中图法分类号:TP391.41　　**文献标识码**:A　　**文章编号**:1006-8961(XXXX)XX-0001-16

论文引用格式:Si Jingzi, Li Na, Tang Song, Zhao Qijun. Observation-reliability-aware track-state adaptationfor multi-object tracking: RA-BoT-SORT[J/OL]. Journal of Image and Graphics, XXXX:1-16. DOI: 10.11834/jig.260605. (司靖梓, 李娜, 唐松, 赵启军. 观测可靠性引导的轨迹状态自适应多目标跟踪:RA-BoT-SORT[J/OL]. 中国图象图形学报,XXXX:1-16. DOI: 10.11834/jig.260605.)[DOI:10.11834/jig.260605]

观测可靠性引导的轨迹状态自适应多目标跟踪:RA-BoT-SORT

司靖梓[1],李娜[1],唐松[2],赵启军[1,2]

1. 四川大学计算机学院(软件学院),成都　610065; 2. 四川大学视觉合成图形图像技术国家级重点实验室,成都　610065

摘　要:**目的**　基于检测的多目标跟踪(multi-object tracking,MOT)在完成关联后仍需更新外观模板、运动状态与轨迹保留期,低可靠性观测可能通过上述状态影响后续身份保持。为限制这一影响,提出观测可靠性引导的轨迹状态自适应多目标跟踪方法(observation-reliability-aware track-state adaptation for multi-object tracking,RA-BoT-SORT)。**方法**　在保持鲁棒关联多行人跟踪(robust associations multi-pedestrian tracking,BoT-SORT)基线检测器、行人重识别(re-identification,ReID)网络和主体关联流程不变的条件下,设计三类有界调节:结合检测置信度与边界框面积比例,以限定权重的指数移动平均(exponential moving average,EMA)控制外观特征写入;以基线测量协方差为下界进行置信度自适应卡尔曼校正;利用轨迹历史置信度调整丢失轨迹的有效保留期。三个模块不增加网络参数或额外训练。**结果**　相较同配置复现基线,在采用相同线性插值后处理的MOT17-test结果中,关联准确率(association accuracy,AssA)、身份F1分数(identification F1 score,IDF1)和高阶跟踪准确率(higher order tracking accuracy,HOTA)分别由65.0%、79.9%和64.8%提高至65.5%、80.5%和65.1%,身份切换次数(identity switches,IDSW)由1358降至1122,减少17.4%;MOT20-test上,上述三项指标分别由61.7%、76.4%和62.7%提高至62.1%、76.7%和62.9%,IDSW由1385降至1346,减少2.8%。单模块消融在两个验证集上均提高AssA并减少IDSW,完整组合体现多指标权衡。跟踪阶段平均时延增加0.396 ms/帧(0.54%),端到端速度约4.63帧/秒,峰值张量分配显存不变。**结论**　在MOT17与MOT20的行人跟踪条件下,针对不同状态对象实施有界调节,能够在保持BoT-SORT主体流程的同时,以有限额外运行开销改善身份一致性;其收益体现为身份保持与检测误差之间的多指标权衡。

关键词:多目标跟踪;观测可靠性;轨迹状态调节;身份一致性;卡尔曼滤波;轨迹管理

Observation-reliability-aware track-state adaptationfor multi-object tracking: RA-BoT-SORT

Si Jingzi[1], Li Na[1], Tang Song[2], Zhao Qijun[1,2]

1. College of Computer Science, Sichuan University, Chengdu 610065, China; 2. National Key Laboratory of Fundamental Science on Synthetic Vision, Sichuan University, Chengdu 610065, China

Abstract: **Objective** Detection-based multi-object tracking (MOT) must maintain consistent identities while estimating the positions of multiple targets in successive frames. A successful association does not necessarily imply that the associ-

收稿日期:2026-09-22;**修回日期**:2026-11-18

*通信作者:赵启军,男,四川大学计算机学院(软件学院)教授、博士生导师、副院长,主要研究方向为生物特征识别、计算机视觉。E-mail:qjzhao@scu.edu.cn;赵启军 qjzhao@scu.edu.cn

基金项目:国家自然科学基金项目(62671402,62176170)

Supported by: National Natural Science Foundation of China (62671402, 62176170)

© 中国图象图形学报版权所有

ated observation is suitable for updating every component of a track with the same strength. An observation can correspond to the correct person while its bounding box is displaced or its appearance feature is degraded by occlusion, blur, or limited image resolution. Writing such information into appearance memory, the motion posterior, or the track-retention policy may influence subsequent associations. Existing confidence-aware tracking methods provide important foundations, but the role and permissible magnitude of confidence-dependent updates differ across these states. The objective is to improve identity consistency through bounded regulation of observation-dependent state updates while retaining the detector, re-identification (ReID) network, and main association procedure of robust associations multi-pedestrian tracking (BoT-SORT). **Method** The proposed observation-reliability-aware track-state adaptation for multi-object tracking (RA-BoT-SORT) comprises composite-reliability-guided appearance updating (CRAU), lower-bound-constrained confidence-adaptive Kalman correction (LBCA-KC), and reliability-aware track retention (RATR). CRAU combines the detection confidence and the ratio of bounding-box area to image area as an operational proxy for appearance updating. The two inputs are averaged and clipped to the unit interval. This proxy adjusts the historical weight of an exponential moving average (EMA) within 0.90 to 0.93, so the coefficient of the current feature remains between 0.07 and 0.10 before feature normalization. The update is applied only when the matched detection has an available re-identification feature; the first available feature initializes the template. The area ratio supplies scale information but is not treated as a direct measurement of visibility or appearance quality. LBCA-KC scales the baseline measurement covariance using the current matched detection confidence. With the adopted noise coefficient and a confidence in the unit interval, the scale lies between one and two. The baseline covariance is retained as a lower bound: high-confidence measurements approach the baseline update, whereas lower-confidence measurements receive enlarged measurement covariance. The constraint applies at the same predicted state and observation model; it does not impose a constant covariance across tracks or frames. RATR maintains an exponential moving average of the confidence scores of successful matches, including reactivations. This historical score controls the effective retention period of a lost track through a bounded mapping. Before integer truncation, the retention factor lies between 0.8 and 1.3 under the stated confidence range. The historical score remains unchanged during a period without successful observations. A longer retention window allows more opportunities for reassociation but neither generates missing detections nor guarantees recovery. The modules share a bounded-update principle while using state-specific inputs and time scales; they do not share a single composite score or form a serial prediction network. All three modules use quantities already available in the tracking pipeline and introduce no additional trainable network parameters. The detection score is retained from the baseline detector and is not interpreted as a calibrated probability of localization correctness. Equal weighting of score and area does not imply equal statistical contributions, because their empirical distributions can differ substantially. The bounds are therefore designed to restrict the magnitude of adaptation rather than guarantee that unreliable observations are always recognized. In particular, appearance updating, measurement correction, and retention control are activated under their respective conditions, while the original two-stage matching and reactivation criteria remain unchanged. Evaluation uses a reproduced BoT-SORT-ReID baseline with matched detector weights, appearance weights, association settings, numerical precision, and post-processing. MOT17 and MOT20 validation experiments include eight module combinations, four appearance-update settings, alternative Kalman correction strategies, and parameter sensitivity analysis. Test-set accuracy is reported after identical sequence-level linear interpolation, which is distinguished from the online tracker update. **Result** On MOT17-test, association accuracy (AssA), identification F1 score (IDF1), and higher order tracking accuracy (HOTA) increase from 65.0%, 79.9%, and 64.8% for the reproduced baseline to 65.5%, 80.5%, and 65.1%, respectively. Identity switches decrease from 1,358 to 1,122, a relative reduction of 17.4%. False positives decrease from 26,651 to 22,509, whereas false negatives increase from 81,870 to 85,758, indicating that the improvement does not extend uniformly to every error component. On MOT20-test, AssA, IDF1, and HOTA increase from 61.7%, 76.4%, and 62.7% to 62.1%, 76.7%, and 62.9%, respectively. Identity switches decrease from 1,385 to 1,346, corresponding to a 2.8% reduction. The smaller identity-switch reduction on MOT20 indicates scene-dependent benefits rather than a uniform improvement across data distributions. Each individual module improves AssA and reduces identity switches on both validation splits. The full configuration improves the principal identity-related metrics relative to the baseline, but some two-module configurations achieve higher AssA. Similarly, the

© 中国图象图形学报版权所有

composite appearance proxy offers a multi-metric compromise rather than consistently outperforming confidence alone or fixed EMA on every metric. These results support the usefulness of state-specific regulation under the evaluated conditions without establishing universally positive interactions between modules. Runtime is measured for the baseline, three single-module configurations, and the full method on the same NVIDIA GeForce RTX 4090 server. Each configuration is run independently three times on seven MOT17 validation videos, excluding the first 30 frames of each video and measuring 2, 442 frames per run. End-to-end timing includes image decoding, preprocessing, detector inference, tracking, data transfers, and output preparation in memory. Tracker-only timing covers the update interface, including re-identification, camera-motion compensation using precomputed transforms, association, and track management. The baseline and full method have mean end-to-end latencies of 215.890 and 215.881 ms per frame, respectively, with processing rates of approximately 4.63 frames per second. Their tracker-only latencies are 72.809 and 73.204 ms per frame, giving a paired increase of 0.396±0.516 ms per frame, or 0.54% relative to the baseline mean. Peak tensor-allocation memory is 4, 340.356 MiB for both methods. The end-to-end paired difference is −0.010±0.986 ms per frame; its small magnitude relative to run-to-run variation does not indicate acceleration. These measurements describe configuration-level overhead, including changes in retained-track populations as well as state-update operations. Published results of other trackers provide benchmark context, whereas performance increments attributed to RA-BoT-SORT are derived from the matched baseline comparison. **Conclusion** Under the evaluated MOT17 and MOT20 pedestrian-tracking conditions, bounded regulation of appearance memory, motion correction, and lost-track retention improves identity consistency without extra network training. Its contribution lies in the state-specific update directions, bounds, and activation conditions, rather than confidence weighting alone. The observed benefits remain subject to detection-association trade-offs and the limitations of score and area proxies. Further work can separate direct module operations from changes in retained-track populations.

Key words: multi-object tracking; observation reliability; track-state adaptation; identity consistency; Kalman filtering; track management

论文引用格式:【DOI:10.11834/jig.260605】

0 引　言

基于检测的多目标跟踪(multi-object tracking, MOT)通过跨帧关联维持目标身份。简易在线实时跟踪(simple online and realtime tracking, SORT)采用卡尔曼预测与匈牙利分配(Bewley等,2016);深度外观度量增强的SORT(DeepSORT)引入外观特征(Wojke等,2017);ByteTrack将低置信度检测纳入第二阶段关联(Zhang等,2022)。鲁棒关联行人跟踪方法(robust associations multi-pedestrian tracking, BoT-SORT)结合相机运动补偿、交并比(intersection over union, IoU)和重识别(re-identification, ReID)特征(Aharon等,2022);其外观增强版本记为BoT-SORT-ReID。

成功关联不意味着观测适合以相同强度更新每一类状态。定位偏差和外观失真可能通过运动后验和历史模板影响后续匹配,固定保留期也未区分丢失前的匹配历史。置信度感知更新已有研究基础;本文关注在保持BoT-SORT关联规则不变时,如何分别为外观、运动和轨迹保留选择输入、约束更新幅度并规定触发条件。

为此,构建观测可靠性引导的轨迹状态自适应跟踪方法(observation-reliability-aware track-state adaptation for multi-object tracking, RA-BoT-SORT)。在既有置信度感知策略基础上,按状态对象区分输入、调节边界与触发条件,具体改进如下。

1)复合可靠性引导的外观更新(composite-reliability-guided appearance updating, CRAU)将检测分数与面积比例复合作为外观更新输入,采用指数移动平均(exponential moving average, EMA),将历史权重限定在0.90～0.93;仅在匹配检测具有ReID特征时更新模板。

2)下界约束的置信度自适应卡尔曼校正(lower-bound-constrained confidence-adaptive Kalman correction, LBCA-KC)以同一预测状态生成的基线测量协方差为下界,仅在匹配后依据检测分数放大测量噪声,保留基线关联门控。

3)可靠性感知轨迹保留(reliability-aware track retention, RATR)以成功匹配分数的历史EMA调节丢失轨迹的有限保留期;普通匹配和重激活时更新

© 中国图象图形学报版权所有

历史量，丢失期间冻结，严格超期后移除，保持基线重激活判据不变。

1 相关工作

1.1 基于检测的在线多目标跟踪

观测中心在线跟踪（observation-centric SORT，OC-SORT）从观测出发改进运动估计（Cao等，2023）。置信度引导的多行人跟踪（confidence-guided multi-person tracking，ConfTrack）使用检测分数修正更新（Jung等，2024）；非线性预测的扩散式多目标跟踪（diffusion-based multiple object tracking with non-linear prediction，DiffMOT）引入学习式运动预测（Lv等，2024）。郭文等（2024）结合自适应IoU回归损失与层级关联，利用未匹配检测框周围的嵌入信息处理遮挡关联。

纯Transformer在线跟踪（pure Transformer for decoupled and online multi-object tracking，PuTR）学习时序关联（Liu等，2025）；McByte将时序掩码作为关联线索（Stanczyk等，2025）；PlugTrack融合卡尔曼预测和数据驱动预测，其与DiffMOT的组合记为PlugTrack - DiffMOT（Kim等，2026）；遮挡感知在线跟踪（occlusion-aware SORT，OA-SORT）修正位置代价和状态估计（Li等，2026）。与这些方法不同，RA-BoT-SORT保留基线的关联规则和运动预测器，仅调节匹配后的状态更新与丢失轨迹保留。

1.2 外观模板与可靠性相关更新

韩清等（2026）梳理了遮挡、换装及跨模态等特殊场景下的目标重识别研究。StrongSORT沿用深度外观跟踪并采用固定EMA，增强版StrongSORT++使用无外观轨迹连接（appearance-free link model，AFLink）和高斯平滑插值（Gaussian-smoothed interpolation，GSI）（Du等，2023）。基于自适应重识别的行人跟踪（Deep OC-SORT）按检测分数相对阈值的位置调整外观EMA，达到阈值时可停止当前特征写入（Maggiolino等，2023）。在多相机多目标跟踪中，Tang等（2026）提出VistaTrack，按目标在不同视角下的动态可靠性加权融合外观特征。在本文的单相机场景中，CRAU在分数外引入面积线索，将当前特征系数限制在0.07～0.10，差异体现为复合输入和有界写入，而非新增外观网络。

1.3 置信度感知的运动状态估计

GIAOTracker和StrongSORT的噪声尺度自适应卡尔曼滤波（noise scale adaptive Kalman filter，NSA-KF）依据置信度缩放测量噪声（Du等，2021，2023）。LBCA-KC以同一预测状态下的基线测量协方差为下界：低分时放大噪声，高分时趋近基线，而不进一步压低噪声。第4.5节以相同置信度输入比较两种校正映射，不将差异等同于完整跟踪器的性能差异。

1.4 置信度在关联与轨迹存续管理中的应用

置信度轨迹预测跟踪（confidence-based trajectory prediction for multi-object tracking，ConfMOT）以累积置信度和未匹配时长调节轨迹删除（Yi等，2025）；Sentinel依据轨迹状态重加权关联代价，并搜索弱检测以帮助临近终止的轨迹恢复（Yang等，2026）。RATR以成功匹配分数的历史EMA直接确定有界保留期，丢失期间冻结历史量，以严格超期触发移除，不改变关联及重激活判据。

2 BoT-SORT基线与问题定义

2.1 基线流程

检测器在第t帧给出的候选集合为$\boldsymbol{D}_t=\left\{\left(\boldsymbol{b}_j,\ s_j\right)\middle|j=1,\ \cdots,\ N_t\right\}$。式中，$\boldsymbol{D}_t$为第$t$帧候选集合；$\boldsymbol{b}_j$和$s_j$分别为第$j$个候选的边界框与置信度；$N_t$为候选数。

运动状态向量与观测向量分别为$\boldsymbol{x}_t=\left[x_c,\ y_c,\ w,\ h,\ \dot{x}_c,\ \dot{y}_c,\ \dot{w},\ \dot{h}\right]^{\mathrm{T}}$与$\boldsymbol{z}_t=\left[x_c,\ y_c,\ w,\ h\right]^{\mathrm{T}}$。式中，$\boldsymbol{x}_t$为状态向量，$\boldsymbol{z}_t$为观测向量，$x_c$、$y_c$为框中心坐标，$w$和$h$为框宽和框高，上方圆点表示帧间变化率。全局运动补偿（global motion compensation，GMC）用仿射变换修正预测状态与协方差。

在数据关联阶段，BoT-SORT首先依据检测置信度将候选目标划分为高置信检测集合和低置信检测集合，并仅为高置信检测提取ReID特征。在第一阶段关联中，分别计算轨迹与高置信检测之间的IoU距离和外观余弦距离，并利用外观阈值及IoU邻近条件对ReID距离进行门控；随后，对门控后的外观距离与IoU距离逐元素取较小值，构造最终关联代价矩阵，并通过匈牙利算法完成匹配。第二阶段主要依据IoU距离，将第一阶段剩余的已确认轨迹与低置信检测进行匹配。两个关联阶段的成功匹

© 中国图象图形学报版权所有

配均用于校正卡尔曼状态，但只有包含 ReID 特征的高置信匹配会更新轨迹的外观 EMA。未能匹配的已确认轨迹被标记为丢失状态，并在超过预设缓冲期限后移除。

RA-BoT-SORT沿用上述检测划分、关联代价和分配规则，只调节匹配后的外观模板、测量校正以及丢失轨迹的保留期。

2.2 可靠性代理与状态更新

检测置信度沿用检测器的输出处理：同时提供目标性分数和类别分数时取二者乘积。框面积比例描述相对尺度，历史置信度汇集成功匹配的分数。三者是状态调节的操作性代理，不是经真值校准的质量概率。

式(1)—(7)按单条轨迹书写，当前检测指该轨迹在当前帧的匹配观测；式(8)和算法1列出必要的轨迹索引。

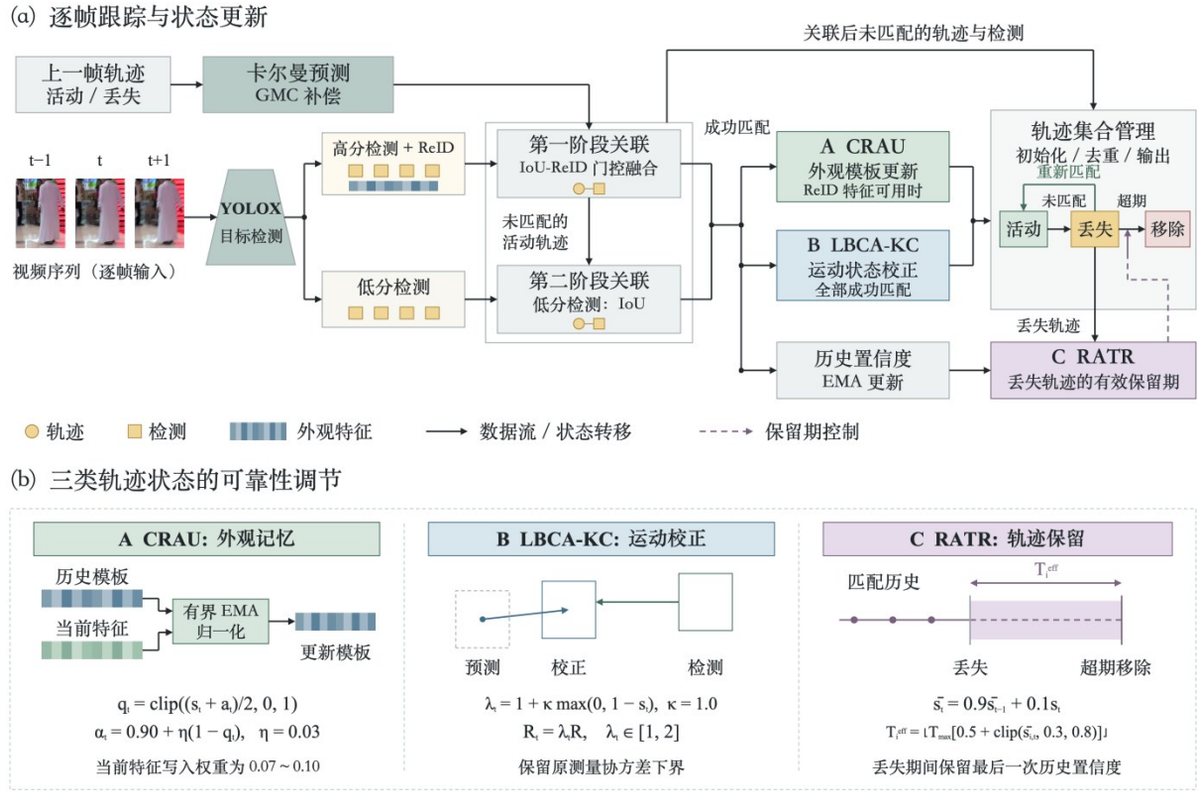
固定外观权重、测量噪声和保留期分别决定错误如何写入模板、影响运动状态和延续至后续关联。下文针对这三类状态规定调节方向与幅度。

3 轨迹状态调节方法

3.1 总体框架与设计约束

图1给出总体流程。图中输入示例取自更快速的区域卷积神经网络(faster region-based convolutional neural network，FRCNN)检测条目 MOT17-12-FRCNN 的第58、59和60帧。CRAU限制当前特征写入外观模板的幅度，LBCA-KC调节匹配后的测量校正，RATR根据历史分数调整丢失轨迹保留期。三者分别使用与状态相关的线索，不共用单一质量分数。

更新后的模板影响后续外观距离，运动后验影响预测框和邻近门控，保留期影响参与关联的候选轨迹。关联规则不变并不意味着匹配结果不变。三项调节通过后续匹配发生联系，不构成串行预测网络；其独立与组合效果由第4.3节分别检验。



(a)总体跟踪与更新流程；(b)三类轨迹状态的可靠性调节机制

(a) Overall tracking and update pipeline; (b) Reliability-guided regulation of three track states.

图1 RA-BoT-SORT总体框架

Fig. 1 Framework of RA-BoT-SORT

© 中国图象图形学报版权所有

图1(a)中,YOLOX目标检测器按分数输出高、低置信检测,仅为高分检测提取ReID特征;活动和丢失轨迹经预测及GMC后参与原有两阶段匹配。

图1(b)展示三个模块的触发条件。跟踪器按时间顺序逐帧处理,不使用未来观测。

3.2 复合可靠性引导的外观更新(CRAU)

先以框面积比例和检测分数构造外观更新代理:

$$a_t = \frac{\max\left[\left(x_2 - x_1\right)\left(y_2 - y_1\right), 1\right]}{\max\left(HW, 1\right)}$$

$$q_t = g\left(\frac{s_t + a_t}{2}, 0, 1\right)。 \tag{1}$$

式中,a_t为框面积占图像面积的比例;x_1、y_1、x_2、y_2依次为匹配框左上、右下顶点的横纵坐标;H和W分别为图像高和宽;q_t为复合代理,g为区间裁剪函数(图1中记作clip),s_t为匹配检测置信度。

以该代理计算历史权重,并更新外观模板:

$$\alpha_t = 0.90 + \eta\left(1 - q_t\right),\quad \eta = 0.03, \tag{2}$$

式中,α_t为历史权重,η为调节幅度。

$$\bar{\boldsymbol{f}}_t = n\left[\alpha_t \bar{\boldsymbol{f}}_{t-1} + \left(1 - \alpha_t\right)\boldsymbol{f}_t\right]。 \tag{3}$$

式中,$\bar{\boldsymbol{f}}_t$为当前模板,n为二范数归一化函数,$\bar{\boldsymbol{f}}_{t-1}$为上一帧模板,$\boldsymbol{f}_t$为已归一化的当前外观特征。

历史权重在0.90~0.93内变化,当前特征的写入系数为0.07~0.10。首个可用特征初始化模板;后续仅在匹配检测带有ReID特征时执行更新。

面积比例仅反映尺度,不等同于可见面积或外观质量;等权平均也不意味着两项输入具有相同的信息量。

3.3 下界约束的置信度自适应卡尔曼校正(LBCA-KC)

匹配后的测量协方差按检测分数缩放:

$$\lambda_t = 1 + \kappa\max\left(0, 1 - s_t\right),$$

$$\boldsymbol{R}_t = \lambda_t \boldsymbol{R},\quad \kappa = 1.0。 \tag{4}$$

式中,λ_t为测量噪声倍率,κ为调节系数,$\boldsymbol{R}_t$为缩放后的测量协方差,$\boldsymbol{R}$为同一预测状态生成的基线测量协方差。

当检测置信度位于[0,1]时,缩放系数在1~2之间。高分匹配趋近基线更新,低分匹配增大测量噪声;门控仍沿用基线观测模型。该倍率范围依赖于检测分数的取值范围。

在同一预测状态下,创新协方差及卡尔曼增益为

$$\boldsymbol{S}_t = \boldsymbol{H}\boldsymbol{P}^{-}\boldsymbol{H}^{\mathrm{T}} + \lambda_t \boldsymbol{R},\quad \boldsymbol{K}_t = \boldsymbol{P}^{-}\boldsymbol{H}^{\mathrm{T}}\boldsymbol{S}_t^{-1}。 \tag{5}$$

式中,$\boldsymbol{S}_t$为创新协方差,$\boldsymbol{H}$为观测矩阵,$\boldsymbol{P}^{-}$为预测协方差,上标T表示转置,$\boldsymbol{K}_t$为卡尔曼增益,上标-1表示矩阵求逆。设预测协方差半正定、基线测量协方差正定,则有

$$\lambda_t \geqslant 1 \Rightarrow \boldsymbol{S}_t \succcurlyeq \boldsymbol{S}_0 \Rightarrow \boldsymbol{S}_t^{-1} \preccurlyeq \boldsymbol{S}_0^{-1}。 \tag{6}$$

式中,$\boldsymbol{S}_0$为不缩放测量噪声时的创新协方差,符号≽和≼表示半正定序。

式(6)只比较同一预测状态下的创新协方差及其逆,不能推出一般矩阵情形下卡尔曼增益的每个元素均单调减小。

3.4 可靠性感知轨迹保留(RATR)

轨迹建立时以首个检测置信度初始化历史分数,普通匹配与重激活时更新:

$$\bar{s}_t = \rho\,\bar{s}_{t-1} + (1 - \rho)s_t。 \tag{7}$$

式中,$\bar{s}_t$为更新后的历史置信度,ρ为历史平滑权重(取0.9),$\bar{s}_{t-1}$为更新前的历史置信度。丢失期间不更新历史分数。

以历史分数确定丢失轨迹的有效保留期:

$$T_i^{\mathrm{eff}} = \left\lfloor T_{\max}\cdot\left[0.5 + g\left(\bar{s}_{i,t}, 0.3, 0.8\right)\right]\right\rfloor。 \tag{8}$$

式中,T_i^{eff}为第i条轨迹的有效保留期,$T_{\max}$为基准保留期,$\bar{s}_{i,t}$为该轨迹当前保存的历史置信度,外层符号表示向下取整。

取整前的保留倍率为0.8~1.3;取整后实际倍率可能略低于0.8。只有丢失帧数严格超过有效保留期时才移除轨迹。延长窗口仅增加重关联机会,不改变匹配判据。

3.5 单帧更新流程

算法1给出单帧更新过程。两个阶段的成功匹配均参与测量校正和历史分数更新,仅具有ReID特征的匹配参与外观更新。“维护轨迹集合”沿用基线的未确认轨迹处理、建轨、重激活和集合维护规则。

式中,$\boldsymbol{I}_t$为当前帧图像;$\boldsymbol{T}_t$表示轨迹集合,上标A、L、R分别指活动、丢失和移除,波浪线表示预测结果。$\boldsymbol{M}_t$为匹配集合,$\boldsymbol{U}$为未匹配集合,上标T和D分别指轨迹与检测。τ_i和d_j分别为第i条轨迹与第j个检测对象,$\boldsymbol{R}_j$为该匹配的测量协方差。算法中的检测索引与式(1)的匹配观测对应。

算法1 RA-BoT-SORT单帧更新流程

输入:当前帧图像$\boldsymbol{I}_t$;上一帧活动轨迹$\boldsymbol{T}_{t-1}^{\mathrm{A}}$;上

© 中国图象图形学报版权所有

一帧丢失轨迹$\boldsymbol{T}_{t-1}^{\mathrm{L}}$；上一帧移除轨迹$\boldsymbol{T}_{t-1}^{\mathrm{R}}$；按预测状态确定的基线测量协方差$\boldsymbol{R}$；基准保留期$T_{\max}$.

输出：活动轨迹$\boldsymbol{T}_t^{\mathrm{A}}$；丢失轨迹$\boldsymbol{T}_t^{\mathrm{L}}$；移除轨迹$\boldsymbol{T}_t^{\mathrm{R}}$.

$\boldsymbol{D}_t \leftarrow$ 检测($\boldsymbol{I}_t$)

$\boldsymbol{T}_t \leftarrow$ 预测与补偿($\boldsymbol{T}_{t-1}^{\mathrm{A}} \cup \boldsymbol{T}_{t-1}^{\mathrm{L}}$)

$(\boldsymbol{M}_t, \boldsymbol{U}_t^{\mathrm{T}}, \boldsymbol{U}_t^{\mathrm{D}}) \leftarrow$ 两阶段关联($\boldsymbol{T}_t, \boldsymbol{D}_t$)

遍历匹配对$(\tau_i, d_j) \in \boldsymbol{M}_t$，执行以下更新

$s_j \leftarrow$ 检测置信度(d_j)

$\boldsymbol{R}_j \leftarrow [1 + \kappa(1 - s_j)]\boldsymbol{R}$

$\tau_i \leftarrow$ 卡尔曼校正($\tau_i, d_j, \boldsymbol{R}_j$)(模块B)

若外观特征$\boldsymbol{f}_j$对检测d_j可用，则

$a_j \leftarrow$ 按式(1)计算框面积比例

$q_j \leftarrow$ 按式(1)计算外观更新代理

$\bar{\boldsymbol{f}}_{i,t} \leftarrow$ 按式(2)和式(3)更新模板(模块A)

结束条件判断

$\bar{s}_{i,t} \leftarrow \rho \bar{s}_{i,t-1} + (1-\rho)s_j$

结束遍历

$(\boldsymbol{T}_t^{\mathrm{A}}, \boldsymbol{T}_t^{\mathrm{L}}, \boldsymbol{T}_t^{\mathrm{R}}) \leftarrow$ 维护轨迹集合$(\boldsymbol{M}_t, \boldsymbol{U}_t^{\mathrm{T}}, \boldsymbol{U}_t^{\mathrm{D}}, \boldsymbol{T}_{t-1}^{\mathrm{R}})$

遍历丢失轨迹$\tau_i \in \boldsymbol{T}_t^{\mathrm{L}}$，执行保留检查

$T_i^{\mathrm{eff}} \leftarrow$ 按式(8)计算有效保留期(模块C)

若丢失帧数$(\tau_i) > T_i^{\mathrm{eff}}$，则

将轨迹τ_i从$\boldsymbol{T}_t^{\mathrm{L}}$移至$\boldsymbol{T}_t^{\mathrm{R}}$

结束条件判断

结束遍历

返回$\boldsymbol{T}_t^{\mathrm{A}}, \boldsymbol{T}_t^{\mathrm{L}}, \boldsymbol{T}_t^{\mathrm{R}}$

4 实　验

4.1 数据集、指标与实现设置

主实验采用MOTChallenge的MOT17和MOT20行人跟踪基准及官方协议(Dendorfer等，2020，2021)。指标为高阶跟踪准确率(higher order tracking accuracy，HOTA)及其关联准确率(association accuracy，AssA)(Luiten等，2021)、身份F1分数(identification F1 score，IDF1)(Ristani等，2016)、多目标跟踪准确率(multiple object tracking accuracy，MOTA)、身份切换次数(identity switches，IDSW)、误检数(false positives，FP)和漏检数(false negatives，FN)。前四项越高越好，后三项越低越好。

MOT17测试集与验证集分别采用bytetrack_x_mot17. pth. tar和bytetrack_ablation. pth. tar，MOT20采用bytetrack_x_mot20. tar，各自配套FastReID权重(He等，2023)。同配置比较固定检测与ReID权重、关联和序列参数，采用16位浮点(16-bit floating point，FP16)与批量大小1。MOT17和MOT20默认输入为800×1440和896×1600像素，MOT20-06和MOT20-08为736×1920像素；MOT17测试集的高分、低分和新轨迹阈值为0. 65、0. 10和0. 75。

两个验证集由训练序列后半段构成；MOT17的7段视频分别按可变形部件模型(deformable part model，DPM)、FRCNN和尺度依赖池化(scale-dependent pooling，SDP)检测条目展开，MOT20包含4段视频。验证集用于参数选择，由TrackEval评测；测试集仅用于最终评测。两种方法均在序列结束后采用相同线性插值，不计入逐帧跟踪时延。

准确性实验采用Ubuntu 18. 04、Python 3. 8、PyTorch 1. 8. 1、统一计算设备架构(Compute Unified Device Architecture，CUDA) 11. 1、RTX 4090(24 GB)、16核Intel Xeon Platinum 8352V及90 GB内存。效率测量配置见第4. 7节。

同配置复现的BoT-SORT-ReID是直接比较基线，MOT17和MOT20测试集结果由官方服务器评测。其他方法的公开结果用于展示基准水平，不作为模块增量的依据。

MOT17、MOT20数据按知识共享署名－非商业性使用－相同方式共享3. 0许可(Creative Commons Attribution-NonCommercial-ShareAlike 3. 0，CC BY-NC-SA 3. 0)及第三方序列原许可使用。

4.2 与代表性方法的比较

表1列出私有检测协议下的测试集结果。公开结果可能采用不同的检测器、训练数据、外观模型和后处理，方法的直接增量仅由复现基线与RA-BoT-SORT的同配置结果计算。McByte未公开报告MOT20结果，故仅列MOT17。

MOT17测试集中，HOTA、IDF1和AssA分别提高0. 3、0. 6和0. 5个百分点，IDSW由1358降至1122(减少17. 4%)。FP由26651降至22509，FN由81870增至85758，表明身份保持改善伴随检测误差构成变化。

MOT20测试集中，HOTA、IDF1和AssA分别提

© 中国图象图形学报版权所有

表 1　MOT17测试集跟踪性能

Table 1　Tracking performance on the MOT17 test set

方法	HOTA/%↑	MOTA/%↑	IDF1/%↑	AssA/%↑	IDSW↓	FP↓	FN↓
MOT17-test							
ByteTrack(Zhang等,2022)	63.1	80.3	77.3	62.0	2196	25491	83721
OC-SORT(Cao等,2023)	63.2	78.0	77.5	63.4	1950	**15129**	107055
StrongSORT++*(Du等,2023)	64.4	79.6	79.5	64.4	1194	27876	86205
BoT-SORT-ReID(作者公开)(Aharon等,2022)	65.0	80.5	80.2	65.5	1212	22521	86037
ConfTrack(Jung等,2024)	**65.4**	80.0	**81.2**	**66.3**	1155	42411	**69411**
DiffMOT(Lv等,2024)	64.5	79.8	79.3	64.6	—	—	—
PuTR(Liu等,2025)	62.1	78.8	75.6	60.5	—	—	—
McByte(Stanczyk等,2025)	64.2	80.2	79.4	—	—	—	—
PlugTrack - DiffMOT(Kim等,2026)	64.2	79.2	79.0	64.4	2361	—	—
OA-SORT(Li等,2026)	64.2	79.6	79.1	64.0	—	—	—
BoT-SORT-ReID(同配置复现)(Aharon等,2022)	64.8	80.5	79.9	65.0	1358	26651	81870
RA-BoT-SORT	65.1	**80.6**	80.5	65.5	**1122**	22509	85758

注:↑、↓分别表示越高、越低越好;各列粗体表示最优(含并列)。"—"表示未报告或口径未能核实;"*"表示采用离线AFLink和GSI;约数以"≈"标记。公开值和复现值不可混同。

表 1　(续) MOT20测试集跟踪性能

Table 1　(continued) Tracking performance on the MOT20 test set

方法	HOTA/%↑	MOTA/%↑	IDF1/%↑	AssA/%↑	IDSW↓	FP↓	FN↓
MOT20-test							
ByteTrack(Zhang等,2022)	61.3	**77.8**	75.2	59.6	1223	26249	**87594**
OC-SORT(Cao等,2023)	62.4	75.7	76.3	62.5	942	19067	105894
StrongSORT++*(Du等,2023)	62.6	73.8	77.0	64.0	770	**16632**	117920
BoT-SORT-ReID(作者公开)(Aharon等,2022)	63.3	**77.8**	77.5	62.9	1313	24638	88863
ConfTrack(Jung等,2024)	**64.8**	77.2	**80.2**	**66.2**	**702**	27183	90313
DiffMOT(Lv等,2024)	61.7	76.7	74.9	60.5	—	—	—
PuTR(Liu等,2025)	61.4	75.6	74.6	60.4	—	—	—
PlugTrack - DiffMOT(Kim等,2026)	61.8	76.4	75.2	60.9	1549	—	—
OA-SORT(Li等,2026)	62.6	76.5	76.6	62.1	1274	≈35900	≈84600
BoT-SORT-ReID(同配置复现)(Aharon等,2022)	62.7	**77.8**	76.4	61.7	1385	24447	89057
RA-BoT-SORT	62.9	**77.8**	76.7	62.1	1346	24394	88997

注:↑、↓分别表示越高、越低越好;各列粗体表示最优(含并列)。"—"表示未报告或口径未能核实;"*"表示采用离线AFLink和GSI;"≈"表示近似值,不参与精确排序。作者公开值与复现值口径可能不同;OC-SORT取官方MOT-5230-10和MOT-5264-13记录,PlugTrack - DiffMOT的IDSW取原文表S3。

高0.2、0.3和0.4个百分点,IDSW由1385降至1346　(减少2.8%),MOTA保持77.8%。身份保持改善的

© 中国图象图形学报版权所有

幅度在两个数据集上不同。

图 2 取表 1 的 MOT17 测试集数据。横轴为 IDF1，纵轴为 MOTA，气泡面积与 HOTA 成正比。空心气泡表示作者公开的 BoT-SORT-ReID 结果(1)或含离线后处理的 StrongSORT++；蓝色实心气泡表示同配置复现基线(2)，箭头由该基线指向 RA-BoT-SORT。箭头仅比较这两组结果，不代表与其他公开方法进行受控比较。

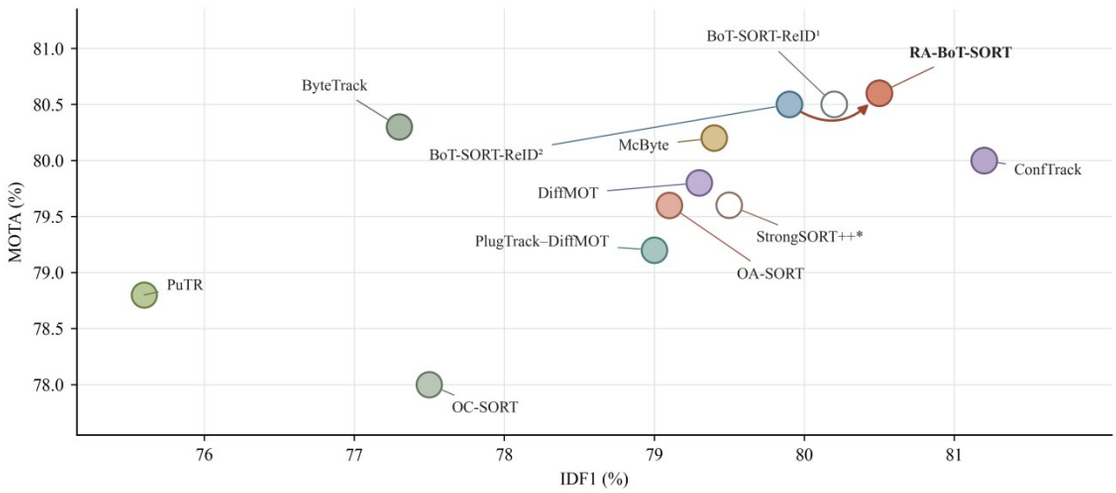


图 2　MOT17 测试集跟踪性能

Fig. 2　Tracking performance on the MOT17 test set

4.3　模块组合消融分析

表 2 比较固定基线、单模块及模块组合。A、B、C 分别表示 CRAU、LBCA-KC 和 RATR；同一数据集仅改变模块的启用状态。

完整方法在 MOT17 和 MOT20 验证集上的 AssA 分别由 71.258% 提高至 71.320% 和由 74.604% 提高至 74.906%，IDSW 分别减少 32 和 28 次。三个单模块在两个验证集上均提高 AssA 并减少 IDSW，但其他指标并非一致改善。

在 A+B 上加入 C 后，MOT17 的 IDSW 由 469 降至 464，但 AssA 由 71.352% 降至 71.320%；MOT20 的 IDSW 由 679 降至 667，AssA 由 74.701% 升至 74.906%。完整配置在两个验证集上均取得最低 IDSW，而最高 AssA 分别由 A+C 和 B+C 取得，体现状态调节之间的指标取舍，不构成统计协同证据。

按原始视频比较基线与完整方法：MOT17 验证集先对同一视频的三个检测条目求均值，7 段视频中 HOTA、IDF1 和 AssA 均为 5 段改善、2 段下降（MOT17-02、MOT17-05）；IDSW 为 4 段减少、3 段持平。MOT20 验证集 4 段视频的前三项均改善，IDSW 为 3 段减少、1 段持平。该分析描述固定划分内的差异，不替代总体评测或支持显著性推断。

4.4　CRAU 可靠性线索的受控比较

保持 LBCA-KC 和 RATR 不变，表 3 比较固定 EMA、仅置信度、仅面积及复合输入四种设置。

相较固定 EMA，复合输入改善 MOT17 验证集的七项指标；在 MOT20 验证集上改善 MOTA、IDF1、IDSW、FP 和 FN，但 HOTA 和 AssA 下降。仅使用置信度在 MOT17 取得最高 HOTA、AssA 及最低 IDSW；仅面积使两个验证集的 IDF1 和 AssA 均低于固定 EMA。复合输入是多指标取舍，并不等同于更准确的外观质量估计。

4.5　与 NSA-KF 的同配置对照

表 4 在相同的跟踪配置和置信度输入下比较原始卡尔曼滤波、NSA-KF 和 LBCA-KC。仅改变测量噪声映射，LBCA-KC 采用式(4)且噪声调节系数为 1。

与 NSA-KF 相比，LBCA-KC 在 MOT17、MOT20 验证集的 AssA 分别提高 0.622 和 0.473 个百分点，IDSW 分别减少 111 和 41 次。与原始卡尔曼滤波相比，LBCA-KC 的 AssA 在两组中均提高，但 MOT17 的 IDF1 略降。

MOT20-03 上，NSA-KF 的 HOTA 与 IDF1 分别高 0.164 和 0.242 个百分点，且 IDSW 少 5 次；LBCA-KC

© 中国图象图形学报版权所有

表 2　模块消融结果

Table 2　Module ablation results

数据集	配置	HOTA/% ↑	MOTA/% ↑	IDF1/% ↑	AssA/% ↑	IDSW ↓	FP ↓	FN ↓
MOT17-val	Baseline	68.269	77.159	80.918	71.258	496	12053	24482
	A	68.345	77.333	80.989	71.307	481	11879	24389
	B	68.313	77.361	80.896	71.315	481	11795	24428
	C	68.321	77.188	81.057	71.342	491	12038	24455
	A+B	68.387	77.542	80.960	71.352	469	11624	24317
	A+C	**68.389**	77.355	**81.135**	**71.381**	479	11867	24368
	B+C	68.327	77.407	80.915	71.305	470	11762	24397
	A+B+C	68.386	**77.579**	80.983	71.320	**464**	**11594**	**24292**
MOT20-val	Baseline	77.259	93.842	89.789	74.604	695	9840	27374
	A	77.358	93.850	89.993	74.775	677	9797	27385
	B	77.350	93.871	89.948	74.749	682	9702	27347
	C	77.339	93.840	89.929	74.748	689	9853	27377
	A+B	77.339	93.886	89.967	74.701	679	9689	**27267**
	A+C	77.328	93.850	89.929	74.705	689	9834	27336
	B+C	**77.466**	93.876	90.142	**74.962**	671	9694	27336
	A+B+C	77.446	**93.888**	**90.146**	74.906	**667**	**9688**	27272

注：↑、↓分别表示越高、越低越好；每个数据集内粗体表示最优（含并列）。A、B、C分别表示CRAU、LBCA-KC和RATR。

表 3　外观更新线索比较

Table 3　Comparison of appearance update cues

数据集	外观更新	HOTA/% ↑	MOTA/% ↑	IDF1/% ↑	AssA/% ↑	IDSW ↓	FP ↓	FN ↓
MOT17-val	固定EMA	68.327	77.407	80.915	71.305	470	11762	24397
	仅置信度	**68.502**	77.522	80.956	**71.678**	**461**	11660	24322
	仅面积	68.235	77.561	80.801	71.011	472	11599	24308
	复合可靠性	68.386	**77.579**	**80.983**	71.320	464	**11594**	**24292**
MOT20-val	固定EMA	**77.466**	93.876	90.142	**74.962**	671	9694	27336
	仅置信度	77.432	93.881	90.083	74.893	671	**9669**	27327
	仅面积	77.405	93.863	89.999	74.829	675	9734	27369
	复合可靠性	77.446	**93.888**	**90.146**	74.906	**667**	9688	**27272**

注：↑、↓分别表示越高、越低越好；每个数据集内粗体表示最优（含并列）。固定EMA表示关闭CRAU调节。

则减少69个FP和30个FN。保留下界并非在所有序列中均优；当预测偏差较大而当前测量准确时，较保守的校正也可能延缓恢复。

4.6　参数敏感性

在MOT17验证集上分别扫描三个调节参数，每次仅改变一项。图3和图4分别给出外观更新系数和轨迹保留倍率上界的敏感性结果。

外观系数取0.01时HOTA最高（68.492%），取0.03时其余五项指标更优。继续增至0.10时，HOTA降至68.098%，IDF1降至80.230%，IDSW增至522次。主配置采用0.03。

表5比较噪声调节系数取1、2和3时的表现，其余设置不变。

© 中国图象图形学报版权所有

表 4　卡尔曼校正策略比较

Table 4　Comparison of Kalman correction strategies

数据集	方法	HOTA/% ↑	MOTA/% ↑	IDF1/% ↑	AssA/% ↑	IDSW ↓	FP ↓	FN ↓
MOT17-val	原始卡尔曼滤波	68.269	77.159	**80.918**	71.258	496	12053	24482
	NSA-KF(Du等,2021)	67.867	76.832	80.435	70.693	592	12032	24938
	LBCA-KC	**68.313**	**77.361**	80.896	**71.315**	**481**	**11795**	**24428**
MOT20-val	原始卡尔曼滤波	77.259	93.842	89.789	74.604	695	9840	27374
	NSA-KF(Du等,2021)	77.041	93.807	89.508	74.276	723	9878	27519
	LBCA-KC	**77.350**	**93.871**	**89.948**	**74.749**	**682**	**9702**	**27347**

注：↑、↓分别表示越高、越低越好；每个数据集内粗体表示最优(含并列)。各组仅改变测量校正策略。

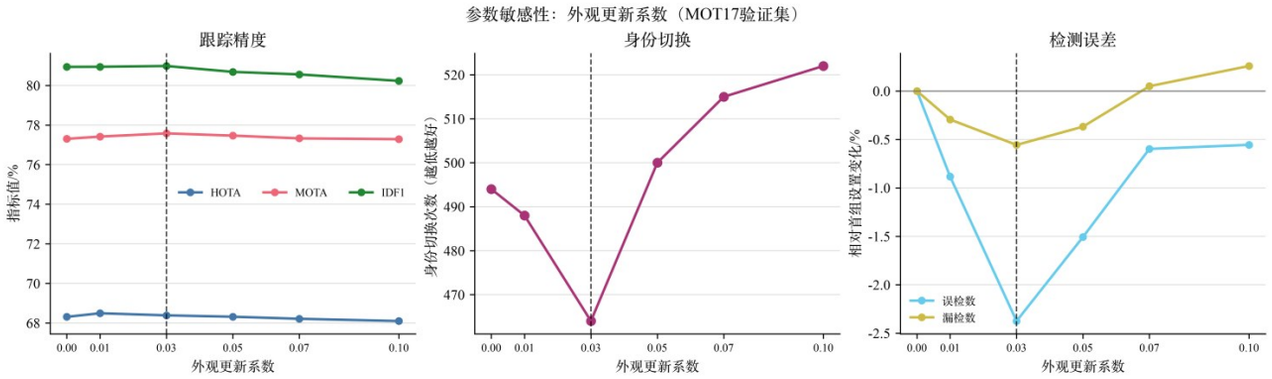


图 3　外观更新系数敏感性

Fig. 3　Sensitivity to the appearance-update coefficient

表 5　卡尔曼噪声系数敏感性

Table 5　Sensitivity to the Kalman noise coefficient

数据集	κ	HOTA/% ↑	MOTA/% ↑	IDF1/% ↑	IDSW ↓	FP ↓	FN ↓
MOT17-val	1	68.386	**77.579**	80.983	464	**11594**	24292
	2	68.369	77.414	81.144	464	11843	24310
	3	**68.404**	77.387	**81.207**	**452**	11942	**24268**
MOT20-val	1	77.446	**93.888**	90.146	**667**	**9688**	27272
	2	77.386	93.862	89.989	682	9808	27293
	3	**77.576**	93.887	**90.272**	**667**	9728	**27236**

注：↑、↓分别表示越高、越低越好；每个数据集内粗体表示最优(含并列)。系数为1时对应表2完整配置。

与系数取1相比，取3在两个验证集上均提高HOTA和IDF1、降低FN，但增加FP且降低MOTA。主配置取1，将测量噪声限制在基线的1～2倍。这些结果用于固定验证划分内的参数选择，不说明全局最优。

保留倍率上界取1.3时，所测六项指标均最优；取1.5、1.7或2.0时结果相同。主配置取1.3。

4.7　复杂度与运行效率

CRAU调整外观EMA权重，LBCA-KC缩放测量协方差，RATR增加一个历史置信度标量并据此计算保留期。三个模块不增加网络或训练，关联求解器不变；轨迹保留策略可能改变候选数量及后续计算量。

© 中国图象图形学报版权所有

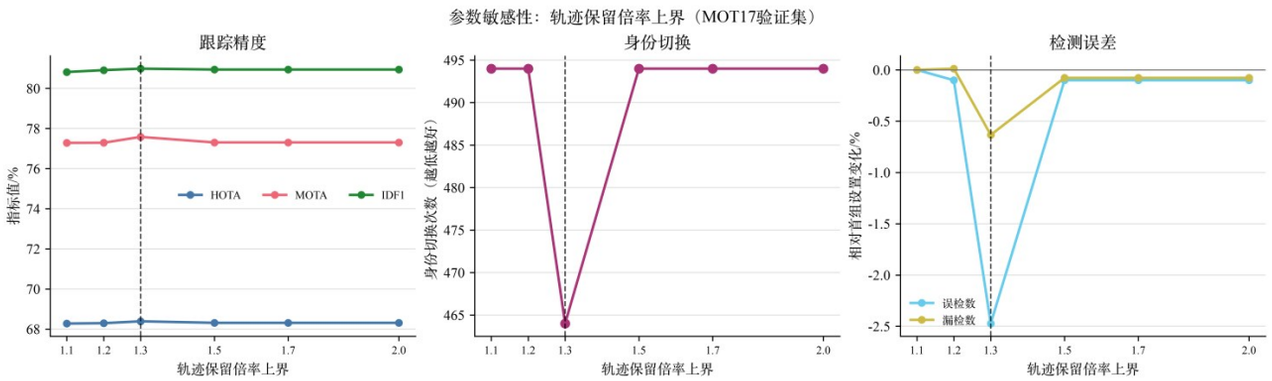


图 4　轨迹保留倍率上界敏感性

Fig. 4　Sensitivity to the upper retention multiplier

表 6　各配置的时延与配对差值

Table 6　Latency and paired differences by configuration

配置	端到端/(ms/帧)↓	Δ端到端/(ms/帧)	仅跟踪/(ms/帧)↓	Δ仅跟踪/(ms/帧)	峰值显存/MiB
基线	215.890±3.360	—	**72.809　1.183**	—	4340.356
基线+A	216.261±2.774	+0.371±0.652	73.032±0.520	+0.223±0.672	4340.356
基线+B	**215.780　1.517**	−0.110±2.161	72.928±0.403	+0.119±0.866	4340.356
基线+C	217.620±0.754	+1.730±3.323	73.435±0.348	+0.626±1.358	4340.356
基线+A+B+C	215.881±2.386	−0.010±0.986	73.204±0.670	+0.396±0.516	4340.356

注：A、B、C分别为CRAU、LBCA-KC和RATR；↓表示时延越小越好。数值为3次运行的均值±样本标准差；Δ为同轮相对基线的差，正值表示时延增加；“—”表示基线无需计算差值。绝对时延列中粗体为最小均值，不表示统计显著。

在配备NVIDIA GeForce RTX 4090的同一服务器上，以MOT17验证集的7个FRCNN序列测试基线、三个单模块和完整方法。每配置独立运行3次并按轮次交错执行；每次处理2652帧，各序列前30帧预热，统计2442帧。

软件环境为Python 3.8.10、PyTorch 1.8.1+cu111及CUDA 11.1。各组固定YOLOX-X检测权重、MOT17 ReID权重、800×1440像素输入、预计算GMC、关联参数、批量大小1、FP16和卷积与批归一化融合。采用单线程、随机种子0，关闭自动调优，无检测帧也调用跟踪更新。

计时采用CUDA同步后的墙钟时间。端到端范围包含图像读取与解码、预处理、检测推理、ReID、GMC状态补偿、关联、轨迹更新及输出筛选与内存整理，并计入上述过程的中央处理器（central processing unit，CPU）与图形处理器（graphics processing unit，GPU）间的数据传输；不包含模型初始化、可视化、磁盘写入或离线插值。GMC采用预计算仿射变换，其离线估计不计时。仅跟踪（Tracker-only）统计BoTSORT. update()及其结束同步，包含接口内的ReID、GMC、关联和轨迹管理，不含检测器推理与接口外的输出整理。

表6显示，在上述配置下，完整方法仅跟踪平均时延较基线增加0.396 ms/帧（0.54%）；端到端速度约4.63帧/s，不满足30帧/s逐帧实时处理要求。两组峰值张量分配显存均为4340.356 MiB（1 MiB=1024^2字节）。

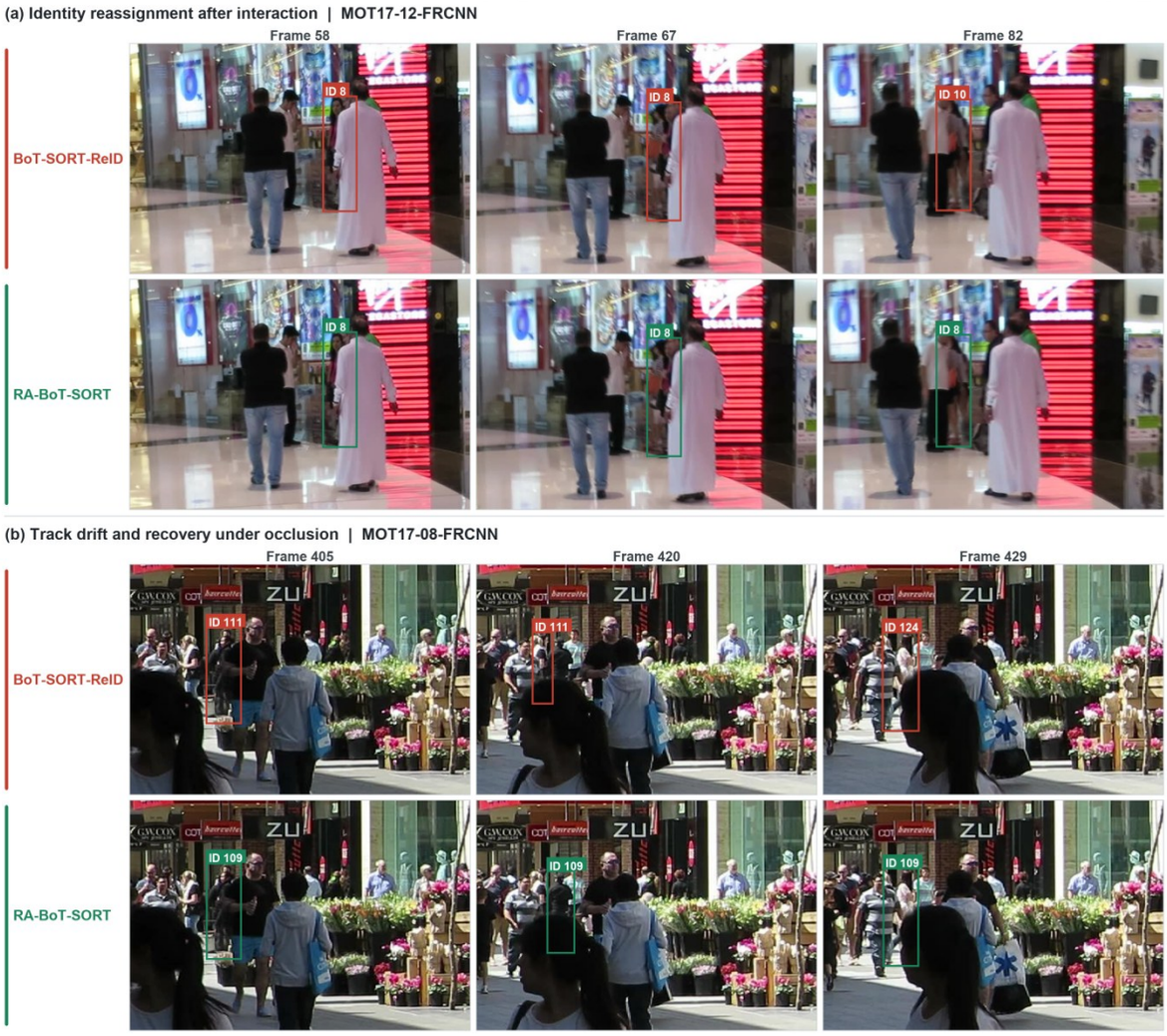
仅启用A、B或C时，仅跟踪时延相对基线的配对平均增量分别为0.223、0.119和0.626 ms/帧，约占基线时延的0.31%、0.16%和0.86%。

单模块的配对差值是配置级净开销，含候选轨迹规模变化，不能相加估计完整方法成本。完整方法的端到端配对差值为−0.010±0.986 ms/帧，不解释为加速或零开销。三次重复仅反映既定平台、配置和序列的运行波动，不保证在其他硬件、检测器或目标密度下有相同效率。

© 中国图象图形学报版权所有

4.8 定性结果

图5展示MOT17测试集中的交互和遮挡片段。基线与本文方法使用相同原始帧和裁剪范围，比较同一方法在不同时刻的身份编号，不横向比较不同方法分配的编号。



(a)交互后的身份变化；(b)遮挡时的轨迹恢复

(a) Identity reassignment after interaction; (b) Track recovery under occlusion.

图5 交互与遮挡场景跟踪结果

Fig. 5 Tracking results in interaction and occlusion scenes

MOT17-12-FRCNN第58、67和82帧中，基线将目标由8号改为10号，本文方法保持8号。MOT17-08-FRCNN第405、420和429帧中，基线的111号轨迹漂移后以124号恢复，本文方法在所示时刻保持109号。两例仅展示特定片段，不能单独区分模块贡献。

5 讨论与局限

5.1 结果分析与适用范围

同配置测试中，两个数据集的AssA、IDF1和HOTA均提高，IDSW均减少，但改善幅度不同。HOTA同时考虑检测与关联，不能由身份切换次数的下降直接推导同等幅度的HOTA增益。表7从原始视频层面补充验证集上的配对结果。

MOT17包含多种视角、运动和交互条件，

© 中国图象图形学报版权所有

表 7　验证集逐视频指标变化

Table 7　Paired changes by validation video

原始视频	ΔHOTA	ΔIDF1	ΔAssA	ΔIDSW（均值）
MOT17-02	−1.262	−0.708	−2.841	−2.000
MOT17-04	+0.393	+0.317	+0.791	0.000
MOT17-05	−0.222	−0.520	−0.483	−1.667
MOT17-09	+1.642	+0.454	+2.509	−5.000
MOT17-10	+0.077	+0.147	+0.084	0.000
MOT17-11	+0.054	+0.009	+0.043	0.000
MOT17-13	+0.175	+0.392	+0.122	−2.000
MOT20-01	+0.725	+0.920	+1.326	−3.000
MOT20-02	+0.708	+1.220	+0.808	−21.000
MOT20-03	+0.060	+0.110	+0.094	−4.000
MOT20-05	+0.114	+0.244	+0.273	0.000

注：Δ=完整方法−同配置基线；HOTA、IDF1、AssA为百分点，IDSW为每条目平均次数变化，负值表示减少。MOT17先在同视频的DPM、FRCNN、SDP三个条目内取算术平均，故IDSW增量可为小数；MOT20每视频仅一个条目。两组均采用相同线性插值，不以视频均值替代TrackEval汇总。

MOT20突出高密度人群及持续遮挡。表7显示，MOT17逐视频HOTA增量介于−1.262至+1.642个百分点。MOT17-02的HOTA、IDF1和AssA分别下降1.262、0.708和2.841个百分点，但IDSW每条目平均减少2次；切换次数因而不能替代整体关联质量。MOT20的HOTA增量介于+0.060至+0.725个百分点，减少的28次IDSW中21次来自MOT20-02，占75.0%，说明改善并非均匀分布。保留轨迹不能补回缺失检测，恒速预测和相似外观也可能限制重关联，但这些仅为机制上的可能解释，不能据此将测试集收益差异归因于某一种场景因素。

5.2　DanceTrack验证集对照

DanceTrack以相似外观、复杂运动和频繁位置交换为特征（Sun等，2022）。验证集的同配置对照覆盖25个序列、25 508帧。基线与完整方法使用相同的YOLOX-X检测权重、公开发布的DanceTrack专用ReID权重、跟踪参数、相机运动补偿、线性插值参数及评测设置。检测准确率（detection accuracy，DetA）与其余四项指标见表8。专用ReID权重的实际训练划分未得到直接确认，因此该对照不作为已确认训练与验证划分独立的泛化证据。

表 8　DanceTrack验证集跟踪性能

Table 8　Tracking performance on the DanceTrack validation set

方法	HOTA/% ↑	DetA/% ↑	AssA/% ↑	IDF1/% ↑	MOTA/% ↑
BoT-SORT基线	**56.497**	**76.523**	**41.905**	**57.145**	**88.008**
RA-BoT-SORT	55.727	76.292	40.897	55.925	88.007
差值/百分点	−0.770	−0.231	−1.008	−1.220	−0.001

注：↑表示越高越好，粗体表示两种方法中的最优值；差值为RA-BoT-SORT减基线。

完整方法的HOTA、AssA和IDF1分别降低0.770、1.008和1.220个百分点，DetA降低0.231个百分点，MOTA仅相差0.001个百分点。DetA与MOTA变化较小，而AssA和IDF1下降，说明MOT17和MOT20的身份保持收益未延续到该配置。相似外观可能减弱检测分数及面积与特征可分性的对应关系，保守更新也可能延缓对快速运动变化的适应；这些解释尚需验证，总体指标不能确定具体原因或将其归于单个模块。

5.3　可靠性代理的局限

检测分数、框面积和匹配历史不等同于定位精度、特征质量或身份正确率。三类调节对代理失准的响应有所不同。

CRAU中，退化观测被高估会提高当前特征的写入权重，可能污染模板；清晰小目标被低估则可能延缓外观适应。0.90—0.93的历史权重将当前特征的写入系数限制在0.07—0.10，但不识别全部退化观测，也不直接限制特征归一化后的误差。

LBCA-KC中，高分但偏移的框仍可能使运动后验偏离真实状态；基线协方差下界只避免因高分进一步压低噪声，不能排除定位错误。准确但低分的观测可能被过度降权，延缓对真实运动变化的校正。

RATR中，错误匹配若积累高分历史，可能延长错误身份的保留；真实轨迹若持续低分，也可能较早

© 中国图象图形学报版权所有

移除。历史EMA减弱单次波动，但不能消除持续偏差；丢失时冻结历史量避免以缺失观测直接更新分数，却不构成身份正确性的重新验证。

在其余输入、预测状态和匹配固定时，若当前检测置信度的变化幅度不超过0.1，三种映射的单次标量响应分别不超过0.0015（外观历史权重）、0.1（测量噪声倍率）和0.01（历史分数及取整前保留倍率）。该解析示例不构成质量校准，也不约束多帧误差、实际轨迹保留或身份错误。

6 结 论

RA-BoT-SORT在BoT-SORT既有关联流程上，对外观记忆、运动校正和轨迹保留分别设置与状态相关的输入、更新边界及触发条件。具体改进是分数与尺度线索驱动的窄范围外观写入、保留基线协方差下界的匹配后校正，以及由成功匹配历史决定的有界保留期。贡献在于上述策略的具体设计与组合，而非新的观测质量估计机制。

同配置对照支持该方法在MOT17和MOT20所测行人跟踪条件下改善身份保持，但该收益未在DanceTrack验证集上延续。在RTX 4090和所测MOT17工作负载下，仅跟踪平均时延较基线增加0.396 ms/帧（0.54%），端到端速度约4.63帧/s，不满足30帧/s逐帧实时处理。较小额外开销不等于系统实时能力，效率结论限于所测平台和配置。

参考文献（References）

Aharon N, Orfaig R and Bobrovsky B Z. 2022. BoT-SORT: robust associations multi-pedestrian tracking[EB/OL]. (2022-07-07)[2026-10-30]. https://arxiv.org/abs/2206.14651v2 [DOI: 10.48550/arXiv.2206.14651]

Bewley A, Ge Z, Ott L, Ramos F and Upcroft B. 2016. Simple online and realtime tracking//Proceedings of 2016 IEEE International Conference on Image Processing. Phoenix, USA: IEEE: 3464-3468 [DOI: 10.1109/ICIP.2016.7533003]

Cao J, Pang J, Weng X, Khirodkar R and Kitani K. 2023. Observation-centric SORT: rethinking SORT for robust multi-object tracking//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, Canada: IEEE: 9686-9696 [DOI: 10.1109/CVPR52729.2023.00934]

Dendorfer P, Rezatofighi H, Milan A, Shi J, Cremers D, Reid I, et al. 2020. MOT20: a benchmark for multi object tracking in crowded scenes[EB/OL]. (2020-03-19)[2026-10-30]. https://arxiv.org/abs/2003.09003 [DOI: 10.48550/arXiv.2003.09003]

Dendorfer P, Ošep A, Milan A, Schindler K, Cremers D, Reid I, et al. 2021. MOTChallenge: a benchmark for single-camera multiple target tracking. International Journal of Computer Vision, 129(4): 845-881 [DOI: 10.1007/s11263-020-01393-0]

Du Y H, Wan J F, Zhao Y Y, Zhang B Y, Tong Z H and Dong J H. 2021. GIAOTracker: a comprehensive framework for MCMOT with global information and optimizing strategies in VisDrone 2021//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision Workshops. Montreal, Canada: IEEE: 2809-2819 [DOI: 10.1109/ICCVW54120.2021.00315]

Du Y H, Zhao Z C, Song Y, Zhao Y Y, Su F, Gong T, et al. 2023. StrongSORT: make DeepSORT great again. IEEE Transactions on Multimedia, 25: 8725-8737 [DOI: 10.1109/TMM.2023.3240881]

Guo W, Liu Q G and Ding X M. 2024. Multi-object tracking using adaptive-IoU loss and hierarchical association. Journal of Image and Graphics, 29(7): 1970-1983 (郭文, 刘其贵, 丁昕苗. 2024. 自适应IoU损失和层级关联的多目标跟踪. 中国图象图形学报, 29(7): 1970-1983) [DOI: 10.11834/jig.230390]

Han Q, Li L F and Min W D. 2026. Research progress on object re-identification in special scenarios. Journal of Image and Graphics, 31(6): 1795-1822 (韩清, 李龙飞, 闵卫东. 2026. 特殊场景下的目标重识别研究进展. 中国图象图形学报, 31(6): 1795-1822) [DOI: 10.11834/jig.260117]

He L X, Liao X Y, Liu W, Liu X C, Cheng P and Mei T. 2023. FastReID: a PyTorch toolbox for general instance re-identification//Proceedings of the 31st ACM International Conference on Multimedia. Ottawa, Canada: ACM: 9664-9667 [DOI: 10.1145/3581783.3613460]

Jung H C, Kang S J, Kim T G and Kim H K. 2024. ConfTrack: Kalman filter-based multi-person tracking by utilizing confidence score of detection box//Proceedings of 2024 IEEE/CVF Winter Conference on Applications of Computer Vision. Waikoloa, USA: IEEE: 6569-6578 [DOI: 10.1109/WACV57701.2024.00645]

Kim S J, Lee S J and Cho M A. 2026. PlugTrack: multi-perceptive motion analysis for adaptive fusion in multi-object tracking//Proceedings of the 40th AAAI Conference on Artificial Intelligence. Singapore: AAAI Press: 5709-5716 [DOI: 10.1609/aaai.v40i7.37491]

Li C J, Ma J B, Shen L, Chen Y R and Chen L Y. 2026. Occlusion-aware SORT: observing occlusion for robust multi-object tracking [EB/OL]. (2026-03-09)[2026-10-30]. https://arxiv.org/abs/2603.06034v2 [DOI: 10.48550/arXiv.2603.06034]

Liu C W, Li H J, Wang Z H and Xu R. 2025. Is a pure transformer effective for separated and online multi-object tracking? ACM

© 中国图象图形学报版权所有

Transactions on Multimedia Computing, Communications, and Applications, 21(12): 343:1-343:21 [DOI: 10.1145/3749105]

Luiten J, Ošep A, Dendorfer P, Torr P, Geiger A, Leal-Taixé L, et al. 2021. HOTA: a higher order metric for evaluating multi-object tracking. International Journal of Computer Vision, 129(2): 548-578 [DOI: 10.1007/s11263-020-01375-2]

Lv W Y, Huang Y H, Zhang N, Lin R S, Han M and Zeng D. 2024. DiffMOT: a real-time diffusion-based multiple object tracker with non-linear prediction//Proceedings of 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 19321-19330 [DOI: 10.1109/CVPR52733.2024.01828]

Maggiolino G, Ahmad A, Cao J and Kitani K. 2023. Deep OC-SORT: multi-pedestrian tracking by adaptive re-identification//Proceedings of 2023 IEEE International Conference on Image Processing. Kuala Lumpur, Malaysia: IEEE: 3025-3029 [DOI: 10.1109/ICIP49359.2023.10222576]

Ristani E, Solera F, Zou R, Cucchiara R and Tomasi C. 2016. Performance measures and a data set for multi-target, multi-camera tracking//Computer Vision-ECCV 2016 Workshops. Amsterdam, The Netherlands: Springer: 17-35 [DOI: 10.1007/978-3-319-48881-3_2]

Stanczyk T, Yoon S and Bremond F. 2025. No train yet gain: towards generic multi-object tracking in sports and beyond//Proceedings of 2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Nashville, USA: IEEE: 6029-6038 [DOI: 10.1109/CVPRW67362.2025.00601]

Sun P Z, Cao J K, Jiang Y, Yuan Z H, Bai S, Kitani K, et al. 2022. DanceTrack: multi-object tracking in uniform appearance and diverse motion//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 20961-20970 [DOI: 10.1109/CVPR52688.2022.02032]

Tang S, Yang Y and Zhao Q J. 2026. Learning dynamic view reliability for multi-camera multi-object tracking. IEEE Transactions on Circuits and Systems for Video Technology, 36(10): 16036-16050 [DOI: 10.1109/TCSVT.2026.3707405]

Wojke N, Bewley A and Paulus D. 2017. Simple online and realtime tracking with a deep association metric//Proceedings of 2017 IEEE International Conference on Image Processing. Beijing, China: IEEE: 3645-3649 [DOI: 10.1109/ICIP.2017.8296962]

Yang H S, Park S W, Sim C B and Jung S H. 2026. Sentinel for confidence-aware multi-object tracking. Scientific Reports, 16(1): 13571 [DOI: 10.1038/s41598-026-43938-2]

Yi K, Li J R and Zhang Y. 2025. Multi-object tracking with confidence-based trajectory prediction scheme. Sensors, 25(23): 7221 [DOI: 10.3390/s25237221]

Zhang Y F, Sun P Z, Jiang Y, Yu D D, Weng F C, Yuan Z H, et al. 2022. ByteTrack: multi-object tracking by associating every detection box//Computer Vision-ECCV 2022. Tel Aviv, Israel: Springer: 1-21 [DOI: 10.1007/978-3-031-20047-2_1]

作者简介

(Author Biographies

李娜,女,主要研究方向为计算机视觉、多目标跟踪。E-mail: lina_scu@stu.scu.edu.cn

唐松,男,主要研究方向为计算机视觉、多目标跟踪。E-mail: tangsong@stu.scu.edu.cn

© 中国图象图形学报版权所有