
中国科学院上海技术物理研究所刘银年团队-CloudAHSI：面向高分五号AHSI云分割的高光谱数据集

MDPI Remote Sensing

作者：writer 来源：科学网

本文原地址：<https://www.iikx.com/news/progress/41781.html>

本文仅供学习交流之用，版权归原作者所有，请勿用于商业用途！

中国科学院上海技术物理研究所刘银年团队-

CloudAHSI：面向高分五号AHSI云分割的高光谱数据集 MDPI Remote

Sensing。论文标题：CloudAHSI: A Hyperspectral Dataset for Cloud Segmentation from GF-5 AHSI

论文链接：<https://www.mdpi.com/2072-4292/18/9/1269>

期刊名：Remote Sensing

期刊主页：<https://www.mdpi.com/journal/remotesensing>

2018年5月9日2时28分，高分五号01（GF-5 01）卫星在太原卫星发射中心成功发射，是全球首颗实现对大气和陆地综合观测的全谱段高光谱卫星。其搭载的可见短波红外高光谱相机(Advanced Hyperspectral Imager, AHSI)，是国际上首台同时兼顾宽覆盖和宽谱段的高光谱相机，已在碳排放监测、土壤有机质与重金属污染评估、水质微量污染检测及矿物精细识别等应用中取得多项突破。然而，高光谱成像过程难以避免云污染，云检测作为光学遥感数据预处理中的关键环节，对后续应用质量具有重要影响。但是，传统的云检测数据集存在光谱覆盖有限、注释尺度小以及场景多样性不足等问题，阻碍了高光谱云检测算法的发展。当前，面向云检测的高光谱数据集仍较为匮乏。针对这一现状，中国科学院上海技术物理研究所刘银年团队研发出了面向高分五号AHSI云分割任务的高光谱数据集——CloudAHSI，成果发表在了Remote Sensing期刊。

图1.CloudAHSI数据集的构建流程：数据获取与场景选择、预处理、标注、切片与增强。

研究过程与结果

针对高光谱云检测中波段冗余度高、云与地物光谱混淆严重以及现有数据集缺乏短波红外波段覆盖等问题，作者构建了首个基于高分五号01星AHSI传感器的覆盖400 – 2500 nm高光谱云检测数据集CloudAHSI，并提出了分光谱区间主成分分析（S-PCA）的云敏感光谱特征降维方法。数据集包含全球45景原始影像及扩展子场景，涵盖裸土、森林、草地/农田、灌木林、雪/冰、城市、湿地和水体八类典型地物。预处理阶段，先对原始330个波段完成辐射定标与TOA反射率转换，再采用S-PCA将数据降维至12个通道，有效保留云敏感信息的同时显著降低计算负担。标注环节设计了光谱先验粗标注+人工精修的半监督框架：基于多阈值自动生成初始云掩膜，借助SWIR假彩色合成（R:2200 nm, G:1620 nm, B:1380 nm）辅助判据，由人工在Photoshop中逐像素精修，最终生成厚云、薄云、晴空三类像素级标签。数据集经滑动窗口裁剪为2052个512×512影像块（训练1792/测试260），并按云覆盖比例与地物类型对测试子集细化划分。

与现有数据集相比，CloudAHSI具有三方面优势：（1）400 – 2500 nm连续谱段覆盖，填补了短波红外波段在云检测数据中的空白；（2）大规模多样本地物类型与领先的像素级规模，支持算法在特定场景下的精细化评估；（3）多尺度兼容性，同时提供原始场景与标准影像块，兼顾全局上下文与小样本高效训练。

基于该数据集，作者构建了涵盖不同云层厚度、多种地表类型及极端场景的测试基准，为系统评估云检测模型的性能、场景适应能力及算法稳定性提供了有力支撑。选用七种主流方法开展基准实验，结果表明所有模型均取得稳健的基线性能（F1：77.28% – 87.91%，MIoU：72.12% – 83.31%）。厚云场景下各方法F1达84.52% – 94.04%，而薄云场景则显著下降至62.92% – 81.14%，揭示了模型对云厚度变化的高度敏感性，同时各模型间的性能差异也从侧面验证了数据集设计的合理性与区分度。消融实验进一步表明，相较全局PCA，S-PCA使UNet的F1提升2.57%、召回率提升10.00%，证实了分光谱降维策略在保留云敏感光谱信息方面的显著优势。

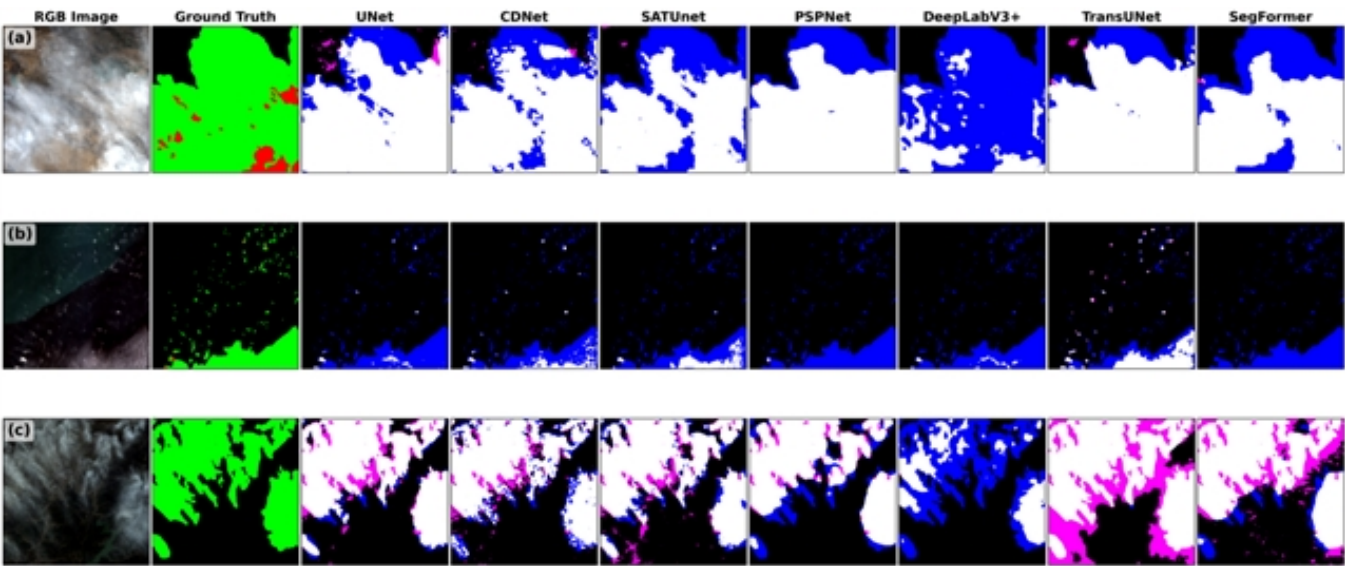


图2.不同云检测方法在薄云主导场景结果示例：其中，Ground Truth中红色为厚云，绿色为薄云，黑色为非云。检测结果中白色为正确检测，蓝色为漏检，紫色为误检。

研究总结

本文介绍了首个基于高分五号01星AHSI数据构建的400 – 2500 nm高光谱云检测数据集CloudAHSI，系统介绍了数据集的构建流程、标注策略和基准实验。实验表明，CloudAHSI在复杂地表背景

下的云检测任务中有效支持深度学习模型，特别是在薄云的检测和评估中展现了显著的数据价值，从而满足从像素分割到场景理解等多层次云检测需求。该数据集的发布为克服高光谱云检测中的频谱混淆瓶颈和提升高光谱遥感数据的有效利用提供了关键数据基础。

Jia, Y.; Zhao, S.; Liu, X.; Liu, Y. CloudAHSI: A Hyperspectral Dataset for Cloud Segmentation from GF-5 AHSI. Remote Sens. 2026, 18, 1269.

期刊介绍

主编：Prasad S. Thenkabail, USGS Western Geographic Science Center (WGSC), USA;

Dongdong Wang, Peking University, China

期刊范围涵盖遥感科学所有领域，从传感器的设计、验证和校准，到遥感在地球科学、环境生态、城市建筑等各方面的广泛应用。

2025 Impact Factor：4.3

2025 CiteScore：9.4

Time to First Decision：22 Days

Acceptance to Publication：2.9 Days

来源：Remote Sensing

更多 科学进展 请访问 <https://www.iikx.com/news/progress/>

本文版权归原作者所有，请勿用于商业用途，[爱科学iikx.com](https://www.iikx.com)转发