A Survey of Object Tracking in UAV Videos*

LIU Liyuan, ZENG Qian*, ZHOU Shengbang, WEI Geng, JIA Yuheng, CHENG Ziyao, WANG Zhuoyue
(Key Laboratory of Intelligent Information Processing and Communication,
Nanning Normal University, Nanning Guangxi 530100, China)

Abstract: The research on object tracking algorithms for UAV aerial videos has gained significant attention due to the rapid development of UAV and related technologies. However, the increasing volume of video data presents numerous challenges that must be addressed for effective target tracking. A comprehensive review of the algorithms used for object tracking based on UAV is provided. Specifically, the three main target tracking algorithms based on correlation filtering, siamese network and tranformer are analyzed and discussed in depth. Furthermore, the development and application requirements of UAV visual target tracking are summarized, and the corresponding prospects are depicted.

Key words: UAV; object tracking; correlation filtering tracking algorithm; siamese network tracking algorithm; transformer tracking algorithm
EEACC: 6135 **doi**: 10.3969/j.issn.1004-1699.2025.01.002

无人机视频中的目标跟踪研究综述*

刘丽媛, 曾 倩*, 周省邦, 韦 耿, 贾宇恒, 陈紫瑶, 王卓越
(南宁师范大学 智能信息处理与通信重点实验室, 广西 南宁 530100)

摘 要: 随着无人机等技术的迅速发展, 针对无人机航拍视频的目标跟踪算法研究逐渐成为热点, 然而随着海量视频的出现, 视频目标跟踪还面临诸多难题亟需解决。对基于无人机的视频目标跟踪的相关算法进行综述, 围绕基于相关滤波、孪生网络以及 Tranformer 三种主流目标跟踪算法进行深入分析与讨论, 并从无人机视觉目标跟踪的发展与应用需求角度出发, 进行总结与展望。

关键词: 无人机; 目标跟踪; 相关滤波跟踪算法; 孪生网络跟踪算法; Transformer 跟踪算法
中图分类号: TP391 **文献标识码**: A **文章编号**: 1004-1699(2025)01-0021-10

视觉目标跟踪是计算机视觉领域中的一个重要研究方向, 自 20 世纪 80 年代起, 目标跟踪技术开始应用于各个领域, 包括监控领域[1]、巡检领域[2]、精准农业[3]、人机交互[4]和智慧工业[5]等。该技术被用来在一段连续的视频序列中, 获取指定目标的运动轨迹并进行持续跟踪。但是, 传统视觉目标跟踪方法的移动性有限, 且物体捕获精度低、设备静态、摄像范围有限、易丢失信息等, 难以进行长时跟踪。

相比传统的跟踪平台, 无人机体积小, 可以在空中飞行, 受地理条件约束少, 可自由垂直起降, 还能在空中悬停, 能够在特殊环境中灵活采集视频数据, 已经在灾害救援、环保监测、交通监视、农业植保和电力巡检等领域发挥了重要作用。基于无人机的视觉目标跟踪在视觉定位[6]、灾害监测系统[7]和交通监控[8]等领域都有成功应用。但是, 目前很多性能优越的目标跟踪算法只能解决单一或少数的问题, 无法解决复杂场景下的问题, 特别是无人机计算资源有限, 需要在保证精度的同时尽量减少复杂度。因此, 进一步研究无人机视觉目标跟踪算法需求迫切且充满挑战。

本文总结了近几年视觉目标跟踪在无人机平台上的应用发展。对国内外无人机视觉目标跟踪算法、基准数据集和性能参数等的发展进行了综合分析, 并进行了展望, 为目标跟踪技术应用于无人机平台提供研究和发展思路。

1 国内外研究现状

1.1 无人机应用发展

随着传感技术、通信和导航等技术的不断进步, 无人机的自动化和智能化技术得到了极大提高, 该

项目来源: 国家自然科学基金, 地区科学基金项目(62161031); 广西自然科学基金委员会, 青年基金项目(2020GXNSFBA297097) 桂林电子科技大学, 广西无线宽带通信与信号处理, 重点实验室基金项目(GXKL06230207); 广西高校中青年教师科研基础能力提升项目(2022KY0381); 广西自然科学基金, 面上项目(2020GXNSFAA297184)
收稿日期: 2023-02-27 **修改日期**: 2023-07-05

技术在城市、工业和农业等领域备受青睐,并具有广泛的应用范围。例如,在 2018 年,Kim 等[1]利用多个异构智能无人机设计了一个差分框架,用于对智慧城市和广阔海洋进行监控;在 2019 年,Fernández 等[9]提出了一种基于无人机的设计和评估系统,旨在自动化库存任务和保持无线射频识别标签上的工业物品的可追溯性;在 2020 年,Popescu 等[3]提出了一种基于无人机和联合无线传感器网络协作的分级结构,用于精准农业作物监测;在 2021 年,Xie 等[2]提出了一种基于检测的跟踪系统,用于自动处理海上船舶检查视频并预测可能存在裂缝的可疑区域;在 2022 年,Li 等[10]提出一种具有光照自适应和抗黑暗能力的新型跟踪器,显著拓宽无人机智能部署的范围。显而易见,在未来,无人机也将在各个领域大展身手。

1.2 国外研究进展

目前,国外对视觉目标跟踪的研究单位主要有卡内基梅隆大学、佛罗里达大学、南加州大学机器人与智能系统研究所、雷丁大学联合法国信息与自动化研究所等,表 1 整理了国外相关研究进展[11-15]。

表 1 无人机视觉目标跟踪国外相关研究进展

研究机构	主要研究
卡内基梅隆大学[11]	VSAM(Video Surveillance and Monitoring)项目,高密度环境下的行人、车辆等目标的实时监控与跟踪
佛罗里达大学[12]	基于无人机平台的 COCOA 监控系统,针对车辆、人等多种类型运动目标的自动检测、轨迹跟踪、视频检索等功能
雷丁大学联合法国信息与自动化研究所[13]	ADVISOR 系统,针对停机场内的行人、飞机和行驶车辆,实现 24 h 全方位监控,对可疑目标进行精准的检测和跟踪
南加州大学机器人与智能系统研究所[14]	广域空中监视系统,建立一个高效且可扩展的框架,包括图像拼接、目标检测与跟踪、广域航空视频行为分析等功能
马里兰大学计算机视觉实验[15]	W4 实时监控系统,可以基于人体不同部位来实现多人跟踪和物品检测

卡内基梅隆大学在 VSAM 项目[11]之后,又开展了 VIVID 项目的研究;COCOA 系统[12]是早期的典型无人机目标跟踪系统,但在视频图像处理阶段使用帧差法,只能处理背景单调的情况;之后,ADVISOR 系统[13]、广域空中监视系统[14]和 W4 实时监控系统[15]等都提高了检测和跟踪的精度,可应用于复杂背景的情况。2021 年,Balamuralidhar 等[8]提出了一种交通监控系统 MultEYE(Multi-task Entwined YOLO and ENet),它可以检测、跟踪和估计一系列航拍图像中的车辆速度,并在无人机上的嵌入式计算机上实时执行这些任务。

1.3 国内研究进展

近年来,国内众多研究所和高校陆续成立了相关课题组、实验室等对无人机视觉跟踪展开研究。如大连理工大学通信与信号处理研究所、同济大学等。国内相关研究进展归纳如表 2 所示[16-20]。

表 2 无人机视觉目标跟踪国内相关研究进展

研究机构	主要研究
中国科学院自动化研究所[16]	智能视频监控系统科奥森,已经成功应用于上海世博会、北京奥运会等多个国家重点项目
阿里巴巴集团[17]	“城市大脑”项目,使用人工智能等技术来处理城市中的海量实时数据,继而实现目标的检测跟踪和红绿灯优化等功能
同济大学[18]	自动时空正则化的无人机高性能视觉跟踪系统,引入空间局部响应图变化作为空间正则化
大连理工大学通信与信号处理研究所[19]	新的视觉跟踪方法:Alpha-Refine,以即插即用的方式有效提高不同类型跟踪器的跟踪性能
香港中文大学(深圳)[20]	鼓励探索自然语言特征和音频线索来进行无人机跟踪,提供自然语言规范和音频描述来丰富提出的无人机目标跟踪新基准 WebUAV-3M

中国科学院自动化研究所研发的中科奥森智能视频监控系统[16],目前已经有了很大的应用市场;之后,阿里巴巴集团开展了“城市大脑”项目[17],实时处理大量交通数据,为城市交通提供便利。近年来,为进一步提高跟踪性能,同济大学、大连理工大学和香港中文大学相继提出了 AutoTrack[18]、Alpha-Refine[19]和 WebUAV-3M[20]无人机跟踪系统和基准,用于深度无人机目标跟踪。

2　无人机视觉目标跟踪算法

2.1　跟踪框架

视觉目标跟踪系统的框架主要由特征表示、搜索策略、观测模型和模型更新等模块组成,其中观测模型对跟踪结果至关重要,大多数跟踪方法集中于对观测模型的设计。目标跟踪方法可分为生成式方法和判别式方法[21]:

①生成式方法首先对当前帧中目标进行特征提取,然后构建一个模型对目标进行表示,最后在后续帧中进行图像匹配搜索,找到与模型最为相似的结果,框架如图 1 所示。与判别式方法相比,该方法的缺点是忽略了跟踪过程中的背景信息。

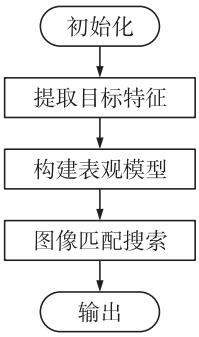

图 1　生成式算法框架

②判别式方法将背景信息考虑在内,利用分类器来区分前景,将跟踪问题一分为二,化为跟踪目标和背景决策边界两个问题,其框架如图 2 所示。目前,在空中监视跟踪中,判别法由于具有较强的鲁棒性而受到广泛的欢迎。

2.2　目标跟踪算法

近年来,大量的高精度的跟踪方法已经得到发展,由于使用快速傅里叶变换的相关滤波器具有很高的计算效率和良好的定位性能,基于相关滤波的跟踪算法在无人机跟踪领域取得了突出的性能。与此同时,随着深度学习技术的发展,近几年来基于孪生网络、Transformer 的算法在跟踪领域也有着突出表现,本节根据模型框架的分类,对基于相关滤波、孪生网络和 Transformer 的三类跟踪算法的发展进行了归纳。

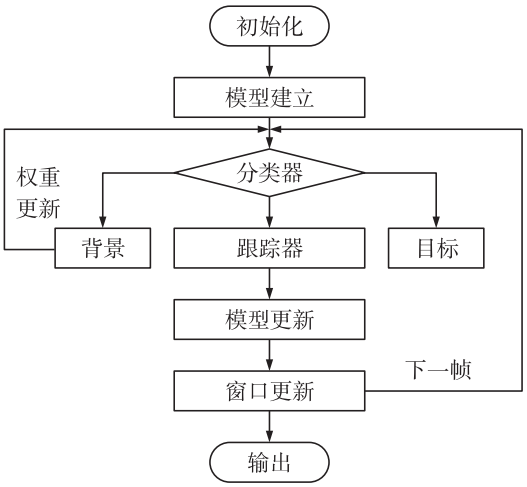

图 2　判别式算法框架

2.2.1　基于相关滤波的跟踪算法

相关滤波跟踪的原理是:使用相关滤波器对目标图像进行训练,得到一个滤波模板,然后利用该滤波器对模板与候选样本做互相关计算,得到跟踪结果的置信图,最大置信度的区域即为当前帧的目标位置。基于相关滤波的跟踪算法的执行过程如图 3 所示,其中⊙表示乘法运算,FFT(Fast Fourier Transformation)表示快速傅里叶变换,IFFT(Inverse Fast Fourier Transform)表示快速傅里叶逆变换。

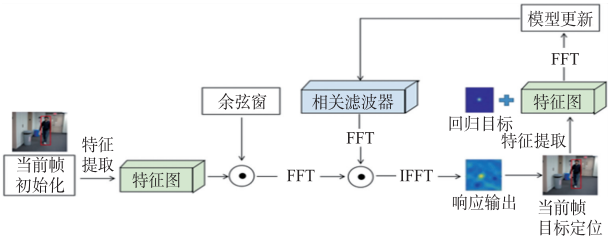

图 3　基于相关滤波的跟踪算法流程

相关滤波在目标跟踪领域的应用是由 Bolme[22]提出的相关滤波器 MOSSE(Minimum Output Sum of Squared Error Filter)开启,该跟踪器在保持以往跟踪效果的同时,速度提升了许多,且对光照、姿态等的变化具有较好的效果。此后,相关滤波被越来越多的学者用于跟踪,相继提出了许多改进的算法,跟踪的效果也越来越好。Danelljan 等[23]提出了一种基于检测跟踪框架的多尺度跟踪相关滤波器 DSST(Discriminative Scale Space Tracker),该算法简单,性能优良,可移植性高。对于振荡较强的目标,DSST 仍能实现较好的跟踪,因此 DSST 对无人机跟踪时背景和目标都在运动这种特殊场景具有较强的鲁棒性。

目前,基于相关滤波的跟踪算法在无人机应用中的优化,大多在 MOSSE 的基础上,结合特征的选择和尺度自适应进行改进。表 3 所示为基于相关滤波的

经典跟踪算法的总结对比[22-28]。

基于相关滤波的经典跟踪算法已经取得了优异的性能,但它们没有充分考虑到无人机场景的独特性——高速飞行或摄像机旋转会导致捕获对象的密度不均匀、目标形变更大,而基于相关滤波的经典跟踪算法对形变较大的目标不敏感,无人机场景的跟踪性能较差,尤其是外部环境的快速变化,导致航拍图像模糊,这将不可避免地导致图像质量下降。

为了更好地应对无人机场景下的目标跟踪,改进基于相关滤波的跟踪方法,Zheng 等[29]提出了一种包含多任务卷积操作和目标检测的基于相关滤波器的无人机视觉跟踪方法。与以往的方法相比,该算法不仅探讨了深度特征之间的相互依赖关系,而且还考虑了卷积滤波器内的相关性。此外,该算法还引入了目标检测来处理目标损失,以提高跟踪性能。另外,考虑到无人机计算资源、电池容量和功耗的限制,Li 等[30]针对基于判别跟踪器的判别尺度估计问题,提出了一种尺度细化策略。为了避免复杂无人机场景中过拟合和滤波器退化的问题,Lin 等[31]提出了一种新的自适应响应推理方法,这个计算方法利用时间信息自适应地学习以前滤波器与当前滤波器之间的相关性,显著提高了跟踪器的鲁棒性。但忽略了不同的判别特征,会影响跟踪性能。为了解决这个问题,Chen 等[32]提出了比例正则化背景感知相关滤波器 TRBCF(Temporally Regularized Background-aware Correlation Filter),通过引入一个时间正则化项提高了空间域目标和背景的可识别性,并使用真实的背景块作为负样本来实现连续无人机视频中的目标跟踪。

表 3 基于相关滤波的经典跟踪算法对比

跟踪算法	优点	缺点	参考文献
MOSSE	在保持较好的跟踪效果的同时,其速度远超同时期的其它跟踪算法	采用最原始的灰度特征,特征表示能力较弱,使用线性滤波器,分类性能一般	[22]
DSST	将跟踪分为平移跟踪和尺度跟踪两个问题,可有效处理目标尺度变化问题	只适用于目标尺度变化小的跟踪,目标尺度变化大时则失效	[23]
CSK(Circulant Structure of Tracking-by-Detection with Kernels)	采用循环结构进行视频中相邻帧的相关性检测,大幅度减少运算量	对尺度变化适应性弱,扛遮挡能力差,导致跟踪效果不佳	[24]
KCF(Kernel Correlation Filter)	引入核技巧和多通道特征,且采用循环矩阵扩充负样本,减少复杂度	目标跟踪框固定,当目标大小发生变化时,易导致目标框漂移	[25]
SAMF(Scale Adaptive with Multiple Features Tracker)	将单一的特征扩展为多特征融合,采用尺度池的方法实现目标的自适应跟踪	当出现目标遮挡和背景相似等情况时,跟踪精度下降	[26]
BACF(Background Aware Correlation Filter)	循环结构中加入补零操作,克服了循环样本带来的边界效应,扩大了搜索域	当目标被遮挡时,或出现重叠情况时,无法准确识别目标位置	[27]
LMCF (Large Margin Object Tracking with Circulant Feature Maps)	使用相关滤波来优化结构化 SVM (Support Vector Machine),提高了计算速度	求解结构化 SVM 时,计算量大,因此算法的实时性效果较差	[28]

相关滤波在目标跟踪领域中具有以下优势:①快速计算:相关滤波方法通过在频域进行计算,能够快速实现目标模板的更新和目标搜索,适用于对实时性要求较高的应用场景;②简单且易于实现:相关滤波方法基于线性滤波的原理,算法简单易懂,实现和调优相对容易。综上所述,相关滤波在目标跟踪领域中具有快速计算、简单易实现以及适应尺度变化等优势。

2.2.2 基于孪生网络的跟踪算法

卷积神经网络由于其良好的特征提取能力和泛化能力,在跟踪领域也取得了巨大的成功。其中,孪生神经网络作为一种特殊的神经网络,其模型结构简单,由两个分支组成,包含模板分支和搜索分支,如图 4 所示。孪生网络通过这两个分支分别训练模板图像和搜索图像,基于孪生神经网络的跟踪算法通过端到端的学习,计算两条分支的相似程度。从原理上来看,这种方法与相关滤波方法非常相似。

Bertinetto 等[33]提出了全卷积孪生跟踪网络 SiamFC(Fully-Convolutional Siamese Network),该跟踪算法极其简单,基于模板匹配的思想,通过计算相

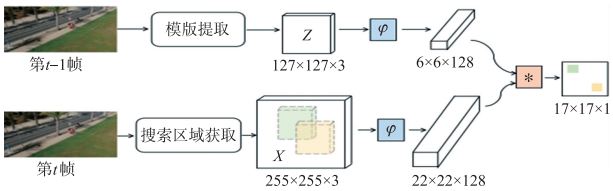

图 4　孪生网络目标跟踪框架[33]

似度来确定目标的位置,进而实现目标跟踪。在跟踪鲁棒性和实时性上比较均衡,实现了较以往算法最先进的性能,达到了实时跟踪的效果。图 5 所示为基于孪生网络的目标跟踪算法流程。

在 SiamFC 的基础上,许多研究人员陆续进行改进,提出了许多性能优越的算法。基于孪生网络的经典跟踪算法的总结对比如表 4 所示[33-38]。

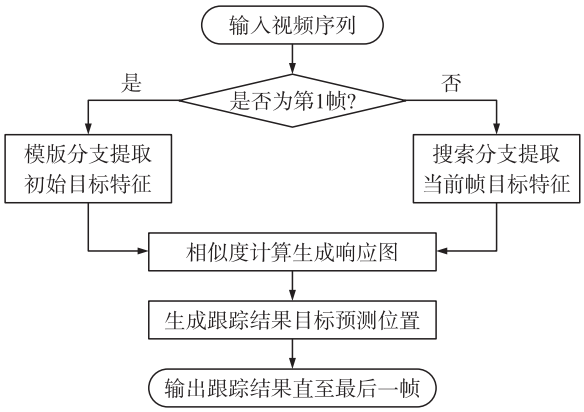

图 5　基于孪生网络的 SiamFC 跟踪算法流程

表 4　基于孪生网络的经典目标跟踪算法对比

跟踪算法	优点	缺点	参考文献
SiamFC	采用 AlexNet 作为主干网络,网络结构简单,速度快	只采用第一帧作为模板,图像的背景信息没有得到全面的利用	[33]
SiamRPN(Siamese Region Proposal Network)	引入 RPN 网络,回归分支直接回归出候选框的位置,解决了尺度问题	网络只学习了区分背景和前景的能力,对于目标相似的物体无法区分	[34]
SiamRPN++	采用深层基准网络 ResNet 等,并加入位置均衡的采样策略,消除了位置偏见	对于尺度变化较大的物体,容易产生跟踪漂移	[35]
SiamMask	使用加深的特征提取网络 ResNet-50,获得更加丰富的信息,可以同时实现视频目标跟踪和目标分割两个任务	为实现目标跟踪和分割两个任务,采用了大量的空间和通道信息,对于跟踪来说存在信息冗余的情况	[36]
SiamFC++	在 SiamFC 的基础上添加了位置回归和质量分数,大大提高了跟踪性能	基于单帧的跟踪,对于快速运动的目标无法准确地跟踪其位置	[37]
SiamDW (Deeper and Wider Siamese Networks)	设计了内部裁剪残差模块单元,消除了 padding 带来的影响,网络判别性增强	网络层次增加,算法的复杂度增大,导致跟踪的实时性降低	[38]

Bo 等[34]在 SiamFC 的基础上,加入了 RPN(Region Proposal Network)网络,大大降低了运行时间,算法性能提高的同时跟踪速度也得到提升。但是,这些网络都使用浅层的类 AlexNet 作为基准特征提取器,这是因为直接使用预训练好的深层网络反而会导致跟踪算法精度的下降。随后,Li 等[35]将深层基准网络 ResNet、Inception 等应用到基于孪生网络的跟踪网络中,很大程度上降低无人机计算代价和内存使用;Wang 等[36]在 Siamese 网络架构中额外增加一个 Mask 分支,提出了对视觉目标跟踪和视频目标分割的统一框架 SiamMask,进一步解决在精度上存在的问题。但是对于较大的尺度变化等追踪问题来说,锚点与框的先验因素限制了追踪器的鲁棒性,容易产生损失。因此,Xu 等[37]提出了一种基于 SiamFC 的新型孪生网络 SiamFC++,该算法引入无先验知识跟踪分支,使得追踪不再依赖于类似于长宽比等先验知识,大大提高了跟踪性能。

为了解决基于模板匹配的跟踪器不能很好地适应无人机视频数据集中频繁的外观变化的问题,Sun 等[39]提出了一种模板驱动的孪生网络 TDSiam(Template-Driven Siamese Network),该网络中加入了一个针对特征提取子网络的模板库分支,以适应目标的可变外观。针对小目标和快速运动目标跟踪性能较差的问题,Zhu 等[40]提出了一个新的多层预测孪生网络 MLPS(Multi-Level Prediction Siamese Network)。其中,多层预测模块可以充分利用每层的特征,提高对不同尺度目标的鲁棒性。

综上所述,孪生网络架构在目标跟踪领域得到了广泛的应用和研究。通过优化主干网络,孪生网络在运行速度和准确率等方面取得了显著的提升。深层次主干网络的应用使得网络能够更好地提取目标的深层次特征信息,这些特征包含了丰富的语义信息和高级表征。在目标跟踪任务中,孪生网络充分利用目标的外观信息和语义信息,从而实现更准确的跟踪结果。孪生网络的优势还体现在结合目标的浅层特征和深层特征方面。浅层特征对目标的细

节和纹理信息更敏感,而深层特征对目标的整体形状和语义信息更敏感。通过综合利用这些特征,孪生网络能够在复杂场景下更好地跟踪目标,并应对遮挡、形变和快速运动等挑战。

2.2.3 基于 Transformer 的跟踪算法

相关性在跟踪领域起着至关重要的作用,尤其是在最近流行的基于孪生网络的跟踪器中。但相关操作本身是一个局部线性匹配过程,易引起语义信息和全局信息丢失,导致陷入局部最优。针对上述问题,Chen 等[41]提出一种新颖的基于注意力的特征融合网络 TransT(Transformer Tracking),该网络首次展示了一种基于查询机制的简单有效的方法,通过建立非线性语义融合和挖掘远距离特征关联,有效聚合目标和搜索区域的全局信息,显著提升了算法的精度。TransT 的网络结构如图 6 所示,从图中可以看出,一个基于注意力机制的特征融合网络取代了 Siamese 框架中的相关操作。

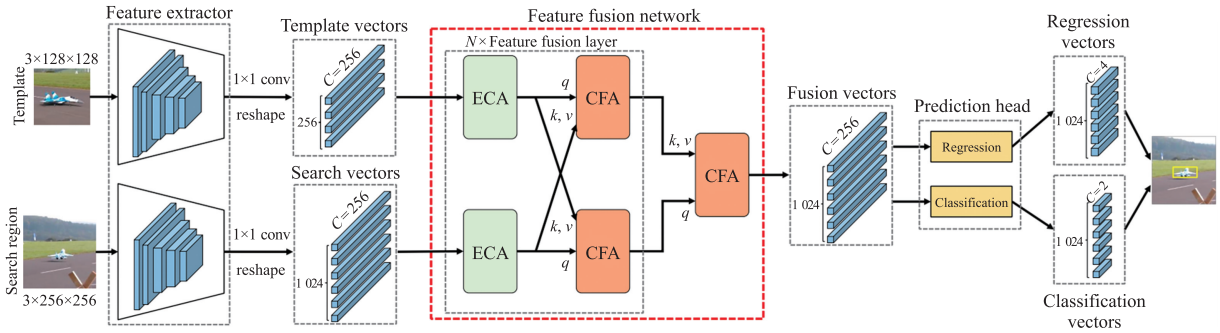

图 6 TransT 的网络结构[41]

在视频对象跟踪中,连续帧之间存在丰富的时序上下文,现有的跟踪器大大忽略了这些上下文。为了解决这个问题,Wang 等[42]提出了一种基于 Transformer 框架的 TMT(Transformer Meets Tracker)跟踪算法,来进行视频中多帧信息的传递。无人机影像中视野面积广阔,导致目标尺寸小,包含的信息有限,而该方法添加 Transformer 结构,加强帧与帧之间的联系,结合上下文信息,可以起到很好的追踪效果,提高跟踪精度。但是 TMT 方法仅使用了空间信息,对时间信息的使用非常少,这导致追踪的效果不佳。所以,Yan 等[43]提出了一种基于 Transformer 的利用时空间信息的追踪架构:STARK 算法(Spatio-Temporal Transformer for Visual Tracking)。STARK 算法通过 Transformer 学习鲁棒的时空联合表示,并将目标跟踪建模成一个直接的边界框预测问题,显著提高跟踪性能。近年的无锚框跟踪器提供了一种高效的回归机制,但是仍然难以精确地估计目标框。因此,Cui 等[44]提出了一个类似于 Transformer 的目标转换回归分支 TREG(Target Transformed Regression),以实现精确的无锚框跟踪。目标图像的上下文表示能够增强目标的相关信息,帮助精确定位框边界,且 Transformer 局部密集的匹配机制,在一定程度上处理了目标的变形,提高了对目标外观变化和几何变形的鲁棒性。

以往考虑的目标跟踪主要是在白天的场景下进行的,而夜间跟踪比较困难,这限制了无人机应用的发展。为了解决这个问题,Ye 等[45]提出了一种称为 SCT(Spatial Channel Transformer)的弱光增强器。SCT 通过跟踪方法前的训练和插入,实现了夜间可靠的无人机跟踪。同时,引入了一种新的空间通道注意模块,对局部环境和全局信息进行建模。并通过鲁棒非线性曲线投影对夜间图像进行去噪和照明。为了在计算资源有限的移动平台(如无人机)上实现实时处理,Xing 等[46]提出了一种新的目标跟踪方法 SiamTPN(Siamese Transformer Pyramid Network),该方法结合了深度神经网络和 Transformer 架构的优点。具体来说,使用 ShuffleNetV2 作为一个轻量级网络,并将其与 Transformer 架构相结合,构建了一个目标特定的外观模型。Lu 等[47]提出了一种即插即用的级联 Transformer 结构 CDT(Cascaded Denoising Transformer),以抑制杂波和复杂的真实噪声,提高无人机夜间跟踪性能。

Transformer 架构在目标跟踪领域中具有多个显著的优势。首先,Transformer 能够有效地建模目标在连续帧之间的时空关系。通过自注意力机制,Transformer 能够对不同帧之间的长期依赖进行建模,捕捉目标的运动和演变。这使得 Transformer 在处理目标跟踪任务时能够更好地理解目标的动态变化,提高跟踪的准确性和鲁棒性。其次,Transformer 具有较强的特征融合能力。Transformer 架构通过自注意力机制,可以灵活地对不同层次、不同位置的特征进行加权融合,实现全局信息的交互和整合。这

种特征融合机制使得 Transformer 能够更好地综合利用目标的多尺度特征,提升目标跟踪的性能。此外,Transformer 还具备较强的泛化能力和可解释性。由于 Transformer 模型在训练过程中没有显式的位置信息,它可以处理任意长度的序列数据,具备较强的泛化能力。最后,Transformer 的自注意力机制使得模型能够自动学习到特征之间的关系,提高了模型的可解释性。

综上所述,Transformer 架构在目标跟踪领域具有建模时空关系、特征融合能力强、泛化能力强和可解释性高等优势,为无人机视频的目标跟踪任务的准确性和鲁棒性提供了有效的解决方案。

3　目标跟踪数据集和评价指标

3.1　数据集

近年来,随着数据驱动深度学习方法的发展,目标跟踪算法在无人机上的应用逐步加深,研究人员也陆续发布了多个无人机场景下的目标跟踪数据集。几种常见的基于视觉的无人机目标跟踪数据集如表 5 所示[48-52]。

表 5　常见的无人机目标跟踪数据集

数据集	属性					
	视频序列	帧数	标记框	分辨率/像素	任务	发表年份
UAV123[48]	123	112 578	112 578	720×720	SOT	2016
Campus[49]	100	929 499	18 500	1 400×1 904	MOT	2016
DTB70[50]	70	15 900	15 900	1 280×720	SOT	2017
UAVDT[51]	100	80 000	841 000	1 080×540	MOT	2018
VisDrone-2021[52]	400	265 228	2 600 000	2 000×1 500	MOT	2021

3.2　评估指标

对于来自无人机的机载视频中的目标跟踪,分为单目标跟踪 SOT(Single Object Tracking)和多目标跟踪 MOT(Multiple Object Tracking)。其中应用较多的为多目标跟踪,常用评估方法包括多目标跟踪准确度 MOTA(Multi-Object Tracking Accuracy)、多目标跟踪精度 MOTP(Multi-Object Tracking Precision)、识别精度 IDP(Identification Precision)、正确识别的检测与地面实况和计算检测的平均数量之比 IDF(Identification F-Score)。MOTA 表示为:

$$\mathrm{MOTA}=\frac{1-\sum(\mathrm{FN}+\mathrm{FP}+\mathrm{IDSW})}{\sum \mathrm{GT}} \tag{1}$$

式中:FN(False Negative)为假阴性的数量,FP(False Positive)为假阳性的数量,IDSW(ID Switch)为所分配的 ID 发生变化的次数,GT(Ground Truth)为正确标注目标的数量。定义 MOTP 为:

$$\mathrm{MOTP}=\frac{\sum_{t,i} d_{t,i}}{\sum_{t} c_t} \tag{2}$$

式中:d 为跟踪目标和给它分配的真值之间在所有帧中的平均度量距离,c 为在当前帧匹配成功的数目。定义 IDP 为:

$$\mathrm{IDP}=\frac{\mathrm{IDTP}}{\mathrm{IDTP}+\mathrm{IDFP}} \tag{3}$$

定义 IDF 得分为:

$$\mathrm{IDF}=\frac{2\mathrm{IDTP}}{2\mathrm{IDTP}+\mathrm{IDFP}+\mathrm{IDFN}} \tag{4}$$

式中:IDTP(Identity True Positives)代表在整个视频中检测目标被正确分配的数量,IDFN(Identity False Negetive)为在整个视频中检测目标被漏分配的数量,IDFP(Identity False Positives)为在整个视频中检测目标被错误分配的数量。

对于单目标跟踪,使用最多的是 Wu 等[53]提出的评价指标:中心位置误差 CLE(Center Location Error)和区域重叠率 OP(Overlap Precision)。CLE 表示预测目标中心位置与真实目标中心位置的像素距离,该值越大,说明误差越大。定义 CLE 为:

$$\mathrm{CLE}=\sqrt{(X_2-X_1)^2-(Y_2-Y_1)^2} \tag{5}$$

式中:X_1、X_2 和 Y_1、Y_2 分别表示预测目标中心与真实目标中心的横纵坐标。OP 以面积来衡量,指预测的目标边界框面积与真实目标边界框面积的比率。定义 OP 为:

$$\mathrm{OP}=\frac{R_1\cap R_2}{R_1\cup R_2} \tag{6}$$

式中:R_1、R_2 分别表示算法预测边界框区域和人工标注框区域。

4　未来展望

随着深度学习技术的发展,目标跟踪在无人机应用中将越来越重要,市场应用范围广、潜力巨大。对无人机视觉目标跟踪未来的展望有如下几个方面:①未来的研究应着眼于增强特征信息,以提高目标跟踪算法的准确性和鲁棒性。一种方法是整合更多的

上下文和语义信息,例如将深度特征和语义信息集成到相关的过滤算法中。②此外,联合学习目标检测和跟踪任务也是未来研究的重点,因为它可以提高算法的目标检测和跟踪性能,从而使算法更能适应复杂的场景。③最后,动态尺度调整、变形建模和自适应学习也是未来研究的重点。例如,可以探索如何通过变形建模来解决目标变形问题,以提高基于相关滤波、孪生网络和 Transformer 算法的鲁棒性。

尽管现今面临着许多难题,相信随着技术的发展无人机视觉目标跟踪将在跟踪速度、跟踪精度等方面取得巨大的成功。

5 结束语

本文介绍了无人机视觉目标跟踪的研究现状,并总结与分析相关工作,结合无人机的发展与应用需求,指出未来的研究可以从增强特征信息、联合学习目标检测和跟踪任务、动态尺度调整、变形建模和自适应学习等方面考虑。综合而言,若要满足无人机视觉目标跟踪的需求,跟踪算法应具备更高的运算速率、准确率以及更低的计算内存。

参考文献:

[1] Kim H, Mokdad L, Ben-Othman J. Designing UAV Surveillance Frameworks for Smart City and Extensive Ocean With Differential Perspectives[J]. IEEE Communications Magazine, 2018, 56(4): 98-104.

[2] Xie J, Stensrud E, Skramstad T. Detection-based Object Tracking Applied to Remote Ship Inspection[J]. Sensors, 2021, 21(3): 761.

[3] Popescu D, Stoican F, Stamatescu G, et al. Advanced UAV-WSN System for Intelligent Monitoring in Precision Agriculture[J]. Sensors, 2020, 20(3): 817.

[4] Nayak S, Nagesh B, Routray A, et al. A Human-Computer Interaction Framework for Emotion Recognition Through Time-series Thermal Video Sequences[J]. Computers & Electrical Engineering, 2021, 93: 107280.

[5] Liu S, Liu D, Srivastava G, et al. Overview and Methods of Correlation Filter Algorithms in Object Tracking[J]. Complex & Intelligent Systems, 2021, 7(4): 1895-1917.

[6] Yang Z Y, Yu X Y, Dedman S, et al. UAV Remote Sensing Applications in Marine Monitoring: Knowledge Visualization and Review[J]. Science of the Total Environment, 2022, 838: 155939.

[7] Nikhil N, Shreyas S M, Vyshnavi G, et al. Unmanned Aerial Vehicles (UAV) in Disaster Management Applications [C]//2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT), Tirunelveli, India, IEEE, 2020: 140-148.

[8] Balamuralidhar N, Tilon S, Nex F. MultEYE: Monitoring System for Real-time Vehicle Detection, Tracking and Speed Estimation from UAV Imagery on Edge-computing Platforms[J]. Remote sensing, 2021, 13(4): 573.

[9] Fernández-Caramés T M, Blanco-Novoa O, Froiz-Míguez I, et al. Towards an Autonomous Industry 4.0 Warehouse: A UAV and Blockchain-based System for Inventory and Traceability Applications in Big Data-driven Supply Chain Management[J]. Sensors, 2019, 19(10): 2394.

[10] Li B W, Fu C H, Ding F Q, et al. All-day Object Tracking for Unmanned Aerial Vehicle[J]. IEEE Transactions on Mobile Computing, 2023, 22(8): 4515-4529.

[11] 戴家树. 视觉场景下的运动目标发现与跟踪算法研究[D]. 哈尔滨:哈尔滨工程大学,2015.

[12] Ali S, Shah M. COCOA: Tracking in Aerial Imagery [C]//2016 Defense and Security symposium, Orlando, FL, USA, 2006: 105-114.

[13] Siebel N T, Maybank S. The ADVISOR Visual Surveillance System[C]//ECCV 2004 Workshop on Applications of Computer Vision, Prague, Czech, 2004: 103-111.

[14] Prokaj J, Zhao X M, Medioni G. Tracking Many Vehicles in Wide Area Aerial Surveillance[C]//2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, Providence, RI, USA, IEEE, 2012: 37-43.

[15] 管皓,薛向阳,安志勇. 深度学习在视频目标跟踪中的应用进展与展望[J]. 自动化学报,2016,42(6):834-847.

[16] Huang K, Tan T. Vs-star: A Visual Interpretation System for Visual Surveillance[J]. Pattern Recognition Letters, 2010, 31(14): 2265-2285.

[17] 王国坤. 复杂背景下基于深度学习的单目标跟踪算法研究[D]. 合肥:中国科学技术大学,2018.

[18] 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, Seattle, WA, USA, 2020: 11920-11929.

[19] Yan B, Zhang X Y, Wang D, et al. Alpha-refine: Boosting Tracking Performance by Precise Bounding Box Estimation [C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 2021: 5285-5294.

[20] Zhang C H, Huang G J, Liu L, et al. WebUAV-3M: A Benchmark Unveiling the Power of Million-scale Deep UAV Tracking[J]. Pattern Analysis and Machine Intelligence, 2023, 45(7): 9186-9205.

[21] 李玺,查宇飞,张天柱,等. 深度学习的目标跟踪算法综述[J]. 中国图象图形学报,2019,24(12):2057-2080.

[22] Bolme D S, Beveridge J R, Draper B A, et al. Visual object Tracking Using Adaptive Correlation Filters[C]//2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, CA, USA, IEEE, 2010:2544-2550.

[23] Danelljan M, Häger G, Khan F, et al. Accurate Scale Estimation for Robust Visual Tracking[C]//British Machine Vision Conference 2014, Nottingham, UK, 2014:C.28.65.

[24] Henriques J F, Caseiro R, Martins P, et al. Exploiting the Circulant Structure of Tracking-by-detection With Kernels [C]//12th European Conference on Computer Vision, Florence, Italy, Springer, 2012:702-715.

[25] 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, 2014, 37(3):583-596.

[26] Li Y, Zhu J K. A Scale Adaptive Kernel Correlation Filter Tracker With Feature Integration[C]//Computer Vision-ECCV 2014 Workshops, Zurich, Switzerland, Springer, 2015:254-265.

[27] Galoogahi H K, Fagg A, Lucey S. Learning Background-aware Correlation Filters for Visual Tracking[C]//2017 IEEE International Conference on Computer Vision, Venice, Italy, 2017:1144-1152.

[28] Wang M M, Liu Y, Huang Z Y. Large Margin Object Tracking With Circulant Feature Maps[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 2017:4800-4808.

[29] Zheng Y, Liu X, Xiao B, et al. Multi-task Convolution Operators With Object Detection for Visual Tracking[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2021, 32(12):8204-8216.

[30] Li S, Liu Y, Zhao Q, et al. Learning Residue-aware Correlation Filters and Refining Scale for Real-time UAV Tracking[J]. Pattern Recognition, 2022, 127:108614.

[31] Lin F, Fu C, He Y, et al. ReCF:Exploiting Response Reasoning for Correlation Filters in Real-time UAV Tracking [J]. IEEE Transactions on Intelligent Transportation Systems, 2021, 23(8):10469-10480.

[32] Chen L, Liu Y. A Robust Spatial-temporal Correlation Filter Tracker for Efficient UAV Visual Tracking[J]. Applied Intelligence, 2023, 53(4):4415-4430.

[33] Bertinetto L, Valmadre J, Henriques J F, et al. Fully-convolutional Siamese Networks for Object Tracking[C]//14th European Conference on Computer Vision Amsterdam Netherland, Springer, 2016:850-865.

[34] 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, UT, USA, IEEE, 2018:8971-8980.

[35] Li B, Wu W, Wang Q, et al. SiamRPN++:Evolution of Siamese Visual Tracking With Very Deep Networks[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR), Long Beach, CA, USA, IEEE, 2020:4277-4186.

[36] Wang Q, Zhang L, Bertinetto L, et al. Fast Online Object Tracking and Segmentation: A Unifying Approach[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 2019:1328-1338.

[37] Xu Y D, Wang Z Y, Li Z X, et al. SiamFC++:Towards Robust and Accurate Visual Tracking With Target Estimation Guidelines[J]. 34th AAAI Conference on Artificial Intelligence, New York, NY, USA, 2020, 34(7):12549-12556.

[38] Zhang Z P, Peng H W. Deeper and Wider Siamese Networks for Real-time Visual Tracking[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 2019:4586-4595.

[39] Sun L, Yang Z, Zhang J, et al. Visual Object Tracking for Unmanned Aerial Vehicles Based on the Template-Driven Siamese Network[J]. Remote Sensing, 2022, 14(7):1584.

[40] Zhu M, Zhang H, Zhang J, et al. Multi-level Prediction Siamese Network for Real-time UAV Visual Tracking[J]. Image and Vision Computing, 2020, 103:104002.

[41] Chen X, Yan B, Zhu J W, et al. Transformer Tracking [C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 2021: 8126-8135.

[42] Wang N, Zhou W G, Wang J, et al. Transformer Meets Tracker: Exploiting temporal Context for Robust Visual Tracking[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 2021:1571-1580.

[43] Yan B, Peng H W, Fu J L, et al. Learning Spatio-temporal Transformer for Visual Tracking[C]//2021 IEEE/CVF International Conference on Computer Vision, Montreal, QC, Canada, 2021:10428-10437.

[44] Cui Y T, Jiang C, Wang L M, et al. Target Transformed Regression for Accurate Tracking[EB/OL]. (2021-04-01). https://arxiv.org/abs/2104.00403.

[45] Ye J, Fu C, Cao Z, et al. Tracker Meets Night: A Transformer Enhancer for UAV Tracking[J]. IEEE Robotics and Automation Letters, 2022, 7(2):3866-3873.

[46] 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 Computer Vision, Waikoloa, HI, USA, 2022:2139-2148.

[47] Lu K, Fu C, Wang Y, et al. Cascaded Denoising Transformer for UAV Nighttime Tracking[J]. IEEE Robotics and Automation Letters, 2023, 8(6): 3142-3149.

[48] Robicquet A, Sadeghian A, Alahi A, et al. Learning Social Etiquette: Human Trajectory Understanding in Crowded Scenes[C]//14th European Conference on Computer Vision, Amsterdam, Netherlands, Springer, 2016: 549-565.

[49] Mueller M, Smith N, Ghanem B. A Benchmark and Simulator for UAV Tracking[C]//14th European Conference on Computer Vision, Springer, 2016: 445-461.

[50] Li S Y, Yeung D Y. Visual Object Tracking for Unmanned Aerial Vehicles: A Benchmark and New Motion Models [C]//31st AAAI Conference on Artificial Intelligence, San Francisco, CA, USA, 2017, 31(1): 4140-4146.

[51] Du D W, Qi Y K, Yu H Y, et al. The Unmanned Aerial Vehicle Benchmark: Object Detection and Tracking[C]// 15th European Conference on Computer Vision, Munich, Germany, 2018: 375-391.

[52] Zhu P F, Wen L Y, Du D W, et al. Detection and Tracking Meet Drones Challenge[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2022, 44 (11): 7380-7399.

[53] Wu Y, Lim J, Yang M H. Online Object Tracking: A Benchmark[C]//2013 IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR, USA, 2013: 2411-2418.

刘丽媛(1999—),女,南宁师范大学在读硕士研究生,主要研究方向为空天信息网络,2729091518@qq.com;

曾 倩(1989—),女,通讯作者,博士,南宁师范大学讲师,主要研究方向为空天信息网络、协作通信和传感器网络,zengqian617@foxmail.com。