

# 基于注意力机制和生成对抗网络的遥感影像云去除

郑其光<sup>1</sup>, 王仁芳<sup>1,2</sup>, 邱虹<sup>2</sup>, 王峰<sup>2</sup>, 高广<sup>2,3</sup>, 吴敦<sup>2,3</sup>

1. 中国海洋大学 计算机科学与技术学院, 青岛 266100;

2. 浙江万里学院 大数据与软件工程学院, 宁波 315100;

3. 宝略科技(浙江)有限公司, 宁波 315042

**摘要:** 受成像环境、传感设备等因素的限制, 高分辨率卫星遥感影像存在云干扰的现象, 其严重影响了影像质量及后续的智能解译。深度学习因其强大的特征学习能力, 近年来在在遥感影像去云中得到了广泛应用。针对现有遥感影像去云方法往往难以更好地保留图像边缘纹理细节的问题, 本文提出一种基于注意力机制的生成对抗网络去云算法。首先设计傅里叶残差模块同时集成云层影像的低频和高频残差信息; 然后利用注意力机制从通道和空间两个角度进行特征关注而获得云层空间分布信息, 并引导云层去除; 最后在构造的联合损失函数约束下训练模型, 确保生成图像和真实影像在内容上的一致性。在遥感影像数据集 RICE 上的实验结果表明, 本文提出的基于注意力机制和生成对抗网络的遥感影像云去除方法能够有效地保留影像边缘纹理细节, 且达到了优异的定量指标和视觉效果。

**关键词:** 遥感影像去云, 傅里叶残差, 通道-空间注意力, 生成对抗网络, 云层空间分布

中图分类号: TP2

引用格式: 郑其光, 王仁芳, 邱虹, 王峰, 高广, 吴敦. XXXX. 基于注意力机制和生成对抗网络的遥感影像云去除. 遥感学报, XX(XX): 1-10

Zheng Qiguang, Wang Renfang, Qiu Hong, Wang Feng, GAO Guang, Wu Dun. XXXX. A remote sensing image cloud removal network based on attention mechanism and generate adversarial network. National Remote Sensing Bulletin, DOI:10.11834/jrs.20243372]

## 1 引言

高分辨率卫星遥感影像在土地使用 (Wan 等, 2021)、环境污染 (Hadjimitsis 等, 2008)、国防安全 (Li 等, 2021)、城市规划 (Judith 等, 2013) 等领域得到了广泛应用。受成像环境、传感设备等因素的限制, 通常卫星等传感设备获取的遥感影像会受到云层的干扰, 致使影像质量降低、细节丢失和信息熵退化等 (Wei 等, 2018), 严重地影响了后续遥感图像分类 (Fan 等, 2022)、变化检测 (Ming 等, 2023) 和超分辨率重建 (Yang 等, 2022) 等解译结果的精度。因此, 有效的去云技术在遥感影像解译领域得到了广泛关注。

传统的去云方法主要包括基于频率域变换的

去云方法和基于先验知识的去云方法。基于频率域变换的去云方法认为遥感影像上的云是低频噪声, 通过在频率域中抑制低频信息, 增强高频信息, 从而降低云层的影响, 增强地面信息的表达。如小波变换 (Wang 等, 2013)、同态滤波 (Shen 等, 2014) 等。基于先验知识的去云方法通过对大量有薄云的遥感图像和无云影像进行统计, 得出二者之间的关系, 并利用这种转换关系来去除薄云得到清晰的遥感影像。如直方图匹配法 (Chen 等, 2018)、暗通道先验理论 (Liu 等, 2016) 等。上述去云方法通常需要人为设定先验条件, 非常容易受到环境因素的影响, 效果并不理想。

近年来, 深度学习因其强大的特征学习能力, 能够学习复杂场景中的复杂关系, 因而在遥感影

收稿日期: XXXX-XX-XX; 预印本: XXXX-XX-XX

基金项目: 国家自然科学基金(编号: 61906170); 浙江省科技计划项目(编号: LGF21F020023, Y24F030030); 宁波市科技计划项目(编号: 2022Z233, 2021Z050, 2022S002, 2023J403)

第一作者简介: 郑其光, 研究方向为遥感智能解译。E-mail: zhengqiguang321@163.com

通信作者简介: 王仁芳, 研究方向为数字图像处理、机器学习、遥感智能解译。E-mail: renfang\_wangac@126.com

像去云中得到了广泛使用。现有的深度学习方法根据处理的对象可以分为基于多幅的遥感影像处理方法和基于单幅的遥感影像处理方法。基于多幅的遥感影像处理方法通常使用多源、多时相数据,用同一地区的无云影像替代云覆盖区域的影像。Mirza 等人 (Mirza 等, 2014) 提出了一种多光谱条件模型 McGANs, 该模型使用合成的云卫星图像和近红外波段图像对网络进行训练, 使其能够从云图中自动生成无云图像。Ren 等人 (Ren 等, 2018) 通过应用白平衡、对比度增强和 Gamma 校正从原始模糊图像中获得三个输入, 并根据这些不同输入之间的外观差异计算像素级置信度图, 通过门控派生输入的重要特征得到最终的去雾图像。Li 等人 (Li 等, 2023) 基于光学和 SAR 遥感影像的层次融合, 提出了多级光谱和结构保持融合网络 HS<sup>2</sup>P, 该网络提出了具有堆叠残差群的深度分层架构, 同时使用信道注意机制来自适应选择具有信息性的特征进行融合。这类方法在去云任务中表现不错, 但是需要解决影像配准、几何和辐射校正问题, 而且需要保证影像相对的时间差小 (即保证两幅影像中该地区地物的变化小), 同时云区也不可重叠。

基于单幅的遥感影像处理方法使用单源、单时相数据, 利用图像补绘或图像生成技术来恢复云层影像中的缺失信息, 可以有效地将云层影像转换为无云影像。Li 等人提出了去雾深度神经网络 AOD-Net (Li 等, 2017), 该模型避免了估计传输矩阵和大气光参数等步骤, 通过轻量级 CNN 直接生成干净的图像, 这种端到端设计可以很容易地将该网络嵌入到其他深度模型。Liu 等人基于注意力机制提出了多尺度网络模型 GridDehazeNet (Liu 等, 2019), 该网络不依赖于大气散射模型, 可以高效地交换不同尺度的信息, 从而可缓解多尺度估计的瓶颈问题。Dong 等人将 U-Net 和 GAN 相结合, 提出了 two-stage (Dong 等, 2020) 模型, 该模型可有效去除厚云并修复影像。Pan 将空间注意力模块与 GAN 网络结合提出了 SpA GAN (Pan, 2020), 利用空间注意力机制生成云层的分布图, 并指导后续模块去云。Zhou 等人提出了遥感影像去云的多尺度残差神经网络模型 MSAR-DefogNet (Zhou 等, 2022), 使用具有更大卷积核的多尺度卷积模块扩大感受野实现节省计算资源的目的。Ding 等人 (Ding 等, 2023) 提出了基于

条件变分自编码器的薄云去除网络, 为每个输入云层影像生成多个合理的无云影像, 通过不确定性分析, 在这些多重预测的基础上生成清晰影像。Zi 等人 (Zi 等, 2023) 基于编码器-解码器架构提出了小波集成卷积神经网络模型 WaveCNN-CR, 通过小波变换提取多尺度和多频率分量, 设计增强特征提取模块对不同尺度下的高频分量提取特征, 并从全局和局部角度对特征进行增强。上述利用残差块的单幅遥感影像处理方法能够取得有效的去云效果, 然而尽管这种使用传统的残差块来捕获特征的方法能够很好地捕获高频特征, 但往往忽视了云层影像和真实影像对之间的低频信息在图像整体外观上的恢复; 同时对通道、空间特征关注较少, 使得图像的特征未得到充分利用, 从而难以有效的保留纹理细节。

针对上述基于单幅的遥感影像去云方法存在的问题, 本文提出一种基于注意力机制的生成对抗网络去云算法。该算法基于残差学习设计了傅里叶残差模块, 使模型在浅层便能获取全局上下文信息的同时, 集成去图低频和高频残差信息; 为了获得充分的云层分布信息, 利用注意力机制设计通道-空间注意力模块; 为了使生成图像和真实影像在内容上保持一致性, 联合 condition GAN 损失、L1 损失、注意力损失和感知损失构造损失函数, 从而使模型能够充分利用图像的特征信息, 确保生成的去云图像能够很好的保留纹理细节。

## 2 研究方法

在本章节, 介绍本文提出的基于注意力机制和生成对抗网络的遥感影像云去除方法的整体网络模型结构、设计的傅里叶残差模块和通道-空间注意力模块和构造的联合损失函数。

### 2.1 网络框架

基于注意力机制的生成对抗网络去云模型 (AtM-GAN) 包括 2 部分: 用于生成重建图像的生成器; 用于判别生成器生成图像质量的判别器。

生成器的结构如图 1 (a) 所示, 在生成器中, 设计的傅里叶残差模块 (FRB) 用于建模云层影像和真实影像对之间的高频和低频差异, 以及云层去除后重构干净背景; 云层去除模块 (CRB) 的结构如图 1 (c) 所示, CRB 由通道-空间注意力模块 (CSAM) 和三个深度傅里叶变换残差模块

(DFTRB) 并联组成。CSAM从通道和空间的角度生成云层分布图；DFTRB在云层分布图的引导下去云，DFTRB的结构如图1 (d)，其中FTB和CB分别为FRB中的FTB和CB。

判别器的结构如图2所示，由卷积层、非线性激活层 (ReLU) 和归一化层构成，输入是生成影像和真实影像，输出的结果表示判别器认为生成影像是真实影像的概率。

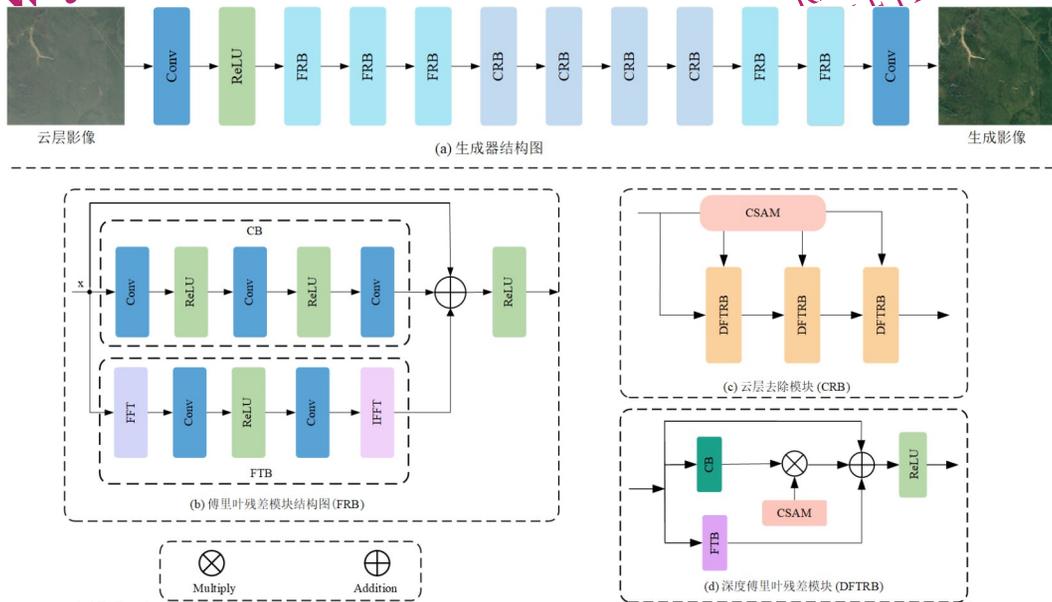


图1 (a)是生成器结构图，用三个FRB提取特征，四个CRB去云，最后用两个FRB重构干净背景；(b)是傅里叶残差块 (FRB)，FFT表示傅里叶变换，IFFT表示傅里叶反变换；(c)是云层去除块 (CRB)；(d)是深度傅里叶变换残差块 (DFTRB)

Fig. 1 (a) is the generator includes three FRBs for extracting features, four CRBs for removing clouds, and two FRBs for reconstructing the clean background; (b) is the Fourier residual block (FRB), FFT denotes Fourier transform, and IFFT represents Inverse Fourier transform; (c) is the cloud removal block (CRB); (d) is the deep Fourier transform residual block (DFTRB)

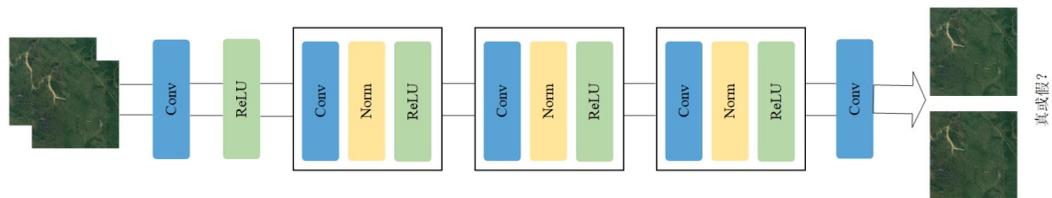


图2 判别器结构图

Fig.2 The structure of the discriminator

## 2.2 傅里叶残差模块

残差网络 (He 等, 2015) 常被用于构建遥感影像去云神经网络以学习云层影像和无云影像之间的差异。残差学习策略可以较好地捕获图像中的高频特征，Mao 等人通过在残差网络中引入傅里叶变换、ReLU 操作和傅里叶反变换提出了 Res FFT-ReLU Block，发现将 Res FFT-ReLU Block 加入到去模糊模型后性能有了较大的提升，其认为在频率域中应用 ReLU 和卷积操作后，再进行傅里叶反变换可以使神经网络灵活的选择频率阈值，同时在频率域中引入非线性操作能为网络带来全

局上下文学习能力 (Mao 等, 2022)。受此启发，为使模型可以同时提取关于遥感影像的高频特征和低频特征，本文构建了傅里叶残差模块 (FRB) 如图 1 (b) 所示。令  $x$  表示 FRB 的输入特征，FRB 可表示为：

$$FRB = \text{ReLU}(CB + FTB + x) \quad (1)$$

$$CB = \text{Conv}_{11}(\text{ReLU}(\text{Conv}_{33}(\text{ReLU}(\text{Conv}_{11}(x)))))) \quad (2)$$

$$FTB = \mathcal{F}^{-1}(\text{Conv}_{11}(\text{ReLU}(\text{Conv}_{11}(\mathcal{F}(x)))))) \quad (3)$$

其中，Conv11 和 Conv33 分别表示卷积核大小为  $1 \times 1$  和  $3 \times 3$  的卷积层， $\mathcal{F}$  表示傅里叶变换， $\mathcal{F}^{-1}$  表示傅里叶反变换。CB 为三个卷积层和两个非线性激活函数 (ReLU) 所组成的卷积块，和 FRB 中第一

条分支构成了残差结构，作用是学习影像中的高频特征；FTB为快速傅立叶变换。卷积层、ReLU激活层和傅立叶反变换所组成的傅里叶变换模块，其作用为学习影像中的低频特征。从图1(b)中可以看到，PRB的输入在FTB中将首先被变换到频域，经过卷积和ReLU层处理后的频率信息通过傅立叶反变换映射到空间域。

### 2.3 通道-空间注意力模块

注意力机制符合人类观察物体的偏好，重点观察突出的特征区域，其中通道注意力机制主要关注特征图中重要的通道。Wang等人(Wang等, 2020)认为避免降维、有适当的通道交互是十分重要的，因此在SENet(Hu等, 2019)的基础上提出了一种不降维的局部跨通道交互模型ECA-Net。为了扩大网络的感受野同时获得更多的通道维度上的特征信息，本文在ECA-Net的基础上加入全局最大池化(Global Max Pooling, GMP)构建通道注意力模块(CAM)，如图3所示。其输入分别经过GMP和全局平均池化(Global Average Pooling, GAP)再相加，并通过卷积层和激活层确定通道的权重，最后将原始输入特征与权重相乘

再相加，得到和输入相同尺寸的输出特征图。CAM的计算过程可以表示为：

$$F_{CAM} = x \otimes \text{sigmoid}(\text{Conv}(\text{GAP}(x) + \text{GMP}(x))) + x \quad (4)$$

其中， $x \in \mathbb{R}^{32 \times H \times W}$ 表示CAM的输入特征， $F_{CAM} \in \mathbb{R}^{32 \times H \times W}$ 表示CAM的输出特征，H和W是图像的高度和宽度。

空间注意力机制主要关注关键区域的位置，本文基于递归神经网络(Hu等, 2020)构建了空间注意力模块(SAM)，如图3所示。其为一个两轮次、四方向(上、下、左、右)具有ReLU激活和单位矩阵初始化的递归神经网络，能够同时捕获局部及全局上下文语义信息。在第一轮中，SAM的目的是总结输入图像每个位置的相邻上下文信息并生成特征图；在第二轮中，SAM进一步收集非局部上下文信息，以生成全局感知特征图。SAM所生成的特征图即为用于引导后续去云的云层空间分布信息，在RICE数据集(Lin等, 2019)上生成的可视化结果如图8所示，颜色越红表示分配越多的注意力，颜色偏蓝表示分配较少的注意力。

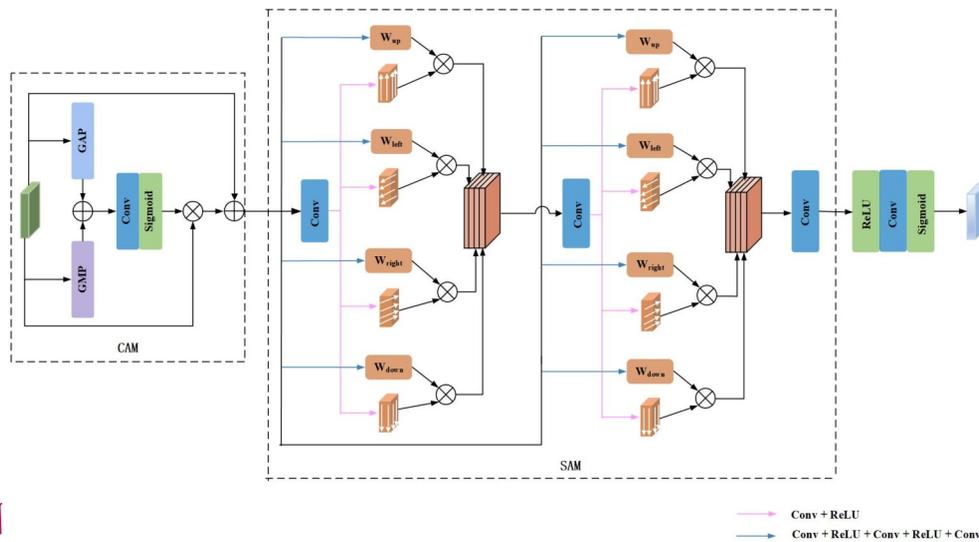


图3 CSAM的结构图

Fig.3 The structure of CSAM

SAM通过两轮次、四方向上的循环卷积分累全局上下文语义信息，如图4所示。在第一轮次中，通过对输入特征图上的每个位置分别在四方向(上、左、下、右)进行循环卷积运算，获得水平和垂直邻域信息；在第二轮次中，通过重复

第一轮的操作，将获得整个特征图的语义信息。令 $h_{i,j}^{right}$ 表示图像在像素 $(i,j)$ 处向右卷积后的特征，那么向右方向上的循环卷积结果可表示为：

$$h_{i,j}^{right} = \max(\alpha_{right} h_{i,j-1}^{right} + h_{i,j}^{right}, 0) \quad (5)$$

其中， $\alpha_{right}$ 是右方向的权重参数，其他三个方向

类似。

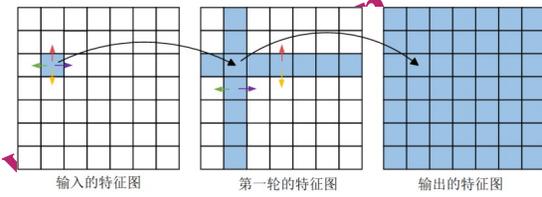


图4 SAM 累计全局上下文语义信息的示例图

Fig.4 Illustration of how SAM accumulates global contextual information

## 2.4 损失函数

为了使生成图像和真实影像保持内容上的一致性，同时加快模型收敛的速度，本文提出的AtM-GAN使用的联合损失函数如公式(6)所示：

$$L_{total} = L_{CGAN} + L_1 + L_{Attention} + L_{per} \quad (6)$$

$L_{CGAN}$  是 condition GAN 损失，定义如公式(7)所示：

$$L_{CGAN}(G,D) = \mathbb{E}_{x,y \sim P_{data}(x,y)} [\log D(x,y)] + \mathbb{E}_{x,y \sim P_{data}(x,y)} [\log(1 - D(x,G(x,z)))] \quad (7)$$

其中， $x$  表示云层影像， $y$  表示生成影像， $y$  表示真实影像， $\mathbb{E}$  表示期望， $P$  表示数据分布， $D$  表示判别器， $G$  表示生成器。 $L_1$  损失用来衡量每一个点的重建效果， $L_1$  的定义如公式(8)所示：

$$L_1 = \frac{\sum_{i=1}^n |y_i - z_i|}{n} \quad (8)$$

式(8)中， $n$  为像素点个数。

注意力损失的定义如公式(9)所示，矩阵  $F$  是通道-空间注意力模块 CSAM 生成的云层分布注意力图， $M$  是通过计算云层影像和真实影像的差值得到云层的二值分布图。

$$L_{Attention} = \|F - M\|_2^2 \quad (9)$$

感知损失定义如公式(10)，它计算真实影像和生成影像之间的相似性。本文使用在 ImageNet 数据集上预训练好的 VGG19 模型提取图像特征， $C_j H_j W_j$  是第  $j$  层对应的特征图大小。

$$L_{per} = \frac{1}{C_j H_j W_j} \|VGG19_j(z) - VGG19_j(y)\|_2^2 \quad (10)$$

## 3 实验结果与分析

为了验证所提出算法的有效性，本文将该算法与主流的去云算法 HALP (He 等, 2023)、Cloud GAN (Singh 等, 2018)、Pix2Pix (Wu 等, 2022)、SpA GAN 在 RICE 数据集上进行了实验

比较。

### 3.1 数据集

RICE 遥感影像数据集包含 RICE1 和 RICE2 两个部分，所有的图像都为  $512 \times 512$  像素。RICE1 是一个在 Google 地球上收集的薄云数据集，总共包含 500 对图像。通过设定一定的时间间隔，RICE2 收集了 Landsat 8 OLI/TIRS 卫星的 736 对数据。将训练集和测试集按照 4:1 的方法划分，对于 RICE1 数据集，分配 400 张用于训练，100 张用于测试；对于 RICE2 数据集，分配 588 张用于训练，148 张用于测试。RICE 数据集的样例如图 5 所示。



图5 RICE 数据样本, 第一行是云层图像, 第二行是对应的真实图像。前两列是 RICE1 数据集的样本, 后两列是 RICE2 数据集的样本

Fig.5 The data samples of RICE, the first row is cloudy image, the second row is real image. The first two columns are samples belong to RICE1, the last two columns are samples belong to RICE2.

### 3.2 训练环境与参数配置

本文的所有实验均在硬件设备为 i7-12700 @ 2.10 GHz CPU, NVIDIA GeForce RTX3060 GPU, 软件环境为 Pytorch 1.12 深度学习框架, Windows 11 64 位的操作系统上运行。使用 Adam 优化器, 训练初始学习率为 0.0004, 采用余弦退火策略对学习率进行周期调整, epoch 为 200 进行批次训练。

### 3.3 评价指标

为了便于同其他算法进行对比, 我们使用峰值信噪比 (PSNR) 和结构相似度指标 (SSIM) 作为评价指标。其中 PSNR 的计算方式如公式 11 和公式 12 所示, 其中  $H$  和  $W$  分别表示图像的高度和宽度,  $Z$  是生成影像,  $Y$  是无云影像, MSE 表示均方误差, MAX 指的是像素值的最大可能范围。

$$PSNR = 10 \log_{10} \left( \frac{MAX^2}{MSE} \right) \quad (11)$$

$$\text{MSE} = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W (Z(i,j) - Y(i,j))^2 \quad (12)$$

SSIM 在亮度、对比度和结构三个方面衡量两张图像的相似性，SSIM 的计算方法如公式 13 所示。

$$\text{SSIM}(Z,Y) = \frac{(2\mu_Z\mu_Y + c_1)(2\sigma_{ZY} + c_2)}{(\mu_Z^2 + \mu_Y^2 + c_1)(\sigma_Z^2 + \sigma_Y^2 + c_2)} \quad (13)$$

其中， $c_1$  和  $c_2$  是常量， $\mu_Z$  和  $\mu_Y$  表示图像的平均亮度，以  $\mu_Z$  为例

$$\mu_Z = \frac{1}{N} \sum_{i=1}^N z_i \quad (14)$$

$\sigma_Z$  和  $\sigma_Y$  表示标准差，以  $\sigma_Z$  为例

$$\sigma_Z = \left( \frac{1}{N-1} \sum_{i=1}^N (z_i - \mu_Z)^2 \right)^{1/2} \quad (15)$$

$\sigma_{ZY}$  表示  $Z$  和  $Y$  的方差，计算方法如公式 16 所示：

$$\sigma_{ZY} = \frac{1}{N-1} \sum_{i=1}^N (z_i - \mu_Z)(y_i - \mu_Y) \quad (16)$$

### 3.4 对比实验

五种算法在 RICE1 上的部分可视化结果如图 6 所示。从左到右，每一列分别表示云层影像、HALP 去云结果、Cloud GAN 去云结果、Pix2Pix 去云结果、SpA GAN 去云结果、AtM GAN 去云结果和对应的真实影像。由于 RICE1 中大部分是薄云影像，地物信息仍然保留了比较清晰的纹理细节，故各个算法生成的结果从视觉上看和对应的真实影像都有一定的相似之处。但是 HALP 和 Cloud GAN 的去云图像整体出现了较大的颜色失真，整体与真实的无云影像存在较大的差异；Pix2Pix 和 SpA GAN 在部分区域也存在一定的颜色差异，本文提出算法的结果与真实影像最为接近。

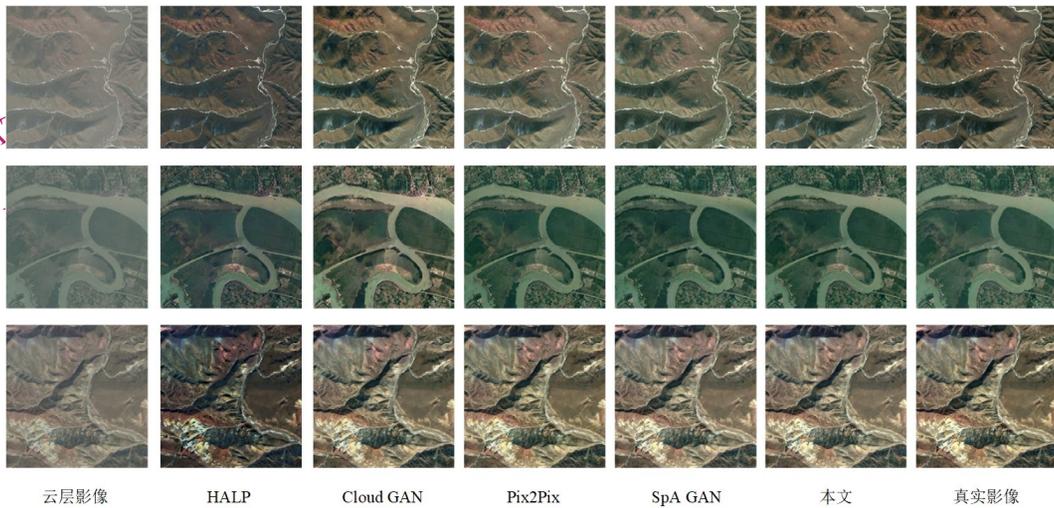


图6 RICE1上的生成结果对比图

Fig.6 Comparison of cloud removal results by different methods on the RICE1 dataset

五种算法在 RICE1 数据集上的 PSNR 和 SSIM 指标如表 1 所示。从表中可以看出本文提出的 AtM-GAN 在 PSNR 和 SSIM 的指标分别是 32.4921 和 0.9746，在对比的几个方法中都达到了最好的效果，得益于建模了云层影像和真实影像对之间的高频和低频差异，云层区域的准确定位和联合损失函数最小化生成影像和真实影像之间的像素差异和结构差异。

五种方法在 RICE2 数据集上的部分可视化结果如图 7 所示。从左到右，每一列分别表示云层影像、HALP 去云结果、Cloud GAN 去云结果、Pix2Pix 去云结果、SpA GAN 去云结果、AtM GAN

去云结果和相对应的真实影像。由于 RICE2 中的图像大多数为厚云，相较于薄云，厚云中被遮挡像素点的数据信息是永久性缺失的，这种伤害也是无法逆转的。从图 7 可以看出，HALP 方法仍然存在大量未去除的云层区域；Cloud GAN 和 Pix2Pix 方法去云后存在较多的云层轮廓；SpA GAN 方法虽然对云层区域定位较为准确，但是也存在局部区域细节丢失的问题；本文方法去云图像无明显的颜色失真，且细节纹理体现较好，与真实的标签影像较为接近。

五种算法在 RICE2 上的 PSNR 和 SSIM 指标如表 2 所示，从表中可以看出本文提出的方法在

PSNR 和 SSIM 的指标分别是 31.3331 和 0.9359，这两个指标在几个对比算法中都达到了最好的效果，两个指标相对于 SpA GAN 分别提升了 3.193 和 0.0322。这表明了注意力机制和傅里叶残差结构在云层去除任务中的有效性。

表 1 在 RICE1 数据集上的云去除性能对比

Table 1 Quantitative comparison of different method for cloud removal on the RICE1 dataset

Model	PSNR	SSIM
HALP	22.8498	0.9033
Cloud GAN	24.2811	0.8349
Pix2Pix	25.5491	0.8793
SpA GAN	29.0656	0.9569
AtM GAN	32.4921	0.9746

### 3.5 消融实验

为了评估傅里叶残差模块和 CSAM 对整个网络去云性能的影响，本文按照网络结构及相关模块设计了消融实验，共包含 5 个实验：(1) WFCS (without FTB, CAM and SAM) 表示 FRB 中没有采用 FTB 分支，同时没用使用 CSAM 中的 CAM 和 SAM。(2) FS (FTB and SAM) 表示 FRB 中引入了 FTB，CSAM 中引入了 SAM，但是 CSAM 中没有引用 CAM。(3) CS (CAM and SAM) 表示 CSAM 中引入了 CAM 和 SAM，但是 FRB 中没有引入 FTB。(4) FC (FTB and CAM) 表示 FRB 中引入了 FTB，CSAM 中引入了 CAM，但是没用引入 SAM。(5) FCS (FTB, CAM and SAM) 表示 FRB 中引入了 FTB，同时 CSAM 中使用了 CAM 和 SAM。

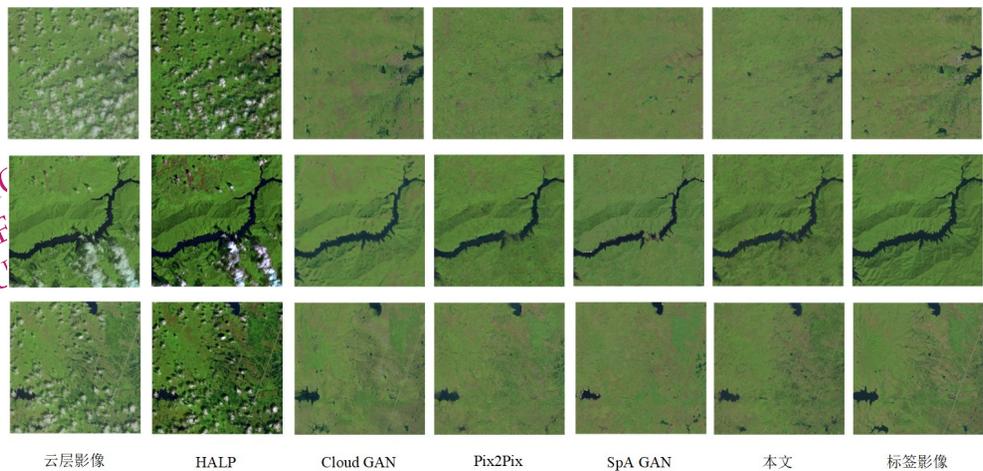


图 7 RICE2 上的生成结果对比图

Fig.7 Comparison of cloud removal results by different methods on the RICE2 dataset

表 2 在 RICE1 数据集上的云去除性能对比

Table 2 Quantitative comparison of different method for cloud removal on the RICE2 dataset

Model	PSNR	SSIM
HALP	20.8498	0.7397
Cloud GAN	24.0205	0.7336
Pix2Pix	23.4597	0.7006
SpA GAN	28.1401	0.9037
AtM GAN	31.3331	0.9359

在 RICE1 和 RICE2 数据集上进行的消融实验结果分别如表 3 和表 4 所示。从表 3 中可以看到 FS 的 SSIM 值高于 FCS，主要由于 RICE1 中的图像大部分区域覆盖了薄云，导致特征在 CAM 中经过 GAP 和 GMP 后值偏大，使得生成图像的像素整体

偏高，从而导致 FCS 中的 SSIM 值偏低。从表中可以看出，FRB、CAM 和 SAM 三者的使用相对于 WFCS 有了明显的提升，说明了傅里叶残差模块和注意力机制在云层去除任务中的有效性。

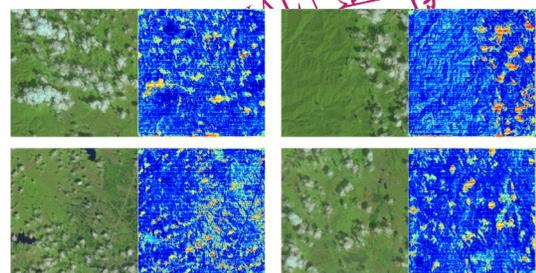


图 8 RICE 数据集生成的云层空间分布信息图

Fig.8 Spatial distribution of cloud generated on the RICE dataset

表3 在 RICE1 数据集上的消融实验

Table 3 Ablation experiments on the RICE1 dataset

Model	PSNR	SSIM
WFCS	27.8127	0.9149
FS	31.9559	0.9748
CS	30.8383	0.9670
FC	30.7825	0.9508
FCS	32.4921	0.9670

表4 在 RICE2 数据集上的消融实验

Table 4 Ablation experiments on the RICE2 dataset

Model	PSNR	SSIM
WFCS	27.4215	0.8196
FS	30.5933	0.9349
CS	28.8533	0.9220
FC	30.9827	0.8638
FCS	31.3331	0.9359

## 4 结论

本文针对现有的云层去除模型无法更好的保留图像边缘纹理细节问题,在生成对抗网络框架下,利用傅里叶残差模块建模云层影像的低频和高频残差信息,使用通道-空间注意力机制来获取云层分布图,同时构建联合损失函数约束生成影像和真实影像的内容一致性。实验结果表明,在 RICE1 和 RICE2 数据集上,本文提出的 AtM-GAN 网络均能实现很好的云层去除效果,在 RICE1 上 PSNR 和 SSIM 指标分别是 32.4921 和 0.9746;在 RICE2 上 PSNR 和 SSIM 指标分别是 31.3331 和 0.9359,这也表明了注意力机制和傅里叶残差块在去除任务中的有效性。

与现有主流去云模型相比,本文提出的算法能够有效保留图像边缘纹理细节,在遥感影像云去除任务中取得良好的效果。由于本文设计的模型针对的是单幅遥感影像云去除,尚不能直接应用到多幅遥感影像云去除任务。因此下一步的主要工作是更深入地探索在多时相和多源数据集上云去除模型的设计及其应用。

## 参考文献 (References)

Chen S L, Chen X H, Chen X, Chen J, Cao X, Shen M G, Yang W and Cui X H. 2018. A Novel Cloud Removal Method Based on IHOT

and the Cloud Trajectories for Landsat Imagery. Remote Sensing, 10(7): 1040 [DOI: 10.3390/rs10071040]

Ding H D, Zi Y and Xie F Y. 2023. Uncertainty-Based Thin Cloud Removal Network via Conditional Variational Autoencoders. Asian Conference on Computer Vision. Berlin: Springer: 52-68 [DOI: 10.1007/978-3-031-26318-2\_41]

Dong H, Pan J S, Xiang L, Hu Z, Zhang X Y, Wang F and Yang M H. 2020. Multi-Scale Boosted Dehazing Network With Dense Feature Fusion. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle: IEEE: 2154-2164 [DOI: 10.1109/CVPR42600.2020.00223]

Fan X N, Yan W, Shi P F and Zhang X W. 2022. Remote sensing image target detection based on a multi-scale deep feature fusion network. National Remote Sensing Bulletin, 26(11): 2292-2303 (范新南, 严炜, 史鹏飞, 张学武. 2022. 多尺度深度特征融合网络的遥感图像目标检测. 遥感学报, 26(11): 2292-2303 [DOI: 10.11834/jrs.20210170])

Hadjimitsis D G and Clayton C R I. 2008. The use of an improved atmospheric correction algorithm for removing atmospheric effects from remotely sensed images using an atmosphere-surface simulation and meteorological data. Meteorological Applications, 15(3): 381-387 [DOI: 10.1002/met.80]

He K M, Zhang X Y, Ren S Q and Sun J. 2015. Deep Residual Learning for Image Recognition. IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas: IEEE: 770-778 [DOI: 10.1109/CVPR.2016.90]

He Y F, Li C L and Li X. 2023. Remote Sensing Image Dehazing Using Heterogeneous Atmospheric Light Prior. IEEE Access, (11): 18805-18820. [DOI: 10.1109/ACCESS.2023.3247967]

Hu J, Shen L and Sun G. 2018. Squeeze-and-Excitation Networks. IEEE Transactions on Pattern Analysis Machine Intelligence 42(8), 2011-2023 [DOI: 10.1109/CVPR.2018.00745]

Hu X W, Fu C W, Zhu L, Qin J and Heng P A. 2020. Direction-aware Spatial Context Features for Shadow Detection and Removal. IEEE Transactions on Pattern Analysis and Machine Intelligence, 42(11): 2795-2808 [DOI: 10.1109/TPAMI.2019.2919616]

Judith C, Schneider J V, Schmidt M, Ortega R, Gaviria J and Zizka G. 2013. Using high-resolution remote sensing data for habitat suitability models of Bromeliaceae in the city of Mérida, Venezuela. Landscape and Urban Planning, 120: 107-118 [DOI: 10.1016/j.landurbplan.2013.08.012]

Li B Y, Peng X L, Wang Z Y, Xu J Z and Feng D. 2017. AOD-Net: All-in-One Dehazing Network. International Conference on Computer Vision (ICCV). Venice: IEEE: 4780-4788 [DOI: 10.1109/ICCV.2017.511]

Li S T, Li C Y and Kang X D. 2021. Development status and future prospects of multi-source remote sensing image fusion. National Remote Sensing Bulletin, 25(1): 148-166 (李树涛, 李聪好, 康旭东. 2021. 多源遥感图像融合发展现状与未来展望. 遥感学报, 25(1): 148-166 [DOI: 10.11834/jrs.20210259])

Li Y S, Wei F Y, Zhang Y J, Chen W, Ma J Y. 2023. HS2P: Hierarchical spectral and structure-preserving fusion network for multimod-

- al remote sensing image cloud and shadow removal. *Information Fusion*, 94: 215-228 [DOI: 10.1016/j.inffus.2023.02.002]
- Liu D Y, Xu G L, Wang X K, Wang Y, Xian S and Fu K. 2019. A Remote Sensing Image Dataset for Cloud Removal. arXiv: 1901.00600
- Liu X H, Ma Y R, Shi Z H and Chen J. 2019. GridDehazeNet: Attention-Based Multi-Scale Network for Image Dehazing. *International Conference on Computer Vision (ICCV)*. Seoul: IEEE: 7313-7322 [DOI: 10.1109/ICCV.2019.00741]
- Liu X, Zhang H, Tang Y Y and Du J X. 2016. Scene-adaptive single image dehazing via opening dark channel model. *IET Image Processing*, 10(11): 877-884 [DOI: 10.1049/iet-ipr.2016.0138]
- Mao X T, Liu Y M, Liu F Z, Li Q L, Shen W and Wang Y. 2022. Intriguing Findings of Frequency Selection for Image Deblurring. *AAAI Conference on Artificial Intelligence*, 37: 1905-1913 [DOI: 10.1609/aaai.v37i2.25281]
- Ming Y S, Liu Q H, Bai H and Huang C. 2023. Classification and change detection of vegetation in the Ruoergai Wetland using optical and SAR remote sensing data. *National Remote Sensing Bulletin*, 27(6): 1414-1425 (明义森, 刘启航, 柏荷, 黄昌. 2023. 利用光学和SAR遥感数据的若尔盖湿地植被分类与变化监测. *遥感学报*, 27(6): 1414-1425 [DOI: 10.11834/jrs.20221767])
- Mirza M and Osindero S. 2014. Conditional Generative Adversarial Nets. [DOI: 10.48550/arXiv.1411.1784]
- Pan H. 2020. Cloud Removal for Remote Sensing Imagery via Spatial Attention Generative Adversarial Network. [DOI: arXiv: 2009.13015]
- Ren W Q, Ma L, Zhang J W, Pan J S, Cao X C, Liu W and Yang M H. 2018. Gated Fusion Network for Single Image Dehazing. *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Seattle: IEEE: 3253-3261 [DOI: 10.1109/CVPR.2018.00343]
- Shen H F, Li H F, Qian Y, Zhang L P and Yuan Q Q. 2014. An effective thin cloud removal procedure for visible remote sensing images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 96: 224-235 [DOI: 10.1016/j.isprsjprs.2014.06.011]
- Singh P and Komodakis N. 2018. Cloud-Gan: Cloud Removal for Sentinel-2 Imagery Using a Cyclic Consistent Generative Adversarial Networks. *International Geoscience and Remote Sensing Symposium*. Valencia: IEEE: 1772-1775 [DOI: 10.1109/IGARSS.2018.8519033]
- Wan Y Q, Fan Y H and Jin M S. 2021. Application of hyperspectral remote sensing for supplementary investigation of polymetallic deposits in Huaniushan ore region, northwestern China. *Scientific Reports*, 11(1): 440 [DOI: 10.1038/s41598-020-79864-0]
- Wang X X, Jiang L S, Chen Y P and Wang J L. 2013. Thin cloud removal from remote sensing images with adaptive thresholds of wavelet transforms. *Journal of University of Electronic Science and Technology of China*, 42(03): 71-74 [DOI: 10.3969/j.issn.1001-0548.2013.03.014]
- Wang Q L, Wu B G, Zhu P F, Li P H, Zuo W M and Hu Q H. 2020. ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks. *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Seattle: IEEE: 11531-11539 [DOI: 10.1109/CVPR42600.2020.01155]
- Wei W, Polap D, Li X H, Wozniak M and Liu J Z. 2018. Study on Remote Sensing Image Vegetation Classification Method Based on Decision Tree Classifier. *Symposium Series on Computational Intelligence (SSCI)*. Piscataway: IEEE: 2292 - 2297 [DOI: 10.1109/SSCI.2018.8628721]
- Wu X, Song W, Zhang X K, Lin G H, Wang H M and Deng Y Y. 2022. Algorithm Development of Cloud Removal from Solar Images Based on Pix2Pix Network. *Computers, Materials & Continua*, 71(2): 3497-3512 [DOI: 10.32604/cmc.2022.022325]
- Yang X, Li F, Lu M, Xu L, Lu X T and Zhang N. 2022. New super-resolution reconstruction method based on Mixed Sparse Representations. *National Remote Sensing Bulletin*, 26(8): 1685-1697 (杨雪, 李峰, 鹿明, 辛蕾, 鲁啸天, 张南. 2022. 混合稀疏表示模型的超分辨率重建. *遥感学报*, 26(8): 1685-1697 [DOI: 10.11834/jrs.20219409])
- Zhou Y, Jing W P, Wang J, Chen G S, Scherer R and Damaševičius R. 2022. MSAR-DefogNet: Lightweight cloud removal network for high resolution remote sensing images based on multi scale convolution. *IET Image Processing*, 16(3): 659-668 [DOI: 10.1049/ipr2.12224]
- Zi Y, Ding H D, Xie F Y, Jiang Z G and Song X D. 2023. Wavelet Integrated Convolutional Neural Network for Thin Cloud Removal in Remote Sensing Images. *Remote Sensing*, 15(3): 781 [DOI: 10.3390/rs15030781]

## A remote sensing image cloud removal network based on attention

## mechanism and generate adversarial network

Zheng Qigang<sup>1</sup>, Wang Renfang<sup>1,2</sup>, Qiu Hong<sup>2</sup>, Wang Feng<sup>2</sup>, GAO Guang<sup>2,3</sup>, Wu Dun<sup>3</sup>

<sup>1</sup>. College of Computer Science and Technology, Ocean University of China, Qingdao 266100, China;

<sup>2</sup>. College of Big Data and Software Engineering, Zhejiang Wanli University, Ningbo 315100, China;

<sup>3</sup>. Popsmart Technology(Zhejiang) Co., Ltd., Ningbo, 315042, China

**Abstract:** High-resolution satellite remote sensing images are often compromised by cloud interference due to constraints such as imaging conditions and sensing equipment. This significantly degrades image quality and hinders intelligent interpretation. In recent years, deep learning has been extensively applied to cloud removal in remote sensing images, thanks to its powerful feature learning capabilities. This paper presents an attention-based generative adversarial network for cloud removal in remote sensing images, which aims to address the challenge of preserving textures. Initially, a Fourier residual block is designed to integrate both low-frequency and high-frequency of cloud images. Subsequently, attention mechanisms are employed from both channel and spatial to capture the spatial distribution information of clouds that guides the network in effectively removing clouds. Finally, the model is trained to minimize the joint loss function, ensuring the coherence of generated images with real images in terms of content. Experimental results derived from the RICE dataset substantiate that the proposed cloud removal method, which integrates attention mechanisms and generative adversarial networks, is capable of effectively eradicating clouds while preserving textures. The following conclusions can be drawn (1) The proposed Fourier residual block plays a pivotal role in adeptly capturing the details of both low-frequency and high-frequency residual information present in cloud images. By merging these diverse features, the block bolsters the network's ability to extract rich features; (2) The attention mechanisms utilized extract features from both spatial and channel dimensions. This enables the model to discern the significance of various features across different dimensions. Moreover, the cloud information obtained through these attention mechanisms allows the model to adaptively present the changes in cloud, thereby enhancing the model's precision and efficiency in cloud removal; (3) Experimental results on the RICE dataset provide compelling evidence of the proposed method's effectiveness in terms of cloud removal and texture preservation.

**Key words:** remote sensing imagery cloud removal, Fourier residual block, channel-spatial attention, generative adversarial networks, spatial distribution of clouds

**Supported by** Supported by National Natural Science Foundation of China (No. 61906170); Science and Technology Plan for Zhejiang Province (No. LGF21F020023, Y24F030030); Plan Project of Ningbo Municipal Science and Technology (Nos. 2022Z233, 2021Z050, 2022S002 and 2023J403)