

结合U-Net和STGAN的多时相遥感图像云去除算法

王卓^{1,2}, 马骏¹, 郭毅³, 周川杰⁴, 柏彬⁴, 李峰²

1. 河南大学 软件学院, 开封 475100;

2. 中国空间技术研究院 钱学森空间技术实验室, 北京 100094;

3. 澳大利亚西悉尼大学, 悉尼 2751;

4. 北京市遥感信息研究所, 北京 100192

摘要: 针对光学遥感图像中云的遮挡可能会降低甚至完全遮挡图像中的某些地面覆盖信息, 限制对地观测、变化检测或土地覆盖分类等问题, 云去除是迫切需要解决的一项重要任务。为了恢复被云遮挡的地面区域, 提出一种基于多时相遥感图像的两阶段云去除算法。第一阶段是云分割, 即直接使用U-Net提取云并去除薄云。第二阶段是图像恢复, 采用时空生成网络(STGAN)去除厚云, STGAN的生成模型采用改进的多输入的U-Net, 通过一次从同一位置的7帧图像序列中提取关键特征恢复相应的不规则厚云覆盖区域。第一阶段的薄云处理有利于后面的STGAN捕捉到更多的地面信息。实验结果表明, 与传统的去云方法和深度学习Pix2Pix等算法相比较, 该算法无论在视觉效果上, 还是峰值信噪比(PSNR)与结构相似性(SSIM)等客观质量评价指标上, 均有显著的提升, 有利于光学遥感图像的进一步利用。

关键词: 遥感图像, 多时相, 云去除, 图像恢复, U-Net, STGAN

中图分类号: TP701/P2

引用格式: 王卓, 马骏, 郭毅, 周川杰, 柏彬, 李峰. 2024. 结合U-Net和STGAN的多时相遥感图像云去除算法. 遥感学报, 28(8): 2089-2100

Wang Z, Ma J, Guo Y, Zhou C J, Bai B and Li F. 2024. Cloud removal in multitemporal remote sensing imagery combining U-Net and spatiotemporal generative networks. National Remote Sensing Bulletin, 28(8): 2089-2100 [DOI:10.11834/jrs.20232255]

1 引言

光学遥感图像越来越多地应用于包括环境污染监测(Peng等, 2016)、作物分类(Kussul等, 2015)、土地覆盖分类(Kussul等, 2017)等各个领域。然而, 卫星图像经常被云层遮挡——地球表面三分之二的面积会被云层遮挡(King等, 2013)。厚厚的云层会隐藏图像的内容, 造成地物信息完全丢失。即使是薄的半透明云层也会扭曲地面, 造成遥感图像中的地物信息发生模糊, 极大地影响卫星图像的使用。因此, 在遥感卫星图像分析中, 解决卫星图像的这种云遮挡是一个非常重要的问题(Saunders和Kriebel, 1988)。

云去除是一个图像缺失信息重建的过程(Li

等, 2019), 对于去云任务, 根据测试阶段输入数据类型不同, 可以分为4类(Shen等, 2015): (1) 基于空间的方法。这种方法需保证整幅图像各像元之间具有空间相关性, 将同一幅遥感图像上未被云层遮挡的像元当作先验信息, 填补云遮挡区域中缺失的像元。Zheng等(2021)提出用U-Net和GAN进行单时相云去除, 但是当厚云缺失区域较大时, GAN的生成模型可能会用不真实的图像填充缺失区域, 对于同一幅图像, 生成器对缺失区域有不同的填充方式。生成器倾向于以简单模式填充缺失区域, 与主要目标(如湖泊、森林和道路)相比, 汽车、行人等小物体的细节有所丧失。Singh和Komodakis(2018)使用CycleGAN来消除薄云, CycleGAN采用循环一致性损失将薄

收稿日期: 2022-05-23; 预印本: 2023-12-19

基金项目: 国家重点研发计划(编号:2020YFA04100)

第一作者简介: 王卓, 研究方向为多时相遥感图像云去除。E-mail: 104754201750@henu.edu.cn

通信作者简介: 李峰, 研究方向为超时相数据处理和压缩感知。E-mail: lifeng@qxslab.cn

云图像转换为无云RGB图像。该方法训练网络时,不使用配对的云一无云数据集,只须输入单张有云图像,因此只适用于解决单时相薄云去除问题。当云层较厚时,模型会生成过度平滑的图像。Mirza和Osindero(2014)提出了条件生成对抗网络(CGAN),而后赵肖肖(2019)利用Pix2Pix(改进的CGAN)方法实现了高分辨率遥感图像去雾。然而,这些方法仅适用于重建较小的薄云区域或只用于单时相遥感图像去云,对于大规模厚云遮挡的复杂区域,重建精度和可靠性通常无法令人满意。(2)基于光谱的方法。该类方法以多光谱遥感数据为输入。利用同一幅多光谱图像的无云波段来估计云波段和无云波段之间的线性或非线性关系,利用估计的关系恢复云遮挡区域。然而,这种方法通常只能处理薄云,因为厚云很大可能性存在多光谱图像的所有波段中。Enomoto等(2017)利用合成的云图数据和原始图像的近红外(NIR)通道,训练了多光谱条件生成对抗网络(MCGAN)来生成无云图像,但是由于输入的训练图像是采用的柏林噪声(Perlin, 2002)仿真的云,对真实的云图不具有适用性。训练图像与真实有云图像上的性能存在显著差异,严重限制了泛化能力。(3)基于时间的方法。该类方法使用来自同一传感器的多时相图像作为输入,通常用于去除厚云层。在时间图像序列中,利用同一位置无云时间段的影像来替换有云遮挡的影像。Wen等(2018)提出了一种两阶段稳健主成分分析(TRPCA)方法来去除序列卫星图像中的云,TRPCA首先用RPCA做分离,提取云及阴影区域进行初始云检测,随后提出一种判别性RPCA算法,对检测到的云像素和无云区域分配相应的权重,以达到云去除效果,但在初始云检测后使用了形态学操作,导致图像序列较短时去云后的图像出现了一些黑色的矩形块,并且该方法云去除后仍有少量云残留。(4)基于时空光谱的混合方法。通过将基于空间、光谱和时间的方法结合起来获得更好的性能。Sarukkai等(2020)提出了用ResNet和U-Net两种分支作生成器的时空生成网络(STGAN),训练了基于RGB和RGB+IR通道的模型来进行云去除,该模型允许接受3帧有云图像,由于该模型在未对有云图像去除薄云的情况

下直接对有云图像序列用STGAN进行云去除,直接利用STGAN没有去除薄云后再利用STGAN捕获的地面信息多,导致去云效果不好。Zhang等(2018)提出了一种基于深度卷积神经网络(CNN)的统一时空光谱框架,用于重构遥感图像中传感器失效导致的缺失信息,并去除遥感图像中的厚云,但是,它需要大量的训练数据集。

目前云去除中基于深度学习的方法,大多使用GAN来去除单张图像的云遮挡问题,由于遥感传感器的重访,很容易获得同一区域的遥感图像序列,但GAN的生成模型很少利用卫星图像提供的时空信息,这样既没有充分利用获取的同一位置的多时相卫星图像,也造成生成的图像往往在先前被厚云遮挡的区域缺乏细节信息。

2 模型及整体架构

受STGAN可以更好地提取多帧时空特征的启发,本文提出了一种基于时空信息的多时相遥感图像云去除算法。在第一阶段,利用U-Net(Ronneberger等, 2015)进行薄云去除,先进行去薄云处理,可以使后面的STGAN捕捉到更多的地面信息;在第二阶段,将第一阶段去薄云后的多时相图像输入到STGAN中进行厚云去除处理。STGAN的生成模型采用多输入的U-Net,通过一次从同一位置的多个图像中提取关键特征,捕捉一个区域内多幅图像的相关性,从而将多帧有云图像的部分信息合成为单张细节丰富的无云图像。通过对特征进行充分的融合,有效提升了生成云图的真实度。遥感图像多时相图像中薄云与厚云通常同时存在(Li等, 2017; Zheng等, 2021),本文算法适用于解决这种薄厚云同时存在的情况。本文方法不需要单独的云检测方法,在去除薄云的同时,也达到了云检测的目的。通过大量实验和验证结果表明该算法的云去除效果较好。

U-Net是一种用于图像分割的卷积神经网络,尤其在生物分割医学领域,得到了广泛的应用。我们将其应用在薄云去除中,视为分割前景云和背景地面,U-Net可以将薄的半透明云进行去除并且将云区域检测出来。而STGAN擅长提取每个图像中的关键特征并充分利用和融合多时相图像的时空信息,生成单个细节丰富的输出图像。我们的两阶段算法将两个网络进行结合,先去除薄云,

再去除厚云, 既可以解决U-Net无法处理厚云区域云遮挡的问题, 又比直接利用STGAN进行云去除能够捕获到更多的地面信息。

2.1 卫星图像云模型

训练图像的数据集采用真实云图与真实无云地面图借助卫星图像的大气散射模型 (Narasimhan和Nayar, 2002) 合成的云图。设 $S, G, C \in R^{n1 \times n2}$, 分别为合成的有云卫星图像 (RGB图像)、地面图像 (RGB图像) 和云图像 (灰度图像)。采用大气散射模型, 将云图像 C 叠加在地面图像 G 上合成有云卫星图像 S :

$$S = (1 - C) \circ G + C \quad (1)$$

式中, S, G, C 的像素归一化到范围 $[0, 1]$, 1 表示一个 $n1 \times n2$ 的全是 1 的矩阵, \circ 表示哈达玛积 (也称为对应元素逐个相乘), 表示取两个相同维数的矩阵 X 和 Y , 生成一个相同维数的矩阵 Z , 使 $Z_{ij} = X_{ij}Y_{ij}$ 。U-Net 是一种用于图像分割的卷积神经网络, 适用于薄云去除, 不适用于厚云去除任务, 本文中经过第一阶段 U-Net 去云后仍然存在云遮挡

的区域, 称为厚云区域。云层厚度 C 对于合成图像 S 有一定的影响, 在有云图像中, 对于云层厚度为 $C_{ij} \approx 0$ 的薄云区域, 则有 $S \approx G$; 而对于云层厚度为 $C_{ij} \approx 1$ 的厚云区域, 则有 $S \approx C$ 。厚云往往不是单独存在, 会掺杂一部分薄云, 先进行去薄云处理有利于 STGAN 捕捉到更多的地面信息, 使得去厚云任务更有针对性。

2.2 整体流程

本文提出了一种两阶段的多时相图像云去除方法, 两阶段方案的过程如图 1 所示。在第一阶段, 将 7 帧采用大气散射模型合成的有云卫星图像 $S_i (i = 1, 2, \dots, 7)$ 输入到 U-Net 中进行云分割, U-Net 将 7 帧有云的卫星图像分别分割为 7 帧地面图像 $G'_i (i = 1, 2, \dots, 7)$ 和 7 帧云图像 $C'_i (i = 1, 2, \dots, 7)$, 来达到薄云去除和云检测的目的。在第二阶段, 通过将得到的 7 帧地面图像 $G'_i (i = 1, 2, \dots, 7)$ 输入到 STGAN 中进行多帧特征提取和特征融合, 得到单张细节丰富且精度较高的无云地面图像 G'' 。

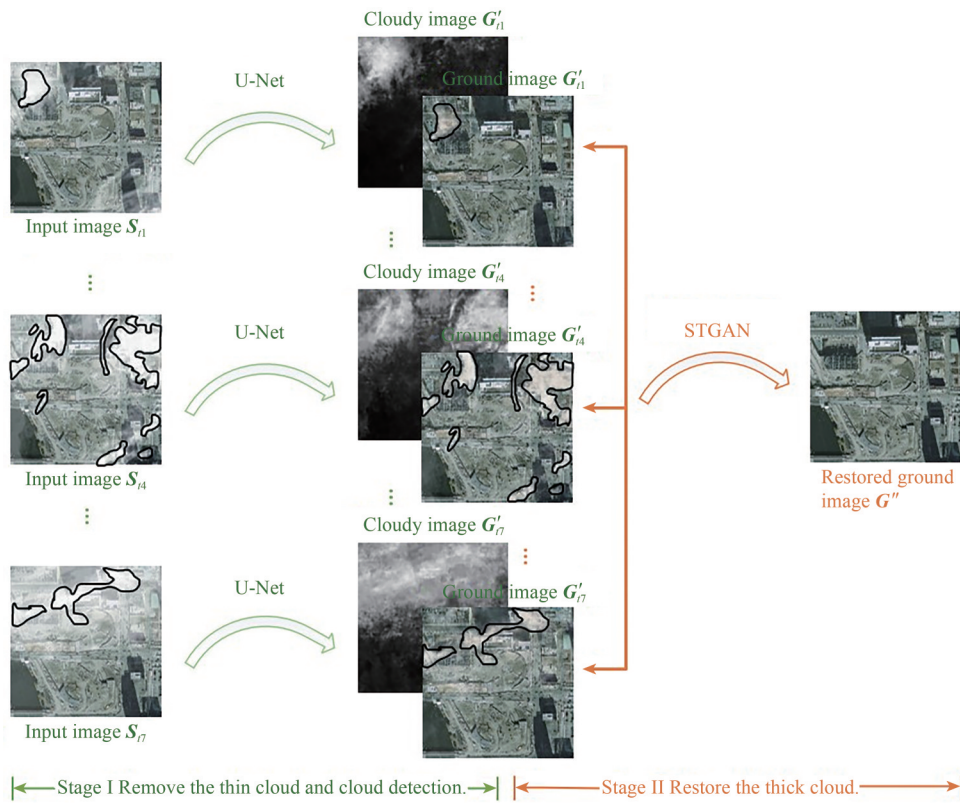


图 1 两阶段的多时相云去除方法整体流程

Fig. 1 The overall process of the two-stage multi-temporal cloud removal algorithm

2.3 网络结构

2.3.1 U-Net 网络结构

第一阶段U-Net的网络结构如图2所示。U-Net网络有11层，前5层形成收缩路径，第6至第10层形成扩展路径。在收缩路径中，通过卷积和下采样来降低图像尺寸，提取一些浅层次的特征，网络第j层的大小为 $n_1/2^j \times n_2/2^j \times (24 \times 2^{j-1})$, $j = 1, 2, 3, 4, 5$, 卷积块（黑色箭头）执行步长为

1或2的多通道2D卷积。在扩展路径（从第6层到第10层）中，通过卷积和上采样来获取一些深层次的特征。由于黑色箭头的复制和连接操作，将编码阶段获得的特征图同解码阶段获得的特征图结合在一起，U-Net的扩展路径每层的通道数增加了一倍，以此结合深层次和浅层次的特征，细化图像，根据得到的特征图进行预测分割。第11层是具有卷积操作和修正线性单元（ReLU）激活函数的标准输出层。

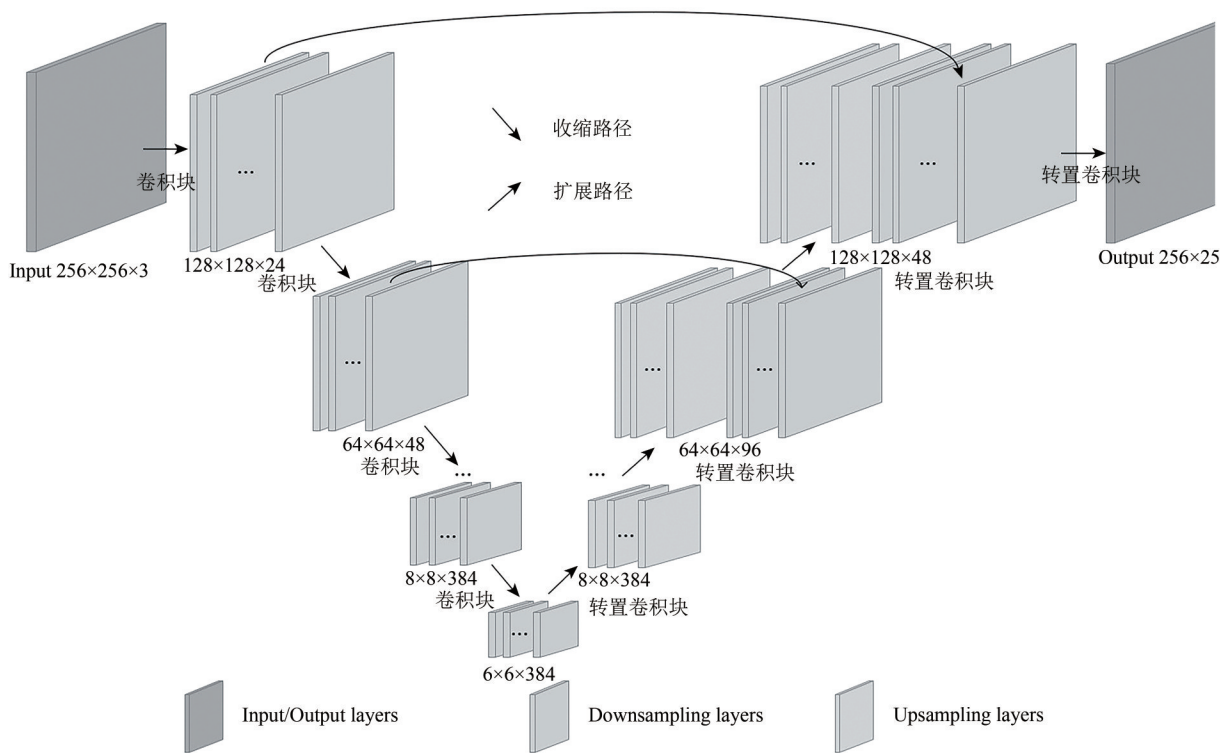


图2 U-Net网络结构

Fig. 2 Encoder-decoder network structure for the U-Net

卷积块由多通道2D卷积和ReLU组成，卷积核大小为3x3，步长为2。转置卷积块由多通道2D转置卷积、批处理归一化（BN）和ReLU组成，卷

积核大小为3x3，步长为2，在转置卷积和ReLU激活函数之间进行BN运算。卷积块和转置卷积块如图3所示。

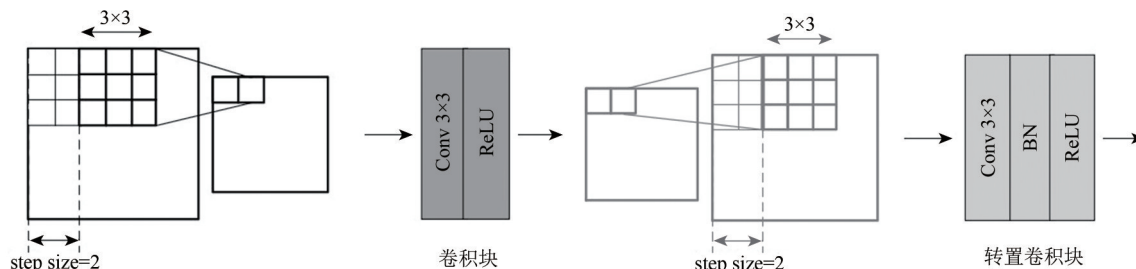


图3 卷积块和转置卷积块

Fig. 3 Conv block and transposed conv block

2D卷积和2D转置卷积是相似的。在滑动窗口中通过在输入张量上移动来执行卷积操作。在2D卷积操作中,卷积核为 $(n1, n2, I, O)$,其中 $n1, n2$ 为卷积核的尺寸, I 为输入特征的通道个数, O 代表卷积核的个数。在2D卷积和2D转置卷积中设置步长为2,分别进行下采样和上采样。

ReLU: 一个非线性的激活函数,可以表示为

$$y = \max(0, x) \quad (2)$$

BN: 为了加速网络的收敛,加快训练,提高泛化能力,利用BN运算使权重张量接近正态分布,具体如下:

$$y = \frac{x - \mu}{\sigma + \delta} \beta_1 + \beta_2 \quad (3)$$

式中, μ, σ 分别为训练时一个批次中每个维度计算的平均值和标准偏差, β_1, β_2 是从训练中学习到的, δ 是一个小常数,以避免数值问题。

2.3.2 STGAN网络结构

为了充分利用和融合时空信息,本文采用了一种改进的GAN生成器(Sarukkai等,2020)结

构,生成器采用多输入的U-Net网络。判别器使用了PatchGAN(Isola等,2017)。如图4所示,生成器修改了传统的U-Net架构,允许每个输入图像分别编码,然后一起解码。多输入的U-Net建立在具有8个下采样块和8个上采样块的通用U-Net框架上,在第 i 个块和第 $n-i$ 个块之间添加了跳跃连接(块的总数 $n=16$)。每个块都由一个ReLU,卷积层和BN层组成。在上采样块和下采样块中,卷积核的大小均为 4×4 ,步长均为2,填充均为1。

7帧输入图像在7个单独的下采样管道中进行下采样,在最终下采样块之后,将得到的所有特征图进行通道拼接,从同一位置的多个图像中提取关键特征,之后将其传递到单个上采样管道。在每个上采样块之后,通过跳跃连接,将输出与对应的7个上采样管道中的每个上采样块的输出连接起来,然后送入下一个上采样块,在最终的上采样块之后,得到最终的输出,即分别对图像进行下采样从每个图像中提取关键特征,之后将图像通道拼接进行上采样以生成单个细节丰富的无云地面图像。

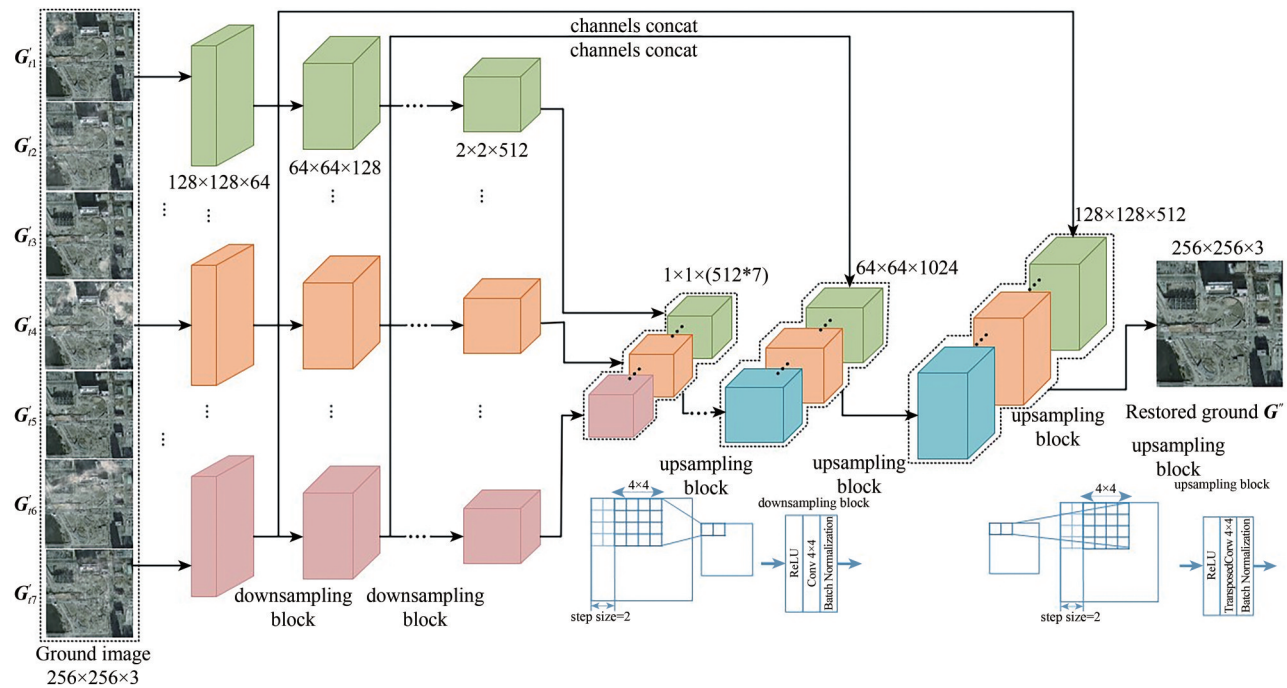


图4 STGAN生成器网络结构

Fig. 4 Network structure for the generator of STGAN

2.4 U-Net训练

U-Net的训练过程如图5所示。在训练U-Net时,将单帧有云卫星图像 $S \in R^{n1 \times n2 \times 3}$ 映射为一帧地面图像 $G' \in R^{n1 \times n2 \times 3}$ 和一帧云图 $C' \in R^{n1 \times n2 \times 1}$,

如图5所示。训练过程中,每次迭代时,对一批图像进行采样,在 $|\Omega| = m$ 的情况下随机采样 $\Omega \subseteq n$,采用式(1)将地面图像 $G \in R^{n1 \times n2 \times 3}$ 和一帧云图 $C \in R^{n1 \times n2 \times 1}$ 合成单帧有云的卫星图像,然后通过

梯度下降法来更新网络权重。损失函数由 L_{cloud} 和 L_{ground} 两部分组成。

$$L_{\text{cloud}} = \frac{1}{m} \sum_{i \in \Omega} \|C_i - C'_i\|_F^2 \quad (4)$$

$$L_{\text{ground}} = \frac{1}{m} \sum_{i \in \Omega} \|(1 - C_i) \circ (G_i - G'_i)\|_F^2 \quad (5)$$

式中, m 表示样本的个数, $\|\cdot\|_F$ 表示 F 范数。对于 L_{cloud} , 表示真实的云图 C 与 U-Net 分割出的云图 C'

之间的差值。对于 L_{ground} , 不直接使用 $\|G_i - G'_i\|_F^2$, 而使用 $\|(1 - C_i) \circ (G_i - G'_i)\|_F^2$ 作为损失函数, 即它们都叠加上云层后, 再进行比较, 在第一阶段后, 厚云仍然存在, 这样即便云层厚度大的区域恢复情况很差, 也不会反映到损失上去, 模型可以更加专注于云层的提取和地面的薄云去除工作。

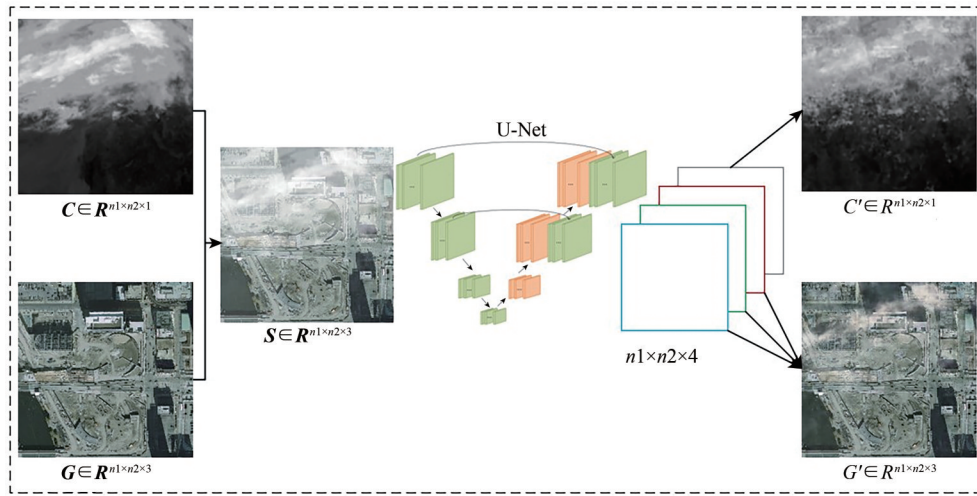


图5 U-Net的训练

Fig. 5 Training of U-Net

2.5 STGAN训练

在训练STGAN时, 生成器G的输入为7帧去薄云后的地面图像 $G'_i \in R^{n1 \times n2 \times 3}$ ($i = 1, 2, 3, 4, 5, 6, 7$), 利用生成器G生成无云的地面图像 $G'' \in R^{n1 \times n2 \times 3}$, 然后将生成器G的参数固定, 训练判别器D, 经过多次迭代, 当判别器D无法区分输入的无云地面图像来自于真实的无云地面图像还是生成器G生成的无云地面图像时, 训练结束。STGAN的训练过程如图6所示。STGAN模型的损失函数在原始对抗损失的基础上加入了回归损失项L1范数约束, 将真实的无云图像与生成的无云图像作L1范数运算, 使生成的无云图像接近目标图像, 即

$$L_{\text{GAN}}(G, D) = E_{(B'_i, B)}(\log D(B'_i, B)) + E_{B'_i}(\log(1 - D(B'_i, G(B'_i)))) \quad (6)$$

$$L_{\text{L1}}(G) = E_{(B'_i, B)}(\|B - G(B'_i)\|_1) \quad (7)$$

$$L_{\text{STGAN}} = \arg \min_G \max_D L_{\text{GAN}}(G, D) + \lambda L_{\text{L1}}(G) \quad (8)$$

式中, B'_i 表示输入的有云地面图像序列帧, B 表示

真实无云地面图像。 $D(B'_i, B)$ 表示输入为有云地面图像序列帧和真实的无云地面图像的概率, $D(B'_i, G(B'_i))$ 表示输入有云地面图像序列帧和生成器G生成的无云地面图像的概率。STGAN模型的损失 L_{STGAN} 由对抗损失 $L_{\text{GAN}}(G, D)$ 和回归损失 $L_{\text{L1}}(G)$ 两部分组成。 λ 为固定常数, 表示对抗损失和回归损失的比值, 本文中 λ 的取值为100。

3 实验及结果分析

3.1 数据集及数据预处理

由于可见光下的多帧云去除任务没有基准数据集, 且Enomoto等(2017)使用柏林噪声仿真的云进行实验时, 柏林噪声仿真的云图训练出的模型不适用于真实的云图, 因此本文选择了与Zheng等(2021)同样的真实云图和地面数据集。使用来自法国国家信息与自动化研究所(INRIA)的无云卫星图像数据集(Maggiore等, 2017)作为地面图像, 使用来自美国国家海洋和大气管理局(NOAA)的纯云层的灰度红外图像作为云图即使

用真实云图数据集, 这样可有效提升模型的泛化能力, 然后利用合成数据集训练本文的两阶段神经网络, 并测试网络性能。

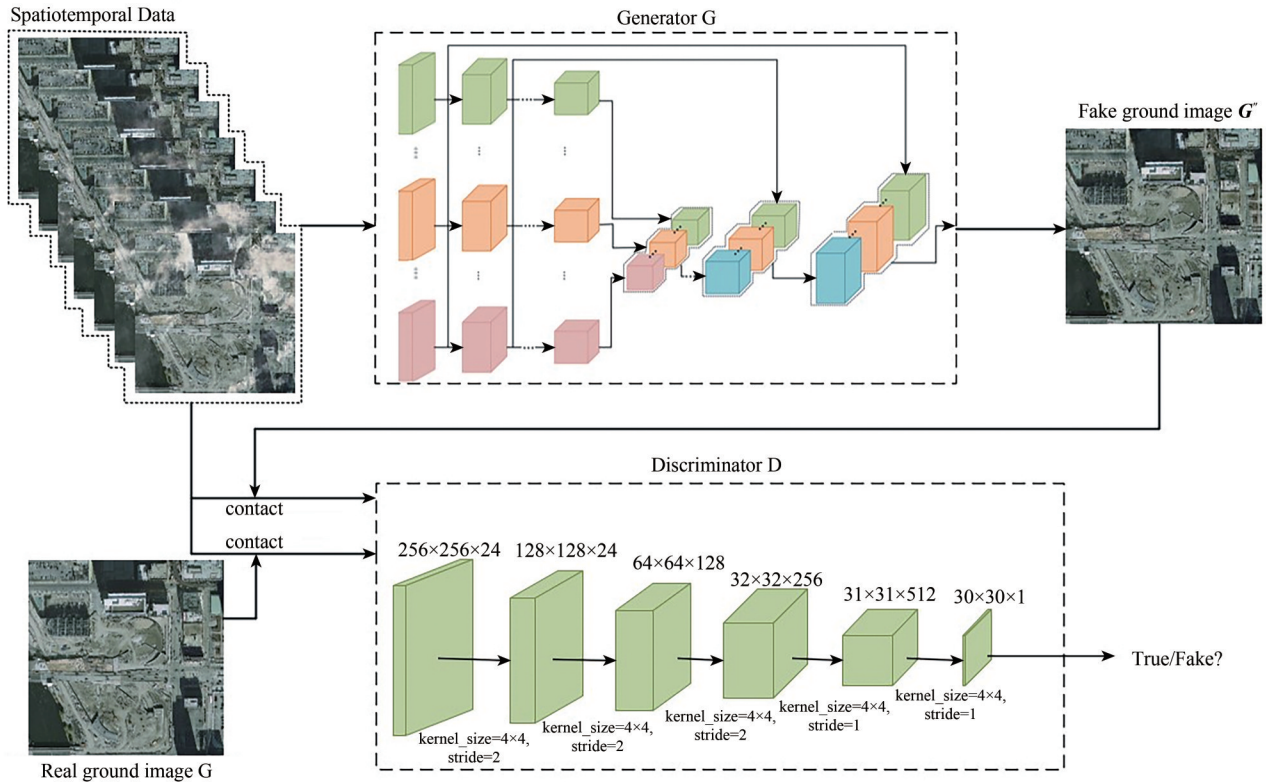


图6 STGAN的训练
Fig. 6 Training of STGAN

为了生成数据集, 取 180 张来自 INRIA 的无云地面图像, 每张图像为 5000×5000 像素的 RGB 图像, 空间分辨率为 0.3 m。首先将其下采样至 1250×1250 像素, 然后随机截取 1024×1024 像素, 将每 1024×1024 像素裁剪为 16 张 256×256 像素的非重叠块, 像素归一化到范围 [0, 1], 这样就得到了 2880 张无云地面图像; 云图像数据取自 NOAA 的 150 张红外图像, 分辨率为 1080×1920 像素, 该卫星每 10 分钟拍摄一张北美大陆的图片。云图像州与州之间有标示行政区的白色边界线, 将这些白线替换为相邻的像素, 达到去除白线的目的, 之后将这些图像预处理为单通道的灰度图像, 在每个图像的底部切割掉显示日期和时间信息的文本, 然后将其下采样到 530×960 像素, 将每幅图像裁剪成 256×256 像素的非重叠块, 像素归一化到范围 [0, 1], 得到 960 张 256×256 像素的云图数据。

3.2 训练过程及实验环境

在第一阶段, 用 2880 张地面图像与 960 张云图数据随机合成 2880 张有云卫星图像作为训练集。

网络训练过程中, 训练批次 $batch_size$ 设置为 32, 训练的总迭代次数为 2^{12} , 使用 Adam 优化器对网络参数进行优化, 设置学习率为 $1E - 4$, 迭代训练至损失不再减小。

在第二阶段, 需要构建时空数据集 $y_{temporal}$, 每一组图像包括 7 帧有云图像和一张无云地面图像, 即 $y_{temporal} = (G'_i, G)$, $i = 1, 2, 3, 4, 5, 6, 7, 8$ 张图像采用同一张地面图像, 每一组 7 帧有云图像通过将同一张地面图像与随机 7 张云图数据合成有云卫星图像, 然后用第一阶段训练好的模型得到去薄云后的 7 帧图像, 最终有 2880 组时空数据集作为第二阶段的数据集, 训练集、验证集与测试集按比例 8 : 1 : 1 随机分配, 得到训练图像组、验证图像组和测试图像组分别为 2304、288、288。训练过程中, 训练批次 $batch_size$ 设置为 1, 总迭代次数为 200 个迭代, 使用 Adam 优化器, 动量 $momentum$ 为 0.5, $beta$ 为 0.99。 $1E - 4$ 的学习率在前 100 个迭代内保持不变, 然后在接下来的 100 个周期内线性衰减为零。

本文实验采用的深度学习框架为 Pytorch，程序语言为 python，实验平台为 NVIDIA GeForce GTX 3060，Intel (R) Core (TM) i7-11800H 处理器，操作系统为 Windows 10。

3.3 评价指标

为了客观、定量的评估光学遥感图像的去云效果，本文采用峰值信噪比 (PSNR) 和结构相似度 (SSIM) (Wang 等, 2004) 两种图像评价指标进行评估。PSNR 基于两幅图像像素点之间的误差衡量两幅图像之间的差异，针对本文中去云后的图像和原始地面图像，评估去云后的图像质量。SSIM 是一种衡量两幅图像相似程度的结构相似性指标，相对 PSNR 而言，结构相似性在评价图像质量上更能符合人类的视觉特性。

PSNR 是基于两幅图像中对应像素之间的平均误差的度量。计算如式 (9) 所示。

$$\text{PSNR}(x,y) = 10 \cdot \log_{10} \left(\frac{\text{MAX}_I^2}{\text{MSE}(x,y)} \right) \quad (9)$$

式中， x 和 y 为两幅图像， $\text{MSE}(x, y)$ 为两幅图像之间的均方误差， MAX_I^2 为图像中任意像素的最大值，最大像素值为 255。PSNR 值的单位是 dB，数值越大表示失真越小，图像质量越好。

SSIM 从亮度 $l(\mathbf{G}, \mathbf{G}')$ 、对比度 $c(\mathbf{G}, \mathbf{G}')$ 和结构 $s(\mathbf{G}, \mathbf{G}')$ 3 个组成部分度量图像之间的结构相似性，反映人眼对图像差异的感知。计算如式 (10) 所示。

$$\text{SSIM}(\mathbf{G}, \mathbf{G}') = l(\mathbf{G}, \mathbf{G}') \cdot c(\mathbf{G}, \mathbf{G}') \cdot s(\mathbf{G}, \mathbf{G}') \quad (10)$$

$$l(\mathbf{G}, \mathbf{G}') = \frac{2\mu_G \mu_{G'} + c_1}{\mu_G^2 + \mu_{G'}^2 + c_1}$$

$$c(\mathbf{G}, \mathbf{G}') = \frac{2\sigma_G \sigma_{G'} + c_2}{\sigma_G^2 + \sigma_{G'}^2 + c_2}$$

$$s(\mathbf{G}, \mathbf{G}') = \frac{2\sigma_{GG'} + c_2}{2\sigma_G^2 \sigma_{G'}^2 + c_2}$$

式中， μ_G 和 σ_G 分别为图像 \mathbf{G} 的均值和标准差， $\sigma_{GG'}$ 为图像 \mathbf{G} 和图像 \mathbf{G}' 的协方差。 c_1 和 c_2 为常数， $c_1 = (0.01L)^2$ ， $c_2 = (0.03L)^2$ ， L 为图像可能的最大像素值。图像 SSIM 取值范围 $[0, 1]$ ，值越大，表示去云后的图像与原始图像的亮度、对比度和结构越相似。

3.4 结果与分析

对于本文提出的两阶段云去除算法，本文主

要设计了两组实验对算法进行验证，第一组实验针对基于 U-Net 和 STGAN 的多时相云去除算法的有效性进行验证，设计了仅利用第一阶段 U-Net 和仅利用第二阶段 STGAN 云去除效果的对比实验；第二组实验针对提出算法的先进性进行验证，设计与当前先进的云去除算法去云效果的对比实验。

3.4.1 两阶段云去除算法有效性定量分析

本小节针对提出的基于 U-Net 和 STGAN 的多时相云去除算法的有效性进行验证，设计了仅利用第一阶段 U-Net 和仅利用第二阶段 STGAN 云去除效果的对比实验，实验结果见表 1。

表 1 两阶段云去除算法有效性定量分析

Table 1 Quantitative analysis of two-stage cloud removal algorithm

方法	PSNR/dB	SSIM
Raw cloudy images	15.92	0.84
Stage I only	25.57	0.91
Stage II only	32.35	0.97
Ours(U-Net+STGAN)	32.64	0.98

针对仅利用第一阶段 U-Net 和仅利用第二阶段 STGAN 分别对 7 帧有云卫星图像作云去除处理，由表 1 可以看出，仅利用第一阶段进行云去除后 PSNR 值为 25.57 dB，SSIM 为 0.91；仅利用第二阶段 STGAN 进行去云处理后，PSNR 值为 30.64 dB，SSIM 为 0.98。仅利用第一阶段或仅利用第二阶段无论是在 PSNR 值还是 SSIM 上都没有两阶段的云去除算法的效果好。两阶段云去除算法较仅利用第一阶段 U-Net 去云 PSNR 值提升了 7.07 dB，较仅利用第二阶段 STGAN 对 7 帧图像去云 PSNR 值提升了 0.29 dB，SSIM 提升了 0.01，证明了本文的两阶段多时相云去除算法的有效性。

3.4.2 算法先进性验证

本节针对提出的基于 U-Net 和 STGAN 的两阶段多时相云去除算法与云去除领域内其他较为先进的云去除算法进行对比，来验证本算法的有效性。选取的对比算法包括 RPCA (Wen 等, 2018)、TRPCA (Wen 等, 2018)、U-Net+GAN (Zheng 等, 2021)、Pix2Pix、STGAN (U-Net) (Sarukkai 等, 2020) 云去除算法。RPCA 利用低秩的地面和云的

稀疏对有云图像做分离达到去云效果, TRPCA在RPCA的基础上对初始检测到的云像素和无云区域分配相应的权重, 再次用RPCA做分离达到云去除效果。U-Net+GAN利用U-Net和GAN分别进行薄云和厚云处理。Pix2Pix模型在图像到图像的配对转换任务中取得了最先进的结果。STGAN(U-Net)用分支网络U-Net做生成器, 对3帧有云卫星图像进行去云。

表2为本文算法与其他对比算法的去云效果定量分析, 由表2可以看出, 本文算法的PSNR值和SSIM均高于其他对比算法。本文算法的PSNR值为32.64 dB, SSIM为0.98, RPCA和TRPCA在图像序列帧为7帧时, 去云效果较差, 本文算法PSNR值远超于RPCA的19.16 dB和TRPCA的18.93 dB, SSIM值超过了RPCA的0.85和TRPCA的0.87。本文算法较U-Net+GAN单时相云去除算法PSNR值提升了7.33 dB, SSIM提升了0.08, 较先进的深度学习方法Pix2Pix单时相云去除算法PSNR值提升了6.25 dB, SSIM提升了0.08。采用STGAN(U-Net)对7帧有云图像的任意3帧做去云处理, 该分支网络的去云效果相对较好, STGAN(U-Net)的PSNR值达到了30.77 dB。本文算法相较于STGAN(U-Net)算法PSNR值提升了1.87 dB, SSIM提升了0.01。经过本文算法与RPCA、TRPCA、U-Net+GAN、Pix2Pix、STGAN(U-Net)等云去除算法进行比较, 充分说明本文算法在去云效果上有明显提升, 验证了本文的云去除算法的先进性。

表2 不同云去除算法去云效果定量分析

Table 2 Quantitative analysis of cloud removal effect of different cloud removal algorithms

方法	PSNR/dB	SSIM
Raw cloudy images	15.92	0.84
RPCA	19.16	0.85
TRPCA	18.93	0.87
U-Net+GAN	25.31	0.90
Pix2Pix	26.39	0.90
STGAN(U-Net)	30.77	0.97
本文方法	32.64	0.98

如图7展示了本文算法与其他去云对比算法的去云视觉效果对比, 共选取了4组7帧的有云图像序列, 每一组的第1行为有云的图像序列 $S_i(i=1, 2, \dots, 7)$, 第2行依次为本文算法、STGAN(U-Net)、Pix2Pix、U-Net+GAN、TRPCA、RPCA算法的去云效果图。STGAN(U-Net)选取了7帧图像序列中的任意3帧进行云去除, Pix2Pix和U-Net+GAN算法分别对图像序列中的7帧进行云去除, 图7显示的为7帧图像中去云效果最好的一帧, TRPCA和RPCA对7帧多时相图像进行去云, 图7选取7帧图像中去云效果最好的一帧进行展示。由图7可以看出, STGAN(U-Net)和Pix2Pix算法在一定程度上能将云层遮挡下的地物显示出来; STGAN(U-Net)算法去云后, 仍有少量的云层残留; Pix2Pix算法去云, 能够生成图像中有遮挡的区域, 并保持这些区域的完整, 同时云遮挡区域的地面进行合理的推断; U-Net+GAN算法去云后仍有少量的云层残留, 并且厚云覆盖区域由于生成器倾向于以简单模式填充缺失区域, 比如城镇区域可能会用森林或植被填充导致生成了不真实的地面信息; 传统算法TRPCA和RPCA算法在对图中的7帧有云图像序列去云时, 由于图像序列较短效果最差, 图像中仍有云层残留, 并且TRPCA由于使用了形态学操作出现了失真现象, TRPCA去云图像出现了一些黑色的矩形块。由此可以看出本文算法的去云效果最好, 去云后的无云地面图像没有云层残留, 图像细节较为丰富, 并且图像中的地物信息较为清晰, 与真实无云地面图像的视觉效果最为接近。当地面图像有大面积区域被厚云遮挡时, 本文算法与其他云去除对比算法的效果都不太理想。这是因为多时相遥感图像云去除往往是根据来自同一传感器的同一地点的多时相有云图像序列进行云去除, 通过同一时间对应的无云区域来估计厚云遮挡区域的缺失信息, 当多时相图像被大面积厚云遮挡时, 无法对厚云缺失区域进行估计, 这种情况在现实生活中很少出现, 在同样的数据集下, Pix2Pix的效果略好一些。为了进一步验证本文算法的适用性, 图8分别展示了不同数据集以及模拟了光谱变化迅

速的地表的实验结果,第1组采用柏林噪声仿真的云图合成的有云图像序列进行云去除,第2组模拟

光谱变化的地表并进行云去除,本文模型依然适用。

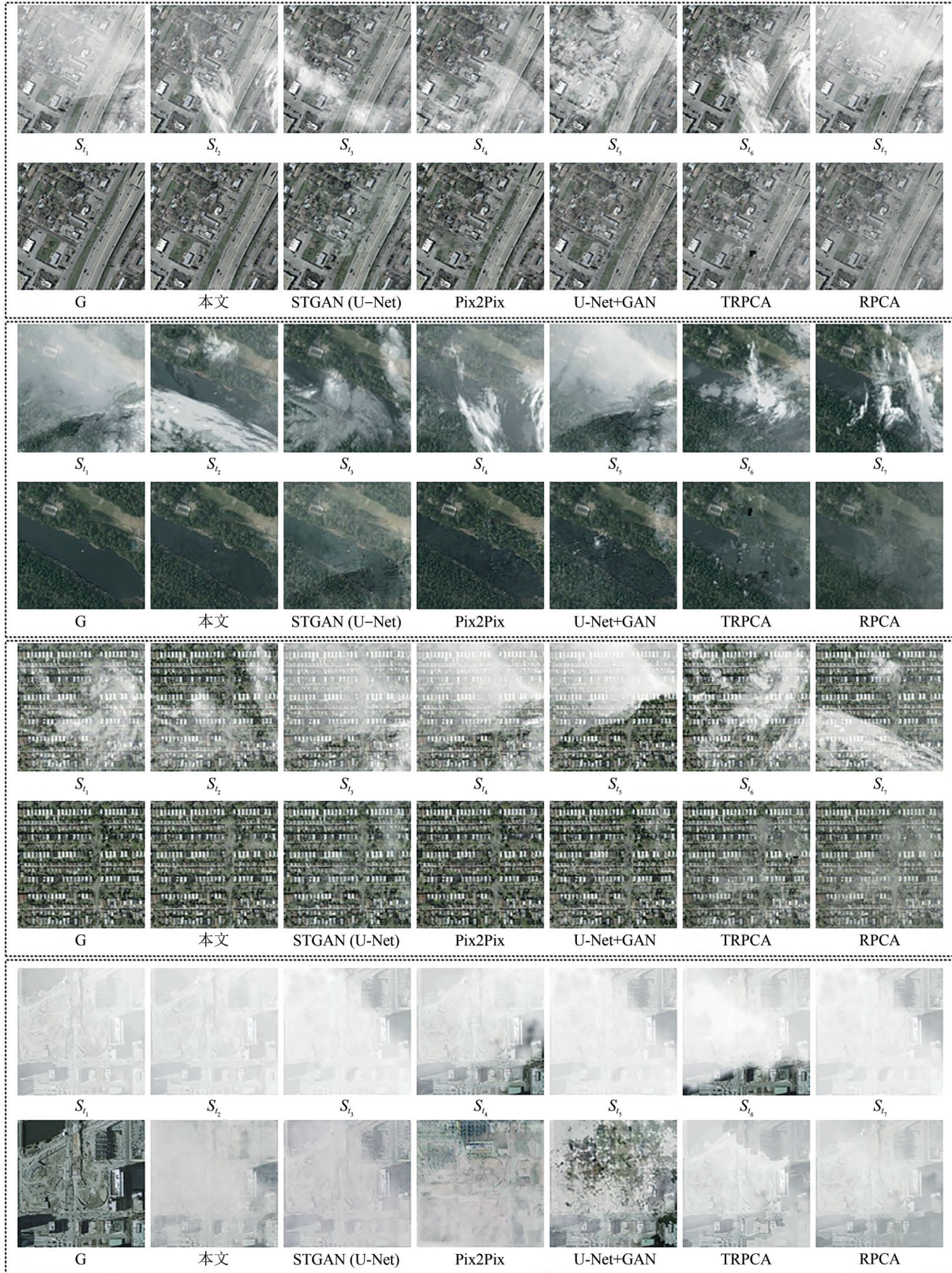


图7 不同云去除算法去云效果图对比

Fig. 7 Comparison of cloud removal effect diagrams of different cloud removal algorithms

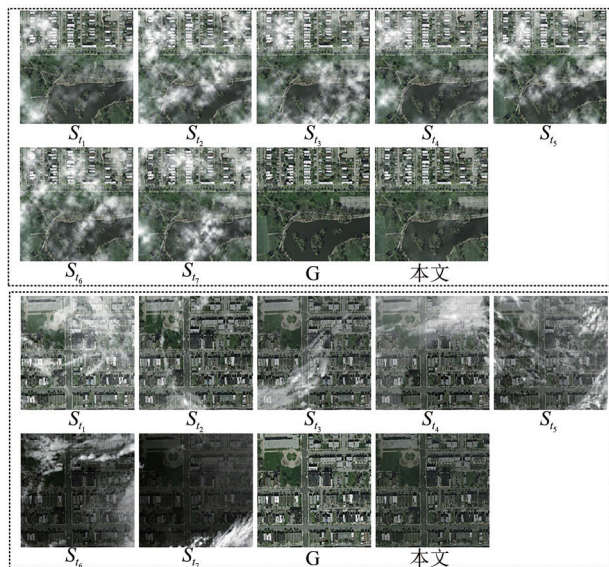


图8 其他情况去云效果图

Fig. 8 Other cases cloud removal effect diagrams

4 结论

本文针对光学遥感图像中常有薄云和厚云遮挡地面的情况,提出了一种基于U-Net和STGAN的多时相遥感图像云去除算法,通过第一阶段U-Net达到了云检测和薄云去除的目的,第一阶段的薄云处理使得第二阶段的STGAN捕捉到了更多的地面细节信息,通过第二阶段STGAN利用同一位置无云时间段的影像来恢复有云遮挡的影像来进行云去除。通过与其他去云算法的对比实验,可以看出本文提出的算法在去云后的地面图像的视觉效果上更好,在PSNR值与SSIM客观评价指标上优于其他对比算法。虽然本文算法取得了相对较好的去云效果,但是也有一定的局限性,当面临厚云大面积覆盖的有云图像序列时,该算法的去云效果不太理想,接下来将探索利用图像序列帧的时空特征更好的重建大面积厚云覆盖区域。

参考文献(References)

Enomoto K, Sakurada K, Wang W M, Fukui H, Matsuoka M, Nakamura R and Kawaguchi N. 2017. Filmy cloud removal on satellite imagery with multispectral conditional generative adversarial nets//Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops. Honolulu: IEEE: 1533-1541 [DOI: 10.1109/cvprw.2017.197]

Isola P, Zhu J Y, Zhou T H and Efros A A. 2017. Image-to-image translation with conditional adversarial networks//Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recogni-

tion. Honolulu: IEEE: 5967-5976 [DOI: 10.1109/cvpr.2017.632]

King M D, Platnick S, Menzel W P, Ackerman S A and Hubanks P A. 2013. Spatial and temporal distribution of clouds observed by MODIS onboard the Terra and Aqua satellites. *IEEE Transactions on Geoscience and Remote Sensing*, 51(7): 3826-3852 [DOI: 10.1109/TGRS.2012.2227333]

Kussul N, Lavreniuk M, Skakun S and Shelestov A. 2017. Deep learning classification of land cover and crop types using remote sensing data. *IEEE Geoscience and Remote Sensing Letters*, 14(5): 778-782 [DOI: 10.1109/LGRS.2017.2681128]

Kussul N, Skakun S, Shelestov A, Lavreniuk M, Yailymov B and Kussul O. 2015. Regional scale crop mapping using multi-temporal satellite imagery. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XL-7-W3: 45-52 [DOI: 10.5194/isprsarchives-XL-7-W3-45-2015]

Li X H, Wang L Y, Cheng Q, Wu P H, Gan W X and Fang L N. 2019. Cloud removal in remote sensing images using nonnegative matrix factorization and error correction. *ISPRS Journal of Photogrammetry and Remote Sensing*, 148: 103-113 [DOI: 10.1016/j.isprsjprs.2018.12.013]

Li Z W, Shen H F, Li H F, Xia G S, Gamba P and Zhang L P. 2017. Multi-feature combined cloud and cloud shadow detection in GaoFen-1 wide field of view imagery. *Remote Sensing of Environment*, 191: 342-358 [DOI: 10.1016/j.rse.2017.01.026]

Maggiore E, Tarabalka Y, Charpiat G and Alliez P. 2017. Can semantic labeling methods generalize to any city? The inria aerial image labeling benchmark//2017 IEEE International Geoscience and Remote Sensing Symposium. Fort Worth: IEEE: 3226-3229 [DOI: 10.1109/IGARSS.2017.8127684]

Mirza M and Osindero S. 2014. Conditional generative adversarial nets. arXiv:1411.1784

Narasimhan S G and Nayar S K. 2002. Vision and the atmosphere. *International Journal of Computer Vision*, 48(3): 233-254 [DOI: 10.1023/A:1016328200723]

Peng J, Chen S, Lü H L, Liu Y X and Wu J S. 2016. Spatiotemporal patterns of remotely sensed PM_{2.5} concentration in China from 1999 to 2011. *Remote Sensing of Environment*, 174: 109-121 [DOI: 10.1016/j.rse.2015.12.008]

Perlin K. 2002. Improving noise. *ACM Transactions on Graphics*, 21(3): 681-682 [DOI: 10.1145/566654.566636]

Ronneberger O, Fischer P and Brox T. 2015. U-net: convolutional networks for biomedical image segmentation//18th International Conference on Medical Image Computing and Computer-Assisted Intervention. Munich: Springer: 234-241 [DOI: 10.1007/978-3-319-24574-4_28]

Sarukkai V, Jain A, Uzgent B and Ermon S. 2020. Cloud removal in satellite images using spatiotemporal generative networks//Proceedings of the 2020 IEEE Winter Conference on Applications of Computer Vision. Snowmass: IEEE: 1785-1794 [DOI: 10.1109/WACV45572.2020.9093564]

Saunders R W and Kriebel K T. 1988. An improved method for detecting clear sky and cloudy radiances from AVHRR data. *International Journal of Remote Sensing*, 9(1): 123-150 [DOI: 10.1080/01431168808954841]

- Shen H F, Li X H, Cheng Q, Zeng C, Yang G, Li H F and Zhang L P. 2015. Missing information reconstruction of remote sensing data: a technical review. *IEEE Geoscience and Remote Sensing Magazine*, 3(3): 61-85 [DOI: 10.1109/MGRS.2015.2441912]
- Singh P and Komodakis N. 2018. Cloud-Gan: cloud removal for Sentinel-2 imagery using a cyclic consistent generative adversarial networks//2018 IEEE International Geoscience and Remote Sensing Symposium. Valencia: IEEE: 1772-1775 [DOI: 10.1109/IGARSS.2018.8519033]
- Wang Z, Bovik A C, Sheikh H R and Simoncelli E P. 2004. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4): 600-612 [DOI: 10.1109/TIP.2003.819861]
- Wen F, Zhang Y J, Gao Z and Ling X. 2018. Two-pass robust component analysis for cloud removal in satellite image sequence. *IEEE Geoscience and Remote Sensing Letters*, 15(7): 1090-1094 [DOI: 10.1109/LGRS.2018.2829028]
- Zhang Q, Yuan Q Q, Zeng C, Li X H and Wei Y C. 2018. Missing data reconstruction in remote sensing image with a unified spatial-temporal-spectral deep convolutional neural network. *IEEE Transactions on Geoscience and Remote Sensing*, 56(8): 4274-4288 [DOI: 10.1109/TGRS.2018.2810208]
- Zhao X X. 2019. Research on High Resolution Remote Sensing Image Dehazing Method Based on Deep Learning. Hohhot: Inner Mongolia University of Technology (赵肖肖. 2019. 基于深度学习的高分辨率遥感图像去雾方法研究. 呼和浩特: 内蒙古工业大学)
- Zheng J H, Liu X Y and Wang X D. 2021. Single image cloud removal using U-net and generative adversarial networks. *IEEE Transactions on Geoscience and Remote Sensing*, 59(8): 6371-6385 [DOI: 10.1109/TGRS.2020.3027819]

Cloud removal in multitemporal remote sensing imagery combining U-Net and spatiotemporal generative networks

WANG Zhuo^{1,2}, MA Jun¹, GUO Yi³, ZHOU Chuanjie⁴, BAI Bin⁴, LI Feng²

1.College of Software, Henan University, Kaifeng 475100, China;

2.Qian Xuesen Space Technology Laboratory, China Academy of Space Technology, Beijing 100094, China;

3.Western Sydney University, Penrith South, NSW 2751, Australia;

4.Beijing Institute of Remote Sensing Information, Beijing 100192, China

Abstract: Cloud occlusion often occurs in optical remote sensing images. Cloud occlusion may reduce or even completely occlude some ground cover information in the images, which limits ground observation, change detection, or land cover classification. Cloud removal is an important task that urgently needs to be solved. Thin and thick clouds usually coexist in optical remote sensing images, and the cloud removal algorithm for single-frame remote sensing images is only suitable for solving the problem of thin cloud occlusion. Therefore, using multi-temporal remote sensing images of the same area at different times to remove clouds has become a major issue. This study aims to fully utilize images in the same location without cloud time period to replace cloud-occluded images for restoring the ground area occluded by clouds. For this purpose, a two-stage cloud removal algorithm for multi-temporal remote sensing images based on U-Net and spatiotemporal generative network (STGAN) is proposed. The first stage is cloud segmentation, which directly uses the U-Net model to extract clouds and remove thin clouds. The second stage is image restoration, which directly uses STGAN to remove thick clouds. It inputs the seven frames of ground images after removing thin clouds into the STGAN model to obtain a single, detail-rich cloud-free ground image. The generative model of STGAN adopts an improved multi-input U-Net to recover the corresponding irregularities in the thick cloud cover area by extracting key features from seven frames of images at the same location at a time. The thin cloud processing in the first stage is beneficial to the subsequent STGAN to capture more ground information. The proposed algorithm can solve the inability of U-Net to handle cloud occlusion in thick cloud areas. It can also capture more ground information than directly using STGAN for cloud removal. It has a better cloud removal effect. The experimental results on our dataset show that only using the first-stage U-Net model and only using the second-stage STGAN model for cloud removal are inferior to the proposed two-stage cloud removal algorithm in terms of subjective visual effects and objective quantitative evaluation indicators such as peak signal-to-noise ratio and structural similarity. This performance fully verifies the effectiveness of the cloud removal algorithm in this study. Compared with traditional cloud removal methods such as RPCA, TRPCA and deep learning algorithms such as Pix2Pix, the proposed algorithm is superior to the comparison algorithm and has a significant improvement, which fully verifies the advancement of the cloud removal algorithm in this study. The proposed algorithm fully utilizes the spatiotemporal information of multi-temporal cloudy satellite images of the same area at different times. It also has good cloud removal performance, which is conducive to the further utilization of optical remote sensing images. Although the proposed algorithm has achieved a relatively good cloud removal effect, it also has certain limitations. The cloud removal effect of the algorithm is not ideal for cloud image sequences with a large area covered by thick clouds. In the follow-up research, the spatiotemporal features of image sequence frames will be explored to better reconstruct large areas covered by thick clouds.

Key words: remote sensing images, multi-temporal, cloud removal, image restoration, U-Net, STGAN

Supported by National Key Research and Development Program of China (No. 2020YFA04100)