果。基于以上优点，TensorFlow深度学习框架深受研究者和互联网科技公司的青睐^[61]。

Keras是基于TensorFlow和Theano的深度学习库，它在Theano和Tensorflow的后端运行，简化了模型构建的复杂度，可通过调用深度模型搭建常用的包，自定义层和优化方法快速设计神经网络^[62]。Keras提供简单的API，并在用户使用错误时提供清晰的、具有可操作性的反馈，除此之外，可以实现在不改动任何代码的前提下实现CPU到GPU的加速训练。Keras最致命的缺陷是无法同时支持多GPU训练，缺少分布式框架。

PyTorch由Facebook人工智能研究实验室开发，建立在基于Lua的机器学习和深度算法的科学计算框架之上，使用Python、CUDA和C/C++库进行处理。PyTorch可将其拆分为Py和Torch，Py代表Python语言，Torch表示一个科学计算框架。近年来，随着Python的生态环境日趋完善，大部分的框架均已提供Python接口，在2017年，Torch团队开发了PyTorch并开源^[63]。Pytorch入门简单，代码逻辑清晰，至今，已成为非常流行的框架。

总之，没有什么框架是一劳永逸，完美无缺的。不同的框架适用的领域不同，在研究过程中，研究者需要根据自己的实际任务来选择框架。

2.5 本章小结

本章从神经网络的发展、人工神经元和常见的激活函数三个方面对神经网络进行了简要的阐述，接着对卷积神经网络相关知识进行介绍，本章重点对卷积神经网络中的核心结构：卷积层、池化层和全连接层基本原理进行了归纳和总结。对于目前深度学习领域应用广泛的注意力机制，本文对注意力思想做了概述，然后对两种典型的注意力网络进行描述分析。对几种代表性激活函数的特点进行介绍，最后对近年来典型的深度模型框架做出总结与评价。

第3章 基于 ME-ANet 的糖尿病视网膜病变分类

3.1 引言

糖尿病视网膜病变 (DR) 是在糖尿病患者中是一种非常严重的眼部疾病, 其发病率随着糖尿病患者的患病时间的增加而不断上升。DR 的潜伏期长短不一, 有的 3 到 5 年, 有的长可达 15 年, 随着时间的推移, 其发病率也会逐渐增高。根据资料显示, 在 15 年以上的糖尿病患者中, 其发病率高达 50%^[24]。在糖尿病患者中, DR 发病初期往往没有什么不适症状, 对患者的日常生活没有什么影响。但是, DR 病变恶化速度较快, 如果前期病变没有控制好, 那么患者眼部将会遭受到严重的视网膜损伤, 影响正常视力, 严重的会导致患者失明。所以, 对于糖尿病患者来说, 定期的视网膜病变检查是非常必要的。在临床中, 医生通常检查患者眼底图像进行诊断, 效率较低, 且非常耗费精力, 若医生长时间的疲劳诊断, 很容易造成误诊、漏诊的情况。特别是在我国拥有数量巨大的糖尿病患者和地区医疗资源分布不均的背景下, 实现糖尿病患者的定期筛查任务变得尤为困难^[64]。基于以上问题, 思考如何利用当前飞速发展的科技更好的解决目前医疗领域出现的问题是当代研究者应该努力的方向, 科技造福人类, 也是其终极目标之一。

早期诊断和治疗是控制糖尿病视网膜病变的关键, 凭借深度学习技术来辅助专业医生实现糖尿病视网膜早期诊断筛查成为研究者关注的重点。本章内容的大致分为三部分: 第一部分是数据集的准备工作, 使用图像处理技术对数据集进行一系列处理; 第二部分就是构建深度分类模型、训练模型, 实现对 DR 病变图像的分类; 第三部分是模型的实验结果与分析。具体来讲, 首先将大量的视网膜眼底图像作为数据集研究对象, 对获取到的图像采用相关的预处理和数据扩充方法进行数据处理, 实现数据的进一步加工。处理后的数据经过深度网络模型进行训练分类, 输出图像对应的等级分类。针对 DR 视网膜不同类别之间差异性小, 部分特征难以提取而造成模型分类精度难以提升的问题, 本文设计一种结合注意力机制与模型集成的 ME-ANet 分类模型, 以此模型对数据进行训练, 提升模型对 DR 五分类的分类精度。

3.2 糖尿病视网膜病变分类算法流程

在糖尿病视网膜分类任务中, 由于数据噪声干扰、分布不均衡等问题, 严重影响模型的训练效果, 本文首先对相关数据进行处理, 随后利用处理后的数据对网络模型进行训练。在数据预处理部分主要采用受限的直方图和高斯滤波等操作增强图像对比度。由于 Kaggle 数据集中, 不同病变等级的图像数据差别很大, 本文通过数据增强等方式扩充病变类别较少的图像, 同时通过随机采样的方式降低病变类别较多的图像数量, 实现数据集各类别数量的相对平衡。接着进行 ME-ANet 网络模型的训练, 实现糖尿病视网

膜病变的五分类任务。最后一个阶段，利用训练好的模型进行测试集的分类，最后计算相应的评价指标并作对比分析，对模型进行性能分析。DR 分类工作流程如图 3.1 所示。

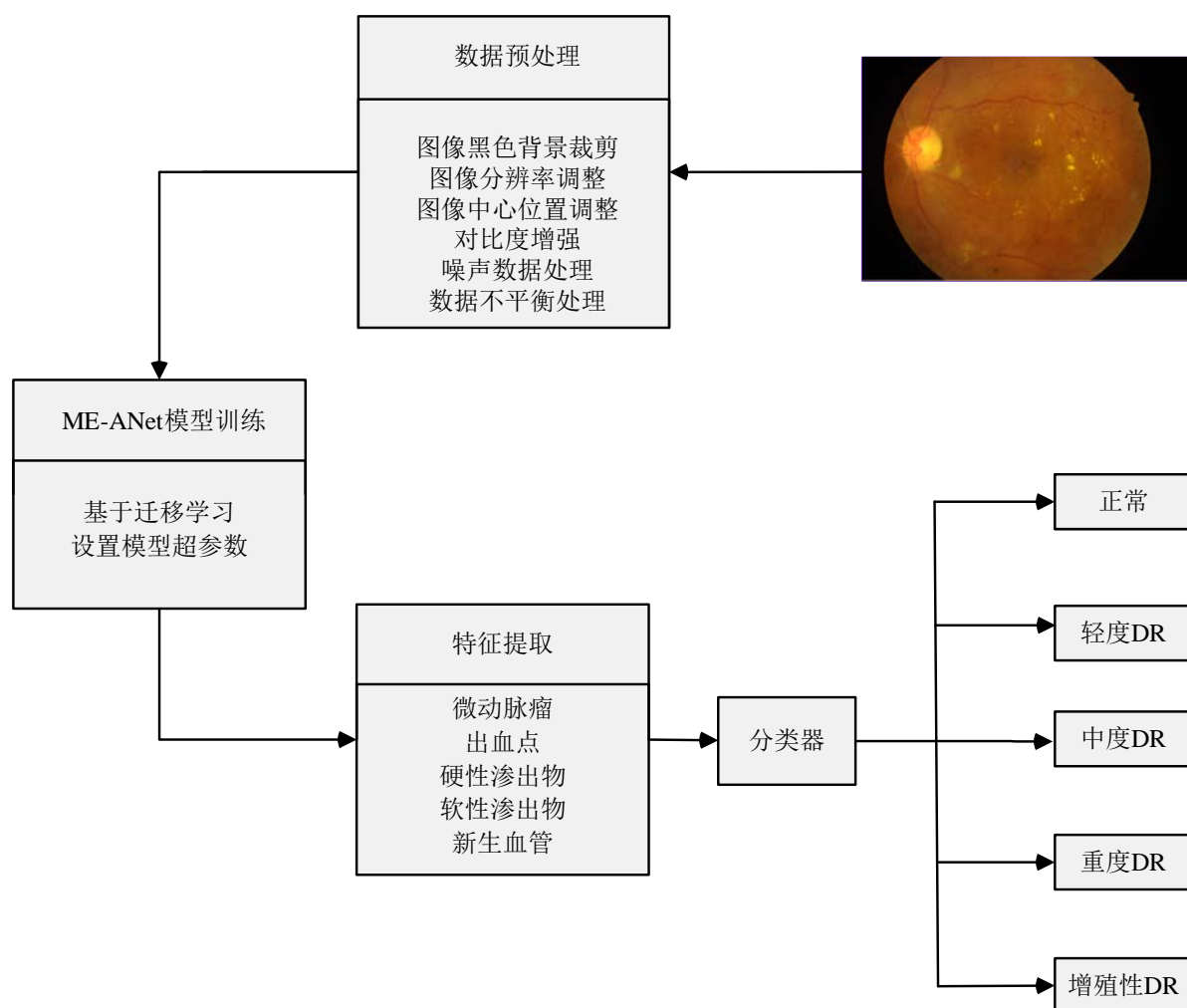


图 3.1 基于深度学习的糖尿病视网膜病变分类流程图

3.3 糖尿病视网膜眼底图像预处理

3.3.1 眼底图像数据集

本实验数据集来源于数据建模和数据分析竞赛平台 Kaggle 中的比赛 Diabetic Retinopathy Detection，数据集包含 35126 张由专业眼科医生诊断为不同严重程度的高分辨率视网膜图像^[65]，分辨率平均在 3300×2800 左右。眼科专家根据糖尿病引发的视网膜病变程度的不同，将其分为 5 个类别，分别为：DR0、DR1、DR2、DR3 和 DR4，分别对应病变程度为正常、轻度、中度、重度和增殖性，具体如图 3.2 中 (a)、(b)、(c)、(d)、(e) 所示。由下图可以看出，不同病变等级的眼底图像之间相似度较高，仅凭肉眼难以准确判断所属类别，另外，一些图片也会出现失焦、伪影或曝光不足的问题。除

此之外，该数据集分布极不均匀如图 3.3 所示，DR0 所属类别占整个数据集的比例约为 73.5%，而属于 DR4 占整个数据集的比例仅 2.0%左右，两类图像相比较 DR0 的数据量是 DR4 数据量的 36 倍左右。具体信息由表 3.1 所示。在模型训练中，保证图像的质量和分布，减少冗余信息的产生，对各类图像的均衡处理可以提高模型的分类效果，因此对眼底图像进行预处理是非常关键的一步。

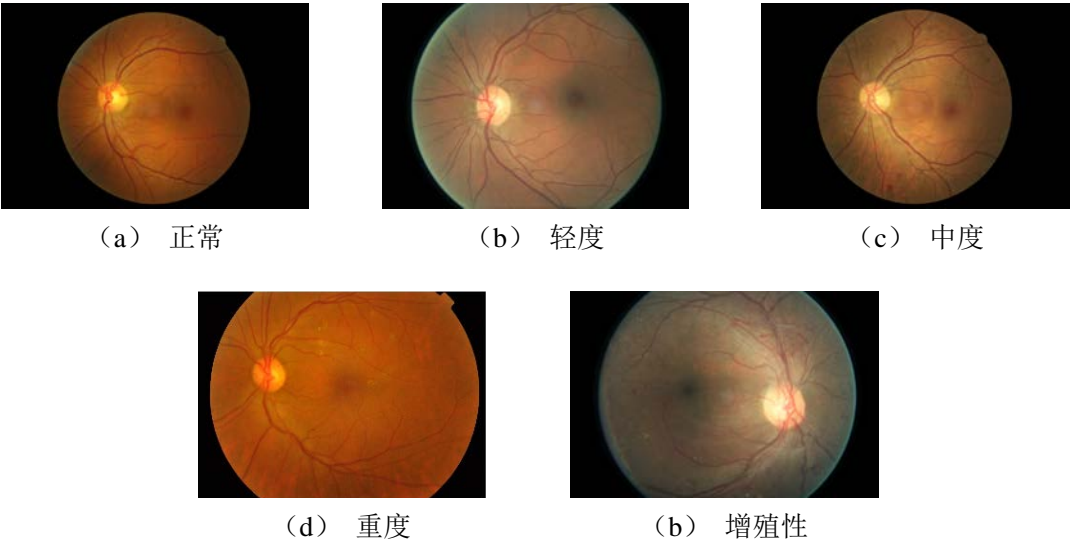


图 3.2 糖尿病视网膜病变程度的五个类别

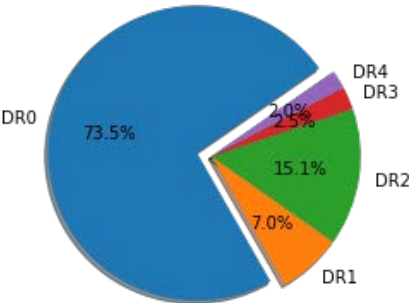


图 3.3 糖尿病视网膜病变数据集类别分布饼状图

表 3.1 数据集数据分布

类别	DR 病变等级	图片数量/张	占比/%
DR0	正常	25810	73.5
DR1	轻度 DR	2443	7.0
DR2	中度 DR	5292	15.1
DR3	重度 DR	873	2.5
DR4	增殖性 DR	708	2.0

3.3.2 数据预处理

数据的准备工作是深度模型训练的关键环节，数据质量的好坏一定程度上影响模型的训练效果。由于眼底图像采集过程中容易受到外界环境的干扰，导致图像质量参差不齐。为了解决这一问题，本文采取一系列图像预处理方法对视网膜图形进行处理，具体处理流程如下图 3.4 所示。首先手动剔除眼底图像质量较差的图像，然后将图像进行归一化处理，具体包括调整切除黑色冗余边框、调整图像分辨率等方式，接着对图像进行对比度调整；最后采用翻转、平移和偏移等增强方式处理，扩增的数据集一部分被划分到测试集，一部分经过对比度调整被划分到训练集。

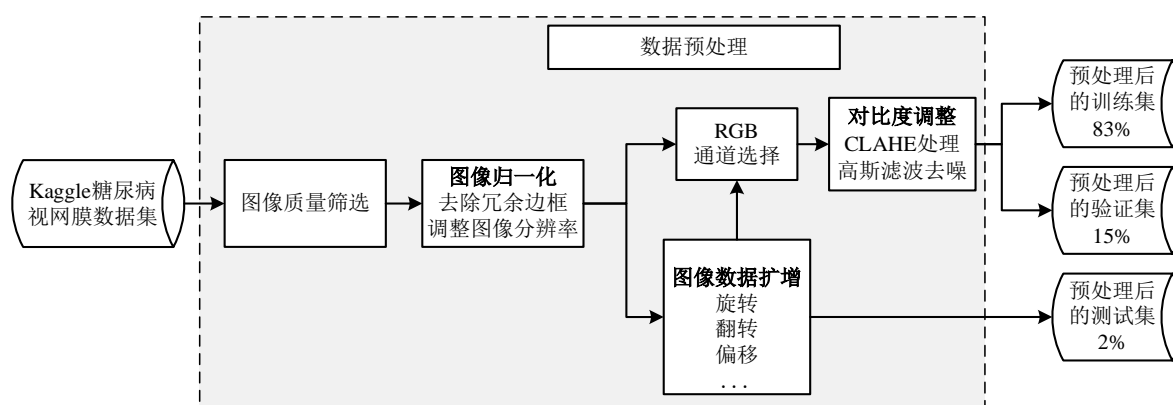


图 3.4 数据预处理流程图

3.3.2.1 图像质量筛选

糖尿病视网膜眼底图像在采集时，图像的像素、曝光、对比度等参数会随着环境、设备和人为等因素的不同而产生不同的差异。如图 3.5 所示，有些图像会出现严重的曝光过度、曝光不足、镜头污染、抖动失焦等问题，这些图像将会严重影响模型的分类效果。而数据预处理的效果会影响模型的学习效果，最终决定模型分类性能的好坏，提高图像质量是决定模型分类效果的重要环节。

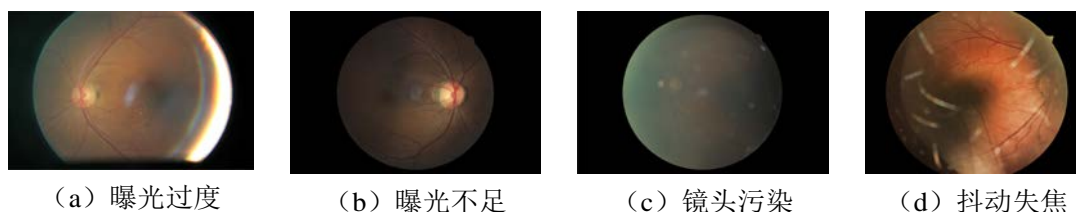


图 3.5 视网膜眼底图像噪声示例图

在眼底图像预处理中，首先手动剔除质量较差的图像，这些图像无法提供完整的视网膜结构信息，不能提供有效的病理特征信息供模型学习。

3.3.2.2 去除冗余边框及大小形状调整

糖尿病视网膜图像中存在黑色冗余边框，在卷积神经网络训练时，这部分区域不仅无法提供有效的特征信息，还大大增加了模型的参数训练。具体的做法是循环遍历图像中的像素点，确定图像中四个方位的第一个非背景区域的有效像素点，然后对图像进行裁剪，由于眼底图像的分辨率较大^[14]，为了平衡实验室服务器的配置，本文将原始图像分辨率进行调整，使用 OpenCV 图像处理库中的 `resize()` 方法统一数据集分辨率为 224×224 像素，缩放后的图像边缘被裁剪，为了恢复完整的圆形视网膜结构，采用图像与运算裁剪图像。处理效果如图 3.6 所示。

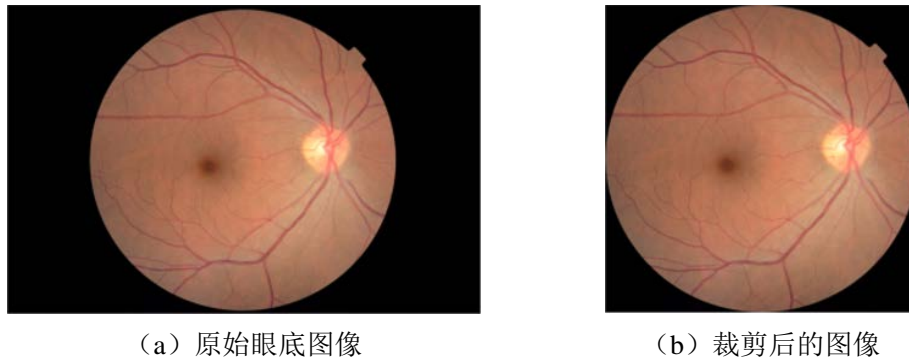


图 3.6 视网膜眼底图像裁剪调整效果对比图

3.3.2.3 受限的对比度自适应直方图均衡

通过观察彩色眼底图像，可以发现在病灶特征中，微动脉瘤的颜色空间与出血病灶非常类似，相近的解剖结构为血管。由于视网膜呈凹面体，这种曲面结构导致了视网膜图像边缘区域的亮度相较于中心区域是比较暗的，造成图像的中心到边缘的对比度相差较大。因此，在进行模型训练之前，对眼底图像对比度增强是非常必要的。

在图像处理中，直方图通常是用来统计各种灰度值的频率，在某种程度上反映了图像亮度的变化规律。直方图均衡是将原始的直方图进行非线性映射，使得其具有均匀的分布特征，因此可以增加灰度值的取值区间，提高整个图像的对比度^[66]。这种做法一定程度上达到图像增强的效果，但是会减少图像的灰度级，导致视盘周围明亮区域和眼底图像较暗区域的一些特征信息被丢失。于是 Pizer S M 等人^[67]提出了自适应直方图均衡（Adaptive Histogram Equalization, AHE）的处理方法来解决上述问题，自适应直方图均衡化是针对局部区域过度明亮或过于暗淡的情况，通过调节图像局部对比度进行优化，具体采用滑动窗口的方式，在每个方向上按一定的步长滑动，在每次滑过的范围内做直方图均衡处理。如果存在重叠的部分，则采用取平均值的方式。自适应直方图均衡技术在提取图像局部细节的同时，也存在图像中灰度分布密集区域的噪声点被过分放大的问题。Zuiderveld K 等人^[68]提出了对比度受限的自适应直方图均衡（CLAHE），限制对比度自适应直方图均衡可以在一定程度上抑制 AHE 中的噪声放大，从而实现对比度的优

化调整。该方法采用预设的剪切阈值进行直方图的剪裁,然后在各个灰度值上平均分配,反复多次,直至各灰度值对应的频率都小于阈值^[69]。具体实施过程是:

- (1) 首先将输入图像其分割成不同的子区,这些子区具有相同的尺寸且互不重叠。
- (2) 在各子区内分别计算直方图分布,采用预定的阈值进行调整和均衡处理。
- (3) 由于前两步处理导致图像在区域间的差别较大,为了解决这一问题,采用双线性插值法对图像进行映射变换,生成最终的直方图。
- (4) 将直方图与原图进行图层滤色混合操作,具体将直方图与原图像中的相应像素值求平均,使图像更加自然和平滑。

本实验使用 CLAHE 算法对糖尿病视网膜眼底图像进行对比度增强预处理,由于采用直方图处理图像时,必须使用单通道灰度图进行处理,如图 3.7 所示,在彩色眼底图像中,可以看到 R、G、B 三种不同的通道呈现出不一样的灰度特性,其中在 G 通道中眼底图像的血管和其他组织表现最为清晰,对比度最好,本文采用 G 单通道图像做对比度受限的自适应直方图均衡化处理,通过多次处理观察,最终设定参数 clipLimit 为 7, tileGridSize 为 6。处理效果及直方图像素分布如图 3.8 所示。

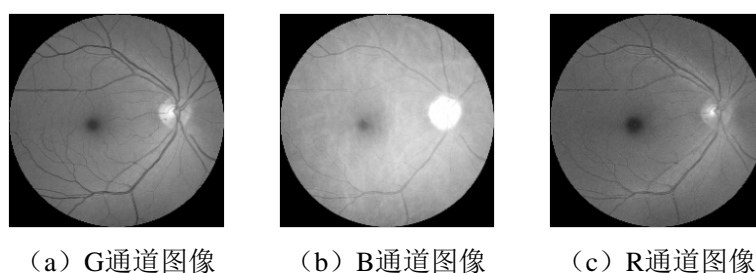


图 3.7 眼底图像 RGB 单通道图像对比图

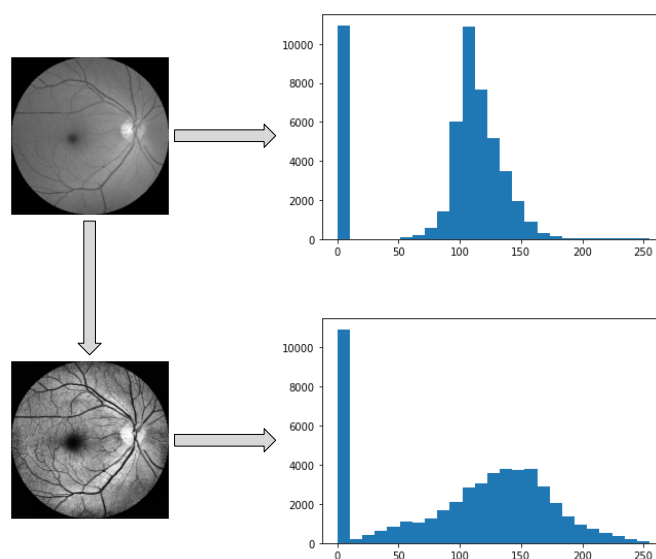


图 3.8 CLAHE 处理前后对比图及其像素值分布图

3.3.2.4 平滑高斯滤波处理

为了减少亮度条件不一致因素的影响,利用高斯滤波对图像进行亮度均衡化和对比度均衡化处理,减少了不同曝光和对比度带来的差异,最大化的获取了眼底图像病变区域的细节特征。在平滑高斯滤波算法中,通过对图像中的每个像素值进行加权平均计算,即利用像素点本身及其周围区域的像素值采用特定的加权平均计算得出。具体公式(3.1)如下。

$$I_i(x, y; \sigma) = \alpha(I(x, y) - G(x, y; \sigma) * I(x, y)) + \beta \quad \text{公式 (3.1)}$$

式中, $G(x, y; \sigma)$ 表示高斯平滑函数, σ 代表标准差,用来判断分析图像的背景像素值,优化像素值的分布,来提高图像的对比度, β 代表强度,像素值分布在 0-255 之间。在处理图像中,当 $\alpha = 4$, $\sigma = r/30$, $\beta = 128$ 处理效果如图 3.3(e)所示。图 3.9 展示了糖尿病视网膜 5 种类别的原图和预处理后的效果图。

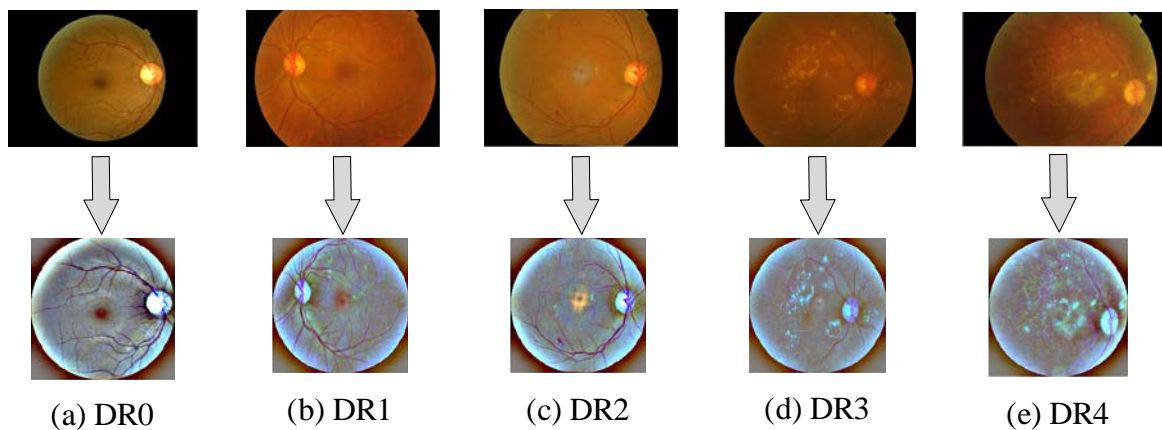


图 3.9 糖尿病视网膜原始图及最终的图像处理效果图

3.3.3 不平衡数据处理

解决数据分布不平衡问题的方法主要分为数据级别和算法级别两类^[70]。针对数据级的处理方法有过采样和欠采样两种方式,过采样针对的是数据集中的小样本数据,对小样本数据集重复采样,以此来扩充数据样本;而欠采样是指对数据集中类别较多的样本范围内进行局部采样^[71]。算法级别的处理方式主要有基于代价敏感^[72]和集成学习两种方式,前者通过加大模型对样本错误分类的惩罚代价来抑制数据不平衡的影响,后者通过使用多个分类器,接着应用策略对部分分类结果进行处理,得到新的样本进行分类。由于深度学习模型训练时对于数据的要求较高,不仅需要大量的数据作为支撑,还需要保证数据的质量和分布的合理性。kaggle 平台提供的 DR 数据集一定程度上满足了深度学习所需要的数据规模,但是该数据集存在严重分布不平衡问题,这对于训练多分类模型是不利的。

本文利用数据增强的方式来扩充小样本数据量，提高模型的鲁棒性。在对图像样本进行扩充时，通常利用 Keras 图片生成器（Image Data Generator）对图像进行以下方式的增强处理。图像生成器中主要采用随机旋转（Rotation_range）、水平翻转（Horizontal_flip）、垂直翻转（Vertical_flip）、比例缩放（Zoom_range）、水平偏移（Width_shift_range）、垂直偏移（Height_shift_range）、剪切变换（Shear_range）、填充处理（Fill_mode）等方法来增加样本数量，最终使得各类别之间数据达到均衡，眼底图像扩增示意图如图 3.10 所示，将扩充后的数据集进行训练集、验证集和测试集进行划分，具体如下表 3.2 所示。

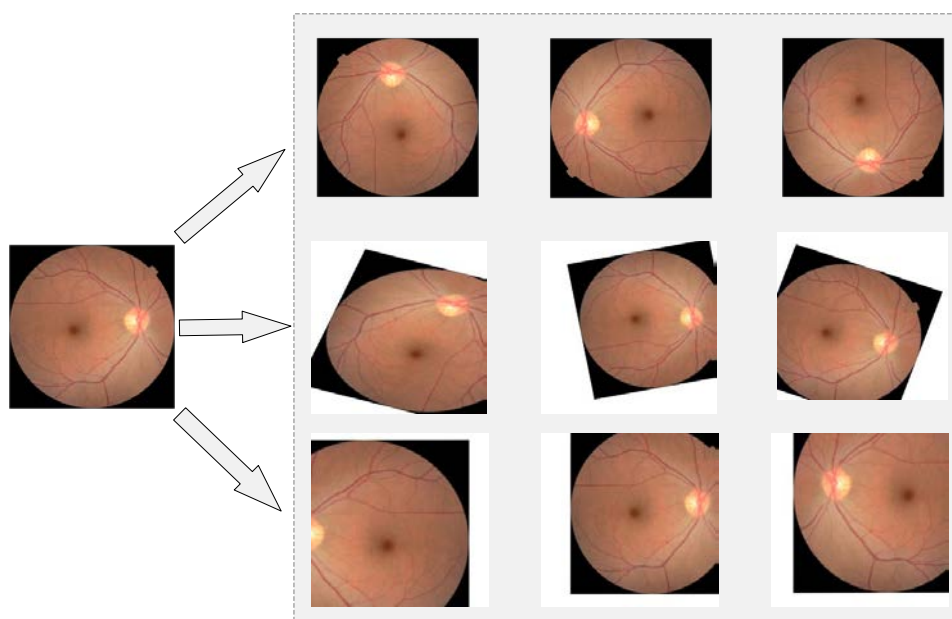


图 3.10 眼底图像扩充示意图

表 3.2 平衡后的数据集划分表

类别	训练集/张	验证集/张	测试集/张	总计
DR0	3840	678	110	4628
DR1	3780	667	110	4557
DR2	3841	678	110	4629
DR3	3832	676	110	4618
DR4	3743	660	110	4513
总计	19036	3359	550	22945

由上表可知，Kaggle 数据集经过预处理和部分类别扩充之后，共得到 22945 张数据集。其中，训练集 19036 张，大约占总数据集的 83%；验证集 3359 张，大约占总数据集的 15%；测试集 550 张，大约占总数据集的 2%。在每一个类别中，训练集、验证集和测试集的比例也大致分别为 83%、15% 和 2%。

3.4 糖尿病视网膜眼底图像分类模型架构设计

模型设计基于 MobileNetV2 和 EfficientNetB0 的模型架构，采用模块集成方式并结合注意力机制构成糖尿病视网膜多分类模型，该模型集成分为模型“头部”（Head）和“主干”（Stem）两部分。在模型训练中，模型头部利用迁移学习的方式提取图像低维特征信息，主干结合注意力模块提取图像的高维特征信息。

3.4.1 MobileNetV2 模型结构

3.4.1.1 深度可分离卷积

MobileNetV2 网络由 Google 研究团队在 2018 年提出，与 MobileNetV1 模型相比，V2 版本的准确率更高且模型参数更小^[74]。整个模型由深度可分离卷积堆叠组成，与传统卷积层相比，深度可分离卷积能够在保证准确率的前提下大幅度地减少了模型参数与运算量。传统卷积过程如图 3.11 所示，假设输入特征矩阵 $F_{input} \in R^{D_F \times D_F \times C_F}$ ，其中特征图的高和宽均为 D_F ，特征图的通道数为 C_F ；接着对输入特征图进行卷积操作，卷积核的个数为 N_K ，大小为 $D_K \times D_K$ ，通道数为 C_K ，在传统卷积中，由于卷积核的通道数必须与输入特征矩阵的通道数相等，故 $C_F = C_K$ ；最后得到输出特征矩阵 $F_{output} \in R^{D_F \times D_F \times N_K}$ ，其中输出特征矩阵的高和宽与输入特征图相等均为 D_F ，特征通道个数为 N_K 。

深度分离卷积由深度卷积（DW）和逐点卷积（PW）构成，具体如图 3.12 所示，假设输入特征 $F_{input} \in R^{D_F \times D_F \times C_F}$ ，该特征通过一步深度卷积处理，得到过渡特征图 $F_{temp} \in R^{D_F \times D_F \times C_F}$ ，在深度卷积中卷积核的通道数为 1，卷积核的尺度大小为 $D_K \times D_K$ ，且卷积核的个数与输入特征矩阵的通道数相同^[75]，即 $N_K = C_F$ 。在逐点卷积中，卷积核的大小为 1×1 ，通道数为 N_K 。过渡特征经过逐点卷积后的输出为 $F_{output} \in R^{D_F \times D_F \times N_K}$ 。

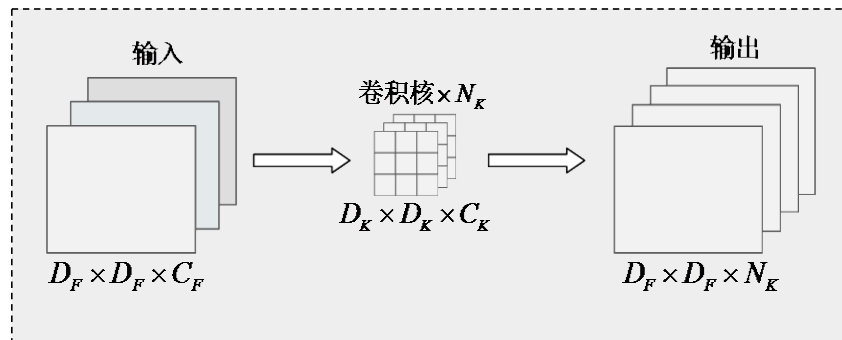


图 3.11 传统卷积过程示意图

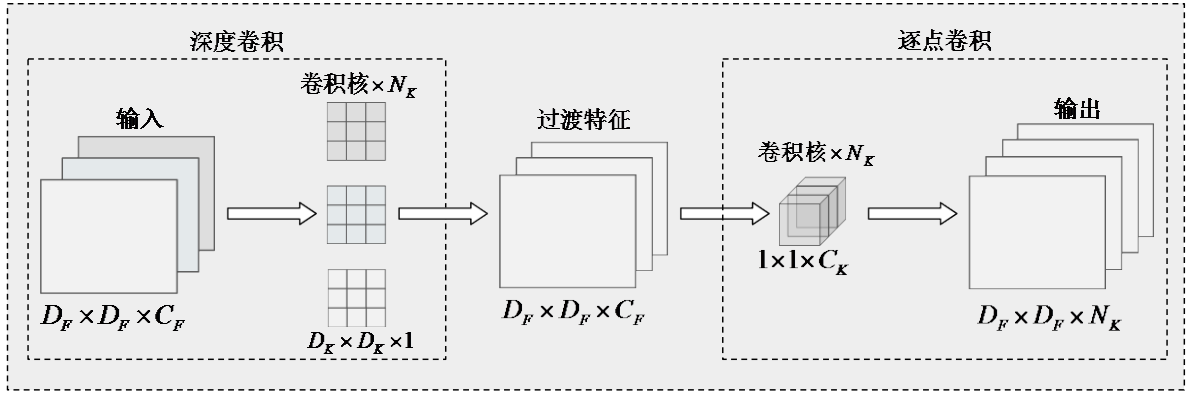


图 3.12 深度可分离卷积过程示意图

按照上面的参数约定，传统卷积的计算量 C_1 ，深度可分离卷积的计算量 C_2 ，其计算公式分别为式(3.2)和式(3.3)。另外，深度可分卷积与传统卷积的计算量之比如式(3.4)所示。

$$C_1 = D_K \times D_K \times C_F \times D_F \times D_F \times N_K \quad \text{公式 (3.2)}$$

$$C_2 = D_K \times D_K \times C_F \times D_F \times D_F + C_F \times N_K \times D_F \times D_F \quad \text{公式 (3.3)}$$

$$\frac{C_2}{C_1} = \frac{1}{N_K} + \frac{1}{D_K \times D_K} \quad \text{公式 (3.4)}$$

在模型训练中，公式(3.4)中的 N_K 是一个较大的数，且深度可分离卷积中的卷积核的大小 D_K 通常设置为 3×3 大小，故理论上讲，相较于传统卷积运算，深度可分离卷积的计算量降至原来的约 $1/9$ 。

3.4.1.2 线性瓶颈层的倒残差结构

该模型为了防止非线性层损失，引入了线性瓶颈层，同时借鉴 ResNet^[76]和 DenseNet 等网络采用捷径分支结构同时结合深度卷积，提出倒残差结构（Inverted Residual）。与经典的残差模块先降维再升维的方式相反，倒残差结构采用先升维再降维的方式进行特征提取，对特征矩阵进行升维更适合深度可分离卷积进行特征提取。最后为了减少非线性激活函数对压缩通道特征信息的丢失，在最后的逐点卷积操作后，使用线性激活函数 Linear。瓶颈层 Bottleneck 具体结构如图 3.13 (a) 所示，输入特征经过 1×1 的 conv 层同时采用激活函数 Relu6 将维度从 k 维扩增到 tk 维，之后通过一个 3×3 的深度卷积同时采用激活函数 Relu6 对特征进行下采样（stride>1 时），最后通过一个 1×1 的 conv 层进行降维，此时不采用激活函数，最终通道数由 tk 维降低到 k 维。对于瓶颈层而言，只有当步长为 1 时才会有捷径分支的出现，如下图 3.13 (b) 所示。

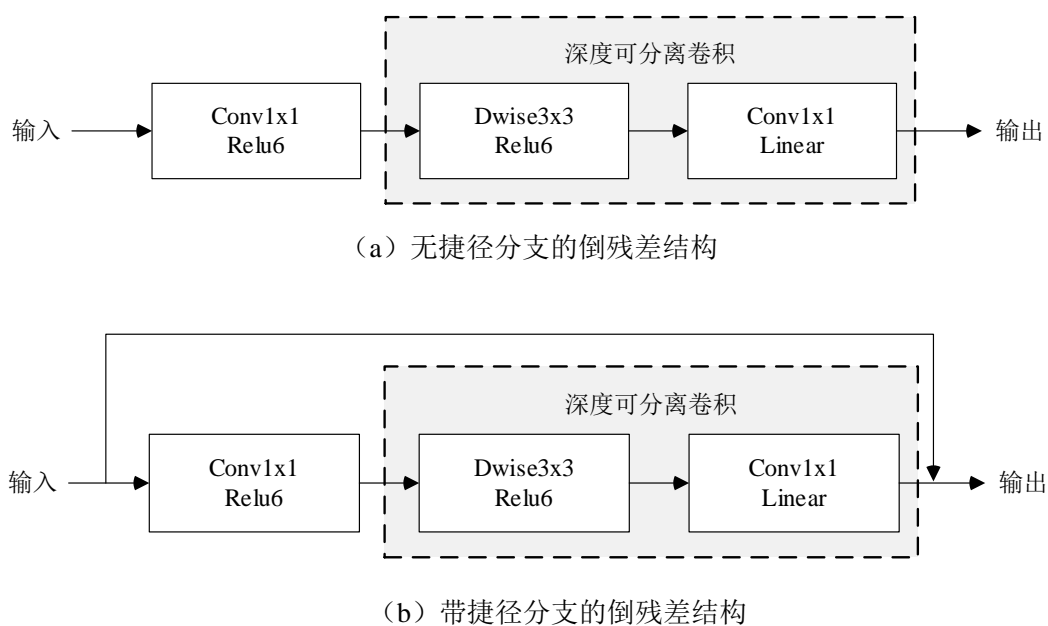


图 3.13 Bottleneck 卷积结构示意图

3.4.1.3 模型参数配置

MobileNetV2 的模型作为集成模型之一,模型由一系列线性瓶颈的倒残差结构组成,配置如下表 3.3 所示。

表 3.3 MobileNetV2 核心结构配置

输入分辨率	卷积模块	扩展倍数	输出通道数	层数	步长
224x224	Conv2d	-	32	1	2
112x112	Bottleneck	1	16	1	1
112x112	Bottleneck	6	24	2	2
56x56	Bottleneck	6	32	3	2
28x28	Bottleneck	6	64	4	2
14x14	Bottleneck	6	96	3	1
14x14	Bottleneck	6	160	3	2
7x7	Bottleneck	6	320	1	1
7x7	Conv2d	-	1280	1	1

3.4.2 EfficientNetB0 模型结构

3.4.2.1 模型尺度缩放

传统的模型性能改进策略是对模型进行加深扩展,ResNet 结构有效解决了模型加深过程中产生的梯度弥散问题,例如 ResNet-18 被不断的加深至 ResNet-200。但是仅仅在模型深度方向加深已经不能更加有效的提升模型性能,更深的模型虽然学习到更高维的

特征信息，拥有更大的感受野，但是在具体的分类、分割任务中，容易忽略低维信息，难以有效学习一些小样本、小目标特征信息，严重影响模型性能的进一步提升。在模型中，特征图通道数越多可以理解为模型的宽度越宽，增加特征图的宽度数虽可以获得更多低层次的语义特征信息，丰富特征空间表达能力，但宽度过大且深度较浅的模型很难学习到深层次的语义特征，不利于特征信息的组合。高分辨率的图像输入可以使模型学习到更多细粒度的信息，能够扩大网络的感受野进而提升网络性能，但是会增大模型的计算量。EfficientNet 网络模型是 Google 团队通过机器搜索得到的深度模型^[77]，该模型将网络宽度、深度及提高图像的分辨率通过缩放系数对分类模型进行 3 个维度的缩放^[78]，如图 3.14 所示。模型通过自适应地优化网络结构，不仅提升了模型对更丰富、更复杂特征的学习能力，并由此提高模型精度，还可以充分利用设备性能，加速模型的训练速度。

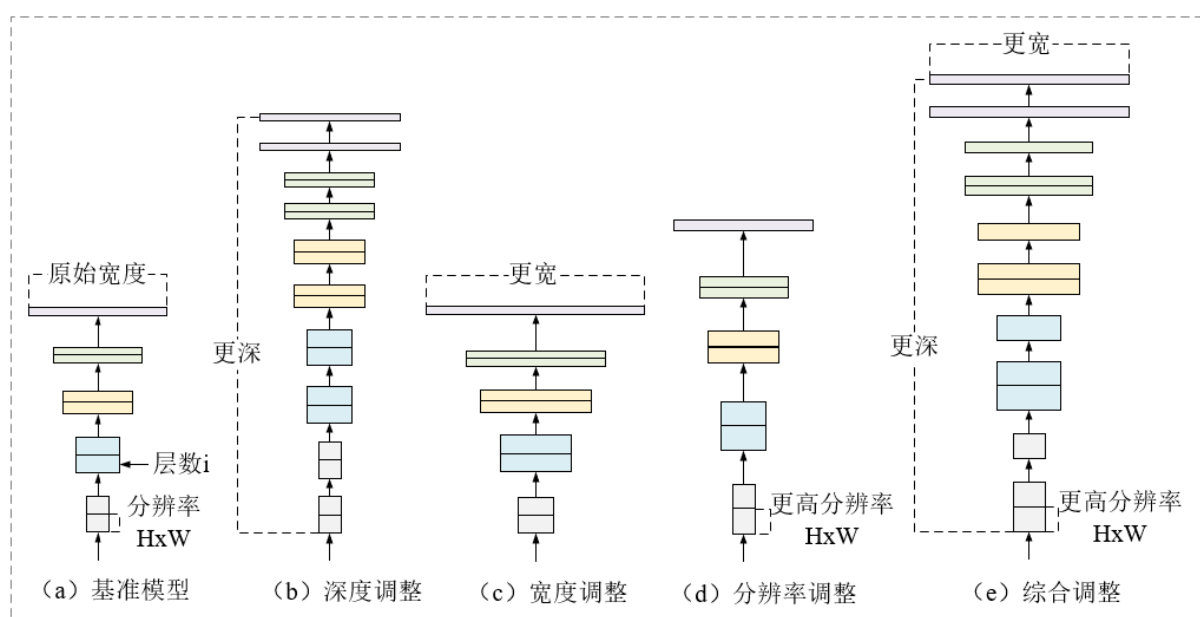


图 3.14 EfficientNet 模型调整示意图

图 3.14 中，(a) 为基准模型，(b) 为对模型进行单一的加深扩展，(c) 为增加特征图的通道数进行模型宽度调整，(d) 为增加输入分辨率进行调整，(e) 为同时对深度、宽度和分辨率进行综合调整。

为了平衡三个维度的调整幅度，模型使用了组合缩放系数。模型参数调整过程中，各维度之间是相互联系的，一般来讲，输入图像的分辨率越高，那么需要的模型更深和宽度更宽。假设模型的三个维度的缩放基数为 α 、 β 和 γ ，分别对应深度 d 、宽度 w 和分辨率 r ，具体扩展公式如式 (3.5) 所示：

$$d = \alpha^{\phi}; w = \beta^{\phi}; r = \gamma^{\phi} \quad \text{公式 (3.5)}$$

上式中， ϕ 为一个单一的组合缩放系数，可以控制资源的使用量，且 α 、 β 和 γ 的必须满足如式 (3.6)、(3.7) 中的条件。

$$\alpha \cdot \beta^2 \cdot \gamma^2 \approx 2 \quad \text{公式 (3.6)}$$

$$\alpha \geq 1, \beta \geq 1, \gamma \geq 1 \quad \text{公式 (3.7)}$$

本实验采用 EfficientNetB0 作为集成模型之一，在该模型中， ϕ 取值为 1，通过网络搜索技术得到 $\alpha=1.2$ 、 $\beta=1.1$ 、 $\gamma=1.15$ 。

3.4.2.2 MBConv 卷积设计

EfficientNet 模型是由一系列 MBConv 卷积结构堆叠而成，其中 MBConv 结构如图 3.15 所示，由一个 1×1 的卷积、一个深度卷积、一个 SE 注意力模块、一个用于降维的 1×1 卷积和一个随机失活层组成，输入特征在经过第一次 1×1 卷积和深度卷积之后都进行了 BN 和 Swish 激活操作，在第二次 1×1 卷积之后只进行 BN 操作，需要注意的是，第一次卷积是为了扩展通道维度，扩展比例设置为 n ，当 $n=1$ 时，不进行通道扩展，直接进行 $k \times k$ 的深度卷积操作。接着将输出特征输入到 SE 注意力模块结构中，再通过卷积进行降维，调整输出通道维度与输入通道维度相等。最后进行随机失活处理，同时在输入和输出之间设置捷径分支，该结构有利于模型融合更多的特征信息，而只有当输入 MBconv 结构的特征矩阵与输出的特征矩阵形状相同时才有捷径分支的存在。

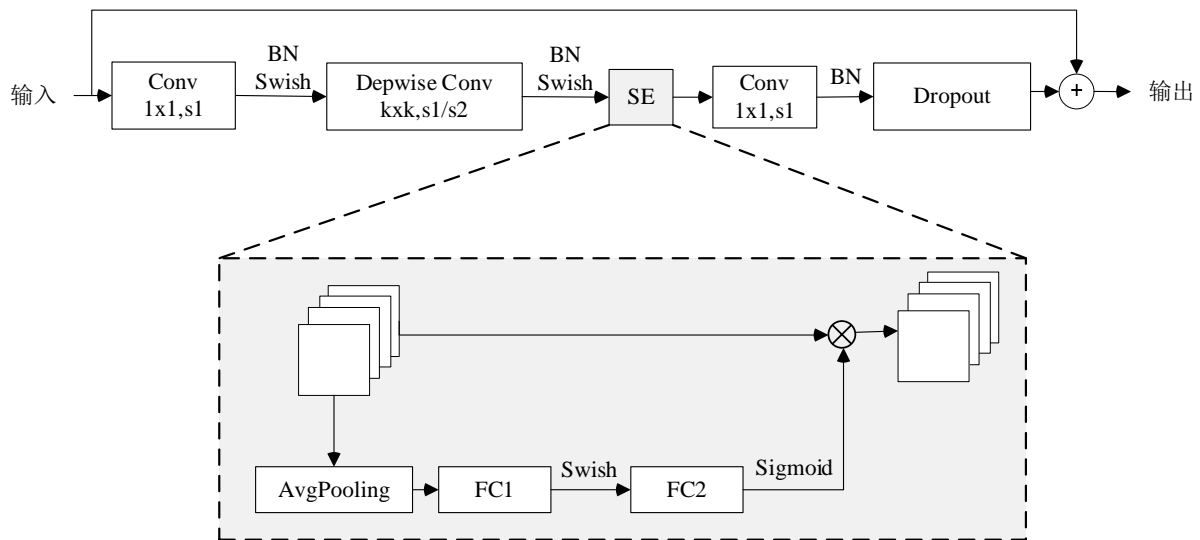


图 3.15 MBConv 结构示意图

在图像分类、图像分割和目标检测任务中，SE 注意力模块效果显著，该模块由一个全局平均池化、两个全连接层构成。其中，FC1 的节点个数是输入该 MBConv 特征矩阵通道数的 $1/4$ ，且使用 Swish 函数做激活处理，FC2 节点个数等于 DW 卷积输出的特征矩阵通道数，同时使用 Sigmoid 激活函数进行激活。假设一个输入特征 $H \times W \times C$ ，首先通过全局平均池化和全连接层将其拉伸为一个 $1 \times 1 \times C$ 的特征向量，然后与原始图像相乘，

对通道特征重新加权计算，从而学习更多目标特征信息。

3.4.2.3 模型参数配置

实验中选取 EfficientNetB0 的部分结构作为主干模型集成部分，该模型完成结构由一个 3×3 的卷积层、16个 MBConv (mobile reversed bottleneck convolutional)、一个 1×1 的卷积层、一个池化层和一个全连接层组成，具体配置如表 3.4 所示。

表 3.4 EfficientNetB0 核心结构配置

阶段	操作	输入分辨率	输出通道数	层数	步长
1	Conv3x3	224x224	32	1	2
2	MBConv1, k3x3	112x112	16	1	1
3	MBConv6, k3x3	112x112	24	2	2
4	MBConv6, k5x5	56x56	40	2	2
5	MBConv6, k3x3	28x28	80	3	2
6	MBConv6, k5x5	14x14	112	3	1
7	MBConv6, k5x5	14x14	192	4	2
8	MBConv6, k3x3	7x7	320	1	1
9	Conv1x1&Pooling&FC	7x7	1280	1	-

3.4.3 GAM 注意力模块设计

在糖尿病视网膜病变分类任务中，微小的病灶点如渗出物和微动脉瘤等类别之间差异较小，造成视网膜不同类别之间相似度较高，模型难以准确识别相似度较高的病变图像，为提高模型分类精度，需要模型具有更加细粒度分类能力，引入更加细粒度分类实现视网膜病变的分级。对于细粒度的分类来说，如何使模型学习到有效的目标特征成为决定细粒度分类模型性能的关键。

在注意力机制中，通道注意力机制通过学习通道注意力权重，判别每个特征通道重要程度的同时抑制信息量较小的通道，而空间注意力机制则是利用学习空间注意力的权值来描述各个空间的重要程度，并与通道注意力机制形成互补机制。基于此，本文设计全局注意力机制模块 (global attention module, GAM) 来学习目标病灶特征，GAM 模块由通道注意力和空间注意力模块组成，采用平均池化单分支结构。全局平均池化 (Global Average Pool, GAP) 可以实现图像通道特征的压缩，对输入特征图的空间变化具有较强的鲁棒性。但仅考虑建模评估特征图中每个通道之间的重要性过于简单，无法准确学习特征之间的长依赖关系。由于空间位置信息决定生成空间选择性注意力图，在计算机视觉任务中，这些空间位置信息往往更有利于模型捕获学习目标特征。故重新设计注意力机制，使用平均池化对图像通道和空间特征信息进行转换处理，具体如图 3.16 所示。

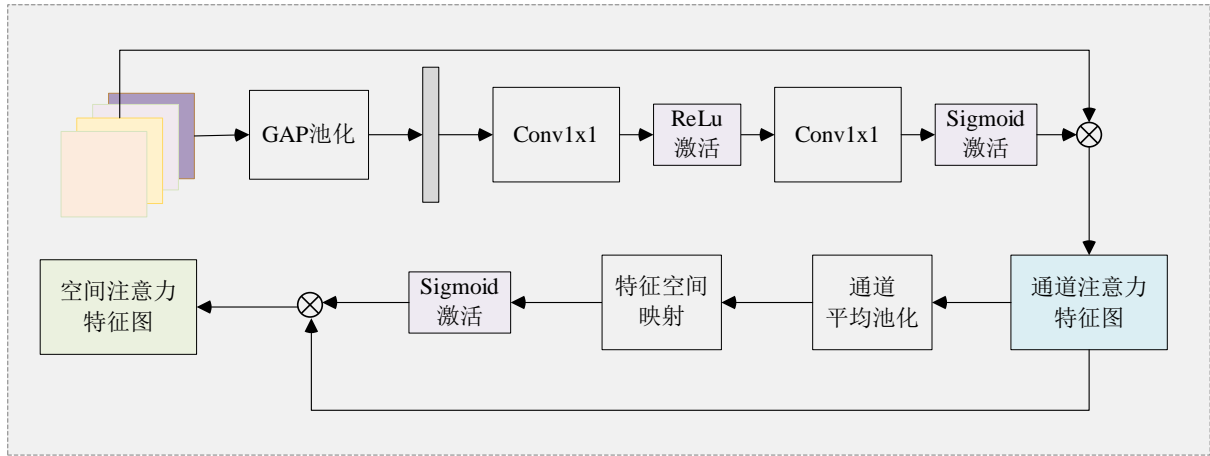


图 3.16 注意力机制模块设计

注意力机制网络将集成模型输出的融合特征作为输入，首先利用公式 (3.8) 计算生成通道注意力特征图 $F_{CA-OUT} \in R^{H \times W \times C}$ 。

$$F_{CA-OUT} = (\sigma(\text{Conv2}(\text{GAP}(F_{GAM-IN})))) \otimes F_{GAB-IN} \quad \text{公式 (3.8)}$$

其中， $\sigma(\cdot)$ 表示 *Sigmoid* 激活函数，将变量值映射到 [0,1] 区间，*Conv2* 表示两个卷积核大小为 1×1 的卷积层，*GAP*(\cdot) 为全局平均池化操作，即在通道维度上进行全局平均池化， \otimes 表示特征同位元素对应相乘。

然后用公式 (3.9) 计算 GAM 的输出特征 $F_{SA-Feature} \in R^{H \times W \times 1}$ ，即空间注意力特征图。

$$F_{SA-Feature} = F_{CA-OUT} \otimes \sigma((C_GAP(F_{CA-OUT}))) \quad \text{公式 (3.9)}$$

上式中，*C_GAP*(\cdot) 为空间维度平均池化操作， $\sigma(\cdot)$ 为 *sigmoid* 激活操作。空间注意力结构简化为单链路，针对模型的原始特征输入，只采用通道平均池化操作，更好地维护了图像的空间信息，有利于对空间注意权值的学习，同时指出每个空间位置的重要性。

3.4.4 模型集成设计

深度学习中，增加网络的深度能够获取更加丰富、复杂的特征，但是网络深度过深不仅会大量增加模型的训练参数，还会出现梯度消失、训练困难等问题。通过增大网络的宽度，可以得到更细粒度的特征信息，但宽度过大且深度较浅的网络将难以学习到更深层的语义信息。为了权衡模型训练的代价和精度，采用模型集成的策略，采用权重代价较小的轻量级网络 *MobileNetV2* 和兼顾网络深度和宽度的 *EfficientNetB0* 模型进行融合作为特征提取网络，具体成分为两部分，即模型头部集成和主干部分集成，两者集成方式略有不同，在主干模型中嵌入了注意力机制。集成模型的总体框架如图 3.17 所示。

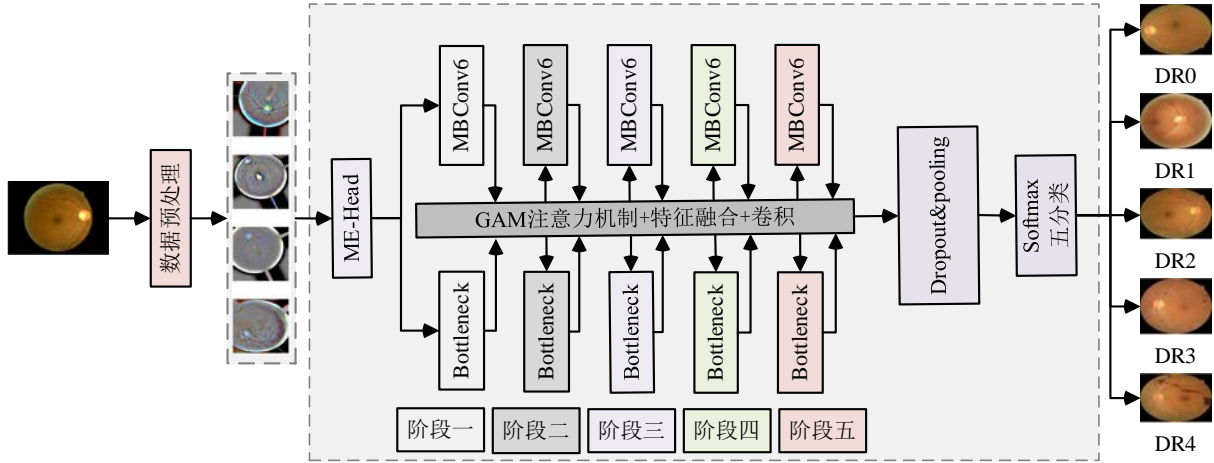


图 3.17 ME-Attention 模型框架

3.4.4.1 模型头部集成

具体的头部集成模型 ME-Head 如下图 3.18 所示，其中头部训练时采用迁移学习的方式加载与训练权重进行特征提取，提高模型的收敛时间。头部集成模型中，输入尺度为 $224 \times 224 \times 3$ ，首先将图像进行一步卷积操作，转化为 MBConv 和 Bottleneck 结构所需要的输入维度 $112 \times 112 \times 3$ ，然后分别选择两种基本模型核心结构的前两层，即对应三组 MBConv 和 Bottleneck 结构，然后进行 Concat 特征融合操作，将两部分的输出特征进行通道叠加，丰富通道特征信息，融合特征尺度为 $56 \times 56 \times 48$ 。最后再进行一次卷积运算，调整输出特征的通道维数，最终得到输入到集成主干模型特征图。

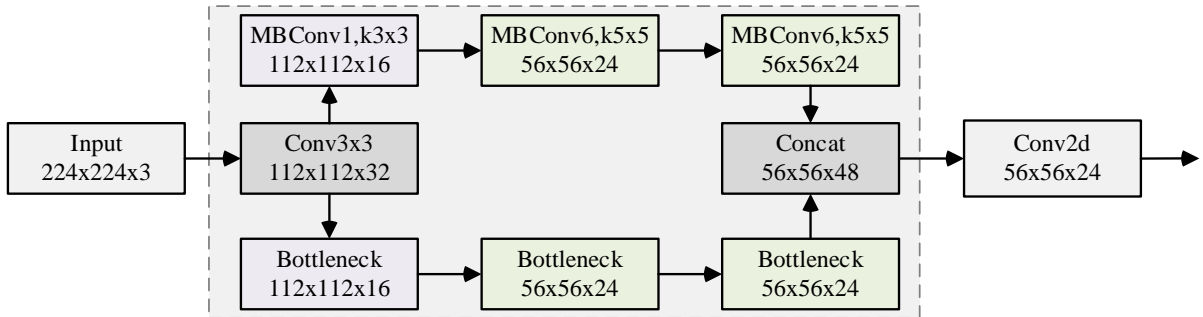


图 3.18 ME-Head 结构

3.4.4.2 模型主干集成

在深度学习中，通常采用加深网络深度来获取医学图像中的深层次语义特征。为了解决网络模型加深带来的浅层网络学习特征学习不充分问题，本文采用集成模型核心结构，并结合注意力机制实现模型对特征的高效学习。随着模型的不断深化，在模型的训练中容易产生梯度消失和爆发问题，解决这一问题的传统方式是进行梯度裁剪和权重正则化操作，但是这种方式容易使网络发生退化现象。而残差结构可以解决这一问题，缓

解网络退化，为此，在主干部分核心结构集成过程中，利用残差结构进行连接。本文根据 MobileNetV2 和 EfficientNetB0 模型的配置结构，本文将模型分为五个阶段进行集成，每个阶段的集成方式如图 3.19 所示。

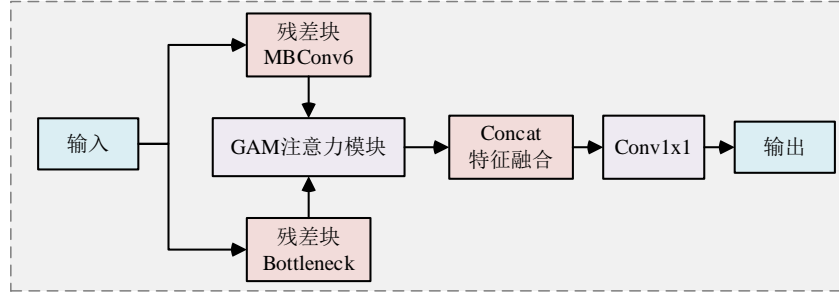


图 3.19 核心结构集成方法示意图

图 3.20 展示了模型集中 EfficientNetB0 核心结构中各阶段的残差结构，由于阶段五在模型设计时只有一个 MBConv 层，故不设计残差连接；图 3.21 展示了基于 MobileNetV2 结构的各阶段残差结构，同样阶段五中只有一个 Bottleneck 层结构，在此不设计残差连接。本文设计分阶段特征融合方案，由图 3.18 可知，头部集成模型的输出特征 $F_{head_out} \in R^{56 \times 56 \times 24}$ ，分别经过 MBConv 残差块和 Bottleneck 残差块进行处理，得到的特征图均通过 GAM 注意力模块处理生成两个注意力特征向量，然后利用 Concat 操作进行特征通道的融合，丰富模型的通道特征，最后经过一个 1×1 的卷积重新调整通道维度，来优化空间特征的映射关系，第一阶段集成结构生成特征图 $F_{stage_1} \in R^{28 \times 28 \times 72}$ 。以此类推，其余各阶段的输出特征分别为： $F_{stage_2} \in R^{14 \times 14 \times 144}$ 、 $F_{stage_3} \in R^{14 \times 14 \times 208}$ 、 $F_{stage_4} \in R^{7 \times 7 \times 352}$ 、 $F_{stage_5} \in R^{7 \times 7 \times 640}$ ，最后进行 dropout 和 pooling 操作后，使用 softmax 函数进行分类。

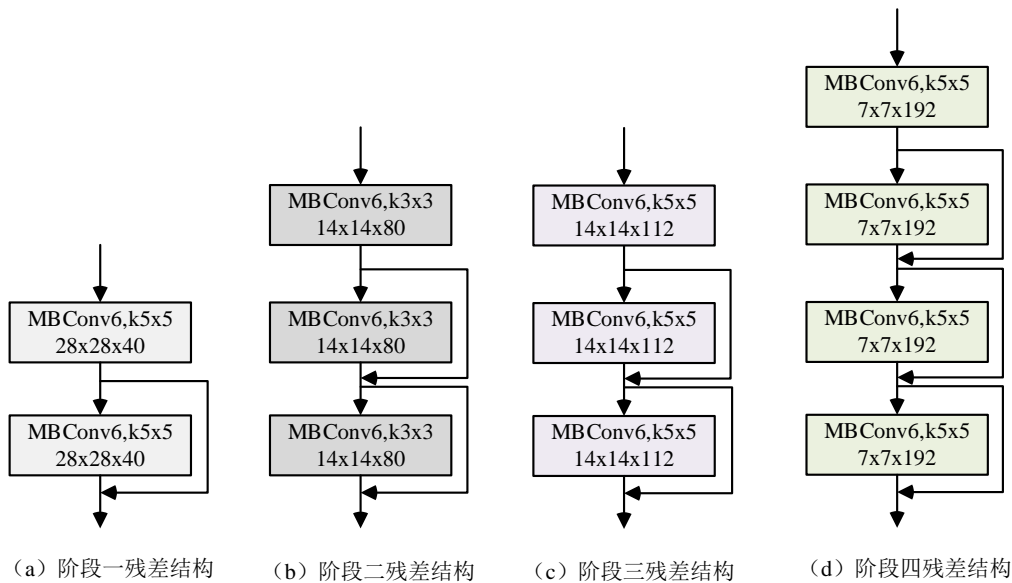


图 3.20 集成模型中基于 EfficientNetB0 的残差结构

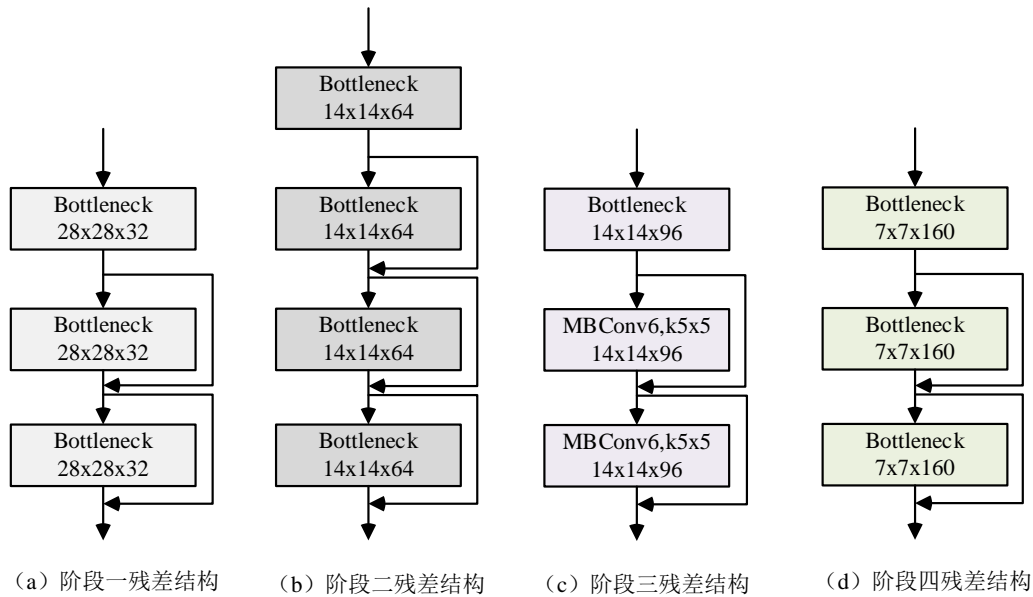


图 3.21 集成模型中基于 MobileNetV2 的残差结构

3.5 实验结果与分析

3.5.1 实验环境

本文实验深度学习框架为 Tensorflow2.0 版本, 采用 Python3 编程语言进行模型框架的搭建, 实验在 GPU 型号为 NVIDIA RTX 2080Ti、16G 显存的 Ubuntu16.04 操作系统, CUDA10.0, cudnn7.6.0 的服务器上训练调试。首先将预处理后的各类数据进行比例划分, 模型输入尺寸调整为 224x224 像素, 训练过程中使用 Adam 优化器进行参数调优, 初始学习率设为 0.00005, Loss 计算过程中采用交叉熵损失函数(Cross Entropy Loss, CE), 损失函数的权重衰减因子设为 0.1, batch_size 设置为 32, 一共训练 60 个 epochs。

3.5.2 实验评估指标

在深度学习中, 评价一个模型分类性能的好坏时, 通常采用分类准确率 (Accuracy)、精确率 (Precision)、特异性 (Specificity)、灵敏度 (Sensitivity)、和二次加权 Kappa 系数对模型分类效果进行评估。在定义以上评估标准前, 首先介绍四种量化统计指标, 统计定义真阳性 (True Positive, TP)、假阳性 (False Positive, FP)、真阴性 (True Negative, TN) 和假阴性 (False Negative, FN) 四个指标^[79]。在本课题研究背景下以糖尿病视网膜病变状况为例, 其中 Positive 样本为病变图像, Negative 为正常眼底图像。TP 表示样本为病变视网膜图像, 模型的预测结果也是病变图像; FP 表示样本为正常眼底图片, 被模型预测为病变图片; TN 表示样本为正常眼底图像, 模型预测也为正常图像; FN 表示样本本身为病变视网膜图像, 但是被模型预测为正常眼底图像。下表 3.5 以表格的形式

对以上四种量化指标进行总结，评价指标定义如式（3.10）~（3.13）所示。

表 3.5 四种量化统计指标

类别\类别	真实病变图像	真实非病变图像
预测病变图像	真阳性（TP）	假阳性（FP）
预测非病变图像	假阴性（FN）	真阴性（TN）

$$Precision = \frac{TP}{TP + FP} \quad \text{公式（3.10）}$$

$$Sensitivity = \frac{TP}{TP + FN} \quad \text{公式（3.11）}$$

$$Specificity = \frac{TN}{TN + FP} \quad \text{公式（3.12）}$$

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad \text{公式（3.13）}$$

精确率又称为阳性预测值（Positive Predictive Value, PPV），表示真实患病样本占预测结果为患病样本中的比例；特异性也叫真阴性率，它表示在所有真实为正常样本的数据中，模型预测结果也为正常样本的比例，其值越高，说明模型诊断非病变的能力较强，其值越低，表明有更多的假阳性出现，造成医疗资源的浪费，对患者造成无端的恐慌；灵敏度既可称为真阳性率，也可称为召回率，它表示在所有真实为患病的样本中，模型同样也预测为患病样本的比例，其值越高，表示模型性能越好，其值越低，则说明假阴性样本较多，也就是出现误诊结果，延误患者就诊。准确率是指模型正确分类样本数占总样本数的比例，即真阳性和真阴性数据在总数据中的比例。值得注意的是，特异性和灵敏度是针对某一具体类别时的评价计算值，而准确率是针对所有类别来进行统计。本文采用二次加权 Kappa 系数（QWK）作为评估指标，Kappa 系数是统计学中度量一致性的指标，对于分类问题，一致性就是模型预测结果和实际分类结果是否一致^[80]。具体评价标准如下表 3.6。

表 3.6 Kappa 评价指标

Kappa 值	一致性
0.00 - 0.20	极低，不可接受
0.21 - 0.40	一般，可接受
0.41 - 0.60	中度，中等
0.61 - 0.80	高度，较大
0.81 - 1.00	几乎完全一致，几乎一致

Kappa 系数是根据混淆矩阵中的统计量的值计算得出的, 取值范围为[-1,1], 它的值一般大于 0。QWK 被设计用于衡量两个评分者在具有序列等级的标签上是否具有一致性, 并且已经被应用在糖尿病视网膜病变等级分类中, QWK 的数学表达如下:

$$k_w = 1 - \frac{\sum_{i,j} w_{i,j} O_{i,j}}{\sum_{i,j} w_{i,j} E_{i,j}} \quad \text{公式 (3.14)}$$

$$w_{i,j} = \frac{(i-j)^2}{(N-1)^2} \quad \text{公式 (3.15)}$$

$$E_{i,j} = \sum_j O_{i,j} \times (\sum_j O_{i,j})^T \quad \text{公式 (3.16)}$$

上式中 $N=5$, 表示样本类别个数。 $O_{i,j}$ 表示一个 $N \times N$ 的矩阵, 代表真正的 i 类被错分为 j 类的样本数; $w_{i,j}$ 代表权重系数, 通过加权来惩罚错分项进而提高模型的分类效果。

另外, 本文引入混淆矩阵 (Confusion Matrix) 来判断模型效果, 混淆矩阵是一种可视化方法, 主要用于比较分类结果和实际预测值, 并可以把分类结果的精度显示在一个混淆矩阵中, 它是评判模型结果的一种指标, 常用于评判分类器模型的优劣。

3.5.3 实验设计与分析

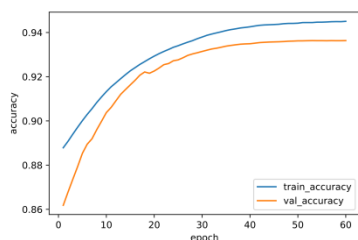
(1) 不同模型结构的性能对比

为了验证集成模型中特征提取网络的有效性。实验利用 3.3 节中预处理后的数据集对模型训练, 实验对比中, 特征提取网络分别取 MobileNetV2、EfficientNetB0 以及两者之间的集成模型, 命名为 ME-Net。ME-Net 与 ME-ANet 集成方式类似, 前者只是按照如图 3.19 中删除 GAM 注意力模块的方式集成, 即与本文的 ME-ANet 模型相比只是缺少了 GAM 注意力模块结构的嵌入。

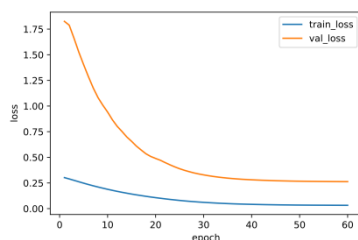
模型的实验平台配置和模型训练超参数设置均与本章 3.5.1 描述一致, 且实验均采用相同的迁移学习训练策略。训练结果如下表 3.7 所示, 集成模型特异性为 0.930, 比 EfficientNetB0 模型低 0.2 个百分点。除此之外, 集成模型的准确率、灵敏度和 Kappa 值均达到最高, 性能明显优于单个模型。表明将模型融合后作为主干网络优于单个基础模型, 也即证明本文提出的模型集成策略的有效性, 适合糖尿病视网膜病变图像的分类任务。图 3.22 展示了 ME-Net 模型的训练曲线图, 由图 3.22 (a) 分析可知, 模型在第 40 个 epoch 左右时就已经达到了收敛状态, 训练时准确率稳定在 0.94 左右, 在验证阶段, 模型准确率达到 0.93 左右; 无论在训练阶段还是在验证阶段, Loss 曲线下降平缓, 没有发生波动, 最后得到的 Kappa 值达到 0.84。最后利用 550 张糖尿病视网膜眼底数据集对模型进行测试, 并将测试结果绘制为混淆矩阵, 如图 3.23 所示。

表 3.7 不同特征提取网络性能对比

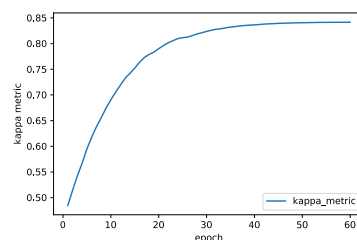
模型	准确率	特异性	灵敏度	Kappa 值
MobileNetV2	0.896	0.914	0.842	0.826
EfficientNetB0	0.857	0.932	0.854	0.812
ME-Net	0.928	0.930	0.868	0.843



(a) Accuracy 曲线图



(b) Loss 曲线图



(c) Kappa 曲线图

图 3.22 ME-Net 模型训练曲线图

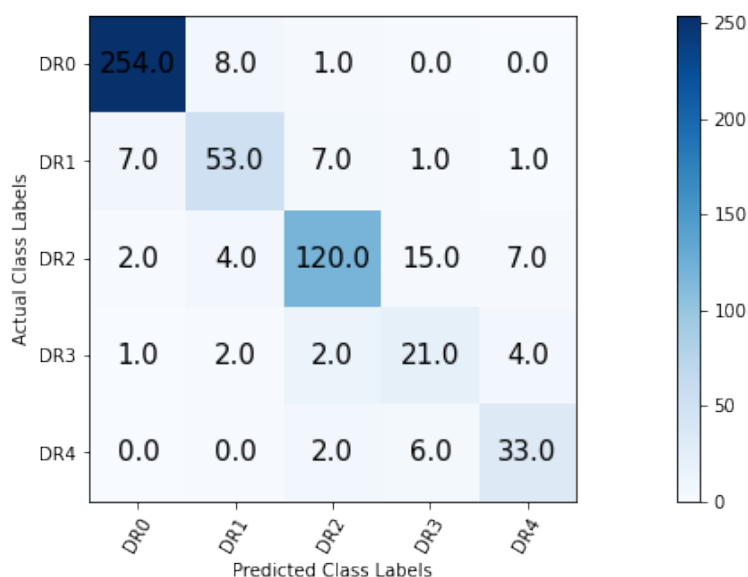


图 3.23 ME-Net 模型预测的混淆矩阵

从混淆矩阵中可以看到被正确预测的 DR 图像的类别和数量，在 550 张测试集中，有 481 张图片被模型正确预测，模型的整体测试准确率达到 87.5%，进一步分析可知，由于视网膜病变眼底图像相邻类别之间差异性较小，给模型精确分类带来了困难，造成模型分类错误的类别主要集中在病变等级相邻的图像之间。表 3.8 展示了模型对视网膜病变图像中每个类别的精确率、召回率和特异度。分析可知，除了模型在 DR3 类别中的精确率不足 0.6，在其余各类中准确率均超过 0.70，从准确率的角度分析，模型对 DR0 和 DR2 的分类较为精准；除此之外，模型的特异度中，在五个类别中均超过 0.95，达到较理想的效果。

表 3.8 ME-Net 模型性能对比

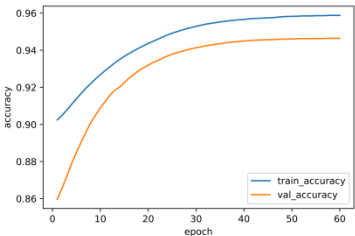
类别	精确率	灵敏度	特异性
DR0	0.962	0.966	0.965
DR1	0.791	0.768	0.971
DR2	0.909	0.811	0.970
DR3	0.488	0.700	0.958
DR4	0.733	0.805	0.976

(2) 不同模型结构加入 GAM 注意力机制的性能对比

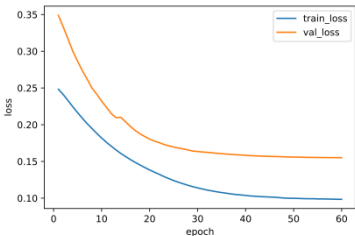
为了探究本文所提出的 GAM 注意力模块的分类效果，分别以 MobileNetV2、EfficientNetB0 和 ME-Net 作为基本的特征提取器，并加载各自的初始化预训练权重，将 3 个模型的特征输出最终分别输入到一个 GAM 模块中，将特征提取网络与 GAM 注意力模块组合为 1 对 1 的串联结构来分析模型效果。如表 3.9 所示，总体来看，相较于表 3.7 中的数据，加入该注意力模块后，在分类准确率、特异性、灵敏度和 Kappa 值上均有所提升，特别是在 ME-Net 特征提取网络下，Kappa 值达到 0.87，提升了近 3 个百分点，说明所提出的 GAM 模块更适合直接与特征提取器结合进行分类任务。模型训练曲线如图 3.24 所示，模型的训练准确率达到 0.96，验证准确率达到 0.94 以上，提升了 1.2 个百分点。模型训练 loss 曲线由 0.25 下降到 0.10，验证阶段 loss 由 0.35 下降到 0.16 左右，无较大波动发生。模型在第 40 个训练 epoch 时接近收敛状态，总体来讲，分类评价指标数值有所提升，GAM 注意力模块发挥了一定的效果。

表 3.9 加入注意力机制模型性能对比

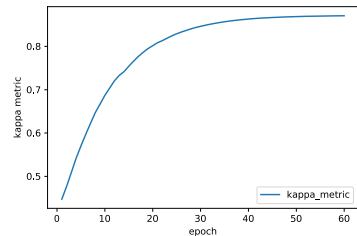
模型	准确率	特异性	灵敏度	Kappa 值
MobileNetV2+GAM	0.923	0.921	0.862	0.830
EfficientNetB0+GAM	0.917	0.924	0.870	0.825
ME-Net+GAM	0.942	0.936	0.889	0.872



(a) Accuracy 曲线图



(b) Loss 曲线图



(c) Kappa 曲线图

图 3.24 加入注意力机制的模型结构训练曲线图

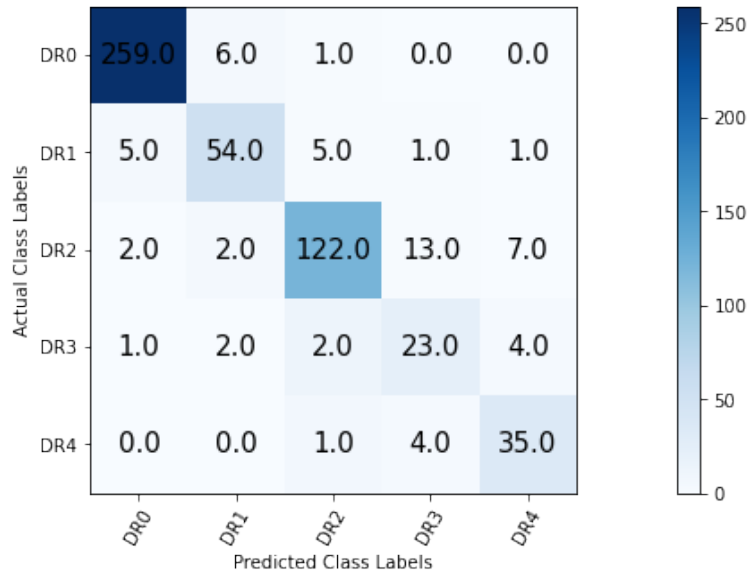


图 3.25 ME-Net+GAM 模型预测的混淆矩阵

从混淆矩阵分析得出，在 550 张测试集中，有 493 张图片被模型正确预测，模型的整体测试准确率达到 89.6%，较 ME-Net 模型提高约 2 个百分点。表 3.10 展示了视网膜病变每个类别的精确率、召回率和特异度，除了模型在 DR3 类别中的精确率不足 0.6，在其余各指标中均超过 0.7，特别是模型的特异度中，在五个类别中均超过 0.95，与表 3.8 中 ME-Net 模型分类性能相比，添加注意力模块的性能更优。

表 3.10 ME-Net+GAM 模型性能对比

类别	精确率	灵敏度	特异性
DR0	0.970	0.974	0.972
DR1	0.844	0.818	0.979
DR2	0.931	0.836	0.978
DR3	0.561	0.719	0.965
DR4	0.745	0.854	0.976

(3) ME-ANet 模型训练策略性能对比

进一步探究头部迁移学习策略对模型的影响，设计一组对照实验，对于本文所设计的 ME-ANet 模型训练时，对于其集成头部 ME-Head 采取加载预训练权重和未加载预训练权重进行对比实验，结果如下表 3.11 所示。

表 3.11 加 ME-ANet 模型训练策略性能对比

训练策略	准确率	特异性	灵敏度	Kappa 值
非迁移学习	0.938	0.945	0.884	0.904
迁移学习	0.952	0.956	0.905	0.914

由上表分析可知, 无论是否采用迁移学习的方式, 本文设计的模型性能指标均达到不错的效果, 在迁移学习训练策略下, 模型的准确率为 0.952, 特异性为 0.956, 灵敏度为 0.905, Kappa 值达到 0.914, 各项指标均超过 0.9, 各项指标值均超过非迁移的训练策略, 表明头部集成的迁移训练策略更进一步提升了模型的性能。Kappa 值达到 0.914, 比非迁移训练方式高 1 个百分点, 表明模型的训练和预测的一致性较高, 模型预测稳定性较好。图 3.26 展示了 ME-ANet 模型的迁移训练曲线图, 模型的训练准确率在 0.975 左右, 验证准确率达到 0.95 以上; 模型的训练 loss 最终下降到 0.05, 验证 loss 下降到 0.15, 均到达最低。与上述模型相比, 本模型获得更高的评价指标值, 也证明了本文设计模型的有效性, 及模型训练策略的有效性。

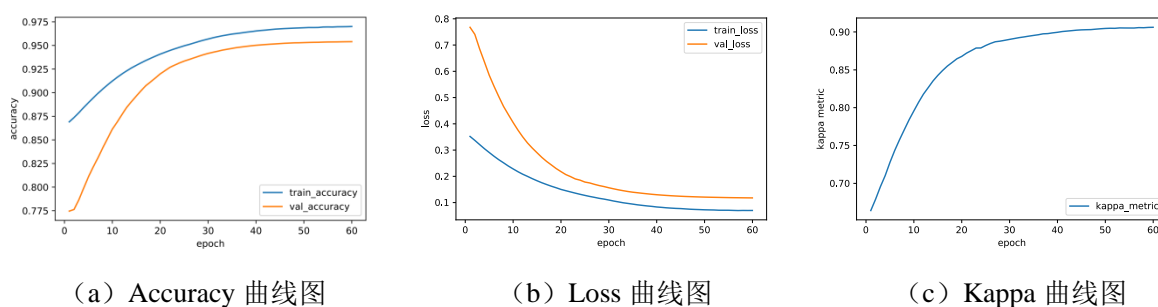


图 3.26 ME-ANet 迁移训练曲线图

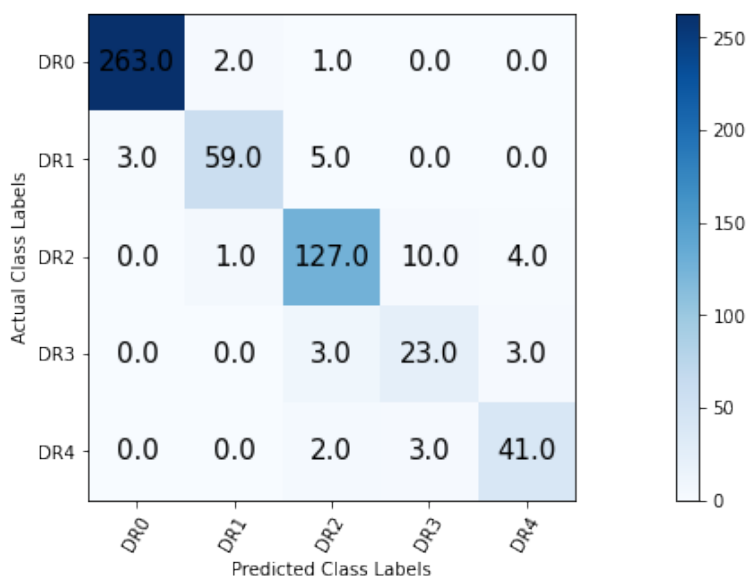


图 3.27 ME-ANet 模型预测的混淆矩阵

从混淆矩阵分析得出, 在 550 张测试集中, 有 513 张图片被模型正确预测, 模型的测试准确率达到 0.933, 与无 GAM 注意力模块相比, 模型的整体测试准确率由 0.875 上升到 0.933, 提高了近 6 个百分点; 相比于 ME-Net 模型的 0.896 的准确率, 本模型提高了近 4 个百分点。进一步分析可知, 视网膜病变眼底图像相邻类别错分的数量显著减少, 说明所设计的 GAM 注意力机制更适用于本课题的分类任务。表 3.12 展示了视网膜病变

每个类别的精确率、灵敏度和特异性，模型在 DR3 类别中的精确率超过 0.6，在其余各指标中均超过 0.85，特别是模型的特异度中，在五个类别中均超过 0.97，达到较好的分类效果。

表 3.12 ME-ANet 模型性能对比

类别	精确率	灵敏度	特异性
DR0	0.989	0.989	0.989
DR1	0.952	0.881	0.994
DR2	0.920	0.894	0.973
DR3	0.639	0.793	0.975
DR4	0.854	0.891	0.986

最后，为了评估本文分类模型 ME-ANet 的分类性能，本文使用相同的数据集和实验训练参数对近年来经典的方法进行训练，为了保证实验的客观性，在模型训练过程中我们不采用迁移学习的策略，并进行实验性能对比，如下表 3.13 所示，在准确率、特异性和二次加权 Kappa 值均达到了最大值，都超过了 0.900，并且灵敏度达到 0.874。进一步表明，本文的模型分类达到了不错的效果。

表 3.13 本文模型与近年来现有深度模型性能对比

模型	准确率	特异性	灵敏度	Kappa 值
AlexNet	0.897	0.910	0.812	0.814
GoogleNet	0.913	0.924	0.876	0.858
ResNet50	0.925	0.935	0.818	0.793
InceptionV3	0.883	0.889	0.782	0.872
本文方法	0.938	0.945	0.874	0.904

3.6 本章小结

本章所提出的一种基于模型融合的混合注意力机制网络用于糖尿病视网膜病变分类，利用 Kaggle 数据集进行模型训练，为了解决数据不充分及分布不平衡的问题，本文采用数据预处理及数据扩增等方式做均衡处理。采用模型集成方式构建特征提取器，并基于迁移学习的方式训练模型，以充分提取低层特征信息。同时改进设计注意力机制，提高模型提取微小病灶特征信息的能力，克服眼底图像之间较小的差异性而导致的模型分类准确率无法进一步提升的困难，使网络更好的学习到 DR 病变类型之间的细微差异，达到一种细粒度分类的效果。实验部分设计三组对比实验分别验证模型结构、注意力机制和模型训练策略对分类模型的影响，同时采用准确率、灵敏度、特异性、Kappa 值和混淆矩阵来评价模型。实验结果表明，本文设计的糖尿病视网膜分类模型效果较好。

第4章 基于深度学习的糖尿病视网膜渗出物分割

4.1 引言

本文的上一章节设计深度学习分类模型并完成对糖尿病视网膜眼底图像的五分类任务，而在临床诊断中，仅仅根据糖尿病视网膜彩色眼底图像实现简单的分类并不能很高效的辅助医生进行诊断，在正确分类的基础上，若能够实现对关键病灶点的区域分割，将会更有助于辅助医生诊断，极大提供 DR 的诊断效率和准确率。为此本文进一步分析视网膜病变图像中的关键病灶，实现对糖尿病视网膜渗出物区域的自动识别分割。

微动脉瘤是 DR 早期病变的标志性病灶点之一。虽然对微动脉瘤的准确诊断能够在第一时间发现病情，但是在 DR 病变早期，该病灶点非常小，不论是采用人工标记还是利用机器学习的特征提取，实现精细的分割都是比较困难的。渗出物包括硬性渗出物和软性渗出物，该病灶呈黄白色相对来说易于观察。在实际临床中，渗出物通常被视为诊断 DR 病变分类的重要参考指标^[81]，因此，利用深度分割模型进行渗出物的精确分割，对实现计算机辅助医疗诊断具有非常重要的现实意义。DR 图像中渗出物如图 4.1 所示。

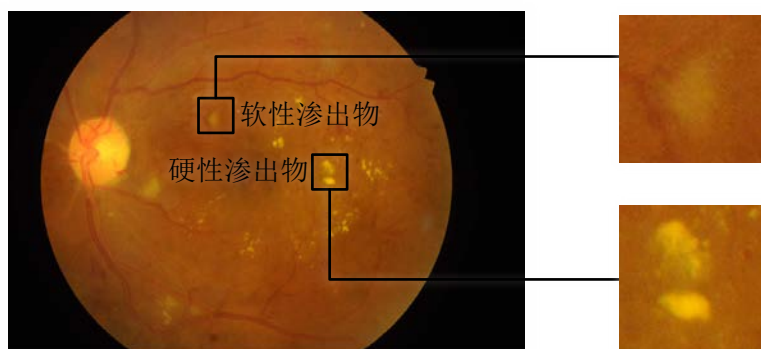


图 4.1 含有渗出物的糖尿病视网膜眼底图像

图 4.1 可以清楚的观察到两种渗出物的分布情况，渗出物在原始图像上占的比例非常小，形状不规则并且其边界比较模糊，这种分散的小目标分割给计算机带来了一定的挑战，另外，从图中可以看出，渗出物的亮度较高，与视盘的亮度相近，也给分割带来了一定的困难。

本章节同样分为模型构建、数据准备、模型训练与结果分析三部分来论述并验证所提出的分割方法模型。首先构造基于深度学习的渗出物分割模型，深度模型具有自动提取目标特征信息的能力，而无需人工对渗出物病灶区域进行复杂的特征提取，凭借此优势，所提出的分割方法能够从大量的 DR 眼底图像中自动学习并提取关键的病灶点特征信息。具体来讲，本章模型构建基于经典的 U-Net 编码器-解码器结构，有效整合目标区域特征的上下文特征信息，实现了病灶像素值的精确定位。在数据的准备方面，数据集

的处理采用第三章预处理方式，最后设计不同的实验对比模型进行训练，并根据实验结果进行分析。

4.2 视网膜渗出物分割模型

4.2.1 模型总体结构

U-Net 由 Ronneberger 等人^[82]提出，使用编码收缩和解码扩展路径进行生物医学图像分割。模型编码端进行特征提取，解码端进行上采样恢复图像分辨率。U-Net 充分利用每个卷积的特征输出，获得不同尺度的上下文特征信息，特别是模型中间的跳跃连接，弥补了上采样过程中特性信息丢失的问题，使 U-Net 比全卷积神经网络（FCN）学习到更加全面的特征信息。具体来讲，从收缩路径到扩展路径的跳跃连接有利于产生更加精确的分割效果。收缩路径由四个编码器单元组成，每个编码器单元都有两个卷积层，紧接一个最大池化层。每一次池化操作后，特征数量翻倍。收缩和扩展路径之间的中间部分称为瓶颈层，该部分由两个卷积层和一个上采样组成。四个解码器单元构成右侧 U-Net 的扩展路径，每个路径由反卷积、来自相应编码器输出特征图的跳跃连接和两个卷积层组成。

改进 U-Net 左侧的收缩路径，即编码端，将 ResNet50 架构取代原有的编码器结构，且 ResNet50 架构已经在 ImageNet 上进行了预训练；解码器由多个 U-Attention 组成，在 U-Attention 后设置特征融合层，实现多尺度特征融合；另外，模型将 LeakyReLu 作为激活层，替换了 Relu 激活层，这一做法有效解决了 Relu 反向传播过程中遇到的“死”神经元节点的问题，同时加快了梯度的传播速度。具体如图 4.2 所示。

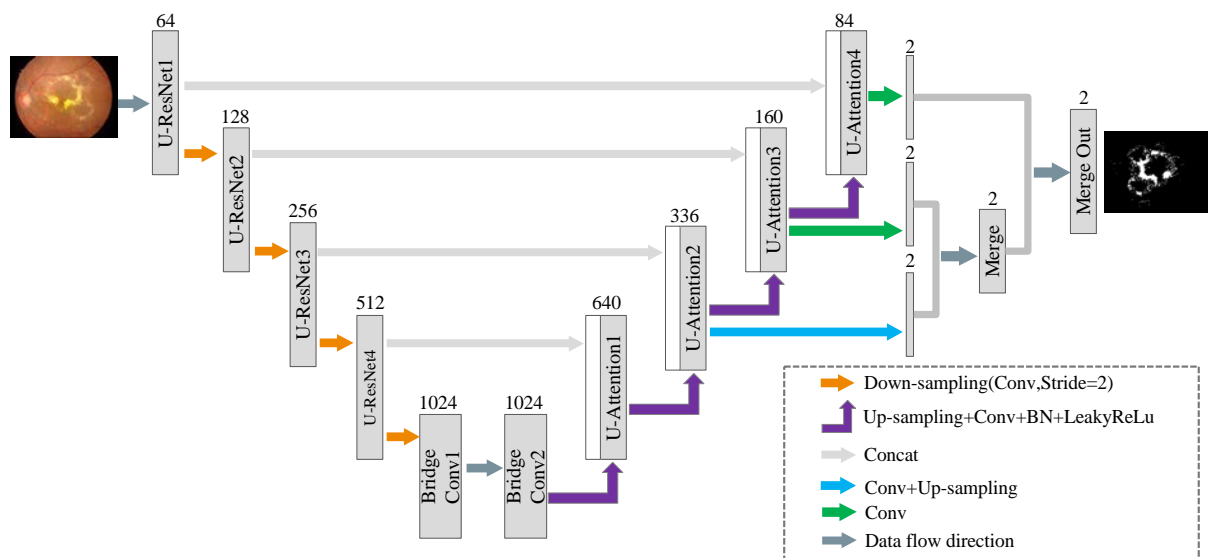


图 4.2 改进 U-Net 的渗出物分割网络模型结构

4.2.2 具体模型设计

编码器由预训练的 ResNet50 架构组成，与其他预训练模型相比，残差网络在分类问题上的准确率更好，除此之外，残差网络能够在保证梯度有效传播的前提下，加深模型的深度，减小梯度消失的问题。编码器由 4 个 ResNet 块构成，在每个 ResNet 块中由多个改进的 bottleneck 块组成，即称为 U-Bottleneck，具体如下图 4.3 (a) 所示。编码端构成如下表 4.1 所示。在模型的底部，有两个中间卷积层，在这里将其命名为 Bridge Conv1 和 Bridge Conv2，用于连接编码端与解码端，该结构包含两个连续的卷积运算，每次卷积运算结束后，利用 LeakyReLU 函数进行一次激活，然后再进行一次批量归一化操作。具体如图 4.3 (b) 所示。解码端由 4 个 U-Attention 单元组成，如下图 4.3 (c) 所示，将第 3 章设计 GAM 注意力模块加入解码单元中，将其记为 GAM-U-Attention 解码单元。将上采样的输出特征输入到注意力模块中，强化学习目标特征，最后输出的结果与编码端对应的原始高分辨特征相结合，充分利用生成的上下文特征信息同时减少上采样过程中的特征损失问题，从而提高模型对视网膜病灶特征的细粒度分割。

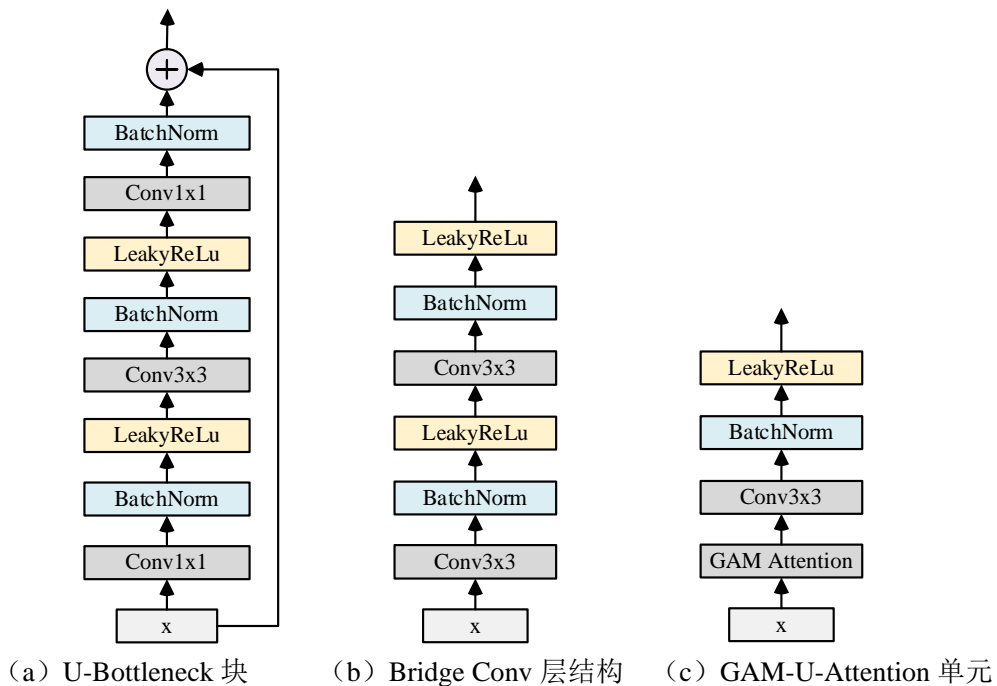


图 4.3 改进 U-Net 具体结构示意图

表 4.1 改进编码器结构配置表

编码单元名称	编码单元组成	数量
U-Resnet1	U-Bottleneck	3
U-Resnet2	U-Bottleneck	4
U-Resnet3	U-Bottleneck	6
U-Resnet4	U-Bottleneck	3

4.3 糖尿病视网膜眼底图像处理

4.3.1 实验数据集

本章对于模型的训练采用 IDRID (Indian Diabetic Retinopathy Image Dataset), IDRID 数据集是在 IEEE 国际生物医学影像研讨会 (ISBI-2018) 举办的“糖尿病性视网膜病变分割和分级挑战”竞赛上公开的^[29], 该数据集由三部分组成, 分别用于病变分级、分割和定位任务, 总共有 597 张眼底图像, 图像是使用 Kowa VX-10a 数字眼底照相机拍摄的, 尺寸为 4288×2848 , 存储格式为 jpeg。本实验使用的是分割数据集, 此部分共 81 张眼底图像, 该数据集分为训练集 54 张和测试集 27 张, 每张眼底图片都有对应病灶区域的像素级别精准标注, 这里的病灶主要包括四类: 微动脉瘤、软性渗出物、硬性渗出物 and 出血点^[80]。图 4.4 展示了该数据集中的原始图像及渗出物标注区域。

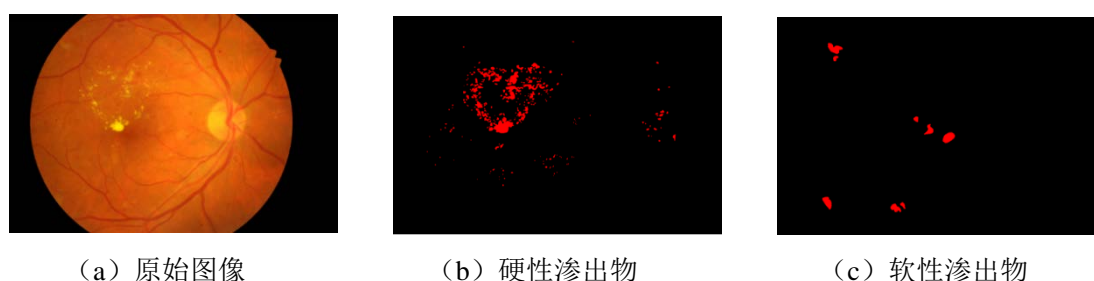


图 4.4 IDRID 数据集样例以对应渗出物标注图

硬性渗出物和软性渗出物通常为白色或者黄白色, 将其称为白色病变区域。数据集构建时, 本文创建一个新的空白 Mask, 将病灶分割图按位运算, 将含有渗出物区域进行整合, 同时将该区域的像素值设为 1。处理后的彩色眼底图片及相应的掩码如下图 4.5 所示。

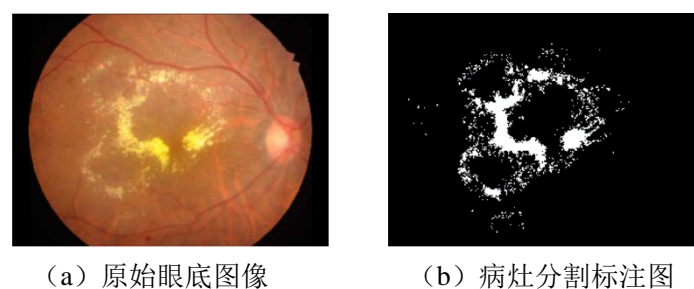


图 4.5 眼底图像及病灶分割图

4.3.2 数据预处理

随着技术的发展, 医疗设备采集到的医学图像质量不断增高, 特别是对于眼底图像的采集, 设备的抵抗干扰能力也不断增强。本章节使用的 IDRID 数据集是由专家统一审核标注的, 图像质量较高, 无需人工剔除质量较差的图片。在数据预处理方面主要是借

鉴第三章数据预处理中所涉及图像对比度增强和图像去噪的方法，只是本章图像处理后的分辨率设置为 1024×1024 。通过预处理，提高了图像的对比度，减少了图像噪声，进一步增强了图像的质量，有利于提升渗出物分割效果。具体处理流程如下图 4.6 所示。

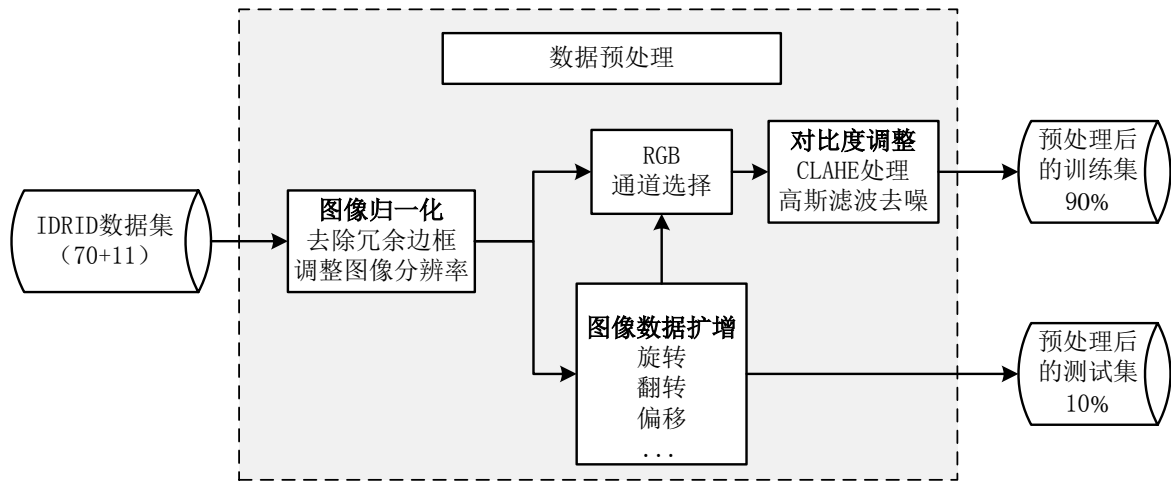


图 4.6 IDRID 眼底图像预处理流程图

由上图可知，本文将原始的 81 张 IDRID 数据集分 70 张训练集和 11 张测试集，整个数据集扩充到 1000 张图像，其中最初原始的 70 张训练集扩充到 900 张图像，作为训练集，最初的 11 张测试集采用图像扩增方式增加到 100 张。数据预处理效果如图 4.7 所示。

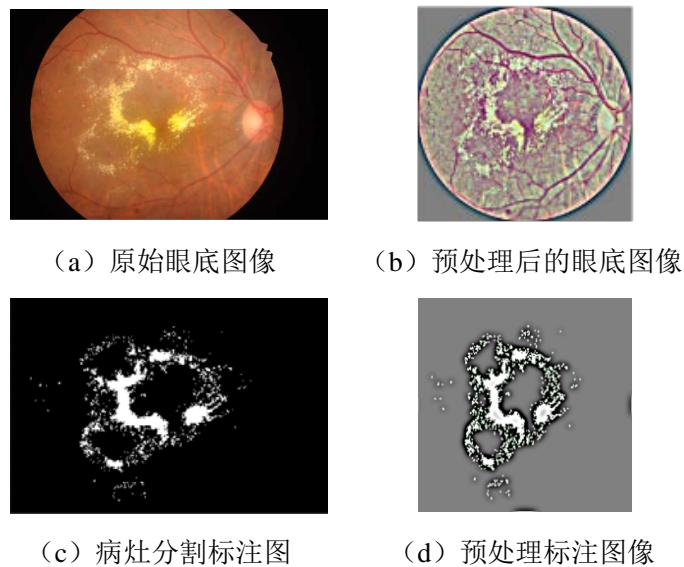


图 4.7 眼底图像及病灶分割处理效果图

在图像增强处理中，数据集集中的原始眼底图像和人工标注的渗出物分割图必须进行一致的处理。在使用受限的对比度直方图增强对比度时，通过多次处理观察，最终设定参数 clipLimit 为 3，tileGridSize 为 8，其余处理参数与第三章 3.3.2 数据预处理一致。经过预处理后，图像的对比度显著提高，如图 4.7 (b) 所示，图像中的血管及渗出物更加

明显，在训练中有利于卷积对渗出物特征的自动提取。图 4.7 (d) 展示了预处理后的渗出物标注图像，与图 4.7 (c) 相比，处理后的图像渗出物轮廓更加清晰。

4.4 实验结果与分析

4.4.1 实验环境

实验深度学习框架为 Tensorflow2.0 版本，采用 Python3 编程语言进行模型框架的搭建，实验在 GPU 型号为 NVIDIA RTX 2080Ti、16G 显存的 Ubuntu16.04 操作系统，CUDA10.0, cudnn7.6.0 的服务器上进行训练调试。模型训练过程中使用 Adam 优化器进行参数调优，初始学习率设为 0.00005，batch_size 设置为 1，Loss 计算过程中采用 Focal loss，同时采用可变的学习率，学习率衰减因子设置为 0.01，且调整学习率等待周期为 10 个 epoch。训练时采用 5 折交叉验证策略，从而选出性能指标最佳的模型。

4.4.2 损失函数

在训练模型的过程中，由于眼底图像中病灶点如渗出物占整幅图像比例不等，绝大部分图片中的渗出物像素值不足总像素值的十分之一。在深度学习中，这种分布严重不均衡的数据会影响模型的分类、分割精度，使模型倾向于预测样本数量较多物体。为缓解模型训练中由于样本正、负像素点分布不均衡而导致模型训练效果不佳的问题，本文采用 Focal Loss 作为模型优化的目标函数。Focal Loss 损失函数最初用于目标识别，其目的在于解决由于样本数目差异过大而导致的分类不均衡问题。Focal Loss 是受交叉熵损失函数（Cross Entropy）启发而设计的，交叉熵损失公式如下：

$$C(p, y) = \begin{cases} -\log(p), & y = 1 \\ -\log(1-p), & y \neq 1 \end{cases} \quad \text{公式 (4.1)}$$

式 (4.1) 是二分类的交叉熵损失函数的表达形式，其中图像标签用 y 表示，图像的预测概率用 p 表示^[83]。在二分类，模型最后需要预测的结果只有两种情况，对于每一个类别，预测得到的概率为 p 和 $1-p$ ^[84]。另外，利用式 (4.2) 可将公式 (4.1) 公式进一步转换，可以得到交叉熵损失的另一种表示形式，具体如式 (4.3) 所示。

$$P = \begin{cases} p, & y = 1 \\ 1-p, & y \neq 1 \end{cases} \quad \text{公式 (4.2)}$$

$$C(p, y) = C(P_i) = -\log(P_i) \quad \text{公式 (4.3)}$$

由式 (4.3) 分析可知，交叉熵损失函数无法解决样本数量不均衡的问题。为此，学

者提出了加入权重约束的策略,通过权重大小进而控制正负样本的数量,缓解数据不平衡问题对模型的影响,加权损失函数如式(4.4)所示。

$$L = \frac{1}{N} \sum_i L_i = -\frac{1}{N} \sum_i [\beta_1 \cdot y_i \cdot \log(p_i) + \beta_2 \cdot (1 - y_i) \cdot \log(1 - p_i)] \quad \text{公式 (4.4)}$$

其中, N 为样本总量; β_1 、 β_2 为自设定权重, 其中 $\beta_1 + \beta_2 = 1$; y_i 表示样本 i 的真实标签, 正类为 1, 负类为 0; p_i 表示样本 i 预测为正类的概率。上式交叉熵损失函数有效解决了样本数据量层面不平衡的问题, 但是对于单个样本中所含有的目标特征如本章研究的渗出物特征像素点仅仅占整个背景像素的很小一部分, 非目标像素值占了绝大部分, 使得模型难以有效提取到所有的目标特征。大多数非渗出物的特征被模型学习, 对于提高模型精度和网络的泛化帮助不大^[85]。

由于改进后的加权交叉熵损失函数无法实现对图像中小目标且难以区分样本的强化学习, 故提出 Focal Loss, 该函数也是基于交叉熵损失函数改进而来。Focal Loss 损失函数能够有效解决图像中有用样本与无用样本的分布不均衡问题, 该损失函数利用指数参数作为权重, 根据之前预设的经验值, 调整难易样本所占的比例。具体使用 γ 使损失函数更加关注难以区分的目标, 本文所使用的公式如下(4.5)所示:

$$FL_{loss} = -\alpha(1 - p_i)^\gamma \log(p_i) \quad \text{公式 (4.5)}$$

其中, α 为平衡焦点损失, 调整正负样本的权重, p_i 是图像中像素分类的概率, γ 是调节因子。当 γ 为 0 时, 焦点损失函数等同于分类交叉熵。随着 γ 的增加, 调节因子的影响同样增加, 实验证明, 当 α 取值为 0.25, γ 取值为 2 时, 效果最好。

4.4.3 实验评估指标

医学上对评价对象存在四个统计参数: 真阳性 (TP) 是指像素点的实际标签为渗出物, 在模型预测的结果也为渗出物; 假阳性 (FP) 是指像素点的实际标签为非渗出物, 但在模型预测结果中为渗出物; 真阴性 (TN) 是指像素点的实际标签为非渗出物, 在模型预测结果中也为非渗出物; 假阴性 (FN) 是指像素点的实际标签为渗出物, 但在模型预测结果中为非渗出物^[86]。具体的渗出物分类表如表 4.2 所示。

表 4.2 渗出物四种量化统计指标

类别\类别	渗出物标记像素点	非渗出物标记像素点
预测渗出物像素点	真阳性 (TP)	假阳性 (FP)
非预测渗出物像素点	假阴性 (FN)	真阴性 (TN)

由上表可知, 模型的 TP、TN 值越高, 表明模型的性能越好, FP、FN 值越低, 模型效果越差。在实际诊断中, FN 要尽量控制到最低, 这种诊断错误会导致患者由于模

型的诊断错误而错过早期最佳治疗时期。另外,根据上述四个统计量^[87],可以计算出三个评价指标:灵敏度(SE)、特异性(SP)、精确率(PPV),具体公式在第三章3.5.2章节中已给出,此处不再赘述。另外,本文增加 F1-Score 和 Dice 评价指标,具体如式(4.6)、(4.7)所示,以便更加全面的评价模型。

$$F1-Score = \frac{2TP}{2TP + FN + FP} \quad \text{公式 (4.6)}$$

$$Dice = \frac{2TP}{2TP + FP + FN} \quad \text{公式 (4.7)}$$

灵敏度用来评估模型正确预测出渗出物的比例^[88],由其公式分析可知,若将 TP 视为定值,则灵敏度的高低是由 FN 决定的,即 FN 越低,则灵敏度越高,进一步说明模型诊断错误的概率降低,模型对渗出物诊断的能力越强;特异性表示正确预测出非渗出物的比例,它的高低体现了模型对于噪声的抗干扰能力,特异性越高,说明模型的抗噪能力越强;精确率表示真实渗出物标记值占有所有预测结果为渗出物像素值的比例,精确率越高,说明模型的分割精度越高。F1-Score 是精确率与召回率的调和平均值,取值范围在 0-1 之间,模型的值越高,表示模型的分割效果越好。Dice 是在医学图像比赛中经常用到的评价指标,它是一种衡量模型分割效果与专家真实标注图之间相似度的指标,取值范围也在 0-1 之间,分割效果最好时值为 1,最差时值为 0。

4.4.4 实验设计与分析

(1) 5 折交叉验证测试

在模型训练中,为了选择最佳的模型,我们需要利用不同的数据集去验证模型的效果,评估模型对新数据的泛化程度。然而,在数据量不足的情况下,若对同一部分数据进行训练和测试并不能客观的验证模型效果,而交叉验证可以很好的解决这个问题。交叉验证每次使用数据集的一部分用于训练,另一部分用于测试,避免两种数据集的交叉,可以客观的评估模型性能。在机器学习中有多种方式可以实现交叉验证,有留一法、分层 k 折交叉验证、随机拆分交叉验证、k 折交叉验证^[89]等。本文采用 k 折交叉验证方法训练模型,训练时 k 取值设置 5,即 5 折交叉验证,最后从训练后的 5 个模型中选择效果最好的模型参数,具体如下图 4.8 所示。

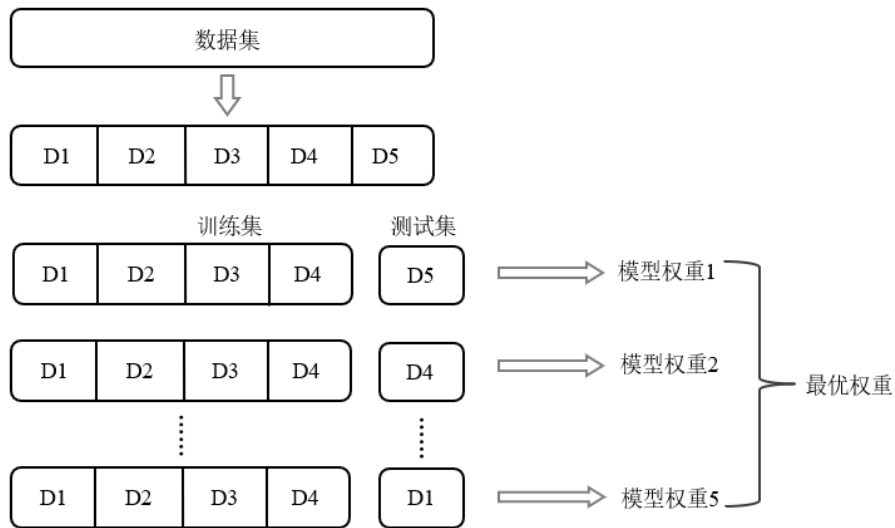


图 4.8 五折交叉验证示意图

模型对于新数据的适应能力越强,其可靠性越高。本文模型采用5折交叉验证方式,利用 IDRID 数据集,每轮的训练过程评价指标如下表 4.3 所示。由表中数据可知,在第二轮交叉验证时,模型在 IDRID 数据集训练下的特异性、灵敏度、精确率、F1-Score 和 Dice 性能指标分别为 0.973、0.982、0.971、0.875 和 0.984,此时模型状态达到最优,故保留模型训练参数。

表 4.3 每一轮交叉验证性能对比果

轮数	特异性	灵敏度	精确率	F1-Score 值	Dice 值
1	0.942	0.961	0.951	0.868	0.962
2	0.973	0.982	0.971	0.875	0.984
3	0.921	0.981	0.953	0.859	0.958
4	0.941	0.978	0.965	0.868	0.966
5	0.961	0.975	0.970	0.871	0.980

(2) Focal Loss 损失函数参数测试

使用 Focal Loss 损失函数进行训练,设置不同的 α 和 γ 的值,分析模型的性能。在 Focal Loss 原始论文实验中,发现当 $\gamma=2$ 时模型效果最佳,故本文在实验中选择默认的 γ 值,只是调整 α 来分析效果。检测结果如下表 4.4 所示,由表中数据分析可知,当 $\alpha=0.25$, $\gamma=2$ 时, Dice 值达到 0.984,模型效果最佳。

表 4.4 Focal Loss 不同参数性能对比果

参数设置	特异性	灵敏度	精确率	F1-Score 值	Dice 值
$\alpha=0.25$	0.973	0.982	0.971	0.875	0.984
$\alpha=0.50$	0.942	0.935	0.967	0.835	0.974
$\alpha=0.75$	0.921	0.938	0.965	0.859	0.958

(3) 损失函数参数测试

通过大量的实验证明,在加权损失函数中, $\beta_1=0.8$, $\beta_2=0.2$ 时效果最好,故在下表的加权损失函数对比中,也选中上述权重参数进行实验对比,具体在不同的损失函数下的模型训练性能指标如下表 4.5 所示。

表 4.5 不同损失函数性能对比果

损失函数类型	特异性	灵敏度	精确率	F1-Score 值	Dice 值
交叉熵函数	0.916	0.825	0.913	0.823	0.965
加权损失函数	0.923	0.905	0.923	0.839	0.973
焦点损失函数	0.973	0.982	0.971	0.875	0.984

从上表分析可知,传统的交叉熵损失函数无法处理数据不平衡问题^[90],难以实现准确的预测结果,加权的交叉熵损失函数解决了数据层面分布不均的问题,但是无法解决单个图像中容易和困难样本的学习。上表数据表明,焦点损失函数很好的解决了数据不平衡和困难样本的学习问题,大幅度提升了模型的训练效果。

本文所设计的糖尿病视网膜病灶分割模型可以实现对渗出物病灶进行比较精确的分割,并且与眼科专家标注的像素级标签图像的一致性较高。另外,本文设计的全局注意力机制进一步提升模型的分割效果,为了表明本文所设计注意力模型的有效性,构建 SE 注意力层嵌入到模型 U-Attention 层结构中,本文将此结构命名为 SE-U-Attention,与本文设计的 GAM-U-Attention 进行分割效果对比。具体分割效果对比如图 4.9 所示。

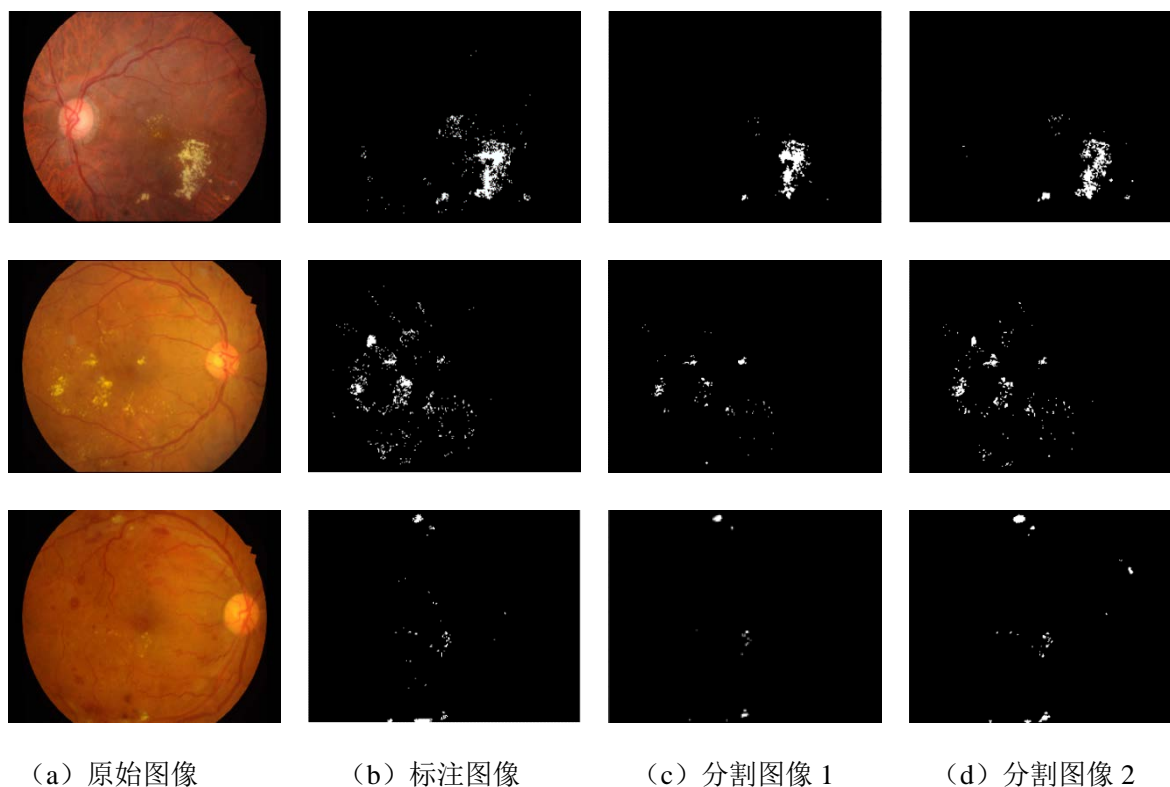


图 4.9 眼底图像渗出物分割效果图

图 4.9 (c) 分割图像 1 所对应的是含有 SE-U-Attention 的分割模型, 4.9 (d) 分割图像 2 所对应的是含有 GAM-U-Attention 的分割模型, 通过观察两模型的分割效果图, 可以看出, 本文设计的模型在渗出物分割效果上更接近原始标注图像, 特别是对于分散的点状渗出物分割效果较好, 且边界较为清晰。表明本文所设计的注意力机制更适合于糖尿病视网膜渗出物分割任务。

为了进一步说明本文设计的注意力模块有助于提升渗出物分割效果, 图 4.10 展示了分割的具体细节对比, 如下图所示。

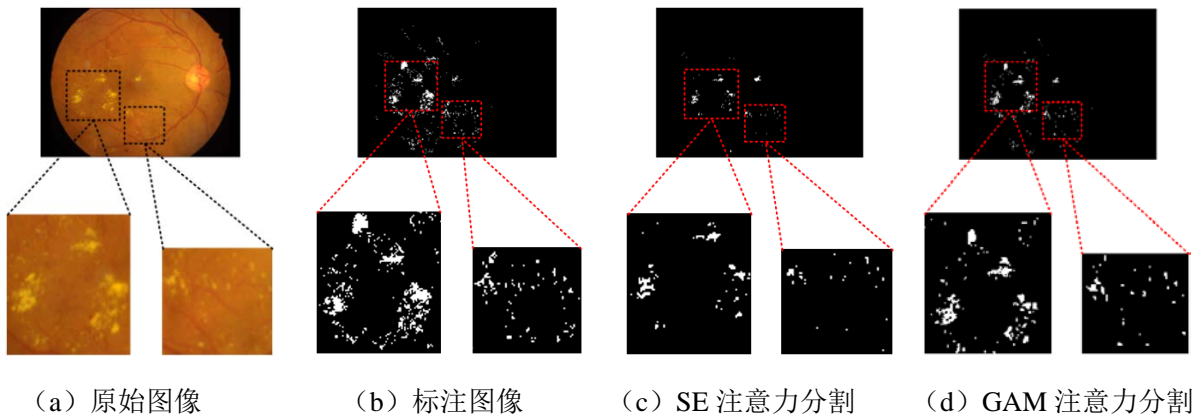


图 4.10 模型分割细节对比图

图 4.10 (a) 为原始眼底图像, 图 4.10 (b) 为原始的标注图像, 图 4.10 (c) 为带有 SE-U-Attention 分割模型的预测分割图, 图 4.10 (d) 为本文带有 GAM 的分割模型的预测分割图。从放大的分割细节图分析可知, 嵌入 GAM 注意力层结构的分割模型效果与标注图像基本一致, 对于渗出物区域的边缘分割精度较高, 相比于 SE 注意力机制, 本模型病灶分割的主体轮廓基本清晰, 对于渗出物主体之外的点状渗出, 模型依然能够有效的分割出来, 且分割效果远超过 SE 注意力机制的分割。进一步说明了 GAM 注意力机制能更好地结合上下文信息, 维持较长的特征依赖关系, 对于提高分割的细节和精度有显著作用。

最后, 将本文所提出的算法与近几年病灶分割算法进行比较, 比较结果如下表 4.6 所示, 除了在 F1-Score 的得分比刘等人的方法低 1.8 个百分点外, 其他各项指标均获得了较高水平, 本文模型的 Dice 值达到 0.984, 远高于其他模型, 进一步说明本文模型的分割效果较好, 与专家标注的图像一致性最高。

表 4.6 不同模型算法性能对比

方法	特异性	灵敏度	精确率	F1-Score 值	Dice 值
刘等人 ^[41]	0.921	0.897	0.867	0.893	0.935
Guo 等人 ^[39]	0.937	0.971	0.925	0.853	0.905
Wang 等人 ^[43]	0.964	0.899	0.927	0.870	0.947
本文方法	0.973	0.982	0.971	0.875	0.984

4.5 本章小结

为了更加精确的实现糖尿病视网膜渗出物的分割,为了在糖尿病视网膜病变分类的基础上进一步实现关键病灶点渗出物的分割,本文构建了一种基于深度学习的糖尿病视网膜渗出物分割模型,该模型以 U-Net 编码器-解码器结构为基本网络架构,同时结合 ResNet50 作为编码端的特征提取网络,实现对特征提取网络的加深,避免了模型训练中梯度消失的问题;在解码端,设计注意力层结构,并嵌入到解码单元中,来提升模型对小病灶点的识别能力;最后在模型解码端构建多特征尺度融合层,实现模型多尺度语义特征的学习。实验过程中利用 IDRID 数据集进行训练,训练时采用 Focal loss 焦点损失函数,进一步强化模型对样本中小目标物体特征的学习优化。与其他渗出物分割算法相比,该算法分割性能较优,且具有普遍适用性。

第5章 总结与展望

5.1 全文总结

随着人类物质文明的极大发展,人们的生活方式已经彻底改变,快节奏逐渐成为社会发展的趋势,而潜在的各种慢性疾病也不断涌现而来,影响着人们的健康。其中,糖尿病是就是严重威胁现代人健康的疾病之一,而糖尿病视网膜病变是就是其典型的并发症之一,身体长期维持在高血糖水平下会造成人类眼部血管发生异常,如果患者诊断、治疗不及时则会造成眼部病情快速恶化,轻者影响正常视力,严重的可导致失明。现阶段,我国已经成为世界上糖尿病患者数量最多的国家,在此背景下,对糖尿病患者的定期筛查成为一项迫切且艰巨的任务。传统的 DR 筛查效率低且严重依赖眼科医生的诊断经验,无法满足准确、高效的诊断。近年来,随着计算机科学的高速发展,进一步推动了深度学习在医学领域中的应用进程。尤其是在医学影像领域,深度学习逐渐成为一个热门课题。近年来,基于深度学习的 DR 图像分析已经在国内外已经取得了不错的成果,在前人研究的基础上,本文利用深度学习模型对 DR 图像进行了分类和渗出物区域分割,更好的实现计算机辅助医生诊断的目的,具有一定的研究价值。本文利用深度学习对 DR 图像进行分析研究,具体工作如下:

(1) 针对糖尿病视网膜病变图像的分类与渗出物分割公开数据集数量有限等问题,为满足深度模型训练任务中需要大量的数据集作为支撑。本文针对本课题使用的分类、分割数据集进行一系列的预处理,采用人工筛选和现有的图像处理技术对公开的数据集进一步加工,如必要的质量筛选,对比度调整、图像平滑去噪、数据增强等方法。同时采用图像扩增技术对数据进行平衡扩增,克服数据量不足和分布不均衡问题。

(2) 针对现有的 DR 分类模型对病变类别之间的分类准确率较低等问题,本文提出了一种基于模型集成的全局注意力机制分类网络: ME-ANet。该网络有效解决了深度模型由于 DR 病变图像之间相似度较高而难以准确分类的问题。具体来讲,采用集成的思想,将 MobileNetV2 和 EfficientNetB0 的核心结构分阶段进行集成融合,组建全新的特征提取网络,能够实现对视网膜图像多尺度特征的学习;同时,设计并结合全局注意力机制嵌入至集成模块中,进一步获取病变特征的权重信息,抑制非病变特征噪声进行关注学习;最终,网络实现了对图像底层特征信息的充分提取,减少了微病灶特征信息在训练过程中的卷积丢失问题,有效提升了模型的性能。本文所设计的模型利用 Kaggle 数据集进行训练,最后准确率、特异性、灵敏度和 Kappa 值分别达到 0.938、0.945、0.874 和 0.904,与其他模型对比可知,本文构建的 DR 分类模型效果较好。

(3) 为了在糖尿病视网膜病变分类的基础上进一步实现模型对关键病灶点渗出物区域的分割,本文构建了一种基于深度学习的糖尿病视网膜渗出物分割模型,该模型以

U-Net 编码器-解码器结构为基本网络架构，同时结合 ResNet50 作为编码端的特征提取网络，不仅实现对特征提取网络的加深，还避免了模型训练中梯度消失的问题；在解码端，设计注意力层结构，并嵌入到解码单元中，来提升模型对小病灶点的识别能力；最后在模型解码端构建多特征尺度融合层，实现模型多尺度语义特征的学习。实验过程中利用 IDRID 数据集进行训练，训练时采用 Focal loss 焦点损失函数，进一步强化模型对样本中小目标物体特征的学习优化。模型最后在评价指标特异性、灵敏度、精确率、F1-Score 和 Dice 中分别达到 0.973、0.982、0.971、0.875 和 0.984，与其他先进算法相比，该算法分割性能较优，且具有普遍适用性。

5.2 未来工作展望

本文提出的糖尿病视网膜分类模型和渗出物分割模型取得了一定的成果，有利于实现计算机辅助医生高效诊断的目的，为眼科医生日常的诊断提供了重要参考。然而，对于本文的研究工作来说，仍存在一些不足，也是未来研究中需要克服的问题：

(1) 本文模型训练所使用的眼底数据集都是国外公开的眼底图片，由于人种的差异，人们的视网膜组织也会有所不同。利用国外公开的眼底数据集训练出的模型，对我国的视网膜进行分析诊断是不客观的。在接下来的任务中，应积极主动地和各地眼科医院联系，搜集国内不同地区的糖尿病患者的眼底图像，并和眼科专家探讨眼底所属的类别和相关病灶的标注，构建国内数据集并进行模型训练。

(2) 技术的研发需要落地应用到实际生活中，不能总停留在实验室阶段。本文所构建的深度模型虽然取得了不错的效果，但却是以训练大量的参数、消耗大量的计算资源为代价，接下来的任务，将继续研究优化深度模型的方法，试图构建更加轻量化的深度模型，节约计算资源，使其更容易地嵌入到移动设备当中，这对于解决我国各地区相关医疗资源分布不均衡问题将会有非常重要的意义。

参考文献

- [1] 中国微循环学会糖尿病与微循环专业委员会, 中国医疗保健国际交流促进会基层卫生分会基层糖尿病学部, 江苏省基层内分泌特色科室孵化联盟. 基层糖尿病微血管病变筛查与防治专家共识[J]. 中国医学前沿杂志, 2021, 13(6): 16-38.
- [2] 何蓓蕾, 何媛. 糖尿病与非视网膜眼部并发症相关性的研究进展[J]. 国际眼科杂志, 2021, 21(04): 623-627.
- [3] 张小玲, 邱曙东, 陈艳炯, 等. 糖尿病性视网膜病变发病机制研究进展[J]. 国际眼科杂志, 2015, 5(6): 1239-1242.
- [4] 中华医学会眼科学会眼底病学组. 我国糖尿病视网膜病变临床诊疗指南[J]. 中华眼科杂志, 2014, 50(11): 851-865.
- [5] 范家伟, 张如, 陆萌, 等. 深度学习方法在糖尿病视网膜病变诊断中的应用[J]. 自动化学报, 2021, 47(05): 985-1004.
- [6] Huang C X. Long-term effects of pattern scan laser pan-retinal photocoagulation on diabetic retinopathy in Chinese patients: a retrospective study[J]. International Journal of Ophthalmology, 2020, 13(2): 239-245.
- [7] 于凤丽. 基于深度学习的糖尿病性视网膜图像分类研究[D]. 南京: 南京航空航天大学, 2018: 18-39.
- [8] Sadda S R, Nittala M G, Taweebanjongsin W, et al. Quantitative assessment of the severity of diabetic retinopathy[J]. American Journal of Ophthalmology, 2020, 218: 342-352.
- [9] Ehlers, Justis P, Jiang, et al. Quantitative Ultra-Widefield Angiography and Diabetic Retinopathy Severity An Assessment of Panretinal Leakage Index, Ischemic Index and Microaneurysm Count[J]. Ophthalmology, 2019, 126(11): 1527-1532.
- [10] Hipwell J H, Strachan F, Olson J A, et al. Automated detection of microaneurysms in digital red-free photographs: a diabetic retinopathy screening tool[J]. Diabetic medicine, 2000, 17(8):588-594.
- [11] Zhang B, Karray F, Wu X, et al. Hierarchical detection of red lesions in retinal images by multiscale correlation filtering[J]. Proceedings of SPIE The International Society for Optical Engineering, 2009, 7260: 72601L-72601L-12P.
- [12] 马文俊. 基于机器学习的糖尿病视网膜病变分级研究[D]. 哈尔滨: 哈尔滨工程大学, 2017: 16-48.
- [13] Usman Akram M, Khalid S, Tariq A, et al. Detection and classification of retinal lesions for grading of diabetic retinopathy[J]. Computers in Biology & Medicine, 2014, 45(2): 161-171.
- [14] 孙晨. 基于深度学习的糖尿病视网膜眼底图像分析[D]. 上海: 华东师范大学, 2019: 12-45.
- [15] 赵刚. 基于深度学习的糖尿病视网膜病变分类研究[D]. 兰州: 兰州理工大学, 2021: 2-56.
- [16] 范家伟, 张如如, 陆萌等. 深度学习方法在糖尿病视网膜病变诊断中的应用[J]. 自动化学报, 2019, 45(7): 1-21.
- [17] Adarsh P, Jeyakumari D. Multiclass SVM-Based Automated Diagnosis of Diabetic Retinopathy[C]. IEEE Communications and Signal Processing (ICCSP), 2013 International Conference, 2013: 206-210.

- [18] Shahin E M, Taha T E, Al-Nuaimy W, et al. Automated detection of diabetic retinopathy in blurred digital fundus images[C]. International Computer Engineering Conference (ICENCO), 2012: 20-25.
- [19] Nayak J, Bhat P S, Acharya R, et al. Automated identification of diabetic retinopathy stages using digital fundus images[J]. Journal of medical systems, 2008, 32(2): 107-115.
- [20] Ahuja A, Ashish K. Commentary: Change in Trends of Imaging the Retina[J]. Indian Journal of Ophthalmology, 2018, 66(11): 1620-1628.
- [21] Wang X, Lu Y, Wang Y, et al. Diabetic Retinopathy Stage Classification Using Convolutional Neural Networks[C]. IEEE International Conference on Information Reuse and Integration (IRI), 2018: 465-471.
- [22] Gulshan V, Peng L, Coram M, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs[J]. Journal of the American Medical Association, 2016, 316(22): 2402-2410.
- [23] 李琼, 柏正尧, 刘莹芳. 糖尿病性视网膜图像的深度学习分类方法[J]. 中国图像图形学报, 2018, 23(10): 1594-1603.
- [24] 丁蓬莉, 清勇, 张振, 等. 糖尿病性视网膜图像的深度学习神经网络分类方法[J]. 计算机应用, 2017, 7(3): 699-704.
- [25] Zhou K, Gu Z W, Liu W, Luo W X, et al. Multi-cell multi-task convolutional neural networks for diabetic retinopathy grading[C]. 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, 2018: 2724-2727.
- [26] Pratt H, Coenen F, Broadbent D M, et al. Convolutional Neural Networks for Diabetic Retinopathy[J]. Procedia Computer Science, 2016, 90: 200-205.
- [27] 李轩屹, 朱晓军. 多尺度深度CNN的糖尿病视网膜病变分类[J]. 计算机工程与设计, 2020, 41(11): 3229-3234.
- [28] 闫育铭, 李峰, 罗德名, 等. 基于深度迁移学习的糖尿病视网膜病变的检测[J]. 光学仪器, 2020, 42(05): 33-42.
- [29] 聂永琦, 曹慧, 杨锋, 刘静. 深度学习在糖尿病视网膜病灶检测中的应用综述[J]. 计算机工程与应用, 2021, 57(20): 25-41.
- [30] Yi X, Walia E, Babyn P. Generative adversarial network in medical imaging: A review[J]. Medical image analysis, 2019, 58: 101552.
- [31] Zhang X, Thibault G, Decenciere E, et al. Exudate detection in color retinal images for mass screening of diabetic retinopathy. Medical image analysis, 2014, 18(7): 1026-1043.
- [32] Harangi B, Hajdu A. Automatic exudate detection by fusing multiple active contours and regionwise classification[J]. Computers in biology and medicine, 2014, 54: 156-171.
- [33] 高玮玮, 沈建新, 程武山, 等. 基于改进的模糊 C-均值聚类算法及支持向量机的眼底图像中硬性渗出检测方法[J]. 北京生物医学工程, 2017, 36(4): 331-337.
- [34] Pereira C, Gongalves L, Ferreira M. Exudate segmentation in fundus images using an ant colony optimization approach[J]. Information Sciences, 2015, 296: 14-24.
- [35] Chen X, Bu W, Wu X, et al. A novel method for automatic hard exudates detection in color retinal

- images[C]. 2012 International Conference on Machine Learning and Cybernetics. IEEE, 2012, 3: 1175-1181.
- [36] Kang Y, Fang Y, Lai X. Automatic Detection of Diabetic Retinopathy with Statistical Method and Bayesian Classifier[J]. Journal of Medical Imaging and Health Informatics, 2020, 10(5): 1225-1233.
- [37] Jiang Y, Tan N, Peng T, et al. Retinal vessels segmentation based on dilated multi-scale convolutional neural network[J]. IEEE Access, 2019, 7: 76342-76352.
- [38] 叶显一. 基于深度学习的糖尿病视网膜病变图像中渗出物的检测方法研究[D]. 武汉: 武汉工程大学, 2018: 21-46.
- [39] Guo S, Wang K, Kang H, et al. Bin loss for hard exudates segmentation in fundus images[J]. Neurocomputing, 2020, 392: 314-324.
- [40] Guo S, Li T, Wang K, et al. A Lightweight Neural Network for Hard Exudate Segmentation of Fundus Image[C]. International conference on artificial neural networks, 2019: 189-199.
- [41] 刘梦颖. 基于深度学习的糖尿病视网膜图像分析算法研究[D]. 哈尔滨: 哈尔滨理工大学, 2020: 13-36.
- [42] Costa P, Galdran A, Smailagic A, et al. A weakly-supervised framework for interpretable diabetic retinopathy detection on retinal images[J]. IEEE Access, 2018, 6: 18747-18758.
- [43] Wang H, Yuan G, Zhao x, et al. Hard exudate detection based on deep model learned information and multi-feature joint representation for diabetic retinopathy screening[J]. Computer methods and programs in biomedicine, 2020, 191: 105398.
- [44] 朱均安. 基于深度学习的视觉目标跟踪算法研究[D]. 北京: 中国科学院大学, 2020: 12-56.
- [45] Rosenblatt F. The perceptron: a probabilistic model for information storage and organization in the brain[J]. Psychological review, 1958, 65(6): 386.
- [46] D. H. Hubel, and T. N. Wiesel. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex[J]. Journal of Physiology, 1962, 160(1): 106-154.
- [47] Werbos P. Beyond regression: New tools for prediction and analysis in the behavioral sciences[D]. Harvard University, 1974, 128(4): 34-98.
- [48] Fukushima K. Neocognitron: a self organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. Biol Cybern, 1980, 36(4): 193-202.
- [49] LeCun Y, Boser B, Denker J S, et al. Backpropagation applied to handwritten zip code recognition[J]. Neural computation, 1989, 1(4): 541-551.
- [50] Hinton G E, Osindero S, The Y W. A fast learning algorithm for deep belief nets[J]. Neural Computation, 2006, 18(7): 1527-1554.
- [51] Xu K, Ba J, Kiros R, et al. Show, attend and tell: Neural image caption generation with visual attention[C]. International conference on machine learning, 2015: 2048-2057.
- [52] Rush A M, Chopra S, Weston J. A neural attention model for abstractive sentence summarization[J]. arXiv preprint arXiv: 1509.00685, 2015: 45-78.
- [53] 蒋留兵, 张点, 潘波, 等. 基于生成对抗网络的多聚焦图像融合[J]. 计算机辅助设计与图形学报, 2021, 33(11): 1715-1725.

- [54] Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate[J]. Computer Science, 2014: 34-49.
- [55] Hu J, Shen L, Sun G, et al. Squeeze-and-excitation networks[C]. Proceedings of the IEEE conference on computer vision and pattern recognition, 2018: 7132-7141.
- [56] Wang J, Lv P, Wang H, et al. SAR-U-Net:squeeze-and-excitation block and atrous spatial pyramid pooling based residual U-Net for automatic liver CT segmentation[J]. Computer Methods and Programs in Biomedicine, 2021, 208: 106268.
- [57] Woo S, Park J, Lee J Y, et al. Cbam: Convolutional Block Attention Module[C]. Proceedings of the European conference on computer vision (ECCV). 2018: 3-19.
- [58] Zhao B, Wu X, Feng J, et al. Diversified visual attention networks for fine-grained object classification[J]. IEEE Transactions on Multimedia, 2017, 19(6): 1245-1256.
- [59] Al-Bdour G, Al-Qurran R, Al-Ayyoub M, et al. Benchmarking open source deep learning frameworks[J]. International Journal of Electrical and Computer Engineering, 2020, 10(5): 5479.
- [60] Jia Y, Shelhamer E, Donahue J, et al. Caffe: Convolutional architecture for fast feature embedding[C]. Proceedings of the 22nd ACM international conference on Multimedia, 2014: 675-678.
- [61] 费宁, 张浩然. TensorFlow 架构与实现机制的研究[J]. 计算机技术与发展. 2019(09): 34-89.
- [62] Manaswi N K. Understanding and working with Keras[M]. Deep Learning with Applications Using Python. Apress, Berkeley, CA, 2018: 31-43.
- [63] Ketkar N, Moolayil J. Introduction to pytorch[M]. Deep learning with python. Apress, Berkeley, CA, 2021: 27-91.
- [64] 张德彪. 基于深度学习的糖尿病视网膜病变分类和病变检测方法的研究[D]. 哈尔滨: 哈尔滨工业大学, 2017: 25-31.
- [65] 李轩屹. 基于卷积神经网络的糖尿病视网膜病变分类方法研究[D]. 太原: 太原理工大学, 2020: 15-31.
- [66] 张子振. 基于深度学习的糖尿病视网膜图像的分析与研究[D]. 保定: 河北大学, 2020: 7-56.
- [67] Pizer S M, Amburn E P, Austin J D, et al. Adaptive histogram equalization and its variations[J]. Computer Vision Graphics & Image Processing, 1987, 39(3): 355-368.
- [68] Zuiderveld K. Contrast limited adaptive histogram equalization[M]. Academic Press Professional, 1994: 34-109.
- [69] Stark J A. Adaptive image contrast enhancement using generalizations of histogram equalization[J]. IEEE Transactions on Image Processing, 2000, 9(5): 889-896.
- [70] Akkasi A, Varoglu E, Dimililer N. Balanced undersampling: a novel sentence-based undersampling method to improve recognition of named entitiesin chemical and biomedical text[J]. Applied Intelligence, 2018, 48(8): 1965-1978.
- [71] Cartus A R, Bodnar L M, Naimi A I. The impact of under sampling on the predictive performance of logistic regression and machine learning algorithms: A simulation study[J]. Epidemiology, 2020, 31(5): 42-44.
- [72] Zhu Z, Wang Z, Li D, et al. Tree-based space partition and merging ensemble learning framework for

- imbalanced problems[J]. Information Sciences, 2019, 503: 1-22.
- [73] Wang Z, Wang B, Cheng Y, et al. Cost-sensitive fuzzy multiple kernel learning for imbalanced problem[J]. Neurocomputing, 2019, 366: 178-193.
- [74] 冯宇翔. 基于改进 PSPNet 的语义分割方法研究[D]. 鞍山: 辽宁科技大学, 2021: 12-48.
- [75] 梁佳雯. 基于轻量级网络的遥感图像实时语义分割算法研究[D]. 成都: 电子科技大学, 2021: 26-47.
- [76] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]. Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.
- [77] Tan M, Le Q. Efficientnet: Rethinking model scaling for convolutional neural networks[C]. International conference on machine learning. PMLR, 2019: 6105-6114.
- [78] 张子振, 刘明, 朱德江. 融合注意力机制和高效网络的糖尿病视网膜病变识别与分类[J]. 中国图象图形学报, 2020, 25(8): 1708-1718.
- [79] 李郭. 基于深度学习的糖尿病视网膜眼底图像病变分析算法研究[D]. 上海: 上海交通大学, 2018: 21-78.
- [80] 徐宏. 基于眼底图像的糖尿病视网膜病变智能诊断[D]. 成都: 电子科技大学, 2019: 12-46.
- [81] 白杰, 张赛, 李艳萍. 基于改进交错组卷积的眼底硬性渗出物自动分割[J/OL]. 计算机工程, 2022, 1-10[2022-04-09].
- [82] Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation[C]. International Conference on Medical image Computing and Computer-Assisted Intervention. Springer international publishing, 2015, 9351: 234-241.
- [83] 张忠祥. 基于深度学习的视网膜病变硬性渗出物分割[D]. 济南: 山东师范大学, 2020: 12-56.
- [84] 倪启业. 基于特征联合深度学习的彩色图像拼接检测研究[D]. 南京: 南京信息工程大学, 2020: 21-56.
- [85] Mateen M, Wen J, Nasrullah N, et al. Exudate Detection for Diabetic Retinopathy Using Pretrained Convolutional Neural Networks[J]. Complexity, 2020, 13(6): 547.
- [86] 陈志远, 金炜, 李纲. 注意力残差链式融合的彩色眼底图像硬性渗出检测[J]. 光电子·激光, 2020, 31(10): 1044-1053.
- [87] 连超铭, 钟舜聪, 张添福, 等. 光学相干断层扫描视网膜图像的迁移学习分类[J]. 激光与光电子学进展, 2021, 58(1): 262-268.
- [88] 孟凡奎, 银温社, 贺建峰. 基于深度学习的眼底图像出血点检测方法[J]. 山东大学学报(理学版), 2020, 55(9): 62-71.
- [89] 鲁品肃. 基于目标检测的糖尿病视网膜病变多模型检测[D]. 成都: 电子科技大学, 2020: 23-38.
- [90] 谢云霞, 黄海于, 胡建斌. 基于深度卷积神经网络的糖尿病视网膜病变分期及病灶检测[J]. 计算机应用, 2020, 40(8): 2460-2464.

攻读学位期间的成果及发表的学术论文

一、个人简历

李贺军，男，出生于1995年2月6日，于2018年6月在合肥学院获得工学学士学位。

二、在攻读学位期间参加的研究项目

1、项目来源：国家自然科学基金青年科学基金项目

项目名称：《基于表情与眼动双模态协同计算的情感状态智能分类方法研究》

项目编号：61906051

项目时间：2020.1-2022.12

2、广西自然科学基金资助项目（编号：2018GXNSFAA281235）

3、广西中青年教师基础能力提升项目（编号：2018KY0248，2020KY06026）

三、在攻读学位期间的获奖情况

（1）2020年获得桂林理工大学研究生学业奖学金三等奖

（2）2020年通过大学英语六级考试

（3）2020年获得广西人工智能大赛三等奖

（4）2020年获得全国大学生英语作文大赛国家级二等奖

（5）2021年获得桂林理工大学研究生学业奖学金三等奖

四、在攻读学位期间的发表论文

- [1] 程小辉，李贺军，邓昀，陶小梅，黎辛晓. 基于ME-ANet模型的糖尿病视网膜病变分级[J]. 广西科学（第二作者）（中文核心，已录用）
- [2] 程小辉，李贺军，邓昀，陶小梅，黎辛晓. 基于改进U-Net的视网膜渗出物分割方法[J]. 计算机工程与设计（第二作者）（中文核心，已录用）
- [3] 程小辉，李贺军，邓昀，陶小梅，黎辛晓. 基于改进U-Net的糖尿病视网膜图像血管分割研究[J]. 桂林理工大学学报（第二作者）（在外审阶段中）

五、在攻读学位期间的专利申请

- [1] 一种基于改进U-Net网络的视网膜图像血管分割方法，申请编号：202210477599.2
- [2] 一种基于嵌入残差卷积和注意力模块的U型糖尿病视网膜渗出物分割算法，申请编号：202210477600.1

致 谢

行文至此，落笔为终，思绪万千。求学二十余载，已近而立之年，今求学生涯即将落幕，心中更是五味杂陈，不免有一丝丝的惶恐，但更多的是喜悦。从2019年仲秋，到2022年盛夏，硕士三年时光，一晃而过，当时参加复试的场景还历历在目。如今，已然到了说再见的时间，再见这里的人们，告别这里的山水。此刻，在群山环抱的雁山校园中，伴随着鸟语花香，致谢那些在求学路上给予我帮助和支持的最可爱的人。

桃李不言，下自成蹊。首先，十分感谢我的导师程小辉教授，在过去的三年中，无论是在学习上还是生活中，都给予了我极大的鼓励与帮助，使我有信心和勇气面对学习生活中遇到的各种困难。在学习中，程教授博学多识，治学严谨，对我的课题研究和论文撰写工作给予了密切关注，并在这期间提供了许多专业性的建议与指导。从最开始的论文选题到如今最终定稿的完成，程教授付出了大量的时间和精力，从内容到格式，不停地审阅修改并提出宝贵意见。在生活中，程教授幽默风趣，为人随和，让我深刻感受到了如家人般的温暖与关怀，也让我明白了一些待人接物、为人处世的态度和原则。其次，我要感谢谢晓兰老师、董明刚老师和答辩组的老师们，感谢他们在论文开题和预答辩期间对我论文的内容和结构提出了许多宝贵的意见，让我能够更加合理从容地学习研究相关课题。

羊有跪乳之恩，鸦有反哺之义。感谢父母对我二十多年的养育之恩，感谢家人二十余年的教养和陪伴，感谢你们不求回报的默默付出，才让我在学习这条道路上走到了现在，你们是我勇敢奔赴远方的坚强后盾。生养之恩，此生无以回报，希望今后岁月安好，让你们能够在我的陪伴下健康幸福地生活。

丈夫非无泪，不洒离别间。感谢我的师兄师姐：童辉辉、赵君、刘鹏飞、牛童、冯险锋，感谢他们对我学习和生活上的帮助；感谢同门：张皓然、邵平、刘天承、汪跃，三年来我们相互激励，共同努力。感谢师弟师妹：罗源敏、丁黄婧、许元号、李钰、吴喜凯，带给我喜悦和力量。另外，感谢我的博士师姐：张宇师姐，给予我学术上的指导和鼓励。

执子之手，与子偕老。感谢我的女朋友：万贺迎。三年异地生活已近尾声，感谢她这三年对我的理解和包容，感谢她的陪伴与支持。

最后，感谢所有给予我帮助的可可爱的人们，愿我们永远心中有梦，眼里有光。