

重庆邮电大学  
CHONGQING UNIVERSITY OF POSTS AND TELECOMMUNICATIONS

硕士学位论文  
DOCTORAL DISSERTATION



论文题目      基于 SAR 引导的光学图像  
                 去云方法研究

学科专业      计算机科学与技术

学      号      S230201040

作者姓名      胡美琳

指导教师      罗小波

学      院      计算机学院

学校代码 10617 UDC XXXXXX  
分 类 号 XXXXXX 密级 公开

# 学 位 论 文

基于 SAR 引导的光学图像去云方法研究

胡 美 珑

指导教师 罗小波 教 授

申请学位级别 硕士 学科专业 计算机科学与技术  
专业学位领域 计算机视觉  
答辩委员会主席 某某某 教授 论文答辩日期 2021 年 5 月 20 日  
学位授予单位和日期 重庆邮电大学 2021 年 6 月

# **Research on SAR-guided optical image cloud removal methods**

A Master Dissertation Submitted to  
Chongqing University of Posts and Telecommunications

Discipline	<b>Computer Science and Technology</b>
Student ID	<b>S230201040</b>
Author	<b>HU meilin</b>
Supervisor	<b>Prof. LUO Xiaobo</b>
School	<b>School of Computer Sciences and Technology</b>

## **重庆邮电大学**

### **学位论文独创性声明**

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。除文中已经注明引用的内容外，本论文中不包含其他个人或集体已经发表或撰写过的作品成果。对本文的研究做出重要贡献的个人和集体，均已在论文中以明确方式标明。本人完全知晓本声明的法律后果由本人承担。

学位论文作者签名：

日期： 年 月 日

## **重庆邮电大学**

### **学位论文使用授权书**

本人同意学校保留并向国家有关部门或机构送交论文的复印件和电子版，允许论文被查阅和借阅。

本学位论文属于：

公开论文

涉密论文，保密\_\_\_\_年，过保密期后适用本授权书。

(请在以上方框内选择打“√”)

作者签名：

导师签名：

日期： 年 月 日

## 摘要

光学遥感影像在对地观测与资源环境监测中具有重要应用价值，但受云遮挡及复杂大气条件影响，影像中往往存在大面积信息缺失，严重制约了其在精细化遥感应用中的可靠性与实用性。合成孔径雷达（Synthetic Aperture Radar, SAR）影像具备全天时、全天候成像能力，能够在云遮挡条件下稳定获取地表结构信息，为光学遥感影像云去除提供了重要的辅助信息来源。然而，SAR 与光学影像在成像机理、空间分辨特性及信息表达形式上的显著差异，使得多模态信息的有效协同与融合仍面临较大挑战。

针对上述问题，本文围绕 SAR 和光学协同的多模态遥感影像云去除任务，系统分析了遥感影像退化机理及信息缺失特性，提出了一种基于 SAR 引导的遥感图像云去除网络 (SAR-Guided Dual-Branch Network for Cloud Removal, SGN-CR)。该网络采用双分支编码架构，分别对 SAR 影像与含云光学影像进行特征建模，并以 SAR 分支提取的结构信息作为先验，引导光学分支的特征学习过程。通过在特征提取阶段引入 SAR 引导的注意力调制机制，SGN-CR 能够在云遮挡区域有效强化结构感知能力，抑制云干扰对光学特征建模的负面影响。

在此基础上，本文进一步设计了分层协同的跨模态特征融合策略，实现浅层结构约束与深层语义互补的有机结合，从而提升厚云遮挡条件下光学影像重建的稳定性与结构一致性。基于公开遥感数据集开展的大量对比实验结果表明，所提出的 SGN-CR 方法在峰值信噪比、结构相似性及光谱一致性等多项定量评价指标上均优于现有主流方法，尤其在大面积云遮挡场景下表现出更强的结构恢复能力与鲁棒性。进一步的消融实验证明了各关键模块设计的有效性。针对实际应用中对计算效率的需求，本文还提出了一种轻量化改进模型 Lite-SGN-CR，在显著降低模型参数量与计算复杂度的同时，保持了较为稳定的云去除性能。

**关键词：**遥感影像，云去除，SAR-光学多模态，结构引导，深度学习，轻量化

## ABSTRACT

Optical remote sensing imagery has significant application value in Earth observation and resource and environmental monitoring. However, due to cloud cover and complex atmospheric conditions, large areas of information are often missing in these images, severely limiting their reliability and practicality in refined remote sensing applications. Synthetic Aperture Radar (SAR) imagery possesses all-weather, day-and-night imaging capabilities, and can stably acquire surface structure information even under cloud cover conditions, providing an important auxiliary information source for cloud removal from optical remote sensing imagery. However, the significant differences between SAR and optical imagery in imaging mechanisms, spatial resolution characteristics, and information representation formats mean that the effective coordination and fusion of multimodal information still faces considerable challenges.

To address these issues, this paper focuses on the task of cloud removal from multimodal remote sensing imagery using SAR and optical co-processing. It systematically analyzes the degradation mechanism and information loss characteristics of remote sensing images and proposes a SAR-Guided Dual-Branch Network for Cloud Removal (SGN-CR). This network employs a dual-branch coding architecture, performing feature modeling on SAR and cloud-covered optical images separately. The structural information extracted from the SAR branch serves as prior knowledge, guiding the feature learning process of the optical branch. By introducing a SAR-guided attention modulation mechanism during feature extraction, SGN-CR effectively enhances structure perception capabilities in cloud-covered areas, suppressing the negative impact of cloud interference on optical feature modeling.

Building upon this, this paper further designs a hierarchical collaborative cross-modal feature fusion strategy, achieving an organic combination of shallow structural constraints and deep semantic complementarity, thereby improving the stability and structural consistency of optical image reconstruction under thick cloud cover conditions. Extensive comparative experiments based on publicly available remote sensing datasets demonstrate that the proposed SGN-CR method outperforms existing mainstream methods in multiple quantitative evaluation metrics, including peak signal-to-noise ratio, structural similarity, and spectral consistency, exhibiting stronger structural recovery capabilities and robustness, especially in large-area cloud-covered scenarios. Further ablation experiments vali-

## ABSTRACT

---

date the effectiveness of the key module design. To address the computational efficiency requirements in practical applications, this paper also proposes a lightweight improved model, Lite-SGN-CR, which significantly reduces the number of model parameters and computational complexity while maintaining relatively stable cloud removal performance.

**Keywords:** Remote sensing imagery, cloud removal, SAR-optical multimodal, structure-guided, deep learning, lightweight

## 目 录

摘 要 .....	I
ABSTRACT .....	II
图目录 .....	V
表目录 .....	VI
第 1 章 绪论 .....	1
1.1 研究背景 .....	1
1.2 国内外研究现状 .....	2
1.2.1 单时相光学遥感图像云去除方法 .....	2
1.2.2 多时相光学遥感图像云去除方法 .....	4
1.2.3 SAR 辅助光学遥感图像云去除方法 .....	5
1.3 论文研究主要内容 .....	7
1.4 论文组织结构 .....	8
第 2 章 相关理论基础 .....	9
2.1 本章引言 .....	9
2.2 光学遥感图像退化原理 .....	9
2.2.1 云遮挡对光学遥感成像影响 .....	9
2.2.2 大气传输过程对光学遥感图像成像影响 .....	10
2.3 SAR 图像成像原理与特性 .....	10
2.3.1 SAR 图像成像原理 .....	10
2.3.2 SAR 图像的相干斑噪声影响 .....	11
2.4 深度学习理论基础 .....	12
2.4.1 卷积神经网络 .....	12
2.4.2 注意力机制 .....	15
2.4.3 Transformer 理论基础 .....	16
2.4.4 CNN 与 Transformer 在图像领域的应用 .....	18
2.4.5 深度学习模型轻量化的理论基础 .....	19
2.4.6 遥感图像重建质量评价指标 .....	22
2.5 本章小结 .....	24
第 3 章 SAR 引导下的双分支去云网络设计 .....	25
3.1 本章引言 .....	25

3.2 研究方法 .....	25
3.2.1 SGN-CR 网络总体架构.....	25
3.2.2 SAR 引导的双分支特征编码结构 .....	28
3.2.3 SAR 引导的注意力调制机制（SGAM） .....	30
3.2.4 层级协同的跨模态特征融合策略.....	32
3.2.5 基于跨轴注意力的全局建模方法（CAA） .....	35
3.2.6 训练目标与损失函数设置 .....	36
3.3 实验环境与数据集说明 .....	38
3.3.1 数据集说明 .....	38
3.3.2 实验实现细节.....	38
3.3.3 对比方法设置.....	38
3.4 实验结果与分析.....	39
3.4.1 定量实验结果分析.....	39
3.4.2 可视化实验结果分析 .....	41
3.4.3 关键模块消融实验分析.....	42
3.4.4 SGN-CR 实验结果综合讨论.....	47
3.5 本章小结 .....	48
3.6 引用参考文献 .....	49
<b>第 4 章 基于 SGN-CR 的轻量化遥感图像云去除方法研究 .....</b>	<b>50</b>
4.1 本章引言 .....	50
4.2 SGN-CR 模型复杂度分析与轻量化动机 .....	51
4.2.1 SGN-CR 的整体复杂度分析.....	51
4.2.2 各功能模块的计算特性分析.....	52
4.2.3 小结.....	53
4.3 Lite-SGN-CR 轻量化网络设计 .....	54
4.3.1 Lite-SGN-CR 的整体架构设计 .....	54
4.3.2 基于深度可分离卷积的轻量化模块设计 .....	58
4.3.3 渐进式去云的推理增强策略.....	64
4.3.4 小结.....	65
4.4 实验结果与分析 .....	65
4.4.1 实验设置与评价指标 .....	65
4.4.2 Lite-SGN-CR 与现有方法综合对比 .....	67
4.4.3 性能-效率权衡分析.....	68
4.4.4 轻量化设计消融实验分析 .....	68

## 目 录

---

4.4.5 演进式去云策略性能分析 .....	69
4.5 本章小结 .....	69
<b>第 5 章 总结与展望 .....</b>	<b>71</b>
5.1 主要工作总结 .....	71
5.2 研究展望 .....	71
<b>参考文献 .....</b>	<b>73</b>
<b>作者简介 .....</b>	<b>77</b>
1. 基本情况 .....	77
2. 教育和工作经历 .....	77
3. 攻读学位期间的研究成果 .....	77
3.1 发表的学术论文和著作 .....	77
3.2 申请（授权）专利 .....	77
3.3 参与的科研项目及获奖 .....	77
<b>致 谢 .....</b>	<b>78</b>

## 图目录

图 3-1 SGN-CR 整体网络结构示意图 .....	26
图 3-2 Restore-block 结构示意图 .....	27
图 3-3 SAR-block 结构示意图 .....	29
图 3-4 Opt-block 结构示意图 .....	30
图 3-5 协同特征融合模块 .....	33
图 3-6 SEN12MS-CR 测试集上不同方法的云去除对比结果 .....	43
图 4-1 Lite-SGN-CR 整体网络结构示意图 .....	55
图 4-2 Lite-SAR-block 结构示意图 .....	60
图 4-3 Lite-CMCA 结构示意图 .....	61
图 4-4 Lite-restore-block 结构示意图 .....	63
图 4-5 SEN12MS-CR 测试集上不同方法的云去除对比结果 .....	69

## 表目录

表 3-1 不同云去除方法在 SEN12MS-CR 数据集上的定量对比结果 .....	40
表 3-2 SAR 引导注意力调制机制消融实验结果 .....	44
表 3-3 不同注意力机制与轴向计算策略消融实验结果 .....	45
表 3-4 不同特征融合策略组合的消融实验结果 .....	46
表 4-1 SGN-CR 整体复杂度统计 .....	51
表 4-2 SGN-CR 模块级复杂度分布 .....	52
表 4-3 Lite-SGN-CR 与不同模型重建性能对比 .....	67
表 4-4 Lite-SGN-CR 与不同模型复杂度与推理效率对比 .....	68
表 4-5 Lite-SGN-CR 与 SGN-CR 收益对比 .....	70

# 第1章 绪论

## 1.1 研究背景

遥感技术是通过传感器在非接触条件下获取地表信息的重要技术手段，其基本原理是利用地物对电磁波的反射、辐射或散射特性，对地表目标进行成像与分析。相较于传统地面观测方式，遥感技术具备覆盖范围广、信息获取效率高以及能够长期、连续观测等显著优势，已广泛应用于国土资源调查、生态环境监测、农业生产评估、灾害应急响应以及全球变化研究等多个领域。随着对地观测系统的持续发展，遥感影像在空间分辨率、光谱维度和时间分辨率等方面不断提升，高分辨率、多光谱乃至高光谱遥感影像逐渐成为支撑精细化遥感应用的重要数据基础。

在众多遥感成像方式中，光学遥感影像因其能够直接反映地表物质的光谱反射特性，在地物识别、分类与定量分析等任务中具有不可替代的优势。多光谱光学影像不仅包含丰富的空间结构信息，还能够在不同波段上刻画植被、水体、裸地和人工建筑等典型地物的光谱差异，是当前遥感应用中最为常用的数据类型之一。然而，这种高度依赖地表反射辐射信息的成像机制，使光学遥感成像过程对大气条件极为敏感。云层、薄雾以及云影等大气现象会显著干扰传感器对地表的有效观测，导致成像质量下降甚至影像失效。统计结果表明<sup>[4]</sup>，全球陆地区域中相当比例的光学遥感影像在不同程度上受到云及云影的覆盖，该现象在热带及季风气候区尤为突出。云遮挡不仅会造成地表信息的局部缺失，还会引入明显的光谱畸变和结构模糊，从而严重影响后续遥感解译与定量分析任务的可靠性。

从成像机理角度来看，云层对光学遥感影像可用性的影响主要体现在多个层面。首先，云具有较强的反射与散射能力，会直接遮挡传感器对地表的观测视线，使云覆盖区域内的真实地表辐射信息完全缺失，形成典型的“信息空洞”区域。其次，云层引起的复杂大气散射效应会对云边缘或薄云区域的像素光谱造成污染，表现为亮度异常、色彩偏移或对比度降低，从而破坏地表物质原有的光谱一致性。此外，云影的存在会在地表形成非均匀的辐射衰减区域，进一步加剧影像内部的光照不一致性。这些因素共同作用，使受云影响的光学遥感影像在空间结构表达和光谱特性表征方面均出现显著失真，难以直接满足定量遥感分析与精细化应用的需求。

从遥感应用层面来看，云遮挡问题的影响并不仅限于单幅影像质量的下降，还会对多时相分析、变化检测以及长时间序列建模等任务产生连锁效应。在高云覆盖区域，能够满足应用需求的无云光学影像获取周期往往显著延长，甚至在关键时间窗口内完全缺失，从而削弱遥感数据在农业监测、灾害评估等时效性要求较高场景中的应用价值。因此，在受云污染的观测条件下实现地表信息的有效恢复，

已成为提升光学遥感数据利用率和保障下游应用可靠性的基础性问题。

遥感图像云去除任务正是在上述背景下提出，其核心目标是在不依赖额外物理观测条件的情况下，对受云遮挡的光学影像进行信息重建。从问题本质上看，该任务属于一种高度不确定的信息缺失图像重建问题，其难点在于云覆盖区域内的真实地表信息往往被完全遮蔽，模型需要在缺乏直接观测的条件下推断合理的空间结构与光谱分布。尤其在厚云或大范围连续云覆盖场景中，地表纹理、边缘结构以及光谱分布均遭到严重破坏，传统基于局部统计或经验假设的方法难以获得稳定可靠的重建结果。这也使得云去除问题成为遥感图像处理领域中兼具理论挑战性和实际应用价值的重要研究方向。

随着遥感数据获取能力和应用需求的不断提升，云去除技术已不再仅服务于视觉层面的影像修复，而是逐渐演变为变化检测、时序分析和精细地物识别等高层遥感应用的关键前置环节。这对云去除结果在空间结构完整性、光谱保真性以及整体一致性等方面提出了更为严格的要求，也推动相关研究从传统的经验模型方法逐步向数据驱动的智能重建方法发展。

## 1.2 国内外研究现状

### 1.2.1 单时相光学遥感图像云去除方法

单时相光学遥感图像云去除方法是仅依赖单幅受云污染影像本身所包含的信息来进行地表重建，不引入额外的时间序列或多源辅助数据。由于其数据获取成本低、处理流程相对简洁，该类方法在早期遥感图像云去除研究中占据重要地位，至今仍在轻薄云覆盖场景下具有一定应用价值。根据方法建模思想的不同，单时相云去除技术大致可分为基于物理模型与先验假设的传统方法，以及基于深度学习的数据驱动方法。

#### (1) 基于物理模型与先验假设的传统方法

传统单时相云去除方法通常基于对成像机理或统计特性的先验假设，认为云层与地表在空间分布、频率特性或物理成像参数上存在可区分的差异。典型方法包括空间域插值与相似像元替换方法、频率域滤波方法以及大气散射物理模型方法。

空间域方法<sup>[5-7]</sup>假设云覆盖区域与其邻域的云自由区域在统计特性或纹理结构上具有相似性，并且这种相似性有助于估计缺失数据以产生视觉上一致的无云图像，通过插值、回归或相似像元替换的方式对缺失区域进行估计。该类方法实现简单，但仅限于去除斑点云，在大范围连续云覆盖或复杂地物场景下，重建结果往往偏离真实地表语义。

频率域方法则利用薄云在低频成分中占优的特性，通过设计低通或多尺度滤

波器分离云层与地表信息<sup>[8]</sup>。然而，该类方法对滤波参数高度敏感，不合适的频率可能导致无云区域中原始低频信息的丢失，引入细节损失或产生过度平滑现象。

基于物理模型的方法多借鉴大气散射模型，显式建模云层对辐射传输过程的影响，通过估计透射率和大气光参数恢复无云影像。通过精确估计和调整大气参数，这些方法可以有效地去除薄云。例如，基于先验的方法基于先验假设来估计透射图通过全局大气散射模型来恢复无云图像<sup>[9]</sup>。尽管该类方法具有较强的物理可解释性，但其有效性高度依赖于模型假设的准确性，在非均匀云分布或厚云场景下往往难以满足建模前提。

总体而言，传统方法在处理均匀薄云场景时具有一定优势，但受限于先验假设的表达能力，其在复杂云结构和语义重建方面存在明显瓶颈。

## (2) 基于深度学习的单时相方法

随着深度学习技术的发展，随着深度学习的不断进步，基于它的单图像去云已经获得了广阔的前景<sup>[10]</sup>。研究者开始利用卷积神经网络（CNN）直接学习受云影像与无云影像之间的非线性映射关系。CNN 通过多层特征提取能够有效建模局部纹理和光谱模式，在薄云去除任务中取得了显著进展。为增强特征表达能力，部分研究引入多尺度结构<sup>[11,12]</sup>、残差连接<sup>[13]</sup> 和注意力机制<sup>[14]</sup>，以提升细节恢复效果。

在此基础上，生成对抗网络（GAN）被引入单时相云去除任务中，通过对抗学习缓解对成对标注数据的依赖<sup>[15–18]</sup>。CycleGAN 等<sup>[19–21]</sup> 无监督框架通过循环一致性约束实现跨域映射，在一定程度上提升了真实场景下的泛化能力。然而，由于缺乏显式结构约束，GAN 类方法在大面积云遮挡区域仍容易产生伪影或色彩失真。

近年来，Transformer 架构凭借其全局建模能力被用于单时相去云任务<sup>[22]</sup>。自注意力机制有助于捕获长程依赖关系，缓解 CNN 感受野受限的问题。然而，该类方法通常伴随着较高的计算复杂度，在高分辨率遥感影像场景下对算力和存储资源提出了较高要求。

尽管深度学习方法显著提升了单时相云去除的表现，但其本质仍受制于单幅影像的信息上限。当厚云完全遮挡地表区域时，网络难以从输入中推断缺失的真实结构与语义信息。

## (3) 扩散模型方法

扩散模型作为近年来兴起的一类生成式模型，凭借其强大的分布建模能力，在图像恢复与生成任务中展现出优异性能。部分研究尝试将条件扩散过程引入单时相或弱辅助云去除场景，通过逐步去噪的方式生成无云影像。

相较于 GAN，扩散模型在生成稳定性和细节一致性方面具有一定优势，但其多步迭代采样机制导致推理效率较低，难以直接适配大尺度、高分辨率遥感影像

处理需求。此外，扩散模型对训练数据规模和计算资源的依赖较高，在工程部署和实时应用场景中难以真正应用。因此，尽管扩散模型在视觉生成任务中展现出潜力，但在高分辨率、多光谱遥感云去除任务中，其应用仍面临效率与稳定性方面的挑战。

综上所述，单时相云去除方法在无需额外数据的前提下具备良好的灵活性和适用性，但在厚云覆盖和复杂地物场景中，其重建能力受到信息缺失的根本限制。这一局限性促使研究逐步向引入多源辅助信息的方向发展。

### 1.2.2 多时相光学遥感图像云去除方法

为克服单时相光学影像在厚云覆盖区域中信息严重缺失的局限，研究者逐渐将注意力转向基于多时相光学影像的云去除方法。多时相遥感图像云去除方法利用同一区域在不同时间获取的多幅影像作为辅助信息<sup>[23]</sup>，通过挖掘时间维度上的互补特征来恢复被云遮挡的地表区域。相比单时相方法，多时相方法在厚云或大面积连续云覆盖场景下具备更强的信息恢复能力，是解决严重信息缺失问题的重要研究方向之一。

多时相云去除方法的核心思想是利用地表目标在短时间尺度内相对稳定的假设，通过对多时相影像进行配准与融合，在云覆盖区域引入来自其他时相的无云或低云观测信息。根据是否依赖云掩膜信息，多时相云去除方法通常可分为基于掩膜的非盲方法与无需显式掩膜的盲方法。

#### (1) 基于掩膜的多时相云去除方法

非盲多时相方法假设云覆盖区域可以通过人工标注或自动云检测算法准确获得，并利用云掩膜明确区分受损区域与有效区域<sup>[24,25]</sup>。早期研究多采用相似像元替换、字典学习或优化模型<sup>[?,26,28]</sup>，在多时相影像中搜索与云遮挡区域最相似的云自由块进行填充。

为缓解光谱差异和时间变化带来的影响，部分研究引入低秩建模、稀疏表示或张量分解方法，对多时相影像中的冗余信息进行建模，从而实现更稳健的缺失重建<sup>[29,30]</sup>。这些方法在假设地表变化较小的前提下，能够较好地保持光谱一致性和结构连续性。

然而，非盲方法对云掩膜精度高度敏感，而复杂云形态和云影干扰使得云检测在实际应用中难以完全可靠。此外，当多时相影像间存在显著地物变化时，基于历史信息的替换策略容易引入结构错位或语义失真。

#### (2) 多时相盲云去除方法

为降低对云掩膜的依赖，研究者提出了多时相盲云去除方法，将云检测与云去除过程统一建模<sup>[24]</sup>。部分模型驱动方法通过矩阵分解或低秩—稀疏分离，将云成

分视为稀疏噪声进行估计<sup>[31,32]</sup>。该类方法具有一定理论可解释性，但对参数选择和数据分布较为敏感<sup>[33]</sup>。

随着深度学习的发展，数据驱动的多时相方法逐渐成为主流。典型方法采用时序卷积网络或编码器—解码器结构，对多时相影像进行联合建模，以捕获时间维度上的相关性<sup>[35]</sup>。部分研究进一步引入时间注意力机制，以增强对关键无云时刻的选择能力。

尽管深度学习方法在多时相场景中展现出较强的重建能力，但其性能高度依赖于时序数据的完整性和时间分布。当可用时相数量有限或时间间隔过长时，模型容易受到地表变化和配准误差的影响<sup>[34]</sup>，导致重建结果不稳定。

### (3) 多时相方法的局限性分析

总体而言，多时相云去除方法通过引入时间冗余信息有效缓解了单时相信息不足的问题，在处理厚云和连续云覆盖方面具有明显优势。然而，该类方法在实际应用中仍面临多方面挑战：首先，多时相影像的获取受限于传感器重访周期，难以保证在关键时间节点获得高质量辅助数据；其次，地表变化、季节差异及成像条件不一致会破坏时序一致性假设；此外，多源时序数据的配准与预处理过程增加了系统复杂度。

上述问题在一定程度上限制了多时相方法在复杂场景下的稳定性与通用性，也促使研究逐步探索引入具有物理互补特性的多模态辅助信息，以进一步提升云去除的可靠性。

### 1.2.3 SAR 辅助光学遥感图像云去除方法

合成孔径雷达（Synthetic Aperture Radar, SAR）作为一种主动微波成像方式，能够在全天时、全天候条件下获取地表信息<sup>[36]</sup>，其成像过程几乎不受云层和光照条件的影响。这一独特优势使 SAR 数据在光学遥感图像云去除任务中具备重要的辅助价值<sup>[37]</sup>。研究者尝试通过引入 SAR 影像作为结构或纹理先验，来弥补光学影像在厚云遮挡条件下信息严重缺失的问题，从而显著提升云去除结果的完整性与稳定性。

需要指出的是，除将 SAR 作为辅助信息参与多模态融合外，部分研究也尝试仅依赖 SAR 数据直接恢复对应的光学影像，将该问题视为一种跨模态重建或图像翻译任务。此类方法通常基于深度生成模型学习 SAR 与光学影像之间的映射关系，在特定场景下能够生成具有一定视觉合理性的光学结果<sup>[44]</sup>。

然而，从成像机理角度来看，SAR 与光学影像在物理基础和信息表达形式上存在显著差异。SAR 成像主要反映地表的几何结构、粗糙度和介电特性，而光学影像则以地物的光谱反射特性为主，两者之间并不存在严格的一一对应关系。这

种模态不对称性使得仅凭 SAR 数据难以准确推断地表的真实光谱分布，尤其在植被覆盖、复杂地物混合或材质相近区域中，重建结果往往存在显著不确定性。

从实践角度看，基于 SAR 的单模态重建方法对训练数据分布高度敏感，其生成结果在跨区域或跨传感器场景下泛化能力有限。此外，SAR 影像中普遍存在的斑点噪声也容易在生成过程中被放大，进一步影响重建光学影像的稳定性和物理一致性。因此，单纯依赖 SAR 数据进行光学影像恢复在理论完备性和实际应用可靠性方面均存在明显局限。

基于上述分析，当前研究普遍倾向于将 SAR 数据作为结构先验或辅助信息，引导光学影像的云去除与重建过程，而非完全替代光学观测本身。这种多模态协同策略能够在保留光学影像光谱表达优势的同时，引入 SAR 提供的稳定结构信息，从而在复杂云遮挡场景下实现更为可靠的地表重建。

早期 SAR 辅助的光学云去除研究多采用基于生成模型的跨模态映射策略，将云去除问题视为 SAR 到光学影像的条件生成或翻译任务。典型方法包括条件生成对抗网络（cGAN）及其变体，通过对抗训练学习 SAR 与光学影像之间的非线性映射关系<sup>[38]</sup>。该类方法能够在一定程度上恢复被厚云遮挡的地表结构，在视觉效果上取得了较为理想的结果。

然而，基于生成式对抗学习的 SAR 辅助去云的方法通常将光学和 SAR 两种模态视为整体输入进行端到端映射，忽略了 SAR 与光学影像在成像机理、噪声分布以及信息表达形式上的本质差异。一方面，SAR 影像中普遍存在的斑点噪声容易在特征融合或生成过程中被放大，进而引入伪影；另一方面，纯生成式映射缺乏显式的结构约束，在复杂地物区域容易产生语义不一致或纹理失真现象。

因此部分研究引入注意力机制和多尺度特征融合策略，以增强关键区域的信息交互能力。例如，通过通道注意力或空间注意力引导网络重点关注云覆盖区域，从而提升云去除的针对性<sup>[39]</sup>。此外，一些方法尝试在网络中分别建模 SAR 与光学特征，再通过特征级融合实现跨模态协同<sup>[48]</sup>。

尽管上述方法在一定程度上提升了 SAR 辅助云去除的效果，但现有融合策略仍普遍存在两个方面的不足。首先，多数方法在特征提取阶段采用先分别并行的提取 SAR 和光学图像特征结构，再在深层进行跨模态的信息交互，也就是说通常发生在网络的中后期。而这就导致 SAR 提供的结构先验未能对光学特征的早期建模形成有效约束。其次，部分方法在融合过程中对 SAR 特征采取简单叠加或全局注意方式，缺乏对不同层级特征语义差异的针对性建模，导致浅层噪声干扰和深层语义补全之间难以平衡。

近年来，Transformer 结构被引入 SAR 辅助的光学云去除任务中，以增强跨模态全局依赖建模能力<sup>[40]</sup>。自注意力机制能够在一定程度上缓解局部感受野限制问

题，但其计算复杂度随特征分辨率呈平方增长，使得在高分辨率遥感影像场景中的应用面临显著的算力和存储压力。此外，纯基于全局注意的跨模态交互方式在缺乏显式结构引导的情况下，仍难以避免噪声信息的无差别传播。

综合来看，SAR 辅助光学的多模态云去除方法通过引入具有物理互补性的辅助信息，在厚云去除和结构恢复方面展现出明显优势。然而，如何在有效抑制 SAR 噪声干扰的同时，充分发挥其结构先验价值，并在不同语义层级实现合理的跨模态协同，仍是当前研究面临的关键问题<sup>[41]</sup>。这一问题也为后续基于结构引导的跨模态云去除模型设计提供了重要研究动机。

### 1.3 论文研究主要内容

针对光学遥感影像在云遮挡条件下面临的信息缺失与结构破坏问题，本文围绕 SAR 和光学多模态数据协同建模这一研究方向，系统开展遥感图像云去除方法研究。通过对现有单时相、多时相及多模态方法的分析可以发现，尽管引入辅助信息能够在一定程度上缓解云遮挡带来的影响，但在结构约束方式、跨模态协同机制以及模型实用性等方面仍存在不足。基于上述认识，本文的研究工作主要集中在以下三个方面：

(1) 针对单时相方法信息不足、多时相方法稳定性受限的问题，本文从遥感成像机理和信息互补性的角度出发，将 SAR 数据引入光学遥感图像云去除任务中，重点研究如何利用 SAR 在厚云条件下仍然稳定存在的结构信息，为光学影像重建提供有效约束。通过对 SAR 与光学影像差异特性的分析，本文探索一种以光学影像重建为目标、以 SAR 结构信息为辅助的多模态云去除建模思路，为后续方法设计奠定基础。

(2) 在多模态建模框架下，本文进一步关注跨模态信息如何在特征层面实现有效协同的问题。针对现有方法中 SAR 结构信息介入滞后、噪声干扰易传播以及不同层级特征语义差异未被充分利用等现象，本文从特征层级和信息属性的角度，对跨模态协同机制在云去除过程中的作用进行研究。通过分析不同层级特征在纹理抑制与结构补全中的功能差异，探讨在保证光谱一致性的前提下，引入结构先验以提升重建稳定性的有效策略。

(3) 考虑到多模态深度学习模型在参数规模和计算复杂度方面面临的实际约束，本文在上述研究基础上进一步开展模型效率优化相关工作。通过分析多模态云去除模型中不同功能模块的计算开销与性能贡献关系，探索在尽量保持云去除质量和结构恢复能力的前提下，降低模型复杂度的可行方法，为模型在资源受限场景下的实际应用提供支持。

综上，本文围绕 SAR 与光学多模态遥感图像云去除问题，从问题建模、跨模态协同机制分析以及模型实用性优化等层面展开研究，力求在复杂云覆盖场景下实现结构可靠、性能稳定且具有应用潜力的云去除方法。

## 1.4 论文组织结构

本文围绕遥感图像云去除问题，结合具有物理互补特性的多模态遥感观测数据，系统开展方法设计与应用研究。全文共分为五个章节，各章节内容安排如下。

第一章为绪论部分，主要介绍遥感图像云去除研究的背景与意义，对国内外相关研究现状进行系统梳理，并在此基础上明确本文的研究内容与总体研究思路，为后续章节奠定理论基础和研究框架。

第二章介绍本文研究所涉及的相关理论基础与关键技术，包括遥感成像基本原理、云遮挡对光学影像的影响特性以及深度学习在遥感图像重建中的应用，为后续方法设计提供必要的理论支撑。

第三章围绕多模态遥感图像云去除方法展开研究，重点针对厚云遮挡场景下光学影像结构信息严重缺失的问题，构建基于结构先验引导的云去除模型框架。通过对多模态特征协同建模与重建机制的系统分析，详细介绍模型整体设计、关键模块构成以及相应的实验验证结果，全面评估所提出方法在云去除精度和结构恢复方面的有效性。

第四章在第三章研究工作的基础上，进一步面向实际应用场景，对所提出模型进行轻量化与效率优化研究。通过对模型计算复杂度与性能表现的分析，探索在保证云去除效果的前提下，降低模型参数规模和计算开销的可行策略，并通过实验验证轻量化模型在性能与效率之间的平衡效果。

第五章是总结与展望。回顾了研究的核心工作，并对关键研究内容进行了总结。同时，还深入探讨了目前研究的局限性，并展望了未来研究的潜在发展路径

通过上述章节安排，本文从问题分析、方法设计到工程优化等多个层面，对多模态遥感图像云去除问题进行了系统研究。

## 第2章 相关理论基础

### 2.1 本章引言

遥感图像云去除任务涉及复杂的成像退化过程与信息缺失问题，其重建效果不仅受制于云与大气对光学成像的物理干扰机理，还与不同遥感数据源的信息表达特性及模型建模能力密切相关。为从理论层面支撑后续方法设计，有必要对云遮挡条件下的遥感图像退化特性、多源遥感数据的成像机理差异，以及深度学习模型在图像重建任务中的建模特性进行系统梳理。同时，明确遥感图像重建质量的评价依据，也是开展定量实验分析的重要前提。基于此，本章围绕遥感图像云去除所涉及的相关理论基础展开，为后文提出的方法模型的提出奠定理论基础。

### 2.2 光学遥感图像退化原理

在光学遥感成像过程中，传感器所获取的影像并非地表真实反射特性的直接记录，而是地表辐射信号在穿过云层与大气介质后形成的观测结果。当成像路径中存在云或复杂大气条件时，地表辐射信号在传播过程中会发生显著改变，导致观测影像与真实地表之间产生偏差。这种偏差在遥感影像中表现为信息退化，其结果不仅体现在图像质量下降，更重要的是造成地表信息在空间结构和光谱维度上的缺失。本节将从云遮挡和大气传输两个方面，对遥感图像退化产生的原因、作用机制及其带来的信息缺失后果进行系统分析。

#### 2.2.1 云遮挡对光学遥感成像影响

云遮挡是导致光学遥感图像退化的主要因素之一，其本质原因在于云层对电磁辐射传播路径的直接阻断。光学遥感成像依赖地表目标对太阳辐射的反射信号，而当云层位于传感器与地表之间时，地表反射辐射在传播过程中会被云体大量反射和散射，难以有效到达传感器。此时，传感器接收到的辐射信号主要来源于云体本身，而非真实地表目标。

(TODO: 传感器、云层、地表反射的图，以及云污染图像)

从成像机制角度看，云遮挡引发的退化并非简单的亮度变化，而是导致地表信息在观测层面被直接“替换”。在厚云覆盖区域，来自地表的辐射信号几乎完全被云层屏蔽，使得该区域在光学影像中不再包含任何有效的地表观测信息。这意味着，对于这些区域而言，观测影像与真实地表之间不再存在可逆映射关系，地表信息无法通过单幅受云影响的影像直接恢复。

这种退化对遥感图像的信息表达造成了两方面的后果。首先，在空间维度上，云遮挡破坏了地物原有的连续结构，使道路、河流和地物边界等几何特征在影像中出现断裂或消失。其次，在光谱维度上，不同地物之间原本稳定的多光谱反射关系被云体反射特性所掩盖，导致光谱信息整体失真。由于云覆盖通常具有空间非均匀性，不同区域的退化程度存在显著差异，这进一步增加了云去除过程中结构恢复和光谱重建的难度。

### 2.2.2 大气传输过程对光学遥感图像成像影响

除云遮挡外，大气传输过程也是造成遥感图像退化的重要因素。即使在无明显云覆盖的情况下，地表辐射信号在传播过程中仍需穿过由气体分子、气溶胶和水汽组成的大气介质。这些成分会对电磁辐射产生吸收和散射作用，从而导致辐射能量衰减以及信号特性的改变。

从物理机制上看，大气吸收会降低辐射信号的整体强度，而散射过程则会引入额外的背景辐射成分。这种背景辐射叠加在地表反射信号之上，使得传感器接收到的信号同时包含目标信息和大气干扰信息。由于不同波段对大气成分的敏感程度不同，各光谱通道受到的影响并不一致，从而破坏了多光谱影像中原有的光谱一致性。

与云遮挡造成的信息“直接缺失”不同，大气传输引发的退化更多表现为信息质量的下降和不确定性的增加。一方面，大气状态具有明显的时空变化特性，其影响难以通过统一参数进行精确描述；另一方面，大气效应与地表反射特性之间往往呈现复杂的非线性关系，使得观测影像与真实地表之间的映射关系更加难以建模。这种不确定性会削弱影像的对比度和细节表达能力，使得细粒度结构信息和微弱光谱差异难以被准确辨识。

综合来看，云遮挡和大气传输共同作用，导致光学遥感影像在空间结构和光谱特征两个层面同时发生退化。其中，云遮挡主要引发信息的不可逆缺失，而大气传输则进一步加剧信息衰减和建模不确定性。这种双重退化机制使遥感图像云去除任务面临显著挑战，也为后续重建方法在结构恢复可靠性和光谱保真性方面提出了更高要求。

## 2.3 SAR 图像成像原理与特性

### 2.3.1 SAR 图像成像原理

SAR 图像是一种基于微波主动发射与回波接收的成像系统。传感器向地表发射微波信号，并接收由地物散射返回的回波信息，通过合成孔径技术在飞行方向

上获得较高的空间分辨率。与依赖太阳辐射的光学成像不同，SAR 图像成像过程不受光照条件限制，且微波波段对云层和大气的穿透能力较强，因此能够在多云、阴雨等复杂气象条件下持续获取地表观测数据。

由于成像机理的不同，SAR 图像所表达的信息与光学影像存在显著差异。SAR 回波强度主要由地表目标的几何形态、电磁散射特性以及空间取向共同决定，而不是由地物的光谱反射特性主导。因此，SAR 图像对地表目标的表达更偏向于结构层面，尤其对具有明显几何形态的目标较为敏感。

在实际成像结果中，这种结构敏感性表现为对空间轮廓和连续形态的良好刻画。例如，道路、水体边界以及建筑物轮廓等具有明确几何结构的地物，在 SAR 图像中通常能够保持较好的空间连续性。即使在光学影像中被厚云完全遮挡的区域，这类结构信息在 SAR 图像中仍然往往是可观测的。这说明，相较于光学影像在云遮挡条件下面临的结构“信息缺失”，SAR 图像能够提供相对稳定的结构观测结果。

上述成像特性使 SAR 在云遮挡场景下具备显著优势：其能够在光学信息严重缺失的区域，为地表结构提供可靠的空间约束。这一优势为遥感图像云去除任务提供了新的信息来源，使得重建过程不再完全依赖光学影像中残存的局部纹理或统计先验。

然而，需要值得注意的是，SAR 图像并不直接包含光学影像所具备的光谱信息，其成像结果主要反映结构和物理属性，而非颜色和光谱特征。因此，SAR 更适合作为结构辅助信息，用于补充光学影像中缺失的空间结构，而难以单独完成光学影像的高质量重建。

(TODO: sar 和光学图像展示，以及 SAR 的噪声影响)

### 2.3.2 SAR 图像的相干斑噪声影响

尽管 SAR 在结构信息获取方面具有明显优势，但其成像过程不可避免地会引入相干斑噪声。相干斑噪声产生于微波在同一分辨单元内与多个散射体发生相干叠加，其结果是在图像中形成具有随机性的颗粒状纹理。这种噪声并非简单的加性干扰，而是一种与回波强度相关的乘性噪声，其统计特性较为复杂。

从成像结果来看，相干斑噪声在一定程度上增强了结构边缘的对比度，使部分轮廓更加明显，但同时也在局部区域引入了强烈的随机波动。这种随机性破坏了图像的平滑性和一致性，使得真实结构信息与噪声成分在空间尺度上相互混叠。在高分辨率遥感场景中，这一问题尤为突出。

在云去除等图像重建任务中，如果在缺乏有效约束的情况下将 SAR 图像直接引入光学影像重建过程，噪声成分可能被误识别为可靠结构信息，从而在重建结果

中引入不必要的纹理干扰，甚至影响光学影像原有的光谱一致性。此外，相干斑噪声的强度和分布受成像角度、地物类型及极化方式等因素影响，在不同场景中表现出较大的不稳定性，这进一步增加了 SAR 信息在跨模态利用过程中的不确定性。

因此，尽管 SAR 图像在云遮挡条件下能够提供稳定的结构观测能力，其噪声特性决定了该类信息不宜被无选择地使用。如何在充分利用 SAR 结构优势的同时，有效抑制相干斑噪声的负面影响，是多源遥感数据协同建模中必须重点考虑的问题。

## 2.4 深度学习理论基础

前述分析表明，在云遮挡条件下，光学遥感影像面临严重的信息缺失，而 SAR 图像虽能够提供稳定的结构观测能力，但其成像结果在信息表达形式和噪声特性上与光学影像存在显著差异。这种差异决定了云去除重建过程不仅需要补充缺失信息，还需要在多源信息之间实现有效选择与协同建模。因此，重建模型必须具备较强的非线性表达能力，能够在复杂退化条件下挖掘不同尺度、不同语义层级的有效特征。

深度学习模型凭借其端到端建模能力和层级化特征表达机制，为遥感图像重建任务提供了新的解决思路。然而，不同模型结构在特征建模范围、信息聚合方式以及计算复杂度等方面存在显著差异，其适用性和局限性有必要从理论层面进行分析。基于此，下面将从卷积神经网络、注意力机制以及 Transformer 等方面，对深度学习模型在图像重建任务中的建模特性进行系统阐述。

### 2.4.1 卷积神经网络

卷积神经网络（Convolutional Neural Network, CNN）是一类专门用于处理图像数据的深度学习模型，其设计目标是从原始像素空间中逐层提取具有判别能力的空间特征，并建立输入图像与目标输出之间的映射关系。由于图像具有显著的空间结构和局部相关性特征，CNN 在模型设计上显式引入了针对空间数据的建模假设，使其能够更高效地学习图像中的结构模式。与传统方法依赖人工设计特征不同，CNN 通过数据驱动的方式自动学习特征表示，在图像复原、修复与重建等低层视觉任务中展现出良好的适应性和泛化能力。

#### 2.4.1.1 卷积操作

在图像处理中，一幅图像可以看作由像素点组成的规则二维网格，相邻或局部像素之间通常具有明显的统计相关性和结构连续性。卷积神经网络的卷积操作，正是基于这一先验假设进行设计的。

卷积层是卷积神经网络中最基本的组成单元，其核心计算单元为卷积核。卷积核可以理解为一个尺寸较小的权值矩阵，用于在图像的局部区域内对像素进行加权组合。对一个图像进行卷积时，通过滑动图像上的卷积核，在每一个位置对局部像素进行加权求和，从而将原始像素映射为对应的局部特征响应。这一过程可以理解为在局部邻域内对图像结构进行感知和编码，使模型能够有效捕获边缘、纹理以及局部形态变化等基础视觉特征。

此外，同一个卷积核会在整幅图像上重复使用，卷积层对相同的局部结构模式在不同位置具有一致的响应能力，这种机制称为权值共享。权值共享一方面显著减少了模型参数数量，降低了过拟合风险；另一方面使模型在不同空间位置上对相同结构模式具有一致的响应能力，从而增强了模型对平移变化的适应性。这一特性使 CNN 在高分辨率图像处理中具备较高的计算效率和稳定性。

通过设置多个不同的卷积核，卷积层可以同时提取多种类型的局部空间特征，为后续特征表示提供基础。

(TODO: 卷积操作图)

#### 2.4.1.2 激活函数

卷积运算从本质上来看是一种线性变换，仅依赖卷积层中的线性加权运算，模型的表达能力将受到较大限制。为增强网络对复杂映射关系的建模能力，CNN 通常在卷积运算之后引入激活函数，对特征进行非线性变换。激活函数的作用在于打破线性叠加的限制、增强模型的表示能力，使网络能够逼近更加复杂的非线性函数。

在图像重建任务中，退化过程往往涉及噪声干扰、信息缺失以及非线性失真等因素。非线性变换的引入，不再局限于简单的线性滤波或局部平滑操作，而是能够学习从退化图像到目标图像之间更为复杂的映射关系，为恢复细节结构和真实纹理提供必要的建模能力。

在实际应用中，常见的激活函数包括 Sigmoid、Tanh 以及 ReLU 等。其中，ReLU 函数因其形式简单、计算效率高、在一定程度上缓解梯度消失问题，而在图像重建任务中得到广泛应用。通过在多层网络中引入激活函数，CNN 能够逐步学习从退化图像到目标图像之间的复杂非线性关系，而不再局限于线性滤波或简单平滑操作。

(TODO: 激活函数图)

#### 2.4.1.3 层级特征

卷积神经网络通过多层卷积与非线性变换的逐级堆叠，形成层级化的特征表示机制。网络中不同层的特征并非人为设定，而是在训练过程中由数据驱动逐步学习得到的，其语义层级与特征所对应的空间尺度密切相关。

在网络的浅层阶段，特征由卷积核在高空间分辨率特征图上提取，更关注局部空间变化。这类特征主要响应于边缘、局部对比度变化以及细微纹理等局部结构信息，能够较好地刻画地物边界和细节特征。随着网络向更深层推进，卷积运算通常伴随下采样操作，使特征图的空间分辨率逐步降低，同时每个特征响应所对应的输入区域范围不断扩大，即有效感受野逐渐增大。

通过下采样和多层卷积的共同作用，浅层提取的局部结构信息在更大空间范围内被逐步组合和抽象，形成对更复杂空间模式的响应，即深层特征。相较于浅层特征，深层特征对整体空间结构和上下文关系更加敏感，而对局部细节变化的依赖相对减弱。

在卷积神经网络中，空间分辨率的变化是实现多尺度特征建模的重要手段。通过对特征图进行上下采样操作，网络能够在不同空间尺度上对图像信息进行建模，从而在局部细节刻画与全局结构感知之间取得平衡。

下采样操作的主要作用是降低特征图的空间分辨率，以扩大单个特征响应所对应的输入区域范围，即有效感受野。通过在较低分辨率的特征图上进行卷积运算，模型能够在更大空间范围内聚合上下文信息，从而提升对整体结构和长程空间关系的建模能力。在卷积神经网络中，下采样通常通过池化操作或步幅卷积实现。

与下采样相对应，上采样操作用于在重建阶段逐步恢复特征图的空间分辨率，使输出结果与输入图像尺寸一致。上采样可以通过插值、反卷积或其他特征重排方式实现，其核心目标是在恢复空间分辨率的同时，尽可能保留已学习的结构和语义信息。然而，由于下采样阶段已不可避免地丢失部分高频细节信息，单纯依赖上采样操作往往难以完全恢复精细结构。

因此，在图像重建任务中，上下采样操作通常需要与跨层特征传递机制相结合，通过融合不同尺度、不同分辨率的特征信息，在保证整体结构一致性的同时增强细节恢复能力。这种空间尺度变换与特征重组机制，为卷积神经网络在复杂场景下进行有效图像重建提供了重要支撑。

对于遥感图像云去除任务而言，这种层级特征表示具有重要意义。浅层特征有助于保留地物边界和局部细节，而通过下采样获得的深层特征则能够在更大空间尺度上建模地物结构和上下文信息，为在大面积云遮挡区域进行合理推断提供支持。通过在重建过程中合理利用不同层级的特征信息，CNN 能够在细节保真性与

结构一致性之间取得平衡。

#### 2.4.1.4 局限性与发展需求

尽管卷积神经网络在局部特征提取和细节恢复方面具有显著优势，但其建模过程主要依赖局部邻域信息的逐层传播，难以显式刻画图像中远距离区域之间的依赖关系。在云遮挡较为严重的场景中，光学遥感影像中往往缺乏可靠的局部参考信息，此时仅依赖 CNN 的局部建模机制，难以对大尺度结构一致性进行有效约束。

这一局限性表明，传统卷积神经网络在复杂信息缺失重建任务中仍存在不足，有必要进一步引入能够建模更大范围上下文关系的机制，以提升模型在遥感图像云去除任务中的重建可靠性。

#### 2.4.2 注意力机制

注意力机制（Attention Mechanism）最初源于人类视觉与认知过程中的选择性关注行为，其核心思想是在大量信息中动态分配有限的建模能力，使模型能够优先关注对当前任务更为重要的部分。在深度学习框架下，注意力机制通过显式学习特征的重要性权重，对特征表示进行加权调节，从而实现对关键信息的突出建模与对冗余信息的抑制。这一机制为深度模型提供了一种超越固定结构计算的自适应特征选择能力。

在卷积神经网络中，特征提取主要依赖局部卷积运算和权值共享机制。尽管这种设计使模型在计算效率和局部结构建模方面具有显著优势，但也带来了固有的局限性。具体而言，标准卷积操作在同一层内对所有空间位置和特征通道采用统一的处理方式，默认各位置、各通道在特征表达中的重要性是等价的。这种“均匀建模”的假设在复杂视觉任务中往往并不成立，尤其是在存在遮挡、噪声或信息缺失的场景下，不同区域和不同特征对最终重建结果的贡献存在显著差异。

正是由于卷积操作难以显式建模这种特征重要性的差异，注意力机制被引入到卷积神经网络中，用以弥补其在特征选择层面的不足。通过在特征空间中学习一组自适应权重，注意力机制能够对原始特征进行重新分配，使模型在保持卷积结构优势的同时，具备对关键信息进行重点建模的能力。这一思想并不改变卷积的基本计算形式，而是在特征层面对卷积结果进行调制，因此具有良好的通用性和可插拔性。

根据建模维度的不同，现有注意力机制主要可以分为通道注意力和空间注意力两类。通道注意力侧重于刻画不同特征通道之间的重要性差异，其核心思想是通过建模通道间的相关性，增强对任务更有判别力的特征表示；空间注意力则关注特征图中不同空间位置的重要性分布，使模型能够更加聚焦于关键区域而忽略背

景或干扰区域。在实际应用中，这两类注意力机制常被组合使用，通过从通道维度和空间维度对特征进行联合建模，实现更全面的特征增强。

在此基础上，通道一空间联合注意力模块逐渐成为图像重建任务中的主流选择。其中，典型代表如 SE (Squeeze-and-Excitation) 模块主要通过通道维度建模特征重要性，而 CBAM (Convolutional Block Attention Module) 则进一步引入空间注意力，对特征进行顺序或并行的多维度加权。这类模块在结构上相对轻量，能够在不显著增加计算复杂度的前提下，有效提升特征表示的判别性，因此在图像复原与重建任务中得到广泛应用。

(TODO: 注意力模块图)

对于遥感图像云去除任务而言，注意力机制具有重要的理论意义。一方面，云遮挡导致图像中不同空间区域的信息完整性存在显著差异，注意力机制能够引导模型重点关注云覆盖区域及其结构边界，提高对关键缺失区域的建模能力；另一方面，在多源或多模态特征融合场景中，注意力机制可作为一种自适应调节手段，缓解不同特征之间信息分布不均或噪声干扰的问题，从而提升重建结果的稳定性与可靠性。然而，需要指出的是，注意力机制本质上仍然依附于局部特征建模框架，其作用主要体现在特征选择与增强层面，对建模范围的提升仍然有限。

#### 2.4.3 Transformer 理论基础

随着深度学习在视觉领域的不断发展，研究者逐渐认识到，仅依赖卷积神经网络的局部建模机制，在复杂场景下难以充分刻画图像中大范围区域之间的关联关系。尤其是在存在大面积信息缺失或结构推断需求的任务中，模型不仅需要感知局部纹理和边缘信息，还需要在更大空间范围内建立全局一致性的结构约束。在此背景下，Transformer 模型应运而生，其核心目标是通过显式建模长程依赖关系，突破传统卷积结构在建模范围上的固有局限。

Transformer 最初提出于自然语言处理领域，用以解决循环神经网络在自然语言处理中长序列建模中存在的效率低、依赖建模受限等问题。其关键思想在于完全摒弃递归结构，转而采用基于注意力机制的全局特征交互方式，使序列中任意位置的特征都可以直接建立联系。这一思想随后被引入到视觉任务中，并逐渐发展为一类以全局建模为核心优势的通用特征建模框架。

与卷积神经网络相比，Transformer 在建模机理上存在本质差异。卷积神经网络通过局部卷积与逐层传播实现特征交互，其建模过程具有明显的局部性和层级性；而 Transformer 则通过自注意力机制在单层结构中直接建立全局依赖关系，使特征之间的交互不再受限于空间距离。这种差异使 Transformer 在捕获长程结构关系和全局上下文信息方面具有天然优势，但同时也带来了计算复杂度和数据需求

方面的挑战。

Transformer 的核心组成单元是自注意力机制 (Self-Attention)。自注意力的基本思想是：对于特征序列中的每一个元素，模型通过计算其与序列中所有其他元素之间的相关性，自适应地聚合全局信息，从而生成新的特征表示。在这一过程中，输入特征首先被映射为查询 (Query, Q)、键 (Key, K) 和值 (Value, V) 三组表示。Q 用于刻画当前特征对其他特征的关注需求，K 用于描述各特征的属性，而 V 则承载被聚合的实际信息内容。通过计算 Q 与 K 之间的相似度，自注意力机制能够为不同位置分配不同的权重，并据此对 V 进行加权求和，实现全局范围内的信息交互。

(TODO 注意力机制图、多头、qkv)

为了进一步增强模型的表达能力，Transformer 通常采用多头注意力机制 (Multi-Head Attention)。多头注意力是在自注意力机制基础上的一种扩展，使模型能够同时关注来自不同表示子空间 (Representation Subspaces) 和不同位置的信息。具体而言，该机制将输入的 Q、K 和 V 通过线性变换映射到  $h$  个独立的并行 “头” (Head) 中，也就是将特征映射到不同的表示子空间，在每个子空间中分别计算注意力权重，从而获得多种互补的特征交互结果，捕获特定的上下文依赖关系。最终，所有头的输出被拼接并通过线性投影融合，从而生成包含丰富语义和结构信息的特征表示，这种并行建模机制使模型能够从不同角度理解特征之间的关系。例如，不同的注意力头可以分别侧重于局部结构关联、长距离依赖关系或不同语义层级的特征交互。通过对多头注意力输出进行融合，模型能够在保持全局建模能力的同时，提升对复杂空间结构和多样依赖模式的表达能力。对于图像重建等需要同时兼顾局部细节与整体结构一致性的任务而言，多头注意力为模型提供了一种更加灵活且有效的全局特征建模方式。

除自注意力模块外，Transformer 还包含若干关键组件以保证模型的稳定性和表达能力。其中，前馈网络用于对注意力输出进行非线性变换，增强特征表达能力；残差连接和归一化操作则用于缓解深层结构中的训练困难，保证梯度传播的稳定性。这些组件共同构成了 Transformer 的基本计算单元，使其能够在保持全局建模能力的同时实现高效训练。

在视觉任务中，Transformer 通常需要结合位置编码或结构约束，以弥补其对空间结构感知能力不足的问题。通过引入显式或隐式的位置信息，Transformer 能够在全局建模的基础上保留一定的空间结构感知能力，从而更好地适应图像数据的特性。这一改进使 Transformer 在图像分类、分割以及图像重建等任务中展现出良好的潜力。

对于遥感图像云去除任务而言，Transformer 的理论优势主要体现在其对大范

围上下文关系的建模能力。通过自注意力机制，模型能够在全局范围内整合结构信息，为大面积云遮挡区域的结构推断提供更强的约束。然而，由于 Transformer 在计算复杂度和数据依赖方面的特点，其在实际应用中往往需要与卷积结构或多尺度机制相结合，以在全局建模能力与计算效率之间取得平衡。上述特性为后续基于卷积与注意力协同设计的模型提供了重要的理论基础。

#### 2.4.4 CNN 与 Transformer 在图像领域的应用

在计算机视觉领域的发展过程中，卷积神经网络长期作为主流模型架构被广泛应用于各类图像理解与重建任务。其成功主要得益于卷积操作对局部空间结构的高效建模能力以及较为成熟的网络设计范式。通过多层卷积、下采样与特征融合，CNN 能够逐级提取从局部纹理到高层结构的多尺度特征表示，在图像分类、目标检测、语义分割以及图像复原等任务中取得了稳定且可靠的性能表现。尤其在图像重建类问题中，卷积神经网络凭借对局部连续性和纹理一致性的良好建模能力，成为早期研究中最常采用的基础模型。

然而，随着任务复杂度的提升，研究者逐渐认识到仅依赖局部卷积运算难以充分刻画图像中远距离区域之间的关联关系。在需要全局上下文建模或大尺度结构推断的视觉任务中，CNN 的逐层局部传播机制在建模效率和表达范围上均受到一定限制。为缓解这一问题，部分研究通过引入注意力机制、多尺度结构或更深层的网络设计，对卷积模型进行改进，但其全局建模能力仍然受限于卷积操作的局部性假设。

在此背景下，Transformer 架构开始被引入到计算机视觉领域。Transformer 最初在自然语言处理任务中取得显著成功，其核心思想是通过自注意力机制显式建模序列中任意元素之间的依赖关系，从而突破传统模型在建模范围上的限制。Vision Transformer (ViT) 是将 Transformer 架构直接应用于图像分类任务的代表性工作之一。该方法将输入图像划分为若干固定大小的图像块，并将其视作序列化的 Token 输入 Transformer 编码器进行处理，从而实现对整幅图像的全局建模。ViT 的提出表明，基于注意力机制的模型在视觉任务中同样具备较强的建模潜力。

在 ViT 之后，研究者针对 Transformer 在视觉任务中存在的计算复杂度高、对数据规模依赖较强以及空间结构建模能力不足等问题，提出了一系列改进模型。例如，引入局部窗口机制以降低计算开销、结合层级结构增强空间建模能力，或通过蒸馏与自监督策略提升训练效率。这些方法在保持 Transformer 全局建模思想的同时，对其结构形式进行了不同程度的调整，使其更好地适应图像数据的特点。

从整体发展趋势来看，卷积神经网络与 Transformer 在图像领域呈现出各自优势互补的特征。CNN 在局部结构建模、参数效率以及训练稳定性方面具有明显优

势，而 Transformer 在全局依赖建模和长程关系刻画方面表现突出。因此，近年来部分研究开始探索将卷积结构与 Transformer 机制相结合的混合建模思路，以在局部细节建模与全局结构感知之间取得平衡。这类方法通过在不同层级或不同模块中引入不同的建模机制，为复杂视觉任务提供了更加灵活的特征建模方式。

对于遥感图像云去除与重建任务而言，图像中往往同时存在局部纹理缺失与大范围结构不完整等问题，不同模型结构在应对这些挑战时各具特点。卷积神经网络和 Transformer 在图像领域的演化与应用，为理解不同建模机制在复杂场景下的适用性提供了有益参考，其具体效果仍需结合任务特性和模型设计进行进一步分析。

## 2.4.5 深度学习模型轻量化的理论基础

### 2.4.5.1 模型能力与复杂度

从理论角度看，深度学习模型的建模能力来源于其对输入数据进行多层非线性变换和特征组合的能力。通过增加网络深度、拓宽特征通道维度或引入更复杂的特征交互机制，模型能够在高维特征空间中表示更加复杂的映射关系，从而提升对复杂模式的拟合能力。然而，这种能力提升通常伴随着模型复杂度的显著增长，使得模型在实际应用中面临计算与资源层面的约束。

模型复杂度通常可以从两个层面进行理解。一方面，复杂度体现在模型的参数规模上，即网络中可学习参数的数量，它直接决定了模型在存储和内存占用方面的需求；另一方面，复杂度还体现在计算复杂度上，通常以模型在一次前向传播过程中所需的计算量来衡量，该指标直接影响模型的推理时延和能耗。在实际应用中，这两种复杂度往往相互关联，共同决定了模型在特定硬件条件下的可部署性。

随着模型规模的不断增长，模型能力的提升往往依赖于更深的网络结构和更高维度的特征表示。在卷积神经网络中，增加网络深度和通道数虽然能够提升特征表达能力，但同时会导致卷积运算次数成倍增加，尤其是在高分辨率输入条件下，特征图尺寸较大，计算代价的累积效应尤为明显。在基于注意力机制或 Transformer 的模型中，模型能力的提升通常依赖于更大范围的特征交互，这类全局或半全局建模方式在理论上具有更强的表达能力，但其计算复杂度往往随特征维度和空间规模快速增长，使得模型在资源受限场景下面临较大挑战。

进一步来看，模型复杂度的增长并不完全等价于有效建模能力的提升。在深层网络中，不同特征通道或不同层级之间可能存在较高的信息冗余，部分计算对最终任务的贡献有限，但仍然参与了完整的前向计算过程。这种冗余计算在一定程度上提高了模型的表达上限，却同时显著增加了计算负担，使模型在推理阶段效率下

降。此外，复杂度的过度增长还可能带来训练稳定性下降、模型过拟合风险增加以及对训练数据规模和硬件资源依赖加重等问题。

在遥感图像重建与云去除任务中，模型复杂度增长所带来的问题尤为突出。一方面，高分辨率遥感影像本身具有较大的空间尺寸和多通道特性，使模型在输入阶段即面临较高的计算压力；另一方面，为了应对复杂的退化机理和大范围信息缺失，模型往往需要具备较强的特征建模能力，从而进一步推高网络规模。这种对模型能力的需求与计算资源约束之间的矛盾，使得直接采用高复杂度模型在实际应用中存在一定局限性。

因此，在保证模型具备必要建模能力的前提下，如何控制模型复杂度、提升计算效率，成为模型设计中必须面对的问题。轻量化研究正是在这一背景下提出，其核心目标并非单纯压缩模型规模，而是通过识别模型中对任务贡献有限但计算代价较高的结构成分，对网络结构进行有针对性的优化，从而在性能与效率之间取得更加合理的平衡。这一思想对于遥感图像处理等对计算资源和推理效率具有较高要求的应用场景具有重要意义。

#### 2.4.5.2 常见轻量化网络结构设计思想

在明确模型复杂度主要来源于网络结构设计之后，轻量化网络的研究逐渐从简单的参数压缩，转向对模型结构与学习过程的系统性优化。从理论角度看，轻量化设计的核心目标并非削弱模型的建模能力，而是通过减少冗余计算或引入有效约束，使有限的计算资源更多地集中于对任务具有关键作用的特征建模过程。

一种具有代表性的轻量化思想是基于通道维度的压缩与重组。在卷积神经网络中，特征通道数直接决定了特征空间的维度规模。尽管增加通道数能够提升模型的表达能力，但在实际网络中，不同通道之间往往存在较高的信息冗余。基于这一观察，研究者提出通过通道压缩或重标定的方式，降低高维特征映射中的冗余计算，从而在减少参数量和计算量的同时，尽量保留关键特征信息。

另一类重要的轻量化设计思想围绕卷积计算结构本身展开。标准卷积在空间维度和通道维度上同时进行特征交互，这在高分辨率和高通道数条件下会带来显著的计算开销。通过对卷积操作进行分解、分组或重构计算路径，可以在不改变卷积基本建模假设的前提下，有效降低计算复杂度。这类方法通过约束特征交互方式，实现了对计算代价的结构级控制，是轻量化卷积网络设计中的经典思路。

在深层网络结构中，特征在不同层级之间往往具有较强的相关性，部分中间特征在多个阶段中被重复计算。基于这一现象，一些轻量化方法通过增强特征重用减少冗余计算，例如通过跨层连接或逐级特征精炼机制，使模型在保持网络深度的同时，避免对相似信息的反复建模，从而提升整体计算效率。

除结构级调整外，知识蒸馏为轻量化模型设计提供了一种从学习过程角度提升模型有效能力的策略。与通过压缩网络结构降低复杂度不同，知识蒸馏主要用于模型训练阶段，其核心思想是在训练轻量模型时，引入一个性能较强的教师模型，对学生模型的学习过程施加额外约束。在具体机制上，知识蒸馏通常通过引导学生模型模仿教师模型的输出行为或中间特征表示来实现。教师模型在训练完成后，能够在高维特征空间中形成更加平滑且具有判别性的表示，其输出结果不仅包含最终预测信息，还隐含了不同类别或不同特征之间的相对关系。通过将这些信息作为软目标提供给学生模型，蒸馏过程使学生模型不再仅依赖于真实标签进行学习，而是同时受到教师模型行为的引导。从特征学习角度看，轻量模型由于结构容量受限，在训练过程中往往难以稳定地探索高维特征空间，容易陷入局部最优或学习到不充分的特征表示。教师模型所提供的软约束能够显式引导学生模型关注对任务判别更为关键的特征模式，从而减少无效特征的学习，提高训练效率和特征利用率。

随着注意力机制和 Transformer 结构在视觉任务中的应用，轻量化研究也逐渐扩展到全局特征交互过程的复杂度控制。由于全局自注意力在理论上具有较高的计算复杂度，部分研究通过限制注意力计算的空间范围、降低特征维度或引入分层建模策略，对全局建模过程进行约束。这类方法在保留关键依赖关系建模能力的同时，显著降低了注意力机制带来的计算开销，使相关模型在实际应用中更具可行性。

总体而言，现有轻量化方法既包括针对网络结构的复杂度削减，也包括通过知识迁移和学习约束提升轻量模型有效能力的策略。不同轻量化思想在削减计算开销和保持模型性能方面各有侧重，其适用性需要结合具体任务特性和应用场景进行综合考量。

#### 2.4.5.3 精度与效率的权衡

在深度学习模型设计中，精度与效率是两个密切相关但往往相互制约的目标。精度通常用于衡量模型在特定任务上的预测或重建能力，在图像重建与云去除任务中，精度不仅体现在像素层面的误差大小，还反映在结构完整性、纹理一致性以及光谱保真性等多个方面。高精度模型通常能够更准确地恢复被遮挡区域的地物结构和细节特征，从而提高重建结果的可靠性。

效率则主要反映模型在实际运行过程中的资源消耗情况，通常包括推理阶段的计算量、模型参数规模、内存占用以及推理时延等因素。对于高分辨率遥感影像处理任务而言，效率直接决定了模型在有限计算资源条件下的可用性，尤其是在大规模数据处理或端侧部署场景中，模型的计算效率往往成为影响其实际应用价值

的重要因素。

在实际应用中，精度与效率往往同时被提出要求。一方面，云去除和图像重建任务对结果质量具有较高要求，过低的精度会导致结构失真或细节缺失，从而影响下游应用的可靠性；另一方面，遥感影像数据规模大、分辨率高，若模型计算复杂度过高，将难以满足实际应用中的时效性和资源约束。因此，单纯追求高精度或极致效率，均难以满足实际需求。

精度与效率之间的冲突主要源于模型能力与复杂度之间的内在关系。提升模型精度通常依赖于更强的特征表达能力，而这往往通过增加网络深度、拓宽特征通道或引入复杂的特征交互机制来实现，这些设计会直接导致计算量和参数规模的增加。相反，为提升模型效率而削减网络结构，往往会限制模型的表达能力，使其难以充分刻画复杂场景中的细节和结构，从而对重建精度产生不利影响。

因此，模型轻量化的核心并非在精度和效率之间进行简单取舍，而是在两者之间寻找合理平衡。通常的做法是在满足任务最低精度需求的前提下，尽可能降低模型复杂度。例如，可以优先保留对重建质量贡献较大的关键结构，削减冗余计算较多但对性能提升有限的部分；或通过引入辅助约束和训练策略，在不增加推理阶段复杂度的情况下，弥补因结构压缩带来的精度损失。这种以任务需求为导向的权衡策略，使模型在性能和效率之间达到可接受的折中状态。

在遥感图像云去除任务中，这种权衡尤为重要。云遮挡条件复杂、地物类型多样，对模型精度提出了较高要求；同时，遥感影像通常具有较大的空间尺度和数据规模，对模型效率也构成了现实约束。因此，轻量化模型的设计需要在保证基本重建质量和结构一致性的前提下，控制计算开销和模型规模，使模型能够在实际应用环境中稳定运行。上述权衡原则为后续轻量化模型的结构设计与实验分析提供了重要的理论指导。

#### 2.4.6 遥感图像重建质量评价指标

在遥感图像云去除与重建任务中，模型性能的优劣不仅体现在视觉效果上，还需要通过定量指标进行客观评估。由于云遮挡会导致像素缺失、结构破坏以及光谱信息失真，单一评价指标往往难以全面反映重建结果的质量。因此，合理选择和理解评价指标的物理含义与侧重点，是对不同方法进行公平比较和性能分析的重要前提。

为对遥感图像云去除与重建任务中的模型性能进行定量评估，通常需要从像素精度、结构一致性、光谱保真性以及误差幅度等多个角度对重建结果进行综合分析。基于上述考虑，本文采用峰值信噪比（Peak Signal-to-Noise Ratio, PSNR）、结构相似性指数（Structural Similarity Index Measure, SSIM）、光谱角映射（Spectral

Angle Mapper, SAM) 以及平均绝对误差 (Mean Absolute Error, MAE) 作为评价指标, 对去云结果的重建质量进行系统评估。

设  $x$  表示模型生成的去云光学影像,  $y$  表示对应的真实无云影像,  $n$  为图像中的像素总数, 并且实验中所有影像都归一化至  $[0, 1]$  区间时, 各评价指标的定义如下。

首先, PSNR 是图像质量评估中最常用的指标之一, 其数值越大表示重建结果在像素层面越接近真实影像。PSNR 基于均方根误差 (Root Mean Square Error, RMSE) 计算, PSNR 计算公式为:

$$PSNR(x, y) = 20 \cdot \log_{10} \left( \frac{1}{RMSE(x, y)} \right) \quad (2-1)$$

其中, RMSE 定义为:

$$RMSE(x, y) = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - y_i)^2} \quad (2-2)$$

为评估去云结果在感知层面的结构相似性, 本文进一步采用 SSIM 指标。该指标从亮度、对比度和结构三个方面衡量两幅图像之间的相似程度。设  $\mu_x$  和  $\mu_y$  分别表示图像  $x$  和  $y$  的均值,  $\sigma_x^2$  和  $\sigma_y^2$  表示方差,  $\sigma_{xy}$  表示协方差,  $\varepsilon_1$  与  $\varepsilon_2$  为防止分母为零而引入的常数, 则 SSIM 的计算公式为:

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + \varepsilon_1)(2\sigma_{xy} + \varepsilon_2)}{(\mu_x^2 + \mu_y^2 + \varepsilon_1)(\sigma_x^2 + \sigma_y^2 + \varepsilon_2)} \quad (2-3)$$

SSIM 的取值范围为  $[0, 1]$ , 其数值越大表示结构相似性越高。

考虑到遥感影像通常具有多光谱特性, 仅依赖像素误差和结构指标难以全面反映光谱保持能力, 本文引入 SAM 作为光谱一致性评价指标。SAM 通过计算预测光谱向量与真实光谱向量之间的夹角来衡量光谱形态的一致性, 其定义如下:

$$SAM(x, y) = \cos^{-1} \left( \frac{\sum_{i=1}^n x_i y_i}{\sqrt{\sum_{i=1}^n x_i^2 \cdot \sum_{i=1}^n y_i^2}} \right) \quad (2-4)$$

SAM 值越小, 表示光谱失真越小, 去云结果在光谱层面越接近真实影像。

此外, 本文采用 MAE 对像素级误差幅度进行补充评估。相较于均方误差, MAE 对异常值具有更好的鲁棒性, 其定义为:

$$MAE(x, y) = \frac{1}{n} \sum_{i=1}^n |x_i - y_i| \quad (2-5)$$

综合上述四项指标, 可以从像素精度、结构一致性以及光谱保真性等角度对去

云结果的重建质量进行全面评价，为后续实验结果分析提供评价依据。

## 2.5 本章小结

本章围绕遥感图像云去除任务所涉及的相关理论基础展开了系统阐述。从遥感图像与 SAR 图像特性出发，分析各自的优缺点。对深度学习模型中的卷积神经网络、注意力机制以及 Transformer 等结构进行梳理。并在此基础上，进一步从模型能力与复杂度的角度探讨了深度学习模型轻量化的理论背景，为后续轻量化模型设计提供了理论依据。最后介绍了遥感图像云去除中常用的评价指标，为后续实验结果的定量分析与方法对比提供了统一的评价基础。基于本章所述的理论分析，下一章将结合具体任务需求，进一步介绍所提出的遥感图像云去除模型及其网络结构设计。

## 第3章 SAR引导下的双分支去云网络设计

### 3.1 本章引言

SAR影像具备全天时、全天候成像能力，能够穿透云层获取稳定的地物散射信息，在云去除任务中展现出独特优势。近年来，基于SAR与光学影像融合的多模态云去除方法逐渐成为研究热点。然而，现有方法多采用并行特征提取或简单特征融合策略，SAR与光学模态之间的交互往往局限于后期阶段，难以在特征建模早期有效引入结构约束；同时，SAR影像固有的斑点噪声若缺乏有效抑制机制，也可能对光学重建结果造成干扰，影响光谱一致性。

针对上述问题，本章提出一种基于SAR引导的双分支遥感图像云去除网络（SAR-Guided Dual-Branch Network for Cloud Removal, SGN-CR）。该模型从多模态协同建模的角度出发，构建由SAR分支和光学分支组成的异构双分支结构，通过显式结构引导机制在特征提取阶段建立跨模态约束关系，使SAR提供的稳定几何结构信息能够主动参与光学特征建模过程，从而有效缓解厚云区域结构缺失与纹理模糊问题。

在网络结构设计上，SGN-CR结合卷积神经网络与Transformer的优势，分别针对SAR与光学影像的成像特性进行特征建模。同时，引入层级协同的跨模态特征融合策略，在不同尺度和语义层级下实现结构引导与语义补全的有机结合。此外，为在保证全局建模能力的同时控制计算开销，模型采用高效的注意力建模方式以捕获长程依赖关系，提升大范围云遮挡场景下的重建一致性。

本章将围绕SGN-CR网络的整体架构与关键模块展开详细介绍。首先给出模型的总体结构设计，其次依次阐述双分支特征编码机制、SAR引导注意力调制方法、跨模态融合策略以及全局建模方法，最后通过系统实验验证所提出方法在云去除性能方面的有效性，为后续章节的进一步研究奠定基础。

### 3.2 研究方法

#### 3.2.1 SGN-CR网络总体架构

为充分挖掘SAR与光学影像在云去除任务中的互补信息，本章提出的SGN-CR采用一种基于结构引导的双分支网络架构。该架构以“SAR结构先验引导光学重建”为核心思想，通过显式建模SAR图像特征和光学图像特征，在特征提取阶段即建立跨模态约束关系，从而提升厚云遮挡条件下的结构恢复能力与重建稳定性。

如图 3-1 所示, SGN-CR 的整体网络结构由四个主要部分组成: SAR 编码分支、光学编码分支、跨模态特征融合模块以及解码重建模块。网络输入包括一幅含云光学影像和一幅与之空间对齐的 SAR 影像, 二者分别进入对应的编码分支进行特征提取。在保持各自模态特性独立建模的同时, 通过引导与融合机制实现多层次的信息交互。

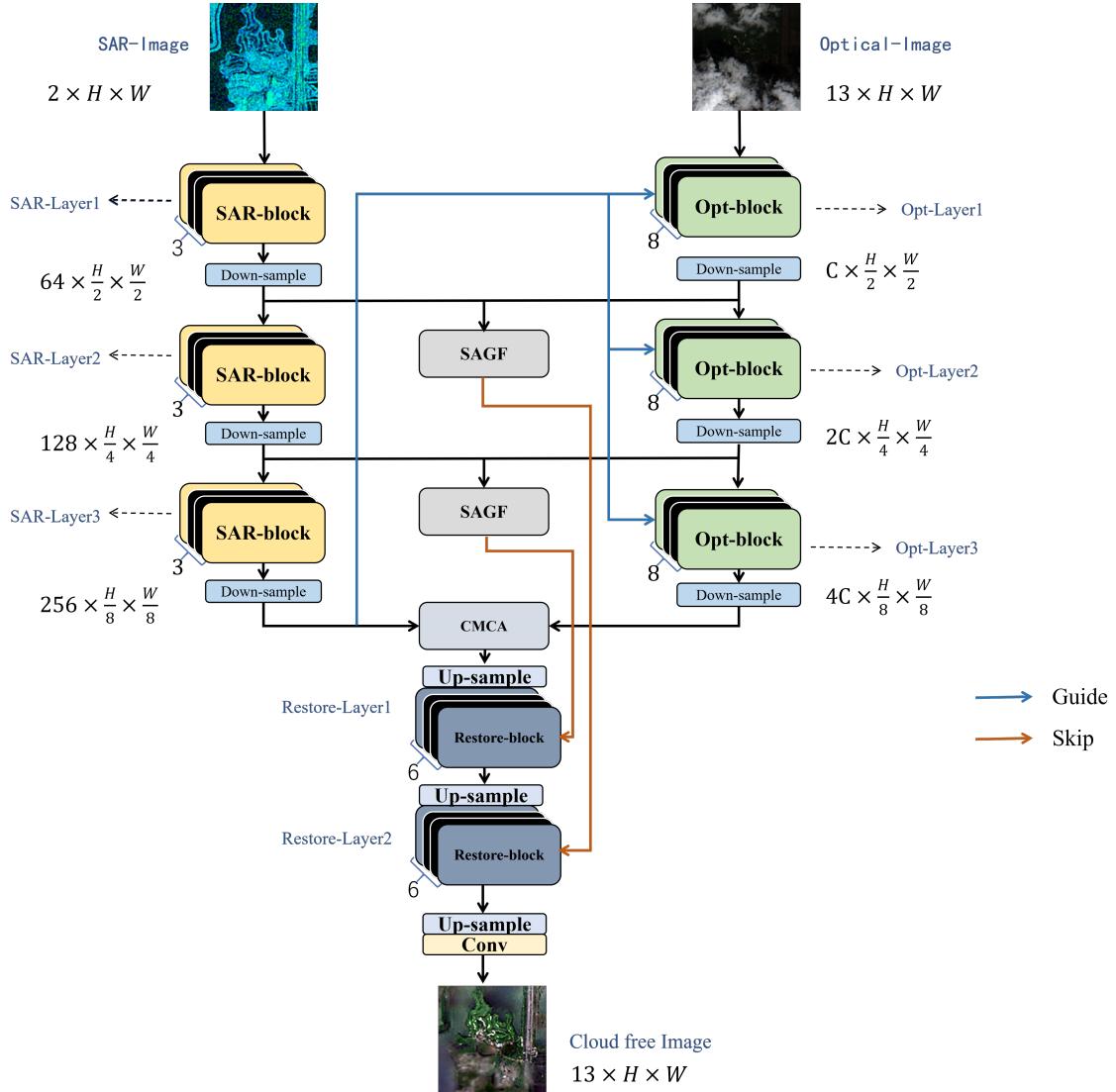


图 3-1 SGN-CR 整体网络结构示意图

Fig. 3-1 SGN-CR Overall Network Structure Diagram

SAR 编码分支采用基于卷积的结构, 用于从 SAR 影像中提取稳定的几何结构特征。该分支重点关注地物的轮廓、边界以及空间连续性等结构信息, 并通过多尺度特征提取逐步增强结构表达能力。由于 SAR 影像不受云层遮挡影响, 该分支能够在厚云区域提供可靠的结构先验, 为后续光学重建提供约束条件。经过编码后, SAR 分支输出的高层特征不仅参与跨模态融合, 同时作为结构引导信号传递至光

学分支。

光学编码分支采用基于注意力机制的特征建模方式，旨在从含云光学影像中恢复被遮挡区域的光谱与语义信息。考虑到云遮挡会导致局部信息严重缺失，该分支在特征提取过程中引入 SAR 引导机制，使光学特征在建模时能够感知潜在的地物结构分布，从而避免在厚云区域产生无约束的纹理推断。同时，通过多层次的编码结构逐步扩大感受野，以增强对大范围云遮挡区域的上下文建模能力。

在完成双分支特征编码后，SGN-CR 通过层级化的跨模态特征融合模块对不同尺度和语义层级的特征进行协同整合。浅层特征融合侧重于空间层面的结构注入与噪声抑制，以改善局部纹理连续性；深层特征融合则关注语义层面的对齐与补全，利用 SAR 提供的稳定结构信息辅助光学分支恢复被遮挡区域的高层语义表示。通过这种分层协同的融合方式，模型能够在不同尺度上充分发挥两种模态的互补优势。

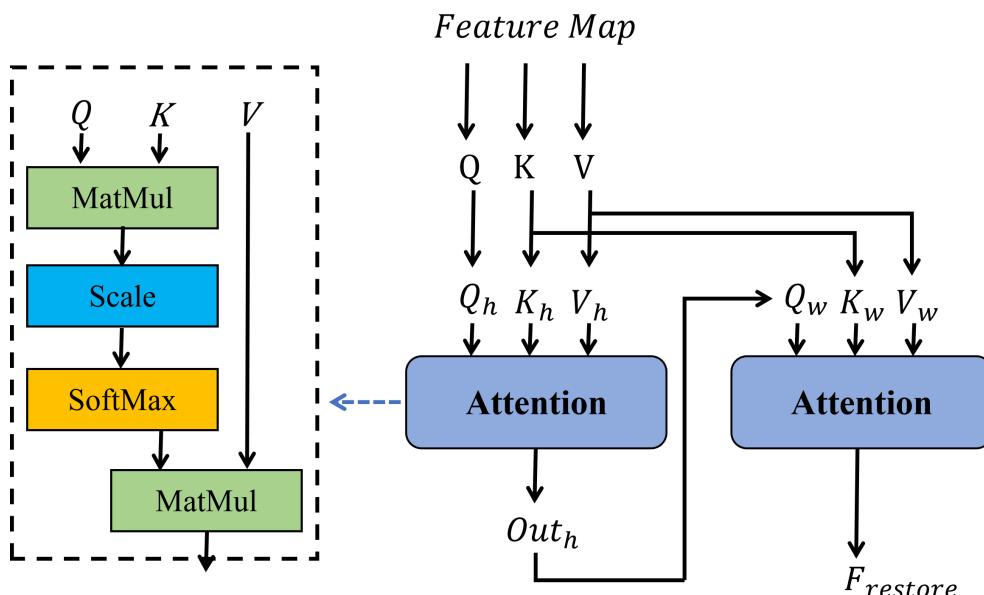


图 3-2 Restore-block 结构示意图

Fig. 3-2 ORestore-block structure diagram

为进一步说明解码重建模块的实现形式，本文采用 **Restore-block** 作为解码阶段的基本恢复单元。如图 3-2 所示，**Restore-block** 的核心计算遵循标准注意力建模流程：首先通过矩阵乘法计算查询与键之间的相关性，并进行尺度缩放与 Softmax 归一化以得到注意力权重，随后将权重与值向量进行加权聚合，实现特征重整与细节补偿。该模块主要用于在上采样恢复空间分辨率的同时，对云遮挡导致的局部纹理缺失区域进行自适应增强，从而提升解码阶段的结构连续性与边缘清晰度。需要指出的是，**Restore-block** 仅作用于融合后的特征恢复过程，以恢复空间分辨率

和细节结构为主要目标，不再引入新的跨模态交互机制，与编码端的 SAR 引导与跨模态融合形成明确分工。在保持整体结构一致性的同时，进一步细化纹理与边缘信息，使输出结果在结构连续性和光谱一致性方面均接近真实无云影像。

总体而言，SGN-CR 通过构建“编码—引导—融合—重建”的整体框架，在网络结构层面实现了 SAR 结构信息对光学重建过程的深度参与。该总体架构为后续各关键模块的设计提供了统一的组织形式，也为多模态信息在云去除任务中的有效协同奠定了基础。

### 3.2.2 SAR 引导的双分支特征编码结构

针对 SAR 与光学影像在成像机理和信息表达方式上的显著差异，SGN-CR 在特征编码阶段采用双分支结构，对两种模态分别进行特征建模，并通过引导机制在编码阶段建立跨模态约束关系。该设计避免了对异构模态进行同构处理所带来的信息混淆，使网络能够在保持模态特性的同时实现有效协同。

双分支编码结构由 SAR 分支和光学分支组成。其中，SAR 分支侧重于提取稳定的几何结构特征，为云遮挡区域的光学重建提供结构先验；光学分支则负责对含云影像进行语义与光谱特征建模，是最终去云结果的主要信息来源。二者在功能定位和网络结构上均存在明显差异，但在编码阶段通过显式引导机制形成协同关系。

#### 3.2.2.1 SAR 分支的结构特征提取网络

SAR 影像基于微波散射成像，不受云层遮挡影响，能够稳定反映地物的几何轮廓和空间布局特征。然而，其影像中通常包含较强的斑点噪声，且缺乏精细的光谱信息。因此，在 SGN-CR 中，SAR 分支的设计目标并非进行语义推理或光谱重建，而是提取具有稳定性和连续性的结构先验，用于引导光学分支的特征建模过程。

为此，SGN-CR 的 SAR 分支采用基于卷积神经网络的结构特征提取网络，通过局部感受野建模和层级下采样逐步强化地物结构表达能力。如图 3-3 中所示，SAR 分支由若干连续的卷积块堆叠而成，每个卷积块均采用“卷积-归一化-非线性激活”的基本结构形式，即 Conv-BatchNorm-ReLU。该结构能够在有效抑制斑点噪声的同时，突出地物边缘、轮廓和结构走向等几何信息。

在网络层级设计上，SAR 分支通过多尺度特征提取逐步扩大感受野，使高层特征能够聚合更大范围内的结构信息。随着网络深度的增加，特征表示由局部纹理逐渐过渡为抽象的结构描述，从而为厚云遮挡区域提供可靠的几何约束。由于 SAR 分支的功能明确限定为结构先验提取，其输出特征主要用于后续的引导与融合过程，而不直接参与最终影像的重建。

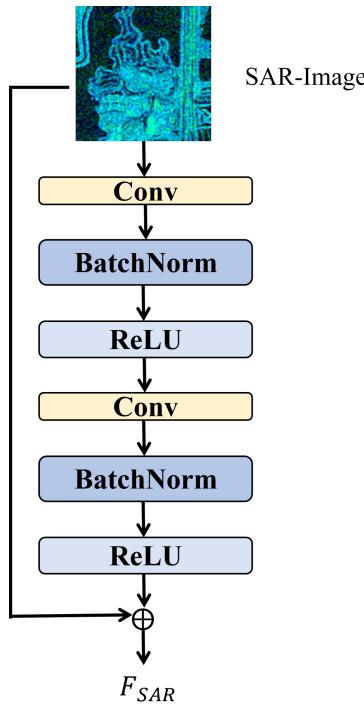


图 3-3 SAR-block 结构示意图

Fig. 3-3 SAR-block structure diagram

通过上述结构设计，SAR 分支能够在保持网络形式相对简洁的前提下，稳定提取对云去除任务具有关键约束作用的结构特征，为光学分支在云遮挡区域的特征建模提供必要的先验信息。

### 3.2.2.2 光学分支的语义特征编码网络

与 SAR 分支不同，光学影像包含丰富的光谱和语义信息，是云去除后最终恢复目标所在的模态。然而，在云遮挡区域，光学影像中的地物信息往往严重缺失，仅依赖局部上下文进行推断容易产生结构断裂或不真实的纹理补全。因此，光学分支在 SGN-CR 中承担着更为复杂的特征建模任务，需要同时具备局部细节建模和全局语义推理能力。

为增强对大范围云遮挡区域的建模能力，SGN-CR 的光学分支采用基于注意力机制的特征编码结构，在特征提取过程中显式建模长程依赖关系。通过多层编码结构逐步抽象特征表示，光学分支能够在较大空间范围内捕获上下文信息，为被遮挡区域的语义补全提供依据。同时，该分支在编码过程中保留多尺度特征表示，以兼顾局部纹理细节与高层语义一致性。

在双分支协同框架下，光学分支并非独立运行，而是在编码阶段引入来自 SAR 分支的结构引导信息。具体而言，SAR 分支输出的高层结构特征被用于调制光学

特征的空间响应，使光学编码过程能够感知潜在的地物结构分布。这种设计有效约束了云遮挡区域的特征生成方向，减少了无约束语义推断带来的伪纹理问题。

光学分支整体结构如图 3-4 所示，但需要指出的是，本节重点关注光学分支的整体编码形态及其在双分支结构中的功能定位，具体的结构引导方式和注意力调制机制将在下一节中进一步展开。通过上述语义特征编码网络的设计，SGN-CR 能够在充分利用光学影像语义信息的同时，引入 SAR 提供的结构先验，从而为后续高质量的云去除重建奠定基础。

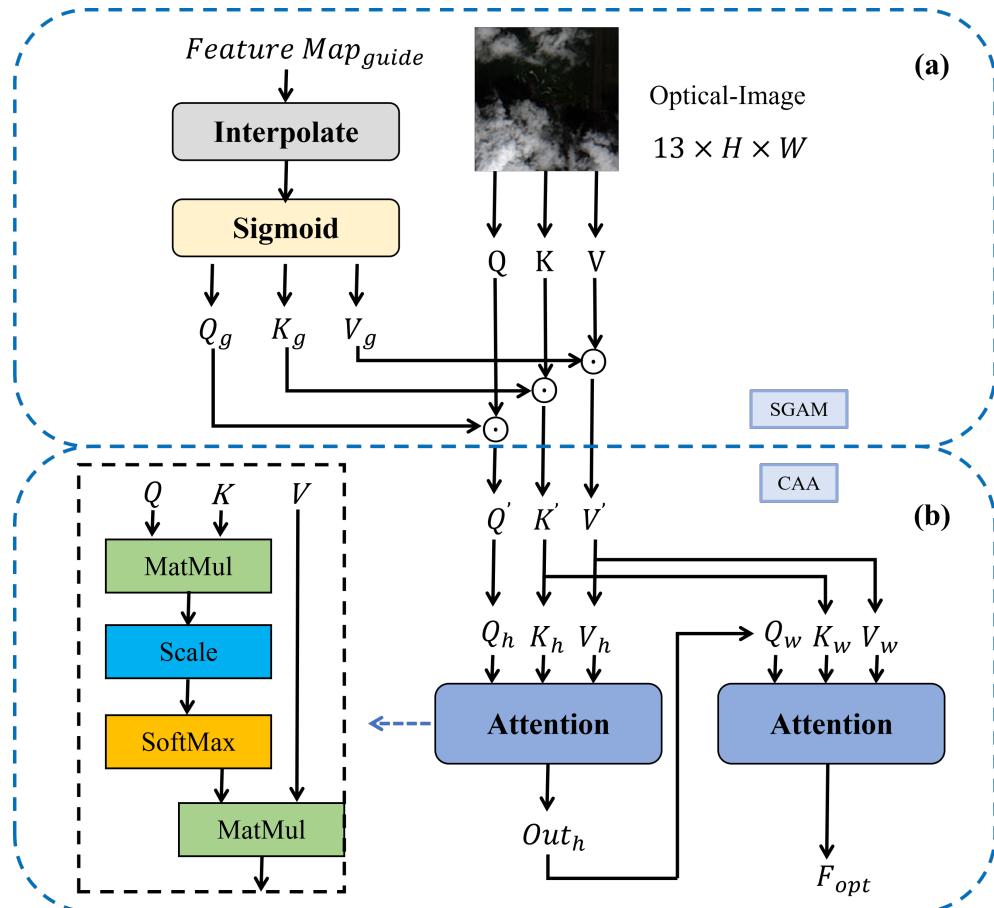


图 3-4 Opt-block 结构示意图

Fig. 3-4 Opt-block structure diagram

### 3.2.3 SAR 引导的注意力调制机制（SGAM）

在双分支特征编码结构中，SAR 分支能够提供稳定的地物结构先验，但若仅在特征融合阶段引入该信息，仍难以从根本上约束光学分支在厚云区域的特征生成过程。为此，SGN-CR 在光学特征编码阶段引入一种 SAR 引导的注意力调制机制（SAR-Guided Attention Modulation, SGAM），通过显式的空间权重调制方式，使 SAR 提供的结构信息能够主动参与光学特征建模过程。

如3-4中SAGM所示，SGAM的核心思想是利用SAR特征生成空间注意力权重，对光学特征进行逐位置调制，从而引导网络在云遮挡区域重点关注潜在的地物结构分布。与直接特征拼接或加权融合不同，SGAM不对两种模态的特征进行同构映射，而是将SAR信息转化为对光学特征响应强度的约束信号，使结构先验以“调制”的形式嵌入光学编码过程。

具体而言，设光学分支在某一编码阶段输出的特征图为 $F_{opt} \in \mathbb{R}^{C \times H \times W}$ ，而SAR分支深层尺度下输出的结构特征为 $F_{sar} \in \mathbb{R}^{C_s \times H_s \times W_s}$ 。SGAM将SAR分支输出的深层结构特征 $F_{sar}$ 作为调制信号 $Y_{guide}$ ，并利用 $Y_{guide}$ 生成用于调制注意力计算的引导项。如图3-4(a)所示，由于两分支在该阶段可能存在空间分辨率差异，首先对 $Y_{guide}$ 进行插值对齐：

$$\tilde{Y}_{guide} = \mathcal{I}(Y_{guide}) \quad (3-1)$$

其中 $\mathcal{I}(\cdot)$ 表示双线性插值算子。

接着，将对齐后的SAR引导特征映射为三路引导张量，用于分别调制光学注意力中的Query、Key与Value。具体地，先通过Sigmoid将引导响应进行归一化处理：

$$M = \sigma(\tilde{Y}_{guide}) \quad (3-2)$$

其中 $\sigma(\cdot)$ 为Sigmoid激活函数。随后，采用三组轻量映射函数将 $M$ 分别映射为 $Q_g$ ， $K_g$ ， $V_g$ ：

$$Q_g = \mathcal{G}_Q(M), \quad K_g = \mathcal{G}_K(M), \quad V_g = \mathcal{G}_V(M) \quad (3-3)$$

其中 $\mathcal{G}_Q(\cdot)$ ， $\mathcal{G}_K(\cdot)$ ， $\mathcal{G}_V(\cdot)$ 表示用于通道对齐与尺度匹配的轻量卷积映射。

另一方面，光学分支在进入自注意力计算前，将其特征 $F_{opt}$ 通过线性投影得到 $Q$ ， $K$ ， $V$ ：

$$Q = \phi_Q(F_{opt}), \quad K = \phi_K(F_{opt}), \quad V = \phi_V(F_{opt}) \quad (3-4)$$

其中 $\phi_Q(\cdot)$ ， $\phi_K(\cdot)$ ， $\phi_V(\cdot)$ 表示对应的线性映射。

最后，将三路SAR引导项以残差式“非抑制增强”的形式作用于光学分支的 $Q$ ， $K$ ， $V$ ，得到调制后的 $Q'$ ， $K'$ ， $V'$ ：

$$\begin{aligned} Q' &= Q \odot (1 + Q_g) \\ K' &= K \odot (1 + K_g) \\ V' &= V \odot (1 + V_g) \end{aligned} \quad (3-5)$$

其中 $\odot$ 表示逐元素乘法。上述 $(1 + \cdot)$ 的残差式调制确保调制系数的取值域为 $(1, 2)$ ，从而在任何情况下都不会抑制原始光学特征：当SAR引导响应较弱时，恒等项保

证  $Q$ 、 $K$ 、 $V$  的基础表征不被削弱；当 SAR 在道路、建筑边缘等区域给出显著结构提示时， $Q_g$ 、 $K_g$ 、 $V_g$  将同步增强对应位置的注意力表征，进而在后续自注意力计算中引导光学分支更聚焦于潜在地物结构而非云层噪声。

通过上述操作，光学特征在结构显著区域得到增强，而在结构不确定或噪声较多的区域受到抑制，从而有效约束云遮挡区域的特征生成过程。

需要强调的是，SGAM 的作用并非直接引入 SAR 的像素级信息，而是利用 SAR 特征的结构一致性生成引导权重，对光学特征进行空间层面的响应调制。这种设计能够在保持光学特征语义表达能力的同时，引入稳定的几何约束，避免厚云区域中出现无物理依据的纹理推断。

此外，SGAM 属于轻量级的引导机制，其主要计算开销来自简单的卷积映射和逐元素运算，不涉及复杂的矩阵乘法或全局注意力计算。因此，该模块可以在不显著增加计算复杂度的前提下，显著提升模型在复杂云遮挡场景下的结构一致性和重建稳定性。

通过在光学特征编码阶段引入 SGAM，SGN-CR 实现了 SAR 结构先验对光学特征建模的早期介入，为后续跨模态特征融合和全局建模提供了更加可靠的特征基础。

### 3.2.4 层级协同的跨模态特征融合策略

在完成双分支特征编码与 SAR 引导调制之后，如何在不同尺度和语义层级下有效整合 SAR 与光学特征，是提升云去除重建质量的关键问题。若仅在单一尺度或单一层级进行跨模态交互，容易导致结构信息利用不充分或语义补全能力受限。为此，SGN-CR 设计了一种层级协同的跨模态特征融合策略，在浅层与深层分别引入针对性的融合机制，实现结构信息与语义信息的互补协同。

该融合策略由两个功能互补的模块构成：空间自适应门控融合模块（Spatially Adaptive Gated Fusion, SAGF）与跨模态交叉注意力融合模块（Cross-Modal Cross Attention, CMCA）。二者分别作用于不同尺度和语义层级，形成由浅入深的跨模态协同过程。

#### 3.2.4.1 空间自适应门控融合模块（SAGF）

在编码网络的浅层阶段，特征主要包含局部纹理与边缘信息，空间分辨率较高。此时，SAR 特征能够提供清晰的地物轮廓和结构连续性，但其斑点噪声若直接注入光学特征，可能对局部纹理恢复造成干扰。因此，SGN-CR 在浅层采用 SAGF 模块对 SAR 信息进行选择性引入。

SAGF 的核心思想是根据光学特征与 SAR 特征的联合响应，自适应生成空间

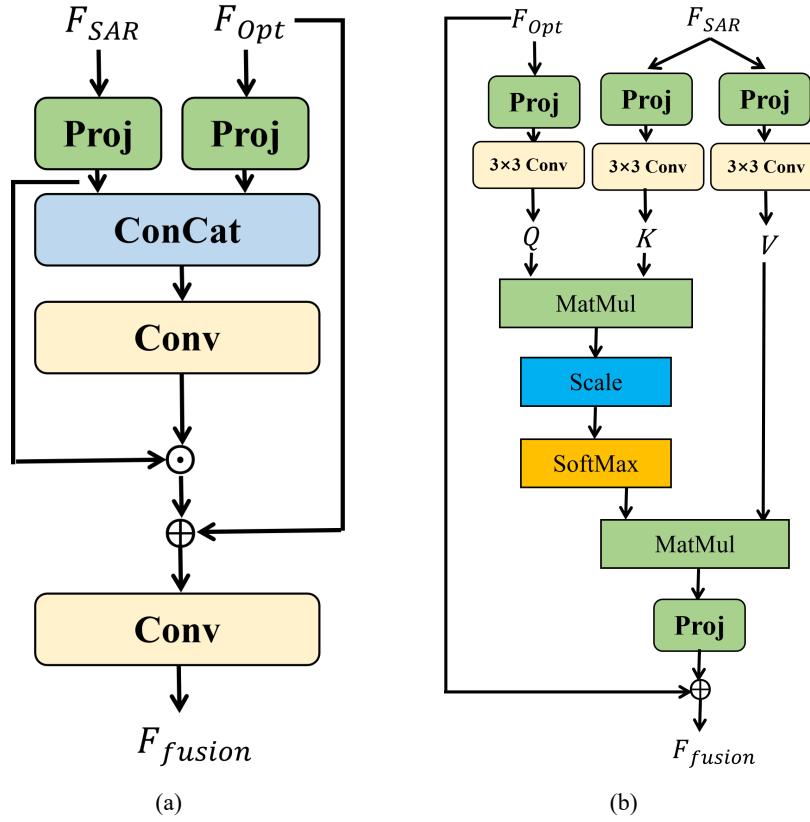


图 3-5 协同特征融合模块 (a) SAGF 模块示意图 (b) CMCA 模块示意图

Fig. 3-5 Collaborative Feature Fusion (a) SAGF module diagram (b) CMCA module diagram

门控权重，从而决定 SAR 结构信息在不同空间位置的注入强度。设某一尺度下光学特征与 SAR 特征分别为  $F_{opt}^{(l)} \in \mathbb{R}^{C_l \times H_l \times W_l}$ ,  $F_{sar}^{(l)} \in \mathbb{R}^{C_s \times H_l \times W_l}$ ，首先在通道维度上对两种模态特征进行拼接，得到联合特征表示：

$$F_{cat}^{(l)} = [F_{opt}^{(l)}, F_{sar}^{(l)}] \quad (3-6)$$

随后，通过轻量级卷积映射生成空间门控权重图：

$$G^{(l)} = \sigma(\mathcal{H}(F_{cat}^{(l)})) \quad (3-7)$$

其中  $\mathcal{H}(\cdot)$  表示由卷积层构成的映射函数， $\sigma(\cdot)$  为 Sigmoid 激活函数， $G^{(l)} \in \mathbb{R}^{1 \times H_l \times W_l}$ 。最终，SAGF 通过门控方式对 SAR 特征进行加权注入：

$$F_{fuse}^{(l)} = F_{opt}^{(l)} + G^{(l)} \odot F_{sar}^{(l)} \quad (3-8)$$

该公式设计的有效性建立在残差学习与不确定性建模的理论基础之上，门控掩膜  $G$  实质上充当了一个像素级的不确定性评估器。网络通过联合分析  $F_{opt}$  的纹理缺

失程度与  $F_{sar}$  的结构显著性，自适应地调节融合策略。

在厚云区域，门控值  $G$  趋近于 1，门控网络识别出光学信息的高不确定性，从而全强度激活 SAR 特征的注入权重。值得注意的是，由于  $F_{sar}$  已通过 SAR 分支进行了特征抽象与降噪处理，此时注入的是地物几何骨架特征而非原始的相干斑噪声，从而有效填补了纹理空洞。

在无云或者清晰区域，门控值  $G$  趋近于 0，此时光学特征  $F_{opt}$  包含高保真的地表光谱，对 SAR 特征进行空间阻断，模型能够无损地保留光学图像中未受污染的原始光谱信息（如颜色与亮度），防止 SAR 模态特征干扰高保真的光谱分布，对无云区域造成不必要的辐射扰动，有效维持了光谱的真实性。

而在过渡区域，此时一般是在云边缘或薄云区， $G$  处于中间值，网络自适应地平衡两者，实现纹理的平滑过渡。

这种机制成功地将 SAR 的结构优势与噪声劣势在空间域上进行了解耦，确保了模型仅在必要的位置引入必要的辅助信息，从而在最大化结构恢复能力的同时，实现了融合伪影的最小化。通过上述设计，SAGF 能够在结构显著区域增 SAR 信息的引导作用，同时在噪声或不确定区域抑制其干扰，从而在浅层阶段实现结构信息的稳健注入。

### 3.2.4.2 跨模态交叉注意力融合模块（CMCA）

在网络的深层阶段，特征逐渐具备较强的语义表达能力，空间分辨率相对降低。此时，云遮挡区域的主要问题由局部纹理缺失转变为高层语义不完整，简单的逐像素融合已难以满足语义补全需求。为此，SGN-CR 在深层引入 CMCA 模块，以实现更高层次的跨模态语义协同。

CMCA 采用交叉注意力机制，使光学特征能够从 SAR 特征中检索与当前语义位置相关的结构信息。设光学与 SAR 的深层特征分别为  $F_{opt}^{(h)}$  与  $F_{sar}^{(h)}$ ，CMCA 以光学特征作为查询（Query），以 SAR 特征作为键（Key）和值（Value），其计算过程可表示为：

$$Q = \mathcal{W}_q F_{opt}^{(h)}, \quad K = \mathcal{W}_k F_{sar}^{(h)}, \quad V = \mathcal{W}_v F_{sar}^{(h)} \quad (3-9)$$

其中  $\mathcal{W}_q$ 、 $\mathcal{W}_k$  和  $\mathcal{W}_v$  表示线性映射矩阵，并且将深层光学特征  $F_{opt}$  定义为 Q 的来源，将具备完整结构信息的 SAR 特征  $F_{sar}$  定义为 K 和 V 的来源。输入的特征通过投影层和卷积层在注意力计算前补充局部上下文信息，随后通过矩阵运算，生成跨模态注意力图：

$$A = \text{Softmax} \left( \frac{QK^\top}{\sqrt{d}} \right) V \quad (3-10)$$

其中  $d$  为特征维度,  $\text{Softmax}\left(\frac{QK^T}{\sqrt{d}}\right)$  为交叉注意力权重。最终融合特征由加权求和得到:

$$F_{fuse}^{(h)} = A + F_{opt}^{(h)} \quad (3-11)$$

通过 CMCA, 光学特征能够在语义层面有选择地吸收 SAR 中与当前上下文相关的结构信息, 从而提升厚云遮挡区域的语义一致性与结构合理性。

综上所述, SAGF 与 CMCA 分别从空间层面和语义层面对 SAR 与光学特征进行协同建模, 构成一种由浅入深的层级跨模态融合策略。该设计使结构先验在不同特征层级中均能够被有效利用, 为后续光学特征的全局建模与解码重建提供更加可靠的融合特征表示。

### 3.2.5 基于跨轴注意力的全局建模方法 (CAA)

在完成 SAR 引导的特征编码与跨模态融合之后, 光学特征已经在多尺度层面融合了结构先验信息。然而, 对于大范围厚云遮挡区域, 仅依赖局部卷积或逐像素融合仍难以充分建模长程依赖关系。为增强光学特征在空间维度上的全局建模能力, SGN-CR 在光学分支中引入一种基于跨轴注意力的全局建模方法 (Cross-Axis Attention, CAA), 用于在较低计算代价下捕获远距离像素之间的相关性。

传统二维自注意力通常在空间维度上对所有位置两两计算相关性, 其计算复杂度随空间分辨率呈平方增长, 在高分辨率特征图上计算代价较高。CAA 通过将二维空间注意力分解为沿单一轴方向的一维注意力计算, 在保持全局建模能力的同时有效降低计算复杂度。在跨轴注意力中, 注意力计算被分解为沿高度方向和宽度方向的两个一维注意力过程。如图3-4(b) 所示, CAA 模块以 SAGM 模块输出的经调制后的  $Q'$ 、 $K'$  和  $V'$  作为输入, 按顺序对高度和宽度方向进行注意力计算。

沿高度轴进行注意力计算是对特征图的每一列像素进行注意力计算, 能捕获垂直方向上的长程上下文依赖。这里的“长程依赖”指的是位于图像顶部和底部的像素之间的语义关联 (例如贯穿整幅图像的河流或道路), 这是由于云层遮挡可能导致局部纹理断裂, 需通过全局上下文进行推断, 复杂度为  $O(H \times W \times H)$ 。而宽度轴与高度轴同理, 但是对每一行像素进行计算, 捕获水平方向的长程依赖。复杂度为  $O(H \times W \times W)$ 。通过以上这种分解, CAA 的计算复杂度从  $O((HW)^2)$  降低为  $O(H^2W + HW^2)$ , 当  $H = W = N$  时, 复杂度从  $O(N^4)$  降至  $O(N^3)$ 。

具体实现是将  $Q'$ ,  $K'$ ,  $V'$  做维度变换得到  $Q_h$ ,  $K_h$ ,  $V_h$ , 沿高度轴计算注意力权重并进行特征聚合:

$$Out_h = \text{Softmax}\left(\frac{Q_h K_h^T}{d}\right) V_h \quad (3-12)$$

其中  $d$  表示特征维度,  $Q_h$  和  $K_h$  分别表示沿高度轴展开后的查询与键特征,

$\text{Softmax}\left(\frac{Q_h K_h^T}{d}\right)$  是沿高度轴计算的注意力权重。之后将  $K'$ ,  $V'$  进行维度变换作为下一步的  $K_w$ ,  $V_w$ , 将刚刚得到的  $Out_h$  作为  $Q_w$ 。同样在宽度方向上计算一维注意力后得到宽度方向的聚合特征:

$$Out_w = \text{Softmax}\left(\frac{Q_w K_w^T}{d}\right) V_w \quad (3-13)$$

最终, 将  $Out_w$  作为最终的跨轴注意力模块输出。

通过上述方式, CAA 能够在不显著增加计算开销的前提下, 使光学特征在空间维度上建立远距离依赖关系, 从而提升大范围云遮挡场景下的语义一致性和结构连贯性。需要指出的是, CAA 仅作用于光学特征内部, 不涉及 SAR 信息的直接参与, 其主要作用在于增强光学分支自身的全局建模能力, 与前述 SAR 引导机制和跨模态融合模块在功能上形成互补。

### 3.2.6 训练目标与损失函数设置

遥感图像云去除任务不仅要求模型在像素层面准确重建被云遮挡区域的地物信息, 还需同时保持地物结构的连续性与多光谱影像的光谱一致性。针对这一多目标约束问题, 单一损失函数往往难以全面刻画去云结果的质量。为此, 本文在训练阶段采用多项联合损失函数, 对网络输出从像素精度、结构纹理以及光谱保真性三个层面进行综合约束。

考虑到该类联合损失形式已在遥感图像云去除与重建任务中得到广泛验证, 本文在损失函数设计上未引入额外的复杂约束, 而是基于已有研究中成熟且稳定的联合损失框架进行继承与应用。具体而言, 本文采用文献 HPN<sup>[43]</sup> 中提出的联合损失函数作为训练目标, 其定义如下:

$$\mathcal{L}_{total}(P, T) = \alpha \mathcal{L}_{SmoothL_1}(P, T) + (1 - \alpha) \mathcal{L}_{MS-SSIM}(P, T) + \beta \mathcal{L}_{SAM}(P, T) \quad (3-14)$$

其中,  $P$  与  $T$  分别表示网络预测的去云光学影像与对应的真实无云影像, 二者维度均为  $C \times H \times W$ ;  $\alpha$  与  $\beta$  为用于平衡各损失项贡献的超参数。

#### 3.2.6.1 像素级重建损失 ( $SmoothL_1$ )

像素级重建损失用于直接约束模型输出在数值层面逼近真实无云影像, 是云去除任务中最基础的监督信号。本文采用  $SmoothL_1$  损失作为像素级约束项。相较于  $L_2$ ,  $SmoothL_1$  在大误差区域更具鲁棒性, 能够缓解云边缘和高反射区域可能带来的异常梯度, 从而促进训练过程稳定收敛并提高像素层面的重建精度。

### 3.2.6.2 结构相似性约束损失（MS-SSIM）

仅依赖像素级损失容易导致模型在云遮挡区域出现结构模糊或纹理断裂。为增强对结构与纹理细节的约束能力，本文引入多尺度结构相似性损失（Multi-Scale Structural Similarity, MS-SSIM），其从不同尺度对亮度、对比度与结构信息进行联合评估，有助于抑制重建过程中的过度平滑并提升纹理细节一致性。对应的损失形式为：

$$\mathcal{L}_{MS-SSIM}(P, T) = 1 - MS-SSIM(P, T) \quad (3-15)$$

### 3.2.6.3 光谱一致性约束损失（SAM）

遥感影像云去除不仅要求结构恢复合理，还需尽可能保持地物的真实光谱特性。为此，本文引入光谱角映射（Spectral Angle Mapper, SAM）作为光谱一致性约束，通过最小化预测光谱向量与真实光谱向量之间的夹角，降低多光谱通道间的相对失真风险。其定义为：

$$\mathcal{L}_{SAM}(P, T) = \cos^{-1} \left( \frac{\sum_{c,h,w} p_{c,h,w} t_{c,h,w}}{\sqrt{\sum_{c,h,w} p_{c,h,w}^2} \sqrt{\sum_{c,h,w} t_{c,h,w}^2}} \right) \quad (3-16)$$

其中， $p_{c,h,w}$  与  $t_{c,h,w}$  分别表示预测影像与真实影像在通道  $c$ 、空间位置  $(h, w)$  的像素值。该约束能够有效抑制去云过程引入的颜色偏移与光谱形态失真，从而提升结果在后续遥感解译任务中的可用性。

### 3.2.6.4 联合损失设置与分析

通过式 (3-14) 的联合优化，网络在训练过程中能够同时受到来自像素精度、空间结构与光谱分布三个层面的约束。该训练目标与本文 SGN-CR 的网络设计思想保持一致：SAR 分支与结构引导机制侧重于提升结构一致性，光学分支的注意力建模侧重语义与上下文推理，而联合损失则在监督层面对结构恢复与光谱保持提供互补约束，从而促进模型学习到更符合遥感成像特性的映射关系。

在实际训练中，本文参考相关工作经验并结合任务特点设置损失权重：为强化结构信息在去云重建中的约束作用，将  $\alpha$  设置为 0.2，使结构相关的 MS-SSIM 项在整体优化中占据更高权重；同时将  $\beta$  设置为 0.005，以保证 SAM 损失在数值量级上与其他损失项保持平衡。相关权重配置对模型性能的影响将在后续实验章节中进一步验证与讨论。

### 3.3 实验环境与数据集说明

#### 3.3.1 数据集说明

为验证所提出 SGN-CR 方法在遥感图像云去除任务中的有效性，本文选用公开的多模态遥感云去除数据集 SEN12MS-CR<sup>[42]</sup> 进行实验验证。该数据集包含来自全球不同区域的 169 个感兴趣区域（Regions of Interest, ROIs），覆盖多种地物类型、季节变化及气象条件，具有较强的多样性与代表性。

每个 ROI 覆盖的地面范围约为  $52 \times 40$  km，对应原始影像分辨率约为  $5200 \times 4000$  像素。为适应深度学习模型的输入需求，本文将原始影像裁剪为大小为  $256 \times 256$  的不重叠图像块。经处理后，数据集共包含 122,218 组样本，每组样本由以下三部分组成：(1) 一幅来自 Sentinel-2 的含云多光谱光学影像（包含 13 个光谱波段）；(2) 与之对应的无云光学参考影像；(3) 一幅来自 Sentinel-1 的 SAR 影像，包含 VV 与 VH 双极化通道。

在数据预处理阶段，本文采用与 HPN-CR 方法<sup>[43]</sup> 一致的处理流程。具体而言，对 SAR 数据进行强度截断以去除异常值，其中 VV 极化通道截断至  $[-25, 0]$  dB，VH 极化通道截断至  $[-35, 0]$  dB，随后统一归一化至  $[0, 1]$  区间，以减小不同通道间的数值尺度差异。

为评估模型的泛化能力，本文按照 ROI 级别对数据集进行划分，将全部 ROI 按照 149:10:10 的比例随机划分为训练集、验证集和测试集。最终，训练集包含 107,143 个样本，验证集包含 7,176 个样本，测试集包含 7,899 个样本。

#### 3.3.2 实验实现细节

本文所有实验均基于 PyTorch 深度学习框架实现，并在单张 NVIDIA GeForce RTX 4090 D GPU（显存 24 GB）上完成模型训练与测试。模型训练过程中采用 AdamW 优化器，初始学习率设置为  $1 \times 10^{-4}$ ，并使用余弦退火（Cosine Annealing）策略对学习率进行动态调整。

在训练设置方面，batch size 设置为 16，模型共训练 15 个 epoch。在保证训练过程稳定收敛的前提下，上述配置能够在训练效率与性能表现之间取得较好的平衡。除网络结构差异外，本文在所有对比实验中均保持一致的训练策略与参数设置，以确保实验结果的公平性与可比性。

#### 3.3.3 对比方法设置

为全面评估所提出 SGN-CR 方法在遥感图像云去除任务中的性能表现，本文选取了多种具有代表性的已有方法作为对比基线。根据模型所采用的输入模态形式

以及网络结构设计思想的不同，对比方法可大致分为以下三类。本文提出的 SGN-CR 属于基于 SAR 与光学影像协同建模的多模态深度学习方法，其设计目标是在保持结构恢复能力的同时提升重建稳定性，因此主要与第二类和第三类方法进行对比分析。

**(1) 单模态与早期生成式方法。**该类方法主要用于验证在云去除任务中引入多模态信息的必要性。具体包括 GANs<sup>[44]</sup>，其利用生成对抗网络实现从 SAR 到光学影像的直接映射；以及仅基于光学影像的注意力模型 AMGAN-CR<sup>[45]</sup>，通过挖掘光学影像内部上下文信息进行云遮挡区域的重建。

**(2) 传统多模态深度融合方法。**该类方法利用深度神经网络对 SAR 与光学影像进行特征级融合，代表了多模态遥感云去除研究中的主流技术路线。本文选取了多种具有代表性的模型进行对比，包括基于卷积神经网络的统一空间 - 光谱残差网络 USSRN-CR<sup>[46]</sup>，以及结合条件生成对抗网络的多模态生成模型 SAR-Opt-cGAN<sup>[47]</sup>。此外，还选取了基于 Transformer 架构的融合方法 GLF-CR<sup>[48]</sup> 和 Former-CR<sup>[49]</sup>，以对比不同注意力建模机制在云去除任务中的效果差异。

**(3) 异构并行网络方法。**为进一步验证本文所提出网络结构设计的有效性，本文选取最新的异构并行网络模型 HPN-CR<sup>[43]</sup> 作为核心对比方法。该模型通过并行建模不同模态特征并引入结构引导机制，在当前多模态云去除研究中具有代表性。与 HPN-CR 相比，本文提出的 SGN-CR 同样采用双分支结构，但在结构引导方式和特征协同机制上进行了针对性改进，相关性能对比结果将在后续实验中详细给出。

## 3.4 实验结果与分析

### 3.4.1 定量实验结果分析

表 3-1 给出了本文提出的 SGN-CR 与多种代表性遥感图像云去除方法在 SEN12MS-CR 数据集上的定量对比结果。对比方法涵盖单模态方法、传统多模态深度融合方法以及最新的异构并行网络模型。评价指标包括 PSNR、SSIM、SAM 和 MAE，分别从像素重建精度、结构一致性、光谱保持性和像素级误差等多个角度对去云效果进行综合评估。

从表 3-1 可以看出，不同技术路线在四项指标上的表现呈现出较为一致的规律：单模态方法在厚云区域受信息缺失约束，光谱或结构维度易出现明显退化；传统融合方法能够显著提升像素与结构指标，但往往面临“结构先验引入”与“斑点噪声污染”之间的权衡；先进异构网络在整体性能上更具竞争力，但若缺乏编码阶段的显式跨模态交互，其跨模态对齐精度与光谱一致性仍存在进一步提升空

表 3-1 不同云去除方法在 SEN12MS-CR 数据集上的定量对比结果

Table 3-1 Quantitative comparison of different cloud removal methods on the SEN12MS-CR dataset

模型	输入模态		PSNR(dB)↑	SSIM↑	SAM(°)↓	MAE↓
	光学	SAR				
GANs <sup>[44]</sup>	×	✓	24.1163	0.8455	9.7653	0.03792
AMGAN-CR <sup>[45]</sup>	✓	×	28.1985	0.8701	9.2323	0.02991
SAR-Opt-cGAN <sup>[47]</sup>	✓	✓	27.1266	0.8364	8.8707	0.03960
GLF-CR <sup>[48]</sup>	✓	✓	28.8497	0.8580	8.5006	0.02742
Former-CR <sup>[49]</sup>	✓	✓	28.5932	0.8799	8.2512	0.02814
USSRN-CR <sup>[46]</sup>	✓	✓	28.6043	0.8532	9.1736	0.02856
HPN-CR <sup>[43]</sup>	✓	✓	29.9422	0.8973	7.9434	0.02515
<b>SGN-CR</b>	✓	✓	<b>30.5503</b>	<b>0.8990</b>	<b>7.5781</b>	<b>0.02379</b>

间。本文 SGN-CR 通过“编码阶段显式引导 + 分层协同融合 + 低复杂度全局建模”的组合设计，在像素精度、结构一致性与光谱保持性三方面实现同步提升。所以 SGN-CR 在四项评价指标上均取得了最优的性能表现，验证了本文所提出异构双分支结构及 SAR 引导机制在遥感云去除任务中的有效性。

首先，与单模态方法相比，多模态信息的引入对提升云去除性能具有显著作用。仅依赖 SAR 输入的 GANs 方法虽然能够基于雷达后向散射特性恢复部分几何轮廓，但由于完全缺乏光学光谱参考，其重建结果在光谱一致性方面表现较差，PSNR 仅为 24.12 dB，SAM 高达 9.77°，存在明显的颜色失真问题。光学单模态方法 AMGAN-CR 将云去除视为图像修复任务，在薄云区域能够取得一定效果，但在厚云遮挡场景下缺乏穿透性结构信息支撑，难以恢复真实地物细节。相比之下，SGN-CR 通过融合 SAR 与光学影像，在 AMGAN-CR 的基础上实现了约 2.35 dB 的 PSNR 提升，表明多模态协同对于厚云场景下的云去除任务具有不可替代的重要性。

其次，在传统多模态深度融合方法中，SAR-Opt-cGAN、USSRN-CR 等模型通常采用通道拼接或简单特征叠加的方式进行硬融合。这类非选择性的融合策略容

易将 SAR 影像中固有的斑点噪声引入光学重建结果，这种噪声耦合会直接抬升像素误差 (MAE) 并破坏局部纹理一致性 (SSIM)，同时在光谱维度造成通道间相对关系扰动，使得 SAM 难以下降到较低水平，并且在无云或弱云区域产生明显颗粒感，影响整体图像质量。相比之下，SGN-CR 在特征融合阶段引入了空间自适应门控机制，根据云分布动态调节 SAR 特征的注入强度，在抑制噪声传播的同时充分利用 SAR 的结构先验信息。另外，本章提出的 SAGF 在浅层以像素级门控方式抑制噪声扩散，使 SAR 的贡献更集中于厚云区域的结构补偿，从而避免“引入结构先验的同时污染无云区域”的常见问题。最终使 SGN-CR 在结构一致性指标上取得了最高的 SSIM 值 (0.8990) 势。

进一步与异构并行网络方法 HPN-CR 对比可以发现，尽管 HPN-CR 通过并行建模不同模态特征，在多项指标上已取得较强性能 (PSNR 为 29.94 dB)，SGN-CR 仍在此基础上实现了进一步提升。具体而言，SGN-CR 在 PSNR 上提升约 0.60 dB，在 SAM 指标上降低约 0.36°，同时在 MAE 上取得更低误差。这表明 SGN-CR 在保持整体重建精度的同时，对光谱形态的保持更为准确。因此，相比于“并行提取、后期融合”的异构范式，本章进一步证明了“编码阶段显式引导、分层协同融合”的设计能够更充分释放 SAR 结构先验的约束价值，并在保持光谱一致性的同时实现更稳定的结构恢复。

上述性能提升并非源于简单的参数规模增加，而是得益于 SGN-CR 在结构设计上的多层次改进。一方面，SGN-CR 通过在特征提取阶段引入 SAR 引导的注意力调制机制，使 SAR 提供的深层结构先验能够直接参与光学特征的建模过程，从而在早期阶段实现更精确的跨模态对齐；另一方面，引入的跨轴注意力模块能够有效建模长程光谱依赖关系，在大范围云遮挡区域（如河流、农田等）中提升光谱一致性表现。此外，SGN-CR 采用分层协同的融合策略，在浅层抑制噪声、在深层补全语义，从整体上提升了多模态互补效果。

综合上述分析，定量实验结果表明，SGN-CR 在像素精度、结构一致性与光谱保持性等方面均优于现有代表性方法，验证了其在遥感图像云去除任务中的有效性与鲁棒性。

### 3.4.2 可视化实验结果分析

为直观评估不同方法在复杂云遮挡场景下的去云效果，本文从 SEN12MS-CR 测试集中选取若干具有代表性的样本进行可视化对比分析。这些样本覆盖了不同地物类型与不同云覆盖程度，其中部分区域受到厚云严重遮挡，地表信息几乎完全缺失，能够充分反映模型在极端条件下的重建能力。

图 3-6 展示了各方法在相同测试样本上的去云结果。图像从左至右依次为 SAR

输入影像、含云光学影像、无云参考影像以及不同对比方法和本文方法的重建结果。通过可视化对比，可以更加直观地观察不同模型在结构恢复、纹理连续性和光谱一致性方面的差异。

从单模态方法的结果可以看出，仅依赖 SAR 的方法（如 GANs）虽然能够利用雷达回波信息恢复道路、水体等地物的基本轮廓，但由于缺乏光学光谱先验，其重建结果普遍存在明显的颜色失真问题，整体色调偏暗且单一，难以真实反映地表覆盖类型。仅基于光学影像的 AMGAN-CR 方法在薄云区域具有一定修复能力，但在厚云遮挡区域（图中红色标注区域）由于无法获取穿透性结构信息，往往出现纹理模糊甚至伪影生成的现象。

对于多模态融合方法，不同模型在可视化效果上仍存在较大差异。部分 Transformer 融合方法（如 GLF-CR）在整体色彩一致性方面有所改善，但容易对局部高频细节产生过度平滑，导致道路边缘和建筑轮廓不够清晰。基于卷积结构的融合方法（如 SAR-Opt-cGAN 和 USSRN-CR）在一定程度上引入了 SAR 结构信息，但由于缺乏有效的噪声抑制机制，重建结果中仍可观察到明显的斑点噪声残留，尤其是在无云区域更为明显。

相比之下，本文提出的 SGN-CR 在不同场景下均表现出更优的视觉效果。得益于 SAR 结构引导机制与层级协同融合策略，SGN-CR 能够在厚云遮挡区域准确恢复道路网络、农田边界等复杂几何结构，同时在光谱层面保持与无云参考影像高度一致的色彩分布。在图中红色标注的关键区域内，SGN-CR 有效避免了光谱偏移和噪声残留问题，在保证图像清晰度的同时实现了自然平滑的纹理过渡。

总体来看，可视化实验结果与前述定量评价指标的结论保持一致，进一步验证了 SGN-CR 在结构恢复能力、光谱保持性以及整体视觉质量方面的综合优势。

### 3.4.3 关键模块消融实验分析

为了全面验证 SGN-CR 网络中各核心组件的有效性和具体贡献，我们在 SEN12MS-CR 数据集上进行了系统的消融实验。这些实验旨在解答 SAR 引导的注意力调节是否比无引导的并行特征提取更有效、CAA 在平衡全局建模和计算效率方面是否优于传统机制、协同特征融合策略是否是特征融合的最优方案这三个问题。这些实验的定量比较结果分别总结在表 3-2、3-3 和 3-4 中。

#### 3.4.3.1 SAR 引导注意力调制机制的有效性

为验证 SAR 引导注意力调制机制（SGAM）在特征编码阶段的实际作用，本文在保持双分支网络结构和后续跨模态融合策略不变的前提下，构建了一种不引入 SAR 引导的对比模型。该变体仅保留 SAR 与光学分支的并行特征提取过程，并

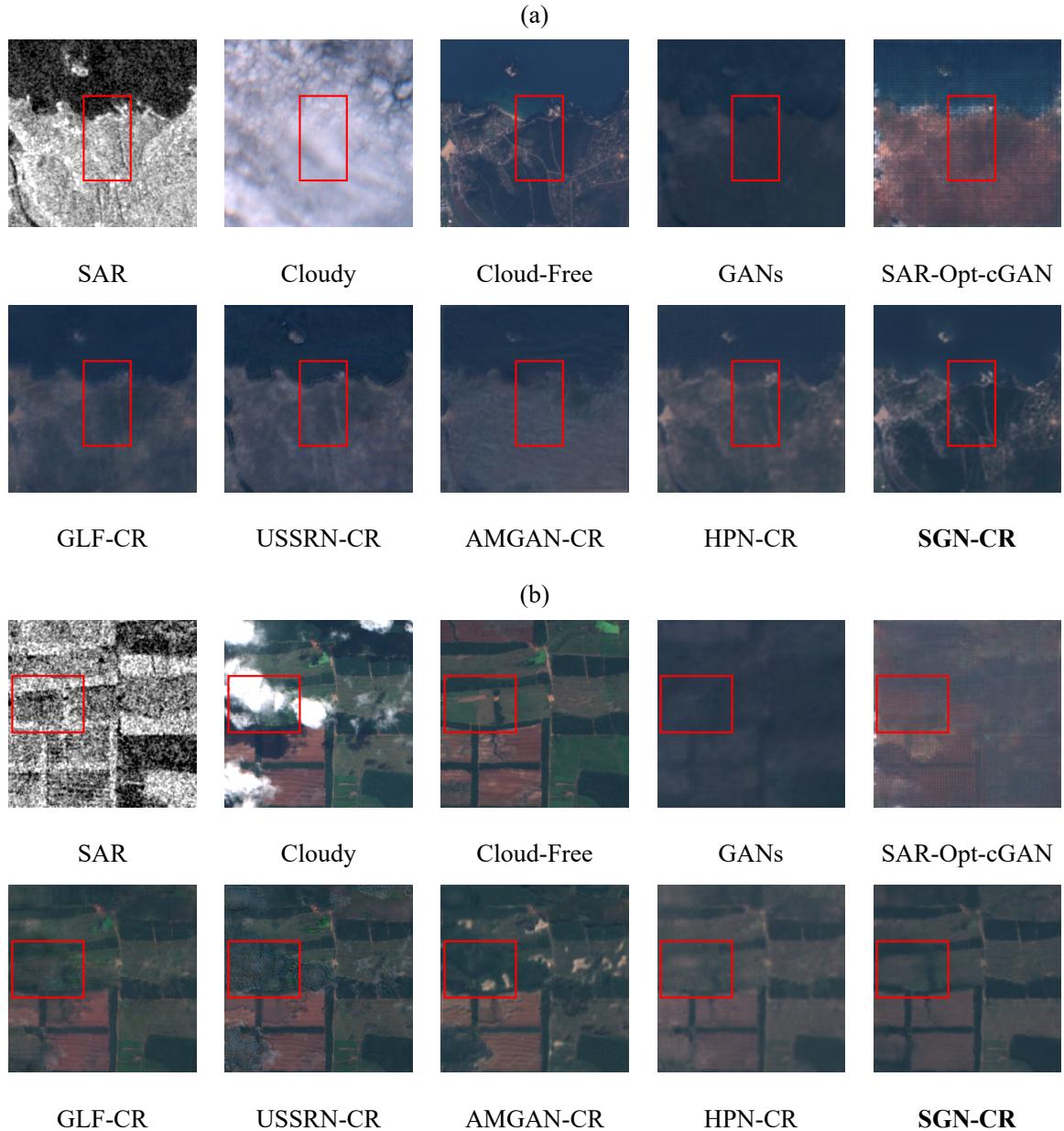


图 3-6 SEN12MS-CR 测试集上不同方法的云去除对比结果。(a) (b) 为不同代表性场景。

Fig. 3-6 Comparison of cloud removal results on representative scenes from the SEN12MS-CR test set. (a) and (b) denote two representative scenes.

在后续阶段进行协同融合，其结构示意如图 3-4(a) 所示。

表 3-2 给出了是否引入 SGAM 的定量对比结果。可以观察到，在移除 SGAM 后，模型的 PSNR 由 30.5503 dB 降至 29.6152 dB，SSIM 由 0.8990 降至 0.8957，同时 SAM 和 MAE 均出现明显劣化。

表 3-2 SAR 引导注意力调制机制消融实验结果

Table 3-2 Ablation results of SAR-guided attention modulation

是否引入 SGAM	PSNR(dB)↑	SSIM↑	SAM(°)↓	MAE↓
是	<b>30.5503</b>	<b>0.8990</b>	<b>7.5781</b>	<b>0.02379</b>
否	29.6152	0.8957	8.1098	0.02578

该结果表明，仅依赖后续特征融合难以弥补光学分支在特征编码阶段缺乏结构约束所带来的误差累积问题。在厚云遮挡区域，光学分支若无法感知潜在地物结构，其特征建模容易受到云层噪声干扰，从而影响后续重建质量。

相比之下，SGAM 通过在特征编码阶段引入 SAR 深层结构先验，对光学注意力计算过程进行显式调制，使光学特征在早期即获得几何约束。这种“先引导、后融合”的设计有效抑制了无约束特征传播，验证了显式结构引导优于仅依赖并行特征提取与隐式融合的建模方式。

### 3.4.3.2 跨轴注意力机制（CAA）的有效性

为评估不同注意力机制在高分辨率遥感场景下的性能与计算效率，本文将光学分支中的注意力模块分别替换为标准全局注意力<sup>[1]</sup>（Standard Attention）、窗口注意力<sup>[2]</sup>（Window Attention）以及所提出的跨轴注意力（CAA）。此外，针对 CAA 的具体计算方式，进一步对比了并行计算策略（ $H \parallel W$ ）以及两种顺序分解策略，即宽度优先（ $W \rightarrow H$ ）和高度优先（ $H \rightarrow W$ ），以分析轴向计算顺序对模型性能的影响。

考虑到推理时间高度依赖具体硬件平台和运行环境，难以进行跨方法的公平对比，本文采用浮点运算次数（FLOPs）作为衡量模型计算复杂度的核心指标。FLOPs 能够量化网络单次前向传播所需的理论计算开销，具有良好的硬件无关性。尤其在星载或边缘遥感应用场景中，模型部署通常受到算力与功耗预算的严格限制，因此基于 FLOPs 的复杂度评估更能真实反映算法的工程可行性。为保证对比的公平性，表 3-3 中所有 FLOPs 均在相同配置条件下统计，包括特征维度为 256、注意力头数为 4、特征图尺寸为  $64 \times 64$ ，窗口注意力的窗口大小设为 16。

表 3-3 汇总了不同注意力机制及轴向计算策略的定量对比结果。可以观察到，标

表3-3 不同注意力机制与轴向计算策略消融实验结果

Table 3-3 Ablation results of attention mechanisms and axial strategies

注意力机制	FLOPs(G)↓	PSNR(dB)↑	SSIM↑	SAM(°)↓	MAE↓
Standard Attention	9.237	29.8041	0.8972	7.9097	0.02531
Window Attention	1.126	30.1116	0.8971	7.9984	0.02441
CAA (并行计算)	0.585	30.3405	0.8984	7.8893	0.02392
CAA (宽度优先)	0.316	30.3457	0.8982	7.8926	0.02418
CAA (高度优先)	<b>0.316</b>	<b>30.5503</b>	<b>0.8990</b>	<b>7.5781</b>	<b>0.02379</b>

准全局注意力虽然具备完整的空间建模能力，但其计算复杂度高达 9.237 GFLOPs。如此高昂的计算开销不仅显著增加了训练和推理成本，也在一定程度上制约了模型的优化过程，最终未能取得最优的重建性能，其 PSNR 仅为 29.8041 dB。窗口注意力通过局部窗口划分有效降低了计算复杂度（1.126 GFLOPs），但由于不同窗口之间缺乏直接的长程交互，其在建模大尺度光谱一致性方面存在天然缺陷，导致 SAM 指标表现不佳。

相比之下，本文提出的跨轴注意力机制在显著降低计算复杂度的同时，实现了更优的重建性能。CAA 通过将二维全局注意力分解为沿空间轴向的一维注意力计算，有效避免了全局注意力中高维矩阵运算带来的计算爆炸问题。在顺序分解策略下，CAA 的计算复杂度进一步降低至 0.316 GFLOPs，相比标准全局注意力减少约 29 倍，相比窗口注意力降低约 3.5 倍，同时在 PSNR、SSIM 和 SAM 等指标上均取得最优结果。

进一步对不同轴向计算顺序的分析表明，并行计算策略虽然能够同时建模水平与垂直方向的上下文关系，但需要分别计算两个轴向的注意力，其 FLOPs 达到 0.585 GFLOPs，计算开销仍然高于顺序分解方案。相比之下，顺序分解策略通过将二维建模过程拆解为两个串行的一维操作，在保证信息逐步传播的同时显著降低了计算成本。

本文进一步参考 Axial-DeepLab<sup>[3]</sup> 中提出的正交分解思想，对顺序分解策略中的轴向执行顺序进行了系统比较。实验结果表明，高度方向优先的顺序分解策略在本数据集上取得了最佳性能，其 PSNR 达到 30.5503 dB，SAM 降至 7.5781°，同时保持最低的计算复杂度。这一结果表明，在遥感影像中，优先聚合垂直方向的上

下文信息有助于为后续的水平方向特征传播提供更加稳定的语义基础，从而提升整体结构恢复与光谱一致性。

基于上述分析，本文最终采用高度方向优先的跨轴注意力计算策略。该设计在继承正交分解高效建模优势的同时，实现了全局上下文建模能力与计算效率之间的良好平衡，为高分辨率遥感云去除任务提供了一种兼具性能与可部署性的注意力建模方案。

### 3.4.3.3 协同特征融合策略的合理性

为验证不同特征层级采用差异化融合策略的必要性，本文针对浅层与深层特征的融合模块组合方式进行了系统消融分析，相关定量结果汇总于表 3-4。通过对不同融合策略配置的对比，可以清晰地揭示融合模块功能与特征层级属性之间的内在匹配关系。

表 3-4 不同特征融合策略组合的消融实验结果

Table 3-4 Ablation results of different fusion strategy combinations

Fusion1	Fusion2	Fusion3	PSNR(dB)↑	SSIM↑	SAM(°)↓	MAE↓
Concat	Concat	Concat	28.8195	0.8810	8.7375	0.03232
SAGF	SAGF	SAGF	29.8027	0.8974	7.8453	0.02516
Concat	Concat	CMCA	30.1195	0.8979	7.7314	0.02472
SAGF	Concat	CMCA	30.2255	0.8964	7.7110	0.02433
Concat	SAGF	CMCA	30.3627	0.8987	7.8096	0.02408
<b>SAGF</b>	<b>SAGF</b>	<b>CMCA</b>	<b>30.5503</b>	<b>0.8990</b>	<b>7.5781</b>	<b>0.02379</b>

当在浅层阶段引入空间自适应门控融合模块（SAGF）后，模型性能得到显著提升。相比全拼接方案，PSNR 提升约 0.98 dB，同时 SSIM 和 MAE 均明显改善。这一结果表明，基于空间不确定性的门控机制能够有效感知云分布与局部结构可靠性，在浅层特征中动态抑制 SAR 斑点噪声对光学纹理的干扰，从而为后续重建提供更加干净且结构一致的特征表示。

进一步在深层阶段引入跨模态通道注意力模块（CMCA）后，模型性能得到进一步提升。通过对比仅使用 SAGF 的方案与最终采用 SAGF+CMCA 的配置，可以观察到 PSNR 从 29.80 dB 提升至 30.55 dB，同时 SAM 明显降低。这一改进表明，

深层特征主要承载的是语义与拓扑结构信息，仅依赖空间门控已难以弥补厚云区域的语义缺失问题，而 CMCA 通过通道维度的跨模态注意力交互，能够更有效地从 SAR 特征中检索并补全缺失的高层语义结构，而非局限于像素级的局部筛选。

进一步对比不同融合组合（表中第 4-6 行）可以发现，仅当浅层采用 SAGF、深层采用 CMCA 时，模型在 MAE（0.02379）和 SSIM（0.8990）等指标上同时达到最优。这一结果充分验证了分层协同融合策略的合理性：浅层特征以高频纹理与边缘信息为主，更适合采用空间门控机制进行噪声抑制与结构筛选；而深层特征更侧重语义表达与全局结构关系，适合通过跨模态注意力机制进行语义补全与拓扑恢复。

需要指出的是，本文未在浅层阶段直接引入 CMCA，主要基于计算效率与特征属性两方面的综合考量。一方面，浅层特征图通常具有较高空间分辨率（如  $H/2$ 、 $H/4$ ），若在该阶段执行跨模态注意力投影与高维矩阵运算，将带来显著的额外计算与显存开销，不利于大规模遥感影像训练与部署。另一方面，浅层特征中包含大量局部高频纹理，同时也伴随较强的 SAR 斑点噪声。在该阶段进行强语义对齐，容易导致注意力机制对噪声模式产生过度响应，从而引入冗余建模甚至过拟合风险。相比之下，SAGF 的空间敏感性使其更适合完成浅层“噪声抑制与结构筛选”的任务，而 CMCA 则更适合在语义更加明确的深层特征上执行“跨模态语义补全”。

基于上述分析，本文最终采用浅层 SAGF 与深层 CMCA 相结合的分层协同融合策略，在结构引导、噪声抑制与语义补全之间实现了有效平衡。

综合三组消融实验结果可以进一步总结如下结论：首先，SGAM 的引入表明跨模态交互应尽可能前置，通过在特征编码阶段显式注入 SAR 结构先验，可有效降低厚云区域的重建不确定性；其次，CAA 通过轴向分解在显著降低计算复杂度的同时保留了全局上下文传播能力，实现了遥感影像大尺度建模与计算效率之间的平衡；最后，分层协同融合策略证明了“融合模块功能应与特征层级属性相匹配”的设计原则，浅层门控去噪与深层语义对齐的协同作用，有效缓解了 SAR 噪声污染与语义缺失之间的矛盾，从而解释了表 3-1 中 SGN-CR 在 SSIM、SAM 与 MAE 等指标上的综合优势来源。

### 3.4.4 SGN-CR 实验结果综合讨论

结合前述定量评价、可视化结果以及消融实验分析，可以对 SGN-CR 在遥感图像云去除任务中的整体表现与内在机理进行进一步综合讨论。从实验结果来看，SGN-CR 在结构恢复精度、光谱一致性以及整体重建稳定性等方面均显著优于对比方法，这一性能优势并非来源于单一模块的叠加，而是多项针对性设计在不同阶段协同作用的结果。

首先，从结构恢复角度分析，SGN-CR 在厚云遮挡区域能够稳定重建道路、河流及地块边界等几何结构，这与 SAR 结构先验在特征编码阶段的早期介入密切相关。与多数仅在融合阶段引入 SAR 信息的方法不同，SGN-CR 通过 SAR 引导注意力调制机制（SGAM），在光学分支特征提取过程中显式引入结构约束，使光学特征在编码初期即感知潜在地物轮廓。这种“编码阶段引导”的设计有效避免了无约束纹理推断带来的结构漂移问题，为后续重建奠定了稳定的几何基础。

其次，从光谱保持与噪声抑制的角度来看，实验结果表明，SGN-CR 在引入 SAR 信息的同时并未造成光谱失真或噪声放大，这一点在 SAM 与 MAE 指标以及可视化结果中均得到验证。其关键原因在于模型采用了层级协同的跨模态特征融合策略：在浅层阶段通过空间自适应门控融合模块（SAGF）对 SAR 特征进行选择性注入，有效抑制斑点噪声向光学特征的传播；在深层阶段则利用跨模态交叉注意力模块（CMCA）完成语义层面的结构补全。这种分层匹配的融合方式避免了单一融合所带来的干扰，使结构引导与光谱保真之间达到较好的平衡。

再次，从全局建模能力与计算效率的综合表现来看，跨轴注意力机制（CAA）在实验中展现出明显优势。消融实验结果表明，CAA 在显著降低计算复杂度的同时，仍能够有效捕获长程空间依赖关系，从而提升大范围云遮挡区域的语义一致性。这一现象说明，在高分辨率遥感影像场景下，采用轴向分解的全局建模方式能够在性能与效率之间取得更合理的折中，为复杂场景下的云去除任务提供了一种可行的建模思路。

综合来看，SGN-CR 的性能提升并非依赖单一模块的增强，而是源于整体架构层面的协同设计：SAR 结构先验通过早期引导约束光学特征生成方向，分层融合策略在不同特征层级上实现噪声抑制与语义补全，而高效的全局建模机制进一步增强了模型在大尺度遮挡场景下的稳定性。这种从“编码—引导—融合—建模”多个环节协同优化的设计，使得 SGN-CR 能够在复杂云遮挡条件下实现结构一致、光谱可信且稳定可靠的去云重建结果。

上述综合实验结果充分验证了本文提出方法在遥感图像云去除任务中的有效性，也为后续轻量化设计与工程化应用提供了明确的性能基准和结构依据。

### 3.5 本章小结

本章围绕所提出的 SGN-CR 遥感图像云去除方法，从实验设计、定量评价、可视化分析以及消融实验等多个角度，对模型的整体性能与关键结构设计进行了系统验证与分析。实验结果表明，SGN-CR 在复杂云遮挡场景下能够同时兼顾结构一致性、光谱保真性与重建稳定性，在多项评价指标上均优于现有代表性方法，验证

了所提出方法的有效性。

通过定量实验与可视化结果可以看出，SGN-CR 在厚云区域表现出更强的结构恢复能力，能够稳定重建道路、地块边界等几何结构，同时保持较为自然的光谱分布。进一步的消融实验表明，该性能提升并非源于单一模块，而是由多项针对性设计协同作用的结果。其中，SAR 引导注意力调制机制在特征编码阶段引入显式结构约束，有效抑制了云噪声对光学特征建模的干扰；层级协同的跨模态特征融合策略在不同特征层级上实现了噪声抑制与语义补全的合理分工；跨轴注意力机制则在保证全局建模能力的同时显著降低了计算复杂度，为大尺度遥感场景下的高效建模提供了可行方案。

综合分析可知，SGN-CR 的优势来源于整体架构层面的系统性设计，而非简单的网络堆叠或参数扩展。通过在编码、引导、融合与全局建模等多个阶段引入针对遥感云去除任务特性的结构约束与信息交互机制，模型能够在复杂遮挡条件下实现更加可靠的去云重建。

需要指出的是，尽管 SGN-CR 在性能上取得了较为理想的结果，其双分支异构结构与多模块协同设计也不可避免地带来了较高的计算复杂度和参数规模。这在一定程度上限制了模型在资源受限场景下的部署与应用。基于此，下一章将在本章实验结果与结构分析的基础上，进一步探讨 SGN-CR 的轻量化设计思路，提出面向高效推理的 Lite-SGN-CR 网络结构，以在保证去云性能的前提下降低计算开销。

### 3.6 引用参考文献

参考文献引用示例：单篇引用 [?], 单篇多次引用 [?]<sup>55</sup>, 多篇同处引用 [?, ?, ?, ?]

## 第4章 基于 SGN-CR 的轻量化遥感图像云去除方法研究

### 4.1 本章引言

第三章围绕遥感图像云去除任务，提出了一种基于 SAR 引导的双分支深度学习模型 SGN-CR。该模型通过引入主动式结构引导机制与层级协同特征融合策略，有效缓解了厚云遮挡条件下光学影像中结构缺失与纹理失真问题，并在 SEN12MS-CR 数据集上取得了优于现有方法的重建性能。相关实验结果表明，SAR 所提供的稳定几何结构信息在云去除过程中具有重要作用，能够为光学影像恢复提供可靠的结构先验。

在模型性能不断提升的同时，SGN-CR 的网络结构也逐渐趋于复杂。该模型采用异构双分支架构，在光学分支中引入多层注意力模块以建模长程依赖关系，并在多个阶段进行跨模态特征交互。这类设计显著增强了模型的特征表达能力，但同时也带来了参数规模和计算复杂度的持续增长。上述以复杂度换取性能的设计思路在高性能计算平台上具有一定可行性，但在实际遥感应用场景中仍面临较为明显的限制。

在星载处理、无人机遥感以及灾害应急等端侧应用场景中，模型通常运行于算力受限且功耗敏感的硬件平台，对推理效率和模型规模提出了更为严格的要求。当模型计算复杂度过高时，不仅会导致推理延迟显著增加，还可能直接限制其在实际系统中的部署与推广。因此，在此类应用背景下，仅关注云去除精度已难以满足实际需求，如何在保证重建质量的同时兼顾模型效率，逐渐成为遥感云去除研究中亟需面对的问题。

进一步分析可以发现，SGN-CR 所采用的多模态结构在性能提升方面具有明显优势，但其各组成模块在功能定位和复杂度贡献上并不均衡。光学分支主要承担光谱与语义信息恢复任务，计算开销相对较大。SAR 分支侧重于提供结构先验，其对整体重建性能的贡献并不完全依赖于复杂的网络结构。一些引导与融合模块虽然计算代价较低，但在抑制伪纹理和增强结构一致性方面发挥了关键作用。这一现象表明，SGN-CR 的整体结构中仍存在可进一步优化和压缩的空间。

基于上述分析，本章在第三章工作的基础上，进一步关注模型计算效率与实际部署可行性问题，围绕 SGN-CR 网络结构展开深入研究，探索一种面向端侧遥感应用的轻量化设计方案。与统一压缩策略不同，本章充分考虑 SAR 与光学模态在云去除任务中的功能差异，从模块选择性轻量化的角度出发，对网络架构进行有针对性的简化与重设计。

本章的主要工作包括对 SGN-CR 模型复杂度及各模块特性进行系统分析，明

确影响模型效率的关键因素。在此基础上，提出一种模块选择性的轻量化网络 Lite-SGN-CR，在保留核心引导机制和跨模态协同能力的前提下显著降低模型复杂度。最后，通过定量实验与可视化分析对所提出方法在性能与效率之间的权衡效果进行验证。实验结果表明，该轻量化设计在保持云去除质量基本稳定的同时，有效提升了模型在资源受限平台上的应用潜力。

## 4.2 SGN-CR 模型复杂度分析与轻量化动机

为明确第三章所提出 SGN-CR 模型在端侧部署场景下面临的实际问题，并为后续轻量化设计提供依据，本节从模型整体复杂度、模块级计算分布以及功能一复杂度匹配关系等方面，对 SGN-CR 网络结构进行进一步分析。

TODO:(不一定使用这个工具) 本章采用由 Facebook AI Research 提供的 fvcore 模型分析工具包对 SGN-CR 的参数量与 FLOPs 进行统计。参数量 (TODO: 注意之前是否会提及, Params) 按可训练参数总数计算; (TODO: 注意之前是否会提及) FLOPs 在输入分辨率为  $256 \times 256$ 、batch size=1 条件下统计单次前向传播的浮点运算次数，并进一步按模块路径归并至 SAR 编码器、光学编码器、跨模态融合模块与解码器，以获得模块级占比分析。

### 4.2.1 SGN-CR 的整体复杂度分析

首先，对 SGN-CR 的整体参数规模和计算复杂度进行统计分析。以输入分辨率为  $256 \times 256$ 、batch size 为 1 的设置为例，对模型单次前向推理过程中的参数量 (Params) 和浮点运算量 (FLOPs) 进行统计，结果如表 4-1 所示。

表 4-1 SGN-CR 整体复杂度统计

Table 4-1 SGN-CR Overall Complexity Statistics

模型	参数量 Pramas(M)	FLOPs(G)
SGN-CR	XX.XX	XXX.X

从表中可以看出，SGN-CR 在取得较高云去除性能的同时，模型参数量和计算复杂度均处于较高水平。该复杂度在 GPU 等高性能计算环境下尚可接受，但在算力受限、功耗敏感的端侧遥感平台上仍可能带来较大的部署压力。这表明，仅从整体层面来看，SGN-CR 仍存在进一步优化模型效率的必要性。

#### 4.2.2 各功能模块的计算特性分析

为了进一步定位 SGN-CR 的计算瓶颈，本研究对模型中各主要子模块的 Parameters 和 FLOPs 进行统计，并分析其在整体模型中的占比分布情况。根据网络结构，SGN-CR 可划分为以下几个主要组成部分：SAR 编码器（SAR Encoder）、光学编码器（Optical Encoder）、注意力模块（SGAM 与 CAA）、跨模态融合模块（SAGF 与 CMCA）、解码器与输出层（Decoder & Head）。对应的模块级复杂度统计结果如表 4-2 所示。

表 4-2 SGN-CR 模块级复杂度分布

Table 4-2 SGN-CR Module-Level Complexity Distribution

模块	Parameters (M)	Parameters 占比 (%)	FLOPs (G)	FLOPs 占比 (%)
SAR Encoder	x.x	x.x	x.x	x.x
Optical Encoder	x.x	x.x	x.x	x.x
SGAM + CAA	x.x	x.x	x.x	x.x
SAGF + CMCA	x.x	x.x	x.x	x.x
Decoder & Head	x.x	x.x	x.x	x.x
总计	x.x	x.x	x.x	x.x

从表 4-2 可以清晰观察到，光学编码器在 SGN-CR 中占据了主要的计算开销。其 FLOPs 占比通常超过整体计算量的一半，明显高于其他模块。这一现象主要源于光学分支采用 Transformer 结构进行特征建模，其多头注意力机制和高维特征投影在较大空间分辨率下会引入大量矩阵乘法运算。

尽管 Transformer 结构在建模长程依赖和全局语义关系方面具有显著优势，对于大范围云遮挡区域的语义补全尤为重要，但其计算复杂度随特征维度和网络深度快速增长，成为限制模型高效推理的主要瓶颈。因此，在轻量化设计中，如何在保留必要全局建模能力的前提下，降低光学分支中注意力模块的计算成本，是需要重点解决的问题。

需要进一步指出的是，计算瓶颈的存在并不等同于该模块适合被直接轻量化。尽管光学编码器中的 Transformer 结构在整体 FLOPs 中占据主导地位，但其所承担的全局语义建模功能在厚云遮挡场景下具有不可替代性。第三章的实验结果已表明，削弱该部分建模能力会显著影响大范围云遮挡区域的语义补全效果。因此，在轻量化设计中，若对光学分支的注意力结构进行激进压缩，可能导致性能退化风险显著高于其他模块。

基于上述考虑，本章未对光学编码器的核心 Transformer 结构进行直接简化，而是从系统层面重新分配模型的计算负担，通过减少冗余的跨模态交互、压缩结构先验提取分支以及简化解码与恢复阶段来实现整体复杂度的有效下降。

与光学分支不同，SAR 编码器在 SGN-CR 中的主要作用是提供稳定的几何结构先验，引导光学分支在云遮挡区域恢复潜在地物结构。第三章的消融实验结果表明，引入 SAR 分支能够显著改善厚云区域的重建质量，尤其在结构连续性和边缘完整性方面具有明显优势。

然而，从模块级复杂度统计结果可以看出，SAR 编码器在参数量上占据了一定比例，但其 FLOPs 占比相对有限。这说明 SAR 分支的计算密集度较低，其网络结构在一定程度上存在参数冗余。在保证结构信息有效表达的前提下，通过采用更加紧凑的网络结构提取 SAR 特征，有望在不显著影响引导效果的情况下进一步降低模型规模。

上述分析表明，SAR 分支具备进行更激进压缩的潜力，这也为后续采用模态不对称的轻量化策略提供了重要依据。

值得注意的是，SGN-CR 中的 SGAM 与 SAGF 等引导与融合模块在整体参数量和 FLOPs 中所占比例极低，通常不足 5%。然而，第三章的消融实验已经验证，移除这些模块会导致模型在 PSNR、SSIM 以及 SAM 等指标上出现明显下降，尤其在复杂云遮挡区域更为显著。

这一现象表明，上述模块虽然计算代价较小，但在结构引导、噪声抑制以及跨模态信息协同方面发挥了关键作用，属于典型的“高性价比模块”。因此，在轻量化设计过程中，若对所有模块采用统一的压缩策略，容易误删这些关键组件，从而造成不必要的性能损失。

### 4.2.3 小结

综上分析可以得出以下结论：

(1) SGN-CR 的整体计算开销主要来源于光学分支的全局特征建模过程，尤其是基于 Transformer 的注意力计算。然而，光学分支在厚云遮挡条件下承担着关键的语义补全任务，其核心建模能力不宜被过度削弱。

(2) SAR 分支在云去除任务中对结构恢复具有重要贡献，但其计算密集度相对较低，在参数规模上存在进一步压缩空间。因此，可通过采用更紧凑的网络结构来降低其复杂度，而不会显著影响整体性能。

(3) SGAM 与 SAGF 等引导与融合模块计算代价较小，但在结构一致性维护和伪纹理抑制方面具有显著性能收益，应在轻量化过程中重点保留。

基于上述认识，本章轻量化设计并未对光学编码器的核心注意力结构进行激

进简化，而是从整体网络结构层面出发，通过模态不对称压缩、跨模态交互策略简化以及解码阶段轻量化等方式降低整体复杂度，并在此基础上提出一种兼顾重建性能与计算效率的轻量化模型 Lite-SGN-CR。

### 4.3 Lite-SGN-CR 轻量化网络设计

基于上一节对 SGN-CR 模型复杂度瓶颈及各功能模块特性的系统分析，可以发现原模型在不同模态分支和功能模块之间存在明显的计算负载分布不均现象。尤其是在光学分支中引入 Transformer 结构进行全局特征建模，虽然显著提升了厚云场景下的语义补全能力，但同时也成为整体计算复杂度的主要来源。另一方面，SAR 分支及部分引导与融合模块在云去除性能中发挥了关键作用，其计算代价却相对较低，表现出较高的性价比。这一分析结果表明，SGN-CR 的整体结构仍存在通过合理重构实现效率优化的空间。

在此基础上，为在不破坏原模型核心建模能力的前提下降低整体计算开销，本节提出一种基于 SGN-CR 的轻量化网络 Lite-SGN-CR。与直接对光学分支中 Transformer 结构进行激进压缩不同，Lite-SGN-CR 从系统层面出发，通过重新分配不同模态分支与功能模块的计算负担，实现整体复杂度的有效下降。具体而言，本文充分考虑 SAR 与光学模态在云去除任务中的功能差异，采用模态不对称的轻量化设计策略，在保持光学分支关键全局建模框架稳定的前提下，对 SAR 编码分支、跨模态交互方式以及解码与输出阶段进行针对性简化与重设计。

Lite-SGN-CR 的设计目标主要体现在三个方面。首先，在保证网络整体结构稳定的前提下，显著降低模型的参数规模与浮点运算量，使其更适合算力受限的端侧遥感应用场景。其次，在轻量化过程中重点保留 SAR 引导机制与层级协同融合策略，确保结构先验能够有效注入光学特征表示，避免因过度压缩导致地物结构信息丢失或伪纹理增强。最后，通过在系统层面引入合理的效率-性能权衡策略，使模型在重建精度下降可控的条件下尽可能提升推理效率，实现云去除性能与计算成本之间的平衡。

本节将围绕 Lite-SGN-CR 的整体网络架构展开详细说明。不同于第三章中以性能最优为主要目标的 SGN-CR，Lite-SGN-CR 以实际部署需求为导向，在继承原模型双分支框架与核心引导思想的基础上，对网络结构与模块实现进行系统性简化与重设计。

#### 4.3.1 Lite-SGN-CR 的整体架构设计

如图3-1和图4-1所示，Lite-SGN-CR 与原始 SGN-CR 网络在整体架构上进行了针对性的轻量化改造。在保留原有双分支多模态融合框架的基础上，Lite-SGN-CR

对各个模块采取了系统性的结构压缩策略。下面将对各部分的改进设计逐一进行分析说明。

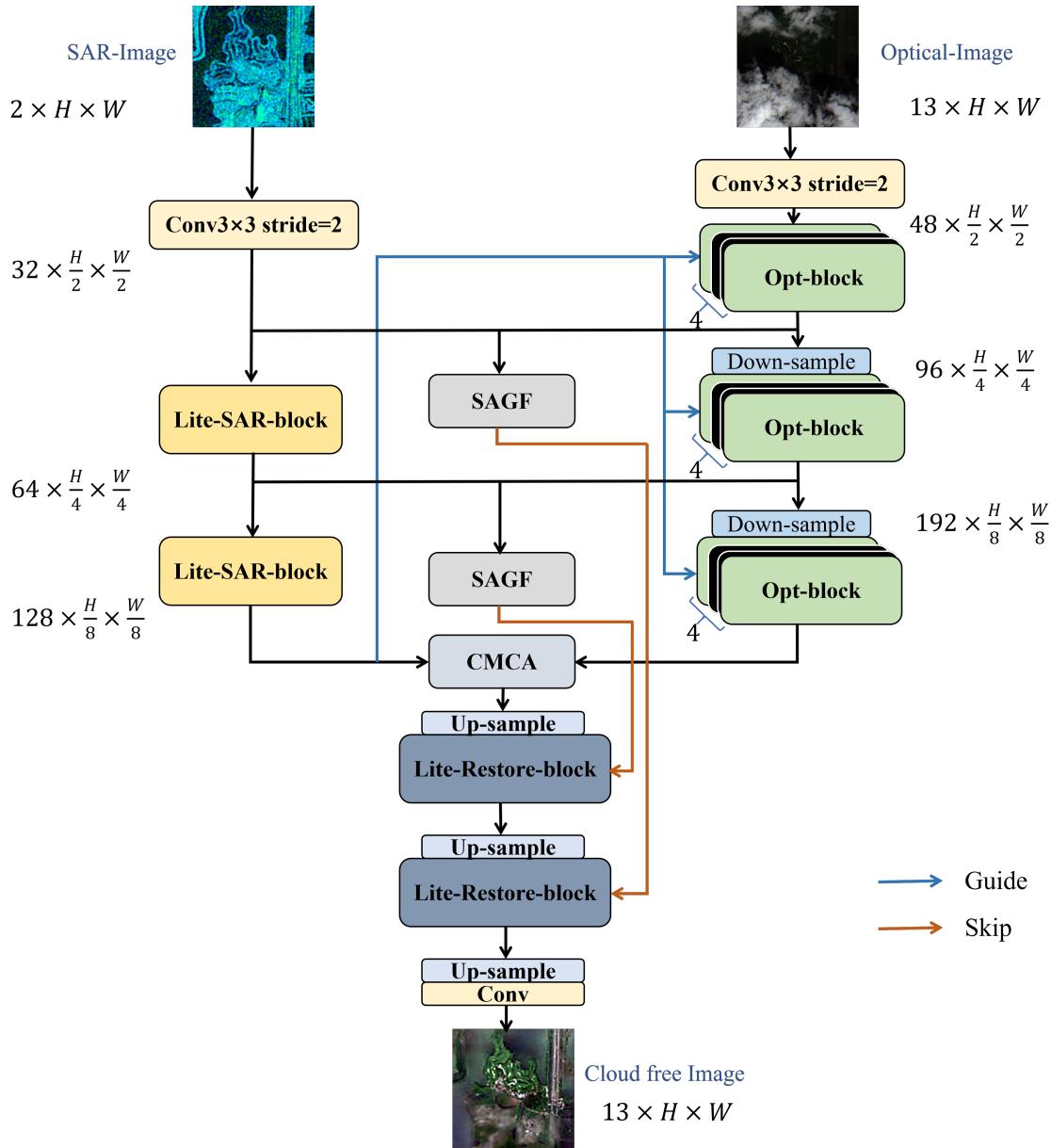


图 4-1 Lite-SGN-CR 整体网络结构示意图

Fig. 4-1 Lite-SGN-CR Overall Network Structure Diagram

### (1) 轻量化 SAR 编码分支

考虑到 SAR 与光学影像在成像机理和信息表达上的差异, Lite-SGN-CR 采用模态不对称的轻量化设计策略, 对 SAR 编码分支赋予明确的功能定位。具体而言, SAR 分支主要用于提供稳定的结构先验, 其输出特征强调地物的几何轮廓与空间连续性, 而不直接承担光谱或语义重建任务。因此, 在保证结构表达能力的前提下, 通过压缩网络深度与通道规模可有效降低 SAR 特征提取分支的计算开销。

在原 SGN-CR 中，SAR 分支依次由三个尺度的层级组成，对应的特征通道分别为 64、128、256，其中每一个尺度的层级均堆叠 3 个 ResNet 风格的 SAR-block。这种包含多个残差卷积块的三层级特征提取网络能获得较大的感受野，但同时也存在感受野重叠冗余，增加了计算开销。

(TODO：这里是否可以用公式表示一下计算开销的复杂度？)

为此，在 Lite-SGN-CR 中，将 SAR 分支重构为一个初始特征嵌入层和两个逐级下采样编码层组成的浅层结构。如图4-1所示，输入 SAR 分支的 SAR 图像首先通过一个  $3 \times 3$ 、 $\text{stride}=2$  的卷积得到  $32 \times \frac{H}{2} \times \frac{H}{2}$  的初始特征，随后仅使用两个 Lite-SAR-block 分别产生  $64 \times \frac{H}{4} \times \frac{H}{4}$  与  $128 \times \frac{H}{8} \times \frac{H}{8}$  的多尺度结构表示。其中 Lite-SAR-block 用轻量级的 DWConv (Depthwise Convolution，深度卷积) 结构替代了原 block 中 ResNet 风格的卷积模块，具体操作及原因将在下一节中探讨。同时，SAR 分支的层级数由 3 层压缩为 2 层，每层 block 的堆叠数也由 3 降为 1，大幅减少了网络深度和参数量。这一改进在降低模型复杂度的同时，仍充分保留了 SAR 分支“结构先验引导”的功能，即利用 SAR 图像提供的显著几何结构信息来指导光学分支的特征提取。

这样压缩是因为 SAR 分支仅承担结构骨架提取与引导信息，过深的层级堆叠反而导致特征冗余，因此通过减少层级数量与通道规模能在较小性能损失的前提下获得显著的复杂度收益。

## (2) 轻量化光学编码分支

相比之下，光学分支需要完成云去除后的光谱重建与语义补全任务，对特征建模能力要求更高。为此，光学编码分支在 Lite-SGN-CR 中沿用了原网络的三级尺度金字塔结构，即保留三个尺度的编码过程和关键的注意力建模能力，以维持对长程依赖与全局语义关系的基本建模能力，同时通过压缩网络规模实现复杂度控制。

在原 SGN-CR 中，光学分支在三个尺度层级，分别堆叠 8 个 Opt-block，并采用  $C$ ， $2C$ ， $4C$  的通道扩展方式进行特征建模。而在 Lite-SGN-CR 中保留了三级下采样的层级结构，但将每一尺度的模块堆叠次数由 8 减少为 4，并将三个阶段的输出通道分别由原网络的配置缩减至 48、96、192，从而在维持基本多尺度表征能力的同时显著降低注意力相关计算的总体开销。

同时，Transformer 注意力模块的超参数（例如多头注意力的头数）也相应减少，以适应收缩后的通道维度。这种通道裁剪与参数精简策略在保证模型轻量化的同时，仍然能保留原光学编码分支中关键的注意力机制。例如，Lite-SGN-CR 继续采用了 CAA 跨轴注意力等全局建模模块，只是在计算代价上进行了优化。而保留 Transformer 式注意力结构的原因是，对于大幅云遮挡的遥感图像来说，恢复纹理和语义信息需要长距离依赖建模和全局上下文信息。

通过在压缩通道的同时优化注意力模块，Lite-SGN-CR 在全局语义建模能力与模型轻量化之间取得了平衡：既避免了原网络中过多冗余特征表示，提高了效率，又确保了跨大范围图像的特征关联和语义对齐不致缺失，契合端侧算力受限条件下对效率与精度平衡的实际需求。

### (3) 轻量化跨模态特征融合

在跨模态交互方面，Lite-SGN-CR 保留了原 SGN-CR 中提出的分层次融合机制，包括空间自适应门控融合模块（SAGF）和跨模态交叉注意模块（CMCA），整体框架与图3.5(a) 和图 3.5(b) 中的原始设计一致。

在浅层，仍然采用 SAGF 模块对光学与 SAR 的浅层特征进行逐像素的门控融合，以滤除 SAR 斑点噪声并选择性注入结构信息；在深层，则利用 CMCA 模块对高层语义特征执行跨模态的注意力融合，从 SAR 分支检索补充光学分支缺失的语义细节。但与原始模型相比，Lite-SGN-CR 对这些融合模块的内部进行了轻量化改进：一方面，由于前端编码器通道数的压缩，输入到 SAGF 和 CMCA 的特征维度相应减少，直接降低了融合计算的参数量；另一方面，在 CMCA 模块中，将原先的标准  $3 \times 3$  卷积运算替换为等尺寸的深度卷积，以大幅削减卷积参数和计算开销（图 4-3 中所示）。采用深度可分离卷积能够在保持空间局部建模能力的同时，以更少的参数实现类似的特征交互效果，从而更加符合轻量化的要求。

需要强调的是，SAGF 及 SAR 引导调制模块 SGAM 在原模型中本身具有较高的性能和复杂度性价比，对抑制噪声和补全语义起着不可或缺的作用，其计算开销相对较低但对结构一致性贡献显著。因此，Lite-SGN-CR 在轻量化过程中未对上述模块的基本交互形式进行结构性删减，只是对其内部结构做简化处理，以达到在保证融合有效性的前提下尽可能减轻计算负担的目的，并且通过骨干网络通道压缩与模块堆叠次数减少的方式，降低其所依赖特征张量的维度，从系统层面实现跨模态交互开销的同步下降。相关实现细节将在后续章节的模块描述中进一步阐述，在此不再展开。

与此同时，Lite-SGN-CR 仍保持在各尺度光学特征提取阶段均引入 SAR 引导与融合机制，使结构先验能够持续注入光学特征表征，避免仅在单一尺度引导可能导致的结构不连续或细节断裂问题。

### (4) 轻量化解码器

在解码器设计方面，Lite-SGN-CR 在原模型复杂的恢复模块层级方面进行了简化，其核心改动体现在解码层级数量与模块堆叠方式的显式简化。如 3-1 所示，原 SGN-CR 的解码器由两层 Restore-Layer 组成，且每一层均堆叠多个 Restore-block，形成深层级、强建模能力的恢复网络。在该结构中，解码端不仅承担空间分辨率恢复任务，还通过多次特征变换参与语义重整与细节增强。然而，这种“多层级  $\times$  多

block”的恢复堆叠方式在高分辨率特征图上会引入大量卷积运算与特征交互，成为整体计算复杂度和推理时间的重要来源。

针对这些问题，Lite-SGN-CR 将解码流程重新设计为三阶段的逐级上采样过程，如 4-1 所示：从编码后的  $\frac{1}{8}$  尺度特征开始，依次上采样恢复到  $\frac{1}{4}$ 、 $\frac{1}{2}$  和最终的原始分辨率，并仅在其中两个过渡阶段插入 Lite-Restore-block 进行轻量的特征重建。

精简后的解码器仅使用 2 个 Restore 模块代替了原先的 6 个，大幅减少了卷积运算次数。在逐级上采样过程中，网络以更渐进的方式重建细节，避免了一步到位上采样可能出现的粗糙过渡，降低了产生伪纹理的风险。同时，通过将解码器由“多层、多 block 的重型恢复结构”重构为“两层、单 block 的逐级恢复结构”。这一设计将主要的模型容量和计算资源重新分配给编码与融合部分，使网络将重点放在多模态特征提取与融合上，从源头提取更高质量的表征，解码器的功能也明确限定为空间分辨率恢复与必要的细节校正。这种层级与堆叠数量的同步压缩，在保证重建精度下降可控的前提下，有效降低了推理复杂度，并提升了整体模型在端侧场景下的实用性。

综上所述，Lite-SGN-CR 围绕编码器、融合、解码器三个方面实施的结构压缩策略，实现了模型复杂度的全面削减和模块协同优化。在保持原网络多尺度特征表示和多模态语义融合优势的前提下，Lite-SGN-CR 大幅降低了模型的参数量和计算量，提高了推理效率和资源利用率。这种设计使模型在保证云层去除任务性能的同时，更具实际部署价值。

### 4.3.2 基于深度可分离卷积的轻量化模块设计

在上一节从网络层级深度与模块配置角度对 Lite-SGN-CR 的整体架构进行压缩后，进一步降低模型计算复杂度仍需从具体算子与模块实现层面入手。对第三章 SGN-CR 模型的计算构成进行分析可以发现，其主要计算开销集中在多尺度编码、跨模态融合及解码恢复阶段的卷积运算中，尤其是标准卷积在高分辨率特征图上的反复堆叠，对参数规模与推理效率造成了显著压力。

为此，Lite-SGN-CR 在保持原网络功能分工和信息流结构不变的前提下，引入深度可分离卷积（Depthwise Separable Convolution）作为核心轻量化算子，对多个关键模块进行系统性的卷积替换，以实现模型复杂度的进一步压缩。

#### 4.3.2.1 深度可分离卷积的原理与适用性分析

深度可分离卷积是一种将标准卷积操作分解为深度卷积（Depthwise Convolution, DWConv）和逐点卷积（Pointwise Convolution, PWConv）的轻量化卷积形式，主要用于减少模型的参数量和计算量。具体来说，深度可分离卷积将标准卷积重

新分解为两个步骤：首先 DWConv 对输入的每一个通道独立使用一个  $K \times K$  的卷积核进行空间特征提取，由于每个卷积核仅作用于一个通道，该步骤并不改变特征图的通道数。然后 PWConv 利用  $1 \times 1$  的卷积核将 DWConv 输出的特征图在深度方向上进行线性组合，该步骤主要负责通道间的信息融合与维度变换，从而实现跨通道的信息融合与特征重构。

第三章中使用的标准卷积在工作时试图同时学习“空间特征”（如  $K \times K$  范围内的边缘、纹理）和“通道特征”（如 R、G、B 通道的混合，或高层语义特征的组合），但图像的空间相关性（局部像素的关系）和跨通道相关性（特征图之间的关系）在很大程度上是独立的。既然它们是独立的，就不需要用一个  $K \times K \times C_{in}$  这种庞大的三维滤波器去同时映射它们，因此可以先把二维的空间关系提取出来，也就是 DWConv 操作，再单独处理一维的通道关系，也就是 PWConv 操作。所以解耦通道和空间的深度可分离卷积方法是一种将映射过程分解的策略。

而从图像处理角度来看，DWConv 和 PWConv 可以看作图像处理的两个经典步骤。DWConv 看做滤波器，作用仅仅是提取特征。比如，在第 1 个通道提取垂直边缘，在第 2 个通道提取水平边缘，它只关心“形状”和“纹理”，而不改变特征的数量，也不融合特征。PWConv 看做混合器，作用是特征融合。它同时看着同一个像素点上的所有通道，来判断：“这里既有垂直边缘又有红色特征，那这里可能是一个红色的柱子”，主要负责通过线性组合生成新的高层语义。

从任务特性角度看，遥感图像云去除尤其依赖于地物结构连续性、边界形态以及空间布局信息的恢复，而非对复杂高维语义的精细分类建模。因此，在多个以“结构建模”和“细节重建”为主的模块中，采用 DWConv 进行空间特征提取、再通过 PWConv 完成必要的通道融合，能够在降低计算开销的同时满足特征建模需求，具有良好的任务适配性。

通过这种分解，深度可分离卷积显著减少了卷积操作中的参数量和计算量。具体而言，假设输入特征图的尺寸为  $H \times W \times C_{in}$ ，输出特征图的尺寸为  $H \times W \times C_{out}$ ，标准卷积的计算复杂度为  $K \times K \times C_{in} \times C_{out} \times H \times W$ ，而深度可分离卷积的计算复杂度为  $K \times K \times C_{in} \times H \times W + C_{in} \times C_{out} \times H \times W$ 。当  $C_{in}$  和  $C_{out}$  较大时，深度可分离卷积能够显著降低计算开销。深度可分离卷积的以上特性，实现了在保持感受野和基本特征表达能力的同时，大幅降低参数量与浮点运算次数，使其成为轻量化网络设计中常用的基础算子。

需要强调的是，原始的 SGN-CR 网络并未使用任何深度可分离卷积，其各个分支和模块均采用标准卷积架构。而 Lite-SGN-CR 中引入深度可分离卷积作为一种系统性的轻量化替换策略，对原网络中的主要卷积模块进行了全面改造。在保持原有信息流和功能不变的前提下，我们将高计算成本的标准卷积替换为计算高

效的 DWConv + PWConv，从而大幅削减模型复杂度。

具体而言，Lite-SGN-CR 在以下关键组件中采用了 DWConv 替换原有卷积，实现模块级的轻量化设计：SAR 编码分支、跨模态融合模块、解码器模块。下面针对上述每个模块的改进逐一说明其基于 DWConv 的结构设计、适配动因及带来的效益。

#### 4.3.2.2 轻量化 SAR 编码模块（Lite-SAR-block）

在 SAR 编码分支中，Lite-SGN-CR 对原有基于 ResNet 风格卷积块的 SAR-block 进行了重点重构，引入基于深度可分离卷积的 Lite-SAR-block 模块。

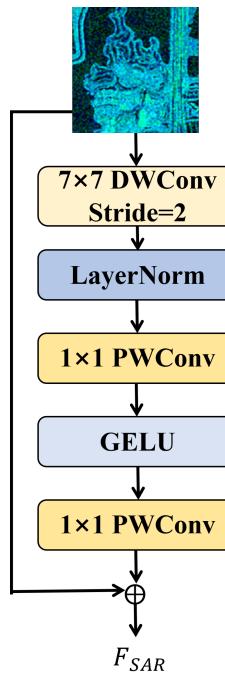


图 4-2 Lite-SAR-block 结构示意图

Fig. 4-2 Schematic diagram of lite-SAR-block structure

如图4-2所示，每个 Lite-SAR-block 由一层大核深度卷积和两层逐点卷积组成：首先采用  $7 \times 7$  的 DWConv 对输入特征进行逐通道空间建模，并通过步长为 2 的设置完成下采样操作。 $7 \times 7$  的 DWConv 能以极小的开销覆盖较大的感受野，在每个通道上提取云覆盖场景的骨架结构特征（如地物的轮廓和边缘）。随后，利用  $1 \times 1$  的 PWConv 在通道维度上对空间特征进行融合，并结合非线性激活函数增强特征表达能力。

该设计的核心动机在于 SAR 分支在 Lite-SGN-CR 中主要承担结构先验提取与引导信息提供的功能，其关注重点在于地物的几何轮廓、边界走向与空间连续性，而非复杂的高层语义推理。因此，采用大核 DWConv 即可在较低计算代价下获得

足够大的感受野，以捕获稳定的结构骨架信息；PWConv则负责对通道信息进行必要的整合，避免逐通道卷积带来的特征割裂问题。

通过以DWConv+PWConv替代原有多层标准卷积，Lite-SAR-block在显著降低参数量与计算复杂度的同时，仍能够保持对结构信息的有效建模能力，为后续光学分支的去云重建提供可靠的结构引导。

#### 4.3.2.3 基于深度可分离卷积的跨模态融合模块（Lite-CMCA）

在跨模态特征融合阶段，Lite-SGN-CR延续了原SGN-CR中的CMCA模块整体框架，但对其内部卷积运算进行了轻量化改造，形成Lite-CMCA。CMCA是用于光学-SAR特征融合的跨模态注意力模块，原始设计中，图3.5(b)，该模块在计算注意力权重前通常包含一个 $3\times 3$ 的卷积操作，用于对局部邻域特征进行建模融合。Lite-SGN-CR中将这一卷积替换为等价尺寸DWConv，构成精简的Lite-CMCA，如图4-3。

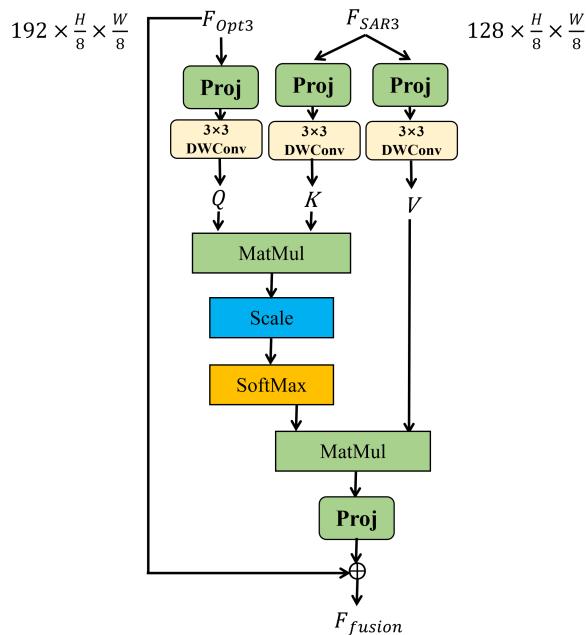


图4-3 Lite-CMCA结构示意图

Fig. 4-3 Schematic diagram of Lite-CMCA structure

在此处，DWConv承担局部模式建模的职责。对于来自光学和SAR的特征图，DWConv提取局部几何特征，而不进行通道间的线性组合。这里不使用完整的深度可分离卷积主要基于两个考虑，一是由于后续的跨模态注意力机制本质上已经完成了模态间与通道间的信息交互，如特征图的加权与相乘等操作，因此此处由PWConv承担的卷积阶段的通道融合显得冗余。二是省略PWConv使得参数量和计算量为标准卷积的 $\frac{1}{C_{out}}$ ，降至了最低，极大地减轻了融合模块的硬件开销。

由于深度卷积不混合通道，计算开销显著降低，使注意力模块能够在减少融合代价的同时完成必要的特征对齐与融合，随后直接利用提取的空间特征生成注意力图。这样的改动大幅减轻了跨模态注意力的计算负担，但并不改变原有注意力机制的作用流程，即 SAR 特征对光学特征的引导补充仍然有效。Lite-CMCA 保留了原模块的跨模态特征对齐和注意力引导功能，只是在更低复杂度下完成这些操作，从而提高了模态融合阶段的效率。

#### 4.3.2.4 轻量化解码模块（Lite-Restore-block）

在解码器设计方面，Lite-SGN-CR 对原始 SGN-CR 中的解码结构进行了结构级重构，其核心变化如下。

首先在轻量化的解码器模块中，移除了解码端的显式注意力建模模块，用轻量化的卷积结构替代了原有的重型特征交互过程。如图3-2所示，原 SGN-CR 的 Restore-block 并非简单的上采样恢复模块，而是包含 Attention 机制的重型恢复结构，其内部通过 MatMul、Scale 和 Softmax 等操作对特征进行显式的全局交互建模。该设计在提升重建精度的同时，也引入了显著的计算开销，尤其是在高分辨率特征图上执行注意力运算，会显著增加推理延迟，并不利于端侧部署。

根据前述复杂度分析可以发现，在 SGN-CR 中，跨模态语义补全与全局结构约束主要由编码端与跨模态融合模块完成，解码阶段继续引入 Attention 机制在一定程度上存在功能重叠，其对最终重建效果的边际收益相对有限。与此同时，解码端 Attention 的计算成本却随着空间分辨率的提升呈指数级增长，成为整体推理效率的重要瓶颈。

针对上述问题，设计的 Lite-SGN-CR 在解码阶段有意识地移除了 Attention 模块，将解码器的功能明确限定为空间分辨率恢复与局部细节重建。如图4-4所示，Lite-SGN-CR 的解码器采用逐级上采样的方式，从  $\frac{1}{8}$  分辨率特征开始，依次恢复至  $\frac{1}{4}$ 、 $\frac{1}{2}$  及原始分辨率。在每一级解码层中，仅使用一个 Lite-Restore-block 对上采样后的特征进行轻量化修正。

每个 Lite-Restore-block 由 DWConv + PWConv 组成的深度可分离卷积构成，其中 DWConv 负责在逐通道层面提取空间细节与结构残差信息，PWConv 则用于对通道特征进行融合与调整。相比原始解码器中的注意力建模方式，该结构能够在显著降低参数量与计算复杂度的同时，满足空间细节恢复的基本需求。

这种解码器重构策略将模型的主要计算资源进一步集中于编码端和跨模态融合阶段，使网络在源头获得更高质量的多模态特征表征，而解码端则以轻量、稳定的方式完成分辨率恢复。通过移除高开销的注意力模块并引入深度可分离卷积，Lite-SGN-CR 的解码器在保证重建精度下降可控的前提下，实现了推理效率的显著

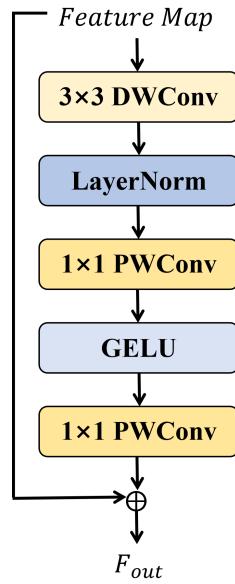


图 4-4 Lite-restore-block 结构示意图

Fig. 4-4 Schematic diagram of lite-Restore-block structure

提升，更加适合资源受限的端侧部署需求。

#### 4.3.2.5 复杂度分析

为进一步量化深度可分离卷积在 Lite-SGN-CR 中带来的复杂度优势，本文从卷积算子层面对标准卷积与深度可分离卷积的参数量进行对比分析。对于一个输入通道数为  $C_{in}$ 、输出通道数为  $C_{out}$ 、卷积核大小为  $k \times k$  的标准卷积，其参数量为：

$$\text{Params}_{\text{Conv}} = k^2 \cdot C_{in} \cdot C_{out} \quad (4-1)$$

相比之下，深度可分离卷积将该过程分解为逐通道的深度卷积和逐点卷积，其总参数量为：

$$\text{Params}_{\text{DS}} = k^2 \cdot C_{in} + C_{in} \cdot C_{out} \quad (4-2)$$

二者的参数量比值可表示为：

$$\frac{\text{Params}_{\text{DS}}}{\text{Params}_{\text{Conv}}} = \frac{1}{C_{out}} + \frac{1}{k^2} \quad (4-3)$$

在 Lite-SGN-CR 的具体配置中，卷积核大小通常取  $k = 3$ ，且各模块的输出通道数普遍大于 32，因此深度可分离卷积在单层卷积中的参数量约为标准卷积的  $\frac{1}{8} \sim \frac{1}{9}$ 。这一差异在多尺度编码与高分辨率解码阶段被进一步放大，使得整体模型的参数规模与 FLOPs 得到显著压缩。

结合前一节的结构分析可以看出，Lite-SGN-CR 在 SAR 编码模块、跨模态融合

模块以及解码恢复模块中系统性地将标准卷积替换为深度可分离卷积和 DWConv，同时配合网络层级数量与模块堆叠次数的减少，实现了算子级与结构级的协同轻量化。特别是在解码阶段，由于特征图空间分辨率较高，卷积计算开销随分辨率平方增长，上述替换策略对整体推理效率的提升尤为显著。

需要强调的是，该复杂度削减是在保持网络主干信息流与多模态融合机制不变的前提下实现的。通过将计算资源从冗余的卷积堆叠中释放出来，Lite-SGN-CR 能够将更多模型容量集中于光学分支的关键注意建模与跨模态交互阶段，从而在效率与重建性能之间取得良好平衡。后续实验章节将通过参数量、FLOPs 及推理时间的对比结果，对上述分析进行进一步验证。

### 4.3.3 渐进式去云的推理增强策略

渐进学习已被引入图像修复（TODO）和图像恢复（TODO）任务，并取得了不错的性能。在遥感图像云去除任务中，云层厚度、分布形态及其与地物结构的耦合程度具有显著差异。对于薄云或局部遮挡区域，单次前向推理通常能够获得较为理想的重建效果；然而在厚云覆盖或云与地物边界复杂的场景下，一次去云过程往往难以完全恢复被遮挡区域的光谱细节与结构连续性。这种现象在轻量化模型中更为明显，其主要原因在于模型容量受限，难以在单次推理中同时完成大尺度结构补全与细节精修。

基于上述观察，本文在 Lite-SGN-CR 的基础上进一步引入一种渐进式去云的推理增强策略，通过多次迭代逐步细化去云结果，以在计算预算允许的情况下提升重建质量。需要强调的是，渐进式云移除策略是一种可选的推理时间增强措施，而不是 Lite-SGN-CR 架构的组成部分。

#### 4.3.3.1 渐进式去云思想

渐进式去云策略的核心思想是将去云过程视为一个逐步逼近的重建问题。在第  $t$  次推理中，模型以前一次的去云结果作为输入，对残留云区域和不确定区域进行进一步修正。具体而言，设  $f(\cdot)$  表示 Lite-SGN-CR 网络， $I_{sar}$  表示对应的 SAR 图像，则渐进式去云过程可形式化表示为：

$$I^{(t)} = f(I^{(t-1)}, I_{sar}), \quad t = 1, 2, \dots, T - 1 \quad (4-4)$$

其中  $I^{(0)}$  表示原始含云光学影像， $I^{(t)}$  为第  $t$  次推理后的去云结果。通过多次迭代，网络能够在前一次结果的基础上逐步修正光谱偏差并增强结构一致性。需要强调的是，在该渐进式去云过程中，各次推理阶段共享同一组网络参数，模型参数规模保持不变，因而该策略并不引入额外的模型存储开销。

#### 4.3.3.2 复杂度与性能、效率权衡

从计算复杂度角度来看，渐进式去云策略主要在推理阶段引入额外开销。若完整执行  $T$  次前向推理，则整体浮点运算量与推理时间近似呈线性增长关系。尽管如此，该策略具有良好的可调节性：在资源受限或对实时性要求较高的场景下，可采用单次推理模式；而在对去云质量要求更高、计算预算相对宽松的应用中，则可通过增加迭代次数以获得更优的重建结果。

此外，由于 Lite-SGN-CR 本身已采用轻量化设计，其单次推理成本显著低于原始 SGN-CR 模型，因此在合理的迭代次数下（如  $T=3$ ），渐进式去云仍能够保持较为可接受的计算复杂度。

需要指出的是，渐进式去云并非 Lite-SGN-CR 的核心网络结构组成部分，而是一种可选的推理阶段增强策略。Lite-SGN-CR 的单次推理模式已能够在大多数场景下提供稳定的去云效果；渐进式去云主要适用于厚云覆盖或高精度重建需求场景，用于在不增加模型参数的前提下进一步提升去云质量。

在后续消融实验中，本文探索了单次推理与多次推理对最终结果的影响（TODO：消融实验表），定量分析渐进式去云策略在性能提升与计算开销之间的权衡关系，并且选择最佳的阶段数来平衡性能和效率。

#### 4.3.4 小结

本节围绕 Lite-SGN-CR 网络的轻量化设计展开，系统介绍了其整体架构、关键模块实现以及推理阶段的增强策略。首先，在整体架构层面，Lite-SGN-CR 在继承 SGN-CR 核心思想的基础上，通过模态不对称设计、分层跨模态融合机制，实现了在显著降低网络复杂度的同时对结构先验的有效利用。其次，在模块实现层面，本文通过引入基于深度可分离卷积的轻量化 SAR 编码器与解码器，有效减少了冗余计算，使计算资源更加集中于光学分支的关键注意建模与语义补全过程。最后，针对轻量化模型在厚云场景下可能存在的重建不足问题，本文提出了一种渐进式去云的推理增强策略，在不增加模型参数的前提下进一步提升去云质量。

通过上述设计，Lite-SGN-CR 在结构完整性、计算效率与重建性能之间实现了良好的平衡，为后续实验验证与端侧应用提供了可靠的网络基础。

### 4.4 实验结果与分析

#### 4.4.1 实验设置与评价指标

为验证所提出 Lite-SGN-CR 网络在遥感图像云去除任务中的重建性能保持能力与计算效率提升效果，本文在与第三章完全一致的数据集与实验环境下开展对

比实验，以确保实验结果的公平性与可比性。实验数据集由成对的含云光学影像、对应的无云光学影像以及配套的 SAR 辅助数据构成，覆盖不同云厚度、地物类型及成像场景，能够较为全面地评估模型在复杂条件下的去云重建能力。

在评价指标方面，本文沿用第三章中使用的峰值信噪比（PSNR）、结构相似性指数（SSIM）、光谱角映射（SAM）以及平均绝对误差（MAE）作为重建性能评价指标，用于从图像质量、结构一致性、光谱保持性和像素级误差等多个角度对去云结果进行定量分析。其中，PSNR 和 SSIM 主要反映整体重建质量与结构保真度，SAM 用于衡量光谱信息的保持能力，MAE 则用于评估重建结果与真实无云影像之间的像素级偏差。

此外，为系统评估 Lite-SGN-CR 在轻量化设计上的有效性，本文进一步引入模型参数量（Params）与浮点运算量（FLOPs）作为计算复杂度评价指标。其中，参数量用于反映模型的存储开销，FLOPs 用于衡量模型在单次前向推理过程中所需的计算量。通过将上述复杂度指标与重建性能指标结合分析，可以全面评估 Lite-SGN-CR 在保证云去除精度的同时，对计算成本的压缩效果，为后续端侧部署与实时应用提供定量依据。

所有模型均在相同的数据划分、训练策略与优化设置下进行训练，以确保对比结果的公平性与可复现性。todo 具体而言，模型训练均采用  $\times\times$  优化器，初始学习率设为  $\times\times$ ，并在训练过程中采用  $\times\times$  学习率衰减策略；模型共训练 15 个 epoch，batch size 设为 16。除网络结构差异外，其余训练超参数均保持一致。

在对比模型选择方面，本文首先以第三章提出的 SGN-CR 作为性能参考模型，用于评估轻量化设计在保持去云重建精度方面的影响。通过将 Lite-SGN-CR 与 SGN-CR 进行对比，可以直观分析在网络结构压缩与算子级轻量化后，模型在重建性能与计算复杂度之间的权衡关系。

此外，为进一步验证所提出 Lite-SGN-CR 在轻量化云去除场景下的竞争力，本文还选取了若干具有代表性的轻量化遥感图像云去除方法作为对比模型（TODO：如  $\times\times$ 、 $\times\times$  等）。这些方法在网络规模、参数量或推理效率方面具有明显的轻量化特征，能够从不同角度反映当前云去除任务中性能-效率折中的主流设计思路。通过与上述方法的系统比较，可以更全面地评估 Lite-SGN-CR 在轻量化设计下的性能优势与适用性。

## 4.4.2 Lite-SGN-CR与现有方法综合对比

### 4.4.2.1 重建性能对比

为验证Lite-SGN-CR在遥感图像云去除任务中的重建能力，本文在与第三章相同的数据集与实验设置下，将其与现有多模态云去除模型进行对比实验。评价指标包括PSNR、SSIM、SAM与MAE，从像素精度、结构一致性与光谱保真性等方面进行综合分析。

表4-3给出了各模型在重建质量与基础复杂度方面的对比结果。

表4-3 Lite-SGN-CR与不同模型重建性能对比

Table 4-3 Performance Comparison of Lite-SGN-CR with Different Models

模型	Params(M)	FLOPs(G)	PSNR(dB)↑	SSIM↑	SAM(°)↓	MAE↓
SAR-Opt-cGAN <sup>[47]</sup>	xx	xx	27.1266	0.8364	8.8707	0.03960
GLF-CR <sup>[48]</sup>	xx	xx	28.5932	0.8799	8.2512	0.02814
USSRN-CR <sup>[46]</sup>	xx	xx	28.6043	0.8532	9.1736	0.02856
GCEPANet <sup>[50]</sup>	xx	xx	30.2255	0.8964	7.7110	0.02433
SGN-CR	xx	xx	30.5503	0.8990	7.5781	0.02379
<b>Lite-SGN-CR</b>	<b>xx</b>	<b>xx</b>	<b>30.5503</b>	<b>0.8990</b>	<b>7.5781</b>	<b>0.02379</b>

从表中可以看出，Lite-SGN-CR在PSNR与SSIM指标上与SGN-CR保持接近水平，在SAM与MAE上未出现明显退化，说明轻量化设计并未显著削弱模型的重建能力。

### 4.4.2.2 复杂度与效率对比

在保证重建质量基本稳定的前提下，本文进一步分析Lite-SGN-CR在模型复杂度与推理效率方面的改进效果。统计指标包括参数量（Params）、FLOPs、推理延迟（Latency）、帧率（FPS）以及显存占用（Memory）。

表4-4展示了不同模型在复杂度与推理效率方面的对比结果。

可以观察到，Lite-SGN-CR在参数量与FLOPs上较SGN-CR明显下降，同时推理延迟显著降低、FPS提升，显存占用亦得到有效控制。这表明所提出的轻量化设计在保证结构表达能力的同时，实现了更优的计算效率。

表 4-4 Lite-SGN-CR 与不同模型复杂度与推理效率对比

Table 4-4 Comparison of Lite-SGN-CR with different model complexity and inference efficiency

模型	Params(M)	FLOPs(G)	Latency(ms)↓	FPS↑	Memory(MB)↓
SAR-Opt-cGAN <sup>[47]</sup>	xx	xx	27.1266	0.8364	8.8707
GLF-CR <sup>[48]</sup>	xx	xx	28.5932	0.8799	8.2512
USSRN-CR <sup>[46]</sup>	xx	xx	28.6043	0.8532	9.1736
GCEPANet <sup>[50]</sup>	xx	xx	30.2255	0.8964	7.7110
SGN-CR	xx	xx	30.5503	0.8990	7.5781
<b>Lite-SGN-CR</b>	<b>xx</b>	<b>xx</b>	<b>30.5503</b>	<b>0.8990</b>	<b>7.5781</b>

#### 4.4.2.3 可视化对比

为更加直观地展示不同模型的重建效果，本文选取典型云遮挡场景进行可视化对比。对比内容包括厚云区域、云边界过渡区域以及复杂地物结构区域。

图4-5通过视觉结果可以观察到，Lite-SGN-CR 在保持结构连续性的同时，有效抑制了伪影与光谱失真现象，与 SGN-CR 的视觉效果基本一致，进一步验证了轻量化设计的有效性。

#### 4.4.3 性能-效率权衡分析

为了更加直观地刻画轻量化带来的收益，本文进一步对 SGN-CR 与 Lite-SGN-CR 进行对比分析，并计算各项指标的下降比例。

表 4-5 给出了两种模型在参数规模、计算复杂度、推理延迟及重建性能方面的变化情况。

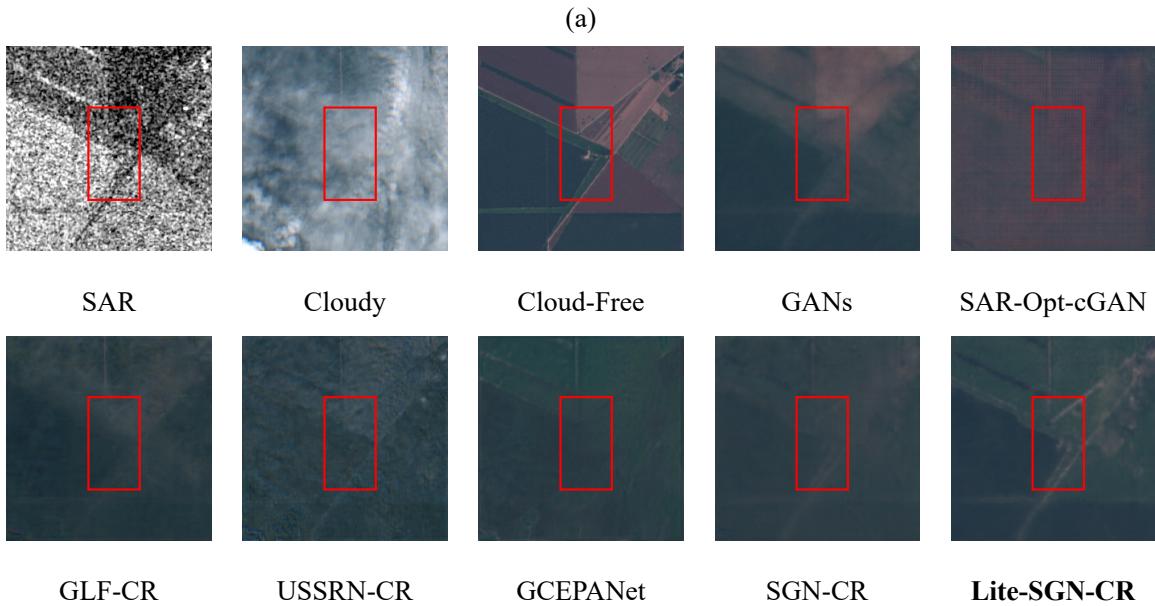
实验结果表明，在仅产生有限 PSNR 波动的情况下，Lite-SGN-CR 实现了显著的参数压缩与计算量降低。这种性能-效率权衡关系说明所提出的轻量化策略具有良好的工程应用潜力。

#### 4.4.4 轻量化设计消融实验分析

为进一步验证各轻量化策略对模型性能的具体影响，本文开展消融实验，从以下几个方面进行分析：

图4-5 SEN12MS-CR测试集上不同方法的云去除对比结果

Fig. 4-5 Comparison of cloud removal results of different methods on the SEN12MS-CR test set



- SAR 编码分支结构替换对模型性能的影响；
- 光学分支通道压缩与注意力头数调整的影响；
- 跨模态融合模块轻量化改进的贡献；
- 对齐方式调整对结构一致性的影响。

通过逐项替换与对比，可以明确各设计在降低复杂度的同时对重建质量所产生的影响，从而验证轻量化方案的合理性与必要性。

#### 4.4.5 演进式去云策略性能分析

除结构轻量化外，Lite-SGN-CR仍保留演进式去云策略以增强结构恢复的稳定性。为分析该策略对最终结果的影响，本文比较不同阶段输出结果及单阶段重建结果的性能差异。

实验结果表明，演进式优化有助于逐层恢复受遮挡区域的空间结构，从而在复杂云层场景下提升重建稳定性。

### 4.5 本章小结

表 4-5 Lite-SGN-CR 与 SGN-CR 收益对比

Table 4-5 Comparison of Lite-SGN-CR and SGN-CR performance

指标	SGN-CR	Lite-SGN-CR	Reduction(%)
Params(M)	xx	xx	<b>27.1266</b>
FLOPs(G)	xx	xx	<b>27.1266</b>
PSNR	xx	xx	<b>27.1266</b>

## 第 5 章 总结与展望

### 5.1 主要工作总结

遥感图像云去除是光学遥感数据应用中的关键问题之一，云遮挡会导致地物信息缺失和结构破坏，严重影响后续解译与分析任务的可靠性。针对传统光学方法在厚云覆盖条件下重建能力受限的问题，本文围绕多源遥感数据协同建模与深度学习方法展开研究，重点探讨了基于 SAR 与光学影像融合的遥感图像云去除方法及其轻量化设计。

本文系统分析了遥感图像在云遮挡和大气传输条件下的退化机理，明确了云去除任务在信息缺失、结构不连续以及光谱失真等方面所面临的主要挑战。在此基础上，结合 SAR 与光学遥感数据在成像机理和信息表达上的差异，阐述了利用 SAR 数据作为结构先验辅助光学影像重建的理论可行性，为后续模型设计提供了物理层面的支撑。

基于前文的分析，本文提出了一种基于 SAR 引导的多模态遥感图像云去除模型 SGN-CR。SGN-CR 通过构建双分支特征提取结构，分别对 SAR 与光学影像进行建模，并在特征层面引入跨模态引导与协同机制，使 SAR 图像中的稳定结构信息能够有效约束光学影像的重建过程。通过多层次特征融合与重建，模型在厚云遮挡条件下表现出较强的结构恢复能力和重建稳定性。实验结果表明，所提出方法在多项定量评价指标和视觉效果上均取得了较为理想的表现。

在此基础上，针对所提出模型在计算复杂度和实际部署方面的潜在限制，本文进一步开展了轻量化模型设计研究。通过对模型能力与复杂度关系的分析，结合轻量化网络结构设计思想，对原有模型结构进行了有针对性的优化，提出了轻量化版本模型。在保证重建性能基本不受明显影响的前提下，有效降低了模型参数规模和计算开销，提高了模型在资源受限场景下的应用可行性。

综上所述，本文围绕遥感图像云去除任务，从理论分析、模型设计到轻量化优化等多个层面展开研究，验证了多源遥感数据协同建模在复杂云遮挡条件下的有效性，并为后续相关研究提供了可参考的技术路线和实验基础。

### 5.2 研究展望

尽管本文围绕多源遥感图像云去除方法及其轻量化设计进行了系统研究，但受限于任务本身的复杂性和现有研究条件，仍存在若干值得进一步深入探索的问题。

(1) 当前云去除方法在面对极端厚云遮挡或完全信息缺失区域时,重建结果仍不可避免地依赖模型的统计推断能力。未来研究可进一步关注如何对这类区域的重建结果进行可信度建模与不确定性刻画,使模型不仅能够给出重建结果,还能够反映不同区域预测的可靠程度,从而为后续遥感解译任务提供更安全的使用依据。

(2) 现有方法大多在理想配准条件下开展研究,而在实际应用中,多源遥感数据往往存在分辨率差异、几何偏差或观测条件不一致等问题。如何在存在配准误差或模态差异的情况下,实现更加稳健的跨模态信息利用,是影响云去除方法工程可行性的重要问题,有待进一步研究。

(3) 当前云去除模型的训练与评估主要依赖有限的数据集,其云型分布、地物类型和观测条件仍难以覆盖真实应用中的复杂情况。未来可通过引入更具代表性的实观测数据,或探索弱监督、自监督等学习范式,降低模型对高质量成对数据的依赖,以提升方法在不同场景下的泛化能力。

(4) 从应用角度看,云去除结果的价值最终体现在对下游遥感任务的支撑作用上。未来研究可进一步探讨云去除结果对地物分类、变化检测等任务的影响机制,并在此基础上构建面向应用目标的联合优化框架,使云去除过程更好地服务于实际遥感应用需求。

## 参考文献

- [1] Dosovitskiy A. An image is worth 16x16 words: Transformers for image recognition at scale[J]. arXiv preprint arXiv:2010.11929, 2020.
- [2] Liu Z, Lin Y, Cao Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows[C]//Proceedings of the IEEE/CVF international conference on computer vision. 2021: 10012-10022.
- [3] Wang H, Zhu Y, Green B, et al. Axial-deeplab: Stand-alone axial-attention for panoptic segmentation[C]//European conference on computer vision. Cham: Springer International Publishing, 2020: 108-126.
- [4] King M D, Platnick S, Menzel W P, et al. Spatial and temporal distribution of clouds observed by MODIS onboard the Terra and Aqua satellites[J]. IEEE transactions on geoscience and remote sensing, 2013, 51(7): 3826-3852.
- [5] Zhu X, Gao F, Liu D, et al. A modified neighborhood similar pixel interpolator approach for removing thick clouds in Landsat images[J]. IEEE Geoscience and Remote Sensing Letters, 2011, 9(3): 521-525.
- [6] Yu C, Chen L, Su L, et al. Kriging interpolation method and its application in retrieval of MODIS aerosol optical depth[C]//2011 19th international conference on geoinformatics. IEEE, 2011: 1-6.
- [7] Xu M, Jia X, Pickering M, et al. Thin cloud removal from optical remote sensing images using the noise-adjusted principal components transform[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2019, 149: 215-225.
- [8] Wan M, Li X. Removing thin cloud on single remote sensing image based on SWF[C]//2016 IEEE International Conference of Online Analysis and Computing Science (ICOACS). IEEE, 2016: 397-400.
- [9] He M, Wang B, Sheng W, et al. Thin cloud removal method in color remote sensing image[J]. Opt. Tech, 2017, 43: 503-508.
- [10] Tao C, Fu S, Qi J, et al. Thick cloud removal in optical remote sensing images using a texture complexity guided self-paced learning method[J]. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: 1-12.
- [11] Yang Q, Wang G, Zhao Y, et al. Multi-scale deep residual learning for cloud removal[C]//IGARSS 2020-2020 IEEE International Geoscience and Remote Sensing Symposium. IEEE, 2020: 4967-4970.

- [12] Shao Z, Pan Y, Diao C, et al. Cloud detection in remote sensing images based on multiscale features-convolutional neural network[J]. IEEE Transactions on Geoscience and Remote Sensing, 2019, 57(6): 4062-4076.
- [13] Zhang K, Zuo W, Chen Y, et al. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising[J]. IEEE transactions on image processing, 2017, 26(7): 3142-3155.
- [14] Wen X, Pan Z, Hu Y, et al. An effective network integrating residual learning and channel attention mechanism for thin cloud removal[J]. IEEE Geoscience and Remote Sensing Letters, 2022, 19: 1-5.
- [15] Ma D, Wu R, Xiao D, et al. Cloud removal from satellite images using a deep learning model with the cloud-matting method[J]. Remote Sensing, 2023, 15(4): 904.
- [16] Zhao Y, Shen S, Hu J, et al. Cloud removal using multimodal GAN with adversarial consistency loss[J]. IEEE Geoscience and Remote Sensing Letters, 2021, 19: 1-5.
- [17] Toizumi T, Zini S, Sagi K, et al. Artifact-free thin cloud removal using gans[C]//2019 IEEE International Conference on Image Processing (ICIP). IEEE, 2019: 3596-3600.
- [18] Wen X, Pan Z, Hu Y, et al. Generative adversarial learning in YUV color space for thin cloud removal on satellite imagery[J]. Remote Sensing, 2021, 13(6): 1079.
- [19] Singh P, Komodakis N. Cloud-gan: Cloud removal for sentinel-2 imagery using a cyclic consistent generative adversarial networks[C]//IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium. IEEE, 2018: 1772-1775.
- [20] Mo Y, Li C, Zheng Y, et al. DCA-CycleGAN: Unsupervised single image dehazing using dark channel attention optimized CycleGAN[J]. Journal of Visual Communication and Image Representation, 2022, 82: 103431.
- [21] Ye H, Xiang H, Xu F. Cycle-Gan network incorporated with atmospheric scattering model for dust removal of martian optical images[J]. IEEE Transactions on Geoscience and Remote Sensing, 2024.
- [22] Ma J, Chen Y, Pan J, et al. SCT-CR: A synergistic convolution-transformer modeling method using SAR-optical data fusion for cloud removal[J]. International Journal of Applied Earth Observation and Geoinformation, 2024, 130: 103909.
- [23] Zou Z, Chen L, Jiang X. Spectral-temporal low-rank regularization with deep prior for thick cloud removal[J]. IEEE Transactions on Geoscience and Remote Sensing, 2023, 62: 1-16.
- [24] Lin J, Huang T Z, Zhao X L, et al. Robust thick cloud removal for multitemporal remote sensing images using coupled tensor factorization[J]. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: 1-16.
- [25] Li H, Zhang L, Shen H. A principal component based haze masking method for visible images[J]. IEEE Geoscience and Remote Sensing Letters, 2013, 11(5): 975-979.

- [26] Lin C H, Tsai P H, Lai K H, et al. Cloud removal from multitemporal satellite images using information cloning[J]. IEEE transactions on geoscience and remote sensing, 2012, 51(1): 232-241.
- [27] Cheng Q, Shen H, Zhang L, et al. Cloud removal for remotely sensed images by similar pixel replacement guided with a spatio-temporal MRF model[J]. ISPRS journal of photogrammetry and remote sensing, 2014, 92: 54-68.
- [28] Li X, Shen H, Zhang L, et al. Recovering quantitative remote sensing products contaminated by thick clouds and shadows using multitemporal dictionary learning[J]. IEEE Transactions on Geoscience and Remote Sensing, 2014, 52(11): 7086-7098.
- [29] Zhang Q, Sun F, Yuan Q, et al. Thick cloud removal for sentinel-2 time-series images via combining deep prior and low-rank tensor completion[C]//2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS. IEEE, 2021: 2675-2678.
- [30] Imran S, Tahir M, Khalid Z, et al. A deep unfolded prior-aided RPCA network for cloud removal[J]. IEEE Signal Processing Letters, 2022, 29: 2048-2052.
- [31] Xu M, Jia X, Pickering M, et al. Cloud removal based on sparse representation via multitemporal dictionary learning[J]. IEEE Transactions on Geoscience and Remote Sensing, 2016, 54(5): 2998-3006.
- [32] Li X, Wang L, Cheng Q, et al. Cloud removal in remote sensing images using nonnegative matrix factorization and error correction[J]. ISPRS journal of photogrammetry and remote sensing, 2019, 148: 103-113.
- [33] Ji T Y, Chu D, Zhao X L, et al. A unified framework of cloud detection and removal based on low-rank and group sparse regularizations for multitemporal multispectral images[J]. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: 1-15.
- [34] Ji S, Dai P, Lu M, et al. Simultaneous cloud detection and removal from bitemporal remote sensing images using cascade convolutional neural networks[J]. IEEE Transactions on Geoscience and Remote Sensing, 2020, 59(1): 732-748.
- [35] Zou X, Li K, Xing J, et al. Pmaa: A progressive multi-scale attention autoencoder model for high-performance cloud removal from multi-temporal satellite imagery[J]. arXiv preprint arXiv:2303.16565, 2023.
- [36] Gao J, Zhang H, Yuan Q. Cloud removal with fusion of SAR and Optical Images by Deep Learning[C]//2019 10th International Workshop on the Analysis of Multitemporal Remote Sensing Images (MultiTemp). IEEE, 2019: 1-3.
- [37] Chen S, Zhang W, Li Z, et al. Cloud removal with SAR-optical data fusion and graph-based feature aggregation network[J]. Remote Sensing, 2022, 14(14): 3374.

- [38] Li Y, Fu R, Meng X, et al. A SAR-to-optical image translation method based on conditional generation adversarial network (cGAN)[J]. Ieee Access, 2020, 8: 60338-60343.
- [39] Hao Y, Jiang W, Liu W, et al. Selecting information fusion generative adversarial network for remote-sensing image cloud removal[J]. IEEE Geoscience and Remote Sensing Letters, 2023, 20: 1-5.
- [40] Li C, Liu X, Li S. Transformer meets GAN: Cloud-free multispectral image reconstruction via multisensor data fusion in satellite images[J]. IEEE Transactions on Geoscience and Remote Sensing, 2023, 61: 1-13.
- [41] Zhang J, Zhou J, Lu X. Feature-guided SAR-to-optical image translation[J]. Ieee Access, 2020, 8: 70925-70937.
- [42] Ebel P, Xu Y, Schmitt M, et al. SEN12MS-CR-TS: A remote-sensing data set for multimodal multitemporal cloud removal[J]. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: 1-14.
- [43] Gu P, Liu W, Feng S, et al. HPN-CR: Heterogeneous Parallel Network for SAR-Optical Data Fusion Cloud Removal[J]. IEEE Transactions on Geoscience and Remote Sensing, 2025.
- [44] Darbaghshahi F N, Mohammadi M R, Soryani M. Cloud removal in remote sensing images using generative adversarial networks and SAR-to-optical image translation[J]. IEEE Transactions on Geoscience and Remote Sensing, 2021, 60: 1-9.
- [45] Xu M, Deng F, Jia S, et al. Attention mechanism-based generative adversarial networks for cloud removal in Landsat images[J]. Remote sensing of environment, 2022, 271: 112902.
- [46] Wang Y, Zhang B, Zhang W, et al. Cloud removal with SAR-optical data fusion using a unified spatial-spectral residual network[J]. IEEE Transactions on Geoscience and Remote Sensing, 2023, 62: 1-20.
- [47] Grohnfeldt C, Schmitt M, Zhu X. A conditional generative adversarial network to fuse SAR and multispectral optical data for cloud removal from Sentinel-2 images[C]//IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium. IEEE, 2018: 1726-1729.
- [48] Xu F, Shi Y, Ebel P, et al. GLF-CR: SAR-enhanced cloud removal with global-local fusion[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2022, 192: 268-278.
- [49] Han S, Wang J, Zhang S. Former-CR: A transformer-based thick cloud removal method with optical and SAR imagery[J]. Remote Sensing, 2023, 15(5): 1196.
- [50] Zhou Q, Wang X, Fang J, et al. GCEPANet: A Lightweight and Efficient Remote Sensing Image Cloud Removal Network Model for Optical-SAR Image Fusion[J]. Information Fusion, 2025: 104090.

## 作者简介

### 1. 基本情况

胡美琳，女，四川人，2001年1月出生，重庆邮电大学计算机学院计算机科学与技术专业2023级硕士研究生。

### 2. 教育和工作经历

2019.09~2023.06 成都工业学院网络与通信工程学院，本科，专业：网络工程

2023.09~2026.06 重庆邮电大学计算机科学与技术学院，硕士研究生，专业：计算机科学与技术

### 3. 攻读学位期间的研究成果

#### 3.1 发表的学术论文和著作

#### 3.2 申请（授权）专利

[1] 罗小波，胡美琳. 基于 CNN-Transformer 异构架构与 SAR 引导的光学图像云去除方法：中国，发明专利申请 202511875861.9 [P]. (已受理，尚未公开) .

#### 3.3 参与的科研项目及获奖

[1] 胡美琳. 硕士一年级学业三等奖学金, 2023.

[2] 胡美琳. 硕士二年级学业二等奖学金, 2024.

## 致 谢

行文至此，心中感慨良多。忽而想起儿时盛夏老家的泥巴路，两旁种有高大的柏树，枝干挺拔，偶有一阵风过，树叶沙沙摩擦，投出的树荫伴着蝉鸣摇摆。晃动的斑驳树荫间一个小孩，坐着小板凳，支着膝盖放着敞开的铁皮文具盒，嘴里飞速念着文具盒里印着的九九乘法表，也不觉耳边的聒噪蝉鸣讨厌。恍惚一梦，竟已匆匆二十年。二十年漫漫求学路，从小山村的蜿蜒泥巴路，到县城公交开过的水泥路，再到火车才能到的省城铁路，居然已经走出了这么远。

感谢一路上遇见的每一位老师，你们让我成为今天的我，你们不同的人生阅历，让我看到成长的意义。感谢罗小波老师，三年前同样的春夏交接时，在你的办公室我们第一次相见，我成为你课题组的一员。你知识渊博但对我们从不苛责，在课题上给了我很大的自由和发挥空间。

感谢我的家人，你们从不过问太多，但总为我兜底。尤其感谢我的母亲，这二十年读书路，我若努力有一份、辛苦有一份，那你必是努力十份、辛苦十份。若不是你的坚持，我还留在老家的泥巴路上面，走不出，走不脱。你总是很豁达，在我迷茫低落时说出一句，“怕什么，还有你老娘呢。”你让我有勇气一步步走到今天，并且还将一步步走到更远。

感谢我的朋友们，一路上遇见你们是我的意外之喜。我很庆幸我有一群不管多久不联系都不生疏的朋友，不管我们走出去多远，都不会忘记拥有共同回忆的地方。感谢在我学业不顺时，为我出谋划策的你；感谢在我不善交际时，率先伸出手的你；感谢在我絮叨吐槽时，比我更先义愤填膺的你。感谢每一个出现在我生活中的你。

对自己和未来，心中有万千期盼与不安，皆化作六年级同样盛夏毕业时，郭老师给出的最后一句评语，“须知学如逆水行舟，不进则退。”愿自己，不要丢失学习的初心，不要丢失持续学习的勇气。坐在泥巴路板凳上的小孩，还会走到更多、更广阔的路上去。

最后，谨向参与本论文评审与审阅的各位老师致以诚挚的感谢，感谢你们在百忙之中对本文提出宝贵的意见与指导。