

中图分类号:

论文编号:

安徽大學

硕士学位论文

# 基于粗粒度注意力的脑卒中病灶 分割方法研究

作者姓名	马昆鹏
一级学科	电子信息
二级学科	计算机技术
指导教师	张燕平教授 杜秀全副教授

# **Research on Segmentation Method of Stroke Lesion Based on Coarse-Grained Attention**

A Dissertation Submitted for the Degree of Master

**Candidate: Ma Kunpeng**

**Supervisor: Prof. ZHANG Yanping、Associate Prof. DU Xiuquan**

图分类号：

论文编号：

## 硕 士 学 位 论 文

# 基于粗粒度注意力的脑卒中病灶分割方法研究

作者姓名	马昆鹏	申请学位级别	工学硕士
指导教师姓名	张燕平、杜秀全	职 称	教授/副教授
学科专业	计算机技术	研究方向	医学图像处理
学习时间自	2020 年 9 月 1 日	起至	2023 年 6 月 1 日止
论文提交日期	2023 年 4 月 30 日	论文答辩日期	2023 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

第一章 绪论.....	1
1.1 研究背景及意义.....	1

1.2 国内外研究现状与研究思路.....	2
1.2.1 国内外脑卒中分割方法.....	2
1.2.2 深度学习脑卒中分割面临的挑战及研究思路.....	3
1.3 主要研究内容.....	6
1.4 论文组织结构.....	7
<b>第二章 相关理论与基础.....</b>	<b>9</b>
2.1 图像分割的网络结构.....	9
2.2 经典网络 U-Net 和 D-UNet.....	9
2.3 图像分割中的注意力机制.....	11
2.4 医学图像分割中的损失函数.....	12
2.5 本章小结.....	14
<b>第三章 粗粒度注意力引导的脑卒中分割网络.....</b>	<b>15</b>
3.1 注意力引导多尺度恢复网络整体框架.....	15
3.2 粗粒度图像块注意力模块.....	15
3.3 跨维度特征融合模块.....	18
3.4 多尺度反卷积上采样模块.....	20
3.5 实验细节.....	21
3.5.1 数据集和实施细节.....	21
3.5.2 评估指标.....	21
3.5.3 损失函数计算.....	22
3.6 实验结果以及分析.....	23
3.6.1 注意力引导多尺度恢复网络的消融实验.....	23
3.6.2 注意力引导多尺度恢复网络与最先进方法的比较.....	23
3.6.3 粗粒度图像块注意力模块的实验分析.....	26
3.6.4 跨维度特征融合模块的实验分析.....	28
3.6.5 多尺度反卷积上采样模块的消融实验.....	30
3.7 本章小结.....	31
<b>第四章 基于粗粒度残差学习的脑卒中分割网络.....</b>	<b>32</b>
4.1 目标感知监督残差学习网络整体框架.....	32



4.2 粗粒度残差学习模块.....	33
4.3 目标感知损失函数.....	34
4.4 实验细节.....	36
4.4.1 数据集.....	36
4.4.2 实验指标.....	37
4.4.3 损失函数的计算.....	37
4.5 实验结果及分析.....	38
4.5.1 消融研究.....	38
4.5.2 目标感知监督残差学习网络的分析实验.....	38
4.5.3 目标感知损失的实验分析.....	42
4.5.4 粗粒度残差学习模块的实验分析.....	45
4.6 本章小结.....	47
总结与展望.....	48
参考文献.....	50
攻读硕士期间取得的学术成果.....	59
致谢.....	60

## 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
图 1.2	细粒度和粗粒度注意力的对比图.....	4
图 1.3	目标感知损失中聚焦样本的扩张图.....	7
图 2.1	U-Net 网络模型.....	10
图 2.2	D-UNet 网络模型.....	11
图 2.3	通道注意力和空间注意力.....	12
图 3.1	注意力引导多尺度恢复网络模型.....	15
图 3.2	粗粒度图像块注意力结构图.....	16
图 3.3	注意力图的上采样过程.....	17
图 3.4	跨维度特征融合模块的结构图.....	19
图 3.5	3*3 反卷积的缺陷表示图.....	20
图 3.6	不同方法的分割结果图.....	25
图 3.7	边界分割结果对比.....	25
图 3.8	AGMR-Net 与其他模型的箱型图 .....	26
图 3.9	粗粒度注意力的可视化效果图.....	28
图 3.10	CPA 模块对特征权重的影响 .....	28
图 4.1	目标感知监督残差学习网络结构.....	32
图 4.2	粗粒度残差学习模块结构图.....	33
图 4.3	损失函数聚焦区域扩张图.....	35
图 4.4	聚焦因子的作用图.....	35
图 4.5	不同方法预测结果的比较.....	40
图 4.6	AGMR-Net 和 TSRL-Net 在召回率上的比较.....	41
图 4.7	不同模型的 DSC 结果的箱形图.....	41
图 4.8	$\beta$ 在目标感知损失函数中的影响 .....	43
图 4.9	不同残差学习的侧输出的对比.....	46

## Figure Index

Figure 1.1	Feature maps of encoder to decoder stages in the U-Net model .....	4
Figure 1.2	Comparison of fine-grained and coarse-grained attention maps .....	4
Figure 1.3	Expansion plot of focused samples in target-aware loss .....	7
Figure 2.1	U-Net network model .....	10
Figure 2.2	D-UNet network model .....	11
Figure 2.3	Channel attention and spatial attention.....	12
Figure 3.1	Attention-guided multiscale recovery network model .....	15
Figure 3.2	Coarse-grained patch attention structure figure .....	16
Figure 3.3	The upsampling process of the attentional map .....	17
Figure 3.4	Structure of the cross-dimensional feature fusion module .....	19
Figure 3.5	Disadvantage representation figure of 3*3 deconvolution .....	20
Figure 3.6	Figure of segmentation results of different methods .....	25
Figure 3.7	Comparison of boundary segmentation results .....	25
Figure 3.8	Box plot of AGMR-Net with other models .....	26
Figure 3.9	Visualization of coarse-grained attention .....	28
Figure 3.10	Impact of CPA module on feature weights .....	28
Figure 4.1	Target-aware supervised residual learning network structure .....	32
Figure 4.2	Structure diagram of coarse-grained residual learning module.....	33
Figure 4.3	Loss function focus area expansion map.....	35
Figure 4.4	Map of the role of focus factors .....	35
Figure 4.5	Comparison of prediction results of different methods .....	40
Figure 4.6	Comparison of AGMR-Net and TSRL-Net in terms of Recall .....	41
Figure 4.7	Box plots of DSC results for different models .....	41
Figure 4.8	The effect of $\beta$ in the target-aware loss function .....	43
Figure 4.9	Comparison of the side outputs of different residual learning .....	46

## 表索引

表 2.1	医学图像分割中常用损失函数.....	13
表 3.1	每个模块对骨干网络 D-UNet 的影响.....	23
表 3.2	我们的方法和先前的方法比较.....	24
表 3.3	边界分割结果对比.....	26
表 3.4	空间注意方法的比较.....	27
表 3.5	CFF 与普通特征融合方法的比较.....	29
表 3.6	不同衰减率对基线的提升.....	29
表 3.7	MDU 与单尺度反卷积的比较.....	30
表 3.8	MDU 模块在不同位置的整合.....	30
表 4.1	消融实验的定量结果.....	38
表 4.2	在 ATLAS 和 ISLES 数据集上的对比.....	39
表 4.3	复杂性分析和显著性分析实验.....	42
表 4.4	$\gamma$ 和 $\beta$ 对损失函数的影响.....	43
表 4.5	目标感知损失在其他网络上的适用性.....	44
表 4.6	不同损失函数的对比.....	45
表 4.7	不同的残差学习模块的比较.....	45

## Table Index

Table 2.1	Common loss functions in medical image segmentation.....	13
Table 3.1	Impact of each module on the backbone network D-UNet.....	23
Table 3.2	Comparison of our method with previous methods .....	24
Table 3.3	Comparison of boundary segmentation results .....	26
Table 3.4	Comparison of spatial attention methods.....	27
Table 3.5	Comparison of CFF and common feature fusion methods .....	29
Table 3.6	Enhancement of baseline by different decay rates .....	29
Table 3.7	Comparison of MDU and single-size deconvolution.....	30
Table 3.8	Integration of MDU modules in different locations.....	30
Table 4.1	Quantitative results of ablation experiments.....	38
Table 4.2	Comparison on ATLAS and ISLES datasets.....	39
Table 4.3	Complexity analysis and significance analysis experiments .....	42
Table 4.4	The effect of $\gamma$ and $\beta$ on the loss function .....	43
Table 4.5	Applicability of target-aware loss on other networks .....	44
Table 4.6	Comparison of different loss functions .....	45
Table 4.7	Comparison of different residual learning modules.....	45



# 第一章 绪论

## 1.1 研究背景及意义

脑卒中，又称脑中风，是由于脑血管破裂（脑溢血）或局部血管堵塞<sup>[1]</sup>（脑梗塞）引发的脑部缺血<sup>[2]</sup>。脑卒中疾病具有普遍性和高危性，脑卒中患者不仅会出现沟通困难、肌肉无力和记忆力衰退等不可逆转的残疾<sup>[3]</sup>，而且治疗不及时将会导致脑死亡，是世界卫生组织(WTO)报道的全球第二大疾病死亡原因<sup>[4]</sup>。在病情发作后的 24 小时内，临床医生面临重要的、时间敏感的决定，例如是否进行干预以挽救已经受损组织。脑卒中的亚急性期和慢性阶段，医生需要了解不同干预的反应和大脑的相应变化，这更有利于脑卒中患者的长期康复。在脑卒中急性期，谨慎的临床决策至关重要，因为有效的干预治疗可以保留正常的脑神经组织和促进早期脑部功能恢复。常规获取的结构核磁共振图像(MRI)可以提供大脑结构图像<sup>[5]</sup>，有助于医生做出正确的诊断和紧急的临床决策<sup>[3]</sup>。

因此脑卒中医学图像具有重要的研究价值，它能够帮助医生了解脑卒中的病情，但是其先决条件是医生需要准确地识别出核磁共振扫描图像中脑卒中病灶的区域<sup>[6]</sup>，以便计算病变区域的大小、形状和体积，这些信息对于医生诊断和治疗脑卒中至关重要<sup>[7]</sup>。由医生手动标注的标签被视为金标准，但是这种精准绘制是时间和劳动密集型的过程，需要专业的医学知识，并且依赖于医生的主观判断，因此不适用于大量的数据标注和实时快速的临床需求。高分辨率 T1 加权 (T1W)MRI 更适用于脑卒中的识别和绘制病变组织，现有的用于 T1W MRI 数据的自动脑卒中分割工具中，大多数都不是准确和可靠的，并且需要大量的手动工作来进行分割质量控制和矫正。所以需要一种准确的自动方法来帮助医生分割病变区域<sup>[8]</sup>，估计病灶的形状和大小，这将在脑卒中治疗的前期和后期提供重要的定量和定性线索<sup>[9]</sup>，帮助医生实时定制适合病人的诊疗方案。

机器学习，特别是深度学习算法，已经被用来解决这个问题，它们依靠大型、多样化的训练数据集来创建可推广的模型，这些模型可以在新数据上表现优异。由于脑卒中是全世界承认致残的主要原因，当今基于深度学习辅助脑卒中分割的方法备受关注，具有重要的研究价值。

## 1.2 国内外研究现状与研究思路

### 1.2.1 国内外脑卒中分割方法

利用计算机辅助完成自动化脑卒中病灶分割的方法从早期的传统图像处理方法逐步发展到现在先进的深度学习分割方法。在传统方法中, Prakash 等人<sup>[10]</sup>提出使用阈值分割方法对缺血脑卒中中进行分割。Mitra 等人<sup>[11]</sup>提出利用随机森林方法对多模态 MRI 图像进行脑卒中概率分类。为了改善边界的分割效果, Mujumdar 等人<sup>[12]</sup>采用 Chan-Vese 模型<sup>[13]</sup>提取目标边缘轮廓特征, 对边界轮廓内的脑卒中病灶进行自动分割。

传统的分割方法虽然计算方式简单, 但是分割速率慢且精度低, 不能满足临床上实际的需求, 并且大部分传统算法需要后期的人为干预。深度学习方法旨在模仿人类大脑神经元的突触行为, 构建起能够自动化提取目标特征的神经网络模型, 相较于传统算法, 更加客观和准确。近年来, 深度学习快速的发展, 卷积神经网络(Convolutional Neural Network, CNN)在计算机视觉领域倍受关注, 不仅应用于自然图像, 研究人员也将 CNN 模型扩展到了医学图像分割领域中<sup>[14]</sup>。在医学图像分割领域最广为人知的模型为 U-Net<sup>[15]</sup>模型, 大量研究人员使用 2D 卷积构成的 2D U-Net 的衍生模型应用于脑卒中分割。CLCI-Net<sup>[16]</sup>不仅在编码阶段使用多层的密集跳连接和空洞卷积金字塔模块<sup>[17]</sup>捕获多尺度上下文信息, 而且在解码阶段使用 LSTM<sup>[18]</sup>来推断编码器和解码器特征间的上下文信息, 从而捕获目标的精细的边界信息以解决类间模糊的问题。近年来, 注意力机制在医学图像分割领域被广泛的应用<sup>[19]</sup>, X-Net<sup>[20]</sup>使用了非局部注意力提取了卷积特征间的长距离依赖关系, 同时使用深度可分离卷积<sup>[21]</sup>降低了模型的计算量, 因此训练参数明显少于现有的脑卒中分割模型<sup>[22]</sup>。Jia 等人<sup>[23]</sup>在引入循环神经网络<sup>[24]</sup>的基础上使用注意力卷积门递归单元。Liu 等人<sup>[25]</sup>开发的深度残差注意力网络继承 U-Net 的优点, 并引入了跳连接注意力机制, 提高了分割性能。

上述网络都是对 MRI 图像的单独横截面进行 2D 卷积计算, 但是 MRI 是 3D 数据集。2D 卷积虽然比 3D 卷积复杂度低, 但是不能提取 3D 图像切片间的连续上下文信息<sup>[26]</sup>。在医学图像的 3D 数据分割中, 通过 3D 卷积学习的准确边界信息可以增加到 2D 卷积特征图中, 获得更好的边界分割性能。为此, ComboNet<sup>[27]</sup>



直接合并 2D CNN 和 3D CNN 的分割结果，也可以提升目标边界的分割精度，但 3D CNN 的计算量太大<sup>[28]</sup>。为了解决这个问题，一些网络，如 M2D3DNet<sup>[29]</sup> 和 GGPFN<sup>[30]</sup>，只在编码阶段融合 2D 和 3D 卷积特征图，以促进不同维度特征的互补，同时减少了模型计算量。不幸的是，它们的融合方法过于简单，没有充分考虑到不同任务特征之间的依赖性<sup>[31]</sup>。SENet<sup>[32]</sup>设计了挤压和激励模块，根据相关输入的不同部分来分配通道权重<sup>[33]</sup>。学习每个特征对应通道的依赖性可以增强与任务相关的特征<sup>[34,35]</sup>。D-UNet<sup>[26]</sup>使用挤压和激励模块来改进 2D 和 3D 特征融合，大大提高了网络性能。

### 1.2.2 深度学习脑卒中分割面临的挑战及研究思路

尽管上述的深度学习方法已经在脑卒中病灶分割领域取得了不错的成果，但受限于特征提取能力的不足，不能有效的提取特定于目标的特征，使得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，即病变区域之间存在较大的差异性。网络如果无法准确定位目标，将不能有效的提取目标特征；(2)类间模糊，即脑卒中病灶在外观和灰度值上与正常脑组织相似，导致病灶与脑组织之间的边界不清晰，网络将无法区分正常脑组织和病灶；(3)正负样本不平衡，脑卒中病灶的正样本数量远远小于正常脑组织的负样本数量，会导致网络倾向于提升负样本分类精度而忽略正样本分类精度。所以深度学习方法的应用程度虽然已经远远领先于传统方法，但仍值得进一步的改善。

突出的目标空间（位置）信息可以引导网络挖掘目标相关的特定特征，并克服类内不一致的挑战。如图 1.1 所示，在网络编码阶段，目标的空间信息不如在解码阶段突出，所以一些研究采用细粒度的空间注意力来突出编码阶段的目标空间信息<sup>[36-39]</sup>。然而，如图 1.2 中细粒度注意力图所示，编码阶段产生的细粒度注意力图不加区分地增强了所有脑部区域，这种不准确的注意力方式不能突出目标的准确空间信息。造成编码阶段的细粒度注意力图不准确的原因是：(1)网络在编码阶段学习的目标特征是细节特征，如边或角，没有足够的目标信息来产生准确的细粒度注意力图；(2)在编码阶段，随着级联的池化操作，特征图越来越抽象，如图 1.1 所示，目标特征变的不可预测，无法使用包含空间信息的标签（先验信息），注意力的准确性也就不能通过显式的训练和监督来保证<sup>[40]</sup>，细粒度注意力图精度的缺点也限制了它在网络编码阶段中的作用<sup>[41]</sup>。为了能够让注意力图以显式

直接的方式进行监督，PDNet<sup>[42]</sup>引入一种粗粒度的方法来突出病变区域，即以点击信息为中心画一个覆盖目标的圆，并通过监督其直径来控制注意力图的准确性。这种粗粒度的方式虽然不能完全准确的增强目标样本像素点，但是在编码阶段与细粒度注意力相比能够抑制大部分的背景噪音以突出目标信息。然而，点击信息作为网络模型的输入，必须由医生手动标注。因此，有必要设计一个粗粒度的注意力模块，可以在原数据集的基础上和没有医生辅助标注的情况下进行注意力图的直接监督和训练，以引导网络关注目标区域。

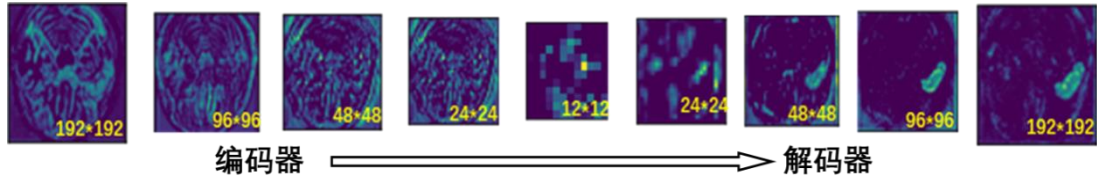


图 1.1 U-Net 模型中编码器到解码器各阶段的特征图

Figure 1.1 Feature maps of encoder to decoder stages in the U-Net model

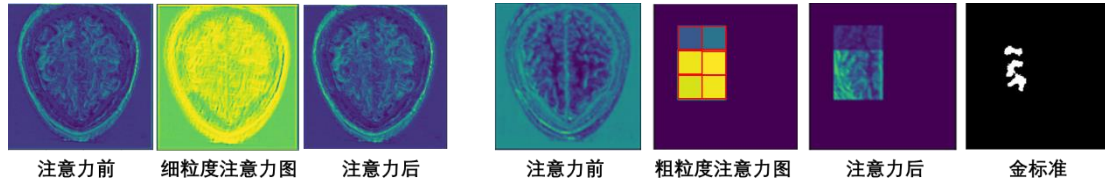


图 1.2 细粒度和粗粒度注意力的对比图

Figure 1.2 Comparison of fine-grained and coarse-grained attention maps

通过编码器和解码器之间的跳连接操作，解码器可以再次学习该阶段还未识别的目标特征，即残差特征，以缓解类内不一致的挑战<sup>[43]</sup>。UNet++<sup>[44]</sup>也受益于不同编码层上采样特征的嵌套互连。RoR<sup>[45]</sup>优化跳连接的映射方式，加快了不同卷积层之间的信息传播。然而，这种简单的跳连接操作会产生冗余的信息，例如，背景噪音和解码器已识别的目标信息。这些信息会稀释目标的残差特征，不利于网络的提取。有研究<sup>[46,47]</sup>在肠息肉的分割中引入了残差学习，在跳连接的基础上，通过反向注意力抑制已经被解码器识别的目标特征。这避免了跳连接中已识别目标信息的重复，有利于引导网络提取目标的残差特征以提高召回率(Recall)。反向注意力虽然避免了已学习目标信息的冗余，但也引入了除目标之外的所有背景噪声，不利于提升结果的精确率(Precision)。Pranet<sup>[46]</sup>通过反向注意力单元增强边界信息的表示，校准了未对齐的目标特征。上述方法通过跳连接在解码阶段有效地学习残差特征，解决了类内不一致的难题，提高了召回率。然而，在解码阶段再次引入了大量的背景噪声，它导致了精确率的下降。因此，我们考虑在反向注意

力的基础上引入能够抑制背景噪音的正向注意力,这样不仅可以提高召回率,还可以提高精确率。

准确的边界特征可以放大类间特征差异,有助于解决类间模糊的挑战<sup>[48]</sup>。以前的工作<sup>[29,30]</sup>使用 3D 卷积弥补 2D 卷积在捕获连续边界信息上的不足,这样可以更准确地进行边界分割。不幸的是,这些方法通过直接求和或串联来融合特征的简单行为可能会产生冗余信息和噪声,从而破坏原始特征<sup>[49]</sup>。为此, D-UNet<sup>[26]</sup>为 2D 和 3D 特征融合引入了两个并行的挤压和激励<sup>[32]</sup>机制,对不同维度的特征进行过滤和筛选以达到更好的融合效果。融合任务是将两个维度的全局特征汇集在一起,然而 D-UNet 只对单个维度的特征通道间局部关系进行建模,却忽略了建模全局特征通道之间的关系,这对不同维度之间的全局特征融合仍是一个障碍。我们认为必须考虑不同维度特征通道间的全局依赖关系,这将有利于不同维度之间特征融合。

在用于医学图像分割的编码器-解码器结构中,边界和空间细节信息会在编码器的下采样中丢失,在解码阶段恢复这些信息可以改善分割效果。U-Net 结构的跳连接将编码阶段的特征信息重新引入到解码阶段,进一步有很多工作关注在编码器和解码器信息的高效融合上<sup>[25]</sup>。例如, Attention Unet<sup>[50]</sup>采用了门控机制, CLCI-Net 使用 Conv LSTM<sup>[51]</sup>将一些细节信息从编码器带到解码器,这些方法都弥补了连续池化和逐步卷积操作造成的信息损失。然而,在使用基本线性插值<sup>[15,26,29]</sup>和反卷积<sup>[52]</sup>来恢复特征图分辨率时,解码器的上采样忽略了空间和边界信息的恢复。通过基于卷积神经网络的技术可以捕捉多尺度特征,用于恢复关于目标特征的空间和边界信息<sup>[52]</sup>。DeepLab V3 中提出空洞卷积金字塔池化模块(ASPP)来学习多尺度的上下文信息<sup>[17]</sup>。PSPNet<sup>[53]</sup>使用一个金字塔池模块(PPM)来构造网络最终层上的全局上下文先验信息。COPLE-Net<sup>[54]</sup>使用 ASPP 模块提取丰富的多尺度特征, CLCI-Net 基于 ASPP 将网络不同阶段的特征图聚合在一起,这些网络都在编码阶段使用了多尺度信息并证明了其有效性。所以在上采样过程中采用参数化的多尺度反卷积来获取空间和边界信息,对于改善密集预测的医学图像分割任务更为有利。

在目标正样本与背景负样本的比例不平衡挑战下,二进制交叉熵损失函数仍然平等地计算正负样本的损失,导致正样本损失和负样本损失比例不平衡<sup>[55]</sup>。过

量的负样本损失将主导反向传播梯度，在预测结果中产生较少的假阳性(FP)，但因此会对目标正样本的监督不足，将产生过多的假阴性(FN)。损失函数被用来控制网络学习有意义的信息，也可以用来解决类不平衡的问题<sup>[56]</sup>。虽然  $\text{dice loss}$ <sup>[43]</sup> 已经被证明可以应对不平衡样本的问题，但其结果是精确率高，召回率低，因为它对 FP 和 FN 的惩罚力度仍然是相同的<sup>[57]</sup>。为此， $\text{focal loss}$ <sup>[58]</sup>通过调整正/负样本损失的动态权重来重建交叉熵损失函数，分类错误的样本（困难样本）损失将得到较大的权重，以缓解类不平衡。不幸的是， $\text{focal loss}$  的理想的静态负样本的权重仍然大于正样本的权重，导致网络由负样本主导。为了提高结果的召回率， $\text{toversky loss}$ <sup>[57]</sup>提升了 FN 惩罚力度的权重，减少了 FP 惩罚力度的权重，但精确率明显降低<sup>[55]</sup>。虽然上述损失函数可以成功地提高精确率或召回率，然而，在设计正负样本损失权重时由于对某一方分配的权重过大，会不可避免地造成了召回率和精确率不平衡的问题。因此，在增加正样本权重以提高召回率的同时，不妨考虑增加部分负样本损失权重来避免负样本损失占比失衡，从而有利于网络挖掘负样本信息以平衡精确率。

### 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力图，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力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 单元的注意力与预测概率成反比，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。

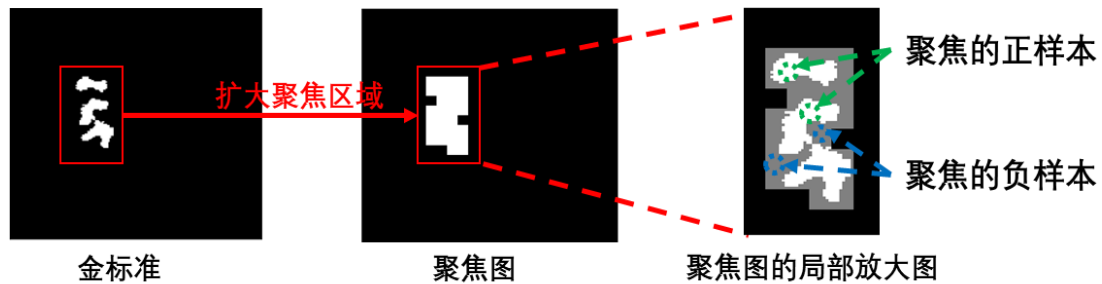


图 1.3 目标感知损失中聚焦样本的扩张图

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

## 1.4 论文组织结构

本文共有五章，以下是各章节主要安排：

第 1 章是本文的绪论，首先，论述了自动化脑卒中分割在临床中的重要意

义。然后，从传统医学图像分割算法和深度学习图像分割算法两方面介绍了国内外脑卒中分割的研究现状。其次，我们讨论了深度学习算法在脑卒中分割中面临的难点和挑战，列出了本文针对这些挑战的研究思路；最后，介绍本文各章内容的安排。

第 2 章首先介绍医学图像分割领域常用的编码器-解码器分割框架以及脑卒中分割中经典的网络：U-Net 和我们的骨干网络 D-UNet。然后，对医学图像分割任务中常见的注意力机制的基本思想和原理进行讨论。最后，介绍了医学图像中常见的损失函数以及改进思路。

第 3 章首先介绍一种基于 D-UNet 改进的注意力引导多尺度恢复网络。其次，详细的介绍粗粒度图像块注意力模块、跨维度特征融合模块以及多尺度反卷积上采样模块。最后阐述了我们实验的细节、指标计算公式、实验所使用的两个脑卒中分割数据集、对比实验和各个模块的分析实验。

第 4 章首先介绍在注意力引导多尺度恢复网络的基础上提出的目标感知监督残差学习网络。其次，介绍所提出的粗粒度残差学习模块和目标感知损失函数。最后介绍了数据集、实验细节、模型评估指标、对比实验和模块的分析实验。

最后是对本文工作的总结和对未来的展望。本章主要包含两个方面，一是对本文提出分割方法的总结，二是对未来脑卒中分割研究工作的展望。

## 第二章 相关理论与基础

### 2.1 图像分割的网络结构

卷积神经网络(CNN)利用卷积的局部感受野<sup>[59]</sup>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)<sup>[60]</sup>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<sup>[61]</sup>,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<sup>[62]</sup>。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<sup>[59]</sup>。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<sup>[63]</sup>和线性插值法<sup>[64]</sup>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 经典网络 U-Net 和 D-UNet

简单的编码器-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<sup>[15]</sup>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<sup>[15]</sup>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  $3 \times 3$  的卷积、ReLU 激活函数和  $2 \times 2$  最大池化组成<sup>[65]</sup>。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。

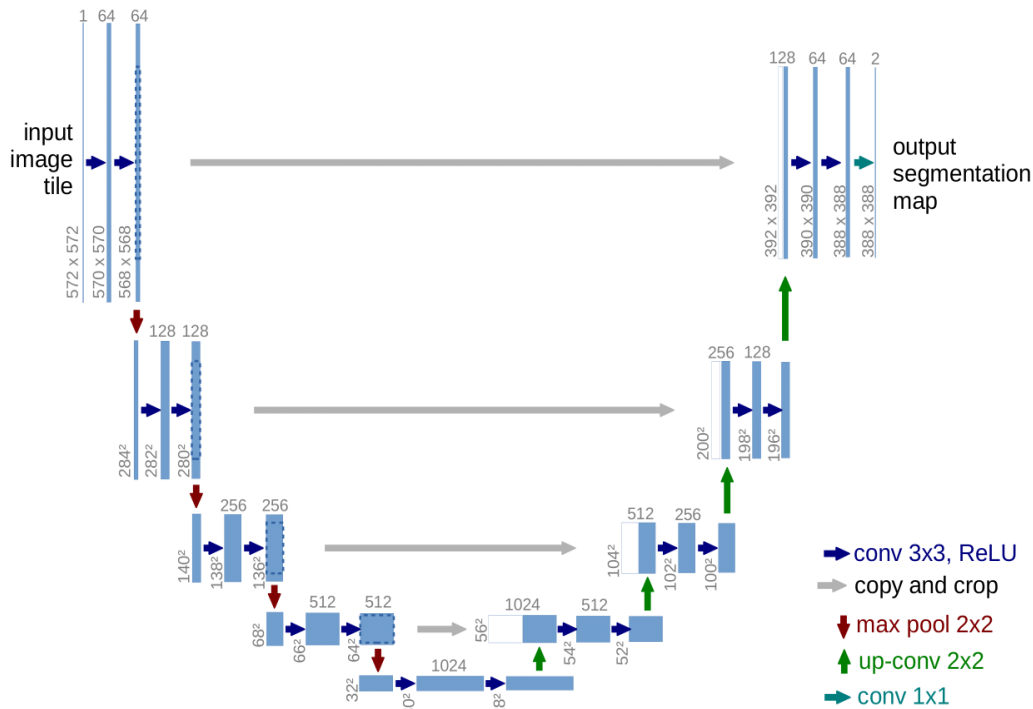


图 2.1 U-Net 网络模型

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<sup>[26]</sup>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 <sup>[26]</sup>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。

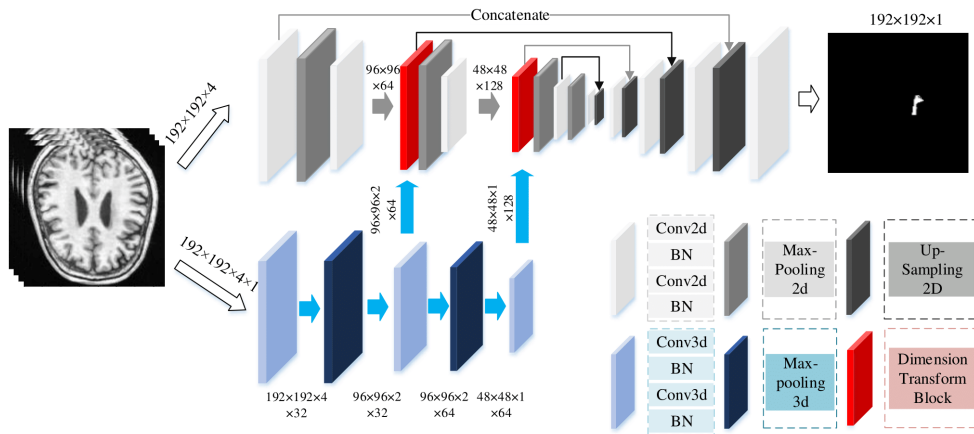


图 2.2 D-UNet 网络模型

Figure 2.2 D-UNet network model

## 2.3 图像分割中的注意力机制

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<sup>[36,37,39]</sup>和通道维度<sup>[26,32,33]</sup>注意力。如图 2.3<sup>[33]</sup>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代表。

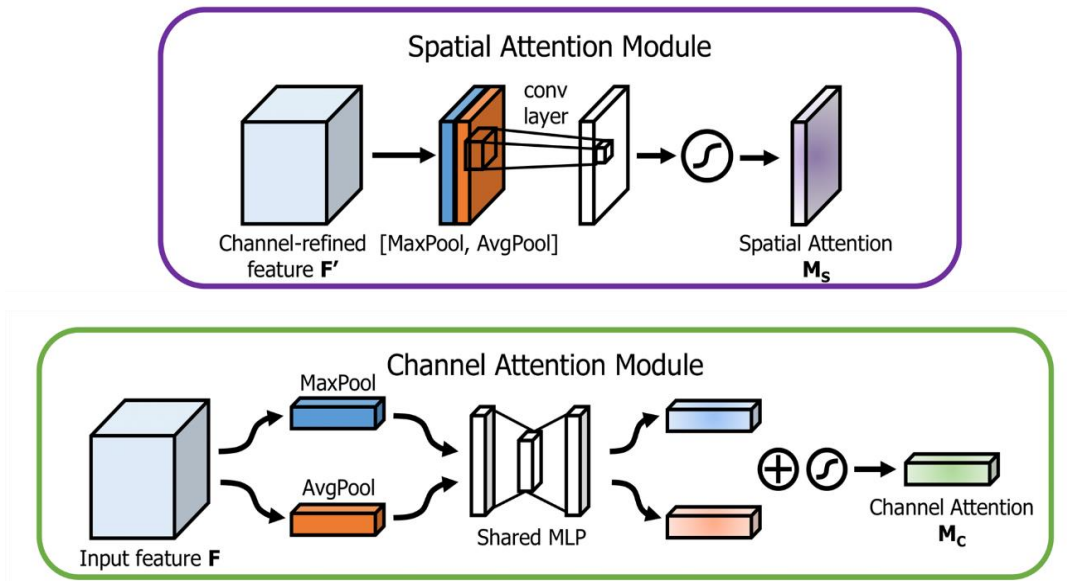


图 2.3 通道注意力和空间注意力

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，显著的提高了解码器-编码器模型在医学图像中的准确性和鲁棒性<sup>[66]</sup>。

## 2.4 医学图像分割中的损失函数

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<sup>[67]</sup>。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<sup>[68]</sup>，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。 $\alpha$  表示正负样本损失的静态权重。 $\gamma$  表示 focal loss 中正负样本损失的动态权重，最后用  $\epsilon$  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<sup>[55]</sup>。更进一步，加权交叉熵(Weighted cross-entropy loss)<sup>[69]</sup>提出了正样本损失的权重  $\alpha$ ，设置  $\alpha > 1$  以减少假阴性， $\alpha < 1$  以减少假阳性。平衡交叉熵(Balanced cross-entropy loss)<sup>[67]</sup>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<sup>[58]</sup>在平衡交叉熵损失函数的基础上进一步提出了动态权重  $\gamma$ ，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

方法	公式
Binary Cross-Entropy loss	$L_{BCE}(y, p) = -(y \log(p) + (1 - y) \log(1 - p))$
Dice Loss	$L_{dice}(y, p) = 1 - \frac{yp + \epsilon}{y + p + \epsilon}$
Weighted cross-entropy loss	$L_{WBCE}(y, p) = -(\alpha y \log(p) + (1 - y) \log(1 - p))$
Balanced Cross-Entropy loss	$L_{BBCE}(y, p) = -(\alpha y \log(p) + (1 - \alpha)(1 - y) \log(1 - p)), \alpha < 1$
Focal loss	$L_{focal}(y, p) = -(\alpha(1 - p)^\gamma y \log(p) + (1 - \alpha)(p)^\gamma (1 - y) \log(1 - p)), 0 < \alpha < 1, \gamma > 1$
Enhanced Mixing Loss	$L_{eml}(y, p) = \frac{L_{focal}(y, p)}{N} - \log(L_{dice}(y, p))$
Tversky Loss	$L_{Tversky}(y, p) = 1 - \frac{yp + \epsilon}{yp + (1 - \alpha)(1 - y)p + \alpha y(1 - p) + \epsilon} \quad 0 < \alpha < 1$

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<sup>[70]</sup>。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<sup>[57]</sup>。Tversky loss<sup>[57]</sup>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<sup>[67]</sup>，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<sup>[71]</sup>。Focal loss 和使用对数运算的 dice loss 被整合到 enhanced mixing loss(EML loss)<sup>[26]</sup>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<sup>[68]</sup>。

## 2.5 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 粗粒度注意力引导的脑卒中分割网络

### 3.1 注意力引导多尺度恢复网络整体框架

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 卷积的交汇点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。

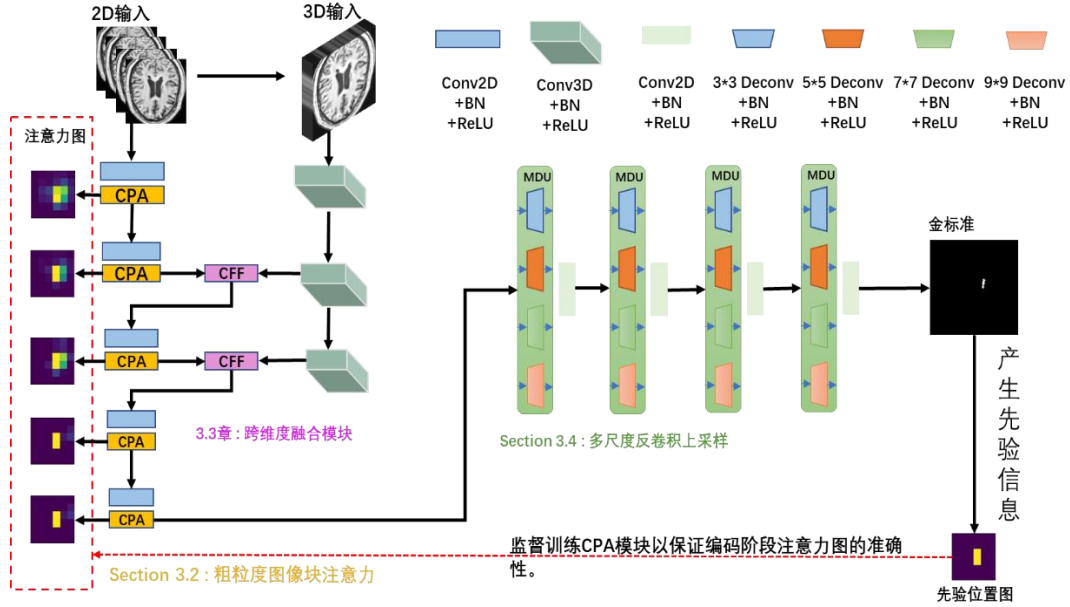


图 3.1 注意力引导多尺度恢复网络模型

Figure 3.1 Attention-guided multiscale recovery network model

### 3.2 粗粒度图像块注意力模块

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<sup>[33]</sup>，由于在是将输入  $1 \times H \times W$  的图像等分成 36 个图像块，所以挤压操作如下所示：

$$F_i'(1, x_i, y_j) = \frac{\sum_{\gamma=1}^C F_i(\gamma, x_i, y_j)}{C} \quad i \in [1, H] \quad (3.1)$$

$$F_g(1, x_i, y_j) = \frac{\sum_{\gamma=1}^h \sum_{\delta=1}^w F_i'(1, x_{(i-1)*h+\gamma}, y_{(j-1)*w+\delta})}{h^2} \quad i \in [1, H] \quad h, w = \frac{H}{6} \quad (3.2)$$

如公式(3.1)所示，其中  $F_i \in \mathbb{R}^{C \times H \times W}$  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  $F_i' \in \mathbb{R}^{1 \times H \times W}$ ，这在对于突出目标区域方面是有效的<sup>[33]</sup>。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  $\frac{H}{6} \times \frac{H}{6}$  的平均池化生成  $F_g \in \mathbb{R}^{1 \times 6 \times 6}$ 。每个池化核覆盖一个图像块并生成图像块描述符。

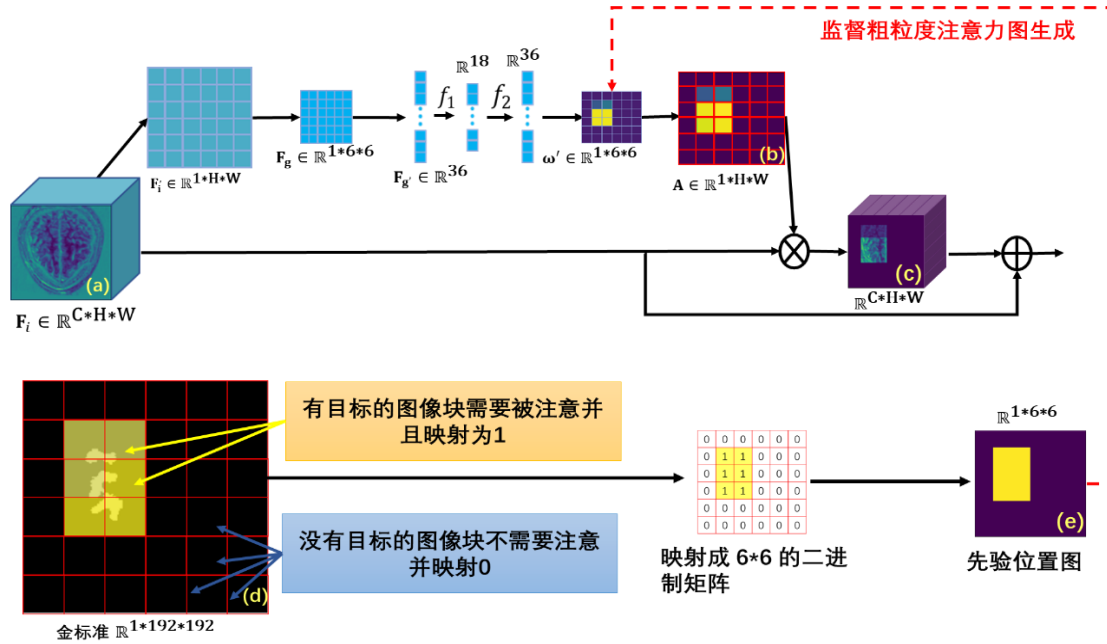


图 3.2 粗粒度图像块注意力结构图

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<sup>[72]</sup>所有图像块之间的依赖关系，如公式(3.3)，其中 $f_1$ 和 $f_2$ 表示两层全连接神经网络， $\varphi$ 表示 sigmoid 激活函数。因此， $F_g$ 被展开为 $F'_g \in \mathbb{R}^{36}$ 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 $\omega$ 。 $\omega$ 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。

$$\omega = \varphi(f_2(f_1(F'_g))) \quad (3.3)$$

将得到的 $\omega \in \mathbb{R}^{36}$ 被重塑为概率图 $\omega' \in \mathbb{R}^{1*6*6}$ ，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  $A \in \mathbb{R}^{1*H*W}$ ，其中上采样公式如下所示：

$$A_{(x-1)*\frac{H}{6}+m,(y-1)*\frac{W}{6}+n} = \omega'_{x,y} \quad x,y \in [1,6] m \in [1,\frac{H}{6}] n \in [1,\frac{W}{6}] \quad (3.4)$$

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  $\text{Output} = A * F_i + F_i$ 。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。

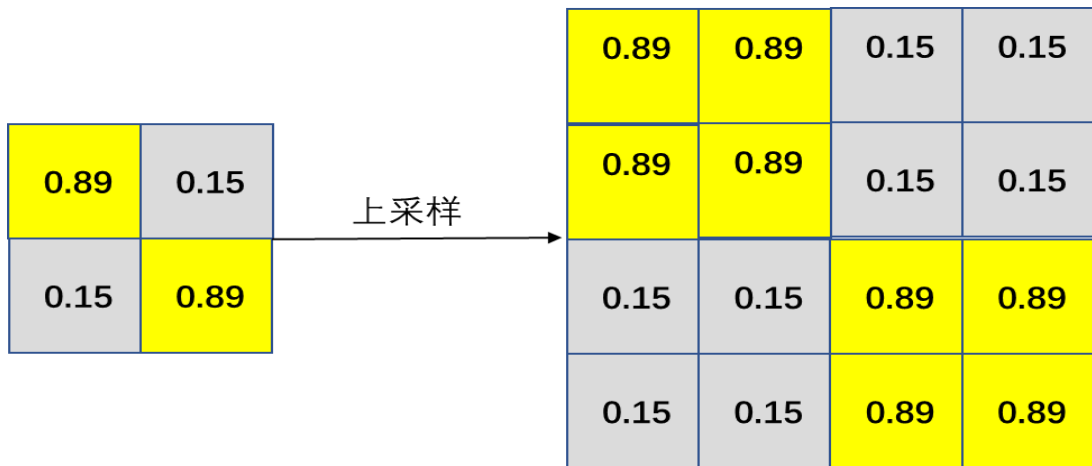


图 3.3 注意力图的上采样过程

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 跨维度特征融合模块

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  $2D_{fm} \in \mathbb{R}^{C*H*W}$  和 3D 特征图  $\mathbb{R}^{C*H*W*D}$  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  $\mathbb{R}^{C*H*W*D}$  转换为单通道  $3D_c \in \mathbb{R}^{1*H*W*D}$ 。然后， $3D_c$  在空间维度上被降维成  $3D_s \in \mathbb{R}^{H*W*D}$ ，并把  $3D_s$  维度置换成为具有  $D$  个通道的  $3D_r \in \mathbb{R}^{D*H*W}$ 。为了保持通道数与 2D 特征图中通道数一致， $3D_r$  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  $3D_{tfm} \in \mathbb{R}^{C*H*W}$ 。2D 特征图  $2D_{fm} \in \mathbb{R}^{C*H*W}$  和  $3D_{tfm}$  被送到后续计算中。

从图 3.4 可以看出，使用全局平均池化  $Avg\_global$  将  $3D_{tfm}$  和  $2D_{fm}$  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  $g \in \mathbb{R}^{2C}$ ，并将  $g$  送入多层全连接神经网络，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  $\omega_2, \omega_3 \in \mathbb{R}^C$  是在经过  $\text{sigmoid}$  激活函数映射后获得的，如公式(3.6)和公式(3.7)所示。 $f_1^2, f_2^2, f_1^3, f_2^3$  表示两层全连接神经网络， $\varphi$  表示  $\text{sigmoid}$  激活函数， $\odot$  表示串联。两种维度的所有通道描述符都包含在  $g$  中，因此多层全连接神经网络学习两种维度之间所有通道之间的关系。

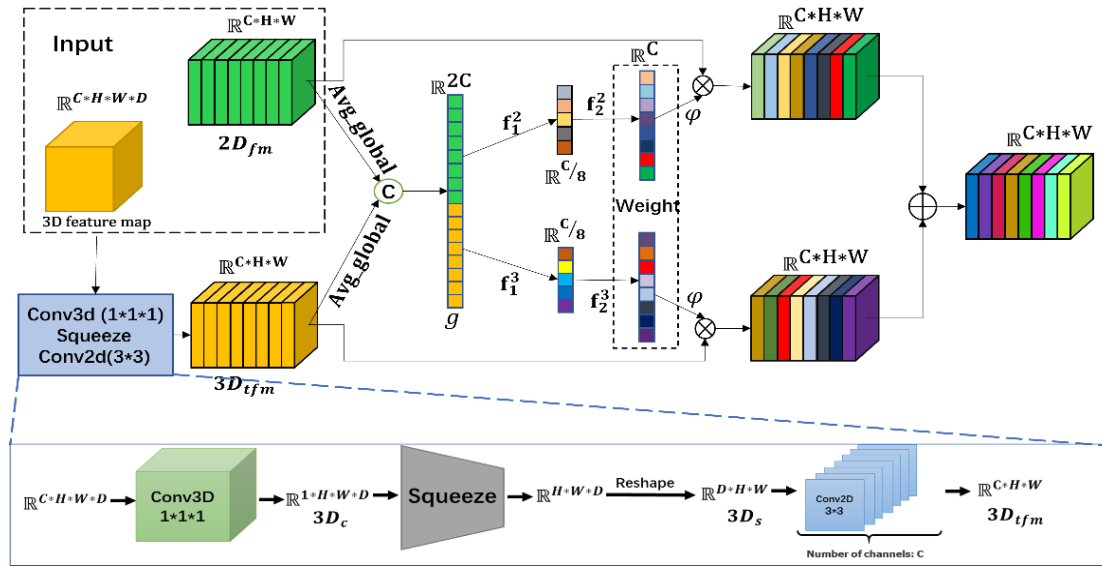


图 3.4 跨维度特征融合模块的结构图

Figure 3.4 Structure of the cross-dimensional feature fusion module

$$g = \text{Avg\_global}(3D_{tfm}) \odot \text{Avg\_global}(2D_{fm}) \quad (3.5)$$

$$\omega_2 = \varphi(f_2^2(f_1^2(g))) \quad (3.6)$$

$$\omega_3 = \varphi(f_2^3(f_1^3(g))) \quad (3.7)$$

$$\text{Output} = 3D_{tfm} * \omega_3 + 2D_{fm} * \omega_2 \quad (3.8)$$

获得的权重  $\omega$  用于重新映射  $3D_{tfm}$  和  $2D_{fm}$  并获得具有融合 2D 和 3D 信息的特征图  $\text{Output} \in \mathbb{R}^{C \times H \times W}$ 。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 多尺度反卷积上采样模块

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,  $9 \times 9$  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  $3 \times 3$  卷积核覆盖的区域, 粉色区域是  $5 \times 5$  卷积核覆盖的区域。图 3.5 (a-c) 表示  $3 \times 3$  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  $3 \times 3$  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  $5 \times 5$  反卷积覆盖。 $5 \times 5$  反卷积也有类似  $3 \times 3$  反卷积的缺点, 但  $7 \times 7$  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  $3 \times 3$ 、 $5 \times 5$ 、 $7 \times 7$  和  $9 \times 9$  的反卷积。

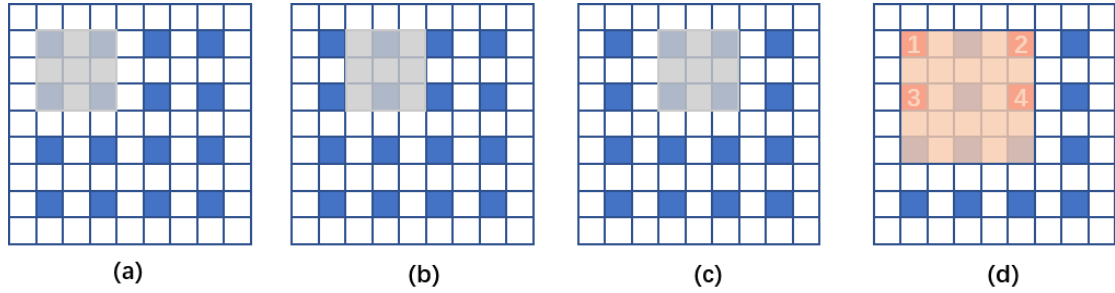


图 3.5  $3 \times 3$  反卷积的缺陷表示图

Figure 3.5 Disadvantage representation figure of  $3 \times 3$  deconvolution

如图 3.1 所示, 对于模块的输入  $\text{input} \in \mathbb{R}^{C \times H \times W}$ , 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  $3 \times 3$ 、 $5 \times 5$ 、 $7 \times 7$  和  $9 \times 9$  的卷积核, 放大后的特征图大小为  $\frac{C'}{4} \times 2H \times 2W$ 。我们将它们串联为  $\text{Output} \in \mathbb{R}^{C' \times 2H \times 2W}$ 。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 实验细节

#### 3.5.1 数据集和实验细节

我们使用两个脑卒中分割数据集来训练和测试我们的网络。

**ATLAS:** 中风后病变的解剖追踪(ATLAS)数据集包含 229 个 MRI T1 模态的病人案例, 每个图像大小为 233\*197, 每个病例有 198 个 2D 图像切片<sup>[73]</sup>。

**ISLES:** 我们使用 ISLES 2015 挑战赛<sup>[74]</sup>中的 SISS 缺血性中风病变数据集, 有 28 个开放的训练案例, 每个案例有 154 个 2D 图像切片。我们使用来自 T1 连续扫描的图像作为网络训练和测试的输入。

**数据预处理:** 我们将图像的每个切片裁剪成 192\*192 大小, 这样就可以去除一些不相关的区域。然后, 我们将每个切片的前两片和后一片结合起来, 形成一个 192\*192\*4 的四通道图像, 输入我们的网络。我们使用 8:2 的比例来划分 ATLAS 数据集, 有 183 个案例用于训练, 46 个案例用于测试。由于 ISLES 只含有 28 案例, 所以我们使用五折交叉验证法来验证我们的模型, 我们将 28 个案例等分为 5 份, 每次取互斥的 1 份用于测试, 其余 4 份用于训练。

**实施细节:** 我们的方法实现是基于 Pytorch<sup>[75]</sup>深度学习框架, 并在 NVIDIA Tesla V100S GPU 上进行训练。每个实验使用的批处理量为 16。训练网络的初始学习率为 0.001。由于我们的方法在 ATLAS 数据集中没有划分验证集, 所以我们使用衰减率为 0.96 的指数衰减法逐渐降低学习率, 以确保网络的最终训练性能是稳定的。并使用 Adam 优化器<sup>[76]</sup>来最小化损失函数。所有的模型都进行了 150 个 epochs 的训练, 以优化方法的性能。

#### 3.5.2 评估指标

在本文中, 我们用五个分割指标来衡量我们方法的性能: (1)Dice 相似性系数(DSC), (2)DSC(G), (3)召回率(Recall), (4)精确率(Precision), 和(5)Jaccard 相似系数。它们被定义为:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (3.9)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (3.10)$$

$$DSC = \frac{2TP}{2TP + FP + FN} \quad (3.11)$$

$$Jaccard = \frac{TP}{TP + FP + FN} \quad (3.12)$$

其中 TP、FP、TN 和 FN 表示测试样本预测结果中真阳性、假阳性、真阴性和假阴性的数量。DSC、Recall(Rec)、Precision(Prec)和 Jaccard(Jac)的指标是基于测试集中每个病人案例的测试结果平均值，而 DSC(G)是基于测试集中所有病例的全局 DSC(DSC(G))<sup>[26]</sup>。召回率能够反应我们网络对于假阴性的改善，表明应对类内不一致问题的有效效应。准确率能够反应我们网络对于假阳性的改善，表明应对类间模糊问题的有效性。

为了验证我们的方法对目标边界的分割效果，我们增加了两个评估指标：平均表面距离(Average surface distance, ASD)和排名为 5%的豪斯多夫距离(95 percent Hausdorff distance, 95HD)，它们被证实可以用来评估边界的分割精度<sup>[77]</sup>。

### 3.5.3 损失函数计算

我们在网络训练中使用了增强混合损失来监督最终的分割结果，如公式(3.13)所示，由 focal loss 和 dice loss 组成，并且我们用 sobel loss<sup>[78]</sup>来监督边界分割的结果。我们用二进制交叉熵损失(BCE\_loss)来监督每一层的粗粒度图像块注意力输出的注意力图，通过损失函数直接监督的方式训练粗粒度图像块注意力模块，提升了注意力图在编码阶段的准确性。

为了防止对最终分割损失的过度影响，我们将粗粒度注意力损失的百分比设定为 0.15。最终，我们的损失函数有三个部分，如公式(3.13-3.15)所示，其中  $p$  是我们的方法预测的结果， $g$  是医生标注的金标准。 $cp$ 表示粗粒度注意力模块的先验位置信息， $cg_i$ 表示在第  $i$  阶段产生的粗粒度图像块注意力图的权重。Focal loss 中的超参数设置： $\gamma=2$ ， $\alpha=0.75$ 。

$$L_1 = \frac{1}{N} Focal\_loss(p, g) - \log(Dice\_loss(p, g)) \quad (3.13)$$

$$L_2 = Sobel\_loss(p, g) \quad (3.14)$$

$$L_{\varepsilon-i} = BCE\_loss(cp, cg_i) \quad i \in 1, 2, 3, 4, 5 \quad (3.15)$$

$$L = L_1 + L_2 + 0.15 \sum_{i=1}^5 L_{\varepsilon-i} \quad (3.16)$$

### 3.6 实验结果以及分析

#### 3.6.1 注意力引导多尺度恢复网络的消融实验

为了分析每个模块对分割精度的影响，我们用它们的消融组合进行了实验，结果如表 3.1 所示。从表中可以看出加入我们提出的任何一个模块都会使 DSC 得分高于骨干网络 D-UNet。我们提出的粗粒度图像块注意力模块，虽然是一种粗粒度的注意力增强方法，但因为粗粒度图像块注意力准确地增强了目标的空间信息，同时抑制了大量的背景噪声，所以在加入粗粒度图像块注意力模型后 Precision 和 Recall 都得到了提高，DSC 提高到了 0.574，这表明粗粒度图像块注意力可以有效地引导网络提取突出区域的目标特征，解决类内不一致。通过添加跨维度特征融合模块，DSC 达到 0.572，95HD 精度提升到 30.197，表明边界分割更加准确。Precision 提高到 0.701，有效地减少了假阳性，表明跨维度特征融合模块可以很好地应对类间模糊的挑战。这两个模块的组合将 DSC 提高到 0.581，并减少了 ASC 和 HD。然而，上述两个模块仅限于编码阶段，它们所提取的空间和边界信息会由于池化操作而丢失。因此我们解码阶段的上采样通过多尺度反卷积上采样模块恢复了高层次语义特征的边界和位置细节信息，所以 DSC，ASD 和 95HD 均得到提升。所有模块的组合取得了最好的结果，这表明了所有组件的有效性。

表 3.1 每个模块对骨干网络 D-UNet 的影响

Table 3.1 Impact of each module on the backbone network D-UNet

CPA	CFF	MDU	DSC(G)	DSC	Rec	Prec	ASD	95HD
			0.726	0.548	0.530	0.665	8.932	36.552
√			0.741	0.574	0.561	0.682	7.833	33.544
	√		0.734	0.572	0.538	0.701	8.670	30.197
		√	0.735	0.570	0.541	0.701	7.972	27.220
√	√		0.744	0.581	0.569	0.667	7.521	29.174
√	√	√	<b>0.754</b>	<b>0.594</b>	<b>0.579</b>	<b>0.713</b>	<b>7.137</b>	<b>27.005</b>

#### 3.6.2 注意力引导多尺度恢复网络与最先进方法的比较

为了证明该方法的先进性，我们将其与一些经典方法(2D U-Net<sup>[15]</sup>, Segnet<sup>[79]</sup>, Pspnet<sup>[53]</sup>, Deeplab v3<sup>[17]</sup>, UNet++<sup>[44]</sup>, 3D Unet<sup>[80]</sup>)、已有的脑卒中分割方法(D-Unet<sup>[26]</sup>, CICI-Net<sup>[16]</sup>, XNet<sup>[20]</sup>)和嵌入了注意力的方法(Attention Unet<sup>[50]</sup>,

ACSNet<sup>[47]</sup>)进行比较。如表 3.2 所示, 在 ATLAS 数据集中, AGMR-Net 在 D-Unet 的基础上将脑卒中分割的 DSC 为 0.594, DSC(G)为 0.754, Recall 为 0.579, Precision 为 0.713。在 ISLES 数据集中, AGMR-Net 在分割脑卒中的 DSC 为 0.614, DSC(G)为 0.753, Recall 为 0.623, Precision 为 0.633。

表 3.2 我们的方法和先前的方法比较

Table 3.2 Comparison of our method with previous methods

方法	ATLAS					ISLES				
	DSC	DSC (G)	Rec	Prec	Jac	DSC	DSC (G)	Rec	Prec	Jac
2D U-Net	0.485	0.712	0.469	0.574	0.368	0.504	0.658	0.512	0.601	0.381
Segnet	0.451	0.662	0.440	0.545	0.337	0.256	0.410	0.242	0.316	0.331
Pspnet	0.468	0.672	0.443	0.539	0.348	0.393	0.480	0.607	0.316	0.191
Deeplab v3	0.415	0.641	0.399	0.492	0.296	0.414	0.632	0.421	0.642	0.234
Unet++	0.511	0.728	0.494	0.638	0.400	0.316	0.384	0.617	0.218	0.205
ACSNet	0.520	0.703	0.462	0.706	0.401	0.442	0.550	0.448	0.554	0.219
Attention Unet	0.536	0.706	0.482	<b>0.721</b>	0.406	0.516	0.685	0.559	0.531	0.396
D-UNet	0.548	0.726	0.530	0.665	0.407	0.595	0.736	0.593	<b>0.774</b>	<b>0.466</b>
X-Net	0.499	0.672	0.466	0.633	0.382	0.437	0.601	0.418	0.552	0.315
CLCI-Net	0.581	-	0.581	0.649	-	0.405	0.643	0.423	0.453	0.305
3D UNet	0.471	0.710	0.474	0.553	-	0.569	0.737	0.577	0.527	0.401
AGMR-Net	<b>0.594</b>	<b>0.754</b>	<b>0.579</b>	0.713	<b>0.468</b>	<b>0.614</b>	<b>0.753</b>	<b>0.623</b>	0.633	0.443

图 3.6 显示了七种方法的预测结果与金标准的差异, 我们将预测结果与金标准做减法运算得到差异图, 差异图中黄色为假阴性, 黑色为假阳性, 这样更方便地显示我们的方法对类内不一致和类间模糊的影响。由于类内不一致, 网络在差异图的黄色区域预测了错误的假阴性, 而由于类间模糊, 在黑色区域预测了错误的假阳性。可以看出, 我们的方法产生了最少的假阴性和假阳性区域, 有效提高了网络的分割性能。

图 3.7 直观地比较了我们和其他方法在脑卒中分割边界分割效果上的比较, 其中红线代表金标准的边界。为了更好地恢复目标边界, UNet++和 CLCI-Net 在编码阶段修改了跳连接, Deeplab V3 使用了多尺度信息, 但边界分割效果仍然不如我们的好。这是由于跨维度特征融合模块在编码阶段捕获了更准确的边界特

征，而且多尺度反卷积上采样在解码阶段很好地恢复了边界特征，我们可以看到我们的分割结果最接近真实边界。

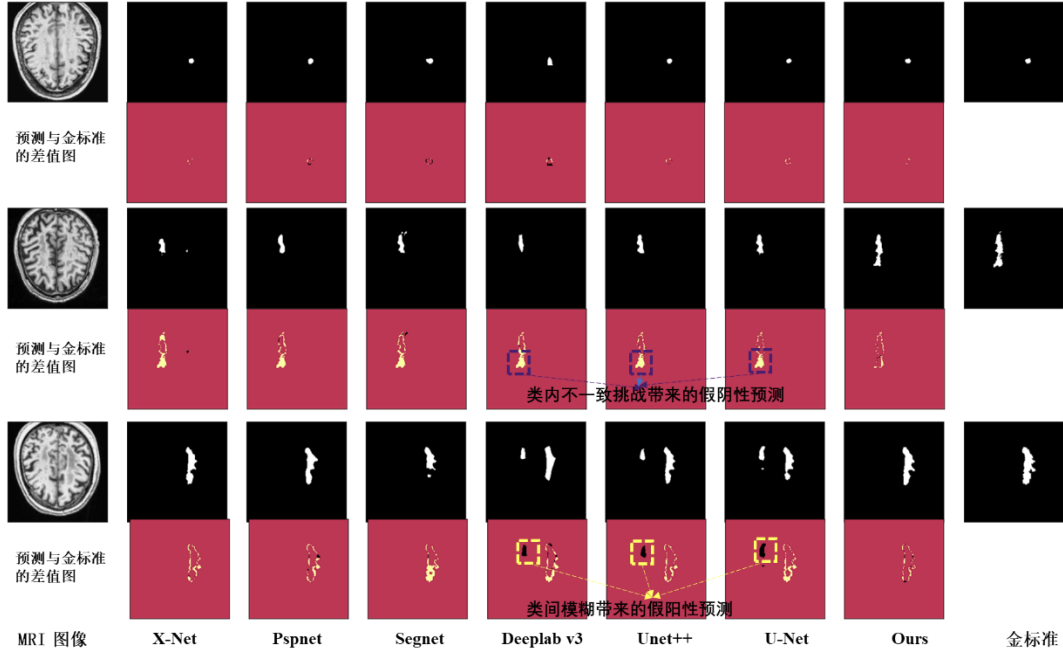


图 3.6 不同方法的分割结果图

Figure 3.6 Figure of segmentation results of different methods

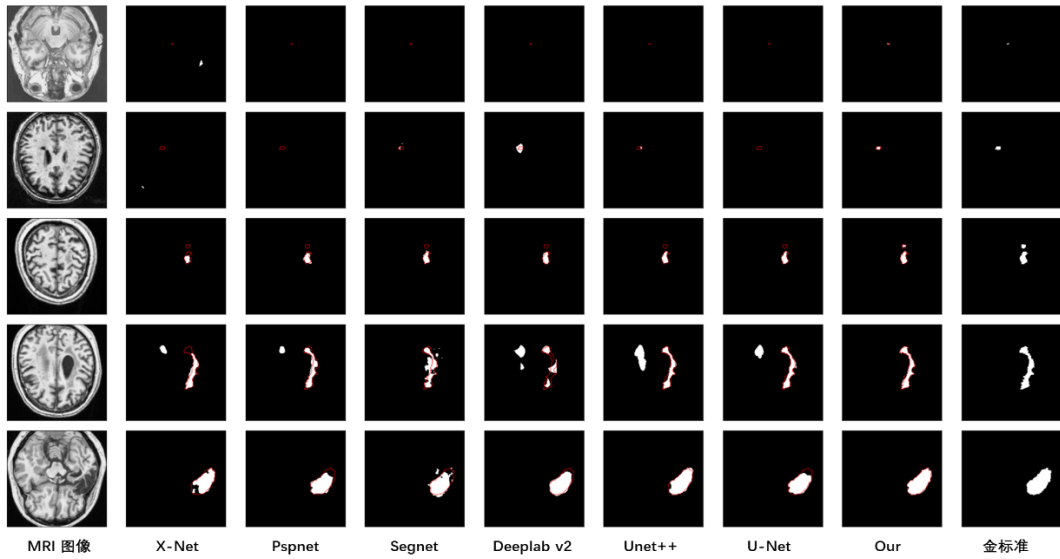


图 3.7 边界分割结果对比

Figure 3.7 Comparison of boundary segmentation results

表 3.3 呈现了我们的方法与其他方法在边界分割结果的定量比较。我们得到的 ASD 为 7.137，95HD 为 27.005。此外，我们将 AGMR-Net 的 ASD 和 95HD 分数减去其他网络的 ASD 和 95HD 分数。如表 3.3 所示，ASD 精度比我们的基线 D-UNet 高 1.795 分，95HD 高 9.550 分。与其他网络相比，证明了我们的方法获得了更好的边界分割效果。我们用 DSC 分数在我们的模型和其他传统模型之

间进行了显著性分析，如表 3.3 所示。实验后的  $p$ -value 都小于 0.05，此外，与骨干网络 D-UNet 相比，我们方法的  $p$ -value 远远小于 0.05。小于 0.05 的  $p$ -value 被认为是 DSC 提升显著的，证明了我们的方法与其他模型相比是有效的。

表 3.3 边界分割结果对比

Table 3.3 Comparison of boundary segmentation results			
方法	ASD	95HD	P-value
Ours-UNet	-8.008	-25.970	1.35E-5
Ours-Segnet	-5.089	-18.669	1.92E-6
Ours-PSPNet	-8.269	-17.908	4.34E-7
Ours-Deeplab v3	-7.058	-14.594	4.77E-7
Ours-Attention UNet	-0.811	-14.637	0.002
Ours-UNet++	-6.827	-10.542	4.06E-4
Ours-D-UNet	-1.795	-9.550	0.001
Ours-X-net	-4.474	-19.497	9.40E-5

图 3.8 显示了使用 DSC 来评估测试集的箱形图。我们的模型具有最高的 DSC 中值，即 0.594，表明与其他模型相比，它的表现很好，并且具有最高的 DSC 分数，表明它的上限很高。而且 DSC 得分的四分位相较于其他网络也在更高的位置，说明我们方法在每个案例上都取得了不错的分割结果，所以 DSC 得分经常集中在精度较高的位置。

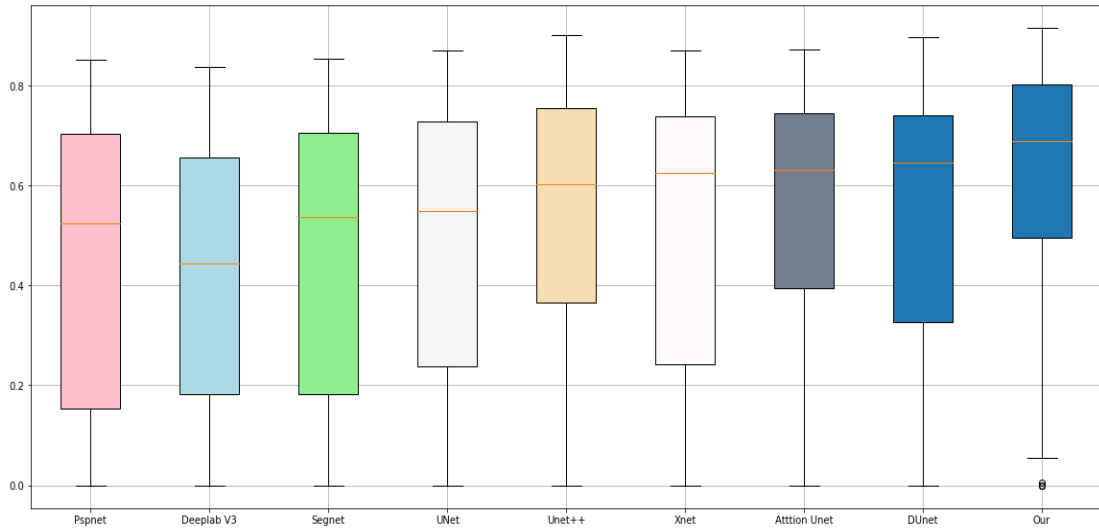


图 3.8 AGMR-Net 与其他模型的箱型图

Figure 3.8 Box plot of AGMR-Net with other models

### 3.6.3 粗粒度图像块注意力模块的实验分析

(1)粗粒度与细粒度注意力的比较。为了评估我们的粗粒度图像块注意力模块的有效性，我们将其与细粒度像素级空间注意方法进行了比较：经典的



Attention-Unet 和 D-UNet + FA(Fine attention)。表 3.4 的结果显示，与其他两种细粒度方法相比，我们的粗粒度图像块注意力方法获得了最好的性能，DSC 和 DSC(G)得分比 D-UNet + FA 高 0.023 和 0.017。这是因为我们的粗粒度图像块注意力方法在编码阶段是通过监督的方式进行学习的，保证了注意力图的准确性。而细粒度的注意力图正如图 1.2 所示，由于在编码阶段注意力的不准确限制了它的性能。

表 3.4 空间注意方法的比较

Table 3.4 Comparison of spatial attention methods

方法	DSC	DSC(G)
Attention-Unet	0.536±0.261	0.706
D-UNet + FA	0.551±0.253	0.724
D-UNet + CPA	<b>0.574±0.272</b>	<b>0.741</b>

(2)粗粒度图像块注意力突出目标空间（位置）信息的过程。如图 3.9 所示，第二行是我们的粗粒度图像块注意力产生的注意力图，其中颜色鲜艳的区域代表网络需要注意的图像块，表示目标区域出现在这些图像块中的概率较高。比较注意前后特征图的背景和病变区域，我们可以发现任务相关（病变）区域明显增强，说明该模块可以有效地捕捉目标的空间信息，从而帮助网络更好地提取目标特征，排除目标不相关特征的干扰，减少噪声。在编码器和解码器的特征融合的过程中，低噪声的特征图也被送到深层的解码器中，这更有利于解码器提取目标特征。同时，粗粒度注意力图的生成可以通过先验知识的监督来训练，并且随着网络层数的加深，注意力图变得更加准确，可以与先验位置图保持一致。如图 3.9 所示，由于目标在每一层特征图中的相对位置变化不大，粗粒度图像块注意力模块能够通过先验知识监督保证生成准确的注意力图，所以图像块注意力能准确地突出每一层目标的空间信息，增强目标和背景之间的对比度。虽然第一列和第二列的注意力图没有后三列的准确，但注意后的特征图中目标区域的图像块与背景图像块相比还是很明显的。此外，我们借用残差网络的思想，使特征图中的重要信息不会因为前期注意力图的错误而丢失。

(3)突出的目标空间信息可以引导网络提取特定于目标的高质量特征。因为与目标相关的特征可以更好地识别目标，所以特征之间的差异性会增加<sup>[81]</sup>。而每个通道都是不同的卷积特征对输入特征图的响应，所以通道之间的差异性和通道权重之间的方差也会增加<sup>[81]</sup>。为了更有效地说明粗粒度图像块注意力模块可以

引导网络提取与目标相关的特征,我们可以通过跨维度特征融合模块为每个通道分配的权重值来定量地支持。如图 3.10 (b)和(d)分别是加入粗粒度图像块注意力模块后两层跨维度特征融合模块得到的 2D 特征图不同通道的权重。观察向量图,加入粗粒度图像块注意力后,向量图的色差增大,这说明通道权重值之间的方差增大。方差的增加证明了加入粗粒度图像块注意力模块后,网络提取了更高质量的与目标相关的特征。特别是在(d)中,向量图的方差更加明显。这是因为随着网络的深入,我们的粗粒度图像块注意力模块提供了更准确的病变位置信息,提取的特征更具有差异性,更与分割目标相关,这也将导致通道权重之间的方差更大。

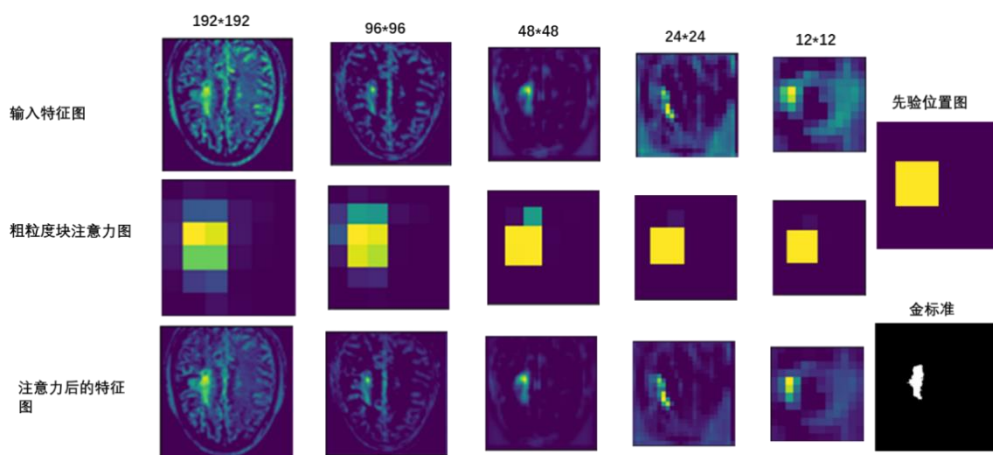


图 3.9 粗粒度注意力的可视化效果图

Figure 3.9 Visualization of coarse-grained attention

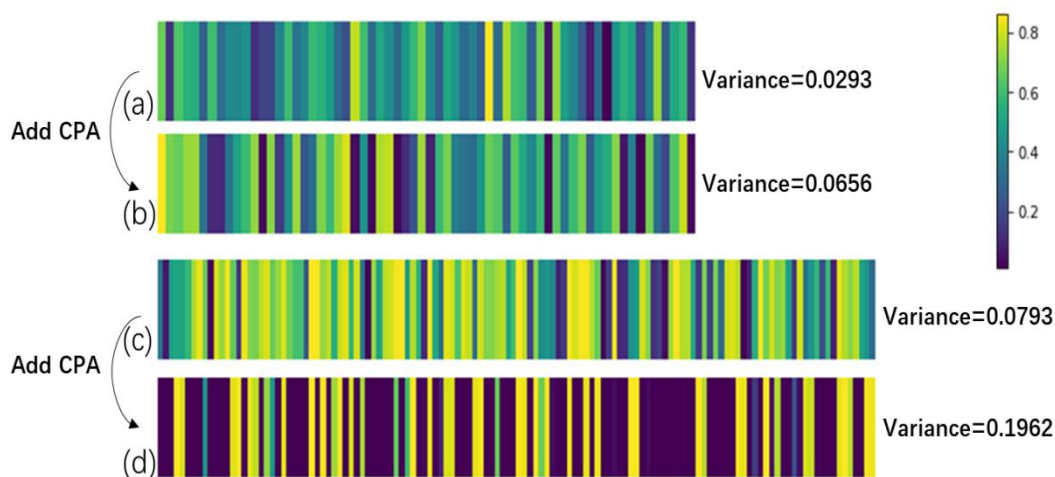


图 3.10 CPA 模块对特征权重的影响

Figure 3.10 Impact of CPA module on feature weights

### 3.6.4 跨维度特征融合模块的实验分析

(1)特征图通道的加权过滤有利于提高结果的准确性: 两个不同维度的特征融合使用简单的串联(Concatenate)或相加(Add)操作会引入无用或冗余的信息。相

比之下,因为建模了不同维度特征图之间全局通道的依赖性并对输出的特征图进行不同的加权过滤,我们的跨维度特征融合模块实现了更准确的预测。如表 3.1 所示,与 D-UNet 相比,DSC、DSC(G)、ASD 和 HD 的准确率提升分别为 0.024、0.008、0.262 和 6.29,因为 D-UNet 中的方法只是简单地对单维特征通道之间的关系进行建模,而跨维度特征融合模块则捕捉到所有维度特征之间的依赖关系。我们也用 add 或 concatenate 代替 AGMR-Net 中的维度特征融合模块来融合 2D 和 3D 特征,从表 3.5 可以看出在特征融合前进行加权过滤的跨维度特征融合优于直接 add 或 concatenate,可以增强特征融合的效果。我们还将跨维度特征融合模块中的 add 操作改为 concatenate(CFF + Concatenate)。虽然 DSC 和 DSC(G)略低于 AGMR-Net,但 concatenate 增加了特征通道的数量和卷积核数量,所以我们提出了使用 Add 的跨维度特征融合模块。

表 3.5 CFF 与普通特征融合方法的比较

Table 3.5 Comparison of CFF and common feature fusion methods

方法	DSC	DSC (G)
Add	0.588±0.248	0.750
Concatenate	0.583±0.257	0.749
CFF-Concatenate	0.593 ± 0.257	0.751
AGMR-Net	<b>0.594±0.273</b>	<b>0.754</b>

(2)不同维度间全局信息更有利于特征的建模和过滤: 为了强调全局信息的重要性,验证跨维度特征融合模块的效果是由于全局信息建模,而不是简单增加模型参数。我们将跨维度特征融合模块中的衰减率降低到 32,以减少参数的数量。跨维度特征融合模块中的参数数量为 $2C * \frac{2C}{32} + 2 \left( \frac{2C}{32} * C \right) = \frac{C^2}{4}$ ,这与 D-UNet 中的特征融合模块的参数数量 $2 \left( C * \frac{C}{16} + \frac{C}{16} * C \right) = \frac{C^2}{4}$ 相同。但从表 3.6 中可以看出,CFF(R=32)模块取得的 DSC 指标高于 D-UNet,并且与 D-UNet 相比跨维度特征融合模块的 p-value 是 0.039,改进效果仍然显著。这表明,CFF 模块的增强是由于全局信息建模,而不是增加模型参数。

表 3.6 不同衰减率对基线的提升

Table 3.6 Enhancement of baseline by different decay rates

方法	DSC (G)	DSC	P-value
Baseline	0.726	0.548±0.264	
+ CFF(R = 32)	0.731	0.566±0.261	0.039
+ CFF(R = 8)	<b>0.734</b>	<b>0.572±0.244</b>	0.001

### 3.6.5 多尺度反卷积上采样模块的消融实验

(1)多尺度信息更利于恢复目标特征。在图 1.1 中，我们可以看到，经过四次下采样后的特征图是非常抽象的。此时，使用简单的插值上采样或单卷积核大小的反卷积过于粗糙。因此需要在完成上采样操作的同时，借助多尺度的上下文信息更好地恢复病灶的细节特征。从表 3.1 可以看出，加入多尺度反卷积上采样模块后，分割效果得到改善，召回率的提高反映了我们方法分割位置的准确性，ASD 和 95HD 反映了分割边界效果的提高。在表 3.7 中，我们将多尺度反卷积上采样模块与单尺度反卷积方法进行比较，结果显示，当卷积核大小为 3、5、7 时，DSC 得分反而卷积核大小的增加而降低。只有当卷积核大小为 9 时，DSC 才会提高。然而，使用所有大小为 9 的卷积核会使网络参数过载，忽略了病变大小的多样性，导致平均 DSC 低于我们的多尺度方法。出于计算量和精准度的衡量，我们使用多尺度反卷积上采样模块。

表 3.7 MDU 与单尺度反卷积的比较

Table 3.7 Comparison of MDU and single-size deconvolution

方法	DSC	DSC(G)	Parameters
AGMR-Net	<b>0.594±0.273</b>	<b>0.754</b>	23799283
Size = 3	0.579±0.278	0.745	<b>12658163</b>
Size = 5	0.552±0.265	0.727	18228723
Size = 7	0.578±0.739	0.735	26584563
Size = 9	0.587±0.251	0.753	37725683

(2)上采样阶段对目标的恢复不应被忽视。许多方法在编码或跳连接处捕获目标的多尺度信息，但忽略了在网络的上采样阶段使用多尺度信息来恢复目标空间和边界信息的必要性。在表 3.8 中，我们把 AGMR-Net 中的多尺度反卷积上采样模块替换到网络的编码阶段或编码器与解码器之间跳连接阶段。虽然在编码或跳连接阶段加入 MDU 模块后，两者都有改善，但分割结果不如嵌入在上采样阶段。

表 3.8 MDU 模块在不同位置的整合

Table 3.8 Integration of MDU modules in different locations

方法	DSC	DSC(G)
编码器阶段	0.586±0.523	0.748
跳连接阶段	0.583±0.278	0.748
AGMR-Net	<b>0.594±0.273</b>	<b>0.754</b>

### 3.7 本章小结

本章中我们介绍了注意力引导多尺度恢复网络 AGMR-Net, 并介绍了为了解决编码阶段类内不一致和类间模糊问题所设计的粗粒度图像块注意力, 跨维度特征融合和多尺度反卷积上采样模块。最后也通过与先进的脑卒中分割方法的对比, 以及对每个方法的消融分析证明了我们方法设计的合理性和必要性。

## 第四章 基于粗粒度残差学习的脑卒中分割网络

### 4.1 目标感知监督残差学习网络整体框架

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。

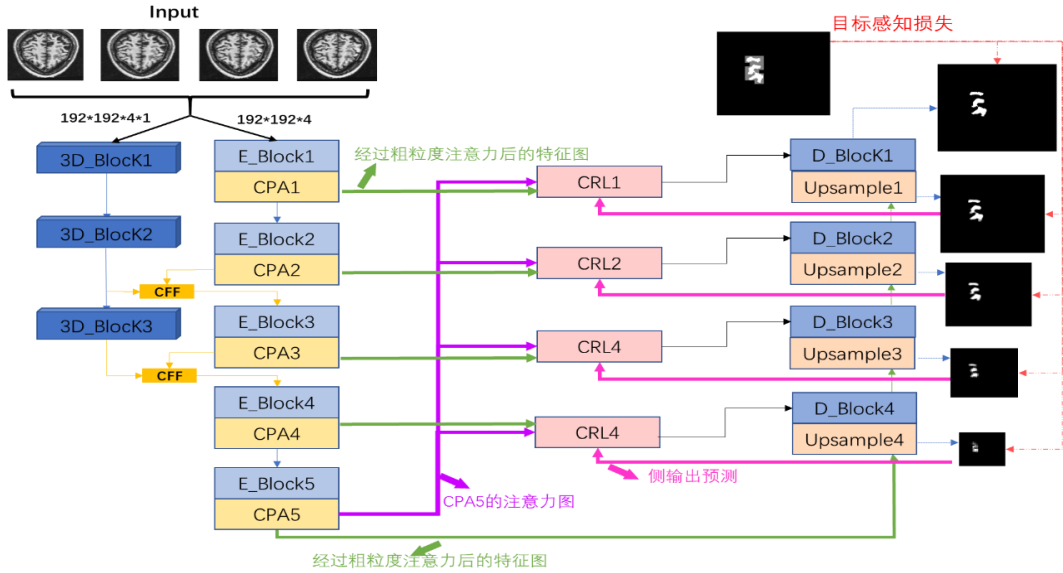


图 4.1 目标感知监督残差学习网络结构

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 粗粒度残差学习模块

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<sup>[15]</sup>，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。

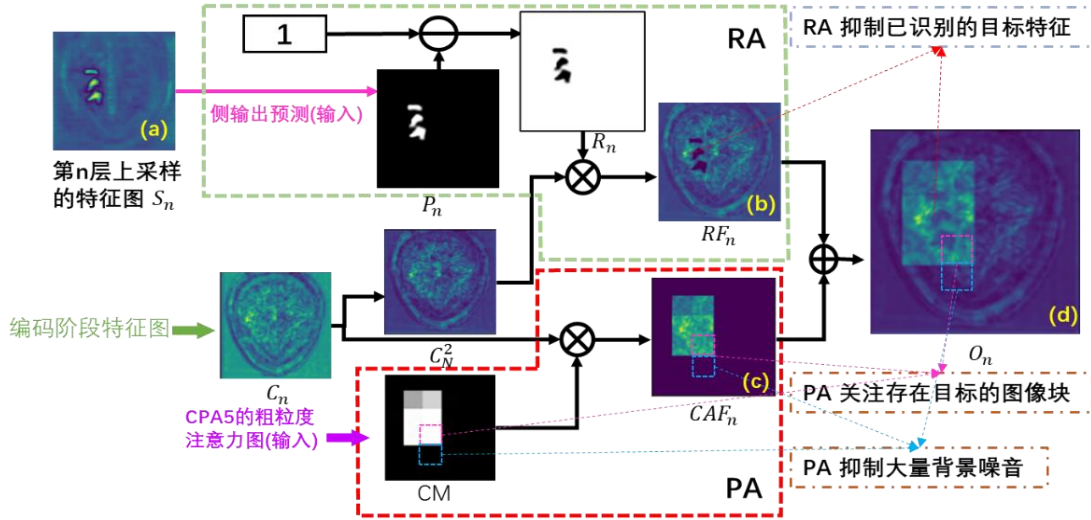


图 4.2 粗粒度残差学习模块结构图

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  $C_n$  和第 n 层上采样的侧输出  $P_n$  作为输入，RA 单元的输出可以被定义为。

$$R_F_n = C_n^2 \otimes (1 - P_n) \quad (4.1)$$

$\otimes$  表示元素相乘。  $C_n^2 = C_n \otimes C_n$  可以突出显示卷积识别的对象。反向注意力图  $R_n$  由  $1 - P_n$  得到。通过观察图 4.2 (b)中  $R_F_n$ ，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

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力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 单元的粗粒度注意力由 CPA5 给出，最终 PA 的输出被定义为：

$$CAF_n = C_n \otimes CPA_5 \quad (4.2)$$

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  $CAF_n$  和  $RF_n$  融合得到最终粗粒度残差学习模块的输出  $O_n$ 。

$$O_n = CAF_n + RF_n \quad (4.3)$$

**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 目标感知损失函数

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。

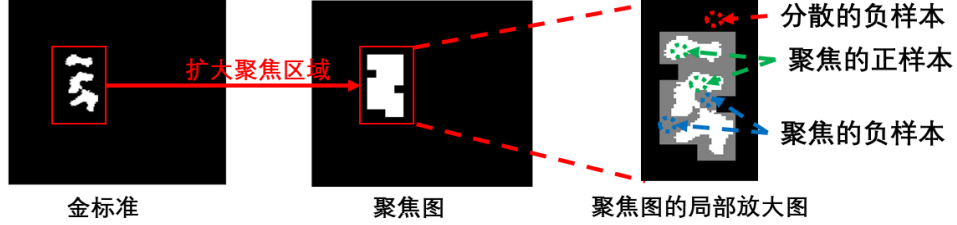


图 4.3 损失函数聚焦区域扩张图

Figure 4.3 Loss function focus area expansion map

**样本划分：**我们用  $\beta$  表示聚焦因子的值，用  $H$  表示图像的高度，用  $g$  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：

$$A_{x*\beta+m,y*\beta+n} = sign\left(\sum_{i=0}^{\beta} \sum_{j=0}^{\beta} g_{x*\beta+i,y*\beta+j}\right) \quad m \in [0, \beta), n \in \left[0, \frac{H}{\beta}\right) \quad (4.4)$$

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  $\beta$  为单位在金标准  $g$  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。

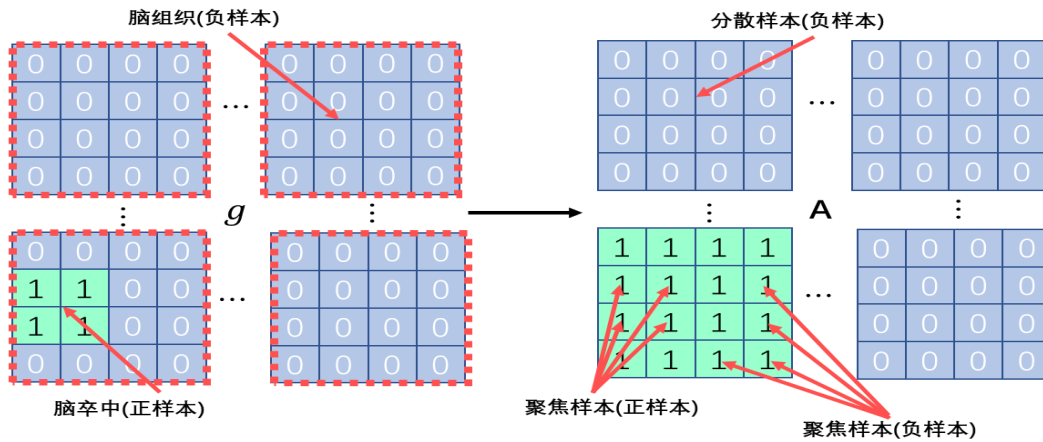


图 4.4 聚焦因子的作用图

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  $p_t$ ，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( $\alpha > 0.5$ )。设置大于 0.5 的  $\alpha$  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  $1-\alpha$ 。我们将新的静态权重映射到 focal loss 上，得到  $\text{loss}_{fp}$ ，其定义为公式(4.6)。

$$p_t = pg + (1 - p)(1 - g) \quad (4.5)$$

$$\text{loss}_{fp} = \frac{\sum_{i=0}^N -(A_i\alpha + |1 - A_i|\alpha)(1 - p_{t_i})^\gamma \log(p_{t_i})}{N} \quad (4.6)$$

其中 $\gamma$ 的取值范围为(0, 5)，用于增加对困难样本的惩罚<sup>[58]</sup>。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：

$$\text{loss}_t = \text{loss}_{fp} + \frac{\sum_{i=0}^N g_i p_i + \epsilon}{\sum_{i=0}^N p_i^2 + \sum_{i=0}^N g_i^2 + \epsilon} \quad (4.7)$$

## 4.4 实验细节

### 4.4.1 数据集

我们在开源的中风后病变解剖追踪(ATLAS)数据集上评估了我们的方法，该数据集包含 229 名中风患者的 T1 加权 MRI 扫描。每个病例的三维图像包含 189 个 233\*197 的切片，每个病例都是在 3T 磁共振扫描仪上采集的，切片厚度为 1.0mm\*1.0mm\*1.0mm，除了队列 1 和 2，它们是在 1.5T 扫描仪上采集的，切片厚度为 0.9mm\*0.9mm\*3.0mm。为了公平地评估该方法的有效性，我们使用了与骨干网络 AGMR-Net 相同的数据划分方式，选择 80% (183 例) 用于训练，其余 20% (46 例) 用于测试。为了适应网络的需要，我们将每个病例的三维图像，沿 Z 轴切片，并裁剪为 180\*180，最后，图像被调整为 192\*192。每个切片连同其前面的两个相邻切片和后面的一个相邻切片，总共四个通道，作为网络的输入。训练和测试集的输入总数分别为 34587\*192\*192\*4 和 8694\*192\*192\*4。

为了进一步评估我们的方法，我们需要在所有三个坐标方向上具有更好的三

维分辨率的图像数据来适应我们的网络结构。我们使用缺血性中风病变分割挑战赛(ISLES 2015)的整个数据集作为辅助验证数据集，并使用相同的 T1 模态作为我们网络的输入。所有的 MRI 序列都进行了头骨剥离，并重新取样到  $1\text{mm} \times 1\text{mm} \times 1\text{mm}$  的切片厚度，每个案例的大小为  $230 \times 230 \times 154$ 。由于数据量小，该数据集被分为五个相同大小的互斥子数据集，我们使用五折交叉验证来验证我们的模型。这个数据集的每个切片都被调整为  $192 \times 192$ ，其余的预处理与 ATLAS 数据集相同。训练集和测试集的输入总数分别为  $3384 \times 192 \times 192 \times 4$  和  $924 \times 192 \times 192 \times 4$ 。

我们在与先进方法对比中增加了两个网络 CPGAN<sup>[82]</sup>和 DPAC-UNet<sup>[83]</sup>，但由于这两份工作并未提供网络的具体细节或者代码，所以我们直接使用他们在论文的实验指标。

#### 4.4.2 实验指标

由于本章的目标监督残差学习网络是在第三章 AGMR-Net 的基础上进行的优化，所以用于评估模型的分割指标保留了第 3.5.2 章节中用的实验指标(DSC, DSC(G), Recall, Precision, Jaccard)，Recall 和 Precision 可以更好的帮助我们衡量本章算法对于假阴性和假阳性的影响。同时，我们在本章增加了评估指标：模型参数量(Parameters, Param)和浮点运算量(Floating point operations, FLOPs)，这两个指标常被用来评估模型的计算复杂度<sup>[84]</sup>。

#### 4.4.3 损失函数的计算

如图 4.1 所示，我们的网络有五个粗粒度图像块注意力的输出、四个上采样的侧输出和一个最终预测输出，对于每一个侧输出和最终预测输出，我们使用目标感知损失函数来计算损失。而每个粗粒度图像块注意力模块的注意力图使用二进制交叉熵函数进行监督，所以最终的损失函数定义为：

$$loss_{total} = l_{end} + \sum_{i=1}^4 ls_i + 0.15 * \sum_{i=1}^5 la_i \quad (4.8)$$

其中  $\gamma$  设为 2， $\alpha$  设为 0.75， $\beta$  设为 4。 $l_{end}$  表示最终预测的损失， $ls_i$  分别对应于四个侧输出的损失。 $la_i$  表示粗粒度图像块注意力模块产生的粗粒度注意力和先验位置图的损失。

对于其他对比实验，我们都使用 EML loss<sup>[26]</sup>函数进行训练。其中  $\alpha$  被设置为 0.75,  $\gamma$  被设置为 2。

## 4.5 实验结果及分析

### 4.5.1 消融研究

为了测试每种方法的有效性，我们移除 AGMR-Net 中解码阶段的多尺度反卷积上采样模块并作为骨干网络，然后逐渐加入我们的方法，实验的定量数据总结于表 4.1。我们首先通过使用目标感知损失函数给正样本损失分配高权重来改善对目标的监督和增加对假阴性的惩罚，以提高正样本的分类精度，因此召回率增加了 0.045，同时我们也补偿了部分负样本的损失来增加对假阳性的惩罚，避免了精确率的下降，所以 DSC 提高了 0.036。然后，加入粗粒度残差学习模块，RA 单元有效地增强了目标残差特征表示，减少了预测结果的假阴性；同时我们用 PA 单元来抑制目标附近的背景噪声，减少了预测结果的假阳性。所以加入粗粒度残差学习模块后，召回率、精确率和 DSC 都得到了提高。最后，在解码阶段，我们把这两个方法结合起来，循环的使用目标感知损失监督侧输出预测并且使用粗粒度残差学习模块学习残差特征，召回率和精确率都得到了提高。

表 4.1 消融实验的定量结果

Table 4.1 Quantitative results of ablation experiments				
方法	DSC(G)	DSC	Rec	Prec
Baseline	0.744	0.581	0.569	0.667
Baseline + CRL	0.755	0.602	0.605	0.673
Baseline + TAL	0.756	0.617	0.639	0.660
Baseline + CRL+TAL	<b>0.760</b>	<b>0.634</b>	<b>0.650</b>	<b>0.678</b>

### 4.5.2 目标感知监督残差学习网络的分析实验

我们通过与先进方法对比的定量指标证明网络的良好性能，如表 4.2 所示，我们的方法在两个脑卒中分割数据集上取得了最佳的平均 DSC、Recall、Jaccard 和 DSC(G)。这些改进是由于目标感知损失函数将监督重点放在目标区域，而且粗粒度残差学习模块在解码阶段增强了未识别的目标残差特征的表达。我们在这两个数据集上的召回率分别为 0.650 和 0.688，表明我们的方法与其他网络相比能够很好地提取目标特征，更有利于克服脑卒中分割结果的假阴性偏高的问题。

此外，精确率分别取得了第二和第四的好成绩，表明我们在提高了召回率的同时很好地平衡了精确率，没有使网络过度地挖掘目标正样本而忽略背景负样本。

表 4.2 在 ATLAS 和 ISLES 数据集上的对比

Table 4.2 Comparison on ATLAS and ISLES datasets

方法	ATLAS					ISLES				
	DSC	DSC (G)	Rec	Prec	Jac	DSC	DSC (G)	Rec	Prec	Jac
2D U-Net	0.485	0.712	0.469	0.574	0.368	0.504	0.658	0.512	0.601	0.381
Segnet	0.451	0.662	0.440	0.545	0.337	0.256	0.410	0.242	0.316	0.331
Pspnet	0.468	0.672	0.443	0.539	0.348	0.393	0.480	0.607	0.316	0.191
Deeplab v3	0.415	0.641	0.399	0.492	0.296	0.414	0.632	0.421	0.642	0.234
Unet++	0.511	0.728	0.494	0.638	0.400	0.316	0.384	0.617	0.218	0.205
ACSNet	0.520	0.703	0.462	0.706	0.401	0.442	0.550	0.448	0.554	0.219
Attention Unet	0.536	0.706	0.482	<b>0.721</b>	0.406	0.516	0.685	0.559	0.531	0.396
D-UNet	0.548	0.726	0.530	0.665	0.407	0.595	0.736	0.593	<b>0.774</b>	0.466
X-net	0.499	0.672	0.466	0.633	0.382	0.437	0.601	0.418	0.552	0.315
CLCI-Net	0.581	-	0.581	0.649	-	0.405	0.643	0.423	0.453	0.305
3D UNet	0.471	0.710	0.474	0.553	-	0.569	0.737	0.577	0.527	0.401
AGMR-Net	0.594	0.754	0.579	0.713	0.468	0.614	0.753	0.623	0.633	0.443
CPGAN	0.617	-	0.595	0.656	-	-	-	-	-	-
DPAC-UNet	0.592	-	0.650	0.678	-	-	-	-	-	-
Ours	<b>0.634</b>	<b>0.760</b>	<b>0.650</b>	0.678	<b>0.504</b>	<b>0.640</b>	<b>0.786</b>	<b>0.668</b>	0.634	<b>0.529</b>

图 4.5 直观的描述了对比实验的预测结果，我们把每个模型的预测结果与金标准作差值运算，所以图中的红色区域代表假阴性区域。很明显，我们的方法分割结果非常好，即使是在目标正样本非常少的情况下，假阴性区域也是最小的。而其他方法的预测结果含有大量的假阴性样本，这是因为在解码阶段，编码特征的简单连接带来的信息冗余或干扰。为此，Unet++改进了跳连接，以获得更好的结果。但是它们缺乏适当的监督，无法正确引导网络提取正样本信息，仍然会导致分类错误。相比之下，目标感知损失函数鼓励网络关注正样本，并与解码器的粗粒度残差学习模块协同工作，因此我们的分割结果具有最少的假阴性。

为了证明我们方法对召回率的改进效果，我们分别测量了 AGMR-Net 和

TSRL-Net 在 ATLAS 测试集上每个案例分割结果的召回率，以进一步评估我们的方法对网络召回率改进的有效性。图 4.6 显示了比较结果，蓝色区域显示 TSRL-Net 的值高于 AGMR-Net，红色表示 TSRL-Net 的值低于 AGMR-Net。可以看出，TSRL-Net 在大多数情况下表现更好，召回率明显提高，只有在极少数情况下，召回率降低。

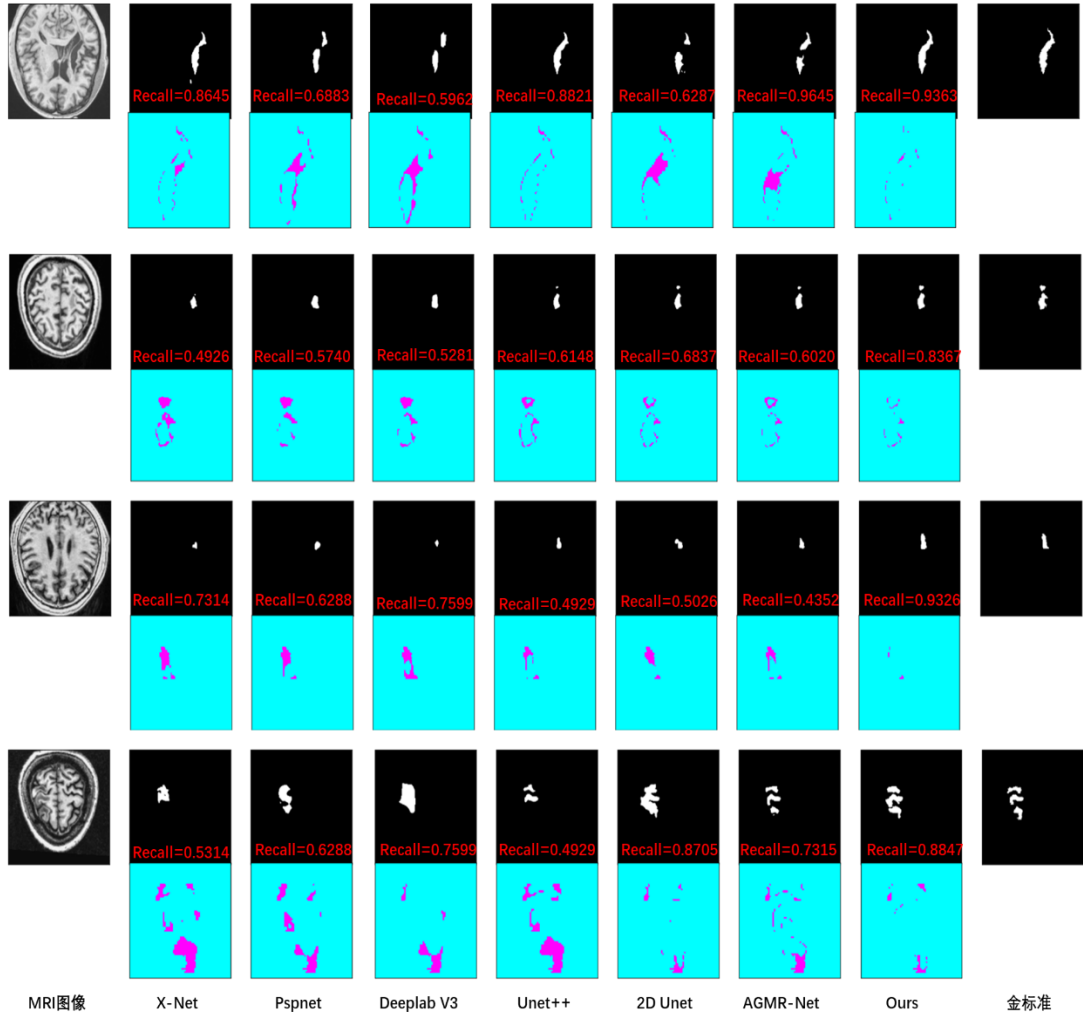


图 4.5 不同方法预测结果的比较

Figure 4.5 Comparison of prediction results of different methods

DSC 反映了最终的分割效果，可以有效地衡量召回率和精确率。如图 4.7 所示，我们使用测试集上所有 MRI 扫描的分割结果的 DSC 分布箱线图来评估不同模型的性能。我们首先与骨干网络 AGMR-Net 和使用 2D 和 3D 融合网络的 D-UNet 进行比较，可以观察到我们的方法的箱线图最小值以及 DSC 分数的下四分位数都有明显增加。对于具有反向注意力的 ACSNet 和具有正向注意力的 Attention Unet 也是如此，表明我们的方法明显提高了网络的分割性能，表明我

们的方法在提高召回率的同时很好地平衡了精确率。

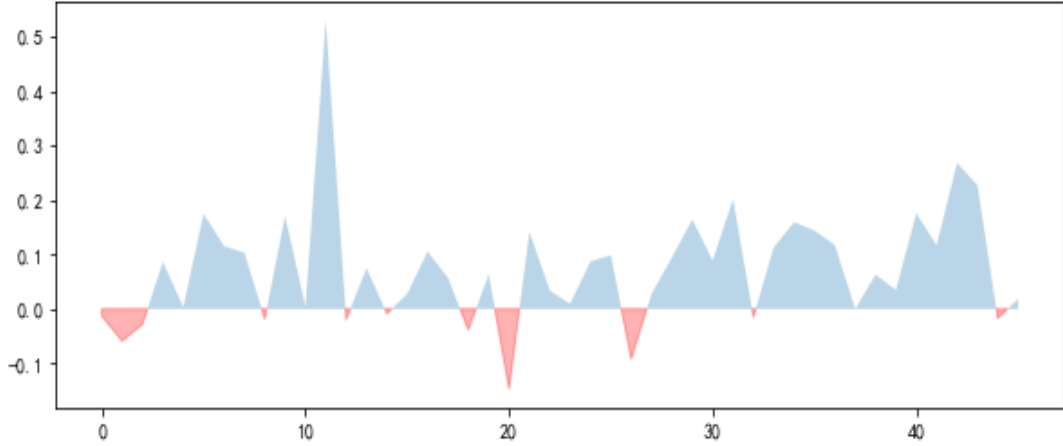


图 4.6 AGMR-Net 和 TSRL-Net 在召回率上的比较

Figure 4.6 Comparison of AGMR-Net and TSRL-Net in terms of Recall

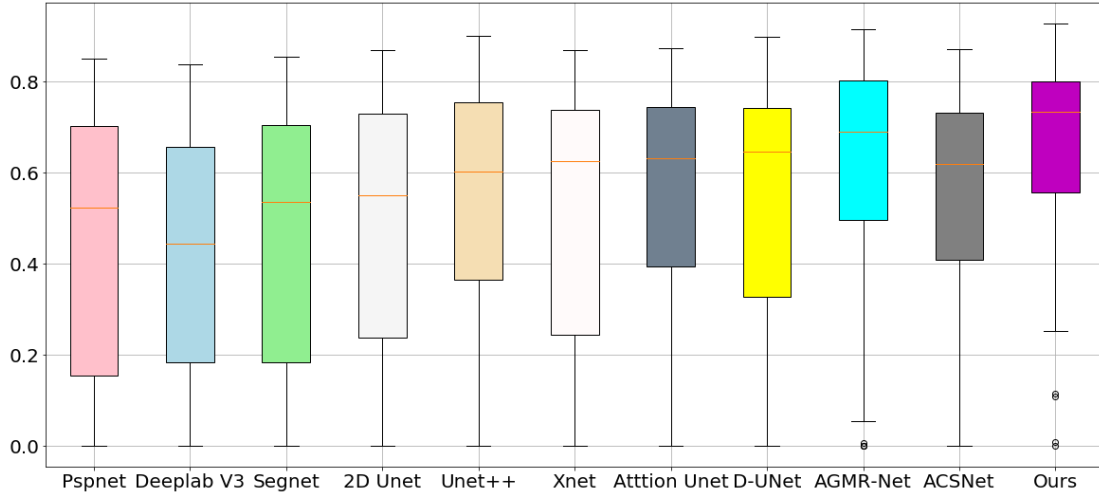


图 4.7 不同模型的 DSC 结果的箱形图

Figure 4.7 Box plots of DSC results for different models

最近的研究<sup>[85,86]</sup>通常使用浮点运算量(FLOPs)和参数量(Param)来证明网络的性能改进不是简单地增加网络的复杂性。如表 4.3 所示, X-Net 使用了深度可分离卷积所以获取了最少的 FLOPs, D-UNet 在减少了每层卷积的通道数, 获得了最少的 Param。我们的方法与骨干网络 AGMR-Net 相比, FLOPs 和 Param 都有所减少, 而且从表 4.2 中可以看出, 我们模型的精度得到很大的提升, 表明我们的方法是一个相对轻量级和高效的网络。与其他网络模型相比, 虽然我们的方法不是最轻量级的, 但在显著性分析实验中, 我们的 p-value 都远远低于 0.05 (p-value 值小于 0.05 说明方法的提升是显著的, 值得改进<sup>[87]</sup>)。医学图像分割是需要高精度的学科, 所以出于在计算量和精度的考量, 我们的方法具有一定的竞争性。

表 4.3 复杂性分析和显著性分析实验

Table 4.3 Complexity analysis and significance analysis experiments

方法	FLOPs	Param	P-value
AGMR-Net	42872026884	23799283	0.005
D-Unet	18089374976	<b>9503779</b>	5.84E-6
X-net	<b>15103904</b>	15072332	2.59E-6
Unet++	42154868736	26897860	2.887E-6
Attention Unet	37437947136	34877421	1.74E-5
Deeplab v3	10454852496	58158145	2.54E-8
Segnet	22531276800	29442433	6.21E-8
Pspnet	22378374144	27494276	2.89E-7
2D UNet	22545469440	17266241	1.85E-6
Ours	23916119454	12854496	-

#### 4.5.3 目标感知损失的实验分析

**(1)聚焦因子  $\beta$  的作用:** 为了说明聚焦因子在损失函数中的作用,我们固定  $\gamma=2$ , 调整  $\beta$  和  $\alpha$  的值训练目标感知监督残差学习网络, 结果如图 4.8 所示。当  $\beta=1$  时, 我们增加  $\alpha$  的值, 增加网络对正样本的关注度以增加了对假阴性的惩罚, 召回率增加, DSC 随之增加, 但是精确率明显下降。这表明增加正样本的权重  $\alpha$ , 可以使正样本损失对总损失的占比增加, 促进网络提取正样本信息, 在正负样本不均衡的情况下有利于 DSC 和召回率的提升。然而, 当  $\alpha=0.75$  时, 网络过于关注正样本而忽略了负样本, 召回率虽然提升, 但是精确率却大幅度的下降, DSC 得分下降。因此, 我们引入聚焦因子  $\beta$ , 通过增加  $\beta$  的值, 扩大聚焦区域, 以弥补负样本的损失, 避免负样本损失的失衡。当  $\alpha=0.75$  时, 增加  $\beta$ , 被聚焦的负样本的数量增加, 负样本分类的错误率降低, 假阳性样本减少, 精确率增加。当  $\beta=4$  时, DSC 和召回率达到最佳值 0.634 和 0.651, 证实了通过  $\beta$  和  $\alpha$  之间的适当权衡增加负样本的损失会获得更好的模型性能, 此时召回率和精确率不像  $\beta=1$  时那样严重失衡, 精确率达到了图中第二高。当  $\beta=8$  时, 过多的负像素被关注, 此时负样本损失在总损失的占比将远远大于正样本损失的占比, 精确率和召回率都下降。

**(2)动态权重  $\gamma$  和聚焦因子  $\beta$  对模型分割结果影响:** 如表 4.4 所示, 我们可以看到, 固定  $\beta$  值不变, 当  $\gamma=2$  时, DSC 是局部最优的。然而, 当  $\gamma$  过高或过低时, DSC 明显下降, 因为  $\gamma$  控制的是损失值的动态权重, 如果动态权重值过大或者过小, 此时难易样本之间得差异也会严重得失衡。当固定  $\gamma$  不变时, 我们增加



$\beta$ , 精确率基本上会增加, 因为我们扩大了对负样本损失的惩罚。当  $\gamma=1$  时, 困难样本和简单样本的损失值的动态权重相差比较小, 所以精确率下降, 但  $\beta$  增加到 8, 精确率也大幅度提高。所以, 我们建议  $\gamma$  值的范围是[1,5],  $\alpha$  得取值范围是(0.25-0.75),  $\beta$  得取值范围是[2-8]。

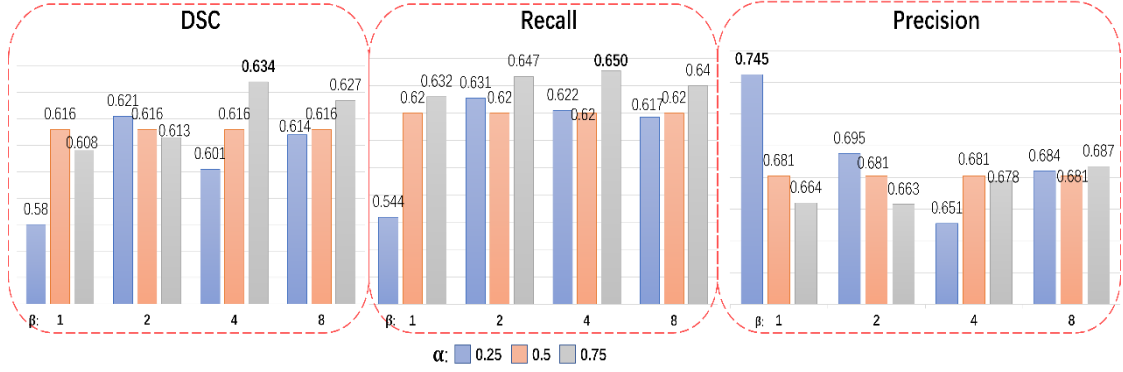


图 4.8  $\beta$  在目标感知损失函数中的影响

Figure 4.8 The effect of  $\beta$  in the target-aware loss function

表 4.4  $\gamma$  和  $\beta$  对损失函数的影响

Table 4.4 The effect of  $\gamma$  and  $\beta$  on the loss function

$r$	$\beta$	DSC	Rec	Prec
0	2	0.487	0.479	0.593
0	4	0.487	0.458	0.632
0	8	0.497	0.456	0.658
1	2	0.472	0.443	0.631
1	4	0.485	0.475	0.600
1	8	0.460	0.398	0.674
2	2	0.613	0.647	0.663
2	4	<b>0.634</b>	<b>0.650</b>	0.678
2	8	0.627	0.640	0.687
3	2	0.441	0.510	0.451
3	4	0.404	0.339	0.663
3	8	0.379	0.297	<b>0.709</b>
4	2	0.400	0.415	0.491
4	4	0.408	0.422	0.488
4	8	0.242	0.317	0.670

(3)目标感知损失函数在其他分割模型上的适用性: 我们利用整合了 focal loss 和 dice loss 的 Eml loss 用于表 4.2 中的部分网络。为了证明目标感知损失函数的适用性, 我们没有局限于我们提出的网络, 而是用目标感知损失函数代替了 Eml loss, 并分别进行了 U-Net、ACSNet、D-UNet 和 Attention Unet 的训练实验, 以

说明我们提出的目标感知损失函数在不同分割网络中的普遍适用性。表 4.5 显示了实验结果，表中加粗字体表示加入目标感知损失函数后提升的指标。当与这些分割网络整合时，目标感知损失增强了其他损失训练模型的性能，同时也提高了每个网络预测结果的召回率，并且我们避免了这些网络分割精确率的大幅下降，所以每个网络的 DSC 得分最终都得到了提升。这些发现支持了我们损失函数的通用性，并表明补偿负样本损失以平衡召回率和精确率是一个可行的选择。

表 4.5 目标感知损失在其他网络上的适用性

Table 4.5 Applicability of target-aware loss on other networks

方法	目标感知损失	DSC	Rec	Prec
U-Net		0.485	0.469	0.574
	✓	<b>0.528</b>	<b>0.493</b>	<b>0.697</b>
ACSNet		0.520	0.462	0.706
	✓	<b>0.530</b>	<b>0.488</b>	0.698
D-UNet		0.548	0.530	0.665
	✓	<b>0.559</b>	<b>0.540</b>	<b>0.769</b>
Attention Unet		0.536	0.482	0.721
	✓	<b>0.562</b>	<b>0.512</b>	0.691
Baseline		0.581	0.569	0.667
	✓	<b>0.617</b>	<b>0.639</b>	0.660

(4)与其他医学图像分割中损失函数的比较：我们使用 TSRL-Net 比较了不同的损失函数，结果如表 4.6 所示，我们使用目标感知损失获得了最佳的 DSC 分数 0.634。因为平等的对待正负样本损失，二进制交叉熵损失函数的 DSC 得分只有 0.545。而针对样本不平衡设计的 Dice loss 获得了 0.600 的 DSC，但是 Recall 得分较低。继承了交叉熵损失和 Dice loss 优点的 Cross-entropy loss + Dice loss 获得了 0.619 的 DSC。在可以控制正负样本损失权重的损失函数中，我们首先设置正样本权重  $\alpha=0.25$ ，负样本权重则为 0.75，毫无疑问，此时的负样本损失占比远远大于正样本损失的占比。所以使用 Tversky loss、EML loss 和 Focal loss 的训练的分割 Recall 和 DSC 分数都较低。但是在目标感知损失中，由于部分正样本损失的权重与负样本相同，所以分割结果的召回率为 0.622。当我们增加  $\alpha=0.75$  时，正样本损失占比增加，每个目标感知损失函数训练的模型分割结果的召回率有所提高，DSC 得分增加。然而，召回率的提高伴随着精确率的降低，所以不能达到最佳结果。在目标感知损失函数中，不仅关注所有正样本损失，以减少假阴性，提高召回率，而且还补偿了负样本损失，增加了对假阳性的惩罚，避免了精确率

的大幅下降，所以我们提出的目标感知损失因此获得了最高的 DSC。

表 4.6 不同损失函数的对比

Table 4.6 Comparison of different loss functions

损失函数	DSC	Rec	Prec
Cross-entropy loss	0.545	0.524	0.663
Dice loss	0.600	0.587	0.693
Cross-entropy loss + Dice loss	0.619	0.613	0.711
Focal loss $\alpha = 0.25$	0.549	0.515	0.652
Focal loss $\alpha = 0.75$	0.568	0.595	0.611
Tversky loss $\alpha = 0.25$	0.534	0.497	0.658
Tversky loss $\alpha = 0.75$	0.615	<b>0.693</b>	0.608
EML loss $\alpha = 0.25$	0.580	0.544	<b>0.745</b>
EML loss $\alpha = 0.75$	0.602	0.605	0.673
Tal $\alpha = 0.25$	0.601	0.622	0.651
Tal $\alpha = 0.75$	<b>0.634</b>	0.650	0.678

#### 4.5.4 粗粒度残差学习模块的实验分析

(1)反向注意力和粗粒度残差学习对召回率和准确率的影响：在解码阶段中，捕捉目标残差的细节特征有助于减少假阴性预测，提高结果的召回率。因此，大多数网络在跳连接使用反向注意力聚集编码器和解码器的特征，但编码阶段的噪声极多，导致网络在学习残差特征时会受到无关信息的影响，从而导致结果的精确率降低。我们在骨干网络的基础上做了一个实验，首先嵌入反向注意力(+Reverse attention)，可以看到召回率提高了，但是精确率下降，所以 DSC 的提升并不高。这是因为反向注意力只抑制了侧输出中已预测的目标特征，这样就相对增强了所有背景噪声的特征表达，导致结果精确率下降。为了解决这个问题，我们提出了粗粒度残差学习模块来指导残差学习。如图 4.2 所示，引入的粗粒度正向注意力大大抑制了背景噪声，减少了噪声给网络带来的干扰，提高了精确率。表 4.7 还给出了定量分析，粗粒度残差学习与反向注意力相比，不仅再次提高了召回率，而且还提高了精确率，说明粗粒度反向注意力模块的引入可以引导网络在更精确的区域进行残差特征的学习，从而使 DSC 得到大幅度提高。

表 4.7 不同的残差学习模块的比较

Table 4.7 Comparison of different residual learning modules

方法	DSC	Rec	Prec
Baseline	0.581	0.569	0.667
+ Reverse Attention	0.587	0.595	0.647
+CRL	<b>0.602</b>	<b>0.605</b>	<b>0.673</b>

**(2)反向注意力和粗粒度残差学习的可视化对比：**如图 4.9 所示，其中 a 行表示含有粗粒度残差学习模块的侧输出，b 行表示只有反向注意力的侧输出。可以看出，由于引入了反向注意力，解码器可以进行有效的残差学习，从侧输出 1 到侧输出 5，随着预测图的分辨率逐渐提高，目标的空间信息和边界信息等细节信息逐渐被修复。粗粒度残差学习模块进一步嵌入了正向注意力模块，抑制了大量的背景噪声，这使得网络可以更加专注于提取残差特征，甚至像 a1,a2 行中绿框中的一些小的残余特征也被修补。而对于 b1 和 b2 行，只有反向注意预测的结果，却没有做到这一点。另一方面，b1,b2 和 b3 行的红框中，有噪声特征的引入，从而降低了结果的精确率。相反，在 a3 行中的嵌入粗粒度残差学习网络的预测结果中，虽然侧输出 2 中有噪声引入，但随后被修正，提高了精确率。

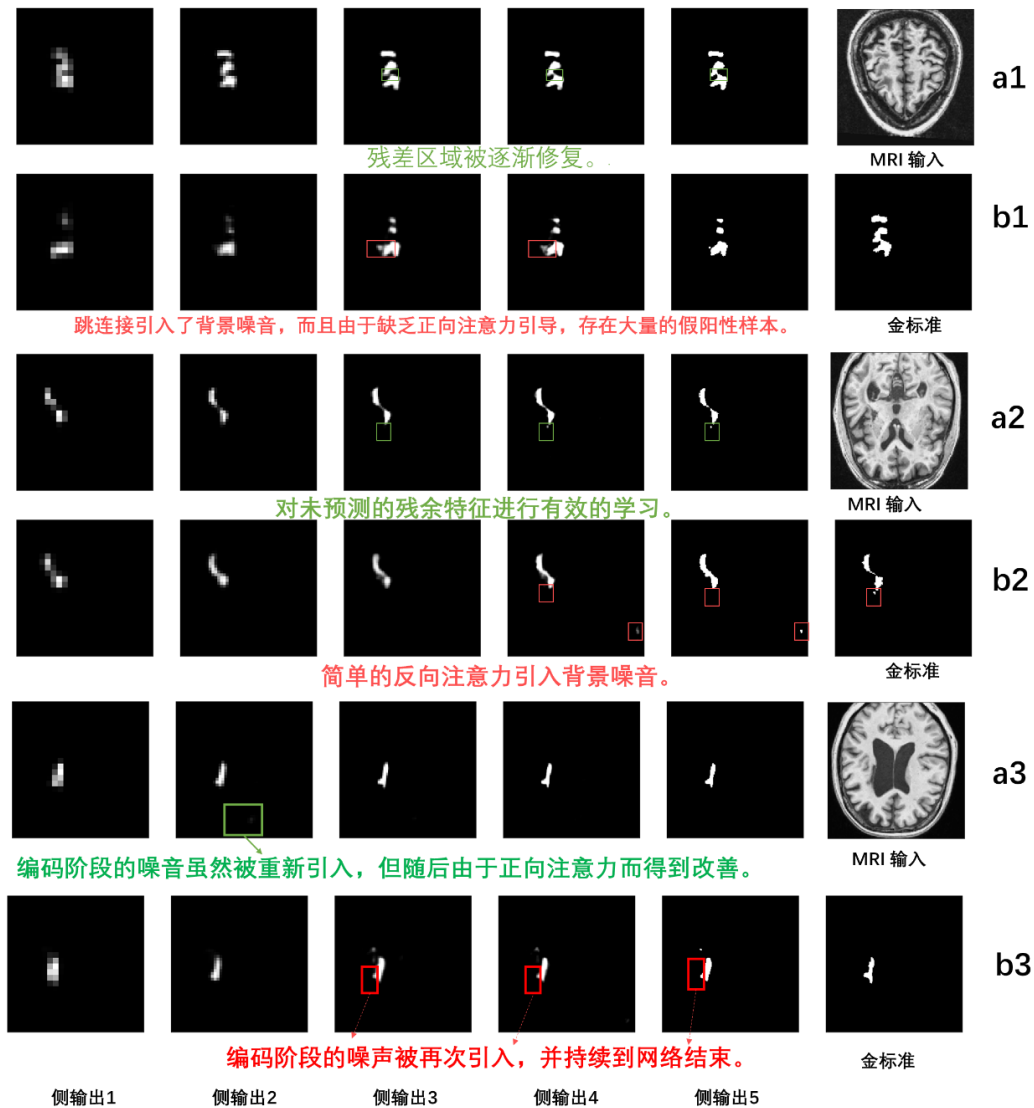


图 4.9 不同残差学习的侧输出的对比

Figure 4.9 Comparison of the side outputs of different residual learning

## 4.6 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，我们更侧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，脑部血管会越来越脆弱，极易引起脑卒中。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力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  $12 \times 12$ ，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  $2 \times 2$ 。所以我们将整个图像等分成 36 个图像块，但这只局限于  $192 \times 192$  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] Khezarpour 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, 王凤喆, 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 年优秀博士硕士学位论文培育项目），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！