引文格式:董燕妮,张乐飞,杜博,等.遥感视频单目标跟踪:进展与展望[J].武汉大学学报(信息科学版),2026,51(10):2180-2195.DOI:10.13203/j.whugis20260276

Citation:DONG Yanni,ZHANG Lefei,DU Bo,et al.Single-Object Tracking in Remote Sensing Video: Advances and Prospects [J].Geomatics and Information Science of Wuhan University,2026,51(10):2180-2195.DOI:10.13203/j.whugis20260276

遥感视频单目标跟踪:进展与展望

董燕妮[1] 张乐飞[2] 杜 博[2] 张良培[3]

1 武汉大学资源与环境科学学院,湖北 武汉,430079

2 武汉大学计算机学院,湖北 武汉,430072

3 武汉大学测绘遥感信息工程全国重点实验室,湖北 武汉,430079

摘 要:遥感视频单目标跟踪通过持续获取运动目标的轨迹与状态变化,推动遥感分析由静态识别走向动态过程理解。以无人机视频、卫星视频和高光谱视频中的单目标跟踪为对象,系统介绍观测平台与数据获取、训练数据来源、数据集及评价协议,梳理3个研究分支的方法演进,并从目标表征、时序状态建模、多源信息利用以及跨域适配与高效部署等方面归纳共性技术问题。分析表明,相关研究已由通用视觉模型的直接迁移逐步转向面向不同观测平台、成像方式和连续观测特点的针对性设计,但专用训练基础尚不充分,长期状态维护的可靠性不足,多源输入适配与实际平台运行之间仍缺少协同考虑。最后,从遥感视频数据与训练基础、序列级目标状态建模以及多源输入适配与端侧运行3个方面进行展望。

关键词:遥感视频;单目标跟踪;目标状态建模;多源信息融合

中图分类号:P237 **文献标识码**:A **收稿日期**:2026-07-21

DOI:10.13203/j.whugis20260276 **文章编号**:1671-8860(2026)10-2180-16

Single-Object Tracking in Remote Sensing Video: Advances and Prospects

DONG Yanni[1] *ZHANG Lefei*[2] *DU Bo*[2] *ZHANG Liangpei*[3]

1 School of Resources and Environmental Sciences, Wuhan University, Wuhan 430079, China

2 School of Computer Science, Wuhan University, Wuhan 430072, China

3 State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China

Abstract: Single-object tracking in remote sensing video continuously estimates the spatial state of a specified moving target, thereby capturing its trajectory and state evolution over time and extending remote sensing analysis from static object recognition to dynamic process understanding. Compared with conventional single-object tracking in natural scenes, single-object tracking in remote sensing video faces additional challenges arising from platform motion, long-range imaging, small target size, limited spatial detail, complex background interference, and heterogeneous spectral observations. The methodological development is therefore shaped by observation platform, imaging mechanism, data availability, and onboard computing resource. Single-object tracking in unmanned aerial vehicle videos, satellite videos, and hyperspectral videos is systematically reviewed in terms of observation platforms and data acquisition, training data sources, benchmark datasets, evaluation protocols, and model development. For unmanned aerial vehicle videos, the methodological evolution has broadly followed the development of generic visual tracking, progressing from online discriminative learning and Siamese matching to Transformer-based interaction and one-stream joint encoding. The related methods increasingly emphasize target representation under aerial views, temporal context modeling, interference suppression, and efficient token processing. Single-object tracking in satellite video has gradually moved beyond the direct transfer of generic trackers. Early studies mainly focused on search-region adjustment, template maintenance, motion constraints, and state repre-

基金项目:国家自然科学基金(U2541203)。

第一作者:董燕妮,博士,教授,研究方向为遥感图像智能解译。dongyanni@whu.edu.cn

通信作者:杜博,博士,教授。dubo@whu.edu.cn

sentation, whereas more recent methods incorporate historical observations into deep matching, preserve high-resolution features for small targets, enhance target-aware interactions, maintain historical states, and explicitly estimate target orientation. Single-object tracking in hyperspectral video seeks to convert spectral responses into discriminative target representations. Representative strategies include explicit material feature construction, band grouping and pseudo-color transformation, joint spatial-spectral modeling, and prompt- or adapter-based transfer from pretrained visual backbones. Several common technical issues arise across these research branches. Small targets and repeated downsampling operations weaken target-specific details, while background clutter and visually similar instances increase the risk of target-distractor confusion. Historical templates, temporal tokens, trajectories, and memory states improve adaptation to appearance and motion changes. However, unreliable predictions can accumulate and propagate over long sequences. Target state modeling therefore needs to jointly organize geometric states, appearance cues, motion trends, visibility, and prediction confidence while controlling error propagation during recursive updates. Multi-source information fusion provides complementary spectral, thermal, material, and semantic cues, but band redundancy, noise, spatial misregistration, missing modalities, and limited training data complicate representation learning and feature interaction. Domain shifts induced by changes in platforms, sensors, illumination, weather, and geographic environments further challenge model generalization. Practical deployment requires joint consideration of tracking accuracy, computational latency, memory consumption, parameter adaptation costs, and robustness to changing inputs. The existing research has shifted from the direct transfer of generic visual models toward specialized designs tailored to different observation platforms, imaging mechanisms, and continuous-observation characteristics. Nevertheless, the dedicated training resources remain limited, long-term target state maintenance remains vulnerable to error accumulation, and multi-source input adaptation has yet to be coordinated with efficient operation on real remote sensing platforms. Future research should strengthen remote sensing video data and training foundations, advance sequence-level target state modeling, and coordinate multi-source input adaptation with resource-aware edge deployment. These advances will help transform continuous remote sensing observations into reliable target trajectories and state information for traffic monitoring, maritime surveillance, emergency response, and dynamic scene understanding.

Key words: remote sensing video; single-object tracking; target state modeling; multi-source information fusion

随着数字化治理、智能交通和应急响应等应用需求不断增长,现代社会对空间信息连续性和时效性的要求日益提高。相较于单幅影像,遥感视频能够以较高频率持续记录特定场景,呈现车辆、舰船和飞机等目标的位置变化与运动过程,为交通运行、海上监管、灾害处置和重点区域巡查中的事件研判提供更加完整的信息[1-3]。

遥感视频单目标跟踪以通用视觉跟踪的任务框架和基本方法为基础[4-5]。受观测平台和成像方式影响,相关研究逐步形成了无人机(unmanned aerial vehicle,UAV)视频、卫星视频和高光谱视频3个主要分支[6-8]。UAV视频和卫星视频主要按观测平台区分,高光谱视频主要按成像模态区分。不同数据在观测视角、目标尺度、运动特征和信息维度上存在差异,各分支的方法研究也围绕相应的数据特点展开。对该领域的梳理既需呈现各分支的方法演进,也需在统一任务框架下归纳共性技术问题。

因此,本文从观测平台与数据获取、训练数据来源、数据集及评价协议等方面介绍遥感视频单目标跟踪的研究基础,分别梳理3个分支的方法演进,并围绕目标表征与背景抑制、时序状态建模与更新、光谱信息利用与多源融合、跨域适配与高效部署等共性技术问题展开分析,最后讨论遥感视频数据与训练基础、序列级目标状态建模以及多源输入适配与端侧运行等发展方向。

1　遥感视频单目标跟踪研究概述

遥感视频单目标跟踪面向连续遥感影像序列,旨在利用首帧标注给出的目标初始位置、尺度和外观参照,在后续帧中持续定位同一目标,

并输出其位置、尺度、方向或轮廓等状态[4,7,9-10]。给定视频序列$\{I_t\}_{t=1}^{T}$及初始帧目标状态s_1,遥感视频单目标跟踪可形式化为:

$$\hat{s}_t=\arg\max_{s\in\Omega_t} f\left(\Phi(I_t,s),\Psi(I_1,s_1)\right) \quad (1)$$

式中,$\hat{s}_t$表示第t帧估计得到的目标状态;s表示待评估的目标状态参数;Ω_t表示第t帧的候选搜索空间;f表示候选状态与目标条件之间的一致性判别函数;$\Phi(I_t,s)$表示当前帧候选状态的视觉表征;$\Psi(I_1,s_1)$表示由首帧目标标注生成的条件表征。该形式概括了遥感视频单目标跟踪的3个基本环节,即目标条件建模、候选状态搜索和目标一致性判别。这些环节建立在具体的观测平台与成像数据之上,其方法设计与性能评价还受到训练样本、标注方式和基准协议的影响。

1.1 观测平台与数据获取

典型遥感视频观测平台与成像载荷实例如表1所示,其中,图像包括红绿蓝(red green blue,RGB)、可见光(visible spectrum,VIS)、近红外(near infrared,NIR)和红光-近红外(red and near infrared,RedNIR)。由表1可知,UAV视频由低空UAV搭载可见光相机获取,也可配置多光谱或红外载荷。低空近距成像能够保留较多空间细节,使目标轮廓、局部纹理和邻近背景较为清晰;飞行高度、姿态、航向和云台角度变化会改变目标尺度、视角和相对位置,并引入明显的背景运动。此外,受机载端存储、传输和计算资源限制,数据获取与模型推理还需满足连续成像和实时处理要求[2,6]。卫星视频由星载光学视频成像载荷获取,主要记录地表目标在俯视或近俯视视角下的短时运动。天基观测覆盖范围大、视角相对稳定,可在道路、水域、港口、机场等较大场景内保持对区域的连续成像。但远距离成像会压缩目标像素占比,局部纹理和边缘细节有限,目标定位更依赖帧间位移、运动连续性和周围背景结构[1,7]。

表1 典型遥感视频观测平台与成像载荷实例

Table 1 Typical Observation Platforms and Imaging Payloads of Remote Sensing Video

观测类型	平台/载荷实例	典型参数	输入形态
UAV视频	低空UAV可见光/多光谱/红外载荷	常见1 080p/4K视频,30~60帧/s,低空斜视或近俯视连续成像	低空斜视/俯视视频序列,尺度、视角和背景变化明显
卫星视频	吉林一号视频卫星	星载光学视频载荷,约25帧/s,地面分辨率约1.12 m	天基俯视RGB视频序列
	SkySat / SkySat Video	星载全动态视频载荷,30~120 s视频片段,单相机连续采集	高分辨率短时对地视频序列
	珠海一号OVS-1	星载高分辨率视频载荷,20帧/s,地面分辨率约1.98 m	天基彩色视频序列
	VividX2/Carbonite-2	小卫星全色/彩色视频载荷,约1 m级高分辨率成像和高清视频	全色或彩色对地视频序列
空间站视频	UrtheCast Iris视频相机	国际空间站搭载4K视频相机,约3帧/s,地面分辨率约1 m/像素	空间站对地视频序列
高光谱视频	XIMEA/IMEC可见光快照式相机	快照式可见光高光谱相机,16波段,460~600 nm,最高120个数据立方体/s	VIS光谱-空间视频序列
	XIMEA/IMEC NIR、RedNIR快照式相机	快照式NIR/RedNIR高光谱相机,NIR 24波段,665~960 nm,RedNIR 15波段,600~860 nm	NIR/RedNIR光谱-空间视频序列
	IMEC CMV2K-SSM 5×5 NIR快照式相机	快照马赛克高光谱相机,24波段,665~960 nm	NIR光谱-空间视频序列

高光谱视频在连续成像过程中同步记录多个窄波段的光谱响应,使跟踪输入由二维外观信息扩展为光谱-空间联合序列。充足的光谱信息能够为弱纹理或背景相似条件下的目标区分提供补充线索,但波段冗余、谱间相关性、噪声波段和高维数据会增加存储、传输和推理开销;波段设置、成像帧率和辐射质量也会影响后续特征压缩和目标判别的稳定性[8,11]。

连续影像流经序列截取、目标实例筛选、首帧初始化、逐帧状态标注和属性整理后,形成围绕单个目标组织的跟踪样本。空间分辨率、帧率、波段设置和观测稳定性影响目标可标注性、状态标注连续性及序列间可比性,并由此约束数据集构建与评价协议。

1.2　数据集与评价协议

现有遥感视频单目标跟踪基准在规模、模态和划分方式上存在明显差异，典型基准及相关数据集见表2，其中，SOT（single-object tracking）为单目标跟踪，HSI（hyperspectral imaging）为高光谱成像，VISO数据集面向卫星视频运动目标检测与多目标跟踪，其轨迹标注可用于构造SOT训练样本。

由表2可知，UAV视频单目标跟踪数据集最为丰富，已由UAV123、UAV20L、DTB70和VisDrone-SOT系列等中小规模评测基准，发展到WebUAV-3M等具有大规模视频、丰富类别和逐帧标注的数据集[6,12-14]。文献[15]构建了面向高空UAV平台的可见光-红外多模态跟踪数据集HiAl，为高空小目标跟踪及多模态跟踪方法评价提供了数据基础。卫星视频数据集以SatSOT、SV248S和OOTB等为代表，目标主要包括车辆、舰船、飞机和列车，序列规模和类别覆盖仍相对有限[7,16-18]。高光谱视频单目标跟踪数据集由早期的同场景高光谱-彩色视频配对序列，逐步发展为具有训练、验证和测试划分的HOTC系列，波段范围由VIS扩展至NIR和RedNIR；BihoT等数据集进一步纳入空间外观相近、光谱响应存在差异的伪装目标[11,19-20]。

表2　典型遥感视频单目标跟踪基准及相关数据集

Table 2　Typical Benchmarks and Related Datasets for Remote Sensing Video Single-Object Tracking

研究分支	数据集	数据规模	数据用途	研究定位
UAV视频	UAV123（2016年）	123个序列，约11万帧	主要用于测试	低空视角通用评测
	UAV20L（2016年）	20个长序列	主要用于长时跟踪测试	UAV长时跟踪评测
	DTB70（2017年）	70个序列，约1.6万帧	主要用于测试	平台运动与尺度变化评测
	VisDrone-SOT系列（2018—2020年）	按届提供视频序列	设有训练、验证和测试划分	密集城市场景与小目标跟踪
	WebUAV-3M（2023年）	4 500个视频，约330万帧	用于训练和测试	大规模UAV专用训练
	HiAl（2025年）	150对可见光-红外视频，超过1.2万帧对	主要用于跟踪方法评价，也用于方法重训练	高空UAV多模态小目标跟踪
卫星视频	SatSOT（2022年）	105个序列，27 664帧	主要用于测试	卫星视频单目标跟踪
	SV248S（2022年）	248个序列	主要用于测试	卫星视频小目标评测
	OOTB（2024年）	110个序列，29 890帧	主要用于测试	旋转目标跟踪评测
	VISO（2021年）	47段视频，约164.6万个标注实例、3 711条轨迹	用于训练和测试	动目标检测与多目标跟踪，可构造目标跟踪样本
高光谱视频	HSI Tracking Benchmark（2020年）	35组高光谱-彩色视频序列	主要用于测试	高光谱与彩色跟踪对照
	HOTC系列（2020年至今）	按届扩展训练集和测试集，覆盖VIS、NIR和RedNIR	用于高光谱跟踪评测	高光谱训练与挑战评测
	BihoT（2025年）	49个序列，41 912帧	27个序列训练，22个序列测试	相似外观与伪装目标跟踪

由于早期卫星视频和高光谱视频数据集多用于性能评测，专用训练样本有限，跟踪器通常先在LaSOT、GOT-10k、TrackingNet和COCO等通用视觉数据集上完成离线训练，再迁移至遥感视频测试[21-24]。这些数据集覆盖较丰富的目标类别、尺度变化、遮挡和运动模式，但其观测视角、目标尺度和背景分布与遥感视频存在差异，直接迁移的性能容易受到限制。随着专用训练数据逐步增加，基于遥感序列的微调、专用数据训练和联合采样逐渐得到应用。WebUAV-3M等大规模数据集可用于UAV专用训练，也可与通用视觉数据联合采样；卫星视频专用训练样本仍较有限，现有方法多沿用通用预训练参数，再利用少量卫星序列微调特征适配与目标定位相关模块；高光谱视频单目标跟踪方法多继承RGB图像跟踪器的主干参数，再利用专用训练集学习波段编码、光谱-空间融合或参数适配模块，以减轻有限样本和高维输入带来的训练压力。

评测通常采用一次通过评估协议（one-pass evaluation，OPE），测试期间不重置跟踪器[9]。现有基准多采用水平矩形框，根据预测框与真实框的交并比绘制成功率曲线，并以曲线下面积（area under the curve，AUC）评价目标重叠程度，以中心位置精度和归一化精度评价定位偏差；部分卫星视频数据采用旋转框，并以旋转交并比同时度

量目标的位置、尺度和方向[7]。跟踪速度通常以每秒处理帧数衡量。卫星视频中的目标尺寸较小,像素级偏移会显著影响交并比和中心位置精度;高光谱视频跟踪方法的波段数量、输入形式及光谱预处理也会改变精度与速度,因此实验结果需同时注明输入尺寸、波段设置、测试硬件和计时范围[11]。

1.3 已有综述与本文定位

现有相关综述已分别从遥感视频跟踪整体发展、UAV视频与卫星视频单目标跟踪、高光谱视频目标跟踪以及Transformer方法应用等角度开展总结[1-3,7-8,11,25]。为了明确本文的研究定位,表3从研究对象、综述重点和组织方式3个方面对相关工作进行比较。

表3 现有遥感视频目标跟踪相关综述对比

Table 3 Comparison of the Existing Reviews on Object Tracking in Remote Sensing Video

文献	研究对象	综述重点	组织方式
LIU等[2]	遥感视频跟踪	方法进展、数据集、评价指标与开放问题	按模型驱动、特征驱动和深度学习等技术范式组织
JIAO等[3]	遥感视频检测与跟踪	Transformer结构及其在目标检测与跟踪中的应用	按Transformer核心设计与任务架构组织
陈琳等[26]	UAV视频目标跟踪	跟踪方法、数据集与评价指标	以相关滤波和深度学习方法为主线
LI等[1]	卫星视频智能处理	目标检测、目标跟踪、场景分类、图像分割与超分辨率等任务	按卫星视频智能处理任务分类梳理
CHEN等[7]	卫星视频单目标跟踪	算法演进、评价基准与旋转目标跟踪	将方法综述与定向目标跟踪基准构建相结合
李洋帆等[27]	卫星视频单目标跟踪	相关滤波、深度学习方法及数据集	按方法范式比较技术特点与适用性
许廷发等[28]	光谱扩展单目标跟踪	可见光-热红外、多光谱和高光谱目标跟踪	按光谱输入形式与融合方式组织
QIAN等[8]	高光谱视频目标跟踪	跟踪方法、数据集及光谱信息利用	按特征表示与跟踪方法类别组织
WANG等[11]	深度高光谱目标跟踪	深度方法分类、实验比较与发展展望	以深度特征建模与实验评测为主线
本文	遥感视频单目标跟踪	UAV、卫星和高光谱视频中的方法进展及共性技术问题	先按三类研究分支纵向梳理,再按共性技术维度横向归纳

相较于现有综述,本文不局限于单一平台、单一成像模态或特定模型体系,而以遥感视频单目标跟踪为统一任务框架,对UAV视频、卫星视频和高光谱视频3个研究分支的方法演进进行纵向梳理,并从目标表征、时序状态建模、多源信息利用、跨域适配和高效部署等方面对共性技术问题进行横向分析。

2 遥感视频单目标跟踪研究进展

本文依次介绍UAV视频、卫星视频和高光谱视频单目标跟踪的代表性方法,相关发展脉络如图1所示。

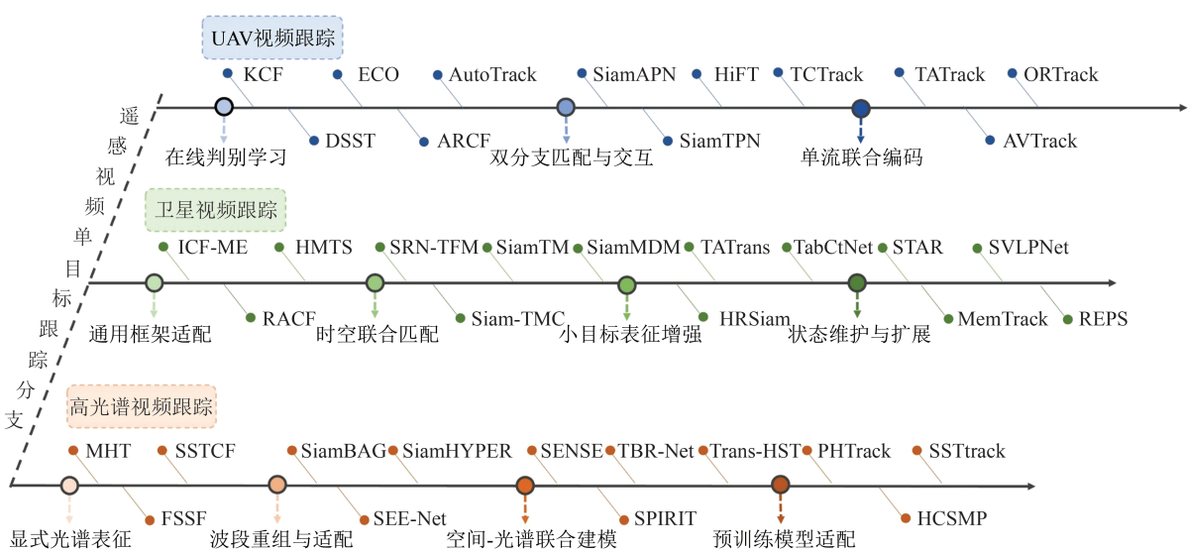


图1 遥感视频单目标跟踪各研究分支发展脉络

Fig. 1 Development of Research Branches of Single-Object Tracking in Remote Sensing Video

2.1　UAV视频单目标跟踪

UAV视频单目标跟踪是遥感视频单目标跟踪中发展较早、方法积累较为丰富的分支。早期研究大多直接采用通用视觉跟踪器,航拍序列主要用于方法测试与性能比较。随着专用基准、应用研究和模型设计逐渐增多,研究重点开始由通用方法的适用性验证转向面向UAV视频的专门建模。其技术发展总体承接了通用视觉跟踪的演进,先后采用在线判别学习、双分支深度匹配和Transformer联合编码等基本框架。

在线相关滤波是早期UAV视频单目标跟踪的主要方法基础。核化相关滤波、判别式尺度空间跟踪器和高效卷积算子跟踪等通用跟踪器利用首帧样本建立判别模型,并在视频过程中持续更新,为这一时期的精度与速度比较提供了基本参照[29-31]。在此基础上,后续工作开始针对航拍序列调整响应约束与更新策略。ARCF通过帧间响应约束减少异常定位[32];AutoTrack则根据响应变化调节空间约束与模型更新,以减轻背景扰动和不可靠观测的影响[33]。针对低照度和光照变化场景,将自适应图像增强与正则化相关滤波相结合,并根据光照质量动态调节时空约束,从而提高复杂光照条件下UAV目标跟踪的稳定性[34]。

大规模自然视频数据为深度跟踪器的离线训练提供了基础,UAV视频单目标跟踪也开始采用模板与搜索区域分支完成通用匹配。SiamFC和SiamRPN等框架将相似性计算、候选生成、目标分类和边界框估计纳入深度网络,确立了双分支跟踪的基本形式[35-36]。面向航拍目标的候选预测与轻量推理需求,SiamAPN以自适应候选生成改进传统Siamese预测流程[37];SiamTPN则在轻量骨干中组织多层特征,兼顾目标表征与推理效率[38]。Transformer被引入双分支框架后,HiFT利用层级相似信息加强模板与搜索区域之间的交互[39]。在空间交互逐渐加深的同时,TCTrack将连续帧上下文接入特征提取与相似性计算,使历史观测能够参与当前目标定位[40]。

双分支框架通常由两个分支分别对模板和搜索区域进行特征编码,再通过后续模块完成二者的匹配或特征交互。单流Transformer则将二者组织为统一序列,在同一视觉Transformer(vision transformer,ViT)骨干中同步完成特征提取与跨区域交互,使UAV专用设计进一步进入联合编码阶段。在这一框架下,AVTrack、ORTrack和TATrack等方法围绕航拍目标表征、遮挡干扰和计算效率,改进骨干内部的信息处理,并将目标相关线索、历史观测或有效Token选择纳入统一编码过程[41-43]。

2.2　卫星视频单目标跟踪

卫星视频单目标跟踪长期依托通用视觉跟踪框架。早期工作多将在自然视频数据上训练的模型直接用于卫星序列,并根据天基连续观测中的目标尺度、位移和方向变化,对搜索范围、模板更新与输出状态作出调整。随着连续帧利用方式增多,历史观测逐渐参与模板构建和目标匹配,相关设计也延伸至骨干特征提取与历史信息建模。

目标位移是早期方法直接利用的连续信息之一。ICF-ME根据相邻帧的位移变化确定当前帧候选区域,使搜索窗口与目标运动范围相适应[44];RACF在位置和尺度之外加入旋转参数,用于描述卫星视角下的目标朝向[45];HMTS则保存历史帧中的目标外观,为后续模板更新提供参考[46]。融合HOG、灰度和高斯曲率等特征,并结合响应置信度驱动的自适应更新与卡尔曼滤波预测,提高了复杂背景和遮挡条件下卫星视频目标跟踪的稳定性[47]。这些改动主要作用于候选搜索、状态参数和模板维护,特征提取与响应定位仍沿用相关滤波框架。在深度匹配模型应用后,运动变化与历史观测直接参与到模板和搜索区域的匹配计算中:SRN-TFM联合外观相似性与运动拟合确定目标位置[48];Siam-TMC也将帧间运动用于当前预测[49]。围绕模板信息的利用,SiamTM在多个特征层级建立模板与搜索区域之间的联系[50],SiamMDM则在首帧模板之外持续加入后续观测,使当前定位能够综合初始外观、近期运动和更新后的目标表征[51]。针对遮挡场景,根据遮挡程度联合采用模板更新、双向验证和轨迹预测,以维持卫星视频中的连续目标跟踪[52];SiamTITP则将动态模板中的时序信息与轨迹预测相结合,增强遮挡条件下的目标定位能力[53]。

随后,研究者更多关注面向卫星小目标的特征提取。HRSiam在骨干中保持较高的空间分辨率,使小尺度目标在深层特征中保留更多位置信息[54]。TATrans将目标相关表示接入模板与搜索区域之间的Transformer交互,使首帧目标条件参与搜索特征的形成[55];TabCtNet则在同一网络中组织局部目标特征与全局场景关系[56]。

近期方法对跟踪状态的利用和描述也更为丰富。例如,STAR通过时空融合连接连续帧观测[57],MemTrack将历史目标表征保存在可更新的记忆状态中[58],SVLPNet利用已有目标位置引导当前预测[59]。除利用跨帧信息外,REPS又将目标方向纳入预测,使跟踪结果由水平框扩展为旋转框,用于描述具有明显朝向的车辆、舰船和飞机等目标[60]。

2.3 高光谱视频单目标跟踪

高光谱视频在空间结构和时间变化之外,还记录了目标及背景在不同谱段上的响应。相关方法的发展主要围绕光谱信息如何由数据属性转化为目标跟踪表征展开,光谱信息的利用方式先后表现为显式光谱表征、波段重组与适配、空间-光谱联合建模和预训练模型适配。

早期方法多将材料光谱响应作为目标外观的补充。MHT以目标材料在不同谱段上的响应构建目标描述,用于区分空间外观相近的目标与背景[19];FSSF结合空间结构与光谱响应,并将构造的特征用于相关滤波定位[61];SSTCF又引入帧间变化,使响应计算同时参考空间、光谱和时间信息[62]。这一阶段的光谱表征多在进入跟踪器前完成构造,材料属性主要用于补充目标判别,定位和更新仍沿用在线滤波流程。

三通道深度跟踪器得到广泛应用后,波段分组和伪彩色转换成为高光谱序列接入现有网络的主要方式,研究重点落在波段的选择与组织。SiamBAG将波段划分为若干组,并利用Siamese网络完成模板匹配[63];SEE-Net根据不同谱段对目标描述的贡献组合多组伪彩色跟踪结果[64];SiamHYPER则通过跨模态适配,将RGB跟踪器中的表征能力迁移到高光谱序列[65];利用光谱分割提取目标光谱信息,通过光谱角距离进行降维,并融合浅层外观特征与深层语义特征完成目标定位与跟踪[66]。

另一类工作把谱段之间的联系放入特征提取和目标匹配过程。SENSE将材料线索与运动信息结合,用于判断当前目标位置[67];SPIRIT在模板与搜索区域之间组织谱段贡献和空间-光谱特征,并利用动态模板补充目标外观[68]。TBR-Net借助Transformer建立远距离谱段之间的联系,同时结合模板与搜索区域的关系精炼特征[69];Trans-HST则在波段组之间传递信息,并融合不同层次的空间-光谱表示[70]。此时,各谱段的作用不再由固定分组完全决定,而会结合指定目标及当前搜索区域进行调整,光谱响应也成为模板匹配中的目标特征。

预训练视觉骨干为高光谱视频单目标跟踪提供了新的模型构建路径,光谱信息可通过提示和适配模块接入通用跟踪模型,高光谱专用学习也相应集中在少量新增参数中。PHTrack通过提示向通用网络传递光谱条件,使高光谱信息参与目标表示的形成[71]。类似地,ADTrack-HOT通过提示学习引入高光谱信息,并利用自适应丢弃机制减少冗余光谱信息[72];HCSMP将跨模态光谱补充与历史提示结合,利用先前观测更新当前表征[73];SSTtrack设置不同适配器,分别处理光谱、空间和时间信息[74]。

综上,表4选取各研究分支的代表性方法,对其技术路线、主要特点和局限性进行比较。

表4 代表性遥感视频单目标跟踪方法对比

Table 4 Comparison of Representative Methods of Single-Object Tracking in Remote Sensing Video

研究分支	代表方法	技术路线	主要特点	主要局限
UAV视频	ARCF[32]	响应约束相关滤波	抑制异常响应	手工特征表征有限
	SiamAPN[37]	孪生候选生成	兼顾候选质量与实时性	时序信息利用不足
	TCTrack[40]	时序上下文建模	联合增强特征与响应	长期状态维护有限
	TATrack[43]	目标感知单流Transformer	强化目标相关特征	显式时序建模不足
卫星视频	ICF-ME[44]	运动辅助相关滤波	运动估计引导搜索	依赖短时运动连续性
	HRSiam[54]	高分辨率孪生特征建模	保留小目标空间细节	时序信息利用不足
	SiamMDM[51]	动态模板与运动建模	联合维护外观与运动	错误模板可能累积传播
	MemTrack[58]	时序记忆Transformer	利用长期历史表征	记忆易受错误观测污染
	REPS[60]	旋转等变孪生网络	支持旋转状态预测	长期状态建模不足
高光谱视频	MHT[19]	材料感知光谱-空间表征	利用材料光谱区分目标	依赖解混与手工特征
	SiamBAG[63]	波段分组与孪生匹配	复用RGB跟踪模型	连续光谱关系受压缩
	SPIRIT[68]	光谱感知交互与动态模板	联合光谱关联与外观更新	动态模板受误差影响
	SSTtrack[74]	光谱-空间-时间条件建模	统一建模多维条件	跨波段配置仍需适配

3　遥感视频单目标跟踪关键问题

上述3个研究分支虽面向不同的观测平台和数据形式,具体方法各有侧重,但其设计均围绕遥感视频中目标信息的表征、保持与利用展开,实际应用还受到数据分布和平台资源条件的影响。基于这些共性,本文从跨分支视角梳理相关方法的技术联系与主要问题,以遥感视频单目标跟踪方法为重点,并参考通用视觉跟踪领域的相关研究。

3.1　目标表征与背景抑制

遥感视频中的目标通常尺寸较小,可供定位和辨识的纹理、边缘与结构信息有限,在多次下采样和深层编码过程中容易进一步衰减。围绕这一问题,HRSiam通过并行高分辨率分支维持较细的空间结构[54];HiFT对不同层级特征形成的相似图进行交互,使浅层位置细节与深层语义信息共同参与匹配[39];SmallTrack则利用小波池化保留不同频率成分,并结合图增强分类改善小目标响应[75];融合多阶段深度特征,并以非对称-膨胀卷积增强目标表征[76]。这些方法分别从空间分辨率保持、层级特征互补、频率信息保留和特征表达增强等方面改善目标表征,但仍难以仅依靠表观信息解决相似目标之间的身份区分问题。

为增强对指定目标的约束,首帧信息由独立模板逐渐融入目标模型和搜索特征的构建过程。ToMP依据目标样本预测专属的分类与边界框回归模型,使首帧条件直接参与候选状态估计[77]。在此基础上,模板与搜索区域之间的交互进一步前移至骨干网络:MixFormer通过混合注意力同步完成特征提取和跨区域信息交互[78],OSTrack则将模板与搜索区域组织为统一序列,在单流骨干中完成联合编码[79]。针对卫星视频中目标像素占比低、表观细节有限的问题,文献[55]提出TATrans,利用目标感知增强突出与指定实例相关的特征。

随着首帧条件对搜索区域的约束不断加深,目标相关信息得到强化,模板中的背景与低质量观测也可能随之传播并产生异常响应。早期研究主要从响应图和在线模型层面限制其影响。ARCF利用相邻帧响应图之间的变化约束异常峰值,降低局部背景扰动造成的定位偏移[32];AutoTrack进一步根据响应变化自适应调节空间正则项和模型更新强度,避免不可靠观测持续污染目标模型[33]。进入Transformer框架后,干扰处理也逐渐由输出端的响应校正前移至特征编码过程。Aba-ViTrack通过背景感知的Token计算减少无关区域参与后续交互,使表征更集中于潜在目标区域[80]。

当搜索区域中存在多个外观相近的实例时,仅削弱背景响应仍难以确认目标身份,干扰处理由此延伸至候选之间的关系判别。KeepTrack不再仅依据单个候选的分类响应作出判断,而是对目标候选与潜在干扰物进行关联,根据候选在连续观测中的对应关系维持指定实例[81]。目标表征由此不仅需要区分前景与背景,还需保留能够辨别相似实例的身份线索;单帧表征不足时,则需要跨帧状态提供补充。

3.2　时序状态建模与更新

视频中的目标状态通常具有时间连续性,前序帧能够从几何和外观两个方面为当前预测提供依据。其中,位置、尺度和方向等几何状态可约束目标的运动范围,外观信息则有助于适应形变、视角和成像条件变化。由于历史模板、轨迹和记忆内容通常依赖前序预测生成,错误结果一旦被保留,便可能持续影响后续定位。

较直接的时序利用方式是显式引入连续帧特征、动态模板或运动轨迹。TCTrack在特征提取和相似图修正阶段融入时序上下文,使前序信息同时参与目标表征与响应调整[40]。另一类方法是联合维护目标的外观与运动状态。MBLT利用前序运动信息约束当前目标位置[82];SiamMDM则结合动态模板、响应图融合和轨迹约束,在适应外观变化的同时保持运动连续性[51]。随着时序跨度扩大,分别由模板、特征、响应图和轨迹承载的历史状态也更难统一组织和传递。

为加强跨帧状态的传递,后续方法开始以提示、Token或Query等紧凑表示承载历史信息。HIPTrack根据历史前景区域和目标特征构建提示,将筛选后的目标线索传递至当前帧[83];ODTrack利用在线传播的时序Token保存定位相关状态,使其随视频过程持续更新[84]。进一步地,AQATrack通过滑动窗口内的自回归Query描述短期外观变化[85];ARTrackV2则将目标位置和外观纳入统一的自回归过程[86]。历史信息从辅助性的多帧输入逐渐转化为直接参与当前预测的递推状态。

递推状态对前序结果的依赖使得历史质量直接影响后续定位。DTPTrack先评估不同历史

状态的可靠性,再将有效信息合成为时序先验,减少低质量预测对当前表征的干扰[87]。当漂移或目标丢失已经发生时,仅筛选历史状态已难以恢复目标。RTracker利用正负样本树状记忆分别保存目标与干扰物证据,并在跟踪失效后重新启动目标定位[88]。前者限制不可靠历史信息的使用,后者则在递推中断后重新建立目标状态。

在筛选和修正历史状态之外,时序信息还可用于预测目标的后续变化。文献[89]提出DreamTrack,根据历史运动和外观生成多种可能的未来状态,并将其作为后续搜索与定位的先验,再结合当前帧观测完成校正。未来预测进一步扩大了历史信息的作用范围,其可靠性仍取决于前序状态质量以及多种候选未来的筛选。

3.3 光谱信息利用与多源融合

现有跟踪器多基于三通道RGB视频构建,而遥感平台还可获取红外、多光谱和高光谱等观测。这些数据在波段数量、响应机理、空间分辨率和配准关系等方面存在差异,难以直接沿用面向RGB输入设计的表征方式。早期高光谱视频单目标跟踪通常先调整输入形式,SiamBAG通过波段分组构造多组三通道图像,以复用已有Siamese网络[63]。该方式降低了高光谱数据接入RGB框架的难度,但波段间的连续关系和材料信息未能得到充分利用。

在显式材料特征的基础上,后续方法开始关注建模谱段关系及其与空间结构的联系。MHT[19]从材料属性出发,利用多维梯度的光谱-空间直方图和材料组分丰度构建目标表征,为空间外观相近的目标与背景区分提供补充信息。SPIRIT进一步在模板与搜索区域之间建立光谱感知交互[68],SSF-Net则结合空间-光谱特征与光谱角一致性完成目标匹配[90]。面向多光谱UAV视频,UNTrack利用光谱提示组织初始模板与连续搜索帧,联合编码光谱、空间和时序信息[91]。随着光谱信息深入特征提取与匹配过程,波段冗余、噪声以及背景光谱混入也更容易影响目标表征。

多波段联合建模通常需要专用网络和训练数据,但现有高光谱及多源跟踪数据规模有限,针对不同输入重新训练完整模型的成本较高。特征对齐为复用RGB预训练知识提供了一种路径,Trans-DAT通过缩小RGB与高光谱表征之间的差异,将已有跟踪能力迁移至高光谱输入[92];PHTrack则以提示形式将光谱信息接入预训练骨干[71]。类似思路也被用于一般多模态跟踪,ViPT通过模态提示引入辅助观测[93];SDSTrack利用适配器和自蒸馏传递RGB主干知识,从而减少对原有网络参数的大幅调整[94]。

完成多源信息的表征适配后,各类输入在视频过程中仍未必始终可用。针对部分模态缺失,FlexTrack利用异构专家和视频级掩膜训练,根据当前可用输入选择相应的计算分支[95]。当多种输入同时存在时,其对指定目标的作用也可能不同。RAGTrack进一步引入目标文本描述和检索信息,通过目标相关Token选择与跨模态通道交换调节可见光和热红外特征,使文本描述和检索结果也能参与目标相关特征的选择[96]。对于多传感器跟踪,空间配准误差、输入质量下降和模态缺失仍会破坏特征对应关系;相关多模态检测研究通过跨模态交互与偏移引导融合缓解不同传感器图像之间的空间错位问题,为多源信息的特征对齐提供了参考[97]。当传感器配置或数据分布发生变化时,既有转换与融合机制往往还需重新适配。

3.4 跨域适配与高效部署

跟踪器由通用数据集走向实际遥感平台时,平台、地域、昼夜、天气和传感器变化会引起目标尺度、辐射特征与背景分布偏移。针对昼夜成像差异,UDAT利用无标注夜间视频构造训练样本,并通过对抗式域对齐缩小日间与夜间特征分布的差异[98]。除光照变化外,通用视频与UAV观测之间还存在明显的目标尺度偏移,Scale-AwareDA进一步将目标尺度及其分布纳入适配过程,使跨域学习更贴近UAV跟踪中的小目标特征[99]。

当目标域标注有限时,域适配还依赖可用样本的发掘与补充。SAM-DA借助分割基础模型从无标注夜间图像中挖掘潜在目标,为后续适配提供伪标注样本[100]。面向雨、雪、雾等多种天气变化,UMDATrack利用文本条件生成目标域训练视频,并通过领域适配器和目标感知置信度对齐,在共享骨干上处理不同天气域[101]。跨域适配由整体特征分布对齐逐渐扩展到目标尺度、样本质量和具体环境变化的联合处理。相关遥感图像域适应研究表明,多级空间特征与平衡测度学习可加强不同数据域之间的分布对齐,并缓解类别不平衡对模型训练的影响[102]。

模型具备目标域适应能力后,其在UAV端或星载设备上的运行仍受到算力、存储和时延约

束。早期研究主要通过轻量化网络结构降低推理成本。LightTrack利用架构搜索联合选择轻量骨干与预测头[103];HiT则采用分层轻量Transformer,并通过特征桥接保留定位所需的信息[104]。除Transformer轻量化外,还可采用视觉状态空间模型构建TMamba,在降低计算量和显存开销的同时维持目标表征与建模能力[105]。随着单流Transformer的应用,轻量化对象由网络结构进一步扩展到Token组织与计算路径。CVTTrack分别从初始模板和历史帧中提取目标Token与时序Token,减少无关状态参与后续匹配[106];UTPTrack则进一步联合裁剪搜索区域、静态模板和动态模板中的冗余Token[107]。除压缩参与交互的Token外,计算裁剪还可直接作用于主干路径,SGLATrack依据层间表征相似性动态跳过重复计算层[108]。

另一方面,模型适应新平台和新数据时的参数更新开销也会影响实际部署。针对不同场景分别进行全量微调,需要重复更新并保存完整模型参数。LoRAT将低秩更新引入跟踪骨干,仅训练少量新增参数,且不额外增加推理时延[109];SPMTrack进一步利用跟踪专用混合专家建模时空关系,同时控制可训练参数规模[110]。目前,跨域适配与高效推理仍多被分别评价,较少检验轻量模型在域偏移条件下的稳定性。实际部署还应在目标设备上联合考察跟踪精度、运行时延、资源占用和参数更新成本。

4　发展展望

综上,遥感视频单目标跟踪已在目标表征、时序状态建模、多源信息利用、跨域适配和高效部署等方面取得较多进展[2-3,7-8]。后续研究可重点关注遥感视频数据与训练基础、序列级目标状态建模,以及多源输入适配与端侧运行3个方面:

1)构建面向遥感视频的数据与训练基础。卫星视频单目标跟踪目前主要依靠通用视觉数据完成训练[51-55],高光谱视频单目标跟踪也常通过波段重组、特征迁移或参数适配复用RGB模型[63,65,71,92],面向遥感视频的训练数据和预训练基础仍不充分[2,7-8,11]。增加测试序列能够扩大方法的评价范围[14-16,20],却不能替代训练阶段对遥感小目标、平台运动和传感器成像变化的学习。数据建设因而需要逐步由测试集扩充延伸至训练资源积累,利用未标注遥感视频学习帧间对应、运动连续性和成像变化规律,并结合弱标注与可控生成补充稀缺目标和少见观测条件[98,100-101]。在此基础上,还应区分通用视觉预训练、遥感视频预训练和特定平台适配的作用,明确各阶段的数据来源及目标域样本使用方式,使方法性能与其训练基础建立清晰的对应关系。

2)由帧间匹配走向序列级目标状态建模。现有框架主要通过模板与搜索区域的匹配估计当前帧目标位置,帧间信息则以模板更新、历史特征或前序结果等形式参与后续预测。随着历史提示、时序Token和自回归Query等方法的发展,连续观测开始被用于刻画目标随时间发生的变化,跟踪建模也逐步延伸到目标状态的序列演化[83-86]。目标的外观、位置、尺度、运动趋势、可见性和预测置信度由此可以在统一的时序状态中组织,并结合平台运动和场景结构约束其变化。由于状态更新通常依赖前序预测,当前观测还需对历史信息进行检验,减少遮挡和漂移造成的误差传播,并为目标重新出现后的身份确认提供可靠依据[87-88]。随着状态信息由几何与外观扩展到目标属性和场景关系,视觉基础模型和视觉语言模型可为初始化、相似候选判别和跟踪失效恢复提供高层信息[96,100],并与候选关联和状态修正机制结合[81,88]。

3)协调多源输入适配与端侧运行。实际运行中,输入可用性、数据分布和设备资源可能同时变化,多源融合、场景适配和模型轻量化之间也会相互影响。模态缺失或波段噪声会改变特征选择与融合结果,参数适配会带来额外的训练、存储和更新开销,Token、网络层和历史状态的裁剪则可能影响遮挡后的目标恢复。因此,后续模型需要根据输入质量和可用性选择参与计算的模态与波段[63-64,95-96],结合数据分布和场景变化调整表征及参数[92,98-101,109-110],并依据设备资源调整搜索范围、Token数量、网络深度和历史信息规模[106-108]。相应实验还应在目标设备上同时考察条件变化下的跟踪精度、运行时延、存储占用和适配开销,以判断模型在实际遥感平台上的运行表现。

5　结　语

遥感视频单目标跟踪研究已由通用视觉模型的直接迁移逐步转向针对不同观测平台、成像方式和连续观测特点的专门设计。相关方法在目标表征、时序状态建模和多源信息利用等方面持续发展,跨域适配与高效部署也逐渐受到关

注,但长期状态维护、输入适配和平台运行仍多被分开考虑。随着低空遥感、视频卫星和新型光谱载荷不断拓展动态观测能力,遥感视频单目标跟踪有望为运动过程分析、异常事件识别和动态场景理解提供更加可靠的目标轨迹与状态信息。

参 考 文 献

[1] LI S Y, SUN X, GU Y F, et al. Recent Advances in Intelligent Processing of Satellite Video: Challenges, Methods, and Applications[J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2023, 16: 6776-6798.

[2] LIU F, WANG J H, JIAO L C, et al. Remote Sensing Video Tracking: Current Status, Challenges, and Future[J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2025, 18: 14338-14367.

[3] JIAO L C, ZHANG X, LIU X, et al. Transformer Meets Remote Sensing Video Detection and Tracking: A Comprehensive Survey[J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2023, 16: 18-45.

[4] SMEULDERS A W, CHU D M, CUCCHIARA R, et al. Visual Tracking: An Experimental Survey [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2014, 36(7): 1442-1468.

[5] 韩瑞泽,冯伟,郭青,等. 视频单目标跟踪研究进展综述[J]. 计算机学报, 2022, 45(9):1877-1907.
HAN Ruize, FENG Wei, GUO Qing, et al. The Research of Single Object Tracking: A Survey [J]. *Chinese Journal of Computers*, 2022, 45(9): 1877-1907.

[6] MUELLER M, SMITH N, GHANEM B. A Benchmark and Simulator for UAV Tracking[C]// The 14th European Conference on Computer Vision (ECCV), Amsterdam, The Netherlands, 2016.

[7] CHEN Y Z, TANG Y Q, XIAO Y, et al. Satellite Video Single Object Tracking: A Systematic Review and an Oriented Object Tracking Benchmark [J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2024, 210: 212-240.

[8] QIAN K, SHEN J L, WANG S Q, et al. Recent Advances in Object Tracking Using Hyperspectral Videos: A Survey[J]. *Multimedia Tools and Applications*, 2024, 83(18): 56155-56181.

[9] WU Y, LIM J, YANG M H. Object Tracking Benchmark [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, 37(9): 1834-1848.

[10] KRISTAN M, MATAS J, LEONARDIS A, et al. The Ninth Visual Object Tracking VOT2021 Challenge Results [C]//IEEE/CVF International Conference on Computer Vision, Montreal, Canada, 2021.

[11] WANG Y, LI X, YANG X, et al. Deep Feature-Based Hyperspectral Object Tracking: An Experimental Survey and Outlook [J]. *Remote Sensing*, 2025, 17(4): 645.

[12] LI S, YEUNG D Y. Visual Object Tracking for Unmanned Aerial Vehicles: A Benchmark and New Motion Models[C]//AAAI Conference on Artificial Intelligence, San Francisco, USA, 2017.

[13] FAN H, WEN L Y, DU D W, et al. VisDrone-SOT2020: The Vision Meets Drone Single Object Tracking Challenge Results[C]//The 16th European Conference on Computer Vision (ECCV), Glasgow, UK, 2020.

[14] ZHANG C H, HUANG G J, LIU L, et al. WebUAV-3M: A Benchmark for Unveiling the Power of Million-Scale Deep UAV Tracking [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023, 45(7): 9186-9205.

[15] 肖云,曹丹,李成龙,等. 基于高空无人机平台的多模态跟踪数据集[J]. 中国图象图形学报, 2025, 30(2):361-374.
XIAO Yun, CAO Dan, LI Chenglong, et al. A Benchmark Dataset for High-Altitude UAV Multimodal Tracking[J]. *Journal of Image and Graphics*, 2025, 30(2): 361-374.

[16] ZHAO M Q, LI S Y, XUAN S Y, et al. SatSOT: A Benchmark Dataset for Satellite Video Single Object Tracking[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2022, 60: 1-11.

[17] LI Y, JIAO L, HUANG Z, et al. Deep Learning-Based Object Tracking in Satellite Videos: A Comprehensive Survey with a New Dataset [J]. *IEEE Geoscience and Remote Sensing Magazine*, 2022, 10(4): 181-212.

[18] YIN Q, HU Q Y, LIU H, et al. Detecting and Tracking Small and Dense Moving Objects in Satellite Videos: A Benchmark [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2022, 60: 1-18.

[19] XIONG F C, ZHOU J, QIAN Y T. Material Based Object Tracking in Hyperspectral Videos[J]. *IEEE Transactions on Image Processing*, 2020, 29: 3719-3733.

[20] WANG H Z, LI W, XIA X G, et al. BihoT: A Large-Scale Dataset and Benchmark for Hyperspec-

tral Camouflaged Object Tracking[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2025, 36(9): 16392-16406.

[21] FAN H, LIN L T, YANG F, et al. LaSOT: A High-Quality Benchmark for Large-Scale Single Object Tracking[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, USA, 2019.

[22] HUANG L H, ZHAO X, HUANG K Q. GOT-10k: A Large High-Diversity Benchmark for Generic Object Tracking in the Wild[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021, 43(5): 1562-1577.

[23] MÜLLER M, BIBI A, GIANCOLA S, et al. TrackingNet: A Large-Scale Dataset and Benchmark for Object Tracking in the Wild[C]//The 15th European Conference on Computer Vision (ECCV), Munich, Germany, 2018.

[24] LIN T Y, MAIRE M, BELONGIE S, et al. Microsoft COCO: Common Objects in Context[C]//The 13th European Conference on Computer Vision (ECCV), Zurich, Switzerland, 2014.

[25] 孙子文, 钱立志, 杨传栋, 等. 基于Transformer的视觉目标跟踪方法综述[J]. 计算机应用, 2024, 44(5):1644-1654.
SUN Ziwen, QIAN Lizhi, YANG Chuandong, et al. Survey of Visual Object Tracking Methods Based on Transformer[J]. *Journal of Computer Applications*, 2024, 44(5): 1644-1654.

[26] 陈琳, 刘允刚. 面向无人机的视觉目标跟踪算法:综述与展望[J]. 信息与控制, 2022, 51(1):23-40.
CHEN Lin, LIU Yungang. UAV Visual Target Tracking Algorithms: Review and Future Prospect [J]. *Information and Control*, 2022, 51(1): 23-40.

[27] 李洋帆, 李伟, 田静, 等. 遥感卫星视频目标跟踪方法综述[J]. 中国空间科学技术(中英文), 2025, 45(5):60-74.
LI Yangfan, LI Wei, TIAN Jing, et al. Object Tracking in Satellite Videos: A Survey[J]. *Chinese Space Science and Technology*, 2025, 45(5): 60-74.

[28] 许廷发, 李天昊, 王颖, 等. 光谱信息扩展的单目标跟踪技术研究进展[J]. 信号处理, 2024, 40(11):1925-1950.
XU Tingfa, LI Tianhao, WANG Ying, et al. Recent Advances in Single Object Tracking Methods with Extended Spectral Information[J]. *Journal of Signal Processing*, 2024, 40(11): 1925-1950.

[29] HENRIQUES J F, CASEIRO R, MARTINS P, et al. High-Speed Tracking with Kernelized Correlation Filters[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, 37(3): 583-596.

[30] DANELLJAN M, HAGER G, KHAN F S, et al. Discriminative Scale Space Tracking[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(8): 1561-1575.

[31] DANELLJAN M, BHAT G, KHAN F S, et al. ECO: Efficient Convolution Operators for Tracking[C]//2017 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, USA, 2017.

[32] HUANG Z Y, FU C H, LI Y M, et al. Learning Aberrance Repressed Correlation Filters for Real-Time UAV Tracking[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, South Korea, 2019.

[33] LI Y M, FU C H, DING F Q, et al. AutoTrack: Towards High-Performance Visual Tracking for UAV With Automatic Spatio-Temporal Regularization[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, USA, 2020.

[34] 邵延华, 陈慧玲, 付贵, 等. 融合图像增强的正则化相关滤波无人机目标跟踪[J]. 中国图象图形学报, 2025, 30(10):3302-3318.
SHAO Yanhua, CHEN Huiling, FU Gui, et al. Fuse Image Enhancement with a Regularized Correlation Filter for Target Tracking of UAVs[J]. *Journal of Image and Graphics*, 2025, 30(10): 3302-3318.

[35] BERTINETTO L, VALMADRE J, HENRIQUES J F, et al. Fully-Convolutional Siamese Networks for Object Tracking[C]//14th European Conference on Computer Vision (ECCV), Amsterdam, The Netherlands, 2016.

[36] LI B, YAN J J, WU W, et al. High Performance Visual Tracking with Siamese Region Proposal Network[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, USA, 2018.

[37] FU C H, CAO Z A, LI Y M, et al. Onboard Real-Time Aerial Tracking With Efficient Siamese Anchor Proposal Network[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2022, 60: 1-13.

[38] XING D T, EVANGELIOU N, TSOUKALAS A, et al. Siamese Transformer Pyramid Networks for Real-Time UAV Tracking[C]//2022 IEEE/CVF Winter Conference on Applications of Com-

puter Vision (WACV), Waikoloa, USA, 2022.

[39] CAO Z A, FU C H, YE J J, et al. HiFT: Hierarchical Feature Transformer for Aerial Tracking [C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, Canada, 2021.

[40] CAO Z A, HUANG Z Y, PAN L, et al. TCTrack: Temporal Contexts for Aerial Tracking [C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, USA, 2022.

[41] WU Y, LI Y X, LIU M Y, et al. Learning an Adaptive and View-Invariant Vision Transformer for Real-Time UAV Tracking[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2026, 36(2): 2403-2418.

[42] WU Y, WANG X C, YANG X Y, et al. Learning Occlusion-Robust Vision Transformers for Real-Time UAV Tracking[C]//2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, USA, 2025.

[43] LI S W, YANG X Y, WANG X C, et al. Learning Target-Aware Vision Transformers for Real-Time UAV Tracking[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2024, 62: 1-18.

[44] XUAN S Y, LI S Y, HAN M F, et al. Object Tracking in Satellite Videos by Improved Correlation Filters With Motion Estimations[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2020, 58(2): 1074-1086.

[45] XUAN S Y, LI S Y, ZHAO Z F, et al. Rotation Adaptive Correlation Filter for Moving Object Tracking in Satellite Videos[J]. *Neurocomputing*, 2021, 438: 94-106.

[46] CHEN S L, WANG T Y, WANG H S, et al. Vehicle Tracking on Satellite Video Based on Historical Model[J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2022, 15: 7784-7796.

[47] 刘耀胜, 廖育荣, 林存宝, 等. 基于特征融合与抗遮挡的卫星视频目标跟踪方法[J]. 北京航空航天大学学报, 2022, 48(12):2537-2547.
LIU Yaosheng, LIAO Yurong, LIN Cunbao, et al. Feature-Fusion and Anti-Occlusion Based Target Tracking Method for Satellite Videos[J]. *Journal of Beijing University of Aeronautics and Astronautics*, 2022, 48(12): 2537-2547.

[48] RUAN L, GUO Y J, YANG D Q, et al. Deep Siamese Network with Motion Fitting for Object Tracking in Satellite Videos[J]. *IEEE Geoscience and Remote Sensing Letters*, 2022, 19: 1-5.

[49] NIE Y D, BIAN C J, LI L G. Object Tracking in Satellite Videos Based on Siamese Network with Multidimensional Information-Aware and Temporal Motion Compensation[J]. *IEEE Geoscience and Remote Sensing Letters*, 2022, 19: 1-5.

[50] YANG J W, PAN Z X, LIU Y H, et al. Single Object Tracking in Satellite Videos Based on Feature Enhancement and Multi-level Matching Strategy [J]. *Remote Sensing*, 2023, 15(17): 4351.

[51] YANG J W, PAN Z X, WANG Z M, et al. SiamMDM: An Adaptive Fusion Network with Dynamic Template for Real-Time Satellite Video Single Object Tracking[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2023, 61: 1-19.

[52] 吕京国, 白颖奇, 王琛. 一种抗遮掩的卫星视频目标跟踪算法[J]. 武汉大学学报(信息科学版), 2024, 49(9):1586-1596.
LÜ Jingguo, BAI Yingqi, WANG Chen. Anti-occlusion Target Tracking Algorithm Based on Satellite Video[J]. *Geomatics and Information Science of Wuhan University*, 2024, 49(9): 1586-1596.

[53] ZHOU J W, DONG Y N, DU B. SiamTITP: Incorporating Temporal Information and Trajectory Prediction Siamese Network for Satellite Video Object Tracking[J]. *IEEE Transactions on Image Processing*, 2025, 34: 4120-4133.

[54] SHAO J, DU B, WU C, et al. HRSiam: High-Resolution Siamese Network, Towards Space-Borne Satellite Video Tracking[J]. *IEEE Transactions on Image Processing*, 2021, 30: 3056-3068.

[55] LAI P J, ZHANG M L, CHENG G, et al. Target-Aware Transformer for Satellite Video Object Tracking[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2024, 62: 1-10.

[56] ZHU Q Q, HUANG X, GUAN Q F. TabCtNet: Target-Aware Bilateral CNN-Transformer Network for Single Object Tracking in Satellite Videos [J]. *International Journal of Applied Earth Observation and Geoinformation*, 2024, 128: 103723.

[57] CHEN Y Z, YUAN Q Q, XIAO Y, et al. STAR: A Unified Spatiotemporal Fusion Framework for Satellite Video Object Tracking[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2025, 63: 1-22.

[58] ZHOU J W, DONG Y N, ZHANG Y X, et al. Incorporating Prior Knowledge and Temporal Memory Transformer Network for Satellite Video Object Tracking [J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2025, 228: 630-647.

[59] WANG J H, LIU F, JIAO L C, et al. Satellite Video Object Tracking Based on Location Prompts [J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2024, 34(7): 6253-6264.

[60] CHEN Y Z, TANG Y Q, YUAN Q Q, et al. REPS: Rotation Equivariant Siamese Network Enhanced by Probability Segmentation for Satellite Video Tracking [J]. *International Journal of Applied Earth Observation and Geoinformation*, 2024, 128: 103741.

[61] CHEN L L, ZHAO Y Q, YAO J X, et al. Object Tracking in Hyperspectral-Oriented Video with Fast Spatial-Spectral Features [J]. *Remote Sensing*, 2021, 13(10): 1922.

[62] XIONG F C, SUN Y L, ZHOU J, et al. Spatial-Spectral-Temporal Correlation Filter for Hyperspectral Object Tracking[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2025, 63: 1-13.

[63] LI W, HOU Z F, ZHOU J, et al. SiamBAG: Band Attention Grouping-Based Siamese Object Tracking Network for Hyperspectral Videos [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2023, 61: 1-12.

[64] LI Z F, XIONG F C, ZHOU J, et al. Learning a Deep Ensemble Network with Band Importance for Hyperspectral Object Tracking [J]. *IEEE Transactions on Image Processing*, 2023, 32: 2901-2914.

[65] LIU Z Q, WANG X Y, ZHONG Y F, et al. SiamHYPER: Learning a Hyperspectral Object Tracker from an RGB-Based Tracker [J]. *IEEE Transactions on Image Processing*, 2022, 31: 7116-7129.

[66] 王永鑫, 夏慧琳, 姜贺, 等. 基于光谱分割的孪生融合高光谱目标跟踪[J]. 红外与激光工程, 2025, 54(2):20240525.
WANG Yongxin, XIA Huilin, JIANG He, et al. Hyperspectral Target Tracking Based on Siamese Fusion with Spectral Segmentation[J]. *Infrared and Laser Engineering*, 2025, 54(2): 20240525.

[67] CHEN Y Z, YUAN Q Q, TANG Y Q, et al. SENSE: Hyperspectral Video Object Tracker *via* Fusing Material and Motion Cues [J]. *Information Fusion*, 2024, 109: 102395.

[68] CHEN Y Z, YUAN Q Q, TANG Y Q, et al. SPIRIT: Spectral Awareness Interaction Network with Dynamic Template for Hyperspectral Object Tracking[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2024, 62: 1-16.

[69] WANG H Z, LI W, XIA X G, et al. Transformer-Based Band Regrouping with Feature Refinement for Hyperspectral Object Tracking[J]. *IEEE Tran-sactions on Geoscience and Remote Sensing*, 2024, 62: 1-14.

[70] GAO L, CHEN L K, LIU P, et al. A Transformer-Based Network for Hyperspectral Object Tracking [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2023, 61: 1-11.

[71] CHEN Y Z, TANG Y Q, SU X, et al. PHTrack: Prompting for Hyperspectral Video Tracking [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2024, 62: 1-18.

[72] ZHOU J W, ZHANG J Q, DONG Y N. Incorporating Prompt Learning and Adaptive Dropping Hyperspectral Information Tracker for Hyperspectral Object Tracking [C]//The 14th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS), Helsinki, Finland, 2024.

[73] JIANG W H, ZHAO D, WANG C, et al. Hyperspectral Video Object Tracking with Cross-Modal Spectral Complementary and Memory Prompt Network [J]. *Knowledge-Based Systems*, 2025, 330: 114595.

[74] CHEN Y Z, YUAN Q Q, TANG Y Q, et al. SSTtrack: A Unified Hyperspectral Video Tracking Framework *via* Modeling Spectral-Spatial-Temporal Conditions[J]. *Information Fusion*, 2025, 114: 102658.

[75] XUE Y L, JIN G D, SHEN T, et al. SmallTrack: Wavelet Pooling and Graph Enhanced Classification for UAV Small Object Tracking[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2023, 61: 1-15.

[76] 孙波, 杨春成, 徐立, 等. 基于多阶段特征和非对称-膨胀卷积的目标跟踪[J]. 武汉大学学报(信息科学版), 2024, 49(9):1712-1722.
SUN Bo, YANG Chuncheng, XU Li, et al. An Object Tracking Method Based on Multi-stage Features and Asymmetric-Dilated Convolution[J]. *Geomatics and Information Science of Wuhan University*, 2024, 49(9): 1712-1722.

[77] MAYER C, DANELLJAN M, BHAT G, et al. Transforming Model Prediction for Tracking [C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, USA, 2022.

[78] CUI Y T, JIANG C, WANG L M, et al. MixFormer: End-to-End Tracking with Iterative Mixed Attention [C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, USA, 2022.

[79] YE B T, CHANG H, MA B P, et al. Joint Feature Learning and Relation Modeling for Tracking: A One-Stream Framework[C]//The 17th European Conference on Computer Vision (ECCV), Tel Aviv, Israel, 2022.

[80] LI S W, YANG Y X, ZENG D, et al. Adaptive and Background-Aware Vision Transformer for Real-Time UAV Tracking[C]//2023 IEEE/CVF International Conference on Computer Vision (ICCV), Paris, France, 2023.

[81] MAYER C, DANELLJAN M, PANI PAUDEL D, et al. Learning Target Candidate Association to Keep Track of What Not to Track[C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, Canada, 2021.

[82] ZHANG W, JIAO L, LIU F, et al. MBLT: Learning Motion and Background for Vehicle Tracking in Satellite Videos[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2022, 60: 1-15.

[83] CAI W R, LIU Q J, WANG Y H. HIPTrack: Visual Tracking with Historical Prompts[C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, USA, 2024.

[84] ZHENG Y, ZHONG B, LIANG Q, et al. ODTrack: Online Dense Temporal Token Learning for Visual Tracking[C]// The 38th Annual AAAI Conference on Artificial Intelligence, Vancouver, Canada, 2024.

[85] XIE J X, ZHONG B N, MO Z Y, et al. Autoregressive Queries for Adaptive Tracking with Spatio-Temporal Transformers [C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, USA, 2024.

[86] BAI Y F, ZHAO Z Y, GONG Y H, et al. ARTrackV2: Prompting Autoregressive Tracker Where to Look and How to Describe[C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, USA, 2024.

[87] HUANG Y, LIN L, ZHUANG W, et al. Drift-Resilient Temporal Priors for Visual Tracking[C]// 2026 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Denver, USA, 2026.

[88] HUANG Y Q, LI X, ZHOU Z K, et al. RTracker: Recoverable Tracking via PN Tree Structured Memory [C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, USA, 2024.

[89] GUO M Z, TAN W P, RAN W Y, et al. DreamTrack: Dreaming the Future for Multimodal Visual Object Tracking[C]//2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, USA, 2025.

[90] WANG H Z, LI W, XIA X G, et al. SSF-Net: Spatial-Spectral Fusion Network With Spectral Angle Awareness for Hyperspectral Object Tracking [J]. *IEEE Transactions on Image Processing*, 2025, 34: 3518-3532.

[91] QIN H L, XU T F, LI T H, et al. MUST: The First Dataset and Unified Framework for Multispectral UAV Single Object Tracking[C]//2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, USA, 2025.

[92] WU Y N, JIAO L C, LIU X, et al. Domain Adaptation-Aware Transformer for Hyperspectral Object Tracking [J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2024, 34(9): 8041-8052.

[93] ZHU J W, LAI S M, CHEN X, et al. Visual Prompt Multi-modal Tracking [C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, Canada, 2023.

[94] HOU X J, XING J Z, QIAN Y J, et al. SDSTrack: Self-Distillation Symmetric Adapter Learning for Multi-modal Visual Object Tracking [C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, USA, 2024.

[95] TAN Y D, SHAO J W, ZAMFIR E, et al. What You Have Is What You Track: Adaptive and Robust Multimodal Tracking [C]//2025 IEEE/CVF International Conference on Computer Vision (ICCV), Honolulu, USA, 2025.

[96] LI H, WANG Y, HAO W, et al. RAGTrack: Language-Aware RGBT Tracking with Retrieval-Augmented Generation[C]//2026 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Denver, USA, 2026.

[97] LIU C, MA X, YANG X C, et al. COMO: Cross-Mamba Interaction and Offset-Guided Fusion for Multimodal Object Detection[J]. *Information Fusion*, 2026, 125: 103414.

[98] YE J J, FU C H, ZHENG G Z, et al. Unsupervised Domain Adaptation for Nighttime Aerial Tracking [C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, USA, 2022.

[99] FU C H, LI T, YE J J, et al. Scale-Aware Do-

main Adaptation for Robust UAV Tracking [J]. *IEEE Robotics and Automation Letters*, 2023, 8(6): 3764-3771.

[100]FU C H, YAO L L, ZUO H B, et al. SAM-DA: UAV Tracks Anything at Night with SAM-Powered Domain Adaptation [C]//2024 International Conference on Advanced Robotics and Mechatronics (ICARM), Tokyo, Japan, 2024.

[101]YAO S Y, ZHU R, WANG Z Q, et al. UMDATrack: Unified Multi-domain Adaptive Tracking Under Adverse Weather Conditions [C]//2025 IEEE/CVF International Conference on Computer Vision (ICCV), Honolulu, USA, 2025.

[102]董燕妮, 秦学翔, 张玉香, 等. 多级空间特征的平衡测度学习遥感图像域适应分类[J]. 武汉大学学报(信息科学版), 2026, 51(6):1182-1193.
DONG Yanni, QIN Xuexiang, ZHANG Yuxiang, et al. Balanced Metric Learning in Multilevel Spatial Features for Domain Adaptation in Remote Sensing Image Classification[J]. *Geomatics and Information Science of Wuhan University*, 2026, 51(6): 1182-1193.

[103]YAN B, PENG H W, WU K, et al. LightTrack: Finding Lightweight Neural Networks for Object Tracking via One-Shot Architecture Search [C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, USA, 2021.

[104]KANG B, CHEN X, WANG D, et al. Exploring Lightweight Hierarchical Vision Transformers for Efficient Visual Tracking[C]//2023 IEEE/CVF International Conference on Computer Vision (ICCV), Paris, France, 2023.

[105]康奔, 陈鑫, 赵洁, 等. TMamba:面向高效目标跟踪的视觉状态空间模型[J]. 中国图象图形学报, 2025, 30(10):3199-3214.
KANG Ben, CHEN Xin, ZHAO Jie, et al. TMamba: A Visual State-Space Model for Efficient Object Tracking [J]. *Journal of Image and Graphics*, 2025, 30(10): 3199-3214.

[106]LI J N, YUAN X Y, QIN H L, et al. CVTTrack: Concentrating on Valid Tokens for One-Stream Tracking[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2025, 35(1): 33-44.

[107]WU H, WANG X, ZHANG J, et al. UTPTrack: Towards Simple and Unified Token Pruning for Visual Tracking[C]//2026 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Denver, USA, 2026.

[108]XUE C C, ZHONG B N, LIANG Q H, et al. Similarity-Guided Layer-Adaptive Vision Transformer for UAV Tracking [C]//2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, USA, 2025.

[109]LIN L T, FAN H, ZHANG Z P, et al. Tracking Meets LoRA: Faster Training, Larger Model, Stronger Performance [C]//The 18th European Con-ference on Computer Vision (ECCV), Milano, Italy, 2024.

[110]CAI W R, LIU Q J, WANG Y H. SPMTrack: Spatio-Temporal Parameter-Efficient Fine-Tuning with Mixture of Experts for Scalable Visual Tracking [C]//2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, USA, 2025.