

引用格式: Li Jianfeng, Liu Siqi, Li Jinbin, *et al.* Research on Cloud Detection Model of Sentinel-2 Image Coupled with SVM and Cloud-Score Algorithm[J]. Remote Sensing Technology and Application, 2022, 37(3): 713-720. [李健锋, 刘思琪, 李劲彬, 等. 耦合 SVM 和 Cloud-Score 算法的 Sentinel-2 影像云检测模型研究[J]. 遥感技术与应用, 2022, 37(3): 713-720.]
DOI: 10.11873/j.issn.1004-0323.2022.3.0713

耦合 SVM 和 Cloud-Score 算法的 Sentinel-2 影像云检测模型研究

李健锋^{1,3,4,5}, 刘思琪^{1,3,4,5}, 李劲彬^{1,3,4,5}, 彭 颢^{1,3,4,5}, 叶虎平²

- (1. 陕西地建土地工程技术研究院有限责任公司, 陕西 西安 710021;
2. 中国科学院地理科学与资源研究所 资源与环境信息系统国家重点实验室, 北京 100101;
3. 陕西省土地工程建设集团有限责任公司, 陕西 西安 710075;
4. 自然资源部退化及未利用土地整治工程重点实验室, 陕西 西安 710075;
5. 陕西省土地整治工程技术研究中心, 陕西 西安 710075)

摘要: 云覆盖阻碍了光学遥感卫星对地观测的有效范围, 快速、准确的云检测是遥感应用产品生成过程中的重要一步。针对 Google Earth Engine 云平台中缺乏适用且高质量的云检测模型, 以热带多云的斯里兰卡为研究区, 构建了耦合 SVM 和 Cloud-Score 算法的 Sentinel-2 影像云检测模型, 通过实验从目视判读与定量分析两个角度对比了其 QA60 法、Cloud-Score 算法以及 Fmask 的云检测精度, 并在海南岛和亚马逊森林两个地区进行了云检测测试。研究结果表明: Fmask 模型的云检测性能最低, 总体精度仅为 63.45%, 存在严重的水体误分为云的现象, 但其漏提率极低; QA60 法对卷云识别不足, 漏提率较高, 同时存在一定的误分现象, 并且低空间分辨率影响了云体边界提取结果的细节性; Cloud-Score 算法的云检测性能明显好于 QA60 法, 总体精度达到了 89.83%, 误提率仅为 2.17%, 但仍存在部分卷云漏提的现象; 相比于其他 3 种云检测方法, 本文提出的云检测模型总体精度最高, 达到了 98.21%, 并且拥有极低的漏提率和误提率, 能比较精准地识别出云体的边界, 可满足 Sentinel-2 遥感产品的云检测预处理需求。

关键词: 云检测; SVM; Cloud-Score 算法; Sentinel-2; Fmask

中图分类号: P407; TP751 **文献标志码:** A **文章编号:** 1004-0323(2022)03-0713-08

1 引言

光学卫星影像中的云覆盖是影响其进行全球范围应用的关键因素, 快速准确的云检测是遥感产品生成过程中的关键一步^[1-3]。近年来, 随着云计算、大数据技术的发展, Google Earth Engine^[3]、PIE-Engine^[4]等遥感云平台逐渐发布, 遥感应用研

究已突破了原先的“影像下载—预处理—模型构建—结果分析”模式, 借助云平台强大的计算和存储能力, 通过在线调用 API 接口对影像进行批量分析计算, 极大地提高了成果转化的速度与效率, 加速了遥感卫星的全球化应用^[4-6]。

针对 Sentinel-2 影像, Google Earth Engine 提供

收稿日期: 2021-04-17; 修订日期: 2022-02-18

基金项目: 国家重点研发项目(2019YFE0126500), 高分辨率对地观测系统国家重大专项(21-Y20B01-9001-19/22), 陕西地建—西安交大土地工程与人居环境技术创新中心开放基金资助项目(2021WHZ0090), 陕西省土地工程建设集团内部科研项目(DJNY 2022-29、DJTD2022-4), 中央高校基本科研业务费资助项目(300102352502)。

作者简介: 李健锋(1994—), 男, 浙江义乌人, 硕士, 工程师, 主要从事定量遥感研究。E-mail: ljf_sxdj@126.com

通讯作者: 叶虎平(1983—), 男, 陕西富平人, 博士, 助理研究员, 主要从事无人机航路规划、水色遥感与水体光学模型研究。
E-mail: yehp@igsrr.ac.cn

了3种云检测方法:QA60波段法^[7]、Cloud-Score算法^[8]以及S2cloudless^[9]。QA60波段由蓝色波段和两个短波红外波段生成,不同于其他地物,云和雪在蓝色波段反射率极高,而在短波红外波段云的反射率明显高于雪,但QA60波段法整体的云检测效果不稳定,通常会导致低估或高估的现象,并且其分辨率为60 m,识别出的云体边界细节性不足^[10-11]。Cloud-Score算法主要原理是利用亮度、温度和归一化雪盖指数(Normalized Difference Snow Index, NDSI)组合计算出云的可能性分数,单个像素的分数越高,属于云的可能性就越大,最终通过设置阈值进行云掩膜处理。复杂的大气条件和Sentinel-2影像中缺少热红外波段等因素会导致Cloud-Score算法低估云的范围^[11-12]。与目前最先进的基于阈值的云检测算法相比,机器学习算法具有更好的云检测性能,如Fmask、Sen2Cor以及MAJA等^[15-18]。S2cloudless^[9]是Sentinel-Hub提供的专门针对Sentinel-2影像的一种基于机器学习的云检测模型,主要基于LightGBM算法^[17],通过世界各地的多个云和非云样本上进行训练,最终通过分类获得云的范围。Google Earth Engine提供了基于S2cloudless模型计算的Sentinel-2影像的云概率图像数据集(“COPERNICUS/S2_CLOUD_PROBABILITY”),但该数据仍在生产中,尚未覆盖所有影像,并且在亚马逊热带雨林地区的云检测测试中,S2cloudless的效果并不理想,仅为52%^[18]。因此,迫切需要一种更加精确且适用于Google Earth Engine云平台上Sentinel-2影像的云检测模型。

斯里兰卡是“21世纪海上丝绸之路”的重要参与国,作为连接亚—非—欧海上航路的枢纽,具有极重要的经济及地理意义^[19]。斯里兰卡属于热带地区,大气环境复杂,近中午前后常会形成各种云覆盖,致使对地观测卫星过境影像常年受云雾覆盖影响。通过Google Earth Engine统计,覆盖斯里兰卡的无云光学影像(Landsat-8、Sentinel-2)占比不到1%。以斯里兰卡为研究区,基于Google Earth Engine云平台,提出了一种耦合SVM和Cloud-Score算法的Sentinel-2影像云检测模型,以期实现准确、可靠的云范围识别。

2 数据与研究区

2.1 研究区概况

斯里兰卡位于南亚次大陆的南端,印度的东南

方向,整个岛屿在印度洋之中,东西宽224 km,南北长432 km,国土面积达65 610 km²^[20]。斯里兰卡地处热带地区,属于热带季风性气候,无四季之分,年平均气温达28℃,终年如夏。除中南部山区外,一年中的气候变化很小,气候变化主要与降雨有关。本文选取了4个站点作为Sentinel-2影像云检测测试点,具体位置如图1所示。

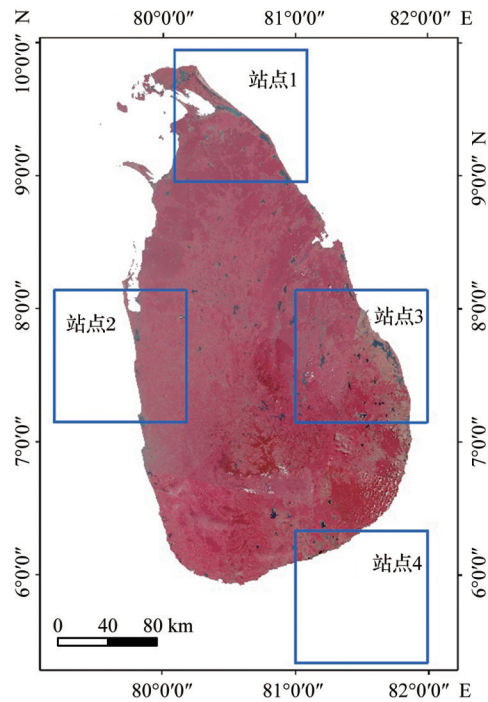


图1 研究区位置

Fig.1 Location of study area

2.2 数据来源及处理

Sentinel-2属于高分辨率多光谱成像卫星,主要由2A和2B两颗卫星组成^[21]。Sentinel-2A发射于2015年6月23日,Sentinel-2B发射于2017年3月7日,双星的重访周期大约为5 d。Sentinel-2携带有多光谱仪,其13个光谱带跨越可见光、近红外和短波红外,幅宽为290 km,高度为756 km。研究采用的Sentinel-2数据来自Google Earth Engine云平台所提供的“COPERNICUS/S2_SR”数据集。针对每个站点,选取两景影像进行实验,影像的具体信息如表1所示。

3 研究方法

3.1 Google Earth Engine

Google Earth Engine是Google提供的对大量全球尺度地球科学资料数据进行在线可视化计算和分析处理的云平台^[5]。该平台主要存储了卫星影

表1 影像具体信息

Table 1 Specific information of images

站点	条带号	日期	传感器	含云量/%
站点1	T44PMR	2019/07/25	Sentinel-2B	1.66
		2020/12/11	Sentinel-2A	17.51
站点2	T44NLP	2019/04/06	Sentinel-2B	6.59
		2020/02/15	Sentinel-2A	12.22
站点3	T44NNP	2019/04/03	Sentinel-2B	8.88
		2020/04/07	Sentinel-2B	8.10
站点4	T44NNM	2019/04/08	Sentinel-2A	6.37
		2020/11/08	Sentinel-2A	25.60

像和其他地球观测数据数据库中的资料,同时提供了足够的运算能力对存储数据进行调用和处理。相比于ENVI、Erdas等传统的影像处理软件,Google Earth Engine云平台可以快速、批量地处理影像,并且不需要下载影像数据,同时提供多种格式数据的导出。Google Earth Engine提供了在线的JavaScript API以及离线的Python API,用户通过调用API可快速搭建基于Google云的Web端服务。

3.2 Cloud-Score算法

Cloud-Score算法主要利用云的光谱和热属性来识别并从遥感影像中进行剔除^[10]。为了简化该过程,Cloud-Score函数使用最小/最大归一化函数在0和1之间重新调整预期反射率(公式(1))。该算法主要原理是将归一化函数应用于计算4种云的光谱特性,最终获得的Cloud-Score值是这些归一化值中的最小值(公式(2)),以此寻求明亮和寒冷的像素,但同时不具有雪的光谱特性。具体公式为:

$$Norm(r, r_{min}, r_{max}) = \frac{r - r_{min}}{r_{max} - r_{min}} \quad (1)$$

CloudScore =

$$\min \left\{ 1.0, Norm(\rho_{Blue}, 0.1, 0.3), Norm(\rho_{Blue} + \rho_{Green} + \rho_{Red}, 0.2, 0.8), Norm(\rho_{NIR} + \rho_{SWIR1} + \rho_{SWIR2}, 0.3, 0.8), Norm\left(\frac{\rho_{Green} - \rho_{SWIR1}}{\rho_{Green} + \rho_{SWIR2}}, 0.8, 0.6\right) \right\} \quad (2)$$

其中:Norm表示归一化函数; r 为反射率; r_{max} 、 r_{min} 分别为反射率的最大、最小值; ρ_{Red} 、 ρ_{Green} 、 ρ_{Blue} 、 ρ_{NIR} 、 ρ_{SWIR1} 、 ρ_{SWIR2} 分别为遥感影像红、绿、蓝、近红外、短波红外1以及短波红外2波段的反射率。

3.3 支持向量机(SVM)

SVM算法是一种建立在统计学习理论的VC

维理论和结构风险最小化原理基础上的分类算法^[24]。相比于神经网络或传统的基于统计的分类方法,SVM通过向量的个数来控制模型的复杂度,不需要通过降维处理削减特征变量来控制模型的复杂度,因此在分类的过程中,SVM分类器不会损失地物目标的特征信息,减少了一些过拟合现象的发生。SVM在精度、泛化性以及高维数据处理等方面均具有明显的优势,在遥感影像分类应用中获得了广泛应用^[25, 26]。

SVM的基本原理是将原始的特征向量变换到高维特征空间,在高维空间中求解最优分类超平面,分类精度主要取决于超平面与超平面两边的边界平面之间的距离,距离越大,分类器的精度就越高,误差就越小。给定一个样本集 (x_i, y_i) , i 为1, 2, 3... N , $x_i \in \mathbb{R}^n$, $y_i \in \{-1, 1\}$,分类线 $H = \omega \cdot x + b = 0$,最优分类线需要有最大的分类间隔,即两类样本之间的距离 $\frac{2}{\|\omega\|}$ 最大,等同于求函数的最小值问题。

$$\phi(\omega) = \frac{1}{2} \|\omega\|^2 \quad (3)$$

本研究中云体的识别属于线性不可分问题,为此需要在上式中引入松弛变量 ξ_i 以及惩罚参数 c ,则SVM可以表示为:

$$\min \left[\frac{1}{2} \|\omega\|^2 + c \sum_{i=1}^l (\xi_i + \xi_i^*) \right] \quad (4)$$

$$\text{s.t.} \begin{cases} ((\omega \cdot x_i) + b - y_i) \leq \varepsilon + \xi_i \\ y_i - ((\omega \cdot x_i) + b) \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases}$$

对于优化问题,需要变换为拉格朗日算子的对偶问题,最终SVM的表达式为:

$$f(x) = \sum_{i=1}^n (\alpha_i - \alpha_i^*) k(x_i, x) + b \quad (5)$$

其中: x_i 为训练样本的向量; b 为线性函数阈值。

3.4 耦合SVM和Cloud-Score算法的云检测模型

为了更加精确地对Sentinel-2影像进行云检测,本文基于Google Earth Engine云平台,借助其强大的分析与处理能力,通过耦合SVM和Cloud-Score算法,构建云检测模型。模型主要原理是在Cloud-Score算法的云检测结果基础上,利用分层随机抽样获得SVM的训练样本,最终通过训练和分类进行云检测。QA60波段法和Cloud-Score算法具有阈值法高效率的特点,而Cloud-Score算法的云检测结果分辨率和精度更高,因此更适合作为样

本的选取图层。图2为耦合SVM和Cloud-Score算法的云检测模型实现过程。首先以时间、含云量、位置为条件,通过编码筛选(ee.ImageCollection.filter())所需云检测的Sentinel-2影像(“COPERNICUS/S2_SR”);其次对筛选后的影像进行重投影(ee.Image.reproject())、裁剪(ee.Image.clip())等预处理操作;然后利用Cloud-Score算法识别云体范围,并设置SVM分类器(ee.Classifier.svm())的相关参数,核函数选择径向基核函数(Radial basis function, RBF),gamma、cost参数默认设置为0.5和10,具体可参考Hus等^[27]提出的SVM参数设置指南;接着利用分层抽样函数(ee.Image.stratifiedSample())进行云和非云样本选取,采样比例设置为10 m,样本数量默认设置为50个;最终通过SVM训练(svm.train())和分类(svm.classify())确定云体范围,并利用 3×3 栅格窗口进行中值滤波(ee.Image.focal_median())去除噪声后得到云检测结果。

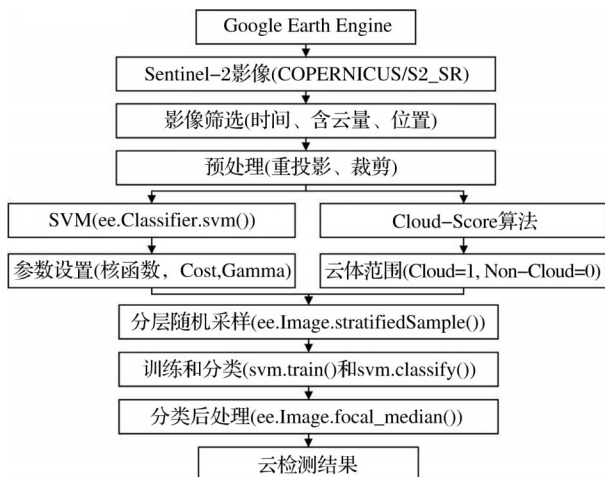


图2 耦合SVM和Cloud-Score算法的云检测模型

Fig.2 Cloud Detection Model coupled with SVM and Cloud-Score algorithm

4 结果与分析

针对斯里兰卡4个站点所选择的8景影像,分别利用QA60波段法、Cloud-Score算法、Fmask以及本文提出的模型进行云检测,每个站点首幅影像的云检测结果如图3所示。通过详尽的视觉分析可以发现,4种方法的云检测结果存在显著差异,总体性能呈现出 $Fmask < QA60 < Cloud-Score < 本文模型$ 。QA60波段法表现得不稳定,4个站点的结果均出现了不同程度的漏提现象,尤其是对卷云识别效果较差;同时还存在误提现象,站点1中将海上部分清晰像素误识别为云;并且其分辨率为60 m,而

Sentinel-2影像的最高分辨率为10 m,因此云边界的识别效果较差。相比于QA60波段法,由于更多的波段参与云可能性分数的计算,Cloud-Score算法在云体边界提取的完整性方面有了明显的提升,并且对密云的认识效果较好,但仍存在部分卷云的漏提现象,尤其是站点4南部海域存在较多的卷云未识别。Sanchez等^[22,28]在不同地区通过实验均证明了Fmask模型的云检测精度在90%以上,然而其在斯里兰卡4个站点的云检测测试中很不理想。Fmask模型的云检测结果呈现出不同程度的误提现象,站点1沿海地区将部分水体识别为云,站点2和站点4的海域中存在大量的误提现象。虽然Fmask模型对密云和卷云均表现出了较好的识别效果,但严重的误提现象影响了云检测的整体精度。相比于其他3种模型的云检测结果,本文提出的云检测方法性能最好,能完整地识别出云体边界,对卷云的识别呈现出了良好的效果,视觉上仅在站点2北部海域出现了部分卷云的漏提现象。

以每景影像的目视云检测结果为基准,分别选取200个云体样本和非云体样本,通过计算混淆矩阵^[29]分别从漏提率、误提率和总体精度3个方面进一步反映不同模型云检测精度,其统计结果如表2所示。从表中可以看出,4种模型中Fmask模型的总体精度最低,仅为63.45%,误提率极高,达到了38.74%,但其漏提率极低;QA60法的漏提率最高,为25.35%,同时存在较高的误提率,总体精度为78.16%;Cloud-Score算法的云检测性能明显好于QA60法,总体精度达到了89.83%,误提率仅为2.17%,但仍存在较高的漏提率;研究提出的耦合SVM和Cloud-Score算法的云检测模型总体精度最高,达到了98.21%,并且拥有极低的漏提率和误提率,整体云检测性能较好,可满足Sentinel-2遥感产品的云检测预处理需求。

5 讨论

本文构建的耦合SVM和Cloud-Score算法的云检测模型在斯里兰卡地区表现出了较高的精度,为了进一步验证其不同地区在不同云量条件的适用性,分别在海南岛和亚马逊热带森林按照含云量0~10%、10%~20%、20%~30%、30%~40%和40%~50%分别选取一景Sentinel-2影像进行云检测,具体信息如表3所示。通过每景影像的云检测结果与原始影像叠加对比分析发现,本文构建的模

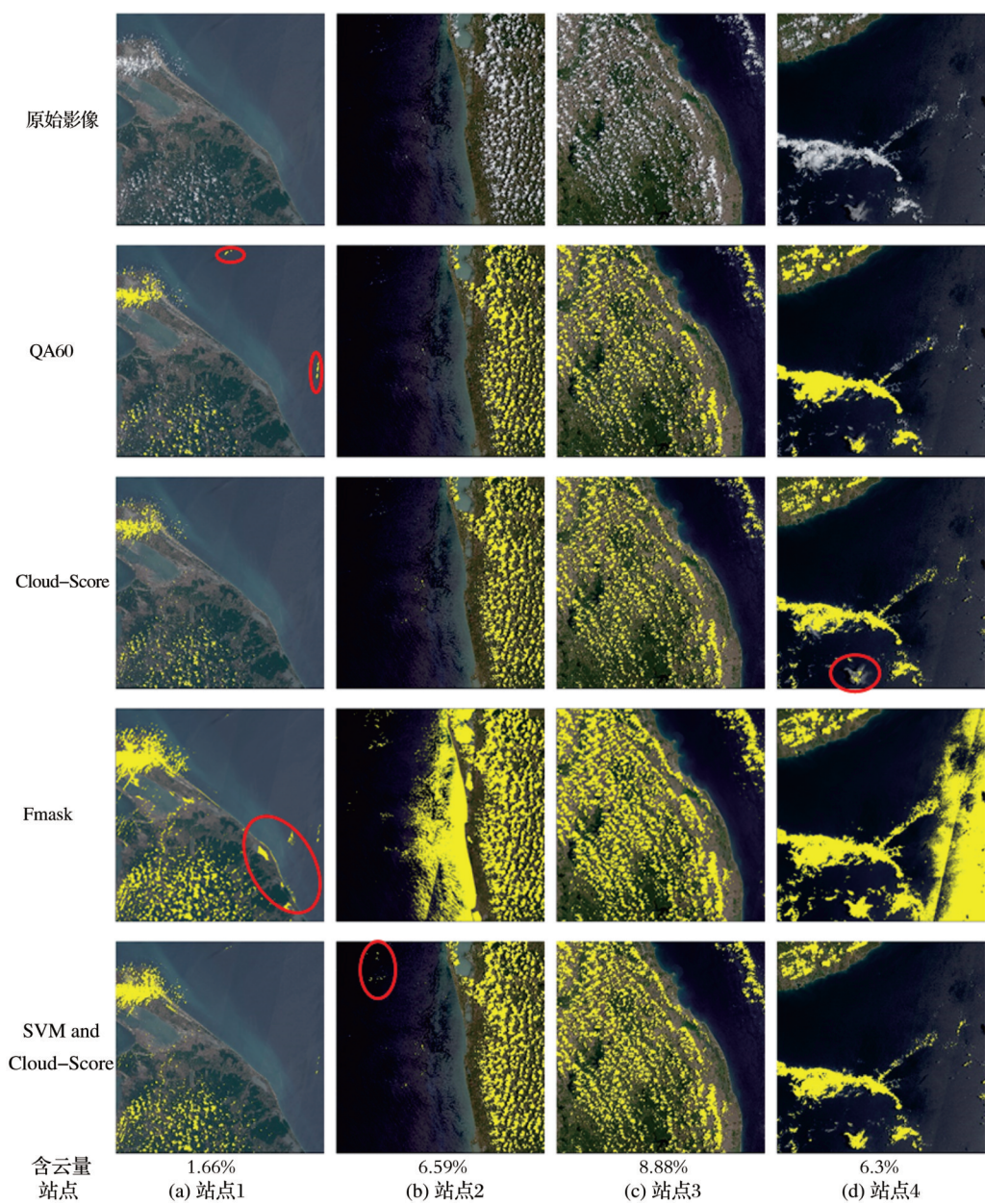


图 3 不同模型云检测结果

Fig.3 Cloud detection results of different models

表 2 不同模型云检测精度

Table 2 Cloud detection accuracy of different models			
模型	漏提率/%	误提率/%	总体精度/%
QA60	25.35	7.21	78.16
Cloud-Score	8.64	2.17	89.83
Fmask	2.27	38.74	63.45
SVM and Cloud-Score	1.06	0.15	98.21

型在两个地区不同含云量条件下的云检测测试中均表现出了较高的精度,能准确地识别出云体的边界。图 4 为两个地区高含云量条件下不同模型的云检测结果。从图中可以看出,不同模型在高云量条件下的云检测结果同样差异显著,QA60 波段法和

表 3 影像具体信息

Table 3 Specific information of images				
地点	条带号	日期	传感器	含云量
海南岛	T49QCB	2018/11/04	Sentinel-2B	8.79%
		2020/01/08	Sentinel-2B	13.23%
		2018/05/23	Sentinel-2A	21.22%
		2019/11/09	Sentinel-2B	30.92%
		2018/01/13	Sentinel-2A	42.60%
亚马逊 热带森林	T18LXP	2020/04/22	Sentinel-2A	4.50 %
		2019/10/25	Sentinel-2A	19.74%
		2018/07/27	Sentinel-2B	27.86%
		2019/06/22	Sentinel-2B	36.21%
		2021/01/17	Sentinel-2A	47.09%

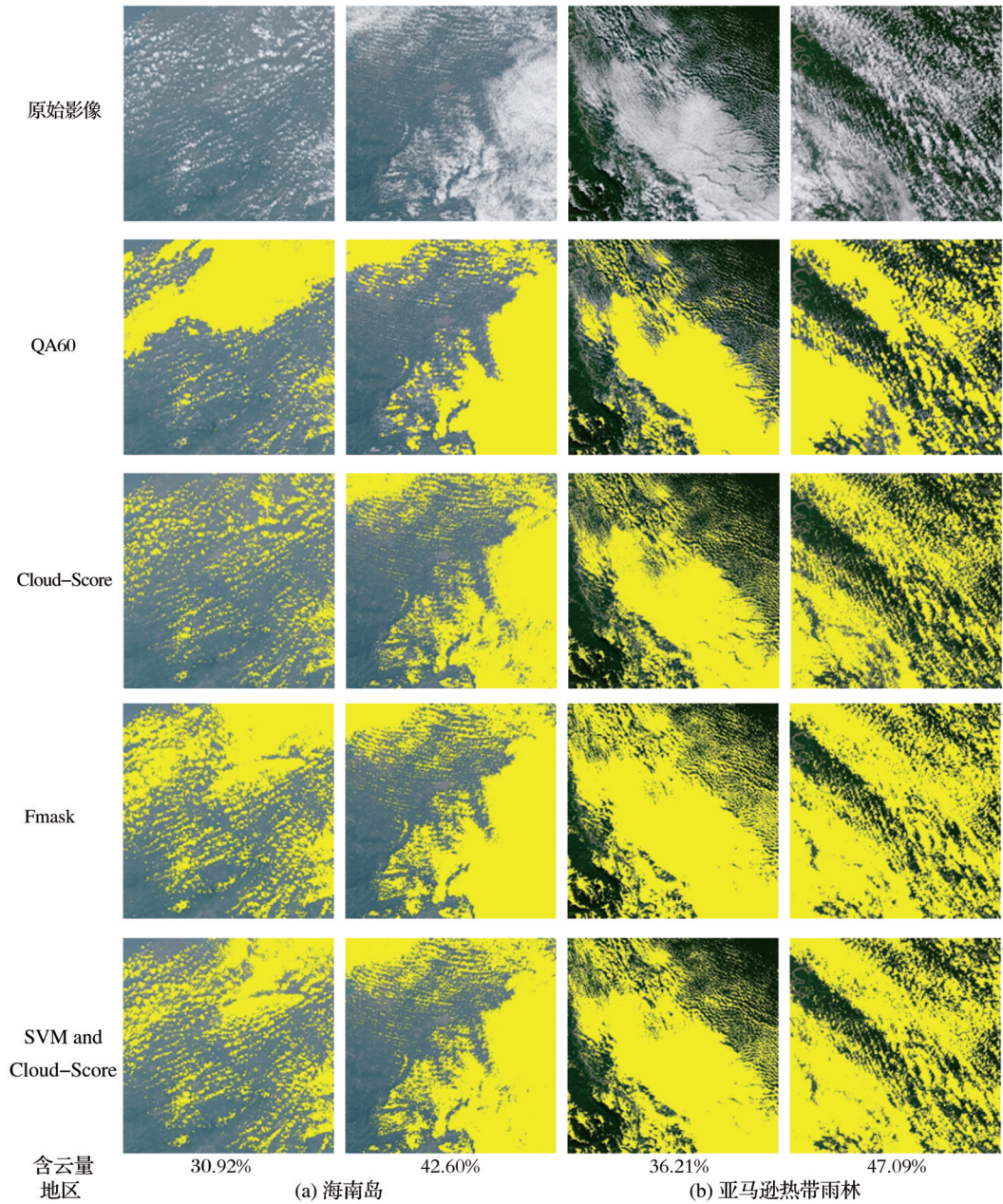


图 4 云检测结果

Fig.4 Cloud detection results

Fmask存在明显的误提现象,将大量存在于云体中间的清晰像元识别为云;QA60波段法和 Cloud-Score算对卷云的识别效果有限,但本文模型对海南岛和亚马逊热带森林的卷云识别效果较好,能比较精准地提取出卷云与其他地物的边界,仅存在部分少量的薄卷云漏提现象。耦合 SVM 和 Cloud-Score算法云检测模型的实现依托于 Google Earth Engine云平台强大的计算能力,确保了模型的运行效率,并且针对单景影像的样本训练和分类满足了精度和通用性要求。同时 Google Earth Engine云平台中易于实现模型的扩展和批量影像的处理操作,

因此在大规模的云检测应用方面具有一定潜力。

6 结 论

以热带多云的斯里兰卡为研究区,通过耦合 SVM 和 Cloud-Score算法,构建了适用于 Google Earth Engine云平台的 Sentinel-2云检测模型;对比了该模型与 QA60法、Cloud-Score算法以及 Fmask的云检测精度,并在海南岛和亚马逊热带森林验证了该模型云检测的通用性和鲁棒性。研究结果表明:Fmask模型的云检测性能最低,总体精度仅为 63.45%,存在严重的水体误分为云的现象,但其漏

提率极低;QA60法对卷云识别不足,漏提率较高,同时存在一定的误分现象,并且云体边界结果的细节性较差;Cloud-Score算法的云检测性能明显好于QA60法,总体精度达到了89.83%,误提率仅为2.17%,但仍存在部分卷云漏提的现象;相比于其他3种云检测方法,本文提出的耦合SVM和Cloud-Score算法的云检测模型总体精度最高,达到了98.21%,并且拥有极低的漏提率和误提率,能比较精准地识别出云体的边界,可满足Sentinel-2遥感产品的云检测预处理需求。在后续的研究中,将探究模型在Landsat系列影像的云检测潜力,同时聚焦大批量影像的云检测。

参考文献(References):

- [1] Foga S, Scaramuzza P L, Guo S, *et al.* Cloud detection algorithm comparison and validation for operational Landsat data products [J]. *Remote Sensing of Environment*, 2017, 194: 379-390. DOI: 10.1016/j.rse.2017.03.026.
- [2] Li Z, Shen H, Cheng Q, *et al.* Deep learning based cloud detection for medium and high resolution remote sensing images of different sensors [J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2019, 150: 197-212. DOI: 10.1016/j.isprsjprs.2019.02.017.
- [3] Sun L, Yang X, Jia S, *et al.* Satellite data cloud detection using deep learning supported by hyperspectral data [J]. *International Journal of Remote Sensing*, 2020, 41(4): 1349-1371. DOI: 10.1080/01431161.2019.1667548.
- [4] Wang C, Liu Y, Guo L, *et al.* A smart agricultural service platform for crop planting, monitoring and management-PIE-engine landscape [C]// 2021 9th International Conference on Agro-Geoinformatics (Agro-Geoinformatics). IEEE, 2021: 1-6. DOI: 10.1109/Agro-Geoinformatics50104.2021.9530347.
- [5] Gorelick N, Hancher M, Dixon M, *et al.* Google Earth Engine: Planetary-scale geospatial analysis for everyone [J]. *Remote sensing of Environment*, 2017, 202: 18-27. DOI: 10.1016/j.rse.2017.06.031.
- [6] Xiong J, Thenkabail P S, Gumma M K, *et al.* Automated cropland mapping of continental Africa using Google Earth Engine cloud computing [J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2017, 126: 225-244. DOI: 10.1016/j.isprsjprs.2017.01.019.
- [7] Carrasco L, O'Neil A W, Morton R D, *et al.* Evaluating combinations of temporally aggregated Sentinel-1, Sentinel-2 and Landsat 8 for land cover mapping with Google Earth Engine [J]. *Remote Sensing*, 2019, 11(3): 288. DOI: 10.3390/rs11030288.
- [8] Wang L, Diao C, Xian G, *et al.* A summary of the special issue on remote sensing of land change science with Google Earth Engine [J]. 2020, 248: 112002. DOI: 10.1016/j.rse.2020.112002.
- [9] ESA. Level-1C Cloud Masks-Sentinel-2 MSI Technical Guide-Sentinel Online [EB/OL]. <https://sentinel.esa.int/web/sentinel/technical-guides/sentinel-2-msi/level-1c/cloud-masks>. 2020.
- [10] Housman I W, Chastain R A, Finco M V. An evaluation of forest health insect and disease survey data and satellite-based remote sensing forest change detection methods: Case studies in the United States [J]. *Remote Sensing*, 2018, 10(8): 1184. DOI: 10.3390/rs10081184.
- [11] EO Research. Cloud Masks at Your Service [EB/OL]. <https://medium.com/sentinel-hub/cloud-masks-at-your-service-6e5b2cb2ce8a>. 2020.
- [12] Clerc S, Devignot O, Pessiot L. S2 MPC data quality report [R]. 2015.
- [13] Coluzzi R, Imbrenda V, Lanfredi M, *et al.* A first assessment of the Sentinel-2 Level 1-C cloud mask product to support informed surface analyses [J]. *Remote Sensing of Environment*, 2018, 217: 426-443. DOI: 10.1016/j.rse.2018.08.009.
- [14] Nguyen M D, Baez-Villanueva O M, Bui D D, *et al.* Harmonization of Landsat and Sentinel 2 for crop monitoring in drought prone areas: Case studies of Ninh Thuan (Vietnam) and Bekaa (Lebanon) [J]. *Remote Sensing*, 2020, 12(2): 281. DOI: 10.3390/rs12020281.
- [15] Qiu S, Zhu Z, He B. Fmask 4.0: Improved cloud and cloud shadow detection in Landsats 4-8 and Sentinel-2 imagery [J]. *Remote Sensing of Environment*, 2019, 231: 111205. DOI: 10.1016/j.rse.2019.05.024.
- [16] Main-Knorn M, Pflug B, Louis J, *et al.* Sen2Cor for Sentinel-2 [C]// Image and Signal Processing for Remote Sensing XXIII. International Society for Optics and Photonics, 2017, 10427: 1042704. DOI: 10.1117/12.2278218.
- [17] Hagolle O, Huc M, Auer S, *et al.* MAJA algorithm theoretical basis document [R]. 2017.
- [18] Zupanc A. Improving cloud detection with machine learning [R]. 2017.
- [19] Ke G, Meng Q, Finley T, *et al.* Lightgbm: A highly efficient gradient boosting decision tree [J]. *Advances in Neural Information Processing Systems*, 2017, 30: 3146-3154. DOI: 10.1080/09512748.2017.1375000.
- [20] Chung C. What are the strategic and economic implications for South Asia of China's Maritime Silk Road initiative? [J]. *The Pacific Review*, 2018, 31(3): 315-332. DOI: 10.1080/09512748.2017.1375000.
- [21] Drusch M, Del Bello U, Carlier S, *et al.* Sentinel-2: ESA's optical high-resolution mission for GMES operational services [J]. *Remote Sensing of Environment*, 2012, 120: 25-36. DOI: 10.1016/j.rse.2011.11.026.
- [22] Sanchez A H, Picoli M C A, Camara G, *et al.* Comparison of Cloud cover detection algorithms on Sentinel-2 images of the amazon tropical forest [J]. *Remote Sensing*, 2020, 12(8): 1284. DOI: 10.3390/rs12081284.
- [23] Li Jianfeng, Ye Huping, Zhang Zongke, *et al.* Spatiotemporal change analysis of Sri Lanka inland water based on Landsat imagery [J]. *Geo-Information Science*, 2019, 21(5): 781-8. [李

- 健锋, 叶虎平, 张宗科, 等. 基于 Landsat 影像的斯里兰卡内陆湖库水体时空变化分析[J]. 地球信息科学学报, 2019, 21(5): 781-788.]
- [24] Deng Naiyang, Tian Yingjie. A new method of data mining-support vector machine [M]. Beijing: Science Publication, 2004.[邓乃扬, 田英杰. 数据挖掘中的新方法: 支持向量机 [M]. 北京: 科学出版社, 2004.]
- [25] Guo Y, Jia X, Paull D. Effective sequential classifier training for SVM-based multitemporal remote sensing image classification[J]. IEEE Transactions on Image Processing, 2018, 27(6): 3036-3048. DOI: 10.1109/TIP.2018.2808767.
- [26] Shaharum N S N, Shafri H Z M, Ghani W A W A K, *et al.* Oil palm mapping over Peninsular Malaysia using Google Earth Engine and machine learning algorithms [J]. Remote Sensing Applications: Society and Environment, 2020, 17: 100287. DOI: 10.1016/j.rsase.2020.100287.
- [27] Hsu C W, Chang C C, Lin C J. A practical guide to support vector classification [R]. Department of Computer Science, 2003.
- [28] Baetens L, Desjardins C, Hagolle O. Validation of copernicus Sentinel-2 cloud masks obtained from MAJA, Sen2Cor, and FMask processors using reference cloud masks generated with a supervised active learning procedure [J]. Remote Sensing, 2019, 11(4): 433. DOI: 10.3390/rs11040433.
- [29] Congalton R G. A review of assessing the accuracy of classifications of remotely sensed data [J]. Remote Sensing of Environment, 1991, 37(1): 35-46. DOI: 10.1016/0034-4257(91)90048-B.

Research on Cloud Detection Model of Sentinel-2 Image Coupled with SVM and Cloud-Score Algorithm

Li Jianfeng^{1,3,4,5}, Liu Siqu^{1,3,4,5}, Li Jinbin^{1,3,4,5}, Peng Biao^{1,3,4,5}, Ye Huping²

(1. *Institute of Land Engineering and Technology, Shaanxi Provincial Land Engineering Construction Group Co., Ltd., Xi'an 710021, China;*

2. *State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China;*

3. *Shaanxi Provincial Land Engineering Construction Group Co., Ltd., Xi'an 710075, China;*

4. *Key Laboratory of Degraded and Unused Land Consolidation Engineering, the Ministry of Natural Resources, Xi'an 710075, China;*

5. *Shaanxi Provincial Land Consolidation Engineering Technology Research Center, Xi'an 710075, China)*

Abstract: Cloud coverage hinders the effective range of earth observation by optical remote sensing satellite. Rapid and accurate cloud detection is an important step in the product generation process of remote sensing application. In view of the lack of suitable and high-quality cloud detection model in Google Earth Engine cloud platform, this study takes tropical cloudy Sri Lanka as the study area, constructs a Sentinel-2 image cloud detection model coupled with SVM and Cloud-Score algorithm. Through experiments, the cloud detection accuracy of this method is compared with that of QA60 method, Cloud-Score algorithm and Fmask from the point of view of visual interpretation and quantitative analysis. The results show that the cloud detection performance of Fmask model is the lowest, and the overall accuracy is only 63.45%. It has a serious phenomenon that water body is mistakenly divided into clouds, but its omission error is very low. The QA60 method has insufficient recognition of cirrus clouds, and the omission error is high. At the same time, it has a certain phenomenon of misclassification, and the low spatial resolution affects the details of cloud boundary extraction results. The cloud detection performance of the Cloud-Score algorithm is obviously better than that of the QA60 method, the overall accuracy is 89.83%, and the commission error is only 2.17%, but there is still a phenomenon that some cirrus clouds are missed. Compared with the other three cloud detection methods, the cloud detection model proposed in this study has the highest overall accuracy, reaching 98.21%, and has extremely low omission error and commission error. The model can accurately identify the boundary of the cloud, and can meet the cloud detection preprocessing requirements of Sentinel-2 remote sensing products.

Key words: Cloud detection; SVM; Cloud-Score algorithm; Sentinel-2; Fmask