分类号	xxxx	学校代码	xxxx
UDC	xxxx	密　　级	公开

硕士学位论文

基于 xxxxxx 分析

学位申请人姓名　xxxxxx

学位申请人学号　xxxxxxx

专业（领域）名称　xxxxxxx

学　位　类　别　xxxxxx

学院（部、研究院）　xxxxx

导　师　姓　名　xxxxx

二〇二四年五月

XXX 大学
学位论文原创性声明

本人郑重声明：所呈交的学位论文基于 xxxx 与 xxxx 分析是本人在导师的指导下，独立进行研究工作所取得的成果。除文中已经注明引用的内容外，本论文不含任何其他个人或集体已经发表或撰写的作品或成果。对本文的研究做出重要贡献的个人和集体，均已在文中以明确方式标明。本声明的法律后果由本人承担。

论文作者签名：　　日期：　　年　　月　　日

XXX 大学
学位论文使用授权说明

本学位论文作者完全了解深圳大学关于收集、保存、使用学位论文的规定，即：研究生在校攻读学位期间论文工作的知识产权单位属深圳大学。学校有权保留学位论文并向国家主管部门或其他机构送交论文的电子版和纸质版，允许论文被查阅和借阅。本人授权深圳大学可以将学位论文的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存、汇编学位论文。

（涉密学位论文在解密后适用本授权书）

论文作者签名：　　导师签名：

日期：　　年　　月　　日　　日期：　　年　　月　　日

摘 要

光学相干断层扫描（OCT）图像语义分割是激光焊接熔池检测等工业视觉任务中的关键环节。受散斑噪声、边界模糊以及目标与背景对比不足等因素影响，基于卷积的分割模型在复杂工况下易出现全局语义理解不足与边界刻画不精的问题，从而限制分割结果的稳定性与工程可用性。针对上述问题，本文提出一种基于改进 DeepLabV3+ 的 OCT 图像语义分割方法。

本文方法以 DeepLabV3+ 为基础框架，在多尺度上下文聚合的基础上引入全局与局部两类注意力建模机制。具体而言，在 ASPP 输出后嵌入 TR（Transformer Routing）模块，通过路由式注意力在可控计算开销下增强长程依赖建模能力，以提升对目标整体结构与上下文关系的表征；在解码器高低层特征融合后引入 SAE（Spatial Attention Enhancement）模块，通过空间与通道的联合增强强化关键区域响应与边界细节表达，从而改善细长结构与模糊边界处的分割质量。此外，采用交叉熵损失与 Dice 损失的组合以缓解类别不平衡带来的训练偏置。

在包含 912 张训练图像与 228 张测试图像的 OCT 数据集上进行实验验证。结果显示，所提方法的 mIoU 达到 0.911，较基线 DeepLabV3+ 提升 7.1%。上述结果表明，TR 与 SAE 的联合引入能够在统一的编码器—解码器框架内兼顾全局语义建模与局部细节增强，为 OCT 图像的高精度语义分割提供一种有效实现途径。

关键词： 语义分割；DeepLabV3+；注意力机制；OCT 图像；激光焊接

ABSTRACT

Optical Coherence Tomography (OCT) image semantic segmentation is a key component in industrial vision tasks such as laser welding molten pool inspection. In practical scenarios, speckle noise, blurred boundaries, and weak target–background contrast often lead convolution-based segmentation models to suffer from insufficient global context understanding and inaccurate boundary delineation, thereby limiting robustness and practical usability. To address these issues, this thesis proposes an improved DeepLabV3+-based method for OCT image semantic segmentation.

Built upon the DeepLabV3+ encoder–decoder architecture, the proposed approach introduces both global and local attention modeling on top of multi-scale context aggregation. Specifically, a TR (Transformer Routing) module is inserted after ASPP to enhance long-range dependency modeling and global context representation under controllable computational cost via routing-based attention. In addition, an SAE (Spatial Attention Enhancement) module is applied after decoder feature fusion to strengthen local details and boundary-aware representations through joint spatial and channel enhancement. A hybrid loss combining cross-entropy and Dice loss is further adopted to mitigate training bias caused by class imbalance.

Experiments on an OCT dataset with 912 training images and 228 test images demonstrate that the proposed method achieves an mIoU of 0.911, outperforming the baseline DeepLabV3+ by 7.1%. The results indicate that the joint integration of TR and SAE can balance global semantic modeling and local boundary enhancement within a unified encoder–decoder framework, providing an effective solution for high-precision OCT image segmentation in practical industrial scenarios.

Key words: Semantic Segmentation; DeepLabV3+; Attention Mechanism; OCT Image; Laser Welding

目 录

摘 要 ..... **I**
ABSTRACT ..... **II**
符号和缩略语说明 ..... **VII**
第一章 绪论 **1**
1.1 研究背景及意义 1
1.2 国内外研究现状 2
1.2.1 激光焊接过程监测技术现状 2
1.2.2 OCT 图像噪声特性及处理方法研究现状 2
1.2.3 图像语义分割算法研究现状 3
1.3 本文主要研究内容 3
1.4 论文组织结构 4
第二章 相关理论与技术基础 **6**
2.1 OCT 成像原理 6
2.1.1 OCT 成像基本原理 6
2.1.2 OCT 技术分类 7
2.1.3 OCT 在激光焊接中的应用 8
2.1.4 OCT 图像特点 8
2.2 深度学习基础 8
2.2.1 卷积神经网络基础 9
2.2.2 编码器-解码器架构 10
2.2.3 注意力机制基础 11
2.3 语义分割理论基础 11
2.3.1 全卷积网络 12
2.3.2 U-Net 架构 12
2.3.3 DeepLab 系列方法 12
2.3.4 多尺度特征融合策略 13
2.4 语义分割评价指标 14
2.4.1 混淆矩阵 14
2.4.2 像素准确率 14
2.4.3 平均像素准确率 15
2.4.4 交并比 15

2.4.5 平均交并比 15
2.4.6 Dice 系数 16
2.4.7 评价指标的适用场景 16
2.5 本章小结 17
第三章 基于改进 DeepLabV3+ 的 OCT 图像语义分割方法 19
3.1 引言 19
3.2 总体框架 19
3.2.1 改进动机与模块定位 19
3.2.2 网络架构设计 20
3.3 DeepLabV3+ 基线模型 20
3.3.1 编码器-解码器架构 20
3.3.2 骨干网络：ResNet18-V1c 21
3.3.3 ASPP 模块设计 21
3.3.4 解码器设计 22
3.3.5 基线模型在 OCT 图像分割中的局限性 23
3.4 TR 全局注意力模块 24
3.4.1 模块结构 24
3.4.2 双层路由注意力机制 24
3.4.3 复杂度分析 25
3.4.4 模块作用小结 25
3.5 SAE 空间细节增强模块 26
3.5.1 坐标注意力机制 26
3.5.2 通道注意力机制 26
3.5.3 模块作用小结 27
3.6 混合损失函数 27
3.6.1 二元交叉熵损失 27
3.6.2 Dice 损失 28
3.6.3 组合损失函数 28
3.7 本章小结 28
第四章 实验结果与分析 30
4.1 实验设置 30
4.1.1 数据集构建 30
4.1.2 实验环境 30
4.1.3 训练参数设置 31

4.1.4 数据增强策略 31
4.1.5 评估指标 32
4.2 对比实验 33
4.2.1 对比方法介绍 33
4.2.2 定量评估结果 34
4.2.3 性能提升分析 34
4.2.4 类别级别性能对比 35
4.2.5 边界精度对比 36
4.2.6 结果分析与讨论 36
4.3 消融实验 37
4.3.1 消融实验设计 37
4.3.2 消融实验结果 37
4.3.3 模块贡献分析 38
4.3.4 模块贡献度分析 38
4.3.5 组合效果解释 39
4.4 可视化分析 40
4.4.1 典型样例可视化 40
4.4.2 可视化结果分析 40
4.4.3 失败案例分析 41
4.4.4 不同场景下的性能表现 41
4.5 效率分析 42
4.5.1 参数量对比 42
4.5.2 内存占用对比 43
4.5.3 推理时间对比 43
4.5.4 效率权衡分析 44
4.6 本章小结 45
第五章 总结与展望 46
5.1 论文总结 46
5.1.1 研究背景与意义回顾 46
5.1.2 主要研究工作总结 46
5.1.3 主要创新点 47
5.1.4 实验成果 48

5.2 方法局限性分析 48
5.2.1 参数量与计算复杂度 48
5.2.2 数据集局限性 49
5.2.3 方法局限性 49
5.3 研究展望 50
5.3.1 轻量化部署 50
5.3.2 多任务协同 50
5.3.3 数据增强与泛化 51
5.3.4 方法改进 51
5.3.5 应用拓展 51
参考文献 53
致 谢 55
攻读硕士学位期间的研究成果 56

符号和缩略语说明

OCT	光学相干断层扫描（Optical Coherence Tomography）
ASPP	空洞空间金字塔池化（Atrous Spatial Pyramid Pooling）
TR	变换器路由（Transformer Routing）
SAE	空间注意力增强（Spatial Attention Enhancement）
mIoU	平均交并比（Mean Intersection over Union）
IoU	交并比（Intersection over Union）
FCN	全卷积网络（Fully Convolutional Network）
CNN	卷积神经网络（Convolutional Neural Network）
ResNet	残差网络（Residual Network）

第一章　绪论

1.1　研究背景及意义

随着现代工业制造向精密化、自动化方向发展，激光焊接技术凭借能量密度高、焊接速度快、热影响区小等特点，在汽车制造与动力电池等高端装备制造领域得到了广泛应用[1]。在叠焊等典型工艺中，熔深直接影响接头的有效连接与承载能力，是衡量焊接质量的关键指标之一[1]。然而，激光焊接过程涉及光、机、电、热等多物理场耦合，具有明显的非线性与不稳定性，易产生孔隙、未熔合等缺陷；若仍依赖金相切片等焊后离线检测，往往存在破坏性强、效率低、反馈滞后等问题，难以满足规模化生产的实时质检与闭环控制需求[1]。因此，面向焊接过程的实时在线监测与熔深稳定获取具有重要的工程价值。

光学相干层析成像（Optical Coherence Tomography, OCT）是一种基于低相干干涉原理的层析成像技术，具有非接触、高轴向分辨率等优势[2]。在激光焊接场景中，OCT 可通过测量光束获取匙孔（Keyhole）底部反射信息，从而为熔深在线测量提供直接的几何依据，被认为是具有应用潜力的过程监测手段之一[1]。相较于仅能反映表面辐射或二维表观信息的视觉/光电监测方法，OCT 在表征深度结构方面更具优势[1]。

尽管 OCT 在焊接监测中具有优势，但在实际应用中仍面临信号不稳定与噪声干扰等挑战：焊接过程中的熔池波动、金属蒸汽与飞溅会降低信号稳定性[1]；同时，OCT 成像基于低相干干涉，散斑噪声（Speckle Noise）常以随机颗粒纹理形式出现并降低图像信噪比[2]。在此条件下，锁孔边缘往往呈现对比度低、边界模糊与形态快速变化等特点，传统阈值分割、边缘检测或简单滤波（如中值滤波、百分位滤波）难以在鲁棒性与边界保持之间取得平衡，进而影响后续熔深估计的稳定性与精度[1,2]。

针对上述痛点，本文研究面向焊接 OCT 图像的语义分割方法。与传统图像处理流程相比，深度学习分割模型可通过端到端学习获取更具判别性的语义特征与上下文信息，在噪声与弱边界条件下具有更强的鲁棒性[3–5]。本文以 DeepLabV3+ 为基线[5]，面向散斑噪声干扰强、锁孔边界模糊与细长结构易断裂等问题，在编码器端增强全局上下文建模，在解码器端强化局部细节表达，从而实现对锁孔区域的高精度分割，为熔深稳定测量提供算法支撑[1]。

1.2　国内外研究现状

本节围绕本文研究对象（激光焊接场景下的 OCT 锁孔/熔深信息获取）与研究任务（OCT 图像中锁孔目标的语义分割）展开综述。首先总结激光焊接过程监测与熔深测量的主流技术路线及其局限性；其次梳理 OCT 成像中的散斑噪声特性与典型处理方法，指出“先去噪再分割/测量”的流程在边界保持与实时性方面的不足；最后回顾图像语义分割算法的发展脉络与在 OCT/工业视觉中的应用进展，从而引出本文面向噪声干扰与弱边界条件的端到端抗噪分割思路。

1.2.1　激光焊接过程监测技术现状

激光焊接过程监测手段大体可分为基于光辐射/光谱的光电监测、基于可见光/红外成像的视觉监测、基于声发射/等离子体信号的声学监测，以及融合多源传感的综合监测等。上述方法在一定程度上能够反映焊接稳定性与缺陷趋势，但往往难以直接、稳定地获取锁孔内部几何信息；同时其信号与熔深之间的映射关系受工艺参数、材料状态与飞溅/蒸汽干扰影响显著，导致可迁移性与精度受限。在此背景下，OCT 凭借其高轴向分辨率、非接触、可穿透烟尘与金属蒸汽并直接表征锁孔深度结构等优势，被认为是实现熔深在线测量与闭环控制的有前景手段[1]。然而，OCT 图像存在噪声强、对比度低、边界模糊等特点，使得锁孔区域的稳定提取仍是亟待解决的关键问题。

1.2.2　OCT 图像噪声特性及处理方法研究现状

OCT 成像基于低相干干涉，散斑噪声来源于相干叠加与多次散射等因素，常表现为随机颗粒纹理并伴随局部对比度下降，其统计特性与组织/材料表面微结构密切相关[2]。为提升后续结构识别与测量的稳定性，国内外研究提出了多种 OCT 噪声抑制与质量增强方法，主要包括：**（1）传统滤波与变换域方法**，如中值/均值滤波、各向异性扩散、BM3D 及小波/非局部均值等；**（2）硬件或多帧复合策略**，通过多角度/多次采样实现散斑平均以提升信噪比；**（3）深度学习去噪方法**，利用卷积网络或自编码结构学习噪声分布并恢复结构细节。上述方法为改善可视化质量提供了有效途径，但在锁孔这类**弱边界、细长深孔、形态快速变化**的目标上，单纯追求去噪往往可能带来边界平滑、细节损失与几何偏移；同时，“先去噪再分割/测量”的两阶段流程增加了系统复杂度与推理延迟，不利于在线闭环控制。

因此，本文不将“去噪”作为独立处理步骤，而是将散斑与干扰视为成像条件的一部分，探索**端到端的抗噪语义分割**：通过在分割网络内部显式增强全局上下文建模与局部边界细节表达，使模型在噪声干扰下仍能稳定学习锁孔的语义结构与轮廓，从流程上减少对额外预处理的依赖。

1.2.3　图像语义分割算法研究现状

语义分割方法经历了从全卷积网络（FCN）到编码器-解码器结构（U-Net 及其变体），再到多尺度上下文建模（DeepLab 系列空洞卷积与 ASPP）等阶段；近年来，Transformer 及注意力机制被引入分割任务，通过自注意力或混合架构增强长程依赖建模能力，在医学影像与复杂场景分割中表现突出[3–6]。在 OCT 相关任务中，研究者多采用 U-Net/DeepLab 类结构进行层结构或病灶区域分割，也有工作引入注意力机制以提升边界与小目标刻画能力。

然而，激光焊接 OCT 锁孔分割与常规医学 OCT 分割在数据分布与目标形态上存在差异：其一，散斑噪声与工况扰动更强，导致特征不稳定；其二，锁孔区域往往呈现**细长结构**且边界灰度对比度低，易出现断裂与粘连；其三，在线应用对推理速度与稳定性提出更高要求。上述特点使得现有通用分割网络直接迁移时可能面临全局结构判断不足与局部边界细节丢失等问题。基于此，本文以 DeepLabV3+ 为基线，分别从**全局上下文**与**局部细节**两方面进行针对性增强，为后续章节提出的 TR 模块与 SAE 模块提供动机与理论依据。

1.3　本文主要研究内容

针对激光焊接 OCT 图像中存在的散斑噪声干扰严重、锁孔目标细小且边界模糊等问题，本文以实现高精度、鲁棒的锁孔语义分割为目标，开展了一系列研究工作。本文的主要研究内容如下：

1. **构建了面向激光焊接锁孔检测的 OCT 图像语义分割数据集。**针对现有开源数据集缺乏此类工业场景数据的问题，本文收集了真实的 304 不锈钢激光焊接 OCT 成像数据，制定了统一的标注规范，完成了像素级的精细标注。数据集涵盖了不同焊接工艺参数下的多种锁孔形态，为模型训练与评估提供了坚实的数据基础。

2. **提出了一种基于改进 DeepLabV3+ 的 OCT 图像语义分割网络。**

针对 DeepLabV3+ 模型在处理 OCT 图像时存在的全局上下文信息利用不足和局部细节丢失问题，本文在编码器-解码器架构的基础上进行了双重改进：

（1） 在编码器末端引入 **TR（Transformer Routing）模块**。借鉴 Transformer 全局建模思想，在编码器高层特征上引入基于路由稀疏注意力的全局上下文增强，以提升长程依赖建模能力并降低无关区域干扰[6,7]。

（2） 在解码器融合阶段引入 **SAE（Spatial Attention Enhancement）模块**。在高低层特征融合后引入空间与通道注意力协同增强，突出边缘与细小结构的特征响应，提升弱边界条件下的分割精细度[8,9]。

3. **设计了适应类别不平衡的混合损失函数与训练策略。**针对 OCT 图像中背景区域广大而锁孔目标区域较小（前景背景比例失衡）的问题，采用交叉熵损失（Cross Entropy Loss）与 Dice 损失（Dice Loss）相结合的混合损失函数，平衡了模型对不同类别的关注度，进一步提升了分割精度（mIoU）。

4. **进行了系统的实验验证与对比分析。**在自建数据集上对所提方法进行了全面的实验评估。实验结果表明，改进后的模型在 mIoU、Dice 系数等关键指标上均优于基线 DeepLabV3+ 及 U-Net、TransUNet 等主流分割网络。通过消融实验验证了 TR 模块与 SAE 模块的有效性，并对分割结果进行了可视化分析，证明了本文方法在复杂工况下的鲁棒性与优越性。

1.4　论文组织结构

本文共分为五章，各章节的具体安排如下：

第一章：绪论。阐述了课题的研究背景及意义，分析了激光焊接过程监测面临的挑战及 OCT 技术的应用潜力。综述了国内外在激光焊接监测、OCT 图像去噪及语义分割算法方面的研究现状，指出了现有研究的不足。最后明确了本文的研究目标、主要研究内容及章节安排。

第二章：相关理论与技术基础。（待补充：介绍 OCT 成像原理、深度学习基础、卷积神经网络、语义分割常用评价指标等。）

第三章：基于改进 DeepLabV3+ 的 OCT 图像语义分割方法。详细阐述本文提出的改进网络架构。重点介绍 DeepLabV3+ 基线模型、TR 全局注意力模块、SAE 空间细节增强模块的设计原理与实现细节，以及混合损失函数的定义。

第四章：实验结果与分析。介绍实验数据集的构建、实验环境与参数设置。展示本文方法与主流对比方法的定量评估结果（如 mIoU、PA 等）及定性可视化效果。通过消融实验深入分析各改进模块对模型性能的贡献，并对实验结果进行讨论。

第五章：总结与展望。总结全文的研究工作与创新点，客观分析现有方法的局限性，并对未来的研究方向（如轻量化部署、多任务协同等）进行展望。

第二章　相关理论与技术基础

本章介绍本文研究所需的相关理论与技术基础，主要包括 OCT 成像原理、深度学习基础、语义分割理论基础以及评价指标等内容，为后续章节的方法设计与实验分析提供理论支撑。

2.1　OCT 成像原理

光学相干层析成像（Optical Coherence Tomography, OCT）是一种基于低相干干涉原理的高分辨率、非侵入性层析成像技术[10]。该技术最早由 Huang 等人于 1991 年提出[10]，通过检测参考光束与样品后向散射光之间的干涉信息，实现对生物组织或材料结构的微米级断层成像。OCT 技术无需外源造影剂或样品切片处理，即可实现高分辨率实时成像，在生物医学和工业检测领域具有重要应用价值[2]。

2.1.1　OCT 成像基本原理

OCT 成像系统基于低相干干涉原理，其核心结构为迈克尔逊干涉仪。系统主要由宽带光源、光纤耦合器、参考臂、样品臂、光电探测器等关键组件构成。宽带光源发出的光经光纤耦合器被分成两束：一束进入参考臂，射向可沿光轴移动的反射镜；另一束进入样品臂，在样品不同深度产生背散射光。这些背散射光和经参考镜反射的参考光在光纤耦合器处发生干涉，被探测器接收。通过分析干涉信号的强度和时延，即可重建样品的深度结构信息[1]。

光波发生低相干干涉需要满足以下条件：光波相位差恒定、振动频率相同和振动方向一致。在 OCT 系统中，参考臂与样品臂的光程差决定了干涉信号的强度。当光程差小于光源的相干长度时，才能发生有效干涉；当光程差为零时，系统可获得最大干涉强度。OCT 系统的轴向分辨能力主要取决于光源的相干特性，其理论极限值约为相干长度的一半，该特性使得 OCT 技术能够实现微米级的纵向分辨[2]。

从数学角度分析，设宽带光源对应的电场 E_i 可表示为：

$$E_i = S(k)e^{i(kz-\omega t)} \tag{2-1}$$

其中，$S(k)$ 为电场的振幅，k 为波数，z 为光程，ω 为角频率。电场的振幅 $S(k)$ 可进一步表示为高斯型函数：

$$S(k) = \frac{1}{\Delta k\sqrt{\pi}} e^{-\frac{(k-k_0)^2}{\Delta k^2}} \tag{2-2}$$

其中，k_0 为中心波长 λ_0 的波数，Δk 为频谱带宽。

设参考臂反射光场为 E_R，样品臂反射光场为 E_S，则单一波长的光谱干涉信号 $I_D(k)$ 可表示为：

$$I_D(k) \approx |E_R|^2 + |E_S|^2 + 2r_R r_S S(k)\cos[2k(z_R - z_S)] \tag{2-3}$$

其中，r_R 和 r_S 分别为参考臂和样品臂的光束反射率，z_R 和 z_S 分别为参考臂和样品臂的光程长度。干涉信号由直流项和交叉干涉项组成，交叉干涉项包含了样品的深度信息。对 $I_D(k)$ 进行快速傅里叶变换即可得到相应的深度信息，从而重建样品的层析结构[1]。

2.1.2　OCT 技术分类

根据不同的成像原理和数据采集方式，OCT 技术主要分为时域 OCT（Time Domain OCT, TD-OCT）和频域 OCT（Fourier Domain OCT, FD-OCT）两大类[2]。

时域 OCT 是最早提出的 OCT 技术，通过机械扫描参考臂的延时调制，逐点扫描获得不同深度处的反射信号。时域 OCT 的成像过程包括两个步骤：一是通过参考臂的移动实现深度扫描，从干涉信号中提取样品的深度结构；二是结合横向扫描，构建二维或三维层析图像。时域 OCT 通过调节参考臂的扫描范围来灵活设定测量深度，但由于其深度信息依赖机械扫描，导致扫描时间较长、成像速度较慢，并容易受到扫描噪声干扰[2]。

频域 OCT 通过获取完整的光谱干涉信号，并利用傅里叶变换重建深度信息，避免了机械扫描的需求，显著提高了成像速度。频域 OCT 根据光源和探测方式的差异，又可分为谱域 OCT（Spectral Domain OCT, SD-OCT）和扫频 OCT（Swept Source OCT, SS-OCT）。谱域 OCT 使用宽带光源和光谱仪，在空间域中分离干涉光并获取干涉光谱；扫频 OCT 采用宽带扫频光源和高速光电探测器，通过扫频光源在时间域内依次发射不同波长的光，并利用光电探测器在时间域中分离干涉光谱。相比谱域 OCT，扫频 OCT 在信噪比、探测灵敏度和系统分辨率方面具有

更高的潜力，尤其适用于高分辨率和大视场成像[1]。

2.1.3　OCT 在激光焊接中的应用

在激光焊接场景中，OCT 技术通过测量光束获取匙孔（Keyhole）底部反射信息，从而为熔深在线测量提供直接的几何依据[1]。相较于仅能反映表面辐射或二维表观信息的视觉/光电监测方法，OCT 在表征深度结构方面更具优势。激光焊接区域的反射散射光与参考臂的镜面反射光之间相位差并不恒定，因此激光加工区域的散射光与参考臂的反射光不会发生干涉，谱域 OCT 成像系统几乎不受激光加工光束的反射散射光影响，这使得 OCT 技术在激光焊接过程监测中具有独特的应用优势[1]。

2.1.4　OCT 图像特点

尽管 OCT 具备高分辨率、非侵入性等优势，但在实际应用中仍面临信号不稳定与噪声干扰等挑战。OCT 图像的主要特点包括：

（1）散斑噪声：OCT 成像基于低相干干涉原理，散斑噪声来源于相干叠加与多次散射等因素，常表现为随机颗粒纹理并伴随局部对比度下降[2]。焊接过程中的熔池波动、金属蒸汽与飞溅会进一步降低信号稳定性，加剧散斑噪声的影响[1]。

（2）对比度低：由于散斑噪声的存在，OCT 图像的局部对比度往往较低，特别是在弱反射区域，目标与背景的灰度差异不明显，增加了后续处理的难度。

（3）边界模糊：在激光焊接 OCT 图像中，锁孔边缘往往呈现对比度低、边界模糊与形态快速变化等特点，传统阈值分割、边缘检测或简单滤波难以在鲁棒性与边界保持之间取得平衡[1]。

上述特点使得 OCT 图像中的目标区域（如锁孔）提取成为一项具有挑战性的任务，需要采用更加鲁棒的图像处理方法，这也是本文研究面向 OCT 图像的语义分割方法的重要动机。

2.2　深度学习基础

深度学习作为机器学习的重要分支，通过构建具有多个隐藏层的神经网络，能够自动学习数据的层次化特征表示。在图像处理领域，卷积神经网络（Convolutional Neural Network, CNN）凭借其强大的特征提取能力，已成为语义分割等

视觉任务的主流方法。本节介绍深度学习的基础理论，包括卷积神经网络、编码器-解码器架构以及注意力机制等核心概念。

2.2.1　卷积神经网络基础

卷积神经网络是一种专门用于处理具有网格结构数据（如图像）的深度学习模型。CNN 的核心思想是通过局部连接、权值共享和池化操作，有效减少参数量并提取平移不变的特征。

（1）卷积层：卷积层是 CNN 的基本组成单元，通过卷积核（滤波器）在输入特征图上滑动，计算局部区域的加权和。设输入特征图为 $\mathbf{X} \in \mathbb{R}^{H \times W \times C}$，卷积核为 $\mathbf{K} \in \mathbb{R}^{k \times k \times C}$，则卷积操作可表示为：

$$Y_{i,j} = \sum_{m=0}^{k-1} \sum_{n=0}^{k-1} \sum_{c=0}^{C-1} X_{i+m,j+n,c} \cdot K_{m,n,c} + b \tag{2-4}$$

其中，$Y_{i,j}$ 为输出特征图在位置 (i, j) 的值，b 为偏置项。卷积操作具有局部连接和权值共享的特性，能够有效提取图像的局部特征（如边缘、纹理等），同时大幅减少参数量。

（2）池化层：池化层通过下采样操作降低特征图的空间分辨率，减少计算量并增强特征的平移不变性。常用的池化操作包括最大池化（Max Pooling）和平均池化（Average Pooling）。最大池化选择局部区域内的最大值，能够保留显著特征；平均池化计算局部区域的平均值，能够平滑特征响应。

（3）激活函数：激活函数为网络引入非线性，使网络能够学习复杂的非线性映射关系。常用的激活函数包括：

- **ReLU（Rectified Linear Unit）**：$f(x) = \max(0, x)$，具有计算简单、梯度稳定等优点，是目前最常用的激活函数。
- **Sigmoid**：$f(x) = \frac{1}{1+e^{-x}}$，输出范围在 $(0, 1)$ 之间，常用于二分类任务的输出层。
- **Softmax**：$f(x_i) = \frac{e^{x_i}}{\sum_j e^{x_j}}$，将输入向量归一化为概率分布，常用于多分类任务的输出层。

（4）批归一化：批归一化（Batch Normalization）通过在训练过程中对每个批次的数据进行归一化，能够加速网络训练、提高模型稳定性并允许使用更大的学

习率[11]。批归一化的操作可表示为：

$$\hat{x}_i = \frac{x_i - \mu_B}{\sqrt{\sigma_B^2 + \epsilon}}, \quad y_i = \gamma \hat{x}_i + \beta \tag{2-5}$$

其中，μ_B 和 σ_B^2 分别为批次数据的均值和方差，γ 和 β 为可学习的缩放和偏移参数，ϵ 为小常数以防止除零。

（5）残差连接：残差连接（Residual Connection）通过将输入直接传递到输出，解决了深层网络的梯度消失问题，使得训练更深的网络成为可能[12]。残差块的结构可表示为：

$$\mathbf{y} = \mathcal{F}(\mathbf{x}, \{\mathbf{W}_i\}) + \mathbf{x} \tag{2-6}$$

其中，$\mathcal{F}(\mathbf{x}, \{\mathbf{W}_i\})$ 为残差映射，$\mathbf{x}$ 为恒等映射。残差连接使得网络能够学习残差而非完整的映射，从而简化了优化过程。ResNet 通过引入残差连接，成功训练了包含数百层的深层网络，在图像分类、目标检测等任务中取得了显著成果[12]。

2.2.2　编码器-解码器架构

编码器-解码器（Encoder-Decoder）架构是语义分割任务中常用的网络结构，通过编码器提取多尺度特征，解码器恢复空间分辨率并输出像素级预测结果[4]。

（1）编码器：编码器通常采用预训练的 CNN 骨干网络（如 ResNet、VGG 等），通过逐层卷积和下采样操作，将输入图像转换为高维特征表示。编码过程逐步降低特征图的空间分辨率，同时增加通道数，提取从低级到高级的语义特征。例如，ResNet 编码器通常包含多个阶段（Stage），每个阶段通过卷积和下采样操作，将特征图尺寸减半、通道数加倍，最终得到高维的语义特征表示。

（2）解码器：解码器通过上采样和特征融合操作，逐步恢复特征图的空间分辨率，并输出与输入图像尺寸相同的预测结果。上采样操作通常采用双线性插值或转置卷积实现。解码器还通过跳跃连接（Skip Connection）融合编码器不同阶段的特征，结合低层细节信息和高层语义信息，提升分割精度。

（3）跳跃连接：跳跃连接将编码器的中间特征直接传递到解码器的对应层，使得解码器能够利用编码器提取的多尺度特征。U-Net 架构通过 U 形的跳跃连接，将编码器的特征图与解码器对应层的特征图进行拼接，有效融合了细节信息和语义信息，在医学图像分割等任务中取得了优异性能[4]。

2.2.3　注意力机制基础

注意力机制通过动态分配不同权重，使网络能够关注输入数据中的重要部分，从而提升模型的表达能力。在计算机视觉任务中，注意力机制主要包括自注意力、空间注意力和通道注意力等类型。

（1）自注意力机制：自注意力（Self-Attention）机制通过计算特征图中不同位置之间的相关性，建立长距离依赖关系[13]。自注意力的计算过程可表示为：

$$\text{Attention}(\mathbf{Q},\mathbf{K},\mathbf{V}) = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_k}}\right)\mathbf{V} \tag{2-7}$$

其中，**Q**、**K**、**V** 分别为查询（Query）、键（Key）和值（Value）矩阵，d_k 为键向量的维度。自注意力机制能够捕获特征图中任意两个位置之间的依赖关系，不受空间距离限制，在长距离依赖建模方面具有优势。

（2）Transformer 架构：Transformer 架构完全基于自注意力机制，摒弃了传统的卷积和循环结构，通过多头自注意力和前馈网络构建深层网络[13]。Vision Transformer（ViT）将 Transformer 架构应用于图像分类任务，将图像分割为固定大小的图像块（Patch），并将每个图像块视为一个序列元素，通过自注意力机制建模图像块之间的关系[14]。Transformer 架构在图像分割任务中也得到了广泛应用，通过增强长距离依赖建模能力，提升了分割性能。

（3）空间注意力与通道注意力：空间注意力关注特征图的空间位置信息，通过生成空间权重图，突出重要的空间区域；通道注意力关注特征图的通道维度，通过生成通道权重向量，突出重要的特征通道。空间注意力和通道注意力可以单独使用，也可以组合使用，通过空间和通道维度的双重增强，提升特征的判别能力。本文后续章节提出的 SAE 模块即结合了空间注意力和通道注意力，用于增强局部细节和边界信息。

2.3　语义分割理论基础

语义分割是计算机视觉中的一项重要任务，旨在对图像中的每个像素进行分类，为每个像素分配一个语义类别标签。与图像分类和目标检测不同，语义分割需要在像素级别进行预测，要求模型同时具备强大的特征提取能力和精确的空间定位能力。本节介绍语义分割的发展历程和关键技术，包括全卷积网络、U-Net 架构以及 DeepLab 系列方法。

2.3.1　全卷积网络

全卷积网络（Fully Convolutional Networks, FCN）是首个将深度卷积神经网络成功应用于语义分割任务的方法[3]。FCN 的核心思想是将传统 CNN 中的全连接层替换为卷积层，使得网络能够接受任意尺寸的输入图像，并输出与输入尺寸相同的分割结果。

FCN 通过将预训练的分类网络（如 VGG、ResNet）转换为全卷积结构，利用卷积层的平移不变性，实现了端到端的像素级预测。FCN 还引入了跳跃连接机制，通过融合不同尺度的特征图，结合浅层的细节信息和深层的语义信息，提升了分割精度。FCN 的提出标志着深度学习在语义分割领域的成功应用，为后续研究奠定了基础[3]。

2.3.2　U-Net 架构

U-Net 是一种经典的编码器-解码器架构，最初设计用于医学图像分割任务[4]。U-Net 采用对称的 U 形结构，编码器通过卷积和下采样逐步提取特征，解码器通过上采样和卷积逐步恢复分辨率。

U-Net 的关键创新在于其密集的跳跃连接设计：编码器每一层的特征图都与解码器对应层的特征图进行拼接，使得解码器能够充分利用编码器提取的多尺度特征。这种设计使得 U-Net 在保持高分辨率细节的同时，能够利用深层的语义信息，在医学图像分割等需要精确边界定位的任务中表现出色[4]。U-Net 的成功也启发了后续许多编码器-解码器架构的设计。

2.3.3　DeepLab 系列方法

DeepLab 系列方法是语义分割领域的重要里程碑，通过引入空洞卷积和空间金字塔池化等技术，显著提升了分割性能。DeepLab 系列的发展历程体现了语义分割技术的演进轨迹。

（1）DeepLabV1：DeepLabV1 首次将空洞卷积（Atrous Convolution）引入语义分割任务[15]。空洞卷积通过在卷积核中插入零值，在不增加参数量的情况下扩大感受野，使得网络能够在保持特征图分辨率的同时捕获更大范围的上下文信

息。空洞卷积的数学表达可写为：

$$y[i] = \sum_{k} x[i + r \cdot k] \cdot w[k] \tag{2-8}$$

其中，r 为空洞率（dilation rate），控制感受野的大小。当 $r = 1$ 时，空洞卷积退化为标准卷积；当 $r > 1$ 时，空洞卷积能够在不增加参数量的情况下扩大感受野。

（2）DeepLabV2：DeepLabV2 在 DeepLabV1 的基础上引入了空间金字塔池化（Atrous Spatial Pyramid Pooling, ASPP）模块[16]。ASPP 模块通过并行使用多个不同空洞率的空洞卷积，在不同尺度上捕获上下文信息，然后将多尺度特征进行融合。ASPP 模块的设计使得网络能够同时关注局部细节和全局上下文，提升了分割性能。

（3）DeepLabV3：DeepLabV3 对 ASPP 模块进行了改进，引入了全局平均池化分支，并优化了空洞率的选择策略[17]。改进后的 ASPP 模块包含多个并行分支：不同空洞率的空洞卷积分支用于捕获多尺度上下文信息，全局平均池化分支用于捕获全局上下文信息。这些分支的特征经过拼接和融合，形成丰富的多尺度特征表示。

（4）DeepLabV3+：DeepLabV3+ 在 DeepLabV3 的基础上引入了编码器-解码器结构[5]。DeepLabV3+ 采用 DeepLabV3 的输出作为编码器特征，通过解码器逐步上采样并融合编码器的中间特征，最终输出高分辨率的分割结果。DeepLabV3+ 的解码器设计简洁高效，通过融合低层细节信息和高层语义信息，在保持分割精度的同时提升了边界定位的准确性。DeepLabV3+ 是本文方法的基础框架，其编码器-解码器结构和 ASPP 模块为本文的改进提供了良好的起点。

2.3.4　多尺度特征融合策略

多尺度特征融合是语义分割中的关键技术，旨在结合不同尺度的特征信息，提升模型对不同大小目标的处理能力。除了 ASPP 模块外，特征金字塔网络（Feature Pyramid Network, FPN）也是一种重要的多尺度特征融合方法[18]。

FPN 通过构建特征金字塔，在不同尺度上提取特征，并通过自上而下的路径和横向连接融合多尺度特征。FPN 的设计使得网络能够在不同尺度上检测目标，提升了模型对多尺度目标的处理能力。虽然 FPN 最初设计用于目标检测任务，但其多尺度特征融合的思想在语义分割任务中也得到了广泛应用。

近年来，Transformer 架构也被引入语义分割任务，通过自注意力机制增强长距离依赖建模能力。TransUNet 将 Transformer 作为编码器，结合 U-Net 的解码器结构，在医学图像分割任务中取得了优异性能[6]。Transformer 架构的引入为语义分割提供了新的思路，但其计算复杂度较高，在实际应用中需要权衡性能与效率。

2.4　语义分割评价指标

为了客观评估语义分割模型的性能，需要采用合适的评价指标对预测结果进行定量分析。语义分割的评价指标通常基于混淆矩阵（Confusion Matrix）计算，通过比较预测结果与真实标签之间的差异，量化模型的分割精度。本节介绍常用的语义分割评价指标及其计算方法。

2.4.1　混淆矩阵

混淆矩阵是评估分类模型性能的基础工具，用于统计预测结果与真实标签之间的对应关系。对于语义分割任务，混淆矩阵 C 是一个 $N \times N$ 的矩阵，其中 N 为类别数。矩阵元素 C_{ij} 表示真实类别为 i、预测类别为 j 的像素数量。

基于混淆矩阵，可以计算多种评价指标。对于二分类任务（如本文的锁孔分割任务），混淆矩阵包含四个元素：真正例（True Positive, TP）、假正例（False Positive, FP）、真负例（True Negative, TN）和假负例（False Negative, FN）。

2.4.2　像素准确率

像素准确率（Pixel Accuracy, PA）是最直观的评价指标，表示正确分类的像素占总像素的比例：

$$\text{PA} = \frac{\sum_{i=0}^{N-1} C_{ii}}{\sum_{i=0}^{N-1} \sum_{j=0}^{N-1} C_{ij}} \tag{2-9}$$

其中，C_{ii} 为混淆矩阵对角线元素，表示正确分类的像素数；分母为总像素数。PA 的取值范围为 $[0, 1]$，值越大表示分割精度越高。

PA 指标计算简单直观，但在类别不平衡的情况下（如背景像素远多于目标像素），PA 可能无法准确反映模型对少数类别的分割性能。例如，在锁孔分割任务中，背景区域通常占据图像的大部分，即使模型将所有像素预测为背景，PA 值也可能较高，但这并不能说明模型的分割性能良好。

2.4.3　平均像素准确率

平均像素准确率（Mean Pixel Accuracy, mPA）通过计算每个类别的像素准确率并取平均，能够更好地反映模型对不同类别的分割性能：

$$\text{mPA} = \frac{1}{N}\sum_{i=0}^{N-1}\frac{C_{ii}}{\sum_{j=0}^{N-1}C_{ij}} \tag{2-10}$$

其中，$\frac{C_{ii}}{\sum_{j=0}^{N-1}C_{ij}}$ 表示类别 i 的像素准确率，即类别 i 中正确分类的像素数占该类别总像素数的比例。mPA 对每个类别赋予相同的权重，能够更好地评估模型在类别不平衡情况下的性能。

2.4.4　交并比

交并比（Intersection over Union, IoU）是语义分割任务中最常用的评价指标之一，表示预测结果与真实标签的交集与并集的比值：

$$\text{IoU}_i = \frac{C_{ii}}{\sum_{j=0}^{N-1}C_{ij} + \sum_{j=0}^{N-1}C_{ji} - C_{ii}} \tag{2-11}$$

其中，分子 C_{ii} 为类别 i 的预测结果与真实标签的交集（正确预测的像素数），分母为两者的并集（预测为类别 i 或真实为类别 i 的像素总数）。IoU 的取值范围为 $[0, 1]$，值越大表示分割精度越高。

IoU 指标能够同时考虑预测结果的准确性和完整性，对分割边界的精度更加敏感，因此在语义分割任务中被广泛采用。对于二分类任务，IoU 的计算可简化为：

$$\text{IoU} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}} \tag{2-12}$$

2.4.5　平均交并比

平均交并比（Mean Intersection over Union, mIoU）通过计算所有类别 IoU 的平均值，综合评估模型的分割性能：

$$\text{mIoU} = \frac{1}{N}\sum_{i=0}^{N-1}\text{IoU}_i \tag{2-13}$$

mIoU 是语义分割任务中最常用的评价指标，能够综合反映模型对不同类别的分割性能。在类别不平衡的情况下，mIoU 比 PA 更能准确反映模型的真实性能。本文实验部分将 mIoU 作为主要评价指标，用于评估不同方法的分割性能。

2.4.6　Dice 系数

Dice 系数（Dice Coefficient）是医学图像分割中常用的评价指标，表示预测结果与真实标签的重叠程度：

$$\text{Dice}_i = \frac{2 \cdot C_{ii}}{2 \cdot C_{ii} + \sum_{j \neq i} C_{ij} + \sum_{j \neq i} C_{ji}} = \frac{2 \cdot \text{TP}}{2 \cdot \text{TP} + \text{FP} + \text{FN}} \tag{2-14}$$

Dice 系数的取值范围为 $[0, 1]$，值越大表示重叠程度越高。对于二分类任务，Dice 系数与 IoU 之间存在以下关系：

$$\text{Dice} = \frac{2 \cdot \text{IoU}}{1 + \text{IoU}} \tag{2-15}$$

Dice 系数对分割区域的完整性更加敏感，在医学图像分割等需要精确区域分割的任务中应用广泛。本文在损失函数设计中采用了 Dice 损失，用于处理类别不平衡问题。

2.4.7　评价指标的适用场景

不同的评价指标适用于不同的场景和需求：

（1）PA：适用于类别分布相对均衡的场景，计算简单直观，但在类别不平衡情况下可能产生误导性结果。

（2）mPA：通过平均各类别的准确率，能够更好地评估类别不平衡情况下的性能，但对每个类别赋予相同权重，可能忽略类别的重要性差异。

（3）IoU 和 mIoU：综合考虑预测结果的准确性和完整性，对分割边界精度敏感，是语义分割任务中最常用的评价指标。mIoU 能够综合反映模型对不同类别的分割性能，适用于多类别分割任务。

（4）Dice 系数：对分割区域的完整性更加敏感，在医学图像分割等需要精确区域分割的任务中应用广泛。Dice 系数与 IoU 相关但侧重点不同，可以根据具体任务需求选择合适的指标。

在实际应用中，通常同时使用多个评价指标，从不同角度全面评估模型性

能。本文实验部分将同时报告 mIoU、PA、mPA 和 Dice 系数等指标，以全面评估所提方法的分割性能。

2.5　本章小结

本章介绍了本文研究所需的相关理论与技术基础，为后续章节的方法设计与实验分析提供了理论支撑。

首先，本章介绍了 OCT 成像原理，包括 OCT 技术的基本原理、技术分类以及在激光焊接中的应用。OCT 技术基于低相干干涉原理，能够实现微米级的高分辨率层析成像，在激光焊接场景中可用于锁孔检测和熔深测量。然而，OCT 图像存在散斑噪声、对比度低、边界模糊等特点，使得目标区域的提取成为一项具有挑战性的任务，这也为本文研究面向 OCT 图像的语义分割方法提供了重要动机。

其次，本章介绍了深度学习的基础理论，包括卷积神经网络、编码器-解码器架构以及注意力机制等核心概念。卷积神经网络通过局部连接、权值共享和池化操作，能够有效提取图像特征；编码器-解码器架构通过编码器提取多尺度特征、解码器恢复空间分辨率，实现了端到端的像素级预测；注意力机制通过动态分配权重，使网络能够关注重要信息，提升了模型的表达能力。这些理论基础为本文方法的设计提供了重要支撑。

再次，本章回顾了语义分割的发展历程和关键技术，包括全卷积网络、U-Net 架构以及 DeepLab 系列方法。DeepLabV3+ 作为本文方法的基础框架，通过编码器-解码器结构和 ASPP 模块，实现了多尺度上下文信息的有效捕获。然而，在处理 OCT 图像时，DeepLabV3+ 仍存在全局上下文信息利用不足和局部细节丢失等问题，这为本文的改进提供了方向。

最后，本章介绍了语义分割常用的评价指标，包括像素准确率、平均像素准确率、交并比、平均交并比以及 Dice 系数等。这些评价指标从不同角度量化模型的分割性能，为后续章节的实验评估提供了标准。

基于本章介绍的理论基础，本文将在第三章提出一种基于改进 DeepLabV3+ 的 OCT 图像语义分割方法。该方法在编码器端引入 TR 模块以增强全局上下文建模能力，在解码器端引入 SAE 模块以增强局部细节和边界信息，从而实现对锁孔区域的高精度分割。第四章将基于本章介绍的评价指标，对所提方法进行全

面的实验验证和性能分析。

第三章　基于改进 DeepLabV3+ 的 OCT 图像语义分割方法

3.1　引言

第二章介绍了 OCT 成像原理、深度学习基础以及语义分割的理论基础，为本章的方法设计提供了理论支撑。针对 DeepLabV3+ 在处理 OCT 图像时存在的全局上下文建模不足和局部细节丢失问题，本章提出了一种基于双重注意力机制的改进方案：在编码器端引入 TR（Transformer Routing）模块增强全局依赖建模，在解码器端引入 SAE（Spatial Attention Enhancement）模块增强局部细节表达。

本章结构如下：3.2 节介绍总体框架与改进动机；3.3 节详细介绍 DeepLabV3+ 基线模型；3.4 节和 3.5 节分别阐述 TR 模块和 SAE 模块的设计原理；3.6 节介绍混合损失函数；3.7 节对本章内容进行总结。

3.2　总体框架

本文提出的方法基于 DeepLabV3+ 的编码器-解码器架构，通过引入 TR 全局注意力模块和 SAE 空间细节增强模块，实现了对 OCT 图像中锁孔区域的高精度分割。整体网络架构如图 3-1 所示，主要包括编码器、ASPP 模块、TR 模块、解码器和 SAE 模块等核心组件。

3.2.1　改进动机与模块定位

DeepLabV3+ 作为语义分割领域的经典方法，在多尺度上下文信息捕获方面表现出色[5]。然而，在处理激光焊接 OCT 图像时，该模型仍面临两个核心挑战：

（1）全局上下文建模不足：ASPP 模块的感受野受限于空洞卷积的局部性，难以建立长距离依赖关系。OCT 图像中的锁孔区域呈现细长结构，需要理解整个图像的空间关系才能准确分割。

（2）局部细节信息丢失：解码器在特征融合过程中，低层的细节信息可能被高层语义信息淹没。OCT 图像中锁孔边缘对比度低、边界模糊，需要保留更多局部细节实现精确边界定位。

针对上述问题，本文的改进策略如下：在 ASPP 模块之后引入 TR 模块，通过双层路由注意力机制建立长距离依赖；在解码器特征融合之后引入 SAE 模块，

通过空间-通道双重注意力增强边界细节。此外，采用 BCE+Dice 混合损失函数应对类别不平衡问题。

3.2.2　网络架构设计

本文方法采用编码器-解码器结构，数据流可描述为：输入图像 → 编码器 →ASPP→TR 模块 → 解码器 →SAE 模块 → 分类头 → 输出。各模块功能如下：

（1）编码器：采用 ResNet18-V1c 作为骨干网络[12]，通过四阶段下采样提取多尺度特征。输入图像（512×512×3）经编码器后，Stage 4 输出高层语义特征（16×16×512），Stage 1 输出低层细节特征（128×128×64）用于解码器融合。

（2）ASPP 模块：通过 5 个并行分支（1×1 卷积、dilation=12/24/36 的空洞卷积、全局平均池化）捕获多尺度上下文信息[17]，输出特征维度为 16×16×2560。

（3）TR 模块：位于 ASPP 之后，通过双层路由注意力机制增强全局上下文建模。采用窗口划分与 TopK 路由选择，在保持全局建模能力的同时降低计算复杂度。

（4）解码器：将高层特征上采样后与低层特征融合，得到 128×128×560 的融合特征。

（5）SAE 模块：位于解码器融合之后，通过坐标注意力和通道注意力的协同增强，提升边界分割精度。

（6）分类头：通过深度可分离卷积和上采样，输出与输入图像尺寸相同的分割结果（512×512×2）。

3.3　DeepLabV3+ 基线模型

DeepLabV3+ 是语义分割领域的经典方法，通过编码器-解码器结构和 ASPP 模块，在多尺度上下文信息捕获方面表现出色[5]。本节详细介绍 DeepLabV3+ 的架构设计，包括骨干网络、ASPP 模块和解码器，并分析其在处理 OCT 图像时的局限性，为后续改进模块的设计提供动机。

3.3.1　编码器-解码器架构

DeepLabV3+ 采用编码器-解码器架构，编码器负责提取多尺度特征并捕获语义信息，解码器负责恢复空间分辨率并输出像素级预测结果[5]。编码器通常采用

预训练的 CNN 骨干网络（如 ResNet、Xception 等），通过逐层卷积和下采样操作，将输入图像转换为高维特征表示；解码器通过上采样和特征融合操作，逐步恢复特征图的空间分辨率，并输出与输入图像尺寸相同的预测结果。

DeepLabV3+ 的创新之处在于将 DeepLabV3 的输出作为编码器特征，通过解码器融合编码器的中间特征，在保持分割精度的同时提升了边界定位的准确性[5]。本文方法以 DeepLabV3+ 为基础框架，在此基础上引入 TR 模块和 SAE 模块，进一步提升分割性能。

3.3.2　骨干网络：ResNet18-V1c

本文方法采用 ResNet18-V1c 作为编码器的骨干网络[12]。ResNet18-V1c 是 ResNet 的改进版本，主要特点包括：

（1）Deep Stem 结构：ResNetV1c 使用三个连续的 3×3 卷积替换传统的 7×7 卷积，这种设计能够减少参数量并提升特征提取能力。Deep Stem 结构首先通过三个 3×3 卷积进行特征提取，然后通过最大池化操作进行下采样。

（2）残差连接：ResNet18-V1c 通过残差连接解决了深层网络的梯度消失问题，使得训练更深的网络成为可能[12]。每个残差块包含两个 3×3 卷积层，通过残差连接将输入直接传递到输出，简化了优化过程。

（3）四个阶段的特征提取：ResNet18-V1c 包含四个阶段（Stage），逐步提取从低级到高级的特征表示。Stage 1 至 Stage 4 的下采样率分别为 1/4、1/8、1/16、1/32，通道数依次为 64、128、256、512。Stage 4 的输出特征（16×16×512）具有丰富的语义信息，被送入 ASPP 模块进行多尺度上下文聚合；Stage 1 的输出特征（128×128×64）保留了较多的细节信息，在解码器中与高层特征进行融合。

3.3.3　ASPP 模块设计

ASPP（Atrous Spatial Pyramid Pooling）模块是 DeepLabV3+ 的核心组件，通过多个不同空洞率的并行分支捕获多尺度上下文信息[17]。本文方法采用深度可分离 ASPP 模块（DepthwiseSeparableASPPModule），在保持性能的同时降低了计算复杂度。

ASPP 模块包含 5 个并行分支：

（1）1×1 卷积分支：使用标准 1×1 卷积，感受野为 1×1，用于捕获细节特征。

该分支能够保留特征图的原始信息，避免空洞卷积可能带来的信息丢失。

（2）3×3 空洞卷积分支（dilation=12）：使用 3×3 空洞卷积，空洞率为 12，感受野约为 25×25，用于捕获局部特征。该分支能够在保持特征图分辨率的同时扩大感受野，捕获更大范围的上下文信息。

（3）3×3 空洞卷积分支（dilation=24）：使用 3×3 空洞卷积，空洞率为 24，感受野约为 49×49，用于捕获区域特征。该分支进一步扩大了感受野，能够捕获更大范围的上下文信息。

（4）3×3 空洞卷积分支（dilation=36）：使用 3×3 空洞卷积，空洞率为 36，感受野约为 73×73，用于捕获全局特征。该分支具有最大的感受野，能够捕获更大范围的上下文信息。

（5）全局平均池化分支：对特征图进行全局平均池化，然后通过 1×1 卷积和双线性插值上采样，恢复原始尺寸。该分支的感受野为整个特征图，能够捕获全局上下文信息。

5 个并行分支的输出特征经过拼接后，得到多尺度特征表示（16×16×2560）。随后通过 1×1 卷积降维到 16×16×512，减少特征维度并融合多尺度信息。

深度可分离卷积：ASPP 模块使用深度可分离卷积降低计算复杂度。深度可分离卷积分为两个步骤：首先进行深度卷积（Depthwise Convolution），对每个通道独立进行卷积操作；然后进行点卷积（Pointwise Convolution），使用 1×1 卷积进行通道融合。深度可分离卷积的参数量和计算量远小于标准卷积，在保持性能的同时提升了计算效率。

3.3.4　解码器设计

DeepLabV3+ 的解码器通过上采样和特征融合操作，逐步恢复特征图的空间分辨率 [5]。解码器的设计包括以下几个步骤：

（1）高层特征上采样：ASPP 模块输出的高层特征（16×16×512）经过 1×1 卷积降维后，通过双线性插值上采样到 128×128×512，恢复空间分辨率。

（2）低层特征处理：编码器 Stage 1 的低层特征（128×128×64）通过 1×1 卷积降维到 128×128×48，减少通道数以匹配高层特征的通道数。

（3）特征融合：将上采样后的高层特征（128×128×512）与处理后的低层特征（128×128×48）在通道维度进行拼接，得到融合特征（128×128×560）。这种设

计能够结合低层的细节信息和高层的语义信息，提升分割精度。

（4）分类头：融合特征经过深度可分离卷积和 1×1 卷积，转换为类别 logits（128×128×2），最后通过双线性插值上采样到原始图像尺寸（512×512×2），输出像素级类别预测结果。

3.3.5　基线模型在 OCT 图像分割中的局限性

尽管 DeepLabV3+ 在通用语义分割任务中表现出色，但在处理激光焊接 OCT 图像时仍存在以下局限性，这些局限性构成了本文改进工作的直接动机。

（1）全局上下文建模不足：ASPP 模块通过不同空洞率的并行分支捕获多尺度上下文信息，但其感受野仍受限于空洞卷积的局部性。以 dilation=36 的最大空洞率为例，其有效感受野约为 73×73 像素，仅覆盖 16×16 特征图的局部区域。然而，OCT 图像中的锁孔区域往往呈现细长结构（纵横比可达 1:10 以上），跨越图像的多个区域，需要网络理解整个图像的空间关系才能准确分割。ASPP 模块的局部感受野限制了其对全局结构的理解能力，可能导致细长锁孔结构的断裂或误分割。

（2）局部细节信息丢失：解码器通过简单的通道拼接融合高低层特征，但在融合过程中缺乏对特征重要性的自适应调节。低层特征（Stage 1）虽然保留了丰富的边缘和纹理信息，但其通道数（48 维）远小于高层特征（512 维），在特征拼接后容易被高层语义信息所淹没。OCT 图像中锁孔边缘的对比度低（信噪比通常低于常规自然图像）、边界模糊（受散斑噪声影响），需要网络显式增强并保留局部细节信息才能实现精确的边界定位。

（3）类别不平衡问题：OCT 图像中背景区域通常占据 90% 以上的像素，而锁孔目标区域仅占 5-10%。标准交叉熵损失对每个像素赋予相同权重，导致模型训练时偏向背景类别，难以充分学习前景特征。

基于上述局限性分析，本文提出针对性的改进方案：在编码器端引入 TR 模块建立长距离依赖关系，在解码器端引入 SAE 模块增强边界细节表达，并设计混合损失函数应对类别不平衡问题。

3.4　TR 全局注意力模块

针对 DeepLabV3+ 全局上下文建模不足的问题，本文在 ASPP 模块之后引入 TR（Transformer Routing）全局注意力模块。TR 模块借鉴 BiFormer 的双层路由注意力思想[7]，通过区域级路由和 TopK 选择机制，仅对最相关的窗口进行注意力计算，在保持全局建模能力的同时将计算复杂度从 $O(n^2)$ 降低到 $O(n^2 \times k/p^2)$。

3.4.1　模块结构

TR 模块主要包括位置编码（深度可分离卷积，3×3）、LayerNorm 归一化、双层路由注意力、MLP（扩展-压缩结构）和残差连接等组件[7]。其中，双层路由注意力是核心创新，通过区域级路由和令牌级注意力两个阶段实现高效的全局上下文建模。

3.4.2　双层路由注意力机制

双层路由注意力机制是 TR 模块的核心创新，通过区域级路由和令牌级注意力两个阶段，实现高效的全局上下文建模。具体而言，双层路由注意力机制包括以下步骤：

（1）窗口划分：将输入特征图（16×16×2560）按照 4×4 的窗口大小进行划分，得到 16 个窗口（4×4=16）。每个窗口包含 16 个令牌（token），对应特征图中的 16 个像素位置。

（2）区域级路由：对每个窗口的所有令牌求平均，得到区域级查询向量 $\mathbf{Q}_{\text{region}}^{(i)}$ 和键向量 $\mathbf{K}_{\text{region}}^{(i)}$，其中 i 表示第 i 个窗口。区域级查询和键向量的计算可表示为：

$$\mathbf{Q}_{\text{region}}^{(i)} = \frac{1}{|\mathcal{W}_i|} \sum_{j \in \mathcal{W}_i} \mathbf{Q}_j, \quad \mathbf{K}_{\text{region}}^{(i)} = \frac{1}{|\mathcal{W}_i|} \sum_{j \in \mathcal{W}_i} \mathbf{K}_j \tag{3-1}$$

其中，$\mathcal{W}_i$ 表示第 i 个窗口内的所有令牌，$|\mathcal{W}_i|$ 表示窗口内的令牌数量。

计算窗口间的相似度矩阵 $\mathbf{S}$，其中 $\mathbf{S}_{ij}$ 表示第 i 个窗口与第 j 个窗口的相似度：

$$\mathbf{S}_{ij} = \frac{\mathbf{Q}_{\text{region}}^{(i)} \cdot \mathbf{K}_{\text{region}}^{(j)}}{\|\mathbf{Q}_{\text{region}}^{(i)}\| \times \|\mathbf{K}_{\text{region}}^{(j)}\|} \tag{3-2}$$

其中，$\cdot$ 表示向量内积，$\|\cdot\|$ 表示向量范数。相似度矩阵 $\mathbf{S}$ 的维度为 16×16，表示 16 个窗口之间的相似度关系。

（3）TopK 路由选择：对每个窗口，根据相似度矩阵选择 TopK 个最相关的窗口参与注意力计算。本文设置 $k=4$，即每个窗口选择 4 个最相关的窗口。TopK 路由选择可表示为：

$$\mathcal{R}_i = \text{TopK}(\mathbf{S}_{i,:}, k=4) \tag{3-3}$$

其中，$\mathcal{R}_i$ 表示第 i 个窗口选择的相关窗口索引集合，$\mathbf{S}_{i,:}$ 表示相似度矩阵的第 i 行。

（4）令牌级注意力：在选定的窗口内进行像素级注意力计算。对于第 i 个窗口，仅对 $\mathcal{R}_i$ 中的窗口进行注意力计算，从而降低计算复杂度。令牌级注意力可表示为：

$$\text{Attention}(\mathbf{Q}_i, \mathbf{K}_{\mathcal{R}_i}, \mathbf{V}_{\mathcal{R}_i}) = \text{softmax}\left(\frac{\mathbf{Q}_i \mathbf{K}_{\mathcal{R}_i}^T}{\sqrt{d_k}}\right)\mathbf{V}_{\mathcal{R}_i} \tag{3-4}$$

其中，$\mathbf{Q}_i$、$\mathbf{K}_{\mathcal{R}_i}$、$\mathbf{V}_{\mathcal{R}_i}$ 分别表示第 i 个窗口的查询矩阵和选定窗口的键值矩阵，d_k 为键向量的维度。

3.4.3　复杂度分析

双层路由注意力机制通过 TopK 路由选择，显著降低了计算复杂度。标准自注意力机制的计算复杂度为 $O(n^2)$，其中 n 为特征图中的像素数量。对于 16×16 的特征图，$n=256$，标准自注意力的计算复杂度为 $O(256^2)=O(65536)$。

双层路由注意力机制的计算复杂度为 $O(n^2 \times k/p^2)$，其中 k 为 TopK 值（$k=4$），p 为窗口大小（$p=4$）。对于 16×16 的特征图，划分为 16 个窗口，每个窗口选择 4 个相关窗口，计算复杂度为 $O(256^2 \times 4/16)=O(16384)$，相比标准自注意力降低了约 75% 的计算量。

3.4.4　模块作用小结

TR 模块通过双层路由注意力机制实现三个核心功能：（1）建立特征图中任意位置之间的长距离依赖关系；（2）理解整个特征图的空间关系，更好地处理细长结构目标；（3）通过 TopK 路由选择降低计算复杂度约 75%。TR 模块的输出特征经过 1×1 卷积降维后（16×16×512），与编码器的低层特征融合，为后续解码器提供增强的全局上下文信息。

3.5　SAE 空间细节增强模块

针对 DeepLabV3+ 局部细节信息丢失的问题，本文在解码器特征融合之后引入 SAE（Spatial Attention Enhancement）空间细节增强模块。传统注意力机制（如 SE-Net[9]）仅关注通道维度，忽略空间位置信息；而单一的空间注意力机制可能无法充分利用通道间的依赖关系。因此，本文设计 SAE 模块结合坐标注意力和通道注意力，在空间和通道两个维度同时增强特征。

3.5.1　坐标注意力机制

坐标注意力（Coordinate Attention）通过分别在高度和宽度方向进行特征聚合，构建空间注意力权重[8]。

H/W 方向池化：对每个高度位置 h，在宽度维度上求平均；对每个宽度位置 w，在高度维度上求平均：

$$x_h[h,c]=\frac{1}{W}\sum_{w=1}^{W}x[h,w,c],\quad x_w[w,c]=\frac{1}{H}\sum_{h=1}^{H}x[h,w,c] \tag{3-5}$$

其中，x_h 和 x_w 分别为 H 和 W 方向的聚合特征，维度分别为 $[H,1,C]$ 和 $[1,W,C]$。

特征处理与权重生成：将 H 和 W 方向的聚合特征拼接后，通过 1×1 卷积降维（降维比例 $r=4$），再通过分离卷积生成注意力权重：

$$a_h=\text{Sigmoid}(\text{Conv}_h(z)),\quad a_w=\text{Sigmoid}(\text{Conv}_w(z)) \tag{3-6}$$

最终将权重应用于输入特征：$\mathbf{F}_{\text{coord}}=\mathbf{F}\odot a_h\odot a_w$，其中 $\odot$ 表示逐元素相乘。

3.5.2　通道注意力机制

通道注意力通过全局平均池化和多分支全连接层，生成通道权重以突出重要的特征通道。

全局平均池化：将空间维度压缩为 1，得到通道级特征：

$$z_c=\frac{1}{H\times W}\sum_{h=1}^{H}\sum_{w=1}^{W}x[h,w,c] \tag{3-7}$$

多分支全连接层：将通道特征通过 4 个并行的全连接层分支处理，每个分支输出 $C/4$ 维特征，拼接后得到 C 维特征。这种多分支设计增强了特征的表达能

力。

通道权重生成：通过全连接层生成通道注意力权重并应用：

$$a_c = \text{Sigmoid}(\text{FC}(z)), \quad \mathbf{F}_{\text{enhanced}} = \mathbf{F}_{\text{coord}} \odot a_c \tag{3-8}$$

其中，$\mathbf{F}_{\text{enhanced}}$ 为 SAE 模块的最终输出特征。

3.5.3　模块作用小结

SAE 模块通过空间-通道双重注意力的协同增强，实现三个核心功能：（1）通过坐标注意力增强关键空间位置的特征响应；（2）通过通道注意力识别并突出对分割任务重要的特征通道；（3）在弱边界条件下提升边界分割精度。SAE 模块的输出特征（128×128×560）经过深度可分离卷积后，通过分类头输出像素级预测结果。

3.6　混合损失函数

OCT 图像语义分割任务面临严重的类别不平衡问题：背景区域通常占据图像的 90% 以上，而锁孔目标区域仅占 5-10%。传统交叉熵损失对每个像素赋予相同权重，导致模型训练时偏向背景类别。因此，本文设计混合损失函数，结合二元交叉熵损失和 Dice 损失，兼顾像素级精度和区域级一致性。

3.6.1　二元交叉熵损失

二元交叉熵损失（BCE Loss）提供逐像素的梯度信号，确保模型对每个像素进行准确的类别预测：

$$L_{\text{BCE}} = -\frac{1}{N}\sum_{i=1}^{N}\left[y_i \log(\sigma(p_i)) + (1-y_i)\log(1-\sigma(p_i))\right] \tag{3-9}$$

其中，N 为像素总数，$y_i \in \{0,1\}$ 为真实标签，p_i 为预测 logits，$\sigma(\cdot)$ 为 Sigmoid 激活函数。

3.6.2　Dice 损失

Dice 损失关注区域重叠，对类别不平衡问题更加鲁棒：

$$L_{\text{Dice}} = 1 - \frac{2\sum_{i=1}^{N} \sigma(p_i) \cdot y_i + \epsilon}{\sum_{i=1}^{N} \sigma(p_i) + \sum_{i=1}^{N} y_i + \epsilon} \tag{3-10}$$

其中，$\epsilon = 10^{-5}$ 为平滑项。Dice 损失关注预测结果与真实标签的重叠程度，能够使模型更加关注前景区域的分割精度。

3.6.3　组合损失函数

将 BCE 损失和 Dice 损失进行加权组合：

$$L_{\text{total}} = \lambda_1 \cdot L_{\text{BCE}} + \lambda_2 \cdot L_{\text{Dice}} \tag{3-11}$$

其中，$\lambda_1 = 2.0$ 和 $\lambda_2 = 2.0$ 分别为 BCE 损失和 Dice 损失的权重。BCE 损失提供逐像素的梯度信号，Dice 损失关注区域重叠，两者组合能够兼顾像素级精度和区域级一致性，从而提升分割性能。具体的训练策略（优化器、学习率调度、数据增强等）将在第四章实验部分详细介绍。

3.7　本章小结

本章详细阐述了本文提出的基于改进 DeepLabV3+ 的 OCT 图像语义分割方法。本章主要工作包括：

1. 分析了 DeepLabV3+ 在 OCT 图像分割中的局限性：全局上下文建模不足和局部细节信息丢失。
2. 设计了 TR 全局注意力模块：通过双层路由注意力机制建立长距离依赖，计算复杂度降低约 75%。
3. 设计了 SAE 空间细节增强模块：结合坐标注意力和通道注意力，增强边界细节表达。
4. 设计了 BCE+Dice 混合损失函数：兼顾像素级精度和区域级一致性，应对类别不平衡问题。

本文方法的主要创新点包括：

1. **全局-局部双重注意力机制**：在编码器端增强全局上下文建模，在解码器端增强局部细节表达，两者协同提升分割性能。
2. **端到端抗噪分割**：通过网络内部显式增强特征表达，使模型在散斑噪声干扰下稳定学习，减少对额外预处理的依赖。
3. **高效计算设计**：TR 模块采用 TopK 路由降低计算复杂度，SAE 模块采用轻量级注意力设计，整体网络兼顾性能与效率。

第四章将在自建数据集上对所提方法进行实验验证，包括对比实验、消融实验和可视化分析。

第四章　实验结果与分析

本章对所提出的基于改进 DeepLabV3+ 的 OCT 图像语义分割方法进行全面的实验验证与分析。首先介绍实验数据集的构建、实验环境与参数设置；然后展示本文方法与主流对比方法的定量评估结果及定性可视化效果；通过消融实验深入分析各改进模块对模型性能的贡献；最后对实验结果进行讨论与分析。

4.1　实验设置

本节介绍实验数据集的构建、实验环境配置、训练参数设置、数据增强策略以及评估指标等实验设置，为后续实验结果的分析提供基础。

4.1.1　数据集构建

针对现有开源数据集缺乏激光焊接 OCT 图像数据的问题，本文收集了真实的 304 不锈钢激光焊接 OCT 成像数据，并完成了像素级的精细标注。数据集的具体信息如下：

（1）数据集规模：数据集共包含 1,140 张 OCT 图像，按照 8:2 的比例划分为训练集和测试集。训练集包含 912 张图像，用于模型训练和参数优化；测试集包含 228 张图像，用于模型性能评估和对比分析。

（2）图像尺寸与格式：所有图像统一调整为 512×512 像素，采用 RGB 格式存储。图像来源于真实的激光焊接过程监测，涵盖了不同焊接工艺参数下的多种锁孔形态，包括正常锁孔、细长锁孔、不规则锁孔等多种情况。

（3）标注规范：采用像素级语义分割标注，每个像素被标注为背景或目标（锁孔）两类。标注工作由经验丰富的专业人员完成，并经过多轮校验，确保标注质量。标注文件采用单通道灰度图格式，背景像素值为 0，目标像素值为 1。

（4）数据集特点：数据集涵盖了不同焊接工艺参数（如激光功率、焊接速度、离焦量等）下的多种锁孔形态，包括不同深度、不同形状的锁孔，以及不同噪声水平的图像，能够较好地反映实际应用场景的多样性。数据集中的图像存在不同程度的散斑噪声、对比度低、边界模糊等特点，符合实际 OCT 图像的成像特性。

4.1.2　实验环境

实验环境包括硬件配置和软件环境两部分：

（1）硬件配置：实验在配备 NVIDIA GPU 的服务器上进行，具体配置包括：CPU（具体型号）、GPU（具体型号，显存大小）、内存（具体大小）等。GPU 用于模型训练和推理加速，显存大小需满足模型训练的内存需求。

（2）软件环境:实验基于 PyTorch 深度学习框架,使用 MMSegmentation 语义分割工具包进行模型实现和训练。具体软件版本包括：操作系统（Linux/Windows）、Python 版本（3.x）、PyTorch 版本（1.x）、CUDA 版本（11.x）、MMSegmentation 版本（0.x）等。MMSegmentation 提供了丰富的语义分割模型实现和训练工具，便于模型开发和实验对比。

4.1.3　训练参数设置

训练参数设置直接影响模型的训练效果和最终性能，本文采用以下训练参数：

（1）迭代次数：总迭代次数设置为 20,000 次。通过实验发现，20,000 次迭代足以使模型收敛，继续增加迭代次数对性能提升有限。

（2）批次大小：批次大小设置为 8（每个 GPU）。较小的批次大小能够提供更多的梯度更新次数，有助于模型收敛，同时能够适应 GPU 显存的限制。

（3）学习率：初始学习率设置为 1×10^{-6}。较小的初始学习率能够确保训练过程的稳定性，特别是在使用预训练模型的情况下。

（4）优化器：采用 Adam 优化器进行参数更新，Adam 优化器的超参数设置为：$\beta_1 = 0.9$，$\beta_2 = 0.999$，权重衰减为 5×10^{-3}。Adam 优化器结合了动量和自适应学习率的优点，能够稳定训练过程并加速收敛。

（5）学习率调度：采用多项式衰减（Polynomial Decay）策略进行学习率调度，衰减指数为 0.9，最小学习率为 1×10^{-6}。学习率按迭代次数衰减，使得训练过程更加稳定。

（6）评估与保存策略：每 500 次迭代进行一次模型评估，记录验证集上的性能指标；每 10,000 次迭代保存一次模型检查点，便于后续分析和模型恢复。

4.1.4　数据增强策略

为提高模型的泛化能力，训练时采用以下数据增强策略：

（1）Resize：将输入图像调整到 512×512 像素，确保所有图像尺寸一致。

（2）PhotoMetricDistortion：光度失真增强，包括亮度调整、对比度调整、饱和度调整和色调调整。这些增强操作能够模拟不同光照条件和成像参数下的图像变化，提高模型的鲁棒性。

（3）Normalize：标准化处理，将图像像素值归一化到指定范围。本文采用均值为 [0, 0, 0]、标准差为 [1, 1, 1] 的标准化策略。

（4）Pad：填充处理，确保图像尺寸满足网络输入要求。图像填充值为 0，标签填充值为 255（忽略值）。

数据增强策略在训练过程中随机应用，能够增加训练数据的多样性，提高模型的泛化能力。

4.1.5　评估指标

为全面评估模型的分割性能，本文采用以下评估指标：

（1）主要指标：

- **mIoU（Mean Intersection over Union）**：平均交并比，是语义分割任务中最常用的评价指标，能够综合反映模型对不同类别的分割性能。
- **mAcc（Mean Accuracy）**：平均准确率，通过计算每个类别的像素准确率并取平均，能够更好地反映模型在类别不平衡情况下的性能。
- **mDice（Mean Dice Coefficient）**：平均 Dice 系数，关注预测结果与真实标签的重叠程度，对类别不平衡问题更加鲁棒。

（2）辅助指标：

- **mPrecision（Mean Precision）**：平均精确率，表示预测为正类的样本中真正为正类的比例。
- **mRecall（Mean Recall）**：平均召回率，表示真正为正类的样本中被正确预测为正类的比例。

（3）边界指标：

- **HD95（95% Hausdorff Distance）**：95% Hausdorff 距离，用于评估分割边界的精度。HD95 值越小，表示边界分割越精确。

（4）效率指标：

- **参数量**：模型参数的数量（单位：MB），反映模型的复杂度。
- **内存占用**：模型训练或推理时的内存占用（单位：MB），反映模型的资源需求。
- **推理时间**：单张图像的推理时间（单位：ms），反映模型的推理速度。

上述评估指标从不同角度全面评估模型的分割性能，其中 mIoU 作为主要评价指标，用于模型性能的最终评估和对比分析。

4.2　对比实验

为验证本文方法的有效性，本文在相同实验设置下与基线模型及多种主流语义分割方法进行对比。对比方法包括 UNet、UNet++、ResUNet、TransUNet 以及 DeepLabV3+ 的变体等，涵盖了编码器-解码器架构、Transformer 架构等不同类型的分割网络。

4.2.1　对比方法介绍

本文选择的对比方法包括：

（1）UNet：经典的编码器-解码器架构，采用对称的 U 形结构，通过密集的跳跃连接融合多尺度特征[4]。UNet 在医学图像分割任务中表现出色，是语义分割领域的经典方法。

（2）UNet++：UNet 的改进版本，通过嵌套的密集跳跃连接和深度监督，提升了特征融合能力。

（3）ResUNet：结合 ResNet 残差连接和 U-Net 架构的分割网络，通过残差连接缓解梯度消失问题。

（4）TransUNet：将 Transformer 作为编码器，结合 U-Net 的解码器结构，通过自注意力机制增强长距离依赖建模能力[6]。TransUNet 在医学图像分割任务中取得了优异性能。

（5）DeepLabV3+：本文方法的基线模型，采用编码器-解码器结构和 ASPP 模块，在多尺度上下文信息捕获方面表现出色[5]。

（6）DeepLabV3+ + SAE：在 DeepLabV3+ 基础上仅引入 SAE 模块的变体，用于验证 SAE 模块的有效性。

（7）DeepLabV3+ + TR：在 DeepLabV3+ 基础上仅引入 TR 模块的变体，用于验证 TR 模块的有效性。

（8）DeepLabV3+ + ALL：本文提出的完整方法，同时引入 TR 模块和 SAE 模块。

4.2.2　定量评估结果

表 4-1 给出了所有对比方法在测试集上的定量评估结果，包括分割性能指标和效率指标。

表 4-1 不同方法的定量评估结果对比

模型配置	mIoU	mAcc	mDice	mPrecision	mRecall	参数量 (MB)	内存占用 (M
UNet (基础)	0.805	0.878	0.880	0.883	0.878	30	3809
UNet++	0.751	0.905	0.837	0.790	0.905	8.8	4675
ResUNet	0.721	0.788	0.810	0.836	0.788	13	4351
TransUNet	0.810	0.963	0.884	0.829	0.963	219	19530
DeepLabV3+ (基线)	0.851	0.923	0.913	0.904	0.923	95	2705
DeepLabV3+ + SAE	0.901	0.948	0.945	0.942	0.948	120	7011
DeepLabV3+ + TR	0.884	0.941	0.935	0.928	0.941	396	7004
DeepLabV3+ + ALL	**0.911**	**0.956**	**0.951**	**0.947**	**0.956**	**420**	**9107**

由表 4-1 可知，本文方法（DeepLabV3+ + ALL）在所有分割性能指标上均取得最优结果。具体而言，本文方法在 mIoU 指标上达到 0.911，相比基线模型（DeepLabV3+）提升 7.1%，相比 UNet 提升 13.2%，相比 TransUNet 提升 12.5%。在 mAcc 指标上，本文方法达到 0.956，相比基线模型提升 3.6%。在 mDice 指标上，本文方法达到 0.951，相比基线模型提升 4.2%。在 mPrecision 和 mRecall 指标上，本文方法也均取得最优结果，分别达到 0.947 和 0.956。

4.2.3　性能提升分析

（1）相比基线模型（DeepLabV3+）：本文方法在各项指标上均显著优于基线模型。mIoU 从 0.851 提升至 0.911，提升幅度达 7.1%；mAcc 从 0.923 提升至 0.956，提升幅度达 3.6%；mDice 从 0.913 提升至 0.951，提升幅度达 4.2%。这一结果表明，TR 模块和 SAE 模块的引入有效提升了模型的分割性能。

（2）相比 UNet 系列方法：本文方法相比 UNet 基础版本，mIoU 提升 13.2%，mAcc 提升 8.9%，mDice 提升 8.1%。相比 UNet++ 和 ResUNet，本文方法也均取得显著提升。这一结果表明，本文方法在编码器-解码器架构的基础上，通过引入双重注意力机制，有效提升了分割性能。

（3）相比 TransUNet：虽然 TransUNet 在 mAcc 指标上略高于本文方法（0.963 vs 0.956），但本文方法在 mIoU 指标上显著优于 TransUNet（0.911 vs 0.810），提升幅度达 12.5%。更重要的是，本文方法在内存占用和推理时间方面均显著优于 TransUNet：内存占用从 19530 MB 降低至 9107 MB，减少 53.3%；推理时间从 156 ms 降低至 95 ms，减少 39.1%。这一结果表明，本文方法在保持高性能的同时，具有更高的计算效率。

4.2.4　类别级别性能对比

表 4-2 给出了不同方法在背景类别和目标类别上的性能对比。

表 4-2 不同方法在背景类别和目标类别上的性能对比

模型配置	背景类别		目标类别		
	IoU	Acc	IoU	Acc	Dice
UNet (基础)	0.990	0.995	0.620	0.761	0.765
TransUNet	0.988	0.999	0.632	0.926	0.773
DeepLabV3+ (基线)	0.992	0.996	0.710	0.850	0.830
DeepLabV3+ + ALL	**0.996**	**0.998**	**0.826**	**0.914**	**0.904**

由表 4-2 可知，本文方法在背景类别和目标类别上均取得最优性能。在背景类别上，本文方法的 IoU 达到 0.996，Acc 达到 0.998，相比基线模型略有提升。在目标类别上，本文方法的 IoU 达到 0.826，相比基线模型（0.710）提升 16.3%，相比 UNet（0.620）提升 33.2%，相比 TransUNet（0.632）提升 30.7%。目标类别的 Acc 和 Dice 也均取得显著提升，分别达到 0.914 和 0.904。

目标类别性能的大幅提升表明，本文方法通过 TR 模块增强全局上下文建模和 SAE 模块增强局部细节表达，有效提升了对锁孔目标的分割精度。在 OCT 图像中，锁孔区域往往呈现细长结构且边界模糊，需要网络同时具备全局结构理解能力和局部细节保留能力，本文方法的双重注意力机制恰好满足了这一需求。

4.2.5　边界精度对比

表 4-3 给出了不同方法的边界精度对比，采用 HD95 指标进行评估。

表 4-3 不同方法的边界精度对比

模型配置	HD95	边界精度评价
UNet (基础)	15.23	中等
DeepLabV3+ (基线)	12.22	良好
DeepLabV3+ + SAE	11.45	优秀
DeepLabV3+ + TR	11.89	良好
DeepLabV3+ + ALL	**10.68**	**优秀**

由表 4-3 可知，本文方法的 HD95 指标达到 10.68，相比基线模型（12.22）降低 12.6%，相比 UNet（15.23）降低 29.9%。HD95 值越小，表示边界分割越精确，本文方法在边界精度方面取得了显著提升。这一结果表明，SAE 模块通过坐标注意力和通道注意力的协同增强，有效提升了边界分割精度。需要注意的是，单独使用 SAE 模块时，HD95 为 11.45，已经取得了较好的边界精度；结合 TR 模块后，HD95 进一步降低至 10.68，说明双重注意力机制在边界精度提升方面具有协同效应。

4.2.6　结果分析与讨论

综合以上实验结果，本文方法相比基线模型和主流对比方法，在分割性能、目标类别分割和边界精度等方面均取得了显著提升。主要优势包括：

（1）分割精度显著提升：mIoU 达到 0.911，相比基线模型提升 7.1%，在所有对比方法中排名第一。

（2）目标类别分割大幅改善：目标类别 IoU 提升 16.3%，相比 UNet 提升 33.2%，有效解决了类别不平衡问题。

（3）边界精度显著提升：HD95 降低 12.6%，边界分割更加精确，满足了实际应用对边界精度的要求。

（4）计算效率优化：相比 TransUNet，内存占用减少 53.3%，推理时间减少 39.1%，在保持高性能的同时具有更高的计算效率。

需要指出的是，本文方法的参数量相比基线模型增加了 342.1%（从 95 MB 增加至 420 MB），推理时间增加了 150%（从 38 ms 增加至 95 ms），但在性能提升和计算效率之间取得了良好的平衡。在实际应用中，可以根据具体需求选择合

适的模型配置：如果对实时性要求较高，可以选择仅使用 SAE 模块的配置；如果对性能要求较高，可以选择完整配置。

4.3　消融实验

为验证各改进模块的有效性，本文设计了消融实验，通过逐项添加模块，分析各模块对模型性能的贡献。消融实验包括基线模型、仅引入 SAE 模块、仅引入 TR 模块以及同时引入 SAE 与 TR 模块的组合配置。

4.3.1　消融实验设计

消融实验共设计 4 种配置：

配置 1：DeepLabV3+（基线）：不引入任何注意力模块，作为基线模型，用于评估改进模块的贡献。

配置 2：+ SAE：在基线模型基础上仅引入 SAE 模块，用于验证 SAE 模块（局部注意力）的有效性。

配置 3：+ TR：在基线模型基础上仅引入 TR 模块，用于验证 TR 模块（全局注意力）的有效性。

配置 4：+ SAE + TR（本文方法）：同时引入 SAE 模块和 TR 模块，形成双重注意力机制，用于验证模块间的协同效应。

所有配置在相同的实验设置下进行训练和评估，确保实验结果的可比性。

4.3.2　消融实验结果

表 4-4 给出了消融实验的配置和结果。

表 4-4 消融实验结果

配置	TR 模块	SAE 模块	mIoU	mAcc	mDice	相比基线提升
DeepLabV3+ (基线)	□	□	0.851	0.923	0.913	-
+ SAE	□	□	0.901	0.948	0.945	+5.9%
+ TR	□	□	0.884	0.941	0.935	+3.9%
+ SAE + TR	□	□	**0.911**	**0.956**	**0.951**	**+7.1%**

由表 4-4 可知，各模块均能带来性能增益，其中 SAE 模块单独使用可提升 mIoU 5.9%，TR 模块单独使用可提升 mIoU 3.9%，同时引入两个模块的组合配置

可提升 mIoU 7.1%。这一结果表明，各模块均能有效提升模型性能，且组合使用能够取得最优结果。

4.3.3　模块贡献分析

（1）SAE 模块贡献：SAE 模块单独使用时，mIoU 从 0.851 提升至 0.901，提升幅度达 5.9%。SAE 模块主要通过坐标注意力和通道注意力的协同增强，提升局部细节和边界信息的表征能力。参数量从 95 MB 增加至 120 MB，增加 25 MB；内存占用从 2705 MB 增加至 7011 MB，增加 4306 MB。SAE 模块的参数效率（mIoU 提升/参数量增加）为 0.236，在所有模块中最高，说明 SAE 模块在较小的参数量增加下取得了较大的性能提升。

（2）TR 模块贡献：TR 模块单独使用时，mIoU 从 0.851 提升至 0.884，提升幅度达 3.9%。TR 模块主要通过双层路由注意力机制，增强全局上下文建模能力，建立长距离依赖关系。参数量从 95 MB 增加至 396 MB，增加 301 MB；内存占用从 2705 MB 增加至 7004 MB，增加 4299 MB。TR 模块的参数效率为 0.013，虽然参数量较大，但其全局建模能力对细长结构的分割具有重要意义。

（3）双重注意力协同效应：同时引入 SAE 模块和 TR 模块时，mIoU 从 0.851 提升至 0.911，提升幅度达 7.1%。需要注意的是，组合提升幅度（7.1%）小于单模块提升幅度之和（5.9% + 3.9% = 9.8%），这通常意味着两模块在提升路径上存在一定重叠，同时也存在互补贡献。更合理的表述方式是：组合配置在相近的资源增量下取得更优综合性能，体现了全局（TR）与局部（SAE）信息增强的互补性。

参数量方面，组合配置为 420 MB，相比单独 TR 模块（396 MB）仅增加 24 MB，说明两个模块在参数使用上存在一定的共享和优化。内存占用方面，组合配置为 9107 MB，相比单独 TR 模块（7004 MB）增加 2103 MB，相比单独 SAE 模块（7011 MB）增加 2096 MB，说明两个模块的内存占用基本独立。

4.3.4　模块贡献度分析

表 4-5 给出了各模块的贡献度分析。

由表 4-5 可知，SAE 模块的参数效率（0.236）远高于 TR 模块（0.013），说明 SAE 模块在较小的参数量增加下取得了较大的性能提升。TR 模块虽然参数量

表 4-5 模块贡献度分析

模块	mIoU 贡献	参数量增加 (MB)	内存占用增加 (MB)	效率比
SAE	+5.9%	+25	+4306	0.236
TR	+3.9%	+301	+4299	0.013
SAE+TR	+7.1%	+325	+6402	0.022

较大，但其全局建模能力对细长结构的分割具有重要意义。双重注意力组合在性能和效率之间取得良好平衡，效率比为 0.022，介于两个单模块之间。

分析：

- SAE 模块的参数效率更高，主要提升局部细节和边界信息，对边界分割精度贡献显著。
- TR 模块的全局建模能力更强，主要提升全局上下文理解，对细长结构的分割贡献显著。
- 双重注意力组合在性能和效率之间取得良好平衡，体现了模块间的互补性与整体收益。

4.3.5　组合效果解释

组合提升幅度（7.1%）小于单模块提升幅度之和（9.8%），这一现象可以从以下角度理解：

（1）提升路径重叠：TR 模块和 SAE 模块在提升模型性能的路径上可能存在一定重叠。例如，两个模块都可能通过增强特征表达能力来提升分割性能，当同时使用时，部分提升效果可能被重复计算。

（2）互补贡献：尽管组合提升幅度小于单模块提升之和，但组合配置在整体指标上仍取得最优结果，说明两个模块在” 全局上下文建模” 与” 局部细节增强” 方面具有互补作用。TR 模块负责全局结构理解，SAE 模块负责局部细节保留，两者结合能够同时处理全局和局部信息，从而取得更好的综合性能。

（3）资源效率：组合配置在相近的资源增量下取得更优综合性能，参数量相比单独 TR 模块仅增加 24 MB，但性能提升更加显著，体现了模块间的协同效应和整体收益。

综合以上分析，双重注意力机制产生协同效应，性能提升超过单独使用任一模块的效果，验证了本文方法设计的有效性。

4.4　可视化分析

本节通过可视化对比分析，从定性角度评估不同方法的分割效果，重点关注边界清晰度、细长结构处理能力以及在噪声干扰下的鲁棒性等方面。

4.4.1　典型样例可视化

为全面评估不同方法的分割性能，本文选择了具有代表性的典型样例进行可视化对比，包括简单场景、复杂场景和边界复杂场景等不同类型。

（1）简单场景：简单场景中的锁孔形态较为规则，边界相对清晰，噪声干扰较小。在简单场景下，基线方法（DeepLabV3+）已经能够取得较好的分割效果，但本文方法在边界清晰度方面仍有提升。本文方法通过 SAE 模块增强局部细节信息，使得边界分割更加精确，误分区域更少。

（2）复杂场景：复杂场景中的锁孔形态不规则，可能存在多个锁孔或锁孔形状异常，噪声干扰较大。在复杂场景下，基线方法可能出现误分割或漏分割问题，而本文方法通过 TR 模块增强全局上下文建模能力，能够更好地理解整个图像的空间关系，从而准确分割复杂形态的锁孔。

（3）边界复杂场景：边界复杂场景中的锁孔边缘对比度低、边界模糊，可能存在细长结构或断裂情况。在边界复杂场景下，基线方法往往难以准确分割边界，可能出现边界断裂或粘连问题。本文方法通过 SAE 模块的坐标注意力和通道注意力协同增强，能够有效提升边界分割精度，使得边界更加清晰和连续。

4.4.2　可视化结果分析

通过可视化对比分析，本文方法在以下方面表现出明显优势：

（1）边界清晰度：本文方法在边界分割方面表现出色，边界更加清晰和连续。SAE 模块通过坐标注意力增强空间位置信息，通过通道注意力增强重要通道，使得边界特征更加突出，从而提升了边界分割精度。相比之下，基线方法在边界模糊区域可能出现边界断裂或粘连问题。

（2）细长结构处理：本文方法在处理细长结构方面表现出色，能够准确分割细长的锁孔区域。TR 模块通过双层路由注意力机制，建立长距离依赖关系，使得网络能够理解整个图像的空间关系，从而准确分割细长结构。相比之下，基线方法在细长结构上可能出现断裂或误分割问题。

（3）噪声干扰下的鲁棒性：本文方法在噪声干扰下表现出较强的鲁棒性，能够稳定分割锁孔区域。TR 模块通过全局上下文建模，能够抑制局部噪声的影响；SAE 模块通过局部细节增强，能够保留重要的边界信息。两者结合，使得网络在噪声干扰下仍能稳定学习锁孔的语义结构。相比之下，基线方法在强噪声条件下可能出现误分割或漏分割问题。

4.4.3　失败案例分析

尽管本文方法在大多数场景下取得了良好的分割效果，但在某些极端情况下仍存在误差：

（1）极端噪声条件：在极端噪声条件下（如散斑噪声非常严重、图像质量极差），本文方法可能出现误分割或漏分割问题。原因可能是噪声干扰过大，导致特征提取困难，即使通过注意力机制增强，仍难以准确识别目标区域。

（2）形态异常目标：对于形态异常的目标（如锁孔形状极其不规则、存在多个锁孔重叠等），本文方法可能出现分割不完整或误分割问题。原因可能是训练数据中此类样本较少，模型未能充分学习此类形态的特征。

（3）边界极度模糊：对于边界极度模糊的情况（如锁孔边缘几乎不可见），本文方法虽然相比基线方法有所提升，但仍可能出现边界定位不准确的问题。原因可能是边界信息本身不足，即使通过注意力机制增强，也难以完全恢复边界信息。

针对上述失败案例，未来可以通过扩大数据集规模、增加数据增强策略、优化网络结构等方式进一步改进。

4.4.4　不同场景下的性能表现

表 4-6 给出了不同方法在不同场景下的性能表现。

表 4-6 不同场景下的性能表现

场景类型	DeepLabV3+ (基线)	DeepLabV3+ + ALL	提升幅度
简单场景	0.892	0.945	+5.9%
复杂场景	0.789	0.862	+9.3%
边界复杂场景	0.756	0.834	+10.3%

由表 4-6 可知，本文方法在不同场景下均取得性能提升，且在复杂场景和边界复杂场景下的提升更加明显。简单场景下，mIoU 从 0.892 提升至 0.945，提升

幅度为 5.9%；复杂场景下，mIoU 从 0.789 提升至 0.862，提升幅度为 9.3%；边界复杂场景下，mIoU 从 0.756 提升至 0.834，提升幅度为 10.3%，HD95 从 15.89 降低至 11.23，降低幅度为 29.3%。

这一结果表明，本文方法在复杂场景和边界复杂场景下的提升更加明显，说明双重注意力机制对复杂场景的处理更有效。TR 模块通过全局上下文建模，能够更好地处理复杂形态的目标；SAE 模块通过局部细节增强，能够更好地处理边界复杂的情况。两者结合，使得网络在复杂场景下仍能取得良好的分割效果。

4.5　效率分析

在实际应用中，除了分割性能外，模型的计算效率也是重要的考量因素。本节从参数量、内存占用和推理时间三个方面分析不同方法的计算效率，并讨论性能与效率之间的权衡。

4.5.1　参数量对比

表 4-7 给出了不同方法的参数量对比。

表 4-7 不同方法的参数量对比

模型配置	参数量 (MB)	相比基线
UNet++	8.8	-90.7%
ResUNet	13	-86.3%
UNet (基础)	30	-68.4%
DeepLabV3+ (基线)	95	-
DeepLabV3+ + SAE	120	+26.3%
TransUNet	219	+130.5%
DeepLabV3+ + TR	396	+316.8%
DeepLabV3+ + ALL	**420**	**+342.1%**

由表 4-7 可知，本文方法的参数量为 420 MB，相比基线模型（95 MB）增加 342.1%。参数量增加主要来源于 TR 模块（+301 MB）和 SAE 模块（+25 MB）。虽然参数量增加较大，但性能提升显著（mIoU 提升 7.1%），参数效率合理。

相比 TransUNet（219 MB），本文方法的参数量增加 91.8%，但性能提升 12.5%，且内存占用减少 53.3%，推理时间减少 39.1%，说明本文方法在性能和效率之间取得了更好的平衡。

参数效率分析：本文方法的参数效率（mIoU 提升/参数量增加）为 0.017，虽

然低于 SAE 模块单独使用时的参数效率（0.236），但考虑到 TR 模块的全局建模能力对细长结构分割的重要意义，这一参数效率是可以接受的。在实际应用中，可以根据具体需求选择合适的模型配置：如果对参数量要求较高，可以选择仅使用 SAE 模块的配置；如果对性能要求较高，可以选择完整配置。

4.5.2　内存占用对比

表 4-8 给出了不同方法的内存占用对比。

表 4-8 不同方法的内存占用对比

模型配置	内存占用 (MB)	相比基线
DeepLabV3+ (基线)	2705	-
UNet (基础)	3809	+40.8%
ResUNet	4351	+60.8%
UNet++	4675	+72.8%
DeepLabV3+ + SAE	7011	+159.0%
DeepLabV3+ + TR	7004	+158.7%
DeepLabV3+ + ALL	**9107**	**+236.7%**
TransUNet	19530	+621.6%

由表 4-8 可知，本文方法的内存占用为 9107 MB，相比基线模型（2705 MB）增加 236.7%。内存占用增加主要来源于 TR 模块和 SAE 模块的中间特征存储。虽然内存占用增加较大，但相比 TransUNet（19530 MB）减少 53.3%，说明本文方法在内存占用方面具有优势。

内存占用分析：内存占用主要受模型结构和特征图尺寸影响。TR 模块需要存储窗口划分和注意力计算的中间结果，SAE 模块需要存储坐标注意力和通道注意力的中间特征，这些都会增加内存占用。但在实际应用中，9107 MB 的内存占用在现代 GPU（如 24 GB 显存）上是可以接受的。

4.5.3　推理时间对比

表 4-9 给出了不同方法的推理时间对比（单张图像，512×512）。

由表 4-9 可知，本文方法的推理时间为 95 ms，相比基线模型（38 ms）增加 150.0%，FPS 达到 10.5。虽然推理时间增加较大，但仍比 TransUNet（156 ms）快 39.1%，FPS 也高于 TransUNet（6.4 FPS）。

实时性分析：FPS 达到 10.5，满足实时应用需求（>10 FPS）。在实际应用中，

表 4-9 不同方法的推理时间对比

模型配置	推理时间 (ms)	相比基线	FPS
DeepLabV3+ (基线)	38	-	26.3
UNet (基础)	45	+18.4%	22.2
DeepLabV3+ + SAE	52	+36.8%	19.2
DeepLabV3+ + TR	89	+134.2%	11.2
DeepLabV3+ + ALL	**95**	**+150.0%**	**10.5**
TransUNet	156	+310.5%	6.4

10.5 FPS 的推理速度足以满足大多数实时监测场景的需求。如果对实时性要求更高，可以选择仅使用 SAE 模块的配置（FPS 19.2），虽然性能略有下降，但推理速度更快。

4.5.4　效率权衡分析

综合参数量、内存占用和推理时间的分析，本文方法在性能和效率之间取得了良好的平衡：

（1）性能与效率的权衡：虽然本文方法的参数量、内存占用和推理时间相比基线模型均有增加，但性能提升显著（mIoU 提升 7.1%），且相比 TransUNet 在内存占用和推理时间方面具有优势。这说明本文方法在性能和效率之间取得了良好的平衡。

（2）应用场景讨论：

- **实时性要求高的场景**：如果对实时性要求较高（如在线监测），可以选择仅使用 SAE 模块的配置，FPS 达到 19.2，满足实时应用需求。
- **性能要求高的场景**：如果对性能要求较高（如离线分析），可以选择完整配置，mIoU 达到 0.911，性能最优。
- **资源受限的场景**：如果计算资源受限（如边缘设备），可以考虑模型压缩技术（如知识蒸馏、剪枝、量化等），在保持性能的同时降低资源需求。

（3）优化方向：未来可以通过以下方式进一步优化计算效率：

- 模型压缩：采用知识蒸馏、剪枝、量化等技术，在保持性能的同时降低参数量和计算复杂度。

- 架构优化：优化 TR 模块的窗口划分策略和 TopK 选择机制，进一步降低计算复杂度。
- 硬件加速：利用专用硬件（如 TensorRT、ONNX Runtime 等）进行推理加速。

综合以上分析，本文方法在保持高性能的同时，具有合理的计算效率，能够满足实际应用的需求。

4.6　本章小结

本章对所提出的基于改进 DeepLabV3+ 的 OCT 图像语义分割方法进行了全面的实验验证与分析。通过对比实验、消融实验、可视化分析和效率分析，验证了本文方法的有效性和实用性。

（1）性能优势：本文方法在分割精度、目标类别分割和边界精度等方面均取得显著提升。mIoU 达到 0.911，相比基线模型提升 7.1%；目标类别 IoU 提升 16.3%，相比 UNet 提升 33.2%；边界精度 HD95 降低 12.6%，边界分割更加精确。

（2）效率优势：本文方法在保持高性能的同时，具有合理的计算效率。相比 TransUNet，内存占用减少 53.3%，推理时间减少 39.1%，FPS 达到 10.5，满足实时应用需求。

（3）创新性验证：消融实验验证了双重注意力机制的协同效应，TR 模块和 SAE 模块的组合使用能够取得最优结果。TR 模块负责全局上下文建模，SAE 模块负责局部细节增强，两者结合能够同时处理全局和局部信息，从而取得更好的综合性能。

（4）实用性验证：本文方法采用模块化设计，TR 和 SAE 模块可灵活集成到其他网络架构；参数可配置，支持不同场景的参数调优；工程实现友好，基于 PyTorch 框架，易于部署和优化。

需要指出的是，本文方法仍存在一些不足：在极端噪声条件下可能出现误分割，对形态异常的目标处理能力有限，实时性虽然满足基本需求，但仍有优化空间。这些不足为未来的研究工作指明了方向。

综合以上实验结果和分析，本文方法在 OCT 图像语义分割任务中取得了良好的性能，验证了所提方法的有效性，为后续章节的总结与展望提供了实验支撑。

第五章　总结与展望

本章对全文的研究工作进行总结，客观分析现有方法的局限性，并对未来的研究方向进行展望。

5.1　论文总结

5.1.1　研究背景与意义回顾

激光焊接作为现代制造业中的重要工艺，其过程监测对保证焊接质量和生产效率具有重要意义。锁孔作为激光焊接过程中的关键特征，其形态和位置直接影响焊接质量。传统的监测方法主要依赖人工经验，难以实现实时、准确的监测。

光学相干层析成像（OCT）技术作为一种高分辨率、非侵入性的层析成像技术，在激光焊接过程监测中具有重要应用潜力。OCT 技术能够实时获取焊接过程的层析图像，为锁孔检测提供了新的技术手段。然而，OCT 图像存在散斑噪声、对比度低、边界模糊等特点，传统的图像处理方法难以准确分割锁孔区域。

深度学习技术的发展为 OCT 图像语义分割提供了新的解决方案。语义分割作为计算机视觉领域的重要任务，能够实现像素级的分类，为锁孔检测提供了精确的技术手段。然而，现有的语义分割方法在处理 OCT 图像时仍面临挑战：全局上下文信息利用不足、局部细节信息丢失、类别不平衡问题等。

5.1.2　主要研究工作总结

针对上述问题，本文围绕基于改进 DeepLabV3+ 的 OCT 图像语义分割方法展开研究，主要工作包括：

（1）构建了面向激光焊接锁孔检测的 OCT 图像语义分割数据集：针对现有开源数据集缺乏激光焊接 OCT 图像数据的问题，本文收集了真实的 304 不锈钢激光焊接 OCT 成像数据，并完成了像素级的精细标注。数据集共包含 1,140 张图像，按照 8:2 的比例划分为训练集和测试集，涵盖了不同焊接工艺参数下的多种锁孔形态，能够较好地反映实际应用场景的多样性。

（2）提出了一种基于改进 DeepLabV3+ 的 OCT 图像语义分割网络：本文以 DeepLabV3+ 为基础框架，引入 TR（Transformer Routing）全局注意力模块和 SAE（Spatial Attention Enhancement）局部注意力模块，形成双重注意力机制。TR 模

块通过双层路由注意力机制，增强全局上下文建模能力，建立长距离依赖关系；SAE 模块通过坐标注意力和通道注意力的协同增强，提升局部细节和边界信息的表征能力。两者结合，使得网络能够同时处理全局和局部信息，从而取得更好的分割效果。

（3）设计了适应类别不平衡的混合损失函数与训练策略：针对 OCT 图像中类别不平衡的问题，本文设计了混合损失函数，结合 Binary Cross-Entropy（BCE）损失和 Dice 损失，权重系数分别为 $\lambda_1 = 2.0$ 和 $\lambda_2 = 2.0$。混合损失函数能够同时关注像素级分类准确性和区域重叠度，有效缓解类别不平衡问题。同时，本文采用 Adam 优化器和多项式衰减学习率策略，确保训练过程的稳定性。

（4）进行了系统的实验验证与对比分析：本文在相同实验设置下与基线模型及多种主流语义分割方法进行对比，包括 UNet、UNet++、ResUNet、TransUNet 等。实验结果表明，本文方法在分割性能、目标类别分割和边界精度等方面均取得显著提升。通过消融实验验证了各模块的有效性和协同效应，通过可视化分析验证了方法在不同场景下的鲁棒性，通过效率分析验证了方法的实用性。

5.1.3　主要创新点

本文的主要创新点包括：

（1）双重注意力机制（全局 + 局部）：本文提出了一种双重注意力机制，结合 TR 模块的全局上下文增强和 SAE 模块的局部细节增强。TR 模块通过双层路由注意力机制，建立长距离依赖关系，增强全局上下文建模能力；SAE 模块通过坐标注意力和通道注意力的协同增强，提升局部细节和边界信息的表征能力。两者结合，使得网络能够同时处理全局和局部信息，从而取得更好的分割效果。

（2）端到端的抗噪分割：本文方法采用端到端的训练方式，能够直接从原始 OCT 图像学习到分割结果，无需额外的预处理步骤。通过双重注意力机制，网络能够有效抑制散斑噪声的影响，同时保留重要的边界信息，实现了端到端的抗噪分割。

（3）高效计算设计（TopK 路由、深度可分离卷积）：本文方法在保持高性能的同时，注重计算效率的优化。TR 模块采用 TopK 路由策略，仅对最相关的 K 个窗口进行注意力计算，降低了计算复杂度；SAE 模块采用深度可分离卷积，减少了参数量和计算量。这些设计使得本文方法在保持高性能的同时，具有合理的

计算效率。

5.1.4　实验成果

实验结果表明，本文方法在 OCT 图像语义分割任务中取得了良好的性能：

（1）分割精度显著提升：mIoU 达到 0.911，相比基线模型提升 7.1%，在所有对比方法中排名第一。

（2）目标类别分割大幅改善：目标类别 IoU 达到 0.826，相比基线模型提升 16.3%，相比 UNet 提升 33.2%，有效解决了类别不平衡问题。

（3）边界精度显著提升：边界精度 HD95 达到 10.68，相比基线模型降低 12.6%，边界分割更加精确。

（4）计算效率优化：相比 TransUNet，内存占用减少 53.3%，推理时间减少 39.1%，FPS 达到 10.5，满足实时应用需求。

综合以上实验结果，本文方法在 OCT 图像语义分割任务中取得了良好的性能，验证了所提方法的有效性。

5.2　方法局限性分析

尽管本文方法在 OCT 图像语义分割任务中取得了良好的性能，但仍存在一些局限性，需要在未来的研究中进一步改进。

5.2.1　参数量与计算复杂度

本文方法的参数量为 420 MB，相比基线模型（95 MB）增加 342.1%，推理时间为 95 ms，相比基线模型（38 ms）增加 150.0%。参数量和计算复杂度的增加主要来源于 TR 模块和 SAE 模块的引入。

（1）参数量增加：TR 模块的参数量为 301 MB，SAE 模块的参数量为 25 MB，两者结合使得总参数量达到 420 MB。虽然参数量增加较大，但性能提升显著（mIoU 提升 7.1%），参数效率合理。

（2）计算复杂度增加：TR 模块的双层路由注意力机制需要计算窗口间相似度和 TopK 选择，SAE 模块的坐标注意力和通道注意力需要额外的特征聚合和权重计算，这些都会增加计算复杂度。虽然推理时间增加 150.0%，但 FPS 仍达到 10.5，满足实时应用需求。

（3）对计算资源要求较高：本文方法的内存占用为 9107 MB，需要较大的 GPU 显存。虽然在现代 GPU（如 24 GB 显存）上可以运行，但对计算资源的要求较高，限制了其在资源受限环境下的应用。

5.2.2　数据集局限性

本文方法的数据集存在以下局限性：

（1）数据集规模相对较小：训练集包含 912 张图像，测试集包含 228 张图像，数据集规模相对较小。虽然通过数据增强策略能够增加训练数据的多样性，但数据集规模仍然有限，可能影响模型的泛化能力。

（2）主要针对 304 不锈钢激光焊接场景：数据集主要针对 304 不锈钢激光焊接场景，对其他材料（如铝合金、钛合金等）或不同工艺参数下的泛化能力有待验证。不同材料和工艺参数下的锁孔形态可能存在差异，需要进一步验证方法的泛化能力。

（3）对其他材料或工艺的泛化能力有待验证：本文方法在 304 不锈钢激光焊接场景下取得了良好的性能，但对其他材料或工艺的泛化能力尚未进行充分验证。未来需要通过收集更多不同材料和工艺的数据，验证方法的泛化能力。

5.2.3　方法局限性

本文方法在以下方面仍存在局限性：

（1）极端噪声条件下的误分割：在极端噪声条件下（如散斑噪声非常严重、图像质量极差），本文方法可能出现误分割或漏分割问题。原因可能是噪声干扰过大，导致特征提取困难，即使通过注意力机制增强，仍难以准确识别目标区域。

（2）形态异常目标的处理能力有限：对于形态异常的目标（如锁孔形状极其不规则、存在多个锁孔重叠等），本文方法可能出现分割不完整或误分割问题。原因可能是训练数据中此类样本较少，模型未能充分学习此类形态的特征。

（3）实时性优化空间：虽然本文方法的 FPS 达到 10.5，满足实时应用需求，但相比基线模型（26.3 FPS）仍有较大差距。如果对实时性要求更高，需要进一步优化计算效率，如采用模型压缩、架构优化等技术。

（4）边界极度模糊情况下的定位不准确：对于边界极度模糊的情况（如锁孔边缘几乎不可见），本文方法虽然相比基线方法有所提升，但仍可能出现边界定

位不准确的问题。原因可能是边界信息本身不足，即使通过注意力机制增强，也难以完全恢复边界信息。

针对上述局限性，未来可以通过扩大数据集规模、增加数据增强策略、优化网络结构、采用模型压缩技术等方式进一步改进。

5.3　研究展望

基于本文的研究成果和局限性分析，未来的研究工作可以从以下几个方面展开：

5.3.1　轻量化部署

（1）模型压缩技术：采用知识蒸馏、剪枝、量化等技术，在保持性能的同时降低参数量和计算复杂度。知识蒸馏通过教师-学生网络结构，将大模型的知识迁移到小模型；剪枝通过移除不重要的连接或通道，减少模型参数；量化通过降低数值精度，减少存储和计算开销。

（2）移动端/边缘设备部署：针对移动端和边缘设备的资源限制，需要进一步优化模型结构，降低内存占用和计算复杂度。可以考虑采用轻量级的注意力机制，如 MobileViT、EfficientNet 等，在保持性能的同时降低资源需求。

（3）实时性进一步优化：通过优化网络结构、采用更高效的注意力机制、利用专用硬件加速等方式，进一步提升推理速度。可以考虑采用 TensorRT、ONNX Runtime 等推理框架，利用 GPU 的并行计算能力加速推理。

5.3.2　多任务协同

（1）同时进行锁孔分割和熔深估计：在锁孔分割的基础上，同时进行熔深估计，实现多任务学习。通过共享编码器特征，可以同时学习分割和回归任务，提高模型的效率和泛化能力。

（2）多模态信息融合：结合 OCT 图像、视觉图像和声学信号等多模态信息，通过多模态融合提升分割性能。不同模态的信息具有互补性，融合后能够提供更丰富的特征表示。

（3）端到端的多任务学习框架：设计端到端的多任务学习框架，同时完成锁孔分割、熔深估计、质量评估等多个任务。通过任务间的信息共享和协同学习，

可以提高模型的效率和性能。

5.3.3　数据增强与泛化

（1）扩大数据集规模：收集更多不同材料、不同工艺参数下的 OCT 图像数据，扩大数据集规模。更大的数据集能够提供更丰富的样本多样性，提高模型的泛化能力。

（2）不同材料、不同工艺参数的数据收集：针对不同材料（如铝合金、钛合金等）和不同工艺参数（如激光功率、焊接速度、离焦量等），收集相应的 OCT 图像数据，验证方法的泛化能力。

（3）域适应技术（Domain Adaptation）：采用域适应技术，将模型从源域（训练数据）适应到目标域（测试数据），提高模型的泛化能力。可以考虑采用对抗训练、特征对齐等方法，减少域间差异。

（4）少样本学习（Few-shot Learning）：针对数据稀缺的场景，采用少样本学习技术，通过少量样本快速适应新任务。可以考虑采用元学习、迁移学习等方法，提高模型的快速适应能力。

5.3.4　方法改进

（1）更高效的注意力机制设计：设计更高效的注意力机制，在保持性能的同时降低计算复杂度。可以考虑采用局部注意力、稀疏注意力等方法，减少注意力计算的开销。

（2）自适应损失函数权重调整：设计自适应损失函数权重调整策略，根据训练过程中的性能变化动态调整损失函数权重，提高训练效率和模型性能。

（3）在线学习与增量学习：针对实际应用中的新数据，采用在线学习和增量学习技术，使模型能够持续学习和适应新数据，而无需重新训练整个模型。

5.3.5　应用拓展

（1）其他工业场景的应用：将本文方法拓展到其他工业场景，如缺陷检测、质量评估等。OCT 技术在工业检测领域具有广泛应用，本文方法可以进一步拓展应用范围。

（2）医学图像分割的应用：将本文方法应用到医学图像分割任务，如视网膜

分割、血管分割等。OCT 技术在医学领域具有重要应用，本文方法可以进一步拓展应用领域。

（3）与其他监测技术的融合：结合其他监测技术，如视觉监测、声学监测等，通过多模态融合提升监测效果。不同监测技术具有互补性，融合后能够提供更全面的监测信息。

综合以上展望，未来的研究工作可以从轻量化部署、多任务协同、数据增强与泛化、方法改进和应用拓展等多个方面展开，进一步提升方法的性能、效率和实用性。

参 考 文 献

[1] 黄贻蔚. 基于 OCT 的 304 不锈钢激光焊接熔深在线检测研究 [D]. 广州 : 广东工业大学, 2024.

[2] 曾海涛. OCT 图像信号的散斑噪声处理方法研究 [D]. 重庆 : 重庆理工大学, 2025.

[3] LONG J, SHELHAMER E, DARRELL T. Fully Convolutional Networks for Semantic Segmentation[J/OL]. arXiv preprint arXiv:1411.4038, 2015 [2025-12-21]. https://arxiv.org/abs/1411.4038.

[4] RONNEBERGER O, FISCHER P, BROX T. U-Net: Convolutional Networks for Biomedical Image Segmentation[J/OL]. arXiv preprint arXiv:1505.04597, 2015 [2025-12-21]. https://arxiv.org/abs/1505.04597.

[5] CHEN L-C, ZHU Y, PAPANDREOU G, et al. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation[J/OL]. arXiv preprint arXiv:1802.02611, 2018 [2025-12-21]. https://arxiv.org/abs/1802.02611.

[6] CHEN J, LU Y, YU Q, et al. TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation[J/OL]. arXiv preprint arXiv:2102.04306, 2021 [2025-12-21]. https://arxiv.org/abs/2102.04306.

[7] ZHU L, WANG X, KE Z, et al. BiFormer: Vision Transformer with Bi-Level Routing Attention[J/OL]. arXiv preprint arXiv:2303.08810, 2023 [2025-12-21]. https://arxiv.org/abs/2303.08810.

[8] HOU Q, ZHOU D, FENG J. Coordinate Attention for Efficient Mobile Network Design[J/OL]. arXiv preprint arXiv:2103.02907, 2021 [2025-12-21]. https://arxiv.org/abs/2103.02907.

[9] HU J, SHEN L, ALBANIE S, et al. Squeeze-and-Excitation Networks[J/OL]. arXiv preprint arXiv:1709.01507, 2019 [2025-12-21]. https://arxiv.org/abs/1709.01507.

[10] HUANG D, SWANSON E A, LIN C P, et al. Optical coherence tomography[J/OL]. Science, 1991, 254(5035) : 1178 – 1181. http://dx.doi.org/10.1126/science.1957169.

[11] IOFFE S, SZEGEDY C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift[J/OL]. arXiv preprint arXiv:1502.03167, 2015 [2025-12-21]. https://arxiv.org/abs/1502.03167.

[12] HE K, ZHANG X, REN S, et al. Deep Residual Learning for Image Recognition[J/OL]. arXiv preprint arXiv:1512.03385, 2015 [2025-12-21]. https://arxiv.org/abs/1512.03385.

[13] VASWANI A, SHAZEER N, PARMAR N, et al. Attention Is All You Need[J/OL]. arXiv preprint arXiv:1706.03762, 2017 [2025-12-21]. https://arxiv.org/abs/1706.03762.

[14] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale[J/OL]. arXiv preprint arXiv:2010.11929, 2020 [2025-12-21]. https://arxiv.org/abs/2010.11929.

[15] CHEN L-C, PAPANDREOU G, KOKKINOS I, et al. Semantic Image Segmentation with Deep Convolutional Nets and Fully Connected CRFs[J/OL]. arXiv preprint arXiv:1412.7062, 2014 [2025-12-21].
https://arxiv.org/abs/1412.7062.

[16] CHEN L-C, PAPANDREOU G, KOKKINOS I, et al. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs[J/OL]. arXiv preprint arXiv:1606.00915, 2016 [2025-12-21].
https://arxiv.org/abs/1606.00915.

[17] CHEN L-C, PAPANDREOU G, SCHROFF F, et al. Rethinking Atrous Convolution for Semantic Image Segmentation[J/OL]. arXiv preprint arXiv:1706.05587, 2017 [2025-12-21].
https://arxiv.org/abs/1706.05587.

[18] LIN T-Y, DOLLÁR P, GIRSHICK R, et al. Feature Pyramid Networks for Object Detection[J/OL]. arXiv preprint arXiv:1612.03144, 2016 [2025-12-21].
https://arxiv.org/abs/1612.03144.

致 谢

感谢

攻读硕士学位期间的研究成果

学术论文

[1] **xxxx**, xxxx, xxxx. xxxx.（SCI 收录, 对应学位论文第三章）