

分类号 _____

编号 _____

U D C _____

密级 _____



南方科技大学
SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

本科生毕业设计（论文）

题 目： 基于多任务的糖尿病视网膜病变

识别算法

姓 名： 陈晓珊

学 号： 11910233

系 别： 计算机科学与工程系

专 业： 计算机科学与技术

指导教师： 刘江 教授

2023 年 5 月 5 日

CLC _____

Number _____

UDC _____

Available for reference ☐Yes ☐No



SUSTech Southern University
of Science and
Technology

Undergraduate Thesis

Thesis Title: Multi-Task Learning for Diabetic Retinopathy
Classification and Lesion Segmentation

Student Name: Xiaoshan Chen

Student ID: 11910233

Department: Department of Computer Science
and Engineering

Program: Computer Science and Technology

Thesis Advisor: Professor Jiang Liu

Date: May 5, 2023

诚信承诺书

1. 本人郑重承诺所呈交的毕业设计（论文），是在导师的指导下，独立进行研究工作所取得的成果，所有数据、图片资料均真实可靠。

2. 除文中已经注明引用的内容外，本论文不包含任何其他人或集体已经发表或撰写过的作品或成果。对本论文的研究作出重要贡献的个人和集体，均已在文中以明确的方式标明。

3. 本人承诺在毕业论文（设计）选题和研究内容过程中没有抄袭他人研究成果和伪造相关数据等行为。

4. 在毕业论文（设计）中对侵犯任何方面知识产权的行为，由本人承担相应的法律责任。

作者签名:

_____年____月____日

COMMITMENT OF HONESTY

1. I solemnly promise that the paper presented comes from my independent research work under my supervisor's supervision. All statistics and images are real and reliable.
2. Except for the annotated reference, the paper contents no other published work or achievement by person or group. All people making important contributions to the study of the paper have been indicated clearly in the paper.
3. I promise that I did not plagiarize other people's research achievement or forge related data in the process of designing topic and research content.
4. If there is violation of any intellectual property right, I will take legal responsibility myself.

Signature:

Date:

基于多任务的糖尿病视网膜病变 识别算法

陈晓珊

(计算机科学与工程系 指导教师: 刘江)

[摘要]: 糖尿病视网膜病变 (Diabetic retinopathy, DR) 是导致劳动年龄人口失明的主要疾病之一, 快速且准确的诊断需要医生具备足够的经验, 而一些基层地区缺乏这样的医生, 使得一些患者因无法得到及时的诊断而造成无法挽回的失明, 因此开发一个糖尿病视网膜病变自动化诊断算法实现眼底图像病灶的分割以及疾病的诊断, 以减少筛查时间、成本和工作量, 将具有显著的潜在效益。目前已有大量工作对于 DR 分类任务以及病灶分割任务进行了研究, 但是基于单任务的训练, 使得网络难以从小样本中充分挖掘图像中的信息。因此本文探究如何设计合理的多任务模型, 以利用多任务学习的优势, 提升分割与分类任务的效果并缓解过拟合的问题。本文提出的模型利用 VGG16 网络进行 DR 级别的诊断, 利用 U-Net VGG16 实现病灶分割, 两个网络之间通过共享特征提取网络建立任务关联性, 并设计了两个基于注意力机制的特征融合模块, 实现两个任务特征的传递与融合, 以充分利用两个任务所提取的病灶特征。本文在 IDRI 数据集上进行了一系列消融与对比实验, 在 DDR 以及 EyePACS 数据集上进行了模型泛化性检验实验, 结果表明, 本文提出的多任务模型, 可以提升分类以及分割网络的效果, 能够在小样本训练的情况挖掘图像信息, 提升模型泛化能力。

[关键词]: 糖尿病视网膜病变; 眼底病灶分割; 多任务学习; 注意力机制

[ABSTRACT]: Diabetic retinopathy (DR) is one of the leading diseases causing blindness in the working age population. However, the diagnosis process is time consuming and labour intensive for doctors. Therefore, developing an automated diagnostic algorithm for DR diagnosis and lesion segmentation to reduce the time, cost and effort of screening would be significant. A large amount of work has been done on the DR classification task and the lesion segmentation task, but training based on a single task makes it difficult for the network to fully exploit the information in the images from small samples. This paper therefore investigates how to design a reasonable multi-task model to take advantage of multi-task learning to improve the effectiveness of segmentation and classification tasks and to alleviate the overfitting problem. The model proposed in this paper utilises the VGG16 network for DR diagnosis and the U-Net VGG16 for lesion segmentation. A shared feature extraction network is designed to establish task correlation, and two feature fusion modules based on attention mechanism is designed to enable the transfer and fusion of features from the two tasks to make full use of the lesion features extracted by both tasks. A series of ablation and comparison experiments were conducted on the IDRiD dataset, and model generalisability testing experiments were carried out on the DDR and Eye-PACS datasets. The results show that the multi-task model proposed in this paper, which can improve the effectiveness of classification as well as segmentation networks, can mine image information and improve model generalisation with small sample training.

[Key words]: Diabetic retinopathy; Fundus image lesion segmentation; Multitasking learning; Attention mechanisms

目录

1. 绪论	1
1.1 研究背景	1
1.2 研究内容	2
1.3 全文结构	3
2. 国内外研究现状	4
2.1 DR 级别分类网络	4
2.2 眼底多病灶分割网络	4
2.3 分割与分类多任务网络	5
3. DR 级别分类与病灶分割多任务网络	7
3.1 多任务模型架构	7
3.1.1 特征提取网络	7
3.1.2 基于 VGG16 的 DR 级别分类网络	8
3.1.3 基于 U-Net VGG16 的眼底多病灶分割网络	8
3.2 特征融合模块	8
3.2.1 深层特征图融合模块	9
3.2.2 浅层特征图融合模块	11
3.3 多任务网络损失函数	11
4. 实验结果与分析	13
4.1 实验数据集	13
4.1.1 DR 级别分类数据集	13
4.1.2 眼底多病灶分割数据集	13

4.1.3 数据预处理与数据增强	14
4.2 实验评价标准	14
4.2.1 分类评价标准	14
4.2.2 语义分割评价标准	15
4.3 实现细节	15
4.4 消融实验	16
4.4.1 共同特征提取网络	16
4.4.2 病灶特征融合模块	17
4.4.3 各成分消融实验结果	20
4.5 对比实验	21
4.5.1 DR 级别分类任务的实验结果	21
4.5.2 眼底多病灶分割任务的实验结果	22
4.6 模型泛化能力检验实验	23
5. 结论	25
参考文献	26
致谢	29

1. 绪论

1.1 研究背景

糖尿病视网膜病变 (Diabetic Retinopathy, DR) 是劳动年龄人口失明的主要原因之一^[1], 该疾病在发展的不同阶段会在视网膜出现不同的病灶, 如图1所示, 该图展示了眼底微血管瘤、出血、软性渗出以及硬性渗出的局部放大图, 医生可以根据眼底彩照中显示的症状对病情的严重程度做出诊断, 从而给出相应的治疗方案。按照国际临床糖尿病黄斑水肿严重程度分级标准^[2], 糖尿病视网膜病变的严重程度可以分为五个级别, 正常、轻度、中度、重度非增殖性和增殖性, 严重程度分级与眼底图像上病灶的表征有很强的相关性, 其等级划分以及出现的眼底病灶如表1所示^[3]。该病变的轻度等级的表现为出现微动脉瘤^[4], 为视网膜浅层上的小红点。微血管瘤容易破裂, 破裂后将导致不同大小和形状的出血。受损的视网膜血管接着使脂蛋白沉淀泄漏, 显现为黄色渗出, 属于病情发展后期出现的病灶^[5]。因此对于这些特定病灶的识别有助于确定视网膜病变的阶段, 以便进行诊断以及治疗。然而, 对眼科医生来说, 准确地对糖尿病视网膜病变进行分级, 尤其是初期阶段的诊断, 是一项耗时的工作, 此外, 一些基层眼科医疗存在力量不足的问题, 眼科服务供不应求, 患者无法得到及时的诊疗, 将容易造成不可挽回的视力损伤和失明。因此, 开发一个糖尿病视网膜病变自动化诊断算法, 以减少筛查时间、成本和工作量, 将具有显著的潜在效益。

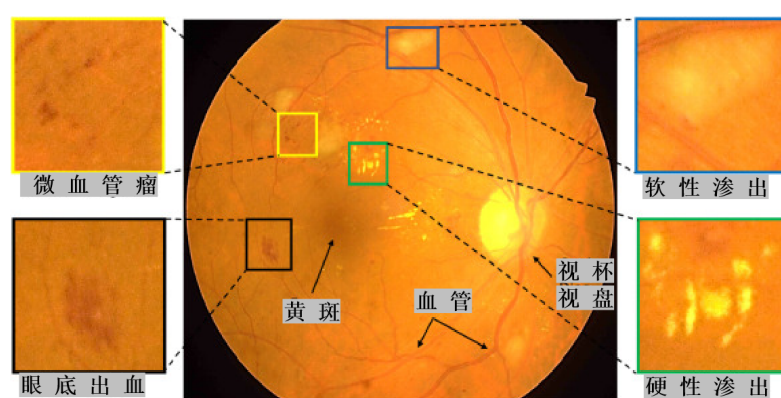


图 1 眼底彩照结构与病灶示意图¹

该算法的开发涉及两项重要的任务, DR 级别分类以及病灶分割, 分类任务能够实现对眼底图像的 DR 级别分类, 分割能够实现对图像像素级别的分类, 能够识

¹ 图片来源: <https://doi.org/10.1016/j.media.2019.101561>

表 1 DR 级别诊断依据

等级	症状
正常	眼底无任何异常
轻度非增殖性 DR	微血管瘤以及小的出血点
中度非增殖性 DR	硬性渗出以及出血斑
重度非增殖性 DR	软性渗出以及出血斑
增殖性 DR	新生血管

别病灶并判断所属的类别，以及勾画出病灶区域。实现可靠且精准的病灶分割算法与分类算法，将能够辅助医生进行诊断，提高诊断效率。目前已有大量工作对两个任务进行了研究，但依然面临以下挑战，在 DR 级别分类任务中，类别间差异较小，例如级别为 1 级的 DR 图像可能只包含少数微血管瘤^[6]，即使是专家也难以识别，导致其常常被模型错分为正常类，而在多病灶分割任务中，即使模型能够准确识别病灶，但由于病灶间相似性大，往往会判错病灶类型。此外，由于对图像进行像素级别的标注比较困难耗时，现有的标注数据较少，目前的分割方法难以达到较高的精度，同时也容易出现过拟合的问题。尽管一些方法针对上述问题提出了一些解决方式，但是在数据量少以及采用单任务进行训练的情况下，网络依然难以学习到眼底图像中的病灶的复杂特征，DR 场景下分类任务以及分割任务的关联性也被忽略。

引入多任务学习机制，则能够一定程度上缓解上述问题。多任务学习是一种推导迁移学习方法，在任务具备关联性的基础上，设计合理的多任务框架够提升各个子任务的精度。多任务学习使得模型在样本较少的情况下充分地挖掘图像所蕴含的信息，引导模型将注意力放在与任务相关的特征信息上，因此能帮助缓解数据标注稀疏而引发的模型过拟合问题，提高模型泛化能力。在所研究的 DR 疾病诊断场景下，眼底多病灶分割任务与 DR 级别分类任务具备一定的关联性。DR 的分级需要依据病灶信息，引入来自于分割网络的病灶信息可以为分类网络提供先验信息，对于分割任务，引入 DR 级别的信息，则有助于网络捕获病灶相关的语义信息，增强对于特定区域的注意力，提高分割网络识别病灶以及判断病灶类别的能力。

1.2 研究内容

本文主要研究如何构建 DR 级别分类任务以及眼底病灶分割任务的多任务学习网络，并设计一种有效的基于注意力机制的特征融合模块实现两个任务间病灶

特征的融合，以充分利用两个任务在病灶信息上的关联性以及多任务学习的优势，达到增强网络对于病灶信息的关注的效果，从而提升分割精度以及分级的准确性并增强模型的泛化性。网络设计上的主要研究工作如下，首先是探究特征提取网络的层数，通过共享特征，合理构建任务间的关联性，其次是在非共享部分，设计特征融合模块，以最大化利用两个任务所提取的病灶特征，来提升子任务的效果，该部分研究重点包括特征融合方式以及所选取的用于融合的特征图。在模型设计的过程中，在 IDRiD-D 以及 IDRiD-S 两个数据集上进行相应实验，以验证模型设计的合理性以及有效性，完成模型设计后，与现有方法进行比对，并在 DDR 以及 EyePACS 数据集上进行测试以检测多任务学下模型的泛化能力。

1.3 全文结构

本文分为五章，主要结构如下：第一章为绪论，主要介绍了本文的研究背景以及研究内容，首先讲述了糖尿病视网膜病变以及引入计算机辅助诊断系统的必要性，接着介绍了现有的自动化诊断算法存在的问题，接着描述了在 DR 级别诊断以及病灶分割上引入多任务学习的合理性与优势，最后是本文的研究内容。第二章为对所研究内容的国内外研究现状的概述，包括 DR 级别分类任务、眼底多病灶分割任务以及医学领域上分割与分类的多任务网络的现有研究成果。第三章为对本文提出的基于多任务学习的分割分类网络的介绍，首先对于多任务网络的整体框架进行了介绍，包括使用到的特征提取网络、分类网络以及分割网络，接着对于特征融合模块进行了介绍。第四章为对实验内容的介绍与结果分析，首先是描述实验使用的数据集，接着介绍了分割以及分类任务的实验评价标准，接着介绍了网络训练细节，最后是对消融实验、对比试验以及模型泛化性检验的实验结果分析。第五章总结了全文的内容，分析了模型存在的一些不足，并指出了下一步工作的研究方向。

2. 国内外研究现状

本章为对本文所涉及研究内容的研究现状进行介绍，依次介绍 DR 级别分类网络、眼底多病灶分割网络以及分割与分类的多任务网络，最后是对目前方法的概括与分析。

2.1 DR 级别分类网络

利用卷积神经网络，通过输入图像得到分级结果的端到端的深度学习算法被广泛研究以及应用。Shaohua Wan 等人采用 AlexNet^[7]、VggNet^[8]、GoogleNet^[9]、ResNet^[10]等模型进行 DR 分类，并分析这些模型对 DR 图像分类的效果^[11]，Zhixi Li 等人则使用了四种 Inception-v3^[12]网络来对 DR 进行分类^[13]。Khan 等人基于 VGG16 网络引入空间金字塔池化以及微网络来解决 DR 分级中的多尺度问题^[14]。Riaz 等人则采用深度以及密集连接网络进行分级，并验证了卷积层间的连接使得网络能够利用深度监督，学习重要的特征信息^[15]。

2.2 眼底多病灶分割网络

DR 的不同发展阶段会在视网膜上出现不同的病灶，病灶的大小、分布位置以及数量都会随着病情的变化而变化，因此对病灶的准确分割，能为病变的分级提供不可或缺的参考。眼底多病灶分割主要存在以下三个难点，分别是病灶特征复杂、病灶通常较小难以识别以及病灶间相似性大。现有方法针对以上三个难点提出了一些改进方式，为学习到病灶复杂的特征，在文献^[16]中，作者对网络五层特征图进行整合，并设计了 Scale-Aware Attention(SAA) 模块来动态加权每一层的关键信息。在文献^[17]中，作者提出了 Global Pyramid Guidance (GPG) 模块进行深浅层全局信息的整合，引入 Scale-Aware Pyramid Fusion (SAPF) 模块对多尺度信息进行选择性融合。PMCNet 的作者则是对邻近层的特征图进行信息整合，并采用 Dynamic Attention Block(DAB) 对提取的特征进行融合^[18]。文献^[19]中，作者则是利用通道注意力机制挖掘浅层特征图的细节信息以提取微小病灶特征，此外，引入基于空间注意力机制的自注意力模块进一步提取病灶特征，实现更为准确的病灶类别判断。文献^[20]中，作者在 U-Net 的四层与五层的跳跃连接部分引入空间以及通道双注意力机制捕捉更加丰富的病灶信息，有助于模型更准确判断病灶类别。

2.3 分割与分类多任务网络

考虑到分割以及分类任务的关联性以及多任务学习的优势, 研究人员探索了基于多任务学习的分割分类网络在不同医学图像的应用, 从多任务网络的架构以及任务间信息的传递与融合两个方面进行了模型设计。一些方法构建任务独立的双流网络, 两个任务之间没有权值共享, 并在特定位置引入了信息融合模块。在文献^[21]中, 作者采用双流结构并对分割任务得到的病灶分割结果进行全局平均池化, 与分类网络深层特征图进行融合。在文献^[22]中, 作者以 EfficientNet^[23]为基础网络构建胸腔 CT 图像的新冠诊断以及病灶分割的多任务网络, 在网络的每一层通过一个基于 Squeeze-and-Excitation(SE)^[24]的通道注意力模块实现特征图的传递与融合。然而, 两个任务在浅层所提取的特征是能够共享的, 前两篇文献中的多任务网络设计则忽略了两个任务特征的相似性, 一些方法因此引入了一个特征提取网络用于提取两个任务的共同特征, Alex 等人使用编码器-解码器网络进行 DR 分级和病灶分割, 创建了一个双流网络, 两个任务共享一个编码器, 其中 VGG-16 编码器输出 DR 分级, 解码器输出病灶分割^[25]。然而, 完全共享参数则未考虑两个任务在网络深层所提取特征的差异性, 因此一些方法只将分类网络的浅层作为共同特征提取模块, 并设计模块进行特征融合。Xiaofei Wang 等人设计了一个用于诊断新冠病毒的病灶分割分类联合深度学习模型 DeepSC-COVID, 该方法引入了一个跨任务特征自网络来联合提取三维特征, 用于后续的三维病灶分割和分类子网, 分割网络的输出与原图像进行通道叠加后输入到分类网络中^[26]。Jinkui Hao 等人提出了一种用于光学相干层析血管成像的联合分割、检测和分类模型, 该模型包括共享的特征提取器, 以及针对特定任务的控制模块用于学习如何进行特征选择和融合^[27]。以上方法都只在网络的浅层进行了特征融合, 一些方法则是在网络的深层进行特征融合, 如 Sheng Chen 等人采用两个任务共享一个特征提取模块的方式进行多任务学习, 并在网络深层设计了一个特征传递模块, 传递两个任务所提取的特征信息, 同时使用门控函数来控制信息的传输以确保传递信息的有效性^[28]。

基于以上总结与分析, 本文在多任务框架设计上, 引入一个特征提取网络以实现任务之间权值共享, 采用 VGG16 的前三层网络作为提取网络, 提取共同特征建立任务关联性, 同时保留任务之间的差异性。在融合模块的设计上, 本文对于现有方法在融合信息以及融合方式上进行了分析, 在融合信息的选取上, 现有方法

通常利用分割的输出结果^[21,29]或分类的类激活图^[30]，并在网络的浅层进行信息交互，原因在于浅层特征图包含更多的细节信息，能够更为充分地利用病灶相关信息，部分方法则在网络的深层设计了交互^[22,28]，从而能够利用到深层特征图所包含的更加丰富的语义信息，此外深层特征图往往分辨率较低，在深层进行信息交互不会产生较大的计算量。基于以上分析，本文同时引入了网络的深层与浅层特征图进行特征融合，以充分利用浅层特征所包含的细节信息以及深层特征图中丰富的语义信息。在特征融合方式上，由于在经过下采样进行特征提取的过程中，关于病灶的特征往往存在于特定通道上，因此现有方法往往采用通道叠或是引入通道注意力机制对不同特征图信息进行加权整合的方法，以进行信息补充^[21-22,30]，然而这种方法忽略了特征图的空间信息，此外，在网络的深层，病灶相关信息也只会出现在特定通道上的特定位置上，因此进一步引入空间注意力模块捕捉病灶特征是有必要的，基于此，本文在采用基于通道以及空间注意力的模块进行特征融合，以引导网络关注来自两个任务的重要特征。

3. DR 级别分类与病灶分割多任务网络

本文所提出的多任务网络包含以下四个部分，特征提取网络，分类网络，分割网络以及特征融合模块。本章首先描述多任务模型的架构，包括两个任务所共享的特征提取网络、分类网络以及分割网络，接着对于本文设计的两个特征融合模块进行详细介绍，最后对于模型的损失函数进行介绍。

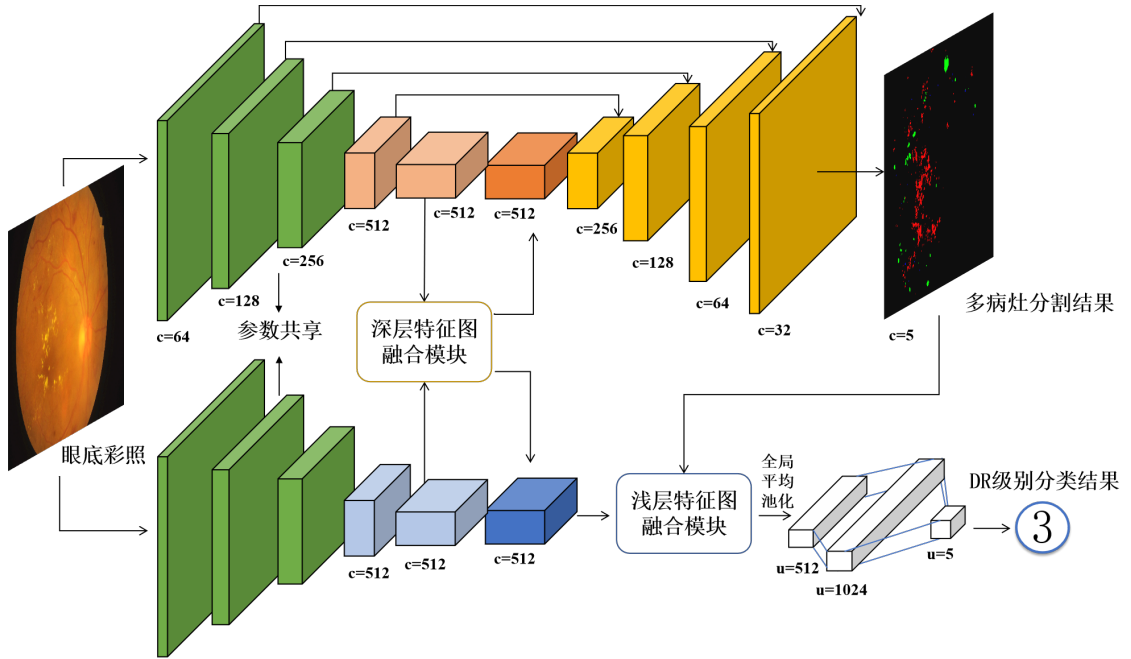


图2 病灶分割与 DR 级别分类多任务模型结构图，c 表示通道数，u 表示全连接层单元数

3.1 多任务模型架构

算法框架示意图如图2所示，首先使用特征提取网络提取两个任务的共享特征，所得到的特征经过分类网络得到 DR 级别诊断结果，经过分割网络得到多病灶分割图，而所提出的病灶特征融合模块则用于连接分割网络与分类网络，使两个任务所提取的特征可以进行传递以及融合，以提升子任务的性能。本节将依次描述使用的特征提取网络、分类网络以及分割网络。

3.1.1 特征提取网络

由于分割与分类的所提取的特征在浅层存在一定的相似性，而在深层则存在一定的差异性，基于此，项目探究一个合理的参数共享层数，提取两个任务的共同特征，在建立任务关联性的同时又保留一定的任务独立性。本文选择 VGG16 网络作为特征提取网络，VGG16 在一般图像分类和医学图像分级任务中都表现出优

异的性能，能够提取丰富的语义特征，此外，VGG16 在 ImageNet 数据集^[31]上预训练权重的可用性使模型能够利用迁移学习在图像分类和分割中实现成功的改进。在网络层数设置上，本文经过相应的实验，确定采用 VGG16 网络的前三层网络，即图2中的绿色模块，用于提取两个任务的共同特征。VGG16 网络由五个 VGG 模块、最大值池化以及全连接层组成，其中 VGG 模块包含 2 个 3x3 的卷积层，VGG 模块之间为池化层用于对特征图进行下采样，最后的全连接层则是基于全局特征输出多类别概率，判断图像所属的 DR 级别。

3.1.2 基于 VGG16 的 DR 级别分类网络

本文的分类网络为 VGG16 网络的剩余部分，来自特征提取网络的特征图，经过两个 VGG 模块进一步提取分类任务相关的特征，在本文多任务模型中，则会经过后续的深层特征融合模块以及浅层特征融合模块，进一步挖掘病灶的语义特征，之后经过全局平均池化与全连接层得到最终的分类结果。

3.1.3 基于 U-Net VGG16 的眼底多病灶分割网络

本文的分割网络由 VGG16 网络的最后两层卷积层以及 U-Net VGG16 网络的解码器部分组成，U-Net VGG16 网络中，编码器为 VGG16 网络的卷积层，解码器的部分与 U-Net 类似，同样采用了跳跃连接结构，融合来自编码器同层的特征图信息，不同的是每一层的输出特征通道数。在本文多任务模型中，编码器最后一层提取的特征将输入到深层特征融合模块，模块所输出的分割特征图则作为解码器的输入，得到分割结果。为充分利用分割结果中所包含的病灶的细节信息，该输出结果会作为浅层特征融合模块的输入，与分类特征图进行融合。

3.2 特征融合模块

考虑到网络的深层包含更加丰富的语义信息，而网络的浅层则包含更多的细节信息，为充分利用网络所提取这两类信息，本文针对深浅层的特征图设计了融合模块，以进行更加全面且有效的特征融合，并采用基于空间以及通道的双注意力机制设计特征融合模块以引导网络提取重要的病灶特征。因此，本文设计的特征融合模块包括两部分，一部分是分割网络深层特征图与分类网络深层特征图的融合模块，该模块引入了基于空间与通道的双注意力的模块，另一部分则是分割网

络浅层特征图与分类网络深层特征图的融合模块，由于浅层特征图的引入是为了引导网络关注更多的细节信息，因此该模块仅引入了基于空间注意力的模块。接下来将依次对两个模块的内部结构进行详细介绍。

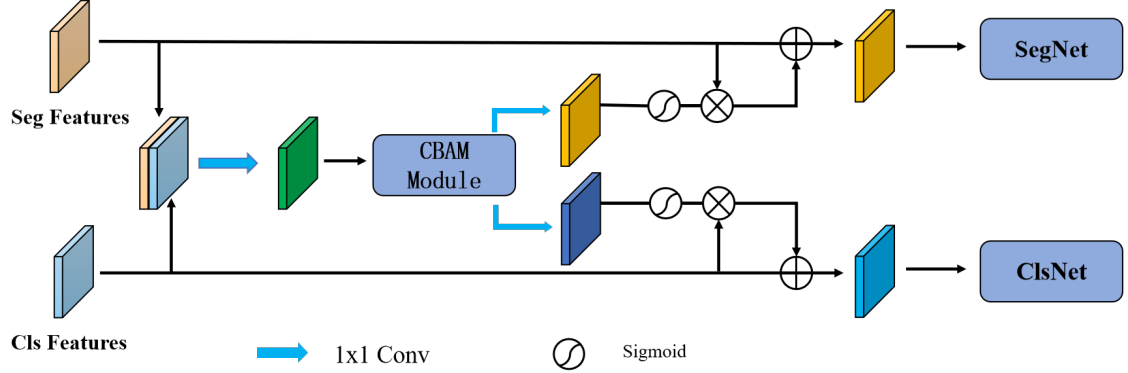


图3 深层特征图融合模块

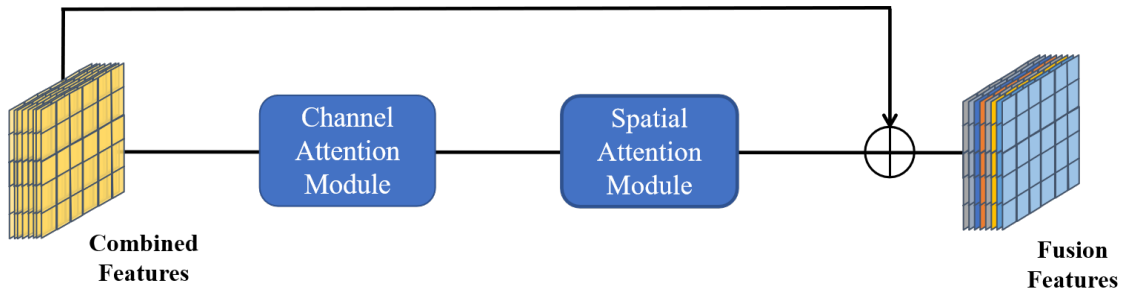


图4 CBAM 融合模块示意图

3.2.1 深层特征图融合模块

两个网络在深层所提取的抽象特征存在一定的差异性，设计该模块实现特征的融合以帮助两个任务提取语义更为丰富的特征。该模块的结构如图3所示，模块的输入为分割网络的第五层特征图（Seg₅）及分类网络的第五层特征图（Cls₅），模块内部首先对于两个任务的特征图进行通道上的叠加，并利用一个1x1的卷积整合通道信息，同时对通道数进行压缩，接着经过一个基于空间与通道的双注意力机制的Convolutional Block Attention Module（CBAM）模块^[32]进行特征融合，得到的融合特征经过1x1卷积分别映射到分割特征图（Seg₅）和分类特征图（Cls₅）所在的空间，得到的两个新的特征图经过Sigmoid激活函数后，与原先的输入特征图进行相乘得到加权的特征图，该特征图再与输入特征图进行加和，得到模块的最终输出特征图，输出的特征图将会作为后续网络的输入特征。其中使用到的

CBAM 模块的结构图如图4所示，该模块由通道注意力机制模块和空间注意力模块组成，输入的特征依次经过两个模块，并与输出特征进行加和最终得到两个任务的融合特征。两个模块的结构图如图6和图5所示。通道注意力模块首先使用全局平均池化以及全局最大值池化，整合图像的全局信息，并采用基于压缩与扩张思想的三层全连接层，用于学习网络需要关注的重要通道，对得到的全局特征进行加和后利用 Sigmoid 激活函数得到相应的权重值，对原特征图各个通道进行加权，从而得到激活特定通道的特征图。而空间注意力机制则是首先基于通道进行平均池化以及最大值池化得到两张通道数为 1 的特征图，对特征图进行通道叠加后，利用 7x7 的卷积提取重要的空间信息，引导网络关注特定位置的信息，接着采用 Sigmoid 激活函数得到各个图像位置的权重值，对原特征图各个位置进行加权，从而得到激活特定位置的特征图。两个模块的串联排布，使得 CBAM 模块能够序列化地在通道和空间两个维度上产生注意力特征图信息，帮助网络聚焦图像的重要特征，抑制不必要的区域响应。

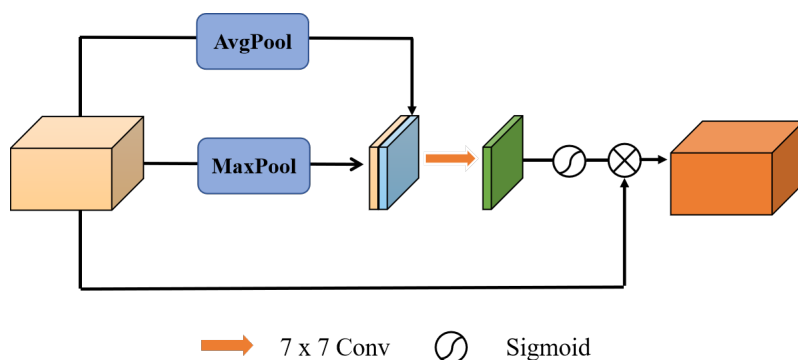


图 5 空间注意力模块示意图

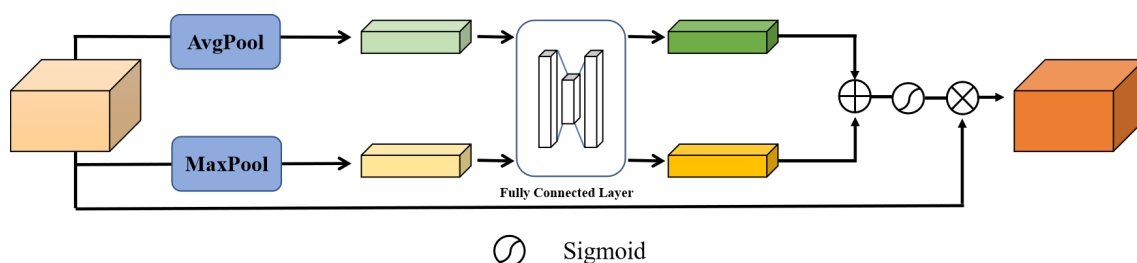


图 6 通道注意力模块示意图

3.2.2 浅层特征图融合模块

浅层特征图融合模块的引入该模块构造如图7所示，模块的输入特征为经过 SoftMax 操作的分割输出结果以及经过深层特征融合模块的分类特征图，分割浅层特征图首先经过一个基于卷积的下采样操作，得到与分类特征图同等大小的特征图，接着对两个特征图进行通道叠加，并采用一个 3x3 的卷积进行信息整合，整合后的特征经过一个空间注意力模块实现特征的融合与提取，接着使用 1x1 的卷积将输出特征映射到分类特征所在的空间上，并与输入的分类特征进行相乘，不同于采用 Sigmoid 函数对融合信息进行激活，该模块采用 1x1 的卷积动态学习激活特征，最后再与输入的分类特征图进行加和，得到融合了分割网络浅层特征的分类特征图，该特征图将作为分类网络后续全连接层的输入。

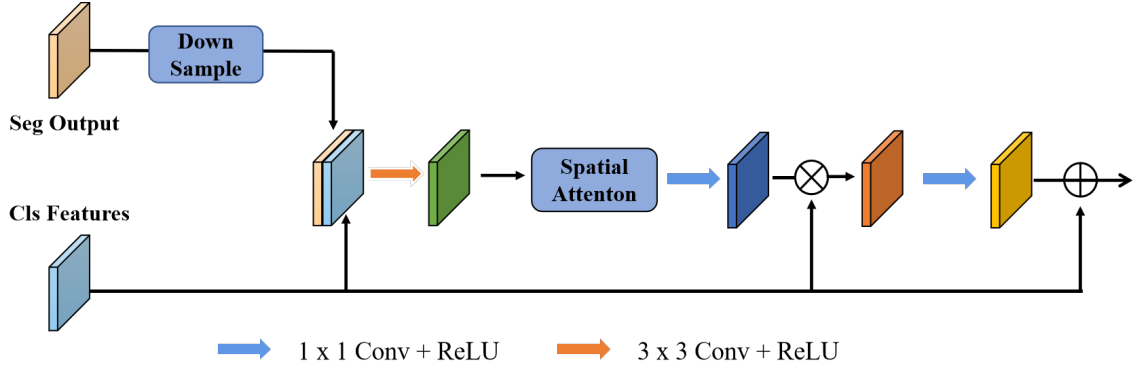


图 7 浅层特征图融合模块

3.3 多任务网络损失函数

对于分类网络，本文采用 Cross Entropy(CE), 即交叉熵损失，其计算方式如公式1所示。

$$\mathcal{L}_{CE} = \frac{1}{n} \sum_{i=0}^n y_i \log \hat{y} \quad (1)$$

其中 n 表示样本数， i 表示第 i 个样本， y 为第 i 个样本的类别标签， \hat{y} 则为分类网络的输出类别。

对于分割网络，由于眼底病灶在图像中的占比偏小，而背景区域较大，存在样本分布不平衡的问题，导致了类别间损失的不均衡，采用交叉熵损失容易导致模型忽略微小的病灶。为了缓解这个问题，本文分割损失采用 Class-balanced Cross Entropy(CBCE) 损失^[33]，其计算方式如公式2所示，该损失能够动态地依据病灶在

图像中的占比调整对应的损失权重，对于占比小的类别，如微血管瘤，增大其损失权重，对于占比较大的类别，如背景，则减小其损失权重。

$$L_{CBCE} = \frac{1}{C} \sum_{i=0}^C -\frac{1-\beta}{1-\beta^{n_i}} y_i \log(z_i) \quad (2)$$

其中， C 代表类别数目， β 是一个超参数，本文在模型训练过程中将其设置为 0.999， n_i 代表图像中类别为 i 的像素点的个数， y_i 为类别为 i 的真实标注， z 为分割输出结果经过 SoftMax 函数后的结果。

因此，多任务网络的整体损失为两个网络的损失之和。

$$L_{total} = L_{CE} + L_{CBCE} \quad (3)$$

4. 实验结果与分析

本章对于模型的实验内容进行介绍，包括实验所使用到的数据集、实验评价标准、实验的实现细节、消融实验结果、对比实验结果以及模型泛化性实验结果。

4.1 实验数据集

本文使用的数据集包括 DR 级别分类数据集和眼底多病灶分割数据集，分类数据集包括 IDRiD-D^[3]和 EyePACS^[34]，分割数据集包括 IDRiD-S^[3]和 DDR^[35]。针对多任务训练，在 IDRiD 数据集上进行网络预训练得到伪标签以构建同时具备分类以及分割标签的样本，并引入数据增强扩充样本。

4.1.1 DR 级别分类数据集

IDRiD-D: 该数据集依照国际临床糖尿病黄斑水肿严重程度分级标准^[2]，将 DR 疾病分为五级，依次是正常、轻度非增殖性 DR、中度非增殖性 DR、重度非增殖性 DR 以及增殖性 DR。该数据集包含 413 张训练图像与 103 张测试图像，实验中，训练阶段进一步将训练集按照 8:2 的比例划分为训练集以及验证集，划分时，在每一个类别上按照划分比例进行划分，以保证在训练集以及验证集上类别数目分布一致。

EyePACS: 该数据集由 35126 张训练图像和 53576 张测试图像组成，分级依据与 IDRiD-D 数据集相同，有五个 DR 类别。然而，该数据集中的图像是由不同类型的相机在不同的光照条件下拍摄得到的，且该数据集只有图像级别的分级标签，无法用于分割分类多任务的模型训练，因此本文只使用该数据集的测试集图像进行模型泛化能力检验。

4.1.2 眼底多病灶分割数据集

IDRiD-S: 该数据集包含 81 张具有 DR 症状的眼底彩照，其中 54 张用于训练，27 张用于测试。标注的 DR 病变包括硬性渗出、软性渗出、微血管瘤以及眼底出血四类病灶，实验中，训练阶段进一步将训练集按照 8:2 的比例划分为训练集以及验证集。

DDR: 该数据集包含 757 张具有 DR 症状的眼底彩照，数据集按照 5:2:3 的比例被划分为训练集、验证集与测试集。所含的标注病灶类别与 IDRiD-S 相同。本

文使用该数据集进行模型泛化能力测试。

4.1.3 数据预处理与数据增强

本文提出的多任务模型训练需要图像同时带有 DR 级别标签以及多病灶分割标签，而 IDRiD 数据集上缺乏同时带有两种标签的图像，因此实验采用预训练网络生成伪标签的方式来生成完整的数据集进行模型训练，具体而言，对于 IDRiD-S 数据集上的图像，采用在 IDRiD-D 数据集上预训练的 VGG16 网络生成对应的分类标签，对于 IDRiD-D 数据集上的图像，采用在 IDRiD-S 数据集上预训练的 U-Net VGG16 网络生成对应的分割标签，对于类别标签为 0 的图像，忽略分割网络输出的分割结果，设置像素值全为 0 的图像为其分割标签。实验中统一将图像缩放至 512x512 的大小，此外，所用数据集样本数量较少，模型容易在小数据集上过拟合，导致模型泛化能力较差，实验针对训练集使用实时数据增强算法进行了数据扩充，使用到的数据增强方法包括：随机水平翻转、随机垂直翻转、平移、随机缩放、随机旋转和随机对比度增强。

4.2 实验评价标准

4.2.1 分类评价标准

本文采用准确率 (Accuracy)、F1 值 (F1-score) 以及 ROC 曲线下面积 (ROC_AUC, area under the curve of receiver operating characteristic) 三种指标来评价分类模型的效果。其中准确率以及 F1 值的计算公式如公式4-5所示, 所给出得公式中的 TP (真阳) 代表正样本被分类正确的数量, FN (假阴) 代表正样本中被分类为负样本的数量, TN (真阴) 代表负样本中被分类正确的样本, FP (假阳) 表示负样本中被分类为正样本的数量。准确率为预测正确的概率, 是分类任务中最常用的指标, 但是当出现样本不均衡的情况时, 并不能合理反映模型的预测能力, 而 IDRiD-D 数据集上各类别的数目不一致, 因此, 同时引入 F1 值以及 ROC_AUC 作为分类评价标准, F1 值的计算采用 Macro-F1, 该标准分别计算每一类的 F1 值, 最后取平均值得到 Macro-F1。ROC 曲线是一种用于表示分类模型性能的图形工具。ROC_AUC 通过将真阳性率和假阳性率作为横纵坐标, 依据不同阈值绘制 ROC 曲线, 以描绘分类器在不同阈值下的性能, ROC_AUC 则是该 ROC 曲线下的面积, 该值越接近

1, 表示分类器性能越好。

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

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

$$Precision = \frac{TP}{TP + FP} \quad (6)$$

$$Recall = \frac{TP}{TP + FN} \quad (7)$$

4.2.2 语义分割评价标准

本文采用交并比 (Intersection over Union, IoU)、Dice 以及 PR_AUC(area under the curve of precision and recall) 三种指标来评价分割模型效果, 其中, IoU 与 Dice 的计算公式如公式8-9所示。IoU 与 Dice 都是基于集合进行计算的相似度量函数, 通常用于计算两个样本的相似度, 能够反映分割结果与真实标注之间的重叠程度, 重叠部分越多, 数值越接近于 1。此外, 在分割任务上本文使用 PR_AUC 而非 ROC_AUC 作为评价标准, 因为在分类任务上, 每一个类别的判别正确与否具备同等的重要性, 而分割任务则加关注模型对于分割目标 (正样本) 判别的准确性, 即图像的病灶区域而非背景区域的判别, 基于 Precision 和 Recall 衡量的 AUC 则能够反映模型定位分割正样本的能力。

$$IoU = \frac{TP}{TP + FN + FP} \quad (8)$$

$$Dice = \frac{2 \times TP}{2 \times TP + FN + FP} \quad (9)$$

4.3 实现细节

本文基于 Pytorch 框架进行网络的实现与实验, 所有的实验都在 NVIDIA RTX A6000 显卡上运行。对于分类任务, 网络在 IDRiD-D 数据集上进行训练, 训练轮数为 100, 批处理量为 4, 初始学习率设置为 0.001, 对于分割任务, 网络在 IDRiD-S 数据集上进行训练, 训练轮数为 300, 批处理量为 2, 初始学习率设置为 0.01。两个任务上均采用动量为 0.9、权值衰减为 0.0001 的小批量随机梯度下降方法对模型进行优化, 采用 Poly 学习速率调整策略^[36]在训练过程中动态设置学习率。

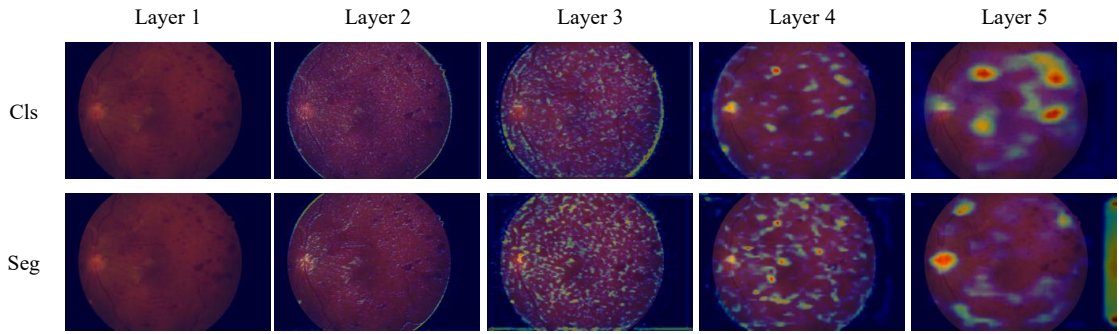


图 8 分类网络与分割网络各层类激活图，Cls 表示分类网络，Seg 表示分割网络

4.4 消融实验

该部分包括两部分的实验内容介绍，首先是本文提出的多任务模型设计的探索实验，包括共同特征提取网络的层数设置以及融合模块的设计，接着是网络组成成分的消融实验，以检验各个组成成分的有效性以及模型设计的合理性。

4.4.1 共同特征提取网络

本实验探究针对于分割以及分类任务的合理共享层数，以提取共同特征建立任务关联性，同时保留任务特征提取上的独立性。神经网络不同层所提取的特征不同，由浅层到深层，所提取的特征由具象到抽象，浅层较为关注边缘等信息，而深层则提取的是任务相关的特征，对于分割以及分类网络，在深层都会提取病灶相关的特征，但不同的任务，相同层上所提取的特征也会有所不同，对于分割以及分类网络，即使都关注病灶信息，但是同层上的病灶特征也会有所差异。为了分析两个任务在在各层所关注信息的相似性以及差异性，本文使用 GradCAM^[37]可视化技术对于两个任务在网络不同层的类激活图进行了可视化，实验所选取的 VGG16 以及 U-Net VGG16 网络在特征提取阶段均有五层，网络各层的类激活如图8所示，可以看出两个任务在前三层关注的特征信息类似，网络提取的为图像上的边缘信息，分割较之于分类对于边缘信息的关注度更高，而从第四层开始关注病灶区域，两个任务所关注的病灶区域也有所不同。

基于上述可视化结果的观察与分析，分别在分割任务以及分类任务上进行了层数共享实验，以探究最优的共同特征提取网络层数，从共享前三层参数开始，逐步增加两个任务之间的共享层数，表2与表3分别展示了不同共享层数的多任务模型在分类以及分割任务上的实验结果。可以看出，在两个任务上，共享前三层参

表 2 DR 级别分类任务上的层数实验结果，Layer 表示共同特征提取网络的层数，也表示分割与分类网络的共享层数

Layer	Acc(%)	F1-score(%)	AUC(%)
3	64.27±1.55	60.64±1.32	68.62±0.58
4	59.42±2.56	56.19±2.37	66.12±1.17
5	61.55±0.78	57.82±1.19	66.98±0.59

表 3 分割任务上的层数实验结果，Layer 表示共同特征提取网络的层数，也表示分割与分类网络的共享层数

Layer	mDice(%)	mIoU(%)	mAUC(%)
3	51.11±0.69	36.28±0.62	60.71±0.65
4	50.08±1.04	35.44±0.80	60.33±0.81
5	49.14±1.08	34.49±1.11	58.45±1.16

数在各项指标上都达到最高值，该结果也符合对于两个任务各层特征图可视化的分析结果，从网络第四层开始，两个任务所提取的特征图之间便出现了差异性，进行参数共享反而会降低两个任务的效果，基于该实验结果，本文将共同特征网络的层数设置为三，即分割与分类网络共用 VGG16 网络的前三层以提取共同特征。

4.4.2 病灶特征融合模块

从图8所示的类激活图中，可以看出分割与分类网络在第四层开始将注意力放置在相关病灶上，但是两个任务所关注的特征会有所不同，对于分类任务而言，网络的判别依据为特定的病灶的出现以及出现的数量，结合来自分割网络所提取的分类网络未注意到的相关病灶，能够为分类网络提供先验信息，提升分类网络的 DR 级别诊断能力，此外，由于分类网络对于一张图像往往会关注特定类别的病灶，从而做出类别判断，对于分割任务而言，引入来自分类网络所提取的特征则有助于增强其对病灶类别判断能力，减少病灶类别误判的产生。为充分利用两个任务所提取的病灶特征，以提升两个子任务的准确性，本文在确定共同特征提取网络的基础上，进一步提取了分割网络的深层特征图和浅层特征图与分类网络的深层特征图进行特征融合，且针对分割网络的两类特征图设计相应的融合模块以实现有效的特征融合。对于深层特征图的融合模块，从以下两个方面设计相应的实验探究有效的特征融合方式，一是特征融合模块的设计，在这一方面主要探究采用什么机制对两个任务的特征进行有效融合，二是融合特征图的选取，探究进行特

征融合的最优特征图。对于浅层特征图的融合模块，则是探究合适的融合机制。

深层特征融合模块设计：所利用的分割网络深层特征图与分类网络的深层特征图具有一定的相似性，都关注病灶区域，但是所关注的病灶区域有所不同，对于分类任务而言，分割任务所关注的病灶特征，并不是全部都有助于其进行类别判断的，例如，对于一张图像，分类网络关注在出血以及软性渗出两种病灶上，而分割任务会同时提取硬性渗出这一病灶的特征，进行融合的过程中，若不对分割网络所提取的不同病灶的特征加以区分，反而会影响分类网络的准确率，对于分割网络，也需要对于分类网络所提取的特征加以区分，因此融合模块的设计需要能够使得两个网络能够对图像特征所包含的重要信息进行动态选择，同时抑制无关信息。基于此，本文采用基于空间与通道双注意力机制的 CBAM 模块进行特征融合，也设置了实验对比基于其他注意力机制的模型以验证双注意力机制的有效性。实验选取了语义信息更为丰富的第五层特征图进行特征融合，比对的注意力机制包括，空间注意力^[38]、基于 Squeeze-and-Excitation (SE) 的通道注意力^[24]和自注意力^[38]，表6展示了在 DR 级别分类任务上不同注意力机制的实验结果，从表中结果可以看出，基于双注意力机制的 CBAM 模块优于单一的注意力机制，验证了基于双注意力机制的融合模块能够更加有效地融合两个任务所提取到的病灶特征，对于网络的深层特征图，重要信息分布在特定通道的部分区域上，因此需要结合通道以及空间注意力机制，引导网络关注特定特征的特定位置，从而利用病灶的重要特征。

表 4 DR 级别分类任务上不同注意力机制的融合模块实验结果

Fusion block	Acc(%)	F1-score(%)	AUC(%)
Spatial Attention	63.11±1.62	59.43±2.19	68.35±1.76
SE	62.78±2.10	59.21±2.41	68.24±1.49
Self Attention	64.08±2.10	61.54±2.41	69.35±1.49
CBAM	65.53±1.61	61.90±2.38	69.45±1.72

融合特征图选取：经过多次的下采样，网络的深层能够整合关注目标的全局以及局部特征，从而提取更加丰富的语义信息。所选取的用于特征融合的特征图对于实现两个任务的有效交互也有着一定影响。因此，本文设计实验在分类任务上，对于特征融合位置进行探索，比对了如图9展示的三处交互位置，在网络的第四层存在以下两处交互位置，分别是分割网络编码器的第四层特征图（EN_4）与

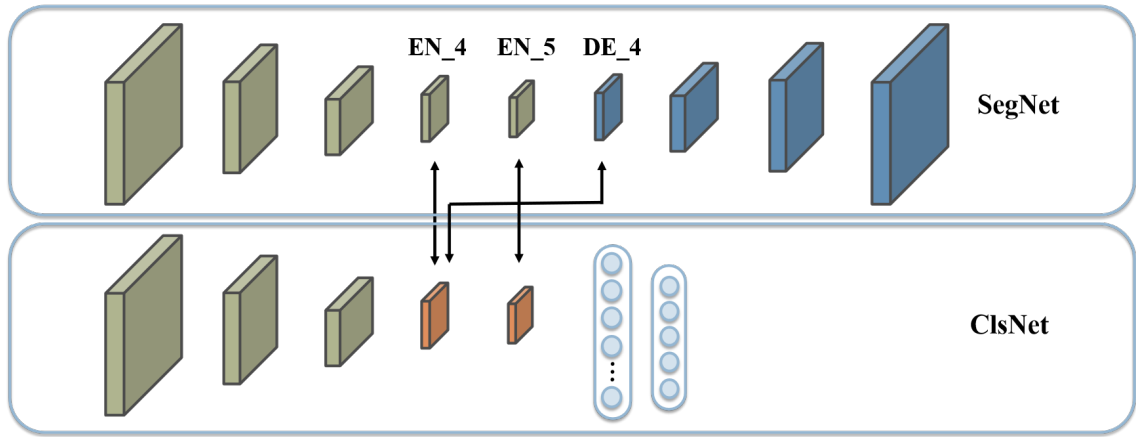


图 9 网络深层可交互位置

分类网络的第四层特征图进行交互，以及分割网络解码器第四层特征图（DE_4）与分类网络的第四层特征图进行交互，在网络的第五层则只有一处交互位置，即分割网络的第五层特征图（EN_5）与分类网络第五层特征图间的交互。实验结果如表5所示，可以看出选取网络的第五层特征图进行交互的模型，在各项指标上高于不加融合模块的模型，也高于其他两处的交互位置，表明了网络的深层能够学习到更具表征的病灶特征，适合用于进行特征融合，为两个任务的特征融合引入正确且重要的病灶信息。此外，在网络的第四层进行交互的模型，相较于不加融合模块的多任务网络，在数值上有所降低，由于第四层网络所提取的特征可能仍然存在一些干扰信息，且所提取的语义信息表征度要弱于第五层特征，基于 CBAM 的融合模块无法正确提取重要信息，从而干扰了网络的对于类别的判别。

表 5 DR 级别分类任务上不同融合特征比对实验结果

Seg features	Acc(%)	F1-score(%)	AUC(%)
-	64.27±1.55	60.64±1.32	68.62±0.58
EN_4	63.59±1.75	59.97±1.74	68.68±1.08
DE_4	63.35±0.42	59.67±0.27	68.18±0.28
EN_5	65.53±1.61	61.90±2.38	69.45±1.72

浅层特征融合模块: 尽管网络的深层特征图包含较为丰富的语义信息，但是网络的下采样过程会造成一些细节丢失，导致微小病灶的特征学习往往需要依赖浅层的特征图，因此，本文设计了浅层特征融合模块，以进一步引入分割网络的浅层特征以提高分类网络对于微小病灶的关注，由于需要引导网络关注特定区域，该模块采用基于空间注意力的方式对特征进行融合。在这部分设计了实验比对了基于空间注意力机制与自注意力机制的融合模块的效果，其比对结果如表6所示，可

以看出基于空间注意力机制的融合模块优于基于自注意力机制的融合模块，尽管自注意力机制能够整合全局信息，但是由于该方法缺乏偏置信息，需要较多的样本进行学习，而本文使用的数据集样本量较小，导致该模块无法得到充分学习，因此难以有效融合两个任务的特征。

表 6 注意力机制比较结果

Attention Module	Acc(%)	F1-score(%)	AUC(%)
Self Attention	64.56±1.61	61.67±1.72	69.24±0.89
Spatial Attention	67.18±2.16	64.76±2.08	71.24±1.25

4.4.3 各成分消融实验结果

本文模型包括共同特征提取网络、分割网络、分类网络、深层特征融合模块与浅层特征融合模块，在分类任务以及分割任务上都设置了消融实验，以检验各组成成分的有效性。在分类任务上，采用 VGG16 作为分类的基础模型，在分割任务上，则以 U-Net VGG16 作为基础模型，在两个任务上的消融实验结果如表7和表8所示，可以看出在引入共同特征提取网络，实现多任务学习的情况下，该模型在两个任务上均优于对应的基础模型，在各项指标上都有较大的提升，一方面说明了分割与分类任务之间存在一定的关联性，采用提取共同特征的方式，能够在网络特征学习上构建关联性，此外，也说明了在合理的多任务网络框架能够提升子任务的效果。此外，在分类任务上，进一步引入来自分割的浅层特征图的模型相比于仅仅在网络深层进行特征融合的模型，有较大的提升，说明了浅层特征图所包含的细节特征有助于分类模型增强对于病灶的关注。而在分割任务上，引入融合模块后，所带来的模型分割精度提升，相对较小，但是引入浅层特征图进行融合后，模型的稳定性有所提升。

表 7 DR 级别分类任务上多任务模型各成分消融实验结果, SegNet 表示分割网络, Deep Fusion 表示深层特征融合模块, Shallow Fusion 表示浅层特征融合模块

SegNet	Counterpart		Acc(%)	F1-score(%)	AUC(%)
	Deep Fusion	Shallow Fusion			
×	×	×	61.94±2.07	57.31±2.53	66.67±1.45
√	×	×	64.27±1.55	60.64±1.32	68.62±0.58
√	√	×	65.53±1.61	61.90±2.38	69.45±1.72
√	√	√	67.18±2.16	64.76±2.08	71.24±1.25

表 8 眼底多病灶分割任务上多任务模型各成分消融实验结果，ClsNet 表示分类网络，Deep Fusion 表示深层特征融合模块，Shallow Fusion 表示浅层特征融合模块

ClsNet	Counterpart		mDice(%)	mIoU(%)	mAUC(%)
	Deep Fusion	Shallow Fusion			
×	×	×	49.83±0.95	35.16±0.82	57.11±1.00
√	×	×	51.11±0.69	36.28±0.62	60.71±0.65
√	√	×	51.01±1.42	36.41±0.89	61.27±1.03
√	√	√	51.45±0.72	36.92±0.43	61.82±0.65

4.5 对比实验

为了评估本文所提出的多任务模型设计的有效性，在 IDRiD-S 以及 IDRiD-D 两个数据集上进行了与现有方法的对比实验。下文将分别介绍在分类任务以及分割任务的对比实验结果。

4.5.1 DR 级别分类任务的实验结果

在分类任务上，本文使用 VGG 16^[8]作为基础模型来构建多任务框架，将多模型网络与其他现有方法进行了比较，包括 4 种单任务下的用于 DR 级别分类的网络（VGG16^[8]，ResNet50^[10]，ResNet152^[10]和 EfficientNetB0^[23]）和 4 种分割与分类多任务网络（JCSNet^[22]，LADEN^[21]，MTUnet^[25]和 Chen 等人提出的网络^[28]），每一个网络都在相同的实验设置下进行训练以及测试。DR 级别分类网络对比结果如表9所示，可以观察到本文提出的算法在分类任务各项指标上都是取得了最高值，与单任务下的分类网络相比有较大的提升，说明了多任务模型的有效性。此外，JCS^[22]以及 LADEN^[21]这两种多任务网络的效果反而要低于一些单任务网络，不同于 MTUnet^[25]，Chen 等人提出的模型^[28]以及本文提出的算法，这两个网络并没有采用共同特征提取网络，而是采用独立的双流框架，分类以及分割任务之间并没有共享相关特征，JCS^[22]在分割与分类网络的每一层引入 SE 模块^[24]以实现任务间参数的共享，但是这种做法忽略了两个任务间在浅层的相似性，此外，在每一层引入融合模块导致模型参数量大，容易产生过拟合的问题，而 LADEN^[21]中在特征交互上，则只是将分割的输出结果进行全局平均池化，将该特征与分类网络特征进行通道叠加，这种方式一方面无法有效融合特征，另一方面也未利用到分割深层的特征。MTUnet^[25]以及 Chen 等人提出的模型^[28]与本文所提出的模型类似，都有参数共享层，但是 MTUnet 则是共享所有层的参数，忽略了深层网络特征的差异

性，Chen 等人提出的模型^[28]中所采用的门机制进行特征选择，在眼底彩照这一复杂的语义场景下是不适用的，而本文提出的基于注意力机制的融合模块则能够引导网络提取任务的重要特征，同时抑制无关的特征。

表 9 DR 级别分类网络对比实验结果

Model	Acc(%)	F1-score(%)	AUC(%)
VGG16	61.94±2.07	57.31±2.53	66.67±1.45
ResNet50	54.76±2.00	52.60±1.51	64.66±0.76
ResNet152	54.56±2.84	52.23±3.47	64.57±2.50
EfficientNetB0	60.97±4.32	58.63±4.51	68.42±2.77
JCSNet	58.45±2.33	54.40±2.02	64.92±1.19
LADEN	60.00±1.29	56.36±1.74	66.27±1.10
MTUnet	61.55±0.78	57.82±1.19	66.98±0.59
Chen et al.	62.33±1.43	58.81±2.00	67.63±1.19
Proposed	67.18±2.16	64.76±2.08	71.24±1.25

4.5.2 眼底多病灶分割任务的实验结果

本文在 IDRiD-S 数据集上进行眼底多病灶分割任务的实验，将本文所提出的算法与现有方法进行了比对，包括 4 个单任务下的医学影像通用分割网络（U-Net VGG16^[8]，U-Net^[39]，DeepLabv3+^[40]和 DenseUNet^[41]），三个针对眼底多病灶分割提出的网络（FCRN^[42]，CASENet^[43]和 PMCNet^[18]）和 4 个分割与分类多任务网络（JCS^[22]，LADEN^[21]，MTUnet^[25]和 Chen 等人提出的模型^[28]），其中 FCRN 以及 CASENet 网络的结果引用了文献^[44]中的结果，其余的网络都在相同的实验设置下进行训练以及测试。不同网络的实验结果如表9所示，可以观察到本文提出的算法在各项指标上都要高于单任务的分割网络，在多任务网络上，与 LADEN^[21]取得相似的值，在 mDice 以及 mIoU 两个指标上稳定性则更好。对比网络的可视化结果如图10所示，比对各个网络分割结果图下方两个标注区域，可以发现本文提出的网络对于病灶的类别判断的准确度高于其他网络，眼底病灶中，硬性渗出与软性渗出极为相似，多数网络容易将硬性渗出判断为软性渗出，而本文提出模型则能准确判别病灶类型。此外，相比于 LADEN 模型，本文提出的模型对于小病灶的识别能力较强，能够识别并定位到微小病灶，如图中所展示的微小出血以及微血管瘤，验证了浅层特征图融合模块的有效性。

表 10 眼底多病灶分割网络对比实验结果

Model	mDice(%)	mIoU(%)	mAUC(%)
U-Net VGG16	49.83±0.95	35.16±0.82	57.11±1.00
U-Net	48.44±0.98	33.51±0.94	58.09±1.63
DeepLabv3+	46.71±0.97	32.45±0.87	48.64±1.32
DenseUNet	40.24±0.87	26.79±0.56	49.10±1.17
FCRN	-	-	45.52
CASENet	-	-	48.23
PMCNet	46.33±2.44	32.37±1.75	55.83±2.28
MTUNet	49.14±1.08	34.49±1.11	58.45±1.16
JCSNet	48.43±2.35	34.79±1.52	59.16±1.11
LADEN	51.48±2.47	37.08±1.57	62.13±0.53
Chen et al.	47.55±2.21	33.48±1.36	58.72±0.74
Proposed	51.45±0.72	36.92±0.43	61.82±0.65

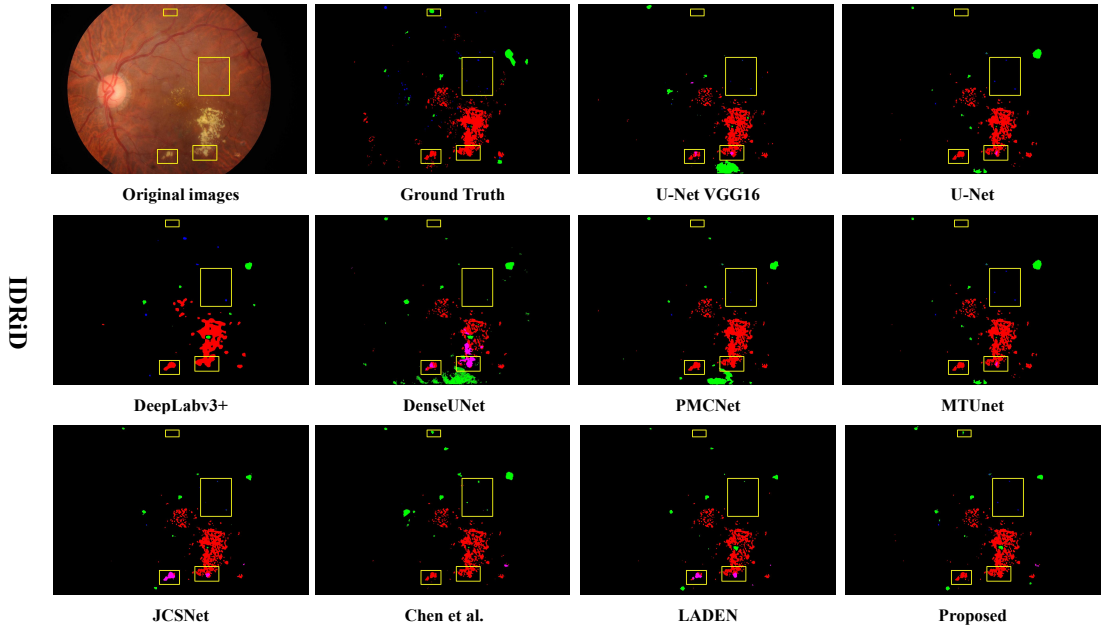


图 10 IDRiD-S 数据集上不同网络可视化结果，图像中，红色为硬性渗出，粉色为软性渗出，绿色为眼底出血，蓝色为微血管瘤

4.6 模型泛化能力检验实验

模型泛化性指模型经过训练后，应用到新数据并做出准确预测的能力。不同数据集在采集设备、采集环境等因素存在差异，这也导致图像在质量、亮度和对比度等方面存在差异性，模型能够在差异性存在的情况依然保持较高的预测能力，将使得模型在不同场景下具有通用性，这对于模型的临床应用具有重要意义。而本文所使用的多任务学习可以使得模型在样本较少的情况下充分地挖掘图像所蕴含的信息，引导模型将注意力放在与任务相关的特征信息上，从而缓解数据标注稀疏而引发的模型过拟合问题，提高模型泛化能力。为检验多任务模式下模型的泛化能力，本文引入额外数据集作为测试集，对模型的泛化性进行评估。对于分类

任务，本文使用 EyePACS 数据集作为额外的数据集来评估模型。对于眼底病灶分割任务，本文使用 DDR 数据集作为评估数据集。比对结果如表11和表12所示，可以看到，在分类任务上，本文所提出的多任务模型虽然在 Acc 指标上低于基础模型，但是在 F1 值以及 AUC 指标上均有提升，且模型稳定性增强。在分割任务上，多任务学习引入的优势更为突出，多任务模型相比于单任务的基础模型，在各项指标上均有提升，尤其是在 mAUC 指标上有大于 5% 的提升，且稳定性有所提升。模型训练所使用到分割任务的训练集仅仅包含 43 张图像，单任务模型容易出现过拟合的情况，引入多任务学习后模型的泛化能力有较为明显的提升，验证了多任务学习能够在小样本上充分挖掘重要信息。

表 11 DR 级别分类任务模型泛化能力测试比对结果

Dataset	Model	Acc(%)	F1-score(%)	AUC(%)
EyePACS	VGG 16	73.03±0.85	68.65±0.72	56.64±1.63
	Proposed	72.59±1.19	68.82±0.35	57.54±0.69

表 12 眼底多病灶分割任务模型泛化能力测试比对结果

Dataset	Model	mDice(%)	mIoU(%)	mAUC(%)
DDR	UNet VGG16	25.94±2.56	15.25±1.83	15.05±2.73
	Proposed	26.60±1.17	15.65±0.89	20.17±1.74

5. 结论

本文的主要内容是研究基于多任务学习的深度学习模型在 DR 级别分类和眼底病灶分割的应用。现有的方法通常基于单任务学习对网络进行改进，但由于眼底图像中带有标注的图像较少，基于单任务的学习难以挖掘重要的图像信息，模型容易出现过拟合的问题，导致模型泛化性较差，而基于多任务学习则能一定程度上缓解这个问题，由于眼科医生根据眼底彩照中出现的病灶对于 DR 级别进行诊断，因此两个任务具备高度的关联性，引入多任务学习对两个任务进行研究，在构建合理的多任务模型的情况下，则能够提升两个任务的效果并提升模型泛化性。本文对于现有的多任务网络进行了方法总结与问题分析，设计了适用于 DR 级别分类与病灶分割的多任务模型，模型包括一个特征提取网络用于建立任务关联性，提取的共同特征分别输入到基于 U-Net VGG16 的分割网络以及基于 VGG16 的分类网络，同时设计了两个基于注意力机制的特征融合模块，将分割网络的深浅层特征与分类网络的深层特征进行融合。本文在 IDRiD 数据集上进行了一系列的消融实验，验证了模型设计的合理性以及引入多任务学习和融合模块的有效性，对于分类任务而言，引入分割网络进行多任务学习，网络的分类准确率有较大的提升，对于分割任务而言，网络对于病灶的类别判断的准确率有所提高，同时对于小病灶的识别定位能力也有一定的提升。本文在 DDR 以及 EyePACS 数据集上的测试结果也表明了多任务学习能够提升模型的泛化性。

本文提出的多任务模型，虽然在 DR 级别分类任务与病灶分割任务上都取得了比较好的效果，但仍有可以优化和改进的空间，后续仍可以从以下几个方面对于模型进行研究与改进：

(1) 目前模型在分割任务上的提升相对较小，观察其分割结果发现，仍然存在较多类别误判的情况，可以进一步在分割网络中引入来自于分类网络的类别信息，如类激活图等，或是在深层进一步提取病灶特征，以提升网络的对于不同病灶的区分能力。

(2) 目前方法采用先使用预训练的网络得到完整标签，但是其可信度受限于模型精度，可以采用混合分类以及分割数据集的方式，对于无标签数据，采用一致性损失进行约束，以最大化利用已有数据，并调整相应的损失权重，找到两个任务联合训练下的最优损失函数设置。

参考文献

- [1] BARTH T, HELBIG H. Diabetische Retinopathie[J]. Augenheilkunde up2date, 2021, 11(03): 231-247.
- [2] WU L, FERNANDEZ-LOAIZA P, SAUMA J, et al. Classification of diabetic retinopathy and diabetic macular edema[J]. World journal of diabetes, 2013, 4(6): 290.
- [3] PORWAL P, PACHADE S, KOKARE M, et al. Idrid: Diabetic retinopathy–segmentation and grading challenge[J]. Medical image analysis, 2020, 59: 101561.
- [4] DAS D, BISWAS S K, BANDYOPADHYAY S. A critical review on diagnosis of diabetic retinopathy using machine learning and deep learning[J]. Multimedia Tools and Applications, 2022, 81(18): 25613-25655.
- [5] RÍOS H, RODRÍGUEZ F J, PERDOMO O J, et al. A deep learning model for classification of diabetic retinopathy in eye fundus images based on retinal lesion detection[C]//17th International Symposium on Medical Information Processing and Analysis: vol. 12088. 2021: 253-260.
- [6] LIN Z, GUO R, WANG Y, et al. A framework for identifying diabetic retinopathy based on anti-noise detection and attention-based fusion[C]//Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part II 11. 2018: 74-82.
- [7] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6): 84-90.
- [8] SIMONYAN K, ZISSERMAN A. Very Deep Convolutional Networks for Large-Scale Image Recognition[C/OL]//BENGIO Y, LECUN Y. 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings. 2015. <http://arxiv.org/abs/1409.1556>.
- [9] SZEGEDY C, LIU W, JIA Y, et al. Going deeper with convolutions[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 1-9.
- [10] 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.
- [11] WAN S, LIANG Y, ZHANG Y. Deep convolutional neural networks for diabetic retinopathy detection by image classification[J]. Computers & Electrical Engineering, 2018, 72: 274-282.
- [12] SZEGEDY C, VANHOUCKE V, IOFFE S, et al. Rethinking the inception architecture for computer vision[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 2818-2826.
- [13] LI Z, KEEL S, LIU C, et al. An automated grading system for detection of vision-threatening referable diabetic retinopathy on the basis of color fundus photographs[J]. Diabetes care, 2018, 41(12): 2509-2516.
- [14] KHAN Z, KHAN F G, KHAN A, et al. Diabetic retinopathy detection using VGG-NIN a deep learning architecture[J]. IEEE Access, 2021, 9: 61408-61416.
- [15] RIAZ H, PARK J, CHOI H, et al. Deep and densely connected networks for classification of diabetic retinopathy[J]. Diagnostics, 2020, 10(1): 24.

- [16] BO W, LI T, LIU X, et al. SAA: scale-aware attention block for multi-lesion segmentation of fundus images[C]//2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI). 2022: 1-5.
- [17] FENG S, ZHAO H, SHI F, et al. CPFNet: Context pyramid fusion network for medical image segmentation[J]. IEEE transactions on medical imaging, 2020, 39(10): 3008-3018.
- [18] HE A, WANG K, LI T, et al. Progressive Multiscale Consistent Network for Multiclass Fundus Lesion Segmentation[J]. IEEE transactions on medical imaging, 2022, 41(11): 3146-3157.
- [19] HUANG S, LI J, XIAO Y, et al. RTNet: relation transformer network for diabetic retinopathy multi-lesion segmentation[J]. IEEE Transactions on Medical Imaging, 2022, 41(6): 1596-1607.
- [20] WAN C, CHEN Y, LI H, et al. EAD-net: a novel lesion segmentation method in diabetic retinopathy using neural networks[J]. Disease Markers, 2021, 2021.
- [21] KAMBLE R, SRIVASTAVA A, SINGHAL N. LADEN: Lesion-Aware Adversarial Deep Network for Grading of Macular Diseases Using Color Fundus Images[C]//2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI). 2022: 1-4.
- [22] WU Y H, GAO S H, MEI J, et al. Jcs: An explainable covid-19 diagnosis system by joint classification and segmentation[J]. IEEE Transactions on Image Processing, 2021, 30: 3113-3126.
- [23] TAN M, LE Q. Efficientnet: Rethinking model scaling for convolutional neural networks[C]//International conference on machine learning. 2019: 6105-6114.
- [24] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 7132-7141.
- [25] FOO A, HSU W, LEE M L, et al. Multi-task learning for diabetic retinopathy grading and lesion segmentation[C]//Proceedings of the AAAI conference on artificial intelligence: vol. 34: 08. 2020: 13267-13272.
- [26] WANG X, JIANG L, LI L, et al. Joint learning of 3D lesion segmentation and classification for explainable COVID-19 diagnosis[J]. IEEE transactions on medical imaging, 2021, 40(9): 2463-2476.
- [27] HAO J, SHEN T, ZHU X, et al. Retinal structure detection in OCTA image via Voting-Based Multitask learning[J]. IEEE Transactions on Medical Imaging, 2022, 41(12): 3969-3980.
- [28] CHEN S, WANG Z, SHI J, et al. A multi-task framework with feature passing module for skin lesion classification and segmentation[C]//2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018). 2018: 1126-1129.
- [29] ZHOU Y, HE X, HUANG L, et al. Collaborative learning of semi-supervised segmentation and classification for medical images[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019: 2079-2088.
- [30] XIE Y, ZHANG J, XIA Y, et al. A mutual bootstrapping model for automated skin lesion segmentation and classification[J]. IEEE transactions on medical imaging, 2020, 39(7): 2482-2493.
- [31] DENG J, DONG W, SOCHER R, et al. Imagenet: A large-scale hierarchical image database [C]//2009 IEEE conference on computer vision and pattern recognition. 2009: 248-255.

- [32] 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.
- [33] CUI Y, JIA M, LIN T Y, et al. Class-balanced loss based on effective number of samples[C] //Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019: 9268-9277.
- [34] CUADROS J, SIM I. EyePACS: An Open Source Clinical Communication System For Eye Care[C/OL]//FIESCHI M, COIERA E W, LI Y (. Studies in Health Technology and Informatics: MEDINFO 2004 - Proceedings of the 11th World Congress on Medical Informatics, San Francisco, California, USA, September 7-11, 2004: vol. 107. IOS Press, 2004: 207-211. <https://doi.org/10.3233/978-1-60750-949-3-207>. DOI: 10.3233/978-1-60750-949-3-207.
- [35] LI T, GAO Y, WANG K, et al. Diagnostic assessment of deep learning algorithms for diabetic retinopathy screening[J]. Information Sciences, 2019, 501: 511-522.
- [36] LIU W, RABINOVICH A, BERG A C. ParseNet: Looking Wider to See Better[J/OL]. CoRR, 2015, abs/1506.04579. arXiv: 1506.04579. <http://arxiv.org/abs/1506.04579>.
- [37] SELVARAJU R R, COGSWELL M, DAS A, et al. Grad-cam: Visual explanations from deep networks via gradient-based localization[C]//Proceedings of the IEEE international conference on computer vision. 2017: 618-626.
- [38] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[J]. Advances in neural information processing systems, 2017, 30.
- [39] RONNEBERGER O, FISCHER P, BROX T. U-net: Convolutional networks for biomedical image segmentation[C]//Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18. 2015: 234-241.
- [40] CHEN L C, ZHU Y, PAPANDREOU G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation[C]//Proceedings of the European conference on computer vision (ECCV). 2018: 801-818.
- [41] CAO Y, LIU S, PENG Y, et al. DenseUNet: densely connected UNet for electron microscopy image segmentation[J]. IET Image Processing, 2020, 14(12): 2682-2689.
- [42] MO J, ZHANG L, FENG Y. Exudate-based diabetic macular edema recognition in retinal images using cascaded deep residual networks[J]. Neurocomputing, 2018, 290: 161-171.
- [43] YU Z, FENG C, LIU M Y, et al. Casenet: Deep category-aware semantic edge detection[C]// Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 5964-5973.
- [44] GUO S, LI T, KANG H, et al. L-Seg: An end-to-end unified framework for multi-lesion segmentation of fundus images[J]. Neurocomputing, 2019, 349: 52-63.

致谢

本文不仅体现了本人几年来的学习与研究成果，同时它也凝结了老师、同学、朋友的支持与帮助，在这里我向他们表示由衷的感谢。

首先要感谢我的导师刘江教授，刘老师专业知识渊博，治学态度严谨，老师所提出的科研的四个步骤也让我受益匪浅，我在本科的科研经历中也一直遵循这几个步骤。此外，老师对学生认真负责，关怀备至，当我对未来规划感到困惑时，老师耐心地与我交谈并分享他的一些经历与看法，教会我如何进行选择，帮助我确定自己的方向。在做毕业设计的每个阶段，老师都尽心尽力地进行了指导，耐心地指出每一步存在的不足，并对后续改进提出宝贵的意见，十分感激老师一路上给予我的帮助与支持。

另外，要感谢实验室的师兄师姐们在我进组之后，在学习与科研上给予我的指导与帮助。每当我遇到一些技术上的问题时，总会认真地帮我寻找解决方案，并提供我一些相应的解决思路，引导我自己去探索，帮助我一步步学会如何发现与解决问题。十分感谢你们一路上对于我的关心与支持。

最后，向这几年在我的学习和生活中曾给予我支持、教导、扶持和帮助的老师 and 同学表示深深的谢意。