

文章编号: 1671-7449(2025)06-0696-10

J. Test Measur. Technol., 2025, 39(6): 696-705.

基于无人机视角的地面多军事目标跟踪算法

马泽兵, 党长营*, 曾志强, 李忠华, 李建素, 陈雪丽, 常文彪, 王文媛

(中北大学 机械工程学院, 山西 太原 030051)

摘要: 针对当前多目标跟踪算法在军事无人机视角的战场感知中存在轨迹中断和ID跳转问题, 提出一种基于无人机视角的地面多军事目标跟踪算法 VA-ByteTrack。首先, 将EfficientFormerV2网络作为检测器YOLOX的主干网络, 并在输出端添加卷积块注意力模块(Convolutional Block Attention Module, CBAM), 以提高检测器对小尺寸目标的特征提取能力, 解决了小目标特征模糊导致的轨迹中断问题。其次, 引入运动匹配代价并比(Intersection Over, Union, IOU)和检测置信度 S , 构建了基于sigmoid函数的自适应卡尔曼滤波, 以平衡高分检测框和低分检测框的不同协方差需求, 解决了目标密集且相互遮挡所导致的频繁ID跳转问题。实验结果表明, 相比SORT、DeepSORT、ByteTrack等主流算法, 所提算法跟踪准确度分别提升1.7%、1.4%和1.0个百分点, 跟踪精度分别提升0.7%、0.5%和1.9个百分点, 轨迹中断问题分别降低49.4%、7.7%和7.2%, ID跳转问题分别降低46.3%、15.9%和12.1%, 满足战场动态感知的实际需求。

关键词: 多军事目标跟踪; 无人机视角; 自适应卡尔曼滤波; 目标检测

中图分类号: TB96

文献标识码: A

doi: 10.62756/csjs.1671-7449.2025072

引用格式: 马泽兵, 党长营, 曾志强, 等. 基于无人机视角的地面多军事目标跟踪算法[J]. 测试技术学报, 2025, 39(6): 696-705.

MA Zebing, DANG Changying, ZENG Zhiqiang, et al. Ground multi-military target tracking algorithm based on perspective of UAV[J]. Journal of Test and Measurement Technology, 2025, 39(6): 696-705.

Ground Multi-Military Target Tracking Algorithm Based on Perspective of UAV

MA Zebing, DANG Changying*, ZENG Zhiqiang, LI Zhonghua, LI Jiansu, CHEN Xueli,
CHANG Wenbiao, WANG Wenyuan

(School of Mechanical Engineering, North University of China, Taiyuan 030051, China)

Abstract: Aiming at the problems of track interruption and ID jump in the battlefield perception from the perspective of military UAVs, this paper proposes a ground multi-target tracking algorithm VA-ByteTrack based on the perspective of UAVs. First, EfficientFormerV2 network was used as the backbone network of detector YOLOX, and CBAM attention mechanism was added to the output to improve the feature extraction capability of detector for small-size targets, and to solve the problem of track interruption caused by fuzzy features of small targets. Secondly, by introducing motion matching cost IOU and detection confidence S , an adaptive Kalman filter AKF based on sigmoid function is constructed to balance the different covariance requirements of high-score detection boxes and low-score detection boxes, and to

收稿日期: 2024-12-04

基金项目: 山西省基础研究计划面上项目(202103021224199, 20210302123047)

作者简介: 马泽兵(2002—), 男, 硕士生, 主要从事机器视觉研究。E-mail: 1420738901@qq.com。

* 通信作者: 党长营(1983—), 男, 副教授, 博士, 主要从事人工智能、精密测量、机器视觉研究。E-mail: dcy@nuc.edu.cn。

solve the problem of frequent ID jump caused by dense targets and mutual occlusion. Experimental results show that compared with mainstream methods such as SORT, DeepSORT and ByteTrack, the tracking accuracy of the proposed method is increased by 1.7, 1.4 and 1.0 percent point, the tracking accuracy is increased by 0.7, 0.5 and 1.9 percent point, and the track interruption problem is reduced by 49.4%, 7.7% and 7.2%, respectively. The ID jump problem is reduced by 46.3%, 15.9% and 12.1% respectively, which meets the actual demand of battlefield dynamic perception.

Key words: multi-military target tracking; perspective of UAV; adaptive Kalman filtering; object detection

0 引言

随着计算机视觉技术的迅速发展,基于无人机视角的多目标检测与跟踪技术在战场动态感知^[1-3]、智慧城市管理^[4-5]、智慧交通调控^[6]等重要领域展现出广泛的应用价值。然而,传统的多目标跟踪技术在基于无人机视角的战场动态感知中,往往受到目标过小、特征模糊、目标密集且相互遮挡等因素的干扰,出现轨迹中断、频繁身份ID跳转和跟踪精度严重下滑等问题^[7-8]。因此,研究一种基于无人机视角的在线实时地面多军事目标跟踪算法迫在眉睫。

目前,多目标跟踪(Multiobject Tracking, MOT)可分为基于检测的跟踪范式(Defect-Before-Trace, DBT)和无检测跟踪范式(Detection-Free Tracking, DFT)^[9]。近年来,随着检测器技术的日益成熟,相比于复杂繁琐的DFT, DBT已被证明是一种简单、直接、有效的方法^[10]。2016年, Bewley等^[11]首次提出在线实时多目标跟踪算法(Simple Online and Realtime Tracking, SORT),但该算法对遮挡目标的跟踪效果较差。2017年, Wojke等^[12]提出DeepSORT算法,引入REID网络提取目标外观特征以改善对遮挡目标的跟踪效果。之后,罗茜等^[13]结合YOLOv5和DeepSORT,将含有时空注意力模块的残差网络作为跟踪模块中的特征提取网络,得到了具备较强抗干扰能力的无人机视角跟踪算法。高质桐^[14]将提取的车辆边缘和颜色信息作为主要特征,以确保对特征模糊目标的有效跟踪。尽管这些基于Deep SORT的算法及优化算法都取得了不错的跟踪效果,但是特征提取网络的引入会导致跟踪速度的大幅下降和计算量的增加,因此,不适于无人机边缘设备的部署。2022年, Zhang等^[15]考虑到通过分数阈值过滤低分检测框的方式会导致大量轨迹的中断,提出新的多目标跟踪算法ByteTrack,通过增加低分检测框关联阶段,在不影响跟踪速度的前提下,提高了算法的整体性能。朱奇光等^[16]在ByteTrack

算法的基础上,将卡尔曼滤波的状态向量融入加速度信息,以适应车辆速度快速变化的情况,提升了无人机视角下的车辆跟踪效果。然而,传统卡尔曼滤波的固定噪声设置难以平衡低噪声干扰的高分检测框和高噪声干扰的低分检测框的不同协方差需求,再加之航拍视频背景过于复杂,使得该算法在无人机视角应用中仍出现频繁ID切换与跟踪精度严重下滑问题^[17]。

此外,由于检测器性能直接决定了跟踪算法的实时性、精度和鲁棒性,学者们通常会对检测器进行改进。例如, Zhang等^[18]采用轻量化检测器,解决了模型过大不利于边缘部署问题; Yuan等^[19]通过增加检测器的检测层,解决了小特征不明显问题,提高了跟踪精度。

综上所述, SORT算法速度最优,但在遮挡情况下的跟踪性能不足; DeepSORT引入特征网络改善对遮挡目标的跟踪效果,但计算量大,难以在边缘设备部署; ByteTrack算法可以同时兼顾跟踪速度和跟踪效果,但其卡尔曼滤波的固定噪声设置却导致了跟踪精度下滑。因此,本文提出一种基于无人机视角的地面多军事目标跟踪算法(VA-ByteTrack)。在检测器方面,引入ViT网络EfficientFormerV2和卷积块注意力模块(Convolutional Block Attention Module, CBAM),提高模型的特征提取能力,减少轨迹中断问题;在跟踪方面,引入运动匹配代价交并比(Intersection Over Union, IOU)和检测置信度 S ,构造基于sigmoid函数的自适应卡尔曼滤波(Adaptive Kalman Filter, AKF),提高算法跟踪精度以及降低ID切换次数。

1 算法原理

本文所提算法VA-ByteTrack是一种基于检测的跟踪策略,由优化的检测器YOLOX、AKF、匈牙利算法3个核心部分组成(见图1)。检测器实现地面多军事目标的高效精准检测; AKF通过预测和更

新来估计目标的运动状态与位置信息；匈牙利算法完成预测跟踪轨迹与检测器输出之间的数据关联，进而创建目标的跟踪。

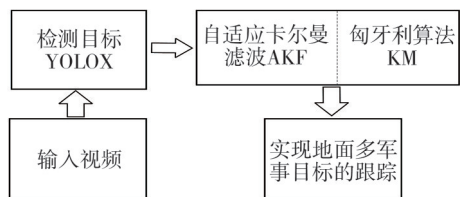


图1 算法原理图

Fig.1 Algorithm schematic diagram

1.1 优化的检测器

检测器性能对算法跟踪效果和跟踪速度效果有着直接的影响。检测器性能不足往往会引发跟踪过程中的漏检和轨迹中断问题。因此，本文首

先对 ByteTrack 算法初始检测器 YOLOX^[20] 进行优化，以提升无人机视角下的目标检测效果，优化后的模型结构如图 2 所示，优化措施如下。

1) 融合 ViT 网络 EfficientFormerV2

将 YOLOX 的骨干网络 CSP Darknet 替换为 ViT^[21] 网络 EfficientFormerV2^[22]。该网络采用四阶段分层设计，其获得特征尺寸为输入分辨率的 1/4, 1/8, 1/16, 1/32。前两阶段为捕捉高分辨率下的局部信息，仅使用前馈网络 (Feedforward Neural Network, FFN)；后两个阶段则使用 FFN 和多头自注意力机制 (Multi-Head Self-Attention, MHSA)。此外，该网络通过细粒度联合搜索方法同时优化模型大小和速度，灵活配置网络中的网络深度、宽度以及 FFN 扩展率，不仅拥有强大的特征提取能力，而且具备最优参数量和延时。

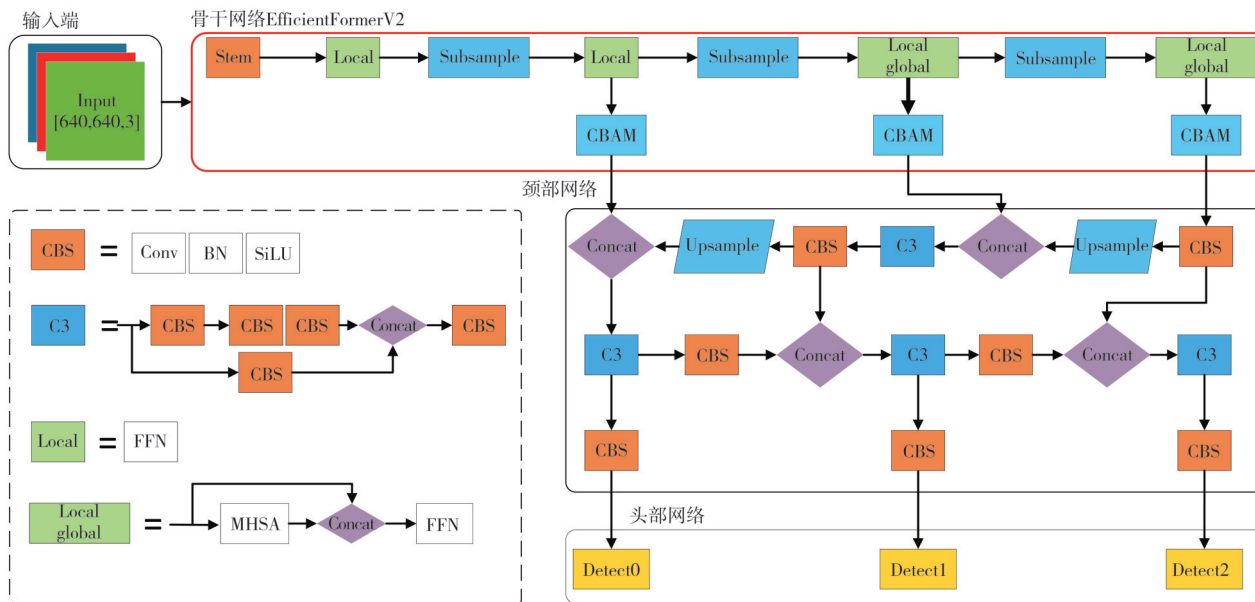


图2 优化的YOLO X模型结构

Fig.2 Improved YOLO X model structure

将 EfficientFormerV2 作为检测器 YOLOX 的骨干网络，可利用其优秀的自注意力机制，有效捕捉图像中的全局特征，显著增强网络对小尺寸目标的提取能力；同时，还具备较低的参数量和快速的推理速度，完美契合无人机边缘设备的部署，确保了在实际应用中的高性能和低延迟表现。

2) 添加 CBAM 注意力机制

在骨干网络的输出端添加混合注意力机制 CBAM，它是一种用于前反馈卷积神经网络的简单有效的注意模块，由通道注意力模块 (Channel Attention Module, CAM) 和空间注意力模块 (Spatial Attention Module, SAM) 两部分组成^[23]，如

图 3 所示。

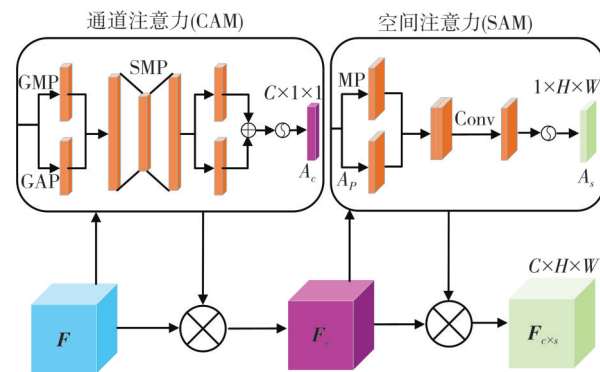


图3 CBAM结构

Fig.3 Structure of CBAM

CAM使网络能够自动选择对当前任务最有用的特征通道, SAM使网络能够聚焦于图像中最重要区域。在骨干网络输出端添加CBAM可使网络通过特征和空间两个维度推断关注图, 进而提高对目标关键特征的提取能力和降低航拍视频中复杂背景的干扰, 提高检测效果。

1.2 自适应卡尔曼滤波

ByteTrack算法通过引入低分检测框的关联, 可以恢复一部分因遮挡而跟踪失败的目标, 显著提升多目标跟踪的准确性和鲁棒性。但是这种关联会导致跟踪精度下降和ID跳转问题, 原因在于低分检测框通常涉及高水平的遮挡噪声, 而卡尔曼滤波的固定噪声设置通常只考虑高分检测框, 难以平衡高分和低分检测框的不同需求^[24]。

基于以上分析, 本文实现了一种基于sigmoid函数的低成本、高效AKF模型, 以满足高分和低分检测框的不同需求, 提高跟踪精度, 降低ID切换次数。首先, 将运动匹配代价IOU和检测置信度S分别引入线性卡尔曼滤波作为过程噪声尺度参数和观测噪声尺度参数, 之后通过sigmoid函数构造两个非线性适应因子 κ 和 ξ , 进而调整不同状态轨迹的过程和观测协方差矩阵。

首先, 设定卡尔曼滤波的状态向量 X 为

$$X = (x, y, r, h, v_x, v_y, v_r, v_h), \quad (1)$$

式中: x 和 y 为目标框的中心坐标; r 为目标框的长宽比值; h 为目标框的长度; 后4个值分别为对应参数的速度。令 $S = (x, y, r, h)$, $V = (v_x, v_y, v_r, v_h)$, 基于此, 状态向量 X 在时间 t 的预测可以表示为

$$\widetilde{X}_t = \begin{bmatrix} 1 & \Delta t \\ 0 & 1 \end{bmatrix}_{8 \times 8} \times \begin{bmatrix} S \\ V \end{bmatrix}_{t-1}. \quad (2)$$

整个卡尔曼滤波算法的预测过程为

$$\begin{cases} \widetilde{X}_t = X \widetilde{X}_{t-1}, \\ P_t = A P_{t-1} A^T + \kappa_t Q, \\ \kappa_t = \frac{1}{1 + e^{-40 * IOU + 32}}, \end{cases} \quad (3)$$

式中: \widetilde{X} 为目标估计状态; \widetilde{X}_t 为对 \widetilde{X} 的预测值; A 为目标状态转移矩阵; P_t 为目标信息的状态估计协方差矩阵; P_t 为 P_t 的预测值; Q 为过程噪声协方差矩阵; κ 为过程噪声自适应因子; IOU 为上一时刻的运动匹配代价, 即检测框与预测轨迹之间的交并比。公式中, 首先对 IOU 值进行线性变换, 然后将变换后的结果输入到sigmoid函数中,

以充分利用sigmoid函数优秀的非线性处理能力优化和调整过程噪声自适应因子 κ 。在增加过程噪声自适应因子后, 对单个目标预测过程的计算量由1 216 FLOPs增加至1 310 FLOPs, 并不会造成过大的计算负担。

κ 与 IOU 的变换关系如图4所示, κ 随着 IOU 值的增加逐渐减小。当 $0 < IOU < 0.7$ 时, κ 趋向于1, 此时可以为高偏差预测分配极大的过程协方差, 以抑制高偏差预测带来的干扰; 当 $0.8 < IOU < 1$ 时, 为低偏差分配较小的过程协方差, 以维持低偏差预测的稳定性, 使系统能够更准确地跟踪目标。

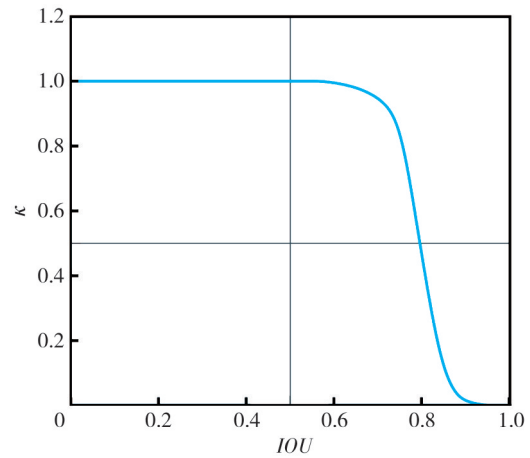


图4 过程噪声自适应因子 κ 变化趋势

Fig. 4 Trend of process noise adaptive factor κ

卡尔曼滤波更新过程为

$$\begin{cases} K_t = P_t H^T (H P_t H^T + \xi_t R)^{-1}, \\ \widetilde{X}_t = \widetilde{X}_t + K_t (Z_t - H \widetilde{X}_t), \\ P_t = (I - K_t H) P_t, \\ \xi_t = \frac{1}{1 + e^{-40 * s + 8}}, \end{cases} \quad (4)$$

式中: H 为观测矩阵; R 为观测噪声协方差矩阵; I 为单位矩阵; K 为卡尔曼滤波增益系数; ξ 为观测噪声自适应因子; s 为该时刻检测框的置信度。观测噪声自适应因子 ξ 的构造过程与上文相似, 通过线性变换和sigmoid函数调整 ξ 值。在增加观测噪声自适应因子后, 对单个目标的卡尔曼滤波更新过程计算量由3 298 FLOPs增长至3 392 FLOPs。

ξ 与 s 的变换关系如图5所示, ξ 随着 s 值的增加而降低。当 $0.3 < s < 1$ 时, ξ 趋向于0, 这意味着可以维持高置信度检测事件以较低的观测协方差; 当 $s < 0.3$ 时, ξ 随着 s 的减小快速增大, 且 $s < 0.2$ 时, ξ 趋向于1, 表明可以分配低分检测事件以

较大的观测协方差,以避免包含大量噪声或错误的低分检测事件所带来的干扰。

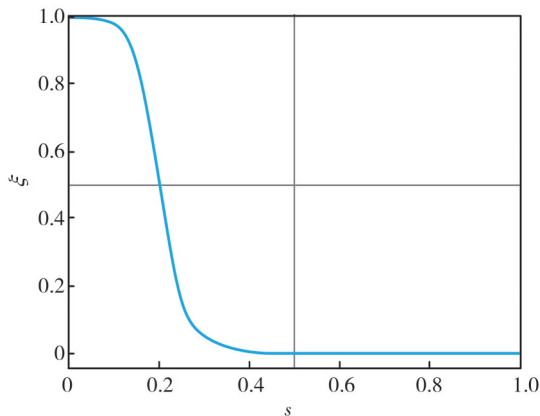


图5 观测噪声自适应因子 ξ 变化趋势

Fig. 5 Trend of process observation adaptive factor ξ

1.3 算法流程

本文算法流程如图6所示。

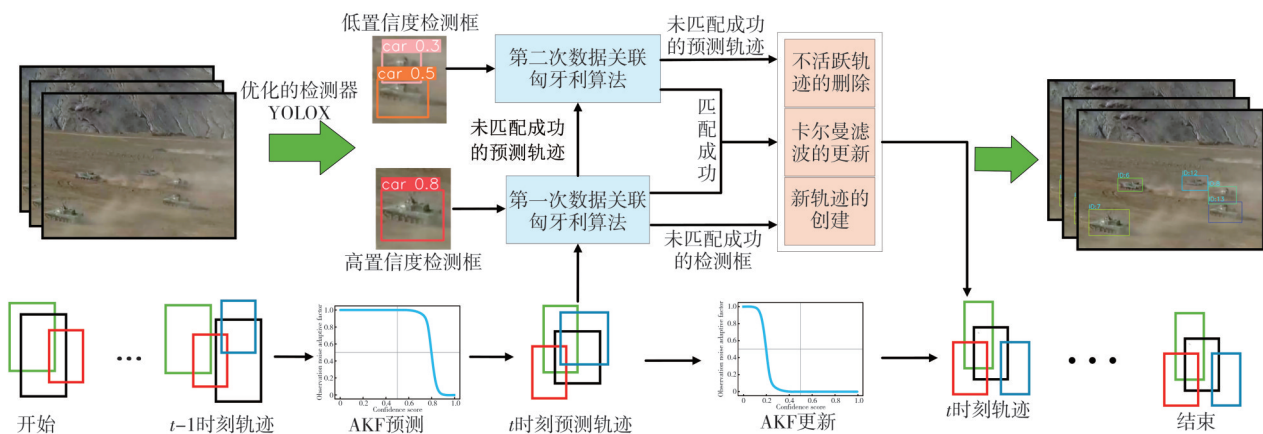


图6 算法流程图

Fig. 6 Algorithm flow chart

2 实验结果与分析

2.1 数据集与评价指标

本文采用大型公开的无人机航拍视频 VisDrone MOT^[25] 训练和测试本文算法 VA-ByteTrack。VisDrone MOT 数据集由天津大学机器学习和数据挖掘实验室采集,共包含训练视频56个、验证视频序列7个和测试视频序列33个,广泛涵盖了多种场景和不同高度下的航拍视频。同时,还在战场航拍视频资源上展开测试,以验证本文算法在战场环境下对地面多军事目标的跟踪效果。

实验采用多项评价指标对所提算法进行定量分析,其中,平均精度均值(mean Average Preci-

首先,目标检测器执行检测任务,将输出结果根据给定阈值划分为高分检测框和低分检测框。同时,卡尔曼滤波器启动并初始化参数。

随后,进行两次数据关联操作,第一次为高分检测框与卡尔曼滤波器预测所得的跟踪轨迹进行匹配,若匹配成功,则进行跟踪轨迹的保存和卡尔曼滤波的更新;若高分检测框未匹配成功,系统为其创建新的跟踪轨迹,若跟踪轨迹未匹配成功,则进入第二次数据关联阶段。

此时,将第一次数据关联操作时未匹配成功的跟踪轨迹与低分检测框进行第二次匹配。若此次匹配成功,同样保存该轨迹,并对卡尔曼滤波进行更新;若匹配依旧失败,则暂时保留该跟踪轨迹。

当某条跟踪轨迹连续30次未能匹配成功时,系统则视为不活跃轨迹并予以删除。

sion, mAP)、小目标平均精度(Average Precision, AP_s)用于目标检测任务;多目标跟踪准确度(Multiple Object Traking Accuracy, MOTA)、多目标跟踪精度(Multiple Object Tracking Precision, MOTP)、保持身份一致性分数(Identification F1-score, IDF1)、目标跟踪轨迹占真实长度80%以上的轨迹数目百分比(Mostly Tracked, MT)、目标跟踪轨迹占真实长度最多20%的轨迹数目百分比(Mostly Lost, ML)、身份切换次数(IDs)、轨迹中断(Frag)、测试环境下算法运行速度(Frames per Second, FPS)用于多目标跟踪任务。为了更加直观地反应性能优劣,以下最优数据加粗表示。

2.2 实验平台与参数设置

实验在 Windows11 操作系统下运行, 安装有 CUDA 11.1 版本的并行计算框架, 采用 PyTorch 1.9.0 版本的深度学习框架, 处理器为 i5-12600kf, 显卡使用 NVIDIA GeForce RTX3060。检测器训练过程中, 每批样本数(batch size)设置为 4, 多线程进程(num workers)设置为 4, 权值初始学习率(learning rate)设置为 0.01, 动量(momentum)设置为 0.937, 优化器为 SGD。在跟踪阶段, 高检测分数阈值设置为 0.6, 低检测分数阈值为 0.1, 跟踪分数阈值为 0.7^[26]。

2.3 检测器性能实验

本节在 VisDrone 上展开针对检测器性能的消融实验, 以验证对检测器 YOLOX 所作优化的重要性, 实验具体内容如表 1 所示。在实验过程中, 逐步将 EfficientFormerV2 和 CBAM 模块集成到 YOLOX-S 网络中, 并与 YOLOX-M、YOLOX-L 和 YOLOX-S 进行比较。从表 1 第 4 行可以看出, 将骨干网络替换 ViT 网络 EfficientFormerV2 后, 模型在保持参数量 12.03 M、计算量 41.5 GFLOPs 的同时, 具备了较强的特征提取能力, 达到接近 YOLOX-M 的检测效果。这是因为 ViT 网络 EfficientFormerV2 采取自注意力机制和前馈神经网络的参数共享策略, 大幅度降低了检测器的空间复杂度, 优化了模型的存储需求。同时, 其相比卷积神经网络大量减少了卷积层的使用, 显著降低了模型的时间复杂度。此外, 得益于 EfficientFormerV2 网络中的自注意力机制可以有效捕捉图像的全局特征, 使得对小目标的检测效果也得到显著提升。由表 1 中第 5 行可知, 在主干网络输出端添加 CBAM, 可以在空间和通道两个维度上驱使网络正确地聚焦于目标物体, 进一步提升检测效果, 而网络参数仅增加 0.3 M。

表 1 检测器消融实验结果

Tab. 1 Results of ablation experiment

算法	参数量/M	计算量/ GFLOPs	mAP/%	AP _s /%
YOLOX-S	8.94	43.0	37.7	15.8
YOLOX-M	25.28	118.7	38.6	16.5
YOLOX-L	56.30	155.6	42.1	17.1
YOLOX-S+V2	12.03	41.5	38.4	16.9
YOLOX-S+ V2+CBAM	12.33	43.2	40.1	18.2

综上所述, YOLOX-L 性能最优, 但其网络参数量过大, 不利于在无人机等计算能力有限的边

缘设备上部署。而本文优化的检测器具有较低的时间复杂度和空间复杂度, 可以在参数量保持在 12.33 M、计算量 43.2 GFLOPs 基础上, 实现优异的检测效果, mAP 和 AP_s 分别达到 40.1% 和 18.2%, 在无人机视角应用中表现优秀。

2.4 多目标跟踪性能实验

本节将优化的 YOLOX 网络与优化的 ByteTrack 算法相结合形成一个完整的 MOT 算法 VA-ByteTrack, 并进行消融实验和对比实验验证本文算法的优越性。

首先, 通过消融实验验证本文所作优化对跟踪效果的提升程度。实验结果如表 2 所示, 其中 V2-CBAM 表示对检测器 YOLOX 的优化, AKF 表示对 ByteTrack 算法中卡尔曼滤波的优化。通过对比第 1 行和第 2 行数据, 可以看出在对检测器进行优化后, 由于检测器具备了更强的小目标检测能力, 各项跟踪指标得到了提升, MOTA 和 IDF1 分别提升了 0.5 和 1.8 个百分点, MOTP 下降了 0.1 个百分点。表中第 1 行和第 3 行对比可知, 引入基于 sigmoid 变换的 AKF 可以在略微牺牲处理速度的情况下显著提高算法的整体性能, MOTA、MOTP、IDF1 分别提升 0.6、1.9 和 1.1 百分点。由表中第 4 行可以看出, 优化的 YOLOX 网络与优化的 ByteTrack 算法结合所成的 VA-ByteTrack 算法, 实现了更为优异的多目标跟踪性能。相较于基准算法 ByteTrack, MOTA、MOTP 和 IDF1 分别提升了 1.0、1.9 和 2.7 百分点, FPS 下降了 0.38 帧/s。处理速度的下降主要归因于自适应因子的引入, 尽管在单个目标的跟踪中, 引入自适应因子所增加的参数量只有不到 100 FLOPs, 但是在多目标跟踪场景的应用中, 仍会导致时间复杂度的提升和跟踪速度的下降, 但考虑到实时性的要求, 损耗可以接受。

表 2 算法跟踪消融实验

Tab. 2 Algorithms track ablation experiments

算法	MOTA/%	MOTP/%	IDF1/%	FPS/ (帧·s ⁻¹)
ByteTrack (基准算法)	36.3	75.5	47.6	6.42
V2-CBAM 优化	36.8	75.4	49.4	6.45
AKF 优化	36.9	77.4	48.7	5.98
V2-CBAM+ AKF 优化	37.3	77.4	50.3	6.04

其次, 为了进一步验证本文算法的优越性, 开展了与主流 MOT 算法的对比实验, 实验结果如

表3所示。DeepSORT算法通过提取外观特征可以有效提高跟踪性能,但处理速度大幅下降;ByteTrack算法通过对低分检测框的有效利用,可以在保证速度的情况下有效提高跟踪效果,但其固定噪声设置限制了其跟踪精度的进一步提升;本文算法以ByteTrack算法为基准,通过对检测器YOLOX的优化和AKF的应用,取得了优异的多

目标跟踪效果。相比SORT、DeepSORT、ByteTrack等主流算法,本文算法MOTA分别提升1.7、1.4和1.0个百分点,MOTP分别提升0.7、0.5和1.9个百分点,Frag分别降低49.4%、7.7%和7.2%,IDs跳转问题分别降低46.3%、15.9%和12.1%,在无人机视角的多目标跟踪性能表现出色,满足了战场动态感知的实际需求。

表3 与主流MOT算法实验结果对比

Tab. 3 Comparison of ablation results

跟踪算法	MOTA/%	MOTP/%	IDF1/%	MT/%	ML/%	Frag	IDs	FPS/(帧·s ⁻¹)
SORT	35.6	76.7	47.1	22.9	52.3	1 823	1 496	6.56
DeepSORT	35.9	76.9	48.7	29.6	50.1	998	956	4.60
ByteTrack	36.3	75.5	47.6	30.1	53.0	993	915	6.42
ByteTrack+Re-ID	36.7	75.8	48.1	28.4	51.0	980	1 094	4.85
VA-ByteTrack(本文算法)	37.3	77.4	50.3	30.6	50.4	921	804	6.04

2.5 算法跟踪效果可视化分析

为了更直观地展示本文算法的优越性,将其与ByteTrack算法在截取的3段长约10 s的视频序列中进行跟踪效果的直接对比。其中,前两段视频序列属于公共数据集VisDrone MOT,最后一段视频序列属于战场航拍视频数据集。表4详细描述了两种算法在这3个视频序列中的跟踪表现,包括IDs、Frag和漏检次数。从表中可以清晰看出,本文算法可以显著降低轨迹中断、ID跳转和漏检问题。在白天密集道路场景中,轨迹中断次数由16次大幅减少至6次,ID切换次数由12次降至4次,漏检次数由14次降至7次;在夜晚光线质量差的场景中亦是如此,轨迹中断次数下降

3次,ID切换次数下降3次;战场环境下,本文算法在降低漏检次数的同时,完全消除了轨迹中断和ID切换问题。

表4 跟踪效果对比

Tab. 4 Tracking effect comparison

序列及目标总数	算法	Frag	IDs	漏检
1(白天,无人机快速移动,总数125)	ByteTrack	16	12	14
	本文算法	6	4	7
2(夜晚,光线质量差,总数52)	ByteTrack	5	3	2
	本文算法	2	0	2
3(战场,浓烟环绕,总数15)	ByteTrack	1	1	4
	本文算法	0	0	3

图7~图9展示了两种算法在视频序列中不同帧的实际跟踪效果,并将需要重点关注的目标以红色三角形标注出,进一步凸显两者之间的差异与优劣。

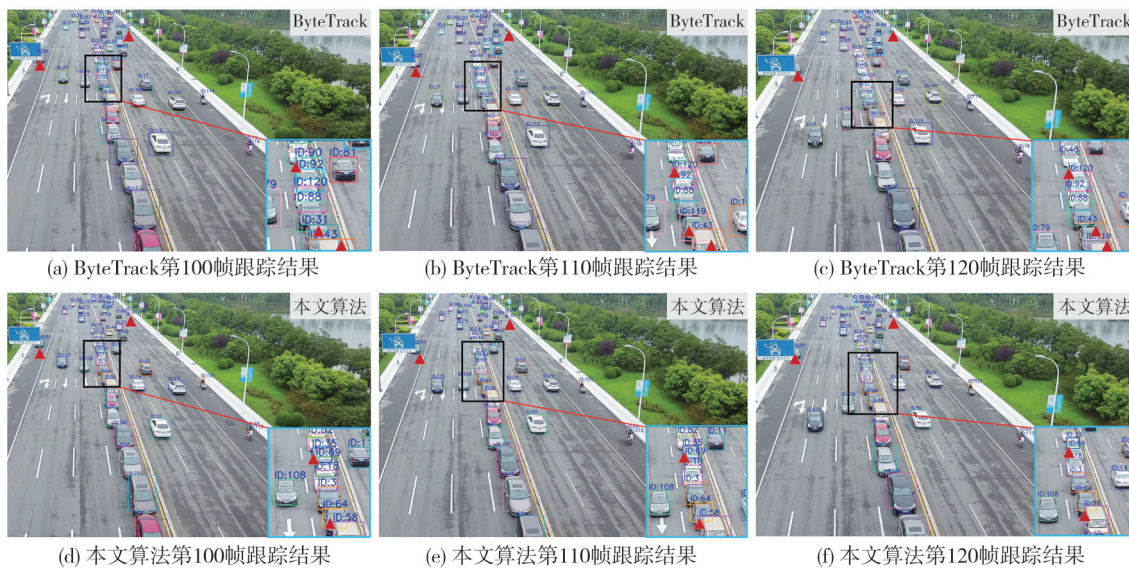


图7 密集环境跟踪效果对比

Fig. 7 Intensive environment tracking effect comparison

从图 7 可以看出,无人机航拍视频中目标小、特征模糊、遮挡问题严重,传统 ByteTrack 算法在实际应用过程中效果较差。例如,最左侧被广告牌遮挡的车辆被漏检,左数第四车道因车辆间的相互遮挡出现了 ID 跳转和漏检问题,最右侧

远端行人因特征模糊也出现了不同程度的漏检问题。而本文算法在实验中取得了不错的效果,无论是左侧广告牌遮挡的车辆、左数第四车道相互遮挡的车辆还是右侧道路行人均实现了高质量的跟踪。



图 8 夜晚跟踪效果对比

Fig. 8 Night tracking effect comparison

在黑暗场景的跟踪效果如图 8 所示,该视频片段由于无人机的旋转导致图片模糊,再加上夜晚光线较差,传统 ByteTrack 算法在对上方道路新出现的摩托车、左下角人群中的小女孩均出现了 ID 跳转问题或漏检问题。相比之下本文算法在跟踪过程中未出现异常。

军事目标的跟踪效果对比,可以清楚观察到,本文算法具备较强的抗干扰能力,即使在浓烟密布的环境下依然能保持稳定的跟踪,未出现 ID 跳转问题。同时,AKF 通过给予高分检测框以较低的观测协方差,使得跟踪框更加接近对象本体、精度更高,这在基于无人机的战场动态感知的实际应用中将发挥重要的作用。

图 9 展示在战场航拍视频资源上的对地面多

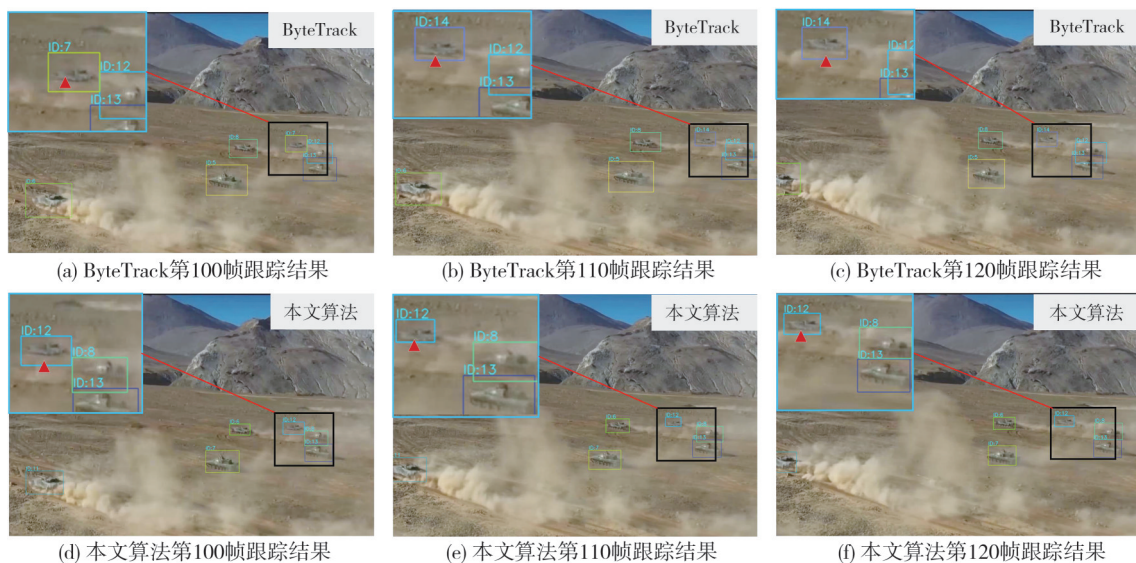


图 9 战场跟踪效果对比

Fig. 9 Battlefield tracking effect comparison

综上,本文算法可以降低小目标特征模糊、目标密集且相互遮挡、光线偏暗、浓烟干扰等因素的影响,在无人机视角应用中表现出色,满足基于无人机视角的战场动态感知的实际需求。

3 结论

针对当前多目标跟踪算法在军事无人机视角场景应用时受到目标小、特征模糊、目标密集且相互遮挡等因素干扰所导致的轨迹中断问题和ID跳转问题,本文提出一种基于无人机视角的地面多军事目标跟踪算法,利用融合ViT网络和添加CBAM注意力机制的检测器YOLOX,提高对小尺寸目标的特征提取能力,利用基于sigmoid函数AKF平衡高分检测框和低分检测框的不同协方差需求,实现了无人机平台下地面多军事目标的稳定跟踪。

实验结果表明,本文算法可以有效改善跟踪质量,相比SORT、DeepSORT、ByteTrack等主流算法,本文算法跟踪准确度分别提升1.7%、1.4%和1.0百分点;跟踪精度分别提升0.7%、0.5%和1.9百分点;轨迹中断问题分别降低49.4%、7.7%和7.2%;ID跳转问题分别降低46.3%、15.9%和12.1%。同时,该算法在战场航拍视频资源的实际测试中也展现了卓越的性能,满足基于无人机视角的战场动态感知的实际需求。

参考文献:

- [1] 蒋超,崔玉伟,王辉. 基于图像的无人机战场态势感知技术综述[J]. 测控技术, 2021, 40(12): 14-19.
JIANG Chao, CUI Yuwei, WANG Hui. Review on battlefield situational awareness of UAV based on image technology[J]. Measurement & Control Technology, 2021, 40(12): 14-19. (in Chinese)
- [2] 阮国庆,易侃,孙家栋,等. 智能战场感知技术研究现状与发展趋势[J]. 指挥信息系统与技术, 2022, 13(3): 17-22.
RUAN Guoqing, YI Kan, SUN Jiadong, et al. Research status and development trend of intelligent battlefield perception technology[J]. Command Information System and Technology, 2022, 13(3): 17-22. (in Chinese)
- [3] 吴立珍. 面向UAV战场感知的目标特征建模与应用研究[D]. 长沙:国防科学技术大学, 2013.
- [4] 孙志强,宋丽丽. 基于无人机的安防监控系统在智慧城市中的应用前景与实现[J]. 城市建设理论研究(电子版), 2024(3): 211-213.
- [5] SUN Zhiqiang, SONG Lili. Application prospect and realization of security monitoring system based on UAV in smart city[J]. Theoretical Research in Urban Construction, 2024(3): 211-213. (in Chinese)
- [5] 张怡,陈平. 基于深度置信网络的无人机航拍图像变化检测[J]. 测试技术学报, 2020, 34(3): 190-196.
ZHANG Yi, CHEN Ping. Change detection of unmanned aerial vehicle images based on deep belief networks[J]. Journal of Test and Measurement Technology, 2020, 34(3): 190-196. (in Chinese)
- [6] ZHANG Y, MA R G, LIANG C. Road target detection algorithm based on improved YOLOv5 in UAV images [J]. Journal of Measurement Science and Instrumentation, 2024, 15(1): 128-139.
- [7] 欧阳权,张怡,马延,等. 基于深度学习的无人机航拍目标检测与跟踪方法综述[J]. 电光与控制, 2024, 31(3): 1-7.
OUYANG Quan, ZHANG Yi, MA Yan, et al. A review of UAV aerial photography target detection and tracking methods based on deep learning[J]. Electronics Optics & Control, 2024, 31(3): 1-7. (in Chinese)
- [8] 李松林,江剑. 改进YOLOv5的无人机小目标检测算法[J]. 测试技术学报, 2024, 38(4): 354-362.
LI Songlin, JIANG Jian. Improved UAV small object detection algorithm based on YOLOv5[J]. Journal of Test and Measurement Technology, 2024, 38(4): 354-362. (in Chinese)
- [9] 梁天一. 基于深度学习的视觉多目标跟踪方法研究[D]. 长沙:国防科技大学, 2022.
- [10] 羊洁明. 基于深度学习的视频多目标跟踪算法研究[D]. 无锡:江南大学, 2022.
- [11] BEWLEY A, GE Z, OTT L, et al. Simple online and realtime tracking [C]//2016 IEEE International Conference on Image Processing (ICIP), 2016: 3464-3468.
- [12] WOJKE N, BEWLEY A, PAULUS D. Simple online and realtime tracking with a deep association metric [C]//2017 IEEE International Conference on Image Processing (ICIP), 2017: 3645-3649.
- [13] 罗茜,赵睿,庄慧珊,等. YOLOv5与Deep-SORT联合优化的无人机多目标跟踪算法[J]. 信号处理, 2022, 38(12): 2628-2638.
LUO Xi, ZHAO Rui, ZHUANG Huishan, et al. UAV multi-target tracking algorithm jointly optimized by YOLOv5 and deep-SORT [J]. Journal of Signal Processing, 2022, 38(12): 2628-2638. (in Chinese)
- [14] 高质桐. 基于航拍图像的城市交通场景自然驾驶数

- 据集构建方法研究[D]. 长春: 吉林大学, 2024.
- [15] ZHANG Y, SUN P, JIANG Y, et al. ByteTrack: multi-object tracking by associating every detection box [M]//Computer Vision-ECCV 2022. Cham: Springer Nature Switzerland, 2022: 1-21.
- [16] 朱奇光, 商健, 刘博, 等. 基于无人机航拍视频车辆多目标跟踪算法研究[J]. 计量学报, 2024, 45(12): 1772-1779.
- SHANG Jian, LIU Bo, et al. Research on vehicle multi-target tracking algorithm based on UAV aerial video [J/OL]. Acta Metrologica Sinica, 2024, 45(12): 1772-1779. (in Chinese)
- [17] LIU Y, LI Y, XU D, et al. Adaptive Kalman Filter with power transformation for online multi-object tracking [J]. Multimedia Systems, 2023, 29(3): 1231-1244.
- [18] ZHANG Y, TAO Q, YIN Y. A lightweight man-overboard detection and tracking model using aerial images for maritime search and rescue [J]. Remote Sensing, 2023, 16(1): 165.
- [19] YUAN K, WANG Q, MI Y, et al. Improved feature fusion in YOLOv5 for accurate detection and counting of Chinese flowering cabbage (*brassica campestris* L. ssp. *chinensis* var. *utilis* tsen et lee) buds [J]. Agronomy, 2023, 14(1): 42.
- [20] GUO X, ZHAO X. Oriented-YOLOX: an improved YOLOX for rotated object detection[C]//Proceedings of the 2024 5th International Conference on Computing, Networks and Internet of Things, 2024: 261-265.
- [21] ASIRI A A, SHAF A, ALI T, et al. Advancing brain tumor classification through fine-tuned vision transformers: a comparative study of pre-trained models[J]. Sensors, 2023, 23(18): 7913.
- [22] LI Y, HU J, WEN Y, et al. Rethinking vision transformers for MobileNet size and speed [C]//2023 IEEE/CVF International Conference on Computer Vision (ICCV), 2023: 16843-16854.
- [23] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module [C]//2018 IEEE/CVF International Conference on Computer Vision (ICCV), 2018: 3-19.
- [24] 李顺芳. 多目标跟踪用补偿-自适应卡尔曼滤波器设计[J]. 滁州学院学报, 2023, 25(5): 30-34.
- LI Shunfang. Design of compensation-adaptive Kalman filtering for online multi object tracking[J]. Journal of Chuzhou University, 2023, 25(5): 30-34. (in Chinese)
- [25] CHEN G, WANG W, HE Z, et al. VisDrone-MOT2021: the vision meets drone multiple object tracking challenge results[C]//2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW), 2021: 2839-2846.
- [26] XU X, FENG Z, CAO C, et al. STN-track: multiobject tracking of unmanned aerial vehicles by swin transformer neck and new data association method[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2022, 15: 8734-8743.