

文章编号 1004-924X(2025)18-2996-12

## 遥感图像双模态融合去云方法

柳长源\*, 曹 青, 刘金凤

(哈尔滨理工大学 测控技术与通信工程学院, 黑龙江 哈尔滨 150080)

**摘要:**针对合成孔径雷达(SAR)图像结合光学图像数据融合去云过程中不能准确处理云与背景的差异问题,本文提出了一种基于双分支结构的SAR图像与光学遥感图像结合的去云网络。在特征提取阶段,引入多尺度注意力机制,有效捕捉图像中全局与局部信息,为后续的融合和去云操作提供更具有代表性的特征信息表示。重新设计局部融合分支与差分支的并联结构,并通过门控机制动态平衡两者贡献,充分挖掘两者的互补性,更加细化光学有云图像中的云层边缘,从而恢复出更加精细的无云光学图像。采用位置感知增强的Swin Transformer将局部特征密集连接,使网络模型面对复杂环境具有更好的鲁棒性。所提模型的峰值信噪比(PSNR)和结构相似性指数(SSIM)较目前最优算法分别高出0.8331 dB和0.0246。证明本文算法在图像去云任务中对比其他方法具有更优性能。

**关键词:**遥感图像;合成孔径雷达图像;光学图像;多模态;数据融合;云去除

**中图分类号:**TP751 **文献标识码:**A

**doi:**10.37188/OPE.20253318.2996

**CSTR:**32169.14.OPE.20253318.2996

## A bimodal fusion method for remote sensing images to cloud removal

LIU Changyuan\*, CAO Qing, LIU Jinfeng

(College of Measurement and Control Technology and Communication Engineering,  
Harbin University of Science and Technology, Harbin 150080, China)

\* Corresponding author, E-mail: liuchangyuan@hrbust.edu.cn

**Abstract:** In order to solve the problem that the difference between cloud and background could not be accurately handled in the process of cloud removal from Synthetic Aperture Radar (SAR) images combined with optical image data, this paper proposed a cloud removal network based on the combination of SAR images and optical remote sensing images with a double-branched structure. In the feature extraction stage, a multi-scale attention mechanism was introduced to effectively capture the global and local information in the image, and provide a more representative representation of feature information for subsequent fusion and cloud removal operations. The parallel structure of the local fusion branch and the differential branch was redesigned, and the contribution of the two was dynamically balanced through the gating mechanism, so as to fully explore the complementarity of the two, and refine the cloud edge in the optical cloudy image, so as to restore a more fine cloudless optical image. The location-aware enhanced Swin Transformer was used to densely connect local features, so that the network model had better robustness

收稿日期:2025-05-27;修订日期:2025-07-04.

基金项目:黑龙江省自然科学基金(No. LH2023E086);黑龙江省交通运输厅科技项目(No. HJK2024B002)

in the face of complex environments. The Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM) of the proposed model are 0.833 1 dB and 0.024 6 higher than those of the current optimal algorithm, respectively. It is proved that the proposed algorithm has better performance than other methods in the image removal task.

**Key words:** remotely sensed imagery; synthetic aperture radar image; optical images; multimodal; data fusion; cloud removal

## 1 引 言

卫星遥感技术已成为获取地理信息数据的重要手段<sup>[1]</sup>,但云层的存在会导致数据失真,增加了对遥感图像分析解译的难度,进而限制了光学遥感图像的使用价值与发展潜力<sup>[2-4]</sup>,所以遥感图像去云是地球观测领域的核心挑战。去云方法<sup>[5-13]</sup>可分为基于光谱<sup>[5]</sup>、空间<sup>[9]</sup>及时间<sup>[12]</sup>三类。基于光谱的方法通过分析遥感图像的光谱特征,利用光谱信息恢复云层遮挡区域的地表特征。但由于厚云会笼罩多光谱的所有波段,无法针对厚云进行去云任务。基于空间的方法依赖图像的局部空间结构,通过修复或替换云覆盖区域的内容达到去云目的,只适用于薄云或局部遮挡的场景,可能会因为空间信息不足导致修复结果失真。基于时间的方法常见于利用多时相遥感数据的时空连续性,通过对比或融合不同时间点的影像数据去云,对数据获取频率和时空一致性要求较高。

光学遥感图像具有高分辨率、视觉特征明显及可解释性强等优点,但极易受到云雾天气影响,致使图像质量下降,增加了对地表信息判别区分和编码解码的难度,大大限制了其应用价值与潜力。同比单独光学遥感影像,合成孔径雷达(Synthetic Aperture Radar, SAR)图像具备全天时、全天候的优势<sup>[14]</sup>,与光学遥感图像融合可以取长补短,提供有效可靠的地表信息,提升遥感数据的实际使用价值,尤其在复杂云覆盖场景下,实现跨模态数据的鲁棒融合,恢复出高质量地表信息具有重要意义<sup>[15]</sup>。因此近年来将SAR图像与光学图像结合去云的方法备受关注,随着各类网络架构的兴起,便有很多学者将这些网络架构作为主干网络使SAR图像加入遥感图像去云任务行列中。Liu(2018)等人<sup>[16]</sup>利用改进的循环一致性损失生成对抗网络(cycle GAN)将

SAR图像直接转换成光学图像,代替受云层污染的光学遥感图像,但结果与真实图像有较大的差异。Gao(2021)等人<sup>[17]</sup>利用CNN建立多时相SAR和光学图像之间的关系,在没有大量数据的情况下得到去云效果,反映地表信息变化,并未使用多时态数据加强模型的泛化能力。龚循强等人<sup>[18]</sup>提出结合区域能量、滚动引导滤波以及双通道单位连接脉冲耦合神经网络将SAR图像与光学图像进行融合,在抵抗SAR图像的噪声干扰的同时高效提取影像的高频细节和纹理,只能针对全色影像进行去云任务,无法用于多光谱。由于不同传感器图像之间存在成像机制不同,数据不一致和配准差异,导致SAR图像与光学图像在复杂环境中的边缘提取不准确。

针对上述问题,本文提出一种采用双分支结构的SAR图像与光学遥感图像结合的去云方法。通过引入多尺度注意力结构获取多尺度特征信息,提取不同感受野的空间特征,有效捕捉图像中的局部和全局信息;重新设计SAR图像与光学图像的对齐融合方式,充分挖掘两者的互补性;在Swin Transformer的窗口注意力中引入可学习位置感知增强操作,加强关注局部窗口内的空间分布,增强对复杂环境中云层边界恢复的能力。

## 2 成像原理

### 2.1 光学遥感成像原理

光学遥感属于被动遥感方式成像,通过由多个相机组成的光学传感器接收地表反射的太阳光,利用不同地物对光波的反射特性一次性生成二维图像,光学遥感图像可以显示地物的可见特征,能够为水体、植被等地物信息提供丰富的色彩信息<sup>[19]</sup>。

当太阳光照射地表时,不同地物反射不同波

长的光波,辐射亮度如公式(1)所示。

$$L = \frac{E \cdot \rho}{\pi}, \quad (1)$$

其中: $E$ 为入射光辐射强度, $\rho$ 为地物反射率。

光学传感器接收反射光波,记录光波的强度和波长信息,像元亮度如公式(2)所示。

$$B = \int_{\lambda_1}^{\lambda_2} L(\lambda) \cdot S(\lambda) d\lambda, \quad (2)$$

其中: $B$ 为像元亮度, $\lambda_1$ 和 $\lambda_2$ 为传感器的光谱范围, $L(\lambda)$ 为波长为 $\lambda$ 的辐射亮度, $S(\lambda)$ 为传感器的光谱响应函数。

最后通过数据处理和矫正,生成地表的有色光学图像。

## 2.2 合成孔径雷达成像原理

合成孔径雷达属于主动遥感方式成像,通过发射电磁波脉冲并接收目标回波信号,利用雷达与目标之间的相对运动并进行相关解调压缩处理。SAR成像示意图如图1所示。

雷达在不同位置上接收同一地物信息的回波信号,此运动可将短雷达天线等效为长雷达天线,在提高雷达系统的方位分辨率的同时确保距离不影响分辨率<sup>[20]</sup>。SAR成像不同于传统光学遥感一次成像,需要距离向聚焦和方位向聚焦,雷达的视线方向为距离向,前进方向为方位向。

## 3 本文方法

本文的整体结构如图2所示,首先,将SAR图像和有云光学图像输入到像素重排模块,以

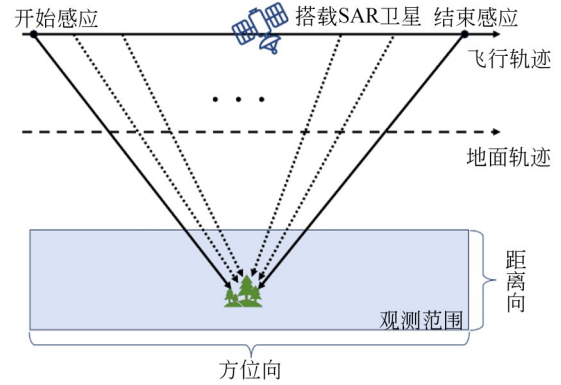


图1 SAR成像示意图

Fig. 1 Schematic diagram of SAR imaging

维度变换的角度进行像素重排,以张量的形式进行特征提取,分别得到关于SAR图像和光学图像的特征数据信息,将两种数据信息交叉引入到分支并联结构(Diff-LFF Branch, DLB)中,双模态特征的补偿性(局部融合分支)与差异性(差分支)并联结构可以将特征进一步的对齐融合和验证,将各层得到的光学特征信息进行拼接,紧接着输入到全局特征模块中在全局层面上进行特征细化,最后经过上采样后添加最初的光学图像,得到高质量光学无云影像。为了高效完成无云影像的重建任务,该网络在特征提取的过程中加入多尺度注意力机制,并在金字塔残差结构中构建了分支并联结构,最后引入可学习位置感知增强操作更好地适应复杂环境的去云任务。

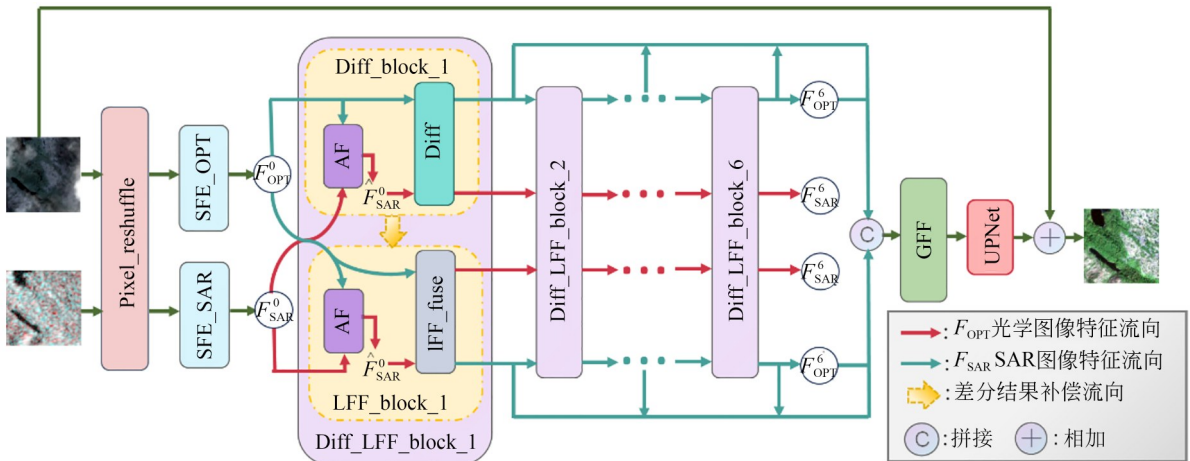


图2 整体结构图

Fig. 2 Overall structure diagram

### 3.1 多尺度注意力特征提取

随着网络的不断深入,大多数方法忽略了网络对多尺度注意力的适应性,在 ResNet 特征提取网络中加入了多尺度注意力结构,使其关注所需关键信息,忽略未决特征的无关噪声,以便更好地利用多尺度信息。根据光学图像和 SAR 图像的特征信息不同的特点,确保特征提取不充分和特征丢失这类问题得到解决,在最新的视觉多尺度注意研究<sup>[21]</sup>的启发下,引入多尺度注意力网络(Multi-Scale Attention Network, MSAN)解决上述问题。本文多尺度注意力特征提取模块结构如图 3 所示。

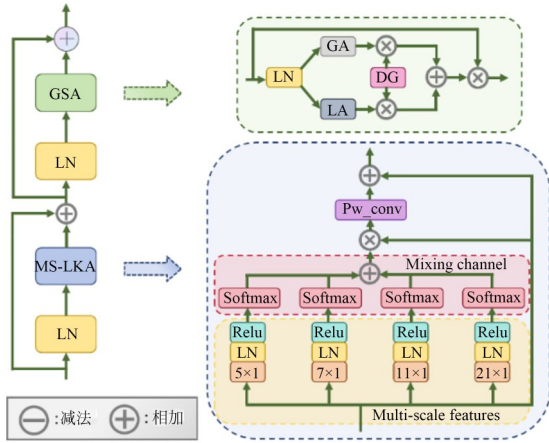


图 3 多尺度注意力特征提取模块

Fig. 3 Multi-scale attention feature extraction module

MSAN 采用了多尺度大核注意力单元(Multi-Scale Large-Kernel Attention, MS\_LKA)和门控空间注意力单元(Gated Spatial Attention, GSA),这两个单元在特征提取的基础卷积块中加入不同尺度的深度核,可以提取不同感受野范围的空间特征,有效捕捉图像中全局和局部信息,并将简单空间注意和门控线性单元整合搭配注意门控空间单元中,达到在复杂性的情况下去除线性层并捕获局部连续性的效果。

将特征信息  $X$  输入到多尺度注意力模块中,具体过程如式(3)所示:

$$\begin{cases} Z = \text{LN}(X) \\ Z_1 = X + \text{MS\_LKA}(f_i(Z)) \\ Z_2 = \text{LN}(Z_1) \\ X = Z_1 + \text{GSA}(Z_2) \end{cases}, \quad (3)$$

其中:  $\text{LN}(\cdot)$  是四层归一化因子, MS\_LKA 是多尺

度的大核分解模块, GSA 是动态门控空间注意力单元,  $f_i(\cdot)$  表示第  $i$  点卷积, 始终保持维度不变。

将特定信息映射到  $C, H, W$  中, 分别用四组不同尺寸的大核卷积进行处理, 利用加权机制动态调整 MS\_LKA 的输出, 以此放大高响应区域。最后利用注意门控空间单元实现动态门控机制, 减少参数和计算。

### 3.2 分支并联结构的特征融合与重建

光学图像和 SAR 图像在物理特征、噪声类型、空间分辨率和语义表达上存在显著差异, 单纯依赖特征对齐无法充分挖掘互补性。本文在基线模型 M3R-CR<sup>[22]</sup> 的基础上提出 DLB 结构, 显式构建了双模态特征的补偿性(局部融合分支)与差异性(差分支)并联结构, 并通过门控机制动态平衡了两者的贡献。

#### 3.2.1 局部融合分支

局部融合分支如下图所示, SAR 图像作为源特征, 光学图像作为参考特征, 两特征在经过上一模块后首先进行对齐操作, 将对齐后的特征输入到关于 SAR 图像和光学的双通道中, 双通道分别将每个流中最后的状态视为前一个特征的残差, 将前后特征拼接在一起视为对前一个特征的矫正, 使新的特征图更加对齐。各通道局部融合后的特征的计算如公式(4)所示:

$$LFF^i = \text{Conv}(\text{concat}(F^{i-1}, F^i)), \quad (4)$$

其中:  $\text{concat}$  表示在通道维度上直接拼接两个张量,  $\text{Conv}$  表示使用  $1 \times 1$  卷积对拼接后的特征进行处理, 恢复原有的通道数。

局部融合分支由两个具有权重共享的密集连接流构建而成, 内部使用相同卷积核, 通过这种结构增加模型的鲁棒性, 防止由于特征不对齐而导致的伪变化。具体操作流程如图 4 所示。

Swin Transformer 将局部融合后的特征进行展开和转置, 帮助模型更好地理解两特征之间的关系。通过局部融合分支可以将 SAR 图像像素

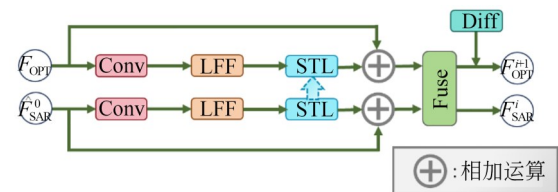


图 4 局部融合分支模块

Fig. 4 Local fusion branch module

点均匀分布在光学有云图像的对应像素点上,更加细化光学有云图像中的云层边缘,从而恢复出更加精细的无云光学图像。

### 3.2.2 差分支

差分支的目的是处理光学特征和 SAR 特征之间的差异,用于提取模态差异信息。光学特征中的纹理信息映射在 SAR 特征中效果并不明显,使用差值计算能够直观地反映出两者之间的差异<sup>[23]</sup>,突出纹理细节的差异,强制网络关注光学图像相较于 SAR 图像独有的特征细节。计算如公式(5)所示:

$$\begin{aligned} \text{Diff}_{\text{feat}} &= C_{\text{diff}}(|\text{OPT}_f - \text{SAR}_f|) \\ C_{\text{diff}} &\in R^{G_0 \times G_0 \times 3 \times 3}, \end{aligned} \quad (5)$$

其中,绝对值运算 $|\cdot|$ 确保差异值非负,抑制符号干扰。

将得到的差异信息使用 $3 \times 3$ 卷积核组成的多个卷积层进行特征提取,每个卷积层都可以捕捉到不同层次的特征表示,卷积层最开始捕捉的差异特征是一些边缘信息,后续卷积层则进一步提取更加复杂的纹理和形状。采用密集连接进行处理,将每一层的输出结果都连接起来,可以丰富历史信息并且极大程度上解决了梯度消失问题,进一步增强特征差异性表示。差分支模块如图 5 所示。

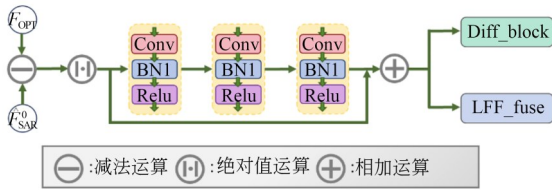


图 5 差分支模块

Fig. 5 Difference branching module

差分支处理结果分别输入到下一层差分块和融合到光学特征中,融合操作可以使光学特征更好地吸收 SAR 特征中有用信息的同时还能保留自身的特征优势,动态捕捉和利用两者的互补信息,提高网络对不同数据的适应性和泛化能力。

### 3.2.3 局部融合分支与差分支分配特征权重

局部融合分支和差分支通过门控机制分配分支特征权重,如果遇到云覆盖区域侧重差分支,清晰区域侧重局部融合分支,避免简单加权导致伪影出现。首先将局部融合分支得到的特

征和差分支得到的特征通过卷积操作计算注意力得分,注意力得分反映了每个特征对最终输出的重要程度,注意力得分计算如公式(6)所示:

$$F_{\text{score}} = \text{conv}(LFF_{\text{feat}}, \text{Diff}_{\text{feat}}) \in R^{2G_0 \times H \times W}, \quad (6)$$

其中: $F_{\text{score}}$ 表示注意力得分, $\text{conv}$ 表示卷积操作, $LFF_{\text{feat}}$ 和 $\text{Diff}_{\text{feat}}$ 分别是局部融合分支和差分支的特征信息。

将注意力得分通过激活函数进行归一化,得到注意力权重,注意力权重反映了每个特征对最终输出的贡献程度,注意力权重计算如公式(7)所示:

$$\text{GateWeight} = \sigma(\lambda_{\text{fuse}}(F_{\text{score}})) \in R^{1 \times H \times W}, \quad (7)$$

其中: $\sigma(\cdot)$ 为激活函数, $\lambda_{\text{fuse}}$ 为 $1 \times 1$ 卷积。

分配权重后融合局部融合分支和差分支,计算如公式(8)所示:

$$\begin{aligned} F_{\text{combined}} &= LFF_{\text{feat}} \cdot \text{GateWeight} + \\ &\quad \text{Diff}_{\text{feat}} \cdot (1 - \text{GateWeight}), \end{aligned} \quad (8)$$

通过动态滤波和门控修正得到新的双模态特征,动态滤波器可以基于光学与 SAR 特征动态生成卷积核,调整滤波参数以突出云覆盖区域的 SAR 差异特征或光学补偿特征,确保跨模态特征在关键区域的互补性,具体计算过程如公式(9)所示:

$$\begin{cases} \kappa_{\text{SAR}} = \text{DFG}(\text{OPT}, \text{SAR}) \in R^{\frac{G_0}{2} \times ks^2} \\ F_{\text{SAR}}^{\text{filt}} = \text{conv2d}(F_{\text{SAR}}, \kappa_{\text{SAR}}) \end{cases}, \quad (9)$$

其中: $ks = 5$ , $F_{\text{SAR}}^{\text{filt}}$ 为修正后的 SAR 特征。

通过差异去驱动门控修正得到双模态特征,关于更新后的光学特征和 SAR 特征计算公式(10)和公式(11)所示:

$$\begin{aligned} \text{OPT}_{\text{new}} &= \\ \text{OPT} &+ (F_{\text{SAR}}^{\text{filt}} - \text{OPT}) \cdot \sigma(\zeta_{\text{SAR}} + F_{\text{combined}}), \end{aligned} \quad (10)$$

$$\begin{aligned} \text{SAR}_{\text{new}} &= \text{SAR} + (\text{OPT}_{\text{new}} - F_{\text{SAR}}^{\text{filt}}) \cdot \\ &\quad \sigma(\zeta_{\text{OPT}} + F_{\text{combined}}), \end{aligned} \quad (11)$$

其中, $\zeta_{\text{SAR/OPT}}$ 为 SAR 和光学分支的门控卷积层。

局部融合分支与差分支损失权重同样通过注意力权重得到,将其注意力权重作为初始损失权重。在反向传播时,记录两分支的梯度向量,定义梯度冲突指标如公式(12)所示:

$$\cos \theta = \frac{\nabla L_{\text{LFF}} \cdot \nabla L_{\text{Diff}}}{\|\nabla L_{\text{LFF}}\| \cdot \|\nabla L_{\text{Diff}}\| + \epsilon}, \quad (12)$$

其中: $\nabla L_{\text{LFF}}$ 为局部融合分支的损失梯度向量, $\nabla L_{\text{diff}}$ 为差分支的损失梯度向量, $\cdot$ 为向量点积运算, $\|\cdot\|$ 为向量的 L2 范数, $\epsilon$ 是一个很小的正数,防

止除零。

若  $\cos \theta < 0$ , 表明梯度方向冲突, 则根据梯度冲突强度触发权重修正机制调整损失权重。动态权重调整如公式(13)所示:

$$\begin{cases} \lambda_{LFF} = GateWeight \cdot (1 - \gamma \cdot |\cos \theta|) \\ \lambda_{Diff} = (1 - GateWeight) \cdot (1 + \gamma \cdot |\cos \theta|) \end{cases}, \quad (13)$$

其中:  $\gamma \in [0, 1]$  为冲突敏感系数, 用于控制梯度冲突对权重的影响程度。

最后将动态权重与双分支的损失结合, 实现损失权重自适应增强。双分支权重自适应增强损失函数如公式(14)所示:

$$L_t = \lambda_{LFF} \cdot L_{LFF} + \lambda_{Diff} \cdot L_{Diff}. \quad (14)$$

权重变化与云覆盖比例、边缘敏感度直接相关, 通过注意力权重得到双分支的损失权重, 无需额外参数, 减少模型负担。通过与初始信息进行残差连接可以确保低阶特征有效结合高阶特征, 将特征信息在多阶层之间有效连通, 促进图像的基本信息与高阶信息的融合, 极大程度地解决梯度消失的问题。

### 3.3 位置感知增强的 Transformer

Swin Transformer<sup>[24]</sup>作为强大的视觉模型, 在诸多任务中以其独特的窗口注意力机制在处理局部特征关系中展现出了优异的性能, 传统的注意力机制主要基于相对位置偏差来考虑位置信息, 本文在此基础上进行了扩展, 引入了位置感知增强操作 (Enhanced Location Awareness, ELA), 使其具有更高的灵活性和可学习性, 能够更精确地引导注意力在窗口内的分布, 从而更好

地建模相对位置关系。

普通自注意力计算<sup>[24]</sup>如公式(15)所示:

$$Attention = \text{softmax} \left( \frac{QK^T}{\sqrt{d}} + Bias_{re} \right) V, \quad (15)$$

其中:  $Q, K, V$  分别是查询矩阵, 键值矩阵, 值矩阵,  $Bias_{re}$  是根据元素间相对位置计算的偏置项,  $d$  是键值的维度。

Swin Transformer 采用窗口划分后, 每个窗口的位置坐标局部连续, 通过生成窗口内坐标的网络, 计算出所有位置之间的相对坐标, 进一步得到相对位置索引, 通过该索引在相对位置偏差表中查找对应的相对位置偏差, 从而在注意力计算中加强相对位置关系。为增强模型对位置信息的敏感性, 将可学习位置信息直接融合到查询中, 使查询在最初阶段就携带了位置特征, 更直接地影响注意力的计算。

位置感知增强操作主要在窗口自注意力中, 在查询向量  $Q$  和键向量  $K$  上加入可学习的位置嵌入操作, 具体计算过程如公式(16)所示:

$$\begin{cases} Q_i = Q + \text{emb}(N_i) \\ K_i = K + \text{emb}(N_i) \end{cases}, \quad (16)$$

其中:  $\text{emb}$  是位置感知信息嵌入操作,  $N$  是窗口内的元素数量,  $i$  对应元素的依赖位置。

普通注意力机制在注意力计算后, 位置信息才添加进来用作结果微调, 侧重于捕捉相对位置之间的关系, 而位置感知增强是在查询阶段就深度融合位置信息, 可以自动调整位置信息的表示, 更好地适应任务, 注意力机制如图 6 所示。

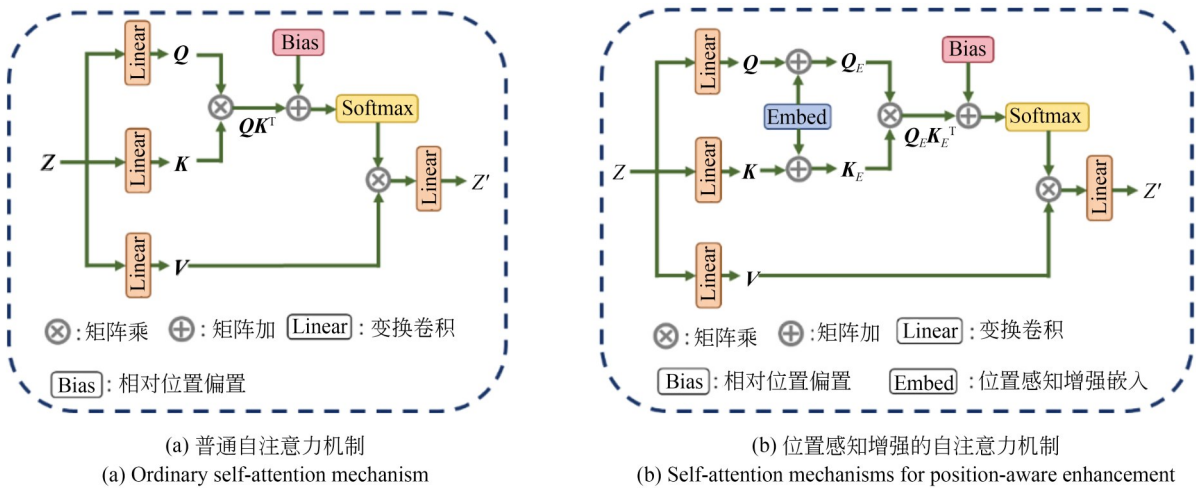


图 6 注意力机制

Fig. 6 Attention mechanisms

将增加位置感知后的  $Q$  和  $K$  代入到原始 Swin Transformer 的注意力公式中,具体计算如公式(17)所示:

$$\text{Attention}_{RS} = \text{softmax} \left( \frac{Q_E K_E^T}{\sqrt{d}} + \text{Bias}_{re} \right) V, \quad (17)$$

其中,  $Q_E$  和  $K_E$  分别是增加位置感知后的  $Q$  和  $K$ 。

图(a)使普通自注意力机制,位置感知增强的自注意力机制如右图(b)所示。位置感知增强操作使窗口在面对数据变化时更加稳定,确保注意力计算和特征处理能够准确反映图像中云层与地表信息的真实结构和位置关系,加强分辨云层和地表信息距离衰减关系的能力,增强网络模型的鲁棒性,进而提高去云的可靠性和稳定性。

### 3.4 损失函数

传统的损失函数通常对整个图像区域采用相同的权重,无法区分云层区域和非云层区域。EC-Charbonnier 损失函数<sup>[25]</sup>通过引入云层掩码和权重系数,对云层区域的预测误差进行加权,使网络在完成去云任务时更加关注云层区域的预测误差,在优化过程中,既能关注全局的去云效果,又能细致处理云层边缘和覆盖区域的细节。

首先计算预测无云图像与真实图像的 Charbonnier 函数,计算如公式(18)所示:

$$L_C = \left( \left( \hat{X} - X \right)^2 + \epsilon^2 \right)^\alpha, \quad (18)$$

其中:  $\hat{X}$  为预测无云图像,  $X$  为真实无云图像,  $\epsilon$  是一个很小的正数,防止预测无云图像与真实无云图像差异为零时,损失函数的导为无穷大,  $\alpha = 0.45$ , 用于控制损失函数的鲁棒性。

通过加权损失将云层区域的损失值放大,而非云层区域的损失值保持不变,加权损失计算如公式(19)所示:

$$L_w = (1 + EC \times M) \times L_C, \quad (19)$$

其中:  $EC$  为权重系数,增加云层区域的损失权重,在进行去云任务时更加关注云层区域的预测误差。  $M$  为云层掩码,设置云层区域为 1,非云区域为 0。  $(\cdot)$  为权重矩阵。

计算所有像素的加权损失的平均值,得到最终的损失值,计算如式(20)所示:

$$L = \frac{1}{N} \sum_{i=1}^N L_w, \quad (20)$$

其中,  $N$  为图像的像素总数。

通过 EC-Charbonnier 损失函数,不仅考虑预测误差的大小,还通过权重矩阵对云层区域的预测误差进行加权,优化本网络在去云任务中的性能。

## 4 实验与分析

### 4.1 实验数据集

本文实验采用数据集 M3M-CR<sup>[22]</sup> 为建立的模型提供数据训练、验证和实验测试,该数据集具有全球采样的高分辨率光学观测数据,主要使用高分辨率 Planet Scope 数据构建并由哨兵 1 号 (Sentinel-1) 卫星的 SAR 数据作为提升去云性能的辅助信息,针对全球各大洲选取了 780 个非重叠 AOI 地理空间位置进行采样,在不同季节中共收集了 63 000 个样本,其中光学图像的空间分辨率为 3 m,与目前现有的公开去云系列数据集相比,填补了高分辨率去云数据集的空缺,满足高分辨率去云影像数据集的要求。此数据集满足云层的多样性,由半透明到密集,由轻云到重云覆盖,按云层的覆盖率等比例分布。使用了 53 506 个训练样本、6 494 个验证样本及 3 000 个测试样本。在训练集上训练深度学习网络模型,并将本方法与其他现有方法进行对比。

本文实验采用数据集 SEN12MS-CR-TS<sup>[26]</sup> 作为验证本文方法的泛化能力,此数据集具有多模态、多时相的全球分布式数据集,其中的数据通过哥白尼任务的哨兵 1 号 (Sentinel-1) 卫星和哨兵 2 号 (Sentinel-2) 卫星从各个大陆中选取了 53 个非重叠 AOI 区域,在不同季节中共收集 15 578 个包含不同类型真实云层的样本。与 M3M-CR 不同的是,SEN12MS-CR-TS 的云覆盖率分布为双峰分布,即 0% 云遮挡率和 100% 云遮挡率数量最多,该数据集的平均云遮挡率为 50%。其中光学图像和 SAR 图像的空间分辨率都为 10 m,在此数据集中使用了 11 490 个训练样本、1921 个验证样本及 2167 个测试样本。

### 4.2 实验环境参数设置及评价指标

实验硬件环境采用单张 NVIDIA GeForce RTX 4060Ti 显卡,显存为 16G,使用 Ubuntu20.04 操作系统,Pytorch1.10 深度学习框架,在实验过程中采用 Adam 优化器来训练模型。基础学的学习

率设置为 $2 \times 10^{-4}$ ,采用基于步长衰减方法来调整学习率,训练轮数设置为 30,批处理大小设置为 4。

在实验中使用的评价指标主要包括两个方面,一是基于两幅图像对应像素之间的平均误差度量,即峰值信噪比(Peak Signal-to-Noise Ratio, PSNR)来评估去云后的图像的失真度及图像的质量,PSNR 值越高说明恢复质量越好。二是从图像的亮度、对比度和结构相似性方面,反映人眼对视觉质量的感知,即结构相似度(Structural Similarity Index, SSIM)来评估去云后的图像与原图像在亮度、对比度和结构方面的相似

度,SSIM 的取值范围在 0~1 之间,在此区间内,值越大,表示相似度越高。

4.3 实验

4.3.1 消融实验及分析

为了验证各个模块的有效性,本文设计了消融实验,在基线模型上依次引入分支并联结构(DLB)、位置感知增强的 Swin Transformer(ELA)及多尺度注意力网络(MSAN),在 M3M-CR 数据集上进行一系列的消融实验并根据 PSNR 和 SSIM 两项指标验证模块的有效性,实验结果如表 1 所示。

表 1 M3M-CR 测试集上各模块消融实验  
Tab. 1 Each module on the M3M-CR test set was ablatively solidified

| Ablation | Baseline | DLB | ELA | MSAN | PSNR/dB  | SSIM    |
|----------|----------|-----|-----|------|----------|---------|
| Baseline | ✓        |     |     |      | 30.529 6 | 0.910 5 |
| 1        | ✓        | ✓   |     |      | 30.927 5 | 0.921 4 |
| 2        | ✓        |     | ✓   |      | 30.783 1 | 0.920 1 |
| 3        | ✓        |     |     | ✓    | 30.749 6 | 0.919 3 |
| 4        | ✓        | ✓   |     | ✓    | 31.080 4 | 0.931 4 |
| 5        | ✓        | ✓   | ✓   | ✓    | 31.362 7 | 0.935 1 |

根据表 1 所示的实验结果,这三项改进均展现了不错的效果,首先依次单独引入 DLB 分支结构、ELA 操作、MSAN 模块,验证各模块的提升有效性;其次同时引入 DLB 分支结构及 MSAN 模块,在 PSNR 和 SSIM 这两项指标分别提升 0.550 8 dB,0.020 9,最后在 Swin Transformer 中引入 ELA 位置感知增强操作,使局部特征连接更加密集,实验结果显示各项指标数据都得到了进一步的提升;本文方法与基线模型相比,其指标效果分别提升 0.833 1 dB,0.024 6。这成功地验证了各个模块对提升网络性能的有效性。

由表 2 本文方法与基线模型进行参数量与计算量的对比,本文方法虽然参数量和计算量较基

线模型有一定的增加,但增加幅度很小,不影响实际中的使用。

4.3.2 对比实验分析

为了验证本网络的去云处理性能,从 PSNR 和 SSIM 两个评价指标与其他 4 种主流去云模型进行对比,本次对比实验使用 M3R-CR 数据集进行定量分析,实验结果如表 3 所示,其中,最优结果以粗体显示,次优结果以下划线显示。

SAR-opt-cGAN<sup>[27]</sup>, Pix2pix<sup>[28]</sup> 和 SpA GAN<sup>[29]</sup> 基于在平原地区且体量小的数据集进行训练,在本实验数据集中的泛化能力表现不佳。通过表 3 可以看出,本文方法以 M3R-CR 作为基线模型,在同一数据集的情况下,与基线模型在 PSNR 高出 0.833 1 dB,在 SSIM 更接近于 1。本网络生成结果的客观评价均优于其他 4 种办法,体现了本文方法在去云场景中的优越性。

为对比不同方法的去云效果,验证本文方法的有效性,选取 7 对光学图像和同一地区的 SAR 图像进行可视化分析,将同一区域临近时

表 2 与基线模型的参数量与计算量对比结果  
Tab. 2 Comparison of params and FLOPs for baseline

| Method | Params/M | FLOPs/G |
|--------|----------|---------|
| M3R-CR | 43.53    | 174.94  |
| Ours   | 47.85    | 202.05  |

表 3 M3M-CR 测试集上主流网络整体性能定量对比结果

Tab.3 Quantitative comparison of the overall performance of mainstream networks on the M3M-CR dataset

| Method       | PSNR/dB  | SSIM    |
|--------------|----------|---------|
| SAR-opt-cGAN | 27.470 0 | 0.839 0 |
| Pix2pix      | 25.739 5 | 0.792 8 |
| M3R-CR       | 30.529 6 | 0.910 5 |
| SpA GAN      | 26.220 0 | 0.839 0 |
| Ours         | 31.362 7 | 0.935 1 |

间的光学无云图像作为目标图像,结果如图 7 所示。

图(7)中,(a)至(c)分别是 SAR 图像、光学有云图像和光学无云图像,(d)至(h)分别是 SAR-opt-cGAN 模型、Pix2pix 模型、M3R-CR 模型、SpA GAN 模型及本文方法的可视化结果。光学图像虽然被云层污染,但根据 SAR 图像提供的地表信息,本文方法能较好地恢复被污染的区域细节。

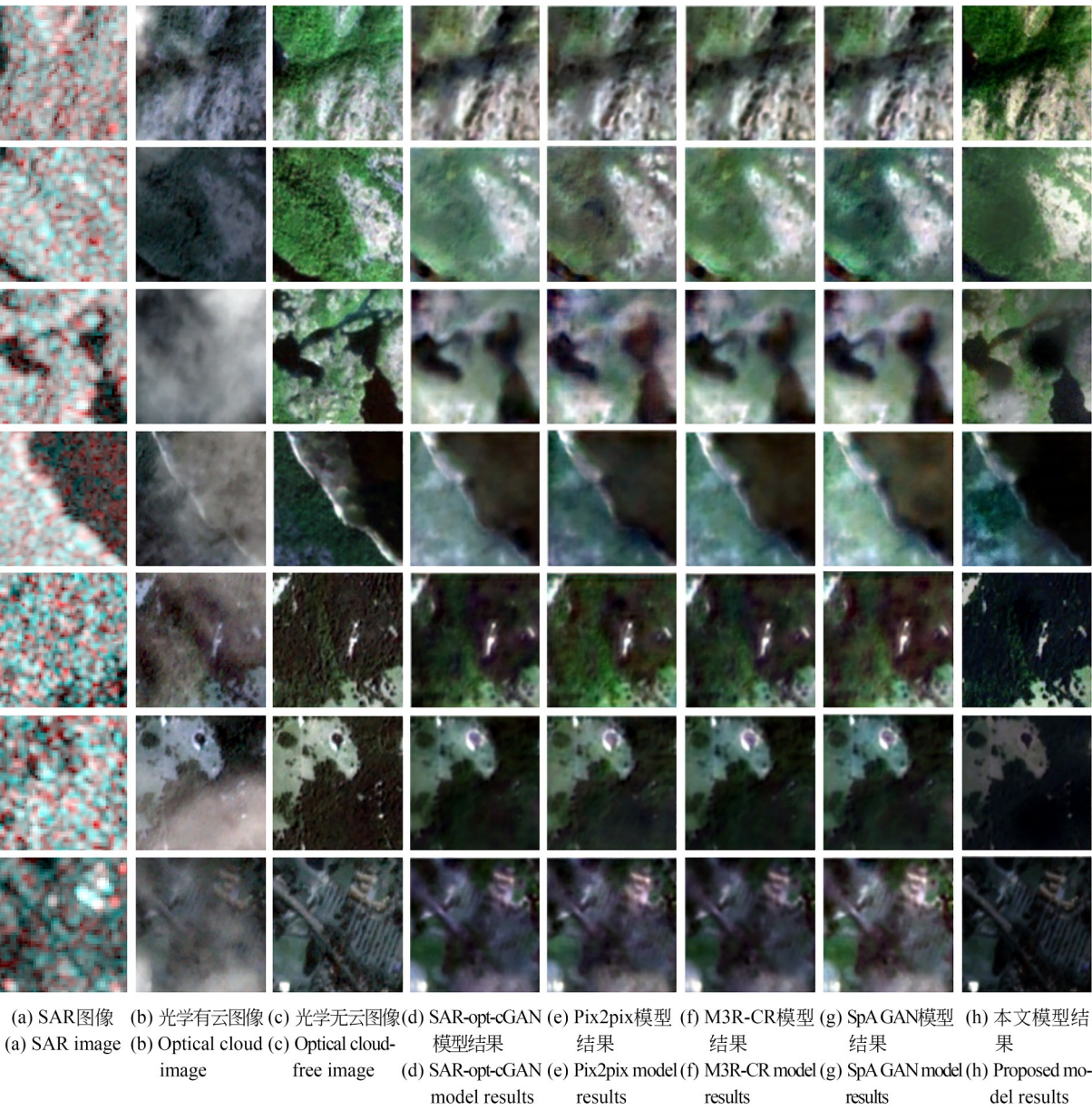


Fig. 7 Comparison of cloud removal results of different networks in the M3M-CR dataset

4.3.3 泛化性能测试分析

为了进一步验证本文方法的泛化性能,使用 SEN12MS-CR-TS 数据集进行测试,结果如表 4 所示。

表 4 SEN12MS-CR-TS 测试集上主流网络整体性能定量对比结果

Tab. 4 Quantitative comparison results of the overall performance of mainstream networks on the SEN12MS-CR-TS dataset

| Method       | PSNR/dB  | SSIM    |
|--------------|----------|---------|
| SAR-opt-cGAN | 27.157 7 | 0.650 3 |
| Pix2pix      | 24.739 5 | 0.528 3 |
| M3R-CR       | 30.161 8 | 0.907 0 |
| SpA GAN      | 25.682 2 | 0.813 9 |
| Ours         | 30.934 7 | 0.921 4 |

本文方法在 SEN12MS-CR-TS 测试集上的 PSNR 达到了 30.934 7,与其他网络相比,PSNR 达到最高,高出基线模型 M3R-CR 0.772 9 dB。本文方法在第二个数据集达到较好效果,这说明本文方法在光学遥感图像结合 SAR 图像去云应用中具有较好的泛化能力。为更好地观测本文方法的泛化效果,将本文方法在 M3M-CR 测试集与 SEN12MS-CR-TS 测试集上泛化效果做对比,如表 5 所示。

表 5 本文方法在 M3M-CR 测试集与 SEN12MS-CR-TS 测试集泛化对比结果

Tab. 5 Generalized and compared between the M3M-CR dataset and the SEN12MS-CR-TS dataset

| Dataset       | PSNR/dB  | SSIM    |
|---------------|----------|---------|
| M3M-CR        | 31.362 7 | 0.935 1 |
| SEN12MS-CR-TS | 30.934 7 | 0.921 4 |

本文方法引入多尺度注意力网络增强捕捉全局和局部信息,忽略不相关的背景信息,加强特征提取能力,并且使用差分支突出云边缘的差

异性,利用位置感知增强的 Swin Transformer 准确反映图像中云层与地表信息的真实结构和位置关系,提升了去云的可靠性和稳定性,增强了模型的泛化能力。

5 结 论

本文充分利用 SAR 图像和光学图像的互补信息,通过改进 M3R-CR,提出了并联结构的 SAR 图像结合光学图像去云方法。为了更准确的识别和提取云层及其覆盖的地表信息,本文在特征提取网络中引入了大核多尺度注意力,提升特征提取网络表达能力,整合更大范围的局部信息,为后续的融合和去云操作提供更具有代表性的特征信息表示。同时为了加强云层边缘及覆盖区域的差异性,更准确地确定云层的位置并进行去除,设计了差分支与局部融合并联结构,针对不同的局部区域的特征进行细致处理,避免因全局融合而丢失重要细节。在 Swin Transformer 中引入位置感知增强操作,使模型在计算注意力时更好地考虑元素之间的相对位置关系,从而提高对不同数据集和场景的泛化能力,不会因为场景的变化而出现性能大幅下降的情况。在数据集 M3M-CR 上进行消融实验和对比实验,PSNR 值达到 31.362 7 dB,SSIM 值也达到 0.935 1,在数据集 SEN12MS-CR-TS 上进行泛化实验 PSNR 值和 SSIM 值分别达到 30.934 7 dB 和 0.921 4。消融实验结果验证本文提出的改进策略具备有效性;对比实验的客观评价结果和主观视觉效果对比其他方法均有提升;泛化实验验证了本文方法能够在不同数据集上表现出较好的稳定性。

作者贡献声明:

柳长源:论文方法构思和撰写,论文审核;  
曹青:实验设计及数据分析,论文编辑;  
刘金凤:为项目研究成果成功发表而去争取并获得资助。

参考文献:

[1] CHANG J X, GAO X J, YANG Y W, *et al.* Object-oriented building contour optimization methodol-

ogy for image classification results *via* generalized gradient vector flow snake model[J]. *Remote Sensing*, 2021, 13(12): 2406.

- [2] GAO X J, ZHANG G B, YANG Y W, *et al.* Two-stage domain adaptation based on image and feature levels for cloud detection in cross-spatiotemporal domain[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2024, 62: 5610517.
- [3] 王睿. 基于参考影像信息融合的遥感影像云检测和云去除算法研究[D]. 上海: 上海大学, 2020.  
WANG R. *Research on Cloud Detection and Cloud Removal Algorithm of Remote Sensing Image based on Reference Image Information Fusion*[D]. Shanghai: Shanghai University, 2020. (in Chinese)
- [4] 徐萌, 王思涵, 郭仁忠, 等. 遥感影像云检测和云去除方法综述[J]. 计算机研究与发展, 2024, 61(6): 1585-1607.  
XU M, WANG S H, GUO R Z, *et al.* Review of cloud detection and cloud removal methods for remote sensing images[J]. *Journal of Computer Research and Development*, 2024, 61(6): 1585-1607. (in Chinese)
- [5] ZHANG C, LI H F, SHEN H F. A scattering law based cirrus correction method for Landsat 8 OLI visible and near-infrared images[J]. *Remote Sensing of Environment*, 2021, 253: 112202.
- [6] GLADKOVA I, GROSSBERG M D, SHAHRIAR F, *et al.* Quantitative restoration for MODIS band 6 on aqua[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2012, 50(6): 2409-2416.
- [7] WANG L X, WANG Q M. Fast spatial-spectral random forests for thick cloud removal of hyperspectral images[J]. *International Journal of Applied Earth Observation and Geoinformation*, 2022, 112: 102916.
- [8] ZHOU B X, WANG Y. A thin-cloud removal approach combining the cirrus band and RTM-based algorithm for Landsat-8 OLI data[C]. *IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium. July 28 - August 2, 2019, Yokohama, Japan*. IEEE, 2019: 1434-1437.
- [9] YU C, CHEN L F, SU L, *et al.* Kriging interpolation method and its application in retrieval of MODIS aerosol optical depth[C]. *2011 19th International Conference on Geoinformatics. June 24-26, 2011, Shanghai, China*. IEEE, 2011: 1-6.
- [10] MAALOUF A, CARRE P, AUGEREAU B, *et al.* A bandelet-based inpainting technique for clouds removal from remotely sensed images[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2009, 47(7): 2363-2371.
- [11] SHEN H F, ZHANG L P. A MAP-based algorithm for destriping and inpainting of remotely sensed images[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2009, 47(5): 1492-1502.
- [12] XIE S, LIU L Y, YANG J N. Enhanced Landsat surface reflectance prediction considering land cover change by using an ensemble of spectro-temporal and spectro-spatial predictions[J]. *Advances in Space Research*, 2022, 69(7): 2697-2710.
- [13] XU M, JIA X P, PICKERING M, *et al.* Cloud removal based on sparse representation via multi-temporal dictionary learning[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2016, 54(5): 2998-3006.
- [14] 陶亮亮, 王雨琦. 基于雷达和光学遥感数据的云污染区域光谱重建算法[J]. 农业工程学报, 2024, 40(17): 216-223.  
TAO L L, WANG Y Q. Reconstructing spectral information in cloud contaminated regions using object-oriented approach[J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2024, 40(17): 216-223. (in Chinese)
- [15] RATHORE K, MAURYA A K, SINGH D. Potential assessment of SAR and optical data with machine learning to monitor temporal changes in tall vegetation[J]. *Journal of the Indian Society of Remote Sensing*, 2024, 52(12): 2807-2824.
- [16] LIU L, LEI B. Can SAR images and optical images transfer with each other? [C]. *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium. July 22-27, 2018, Valencia, Spain*. IEEE, 2018: 7019-7022.
- [17] GAO J H, YUAN Q Q, LI J, *et al.* Cloud removal with fusion of high resolution optical and SAR images using generative adversarial networks[J]. *Remote Sensing*, 2020, 12(1): 191.
- [18] 龚循强, 方启锐, 侯昭阳, 等. 一种光学与合成孔径雷达影像融合去云方法[J]. 光学学报, 2024, 44(24): 193-204.  
GONG X Q, FANG Q R, HOU Z Y, *et al.* A method for cloud removal using optical and synthetic aperture radar image fusion[J]. *Acta Optica Sinica*, 2024, 44(24): 193-204. (in Chinese)

- [19] 强德霞. 遥感技术[M]. 重庆: 重庆大学出版社, 2024.  
QIANG D X. *Remote Sensing Technology* [M]. Chongqing: Chongqing University Press, 2024. (in Chinese)
- [20] 王超, 石爱业, 陈嘉琪. 高分辨率遥感影像变化检测[M]. 北京: 人民邮电出版社, 2018.  
WANG C, SHI A Y, CHEN J Q. *High Resolution Remote Sensing Image Change Detection*[M]. Beijing: Posts & Telecom Press, 2018. (in Chinese)
- [21] WANG Y, LIU Y, ZHAO S J, *et al.* CAMixerSR: Only Details Need More "Attention" [C]. 2024 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. June 16-22, 2024, Seattle, WA, USA. IEEE, 2024: 25837-25846.
- [22] XU F, SHI Y L, EBEL P, *et al.* Multimodal and multiresolution data fusion for high-resolution cloud removal: a novel baseline and benchmark [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2023, 62: 5600615.
- [23] CHEN P, ZHANG B, HONG D F, *et al.* FCCDN: Feature constraint network for VHR image change detection[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2022, 187: 101-119.
- [24] LIU Z, LIN Y T, CAO Y, *et al.* Swin transformer: hierarchical vision transformer using shifted windows[C]. 2021 *IEEE/CVF International Conference on Computer Vision (ICCV)*. October 10-17, 2021, Montreal, QC, Canada. IEEE, 2021: 9992-10002.
- [25] LAI W S, HUANG J B, AHUJA N, *et al.* Deep laplacian pyramid networks for fast and accurate super-resolution [C]. 2017 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. July 21-26, 2017, Honolulu, HI, USA. IEEE, 2017: 5835-5843.
- [26] EBEL P, XU Y J, 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: 5222414.
- [27] GROHNfeldt C, SCHMITT M, ZHU X 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*. July 22-27, 2018, Valencia, Spain. IEEE, 2018: 1726-1729.
- [28] ISOLA P, ZHU J Y, ZHOU T H, *et al.* Image-to-image translation with conditional adversarial networks[C]. 2017 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. July 21-26, 2017, Honolulu, HI, USA. IEEE, 2017: 5967-5976.
- [29] PAN H. Cloud removal for remote sensing imagery via spatial attention generative adversarial network [J]. *arXiv preprint arXiv: 2009.13015*, 2020.

#### 作者简介:



柳长源(1970—),男,黑龙江哈尔滨人,博士,副教授,硕士生导师,1993年于吉林大学获得学士学位,2005年于哈尔滨理工大学获得硕士学位,2013年于哈尔滨工程大学获得博士学位,2016-2017年普渡大学计算机与电子工程系访问学者,主要从事模式识别、人工智能与机器学习、数字图像处理领域研究。E-mail: liuchangyuan@hrbust.edu.cn



曹青(2000—),女,安徽阜阳人,硕士研究生,2023年于安徽信息工程学院获得学士学位,现就读于哈尔滨理工大学,主要从事遥感图像处理及多模态数据融合。E-mail: 2320610125@stu.hrbust.edu.cn