



引文格式: 张永泽, 达飞鹏. 膨胀区域匹配和自适应轨迹管理策略的多目标跟踪方法[J]. 武汉大学学报(信息科学版), 2024, 49(4): 572-581. DOI: 10.13203/j.whugis20230359

**Citation:** ZHANG Yongze, DA Feipeng. A Multi-object Tracking Method Based on Dilatation Region Matching and Adaptive Trajectory Management Strategy[J]. Geomatics and Information Science of Wuhan University, 2024, 49(4): 572-581. DOI: 10.13203/j.whugis20230359

# 膨胀区域匹配和自适应轨迹管理策略的多目标跟踪方法

张永泽<sup>1,2,3</sup>   达飞鹏<sup>1,2</sup>

1 东南大学自动化学院, 江苏 南京, 210018

2 东南大学复杂工程系统测量与控制教育部重点实验室, 江苏 南京, 210018

3 江苏电力信息技术有限公司, 江苏 南京, 210018

**摘要:** 针对由于目标频繁遮挡、不规则运动导致的外观特征不可靠和运动特征难以获取的问题, 提出一种基于膨胀交并比区域(dilatation intersection over union, DIOU)匹配和自适应轨迹管理策略的多目标跟踪算法。DIOU模块通过膨胀匹配区域, 提升轨迹级联匹配的精度。自适应轨迹管理策略利用目标检测置信度动态调整轨迹生命周期, 显著减少了异常跟踪和身份跳变。在公开数据集 MOT17、MOT20 和 DanceTrack 上进行了验证与测试, 其在测试集上的高阶跟踪精度平均提升了 2.4%, 实验结果证明了所提方法的有效性。

**关键词:** 计算机视觉; 多目标跟踪; 不规则运动; 数据关联; 轨迹管理

中图分类号:P208

文献标识码:A

收稿日期: 2023-09-26

DOI: 10.13203/j.whugis20230359

文章编号: 1671-8860(2024)04-0572-10

## A Multi-object Tracking Method Based on Dilatation Region Matching and Adaptive Trajectory Management Strategy

ZHANG Yongze<sup>1,2,3</sup>   DA Feipeng<sup>1,2</sup>

1 School of Automation, Southeast University, Nanjing 210018, China

2 Key Laboratory of Measurement and Control of Complex Systems of Engineering, Ministry of Education, Southeast University, Nanjing 210018, China

3 Jiangsu Electric Power Information Technology Co. Ltd, Nanjing 210018, China

**Abstract:** Objectives: Multi-object tracking (MOT) is a pivotal research area within the computer vision domain. Despite significant strides in MOT research, the field continues to grapple with formidable challenges: Indistinct appearance attributes of objects, objects exhibit irregular motion, anomalies in tracking arising from rigid trajectory lifecycle management strategies. These elements substantially undermine the precision and robustness of multi-object tracking endeavors. Methods: In response to these challenges, we present an advanced multi-object tracking algorithm that integrates dilatation intersection over union (DIOU) matching with an adaptive trajectory management approach. Initially, we introduce a metric based on a refined DIOU area for the primary matching between active trajectories and high-confidence detections, thereby improving the direct matching performance for high-quality detection boxes. Subsequently, for the re-matching of active trajectories with low-confidence detections, we implement a metric centered on a moderately dilated DIOU area, enhancing the tracking continuity of these detections. Furthermore, for reconnecting inactive trajectories with unmatched high-confidence detections, we employ a metric utilizing an extensively dilated DIOU area to bolster the probability of reactivating dormant trajectories. Lastly, an

基金项目: 国家自然科学基金(51475092); 江苏省前沿引领技术基础研究专项(BK20192004C)。

第一作者: 张永泽, 博士, 从事计算机视觉技术与多目标跟踪技术研究。820944509@qq.com

通讯作者: 达飞鹏, 博士, 教授。dafp@seu.edu.cn

adaptive trajectory management strategy predicated on detection confidence scores is deployed to dynamically modulate the lifespan of trajectories, thereby mitigating the incidence of tracking anomalies and identity switches induced by occlusions and misidentifications. **Results:** (1) The application of the DIOU-based matching framework has yielded 5.4% increase in HOTA (higher order tracking accuracy) and a 1.5% increase in MOTA (multiple object tracking accuracy) on the DanceTrack dataset, corroborating the method's efficacy in densely populated scenes and complex motion environments. (2) The implementation of the adaptive trajectory management module has further resulted in 4.6% rise in HOTA, 0.8% elevation in MOTA, and 2.1% improvement in IDF1 (identification *F*-score) on the DanceTrack dataset, demonstrating its capacity to efficiently counteract the limitations of fixed lifecycle sensitivities to false detections and missed detections. **Conclusions:** Although the refinement of data association and trajectory management strategies has led to a surge in tracking accuracy, the layering of multiple strategies has introduced a trade-off with computational efficiency, curtailing the peak performance of the tracking system.

**Key words:** computer vision; multi-object tracking; irregular motion; data association; trajectory management

多目标跟踪(multi-object tracking, MOT)是计算机视觉领域的一项重要研究课题,是自动驾驶、交通监控、电力巡检、机器人定位导航、军事侦察等领域应用的关键技术<sup>[1-6]</sup>。MOT任务旨在解析跟踪对象的行为,通过对视频中连续多帧的多个目标进行持续定位,保持跟踪目标的身份一致并生成相应运动轨迹。过去十年里,MOT问题受到了众多学者的广泛关注,形成了很多重要的研究成果。先检测后跟踪<sup>[7]</sup>(tracking-by-detection, TBD)已成为MOT任务中最具竞争力的范式之一。TBD范式包含两个子任务,即目标检测<sup>[8]</sup>和轨迹匹配<sup>[9]</sup>。轨迹匹配由两个部分组成:一是运动模型和状态估计用于预测后续帧中轨迹的目标区域,通常使用卡尔曼滤波器(Kalman filter, KF);二是将当前新的检测结果与待匹配的轨迹进行关联。目前有两种主流的方法来处理关联任务:(1)将预测框与检测框的交并比<sup>[10]</sup>(intersection over union, IOU)作为度量指标;(2)通过提取目标的外观特征来解决重识别(re-identification, ReID)<sup>[11]</sup>任务,上述两种方法都通过量化为距离指标解决全局分配问题。

2016年,Bewley等<sup>[12]</sup>将KF和匈牙利算法结合,搭建TBD框架。2017年,Wojke等<sup>[13]</sup>加入深度外观特征提取模型,使用马氏距离加权求和的级联匹配策略,减少了45%的身份跳变次数。2018年,Chen等<sup>[14]</sup>利用候选区域处理冗余的不可靠检测。2019年,Wang等<sup>[15]</sup>提出端到端的联合检测和嵌入框架,将外观模型嵌入单阶段检测器,计算成本显著降低,运行速度可达40帧/s。2022年,Zhang等<sup>[16]</sup>使用低分检测框与轨迹的相似性来恢复跟踪轨迹并过滤掉背景检测,在

MOT17数据集上实现了80.3 MOTA和63.1 HOTA的跟踪效果。

虽然MOT研究已经有了很大的发展,但仍然面临巨大挑战:(1)目标存在难以辨识外观特征的问题。一方面不同目标频繁交互时,跟踪器难以利用外观特征和位置信息区分不同目标;另一方面当目标被背景中的遮挡物遮挡时,导致目标出现短暂或长期的消失,影响目标检测准确率,同时由于外观信息不准确,导致跟踪性能急剧下降。(2)目标不规则运动引起的运动估计偏差导致目标丢失的问题。现有算法大多考虑目标规则运动的情况,当目标出现突然加速或转弯等状态,导致运动估计不准确,存在交并比等于0的情况,使目标失去匹配机会。(3)轨迹固定生命周期管理策略对误检、漏检过于敏感的问题。当前的MOT方法中基本都会采用基于计数的固定生命周期策略,但基于计数的方法对所有检测都一视同仁,未能体现检测置信度对跟踪结果的影响。

针对上述目标不规则运动和轨迹固定生命周期策略问题,本文提出了一种基于膨胀交并比区域(dilatation intersection over union, DIOU)匹配和自适应轨迹管理策略的MOT算法。首先,对于激活轨迹和高分检测首次匹配,设计了一种基于小DIOU的度量指标,提升了高分检测框直接匹配的性能;其次,对于激活轨迹和低分检测进行二次匹配,设计了一种基于中DIOU的度量指标,提高了低分检测框的追踪性能;然后,对于非激活轨迹和未匹配高分检测再次进行关联,设计了一种基于大DIOU的度量指标,提高了非激活轨迹追踪的可能性;最后,对于依然未能匹配

的轨迹,设计了一种基于检测置信度分数的自适应轨迹管理策略,动态地调整了轨迹的生命周期,减少了由于遮挡和误检导致的异常跟踪和身份跳变。本文选择 MOT17、MOT20 和 Dance-Track<sup>[17]</sup>作为验证数据集,实验结果表明,相比于现有的 MOT 算法,本文所提出的算法具有更好的精度和更强的鲁棒性。

## 1 MOT 算法

### 1.1 算法流程

本文提出的算法根据检测结果完成 MOT 任务,流程如图 1 所示,主要步骤如下:

- 1) 输入视频数据。
- 2) 使用目标检测器输出检测框,并将检测结果分为高分检测与低分检测。
- 3) 使用 KF 获取前一帧的预测框;
- 4) 采用级联匹配策略,基于 DIOU 方法,对于高分检测与激活轨迹,采用小 DIOU 计算检测框与预测框的相似度;对于低分检测与激活轨迹,采用中 DIOU 计算检测框与预测框的相似度;对于还未匹配的高分检测与非激活轨迹,采用大 DIOU 计算检测框与预测框的相似度,并使用贪婪算法获取最优匹配结果。
- 5) 使用 KF 对匹配成功的轨迹进行状态更新。
- 6) 对于未匹配成功的轨迹进行动态生命周期管理,对于小于或等于最大生命周期的轨迹,加入待匹配轨迹集合,对于超过最大生命周期的轨迹,直接删除。
- 7) 重复 2)~6),直到视频全部结束,输出目标身份(identification, ID)和目标框。

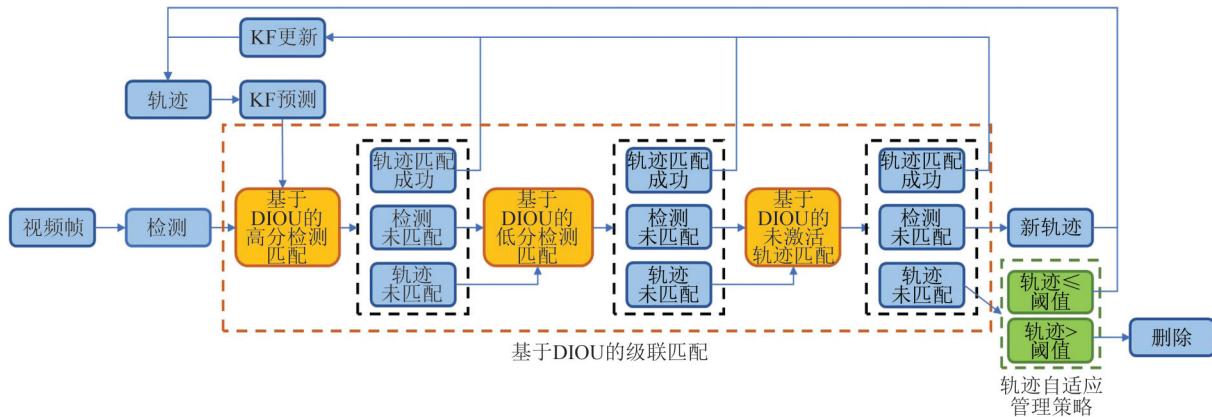


图 1 基于膨胀区域匹配和自适应轨迹管理策略的多目标算法流程图

Fig. 1 Diagram of Multi-object Tracking Algorithm Based on Dilatation Region Matching and Adaptive Trajectory Management Strategy

### 1.2 卡尔曼滤波器

卡尔曼滤波器<sup>[18]</sup>是一种高效率的递归滤波器(自回归滤波器),能够从包含噪声的测量中,估计动态系统的状态。目前,KF 算法广泛应用于 MOT 中的运动估计任务,其算法核心方程如下所示。

- 1) 系统状态方程:

$$\begin{cases} X_t = AX_{t-1} + BU_{t-1} + W_{t-1} \\ W_t \sim N(0, Q_t) \end{cases} \quad (1)$$

式中, $X_t$  表示系统状态矩阵; $A$  表示状态转移矩阵; $X_{t-1}$  表示  $t-1$  时刻的状态输出值; $B$  表示控制输入矩阵; $U_{t-1}$  表示  $t-1$  时刻的控制输入值; $W_{t-1}$  表示  $t-1$  时刻的过程噪声,且相互独立并服从正态分布,表现为期望为 0、协方差为  $Q$  的高斯白噪声。

- 2) 系统观测方程:

$$\begin{cases} Z_t = HX_t + V_t \\ V_t \sim N(0, R_t) \end{cases} \quad (2)$$

式中, $Z_t$  表示  $t$  时刻状态矩阵的观测值; $H$  为状态观测矩阵; $V_t$  表示  $t$  时刻测量噪声,且相互独立并服从正态分布,表现为期望为 0、协方差为  $R$  的高斯白噪声。

- 3) 预测过程中的系统先验估计方程:

$$\hat{x}_t = A\hat{x}_{t-1} + Bu_{t-1} \quad (3)$$

式中, $\hat{x}_t$  表示  $t$  时刻的先验状态估计值; $\hat{x}_{t-1}$  表示  $t-1$  时刻的后验状态估计值; $u_{t-1}$  表示  $t-1$  时刻的控制输入矩阵。

- 4) 预测过程中先验估计协方差估计方程:

$$P_t = AP_{t-1}A^T + Q \quad (4)$$

式中, $P_t$  表示  $t$  时刻的先验估计协方差; $P_{t-1}$  表示  $t-1$  时刻后验估计协方差; $Q$  为过程噪声  $W_t$  的协方差矩阵。

5)更新过程中的卡尔曼增益方程:

$$K_t = \frac{P_t H^T}{H P_t H^T + R} \quad (5)$$

式中,  $K_t$  表示  $t$  时刻卡尔曼增益;  $R$  为测量噪声  $V_t$  的协方差矩阵。

6)更新过程中的后验估计方程:

$$\hat{x}_t = \hat{x}_{t-1} + K_t (Z_t - H \hat{x}_{t-1}) \quad (6)$$

式中,  $\hat{x}_t$  表示  $t$  时刻经过卡尔曼滤波后的最优估计值; 利用卡尔曼增益  $K$ , 表征了状态最优估计过程中模型预测误差和量测误差的比重。

7)更新过程中的后验估计协方差估计方程:

$$P_t = (I - K_t H) P_{t-1} \quad (7)$$

式中,  $P_t$  表示  $t$  时刻卡尔曼滤波的后验估计协方差, 更新后作为下一次迭代的输入;  $I$  为单位矩阵。KF 算法正是利用这种不断更新的机制, 以此来减少随机噪声的影响。

### 1.2.1 状态预测

为了预测目标在视频帧中的状态, 本文算法使用线性恒速模型来估计目标在帧间的位移, 每个目标的状态被建模为一个 8 维向量, 即:

$$X = [x \ y \ w \ h \ \dot{x} \ \dot{y} \ \dot{w} \ \dot{h}]^T \quad (8)$$

式中,  $x$  表示目标中心的横坐标;  $y$  表示目标中心

$$\left\{ \begin{array}{l} D_{\text{high-match}} = \{D_{\text{high-match}}^1, D_{\text{high-match}}^2, \dots, D_{\text{high-match}}^{s_{\text{high-t}}} \} \\ D_{\text{low-match}} = \{D_{\text{low-match}}^1, D_{\text{low-match}}^2, \dots, D_{\text{low-match}}^{s_{\text{low-t}}} \} \\ T_{\text{act-match}} = \{T_{\text{act-match}}^1, T_{\text{act-match}}^2, \dots, T_{\text{act-match}}^{s_{\text{act-t}}} \} \\ T_{\text{deact-match}} = \{T_{\text{deact-match}}^1, T_{\text{deact-match}}^2, \dots, T_{\text{deact-match}}^{s_{\text{deact-t}}} \} \\ D_{\text{high-unmatch}} = \{D_{\text{high-unmatch}}^1, D_{\text{high-unmatch}}^2, \dots, D_{\text{high-unmatch}}^{m_t - s_{\text{high-t}}} \} \\ D_{\text{low-unmatch}} = \{D_{\text{low-unmatch}}^1, D_{\text{low-unmatch}}^2, \dots, D_{\text{low-unmatch}}^{m_t - s_{\text{low-t}}} \} \\ T_{\text{act-unmatch}} = \{T_{\text{act-unmatch}}^1, T_{\text{act-unmatch}}^2, \dots, T_{\text{act-unmatch}}^{n_{t-1} - s_{\text{act-t}}} \} \\ T_{\text{deact-unmatch}} = \{T_{\text{deact-unmatch}}^1, T_{\text{deact-unmatch}}^2, \dots, T_{\text{deact-unmatch}}^{n_{t-1} - s_{\text{deact-t}}} \} \end{array} \right. \quad (11)$$

式中,  $D_{\text{high-match}}$ 、 $D_{\text{low-match}}$ 、 $T_{\text{act-match}}$ 、 $T_{\text{deact-match}}$  分别表示匹配成功的高分检测目标、低分检测目标、激活轨迹、非激活轨迹;  $s_{\text{high-t}}$ 、 $s_{\text{low-t}}$ 、 $s_{\text{act-t}}$ 、 $s_{\text{deact-t}}$  分别表示当前时刻匹配成功的数量;  $D_{\text{high-unmatch}}$ 、 $D_{\text{low-unmatch}}$ 、 $T_{\text{act-unmatch}}$ 、 $T_{\text{deact-unmatch}}$  分别表示匹配失败的高分检测目标、低分检测目标、激活轨迹、非激活轨迹;  $n_{t-1}$  表示上一时刻的所有跟踪目标数量;  $m_t$  表示当前时刻的检测目标数量。当匹配完成后, 使用匹配成功的目标  $D_{\text{high-match}}$  和  $D_{\text{low-match}}$  根据朴素贝叶斯规则更新对应的目标状态。

### 1.3 基于 DIOU 的贪婪匹配算法

如图 2 所示, 在 MOT 任务中, 由于遮挡、目标交互、运动不规则等变化状态, 存在外观特征和运动特征难以获取的情况, 即 IOU 为 0, 导致轨迹匹配失败。为解决传统 IOU 度量失效问题, 本

的纵坐标;  $w$  表示目标框的宽度;  $h$  表示目标框的高度,  $\dot{x}$ 、 $\dot{y}$ 、 $\dot{w}$ 、 $\dot{h}$  分别表示它们在图像坐标系中的变化速率。当检测与轨迹匹配成功时, 使用检测框更新目标状态; 如果检测与轨迹未匹配成功时, 则简单地预测目标状态。

第  $t-1$  帧所有目标状态的集合为:

$$\{\mu_{t-1}^1, \mu_{t-1}^2, \dots, \mu_{t-1}^{n_{t-1}}\} \quad (9)$$

式中,  $\mu_{t-1}^i$  表示  $t-1$  在时刻的第  $i$  个目标状态;  $n_{t-1}$  表示在  $t-1$  时刻存在的目标数量。通过  $t-1$  时刻的目标状态, 根据线性恒速模型可以估计出第  $t$  帧的目标状态, 则目标的预测位置为:

$$\begin{cases} x_t = x_{t-1} + v_x \\ y_t = y_{t-1} + v_y \end{cases} \quad (10)$$

对于  $t-1$  时刻的每个目标状态  $\mu_{t-1}^i$ , 都可以给出  $t$  时刻的预测状态  $\mu_t^i$ 。

#### 1.2.2 状态更新

根据高低分检测与不同状态轨迹匹配结果, 可以将目标状态分为 8 种: 高分检测匹配成功、低分检测匹配成功、激活轨迹匹配成功、非激活轨迹匹配成功、高分检测匹配失败、低分检测匹配失败、激活轨迹匹配失败、非激活轨迹匹配失败。具体形式为:

文设计了基于 DIOU 的度量指标, 为初始不重叠的检测区域与轨迹构建时空相似性, 在不改变原始位置中心点、长宽比例和形状的基础上, 扩大两者的匹配空间。

IOU 度量指标如图 3 所示, 具体形式为:

$$D_i = \frac{A \cap B}{A \cup B} \quad (12)$$

式中,  $A$ 、 $B$  分别代表两个原始框。分子由  $A$ 、 $B$  两个框之间的交集组成, 分母由  $A$ 、 $B$  两个框之间的并集组成, IOU 指标由两者之间比值得到。

本文设计的基于 DIOU 的计算方法, 在原始框的周围添加一个成比例的膨胀区域, 如图 4 所示, 具体形式为:

$$D_{di} = \frac{C \cap D}{C \cup D} \quad (13)$$

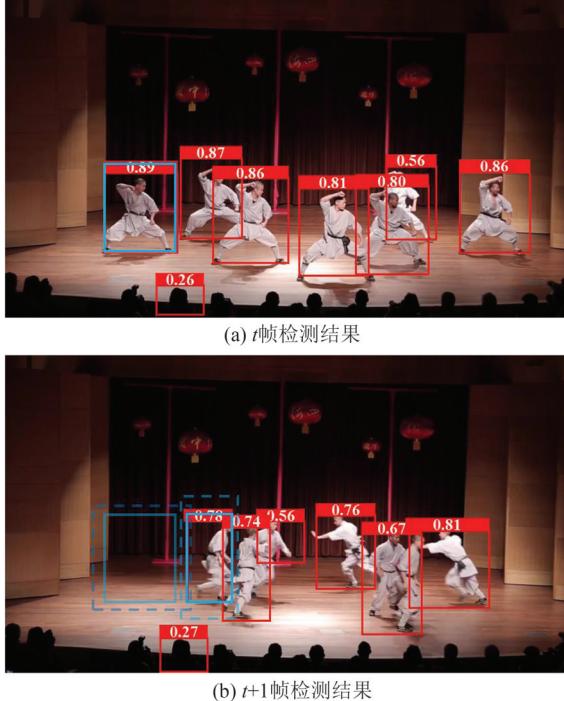


图2 膨胀区域匹配示意图

Fig. 2 Diagram of Dilatation Region Matching

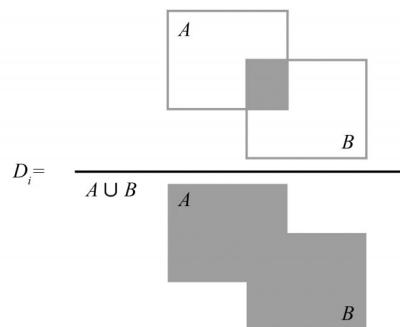


图3 传统IOU匹配示意图

Fig. 3 Diagram of Traditional IOU Matching

式中,  $C$ 、 $D$  分别代表基于  $A$ 、 $B$  两个原始框膨胀后的两个膨胀框。分子由  $C$ 、 $D$  两个膨胀框之间的交集组成, 分母由  $C$ 、 $D$  两个框之间的并集组成, DIOU 指标由两者之间比值得到。

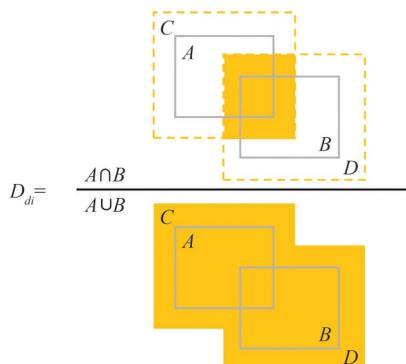


图4 DIOU匹配示意图

Fig. 4 Diagram of DIOU Matching

膨胀区域如图 5 所示, 膨胀框中心坐标与原始框共享, 具体形式为:

$$d = \frac{D_a - a}{a} = \frac{D_b - b}{b} \quad (14)$$

式中,  $d$  表示膨胀区域比例参数;  $D_a$  表示膨胀框的宽度;  $a$  表示原始框的宽度;  $D_b$  表示膨胀框的高度;  $b$  表示原始框的高度。

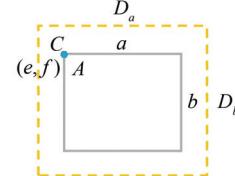


图5 膨胀区域示意图

Fig. 5 Diagram of Dilatation Region

假设原始匹配状态矩阵  $A$  表示为:

$$A = [e \ f \ a \ b] \quad (15)$$

式中,  $e$  表示检测框左上角点的横坐标;  $f$  表示检测框左上角点的纵坐标。膨胀区域比例参数为  $d$ , 则有膨胀后状态矩阵  $A_d$  表示为:

$$A_d = [(e - \frac{d_a}{2}) \ (f + \frac{d_b}{2}) \ a(1+d) \ b(1+d)] \quad (16)$$

式中,  $d_a$  表示膨胀增加的宽度;  $d_b$  表示膨胀增加的高度。本文方法使用三级关联的级联匹配策略, 分别有小膨胀区域比例超参数  $d_1$ 、中膨胀区域比例超参数  $d_2$ 、大膨胀区域比例超参数  $d_3$ , 且  $d_1 < d_2 < d_3$ , 在  $[0.2, 0.7]$  范围内搜索  $d_1$ 、 $d_2$ 、 $d_3$  的组合, 组合种类是有限的, 因此, 通过网格搜索确定最佳数值  $d_1$ 、 $d_2$ 、 $d_3$ 。

#### 1.4 自适应轨迹管理策略

在 MOT 任务中, 由于跟踪目标存在离开视野或者可能有新目标进入视野的情况出现, 需要有一个轨迹管理模块, 用于决定轨迹的生成和终止。一般做法是采用基于计数的固定生命周期策略, 将未匹配成功的检测目标视为潜在的新轨迹, 为提高跟踪准确性, 只有当该目标连续匹配成功至少最小命中帧后, 才将其认为是一个新的轨迹; 将未匹配成功的跟踪目标视为离开视野的轨迹, 为避免误删轨迹, 只有当该目标连续未匹配成功超过固定生命周期帧后, 才将其认为是跟踪目标丢失视野并删除其对应的轨迹。这种基于计数的固定周期管理策略, 保留了因单帧漏检未能匹配的轨迹, 并删除了确认离开视野的目标轨迹, 一定程度上减少了身份跳变, 保持了身份的唯一性。在实际过程中, 目标检测器往往存在误检、漏检的情况, 采用固定周期的轨迹管理策

略,将产生错误的跟踪轨迹,未能有效利用检测置信度信息对所有的目标进行相同的管理操作,导致高置信度的检测结果需要连续多帧匹配后才能确认跟踪轨迹,其间如果出现遮挡情况,有可能被删除;而低置信度的检测结果需要等待多帧匹配失败后,才会被删除。

基于以上原因,本文提出一种基于置信度的轨迹自适应管理策略。首先,根据检测器的置信度结果,动态调整轨迹的生命周期。对于检测与轨迹匹配失败的情况,轨迹置信度衰减;对于检测与轨迹匹配成功的情况,轨迹置信度增加。其次,根据当前帧与上一帧的检测框与视野中心距离比值,判断目标状态。如果比值大于 1,则代表目标在远离视野,属于逃逸状态,轨迹置信度下降;如果比值小于 1,则代表目标靠近视野,属于接近状态,轨迹置信度上升。轨迹状态具体形式为:

$$\begin{cases} S_t = \sqrt{(x_t - x_0)^2 + ((y_t - y_0) \times k)^2} \\ \Delta S = \frac{S_t}{S_{t-1}} \end{cases} \quad (17)$$

式中, $S_t$  为检测框中心与视野中心的距离; $(x_t, y_t)$  为  $t$  时刻的检测框中心的横、纵坐标; $(x_0, y_0)$  为视野中心的横、纵坐标; $k$  为常数,代表视野图像的宽高比; $\Delta S$  为轨迹状态判定系数,由当前帧与前一帧到视野中心的距离比值组成。

基于检测置信度的轨迹分数公式为:

$$T_t = \begin{cases} T_{t-1} - \epsilon \times \Delta S, & \text{未匹配} \\ 1 - ((1 - T_{t-1}) \times (1 - D_t) \times \Delta S), & \text{匹配} \end{cases} \quad (18)$$

式中, $T_t$  为  $t$  时刻的轨迹分数; $T_{t-1}$  为  $t-1$  时刻的轨迹分数; $\epsilon$  为分数衰减常数; $D_t$  是  $t$  时刻目标检测的置信度。根据当前时刻的轨迹置信度,则有:

$$C_{dmax} = C_{max} \times (\sigma(\frac{T_t}{T_{t-1}}) + \lambda) \quad (19)$$

式中, $C_{dmax}$  为动态调整后的最大生命周期; $C_{max}$  为最大生命周期; $\sigma()$  为 sigmoid 函数; $\lambda$  为偏移系数。通过选取合适的超参数  $\epsilon$  和  $\lambda$ ,利用 sigmoid 函数 S 型特征曲线,动态调整最大生命周期,进一步提升跟踪效果。当目标连续匹配成功时,轨迹接近视野中心,轨迹置信度持续上升,则生命周期越长;当目标匹配失败时,轨迹逃离视野中心,轨迹置信度持续下降,则轨迹生命周期越短,从而实现轨迹生命周期的动态调整。

## 2 DIOU 与自适应轨迹管理模块验证分析

### 2.1 数据集与评估指标

#### 2.1.1 数据集

本文实验基于 3 个公开的 MOT 数据集。MOT17 数据集是用于 MOT 的最常用的公开数据集之一,包含 28 个视频序列,其中 14 个视频为训练集(并提供了 3 种公开检测器 DPM、FRCNN、SDP),14 个视频为测试集,所有视频均以视频帧的形式提供,28 个视频共包含 11 235 帧,用于训练和测试的视频帧数分别为 5 316 帧和 5 919 帧。本文实验使用 MOT17 中训练集的每个视频前半段训练,后半段用于验证。

MOT20 数据集重点关注极端密集人群场景,其视频最多可达单帧 246 人,可以很好地验证算法在频繁遮挡、交互场景下的跟踪效果。MOT20 数据集共包含 8 个视频,其中 01、02、03、05 为训练集,04、06、07、08 为测试集,所有视频均以视频帧的形式提供,8 个视频共包含 13 410 帧,用于训练和测试的视频帧数分别为 8 931 帧和 4 479 帧。本文实验将 MOT20 中训练集的每个视频前半段用于训练,后半段用于验证。

DanceTrack 数据集包括集体舞蹈、功夫、体操等视频,具有外观相似且运动不规则的特点,减少了外观特征依赖,降低了目标检测的权重,提升数据关联的重要性,非常适合验证算法在数据关联方面的性能。DanceTrack 数据集包含 100 个视频,其中 40 个视频为训练集,25 个视频为验证集,35 个为测试集,所有视频均以视频帧的形式提供,100 个视频共包含 105 834 帧,用于训练、验证和测试的视频帧数分别为 41 775 帧、25 508 帧和 38 551 帧。本文实验使用 DanceTrack 中的训练集和验证集作为合集,每个视频前半段用于训练,后半段用于验证。

#### 2.1.2 软硬件配置

实验硬件配置为:CPU 为 Intel (R) Core (TM)i9-10900X,主频 3.70 GHz, GPU 为 NVIDIA GeForce RTX3090。实验软件环境为:ubuntu20, CUDA11.6 和 PyTorch 1.13.1。目标检测器使用 YOLOX,其中高分检测阈值为 0.6,低分检测阈值为 0.1。

在数据关联阶段,先使用本文提出的 DIOU 计算检测框和预测框之间的相似性,再使用贪婪算法进行轨迹匹配,设置相似性阈值为 0.7。大于

阈值则表示匹配成功,匹配成功的轨迹进入KF进行状态更新;否则代表匹配失败,对于检测未匹配的轨迹,作为新轨迹进入待匹配轨迹集合,对于轨迹未匹配,进入轨迹自适应管理模块。轨迹自适应管理模块根据目标置信度,动态调整轨迹最大生命周期,超过最大生命周期的轨迹直接删除,否则送入待匹配轨迹集合。

### 2.1.3 评估指标

本文实验使用高阶跟踪精度(higher order tracking accuracy, HOTA)<sup>[19]</sup>、MOTA<sup>[20]</sup>(multiple object tracking accuracy)、IDF1<sup>[21]</sup>(identification F-score)作为跟踪算法的评估指标。

HOTA是一个均衡考虑检测准确率、匹配准确率、定位准确率的综合评估指标。

MOTA指MOT准确度,是直观地衡量跟踪器在检测物体真阴性(true negatives, TN)、假阳性(false positives, FP)和保持轨迹身份错误切换(identification switch, IDS)次数的指标,也是衡量MOT算法精确性方面最重要的指标之一,以1

为最佳情况,数值越高代表跟踪精确度越好。

IDF1指目标身份确认的F1值,代表被检测和跟踪的目标中获取正确ID的检测目标比例,此指标更聚焦于跟踪算法跟踪某个目标的时间长短,考察跟踪的连续性和重识别的准确性, IDF1指标数值越高,代表跟踪特定目标的精度越好。本文从3个不同角度评估跟踪结果,其中,HOTA是最重要的评估指标。

## 2.2 实验结果

### 2.2.1 基于DIOU的匹配模块

在DIOU匹配模块中,膨胀系数 $d_1, d_2, d_3$ 是关键超参数。在DanceTrack数据集中,在取值范围[0.2, 0.7]之间形成 $d_1, d_2, d_3$ 的组合,同时评估不同组合的跟踪性能。由于 $d_1 < d_2 < d_3$ ,存在20种组合,如图6所示,当 $d_1 = 0.3, d_2 = 0.4, d_3 = 0.5$ 时,跟踪器获得了最高的HOTA值74.6;当使用IOU度量,其HOTA值为69.2,小于上述所有组合,因此,使用DIOU度量会显著影响跟踪性能。

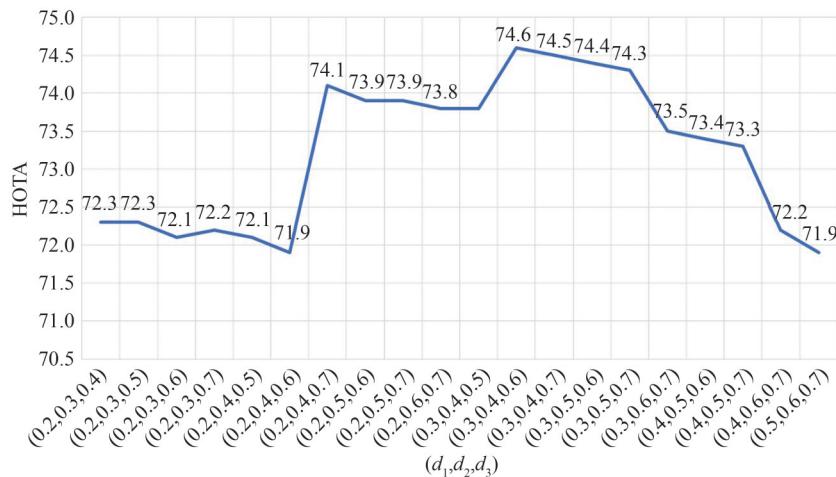


图6 不同 $d_1, d_2, d_3$ 组合在DanceTrack数据集上的验证结果

Fig. 6 Validation Results of Different  $d_1, d_2, d_3$  Combinations on DanceTrack Dataset

### 2.2.2 自适应轨迹管理模块

在自适应轨迹管理模块中,核心超参数为分数衰减常数 $\epsilon$ 和偏移系数 $\lambda$ ,本文使用 $\epsilon = \{0.1, 0.2, 0.3, 0.4, 0.5\}$ 和 $\lambda = \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$ 进行网格搜索,如图7所示,当 $\epsilon = 0.1, \lambda = 0.5$ 时,取得最佳HOTA值73.8。通过对比可知,分数衰减常数随着性能增大下降,而偏移系数则随着数值增大,性能趋势呈现先升后降。当使用固定生命周期策略时,即 $\epsilon = 0, \lambda = 0$ ,HOTA得分为69.2,小于所有使用自适应管理策略的得分,因此,使用本文提出的自适应管理策略对提升跟踪性能十分必要。

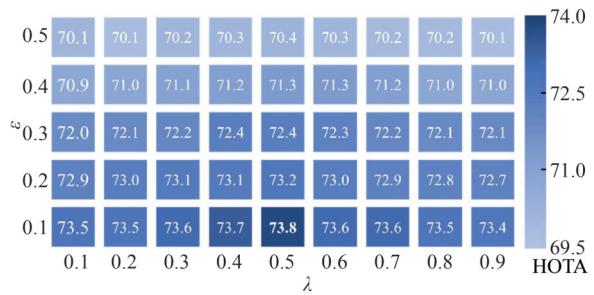


图7 不同 $\epsilon, \lambda$ 组合的跟踪结果对比

Fig. 7 Comparison of Results of Different  $\epsilon, \lambda$  Combinations

## 2.3 消融实验

为验证基于DIOU的匹配模块和自适应轨迹

管理模块对跟踪结果的影响,并量化各个模块的贡献,本文在 MOT17、MOT20 和 DanceTrack 数据集上进行验证。设置了 4 种组合方法:方法 A 是指使用基于传统 IOU 的匹配方法和固定生命周期方法,方法 B 是指使用基于传统 IOU 的匹配方法和本文提出的自适应轨迹管理方法,方法 C 是指使用本文提出的基于 DIOU 的匹配方法和固定生命周期方法,方法 D 是指使用本文提出的基于 DIOU 的匹配方法和自适应轨迹管理方法。

首先,对整体算法进行综合性能分析,表 1 列出了本文算法与各类先进 MOT 算法在 MOT17、MOT20 和 DanceTrack 数据集上的对比结果。从表 1 中可以发现,在 MOT17 数据集上,相较于 BotSORT,本文算法的 HOTA 有 2.6% 的提升,MOTA 有 2.2% 的提升,IDF1 有 1.2% 的提升,获得了平均 2% 的提升。在 MOT20 数据集上,本文算法的 HOTA 比第二名的 BotSORT 方法高出 2.5%,MOTA 比第二名 ByteTrack 高出 1.8%,而 IDF1 比 BotSORT 则高出 1.3%,3 项主要指标均获得了约 2% 的提升。在 DanceTrack 数据集上,本文算法的 HOTA 有 2.2% 的提升,IDF1 有 1.5% 的提升,而 MOTA 指标比第一名 SORT 低 0.5%,获得了 2 项指标提升、1 项指标接近的结果。HOTA、MOTA、IDF1 是反映检测和跟踪性能的主要指标,证明了本文整体方法在极度密集人群中和在不规则运动场景下跟踪的有效性。

其次,对基于 DIOU 的匹配模块进行分析。详见表 2,与方法 A 对比,方法 C 的 3 项指标在

MOT17 数据集约有 4.5% 的提升,在 MOT20 数据集约有 2.7% 的提升,在 DanceTrack 数据集约有 3.3% 的提升,证明了基于 DIOU 的匹配方法对跟踪性能的贡献。特别说明,对于 DanceTrack 数据集,由于不规则运动,更加容易引起检测框与预测框不重叠的情况,即  $\text{IOU}=0$ ,针对这种情况,本文提出的 DIOU 方法通过膨胀匹配区域克服了传统 IOU 失效的问题,获得可靠的相关结果,从而提升跟踪的精度与稳定性。

**表 1 不同跟踪算法在 MOT17、MOT20 和 DanceTrack 数据集上的定量比较**

Tab. 1 Quantitative Comparison of Different Tracking Algorithms on MOT17, MOT20 and DanceTrack Dataset

数据集	算法	HOTA/%	MOTA/%	IDF1/%
MOT17	SORT	62.1	73.8	69.8
	DeepSORT	62.4	74.4	71.2
	ByteTrack	63.1	80.3	77.3
	BotSORT	63.6	80.6	79.5
	本文算法	66.2	82.8	80.7
MOT20	SORT	62.3	73.4	70.8
	DeepSORT	61.7	72.6	70.5
	ByteTrack	61.3	77.8	75.2
	BotSORT	62.6	77.7	76.3
	本文算法	65.1	79.6	77.6
Dance-Track	SORT	67.9	89.8	70.8
	DeepSORT	65.6	85.8	67.9
	ByteTrack	71.5	87.5	72.5
	BotSORT	74.3	86.9	74.1
	本文算法	76.5	89.3	75.6

**表 2 不同组合方法在 MOT17、MOT20 和 DanceTrack 数据集上的指标对比**

Tab. 2 Performance Comparison of Different Combination Methods on MOT17, MOT20 and DanceTrack Datasets

数据集	组合方法	DIOU	自适应生命周期策略	HOTA/%	MOTA/%	IDF1/%
MOT17	A	×	×	61.7	74.7	71.1
	B	×	√	63.4	77.3	74.2
	C	√	×	64.6	79.6	76.8
	D	√	√	66.2	82.8	80.7
MOT20	A	×	×	62.4	73.8	73.4
	B	×	√	63.7	75.9	75.8
	C	√	×	63.5	77.5	76.8
	D	√	√	65.1	79.6	77.6
DanceTrack	A	×	×	69.2	87.3	71.7
	B	×	√	73.8	88.1	73.8
	C	√	×	74.6	88.8	74.7
	D	√	√	76.5	89.3	75.6

然后,对自适应轨迹管理模块进行分析。通过与方法 A 相比,方法 B 的 3 项指标在 MOT17 数

据集约有 2.5% 的提升,在 MOT20 数据集约有 1.9% 的提升,在 DanceTrack 数据集约有 2.5% 的

提升,证明了自适应生命周期方法对跟踪性能的贡献。特别说明,在IDF1指标均有超过2%的显著提升,这是因为MOT20和DanceTrack的数据特性,容易出现由于误检、漏检导致匹配失败的情况,使用本文提出的自适应生命周期管理方法,可以有效克服固定生命周期对于误检、漏检过于敏感的缺点。

最后,对基于DIoU的匹配模块和自适应轨迹管理模块两个整体的贡献进行分析。通过对方法A和D、B和D、C和D,3项指标在3个验证数据集上均有不同程度的提升,与方法A相比,方法D的HOA指标在MOT17、MOT20和DanceTrack数据集分别有4.5%、2.7%、7.3%的显著提升,消融实验表明,本文方法可以明显改善MOT结果。

#### 2.4 定性分析

对上述方法做定性分析,图8为DanceTrack0001数据集中第283~287帧,图9为DanceTrack0001数据集中第141~146帧。

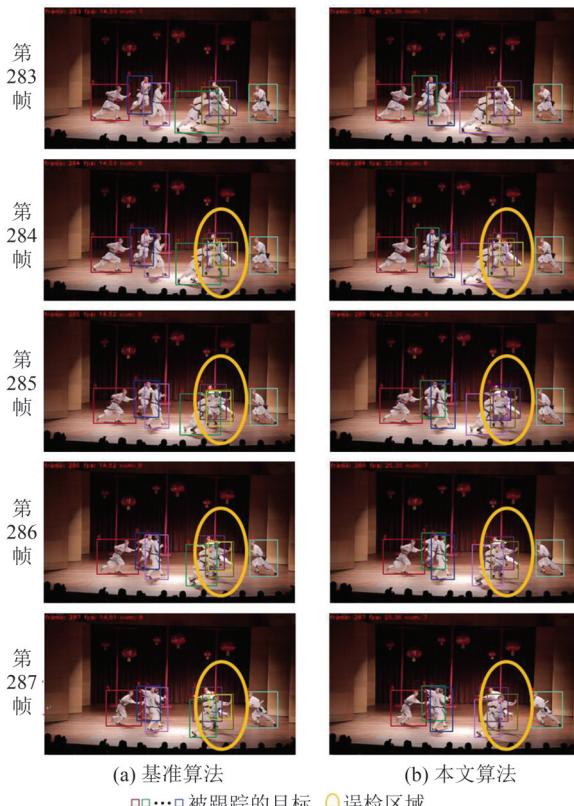


图8 错检场景下基准算法与本文算法跟踪结果对比

Fig. 8 Comparison of Tracking Results Between Benchmark Algorithm and Our Algorithm in False Detection Scenes

如图8中第284帧,基准算法和本文算法都出现了错检的情况(用黄圈标识),由于检测器误检新出现了错检目标(淡黄色方框,编号14),本

文算法采用了动态自适应生命周期管理方法,对于突然出现的不可靠目标给予轨迹置信度惩罚,极快地减少了其生命周期,两帧后,即第286帧,便删除了错检轨迹(编号14),而基准算法依然在持续跟踪错检轨迹,直到生命周期结束。上述结果表明,本文所提出的自适应生命周期方法,可以快速地减少错检轨迹的生命周期,有效地降低错检次数。



图9 错检场景下基准算法与本文算法跟踪结果对比  
Fig. 9 Comparison of Tracking Results Between Benchmark Algorithm and Our Algorithm in Missed Detection Scenes

对于漏检的情况,如图9中第142帧,由于被其他目标遮挡导致漏检(左列图中的编号3漏检,绿色方框;右列的编号2漏检,紫色方框),基准算法和本文算法均出现了漏检的情况,本文算法由于采用自适应轨迹管理策略,前期多帧匹配成功,扩大了生命周期,持续保存了目标的特征和ID,同时由于膨胀区域,提升了目标再次匹配的可能性,因此在4帧后,即第146帧,重新恢复为正确的ID(恢复了漏检目标编号2, ID未发生跳变),而基准算法达到最大生命周期就删除了目

标信息,没有正确恢复目标 ID。本文提出的膨胀区域匹配和自适应轨迹管理策略,对于漏检情况可以提升目标匹配的可能性,恢复正确目标身份,减少了漏检导致的目标身份跳变。

### 3 结语

本文提出了一种基于 DIOU 匹配和自适应轨迹管理策略的 MOT 算法,以解决目标频繁遮挡、交互导致的外观特征和运动特征难以获取的问题。首先,利用小膨胀区域提升高分检测与激活轨迹的匹配概率;然后,构建中膨胀匹配区域,提高低分检测与激活轨迹的匹配概率;接着,构建大膨胀匹配区域,提高未匹配高分检测与未激活轨迹数据关联的准确性;最后,使用自适应轨迹管理策略,依据目标检测置信度和轨迹状态,动态管理轨迹的生命周期,减少由于固定生命周期导致的错误关联问题。本文在 MOT17、MOT20 和 DanceTrack 公开数据集上进行综合评估,结果表明本文提出的算法相比于现有的 MOT 算法具有更好的精度与稳定性。

### 参 考 文 献

- [1] Zou Beiji, Li Bozhou, Liu Shu. A Multi-pedestrian Tracking Algorithm Based on Center Point Detection and Person Re-identification[J]. *Geomatics and Information Science of Wuhan University*, 2021, 46(9): 1345–1353. (邹北骥, 李伯洲, 刘姝. 基于中心点检测和重识别的多行人跟踪算法[J]. 武汉大学学报(信息科学版), 2021, 46(9): 1345–1353.)
- [2] Luo Xiaoyue, Wang Yanhui, Zhang Xingguo. A Violation Analysis Method of Traffic Targets Based on Video and GIS[J]. *Geomatics and Information Science of Wuhan University*, 2023, 48(4): 647–655. (罗霄月, 王艳慧, 张兴国. 视频与 GIS 协同的交通违规行为分析方法[J]. 武汉大学学报(信息科学版), 2023, 48(4): 647–655.)
- [3] Zhang Xing, Liu Tao, Sun Longpei, et al. A Visual-Inertial Collaborative Indoor Localization Method for Multiple Moving Pedestrian Targets[J]. *Geomatics and Information Science of Wuhan University*, 2021, 46(5): 672–680. (张星, 刘涛, 孙龙培, 等. 一种视觉与惯性协同的室内多行人目标定位方法[J]. 武汉大学学报(信息科学版), 2021, 46(5): 672–680.)
- [4] Bai H X, Cheng W S, Chu P, et al. GMOT-40: A Benchmark for Generic Multiple Object Tracking [C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, USA, 2021.
- [5] Cao J K, Pang J M, Weng X S, et al. Observation-Centric SORT: Rethinking SORT for Robust Multi-object Tracking [C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, Canada, 2023.
- [6] Zeng F G, Dong B, Zhang Y A, et al. MOTR: End-to-End Multiple-Object Tracking with Transformer [M]//The 17th European Conference on Computer Vision, Tel-Aviv, Israel, 2022.
- [7] Yu F W, Li W B, Li Q Q, et al. POI: Multiple Object Tracking with High Performance Detection and Appearance Feature [M]//The 14th European Conference on Computer Vision, Amsterdam, Netherlands, 2016.
- [8] Ren S, He K, Girshick R, et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks [J]. *Advances in Neural Information Processing Systems*, 2015, 28(10): 489–502.
- [9] Zhang Y F, Wang C Y, Wang X G, et al. Fair-MOT: On the Fairness of Detection and Re-identification in Multiple Object Tracking[J]. *International Journal of Computer Vision*, 2021, 129(11): 3069–3087.
- [10] Bochinski E, Eiselein V, Sikora T. High-Speed Tracking-by-Detection Without Using Image Information[C]//The 14th IEEE International Conference on Advanced Video and Signal Based Surveillance, Lecce, Italy, 2017.
- [11] Zheng L, Bie Z, Sun Y F, et al. MARS: A Video Benchmark for Large-Scale Person Re-identification [C]//The 14th European Conference on Computer Vision, Amsterdam, Netherlands, 2016.
- [12] Bewley A, Ge Z Y, Ott L, et al. Simple Online and Realtime Tracking [C]//The 23rd IEEE International Conference on Image Processing, Phoenix, USA, 2016.
- [13] Wojke N, Bewley A, Paulus D. Simple Online and Realtime Tracking with a Deep Association Metric [C]//IEEE International Conference on Image Processing, Beijing, China, 2017.
- [14] Chen L, Ai H Z, Zhuang Z J, et al. Real-Time Multiple People Tracking with Deeply Learned Candidate Selection and Person Re-Identification [C]//IEEE International Conference on Multimedia and Expo, San Diego, USA, 2018.
- [15] Wang Z D, Zheng L, Liu Y X, et al. Towards Real-Time Multi-object Tracking[C]//The 18th European Conference on Computer Vision, Munich, Germany, 2018.