扫描_1

**中图分类号：**

**论文编号：**

作者姓名 陈家佳

一级学科 电子信息

二级学科 计算机技术

指导教师 杜秀全教授

**基于图像级标注的弱监督结肠息肉检测方法研究**

**Research on Weakly Supervised Detection Method of Colon Polyp Based on Image-level Annotation**

A Dissertation Submitted for the Degree of Master

**Candidate：Chen Jiajia**

**Supervisor：Prof. Du Xiuquan**

**图分类号：**

**论文编号：**

硕 士 学 位 论 文

**基于图像级标注的弱监督结肠息肉检测方法研究**

作者姓名 陈家佳 申请学位级别 工学硕士

指导教师姓名 杜秀全 职 称 教授

学科专业 计算机技术 研究方向 医学图像处理

学习时间自 2022年 9 月 1 日 起至 2025年 6 月 1 日止

论文提交日期 2025年 4 月 30 日 论文答辩日期 2025年 5 月16 日

独创性声明

本人声明所呈交的学位论文是本人在导师指导下进行的研究工作及取得的研究成果。据我所知，除了文中特别加以标注和致谢的地方外，论文中不包含其他人已经发表或撰写过的研究成果，也不包含为获得安徽大学或其他教育机构的学位或证书而使用过的材料。如有学术不端行为，一切后果由本人承担，与导师和安徽大学无关。

学位论文作者签名：　　　　　　　　　签字日期：　　　　年　　月　　日

学位论文版权使用授权书

本学位论文作者完全了解安徽大学有关保留、使用学位论文的规定，有权保留并向国家有关部门或机构送交论文的复印件和磁盘，允许论文被查阅和借阅。本人授权安徽大学可以将学位论文的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存、汇编学位论文。

（保密的学位论文在解密后适用本授权书）

学位论文作者签名：　　　　　　　　　导师签名：

签字日期：　　 年　　月　　日　 签字日期：　　　　　年　　月　　日

**摘 要**

自动和准确的脑卒中病灶分割对于临床评估脑卒中的病变状态至关重要，是建立脑疾病计算机辅助诊断系统的关键任务。尽管现有的深度学习方法已经在医学图像病灶分割领域取得了显著的效果，但它们在脑卒中分割上的进一步应用仍然受到以下因素的阻碍：(1)类内不一致，病灶区域内的特征存在较大的差异性；(2)类间模糊，病灶区域在外观上与正常脑组织相似；(3)正负样本不平衡，代表病灶的正样本数量远远少于代表正常脑组织的负样本数量。为了应对脑卒中分割中的这些挑战，实现病灶的更精准分割。本文的主要工作如下：

1.针对类内不一致和类间模糊问题，本文提出注意力引导多尺度恢复网络。首先，在网络编码阶段使用粗粒度图像块注意力模块，以多级和显式的监督方式获得基于图像块的粗粒度注意力图，通过基于图像块级别的加权技术实现目标空间上下文信息的显著表示，缓解了类内不一致在编码阶段的影响。其次，为了获得更精准的边界分割以应对类间模糊挑战，使用新设计的跨维度特征融合模块捕获不同维度特征通道间的全局上下文信息，以进一步指导2D和3D特征的选择性融合，弥补2D卷积边界学习能力的不足。最后，由于编码阶段的池化操作会造成边界和空间信息的丢失，所以在解码阶段，使用多尺度反卷积上采样代替线性插值恢复特征图分辨率，利用多尺度上下文信息来恢复丢失的细节信息。

2.针对正负样本不平衡和解码阶段的类内不一致挑战造成的分割结果假阴性偏高问题，我们提出了目标感知监督残差学习网络。针对正负样本不平衡问题，设计了一个目标感知损失函数，通过聚焦因子扩展损失计算的高权重区域，在高度关注正样本损失的同时补偿目标周围的负样本损失。提升了分割结果的召回率，同时避免了精确率的失衡。然后，设计了粗粒度残差学习模块，在解码阶段逐步修复丢失的残差特征，缓解了解码阶段中类内不一致导致的假阴性偏高的问题。其中，反向注意力单元和正向注意力单元分别抑制冗余的已识别目标和背景噪声，使网络更集中的提取残差目标区域中的重要特征。

我们在ATLAS和ISLES 2015公共脑卒中数据集上进行了大量实验，结果表明，我们提出的每个模块是有效的，在脑卒中诊断方面具有巨大的潜力。

**关键字：**粗粒度注意力，特征融合，多尺度信息，损失函数，残差学习

**Abstract**

Automatic and accurate stroke lesion segmentation is crucial for clinical assessment of the lesion status of stroke and is a key task for establishing computer-aided diagnosis systems for brain diseases. Although current deep learning methods have achieved remarkable results in the field of medical image lesion segmentation, their further application to stroke segmentation is still hampered by (1) intra-class inconsistency, with large variability between lesion regions; (2) inter-class ambiguity, with lesion regions resembling normal in appearance; and (3) positive and negative sample imbalance, with the number of positive samples representing lesions is much smaller than the number of negative samples representing normal brain tissue. To address these challenges in stroke segmentation and achieve more accurate segmentation of lesions. The main work of this paper is as follows:

1. For the intra-class inconsistency and inter-class ambiguity problems, this paper proposes a new stroke segmentation method, the attention-guided multiscale recovery network. Firstly, a coarse-grained patch attention module in the encoding is adopted to obtain a patch-based coarse-grained attention map in a multistage, explicitly supervised manner, enabling target spatial context saliency representation with a patch-based weighting technique that eliminates the effect of intraclass inconsistency in the encoding stage. Secondly, to obtain more accurate boundary segmentation to meet the challenge of interclass indistinction, a newly designed cross-dimensional feature fusion module is used to capture the global context information between different dimensional feature channels to further guide the selective aggregation of 2D and 3D features, which can compensate for the lack of boundary learning capability of 2D convolution. Lastly, since the pooling operation in the encoding stage causes the boundary and spatial information to be lost, in the decoding stage, multi-scale deconvolution upsampling is used instead of linear interpolation to recover the feature map resolution, and multi-scale contextual information is used to recover the lost detail information.

2. We propose a target-aware supervised residual learning network in order to cope with the high false negative problem of segmentation results caused by the positive and negative sample imbalance challenge and the intra-class inconsistency challenge in the decoding stage. Considering the problem of imbalance of positive and negative samples, a creatively target-aware loss function is designed to dilate the high-weight region of the loss calculation with a focus factor, pay high attention to the positive sample loss, and compensate for the negative samples around the target. The Recall of the segmentation results is improved while avoiding the imbalance of the Precision. Then, a coarse-grained residual learning module is developed to gradually repair the lost residual features during the decoding phase, which alleviates the problem of high false negatives caused by intra-class ambiguities. Here, our reverse/forward attention unit suppresses redundant target/background noise and allows relatively more focused highlighting of important features in the target residual region.

We conducted extensive experiments on the ATLAS and ISLES 2015 public stroke datasets, and the results show that each of our proposed modules is effective and has great potential for stroke diagnosis.

**Key words:** Coarse-grained attention, Feature fusion, Multiscale information, Loss function, Residual learning

**目 录**

[**摘 要** V](#_Toc171495248)

[第一章 绪论 1](#_Toc171495249)

[1.1 研究背景及意义 1](#_Toc171495250)

[1.2 国内外研究现状与研究思路 3](#_Toc171495251)

[1.2.1 息肉检测方法的研究现状 3](#_Toc171495252)

[1.2.2 图像级弱监督方法的研究现状 5](#_Toc171495253)

[1.3 主要研究内容 9](#_Toc171495254)

[1.4 论文组织结构 10](#_Toc171495255)

[第二章 相关理论与基础 12](#_Toc171495256)

[2.1 息肉相关理论知识 12](#_Toc171495257)

[2.2 息肉检测中的监督方式 13](#_Toc171495258)

[2.2.1 全监督息肉检测 13](#_Toc171495259)

[2.2.2 半监督息肉检测 13](#_Toc171495260)

[2.2.3 弱监督息肉检测 13](#_Toc171495261)

[2.4 深度学习技术 16](#_Toc171495262)

[2.4.1 卷积神经网络的结构 16](#_Toc171495263)

[2.4.2 经典网络WSDDN 16](#_Toc171495264)

[2.5 弱监督息肉检测相关技术 16](#_Toc171495265)

[2.5.1 损失函数 16](#_Toc171495266)

[2.5.2 非极大值抑制 16](#_Toc171495267)

[2.5.3 评价指标 16](#_Toc171495268)

[2.6 本章小结 18](#_Toc171495269)

[第三章 形态驱动的弱监督息肉检测网络 19](#_Toc171495270)

[3.1 研究动机 19](#_Toc171495271)

[3.2 MDNet模型 20](#_Toc171495272)

[3.2.1 跨域参考模块 20](#_Toc171495273)

[3.2.2 空间类别模块 20](#_Toc171495274)

[3.2.3 多示例分支 20](#_Toc171495275)

[3.2.4 双阈值后处理策略 20](#_Toc171495276)

[3.3 实验细节 23](#_Toc171495277)

[3.3.1 数据集 23](#_Toc171495278)

[3.3.2 网络结构和参数 23](#_Toc171495279)

[3.3.3 损失函数 23](#_Toc171495280)

[3.3.4 评估指标 23](#_Toc171495281)

[3.3.5 对比算法介绍 23](#_Toc171495282)

[3.4 实验结果以及分析 25](#_Toc171495283)

[3.4.1与现有方法的比较 25](#_Toc171495284)

[3.4.2 消融实验：跨域参考模块 25](#_Toc171495285)

[3.4.3 消融实验：空间类别模块 25](#_Toc171495286)

[3.4.4 消融实验：双阈值后处理策略 25](#_Toc171495287)

[3.5 本章小结 26](#_Toc171495288)

[第四章 基于合作的弱监督息肉检测网络 27](#_Toc171495289)

[4.1 研究动机 27](#_Toc171495290)

[4.2 CBNet模型 28](#_Toc171495291)

[4.2.1 基于合作的区域生成网络 28](#_Toc171495292)

[4.2.2 自适应感兴趣区域融合模块 28](#_Toc171495293)

[4.2.3 多示例检测网络 28](#_Toc171495294)

[4.3 实验细节 30](#_Toc171495295)

[4.3.1 数据集 30](#_Toc171495296)

[4.3.2 网络结构和参数 30](#_Toc171495297)

[4.3.3 损失函数 30](#_Toc171495298)

[4.3.4 评估指标 30](#_Toc171495299)

[4.3.5 对比算法介绍 30](#_Toc171495300)

[4.4 实验结果及分析 32](#_Toc171495301)

[4.4.1 与现有方法比较 32](#_Toc171495302)

[4.4.2 消融实验：基于合作的区域生成网络 32](#_Toc171495303)

[4.4.3 消融实验：自适应感兴趣区域融合模块 32](#_Toc171495304)

[4.5 本章小结 32](#_Toc171495305)

[总结与展望 33](#_Toc171495306)

[参考文献 35](#_Toc171495307)

[攻读硕士期间取得的学术成果 45](#_Toc171495308)

[致谢 46](#_Toc171495309)

**图索引**

[图1.1 U-Net模型中编码器到解码器各阶段的特征图 4](#_Toc130471991)

[图1.2 细粒度和粗粒度注意力的对比图 4](#_Toc130471992)

[图1.3 目标感知损失中聚焦样本的扩张图 7](#_Toc130471993)

[图2.1 U-Net网络模型 10](#_Toc130471994)

[图2.2 D-UNet网络模型 11](#_Toc130471995)

[图2.3 通道注意力和空间注意力 12](#_Toc130471996)

[图3.1 注意力引导多尺度恢复网络模型 15](#_Toc130471997)

[图3.2 粗粒度图像块注意力结构图 16](#_Toc130471998)

[图3.3 注意力图的上采样过程 17](#_Toc130471999)

[图3.4 跨维度特征融合模块的结构图 19](#_Toc130472000)

[图3.5 3\*3反卷积的缺陷表示图 20](#_Toc130472001)

[图3.6 不同方法的分割结果图 25](#_Toc130472002)

[图3.7 边界分割结果对比 25](#_Toc130472003)

[图3.8 AGMR-Net与其他模型的箱型图 26](#_Toc130472004)

[图3.9 粗粒度注意力的可视化效果图 28](#_Toc130472005)

[图3.10 CPA模块对特征权重的影响 28](#_Toc130472006)

[图4.1 目标感知监督残差学习网络结构 32](#_Toc130472007)

[图4.2 粗粒度残差学习模块结构图 33](#_Toc130472008)

[图4.3 损失函数聚焦区域扩张图 35](#_Toc130472009)

[图4.4 聚焦因子的作用图 35](#_Toc130472010)

[图4.5 不同方法预测结果的比较 40](#_Toc130472011)

[图4.6 AGMR-Net和TSRL-Net在召回率上的比较 41](#_Toc130472012)

[图4.7 不同模型的DSC结果的箱形图 41](#_Toc130472013)

[图4.8 β在目标感知损失函数中的影响 43](#_Toc130472014)

[图4.9 不同残差学习的侧输出的对比 46](#_Toc130472015)

**Figure Index**

[Figure 1.1 Feature maps of encoder to decoder stages in the U-Net model 4](#_Toc130759016)

[Figure 1.2 Comparison of fine-grained and coarse-grained attention maps 4](#_Toc130759017)

[Figure 1.3 Expansion plot of focused samples in target-aware loss 7](#_Toc130759018)

[Figure 2.1 U-Net network model 10](#_Toc130759019)

[Figure 2.2 D-UNet network model 11](#_Toc130759020)

[Figure 2.3 Channel attention and spatial attention 12](#_Toc130759021)

[Figure 3.1 Attention-guided multiscale recovery network model 15](#_Toc130759022)

[Figure 3.2 Coarse-grained patch attention structure figure 16](#_Toc130759023)

[Figure 3.3 The upsampling process of the attentional map 17](#_Toc130759024)

[Figure 3.4 Structure of the cross-dimensional feature fusion module 19](#_Toc130759025)

[Figure 3.5 Disadvantage representation figure of 3\*3 deconvolution 20](#_Toc130759026)

[Figure 3.6 Figure of segmentation results of different methods 25](#_Toc130759027)

[Figure 3.7 Comparison of boundary segmentation results 25](#_Toc130759028)

[Figure 3.8 Box plot of AGMR-Net with other models 26](#_Toc130759029)

[Figure 3.9 Visualization of coarse-grained attention 28](#_Toc130759030)

[Figure 3.10 Impact of CPA module on feature weights 28](#_Toc130759031)

[Figure 4.1 Target-aware supervised residual learning network structure 32](#_Toc130759032)

[Figure 4.2 Structure diagram of coarse-grained residual learning module 33](#_Toc130759033)

[Figure 4.3 Loss function focus area expansion map 35](#_Toc130759034)

[Figure 4.4 Map of the role of focus factors 35](#_Toc130759035)

[Figure 4.5 Comparison of prediction results of different methods 40](#_Toc130759036)

[Figure 4.6 Comparison of AGMR-Net and TSRL-Net in terms of Recall 41](#_Toc130759037)

[Figure 4.7 Box plots of DSC results for different models 41](#_Toc130759038)

[Figure 4.8 The effect of β in the target-aware loss function 43](#_Toc130759039)

[Figure 4.9 Comparison of the side outputs of different residual learning 46](#_Toc130759040)

**表索引**

[表2.1 医学图像分割中常用损失函数 13](#_Toc130472041)

[表3.1 每个模块对骨干网络D-UNet的影响 23](#_Toc130472042)

[表3.2 我们的方法和先前的方法比较 24](#_Toc130472043)

[表3.3 边界分割结果对比 26](#_Toc130472044)

[表3.4 空间注意方法的比较 27](#_Toc130472045)

[表3.5 CFF与普通特征融合方法的比较 29](#_Toc130472046)

[表3.6 不同衰减率对基线的提升 29](#_Toc130472047)

[表3.7 MDU与单尺度反卷积的比较 30](#_Toc130472048)

[表3.8 MDU模块在不同位置的整合 30](#_Toc130472049)

[表4.1 消融实验的定量结果 38](#_Toc130472050)

[表4.2 在ATLAS和ISLES数据集上的对比 39](#_Toc130472051)

[表4.3 复杂性分析和显著性分析实验 42](#_Toc130472052)

[表4.4 γ和β对损失函数的影响 43](#_Toc130472053)

[表4.5 目标感知损失在其他网络上的适用性 44](#_Toc130472054)

[表4.6 不同损失函数的对比 45](#_Toc130472055)

[表4.7 不同的残差学习模块的比较 45](#_Toc130472056)

**Table Index**

[Table 2.1 Common loss functions in medical image segmentation 13](#_Toc130472057)

[Table 3.1 Impact of each module on the backbone network D-UNet 23](#_Toc130472058)

[Table 3.2 Comparison of our method with previous methods 24](#_Toc130472059)

[Table 3.3 Comparison of boundary segmentation results 26](#_Toc130472060)

[Table 3.4 Comparison of spatial attention methods 27](#_Toc130472061)

[Table 3.5 Comparison of CFF and common feature fusion methods 29](#_Toc130472062)

[Table 3.6 Enhancement of baseline by different decay rates 29](#_Toc130472063)

[Table 3.7 Comparison of MDU and single-size deconvolution 30](#_Toc130472064)

[Table 3.8 Integration of MDU modules in different locations 30](#_Toc130472065)

[Table 4.1 Quantitative results of ablation experiments 38](#_Toc130472066)

[Table 4.2 Comparison on ATLAS and ISLES datasets 39](#_Toc130472067)

[Table 4.3 Complexity analysis and significance analysis experiments 42](#_Toc130472068)

[Table 4.4 The effect of γ and β on the loss function 43](#_Toc130472069)

[Table 4.5 Applicability of target-aware loss on other networks 44](#_Toc130472070)

[Table 4.6 Comparison of different loss functions 45](#_Toc130472071)

[Table 4.7 Comparison of different residual learning modules 45](#_Toc130472072)

# 第一章 绪论

## 1.1 研究背景及意义

结直肠癌（CRC）是全球第三大最危险的癌症[1]，造成严重的公共卫生负担。如图1.1 [2]所示，在全球各地区，结直肠癌都具有很高的发病率和死亡率。根据《2024年癌症统计报告》[3]，预计在2024年美国将有超过2百万新增癌症确诊病例和60万例因癌症导致的死亡。我国的癌症中心基于肿瘤登记及随访监测的最新数据[4]统计， 2022年我国全癌种新发病例482.47万人，其中结直肠癌新发病例高达51.71万人，占全部恶性肿瘤发病的10.7%；由结直肠癌导致的死亡病例24万人，占全部恶性肿瘤死亡的9.3%。

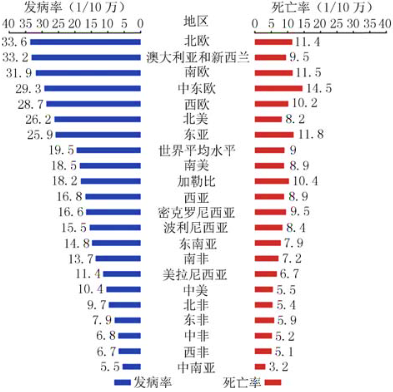


图1.1 全球各地区结直肠癌发病率和死亡率[2]

Figure 1.1 Incidence and mortality of colorectal cancer in the world by region[2]

结直肠癌的发生与多种因素相关，其中，结肠粘膜过度增生而形成的赘生物息肉，是引发结直肠癌的主要因素[5]。这些息肉发病隐匿，起初是消化道的一部分，是结肠黏膜表面突出到肠腔内的隆起状病变，它们初始时可能是良性的，但在时间的推移下有可能逐渐发展为腺瘤，最终恶化成癌症，威胁患者的生命。因此通过早期筛查并切除必要的结肠息肉，可以有效预防结直肠癌的发生和发展，为患者提供更好的生存机会。在临床实践中，结肠镜检查是预防结肠癌的金标准[6]，结肠镜内置照明装置能获取清晰的肠腔视野并提供结肠中息肉的位置和外观的信息，内镜医生通过操作结肠镜可以检查整个结肠，同时立即去除癌前息肉。

然而，结肠镜检查要求操作者有较高的经验水平和专业知识。因为结肠壁的粘膜会形成大量褶皱，其中一些褶皱与结肠息肉相似，使息肉难以识别；结肠息肉尺寸差异很大，尤其是处于初期凸起阶段的息肉很小，极其不易察觉。除此之外，在检查过程中包含息肉的视频帧远少于不包含息肉的视频帧，如果只靠医生用肉眼在大量的视频画面中查找尺寸形状不一的息肉将无疑是一项繁琐且极具挑战的工作。最近的研究表明，在接受结肠镜检查的患者中，有高达26%[7]的息肉被遗漏，这些息肉有可能会发展成晚癌，从而降低患者的生存率。因此，借助计算机对结肠镜视频进行息肉的自动检测的计算机辅助诊断（Computer Aided Diagnosis，CAD）在医疗领域受到了极大地推动。

得益于自动学习数据特征的能力、卓越的数据拟合性能以及丰富的监督学习策略，深度学习方法已成为各领域CAD的基础[8,9]。这些方法能够从数据中自动学习到低层次的语义特征和高层次的抽象特征，并通过海量的标注数据进行端到端的训练。在这些方法中基于边界框的全监督方法取得了优秀的效果，然而，截至目前为止，息肉全监督检测的“金标准”仍然是内镜医生手动标注的边界框。这种对象级别的标注需要标注人员利用标注工具为每一个目标对象打上精准的边界框，这需要大量的标注时间。更重要的是，对于息肉图像而言，标注者需要具备一定的医学知识，才能判断图像中是否存在息肉以及息肉的类型。由于内镜医生的供不应求[10]以及标注工作的重负，息肉检测领域尚缺充足且高质量的标注图像，这阻碍了全监督学习方法在息肉自动检测领域上的应用。为了降低对注释的依赖，基于多示例学习的弱监督息肉检测任务孕育而生。如图1.2所示，根据弱标注的不同形式，弱监督方法又进一步被分为了点、涂鸦以及图像级三种[11-13]。其中，图像级的标注仅需要标注人员标注目标的类别即可，1表示图像中存在该类别的目标，0则表示不存在，是标注成本最低的弱监督方法，具有极大的应用价值，有助于推动智能医疗的发展。

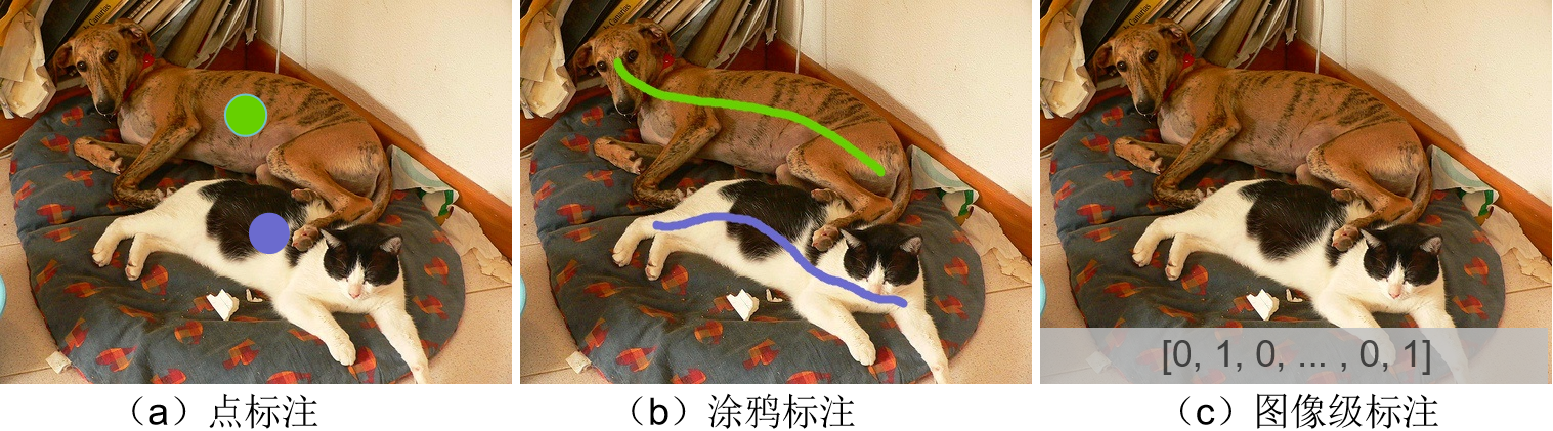


图1.2 检测任务常见的弱标注形式[11-13]（示例来自PASCAL VOC 2012数据集）

Figure 1.2 Common forms of weak annotation in detection tasks [11-13](examples from PASCAL VOC 2012 dataset)

因此，本研究重点探讨如何利用图像级标注来检测结肠图像中的息肉。方法仅根据图像中是否存在该类息肉就能实现自动检测，包括定位息肉位置、评估息肉所属类别的概率。这显著降低对于标注精确度的需求，拓宽实际应用的范围，将有力促进医学图像自动化分析领域的发展和普及，具有十分重要的研究价值。

## 1.2 国内外研究现状与研究思路

随着人工智能的发展和医学图像处理技术的日益成熟，结肠图像的自动化息肉检测以极具发展前景的优势吸引了国内外众多科研人员，如今已取得许多优秀的成果。本章节主要对国内外结肠图像息肉检测方法以及图像级弱监督方法的研究现状进行归纳与总结，主要分为两小节：1）息肉检测方法的研究现状；2）图像级弱监督方法的研究现状。

### 1.2.1 息肉检测方法的研究现状

早期的息肉检测技术主要依赖传统图像处理技术：提取图像边缘、定向梯度直方图（Histograms of Oriented Gradients，HOG）、颜色、形状、尺度不变特征变换（Scale-invariant feature transform，SIFT）等，通过传统分类器进行识别。例如，Hasan等人[14]融合轮廓小波变换和神经特征，采用最小冗余最大相关（Max-Relevance and Min-Redundancy，MRMR）降维方法从图像中提取深层特征，利用支持向量机对胃肠道息肉进行诊断，并标记检测到的息肉区域。然而，这种特征的设计依据主要来自医生的先验知识，受医生阅历和主观性的影响，缺乏通用性，难以处理复杂场景的检测任务和端到端的学习[15]。此外，由于息肉图像形态多样性、光照变化多样性和背景多样性，特征提取方法的鲁棒性较差。例如，由于某些息肉表面光滑，在内镜图像中表现出与正常衬里相同的形状-纹理特征，传统的方法容易忽略这类息肉;正常的结肠内壁呈凸起结构，传统的方法容易将其误诊为息肉。因此，传统的息肉检测方法很难很好地完成检测任务[16]。

为了克服传统方法的不足，研究开始转向无需人工设计，可自动从原始数据学习特征的深度学习[17]。Tashk等人[18]调整区域建议网络，从结肠胶囊内窥镜获得的视频帧中定位息肉，并为其恶性肿瘤的风险提供预测评分。Liew等人[19]基于残差网络，结合主成分分析和AdaBoost集成学习，应用对比度增强、图像阈值分割和中值滤波等技术来减少噪声的干扰，实现更为精准的息肉检测。为了实现在息肉边界与背景模糊的情况下也能实现很好的检测效果，Qadir等人 [20] 在单前馈网络模型中引入二维高斯掩模来降低息肉的漏检率。此外，还有一系列基于YOLO（You Only Look Once）的方法 [21-23]，他们从减少时间、小息肉、纹理等角度对基础的YOLO方法进行优化，进一步提升了检测的实时性与准确性。随着视觉大模型的发展，近年来还涌现出了一些借助SAM（Segment Anything）进行息肉检测的方法[24-27]。与传统方法相比这些全监督方法的准确率和泛化能力虽然都有了很大的提高，但仍存在以下缺点：1）硬标注，方法需要与具有实例级边界框标注的数据集相结合。然而医学图像的复杂性和息肉的多样性，标注工作费时费力且极具挑战，这限制了全监督方法在息肉检测领域的应用；2）预测信息不完整，息肉的种类、类别置信度和位置能提供不同的信息，具有同等重要程度，应同时给出。而已有的方法给出的预测结果在临床应用中存在不同组合的预测内容缺失；3）息肉的类别划分不合理，目前的方法多以病理学的增生型或腺瘤型进行分类并通过这两类息肉的mAP（Mean Average Precision）来评估性能的好坏，只关注于不同病理下息肉的检出率忽略息肉不同形态下检测性能的差异。但实际的研究表明息肉的漏检的影响因素主要是其形态（扁平的，带蒂的，边缘的）[28]。因此，基于病理的性能评估方式是不合理的，合理的评估应基于形态学。

为了解决全监督方法在训练过程中需要大量硬标注数据的困难，人们开始将目光转向弱监督方法。如图1.3所示，这些方法可以概括为两类，一类通过类激活映射（Class Activation Mapping，CAM），另一种通过多示例学习（Multiple Instance Learning，MIL）。

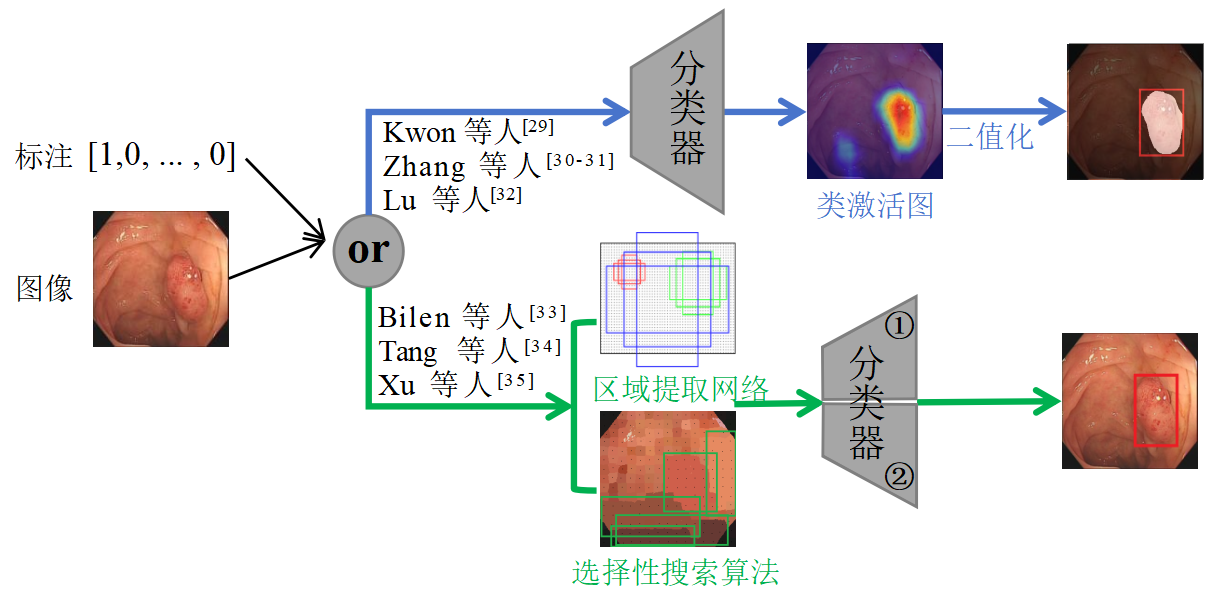


图1.3 弱监督方法的分类

Figure 1.3 Classification of weak supervision methods

1. 基于CAM的方法通过将网络中最后一个卷积层的特征图与其对应的类别权重进行加权求和生成热力图，突出显示图像中对分类决策最为重要的区域，实现对目标的精确定位。Kwon等人[29]设计了一种多任务学习的结肠癌组织学定位方法，结合组织学分类和病变定位任务，在息肉分割的数据集上进行训练，有效提高了病变定位的准确性，并提升了组织学分类的准确率。Zhang等人[30-31]提出了I2C方法引入像素级关联约束，提高定位图的质量；还提出了SEM方法，通过目标子区域内的点相似性来细化定位图。Lu等人[32]将几何形状信息纳入监督的范围，提出了一种多任务协作损失函数。
2. 基于MIL的方法将输入图像分为多个示例包，并为其构建对应的分类以及检测器，将检测问题转化为筛选类别正示例的分类问题。Bilen等人[33]首次提出了基于多示例学习的弱监督目标检测框架，通过并行的分类与定位分支成功地将图像级的类别标注转换为对示例级别候选框得分的监督。Tang等人[34]在WSDDN的基础上进一步提出了用于细化分类分支的在线示例分类细化机制，对检测结果进一步细化保证了检测结果完整性。Xu等[35]在MIL中融入了聚类方法，有效利用图像级别的标记信息，实现了精确的分类与分割任务；引入上下文限制作为先验知识，增强了网络对多种尺寸和形状特征的辨识能力，降低了多个示例间的相互干扰或重叠问题。

以上弱监督方法虽然从不同的角度提供了图像级弱监督检测的思路，但仍存在着不足之处。基于CAM方法的不足：1）过度依赖分类器，导致结果过度聚焦目标最具判别性的部分，目标的覆盖缺乏完整性；2）训练过程中采用的损失函数仅能确保分类结果的正确性，缺乏对类激活区域的有效监督，导致激活区域错误。基于MIL的方法虽然避免了CAM类方法的短板，但由于伪标注框的引入，训练数据存在大量的噪声，这不仅加大了训练时网络的学习难度同时降低了检测结果的精度。

综上，目前的息肉检测方法存在的挑战与不足如图1.4所示。A）传统方法：受主观影响大、缺乏泛化性；B）全监督方法：硬标注、预测不完整、类别划分不合理；C）弱监督方法：CAM类方法检测缺乏完整性、激活区域可能出错，MIL类方法存在大量噪声、检测精度低。



图1.4 息肉检测不同方法的优缺点分析图

Figure 1.4 Analytical plots of the advantages and disadvantages of different methods for polyp detection

### 1.2.2 图像级弱监督方法的研究现状

数据标注的精确性和全面性对于模型的性能至关重要，然而，传统的对象级标注方法因其昂贵的成本，限制了其在广泛场景下的应用潜力。为了突破这一瓶颈，研究界开始探索弱监督方法，其中，图像级弱监督学习以其极低的标注成本，吸引了学术界的高度关注和深入研究。然而，由于检测结果过拟合问题、多目标漏检问题以及候选框问题，现有弱监督目标检测的最好mAP仅有58.1%，远远低于全监督mAP的89.3%。为了解决限制检测效果的这三大难题，研究人员分别从优化伪标注生成的算法[36-38]、结合弱监督语义分割[39-42]以及自训练[43-50]角度进行了改进。

基于伪标注生成算法的改进：通过结合选择性搜索（Selective Search，SS）与相似性区域生成，将原本数以千记的伪标注减少到数十个，提升伪标注质量的同时提高算法的效率。例如，Tang等人[36]提出了一种弱监督可用的区域建议网络WSRPN，如图1.5（a）所示，首先利用图像的底层特征提供的语义信息为每个初始的伪标注打分得到对应的似物性分数并进行初步过滤得到粗糙的伪标注，然后利用训练好的基于区域的CNN分类器[43]对粗糙的伪标注进一步筛选得到精炼的伪标注。Cheng等人[37]提出了PG-PS：结合SS和类激活图生成更高召回率的初始伪标注，在筛选时尽可能多的保证正样本，与此同时仅挑选类别界定清晰的难分负样本，并通过调整权重让网络关注更具标识度的负示例，如图1.5（b）所示。

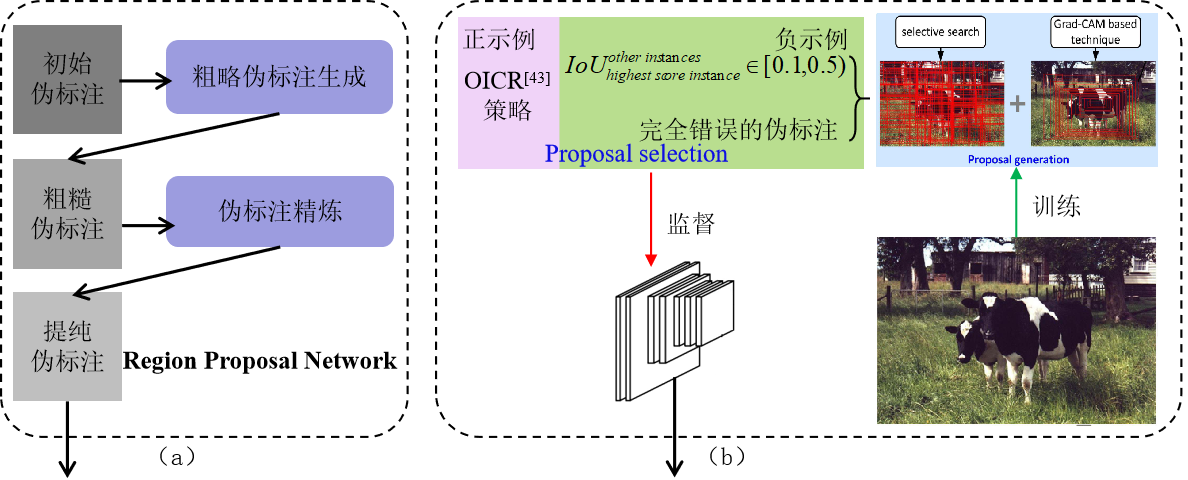


图1.5伪标签优化策略示意图（示例WSRPN和PG-PS方法）

Figure 1.5 Schematic diagram of pseudo-label optimization strategy (examples from WSRPN and PG-PS)

结合弱监督语义分割的改进：在弱监督检测网络的主干上增加一个用于指导检测任务的分割分支。如图1.6所示，分割的协作策略进一步可以细分为仅为：1）仅提供先验知识的单向策略；2） 与检测相互协作的双向策略。例如，在单向策略中，Gao等人[40]提出了一种基于多示例检测网络的对抗擦除策略，利用弱监督分割的结果辅助筛选伪标注，同时维持伪标注在对抗擦除过程中的多示例学习约束，最后融合两个多示例检测网络输出，克服了检测不完整的问题。Wei等人[39] 通过利用分割上下文挖掘紧密贴合的伪标注，训练分类网络并通过类激活映射出用于训练分割网络的伪标签，训练分割网络以生成分割置信图，进而利用该置信图提取更精确的目标物体伪标注。在双向策略中，Li等人[42]提出了一种相互协作的网络，检测为分割提供实例级的热力图作为监督，分割生成分割图为检测提供先验矩阵筛选伪标注，二者通过动态协作形成循环，相辅相成。

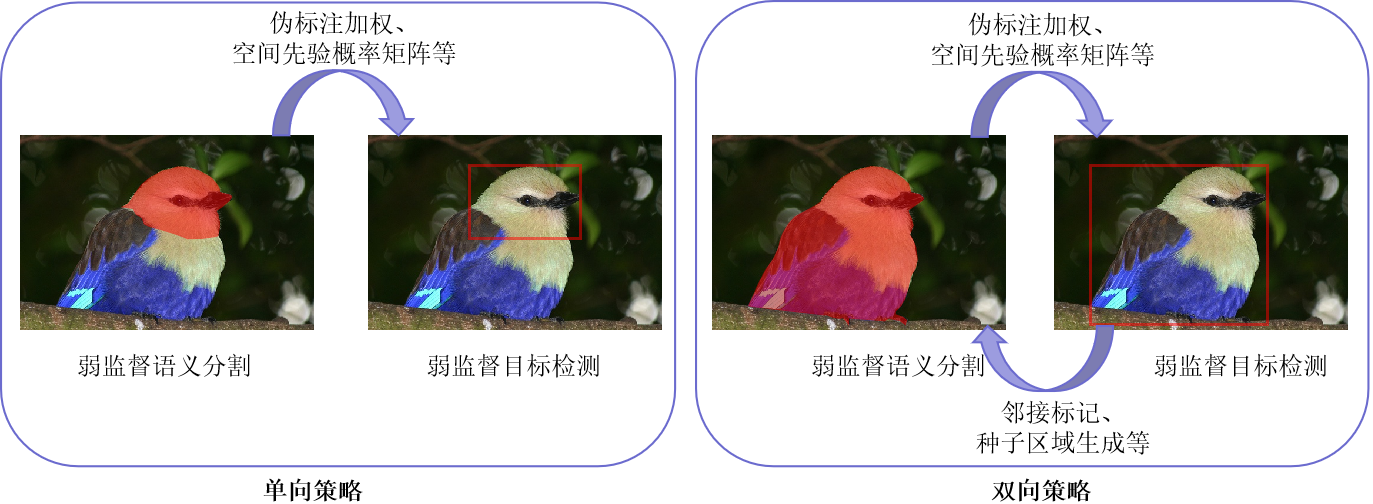


图1.6 分割协作策略示意图

Figure 1.6 Schematic diagram of Segmentation collaboration strategy

基于自训练的改进：以基础框架中的图像分类器作为起点，输出的伪标注得分为基础，为每个伪标注创建用于训练实例级分类器的伪标签。随后，多次迭代这个过程，训练出多个实例级分类器，并通过知识蒸馏技术不断优化并提升伪标注得分的精确度。如图1.7所示，根据自训练阶段的不同特点又可以细分为：1）优化伪标签的方法；2）正则化约束方法；3）结合边界框回归的方法。 在优化伪标签的方法中，最经典的为Tang等人[43]提出的k串行的实例分类器方法OICR：下一级实例级分类器的伪标签由上一级的分类器决定，每个类别最高分的伪标注及与其高交并比的其他伪标注被赋予正样本标签、其余则为负样本标签。在正则化约束的方法中，Huang等人[51]在OICR的基础上进行改进，综合网络多层级和图像多个变换特征图的注意力图用于正则化约束，进而保证物体检测的完整性和小目标的识别。在结合边界框回归的方法中，Chen等人[49]以OICR为基干，在其基础上串行设计了一个空间似然投票模块，将OICR的输出作为模块的输入进行实例挑选、概率积累、似然区域投票并将结果作用于后续的边界框回归。

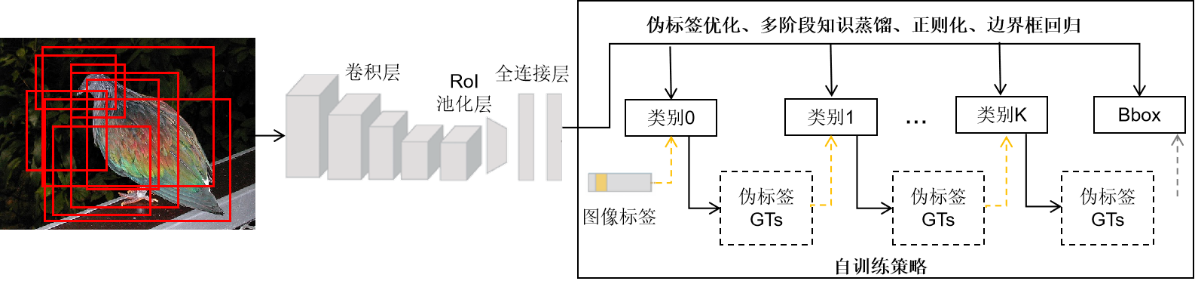


图1.7 自训练策略示意图

Figure 1.7 Schematic diagram of Segmentation collaboration strategy

表1.1汇总了当前主流的弱监督方法在VOC 2007数据集上的检测结果。分析整体结果mAP可以看出，不同策略的方法都有助于提升检测性能，但目前的mAP仍仅有50%~60%，与全监督相比还有很大的提升空间。深入分析每个类别最高的AP（Average Percision）还可以发现这些方法在一些非刚性物体（例如人、植物）和形状多变的物体（例如船、瓶子、椅子）中检测效果只有不到40%，因为这些类别容易出现多目标遮挡并且形变更复杂。同时对比不同的优化策略可以进一步发现，类别最高AP大部分都出现于自训练策略。

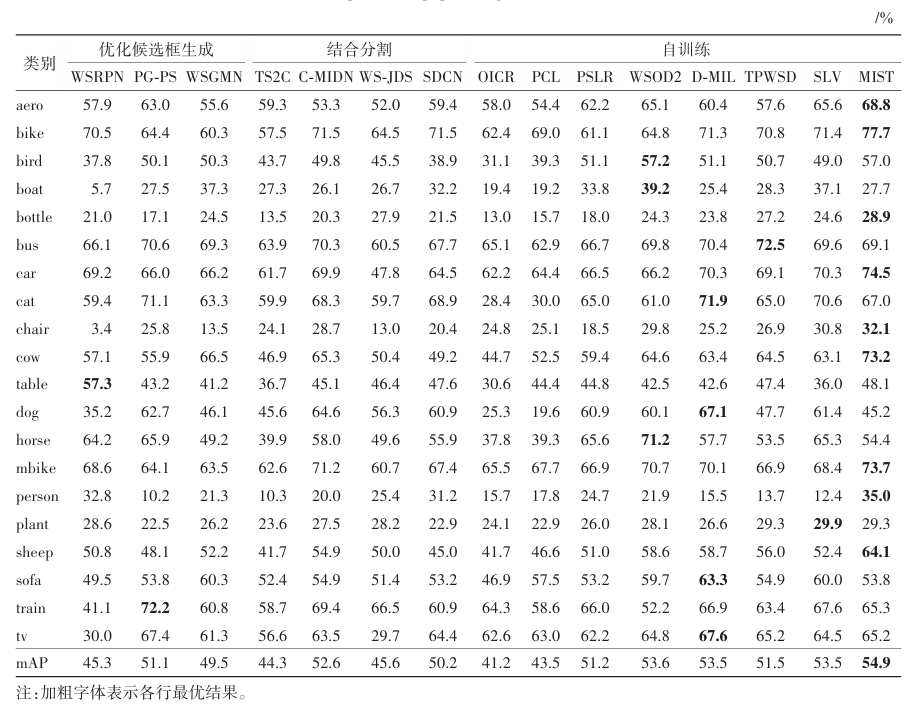


图1.7 自训练策略示意图

Figure 1.7 Schematic diagram of Self-training strategy

虽然以上图像及弱监督检测方法均采用不同的策略提升了检测的效果，但它们都是基于自然图像的，在训练时所采用的伪标注都是提供的。

低然而不幸的是，虽然上述方法在自然图像领域取得了显著的进展，但它们却未能成功地应用于医学图像领域。如表1.2所示，我们从自训练策略中挑选了OICR和WSOD2进行迁移训练并应用于息肉的弱监督检测，并同时挑选了WSDDN作为基干网络进行对比。分析实验结果可以发现，无论在息肉的哪个数据集上，这三种方法的效果都有极大的降低。

## 1.3 主要研究内容

为了能够提取更多特定于目标的特征，本文主要针对脑卒中分割中面临的类间模糊，类内不一致和正负样本不平衡问题进行研究，并基于深度学习方法构建脑卒中自动分割模型。主要研究内容如下：

(1)我们提出了一个注意力引导多尺度恢复网络(AGMR-Net)。在编码阶段，粗粒度的注意力用于突出目标空间信息缓解类内不一致的问题，通过3D卷积学习的精确边界信息解决了类间模糊的问题，而解码阶段的多尺度信息被用来恢复因池化而丢失的空间和边界信息。具体来说，AGMR-Net在编码阶段使用2D和3D卷积进行脑卒中病灶特征提取。为了便于在编码阶段有效地提取目标特征，粗粒度图像块注意力(CPA)模块生成基于图像块的粗粒度注意图，以粗粒度加权的方式突出目标空间信息。在图1.2的粗粒度注意力图中，每个红色图像块上都含有目标的部分区域，我们增强每个目标图像块内特征的表示，引导网络更专注于提取目标特征，解决类内不一致的问题。此外，我们在粗粒度图像块注意力中引入了先验位置图，因为先验知识的监督可以使注意力更加准确，粗粒度图像块注意力也更容易训练。之后，跨维度特征融合(CFF)模块对不同维度的全局特征通道依赖性进行建模，自适应地融合2D和3D特征，以完善2D特征的边界表示，解决类间模糊的挑战。然而，上述模块捕获的特征受限于池化引起的信息损失，为此我们提出多尺度反卷积上采样模块(MDU)使用多尺度特征信息来恢复池化操作丢失的目标空间和边界信息。

(2)在AGMR-Net的基础上，我们解决由类不平衡和类内不一致两个方面引起的假阴性结果偏高的挑战：一是重构正负样本损失，使网络集中于提取目标信息；二是利用正向和反向注意来指导残差特征的学习。因此，我们提出了一个用于脑卒中分割的目标感知监督残差学习网络(TSRL-Net)。具体来说，我们首先设计了目标感知损失函数(TAL)来控制一个聚焦因子，然后将整个图像样本分为聚焦样本和分散样本。如图1.3所示，在金标准中只有白色的目标区域会被赋予高权重。但是经过我们损失函数的重新划分后，在聚焦图中，白色区域的聚焦样本（正样本）和灰色区域的聚焦样本（负样本）都赋予高权重，增加聚焦正样本的损失以减少假阴性来提高召回率，同时补偿聚焦负样本的损失以平衡精确率；黑色区域的分散样本（负样本）被赋予低权重，可以避免负样本损失主导梯度。然后，在网络的解码阶段，我们使用一个粗粒度残差学习(CRL)模块，使用正向注意力(PA)和反向注意力(RA)单元。RA单元的注意图与预测概率成反比，使网络能够检测编码器特征图中的残余特征并提高召回率。PA单元以粗粒度的形式强调目标及其周围的细节特征，以抑制背景噪声和提升精确率。

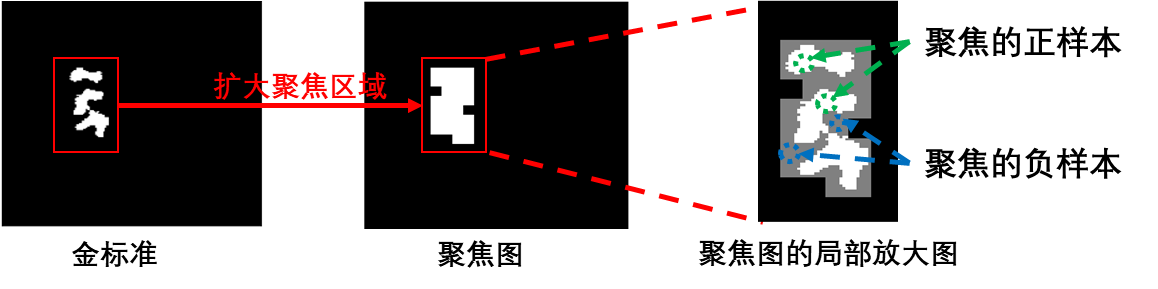


图1.3 目标感知损失中聚焦样本的扩张图

Figure 1.3 Expansion plot of focused samples in target-aware loss

## 1.4 论文组织结构

本文主要研究基于多示例学习的弱监督结肠息肉检测方法，主要涉及多示例学习以及弱监督技术。全文共有五章，结构安排如下：

第1章：绪论。本章首先介绍了结肠图像自动化检测对于提高息肉识别率、辅助临床医生工作的重要意义，并且介绍了弱监督目标检测方法在息肉检测的应用价值。然后总结并概括了国内外基于全监督的息肉检测方法、基于图像级标签的弱监督图像检测方法：基于类激活图和多示例学习的弱监督目标检测方法，探讨了它们存在的优势以及不足。最后梳理了本文的主要研究内容以及组织结构。

第2章：相关理论与基础。本章首先阐述了息肉相关理论知识，而后介绍息肉检测中的监督方式，然后介绍弱监督方法中常用的多示例学习以及深度学习技术，最后介绍弱监督息肉检测的相关技术。

第3章：形态驱动的弱监督息肉检测网络。本章首先分析已有全监督和弱监督方法在息肉检测领域的不足，引入形态驱动以及弱监督的研究动机，而后提出网络的整体框架，主要包含四个部分：1）跨域参考模块，2）空间类别模块，3）多示例分支，4）双阈值后处理策略，并对每个部分进行详细阐述。最后详细介绍了实验的相关细节，并通过实验证明本文提出网络的有效性，同时通过消融实验验证了创新部分的有效性。

第4章：基于合作的弱监督息肉检测网络。本章首先分析已有弱监督方法和SAM互补的优缺点，引入二者合作的研究动机，而后提出网络的整体框架，主要包含三个部分：1）基于合作的区域生成网络，2）自适应感兴趣区域融合模块，3）多示例检测网络，并对每个部分进行详细阐述。最后详细介绍了实验的相关细节，并通过实验证明本文提出网络的有效性，同时通过消融实验验证了创新部分的有效性。

第5章：总结与展望。本章主要总结了本文所提出的两个创新方法，对其优势进行了概括，同时指出了其中的不足，并对进一步的研究工作进行了展望。

# 第二章 相关理论与基础

## 2.1 息肉相关理论知识

卷积神经网络(CNN)利用卷积的局部感受野[[59](#_ENREF_59)]特性将图像划分为多个区域进行目标特征的提取，是常用于计算机视觉方向的神经网络框架，例如，图像分类，图像分割和目标检测。目前在医学图像分割领域的最先进模型是全卷积神经网络(FCNN)[[60](#_ENREF_60)]的变体，FCNN通常由两个组件组成：编码器和解码器。

在图像分割任务中，编码器通常由若干个卷积层组成，用于从输入图像中提取目标的细节特征[[61](#_ENREF_61)]，例如纹理、边缘和棱角等。然后将卷积后的特征图送入到下采样模块中。例如：最大池化层和平均池化层[[62](#_ENREF_62)]。下采样会降低特征图的分辨率，减小后续卷积的运算量，同时会扩大卷积的全局感受野[[59](#_ENREF_59)]。不同的下采样操作有助于捕获有关目标的不同信息，最大池化有利于捕获目标和边界信息，平均池化有利于捕获背景信息。在下采样后，还可以继续进行卷积和下采样操作，目的是尽可能细化特征表示和扩大卷积的感受野，从而捕获更多的目标的细节信息。

解码器同样由一系列卷积层组成，从编码器输出的小分辨率特征图中提取目标的语义特征，并生成最终预测。由于解码器靠近输出层，这些特征相较于编码器特征更加的抽象，具有较高的辨别性。与编码器不同，解码器中的卷积与上采样模块互相搭配，上采样通常由反卷积层[[63](#_ENREF_63)]和线性插值法[[64](#_ENREF_64)]组成。通过卷积和上采样级联逐渐将编码器生成的小分辨率特征图上采样到其原始大小。解码器的最终输出是单通道或者多通道的预测图，其中每个像素都被分配一个类别概率，并与医生标注的金标准计算损失。

编码器和解码器通常使用反向传播和梯度下降进行端到端训练，以最小化损失函数的方式提取图片信息。总体而言，编码器和解码器是全卷积神经网络的关键组件，它们的级联所构成的网络能够提取图像的复杂特征并生成有意义的输出。

## 2.2 检测中的监督方式

### 2.2.1 全监督检测

### 2.2.2 半监督检测

### 2.2.3 弱监督检测

**不完全监督**

**不确切监督**

**不精确监督**

单的编码器-解码器结构虽然能够从低层到高层不断的学习细节特征和语义特征，但是池化操作会造成目标部分细节信息的丢失。为了避免信息的丢失，现有的先进编码器-解码器模型U-Net[[15](#_ENREF_15)]利用长距离跳连接将编码器的细节特征和解码器的语义特征相互融合，能够在网络的深层避免细节信息的丢失，在复杂的医学图像中完成精细的分割任务。

U-Net结构图如图2.1 [[15](#_ENREF_15)]所示。左边为编码器，由四层编码模块构成。每个编码模块由两个卷积核为3\*3的卷积、ReLU激活函数和2\*2最大池化组成[[65](#_ENREF_65)]。它们可以通过卷积和池化的步长对特征图进行下采样，逐渐的减小特征图分辨率，并且逐渐的增加卷积核通道数以捕获更多的细节特征。右侧的解码器同样由四个解码块构成，为了生成与金标准同尺寸和同通道的预测图，解码块在增加特征图分辨率的同时减小特征通道的数量。在编码器和解码器的每个模块之间都有长距离的跳连接，通过通道维度的拼接融合了低层和高层的特征。这样可以使解码器卷积在编码器的特征图中提取细节特征，避免了低层细节信息的丢失。

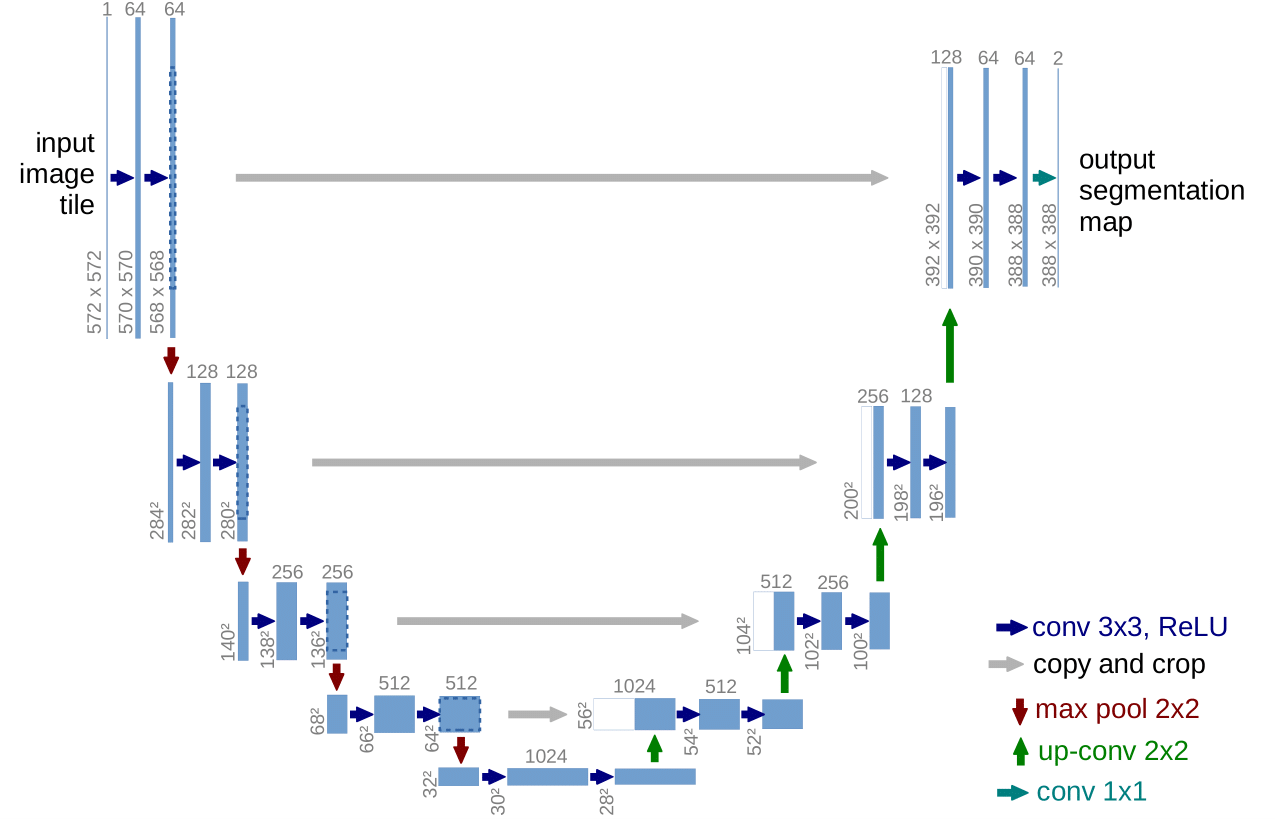


图2.1 U-Net网络模型

Figure 2.1 U-Net network model

U-Net在医学图像分割领域是最受欢迎的网络之一，已经被广泛的应用于器官分割和病灶分割。但是在U-Net模型中，3D的医学图像数据被转换成多个2D平面切片，并将独立的切片送入网络进行训练和预测。这样的作法忽略了连续切片之间的空间维度特征，因此不利于捕获目标的边界信息，预测也是不连续的。虽然完全由3D卷积构成的U-Net网络(3D U-Net)能够提取空间维度信息，但是3D CNN计算量和存储量要求过大，因此没有被广泛的应用。

为了能够捕获空间维度特征同时降低模型复杂性，Zhou等人[[26](#_ENREF_26)]提出了D-UNet模型使3D卷积提取的空间维度信息能够在以2D卷积为主体的网络框架中得到应用。如图2.2 [[26](#_ENREF_26)]所示，网络的基本架构由U-Net组成，不同的是D-UNet的编码器由2D和3D两种维度的卷积构成，所以网络的输入是四张连续的切片。2D卷积和3D卷积在完成特征提取后，通过红色模块进行特征融合，使得后续的2D网络能够融入3D的空间维度信息，细化目标的边界特征。特别的是网络的输入只有四张连续的切片，而且在解码器不使用3D卷积，这避免了模型计算量大幅度提升。

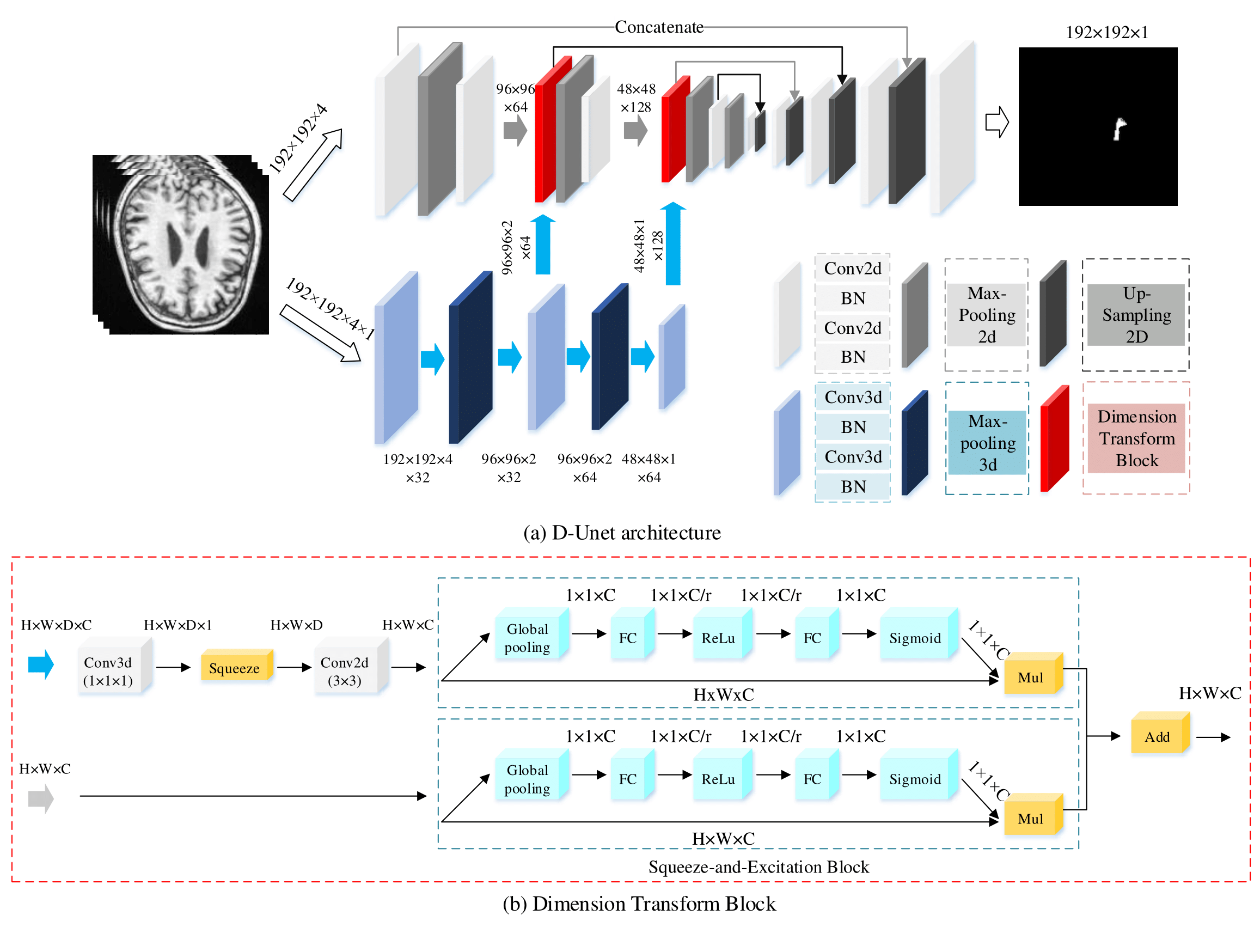


图2.2 D-UNet网络模型

Figure 2.2 D-UNet network model

编码器-解码器结构非常灵活，可以根据不同的分割任务和挑战进行修改。例如，Attention Unet使用注意力机制，进一步的提高分割的准确率。注意力机制模仿人类视觉对外界事物的聚焦行为，从而快速的获取感兴趣信息或任务相关的信息，降低无关信息的干扰。对于卷积神经网络而言，注意力机制把不同的权重赋予特征图中不同的区域，对于任务高度相关的区域分配较高的权重，对于任务不相关区域分配低权重。通过不同的加权值，可以突出特征图中的感兴趣区域或者重要区域。这使得模型有区别的对待特征图中信息，卷积在需要关注的区域获得高响应值，在不需要关注的区域获得低响应值，从而有效的使得模型更多的关注任务相关区域。注意力模型在涉及处理医学图像中的复杂组织或易混淆的边界时特别有用，例如，在胰腺分割中，Attention Unet使用注意力集中在胰腺区域而忽略腹部其他组织。

注意力通过对不同维度的加权方式实现了不同效果的注意力模型：空间（位置）维度注意力[[36](#_ENREF_36),[37](#_ENREF_37),[39](#_ENREF_39)]和通道维度[[26](#_ENREF_26),[32](#_ENREF_32),[33](#_ENREF_33)]注意力。如图2.3 [[33](#_ENREF_33)]所示，空间注意力(Spatial attention)机制根据输入特征图生成注意力图，注意力图中每个像素点的值对应输入特征图中每个像素点的权重值。通过对不同像素点分配不同的权重，使模型选择性地关注输入图像的空间（位置）区域。在实际应用中，空间注意力也可分为正向注意力和反向注意力。正向注意力用于增强特征图中目标信息，促进网络提取目标特征。而反向注意力用于抑制特征图已经识别出的目标信息，增强未识别的残差目标信息，促进网络提取残差特征。通道注意力(Channel attention)机制建模多通道特征图中的依赖性，生成对应每个特征图通道的权重，增强与任务最相关的通道特征的表示。

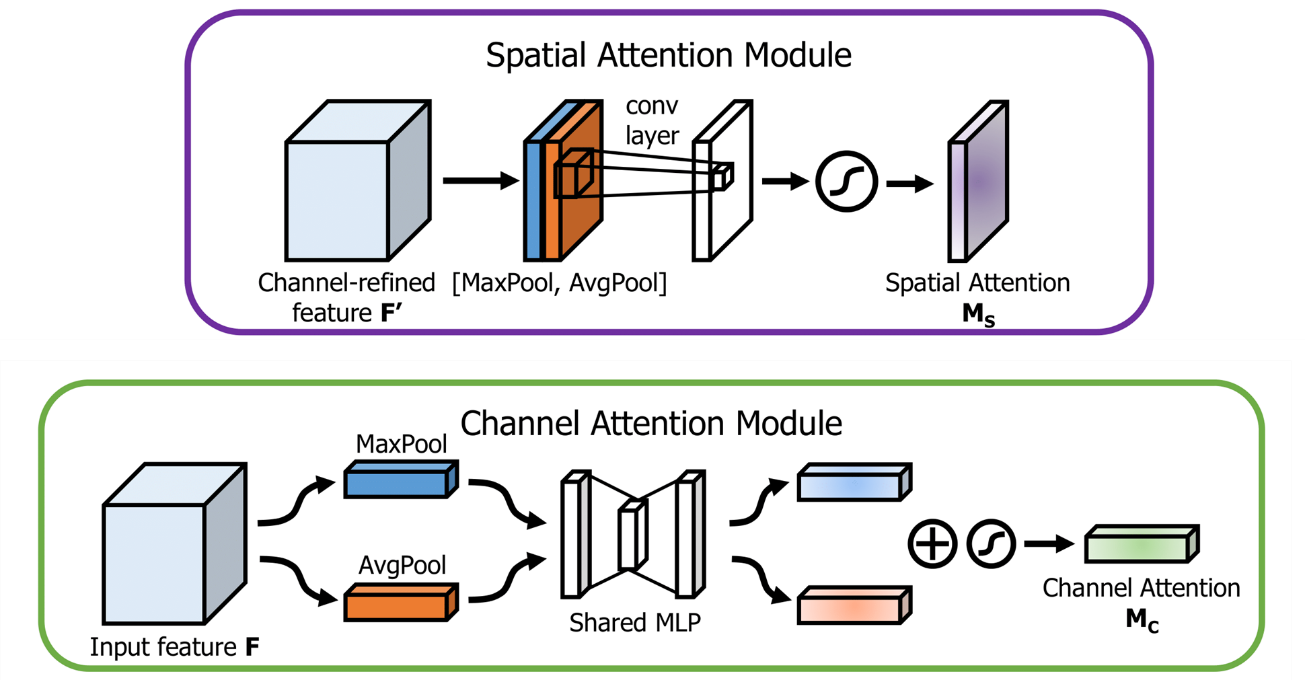


图2.3 通道注意力和空间注意力

**Figure 2.3 Channel attention and spatial attention**

在最近的研究中，通常结合空间注意力和通道注意力。这样不仅选择性的关注输入特征图中特定的区域，还可以对特征通道进行不同程度关注，显著的提高了编码器-解码器模型在医学图像中的准确性和鲁棒性[[66](#_ENREF_66)]。

## 2.4 深度学习技术

### 2.4.1 卷积神经网络的结构

### 2.4.2 经典的弱监督目标检测

**WSDDN**

**OICR**

## 2.5 弱监督息肉检测相关技术

### 2.5.1 损失函数

### 2.5.2 非极大值抑制

### 2.5.3 评价指标

当设计一个用于分割医学图像的深度学习模型时，损失函数的选择非常重要，因为它们会刺激算法的反向传播过程[[67](#_ENREF_67)]。在二分类的分割任务中，较为先进的损失函数会在训练期间为假阳性和假阴性分配不同的惩罚力度，也即正负样本损失的权重。例如，增加正样本损失的权重就会增加对假阴性的惩罚[[68](#_ENREF_68)]，会促进网络提取正样本信息以减少假阴性预测；增加负样本损失的权重会增加对假阳性的惩罚。所以损失函数决定了网络调节假阴性和假阳性的能力。

我们在表2.1中列出了常用于医学图像的损失函数的名称和公式, 表中公式的y表示医生标注的二进制金标准，p表示模型的预测值。表示正负样本损失的静态权重。表示focal loss中正负样本损失的动态权重，最后用来防止分母为0。二进制交叉熵损失函数(Binary cross-entropy loss)对图像中的样本损失进行同等学习，得到的训练结果将由拥有更多数量的类主导，这无法应对医学图像中经常出现的类不平衡现象[[55](#_ENREF_55)]。更进一步，加权交叉熵(Weighted cross-entropy loss)[[69](#_ENREF_69)]提出了正样本损失的权重，设置 >1以减少假阴性， <1以减少假阳性。平衡交叉熵(Balanced cross-entropy loss)[[67](#_ENREF_67)]也被用来对负样本进行加权，可以用来调整假阴性和假阳性。Focal loss[[58](#_ENREF_58)]在平衡交叉熵损失函数的基础上进一步提出了动态权重，使简单样本得到的损失变小，困难样本的损失变大，从而加强了对困难样本的关注，使模型能够更专注于学习困难样本。上述方法都是基于像素分类正确与否计算出的损失值，从评估指标上看，这不一定能带来好结果。

表2.1 医学图像分割中常用损失函数

Table 2.1 Common loss functions in medical image segmentation

|  |  |
| --- | --- |
| 方法 | 公式 |
| Binary Cross-Entropy loss |  |
| Dice Loss |  |
| Weighted cross-entropy loss |  |
| Balanced Cross-Entropy loss | *,* |
| Focal loss | *,* |
| Enhanced Mixing Loss |  |
| Tversky Loss |  |

在图像分割领域的评价中通常使用dice指标。在样本极度不平衡的情况下，dice loss也被使用[[70](#_ENREF_70)]。不幸的是，dice loss的一个局限性是，它对假阳性(FP)和假阴性(FN)的惩罚力度相同[[57](#_ENREF_57)]。Tversky loss[[57](#_ENREF_57)]在dice loss的基础上为FP和FN引入了两个权重，灵活地调节召回率和精确率。但是，由于dice系数本身是非凸的[[67](#_ENREF_67)]，因此它被广泛地与改进的二进制交叉熵损失相结合，这样即利用了dice loss对于dice指标的贡献，同时使用交叉熵损失函数改进了梯度下降的平滑度[[71](#_ENREF_71)]。Focal loss和使用对数运算的dice loss被整合到enhanced mixing loss(EML loss) [[26](#_ENREF_26)]中，以获得更快的收敛率。然而，focal loss的最佳参数设置导致分割结果的假阳性很高。为了提高召回率，应该增加正样本损失的权重来增加对假阴性样本的惩罚，但是我们认为权重的设置不应该过度偏向于正样本损失，而应对负样本损失进行补偿[[68](#_ENREF_68)]。

## 2.6 本章小结

本章中，我们首先介绍编码器-解码器分割网络结构的基本概念，然后，对本文所设计的相关算法的基本概念和改进思路进行介绍，包括医学图像分割网络(U-Net, D-UNet)，注意力机制（正向和反向空间注意力机制，通道注意力机制）和损失函数。

# 第三章 形态驱动的弱监督息肉检测网络

## 3.1 研究动机

在本章中，我们提出了一个新颖的注意力引导多尺度恢复网络，如图3.1所示。由于脑卒中数据集是3D数据，所以我们选择D-UNet作为我们的骨干网络，网络的输入是四张连续的切片，分别送入2D卷积和3D卷积中进行特征的提取。为了更准确的脑卒中分割，首先，我们引入了一个粗粒度图像块注意力(CPA)模块，它生成了一个粗粒度的注意力图，突出了目标的空间信息，引导网络挖掘目标特征，解决了类内不一致的问题。并且基于标签生成的先验位置图作为先验知识有利于粗粒度图像块注意力的监督训练。其次，为了完善2D特征的边界信息，我们在2D卷积和3D卷积的交汇处提出了一个跨维度特征融合(CFF)模块融合2D和3D特征，它自适应的选择有利于分割的不同维度的互补特征。最后，在解码阶段设计了一个多尺度反卷积上采样(MDU)模块来捕捉多尺度的上下文信息，并恢复多次池化丢失的目标空间和边界信息。

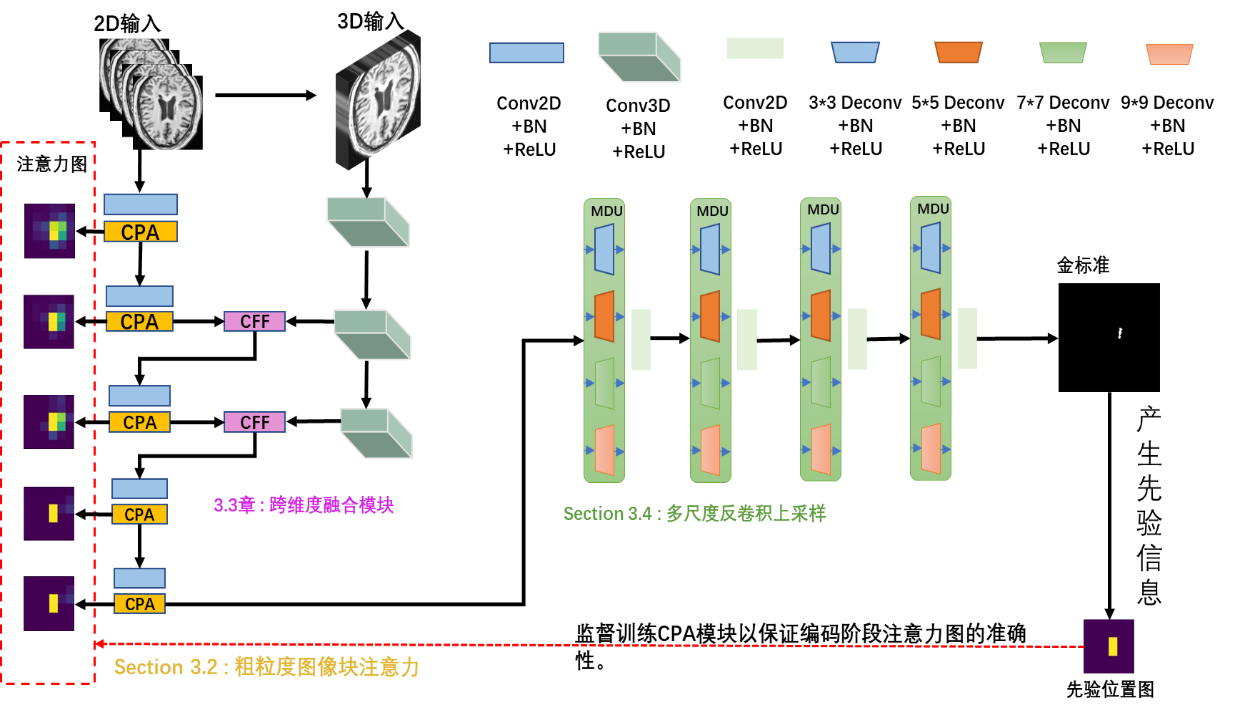


图3.1 注意力引导多尺度恢复网络模型

Figure 3.1 Attention-guided multiscale recovery network model

## 3.2 MDNet模型

### 3.2.1 跨域参考模块

### 3.2.2 空间类别模块

### 3.2.3 多示例分支

### 3.2.4 双阈值后处理策略

在网络的2D编码层，我们利用先验知识设计粗粒度图像块注意力，以增强特征图中的目标位置线索，减少背景中的噪声信息，从而引导网络提取高质量的特征。接下来，我们将介绍粗粒度图像块注意力的细节。

2D编码中每个阶段的卷积生成的特征图（图3.2(a)）被输入到模块中，以生成相同分辨率的粗粒度注意力图（图3.2 (b)）。如图3.2 (b)和(c)所示，我们使用图像块级别的空间注意力，注意力图描述了特征图中不同图像块中存在目标的概率，揭示了在特征图中需要增强或抑制哪些图像块。但是在进行权重评估时，由于卷积的局部感受野太小，不能获得所有图像块之间的空间上下文关系。所以我们首先将图像块全局空间信息挤压为图像块描述符[[33](#_ENREF_33)]，由于在是将输入1\*H\*W的图像等分成36个图像块，所以挤压操作如下所示：

如公式(3.1)所示，其中表示第i层编码阶段两层卷积输出的特征图，C、H和W分别是特征图的通道数、高度和宽度。我们沿通道维度使用平均池化获得单通道的特征图，这在对于突出目标区域方面是有效的[[33](#_ENREF_33)]。然后，如公式(3.2)所示，为了聚合空间信息以获得特征图中每个图像块的全局信息，我们使用池化核大小为的平均池化生成 。每个池化核覆盖一个图像块并生成图像块描述符。

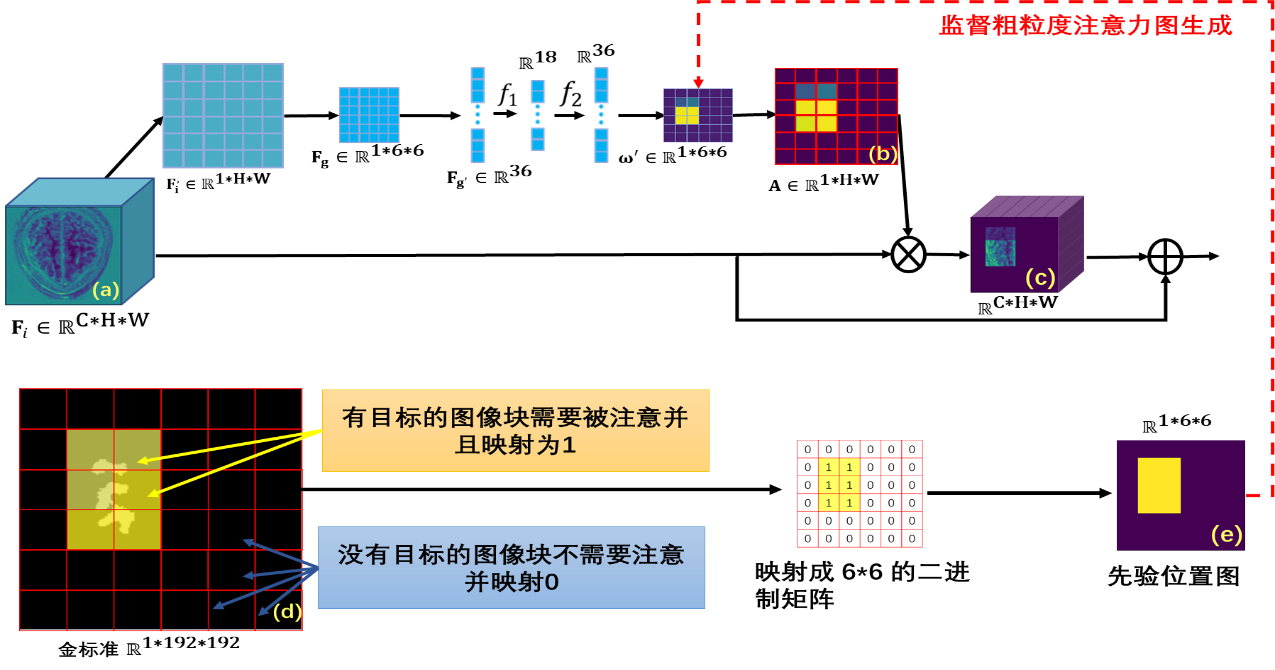


图3.2 粗粒度图像块注意力结构图

Figure 3.2 Coarse-grained patch attention structure figure

我们使用多层全连接神经网络来显式建模[[72](#_ENREF_72)]所有图像块之间的依赖关系，如公式(3.3)，其中表示两层全连接神经网络，表示sigmoid激活函数。因此，被展开为并送入多层全连接神经网络中，中间层和最后一层分别有18个和36个神经元，并通过sigmoid函数映射为0到1之间的连续值。中的值表示每个图像块存在目标的概率，也即该图像块需要被关注的程度。概率越高说明图像块存在目标的概率越大，应该增强概率高的图像块的特征表示，引导网络提取目标的特征。

将得到的被重塑为概率图，并通过线性插值上采样恢复大小为1\*H\*W粗粒度图像块注意力图，其中上采样公式如下所示：

此时把粗粒度图像块注意力图A与输入的编码器特征图逐元素相乘，会将含有目标区域的图像块赋予高权重，而不含目标的背景图像块赋予低权重。这将促进网络专注于含有目标的高权重图像块，有利于卷积提取特定于目标的特征。与基于像素点的细粒度注意力图中权重分数不同，我们的权重分数是基于图像块。例如，如图3.3所示，我们可以看到，在上采样后的4 \* 4粗粒度注意力图中，每个2\*2的图像块中所有像素点的权重得分是相同的。每个编码阶段的最终输出为。注意，我们使用残差网络的思想将注意力后特征图添加到输入特征图中，这提高了它们的容错能力。

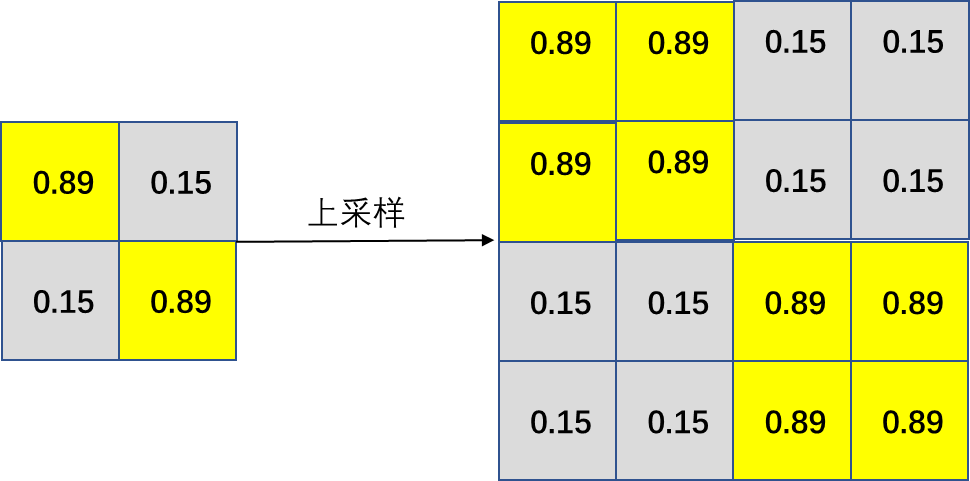


图3.3 注意力图的上采样过程

Figure 3.3 The upsampling process of the attentional map

为了保证生成的注意力的准确性，粗粒度图像块注意力模块以直接显式监督的方式训练学习，我们在金标准上使用通道上的最大池化生成一个1\*6\*6的二进制注意力矩阵。如图3.2 (d)所示金标准大小为192\*192，每个池化核大小为32 \*32，当有病变区域时，每个池化核覆盖的金标准上图像块映射为1，表示该图像块需要被注意，如果没有病变则映射为0。生成的先验图如图3.2 (e)所示，将二进制注意力矩阵可视化为先验位置图，我们将其作为先验知识，并用二进制交叉熵损失监督粗粒度注意力图的生成，从而获得更准确的预测结果。更重要的是，如图1.1所示，尽管目标特征随着池化而变得抽象，但目标在特征图中的相对位置不会改变。因此，我们通过池化获得的先验信息，可以监督每个阶段的粗粒度图像块注意力图的生成，以确保注意力的准确性。我们在粗粒度图像块注意力的消融实验（3.6.3章）中可视化了粗粒度注意力对特征图的影响。

## 3.3 实验细节

### 3.3.1 数据集

### 3.3.2 网络结构和参数

### 3.3.3 损失函数

### 3.3.4 评估指标

### 3.3.5 对比算法介绍

虽然粗粒度图像块注意力帮助网络定位目标所在的图像块位置，从中提取特定于目标的特征。但是仍然受到2D卷积不能捕获3D空间维度信息的限制。2D和3D特征的跨维度特征融合模块实现了两个不同维度特征的自适应融合，取代了简单的直接加法和串联，细化了2D特征图的边界特征。跨维度特征融合模块以2D特征图和3D特征图为输入。在每个跨维度特征融合模块中，我们首先对3D特征图进行降维，然后对2D和3D维度间的特征图进行建模，并研究两个维度所有特征通道之间的相关性，以增强网络对不同维度特征的敏感度。

如图3.4所示，为了降低3D特征图的维数，我们使用一个卷积核大小1\* 1\*1的单通道的3D卷积将输入的3D特征图转换为单通道。然后，在空间维度上被降维成，并把 维度置换成为具有D个通道的。为了保持通道数与2D特征图中通道数一致，经过卷积核大小3 \*3的C通道2D卷积运算，得到有C通道的特征图。 2D特征图和 被送到后续计算中。

从图3.4可以看出，使用全局平均池化将和的每个通道的全局信息压缩到通道描述符中，并通过多层全连接神经网络捕获通道间的依赖关系。如公式(3.5)所示我们串联不同维度的特征描述符获得，并送入多层全接连神经网络，对不同维度上所有特征通道的相关性进行建模。为了减少计算，中间层的神经元数量是C/R，其中R是衰减率，最后一层有C个神经元。每个维度的通道权重是在经过sigmoid 激活函数映射后获得的，如公式(3.6)和公式(3.7)所示。 表示两层全连接神经网络，表示sigmoid 激活函数， ©表示串联。两种维度的所有通道描述符都包含在中，因此多层全接连神经网络学习两种维度之间所有通道之间的关系。



图3.4 跨维度特征融合模块的结构图

Figure 3.4 Structure of the cross-dimensional feature fusion module

获得的权重用于重新映射和并获得具有融合2D 和 3D 信息的特征图。跨维度特征融合模块应用不同维度特征的全局上下文信息，如果融合时通道的权重高，则应强调其特征表示。在特征融合中，通过增强不同维度的重要特征，抑制不相关的特征，从而增强通道特征图之间的上下文语义依赖性以及融合时特征图的判别力。

## 3.4 实验结果以及分析

### 3.4.1与现有方法的比较

为

### 3.4.2 消融实验：跨域参考模块

### 3.4.3 消融实验：空间类别模块

### 3.4.4 消融实验：双阈值后处理策略

在U形结构的医学图像分割网络中，无参数的线性插值和简单的反卷积不能有效的恢复因编码阶段连续池化而丢失的空间和边界信息，使得分割具有挑战性。为了解决这个问题，我们提出了一个多尺度反卷积上采样模块，该模块使用多尺度上下文信息来恢复病变区域的空间和边界特征。

在图3.5中，9\*9的白蓝相间特征图是反卷积过程中填充后的特征图，白色区域是反卷积填充的0值，蓝色表示原始输入特征图的有效区域，灰色区域表示3\*3卷积核覆盖的区域，粉色区域是5\*5卷积覆盖的区域。图3.5 (a-c)表示3\*3卷积移动过程。在灰色区域，只有不到一半的有效信息，这无疑是不利于目标区域细节特征的恢复，并且这对于小目标来说是致命的。比较图3.5中(a)–(c)，我们可以发现，在3\*3卷积过程中，由于感受野的限制，例如图3.5 (d)中的红色像素点1-4之间的关系被忽略了。但是，这些像素点可以被5\*5反卷积覆盖。5 \* 5反卷积也有类似3\*3反卷积的缺点，但7 \* 7反卷积可以弥补它。因此，我们使用了多个具有不同卷积核大小的反卷积，更好地恢复病变区域细节特征和改善分割结果的整体结构。考虑到上采样的计算量和特征图的大小，我们只使用3\*3、5\*5、7\*7和9\*9的反卷积。

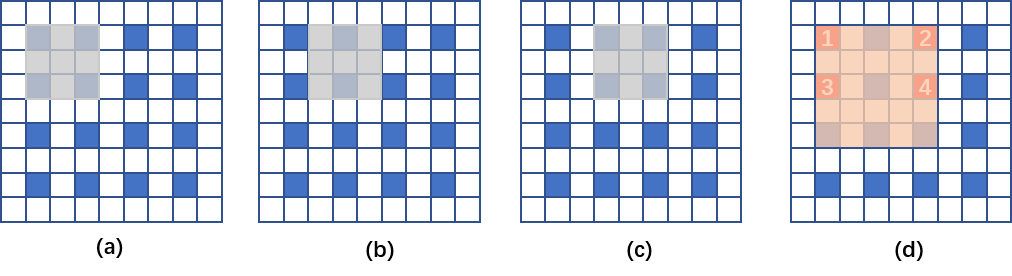


图3.5 3\*3反卷积的缺陷表示图

Figure 3.5 Disadvantage representation figure of 3\*3 deconvolution

如图3.1所示，对于模块的输入，我们使用不同卷积核大小的反卷积来增加特征图分辨率。具体来说，我们使用 3\*3、5\*5、7\*7和9 \* 9 的卷积核，放大后的特征图大小为。我们将它们串联为。每个反卷积之后是批量归一化(BN)、ReLU 和 dropout 操作，其中 dropout 层可以缓解因多尺度卷积大量参数造成的过拟合。

## 3.5 本章小结

# 第四章 基于合作的弱监督息肉检测网络

## 4.1 研究动机

在医学图像分割中，假阴性的预测可能是致命的。因为在实际的临床中如果不把病灶区域切除干净，这可能导致病变的再次复发，影响疾病的治疗和痊愈。所以在本章节，我们更注重减少模型预测结果的假阴性，同时避免产生过多的假阳性。为了解决类不平衡和类间不一致造成的假阴性偏高问题，在AGMR-Net的基础上，我们提出了目标感知监督残差学习网络(TSRL-Net)，能够提高脑卒中语义分割的召回率，同时平衡精确率。图4.1显示了TSRL-Net的整体结构。

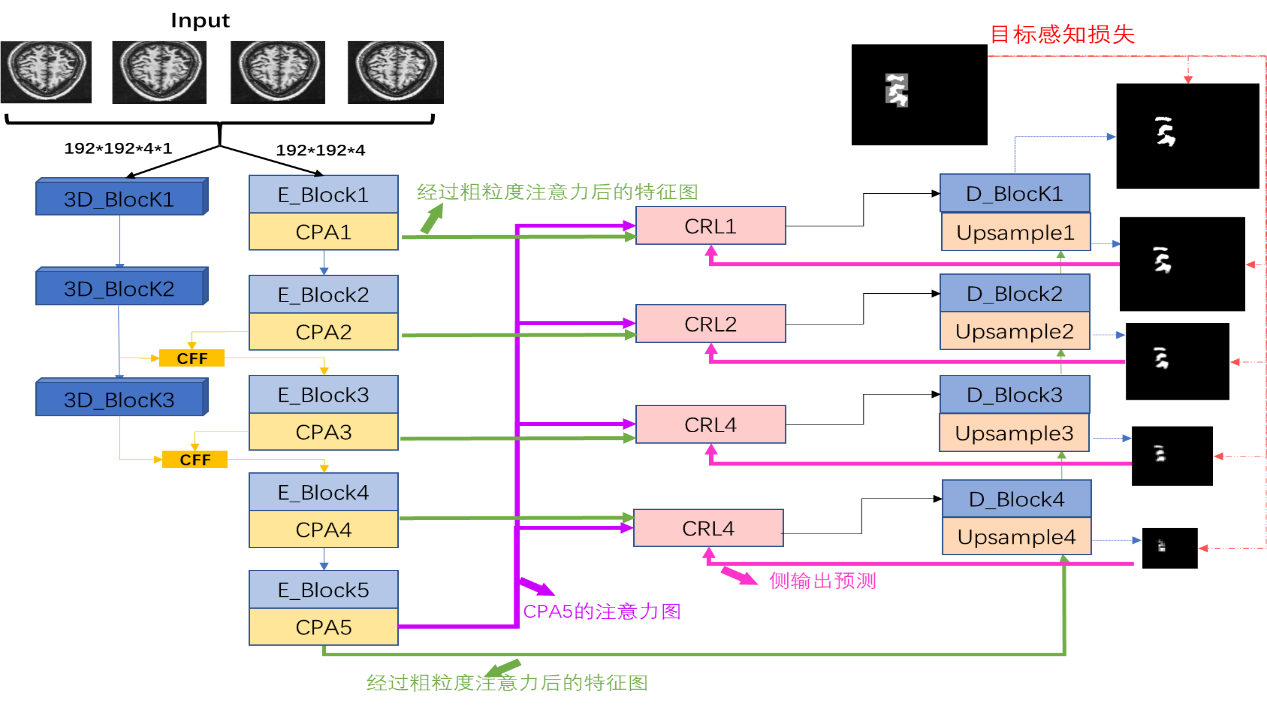


图4.1 目标感知监督残差学习网络结构

Figure 4.1 Target-aware supervised residual learning network structure

因为2D和3D的融合网络可以在编码阶段充分学习目标边界细节特征。网络的2D编码器(E\_Block)有一个192\*192\*4的四通道输入（四个相邻的切片，每个待分割的真实图像的前一个切片和后两个切片），经过维度扩展后作为单通道192\*192\*4\*1的输入输入到三维编码器(3D\_Block)。每个E\_Block的输出由粗粒度图像块注意力模块处理，该模块用粗粒度图像块级注意力增强目标所在图像块的特征表示，并使用跨维度特征融合模块自适应融合3D\_Block的特征，以微调目标边界特征的表示。然后，在通过上采样恢复目标的分辨率后，将其送入二维解码器(D\_Block)以学习目标的高级语义特征。每个上采样输出一个由目标感知损失函数监督的侧输出预测图，指导网络挖掘目标信息。已经预测的目标信息将在下阶段被粗粒度残差学习模块中的反向注意力单元抑制，剩余（残差）的目标信息通过反向和正向注意力得到增强，背景噪声被抑制，然后被发送到D\_Block。通过从下到上循环进行目标监督和目标残差学习，以提高预测的召回率和精确率，从而提高网络性能。

## 4.2 CBNet模型

### 4.2.1 基于合作的区域生成网络

### 4.2.2 自适应感兴趣区域融合模块

### 4.2.3 多示例检测网络

使用跳连接可以让解码器再次学习编码器特征图中目标的细节特征[[15](#_ENREF_15)]，达到提升召回率的目的。但如图4.2 (a)所示，编码器特征图不仅包含解码器已经识别的目标区域，还包含大量的背景噪声。这些信息阻碍了网络对目标残差特征的提取，不利于应对类内不一致的挑战。为此，我们提出了一个粗粒度残差学习模块，它由粗粒度的正向注意(PA)和反向注意(RA)单元组成。



图4.2 粗粒度残差学习模块结构图

Figure 4.2 Structure diagram of coarse-grained residual learning module

**RA单元**：我们使用RA单元来抑制解码器已识别的目标信息，以避免跳连接过程中已识别信息的冗余，这更有利于网络去挖掘目标的残差特征，以减少假阴性提高召回率，如图4.2中的绿框所示。第n层编码器(E\_Blockn)的特征图和第n层上采样的侧输出作为输入，RA单元的输出可以被定义为。

表示元素相乘。 可以突出显示卷积识别的对象。反向注意力图 由得到。通过观察图4.2 (b)中，我们可以看到，图中已经识别的目标区域的特征被抑制，未预测的残差特征得到了保留。因此，RA单元的目的是增强未预测的目标区域特征，以引导网络发现与缺失目标区域相关的残差特征。

**PA单元**：受第三章粗粒度图像块注意力工作的启发，在本章中，我们使用粗粒度注意图来增强目标所在图像块的特征表示，抑制大量不存在目标的图像块的特征表示，来达到抑制背景噪声的目的。这可以避免了网络受到噪声的干扰，以减少假阳性，平衡精确率。PA单元的粗粒度注意图由CPA5给出，最终PA的输出被定义为：

PA单元对目标和其附近的少量背景样本给予高度关注，但抑制了大量的背景噪声。我们把 和 融合得到最终粗粒度残差学习模块的输出 。

**可视化分析：**图4.2显示了使用粗粒度残差学习过程中特征图的变化。如图4.2(a)所示，编码阶段的特征图不仅包含解码器未识别的目标残差特征，还包含解码已经识别的目标特征和大量的背景噪声。出于这个原因，在RA单元中，我们用反向注意抑制了网络已经识别的目标区域。然而，如图4.2 (b)所示，我们可以看到，所有目标区域之外的噪音都被重新引入到解码阶段，这不可避免地引入了太多的干扰信息。为了更进一步抑制这些干扰信息，我们引入了正向注意力，利用粗粒度的正向注意来增强存在目标的图像块特征信息和抑制不含目标的背景噪音图像块特征。然而，如图4.2 (c)所示，我们可以观察到特征图中解码器已识别的目标特征仍然处于高度注意状态，造成了已识别目标信息的冗余。为此，我们将RA和PA单元结合起来，如图4.2 (d)所示，不仅抑制了已识别的目标特征以避免信息的冗余，还抑制大量的背景特征以避免引入大量的干扰信息。这使得目标的残差区域得到增强，由于网络总是关注特征表示强烈的区域，图4.2 (d)更有利于网络提取残余特征以提高召回率。同时，由于大量的背景噪声被抑制，可以达到平衡精确率的目的。

## 4.3 实验细节

### 4.3.1 数据集

### 4.3.2 网络结构和参数

### 4.3.3 损失函数

### 4.3.4 评估指标

### 4.3.5 对比算法介绍

面对正负样本不平衡的挑战，损失函数必须聚焦于正样本损失，以引导网络挖掘目标的正样本信息。然而，不考虑负样本将导致召回率和精确率的不平衡。我们提出了目标感知损失函数，通过一个可调节的聚焦因子来扩大聚焦样本区域，如图4.3所示，原本只需要关注目标正样本的聚焦区域被扩展到包含了部分负样本。此时关注聚焦区域内所有样本的损失，可以提高对正样本损失的关注度，也即增加对假阴性样本的惩罚，提高召回率。同时，它还可以高度关注了负样本的损失，提高对假阳性样本的惩罚，以平衡精确率。具体算法如下。

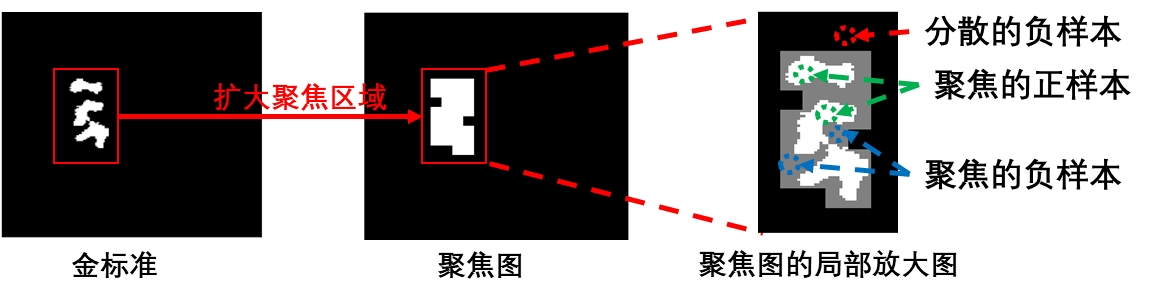


图4.3 损失函数聚焦区域扩张图

Figure 4.3 Loss function focus area expansion map

**样本划分：**我们用β表示聚焦因子的值，用H表示图像的高度，用表示二进制的金标准。A是生成的聚焦图，通过*sign*函数被映射为0或1，1表示聚焦样本，0表示分散样本。然后，划分聚焦和分散样本的过程被定义为：

我们用图4.4来帮助理解聚焦因子的作用。形象地说，聚焦因子控制着聚焦窗口（图4.4中的红色虚线框）的大小以及步长，它以β为单位在金标准中滑动。聚焦窗口在滑动过程中检测窗口内是否存在正样本，如果存在，这个窗口内的所有样本将被设置为聚焦样本（绿色区域）并映射为1。否则，它们将被设置为分散样本（蓝色区域）并被映射为0。

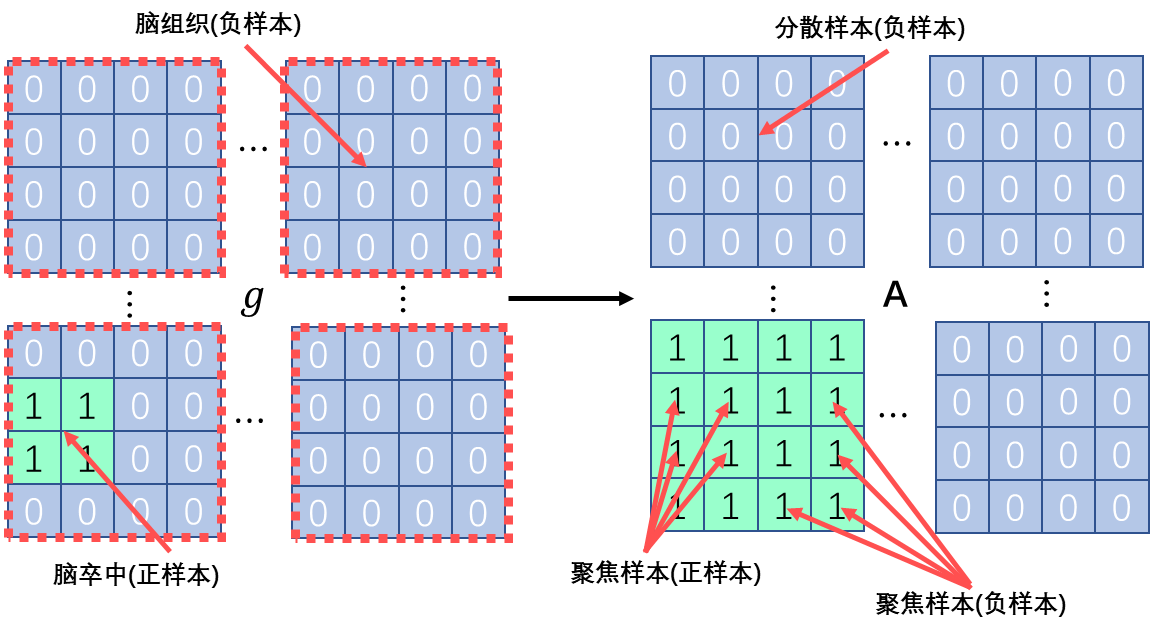


图4.4 聚焦因子的作用图

Figure 4.4 Map of the role of focus factors

**损失计算**：我们将模型的最终预测概率p改写为 ，如公式(4.5)所示。为了实现重点关注正样本损失和补偿负样本损失的目标，我们给聚焦样本中的正样本和负样本都分配了高权重(α>0.5)。设置大于0.5的α可以提升对聚焦样本损失的关注，增加了训练期间对目标及其周围的假阴性和假阳性的惩罚力度，以此来提高召回率和平衡精确率。而对于分散样本，我们分配了一个低权重，即1-α。我们将新的静态权重映射到focal loss上，得到 ，其定义为公式(4.6)。

其中的取值范围为(0, 5)，用于增加对困难样本的惩罚[[58](#_ENREF_58)]。N表示一张金标准图像块中的样本总数。另外，我们还增加了dice loss，以应对极端的样本不平衡问题，因此，目标感知损失定义为：

## 4.4 实验结果及分析

### 4.4.1 与现有方法比较

### 4.4.2 消融实验：基于合作的区域生成网络

我

### 4.4.3 消融实验：自适应感兴趣区域融合模块

## 4.5 本章小结

由于在医学领域，假阴性比假阳性更具有破坏性，因为假阴性可能会延误病情或者导致病灶切除不彻底。所以在本章中，我们更注重于提升分割结果的召回率，同时保证精确率不会大幅下降。为此我们设计了目标感知损失函数和粗粒度残差学习模块来解决假阴性偏多的问题。在实验章节中，我们也证明了我们的方法可以有效的解决类间不一致和类不平衡的问题，有效的减少了分割结果的假阴性预测。

# 总结与展望

随着人们年龄的增长，脑部血管会越加脆弱，极容易引起脑卒中。大脑是带给人类多姿多彩生活的重要器官，所以识别脑卒中病灶十分重要，能够及早的发现脑卒中病灶并进行病情评估。基于上述的医学背景知识，从MRI图像中自动化识别脑卒中并对病灶区域进行分割是十分具有研究意义的。本文首先对脑卒中病灶分割的背景、方法和面临的挑战进行了介绍。然后，对本文中的一些基础网络架构、注意力方法和损失函数进行了阐述。最后针对脑卒中分割所面临的挑战，设计出两个新的分割模型，具体如下：

(1)为了在编码阶段捕获目标的空间和边界信息，我们设计了注意力引导多尺度恢复网络(AGMR-Net)，它整合了粗粒度图像块注意力模块和跨维度特征融合模块。粗粒度图像块注意力使用图像块级别的粗粒度注意图引导网络的注意力集中在目标区域，以提取与目标高度相关的细节特征，解决了编码阶段存在的类内不一致问题。跨维度特征融合模块结合不同维度的全局通道信息自适应地融合2D和3D特征，以细化2D特征图的边界，有助于解决类间模糊问题。多尺度反卷积上采样用于解码阶段，从经过多次下采样的特征图中恢复目标的空间和边界信息。AGMR-Net的主要优势在于使用了不同图像块之间的空间上下文信息、不同维度之间的全局通道上下文信息和多尺度上下文信息，表明了上下文信息的应用对提高分割性能的重要性，与现有方法相比具有很大优势。

(2)在AGMR-Net的基础上，为了更好的解决正负样本不平衡和解码器阶段存在的类内不一致问题，我们提出了目标感知监督残差学习网络(TSRL-Net)。首先，我们提出了一个目标感知损失函数，通过划分聚焦样本和分散样本并对其损失进行不同程度的加权，更好的平衡召回率和精确率。我们进一步结合正向和反向注意力，提出了一个粗粒度残差学习模块，在解码阶段提供目标残差特征的空间信息，帮助网络恢复目标的残差细节信息以提高召回率，同时避免了网络受到背景噪声的干扰以提高精确率。

虽然我们的方法在脑卒中分割领域取得了一定的成果，但是我们的注意力方式和损失函数的设计都是采用了一种粗粒度的思想，仍然存在一些有待解决的问题。而且基于现在的实验方式和数据集的获取方式，我们认为仍然不能应对临床中脑卒中面临的实际问题。所以对未来的研究工作和方向进行了展望：

(1)粗粒度方法的缺陷。因为我们的注意力和损失函数的聚焦区域扩展方式都可以理解为是一种图像块级的粗粒度方式，但是粗粒度有它自身的缺陷，就是面对大目标分割时是无用的。而且在本文中，由于经过四次的池化，最终编码器的特征图大小是12\*12，为了仍然保持一种粗粒度的方式，此时粗粒度图像块的大小为2\*2。所以我们把整个图像等分成36个图像块，但这只局限于192\*192和四次池化的网络。粗粒度图像块的大小选择也是值得探讨的问题，我们认为可以充分的考虑多粒度图像块的信息来辅助分割，例如，每层的注意力粒度不同或者一层内集成多种粒度的注意力。

(2)数据集采集的缺陷。医学图像标注的成本是十分巨大的，不仅涉及病人的隐私，也需要消耗专业医生大量的时间，非医学类的科研人员无法对难识别的病灶进行标注，所以造成了脑卒中数据集的短缺，而少量的数据并不能使模型有效的学习目标的特征。在本文中，我们采用的是公开的脑卒中分割数据集，他们都是经过专家的筛选和标注的。但是在实际的数据集采集中，由于检测的机器不同或者病人的姿势的问题，造成的图像缺陷，我们的方法无疑是不能有效应对这些图像的。所以大量有效和符合临床应用的数据集对于进一步的优化模型和推动模型的实际应用是十分有效的。

# 参考文献

[1] Virani S S, Alonso A, Aparicio H J, et al. Heart disease and stroke statistics—2021 update: a report from the American Heart Association[J]. Circulation, 2021,143(8): e254-e743.

[2] Grysiewicz R A, Thomas K, Pandey D K. Epidemiology of Ischemic and Hemorrhagic Stroke: Incidence, Prevalence, Mortality, and Risk Factors[J]. Neurologic Clinics, 2008,26(4): 871-895.

[3] Sirsat M S, Fermé E, Câmara J. Machine learning for brain stroke: a review[J]. Journal of Stroke and Cerebrovascular Diseases, 2020,29(10): 105162.

[4] Wang Y, Liu H, Liu Y, et al. Deep learning framework for hemorrhagic stroke segmentation and detection[C]. BIBE 2018; International Conference on Biological Information and Biomedical Engineering, 2018: 1-6.

[5] Demeestere J, Wouters A, Christensen S, et al. Review of perfusion imaging in acute ischemic stroke: from time to tissue[J]. Stroke, 2020,51(3): 1017-1024.

[6] Goyal M, Menon B K, Van Zwam W H, et al. Endovascular thrombectomy after large-vessel ischaemic stroke: a meta-analysis of individual patient data from five randomised trials[J]. The Lancet, 2016,387(10029): 1723-1731.

[7] Zhang L, Song R, Wang Y, et al. Ischemic stroke lesion segmentation using multi-plane information fusion[J]. IEEE Access, 2020,8: 45715-45725.

[8] Alquhayz H, Tufail H Z, Raza B. The multi-level classification network (MCN) with modified residual U-Net for ischemic stroke lesions segmentation from ATLAS[J]. Computers in Biology and Medicine, 2022,151: 106332.

[9] Liu C, Zeng X, Liang K, et al. Improved Brain Lesion Segmentation with Anatomical Priors from Healthy Subjects[C]. International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI), 2021: 186-195.

[10] Prakash K N B, Gupta V, Hu J, et al. Automatic processing of diffusion-weighted ischemic stroke images based on divergence measures: slice and hemisphere identification, and stroke region segmentation[J]. International Journal of Computer Assisted Radiology and Surgery, 2008,3(6): 559-570.

[11] Mitra J, Bourgeat P, Fripp J, et al. Lesion segmentation from multimodal MRI using random forest following ischemic stroke[J]. NeuroImage, 2014,98: 324-335.

[12] Mujumdar S, Varma R, Kishore L T. A novel framework for segmentation of stroke lesions in Diffusion Weighted MRI using multiple b-value data[C]. Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012), 2012: 3762-3765.

[13] Chan T F, Vese L A. Active contours without edges[J]. IEEE Transactions on Image Processing, 2001,10(2): 266-277.

[14] Qin X, Fan D-P, Huang C, et al. Boundary-Aware Segmentation Network for Mobile and Web Applications[J]. ArXiv, 2021,abs/2101.04704.

[15] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation[C]. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, 2015: 234-241.

[16] Yang H, Huang W, Qi K, et al. CLCI-Net: Cross-Level Fusion and Context Inference Networks for Lesion Segmentation of Chronic Stroke[C]. Medical Image Computing and Computer Assisted Intervention – MICCAI 2019, 2019: 266-274.

[17] Chen L-C, Papandreou G, Schroff F, et al. Rethinking Atrous Convolution for Semantic Image Segmentation[J]. ArXiv, 2017,abs/1706.05587.

[18] Hochreiter S, Schmidhuber J. Long short-term memory[J]. Neural computation, 1997,9(8): 1735-1780.

[19] Sinha A, Dolz J. Multi-Scale Self-Guided Attention for Medical Image Segmentation[J]. IEEE Journal of Biomedical and Health Informatics, 2021,25(1): 121-130.

[20] Qi K, Yang H, Li C, et al. X-Net: Brain Stroke Lesion Segmentation Based on Depthwise Separable Convolution and Long-Range Dependencies[C]. Medical Image Computing and Computer Assisted Intervention – MICCAI 2019, 2019: 247-255.

[21] Chollet F. Xception: Deep Learning with Depthwise Separable Convolutions[C]. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017: 1800-1807.

[22] Khezrpour S, Seyedarabi H, Razavi S N, et al. Automatic segmentation of the brain stroke lesions from MR flair scans using improved U-net framework[J]. Biomedical Signal Processing and Control, 2022,78: 103978.

[23] 贾小慧, 张雪英, 王夙喆, 等. 基于双向循环U-Net模型的脑卒中病灶分割方法[J]. 太原理工大学学报, 2022,53(06): 1127-1133.

[24] Zaremba W, Sutskever I, Vinyals O. Recurrent Neural Network Regularization[J]. ArXiv, 2014,abs/1409.2329.

[25] Liu L, Kurgan L, Wu F-X, et al. Attention convolutional neural network for accurate segmentation and quantification of lesions in ischemic stroke disease[J]. Medical Image Analysis, 2020,65: 101791.

[26] Zhou Y, Huang W, Dong P, et al. D-UNet: A Dimension-Fusion U Shape Network for Chronic Stroke Lesion Segmentation[J]. IEEE/ACM Transactions on Computational Biology and Bioinformatics, 2021,18(3): 940-950.

[27] Akal O, Peng Z, Hermosillo G. ComboNet: Combined 2D & 3D Architecture for Aorta Segmentation[J]. ArXiv, 2020,abs/2006.05325.

[28] Yalçın S, Vural H. Brain stroke classification and segmentation using encoder-decoder based deep convolutional neural networks[J]. Computers in Biology and Medicine, 2022,149: 105941.

[29] Wang H, Cao J, Feng J, et al. Mixed 2D and 3D convolutional network with multi-scale context for lesion segmentation in breast DCE-MRI[J]. Biomedical Signal Processing and Control, 2021,68: 102607.

[30] Fang C, Li G, Pan C, et al. Globally Guided Progressive Fusion Network for 3D Pancreas Segmentation[C]. Medical Image Computing and Computer Assisted Intervention – MICCAI 2019, 2019: 210-218.

[31] Wu H, Wang W, Zhong J, et al. SCS-Net: A Scale and Context Sensitive Network for Retinal Vessel Segmentation[J]. Medical Image Analysis, 2021,70: 102025.

[32] Hu J, Shen L, Albanie S, et al. Squeeze-and-Excitation Networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020,42(8): 2011-2023.

[33] Woo S, Park J, Lee J-Y, et al. CBAM: Convolutional Block Attention Module[C]. Computer Vision – ECCV 2018, 2018: 3-19.

[34] Wang Q, Wu B, Zhu P, et al. ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks[C]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020: 11531-11539.

[35] Song Y, Du X, Zhang Y, et al. Two-stage segmentation network with feature aggregation and multi-level attention mechanism for multi-modality heart images[J]. Computerized Medical Imaging and Graphics, 2022,97: 102054.

[36] Dou H, Karimi D, Rollins C K, et al. A Deep Attentive Convolutional Neural Network for Automatic Cortical Plate Segmentation in Fetal MRI[J]. IEEE Transactions on Medical Imaging, 2021,40(4): 1123-1133.

[37] Li Y, Li H, Fan Y. ACEnet: Anatomical context-encoding network for neuroanatomy segmentation[J]. Medical Image Analysis, 2021,70: 101991.

[38] Wang F, Jiang M, Qian C, et al. Residual Attention Network for Image Classification[C]. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017: 6450-6458.

[39] Liu Y-C, Shahid M, Sarapugdi W, et al. Cascaded atrous dual attention U-Net for tumor segmentation[J]. Multimedia Tools and Applications, 2021,80(20): 30007-30031.

[40] Yu C, Wang J, Gao C, et al. Context Prior for Scene Segmentation[C]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020: 12413-12422.

[41] Huang J, Shen H, Chen B, et al. Segmentation of Paraspinal Muscles at Varied Lumbar Spinal Levels by Explicit Saliency-Aware Learning[C]. Medical Image Computing and Computer Assisted Intervention – MICCAI 2020, 2020: 652-661.

[42] Tang Y, Yan K, Cai J, et al. Lesion Segmentation and RECIST Diameter Prediction via Click-Driven Attention and Dual-Path Connection[C]. Medical Image Computing and Computer Assisted Intervention – MICCAI 2021, 2021: 341-351.

[43] Milletari F, Navab N, Ahmadi S A. V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation[C]. 2016 Fourth International Conference on 3D Vision (3DV), 2016: 565-571.

[44] Zhou Z, Rahman Siddiquee M M, Tajbakhsh N, et al. UNet++: A Nested U-Net Architecture for Medical Image Segmentation[C]. Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, 2018: 3-11.

[45] Zhang K, Sun M, Han T X, et al. Residual Networks of Residual Networks: Multilevel Residual Networks[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2018,28(6): 1303-1314.

[46] Fan D-P, Ji G-P, Zhou T, et al. PraNet: Parallel Reverse Attention Network for Polyp Segmentation[C]. Medical Image Computing and Computer Assisted Intervention – MICCAI 2020, 2020: 263-273.

[47] Zhang R, Li G, Li Z, et al. Adaptive Context Selection for Polyp Segmentation[C]. Medical Image Computing and Computer Assisted Intervention – MICCAI 2020, 2020: 253-262.

[48] Yu C, Wang J, Peng C, et al. Learning a Discriminative Feature Network for Semantic Segmentation[C]. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018: 1857-1866.

[49] Qin Y, Kamnitsas K, Ancha S, et al. Autofocus Layer for Semantic Segmentation[C]. Medical Image Computing and Computer Assisted Intervention – MICCAI 2018, 2018: 603-611.

[50] Oktay O, Schlemper J, Folgoc L L, et al. Attention U-Net: Learning Where to Look for the Pancreas[J]. ArXiv, 2018,abs/1804.03999.

[51] Shi X, Chen Z, Wang H, et al. Convolutional LSTM Network: a machine learning approach for precipitation nowcasting[C].Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 1,2015: 802–810.

[52] Zhang Y, Li H, Du J, et al. 3D Multi-Attention Guided Multi-Task Learning Network for Automatic Gastric Tumor Segmentation and Lymph Node Classification[J]. IEEE Transactions on Medical Imaging, 2021,40(6): 1618-1631.

[53] Zhao H, Shi J, Qi X, et al. Pyramid Scene Parsing Network[C]. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017: 6230-6239.

[54] Wang G, Liu X, Li C, et al. A Noise-Robust Framework for Automatic Segmentation of COVID-19 Pneumonia Lesions From CT Images[J]. IEEE Transactions on Medical Imaging, 2020,39(8): 2653-2663.

[55] Wang L, Wang C, Sun Z, et al. An Improved Dice Loss for Pneumothorax Segmentation by Mining the Information of Negative Areas[J]. IEEE Access, 2020,8: 167939-167949.

[56] Ma J, Chen J, Ng M, et al. Loss odyssey in medical image segmentation[J]. Medical Image Analysis, 2021,71: 102035.

[57] Abraham N, Khan N M. A Novel Focal Tversky Loss Function With Improved Attention U-Net for Lesion Segmentation[C]. 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), 2019: 683-687.

[58] Lin T Y, Goyal P, Girshick R, et al. Focal Loss for Dense Object Detection[C]. 2017 IEEE International Conference on Computer Vision (ICCV), 2017: 2999-3007.

[59] Luo W, Li Y, Urtasun R, et al. Understanding the effective receptive field in deep convolutional neural networks[C].Proceedings of the 30th International Conference on Neural Information Processing Systems,2016: 4905–4913.

[60] Shelhamer E, Long J, Darrell T. Fully Convolutional Networks for Semantic Segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017,39(4): 640-651.

[61] Wu Z, Zhang X, Li F, et al. Multi-scale long-range interactive and regional attention network for stroke lesion segmentation[J]. Computers and Electrical Engineering, 2022,103: 108345.

[62] Yu D, Wang H, Chen P, et al. Mixed Pooling for Convolutional Neural Networks[C]. Rough Sets and Knowledge Technology, 2014: 364-375.

[63] Zeiler M D, Taylor G W, Fergus R. Adaptive deconvolutional networks for mid and high level feature learning[C]. 2011 International Conference on Computer Vision, 2011: 2018-2025.

[64] Wang Z, Chen J, Hoi S C H. Deep Learning for Image Super-Resolution: A Survey[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021,43(10): 3365-3387.

[65] Xie L, Huang J, Yu J, et al. CNTSeg: A multimodal deep-learning-based network for cranial nerves tract segmentation[J]. Medical Image Analysis, 2023,86: 102766.

[66] Mnih V, Heess N, Graves A, et al. Recurrent models of visual attention[C].Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2,2014: 2204–2212.

[67] Jadon S. A survey of loss functions for semantic segmentation[C]. 2020 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB), 2020: 1-7.

[68] Ho Y, Wookey S. The Real-World-Weight Cross-Entropy Loss Function: Modeling the Costs of Mislabeling[J]. IEEE Access, 2020,8: 4806-4813.

[69] Pihur V, Datta S, Datta S. Weighted rank aggregation of cluster validation measures: a Monte Carlo cross-entropy approach[J]. Bioinformatics, 2007,23(13): 1607-1615.

[70] Sudre C H, Li W, Vercauteren T, et al. Generalised Dice Overlap as a Deep Learning Loss Function for Highly Unbalanced Segmentations[C]. Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, 2017: 240-248.

[71] Taghanaki S A, Zheng Y, Kevin Zhou S, et al. Combo loss: Handling input and output imbalance in multi-organ segmentation[J]. Computerized Medical Imaging and Graphics, 2019,75: 24-33.

[72] Roy A G, Navab N, Wachinger C. Concurrent Spatial and Channel ‘Squeeze & Excitation’ in Fully Convolutional Networks[C]. Medical Image Computing and Computer Assisted Intervention – MICCAI 2018, 2018: 421-429.

[73] Liew S-L, Anglin J M, Banks N W, et al. A large, open source dataset of stroke anatomical brain images and manual lesion segmentations[J]. Scientific Data, 2018,5(1): 180011.

[74] Maier O, Menze B H, Von Der Gablentz J, et al. ISLES 2015 - A public evaluation benchmark for ischemic stroke lesion segmentation from multispectral MRI[J]. Medical Image Analysis, 2017,35: 250-269.

[75] Paszke A, Gross S, Chintala S, et al. Automatic differentiation in PyTorch[J], 2017.

[76] Kingma D P, Ba J. Adam: A Method for Stochastic Optimization[J]. CoRR, 2014,abs/1412.6980.

[77] Xu Z, Lu D, Wang Y, et al. Noisy Labels are Treasure: Mean-Teacher-Assisted Confident Learning for Hepatic Vessel Segmentation[C]. Medical Image Computing and Computer Assisted Intervention – MICCAI 2021, 2021: 3-13.

[78] Guo L, Lei B, Chen W, et al. Dual attention enhancement feature fusion network for segmentation and quantitative analysis of paediatric echocardiography[J]. Medical Image Analysis, 2021,71: 102042.

[79] Badrinarayanan V, Kendall A, Cipolla R. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017,39(12): 2481-2495.

[80] Çiçek Ö, Abdulkadir A, Lienkamp S S, et al. 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation[C]. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016, 2016: 424-432.

[81] Yang Z, Zhu L, Wu Y, et al. Gated Channel Transformation for Visual Recognition[C]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020: 11791-11800.

[82] Wang S, Chen Z, You S, et al. Brain stroke lesion segmentation using consistent perception generative adversarial network[J]. Neural Computing and Applications, 2022,34(11): 8657-8669.

[83] Hui H, Zhang X, Wu Z, et al. Dual-Path Attention Compensation U-Net for Stroke Lesion Segmentation[J]. Intell. Neuroscience, 2021,2021: 16.

[84] Ji Y, Zhang R, Wang H, et al. Multi-compound Transformer for Accurate Biomedical Image Segmentation[C]. Medical Image Computing and Computer Assisted Intervention – MICCAI 2021, 2021: 326-336.

[85] Wu H, Liu J, Xiao F, et al. Semi-supervised segmentation of echocardiography videos via noise-resilient spatiotemporal semantic calibration and fusion[J]. Medical image analysis, 2022,78: 102397.

[86] Xiang T, Zhang C, Wang X, et al. Towards bi-directional skip connections in encoder-decoder architectures and beyond[J]. Medical Image Analysis, 2022,78: 102420.

[87] Boutillon A, Conze P-H, Pons C, et al. Multi-task, Multi-domain Deep Segmentation with Shared Representations and Contrastive Regularization for Sparse Pediatric Datasets[C]. Medical Image Computing and Computer Assisted Intervention – MICCAI 2021, 2021: 239-249.

# 攻读硕士期间取得的学术成果

**学术论文：**

[1] Du X, Ma K, Song Y. AGMR-Net: Attention-guided multiscale recovery framework for stroke segmentation[J]. Computerized Medical Imaging and Graphics, 2022,101: 102120.

[2] Li L, Ma K, Song Y, Du X. TSRL-Net: Target-aware supervision residual learning for stroke segmentation. Cresomput Biol Med. 2023 Jun;159:106840. doi: 10.1016/j.compbiomed.2023.106840. Epub 2023 Mar 28. PMID: 37116236.

**参与的科研项目：**

[1] 基于注意力的深度学习脑卒中分割方法研究，安徽大学计算机科学与技术学院高峰学科建设项目（2021年优秀博士硕士论文培育项目）,Z010111016

# 致谢

不知不觉间，学生生涯即将结束。三年硕士期间对于医学图像分割领域的研究也弥补了我本科没有进入医学专业的遗憾，让我通过另一种形式重新进入医学领域。我非常感谢张燕平导师、杜秀全导师的谆谆教导，以及实验室伙伴、父母、女友的陪伴。

在我的整个科研过程中，张燕平教授是一位杰出的导师，为我的学术研究提供了至关重要的灵感。她在商空间与粒计算领域的专业知识对我来说非常宝贵，她独特的知识和见解，帮助我对粗粒度的思想有了深度的了解。我从她的《粒计算与商空间理论》课程中学到了很多东西。

同样，杜秀全副教授也是一位非凡的导师，他总是可以回答困扰我们的问题并引导我们发现解决问题的思路，他的引导和支持帮助我克服了许多学术上障碍。我感谢杜秀全老师花了无数个小时审查我的工作，并对工作内容的逻辑和论文的语法提供反馈，促使我能够在国际期刊中发表论文。他的指导非常重要，提供了有效的改进思路，磨练了我的科研能力，并加深了我对神经网络领域的理解。非常幸运能与两位和蔼可亲、才华横溢的导师一起度过三年研究生生涯！

感谢我的师兄师姐，在我刚入学的时候带我入门医学图像分割领域，帮我解决代码上的烦恼，给我分享科研的经验，让我在研一时少走很多弯路。感谢组内成员们的集体知识加深了我对医学图像分割领域的理解，我们互相分享科研过程中的想法，彼此讨论方法的优点，缺点和可行性。当然还有实验室的成员，他们带给了实验室欢声笑语，是组与组之间沟通的桥梁。

我非常感谢我的父母！他们一直在我身边，支持我考研，支持我的生活，让本该工作的年纪继续在学校内生活。我很幸运能得到他们坚定不移的支持。我感谢他们的牺牲以及他们在我研究生学习期间以多种方式支持我。

最后，我感谢我的女朋友！在本科以及整个研究生生涯中，她的理解和认可帮助我保持努力向前和科研动力，即使在考研艰难的时候也提供坚定不移的支持和鼓励！

最后感谢各位专家们在论文评审和答辩中的评论！非常荣幸有机会在您们的指导下学习和成长！我期待着将我获得的知识和能力应用到我人生下一阶段！