

单位代码：10610

送审编号：2022226040008

# 四川大学

## 硕士研究生学位论文

(学术学位)

题目：

基于小样本学习的

医学图像分割方法研究

学位类别：

工学硕士

学科专业：

软件工程

学位论文完成时间

二〇二五年四月



## 摘要

随着医学成像技术与人工智能的发展，基于深度学习的医学图像分割技术在临床诊断与辅助治疗中发挥着重要的作用。然而，医学图像通常面临标注成本高昂、数据规模有限以及器官结构多样化等挑战，导致传统分割模型在小样本场景下难以取得理想表现。围绕“小样本医学图像分割”这一核心议题，本文主要探究如何增强对解剖边界与关键语义的关注，以在少量标注资源下提升分割精度与鲁棒性。

本文以小样本通用图像分割为基础，研究多尺度表征与边缘信息对模型泛化能力的影响并结合医学图像的特殊性，提出一套高效且可扩展的小样本医学图像分割方法，并与分割大模型进行对比分析。主要贡献如下：

(1) 针对小样本通用图像分割中的多尺度目标识别困难与边缘信息易丢失的问题，本文提出了基于多尺度记忆协调与边缘上下文聚合的小样本通用图像分割方法 (Multi-Scale Edge Context Aggregation Network, MS-ECANet)。多尺度记忆协调模块通过融合当前任务原型与历史记忆原型，缓解了单一支持样本信息不足的局限；边缘上下文聚合模块则结合 Sobel 边缘算子与通道注意力机制，显式提取并强化边缘特征。实验结果表明，MS-ECANet 在小样本条件下表现出较高的精度和鲁棒性；在医学数据上的初步验证表明，MS-ECANet 具有良好的边缘增强能力，为后续医学图像分割的深入研究提供了有力支撑。

(2) 针对医学图像中器官边界模糊、对比度低等问题，本文提出了基于双提示协同原型与 Mamba 交互的小样本医学图像分割方法 (Dual-Prompt Collaborative-Prototype Mamba Network, DPCP-MNet)。在特征采集与处理阶段，模型基于视觉提示策略，引入边缘感知几何提示与中心化语义提示，分别提取几何轮廓与核心语义信息，进而生成融合结构与语义的协同原型，为后续匹配提供了高质量的原型；在全局特征交互阶段，引入基于状态空间建模的协同原型 Mamba 交互模块，以线性复杂度对全局特征建模，使协同原型能够灵活适配查询样本，从而兼顾推理效率与跨类别泛化能力；最后，通过多层次融合解码器逐级复原高分辨率特征，实现精确的边界还原。实验结果表明，DPCP-MNet 在 Abdominal-CT、Abdominal-MRI、Cardiac-MRI 等多组医学数据上均获得了性能提升，并在少量标注条件下对比分割大模型（如 MedSAM2、SAM2）时展现出更高的准确率和鲁棒性。

综上，本文提出的 MS-ECANet 与 DPCP-MNet 分别面向通用图像与医学图像两大场景，在小样本分割任务中有效提升了模型的精度与鲁棒性。研究结果不仅展现了小样本分割技术在医疗影像分析中的应用潜力，也为后续更复杂的病灶检测与临床诊断提供了新思路。

**关键词：**小样本学习；医学图像分割；记忆原型；视觉提示



## Abstract

With the advancement of medical imaging technologies and artificial intelligence, deep learning-based medical image segmentation has played a crucial role in clinical diagnosis and treatment assistance. However, medical images typically face challenges such as high annotation costs, limited data scale, and diverse organ structures, making it difficult for traditional segmentation models to achieve ideal performance in few-shot scenarios. Centering on the core issue of "few-shot medical image segmentation," this paper primarily explores how to enhance attention to anatomical boundaries and key semantics to improve segmentation accuracy and robustness with limited annotation resources.

Based on few-shot general image segmentation, this paper investigates the impact of multi-scale representation and edge information on model generalization ability. Combining the specificity of medical images, we propose an efficient and extensible few-shot medical image segmentation method, and conduct comparative analysis with segmentation foundation models. The main contributions are as follows:

(1) Addressing the challenges of multi-scale target recognition and edge information loss in few-shot general image segmentation, we propose the Multi-Scale Edge Context Aggregation Network (MS-ECANet). The multi-scale memory coordination module mitigates the limitations of insufficient information from single support samples by fusing current task prototypes with historical memory prototypes. The edge context aggregation module explicitly extracts and enhances edge features by combining Sobel edge operators with channel attention mechanisms. Experimental results demonstrate that MS-ECANet exhibits high accuracy and robustness under few-shot conditions. Preliminary validation on medical data indicates that MS-ECANet possesses excellent edge enhancement capabilities, providing strong support for further research in medical image segmentation.

(2) Targeting issues such as blurred organ boundaries and low contrast in medical images, we propose the Dual-Prompt Collaborative-Prototype Mamba Network (DPCP-MNet) for few-shot medical image segmentation. During feature acquisition and processing, the model introduces edge-aware geometric prompts and centralized semantic prompts based on visual prompting strategies, extracting geometric contours and core semantic information respectively to generate collaborative prototypes that integrate structure and semantics, providing high-quality prototypes for subsequent matching. In the global feature interaction stage, we introduce a collaborative prototype Mamba interaction module based on state space modeling, modeling global features with linear complexity, enabling collaborative prototypes to flexibly adapt to query samples, thereby balancing inference efficiency and cross-category generalization abil-

ity. Finally, a multi-level fusion decoder gradually restores high-resolution features, achieving precise boundary reconstruction. Experimental results show that DPCP-MNet achieves performance improvements on multiple medical datasets including Abdominal-CT, Abdominal-MRI, and Cardiac-MRI, and demonstrates higher accuracy and robustness compared to segmentation foundation models (such as MedSAM2, SAM2) under limited annotation conditions.

In conclusion, the MS-ECANet and DPCP-MNet proposed in this paper target general and medical image scenarios respectively, effectively improving model accuracy and robustness in few-shot segmentation tasks. The research findings not only demonstrate the application potential of few-shot segmentation technology in medical image analysis but also provide new insights for subsequent more complex lesion detection and clinical diagnosis.

**Keywords:** few shot learning; medical image segmentation; memory prototype; visual prompts;

## 目 录

摘 要.....	I
Abstract .....	III
<b>第 1 章 绪论.....</b>	<b>1</b>
1.1 研究背景与意义.....	1
1.2 国内外研究现状.....	1
1.2.1 深度学习图像分割.....	2
1.2.2 小样本通用图像分割现状.....	3
1.2.3 小样本医学图像分割现状.....	6
1.2.4 分割大模型医学应用现状.....	7
1.3 本文主要研究内容 .....	8
1.4 本文结构安排.....	9
<b>第 2 章 相关理论和技术 .....</b>	<b>11</b>
2.1 图像分割经典模型 .....	11
2.2 小样本图像分割.....	13
2.2.1 小样本通用图像分割 .....	13
2.2.2 小样本医学图像分割 .....	15
2.3 其他相关技术.....	16
2.3.1 视觉提示.....	16
2.3.2 Mamba.....	17
2.4 分割大模型.....	18
2.4.1 SAM 系列模型的结构与特点 .....	18
2.4.2 MedSAM 医学分割大模型.....	19
2.5 图像分割的评价指标 .....	21
2.6 本章小节 .....	21
<b>第 3 章 基于多尺度记忆协调与边缘上下文聚合的小样本通用图像分割方法 .....</b>	<b>23</b>
3.1 引言 .....	23
3.2 多尺度记忆协调与边缘上下文聚合网络的整体框架 .....	24
3.2.1 模型整体框架 .....	24
3.2.2 空洞空间金字塔池化模块.....	25

3.2.3 多尺度记忆协调模块 .....	26
3.2.4 边缘上下文聚合模块 .....	27
3.2.5 原型匹配.....	29
3.2.6 解码器 .....	29
3.2.7 MS-ECANet 损失函数.....	31
3.3 相关实验及结果.....	32
3.3.1 数据集 .....	32
3.3.2 实验配置.....	32
3.3.3 对比实验与分析.....	33
3.3.4 消融实验.....	34
3.3.5 在医学图像上的适应性分析 .....	37
3.4 本章小结 .....	40
<b>第 4 章 基于双提示协同原型与 Mamba 交互的小样本医学图像分割方法 .....</b>	<b>41</b>
4.1 引言 .....	41
4.2 双提示协同原型与 Mamba 交互网络的整体框架 .....	42
4.2.1 模型整体框架 .....	42
4.2.2 边缘感知提示生成器 .....	42
4.2.3 中心化语义提示生成器 .....	44
4.2.4 协同原型生成模块.....	45
4.2.5 协同原型 Mamba 交互模块 .....	46
4.2.6 层次化融合解码器.....	48
4.2.7 DPCP-MNet 损失函数.....	49
4.3 相关实验及结果.....	50
4.3.1 数据集 .....	50
4.3.2 实验设置.....	51
4.3.3 对比实验与分析.....	51
4.3.4 消融实验.....	54
4.3.5 方法与分割大模型的性能对比 .....	57
4.4 本章小结 .....	60
<b>第 5 章 总结与展望.....</b>	<b>61</b>
5.1 研究成果总结 .....	61
5.2 未来研究方向 .....	61
<b>参考文献 .....</b>	<b>63</b>

攻读学位期间取得的研究成果..... 69



## 插图目录

图 1.1	技术路线图	8
图 2.1	FCN 网络架构图 <sup>[5]</sup>	11
图 2.2	U-Net 网络架构图 <sup>[6]</sup>	12
图 2.3	PANet 网络架构图 <sup>[14]</sup>	13
图 2.4	MENUA 网络架构图 <sup>[66]</sup>	14
图 2.5	PGP-SAM 网络架构图 <sup>[66]</sup>	15
图 2.6	Bro 网络架构图 <sup>[66]</sup>	16
图 2.7	SAM 网络架构图 <sup>[59]</sup>	19
图 2.8	MedSAM 网络结构示意图 <sup>[64]</sup>	20
图 3.1	模型整体设计思路图	23
图 3.2	MS-ECANet 整体框架图	24
图 3.3	多尺度记忆协调模块结构图	27
图 3.4	边缘上下文聚合模块结构图	28
图 3.5	解码器结构图	30
图 3.6	PASCAL-5 <sup>i</sup> 数据集的模型分割结果	33
图 3.7	t-SNE 特征分布可视化	36
图 3.8	ECAM 模块医学图像可视化	39
图 4.1	模型整体设计思路图	41
图 4.2	DPCP-MNet 整体框架图	42
图 4.3	EA-Prompt 模块结构图	43
图 4.4	CS-Prompt 模块结构图	45
图 4.5	协同原型生成模块结构图	46
图 4.6	协同原型 Mamba 交互模块	47
图 4.7	层次化融合解码器结构图	49
图 4.8	腹部 MRI 数据集的模型分割结果	52
图 4.9	腹部 MRI 数据集的双提示策略可视化	55



## 表格目录

表 3.1	不同权重组合对 PASCAL-5 <sup>i</sup> 分割性能的影响 (mIoU, %)	32
表 3.2	PASCAL-5 <sup>i</sup> 数据集的各模型结果对比 (mIoU, %)	33
表 3.3	COCO-20 <sup>i</sup> 数据集的各模型结果对比 (mIoU, %)	34
表 3.4	PASCAL-5 <sup>i</sup> 数据集的各模块消融实验结果 (mIoU, %)	35
表 3.5	不同原型方法结果对比 (mIoU, %)	37
表 3.6	腹部 MRI 数据集 Setting-1 设置下的各模型结果对比 (Dice, %)	38
表 3.7	腹部 MRI 数据集 Setting-1 设置下的各模块消融实验结果 (Dice, %)	39
表 4.1	不同权重组合对腹部 MRI 数据集分割性能的影响 (Dice, %)	50
表 4.2	腹部数据集 Setting-1 设置下各模型结果对比 (Dice, %)	52
表 4.3	心脏 MRI 数据集 Setting-1 设置下各模型分割结果 (Dice, %)	53
表 4.4	腹部数据集 Setting-2 设置下各模型分割结果 (Dice, %)	53
表 4.5	腹部 MRI 数据集的各模块消融实验结果 (Dice, %)	54
表 4.6	层次化融合解码器消融实验结果 (Dice, %)	56
表 4.7	主要实验参数	58
表 4.8	不同样本量下的分割性能 (Dice, % / HD95, mm)	59
表 4.9	模型选择策略	59



# 第 1 章 绪论

## 1.1 研究背景与意义

随着医学影像设备和数字化技术的快速发展，如何从 CT、MRI、超声等多模态医学影像中精准分割出病灶或器官的轮廓，已成为提升医疗效率与质量的热点研究问题。近年来，深度学习模型在计算机视觉领域取得了显著进展，尤其在图像分割任务上表现优异，已超越基于手工特征的传统提取方法，成为医学图像处理的主流技术路线。

与通用图像分割不同，医学图像分割领域普遍面临数据不足的瓶颈，主要原因可归结为以下两方面：（1）医学图像涉及患者隐私，获取与共享渠道受限。（2）专业医生的标注成本高昂，罕见病或特殊病例的标注数据尤为稀缺。受限于“小样本”现状，常规深度学习模型容易发生过拟合，导致泛化能力不足，难以满足多样化的临床需求。因此，如何在有限数据条件下仍能保障医学图像分割的准确性与鲁棒性，成为该领域亟待解决的核心问题。

针对数据规模不足的问题，已有研究<sup>[1][2]</sup>在通用计算机视觉场景中提出了多种小样本分割方法，如元学习、少样本迁移学习及对比学习等，并在若干自然图像分割任务中取得了良好效果。这些方法能够在数据规模受限的情况下挖掘更多潜在信息，从而提升模型的泛化能力。受上述成果启发，相关工作<sup>[3][4]</sup>尝试将通用小样本分割策略迁移至医学图像领域。然而，受限于医学图像成像模式的复杂性以及对分割精度的高要求，现有方法在实际应用中的表现仍有待进一步验证与提升。综上，实现具有临床实用性的小样本医学图像分割仍面临诸多关键技术挑战，有待进一步深入研究与攻关。

基于上述背景，本文将聚焦小样本医学图像分割，探讨在少量标注数据条件下实现可靠的医学图像分割的可行途径。通过系统分析小样本分割的理论基础与医学场景下的特定挑战，本文旨在提出兼具高精度与强鲁棒性的小样本医学图像分割方案，为临床诊断与治疗规划提供有力技术支撑。

## 1.2 国内外研究现状

随着医学影像分析技术的不断发展，医学图像分割作为其中的关键任务之一，得到了国内外研究学者的广泛关注。特别是在医学图像领域，由于标注数据的匮乏和标注成本的高昂，如何在小样本条件下实现高效的图像分割已成为重要的研究方向。小样本学习（Few-Shot Learning）技术，凭借其在有限标注数据情况下的出色表现，为解决这一难题提供了新的思路。

本节分为三个主要部分：首先回顾国内外深度学习图像分割的研究现状；然后分别探讨小样本通用图像分割与小样本医学图像分割的技术发展，最后关注了 SAM 分割大模型在医学分割领域的应用现状。

### 1.2.1 深度学习图像分割

传统的图像分割方法通常依赖于手工设计的特征，如颜色、纹理、边缘等低级特征，并结合一些阈值或规则来实现图像的区域划分。虽然这些方法在某些简单场景下能够取得不错的效果，但它们在处理复杂背景、遮挡或多目标分割时往往表现不佳。

随着卷积神经网络（CNN）和深度学习方法的兴起，传统的图像分割方法逐渐被基于数据驱动的深度学习方法所替代。特别是全卷积网络（FCN）和 U-Net 等模型的提出，标志着语义分割技术的一次重大突破。与传统方法相比，深度学习模型通过自动学习图像特征，显著提高了分割的精度和鲁棒性。

2015 年，Long 等人<sup>[5]</sup>提出了全卷积网络（Fully Convolutional Network, FCN），它通过用卷积层替代传统的全连接层，使得网络可以接受任意大小的输入图像，并输出与输入图像大小相同的分割结果。FCN 的引入标志着深度学习在图像分割领域的重要突破。其采用了多层卷积和反卷积（上采样）过程，并通过跳接连接结合低层特征和高层语义特征，提升了分割精度。Ronneberger 等人<sup>[6]</sup>提出了 U-Net，这是一种基于 FCN 的架构，专门为生物学图像分割设计。U-Net 通过对称的编码器-解码器结构以及跳接机制，实现了高精度的图像分割。与 FCN 相比，U-Net 在网络结构中增加了更多的细节特征，使其在医学图像分割中取得了极好的效果，尤其在小样本数据集上具有较强的优势。

2017 年，He 等人<sup>[7]</sup>提出了 Mask R-CNN，该模型是基于 Faster R-CNN<sup>[8]</sup> 框架进行改进的。Mask R-CNN 不仅能够进行目标检测，还能够生成目标的像素级分割掩膜。在原有的 Faster R-CNN 基础上，Mask R-CNN 增加了一个分支，负责生成每个目标的二进制掩膜。通过这种方式，Mask R-CNN 在实例分割任务中表现出了强大的能力，尤其在复杂场景中能够有效地分割重叠或紧密相邻的目标。

2019 年，Liu 等人<sup>[9]</sup>提出了 Swin Transformer，这是一种基于 Transformer 的图像分割模型。与传统的卷积神经网络不同，Swin Transformer 采用了基于自注意力机制的局部窗口操作，并通过分层的结构逐步构建全局上下文信息。该模型能够更好地捕捉图像的全局依赖，提升了分割任务的效果，尤其是在处理复杂的医学图像时，具有较强的表现力。

2020 年，Yuan 等人<sup>[10]</sup>提出了 Vision Transformer (ViT)，该模型将 Transformer 架构应用于计算机视觉任务，并通过将图像划分为固定大小的块，将每个图像块作为序列输入模型。这种方法使得 Transformer 能够有效捕捉全局信息，极大提高了图像分割的精度。虽然 ViT 在大规模数据集上表现优异，但在小样本学习任务中，由于其需要大量的标注数据来进行训练，因此面临一定的挑战。

此外，2020 年以后，Swin Transformer<sup>[9]</sup> 和 Vision Transformer<sup>[10]</sup> 等 Transformer 架构在图像分割任务中逐渐展现出强大的能力。Transformer 模型由于其强大的全局建模能力，在医学图像分割中表现出色，尤其是在图像细节和全局语义信息的理解方面，能够超越传统的卷积神经网络。然而，Transformer 架构在计算资源和数据量上的要求较

高，这对于小样本医学图像分割任务仍然是一个需要克服的挑战。

2021年，Xie 等人<sup>[11]</sup>提出了 SegFormer，该模型通过分层 Transformer 结构来高效编码多尺度特征，并采用轻量级 MLP 对多尺度特征进行聚合解码，从而生成高质量的分割掩码。SegFormer 在显著提升推理速度的同时，也保持了较高的精度。

2021年，Cheng 等人<sup>[12]</sup>提出了 Mask2Former，统一了语义分割、实例分割和全景分割三大任务。其核心在于掩码注意力机制与 Transformer 解码器的结合，实现了多任务分割范式的统一，成为通用分割架构的里程碑。

2023年，Wang 等人<sup>[13]</sup>提出了 SegGPT，将大规模预训练的思想引入分割领域。该模型不仅能够处理少样本甚至零样本条件下多种分割任务，而且展示出了惊人的泛化能力，为未来的通用分割研究提供了新的方向。

这些深度学习模型的提出和发展，不仅推动了语义分割技术的进步，也为医学图像分析提供了强有力的工具。随着大规模预训练模型（如 Vision Transformer、Swin Transformer 等）的不断发展，它们在医学图像分割中展现出了巨大的潜力，尤其是在处理复杂、异质的医学图像数据时，能够提取更为丰富的特征信息，为分割任务提供了更高的精度和鲁棒性。

### 1.2.2 小样本通用图像分割现状

在许多对标注数据依赖度较高且获取成本昂贵的应用场景（如医学影像分析、遥感图像处理 and 自动驾驶等）中，小样本分割技术展现出较大的应用潜力。借助深度学习强大的特征提取能力，小样本分割方法可在仅有极少量标注样本的条件下，依然学得有效的分割策略，因而成为应对数据稀缺问题的可行方案，并在实际任务中获得了广泛关注。总体而言，小样本分割方法具备以下主要优势：（1）数据需求相对较低。通过对少量样本进行更有效的特征学习，小样本分割能够在有限标注的前提下挖掘更多信息，从而提高模型的利用效率。（2）任务适应性较强。常见的小样本分割常结合元学习、迁移学习等思路，使得模型可以在任务或领域切换时进行快速微调，从而更好地适配新的数据分布。（3）有效利用标注信息。小样本分割通过提炼少量样本中的语义特征，实现对目标区域的精细刻画，在增强模型辨别力的同时，进一步提升了分割的准确度。

伴随小样本学习技术的持续发展，学者们针对数据稀缺、标注困难等问题提出了多种小样本图像分割方法<sup>[14][15][16]</sup>。其中，基于生成模型（如 GAN、VAE）的方法虽然可以为小样本分割提供额外的伪标记样本，但也带来了更高的训练难度与推理成本；相比之下，判别式方法往往通过直接优化模型的判别能力，在给定的少量标注数据上更为高效。基于此思路的小样本分割技术，已经在医学图像、遥感影像以及自动驾驶感知等领域取得了良好的应用成效，为传统分割算法在数据稀缺与标注复杂度较高的场景下提供了新的解决思路。

### 1.2.2.1 基于度量学习的方法

度量学习在处理少镜头 ISS 中起着至关重要的作用，其中基于原型网络<sup>[17]</sup>的方法<sup>[14-16]</sup>，处于主导位置。与传统的基于学习的方法<sup>[18-20]</sup>不同，其中类的学习原型是最优原型的近似估计，这些少镜头方法<sup>[21-24]</sup>的目标是获得一个特定于类的原型，这可能不是最优原型的近似值，只要它可以提供对象的信息，并为具有与原型具有相同语义类的查询特征获得更高的相似度分数。然而，仅通过向量原型来描述一个类别是不够的，这激发了一些方法<sup>[25-27]</sup>，旨在为每个类生成多个原型。此外，直接执行支持和查询特征之间的元素级密集匹配<sup>[28-30]</sup>也是突破这一障碍的可行方法。相关方法描述如下。在早期的研究中，尝试用单个类别描述符进行特征匹配。

原型网络首次在<sup>[14]</sup>中被用于通过少量标注样本对未见过的对象进行分割。具体来说，设计了一个双分支网络来为查询图像估计分割图。第一分支以支持集为输入，输出一个全局类别描述符，而第二分支利用生成的原型作为指导来调整查询集的分割结果。PANet 的工作启发了后续研究<sup>[15]</sup>，其直接基于查询特征和类别特定原型之间的余弦距离完成了二元分割。为了充分利用支持信息，查询样本及其预测的掩码被进一步视为新的支持集，用来指导原始支持样本的分割。与上述<sup>[14]</sup>中应用固定距离函数的方法不同，其他方法<sup>[21-24]</sup>利用一个可学习的神经网络预测分割图，可被视为一种可学习的距离度量，用于隐式测量支持和查询特征之间的相似性。这些方法融合了目标类别的支持提示与查询特征，并解码融合特征以输出最终的分割结果。一种常见的融合策略是将查询特征与平铺的原型<sup>[16][21]</sup>或支持特征图<sup>[22]</sup>沿通道维度拼接，结合多尺度特征<sup>[23]</sup>和多类标签信息<sup>[24]</sup>以增强查询样本的表示能力。除了上述提到的直接沿通道维度拼接，其他方法还包括元素级相加<sup>[31]</sup>、通过注意力图重新加权<sup>[32]</sup>、相似性指导<sup>[33-34]</sup>等，都是支持和查询特征融合的可行方式。

尽管上述方法对小样本图像分割做出了不可否认的贡献，但其中一些方法<sup>[21-23]</sup>通过掩码平均池化操作为每个语义类别生成整体描述符，这带来了以下问题。首先，标注类别样本的不足导致原型学习器无法输出鲁棒的类别表示。其次，由于支持样本和查询样本之间的外观变化，仅通过全局特征向量难以捕获丰富的细粒度语义信息。

为了解决这一困境，一些后续研究尝试为每个语义类别生成多个原型<sup>[25-26,35]</sup>，或对支持和查询图像进行密集匹配<sup>[28-30]</sup>。

从原型生成的角度看，相似性度量在每个生成的原型与查询特征之间进行。一种常见的方法是根据特定机制将对象分解为不同部分，并为每部分生成对应的原型。Li 等人<sup>[25]</sup>通过分解对象语义辅助生成多个原型。Liu 等人<sup>[26]</sup>提出了一种自适应调整原型数量的方法，通过对具有不同空间位置的相似支持特征向量进行分组来生成特定原型。此外，Zhang 等人<sup>[35]</sup>进一步利用未标注样本优化多部分感知原型，以获得更细粒度的特征表示。Wang 等人<sup>[36]</sup>设计了三种从多个角度生成的不同描述符，用于与查询特征进行匹配。Zhang 等人<sup>[37]</sup>不再生成确定性原型，而是通过估计生成原型的分布，来模拟因有限训练图像和对象变化而导致的不确定性，从而提高分割模型的鲁棒性。

从密集匹配的角度看，不同研究引入了不同的机制。例如，Wang 等人<sup>[28]</sup>提出金字塔图网络，以捕捉支持和查询特征在不同尺度上的密集对应关系。Liu 等人<sup>[29]</sup>提出民主注意力网络，关注对象所在的像素区域，在支持和查询图像之间建立鲁棒的对应关系。Zhang 等人<sup>[30]</sup>提出谐波特征激活策略，联合利用独特的支持特征进行像素级语义匹配。Gao 等人<sup>[38]</sup>提出一种新型的交叉注意机制，将支持图像中更多相关的像素级特征聚合到查询图像中。Shi 等人<sup>[39]</sup>构建二分图，并应用图注意力机制及权重调整策略，以促使更多目标对象像素参与查询图像的分割。Zhuge 等人<sup>[40]</sup>探讨前景和背景的密集相关性，从而缓解原型学习和前景特征对密集匹配信息丢失的影响。

### 1.2.2.2 基于参数预测的方法

参数预测 (parameter prediction) 方法通过显式生成或调整分类器权重来实现跨类别迁移，无需迭代反向传播微调，因而在在线阶段具备更低的计算开销，并有效缓解灾难性遗忘。其核心思路在于：当支持-查询 (support-query) 对到来时，利用支持信息 (或支持与查询的组合) 即时“预测”一组适配目标类别的新权重，再以这些权重对查询图像进行像素级分类。现有研究主要可分为以下三类：

(1) 条件分支生成型 Nguyen 等人<sup>[41]</sup>最早提出了双分支网络框架：其中，条件分支接收特定类别的支持图像及其掩码，输出逻辑回归 (LR) 层的权重；分割分支则专注于对查询图像进行高层语义特征提取。通过用条件分支预测的权重替换原 LR 层参数，网络可在一次前向传播中完成对新类别的自适应，从而快速生成像素级语义标签。此举开创性地将 few-shot 识别中的“权重生成”思想引入到像素级分割任务中。

(2) 支持-查询联合生成型鉴于查询图像本身也蕴含判别性上下文，Liu 等人<sup>[42]</sup>将支持与查询特征共同输入权重生成器，以进一步提升对外观多变目标的适配能力。其核心在于在权重预测网络中显式建模“支持-查询相似性”，使得在极少支持样本的条件下依旧能保持稳健分割效果。

(3) 动态增量式权重调整型直接替换分类器参数虽然高效，却可能削弱模型对基类的记忆。为此，Yang 等人<sup>[43]</sup>提出权重累加 (add-on) 策略：先保留原有基类权重，再将预测得到的新类权重按一定系数动态叠加到分类器中，既维持基类决策边界，又为模型注入辨别新类的能力，从而显著缓解灾难性遗忘。在 1-shot 实验中，该方法将 mIoU 从 64.6% 提升至 69.8%，充分证明了增量式权重注入的优势。

### 1.2.2.3 基于微调的方法

基于微调的小样本图像分割方法旨在通过优化算法调整预训练分割网络的参数，从而学习未见类别。Tian 等人<sup>[44]</sup>通过最小化支持样本预测结果与其对应掩码之间的误差，逐步优化分割网络的参数。参数微调有效缓解了离线和在线阶段的类别间差异导致的性能下降。

Boudiaf 等人<sup>[45]</sup>提出了一种嵌入网络与可微分线性分类模型相结合的方法，能够更高效地更新分类模型参数，同时嵌入网络具备良好的类别泛化能力。

不同于上述方法在离线阶段采用 episode 训练策略，Wu 等人<sup>[46]</sup>采用了一种基于标准监督学习的推理策略，在基础类别上训练特征提取器。在推理阶段，该方法通过最小化基于支持图像标签的损失函数以及无标签查询图像的统计特性，对线性分类器进行微调。

此外，该研究不仅关注跨类别适应问题，还进一步探讨了训练数据与推理数据之间由于分布多样性导致的偏移问题，从而提升了方法在实际应用中的适应性。

#### 1.2.2.4 基于记忆的方法

在基于记忆的小样本图像分割方法中，利用外部存储器保留先前的相关信息，以帮助对查询样本进行分割。Xie 等人<sup>[47]</sup>通过存储具有显著视觉差异的不同类别的常见属性和先验信息，实现对未见类别的标签迁移。然而，随着存储器大小的增长，模型的计算负担也相应增加。

Mottaghi 等人<sup>[48]</sup>采用了一种不同的策略，将目标类别的多分辨率特征存储，以减少计算开销。这些存储的特征随后被提取，以获取更多跨分辨率信息，从而生成更准确的分割结果。

### 1.2.3 小样本医学图像分割现状

近年来，小样本医学图像分割（FSMIS）因其在数据稀缺场景中的重要性，吸引了大量研究者的关注。受法律、伦理及隐私保护等因素的限制，医学图像领域难以获取大规模标注数据，这对传统深度学习方法构成了挑战。小样本学习技术能够在有限标注数据的条件下有效实现医学图像分割任务，为医学影像分析提供了新的解决方案。根据其技术实现，FSMIS 方法主要可分为两大类：基于双分支交互结构的方法和基于原型网络结构的方法。

#### 1.2.3.1 基于双分支交互结构的方法

基于双分支交互结构的方法通过设计支持分支和查询分支之间的交互模块，增强支持图像与查询图像之间的信息传递能力，从而更精准地捕捉目标的相关特征。其核心思想是利用支持图像中的特征信息来指导查询图像的分割。Guha Roy 等人<sup>[49]</sup>提出了 SE-Net，这是首个专门针对医学图像的小样本分割模型。该模型在条件分支和分割分支之间引入压缩-激励模块（Squeeze-and-Excitation Module），有效提升了支持图像与查询图像的交互性。Feng 等人<sup>[50]</sup>提出了 MRrNet，该模型结合多分辨率特征与交互机制，在复杂医学场景中显著提升了分割性能，尤其在细粒度目标的分割任务中表现突出。Sun 等人<sup>[51]</sup>研发了 GCN-DE，该模型通过引入全局相关性模块，增强了支持图像与查询图像之间的全局信息交互，从而提升了分割任务的鲁棒性。Wu 等人<sup>[52]</sup>提出了 AAS-DCL，针对支持图像与查询图像的特征对齐问题，该方法利用对比学习机制，在不同原型之间构建对比关系，有效增强了分割性能，尤其在多类别医学图像分割任务中展现了优异的适应能力。

这些基于双分支交互结构的方法在捕捉医学图像中的局部特征和全局语义关系方面表现优异，特别适用于复杂目标结构的分割任务。然而，此类方法通常依赖复杂的交互机制，这可能会增加计算成本，从而限制其在资源受限场景中的应用。

### 1.2.3.2 基于原型网络结构的方法

基于原型网络结构的方法通过学习支持样本的语义原型 (Prototype)，并将其作为查询图像的分割参考。这类方法通过构建和优化原型表示，在小样本医学图像分割任务中展现出良好的效果。Ouyang 等人<sup>[53]</sup>提出了 SSL-ALPNet，该模型结合自监督学习与小样本学习技术，并引入自适应局部原型池化模块 (ALP)，能够提取更具泛化能力的局部特征信息，从而提升模型在不同医学任务中的适应性。Hansen 等人<sup>[54]</sup>研发了 ADNet，该模型从异常检测的角度出发，仅使用前景原型进行分割，有效减少了背景特征对分割性能的干扰，从而显著增强了目标区域的辨识能力。Shen 等人<sup>[55]</sup>设计了 Q-Net，在 ADNet 的基础上进一步提出原型优化模块。该模块能够在测试阶段利用查询图像的信息更新语义类别原型，从而进一步提升分割精度。Ding 等人<sup>[56]</sup>提出了 CRAPNet，该模型采用循环一致性注意机制，保持了查询图像与支持图像像素之间的空间对应关系，从而增强了特征对齐能力。

基于原型网络结构的方法充分利用了支持样本的语义信息，具有结构简单、训练高效的特点，因此适用于多样化的医学场景。然而，其分割性能依赖于原型表示的表达能力，在特征复杂或数据噪声较大的场景下，效果可能受到一定限制。

### 1.2.4 分割大模型医学应用现状

在深度学习领域，Transformer 架构<sup>[9]</sup>最初在自然语言处理 (NLP) 中取得突破性成果，随后逐步被引入视觉任务，诞生了 Vision Transformer (ViT)<sup>[10]</sup>等全新视觉模型。这类 Transformer 视觉模型在捕捉长距离依赖关系方面远优于传统卷积神经网络 (CNN)，尤其在图像分割、目标检测等任务中展现出显著的性能提升，如 DETR<sup>[57]</sup>、SegFormer<sup>[58]</sup>和 Swin Transformer<sup>[9]</sup>系列，推动了视觉分割领域的发展。

此背景下，2023 年 Kirillov 等人<sup>[59]</sup>提出了 Segment Anything Model (SAM)，开创了“可提示的图像分割 (Promptable Image Segmentation)”新范式。SAM 基于超大规模的 SA-1B 数据集 (超过 11 亿分割掩码) 进行训练，展现出卓越的泛化能力和零样本分割能力。通过提示机制 (如点、框、文本)，SAM 实现了在不同领域任务中的灵活适配和快速分割。近期研究也持续推进 SAM 模型的优化，例如，Ke 等人<sup>[60]</sup>通过引入高质量输出标记，针对细粒度掩码训练提升了 SAM 在精细分割任务中的表现；Xiong 等人<sup>[61]</sup>和 Zhang 等人<sup>[62]</sup>通过架构优化与稀疏注意力机制大幅提高了 SAM 的计算效率；此外，多模态扩展策略将 SAM 与自然语言处理任务结合，使其实现了基于语义的动态分割。

在医学图像领域，SAM 模型同样得到了积极探索与广泛应用。Wu 等人<sup>[63]</sup>提出了医学 SAM 适配器 (Medical SAM Adapter)，展示了一种可能的解决方案，即通过适配器

根据参数有效的微调范例对预训练的 SAM 模型进行微调，在医学图像分割上显示出惊人的良好性能。Ma 等人<sup>[64]</sup> 则进一步优化了 SAM 的提示策略, 提出 MedSAM, 展示了 SAM 在医学领域应用的潜力。这些研究初步证实了 MedSAM 模型在医学图像小样本场景中的良好适配潜力。然而，医学影像数据自身存在诸多特殊挑战，如数据稀缺、边界模糊、类别不平衡及模态差异等，使得 MedSAM 等大模型的实际性能和适用范围仍需进一步探索。

因此，在大模型时代下，医学图像分割研究的重要趋势之一是明确传统专用小样本模型与 MedSAM 等通用大模型在医学图像小样本条件下的表现差异和适用条件。基于这一需求，本论文将围绕医学图像的小样本分割任务，展开专用模型与 MedSAM 模型的系统对比研究，以揭示两种技术路线的优势与不足。这不仅有助于深入理解医学图像分割方法的有效性与局限性，还能为医学图像分割技术的发展方向提供理论依据和应用指导，具有重要的学术价值与现实意义。

### 1.3 本文主要研究内容

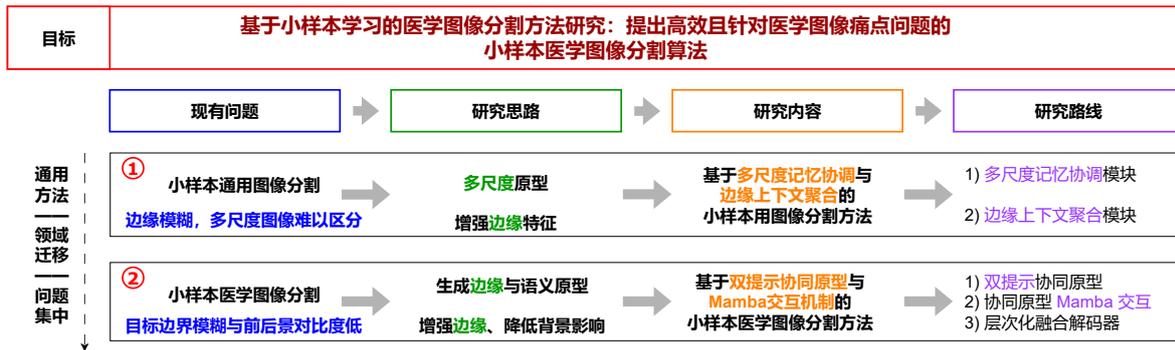


图 1.1 技术路线图

Figure 1.1 The technical roadmap

本文针对小样本图像分割的现实需求，围绕通用场景与医学场景两大方向展开研究，力求在有限标注数据条件下实现精度与鲁棒性的统一。通过设计多种具有针对性的模块与网络结构，实现了对目标外部轮廓与内部语义特征的多角度建模，从而有效提升模型的分割精度与表达能力。具体研究内容如下：

(1) 提出了基于多尺度记忆协调与边缘上下文聚合的小样本通用图像分割方法 (Multi-Scale Edge Context Aggregation Network, MS-ECANet)。该方法通过多尺度记忆协调模块 (Multi-Scale Memory Coordinate, MSMC) 有效存储并整合不同尺度支持样本的原型表征，弥补了单一支持样本导致的表征不足；同时引入边缘上下文聚合模块 (Edge Context Aggregation Module, ECAM)，结合基于 Sobel 的边缘检测和通道注意力机制，强化了对目标边缘的判别能力。该方法在通用任务中展现出较高的鲁棒性，同时

在医学数据集上的初步探索表明，该网络所采用的边缘上下文聚合的思路适用于医疗图像场景，这也为后续医学图像场景的探索打下了基础。

(2) 提出了基于双提示协同原型与 Mamba 交互的小样本医学图像分割方法 (Dual-Prompt Collaborative-Prototype Mamba Network, DPCP-MNet)。该方法通过双提示协同原型模块 (Dual-Prompt Collaborative-Prototype, DPCP)，生成边缘感知提示和中心化语义提示，用于捕获器官的外部轮廓与内部语义信息。在后续原型生成阶段，双提示实现协同融合，有效提升了对病灶或器官的多视角刻画能力；同时在网络结构上引入 Mamba 机制，兼顾了对高分辨率医学图像的全局注意力建模与计算效率。该方法在医学图像分割任务展现出较优的表现。相较于依赖大规模预训练的分割大模型，该方法在极少标注场景中更能保持稳定表现，并为医学影像分析的临床应用提供了可行的技术路径。

综上所述，本文紧扣小样本学习在医学图像分割中的核心需求，先后提出了 MS-ECANet 和 DPCP-MNet 两种创新方法：前者从通用小样本分割入手，通过边缘上下文聚合与多尺度记忆协调等策略，为后续医学研究提供了可迁移的理论与技术基础；后者则重点针对医学图像边界模糊、对比度低等痛点，提出结合双协同原型生成与 Mamba 交互机制的 DPCP-MNet，有效提升了医学小样本分割的精度与效率。通过这一系列研究，本文为小样本医学图像分割提出了系统化的解决方案，并进一步拓展了深度学习在医学辅助诊断和精准治疗中的应用范围。

## 1.4 本文结构安排

第一章绪论。本章主要介绍基于小样本医学图像分割技术的研究背景与意义，阐明小样本学习与医学图像分割在计算机视觉领域中的紧密关联，分析国内外相关研究的发展现状与趋势。随后，明确本文拟解决的关键问题，概述研究内容与技术路线，并简要说明论文的整体结构，帮助读者对全文内容形成概览。

第二章相关理论和技术。本章先回顾了 FCN、U-Net 等经典分割模型对小样本研究的启发；接着分析了 PANet、MENUA 等代表性小样本分割网络在度量学习和原型对齐等方面的技术路径；随后探讨了视觉提示和 Mamba 在提升分割方法能力上的作用；并进一步分析 SAM 系列模型在极少样本场景中的适应性表现；最后总结了 IoU、Dice 等常用评价指标的体系与意义。这些理论梳理为后续实验研究奠定了技术基础。

第三章基于多尺度记忆协调与边缘上下文聚合的小样本通用图像分割方法。本章针对目标尺度多样与边界模糊的问题，提出了一种结合多尺度记忆协调与边缘上下文聚合的分割网络架构。通过在 PASCAL-5<sup>i</sup> 和 COCO-20<sup>i</sup> 数据集上的实验验证，表明该方法在小样本条件下具有较高的精度与稳健性。鉴于医学影像同样面临模糊边界与尺度多变的挑战，本章对医学数据进行了初步测试，结果显示边缘增强技术在医学场景的分割任务中也具备良好的适应性，为后续医学图像分割研究提供了技术支撑。总体而言，实验结果证明了多尺度特征融合与边缘信息强化在小样本分割中的关键作用。

第四章基于双提示协同原型与 Mamba 交互的小样本医学图像分割方法。本章针

对医学影像中常见的边缘模糊、对比度低等难题，提出了一种结合双提示协同原型与 Mamba 交互的分割网络架构。该方法通过融合边缘感知几何提示与中心化语义提示，并结合高效的 Mamba 交互框架与层次化解码器，在提升推理效率的同时，显著优化了分割边界质量。多组医学数据集上的实验结果表明，所提网络具备良好的泛化能力。本章还就极少量标注数据情形进行了对比测试，分析了 DPCP-MNet 与 MedSAM2、SAM2 等分割大模型在小样本场景下的性能差异。这些研究结果不仅为医学分割技术的模型选型提供了重要的实证依据，也进一步丰富了小样本分割技术在医学领域的应用思路。

第五章总结与展望。本章归纳了全文的研究成果，并对未来的发展方向进行了展望。针对小样本图像分割的挑战，本文提出两套方案：通用场景下的 MS-ECANet 网络通过多尺度记忆协调与边缘上下文聚合解决尺度变化与边缘模糊问题；医学场景中的 DPCP-MNet 网络基于双提示协同原型与 Mamba 交互机制应对高分辨率、形态多变及噪声干扰。未来研究可从增强复杂场景鲁棒性、融合小样本与大模型优势以及发展多模态医学图像技术三方面拓展。

## 第2章 相关理论和技术

为系统阐述本文在小样本医学图像分割中的研究基础与思路，本章将从经典图像分割模型、小样本分割方法、其他相关技术、大模型分割网络以及评价指标五个方面进行综述与分析，并介绍了所提到的模型与本文的相关联系。通过以上内容的梳理，本章为后续提出的小样本医学分割模型奠定了必要的理论基础和技术背景，并为实验评估提供了统一的指标体系支撑。

### 2.1 图像分割经典模型

图像分割任务的核心在于对图像进行像素级分类，提取目标区域的精确掩码。在小样本分割任务中，由于训练数据有限，如何设计有效的网络结构来最大化学习信息至关重要。基于全卷积网络 (FCN) 和编码-解码结构 (U-Net) 的方法是图像分割领域最具代表性的经典网络架构，它们也为小样本分割提供了重要的框架基础。由于小样本分割方法往往继承了 FCN 的全卷积架构，本文研究工作与 FCN 和 U-Net 在网络结构上的共性使得它们成为重要的参考模型。下文将重点介绍 FCN 与 U-Net 的核心思想、优劣势，并简要说明它们对本文研究的启示，尤其是在有限数据情况下如何优化分割精度。

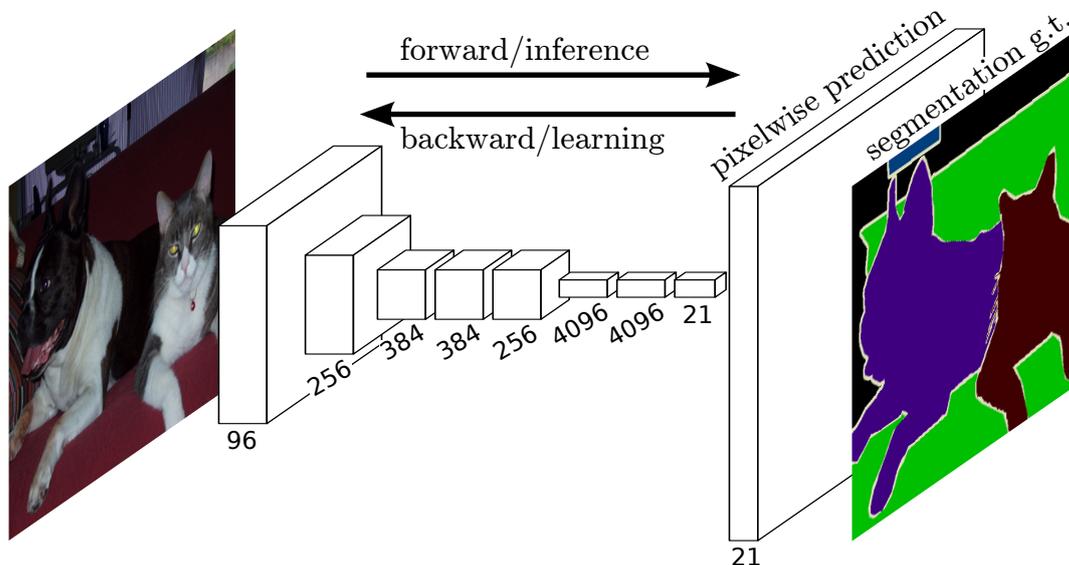


图 2.1 FCN 网络架构图<sup>[5]</sup>

Figure 2.1 Network architecture diagram of FCN

FCN<sup>[5]</sup> (全卷积网络) 首次提出时，主要应用于自然场景图像的语义分割任务。由于其完全由卷积层构成，能够接受任意尺寸输入图像并输出与输入尺寸对应的分割图像，因此广泛用于像素级预测任务中。如图2.1所示，FCN 架构由卷积编码器和逐步上采样解码器组成。编码器部分通常采用经典卷积神经网络 (如 VGG、ResNet)，用于提

取高层次语义特征，而解码器则通过反卷积或插值逐步恢复至原始图像尺寸。FCN 的优势在于其简洁的结构和处理任意大小图像的能力，但其劣势在于过度依赖逐级上采样，导致对图像边缘和细节的恢复能力有限，容易出现边缘模糊。由于小样本分割需要有效利用有限数据，FCN 这一全卷积网络架构对小样本分割模型设计具有重要启示。例如，OSLSM<sup>[65]</sup> (One-Shot Learning for Semantic Segmentation) 就是在 FCN 思路的基础上提出的早期小样本分割网络之一。该方法采用双分支结构：一支处理少量带标注的支持图像，用于学习判别特征；另一支则通过这些特征对查询图像进行像素级预测。OSLSM 为在小样本条件下实现图像分割提供了启发，其框架和思路对本文的小样本分割方法具有直接关联。

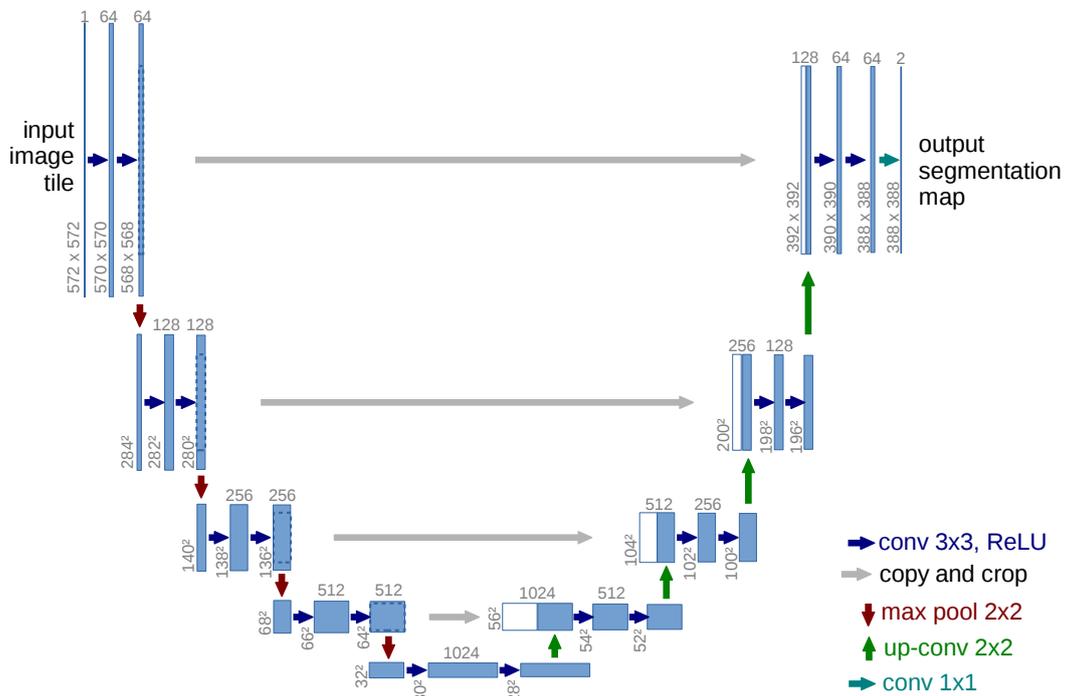


图 2.2 U-Net 网络架构图<sup>[6]</sup>

Figure 2.2 Network architecture diagram of U-Net

U-Net<sup>[6]</sup> 最初应用于医学图像分割，采用对称的编码-解码结构，并在编码器和解码器之间建立跳跃连接 (Skip Connections)。编码器通过卷积和池化操作逐步缩小特征图尺寸，提取不同层次的语义信息；解码器则通过上采样和卷积将特征图恢复至原始图像尺寸。跳跃连接能将浅层的细节特征（如边缘和纹理信息）直接传递给解码器，显著提升边界的刻画与恢复能力。U-Net 的优势在于结构简洁、参数量较少，并且在医学图像小样本分割中表现出色。然而，其解码器部分依赖简单的上采样操作，缺乏对复杂全局上下文的建模能力，当目标在尺度、形状或纹理上差异较大时，分割精度可能下降。受 U-Net 结构启发，本文在解码端对其进行改进，引入更充分的特征融合和精细化策略，以增强对多尺度目标的分割能力并改善目标边缘的预测精度，尤其是在小样本条件下，期望能有效克服传统 U-Net 在有限数据下的精度下降问题。

上述经典模型 (FCN 和 U-Net) 为本文网络设计提供了重要参考, 然而它们在小样本条件下仍面临以下挑战: FCN 过度依赖逐级上采样, 导致边缘模糊和细节恢复能力不足; 而 U-Net 虽能有效利用跳跃连接捕捉边缘细节, 但解码端的上采样操作缺乏全局上下文建模能力, 容易导致分割精度下降。针对这些不足, 本文的核心目标是提出一种新型网络架构, 既能在有限数据下有效融合全局上下文信息, 又能精确恢复细节边界, 从而提高小样本图像分割的精度与鲁棒性。

## 2.2 小样本图像分割

小样本学习 (Few-Shot Learning) 的目标是使模型在极少监督信息下仍具备良好的泛化能力, 模拟人类“少量经验即可举一反三”的学习方式。在图像分割任务中, 由于标注数据的高昂成本, 小样本学习为解决数据稀缺问题提供了有效路径。当前主流做法通过支撑集 (support set) 和查询集 (query set) 构成双分支结构, 借助元学习反复构造小样本任务, 使得网络习得“如何快速学习”的能力, 从而可以迁移到新类别或新场景的分割任务中。围绕这一目标, 研究者提出了多种技术路径, 例如度量学习、记忆增强、提示生成与背景建模等, 它们从不同角度缓解了数据稀缺带来的表征不足问题。

本节将现有的小样本分割研究按应用场景分为两大方向: 小样本通用图像分割与小样本医学图像分割。这两类方法面临的挑战有所不同, 前者主要解决自然场景中多类物体的分割问题, 而后者则专注于解决医学图像中常见的类内差异与边界模糊问题。为了更好地展示这些方法的应用, 我们分别选取了具有代表性的网络进行讨论, 在小样本通用图像分割中, 本节选择了 PANet 与 MENUA 网络进行了探讨, 在小样本医学图像分割中我们选择了 PGP-SAM 与 Bro 网络进行了探讨, 四种方法分别代表了原型对齐、记忆重编码、提示学习和背景融合四条思路, 后文将依次介绍它们的核心机制、优势局限及与本文工作的关联。

### 2.2.1 小样本通用图像分割

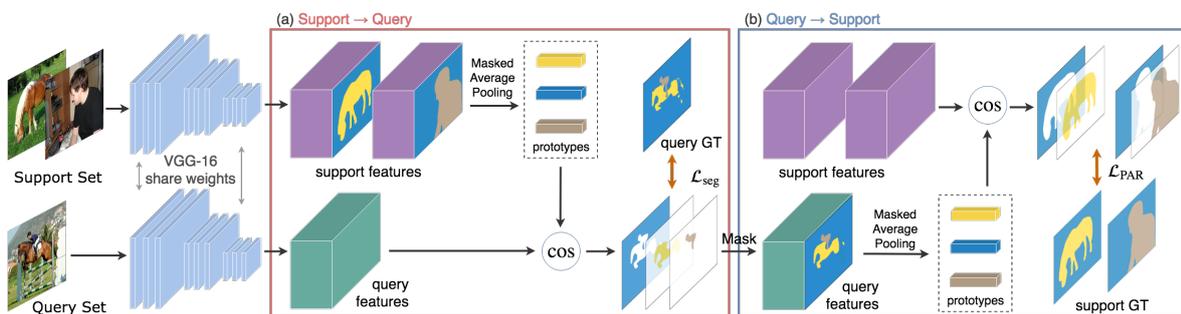


图 2.3 PANet 网络架构图<sup>[14]</sup>

Figure 2.3 Network architecture diagram of PANet

小样本通用图像分割旨在解决在自然场景或多类物体类别下, 依然能通过极少标注样本获得准确的分割掩码。PANet<sup>[14]</sup> (Prototype Alignment Network) 是小样本学习领域

的一项里程碑式工作。该方法在原型网络（Prototypical Network）基础上提出了原型对齐机制，通过将查询图像的特征与支撑集中的类别原型进行对齐，成功增强了小样本场景下像素级分类的精度。这一方法通过强化特征对齐，缓解了数据稀缺带来的表征不足问题。具体而言，PANet 从支撑集中提取每个类别的原型向量，这些原型代表了类别在特征空间中的分布中心；然后将查询图像的特征与原型对齐，通过计算相似度来完成像素级分类。与传统度量学习类似，PANet 也充分利用了度量空间“类内紧密、类间分散”的特性，但它进一步强化了支撑与查询之间的特征对齐过程，从而在样本有限的场景下依然能保持较高的分割精度。PANet 的多尺度特征融合模块也有助于提升对不同大小目标的检测能力。其主要优势在于：结构清晰、易于训练，同时适应多类物体的场景；局限在于：对跨域场景或极端外观变化仍然有一定敏感度。该工作提出的原型网络，被视为小样本学习经久不衰的经典网络模型，在本文中也视为一个基本思路构成。

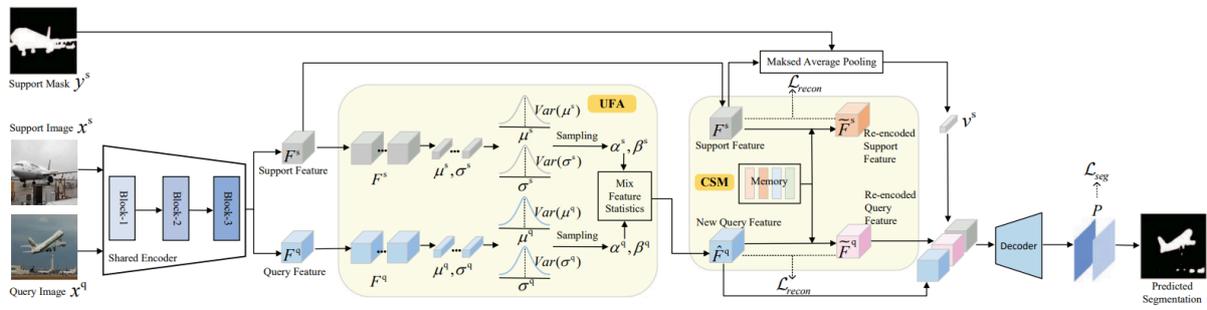


图 2.4 MENUA 网络架构图<sup>[66]</sup>

Figure 2.4 Network architecture diagram of MENUA

如图 2.4 所示，基于不确定性特征增强的记忆引导网络（MENUA）<sup>[66]</sup> 通过引入“类共享记忆”（Class-Shared Memory, CSM）模块与“不确定性特征增强”（Uncertainty-based Feature Augmentation, UFA）模块，来应对新旧类别分布差异与类内差异带来的挑战。其核心思想在于：一方面，记忆模块以可学习向量（memory vectors）的形式，捕捉并存储通用的局部模式或类别无关的“基础元素”，在推理阶段用以重编码查询图像特征，降低基类与新类在特征表示上的差异；另一方面，针对小样本场景中可能出现的外观变形、尺度变化等类内差异，UFA 模块通过不确定性建模来在特征层面进行数据增强，从而提升网络对查询图像的鲁棒性与泛化能力。

但该网络也存在一些局限：（1）记忆向量的设计与数量需要在“充分表达”与“额外开销”之间平衡，若记忆向量过少，难以覆盖复杂场景中的多样局部模式；若过多，则会带来额外存储及检索负担。（2）网络在推理阶段对查询特征进行重编码时，仍需假设训练阶段所学到的“基础元素”可有效转移到目标新类，但若新类的形态与基类差异极端显著，记忆向量的泛化能力可能受到限制。本文提出的多尺度记忆协调模块（MSMC）沿用了记忆向量重编码的基本思想，但进一步从多尺度特征融合角度扩展。

## 2.2.2 小样本医学图像分割

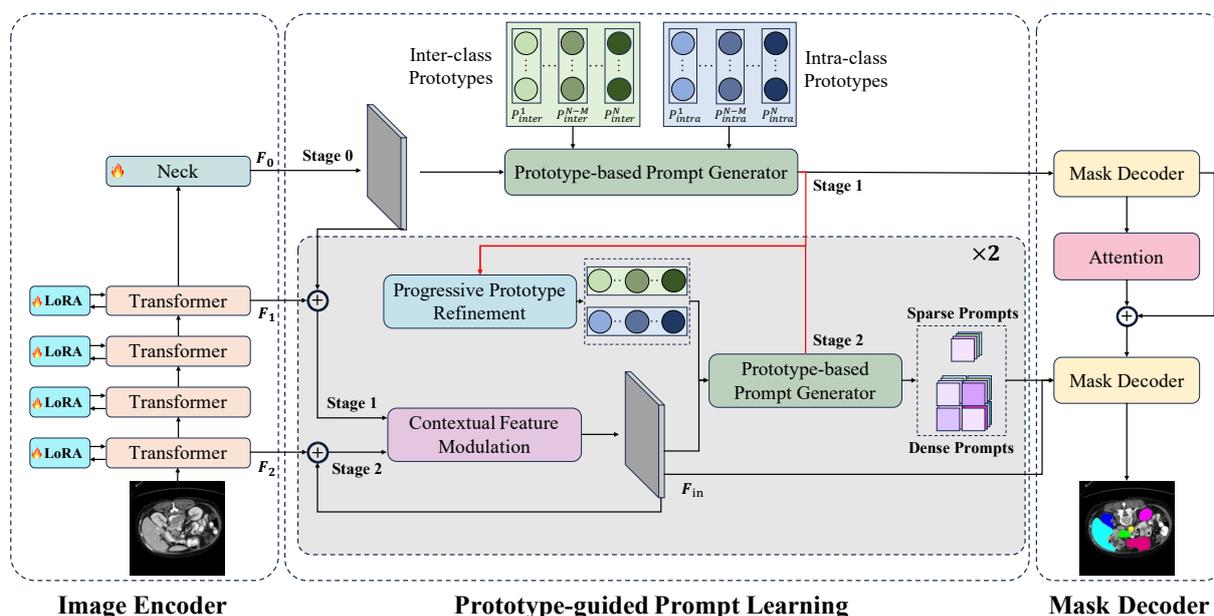
图 2.5 PGP-SAM 网络架构图<sup>[66]</sup>

Figure 2.5 Network architecture diagram of PGP-SAM

PGP-SAM<sup>[67]</sup> 针对小样本医学图像分割任务，对 Segment Anything Model 进行了创新优化。该方法的创新之处在于无需依赖人工标注的提示（如点、框或文字提示），而是通过自我生成类特异性提示来引导模型进行分割。在医学图像中，尤其是对稀有病灶或边界模糊的挑战尤为突出，PGP-SAM 通过上下文调制与渐进式原型优化，能有效提高对复杂病灶的分割鲁棒性。如图 2.5 所示，该网络实现路径分成两步：首先借助 CFM (Contextual Feature Modulation) 在通道与空间维度融入全局上下文，使网络自动突出前景区域、削弱冗余纹理；随后通过 PPR (Progressive Prototype Refinement) 把“类内原型”和“类间原型”放进双路径交叉注意力框架中，让提示向量在多轮交互中逐渐细化，最终驱动 SAM 的掩码解码器输出高精度结果。该方案的优点为：(1) 去除了对精细提示的人工依赖，简化了临床落地流程；(2) 交叉注意力在优化原型的同时也兼顾了跨类差异，提高了对罕见病灶的分割鲁棒性。不过，它仍需大量特征交互与注意力运算，计算成本不低；且当类别之间的语义距离过大、样本极度稀缺时，类间原型的参考价值会被削弱。本文借鉴了“提示可反向促进原型质量”的思路，将提示机制与前景特征聚焦结合，用于进一步提高原型的判别力以达到解决边界模糊与纹理相似的问题。

Bro<sup>[68]</sup> (pluggable Background-fused prototype) 使用了另一种思路，利用背景信息改进小样本医学分割。原文作者观察到医学图像前后景纹理极为相似，单纯依赖前景原型容易混淆，于是提出 FeaC (Feature Similarity Calibration) 和 HiCA (Hierarchical Channel-Adversarial Attention) 两个模块。如图 2.6 所示，FeaC 通过支持-查询交叉注意力把支持图中的背景噪声先行抑制，再提升前景清晰度；HiCA 把背景特征按通道分组，从粗到

细地对各组信息做对抗融合，最终得到一套更具区分度的“背景原型”，与前景原型共同参与像素分类。优势在于能显式建模背景，尤其适合边界模糊、纹理相近的病灶；局限则体现在通道分组数量与对抗正则参数需要手动调节，跨数据集泛化时需重新搜索最优设置。本文在 Bro 的启发下，在保证前景原型准确性的前提下加入了对背景的动态建模，以增强前后景分割边界的清晰度，特别针对纹理相似度高的医学场景。

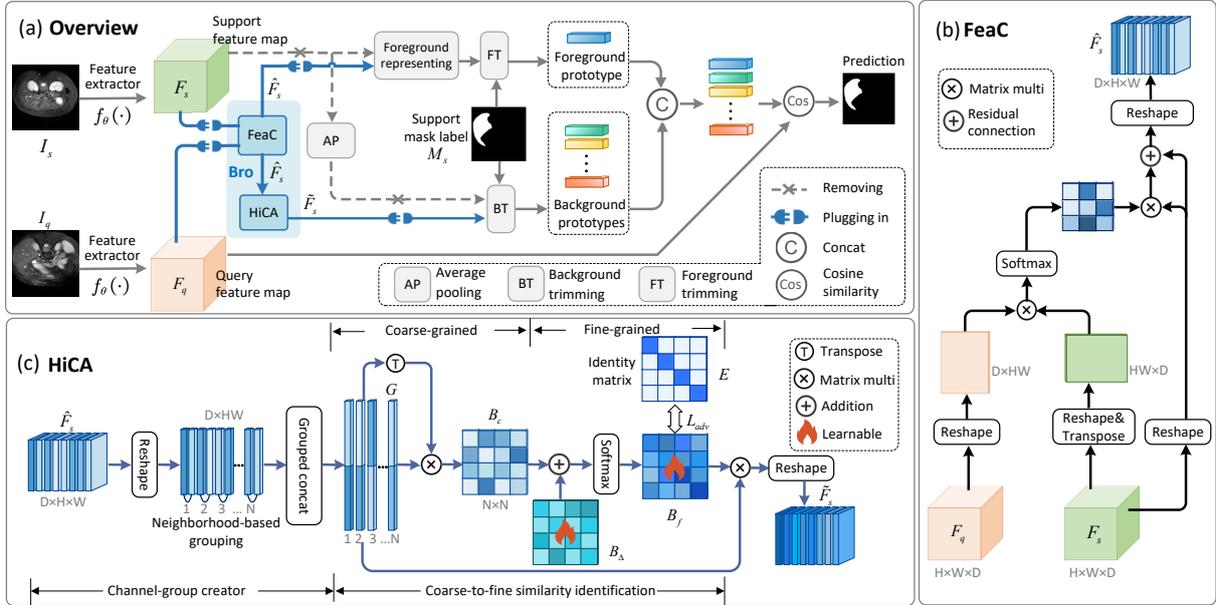


图 2.6 Bro 网络架构图<sup>[66]</sup>

Figure 2.6 Network architecture diagram of Bro

## 2.3 其他相关技术

本文所提出的核心方法与视觉提示机制、线性时间序列模型等技术思路有一定衔接。为了让读者更好地理解相关概念与应用场景，本节将先介绍两类与本文后续章节紧密相关的技术：视觉提示（Visual Prompt）与 Mamba。其中，视觉提示主要探讨如何通过插入少量可学习向量来高效适配预训练模型，而 Mamba 则聚焦在长序列任务的线性复杂度建模。

### 2.3.1 视觉提示

在计算机视觉任务中，“视觉提示”（Visual Prompt）是一类通过在模型输入或网络结构中插入少量可训练向量（提示向量），从而高效适配预训练模型的方法。相比于传统的全模型微调方式，视觉提示不仅能显著减少需要微调的参数规模，同时在诸多视觉识别场景下能取得与全模型微调相当甚至更优的精度。以下介绍两种具有代表性的视觉提示方法。

VPT<sup>[69]</sup> 提出在 Transformer 模型（如 Vision Transformer, ViT）的输入端引入一组任务相关且可学习的提示向量，并在保持预训练主干网络冻结的情况下，仅对这少量提示

向量和输出头 (head) 进行训练。相比传统微调时数以千万计的可训练参数, VPT 所需的参数量通常小于 1%, 大幅降低了模型微调时的存储与计算成本。更为重要的是, VPT 在图像分类等多项视觉任务上表现出与全模型微调相近甚至超越的性能, 特别是在数据量有限的情况下, VPT 的优势更加明显。

在具体实现上, VPT 的提示向量会被插入到 Transformer 每一层或仅第一层的输入序列中 (分别称为“深层 VPT”和“浅层 VPT”)。深层 VPT 在所有层都插入提示, 能充分为下游任务贡献可学习的表示能力; 浅层 VPT 仅在最初的输入层插入提示, 虽然略逊于深层 VPT, 但可进一步减少可训练参数量, 适用于在多任务或超大规模模型场景下的部署。

边缘提示 (Edge Prompt) 与 VPT 将提示聚焦于语义或注意力特征不同, SPPEL<sup>[70]</sup> (Self-Prompting Perceptual Edge Learning) 则将图像的边缘特征作为可学习提示, 用于提升密集预测任务 (如图像分割、目标检测) 的效果。其核心思想是在网络的特征图中, 通过可训练的“传统算子” (如 Kirsch、HOG) 自动萃取几何与语义相融合的边缘特征, 并在模型的多尺度特征层中插入这些边缘提示, 从而让下游解码器或检测头更好地感知目标轮廓、纹理等关键信息。

具体而言, SPPEL 首先通过手工算子的可训练版本提取出初步的“感知边缘”特征; 然后通过一种对抗结构来度量不同边缘算子间的相似性, 并动态融合其输出, 以得到更精细的边缘提示。最后将该提示与原图像特征结合, 引导解码器在器官/病灶分割、目标检测等多种视觉任务中获取更准确的空间定位和语义划分能力。这种“边缘提示”的方式, 在医学图像细节要求高、难以获取足够训练数据的情况下, 优势尤为显著。

相比传统全模型微调, 视觉提示在参数效率与特征适配方面具有突出优势; 在少量标注样本或高精度分割等场景中, 基于视觉提示的思路值得关注。本文在第4章引入了类似的提示机制以强化对边界细节的感知。

### 2.3.2 Mamba

Mamba<sup>[71]</sup> 是一类基于状态空间模型 (SSM) 的线性时间序列模型, 旨在在大规模或超长序列任务中兼顾效率与高性能。传统自注意力模型 (Transformer) 在处理长序列时通常需要二次方级别的计算与显存开销, 而 Mamba 通过引入可选择性状态空间 (Selective State Space), 在保持线性时间复杂度的同时, 仍能有效捕捉复杂序列中的上下文信息。

在模型结构方面, Mamba 借鉴了线性递归和卷积的核心思想, 将其融合进统一的状态空间框架之中。与通用的 SSM 不同的是, Mamba 通过“可选择性机制” (Selection Mechanism) 使状态更新依赖于输入, 从而可以在长序列中“选择性”保留关键信息、过滤噪声。这种灵活性尤其适合在存在不规则间隔或冗余片段的时序数据里进行信息提取, 比如大规模语言建模或基因组序列分析。

在硬件加速层面, Mamba 摒弃了依赖全局卷积的“伪线性”计算路径, 转而采用以

并行扫描为核心的复合算子融合，在 GPU 上显著降低了读写开销。这使得 Mamba 在极长序列的推理阶段具备近似线性或常数级的时间复杂度，与自注意力相比具有明显优势。在大模型规模下，Mamba 采用简化统一的网络结构，将状态空间单元与 MLP 层共同堆叠，从而在多种场景下均能保持较好扩展性和性能。

基于上述特性，Mamba 在多个领域的长序列任务上都表现出极大潜力。例如在语言模型中，Mamba 可以在百万级别长度的上下文窗口内保持高效推理与良好泛化；在语音与基因组等连续或离散信号场景下，Mamba 同样具备优秀的表示能力和长距离建模优势。尽管如此，Mamba 仍面临少量挑战：一方面，对跨模态与多任务的适配尚在探索阶段，另一方面，如何在复杂数据分布下确保模型的稳定训练与精度仍是持续研究的方向。

借助对可选择性状态空间的灵活建模，Mamba 在长序列任务中展现出强大的泛化能力与线性复杂度。但其在跨模态和多任务场景下的表现仍有待进一步研究。本文在第4章也将对 Mamba 的长序列表示能力进行探讨，期望在极少样本和不规则结构数据下获得高效的上下文建模性能。

## 2.4 分割大模型

随着 Transformer 在视觉领域的快速崛起，大规模预训练模型在图像分割任务中展现出了卓越的性能与强大的泛化能力。近年来，作为通用分割大模型的代表，Segment Anything Model (SAM) 系列迅速引起了学术界和工业界的广泛关注。该系列模型在自然场景、交互式提示等多个方面取得了显著进展，并逐步扩展至医学影像分析领域，衍生出针对医学分割优化的 MedSAM 等版本。为评估小样本分割方法在极少数据场景下的实际表现，并明确小样本分割模型相较于大模型的优势区域，本文在第 4 章的实验中，将 SAM2 和 MedSAM2 作为对比模型，系统分析了它们在医学图像分割任务中的精度与鲁棒性。

### 2.4.1 SAM 系列模型的结构与特点

Segment Anything Model (SAM) 是由 Kirillov 等人<sup>[59]</sup>于 2023 年提出的一种通用视觉分割模型，其设计旨在提供更灵活、高效且精确的图像分割解决方案。SAM 的核心目标是克服传统图像分割方法在处理多样化输入和复杂场景时的局限性。相较于现有的分割技术，SAM 提供了一个更为通用和强大的框架，能够在不同任务和输入类型下实现高效的图像分割。

如图 2.7 所示，SAM 模型的架构由三个主要组件组成：图像编码器、提示编码器和掩码解码器。这些组件紧密协作，确保了精确的图像分割。

SAM 模型中的图像编码器采用了基于 Vision Transformer (ViT) 的结构。ViT 是一种深度学习架构，能够有效捕捉图像的全局特征并处理复杂的视觉数据。得益于 ViT 的自注意力机制，图像的不同区域可以相互传递信息，这使得 SAM 能够在复杂背景或包

含多样物体的图像中提取出高质量的特征。ViT 在捕捉长距离依赖关系方面的优势使得 SAM 在各类图像数据中均能展现出出色的性能。

SAM 的提示编码器赋予模型处理多种类型输入提示的能力，包括点、框和文本等。用户可以通过交互式提示引导模型关注图像中的特定区域或物体。提示编码器的设计旨在增强模型的灵活性，使其能够根据用户的需求调整分割结果，从而实现对多种任务和环境的定制化图像分割。这一功能使得 SAM 在处理不同场景时，能够以更高的效率和准确性完成图像分割任务。

该结构中的掩码解码器则负责将图像编码器与提示编码器输出的特征进行融合，从而生成最终的分割掩码。掩码解码器的作用是通过精细解码图像特征，输出与输入图像相匹配的分割掩码。这一过程确保了 SAM 在保持高效性的同时，能够维持图像分割的高精度。

SAM 模型的一大优势在于其在超大规模数据集上的预训练。SAM 使用了一个名为 SA-1B 的数据集，该数据集包含超过 11 亿个分割掩码。通过在如此大规模的数据集上进行预训练，SAM 获得了卓越的零样本泛化能力。这使得即便没有针对特定任务的额外训练数据，SAM 也能够新的任务中实现高质量的分割。这种零样本学习能力使得 SAM 在实际应用中具有极强的适应性，能够处理各种新的、未见过的图像数据。

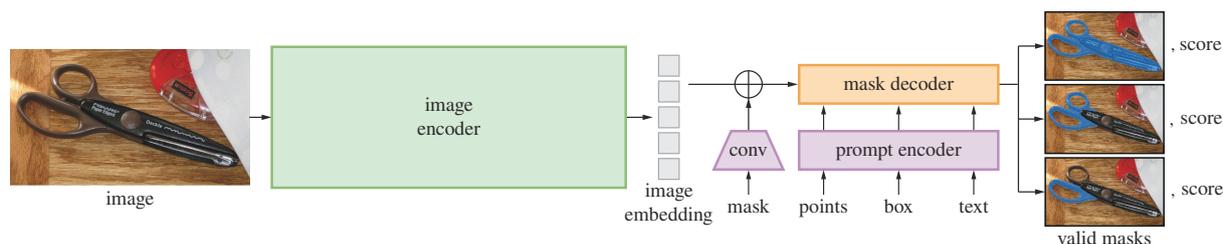


图 2.7 SAM 网络架构图<sup>[59]</sup>

Figure 2.7 Network architecture diagram of SAM

SAM 的灵活提示机制为用户提供了极大的便利。无论是点、框还是文本提示，SAM 都能够快速响应并生成精准的分割结果。该提示机制不仅使得图像分割过程更加交互式 and 高效，还使得用户能够根据需求灵活调整分割目标，而无需进行复杂的模型训练或调参。

#### 2.4.2 MedSAM 医学分割大模型

MedSAM<sup>[64]</sup> 是一款针对医学图像分割任务开发的 SAM 扩展模型。与通用的 Segment Anything Model (SAM) 不同，MedSAM 通过结合医学影像的领域特征与知识，进行了一系列优化与适配，以便在医学图像分割中发挥更大的优势。通过针对医学影像的特殊需求进行精确调整，MedSAM 在分割精度和效率上表现出显著提升，特别是在处理复杂医学影像时，展现出了较通用模型更优的性能。

在整体架构设计方面，MedSAM 延续了 SAM 的核心结构，但对多个关键组件进行了针对医学场景的优化。图 2.8 展示了 MedSAM 的网络结构，以下几方面的改进尤为突出。

在图像编码器部分，MedSAM 基于 Vision Transformer (ViT) 架构进行了微调，以便更好地适应医学影像的特征。医学影像的特殊性，如 CT、MRI 和超声图像，通常具有较高的分辨率和独特的结构特征，因此，MedSAM 通过对 ViT 架构的调整，使得模型在处理这些高分辨率图像时能够更精确地捕捉细节信息。这一改进显著提高了模型对复杂器官和病灶的分割能力，尤其是在细节呈现和解剖结构的捕捉方面，MedSAM 取得了显著优势。

在提示机制的设计上，MedSAM 针对医学影像的特性对传统的点、框和文本提示进行了优化。通过增强点提示功能，MedSAM 能够精准地处理微小病灶或肿瘤的精细分割，尤其适用于边界模糊或细小的目标区域。框提示则在整体器官的分割任务中表现出色，大大简化了人工标注的工作量；文本提示增强了模型对语义信息的理解，使得 MedSAM 在处理多模态数据时，能够有效整合不同来源的信息，从而进一步提高分割精度。

掩码解码器方面，MedSAM 进行了专门优化，以提升分割结果的空间连续性和细节准确性。医学影像分割不仅要求高精度，还必须确保分割结果的一致性和可靠性。因此，MedSAM 引入了一致性损失函数，以强化输出分割掩码的稳定性。该优化确保了模型在处理复杂病变或器官分割时，能够生成平滑且连贯的分割结果。

尽管 MedSAM 在医学影像分割领域取得了显著进展，但仍面临若干挑战。首先，尽管模型已在庞大的数据集上进行预训练，MedSAM 对医学领域特定数据的需求依然较高，尤其是在面对稀有疾病或特殊病灶时，数据的不足可能会影响模型的表现。其次，MedSAM 在小样本学习中的泛化能力有待进一步提升，尤其在数据量受限的情况下，如何保证分割精度和模型稳定性仍是一个挑战。此外，MedSAM 在跨模态数据适配上仍需进一步优化，以提高其对不同模态影像的适应能力和分割效果。随着临床需求的不断演变，MedSAM 需要在精度与可解释性方面持续优化，努力提升对复杂病灶与多模态数据的适应能力。

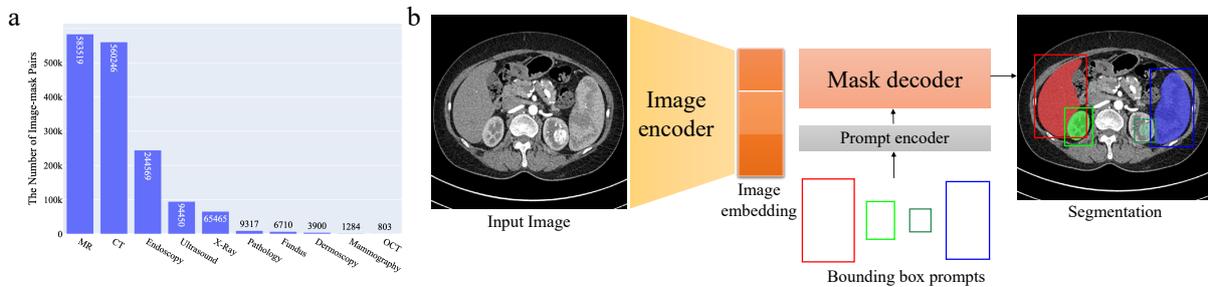


图 2.8 MedSAM 网络结构示意图<sup>[64]</sup>

Figure 2.8 Network architecture diagram of MedSAM

## 2.5 图像分割的评价指标

在图像分割任务中，评价指标能直接反映模型对目标区域或边界的预测精度。本文主要关注以下三种常见指标：交并比 (IoU)、Dice 系数 (DSC) 和 95% 百分位 Hausdorff 距离 (HD95)。它们分别从区域重叠度和边界距离两方面衡量分割效果，尤其在医学影像分割中具有广泛应用。

(1) 交并比 (IoU)。IoU 衡量预测区域与真实区域之间的重叠面积与其并集面积的比例，计算公式如下：

$$\text{IoU} = \frac{|A \cap B|}{|A \cup B|} \quad (2-1)$$

IoU 也是介于 0 到 1 之间的数值，值越大表示分割效果越好。

(2) Dice 系数 (DSC)。Dice 系数用于衡量预测区域与真实区域之间的重叠程度，公式如下：

$$\text{DSC} = \frac{2|A \cap B|}{|A| + |B|} \quad (2-2)$$

其中， $A$  表示模型预测的区域， $B$  表示真实标注的区域。Dice 系数范围为 0 到 1，值越接近 1 表示模型分割效果越好。

(3) 95% 百分位 Hausdorff 距离 (HD95)。用于衡量预测分割边界与真实分割边界之间的最大空间距离，公式如下：

$$\text{HD95}(A, B) = \max\{h_{95}(A, B), h_{95}(B, A)\} \quad (2-3)$$

其中， $A$  表示预测分割边界点集， $B$  表示真实标注边界点集， $h_{95}(A, B)$  表示从集合  $A$  中的点到集合  $B$  中最近点的所有距离值的 95% 百分位数。HD95 的单位通常为毫米 (mm)，值越小表示模型分割的边界精确度越高。

## 2.6 本章小节

综上所述，本章从多角度回顾了图像分割的经典模型与算法，探讨了小样本图像分割的方法演进与挑战，并结合视觉提示、Mamba 等关键技术，进一步扩展到大模型分割网络 (SAM 系列) 的最新进展，并且介绍图像分割的评价指标。具体归纳如下：

(1) 经典语义分割模型，FCN 与 U-Net 等网络深刻影响了后续分割方法的设计思路，但其在小样本条件下普遍存在对边缘精细度和全局上下文的不足，这为后续改进提供了动力。(2) 小样本分割方法，PANet、MENUA、PGP-SAM、Bro 等方法分别从原型对齐、记忆增强、提示生成、背景融合等不同角度，解决少数据带来的泛化难题，也展现了医学图像分割中独特的场景需求。(3) 其他相关技术，视觉提示 (VPT、Edge Prompt) 在参数效率与边界检测方面提供了新思路，Mamba 则在长序列建模中兼顾了效率与性能。(4) 大模型分割，另一方面，SAM 系列及其医学扩展 MedSAM 依托大规模预训练与提示机制，在通用分割与医学场景中均取得突破性进展，但在小样本与跨模态适配方

面仍存局限。(5) 图像分割评价指标, IoU、Dice 系数及 HD95 从不同维度衡量分割性能, 为后续实验设计与模型评估提供了客观量化标准。本章的综述与分析为后续研究奠定了坚实的理论与实践基础。后续实验章节将基于上述模型与技术思路, 结合小样本图像分割的实际需求, 详细阐述相关的网络结构与优化策略, 并在实验中基于这些评价指标对相关网络的性能进行检验。

## 第3章 基于多尺度记忆协调与边缘上下文聚合的小样本通用图像分割方法

### 3.1 引言

小样本通用图像分割在自动驾驶感知与目标检测等领域中具有广阔的应用前景。然而，当可用标注数据极其有限时，目标的尺度、形状及背景等方面的多样性会显著制约模型分割性能。传统方法（如 PFENet、ASGNet）通常仅依赖单一尺度的原型进行特征匹配，却容易忽视跨尺度差异与历史经验的积累，难以在复杂场景下保持泛化性。同时，背景噪声和边缘模糊的干扰又常导致边缘定位不准确。

针对上述问题，本章提出了一种基于多尺度记忆协调与边缘上下文聚合的小样本通用图像分割网络（Multi-Scale Edge Context Aggregation Network, MS-ECANet），以兼顾多尺度目标的精准识别与边缘细节的优化。具体而言，首先设计了多尺度记忆协调模块（Multi-Scale Memory Coordinate, MSMC）用于融合当前任务原型与历史记忆原型，从而有效缓解单一支持样本带来的语义信息不足问题；其次，提出边缘上下文聚合模块（Edge Context Aggregation Module, ECAM），利用 Sobel 算子与通道注意力机制显式提取并增强目标边缘特征；最后，通过构建原型匹配机制与特征融合解码器，在逐像素相似度计算基础上，生成高分辨率且清晰的目标掩码。为验证方法的有效性，本章将在 PASCAL-5<sup>i</sup>、COCO-20<sup>i</sup> 等标准数据集上开展对比与消融实验，并在 Abdominal-MRI 医学数据集上开展进一步测试，以验证该方法能否顺利迁移到医学图像中。图 3.1 展示了 MS-ECANet 模型整体设计思路。

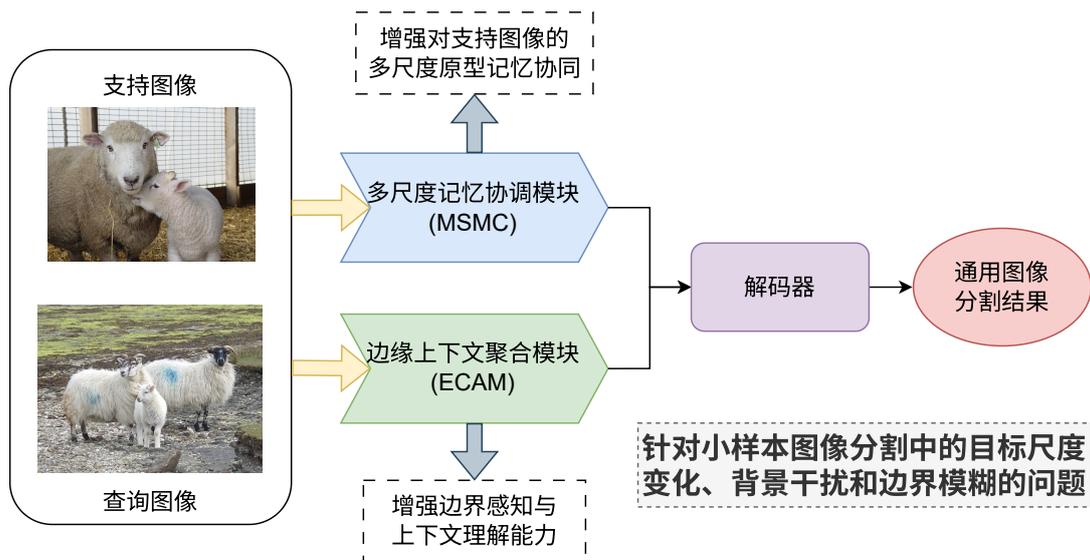


图 3.1 模型整体设计思路图

Figure 3.1 Diagram of model overall design idea

### 3.2 多尺度记忆协调与边缘上下文聚合网络的整体框架

#### 3.2.1 模型整体框架

如图 3.2 所示，多尺度记忆协调与边缘上下文聚合网络（Multi-Scale Edge Context Aggregation Network, MS-ECANet）旨在解决小样本图像分割中目标尺度多变、背景干扰强及边界模糊等痛点。

网络以 ResNet-101<sup>[72]</sup> 为骨干网络，并引入空洞空间金字塔池化（ASPP）以捕获全局多尺度上下文，为后续模块提供丰富语义先验。多尺度记忆协调模块通过支持图像掩码筛选目标区域特征，并与记忆库中历次任务的原型进行门控融合，形成自适应且具泛化能力的类别原型。该在情景训练过程中，该原型通过动态更新机制，利用新增样本和历史信息的迭代融合，有效减轻了小样本条件下的类别特征漂移。针对查询图像，边缘上下文聚合模块首先利用通道注意力强调与目标相关的高阶语义，再通过可分离卷积提取细粒度边缘特征，最终在空间维度进行逐像素加权融合。在解码端，网络将多尺度记忆原型与查询特征进行余弦相似度匹配，得到匹配分数图。随后将其与查询的低层特征级联，经两级反卷积逐步恢复至原分辨率，并融合输出的边缘特征以校正细微边界。为突出 MSMC 与 ECAM 在小样本分割中的互补作用，本章设计的总损失函数为加权求和形式，包含以下三项：主分割损失、边缘一致性损失及原型稳定性损失。

MS-ECANet 的核心结构包含以下模块：空洞空间金字塔池化模块、多尺度记忆协调模块和边缘上下文聚合模块等。本节将逐一分析其设计细节，最后讨论损失函数的设置。

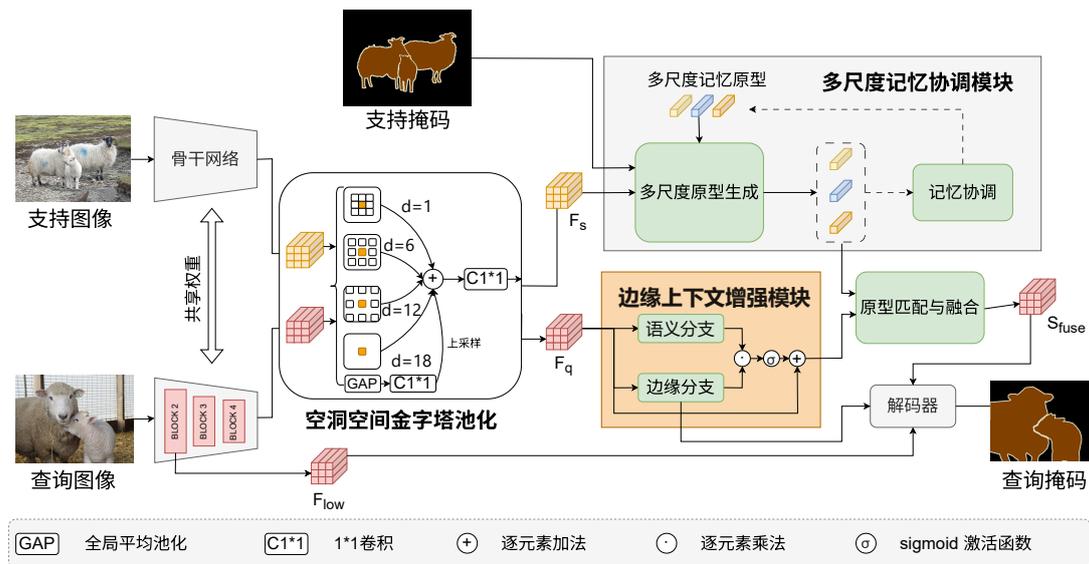


图 3.2 MS-ECANet 整体框架图

Figure 3.2 The overall framework diagram of MS-ECANet

### 3.2.2 空洞空间金字塔池化模块

空洞空间金字塔池化 (Atrous Spatial Pyramid Pooling, ASPP) 是一种经典的多尺度特征提取模块, 广泛应用于语义分割和目标检测等计算机视觉任务中。其核心思想是利用不同膨胀率 (dilation rate) 的空洞卷积, 在不增加额外参数或显著计算成本的前提下, 有效扩展卷积核的感受野, 从而提升模型对多尺度上下文信息的建模能力。

传统卷积在提取局部特征方面表现良好, 但对于需要全局语义理解的任务, 如语义分割, 单一尺度的卷积感受野往往不足以捕捉不同尺寸目标的上下文关系。ASPP 模块通过在多个分支中设置不同的扩张率, 使每个分支能够关注图像中不同空间尺度的特征响应, 从而在不牺牲分辨率的情况下, 获取更丰富的上下文信息。

如图 3.2 所示, 当骨干网络 (如 ResNet-101) 输出特征图  $F \in \mathbb{R}^{\frac{H}{4} \times \frac{W}{4} \times C}$  后, ASPP 模块会并行应用多个不同膨胀率 (如 6、12、18) 的空洞卷积分支, 以分别捕捉中、远距离的上下文信息。此外, 为保持细粒度的空间结构, 模块还引入一个  $1 \times 1$  卷积分支来强化局部感知能力。为了进一步增强模型的全局理解能力, 还通过全局平均池化 (GAP) 提取整幅图像的语义特征, 后续通过  $1 \times 1$  卷积降维并上采样至原特征图大小, 作为一个全局特征补充分支。

这些分支的输出特征在通道维度上进行拼接, 形成一个融合了多尺度信息的联合特征表示。随后, 使用一个  $1 \times 1$  卷积对拼接后的特征图进行压缩与融合, 以降低通道数并增强特征之间的交互, 最终得到结构紧凑且语义丰富的多尺度特征图  $F_{aspp}$ , 为后续的解码器或判别模块提供高质量的输入。

如式 (3-1) 所示, 空洞空间金字塔池化的具体公式为:

$$F_{aspp} = \text{Conv}_{1 \times 1} (\text{Concat}(F_{base}, F_6, F_{12}, F_{18}, G)) \quad (3-1)$$

其中,  $F_{base}$  为原始输入特征图,  $F_6, F_{12}, F_{18}$ , 分别是通过不同扩张率的空洞卷积得到的特征图,  $G$  是通过全局平均池化和上采样得到的全局特征图, 最终通过  $1 \times 1$  卷积融合得到的  $F_{aspp}$  多尺度信息融合后的特征图。通过这种多尺度与全局结合的方式, ASPP 能够在保留局部细节的同时, 最大化地利用上下文线索。

ASPP 在支持图像和查询图像中的应用有所不同, 因此分别针对它们的任务特点进行优化: (1) 支持图像: 在支持图像的处理过程中, ASPP 特征会经过掩码引导的空间加权, 以抑制背景干扰, 只保留目标区域的信息。处理后的特征图被送入多尺度记忆协调模块 (MSMC), 用于计算最终的支持原型, 从而保证模型能够更准确地进行目标分割。(2) 查询图像: 对于查询图像, ASPP 特征直接进入边缘上下文聚合模块 (ECAM), 该模块进一步提取边缘特征并增强语义表达。这样, ASPP 模块帮助查询图像聚焦于边缘信息, 提高分割精度, 并增强语义理解。

综上, ASPP 为 MS-ECANet 提供了多尺度与全局相结合的特征表达, 增强了模型对不同尺度目标的感知能力, 有助于捕获更丰富的上下文信息。这种多尺度特征提取机制结合模型其他组件, 共同提升了小样本分割场景下的性能表现。

### 3.2.3 多尺度记忆协调模块

多尺度记忆协调模块 (Multi-Scale Memory Coordinate, MSMC) 用于结合当前任务的多尺度原型与记忆库中的历史原型, 以提升小样本分割的泛化能力。在小样本学习中, 原型指的是由支持集中特定类别的像素特征聚合而成的向量, 它刻画了该类别在特征空间中的中心表示。记忆库则用来长期存储来自不同任务的历史原型及其统计信息, 以便在新任务中进行快速检索与融合, 从而提升模型的跨任务泛化能力。

该模块主要包含三个步骤: 首先计算当前任务的多尺度原型; 然后根据相似度与历史精度进行当前原型与历史原型的融合; 最后动态更新记忆库, 以持续优化跨任务的整体表现。如图 3.3 所示, MSMC 先利用 ASPP 生成的多尺度特征, 并通过掩码筛选目标区域后得到初步的支持原型。随后, 模型从记忆库中检索潜在相关的历史原型, 进行相似度计算和融合; 在此基础上, 再对记忆库进行更新, 以保证后续任务能复用更稳定的原型信息。

计算当前任务的多尺度原型时, 首先利用支持图像的掩码去除背景噪声, 仅保留目标区域。随后, 采用不同尺度的池化窗口 (如  $3 \times 3, 5 \times 5, 7 \times 7$ ) 计算支持原型, 以增强对不同尺度目标的适应性, 最终获得一组多尺度原型。为确保这些多尺度特征能够在同一特征空间内进行比较和融合, 模块采用插值或上采样等方式对不同尺度输出的特征图进行尺寸对齐。完成对齐后, 通过自注意力机制为各尺度特征分配权重, 最后采用加权平均的方式进行聚合, 从而得到一组能够综合多尺度信息原型表示。

MSMC 通过融合当前任务的多尺度原型与记忆库中的历史原型来提高原型的稳定性。记忆库中不仅存储了先前任务的原型, 还跟踪其调用频率和匹配精度, 以评估其有效性。假设  $P_k^{s, \text{current}}$  表示当前任务第  $k$  类的多尺度原型,  $P_k^{s', \text{hist}}$  表示历史原型向量。两者的相似度由余弦相似度定义, 如式 (3-2) 所示:

$$\text{Sim}(s, s') = \frac{P_k^{s, \text{current}} \cdot P_k^{s', \text{hist}}}{\|P_k^{s, \text{current}}\| \cdot \|P_k^{s', \text{hist}}\|} \quad (3-2)$$

接着将该相似度与  $\text{Accuracy}(s')$  等历史统计信息综合考虑, 得到每个历史原型的融合权重, 如式 (3-3) 所示:

$$\beta_{s'} = \frac{\text{Sim}(s, s') \cdot \text{Accuracy}(s')}{\sum_{s'} (\text{Sim}(s, s') \cdot \text{Accuracy}(s'))} \quad (3-3)$$

其中  $\beta_s$  越大代表该历史原型的贡献度越大。对当前与历史原型进行加权求和, 生成最终的支持原型, 如式 (3-4) 所示:

$$P_k^{\text{fused}} = \sum_{s=1}^3 P_k^{s, \text{current}} + \sum_{s'} \beta_{s'} P_k^{s', \text{hist}} \quad (3-4)$$

其中  $\beta_{s'}$  较低的历史原型在融合中权重较小, 以避免干扰最终表示。记忆库的动态更新基于滑动平均策略。当某一历史原型与当前原型的相似度高于设定阈值  $\tau$  (如 0.7) 时,

对该原型进行更新，如式 (3-5) 所示：

$$P_k^{s',hist} \leftarrow \alpha P_k^{s',hist} + (1 - \alpha) P_k^{fused}, \quad (\alpha = 0.9) \quad (3-5)$$

其中  $\alpha$  用于平滑，保证更新过程中历史信息得以保留。

为进一步提升检索效率，模块还结合以下策略：使用 PCA 降维减少存储维度；计算重要性评分（综合调用频率、匹配精度等）筛除低效原型；若多个原型与当前原型高度相似，则通过加权或池化进行合并，从而减少冗余。

得益于多尺度原型计算、历史原型融合与记忆库维护三者的协同，MSMC 能够自适应地保留高价值原型并淘汰无用原型，在小样本分割任务中显著提升了跨任务的稳定性与泛化性能。

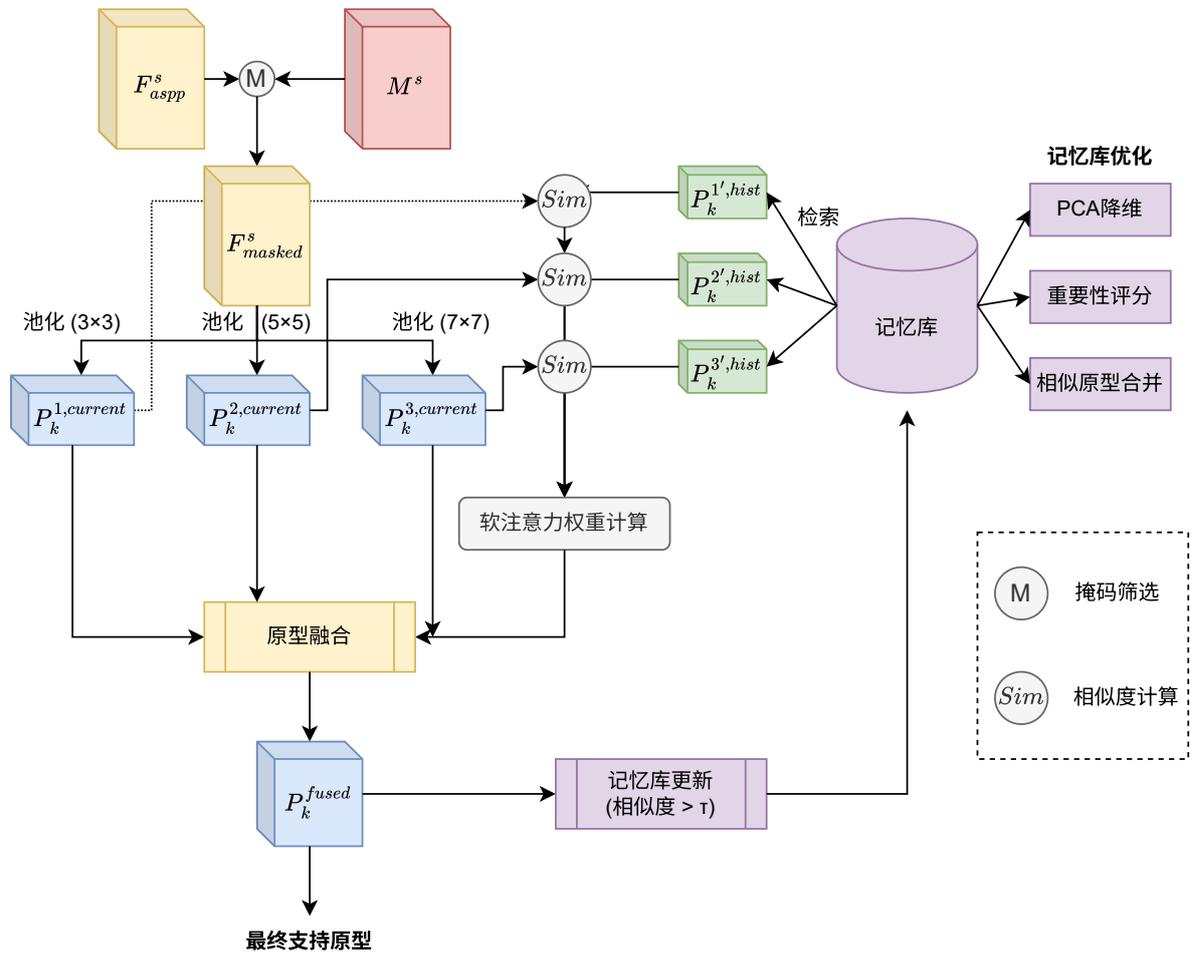


图 3.3 多尺度记忆协调模块结构图

Figure 3.3 The structure diagram of Multi-scale memory coordination module

### 3.2.4 边缘上下文聚合模块

边缘上下文聚合模块 (Edge Context Aggregation Module, ECAM) 通过结合边缘信息与语义特征，提高小样本环境下的目标区域辨识能力和边界清晰度。该模块包含以下

三个主要步骤：(1) 使用 Sobel 算子提取边缘特征；(2) 采用通道注意力机制增强语义特征；(3) 利用门控机制融合边缘和语义特征，从而抑制背景噪声并增强前景目标边缘。

如图 3.4 所示，ECAM 以 ASPP 生成的查询特征  $F_{aspp}^q$  为输入，首先通过 Sobel 算子提取边缘特征，随后应用通道注意力机制对语义特征进行增强，最后使用门控机制融合边缘特征与语义特征，最终得到增强后的查询特征  $F_{enhanced}^q$ 。

边缘信息的提取采用 Sobel 算子，该算子通过计算图像中水平方向与垂直方向的梯度，突出目标物体的边界。Sobel 算子计算公式，如式 (3-6) 所示：

$$F_{edge} = \sqrt{(\text{Sobel}_x(F_q))^2 + (\text{Sobel}_y(F_q))^2} \quad (3-6)$$

其中  $\text{Sobel}_x$  和  $\text{Sobel}_y$  分别表示水平方向和垂直方向的 Sobel 过滤操作。由于 Sobel 操作输出的特征通常为单通道或少通道特征，需通过  $1 \times 1$  卷积进行通道扩展，并对输出特征进行尺寸对齐，通常通过双线性插值进行上采样或下采样，使其与  $F_{aspp}^q$  具有相同的空间尺寸。

语义特征的提取采用高效通道注意力，该机制通过全局平均池化提取通道级特征，并使用多层感知机 (MLP) 计算通道级的注意力权重，如式 (3-7) 所示：

$$F_{sem} = \sigma(\text{MLP}(\text{GAP}(F_q))) \odot F_q \quad (3-7)$$

其中  $\text{GAP}(F_q)$  代表对特征图进行全局平均池化操作， $\sigma(\cdot)$  为 Sigmoid 激活函数， $\odot$  代表逐通道点乘运算。该操作确保模型能够在通道维度上自动学习重要特征，提高目标区域的辨识能力。本工作中，本模块主要负责优化 ASPP 输出的特征，进一步突出多尺度特征在每个通道上的重要性，并且聚焦边缘信息。

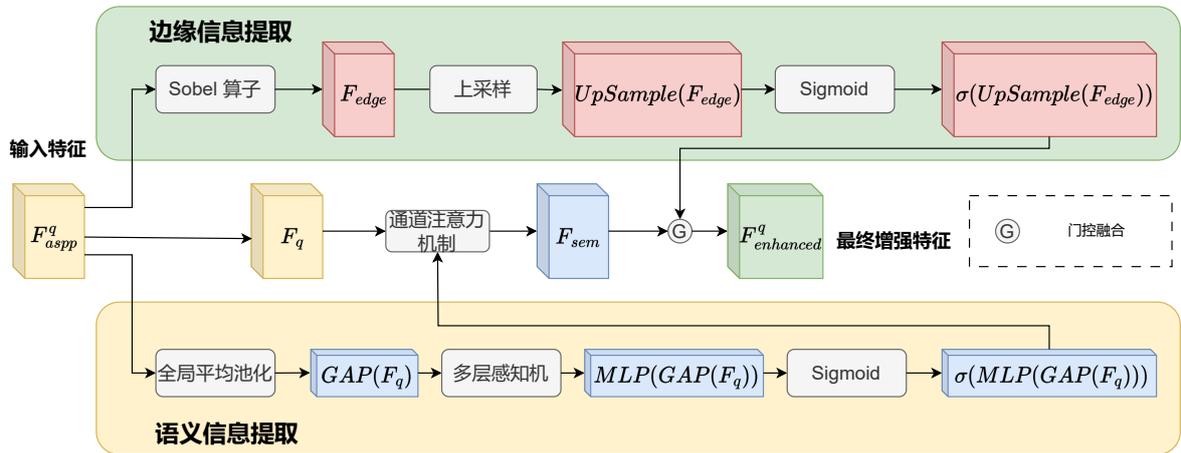


图 3.4 边缘上下文聚合模块结构图

Figure 3.4 The structure diagram of edge context aggregation module

在边缘信息和语义信息的融合过程中，ECAM 采用门控机制确保边缘特征不会过度影响全局语义特征，同时有效抑制背景噪声。融合过程如式 (3-8) 所示：

$$F_{enhanced} = F_{sem} \odot \sigma(\text{UpSample}(F_{edge})) \quad (3-8)$$

门控机制的核心思想是通过 Sigmoid 输出的边缘特征作为权重，使得语义特征在目标边界区域得到增强，而背景区域则被抑制。这里将  $F_{\text{edge}}$  经过插值操作与  $F_{\text{sem}}$  对齐，然后进行逐像素点乘运算，从而增强前景边缘。

通过这些步骤，ECAM 最终获得的增强特征  $F_{\text{enhanced}}^q$  在局部边缘和全局语义方面都得到强化，为后续的原型匹配与解码过程提供了更加鲁棒且精细的特征表示。

### 3.2.5 原型匹配

原型匹配是 MS-ECANet 进行小样本分割的核心步骤，通过计算查询图像每个像素与类别原型的相似度，生成匹配分数图，并最终用于目标区域的预测。该过程依赖高质量的类别原型，同时结合查询图像的多尺度特征，以确保不同尺寸的目标均能得到良好的匹配。

原型匹配的输入包括支持集生成的融合原型  $P^{\text{fused}}$  和查询图像的增强特征  $F_{\text{enhanced}}^q$ 。其中， $P^{\text{fused}}$  由 MSMC 计算得到，包含多尺度信息，而  $F_{\text{enhanced}}^q$  则由 ASPP 和 ECAM 处理后获得，包含查询图像的全局语义信息和局部边缘特征。匹配过程通过余弦相似度计算查询特征与类别原型的相似度，从而生成逐像素的匹配分数图，如式 (3-9) 所示：

$$\text{Sim}(s, s') = \frac{P_k^{s, \text{current}} \cdot P_k^{s', \text{hist}}}{\|P_k^{s, \text{current}}\| \|P_k^{s', \text{hist}}\|} \quad (3-9)$$

其中， $F_{\text{enhanced}}^q(x, y)$  表示查询图像在  $(x, y)$  处的特征向量，计算结果  $S_{\text{fuse}}(x, y)$  代表该像素属于目标类别的匹配置信度。为了优化匹配结果，计算完成后结合 ECAM 生成的边缘信息进行加权，以增强目标边界的清晰度，如式 (3-10) 所示：

$$S_{\text{fuse}} = S_{\text{fuse}} \odot \sigma(\text{UpSample}(F_{\text{edge}})) \quad (3-10)$$

其中， $F_{\text{edge}}$  是 ECAM 计算得到的边缘特征， $\sigma(\cdot)$  为 Sigmoid 函数， $\text{UpSample}(\cdot)$  表示上采样操作，确保边缘特征与匹配分数图具有相同的空间尺寸。通过这一优化，模型能够在目标边界处进行更加精准的匹配，减少背景误分类。最终，优化后的  $S_{\text{fuse}}$  被输入到解码器中，用于生成最终的目标掩码。

### 3.2.6 解码器

解码器用于将原型匹配生成的匹配分数图转换为最终的分割掩码。通过融合查询图像的低层特征、边缘信息和匹配分数，该模块能够增强目标区域的细节表达，提高分割精度，尤其是在边界模糊或复杂背景干扰的情况下。解码器的主要目标包括：(1) 融合低层特征以恢复目标的空间细节；(2) 利用边缘信息优化目标边界；(3) 通过卷积预测像素类别，生成最终的分割掩码。

如图 3.5 所示，解码器的输入包括匹配分数图  $S_{\text{fuse}}$ 、查询图像的低层特征  $F_{\text{low}}^q$  以及边缘信息  $F_{\text{edge}}$ 。首先，对匹配分数图进行上采样，使其空间尺寸与低层特征对齐，并在

通道维度进行拼接。随后，引入上采样后的边缘特征，以增强目标边界信息。最终，通过卷积预测每个像素的类别概率，生成分割掩码。

匹配分数图  $S_{fuse}$  由原型匹配模块生成，表示查询图像中每个像素与多尺度记忆原型的匹配置信度。查询图像的低层特征  $F_{low}^q$  由 ResNet-101 的 Block2 层提取，包含高分辨率的空间信息。由于  $S_{fuse}$  在前向传播中通常经历了下采样操作，为了保证它与  $F_{low}^q$  在逐像素维度上的准确对齐，需要首先采用双线性插值进行上采样，如式 (3-11) 所示：

$$S_{up} = \text{UpSample}(S_{fuse}) \quad (3-11)$$

双线性插值方法能够平滑地恢复空间分辨率，适用于细粒度特征的恢复。将上采样后的  $S_{up}$  与低层特征在通道维度上拼接，补充高层特征中的细节信息，如式 (3-12) 所示：

$$F_{skip} = \text{Concat}(F_{low}^q, S_{up}) \quad (3-12)$$

边缘信息  $F_{edge}$  由 ECAM 提供，突出查询图像的边缘与轮廓。ECAM 通过卷积神经网络从低层特征中提取高频边缘信息。为了与  $F_{skip}$  保持一致的分辨率，需要对  $F_{edge}$  进行上采样，如式 (3-13) 所示：

$$F_{edge, up} = \text{UpSample}(F_{edge}) \quad (3-13)$$

通过插值操作确保边缘特征的分辨率与  $F_{skip}$  相同，然后将其与  $F_{skip}$  拼接，进一步增强目标边界信息，如式 (3-14) 所示：

$$F_{decode} = \text{Concat}(F_{skip}, F_{edge, up}) \quad (3-14)$$

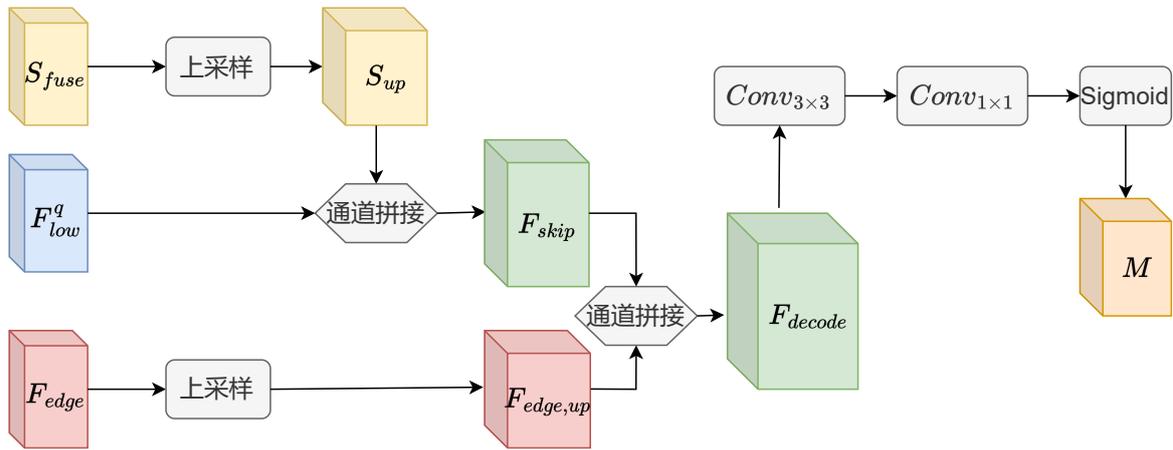


图 3.5 解码器结构图

Figure 3.5 The structure diagram of decoder

最后，解码器使用  $3 \times 3$  和  $1 \times 1$  卷积对融合后的特征进行提取与通道压缩，并通过 Sigmoid 激活函数输出分割掩码，如式 (3-15) 所示：

$$M = \sigma(\text{Conv}_{1 \times 1}(\text{Conv}_{3 \times 3}(F_{decode}))) \quad (3-15)$$

其中,  $\text{Conv}_{3\times 3}$  提取局部细节特征,  $\text{Conv}_{1\times 1}$  将通道数映射为单通道输出,  $\sigma(\cdot)$  将结果归一化到  $[0, 1]$ 。

在小样本分割场景中, 训练样本有限, 模型容易受到背景噪声和边界模糊等问题的影响。通过融合高分辨率的低层特征和匹配分数图, 解码器能够在空间细节方面进行补偿; 同时引入边缘特征赋予目标边界更高权重, 有效抑制背景干扰并强化边界区域。这样的多源特征结合有效缓解了小样本条件下的复杂场景适应问题, 并提高了最终分割结果的精度和稳定性。

### 3.2.7 MS-ECANet 损失函数

此外, 在小样本分割中, 由于标注数据稀缺, 模型更依赖于结构性设计与训练策略的配合。仅靠网络结构的增强仍不足以完全解决特征稀疏、边界模糊等问题。为了进一步引导模型高效学习多尺度与边缘信息, 合理设计损失函数成为提升性能的另一关键因素。为充分发挥多尺度记忆协调模块 (MSMC) 与边缘上下文聚合模块 (ECAM) 在小样本分割中的作用, 本文将损失函数设计为多个子损失的加权和, 如式 (3-16) 所示:

$$\mathcal{L}_{\text{MS-ECA}} = \alpha \cdot \mathcal{L}_{\text{seg}} + \beta \cdot \mathcal{L}_{\text{edge}} + \gamma \cdot \mathcal{L}_{\text{proto}} \quad (3-16)$$

具体而言, 各子损失及其作用如下所示:

- $\mathcal{L}_{\text{seg}}$ : 主分割损失 (如 Dice、BCE 或 CE), 用于约束最终的目标掩码预测;
- $\mathcal{L}_{\text{edge}}$ : 边缘一致性损失, 可利用 Sobel/方向梯度等方法度量预测掩码与真实边缘的吻合度, 强化 ECAM 模块的边界感知;
- $\mathcal{L}_{\text{proto}}$ : 原型稳定性损失, 用以在 MSMC 中约束多尺度原型与记忆库原型的偏差, 减少跨任务原型漂移;
- $\alpha, \beta, \gamma$ : 超参数权重, 用于平衡分割与边缘、原型各项的贡献度, 经测试选择系数为:  $\alpha = 1.0, \beta = 0.3, \gamma = 0.1$ 。

在表 3.1 中, 不同损失权重组合在 PASCAL-5<sup>i</sup> 5-shot 分割上的对比结果表明: 当取 ( $\alpha = 1.0, \beta = 0.3, \gamma = 0.1$ ) 时, mIoU 达到 75.5%, 相较其他组合取得了更好的平衡性并显著提升分割性能。

当  $\beta$  和  $\gamma$  的权重增大时, 例如组合 (1.0, 0.5, 0.1), 模型的性能略有下降 (72.9%), 可能是因为较高的边缘一致性损失对分割精度的贡献有限, 而更高的原型损失权重并没有带来更大提升。

通过这些实验可以看出, 在小样本分割任务中, 合理的损失权重设置能够有效地平衡不同模块的作用, 进而提高分割精度。

表 3.1 不同权重组合对 PASCAL-5<sup>i</sup> 分割性能的影响 (mIoU, %)

Table 3.1 Effect of different weight combinations on segmentation performance of PASCAL-5<sup>i</sup> (mIoU, %)

实验编号	损失权重设置 ( $\alpha, \beta, \gamma$ )	mean (%)
1	(1.0, 0.0, 0.0)	69.3
2	(1.0, 0.3, 0.0)	73.8
3	<b>(1.0, 0.3, 0.1)</b>	<b>75.5</b>
4	(1.0, 0.5, 0.1)	72.9
5	(1.0, 0.3, 0.2)	74.1
6	(1.0, 0.1, 0.1)	73.6
7	(0.8, 0.3, 0.1)	74.3
8	(1.2, 0.3, 0.1)	75.9
9	(1.0, 0.0, 0.1)	71.4

### 3.3 相关实验及结果

#### 3.3.1 数据集

本实验采用 PASCAL-5<sup>i</sup>[65] 和 COCO-20<sup>i</sup>[41] 两个小样本语义分割基准数据集。PASCAL-5<sup>i</sup> 是由 PASCAL VOC 2012[73] 与 SDS 的扩展标注构建而成。它将 20 个类别平均划分为 4 个交叉验证折 (split0、split1、split2、split3)，每折包含 5 个类别；在每个测试阶段，遵循 OSLSM 标准协议，随机采样 1000 个查询支持对进行评估。

COCO-20<sup>i</sup> 数据集源自 MS-COCO[74]，按照文献[41] 的方式将 80 个类别分为同样的 4 个折 (split0~3)。具体地，其类别索引集合定义为  $\{4x - 3 + i\}$  (其中  $x \in \{1, \dots, 20\}$ ,  $i \in \{0, 1, 2, 3\}$ )。在此基础上，为保证结果稳定，每折评估时均从包含 40,137 张验证图像的数据集中随机采样 20,000 个查询支持对。测试过程同样采用交叉验证策略：在对某一折 (split $i$ ) 进行评估时，模型使用剩余三折的数据进行训练。

#### 3.3.2 实验配置

所有实验均在 Ubuntu 20.04 操作系统下，基于 PyTorch 1.12 深度学习框架进行实现，硬件环境为 NVIDIA RTX 4090 显卡 (24GB 显存)，并启用混合精度训练加速。所有模型的训练与推理均在单卡环境中完成。训练时采用随机梯度下降 (SGD) 优化器，动量系数为 0.9、权重衰减系数为  $1 \times 10^{-4}$ 。在 PASCAL-5<sup>i</sup> 上训练 200 个 epoch，初始学习率为  $2.5 \times 10^{-3}$ ，批次大小为 4；针对更大规模的 COCO 数据集，训练周期调整为 50 个 epoch，初始学习率升至  $5 \times 10^{-3}$ ，批次大小为 8。骨干网络参数在训练过程中保持固定，仅更新其他模块参数。在将图片输入网络之前，将所有图像都调整为  $384 \times 384$ ，并使用 ImageNet 均值和均方差进行归一化。遵循 HSNet[75] 的协议，不应用额外的数据增强 (例如，翻转、缩放或颜色抖动) 以确保公平比较。

### 3.3.3 对比实验与分析

在以 ResNet-101 为骨干网络的实验中，MS-ECANet 在 PASCAL-5<sup>i</sup> 数据集的 1-shot 设置下取得了 77.8% 的 mIoU，较现有最优方法提升 7.3% 的 mIoU，刷新了该数据集下的最新性能记录。进一步观察各折结果可见，MS-ECANet 在 split3 测试集上达到 73.8% 的分割精度，与 HSNet 相比实现了 10.7% 的显著增益；而在 5-shot 设置下，模型 mIoU 继续提升至 79.2%。实验结果表明，本文提出的两个核心模块显著增强了模型对目标细节与上下文信息的捕获能力，使其在面对分割任务时展现出较好的性能。

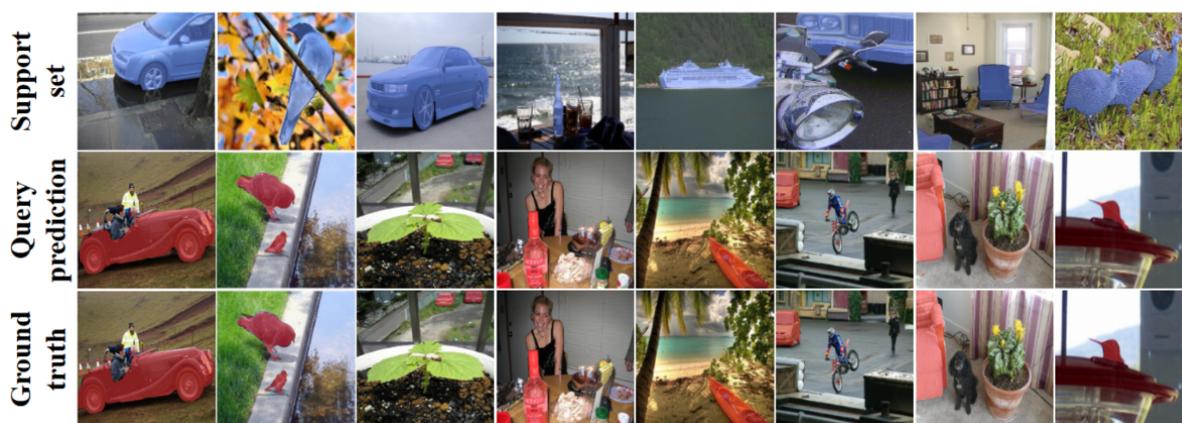


图 3.6 PASCAL-5<sup>i</sup> 数据集的模型分割结果

Figure 3.6 segmentation results of the model of PASCAL-5<sup>i</sup> dataset

表 3.2 PASCAL-5<sup>i</sup> 数据集的各模型结果对比 (mIoU, %)

Table 3.2 Comparison of each model result of PASCAL-5<sup>i</sup> dataset (mIoU, %)

模型	骨干网络	1-shot					5-shot				
		split0	split1	split2	split3	mean	split0	split1	split2	split3	mean
DAN <sup>[28]</sup>	Res-101	54.7	68.6	57.8	51.6	58.2	57.9	69.0	60.1	54.9	60.5
PMM <sup>[76]</sup>		54.7	68.6	57.8	51.6	58.2	57.9	69.0	60.1	54.9	60.5
PFENet <sup>[22]</sup>		60.5	69.4	54.4	55.9	60.1	62.8	70.4	54.9	57.6	61.4
CyCTR <sup>[30]</sup>		67.2	71.1	57.6	59.0	63.7	71.0	75.0	58.5	65.0	67.4
HSNet <sup>[75]</sup>		67.3	72.3	62.0	63.1	66.2	71.8	74.4	67.0	68.3	70.4
ASGNet <sup>[25]</sup>		59.8	67.4	55.6	54.4	59.3	64.6	71.3	64.2	57.3	64.4
SSP <sup>[77]</sup>		63.7	70.1	66.7	55.4	64.0	70.3	76.3	<b>77.8</b>	65.5	72.5
DCAMA <sup>[39]</sup>		65.4	71.4	63.2	58.3	64.6	70.7	73.7	66.8	61.9	68.3
SGT <sup>[78]</sup>		70.3	<b>76.6</b>	70.4	64.8	70.5	73.2	<b>78.0</b>	73.7	66.2	72.8
RiFeNet <sup>[79]</sup>		71.6	73.5	58.0	61.2	66.1	75.3	76.9	59.6	65.1	69.2
<b>MS-ECANet (Ours)</b>	Res-101	<b>73.4</b>	73.6	<b>72.7</b>	<b>71.8</b>	<b>72.9</b>	<b>76.7</b>	75.6	75.4	<b>74.2</b>	<b>75.5</b>

与之形成对比，传统的原型匹配方法（如 DAN、PMM）在 split1 至 split3 的跨域测试中平均精度大多低于 60%。即使是结合跨样本注意力机制的 CyCTR，其 1-shot 性能也仅为 63.7%，在高纹理复杂度或类别差异较大的场景中仍面临较大挑战。而 MS-ECANet 则在分割边界与局部细节的处理上展现出更具优势的表现，进一步验证了融合多尺度特征与注意力机制的必要性。

此外，如图 3.6 所示，本节给出了 MS-ECANet 在 PASCAL-5<sup>i</sup> 数据集 1-shot 设置下的分割掩码示例。图中可见，模型输出的目标边缘较为连续，细节保持良好，对于远景小目标亦无明显漏检；这些直观结果与前述定量指标基本一致，进一步说明了模型在复杂纹理与尺度变化条件下的稳健性。

在更大规模的 COCO-20<sup>i</sup> 数据集测试中，MS-ECANet 依然表现出较高的分割性能。与 HDMNet 相比，在 1-shot 和 5-shot 设置下，模型的 mIoU 分别提升了 1.8 和 1.4 个百分点；相较于 PFENet，其 mIoU 提升幅度超过 10 个百分点。这说明该方法在类别数量更多、数据分布更复杂的情形下仍能保持较高精度，进一步证明了其在小样本分割任务中的泛化能力。此外，COCO-20<sup>i</sup> 的评估结果还表明，在多类目标并存且标注噪声较高的环境中，该方法依旧保持了良好的性能，表现出较好的抗干扰能力，进一步彰显了本模型设计的合理性。

表 3.3 COCO-20<sup>i</sup> 数据集的各模型结果对比 (mIoU, %)

Table 3.3 Comparison of each model result of COCO-20<sup>i</sup> dataset (mIoU, %)

模型	骨干网络	1-shot					5-shot				
		split0	split1	split2	split3	mean	split0	split1	split2	split3	mean
PPNet <sup>[26]</sup>		28.1	30.8	29.5	27.7	29.0	39.0	40.8	37.1	37.3	38.5
PMM <sup>[76]</sup>		29.3	34.8	27.1	27.3	29.6	33.0	40.6	30.3	33.3	34.3
CyCTR <sup>[30]</sup>		38.9	43.0	39.6	39.8	40.3	41.1	48.9	45.2	47.0	45.6
HSNet <sup>[75]</sup>	Res-50	36.3	43.1	38.7	38.7	39.2	43.3	51.3	48.2	45.0	46.9
SSP <sup>[77]</sup>		46.4	35.2	27.3	25.4	33.6	53.8	41.5	36.0	33.7	41.3
DCAMA <sup>[39]</sup>		41.9	45.1	44.4	41.7	43.3	45.9	50.5	50.7	46.0	48.3
SGT <sup>[78]</sup>		49.3	55.8	49.6	49.7	49.9	49.9	61.2	55.2	55.0	55.3
RiFeNet <sup>[79]</sup>		39.1	47.2	44.6	45.4	44.1	44.3	52.4	49.3	48.4	48.6
PI-CLIP(HDMNet) <sup>[80]</sup>		49.3	65.7	55.8	56.3	56.8	56.4	66.2	55.9	58.0	59.1
<b>MS-ECANet (Ours)</b>	Res-50	<b>50.9</b>	<b>67.3</b>	<b>57.4</b>	<b>57.9</b>	<b>57.4</b>	<b>57.8</b>	<b>67.6</b>	<b>57.3</b>	<b>59.4</b>	<b>60.5</b>

### 3.3.4 消融实验

本节在 1-shot 设置下使用 PASCAL-5<sup>i</sup> 上的 ResNet101 主干进行了一系列消融测试。该测试可以评估每个组件对分割性能的影响并验证其有效性。除非另有说明，所有消融实验均基于 PASCAL-5<sup>i</sup> 数据集并在 1-shot 设置下使用 ResNet-101 主干网络进行。

### 3.3.4.1 MSMC 与 ECAM 消融实验

为评估多尺度记忆协调模块 (MSMC) 与边缘上下文聚合模块 (ECAM) 对模型性能贡献, 本文在 PASCAL-5<sup>i</sup> 数据集上开展了 1-shot 消融实验。实验基线选用 DCAMA (mIoU=64.6%), 随后逐步引入各模块以评估其对分割性能的影响, 相关结果见表 3.4。

首先, 单独引入 ECAM 模块后, 模型的 mIoU 提升了 3.4 个百分点, 达到了 68.0%。其中, ECAM 在 split2 子集上的表现尤为突出, 提升了 3.0 个百分点 (66.2% vs. 63.2%), 表明 ECAM 通过强化边缘特征的机制, 有效提升了目标边界模糊区域的辨识能力。

相对于的, 当单独引入 MSMC 时, 模型在 split2 子集上的 mIoU 达到 72.7%, 比基线提高 8.1 个百分点, 证明其在跨尺度特征融合方面具备较强的适应能力。然而, 在 split3 子集中, MSMC 的表现略逊于 ECAM, 表明 MSMC 在面对复杂背景时尚无法完全发挥自身优势。

当两者组合使用时, 模型分割精度明显提高: 在 split1 子集达 73.6%, 整体 mIoU 提高至 72.9%, 相较基线增幅达 8.3 个百分点。这种大幅提升主要得益于 ECAM 与 MSMC 在特征提取上的互补性: 前者强化边缘轮廓, 后者聚合多尺度特征与历史信息, 从而兼顾局部细节与全局一致性。

表 3.4 PASCAL-5<sup>i</sup> 数据集的各模块消融实验结果 (mIoU, %)

Table 3.4 Ablation results of each module of PASCAL-5<sup>i</sup> dataset (mIoU, %)

模块	split0	split1	split2	split3	mean
Baseline (DCAMA)	65.4	71.4	63.2	58.3	64.6
Baseline+ECAM (only)	67.3	71.8	66.2	66.8	68.0
Baseline+MSMC (only)	69.5	72.1	71.3	66.4	69.8
Baseline+ECAM+MSMC	<b>73.4</b>	<b>73.6</b>	<b>72.7</b>	<b>71.8</b>	<b>72.9</b>

### 3.3.4.2 多尺度原型与单尺度全局原型的对比

为评估多尺度原型对模型性能的影响, 本文在 PASCAL-5<sup>i</sup> 数据集上进行了 1-shot 消融实验, 重点对多尺度记忆协调模块在多尺度原型与单尺度全局原型两种策略下的表现进行了对比分析 (两者均采用记忆协调方法)。多尺度原型通过在不同尺度下提取特征, 以适应多样化的目标尺寸; 而单尺度全局原型则依赖固定尺度特征, 在跨尺度目标时可能存在一定局限。

如表 3.5 所示, 采用单尺度全局原型的模型在 split0、split1 和 split3 子集上的 mIoU 均低于多尺度原型策略, 整体 mIoU 为 67.8%。相比之下, 多尺度原型策略的适应性更佳, 整体 mIoU 提高 2.0 个百分点, 达 69.8%。

从各子集的增益来看, 多尺度原型在小目标比例较高的 split0 与 split2 中分别带来 1.9% 和 6.3% 的提升, 说明跨尺度特征对增强小目标分割精度具有积极作用。然而, 在

以大目标为主的 split3 子集中，两者的性能差距仅 1.4% (66.4% 对 65.0%)，表明当目标尺寸相对统一时，单尺度全局原型也能提供较为相对充分的类别表征。

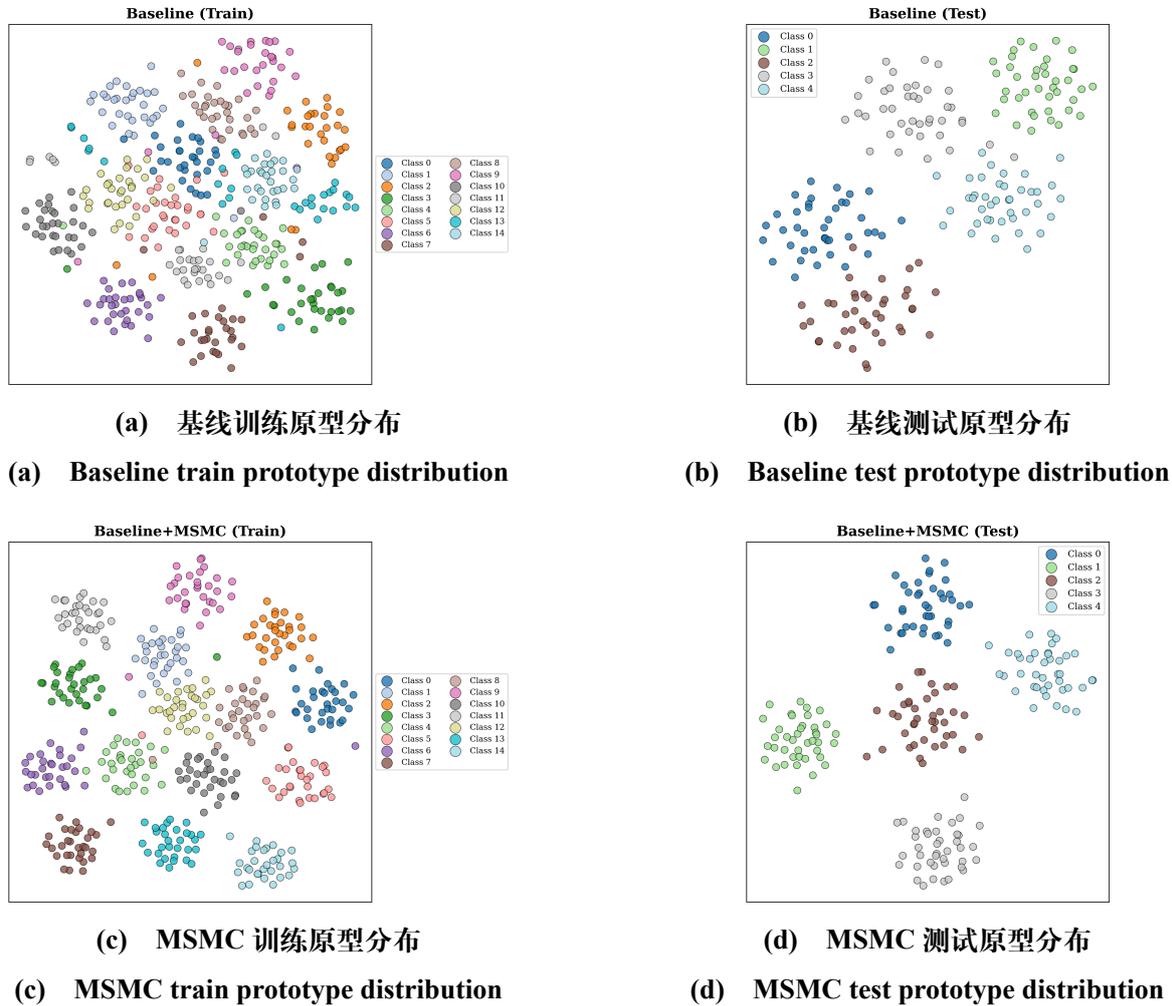


图 3.7 t-SNE 特征分布可视化

Figure 3.7 Visualization of t-SNE feature distribution

为进一步探究两种原型方案在特征空间的分布情况，本文对训练测试过程中的原型向量进行了 t-SNE 聚类可视化（见图 3.7）。图 3.7a 与图 3.7c 分别展示了训练阶段在 15 个类别上的原型分布，多尺度原型具有更清晰的类间边界和更紧凑的类内分布；图 3.7b 与图 3.7d 则对应测试阶段的原型分布，尽管整体紧凑度有所降低，但多尺度原型仍保持了较为明确的判别结构，表明其在跨尺度目标条件下具有较好的泛化能力。

总体而言，实验结果表明，多尺度原型在处理多样化目标时能够有效提升模型的鲁棒性与分割精度，尤其在小目标和尺度差异较大的场景中表现更优。相比之下，单尺度全局原型由于固定感受野限制，在跨尺度任务中性能相对不足，进一步说明多尺度特征表达对于提升图像分割性能具有重要意义。

表 3.5 不同原型方法结果对比 (mIoU, %)

Table 3.5 Comparison of results of different prototyp methods (mIoU, %)

方法	split0	split1	split2	split3	mean
单尺度全局原型	67.6	71.7	68.9	65.0	67.8
多尺度原型 (本方法)	<b>69.5</b>	<b>72.1</b>	<b>71.3</b>	<b>66.4</b>	<b>69.8</b>

### 3.3.5 在医学图像上的适应性分析

在之前的实验中, 本章已经证明了 MS-ECANet 在自然图像小样本分割任务上的有效性, 且其在 PASCAL-5<sup>i</sup> 和 COCO-20<sup>i</sup> 数据集上表现出了较好的性能。然而, 医学影像与自然图像在成像模式、数据分布、解剖结构等方面存在显著差异, 直接将通用小样本分割算法迁移到医学图像领域时, 往往会面临许多挑战。

医学图像模态多样 (CT、MRI、超声等), 灰度分布与纹理特征差异明显, 且常伴噪声、运动伪影及背景干扰, 使特征提取与目标分割更为困难。解剖结构个体差异大, 同一器官或病灶在不同患者中形态、尺寸各异, 进一步加剧定位难度。与此同时, 医学影像的高质量标注通常依赖专业医师的人工标注, 导致标注成本高昂且公开数据集稀缺。对于罕见病或特殊临床场景, 可用的公开标注数据尤为匮乏。

因此, 尽管小样本学习为解决数据稀缺问题提供了新思路, 但通用小样本分割方法在医学领域的适用性仍需深入探讨。本研究将通过系统实验与分析, 全面评估通用小样本分割算法在医学领域的适应性, 主要从以下三个方面展开: (1) 特征处理与匹配策略。通过与医学图像分割方法对比, 分析通用小样本分割算法在医学数据上的表现, 识别哪些特征处理与匹配策略在医学场景中依然有效, 哪些则需要进行特殊定制或改进; (2) 主要问题与挑战。识别出通用模型在医学图像数据上表现的主要问题, 例如在处理复杂背景、模糊边缘或解剖变异时的不足; (3) 为后续优化提供依据。为第四章即将提出的医学图像分割专用模块提供实践佐证与改进思路, 以便在设计新的医学图像分割算法时, 能够更有针对性地优化对医学图像的适应性。

综上所述, 本节将结合简单的迁移实验与理论分析, 从定量、定性及消融实验的角度, 评估通用小样本分割算法在医学影像上的适应程度。这一分析将为后续第四章提出的医学分割专用方法提供必要的背景和理论依据。

#### 3.3.5.1 相关实验及结果

本实验基于腹部 MRI 数据集<sup>[81]</sup> 进行评估, 数据集来源于 MICCAI 2015 多模态腹部图像分割挑战赛, 包含多个关键器官 (如肝脏、脾脏、左肾、右肾) 的标注。为确保实验的公平性, 并适配小样本学习的特点, 本实验严格遵循 Q-Net 等先进方法的标准化预处理流程。首先, 对影像直方图进行顶部 0.5% 的强度截断, 以减少极端噪声对模型学习的干扰。随后, 根据不同解剖部位选择标准切面: 腹部数据集采用轴向切面, 心脏数据集采用短轴切面, 并在保持原始分辨率的基础上进行重采样。在数据输入层面, 单

通道医学图像在通道维度复制三次，以适配主流深度学习网络的 RGB 输入格式，最终统一裁剪至 256×256 像素。该预处理策略既能保留影像的关键解剖信息，又能确保不同数据集之间的尺度一致性，为模型的跨数据集评测提供了可靠的基准。

表 3.6 腹部 MRI 数据集 Setting-1 设置下的各模型结果对比 (Dice, %)

Table 3.6 Comparison of results of each model on abdominal MRI dataset under Setting-1 (Dice, %)

模型	Liver	RK	LK	Spleen	Mean
SE-Net <sup>[49]</sup>	29.02	47.96	45.78	47.30	42.51
ALPNet <sup>[82]</sup>	62.35	48.42	44.73	49.61	51.28
SSL-PANet <sup>[82]</sup>	71.73	60.81	58.83	61.32	63.17
SSL-ALPNet <sup>[82]</sup>	76.10	85.18	81.92	72.18	78.84
ADNet <sup>[54]</sup>	82.11	85.08	73.86	72.29	78.51
AAS-DCL <sup>[52]</sup>	72.33	86.11	80.37	76.24	78.76
SR&CL <sup>[83]</sup>	73.94	76.01	80.87	75.60	77.00
Q-Net <sup>[55]</sup>	81.74	87.98	78.36	75.99	81.02
CRAPNet <sup>[56]</sup>	76.46	86.15	84.95	74.32	79.79
PAMI <sup>[84]</sup>	<b>82.59</b>	<b>88.73</b>	81.83	76.37	82.3
DGAPNet <sup>[85]</sup>	81.31	86.99	<b>85.84</b>	<b>80.64</b>	<b>83.70</b>
MS-ECANet (Ours)	79.61	84.25	83.13	78.02	81.25

数据集采用 5 折交叉验证（每折从 5 份数据中随机选 1 作为测试集，其余作为训练集）。实验使用 1-shot 设置（即每个类别仅提供一个支持样本），以考察小样本条件下模型的分割能力。实验指标：本实验主要采用 Dice 指标，以评估模型在不同解剖区域的分割精度。

在此部分，实验对比了 MS-ECANet 与其他当前主流医学图像分割方法（如 SE-Net、Q-Net 等）在腹部 MRI 数据集上的分割性能。下表展示了不同模型在腹部 MRI 数据集上的 Dice 分数，提供了与现有方法的基准对比。

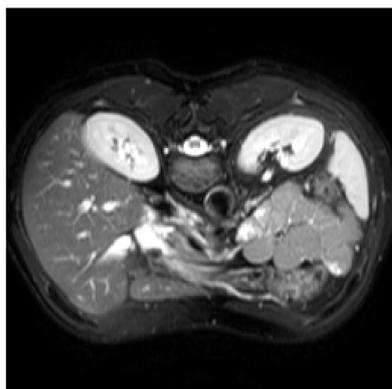
由表中可见，MS-ECANet 在各关键器官的分割任务中均优于 SE-Net 和 Q-Net，尤其在左肾与脾脏上表现出较高的 Dice 分数，整体性能具有竞争力。随后，本节通过模块消融实验进一步验证第三章提出的多尺度记忆协调模块 (MSMC) 和边缘上下文聚合模块 (ECAM) 对医学图像分割性能的贡献。

在消融实验中，本文比较了四种模型配置：(1) 基线模型。不使用 MSMC 和 ECAM，仅使用标准卷积层进行特征提取。(2) 移除 MSMC。仅保留 ECAM 模块，测试其对模型性能的影响。(3) 移除 ECAM。仅保留 MSMC 模块，测试其对模型性能的影响。(4) MS-ECANet 完整模型。包含多尺度记忆协调模块 (MSMC) 与边缘上下文聚合模块 (ECAM)。

表 3.7 腹部 MRI 数据集 Setting-1 设置下的各模块消融实验结果 (Dice, %)

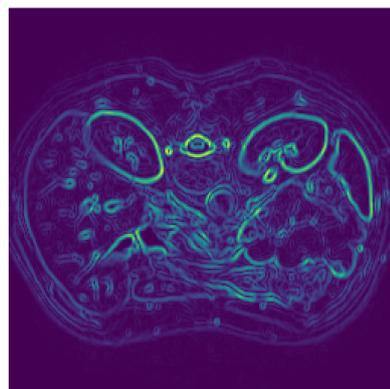
Table 3.7 Ablation results of each module of abdominal MRI dataset under Setting-1 (Dice, %)

Components	LK	RK	Spl.	Li.	Mean
Baseline (DCAMA)	67.92	66.88	64.73	60.42	76.99
Baseline+ECAM (only)	79.58	82.76	82.15	77.12	80.40
Baseline+MSMC (only)	77.33	81.91	81.42	75.48	79.04
Baseline+ECAM+MSMC	<b>79.61</b>	<b>84.25</b>	<b>83.13</b>	<b>78.02</b>	<b>81.25</b>



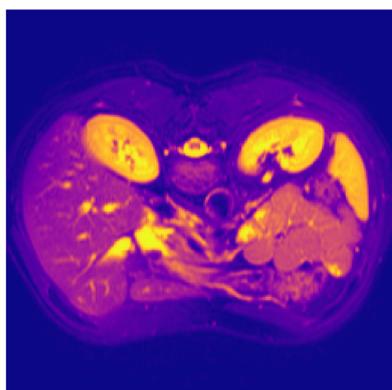
(a) 输入图像

(a) Input image



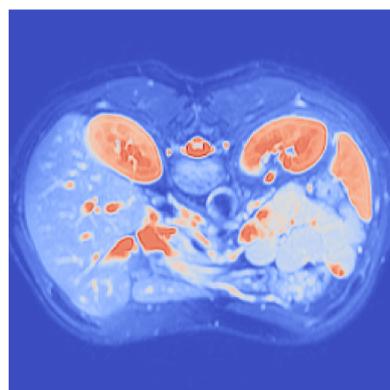
(b) Sobel 算子提取的图片

(b) Image extracted by Sobel operator



(c) 通道增强特征

(c) Channel enhancement feature



(d) 最终增强特征

(d) Final enhancement feature

图 3.8 ECAM 模块医学图像可视化

Figure 3.8 Visualization of medical images with ECAM module

如表 3.7 所示, 基线模型的整体精度相对较低, 特别是在左肾和脾脏等器官上表现欠佳。当仅保留 ECAM (移除 MSMC) 时, 模型在处理器官边缘模糊和对比度较低场景时有了大幅提升, 左肾和脾脏的 Dice 分数明显提高; 而当仅保留 MSMC (移除 ECAM) 时, 模型对多尺度特征的捕捉更为出色, 对右肾和肝脏有一定增益, 但在精细边界复原方面略显不足。进一步地, 完整的 MS-ECANet (同时包含 MSMC 与 ECAM) 在所有器官上均取得了最高的平均 Dice 分数, 展现出两大模块在小样本分割条件下的互补特性。

具体而言, MSMC 提供了多尺度记忆与类别原型的动态匹配, 而 ECAM 则强化了对边缘细节和上下文信息的聚合, 两者协同作用使模型在多种器官和病灶分割中都具备更好的鲁棒性和精度。值得注意的是, 当移除 ECAM 时, 模型在左肾和脾脏上的 Dice 分数下降约 5~8%, 进一步说明边缘上下文聚合模块对器官边界模糊和低对比度的医学图像具有关键作用。

为深入理解 ECAM 模块在医学图像分割中的作用, 我们进行了可视化分析。图 3.8 展示了 ECAM 模块的关键处理流程: 图 3.8a 为原始医学查询图像; 图 3.8b 为 Sobel 算子提取的边缘特征, 突出图像轮廓; 图 3.8c 展示通道注意力增强的语义特征, 强调关键目标区域; 图 3.8d 为边缘与语义特征融合后的最终增强特征  $F_{enhanced}^q$ , 同时保持边界清晰度和内部一致性。

可视化结果表明, ECAM 模块通过有效整合局部边缘信息和全局语义信息, 为分割任务提供了更加鲁棒且精细的特征表示, 尤其在目标边界区域表现突出。

基于本节的实验结果与分析, 本文可以得出以下结论: MS-ECANet 在腹部 MRI 数据集上展现了具有竞争力的性能, 并显著优于 SE-Net 和 Q-Net 等传统方法。消融实验表明, 边缘上下文聚合模块 (ECAM) 对医学图像性能的提升起到了较为重要的作用, 尤其在肾脏、脾脏等低对比度区域表现明显。定性分析进一步验证了 MS-ECANet 在应对医学图像边缘模糊、器官形态差异等挑战方面的有效性, 尤其得益于 ECAM 模块。上述结果为第四章的医学图像专用方法优化提供了必要的实验支持; 后续将在此基础上继续完善网络结构, 以提升模型在其他医学图像数据集上的泛化能力。

### 3.4 本章小结

本章提出了基于多尺度记忆协调 (MSMC) 与边缘上下文聚合 (ECAM) 双重策略的 MS-ECANet, 可在小样本通用图像分割场景下同时兼顾不同尺度目标的识别需求与边缘细节的精确刻画。通过多尺度特征提取与历史原型记忆融合, MSMC 模块有效缓解了单一原型信息不足的问题; ECAM 则结合梯度检测与通道注意力, 进一步强化了目标边缘特征的表达。

在通用数据集的少样本对比实验中, MS-ECANet 相比基准方法取得了在准确度方面有所增强。证明了其在应对多尺度目标与边缘模糊问题方面具备良好的鲁棒性。此外, 第 3.3.5 小节针对 Abdominal-MRI 医学数据集进行的适应性分析表明: MSMC 在医学影像上的性能提升相对有限, 而 ECAM 的边缘增强策略在低对比度、轮廓模糊的器官分割场景下依然有效。因此, 在后续面向医学图像的研究中, 应保留边缘强化的设计思路, 以进一步挖掘本章方法在医学分割领域的潜力。

简言之, MS-ECANet 通过多尺度记忆与边缘上下文增强的双重策略, 显著提升了小样本分割的整体精度与鲁棒性, 并在医学图像场景中表现出良好的迁移能力。本文将在后续章节中针对更复杂的医学场景 (如器官形态变异与噪声干扰) 开展优化, 以更好地满足临床诊断与手术辅助的精细化需求。

## 第4章 基于双提示协同原型与 Mamba 交互的小样本医学图像分割方法

### 4.1 引言

近年来，小样本医学图像分割取得了较大进展，现有研究多以原型网络和元学习为基础，能够在少量支持样本条件下实现医学图像像素级分割。然而，由于医学解剖结构具有较大的个体差异与形态变异，传统静态原型匹配难以刻画精细几何关系；再加上器官轮廓边缘模糊、对比度低，易受噪声干扰而降低分割精度。同时，高分辨率医学影像带来的计算负担，也使现有卷积或注意力机制难以兼顾全局建模与高效推理。

结合第三章的实验启示，本章进一步聚焦图像边缘与背景干扰问题，提出双提示协同原型与 Mamba 交互网络 (Dual-Prompt Collaborative-Prototype Mamba Network, DPCP-MNet)，以提升医学小样本场景下的边缘识别与前后景区分能力。具体而言：(1) 通过边缘感知几何提示与中心化语义提示，分别捕获器官的边界细节与主体语义，构成双重视觉提示，并在此基础上生成包含边界与语义协同信息的原型表示；(2) 在全局交互阶段，引入协同原型 Mamba 交互模块，通过状态空间建模以高效处理高分辨率图像，并借助协同原型融合提升查询特征的适配能力；(3) 采用层次化融合解码器逐级恢复目标细节，进一步强化边缘语义一致性。

为验证方法有效性，本章在 Abdominal-CT、Abdominal-MRI 及 Cardiac-MRI 数据集上开展小样本医学图像分割实验，DPCP-MNet 均获得较为稳定的性能提升。此外，本章为验证小样本模型的现有优势，将模型与 SAM2, Med-SAM2 等分割大模型进行了对比实验，考察了整体分割精度与边缘质量，为后续医学分割模型的选型提供了参考。本章模型整体设计思路如图 4.1 所示。

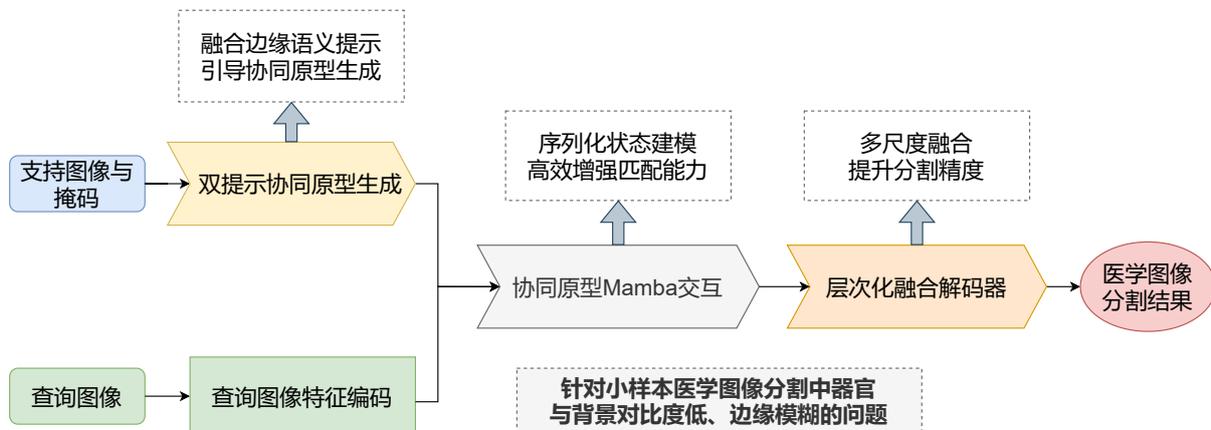


图 4.1 模型整体设计思路图

Figure 4.1 Diagram of model overall design idea

## 4.2 双提示协同原型与 Mamba 交互网络的整体框架

### 4.2.1 模型整体框架

本研究提出的双提示协同原型与 Mamba 交互网络，整体架构如图 4.2 所示。网络沿用双路径结构，以支持图像和查询图像为输入，通过协同原型学习与状态空间交互，实现小样本医学图像的高精度分割。首先，编码器在双分支上提取多尺度空间与语义特征。支持分支进一步融合边缘感知提示与中心化语义提示，随后经全局池化与非线性映射生成协同原型，有效整合边界与主体信息。其次，将协同原型与查询分支的深层特征共同输入 Mamba 交互模块。模块利用线性状态空间建模与双向信息传播机制捕获长程依赖，在较低计算开销下输出增强的查询特征，进一步通过与原型的余弦相似度生成相似度图。最后，解码器采用层次化融合策略，将增强查询特征、相似度图以及从浅层到深层的编码特征逐级融合并恢复分辨率。相似度图作为空间注意力权重，引导网络聚焦目标区域，提升分割精度与边界完整性。

DPCP-MNet 的框架有以下模块组成：边缘感知提示生成器、中心化语义提示生成器、协同原型生成模块、协同原型 Mamba 交互模块、层次化融合解码器。本节将逐一分析其设计细节，最后讨论损失函数的设置。

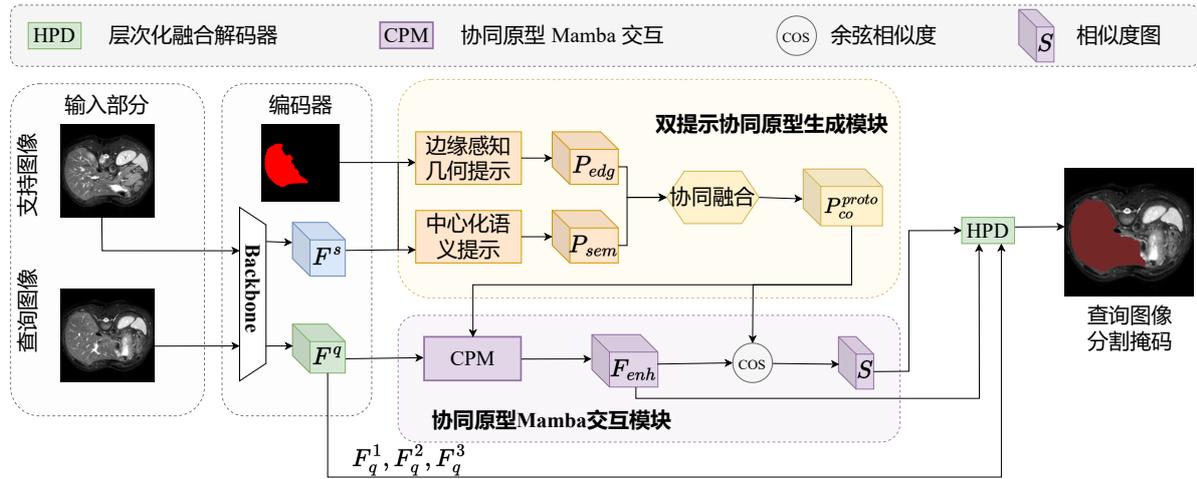


图 4.2 DPCP-MNet 整体框架图

Figure 4.2 The overall framework diagram of DPCP-MNet

### 4.2.2 边缘感知提示生成器

EA-Prompt 旨在强化支持图像中的边缘信息，从而帮助模型更准确地感知目标器官的几何结构。医学图像中的解剖目标通常边界模糊，尤其在小样本场景下，依赖全局语义信息可能会导致边缘细节的丢失。为此，EA-Prompt 提取方向敏感的边缘特征，并结合深层语义信息，引导分割网络专注于目标轮廓，同时有效抑制背景噪声。

为了确保 EA-Prompt 专注于目标器官的边缘信息，网络首先使用支持掩码对骨干网络提取的特征进行 Mask-Guided 过滤。具体而言，令  $F_{s1}$  和  $F_{s3}$  分别表示支持图像的

浅层特征与深层特征，在边缘提取之前，通过掩码  $M_s$  进行空间加权操作，如式 (4-1) 所示：

$$F'_{s1} = F_{s1} \odot M_s, \quad F'_{s3} = F_{s3} \odot M_s \quad (4-1)$$

这一过程有效去除背景干扰，确保 EA-Prompt 仅关注目标区域。在此之后，就正式进入 EA-Prompt 的处理流程，整体流程如图 4.3 所示，包含两个关键步骤：

方向敏感边缘提取。首先，采用 8 方向可微分 Sobel 过滤器计算边缘梯度特征，以提取不同方向的边界信息。对于每个方向  $k$ ，计算，如式 (4-2) 所示：

$$E_k = \text{Sobel}_k(F'_{s1}) \quad (4-2)$$

得到 8 个方向的边缘特征图  $\{E_k\}_{k=1}^8$ ，提升对多样化目标轮廓的感知能力。

多尺度边缘增强。随后，将边缘特征  $\{E_k\}$  与上采样后的深层特征  $\text{Up}(F'_{s3})$  进行拼接，并通过  $3 \times 3$  卷积和 Sigmoid 激活函数生成最终的边缘提示，如式 (4-3) 所示：

$$P_{\text{edge}} = \sigma(W_g * [E_1, \dots, E_8, \text{Up}(F'_{s3})]) \quad (4-3)$$

其中， $W_g$  为可学习卷积核。该设计不仅融合了方向敏感的边缘信息，同时结合了高层语义特征，使生成的边缘提示更加鲁棒，有助于提高分割网络的边界感知能力。

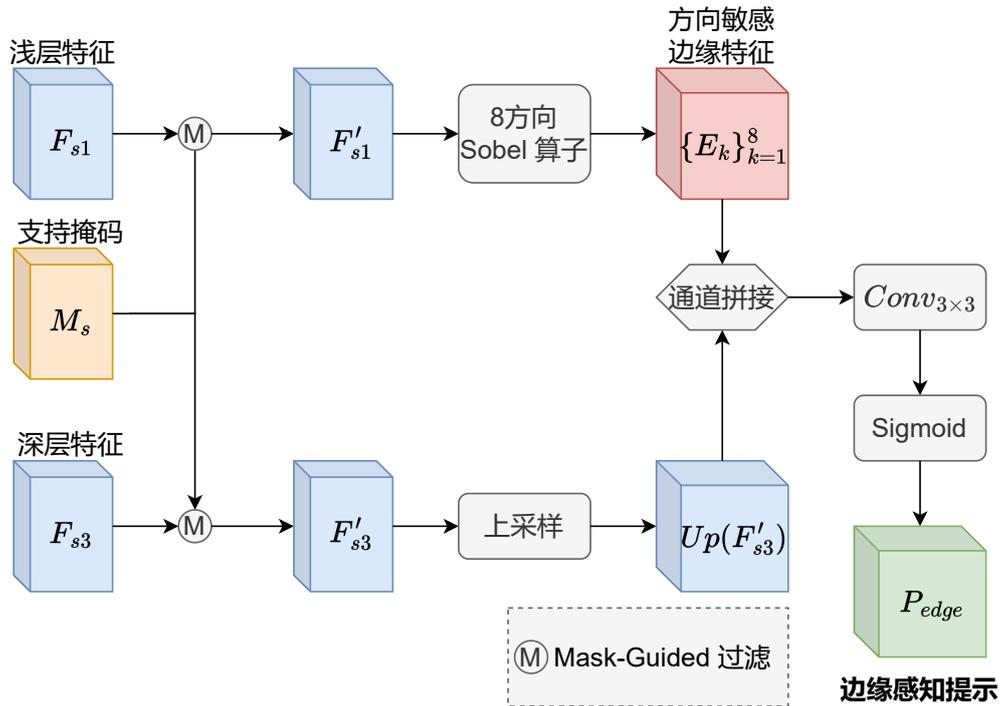


图 4.3 EA-Prompt 模块结构图

Figure 4.3 Architecture of the EA-Prompt module

上采样方式与特征对齐：在边缘特征与深层特征融合时，深层特征  $F'_{s3}$  通过双线性插值上采样至与浅层特征  $F'_{s1}$  相同的空间分辨率，从而确保空间尺寸的一致性，为多尺度特征融合提供基础。

EA-Prompt 设计使分割网络能够同时关注目标区域的边缘细节和整体结构，减少背景噪声的干扰，提升分割精度。通过引入方向敏感性和多尺度信息融合，该方法在小样本场景下增强了对模糊边界的辨识能力，从而提高了分割任务的稳定性与准确性。

### 4.2.3 中心化语义提示生成器

CS-Prompt 旨在增强目标区域的语义信息，使分割模型能够更精准地聚焦于解剖结构的主体区域。医学图像中的解剖目标常常存在对比度低、边界模糊等问题，而传统全局池化方法易受背景信息干扰，导致特征表达不稳定。为克服这一问题，CS-Prompt 通过通道注意力机制增强目标区域的特征表达，并结合 Mask-Guided 过滤机制，有效抑制背景信息干扰，提升模型对目标区域的感知能力。

CS-Prompt 的预处理逻辑与 EA-Prompt 保持一致。首先，通过支持掩码  $M_s$  对骨干网络提取的深层特征  $F_{s3}$  进行 Mask-Guided 过滤，以去除无关区域的特征信息，如式 (4-4) 所示：

$$F'_{s3} = F_{s3} \odot M_s \quad (4-4)$$

这一操作可有效减少背景干扰，使 CS-Prompt 仅聚焦于解剖目标的主体区域。如图 4.4 所示，CS-Prompt 包括两个核心模块：通道注意力计算。首先，对 Mask-Guided 过滤后的特征  $F'_{s3}$  施加通道式全局平均池化 (GAP)，以计算每个通道的重要性权重，如式 (4-5) 所示：

$$w_c = \sigma \left( \text{MLP}(\text{GAP}(F'^{(c)}_{s3})) \right), \quad c = 1, \dots, 512 \quad (4-5)$$

其中，MLP 为通道注意力学习模块， $\sigma(\cdot)$  为 Sigmoid 激活函数，用于归一化通道权重至  $[0, 1]$  区间，从而调整不同通道对目标区域的贡献，值得注意的是，这里使用了轻量化的通道加权策略，并非基于 Transformer 的多头自注意力结构，避免了复杂的注意力运算流程。

中心化语义提示生成。在通道注意力计算完成后，所得权重  $w_c$  作用于原始特征的通道维度，以生成加权特征，如式 (4-6) 所示：

$$F''_{s3} = \sum_{c=1}^{512} w_c F'^{(c)}_{s3} \quad (4-6)$$

通过  $1 \times 1$  卷积进一步压缩通道维度，得到最终的中心化语义提示，如式 (4-7) 所示：

$$P_{\text{sal}} = \sigma(W_s * F''_{s3}) \in \mathbb{R}^{\frac{H}{16} \times \frac{W}{16} \times 1} \quad (4-7)$$

其中， $W_s$  为  $1 \times 1$  卷积核，用于整合通道信息并生成最终的单通道语义提示。

通过通道注意力增强与 Mask-Guided 过滤相结合，CS-Prompt 能有效提升目标区域的语义表达，并减少背景噪声干扰，提高分割精度和模型泛化能力。

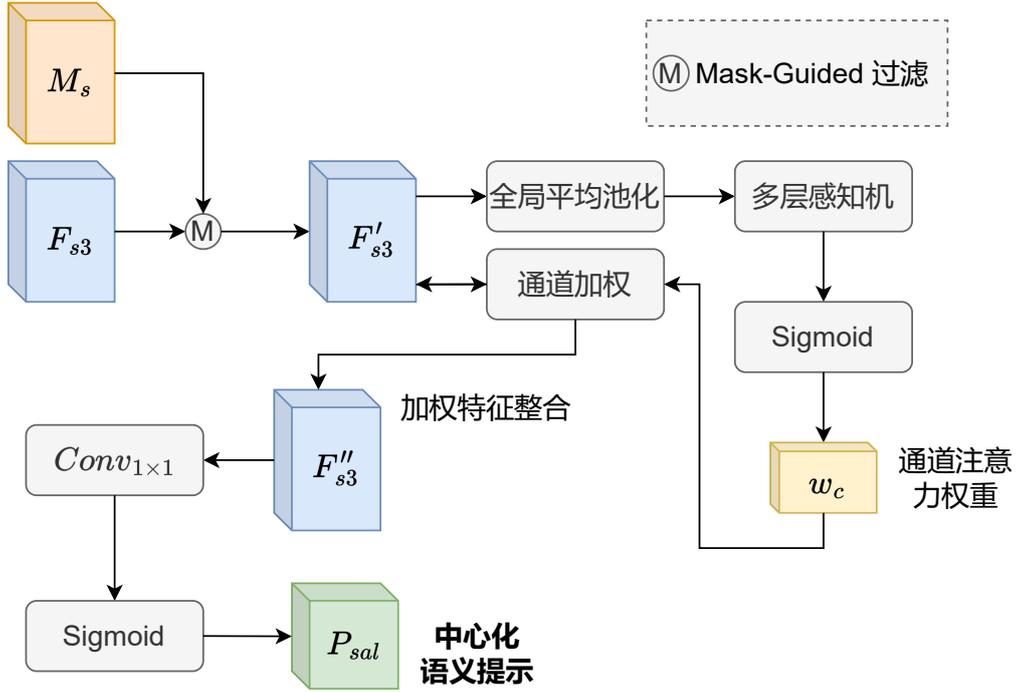


图 4.4 CS-Prompt 模块结构图

Figure 4.4 Architecture of the CS-Prompt module

#### 4.2.4 协同原型生成模块

本模块旨在融合 EA-Prompt 生成的边缘提示  $P_{edge}$  与 CS-Prompt 生成的语义提示  $P_{sal}$ ，并构建统一的协同原型  $P_{co}^{proto}$ ，为后续 CP-Mamba 的状态空间建模提供融合特征。如图 4.5 所示，主要包含以下三个步骤：

首先，对边缘提示  $P_{edge}$  与语义提示  $P_{sal}$  进行特征对齐，以确保二者处于相同表示空间。具体地，分别通过可学习的卷积或线性变换映射，如式 (4-8) 所示：

$$\hat{P}_{edge} = W_e * P_{edge}, \quad \hat{P}_{sal} = W_s * P_{sal} \quad (4-8)$$

其中， $W_e$  和  $W_s$  分别为边缘提示与语义提示的变换参数， $*$  代表卷积操作或逐通道线性变换。积操作通常适合捕捉局部空间相关性，而线性变换更侧重逐通道特征的独立建模。在实际应用中，可根据输入特征的大小、分辨率和计算开销来选择恰当的方式。参数  $W_e$  和  $W_s$  在训练时可随网络的反向传播一起更新，使得网络能够在融合前尽量消除原始特征分布的不一致性，保证后续交互的有效性。

接下来，通过跨提示交互机制，建立边缘与语义提示之间的关联度。可采用点乘或多头注意力机制，计算在不同位置和通道维度上的相似性，如式 (4-9) 所示：

$$\hat{P}'_{edge} = \text{Attn}(\hat{P}_{edge}, \hat{P}_{sal}), \quad \hat{P}'_{sal} = \text{Attn}(\hat{P}_{sal}, \hat{P}_{edge}), \quad (4-9)$$

其中， $\text{Attn}(\cdot, \cdot)$  表示跨特征注意力函数，通过点乘或多头方式来度量两个特征在位置和通道上的相关性。在多头注意力中，不同头可以关注特征中的不同部分，从而挖掘更丰

富的互补信息。该机制使得边缘提示能够更加聚焦于关键轮廓区域，而语义提示则得到增强，以提供更加稳健的全局特征。注意力权重在训练中同样是可学习的，能够根据损失函数的反馈不断优化。

最后，将交互后的特征融合，并通过全局池化和 MLP 提取最终的协同原型，如式 (4-10) 所示：

$$P_{co} = \phi([\hat{P}'_{edge}, \hat{P}'_{sal}]), \quad P_{co}^{proto} = \text{MLP}(\text{GAP}(P_{co})) \quad (4-10)$$

其中， $[\cdot, \cdot]$  表示拼接操作， $\phi(\cdot)$  为卷积或 MLP 映射函数，GAP 表示全局平均池化。得到的协同原型  $P_{co}^{proto}$  结合了边缘和语义信息，可在后续分割任务中为目标形态与语义属性提供一致且稳定的特征表示。

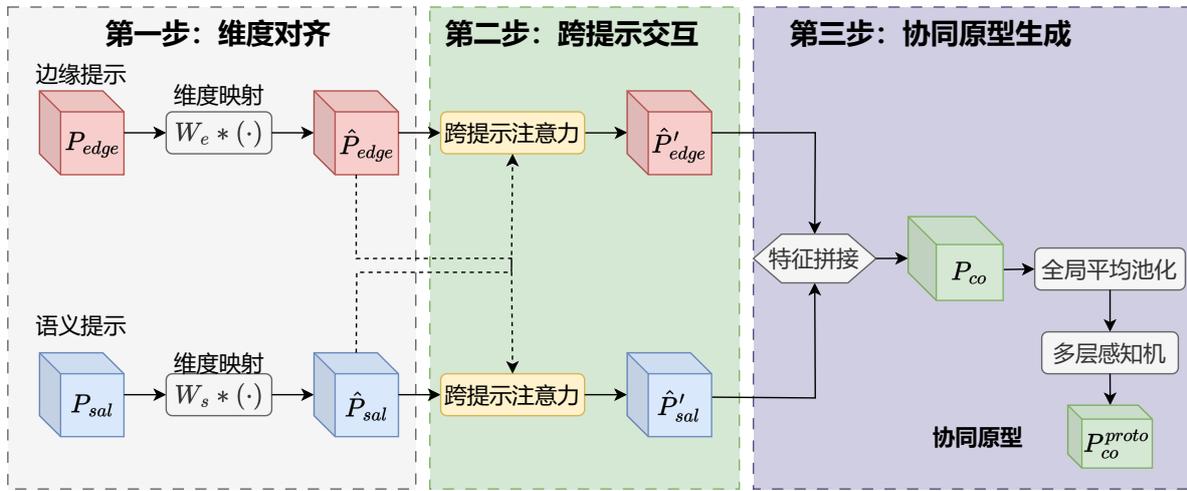


图 4.5 协同原型生成模块结构图

Figure 4.5 The structure diagram of co-prototype generation module

#### 4.2.5 协同原型 Mamba 交互模块

协同原型 Mamba 交互模块是基于近期提出的 Mamba 架构，通过高效的选择性状态空间模型 (SSM) 将协同原型与查询特征深度融合，且保持线性计算复杂度。该模块主要优势包括：(1) 捕捉长程依赖关系，弥补卷积操作的局部感受野限制；(2) 双向递归扫描提升特征表达能力；(3) 相比传统 Transformer 显著降低计算成本。

##### 4.2.5.1 序列化与特征融合

CP-Mamba 首先将协同原型  $P_{co}^{proto}$  与查询图像的深层特征  $F_{q3}$  组织成序列，以完成特征融合：

(1) 特征预处理：对深层特征  $F_{q3} \in \mathbb{R}^{H \times W \times C}$  进行空间展平，转换为序列形式  $\text{Flatten}(F_{q3}) \in \mathbb{R}^{HW \times C}$ ，保留每个空间位置的通道信息。在此过程中，模块通过简单的维度变换保证查询特征的形状与后续的序列操作保持一致，同时预留空间与原型进行拼接。(2) 原型扩展：为使原型与特征维度对齐，将协同原型  $P_{co}^{proto} \in \mathbb{R}^{1 \times C}$  通过复

制扩展为  $\tilde{P}_{co}^{proto} \in \mathbb{R}^{L_p \times C}$ ，其中  $L_p$  为设定的原型序列长度，通常远小于特征序列长度。该扩展过程可以理解为将协同原型在序列方向广播，使其能够与后续每个特征位置进行充分的交互。(3) 序列拼接：将扩展后的原型与展平的特征拼接得到混合序列，如式 (4-11) 所示：

$$X = [\tilde{P}_{co}^{proto}, \text{Flatten}(F_{q3})] \in \mathbb{R}^{(L_p+HW) \times C} \quad (4-11)$$

(4) 位置编码：添加可学习的位置编码  $P_{pos}$ ，增强序列中的位置信息，如式 (4-12) 所示：

$$X_{emb} = X + P_{pos} \quad (4-12)$$

在本模型中，采用可训练的参数矩阵来实现位置编码，与展平后每个序列元素逐元素相加。通过这种方式，网络能够保留并学习特征在原图空间位置上的依赖关系，进一步提升融合特征的表达能力。

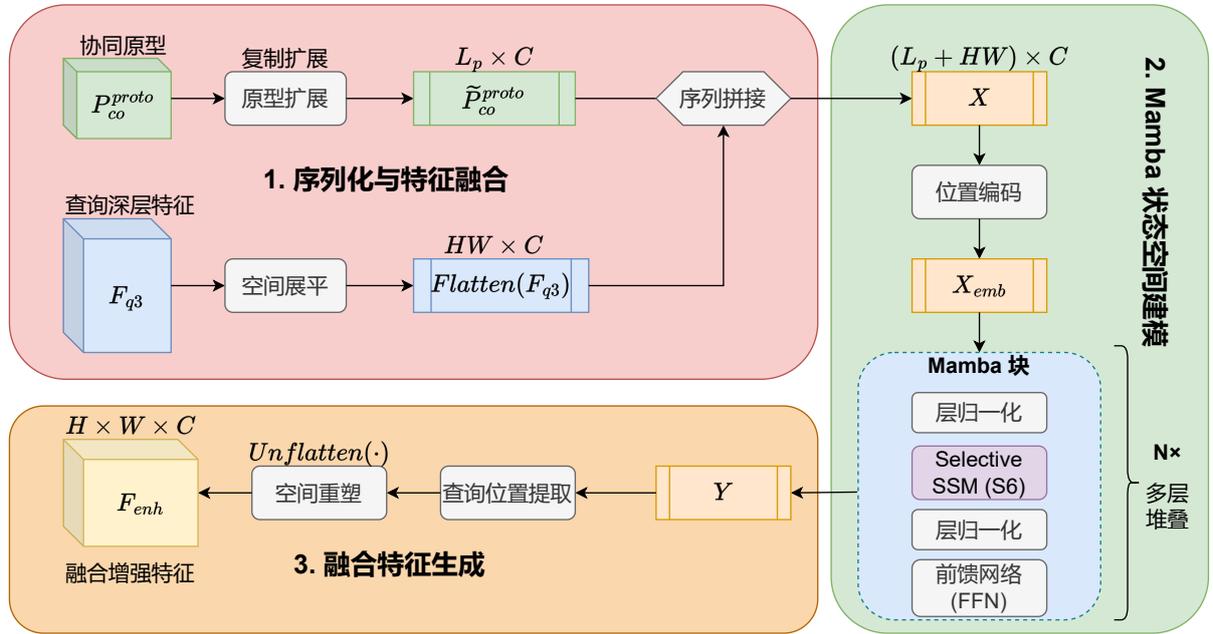


图 4.6 协同原型 Mamba 交互模块

Figure 4.6 The structure diagram of collaborative prototype - mamba interaction module

#### 4.2.5.2 Mamba 状态空间建模

在传统 SSM 框架中，输入-输出关系为连续或离散形式。Mamba 采用参数化方法，使系统矩阵  $A, B$  动态适配输入，并通过选择性机制实现，如式 (4-13) 所示：

$$S6(x) = \sigma(P_{gate}(x)) \odot (P_{value}(x) \star SSM(P_{SSM}(x))) \quad (4-13)$$

其中  $\odot$  为逐元素乘法， $\star$  为 SSM 卷积操作， $\sigma$  为 SiLU 激活函数， $P_{gate}$ 、 $P_{value}$ 、 $P_{SSM}$  均为线性投影。

Mamba 的整体流程与 Transformer 的“注意力 + 前馈”类似，但用状态空间和选择性机制替代多头自注意力，在捕捉远程依赖和减少计算成本方面更具优势。对于低对比

度、模糊边界的医学影像，Mamba 可以更稳定地聚焦关键部位，提高分割边缘的准确度与一致性。

#### 4.2.5.3 融合特征生成

经过 Mamba 处理后，输出序列  $Y$  包含了协同原型与查询特征的交互信息。提取与原查询特征对应的部分并恢复空间结构，如式 (4-14) 所示：

$$F_{\text{enh}} = \text{Unflatten}(Y_{L_p:L_p+HW}) \in \mathbb{R}^{H \times W \times C} \quad (4-14)$$

将增强后的特征  $F_{\text{enh}}$  输入后续解码或细化模块，完成最终分割预测。该增强特征不仅包含查询图像的局部细节，也融合了协同原型的全局语义先验，因此在对比度低、边界模糊的医疗场景中表现出更好的分割精度和一致性。

#### 4.2.6 层次化融合解码器

HP-Decoder 旨在通过逐级融合多尺度特征信息，提高医学图像分割的边界精度与目标完整性。在小样本分割任务中，分割模型通常依赖支持图像提供的结构信息，而仅基于相似度计算直接生成分割掩码，往往导致分割结果过于平滑或细节缺失。为克服这一问题，HP-Decoder 采用多尺度融合策略，在解码过程中逐步恢复高分辨率信息，并结合查询图像特征进行精确重建，从而提升分割质量。

如图 4.7 所示，HP-Decoder 通过逐级融合不同尺度的特征，实现目标边缘与精细结构的高精度恢复。在小样本医学图像分割任务中，直接基于全局相似度计算生成掩码可能会导致细节丢失。为此，HP-Decoder 引入分层融合与原型引导的空间注意力机制，以增强解码过程中对边缘结构的恢复能力。

该解码器以 CP-Mamba 生成的增强特征  $F_{\text{enh}}$ 、协同原型  $P_{\text{co}}^{\text{proto}}$  以及查询特征  $F_q$  与其相似度图  $S$  作为输入。解码过程分三阶段（例如  $\times 2, \times 2, \times 4$ ）逐步上采样，并在每阶段融合编码器的相应尺度特征以补充局部空间细节。设第  $i$  级解码特征为  $F_{\text{dec}}^{(i)}$ ，对应的编码器特征为  $F_q^{(i)}$ 。为对齐不同尺度，先用双线性插值统一分辨率，随后采用拼接或加权相加进行融合，如式 (4-15) 所示：

$$F_{\text{dec}}^{(i)} = \phi([F_{\text{enh}}^{(i)}, F_q^{(i)}, S^{(i)}]) \quad (4-15)$$

其中， $\phi(\cdot)$  代表卷积与激活函数组成的特征融合操作， $S^{(i)}$  为与当前尺度相对应的相似度图。对于融合方式，采用加权相加的方式将每一层特征按其重要性加权，以避免过度平滑的效果。

为进一步关注目标区域，HP-Decoder 利用相似度图  $S$  生成空间注意力权重，强化对关键部位的表达，如式 (4-16) 所示：

$$F_{\text{dec}}^{(i)'} = F_{\text{dec}}^{(i)} \odot \sigma(S^{(i)}) \quad (4-16)$$

其中， $\sigma(\cdot)$  为 Sigmoid 激活函数， $\odot$  表示逐元素乘法。该步骤能够自适应地抑制背景噪

声,同时增强解剖结构的关键特征。相似度图  $S$  是通过计算查询特征  $F_q$  和支持特征  $F_{enh}$  之间的余弦相似度获得的,该相似度图能够为解码器提供更加精确的目标区域定位。

经过多尺度融合与空间注意力增强后,HP-Decoder 通过卷积层生成最终的分割掩码,如式 (4-17) 所示:

$$M_q = \text{Conv}_{\text{out}}(F_{\text{dec}}^{\text{final}}) \quad (4-17)$$

与直接基于全局相似度生成掩码相比,HP-Decoder 的层次化策略更能保留细节并优化边缘轮廓。在不同模态(如 MRI、CT)的实验结果也表明,该解码器具有良好的适应性和泛化能力,为小样本医学图像分割提供了更稳定的高精度输出。

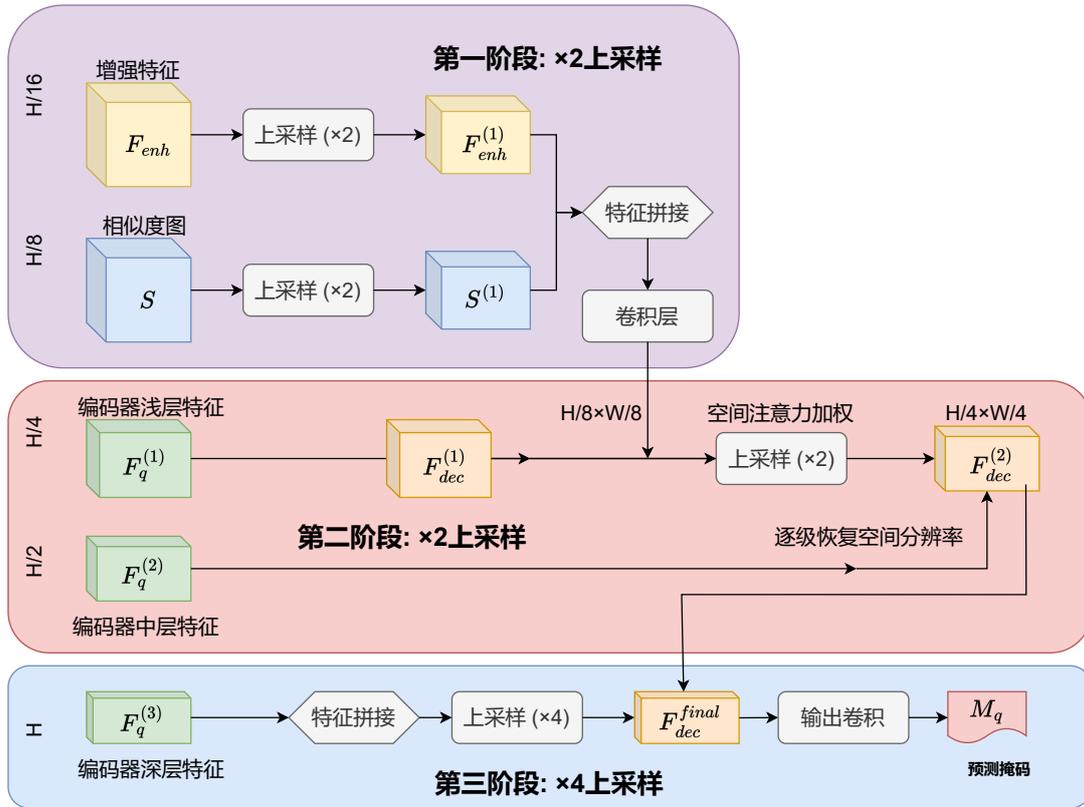


图 4.7 层次化融合解码器结构图

Figure 4.7 The structure diagram of hierarchical fusion decoder

#### 4.2.7 DPCP-MNet 损失函数

针对小样本医学图像分割,DPCP-MNet 结合双提示生成协同原型与协同原型 Mamba 交互的设计思路,并考虑原型融合等因素,设计了如下综合损失函数,如式 (4-18) 所示:

$$\mathcal{L}_{\text{DPM}} = \alpha \cdot \mathcal{L}_{\text{seg}} + \beta \cdot \mathcal{L}_{\text{proto}}, \quad (4-18)$$

具体而言,各子损失及其作用如下所示:

- $\mathcal{L}_{\text{seg}}$ : 主分割损失,可选用 Dice 或 Focal 等医学图像常用度量,用于监督分割主干结果;

- $\mathcal{L}_{\text{proto}}$ : 协同原型一致性损失, 用以约束边缘感知提示 (EA-Prompt) 与中心化语义提示 (CS-Prompt) 融合后的原型在跨病例中的稳定性;
- $\alpha, \beta$ : 相应子损失的平衡系数, 可针对医学图像中的边缘模糊、噪声多等问题进行调整。经测试选择系数为:  $\alpha = 1.0, \beta = 0.35$ 。

表 4.1 展示了 DPCP-MNet 在腹部 MRI 数据集 setting-2 设置下, 不同损失权重组合的性能表现。从结果可以看出, 当  $\beta = 0.35$  时, 模型的 Dice 指标达到了最佳值 73.81%。这一组合的表现明显优于其他组合。

此外, 随着  $\beta$  值的增加, 模型性能在一定范围内提高 (例如  $\beta = 0.4$  时为 73.26%), 但在  $\beta$  值过大时, 性能有所下降, 如  $\beta = 0.6$  时仅为 69.47%。这表明, 过高的原型损失权重可能会导致过拟合或其他负面影响。

因此, 合理的损失权重调整在医学图像分割中至关重要, 能够有效平衡不同损失项, 从而优化模型的表现。

表 4.1 不同权重组合对腹部 MRI 数据集分割性能的影响 (Dice, %)

Table 4.1 Effect of different weight combinations on segmentation performance of abdominal MRI dataset(Dice, %)

实验编号	损失权重设置 ( $\alpha, \beta$ )	Mean (%)
1	(1.0, 0.1)	70.29
2	(1.0, 0.2)	71.93
3	<b>(1.0, 0.35)</b>	<b>73.81</b>
4	(1.0, 0.3)	73.58
5	(0.8, 0.35)	71.95
6	(1.0, 0.4)	73.26
7	(1.0, 0.5)	71.04
8	(1.0, 0.6)	69.47

## 4.3 相关实验及结果

### 4.3.1 数据集

为全面评估所提出的 DPCP-MNet 在小样本医学图像分割任务中的有效性, 本研究选取了三组涵盖不同模态的公开医学影像数据集进行实验验证, 包括腹部 CT、腹部 MRI 及心脏 MRI 数据。这些数据集在解剖结构的多样性及影像模态的异质性方面具有代表性, 为模型在低资源环境下的泛化能力提供了充分的测试基础。

Abdominal-CT<sup>[86]</sup> 数据集来源于 MICCAI 2015 Multi-Atlas 腹部器官分割挑战赛, 包含 30 例三维腹部 CT 扫描数据, 标注涵盖肝脏、左肾、右肾及脾脏四类关键器官。Abdominal-MRI<sup>[81]</sup> 数据集选自 ISBI 2019 CHAOS 挑战赛, 包含 20 例三维 T2-SPIR 序列的腹部 MRI 扫描数据, 每例平均包含 36 层切片, 标注类别与 Abdominal-CT 数据集

保持一致。Cardiac-MRI<sup>[87]</sup> 数据集则取自 MICCAI 2019 多序列心脏 MRI 分割挑战赛的 bSSFP 序列，包含 35 例三维心脏 MRI 扫描，每例平均包含 13 层短轴切片，重点关注心肌及心室结构的分割精度。

为确保实验的公平性，并适配小样本学习的特点，本研究严格遵循 Q-Net<sup>[55]</sup> 等方法的标准化预处理流程。首先，对影像直方图进行顶部 0.5% 的强度截断，以减少极端噪声对模型学习的干扰。随后，根据不同解剖部位选择标准切面：腹部数据集采用轴向切面，心脏数据集采用短轴切面，并在保持原始分辨率的基础上进行重采样。在数据输入层面，单通道医学图像在通道维度复制三次，以适配主流深度学习网络的 RGB 输入格式，最终统一裁剪至  $256 \times 256$  像素。该预处理策略既能保留影像的关键解剖信息，又能确保不同数据集之间的尺度一致性，为模型的跨数据集评测提供了可靠的基准。

本研究将上述数据集的三维影像转换为二维切片进行处理，这不仅有助于缓解小样本条件下的数据稀缺问题，同时也能检验模型在跨切片解剖结构建模方面的能力。实验过程中，将严格按照小样本学习协议构建支持集与查询集，并通过随机采样确保模型在有限标注条件下的泛化能力能够得到客观评估。

#### 4.3.2 实验设置

为全面验证 DPCP-MNet 在小样本医学图像分割中的性能，本节从模型架构、训练策略、数据预处理及损失函数设计等方面系统阐述实验配置，确保实验过程的可复现性与结果的科学性。DPCP-MNet 以 ResNet-101 为双分支编码器 (DS-Encoder) 的主干网络，其权重通过 MS-COCO 数据集预训练初始化。输入图像尺寸统一为  $3 \times 256 \times 256$ ，编码器输出多尺度特征图 (分辨率  $256 \times 256$ ，通道数 32)。模型训练采用随机梯度下降 (SGD) 优化器，初始学习率为  $1 \times 10^{-3}$ ，动量参数为 0.9，权重衰减系数为  $1 \times 10^{-4}$ 。学习率每 1000 次迭代基于验证集性能动态调整，训练过程遵循 1-way 1-shot 范式，共进行 30k 次迭代。实验基于 PyTorch 框架，硬件平台为 NVIDIA RTX 4090 GPU。

实验采用 5 折交叉验证，每类数据集均分为 5 份，每轮训练选取 1 份作为测试集。测试阶段随机选择 1 例扫描作为支持样本，其余作为查询样本。实验设置分为两种模式：

- **Setting 1:** 允许目标类在训练数据中以背景形式存在，适用于腹部 MRI/CT 及前列腺 MRI
- **Setting 2:** 严格排除含目标类的训练切片，仅用于腹部 MRI/CT。心脏 MRI 因解剖包裹性仅采用 Setting 1

#### 4.3.3 对比实验与分析

为系统评估 DPCP-MNet 的有效性，本文将其与 SE-Net、ALPNet、SSL-PANet、SSL-ALPNet、Q-Net、ADNet、AAS-DCL、SR&CL、CRAPNet、PAMI 以及最新的 DGAPNet 等方法进行了对比。其中，SSL-ALPNet 和 SSL-PANet 采用超像素自监督策略以缓解标注缺失，Q-Net 与 ADNet 则借助三维超体素生成伪标签扩展训练数据；其余方法均基于

公开代码或作者提供的权重实现。本实验在腹部数据集上严格按照原论文设置复现了 Q-Net 的基准，以保证结果的公平性。模型定量结果如表 4.2 表 4.4 表 4.3 所示，可视化示例参见图 4.8。

表 4.2 腹部数据集 Setting-1 设置下各模型结果对比 (Dice, %)

Table 4.2 Comparison of the results of each model on abdominal datasets under Setting-1 (Dice, %)

模型	Abdominal-CT					Abdominal-MRI				
	Liver	RK	LK	Spleen	Mean	Liver	RK	LK	Spleen	Mean
SE-Net <sup>[49]</sup>	35.42	12.51	24.42	43.66	29.00	29.02	47.96	45.78	47.30	42.51
ALPNet <sup>[82]</sup>	65.07	31.32	29.12	41.00	41.63	62.35	48.42	44.73	49.61	51.28
SSL-PANet <sup>[82]</sup>	60.86	50.42	56.52	55.72	57.88	71.73	60.81	58.83	61.32	63.17
SSL-ALPNet <sup>[82]</sup>	78.29	71.81	72.36	70.96	73.35	76.10	85.18	81.92	72.18	78.84
ADNet <sup>[54]</sup>	77.24	71.13	63.48	78.02	72.97	82.11	85.08	73.86	72.29	78.51
AAS-DCL <sup>[52]</sup>	78.04	73.19	74.58	73.20	74.52	72.33	86.11	80.37	76.24	78.76
SR&CL <sup>[83]</sup>	76.06	71.22	73.45	73.41	73.53	73.94	76.01	80.87	75.60	77.00
Q-Net <sup>[55]</sup>	77.08	71.87	76.89	76.31	75.54	81.74	87.98	78.36	75.99	81.02
CRAPNet <sup>[56]</sup>	75.41	74.18	74.69	70.37	73.66	76.46	86.15	84.95	74.32	79.79
PAMI <sup>[84]</sup>	<b>81.32</b>	<b>80.57</b>	76.52	72.38	77.69	82.59	88.73	81.83	76.37	82.38
DGAPNet <sup>[85]</sup>	65.59	79.56	82.67	<b>83.28</b>	77.77	81.31	86.99	<b>85.84</b>	<b>80.64</b>	83.70
DPCP-MNet (Ours)	74.89	76.43	<b>83.24</b>	80.95	<b>77.88</b>	<b>82.67</b>	<b>88.46</b>	85.51	78.89	<b>83.88</b>

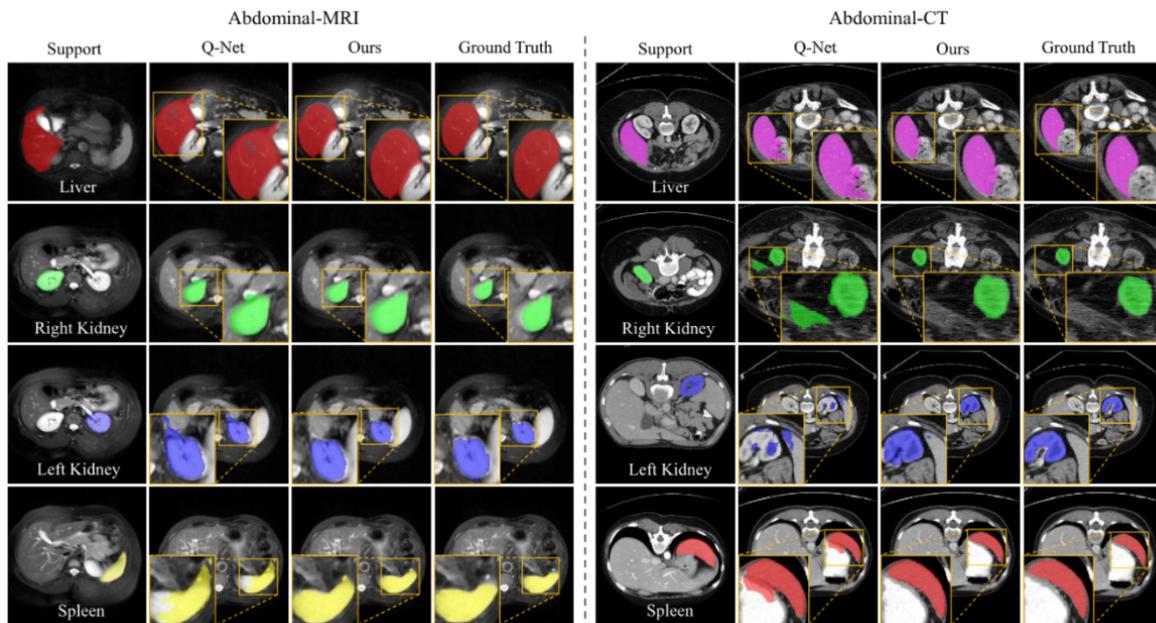


图 4.8 腹部 MRI 数据集的模型分割结果

Figure 4.8 Segmentation results of the model of the abdominal dataset

如表 4.2 所示，DPCP-MNet 在两种模态均取得了最高均值：腹部 CT 达 77.88%，腹

部MRI达83.88%。与当前最优模型DGAPNet相比,平均分别提升0.19%和1.50%。就器官层面而言,左肾(83.24%)与脾脏(80.95%)均达到当前最优水准,较SSL-ALPNet分别提升10.88%与9.99%。图4.8展示了该设置下的部分分割结果。此外,从器官分布来看,DPCP-MNet在较大解剖结构(肝脏)与较小形变脏器(肾脏)上均表现出较高精度,对不同形态器官具有较强适应性;这表明该模型在腹部场景中兼顾了局部细节刻画与全局判别能力,提升了对形变器官的鲁棒性。

表4.3 心脏MRI数据集Setting-1设置下各模型分割结果(Dice, %)

Table 4.3 Comparison of the results of each model on Cardiac MRI datasets under Setting-1 (Dice, %)

模型	Cardiac-MRI			
	LV-MYO	LV-BP	RV	Mean
SE-Net <sup>[49]</sup>	25.18	58.04	12.86	32.02
ALPNet <sup>[82]</sup>	49.53	73.08	58.50	60.34
SSL-PANet <sup>[82]</sup>	46.79	70.43	69.52	62.25
SSL-ALPNet <sup>[82]</sup>	66.74	83.99	79.96	76.90
ADNet <sup>[54]</sup>	62.43	81.53	77.31	75.76
AAS-DCL <sup>[52]</sup>	64.03	85.21	79.13	76.12
SR&CL <sup>[83]</sup>	65.83	84.72	78.41	76.32
Q-Net <sup>[55]</sup>	65.92	90.25	78.19	78.12
CRAPNet <sup>[56]</sup>	65.48	83.02	78.27	75.59
PAMI <sup>[84]</sup>	66.82	89.57	<b>80.17</b>	78.85
DGAPNet <sup>[85]</sup>	<b>67.62</b>	89.82	80.09	79.18
<b>DPCP-MNet (Ours)</b>	67.58	<b>90.65</b>	79.94	<b>79.39</b>

表4.4 腹部数据集Setting-2设置下各模型分割结果(Dice, %)

Table 4.4 Comparison of the results of each model on abdominal datasets under Setting-2 (Dice, %)

模型	Abdominal-CT					Abdominal-MRI				
	Liver	RK	LK	Spleen	Mean	Liver	RK	LK	Spleen	Mean
SE-Net <sup>[49]</sup>	0.27	14.34	32.83	0.23	11.91	27.43	61.32	62.11	51.80	50.66
ALPNet <sup>[82]</sup>	41.71	11.61	13.90	16.39	20.90	37.32	58.99	53.21	52.18	50.43
SSL-PANet <sup>[82]</sup>	61.71	34.69	37.58	43.73	44.42	64.99	47.95	47.71	58.73	54.85
SSL-ALPNet <sup>[82]</sup>	73.65	54.82	63.34	60.25	63.02	73.05	<b>78.39</b>	73.63	67.02	73.02
ADNet <sup>[54]</sup>	70.63	40.52	48.41	50.97	52.63	77.03	56.68	59.64	59.44	63.20
Q-Net <sup>[55]</sup>	<b>77.07</b>	51.47	65.67	63.38	64.40	<b>78.25</b>	65.94	64.81	65.37	68.59
DGAPNet <sup>[85]</sup>	65.56	<b>68.06</b>	<b>74.10</b>	65.91	68.41	75.93	75.96	73.76	69.21	73.72
<b>DPCP-MNet (Ours)</b>	74.64	62.61	72.25	<b>66.42</b>	<b>68.98</b>	75.54	75.18	<b>74.15</b>	<b>70.36</b>	<b>73.81</b>

如表 4.3 所示, DPCP-MNet 在心脏 MRI 的平均 Dice 达到 79.39%, 略高于 DGAPNet (79.18%) 和 PAMI (78.85%)。其中, LV-BP 分割精度达到 90.65%, 超越了现有最佳结果。

在更具挑战性的 Setting-2 (如表 4.4) 中, DPCP-MNet 继续领先: 腹部 CT 均值为 68.98%, 比 DGAPNet 提升 0.57%; 腹部 MRI 均值为 73.81%, 相比 SSL-ALPNet 高出 0.79%。左肾与脾脏依旧是提升最明显的器官, 分别达到 72.25% 和 70.36%。

DPCP-MNet 在三套评估中均取得最高均值, 平均相对提升约均在一个百分点以上。无论在 CT/MRI 模态还是低标注的 Setting-2 场景下, 模型均能保持稳定优势, 表明其在器官形态变化与标注稀缺条件下依旧具有良好的鲁棒性。

#### 4.3.4 消融实验

以下所有消融实验均基于 Abdominal-MRI 数据集, 并在 Setting2 的实验配置下进行, 以确保实验结果的可比性和稳定性。

##### 4.3.4.1 双提示策略、协同原型生成与 CP-Mamba 的协同作用

本实验主要研究 DPCP-MNet 各关键模块对模型性能的作用, 重点分析双提示策略、协同原型生成以及 CP-Mamba 的交互对最终分割性能的影响。为评估这些组件在增强语义一致性与优化全局匹配方面的独立贡献, 本文逐步消融不同模块并记录其对分割精度的影响。

表 4.5 腹部 MRI 数据集的各模块消融实验结果 (Dice, %)

Table 4.5 Ablation results of each module of abdominal MRI dataset(Dice, %)

模块	Abdominal-MRI				
	LK	RK	Spl.	Li.	Mean
仅 GAP 生成原型 (无协同提示交互)	71.12	70.85	69.02	65.34	69.58
仅 EA-Prompt (无 CS-Prompt)	72.05	71.34	70.11	66.91	70.60
仅 CS-Prompt (无 EA-Prompt)	70.86	69.93	68.34	64.72	68.96
仅使用 CNN 进行融合 (无 CP-Mamba)	71.45	70.12	69.21	65.78	69.64
仅使用直接相似度匹配 (无 CP-Mamba)	69.33	68.47	67.12	63.21	67.53
完整 DPCP-MNet	<b>73.94</b>	<b>73.58</b>	<b>72.55</b>	<b>68.76</b>	<b>72.21</b>

实验以完整的 DPCP-MNet 作为基准, 并设计以下五种对照配置: (1) 移除协同原型生成模块。不进行提示交互, 直接对 EA-Prompt 与 CS-Prompt 进行全局池化 (GAP) 以独立生成原型。(2) 移除 CS-Prompt。仅保留 EA-Prompt 计算原型, 用于评估边缘信息在匹配过程中的作用。(3) 移除 EA-Prompt。仅保留 CS-Prompt 计算原型, 用于评估语义信息对分割性能的影响。(4) 移除 CP-Mamba。使用  $1 \times 1$  卷积进行融合, 用于分析

全局建模对特征交互的影响。(5) 同时移除 CP-Mamba 并采用直接相似度匹配。仅根据查询特征与原型的余弦相似度进行分割，以考察 CP-Mamba 对查询特征优化的作用。

如表 4.5 所示，当移除协同原型生成模块后，所有解剖结构的分割精度均有所下降，整体均值降低 2.63%，表明协同提示交互在增强原型表达稳定性与减少匹配误差方面发挥了关键作用。进一步分析显示，当缺少 EA-Prompt 或 CS-Prompt 时，分割精度显著下降，尤其在缺乏边缘信息 (EA-Prompt) 时，Dice 分数呈明显退化。然而，仅使用语义提示 (CS-Prompt) 时，结果也不如全局池化方式所生成的原型。在以普通 CNN 替换 CP-Mamba 进行特征融合时，分割精度下降 2.57%，说明全局建模能更有效地提升特征适配性。若完全移除 CP-Mamba 并仅采用直接相似度匹配，则降幅更为明显，Dice 分数减少 4.68%，突出 CP-Mamba 在优化查询特征过程中的关键地位。

综上所述，双提示策略、协同原型生成与 CP-Mamba 的交互建模在小样本医学图像分割中展现出显著的协同增益。移除任一模块均会导致匹配不稳定并降低特征适配性，从而影响最终分割精度。上述结果进一步表明 DPCP-MNet 在低资源医学影像分割场景下具有较高可行性，也为后续改进方向提供了理论支持。

为了进一步阐释双提示策略的作用机制，本节有必要通过可视化的方式展示边缘提示和语义提示的生成过程及其在分割任务中的贡献。

图 4.9 展示本章提出的双重提示生成机制，该机制以边缘提示和语义提示协同引导分割过程，从而提升整体精度与鲁棒性。如图 4.9a 所示，原始输入图像经掩码处理后保留了目标区域的纹理与强度信息。图 4.9b 展示经多方向 Sobel 算子提取并加权增强的边缘特征，引导模型聚焦目标轮廓并抑制内部纹理干扰。图 4.9c 给出了融合边缘与深层特征后，通过卷积运算与 Sigmoid 激活生成的边缘提示，用以帮助模型更精确地定位目标边界。图 4.9d 所示的语义提示融合通道注意力与距离变换，在目标中心呈现更高响应，引导模型关注核心区域。该双重提示策略充分发挥边缘与语义信息的互补优势：边缘提示聚焦目标轮廓，提升分割细节；语义提示强调目标内部，提高结果一致性。二者协同作用在边界复杂或对比度较低的场景下尤为有效，能够有效提升分割效果。

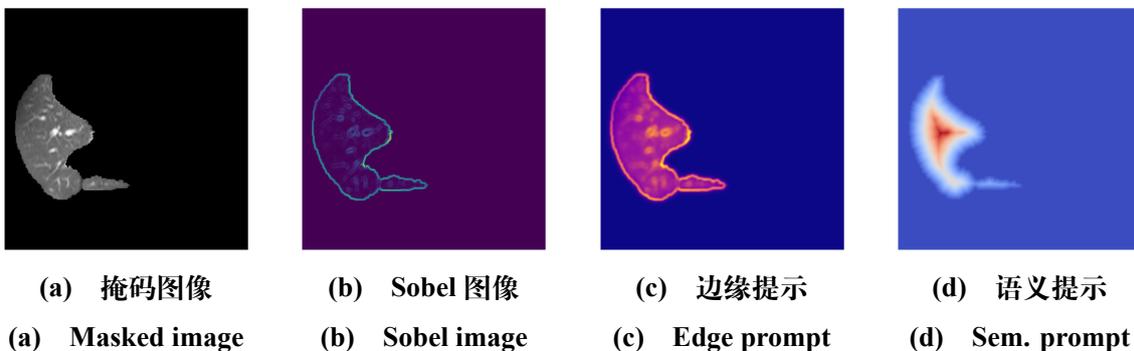


图 4.9 腹部 MRI 数据集的双提示策略可视化

Figure 4.9 Dual-prompt strategy visualization of abdominal MRI dataset

#### 4.3.4.2 HP-Decoder 的必要性

本实验旨在探讨层次化融合解码器 (HP-Decoder) 在多尺度特征恢复、边界优化以及目标完整性方面的作用。传统的小样本分割方法往往基于原型相似度直接生成掩码, 在空间信息利用上存在不足; 而 HP-Decoder 通过融合编码器不同层级的特征, 能够增强目标区域的边界清晰度和局部细节保留能力, 从而在小样本环境下实现更稳健的分割表现。本文通过对 HP-Decoder 及其核心机制的逐步消融, 系统评估其对分割质量的影响。

本研究以完整的 DPCP-MNet 为基准, 并采用以下三种对比方案: (1) 移除 HP-Decoder: 直接基于相似度计算生成掩码, 跳过层次化解码过程, 以评估 HP-Decoder 在恢复局部空间细节与边界特征方面的作用; (2) 仅采用 U-Net 风格跳跃连接: 使用一般的 U-Net 层次化解码方式; (3) 移除层次化融合: 仅在最终阶段进行单层上采样, 用以探讨多尺度特征融合对分割精度的影响。

在评价指标方面, 实验主要采用 Dice 分数衡量对左肾、右肾、脾、肝等解剖结构的整体分割质量。

表 4.6 层次化融合解码器消融实验结果 (Dice, %)

Table 4.6 Results of the ablation experiment of hierarchical fusion decoder (Dice, %)

模块	Abdominal-MRI				
	LK	RK	Spl.	Li.	Mean
直接相似度计算掩码 (无解码器)	67.92	66.88	64.73	60.42	64.99
U-Net 跳跃连接 (普通解码)	71.58	70.76	69.15	65.12	69.15
仅使用单层上采样 (无层次融合)	72.33	71.95	70.02	66.48	70.69
完整 DPCP-MNet (HP-Decoder)	<b>73.94</b>	<b>73.58</b>	<b>72.55</b>	<b>68.76</b>	<b>72.21</b>

如表 4.6 所示, 移除 HP-Decoder 并直接基于相似度计算掩码时, 各解剖结构的 Dice 分数均明显降低, 整体均值下降 7.22%, 表明 HP-Decoder 在细节恢复与边界精确度方面发挥了关键作用。进一步分析显示, 在采用 U-Net 风格跳跃连接后, 尽管分割性能有所提升, 但整体均值依然比完整 HP-Decoder 低 3.06%。这说明跳跃连接虽能补充部分局部细节, 但因缺乏专门的层次化解码设计, 目标还原能力仍受限制, 导致分割精度不及完整方案。

此外, 仅在最后阶段进行单层上采样 (移除层次融合) 时, 整体均值下降 1.52%。这一结果进一步证明多尺度特征融合对提升边界精准度的重要性: 分阶段融合高低层次特征, 可更充分地保留空间细节并维持语义一致性。

综上所述, 实验结果充分验证了 HP-Decoder 在小样本医学图像分割任务中的有效性。通过层次化特征融合与原型引导的注意力机制, HP-Decoder 在恢复目标区域高分辨率特征的同时, 显著增强了边界完整性。与传统跳跃连接结构或直接相似度计算方式

相比，HP-Decoder 在分割稳定性与质量方面均展现出更优表现，也为低资源条件下的医学影像分割提供了更具竞争力的解码器设计思路。

### 4.3.5 方法与分割大模型的性能对比

近年来，医学成像与人工智能技术迅速发展，为医学图像分割提供了有力支持。然而，医疗影像获取与标注的高成本，以及罕见病或极度受限的临床场景（仅能获得少量有效标注）往往让传统依赖海量数据的模型难以奏效。由此催生的小样本医学图像分割方法，能够在仅有极少量标注资源的前提下，快速构建高效率、低负担的分割模型，往往只需要几张带注释的医学图像即可取得令人满意的分割效果。

与此同时，计算机视觉领域涌现出以 Segment Anything Model (SAM) 和 MedSAM 为代表的大规模预训练分割模型，这些模型能够在自然图像或混合领域上学习到通用特征，从而部分减轻对大规模标注数据的依赖。然而，在不同的疾病情况下，病灶形态模糊，位置多变，且医学影像与常规的自然图像分布存在较大差异，导致这些预训练大模型在极少量标注时可能出现边界精细度不足或推理速度难以满足实时需求等问题。相比之下，专门为小样本情形设计的模型往往能在训练初期就快速收敛，对病变区域做出更精细的边界预测。

基于以上背景，本节将深入探讨当模型只有极少量标注（如罕见病场景）可以训练时，通用分割大模型（如 SAM2、MedSAM2）与小样本专用分割模型（DPCP-MNet）在医学图像分割中的具体差异及优缺点。

#### 4.3.5.1 实验设计

##### (1) 实验数据与设置

本实验主要使用的数据集为：Abdominal-MRI（选自 ISBI 2019 CHAOS 挑战赛）。并在 Ubuntu 20.04、PyTorch 1.12 的环境下，于 NVIDIA RTX 4090（24GB 显存）平台上完成全部实验。为保持与其他研究一致，数据集按 4:1 的比例划分训练集与验证集，并使用官方默认测试集。所有图像在输入模型前统一进行重采样、归一化等预处理操作，具体细节可参照本研究第 4.3.1 小节。

为评估模型在极端小样本场景下的适应能力，本研究在训练和验证阶段分别选取 1、3、5 个带注释样本进行对比。在每一种样本数设置下，均进行了 3 次独立重复实验，并取平均性能与标准差作为最终指标，以减少偶然性。所有对比模型均使用相同的随机种子以确保可重复性。

##### (2) 对比方法与实施细节

本研究对比了两类模型：小样本医学图像分割模型（DPCP-MNet）与分割大模型（SAM2、Med-SAM2）。为保证公平性，三者在同一数据集和实验设置下进行训练与测试，主要差异体现在网络结构及训练策略上。具体训练策略如下：

- 通用分割大模型（SAM2、Med-SAM2）。由于原版的 SAM2 与 Med-SAM2 并未在医学数据上进行大规模预训练，为减少医学域分布差异带来的影响，本研究先用

Abdominal-MRI 对其进行初步微调，使模型熟悉 MRI 图像的整体灰度和组织结构特征。随后，在极少量标注（1、3、5 张）条件下再进行二次微调，与 DPCP-MNet 使用相同数量的带注释数据，确保可比性。

- 小样本医学图像分割模型。DPCP-MNet 直接在相同的 Abdominal-MRI 数据集上，采用标准的小样本医学图像训练流程进行端到端训练，不包含额外的预训练环节。

表 4.7 统计了实验的主要参数设置，包括学习率、优化器与训练轮数等。可以看出，DPCP-MNet 与 SAM2 / Med-SAM2 的差异主要体现在网络结构和训练策略上，其他实验条件保持一致。

关于训练轮数，本研究在初步实验中通过观察验证集上的 Dice 系数及 HD95 随训练轮数变化的曲线，发现对于 DPCP-MNet 来说，在 80~100 轮之间还存在一定的性能提升空间；而对于已预训练的大模型（SAM2 / Med-SAM2）来说，在约 50 轮时就趋于收敛。

表 4.7 主要实验参数

Table 4.7 Main Experimental Parameters

参数	DPCP-MNet	SAM2	Med-SAM2
学习率	0.001	0.0005	0.0005
训练轮数	100	50	50
优化器	AdamW	AdamW	AdamW
训练方式	端到端	预训练微调	预训练微调

### (3) 实验指标与资源开销说明

本实验选用的指标为 Dice 系数与 HD95（95% Hausdorff 距离）。Dice 用于衡量预测区域与真实区域的重叠度，HD95 则用于表征在极端情形下预测边界与真实边界之间的最大距离偏差。

与此同时推理速度与显存占用等因素对临床部署非常关键，但当前主流 GPU（如 RTX 4090）基本能满足常规推理需求，故本研究暂不对此做详细对比。

#### 4.3.5.2 实验结果与分析

为了评估模型的分割能力，表 4.8 统计了不同标注样本量条件下，各模型的 Dice 系数和 HD95 结果。

如表 4.8 所示，(1) 极少量标注（1~3 张）时，DPCP-MNet 在 Dice 和 HD95 上均优于 SAM2 与 MedSAM2，体现出其针对小样本设计的优势。(2) 当标注量增至 5 张后，Med-SAM2 的 Dice 得分达到 85.2%，略高于 DPCP-MNet 的 83.4%，显示出预训练大模型在有一定数据量支撑时可发挥更强的跨域特征迁移能力。(3) 对于形态复杂、边界模糊的病灶，HD95 指标对局部偏移更为敏感，DPCP-MNet 对边缘的关注获得更好的指标表现，在三种样本条件下 HD95 均保持更低。

表 4.8 不同样本量下的分割性能 (Dice, % / HD95, mm)

Table 4.8 Segmentation Performance under Different Sample Sizes (Dice, % / HD95, mm)

样本量	DPCP-MNet (Dice / HD95)	SAM2 (Dice / HD95)	Med-SAM2 (Dice / HD95)
1	78.3 / 3.5 mm	62.5 / 5.2 mm	71.7 / 4.1 mm
3	82.5 / 3.2 mm	68.1 / 4.8 mm	79.3 / 3.5 mm
5	83.4 / 3.0 mm	72.5 / 4.5 mm	85.2 / 3.1 mm

## 4.3.5.3 讨论

本节围绕实验结果展开分析，主要探讨数据量对模型选择的影响，并总结当前研究的局限性和可能改进方向。

## (1) 模型选择策略分析

实验结果表明，在极小样本条件 ( $N \leq 3$ ) 下，DPCP-MNet 在 Dice 和 HD95 上更具优势，说明其在稀缺标注场景中拥有快速收敛与捕捉医学细节的能力。若标注量增至 5 张或更多，Med-SAM2 依托其大规模预训练所获得的泛化特征，在 Dice 指标上实现了更高水平，表明大模型在样本量增多时潜力可逐步被激活。

至于 SAM2，在本研究的设定下性能相对一般，并不代表其在所有医学场景都欠佳。若对其进行更深入的医学域预训练或采用更针对性的微调策略，也可能获得更优表现。表 4.9 对三种模型的适用场景做了简要归纳。

表 4.9 模型选择策略

Table 4.9 Model Selection Strategy

数据量	推荐模型	适用场景
$N \leq 3$	DPCP-MNet	少样本、小规模数据集
$N \geq 5$	Med-SAM2	充分标注、数据丰富的场景
任意 $N$	SAM2 (不推荐)	当前实验设定下效果一般

(2) 实验局限性与改进方向由于本研究聚焦“极小样本”环境下的可行性验证，未进一步扩展至更多标注数量（如 10、20 张）或更多模态（CT、超声等）。后续可考虑：(1) 大规模数据测试。在更丰富的器官和病种（如脑、肺、骨骼等）上评估小样本模型与大模型的泛化能力，探讨它们在不同解剖结构下的分割差异。(2) 模型轻量化与多平台部署。通过剪枝、量化等手段，减少分割大模型推理时的内存与计算需求，适配临床床旁设备或移动端场景。

综上所述，虽然数据规模和标注数量相对有限，但实验结果初步验证了 DPCP-MNet 在极少量标注条件下的优势，同时也显示了预训练大模型（Med-SAM2）在更丰富数据条件下的潜能。未来可在更大规模、多模态的医学影像场景中，系统评估“预训练大模型 vs. 小样本模型”的平衡点，为小样本医学图像分割的研究与临床应用提供更全面的实证依据。

## 4.4 本章小结

本章提出 DPCP-MNet，一种基于双提示协同原型与协同原型 Mamba 交互的小样本医学图像分割网络，旨在缓解器官形态多样和边缘模糊导致的局限。其核心在于：边缘提示用于强化轮廓结构，语义提示专注于内部语义的提取，两者协同引导协同原型构建，从而同时保留空间边界与语义中心信息；同时，Mamba 交互模块以线性复杂度实现全局特征建模，并通过协同原型融合增强查询特征的表达能力和适配性。在多组医学数据集的小样本分割实验中，DPCP-MNet 相比现有方法表现出更稳定的边缘识别和更高的分割精度。初步对比通用分割大模型的结果显示：在充足数据时大模型迁移能力更强，而在极少标注情境下，DPCP-MNet 依托可定制的边缘感知与稀少标注利用策略，仍展现出可观的应用潜力。综上，DPCP-MNet 在小样本医学图像分割中兼顾了几何边界刻画与全局建模效率，实验结果进一步验证了双提示协同原型与 Mamba 交互机制的有效性。后续可在更大规模或多模态医学数据上评估其泛化性能，并在推理阶段结合主动学习或在线微调，以更好满足临床实时性与病灶多变性的需求。

## 第5章 总结与展望

### 5.1 研究成果总结

小样本图像分割正逐渐成为计算机视觉和医疗辅助诊断等领域的重要技术。尽管深度学习显著推动了分割研究，小样本场景仍面临目标尺度多样、边缘模糊、噪声干扰及计算效率等挑战。针对这些问题，本文提出了 MS-ECANet 与 DPCP-MNet 两套方法，分别面向通用和医学场景的小样本分割，并取得以下主要成果：

本文围绕小样本图像分割展开研究，提出了创新的方法并取得了以下主要成果：

(1) 为了更好地兼顾不同尺度目标的精确识别与边缘细节的刻画，本文提出了基于多尺度记忆协调 (MSMC) 与边缘上下文聚合 (ECAM) 的小样本通用图像分割方法 (MS-ECANet)。MSMC 通过构建层级原型记忆池，实现了尺度自适应匹配，而 ECAM 利用梯度调制与特征融合技术强化了边缘特征表达。在优化目标分割精度与边缘细节表现的同时，模型在小样本环境下展现了显著优势。实验结果显示，MS-ECANet 在 PASCAL-5<sup>i</sup>、COCO-20<sup>i</sup> 等通用数据集上能够在处理目标尺度变化与边缘模糊时展现更好的鲁棒性。同时，为验证通用方法在医学分割上的潜力，本文也在 Abdominal-MRI 数据集上进行了附加实验，结果表明 ECAM 的边缘增强策略在低对比度、轮廓模糊的器官分割中依然有效，但 MSMC 对医学影像的增益相对有限。

(2) 在医学图像领域中，针对高分辨率、目标形态多变及噪声干扰等问题，本文提出了基于双提示协同原型与 Mamba 交互的小样本医学图像分割方法 (DPCP-MNet)。该模型通过边缘感知几何提示 (EA-Prompt) 和中心化语义提示 (CS-Prompt) 分别增强了边界与语义信息，有效克服了边界模糊与目标形态复杂等挑战；通过 Mamba 交互模块以线性复杂度实现全局特征建模，并凭借协同原型融合增强查询特征的表达能力和适配性。实验结果表明，DPCP-MNet 在 Abdominal-CT、Abdominal-MRI、Cardiac-MRI 等多组医学数据集上均取得了显著性能提升，且与分割大模型（如 SAM2、MedSAM2）进行对比时，在小样本标注条件下具有更高的性能与边缘精度。

综上所述，本文通过 MS-ECANet 与 DPCP-MNet 两种不同的思路，分别探索了小样本通用分割与小样本医学图像分割在多尺度表示、边缘增强以及高效交互建模方面等关键技术环节，为后续在复杂场景下进一步提高小样本分割精度与泛化能力提供了可行方案。

### 5.2 未来研究方向

尽管本研究在小样本图像分割领域取得了初步成果，但仍存在多项技术挑战和广阔的研究空白。在此基础上，后续工作可从以下方向进一步深入：

(1) 提升模型在复杂真实场景中的鲁棒性：现有研究多基于理想化数据，而在实际

应用中，图像常受到噪声、遮挡及病变异常等因素干扰。可设计更具鲁棒性的数据增强与特征强化策略，利用多尺度、异质扰动模拟真实复杂环境，以提升模型在上述场景中的分割性能。

(2) 融合小样本学习与大模型优势：小样本学习具备模型轻量、推理快速、迁移便捷等特点；而大模型凭借庞大参数容量，在跨域、跨模态任务中展现出更强的特征提取与微调能力。后续工作可尝试采用迁移学习、参数共享或知识蒸馏等策略，将大模型的丰富先验与小样本学习的高效性结合。在保持低计算开销与高推理速度的同时，实现模型对新领域的快速适应。

(3) 多模态图像融合：随着多模态成像技术的普及，如何有效整合 CT、MRI、超声等异质信息，以提升小样本分割的精度与泛化能力，是未来的关键课题。可进一步研究跨模态对齐、共享编码器等策略，在获取丰富上下文的同时，尽可能降低对大规模标注数据的依赖。

总之，随着小样本分割研究的深入与大模型技术的飞速发展，如何融合二者优势并攻克复杂场景中的挑战，仍是一条长期且富有价值的研究道路。期待通过以上方向的深入研究，为通用场景和医疗应用提供精度更高、效率更优、推广更便捷的小样本分割方案。

## 参考文献

- [1] 陈善娟, 于云龙, 李英明. 小样本语义分割研究现状与分析[J]. 计算机学报, 2024, 47(10):2417-2451.
- [2] Ren W, Tang Y, Sun Q, et al. Visual semantic segmentation based on few/zero-shot learning: An overview[J]. IEEE/CAA Journal of Automatica Sinica, 2023.
- [3] Pachetti E, Colantonio S. A systematic review of few-shot learning in medical imaging[J]. Artificial intelligence in medicine, 2024:102949.
- [4] Kim Y, Kang D, Mok Y, et al. A review on few-shot learning for medical image segmentation[C]//2023 International Conference on Electronics, Information, and Communication (ICEIC). IEEE, 2023: 1-3.
- [5] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 3431-3440.
- [6] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation [C]//Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18. Springer, 2015: 234-241.
- [7] He K, Gkioxari G, Dollár P, et al. Mask r-cnn[C]//Proceedings of the IEEE international conference on computer vision. 2017: 2961-2969.
- [8] Girshick R. Fast r-cnn[C]//Proceedings of the IEEE international conference on computer vision. 2015: 1440-1448.
- [9] Liu Z, Lin Y, Cao Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows [C]//Proceedings of the IEEE/CVF international conference on computer vision. 2021: 10012-10022.
- [10] Yuan L, Chen Y, Wang T, et al. Tokens-to-token vit: Training vision transformers from scratch on imagenet[C]//Proceedings of the IEEE/CVF international conference on computer vision. 2021: 558-567.
- [11] Xie E, Wang W, Yu Z, et al. Segformer: Simple and efficient design for semantic segmentation with transformers[J]. Advances in neural information processing systems, 2021, 34:12077-12090.
- [12] Cheng B, Choudhuri A, Misra I, et al. Mask2former for video instance segmentation[J]. arXiv preprint arXiv:2112.10764, 2021.
- [13] Wang X, Zhang X, Cao Y, et al. Seggpt: Towards segmenting everything in context[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2023: 1130-1140.
- [14] Wang K, Liew JH, Zou Y, et al. Panet: Few-shot image semantic segmentation with prototype alignment [C]//proceedings of the IEEE/CVF international conference on computer vision. 2019: 9197-9206.
- [15] Zhang C, Lin G, Liu F, et al. Canet: Class-agnostic segmentation networks with iterative refinement and attentive few-shot learning[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019: 5217-5226.
- [16] Lang C, Cheng G, Tu B, et al. Learning what not to segment: A new perspective on few-shot segmentation[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022: 8057-8067.
- [17] Zhou T, Wang W, Konukoglu E, et al. Rethinking semantic segmentation: A prototype view[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022: 2582-

- 2593.
- [18] Wang W, Zhou T, Yu F, et al. Exploring cross-image pixel contrast for semantic segmentation[C]// Proceedings of the IEEE/CVF international conference on computer vision. 2021: 7303-7313.
  - [19] Michieli U, Ozay M. Prototype guided federated learning of visual feature representations[J]. arXiv preprint arXiv:2105.08982, 2021.
  - [20] Kendall A, Gal Y. What uncertainties do we need in bayesian deep learning for computer vision?[J]. Advances in neural information processing systems, 2017, 30.
  - [21] Rakelly K, Shelhamer E, Darrell T, et al. Conditional networks for few-shot semantic segmentation [J]. International Conference on Learning Representations, International Conference on Learning Representations, 2018.
  - [22] Tian Z, Zhao H, Shu M, et al. Prior guided feature enrichment network for few-shot segmentation[J]. IEEE transactions on pattern analysis and machine intelligence, 2020, 44(2):1050-1065.
  - [23] Chen T, Xie G S, Yao Y, et al. Semantically meaningful class prototype learning for one-shot image segmentation[J]. IEEE Transactions on Multimedia, 2021, 24:968-980.
  - [24] Hu T, Yang P, Zhang C, et al. Attention-based multi-context guiding for few-shot semantic segmentation [C]//Proceedings of the AAAI conference on artificial intelligence: volume 33. 2019: 8441-8448.
  - [25] Li G, Jampani V, Sevilla-Lara L, et al. Adaptive prototype learning and allocation for few-shot segmentation[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021: 8334-8343.
  - [26] Liu Y, Zhang X, Zhang S, et al. Part-aware prototype network for few-shot semantic segmentation [C]//Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part IX 16. Springer, 2020: 142-158.
  - [27] Wang H, Yang Y, Cao X, et al. Variational prototype inference for few-shot semantic segmentation[C]// Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 2021: 525-534.
  - [28] Wang H, Zhang X, Hu Y, et al. Few-shot semantic segmentation with democratic attention networks [C]//Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XIII 16. Springer, 2020: 730-746.
  - [29] Liu B, Jiao J, Ye Q. Harmonic feature activation for few-shot semantic segmentation[J]. IEEE Transactions on Image Processing, 2021, 30:3142-3153.
  - [30] Zhang G, Kang G, Yang Y, et al. Few-shot segmentation via cycle-consistent transformer[J]. Advances in Neural Information Processing Systems, 2021, 34:21984-21996.
  - [31] Yang Y, Meng F, Li H, et al. A new local transformation module for few-shot segmentation[C]// MultiMedia Modeling: 26th International Conference, MMM 2020, Daejeon, South Korea, January 5–8, 2020, Proceedings, Part II 26. Springer, 2020: 76-87.
  - [32] Zhang X, Wei Y, Yang Y, et al. Sg-one: Similarity guidance network for one-shot semantic segmentation [J]. IEEE transactions on cybernetics, 2020, 50(9):3855-3865.
  - [33] Liu Y, Jiang B, Xu J. Axial assembled correspondence network for few-shot semantic segmentation[J]. IEEE/CAA Journal of Automatica Sinica, 2022, 10(3):711-721.
  - [34] Yang B, Wan F, Liu C, et al. Part-based semantic transform for few-shot semantic segmentation[J]. IEEE Transactions on Neural Networks and Learning Systems, 2021, 33(12):7141-7152.
  - [35] Zhang X, Wei Y, Li Z, et al. Rich embedding features for one-shot semantic segmentation[J]. IEEE

- Transactions on Neural Networks and Learning Systems, 2021, 33(11):6484-6493.
- [36] Wang H, Yang Y, Cao X, et al. Variational prototype inference for few-shot semantic segmentation[C]// Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 2021: 525-534.
- [37] Zhang C, Lin G, Liu F, et al. Pyramid graph networks with connection attentions for region-based one-shot semantic segmentation[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019: 9587-9595.
- [38] Gao H, Xiao J, Yin Y, et al. A mutually supervised graph attention network for few-shot segmentation: The perspective of fully utilizing limited samples[J]. IEEE Transactions on neural networks and learning systems, 2022, 35(4):4826-4838.
- [39] Shi X, Wei D, Zhang Y, et al. Dense cross-query-and-support attention weighted mask aggregation for few-shot segmentation[C]//European Conference on Computer Vision. Springer, 2022: 151-168.
- [40] Zhuge Y, Shen C. Deep reasoning network for few-shot semantic segmentation[C]//Proceedings of the 29th ACM International Conference on Multimedia. 2021: 5344-5352.
- [41] Nguyen K, Todorovic S. Feature weighting and boosting for few-shot segmentation[C]//Proceedings of the IEEE/CVF international conference on computer vision. 2019: 622-631.
- [42] Liu L, Cao J, Liu M, et al. Dynamic extension nets for few-shot semantic segmentation[C]//Proceedings of the 28th ACM international conference on multimedia. 2020: 1441-1449.
- [43] Yang X, Wang B, Chen K, et al. Brinet: Towards bridging the intra-class and inter-class gaps in one-shot segmentation. arxiv 2020[J]. arXiv preprint arXiv:2008.06226, 2020.
- [44] Tian P, Wu Z, Qi L, et al. Differentiable meta-learning model for few-shot semantic segmentation[C]// Proceedings of the AAAI Conference on Artificial Intelligence: volume 34. 2020: 12087-12094.
- [45] Boudiaf M, Kervadec H, Masud Z I, et al. Few-shot segmentation without meta-learning: A good transductive inference is all you need?[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021: 13979-13988.
- [46] Wu Z, Shi X, Lin G, et al. Learning meta-class memory for few-shot semantic segmentation[C]// Proceedings of the IEEE/CVF international conference on computer vision. 2021: 517-526.
- [47] Xie G S, Xiong H, Liu J, et al. Few-shot semantic segmentation with cyclic memory network[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021: 7293-7302.
- [48] Mottaghi R, Chen X, Liu X, et al. The role of context for object detection and semantic segmentation in the wild[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2014: 891-898.
- [49] Roy A G, Siddiqui S, Pölsterl S, et al. ‘squeeze & excite’ guided few-shot segmentation of volumetric images[J]. Medical image analysis, 2020, 59:101587.
- [50] Feng R, Zheng X, Gao T, et al. Interactive few-shot learning: Limited supervision, better medical image segmentation[J]. IEEE Transactions on Medical Imaging, 2021, 40(10):2575-2588.
- [51] Sun L, Li C, Ding X, et al. Few-shot medical image segmentation using a global correlation network with discriminative embedding[J]. Computers in biology and medicine, 2022, 140:105067.
- [52] Wu H, Xiao F, Liang C. Dual contrastive learning with anatomical auxiliary supervision for few-shot medical image segmentation[C]//European Conference on Computer Vision. Springer, 2022: 417-434.
- [53] Ouyang C, Biffi C, Chen C, et al. Self-supervised learning for few-shot medical image segmentation [J]. IEEE Transactions on Medical Imaging, 2022, 41(7):1837-1848.

- [54] Hansen S, Gautam S, Jenssen R, et al. Anomaly detection-inspired few-shot medical image segmentation through self-supervision with supervoxels[J]. *Medical Image Analysis*, 2022, 78:102385.
- [55] Shen Q, Li Y, Jin J, et al. Q-net: Query-informed few-shot medical image segmentation[C]//*Proceedings of SAI Intelligent Systems Conference*. Springer, 2023: 610-628.
- [56] Ding H, Sun C, Tang H, et al. Few-shot medical image segmentation with cycle-resemblance attention [C]//*Proceedings of the IEEE/CVF winter conference on applications of computer vision*. 2023: 2488-2497.
- [57] Zhu X, Su W, Lu L, et al. Deformable detr: Deformable transformers for end-to-end object detection. arxiv 2020[J]. *arXiv preprint arXiv:2010.04159*, 2010, 3.
- [58] Xie E, Wang W, Yu Z, et al. Segformer: Simple and efficient design for semantic segmentation with transformers[J]. *Advances in neural information processing systems*, 2021, 34:12077-12090.
- [59] Kirillov A, Mintun E, Ravi N, et al. Segment anything[C]//*Proceedings of the IEEE/CVF international conference on computer vision*. 2023: 4015-4026.
- [60] Ke L, Ye M, Danelljan M, et al. Segment anything in high quality[J]. *Advances in Neural Information Processing Systems*, 2023, 36:29914-29934.
- [61] Xiong Y, Varadarajan B, Wu L, et al. Efficientsam: Leveraged masked image pretraining for efficient segment anything[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2024: 16111-16121.
- [62] Zhang C, Han D, Qiao Y, et al. Faster segment anything: Towards lightweight sam for mobile applications[J]. *arXiv preprint arXiv:2306.14289*, 2023.
- [63] Wu J, Fu R, Fang H, et al. Medical sam adapter: Adapting segment anything model for medical image segmentation. arxiv 2023[J]. *arXiv preprint arXiv:2304.12620*, 2023.
- [64] Ma J, He Y, Li F, et al. Segment anything in medical images[J]. *Nature Communications*, 2024, 15(1): 654.
- [65] Shaban A, Bansal S, Liu Z, et al. One-shot learning for semantic segmentation[J]. *arXiv preprint arXiv:1709.03410*, 2017.
- [66] Chen X, Shi M. Memory-guided network with uncertainty-based feature augmentation for few-shot semantic segmentation[C/OL]//*2024 IEEE International Conference on Multimedia and Expo (ICME)*. 2024: 1-6. DOI: 10.1109/ICME57554.2024.10688161.
- [67] Yan Z, Yin Z, Lin T, et al. Pgp-sam: Prototype-guided prompt learning for efficient few-shot medical image segmentation[J]. *arXiv preprint arXiv:2501.06692*, 2025.
- [68] Tang S, Zu C, Su W, et al. Is foreground prototype sufficient? few-shot medical image segmentation with background-fused prototype[J]. *arXiv preprint arXiv:2412.02983*, 2024.
- [69] Jia M, Tang L, Chen B C, et al. Visual prompt tuning[C]//*European conference on computer vision*. Springer, 2022: 709-727.
- [70] Chen H, Dong Y, Lu Z M, et al. Self-prompting perceptual edge learning for dense prediction[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2023, 34(6):4528-4541.
- [71] Gu A, Dao T. Mamba: Linear-time sequence modeling with selective state spaces[J]. *arXiv preprint arXiv:2312.00752*, 2023.
- [72] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//*Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016: 770-778.

- [73] Everingham M, Van Gool L, Williams C K, et al. The pascal visual object classes (voc) challenge[J]. *International journal of computer vision*, 2010, 88:303-338.
- [74] Lin T Y, Maire M, Belongie S, et al. Microsoft coco: Common objects in context[C]//Computer vision—ECCV 2014: 13th European conference, zurich, Switzerland, September 6-12, 2014, proceedings, part v 13. Springer, 2014: 740-755.
- [75] Min J, Kang D, Cho M. Hypercorrelation squeeze for few-shot segmentation[C]//Proceedings of the IEEE/CVF international conference on computer vision. 2021: 6941-6952.
- [76] Yang B, Liu C, Li B, et al. Prototype mixture models for few-shot semantic segmentation[C]//Computer Vision—ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part VIII 16. Springer, 2020: 763-778.
- [77] Fan Q, Pei W, Tai Y W, et al. Self-support few-shot semantic segmentation[C]//European Conference on Computer Vision. Springer, 2022: 701-719.
- [78] Ai K, Hu H, Zhou Q, et al. Sgt: Self-guided transformer for few-shot semantic segmentation [C]//ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2024: 5935-5939.
- [79] Bao X, Qin J, Sun S, et al. Relevant intrinsic feature enhancement network for few-shot semantic segmentation[C]//Proceedings of the AAAI Conference on Artificial Intelligence: volume 38. 2024: 765-773.
- [80] Wang J, Zhang B, Pang J, et al. Rethinking prior information generation with clip for few-shot segmentation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2024: 3941-3951.
- [81] Kavur A E, Gezer N S, Barış M, et al. Chaos challenge-combined (ct-mr) healthy abdominal organ segmentation[J]. *Medical image analysis*, 2021, 69:101950.
- [82] Ouyang C, Biffi C, Chen C, et al. Self-supervision with superpixels: Training few-shot medical image segmentation without annotation[C]//Computer Vision—ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXIX 16. Springer, 2020: 762-780.
- [83] Wang R, Zhou Q, Zheng G. Few-shot medical image segmentation regularized with self-reference and contrastive learning[C]//International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, 2022: 514-523.
- [84] Zhu Y, Wang S, Xin T, et al. Partition-a-medical-image: Extracting multiple representative sub-regions for few-shot medical image segmentation[J]. *IEEE Transactions on Instrumentation and Measurement*, 2024.
- [85] Shen Y, Fan W, Wang C, et al. Dual-guided prototype alignment network for few-shot medical image segmentation[J]. *IEEE Transactions on Instrumentation and Measurement*, 2024.
- [86] Landman B, Xu Z, Igelsias J, et al. Miccai multi-atlas labeling beyond the cranial vault—workshop and challenge[C]//Proc. MICCAI multi-atlas labeling beyond cranial vault—workshop challenge: volume 5. Munich, Germany, 2015: 12.
- [87] Zhuang X. Multivariate mixture model for cardiac segmentation from multi-sequence mri[C]//International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, 2016: 581-588.



## 攻读学位期间取得的研究成果

攻读学位期间所取得的研究成果，以第一作者身份已在 CCF-C 会议上正式发表论文一篇；以第三作者身份已在 CCF-C 会议上正式发表论文一篇；以第二作者身份，导师第一作者，申请发明专利一项，该专利处于实审阶段。按照双盲评阅的要求，成果的详细信息未在此匿名评审论文中列出。