

中图法分类号: TP391 文献标识码: A 文章编号: 1006-8961(XXXX)XX-0001-14

论文引用格式: Xu Fang, Xu Jian, Yang Wen, Xia Gui-Song. BatonTrack1.0: A Benchmark for Multi-Object Tracking Based on Multi-Satellite Cooperative Observation[J/OL]. Journal of Image and Graphics, XXXX: 1-14. DOI: 10.11834/jig.260390. (徐芳, 徐健, 杨文, 夏桂松. 面向多卫星协同观测的多目标跟踪基准数据集 BatonTrack1.0[J/OL]. 中国图象图形学报, XXXX: 1-14. DOI: 10.11834/jig.260390.) [DOI: 10.11834/jig.260390]

面向多卫星协同观测的多目标跟踪基准数据集 BatonTrack1.0

徐芳¹, 徐健², 杨文³, 夏桂松¹

1. 武汉大学人工智能学院, 武汉 430072; 2. 武汉大学计算机学院, 武汉 430072; 3. 武汉大学电子信息学院, 武汉 430072

摘要: 目的 现有遥感卫星视频目标跟踪数据集主要面向单星连续观测, 受单星短时凝视观测限制, 难以支撑长时间目标持续跟踪研究。针对上述问题, 本文构建了首个面向多卫星协同观测的多目标跟踪基准数据集 BatonTrack1.0。方法 基于吉林一号 GF03C01、GF03C02 和 GF03C03 三颗视频卫星对同一区域的真实接续成像数据, 构建跨卫星多目标跟踪序列。通过动态裁剪方式保留目标在不同卫星观测片段中的运动延续关系, 并为每个序列提供标准多目标跟踪标注和跨片段时空关系标注, 包括目标位置、身份编号、帧时间戳及片段间投影变换关系。在此基础上, 构建仿真多卫星观测数据作为训练数据, 并选取 12 种主流多目标跟踪算法在 BatonTrack1.0 上进行统一评测, 涵盖基于运动建模、外观表征以及运动-外观联合关联的代表性方法。结果 BatonTrack1.0 共构建 8 个多目标跟踪序列, 标注 192 架飞机目标, 包括静止、起飞、降落、进出港滑行和避让等多种典型运行状态; 相邻跨卫星观测片段之间存在明显时间空窗, 目标在未观测期间持续运动, 跨片段位置和运动状态可能发生显著变化。实验结果表明, 现有方法在单个卫星观测片段内能够取得较好的目标定位与短时跟踪效果, 但在跨卫星片段切换时难以保持目标身份一致性。结论 BatonTrack1.0 能够为真实多卫星协同观测条件下的跨卫星目标关联、身份保持和长时间持续跟踪研究提供数据支撑与统一评测基准, 推动相关算法的设计、验证与创新, 对于支持广域目标持续感知能力建设具有重要意义。数据集获取链接: <https://doi.org/10.57760/sciencedb.46325>。

关键词: 目标检测; 目标跟踪; 多目标跟踪; 遥感视频; 多卫星协同观测

BatonTrack1.0: A Benchmark for Multi-Object Tracking Based on Multi-Satellite Cooperative Observation

Xu Fang¹, Xu Jian², Yang Wen³, Xia Gui-Song¹

1. School of Artificial Intelligence, Wuhan University, Wuhan 430072, China; 2. School of Computer Science, Wuhan University, Wuhan 430072, China; 3. School of Electronic Information, Wuhan University, Wuhan 430072, China

Abstract: **Objective** Wide-area persistent surveillance is crucial for a broad range of applications, including maritime traffic monitoring, air traffic surveillance, emergency response, and national security. Compared with ground-based, air-borne, radar-based, and other observation systems, video satellites feature extensive spatial coverage and are subject to fewer constraints in data acquisition, providing valuable observations for multi-object tracking (MOT) in complex scenes.

收稿日期: 2026-06-30; 修回日期: 2026-09-04

基金项目: 国家重点研发计划(2024YFF1401003); 国家自然科学基金(62325111, 62401406); 湖北省自然科学基金(2026AFA027); 测绘遥感信息工程全国重点实验室专项科研经费资助

Supported by: National Key Research and Development Program of China (2024YFF1401003); National Natural Science Foundation of China (62325111, 62401406); Hubei Provincial Natural Science Foundation of China (2026AFA027); LIESMARS Special Research Funding

©中国图象图形学报版权所有

In recent years, satellite video object tracking has received increasing attention, with continued advances in tracking methods and benchmark datasets. However, most existing studies focus on continuous observation by a single satellite. Restricted by satellite orbital motion, imaging field of view, staring capability, and other factors, the continuous observation window of a single satellite typically lasts only tens of seconds, making it difficult to fully characterize the long-term spatiotemporal evolution of objects. With the rapid development of satellite constellations, Earth observation is shifting from independent single-satellite observation toward cooperative multi-satellite observation. Multiple satellites can successively observe the same objects at different times, orbital positions, and viewing angles, thereby improving the temporal continuity and spatial coverage of observations. In practice, however, successive multi-satellite imaging is constrained by factors such as constellation resource scheduling, orbital distribution, satellite overpass times, and attitude maneuvering capabilities. Consequently, observations acquired by different satellites are generally discontinuous in time, with temporal gaps frequently occurring between adjacent observation segments, making it difficult to maintain trajectory continuity and identity consistency. Therefore, fully exploiting multi-satellite cooperative observation to achieve cross-satellite object association and long-term continuous tracking remains a key challenge in Earth observation. To support research on this challenge, we construct BatonTrack1.0, the first benchmark dataset for multi-object tracking under multi-satellite cooperative observation. **Method** BatonTrack1.0 is built from real successive observations acquired by three Jilin-1 video satellites (GF03C01, GF03C02, and GF03C03) at two representative aviation hubs, Abu Dhabi Airport and New Delhi Airport, with coverage spanning key functional areas, including runways, taxiways, aprons, and arrival and departure zones. Each video segment lasts approximately 14 s and has a frame rate of 5 frames per second (fps). Since the satellite overpasses are not temporally contiguous, adjacent video segments contain observation gaps ranging from approximately 1 min 44 s to 2 min 1 s. To preserve the continuity of object motion across successive satellite observations, we construct cross-satellite multi-object tracking sequences by dynamically cropping each segment according to the locations of the objects, rather than extracting the same fixed geographic area across all segments. Each sequence is annotated with both standard multi-object tracking labels and cross-segment spatiotemporal annotations, including object locations, identity labels, frame time-stamps, and inter-segment projection transformations. As BatonTrack1.0 is constructed from real successive multi-satellite observations, it is relatively limited in scale. We therefore construct an additional simulated multi-satellite video dataset based on the SAT-MTB dataset for training, and reserve BatonTrack1.0 exclusively for testing to evaluate algorithm performance under real-world multi-satellite observation conditions. We benchmark 12 representative multi-object tracking algorithms that rely on motion cues, appearance cues, or both for object association. Unlike conventional multi-object tracking under single-satellite observation, multi-object tracking under multi-satellite cooperative observation requires algorithms to maintain trajectory continuity within individual observation segments while preserving object identities across segments. To determine where the primary performance bottleneck lies in this task, we report results for both complete cross-satellite sequences and isolated single-satellite video segments. **Result** BatonTrack1.0 contains 8 multi-object tracking sequences, with each comprising two or three satellite video segments. The spatial coverage of each segment is dynamically adjusted according to the locations of the objects, and the segments correspond to temporally non-overlapping short-term observation windows. All video frames have a spatial resolution of 1,000×1,000 pixels. Together, the sequences contain annotations for 192 aircraft, covering typical operational states such as takeoff, landing and taxiing. Existing methods exhibit favorable performance in object localization and short-term tracking within individual satellite video segments, but struggle to maintain identity consistency across segments acquired by different satellites. A comparison between complete-sequence and individual-segment evaluations shows that Multiple Object Tracking Accuracy (MOTA), which is strongly influenced by detection performance, differs only slightly, whereas Identity Switches (IDS), which reflect identity association errors, decrease substantially under individual-segment evaluation. It can be attributed to three main factors. First, temporal gaps occur between adjacent satellite video segments, during which objects remain unobserved while continuing to move, thereby reducing the reliability of trajectory prediction based on short-term motion continuity. Second, disparities in viewing angles, imaging geometry, and spatial coverage across satellites lead to significant object position shifts between segments, further complicating cross-segment identity association. In addition, aircraft appear as tiny objects with limited textual detail in satellite imagery, weakening appearance-based identity discrimination. When multiple visually similar air-

craft congregate on taxiways, runways, or aprons, association strategies relying solely on local motion cues or appearance similarity are prone to identity confusion. Therefore, robust and reliable identity association across video segments under temporal discontinuities, viewpoint variations, and weak appearance representations remains an important unresolved problem in multi-object tracking with successive multi-satellite observations. Future research could integrate scene priors, object attributes and behavioral states, long-term trajectory prediction, and inter-object relationship modeling to improve the reliability and robustness of persistent monitoring of objects in multi-satellite observation scenarios. **Conclusion** We propose BatonTrack1.0, the first public benchmark dataset tailored for multi-object tracking under cooperative multi-satellite observation. Derived from real-world sequential satellite data, BatonTrack1.0 provides a standardized foundation to develop, validate, and compare advanced algorithms capable of preserving object identities across spatial-temporal gaps. BatonTrack1.0 is publicly available at <https://doi.org/10.57760/sciencedb.46325>. In future work, we will further expand the dataset with real multi-satellite cooperative observation sequences, broaden its coverage of observation scenarios and object categories, and include more samples spanning diverse temporal gaps, viewpoint variations, and imaging geometries.

Key words: object detection; object tracking; multi-object tracking; satellite video; multi-satellite cooperative observation

论文引用格式: Xu Fang, Xu Jian, Yang Wen, Xia Gui-Song. BatonTrack1.0: A Benchmark for Multi-Object Tracking Based on Multi-Satellite Cooperative Observation [J/OL]. Journal of Image and Graphics, XXXX: 1-14. DOI: 10.11834/jig.260390. (徐芳, 徐健, 杨文, 夏桂松. 面向多卫星协同观测的多目标跟踪基准数据集 BatonTrack1.0 [J/OL]. 中国图象图形学报, XXXX: 1-14. DOI: 10.11834/jig.260390.) [DOI: 10.11834/jig.260390]

论文引用格式: Xu Fang, Xu Jian, Yang Wen, Xia Guisong. BatonTrack1.0: A Multi-Object Tracking Benchmark Dataset for Multi-Satellite Collaborative Observation [J/OL]. Journal of Image and Graphics. DOI: 10.11834/jig.260390. (徐芳, 徐健, 杨文, 夏桂松. 面向多卫星协同观测的多目标跟踪基准数据集 BatonTrack1.0 [J/OL]. 中国图象图形学报. DOI: 10.11834/jig.260390.)

0 引言

广域目标持续感知是海上交通监测、航空目标监管、应急救援及国防安全等国家重大需求的关键支撑(Liu等, 2025; Zhi等, 2026)。与地面监控、机载平台和雷达探测等手段相比, 遥感视频卫星具有覆盖能力强、观测限制少和场景适应性高等优势, 为复杂区域内多目标跟踪提供了重要数据来源。近年来, 遥感卫星视频目标跟踪受到广泛关注, 相关方法

与数据集研究不断推进(Li等, 2023; He等, 2022; Xiao等, 2024; Lu等, 2025; Gao等, 2025; Xiao等, 2026)。然而, 现有研究主要聚焦于单星观测场景。受卫星轨道运动、成像视场和凝视时间等因素限制, 单星连续凝视观测窗口通常仅为数十秒, 难以完整刻画目标在长时空维度上的连续演化过程。

在星座组网与空天信息产业快速发展的背景下, 对地观测正由单星独立观测逐步向多星协同观测模式转变(Li等, 2024; Kang等, 2024)。多颗卫星能够在不同时间、不同轨道位置 and 不同观测视角下对目标进行接续成像, 从而显著提升目标观测的时间连续性与空间覆盖范围, 如图1所示。然而, 在实际观测过程中, 多星接续成像需要综合考虑星座资源调度、轨道分布、过境时刻和姿态机动能力等条件, 不同卫星观测在时间上难以连续衔接, 相邻观测片段之间往往存在时间空窗, 导致目标轨迹连续性和身份一致性难以保持。因此, 如何充分利用多星协同观测能力, 实现地表目标的跨卫星关联与

长时连续跟踪, 已成为遥感对地观测领域亟待解决的关键问题。

目前, 面向多卫星协同观测的多目标跟踪研究尚处于起步阶段。Li等人(2025)提出了跨视频微小目标跟踪方法CVTOT, 并构建了初步基准CVOT, 用于评估非连续视频片段之间的目标身份关联能力。然而, 该基准主要通过对已有单星视频进行片段切分来模拟跨视频场景, 对真实多卫星协同观测场景的刻画不足。一方面, 切分得到的视频片段来源于

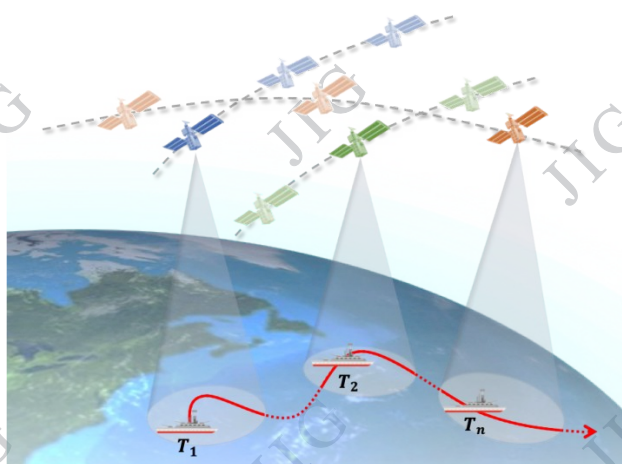


图1 多卫星协同观测下的目标跟踪示意图

Fig. 1 Illustration of object tracking under multi-satellite cooperative observation.

同一颗卫星、同一次观测过程,忽略了不同卫星在轨道位置、观测姿态、成像角度和几何投影等方面的变化,难以体现多星数据之间的内在差异;另一方面,受单星视频观测时长约束,切片片段之间通常只能形成相对有限的时间间隔,难以覆盖真实多星接续观测中的长时间空窗,因而不足以反映长时间空窗下目标持续运动所导致的身分关联不确定性增强问题。因此,构建一个能够真实反映多卫星协同观测复杂性的多目标跟踪数据集,对于支撑跨卫星目标关联、长时轨迹接续和动态目标持续感知算法研究具有重要意义。

为解决上述问题,本文构建了一个面向多卫星协同观测的多目标跟踪基准数据集 BatonTrack1.0。该数据集基于吉林一号 GF03C01、GF03C02 和 GF03C03 三颗视频卫星的真实接续观测数据构建,覆盖阿布扎比机场和新德里机场两个典型场景。与基于单星视频切分所获得的跨视频片段不同, BatonTrack1.0 直接来源于多颗视频卫星对同一区域的任务调度与接续成像,能够更真实地反映多卫星协同观测条件下的时间不连续、视角变化和成像几何差异等复杂观测特性。数据集中每段视频对地凝视时长约为 14 秒,相邻跨卫星观测片段之间存在明显时间空窗,其中最大空窗为 2 分 1 秒,最小空窗为 1 分 44 秒。BatonTrack1.0 共标注 192 架飞机目标,涵盖静止、起飞、降落、进出港滑行、避让等多种典型行为状态,为跨卫星多目标跟踪研究提供了真实数据基础。本文的主要贡献如下:

(1)据我们所知, BatonTrack1.0 是首个真实多卫星协同观测条件下的多目标跟踪基准数据集。该数据集由多颗吉林一号视频卫星接续观测获得,可用于开发和评估跨卫星目标关联、身份保持和长时间目标跟踪算法。后续我们将持续扩展数据集规模与场景覆盖范围,以更好支撑真实多星协同观测条件下的动态目标持续感知研究。

(2)本文在 BatonTrack1.0 上系统评测了多种主流多目标跟踪算法,包括基于运动信息关联的方法和基于外观信息关联的方法,建立了面向多卫星协同观测场景的初步性能基准,为后续算法设计、性能对比与方法改进提供统一参考。

1 相关工作

本文从三个方面介绍相关研究:首先,分析现有遥感卫星视频目标跟踪数据集及其在支撑长时间目标持续跟踪研究方面的不足;其次,概述多目标跟踪技术,阐明当前主流跟踪范式的核心机制及其在多卫星协同观测条件下的局限性;最后,讨论与跨卫星身份关联密切相关的多摄像头多目标跟踪技术,并分析其与多卫星接续观测在时空关系和成像条件方面的差异。

1.1 遥感卫星视频目标跟踪数据集

高分辨率对地观测技术快速发展,遥感视频卫星(如 SkySat-1、Jilin-1、LuoJia-3-01 等)通过凝视成像,突破了传统静态成像的局限,能够在一定时间窗口内连续记录地表目标的运动过程,成为目标动态监测的重要技术手段和数据来源。为推动该领域发展,研究者陆续构建了多个遥感卫星视频目标跟踪数据集用于算法研究,主要可分为单目标跟踪(Single Object Tracking, SOT)数据集和多目标跟踪(Multiple Object Tracking, MOT)数据集两类,如表 1 所示。

单目标跟踪数据集通常以单一指定目标为研究对象,主要用于评估算法在给定初始位置条件下对目标进行连续定位与状态估计的能力。例如, Sat-SOT 数据集(Zhao 等, 2022)由 105 个卫星视频序列组成,覆盖 26 列火车、65 辆汽车、9 架飞机和 5 艘轮船,为遥感卫星视频单目标跟踪提供了较早的标准评测基准。SV248S(Li 等, 2022)进一步扩展了卫星视频单目标跟踪数据规模,包含 248 个卫星视频序

表 1 现有遥感卫星视频目标跟踪数据集与 BatonTrack1.0
Table 1 Comparison of BatonTrack1.0 with existing satellite video object tracking datasets

数据集	任务	多卫星协同观测	最大观测时长(秒)
SatSOT	SOT	■	30
SV248S	SOT	■	30
OOTB	SOT	■	120
VISO	MOT	■	75
AIR-MOT	MOT	■	65
SAT-MTB	MOT	■	43
BatonTrack1.0	MOT	■	277

列,重点关注小目标在复杂背景下的稳定跟踪问题。OOTB(Chen 等, 2024)则在标注形式上引入旋转框,通过提供目标方向信息促进遥感卫星视频目标跟踪性能的提升。

然而,在机场、港口、道路及海域等实际遥感应用场景中,观测区域内往往包含多个同时运动且相互作用的目標。单目标跟踪仅关注给定目标的连续定位,难以刻画复杂场景中的群体动态特性。多目标跟踪旨在同时获取多个目标的位置、身份和轨迹信息,能够刻画目标群体的时空分布、运动规律和交互关系,为区域交通流分析、机场运行监测、海上船舶监管、异常事件发现和态势感知等应用提供基础数据支撑。围绕上述需求,VISO(Yin 等, 2021)、AIR-MOT(He 等, 2022)和 SAT-MTB(Li 等, 2023)等多目标跟踪数据集陆续被构建,为遥感卫星视频中多目标定位、跨帧关联和身份一致性保持等研究提供了重要基础。

总体而言,现有数据集推动了遥感卫星视频目标跟踪研究的发展,但主要基于单星连续观测视频构建,关注同一观测片段内的目标定位与跨帧关联,受单星凝视时间限制,难以支撑长时间目标持续跟踪研究。多星协同观测能够突破单星观测时长限制,为实现目标持续感知提供了重要技术路径,但目前仍缺乏面向真实多卫星接续观测场景的公开多目标跟踪数据集。为此,本文构建了 BatonTrack1.0,一个面向真实多卫星协同观测场景的多目标跟踪基准数据集,为跨卫星目标关联与长时间连续跟踪研究提供数据支撑。

1.2 多目标跟踪技术

多目标跟踪旨在连续观测序列中同时估计多个目标的位置与身份,并保持轨迹的一致性,其核心在于跨时刻的目标关联。现有方法大体可分为“先检测后跟踪”(Tracking-by-Detection, TBD)范式和“联合检测与跟踪”(Joint Detection and Tracking, JDT)范式。

在 TBD 范式中,数据关联通常作为检测后的独立步骤完成(Bewley 等, 2016; Wojke 等, 2017; Zhang 等, 2022; Qin 等, 2023; Lv 等, 2024; Li 等, 2024)。该类方法首先利用目标检测器获得每一帧中的目标候选框,再依据运动信息、外观特征及历史轨迹等信息完成跨帧目标关联。例如, FineTrack(Ren 等, 2023)将外观特征作为目标身份判别线索,并通过挖掘目标局部细节信息增强外观表征的判别性,从而缓解遮挡、目标重现和相似运动造成的身份混淆。TOPIC(Cao 等, 2025)则联合利用运动信息和外观信息构建关联关系,并根据目标运动复杂程度自适应选择更可靠的匹配线索,以提升复杂运动和外观混淆场景下的身份保持能力。

JDT 范式在统一框架中联合建模目标检测与身份关联,通过协同优化缓解检测与跟踪模块分离带来的误差传播(Zhou 等, 2020)。部分研究开始由基于相似度计算的显式匹配转向基于目标状态传播的关联方式。例如, MOTR(Zeng 等, 2022)利用 Transformer(Vaswani 等, 2017)维护目标实例表示,使同一目标的状态能够在连续帧中持续更新,从而在检测过程中隐式完成身份关联。针对长时遮挡和目标重现造成的身份不稳定问题, MeMOTR(Gao 等, 2023)进一步引入长期记忆机制,将历史目标信息用于当前帧的身份判断,以增强长时关联稳定性。在此基础上, SambaMOTR(Sege 等, 2025)进一步关注多目标之间的相互关系,通过多轨迹联合建模同时捕获单个目标的长期运动规律和目标间交互关系,以提升复杂交互场景下的身份保持能力。

总体而言,现有多目标跟踪方法大多面向连续视频帧设计,主要依赖外观相似性、运动连续性或帧间状态传播完成目标关联。然而,在多卫星协同观测条件下,同类目标视觉表征高度相似,外观匹配易导致身份混淆;同时,观测序列在时间上呈现稀疏与不连续特性,长时观测空窗使得目标运动偏离短时连续性假设,基于短时运动连续性或连续状态传播

的关联机制难以适用。

1.3 多摄像头多目标跟踪技术

多摄像头多目标跟踪旨在跨多个摄像头持续保持目标身份一致性。相关研究主要面向两类场景：一类依赖多摄像头同步观测，多个摄像头同时覆盖同一区域，通过跨视角目标关联缓解单一视角下的遮挡和轨迹丢失问题(Huang等, 2023; Specker等, 2023)。例如，MIA-Net(Liu等, 2023)面向多无人机协同观测场景，通过局部-全局匹配建模不同无人机视角下的目标拓扑对应关系，实现跨无人机身份关联，从而提升复杂遮挡场景下的轨迹连续性。另一类主要面向异步观测场景，不同摄像头的观测时间不同步，通常依赖目标重识别特征、摄像头拓扑关系和时空一致性约束，对不同摄像头中的轨迹片段进行匹配与连接，以缓解单摄像头轨迹断裂对全局跟踪的影响(Tang等, 2019; Cheng等, 2023)。例如，Nguyen等人(2024)面向城市交通场景，将单摄像头内生成的车辆轨迹片段作为关联单元，通过比较不同摄像头中轨迹片段的视觉与时空结构一致性实现跨摄像头车辆关联。

总体而言，现有多摄像头多目标跟踪方法主要面向地面固定摄像头或无人机近距观测场景，通常建立在跨视角观测在时间上近似同步的假设之上，通过利用具有判别性的目标外观特征以及可建模的时空一致性约束，实现跨视角的目标关联，其核心目标在于缓解单一视角下的遮挡问题。相比之下，多卫星协同观测在时间上呈现显著的稀疏与异步特性，不同观测片段之间缺乏连续的时空对应关系，同时卫星姿态与成像几何动态变化，且遥感目标尺度小、同类目标外观高度相似，使得依赖时空一致性与外观重识别的多摄像头跟踪方法难以直接适用于跨卫星身份关联任务。

2 BatonTrack1.0数据集构建

2.1 数据收集

BatonTrack1.0数据集涵盖阿布扎比机场和德里机场两个典型航空枢纽场景，观测范围覆盖跑道、滑行道、停机坪和进出港区域等关键功能区。针对每个场景，数据集采集了吉林一号GF03C01、GF03C02和GF03C03三颗视频卫星先后过境时获取的真实观测视频。每段视频时长约为14秒，帧率为

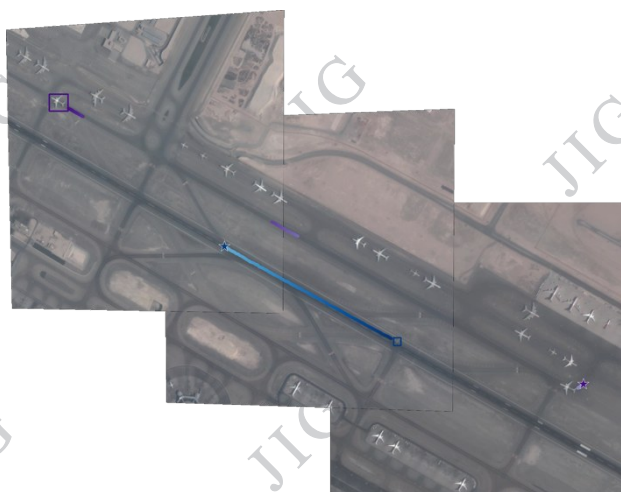
5fps(frames per second);由于三颗视频卫星的过境时间并不连续，相邻视频片段之间存在约1分44秒至2分1秒的时间间隔。

为记录目标在多颗卫星接续观测中的运动延续关系，本文根据目标在不同观测片段中的实际活动位置进行动态裁剪，而非在各片段中固定截取同一地理区域。基于上述策略，本文共构建了8个多目标跟踪序列，每个序列由2段或3段卫星视频片段组成。各片段的空间覆盖范围随目标位置动态调整，并分别对应不同的短时观测窗口；所有片段的帧图像尺寸均为1000×1000像素。8个序列共包含192架飞机目标，涵盖静止、滑行、起飞和降落等多种运行状态。

2.2 数据标注

针对每个多目标跟踪序列，BatonTrack1.0提供两类标注信息。第一类为标准多目标跟踪标注，用于描述每一帧中飞机目标的位置与身份信息。具体而言，每条标注表示为(*frame*, *id*, *bb_left*, *bb_top*, *bb_width*, *bb_height*, *conf*, *class*, *visibility*)，其中*frame*表示帧编号，跨片段帧编号按序连续递增；*id*表示目标身份编号，同一目标在同一序列的不同帧及不同卫星视频片段中保持一致的身份编号；(*bb_left*, *bb_top*, *bb_width*, *bb_height*)表示目标边界框的左上角坐标、宽度和高度；(*conf*, *class*, *visibility*)分别表示标注有效性标志、目标类别编号和目标可见比例，由于BatonTrack1.0采用人工标注且目标类别均为飞机，本文将其仅作为格式兼容项保留，并统一设为默认值-1。

值得注意的是，由于相邻卫星视频片段之间存在观测空窗，目标在此期间不可见，因而无法依据连续轨迹直接建立前后片段的身份对应关系。为保证跨卫星身份标注的准确性，BatonTrack1.0采用“时间一致性-空间连续性-运动合理性”相结合的人工标注策略。具体而言，首先依据各卫星视频片段的实际采集时间，确定不同观测片段的先后关系及时间间隔；随后，在统一地理坐标系下，结合跑道、滑行道和停机坪等机场功能区域之间的拓扑连通关系，推断目标在观测空窗内的可行运动路径和可能到达区域；最后，综合目标所在功能区域、运动方向、速度、轨迹趋势及其与周围目标的交互关系，判断候选对应关系是否符合机场场景约束和目标运动规律。全部标注结果均经过人工复核，对于无法唯一确认



(a)



(b)

图2 BatonTrack1.0中两组多卫星协同观测序列示例

Fig. 2 Two examples of multi-satellite cooperative observation sequences from BatonTrack1.0.

身份的目标予以排除,以保证身份关联评估结果的可靠性。

此外,BatonTrack1.0为每个多目标跟踪序列进一步提供跨片段时空关系标注,表示为(sat_num , $frame_timestamp$, $projection_matrix$),用于描述不同卫星视频片段之间的时间关系和空间对应关系,为跨卫星目标关联、轨迹接续以及时空一致性建模提供辅助信息。其中, sat_num 表示序列所含卫星视频片段数量; $frame_timestamp$ 表示各片段内每一帧的相对时间戳,以序列中的最小采样间隔作为基本时间单位,并以第一帧作为时间起点,用于记录各帧在真实观测时间轴上的相对位置; $projection_matrix$ 表示相邻片段间的投影变换矩阵,由相邻片段中手动标注的四组对应参考点估计得到,用于描述前一片段到后一片段的图像平面映射关系。

图2展示了其中两组典型序列示例。为便于展示,根据相邻观测片段之间的投影变换矩阵,将同一序列中的不同观测片段映射至统一坐标系。图中,★表示目标在该序列中的首次出现位置,同一目标在不同观测片段中的轨迹采用相同颜色表示,颜色由浅到深表示时间推进。可以发现,相较于单星观测,多卫星协同观测显著延长了目标的持续观测时长,能够更加完整地记录目标的状态变化和轨迹演化过程。

2.3 数据特性

跨片段时间不连续。BatonTrack1.0基于多颗视频卫星对同一区域的真实接续观测构建。与现有遥感卫星视频跟踪数据集主要面向单星短时连续观测不同,BatonTrack1.0中相邻视频片段对应不同卫星的短时观测窗口,片段之间存在时间间隔。由于目标在该时间间隔内仍会持续运动,跨片段目标位置和运动状态可能发生显著变化。如图2(a)所示,紫色轨迹对应的飞机在前一卫星观测结束后继续运动,再次出现在后一卫星视频中时,其位置已发生明显变化。因此,相邻片段之间难以直接依赖短时运动连续性完成目标关联,加剧了目标身份接续和轨迹关联的不确定性。

目标尺度小、纹理弱。BatonTrack1.0面向遥感卫星视频中的飞机目标。受卫星成像距离和空间分辨率限制,飞机目标在图像中占比较小,纹理细节和局部结构信息有限,难以提供稳定的外观判别线索。如图2所示,不同飞机目标之间的外观差异不显著,尤其是相同型号飞机,往往呈现高度相似的视觉特征。因此,目标之间易发生身份混淆,对跟踪算法学习稳定的目标判别线索、降低跨片段关联不确定性提出了更高要求。

目标间相互作用复杂。BatonTrack1.0中多个飞机目标可能同时出现在滑行道、停机坪和跑道等机场关键区域,并伴随会车、避让、等待和进出港滑行等交互过程。在此过程中,飞机目标的运动状态会受到机场运行约束及周围目标影响而动态调整,使目标运动不再表现为简单、稳定的连续过程。如图2(b)所示,蓝色轨迹对应的飞机在第二颗卫星观测窗口中向下运动,紫色轨迹对应的飞机为避让该目标,在第二颗卫星观测窗口内保持静止,在第三颗卫星观测窗口中继续运动。这种受交互约束影响的非匀速、非连续运动模式,进一步削弱了基于短时运

动连续性假设的轨迹预测可靠性,对算法的运动建模能力提出了更高要求。

3 实验

3.1 实验设置

考虑到 BatonTrack1.0 来源于真实多卫星接续成像,数据规模有限,本文进一步构建仿真多卫星视频数据集作为训练集,并将 BatonTrack1.0 作为真实测试集,评估算法在真实多卫星观测场景下的性能。具体而言,本文基于 SAT-MTB 数据集(Li 等, 2023)构造仿真多卫星视频序列,模拟多颗卫星在不同时间窗口内对同一区域进行协同成像的过程。SAT-MTB 与 BatonTrack1.0 均来源于吉林一号系列卫星,有助于减小基础成像特征的域差异,提高仿真训练数据与真实测试场景之间的可迁移性。SAT-MTB 数据集包含飞机、车辆、船舶等多类运动目标。鉴于 BatonTrack1.0 当前仅面向飞机目标,本文在仿真数据构建过程中仅使用其中的飞机类别数据。对于 SAT-MTB 数据集中的每段原始单星连续观测视频,首先从不同时间位置截取若干短时视频片段,以模拟相邻跨卫星观测片段之间的时空窗;随后对不同片段进行裁剪、平移和几何变换,以模拟跨卫星观测中的视野偏移与成像几何差异。最终,仿真训练集共构建了 36 个多目标跟踪序列,每个序列由 2~3 段视频片段组成,共包含 496 个目标实例。

在实验中,本文选取 12 种主流多目标跟踪算法进行对比评测,包括 JDT 范式方法 CenterTrack(Zhou 等, 2020)、FairMOT(Zhang 等, 2021)、MOTR(Zeng 等, 2022)、TGraM(He 等, 2022)、MeMOTR(Gao 等, 2023)、MP2Net(Zhao 等, 2024)和 TBD 范式方法 ByteTrack(Zhang 等, 2022)、MotionTrack(Qin 等, 2023)、MambaTrack(Li 等, 2024)、DiffMOT(Lv 等, 2024)、TOPIC(Cao 等, 2025)和 TrackTrack(Shim 等, 2025),涵盖基于运动建模、外观表征以及运动-外观联合关联的代表性方法。对于 TBD 范式方法,其检测模块和关联模块通常相互独立。本文统一采用 YOLOX-X(Ge 等, 2022)生成目标检测结果,并采用 FastReID(He 等, 2023)提取目标外观嵌入特征;各跟踪方法在相同检测输入和外观特征条件下执行轨迹关联,以保证不同 TBD 范式方法之间的公平比较。为评估各算法在多卫星协同观测场景下的适用

性,本文保持各算法原始跟踪框架,仅以图像帧序列为输入,不额外引入跨片段时空关系等先验信息。由于所评估算法均面向连续视频帧设计,对于每个多目标跟踪序列,本文按照实际观测顺序将各卫星视频片段依次拼接为连续帧序列,并保持跨片段帧编号连续,从而形成算法可直接处理的单一输入序列。片段切换时,跟踪器沿用前一片段结束时的跟踪状态,由各算法依据自身的关联机制完成跨片段身份关联。

所有实验均在 NVIDIA Tesla V100 GPU 上进行,各方法均采用原文中的参数设置,仅 MOTR 和 MeMOTR 两种基于 Transformer 架构的方法根据显存限制对 batch size 进行了调整。

3.2 评价指标

本文采用多目标跟踪领域常用的评价指标对算法性能进行评估,包括 MOTA(Multiple Object Tracking Accuracy)、IDF1(Identification F1 Score)、IDS(Identity Switches)和 MT(Mostly Tracked Targets)等。其中,MOTA 用于衡量整体跟踪精度,综合考虑漏检、误检和身份切换情况:

$$MOTA = 1 - \frac{\sum_t (FN_t + FP_t + IDS_t)}{\sum_t (GT_t)} \quad \#(1)$$

式中, FN_t 、 FP_t 和 IDS_t 分别表示第 t 帧中的漏检数、误检数和身份切换次数, GT_t 表示第 t 帧中的真实目标数。MOTA 越高,算法整体跟踪精度越好。IDF1 用于评估目标身份关联准确性:

$$IDF1 = \frac{2IDTP}{2IDTP + IDFP + IDFN} \quad \#(2)$$

式中,IDTP 表示身份匹配正确的目标数,IDFP 表示身份匹配错误的预测目标数,IDFN 表示未被正确匹配的真实目标数。IDF1 越高,说明算法在目标身份关联方面表现越稳定。IDS 表示跟踪过程中目标身份发生切换的总次数:

$$IDS = \sum_t IDS_t \quad \#(3)$$

IDS 越低,说明算法在连续跟踪过程中越不容易发生身份混淆。MT 表示在超过 80% 生命周期内被成功跟踪的目标数量:

$$MT = \sum_i \mathbf{I}(r_i \geq 0.8) \quad \#(4)$$

式中, r_i 表示第 i 条真实轨迹被成功跟踪的比例, $\mathbf{I}(\cdot)$ 为指示函数。MT 越高,说明算法能够完整跟踪更多目标轨迹。

区别于传统单星多目标跟踪,多卫星协同观测不仅需要保证单个观测片段内轨迹的连续性,还需要实现不同片段之间目标身份的一致关联。为分析多卫星协同观测场景下多目标跟踪的主要瓶颈,本文分别在完整序列和独立片段两种评估方式下计算上述指标:前者同时评估片段内跟踪与跨片段身份关联,后者仅评估片段内跟踪。

3.3 实验结果

表2给出了12种主流多目标跟踪算法在BatonTrack1.0上的对比结果。可以观察到,TBD范式方法整体优于JDT范式方法,其主要原因在于,TBD范式方法将目标检测与轨迹关联解耦,先获得相对稳

定的检测结果,再基于空间位置关系、运动预测结果和外观特征相似性完成目标匹配与轨迹更新。在BatonTrack1.0中,即使相邻片段之间存在时间空窗,TBD方法仍可在后一片段中重新利用检测结果建立候选目标集合,并通过关联策略完成跨片段匹配,因此对连续帧状态传播的依赖相对较弱。相比之下,JDT范式方法通常在统一网络中同时完成目标检测和身份关联,其关联过程依赖网络学习到的目标表示、中心偏移、目标查询或历史记忆状态。这类方法在连续视频中能够利用帧间状态传播实现目标关联,但BatonTrack1.0中相邻视频片段之

表 2 BatonTrack1.0 数据集上的基准结果
Table 2 Baseline Results on BatonTrack1.0

方法	发表信息	完整序列				独立片段			
		MOTA ↑	IDF1 ↑	IDS ↓	MT ↑	MOTA ↑	IDF1 ↑	IDS ↓	MT ↑
CenterTrack	ECCV'2020	37.6	39.8	689	69	38.3	57.3	654	139
FairMOT	IJCV'2021	54.8	55.5	128	136	55.1	77.7	48	212
MOTR	ECCV'2022	14.5	40.2	36	91	14.7	57.7	2	163
TGraM	TGRS'2022	72.7	60.7	85	159	73.6	86.9	1	241
MeMOTR	ICCV'2023	48.6	52.2	42	116	48.8	73.1	5	197
MP2Net	TGRS'2024	-13.0	29.8	42	65	-12.8	40.0	6	96
ByteTrack	ECCV'2022	90.9	64.2	88	183	91.3	95.7	0	273
MotionTrack	CVPR'2023	90.8	62.3	87	183	91.3	95.7	0	273
MambaTrack	ACM MM'2024	88.5	63.8	145	165	88.9	92.0	64	247
DiffMOT	CVPR'2024	90.5	64.6	92	183	90.9	95.4	4	273
TOPIC	TIP'2025	88.4	63.7	88	182	88.8	94.4	0	272
TrackTrack	CVPR'2025	88.6	64.8	88	182	89.1	94.5	0	272

间存在时间空窗,目标在空窗期间仍会持续运动,跨片段位置和运动状态可能发生明显变化。此时,基于连续帧状态传播的关联假设被削弱,目标查询或历史状态难以准确延续到后一卫星片段,容易导致轨迹中断或目标漏检。因此,JDT范式方法整体表现受限。以MP2Net为例,其核心机制较依赖连续视频中的掩膜传播和运动预测。BatonTrack1.0中相邻片段之间不仅存在时间空窗,还伴随视野偏移和成像几何差异。当利用前一片段的目标状态预测后一片段目标位置时,时间间隔内的持续运动会放大预测误差,导致目标位置估计不准、轨迹延续失

败或错误匹配。同时,飞机目标尺度较小,局部区域定位和掩膜传播对误差较为敏感,进一步增加了跨片段关联难度。对于TBD范式方法,各算法在MOTA指标上的差异相对较小。这是因为本文统一采用相同权重的YOLOX-X作为检测器,并保持一致的推理置信度阈值,使不同TBD方法具有基本一致的检测输入。由于后续关联策略不同,各方法在IDF1和IDS指标上存在一定差异,其中TOPIC和TrackTrack表现相对较优。TOPIC采用并行关联策略,根据目标运动复杂程度在不同关联线索之间进行选择,能够在运动可靠时优先利用运动约束,在运

动不确定时引入外观信息辅助匹配,从而降低单一关联线索失效带来的身份混淆;TrackTrack 更加关注轨迹本身的匹配关系,通过轨迹级信息完成目标关联,而非仅依赖相邻帧检测框的局部匹配,因此能够利用更完整的历史轨迹状态维持目标身份一致性。进一步比较各方法在完整序列和独立片段上的结果可以发现,相较于完整序列评估,独立片段评估下受检测性能影响较大的 MOTA 指标差异较小,而反映身份关联性能的 IDF1 普遍提高、IDS 显著减少。以 TrackTrack 为例,其 IDS 由完整序列评估下的 88 降至独立片段评估下的 0。这表明,跨卫星片段间的身份关联是多卫星协同观测多目标跟踪的主要难点。

图 3 展示了 TOPIC 和 TrackTrack 在图 2 所示两组多卫星接续观测序列上的跟踪结果。每幅图最左列为目标轨迹可视化结果,其余列展示了不同卫星观测片段首帧和末帧的跟踪结果。同一序列中,相同目标身份采用相同颜色表示。可以观察到, TOPIC 和 TrackTrack 在单个卫星观测片段内均能够较好地完成目标定位与跟踪,但在片段切换后,跟踪结果中的目标身份与真实标注并不一致。具体而言,同一目标在后一卫星片段中被赋予新的身份编号,导致轨迹在跨片段处发生身份重分配。对于 TOPIC,其关联过程依赖运动状态与外观线索的并行匹配。然而,在相邻片段之间存在时间空窗的条件下,目标在未观测期间持续运动,重新出现时已偏离前一轨迹的预测位置;同时,飞机目标尺度小、纹理弱,外观特征难以提供稳定的身份判别信息,使运动线索和外观线索均难以形成可靠关联。对于 TrackTrack,其主要利用轨迹级历史信息进行目标关联,但跨卫星片段之间的时间空窗会破坏轨迹状态的连续积累;同时,跨卫星视野偏移和成像几何差异进一步降低了历史轨迹对后一片段目标匹配的约束能力。因此,两种方法虽然在短时连续观测片段内具有较好的跟踪性能,但在多卫星接续观测场景下难以保持跨片段目标身份一致性。

4 现存挑战与未来展望

尽管现有多目标跟踪方法在单个卫星观测片段内能够取得较好的目标定位与短时跟踪效果,但在多卫星接续观测场景下仍面临显著挑战。首先,相

邻卫星观测片段之间存在时间空窗,目标在未观测期间持续运动,导致基于短时运动连续性假设的轨迹预测难以保持稳定。其次,不同卫星之间存在视角变化、成像几何差异及空间覆盖偏移,使目标在图像平面中的位置分布发生明显变化,进一步增加了跨片段身份关联的复杂性。此外,遥感视频中的飞机目标尺度较小、纹理细节有限,外观特征缺乏稳定判别能力。当多个外观相似目标同时分布于滑行道、跑道和停机坪等区域时,仅依赖局部运动信息或外观相似性的关联策略容易产生身份混淆。因此,如何在时间非连续、视角异构和目标弱外观表征条件下,实现稳定可靠的跨片段身份接续,仍是多卫星接续观测多目标跟踪亟待解决的关键问题。未来研究可从以下几个方面进一步推进:

一是可进一步融合场景先验与目标语义信息,提升跨片段身份关联能力。不同观测场景中目标运动通常受到地理环境、空间结构、通行规则和任务背景等因素约束,不同类型目标在尺度、形态、运动模式和行为特征上也存在一定差异。未来可将场景结构、地理知识、目标属性和任务先验等高层语义信息与视觉观测结果进行联合建模,构建面向场景认知的跨片段关联框架,从而缩小候选匹配空间,降低外观相似目标之间的身份混淆,提高复杂场景下的身份判别能力。

二是可进一步研究长时间跨度轨迹预测与目标行为意图推理,提高时间空窗条件下的轨迹延续能力。由于相邻卫星观测片段之间存在非连续观测间隔,目标在未观测期间可能发生位置、速度和运动方向的显著变化,传统基于短时连续性假设的运动模型难以准确估计目标状态。未来可综合利用目标历史轨迹、场景约束、运动规律和潜在行为模式,对目标在未观测期间的可能运动状态进行长期预测,并进一步推断其潜在运动意图。通过将轨迹预测与行为推理相结合,可为后续观测片段中的身份关联提供更加可靠的时序先验。

三是可进一步引入目标群体关系建模,实现全局一致性的身份关联优化。复杂场景中的多个目标通常并非完全独立,其运动过程可能受到相互避让、队列行进、空间占用和任务交互等关系影响。未来可从单目标独立建模扩展到多目标关系建模,充分利用目标之间的相对位置、运动趋势、时空交互和群体行为一致性约束,从整体层面对跨片段身份关联

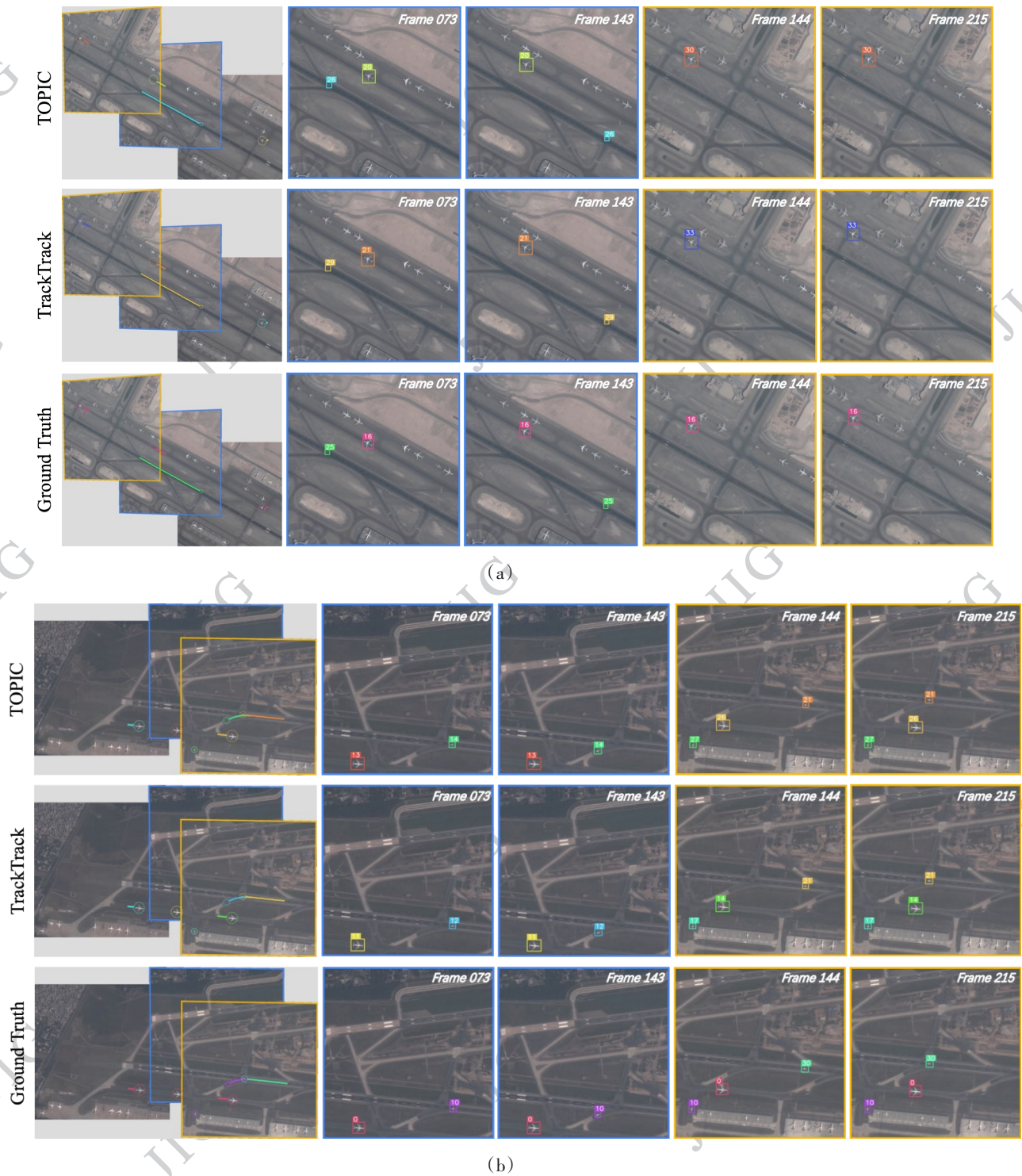


图3 TOPIC与TrackTrack在BatonTrack1.0中两组多卫星协同观测序列上的跟踪结果可视化

Fig. 3 Visualization of tracking results of TOPIC and TrackTrack on two examples of multi-satellite cooperative observation sequences from BatonTrack1.0

进行联合优化。相比仅依赖局部目标特征进行匹配,群体关系建模有助于抑制局部误关联导致的身份漂移,提高密集复杂场景中的整体关联稳定性。

此外,需关注真实多星数据稀缺条件下的模型

训练与泛化问题。真实多星接续观测数据获取困难、数量有限。本文将 BatonTrack1.0 全部作为测试集,并基于单星卫星视频跟踪数据构建仿真训练集。仿真数据通过间隔截取、裁剪、平移和几何变换模拟

时空窗、视野偏移与成像几何差异,但难以充分表征真实跨卫星成像差异及长时间空窗下的身份关联不确定性。未来一方面可研究面向真实多星成像过程的生成式仿真方法,提高仿真数据对复杂观测条件的表征能力;另一方面可探索基于零样本思维链推理的跨片段身份关联方法,引导预训练多模态模型综合时空关系、运动状态和场景约束进行分步推理,降低对真实多星训练数据的依赖。

随着遥感卫星星座规模不断扩大,多卫星协同观测正逐步向多平台、多传感器、多时相联合观测发展。未来可进一步融合光学、SAR(Synthetic Aperture Radar)、红外等多源遥感数据,以及卫星轨道参数、姿态信息、地图数据和地理知识等多维信息,构建统一的时空感知与目标关联框架,实现跨平台、跨传感器、跨时空尺度的持续目标跟踪。同时,结合遥感基础模型、多模态大模型和智能体等新兴技术,探索具备长期记忆、知识推理和自主决策能力的遥感多目标跟踪,进一步突破传统跟踪方法对局部视觉特征和短时运动信息的依赖,实现更加鲁棒和持续的多卫星协同目标感知。

5 结 论

本文针对单颗卫星凝视时间有限、难以支撑长时间持续跟踪的问题,研究了多卫星协同观测条件下的长时序多目标跟踪任务,构建了首个面向真实多卫星协同观测的多目标跟踪基准数据集 BatonTrack1.0。该数据集由吉林一号多颗视频卫星对同一区域获取的真实观测片段构成。区别于单星连续观测数据,相邻片段之间存在时空窗,并伴随视角变化和成像几何差异。数据集共构建8个多目标跟踪序列,标注192架飞机目标,涵盖静止、起飞、降落、进出港滑行和避让等多种典型运行状态,为跨卫星目标关联与长时间连续跟踪研究提供了真实数据支撑。

考虑到BatonTrack1.0来源于真实多卫星接续成像,数据规模有限,本文基于SAT-MTB数据集补充构建了仿真多卫星视频数据集作为训练集,并将BatonTrack1.0作为真实评测基准。在此基础上,本文系统评测了12种主流多目标跟踪方法,分别从完整序列和独立片段两个层面,对其在真实多卫星观测场景下的性能进行了统一评估。实验结果表明,

现有方法在单个卫星观测片段内能够取得较好的目标定位与短时跟踪效果,但在跨卫星片段切换时仍难以稳定保持目标身份一致性。其主要原因在于,现有多目标跟踪方法通常依赖运动连续性或外观相似性完成目标关联。然而,目标在时空窗内持续运动,使基于短时运动连续性的状态预测可靠性下降;同时,卫星视频中的飞机目标尺度小、纹理信息有限,难以提取具有稳定判别能力的外观特征。未来可结合场景先验、目标语义信息、长时轨迹预测和群体关系建模等,提升多卫星观测条件下动态目标持续感知的可靠性与稳定性。

后续工作将进一步扩充真实多卫星协同观测序列的规模,覆盖更广泛的观测场景与目标类别,并丰富不同时空窗、视角变化和成像几何差异条件下的观测样本。

参考文献(References)

- Liu F, Wang J, Jiao L, Zhang J, Wang H, Li S, et al. 2025. Remote sensing video tracking: Current status, challenges and future. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 18: 14338-14367 [DOI: 10.1109/JSTARS.2025.3573572]
- Li S, Zhou Z, Zhao M, Yang J, Guo W, Lv Y, et al. 2023. A multitask benchmark dataset for satellite video: Object detection, tracking, and segmentation. IEEE Transactions on Geoscience and Remote Sensing, 61: 1-21 [DOI: 10.1109/TGRS.2023.3278075]
- He Q, Sun X, Yan Z, Li B, and Fu K. 2022. Multi-object tracking in satellite videos with graph-based multitask modeling. IEEE Transactions on Geoscience and Remote Sensing, 60: 1-13 [DOI: 10.1109/TGRS.2022.3152250]
- Xiao C, An W, Zhang Y, Su Z, Li M, Sheng W, et al. 2024. Highly efficient and unsupervised framework for moving object detection in satellite videos. IEEE Transactions on Pattern Analysis and Machine Intelligence, 46 (12): 11532-11539 [DOI: 10.1109/TPAMI.2024.3409824]
- Lu W, Sun B, Li S and Li X. 2025. FMA-Net: Flow-driven motion-aware network for multi-object tracking in satellite videos. IEEE Transactions on Geoscience and Remote Sensing, 2025, 63: 1-15 [DOI: 10.1109/TGRS.2025.3624205]
- Xiao Z, Xu F, Liu C, Yang W and Xia G-S. 2026. SPAX: Fully sparse framework with hierarchical spatio-temporal fusion for moving object tracking in satellite videos. IEEE Transactions on Geoscience and Remote Sensing, 64: 1-15 [DOI: 10.1109/TGRS.2025.3643466]
- Li D, Wang M and Wu Q. 2024. Fast, accurate and smart applications

- of intelligent remote sensing satellites. *Advanced Small Satellite Technology*, 1(1): 1-9 (李德仁, 王密, 仵倩玉. 2024. 论智能遥感卫星的“快、准、灵”应用服务. 先进小卫星技术(中英文), 1(1): 1-9)
- Kang L, Tian J and Jiang B. 2024. Challenges and research on remote sensing satellite application technology in the Giant Constellation Era. *National Remote Sensing Bulletin*, 28(7): 1658-1666 (康利鸿, 田菁, 江碧涛. 2024. 巨星座时代遥感卫星应用技术挑战与思考. 遥感学报, 28(7): 1658-1666) [DOI: 10.11834/jrs.20233248]
- Li Z, Yue P and Zhang C. 2025. CVTOT: A cross-video tiny object tracking method and dataset for Earth observation videos. *IEEE Transactions on Geoscience and Remote Sensing*, 63: 1-19 [DOI: 10.1109/TGRS.2025.3630601]
- Zhao M, Li S, Xuan S, Kou L, Gong S and Zhou Z. 2022. SatSOT: A benchmark dataset for satellite video single object tracking. *IEEE Transactions on Geoscience and Remote Sensing*, 60: 1-11 [DOI: 10.1109/TGRS.2022.3140809]
- Li Y, Jiao L, Huang Z, Zhang X, Zhang R, Song X, et al. 2022. Deep learning-based object tracking in satellite videos: A comprehensive survey with a new dataset. *IEEE Geoscience and Remote Sensing Magazine*, 10(4): 181-212 [DOI: 10.1109/MGRS.2022.3198643]
- Chen Y, Tang Y, Xiao Y, Yuan Q, Zhang Y, Liu F, et al. 2024. Satellite video single object tracking: A systematic review and an oriented object tracking benchmark. *ISPRS Journal of Photogrammetry and Remote Sensing*, 210: 212-240 [DOI: 10.1016/j.isprsjprs.2024.03.013]
- Yin Q, Hu Q, Liu H, Zhang F, Wang, Y, Lin Z, et al. 2021. Detecting and tracking small and dense moving objects in satellite videos: A benchmark. *IEEE Transactions on Geoscience and Remote Sensing*, 60: 1-18 [DOI: 10.1109/TGRS.2021.3130436]
- Bewley A, Ge Z, Ott L, Ramos F, and Upcroft B. 2016. Simple online and realtime tracking// *Proceedings of the International Conference on Image Processing*. Phoenix: IEEE: 3464-3468 [DOI: 10.1109/ICIP.2016.7533003]
- Wojke N, Bewley A, and Paulus D. 2017. Simple online and realtime tracking with a deep association metric// *Proceedings of the International Conference on Image Processing*. Beijing: IEEE: 3645-3649 [DOI: 10.1109/ICIP.2017.8296962]
- Zhang Y, Sun P, Jiang Y, Yu D, Yuan Z, Luo P, et al. 2022. ByteTrack: multi-object tracking by associating every detection box// *Proceedings of the European conference on Computer Vision*. Tel Aviv: Springer: 1-21 [DOI: 10.1007/978-3-031-20047-2_1]
- Qin Z, Zhou S, Wang L, Duan J, Hua G, and Tang W. 2023. MotionTrack: learning robust short-term and long-term motions for multi-object tracking// *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Vancouver: IEEE: 17939-17948 [DOI: 10.1109/CVPR52729.2023.01720]
- Lv X, Li S, Chen Y, Wang K, and Yang L. 2024. DiffMOT: a real-time diffusion-based multiple object tracker with non-linear prediction// *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Seattle: IEEE: 19321-19330 [DOI: 10.1109/CVPR52733.2024.01827]
- Li S, Wang X, Wang H, Wang X, and Liu D. 2024. MambaTrack: a simple baseline for multiple object tracking with state space model// *Proceedings of the ACM International Conference on Multimedia*. Melbourne: ACM: 4082 - 4091 [DOI: 10.1145/3664647.3680944]
- Ren H, Han S, Ding H, Zhang Z, Wang H, and Wang F. 2023. Focus on details: online multi-object tracking with diverse fine-grained representation// *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Vancouver: IEEE: 11289-11298 [DOI: 10.1109/CVPR52729.2023.01086]
- Cao X, Zheng Y, Yao Y, Qin H, Cao X, and Guo S. 2025. TOPIC: a parallel association paradigm for multi-object tracking under complex motions and diverse scenes. *IEEE Transactions on Image Processing*, 34: 743-758 [DOI: 10.1109/TIP.2025.3526066]
- Zhou X, Koltun V, and Krähenbühl P. 2020. Tracking objects as points// *Proceedings of the European Conference on Computer Vision*. Glasgow: Springer: 474-490 [DOI: 10.1007/978-3-030-58548-8_28]
- Zeng F, Dong B, Wang T, Chen X, Zhang Y, and Wei Y. 2022. MOTR: end-to-end multiple-object tracking with transformer// *Proceedings of the European Conference on Computer Vision*. Tel Aviv: Springer: 659-675 [DOI: 10.1007/978-3-031-20047-2_38]
- Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A N, et al. 2017. Attention is all you need// *Proceedings of the International Conference on Neural Information Processing Systems*. Long Beach: Curran Associates: 5998-6008 [DOI: 10.5555/3295222.3295349]
- Gao R, Wang L, Li B, and Wu J. 2023. MeMOTR: long-term memory-augmented transformer for multi-object tracking// *Proceedings of the IEEE/CVF International Conference on Computer Vision*. Paris: IEEE: 9901-9910 [DOI: 10.1109/ICCV51070.2023.00912]
- Segu M, Piccinelli L, Li S, Yang Y H, Van Gool L, and Schiele B. 2025. Samba: synchronized set-of-sequences modeling for multiple object tracking// *Proceedings of the International Conference on Learning Representations*. Singapore: 30057-30070
- Liu Z, Shang Y, Li T, Chen G, Wang Y, Hu Q, et al. 2023. Robust multi-drone multi-target tracking to resolve target occlusion: a benchmark. *IEEE Transactions on Multimedia*, 25: 1462-1476 [DOI: 10.1109/TMM.2023.3234822]
- Nguyen T T, Nguyen H H, Sartipi M, and Fisichella M. 2024. Multi-vehicle multi-camera tracking with graph-based tracklet features. *IEEE Transactions on Multimedia*, 26: 972-983 [DOI: 10.1109/TMM.2023.3274369]
- Specker A. 2024. OCMCTrack: online multi-target multi-camera tracking with corrective matching cascade// *Proceedings of the IEEE/*

- CVF Conference on Computer Vision and Pattern Recognition Workshops. Seattle: IEEE: 7236-7244 [DOI: 10.1109/CVPRW63382.2024.00719]
- Cheng C C, Qiu M X, Chiang C K, and Lai S H. 2023. ReST: a reconfigurable spatial-temporal graph model for multi-camera multi-object tracking// Proceedings of the IEEE/CVF International Conference on Computer Vision. Paris: IEEE: 10051-10060 [DOI: 10.1109/ICCV51070.2023.00924]
- Tang Z, Naphade M, Liu M Y, Yang X, Birchfield S, Wang S, et al. 2019. CityFlow: a city-scale benchmark for multi-target multi-camera vehicle tracking and re-identification// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE: 8789-8798 [DOI: 10.1109/CVPR.2019.00900]
- Huang H W, Yang C Y, Jiang Z, Kim P K, Lee K, Kim K, et al. 2023. Enhancing multi-camera people tracking with anchor-guided clustering and spatio-temporal consistency ID re-assignment// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Vancouver: IEEE: 5239-5249 [DOI: 10.1109/CVPRW59228.2023.00552]
- Zhang Y, Wang C, Wang X, Zeng W and Liu W. 2021. FairMOT: On the fairness of detection and re-identification in multiple object tracking. International Journal of Computer Vision, 129 (11) : 3069-3087 [DOI: 10.1007/s11263-021-01513-4]
- Zhao M, Li S, Wang H, Yang J, Sun Y and Gu Y. 2024. MP2Net: Mask propagation and motion prediction network for multiobject tracking in satellite videos. IEEE Transactions on Geoscience and Remote Sensing, 62: 1-15 [DOI: 10.1109/TGRS.2024.3385406]
- Shim K, Ko K, Yang Y and Kim C. 2025. Focusing on tracks for online multi-object tracking// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE: 11687-11696 [DOI: 10.1109/CVPR52734.2025.01091]
- Ge Z, Liu S, Wang F, Li Z and Sun J. 2022. YOLOX: Exceeding YOLO Series in 2021// Proceedings of the European Conference on Computer Vision. Tel Aviv: Springer: 378 - 394 [DOI: 10.1007/978-3-031-19815-1_22]
- He L, Liao X, Liu W, Liu X, Cheng P and Mei T. 2023. FastReID: A PyTorch Toolbox for General Instance Re-Identification// Proceedings of the ACM International Conference on Multimedia. Ottawa: ACM: 6748 - 6757 [DOI: 10.1145/3581783.3613460]
- Gao T, He Y, Ma X, Lu Z, Peng S and Liu Y. 2025. Single target tracking in satellite videos. Journal of Image and Graphics, 30 (12) : 3740-3759 (高桃峰, 何银鑫, 马学良, 卢自来, 彭世勇, 刘洋. 2025. 卫星视频单目标跟踪综述. 中国图象图形学报, 30(12): 3740-3759) [DOI: 10.11834/jig.240750]
- Zhi Y, Jiang Y, Yang Z, Chen Y, Hao W, Ma M, et al. 2026. Development status and prospects of pretrained foundation models for remote sensing imagery. Journal of Image and Graphics, 31 (4) : 0973-0986 (支元杰, 姜艺伟, 杨知, 陈奕州, 郝文魁, 马明阳, 等. 2026. 面向遥感图像的预训练基础模型发展现状与展望. 中国图象图形学报, 31(4): 0973-0986) [DOI: 10.11834/jig.250424]

作者简介

徐芳, 女, 博士后, 研究方向为多源遥感时空数据智能解译。

E-mail: xufang@whu.edu.cn

夏桂松, 通信作者, 男, 教授, 主要研究方向为计算机视觉、智能无人系统、遥感智能信息处理。E-mail: guisong.xia@whu.edu.cn

徐健, 男, 博士研究生, 主要研究方向为遥感卫星视频目标跟踪方法。E-mail: jian_xu@whu.edu.cn

杨文, 男, 教授, 主要研究方向为图像处理与计算机视觉。Email: yangwen@whu.edu.cn