

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¹, 王仁芳^{1, 2}, 邱虹², 王峰², 高广^{2, 3}, 吴敦^{2, 3}

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: P2 **文献标志码:** A

引用格式: 郑其光,王仁芳,邱虹,王峰,高广,吴敦.2023. 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, X(X):xx-xx

Zheng Q G, Wang R F, Qiu H, Wang F, Gao G and Wu D. 2023. A remote sensing image cloud removal network based on attention mechanism and generate adversarial network. National Remote Sensing Bulletin, X(X):xx-xx[DOI: xxx]

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,

收稿日期: 20xx-xx-xx; 预印本:20xx-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, 1996年生, 男, 研究生, 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, 1974年生, 男, 教授, 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

30 有效的去云技术在遥感影像解译领域得到了广泛关 57 获得三个输入，并根据这些不同输入之间的外观差
31 注。 58 异计算像素级置信度图，通过门控派生输入的重要
32 传统的去云方法主要包括基于频率域变换的去 59 特征得到最终的去雾图像。Li 等人(Li 等, 2023)基于
33 云方法和基于先验知识的去云方法。基于频率域变 60 光学和 SAR 遥感影像的层次融合，提出了多级光谱
34 换的去云方法认为遥感影像上的云是低频噪声，通 61 和结构保持融合网络 HS²P，该网络提出了具有堆叠
35 过在频率域中抑制低频信息，增强高频信息，从而 62 残差群的深度分层架构，同时使用信道注意机制来
36 降低云层的影响，增强地面信息的表达。如小波变 63 自适应选择具有信息性的特征进行融合。这类方法
37 换(Wang 等, 2013)、同态滤波(Shen 等, 2014)等。基 64 在去云任务中表现不错，但是需要解决影像配准、
38 于先验知识的去云方法通过对大量有薄云的遥感图 65 几何和辐射校正问题，而且需要保证影像相对时
39 像和无云影像进行统计，得出二者之间的关系，并 66 间差小（即保证两幅影像中该地区地物的变化小），
40 利用这种转换关系来去除薄云得到清晰的遥感影 67 同时云区也不可重叠。
41 像。如直方图匹配法(Chen 等, 2018)、暗通道先验理 68 基于单幅的遥感影像处理方法使用单源、单时
42 论(Liu 等, 2016)等。上述去云方法通常需要人为设 69 相数据，利用图像补绘或图像生成技术来恢复云层
43 定先验条件，非常容易受到环境因素的影响，效果 70 影像中的缺失信息，可以有效地将云层影像转换为
44 并不理想。 71 无云影像。Li 等人提出了去雾深度神经网络
45 近年来，深度学习因其强大的特征学习能力， 72 AOD-Net(Li 等, 2017)，该模型避免了估计传输矩阵
46 能够学习复杂场景中的复杂关系，因而在遥感影像 73 和大气光参数等步骤，通过轻量级 CNN 直接生成干
47 去云中得到了广泛使用。现有的深度学习方法根据 74 净的图像，这种端到端设计可以很容易地将该网络
48 处理的对象可以分为基于多幅的遥感影像处理方法 75 嵌入到其他深度模型。Liu 等人基于注意力机制提出
49 和基于单幅的遥感影像处理方法。基于多幅的遥感 76 了多尺度网络模型 GridDehazeNet(Liu 等, 2019)，该
50 影像处理方法通常使用多源、多时相数据，用同一 77 网络不依赖于大气散射模型，可以高效地交换不同
51 地区的无云影像替代云覆盖区域的影像。Mirza 等人 78 尺度的信息，从而可缓解多尺度估计的瓶颈问题。
52 (Mirza 等, 2014)提出了一种多光谱条件模型 79 Dong 等人将 U-Net 和 GAN 相结合，提出了
53 McGANs，该模型使用合成的云卫星图像和近红外 80 two-stage(Dong 等, 2020)模型，该模型可有效去除厚
54 波段图像对网络进行训练，使其能够从云图中自动 81 云并修复影像。Pan 将空间注意力模块与 GAN 网络
55 生成无云图像。Ren 等人(Ren 等, 2018)通过应用白 82 结合提出了 SpA GAN(Pan, 2020)，利用空间注意力
56 平衡、对比度增强和 Gamma 校正从原始模糊图像中 83 机制生成云层的分布图，并指导后续模块去云。Zhou

84 等人提出了遥感影像去云的多尺度残差神经网络模
 85 型 MSAR-DefogNet(Zhou 等, 2022), 使用具有更大
 86 卷积核的多尺度卷积模块扩大感受野实现节省计算
 87 资源的目的。Ding 等人(Ding 等, 2023)提出了基于条
 88 件变分自编码器的薄云去除网络, 为每个输入云层
 89 影像生成多个合理的无云影像, 通过不确定性分析,
 90 在这些多重预测的基础上生成清晰影像。Zi 等人
 91 (Zi 等, 2023)基于编码器-解码器架构提出了小波集
 92 成卷积神经网络模型 WaveCNN-CR, 通过小波变换
 93 提取多尺度和多频率分量, 设计增强特征提取模块
 94 对不同尺度下的高频分量提取特征, 并从全局和局
 95 部角度对特征进行增强。上述利用残差块的单幅遥
 96 感影像处理方法能够取得有效的去云效果, 然而尽
 97 管这种使用传统的残差块来捕获特征的方法能够很
 98 好地捕获高频特征, 但往往忽视了云层影像和真实
 99 影像对之间的低频信息在图像整体外观上的恢复;
 100 同时对通道、空间特征关注较少, 使得图像的特
 101 征未得到充分利用, 从而难以有效的保留纹理细节。
 102 针对上述基于单幅的遥感影像去云方法存在的
 103 问题, 本文提出一种基于注意力机制的生成对抗网
 104 络去云算法。该算法基于残差学习设计了傅里叶残
 105 差模块, 使模型在浅层便能获取全局上下文信息的
 106 同时, 集成云图低频和高频残差信息; 为了获得充
 107 分的云层分布信息, 利用注意力机制设计通道-空间
 108 注意力模块; 为了使生成图像和真实影像在内容上
 109 保持一致性, 联合 condition GAN 损失、L1 损失、
 110 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

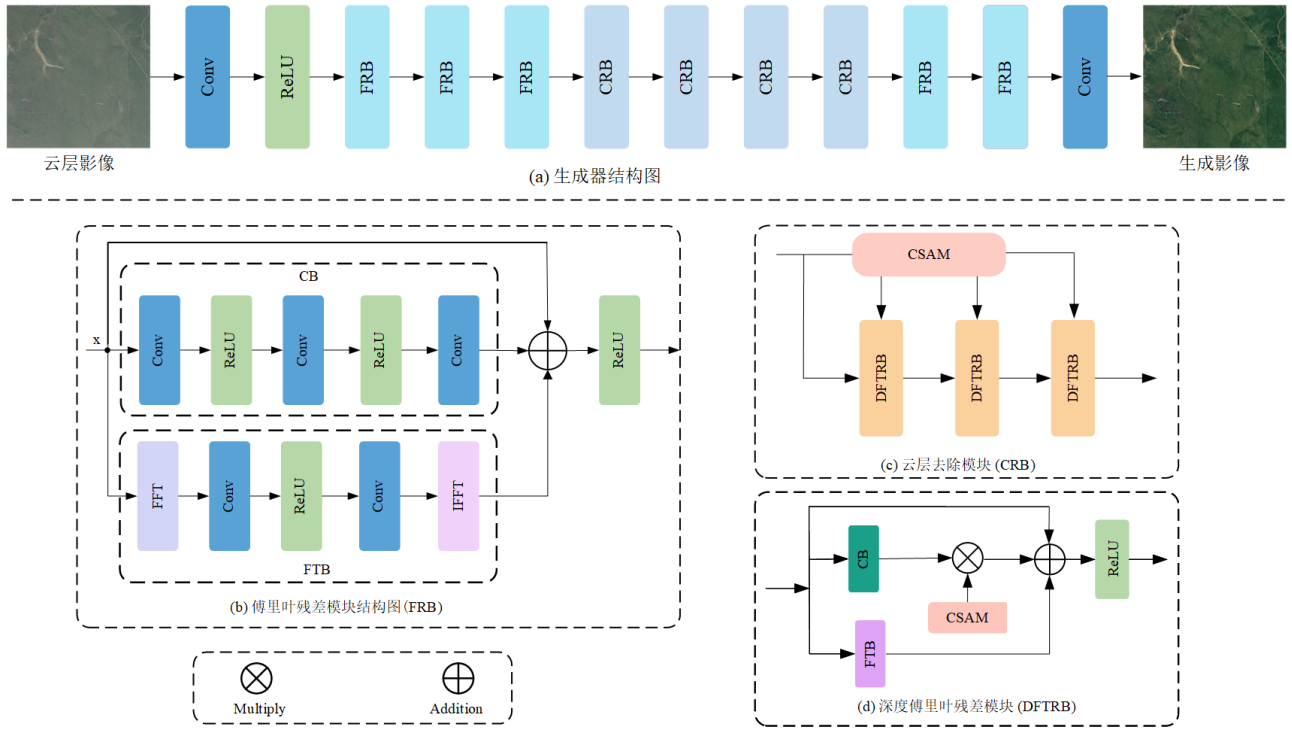
在本章节, 介绍本文提出的基于注意力机制和
 生成对抗网络的遥感影像云去除方法的整体网络模
 型结构、设计的傅里叶残差模块和通道-空间注意力
 模块和构造的联合损失函数。

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。

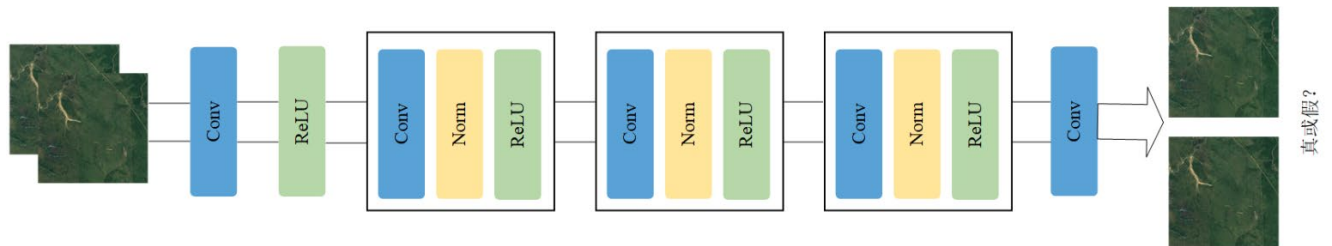
判别器的结构如图 2 所示, 由卷积层、非线性
 激活层 (ReLU) 和归一化层构成, 输入是生成影像
 和真实影像, 输出的结果表示判别器认为生成影像
 是真实影像的概率。



136

137 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
 138 差块 (FRB), FFT 表示傅里叶变换, IFFT 表示傅里叶反变换 (c) 是云层去除模块 (CRB) (d) 是深度傅里叶变换残差
 139 块 (DFTRB)

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



143

144

图 2 判别器结构图

145

Fig.2 The structure of the discriminator

146 **2.2 傅里叶残差模块**

153 模型后性能有了较大的提升, 其认为在频率域中应

147 残差网络(He 等, 2015)常被用于构建遥感影像
 148 去云神经网络以学习云层影像和无云影像之间的差
 149 异。残差学习策略可以较好地捕获图像中的高频特
 150 征, Mao 等人通过在残差网络中引入傅里叶变换、
 151 ReLU 操作和傅里叶反变换提出了 Res FFT-ReLU
 152 Block, 发现将 Res FFT-ReLU Block 加入到去模糊

54 用 ReLU 和卷积操作后, 再进行傅里叶反变换可以
 55 使神经网络灵活的选择频率阈值, 同时在频率域中
 56 引入非线性操作能为网络带来全局上下文学习能力
 57 (Mao 等, 2022)。受此启发, 为使模型可以同时提取
 58 关于遥感影像的高频特征和低频特征, 本文构建了
 59 傅里叶残差模块 (FRB) 如图 1 (b) 所示。令 x 表

160 示 FRB 的输入特征，FRB 可表示为：

$$161 \quad \text{FRB} = \text{ReLU}(\text{CB} + \text{FTB} + x) \quad (1) \quad 185$$

$$162 \quad \text{CB} = \text{Conv11}(\text{ReLU}(\text{Conv33}(\text{ReLU}(\text{Conv11}(x)))))) \quad (2)$$

$$163 \quad \text{FTB} = \mathcal{F}^{-1}(\text{Conv11}(\text{ReLU}(\text{Conv11}(\mathcal{F}(x)))))) \quad (3) \quad 186$$

164 其中， Conv11 和 Conv33 分别表示卷积核大小为 1×1 和

165 和 3×3 的卷积层， \mathcal{F} 表示傅里叶变换， \mathcal{F}^{-1} 表示傅

166 里叶反变换。 CB 为三个卷积层和两个非线性激活函

167 数（ReLU）所组成的卷积块，和 FRB 中第一条分

168 支构成了残差结构，作用是学习影像中的高频特征

169 FTB 为快速傅立叶变换、卷积层、ReLU 激活层和

170 傅立叶反变换所组成的傅里叶变换模块，其作用为

171 学习影像中的低频特征。从图 1(b)中可以看到，FRB

172 的输入在 FTB 中将首先被变换到频域，经过卷积和

173 ReLU 层处理后的频率信息通过傅立叶反变换映射

174 到空间域。

175 2.3 通道-空间注意力模块

176 注意力机制符合人类观察物体的偏好，重点观

177 察突出的特征区域，其中通道注意力机制主要关注

178 特征图中重要的通道。Wang 等人(Wang 等, 2020)

179 认为避免降维、有适当的通道交互是十分重要的，

180 因此在 SENet(Hu 等, 2019)的基础上提出了一种不

181 降维的局部跨通道交互模型 ECA-Net。为了扩大网

182 络的感受野同时获得更多的通道维度上的特征信

183 息，本文在 ECA-Net 的基础上加入全局最大池化

184 （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

187 相加，并通过卷积层和激活层确定通道的权重，最

188 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

190 为：

$$191 \quad F_{CAM} = x \otimes \text{sigmoid}(\text{Conv}(\text{GAP}(x) + \text{GMP}(x))) + x \quad (4)$$

192 其中， $x \in \mathbb{R}^{32 \times H \times W}$ 表示 CAM 的输入特征，

193 $F_{CAM} \in \mathbb{R}^{32 \times H \times W}$ 表示 CAM 的输出特征，H 和 W 是图

194 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

195 文基于递归神经网络(Hu 等, 2020)构建了空间注意

197 力模块（SAM），如图 3 所示。其为一个两轮次、

198 四方向（上，下，左，右）具有 ReLU 激活和单位

199 矩阵初始化的递归神经网络，能够同时捕获局部及

200 全局上下文语义信息。在第一轮中，SAM 的目的是

201 总结输入图像每个位置的相邻上下文信息并生成特

202 征图；在第二轮中，SAM 进一步收集非局部上下文

203 信息，以生成全局感知特征图。SAM 所生成的特征

204 图即为用于引导后续去云的云层空间分布信息，在

205 RICE 数据集(Lin 等, 2019)上生成的可视化结果如

206 图 8 所示，颜色越红表示分配越多的注意力，颜色

207 偏蓝表示分配较少的注意力。

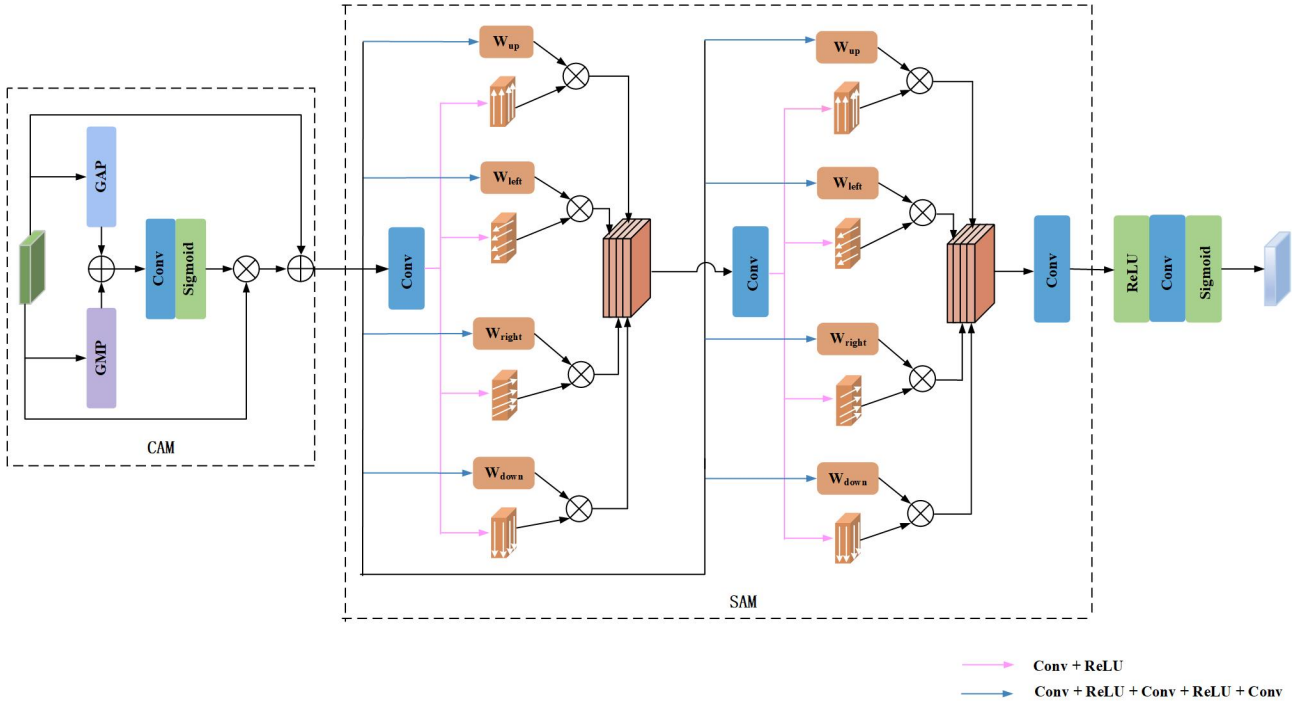


图 3 CSAM 的结构图

Fig.3 The structure of CSAM

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：

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

其中， x 表示云层影像， z 表示生成影像， y 表示真实影像， ζ 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：

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

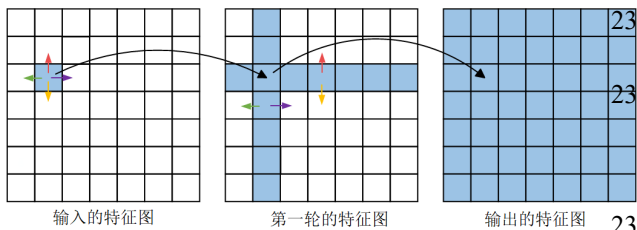


图 4 SAM 累计全局上下文语义信息的示例图

208

209

210

211 SAM 通过两轮次、四方向上的循环卷积累积全

212 局上下文语义信息，如图 4 所示。在第一轮次中，

213 通过对输入特征图上的每个位置分别在四方向(上、

214 左、下、右)进行循环卷积运算，获得水平和垂直邻

215 域信息；在第二轮次中，通过重复第一轮的操作，

216 将获得整个特征图的语义信息。令 $h_{i,j}^{right}$ 表示图像在

217 像素 (i, j) 处向右卷积后的特征，那么向右方向上的

218 循环卷积结果可表示为：

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

220 其中， α_{right} 是右方向的权重参数，其他三个方向类

221 似。

222

223

224

225

226

227

228

229

230

231

232

233

234

235

236

237

238 式 (8) 中, n 为像素点个数。

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

$$243 \quad L_{Attention} = \|F - M\|_2^2 \quad (9)$$

244 感知损失定义如公式 (10), 它计算真实影像
245 和生成影像之间的相似性。本文使用在 ImageNet 数
246 据集上预训练好的 VGG19 模型提取图像特征,
247 $C_j H_j W_j$ 是第 j 层对应的特征图大小。

$$248 \quad L_{per} = \frac{1}{C_j H_j W_j} \|VGG19_j(z) - VGG19_j(y)\|_2^2 \quad (10)$$

249 3 实验结果与分析

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

254 3.1 数据集

255 RICE 遥感影像数据集包含 RICE1 和 RICE2 两
256 个部分, 所有的图像都为 512×512 像素。RICE1 是
257 一个在 Google 地球上收集的薄云数据集, 总共包含
258 500 对图像。通过设定一定的时间间隔, RICE2 收
259 集了 Landsat 8 OLI/TIRS 卫星的 736 对数据。将
260 训练集和测试集按照 4: 1 的方法划分, 对于 RICE1
261 数据集, 分配 400 张用于训练, 100 张用于测试; 284

262 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。

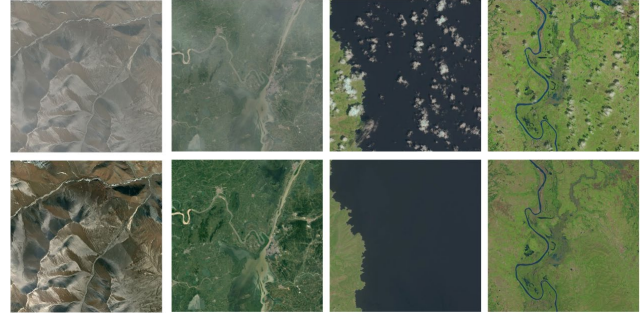


图 5 RICE 数据样本, 第一行是云层图像, 第二行是
对应的真实图像。前两列是 RICE1 数据集的样本, 后两
列是 RICE2 数据集的样本

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.

272 3.2 训练环境与参数配置

273 本文的所有实验均在硬件设备为 i7-12700 @
274 2.10 GHz CPU, NVIDIA GeForce RTX3060GPU, 软
275 件环境为 Pytorch 1.12 深度学习框架, Windows 11 64
276 位的操作系统上运行。使用 Adam 优化器, 训练初
277 始学习率为 0.0004, 采用余弦退火策略对学习率进
278 行周期调整, epoch 为 200 进行批次训练。

279 3.3 评价指标

280 为了便于同其他算法进行对比, 我们使用峰值
281 信噪比(PSNR)和结构相似度指标(SSIM)作为评价指
282 标。其中 PSNR 的计算方式如公式 11 和公式 12 所
283 示, 其中 H 和 W 分别表示图像的高度和宽度, Z 是
284 生成影像, Y 是无云影像, MSE 表示均方误差, MAX

285 指的是像素值的最大可能范围。

$$286 \quad \text{PSNR} = 10 \log_{10} \left(\frac{\text{MAX}^2}{\text{MSE}} \right) \quad (11)$$

$$287 \quad \text{MSE} = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W (Z(i, j) - Y(i, j))^2 \quad (12)$$

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

$$290 \quad \text{SSIM}(Z, Y) = \frac{(2\mu_Z\mu_Y + c_1)(2\sigma_{ZY} + c_2)}{(\mu_Z^2 + \mu_Y^2 + c_1)(\sigma_Z^2 + \sigma_Y^2 + c_2)} \quad (13)$$

291 其中, c_1 和 c_2 是常量, μ_Z 和 μ_Y 表示图像的平均亮度,

292 以 μ_Z 为例

$$293 \quad \mu_Z = \frac{1}{N} \sum_{i=1}^N z_i \quad (14)$$

294 σ_Z 和 σ_Y 表示标准差, 以 σ_Z 为例

$$295 \quad \sigma_Z = \left(\frac{1}{N-1} \sum_{i=1}^N (z_i - \mu_Z)^2 \right)^{1/2} \quad (15)$$

296 σ_{ZY} 表示 Z 和 Y 的方差, 计算方法如公式 16 所示:

$$297 \quad \sigma_{ZY} = \frac{1}{N-1} \sum_{i=1}^N (z_i - \mu_Z)(y_i - \mu_Y) \quad (16)$$

298 3.4 对比实验

299 五种算法在 RICE1 上的部分可视化结果如图 6

300 所示。从左到右, 每一列分别表示云层影像、HALP

301 去云结果、Cloud GAN 去云结果、Pix2Pix 去云结果、

302 SpA GAN 去云结果、AtM GAN 去云结果和对应的

303 真实影像。由于 RICE1 中大部分是薄云影像, 地物

304 信息仍然保留了比较清晰的纹理细节, 故各个算法

305 生成的结果从视觉上看和对应的真实影像都有一定

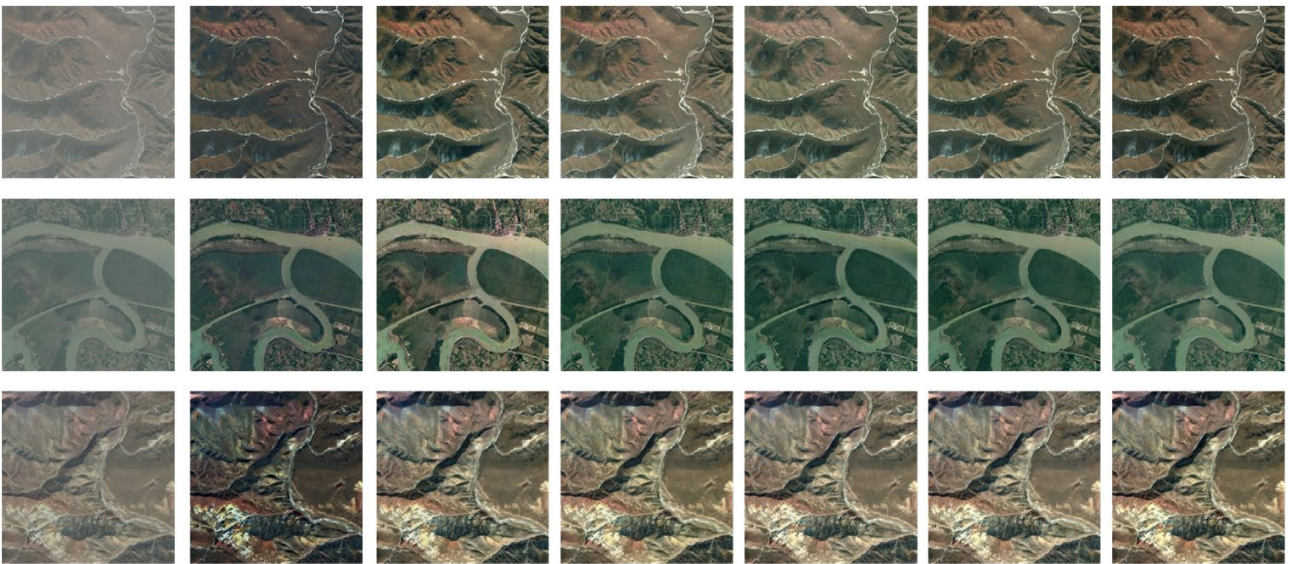
306 的相似之处。但是 HALP 和 Cloud GAN 的去云图像

307 整体出现了较大的颜色失真, 整体与真实的无云影

308 像存在较大的差异; Pix2Pix 和 SpA GAN 在部分区

309 域也存在一定的颜色差异, 本文提出算法的结果与

310 真实影像最为接近。



311 云层影像

312 HALP

Cloud GAN

Pix2Pix

SpA GAN

本文

真实影像

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

314 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

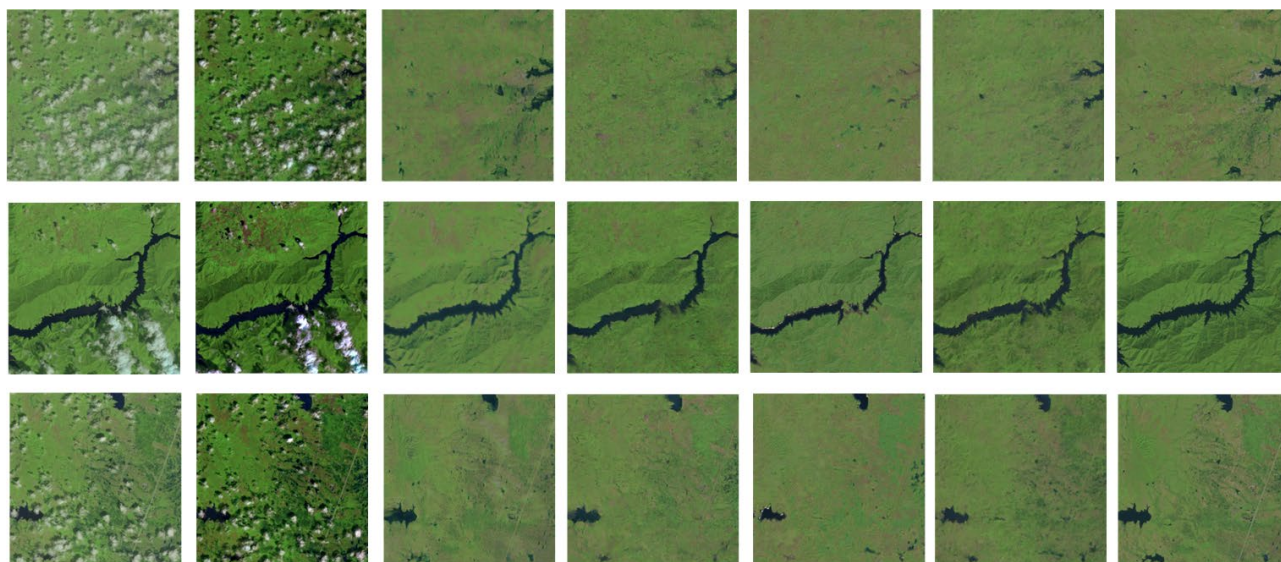
316 AtM-GAN 在 PSNR 和 SSIM 的指标分别是 32.492B25 五种方法在 RICE2 数据集上的部分可视化结果
 317 和 0.9746, 在对比的几个方法中都达到了最好的效26 如图 7 所示。从左到右, 每一列分别表示云层影像、
 318 果, 得益于建模了云层影像和真实影像对之间的高27 HALP 去云结果、Cloud GAN 去云结果、Pix2Pix 去
 319 频和低频差异, 云层区域的准确定位和联合损失函28 云结果、SpA GAN 去云结果、AtM GAN 去云结果
 320 数最小化生成影像和真实影像之间的像素差异和结29 和相对应的真实影像。由于 RICE2 中的图像大多数
 321 构差异。 330 为厚云, 相较于薄云, 厚云中被遮挡像素点的数据

322 表 1 在 RICE1 数据集上的云去除性能对比 331
 323 Table 1 Quantitative comparison of different method
 324 for cloud removal on the RICE1 dataset 332

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

333 信息是永久性缺失的, 这种伤害也是无法逆转的。
 334 从图 7 可以看出, HALP 方法仍然存在大量未去除
 335 的云层区域; Cloud GAN 和 Pix2Pix 方法去云后存
 336 在较多的云层轮廓; SpA GAN 方法虽然对云层区域
 337 定位较为准确, 但是也存在局部区域细节丢失的问
 338 题; 本文方法去云图像无明显的颜色失真, 且细节
 339 纹理体现较好, 与真实的标签影像较为接近。

338



云层影像

HALP

Cloud GAN

Pix2Pix

SpA GAN

本文

标签影像

339

340

341

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

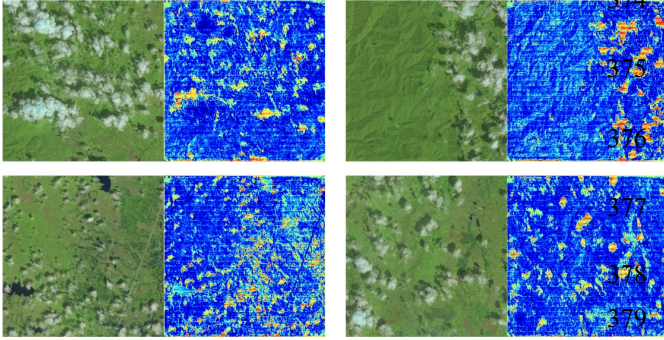
342 五种算法在 RICE2 上的 PSNR 和 SSIM 指标如345 标在几个对比算法中都达到了最好的效果, 两个指
 343 表 2 所示, 从表中可以看出本文提出的方法在 PSNR346 标相对于 SpA GAN 分别提升了 3.193 和 0.0322。这
 344 和 SSIM 的指标分别是 31.3331 和 0.9359, 这两个指47 表明了注意力机制和傅里叶残差结构在云层去除任

348 务中的有效性。

349 表 2 在 RICE1 数据集上的云去除性能对比

350 Table 2 Quantitative comparison of different method
351 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



352 图 8 RICE 数据集生成的云层空间分布信息图

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

354 3.5 消融实验

355 为了评估傅里叶残差模块和 CSAM 对整个网络
356 去云性能的影响，本文按照网络结构及相关模块设

357 计了消融实验，共包含 5 个实验：(1)WFCS(without
358 FTB, CAM and SAM)

359 表示 FRB 中没有采用 FTB 分
360 支，同时没用使用 CSAM 中的 CAM 和 SAM。(2)

361 FS (FTB and SAM)表示 FRB 中引入了 FTB, CSAM
362 中引入了 SAM, 但是 CSAM 中没有引用 CAM。(3)

363 CS (CAM and SAM)表示 CSAM 中引入了 CAM 和
364 SAM, 但是 FRB 中没有引入 FTB。(4)FC (FTB and

365 CAM)表示 FRB 中引入了 FTB, CSAM 中引入了
366 CAM, 但是没用引入 SAM。(5)FCS (FTB, CAM and

367 SAM)表示 FRB 中引入了 FTB, 同时 CSAM 中使
368 用了 CAM 和 SAM。

369 在 RICE1 和 RICE2 数据集上进行的消融实验结

370 果分别如表 3 和表 4 所示。从表 3 中可以看到 FS 的

371 SSIM 值高于 FCS, 主要由于 RICE1 中的图像大部

372 分区域覆盖了薄云, 导致特征在 CAM 中经过 GAP

373 和 GMP 后值偏大, 使得生成图像的像素整体偏高,

374 从而导致 FCS 中的 SSIM 值偏低。从表中可以看出,

375 FRB、CAM 和 SAM 三者的使用相对于 WFCS 有了

376 明显的提升, 说明了傅里叶残差模块和注意力机制

377 在云层去除任务中的有效性。

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 RICE1 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

384 本文针对现存的云层去除模型无法更好的保留

385 图像边缘纹理细节问题, 在生成对抗网络框架下,

- 386 利用傅里叶残差模块建模云层影像的低频和低频残
387 差信息,使用通道-空间注意力机制来获取云层分布
388 图,同时构建联合损失函数约束生成影像和真实影
389 像的内容一致性。实验结果表明,在 RICE1 和 RICE2
390 数据集上,本文提出的 AtM-GAN 网络均能实现很
391 好的云层去除效果,在 RICE1 上 PSNR 和 SSIM 指
392 标分别是 32.4921 和 0.9746;在 RICE2 上 PSNR 和
393 SSIM 指标分别是 31.3331 和 0.9359,这也表明了注
394 意力机制和傅里叶残差块在去除任务中的有效性。
395 与现有主流去云模型相比,本文提出的算法能
396 够有效保留图像边缘纹理细节,在遥感影像云去除
397 任务中取得良好的效果。由于本文设计的模型针对
398 的是单幅遥感影像云去除,尚不能直接应用到多幅
399 遥感影像云去除任务。因此下一步的主要工作是更
400 深入地探索在多时相和多源数据集上云去除模型的
401 设计及其应用。
- 402 **参考文献(References)**
- 403 Chen S L, Chen X H, Chen X, Chen J, Cao X, Shen M G, Yang W
404 and Cui X H. 2018. A Novel Cloud Removal Method Based on
405 IHOT and the Cloud Trajectories for Landsat Imagery. *Remote
406 Sensing*, 10(7): 1040 [DOI: 10.3390/rs10071040]
- 407 Ding H D, Zi Y and Xie F Y. 2023. Uncertainty-Based Thin Cloud
408 Removal Network via Conditional Variational Autoencoders
409 Asian Conference on Computer Vision. Berlin: Springer
410 52-68 [DOI: 10.1007/978-3-031-26313-2_4]
- 411 Dong H, Pan J S, Xiang L, Hu Z, Zhang X Y, Wang F and Yang M
412 H. 2020. Multi-Scale Boosted Dehazing Network With Dense
413 Feature Fusion. *IEEE/CVF Conference on Computer Vision
414 and Pattern Recognition(CVPR)*. Seattle: IEEE: 2154-2164
415 [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

- 453 prospects of multi-source remote sensing image fusion 490
 454 National Remote Sensing Bulletin, 25 (1): 148-166 (李树涛 491
 455 李聪好, 康旭东. 2021. 多源遥感图像融合发展现状与未来 492
 456 展望. 遥感学报, 25 (1): 148-166) [DOI: 493
 457 10.11834/jrs.20210259] 494
 458 Li Y S, Wei F Y, Zhang Y J, Chen W, Ma J Y. 2023. HS2P: 495
 459 Hierarchical spectral and structure-preserving fusion network 496
 460 for multimodal remote sensing image cloud and shadow 497
 461 removal. Information Fusion, 94: 215-228 [DOI: 498
 462 10.1016/j.inffus.2023.02.002] 499
 463 Lin D Y, Xu G L, Wang X K, Wang Y, Xian S and Fu K. 2019. A 500
 464 Remote Sensing Image Dataset for Cloud Removal. 501
 465 arXiv:1901.00600 502
 466 Liu X H, Ma Y R, Shi Z H and Chen J. 2019. GridDehazeNet: 503
 467 Attention-Based Multi-Scale Network for Image Dehazing. 504
 468 International Conference on Computer Vision (ICCV). Seoul: 505
 469 IEEE: 7313-7322 [DOI: 10.1109/ICCV.2019.00741] 506
 470 Liu X, Zhang H, Tang Y Y and Du J X. 2016. Scene - adaptive 507
 471 single image dehazing via opening dark channel model. IET 508
 472 Image Processing, 10(11): 877-884 [DOI: 509
 473 10.1049/iet-ipr.2016.0138] 510
 474 Mao X T, Liu Y M, Liu F Z, Li Q L, Shen W and Wang Y. 2022. 511
 475 Intriguing Findings of Frequency Selection for Image 512
 476 Deblurring. AAAI Conference on Artificial Intelligence, 37: 513
 477 1905-1913 [DOI: 10.1609/aaai.v37i2.25281] 514
 478 Ming Y S, Liu Q H, Bai H and Huang C. 2023. Classification and 515
 479 change detection of vegetation in the Ruergai Wetland using 516
 480 optical and SAR remote sensing data. National Remote 517
 481 Sensing Bulletin, 27 (6): 1414-1425 (明义森, 刘启航, 梅 518
 482 荷, 黄昌. 2023. 利用光学和 SAR 遥感数据的若尔盖湿地植 519
 483 被分类与变化监测. 遥感学报, 27 (6): 1414-1425)[DOI: 520
 484 10.11834/jrs.20221767] 521
 485 Mirza M and Osindero S. 2014. Conditional Generative Adversarial 522
 486 Nets. [DOI: 10.48550/arXiv.1411.1784] 523
 487 Pan H. 2020. Cloud Removal for Remote Sensing Imagery via 524
 488 Spatial Attention Generative Adversarial Network. [DOI: 525
 489 arXiv:2009.13015] 526
 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 &

- 528 Continua, 71(2): 3497-3512 [DOI: 10.1049/ipr2.12224] Zhou Y, Jing W P, Wang J, Chen G S, Scherer R and Damaševičius
 529 10.32604/cmc.2022.022325] 537 R. 2022. MSAR-DefogNet: Lightweight cloud removal
 530 Yang X, Li F, Lu M, Xin L, Lu X T and Zhang N. 2022. New 538 network for high resolution remote sensing images based on
 531 super-resolution reconstruction method based on Mixed Sparse 539 multi scale convolution. IET Image Processing, 16(3): 659-668
 532 Representations. National Remote Sensing Bulletin, 26 (8): 540 [DOI: 10.1049/ipr2.12224]
 533 1685-1697 (杨雪, 李峰, 鹿明, 辛蕾, 鲁啸天, 张南. 2022. 541 Zi Y, Ding H D, Xie F Y, Jiang Z G and Song X D. 2023. Wavelet
 534 混合稀疏表示模型的超分辨率重建. 遥感学报, 26 (8): 542 Integrated Convolutional Neural Network for Thin Cloud
 535 1685-1697) [DOI: 10.11834/jrs.20219409] 543 Removal in Remote Sensing Images. Remote Sensing, 15(3):
 544 781 [DOI: 10.3390/rs15030781]

545

546

A remote sensing image cloud removal network based on attention mechanism and generate adversarial network

547

548

Zheng Qiguang¹, Wang Renfang^{1, 2}, Qiu Hong², Wang Feng², GAO Guang^{2, 3}, Wu Dun^{2, 3}

549

550

1. College of Computer Science and Technology, Ocean University of China, Qingdao 266100, China;

551

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

552

3. Popsmart Technology(Zhejiang) Co., Ltd., Ningbo, 315042, China

553

554

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.

555

556

557

558

559

560

561

562

563

564

565

566 The following conclusions can be drawn: (1) The proposed Fourier residual block plays a pivotal role in
567 adeptly capturing the details of both low-frequency and high-frequency residual information present in cloud
568 images. By merging these diverse features, the block bolsters the network's ability to extract rich features; (2)
569 The attention mechanisms utilized extract features from both spatial and channel dimensions. This enables
570 the model to discern the significance of various features across different dimensions. Moreover, the cloud
571 information obtained through these attention mechanisms allows the model to adaptively present the changes
572 in cloud, thereby enhancing the model's precision and efficiency in cloud removal; (3) Experimental results
573 on the RICE dataset provide compelling evidence of the proposed method's effectiveness in terms of cloud
574 removal and texture preservation.

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

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