

分类号 TP391.41;R73
U D C

密级 公开
编号 10741

兰州财经大学

LANZHOU UNIVERSITY OF FINANCE AND ECONOMICS

硕士学位论文

论文题目 基于卷积反向注意力模块和半监督学习的结肠
息肉分割算法研究

研究生姓名: 宋思良

指导教师姓名、职称: 何江萍 教授

学科、专业名称: 管理科学与工程

研究方向: 信息管理与信息系统

提交日期: 2024年5月31日

独创性声明

本人声明所呈交的论文是我个人在导师指导下进行的研究工作及取得的研究成果。尽我所知，除了文中特别加以标注和致谢的地方外，论文中不包含其他人已经发表或撰写过的研究成果。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示了谢意。

学位论文作者签名： 宋思良 签字日期： 2024年5月31日

导师签名： 何江萍 签字日期： 2024.5.31

关于论文使用授权的说明

本人完全了解学校关于保留、使用学位论文的各项规定， 同意（选择“同意” / “不同意”）以下事项：

- 1.学校有权保留本论文的复印件和磁盘，允许论文被查阅和借阅，可以采用影印、缩印或扫描等复制手段保存、汇编学位论文；
- 2.学校有权将本人的学位论文提交至清华大学“中国学术期刊（光盘版）电子杂志社”用于出版和编入CNKI《中国知识资源总库》或其他同类数据库，传播本学位论文的全部或部分内容。

学位论文作者签名： 宋思良 签字日期： 2024年5月31日

导师签名： 何江萍 签字日期： 2024.5.31

Research on colorectal polyp segmentation algorithm based on convolutional reverse attention module and semi-supervised learning

Candidate : Song Siliang

Supervisor: He jiangping

摘要

结直肠癌是一种常见且致死率较高的全球性癌症。及早进行息肉切除手术可以显著降低癌变的风险，因此定期进行结直肠检查对于预防结直肠癌非常重要。目前，结肠镜检查是最常用的结直肠息肉检测技术。然而，由于息肉的形态各异、大小不一，并且与粘膜的边界模糊，准确地分割息肉对于计算机来说仍然是一个巨大的挑战。随着深度学习的发展，许多息肉分割方法应运而生，在医学实践中极大地帮助医生寻找和识别结直肠息肉。然而，大多数研究使用通用的传统模型，很少有研究考虑到息肉分割和伪装物体检测之间的联系。因此，本文分别从全监督和半监督的角度出发来解决问题。具体而言，本文的主要研究内容包括以下两个方面：

(1) 基于卷积反向注意力模块的息肉分割网络：考虑到息肉的边界与结直肠壁之间的模糊不清，本文借鉴了伪装物体检测的思路中的反向注意力来解决息肉分割的问题。反向注意力可以预测出伪装目标的边缘特征，以往的研究未能考虑到卷积注意力模块（Convolutional Block Attention Module, CBAM）的定位性能，因此本文将其引入到卷积反向注意力中。首先，将息肉图像输入特征提取网络进行特征提取，并利用特征金字塔（Feature Pyramid, FP）模块增强特征。然后，通过聚合解码器（Aggregation Decoder, AD）模块获得粗略的分割结果。最后，利用卷积反向注意力（Convolution Reverse Attention, CRA）模块对粗略的分割结果进行细化。相比几种经典的息肉分割网络，本文提出的方法能够提高分割的准确性。

(2) 基于 mean teacher 的半监督息肉分割网络：鉴于标注数据对于息肉分割任务来说既耗时又昂贵，而无标注的息肉分割数据相对丰富，本文引入了半监督的思路，充分利用无标注数据。具体而言，一方面，本文利用基于数据变换的正则化方法提高网络的鲁棒性。另一方面，引入不确定性感知机制，使得学生模型能够学习到更可靠的信息。在常见的公开数据集上，本文提出的方法在半监督网络性能方面优于经典的半监督网络。

关键词： 息肉分割 全监督 半监督 反向注意力 正则化 不确定性感知

Abstract

Colorectal cancer is a common and deadly cancer worldwide. Early polyp removal surgery can significantly reduce the risk of cancer, so regular colon exams are important to prevent colorectal cancer. Currently, colonoscopy is the most commonly used technique for detecting colorectal polyps. However, because polyps vary in shape, size, and have blurred boundaries with mucous membranes, accurately partitioning polyps remains a huge challenge for computers. With the development of deep learning, many polyp segmentation methods have emerged, which greatly help doctors find and identify colorectal polyps in medical practice. However, most studies have used generic conventional models, and few have considered the link between polyp segmentation and camouflaged object detection. Therefore, this paper starts from the Angle of full supervision and semi-supervision to solve the problem. Specifically, the main research contributions of this paper are as follows:

(1) Polyp segmentation network based on convolutional reverse attention module: Considering the ambiguity between the polyp boundary and the colorectal wall, this paper uses the reverse attention in the idea of camouflage object detection to solve the problem of polyp segmentation. While reverse Attention can predict the edge features of camouflaged objects, previous studies fail to take into account the positioning performance of the Convolutional Block Attention Module (CBAM), so

this paper introduces it into the convolutional block attention module. Firstly, the polyp image was input into the Feature extraction network for feature extraction, and the Feature Pyramid (FP) module was used to enhance the features. Then, rough segmentation results are obtained through the Aggregation Decoder (AD) module. Finally, we use the Convolution Reverse Attention (CRA) module to refine the rough segmentation results. Compared with several polyp segmentation networks, the proposed method can improve the accuracy of segmentation.

(2) Semi-supervised polyp segmentation network based on mean teacher: Considering that labeling data for polyp segmentation tasks is time-consuming and expensive, while unlabeled polyp segmentation data is relatively abundant, this paper introduces a semi-supervised approach to fully utilize unlabeled data. Specifically, on one hand, this paper improves the network's robustness using regularization methods based on data transformations. On the other hand, an uncertainty perception mechanism is introduced, enabling the student model to learn more reliable information. On commonly used public datasets, the proposed method in this paper outperforms classical semi-supervised networks in terms of semi-supervised network performance.

Keywords : Polyp segmentation; Fully supervised; Semi-supervised; Reverse attention; regularizes; Uncertainty perception

目 录

1 绪 论	1
1.1 研究背景.....	1
1.1.1 结直肠息肉简介.....	1
1.1.2 结肠镜介绍.....	2
1.1.3 结直肠息肉自动分割.....	4
1.2 研究现状.....	5
1.2.1 基于机器学习的息肉分割研究现状.....	5
1.2.2 基于深度学习的息肉分割研究现状.....	6
1.2.3 半监督息肉分割研究现状.....	8
1.3 研究内容及创新.....	11
1.4 本文的组织架构.....	12
2 相关理论	13
2.1 深度学习.....	13
2.2 半监督学习.....	14
2.3 注意力机制.....	16
2.4 本章小结.....	17
3 基于卷积反向注意力模块的息肉分割网络	18
3.1 方法.....	19
3.1.1 深度学习模型架构.....	19
3.1.2 特征金字塔模块.....	20
3.1.3 聚合解码器.....	21
3.1.4 卷积反向注意力模块.....	22
3.1.5 损失函数.....	23
3.2 实验.....	23
3.2.1 实验环境.....	23
3.2.2 数据集介绍.....	23
3.2.3 评价指标.....	25
3.2.4 对比实验.....	25
3.2.5 消融实验.....	27
3.3 讨论.....	28
3.4 本章小结.....	29
4 基于 mean teacher 的半监督息肉分割网络	31
4.1 方法.....	32
4.1.1 mean teacher 模型框架.....	32
4.1.2 基于数据变换的正则化.....	33

4.1.3 不确定性感知.....	34
4.1.4 损失函数.....	35
4.2 实验.....	36
4.2.1 实验环境.....	36
4.2.2 数据集介绍.....	36
4.2.3 评价指标.....	37
4.2.4 对比实验.....	38
4.2.5 消融实验.....	40
4.3 讨论.....	40
4.4 本章小结.....	41
5 总结与展望.....	43
5.1 总结.....	43
5.2 展望.....	44
参考文献.....	46
致谢.....	53

1 绪 论

在医疗实践中，结肠镜检查作为筛查结直肠疾病的主要手段一直备受关注，已经成为一项热门的体检项目。然而，这也带来了相应的医疗压力，加重了医生的负担，从而增加漏诊的可能性，因此计算机自动诊断方法作为辅助医生进行结直肠病变筛查的技术就变得十分关键。在筛查结直肠疾病时，结直肠息肉的检测是最常见的检查项目。随着深度学习的快速发展，改进和开发一种高效的结直肠息肉分割方法变得非常必要。本文引入了伪装物体检测的框架模型，提出了基于卷积反向注意力模块的息肉分割网络。这种方法旨在克服结直肠息肉形态各异、大小不一以及与粘膜边界模糊的挑战。然而，深度学习的训练过程通常依赖于大量标记数据，而大部分结直肠息肉数据是没有标记的。这给结直肠息肉分割任务带来了困难。为了解决上述问题，本文考虑从半监督的角度出发，提出了基于 mean teacher 的半监督息肉分割网络。通过利用无标记数据，来进一步提高模型的性能和泛化能力。由此本文引入了基于数据变换的正则化方法以增强网络的鲁棒性，并采用不确定性感知机制，使学生模型能够从无标记数据中学习到的更可靠的信息。

综上所述，本文提出了基于卷积反向注意力模块的息肉分割网络和基于 mean teacher 的半监督方法，旨在改进结直肠息肉分割任务。这些方法对于辅助医生进行结直肠疾病筛查具有重要意义，有望提供更准确、更高效的结直肠息肉检测工具，辅助医生进行早期的诊断和治疗。

1.1 研究背景

1.1.1 结直肠息肉简介

结直肠息肉是一种在结直肠粘膜上突起的病变，通常不会引起一些特别明显症状。然而，一些结直肠息肉有可能会溃烂、出血或导致肠梗阻等其他并发症。在医学上，结直肠息肉主要分为两大类，包括肿瘤性结直肠息肉，例如腺瘤等，以及非肿瘤性结直肠息肉，例如炎性息肉等。

腺瘤是目前最常见的结直肠息肉类型。一般情况下腺瘤的表面是光滑平整的，有时

会呈现出“云状”形态，这都是腺瘤的正常形状。当腺瘤可能发生病变的时候，从组织学的角度来看，腺瘤会呈现出结构异常、细胞增生等特征，有时还可能出现不典型增生的形态学特征。炎性息肉是一种由间质组织、上皮成分和炎症细胞构成的非肿瘤性粘膜突起。炎性息肉被分为两大类，包括炎性假息肉和脱垂型炎性息肉。这两种炎性息肉的一般没有危害性，病变的可能性比腺瘤低。根据临床病例的统计，半数以上的结直肠息肉都是腺瘤，而且三成到五成的结直肠癌患者还存在并发的多种腺瘤^[1]。

根据结直肠癌筛查研究发现年龄是结直肠腺瘤发生的危险因素之一，其中 50 岁年龄段的腺瘤患病率占到总患病率的百分之三十左右^[2]。与此同时，尸检调查也进一步表明在七十多岁时大结直肠腺瘤的发生率已超过百分之五十。但是，青壮年大约年龄在 20 岁到 35 岁时，患病率仅为百分之二左右^[3]。在当今生活水平日益提高的大背景下，除了年龄影响之外，结直肠腺瘤的风险增加与身体质量指数 (Body Mass Index, BMI) 的增加也存在一定的相关性。根据对多家研究机构的研究结果进行整合分析，可以得出结论，当 BMI 每增加五个单位，则患有结直肠腺瘤的风险就将会增加两成^[4]。在实际情况中，通过测量腹部内脏脂肪组织体积得到的腹部肥胖，可能比 BMI 或腰围更能预测患有结直肠腺瘤的风险^[5]。

通过早期检测和切除结直肠息肉，可以预防结直肠癌的发生。结肠镜检查是一种常用的筛查方法，可以发现和切除结直肠息肉。对于高风险人群，如年龄超过 50 岁、有家族史以及存在其他危险因素，定期结肠镜检查非常重要。了解结直肠息肉的类型、危险因素以及早期检测和治疗的的重要性，可以帮助人们预防结直肠癌。因此，及早采取预防措施，并在医院进行定期的结肠镜筛查是至关重要的。

1.1.2 结肠镜介绍

目前用于结直肠疾病筛查和治疗的手段主要是结肠镜检查。结肠镜检查具有非常广泛的应用，也是对结直肠息肉进行定期的筛查的主要方式。一般来说，结肠镜检查是非常安全、准确并且耐受性良好的。结肠镜检查可以通过观察整个大肠和远端回肠的粘膜来评估结直肠的状况。结肠镜检查最重要的应用就是切除结直肠息肉。切除结直肠息肉以后可以有效地降低患结直肠癌的风险。

结肠镜检查是一种常见的内窥镜技术，通过将柔软的光纤束插入直肠，并沿着结肠逐渐推进，这样就可以直接观察并及时诊断结肠内部的粘膜状况^[6]。结肠镜通常由一根

灵活的管道组成，其中包括用于照明的光纤束、用于视频显示的摄像头以及用于组织检查的取样工具。在结肠镜检查中，医生可以观察到结直肠的外观、血管模式、病变、息肉以及其他异常的发现。如果发现任何的异常组织或者可疑的病变，医生可以立即进行组织取样（活检），以便进行进一步的病理学检查，来判断病变良性或者恶性。这使得结肠镜检查成为早期诊断结直肠癌以及其他结直肠疾病的重要手段。

结肠镜检查大多数适用于结直肠癌筛查为阳性、有缺铁性贫血、已进行过息肉切除或癌后切除监测、异常的结直肠影像学结果的患者^[7]。此外，结肠镜检查是常用的结直肠癌筛查方法之一^[8]，是非常热门的体检项目。每年都有大量患者进行结直肠癌筛查^[9]。

结肠镜检查的效果与患者对手术的接受程度息息相关，而这主要取决于患者对肠道准备工作。肠道准备的质量会影响检查的完整性以及是否需要提前停止或二次检查^{[10][11]}。充分的肠道准备是结肠镜检查的主要考虑因素。腺瘤筛查的准确率与精细的准备工作和较长的停药时间之间具有一定的关联性。部分患者经过结肠镜检查以后可能出现不良反应，例如呼吸抑制、血压下降等。当然，医生高超的技术和丰富的临床经验有助于防止息肉切除后产生不良反应。在结肠镜检查中，医生可以通过不同的技术和操作来执行治疗过程。例如，对于发现的息肉，医生可以使用结肠镜上的工具直接切除息肉，以减少患结直肠癌的风险。在临床实践中，这种操作过程被称为息肉切除术。对于有恶性病变风险的肿瘤，结肠镜检查也可以作为引导下的微创手术工具，用于切除病变。尽管结肠镜检查是一项有益的诊断和治疗工具，但由于患者的个体各异也存在一些风险和并发症。一般的并发症主要是包括出血、穿孔和感染^[12]。然而，在实际操作中，这些并发症的发生率相对较低，而且在临床经验丰富的医生操作下，结肠镜检查的风险可以说是无限接近于零。就目前来说，结肠镜检查是一种常用且有效的方法，用于评估结直肠健康状况、诊断疾病，同时还能进行相应的治疗。患者肠道准备的质量、医生的技术和医生的经验以及不良反应都对结肠镜检查的效果和安全性产生相应的影响。

总体而言，结肠镜检查是一种可靠且有效的检查项目，并且广泛应用于结直肠疾病诊断和治疗。结肠镜检查不仅可以帮助医生评估结直肠健康状况和发现异常，还可以进行治疗性干预，如息肉切除。通过结肠镜检查，早期病变可以被及时发现和处理，从而提高患者的治疗成功率和生存率。因此，结肠镜检查在结直肠疾病管理中扮演着重要角色。

1.1.3 结直肠息肉自动分割

图像分割是一项关键的图像处理技术，主要是将图像划分为具有不同特征和不同属性的区域。图像分割在许多领域中都具有极为重要应用，尤其在医学图像的肿瘤检测、自然图像的物体检测等领域备受青睐。相比于传统的手动分割方法，深度学习技术为图像分割带来了突破性的进展，使得分割过程更加高效、快速且准确。

在医学图像的肿瘤检测领域，图像分割的重要性是不可忽视的。医生手动分割器官或病灶区域是一项非常繁琐且耗时的任务，然而深度学习技术在医学图像的实际应用可以减轻医生的负担。通过图像分割，可以将感兴趣的区域从图像中提取出来，用于病灶检测、病灶分类以及后续的感兴趣区域（Region of Interest, ROI）划分。分割过程通常是将整个图像输入到模型中，并生成相应的分割结果。训练分割模型所需的数据集由医学图像和对应的手动分割结果组成，其中医学图像在患者进行检查后整理得到，手动分割结果通常是从先前的医学专家标注中获得。

与图像分类不同，图像分割需要捕捉到整个图像的全局空间上下文，从而实现更精确的分割。肿瘤图像的分割在癌症诊断中发挥着不可替代的作用。例如，在结直肠癌等肿瘤的研究中，对结直肠息肉进行准确的分割是一项对于计算机视觉来说极具挑战性的科研任务，但同时对于医生来说又是十分关键的医学任务。结直肠息肉是结直肠内突出的异常组织，通常是结直肠癌发展的前期病变。通过对结直肠息肉进行分割，医生可以确定息肉位置、形状和大小，从而有助于及时地发现潜在的癌变。早期发现和切除可疑的癌变结直肠息肉可以大大提高结直肠癌的早期诊断率，不仅有预防的作用，还能增加治愈可能性。

此外，结直肠息肉的分割还提供了关于息肉的详细信息，例如大小、形态、位置和组织特征。这些信息对于制定相应的治疗计划具有很大的参考意义。医生可以根据分割结果来确定是否需要手术切除、内镜切除或者其他治疗方法。准确的结直肠息肉分割在手术过程中也具有重要意义，可以提供手术导航的参考，帮助医生确定手术切除的范围和方式，从而最大程度地减少对正常组织的损伤。

除了临床应用，结直肠息肉分割还在科学研究和教育领域发挥着重要作用。通过对大量结直肠息肉进行分割，可以建立大规模结直肠息肉的数据库，用于分析和研究结直肠癌发展的模式和特征。这对于深入了解结直肠癌的发展过程、癌变预测和癌症治疗都具有重要意义。同时，结直肠息肉分割还可以用于医学生和医生的培训，提高他们对结

直肠疾病的认识和诊断能力。通过使用大量的结直肠息肉图像进行分割训练，医学相关的教育者可以为医学生提供有关结直肠息肉的实例，来帮助他们更全面地学习如何准确地识别和分割结直肠息肉，从而提高他们的临床技能。

总而言之，图像分割在生物医学和肿瘤分割领域具有重要意义。特别是在结直肠癌早期诊断、治疗规划和手术导航中，结直肠息肉分割发挥着十分关键的作用。通过准确地分割结直肠息肉，医生可以及早发现潜在的癌变，并制定相应的治疗计划。此外，结直肠息肉分割还为科研和教育提供了宝贵的工具和资源，有助于深入了解结直肠癌的发展过程，并培养医学生和医生的专业技能。随着深度学习和图像处理技术的不断发展，图像分割在医学领域的应用将继续取得更多突破和进展，为患者的诊断和治疗带来更多的帮助。

1.2 研究现状

1.2.1 基于机器学习的息肉分割研究现状

在深度学习尚未兴起之前，在自动息肉分割领域主要是靠传统的机器学习。研究人员通过分析息肉的颜色、形状、纹理等形态学的特征信息，提取出相应的特征，来进行分割任务。这些特征被认为是与息肉区域相关的重要线索，因此被广泛用于传统的机器学习算法中。

Yao 等人^[13]提出的基于模糊聚类和变形模型的 CT 结肠镜结肠息肉分割，是最早的自动息肉分割方法。该方法结合了知识引导强度调整、模糊均值聚类以及可变形模型。通过引入知识引导强度调整，算法能够根据先验知识对特征进行加权，提高了分割的准确性。Lu 等人^[14]提出了基于多阶段概率二值学习和组合模型的三维 CT 结肠镜息肉精确分割，是一种融合低、中水平信息的三阶段概率二值分类方法。该方法在三维 CT 结肠镜图像中实现了有效的息肉分割。Gross 等人^[15]的研究集中于多尺度滤波增强初级边缘的分割算法。该思路是利用了在不同尺寸下对图像的过滤，从而增加了边界信息，进而提高了分离算法的效率。除此之外，Ganz 等人^[16]还采用了 NBI 窄带成像作为先验知识，优化了 gPb-OwT-UCM 算法的尺度选择问题提出了 Shape-UCM 算法。该方法通过结合息肉的形状信息，获得了比以往更好的分割结果。

传统的人工设计特征和机器学习方法存在一些局限性。首先，人工设计特征需要依

赖领域专家的丰富的经验，难以捕捉到数据中的复杂关联性信息和高层次语义特征，而且及其的耗时费力。其次，传统的机器学习算法在处理大规模数据时可能会面临性能瓶颈。然而幸运的是，深度学习的兴起为自动息肉分割带来了新的飞跃。利用深度学习可以从海量的数据中提取到多尺度多维度的特征，以此实现更加精准的图像分割。深度学习模型的引入使得自动息肉分割在准确性和效率方面取得了显著的提升。随着深度学习的发展，自动息肉分割领域的研究将会变得更加多样和创新。通过不断探索和改进算法，可以期待在未来实现更准确、更高效的自动息肉分割技术。

1.2.2 基于深度学习的息肉分割研究现状

传统的方法在息肉分割任务中受到形态学特征有限的信息表示能力的限制。这些方法缺乏泛化能力，不太适合大规模的应用。然而，随着深度学习的兴起之后，出现了卷积神经网络，极大地推动了息肉分割领域的快速发展，人们越来越倾向于使用深度学习提取到的高维特征来处理各种分割任务。深度学习可以通过对息肉图像进行高维特征提取，并通过前向传播和反向传播的迭代学习来获得更准确的息肉分割结果。深度学习在息肉分割任务中的应用是基于大规模标注数据和强大的计算能力的优势。相比于传统的机器学习，深度学习更加智能，更加强大。通过深度学习模型，可以从图像中提取更高层次、更丰富的特征表示，从而实现更准确的息肉分割。

卷积神经网络 (Convolutional Neural Networks, CNN) 的发展，尤其是 U-Net^[17] 的引入给息肉分割任务带来了许多启发性的思考，推动研究者们不断地提出新的模型并展现出令人兴奋的结果。其中最经典的延伸是 UNet++^[18]，该网络综合了长连接和短连接的方案，利用短连接进行模型训练，而长连接则用于获取更多有价值的信息。此外，还有一些结合了 ResNet^[19] 和 Unet^[17] 的模型，如 ResUNet-mod^[20] 和 ResUNet++^[21]，这些都是常用于息肉分割的深度学习模型。

除此之外，还有边界感知模型。边界感知模型最初主要用于目标检测，后来也被越来越多地应用于息肉分割任务。由于医学图像分割的特殊性，准确的边界感知表示具有重要意义。SFA (Selective feature aggregation)^[22] 是其中的较为经典模型，该模型通过构建区域选择性特征聚合网络和边界约束来提高分割性能。该模型主要由共享的编码器、预测息肉区域的解码器和两个相互提取边界信息的解码器组成。通过引入三个级联的组件编码器和解码器以及嵌入选择性内核模块在卷积层中，实现了有选择性的特征

聚合。

为了获得优异的性能，需要对相关信息进行优先排序，而不是像以往一样平等对待所有信息。所以通过整合注意力机制，使得模型能够专注于最重要的特征，从而提高分割性能。例如，CCBANet (cascading context and balancing attention) [23] 利用预测平衡背景、息肉和边界的注意力来调整输出。该模型提出了两个创新的模块，即 CCM (Cascading Context Module) 和 BAM (Balancing Attention Module)。CCM 用于级联连接上下文的模块，并使用 BAM 来实现有效的息肉分割。这种模型的引入能够挖掘更多的边界信息。另外，一些模型还将 Transformer 与息肉分割相结合，如 DS-TransUNet [24]。与之前基于 Transformer 的解决方案不同的是，DS-TransUNet 首次采用基于 Swin Transformer 的双尺度编码器的子网络来提取不同尺度尺度的粗粒度和细粒度特征表示。DS-TransUNet 模型所提出 TIF (Transformer Interactive Fusion) 模块，能够通过 self-attention 机制训练模型学习不同尺度特征之间的全局依赖关系。DS-TransUNet 模型利用基于注意力机制的解码器来生成细致的分割结果。另外一种常用的方法是使用区域提议网络 (Region Proposal Networks, RPN) 和 CNN 相结合。RPN 负责生成候选框 (即目标的可能位置)，而 CNN 负责对这些候选框进行分类和边界框回归。这种方法在深度学习中非常成功，产生了一系列的目标检测算法，如一些经典的算法，包括 Faster R-CNN [25]、YOLO [26] 和 SSD [27]。

Faster R-CNN 是引入了 RPN 来生成候选框。RPN 通过在不同位置和尺度上滑动一个锚框 (anchor) 来生成候选框，然后使用 CNN 对这些候选框进行分类和回归。Faster R-CNN 具有较高的检测精度和较快的检测速度，在许多目标检测竞赛中取得了优秀的业绩。YOLO (You Only Look Once) 是将目标检测问题转化为一个回归问题。YOLO 具有较快的检测速度，但在小目标检测和定位精度方面可能略逊于 Faster R-CNN。SSD (Single Shot MultiBox Detector) 是一种将多个尺度特征图与不同尺度的锚框相结合的目标检测算法。SSD 通过在每个特征图上预测目标的类别和边界框来进行检测。SSD 在检测精度和速度之间取得了一定的平衡。

除此之外，在息肉分割任务中，特征融合也是一种极为有效的方法。特征融合可以处理对象大小变化的问题并利用高层特征和低层特征的融合来提高分割性能。目前存在多种特征融合策略可以实现这一目标。其中，PraNet [28] 是邵岭教授团队提出的一种经典的息肉分割网络。该网络创造性地提出卷积并行思路来融合模型所提取的高级特征，同时通过反向注意力模块对边界线索进行挖掘探索。PraNet 采用了 parallel reverse

attention network 的结构，首先使用 PPD 结构 (parallel partial decoder) 聚集高层的语义特征，然后基于混合的特征生成一个全局的初始化特征指引区域 (guidance area)。此外，PraNet 还引入了 RA 模块 (Reverse attention) 来挖掘对应的边缘信息。通过这种特征融合和边界线索的挖掘，PraNet 能够获得更为准确的分割结果。

另外一个特征融合模型是 CaraNet^[29]，该模型专注于改善小型医疗对象的分割性能。对于早期疾病发现和降低死亡率来说，准确检测微小医疗物体具有十分重要的意义。CaraNet 通过并行部分解码器对高层特征进行聚合，从而获取上下文信息并生成全局映射图用于分割。这个全局映射图能够捕捉到粗略的息肉组织位置。为了进一步挖掘边界信息，CaraNet 采用了反复使用轴向反向注意力 (A-RA) 模块的策略来构建分割目标和分割边界之间的潜在关系。通过区域和边界之间的迭代交互，CaraNet 可以校正预测结果中一些不一致的区域，从而提高分割的准确性。

总的来说，深度学习在息肉分割任务中取得了显著的进展。通过应用 CNN、边界感知模型、注意力模型以及特征融合等技术，研究人员已经开发出一系列高效准确的模型来实现对结直肠息肉的精准分割。这些模型利用深度特征提取和信息聚合的能力，能够更好地捕捉息肉的形态和边界信息，从而提高分割的准确性和稳定性。然而，仍然存在一些挑战，包括数据的类别不充分、专家标注的困难、模型的域迁移等问题，这些问题急需进行更深入的研究。通过克服这些挑战，可以进一步提升息肉分割的性能，并为早期疾病诊断和治疗提供更有效的帮助。

1.2.3 半监督息肉分割研究现状

半监督学习旨在利用大量未标记数据与标记数据相结合的方式训练更高性能的分割模型。对于半监督息肉分割研究，本文的目标是结合标注数据集和未标注数据集构建一个高效的深度学习模型，并使其性能与在同样标记量的数据集上训练的经典模型相当甚至是超越。本文主要将这些半监督息肉分割方法分为以下三种策略。

第一种策略是伪标签的半监督学习^[30]。该方法是利用事先在大型数据集上进行预训练后的网络，来对没有标签的数据进行预测，并将预测的标签当作伪标签。为了提高伪标签的质量，常见的方法是为未标记图像分配伪注释，并且同时使用标记的图像和伪标签的图像来更新训练模型。随着模型迭代次数的增加，会逐步提高伪标签的准确性。

例如，Yao 等人^[31]为了提高来自未知分布的未标记图像的伪标签质量，提出了一种

置信度感知的交叉半监督网络。利用傅里叶变换来学习跨域的低级统计信息，并增强图像以纳入跨域信息。该方法通过伪标签的置信度来指导模型对未标记数据的预测，提高伪标签的准确性。。Shi 等人^[32]提出了一个保守-激进网络 (conservative-radical network, CoraNet)，其中保守设置倾向于预测像素属于背景，而激进设置倾向于预测像素属于前景。该网络有三大模块，一是保守-激进模块 (conservative-radical module, CRM)，二是特定的区域分割网络 (certain region segmentation network, C-SN)，三是不确定的区域分割网络 (uncertain region segmentation network, UC-SN)。Zhang 等人^[33]考虑从伪标签纠正的思路来训练模型，通过在骨干分割网络之后添加一个错误分割网络来纠正未标记数据的分割结果，提出了基于纠错的标签纠错方法，即 ECLR。错误分割网络用于纠正主网络在未标记数据上的预测，从而提高伪标签的质量。这些方法都利用了未标记数据来增强分割模型的性能。通过迭代来生成伪标签并更新模型，可以逐步改善伪标注的质量，从而进一步提升模型的性能。

第二种策略是无监督正则化的半监督学习。这种方法利用未标记数据通过正则化的思路来训练分割模型。常用的正则化方法包括一致性学习、协同训练等手段。这种训练思路不同于伪标签的生成和模型的更新迭代训练。最近几年越来越多的学者在研究通过无监督损失函数等无监督正则化手段来有效地利用无标签数据。

其中，一致性学习是一种常见的无监督正则化方法，一致性学习通过鼓励模型在不同的输入变换下产生一致的预测结果来提高模型的鲁棒性。例如，Samuli 等人^[34]引入了自集成，进而提出了时间集成策略，将未标记数据的指数移动平均 (Exponential Moving Average, EMA) 预测结果与标记数据的预测结果进行一致性对比。这种一致性对比可以通过计算预测结果之间的差异度量来实现，例如均方差或互信息。通过最小化差异度量，模型被迫在不同的输入变换下产生一致的预测结果。该方法的基本思想是在不同的时间点训练多个模型，然后将模型的预测结果结合起来进行最终预测。为了减轻训练过程中维持 EMA 预测的负担，Tarvainen 等人^[35]在时间集成策略的基础上提出使用教师模型和学生模型之间的 EMA 权重进行训练，并且对学生模型和教师模型同时加入相同的数据扰动来使用一致性损失函数训练模型。在半监督研究中，“mean-teacher”架构的方法由于其简单性而被广泛采用。Zeng 等人^[36]在“mean-teacher”架构的基础上对教师模型中 EMA 的加权方式进行提升改进，为了让教师模型可以自发地、有目标地进行更新迭代，他们引入了来自学生模型在标记集上表现的反馈信号。

协同训练是另一种常见的无监督正则化方法。协同训练使用多个模型来协同进行训

练，每个模型都在不同的视角下处理数据。例如，自监督协同训练方法，其中一个模型通过自监督任务（如图像旋转预测）从未标记数据中学习表示，而另一个模型通过分割任务从标记数据中学习。通过这种方式，未标记数据的表示可以通过自监督任务来提供额外的信息，帮助分割模型进行训练。对抗学习是一种使用对抗网络（Generative Adversarial Network, GAN）的无监督正则化方法。在分割任务中，生成器网络用于生成伪分割掩码，而判别器网络则用于区分真实的标记数据和生成的伪分割掩码。通过对抗训练的方式，生成器网络被迫生成逼真的伪分割掩码，以欺骗判别器网络。这种对抗学习的框架可以通过最小化生成器和判别器之间的对抗损失来实现，同时最大化分割模型的性能。熵最小化是一种将未标记数据纳入训练过程的无监督正则化方法。熵最小化通过最小化未标记数据的预测熵来提高模型的鲁棒性。具体而言，模型被迫产生确定性的预测结果，使得熵最小化。这可以通过最小化预测结果的交叉熵损失函数来进行实现。

第三种策略是生成模型的半监督学习。这种方法使用生成模型来生成伪标记数据，并将其与真实标记数据一起用于训练分割模型。生成模型可以是生成对抗网络（GAN）或变分自编码器（VAE）。生成模型可以学习未标记数据的潜在表示，并生成与标记数据类似的伪标记数据。这些伪标记数据可以用于增强分割模型的训练。先验知识的半监督学习是一种利用未标记图像的方法，通过让模型提前学习到分割目标形状、分割数量等先验知识，以此来提高医学图像分割的表示能力。一些半监督算法利用先验知识来增强对新任务的表示能力。然而，在应用先验知识时需要考虑过拟合和不可微分的情况。

例如，为了克服网络在冷启动阶段被随机初始化的问题，huang 等人^[37]添加了并行的重构预训练，提出了一种两阶段、半监督学习的神经元分割方法。这种方法通过使用并行的重构预训练来避免网络从随机初始化开始训练。Wang 等人^[38]提出了一种分离的区域级对比学习方案，即 SepaReg 网络。该网络用超像素将图像划分为区域，并基于对比学习来学到器官内和器官间的表示，然后使用该模型来初始化半监督框架，显著提高了半监督的性能。hu 等人^[39]提出了一种有监督的局部对比损失。该网络将自监督的对比学习引入半监督框架中，来增强模型对解剖先验知识的学习能力。Wu 等人^[40]通过交叉补丁密集对比学习框架来对学习教师模型和学生模型的特征，同时结合一致性正则化和熵最小化技术提升模型的鲁棒性，进一步提升了分割的性能。这些方法的目标都是利用先验知识来提高医学图像分割的性能。

总的来说，半监督学习在息肉分割中的应用，旨在利用未标记数据来增强分割模型的性能。这些方法可以通过生成伪标签、无监督正则化或生成模型等策略来实现。通过

利用未标记数据，可以提高模型的泛化能力和鲁棒性，从而在有限的标记数据下实现更好的分割性能。

1.3 研究内容及创新

本文的研究内容主要集中在结肠镜的息肉分割领域，通过全监督和半监督两个角度展开研究。针对全监督的情况，本文提出了卷积反向注意力网络（CRANet）的结直肠息肉分割方法，与以往的方法相比，该方法引入了卷积注意力模块（Convolutional Block Attention Module, CBAM）^[41]到卷积反向注意力（Convolution Reverse Attention, CRA）模块，使得息肉分割的性能得到了相应的提升。同时还使用特征金字塔模块和聚合解码器模块对结直肠息肉进行粗分割。针对半监督，本文提出了基于 mean teacher 的半监督息肉分割网络。利用基于数据变换的正则化提高网络分割的鲁棒性，利用不确定性感知提升 student 网络的性能。使得网络在标记数据很少的情况下，也能最大程度进行精准的息肉分割。本文的创新点如下：

（1）利用了特征金字塔（Feature Pyramid, FP）和聚合解码器（Aggregation Decoder, AD）：这两个模块用于定位结直肠内的隐匿的息肉。特征金字塔能够对不同尺度的特征进行融合，从而更好地捕捉到细小息肉的信息。聚合解码器则用于将定位到的息肉进行粗分割。

（2）引入了卷积注意力模块（Convolutional Block Attention Module, CBAM）：该模块被嵌入到卷积反向注意力网络（Convolutional Reverse Attention Network, CRANet）中，用于识别已经进行粗分割的息肉，并细化其分割边界。通过引入卷积注意力模块，本文提出的方法在息肉分割的准确性上有所提升。

（3）基于 mean teacher 的半监督息肉分割网络结合了基于数据变换的正则化和不确定性感知的方法。利用数据变换的正则化可以提高网络的鲁棒性，使其在标记数据较少的情况下仍能进行有效的息肉分割。同时，不确定性感知的技术能够提升学生网络的性能，进一步提高半监督息肉分割的准确性。

1.4 本文的组织架构

本文是针对结直肠镜息肉分割的研究，基本的论文结构如下：

第一章主要介绍研究背景及其意义和研究现状。首先，研究背景及其意义分别是对结直肠息肉介绍、结肠镜介绍和结直肠息肉自动分割，然后研究现状主要介绍了基于机器学习的息肉分割研究现状、基于深度学习的息肉分割研究现状和半监督息肉分割研究现状。最后，介绍本文的研究内容及创新和组织架构。

第二章主要介绍深度学习理论背景。包括深度学习、半监督学习、注意力机制的基础知识以及它们常见的模型。

第三章主要介绍基于卷积反向注意力模块的息肉分割网络的研究内容和实验结果。介绍了研究的总体设计和深度学习模型的架构，包括特征金字塔（Feature Pyramid, FP）、聚合解码器（Aggregation Decoder, AD）以及卷积反向注意力（Convolutional Reverse Attention），同时通过消融实验和对比实验证明模型的有效性，最后分析实验结果并且讨论。

第四章主要介绍基于 mean teacher 的半监督息肉分割网络的研究内容和实验结果。首先，介绍半监督网络的总体实验设计和主干网络架构，包括利用数据变换的正则化和不确定性感知，然后分析实验结果，最后对实验结论进行讨论。

第五章主要介绍总结和展望。首先对本文内容进行总结，然后根据研究内容展示其不足与展望。

2 相关理论

2.1 深度学习

在过去的几年里，深度学习（Deep learning）被广泛认为是机器学习（Machine learning）领域的风向标，并在各种应用领域取得了巨大成功。而且深度学习在计算机视觉、自然语言处理等领域中的使用已经非常普遍^{[42][43]}，并且取得了出色的结果，有时甚至能够匹敌或超越人类。随着时间的推移，深度学习领域发展迅速，不断涌现出新的研究成果。这一方面不仅得益于数据获取能力的不断增长，另一方面也受益于硬件技术的惊人进步，比如高性能计算^[44]。

特征工程是机器学习算法性能有效发挥的关键，合适的数据表示可以使得机器学习算法达到更加令人瞩目的效果。自机器学习爆火以来，特征工程一直是这其中最为关键的研究课题，特征工程研究的突破可以为相关研究提供更为强大的性能。特征工程的目标是从原始数据中构建新的特征，然而这通常需要大量的人力和领域知识。以计算机视觉领域为例，如定向梯度直方图（HOG）^[45]、尺度不变特征变换（SIFT）^[46]和词袋（BoW）^[47]。深度学习算法与传统的特征工程相比，能够自动地进行特征提取，这为研究人员提供了使用尽可能少的人力和领域知识来提取判别特征的机会^[48]。深度学习算法具有多层数据表示体系结构，是一种深度的神经网络，包括输入层、隐藏层以及输出层。

深度学习源自传统的神经网络，在其基础上进行了改进和发展。目前深度学习技术在音频和语音处理、视觉数据处理、自然语言处理等领域取得了显著成果^[49]。从最初的AlexNet^[50]网络到目前最先进的人工智能大模型，深度学习网络结构不断更新迭代。这些技术已经在植物病害分类^[51]、目标检测^[52]、医学图像分析^[53]等科研领域得到广泛应用，并推动了相关领域的快速发展。随着时间的推移，可以期待深度学习在更多领域的应用以及进一步的突破。随着深度学习技术的不断发展，可能会看到更多创新的应用和突破性的进展。下面简单介绍一些深度学习的经典方法。

首先是深度学习的开山之作神经网络（Neural Networks）。神经网络是深度学习的基础性工具。神经网络是由多个神经元层组成，每个神经元层包含一组节点或称为神经元。这些神经元通过权重和激活函数的训练学习输入数据的特征，通过隐藏层并到达输出层，从而生成相应的输出。深度学习网络一般是包含许多层，这就使得可以对输入数据进行逐层进行特征提取和特征学习。

其次是深度学习网络训练的关键步骤反向传播 (Backpropagation)。反向传播是一种用于训练神经网络的必备算法。反向传播通过计算预测输出与实际输出之间的误差,并将误差沿着网络进行反向传播,更新每个神经元层之间的权重,以使网络的输出更加靠近标签。以此循环往复,更新迭代,使得网络更为强大。反向传播是深度学习中实现梯度下降优化的关键步骤。

另一个重要的概念是 CNN。CNN 是一种特殊类型的神经网络,主要用于图像和视觉数据处理。CNN 利用卷积层和池化层来有效地提取图像中的特征,并通过全连接层进行分类或回归任务。CNN 在计算机视觉领域的许多任务中都取得了突破性的成果,如图像分类、目标检测和语义分割等领域,击败了传统的机器学习。

此外,还有循环神经网络 (Recurrent Neural Network, RNN)。RNN 是一种具有循环连接的神经网络结构,特别适用于处理序列数据,如文本和语音。RNN 通过利用记忆单元和时间步连接,可以捕捉序列数据中的时序依赖关系。这使得 RNN 在机器翻译、语音识别和文本生成等任务中表现出色。

还有一些其他的深度学习技术,例如生成对抗网络 (GAN),用于生成新的数据样本,以及强化学习 (Reinforcement Learning),用于基于奖励信号进行决策和学习。这些技术不断发展和演进,为各种应用领域提供了解决问题的新途径。深度学习是一种强大的机器学习方法,通过利用大规模数据和复杂的神经网络结构,能够自动地从数据中学习和提取特征,并在各种任务中取得优异的结果。深度学习在计算机视觉、自然语言处理、语音识别等领域的广泛应用,也在推动人工智能技术的进步和创新。

2.2 半监督学习

在许多应用领域中,一个普遍存在的问题是缺乏足够的标记数据。相比之下,未标记数据相对容易获取。然而,在实际情况中获取标记数据却是一项困难的任務,因为这是需要经验丰富的领域专家来对未标记数据进行标记。为了解决这一问题,半监督学习应运而生,作为监督学习和无监督学习之间的一种中间方法。半监督学习方法包括自训练、协同训练、多视图学习和 TSVMs 等技术。传统的半监督学习方法主要分为半监督分类和半监督聚类两类,相较于传统的有监督和无监督学习技术,半监督学习具有更高的准确性。

与有监督学习和无监督学习相比,半监督学习是一种相对较新的学习方法。有监督

学习是从一组给定的示例中进行学习，其中每个示例由问题实例和其标签组成（通常由该领域的专家提供）。例如，在分类问题中，将待分类的数据元素表示为特征向量，并将类别表示为分类标签。示例集也被称为训练集或标记集，用于构建分类器，该分类器可以对任何给定的新数据实例进行分类。在无监督学习中，是不提供任何标记集，而是获得一组未标记的数据。无监督学习的主要任务是发现数据集中存在的结构，例如聚类结构。与有监督学习相比，无监督学习更加困难，因为模型没有基本事实来评估结果的准确性^[54]。

在许多现实世界的应用领域（如图像处理 and 文本处理）中，存在大量未标记数据，而专家对未标记数据进行标记是一项昂贵且费时的任务。所以许多实际应用的数据集都是稀疏的标记数据。在这种情况下，半监督学习成为一种重要的研究方向，提供了将混合标记和未标记数据纳入学习框架的方法，在标记数据的引导下，对未标记的数据进行充分且有效地利用。

过去的研究中提出了许多半监督分类和半监督聚类的方法。半监督学习方法可以分为生成式和判别式方法^{[55][56]}。生成式方法采用期望最大化（EM）等技术，依赖于输入数据的分布。然而，生成式方法可能会在输入数据与分类任务不匹配时报错^[57]。判别式半监督方法^[58]包括概率和非概率方法，如转导支持向量机（TSVM）和基于图的方法，这些方法使用假设类别的密度的方法。

半监督学习解决了监督学习和无监督学习方法的问题，并提供了更加准确地模式识别未标记数据和标记数据的方法。对于标记数据有限或难以获取的情况下尤为重要，通过利用未标记数据的信息，半监督学习能够提高学习算法的性能和泛化能力。下面简单介绍一些半监督的经典方法。

自训练（Self-training）是一种常见的半监督学习方法，自训练采用迭代的方式进行。首先，从已标记数据中构建一个初始分类器，然后利用该分类器对未标记的数据进行预测，并将预测结果中置信度较高的样本添加到已标记数据集中。通过这种方式，新标记的数据可以用于进一步改进分类器的性能。这个过程可以进行多次迭代，逐渐增加已标记数据的数量。

协同训练（Co-training）是另一种常见的半监督学习方法，协同训练假设数据包含多个视图或表示，并且每个视图提供了关于数据不同方面的信息。在协同训练中，有两个独立的分类器，每个分类器只使用其中一个视图的数据进行训练。然后，这两个分类器使用彼此的预测结果进行信息交换，并将置信度较高的样本添加到对方的训练集中。

通过这种方式，两个分类器可以相互促进，提高整体的分类性能。

多视图学习 (Multi-view Learning) 是一种多维度的半监督学习方法，多视图学习利用多个数据视图来提供更全面和丰富的信息。每个数据视图可以是不同的特征表示、不同的数据来源或不同的数据视角。通过综合多个视图的信息，多视图学习可以提高分类器的性能和泛化能力。

转导支持向量机 (Transductive Support Vector Machines, TSVMs) 是一种判别式的半监督学习方法，转导支持向量机通过使用未标记数据来调整分类器的决策边界，以最大程度地减少未标记数据中的错误分类。TSVMs 利用未标记数据的分布信息来提高分类器的鲁棒性和性能。

除了上述提到的方法，还有其他一些半监督学习技术，如基于图的半监督学习、半监督降维以及半监督聚类等方法。这些方法在不同的问题和应用领域中具有广泛的应用。半监督学习是一种弥补有监督学习和无监督学习之间差距的方法。半监督学习利用未标记数据的信息来提高学习算法的性能和泛化能力，对于标记数据有限或难以获取的情况下尤为有用。通过充分利用未标记数据，半监督学习有助于解决实际应用中的数据标记问题，并提供更准确的模式识别和分类能力。

2.3 注意力机制

注意力机制在神经网络研究中扮演着重要的角色，并且注意力机制可以被理解为一种基于提取图像特征的动态权重调整过程。注意力机制的发展主要是计算机视觉领域，其中最多是在自然图像领域的发展。为解决语义分割和三维重建等领域的问题，注意力机制被广泛应用。

除了自然图像领域，注意力机制也在医学图像领域取得了相应的发展。注意力机制已成功应用于医学图像分割，例如，MedT (Medical Transformer)^[59]。使用门控轴向自关注层构建了用于超声和显微镜图像分割的 Local-Global (LoGo) 网络，TransUNet 与 U-Net 和 FCN^[60] 等纯卷积神经网络相比，所有这些基于注意力机制的分割方法都使得分割性能得到了显著的提升。下面简单介绍一些注意力机制的经典方法。

自注意力机制 (Self-Attention) 是注意力机制的一种特殊形式，自注意力机制在序列或图像的不同位置之间建立了交互关系。自注意力机制可以通过计算每个位置与其他位置的相关性得分来为每个位置分配权重。这种机制广泛应用于自然语言处理任务，

如机器翻译、语言建模和文本生成等。

多头注意力机制 (Multi-Head Attention) 是一种扩展的注意力机制形式, 多头注意力机制允许模型在不同的表示子空间中学习多组注意力权重。通过使用多个注意力头, 模型可以同时关注输入的不同方面和语义层次, 从而提高表示能力和模型性能。

视觉注意力 (Visual Attention) 是指在计算机视觉任务中, 模型可以选择性地关注图像的特定区域或特征, 以便更好地解决任务。视觉注意力在目标检测、图像分类、图像生成等任务中得到广泛应用, 视觉注意力使得模型能够自动地学习和集中于图像中最相关和重要的部分。

跨模态注意力 (Cross-Modal Attention) 是指在多模态学习任务中, 模型可以学习不同模态之间的相关性, 并根据这些相关性来调整模态之间的注意力分配。例如, 在图像与文本之间的视觉问答任务中, 模型可以通过跨模态注意力机制来关注图像中与问题相关的区域, 以便生成准确的回答。

注意力可视化 (Attention Visualization) 是一种用于理解和解释注意力机制在模型决策中的作用的技术。通过将注意力权重映射回输入数据, 可以可视化出模型对输入的关注程度和重要性。这种可视化有助于解释模型的决策过程, 并提供对模型行为的观察。

除了上述内容, 注意力机制还在语音识别、推荐系统、强化学习等领域得到了广泛的应用和研究。随着深度学习的快速发展, 注意力机制将继续发挥重要作用, 并在各个领域推动模型的性能和效果提升。

2.4 本章小结

第二章主要介绍深度学习理论背景。包括深度学习、半监督学习以及注意力机制的基础知识。通过对比半监督学习和全监督学习来深化对半监督的认识, 同时介绍了相应主流的方法。在注意力机制的章节介绍计算机视觉注意力的基本概念以及常用的算法。

3 基于卷积反向注意力模块的息肉分割网络

息肉的切除活检是早期筛查癌症的重要手段，对于患者的健康至关重要。在医学实践中，获得准确的息肉分割结果是切除息肉的关键步骤之一。然而，准确的分割并非易事。在进行息肉分割时，面临着一些挑战。一方面，息肉在整个图像中的占比通常不到三分之一，且其形状大小很不规则，难以快速且准确的检测，这给准确定位息肉位置带来了较大的困难。另一方面，息肉与肠道边界之间的界限模糊不清，使得准确分割息肉边界变得更加具有挑战性。这种情况类似于计算机视觉领域中的隐藏物体检测，其中目标物体与背景颜色相似，就像息肉与肠道之间的关系一样。

因此，为了克服这些困难，本文借鉴了隐藏物体检测的思想和方法，并将其应用于息肉分割领域。通过分析以往网络的优势和缺陷，本文采用了基于卷积反向注意力模块的息肉分割网络，即 CRANet (Convolution Reverse Attention Network)。该网络利用卷积反向注意力(Convolution Reverse Attention, CRA)、特征金字塔(Feature Pyramid, FP)以及聚合解码器(Aggregation Decoder, AD)来提取息肉的特征信息。与经典的息肉分割网络相比，本文在分割准确性方面取得了显著提升。

此外，本文还注意到神经网络所提取的特征主要集中在背景信息的提取上，很难聚焦到息肉区域。为了解决这一问题，本文提出了 CRA 模块，具体来说就是在反向注意力中加入卷积注意力模块(Convolutional Block Attention Module, CBAM)，该机制在神经网络中得到了广泛应用，在 PFNet^[61]中的定位模块(positioning module, PM)证明了 CBAM 的小目标定位性能。CBAM 能够有选择性地聚焦于关键信息，从而更好地捕捉目标息肉。通过将卷积注意力模块应用于本文的 CRANet 中，就能够有效地引导网络分割出息肉。

CRANet 对于提高早期息肉分割的准确性具有重要意义。CRANet 结合了隐藏物体检测的思想和注意力机制，有效地克服了分割过程中的困难。通过实验证明，CRANet 在分割任务中表现出色，并有望在临床实践中为早期癌症诊断和治疗提供有力支持。未来，将进一步优化和改进这一方法，以期在医学领域取得更大的突破和进展。

3.1 方法

3.1.1 深度学习模型架构

图 3.1 展示了 CRANet（卷积反向注意力网络）的整体架构。CRANet 模型由两个模块组成，即定位模块和识别模块。定位模块利用特征金字塔（Feature Pyramid, FP）和聚合解码器（Aggregation Decoder, AD）来确定息肉的具体位置，而识别模块则引入卷积反向注意力（Convolution Reverse Attention, CRA）模块来细化息肉区域，进一步对息肉进行精细化分割。

Res2Net^[62] 是 ResNet 的拓展升级，Res2Net 可以在更细粒度级别表示多尺度特征。所以本文选择 Res2Net^[62] 作为 CRANet 的主干网络来对息肉图像进行特征提取。因数据集的数据量不足导致深度学习模型训练不充分、泛化能力不够，是深度学习在医学图像领域常见的限制，为缓解这一问题，本文引入了预训练^[63]。在大型的自然图像数据集 ImageNet 上对 Res2Net 进行大量的训练获得强大的视觉知识^[64]，再通过微调的方式来引导深度学习模型对结直肠息肉进行分割。

通过将上述的模块和策略结合在一起，CRANet 模型能够有效地定位和识别息肉区域，克服了数据稀缺性的挑战，并利用预训练的 Res2Net 网络提供了强大的多尺度表示能力。这使得 CRANet 能够在息肉分割任务中获得更有效的特征信息。

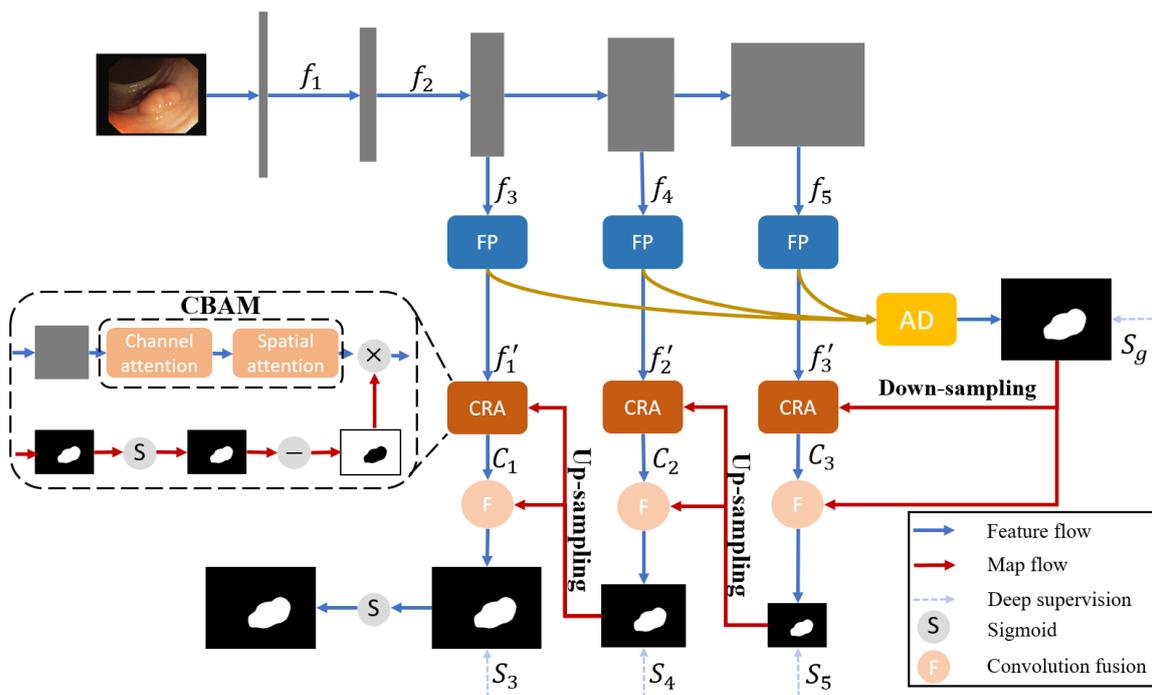


图 3.1 CRANet 模型框架图

接下来, 本文将详细阐述了三个主要模块的细节, 以进一步解释 CRANet 的工作原理和优势。

首先是特征金字塔 (Feature Pyramid, FP) 模块。该模块通过放大的上下文线索捕获细粒度纹理, 从而提供了丰富的多尺度特征。特征金字塔通过在不同的网络层次上提取特征并进行融合, 实现了对图像中不同尺度信息的有效利用。

其次是聚合解码器 (Aggregation Decoder, AD) 模块。该模块能够提供位置信息, 帮助定位模块准确地确定息肉的位置, 得到粗分割结果。聚合解码器通过使用三个解码器分支, 将来自最后的三个高层级的特征进行聚合, 从而提高了定位的准确性和稳定性。

最后是卷积反向注意力 (Convolution Reverse Attention, CRA) 模块。该模块引入了卷积注意力模块, 使得 CRANet 能够精确识别息肉区域。卷积反向注意力模块通过对粗分割的图像进行反向操作, 并将其与特征图进行逐元素相乘, 从而产生更加准确的预测结果。

这些模块相互协作, 通过从更深层细化粗分割结果的方式, 提高了 CRANet 在息肉分割任务中的准确性。

3.1.2 特征金字塔模块

目前, 在神经科学领域已经发现不同大小的感受野可以刺激眼睛中凹的区域。这就意味着不同大小的感受野也就是多尺度的卷积能够捕捉到更有用的信息, 从而提升模型的泛化能力。因此, 在定位阶段使用特征金字塔 (Feature Pyramid, FP) 模型来进行多层次的学习。下面将对 FP 模块进行一个详细的介绍。

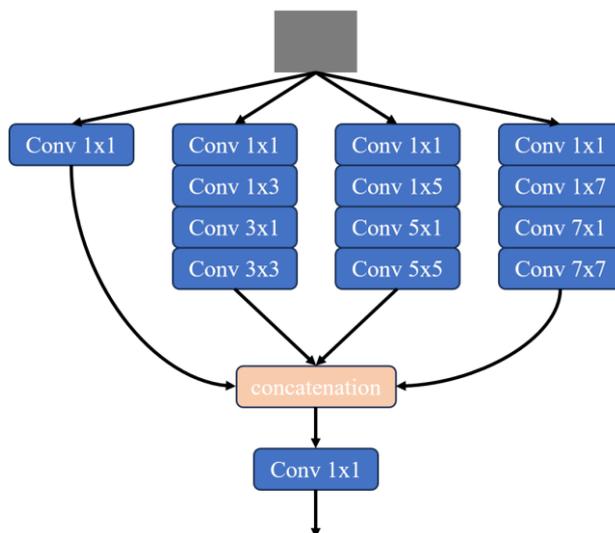


图 3.2 FP 模块框架图

如图 3.2 所示, FP 模块的一共有 4 个通道即 $\{b_i, i=1, 2, 3\}$, 每个通道的卷积核大小分别为 $\{r_i, i=3, 5, 7\}$ 以及卷积核大小为 1×1 卷积的简易通道。在每个通道 b_i 中, 第一个卷积层利用卷积核大小为 1×1 的卷积操作 (Conv 1×1) 减少通道数量。第二个卷积层利用卷积核大小为 $1 \times r_i$ 的卷积操作 (Conv $1 \times r_i$)。第三个卷积层利用卷积核大小为 $r_i \times 1$ 的卷积操作 (Conv $r_i \times 1$)。第四个卷积层利用卷积核大小为 $r_i \times r_i$ 的卷积操作 (Conv $r_i \times r_i$)。然后, 将三个通道 $\{b_i, i=1, 2, 3\}$ 和简易通道拼接起来, 并通过卷积核大小为 1×1 卷积操作将通道进行缩减。最后, 将整个模块使用 ReLU 函数进行激活得到最终的输出特征。

通过构建这样的 FP 模块, 本文能够有效地融合多尺度的高维特征, 以此来捕捉到更丰富的视觉信息。这有助于提高模型在定位阶段的性能, 使其能够更准确地确定息肉的位置。

3.1.3 聚合解码器

虽然目前最先进的分割网络仍然十分依赖于来自编码器的多级特征, 通过对每个层次特征的提取和分析来做出预测。然而在关于级联部分解码器的研究^[65]中发现浅层的特征不仅需要消耗更多的计算资源, 而且还对性能的贡献较小。所以为了提高学习效能, 本文也参考借鉴了这样的思路, 仅考虑聚合后三个高维度的特征。具体而言, 在定位阶段, 本文需要定位难以发现的息肉, 因此从后三个 FP 中获取高维特征进行聚合分析。

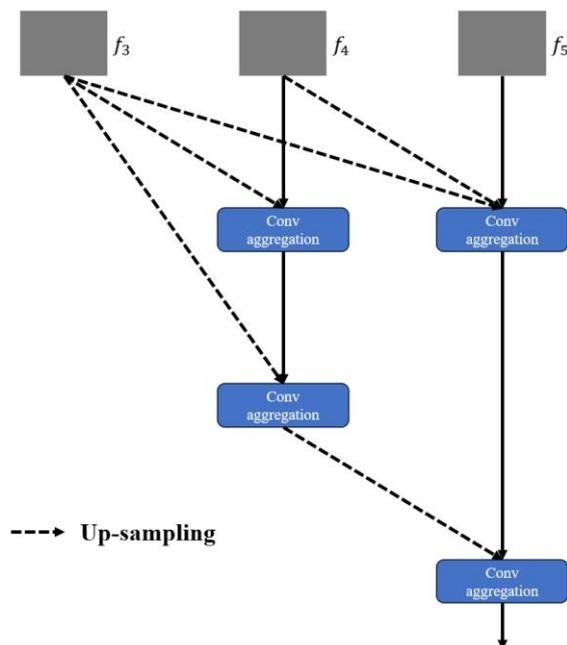


图 3.3 AD 模块框架图

本文采用如图 3.2 所示的聚合解码器来聚合高级特征。首先模型将原始图像输入在 ImageNet 预训练后的 Res2Net，得到对应的五级层次特征。然后，如图 3.2 所示，按照前文提到的聚合策略，通过卷积聚合操作，整合了来自 Res2Net 的后三层高级特征。卷积聚合操作就是先将同一尺度的特征在通道维度上进行拼接，再通过卷积操作进行处理，最后得到相应的特征。本文通过后三层高级特征的特征聚合，最终得到聚合特征。

通过使用聚合解码器，本文能够有效地利用高维特征，而无需过多关注低维特征的计算成本。这样的设计选择既能提高性能，又能保持计算效率。

3.1.4 卷积反向注意力模块

本文的聚合特征器是由最后的三层汇聚而成的，聚合特征器只能定位息肉的相对粗略的位置，忽略了结构和纹理细节。然而，特征金字塔模块只能从已经经过预训练的 Res2Net 模型中提取相应的多维度的特征。为了能够让模型提取出更有价值的特征信息，本文设计了卷积反向注意力（CRA）模块来高效地提取息肉的细节信息。CRA 模块的概述和详细信息见图 3.1 的左侧所示。

CRA 模块的第一行，是注意力机制提取特征信息的过程。其输入是来自于在高维特征 f_i 通过 FP 模块进行加强后得到的多尺度特征 f'_i 。然后，将特征 f'_i 输入到卷积注意力模块来提取显著性信息。其中卷积注意力模块是一种用于 CNN 的简单而有效的注意力模块，具体操作如公式 3-1 所示。

$$CBAM_i = M_{\text{Channel}}(F) \times M_{\text{Spatial}}(F) \quad (3-1)$$

其中 CBAM 是由通道注意力机制和空间注意力机制组合而成。通道注意力机制（Channel Attention Module）是将特征图在空间维度上进行压缩，得到一个一维矢量后再进行操作，具体操作如公式 3-2 所示。空间注意力机制（Spatial Attention Module）是对通道进行压缩，在通道维度分别进行了平均值池化和最大值池化，具体操作如公式 3-3 所示。

$$M_{\text{Channel}}(F) = \sigma \left(MLP(AvgPool(F)) + MLP(MaxPool(F)) \right) \quad (3-2)$$

$$M_{\text{Spatial}}(F) = \sigma(f^{3 \times 3}([AvgPool(F); MaxPool(F)])) \quad (3-3)$$

其中， σ 为 sigmoid 操作， $f^{3 \times 3}$ 表示卷积核的大小。

CRA 模块的第二行，本文应用了反向操作^[66]，针对每一个侧输出 S_i 中检测显著性特

征，都先进行 sigmoid 操作再被 1 减。这样就将背景和息肉的颜色进行反转。反向的操作如公式 3-4 所示。

$$R_i = 1 - \text{Sigmoid}(S_i) \quad (3-4)$$

根据上述的一系列公式，最终得到了卷积反向注意力操作过程，具体来说，就是将经过反向的操作的侧输出与经过卷积注意力模块的显著性特征进行相乘。如公式 3-5 所示。

$$CRA_i = R_i \otimes CBAM_i \quad (3-5)$$

其中 \otimes 是按元素的乘法。

3.1.5 损失函数

本文采用了加权交叉熵（BCE）损失函数用于计算全局损失，加权交并比（IoU）损失函数用于计算局部损失。公式 3-6 中 \mathcal{L} 整合了这两种损失函数。为了通过训练提升 CRANet 的鲁棒性，本文对模型的三个粗分割结果 (S_1, S_2, S_3) 以及对最终精细化分割结果 S_g 进行深度监督。同时，为了更好地处理分割任务中的像素级别的准确性和全局一致性，将四个分割结果通过上采样的方式将尺寸变换到与金标准 G 相同大小来计算 BCE 损失函数和 IoU 损失函数。总损失如公式 3-6 所示。

$$\mathcal{L}_{total} = \sum_{i=1}^3 \mathcal{L}(G, S_i) + \mathcal{L}(G, S_g) \quad (3-6)$$

3.2 实验

3.2.1 实验环境

本文使用了基于 Python 3.10 和 PyTorch 2.0 来对模型进行设计。设置批次大小为 28，并使用 AdamW 优化器来训练模型。模型在一台 Ubuntu 20.04 系统的机器上进行训练。显卡采用的是有 24 GB 显存的 NVIDIA RTX3090 GPU。

3.2.2 数据集介绍

本文选择了多个具有代表性的数据集来评估本文的方法的性能。这些数据集涵盖了

不同来源、不同尺寸和不同注释水平的息肉图像，使得本文能够对算法在各种情况下的表现进行全面评估。本文在五个息肉分割数据集上进行了实验评估，这些数据集分别是 ETIS^[67]、CVC-ClinicDB^[68]、CVC-ColonDB^[69]、EndoScene^[70]和 Kvasir^[71]。其中的 Kvasir 数据集是一个大型的数据集。在表 3.1 中展示了这些数据集的详细信息。

ETIS 数据集包含 196 张不同尺寸的息肉图像以及相应的标注。这个数据集的图像大小和形状多样，涵盖了真实世界中的不同情况。

CVC-ClinicDB 是 MICCAI 2015 结肠镜视频自动息肉检测子挑战赛训练阶段的官方数据集，其中包含了来自 23 位患者的 31 个息肉视频序列，共计 612 张图像。这个数据集提供了丰富的结肠镜图像，其中包含了不同类型和形状的息肉。

CVC-ColonDB 是一个包含了 15 个短视频序列的数据库，每个序列都来自不同的研究，并包括至少一个息肉，共提供了 380 张测试集图像。这个数据集为本文提供了更多不同场景下的息肉图像，有助于评估算法的泛化能力。

EndoScene 是将 CVC-ColonDB 和 CVC-ClinicDB 合并成一个新的数据集，其中包含了来自 36 位患者的 44 个视频序列的 912 张图像。这个数据集通过重新注释旧数据集，提供了更准确和详细的标注信息。

Kvasir-SEG 数据集是一个包含了 1000 张带有注释的息肉图像及其相应掩模的大型数据集。这个数据集有较大的样本规模，为本文提供了更丰富的数据来评估算法的性能，更加全面的验证模型的性能。

表 3.1 息肉分割数据集介绍

数据集名称	图像尺寸	训练集的数量	测试集的数量
ETIS	966×1225		196
CVC-ClinicDB	288×384		62
CVC-ColonDB	500×574	1450	380
CVC-300	500×574		60
Kvasir	1070×1348		100

通过这些多样化的数据集上进行实验评估，本文能够更全面地了解本文提出的模型在不同场景下的表现和鲁棒性。具体而言，本文的训练集由 Kvasir-SEG 中的 900 张图像和 CVC-ClinicDB 的 550 张图像组成，总共 1450 个样本；测试集中包括 Kvasir-SEG 剩余的 100 张图像、CVC-ClinicDB 剩余的 62 张图像、完整的 CVC-ColonDB 数据集（380

个样本)、完整的 ETIS 数据集(196 个样本)以及来自 CVC-300 数据集的 60 张图像(CVC-300 是 EndoScene 的 912 张图像减去 ColonDB 的 612 张图像)。上述的训练测试划分和设置都是依照相关息肉分割工作^{[28][29]}的共识。

3.2.3 评价指标

除了使用平均 Dice 和平均 IoU 这两个指标进行定量评估之外,本文还引入了其他四种测量指标,包括加权 Dice 值 F_B^ω 、MAE、增强对齐度量值 E_ϕ^{max} 和结构度量值 S_α 。

加权 Dice 值 F_B^ω 用于修正 Dice 中的“等重要缺陷”。通过引入加权系数 ω ,可以对不同类型的缺陷赋予不同的重要性,从而更准确地评估模型的性能。

MAE (Mean Absolute Error, 平均绝对误差)用于衡量像素级的精度。通过计算预测结果与真实结果之间的绝对差异的平均值,可以评估模型对每个像素的预测准确程度。

增强对齐度量值 E_ϕ^{max} 是用于评估像素级和全局级的相似性。增强对齐度量值通过计算预测结果与真实结果之间的最大对齐误差来度量两者之间的一致性。

结构度量值 S_α 用于衡量预测结果与真实结果之间的结构相似性。结构度量值通过比较两者之间的结构特征,如边缘、纹理等,来评估模型的性能。

3.2.4 对比实验

本文提出了一种名为 CRANet 的模型,旨在提高息肉分割的性能和准确性。本文在实验中使用了五个公开的息肉数据集,其中包括 ETIS、CVC-ClinicDB、CVC-ColonDB、EndoScene (CVC-T) 和 Kvasir。为了进行全面的对比实验,本文还选取了其他常用的图像分割模型作为比较对象,包括 UNet^[17]、UNet++^[18]、SFA^[22]、PraNet^[28]、CCBANet^[23]、DS-TransUNet^[24]以及 CaraNet^[29]。表 3.2 中的 mDice 代表 mean Dice, mIoU 代表 mean IoU, F_B^ω 代表加权 Dice, S_α 代表结构测量, E_ϕ^{max} 代表增强对齐, MAE 代表平均绝对误差。↑表示越高越好, ↓表示越低越好。

通过表 3.2 的实验结果,本文发现 CRANet 模型在五个公开的息肉分割数据集集中的性能优于以往的网络。尤其是在包含小型息肉的 ETIS 数据集上,CRANet 模型表现也很出色。在五个公开息肉数据集上的结果证明了 CRANet 模型在息肉分割任务中的优越性和鲁棒性。CRANet 模型的成功应用有望促进医学影像领域的进一步发展,并为患者提供

更好的医疗服务。

表 3.2 CRANet 模型在 Kvasir、CVC-ClinicDB、CVC-ColonDB、ETIS 和 CVC-T (EndoScene 测试数据集) 的测试结果。

数据集	模型	mDice \uparrow	mIoU \uparrow	F_B^ω \uparrow	S_α \uparrow	E_ϕ^{max} \uparrow	MAE \downarrow
ETIS	UNet	0.398	0.335	0.366	0.684	0.740	0.036
	UNet++	0.401	0.344	0.390	0.683	0.776	0.035
	SFA	0.297	0.217	0.231	0.557	0.633	0.109
	PraNet	0.628	0.567	0.600	0.794	0.841	0.031
	CCBANet	0.677	0.610	0.640	0.800	0.838	0.028
	DS-TransUNet	0.675	0.592	0.625	0.802	0.859	0.023
	CaraNet	0.747	0.672	0.709	0.868	0.894	0.017
	CRANet	0.784	0.705	0.745	0.864	0.894	0.016
CVC-ClinicDB	UNet	0.823	0.755	0.811	0.889	0.954	0.019
	UNet++	0.794	0.729	0.785	0.873	0.931	0.022
	SFA	0.700	0.607	0.647	0.793	0.885	0.042
	PraNet	0.899	0.849	0.896	0.936	0.979	0.009
	CCBANet	0.909	0.856	0.903	0.939	0.971	0.010
	DS-TransUNet	0.912	0.859	0.908	0.936	0.976	0.007
	CaraNet	0.936	0.887	0.931	0.954	0.991	0.007
	CRANet	0.940	0.894	0.940	0.954	0.985	0.007
CVC-ColonDB	UNet	0.512	0.444	0.498	0.712	0.776	0.061
	UNet++	0.483	0.410	0.467	0.691	0.760	0.064
	SFA	0.469	0.347	0.379	0.634	0.765	0.094
	PraNet	0.709	0.640	0.696	0.819	0.869	0.045
	CCBANet	0.758	0.675	0.736	0.842	0.880	0.042
	DS-TransUNet	0.762	0.682	0.738	0.829	0.872	0.053
	CaraNet	0.773	0.689	0.729	0.853	0.902	0.042
	CRANet	0.791	0.717	0.777	0.855	0.907	0.033

数据集	模型	mDice \uparrow	mIoU \uparrow	F_B^ω \uparrow	S_α \uparrow	E_ϕ^{max} \uparrow	MAE \downarrow
CVC-T	UNet	0.710	0.627	0.684	0.843	0.876	0.022
	UNet++	0.707	0.624	0.687	0.839	0.898	0.018
	SFA	0.467	0.329	0.341	0.640	0.817	0.065
	PraNet	0.871	0.797	0.843	0.925	0.972	0.010
	CCBANet	0.903	0.833	0.881	0.933	0.986	0.001
	DS-TransUNet	0.880	0.798	0.854	0.920	0.978	0.007
	CaraNet	0.903	0.838	0.887	0.940	0.989	0.007
	CRANet	0.913	0.850	0.901	0.940	0.981	0.005
Kvasir	UNet	0.818	0.746	0.794	0.858	0.893	0.055
	UNet++	0.821	0.743	0.808	0.862	0.910	0.048
	SFA	0.723	0.611	0.670	0.782	0.849	0.075
	PraNet	0.898	0.840	0.885	0.915	0.948	0.030
	CCBANet	0.902	0.845	0.887	0.916	0.952	0.032
	DS-TransUNet	0.913	0.857	0.902	0.923	0.963	0.023
	CaraNet	0.918	0.865	0.909	0.929	0.968	0.023
	CRANet	0.920	0.869	0.916	0.930	0.957	0.022

注: mDice, mean Dice; mIoU, mean IoU; F_B^ω , 加权 Dice; S_α , 结构测量; E_ϕ^{max} , 增强对齐; MAE, 平均绝对误差; \uparrow 表示越高越好; \downarrow 表示越低越好。

3.2.5 消融实验

在本文中, 首先对没有 CRA 但有 FP 模块的 CRANet 进行了实验评估。通过表 3.3 的实验结果可以发现, 相比于完整的模型, 有 FP 模块但没有 CRA 模块情况下, 模型的性能在五个数据集上没有急剧下降, 这证明了 FP 模块在息肉识别方面的有效性。尤其值得注意的是, 在 ETIS 数据集上, mDice 指标的表现优异。这个结果有力地证明了 FP 模块可以帮助模型适应不同大小的感受野, 从而提升了模型的整体性能。

其次, 对有 CRA 模块但没有 FP 模块的 CRANet 进行了实验评估。根据表 3.3 的结果显示, 有 CRA 模块但没有 FP 模块情况下, 在五个公共数据集上也没有急剧下降。而且, 相比于没有 CRA 但有 FP 模块的情况, 有 CRA 模块但没有 FP 模块的模型性能表现更为优

异，说明 CRA 模块对模型的性能提升程度比 FP 模块好。

总的来说，CRA 模块和 FP 模块在 CRANet 中的引入都对模型的性能有着积极的影响。CRA 模块能够提高息肉的识别能力，而 FP 模块则能够适应不同大小的感受野，从而进一步提高了模型的性能。验证了 CRANet 方法在自动息肉分割任务中的有效性和优越性。

表 3.3 CRANet 模型在 Kvasir、CVC-ClinicDB、CVC-ColonDB、ETIS 和 CVC-T (EndoScene 测试数据集)的消融实验

模型	FP	CRA	mDice \uparrow	mIoU \uparrow	F_B^ω \uparrow	S_α \uparrow	E_ϕ^{max} \uparrow	MAE \downarrow
ETIS	\sqrt	\times	0.759	0.665	0.707	0.844	0.895	0.016
	\times	\sqrt	0.772	0.695	0.745	0.864	0.884	0.012
	\sqrt	\sqrt	0.784	0.705	0.745	0.864	0.894	0.016
CVC-ClinicDB	\sqrt	\times	0.925	0.875	0.920	0.942	0.976	0.008
	\times	\sqrt	0.933	0.887	0.935	0.949	0.980	0.007
	\sqrt	\sqrt	0.940	0.894	0.940	0.954	0.985	0.007
CVC-ColonDB	\sqrt	\times	0.759	0.688	0.749	0.835	0.870	0.038
	\times	\sqrt	0.761	0.680	0.738	0.831	0.876	0.004
	\sqrt	\sqrt	0.791	0.717	0.777	0.855	0.907	0.033
CVC-T	\sqrt	\times	0.890	0.821	0.875	0.928	0.967	0.007
	\times	\sqrt	0.906	0.837	0.888	0.932	0.979	0.006
	\sqrt	\sqrt	0.913	0.850	0.901	0.940	0.981	0.005
Kvasir	\sqrt	\times	0.894	0.842	0.891	0.909	0.945	0.029
	\times	\sqrt	0.916	0.867	0.910	0.921	0.958	0.024
	\sqrt	\sqrt	0.920	0.869	0.916	0.930	0.957	0.022

注：mDice, mean Dice; mIoU, mean IoU; F_B^ω , 加权 Dice; S_α , 结构测量; E_ϕ^{max} , 增强对齐; MAE, 平均绝对误差; \uparrow 表示越高越好; \downarrow 表示越低越好。

3.3 讨论

本文介绍了一种基于深度学习的分割模型，名为 CRANet，旨在提高自动息肉分割任务的性能。CRANet 通过融合卷积反向注意力（CRA）模块、特征金字塔（FP）模块以及

聚合解码器 (AD) 模块, 增强了分割的准确性和鲁棒性。实验证明, CRANet 在各项评价标准上的整体表现优于其他经典模型。相较于经典网络 PraNet 和 CaraNet, CRANet 提供了更精确的预测。然而, 尽管 CRANet 在医学图像分割任务上取得了较高的性能, 但在模型的细节方面还有提升的空间。

首先, 本文对输出的粗分割结果仅使用简单的插值上采样进行操作。这样的操作可能会丢失许多图像中的重要信息, 一些边界信息可能会因此模糊, 从而导致模型分割不够细致。所以在未来的研究中, 可以更换为计算量较大的反卷积来进行上采样。

其次, 本文的 CRANet 的主干网络 Res2Net 是在自然图像集的 ImageNet 上进行预训练的, 而 ImageNet 中的自然图像与医学图像的分布是不一致的, 存在一定的差异。未来的工作可以考虑使用针对医学图像的预训练模型, 作为骨干网络, 以进一步提高 CRANet 在医学图像分割中的准确性和稳定性。

最后, 随着三维医学图像的广泛应用, 对于三维医学成像分割的需求也越来越迫切。因此, 本文计划未来对 CRANet 进行改进, 以使其适用于三维医学成像分割任务, 并构建出 3D 版本的 CRANet。这将有助于实现对 CT 或 MRI 图像等三维医学数据的准确分割, 进一步提高分割结果的质量和准确性。

CRANet 作为一种基于深度学习的息肉分割模型, 在医学图像分割中展现出显著的改进。然而, 仍需在模型优化方面进行改进, 并探索更适用于医学图像的预训练模型, 以进一步提高分割准确性。此外, 将 CRANet 扩展到三维医学图像分割是未来工作的重点, 以满足日益增长的需求和挑战。

3.4 本章小结

本文提出了一种新的神经网络模型, 名为 CRANet, 用于在医学图像中进行息肉分割任务。传统的计算机视觉方法在处理医学图像分割时存在一定的局限性, 而基于深度学习的方法在这个领域取得了显著的进展。CRANet 模型创新性地融合了卷积反向注意力 (CRA) 模块和特征金字塔 (FP) 模块, 以提高息肉分割的性能。

CRA 模块是 CRANet 的关键组成部分, 通过在不同尺度上对特征图进行注意力调整, 有助于精细化地提取息肉的边界和细节信息。这种注意力机制使得 CRANet 能够更准确地捕捉到息肉的形状和结构, 从而提高了分割的准确性。另一个关键组件是 FP 模块, 通过在网络内部构建多层特征金字塔来实现对不同大小目标的检测适应性。这种金字塔

结构使得 CRANet 能够在不同尺度上对息肉进行有效的分割，并获得良好的结果。通过对 CRANet 进行广泛的实验验证，CRANet 整体表现优于经典的模型。这证明了 CRANet 在息肉分割任务中的性能。

总的来说，CRANet 模型通过利用 CRA 模块、FP 模块以及 AD 模块，提高了医学图像中息肉分割的性能。CRANet 模型在整体分割精度和处理隐匿区域方面都表现出色，为医学图像分割领域的研究和应用提供了有价值的贡献。在未来的工作中，进一步优化 CRANet 模型，并将其扩展到三维医学图像分割领域，将会是一个有意义的方向。上述改进将帮助本文在 CT 或 MRI 图像中更准确地进行分割，提高分割结果的质量和准确性。通过持续的努力和研究，CRANet 有望在医学图像分割领域取得更大的突破。

4 基于 mean teacher 的半监督息肉分割网络

在先前提到的基于卷积反向注意力模块的息肉分割网络的实验中，是假设所有的结肠镜图像都有分割掩码，这是一种全监督的条件。然而，在实际情况中，医学专家对结肠镜图像进行分割是一项繁琐且费时的任务，因此大部分结肠镜图像并没有相应的分割掩码。为了充分利用这些没有分割掩码的结肠镜图像，本文考虑采用半监督方法进行研究。因此，本文提出了一种基于 mean teacher 的半监督息肉分割网络，旨在使用半监督的方式来对没有分割掩码的结肠镜图像进行充分地利用。

模型的主体框架是 mean teacher 网络。mean teacher 网络是一种基于网络蒸馏的方法，通过训练一个主网络和一个移动平均参数的辅助网络来提高网络的泛化性能。主网络和辅助网络在训练过程中相互交互，主网络也就是 student 模型，通过监督数据进行更新，而辅助网络就是 teacher 模型，则通过未标记数据进行更新。这种交互训练可以帮助网络更好地利用未标记数据，并提供更准确的分割结果。

与此同时，基于 mean teacher 的半监督息肉分割网络结合了数据变换的正则化和不确定性感知。相比目前常用的半监督网络，具有更好的性能。数据变换的正则化是一种常用的半监督学习方法，通过对输入数据进行随机变换，增加了模型对输入数据的鲁棒性，提高了泛化能力。而不确定性感知则是指网络可以对自身的预测结果产生一定的不确定性估计，从而能够更好地处理没有标签的数据。

通过将这些方法结合起来，本文的半监督息肉分割网络在利用未标记数据方面具有优势。mean teacher 的框架可以充分利用未标记数据进行训练，提高网络的泛化能力。数据变换的正则化可以增加模型的鲁棒性，降低过拟合的风险。不确定性感知可以提供更准确的不确定性估计，帮助网络对未标记数据进行可信度的评估。

总的来说，本文提出的基于 mean teacher 的半监督息肉分割网络通过结合数据变换的正则化和不确定性感知，充分利用未标记数据，优于当前常用的半监督网络。这种网络结构有望在医学图像分割领域取得更好的性能，提高结肠镜图像的分割准确度，并减轻医学专家的工作负担。未来的工作可以进一步优化和扩展这个半监督网络，探索更多的半监督学习方法，以进一步提高息肉分割的性能和效果。

4.1 方法

4.1.1 mean teacher 模型框架

本文的网络模型是在 mean teacher 模型的基础上进行设计, 因此首先需要对 mean teacher 模型进行详细介绍。在 mean teacher 模型引起轰动之前, Temporal Ensembling 一直是主流的方法。Temporal Ensembling 是对 Π 模型的升级和改进。 Π 模型通过两个损失函数充分利用未标记数据。第一项使用两种不同且随机的数据增强方式对未标记数据进行处理, 并使用均方差计算损失函数。第二项计算已标记数据通过模型得到的预测值与标签之间的交叉熵损失。然而, 这需要进行多次前向传播, 耗时且计算量大, 而且 Π 模型的反向传播也存在不确定性。为了解决这些问题, 引入了自集成的思想, 提出了时序集成的概念, 即 Temporal Ensembling。时序集成将一次前向传播得到的预测值视为过去若干次前向传播预测值的集成。这种方法显著加快了训练速度, 但也增加了空间使用率, 是以空间换时间的策略。

基于对 Temporal Ensembling 的考虑和分析, mean teacher 模型采用平均模型权重来代替平均模型预测值。平均模型权重通常能够产生比直接使用最终权重更准确的模型。在 mean teacher 模型中, teacher 模型使用 student 模型经过 EMA 处理后的权重, 而不是与 student 模型共享权重。这样, mean teacher 模型可以在每个批次训练完成后及时地整合先前批次中的信息, 而不仅仅在每轮训练。此外, 由于 EMA 不仅仅提高了最后一层的输出, 还提高了所有层的输出质量, 因此模型能够更好地展现中间层甚至高层的特征信息。

接下来, 本文对 mean teacher 模型的具体训练过程进行分析。首先, mean teacher 模型通过将 teacher 模型的权重设置为 student 模型的 EMA 后的权重来替代平均模型的预测值。这样做的好处是, EMA 权重可以减少模型的不稳定性, 使模型更加平滑并具有鲁棒性。通过使用 EMA 权重, mean teacher 模型能够在训练过程中及时地整合先前批次中的信息, 从而提高模型的性能。其次, mean teacher 模型不仅仅关注最后一层的输出, 还关注中间层和高层的特征表示。这是因为 EMA 权重不仅仅影响最后一层的预测结果, 还影响了整个模型的中间层表示。通过使用 EMA 权重, mean teacher 模型能够更好地利用这些中间层表示, 从而提高模型的性能和泛化能力。此外, mean teacher 模型在学习过程中使用了两个损失函数: 一个用于监督学习的有标签数据, 另一个用于无标签数据

的自监督学习。这种组合的损失函数能够充分利用所有可用的数据，并促使模型在有标签和无标签数据上都取得良好的表现。最后，mean teacher 模型在半监督学习任务中表现出很好的适应性。通过使用 EMA 权重和多任务损失函数，结合有标签和无标签数据的信息，mean teacher 模型提高了模型的性能和泛化能力。

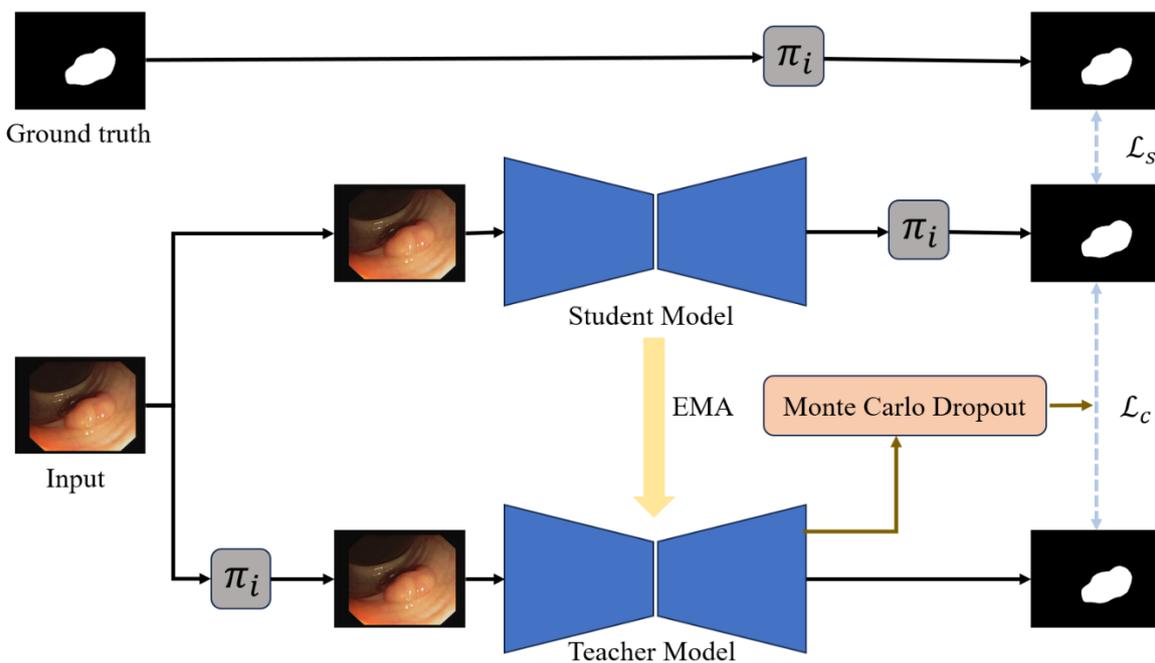


图 4.1 模型框架

本文提出的半监督模型是基于 mean teacher 模型构建的，如图 4.1 所示，结合了基于数据变换的正则化和不确定性感知。接下来，本文将详细介绍上述两个模块。

4.1.2 基于数据变换的正则化

在经典的半监督网络中，数据正则化是利用未标记数据的常见方法。本文引用了一种正则化方法，以此来充分地发挥未标记数据的潜力。在 mean teacher 网络中，该方法鼓励模型在使用不同正则化条件下，即在不同的数据变换的方式下，对相同输入进行一致化的预测，然后通过一致性损失函数对模型进行迭代更新。

在医学分割领域中，对输入图像进行数据变换时，例如对称，放缩等操作，期望的预测也是应该以相同的方式进行数据变换。然而，在实际情况中，CNN 在输入图像进行数据变换时，对应的网络预测并不会以相同的方式进行变换^[72]。为了解决这个问题，本文借鉴了以往的思路，引入了一个无监督损失。具体来说，是通过比较网络在相同输入的不同数据变换下的预测之间的差异。

在半监督学习中，常用的正则化方法都能在一般的医学分类任务中表现出较好的性

能。然而，在医学图像领域，合理应用正则化方法使得模型性能得到相应的提升是一个极具挑战性的任务。这两类任务之间的一个显著区别在于，对于分类问题，输入数据的变换（如平移、对称和放缩）并不会使得分类输出产生相应的变换。相对而言，对于分割问题，输入数据的变换将会对分割结果产生相应的变换。简言之，分类问题是输入变换输出不变的，而分割任务是输入变换输出等变的。

在医学图像分割中，这种现象的存在会严重制约随机数据变换的无监督正则化效果^[34]。为了更有效地将未标记数据和标记数据结合使用进行分割任务并增强正则化的效果，如图 4.1 所示，本文引入了转换一致方案作为半监督正则化。

具体而言，该方案通过在输入和输出同时进行数据变换来实现。对于一个训练输入 x_i ，在对 teacher 网络进行训练时，将数据扰动 π_i 应用于输入图像。而在对 student 网络进行训练时，将数据扰动 π_i 应用于预测图像。为了保证预测和标签的一致，还要将同样的数据扰动 π_i 应用于标签 y_i ，并通过标准的交叉熵损失进行迭代更新。二元交叉熵损失函数为下述 4-1 公式：

$$\mathcal{L}_s(f, y) = -\frac{1}{N} \sum_{i=1}^N y_i \cdot \log f_i + (1 - y_i) \cdot \log(1 - f_i) \quad (4-1)$$

公式 4-1 中 N 是全部体素， i 是单个体素， y_i 是金标准， f_i 是 student 模型的预测结果。

4.1.3 不确定性感知

本文还引入了不确定性感知思路来改进正则化，原本的任务是从 3D MR 图像中分割左心房，但是由于其性能的优越性和即插即用的便利性，本文将其引入到 2D 图像中分割息肉。

不确定性感知方法基于 mean teacher 框架。如果未标记的输入数据不提供基本标签，而是仅靠网络自身进行学习预测生成伪标签。这种情况下 teacher 模型生成的伪标签可能是不可靠的，从而让 student 模型学到了很多不可靠的标签，使得 student 模型的性能不佳。为了解决这个问题，本文参考了以往的思路，引入不确定性感知^[73]，通过对 teacher 模型的不确定性信息进行估计，从而逐步从有意义且可靠的标签中进行学习。也就是说，除了生成目标输出之外，teacher 模型还对每个目标预测进行不确定性程度的估计。基于估计得到的不确定性程度，通过设定相应的阈值来过滤掉不可靠的预测并

仅选择可靠的预测作为 student 模型的目标来计算一致性损失。这样的策略可以为 student 模型提供了更可靠的监督,同时也鼓励 teacher 模型生成更高质量的目标预测。所以对于一个批次的训练图像, teacher 模型不仅需要生成目标预测,还需要进一步估计每个目标的不确定性程度。然后使用一致性损失来优化学生模型,使其在不确定性的指导下训练学习可信目标。

本文参考以往研究的经验,使用蒙特卡洛 Dropout 来近似估计不确定性^[74]。如图 4.1 所示,对于每一个批次的输入图像来说, teacher 模型要执行 M 次的随机前向传播。然后,针对每一个体素,模型可以得到一组相应的概率向量。具体来说,概率向量可以表示为 $\{p_t\}_{t=1}^T$ 。同时本文使用预测熵来实现蒙特卡洛 Dropout 的操作。预测熵可以概括为下述 4-2 公式:

$$\mu_c = \frac{1}{M} \sum_t p_t^c \quad \text{and} \quad \mu = - \sum_c \mu_c \log \mu_c \quad (4-2)$$

公式 4-2 中 p_t^c 为在 t 次预测时第 c 类出现的概率。最终将这些概率全部汇总起来,可以构造一个不确定张量 U , $\{\mu\} \in \mathbb{R}^{H \times W \times D}$ 。在 U 的指导下,模型将根据设定的阈值对不可靠的预测进行剔除,从而让 student 模型可以训练到更为可靠的标签。不确定性感知一致性损失 \mathcal{L}_c 的公式为:

$$\mathcal{L}_c(f', f) = \frac{\sum_v \mathbb{I}(\mu_v < S) \|f'_v - f_v\|^2}{\sum_v \mathbb{I}(\mu_v < S)} \quad (4-3)$$

公式 4-3 中 $\mathbb{I}(\cdot)$ 为指标函数; μ_v 为第 v 个体素的估计不确定度 U ; S 是过滤不确定预测的阈值; f'_v 和 f_v 分别是 teacher 模型和 student 模型在第 v 个体素的预测结果。通过本文在训练过程中不确定性感知的一致性损失, student 模型和 teacher 模型都可以学习到更可靠的知识,从而降低模型的整体不确定性。

4.1.4 损失函数

本文的损失函数由两部分构成,分别是二元交叉熵和均方误差。总损失函数表示如下:

$$\mathcal{L}_{total} = \mathcal{L}_s(f, y) + \mathcal{L}_c(f', f) \quad (4-4)$$

公式 4-4 中, f 是 student 模型的预测结果, f' 是 teacher 模型的预测结果, y 是金标准。

4.2 实验

4.2.1 实验环境

本文使用了基于 Python 3.10 和 PyTorch 2.0 来对模型进行设计。设置批次大小为 50，并使用 AdamW 优化器来训练模型。训练模型的设备系统是 Ubuntu 20.04。显卡采用的是有 24 GB 显存的 NVIDIA RTX3090 GPU。

4.2.2 数据集介绍

本文在两个息肉分割数据集上进行了广泛的测试，数据集包括 CVC-ClinicDB 和 HyperKvasir^[75]。

CVC-ClinicDB 是 MICCAI 2015 结肠镜视频自动息肉检测子挑战赛训练阶段的官方数据集。该数据集包含了来自 23 例患者的 31 个息肉视频序列，总共包含 612 张图像。每张图像都附带有相关的息肉和背景分割掩码。CVC-ClinicDB 是一个被广泛使用的数据集，用于评估结肠镜图像和视频分割任务的性能。

另一个数据集是 HyperKvasir，该数据集是当前可用的最大的胃肠道图像和视频数据集之一。该数据集收集自挪威 Bærum 医院的真实胃镜和结肠镜检查，并且部分数据由经验丰富的胃肠内窥镜医生进行了标记。HyperKvasir 数据集包含了惊人的 110,079 张图像和 374 个视频，涵盖了解剖标志、病理和正常的发现。总体而言，图像和视频帧的数量约为 100 万。HyperKvasir 数据集有规模大以及涵盖面广的特点这使其成为训练息肉分割模型的理想选择。值得注意的是，本文只使用了 HyperKvasir 数据集中的 1000 个带有标签的分割图像进行训练和评估。与此同时，本文还利用了该数据集中 99417 个没有标签的数据，以实现半监督学习的目标。

通过在 CVC-ClinicDB 和 HyperKvasir 数据集上进行广泛的测试，本文能够评估本文的模型在不同场景下的表现，并验证其在实际医学图像分割任务中的有效性。这些数据集的多样性和规模为本文提供了丰富的信息，以更好地理解 and 评估本文的方法。

4.2.3 评价指标

本文对分割性能进行评估时，使用了几个常见的指标，包括 Jaccard 指数 (JA)、Dice 系数 (DI)、像素精度 (AC)、灵敏度 (SE)。

Jaccard 指数 (JA)，也称为 IoU (Intersection over Union)，是一种评估分割结果准确性的度量。Jaccard 指数通过计算预测分割结果与真实分割结果的交集和并集之间的比值来衡量两者之间的重叠程度。Jaccard 指数的取值范围为 0 到 1，值越接近 1 表示预测结果与真实结果的重叠程度越高，分割性能越好，具体操作如公式 4-5 所示。

$$JA = \frac{TP}{TP + FP + FN} \quad (4-5)$$

Dice 系数 (DI) 也是一种常用的分割性能度量指标，与 Jaccard 指数密切相关。Dice 系数通过计算预测分割结果和真实分割结果的重叠区域的两倍与两者的总像素数之间的比值来评估分割的准确性。Dice 系数的取值范围也是 0 到 1，值越接近 1 表示分割结果越准确，具体操作如公式 4-6 所示。

$$DI = \frac{2 \times TP}{2 \times TP + FN + FP} \quad (4-6)$$

像素精度 (AC) 是衡量分割结果准确性的另一个重要指标。像素精度计算预测分割结果中正确分类的像素数与总像素数之间的比值。由于像素精度包括前景像素 (息肉) 和背景像素，而且前景像素的个数远少于背景像素，所以像素精度一直维持在较高水平，并且模型之间精度较为接近。具体操作如公式 4-7 所示。

$$AC = \frac{TP + TN}{TP + FP + TN + FN} \quad (4-7)$$

灵敏度 (SE) 是用于衡量二分类问题中分类器性能的常用指标。在分割任务中，本文可以将分割结果与真实结果进行二值化，然后计算分割结果中正样本 (如肿瘤区域) 的灵敏度。灵敏度表示分类器正确检测到正样本的能力，指标的取值范围也是 0 到 1，值越接近 1 表示分类器性能越好，具体操作如公式 4-8 所示。

$$SE = \frac{TP}{TP + FN} \quad (4-8)$$

上述公式中的 TP、FP、TN、FN 均是由混淆矩阵得到。其中 TP 是被模型预测为正类的正样本，TN 是被模型预测为负类的负样本，FP 是被模型预测为正类的负样本，FN 是被模型预测为负类的正样本

通过使用这些指标，本文能够全面评估分割算法的性能，了解其在不同方面的表现。这些指标在医学图像分割领域被广泛使用，并能够提供对分割结果质量的定量评估，帮

助研究人员进行性能比较。

4.2.4 对比实验

为了验证模型在半监督学习设置下的性能，本文首先在 CVC-ClinicDB 数据集上评估了不同标记量对模型性能的影响。表 4.1 展示了在标记量为 5% (28)、10% (55)、20% (110)、30% (165)、40% (220) 和 50% (275) 时进行的训练和测试结果。其中，“Supervised”表示在相应的标记量下进行有监督训练。通过观察结果，本文可以看出，在标记量逐渐增加时，模型的性能稳步提升。然而，在标记量为 50% (275) 时，性能趋于稳定，本文认为这可能与标记量的阈值有关。

表 4.1 在 CVC-ClinicDB 数据集上的不同标记量的对比实验

标记率	监督方式	DI	JA	AC	SE
5%	Supervised	0.437	0.348	0.918	0.431
	Ours	0.801	0.724	0.974	0.819
10%	Supervised	0.524	0.423	0.933	0.496
	Ours	0.835	0.764	0.983	0.855
20%	Supervised	0.631	0.527	0.957	0.620
	Ours	0.865	0.798	0.985	0.867
30%	Supervised	0.678	0.579	0.963	0.660
	Ours	0.875	0.812	0.986	0.879
40%	Supervised	0.705	0.612	0.939	0.679
	Ours	0.890	0.833	0.985	0.897
50%	Supervised	0.745	0.649	0.970	0.723
	Ours	0.887	0.833	0.984	0.889

注：Ours，基于 mean teacher 的半监督息肉分割网络；DI，Dice 系数；JA，Jaccard 指数；AC，像素精度；SE，灵敏度；由于息肉区域和背景区域的比例不平衡，所以 AC（准确率）一直维持较高水平。

通过这一实验，本文可以得出结论随着标记量的增加，模型的半监督学习性能得到了提升，但在达到一定阈值后，性能的提升速度减缓并且出现趋于稳定的迹象。这可能是因为在标记量达到一定程度后，额外的标记数据对模型的学习贡献逐渐减少，甚至可

能引入额外的噪声。因此，在确定标记量时需要权衡资源成本和性能提升之间的关系，以便找到最佳的标记量设置。

表 4.2 在 CVC-ClinicDB 数据集上用标记量为 10%的经典方法对比试验

模型	DI	JA	AC	SE
Supervised	0.524	0.423	0.933	0.496
Cross Pseudo Supervision	0.664	0.578	0.959	0.650
Deep Adversarial Networks	0.709	0.626	0.903	0.668
Entropy Minimization	0.689	0.609	0.966	0.659
Interpolation Consistency Training	0.743	0.665	0.938	0.737
Mean Teacher	0.747	0.667	0.952	0.723
Ours	0.835	0.764	0.983	0.855

注：Ours，基于 mean teacher 的半监督息肉分割网络；DI，Dice 系数；JA，Jaccard 指数；AC，像素精度；SE，灵敏度；由于息肉区域和背景区域的比例不平衡，所以 AC（准确率）一直维持较高水平。

表 4.3 HyperKvasir 数据集上的对比试验

模型	DI	JA	AC	SE
Supervised	0.697	0.591	0.920	0.682
Cross Pseudo Supervision	0.786	0.702	0.947	0.774
Deep Adversarial Networks	0.755	0.676	0.926	0.762
Entropy Minimization	0.816	0.735	0.944	0.820
Interpolation Consistency Training	0.784	0.703	0.938	0.778
Mean Teacher	0.797	0.715	0.942	0.806
Ours	0.860	0.799	0.963	0.898

注：Ours，基于 mean teacher 的半监督息肉分割网络；DI，Dice 系数；JA，Jaccard 指数；AC，像素精度；SE，灵敏度；由于息肉区域和背景区域的比例不平衡，所以 AC（准确率）一直维持较高水平。

其次，本文用了在五个经典的半监督算法在 CVC-ClinicDB 数据集上用标记量为 10%（55）对比本文的模型，如表 4.2 所示，本文的模型相比于以往的模型也达到了不错的性能。

最后, 本文在一个大型息肉数据集 HyperKvasir 上进行进一步的模型性能测试。同样, 本文的模型相比于以往的模型和全监督模式, 本文所提出的方法在各个指标都有所提升, 如表 4.3 所示。

4.2.5 消融实验

本文首先对没有基于数据变换的正则化进行了实验评估。其次, 对没有不确定性感知进行了实验评估。本文在 CVC-ClinicDB 数据集上, 使用标记量为 10% 的情况进行实验, 从表 4.4 可以看出, 两大模块都发挥了相应的作用。

表 4.4 消融试验

	DI	JA	AC	SE
With regularization	0.815	0.747	0.980	0.830
With uncertainty	0.814	0.742	0.978	0.847
Ours	0.835	0.764	0.983	0.855

注: Ours, 基于 mean teacher 的半监督息肉分割网络; DI, Dice 系数; JA, Jaccard 指数; AC, 像素精度; SE, 灵敏度; 由于息肉区域和背景区域的比例不平衡, 所以 AC (准确率) 一直维持较高水平。

4.3 讨论

全监督的深度学习方法已在医学图像领域的多个任务中证明了其高效性。然而, 这些方法在很大程度上依赖于具有大量标注数据的可用性。然而, 医学图像领域的标注工作既专业又繁琐, 这严重限制了该领域的发展。因此, 开发能够利用有限标注数据的新方法对于推进医学图像在实际临床应用中的应用具有重要意义。本文的主要目标是开发一种半监督学习方法, 用于息肉分割, 该方法有潜力利用大量未标注数据来减轻标注工作的负担。本文通过在 CVC-ClinicDB 数据集和 HyperKvasir 数据集上进行实验证明了该方法的有效性。

首先需要注意的是, 目前本文的模型仅适用于二维数据。由于基于数据变换的正则化方法的存在, 二维和三维数据的表示方式是不同的。然而, 医学图像数据包括二维平面扫描和三维体积数据, 因此这种二维性的限制可能会带来一定的局限性。对于三维半监督学习来说, 考虑到三维体积数据的自然属性来研究具体的方法可能是一个非常前景

的科研方向。因此，未来的研究可以将模型发展成为一个通用的二维和三维框架。

其次，本文的实验的基础是建立在数据的分布是相同的。然而，在实际临床中，标记和未标记的数据之间可能是在不同的域，而且数据分布不一致。一般来说，当未标记数据集包含分布外的样本时，半监督学习方法的性能会显著下降。然而，目前针对未标记数据和标记数据之间的域偏移问题，在医学图像中还未得到妥善地解决。因此，在未来，本文将探索解决域偏移的思路。

最后，本文在获取模型分割结果时使用了医学图像中常用的技术，如填充和上采样。然而，这些操作可能会引入一定的不确定性。因此，学习如何最小化由这些转换引起的输出差异将使网络具有转换一致性，并实现网络的正则化。虽然本文对图像进行了相应的数据扰动，但是还没考虑到这些内部具体的数据变化。这还需要后面对正则化进行进一步的研究。

通过解决这些问题，本文可以进一步提升半监督学习方法在息肉分割任务中的性能，并为医学图像领域的发展做出更大的贡献。在未来的研究中，本文将努力克服当前模型的局限性，使其能够处理更广泛的数据类型。

4.4 本章小结

本文提出了一种基于 mean teacher 的半监督方法。该方法综合运用了基于数据变化的正则化和不确定性感知，以提高半监督模型的稳定性和分割结果的质量。本文采用了一种简单而有效的正则化策略，通过训练模型在不同干扰下还能对相同输入的分割结果保持一致，以此来充分地发挥未标记数据的潜力。同时，本文还探索了模型的不确定性，以进一步提升目标的分割质量。

与其他经典的半监督方法进行了比较后，本文验证了该方法的有效性。在大型息肉数据集 HyperKvasir 上进行的实验证明了本文方法的优越性和可行性。通过与其他方法的比较，本文的方法在分割性能和鲁棒性方面表现出显著的优势。

在未来的研究中，本文将进一步改进和扩展该方法。首先，本文将探索更多复杂的正则化策略，以进一步提高模型的稳定性和泛化能力。其次，本文将进一步探索模型的不确定性，以更准确地评估分割结果的可靠性。此外，本文还将考虑适用于三维结肠镜图像的扩展，以应对更多实际临床场景下的挑战。

通过这些改进和扩展，本文可以进一步提升该方法在结肠镜图像分割任务中的性能，

并为临床医学图像分析领域的发展做出更大的贡献。相信本文的模型将为减轻医学图像标注负担、提高自动分割算法的效果和实用性方面带来积极的影响。

5 总结与展望

5.1 总结

结直肠癌是全世界第三常见的癌症，约占所有癌症病例的 10%。结直肠癌也是全世界癌症相关死亡的第二大原因。结肠镜检查是结直肠癌筛查的主要手段。医生通过结肠镜检查的手段可以通过影像信息对患者的肠道进行诊断，这其中最重要的就是寻找息肉。息肉是结直肠癌早期的发病主要原因之一，所以准确的分割出息肉，来帮助医生进行诊断，可以预防结直肠癌的发生。在实际的操作中，消化科的息肉分割工作量巨大而且繁琐，所以利用深度学习技术开发息肉分割的模型就变得尤为重要。

本文从全监督网络和半监督网络的角度出发对结肠镜图像的息肉分割进行研究。首先，为了更加全面地剖析本文的工作，本文从结直肠息肉介绍、结肠镜介绍以及结直肠息肉自动分割进行了阐述。其次，为了了解各方面研究在息肉分割的进展，本文对基于机器学习的息肉分割研究现状、基于深度学习的息肉分割研究现状以及半监督息肉分割研究现状进行了详细的论述。然后，为了更好地了解本文的工作，本文通过对深度学习、半监督学习以及注意力机制的理论以及基础模型的介绍来对本文的工作进行铺垫。本文具体的研究工作如下：

(1) 本文在息肉分割全监督方面，开发了基于卷积反向注意力模块的息肉分割网络。由于小型息肉以及息肉的隐匿性对息肉分割造成了阻碍，本文在这里参考了伪装物体检测的思路，并将其应用到息肉分割。具体来说，本文通过定位和识别模块来对息肉进行分割。定位模块包括特征金字塔 (Feature Pyramid, FP) 和聚合解码器 (Aggregation Decoder, AD)。识别模块是卷积反向注意力 (Convolution Reverse Attention, CRA) 模块。通过本文的对比实验和消融实验，可以发现本文提出的模型在性能上具有优越性。

(2) 本文在息肉分割的半监督方面，开发了基于 mean teacher 的半监督息肉分割网络。由于在实际情况下，息肉分割难以获得完备的标签，因此使用半监督的方式来利用无标签的数据。本文结合了基于数据变换的正则化和不确定性感知。在常规数据集 CVC-ClinicDB 和大型数据集 HyperKvasir 上，本文的方法均取得了不错的性能。

通过本文的研究工作，本文得出结论的是在结肠镜图像的息肉分割任务中，全监督和半监督网络均具有潜力，可以提高结直肠息肉分割的准确性和效率，为医生提供更好的辅助诊断工具。未来的研究可以进一步探索更先进的深度学习模型和算法，以进一步

提升结直肠癌的早期诊断准确性以及后续治疗效果。

5.2 展望

近一年来，随着大模型的流行，许多传统的模型难以崭露头角。在科技飞速进步的今天，或许强大的算力才是模型性能提升的关键。智慧医疗领域也在逐步发展壮大，并且涌现出更多的技术和研发。这种技术的蓬勃发展有助于缓解医疗资源短缺的瓶颈。因此，将基础性和普遍性的医疗手段进行智能化可能成为推动医疗发展的重要方向之一，其中包括医学图像分割任务。

医学图像分割任务是一项专业性极强且繁琐的医疗工作。如果能够实现高度自动化，将大大提升医疗的效率，并提升整体医疗水平。但目前仍存在一些挑战需要在未来的研究中突破。以下是一些可能的研究方向：

(1) 收集和开发大规模的数据集：数据是科研的生命。正如李飞飞团队公布的 ImageNet 数据集极大地推动了自然图像的发展一样，医学图像领域也需要各位专家学者的共同努力，早日开发出大规模医学图像数据集，从而推动医疗的发展。针对结直肠息肉的数据，要与多方医院进行全面的合作，结肠镜检查是在各大医院非常常见的检查项目，所以要好好的收集和利用结肠镜检查数据，构建大型的结直肠数据集，以此来推动息肉分割领域进一步的发展。

(2) 医学图像的域适应：本文所研究的数据集没有涉及域适应的问题，但是一个强大的医学图像模型应该能够在多个医学中心的数据都能发挥稳定的性能。每个医学中心都有自己的模式和方法，导致最终的图像存在肉眼难以分辨的差异。因此，医学图像的域适应问题是一个十分有意义的研究课题。针对从每一个医院收集到的结肠镜图像，可能受到设备型号不一致，操作流程偏差等影响，会导致结肠镜得到图像有差异，所以分析在各个医院收集到的结肠镜图像，做好域适应的工作。通过域适应的方法，可以使模型在不同医学中心的数据上具有更好的泛化能力。

(3) 多模态数据融合：一位患者的检查可能不仅仅包含图像，还包括血液分析和病理检查等。如果将一位患者的所有检查结果都进行学习，或许可以进一步提升息肉分割的性能，更好地辅助医生做出最终的诊断。在息肉分割的临床实际中，医生一般会给出诊断的意见，而不是分割掩码。如果可以充分利用诊断的意见文本并结合结肠镜图像来对息肉分割进行训练，会进一步提升半监督息肉分割的性能，同时也会降低医生手动分

割的负担。因此，研究人员可以探索如何有效地融合多模态数据，以提高医学图像分割任务的准确性和鲁棒性。

在未来的工作中，重点关注这些方面有望进一步推动医学图像分割领域的发展，从而为智慧医疗带来更多的突破和进步。

参考文献

- [1] 周平红, 姚礼庆, 陈巍峰, 等. 结直肠腺瘤性息肉和早期癌的内镜治疗[J]. 中华外科杂志, 2008(18):1386-1389.
- [2] Cheryl J. Pendergrass, et al. Occurrence of Colorectal Adenomas in Younger Adults: An Epidemiologic Necropsy Study[J]. Clinical Gastroenterology and Hepatology, 2008;6(9):1011-1015.
- [3] Williams AR, Balasooriya BA, Day DW. Polyps and cancer of the large bowel: a necropsy study in Liverpool[J]. Gut. 1982;23(10):835-842.
- [4] Ben, Qiwen et al. Body mass index increases risk for colorectal adenomas based on meta-analysis[J]. Gastroenterology, 2012;142(2): 762-72.
- [5] Nam, Su Youn et al. Abdominal visceral adipose tissue predicts risk of colorectal adenoma in both sexes[J]. Clinical gastroenterology and hepatology: the official clinical practice journal of the American Gastroenterological Association, 2020;8(5): 443-450.e2.
- [6] 张宏博, 丁杰, 樊代明, 等. 结肠镜检查前肠道准备清洁效果的临床对比研究[J]. 中华消化内镜杂志, 2003(1):37-39.
- [7] McLachlan SA, Clements A, Austoker J. Patients' experiences and reported barriers to colonoscopy in the screening context—a systematic review of the literature[J]. Patient Educ Couns, 2012;86: 137 – 46.
- [8] Centers for Disease Control and Prevention. Vital signs: colorectal cancer screening test use—United States, 2012[J]. MMWR Morb Mortal Wkly Rep, 2013;62:881 – 888.
- [9] Peery AF, Dellon ES, Lund J et al. Burden of gastrointestinal disease in the United States: 2012 update[J]. Gastroenterology, 2012;143:1179 – 1187.
- [10] Harewood GC, Sharma VK, de Garmo P. Impact of colonoscopy preparation quality on detection of suspected colonic neoplasia[J]. Gastrointest Endosc, 2003;58:76 – 79.
- [11] Froehlich F, Wietlisbach V, Gonvers JJ et al. Impact of colonic cleansing on quality and diagnostic yield of colonoscopy: the European Panel of Appropriateness of Gastrointestinal Endoscopy European multicenter study[J]. Gastrointest Endosc, 2005;61: 378 – 384.
- [12] 肖定华, 沈守荣, 贾燕, 等. 结肠镜检查肠道准备改良方法探讨[J]. 中国内镜杂志, 2007(2):190-192.

- [13] J. Yao, M. Miller, M. Franaszek, and R. M. Summers. Colonic polyp segmentation in ct colonography-based on fuzzy clustering and deformable models[J]. IEEE TMI,2004;23(11): 1344–1352.
- [14] L. Lu, A. Barbu, M. Wolf, J. Liang, M. Salganicoff, and D. Comaniciu. Accurate polyp segmentation for 3d ct colongraphy using multistaged probabilistic binary learning and compositional model[C]//IEEECVPR, 2008:1–8.
- [15] S. Gross, M. Kennel, T. Stehle, J. Wulff, J. Tischendorf, C. Trautwein, and T. Aach. Polyp segmentation in nbi colonoscopy[J]. Bildverarbeitung fur die Medizin, 2009:252–256.
- [16] M. Ganz, X. Yang, and G. Slabaugh. Automatic segmentation of polyps in colonoscopic narrow-band imaging data[C]//IEEE TBE, 2012;59(8):2144–2151.
- [17] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation[C]//MICCAI, 2015:234–241.
- [18] Zhou, Z., Siddiquee, M. M. R., Tajbakhsh, N., & Liang, J. Unet++: A nested u-net architecture for medical image segmentation[J]. Deep learning in medical image analysis and multimodal learning for clinical decision support, 2018: 3-11.
- [19] He, K. M.; Zhang, X. Y.; Ren, S. Q.; Sun, J. Deep residual learning for image recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770–778.
- [20] Zhang, Z., Liu, Q., & Wang, Y. Road extraction by deep residual u-net[C]//IEEE Geoscience and Remote Sensing Letters, 2018;15(5):749-753.
- [21] Jha, D., Smedsrud, P. H., Riegler, M. A., Johansen, D., De Lange, T., Halvorsen, P., & Johansen, H. D. Resunet++: An advanced architecture for medical image segmentation[C]//2019 IEEE International Symposium on Multimedia, 2018:2225-2255.
- [22] Y. Fang, C. Chen, Y. Yuan, and K.-y. Tong. Selective feature aggregation network with area-boundary constraints for polyp segmentation[J]. MICCAI, 2019:302–310.
- [23] T.-C. Nguyen, T.-P. Nguyen, G.-H. Diep, A.-H. Tran-Dinh, T. V. Nguyen, and M.-T. Tran. Ccbanet: cascading context and balancing attention for polyp segmentation[J]. MICCAI, 2021:633–643.
- [24] Lin, A., Chen, B., Xu, J., Zhang, Z., Lu, G., & Zhang, D. Ds-transunet: Dual swin transformer u-net for medical image segmentation[C]//IEEE Transactions on

- Instrumentation and Measurement,2022;71:1-15.
- [25] Girshick, R. Fast R-CNN[C]//Proceedings of 2015 IEEE International Conference on Computer Vision,2015:1440–1448.
- [26] Redmon, Joseph , et al. You Only Look Once: Unified, Real-Time Object Detection[C]//Computer Vision & Pattern Recognition IEEE, 2016.
- [27] Wei, Liu , et al. SSD: Single Shot MultiBox Detector[C]//Springer, Cham, 2016.
- [28] D.-P. Fan, G.-P. Ji, T. Zhou, G. Chen, H. Fu, J. Shen, and L. Shao. Pranet: Parallel reverse attention network for polyp segmentation[J]. MICCAI, 2020:263–273.
- [29] Lou, A., Guan, S., Ko, H., Loew, M.H. CaraNet: context axial reverse attention network for segmentation of small medical objects[J]. Medical Imaging 2022: Image Processing,2022;12032: 81–92.
- [30] 刘雅芬, 郑艺峰, 江铃焱, 李国和, 张文杰. 深度半监督学习中伪标签方法综述 [J]. 计算机科学与探索, 2022;16(6):1279–1290.
- [31] H. Yao, X. Hu, and X. Li. Enhancing pseudo label quality for semi-supervised domain-generalized medical image segmentation[C]//Proceedings of the AAAI Conference on Artificial Intelligence, 2022;36(3): 3099–3107.
- [32] Y. Shi, J. Zhang, T. Ling, J. Lu, Y. Zheng, Q. Yu, L. Qi, and Y. Gao. Inconsistency-aware uncertainty estimation for semi-supervised medical image segmentation[C]//IEEE Transactions on Medical Imaging, 2021;41(3):608–620.
- [33] Z. Zhang, C. Tian, H. X. Bai, Z. Jiao, and X. Tian. Discriminative error prediction network for semi-supervised colon gland segmentation[J]. Medical Image Analysis, 2022;79:102458, 2022.
- [34] S. Laine and T. Aila. Temporal ensembling for semi-supervised learning[J]. Neural and Evolutionary Computing, 2016.
- [35] A. Tarvainen and H. Valpola. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results[J]. Advances in neural information processing systems, 2017:30.
- [36] X. Zeng, R. Huang, Y. Zhong, D. Sun, C. Han, D. Lin, D. Ni, and Y. Wang. Reciprocal learning for semi-supervised segmentation[J]. International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, 2021:352–361.

- [37] W. Huang, C. Chen, Z. Xiong, Y. Zhang, X. Chen, X. Sun, and F. Wu. Semi-supervised neuron segmentation via reinforced consistency learning[J]. *IEEE Transactions on Medical Imaging*, 2022:1–1.
- [38] J. Wang, X. Li, Y. Han, J. Qin, L. Wang, and Z. Qichao. Separated contrastive learning for organ-at-risk and gross-tumor-volume segmentation with limited annotation[C]// *Proceedings of the AAAI Conference on Artificial Intelligence*, 2022;36(3): 2459–2467.
- [39] X. Hu, D. Zeng, X. Xu, and Y. Shi. Semi-supervised contrastive learning for label-efficient medical image segmentation[C]// *International Conference on Medical Image Computing and Computer Assisted Intervention*, 2021:481–490.
- [40] H. Wu, Z. Wang, Y. Song, L. Yang, and J. Qin. Cross-patch dense contrastive learning for semi-supervised segmentation of cellular nuclei in histopathologic image[C]// *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022: 11 666–11 675.
- [41] Woo, S.; Park, J.; Lee, J.; Kweon, I. S. CBAM: Convolutional block attention module. In: *Computer Vision – ECCV 2018. Lecture Notes in Computer Science*[J]. Ferrari, V.; Hebert, M.; Sminchisescu, C.; Weiss, Y. Eds. Springer Cham, 2018;11211: 3–19.
- [42] Rozenwald MB, Galitsyna AA, Sapunov GV, Khrameeva EE, Gelfand MS. A machine learning framework for the prediction of chromatin folding in *Drosophila* using epigenetic features[J]. *PeerJ Comput Sci*, 2020;6:307.
- [43] Amrit C, Paauw T, Aly R, Lavric M. Identifying child abuse through text mining and machine learning[J]. *Expert Syst Appl*, 2017;88:402–18.
- [44] Pouyanfar S, Sadiq S, Yan Y, Tian H, Tao Y, Reyes MP, Shyu ML, Chen SC, Iyengar S. A survey on deep learning: algorithms, techniques, and applications[J]. *ACM Comput Surv (CSUR)*, 2018;51(5):1–36.
- [45] Dalal N, Triggs B. Histograms of oriented gradients for human detection[J]. *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*, 2005;1:886–93.
- [46] Lowe DG. Object recognition from local scale-invariant features[J]. *Proceedings of the seventh IEEE international conference on computer vision*, 1999;2:1150–1157.
- [47] Wu L, Hoi SC, Yu N. Semantics-preserving bag-of-words models and applications[J].

- IEEE Trans Image Process, 2010;19(7):1908–20.
- [48] LeCun Y, Bengio Y, Hinton G. Deep learning[J]. Nature,2015;521(7553):436–44.
- [49] Adeel A, Gogate M, Hussain A. Contextual deep learning-based audio-visual switching for speech enhancement in real-world environments[J]. Inf Fusion,2020;59:163–70.
- [50] Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks[J]. Commun ACM, 2017;60(6):84–90.
- [51] Hasan RI, Yusuf SM, Alzubaidi L. Review of the state of the art of deep learning for plant diseases: a broad analysis and discussion[J]. Plants, 2020;9(10):1302.
- [52] Xiao Y, Tian Z, Yu J, Zhang Y, Liu S, Du S, Lan X. A review of object detection based on deep learning[J]. Multimed Tools Appl, 2020;79(33):23729–91.
- [53] Ker J, Wang L, Rao J, Lim T. Deep learning applications in medical image analysis[J]. IEEE Access, 2017;6:9375–89.
- [54] A. Jain, M. Murty, and P. Flynn. Data clustering: A review acm computing surveys,1999;31.
- [55] O. Chapelle, B. Scholkopf, and A. Zien. Semi-supervised learning (chapelle, o. et al., eds.; 2006) [J]. IEEE Transactions on Neural Networks, 2009;20(3): 542–542.
- [56] P. K. Mallapragada, University, Some contributions to semisupervised learning[M]. Michigan State 2010.
- [57] J. Bernardo, M. Bayarri, J. Berger, A. Dawid, D. Heckerman, A. Smith, and M. West, Generative or discriminative? Getting the best of both worlds[J]. Bayesian Stat, 2007;8(3):3–24, 2007.
- [58] R. K. Ando and T. Zhang. Two-view feature generation model for semi-supervised learning[J]. Proceedings of the 24th international conference on Machine learning. ACM, 2007:25–32.
- [59] Valanarasu, J. M. J., Oza, P., Hacihaliloglu, I., & Patel, V. M. Medical transformer: Gated axial-attention for medical image segmentation,2021.
- [60] Long, J., Shelhamer, E., & Darrell, T. Fully convolutional networks for semantic segmentation[C]//IEEE conference on computer vision and pattern recognition,2015:3431-3440.
- [61] Haiyang Mei, Ge-Peng Ji, Ziqi Wei, Xin Yang, Xiaopeng Wei, Deng-Ping Fan.

- Camouflaged Object Segmentation With Distraction Mining[C]//The IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021:8772-8781
- [62] Gao, S., Cheng, M. M., Zhao, K., Zhang, X. Y., Yang, M. H., & Torr, P. H. Res2net: A new multi-scale backbone architecture[C]//IEEE transactions on pattern analysis and machine intelligence,2019.
- [63] Dahl, G. E., Yu, D., Deng, L., & Acero, A. Context-dependent pre-trained deep neural networks for largevocabulary speech recognition[C]//IEEE Transactions on audio, speech, and language processing, 2011;20(1):30-42.
- [64] Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L. Imagenet: A large-scale hierarchical image database, 2009[C]//IEEE Conference on Computer Vision and Pattern Recognition,2020:248-255.
- [65] Z. Wu, L. Su, and Q. Huang. Cascaded partial decoder for fast and accurate salient object detection[C]//IEEE Conf. Comput. Vis. Pattern Recog., 2019:3907–3916.
- [66] Chen, S., Tan, X., Wang, B., & Hu, X. Reverse attention for salient object detection[C]//Proceedings of the European Conference on Computer Vision (ECCV),2018:234-250.
- [67] Silva, J., Histace, A., Romain, O., Dray, X., & Granado, B. Toward embedded detection of polyps in wce images for early diagnosis of colorectal cancer[J]. International journal of computer assisted radiology and surgery, 2014;9(2):283-293.
- [68] Bernal, J., Sánchez, F. J., Fernández-Esparrach, G., Gil, D., Rodríguez, C., & Vilariño, F. WM-DOVA maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians[J]. Computerized Medical Imaging and Graphics, 2015;43,:99-111.
- [69] Tajbakhsh, N., Gurudu, S. R., & Liang, J. Automated polyp detection in colonoscopy videos using shape and context information[C]//IEEE transactions on medical imaging, 2015;35(2):630-644.
- [70] Vázquez, D., Bernal, J., Sánchez, F. J., Fernández-Esparrach, G., López, A. M., Romero, A., ... & Courville, A. A benchmark for endoluminal scene segmentation of colonoscopy images[J]. Journal of healthcare engineering, 2017.
- [71] Jha, D., Smedsrud, P. H., Riegler, M. A., Halvorsen, P., de Lange, T., Johansen, D., & Johansen, H. D. Kvasir-seg: A segmented polyp datase[J]. International Conference on

- Multimedia Modeling, 2020: 451-462.
- [72] D. E. Worrall, S. J. Garbin, D. Turmukhambetov, and G. J. Brostow. Harmonic networks: Deep translation and rotation equivariance[J]. Proc. IEEE Conf. Comput. Vis. Pattern Recognit, 2017;2:5028–5037.
- [73] Yu, L., Wang, S., Li, X., Fu, CW., Heng, PA. Uncertainty-Aware Self-ensembling Model for Semi-supervised 3D Left Atrium Segmentation. In: Shen, D., et al. Medical Image Computing and Computer Assisted Intervention – MICCAI 2019[C]//MICCAI 2019, 2019;11765.
- [74] Kendall, A., Gal, Y. What uncertainties do we need in bayesian deep learning for computer vision[J]. NIPS,2017:5574–5584.
- [75] Borgli, H. et al. HyperKvasir, a comprehensive multi-class image and video dataset for gastrointestinal endoscopy[J]. Sci. Data, 2020; 7:283.

致谢

一晃三年转瞬即逝，这三年学会了许多，领悟了许多，成长了许多，自己受益良多，其间所取得的每一点进步都离不开老师的谆谆教诲，同学、朋友的热心帮助和家人的默默支持，谢谢你们！

首先，在这三年的读研生活中，我最想感谢的是我的导师何江萍教授。首先，非常感谢何老师在研究过程中的指导和支持。何老师对于研究思路、实验设计、数据分析和结果解读等方面提供的深刻指导和宝贵建议对于本研究起到了至关重要的作用。其次，非常感激实验室提供的研究环境。最后，我要衷心感谢何老师在整个研究过程中给予我的不断鼓励和激励。何老师对于学术追求的热情和对突破的执着精神，对于我作为研究者的成长起到了重要的推动作用。

其次，我要感谢实验室其他的老师。感谢丁晓阳教授、李强教授、韩金仓教授以及杨佳澄老师对实验室的大力支持，保障实验室各位老师和同学能在舒适的环境中进行科研，同时还经常出差学习见世面。感谢李子荣老师、李兵教授对实验室各位同学科研工作的帮助和指导。

最后，感谢成熟稳重的师兄师姐，有了他们以往的铺垫性工作，才有对本研究的设计和思考。感谢天真可爱的师弟师妹，每次组会相互之间的提问，是促进本研究不断完善的强劲支撑。特别感谢张杨洁同学、郝天宇同学以及陈俊豪师弟对本人科研工作的大力支持。

十分感谢父母在这三年读研生活中提供的经济支持和心理帮助。也感谢所有没能未提及到的老师、同学和家人朋友，感谢大家对我工作生活的帮助。同时也感谢百忙之中参加评阅、答辩的各位专家、教授，谢谢你们给予启迪。