

中图分类号: TP391

单位代号: 10280

密 级: 公开

学 号: 23721494

上海大学



硕士学位论文

SHANGHAI UNIVERSITY
MASTER'S DISSERTATION

题 目	基于涂鸦标注的弱监督复杂纹理 图像分割方法研究
--------	----------------------------

作 者 李子铭

学科专业 计算机应用技术

导 师 张瑞

完成日期 二〇二六年四月

姓 名：李子铭

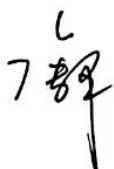
学号：23721494

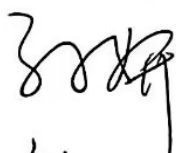
论文题目：基于涂鸦标注的弱监督复杂纹理图像分割方法研究

上海大学

本论文经答辩委员会全体委员审查，确认符合上海大学硕士学位论文质量要求。

答辩委员会签名：

主 席： 

委 员： 


导 师： 

答辩日期： 2026 年 6 月 9 日

姓 名：李子铭

学号：23721494

论文题目：基于涂鸦标注的弱监督复杂纹理图像分割方法研究

上海大学学位论文原创性声明

本人郑重声明：所呈交的学位论文是本人在导师指导下，独立进行研究工作所取得的成果。除了文中特别加以标注和致谢的内容外，论文中不包含其他人已发表或撰写过的研究成果。参与同一工作的其他研究者对本研究所做的任何贡献均已在论文中作了明确的说明并表示了谢意。

学位论文作者签名：李子铭

日期：2026 年 6 月 10 日

上海大学学位论文使用授权说明

本人完全了解上海大学有关保留、使用学位论文的规定，即：学校有权保留论文及送交论文复印件，允许论文被查阅和借阅；学校可以公布论文的全部或部分内容。

(保密的论文在解密后应遵守此规定)

学位论文作者签名：李子铭

导师签名：张瑞

日期：2026 年 6 月 10 日

日期：2026 年 6 月 10 日

上海大学工学硕士学位论文

基于涂鸦标注的弱监督复杂纹理图像
分割方法研究

作者: 李子铭
学科专业: 计算机应用技术
导师: 张瑞

计算机工程与科学学院

上海大学

2026 年 4 月

A Dissertation Submitted to Shanghai University for the Degree
of Master in Engineering

**Research on Weakly Supervised
Segmentation Methods for Complex
Texture Images Based on Scribble
Annotations**

Candidate: Ziming Li

Major: Computer Application Technology

Supervisor: Rui Zhang

**School of Computer Engineering and Science
Shanghai University**

April, 2026

摘 要

复杂纹理图像通常呈现弱对比、强纹理、结构黏连和边界模糊等特征，给高精度像素级分割带来了较大挑战。与此同时，全像素精确标注高度依赖领域专家，获取完整像素级真值往往耗时耗力，严重制约了分割模型的训练与应用。相比全像素标注，涂鸦标注仅需在目标区域内部勾画少量稀疏线条，即可提供类别与局部结构信息，是一种兼顾标注效率与监督有效性的弱监督形式。在涂鸦监督条件下，医学影像与材料显微图像作为复杂纹理图像的两类典型场景，既具有未标注区域约束不足、局部纹理与结构关系复杂以及边界易模糊或黏连等共性特征，又在任务需求上存在一定差异：医学影像更强调形状信息学习、局部纹理刻画与全局结构建模的协同，以及复杂边界区域的判别增强；材料显微图像更强调极小样本条件下的训练稳定性、伪标签可靠性以及复杂相界附近的边界噪声控制。基于此，本文以复杂纹理图像为总体研究背景，选取医学影像与材料显微图像作为具体研究对象，系统研究了基于涂鸦标注的弱监督分割方法。

针对医学影像场景，提出了一种基于一致性约束与对比学习的涂鸦监督分割方法 **FW2SS**。该方法构建了由 **CNN** 与 **Transformer** 组成的异构双分支网络，以增强局部纹理刻画与全局结构建模之间的互补性；进一步结合网络扰动与输入扰动的一致性学习机制，对两个解码器的弱扰动预测结果进行连续权重动态融合，以生成更稳定的像素级伪标签，提升稀疏监督条件下的形状信息学习能力；最后，从跨架构一致且高置信度的像素中构建前景类别原型，并以前景原型为锚点对强扰动视图特征进行像素级原型对比校准，从而增强复杂边界区域的特征判别能力。

针对材料显微图像场景，提出了一种基于多历史投票与边界感知约束的涂鸦监督分割方法 **MVBC**。该方法以采用预训练 **VGG16** 编码器的 **U-Net++** 作为基础分割网络，并结合预训练初始化、浅层冻结和深层微调的迁移学习策略，以提高极小样本条件下的训练稳定性；同时，通过按样本维护的历史预测队列，综合多历史投票、投票占比统计与平均支持置信度估计，并结合边界响应图筛选可靠伪标签，从时间维度提升伪监督的稳定性与可靠性。进一步地，在网络优化过程中联合引入边界平滑与边界锐化约束，从区域内部一致性和边界过渡清晰度两个方面增强模型对复杂

相界的刻画能力。

实验结果表明，所提方法在两类典型复杂纹理图像分割任务中均取得了良好效果。在医学图像分割任务中，FW2SS 在 ACDC 和 MSCMRseg 数据集上的平均 Dice 分别达到 90.0% 和 88.3%。在本文采用的数据划分、二维切片训练和统一实验协议下，该方法优于多种现有涂鸦监督方法，并高于若干常用全监督 U-Net 系列基线方法，整体性能接近本文纳入比较的强全监督方法 nnUNet。在材料图像分割任务中，MVBC 在仅使用极少量训练图像和涂鸦标注的条件下，仍在陶瓷、钛铝合金和碳钢数据集上分别取得了 93.9%/95.3%、85.2%/91.7% 和 81.2%/87.1% 的 mIoU/mDice。上述结果表明，在低标注条件下，针对医学影像场景围绕形状信息学习、局部纹理刻画与全局结构建模协同以及复杂边界区域判别增强进行针对性设计，针对材料显微图像场景围绕训练稳定性、伪标签可靠性以及复杂相界附近边界噪声控制进行针对性设计，能够在显著降低标注成本的前提下有效提升两类典型复杂纹理图像的分割性能，并为相关技术在医学影像智能分析与材料组织自动表征中的应用提供可行思路。

关键词：复杂纹理图像；涂鸦监督；弱监督学习；医学影像分割；材料显微图像分割；伪标签学习

ABSTRACT

Complex texture images are typically characterized by weak contrast, strong texture, structural adhesion, and blurry boundaries, which pose significant challenges to high-precision pixel-level segmentation. Meanwhile, accurate dense pixel-wise annotation relies heavily on domain experts, and obtaining complete pixel-level ground truth is often time-consuming and labor-intensive, severely restricting the training and application of segmentation models. Compared with full pixel-wise annotation, scribble annotation requires only a small number of sparse strokes drawn inside target regions to provide category information and local structural cues, making it a form of weak supervision that balances annotation efficiency and supervisory effectiveness. Under scribble supervision, medical images and material micrographs, as two typical scenarios of complex texture images, not only share common characteristics such as insufficient constraints on unlabeled regions, complex local texture – structure relationships, and blurred or adherent boundaries, but also differ to some extent in task requirements. Medical images place greater emphasis on the coordinated learning of shape information, local texture representation, and global structure modeling, as well as enhanced discrimination in complex boundary regions. In contrast, material micrographs place greater emphasis on training stability under extremely small-sample conditions, pseudo-label reliability, and boundary noise control near complex phase interfaces. Accordingly, taking complex texture images as the overall research background and selecting medical images and material micrographs as specific research objects, this thesis systematically investigates weakly supervised segmentation methods based on scribble annotation.

For the medical imaging scenario, this thesis proposes a scribble-supervised segmentation method based on consistency constraints and contrastive learning. The method constructs a heterogeneous dual-branch network composed of a CNN and a Transformer to enhance the complementarity between local texture representation and global structure modeling. It further incorporates a consistency learning mechanism based on network perturbations and input perturbations, and performs continuously weighted dynamic fusion

of the weakly perturbed predictions from the two decoders to generate more stable pixel-level pseudo-labels, thereby improving shape information learning under sparse supervision. Finally, foreground class prototypes are constructed from pixels that are both cross-architecture consistent and of high confidence, and, using the foreground prototypes as anchors, pixel-level prototype contrastive calibration is performed on the features of the strongly perturbed branch, thereby enhancing feature discrimination in complex boundary regions.

For the material micrograph scenario, this thesis proposes a scribble-supervised segmentation method based on multi-history voting and boundary-aware constraint. The method adopts U-Net++ with a pretrained VGG16 encoder as the base segmentation network, and combines a transfer learning strategy involving pretrained initialization, shallow-layer freezing, and deep-layer fine-tuning to improve training stability under extremely small-sample conditions. Meanwhile, by maintaining a sample-wise historical prediction queue, it integrates multi-history voting, vote-ratio statistics, and average support-confidence estimation, together with a boundary response map, to select reliable pseudo-labels, thereby improving the stability and reliability of pseudo supervision from the temporal perspective. Furthermore, boundary smoothing and boundary sharpening constraints are jointly introduced during network optimization to enhance the model's ability to characterize complex phase boundaries from two aspects: intra-region consistency and the clarity of boundary transitions.

Experimental results demonstrate that the proposed methods achieve favorable performance on two typical complex texture image segmentation tasks. In medical image segmentation tasks, FW2SS achieves average Dice scores of 90.0% and 88.3% on the ACDC and MSCMRseg datasets, respectively, outperforming various existing scribble-supervised methods. Under the same number of training cases and the experimental protocol adopted in this thesis, it achieves higher Dice scores than several commonly used fully supervised U-Net-based baselines, while its overall performance remains close to that of nnUNet, the strongest fully supervised method included in the comparison. In material image segmentation tasks, using only a very small number of training images and scribble annotations, MVBC still achieves mIoU/mDice scores of 93.9%/95.3%, 85.2%/91.7%, and 81.2%/87.1%

on the ceramic, TiAl alloy, and carbon steel datasets, respectively. These results indicate that, under low-annotation conditions, targeted designs for the medical imaging scenario focusing on shape information learning, the coordination of local texture representation and global structure modeling, and enhanced discrimination in complex boundary regions, as well as targeted designs for the material micrograph scenario focusing on training stability, pseudo-label reliability, and boundary noise control near complex phase interfaces, can effectively improve the segmentation performance of these two typical types of complex texture images while significantly reducing annotation cost, and provide a feasible approach for the application of related techniques in intelligent medical image analysis and automatic characterization of material microstructures.

Keywords: complex texture images; scribble supervision; weakly supervised learning; medical image segmentation; material micrograph segmentation; pseudo-label learning

目 录

摘 要	I
ABSTRACT	III
第一章 绪论	1
1.1 课题背景概述	1
1.2 课题研究目的与意义	2
1.3 国内外研究现状.....	4
1.3.1 复杂纹理图像分割算法研究现状	4
1.3.2 基于涂鸦标注的图像分割算法研究现状	6
1.4 论文主要工作	8
1.5 论文组织结构	9
第二章 相关理论与技术概述	11
2.1 基于深度学习的语义分割概述	11
2.2 弱监督分割与涂鸦标注概述.....	13
2.3 伪标签生成与一致性学习概述	15
2.4 语义分割评价指标概述	18
2.5 本章小结	21
第三章 基于一致性约束与对比学习的涂鸦监督医学影像分割方法	22
3.1 方法概述	22
3.1.1 模型框架	22
3.1.2 CNN-Transformer 异构网络	24
3.1.3 涂鸦一致性匹配.....	26
3.1.4 像素级原型对比较准	28
3.1.5 损失函数	31
3.2 实验分析	31
3.2.1 数据集介绍	31
3.2.2 实验设置	33

3.2.3	数据增强设置	33
3.2.4	对比实验	34
3.3	消融实验	39
3.4	本章小结	45
第四章	基于多历史投票与边界感知约束的涂鸦监督材料显微图像分割方法	47
4.1	方法概述	47
4.1.1	模型框架	47
4.1.2	基础分割网络	49
4.1.3	多历史投票伪标签	51
4.1.4	边界感知约束	57
4.1.5	损失函数	61
4.2	实验分析	62
4.2.1	数据集介绍	62
4.2.2	实验设置	64
4.2.3	对比实验	66
4.2.4	消融实验	70
4.2.5	敏感性分析	74
4.3	本章小结	79
第五章	总结与展望	81
5.1	结论	81
5.2	工作展望	83
参考文献		85
攻读硕士学位期间取得的研究成果		96
致 谢		97

第一章 绪论

1.1 课题背景概述

在科技持续进步的背景下，生物医学与材料科学已成为推动国家科技创新和产业升级的重要领域。随着深度学习的快速发展，图像处理技术在医学影像分析和材料组织表征中的应用不断深化，逐步成为促进学科交叉融合与技术创新的重要动力^[1,2]。

在医学影像领域，精准的图像分割是临床诊断辅助、手术规划及疗效评估的重要基础^[3]。近年来，人工智能与医学影像技术的快速发展，使得医学影像数据规模持续扩大、模态类型日益丰富，这种趋势推动了图像分割技术在疾病筛查、病灶识别和治疗评估等方面的广泛应用。国外较早将计算机视觉与深度学习方法应用于医学影像分析，在肿瘤检测、器官分割及病理图像识别等方面取得了显著进展^[4,5]。我国在“健康中国 2030”战略持续推进的背景下，医疗信息化与智慧医疗建设不断深化，CT、MRI、病理切片等多模态医学影像数据快速积累，为医学影像自动分割技术的发展提供了重要支撑。借助计算机视觉技术对病灶区域及器官组织进行自动分割，不仅能够提高分析效率、减轻人工阅片负担，还可为后续定量分析提供较为可靠的形态学依据^[1]。

在材料图像领域，材料显微图像是研究材料微观组织结构的重要数据来源。探索晶粒、第二相、裂纹等微观组织特征与宏观性能之间的映射关系，已成为材料科学中的重要研究任务^[2,6]。近年来，机器学习方法在微观组织建模、多尺度仿真以及工艺—结构—性能关联预测等方向的应用不断深入，为材料微观组织分析从定性描述向量化、智能化方向的转变提供了重要支撑^[7]。随着扫描电子显微镜、光学显微镜等表征技术的发展，材料显微图像数据不断积累，为材料组织结构的自动识别、定量分析和性能预测提供了基础。借助计算机视觉技术，研究人员能够从复杂的金相图像或显微图像中高效提取组织特征，实现材料组织的定量表征，从而加速新材料的发现与设计^[6,8]。

然而，现有医学影像分割和材料显微图像分割方法大多建立在深度学习模型基础之上，而深度学习模型的有效训练通常依赖大量高质量的像素级精细标注。此类标注不仅耗时耗力，而且高度依赖医学专家或材料领域研究人员的专业知识。在医

学影像中,不同器官、病灶或组织结构往往存在形态差异大、边界模糊、灰度对比度低等问题,精确标注难度较高;在材料显微图像中,晶粒、相界、孔洞和裂纹等结构常表现出复杂纹理、尺度变化和边界黏连等特征,同样增加了像素级标注的成本与不确定性。因此,尽管完全监督分割方法在标注充分的条件下能够取得较好效果,但其对高质量稠密标注的依赖限制了其在实际场景中的推广应用。

为降低像素级标注成本,弱监督分割方法逐渐成为医学影像和材料显微图像分析中的重要研究方向。其中,基于涂鸦标注(**Scribble-based Annotation**)的弱监督分割方法仅需标注者在目标区域内部勾画少量稀疏线条,便能够为模型提供目标区域的结构线索与关键语义信息,从而在显著降低标注成本的同时保留一定的监督信息^[9,10]。因此,涂鸦标注较适用于医学影像和材料显微图像这类专业标注成本高、数据规模扩展受限的应用场景。

本文分别针对医学影像和材料显微图像这两类典型复杂纹理图像开展方法设计。具体而言,针对医学影像分割中形状信息学习不足、局部纹理刻画不充分以及全局结构建模能力有限等问题,提出面向医学影像场景的 **FW2SS** 方法;针对材料显微图像分割中小样本训练稳定性不足、伪标签可靠性有限以及复杂相界附近边界噪声较明显等问题,提出面向材料显微图像场景的 **MVBC** 方法,以在降低标注成本的同时提升两类复杂纹理图像弱监督分割的准确性与鲁棒性。

1.2 课题研究目的与意义

本课题以复杂纹理图像的涂鸦监督分割为研究对象,围绕医学影像与材料显微图像两类具有代表性的应用场景,研究低标注成本条件下的高精度图像分割方法。医学影像和材料显微图像虽然来源于不同领域,但均具有纹理结构复杂、目标边界多变、前景与背景对比不明显等特点,且像素级精细标注通常依赖领域专家完成,标注成本较高^[1,2,6]。因此,如何降低对像素级精细标注的依赖,并提升复杂纹理图像分割模型的准确性与鲁棒性,是本课题关注的主要问题。

在弱监督分割任务中,涂鸦标注以少量稀疏线条替代完整的像素级掩码,能够在显著降低标注成本的同时提供目标类别、局部结构和空间分布等关键信息^[9,10]。然而,由于涂鸦标注仅覆盖图像中的少部分区域,大量未标注像素缺乏直接监督,模型训练过程中容易出现监督信息不足、伪标签可靠性有限以及边界区域误差累积等

问题^[11,12]。对于纹理复杂、边界模糊且目标形态多变的复杂纹理图像而言，上述问题更加突出。因此，本课题的核心目的在于探索如何从稀疏涂鸦标注中挖掘有效监督信息，构建可靠的伪监督信号，并提升模型对目标结构和复杂边界区域的表征能力。

具体到医学影像分割场景，目标区域通常具有较强的形态差异和结构变化。不同个体之间的解剖结构存在差异，病灶或器官边界可能受成像分辨率、组织黏连和灰度对比度不足等因素影响而变得模糊。此外，医学影像分割不仅需要关注局部纹理细节，还需要保持目标形状和全局解剖结构的一致性^[1,3]。在涂鸦监督条件下，稀疏标注难以充分约束完整目标区域，容易导致模型对形状信息学习不足、局部纹理刻画与全局结构建模难以协同以及复杂边界区域判别能力有限等问题^[13,14]。

具体到材料显微图像分割场景，图像通常受到材料制备工艺、显微成像条件和组织结构差异的影响，表现出较强的纹理异质性和复杂相界形态。晶粒、第二相、孔洞和裂纹等微观结构之间可能相互交错，边界区域易出现模糊、断裂或黏连现象。与此同时，材料显微图像数据获取和标注成本较高，小样本问题较为突出^[6,15,16]。在涂鸦监督条件下，模型不仅需要从有限样本和稀疏标注中学习有效特征，还需要抑制伪标签噪声在复杂相界附近的传播。因此，材料显微图像分割面临小样本训练稳定性不足、伪标签可靠性有限以及边界噪声易累积等难点^[8,16]。

综合考虑以上问题，医学影像与材料显微图像在涂鸦监督分割条件下均面临标注信息稀疏、未标注区域约束不足以及复杂边界难以准确识别等共性难点，但二者在图像结构、目标形态、数据规模和任务需求等方面存在差异，因而需要结合具体场景进行针对性建模。基于此，本文分别面向医学影像和材料显微图像两类复杂纹理图像场景开展方法设计，研究如何在低标注成本条件下充分利用有限监督信息，提高复杂纹理图像分割的准确性与鲁棒性。

对于医学影像涂鸦监督分割任务，本文主要针对目标形态变化大、边界模糊以及局部纹理与全局结构关系复杂等问题，从结构一致性建模和特征判别增强角度出发，研究基于一致性约束与对比学习的分割方法，以提升模型在稀疏标注条件下对目标区域、形状信息和复杂边界的表达能力。对于材料显微图像涂鸦监督分割任务，本文主要针对样本规模有限、微观组织结构复杂、相界模糊以及伪标签噪声易传播等问题，从历史信息利用和边界感知约束建模角度出发，研究基于多历史投票与边界感知约束的分割方法，以增强模型在小样本和弱标注条件下的训练稳定性与边界

刻画能力。

综上，本文以复杂纹理图像的涂鸦监督分割为研究对象，结合医学影像与材料显微图像的共性问题 and 场景差异，探索低成本、高鲁棒性的弱监督分割方法。在理论层面，本文围绕稀疏监督信息利用、伪标签可靠性提升、结构信息建模和边界判别增强等问题展开研究，有助于丰富弱监督图像分割方法在复杂纹理场景中的研究内容；在应用层面，本文可降低医学影像与材料显微图像分割任务对像素级精细标注的依赖，减轻领域专家的标注负担，并为医学影像辅助诊断和材料微观组织定量表征提供技术支撑。

1.3 国内外研究现状

本文融合图像处理、机器学习与深度视觉表征等技术，旨在面向复杂纹理图像中的典型场景，研究非完全标注条件下的高精度分割方法。本章将从复杂纹理图像分割方法和基于涂鸦标注的图像分割方法两个方面梳理相关研究现状，从而为后续算法设计提供理论依据。

1.3.1 复杂纹理图像分割算法研究现状

复杂纹理图像分割是计算机视觉领域中极具挑战性的研究任务，其对象广泛涵盖生物医学影像和材料显微组织图像。尽管二者应用场景存在显著差异，但由于组织交错、相界模糊，前景与背景之间通常缺乏清晰的对比，其在底层视觉特征上表现出较高的一致性；与此同时，其内部纹理往往呈现出较强的随机性与异构性，并伴随深层语义信息表达不足的问题^[1,17]。纵观国内外研究进展，针对该类图像的分割算法演进大体可归纳为传统机器学习方法和深度表征学习方法两个主要阶段^[2,3,17,18]。

在深度学习兴起之前，复杂纹理图像分割研究主要经历了由基于像素强度的底层处理向基于结构化纹理特征的统计建模的转变。由于医学影像中的病灶区域与材料显微图像中的组织结构通常缺乏显著的亮度差异，早期研究主要依赖手工特征工程，以模拟人类视觉系统对纹理的感知机制。在特征提取层面，研究者通过构建多种数学算子来刻画图像的局部空间相关性。其中，灰度共生矩阵（Gray Level Co-occurrence Matrix, GLCM）^[19]基于不同距离和方向上像素对的联合概率分布，提取能量、熵、对比度和相关性等统计特征；局部二值模式（Local Binary Patterns, LBP）^[20]通过比较中心像素与其邻域像素之间的灰度关系，有效表征了复杂纹理的旋转不变特

性；此外，Gabor 变换^[21]凭借其多尺度、多方向的频率分析能力，被广泛用于捕捉医学组织切片或金属材料相结构中的细微方向性纹理。为进一步引入几何约束，活动轮廓模型（Active Contour Models）^[22]及其改进的 Chan–Vese 模型^[23]相继被提出。这类方法通过构建包含内部能量项和外部能量项的变分泛函，在噪声干扰条件下寻求能量最小化对应的分割边界。在分类与决策层面，上述手工特征通常被输入经典机器学习模型以实现模式识别，如支持向量机（Support Vector Machine, SVM）^[24]、随机森林（Random Forest, RF）^[25]以及模糊聚类算法（Fuzzy C-Means, FCM）^[26]等。然而，传统方法的有效性在很大程度上依赖于领域专家经验、特征设计质量及参数调节策略。当面对形态随机性强、边界弥散的复杂目标时，这类方法由于缺乏高层语义信息的引导，往往容易陷入局部最优，进而导致其在处理跨设备、跨工艺来源图像时的鲁棒性不足^[17]。

自 2012 年 AlexNet^[27]确立卷积神经网络（Convolutional Neural Networks, CNN）在特征自适应学习中的主导地位以来，深度学习逐步取代传统手工特征方法，成为复杂纹理图像分割的主流技术。全卷积网络（Fully Convolutional Networks, FCN）^[28]首次将图像分类网络扩展为端到端的像素级分割模型，奠定了语义分割的基础框架。随后，经典 U-Net 结构^[4]通过对称式编码器—解码器架构及跳跃连接机制，有效实现了高层语义信息与底层空间细节的融合，显著提升了模型对病灶边缘和材料微观组织的表征能力。该结构在医学影像分割领域表现出优异的泛化性能，并衍生出大量改进模型。例如，3D U-Net^[29]将二维卷积操作扩展至三维体数据，能够支持体积分割；U-Net++^[5]通过嵌套式密集跳跃连接进一步增强了多尺度特征融合能力；nnU-Net^[30]则引入自适应网络配置与训练策略，成为多类医学影像分割任务中的重要基准模型。为进一步抑制背景噪声并增强模型对关键区域的关注，Attention U-Net^[31]引入注意力门控机制，通过加性注意力门对跳跃连接特征进行动态加权，从而有效提升分割精度。与此同时，DeepLab 系列^[32]通过空洞卷积和空洞空间金字塔池化模块（Atrous Spatial Pyramid Pooling, ASPP）显著扩大感受野，在捕获多尺度上下文信息的同时保持较高的计算效率，因此被广泛应用于复杂纹理图像分割任务。随着 Transformer 架构在计算机视觉领域的引入与发展，Vision Transformer（ViT）^[33]及其在分割领域的代表性变体 TransUNet^[34]突破了传统卷积算子局部感受野的限制，将自注意力机制引入 U-Net 框架，通过融合 CNN 与 Transformer 的编码结构，实现了全局特征建模与局部细节保留的有机统一。Swin-Unet^[35]采用纯 Transformer 架构，并结合移位窗口注

注意力机制，在保持层次化特征提取能力的同时有效降低了计算复杂度；UNETR^[36]则直接以 **Transformer** 作为三维编码器，并通过跳跃连接与 **CNN** 解码器进行特征融合，特别适用于三维医学体积分割任务，并在多个公开基准数据集上取得了领先性能。

尽管全监督深度学习模型在复杂纹理图像分割任务中取得了显著进展，但其优异性能通常高度依赖大规模像素级精确标注数据。在医学临床影像与材料显微组织图像等应用场景中，全像素标注不仅耗时耗力，而且通常需要具备专业背景的领域专家参与，因而在很大程度上限制了深度分割方法的实际推广与应用。与此同时，复杂纹理图像普遍存在边界模糊、纹理异质性强、样本获取困难以及跨设备、跨工艺条件下数据分布差异显著等问题，进一步加剧了高精度分割需求与低标注成本之间的矛盾。因此，如何在减轻人工标注负担的前提下，有效挖掘有限监督信息，并保持模型对复杂结构及边界区域的表征能力，已成为当前复杂纹理图像分割研究的重要方向。其中，基于非完全标注的弱监督分割方法因能够兼顾标注效率与模型性能，正逐渐受到广泛关注。

1.3.2 基于涂鸦标注的图像分割算法研究现状

在非完全标注的弱监督分割中，监督信号通常包括图像级标签、包围盒、点标注和涂鸦标注等形式^[9,37-39]。相较于仅提供类别存在性的图像级标签，以及仅给出粗略位置范围的包围盒和点标注，涂鸦标注能够以较低的人工代价提供目标的类别信息、局部结构走向以及空间分布线索，因此在标注成本与监督有效性之间取得了较好的平衡^[9,11,40,41]。对于医学影像中的病灶与器官组织以及材料显微图像中的复杂相区与互穿结构而言，涂鸦标注可有效缓解全标注所带来的高昂成本，为后续监督传播与伪标签构造提供较稳定的语义监督^[10,11]。已有研究对弱监督语义分割方法进行了系统综述，从图像级标签出发梳理了由传统方法到视觉基础模型的演进脉络，为不同标注形式下的监督信息利用策略提供了重要参考^[42]。从研究脉络上看，基于涂鸦标注的图像分割最早起源于交互式分割，其代表性工作包括 **Graph-cut**、**GrabCut** 和 **Random Walk** 等^[43-45]。这类方法主要利用像素之间的灰度相似性、空间邻近性以及种子点约束，将少量人工标注逐步扩散到未标注区域。

随着深度学习的发展，涂鸦监督分割的核心问题逐渐从如何传播稀疏标注演化为如何构造可靠的致密伪监督并稳定地约束未标注区域。围绕这一目标，现有方法大体可以归纳为三条技术主线。

第一类是基于监督传播的弱监督分割方法。这类方法主要关注如何在深度学习框架下，将稀疏涂鸦标注扩展为对未标注区域的有效约束。代表性工作通常通过图割优化、归一化割损失、图正则化或区域一致性约束等机制，将已标注种子区域的语义信息逐步传播至未知区域^[9,40,41,46]。其优点在于能够直接利用涂鸦所提供的确定性监督信息，在训练初期快速建立目标区域的大致范围，但这类方法的传播过程通常仍较依赖局部纹理一致性和邻域关系，当图像中存在弱边界、复杂纹理或大范围异质结构时，传播结果容易偏离真实轮廓。这类方法为后续研究奠定了从稀疏监督生成初始致密标签的基础，但在复杂纹理场景下的鲁棒性仍然有限。

第二类方法是基于伪标签迭代与一致性学习的策略。这类方法构成了当前涂鸦监督分割研究的主体，其核心思想是在未标注区域利用模型预测生成伪标签，并通过输入扰动一致性、网络扰动一致性提升伪标签的可靠性^[12,47-50]。其中，一类方法主要基于输入视图扰动展开，例如 **CycleMix** 和 **ShapePU** 通过构造不同的增强视图并约束模型输出保持一致，从而提升网络对外观变化和输入扰动的鲁棒性^[12,47]；另一类方法则侧重于网络扰动或多分支协同，例如 **DMPLS**、**ScribbleVC** 和 **ScribFormer** 利用双网络、双解码器或异构分支之间的差异生成共识伪标签，以缓解单一预测的不稳定性^[48-50]。总体而言，这类方法推动了涂鸦监督分割由单次标签传播向动态自训练与一致性约束相结合的方向发展。然而，基于输入扰动的方法在稀疏监督条件下容易因强增强引入额外噪声，进而导致伪标签坍塌或错误扩散；基于网络扰动的方法虽然能够在一定程度上提高伪标签稳定性，但若缺乏进一步的特征判别增强与边界筛选机制，复杂边界区域中的噪声仍可能被保留并持续累积^[48,49,51]。

第三类是基于结构先验与特征增强的方法。这类方法通常通过强化主干网络的表征能力来弥补稀疏监督带来的信息不足，例如引入 **Transformer**、**Visual Mamba-CNN** 等结构，以增强长程依赖建模、全局上下文聚合和整体轮廓恢复能力^[34,35,52]。对于医学影像中的器官形状建模以及材料显微图像中的复杂组织关系表达，这类方法具有积极意义。与此同时，也有研究通过原型对比学习、多任务约束等方式进一步增强困难区域和复杂边界的判别能力^[53,54]。需要指出的是，结构建模能力的增强并不等同于伪标签可靠性的提升，若缺乏与之配套的伪标签筛选、一致性约束或边界判别机制，主干网络带来的全局建模优势往往难以直接转化为稳定的像素级监督信号。

现有研究表明，基于涂鸦标注的图像分割已经形成了较为清晰的发展链路。早期工作重在利用图模型完成稀疏监督传播，现阶段研究重点转向伪标签生成与一致性

学习，近期方法则进一步强调结构建模与特征增强^[9,40,48,50]。然而，监督传播方法能够提供初始区域线索，但对复杂纹理的适应性不足；伪标签与一致性学习方法显著增强了未标注区域的监督能力，但在复杂边界附近仍容易受到噪声干扰；结构先验与骨干增强方法改善了全局表达，却尚未从根本上解决弱监督条件下伪标签可靠性不足的问题。现有方法大多分别关注监督传播、预测稳定或结构表征中的某一侧面，尚未形成兼顾伪标签构建、结构信息稳定传播与边界有效表达的统一研究框架^[48-50]。

此外，现有研究的主要验证场景仍多集中于医学影像或相对数据较充分的任务设置，而针对材料显微图像小样本涂鸦监督分割的系统研究仍较为有限。已有材料显微图像相关研究表明，该类图像通常具有复杂纹理、模糊边界、样本稀缺以及标注过程对领域知识强依赖等特点，这使得少标注条件下的稳定分割更加困难^[51,55]。与医学影像相比，材料显微图像通常具有更强的纹理异质性、更复杂的相界形态以及更显著的小样本约束，这使得模型单次预测波动更大，伪标签也更容易在复杂边界附近发生退化。特别地，Han 等提出了面向小样本材料显微图像的基于伪标签的弱监督分割方法，将涂鸦标注与伪标签构建直接结合起来，为材料显微图像小样本弱监督分割提供了较有代表性的研究实例^[51]。当研究对象扩展至材料显微图像小样本场景后，问题焦点进一步转化为如何在极有限样本条件下构造高稳定性、高可靠性的伪标签，并避免边界噪声在训练过程中持续累积。

1.4 论文主要工作

面向医学影像与材料显微图像在有限标注条件下的高精度分割需求，本文以复杂纹理图像的涂鸦监督分割为研究对象，围绕稀疏标注条件下伪监督信号可靠性不足、结构信息利用不充分以及复杂边界区域判别困难等问题，分别针对医学影像和材料显微图像两类典型场景开展方法设计与实验验证。本文的总体研究思路如图1.1所示。

本文的主要工作与创新点如下：

(1) 针对涂鸦监督医学影像分割中形状信息学习不足、局部纹理刻画与全局结构建模难以协同以及复杂边界判别困难等问题，提出了基于一致性约束与对比学习的FW2SS方法。该方法构建CNN与Transformer异构双分支网络，结合弱到强一致性学习、双重一致性伪标签生成和像素级原型对比较准模块，在稀疏涂鸦监督条件

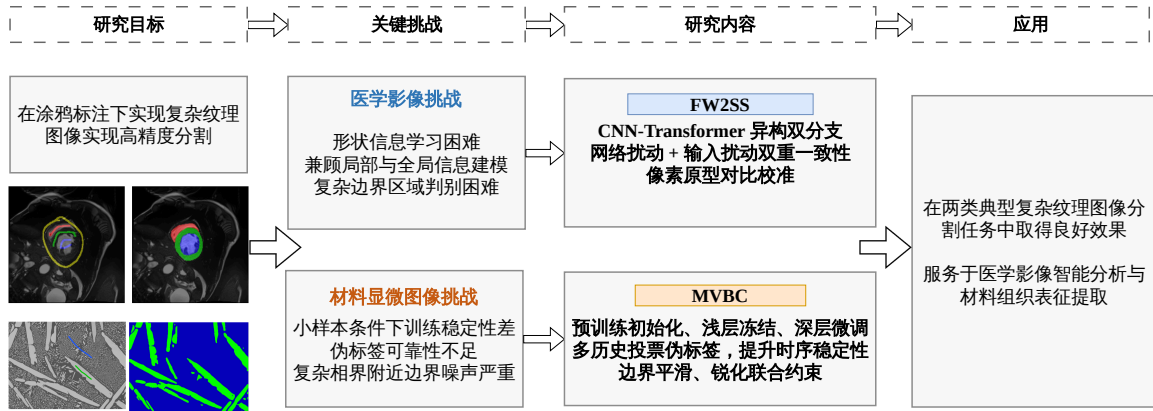


图 1.1 本文总体研究思路

下增强局部纹理与全局结构的互补建模能力，提高伪标签可靠性和复杂边界区域的特征判别能力，从而提升医学影像分割性能。

(2) 针对材料显微图像小样本涂鸦监督分割中训练稳定性不足、伪标签可靠性有限以及复杂相界附近边界噪声易传播等问题，提出了基于多历史投票与边界感知约束的 **MVBC** 方法。该方法通过按样本维护历史预测队列并进行多历史投票，从时间维度筛选稳定可靠的伪标签；同时结合边界响应图、边界平滑约束和边界锐化约束，抑制复杂相界附近的伪标签噪声传播，增强模型对材料显微图像复杂边界的刻画能力。

本文围绕复杂纹理图像的涂鸦监督分割问题，分别面向医学影像与材料显微图像两类应用场景开展方法设计与实验验证，探索了低标注成本条件下可靠伪监督构建、结构信息建模和复杂边界判别增强的有效方法。

1.5 论文组织结构

本论文共分为五章，各章内容安排如下：

第一章为绪论。主要介绍研究背景与研究意义，阐述医学影像和材料显微图像在复杂纹理分割任务中面临的关键难点，并在此基础上系统梳理复杂纹理图像分割方法的发展脉络以及基于涂鸦标注的弱监督分割研究进展，进一步明确本文的研究问题、主要工作与整体技术路线。

第二章为相关理论与技术概述。该章围绕本文所需的关键理论展开，系统介绍基于深度学习的语义分割方法、弱监督分割与涂鸦标注机制、伪标签生成与一致性正则化思想，以及常用分割评价指标，为后续方法设计和实验分析提供理论基础。

第三章为基于一致性约束与对比学习的涂鸦监督医学影像分割方法。针对医学影像在涂鸦监督条件下存在的形状信息学习不足、局部纹理刻画与全局结构建模难以协同以及复杂边界区域判别困难等问题,提出异构双分支一致性学习框架 **FW2SS**;通过 **CNN-Transformer** 异构双分支建模、联合网络扰动与输入扰动的一致性学习,并借助像素级原型对比校准模块,提升形状信息学习能力与边界区域特征判别能力,并在 **ACDC** 与 **MSCMRseg** 数据集上验证方法的有效性。

第四章为基于多历史投票与边界感知约束的涂鸦监督材料显微图像分割方法。针对材料显微图像在小样本涂鸦监督条件下存在的训练稳定性不足、伪标签可靠性有限以及复杂相界附近边界噪声易传播等问题,提出 **MVBC** 方法;该方法通过迁移学习策略增强小样本条件下的训练稳定性,通过多历史投票机制构建可靠伪标签,并结合边界阈值过滤、边界平滑和边界锐化约束增强对复杂相界的建模能力,最终在陶瓷、钛铝合金和碳钢数据集上进行了系统实验验证。

第五章为总结与展望。对全文研究内容进行总结,归纳所提方法的主要贡献、实验结论与不足之处,并围绕复杂纹理图像弱监督分割的后续发展方向,从模型统一化、跨域泛化及实际应用等方面提出进一步研究设想。

第二章 相关理论与技术概述

本章围绕后续研究涉及的相关理论与技术展开。首先，回顾深度学习语义分割的主要模型架构；其次，介绍弱监督学习与涂鸦标注的基本思想，并结合医学影像与材料显微图像这两类典型复杂纹理场景说明其适用背景；随后，概述伪标签生成与一致性学习的基本原理；最后，说明后续实验中采用的评价指标，为后续章节的结果分析提供基础。

2.1 基于深度学习的语义分割概述

语义分割的核心任务是为图像中的每个像素赋予明确的语义类别标签，从而同时实现像素类别分类与空间定位。与图像分类或目标检测相比，语义分割不仅要求模型具备较强的语义判别能力，还要求其能够恢复目标区域的空间边界、局部形态与细粒度结构信息。因此，对于医学影像与材料显微图像这类复杂纹理数据而言，分割性能往往不仅取决于高层语义特征的表达能力，还取决于模型对低层纹理、边界细节及局部几何结构的刻画能力。尤其是在病灶边缘模糊、组织间对比度较低，或显微组织中晶粒、孔洞、第二相和裂纹等结构相互黏连的情况下，如何在语义抽象与空间恢复之间取得平衡，始终是语义分割模型设计中的关键问题^[4,5,28]。

从现有深度学习方法的发展脉络来看，语义分割主干网络大致可以分为两类。第一类是以 FCN、U-Net 及其扩展模型为代表的卷积式编码器—解码器架构^[4,5,28]。FCN 首次将全连接层替换为卷积层，实现了端到端的像素级预测，为现代语义分割奠定了基础^[28]。在此基础上，U-Net 通过对称的编码器—解码器结构以及跳跃连接机制，将浅层高分辨率特征与深层语义特征进行融合，使网络在获得上下文信息的同时仍具备较强的定位能力，因此在医学影像分割中获得了广泛应用^[4]。随后，U-Net++ 进一步通过嵌套式跳跃连接和深监督机制缩小编码器与解码器特征之间的语义差异，从而增强了对边界和细粒度结构的恢复能力^[5]。总体而言，这类卷积式架构在局部纹理建模、边界细节保持以及形态学特征恢复方面具有明显优势，尤其适用于器官轮廓、病灶边缘及材料组织纹理等对局部结构敏感的场景。然而，由于卷积操作本质上依赖有限感受野，其对长距离依赖关系和全局形状一致性的建模能力相对受限，因

此在处理尺度变化显著、空间分布复杂或结构关联较强的目标时，往往难以兼顾局部精细度与整体一致性^[4,28]。作为卷积式编码器—解码器架构的典型代表，U-Net 的基本结构如图2.1所示。

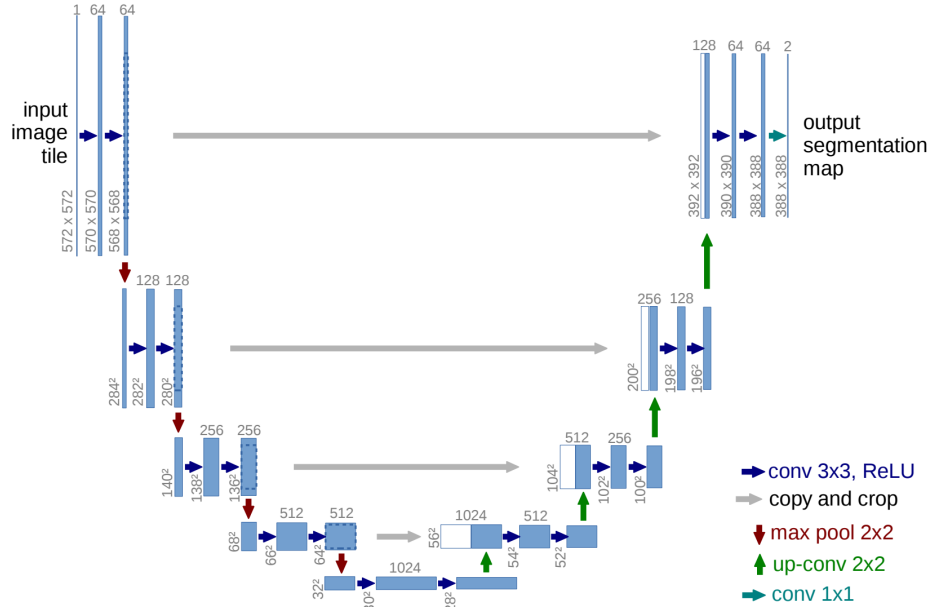


图 2.1 U-Net 模型总览图^[4]

第二类方法是以 Transformer 或 CNN 与 Transformer 异构融合架构为代表的全局建模网络，如 TransUNet、Swin-Unet 和 SegFormer 等^[34,35,56]。这类方法借助自注意力机制建立长距离像素或区域之间的依赖关系，能够在更大范围内整合上下文信息，从而在复杂场景下更好地维持目标整体轮廓的连续性与结构一致性。TransUNet 将 Transformer 编码器与 U-Net 解码器相结合，使模型在全局语义建模的同时保留一定的局部细节恢复能力^[34]；Swin-Unet 则进一步采用层次化窗口注意力机制，在控制计算开销的同时实现多尺度特征提取^[35]；SegFormer 通过层次化 Transformer 编码器与轻量级 MLP 解码器实现了较好的精度与效率平衡，并表现出较强的跨分辨率适应能力^[56]。相较于传统卷积网络，这类模型在刻画长程依赖、整合全局上下文以及建模复杂结构关系方面具有天然优势，对于弱边界、目标黏连、形状破碎或轮廓不完整等问题具有更强的鲁棒性^[34,56]。然而，Transformer 类方法通常对训练样本规模、参数初始化、优化策略及监督信号质量更为敏感，在标注不足、小样本学习或弱监督场景下，其优势并不总能稳定发挥，且模型训练和迁移过程往往具有更高的不确定性^[3,34,57]。作为 CNN 与 Transformer 异构融合架构的代表模型，TransUNet 的基本结

构如图2.2所示。

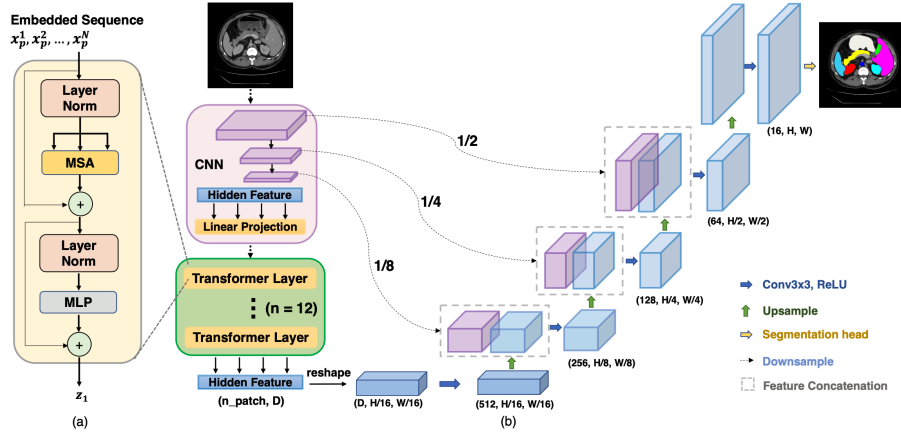


图 2.2 TransUNet 模型总览图^[34]

尽管上述深度语义分割模型已经在自然图像、医学影像和部分工业视觉任务中取得了显著进展，但对于医学影像与材料显微图像这类复杂纹理数据，仅依赖主干网络结构的增强仍难以根本解决弱边界、纹理异质性强、目标尺度变化大以及相邻结构黏连等问题。一方面，卷积网络虽然有利于保留局部边缘与纹理信息，但对长距离依赖和整体结构约束的表达能力不足；另一方面，Transformer 结构尽管能够增强全局上下文建模，却往往对高质量监督和较大规模训练数据存在较强依赖，在有限标注条件下容易出现训练不稳定、伪边界放大或局部细节缺失等现象^[3,35,57]。对于复杂纹理图像分割而言，主干网络设计固然重要，但更关键的问题在于如何在有限标注条件下，通过可靠伪监督构建、结构信息传播与边界约束进一步释放深度模型的表征潜力。也正因如此，近年来弱监督分割、伪标签优化、边界感知学习以及基于结构先验的信息传播策略逐渐成为复杂图像分割研究中的重要方向^[9,10,41,58]。

2.2 弱监督分割与涂鸦标注概述

在语义分割任务中，监督信息的完整程度直接影响模型对目标区域空间结构的学习能力。传统全监督分割依赖逐像素精确标注，虽然能够提供最充分的类别与边界信息，但在医学影像分割任务中，这种标注方式通常需要较强的专业知识支持，标注代价高且一致性难以保证^[14,59]；在材料显微图像分析中，受制于实验制样、数据获取和人工判读成本，全像素标注同样难以大规模开展^[15,16]。因此，在保证分割性能的同时尽可能降低标注代价，已成为复杂纹理图像分割研究中的一个重要问题。

弱监督语义分割的基本思想，是利用不完整、不精确或稀疏的监督信号替代完整像素级标注，引导模型逐步恢复目标区域的空间分布。与全监督学习相比，弱监督学习并不要求在训练开始阶段就获得完整掩码，而是通过监督传播、伪标签更新、一致性约束以及结构先验利用等机制，使模型从有限标注中不断挖掘潜在的区域归属信息^[9,11,37]。其本质是在标注成本与分割精度之间寻求平衡，即在减少像素级精细标注依赖的同时，尽可能保持模型对目标区域和边界结构的分割能力。

从监督形式来看，弱监督分割通常包括图像类别标注、包围盒标注、点标注和涂鸦标注等类型。图像类别标注仅提供目标类别是否存在的信息，这类方式在弱监督分割早期研究中被广泛采用^[37]，但由于缺乏明确的空间约束，往往难以恢复完整区域；包围盒能够提供大致位置范围，相关方法如 **BoxSup** 已验证其可用于分割训练^[38]，但矩形框内常混入大量背景，从而容易引入噪声；点标注能够给出局部类别信息，**Bearman** 等人的工作表明其能够以极低成本提供监督^[39]，但对目标形状、边界和连通结构的刻画能力有限。相比之下，涂鸦标注通过在目标内部或背景区域绘制少量稀疏线条，为模型提供了更连续的局部结构线索和更明确的区域归属信息，因而在标注成本与监督有效性之间取得了较好的平衡^[9,11,40,41]。图2.3展示了在一幅图像上分别使用图像类别标注、包围盒标注、点标注和涂鸦标注四种方式进行标注的实例。

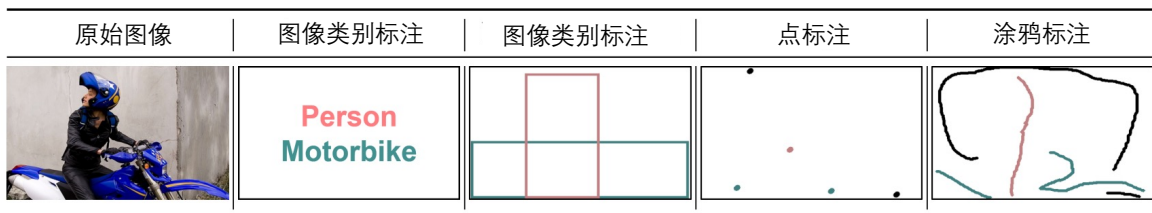


图 2.3 图像类别标注、包围盒标注、点标注和涂鸦标注示意图^[60]

对于医学影像与材料显微图像这两类复杂纹理数据而言，涂鸦标注具有更突出的适用性。一方面，这类图像普遍存在边界模糊、组织或相区黏连、纹理异质性强等问题，严格的像素级边界往往难以稳定界定，而涂鸦标注允许标注者避开歧义较强的边缘区域，仅在确定性较高的位置提供监督，这一点在 **ScribbleSup** 及后续相关研究中已得到体现^[9,10]。另一方面，医学器官、病灶区域以及材料组织常具有细长、分叉、嵌套或不规则扩展等复杂形态，涂鸦线条能够在一定程度上反映目标的延展方向和局部结构走向，因此较点标注和包围盒更有利于复杂结构的学习^[11,40]。

然而，涂鸦标注虽然降低了标注成本，并不意味着分割任务本身得到了简化。由于监督仅覆盖少量局部区域，模型在训练初期往往只能从有限的高置信度像素中学习类别信息，对于大范围未标注区域缺乏直接监督；同时，当图像中存在纹理重复、灰度相近、边界模糊或结构黏连等现象时，稀疏监督还可能在传播过程中引入错误累积，进而影响伪标签质量和模型稳定性。已有研究表明，无论是基于图传播的方法^[40,41]，还是基于伪标签和一致性学习的方法^[12,47-49]，都需要围绕监督稀疏带来的噪声传播问题进行专门设计。因此，在涂鸦监督分割中，核心问题并不只是如何使用稀疏标注，而是如何围绕有限监督构建可靠的密集伪监督，并将局部已知的结构信息稳定传播到未标注区域。

涂鸦标注之所以在弱监督分割中具有重要研究价值，不仅在于其能够显著降低医学与材料领域的标注负担，更在于它为少量确定性监督引导大范围未标注区域学习提供了可行基础。围绕涂鸦监督展开的方法设计，实质上需要回答如何构建可靠伪监督以及如何实现结构信息有效传播这两个关键问题。基于这一目标，下一节将进一步从伪标签生成与一致性学习的角度，对稀疏监督条件下的密集监督构建机制进行分析，并为后续医学影像分割与材料显微图像分割两类方法的展开奠定理论基础^[11,12,48,50]。

2.3 伪标签生成与一致性学习概述

在弱监督语义分割任务中，稀疏标注通常只能为少量像素提供直接监督，而图像中绝大多数区域仍处于未标注状态，对于涂鸦监督场景而言，这一问题尤为突出。尽管涂鸦标注能够提供一定的结构先验和类别锚点，但其覆盖范围远小于目标真实区域，无法直接支撑模型学习完整的边界、形状与区域连通性。因此，如何将有限的显式监督扩展为对未标注区域的有效约束，成为弱监督分割性能提升的关键。围绕这一目标，伪标签生成与一致性学习逐渐成为弱监督与半监督分割中最核心的两类技术路径^[61-64]。前者强调为未标注像素构造可利用的监督信号，后者强调约束模型在不同扰动条件下保持预测稳定，二者相互补充，共同构成了稀疏监督向致密预测过渡的重要理论基础。

伪标签最早源于半监督学习，其基本思想是在人工标注不足的条件下，利用模型对未标注样本的预测结果构造辅助监督信号^[61,62]。对于语义分割任务而言，伪标

签的对象不再是整幅图像类别，而是每个未标注像素的类别估计。因此，在涂鸦监督分割中，伪标签的作用在于突破原始标注覆盖范围的限制，使模型能够从未标注区域中筛选出较可靠的像素作为额外监督目标，从而逐步将训练范围由局部标注扩展至目标主体、边界邻域及部分背景区域。随着训练迭代的推进，模型参数不断更新，伪标签质量也可能同步改善，从而形成预测与训练相互促进的迭代优化过程，其基本流程如图 2.4 所示。

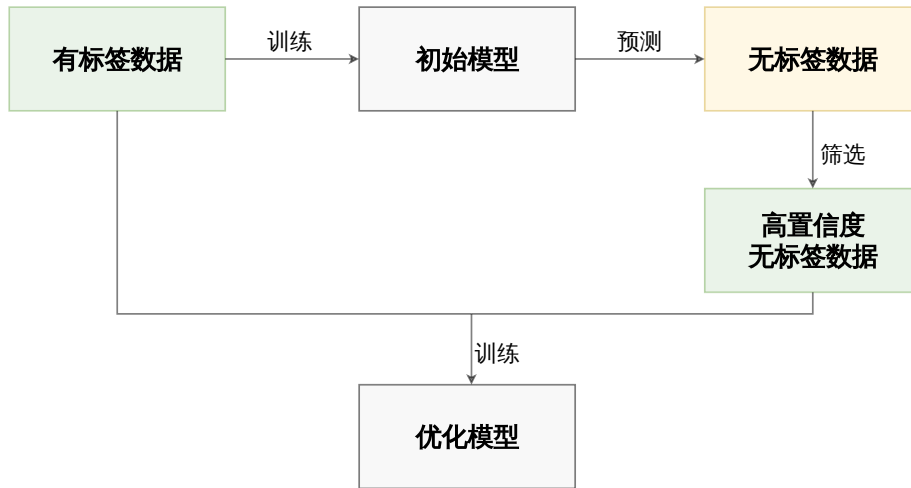


图 2.4 伪标签生成与迭代优化的基本流程示意图

然而，伪标签并不等同于真实标签，其有效性取决于可靠性与覆盖度之间的平衡。一方面，模型在训练早期尚未形成稳定判别能力，所生成的伪标签中可能包含较多错误预测；若直接将这些结果重新用于监督，噪声会在后续训练中不断累积，从而产生确认偏差。另一方面，若仅保留极少数高置信像素，虽然伪标签精度有所提高，但覆盖范围过小，又难以为模型提供足够的新信息。因此，如何在伪标签质量与覆盖范围之间取得平衡，始终是影响算法性能的关键问题。现有研究通常通过置信度筛选、多网络或多分支协同预测，以及空间平滑、边界约束和区域先验等机制提升伪标签的可用性^[12,40,41,46,47,62,65,66]。对于医学影像与材料显微图像这类复杂纹理数据而言，这些机制尤为必要，因为弱对比、边界模糊和局部纹理相似等特征会显著增加边界附近的误分割风险。

已有研究从自训练策略、协同训练、置信度筛选及噪声处理等多个维度对半监督语义分割中的伪标签方法进行了系统综述，指出在伪标签质量与覆盖范围之间取得平衡是提升方法性能的核心问题之一^[58]。

与伪标签生成侧重于为未标注区域构造监督信号不同，一致性学习更强调约束模型在不同扰动条件下保持预测稳定。一致性学习的基本假设是：对于同一幅图像，当其受到不改变语义本质的扰动时，模型输出应尽可能保持一致。也就是说，即使图像经历旋转、翻转、颜色扰动、噪声注入或区域混合等变化，其目标类别和主要结构通常不会发生本质改变，因此鲁棒的分割模型应对这些变化保持稳定预测。对于弱监督任务而言，这一思想尤为重要，因为在缺少密集真值标签时，一致性约束能够为未标注像素提供一种间接监督形式，即通过约束不同扰动视图之间的输出一致性来限制模型决策边界的不稳定波动^[62-64,67]。

现有研究中，一致性学习通常可以概括为基于网络扰动的一致性和基于输入扰动的一致性两类^[62-64]。前者通过不同网络结构、不同参数状态、不同解码分支或教师-学生模型对同一输入产生多个预测结果，并要求这些结果在未标注区域尽可能保持一致。这类方法能够利用不同结构之间的归纳偏置差异提高监督质量，尤其是在 CNN 与 Transformer 等异构结构联合使用时，可在一定程度上兼顾局部纹理细节与全局结构信息^[48-50,65,66]。后者则强调模型在面对不同输入视图时应保持输出稳定，常见形式包括几何变换、颜色扰动、噪声注入、混合增强以及弱增强、强增强策略等。对于语义分割任务而言，这类方法通常以弱增强图像上的预测结果作为伪监督，去约束强增强图像上的预测输出，从而提升模型在复杂数据变化下的鲁棒性与泛化能力^[12,47,62,68]。

对于涂鸦监督分割而言，伪标签生成与一致性学习并不是彼此孤立的模块，而往往是相互依赖、共同演化的。伪标签为未标注区域提供可利用的监督载体，而一致性学习则为这些监督信号提供稳定性约束与可靠性保障。当模型在多个网络分支、多个输入视图或多个扰动条件下对同一区域给出相近预测时，该区域通常可以被视为更可信的伪监督来源；反之，当预测差异较大时，该区域往往对应边界模糊、类别混淆或结构复杂的困难像素，需要被谨慎处理。因此，一致性学习不仅是一种正则化手段，也可以作为伪标签质量评估与筛选的重要依据。

后续两章的方法设计均建立在上述理论基础之上。对于复杂纹理图像的涂鸦监督分割任务而言，稀疏涂鸦标注只能提供局部确定性监督，大量未标注区域需要通过伪标签生成、一致性约束或边界先验等机制获得间接监督。因此，可靠伪监督构建与边界不确定性抑制构成了后续方法设计的共同理论基础。

在这一统一问题背景下，第三章的 FW2SS 方法面向医学影像这一典型复杂纹理

场景，主要从空间域一致性角度出发，通过 CNN-Transformer 异构双分支、弱到强一致性学习和像素级原型对比较准，提升伪标签生成过程的稳定性、目标结构表达能力以及复杂边界区域的特征判别能力。第四章的 MVBC 方法面向材料显微图像这一典型复杂纹理场景，主要从时间域一致性角度出发，通过按样本维护的历史预测队列、多历史投票筛选以及边界感知约束，增强小样本条件下伪标签筛选的稳定性，并抑制复杂相界附近的边界噪声传播。由此，FW2SS 和 MVBC 分别对应可靠伪监督构建问题在空间域和时间域中的两种实现路径，而不是彼此割裂的独立方法。

2.4 语义分割评价指标概述

语义分割任务的目标是对图像中每一个像素进行类别判定，因此其性能评价通常建立在预测结果与真实标签的逐像素比较基础之上。为更清晰地描述各类评价指标的含义，通常可借助混淆矩阵对分割结果进行统计分析。对于第 i 类像素，可将预测结果与真实标签之间的对应关系划分为四种情形：若某像素真实类别为第 i 类，且模型也将其预测为第 i 类，则记为真正例 (True Positive, TP_i)；若某像素真实类别不属于第 i 类，但模型错误地将其预测为第 i 类，则记为假正例 (False Positive, FP_i)；若某像素真实类别为第 i 类，但模型未将其预测为第 i 类，则记为假负例 (False Negative, FN_i)；当像素的真实类别与预测类别均不属于第 i 类时，则记为真负例 (True Negative, TN_i)。基于上述统计量，可以进一步定义语义分割中常用的交并比、Dice 系数以及边界误差类指标。

为了便于理解，以第 i 类为正类时的混淆矩阵统计关系如表 2.1 所示。

表 2.1 以第 i 类为正类时的混淆矩阵示意

真实类别 / 预测类别	第 i 类	非第 i 类
第 i 类	TP_i	FN_i
非第 i 类	FP_i	TN_i

需要指出的是，在语义分割任务中，真负例数量通常远大于真正例、假正例和假负例，尤其是在前景区域占比较小时更为明显。因此，后续常用评价指标主要由 TP_i 、 FP_i 与 FN_i 决定，而较少单独讨论 TN_i 。此外，像素准确率 (Pixel Accuracy, PA) 虽然是语义分割中的常见基础指标，能够从整体上反映像素分类的正确程度，但由于其易受类别分布不均衡的影响，在前景区域较小或背景占比较大的场景下往往

难以充分体现模型对目标区域和复杂边界的分割能力，因此，后续实验不将其作为主要评价指标。

为衡量预测区域与真实区域之间的重叠程度，通常引入交并比（Intersection over Union, IoU）指标。对于第 i 类像素，其交并比定义为预测区域与真实区域交集占并集的比例，即

$$IoU_i = \frac{TP_i}{TP_i + FP_i + FN_i}. \quad (2.1)$$

平均交并比（mean Intersection over Union, mIoU）定义为所有类别交并比的算术平均值，即

$$mIoU = \frac{1}{k} \sum_{i=1}^k IoU_i. \quad (2.2)$$

mIoU 是语义分割任务中最常用的核心评价指标之一。mIoU 同时考虑了漏分割与误分割的影响，通常能够更全面地反映模型对目标区域的定位能力与整体分割质量。特别是在前景与背景比例差异较大、目标边界复杂或类别分布不平衡的情况下，mIoU 往往更具代表性，因此常被用于衡量多类别分割任务的整体性能。

除交并比外，Dice 系数（Dice Similarity Coefficient, DSC）也是图像分割领域广泛采用的重要评价指标。对于第 i 类像素，其 Dice 系数定义为

$$Dice_i = \frac{2TP_i}{2TP_i + FP_i + FN_i}. \quad (2.3)$$

平均 Dice 系数（mean Dice, mDice）定义为全部 k 个类别 Dice 系数的算术平均值，即

$$mDice = \frac{1}{k} \sum_{i=1}^k Dice_i. \quad (2.4)$$

mDice 用于衡量预测区域与真实区域之间的重叠相似性，其所反映的区域重叠特征与 mIoU 较为接近，但对交集区域赋予了相对更高的权重，因此在医学影像分割、材料显微图像分割等任务中应用较为广泛。对于前景区域较小、类别分布不平衡或边界细节较为关键的场景，mDice 通常能够更敏感地反映模型对目标区域的分割效果。

除区域重叠类指标外，在医学影像分割任务中，边界误差类指标同样具有重要意义。对于器官或病灶分割而言，即使预测区域与真实区域具有较高的重叠程度，若局

部边界存在明显偏移，仍可能影响结构测量与临床分析。因此，常进一步引入 **Hausdorff** 距离 (**Hausdorff Distance, HD**)，并采用其鲁棒形式——第 95 百分位 **Hausdorff** 距离 (**95th Percentile Hausdorff Distance, HD95**)——来评价预测边界与真实边界之间的偏差。

设预测分割区域的边界点集为 B_p ，真实标注区域的边界点集为 B_g 。边界点集 B_p 到 B_g 的有向 **Hausdorff** 距离定义为

$$h(B_p, B_g) = \max_{x \in B_p} \min_{y \in B_g} \|x - y\|, \quad (2.5)$$

其中， $\|x - y\|$ 表示边界点 x 与 y 之间的欧氏距离。类似地，可定义从 B_g 到 B_p 的有向 **Hausdorff** 距离 $h(B_g, B_p)$ 。则标准 **Hausdorff** 距离定义为

$$HD(B_p, B_g) = \max(h(B_p, B_g), h(B_g, B_p)). \quad (2.6)$$

由于标准 **Hausdorff** 距离对少量离群点较为敏感，在医学影像分割中通常采用更稳健的 **HD95** 指标。具体而言，先计算预测边界与真实边界之间双向最近邻距离构成的集合：

$$\mathcal{D}(B_p, B_g) = \left\{ \min_{y \in B_g} \|x - y\| \mid x \in B_p \right\} \cup \left\{ \min_{x \in B_p} \|y - x\| \mid y \in B_g \right\}. \quad (2.7)$$

然后取该距离集合的第 95 百分位数，定义为

$$HD95(B_p, B_g) = P_{95}(\mathcal{D}(B_p, B_g)), \quad (2.8)$$

其中， $P_{95}(\cdot)$ 表示第 95 百分位操作。与标准 **HD** 相比，**HD95** 在保留边界偏差度量能力的同时，能够减弱个别异常点对评价结果的过度影响，因此更适用于医学影像分割任务中的边界质量评估。**HD95** 的取值越小，说明预测边界与真实边界之间的偏差越小，模型的边界恢复效果越好。对于三维医学影像分割任务，边界点之间的距离通常基于体素在实际物理空间中的欧氏距离进行计算。

综合来看，**mIoU**、**mDice** 与 **HD95** 分别从区域重叠程度、区域相似性以及边界偏差等角度对语义分割结果进行评价。其中，**mIoU** 更关注预测区域与真实区域之间的重叠质量，通常被用作衡量多类别分割性能的核心指标；**mDice** 对小目标及类别不平衡问题更为敏感，在医学影像分割任务中应用较为广泛；**HD95** 则用于刻画预测边界与真实边界之间的偏差，对边界恢复质量具有较好的敏感性。基于不同任务的

评价需求，医学影像分割实验采用 **Dice** 作为主要评价指标，并结合 **HD95** 对边界质量进行辅助分析；材料显微图像分割实验则采用 **mIoU** 与 **mDice** 作为主要评价指标，以综合评估模型的区域分割质量与类别区分能力。

2.5 本章小结

本章围绕后续研究所涉及的相关理论与关键技术进行了系统阐述。首先，回顾了基于深度学习的语义分割发展脉络，分析了 **FCN**、**U-Net** 及 **Transformer** 等典型模型在医学影像与材料显微图像等典型复杂纹理场景中的适用特点与局限。其次，介绍了弱监督分割的基本思想，重点论述了图像级标注、包围盒标注、点标注与涂鸦标注等不同监督形式的特点，并说明了涂鸦标注在医学影像与材料显微图像中的适用优势。随后，对伪标签生成与一致性学习的基本原理进行了概述，分析了其在稀疏监督条件下扩展监督范围、提升模型鲁棒性的作用机制。最后，介绍了语义分割任务中常用的评价指标，并重点分析了 **mIoU**、**mDice** 与 **HD95** 在区域重叠质量、区域相似性和边界偏差评价中的适用特点，为后续医学影像分割与材料显微图像分割两章的方法设计、实验设置与结果分析奠定了必要的理论基础。

第三章 基于一致性约束与对比学习的涂鸦监督医学影像分割方法

医学影像是复杂纹理图像中的典型类别之一，其目标区域通常具有较强的整体性与解剖关联性。医学影像在涂鸦监督条件下表现出形状信息学习不足、局部纹理刻画与全局结构建模难以协同以及复杂边界区域判别困难等问题：一方面，模型仅能从少量局部标注中学习类别信息，难以充分恢复目标的整体形状与结构关系；另一方面，局部纹理变化、组织黏连与弱边界又使局部特征难以稳定支撑整体结构恢复和边界判别。因此，本章聚焦医学影像这一典型场景，重点讨论在仅有涂鸦标注的条件下，如何有效提升形状信息学习能力，协调局部纹理刻画与全局结构建模，并增强复杂边界区域的特征判别能力。围绕这一思路，本章提出一种基于一致性约束与对比学习的涂鸦监督医学影像分割方法 **FW2SS** (**From Weak Scribble to Strong Segmentation**)。该方法旨在弱涂鸦监督条件下，通过一致性伪监督与特征判别校准，将稀疏、局部的涂鸦标注转化为更强、更完整的分割学习信号。具体而言，**CNN** 与 **Transformer** 异构双分支结构主要用于增强局部纹理与全局结构的互补建模；涂鸦一致性匹配 (**Scribble Consistency Match, SCM**) 通过联合网络扰动与输入扰动生成更稳定的像素级伪标签，从而提升稀疏监督条件下的形状信息学习能力；像素级原型对比校准 (**Pixel-level Prototype Contrastive Calibration, PPCC**) 则主要用于增强复杂边界区域的特征判别能力。

3.1 方法概述

3.1.1 模型框架

本章方法在涂鸦监督条件下构建了一个面向医学影像语义分割的统一训练框架，如图3.1所示，框架由两个核心部分组成：基于 **CNN-Transformer** 异构分割网络的涂鸦一致性匹配模块和像素级原型对比校准模块。其中，涂鸦一致性匹配模块首先引入 **CNN** 与 **Transformer** 组成的异构分割网络，利用 **CNN** 对局部纹理、边界细节的刻画能力以及 **Transformer** 对全局结构、长距离依赖的建模能力，实现局部特征与全局结构信息的互补表达。在训练过程中，输入图像先经过弱扰动生成弱增强视图，并

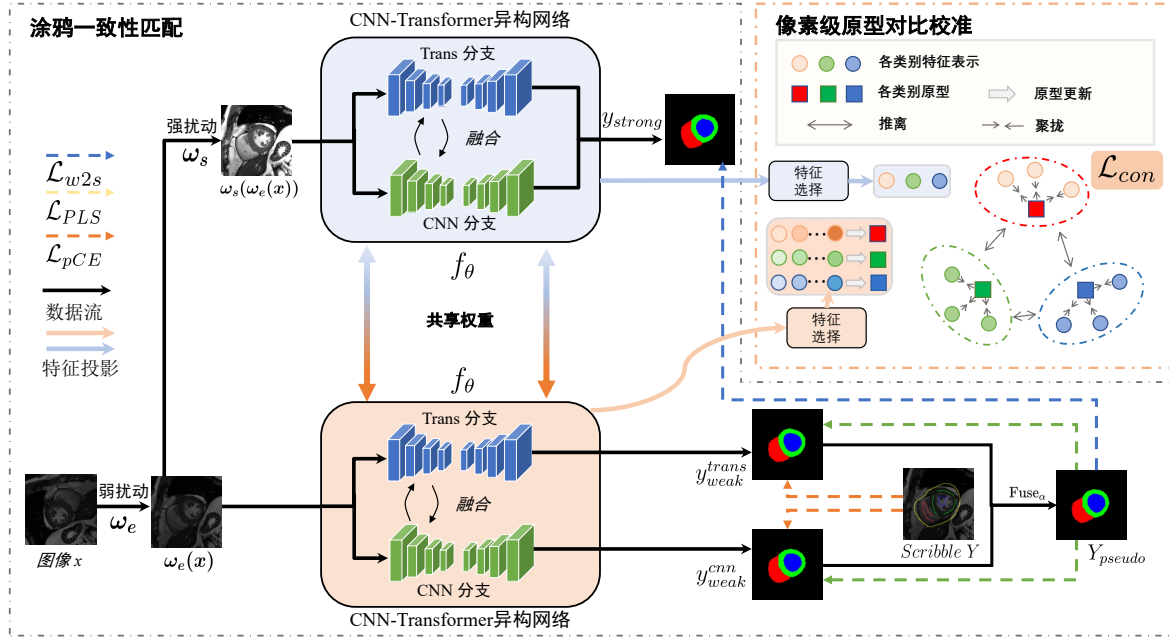


图 3.1 FW2SS 方法的总体框架图

送入异构分割网络，分别得到 CNN 解码器与 Transformer 解码器的预测结果；随后通过动态融合两类预测结果生成像素级共识伪标签，以缓解涂鸦标注稀疏导致的形状信息学习不足问题。在此基础上，方法进一步对弱扰动图像施加强扰动，并将其输入同一异构分割网络，通过弱扰动视图生成的伪标签约束强扰动视图的预测结果，从而同时实现网络扰动一致性与输入扰动一致性，提升模型在复杂扰动条件下的结构恢复能力和鲁棒性；像素级原型对比校准模块依据弱扰动视图下 CNN 解码器与 Transformer 解码器的预测一致性以及像素置信度筛选高可靠前景像素特征，并利用这些特征构建和更新前景类别原型。随后，以前景类别原型作为稳定的类别表征锚点，对强扰动分支中的高置信前景像素特征进行对比校准，使其靠近对应类别原型并远离其他类别原型。通过这种方式，模型能够进一步增强不同前景类别之间的特征可分性，尤其提升复杂边界和弱边界区域的判别能力。最终，涂鸦一致性匹配模块侧重于在稀疏涂鸦监督下生成稳定可靠的伪标签并学习完整结构信息，像素级原型对比校准模块则进一步优化前景特征表示与边界判别能力，二者共同提升涂鸦监督医学影像分割的准确性与稳定性。

为便于后文展开，现对涂鸦监督任务作如下数学定义：设数据集为

$$D = \{(x_n, s_n)\}_{n=1}^N, \quad (3.1)$$

其中, $x_n \in \mathbb{R}^{H \times W}$ 表示第 n 幅输入图像, s_n 表示与其尺寸相同的涂鸦监督标签图。本章统一约定: 标签值 0 表示背景, $1, \dots, C$ 表示 C 个前景目标类别, 255 表示未标注像素。因此, 网络在每个像素位置输出与标签值 $0, 1, \dots, C$ 对应的 $C+1$ 个语义通道, 而标签值 255 仅作为忽略标签使用, 不对应任何预测类别。需要特别说明的是, 在后续像素原型对比校准中, 仅对预测为前景类别且通过置信度筛选的像素计算原型对比损失, 背景像素不参与原型构建、原型更新与对比损失计算。本章中, 符号 C 始终表示前景类别数, 不包含背景类; 因此, 凡涉及原型的类别索引均取 $1, \dots, C$, 凡涉及分割输出的类别索引均取 $0, \dots, C$ 。

3.1.2 CNN-Transformer 异构网络

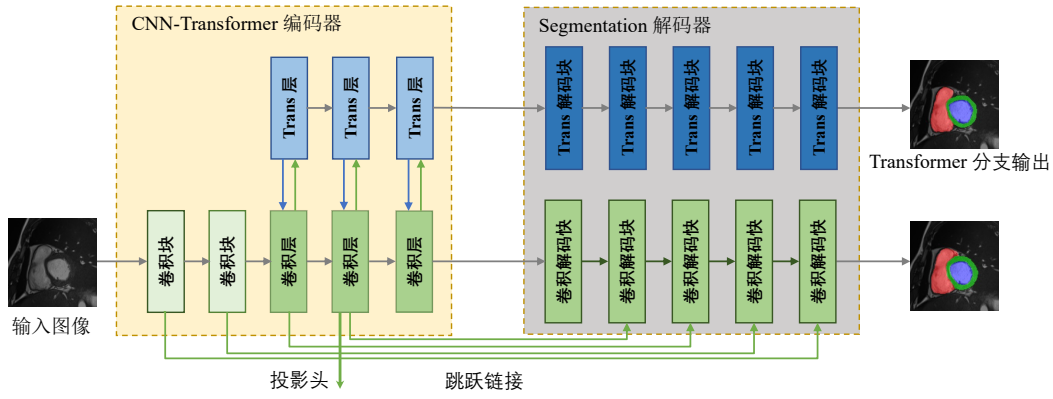


图 3.2 CNN 与 Transformer 异构网络结构示意图

卷积神经网络能够通过卷积核有效捕捉边界和纹理等局部细节, 因此在图像分割领域得到了广泛应用。同时, Transformer 架构通过自注意力机制能够捕捉图像的全局特征, 有助于识别形状、轮廓等结构特征。由于涂鸦监督仅提供有限的局部标注信息, 模型在训练过程中往往难以充分学习目标的整体形状与长距离结构关联, 单纯依赖局部感受野也难以兼顾局部纹理刻画与全局结构建模。因此, 在涂鸦监督医学影像分割任务中, 引入 Transformer 的全局建模能力有助于缓解仅依赖局部涂鸦监督时形状结构学习不足的问题。需要说明的是, 本文采用 CNN 与 Transformer 组成的异构双分支网络, 并非假设两类分支在所有阶段均能够实现无冲突的简单互补, 而是利用二者在感受野、归纳偏置和结构表征上的差异, 在训练阶段构造具有差异性的预测视角和一致性约束信号。如图 3.2 所示, 该网络由 CNN 分支与 Transformer 分支构成, 并分别配置卷积解码器与 Transformer 解码器; 其中, CNN 分支的编码器

与解码器之间通过跳跃连接传递特征。类似的网络架构也曾应用于近期的相关研究中^[49,50]。

CNN 分支编码器由多个卷积模块组成，每个模块包含一系列卷积操作和残差连接，其设计方式类似于残差网络^[69]。卷积操作在逐步对特征图分辨率进行下采样的同时增加通道数，从而提取图像的多尺度特征。下采样通过最大池化或卷积步长来减小空间维度，并保留关键信息。为获得适用于后续原型对比学习的特征表示，本章在 CNN 分支编码器的倒数第二个卷积模块之后显式引入投影头。在结构上，投影头由两个 1×1 卷积层组成，其中间插入批量归一化层和线性整流激活函数，最后进行 L_2 归一化。该模块将高维语义特征映射到一个紧凑的低维嵌入空间中，作为像素原型对比校准模块构建原型并计算对比损失的特征来源。

CNN 分支将每一层提取的局部特征传递给 Transformer 分支，从而增强 Transformer 的局部感知能力。这里采用的跨分支特征传递机制类似于现有文献中提出的特征耦合单元^[70]。同时，Transformer 分支编码器包含与 CNN 分支相同数量的模块，每个模块由一个多头自注意力机制和多层感知机模块组成。多头自注意力机制对图像中像素之间的关系进行建模，从而捕捉其全局特征。同样地，来自 Transformer 分支的全局特征也会被回传至 CNN 分支，以提升其全局感知能力。

在解码器阶段，网络逐步恢复特征图的分辨率以生成精细的分割结果。该网络同时使用卷积解码器和 Transformer 解码器处理由编码器提取的特征表示。卷积解码器主要处理来自卷积编码器的局部特征，通过多级上采样逐步恢复高分辨率特征。卷积解码器还采用了类似于 U-Net^[4] 的跳跃连接，将特征直接从不同的编码器层传递到对应的解码器层。Transformer 解码器主要处理来自 Transformer 编码器的全局特征。这些全局特征捕捉了远处像素之间的依赖关系，有助于建模形状和轮廓等大尺度结构信息。为了保持两类解码器预测视角的差异性，本文不在解码阶段直接进行特征融合，而是分别保留 CNN 解码器与 Transformer 解码器的输出，用于训练阶段的伪标签生成、一致性筛选和原型构建。需要指出的是，这种双分支设计主要服务于训练阶段的伪监督构建，并不意味着在推理阶段对两类输出进行简单平均一定能够获得最优结果。

3.1.3 涂鸦一致性匹配

在半监督学习常用的弱到强训练策略中，弱扰动视图在训练过程中为每幅图像生成的预测结果，通常可作为强扰动视图的伪监督信号。然而，涂鸦监督分割任务与半监督学习任务存在显著差异。由于缺乏用于训练教师网络的完整标签，仅由单一弱扰动视图生成的伪标签质量通常较低。为解决这一问题，本章方法同时考虑网络扰动与输入扰动的一致性学习机制，并构建涂鸦一致性匹配模块，用于在涂鸦监督条件下联合建模网络扰动一致性与输入扰动一致性。

如图 3.1 所示，在弱扰动视图构建过程中，本章方法采用旋转或翻转等简单的空间变换对图像进行数据增强。给定一幅图像 x ，对其施加强扰动 ω_e 以生成 $\omega_e(x)$ 。随后将该受扰动图像输入至 CNN-Transformer 异构网络 f_θ 中，分别获得来自卷积解码器的分割输出 $y_{\text{weak}}^{\text{cnn}}$ 和来自 Transformer 解码器的分割输出 $y_{\text{weak}}^{\text{trans}}$ 。二者均为像素级分类 logits，其通道定义与前文一致。

为避免记号混淆，本章统一将网络解码器的原始输出记为 logits y ，将其经 softmax 归一化后的结果记为概率图 P 。下文中的 Dice 损失均作用于概率图，交叉熵损失则采用基于 logits 的标准形式。为提升伪标签的稳定性，本章方法对来自 CNN 解码器与 Transformer 解码器的弱扰动预测结果进行动态融合。具体而言，首先将两个解码器的输出经过 softmax 归一化，分别记为

$$P_{\text{weak}}^{\text{cnn}} = \text{softmax}(y_{\text{weak}}^{\text{cnn}}), \quad P_{\text{weak}}^{\text{trans}} = \text{softmax}(y_{\text{weak}}^{\text{trans}}), \quad (3.2)$$

在第 t 次训练迭代中，引入动态融合权重

$$\alpha^{(t)} \sim \mathcal{U}(0, 1), \quad (3.3)$$

其中， $\alpha^{(t)}$ 表示在当前迭代中为 CNN 解码器分配的融合权重，而 $1 - \alpha^{(t)}$ 则对应 Transformer 解码器的融合权重。于是，融合后的弱扰动预测概率可表示为

$$P_{\text{fuse}}^{(t)} = \alpha^{(t)} P_{\text{weak}}^{\text{cnn}} + (1 - \alpha^{(t)}) P_{\text{weak}}^{\text{trans}}, \quad (3.4)$$

基于该融合概率分布，伪标签定义为

$$Y_{\text{pseudo}}^{(t)} = \arg \max (P_{\text{fuse}}^{(t)}), \quad (3.5)$$

令融合权重在训练过程中动态变化,有助于模型在不同迭代阶段利用 CNN 分支与 Transformer 分支之间的预测差异构造更具约束作用的伪监督信号。由于卷积神经网络具备较强的局部归纳偏置,而 Transformer 更擅长捕捉长距离依赖与整体轮廓信息,两类分支的加权融合能够在训练阶段一定程度上缓解单一预测视角带来的偏差,使所得伪标签在结构完整性和局部一致性之间取得相对平衡。因此, $Y_{\text{pseudo}}^{(t)}$ 更适合被理解为一种训练阶段的动态共识伪标签,而非推理阶段最优输出形式的直接等价。不同于半监督设定下在两个独立优化网络之间交换伪标签的交叉教学方法^[64],涂鸦一致性匹配模块在同一 CNN-Transformer 异构网络内部利用异构解码器生成单一共识伪标签,并将其与弱到强的输入扰动相结合,从而在稀疏涂鸦标注条件下稳定伪监督过程。

由动态融合得到的伪标签 $Y_{\text{pseudo}}^{(t)}$ 同时约束两个弱扰动解码器输出,使二者在共享伪监督下向一致的结构预测方向收敛。该伪监督损失函数可定义为:

$$L_{\text{PLS}} = \frac{1}{2} \left(L_{\text{dice}} \left(P_{\text{weak}}^{\text{cnn}}, Y_{\text{pseudo}}^{(t)} \right) + L_{\text{dice}} \left(P_{\text{weak}}^{\text{trans}}, Y_{\text{pseudo}}^{(t)} \right) \right) \quad (3.6)$$

其中, L_{dice} 表示 Dice 损失函数,用于衡量概率预测图与由伪标签生成的 one-hot 标注之间的空间相似度。

在强扰动视图构建过程中,基于经弱扰动得到的图像 $\omega_e(x)$,进一步施加强扰动策略 ω_s 。该强扰动策略主要作用于像素值层面的变化。强扰动策略从 RandAugment^[68] 的增强操作中选取,并在人工设计的像素值扰动策略集合中抽取多种变换进行组合应用。因此,强扰动分支接收到的输入图像为 $\omega_s(\omega_e(x))$ 。将该二次受扰动图像输入至 CNN-Transformer 异构网络中,分别得到来自 CNN 解码器和 Transformer 解码器的强扰动输出 $y_{\text{strong}}^{\text{cnn}}$ 和 $y_{\text{strong}}^{\text{trans}}$ 。二者均为像素级分类 logits,其通道定义与弱扰动预测输出一致。进一步,将两个强扰动输出经 softmax 归一化后按 1:1 进行平均,得到最终用于监督和后续对比校准的强分支概率输出:

$$P_{\text{strong}}^{\text{cnn}} = \text{softmax}(y_{\text{strong}}^{\text{cnn}}), \quad P_{\text{strong}}^{\text{trans}} = \text{softmax}(y_{\text{strong}}^{\text{trans}}), \quad (3.7)$$

$$P_{\text{strong}} = \frac{1}{2} (P_{\text{strong}}^{\text{cnn}} + P_{\text{strong}}^{\text{trans}}), \quad (3.8)$$

弱扰动视图生成的伪标签 $Y_{\text{pseudo}}^{(t)}$ 用于约束强扰动视图的输出,通过强制输入扰动一致性实现弱到强伪监督学习。此举确保了强扰动分支在面临剧烈扰动时仍能受到有

效约束，从而促进了对高鲁棒性特征表示的学习。跨分支伪监督损失函数可定义为：

$$L_{w2s} = L_{\text{dice}}(P_{\text{strong}}, Y_{\text{pseudo}}^{(t)}) \quad (3.9)$$

涂鸦一致性匹配模块同时约束了网络扰动与输入扰动的一致性，不仅增强了模型在复杂扰动条件下的鲁棒性，还有效促进了模型学习高鲁棒性的特征表示。

3.1.4 像素级原型对比校准

对于边界复杂或细节精细的目标，粗略的涂鸦标注使模型难以准确学习边界细节。强扰动分支通过像素值扰动对输入图像引入显著的视觉变化，进一步增加了边界像素类别判别的难度。为解决这一问题，本章提出像素级原型对比校准模块。该模块通过从弱扰动视图的预测结果中提取高置信度像素特征来构建前景类别原型（即表征相应前景类别特征的中心向量）。该模块促使强扰动分支的预测与对应前景类别原型保持一致，同时远离不属于该前景类别的区域。该模块有助于提升模型在不同扰动场景下对复杂边界的分割性能。下文所称“原型”均指前景类别原型，仅对应标签值 $1, \dots, C$ 。

像素级原型对比校准模块受 PCC 方法^[53]启发，但针对涂鸦监督医学分割任务进行了重新设计。不同于 PCC 方法通常利用弱监督线索（如由类别激活映射生成的掩码）估计原型，并在不同视图之间执行对称的跨视图对齐，本章方法仅利用弱分支中经一致性筛选的高置信度像素构建原型（即卷积解码器与 Transformer 解码器预测一致的像素），并将这些原型作为锚点，以单向方式（弱扰动 $\rightarrow$ 强扰动）校准强扰动特征，从而避免强视图中的噪声边界污染原型库。

首先，通过一致性和置信度筛选，从经弱扰动输入得到的两个输出 $y_{\text{weak}}^{\text{cnn}}$ 和 $y_{\text{weak}}^{\text{trans}}$ 中选择可靠像素。具体而言，由式 (3.2) 可得两个解码器对应的弱扰动预测概率图 $P_{\text{weak}}^{\text{cnn}}$ 和 $P_{\text{weak}}^{\text{trans}}$ 。随后生成可靠性掩码 M_i ，其中 i 表示单个像素索引，其定义如下：

$$\hat{c}_i^{\text{cnn}} = \arg \max(P_{\text{weak},i}^{\text{cnn}}), \quad \hat{c}_i^{\text{trans}} = \arg \max(P_{\text{weak},i}^{\text{trans}}), \quad (3.10)$$

$$M_i = \begin{cases} 1, & \text{若 } \hat{c}_i^{\text{cnn}} = \hat{c}_i^{\text{trans}} = c, c \in \{1, \dots, C\}, \\ & \text{且 } \max(P_{\text{weak},i}^{\text{cnn}}) \geq \tau_w, \max(P_{\text{weak},i}^{\text{trans}}) \geq \tau_w, \\ 0, & \text{其他情况。} \end{cases} \quad (3.11)$$

具体来说, $\hat{c}_i^{\text{cnn}}$ 和 $\hat{c}_i^{\text{trans}}$ 分别表示在像素 i 处由 CNN 分支和 Transformer 分支预测得到的类别标签。 τ_w 是置信度阈值; 当两个分支在像素 i 处均具有较高置信度, 且共同预测为同一前景类别时, 该像素才被视为可靠像素。

本章方法利用可靠性掩码 M_i 以及模型提取的特征构建高置信度前景类别原型, 并仅为前景类别维护原型记忆库, 记为

$$\mathcal{Q} = \{\mathcal{Q}_1, \mathcal{Q}_2, \dots, \mathcal{Q}_C\}, \quad (3.12)$$

在训练开始时, 每个前景原型 $\mathcal{Q}_c$ ($c = 1, \dots, C$) 被初始化为随机单位向量。在训练过程中, 利用可靠像素的聚合结果更新 $\mathcal{Q}$ 中对应前景类别的原型, 而当前小批量中未出现的前景类别保持不变。详细训练流程总结于算法 3.1。

在训练过程中, 原型采用迭代级更新机制进行更新。由于当 $M_i = 1$ 时有 $\hat{c}_i^{\text{cnn}} = \hat{c}_i^{\text{trans}}$, 因此记其共同预测的前景类别为 $\hat{c}_i$ 。对于每个前景类别 $c = 1, \dots, C$, 定义可靠像素集合

$$\Omega_c = \{i \mid M_i = 1, \hat{c}_i = c\}, \quad N_c = |\Omega_c|, \quad (3.13)$$

则当前小批量上的前景类别原型估计为

$$\hat{\mathcal{Q}}_c = \text{norm}\left(\frac{1}{N_c} \sum_{i \in \Omega_c} f_i\right), \quad c = 1, \dots, C, \quad (3.14)$$

其中, f_i 表示在像素 i 处由投影头提取的特征向量, $\text{norm}(\mathbf{z}) = \mathbf{z}/\|\mathbf{z}\|_2$ 表示对向量进行 L_2 归一化。若 $N_c = 0$, 则当前小批量不更新该前景类别原型, 以避免由噪声或空样本导致的漂移。

为增强前景类别原型的稳定性并降低内存开销, 本章方法采用指数滑动平均方法对原型进行更新。该更新方式能够平滑不同批次之间的波动, 减弱训练噪声及批间差异带来的影响, 从而提高原型更新的鲁棒性。更新公式如下:

$$\mathcal{Q}_c^{(t+1)} = \text{norm}\left((1 - \beta)\mathcal{Q}_c^{(t)} + \beta\hat{\mathcal{Q}}_c^{(t)}\right), \quad c = 1, \dots, C. \quad (3.15)$$

其中, $\mathcal{Q}_c^{(t)}$ 表示当前时刻的前景类别原型, $\mathcal{Q}_c^{(t+1)}$ 表示更新后的前景类别原型, β 是控制更新幅度的平滑系数。

为提高对比学习的稳定性, 在强扰动分支上进行第二阶段筛选。这里直接使用前文定义的最终强分支概率输出 P_{strong} 。记 $\hat{c}_i^s = \arg \max(P_{\text{strong},i})$ 为强扰动分支在像素 i 处的预测类别。仅当像素 i 同时满足 $\max(P_{\text{strong},i}) \geq \tau_s$ 且 $\hat{c}_i^s \in \{1, \dots, C\}$ 时, 特征向量 $\mathbf{v}_i$ 才会被选入原型对比损失的计算; 若 $\hat{c}_i^s = 0$, 则该像素被视为背景预测像素并直接排除。于是, 参与对比学习的像素集合定义为

$$S = \{i \mid \max(P_{\text{strong},i}) \geq \tau_s, \hat{c}_i^s \in \{1, \dots, C\}\}, \quad (3.16)$$

本章方法引入原型对比损失, 以增强特征可分性并提升分割性能, 尤其是在复杂边界和细节丰富的场景中。对于每个被选中的像素 $i \in S$, 其正样本类别由强扰动分支在该像素处的预测类别 $\hat{c}_i^s$ 给出。在此基础上, 该损失用于引导强扰动分支中的预测特征 $\mathbf{v}_i$ 学习更具判别性的表示, 使其趋近于其对应的前景类别原型, 同时远离其他前景类别原型。对于每个被选中的像素 $i \in S$, 逐像素原型对比损失 L_{con}^i 的定义如下:

$$L_{\text{con}}^i = -\log \frac{\exp(\text{sim}(\mathbf{v}_i, \mathcal{Q}_{\hat{c}_i^s})/\tau)}{\sum_{c=1}^C \exp(\text{sim}(\mathbf{v}_i, \mathcal{Q}_c)/\tau)}, \quad i \in S \quad (3.17)$$

其中, $\mathbf{v}_i$ 表示在强扰动分支中于选定像素 i 处获得的特征向量, $\mathcal{Q}_{\hat{c}_i^s}$ 表示其对应的前景类别原型; 分母对全部 C 个前景类别原型求和。 τ 是温度系数, 用于控制对比学习过程中相似度的缩放。

特征向量 $\mathbf{v}_i$ 与原型 $\mathcal{Q}_c$ 之间的余弦相似度定义为:

$$\text{sim}(\mathbf{v}_i, \mathcal{Q}_c) = \frac{\mathbf{v}_i \cdot \mathcal{Q}_c}{\|\mathbf{v}_i\| \|\mathcal{Q}_c\|}, \quad (3.18)$$

其中, $\mathbf{v}_i \cdot \mathcal{Q}_c$ 表示 $\mathbf{v}_i$ 与 $\mathcal{Q}_c$ 之间的点积, 而 $\|\mathbf{v}_i\|$ 和 $\|\mathcal{Q}_c\|$ 分别表示它们各自的欧几里得范数。

原型对比损失通过对集合 S 中所有选定像素的损失取平均来计算:

$$L_{\text{con}} = \frac{1}{|S|} \sum_{i \in S} L_{\text{con}}^i, \quad (3.19)$$

其中, $|S|$ 表示参与前景原型对比学习的前景像素总数; 若某一批次中不存在满足条件的前景像素, 则令 $L_{\text{con}} = 0$ 。

上述像素级原型对比较准模块的完整训练流程伪代码总结于算法 3.1。

3.1.5 损失函数

为缓解涂鸦标注稀疏性带来的挑战，监督学习阶段采用部分交叉熵损失。该方法忽略未标注像素，仅对已标注像素施加监督，从而减小标注噪声对模型训练的影响。

具体而言，对于每个样本 (x, s) ，经弱扰动输入得到的两个 logits 输出 $y_{\text{weak}}^{\text{cnn}}$ 和 $y_{\text{weak}}^{\text{trans}}$ 均通过交叉熵损失进行监督。损失函数定义如下：

$$L_{\text{pCE}}(s, y_{\text{weak}}^{\text{cnn}}, y_{\text{weak}}^{\text{trans}}) = \frac{L_{\text{ce}}(y_{\text{weak}}^{\text{cnn}}, s) + L_{\text{ce}}(y_{\text{weak}}^{\text{trans}}, s)}{2} \quad (3.20)$$

其中， L_{pCE} 表示部分交叉熵损失，用于监督弱扰动视图下的两个预测输出。交叉熵损失的具体形式如下：

$$L_{\text{ce}}(y, s) = - \sum_{i \in \Omega_l} \sum_{k \in K} s_i^k \log((\text{softmax}(y_i))^k) \quad (3.21)$$

其中， $K = \{0, 1, \dots, C\}$ 表示参与分割预测与部分交叉熵计算的语义类别集合，即该集合包含背景类 0 和全部前景类别 $1, \dots, C$ ，但不包含忽略标签 255； $\Omega_l = \{i \mid s_i \neq 255\}$ 表示涂鸦标注中已标注像素的集合； s_i^k 表示像素 i 属于类别 k 的 one-hot 标注值， $(\text{softmax}(y_i))^k$ 表示模型在像素 i 处预测其属于类别 k 的概率。

模型训练的总损失 L_{total} 定义为：

$$L_{\text{total}} = L_{\text{pCE}} + \lambda_1 L_{\text{PLS}} + \lambda_2 L_{\text{w2s}} + \lambda_3 L_{\text{con}} \quad (3.22)$$

3.2 实验分析

3.2.1 数据集介绍

所提出的方法在采用涂鸦监督设置的 ACDC^[71] 和 MSCMRseg^[72,73] 数据集上进行评估。

ACDC 数据集^[71] 包含来自 100 名患者的心脏 MRI 扫描图像，分割目标为左心室 (LV)、右心室 (RV) 和心肌 (MYO)。在模型训练中，采用 Valvano 等人^[11] 提供的涂鸦标注作为弱监督学习的标注信息。按照^[11,12,47] 的设置，所有扫描图像被随机

算法 3.1: 像素级原型对比校准伪代码

Input: $y_c^w, y_t^w, y_c^s, y_t^s \in \mathbb{R}^{B \times (C+1) \times H \times W}$, $f^w, f^s \in \mathbb{R}^{B \times D \times H \times W}$, $Q \in \mathbb{R}^{C \times D}$, 其中 Q 仅存储标签值 $1, \dots, C$ 对应的前景类别原型; $\tau_w, \tau_s, \beta, \tau, \epsilon$

Output: L_{con}, Q

- 1 $p_c \leftarrow \text{softmax}(y_c^w)$, $p_t \leftarrow \text{softmax}(y_t^w)$
- 2 $\hat{c}_c \leftarrow \arg \max(p_c)$, $\hat{c}_t \leftarrow \arg \max(p_t)$
- 3 $M \leftarrow (\hat{c}_c = \hat{c}_t) \wedge (\hat{c}_c \in \{1, \dots, C\}) \wedge (\max(p_c) \geq \tau_w) \wedge (\max(p_t) \geq \tau_w)$ // 仅保留标签值为 $1, \dots, C$ 的可靠前景像素
- 4 **for** $c \leftarrow 1$ **to** C **do**
 - 5 $\mathcal{I}_c \leftarrow \{(b, i, j) \mid M[b, i, j] \wedge \hat{c}_c[b, i, j] = c\}$
 - 6 **if** $|\mathcal{I}_c| > 0$ **then**
 - 7 $\hat{q}_c \leftarrow \text{norm}(\text{mean}_{(b, i, j) \in \mathcal{I}_c}(f^w[b, :, i, j]))$
 - 8 **else**
 - 9 $\hat{q}_c \leftarrow Q[c]$
 - 10 **end**
 - 11 $Q[c] \leftarrow \text{norm}((1 - \beta)Q[c] + \beta\hat{q}_c)$ // EMA update
- 12 **end**
- 13 $p_c^s \leftarrow \text{softmax}(y_c^s)$, $p_t^s \leftarrow \text{softmax}(y_t^s)$
- 14 $p_s \leftarrow \frac{1}{2}(p_c^s + p_t^s)$, $\hat{c}_s \leftarrow \arg \max(p_s)$
- 15 $\mathcal{S} \leftarrow \{(b, i, j) \mid \max(p_s[b, :, i, j]) \geq \tau_s \wedge \hat{c}_s[b, i, j] \in \{1, \dots, C\}\}$ // 仅保留标签值为 $1, \dots, C$ 的强分支高置信前景像素
- 16 **if** $|\mathcal{S}| = 0$ **then**
 - 17 **return** $0, Q$
- 18 **end**
- 19 $L_{\text{con}} \leftarrow 0$
- 20 **foreach** $(b, i, j) \in \mathcal{S}$ **do**
 - 21 $v \leftarrow \text{norm}(f^s[b, :, i, j]); c^+ \leftarrow \hat{c}_s[b, i, j]$
 - 22 $L_{\text{con}} \leftarrow L_{\text{con}} - \log \frac{\exp(\cos(v, Q[c^+])/\tau)}{\sum_{k=1}^C \exp(\cos(v, Q[k])/\tau) + \epsilon}$ // 仅对前景原型进行对比
- 23 **end**
- 24 $L_{\text{con}} \leftarrow L_{\text{con}}/|\mathcal{S}|$
- 25 **return** L_{con}, Q

划分为三个子集，其中 70 例用于训练，15 例用于验证，15 例用于测试。需要说明的是，上述 70/15/15 划分对应 ACDC 数据集的基础训练、验证与测试划分。对于后续主对比实验，实验设置遵循相关工作的常用弱监督比较协议，在 70 例训练病例中进一步采用其中 35 例带涂鸦标注的病例作为训练集进行比较，表 3.1 中“35 例涂鸦标注 + 35 例非配对掩码”和“35 例完整掩码”等设置，均沿用相关文献中的标准实验协议，以保证与已有方法之间的公平比较。

MSCMRseg 数据集^[72,73] 包含来自 45 名心肌病患者的延迟钆增强 MRI 扫描图像，其左心室 (LV)、心肌 (MYO) 和右心室 (RV) 的涂鸦标注由^[12] 提供。与非增强 MRI 图像的分割相比，增强 MRI 图像由于具有更高的复杂性，因此分割任务更具挑战性。为便于对比，按照^[12,47] 的实验设置，45 例扫描图像被随机划分为三个子集，其中 25 例用于训练，5 例用于验证，15 例用于测试。

3.2.2 实验设置

模型在 PyTorch 框架下实现，并在配备 NVIDIA GeForce RTX 3090 GPU (24 GB 显存) 的计算平台上完成训练。对于所有数据集，先将图像切片的像素强度归一化至 0-1，并将 ACDC 和 MSCMRseg 数据集中的图像尺寸分别统一调整为 256×256 和 212×212 。

在训练设置方面，模型在每个数据集上均训练 300 个轮次。学习率和权重衰减通过网格搜索确定，最终分别设为 0.001 和 0.0005。批大小及参数 τ_w 、 τ_s 和 β 分别设为 8、0.9、0.5 和 0.1，损失函数权重 ($\lambda_1, \lambda_2, \lambda_3$) 设为 (0.5, 0.25, 0.01)。训练初期，为保证记忆库中前景类别原型的质量，前 20 个轮次将 λ_3 设为 0；同时，伪标签融合权重 $\alpha^{(t)}$ 在每次迭代中从均匀分布 $\mathcal{U}(0, 1)$ 随机采样。在推理与评价方面，本章采用 CNN 解码器的输出作为最终分割结果，并统一使用 Dice 系数进行性能评估。

3.2.3 数据增强设置

数据增强阶段采用弱扰动和强扰动。弱扰动 ω_e 主要由空间变换构成，包括水平翻转、垂直翻转和固定角度范围内的随机旋转。相比之下，强扰动 ω_s 采用更复杂的像素值增强策略：从人工设计的像素值增强策略集合 Z 中，随机选择 R 种变换并进行组合。具体而言，强扰动基于 RandAugment^[68] 中的操作，默认 $R = 3$ 。像素值增强策略集合 Z 包含如下若干针对像素值的变换：

- **自动对比度增强**：一种对比度增强技术，通过调整图像中亮区与暗区之间的差异来增强视觉效果。调整强度从均匀分布 $U(0.5 - 0.5\delta, 0.5 + 0.5\delta)$ 中采样。
- **直方图均衡化**：一种通过均衡像素直方图来增强动态范围并提升细节可见性的技术。调整强度从均匀分布 $U(0.5 - 0.5\delta, 0.5 + 0.5\delta)$ 中采样。
- **色阶减少**：一种减少图像颜色层级数、从而简化其颜色表示的技术。颜色层级数从均匀分布 $U(3 - 3\delta, 3 + 3\delta)$ 中采样。
- **色彩增强**：一种调节图像饱和度、使颜色更加鲜艳或更加柔和的技术。调整强度从均匀分布 $U(1 - 0.9\delta, 1 + 0.9\delta)$ 中采样。
- **对比度增强**：一种调节图像对比度、增强亮暗区域差异的技术。调整强度从均匀分布 $U(1 - 0.9\delta, 1 + 0.9\delta)$ 中采样。
- **亮度增强**：一种调节图像亮度、以增强或减弱整体亮度感知的技术。调整强度从均匀分布 $U(1 - 0.9\delta, 1 + 0.9\delta)$ 中采样。
- **锐度增强**：一种调节图像锐度、增强细节清晰度的技术。调整强度从均匀分布 $U(1 - 0.9\delta, 1 + 0.9\delta)$ 中采样。

其中， $\delta \in (0, 1]$ 表示扰动强度，默认设置为 1。

3.2.4 对比实验

为全面验证本章所提方法的有效性，在 ACDC 和 MSCMRseg 两个心脏磁共振分割数据集上开展对比实验。实验主要围绕涂鸦监督场景展开，并与多种具有代表性的先进方法进行比较，包括：基于 U-Net 架构的不同涂鸦监督训练策略、基于输入扰动的一致性学习方法、基于网络扰动的伪监督方法，以及采用完整标注的全监督方法。此外，本章还在 ACDC 数据集上进一步比较了利用额外非配对完整标注的方法。

在评价指标方面，本章采用 Dice 系数对左心室 (LV)、心肌 (MYO) 和右心室 (RV) 的分割性能进行定量评估，并以三类目标的平均 Dice 值作为总体性能指标。更高的 Dice 分数表示模型分割结果与真实标注之间具有更高的一致性。基于上述实验设置，可以系统评估本章方法在弱监督医学影像分割任务中的性能表现。

(1) ACDC 数据集对比实验

表 3.1 给出了本章方法在 ACDC 数据集上的详细实验结果，并与多种先进方法进行了系统比较。对比方法主要包括以下几类：(1) 基于 U-Net^[4] 架构的不同涂鸦监督训练策略，例如 Unet_{pce} ^[9]、 Unet_{wpce} ^[40] 和 Unet_{crf} ^[74]；(2) 在相同涂鸦监督损失下

采用不同网络结构的方法, 例如 Unet^{++}_{pce} ^[75] 和 Unet^{++}_{pce} ^[5]; (3) 基于输入扰动的一致性学习方法, 包括 Co-mixup ^[76]、 CutMix ^[77]、 Cutout ^[78]、 PuzzleMix ^[79]、 MixUp ^[80]、 ShapePU ^[47] 和 CycleMix_S ^[12]; (4) 基于网络扰动的伪监督方法, 包括 DMPLS ^[48]、 ScribbleVC ^[49]、 ScribFormer ^[50], 以及近期提出的 QMaxViT-Unet^{+} ^[81]、 EFFDNet ^[82] 和 FDDSeg ^[83]; (5) 利用额外非配对完整标注的方法, 例如 Unet_D ^[11]、 PostDAE ^[84]、 ACCL ^[85] 和 MAAG ^[11]; (6) 采用完整标注的全监督方法, 包括 Unet_F ^[4]、 Unet^{+}_F ^[75]、 Unet^{++}_F ^[5]、 PuzzleMix_F ^[79]、 CycleMix_F ^[12] 和 nnUNet ^[30]。由表 3.1 的上半部分可知, 在涂鸦监督设置下, 本章方法相较于其他基于 U-Net 的训练策略、不同网络结构下的方法、输入扰动方法以及伪监督方法取得了更高的平均 Dice 分数。具体而言, FW2SS 在 ACDC 数据集上取得了 90.0% 的平均 Dice 分数。与表中此前表现最优的基于网络扰动的伪监督方法相比, 平均 Dice 提升了 1.2 个百分点; 与表中此前表现最优的基于输入扰动的一致性学习方法 CycleMix_S 相比, 平均 Dice 提升了 5.2 个百分点。此外, 在 LV、MYO 和 RV 三个类别上, FW2SS 的 Dice 分数均高于表中纳入比较的其他涂鸦监督方法, 表明 FW2SS 对不同心脏结构均具有较好的分割稳定性。

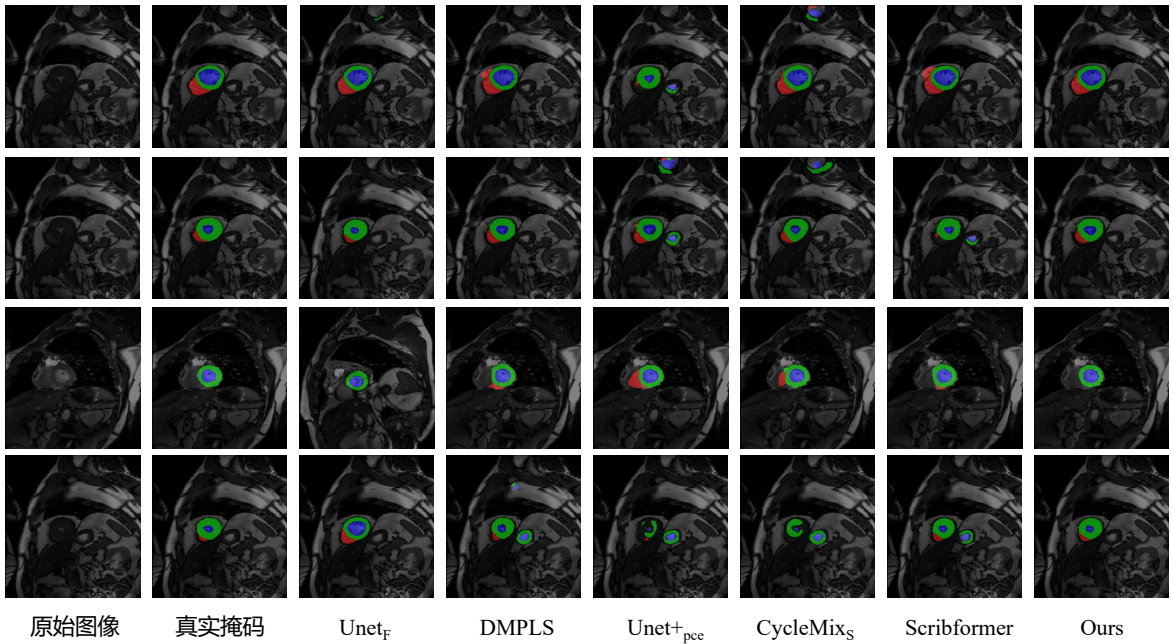


图 3.3 FW2SS 与其他先进方法在 ACDC 数据集上的定性比较

表 3.1 的最后部分进一步比较了 FW2SS 与若干全监督方法在 ACDC 数据集上的性能。在本文采用的二维切片训练、相同训练病例数量及统一预处理设置下,

表 3.1 FW2SS 与其他先进方法在 ACDC 数据集上的性能对比（涂鸦监督设置下最优结果以粗体表示）

ACDC					
方法	监督类型	左心室	心肌	右心室	平均值
35 例涂鸦标注					
Unet _{pce}	涂鸦标注	0.624	0.537	0.526	0.562
Unet _{wpce}	涂鸦标注	0.784	0.675	0.563	0.674
Unet _{crf}	涂鸦标注	0.766	0.661	0.59	0.672
Unet _{++pce}	涂鸦标注	0.785	0.746	0.725	0.752
Unet _{+++pce}	涂鸦标注	0.846	0.652	0.787	0.761
Co-mixup	涂鸦标注	0.622	0.702	0.621	0.648
CutMix	涂鸦标注	0.641	0.734	0.740	0.705
Cutout	涂鸦标注	0.832	0.754	0.812	0.800
Puzzle Mix	涂鸦标注	0.663	0.650	0.559	0.624
MixUp	涂鸦标注	0.803	0.753	0.767	0.774
ShapePU	涂鸦标注	0.860	0.791	0.852	0.834
DMPLS	涂鸦标注	0.908	0.849	0.865	0.874
CycleMix _S	涂鸦标注	0.883	0.798	0.863	0.848
ScribFormer	涂鸦标注	0.922	0.871	0.871	0.888
ScribbleVC	涂鸦标注	0.914	0.870	0.866	0.884
QMax ViT-Unet+	涂鸦标注	0.919	0.874	0.870	0.888
EFFDNet	涂鸦标注	0.856	0.830	0.833	0.840
FDDSeg	涂鸦标注	0.923	0.861	0.877	0.887
FW2SS (本章方法)	涂鸦标注	0.931	0.881	0.887	0.900
35 例涂鸦标注 + 35 例非配对掩码					
UNet _D	混合监督	0.753	0.597	0.404	0.585
PostDAE	混合监督	0.806	0.667	0.556	0.676
ACCL	混合监督	0.878	0.797	0.735	0.803
MAAG	混合监督	0.879	0.817	0.752	0.816
35 例完整掩码					
Unet _F	完整掩码	0.892	0.789	0.830	0.837
Unet _{++F}	完整掩码	0.849	0.817	0.792	0.820
Unet _{+++F}	完整掩码	0.875	0.771	0.798	0.815
Puzzle Mix _F	完整掩码	0.849	0.798	0.771	0.840
CycleMix _F	完整掩码	0.919	0.858	0.882	0.886
nnUNet	完整掩码	0.943	0.901	0.915	0.920

FW2SS 仅使用涂鸦标注进行训练，其平均 Dice 高于若干采用完整掩码训练的 U-Net 系列全监督基线方法，但仍低于本文纳入比较的强全监督方法 nnUNet。这一结果表明，所提方法能够在显著降低标注成本的条件下获得具有竞争力的分割性能，体现了其标注效率优势。需要强调的是，该比较结果仅适用于本文实验协议与所选基线范围，并不说明涂鸦监督方法在一般情况下优于充分优化的全监督方法。

图 3.3 展示了不同方法在 ACDC 数据集上的可视化分割结果。可以观察到，现有涂鸦监督方法由于监督信息不足，普遍存在形状不一致和过分割等问题。当背景区域与目标器官具有相似外观特征时，诸如 $\text{Unet}+_{pce}$ 等方法容易将部分背景误分割为器官区域。相比之下，本章方法得益于双重扰动一致性学习机制，能够更充分地学习目标结构的形状信息，即使在标注有限的情况下，仍可生成与真实标注高度接近的预测结果。

(2) MSCMRseg 数据集对比实验

表 3.2 给出了本章方法在 MSCMRseg 数据集上的对比实验结果。与 ACDC 数据集类似，对比方法主要包括两大类：一类是基于涂鸦监督的弱监督分割方法，另一类是基于完整标注的全监督分割方法。

对于涂鸦监督方法，又可以进一步分为三类：（1）基于 U-Net^[4] 架构并采用不同涂鸦监督损失的训练策略，例如 $\text{Unet}+_{pce}$ ^[75] 和 $\text{Unet}++_{pce}$ ^[5]；（2）基于输入扰动的一致性学习方法，包括 Co-mixup^[76]、CutMix^[77]、Cutout^[78]、PuzzleMix^[79]、MixUp^[80]、ShapePU^[47] 和 CycleMix_S^[12]；（3）基于网络扰动的伪监督方法，例如 DM-PLS^[48]、ScribbleVC^[49]、ScribFormer^[50]，以及近期提出的 QMaxViT-Unet+^[81]、EFFD-Net^[82] 和 FDDSeg^[83]。此外，全监督对比方法包括 Unet_F ^[4]、 $\text{Unet}+_F$ ^[75]、 $\text{Unet}++_F$ ^[5]、PuzzleMix_F^[79]、CycleMix_F^[12] 和 nnUNet^[30]。

由表 3.2 的前半部分可知，在涂鸦监督设置下，本章方法在 MSCMRseg 数据集上同样优于其他基于 U-Net 的训练策略、输入扰动方法以及伪监督方法。FW2SS 取得了 88.3% 的平均 Dice 分数，相比此前表现最优的基于网络扰动的伪监督方法 ScribbleVC 提升了 1.5 个百分点，相比基于输入扰动的一致性学习方法 ShapePU 提升了 3.1 个百分点。值得注意的是，FW2SS 在 LV、MYO 和 RV 三个分割类别上均取得了最优结果，表明其在复杂数据分布下仍具有较强的鲁棒性和泛化能力。

需要说明的是，本文与全监督方法的比较主要用于分析所提方法在低标注成本条件下的标注效率，而非证明涂鸦监督方法在一般意义上优于全监督方法。全监督

方法的性能会受到网络结构、训练策略、数据增强、超参数设置以及是否使用二维或三维信息等因素影响。因此，本文关于全监督基线的比较结论仅适用于当前数据划分、训练协议和纳入比较的方法范围。总体而言，FW2SS 的优势主要体现在仅依赖稀疏涂鸦标注的情况下，仍能够获得接近若干强监督基线的分割表现。

表 3.2 FW2SS 与其他先进方法在 MSCMRseg 数据集上的性能对比（涂鸦监督设置下最优结果以粗体表示）

MSCMRseg					
方法	监督类型	左心室	心肌	右心室	平均值
25 例涂鸦标注					
Unet+ _{pce}	涂鸦标注	0.494	0.583	0.057	0.378
Unet++ _{pce}	涂鸦标注	0.497	0.506	0.472	0.492
Co-mixup	涂鸦标注	0.356	0.343	0.053	0.251
CutMix	涂鸦标注	0.578	0.622	0.761	0.654
Cutout	涂鸦标注	0.459	0.641	0.697	0.599
Puzzle Mix	涂鸦标注	0.061	0.634	0.028	0.241
MixUp	涂鸦标注	0.610	0.463	0.378	0.484
ShapePU	涂鸦标注	0.888	0.813	0.854	0.852
DMPLS	涂鸦标注	0.876	0.779	0.823	0.826
CycleMix _S	涂鸦标注	0.870	0.739	0.791	0.800
ScribFormer	涂鸦标注	0.896	0.813	0.807	0.839
ScribbleVC	涂鸦标注	0.921	0.830	0.852	0.868
QMaxViT-Unet+	涂鸦标注	0.916	0.825	0.861	0.867
EFFDNet	涂鸦标注	0.899	0.808	0.799	0.835
FDDSeg	涂鸦标注	0.915	0.831	0.853	0.866
FW2SS (本章方法)	涂鸦标注	0.924	0.837	0.887	0.883
25 例完整掩码					
Unet _F	完整掩码	0.850	0.721	0.738	0.77
Unet+ _F	完整掩码	0.857	0.720	0.689	0.755
Unet++ _F	完整掩码	0.866	0.745	0.731	0.774
Puzzle Mix _F	完整掩码	0.867	0.742	0.759	0.789
CycleMix _F	完整掩码	0.864	0.785	0.781	0.810
nnUNet	完整掩码	0.944	0.882	0.880	0.902

图 3.4 展示了不同方法在 MSCMRseg 数据集上的可视化分割结果。相较于 ACDC 数据集，MSCMRseg 数据集的分割难度更高，因此其他方法，尤其是 Unet+_{pce}，更容易将大面积背景区域误分为器官区域。同样，CycleMix_S 在部分样本中也会将背景区域误分为 MYO，其他方法则不同程度地存在形状不一致和过分割问题。相比之下，

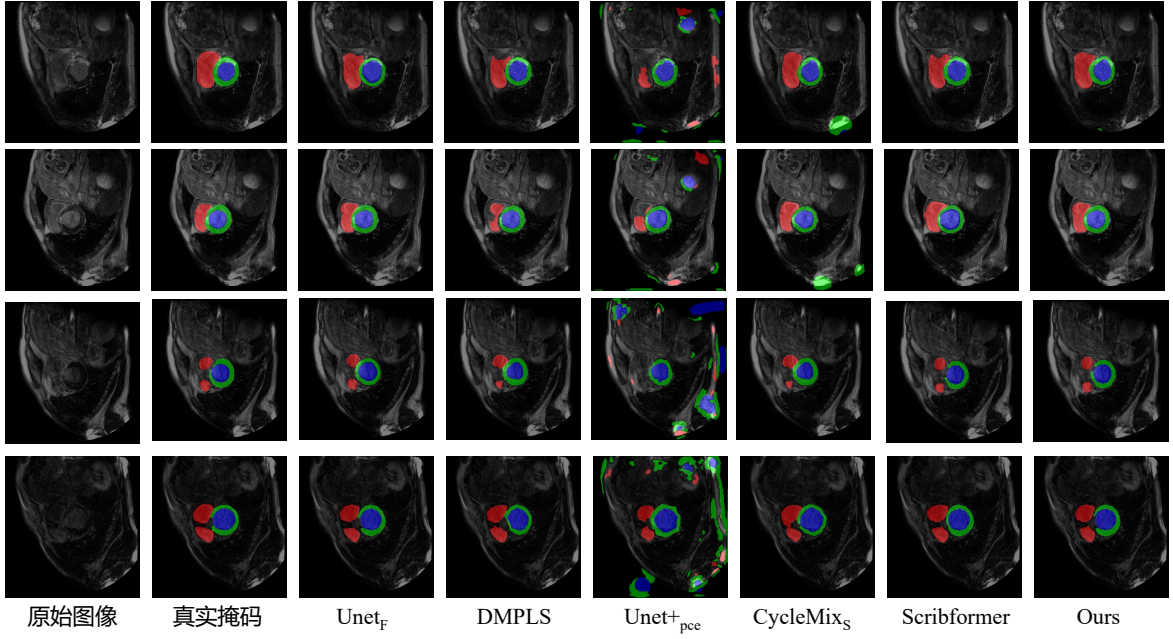


图 3.4 FW2SS 与其他先进方法在 MSCMRseg 数据集上的定性比较

FW2SS 的预测结果与真实标注具有更高的重叠程度。双重一致性学习策略有助于模型充分学习目标结构的形状信息，而像素级原型对比校准模块则进一步提升了边界区域的分割精度，从而有效提升了整体分割质量。

3.3 消融实验

为进一步验证本章所提方法中各个组成模块及关键策略的有效性，在 MSCMRseg 和 ACDC 数据集上开展系统的消融实验。具体而言，消融分析从模块有效性、扰动策略、涂鸦标注数量、涂鸦标注质量、预测阶段不同解码器输出，以及关键超参数敏感性等多个方面展开，对所提方法进行了全面评估。结果显示，各组成模块均对最终性能提升起到了积极作用，同时所设计的扰动机制和参数配置具有较好的合理性与鲁棒性。

首先，本章在 MSCMRseg 数据集的涂鸦标注设置下，对所提模型中各模块的有效性进行了评估，实验结果如表 3.3 所示。当模型仅采用 CNN 分支和弱扰动策略时，由于缺乏额外约束机制，且涂鸦标注提供的监督信息有限，模型性能提升并不明显。当进一步引入强扰动策略后，模型性能仅有小幅提升（由 54.4% 提升至 56.1%）。这是因为在该设置下，仅基于弱扰动预测尚无法生成足够可靠的伪标签，因此无法对强扰动视图的输出形成有效约束。相比之下，在 CNN 分支、弱扰动策略和像素级原

表 3.3 MSCMRseg 数据集上的模块消融实验（最优结果以粗体表示）

涂鸦一致性匹配				对比校准	左心室	心肌	右心室	平均值
网络扰动		输入扰动		\				
CNN	Transformer	弱扰动	强扰动					
✓		✓			0.636	0.408	0.588	0.544
✓		✓	✓		0.653	0.431	0.598	0.561
✓		✓		✓	0.683	0.471	0.658	0.604
✓	✓	✓			0.916	0.831	0.848	0.865
✓	✓	✓	✓		0.918	0.834	0.882	0.878
✓	✓	✓	✓	✓	0.924	0.837	0.887	0.883

型对比校准模块共同作用下，模型性能得到了一定改善。

当同时引入 CNN 和 Transformer 两个网络分支，并结合弱扰动策略时，模型性能得到了明显提升。这种提升主要来源于异构分支在训练阶段提供了不同的预测视角，使模型能够基于两类结构表征之间的一致区域生成相对稳定的伪标签，从而缓解单一分支在稀疏涂鸦监督下可能出现的预测偏差。需要强调的是，这里的性能提升并不意味着 CNN 分支与 Transformer 分支在推理阶段的简单平均必然优于单分支输出，而是说明异构分支作为训练阶段的伪监督构建和一致性约束来源具有积极作用。在此基础上，进一步引入完整的涂鸦一致性匹配模块后，模型性能继续提升。最后，当进一步加入像素级原型对比校准模块后，模型取得了最优的 88.3% 平均 Dice 分数，进一步验证了多重一致性扰动学习框架的有效性。

为研究扰动强度对模型性能的影响，在 ACDC 数据集的涂鸦标注设置下，对强扰动策略 ω_s 进行了不同扰动强度的消融实验。具体而言，实验固定变换组合数 $R = 3$ ，并将扰动强度 δ 依次从 1 减小到 1/2、1/4 和 1/8。由表 3.4 可知，随着扰动强度 δ 的减小，模型在 ACDC 数据集上的性能出现了明显下降。过低的扰动强度可能无法提供足够丰富的特征变换，从而限制模型对数据多样性和复杂性的适应能力。上述现象表明，强扰动策略在本章方法中发挥了关键作用。

表 3.4 ACDC 数据集上的扰动强度消融实验（最优结果以粗体表示）

扰动强度	左心室	心肌	右心室	平均值
1/8	0.923	0.871	0.874	0.889
1/4	0.927	0.874	0.876	0.892
1/2	0.928	0.878	0.881	0.896
1	0.931	0.881	0.887	0.900

类似地，本章在固定扰动强度 $\delta = 1$ 的条件下，进一步评估了变换组合数 R 对模型性能的影响。具体而言，实验将 R 依次设置为 1 至 7，并在 ACDC 数据集上进行验证，结果如图 3.5 所示。可以观察到，随着 R 的增大，模型性能呈现先升高后下降的趋势，并在 $R = 3$ 时达到最佳。适当增加变换组合数能够增强数据多样性，从而提升模型的泛化能力；但当 R 过大时，过多的变换可能会引入无效或不稳定特征，导致表征过于复杂，反而对最终性能产生负面影响。因此，在本章方法中，合理选择变换组合数 R 对于在增强特征多样性的同时避免引入过多噪声至关重要。

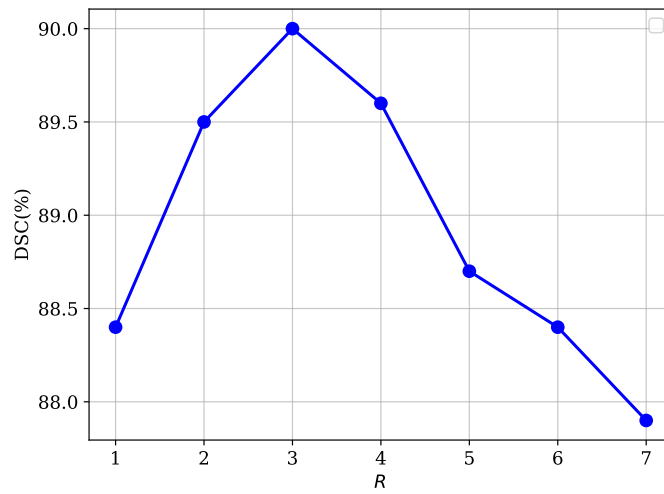


图 3.5 ACDC 数据集上变换组合数 R 的扰动策略消融实验

本章进一步分析了不同训练样本规模对模型性能的影响，实验结果如表 3.5 所示。结果显示，随着涂鸦标注数量的增加，模型分割性能呈现持续上升趋势。当使用全部 70 个涂鸦标注进行训练时，模型取得了最高的 90.6% 平均 Dice 分数。值得注意的是，即使在仅使用 14 个涂鸦标注的情形下，模型仍然取得了 86.8% 的 Dice 分数，表明 FW2SS 在少样本条件下仍具有较强的鲁棒性。这一性能表现可能与 FW2SS 中强扰动视图与弱扰动视图之间的一致性学习机制有关。通过对强增强样本与弱增强样本进行联合优化，模型能够更有效地学习稳健的特征表示，从而在训练样本稀缺的情况下依然具备较强的学习能力和泛化性能。

此外，本章通过对涂鸦标注进行遮挡处理，分析训练样本质量对 FW2SS 模型性能的影响，以评估模型对不同涂鸦长度变化的鲁棒性。具体而言，对于给定的长度比例，首先确定每类涂鸦标注需要被遮挡的像素数量，然后在尽可能保持剩余涂鸦

表 3.5 不同涂鸦标注数量下的分割性能敏感性分析

数据量	左心室	心肌	右心室	平均值
14 条涂鸦标注	0.904	0.850	0.849	0.868
28 条涂鸦标注	0.914	0.858	0.862	0.878
35 条涂鸦标注	0.931	0.881	0.887	0.900
56 条涂鸦标注	0.932	0.888	0.893	0.904
70 条涂鸦标注	0.935	0.888	0.896	0.906

连通性的前提下，对各类像素执行遮挡操作，如图 3.6 所示，随着长度比例的减小，涂鸦标注逐渐缩短。

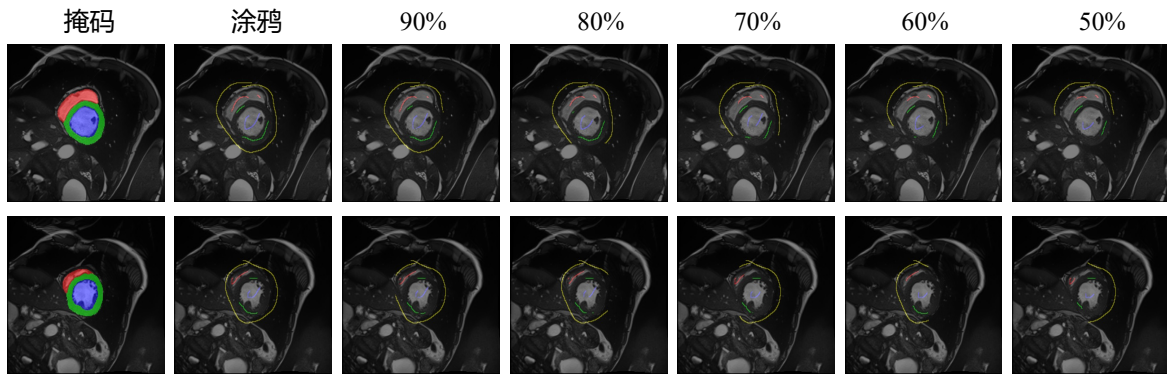


图 3.6 不同长度比例涂鸦标注可视化

如图 3.7 所示，在不同涂鸦比例下，本章方法始终优于基于模型扰动的伪监督方法 ScribFormer 以及基于输入扰动一致性的 CycleMix。尽管随着涂鸦长度比例的减小，模型性能整体会有所下降，但本章提出的稳健一致性学习策略相较于单一一致性学习方法，能够更有效地缓解性能退化。该结果表明，与其他方法相比，本章方法即使在较低质量的涂鸦标注条件下训练，仍能保持较强的分割性能。

为验证最终推理输出选择的合理性，本章进一步比较了 ACDC 数据集上不同解码器输出方案的性能差异，结果如表 3.6 所示。结果显示，在推理阶段，CNN 解码器与 Transformer 解码器之间存在一定的性能差异。值得注意的是，CNN 解码器输出取得了 90.0% 的 Dice 分数，优于 Transformer 解码器和双解码器平均输出（分别为 89.2% 和 89.8%）。

这一现象揭示了本章框架在训练阶段与推理阶段之间存在一定差异。在训练阶段，如表 3.3 所示，CNN 与 Transformer 的双分支组合能够为涂鸦一致性匹配模块提供具有差异性的预测视角，从而辅助生成相对稳定的伪标签，并增强模型对全局结

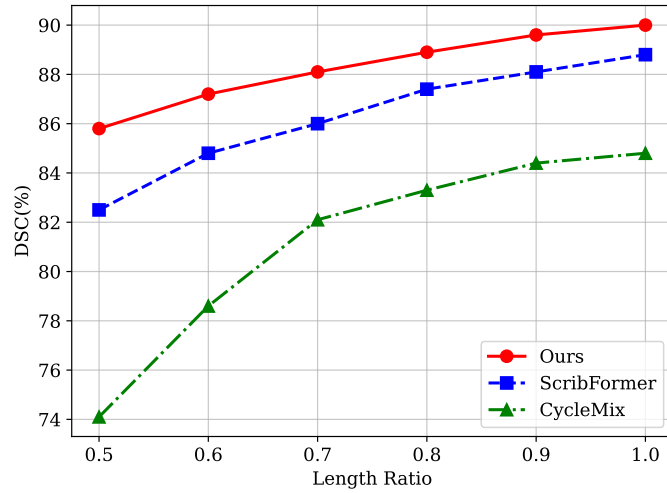


图 3.7 ACDC 数据集上涂鸦标注比例的性能影响

构信息的学习。然而，在推理阶段，最终分割结果不再依赖伪标签生成过程，而更直接地取决于输出分支对局部边界、纹理细节和空间连续性的刻画能力。由于 CNN 具有更强的局部归纳偏置，并通过跳跃连接保留了更多高分辨率空间细节，其解码器在边界精细化和局部纹理恢复方面表现更稳定。相比之下，Transformer 分支侧重全局上下文建模，其输出在部分边界区域可能更平滑；若将 CNN 与 Transformer 输出进行简单平均，可能会引入局部预测与全局预测之间的不一致，导致边界区域被平滑化或产生轻微预测冲突。因此，本文在最终推理阶段采用 CNN 解码器输出作为分割结果。该选择并不否定 Transformer 分支在训练阶段的作用，而是表明异构双分支的主要价值体现在训练阶段的伪标签构建、结构一致性约束和全局上下文辅助学习中。

表 3.6 ACDC 数据集上不同解码器输出的性能对比（最优结果以粗体表示）

解码器	左心室	心肌	右心室	平均值
CNN	0.931	0.881	0.887	0.900
Transformer	0.926	0.871	0.876	0.892
CNN + Transformer	0.931	0.879	0.884	0.898

阈值 τ_w 和 τ_s 对于筛选可靠像素、构建前景类别原型以及进行对比校准至关重要。为验证阈值选择的合理性，并评估模型对这些参数的敏感性，尤其是其对边界分割质量的影响，本章在 ACDC 数据集上开展了敏感性分析。除 Dice 分数外，还引入了第 95 百分位 Hausdorff 距离（HD95）作为评价指标，该指标对边界偏移更加敏

感。

如表 3.7 所示, 可以观察到以下现象: (1) 当固定 τ_w 时, 设置 $\tau_s = 0.5$ 总能取得最佳性能 (平均 Dice 为 0.900, HD95 为 3.7mm)。较高的 τ_s (如 0.7) 会过度限制可用于对比校准的像素数量, 导致监督不足并显著增加边界误差 (HD95 上升至 8.9mm); 而较低的 τ_s (如 0.3) 则会引入更多噪声和不可靠像素, 进而使边界预测不稳定。(2) τ_w 主要影响前景类别原型的可靠性。将 τ_w 从 0.7 提高到 0.9 时, 平均 Dice 有所提升, 且 HD95 显著下降; 但当继续增大到 0.95 时, 性能趋于饱和甚至略有下降 (平均 Dice 为 0.897, HD95 为 4.5mm)。这可能是因为过高的阈值会导致可用于稳定更新原型的像素样本不足, 从而限制模型的泛化能力。综合考虑性能与稳定性, 默认设置采用 $\tau_w = 0.9$ 、 $\tau_s = 0.5$ 。

表 3.7 ACDC 数据集上置信度阈值 (τ_w, τ_s) 敏感性分析

τ_w	τ_s	左心室	心肌	右心室	平均值	HD95(mm)↓
0.7	0.3	0.917	0.875	0.875	0.889	5.9
0.7	0.5	0.923	0.879	0.879	0.894	4.9
0.7	0.7	0.911	0.864	0.871	0.882	9.4
0.9	0.3	0.932	0.879	0.884	0.898	4.3
0.9	0.5	0.931	0.881	0.887	0.900	3.7
0.9	0.7	0.914	0.867	0.873	0.885	8.9
0.95	0.3	0.925	0.876	0.879	0.893	5.3
0.95	0.5	0.929	0.879	0.884	0.897	4.5
0.95	0.7	0.917	0.866	0.861	0.881	9.7

为保证本章框架的可复现性并评估其优化过程的稳健性, 本章进一步研究了式 3.22 中超参数 λ_1 、 λ_2 和 λ_3 的敏感性。这些权重用于平衡不同损失项的贡献。默认取值 ($\lambda_1 = 0.5$, $\lambda_2 = 0.25$, $\lambda_3 = 0.01$) 是通过对训练初期各损失项的数值规模与梯度尺度进行对齐后确定的, 以确保在优化过程中不会出现某一损失项占主导地位的情况。具体而言, 由于对比损失本身的数值尺度较大, 因此为 λ_3 赋予较小权重, 以避免其在训练初期主导整体优化过程, 从而保证数值稳定性以及各项损失之间的梯度平衡。

如表 3.8 所示, 本章在 ACDC 数据集上对各损失权重进行了敏感性实验。结果显示, 默认设置取得了最高的平均 Dice 分数 0.900。当跨分支一致性损失权重 λ_2 增大到 0.5 时, 模型性能明显下降 (平均 Dice 为 0.842), 说明过强的一致性约束可能会抑制模型学习多样特征的能力。同时, 当 λ_3 在较小范围内变化 (如 0.005 到 0.01)

时，模型性能仅有轻微波动。这些结果验证了本章默认超参数配置能够为涂鸦监督学习提供更加稳健且平衡的优化目标。

表 3.8 ACDC 数据集上损失权重 ($\lambda_1, \lambda_2, \lambda_3$) 的敏感性分析结果

λ_1	λ_2	λ_3	左心室	心肌	右心室	平均值
1.0	0.25	0.01	0.915	0.845	0.866	0.875
0.5	0.25	0.01	0.931	0.881	0.887	0.900
0.5	0.5	0.01	0.879	0.815	0.833	0.842
0.5	0.25	0.005	0.927	0.880	0.883	0.897
0.5	0.25	0.1	0.912	0.844	0.853	0.870

3.4 本章小结

作为复杂纹理图像中的典型类别之一，医学影像在涂鸦监督条件下更突出地表现为形状信息学习不足、局部纹理刻画与全局结构建模难以协同以及复杂边界区域判别困难等问题。围绕这些医学影像特有的任务需求，本章提出了一种基于一致性约束与对比学习的涂鸦监督医学影像分割方法 **FW2SS**。其中，**CNN** 与 **Transformer** 异构双分支结构主要用于增强局部纹理与全局结构的互补建模；联合网络扰动与输入扰动的一致性学习机制主要用于动态生成稳定的像素级伪标签，以提升稀疏监督条件下的形状信息学习能力；像素级原型对比校准模块则通过对强扰动分支中的前景预测特征表示进行校准，进一步增强复杂边界区域的特征判别能力，从而提升模型的分割精度。

在实验验证方面，**FW2SS** 在 **ACDC** 和 **MSCMRseg** 两个心脏磁共振数据集上的平均 **Dice** 分数分别达到 90.0% 和 88.3%，在涂鸦监督设置下优于多种代表性弱监督分割方法。同时，在本文采用的相同训练病例数量、二维切片训练和统一实验协议下，**FW2SS** 的平均 **Dice** 高于若干常用全监督 **U-Net** 系列基线方法，并接近本文纳入比较的强全监督方法 **nnUNet**。上述结果主要说明，所提方法能够在显著降低标注成本的前提下获得具有竞争力的分割性能，而不应被理解为涂鸦监督方法在一般情况下优于全监督方法。系统的消融实验进一步表明，异构分支耦合、双重扰动策略及对比校准模块均对最终性能提升发挥了不可或缺的作用，且该方法在标注数量极度稀缺的条件下仍保持了较强的鲁棒性与泛化能力。

此外，所提方法在训练阶段由于采用异构双分支架构与自注意力机制，会带来

一定的计算开销和显存占用，对硬件资源提出一定要求。然而，考虑到该方法仅需低成本的涂鸦标注，便可在本文实验设置下获得接近强全监督基线的分割表现，其在标注成本与分割性能之间实现了较好的平衡，为医学影像低成本分割模型的构建提供了可行思路。

第四章 基于多历史投票与边界感知约束的涂鸦监督材料显微图像分割方法

材料显微图像是复杂纹理图像中的另一类典型代表，其分割结果是材料组织表征、相区识别与性能分析的重要基础。与医学影像相比，材料显微图像在小样本涂鸦监督场景下更突出地表现为训练稳定性不足、伪标签可靠性有限以及复杂相界附近边界噪声易传播等问题。受样本获取困难、组织形貌复杂以及相界模糊黏连等因素影响，模型往往难以形成稳定训练过程，且单次预测波动更容易在复杂相界附近累积为伪标签噪声。因此，本章聚焦材料显微图像这一典型场景，重点讨论如何在极少量涂鸦标注条件下提升训练稳定性、增强伪标签可靠性，并抑制复杂相界附近的边界噪声传播。围绕这一问题，本章提出一种基于多历史投票与边界感知约束的涂鸦监督材料显微图像分割方法，命名为 **Multi-history Voting and Boundary-aware Constraint (MVBC)**。该方法以采用预训练参数初始化与部分冻结策略的 **U-Net++** 网络作为基础分割主干^[5]，通过按样本维护的多历史投票伪标签机制对同一样本在不同训练时刻的预测结果进行投票一致性建模，从时间维度挖掘稳定且高置信度的伪监督信号，以缓解小样本条件下监督不足的问题；同时结合边界阈值过滤以及边界平滑与锐化联合约束，抑制边界区域的伪标签噪声传播，增强模型对复杂相界的刻画能力，从而提升小样本条件下的材料显微图像分割性能。

4.1 方法概述

4.1.1 模型框架

本方法整体框架如图 4.1 所示，主要由基础分割网络、多历史投票伪标签生成模块和边界感知约束模块组成，整体流程包括预处理、训练和推理三个阶段。预处理阶段，首先利用参数冻结的边界检测网络对原始材料显微图像进行离线边界检测，生成与原图坐标对齐的边界响应图，作为固定边界先验用于后续伪标签筛选和边界约束构造。训练阶段，输入图像经过数据增强后输入使用预训练权重的基础分割网络 **U-Net++**，得到像素级分割预测结果。由于训练中包含旋转、翻转等几何增强操作，网络输出需经过相应的几何逆变换恢复至原始图像坐标系，以保证其与涂鸦标

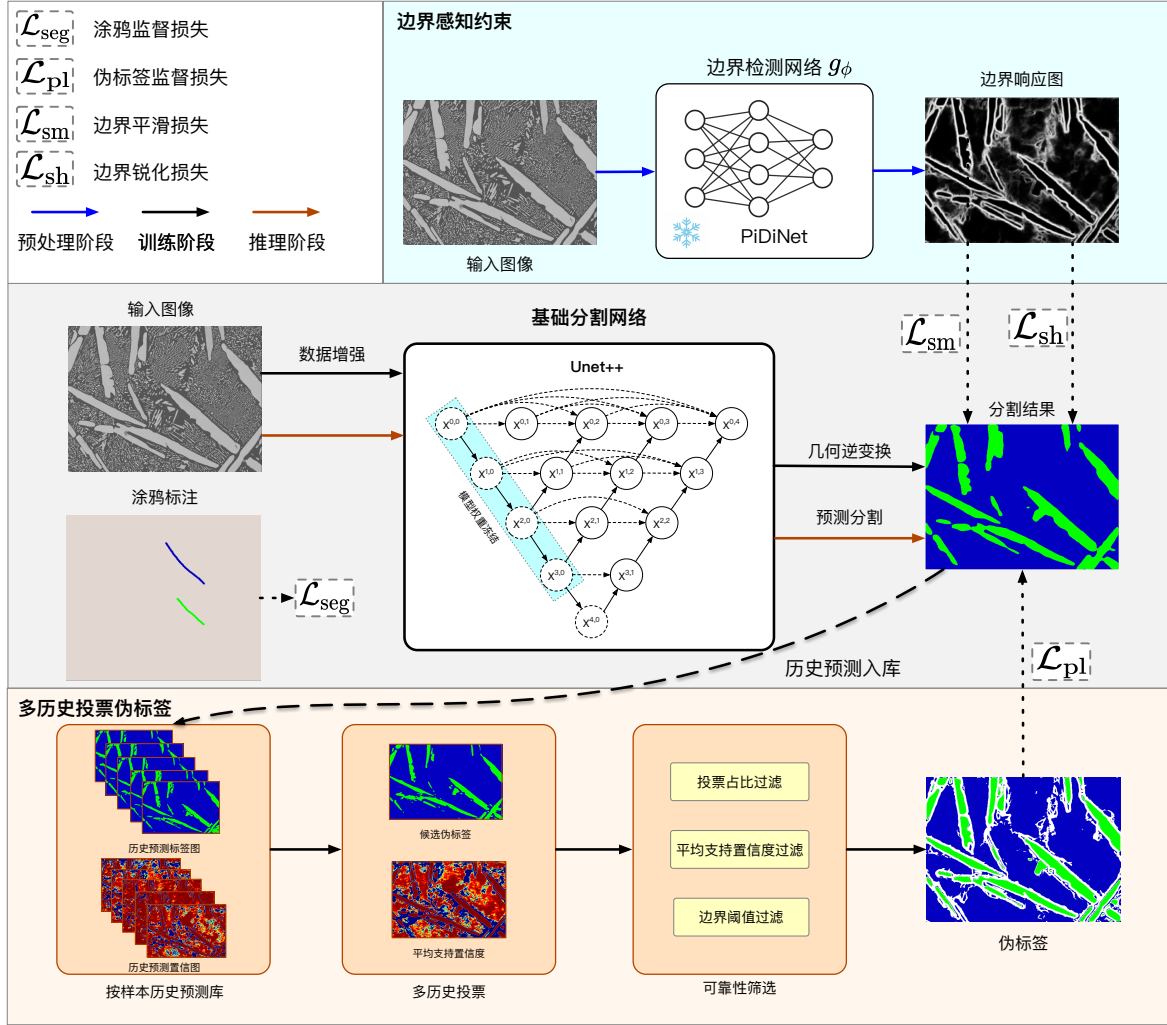


图 4.1 MVBC 方法的整体框架

注、历史预测库和边界响应图在空间位置上对齐。对于已有的涂鸦标注，模型仅在标注像素上计算涂鸦监督损失 $\mathcal{L}_{seg}$ ，作为基础监督信号。在此基础上，MVBC 为每个训练样本维护一个按样本索引的历史预测库，用于保存该样本最近若干次训练访问时的预测标签图和置信度图。当当前样本再次进入训练时，模型首先读取其历史预测结果，并通过多历史投票获得候选伪标签，同时计算像素级投票占比和平均支持置信度。随后，候选伪标签依次经过投票占比过滤、平均支持置信度过滤和边界阈值过滤，只有在历史预测中表现稳定、置信度较高且不位于高响应边界区域的像素，才被赋予最终伪标签并用于计算伪标签监督损失 $\mathcal{L}_{pl}$ 。边界感知约束模块利用预处理阶段得到的边界响应图发挥双重作用：一方面，在伪标签构建阶段通过边界阈值过滤抑制复杂相界附近不可靠伪标签的传播；另一方面，结合分割网络输出构造边界平滑损失 $\mathcal{L}_{sm}$ 和边界锐化损失 $\mathcal{L}_{sh}$ ，分别增强非边界区域的预测一致性和高响应边界

处的类别过渡清晰度。推理阶段，模型仅保留训练完成后的分割网络进行前向预测。

为便于后文表述，首先对材料显微图像分割任务作数学化定义。设训练集为

$$D = \{(x_n, s_n)\}_{n=1}^N, \quad (4.1)$$

其中， $x_n \in \mathbb{R}^{H \times W \times 3}$ 表示第 n 幅材料显微图像， $s_n \in \{0, 1, \dots, C-1, 255\}^{H \times W}$ 表示对应的涂鸦监督标签图。对于每个训练样本 x_n ，MVBC 进一步利用边界检测网络 g_ϕ 从原始图像中提取与其坐标对齐的边界响应图

$$b_n = g_\phi(x_n), \quad b_n \in [0, 1]^{H \times W}, \quad (4.2)$$

需要说明的是， b_n 为由原始图像派生得到的辅助边界先验，不属于训练集的原始标注组成部分。边界响应图由边界检测网络在训练前离线生成，并在训练阶段作为固定边界先验参与伪标签构建阶段的边界阈值过滤与边界感知约束构造。

4.1.2 基础分割网络

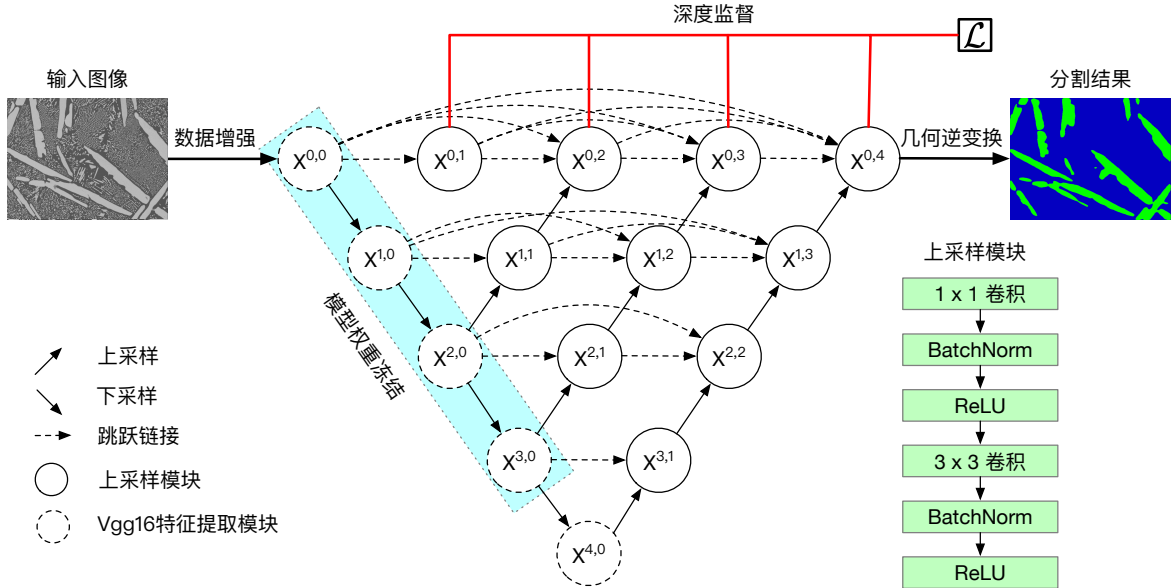


图 4.2 以预训练 VGG16 骨干网络为编码器的 U-Net++ 网络示意图

本节的设计重点不在于提出新的主干网络，而在于为小样本涂鸦监督条件下的稳定训练提供可靠预测基础，从而为后续多历史投票伪标签生成与边界感知约束提供有效支撑。为提升材料显微图像在小样本条件下的分割性能，MVBC 采用编码器—

解码器结构作为基础分割框架。该结构通过编码器逐步提取由浅入深的层次化特征，再由解码器逐级恢复空间分辨率，从而实现像素级分类预测。需要指出的是，MVBC 中的多历史投票伪标签机制与边界感知约束并不依赖某一固定主干网络，理论上可嵌入多种编码器—解码器式分割模型；即使在同一分割框架内部，编码骨干网络也可根据任务需求进行替换。综合考虑材料显微图像的局部纹理复杂、小样本场景下的训练稳定性以及多尺度特征表达能力，MVBC 最终选用以预训练 VGG16 骨干网络为编码器的 U-Net++ 作为基础分割模型^[5,86]。

在 MVBC 所采用的 U-Net++ 结构中，编码器部分由预训练 VGG16 提供初始特征提取能力^[86]。VGG16 通过连续卷积操作能够有效提取边缘、纹理和形状等基础视觉特征，而这类低层视觉信息在自然图像与材料显微图像之间具有一定共通性。对于材料显微图像分割任务而言，训练样本数量通常较少，若完全从头训练编码器，不仅会导致参数学习不充分，还可能因数据规模有限而引发严重过拟合。因此，MVBC 采用在 ImageNet 上预训练的 VGG16 网络参数对编码器进行初始化，以充分利用预训练模型中已经学习到的通用视觉表征，从而提高网络在小样本条件下的收敛速度与训练稳定性。

模型整体结构如图 4.2 所示，其中实线圆圈表示上采样模块，主要由两个核心卷积单元组成，每个单元均遵循“卷积 → 归一化 → 激活”的标准组合；虚线圆圈表示 U-Net++ 所使用的 VGG16 特征提取模块。在输入网络前，材料显微图像首先经过数据增强预处理，以扩充样本多样性并缓解小样本条件下的过拟合问题。具体而言，实验采用旋转与翻转等空间几何增强，以提升模型对方向变化和局部形变的鲁棒性；同时引入亮度调整、对比度调整及归一化等像素值增强操作，以模拟不同成像条件下材料组织的外观变化。经过增强后的图像进入由预训练 VGG16 构建的编码器路径（图中 $X^{0,0}$ 至 $X^{4,0}$ ），在该过程中，特征图的空间分辨率逐步降低，而语义表征能力不断增强。随后，在解码阶段，U-Net++ 通过密集跳跃连接对不同层级的编码特征与解码特征进行多尺度融合，图中以实线圆圈表示的上采样模块通过双层卷积结构逐步恢复图像的空间分辨率，最终输出分割结果。

在训练阶段，为保证多历史投票伪标签的构造与边界感知约束始终在原始图像坐标系下进行，需要对经过旋转、翻转等空间几何增强后的网络输出执行相应的几何逆变换。逆变换后的输出与原始图像、按样本维护的历史预测库以及边界响应图在空间位置上一一对齐，从而保证后续历史比较、伪标签构建和边界感知约束均在

统一坐标系下完成。最终，模型能够输出兼具较高语义精度与精细边缘信息的分割结果，从而实现对材料微观组织的精准识别。需要注意的是，图中红色实线标注的深度监督机制是 U-Net++ 架构的重要特性。损失函数通过将不同尺度的子网络输出 ($X^{0,1}, X^{0,2}, X^{0,3}, X^{0,4}$) 进行监督，使得模型在训练过程中能够获得多层级的梯度反馈。这种设计不仅有效缓解了深层网络中的梯度消失问题，还显著提升了模型在小样本约束下的收敛速度与分割精度。

尽管预训练模型能够提供较好的初始表示，但自然图像与材料显微图像在目标形态、纹理分布和类别语义上仍存在显著差异，因此并非所有预训练参数都能够直接适应材料显微图像分割任务。基于这一考虑，MVBC 并未对全部编码层进行统一微调，而是采用分层迁移与部分冻结相结合的优化策略。具体而言，图中以蓝色阴影标示的前四个阶段编码模块 ($X^{0,0}$ 至 $X^{3,0}$) 主要负责提取边缘、纹理等通用性较强的低级特征。在材料显微图像中，这些特征与晶界、相界等几何轮廓具有较高对应性，因此在跨领域任务中仍具有较好的可迁移性。基于此，在训练过程中固定这部分浅层编码权重，不仅能够保留预训练模型中较稳定的基础表征能力，还能在小样本条件下有效抑制过拟合。相比之下，深层编码模块 ($X^{4,0}$) 更侧重于提取与具体任务相关的高级语义信息，其表达内容与材料显微图像中复杂的组织分布和相区形态密切相关。因此，仅对深层编码模块 $X^{4,0}$ 及所有参与特征融合的解码节点 ($X^{i,j}, j > 0$) 进行微调，使网络在保持通用特征提取能力的同时，能够针对材料显微图像的特定分布特征进行适配。

MVBC 选用以预训练 VGG16 为编码骨干的 U-Net++ 作为基础分割网络，并通过数据增强、预训练初始化、浅层冻结、深层微调的策略缓解材料显微图像小样本训练中的过拟合和表征不足问题。该网络既继承了预训练模型在低层视觉特征提取方面的优势，又借助 U-Net++ 的多尺度嵌套连接结构增强了对复杂材料组织边界及区域结构的刻画能力，从而为后续多历史投票伪标签生成和边界感知约束模块提供可靠的分割基础。

4.1.3 多历史投票伪标签

在涂鸦监督条件下，监督信号仅覆盖少量局部像素，模型虽然能够依赖已标注区域获得初始类别先验，但对于占据主要比例的未标注区域仍缺乏充分而稳定的约束。这一问题在小样本材料显微图像分割中尤为突出：样本规模有限使模型参数更

新更易受偶然噪声和局部扰动影响，不同物相之间纹理相似、灰度接近及相界细长弯曲等特征又进一步提高了单次预测的不确定性。因此，在该场景中，伪标签可靠性的判断不宜仅依赖当前时刻的一次预测结果，而需要结合模型在相邻训练阶段和不同数据增强扰动下的预测稳定性进行经验性筛选。基于此，**MVBC** 按照历史预测信息积累、多历史投票一致性建模和可靠性筛选三个步骤构建伪标签生成机制，以提升小样本材料场景下伪标签的稳定性与抗噪性。

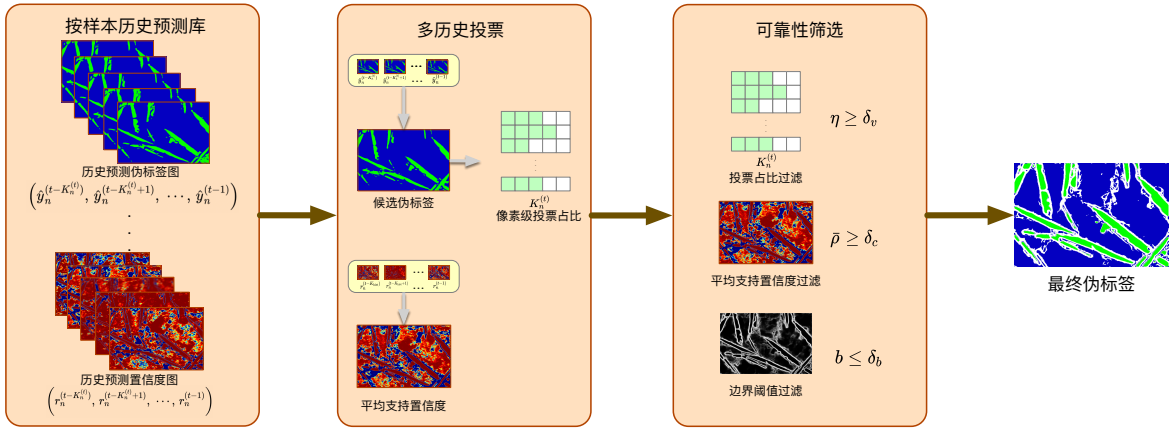


图 4.3 基于按样本历史预测库、多历史投票与三重可靠性筛选的伪标签生成流程图

首先，**MVBC** 为每个训练样本维护一个独立的历史预测库，用于按时间顺序保存该样本最近若干次训练访问时的预测标签图及其对应的置信度图。为与实现中采用的有限长度记忆机制保持一致，该模块将该历史预测库视为按样本维护的有序有限队列。设第 t 次迭代中访问样本 x_n 时，当前模型在原始图像坐标系上的概率输出为

$$\mathbf{p}_n^{(t)} \in [0, 1]^{C \times H \times W}, \quad (4.3)$$

其中， $\mathbf{p}_n^{(t)}$ 表示样本 x_n 在第 t 次迭代时由当前分割网络输出的类别概率图，其第一维对应 C 个分割类别，后两维对应图像空间位置。 $p_{n,c}^{(t)}(u)$ 表示像素位置 u 被预测为类别 c 的概率。考虑到训练阶段对输入图像执行了旋转与翻转等空间增强，而历史预测队列与边界先验均与原始图像坐标严格对齐，**MVBC** 在构造伪标签之前，首先将网络输出恢复到原始图像坐标系，再在统一空间中完成历史比较与后续筛选。则当前位置的硬预测类别与对应置信度分别定义为

$$\hat{y}_n^{(t)}(u) = \arg \max_{c \in \{0, \dots, C-1\}} p_{n,c}^{(t)}(u), \quad r_n^{(t)}(u) = \max_c p_{n,c}^{(t)}(u). \quad (4.4)$$

其中, $\hat{y}_n^{(t)}(u)$ 表示当前模型在像素位置 u 处给出的硬预测类别, 即预测概率最大的类别索引; $r_n^{(t)}(u)$ 表示该硬预测类别对应的最大预测概率, 用于衡量模型对当前位置预测结果的置信程度。基于此, 样本 x_n 在第 t 次迭代时的历史预测队列可表示为

$$\mathcal{H}_n^{(t)} = \left[\left(\hat{y}_n^{(t-K_n^{(t)})}, r_n^{(t-K_n^{(t)})} \right), \dots, \left(\hat{y}_n^{(t-1)}, r_n^{(t-1)} \right) \right], \quad (4.5)$$

其中, $\mathcal{H}_n^{(t)}$ 表示第 n 个训练样本在第 t 次迭代时的历史预测队列; $\hat{y}_n^{(t-l)}$ 表示该样本在历史时刻 $t-l$ 的像素级预测标签图; $r_n^{(t-l)}$ 表示与之对应的像素级预测置信度图; $K_n^{(t)}$ 表示当前时刻该样本已累计的有效历史长度。若设历史预测库的最大容量为 K_{hist} , 则有

$$K_n^{(t)} = \min(|\mathcal{H}_n^{(t)}|, K_{\text{hist}}). \quad (4.6)$$

通过对历史预测队列中的多次历史预测结果进行联合建模, 可以从时间维度表征像素级类别预测的一致性与稳定性, 从而为后续伪标签生成提供依据。需要说明的是, 本文所采用的多历史投票并非建立在历史预测彼此独立同分布的严格统计假设之上。由于相邻历史预测来自同一模型在连续训练阶段的参数状态, 其间不可避免地存在相关性, 因此这里的历史队列不应被理解为独立样本集合, 而应被理解为滑动时间窗口下的预测稳定性记录。换言之, **MVBC** 利用的是同一样本在不同训练时刻和数据增强扰动下的经验一致性, 而非通过独立投票获得严格意义上的统计显著性。

在具体实现上, 该历史预测库按样本身份独立维护, 并以有限长度队列的形式保存最近 K_{hist} 次训练访问时的预测结果, 从而保证同一样本在不同训练访问时刻的历史预测能够被准确追踪; 当历史长度超过上限时, 最早的记录按先进先出策略移除。同时, 该模块采用“先投票、后入库”的更新策略, 即当前时刻的预测结果不直接参与当前时刻伪标签生成, 以避免模型利用自身瞬时输出强化自身, 降低自训练过程中的自确认偏差。

其次, 当样本的历史长度满足 $K_n^{(t)} \geq K_{\min}$ 时, 该模块对历史预测队列中的多次历史结果执行像素级多历史投票。对于任意类别 c 和像素位置 u , 定义其历史票数为

$$v_{n,c}^{(t)}(u) = \sum_{l=1}^{K_n^{(t)}} \mathbb{I}[\hat{y}_n^{(t-l)}(u) = c], \quad (4.7)$$

其中, $v_{n,c}^{(t)}(u)$ 表示第 n 个训练样本在第 t 次迭代时, 像素位置 u 对类别 c 的历史投

票数； $\hat{y}_n^{(t-l)}(u)$ 表示该样本在历史时刻 $t-l$ 时于像素位置 u 的预测类别； $\mathbb{I}[\cdot]$ 为示性函数，当条件成立时取值为 1，否则取值为 0。通过对所有类别的票数进行比较，可得到像素位置 u 的候选伪标签

$$\bar{y}_n^{(t)}(u) = \arg \max_c v_{n,c}^{(t)}(u). \quad (4.8)$$

为了度量候选类别在历史序列中受到的稳定支持程度，本文定义像素级投票占比为

$$\eta_n^{(t)}(u) = \frac{\max_c v_{n,c}^{(t)}(u)}{K_n^{(t)}}. \quad (4.9)$$

其中， $\eta_n^{(t)}(u)$ 表示第 n 个训练样本在第 t 次迭代时，像素位置 u 的候选类别投票占比， $v_{n,c}^{(t)}(u)$ 表示类别 c 在该像素位置上的历史投票数， $K_n^{(t)}$ 表示当前时刻该样本已累计的有效历史长度。该式通过将像素位置 u 处的最大历史票数除以有效历史长度，刻画候选类别在滑动历史窗口内的经验支持比例。 $\eta_n^{(t)}(u)$ 越大，说明该像素在最近若干次历史预测中越倾向于被判定为同一类别，其类别预测在当前训练阶段附近表现出更高的一致性与稳定性。需要指出的是，该指标用于衡量预测结果的相对稳定程度，而非用于构造严格统计意义下的显著性检验。

仅依据票数占优仍不足以全面刻画伪标签的可靠性，该模块进一步统计所有支持候选类别的历史预测置信度，并定义其平均支持置信度为

$$\bar{\rho}_n^{(t)}(u) = \frac{\sum_{l=1}^{K_n^{(t)}} r_n^{(t-l)}(u) \mathbb{I}[\hat{y}_n^{(t-l)}(u) = \bar{y}_n^{(t)}(u)]}{\max_c v_{n,c}^{(t)}(u)}. \quad (4.10)$$

其中， $\bar{\rho}_n^{(t)}(u)$ 表示第 n 个训练样本在第 t 次迭代时，像素位置 u 的平均支持置信度； $r_n^{(t-l)}(u)$ 表示该样本在历史时刻 $t-l$ 时于像素位置 u 的预测置信度； $\hat{y}_n^{(t-l)}(u)$ 表示对应的预测类别； $\bar{y}_n^{(t)}(u)$ 表示通过多历史投票得到的候选伪标签； $\mathbb{I}[\cdot]$ 为示性函数，当历史预测类别与候选伪标签一致时取值为 1，否则取值为 0； $\max_c v_{n,c}^{(t)}(u)$ 表示候选类别在该像素位置上的最大历史票数。该式通过对所有支持候选伪标签的历史预测置信度取平均，从模型输出置信水平角度度量候选类别的支持强度。 $\bar{\rho}_n^{(t)}(u)$ 越大，说明该像素不仅在最近历史窗口内更稳定地被判定为同一类别，而且对应预测通常具有更高的置信水平，因此其伪标签相对更可靠。平均支持置信度并非对历史窗口内全部预测置信度直接取平均，而仅对与候选伪标签类别一致、即对当前候选类别形成支持票的历史预测置信度进行平均。该指标与投票占比共同用于经验性筛选不稳定

预测，而不表示历史预测之间满足独立性假设。

在图 4.3 中，第二步多历史投票同时输出三类信息：候选伪标签 $\bar{y}_n^{(t)}(u)$ 、像素级投票占比 $\eta_n^{(t)}(u)$ 以及平均支持置信度 $\bar{\rho}_n^{(t)}(u)$ 。其中，候选伪标签反映历史投票的多数类别结果，投票占比刻画类别一致性的强弱，平均支持置信度则衡量支持该类别的历史预测是否具有足够高的概率响应。

为了在伪标签构建阶段抑制边界混淆区域的噪声传播，MVBC 预先为每个训练样本生成与原始图像坐标对齐的边界响应图 $b_n \in [0, 1]^{H \times W}$ ，其中， $b_n(u)$ 表示像素位置 u 的边界响应强度，响应值越大表示该位置越可能处于真实相界附近。该边界图由边界检测网络从原始材料显微图像中提取，并在伪标签构建阶段作为边界阈值过滤的先验信息使用；其具体生成方式及在损失函数中的进一步作用，将在后续“边界感知约束”小节中详细说明。

最后，考虑到仅依赖历史一致性与置信度估计，仍可能在复杂边界或局部噪声区域产生不可靠伪标签，MVBC 进一步引入三重可靠性筛选机制，对候选伪标签执行逐像素过滤。具体而言，只有同时满足以下三个条件的像素，才可被视为可靠伪标签像素：其一，候选类别的投票占比不低于阈值 δ_v ，即要求该类别在最近历史窗口内具有足够高的经验一致性；其二，候选类别的平均支持置信度不低于阈值 δ_c ，即要求支持该类别的历史预测具有较高的模型输出置信水平；其三，该像素对应的边界响应不高于阈值 δ_b ，即通过边界阈值过滤避免在边界混淆显著的位置直接传播伪监督。由此，可靠伪标签像素集合定义为

$$\Omega_{n,\text{rel}}^{(t)} = \{u \in \Omega \mid K_n^{(t)} \geq K_{\min}, \eta_n^{(t)}(u) \geq \delta_v, \bar{\rho}_n^{(t)}(u) \geq \delta_c, b_n(u) \leq \delta_b\}, \quad (4.11)$$

其中， $K_{\min}$ 表示启动多历史投票伪标签生成所需的最小历史长度， δ_v 、 δ_c 和 δ_b 分别表示投票占比阈值、平均支持置信度阈值以及边界过滤阈值。基于上述可靠像素集合，可构造最终伪标签图，其定义为

$$\tilde{s}_n^{(t)}(u) = \begin{cases} \bar{y}_n^{(t)}(u), & u \in \Omega_{n,\text{rel}}^{(t)}, \\ 255, & \text{其他}, \end{cases} \quad (4.12)$$

其中， $\tilde{s}_n^{(t)}(u)$ 表示第 t 次迭代时样本 x_n 在像素位置 u 处的最终伪标签； $\bar{y}_n^{(t)}(u)$ 表示由多历史投票得到的候选类别；标签值 255 表示忽略标签，即该位置在当前迭代中不参与伪监督损失计算。由此可见，最终伪标签并非直接来源于历史投票结果，而

是建立在可靠性筛选基础上的选择性赋值过程。只有当某一像素满足可靠像素集合的判定条件时，才将其候选类别写入伪标签图，否则该位置保持忽略状态。

为了在损失计算阶段显式表征伪监督的有效作用区域，该模块定义对应的二值可靠性掩码为

$$w_n^{(t)}(u) = \mathbb{I}\left[u \in \Omega_{n,\text{rel}}^{(t)}\right], \quad (4.13)$$

其中， $\mathbb{I}[\cdot]$ 为示性函数。当像素 u 属于可靠像素集合 $\Omega_{n,\text{rel}}^{(t)}$ 时， $w_n^{(t)}(u) = 1$ ，表示该位置可参与伪标签监督；反之， $w_n^{(t)}(u) = 0$ ，表示该位置在当前轮次中被屏蔽，不对伪标签损失产生贡献。与最终伪标签图相比，可靠性掩码并不记录具体类别信息，而是用于刻画伪监督的作用范围，并在后续伪标签损失计算中充当像素级权重。

图 4.3 中的第三步可靠性筛选本质上是由投票占比过滤、平均支持置信度过滤、边界阈值过滤构成的三重筛选过程。经过该过程后，仅有那些在历史预测中表现出高一致性、高置信度且远离不稳定边界区域的像素，才会被赋予最终伪标签并纳入附加监督。

为直观展示所提出方法在不同材料显微图像上的伪标签构造效果，图 4.4 给出了三个典型数据集上的可视化结果。每一行对应一种材料数据集，从左至右依次为原始图像与涂鸦标注的叠加显示、经过多历史投票与可靠性筛选后的最终伪标签以及模型预测结果。其中，伪标签图中的白色区域表示未通过可靠性筛选的像素，即该位置未被赋予伪标签，在当前迭代中不参与伪监督损失计算。可以观察到，在不同材料数据集上，所生成的伪标签能够较好地保留主要组织区域的判别结果，并在复杂边界附近减少部分不稳定预测的传播，表明所提方法具有一定的跨材料显微图像场景适用性。

综上，MVBC 的多历史伪标签生成机制可以概括为三个连续步骤：首先，通过按样本维护的历史预测队列，按时间顺序保存同一样本最近若干次训练访问时的预测结果及其置信度；其次，通过多历史投票获得候选伪标签，并同步计算像素级投票占比与平均支持置信度；最后，通过投票占比、支持置信度与边界阈值过滤构成的三重可靠性筛选生成最终伪标签。该机制引入了显式的时序一致性建模，将伪标签可靠性的判定依据由单次预测扩展为滑动历史窗口内的经验稳定性判断。需要强调的是，多历史投票并不意味着历史预测之间相互独立，而是利用模型在相邻训练阶段和不同数据增强扰动下是否持续给出一致判断来筛选不稳定伪标签。考虑到材

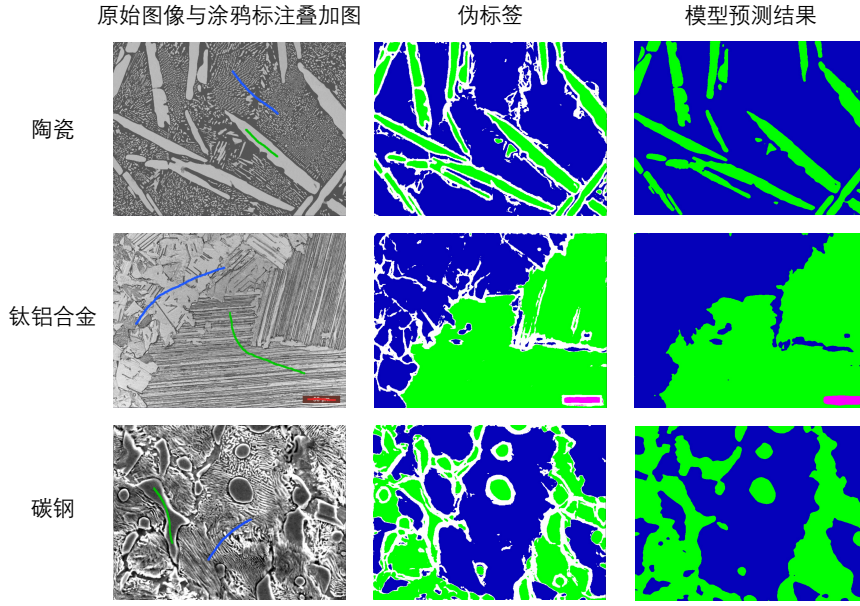


图 4.4 不同材料数据集上的多历史投票伪标签可视化结果

料显微图像小样本场景下随机扰动、样本差异及参数更新所带来的预测震荡更为显著，这种“滑动历史窗口一致性 + 多指标筛选”的联合机制能够在一定程度上抑制偶然一致和错误高置信预测，从而生成更加稳健的伪标签。

4.1.4 边界感知约束

材料显微图像中的相界通常呈现出细长、弯曲、局部对比度不均以及邻近区域纹理相似等特征，因此边界区域往往是最容易产生类别混淆与伪标签噪声的位置。若将边界像素与区域内部像素一并处理，则一方面容易在相界附近生成不可靠伪标签，另一方面也会使模型在优化过程中倾向于过度平滑类别过渡，从而削弱对复杂相界的刻画能力。针对这些问题，MVBC 进一步引入边界感知约束，对边界区域实施区别于普通区域的专门处理。

如前所述，MVBC 中的边界先验并非来自额外人工边界标注，也不是在本章材料数据集上重新训练得到，而是直接采用 PiDiNet^[87] 的公开预训练权重对原始材料显微图像进行离线边界响应提取，生成与原图坐标对齐的边界响应图

$$b_n \in [0, 1]^{H \times W}, \quad (4.14)$$

其中， $b_n(u)$ 的响应值越大，表示像素位置 u 越可能处于真实相界附近。PiDiNet 在 MVBC 中仅作为固定的辅助边界先验提取器使用，其参数在整个分割训练过程中保持不变，不参与 MVBC 分割网络的反向传播与参数更新。因此，MVBC 并未引入

额外人工边界标注监督，而是利用预训练边界检测模型从原始图像中提取结构先验，以辅助伪标签构建阶段的边界阈值过滤与边界感知约束构造。该边界图在训练阶段同时用于伪标签构建阶段的边界阈值过滤与边界损失构造两个环节，从而形成统一的边界感知约束机制。

需要进一步说明的是，边界相关约束中涉及两个不同用途的阈值参数。其一， δ_b 用于伪标签构建阶段的边界阈值过滤，即过滤高边界响应位置，以避免不稳定伪标签沿复杂边界区域传播；其二， τ_{sh} 用于在边界锐化损失中定义高响应边界集合，以突出真实相界附近的类别跃迁。二者功能不同，取值可分别设定。

为直观展示 MVBC 所采用边界先验的来源及其与材料组织相界的对应关系，图 4.5 给出了三类典型材料显微图像上的边界响应图可视化结果。每一行对应一种材料数据集，从左至右依次为原始图像、由 PiDiNet 提取的边界响应图以及原图与边界响应图的叠加显示结果。可以看出，所生成的边界响应能够较好地突出不同材料组织中的相界位置，并在复杂纹理背景下保留主要边界结构，从而为后续伪标签构建中的边界阈值过滤以及边界平滑与锐化约束提供有效先验。

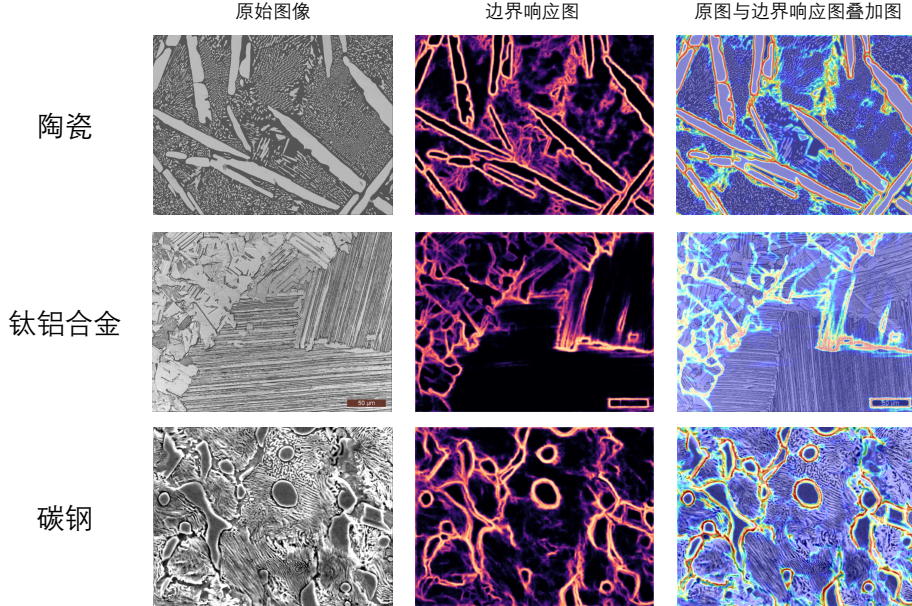


图 4.5 三类材料显微图像上的边界响应图可视化结果

首先，在伪标签生成阶段，边界图被用作边界阈值过滤的判定依据。对于边界响应较高的位置，MVBC 不直接依据历史投票结果构造伪标签，而是通过边界阈值过滤将其保留为忽略状态。其核心动机在于相界附近往往存在真实类别过渡、局部

混相或成像噪声，模型即使在这些区域出现短时一致预测，也未必意味着该预测具备足够可靠性。在伪标签构建阶段显式剔除高边界响应区域，有助于优先在区域内部建立稳定伪监督，再借助后续边界感知约束逐步提升相界刻画能力，从而避免错误伪标签沿边界传播。

为了在网络优化过程中同时增强区域内部一致性与边界过渡清晰度，MVBC 设计了边界平滑损失与边界锐化损失两项约束。设当前样本在原始图像坐标系上的类别概率输出为

$$\mathbf{p}_n \in [0, 1]^{C \times H \times W}, \quad (4.15)$$

其中， $\mathbf{p}_{n,c}(u)$ 表示像素 u 属于第 c 类的预测概率。记 $\mathbf{e}_x$ 和 $\mathbf{e}_y$ 分别为水平方向和垂直方向的单位位移。为与边界平滑项和边界锐化项的实现保持一致，本章分别定义沿两个方向的类别求和差分与类别平均差分：

$$d_{x,n}(u) = \sum_{c=0}^{C-1} |\mathbf{p}_{n,c}(u + \mathbf{e}_x) - \mathbf{p}_{n,c}(u)|, \quad d_{y,n}(u) = \sum_{c=0}^{C-1} |\mathbf{p}_{n,c}(u + \mathbf{e}_y) - \mathbf{p}_{n,c}(u)|, \quad (4.16)$$

$$\bar{d}_{x,n}(u) = \frac{1}{C} d_{x,n}(u), \quad \bar{d}_{y,n}(u) = \frac{1}{C} d_{y,n}(u). \quad (4.17)$$

其中，

$$\Omega_x = \{u \in \Omega \mid u + \mathbf{e}_x \in \Omega\}, \quad \Omega_y = \{u \in \Omega \mid u + \mathbf{e}_y \in \Omega\}, \quad (4.18)$$

分别表示具有有效水平邻域和垂直邻域的像素集合。这里， $d_{x,n}(u)$ 和 $d_{y,n}(u)$ 分别表示像素位置 u 在水平方向和垂直方向上的类别求和差分， $\bar{d}_{x,n}(u)$ 和 $\bar{d}_{y,n}(u)$ 则表示对应的类别平均差分。上述量的值越大，表明相邻像素之间的类别概率分布差异越明显，通常意味着该位置更可能靠近类别边界；其值越小，则表明该区域内部的预测更加平滑一致。

在非边界区域，网络输出应具有较强的区域内部一致性，因此 MVBC 利用边界响应的互补权重构造边界平滑损失

$$\mathcal{L}_{\text{sm}} = \frac{1}{|\Omega_x|} \sum_{u \in \Omega_x} (1 - b_n(u)) \bar{d}_{x,n}(u) + \frac{1}{|\Omega_y|} \sum_{u \in \Omega_y} (1 - b_n(u)) \bar{d}_{y,n}(u), \quad (4.19)$$

其中， Ω_x 和 Ω_y 分别表示具有有效水平邻域和垂直邻域的像素集合。由式中权重项

$(1 - b_n(u))$ 可知，边界响应较弱的位置会被赋予更大的平滑权重，从而更强地抑制区域内部不必要的平均概率振荡；而在边界附近，由于 $b_n(u)$ 较大，该项约束自动减弱，以避免真实类别过渡被过度平滑。因此， $\mathcal{L}_{sm}$ 并非对整幅概率图施加均匀平滑，而是实现了区域内部强平滑、边界邻域弱平滑的结构化约束。

然而，仅依赖边界平滑约束仍不足以获得清晰的相界，因为减弱平滑约束虽有助于保留边界，但并不必然促使模型在边界位置形成足够陡峭的类别过渡。为此，**MVBC** 进一步利用高响应边界区域构造边界锐化损失。定义用于边界锐化的高响应边界集合为

$$\Omega_{n,x}^b = \{u \in \Omega_x \mid b_n(u) > \tau_{sh}\}, \quad \Omega_{n,y}^b = \{u \in \Omega_y \mid b_n(u) > \tau_{sh}\}, \quad (4.20)$$

其中， τ_{sh} 表示边界锐化阈值，用于选取高响应边界区域以计算边界锐化损失。需要注意的是， τ_{sh} 与伪标签构建阶段边界阈值过滤中使用的阈值 δ_b 并不相同：前者用于确定需要被强化的边界区域，后者用于通过边界阈值过滤剔除不适合传播伪标签的高风险边界像素。基于此，边界锐化损失定义为

$$\mathcal{L}_{sh} = \frac{1}{|\Omega_{n,x}^b| + \varepsilon} \sum_{u \in \Omega_{n,x}^b} \exp(-d_{x,n}(u)) + \frac{1}{|\Omega_{n,y}^b| + \varepsilon} \sum_{u \in \Omega_{n,y}^b} \exp(-d_{y,n}(u)). \quad (4.21)$$

其中， $\Omega_{n,x}^b$ 和 $\Omega_{n,y}^b$ 分别表示在水平方向和垂直方向上用于计算边界锐化损失的高响应边界像素集合， ε 为防止分母为零的极小常数。由于函数 $\exp(-x)$ 随 x 的增大而单调减小，最小化 $\mathcal{L}_{sh}$ 等价于鼓励高响应边界区域内的局部概率差分增大，从而增强相邻像素之间的类别跃迁强度，并形成更加清晰的边界过渡。不同于主要作用于区域内部的边界平滑损失， $\mathcal{L}_{sh}$ 更强调在高响应相界附近保留并强化预测概率的突变特性，以提升模型对真实边界位置的刻画能力。

MVBC 中的边界感知约束具有双重作用：一方面，在伪标签构建阶段利用阈值 δ_b 对高风险边界像素执行边界阈值过滤，抑制伪标签噪声传播；另一方面，在网络优化阶段利用边界锐化阈值 τ_{sh} 选取高响应相界区域，并结合边界平滑与边界锐化的协同作用，同时增强区域内部一致性与边界过渡清晰度。前者控制伪监督噪声的注入位置，后者直接塑造预测概率场的空间分布，两者共同提升了模型对复杂材料相界的刻画能力。

4.1.5 损失函数

在训练阶段, **MVBC** 的优化目标由三部分组成: 基于涂鸦标注的部分监督项、基于多历史投票的伪标签监督项以及边界感知正则项。考虑到 **U-Net++** 带有深监督的多输出结构, 本章将所有可用于监督的输出统一记为

$$\{\mathbf{z}_n^{(m)}\}_{m=1}^M, \quad (4.22)$$

其中, M 表示有效监督输出层数, $\mathbf{z}_n^{(m)}(u)$ 表示第 m 个监督输出在像素位置 u 处的类别预测结果。在交叉熵损失计算中, $\mathbf{z}_n^{(m)}(u)$ 与对应像素标签进行匹配监督; 当网络输出尺度与原始标签尺度不一致时, 需要通过上采样或几何逆变换保证二者在空间位置上对齐。当 $M = 1$ 时, 该形式退化为普通单输出分割网络; 当 $M > 1$ 时, 各输出层共同参与训练, 以提高梯度传播的稳定性与多尺度表征能力。

对于涂鸦监督, 仅在已标注像素集合 $\Omega_l = \{u \mid s_n(u) \neq 255\}$ 上计算交叉熵损失, 从而得到部分监督损失

$$\mathcal{L}_{\text{seg}} = \frac{1}{M} \sum_{m=1}^M \frac{1}{|\Omega_l|} \sum_{u \in \Omega_l} \ell_{\text{ce}}(\mathbf{z}_n^{(m)}(u), s_n(u)). \quad (4.23)$$

其中, $\ell_{\text{ce}}(\cdot)$ 表示像素级交叉熵损失, Ω_l 表示当前样本中具有真实涂鸦标注的像素集合。该式表明, 部分监督损失仅在已标注像素位置上计算, 并对所有有效监督输出取平均, 从而保证模型始终以真实涂鸦标注作为基本监督来源。

对于通过多历史投票与三重可靠性筛选得到的可靠像素, **MVBC** 利用上一小节构造的最终伪标签 $\tilde{s}_n^{(t)}$ 及其对应的二值可靠性掩码 $w_n^{(t)}$ 进行附加监督。由于伪标签是在原始图像坐标系下构造的, 因此伪标签损失同样作用于经逆几何变换恢复后的输出结果。记第 m 个监督输出在原始坐标系上的恢复结果为 $\bar{\mathbf{z}}_n^{(m,t)}$, 则伪标签损失可定义为

$$\mathcal{L}_{\text{pl}} = \frac{1}{M} \sum_{m=1}^M \frac{\sum_{u \in \Omega} w_n^{(t)}(u) \ell_{\text{ce}}(\bar{\mathbf{z}}_n^{(m,t)}(u), \tilde{s}_n^{(t)}(u))}{\sum_{u \in \Omega} w_n^{(t)}(u) + \varepsilon}. \quad (4.24)$$

其中, $w_n^{(t)}(u) \in \{0, 1\}$ 用于指示像素位置 u 是否通过由多历史投票、置信度约束和边界阈值过滤构成的三重可靠性筛选; ε 为防止分母为零的极小常数。由该式可见, 只有被判定为可靠的像素才会对伪标签损失产生贡献; 未通过筛选的位置由于 $w_n^{(t)}(u) = 0$, 将被自动屏蔽。因此, $\mathcal{L}_{\text{pl}}$ 本质上是在可靠伪标签区域上施加的加权附

加监督，其作用在于从未标注区域中挖掘稳定的补充监督信号，同时避免不可靠伪标签干扰主监督的稳定优化。

边界感知正则项由边界平滑损失 $\mathcal{L}_{sm}$ 与边界锐化损失 $\mathcal{L}_{sh}$ 组成。前者强调非边界区域的类内一致性，后者强调高响应边界区域的类别过渡清晰度。综合上述各项，MVBC 的总损失函数可写为

$$\mathcal{L}_{total} = \mathcal{L}_{seg} + \lambda_{pl}\mathcal{L}_{pl} + \lambda_{sm}\mathcal{L}_{sm} + \lambda_{sh}\mathcal{L}_{sh}. \quad (4.25)$$

其中， λ_{pl} 、 λ_{sm} 和 λ_{sh} 分别表示伪标签监督项、边界平滑项和边界锐化项的权重系数，用于平衡不同损失项在训练过程中的相对贡献。由该式可知，MVBC 的整体优化目标由真实涂鸦监督、可靠伪标签监督和边界感知正则项三部分共同构成： $\mathcal{L}_{seg}$ 保证模型具备基本的分割判别能力， $\mathcal{L}_{pl}$ 用于补充未标注区域中的有效监督信息，而 $\mathcal{L}_{sm}$ 与 $\mathcal{L}_{sh}$ 则分别从区域内部平滑性与边界清晰性两个角度对预测结果进行结构化约束。

为便于整体理解 MVBC 的训练流程，算法 4.1 给出了基于多历史投票伪标签与边界感知约束的完整优化过程，其中， id_n 表示样本 x_n 在训练集中的唯一标识，用于索引按样本维护的历史预测库。

4.2 实验分析

4.2.1 数据集介绍

为验证 MVBC 在小样本弱监督材料组织分割任务中的有效性与普适性，本章在三类典型材料显微图像数据集上进行了实验，分别为陶瓷数据集、钛铝合金（TiAl）数据集以及碳钢数据集。典型样本及对应的涂鸦标注如图 4.6 所示，图中从左至右依次为陶瓷数据集、钛铝合金数据集和碳钢数据集。彩色线条为专家绘制的稀疏涂鸦标注，其中不同颜色对应不同组织类别。

三类数据集分别对应不同的显微组织分割场景，能够较好覆盖多相材料显微图像中常见的组织形貌差异、相间边界复杂性以及标注稀缺等问题。

其中，陶瓷数据集主要体现典型陶瓷相组织的分布特征。整体来看，各相区域形态相对规则，但图像背景中存在一定纹理干扰，对模型区分细微组织差异的能力提出了较高要求。图中采用不同颜色的涂鸦对相应陶瓷相区域进行标记，以提供稀疏监督信息。

算法 4.1: 基于多历史投票与边界感知约束的训练算法**Input:** 训练集 $D = \{(x_n, s_n, id_n)\}_{n=1}^N$, 预生成边界响应图集合 $\mathcal{B}_{\text{bdry}} = \{b_n\}_{n=1}^N$, 分割网络 f_θ , 超参数 $\{K_{\text{hist}}, K_{\text{min}}, \delta_v, \delta_c, \delta_b, \lambda_{\text{pl}}, \lambda_{\text{sm}}, \lambda_{\text{sh}}\}$ **Output:** 更新后的参数 θ

```

1  初始化历史预测库  $\mathcal{H} \leftarrow \emptyset$ 
2  for  $epoch = 1$  to  $T$  do
3      foreach 小批次  $\mathcal{B} = \{(x_j, s_j, id_j)\}_{j=1}^B$  do
4          读取对应边界响应图  $\{b_j\}_{j=1}^B$ 
5          前向传播得到预测输出  $\{\mathbf{z}_j^{(m)}\}$ , 并恢复至原始坐标系
6          计算涂鸦监督损失  $\mathcal{L}_{\text{seg}}$  与边界感知正则项  $\mathcal{L}_{\text{sm}}, \mathcal{L}_{\text{sh}}$ 
7           $\mathcal{L}_{\text{pl}} \leftarrow 0$ 
8          for  $j \leftarrow 1$  to  $B$  do
9              由当前预测得到  $\hat{y}_j^{(t)}$  和  $r_j^{(t)}$ 
10             if  $|\mathcal{H}[id_j]| \geq K_{\text{min}}$  then
11                 由  $\mathcal{H}[id_j]$  进行多历史投票, 得到  $\bar{y}_j^{(t)}$ 、 $\eta_j^{(t)}$  和  $\bar{\rho}_j^{(t)}$ 
12                 根据  $(\eta_j^{(t)} \geq \delta_v) \wedge (\bar{\rho}_j^{(t)} \geq \delta_c) \wedge (b_j \leq \delta_b)$  执行三重可靠性筛选
13                 (其中  $b_j \leq \delta_b$  对应边界阈值过滤), 并构造伪标签  $\tilde{s}_j^{(t)}$  及掩码
14                  $w_j^{(t)}$ 
15             end
16             将  $(\hat{y}_j^{(t)}, r_j^{(t)})$  写入历史库  $\mathcal{H}[id_j]$ 
17         end
18         根据当前小批次的  $\tilde{s}_j^{(t)}$  和  $w_j^{(t)}$  计算伪标签损失  $\mathcal{L}_{\text{pl}}$ 
19          $\mathcal{L}_{\text{total}} \leftarrow \mathcal{L}_{\text{seg}} + \lambda_{\text{pl}}\mathcal{L}_{\text{pl}} + \lambda_{\text{sm}}\mathcal{L}_{\text{sm}} + \lambda_{\text{sh}}\mathcal{L}_{\text{sh}}$ 
20         反向传播并更新  $\theta$ 
21 end
22 return  $\theta$ 

```

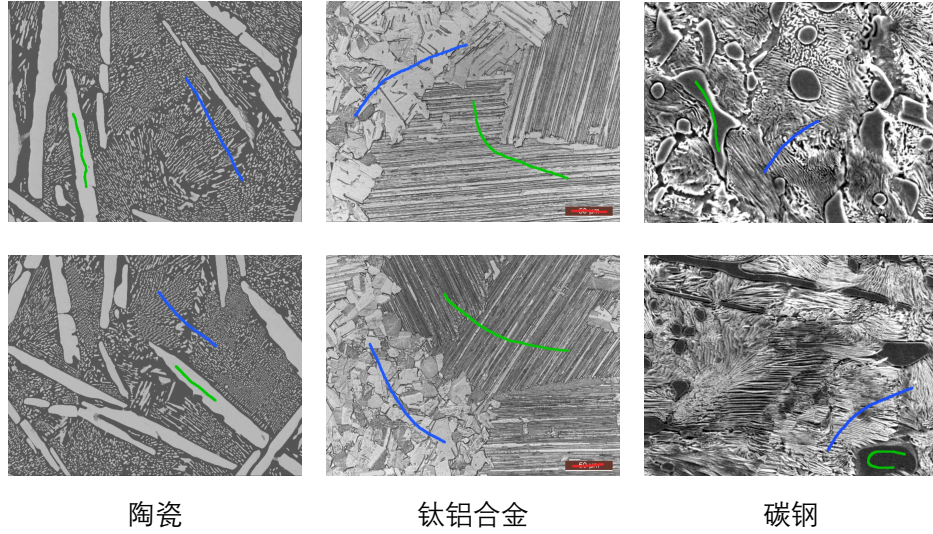


图 4.6 实验所用材料显微图像数据集及涂鸦标注示例

钛铝合金数据集则包含等轴组织和片层组织两种典型显微形貌，同时图像中还包含比例尺区域。对应地，蓝色涂鸦用于指示等轴组织，绿色涂鸦用于指示片层组织，红色涂鸦用于标识比例尺。由于不同组织内部纹理差异较为明显，而部分区域的相界过渡又较为模糊，该数据集能够更有效地检验模型对异质纹理和复杂边界的表征能力。

碳钢数据集主要反映碳钢显微组织中不同组织区域的分布特征。此类图像通常具有纹理细密、局部对比度波动明显以及相界形态复杂等特点，在弱监督条件下，模型容易在相邻组织交界处产生类别混淆。图中的不同颜色涂鸦用于标记对应的碳钢组织类别，从而提供稀疏监督信息。基于这些特征，该数据集能够有效检验模型在复杂金属组织场景下的区域识别能力与边界刻画能力。

考虑到材料显微图像像素级标注成本高且专业性强，实验采用与 PTS 方法^[51]一致的公开材料显微图像数据集及其实验协议。本章关注极小样本涂鸦监督设定，因此每个数据集仅选取极少量图像作为训练样本，其余图像用于测试。训练阶段不使用完整像素级标注作为监督，而仅使用与 PTS 一致的涂鸦监督信号，未被涂鸦覆盖的像素均视为未标注区域，不参与监督损失计算。

4.2.2 实验设置

MVBC 基于 PyTorch 框架实现，所有实验均在配备 NVIDIA GeForce RTX 3090 GPU (24 GB 显存) 的计算平台上完成。原始材料显微图像的尺寸为 640×480 像素。

为保证输入图像、涂鸦标签图与边界响应图在空间几何增强及逆变换过程中的对齐，实验参考 PTS^[51] 的预处理流程，将三者统一调整为 640×640 像素。其中，图像与边界响应图采用双线性插值进行缩放，标签图采用最近邻插值，以避免类别混叠。边界响应图由原始材料显微图像预先生成，并在训练阶段作为辅助先验使用。输入图像像素值归一化至 $[0, 1]$ 区间。在训练阶段，模型仅利用稀疏涂鸦标注作为监督信号，未标注区域被设为忽略标签，不参与监督损失的计算。

在数据划分与网络配置方面，实验遵循小样本弱监督设定。针对每类材料数据，仅使用极少量带涂鸦标注的图像作为训练集，其余样本用于性能评估。MVBC 采用以预训练 VGG16 为编码骨干的 U-Net++ 作为分割主干网络。为提升小样本条件下训练的稳定性，编码器的前四个阶段保持冻结状态，仅对最深层编码模块及所有解码模块进行微调。优化器采用 Adam，初始学习率设为 0.003，批量大小为 1。学习率调度采用 MultiStepLR 策略，并在第 100 个 epoch 时将学习率衰减为原来的 $1/3$ 。总训练轮数统一设置为 120。考虑到极小样本弱监督场景下模型性能易受随机初始化、数据增强顺序以及样本访问顺序等因素影响，每组实验设置下均采用 5 个不同随机种子进行独立重复实验，并以 5 次运行结果的平均值作为最终结果，以增强实验结论的可靠性与稳定性。

训练过程中同时应用空间几何增强与像素值增强。空间增强包括随机旋转与翻转操作，且网络输出需经对应的逆变换恢复至原始坐标系，以确保历史投票伪标签的构造与边界感知约束在统一坐标系下完成。针对多历史投票伪标签机制，MVBC 为每个训练样本维护一个独立更新的有限长度预测队列。默认历史长度 $K_{\text{hist}} = 5$ ，最小启用长度 $K_{\text{min}} = 5$ 。当某一像素在历史窗口内的投票占比不低于 0.8 且支持该类别的平均置信度不低于 0.8 时，该像素方可进入候选伪标签集合。此外，若像素在边界响应图上的值高于边界阈值过滤中使用的阈值 $\delta_b = 0.6$ ，则该位置被视为高风险边界区域，并通过边界阈值过滤从伪标签监督中剔除。

在边界感知约束方面，MVBC 利用预生成的边界响应图作为先验引导。边界正则项由边界平滑项与边界锐化项组成，其中边界锐化阈值设为 $\tau_{\text{sh}} = 0.7$ ，用于选取高响应边界区域以计算边界锐化损失。参数敏感性分析表明，当 $(\lambda_{\text{pl}}, \lambda_{\text{sm}}, \lambda_{\text{sh}}) = (0.1, 0.1, 0.01)$ 时，模型在伪标签监督、区域平滑与边界强化之间取得了较好的平衡。因此，后续实验统一采用该组参数作为默认设置。

4.2.3 对比实验

为验证 MVBC 在小样本涂鸦监督材料显微图像分割任务中的有效性，本章选取多种具有代表性的分割与弱监督分割方法进行对比实验。具体包括基于涂鸦标注训练的 U-Net^[4]、U-Net++^[5]、U-Net 3+^[88] 和 CE-Net^[89]，以及弱监督方法 SPML^[60]、TEL^[90]、CES、PTS^[51] 和 ProCNS^[91]。上述方法既包含经典的编码器—解码器分割网络，也包含面向弱监督场景设计的代表性方法，能够较为全面地呈现所提方法在材料显微图像小样本分割任务中的性能表现。其中，U-Net、U-Net++、U-Net 3+ 和 CE-Net 作为监督分割基线，用于比较不同主干结构在稀疏标注条件下的分割能力；SPML、TEL、CES、PTS 与 ProCNS 则作为弱监督对比方法，用于验证 MVBC 中多历史投票伪标签机制与边界感知约束在提升区域识别稳定性、抑制伪标签噪声传播以及增强复杂边界刻画能力方面的有效性。为尽可能保证比较公平，各方法采用一致的数据划分、训练集规模、输入预处理、训练轮数及评价指标设置，并统一使用 mIoU 和 mDice 进行定量评价。需要说明的是，CES 原方法属于交互式涂鸦分割方法，其完整流程通常包含基于用户反馈的多轮交互修正。考虑到本章关注的是离线的小样本单轮涂鸦监督分割设定，实验中未采用 CES 的多轮人机交互补标流程，而是在与其他方法相同的初始涂鸦标注下进行单轮训练与评估。因此，本章报告的 CES 结果反映的是其在非交互、单轮涂鸦监督设定下的性能，而不代表其完整交互模式下的最优性能上限。各方法在陶瓷、钛铝合金和碳钢数据集上的实验结果分别见表 4.1、表 4.2 和表 4.3。

从表 4.1–表 4.3 可以看出，MVBC 在三类材料显微图像数据集上均取得了最优分割性能，表明所提出的多历史投票伪标签机制与边界感知约束在小样本涂鸦监督场景下具有较强的有效性与稳定性。具体来看，在陶瓷数据集上，MVBC 的 mIoU 和 mDice 分别达到 93.9% 和 95.3%，相较于性能最优的对比方法 PTS，分别提升了 8.0 个百分点和 3.1 个百分点；在钛铝合金数据集上，MVBC 的 mIoU 和 mDice 分别达到 85.2% 和 91.7%，优于所有对比方法，说明 MVBC 在极少标注条件下仍能够获得较强的组织识别能力；在碳钢数据集上，MVBC 的 mIoU 和 mDice 分别达到 81.2% 和 87.1%，较次优方法 PTS 分别提升了 7.6 个百分点和 3.9 个百分点，进一步验证了 MVBC 在复杂纹理和模糊边界场景中的优势。

不同对比方法在三类数据集上的表现存在明显差异。以 U-Net、U-Net++、U-Net

表 4.1 不同方法在陶瓷数据集分割任务上的分割结果

模型	标注方式	mIoU (%)	mDice (%)
U-Net	涂鸦标注	68.9	81.2
U-Net++	涂鸦标注	71.5	82.5
U-Net 3+	涂鸦标注	71.3	82.7
CE-Net	涂鸦标注	62.5	76.2
SPML	涂鸦标注	48.1	63.0
TEL	涂鸦标注	85.3	91.7
CES	涂鸦标注	69.5	80.8
ProCNS	涂鸦标注	82.2	89.9
PTS	涂鸦标注	85.9	92.2
MVBC (所提方法)	涂鸦标注	93.9	95.3

表 4.2 不同方法在钛铝合金数据集分割任务上的分割结果

模型	标注方式	mIoU (%)	mDice (%)
U-Net	涂鸦标注	38.1	51.5
U-Net++	涂鸦标注	37.0	50.8
U-Net 3+	涂鸦标注	41.6	53.9
CE-Net	涂鸦标注	46.1	58.2
SPML	涂鸦标注	62.4	75.4
TEL	涂鸦标注	84.5	91.5
CES	涂鸦标注	32.9	45.3
ProCNS	涂鸦标注	79.1	88.4
PTS	涂鸦标注	78.4	87.9
MVBC (所提方法)	涂鸦标注	85.2	91.7

表 4.3 不同方法在碳钢数据集分割任务上的分割结果

模型	标注方式	mIoU (%)	mDice (%)
U-Net	涂鸦标注	47.1	61.5
U-Net++	涂鸦标注	44.8	58.7
U-Net 3+	涂鸦标注	53.8	65.2
CE-Net	涂鸦标注	40.7	55.3
SPML	涂鸦标注	50.6	63.4
TEL	涂鸦标注	44.4	57.9
CES	涂鸦标注	54.3	65.5
ProCNS	涂鸦标注	64.3	71.1
PTS	涂鸦标注	73.6	83.2
MVBC (所提方法)	涂鸦标注	81.2	87.1

3+ 和 CE-Net 为代表的监督分割基线，在仅有稀疏涂鸦监督的条件下总体性能有限，说明单纯依赖局部标注像素难以有效恢复材料组织的完整区域结构。SPML、TEL、CES、PTS 和 ProCNS 等弱监督方法虽然能够在一定程度上提升分割精度，但其性能在不同数据集之间波动较大，反映出已有方法在面对小样本材料显微图像时，仍然容易受到局部纹理相似、边界模糊以及伪监督噪声累积的影响。相比之下，MVBC 在三类数据集上均保持了较高且稳定的性能水平，说明通过在时间维度上累积并筛选历史预测信息，可以有效缓解单次预测不稳定带来的伪标签误差；同时，边界感知约束能够进一步减少复杂相界附近的误分割与噪声传播，从而提升整体分割质量。

为了更直观地展示不同方法的分割效果，图 4.7、图 4.8 和图 4.9 给出了三组典型样本上的可视化对比结果。可以看到，MVBC 在区域完整性、边界清晰度以及局部细节保留方面均表现出更优的视觉效果。

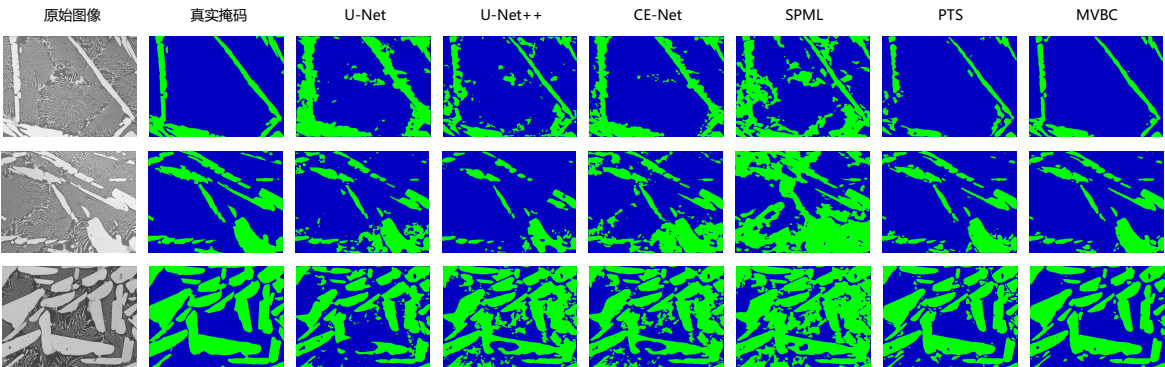


图 4.7 不同模型在陶瓷数据集上的分割结果对比

如图 4.7 所示，在陶瓷数据集中，目标区域整体轮廓相对清晰，但仍存在细长结构、狭窄边界以及局部组织连接处等难分割区域。U-Net、U-Net++ 和 CE-Net 虽然能够恢复主体区域，但在细长相区和边缘转折位置仍出现不同程度的漏分割或边界偏移；SPML 在部分区域出现了较明显的过分割现象，预测结果中包含较多与真实结构不一致的噪声区域；PTS 的结果相对较好，但在个别细窄结构和边界连接处仍存在局部误差。相比之下，MVBC 与真实掩码的形状一致性更高，不仅能够较完整地恢复目标主体区域，还能够在细长边界和局部转折区域保持较好的连续性，体现出所提方法在抑制边界误差传播方面的优势。

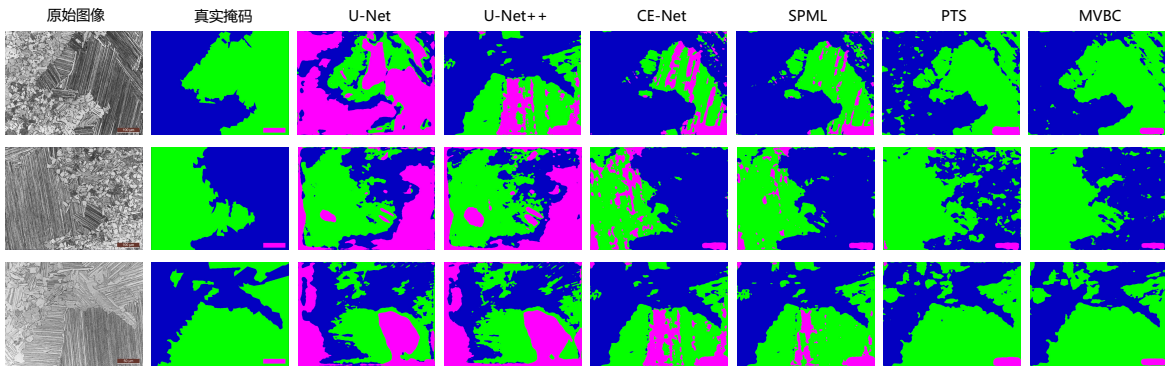


图 4.8 不同模型在钛铝合金数据集上的分割结果对比

如图 4.8 所示，钛铝合金数据集包含多个组织类别及比例尺区域，不同类别之间不仅具有明显的尺度差异，而且在部分边界位置还存在较强的纹理相似性，因此对模型的多类别判别能力和边界定位能力提出了更高要求。对于该类样本，U-Net 和 U-Net++ 在类别交界区域容易出现混淆，导致不同类别之间的分界不够清晰；CE-Net 与 SPML 在局部区域存在明显的类别侵蚀和结构失真；PTS 虽然恢复出了主要组织分布，但在细节边界以及小区域类别识别方面仍存在误差。相比之下，MVBC 不仅能够更准确地区分不同组织类别，还能较好保留比例尺区域与主组织区域之间的清晰分界，预测边界与真实掩码更加一致，说明 MVBC 在复杂多类别场景中同样具有良好的适应能力。

如图 4.9 所示，碳钢数据集具有更复杂的局部纹理和更加破碎的组织分布，属于弱监督条件下更具挑战性的分割场景。可以观察到，在该类样本上，部分方法容易出现大范围噪声散点、区域粘连以及背景误检等问题，其中 SPML 和部分基线模型在未标注区域内生成了较多错误响应，说明其对复杂纹理扰动较为敏感；PTS 虽然能够较好恢复主要区域，但在小尺寸组织和边界邻近位置仍存在一定缺失。相比之下，

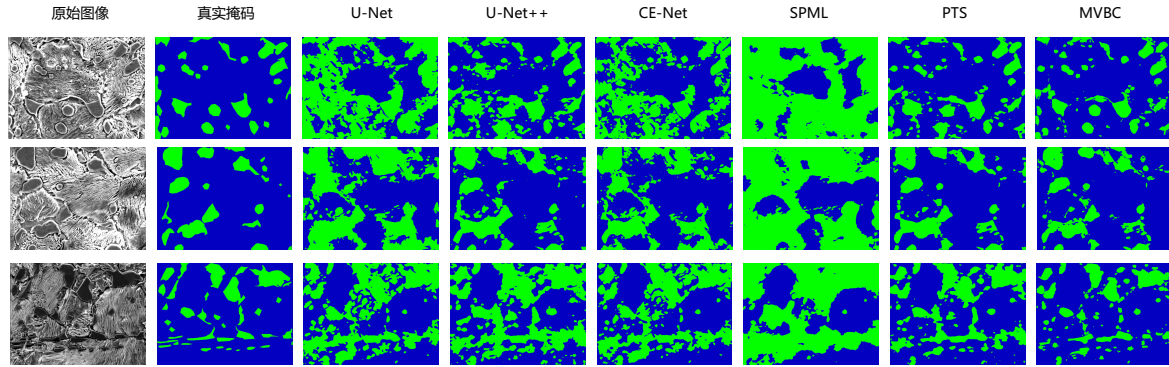


图 4.9 不同模型在碳钢数据集上的分割结果对比

MVBC 所得到的预测结果与真实掩码更加接近，既减少了背景中的离散误检，也更好地保留了细小组织区域的完整性。这说明多历史投票机制能够通过时间一致性筛选不稳定预测，使伪监督主要作用于高置信度且稳定的区域，从而提升复杂纹理场景下的分割鲁棒性。

综合定量结果与可视化结果可见，MVBC 对材料显微图像小样本涂鸦监督分割中的三个关键问题均具有针对性作用：预训练初始化与分层迁移学习为小样本条件下的稳定训练提供了基础，多历史投票机制通过对像素级历史预测结果进行统计投票建模，提升了未标注区域伪标签的时间稳定性与可靠性，边界感知约束则抑制了复杂相界附近的噪声传播并增强了边界表达能力。正因如此，MVBC 能够在极少量涂鸦标注的条件下仍获得稳定且精细的分割结果。

4.2.4 消融实验

表 4.4 MVBC 各模块在陶瓷数据集上的消融实验结果

模型	部分冻结	数据增强	预训练	伪标签	$\mathcal{L}_{sm}$	$\mathcal{L}_{sh}$	mIoU(%)		mDice(%)	
							一张	两张	一张	两张
Baseline							29.7	48.3	45.8	65.1
A1	✓						37.7	54.6	47.5	65.5
A2	✓	✓					47.3	58.3	53.2	69.3
A3	✓	✓	✓				74.3	75.9	82.3	83.9
A4	✓	✓	✓	✓			87.0	89.2	90.0	90.8
A5	✓	✓	✓	✓	✓		89.5	91.9	92.6	93.4
A6	✓	✓	✓	✓		✓	92.2	92.5	93.0	93.5
MVBC	✓	✓	✓	✓	✓	✓	93.9	94.8	95.3	95.9

表 4.4、表 4.5和表 4.6 分别给出了 MVBC 在陶瓷、钛铝合金和碳钢三个材料数据集上的消融实验结果。为系统分析各组成模块对模型性能的影响，实验以基础模型为起点，依次引入部分冻结、数据增强、预训练、伪标签监督、边界平滑损失 $\mathcal{L}_{sm}$ 以及边界锐化损失 $\mathcal{L}_{sh}$ ，并分别在仅使用一张和两张图像作为训练集的条件下进行评估。整体来看，随着各模块逐步加入，模型在三个数据集上的 mIoU 和 mDice 均呈现上升趋势，说明 MVBC 中各项设计均对小样本涂鸦监督材料显微图像分割具有积极作用。

首先，从 Baseline 到 A3 的结果可以看出，部分冻结、数据增强和预训练策略构成了 MVBC 性能提升的重要基础。其中，部分冻结策略在三个数据集上的提升较为稳定，说明在小样本条件下保留预训练编码器浅层通用特征，有助于抑制过拟合并增强训练稳定性。在此基础上，进一步引入数据增强后，模型在仅使用一张图像和两张图像作为训练集时的性能均有所改善，表明几何变换与像素值扰动能够有效扩充样本多样性，增强模型对不同纹理变化和组织形貌差异的适应能力。

当加入预训练初始化后，模型性能在三个数据集上均出现明显提升。以陶瓷数据集为例，A3 模型在仅使用一张图像作为训练集时，mIoU 提升至 74.3%，在使用两张图像作为训练集时进一步达到 75.9%；对应的 mDice 也分别提升至 82.3% 和 83.9%。在钛铝合金数据集上，A3 的 mIoU 在一张和两张训练图像条件下分别达到 67.9% 和 82.6%，mDice 分别达到 80.9% 和 87.4%。在碳钢数据集上，A3 的 mIoU 在两种训练规模下分别为 66.0% 和 71.7%，对应的 mDice 分别为 76.3% 和 81.2%。这些结果表明，预训练特征对于材料显微图像小样本分割任务尤为关键，能够显著增强模型对复杂组织纹理和区域结构的表征能力。

A4 的结果验证了多历史投票伪标签机制的有效性。在三个数据集上，当引入伪标签监督后，模型性能均在 A3 的基础上进一步提升。具体而言，在陶瓷数据集上，A4 在仅使用一张图像作为训练集时，mIoU 达到 87.0%，当训练图像增加到两张时进一步提升至 89.2%；对应的 mDice 分别达到 90.0% 和 90.8%。在钛铝合金数据集上，A4 的 mIoU 在一张和两张训练图像条件下分别达到 77.2% 和 83.9%，mDice 分别达到 83.2% 和 88.1%。在碳钢数据集上，A4 的 mIoU 分别提升至 76.7% 和 81.7%，对应的 mDice 分别达到 79.5% 和 83.5%。这说明，在真实涂鸦监督极为有限的情况下，按样本维护历史预测，并通过多轮投票构造可靠伪标签，能够有效从未标注区域中挖掘稳定的附加监督信号，从而显著提升区域识别能力和整体分割性能。

表 4.5 MVBC 各模块在钛铝合金数据集上的消融实验结果

模型	部分冻结	数据增强	预训练	伪标签	$\mathcal{L}_{sm}$	$\mathcal{L}_{sh}$	mIoU(%)		mDice(%)	
							一张	两张	一张	两张
Baseline							26.1	36.8	41.4	53.8
A1	✓						32.1	49.8	43.8	58.8
A2	✓	✓					41.9	56.4	53.7	66.6
A3	✓	✓	✓				67.9	82.6	80.9	87.4
A4	✓	✓	✓	✓			77.2	83.9	83.2	88.1
A5	✓	✓	✓	✓	✓		81.6	85.7	84.5	90.7
A6	✓	✓	✓	✓		✓	82.6	86.8	86.9	89.6
MVBC	✓	✓	✓	✓	✓	✓	85.2	87.7	91.7	93.1

A5 与 A6 的实验结果表明, 边界感知约束同样对模型性能提升具有重要作用。当加入边界平滑损失 $\mathcal{L}_{sm}$ 后, 模型在三个数据集上均较 A4 取得进一步改善。例如, 在陶瓷数据集上, A5 在一张和两张训练图像条件下的 mIoU 分别提升至 89.5% 和 91.9%, mDice 分别达到 92.6% 和 93.4%。在钛铝合金数据集上, A5 的 mIoU 分别达到 81.6% 和 85.7%, 对应的 mDice 分别为 84.5% 和 90.7%。在碳钢数据集上, A5 的 mIoU 分别达到 79.3% 和 82.1%, mDice 分别达到 84.6% 和 87.0%。这些结果表明, 边界平滑约束能够有效增强非边界区域的类内一致性, 从而减少局部噪声带来的不稳定预测。

相比之下, 仅加入边界锐化损失 $\mathcal{L}_{sh}$ 的 A6 同样表现出明显的性能提升。在陶瓷数据集上, A6 在仅使用一张图像和两张图像作为训练集时, mIoU 分别达到 92.2% 和 92.5%, mDice 分别达到 93.0% 和 93.5%。在钛铝合金数据集上, A6 的 mIoU 分别达到 82.6% 和 86.8%, mDice 分别为 86.9% 和 89.6%。在碳钢数据集上, A6 的 mIoU 分别达到 79.4% 和 83.3%, mDice 分别达到 84.8% 和 87.4%。这一现象说明, 对于材料显微图像中普遍存在的细长、弯曲且模糊的组织边界而言, 边界锐化约束能够更直接地增强边界处的类别跃迁强度, 从而提高复杂相界附近的分割清晰度。

完整模型 MVBC 在三个数据集、两种训练样本规模设置下均取得了最优结果, 进一步验证了各模块之间的协同互补性。具体而言, 在仅使用一张图像作为训练集时, MVBC 在陶瓷、钛铝合金和碳钢数据集上的 mIoU 分别达到 93.9%、85.2% 和 81.2%, 对应的 mDice 分别达到 95.3%、91.7% 和 87.1%; 当训练集扩展为两张图像后, MVBC 在陶瓷、钛铝合金和碳钢数据集上的 mIoU 分别为 94.8%、87.7% 和 85.1%,

表 4.6 MVBC 各模块在碳钢数据集上的消融实验结果

模型	部分冻结	数据增强	预训练	伪标签	$\mathcal{L}_{sm}$	$\mathcal{L}_{sh}$	mIoU(%)		mDice(%)	
							一张	两张	一张	两张
Baseline							42.9	46.1	60.0	63.1
A1	✓						43.1	48.1	61.7	64.4
A2	✓	✓					47.2	60.0	62.8	67.3
A3	✓	✓	✓				66.0	71.7	76.3	81.2
A4	✓	✓	✓	✓			76.7	81.7	79.5	83.5
A5	✓	✓	✓	✓	✓		79.3	82.1	84.6	87.0
A6	✓	✓	✓	✓		✓	79.4	83.3	84.8	87.4
MVBC	✓	✓	✓	✓	✓	✓	81.2	85.1	87.1	89.4

其中陶瓷数据集在训练图像由一张增加到两张后，mIoU 由 93.9% 提升至 94.8%；对应的 mDice 分别达到 95.9%、93.1% 和 89.4%。与仅加入单一边界损失项的模型相比，完整模型在三个数据集上均表现出更优性能，说明边界平滑与边界锐化并非相互替代，而是分别从区域内部一致性和边界过渡清晰性两个方面共同优化预测结果。

表 4.7 多历史投票伪标签机制消融实验结果

伪标签策略	mIoU(%)			mDice(%)		
	陶瓷	钛铝合金	碳钢	陶瓷	钛铝合金	碳钢
不使用伪标签监督	74.3	67.9	66.0	82.3	80.9	76.3
前一轮次单次预测伪标签	74.8	68.1	66.2	82.4	81.6	76.6
多历史投票伪标签	84.6	75.8	74.9	88.4	82.5	78.6
多历史投票 + 置信度筛选	86.1	76.6	75.9	89.3	83.1	79.1
多历史投票 + 置信度筛选 + 边界阈值过滤	87.0	77.2	76.7	90.0	83.2	79.5

表 4.7 给出了基于单张训练图像时不同伪标签策略的消融结果。可以看出，不使用伪标签监督时，模型在三个数据集上的性能相对有限；而直接采用前一轮次单次预测结果作为伪标签时，mIoU 和 mDice 仅有极小幅度提升，说明在极小样本条件下，单次预测容易受到随机扰动和局部噪声影响，难以提供稳定可靠的附加监督。

相比之下，多历史投票伪标签能够明显提升分割性能。以 mIoU 为例，陶瓷、钛铝合金和碳钢数据集分别由单次预测伪标签时的 74.8%、68.1% 和 66.2% 提升至 84.6%、75.8% 和 74.9%，说明相比直接采用单次历史预测结果，结合最近若干训练访问时刻的预测一致性更有助于削弱偶然扰动对伪标签生成的影响。需要注意的是，这里的多

历史投票并非将多个历史预测视为相互独立的统计样本，而是利用同一样本在滑动训练窗口内的预测稳定性来进行经验性筛选。在此基础上进一步加入置信度筛选和边界阈值过滤后，模型性能继续提升，最终分别达到 87.0%、77.2% 和 76.7% 的 mIoU。在单张训练图像的小样本场景下，伪标签质量对模型性能具有重要影响。MVBC 中的筛选策略，能够从时间一致性、预测置信水平和边界阈值过滤三个方面共同约束伪标签生成过程，从而为未标注区域提供更加稳健的监督信号。

4.2.5 敏感性分析

为进一步分析 MVBC 中关键超参数对分割性能的影响，并验证所设参数组合的合理性，本章围绕损失函数权重、多历史投票长度以及伪标签与边界感知约束中的核心阈值开展了参数敏感性分析。考虑到在小样本弱监督场景中，模型对超参数设置较为敏感。若参数设置不当，容易导致伪标签噪声引入过多、边界约束过强或有效监督覆盖不足等问题。因此，有必要系统考察不同参数取值下模型性能的变化规律。除特别说明外，实验均在单张训练图像设定下进行分析，并采用 mIoU 和 mDice 作为评价指标。

表 4.8 损失函数权重的参数敏感性分析结果

变化参数	取值	mIoU(%)	mDice(%)
λ_{pl} ($\lambda_{sm} = 0.1, \lambda_{sh} = 0.01$)	0.01	89.4	90.6
	0.05	92.7	93.2
	0.1	93.9	95.3
	0.5	91.6	92.2
λ_{sm} ($\lambda_{pl} = 0.1, \lambda_{sh} = 0.01$)	0.01	92.5	92.9
	0.05	93.3	93.7
	0.1	93.9	95.3
	0.5	92.4	92.8
λ_{sh} ($\lambda_{pl} = 0.1, \lambda_{sm} = 0.1$)	0.001	91.9	92.4
	0.005	93.0	93.5
	0.01	93.9	95.3
	0.05	92.6	93.0

首先，表 4.8 给出了损失函数中三个权重系数的敏感性分析结果。可以看出，当伪标签监督权重 λ_{pl} 从 0.01 增加到 0.1 时，模型性能持续提升，其中 mIoU 由 89.4%

提升至 93.9%，mDice 由 90.6% 提升至 95.3%。这表明在适当增大伪标签监督强度后，模型能够更充分地利用未标注区域中由多历史投票构造得到的可靠监督信息。然而，当 λ_{pl} 进一步增大到 0.5 时，性能反而下降，说明若伪标签项在总损失中占比过大，会使模型过度依赖附加监督，从而放大少量伪标签误差对训练过程的干扰。

类似地，边界平滑权重 λ_{sm} 与边界锐化权重 λ_{sh} 同样呈现先升后降的变化趋势。当 $\lambda_{sm} = 0.1$ 时，模型取得最优结果；若该值过小，则难以充分抑制区域内部不必要的局部概率振荡；若该值过大，则会导致非边界区域受到过强平滑约束，从而削弱部分细节结构。对于 λ_{sh} 而言，当取值从 0.001 增加至 0.01 时，模型性能逐步提升；但当其增大到 0.05 后，mIoU 和 mDice 均出现下降，说明边界锐化项过强时可能在局部边界处引入过度激进的类别跃迁，反而影响整体分割稳定性。综合来看， $(\lambda_{pl}, \lambda_{sm}, \lambda_{sh}) = (0.1, 0.1, 0.01)$ 能够在伪标签监督、区域平滑与边界强化之间取得较好的平衡，因此被选作后续实验的默认设置。

表 4.9 历史长度 K_{hist} 的参数敏感性分析结果

K_{hist}	mIoU(%)			mDice(%)		
	陶瓷	钛铝合金	碳钢	陶瓷	钛铝合金	碳钢
1	89.1	80.5	76.8	89.6	87.1	82.6
3	92.4	83.6	79.5	92.7	90.2	85.5
5	93.9	85.2	81.2	95.3	91.7	87.1
7	93.2	84.8	80.6	93.7	91.2	86.4
9	92.0	83.5	79.2	92.3	89.9	85.2

表 4.9 给出了历史长度 K_{hist} 对模型性能的影响。可以观察到，当 K_{hist} 从 1 增大到 5 时，三个数据集上的 mIoU 和 mDice 均明显提升；而当其继续增大到 7 和 9 时，性能则有所回落。这说明历史窗口长度对伪标签稳定性与历史信息时效性之间的平衡具有直接影响：窗口过短时，伪标签容易受到单次预测波动影响；窗口过长时，早期不稳定预测又可能被保留在投票过程中，从而干扰当前阶段的伪标签生成。

具体而言，当 $K_{hist} = 1$ 时，伪标签实质上退化为基于单次历史预测的监督，其稳定性较差，难以有效平滑由数据增强、参数更新和局部纹理噪声带来的偶然性误判；当历史长度增加到 3 和 5 时，多历史投票能够在滑动时间窗口内综合最近若干次预测结果，有助于降低单次扰动对伪标签结果的影响，从而提升监督质量。尤其是，当 $K_{hist} = 5$ 时，三个数据集均取得最优结果，说明该设置在预测稳定性和历史

信息时效性之间取得了较好折中。相反,当历史窗口继续增大到 7 或 9 时,更早阶段的预测结果也会被纳入投票过程,而这些预测可能来自模型尚未充分稳定的训练阶段,包含更多边界混淆和类别误判信息,容易削弱当前较优预测的引导作用。因此,本文最终将 K_{hist} 设定为 5。

表 4.10 投票占比阈值 δ_v 的参数敏感性分析结果

δ_v	mIoU(%)			mDice(%)		
	陶瓷	钛铝合金	碳钢	陶瓷	钛铝合金	碳钢
0.6	91.5	82.8	78.5	91.9	89.2	84.3
0.7	92.8	84.1	80.1	93.1	90.7	86.1
0.8	93.9	85.2	81.2	95.3	91.7	87.1
0.9	93.0	84.5	80.4	93.4	91.0	86.3

表 4.10 展示了投票占比阈值 δ_v 对分割性能的影响。总体来看,随着 δ_v 从 0.6 增加到 0.8,模型性能逐步提升;但当其继续增大到 0.9 时,性能则出现轻微下降。该现象表明,投票占比阈值对伪标签可靠性与覆盖率之间的平衡具有直接影响。

当 δ_v 取值较低时,虽然能够保留更多候选伪标签像素,但其中也包含了更多时间一致性不足的位置,使得伪标签噪声更容易被引入训练过程;随着阈值提高,只有在历史窗口内得到更高比例支持的像素才会被保留下来,从而增强伪标签的稳定性与可信度。特别是当 $\delta_v = 0.8$ 时,三个数据集均达到最优结果,说明要求候选类别在大多数历史预测中保持一致,有助于显著降低偶发误判带来的影响。然而,若阈值继续提高到 0.9,三重可靠性筛选条件将变得过于严格,导致部分本身较为可靠但未达到极高一致性的像素被过滤,从而减少有效伪监督覆盖范围。综合来看, $\delta_v = 0.8$ 能够兼顾伪标签的纯净度与覆盖范围,因此被选为默认设置。

表 4.11 平均支持置信度阈值 δ_c 的参数敏感性分析结果

δ_c	mIoU(%)			mDice(%)		
	陶瓷	钛铝合金	碳钢	陶瓷	钛铝合金	碳钢
0.6	91.7	83.1	78.8	92.1	89.5	84.8
0.7	92.9	84.3	80.3	93.2	90.8	86.2
0.8	93.9	85.2	81.2	95.3	91.7	87.1
0.9	92.8	84.2	80.0	93.3	90.8	85.8

表 4.11 给出了平均支持置信度阈值 δ_c 的敏感性分析结果。可以发现，该参数的变化趋势与投票占比阈值 δ_v 基本一致，即在较低取值下模型性能逐步上升，在达到 0.8 时取得最优结果，而继续提高到 0.9 后略有下降。这说明除了历史投票一致性之外，支持候选类别的预测置信水平同样是衡量伪标签可靠性的关键因素。

当 δ_c 较低时，一部分虽然在历史上得到多数投票支持、但对应概率响应并不高的像素也会被纳入伪标签集合，这容易在纹理相似或边界模糊区域引入不够可靠的监督信号。随着 δ_c 提高，模型对支持候选类别的预测置信度提出了更严格要求，从而有助于进一步提升伪标签质量。尤其在三个数据集上， $\delta_c = 0.8$ 时均取得最优性能，表明在本章场景下，将平均支持置信度约束在较高水平，有助于增强伪监督的稳定性。然而，若阈值设置过高，则会使部分具有较好一致性但置信度略低的有效像素被排除，导致伪标签覆盖不足。因此，最终采用 $\delta_c = 0.8$ 作为默认设置。

表 4.12 边界阈值过滤参数 δ_b 的敏感性分析结果

δ_b	mIoU(%)			mDice(%)		
	陶瓷	钛铝合金	碳钢	陶瓷	钛铝合金	碳钢
0.5	93.4	84.6	80.4	93.9	91.0	86.4
0.6	93.9	85.2	81.2	95.3	91.7	87.1
0.7	93.1	84.4	80.2	93.4	91.0	86.0
0.8	91.8	82.9	78.6	92.3	89.4	84.6
0.9	90.5	81.8	77.5	91.9	88.8	83.4

表 4.12 展示了边界阈值过滤参数 δ_b 对模型性能的影响。可以看出，随着 δ_b 从 0.5 增大到 0.6，模型性能持续提升；当阈值继续增大到 0.7、0.8 和 0.9 时，性能则开始下降。这表明，在伪标签构建阶段适度执行边界阈值过滤，有助于减少边界噪声传播，但若边界阈值过滤过弱或过强，都会影响最终分割效果。

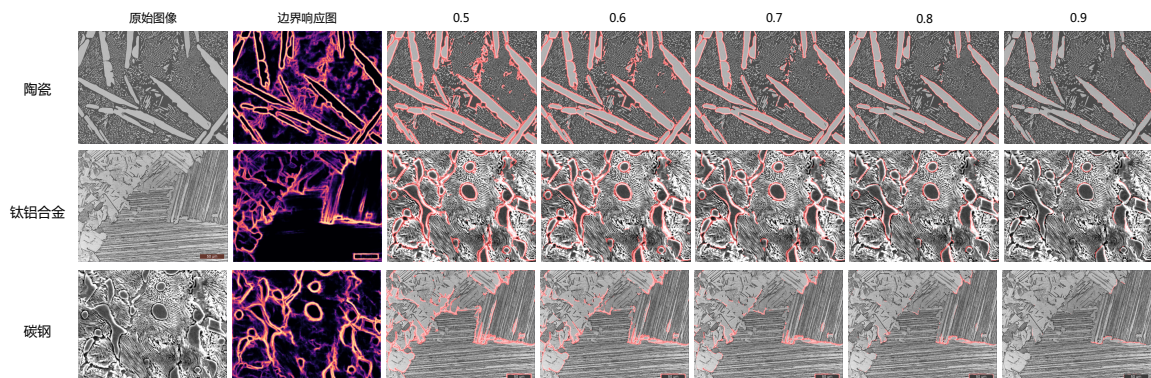
具体而言，当 δ_b 取值较小时，系统会将更多边界邻域像素视为高风险区域并剔除，虽然可以较充分地抑制边界噪声，但同时也会损失部分本可提供有效监督的近边界像素，导致伪标签覆盖范围不足；当 δ_b 适当提高到 0.6 时，模型在三个数据集上均取得最优结果，说明该设置能够较好地地区分应被抑制的高风险边界像素和仍可用于监督的相对稳定像素。而当阈值进一步增大时，更多边界响应较高的位置被保留下来，使复杂相界附近的不稳定伪标签更容易进入训练过程，进而对分割性能产生负面影响。因此，将 δ_b 设置为 0.6。

表 4.13 边界锐化阈值 τ_{sh} 的参数敏感性分析结果

τ_{sh}	mIoU(%)			mDice(%)		
	陶瓷	钛铝合金	碳钢	陶瓷	钛铝合金	碳钢
0.5	92.2	83.3	79.0	92.6	89.9	84.8
0.6	93.3	84.5	80.5	93.6	91.0	86.5
0.7	93.9	85.2	81.2	95.3	91.7	87.1
0.8	93.5	84.8	80.8	94.0	91.2	86.7
0.9	92.9	84.0	79.8	93.3	90.5	85.8

表 4.13 给出了边界锐化阈值 τ_{sh} 的敏感性分析结果。可以观察到，当 τ_{sh} 从 0.5 提高到 0.7 时，三个数据集上的性能逐步提升；当其进一步提高到 0.8 和 0.9 时，性能则略有回落。该结果说明，边界锐化损失作用区域的选取范围对模型边界刻画能力具有重要影响。

当 τ_{sh} 较低时，较大范围的像素会被视为高响应边界区域并参与锐化约束，其中不可避免地包含部分真实边界特征并不明显的区域，从而削弱了边界锐化项的针对性；随着阈值提高，参与锐化损失计算的像素逐渐集中到更接近真实相界的位置，使得该损失能够更有效地增强边界处的类别跃迁强度，因此性能稳步提升。特别是当 $\tau_{sh} = 0.7$ 时，模型在三个数据集上均达到最优结果，说明该设置能够较准确地聚焦于高响应边界区域。然而，若阈值继续提高，则仅有极少数边界强响应位置被保留下来，导致边界锐化约束的作用范围过窄，无法充分覆盖真实相界附近的有效过渡区域，最终使性能略有下降。因此，本文最终将 τ_{sh} 设为 0.7。

图 4.10 不同边界锐化阈值 τ_{sh} 下的边界响应可视化结果

此外，图 4.10 从可视化角度进一步说明了不同阈值下边界锐化作用区域的变化规律，与表 4.13 的定量结论保持一致。每一行分别对应陶瓷、钛铝合金和碳钢数据

集，每一列从左到右依次为原始图像、边界响应图以及在 $\tau_{sh} = 0.5, 0.6, 0.7, 0.8, 0.9$ 时的边界区域叠加显示结果。为保证不同阈值之间结果的可比性，实验在可视化过程中均基于同一原始图像和同一边界响应图，并采用一致的叠加显示方式，仅改变阈值大小。可以看出，随着 τ_{sh} 的增大，被选中参与边界锐化损失计算的区域逐渐由较大范围收缩至边界响应最强的位置：当阈值较低时，锐化约束覆盖范围较大，容易包含部分边界特征不够显著的区域；当阈值较高时，作用区域又过于集中，难以充分覆盖真实相界附近的有效过渡区域。结合表 4.13 的定量结果可以发现， $\tau_{sh} = 0.7$ 时能够较好地聚焦主要边界区域，并在三个数据集上取得最优分割性能。

4.3 本章小结

作为复杂纹理图像中的典型类别之一，材料显微图像在小样本涂鸦监督分割中更突出地表现为训练稳定性不足、伪标签可靠性有限以及复杂相界附近边界噪声易传播等问题。围绕这些材料显微图像特有的任务需求，本章提出了一种基于多历史投票与边界感知约束的涂鸦监督材料显微图像分割方法 **MVBC**。该方法以采用预训练 VGG16 编码器的 U-Net++ 作为基础分割网络，并结合“预训练初始化 + 浅层冻结 + 深层微调”的迁移学习策略，以提升小样本条件下的训练稳定性；在此基础上，构建按样本维护的历史预测队列，并通过多历史投票、投票占比和平均支持置信度估计，从滑动历史窗口中筛选预测相对稳定且置信水平较高的伪标签，以增强伪标签可靠性。需要指出的是，该历史投票机制并不依赖独立同分布假设，而是利用训练过程中预测结果的经验稳定性来降低单次预测波动的影响。结合边界响应图引导的边界阈值过滤以及边界平滑、边界锐化联合约束，**MVBC** 能够抑制复杂相界附近的边界噪声传播并增强边界过渡清晰度。

考虑到材料显微图像中的相界通常具有细长、弯曲、纹理相似和局部对比度不均等特点，**MVBC** 引入了边界感知约束。一方面，利用 PiDiNet 生成的边界响应图在伪标签构建阶段执行边界阈值过滤，抑制高风险边界区域的伪监督传播；另一方面，在网络优化阶段联合引入边界平滑损失与边界锐化损失，在增强区域内部一致性的同时强化边界过渡清晰度。由此，**MVBC** 形成了一个由部分监督、可靠伪监督与边界感知约束共同驱动的统一训练框架，使模型能够在极少量涂鸦标注条件下同时提升区域识别稳定性与复杂相界定位精度。

在实验验证方面,本章在陶瓷、钛铝合金和碳钢三类典型材料显微图像数据集上进行了系统测试。结果表明, **MVBC** 在三个数据集上均取得了最优分割性能, **mIoU** 分别达到 93.9%、85.2% 和 81.2%, **mDice** 分别达到 95.3%、91.7% 和 87.1%。特别是在钛铝合金数据集上, **MVBC** 优于所有参与比较的方法,说明所提出方法能够在极低标注成本下获得较强的组织识别能力。结合定量结果与可视化分析可以看出,多历史投票机制有效提升了伪标签的稳定性,边界感知约束则显著增强了模型对复杂相界和细粒组织结构的刻画能力。

MVBC 在材料显微图像小样本涂鸦监督分割任务中实现了伪标签可靠性与边界表达能力的协同提升,为低标注成本条件下材料显微图像的高精度分割提供了一条可行且有效的思路。与此同时, **MVBC** 仍需为每个样本维护历史预测信息,并额外引入边界先验与边界损失约束,因此其在训练复杂度和参数调节方面相较基础弱监督模型更为复杂。但总体而言, **MVBC** 在标注成本、训练可行性与分割性能之间取得了较好的平衡,为后续面向更复杂材料体系和更大尺度数据场景的弱监督分割研究奠定了基础。

第五章 总结与展望

5.1 结论

本文以复杂纹理图像的涂鸦监督分割为研究对象，选取医学影像与材料显微图像两类典型场景作为具体研究对象，围绕稀疏监督条件下的可靠伪监督构建与复杂边界不确定性抑制问题展开研究。医学影像与材料显微图像虽然在成像机制、目标结构和数据规模方面存在差异，但在涂鸦监督条件下均面临监督信息稀疏、未标注区域伪标签可靠性不足以及复杂边界区域误差易累积等共性挑战。针对上述问题，本文分别从空间域一致性和时间域一致性两个角度进行了方法实例化：面向医学影像场景提出 **FW2SS** 方法，面向材料显微图像小样本场景提出 **MVBC** 方法，并通过系统实验验证了所提方法在两类典型复杂纹理图像分割任务中的有效性。结合全文研究内容与实验结果，可以得到以下结论：

(1) 针对医学影像涂鸦监督分割中形状信息学习不足、局部纹理刻画与全局结构建模难以协同以及复杂边界区域判别困难的问题，本文提出了基于一致性约束与对比学习的 **FW2SS** 方法。该方法构建了由 **CNN** 与 **Transformer** 组成的异构双分支网络，用于增强局部纹理建模与全局结构建模之间的互补性；同时，通过涂鸦一致性匹配模块联合网络扰动与输入扰动生成更稳定的伪监督信号，从而提升稀疏监督条件下的形状信息学习能力与伪监督稳定性；借助像素级原型对比校准模块增强复杂边界区域的特征区分能力。在本文的实验设置下，**FW2SS** 在 **ACDC** 和 **MSCMRseg** 两个心脏 **MRI** 数据集上的平均 **Dice** 分别达到 90.0% 和 88.3%，在涂鸦监督设置下优于多种代表性弱监督分割方法。同时，在相同训练病例数量和统一实验协议下，该方法的平均 **Dice** 高于若干常用全监督 **U-Net** 系列基线方法，并接近本文纳入比较的强全监督方法 **nnUNet**。上述结果表明，在医学影像涂鸦监督场景中，围绕异构结构互补、涂鸦一致性匹配模块与像素级原型对比校准模块开展联合设计，能够在降低标注成本的同时有效缓解弱监督信息不足带来的性能损失。

(2) 针对材料显微图像小样本涂鸦监督分割中训练稳定性不足、伪标签可靠性有限以及复杂相界附近边界噪声易传播的问题，提出了基于多历史投票与边界感知约束的 **MVBC** 方法。该方法通过“预训练初始化、浅层冻结、深层微调”的迁移学习策

略提升小样本条件下的训练稳定性；通过按样本维护的历史预测队列与多历史投票机制，从滑动历史窗口中刻画同一样本预测结果的经验稳定性，并据此筛选相对可靠的伪标签；同时结合边界响应图以及边界平滑、边界锐化联合约束，抑制复杂相界附近高风险边界区域的噪声传播，并增强复杂相界的刻画能力。在仅使用极少量训练图像和涂鸦标注的条件下，MVBC 在陶瓷、钛铝合金和碳钢数据集上的 mIoU 分别达到 93.9%、85.2% 和 81.2%，mDice 分别达到 95.3%、91.7% 和 87.1%。实验结果表明，针对材料显微图像中小样本与复杂边界并存的特点，通过引入基于经验稳定性的多历史伪监督筛选与边界感知约束，能够有效提升弱监督分割的稳定性与精度。

(3) 实验结果表明，基于涂鸦标注的弱监督分割在医学影像与材料显微图像这两类典型复杂纹理场景中均具有较好的应用潜力。更重要的是，本文的实验结果从两个不同场景验证了可靠伪监督构建与边界噪声抑制对于复杂纹理图像涂鸦监督分割的重要性：在医学影像场景中，空间域一致性约束有助于从稀疏涂鸦标注中学习更完整的目标结构；在材料显微图像小样本场景中，时间域历史预测一致性与边界感知约束有助于提升伪标签筛选的稳定性并减少复杂相界附近的误差传播。因此，FW2SS 和 MVBC 虽然分别针对不同应用场景设计，但二者均围绕稀疏监督条件下的伪监督可靠性提升和复杂边界刻画展开，为后续面向更多复杂纹理图像类别开展低标注分割研究提供了有益参考。

总体来看，本文的主要贡献体现在以下三个方面：其一，以复杂纹理图像涂鸦监督分割为统一研究对象，将医学影像与材料显微图像两类典型场景中的共性困难概括为监督信息稀疏、伪标签可靠性不足和复杂边界不确定性突出；其二，围绕可靠伪监督构建与边界噪声抑制这一共同目标，分别从空间域一致性和时间域一致性两个角度提出 FW2SS 和 MVBC 两种方法实例，其中 FW2SS 侧重于医学影像场景中的结构一致性学习与边界特征判别增强，MVBC 侧重于材料显微图像小样本场景中的历史预测稳定性筛选与边界感知约束；其三，通过系统实验验证了在显著降低标注成本的前提下，两类典型复杂纹理场景仍能够获得较高质量的分割结果。上述工作表明，尽管不同复杂纹理图像场景具有各自的数据特征和任务难点，但围绕伪监督可靠性和边界不确定性开展方法设计，能够为低标注成本条件下的复杂纹理图像分割提供具有共性的问题导向和方法参考。

5.2 工作展望

结合本文的范围与实验结果，当前工作在方法统一性、泛化能力和实际应用方面仍有进一步研究空间。后续工作可从以下几个方面展开：

(1) **面向更多典型复杂纹理场景的统一弱监督分割框架**。本文针对医学影像分割与材料显微图像分割两类典型任务分别设计了具有较强针对性的模型结构与监督机制，并在各自场景中取得了良好效果。但从更一般的复杂纹理图像问题出发，当前工作更多是在两个代表性场景上验证了相关思路的有效性，两类方法仍然具有较明显的任务定制特征。后续可进一步探索更加统一的弱监督分割框架，将一致性学习、时序伪标签更新、结构先验建模和边界约束等思想以模块化方式进行融合，提升方法在不同模态、不同纹理类型和不同任务场景之间的可迁移能力。

(2) **伪标签可靠性与不确定性建模的进一步提升**。无论是医学影像中的伪标签生成，还是材料显微图像中的历史投票筛选，伪标签质量始终是影响弱监督分割性能的关键因素。需要指出的是，MVBC 中的历史预测来自同一模型在相邻训练阶段的参数状态，其间存在一定相关性，因此多历史投票更适合作为滑动时间窗口下的经验稳定性筛选机制，而不能等同于严格统计意义上的独立投票过程。后续研究可进一步引入不确定性估计、置信度校准、自步学习以及多教师协同机制，对伪标签进行动态评估与分层利用；同时结合样本难度、自适应阈值和类别分布信息，对不同区域采用差异化的伪监督策略，从而进一步降低错误伪标签对模型优化过程的负面影响。

(3) **交互式标注与可信应用落地**。涂鸦监督具有天然的人机交互属性，未来可进一步将主动学习与交互式分割机制引入当前框架，通过模型自动发现高不确定性区域并引导专家进行增量补标，形成闭环训练过程。同时，面向实际应用，还需要综合考虑模型轻量化、训练与推理效率、资源消耗、结果可解释性以及错误风险提示等问题。后续可从轻量化网络设计、显存友好的训练机制、快速推理策略等方面提升模型的工程可用性与可信度，使弱监督分割技术更好地服务于临床辅助分析和材料检测流程。

复杂纹理图像的涂鸦监督分割仍然具有较大的研究空间，后续工作的重点不应仅停留在单一数据集指标的进一步提升，更在于构建兼具准确性、稳健性、可迁移性与可部署性的弱监督分割方法。围绕这一目标，如何进一步统一模型设计思路、提

升伪标签可靠性，并推动技术在真实场景中的稳定应用，将是值得持续关注的重要方向。

参考文献

- [1] RAYED M E, ISLAM S S, NIHA S I, et al. Deep learning for medical image segmentation: State-of-the-art advancements and challenges[J]. Informatics in medicine unlocked, 2024, 47: 101504.
- [2] HOLME A, COHN R, GAO N, et al. Overview: Computer vision and machine learning for microstructural characterization and analysis[J]. Metallurgical and Materials Transactions A, 2020, 51(12): 5985-5999.
- [3] YAO W, BAI J, LIAO W, et al. From cnn to transformer: A review of medical image segmentation models[J]. Journal of Imaging Informatics in Medicine, 2024, 37(4): 1529-1547.
- [4] 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.
- [5] ZHOU Z, RAHMAN SIDDIQUEE M M, TAJBAKHSN N, et al. Unet++: A nested u-net architecture for medical image segmentation[C]//International workshop on deep learning in medical image analysis. Springer, 2018: 3-11.
- [6] KALININ S V, SUMPTER B G, ARCHIBALD R K. Big-deep-smart data in imaging for guiding materials design[J]. Nature materials, 2015, 14(10): 973-980.
- [7] PENG X L, FATHIDOOST M, LIN B, et al. What can machine learning help with microstructure-informed materials modeling and design?[J/OL]. MRS Bulletin, 2025, 50: 61-79. DOI: 10.1557/s43577-024-00797-4.
- [8] XU Y, ZHANG Y, ZHANG M, et al. Quantitative analysis of metallographic image using attention-aware deep neural networks[J/OL]. Sensors, 2021, 21(1): 43. DOI: 10.3390/s21010043.

- [9] LIN D, DAI J, JIA J, et al. Scribblesup: Scribble-supervised convolutional networks for semantic segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 3159-3167.
- [10] LEE H, JEONG W K. Scribble2label: Scribble-supervised cell segmentation via self-generating pseudo-labels with consistency[C]//Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part I 23. Springer, 2020: 14-23.
- [11] VALVANO G, LEO A, TSAFTARIS S A. Learning to segment from scribbles using multi-scale adversarial attention gates[J]. IEEE Transactions on Medical Imaging, 2021, 40(8): 1990-2001.
- [12] ZHANG K, ZHUANG X. Cyclemix: A holistic strategy for medical image segmentation from scribble supervision[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022: 11656-11665.
- [13] JIAO R, ZHANG Y, DING L, et al. Learning with limited annotations: a survey on deep semi-supervised learning for medical image segmentation[J]. Computers in Biology and Medicine, 2024, 169: 107840.
- [14] TAJBAKSH N, JEYASEELAN L, LI Q, et al. Embracing imperfect datasets: A review of deep learning solutions for medical image segmentation[J]. Medical image analysis, 2020, 63: 101693.
- [15] DECOST B L, JAIN H, ROLLETT A D, et al. Computer vision and machine learning for autonomous characterization of am powder feedstocks[J]. Jom, 2017, 69(3): 456-465.
- [16] NA J, KIM S J, KIM H, et al. A unified microstructure segmentation approach via human-in-the-loop machine learning[J]. Acta Materialia, 2023, 255: 119086.
- [17] ALRFOU K, ZHAO T, KORDIJAZI A. Deep learning methods for microstructural image analysis: the state-of-the-art and future perspectives[J]. Integrating Materials and Manufacturing Innovation, 2024, 13(3): 703-731.
- [18] XIA Q, ZHENG H, ZOU H, et al. A comprehensive review of deep learning for medical image segmentation[J]. Neurocomputing, 2025, 613: 128740.

- [19] HARALICK R M, SHANMUGAM K, DINSTEN I. Textural features for image classification[J]. IEEE Transactions on Systems, Man, and Cybernetics, 1973, SMC-3(6): 610-621.
- [20] OJALA T, PIETIKAINEN M, HARWOOD D. Performance evaluation of texture measures with classification based on kullback discrimination of distributions[C]// Proceedings of 12th International Conference on Pattern Recognition: Vol. 1. IEEE, 1994: 582-585.
- [21] GABOR D. Theory of communication. part 1: The analysis of information[J]. Journal of the Institution of Electrical Engineers-part III: radio and communication engineering, 1946, 93(26): 429-441.
- [22] KASS M, WITKIN A, TERZOPOULOS D. Snakes: Active contour models[J]. International Journal of Computer Vision, 1988, 1(4): 321-331.
- [23] CHAN T F, VESE L A. Active contours without edges[J]. IEEE Transactions on Image Processing, 2001, 10(2): 266-277.
- [24] CORTES C, VAPNIK V. Support-vector networks[J]. Machine Learning, 1995, 20(3): 273-297.
- [25] BREIMAN L. Random forests[J]. Machine Learning, 2001, 45(1): 5-32.
- [26] BEZDEK J C, EHRLICH R, FULL W. Fcm: The fuzzy c-means clustering algorithm [J]. Computers & Geosciences, 1984, 10(2-3): 191-203.
- [27] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet classification with deep convolutional neural networks[C]//Advances in Neural Information Processing Systems: Vol. 25. 2012: 1097-1105.
- [28] LONG J, SELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 3431-3440.
- [29] ÇİÇEK Ö, ABDULKADIR A, LIENKAMP S S, et al. 3d u-net: learning dense volumetric segmentation from sparse annotation[C]//International conference on medical image computing and computer-assisted intervention. Springer, 2016: 424-432.

- [30] ISENSEE F, JAEGER P F, KOHL S A, et al. nnu-net: a self-configuring method for deep learning-based biomedical image segmentation[J]. Nature methods, 2021, 18(2): 203-211.
- [31] OKTAY O, SCHLEMPER J, FOLGOC L L, et al. Attention u-net: Learning where to look for the pancreas[C/OL]//Medical Imaging with Deep Learning. 2018. <https://openreview.net/forum?id=Skft7cijM>. DOI: 10.48550/arXiv.1804.03999.
- [32] CHEN L C, PAPANDREOU G, KOKKINOS I, et al. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs [J]. IEEE transactions on pattern analysis and machine intelligence, 2018, 40(4): 834-848.
- [33] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16x16 words: Transformers for image recognition at scale[C]//9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021. OpenReview.net, 2021.
- [34] CHEN J, MEI J, LI X, et al. TransUNet: Rethinking the U-Net architecture design for medical image segmentation through the lens of transformers[J/OL]. Medical Image Analysis, 2024, 97: 103280. <https://www.sciencedirect.com/science/article/pii/S1361841524002056>. DOI: 10.1016/j.media.2024.103280.
- [35] CAO H, WANG Y, CHEN J, et al. Swin-unet: Unet-like pure transformer for medical image segmentation[C]//European conference on computer vision. Springer, 2022: 205-218.
- [36] HATAMIZADEH A, TANG Y, NATH V, et al. Unetr: Transformers for 3d medical image segmentation[C]//Proceedings of the IEEE/CVF winter conference on applications of computer vision. 2022: 574-584.
- [37] PAPANDREOU G, CHEN L C, MURPHY K P, et al. Weakly-and semi-supervised learning of a deep convolutional network for semantic image segmentation[C]//Proceedings of the IEEE international conference on computer vision. 2015: 1742-1750.

- [38] DAI J, HE K, SUN J. Boxsup: Exploiting bounding boxes to supervise convolutional networks for semantic segmentation[C]//Proceedings of the IEEE international conference on computer vision. 2015: 1635-1643.
- [39] BEARMAN A, RUSSAKOVSKY O, FERRARI V, et al. What's the point: Semantic segmentation with point supervision[C]//European conference on computer vision. Springer, 2016: 549-565.
- [40] TANG M, DJELOUAH A, PERAZZI F, et al. Normalized cut loss for weakly-supervised cnn segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 1818-1827.
- [41] WANG B, QI G, TANG S, et al. Boundary perception guidance: A scribble-supervised semantic segmentation approach[C/OL]//Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence. 2019: 3663-3669. <https://www.ijcai.org/proceedings/2019/508>. DOI: 10.24963/ijcai.2019/508.
- [42] CHEN Z, SUN Q. Weakly-supervised semantic segmentation with image-level labels: From traditional models to foundation models[J/OL]. ACM Computing Surveys, 2025, 57(5): Article 111. DOI: 10.1145/3707447.
- [43] BOYKOV Y, JOLLY M P. Interactive graph cuts for optimal boundary & region segmentation of objects in n-d images[C]//Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001: Vol. 1. IEEE, 2001: 105-112.
- [44] ROTHER C, KOLMOGOROV V, BLAKE A. "grabcut": Interactive foreground extraction using iterated graph cuts[J]. ACM transactions on graphics, 2004, 23(3): 309-314.
- [45] GRADY L. Multilabel random walker image segmentation using prior models[C]//2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05): Vol. 1. IEEE, 2005: 763-770.
- [46] TANG M, PERAZZI F, DJELOUAH A, et al. On regularized losses for weakly-supervised cnn segmentation[C]//Proceedings of the European conference on computer vision (ECCV). 2018: 507-522.

- [47] ZHANG K, ZHUANG X. ShapePU: A new PU learning framework regularized by global consistency for scribble supervised cardiac segmentation[C/OL]//International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, 2022: 162-172. <https://conferences.miccai.org/2022/papers/466-Paper0537.html>. DOI: 10.1007/978-3-031-16452-1_16.
- [48] LUO X, HU M, LIAO W, et al. Scribble-supervised medical image segmentation via dual-branch network and dynamically mixed pseudo labels supervision[C]//International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, 2022: 528-538.
- [49] LI Z, ZHENG Y, LUO X, et al. Scribblevc: Scribble-supervised medical image segmentation with vision-class embedding[C]//Proceedings of the 31st ACM International Conference on Multimedia. 2023: 3384-3393.
- [50] LI Z, ZHENG Y, SHAN D, et al. Scribformer: Transformer makes cnn work better for scribble-based medical image segmentation[J/OL]. IEEE Transactions on Medical Imaging, 2024, 43(6): 2254-2265. DOI: 10.1109/TMI.2024.3363190.
- [51] HAN Y, LI R, WANG B, et al. A pseudo-labeling based weakly supervised segmentation method for few-shot texture images[J]. Expert Systems with Applications, 2024, 238: 122110.
- [52] ZHOU M, REN Z, TAN Q, et al. Visual mamba-cnn for scribble-based segmentation in weakly supervised learning for photoacoustic tomography[J/OL]. Expert Systems with Applications, 2026, 298: 129749. DOI: 10.1016/j.eswa.2025.129749.
- [53] DU Y, FU Z, LIU Q, et al. Weakly supervised semantic segmentation by pixel-to-prototype contrast[C/OL]//IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2022, New Orleans, LA, USA, June 18-24, 2022. IEEE, 2022: 4320-4329. DOI: 10.1109/CVPR52688.2022.00428.
- [54] CHAN G, ZHANG P, DONG H, et al. Scribble-supervised semantic segmentation with prototype-based feature augmentation[C/OL]//Forty-first International Conference on Machine Learning: Vol. 235. PMLR, 2024: 6155-6169. <https://proceedings.mlr.press/v235/chan24b.html>.

- [55] HAN Y, LI R, YANG S, et al. Center-environment feature models for materials image segmentation based on machine learning[J]. Scientific Reports, 2022, 12(1): 12960.
- [56] XIE E, WANG W, YU Z, et al. Segformer: Simple and efficient design for semantic segmentation with transformers[C]//RANZATO M, BEYGELZIMER A, DAUPHIN Y N, et al. Advances in Neural Information Processing Systems 34: Annual Conference on Neural Information Processing Systems 2021, NeurIPS 2021, December 6-14, 2021, virtual. 2021: 12077-12090.
- [57] XIAO H, LI L, LIU Q, et al. Transformers in medical image segmentation: A review [J]. Biomedical Signal Processing and Control, 2023, 84: 104791.
- [58] RAN L, LI Y, LIANG G, et al. Pseudo labeling methods for semi-supervised semantic segmentation: A review and future perspectives[J/OL]. IEEE Transactions on Circuits and Systems for Video Technology, 2025, 35(4): 3054-3080. DOI: 10.1109/TCSVT.2024.3508768.
- [59] ROTH H R, YANG D, XU Z, et al. Going to extremes: weakly supervised medical image segmentation[J]. Machine Learning and Knowledge Extraction, 2021, 3(2): 507-524.
- [60] KE T, HWANG J, YU S X. Universal weakly supervised segmentation by pixel-to-segment contrastive learning[C]//9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021. OpenReview.net, 2021.
- [61] LEE D H, et al. Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks[C]//Workshop on challenges in representation learning, ICML: Vol. 3. Atlanta, 2013: 896.
- [62] SOHN K, BERTHELOT D, CARLINI N, et al. Fixmatch: Simplifying semi-supervised learning with consistency and confidence[C/OL]//Advances in Neural Information Processing Systems: Vol. 33. 2020: 596-608. <https://papers.nips.cc/paper/2020/hash/06964dce9addb1c5cb5d6e3d9838f733-Abstract.html>.
- [63] TARVAINEN A, VALPOLA H. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results[C]//Advances in Neural Information Processing Systems: Vol. 30. 2017: 1195-1204.

- [64] LUO X, HU M, SONG T, et al. Semi-supervised medical image segmentation via cross teaching between cnn and transformer[C]//International conference on medical imaging with deep learning. PMLR, 2022: 820-833.
- [65] CHEN X, YUAN Y, ZENG G, et al. Semi-supervised semantic segmentation with cross pseudo supervision[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021: 2613-2622.
- [66] WANG Y, WANG H, SHEN Y, et al. Semi-supervised semantic segmentation using unreliable pseudo-labels[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022: 4248-4257.
- [67] LAINE S, AILA T. Temporal ensembling for semi-supervised learning[C]//5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings. OpenReview.net, 2017.
- [68] CUBUK E D, ZOPH B, SHLENS J, et al. Randaugment: Practical automated data augmentation with a reduced search space[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops. 2020: 702-703.
- [69] 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.
- [70] PENG Z, HUANG W, GU S, et al. Conformer: Local features coupling global representations for visual recognition[C]//Proceedings of the IEEE/CVF international conference on computer vision. 2021: 367-376.
- [71] BERNARD O, LALANDE A, ZOTTI C, et al. Deep learning techniques for automatic mri cardiac multi-structures segmentation and diagnosis: is the problem solved?[J]. IEEE transactions on medical imaging, 2018, 37(11): 2514-2525.
- [72] 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.

- [73] ZHUANG X. Multivariate mixture model for myocardial segmentation combining multi-source images[J/OL]. IEEE transactions on pattern analysis and machine intelligence, 2019, 41(12): 2933-2946. DOI: 10.1109/TPAMI.2018.2869576.
- [74] ZHENG S, JAYASUMANA S, ROMERA-PAREDES B, et al. Conditional random fields as recurrent neural networks[C]//Proceedings of the IEEE international conference on computer vision. 2015: 1529-1537.
- [75] BAUMGARTNER C F, KOCH L M, POLLEFEYS M, et al. An exploration of 2d and 3d deep learning techniques for cardiac mr image segmentation[C]//Statistical Atlases and Computational Models of the Heart. ACDC and MMWHS Challenges: 8th International Workshop, STACOM 2017, Held in Conjunction with MICCAI 2017, Quebec City, Canada, September 10-14, 2017, Revised Selected Papers 8. Springer, 2018: 111-119.
- [76] KIM J, CHOO W, JEONG H, et al. Co-mixup: Saliency guided joint mixup with supermodular diversity[C]//9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021. OpenReview.net, 2021.
- [77] YUN S, HAN D, OH S J, et al. Cutmix: Regularization strategy to train strong classifiers with localizable features[C]//Proceedings of the IEEE/CVF international conference on computer vision. 2019: 6023-6032.
- [78] DEVRIES T, TAYLOR G W. Improved regularization of convolutional neural networks with cutout[A/OL]. 2017. arXiv: 1708.04552. <https://arxiv.org/abs/1708.04552>.
- [79] KIM J H, CHOO W, SONG H O. Puzzle mix: Exploiting saliency and local statistics for optimal mixup[C]//International conference on machine learning. PMLR, 2020: 5275-5285.
- [80] ZHANG H, CISSÉ M, DAUPHIN Y N, et al. mixup: Beyond empirical risk minimization[C]//6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings. OpenReview.net, 2018.

- [81] NGUYEN-TAT T B, VO H A, DANG P S. Qmaxvit-unet+: A query-based maxvit-unet with edge enhancement for scribble-supervised segmentation of medical images [J/OL]. Computers in Biology and Medicine, 2025, 187: 109762. DOI: 10.1016/j.compbimed.2025.109762.
- [82] LIU J, TAN S Y, YANG X, et al. EFFDNet: A scribble-supervised medical image segmentation method with enhanced foreground feature discrimination[C/OL]//Lecture Notes in Computer Science: Vol. 15975 Medical Image Computing and Computer Assisted Intervention –MICCAI 2025: 28th International Conference, Daejeon, South Korea, September 23–27, 2025, Proceedings, Part XVI. Berlin, Heidelberg: Springer, 2025: 194-204. DOI: 10.1007/978-3-032-05325-1_19.
- [83] ZHANG L, LI W, BI K, et al. FDDSeg: Unleashing the power of scribble annotation for cardiac MRI images through feature decomposition distillation[J/OL]. IEEE Journal of Biomedical and Health Informatics, 2025, 29(1): 285-296. DOI: 10.1109/JBHI.2024.3404884.
- [84] LARRAZABAL A J, MARTÍNEZ C, GLOCKER B, et al. Post-dae: anatomically plausible segmentation via post-processing with denoising autoencoders[J]. IEEE transactions on medical imaging, 2020, 39(12): 3813-3820.
- [85] ZHANG P, ZHONG Y, LI X. ACCL: adversarial constrained-cnn loss for weakly supervised medical image segmentation[A/OL]. 2020. arXiv: 2005.00328. <https://arxiv.org/abs/2005.00328>.
- [86] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[C]//International Conference on Learning Representations. 2015.
- [87] SU Z, LIU W, YU Z, et al. Pixel difference networks for efficient edge detection[C]//Proceedings of the IEEE/CVF international conference on computer vision. 2021: 5117-5127.
- [88] HUANG H, LIN L, TONG R, et al. Unet 3+: A full-scale connected unet for medical image segmentation[C]//ICASSP 2020-2020 IEEE international conference on acoustics, speech and signal processing (ICASSP). Ieee, 2020: 1055-1059.

- [89] GU Z, CHENG J, FU H, et al. Ce-net: Context encoder network for 2d medical image segmentation[J]. IEEE transactions on medical imaging, 2019, 38(10): 2281-2292.
- [90] LIANG Z, WANG T, ZHANG X, et al. Tree energy loss: Towards sparsely annotated semantic segmentation[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022: 16907-16916.
- [91] LIU Y, LIN L, WONG K K, et al. Procnets: Progressive prototype calibration and noise suppression for weakly-supervised medical image segmentation[J/OL]. IEEE Journal of Biomedical and Health Informatics, 2025, 29(4): 2845-2858. DOI: 10.1109/JBHI.2024.3522958.

攻读硕士学位期间取得的研究成果

一、主要作者研究成果

[1] Zhang, R., Li, Z., Chen, Q., Sun, Y., Han, Y*. Scribble consistency match and pixel-level prototype contrastive calibration for weakly supervised medical segmentation. Neurocomputing, 2026, 681: 133258. <https://doi.org/10.1016/j.neucom.2026.133258> (导师一作, 本人二作, 已发表, SCI 二区)

二、软件著作权

[1] 基于涂鸦标注的交互式材料图像分割平台 V1.0, 开发人: 张瑞、李子铭、韩越兴, 登记号: 2024SR1573953, 登记日期: 2024 年 10 月 21 日, 申请人: 上海大学。(导师一作, 本人二作)

[2] 基于划线标注的交互式材料图像分割平台 V1.0, 开发人: 张瑞、李子铭、韩越兴, 登记号: 2025SR2154506, 登记日期: 2025 年 11 月 5 日, 申请人: 上海大学。(导师一作, 本人二作)

致 谢

时光荏苒，三载寒暑转瞬即逝。回首硕士阶段的学习与科研生活，从初入校园时的懵懂青涩，到如今论文付梓时的沉静笃定，这一路走来，有探索中的困惑，也有收获时的欣喜。值此硕士论文完成之际，我的内心充满感激与敬畏。谨以此文，向在我求学道路上给予我关心、帮助与支持的老师、同学、家人致以最诚挚的谢意。

首先，我要向我的导师张瑞老师致以最衷心、最诚挚的感谢。硕士三年，张老师始终在学业与科研道路上给予我悉心指导与耐心帮助。从课题的选定与研究思路的梳理，到实验方案的设计、论文内容的修改完善，张老师都倾注了大量心血。她不仅在专业知识上给予我深刻启发，更在科研态度、治学精神和为人处世方面给予我深远影响。师恩如山，铭记于心，在此谨向张老师致以最崇高的敬意和最由衷的感谢。

同时，我还要特别感谢韩越兴老师、陈侨川老师以及孙妍老师。在三年的科研工作中，三位老师在课题开题、实验设计、理论分析以及论文撰写等多个方面给予了我许多宝贵的指导和帮助。几位老师的言传身教，不仅让我在学术能力上获得提升，也让我更加深刻地理解了科研精神的真正内涵。在此，向各位老师致以由衷的感谢和深深的敬意。

此外，我要感谢实验室的同门在这三年中的陪伴与支持。实验室不仅是我完成课题研究的地方，更是我收获友谊、见证成长的重要场所。感谢大家在科研讨论中给予我的启发与帮助，在实验推进中给予我的协助与鼓励，在日常学习和生活中给予我的关心与包容，正是因为有你们的陪伴，这段求学时光才更加充实而温暖。

我尤其要感谢我的家人，感谢父母多年来无条件的理解、支持与关爱。求学之路并非总是一帆风顺，每当我面对压力与困难时，家人始终是我坚实的后盾。

最后，谨向百忙之中抽空审阅本论文、参与论文评阅和答辩的各位专家致以最崇高的敬意和最诚挚的感谢。感谢各位老师对本论文提出的宝贵意见和建议，使我得以进一步审视自身研究中的不足，并不断完善和提升。

三年的研究生生活即将画上句号，但这段时光所给予我的成长、磨砺与收获，将成为我人生中宝贵的财富。未来无论身处何方，我都将铭记师长的教诲，珍惜同窗的情谊，不忘家人的支持，不负青春，不负所学。