计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
Computer Systems & Applications,2022,31(2):150−160 [doi: 10.15888/j.cnki.csa.008298]

E-mail: csa@iscas.ac.cn
http://www.c-s-a.org.cn
Tel: +86-10-62661041

基于运动信息和再匹配的多目标追踪①

韩　暑[1,2,3], 林　野[6], 郑龙澍[1,2,3], 翁哲鸣[4], 张立华[1,2,5]

[1](复旦大学 工程与应用技术研究院, 上海 200433)
[2](智能机器人教育部工程研究中心, 上海 200433)
[3](上海智能机器人工程技术研究中心, 上海 200433)
[4](中国航天科工飞航技术研究院, 北京 100074)
[5](季华实验室, 佛山 528200)
[6](吉林省人工智能与无人系统工程研究中心, 长春 130021)
通信作者: 张立华, E-mail: lihuazhang@fudan.edu.cn

摘　要: 随着目标检测模型的日趋成熟, 基于检测的追踪成为多目标追踪研究的主要方向. 借助几乎完美的目标检测结果, 在数据关联时可以采用只使用 IoU 信息的方法. 但是在实际使用中, 少量丢失的检测会造成大量的身份互换和轨迹断裂, 进而严重影响追踪效果. 针对这一问题, 该算法引入图像信息, 使用 IoU 模型进行初步追踪, 结合行人特征向量对初步追踪的结果进行校验, 对没有通过校验的轨迹进行再匹配. 对于目标间遮挡的问题, 该算法采用预测目标的运动轨迹, 提前采取措施的方法应对. 该算法采用 MOT16 和 2DMOT15 数据集进行实验, 均取得了较好的效果. 该算法采用在线追踪模式, 更适合在实际应用中使用.

关键词: 多目标追踪; IoU; 运动信息; 再匹配; 计算机视觉

引用格式: 韩暑,林野,郑龙澍,翁哲鸣,张立华.基于运动信息和再匹配的多目标追踪.计算机系统应用,2022,31(2):150–160. http://www.c-s-a.org.cn/1003-3254/8298.html

Multiple Object Tracking Based on Motion Information and Re-matching

HAN Shu[1,2,3], LIN Ye[6], ZHENG Long-Shu[1,2,3], WENG Zhe-Ming[4], ZHANG Li-Hua[1,2,5]

[1](Academy for Engineering and Technology, Fudan University, Shanghai 200433, China)
[2](Engineering Research Center of AI & Robotics, Ministry of Education, Shanghai 200433, China)
[3](Shanghai Engineering Research Center of AI & Robotics, Shanghai 200433, China)
[4](HIWING Technology Academy of CASIC, Beijing 100074, China)
[5](Ji Hua Laboratory, Foshan 528200, China)
[6](Jilin Engineering Research Center of AI & Unmanned Systems, Changchun 130021, China)

Abstract: With the maturity of object detection models, tracking-by-detection has become the mainstream of multi-object tracking research. Assisted by the almost perfect object detection results, data association can be formed only through the IoU information. However, in practice, a small number of missing detections will cause a large number of ID switches and fragmentations, which will seriously affect the tracking results. To solve this problem, the multiple object tracking algorithm is proposed with the introduction of image information. Specifically, preliminary tracking results obtained through the IoU model are verified with the pedestrian feature vector, and for the tracks that have not passed the verification, they are re-matched. For the problem of occlusion, the algorithm adopts the method of predicting the object trajectory and taking measures in advance. Tested on MOT16 and 2DMOT15 datasets, the algorithm has achieved good results, and it is more suitable for practical applications with its online tracking mode.

Key words: multiple object tracking; IoU; motion information; re-matching; computer vision

① 基金项目: 长春市科技创新“双十工程”项目 (19SS012)
收稿时间: 2021-04-07; 修改时间: 2021-05-11; 采用时间: 2021-05-19; csa 在线出版时间: 2022-01-17

1 引言

多目标追踪是一项识别和追踪视频中若干物体或人物位置的计算机视觉任务[1]. 换句话说, 多目标追踪的主要任务是在输入的视频或者一定的图像序列中, 确定若干目标在各帧中的位置, 并在这些帧中维持他们的 ID, 记录他们的轨迹[2].

由于使用的分类参数不同, 多目标追踪的分类方式主要有以下两种.

(1) 根据处理当前视频帧时能否获知接下来几帧的信息, 多目标追踪算法可以分为在线多目标追踪和离线多目标追踪两类[1]. 前者不能获得后续帧的信息, 追踪只能根据当前帧和之前帧的信息进行判断, 而且不能修改之前帧的追踪结果; 后者则一般要求输入整个视频片段, 在处理时可以得到整个视频中的信息, 从而得到全局最优解. 在线追踪的方法更适合实际应用, 所以目前的研究较多. 离线追踪的方法虽然不太适合监控等实际领域, 但是由于可以结合更多的信息, 所以相对追踪精度要更高, 而且通过放松对整体序列的要求, 将部分连续帧作为整体, 把整个视频序列的全局最优解换为若干连续帧的局部最优解, 也能在造成一些延迟的情况下进行实际应用, 并利用其高追踪精度的优势弥补这部分延迟带来的损失. 本文提出的算法由于从最初设计时就希望能贴合实际应用, 所以采用的是在线多目标追踪的方法.

(2) 根据是否需要依靠目标检测模型的输出, 也就是目标初始化的方式, 可以把多目标追踪模型分为需要检测器的追踪模型和无需检测器的追踪模型[2]. 前者的目标是使用检测模型的输出确定的, 因此检测器的能力直接影响追踪结果的质量. 这种方式能够自动完成对特定类别目标的追踪, 不需要人为标注的参与, 但是当检测器出现失误或者要求对非特定类别目标进行追踪时, 很难有效追踪. 后者则需要在视频的第一帧加入人为标注, 而后追踪模型根据这个标注进行目标追踪. 这种方式更类似于对单目标追踪的拓展, 具有目标类别不固定且不需要检测器等特点, 可以降低出现失误的概率. 但是这种方法需要人为对目标进行标注, 所以无法全自动运行, 而且目标在第一帧就确定了, 无法应对视频中出现新目标的情况. 由于研究者们更希望整个系统能够在训练和安装完成后摆脱人力的约束, 所以这两种追踪方式中有关前者的研究相对较多, 本文提出的算法也属于需要检测器的追踪方法.

随着目标检测模型的日趋成熟, 借助几乎完美的目标检测结果, 在数据关联时理论上可以只使用 IoU 信息. 但是在实际使用中, 少量丢失的检测会造成大量的身份互换和轨迹断裂, 进而严重影响追踪效果. 这说明现阶段多目标追踪仍需要图像信息的辅助, 结合使用 IoU 信息和图像信息才能获得更好的追踪效果.

针对上述问题, 本文在 IoU 模型中引入图像信息, 提出如下追踪算法: 采用 CenterNet 模型[3] 检测出当前的行人目标, 再使用 IoU 模型进行追踪. 本文的算法会对 IoU 模型得到的追踪结果进行校验, 同时使用图像信息来缓解轨迹断裂和身份交换的问题. 针对遮挡问题, 本算法预测可能发生的遮挡, 在图像信息相对完整的情况下保留图像信息, 同时预测被遮挡目标的位置, 把这些信息用于在遮挡结束后匹配目标, 从而提升追踪精度. 另外, 本算法对每个目标在下一帧的位置进行预测, 使用预测位置进行交并比的计算, 被遮挡目标再次出现的身份匹配也采用的是该目标运动后的预测位置.

本文的主要工作有: 使用目标检测模型输出的目标特征, 对 IoUtracker 模型的追踪结果进行校验和再匹配; 根据历史帧中的目标运动信息, 预测目标的位置, 并保存可能发生遮挡目标的特征, 用于在遮挡结束后恢复目标身份.

本文接下来会介绍部分多目标追踪的相关研究工作, 并对本算法使用的目标检测模型和本算法主要对比的模型进行简单介绍. 然后会详细介绍本算法, 并展示本算法的相关实验结果与分析. 最后会总结本算法并提出未来改进的方向.

2 相关工作

2.1 多目标追踪研究现状

基于检测的追踪模型的通用框架是: 在获取到视频序列后, 首先使用目标检测器对帧中的目标进行检测, 然后提取目标特征, 之后使用从其它帧中获取到的信息, 利用相似度计算的相关方法, 计算出两个目标是同一物体的概率, 最后根据之前计算的概率为各个目标分配 ID, 完成追踪.

现有的多目标追踪研究中, 有的是在整个流程的各阶段上进行改进提高效果, 有的通过借鉴其它相关领域的方法改进模型, 还有的通过跨领域研究的方式提供更多信息.

在目标检测阶段, 可以使用二维视频帧以外的信息, 或根据视频帧推测出更多信息来增强检测的力度. 例如 Gan 等[4] 同时使用立体声和视觉信息对车辆进行定位和追踪, 并且在光线不足的情况下也能取得较好的效果. Babaee 等[5] 使用 LSTM 进行追踪, 这个模型利用目标的历史位置和速度等信息对未来位置进行预测, 之后对预测结果和检测结果计算 IoU, 进行相似性的度量.

在特征提取阶段, 研究主要集中在对网络骨架的改动上. 例如 Lee 等[6] 将金字塔网络和孪生网络相结合, 根据已有轨迹对目标未来的位置进行预测, 之后使用实际检测结果和 NMS 进行处理, 并从网络的不同隐藏层提取特征, 将这些特征进行融合后得到了较好的特征. Gao 等[7] 将两种负责建模历史目标样本的结构化表示的图卷积网络, 与用于目标外观建模的暹罗框架合并, 得到能够利用当前帧的上下文来学习用于目标定位的自适应特征的模型.

在相似度计算阶段的改进主要是对从网络中得到的特征进行处理. 例如 Ji 等[8] 引入了双向 LSTM, 这个模型使用时序注意力网络得到注意力参数, 使用这一参数对空间注意力网络得到的特征进行加权, 得到了能够缓解遮挡影响的结果. Yoon 等[9] 也使用了双向 LSTM, 并用于计算相似性, 这个模型在全连接层顶端编码了约束框坐标和检测结果置信度等非外观特征, 在无外观特征方法中取得了领先的地位.

给目标分配 ID 实际上就是一个将当前帧中的目标与之前帧中目标进行匹配的过程. 在这一阶段, Ma 等[10] 使用了 GRU 来判断进行轨迹分流的时间点. Wang 等[11] 提出在两个单独的匹配阶段解决鲁棒性和判别力的要求, 这个模型在粗匹配阶段进行通用训练来增强鲁棒性, 而在精细匹配阶段通过远程学习网络增强辨别力. 这两个阶段的联系是后者可以使用前者的输出作为输入, 因此这两个阶段可以串联连接. 如果不使用这一方法, 也可以将它们并行连接, 通过对两者的结果进行优化和融合得到最终结果.

在借鉴相关领域研究方面, 通常是对单目标追踪领域较为成功的方法进行修改后应用在多目标追踪中. 如 Chu 等[12] 提出的模型中, 为了缓解行人交互时产生的遮挡导致追踪算法产生漂移的问题, 他们使用时空域注意力模型, 并判别可能出现的干扰目标. 其中每个目标都独立管理并更新时空域注意力模型, 并选择候选检测结果进行跟踪, 所以在本质上是将单目标追踪扩展到了多目标追踪的研究.

在与跨领域的研究结合的方法中, Ren 等[13] 提出的基于预测-决策网络的联合深度强化学习方法可以作为比较典型的例子. 这个模型将深度强化学习与多目标追踪相结合, 针对目标与目标之间产生遮挡时检测器效果变差影响到追踪效果的问题进行研究, 通过使用决策网络对目标与目标之间、目标与环境之间的联合交互来寻找最优追踪解. 上述这些方法都在许多方面给本算法提供了改进思路和想法.

2.2 目标检测模型

本算法采用的目标检测模型是在 CenterNet 算法的基础上改进的目标检测器, 主要是增加了每个目标的 128 维特征向量的信息输出[14].

目标检测算法主要可以分为二阶段和一阶段两类. 二阶段的目标检测算法如 Faster-RCNN 模型[15], 这类算法会先提取出一定量的候选区域, 再对每个候选区域进行特征提取后完成分类. 一阶段的目标检测算法的代表为 YOLOv3 模型[16], 这类模型一步完成坐标的回归与分类工作, 与二阶段检测模型相比, 在检测精度和检测效率的平衡上有一定的优势. CenterNet 模型则跳出了这两类方法的限制, 提出了零阶段或者无需候选框的新类型目标检测算法. 这个模型将图像直接送入全卷积网络中, 得到一幅热力图, 热力图中的峰值就是目标的中心, 对每个峰值处的图像特征进行回归来预测目标检测框的宽和高, 整个模型是端到端的, 所以检测速度上相较于之前的两类模型有一定的优势, 同时检测精度上也处于较为领先的地位.

2.3 IoUtracker 和 V-IoU 算法简介

IoUtracker 模型是 Bochinski 等[17] 在 2017 年提出的一个多目标追踪模型. 这个模型也是基于检测的模型, 是建立在输入视频或摄像头的帧率达到实时及以上水平, 以及目标检测器不会出现漏检和误检的假设上的多目标追踪模型. 这个模型通过相邻帧间检测框的 IoU 来进行目标的匹配.

IoUtracker 模型的具体流程为: 首先设定阈值 σ_l, 检测得分小于该阈值的检测框将被舍弃. 其次, 对于当前追踪队列中的每一个追踪框, 在当前帧的检测框中找到和该追踪框 IoU 最大的检测框, 并判断该 IoU 是否大于阈值 σ_{IoU}, 如果是, 则将该检测框与该追踪框匹配. 若没有满足上一步中的阈值要求, 则判断该追踪框

在之前帧中的最大检测得分是否大于阈值 σ_h, 以及在该帧之前目标出现的帧数是否大于阈值 $t_{\min}$. 如果满足这一要求, 就认为这个追踪框内是一个正常的追踪物体, 且在当前帧已经消失, 因此将该追踪物体移出追踪队列. 最后, 对于没有匹配上的检测框, 认为是一个新出现的物体, 作为待追踪的物体将对应检测框加入到追踪队列中. 这个方法处理速度非常快, 但是太依赖于检测的结果, 没有考虑目标被遮挡等情况下在部分帧中消失的问题, 所以这个模型的轨迹断裂和身份互换问题较为严重.

针对 IoUtracker 模型的问题, Bochinski 等在 2018 年提出了 V-IoU 模型[18]. 这个模型是在 IoUtracker 模型的基础上增加一个单目标追踪器, 基础想法是在目标消失时, 使用之前帧的图像信息初始化单目标追踪器, 利用这个单目标追踪器继续寻找并对该目标进行追踪.

但是这个模型也存在着一些问题. 例如在目标间发生遮挡的情况下, 在被遮挡的目标最后一次被检测到时, 这两个目标的检测框很可能已经有了较大程度的重叠, 那么使用这样的检测框进行初始化的视觉追踪获得的特征中就会有较多遮挡者的信息, 这会导致视觉追踪就可能受到这些错误特征的影响而追踪遮挡者, 导致后续的身份匹配失败. 另外这个算法是在轨迹断裂后才发挥作用, 可能对断裂前的身份变化的应对能力不足.

3 多目标追踪算法

3.1 本文算法介绍

通过对 IoU 追踪算法的分析可知, 该类型的追踪算法有较大的进步空间和研究价值, 存在的问题主要是目标检测模型的失误和目标间遮挡引发的. 目标检测模型的失误带来的影响很难用追踪模型进行弥补, 所以本文将侧重点放在缓解目标间遮挡带来的影响上.

本算法借鉴 V-IoU 算法中的利用视觉信息对 IoUtracker 模型进行改进的基本思路, 整个算法的介绍如下: 本算法先使用 IoU 算法进行追踪, 然后使用图像信息对 IoU 算法的结果进行校验, 校验不通过的 ID 需要进行再匹配, 最后得到较优的结果. 针对目标间遮挡造成的问题, 采取利用目标间的 IoU 信息预测可能发生的遮挡, 提前获得较为可靠的目标图像信息来帮助在遮挡结束后的目标身份分配的方法, 从而缓解身份互换的问题. 另外, 本算法对每个目标在下一帧的位置进行预测, 使用预测位置进行交并比的计算, 被遮挡目标再次出现的身份匹配也采用该目标运动后的预测位置进行对检测结果的搜索. 因为本算法追踪的目标为行人, 根据社会力模型[19] 和人群运动模式模型[20] 等研究, 行人的运动具有一定规律, 例如更倾向于直线运动等. 而且本算法假定视频序列的帧率较高, 所以借用微分的思想, 认为相邻帧中目标的运动是匀速直线运动, 进而对下一帧中目标的大致位置进行预测.

3.2 算法基本流程

本算法的大致步骤如图 1 所示. 基本流程为第一帧按检测赋 ID, 并保存所有目标的特征向量. 从第二帧开始, 先用 IoUtracker 模型进行追踪, 然后对新增轨迹进行复查, 减少轨迹数量. 接着对每条轨迹进行校验, 校验部分的流程会在后面详细说明. 在把当前帧中所有轨迹校验完成后, 根据上一帧和当前帧同一目标的中心点距离和方向计算运动矢量, 从而得到该目标在下一帧中的预测位置, 如图 2 所示. 在图 2 中, 第 t–1 帧中该目标检测框用细线框标出, 第 t 帧中的检测框则用粗线框表示, 结合这两帧中该目标的检测框信息, 预测第 t+1 帧中该目标位置为虚线框所示. 利用估算的位置和第二帧中的检测框大小计算各目标间的 IoU, 超过阈值 σ_{oc} 的认为可能会发生遮挡, 记录 ID、预测前进方向和距离, 以及第二帧中的特征向量, 把这些和预测遮挡相关的信息放入遮挡列表.

对新增轨迹进行复查是针对被检测到了, 但是由于交并比不足导致被认为是新目标的检测结果, 利用图像信息把新增的目标与在当前帧中没有更新的轨迹进行匹配.

由于本算法是建立在摄像头或视频的帧率较高且无较大变化的假设上的, 所以相邻帧的目标位置变化较小, 参考微分的思想可以认为相邻帧间目标是匀速直线运动的, 这样可以预测下一帧中目标的大致位置. 本算法使用中心点、检测框大小和运动矢量进行目标位置的预测.

当预测的轨迹超出图像范围或当前帧的位置在图像边缘且目标大小出现明显变化时, 认为该目标即将离开图像, 将其标记出来. 当轨迹消失达到一定帧数时, 认为这个轨迹终止, 也会将其标记. 这些被标记的轨迹在较少帧数后模型会停止对这些轨迹的校验和匹配.

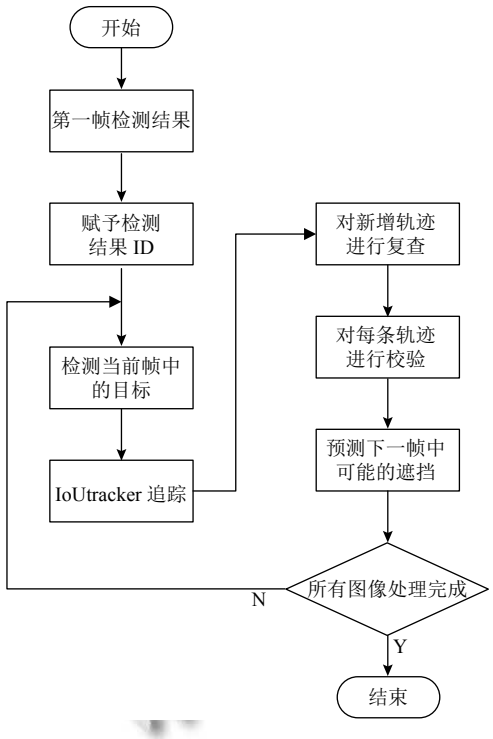
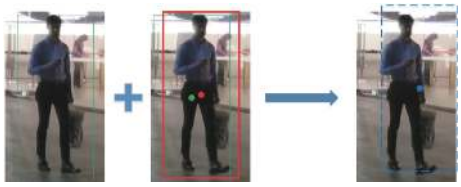


图 1 本算法的基本流程



(a) 第 t−1 帧 (b) 第 t 帧 (c) 第 t+1 帧

图 2 预测目标大致位置的示例

3.3 校验算法流程

对于未终止的轨迹, 流程如图 3 所示. 在校验阶段将当前帧的特征向量与该 ID 之前帧中的特征向量进行比较. 若该 ID 在遮挡列表中, 则增加使用遮挡列表中的特征向量的判别. 校验通过则该检测结果继承 ID, 否则先根据预估位置和运动矢量判断这些目标是否超出图像区域, 如果是的, 则将这些目标的上一帧记录和遮挡列表里的记录删除, 并移出追踪队列. 仍未解决的轨迹以上一帧追踪框的中心点为中心, 上一帧追踪框的顶点向追踪框外移动预测运动矢量后作矩形 R, 若有检测结果与矩形 R 的 IoU 大于阈值 σ_{da}, 则将这个检测结果列为候选对象, 利用图像信息判断该检测结果的 ID 归属.

把原轨迹中上一帧的检测框和预测运动矢量相结合, 可以得到目标的大致范围, 进而提升找到匹配的检测结果的效率. 这一步主要是为了缓解序列帧率较低或目标大幅度改变运动方向等情况造成的预测运动矢量出现偏差的问题.

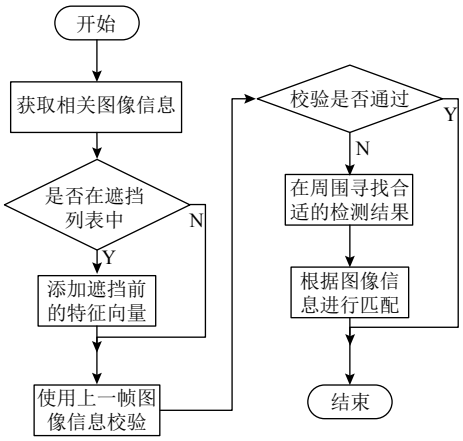


图 3 未终止轨迹校验流程

3.4 模块介绍

本算法主要可以分为 4 个模块, 分别为目标位置预测模块、轨迹整合模块、遮挡预测处理模块和目标校验与再匹配模块.

(1) 目标位置预测模块

本模块主要负责对目标下一帧位置的预测. 如图 2 所示, 在视频序列帧数较高, 目标在相邻若干帧之间位置变化较小的假设基础上, 本算法根据前两帧中目标的位置变化, 结合微分的思想, 以匀速直线运动方式预测目标位置, 形成目标运动向量和目标预测位置处的检测框, 为接下来的模块提供信息.

(2) 轨迹整合模块

本模块负责对 IoUtracker 模型的结果进行预处理.

由于 IoUtracker 模型要求轨迹必须连续, 没有任何对目标间遮挡等造成检测结果中断后目标再次出现的处理方法, 所以该模型输出的新增轨迹中包含大量原有断裂轨迹的目标. 因此本模块在目标再匹配之前, 对 IoUtracker 结果中新增的轨迹进行筛查, 合并能和已有轨迹匹配的新增轨迹, 达到初步对抗遮挡, 减少轨迹总数的效果.

本模块通过遍历新增轨迹, 通过它们与已有轨迹的预测目标位置的交并比和图像特征相似度, 与当前帧中未更新且未被视为结束的已有轨迹尝试匹配的方法, 进行轨迹整合.

(3) 遮挡预测处理模块

本模块主要负责寻找可能发生遮挡的目标, 并在

遮挡发生前保存该目标较为完整的图像信息.

本模块使用目标位置预测模块输出的目标预测位置, 计算下一帧中各目标检测框的交并比, 如果有一组目标的交并比超过阈值 σ_{occ}, 则将它们的身份 ID 和当前帧中的图像特征存入遮挡列表, 直到目标的预测检测框与其它目标的交并比小于 σ_{occ}, 此时本算法认为该目标遮挡过程结束或遮挡可能性消失, 将该目标从遮挡列表中移除.

另外, 如果目标已经在遮挡列表中, 就不更新该目标的信息. 因为此时该目标可能已经进入遮挡过程, 该目标对应的图像信息不一定比之前保存的特征更加有效, 所以本算法选择不更新遮挡列表中该目标的特征.

(4) 目标校验与再匹配模块

本模块负责使用图像特征对 IoUtracker 模型的输出进行校验, 并在目标没有通过校验的情况下为目标找到更适合的检测结果.

本模块的主要流程如图 3 所示. 本模块使用当前帧中轨迹匹配的检测与该轨迹之前帧中最后一个检测和遮挡列表中该轨迹保留的特征相似度进行校验. 如果通过校验, 该检测结果就继承该轨迹身份. 否则该轨迹预测位置框放大其运动矢量大小的矩形框作为搜索依据, 与这个矩形框交并比大于 σ_{reIoU} (进行再匹配时目标检测结果与轨迹框的交并比阈值) 的检测结果再与该轨迹进行图像特征校验, 超过一定阈值时允许该检测结果继承身份. 如果该检测结果已经与另一轨迹匹配, 则该检测结果这两个轨迹之前帧中最后一个检测和遮挡列表中该轨迹保留的特征与该检测结果特征相似度更高者的身份.

4 实验

4.1 实验说明

本文采用 2DMOT15[21] 和 MOT16[22] 数据集进行实验. 2DMOT15 数据集包含了 22 个视频, 这些视频被等分为训练集和测试集两部分. 这些视频来源于其它数据集, 并且使用的录制条件也不同, 例如摄像机的运动, 环境和光照的变化等. 这个数据集一共有 11 283 张分辨率不统一的图像, 共计包含 1 221 个不同的人, 总共有 101 345 个检测框. MOT16 数据集共含有 11 235 帧图像、1 342 个人和 292 733 个检测框. 包含了多个行人目标, 并存在目标交互与遮挡, 其中训练集和测试集各有 7 段视频.

实验中采用的评判标准以正确识别的检测与平均实际结果数和计算的检测数之比 (*IDF*1) 以及多目标追踪精度 (*MOTA*)[23] 两种为主, *IDF*1 综合了 ID 准确率 (IDP) 和 ID 召回率 (IDR), 公式如式 (1) 所示. *MOTA* 公式如式 (2) 所示, 其中 FN 为 false negative, 指追踪算法没有追踪到的真实目标的数量; FP 为 false positive, 指在追踪算法生成了轨迹但真实情况并没有该轨迹的数量; IDS 是指 ID 交换的情况, 也就是虽然检测到了目标并成功追踪到了, 但是对目标的身份识别错误的次数. GT 指标注数据值. *IDF*1 侧重于轨迹中 ID 信息的准确性, 而 *MOTA* 则更全面, 并且能体现出轨迹中 ID 的稳定性.

此外, 本文还采用命中目标轨迹数量 (MT) 和丢失目标轨迹数量 (ML) 等参数对结果进行分析. 其中, 命中目标轨迹数量指结果中与标注轨迹重合超过 80% 的轨迹的数量. 丢失目标轨迹数量则指的是结果中与标注轨迹重合小于 20% 的数量.

$$IDF1 = \frac{2TP}{2TP + FP + FN} \tag{1}$$

$$MOTA = 1 - \frac{(FN + FP + IDS)}{GT} \tag{2}$$

4.2 实验结果与分析

为了验证本算法的效果, 将本算法与其它算法的实验结果进行对比. 本算法与其它算法的实验结果对比是在 MOT16 数据集上进行的, 由于本算法主要在于对 IoUtracker 模型的改进, 所以增加了在 2DMOT15 数据集上与 V-IoU 算法的对比, 以说明本算法对 IoUtracker 改进的效果. 而且由于本算法和 V-IoU 算法都进行了对 IoUtracker 模型的改进, 所以两者在使用 IoUtracker 模型的部分的参数保持一致, 并且均使用 CenterNet 模型得到的检测结果, 这样有利于获得更可靠的比较结果.

4.2.1 与 V-IoU 算法结果的对比

V-IoU 算法和本文算法在 2DMOT15 和 MOT16 数据集上的各项指标结果如表 1 所示. 参数旁边的“↑”表示该参数越高, 该模型的追踪效果越好; “↓”表示该参数越低, 该模型的追踪效果越好. 从表 1 可以看出, 本文算法在 *MOTA*、*IDF*1 等指标上相对于 V-IoU 算法都有一定的提升, 主要原因有两点. 一是因为本文算法可以对 IoUtracker 模型的结果进行修正, 将检测到的结果尽可能连接到真正的轨迹上, 而不是在轨迹断裂后才进行应对. 二是本文算法采用预测目标大致位

置和可能的遮挡的方法, 能够更精确地寻找目标, 并在认为遮挡可能发生时就保存相关信息, 而不是使用目标被遮挡的前一帧处的很可能被污染或大量缺失的信息.

虽然本文算法在 *MOTA* 和 *IDF*1 等指标上取得了一定的提升, 但是本文算法在 FP 和 IDS 指标上表现出了不足. 这是因为 IoUtracker 模型会舍弃较短和整条轨迹中的检测得分不高于 σ_h 的轨迹, 以此来对抗检测结果的错误. 但是一些相对较小的目标的轨迹也会被忽略, 造成 FN 的上升. V-IoU 模型引入单目标追踪器缓解轨迹断裂问题, 但是依然保留对一些目标轨迹的忽略. 而本算法认为, 目前相对成熟的目标检测模型输出的检测结果, 在使用一定阈值的检测得分筛选后, 留下的更多是需要的目标, 所以本算法尽可能利用所有满足阈值的检测结果. 从表 1 中的 FP 和 FN 指标的对比可以发现, 本算法用相对较低的 FP 增加的代价, 得到了 FN 指标上的较多提升, 这也说明了本算法在检测结果利用上的优势. 表 2 是各模型和标注真实值在 MOT16 训练集上使用的检测结果数量, 可以看出本算法的数量更接近真实值, 也说明了本算法的优越性. 大幅降低的 FN 指标缓解了 FP 和 IDS 指标上升带来的负面影响, 所以在 *MOTA* 指标上本算法能相对于 IoUtracker 模型和 V-IoU 模型有一定的提升.

表 1 本文算法与 V-IoU 算法的实验结果对比

数据集	算法	*MOTA* (%) ↑	*IDF*1 (%) ↑	MT (条) ↑	ML (条) ↓	FP (个) ↓	FN (个) ↓	IDS (次)↓
2DMOT15-train	IoUtracker	54.8	49.7	147	203	2 951	14 703	365
	V-IoU	55.4	51.8	158	199	3 577	13 894	327
	本文算法	60.5	57.9	202	116	4 869	10 219	668
2DMOT15-test	IoUtracker	34.7	36.1	76	434	1 431	38 265	419
	V-IoU	36.3	37.7	86	426	1 846	36 922	397
	本文算法	44.4	44.3	189	171	3 572	27 653	2 928
MOT16-train	IoUtracker	57.5	51.5	114	257	4 242	41 950	785
	V-IoU	59.4	53.4	129	228	6 760	37 267	779
	本文算法	73.1	71.1	290	35	7 381	20 258	2 028
MOT16-test	IoUtracker	56.3	46.1	143	436	7 806	70 644	1 182
	V-IoU	56.9	48.5	155	411	11 991	65 582	1 046
	本文算法	66.1	64.5	271	167	12 035	46 904	2 903

表 2 MOT16 训练集上各模型使用的检测结果数量

模型和真实值	数量
IoUtracker模型	72 782
V-IoU模型	79 499
本文算法	97 327
真实值	110 407

由于 IoUtracker 模型只关注连续的轨迹, 所以当目标轨迹断裂后再出现时, 目标身份会发生变化, 所以该模型的 IDS 指标较低. V-IoU 模型虽然使用单目标追踪器对间断进行弥补, 但是为了降低目标漂移的影响只追踪少量帧, 如果在这些帧内目标没有出现, 那么该身份也会被终止, 即使后续该目标再次出现, 模型也会赋予目标新身份, 所以 IDS 也较低. 而本算法尽可能使用满足要求的检测结果, 而且为了减少遮挡带来的影响, 将交并比的判断范围放大, 让每条轨迹可能使用的检测结果增多, 这导致 IDS 指标数值较高.

本文算法在 MOT16 的各个视频序列上的测试结果如表 3 所示, 这里主要考虑的是 *IDF*1 和 *MOTA* 这两个指标, 对本算法进行分析.

分析表 3 的数据可知, 本算法在 MOT16-04 和 MOT16-11 上的效果较好, 而 MOT16-08 和 MOT16-14 序列的结果则不是很理想. 前者说明本文算法对于摄像头的小幅度移动具有一定的鲁棒性. 后者结果较差的可能的原因是在摄像头大幅度运动和在大景深导致的行人大小和交错遮挡形成的复杂场景中, 本文算法不能很好地预测目标的轨迹, 以及不能很好地处理目标大小差距带来相关问题. 结合表 1 中本算法 FP、FN 和 IDS 等指标较高的问题, 说明本算法仍有许多不足, 需要采用一定的方法调整模型. 这些为继续改进本算法提供了方向.

4.2.2 与其它算法结果的对比

本算法在 MOT16 测试集上与其它算法的实验结果对比如表 4 所示. 参数旁边的“↑”表示该参数越高, 该模型的追踪效果越好; “↓”表示该参数越低, 该模型的追踪效果越好. 由表 4 可以看出, 本算法在 *MOTA* 和 *IDF*1 指标上有一定的优势, 在 MT 和 ML 指标上, 本算法也达到了较高的水平, 这说明本算法在追踪的

整体效果上取得了较好的表现.

CNNMTT 模型[24] 基于社会力模型中行人会成群运动的思想, 通过先粗分人群再细分的方法, 结合匈牙利算法对卷积网络提取的特征进行匹配, 取得了较好的效果. 这个算法和本算法都运用了社会力模型的思想, 但是该算法没有充分考虑人群中目标间互相遮挡带来的影响, 所以整体追踪效果受到了一定的影响. smartSORT 模型[25] 将前两帧的目标特征融合后进行特征匹配, 但是对行人空间信息的运用较少, 一定程度上影响了该算法的命中轨迹数量指标. RAR16wVGG 模型[26] 结合递归神经网络和循环神经网络, 根据长期轨迹进行数据关联和目标的时空状态预测, 很好地提高了命中轨迹数量并降低身份互换数量, 但是对于较短轨迹的匹配效果稍差, 进而影响了漏报率指标. VMaxx 模型[27] 使用卷积网络提取的图像特征和空间运动信息, 结合双流长短时记忆网络提供的长期目标图像特征和运动信息进行身份匹配. 这种方法可以有效学习到轨迹的特征, 从而降低丢失轨迹数量, 但是由于缺少身份矫正机制, 一旦身份分配出现错误就会延续下去, 导致 *IDF*1 指标的结果较为不理想.

表 3 本文算法在 MOT16 各序列上的结果

序列	*MOTA*	*IDF*1
MOT16-01	55.0	47.0
MOT16-02	63.6	51.2
MOT16-03	77.0	73.6
MOT16-04	77.4	85.4
MOT16-05	70.6	65.9
MOT16-06	55.9	52.7
MOT16-07	61.9	54.2
MOT16-08	42.7	41.0
MOT16-09	75.4	54.7
MOT16-10	66.9	51.5
MOT16-11	82.1	74.7
MOT16-12	54.4	60.3
MOT16-13	70.1	64.9
MOT16-14	44.8	47.0

表 4 不同算法在 MOT16 数据集上的实验结果对比

算法	*MOTA* (%) ↑	*IDF*1 (%) ↑	MT (条) ↑	ML (条) ↓	FP (个) ↓	FN (个) ↓	IDS (次) ↓
本文算法	66.1	64.5	271	167	12 035	46 904	2 903
CNNMTT[24]	65.2	62.2	246	162	6 578	55 896	946
smartSORT[25]	60.4	56.1	219	161	11 183	59 867	1 135
RAR16wVGG[26]	63.0	63.8	303	168	13 663	53 248	482
VMaxx[27]	62.6	49.2	248	160	10 604	56 182	1 389

本文算法通过使用图像信息对充分利用目标空间信息的 IoUtracker 模型的结果进行校验, 并考虑了遮挡应对机制, 结合目标空间状态预测, 在整体追踪效果上取得了较好的效果. 但是从实验结果对比中可以看出, 本算法的实验结果评估中, IDS 指标不理想, 说明本算法的身份矫正机制较为敏感, 进而造成了这一指标出现的问题.

4.2.3 消融实验

为了更好地分析本算法的性能, 研究各模块对结果的影响, 本文对模型进行了消融实验. 本实验是在 MOT16 的训练集上进行, 分别移除目标位置预测模块、轨迹整合模块、目标再匹配模块和遮挡预测处理模块, 结果如表 5 所示.

表 5 中的结果说明, 对结果影响最大的模块是目标位置预测模块, 而其它模块的作用相对较小. 通过观察视频序列, 发现本数据集中的遮挡情况并不十分严重, 因此后 3 个针对遮挡的模块的效果受到了影响, 而目标位置预测模块作用于所有的轨迹, 所以场景对该模块的影响较小. 另外, 目标再匹配模块移除后对 *IDF*1 指标的影响较大, 说明了在 IoUtracker 中存在大量身份匹配的错误, 本算法借助图像信息纠正了一些这样的错误, 实现了追踪性能的提升.

表 5 移除模块消融实验结果 (%)

移除的模块	*MOTA*	*IDF*1
目标位置预测模块	66.0	56.3
轨迹整合模块	71.4	69.5
目标校验与再匹配模块	72.4	65.1
遮挡预测处理模块	73.0	69.9
遮挡预测持续更新特征	73.1	70.8

另外, 表 5 中最后一行是在为了研究遮挡预测处理模块中, 保留目标第一次进入遮挡列表时的特征是否更加有效的实验结果. 该数据是在其它模块正常工作, 只是当遮挡列表中的目标再一次满足 σ_{occ} 时, 使用该帧目标的特征更新遮挡列表中的特征的情况下得到的. 可以看出虽然 *MOTA* 没有变化, 但是 *IDF*1 出现了下降的情况. 这说明本算法保留目标第一次进入遮挡

列表时的特征的方法更有利于缓解目标间遮挡的影响.

表 6 是逐步添加模块的消融实验结果. 由于本算法处理遮挡的方法允许目标短暂消失, 所以没有采用 IoUtracker 模型中终止轨迹的相关方法, 这使得本算法不会抛弃部分检测结果, 但在没有后续模块时, 不能将一些发生过断裂的轨迹连接起来, 所以在没有添加后续模块时, 相较于 IoUtracker 的结果, 本算法的 *MOTA* 会偏高而 *IDF*1 则偏低.

表 6 的结果再次证明了目标位置预测的重要性, 而其它模块的作用更多地体现在 *IDF*1 指标上, 这说明这些模块在轨迹断裂后恢复目标身份, 维持身份的稳定等方面有一定的效果.

表 6 添加模块消融实验结果 (%)

逐步添加的模块	*MOTA*	*IDF*1
无	62.1	48.5
目标位置预测模块	70.9	63.2
轨迹整合模块	72.4	65.1
目标校验与再匹配模块	73.0	69.9
遮挡预测处理模块	73.1	71.1

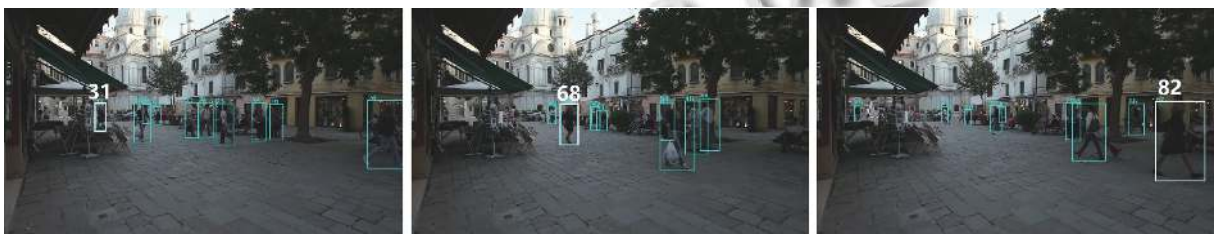
4.2.4 实验结果可视化示例

为了更直观地对这两种多目标追踪算法的效果进行比较, 本文将 MOT16-01 视频序列的部分实验结果可视化, 如图 4 和图 5 所示. 图 4(a) 和图 5(a) 均为 V-IoU 算法的效果, 图 4(b) 和图 5(b) 均是本算法的效果, 而同一列为同一帧图像. 图 4 是一个目标在较大跨度的帧中的追踪情况, 图 5 是一次目标间遮挡前后, 帧跨度较小的追踪情况.

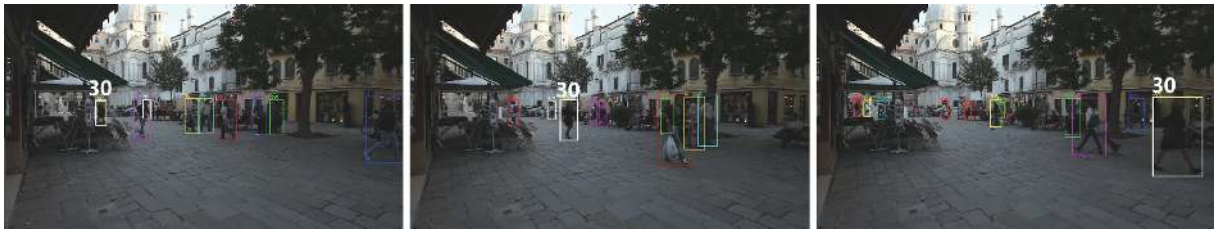
通过图 4 中的对比可以看出, V-IoU 算法在该目标的整个轨迹的处理中出现了间断, 造成该目标 ID 的变化; 而本算法则全程保持了该目标的 ID, 得到了更好的追踪结果. 而图 5 中, 被遮挡的目标再次出现后, V-IOU 算法并没有很好地识别该目标, 甚至出现了两个目标均更换了 ID 的情况; 而本算法通过提前保存信息和预测运动轨迹的方法, 成功在目标再次出现时保持了目标的 ID, 拥有更稳定的追踪表现.

5 结语

本文采用 CenterNet 模型检测当前帧中的行人, 使用 IoUtracker 模型进行初步追踪, 结合检测模型提供的行人特征向量对初步追踪的结果进行校验, 对没有通过校验的轨迹进行再处理, 寻找更适合的检测结果, 最后得到相对较优的匹配结果. 对于目标间遮挡的问题, 本文提出的算法通过预测目标的运动轨迹, 提前判断是否会出现遮挡, 并在遮挡前保存相对完整的目标特征和运动矢量, 为遮挡结束后的身份分配提供更可靠的依据. 本算法采用 MOT16 数据集进行实验, 在测试集上的平均追踪精度达到了 66.1%, 平均 *IDF*1 获得了 64.5% 的结果, 但是在场景特别复杂和摄像头大幅度运动的情况下的追踪效果不是特别理想. 综上, 本算法取得了一定的效果, 但是仍有可以继续改进的部分, 我们会把提高追踪效果和优化算法流程提高处理速度作为重点展开后续的研究.



(a) V-IoU 算法追踪效果一



(b) 本算法追踪效果一

图 4 同一目标在较大跨度帧中的追踪效果

(a) V-IoU 算法追踪效果二

(b) 本算法追踪效果二

图 5 目标间遮挡前后的追踪效果

参考文献

1 Ciaparrone G, Sánchez FL, Tabik S, *et al*. Deep learning in video multi-object tracking: A survey. Neurocomputing, 2019, 381: 61–88.

2 Luo WH, Xing JL, Milan A, *et al*. Multiple object tracking: A literature review. Artificial Intelligence, 2021, 293: 103448. [doi: 10.1016/j.artint.2020.103448]

3 Zhou XY, Wang DQ, Krähenbühl P. Objects as points. arXiv: 1904.07850, 2019.

4 Gan C, Zhao H, Chen PH, *et al*. Self-supervised moving vehicle tracking with stereo sound. The IEEE/CVF International Conference on Computer Vision (ICCV). Seoul: IEEE, 2019. 7052–7061.

5 Babaee M, Li ZM, Rigoll G. Occlusion handling in tracking multiple people using RNN. 2018 25th IEEE International Conference on Image Processing (ICIP). Athens: IEEE, 2018. 2715–2719.

6 Lee S, Kim E. Multiple object tracking via feature pyramid Siamese networks. IEEE Access, 2019, 7: 8181–8194. [doi: 10.1109/ACCESS.2018.2889442]

7 Gao JY, Zhang TZ, Xu CS. Graph convolutional tracking. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach: IEEE, 2019. 4644–4654.

8 Zhu J, Yang H, Liu N, *et al*. Online multi-object tracking with dual matching attention networks. In: Ferrari V, Hebert M, Sminchisescu C, *et al*. eds. Computer Vision-ECCV 2018. Cham: Springer, 2018. 379–396.

9 Yoon K, Kim DY, Yoon YC, *et al*. Data association for multi-object tracking via deep neural networks. Sensors, 2019, 19(3): 559. [doi: 10.3390/s19030559]

10 Ma C, Yang CS, Yang F, *et al*. Trajectory factory: Tracklet cleaving and re-connection by deep Siamese Bi-GRU for multiple object tracking. 2018 IEEE International Conference on Multimedia and Expo (ICME). San Diego: IEEE, 2018. 1–6.

11 Wang GT, Luo C, Xiong ZW, *et al*. SPM-tracker: Series-parallel matching for real-time visual object tracking. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach: IEEE, 2019. 3638–3647.

12 Chu Q, Ouyang WL, Li HS, *et al*. Online multi-object tracking using CNN-based single object tracker with spatial-temporal attention mechanism. 2017 IEEE International Conference on Computer Vision (ICCV). Venice: IEEE, 2017. 4846–4855.

13 Ren LL, Lu JW, Wang ZF, *et al*. Collaborative deep reinforcement learning for multi-object tracking. In: Ferrari V, Hebert M, Sminchisescu C, *et al*. eds. Computer Vision-ECCV 2018. Cham: Springer, 2018. 605–621.

14 Zhang YF, Wang CY, Wang XG, *et al*. A simple baseline for multi-object tracking. arXiv: 2004.01888v4, 2020.

15 Ren SQ, He KM, Girshick R, *et al*. Faster R-CNN: Towards real-time object detection with region proposal networks. Proceedings of the 28th International Conference on Neural Information Processing Systems. Montreal: IEEE, 2017. 1137–1149.

16 Redmon J, Farhadi A. YOLOv3: An incremental improvement. arXiv: 1804.02767v1, 2018.

17 Bochinski E, Eiselein V, Sikora T. High-Speed tracking-by-

detection without using image information. 2017 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS). Lecce: IEEE, 2017. 1–6.

18 Bochinski E, Senst T, Sikora T. Extending IoU based multi-object tracking by visual information. 2018 15th IEEE International Conference on Advanced Video and Signals based Surveillance. Auckland: IEEE, 2019. 1–6.

19 Helbing D, Molnár P. Social force model for pedestrian dynamics. arXiv: cond-mat/9805244, 1998.

20 Hu M, Ali S, Shah M. Detecting global motion patterns in complex videos. 2008 19th International Conference on Pattern Recognition. Tampa: IEEE, 2009. 1–5.

21 Leal-Taixé L, Milan A, Reid I, *et al*. Motchallenge 2015: Towards a benchmark for multi-target tracking. arXiv: 1504.01942v1, 2015.

22 Milan A, Leal-Taixe L, Reid I, *et al*. MOT16: A benchmark for multi-object tracking. arXiv: 1603.00831v2, 2016.

23 Bernardin K, Stiefelhagen R. Evaluating multiple object tracking performance: The CLEAR MOT metrics. EURASIP Journal on Image and Video Processing, 2008: 246309. [doi: 10.1155/2008/246309]

24 Mahmoudi N, Ahadi SM, Rahmati M. Multi-target tracking using CNN-based features: CNNMTT. Multimedia Tools and Applications, 2019, 78(6): 7077–7096. [doi: 10.1007/s11042-018-6467-6]

25 Meneses M, Matos L, Prado B, *et al*. Learning to associate detections for real-time multiple object tracking. arXiv: 2007.06041, 2020.

26 Fang K, Xiang Y, Li XC, *et al*. Recurrent autoregressive networks for online multi-object tracking. 2018 IEEE Winter Conference on Applications of Computer Vision (WACV). Lake Tahoe: IEEE, 2018. 466–475.

27 Wan XY, Wang JJ, Kong ZF, *et al*. Multi-object tracking using online metric learning with long short-term Memory. 2018 25th IEEE International Conference on Image Processing (ICIP). Athens: IEEE, 2018. 788–792.