

中图分类号: TP18; TP391 文献标识码: A 文章编号: 1006-8961(2025)12-3740-20

论文引用格式: Gao T F, He Y X, Ma X L, Lu Z L, Peng S Y 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]

卫星视频单目标跟踪综述

高桃峰¹, 何银鑫^{1,2*}, 马学良¹, 卢自来¹, 彭世勇², 刘洋^{1,2}

1. 四川省交通勘察设计研究院有限公司, 成都 610017; 2. 西南交通大学地球科学与工程学院, 成都 611756

摘要: 基于卫星视频的单目标跟踪受到广泛关注, 在军事和民用领域发挥着重要作用, 可用于城市尺度下的灾害救援、治安防范以及突发事件监控等, 但受目标尺寸小、相似目标干扰、运动模糊和背景复杂等因素影响, 面临诸多挑战。为便于更多学者对该领域进行探索, 本文对卫星视频单目标跟踪研究现状进行梳理和分析。首先, 阐述该领域的挑战与优势。视频卫星视野范围大, 车辆等目标在卫星视频中仅占几个或十几个像素, 特征和纹理稀少, 此外卫星视频中目标数量较多, 目标与干扰物间的可区分性差, 相似性较高, 存在目标的快速移动或卫星平台的抖动导致目标模糊的情况, 以及目标存在于复杂背景中, 当运动目标本身不明显、背景信息比目标信息更加突出时, 会导致跟踪失败。但相对于地面或低空视频, 基于卫星视频的目标跟踪又有一定的优势, 比如: 目标的外部条件, 如摄像机视角等相对稳定、卫星视频中的目标多为刚性目标, 在跟踪过程中很少发生严重变形、目标的运动状态简单明了, 运动轨迹多为直线或平滑的曲线。其次, 总结典型的跟踪方法。包括基于生成式、基于相关滤波和基于深度学习的方法, 介绍解决特征稀少、遮挡、相似目标干扰等问题的相关研究。然后, 归纳现有卫星视频单目标跟踪数据集和常用性能评价指标。卫星视频目标跟踪数据集主要有 XDU-BDSTU、VISO(video satellite objects)、SatSOT、OOTB(the oriented object tracking benchmark)。其中 VISO 数据集规模最大, 包括训练集和测试集, XDU-BDSTU 数据集的影像幅宽大, 适用于长时跟踪, OOTB 数据集的目标标注为旋转框, 更加贴合目标。性能评价指标主要有精确率、成功率和帧率。此外, 在自制的测试集上对典型单目标跟踪方法进行性能评测与分析。实验结果表明, 运动估计、时序信息以及背景信息的利用对卫星视频目标跟踪任务较为重要。最后, 从统一跟踪结果评价指标、更大规模的高质量卫星视频目标跟踪数据集, 以及更适合于卫星视频单目标跟踪的模型、长时跟踪几个方面展望了卫星视频单目标跟踪算法的未来发展趋势。

关键词: 卫星视频; 单目标跟踪; 相关滤波; 深度学习; 吉林 1 号卫星

Single target tracking in satellite videos

Gao Taofeng¹, He Yinxi^{1,2*}, Ma Xueliang¹, Lu Zilai¹, Peng Shiyong², Liu Yang^{1,2}

1. Sichuan Communication Surveying and Design Institute Co., Ltd., Chengdu 610017, China;

2. Faculty of Geosciences and Engineering, Southwest Jiaotong University, Chengdu 611756, China

Abstract: In recent years, single-object tracking in satellite videos has gained substantial attention and plays a pivotal role in military and civilian domains. This tracking has found applications in urban-scale disaster relief, public security surveillance, and the monitoring of emergency events, among others. However, due to a combination of factors, such as small tar-

收稿日期: 2024-12-18; 修回日期: 2025-05-05; 预印本日期: 2025-05-12

* 通信作者: 何银鑫 heyinxi@my.swjtu.edu.cn

基金项目: 四川省科技厅重点研发项目(2021YFS0334); 四川省交通勘察设计研究院有限公司科技项目(232022015)

Supported by: Key R&D Projects of Sichuan Provincial Science and Technology Department (2021YFS0334); Science and Technology Projects of Sichuan Province Transportation Survey and Design Institute Co., Ltd. (232022015)

get size, interference from similar targets, motion blur, and complex backgrounds, single-object tracking in satellite videos presents numerous challenges. Aiming to promote further exploration in this domain by scholars domestically and internationally, this paper comprehensively reviews and critically analyzes the current state of the art in satellite video-based single-object tracking. Considering challenges and advantages, video satellites offer an expansive field of view. Targets such as vehicles typically occupy only a few to a dozen pixels in satellite videos, with limited distinguishing features or textures. Additionally, satellite videos contain many targets, and the distinguishability between the targets of interest and interfering objects is poor, presenting a high degree of similarity. Moreover, target blurring may occur due to their rapid target movement or satellite platform jitter. When the moving target is inconspicuous and background information overshadows target features, tracking failure is likely to occur. However, compared to ground-based or low-altitude videos, satellite video-based object tracking offers certain advantages. For example, external factors related to the target, such as the camera perspective, are relatively stable, aiding tracking algorithms in maintaining a consistent lock on the target. Most objects in satellite videos are rigid and rarely undergo substantial deformation during tracking. Additionally, the aspect ratios of targets remain approximately consistent across video frames, reducing the potential for algorithmic confusion. The motion of targets is typically straightforward, with trajectories generally following straight lines or smooth curves, enabling the prediction of target positions based on historical motion data. Regarding the development of tracking methods, this paper reviews the evolution of single-object tracking methods for satellite videos and highlights typical tracking paradigms, including generative-based approaches, correlation filter-based methods, and deep learning-based techniques. Deep learning-based tracking methods can be further classified into convolutional neural network (CNN)-based and Transformer-based methods. In contrast to the hand-crafted features employed in correlation filter-based methods, CNNs can extract more comprehensive and robust features, thereby enhancing target tracking performance. In recent years, an increasing number of scholars have applied CNNs to satellite video object tracking tasks. However, when processing high-resolution images, long time-series data, and complex backgrounds, which are common in satellite videos, CNNs exhibit certain limitations. Aiming to address these limitations, Transformers have been gradually introduced into satellite video object tracking. Transformers can capture global spatial information and long-term temporal dependencies, offering a promising alternative for improving tracking accuracy in complex scenarios. Regarding datasets and evaluation metrics, this study compiles existing single-object tracking datasets for satellite videos, along with commonly adopted performance evaluation metrics. Prominent datasets in this field include XDU-BDSTU, video satellite objects (VISO), SatSOT, and the oriented object tracking benchmark (OOTB). Among them, the VISO dataset is the largest in scale, comprising training and test subsets. The XDU-BDSTU dataset features images with a large swath width, making it suitable for long-term tracking tasks. The OOTB dataset provides annotations using rotated bounding boxes, which accurately represents the actual target geometry. The main performance evaluation metrics include precision, success rate, and frame rate, which collectively assess tracking methods in terms of tracking accuracy and speed. Aiming to evaluate the applicability of various tracking algorithms across different scenarios, this paper selects 18 algorithms for performance evaluation and analysis on a self-constructed test set. Experimental results highlight the critical roles of motion estimation, temporal information utilization, and background information exploitation in satellite video object tracking. Specifically, the correlation filter with motion estimation (CFME) algorithm leverages historical motion information of the target to enhance tracking performance, while the Trdimp algorithm incorporates temporal and background information, yielding favorable outcomes. When a vehicle makes a turn, the hand-crafted features employed by the correlation filter-based method CFME lack rotational invariance and are poorly equipped to handle changes in the target's bounding box due to rotation, resulting in suboptimal tracking performance. Conversely, methods such as Trdimp and Trsiam directly estimate the target's bounding box, while approaches such as siamese region proposal network (SiamRPN) and SiamRPN++ predefine anchor boxes with different aspect ratios, effectively addressing the challenge of in-plane rotation. Finally, in terms of future perspectives, this paper outlines the anticipated trajectory of single-object tracking algorithms for satellite videos across several key dimensions: standardizing evaluation metrics for tracking results, developing large-scale and high-quality satellite video object tracking datasets, devising models specifically tailored to satellite video tracking challenges, and enabling robust long-term tracking capabilities. In the domain of general video target tracking, commonly used evaluation metrics include those from the OTB and VOT bench-

marks. For satellite video target tracking, scholars predominantly adopt the precision and success rate metrics defined by the OTB evaluation framework. In the OTB metrics for general videos, the precision threshold is customarily set to 20 pixels, and the success rate is evaluated based on the area under the curve (AUC) of the overlap score. However, in satellite video target tracking, researchers often adopt varying threshold settings, which hinders the objective evaluation of algorithms under a unified standard. Thus, standardizing evaluation metrics for tracking results is essential for the advancement of satellite video single-object tracking. Before the emergence of large-scale test datasets, most studies in satellite video object tracking verified algorithms using only a few targets, which restricted comprehensive algorithm performance assessment. Moreover, the use of different test dataset across studies has further hindered direct comparisons between algorithms. Consequently, the development of large-scale, high-quality satellite video object tracking datasets is urgently needed, not only for effective model training, but also for model testing and performance benchmarking. Future research could benefit from rapidly assimilating the latest advancements in general video object tracking domain and adapting them to the unique characteristics of satellite videos. Given the rich background information and the continuous, linear nature of target motion trajectories between adjacent frames in satellite videos, these priors can be fully leveraged to explore global spatial and temporal information, thereby enhancing tracking accuracy. Furthermore, techniques such as knowledge distillation, network pruning, and neural architecture search hold considerable potential for autonomously constructing streamlined, low-complexity models specifically tailored to satellite video single-object tracking. These approaches can enable high-precision, real-time target tracking under constrained computation resources. In contrast to ground-based surveillance videos, satellite videos offer broad coverage, making it possible to track trajectories across entire urban areas. However, in such large-scale scenarios, multiple challenges, such as occlusion, interference from similar objects, motion blur, illumination variation, and target rotation, often occur simultaneously. Aiming to address the demands of real-world applications, the development of satellite video tracking algorithms capable of simultaneously addressing these challenges is imperative.

Key words: satellite video; single object tracking; correlation filtering; deep learning; Jilin-1 satellite

0 引言

卫星视频数据是视频卫星采用“凝视”方式对某一区域进行连续观测获得的时序影像,包含来自地面的静态和动态信息(张过,2016;Wang等,2020)。单目标跟踪是指给定或检测出视频一帧中某个感兴趣目标的初始状态,估计后续图像序列中目标运动状态的过程(Smeulders等,2014)。与普通视频相比,卫星视频可以提供更大范围的动态监测(朱泽尧等,2021;Feng等,2021),基于卫星视频的单目标跟踪更适用于城市尺度下的灾害救援、治安防范以及突发事件监控等(Shao等,2019a)。随着对地观测技术的发展,SkySat、吉林1号等高分辨率视频卫星的数量逐年增多(张作省等,2019),为目标跟踪在卫星视频领域的发展提供了极大机遇。

与其他计算机视觉任务相比,基于卫星视频的目标跟踪的历史相对较短,卫星视频单目标跟踪的文献综述较少。本文回顾基于卫星视频的单目标跟踪的研究进展,对基于卫星视频的目标跟踪方法进行梳理,包括生成式目标跟踪方法、基于相关滤波的

目标跟踪方法和基于深度学习的目标跟踪方法。同时,对卫星视频数据集和评估指标进行说明,将目标跟踪中的一些典型算法在卫星视频上进行测试和对比分析,并指出现存问题和未来发展方向。

1 卫星视频单目标跟踪的挑战与优势

针对地面或低空视频的目标跟踪算法研究已经取得显著进展(Bertinetto等,2016;Li等,2018a;Dong和Shen,2018;Zhu等,2018;Zhang和Peng,2019;Wang等,2019)。但将这些算法直接应用于卫星视频目标跟踪时效果还有待提升。主要原因是卫星视频自身的特点为目标跟踪带来了较大难度和挑战(张过,2016;赵春晖和刘海燕,2019),如图1所示。

卫星视频目标跟踪的特点与挑战如下:1)影像宽幅大,目标尺寸小,如图1(a)(b)所示,视频卫星视野范围大,某一帧影像通常达到上百万个像素,而车辆等目标在卫星视频的一帧影像中仅占几个或十几个像素,且特征和纹理稀少。2)相似物干扰严重。目标数量较多,目标与干扰物间的可区分性差,相似性较高,如图1(c)所示,红色方框中的车辆为待跟

踪目标,而在很小的范围内,与待跟踪目标出现了数量较多的相似目标,相似性较高,易对待跟踪目标造成干扰。3)运动模糊。目标的快速移动或卫星平台的抖动导致目标模糊,目标在图像中呈现出的状态会发生变化,如图1(d)所示。4)背景复杂。当待跟踪目标位于复杂的背景中,待跟踪目标本身不明显,并且背景信息比目标信息更加突出时,易导致跟踪失败,如图1(e)所示,红色矩形内车辆目标的周围背景较复杂,跟踪算法易发生漂移。

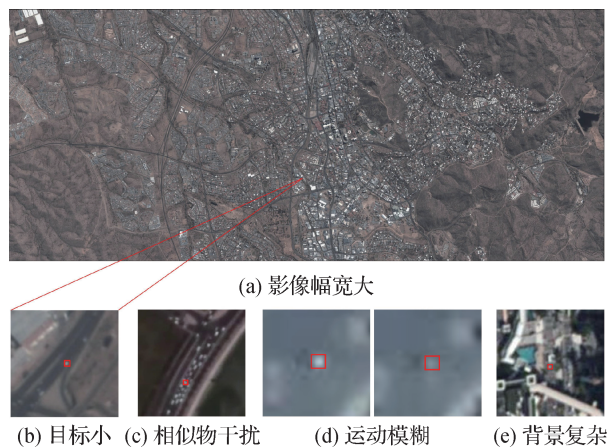


图1 卫星视频目标跟踪的特点与挑战

Fig. 1 Characteristics and challenges of object tracking in satellite videos ((a) large swath width; (b) small target; (c) clutter from similar objects; (d) motion blur; (e) complex background)

相对于地面或低空视频,基于卫星视频的目标跟踪有一定优势(Guo等,2019),充分利用这些优势,克服卫星视频目标跟踪的多重挑战,是卫星视频目标跟踪的关键。图2展示了卫星视频与地面或低空视频在拍摄视角、目标形变以及目标运动状态方面的区别。

从图2可以看出:1)视频卫星拍摄视角相对稳定(如图2(a)所示),使得跟踪算法更易锁定同一目标,也不易出现目标远小近大的情况。2)卫星视频中的目标多为刚性目标(Zhu等,2021),在跟踪过程中很少发生严重变形(如图2(b)所示),在不同的视频帧中,待跟踪目标的长宽基本一致,跟踪过程中不易对算法造成混淆。3)相对于地面视频的行人等目标而言,目标的运动状态简单明了,运动轨迹多为直线或平滑的曲线,有利于根据历史运动信息对目标位置进行预测。

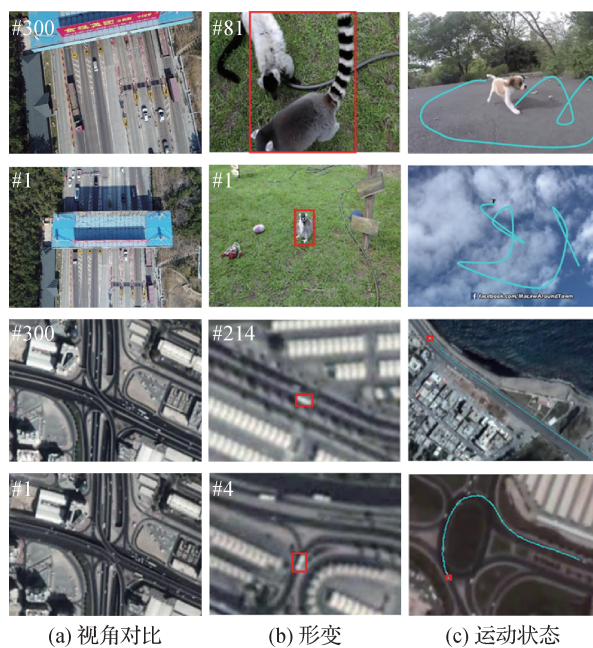


图2 卫星视频目标跟踪的优势

Fig. 2 Advantages of object tracking in satellite videos ((a) viewpoint comparison; (b) rigid/non-rigid deform; (c) motion pattern)

2 卫星视频单目标跟踪方法

根据跟踪策略的不同,现有的目标跟踪方法可分为生成式跟踪方法和判别式跟踪方法。生成式方法首先提取初始视频帧中待跟踪目标的外观特征,然后在后续帧中搜寻与目标最相似的区域作为跟踪结果;判别式方法是建立一个能够判别背景信息和目标信息的分类器,然后基于该分类器判别后续视频帧中的目标和背景,以实现目标跟踪。许多学者结合普通视频目标跟踪方法与卫星视频的特点开展了相关研究。卫星视频单目标跟踪方法发展历程如图3所示。

2.1 基于生成式的卫星视频单目标跟踪

生成式方法如粒子滤波(Isard和Blake,1998)、均值漂移(Comaniciu等,2000)等,首先使用颜色特征、梯度直方图(histogram of oriented gradient, HOG)等手工特征来构造一个对象模板,然后在下一帧中寻找与对象模板相似度最高的区域(Xuan等,2021; Chen和Sui,2019),以此实现目标跟踪。

针对卫星视频中目标尺度小、特征少的问题。王万里(2016)以色度差的倒数作为权值,对Cam-Shift(continuous adaptive mean-shift)算法(Bradski,

1998)进行了改进,抑制背景区域的同时增加对目标本身的关注。吴佳奇等人(2017)则基于卫星视频中的目标在相邻帧中运动平滑的特点,首先使用灰度特征建立基于朴素贝叶斯分类器的条件概率模型,然后采用卡尔曼滤波辅助和优化跟踪结果。为了解决目标与背景相似的问题,Du等人(2019)首先使用

多帧差光流法分离目标与背景,然后将光流场转换为HSV(hue-saturation-value)颜色特征,再进行目标跟踪。

生成式目标跟踪方法通常使用表达能力较弱的手工特征,适用范围小,且生成式方法未考虑背景信息,在面对目标辨识度差等复杂场景时易丢失目标。

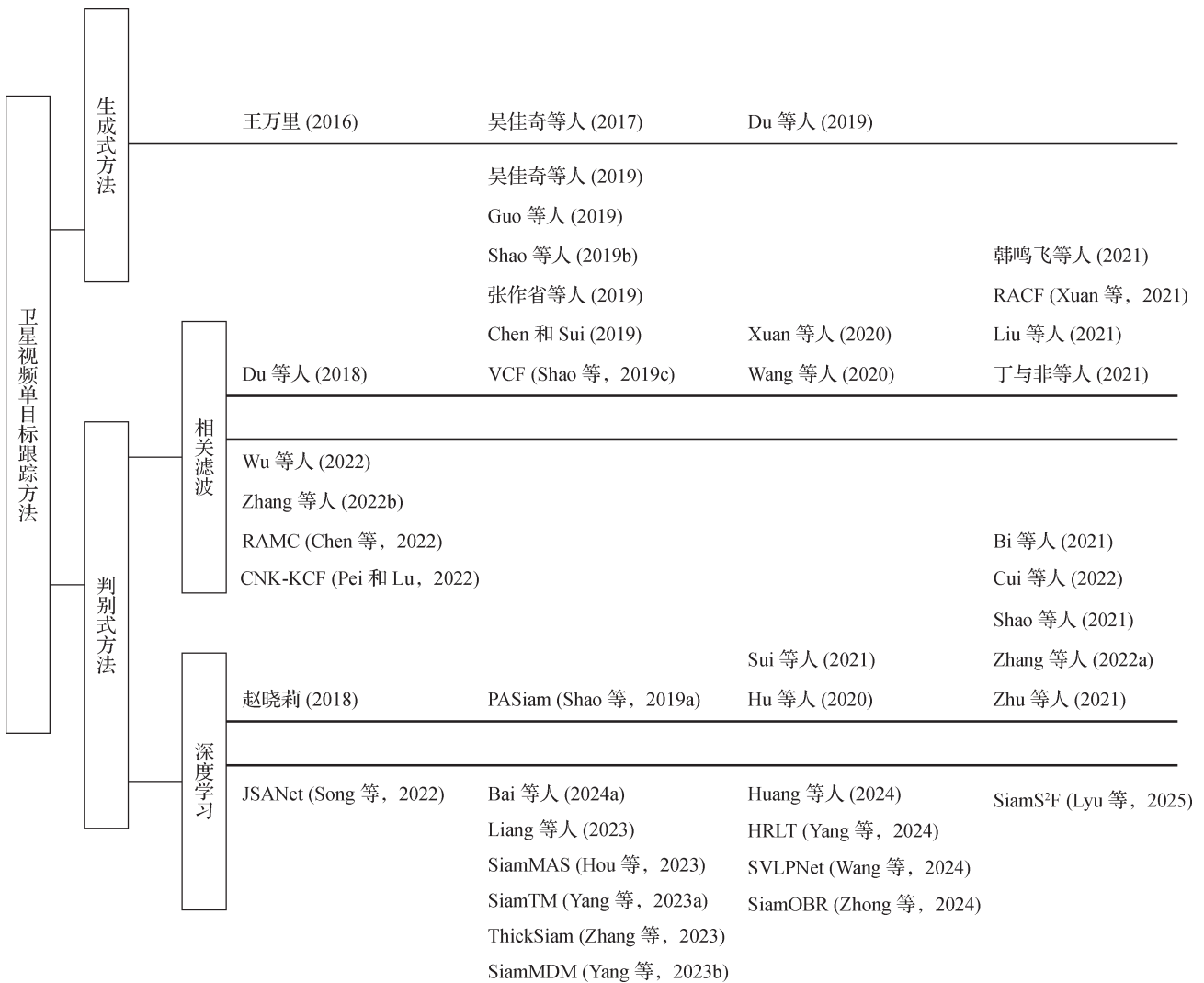


图3 卫星视频单目标跟踪方法发展历程

Fig. 3 Development history of single object tracking methods in satellite video

2.2 基于判别式的卫星视频单目标跟踪

判别式跟踪方法将目标跟踪问题转化为寻求跟踪目标与背景间决策边界的二分类问题,主要包括基于相关滤波的方法和基于深度学习的方法(Li等, 2018b)。

2.2.1 基于相关滤波的卫星视频单目标跟踪

相关滤波方法通过设计一个滤波模板,利用该模板与目标候选区域进行相关运算,在得到的响应

图中,得分最高的位置即为当前帧的目标位置。

针对卫星视频中目标小且与背景相似的问题,Du等人(2018)利用三帧差分法检测出3帧图像之间的变化信息,从背景中突出显示小的运动目标,并与核相关滤波(kernel correlation filter, KCF)(Henriques等, 2015)的跟踪结果相融合以减少模型漂移。Guo等人(2019)、Chen和Sui(2019)使用相关滤波器进行目标跟踪,并对目标跟踪状态进行判断,在跟踪异

常时使用卡尔曼滤波对运动目标的轨迹进行校正。Shao 等人(2019c)提出一种速度相关滤波器(velocity correlation filter, VCF), 首先, 为了更好地区分背景与目标, 利用光流法获取帧间的速度特征进行目标跟踪。此外, 利用相邻帧间目标轨迹呈连续线性的特点, 使用惯性机制减少模型漂移。张作省等人(2019)提出适用于卫星视频的改进时空上下文目标跟踪算法, 实现了实时的高精度目标跟踪。Wang 等人(2020)在 CSK (circulant structure of tracking-by-detection with kernels) 算法(Henriques 等, 2012)中引入 Gabor 滤波器, 增强了目标与背景的对比度, 并提出一个跟踪状态评估指标以避免模型漂移。

仅使用单一特征表示并不能很好地表示卫星视频中的目标, 对此, Shao 等人(2019b)分别利用光流特征和 HOG 特征捕捉目标的变化以及轮廓和纹理信息, 并提出一种自适应融合策略, 以充分利用两种特征的优势。Liu 等人(2021)基于 KCF(Henriques 等, 2015)融合灰度、HOG、局部二值模式(local binary pattern, LBP)特征, 并使用主成分分析来降低融合后的特征维数, 提高目标跟踪的精度与速度。丁与非等人(2021)结合颜色和 HOG 特征, 设计了基于支持向量机(support vector machine, SVM)分类器的重检测模块, 实现长时跟踪。韩鸣飞等人(2021)利用核相关滤波器分别使用原始像素特征和 HOG 特征生成响应图, 通过响应图融合的方式结合两种特征的互补信息, 实现目标跟踪。Pei 和 Lu(2022)提出一种基于 Color-Name 特征和卡尔曼滤波的核相关滤波器 CNK-KCF (kernelized correlation filter based on color-name features and Kalman prediction), 将原始图像映射到 Color-Name 特征空间, 以处理具有多通道颜色特征的图像, 并采用卡尔曼滤波预测目标位置, 根据预测位置确定检测区域, 进而提高跟踪精度。

在卫星视频场景中, 除了目标特征较少的挑战外, 目标旋转以及受遮挡也较为常见。针对目标旋转问题, 吴佳奇等人(2019)将具有旋转不变性的 Hu 不变矩引入相关滤波器, 同时借助卡尔曼滤波对目标运动进行约束, 实现鲁棒跟踪。针对 HOG 特征不能解决旋转问题的缺点, Xuan 等人(2021)提出一种旋转自适应相关滤波(rotation-adaptive correlation filter, RACF)跟踪算法, 通过估计目标旋转角度获得鲁棒特征。Chen 等人(2022)提出一种基于运动约束

的旋转自适应跟踪方法 RAMC (rotation adaptive tracker with motion constraint), 该方法由旋转分支和位移分支组成, 通过将旋转问题分解为位移解决方案, 成功地解耦了旋转和平移运动模式。此外, RAMC 还提出一种内部收缩策略(internal shrinkage strategy, ISS), 用于优化旋转边界框与水平边界框之间的差距。针对遮挡问题, 大部分学者采用“跟踪状态监测 + 轨迹预测”的策略(Wang 等, 2020; Guo 等, 2019; Xuan 等, 2020), 如 Xuan 等人(2020)在核相关滤波算法的基础上, 结合卡尔曼滤波算法和运动轨迹平均算法进行运动估计, 解决遮挡问题的同时还减轻了相关滤波算法的边界效应。Wu 等人(2022)提出一种基于道路信息的改进核相关滤波算法, 通过结合响应图峰值与平均峰值能量重构跟踪置信度模块实现遮挡判别, 并自适应调整卡尔曼滤波的误差, 提高目标位置预测的精度, 同时提出一种以道路信息为辅助的目标搜索策略, 减小了目标消失后的搜索范围, 可在目标被遮挡后快速准确地定位目标。Zhang 等人(2022b)提出一种结合多特征相关滤波器与运动估计的卫星视频目标跟踪算法, 其中运动估计模块结合了卡尔曼滤波器和惯性机制, 在一定程度上解决遮挡问题, 该方法还融合了方向梯度直方图(HOG)和光流(optical flow, OF)特征, 并引入抗干扰机制减弱背景噪声的影响, 提升了对快速旋转和遮挡目标的跟踪精度。

相关滤波算法将计算转换到频域, 利用循环矩阵可以在频域对角化的性质, 大大减少了运算量, 显著提高目标跟踪的速度。部分学者还通过融合多种特征或结合运动估计进一步提高跟踪精度。但由于同样使用手工特征, 在目标特征难以有效提取和描述的复杂情况下, 基于相关滤波的跟踪算法鲁棒性仍然不高。

2.2.2 基于深度学习的卫星视频单目标跟踪

相比于相关滤波方法使用的手工特征, 卷积神经网络(convolutional neural network, CNN)能提取更丰富、更鲁棒的特征, 使得目标跟踪性能获得进一步的提升, 近年来越来越多学者将卷积神经网络应用到卫星视频目标跟踪任务中。

针对卫星视频目标跟踪精度不高的问题, 赵晓莉(2018)在 SiamFC (siamese fully convolutional) (Bertinetto 等, 2016)的基础上引入了模板更新策略, 并使用无人机视频训练网络, 使模型更具有针对性。为

了弥补单一特征的不足, Hu 等人(2020)使用神经网络分别提取外观和运动特征,通过自适应权重进行结合,合理利用二者之间的互补信息,提高跟踪器在卫星视频中的判别能力。针对复杂背景和目标旋转时跟踪精度不高的问题, Bi 等人(2021)基于 SiamFC,首先在 AlexNet 结构中引入了八阶卷积,有效地提高网络的特征表示能力,其次引入了一个可变角度自适应模块来应对目标旋转挑战。Zhu 等人(2021)基于 SiamRPN++ (Li 等, 2019)进行目标跟踪,并提出一种帧间差分质心惯性运动模型以减少模型漂移。Feng 等人(2021)采用帧差法获得物体的运动信息,再将其与外观特征融合,进一步提升了 SiamRPN++ 对小目标的跟踪精度。为了充分学习目标的运动和背景, Zhang 等人(2022a)提出一种基于全卷积网络的预测网络,根据目标之前的运动预测目标在下一帧每个像素中的概率,并引入一种分割方法提取道路背景信息,进一步约束目标可行区域。为了应对卫星视频中目标特征稀少和背景噪声的挑战, Song 等人(2022)提出一种联合孪生注意力感知网络(joint siamese attention-aware network, JSANet)。该网络集成了自注意力和交叉注意力模块,以增强跟踪性能。自注意力模块专注于得分较高的通道,并对空间信息进行转换,实现对关键区域的精准定位。而交叉注意力模块则通过跨通道注意力机制,聚合目标模板与搜索图像之间的有意义上下文依赖关系,有效地挖掘出具有时间关联的对象间的相关性。提高对小目标的识别精度,并增强了在复杂背景下的鲁棒性。Zhang 等人(2023)提出的 Thick-Siam 通过增加残差模块和下采样残差模块的通道数,并裁剪特征图,来丰富 CNN 中语义特征表示,从而提升网络表征能力。该方法还利用卡尔曼滤波同步预测和更新目标的中心点坐标以及目标框大小,进一步提高目标跟踪精度。然而,在目标发生转弯且目标与背景区分度较低的复杂场景下,模型可能出现漂移现象。Yang 等人(2023b)提出一种名为 SiamMDM 的改进孪生跟踪网络,该网络结合了多响应图融合和动态模板更新策略,以实现卫星视频中小目标的精确跟踪。通过融合浅层和深层响应图, SiamMDM 能够更准确地捕捉目标特征,提升跟踪精度。同时,动态模板更新策略有效应对了目标外观的变化,减少了对初始帧的高度依赖。此外,基于卫星视频中目标在短时间内运动近似线性和均匀

的特点,引入了一个评分引导的运动轨迹预测模型,以解决目标完全遮挡时的连续跟踪问题。Yang 等人(2023a)从特征增强和特征匹配策略的角度出发提出 SiamTM,首先提出目标信息增强模块,从水平和垂直维度捕获方向和位置信息,并获取全局通道信息,再将位置信息嵌入到通道注意力中,以增强小目标的特征表达。其次,设计了一个多级匹配模块,将粗粒度的语义抽象信息与细粒度的位置详细信息相结合,有效利用模板信息在搜索区域中准确定位目标,从而提高网络在各种复杂场景下对目标的持续跟踪性能。Zhong 等人(2024)提出一种基于在线背景判别学习和边界框优化的跟踪方法 SiamOBR (siamese network with online background discriminative learning and bounding box optimization)。该方法基于孪生网络架构,引入了在线背景判别学习模块,通过动态更新模板来增强网络区分目标的能力,同时采用边界框二次优化策略提高目标跟踪的精度。此外,为了解决标记框的中心点可能与目标不完全对齐的问题,在训练模型时采用概率回归代替置信回归,并利用交叉熵损失函数纠正小目标的标记错误。Wang 等人(2024)提出的 SVLPNet 通过利用位置提示、改进的特征提取器以及边界框优化模块以应对卫星视频中目标尺寸小、空间分辨率低、外观和纹理信息有限等挑战,并利用运动特征应对光照变化的情况,从而提高跟踪器的整体稳健性。

为了缓解特征图尺寸随着网络的加深而减小、基于深度学习的跟踪器通常需要大量数据进行训练的问题, Sui 等人(2021)设计了一个轻量级网络,使输出的响应图大小与输入图像相同,同时使用迁移学习的策略获取网络参数,仅需初始标注帧进行模型的训练。Shao 等人(2021)将高分辨率网络 HRNet (high-resolution net) (Sun 等, 2019)与 SiamRPN (siamese region proposal network) (Li 等, 2018a)相结合,成功地设计了一种轻量级的并行结构,保证了卫星视频小目标跟踪的精度与速度。Liang 等人(2023)提出一种基于显著性增强机制的轻量级卫星视频目标跟踪方法,通过引入三元组注意力模块增强目标显著性,提高跟踪精度。同时替换传统卷积运算为简单的线性运算,提升网络速度。

针对遮挡问题, Shao 等人(2019a)提出一种基于注意预测的孪生网络(predicting attention inspired siamese network, PASiam),首先构造全卷积孪生网

络以获得目标的细粒度外观特征。此外,采用高斯混合模型检测目标的运动状态,利用卡尔曼滤波器预测和校正目标的位置。此外,越来越多学者基于卫星视频数据的优势,利用时空上下文信息辅助目标跟踪,以抵抗背景、相似目标和遮挡物的干扰,Cui等人(2022)采用深度强化学习的框架进行目标跟踪,使用连续帧之间的时空上下文、网络学习的物体外观模型以及运动向量解决遮挡问题。Hou等人(2023)提出 SiamMAS (siamese multi-scale adaptive search network),该方法通过引入多尺度交叉相关技术,获取更具区分性的特征表示,从而提高目标定位和分类的准确性。同时,结合卡尔曼滤波器与分区搜索策略的自适应搜索模块进一步增强了对目标运动估计的能力。Huang等人(2024)提出一种新型的基于图关联和运动感知的跟踪器(graph association motion-aware tracker, GAMO),利用高斯运动估计器预测目标运动,结合概率采样聚焦于目标最可能出现的区域,并通过图关联模块分析目标与周围物体的空间关系,从而提升了在遮挡和相似干扰场景下的跟踪精度。Lyu等人(2025)提出一种名为 SiamS²F (siamese spectral-spatial-frame)的卫星视频单目标跟踪方法,旨在解决卫星视频中目标尺寸小、复杂遮挡及相似物体干扰等挑战。该方法首先使用孪生骨干网络和通道空间注意力模块提取视频的深层特征。其次,提出一种光谱—空间—帧模块,其中基于图注意力层融合主干的输出特征,然后通过设计的多帧交互注意力机制增强融合特征的帧内和帧间信息。此外,在分类回归头中提出一个额外的中心损失函数,通过限制搜索区域减少目标周围相似物体的影响。

尽管卷积神经网络在目标跟踪中通过局部特征提取表现出色,但在面对卫星视频中的高分辨率图像、长时间序列和复杂背景时,存在一定局限性。为了解决这些问题,Transformer作为一种能够捕捉全局空间信息和长期时序依赖的模型,逐渐引入到卫星视频目标跟踪任务中。Bai等人(2024a)提出一种同步进行在线与离线学习的卫星视频目标跟踪方法 SiamSTO,设计了一个空间到深度特征提取网络(space-to-depth feature extraction network, SDN),以减少小目标跟踪中空间信息的损失。同时,嵌入基于 Transformer 的信息交互(Transformer-based information interaction, TFI)模块的离线学习分支,以增

强目标与背景的区分能力。此外,为弥补离线学习分支仅利用初始帧目标区域的局限性,SiamSTO还结合了在线学习机制,从目标和背景区域动态提取信息,形成更具辨识力的特征表示。Yang等人(2024)提出的 HRLT (high-order relation learning Transformer),通过建模目标与场景上下文的空间和时间高阶关系提高跟踪精度。HRLT包含空间高阶关系推理(spatial high-order relation reasoning, SHR²)和时间高阶关系推理(temporal high-order relation reasoning, THR²)模块,其中,SHR²模块通过整合目标与多个上下文对象的关系,构建目标—上下文的高阶交互,帮助模型在复杂背景中更精确地定位目标,减少背景干扰。THR²模块则负责跨多个帧关联和推理这些空间高阶关系,捕捉目标随时间的动态变化,进一步提高模型的鲁棒性。

3 目标跟踪数据集及评价指标

3.1 卫星视频单目标跟踪数据集

在一个标准的数据集上公平全面地评估一种跟踪方法对卫星视频目标跟踪的发展至关重要,部分研究仅选取有限的目标进行验证,对卫星视频目标跟踪方法性能的评估不够全面。目前公开的卫星视频目标跟踪数据集主要包括 XDU-BDSTU(Zhang等, 2022a)、VISO(video satellite objects)(Yin等, 2022)、SatSOT(Zhao等, 2022)、SV248S(Li等, 2022)和 OOTB(the oriented object tracking benchmark)(Chen等, 2024),详细信息如表1所示。其中,LR表示低分辨率,LF表示特征少,BC表示背景干扰,SO表示相似目标,IPR表示平面内旋转,IV表示光照变化,OCC表示遮挡(POC部分遮挡、FOC完全遮挡),CC表示颜色变化,OV表示超出视野,MB表示运动模糊,BJT表示背景抖动,ARC表示纵横比改变,DEF表示形变。

XDU-BDSTU数据集包含来自9个城市的吉林1号卫星视频,标注了20个具有代表性的车辆目标。数据集总帧数为13 745,包含LR、LF、BC、SO、IPR、IV、OCC等挑战属性。VISO数据集基于吉林1号卫星视频标注,可用于训练和测试,涵盖汽车、火车、轮船和飞机等目标,共3 711个目标,超过百万帧,存在LR、OV、OCC、SO、MB等挑战属性。SatSOT是一个卫星视频目标跟踪测试集,包含吉林1号、Skybox和

Carbonite-2的卫星视频序列,数据集包括65个汽车目标、26个火车目标、5个轮船目标和9个飞机目标,数据量超过2万帧,具有BC、IV、LR、IPR、POC、FOC、SO、ARC等挑战属性。SV248S数据集包括6个视频序列,共248个目标,考虑了OCC、IPR、IV、SO等挑

战属性,并且提供了掩膜级标注。OOTB为定向边界框目标跟踪数据集,包含110个视频序列,共29 890帧,包括45个汽车目标、30个轮船目标、25个飞机目标和10个火车目标。图4展示了开源卫星视频目标跟踪数据集部分示例。

表1 卫星视频目标跟踪公开数据集统计
Table 1 Public datasets of object tracking in satellite videos

数据集	数据类型	目标类型及数量	总帧数	主要挑战属性	发表年份	获取地址
XDU-BDSTU	吉林1号	车辆(20)	10 000+	LR、LF、BC、SO、IPR、IV、OCC	2021	https://github.com/liusiqinqinqin/XDU-BDSTU
VISO	吉林1号	汽车、火车、轮船、飞机(共3 711)	1 000 000+	LR、CC、OV、OCC、SO、MB	2021	https://github.com/QingyongHu/VISO
SatSOT	吉林1号、Skybox、Carbonite-2	汽车(65)、火车(26)、轮船(5)、飞机(9)	20 000+	BC、IV、LR、IPR、POC、FOC、LR、SO、BJT、ARC、DEF	2022	http://www.csu.cas.cn/gb/jggk/kybm/sjlyzx/gcxx_sjj/sjj_wxxl/SV248S
SV248S	吉林1号	轮船、车辆、飞机(共248)	-	OCC、IPR、IV、SO等	2022	https://github.com/xdai-dlgvv/SV248S
OOTB	吉林1号、SkySat-1等	汽车(45)、轮船(30)、飞机(25)、火车(10)	-	IPR、POC、FOC、IV、MB、BC、SO等	2024	https://github.com/YZCU/OOTB

注:“-”表示未统计。



图4 开源卫星视频目标跟踪数据集示例

Fig. 4 Examples of open source satellite video object tracking datasets

3.2 评价指标

对跟踪结果进行评估是目标跟踪的重要一环。卫星视频目标跟踪任务常采用 OTB2013 (Wu 等, 2013) 中的评价准则, 在评估算法时对每一个视频序列使用一次通过评估 (one-pass evaluation, OPE) 方法。跟踪结果的评价通常采用精确率和成功率。

3.2.1 精确率

精确率是指中心定位误差 (center location error, CLE) 小于指定阈值的视频帧占总帧数的百分比 (Wu 等, 2013)。其中, CLE 是指算法预测的目标边界框与真实的目标边界框中心之间的欧氏距离。精度图 (precision plot) 显示了中心定位误差在不同给定阈值下, 满足条件的视频帧占总帧数的百分比。普通视频目标跟踪领域通常使用阈值为 20 像素时的精确率对目标跟踪算法性能进行排序, 在卫星视频目标跟踪领域, 由于目标本身较小, 多数研究者将阈值设定为 3~5 像素 (Wang 等, 2020; Liu 等, 2021; Zhang 等, 2022a; Zhao 等, 2022)。

3.2.2 成功率

成功率是指算法预测的目标边界框与真实的目标边界框的重叠得分 (overlap score, OS) 大于指定阈值的视频帧占总帧数的百分比。其中, 重叠得分计算方法见式 (1)。成功率图 (success plot) 显示了阈值从 0 变化到 1 时的成功率趋势。Wu 等人 (2013) 指出只利用单一阈值 (如 0.5) 下的成功率来评估跟踪算法的性能不具代表性, 使用成功率图的曲线下面积 (area under curve, AUC) 对目标跟踪算法性能进行排序更为合理。在卫星视频目标跟踪领域, 多数研究者使用 AUC 进行评价 (Shao 等, 2019a; Zhu 等, 2021; Chen 和 Sui, 2019; Du 等, 2018; 韩鸣飞等, 2021; Zhang 等, 2022a; Shi 等, 2020)。

$$OS = \frac{a \cap b}{a \cup b} \quad (1)$$

式中, a 为预测的目标边界框, b 为真实的目标边界框。

3.2.3 帧率

帧率 (frame per second, FPS) 是评价算法跟踪速度的重要标准。它表示跟踪器可以在 1 s 内跟踪目标的帧数, 单位为帧/s。通过将跟踪视频序列总帧数除以算法在该序列上的运行时间来计算平均帧率。如果帧率超过 25 帧/s 或 30 帧/s, 则认为该算法满足实时性要求 (Xing 等, 2021)。

4 实验结果与分析

4.1 实验数据集与实验环境

本文实验的目的是在相同的数据集上评估不同跟踪算法的性能, 进而分析各算法在不同场景下的表现差异。为了更好地控制数据集的质量, 本文基于吉林 1 号卫星视频数据, 采用人工标注的方式制作了用于卫星视频单目标跟踪任务的测试集。由于使用普通视频目标跟踪数据训练的模型已经具备一定的特征提取以及相似性度量能力, 因此不对深度学习模型进行训练, 直接使用不同算法官方公布的模型权重进行实验。此外, 研究表明: 相对于飞机和轮船, 跟踪算法在跟踪卫星视频中的车辆目标时更易受到相似目标和背景的干扰 (Zhao 等, 2022), 为了提升跟踪算法精度, 应给予车辆目标更多的关注, 所以自制的数据集中只对车辆目标进行了标注。为了测试算法在不同场景中的性能, 自制的测试集选取了 4 种常见的场景或挑战共 50 个车辆目标进行标注, 具体信息见表 2。

表 2 自制卫星视频目标跟踪数据集

Table 2 Self-made satellite video object tracking dataset

数据集	数据类型	目标与数量	总帧数	属性
测试集	吉林 1 号	车辆 (50)	8 491	常规序列 (不具备挑战属性, 20); 相似干扰物挑战 (10); 平面内旋转 (10); 运动模糊 (10)

本文的实验环境: 操作系统为 Ubuntu 20.04; CPU 为 AMD Ryzen 7 5800X 3.80 GHz; GPU 为 NVIDIA GeForce RTX 3070; 显存大小为 8 GB。

4.2 单目标跟踪算法的评估

为了研究普通视频目标跟踪领域的方法在卫星视频目标跟踪任务的适用性, 本文选取 18 种算法在自制的卫星视频测试集进行评估, 包括基于相关滤波的目标跟踪方法 MOSSE (minimum output sum of squared error) (Bolme 等, 2010)、CSK (Henriques 等, 2012)、CN (color names) (Danelljan 等, 2014)、KCF (Henriques 等, 2015), 基于孪生网络的目标跟踪方法 SiamFC (siamese fully-convolutional) (Bertinetto 等,

2016)、SiamRPN(Li等,2018a)、SiamRPN++(Li等,2019),基于Transformer结构的目标跟踪方法Trdimp(Wang等,2021)、Trsiam(Wang等,2021)、Stark(Yan等,2021)、TransT(Chen等,2021)、HiFT(hierarchical feature Transformer)(Cao等,2021)、SwinTrack(Lin等,2022)、SiamTPN(Xing等,2022)、ARTrack(Wei等,2023)、SeqTrack(Chen等,2023)、ARTrackV2(Bai等,2024b)以及专门针对卫星视频的单目标跟踪方法CFME(Xuan等,2020)。需要说明的是,本节使用的深度学习模型是在通用视频数据集(例如ImageNet VID、YoutubeBB(Real等,2017)、COCO(Lin等,2014)、ImageNet Det(Russakovsky等,2015)等)上进行训练的。

4.2.1 定量评价

首先使用18种目标跟踪算法对自制的50个卫星视频序列进行评价,本文使用阈值为3像素时的精确率以及AUC对跟踪算法进行排序,评价结果如表3和图5所示。

从表3和图5可以看出,在本文自制数据集中,

表3 18种目标跟踪算法在50个卫星视频序列的结果对比

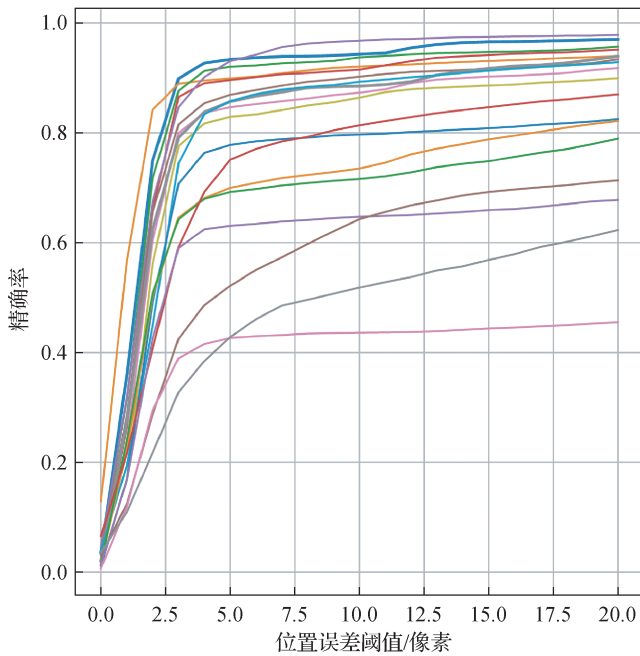
Table 3 Comparison of the results of 18 object tracking algorithms on 50 satellite video sequences

跟踪算法	精确率 (3像素)/%	成功率 (AUC)/%	帧率/(帧/s)	
			GPU	CPU
SiamRPN++	89.88	56.82	56.42	-
CFME	88.94	70.30	-	349.86
Trdimp	87.64	62.07	20.83	-
ARTrackV2	86.53	55.02	-	-
SiamFC	84.69	62.02	81.75	-
Trsiam	81.51	58.62	17.61	-
ARTrack	80.00	51.75	-	-
SiamRPN	79.15	52.35	358.27	-
SwinTrack	77.57	52.64	36.03	-
KCF	74.48	53.85	-	305.49
SeqTrack	70.76	46.61	-	-
HiFT	64.56	42.24	148.19	-
TransT	64.27	42.78	47.33	-
CN	59.25	48.76	-	130.43
Stark	59.05	39.66	24.50	-
MOSSE	42.53	34.71	-	351.03
SiamTPN	38.98	26.57	16.58	-
CSK	32.81	29.66	-	436.57

注:“-”表示未进行测试。

Trackers (CLE@3px)

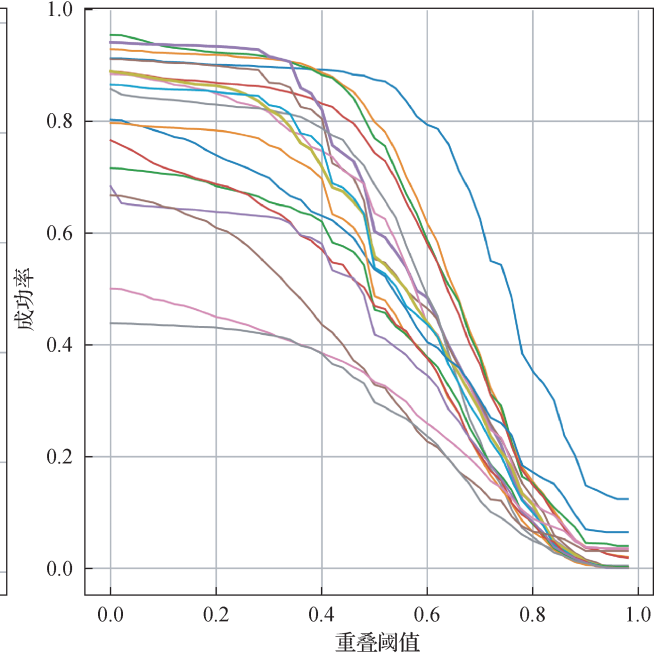
- SiamRPN++ [0.898 8]
- CFME [0.889 4]
- Trdimp [0.876 4]
- ARTrack V2 [0.865 3]
- SiamFC [0.846 9]
- Trsiam [0.815 1]
- ARTrack [0.800 0]
- SiamRPN [0.791 5]
- SwinTrack [0.775 7]
- KCF_gray [0.744 8]
- SeqTrack [0.707 6]
- HiFT [0.645 6]
- TransT [0.642 7]
- CN [0.592 5]
- Stark [0.590 5]
- MOSSE [0.425 3]
- SiamTPN [0.389 8]
- CSK [0.328 1]



(a) 精度图

Trackers (AUC)

- CFME [0.703 0]
- Trdimp [0.620 7]
- SiamFC [0.620 2]
- Trsiam [0.586 2]
- SiamRPN++ [0.568 2]
- ARTrack V2 [0.550 2]
- KCF_gray [0.538 5]
- SwinTrack [0.526 4]
- SiamRPN [0.523 5]
- ARTrack [0.517 5]
- CN [0.487 6]
- SeqTrack [0.466 1]
- TransT [0.427 8]
- HiFT [0.422 4]
- Stark [0.396 6]
- MOSSE [0.347 1]
- CSK [0.296 6]
- SiamTPN [0.265 7]



(b) 成功率图

图5 18种目标跟踪算法在50个卫星视频序列上的精度图和成功率图

Fig. 5 Precision and success plots of 18 object tracking algorithms on 50 satellite video sequences((a)precision plot;(b)success plot)

基于卫星视频单目标跟踪的方法 CFME 取得较好结果, 精确率为 88.94%, 成功率为 70.30%, 证明了运动估计在卫星视频目标跟踪中的有效性。SiamRPN++ 精确率最高, 达到 89.88%, 但其成功率相对较低。Trdimp、ARTrackV2、SiamFC 和 Trsiam 等深度学习方法在本文自制的测试集上都表现出优越性能, 优于 KCF 和 CN 等相关滤波算法。对于跟踪速度, 基于相关滤波器的算法 CFME (349.86 帧/s) 和 KCF (305.49 帧/s) 等在 CPU 上的速度都达到实时性要求。由于引入了 Transformer 结构, Trdimp (20.83 帧/s)、Trsiam (17.61 帧/s) 和 SwinTrack (36.03 帧/s) 等算法的跟踪速度都低于基于孪生网络的算法 SiamFC (81.75 帧/s)、SiamRPN (358.27 帧/s) 和 SiamRPN++ (56.42 帧/s)。

为了评价各跟踪算法在不同场景的适用性, 分别在自制的不同场景序列进行测试, 结果如表 4、图 6 和图 7 所示。可以看出, 在本文自制的测试集

中, CFME 在多数场景下性能都十分优越, 进一步表明了进行运动估计的有效性, 但该算法不能很好地估计目标边界框的尺寸变化, 在平面内旋转挑战序列的精度下降。基于深度学习的方法, 如 Trdimp、ARTrackV2、Trsiam、SiamFC 等都取得了不错的效果, 尤其是基于 Transformer 结构的 Trdimp 算法, 在多数场景下跟踪性能都较好, 分析是由于其利用了视频帧间丰富的时序信息。值得一提的是, SiamRPN、SiamRPN++ 的锚框机制可以很好地解决目标旋转问题, 在应对平面内旋转挑战时精度较高。

4.2.2 定性评价

为了进一步分析跟踪器效果, 本文选取表现较好的 6 个跟踪器对它们在不同场景下的跟踪结果进行可视化, 如图 8 所示。可以看出, 在没有挑战的常规场景中, 6 种跟踪算法都能较好地跟踪目标; 由于目标特征少且与干扰物极为相似, 部分跟踪器在相似干扰物挑战序列中发生了模型漂移, 而 CFME

表 4 18 种目标跟踪算法在不同场景的结果对比

Table 4 Comparison of the results of 18 object tracking algorithms in various scenarios

跟踪算法	常规		相似干扰物		平面内旋转		运动模糊	
	精确率	成功率	精确率	成功率	精确率	成功率	精确率	成功率
SiamRPN++	95.16	67.44	80.47	40.23	97.55	72.88	81.04	36.12
CFME	99.53	81.37	93.67	70.94	61.10	50.98	90.86	68.84
Trdimp	91.18	66.98	75.50	53.76	89.66	65.56	90.68	57.08
ARTrackV2	95.76	68.92	76.79	38.84	79.73	60.00	84.59	38.41
SiamFC	90.13	70.16	71.05	51.74	81.05	59.83	91.08	58.24
Trsiam	88.94	65.92	63.50	45.58	81.94	62.02	84.25	53.65
ARTrack	92.32	65.55	59.40	31.64	76.30	58.72	79.65	37.26
SiamRPN	87.38	64.43	56.32	31.37	96.58	71.93	68.10	29.57
SwinTrack	90.68	64.28	57.86	37.49	75.51	51.83	73.11	45.32
KCF	88.17	65.52	72.30	47.28	79.25	59.47	44.54	31.47
SeqTrack	89.99	64.14	49.63	27.57	62.58	45.93	61.59	31.27
HiFT	85.95	61.71	35.12	18.73	61.22	47.38	54.58	21.67
TransT	84.68	62.15	39.55	21.24	59.93	44.56	52.49	23.79
CN	74.66	63.72	59.18	40.56	61.03	57.14	26.70	18.66
Stark	81.19	59.20	36.77	20.08	44.56	34.51	51.55	25.30
MOSSE	22.70	27.04	74.72	48.59	14.45	21.58	78.08	49.28
SiamTPN	59.59	42.93	23.42	12.62	36.26	27.07	16.06	7.29
CSK	45.43	40.71	25.05	19.22	35.30	37.78	12.83	9.88

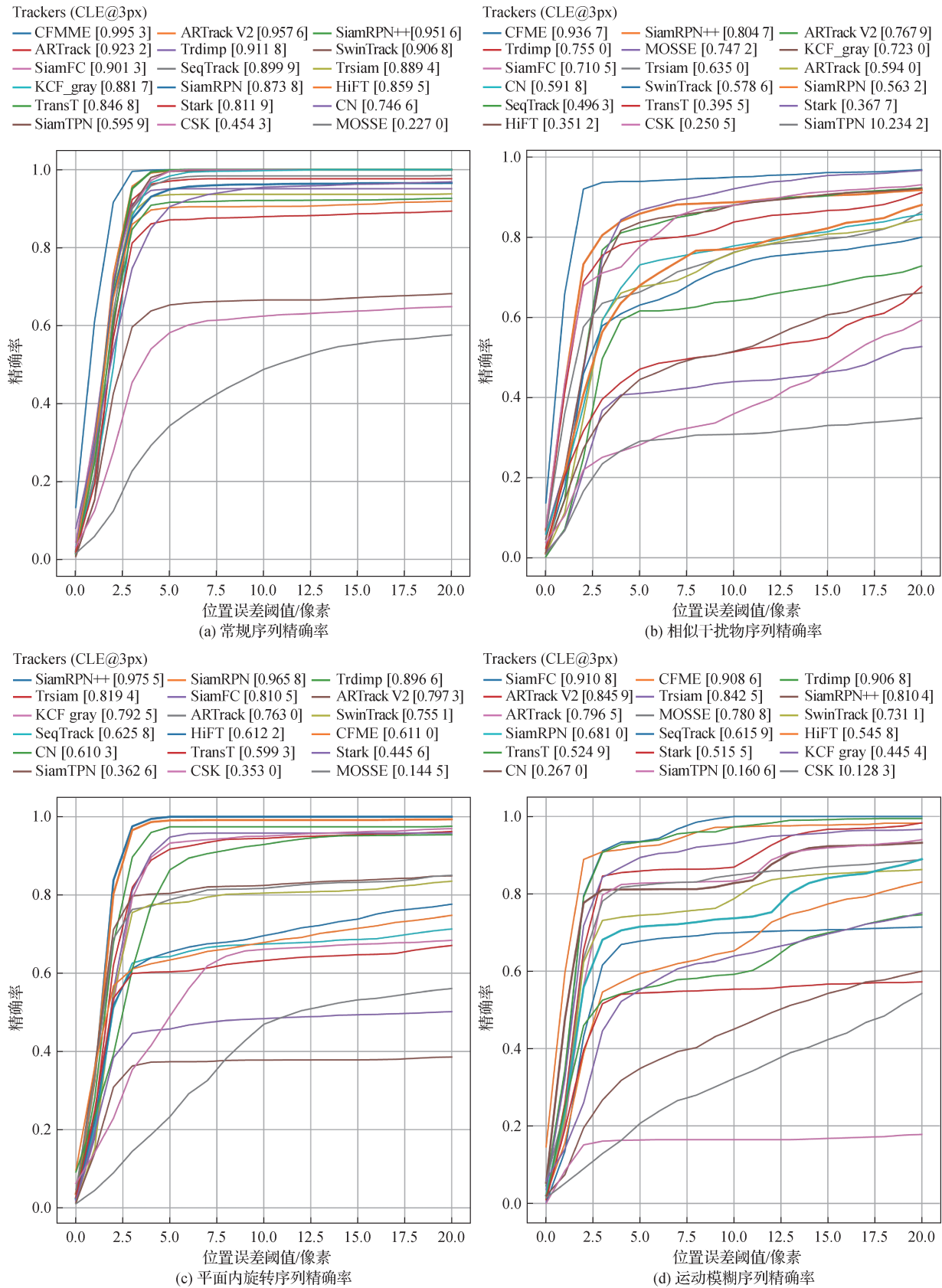


图6 18种目标跟踪算法在不同场景的精度图

Fig. 6 Precision plots of 18 object tracking algorithms in different scenario ((a) precision plot for non-challenging sequences; (b) precision plot for similar distractor challenge sequences; (c) precision plot for in-plane rotation challenge sequences; (d) precision plot for motion blur challenge sequences)

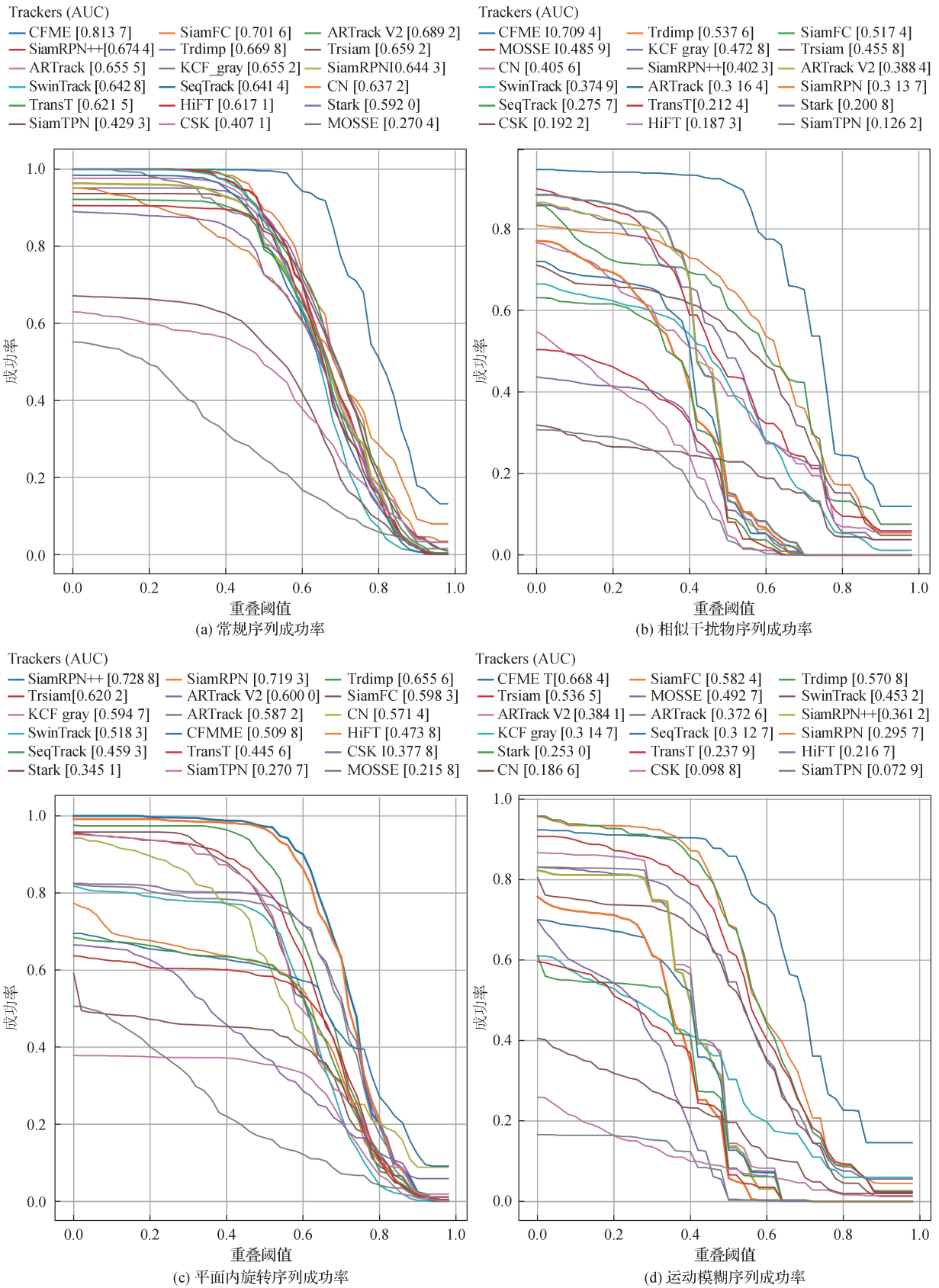


图7 18种目标跟踪算法在不同场景的成功率图

Fig. 7 Success plots of 18 object tracking algorithms in different scenario

((a) success plot for non-challenging sequences; (b) success plot for similar distractor challenge sequences; (c) success plot for in-plane rotation challenge sequences; (d) success plot for motion blur challenge sequences)

利用了目标历史运动信息, Trdimp 考虑了时序信息以及背景信息, 获得了较好的效果; 在车辆转弯时, 基于相关滤波的方法 CFME 使用的手工特征不具备旋转不变性, 且不能很好地处理由于旋转导致的目标边界框的变化, 导致跟踪效果不佳, 而 TrDimp、Trsiam 等直接对目标的边界框进行估计,

可以很好地解决平面内旋转挑战; 在运动模糊挑战中, 车辆目标特征的提取变得更加困难, 多数跟踪器发生了模型漂移, 但 CFME、Trdimp 的跟踪效果依然较好, 进一步说明了运动估计、时序信息以及背景信息的利用在卫星视频目标跟踪任务中的重要性。



图8 精确率前6名跟踪算法在不同场景下的跟踪结果可视化

Fig. 8 Visualization of the tracking results of the top 6 tracking algorithms with precision in different scenarios

5 结 语

尽管在卫星视频单目标跟踪领域中的研究已经取得显著进展, 但还有一些问题需要进一步探讨。本节在回顾现有研究存在问题基础上, 对卫星视频单目标跟踪的具体发展趋势进行讨论:

1) 统一跟踪结果评价指标。普通视频目标跟踪领域常用的评价指标有 OTB (online object tracking benchmark) 和 VOT (visual object tracking) 评价指标。

在卫星视频目标跟踪领域, 学者们多采用 OTB 评价指标中的精确率和成功率对算法进行评价。普通视频 OTB 评价指标中的精确率阈值通常设置为 20 像素, 成功率图的排序依据为曲线下面积 AUC, 但在卫星视频目标跟踪领域, 不同研究者设置的阈值不同, 导致在统一标准下客观评估算法变得困难, 所以统一跟踪结果的评价指标, 将有利于卫星视频单目标跟踪领域的发展。

2) 更大规模的高质量卫星视频目标跟踪数据集。由于缺乏大规模高质量训练集, 卫星视频目标

跟踪任务中现有的深度学习模型多基于普通视频数据进行训练,这阻碍了深度学习在该领域的发展。在大规模测试数据集可用之前,大多数卫星视频目标跟踪研究实验部分仅使用几个目标进行验证,算法性能得不到充分的评估,且不同研究使用的测试数据均不相同,导致不同算法之间不能进行统一对比。所以亟需发展大规模高质量卫星视频目标跟踪数据集,这不仅有利于跟踪器模型训练,对于模型测试和性能比较也至关重要。

3)更适合于卫星视频单目标跟踪的模型。未来的研究可快速融合普通视频目标跟踪领域最新方法,并针对卫星视频的特点进行设计与改进。卫星视频中背景信息丰富且相邻帧间目标运动轨迹呈连续线性,可充分利用这些先验知识,充分挖掘全局空间信息以及时序信息,提升卫星视频单目标跟踪的精度。还可考虑引入知识蒸馏、网络剪枝以及神经结构搜索等技术,自动构建结构精简、复杂度低的适用于卫星视频单目标跟踪任务的模型,实现高精度、高实时性的目标跟踪。

4)长时跟踪。卫星视频相对于地面监控视频的优势在于覆盖范围广,可支撑城市范围内的感兴趣目标轨迹获取。在大范围场景内,可能会同时存在遮挡、相似物干扰、运动模糊、光照变化以及目标旋转等挑战。为了满足实际应用的需求,需要发展一个能同时解决多数挑战的卫星视频目标跟踪方法。

参考文献 (References)

- Bai L, Li Y W, Liu Y and Jiang Z Z. 2024a. SiamSTO: a novel method for object tracking with satellite videos//Proceedings of IET International Radar Conference (IRC 2023). Chongqing, China: IEEE: 786-792 [DOI: 10.1049/iep.2024.1186]
- Bai Y F, Zhao Z Y, Gong Y H and Wei X. 2024b. ARTrackV2: prompting autoregressive tracker where to look and how to describe//Proceedings of 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, USA: IEEE: 19048-19057 [DOI: 10.1109/CVPR52733.2024.01802]
- Bertinetto L, Valmadre J, Henriques J F, Vedaldi A and Torr P H S. 2016. Fully-convolutional Siamese networks for object tracking//Proceedings of 2016 European Conference on Computer Vision. Amsterdam, the Netherlands: Springer: 850-865 [DOI: 10.1007/978-3-319-48881-3_56]
- Bi F K, Sun J Y, Han J H, Wang Y P and Bian M M. 2021. Remote sensing target tracking in satellite videos based on a variable-angle adaptive Siamese network. IET Image Processing, 15 (9): 1987-1997 [DOI: 10.1049/ipr2.12170]
- Bolme D S, Beveridge J R, Draper B A and Lui Y M. 2010. Visual object tracking using adaptive correlation filters//Proceedings of 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Francisco, USA: IEEE: 2544-2550 [DOI: 10.1109/CVPR.2010.5539960]
- Bradski G R. 1998. Real time face and object tracking as a component of a perceptual user interface//Proceedings of the 4th IEEE Workshop on Applications of Computer Vision. WACV'98 (Cat. No.98EX201), Princeton, USA: 214-219 [DOI: 10.1109/ACV.1998.732882]
- Cao Z A, Fu C H, Ye J J, Li B W and Li Y M. 2021. HiFT: hierarchical feature transformer for aerial tracking//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal, Canada: IEEE: 15437-15446 [DOI: 10.1109/ICCV48922.2021.01517]
- Chen X and Sui H. 2019. Real-time tracking in satellite videos via joint discrimination and pose estimation. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLII-2-W12: 23-29 [DOI: 10.5194/isprs-archives-XLII-2-W12-23-2019]
- Chen X, Peng H W, Wang D, Lu H C and Hu H. 2023. Seqtrack: sequence to sequence learning for visual object tracking//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. [s. l.]: IEEE: 14572-14581 [DOI: 10.1109/CVPR52729.2023.01400]
- Chen X, Yan B, Zhu J W, Wang D, Yang X Y and Lu H C. 2021. Transformer tracking//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville, USA: IEEE: 8122-8131 [DOI: 10.1109/CVPR46437.2021.00803]
- Chen Y Z, Tang Y Q, Han T, Zhang Y W, Zou B and Feng H H. 2022. RAMC: a rotation adaptive tracker with motion constraint for satellite video single-object tracking. Remote Sensing, 14(13): #3108 [DOI: 10.3390/rs14133108]
- Chen Y Z, Tang Y Q, Xiao Y, Yuan Q Q, Zhang Y W, Liu F Q, He J and Zhang L P. 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]
- Comaniciu D, Ramesh V and Meer P. 2000. Real-time tracking of non-rigid objects using mean shift//Proceedings of 2000 IEEE Conference on Computer Vision and Pattern Recognition. Hilton Head, USA: IEEE: 142-149 [DOI: 10.1109/CVPR.2000.854761]
- Cui Y Y, Hou B, Wu Q, Ren B, Wang S and Jiao L C. 2022. Remote sensing object tracking with deep reinforcement learning under occlusion. IEEE Transactions on Geoscience and Remote Sensing, 60: #5605213 [DOI: 10.1109/TGRS.2021.3096809]
- Danelljan M, Khan F S, Felsberg M and Van De Weijer J. 2014. Adap-

- tive color attributes for real-time visual tracking//Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus, USA: IEEE: 1090-1097 [DOI: 10.1109/CVPR.2014.143]
- Ding Y F, He H Y, Cao S X and Wang Y. 2021. A long-term tracking algorithm of space-based targets based on multi-feature integration. *Spacecraft Recovery and Remote Sensing*, 42(4): 46-56 (丁与非, 何红艳, 曹世翔, 王钰. 2021. 基于多特征融合的天基面目标长时跟踪算法. *航天返回与遥感*, 42(4): 46-56) [DOI: 10.3969/j.issn.1009-8518.2021.04.006]
- Dong X P and Shen J B. 2018. Triplet loss in Siamese network for object tracking//Proceedings of the 15th European Conference on Computer Vision (ECCV). Munich, Germany: Springer: 472-488 [DOI: 10.1007/978-3-030-01261-8_28]
- Du B, Cai S H and Wu C. 2019. Object tracking in satellite videos based on a multiframe optical flow tracker. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 12(8): 3043-3055 [DOI: 10.1109/JSTARS.2019.2917703]
- Du B, Sun Y J, Cai S H, Wu C and Du Q. 2018. Object tracking in satellite videos by fusing the kernel correlation filter and the three-frame-difference algorithm. *IEEE Geoscience and Remote Sensing Letters*, 15(2): 168-172 [DOI: 10.1109/LGRS.2017.2776899]
- Feng J, Zeng D N, Jia X P, Zhang X R, Li J, Liang Y P and Jiao L C. 2021. Cross-frame keypoint-based and spatial motion information-guided networks for moving vehicle detection and tracking in satellite videos. *ISPRS Journal of Photogrammetry and Remote Sensing*, 177: 116-130 [DOI: 10.1016/j.isprsjprs.2021.05.005]
- Guo Y J, Yang D Q and Chen Z Z. 2019. Object tracking on satellite videos: a correlation filter-based tracking method with trajectory correction by Kalman filter. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 12(9): 3538-3551 [DOI: 10.1109/JSTARS.2019.2933488]
- Han M F, Li S Y, Wan X, Xuan S Y, Zhao Z F, Tan H and Zhang W F. 2021. Integrating multiple features for tracking vehicles in satellite videos. *Journal of Image and Graphics*, 26(11): 2741-2750 (韩鹏飞, 李盛阳, 万雪, 轩诗宇, 赵子飞, 谭洪, 张万峰. 2021. 融合特征的卫星视频车辆单目标跟踪. *中国图象图形学报*, 26(11): 2741-2750) [DOI: 10.11834/jig.200094]
- Henriques J F, Caseiro R, Martins P and Batista J. 2012. Exploiting the circulant structure of tracking-by-detection with kernels//Proceedings of the 12th European Conference on Computer Vision. Florence, Italy: Springer: 702-715 [DOI: 10.1007/978-3-642-33765-9_50]
- Henriques J F, Caseiro R, Martins P and Batista J. 2015. High-speed tracking with kernelized correlation filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(3): 583-596 [DOI: 10.1109/TPAMI.2014.2345390]
- Hou B, Cui Y Y, Ren Z L, Li Z H, Wang S and Jiao L C. 2023. Siamese multi-scale adaptive search network for remote sensing single-object tracking. *Remote Sensing*, 15(17): #4359 [DOI: 10.3390/rs15174359]
- Hu Z P, Yang D Q, Zhang K and Chen Z Z. 2020. Object tracking in satellite videos based on convolutional regression network with appearance and motion features. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13: 783-793 [DOI: 10.1109/JSTARS.2020.2971657]
- Huang Z J, Jiao L C, Zhang J Y, Liu X, Liu F, Zhang X R, Li L L and Chen P H. 2024. A graph association motion-aware tracker for tiny object in satellite videos. *IEEE Transactions on Circuits and Systems for Video Technology*, 34(12): 12907-12922 [DOI: 10.1109/TCSVT.2024.3439371]
- Isard M and Blake A. 1998. Condensation—conditional density propagation for visual tracking. *International Journal of Computer Vision*, 29(1): 5-28 [DOI: 10.1023/A:1008078328650]
- Li B, Wu W, Wang Q, Zhang F Y, Xing J L and Yan J J. 2019. SiamRPN++: evolution of Siamese visual tracking with very deep networks//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, USA: IEEE: 4277-4286 [DOI: 10.1109/CVPR.2019.00441]
- Li B, Yan J J, Wu W, Zhu Z and Hu X L. 2018a. High performance visual tracking with Siamese region proposal network//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 8971-8980 [DOI: 10.1109/CVPR.2018.00935]
- Li P X, Wang D, Wang L J and Lu H C. 2018b. Deep visual tracking: review and experimental comparison. *Pattern Recognition*, 76: 323-338 [DOI: 10.1016/j.patcog.2017.11.007]
- Li Y X, Jiao L C, Huang Z J, Zhang X, Zhang R H, Song X, Tian C X, Zhang Z X, Liu F, Yang S Y, Hou B, Ma W P, Liu X and Li L L. 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]
- Liang J H, Sun J Y, Zhang X M, Bian M M and Bi F K. 2023. Lightweight tracking of satellite video object based on saliency enhancement mechanism. *IEEE Journal on Miniaturization for Air and Space Systems*, 4(2): 100-104 [DOI: 10.1109/JMASS.2023.3234099]
- Lin L T, Fan H, Zhang Z P, Xu Y and Ling H B. 2022. SwinTrack: a simple and strong baseline for transformer tracking [EB/OL]. [2025-01-01]. <https://arxiv.org/pdf/2112.00995.pdf>
- Lin T Y, Maire M, Belongie S, Hays J, Perona P, Ramanan D, Dollár P and Zitnick C L. 2014. Microsoft COCO: common objects in context//Proceedings of the 13th European Conference on Computer Vision. Zurich, Switzerland: Springer: 740-755 [DOI: 10.1007/978-3-319-10602-1_48]
- Liu Y S, Liao Y R, Lin C B, Li Z M, Yang X Y and Zhang A D. 2021. Object tracking in satellite videos based on improved correlation fil-

- ters//Proceedings of the 13th International Conference on Communication Software and Networks. Chongqing, China: IEEE: 323-331 [DOI: 10.1109/ICCSN52437.2021.9463667]
- Lyu P Y, Gao X Y and Zhong Y F. 2025. SiamS²F: satellite video single-object tracking based on a Siamese spectral-spatial-frame correlation network. *IEEE Transactions on Geoscience and Remote Sensing*, 63: #5608213 [DOI: 10.1109/TGRS.2025.3535831]
- Pei W J and Lu X H. 2022. Moving object tracking in satellite videos by kernelized correlation filter based on color-name features and Kalman prediction. *Wireless Communications and Mobile Computing*, 2022: #9735887 [DOI: 10.1155/2022/9735887]
- Real E, Shlens J, Mazzocchi S, Pan X and Vanhoucke V. 2017. Youtube-BoundingBoxes: a large high-precision human-annotated data set for object detection in video//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, USA: IEEE: 7464-7473 [DOI: 10.1109/CVPR.2017.789]
- Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S A, Huang Z H, Karpathy A, Khosla A, Bernstein M, Berg A C and Li F F. 2015. ImageNet large scale visual recognition challenge. *International Journal of Computer Vision*, 115(3): 211-252 [DOI: 10.1007/s11263-015-0816-y]
- Shao J, Du B, Chen W, Gong M M and Liu T L. 2021. HRSiam: high-resolution Siamese network, towards space-borne satellite video tracking. *IEEE Transactions on Image Processing*, 30: 3056-3068 [DOI: 10.1109/TIP.2020.3045634]
- Shao J, Du B, Wu C and Yan P K. 2019a. PASiam: predicting attention inspired Siamese network, for space-borne satellite video tracking//Proceedings of 2019 IEEE International Conference on Multimedia and Expo (ICME). Shanghai, China: IEEE: 1504-1509 [DOI: 10.1109/ICME.2019.00260]
- Shao J, Du B, Wu C and Zhang L F. 2019b. Can we track targets from space? A hybrid kernel correlation filter tracker for satellite video. *IEEE Transactions on Geoscience and Remote Sensing*, 57(11): 8719-8731 [DOI: 10.1109/TGRS.2019.2922648]
- Shao J, Du B, Wu C and Zhang L F. 2019c. Tracking objects from satellite videos: a velocity feature based correlation filter. *IEEE Transactions on Geoscience and Remote Sensing*, 57(10): 7860-7871 [DOI: 10.1109/TGRS.2019.2916953]
- Shi F, Qiu F, Li X, Zhong R F, Yang C K and Tang Y W. 2020. Detecting and tracking moving airplanes from space based on normalized frame difference labeling and improved similarity measures. *Remote Sensing*, 12(21): #3589 [DOI: 10.3390/rs12213589]
- Smeulders A W M, Chu D M, Cucchiara R, Calderara S, Dehghan A and Shah M. 2014. Visual tracking: an experimental survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(7): 1442-1468 [DOI: 10.1109/TPAMI.2013.230]
- Song W, Jiao L C, Liu F, Liu X, Li L L, Yang S Y, Hou B and Zhang W H. 2022. A joint Siamese attention-aware network for vehicle object tracking in satellite videos. *IEEE Transactions on Geoscience and Remote Sensing*, 60: #5625617 [DOI: 10.1109/TGRS.2022.3184755]
- Sui B, Xu M and Li D D. 2021. Object tracking in satellite videos based on a lightweight network. *Journal of Circuits, Systems and Computers*, 30(6): #2150099 [DOI: 10.1142/S0218126621500997]
- Sun K, Xiao B, Liu D and Wang J D. 2019. Deep high-resolution representation learning for human pose estimation//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE: 5686-5696 [DOI: 10.1109/CVPR.2019.00584]
- Wang J H, Liu F, Jiao L C, Gao Y J, Wang H, Li L L, Chen P H, Liu X and Li S. 2024. Satellite video object tracking based on location prompts. *IEEE Transactions on Circuits and Systems for Video Technology*, 34(7): 6253-6264 [DOI: 10.1109/TCSVT.2024.3358549]
- Wang N, Zhou W G, Wang J and Li H Q. 2021. Transformer meets tracker: exploiting temporal context for robust visual tracking//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville, USA: IEEE: 1571-1580 [DOI: 10.1109/CVPR46437.2021.00162]
- Wang Q, Zhang L, Bertinetto L, Hu W M and Torr P H S. 2019. Fast online object tracking and segmentation: a unifying approach//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, USA: IEEE: 1328-1338 [DOI: 10.1109/CVPR.2019.00142]
- Wang W L. 2016. Research on Target Tracking Method Based on Satellite Video. Fuxin: Liaoning Technical University (王万里. 2016. 基于卫星视频的目标跟踪方法研究. 阜新: 辽宁工程技术大学)
- Wang Y M, Wang T Y, Zhang G, Cheng Q and Wu J Q. 2020. Small target tracking in satellite videos using background compensation. *IEEE Transactions on Geoscience and Remote Sensing*, 58(10): 7010-7021 [DOI: 10.1109/TGRS.2020.2978512]
- Wei X, Bai Y F, Zheng Y C, Shi D H and Gong Y H. 2023. Autoregressive visual tracking//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. [s.l.]: IEEE: 9697-9706 [DOI: 10.1109/CVPR52729.2023.00935]
- Wu D, Song H B and Fan C Z. 2022. Object tracking in satellite videos based on improved kernel correlation filter assisted by road information. *Remote Sensing*, 14(17): #4215 [DOI: 10.3390/rs14174215]
- Wu J Q, Wang T Y, Yan J, Zhang G, Jiang X H, Wang Y M, Bai Q and Yuan C. 2019. Satellite video point-target tracking based on Hu correlation filter. *Chinese Space Science and Technology*, 39(3): 55-63 (吴佳奇, 汪韬阳, 颜军, 张过, 蒋晓华, 王韵鸣, 白倩, 袁冲. 2019. 基于Hu相关滤波的光学卫星视频点目标跟踪. 中国空间科学技术, 39(3): 55-63) [DOI: 10.16708/j.cnki.1000-758X.2018.0059]
- Wu J Q, Zhang G, Wang T Y and Jiang Y H. 2017. Satellite video point-target tracking in combination with motion smoothness constraint and grayscale feature. *Acta Geodaetica et Cartographica Sinica*, 46(9): 1135-1146 (吴佳奇, 张过, 汪韬阳, 蒋永华. 2017. 结合

- 运动平滑约束与灰度特征的卫星视频点目标跟踪. 测绘学报, 46(9): 1135-1146 [DOI: 10.11947/j.AGCS.2017.20160599]
- Wu Y, Lim J and Yang M H. 2013. Online object tracking: a benchmark//Proceedings of 2013 IEEE Conference on Computer Vision and Pattern Recognition. Portland, USA: IEEE: 2411-2418 [DOI: 10.1109/CVPR.2013.312]
- Xing D T, Evangeliou N, Tsoukalas A and Tzes A. 2022. Siamese transformer pyramid networks for real-time UAV tracking//Proceedings of 2022 IEEE/CVF Winter Conference on Applications of Computer Vision. Waikoloa, USA: IEEE: 1898-1907 [DOI: 10.1109/WACV51458.2022.00196]
- Xing W W, Liu W B, Wang J, Zhang S L, Wang L H, Yang Y X and Song B W. 2021. Visual Object Tracking from Correlation Filter to Deep Learning. Singapore: Springer: 15-44 [DOI: 10.1007/978-981-16-6242-3]
- Xuan S Y, Li S Y, Han M F, Wan X and Xia G S. 2020. Object tracking in satellite videos by improved correlation filters with motion estimations. IEEE Transactions on Geoscience and Remote Sensing, 58(2): 1074-1086 [DOI: 10.1109/TGRS.2019.2943366]
- Xuan S Y, Li S Y, Zhao Z F, Zhou Z, Zhang W F, Tan H, Xia G S and Gu Y F. 2021. Rotation adaptive correlation filter for moving object tracking in satellite videos. Neurocomputing, 438: 94-106 [DOI: 10.1016/j.neucom.2021.01.058]
- Yan B, Peng H W, Fu J L, Wang D and Lu H C. 2021. Learning spatio-temporal transformer for visual tracking//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal, Canada: IEEE: 10428-10437 [DOI: 10.1109/ICCV48922.2021.01028]
- Yang J W, Pan Z X, Liu Y H, Niu B and Lei B. 2023a. Single object tracking in satellite videos based on feature enhancement and multi-level matching strategy. Remote Sensing, 15(17): #4351 [DOI: 10.3390/rs15174351]
- Yang J W, Pan Z X, Wang Z M, Lei B and Hu Y X. 2023b. SiamMDM: an adaptive fusion network with dynamic template for real-time satellite video single object tracking. IEEE Transactions on Geoscience and Remote Sensing, 61: #5608619 [DOI: 10.1109/TGRS.2023.3271645]
- Yang X Y, Jiao L C, Li Y Y, Liu X, Li L L, Chen P H, Liu F and Yang S Y. 2024. High-order relation learning transformer for satellite video object tracking. IEEE Transactions on Geoscience and Remote Sensing, 62: #5637415 [DOI: 10.1109/TGRS.2024.3444820]
- Yin Q, Hu Q Y, Liu H, Zhang F, Wang Y Q, Lin Z P, An W and Guo Y L. 2022. Detecting and tracking small and dense moving objects in satellite videos: a benchmark. IEEE Transactions on Geoscience and Remote Sensing, 60: #5612518 [DOI: 10.1109/TGRS.2021.3130436]
- Zhang G. 2016. Satellite video processing and applications. Journal of Applied Sciences, 34(4): 361-370 (张过. 2016. 卫星视频处理与应用进展. 应用科学学报, 34(4): 361-370) [DOI: 10.3969/j.issn.0255-8297.2016.04.001]
- Zhang W H, Jiao L C, Liu F, Li L L, Liu X and Liu J. 2022a. MBLT: learning motion and background for vehicle tracking in satellite videos. IEEE Transactions on Geoscience and Remote Sensing, 60: #4703315 [DOI: 10.1109/TGRS.2021.3109028]
- Zhang X D, Zhu K, Chen G Z, Liao P Y, Tan X L, Wang T and Li X W. 2023. High-resolution satellite video single object tracking based on thicksiam framework. GIScience and Remote Sensing, 60(1): #2163063 [DOI: 10.1080/15481603.2022.2163063]
- Zhang Y, Chen D and Zheng Y H. 2022b. Satellite video tracking by multi-feature correlation filters with motion estimation. Remote Sensing, 14(11): #2691 [DOI: 10.3390/rs14112691]
- Zhang Z P and Peng H W. 2019. Deeper and wider Siamese networks for real-time visual tracking//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, USA: IEEE: 4586-4595 [DOI: 10.1109/CVPR.2019.00472]
- Zhang Z X, Zhu R F, Zhong X and Gao F. 2019. Improved real-time spatio-temporal context target tracking algorithm for high-resolution video satellites. Fire Control and Command Control, 44(12): 45-50 (张作省, 朱瑞飞, 钟兴, 高放. 2019. 视频卫星单目标实时跟踪算法. 火力与指挥控制, 44(12): 45-50) [DOI: 10.3969/j.issn.1002-0640.2019.12.010]
- Zhao C H and Liu H Y. 2019. Aircraft object tracking algorithm based on remote sensing satellite bright visual video. Journal of Shenyang University(Natural Science), 31(4): 284-290 (赵春晖, 刘海燕. 2019. 遥感卫星明视视频中飞机目标的跟踪算法. 沈阳大学学报(自然科学版), 31(4): 284-290) [DOI: 10.16103/j.cnki.21-1583/n.2019.04.005]
- Zhao M Q, Li S Y, Xuan S Y, Kou L X, Gong S and Zhou Z. 2022. SatSOT: a benchmark dataset for satellite video single object tracking. IEEE Transactions on Geoscience and Remote Sensing, 60: #5617611 [DOI: 10.1109/TGRS.2022.3140809]
- Zhao X L. 2018. Research on Target Tracking in Remote Sensing Videos Based on Deep Convolutional Network. Changsha: National University of Defense Technology (赵晓莉. 2018. 基于深度卷积网络的遥感视频目标跟踪方法研究. 长沙: 国防科技大学) [DOI: 10.27052/d.cnki.gzjgu.2018.000969]
- Zhong Y F, Fang X T and Shu M. 2024. Online background discriminative learning for satellite video object tracking. IEEE Transactions on Geoscience and Remote Sensing, 62: #5605815 [DOI: 10.1109/TGRS.2024.3353179]
- Zhu K, Zhang X D, Chen G Z, Tan X L, Liao P Y, Wu H Y, Cui X J, Zuo Y N and Lyu Z Y. 2021. Single object tracking in satellite videos: deep Siamese network incorporating an interframe difference centroid inertia motion model. Remote Sensing, 13(7): #1298 [DOI: 10.3390/rs13071298]
- Zhu Z, Wang Q, Li B, Wu W, Yan J J and Hu W M. 2018. Distractor-

aware Siamese networks for visual object tracking//Proceedings of the 15th European Conference on Computer Vision. Munich, Germany: Springer: 103-119 [DOI: 10.1007/978-3-030-01240-3_7]

Zhu Z Y, Hu Q W, Wang Y and Ai M Y. 2021. Research on real-time vehicle tracking method based on Jilin-1 video satellite imagery. Journal of Wuhan University of Technology (Transportation Science Engineering), 45(4): 655-660 (朱泽尧, 胡庆武, 王玥, 艾明耀. 2021. 基于吉林一号视频卫星影像的实时车辆跟踪方法研究. 武汉理工大学学报(交通科学与工程版), 45(4): 655-660) [DOI: 10.3963/j.issn.2095-3844.2021.04.008]

作者简介

高桃峰,男,高级工程师,主要研究方向为空一天一地多源协同交通工程勘测与应用。E-mail: 407000225@qq.com

何银鑫,通信作者,男,工程师,主要研究方向为深度学习、目标跟踪、遥感信息智能处理与应用。

E-mail: heyinxin@my.swjtu.edu.cn

马学良,男,高级工程师,主要研究方向为摄影测量与遥感。

E-mail: 70494669@qq.com

卢自来,男,正高级工程师,主要研究方向为内河航道及公路工程测量。E-mail: 454653012@qq.com

彭世勇,男,助理工程师,主要研究方向为计算机视觉、图像处理和信息融合。E-mail: pengshiyong@my.swjtu.edu.cn

刘洋,男,助理工程师,主要研究方向为图像匹配与配准、遥感信息智能分析与应用。

E-mail: ly895488739@my.swjtu.edu.cn