

文章编号:1007-2780(2026)06-0908-16

结合注意力与抗遮挡策略的无人机航拍目标跟踪

赵 萌¹, 郑滨汐^{2*}, 孙 翥², 马铭阳², 宋 策², 白金成²

(1. 中国人民解放军海军装备部驻沈阳地区军事代表局, 辽宁 沈阳 110000;

2. 中国科学院 长春光学精密机械与物理研究所, 吉林 长春 130033)

摘要:为了解决无人机在目标跟踪任务中存在的复杂干扰、目标遮挡等问题,提出了一种引入注意力机制和目标遮挡判别机制的孪生神经网络目标跟踪算法 SiamAL, 以及其融合目标检测算法的变体 SiamAL-Det。基于 SiamDMU 算法, 首先在主干网络引入注意力机制, 同时采用新的激活函数 h-swish, 提升了复杂干扰条件下的特征表征能力; 其次, 引入基于卡尔曼滤波的运动先验对目标轨迹进行估计, 利用跟踪器置信度联合预测-观测残差对跟踪状态进行分级判别: 在严重遮挡时, 自适应提高运动先验权重, 从而在遮挡期间保持轨迹连续性; 最后设计一种融合检测与跟踪的算法框架来提升实际跟踪效果, 在判定跟踪失败或严重漂移时, 触发轻量检测分支进行重检测与纠正。在数据集 UAV123 上, SiamAL 以及 SiamAL-Det 算法的成功率和精确率分别达到了 65.3%、86.6% 和 68.6%、88.8%, 相比基准算法 SiamDMU 分别提升了 2.5%、3.7% 和 5.8%、5.9%。与多种主流基于孪生网络的跟踪算法对比, 成功率和精确率表现更优, 表明提出的方法在无人机航拍场景中具有良好的应用前景。

关键词:无人机航拍; 孪生目标跟踪; 注意力机制; 卡尔曼滤波

中图分类号: TP391.4 **文献标识码:** A **doi:** 10.37188/CJLCD.2026-0076 **CSTR:** 32172.14.CJLCD.2026-0076

UAV aerial target tracking with attention mechanism and anti-occlusion strategies

ZHAO Meng¹, ZHENG Binxi^{2*}, SUN He², MA Mingyang², SONG Ce², BAI Jincheng²

(1. Shenyang Area Office of the Naval Equipment Department, People's Liberation Army Navy,
Shenyang 110000, China;

2. Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences,
Changchun 130033, China)

Abstract: To address the challenges of complex interference and target occlusion in UAV target tracking, a Siamese network-based tracker named SiamAL is proposed, which incorporates an attention mechanism and an occlusion-discrimination module, along with its variant SiamAL-Det that additionally integrates an object detection module. Based on the SiamDMU algorithm, an attention mechanism is first introduced into the backbone network, and the h-swish activation function is adopted to enhance feature representation under complex interference conditions. A Kalman-filter-based motion prior is then introduced to estimate

收稿日期: 2026-04-30; 修订日期: 2026-05-17.

基金项目: 国家自然科学基金 (No. 62205332, No. 62305335)

Supported by National Natural Science Foundation of China (No. 62205332, No. 62305335)

*通信联系人, E-mail: zhengbinxi@hrbeu.edu.cn

the target trajectory, and the tracking state is evaluated based on tracker confidence and prediction-observation residuals. In cases of severe occlusion, the motion prior weight is adaptively increased to maintain trajectory continuity. Finally, an integrated detection-and-tracking framework is designed to enhance tracking performance. A lightweight detection branch is triggered for re-detection and correction upon tracking failure or severe drift. Experimental results on the UAV123 dataset show that SiamAL achieves a success rate of 65.3% and a precision rate of 86.6%, while SiamAL-Det achieves 68.6% and 88.8%, respectively. In comparison with several mainstream Siamese network-based trackers, SiamAL outperforms in both success rate and precision, demonstrating the promising application prospects of the proposed method in UAV aerial tracking scenarios.

Key words: UAV aerial imagery; siamese tracking; attention mechanism; Kalman filter

1 引 言

目标跟踪是计算机视觉领域的基础问题,其任务是在连续视频序列中根据上下文信息估计目标的实时位置与尺度变化^[1]。随着深度学习的发展,基于孪生网络的跟踪方法通过将跟踪建模为“模板-搜索区域”之间的相似性匹配,在精度与速度之间取得了良好平衡。SiamFC^[2]和SiamRPN++^[3]等代表性工作利用全卷积结构和区域建议网络,将单目标跟踪性能推向新的水平,并在通用场景中逐渐取代传统生成式和相关滤波跟踪器。早期孪生网络跟踪器,如SiamFC将跟踪视为模板与搜索区域特征之间的跨相关运算,依靠端到端训练获得强判别性并实现较高帧率。随后,SiamRPN^[4]系列在此基础上引入区域建议网络(RPN),显著提升了定位精度和尺度估计能力。然而,这类方法通常采用固定模板,难以适应长序列中的外观剧变与遮挡。为此,研究者从模板更新、在线学习与注意力机制等方面对孪生结构进行了持续扩展:SiamATL^[5]通过注意力迁移学习实现模板与环境的在线更新,缓解模板污染问题;SiamBAN^[6]利用块级注意力增强通道间依赖并设计在线更新模块,提升了在UAV123^[7]等基准上的鲁棒性;SiamRPN++在SiamRPN的基础上引入了深层网络作为特征提取器,解决了深层网络带来的平移不变性问题。在SiamRPN++基础上,Su等人同时利用语义掩码和运动掩码增强模板更新并设计了三输入的模板更新网络,提出了SiamDMU算法^[8]。近年来,随着新型光学成像与智能感知技术的发展,目标成像与跟踪不再局限于传统可见光视频

序列,量子成像、超表面成像等新型成像机制也在目标感知、量子目标跟踪和紧凑化光学系统中逐渐展现出潜在应用价值^[9]。另一方面,深度学习与光学传感系统的结合已成为智能感知领域的重要趋势^[10]。因此,面向复杂成像条件的高鲁棒目标跟踪算法具有重要研究意义。

然而,上述方法大多在通用场景或固定视角数据集上验证,其在复杂无人机航拍条件下的性能仍然受限。无人机平台具有成本低、机动性强、部署灵活等优势,已广泛应用于环境监测^[11]、地质勘探^[12]、电力巡检^[13]和搜索救援^[14]等任务,对高鲁棒、低时延的视觉跟踪提出了迫切需求。与地面固定相机不同,UAV航拍多为俯视或大俯仰角视角,目标在成像中像素占比小、形变剧烈并伴有明显畸变,同时还常出现视点急剧变化、遮挡、出视野以及背景中大量相似干扰目标等问题。此外,机载处理器算力与功耗受限,要求跟踪算法在保持高精度的同时具备实时性乃至高速运行能力^[15]。为应对无人机航拍场景的多重挑战,近几年涌现出一批专门面向无人机航拍场景的孪生跟踪方法。

现有方法主要围绕结构设计、时序建模与模板管理以及遮挡与外观变化鲁棒性3个方向展开。在结构设计方面,SiamAD^[16]在SiamRPN++的基础上进行改进,在主干网络加入混合注意力模块融合多层特征,强化对小型无人机目标的判别;并配合层次判别器、重检测分支与模板更新,对跟踪结果进行可靠性评估与失败纠正,相较SiamRPN++的跟踪效果提升明显。SiamSTA^[17]针对野外红外反无人机监测,构建局部搜索与全局重检测的两阶段孪生框架:通过时空注意力约束候选生成以抑制长距离采集中小目标干扰,并在快速运动

导致局部跟踪丢失时,依托基于变化检测的相关滤波器在全局范围重捕获目标,再由状态感知切换机制自适应融合两种模式的优势。李大湘等提出位置敏感的 Transformer 航拍图像目标检测模型,通过多尺度特征融合和位置敏感自注意力增强小目标检测能力,为航拍场景下的目标重检测和跟踪纠偏提供了有益参考^[18]。

考虑长时序信息的重要性,部分方法显式建模时间上下文并引入多模板机制。MT-Track^[19]在孪生框架内设计多步时间建模模块,先利用时间相关模块动态更新模板特征,再通过 Transformer 联合优化历史与当前相关图,在多种无人机航拍数据集上实现了良好的准确率。BACTrack^[20]面向航拍场景构建在线外观模板集,借助混合时间 Transformer 高效建立搜索区域与多模板的时序联系,并通过外观判别器实现自适应模板更新,在多个航拍数据集上表现良好。

此外,也有工作从模板库扩展与遮挡检测等方面提升鲁棒性。TDSiam^[21]为适应无人机视频中频繁的外观变化,在特征提取分支中增加模板库分支与特征对齐模块,通过中心对齐融合改善匹配质量,并设计遮挡检测策略减弱遮挡噪声,在 UAV123 数据集上性能优于多种现有算法。综上,现有方法存在共性短板:模型复杂、难以嵌入低算力平台;遮挡后预测和检测重捕获的协同还不充分;注意力和模板更新多针对通用场景,缺乏与无人机特有成像的深度结合。针对上述方法存在的问题,本文提出了一种更适合应用于无人机航拍场景的算法 SiamAL,以及其融合目标检测算法的变体 SiamAL-Det。

以 SiamDMU 算法为基础,首先提出 H-ResNet-50 特征增强主干网络,将注意力机制与 h-swish^[22]用于航拍小目标和复杂背景下的特征表征;然后提出基于置信度和预测-观测残差联合判别的自适应卡尔曼融合策略^[23];最后提出异常状态触发的检测辅助重捕获框架 SiamAL-Det,而不是每帧检测,在精度和效率之间取得平衡,可以适应复杂的无人机航拍环境。所提算法在无人机航拍视角数据集 UAV123 上进行了测试。实验结果表明,与多种主流基于孪生网络的跟踪算法相比,所提算法在无人机视角跟踪场景中的跟踪性能成功率和精确率表现更优。

2 算法框架和原理

2.1 难点分析

由于无人机的飞行特点,工作过程中得到的航拍图像和常规视角图像有着明显区别:首先,常规视角的跟踪数据集大多数由固定相机拍摄,大部分图像为侧视视角,而无人机在航拍过程中往往是由上而下拍摄的俯视视角,具有的特征并不相同,且航拍过程中的目标往往在图像中像素占比小并存在一定的失真现象;其次,在航拍过程中存在地面、楼房和树木等复杂干扰场景;最后,由于长时跟踪拍摄,目标常被遮挡,为目标跟踪增加了难度^[24]。因此,设计一款针对无人机航拍视角的目标跟踪算法显得尤为重要^[25]。

2.2 算法框架

本文提出了一种更适合无人机航拍目标跟踪的算法 SiamAL。在原有 ResNet-50^[26]主干网络基础上加入 CBAM 注意力机制^[27]和效果更好的激活函数得到改进的主干网络 H-ResNet-50,用来解决复杂背景下低效特征提取导致的跟踪表现不佳问题。之后将 RPN、边界回归和分类分支输出的候选框作为跟踪观测结果,并结合跟踪置信度与预测-观测残差进行状态判别,同时利用基于卡尔曼滤波的运动先验对目标轨迹进行预测,若存在遮挡则启用目标遮挡判别机制对被遮挡的目标位置进行预测,从而有效保持目标轨迹的连续性,算法框架如图 1 所示。图 1 右半部分为融合卡尔曼滤波的遮挡判别与自适应融合模块。SiamAL 跟踪器输出候选框和置信度, Kalman 分支给出运动预测结果,二者共同用于计算预测-观测残差,并结合置信度完成正常跟踪、轻度干扰和严重遮挡 3 种状态判别,最终根据状态选择跟踪器输出、加权融合输出或 Kalman 预测输出。

2.3 主干网络设计

对主干网络的改进包含两部分:在 ResNet-50 主干网络中加入 CBAM 注意力机制;对内部激活函数进行改进。最终得到改进增强的主干网络,称为 H-ResNet50。

2.3.1 CBAM 轻量注意力机制

为了提升无人机航拍在复杂干扰场景下的跟踪效果,使模型更加关注图像中的关键信息,对背景等无用信息的干扰进行有效抑制,从而

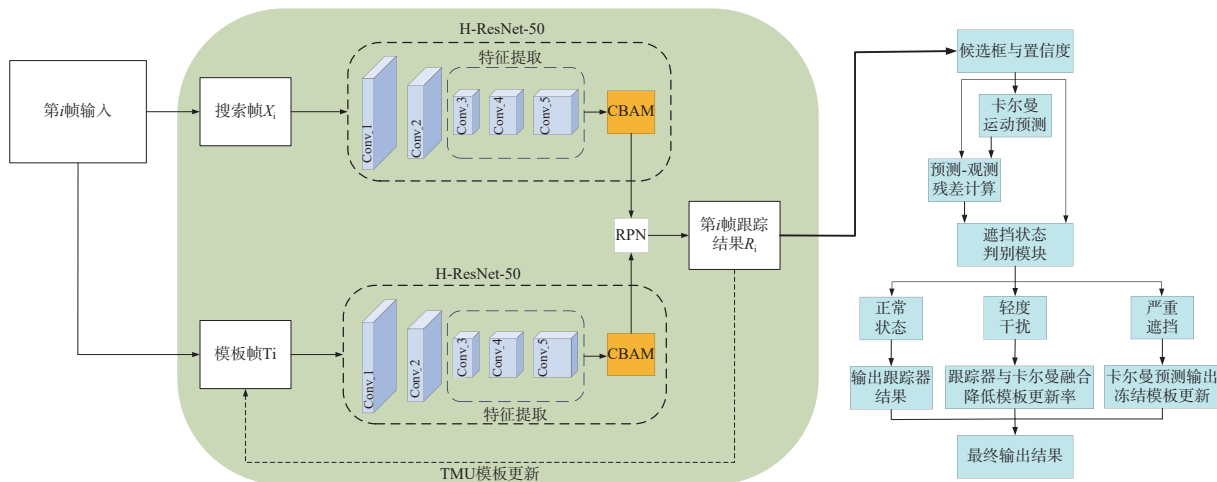


图1 SiamAL总体框架

Fig. 1 Overall structure of SiamAL

进一步提升模型的准确度。在光学图像处理等任务中,注意力机制与残差结构已被广泛用于增强有效特征表达^[28],在主干网络中引入CBAM以突出小目标与抑制背景噪声,已被证明能显著提升模型精度。同时,相较于其他注意力机制,CBAM可以在增强小目标特征和轻量化之间实现平衡。针对边界模糊、特征不明显和背景干扰强的问题,引入注意力机制能够强化难点区域特征表达并抑制虚警信息^[29]。与ECANet等注意力相比,CBAM可同时在通道和空间维度自适应重标定特征,更适合从高空视角突出尺度极小、稀疏分布的目标,并抑制大面积复杂地物背景干扰。相较于GAM等全局注意力机制,CBAM结构简单,可在嵌入式平台上满足实时性要求。因此在SiamAL算法中引入了CBAM注意力机制。CBAM结构如图2所示,包含通道注意力模块(CAM)和空间注意力模块(SAM)两个相互独立的部分,可以在通道维度和空间维度分别进行操作。输入特征 F 先进行通道注意力得到 F' ,再进行空间注意力得到最终输出的 F_a 。

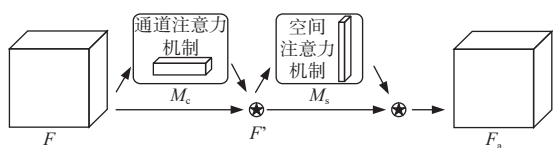


图2 CBAM注意力机制结构框架

Fig. 2 Structure of CBAM attention mechanism

2.3.2 通道注意力机制

如图3所示,将输入的特征图 F 分别经过全局平均池化和全局最大池化,原维度为 $H \times W \times C$ 发生变化,得到两个维度均为 $1 \times 1 \times C$ 的特征。将两个特征分别输入到一个多层感知器当中,其中第一层和第二层的神经元个数分别是 C/r 和 C , r 代表多层感知器的减少率。激活函数选用ReLU函数。对多层感知器的两个输出进行加和再经过激活操作得到最终的通道注意力特征 M_c ,最后将通道注意力特征和原输入特征 F 相乘,得到输出特征 F' 。计算公式如式(1)所示:

$$M_c(F) = \sigma(\text{MLP}(\text{AvgPool}(F)) + \text{MLP}(\text{MaxPool}(F))) = \sigma(W_1(W_0(F_{\text{avg}}^c)) + W_1(W_0(F_{\text{max}}^c))), \quad (1)$$

其中: σ 代表sigmoid函数,MLP代表多层感知操作, W_0 和 W_1 分别代表多层感知器的两层神经网络。

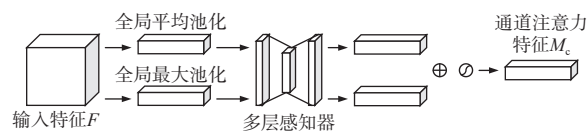


图3 CAM注意力机制结构框架

Fig. 3 Structure of CAM attention mechanism

2.3.3 空间维度注意力机制

将经过通道注意力机制模块输出的特征 F' 作为本模块空间维度注意力机制(图4)的输入。首先对通道分别进行全局最大池化和全局平均

池化,再将结果进行拼接,得到的新特征再经过卷积,此时维度变为 $H \times W \times C$,经过 Sigmoid 函数得到空间注意力特征 M_s ,最后将 M_s 和 F' 相乘得到最终输出的特征 F_a 。计算公式如式(2)所示:

$$M_s(F) = \sigma(f([\text{AvgPool}(F); \text{MaxPool}(F)])) = \sigma(f([F_{\text{avg}}^s; F_{\text{max}}^s])). \quad (2)$$

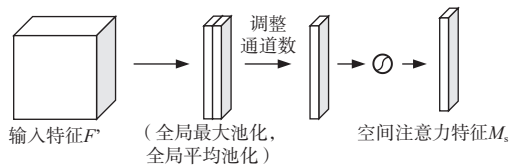


图4 SAM注意力机制结构框架

Fig. 4 Structure of SAM attention mechanism

2.3.4 加入CBAM注意力机制

在主干网络的 conv_5 之后、RPN 之前加入 CBAM 注意力机制,输出的新特征与原始特征相比,分辨率和特征通道数保持不变。

分别选取 ResNet-50 和结合注意力机制的 ResNet-50+CBAM 在无人机航拍视角下提取目标特征,特征热力图如图 5 所示,其中图 5(a)为可见光图像,图 5(b)为原 ResNet-50 提取目标特征的效果,图 5(c)为结合 CBAM 后提取目标特征的效果。可以看出,加入注意力机制后,特征提取时的高响应部分更集中于目标附近,降低了跟踪过程中树木等造成的干扰。

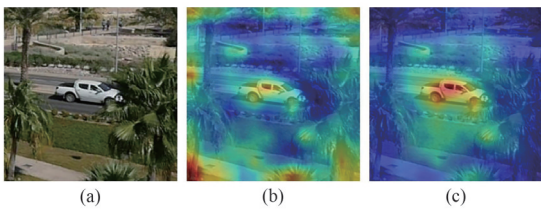


图5 特征热力图。(a)原图;(b)原特征热力图;(c)增强特征图。

Fig. 5 Feature heatmap. (a) Original image; (b) Original feature heatmap; (c) Enhanced feature heatmap.

2.3.5 激活函数的改进

相较于原始主干网络结构,本文在特征提取阶段引入 h-swish 激活函数,对 ResNet-50 进行改造并构建新的 H-ResNet-50 主干网络。h-swish 是 Google 在 MobileNetV3 中提出的激活函数,作

为一种平滑且具有良好非线性表达能力的激活函数,相比 ReLU 等传统激活函数在低响应区域保留了更多有用信息,有利于小目标与弱纹理特征的特征与传播。在主干网络中,仅将 conv_3、conv_4 和 conv_5 部分的 ReLU 函数替换为 h-swish 函数,可以在保持低层纹理特征稳定性的同时增强中高层语义特征表达。依托于 h-swish 的引入,H-ResNet-50 在保证网络深度与参数规模基本不变的前提下,有效提升了特征分布的连续性与可分性,从而增强了对复杂场景下目标外观变化的鲁棒性。h-swish 函数的表达式见式(3),函数图像如图 6 所示。

$$\text{h-swish}(x) = x \times \min(\max((x+3), 0), 6)/6. \quad (3)$$

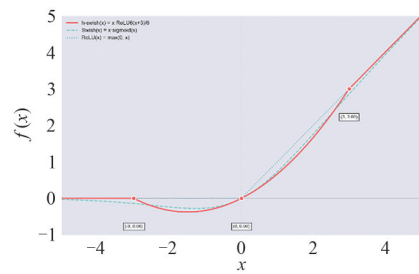


图6 h-swish激活函数

Fig. 6 h-swish activation function

2.4 抗遮挡策略:融合卡尔曼滤波的目标遮挡判别机制

2.4.1 卡尔曼滤波工作原理

卡尔曼滤波是一种在线性系统且噪声服从高斯分布假设下,对动态系统状态序列进行最小均方误差估计的递推算法。针对无人机航拍目标跟踪中出现的短时遮挡情形,由于相邻帧时间间隔较短,通常可采用恒速度模型近似描述目标在局部时间窗内的运动状态,此时卡尔曼滤波的预测功能可以对被遮挡帧中目标的运动状态进行较为准确的估计,从而在遮挡期间保持轨迹的连续性^[17]。在视频序列目标跟踪任务中,可通过建立状态预测方程和观测方程对目标进行估计,其离散形式如式(4)和式(5)所示:

$$x_t = Ax_{t-1} + w_{t-1}, \quad (4)$$

$$z_t = Hx_t + v_t, \quad (5)$$

式中: A 和 H 分别为状态转移矩阵和观测矩阵; x_t 和 z_t 分别代表 t 时刻目标的状态估计和观测向

量,一般由当前帧中目标的观测位置构成。分别使用目标的位置、速度信息以及当前时刻被观测到的位置信息作为二者的参数, w_{t-1} 和 v_t 分别是预测和观测过程中的噪声。根据式(4)和式(5)可以对跟踪过程的目标进行预测。当目标处于遮挡状态时,卡尔曼滤波可以根据前一帧的信息对后续帧的跟踪情况进行可靠预测。

2.4.2 目标遮挡判别机制

在跟踪算法当中,诸多参数都可以用作对目标跟踪效果进行评估和判断,如最大峰值响应、平均峰值相关能量等。然而以上参数的计算涉及整张响应图的统计,较直接使用网络输出的置信度分数更为复杂,在轻量级系统中可能对实时性造成一定影响。相比较而言,SiamAL算法的置信度得分可以对跟踪目标和模板帧的相似程度进行有效的反映。图7选取UAV123数据集中car12序列的前100帧进行测试,并对置信度得分情况进行统计。由图7可知,第16帧、28帧以及87帧跟踪序列出现较明显的目标遮挡情况,此时置信度得分出现了降低情况。另外,置信度分数直接来源于网络前向结果,无需额外复杂运算,相比其他指标几乎不增加额外计算,更适合作为实时场景下的遮挡判别。

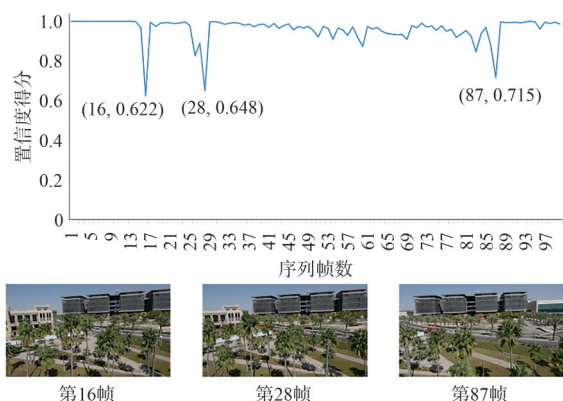


图7 SiamAL算法序列跟踪置信度得分曲线图

Fig.7 Confidence score curve of SiamAL algorithm for sequence tracking

然而在遮挡期间无条件用卡尔曼滤波进行位置预测会导致位置误差累积,在目标遮挡解除后难以及时重捕获。针对该问题,本文提出了更合理的目标遮挡判别机制:SiamAL负责输出目标候选框及对应置信度得分,卡尔曼滤波器对目

标位置进行递推估计;同时结合SiamAL置信度,设计连续而非开关式的自适应更新机制。

2.4.3 融合卡尔曼滤波的跟踪算法

将置信度划分为高置信区间与低置信区间,并以跟踪器的置信度作为观测噪声的调节因子:置信度越低,对应测量噪声协方差矩阵越大,从而自适应减小卡尔曼增益,弱化可疑观测对状态的影响。同时计算观测与预测的欧氏残差,将其与置信度联合构成遮挡分级判据,用于区分正常跟踪、轻度干扰与严重遮挡3种状态。正常状态下执行完整“预测-更新”过程并允许减缓模板更新并直接输出跟踪器的结果;轻度干扰时降低学习率,同时采用跟踪器结果与卡尔曼预测结果加权融合的输出结果;严重遮挡且连续多帧满足条件时,将当前帧视作“无效观测帧”,只保留卡尔曼预测并冻结模板更新。与直接通过阈值触发卡尔曼滤波策略相比,通过置信度调节测量噪声,实现观测与运动先验的连续自适应融合,可以有效避免目标遮挡场景下的跟踪漂移。卡尔曼滤波与遮挡状态判别的相关参数设置见表1。

表1 卡尔曼滤波与遮挡状态判别参数设置

Tab.1 Parameter settings of Kalman filtering and occlusion-state discrimination

参数	含义	设置
Δt	相邻帧时间间隔	1
x_0	初始状态向量	首帧目标框初始化
P_0	初始状态协方差	$\text{diag}(4, 4, 100, 100, 16, 16)$
Q	过程噪声协方差	$\text{diag}(2.25, 2.25, 9, 9, 4, 4)$
λ	遮挡评分权重	0.5
T_1	正常/轻度干扰阈值	0.420
T_2	轻度干扰/严重遮挡阈值	0.720
N_{occ}	严重遮挡连续帧数	3
R_{min}	最小观测噪声协方差	$\text{diag}(4, 4, 16, 16)$
R_{max}	最大观测噪声协方差	$\text{diag}(225, 225, 900, 900)$

由于首帧目标框由初始化框给出,目标中心位置和尺度相对可靠,而初始速度无法由单帧直接确定,因此将位置和尺度分量的初始不确定性设置较小,将速度分量的初始不确定性设置较大,初始状态协方差设置见表1。其中,中心位置、速度和尺度对应的标准差分别为2像素、10像素/帧和4像素。过程噪声协方差用于描述真实目标运动

与恒速度模型之间的偏差。考虑到无人机航拍中存在平台运动、目标尺度变化和短时遮挡等因素,本文设置过程噪声协方差如表1所示,分别对应位置、速度和尺度变化的不确定性。该设置能够在保持运动预测平滑性的同时,为目标短时加速和尺度变化保留调整空间。上述协方差参数根据UAV123遮挡序列上的跟踪稳定性和恢复效果进行经验设置,并在所有测试序列中保持一致。

2.4.4 卡尔曼滤波状态更新过程分析

为了进一步明确卡尔曼滤波在跟踪过程中的作用,本文将SiamAL跟踪器输出的目标框视为观测量,将卡尔曼滤波器预测结果视为运动先验。在第 k 帧中,跟踪器首先输出候选框及其置信度,卡尔曼滤波器根据上一帧状态预测当前帧目标状态并得到预测框。随后计算预测位置与观测位置之间的欧氏残差,并结合置信度进行遮挡状态判别,区分正常跟踪、轻度干扰和严重遮挡3种状态。正常状态下执行完整的预测-更新过程并输出跟踪器结果;轻度干扰状态下增大观测噪声、降低模板更新率;严重遮挡时,将当前观测视为无效观测,跳过量测更新,仅输出卡尔曼预测结果并冻结模板更新,以避免遮挡期间因错误观测导致的轨迹漂移和模板污染。融合卡尔曼滤波的遮挡状态更新流程如图8所示。

2.4.4.1 状态变量定义

针对无人机航拍目标跟踪任务,本文采用恒速度运动模型描述目标在短时间段内的运动状态。第 k 帧目标状态向量定义为:

$$\mathbf{x}_k = [c_x, c_y, v_x, v_y, w, h]^T, \quad (6)$$

其中: c_x 和 c_y 表示目标框中心坐标, v_x 和 v_y 表示目标在水平方向和垂直方向上的速度, w 和 h 分别表示目标框宽度和高度。SiamAL跟踪器在第 k 帧输出的候选框表示为:

$$\mathbf{b}_k^{\text{trk}} = [c_x^{\text{trk}}, c_y^{\text{trk}}, w_k^{\text{trk}}, h_k^{\text{trk}}]. \quad (7)$$

因此,观测向量定义为:

$$\mathbf{z}_k = [c_x^{\text{trk}}, c_y^{\text{trk}}, w_k^{\text{trk}}, h_k^{\text{trk}}]^T. \quad (8)$$

对应的观测矩阵为:

$$\mathbf{H} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}. \quad (9)$$

该观测矩阵表示跟踪器能够直接观测目标中心位置和目标尺度,但不能直接观测目标速度。

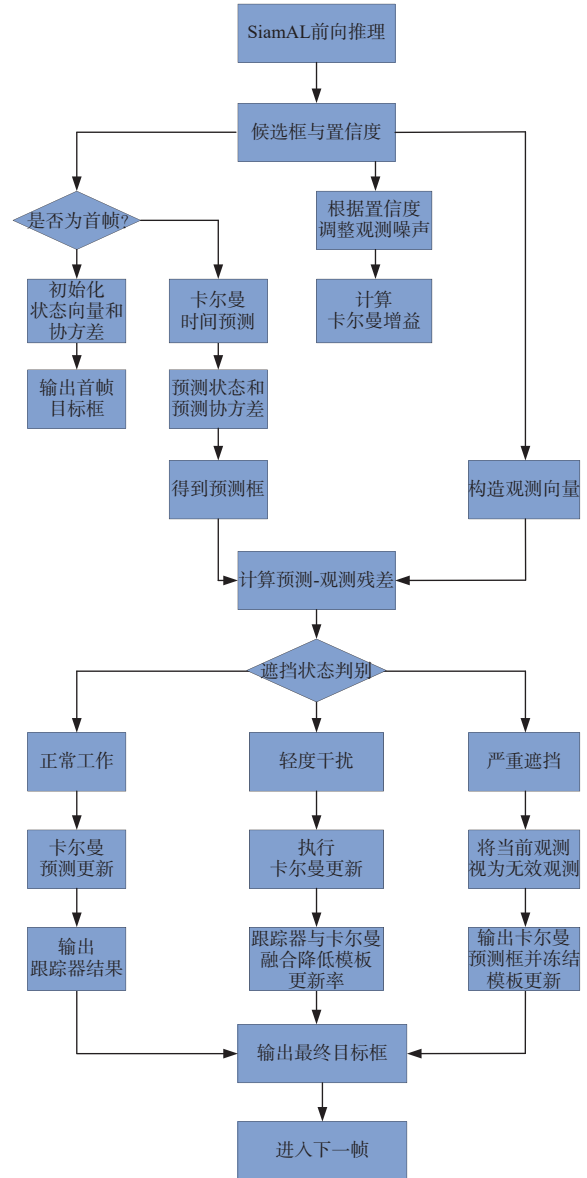


图8 融合卡尔曼滤波的遮挡状态更新流程

Fig. 8 Flowchart of occlusion-aware tracking with Kalman filtering

2.4.4.2 时间预测

在相邻帧间隔较短情况下,目标在局部时间窗口内可近似为匀速运动。相邻帧时间间隔为 Δt ,通常设置为1,状态转移矩阵定义为:

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & \Delta t & 0 & 0 & 0 \\ 0 & 1 & 0 & \Delta t & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}. \quad (10)$$

根据上一帧状态估计 $\mathbf{x}_{k-1}$ 和协方差矩阵 $\mathbf{P}_{k-1}$,

第 k 帧的状态预测和协方差预测分别为:

$$\hat{x}_{k|k-1} = A \cdot \hat{x}_{k-1|k-1}, \quad (11)$$

$$P_{k|k-1} = A \cdot P_{k-1} \cdot A^T + Q, \quad (12)$$

其中, Q 为过程噪声协方差矩阵, 用于描述运动模型的不确定性。

2.4.4.3 观测构建与预测-观测残差计算

卡尔曼滤波器完成时间预测后, 可由预测状态 $\hat{x}_{k|k-1}$ 得到当前目标的预测中心位置 (c_x^{KF}, c_y^{KF}) 。同时, SiamAL 跟踪器输出当前帧观测中心位置 (c_x^{trk}, c_y^{trk}) 。为了衡量当前观测与运动预测之间的一致性, 计算预测-观测欧氏残差:

$$d_k = \sqrt{(c_x^{trk} - c_x^{KF})^2 + (c_y^{trk} - c_y^{KF})^2}. \quad (13)$$

考虑到无人机航拍场景中目标尺度差异较大, 进一步定义归一化残差:

$$r_k = \frac{d_k}{\sqrt{w_k h_k}}. \quad (14)$$

归一化残差能够减弱目标尺度变化对阈值判定的影响, 使遮挡状态判别更加稳定。

2.4.4.4 置信度自适应观测噪声

SiamAL 跟踪器输出的置信度 c_k 反映了当前帧观测结果的可靠性。当目标发生遮挡、背景干扰或相似目标干扰时, 跟踪器置信度通常下降, 此时直接使用观测结果更新卡尔曼状态容易引入错误观测, 导致轨迹漂移。

为此, 本文将跟踪器置信度作为观测噪声协方差的调节因子, 定义自适应观测噪声矩阵如式(15)所示:

$$R_k = (1 - c_k) R_{\max} + c_k R_{\min}, \quad (15)$$

其中, $R_{\min}$ 表示高置信观测下的最小观测噪声协方差, $R_{\max}$ 表示低置信观测下的最大观测噪声协方差。当 c_k 较高时, R_k 接近 $R_{\min}$, 表示当前观测较为可靠; 当 c_k 较低时, R_k 接近 $R_{\max}$, 表示当前观测不可靠, 应降低其对状态更新的影响。卡尔曼增益计算公式为:

$$K_k = P_{k|k-1} H^T (H P_{k|k-1} H^T + R_k)^{-1}. \quad (16)$$

根据式(16)可知, 观测噪声增大时, 卡尔曼增益减小, 滤波器更加依赖运动预测的结果, 从而削弱可疑观测对目标状态估计的影响。

2.4.4.5 跟踪状态分级判别

为了同时考虑外观匹配可靠性和运动一致性, 本文将跟踪器置信度 c_k 与归一化预测-观测残差 r_k 联合构造遮挡评分:

$$S_k = \lambda(1 - c_k) + (1 - \lambda) \min(r_k, 1), \quad (17)$$

其中: 权重系数为 λ , 用于平衡置信度和残差对状态判别的影响; $(1 - c_k)$ 表示外观观测的不可信程度; r_k 表示观测结果与运动预测结果之间的一致程度。根据遮挡评分 S_k , 将状态划分为3类:

$$\text{状态} = \begin{cases} \text{正常工作} & S_k < T_1 \\ \text{轻度干扰} & T_1 < S_k < T_2 \\ \text{严重遮挡} & S_k \geq T_2 \text{ 且连续 } N_{\text{ooc}} \text{ 帧} \end{cases}, \quad (18)$$

其中: T_1 和 T_2 分别为正常跟踪、轻度干扰和严重遮挡之间的状态划分阈值, N_{ooc} 表示严重遮挡状态需要连续满足的帧数。通过设置连续帧判定条件, 可以避免由单帧模糊、快速运动或短时光照变化引起的误判。

2.4.4.6 不同状态下的更新与输出策略

根据跟踪状态的不同, 本文采用不同的状态更新和输出策略。当目标处于正常跟踪状态时, 当前观测可靠, 执行完整的卡尔曼预测-更新过程:

$$x_k = \hat{x}_{k|k-1} + K_k(z_k - H\hat{x}_{k|k-1}), \quad (19)$$

$$P_k = (I - K_k H) P_{k|k-1}. \quad (20)$$

此时最终输出主要采用 SiamAL 跟踪器的结果:

$$b_k = b_k^{trk}. \quad (21)$$

当目标处于轻度干扰状态时, 观测仍具有一定参考价值, 但可靠性下降。此时通过增大观测噪声降低观测权重, 同时采用跟踪器结果与卡尔曼预测结果的加权融合输出:

$$b_k = \alpha_k b_k^{KF} + (1 - \alpha_k) b_k^{trk}, \quad (22)$$

其中, α_k 为运动先验权重, 可根据遮挡评分确定:

$$\alpha_k = \frac{S_k - T_1}{T_2 - T_1}. \quad (23)$$

当 S_k 越接近 T_2 时, 说明当前状态越接近严重遮挡, 运动先验权重越大。该策略可以避免由单帧低置信度造成的硬切换, 使跟踪输出更加平滑。当目标处于严重遮挡状态时, 当前观测被视为无效观测, 跳过量测更新, 仅保留时间预测结果:

$$x_k = \hat{x}_{k|k-1}, \quad (24)$$

$$P_k = P_{k|k-1}. \quad (25)$$

此时最终输出为卡尔曼预测框:

$$b_k = b_k^{KF}. \quad (26)$$

同时冻结模板更新, 避免遮挡区域或背景区域被错误引入模板。当遮挡解除后, 若连续若干帧满足较高置信度和较小预测-观测残差条件,

则算法由严重遮挡状态逐步恢复至轻度干扰状态和正常跟踪状态。

2.5 融合目标检测算法的变体跟踪算法 SiamAL-Det

在无人机航拍场景中,目标经常受到长时遮挡、飞出视野等因素影响,仅依赖单一跟踪器往往难以保持稳定和精确的跟踪效果。许多研究表明,当目标发生严重遮挡或跟踪漂移时,引入轻量化的检测模块,可显著提升重新捕获能力和整体鲁棒性。基于这一思路,可在实际航拍应用中,将目标检测与跟踪过程闭环协同:在常规状态下由高效跟踪算法持续估计目标位置,一旦判定存在跟丢或严重偏移(即连续 30 帧处于严重遮挡状态),则触发检测分支在更大范围内对目标进行重新搜索和校正,从而缓解遮挡、快速运动及出视野造成的跟踪失败。在此基础上,本文设计了一种融合目标检测与孪生跟踪思想的算法框架,并以此命名为 SiamAL-Det:在正常帧利用孪生网络完成快速跟踪,在异常帧切换到重检测模式,实现对目标轨迹的稳健维持与误差修正,兼顾精度与实时性,适应复杂无人机航拍环境。算法流程如图 9 所示。

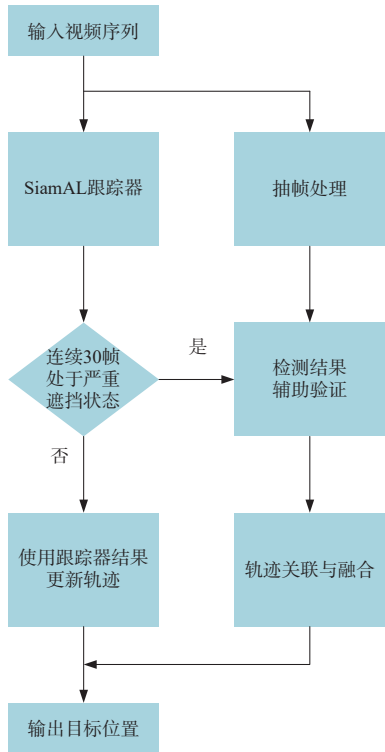


图 9 融合检测算法的变体跟踪算法

Fig. 9 Variant tracking algorithm integrated with detection

检测分支采用 YOLOv5s 作为重检测器。YOLOv5s 属于单阶段目标检测网络,能够在一次前向传播中同时完成目标类别预测和边界框回归,具有较好的检测速度和定位能力,适合无人机航拍场景中小目标、复杂背景和实时处理需求。与两阶段检测器相比,YOLOv5s 不需要显式生成区域提议,推理流程更加简洁,更适合作为异常帧下的轻量重检测模块。

检测器的网络结构选用主干网络(Backbone)、颈部网络(Neck)和检测头(Detection Head)三部分组成。其中,Backbone 用于提取输入图像的多尺度特征;Neck 采用特征金字塔和路径聚合结构进行多尺度特征融合,以增强小目标表征能力;Detection Head 在不同尺度特征图上进行目标类别、置信度和边界框位置预测,从而适应无人机航拍图像中目标尺度变化较大的特点。检测器的输入尺寸为 640×640 ,检测类别根据检测器训练数据集的标注类别确定,主要包括行人、车辆、骑行目标等无人机航拍场景中的常见目标类别。对于不在检测类别集合中的目标,检测分支采用与类别无关的候选框筛选方式,结合检测置信度、空间位置一致性和历史外观相似度选择重检测结果。检测分支采用在航拍检测数据集上预训练的 YOLOv5s 权重,UAV123 数据集仅用于跟踪性能测试,不参与检测器训练。

3 实验结果与分析

3.1 实验环境

SiamAL 算法通过 PyTorch1.7 框架实现,并在单个 RTX 4090 GPU 上进行训练和评估,编译环境为 Python3.7。采用精确率和成功率作为评价指标。精确率代表目标跟踪过程中算法预测的目标中心位置与实际标注的目标中心位置在视频图像上的欧氏距离,计算公式如式(27)所示;成功率代表目标跟踪过程中算法预测的目标框与实际标注的目标框在视频图像上的交并比,计算公式如式(28)所示。

$$e = \sqrt{(x_t - x_0)^2 + (y_t - y_0)^2}, \quad (27)$$

式中: (x_t, y_t) 表示预测目标的中心坐标, (x_0, y_0) 表示实际标注的中心坐标。通常设定阈值 $t=20$ 作为精确率的评价标准,即中心位置误差为 20 个

像素点作为评价依据。

$$os = \frac{A_t \cap A_{gt}}{A_t \cup A_{gt}}, \quad (28)$$

式中: A_t 表示跟踪器预测的区域, A_{gt} 是目标的真实区域。区域重叠率规定的阈值通常情况设为0.5。

3.2 UAV123数据集测试实验

本文对SiamAL在UAV123数据集上进行测试。为了测试提出的SiamAL算法的有效性,选用近年来具有代表性的几种跟踪算法作为比较对象,包括:SiamRPN++^[3]、SiamDMU^[8]、ECM^[29]、面向无人机航拍场景的TDSiam^[30]、带有干扰感知和重搜索机制的DaSiamRPN^[31]以及具有代表性的长时鲁棒跟踪算法DiMP-50^[32]。测试结果如图10所示,图10(a)和图10(b)分别为不同算法在UAV123数据集上的精确率和成功率变化曲线。

SiamAL的成功率为65.3%,精确率达到

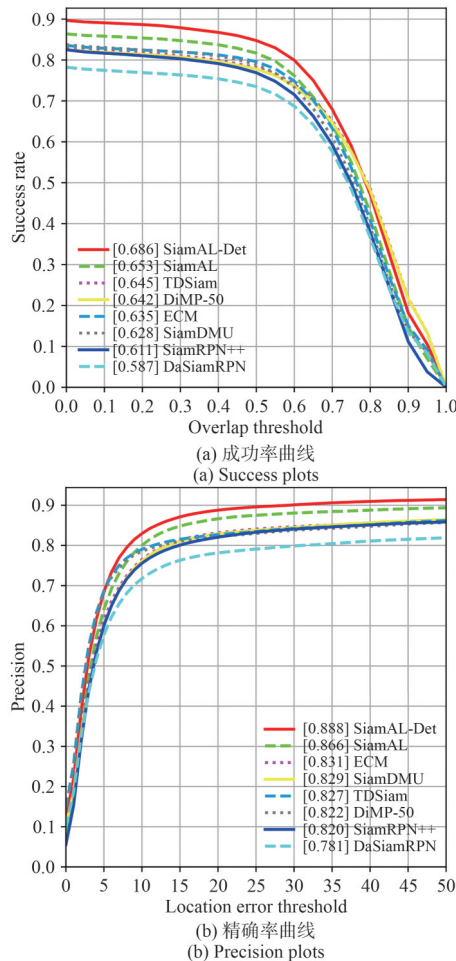


图10 不同算法在UAV123数据集上的测试结果
Fig. 10 Test results of different algorithms on UAV123

86.6%,在SiamAL-Det除外的跟踪算法中表现最好,相比排名第二的ECM算法,其成功率和精确率分别提升了1.8%和3.5%,验证了改进的SiamAL算法在UAV123数据集上的有效性。融合目标检测算法的变体SiamAL-Det,成功率达到68.6%,精确率达到88.8%,相比原SiamAL跟踪算法,成功率和精确率分别提升了3.3%和2.2%,在所有算法中取得了最好的跟踪效果。

3.3 卡尔曼滤波与遮挡判别机制分析

为了进一步验证卡尔曼滤波参与目标跟踪的效果,并验证置信度与预测-观测残差联合判别机制的有效性,本文选取UAV123数据集中的典型序列进行分析。实验记录每一帧的跟踪器置信度,卡尔曼预测中心、跟踪器观测中心、预测-观测残差、遮挡评分、状态判别结果以及最终输出方式,从定量和定性两个方面验证本文遮挡判别与运动先验融合策略的有效性。

3.3.1 遮挡序列轨迹可视化分析

为了进一步定性分析所提方案的有效性,选取UAV123航拍数据集中存在遮挡的car2视频序列的前600帧数据对SiamAL的跟踪轨迹进行可视化定性分析,可视化结果如图11所示。其中蓝色轨迹表示实际标注的目标运动轨迹,红色代表融合了跟踪算法预测的目标跟踪轨迹。绿色圆圈代表实际标注真实值开始位置,绿色方块代

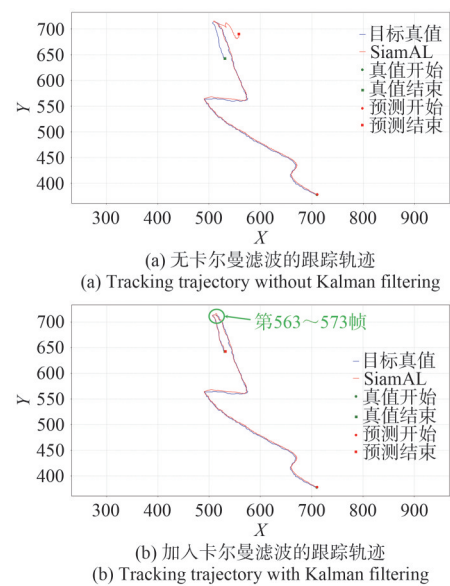


图11 遮挡序列轨迹可视化分析

Fig. 11 Trajectory visualization analysis under occlusion sequences

表实际标注真实值结束位置;红色圆圈代表预测值开始位置,红色方块代表预测值结束位置。

通过观察跟踪轨迹可以发现,在第 563~573 帧出现了严重遮挡情况。图 11(a)中红色轨迹代表未加入卡尔曼滤波跟踪算法的跟踪结果,可以发现在出现遮挡情况时产生了明显的漂移。图 11(b)中红色轨迹代表加入卡尔曼滤波的 SiamAL 跟踪算法的跟踪结果,此时目标严重遮挡且连续多帧满足此条件,当前帧可以视作“无效观测帧”,此时只保留卡尔曼预测并冻结模板更新。在严重遮挡情况结束后,目标置信度提升,可以看出在目标未发生遮挡时,各算法的预测轨迹均能够较好地跟随真实轨迹。当目标进入遮挡区间后,未引入卡尔曼运动先验的方法出

现明显漂移,表明 SiamAL 跟踪算法有效避免了目标遮挡场景下的跟踪漂移,在遮挡场景下发挥了良好的跟踪效果。

3.3.2 关键帧卡尔曼更新过程数据分析

表 2 给出了典型遮挡序列中若干关键帧的卡尔曼滤波过程数据。可以看出,在正常跟踪帧中,置信度较高且预测-观测残差较小,算法输出跟踪器结果;在遮挡发生后,进入轻度遮挡状态,融合结果作为输出;进一步地,置信度下降且残差增大,遮挡评分超过阈值,算法进入严重遮挡状态,当前观测被视为无效观测,仅输出卡尔曼预测结果;在遮挡解除后,残差逐渐减小,系统重新进入融合更新或正常跟踪状态。该结果从数值层面证明了卡尔曼滤波参与了目标状态估计过程。

表 2 关键帧卡尔曼更新过程数据
Tab. 2 Data of the Kalman update process at key frames

帧号	置信度	预测中心	SiamAL 输出位置	归一化残差	遮挡评分	遮挡状态
555	0.620	(517.95, 707.52)	(517.47, 707.32)	0.017	0.199	正常
564	0	(514.29, 714.19)	(511.84, 714.72)	0.098	0.549	轻度
567	0	(514.37, 715.44)	(533.47, 704.18)	0.870	0.935	严重
573	0	(511.80, 717.47)	(532.49, 709.74)	0.937	0.969	严重
580	0.751	(510.80, 707.60)	(514.46, 711.62)	0.231	0.240	正常
590	0.823	(508.23, 641.37)	(509.60, 642.31)	0.071	0.124	正常

3.3.3 卡尔曼融合策略消融实验

为进一步验证自适应卡尔曼融合策略的有效性,比较了无卡尔曼滤波 SiamDMU+H-ResNet-50、固定观测噪声的卡尔曼滤波 SiamAL-KF、依靠置信度调节滤波的 SiamAL-KF-C、依靠残差调节滤波的 SiamAL-KF-R 和置信度+残差联合调节的 SiamAL-KF-C+R 五种策略。为保证消融实验的公平性,5 组实验均采用相同的主干网络、跟踪框架、训练参数和测试序列,仅改变卡尔曼融合策略。数据集选择包含“严重遮挡”属性的视频序列子集,实验结果见表 3。

消融实验结果表明,相比无卡尔曼滤波策略,加入运动先验后遮挡场景下的成功率和精确率均有所提升;相比固定权重卡尔曼滤波策略,本文采用的自适应卡尔曼融合方法在“严重遮挡”属性的视频序列子集的成功率和精确率均取得最好的效果,相比固定权重策略,在成功率和精确率分别提升 2.7% 和 4%。实验结果表明,置信

表 3 不同卡尔曼融合策略在遮挡场景下的消融实验

Tab. 3 Ablation study of different Kalman fusion strategies under occlusion scenarios

方法	策略	成功率	精确率
SiamDMU+H-ResNet-50	仅跟踪器输出	0.471	0.694
SiamAL-KF	固定权重滤波	0.482	0.729
SiamAL-KF-C	依靠置信度调节	0.482	0.740
SiamAL-KF-R	依靠残差调节	0.492	0.737
SiamAL-KF-C+R	置信度+残差调节	0.509	0.769

度+残差调节的自适应卡尔曼融合策略在严重遮挡属性序列中可以有效减弱低可靠观测对状态更新的干扰,从而降低轨迹漂移风险并提升跟踪鲁棒性。

3.4 SiamAL-Det 重检测机制分析

3.4.1 典型序列可视化分析

选取 UAV123 数据集的 person8_1 视频序

列进行分析,如图12所示,该视频序列包含了较长时间的目标飞出场景之外的情况,常规的单一跟踪器往往难以保持稳定和精确的跟踪效果。在第935~1 014帧,目标始终在取景范围之外。1 014帧之后,单一的跟踪器无法对目标重新进行跟踪,而融合目标检测算法的变体跟踪算法SiamAL-Det可以通过重检测模式在目标重回取景范围后进行校正,后半段的跟踪算法可以重新实现对目标的跟踪,进一步提升跟踪性能。

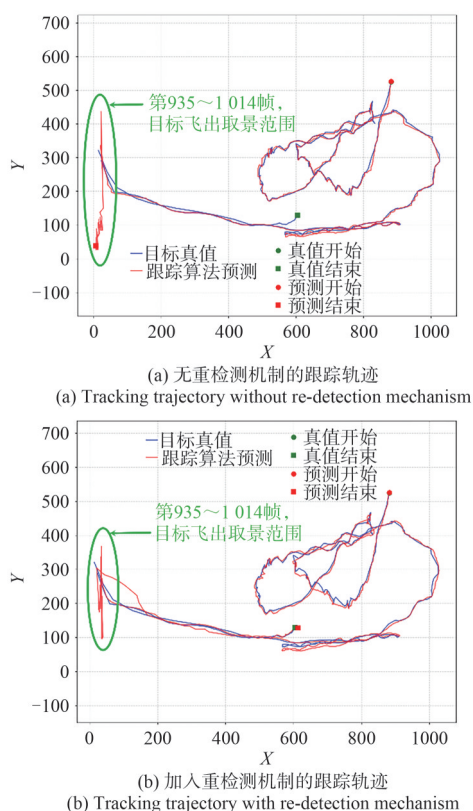


图12 出视野序列重检测轨迹可视化分析

Fig. 12 Trajectory visualization of re-detection under out-of-view sequence

3.4.2 重检测触发策略判别分析

当目标飞出视野或跟踪器漂移时,跟踪置信度 c_k 下降,同时预测-观测残差 d_k 增大,处于严重遮挡状态。当连续30帧处于严重遮挡状态,则判定为跟丢,启动检测器对目标进行重检测。表4展示了不同策略的重检测触发对比结果,其中恢复帧数从检测器触发开始计算。

无重检测机制的SiamAL在目标重新出现后恢复能力较弱,触发后平均恢复帧数超过60帧。

仅采用置信度触发重检测时,虽然恢复帧数缩短至22.4帧,但由于置信度容易受外观变化和背景干扰影响,存在一定误触发情况,导致成功率和精确率略低于SiamAL。仅采用预测-观测残差触发时,成功率和精确率分别提升至0.672和0.870,说明残差能够较好地反映跟踪结果与运动预测之间的不一致。进一步采用置信度与残差联合触发后,成功率和精确率分别达到0.686和0.888,触发后平均恢复帧数缩短至13.5帧,取得最优结果。实验结果表明,联合判别机制能够同时利用观测可靠性和运动一致性信息,更稳定地识别异常帧并触发重检测,从而提升目标重捕获能力。

表4 重检测触发策略消融实验

Tab. 4 Ablation study of the re-detection trigger strategy

方法	触发策略	成功率	精确率	平均恢复帧数
SiamAL	无重检测	0.653	0.866	无重检测, 目标重现后恢复帧数>60
SiamAL-Det-C	仅置信度触发	0.650	0.851	22.4
SiamAL-Det-R	仅残差触发	0.672	0.870	17.8
SiamAL-Det-C+R	置信度+残差联合触发	0.686	0.888	13.5

3.5 消融实验结果分析

为了验证改进主干网络、加入目标遮挡判别机制及融合目标检测算法对总体跟踪性能的影响,选取SiamDMU作为基准算法,在UAV123数据集上开展了多组消融实验,结果如表5所示。

对实验结果分析可得,相较于基准算法SiamDMU,改进主干网络、加入目标遮挡判别机制

表5 UAV123数据集上的消融实验结果

Tab. 5 Results of ablation experiments on UAV123

算法	成功率	精确率
SiamDMU	0.628	0.829
SiamDMU+CBAM	0.633	0.841
SiamDMU+h-swish	0.630	0.829
SiamDMU+H-ResNet-50	0.637	0.841
SiamDMU+目标遮挡判别机制	0.637	0.840
SiamAL	0.653	0.866
SiamDMU+Det	0.677	0.879
SiamAL-Det	0.686	0.888

及融合目标检测算法对总体跟踪性能均产生了积极贡献,其中融合目标检测算法的SiamDMU+Det提升明显,相较于基准算法成功率和精确率分别提升了4.9%和5%。在此基础上,同时加入以上3种改进措施的SiamAL-Det在UAV123数据集上取得了最佳的跟踪效果。

3.6 数据集中不同挑战的测试结果

为了进一步对不同算法进行评估及验证改进的主干网络、目标遮挡判别机制以及融合目标检测算法对跟踪算法的提升效果,本文给出了不同算法在UAV123数据集中12种挑战下的测试结果。12种常见的挑战分别是:全遮挡(Full Occlusion, FOC)、背景杂波(Background Clutters, BC)、相似目标(Similar Object, SOB)、快速运动(Fast Motion, FM)、光照变化(Illumination Variation, IV)、相机运动(Camera Motion, CM)、部分遮挡(Partial Occlusion, POC)、纵横比变化(Aspect Ratio Change, ARC)、观察点改

变(Viewpoint Change, VC)、尺度变化(Scale Variation, SV)、低分辨率(Low Resolution, LR)以及飞出取景范围(Out-of-View, OV)。不同算法在以上挑战下的成功率和精确率结果分别如表6和表7所示。可以看出,SiamAL-Det在多个挑战下均保持了良好的跟踪性能。在成功率方面,SiamAL-Det在尺度变化、相似目标、部分遮挡、低分辨率、光照变化、全遮挡、快速运动、相机运动、背景杂波、纵横比变化等多个场景均表现最佳。在精确率方面,SiamAL-Det在尺度变化、部分遮挡、低分辨率、光照变化、全遮挡、相机运动、背景杂波、纵横比变化、飞出取景范围等多个场景均表现最佳。基于UAV123数据集的12种挑战属性的对比结果表明,所提出的SiamAL-Det在多种典型UAV场景下取得了最高的成功率和精确率,验证了改进主干网络、目标遮挡判别机制与融合目标检测的目标跟踪在复杂干扰场景中的有效性与鲁棒性。

表 6 不同算法在 UAV123 数据集上的成功率结果

Tab. 6 Success rate of different algorithms on the UAV123 dataset

算法	VC	SV	SOB	POC	OV	LR	IV	FOC	FM	CM	BC	ARC	Overall
SiamRPN++	0.650	0.597	0.533	0.532	0.613	0.426	0.595	0.433	0.617	0.631	0.391	0.588	0.611
DaSiamRPN	0.608	0.566	0.521	0.494	0.514	0.437	0.468	0.436	0.516	0.614	0.403	0.551	0.587
DiMP-50	0.648	0.626	0.589	0.550	0.577	0.495	0.570	0.454	0.609	0.624	0.492	0.604	0.642
TDSiam	0.659	0.630	0.593	0.554	0.578	0.500	0.576	0.416	0.616	0.627	0.515	0.607	0.645
SiamDMU	0.619	0.615	0.515	0.535	0.577	0.570	0.585	0.466	0.596	0.614	0.608	0.556	0.628
ECM	0.622	0.621	0.564	0.549	0.585	0.579	0.558	0.477	0.594	0.615	0.624	0.563	0.635
SiamAL	0.664	0.645	0.629	0.568	0.626	0.562	0.590	0.509	0.603	0.652	0.526	0.606	0.653
SiamAL-Det	0.650	0.676	0.633	0.619	0.630	0.639	0.646	0.572	0.636	0.662	0.702	0.632	0.686

表 7 不同算法在 UAV123 数据集上的精确率结果

Tab. 7 Precision rate of different algorithms on the UAV123 dataset

算法	VC	SV	SOB	POC	OV	LR	IV	FOC	FM	CM	BC	ARC	Overall
SiamRPN++	0.838	0.788	0.728	0.749	0.817	0.674	0.811	0.699	0.841	0.840	0.609	0.795	0.820
DaSiamRPN	0.772	0.754	0.717	0.692	0.704	0.663	0.644	0.574	0.707	0.777	0.605	0.736	0.781
DiMP-50	0.830	0.801	0.774	0.733	0.746	0.713	0.750	0.645	0.797	0.804	0.638	0.772	0.822
TDSiam	0.834	0.807	0.780	0.738	0.748	0.722	0.758	0.648	0.807	0.808	0.640	0.776	0.827
SiamDMU	0.797	0.807	0.695	0.730	0.757	0.786	0.771	0.679	0.788	0.818	0.798	0.752	0.829
ECM	0.797	0.810	0.755	0.744	0.767	0.798	0.729	0.690	0.783	0.813	0.808	0.759	0.831
SiamAL	0.864	0.857	0.861	0.797	0.836	0.807	0.794	0.769	0.828	0.873	0.720	0.824	0.866
SiamAL-Det	0.823	0.874	0.853	0.832	0.851	0.872	0.838	0.809	0.828	0.863	0.902	0.839	0.888

对于某些场景,如观察点改变,融合了目标检测的跟踪算法反而表现更差,其原因是由于无

人机视角航拍过程中,检测器出现了误检,对算法的整体性能产生了负面影响。

3.7 跟踪序列对比

对测试结果进行定性实验可以更直接地对算法的有效性进行评价。本文选取 bike2、car15、person14_1 和 person19_2 序列对不同算法进行可视化分析,这4个序列对于无人机航拍场景下不同的目标进行了跟踪,包含尺度变化、全遮挡、部分遮挡、背景杂波和相机运动等常见挑战。为进一步展示各算法跟踪效果之间的差异,其定性分析结果如图13所示。

bike2 和 car15 序列中存在目标遮挡情况, SiamAL-Det 算法可以始终保持良好的跟踪效果,进一步说明 SiamAL 在目标遮挡下具有更加精确的跟踪结果。person14_1 序列在 200 帧前后出现了严重的目标遮挡情况, SiamAL 和 SiamDMU 均产生了较大的跟踪漂移。由于 SiamAL-Det 结合了目标检测算法,可以对目标进行重新跟踪,因此实现了最好的跟踪效果。person19_2 是无人机在移动过程中拍摄的视频,序列前后具有较大的背景变换。该场景下, SiamDMU 算法

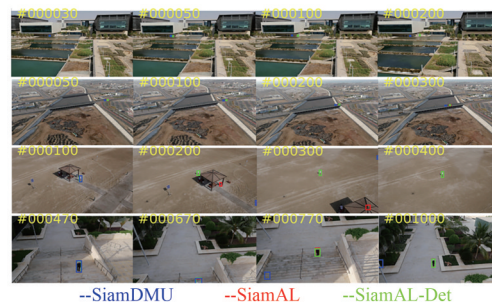


图13 算法定性分析对比(由上至下分别是 bike2、car15、person14_1 和 person19_2)。

Fig. 13 Visualized results of four sequences on UAV123 (from top to down: bike2, car15, person14_1 and person19_2).

产生了较大的跟踪漂移,效果较差。相比之下, SiamAL 和 SiamAL-Det 由于在主干网络中结合了 CBAM 注意力机制,能有效解决无人机航拍过程中经常出现的背景变换干扰。表8给出了这4个序列的定量跟踪结果,与图13展示的定性结果一致。

表8 UAV123中4个序列的跟踪结果

Tab. 8 Tracking results of four sequences on UAV123

序列名称	SiamDMU		SiamAL		SiamAL-Det	
	成功率	精确率	成功率	精确率	成功率	精确率
bike2	0.051	0.166	0.148	0.263	0.792	0.823
car15	0.055	0.244	0.621	0.963	0.607	0.956
person14_1	0.131	0.202	0.128	0.208	0.357	0.520
person19_2	0.581	0.727	0.812	0.951	0.813	0.949

4 结 论

为解决无人机航拍过程中的复杂干扰以及目标遮挡问题,提出了结合注意力机制和目标遮挡判别机制的目标跟踪算法 SiamAL, 以及其融合目标检测算法的变体 SiamAL-Det。在基准算法 SiamDMU 基础上,引入 CBAM 注意力模块与 h-swish 激活函数,显著增强了在复杂背景和干扰条件下的特征表征与鲁棒性。与此同时,利用基于卡尔曼滤波的运动先验对目标轨迹进行预测,并通过跟踪置信度与预测-观测残差联合判别遮挡状态,在严重遮挡时自适应提升运动先验权

重,从而有效保持目标轨迹的连续性。在此基础上,构建了融合目标检测与孪生跟踪的算法框架,提高了在长时序与复杂场景下的整体跟踪稳定性。在 UAV123 无人机航拍数据集上进行测试, SiamAL 以及 SiamAL-Det 算法的成功率和精确率达到了 65.3%、86.6% 和 68.6%、88.8%, 相比基准算法 SiamDMU 分别提升了 2.5%、3.7% 和 5.8%、5.9%,验证了方法在无人机航拍目标跟踪任务中的有效性与应用前景。未来工作中,将针对航拍跟踪过程中目标像素占比较小导致的特征表达不足问题,进一步优化小目标的感知与区分能力。

参 考 文 献:

- [1] 苗宗成,高世严,贺泽民,等. 基于孪生网络的目标跟踪算法[J]. 液晶与显示,2023,38(2):256-266.
MIAO Z C, GAO S Y, HE Z M, *et al.* Single-objective tracking algorithm based on Siamese networks [J]. *Chinese Journal of Liquid Crystals and Displays*, 2023, 38(2): 256-266. (in Chinese)
- [2] BERTINETTO L, VALMADRE J, HENRIQUES J F, *et al.* Fully-convolutional Siamese networks for object tracking [C]//Computer Vision-ECCV 2016 Workshops. Amsterdam: ECCV, 2016: 850-865.
- [3] LI B, WU W, WANG Q, *et al.* SiamRPN++: evolution of Siamese visual tracking with very deep networks [C]//*Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Long Beach: IEEE, 2019: 4277-4286.
- [4] LI B, YAN J J, WU W, *et al.* High performance visual tracking with Siamese region proposal network [C]//*Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City: IEEE, 2018: 8971-8980.
- [5] HUANG B, XU T F, SHEN Z Y, *et al.* SiamATL: online update of Siamese tracking network via attentional transfer learning [J]. *IEEE Transactions on Cybernetics*, 2022, 52(8): 7527-7540.
- [6] CHEN Z D, ZHONG B N, LI G R, *et al.* Siamese box adaptive network for visual tracking [C]//*Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Seattle: IEEE, 2020: 6667-6676.
- [7] MUELLER M, SMITH N, GHANEM B. A benchmark and simulator for UAV tracking [C]. 14th European Conference on Computer Vision. Amsterdam: ECCV, 2016: 445-461.
- [8] LIU J, WANG H, MA C, *et al.* SiamDMU: Siamese dual mask update network for visual object tracking [J]. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 2024, 8(2): 1656-1669.
- [9] MA J Y, REN J L, ZHANG J H, *et al.* Quantum imaging using spatially entangled photon pairs from a nonlinear metasurface [J]. *eLight*, 2025, 5(1): 2.
- [10] HAN G R, ERYILMAZ M, GONCHAROV A, *et al.* Deep learning-enhanced dual-mode multiplexed optical sensor for point-of-care diagnostics of cardiovascular diseases [J]. *Light: Science & Applications*, 2026, 15(1): 190.
- [11] 徐逸,甄佳宁,蒋侠朋,等. 无人机遥感与XGBoost的红树林物种分类[J]. 遥感学报,2021,25(3):737-752.
XU Y, ZHEN J N, JIANG X P, *et al.* Mangrove species classification with UAV-based remote sensing data and XGBoost [J]. *National Remote Sensing Bulletin*, 2021, 25(3): 737-752. (in Chinese)
- [12] 郭庆华,胡天宇,刘瑾,等. 轻小型无人机遥感及其行业应用进展[J]. 地理科学进展,2021,40(9):1550-1569.
GUO Q H, HU T Y, LIU J, *et al.* Advances in light weight unmanned aerial vehicle remote sensing and major industrial applications [J]. *Progress in Geography*, 2021, 40(9): 1550-1569. (in Chinese)
- [13] 吴志成,林秀贵,许家浩,等. 一种采用航拍图像识别高压输电线路防震锤缺失的检测算法[J]. 应用科技,2021, 48(5):35-41,53.
WU Z C, LIN X G, XU J H, *et al.* An algorithm for missing detection of anti-vibration hammer on high voltage transmission line based on aerial image [J]. *Computing magazine of the CCF*, 2021, 48(5): 35-41, 53. (in Chinese)
- [14] 张志豪,厉小润,陈淑涵. 基于改进YOLO11的无人机航拍图像小目标检测算法[J]. 液晶与显示,2025,40(6): 915-930.
ZHANG Z H, LI X R, CHEN S H. Small object detection algorithm in UAV aerial images based on improved YOLO11 [J]. *Chinese Journal of Liquid Crystals and Displays*, 2025, 40(6): 915-930. (in Chinese)
- [15] SOLEIMANITALEB Z, KEYVANRAD M A, JAFARI A. Object tracking methods: a review [C]. 2019 9th International Conference on Computer and Knowledge Engineering (ICCKE). Mashhad: IEEE, 2019: 282-288.
- [16] CHENG F, LIANG Z B, PENG G L, *et al.* An anti-UAV long-term tracking method with hybrid attention mechanism and hierarchical discriminator [J]. *Sensors*, 2022, 22(10): 3701.
- [17] HUANG B, CHEN J J, XU T F, *et al.* SiamSTA: spatio-temporal attention based Siamese tracker for tracking UAVs [C]//*Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision Workshops*. Montreal: IEEE, 2021: 1204-1212.
- [18] 李大湘,辛嘉妮,刘颖. 位置敏感Transformer航拍图像目标检测模型[J]. 光学精密工程,2024,32(5):727-739.

- LI D X, XIN J N, LIU Y. Position-sensitive Transformer aerial image object detection model [J]. *Optics and Precision Engineering*, 2024, 32(5): 727-739.
- [19] YUAN X Y, XU T F, LIU X C, *et al.* Multi-step temporal modeling for UAV tracking [J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2024, 34(8): 7216-7230.
- [20] LIU X C, XU T F, WANG Y, *et al.* BACTrack: building appearance collection for aerial tracking [J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2024, 34(6): 5002-5017.
- [21] PANG H B, SU J, MA R Q, *et al.* Multiple templates transformer for visual object tracking [J]. *Knowledge-Based Systems*, 2023, 280: 111025.
- [22] HOWARD A, SANDLER M, CHU G, *et al.* Searching for MobileNetV3 [C]//*Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision*. Seoul: IEEE, 2019: 1314-1324.
- [23] ZHOU L J, ZHANG J L. Combined Kalman filter and multifeature fusion Siamese network for real-time visual tracking [J]. *Sensors*, 2019, 19(9): 2201.
- [24] FU C H, LU K H, ZHENG G Z, *et al.* Siamese object tracking for unmanned aerial vehicle: a review and comprehensive analysis [J/OL]. *arXiv*, 2022: 2205.04281.
- [25] 黄鹤,张科,陈永安,等. 一种无人机航拍目标的长期跟踪算法[J]. 哈尔滨工业大学学报,2022,54(5):104-116.
HUANG H, ZHANG K, CHEN Y A, *et al.* A long-term tracking algorithm for UAV aerial photography [J]. *Journal of Harbin Institute of Technology*, 2022, 54(5): 104-116. (in Chinese)
- [26] HE K M, ZHANG X Y, REN S Q, *et al.* Deep residual learning for image recognition [C]//*Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas: IEEE, 2016: 770-778.
- [27] WOO S, PARK J, LEE J Y, *et al.* CBAM: convolutional block attention module [C]//*Proceedings of the 15th European Conference on Computer Vision (ECCV)*. Munich: ECCV, 2018: 3-19.
- [28] 闫纲琦,梁宗林,宋延嵩,等. 基于注意力残差网络的快照式多光谱相机图像重构[J]. 中国光学(中英文),2024, 17(6):1316-1328.
YAN G Q, LIANG Z L, SONG Y S, *et al.* Image reconstruction of snapshot multispectral camera based on an attention residual network [J]. *Chinese Optics*, 2024, 17(6): 1316-1328. (in Chinese)
- [29] 王伯霄,宋延嵩,董小娜. 难点注意力感知红外小目标检测网络[J]. 中国光学(中英文),2024,17(3):538-547.
WANG B X, SONG Y S, DONG X N. Indistinguishable points attention-aware network for infrared small object detection [J]. *Chinese Optics*, 2024, 17(3): 538-547. (in Chinese)
- [30] SUN L F, YANG Z, ZHANG J J, *et al.* Visual object tracking for unmanned aerial vehicles based on the template-driven Siamese network [J]. *Remote Sensing*, 2022, 14(7): 1584.
- [31] ZHU Z, WANG Q, LI B, *et al.* Distractor-aware Siamese networks for visual object tracking [C]//*Proceedings of the 15th European Conference on Computer Vision (ECCV)*. Munich: ECCV, 2018: 103-119.
- [32] BHAT G, DANELLJAN M, VAN GOOL L, *et al.* Learning discriminative model prediction for tracking [C]//*Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision*. Seoul: IEEE, 2019: 6181-6190.

作者简介:



赵 萌,男,学士,工程师,2014年于哈尔滨工程大学获得学士学位,主要从事机器视觉方面的研究。E-mail: 635766103@qq.com



郑滨汐,男,硕士,助理研究员,2023年于哈尔滨工程大学获得硕士学位,主要从事深度学习目标跟踪算法的研究。E-mail: zhengbinxi@hrbeu.edu.cn