

基于数据和群体运动仿真模型迭代优化的多目标跟踪算法

赵花蕊[†]

(河南省平台经济发展指导中心,河南 郑州 450008)

摘 要:在线多目标跟踪通过实时分析视频序列中的每一帧信息来估计目标轨迹。尽管深度学习方法在目标稀疏且外观特征明显的场景中表现良好,但在复杂环境下,频繁的遮挡会导致外观特征不可靠,从而影响跟踪准确性。为此,提出了一种基于数据和群体运动仿真模型多目标跟踪算法,将目标运动特征与视觉特征相结合。通过融合这两种特征,构建相似度矩阵以评估目标间的相似性。最终,利用匈牙利匹配算法对相似度矩阵中的检测目标和历史跟踪目标进行匹配,从而得到最终的跟踪结果。实验表明,该方法在线多目标跟踪方面显著提高了性能。

关键词:在线多目标跟踪;群体运动仿真模型;外观特征;匈牙利匹配算法

中图分类号:TP391

文献标识码:A

Multi-object Tracking Algorithm Based on Iterative Optimization of Data and Collective Motion Simulation Model

ZHAO Huarui[†]

(Platform Economy Development Guidance Center of Henan, Zhengzhou, Henan 450008, China)

Abstract: Online multi-object tracking estimates object trajectories by analyzing each frame of video sequences in real time. Although deep learning methods perform well in scenarios where objects are sparse and their appearance features are distinct, frequent occlusions in complex environments can render appearance features unreliable, thereby affecting tracking accuracy. To address this issue, this paper proposes a multi-object tracking algorithm based on data and collective motion simulation models, which combines object motion characteristics with visual features. By integrating these two types of features, a similarity matrix is constructed to evaluate the similarity between objects. Finally, the Hungarian matching algorithm is employed to match detected objects in the similarity matrix with historically tracked objects, leading to the final tracking results. Experimental results demonstrate that this method significantly improves performance in online multi-object tracking.

Key words: online multi-object tracking; collective motion simulation model; appearance features; Hungarian matching algorithm

随着计算机视觉技术的快速发展,多目标跟踪(Multi-object Tracking, MOT)已成为智能视频分析、自动驾驶、人机交互等领域的关键技术之一。在线多目标跟踪尤其具有挑战性,因为它需要在视频序列的每一帧中即时估计多个目标的轨迹,同时

应对各种复杂和动态的环境变化。

传统的多目标跟踪方法^[1]往往依赖于目标的手动特征提取和后续的数据关联,这些方法在目标特征明显且背景相对简单的场景中能够取得较好的效果。然而,在真实世界的应用场景中,目标间

的遮挡、光照变化、目标形变等复杂因素使得基于手动特征的方法难以保持稳定的跟踪性能。

近年来,深度学习技术的引入为多目标跟踪领域带来了新的发展机遇。特别是随着可靠目标检测器的发展,基于检测的多目标跟踪方法在性能方面取得了极大的提升^[2-3]。这些方法通常通过关联检测结果来构建目标的长期轨迹,从而克服跟踪失败的问题。此外,利用检测结果可以大幅缩小目标假设的搜索范围,并自动初始化新轨迹的跟踪。然而,虽然基于检测的跟踪方法在多目标跟踪中具有显著优势,但仍面临一些挑战。例如,目标检测^[4]结果可能存在漏检和误检,这可能导致跟踪器在目标存在遮挡或外观变化时失效。然而,当面对复杂场景,特别是目标频繁遮挡时,仅依赖外观特征进行匹配的方法往往难以准确跟踪目标,因为遮挡会导致目标的外观特征变得不可靠。

为了克服这一难题,本文提出了一种数据和运动仿真模型混合驱动的在线多目标跟踪方法。该方法不仅利用深度学习技术学习目标的外观特征,还引入了一个运动仿真模型来模拟目标未来的运动状态。通过结合目标的外观特征和运动特征,可以得到最终的相似度矩阵,然后使用匈牙利匹配算法对数据进行关联,使得我们能够在复杂场景中更准确地跟踪多个目标。

通过一系列在具有挑战性的公共数据集上的实验,我们验证了所提方法的有效性,并与其他最先进的批量跟踪和在线跟踪方法进行了对比。实验结果表明,本文所提方法不仅在跟踪准确度上有所提升,而且在处理复杂场景和遮挡情况时具有更好的鲁棒性。在接下来的章节中,我们将详细介绍本文所提方法的理论基础、实现细节以及实验结果。首先,我们将回顾多目标跟踪领域的相关研究,分析现有方法的优缺点,并阐述本文方法的研究动机和理论基础。

1 相关工作

在本节中将介绍与本文工作密切相关的 MOT 和群体运动仿真模型方面的前期工作。

1.1 多目标跟踪

在线跟踪方法通常通过逐帧局部关联检测结果来构建长期轨迹,利用粒子滤波^[5]来引导带有检测的多个跟踪器以跟踪遮挡下的目标。此外,基于边缘的零件模型^[6]也被用于描述目标的外观。然而,使用预定义外观模型^[7]的方法在外观剧烈变化时容易出现轨迹漂移问题。

为了解决轨迹漂移问题,一些方法采用了训练

好的在线分类器^[8]来捕捉目标的外观变化,并在遮挡情况下找到目标假设。这些方法构建置信度图^[9],将预训练的检测器和在线训练的分类器的输出进行融合。可变形部件模型^[10]被引入以处理部分遮挡。文献[8]使用强化学习方法学习了一个相似度函数。尽管这些基于局部关联的跟踪方法在许多场景下表现出了改进的性能,但由于它们仅利用了两个连续帧的信息,因此在长时间遮挡下仍然容易产生短而不连续的轨迹片段。

为了在遮挡情况下构建长期轨迹,全局关联方法^[11]通常利用整个序列的所有检测结果来构建最优轨迹。分层关联框架^[12]被设计为逐层生成更长的轨迹。文献[13]和文献[14]采用最小代价流算法解决了全局数据关联问题。文献[15]通过在检测对图中寻找最大加权独立集,将短轨迹合并为长轨迹。文献[16]开发了一个在线学习的 CRF 模型,并通过能量函数最小化来连接轨迹。文献[11]提出了寻找最优轨迹集的能量函数和优化方案。文献[17]对经典的 MHT 算法进行了 MOT 增强。

尽管许多 MOT 算法已经采用卡尔曼滤波方法来校正基于视觉的关联结果,但这些方法都建立在线性运动模型的基础上,无法很好地适应真实世界中目标的运动情况。此外,不同场景中的目标运动具有显著差异,因此使用相同的模型和参数也显得不合理。

1.2 群体运动仿真

多年来,模拟群体和群体行为已经得到了广泛的研究,涌现出各种技术以适应不同的场景和目标。微观方法已被广泛应用于实现行为多样性;而宏观方法则将人群作为一个整体进行建模,因此多样性并非其首要关注点。本文的重点是基于微观的方法,因此我们将其作为调研的核心内容。

基于智能体的人群模拟模型将人群中的每个人视为一个具有独特属性和目标的智能体。每个智能体根据其周围信息在每个时间步骤中做出个体决策。研究人员已经开发了各种微观控制模型^[18],包括基于速度和基于力的模型。在基于速度的模型中,每个智能体通过最小化给定的成本函数来选择速度。这些方法包括速度障碍物(VO)及其几个变体^[19]。在基于力的模型中,每个智能体接收从其与邻居之间的空间或社交关系中产生的虚拟力。Helbing 等提出的社交力模型(SFM)用于行人动态模拟已被广泛应用于人群模拟领域^[20]。除了对人群行为进行详细建模外,还有一些尝试将真实捕获的人群数据引入多智能体模拟。通过从真实世界样本中学习行为模式^[21]并从视频中提取真实世界的轨迹,可以合成自然的人群行

为。基于补丁的方法将原始轨迹从经验数据中转换为补丁,并通过一定规则将这些补丁连接起来。Yersin 等^[22]将运动斑块的概念扩展到密集人群的大环境中。Li 等^[23]通过复制粘贴技术,使用多智能体运动的示例来生成大量人群的动画。Hyun 等^[24]绘制了可变形的运动补丁,以描述多个角色的运动情节。Jordao 等^[25]提出了一种人群雕塑方法,通过直观的变形手势来引导人群的运动。与基于补丁的方法类似,研究人员复制从经验数据中提取的轨迹管来合成新的代理动画。Lai 等^[26]引入群体运动图,对离散的代理群体进行动画化。Lerner 等^[27]通过复制真实人的轨迹并应用于模拟代理,生成看似自然的行为。Chao 等^[28]提出了一种新的数据驱动方法,通过计算每个主体在每帧中的速度来得到。然而,由于试图最小化总体流量纹理能量,这种方法很耗时,因此对于交互式应用程序并不适用。我们的方法是对先前数据驱动方法的补充,并提出了一种将数据驱动和基于物理的多智能体方法相结合的新方法。

基于现有的群体仿真模型,我们利用数据集中的数据对仿真模型的参数进行优化,最终得到符合该场景的仿真模型。

2 方法

本文所提出的方法框架如图 1 所示,旨在实现对视频流进行高效准确的多目标跟踪。在现实世界的场景中,同时跟踪多个目标对于许多计算机视觉应用至关重要,例如交通监控、智能视频分析和自动驾驶等。我们的方法通过结合目标检测和群

体仿真推演的技术,提供了一种强大的多目标跟踪解决方案。

首先,我们针对当前帧的视频图像进行目标检测,以准确地定位和识别场景中的所有目标。通过先进的目标检测算法,我们能够提取目标的视觉特征,包括外观、形状和尺寸等信息。这些视觉特征为后续的跟踪过程提供了重要的线索。与传统的单目标跟踪方法不同,我们的方法还利用历史数据对先前目标进行群体仿真推演。通过分析历史目标的运动模式和行为,我们能够预测它们在当前帧中的运动特征。这种群体仿真推演的方法充分利用了目标之间的相互作用和关联,从而提高了多目标跟踪的准确性和鲁棒性。

在跟踪过程中,我们将运动特征和外观特征融合后进行相似度关联,以建立目标之间的联系。通过计算不同目标之间的相似度分数,我们能够确定最有可能的匹配关系,并获得最终的多目标跟踪结果。这种综合运动特征和外观特征的关联方法能够有效应对目标外观变化、遮挡和交叉等复杂情况,从而提高跟踪的鲁棒性和稳定性。

2.1 定义

对于需要跟踪的视频,总帧数为 T ,当前跟踪第 t 帧。历史跟踪目标为 M_{t-1} ,其中历史跟踪目标对应的历史轨迹点为 $S_{t-1} = \{s_i^{t-1} \mid i \in [1, n]\}$,其中 n 表示在 $t - 1$ 帧中的跟踪目标数。另外 $s_i^{t-1} = (x, y)$ 为目标 i 对应的轨迹点,包括位置和运动速度。历史外观特征为 $E_{t-1} = \{e_i^{t-1} \mid i \in [1, n]\}$ 。当前帧中通过目标检测后得到的检测目标中心点坐标 $D_t = \{d_i^t \mid i \in [1, m]\}$,其中 m 表示当前帧中检测目标的个数。

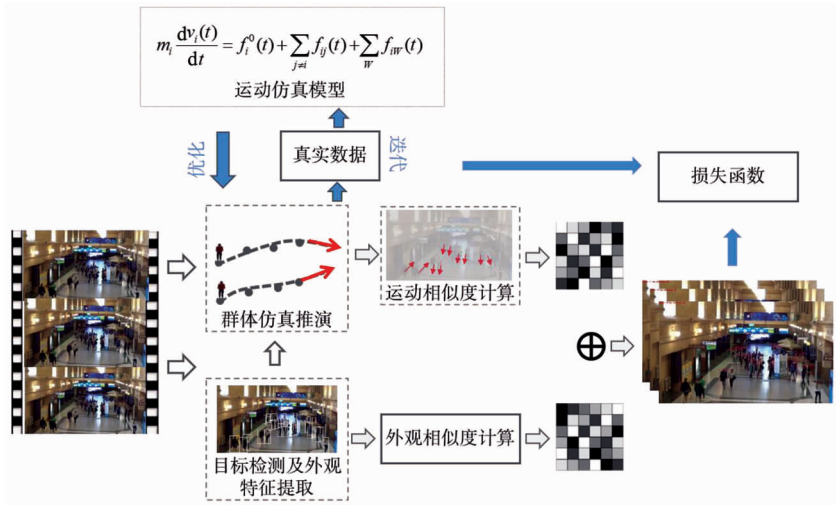


图 1 方法框架

2.2 目标检测及外观特征提取

对于当前帧 t 的图像数据,我们需要首先对其

进行目标检测,获取当前帧中的所有对象。受到 FairMOT 方法的启发,我们将其中的目标检测部

分作为我们的目标检测部分的基础架构。为了在准确性和速度之间取得良好的平衡,它选择将 ResNet-34 作为主干网络。主干网络采用了增强版的 Deep Layer Aggregation (DLA)^[29],以融合多层特征。与原始的 DLA 不同,添加了更多的跳跃连接,类似于特征金字塔网络 (Feature Pyramid Network, FPN)^[30],以增强低级特征和高级特征之间的信息传递。此外,它还对所有上采样模块中的卷积层进行了改进,采用了可变形卷积,使其能够根据物体的尺度和姿态进行动态调整,以提高模型的感受野。

特别地,该部分中三个并行的头被附加到 DLA-34 上,分别用于估计热图、对象中心偏移和边界框大小。每个头都是通过对 DLA-34 的输出特征应用 3×3 卷积(256 通道)来实现的,然后是

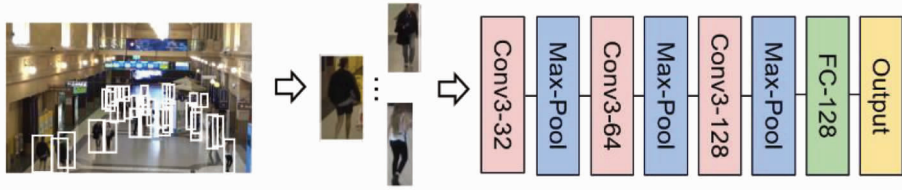


图 2 外观特征提取网络

2.3 群体仿真推演及模型优化

该部分主要涉及两块内容,首先是群体仿真模型构建以及基于数据的模型参数迭代优化。

社会力模型 (Social Force Model, SFM) 主要依据牛顿第二定律^[31],旨在模拟人群中个体之间的相互作用和行为规律,以便更真实地再现拥挤环境中的人流、交通、安全等现象。它基于人类行为学和社会学的原理,并借鉴了物理学中的力学概念。它假设每个个体都具有一定的目标和动力,与周围个体产生相互作用。这种相互作用可以通过各种力来建模,例如排斥力、引力等。通过对这些力的叠加和求和,可以计算出每个个体在给定环境中的运动方向和速度。社会力模型的一个重要应用是人群行为仿真,尤其是在城市规划、交通流分析和紧急情况下的疏散模拟等方面。通过模拟人群在不同场景中的行为,在本文中,我们使用社会力模型来模拟人群的运动,从而提取其对应的运动特征。

社会力模型将行人运动描述为 3 种力的作用结合,表达式如下:

$$m \frac{dv_i(t)}{dt} = f_i^0(t) + \sum_{j \neq i} f_{ij}(t) + \sum_w f_{iw}(t) \quad (1)$$

其中, i, j 为场景中任意两个目标个体; W 为场景中的静态障碍物,例如墙壁、路灯等等。 m 为

生成最终目标的 1×1 卷积层。

最终通过 FairMOT 中的目标检测分支得到了当前帧 t 中的目标检测结果 D_t 。在得到 D_t 后,我们通过外观特征提取网络对每个检测目标对应的 Bound Box 进行外观特征提取,作为该目标对应的外观特征。首先将目标从原图中提取出来并输入外观特征提取网络。该网络结构如图 2 所示。该网络首先使用一个 3×3 卷积(32 通道)来提取目标的浅层细节外观特征,然后通过后面的 3×3 卷积(64 通道)以及 3×3 卷积(128 通道)对其进行深层次语义特征提取。然后通过一个全连接层 (Fully Connected, FC) 得到最终的 128 维外观特征嵌入 $E_t = \{e_1, e_2, \dots, e_m\}$ 。在得到第 t 帧的外观特征嵌入之后,再结合历史外观特征嵌入,即可得到外观相似度矩阵 $V_{n \times m}$ 。

该场景中目标的质量。 $v_i(t)$ 为第 t 帧中第 i 个目标的速度。另外 $f_i^0(t)$ 表示自驱动力,也成为意愿力,其计算公式如下所示:

$$f_i^0(t) = m \frac{v_i^0(t)e_i^0(t) - v_i(t)}{\tau} \quad (2)$$

其中, $v_i^0(t)$ 表示目标 i 的期望速度的标量,而 $e_i^0(t)$ 为行人期望速度的单位方向向量。该自驱动力的作用,是使得行人可以以自身期望的速度向终点运动,期望速度的大小在这里我们将其作为超参数,而期望速度方向则是当前位置指向最终目标点的方向。

社会力模型中的 $f_{ij}(t)$ 表示行人之间的作用力,反映了目标受周围其他目标的影响。其表达式如下:

$$f_{ij}(t) = \{A_i e^{[(r_{ij} - d_{ij})/B_i]} + kg(r_{ij} - d_{ij})\} n_{ij} + \kappa g(r_{ij} - d_{ij}) \Delta v_{ji} t_{ij} \quad (3)$$

目标之间的作用力包含三部分,第一类是行人之间的排斥力;第二类是存在身体接触情况下行人之间产生的挤压力;第三类是当存在身体接触时行人之间的摩擦力。其中 $A_i e^{[(r_{ij} - d_{ij})/B_i]} n_{ij}$ 表示行人之间的斥力,表达的是行人 i 希望与周围行人保持一定的距离的意愿。斥力的作用是避免行人之间的碰撞以及保障行人的运动空间,行人间的距离越小行人间的斥力越大。其中 A_i, B_i 为位置参数; r_{ij} 为行人 i 和行人 j 模型的半径之和(在社会力模型

中假设行人模型是一个半径为 r 的圆盘); d_{ij} 为目标 i 和目标 j 所对应模型的圆心之间的距离; $\mathbf{n}_{ij}$ 是由行人 j 指向行人 i 的方向单位向量。与行人间的斥力不同,行人间的接触力涉及行人之间的物理层面的相互作用。当行人密度达到一定程度时,行人在运动过程中可能发生身体接触。在社会力模型中,我们使用两种力来表达行人间的身体接触:挤压力和摩擦力。挤压力是由于行人之间的物理接触而产生的力量。当行人靠近彼此时,他们可能会互相挤压,这会导致产生一个推开彼此的力。挤压力的计算公式为 $kg(r_{ij} - d_{ij})\mathbf{n}_{ij}$, 其中 k 为挤压系数;当 $r_{ij} - d_{ij} > 0$ 时,意味着行人之间离得很近并且发生了肢体接触,因此会产生挤压力,此时 $g(x) = x$;当 $r_{ij} - d_{ij} \leq 0$ 时,意味着行人直接距离很大,不存在身体接触,此时就没有挤压力,则 $g(x) = 0$ 。最后一类:行人之间的摩擦力 $\kappa g(r_{ij} - d_{ij})\Delta v_{ji}t_{ij}$, 其中 κ 为摩擦力系数; $\Delta v_{ji} = (v_j - v_i) \cdot t_{ij}$ 为行人之间切线方向上的相对速度; t_{ij} 为行人之间的切线方向。

社会力模型中的最后一部分 $f_{iW}(t)$ 为行人与环境中的障碍物(墙壁、路灯等)之间的作用力。与行人之间的相互作用力类似,其表达式如下:

$$f_{iW}(t) = \{A_i e^{[(r_i - d_{iW})/B_i]} + kg(r_i - d_{iW})\}\mathbf{n}_{iW} - \kappa g(r_i - d_{iW})(v_i \cdot t_{iW})t_{ij} \quad (4)$$

可以看出该表达式与行人之间作用力表达式类似,也分为行人和障碍物之间的斥力、摩擦力和挤压力。其中 r_i 为行人 i 的模型半径; d_{iW} 为行人圆心到墙体之间的距离; $\mathbf{n}_{iW}$ 为障碍物指向行人圆心的单位方向向量; t_{iW} 为行人与障碍物边界接触时的切线方向单位向量。

从社会力模型的表达式中可以看到,其中有 8 个参数需要进行手工设定 ($m, r, v^0, \tau, A, B, k, \kappa$)。这里我们不对具体的个体进行区分,假设所有行人公用同一组参数值。我们采取了一种基于数据的参数标定方式,在下一章节中进行详细介绍。

在通过社会力模型根据历史运动特征和当前状态信息得到当前帧的仿真推演结果 S_t , 然后我们基于 S_t, S_{t-1} 以及 D_t 得到最终的运动相似度矩阵。具体的运动相似度计算步骤如图 3 所示。图中的红色向量表示某目标检测结果关于目标 i 的检测运动特征,而蓝色向量表示基于社会力模型推演出的实际运动特征。我们通过余弦相似度来计算两个特征向量之间的相似度,得到最终的运动特征相似度矩阵 $\mathbf{M}_{n \times m}$ 。

2.4 模型执行步骤

本文模型分为训练和应用两部分。训练阶段主要是基于现有的公开或者自己采集的数据集进

行模型学习,通过数据集中的数据提升模型的特征建模能力,从而达到更好的跟踪效果。在完成模型训练之后,我们将模型参数进行保存,然后将训练好的模型进行测试,如果效果达到要求,则可以直接将其运用在实际应用场景中,例如对摄像头实时监控数据中的目标进行跟踪、对视频数据中的目标进行轨迹提取等。由于实际应用的执行步骤和测试阶段的执行步骤一致,因此下文主要从训练和测试两个阶段具体介绍我们所提方法的执行过程。

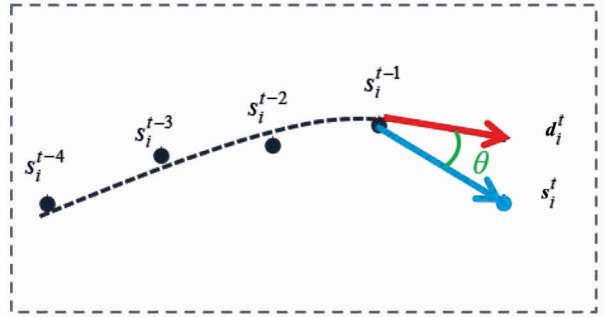


图 3 运动相似度计算

本文模型的训练分为两部分,首先是目标检测及外观特征提取部分,这部分由若干层人工神经网络组成,并且包含三个输出分支:估计热图、对象中心偏移和边界框大小,最终的损失函数也包含三部分,具体表达式可以参考 FairMOT 论文^[2]。另外一部分则是群体仿真模型优化。在训练开始阶段,我们根据经验为 2.3 节中的 8 个参数进行手动初始化。然后得到第 t 帧的推演结果 S_t , 根据每一推演结果与真实数据集的差异,利用梯度下降法优化社会力模型中的 8 个参数,直到模型收敛为止。在训练阶段,两个模型可以分开训练,这样一方面可以提升训练效率,另一方面也可以避免由于模型初始化导致的无法收敛的问题。

在测试阶段,图 1 中的蓝色箭头部分不再执行,该部分仅仅用于训练时优化社会力模型中的参数。测试时,两个模型的执行是有顺序的,具体步骤如图 4 所示。首先将当前第 t 帧的图像输入到目标检测及外观特征提取部分得到 D_t 和 E_t 。同时基于 $t-1$ 帧的数据 S_{t-1} , 群体仿真模型会对其当前运动状态做出推演,得到 S_t 。然后相似度计算得到外观特征相似度矩阵 $\mathbf{V}_{n \times m}$ 和运动特征相似度矩阵 $\mathbf{M}_{n \times m}$ 。然后将两个相似度矩阵进行融合得到 $\mathbf{MV}_{n \times m} = w_M \mathbf{M}_{n \times m} + w_V \mathbf{V}_{n \times m}$, 其中 w_M, w_V 表示特征对应的权重。在得到最终的相似度矩阵之后,我们利用匈牙利匹配算法,对历史跟踪目标和当前检测目标做匹配,得到最终的目标跟踪结果。

与测试阶段的执行步骤相同,实际应用中我们通过车载摄像头或者路测摄像头实时获取视频流,然后

将视频流输入到我们的方法中。该方法将实时进行多目标跟踪并返回每一时刻的跟踪结果。

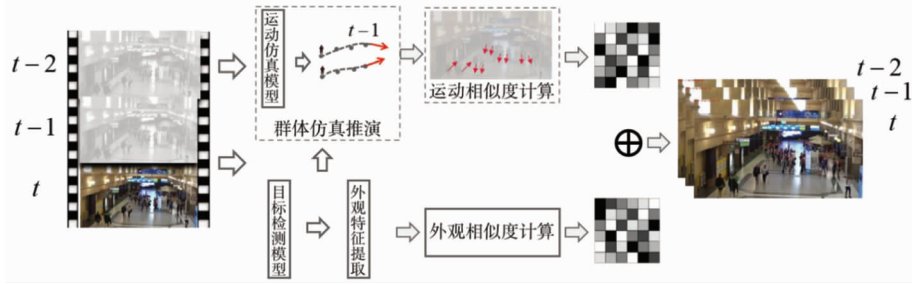


图 4 模型应用阶段执行步骤

3 实验

接下来我们通过实验,验证本文所提方法的有效性。实验环境如表 1 所示

表 1 实验环境配置

实验环境名称		具体配置
CPU	AMD Ryzen 7 5800H with Radeon Graphics(3.20GHz)	
RAM		16 GB
GPU		Nvidia GeForce GTX 3070
OS		Windows 10
Deep learning framework		Pytorch1.5.0
Python		Python3.7.3

3.1 数据集及指标

本文基于 MOT20 数据集^[32]进行多目标跟踪效果测试。该数据集是在 3 个人群密集的场景中拍摄得到的。其中包含 8 段视频,我们将其中的 4 段作为训练集,4 段作为测试集。另外,本文使用以下指标作为度量标准^[33]: ID 的 F1 指数(Identity F1-score IDF1),ID 的假阳性个数(Identity False Positive, IDFP),ID 的假阴性个数(Identity False Negative, IDFN),高阶跟踪精度(Higher Order Tracking Accuracy, HOTA),多目标跟踪精确度(MOT Accuracy, MOTA),假阳性(False Positive, FP),假阴性(False Negative, FN)。

3.2 对比方法

将本文方法与当前流行的 MOT 算法效果作比较,主要与 FairMOT、GSDT_v2 这两个方法进行比较。

3.3 跟踪效果

将上述方法在 MOT20 的测试数据上执行,并

获得最终的多目标跟踪结果,然后利用 3.1 中所提到的评价指标,得到最终的跟踪效果如表 2 所示。从其中的结果可以看出,本文方法基本上都取得了很大的提升。

从表 2 中可以看出,与 FairMOT 以及 GSDT_v2 相比,本文方法在实验中取得了令人满意的效果。具体来说,尽管 MOTA 指标的变化不大,但 IDF1 指标的变化明显。MOTA 指标主要反映了跟踪算法中目标检测部分的效果,而 IDF1 指标则用于评估跟踪算法中数据关联部分的效果。通过分析实验结果,我们可以看出本文方法引入了运动特征,从而显著降低了 IDFP 和 IDFN 指标,这表明每条轨迹中更多的部分被正确地关联起来。值得注意的是,通过增加运动特征,本文的方法能够更好地处理目标在运动过程中的遮挡问题。这个改进使得我们的跟踪算法在目标关联方面表现出更高的准确性和稳定性。通过利用运动特征,我们能够更好地区分目标之间的差异,减少了误关联和漏关联的情况,从而提升了整体的跟踪质量。

下面我们通过一组视频片段的跟踪结果,来展示我们的跟踪效果,如图 5 所示。

图 5 中的第一行展示了我们的方法在连续视频帧中的跟踪结果,而第二行展示了 FairMOT 方法的跟踪结果。我们选择了完整场景图的一部分进行分析,因为该部分出现了现有方法难以解决的遮挡问题。通过对比第一行和第二行的结果,我们可以清楚地观察到在目标之间发生遮挡时,我们的方法展现出更加鲁棒的效果。在目标发生遮挡时,传统方法仅基于外观特征的相似度估计和数据关联可能会导致错误的跟踪结果。这是因为遮挡情况下提取到的外观特征可能受到污染或不完整。然而,我们的方法通过引入运动仿真模型,成功地融合了运动特征和外观特征的相似度估计,从而在遮挡情况下取得了显著的提升。通过运动仿真模

型,我们能够更好地理解目标之间的运动关系,并利用这些信息对遮挡情况进行推断和补偿。这种引入运动特征的策略为外观特征提供了有力的补充,从而在遮挡情况下提升了最终的目标关联结果,进而改善了整体的跟踪效果。在我们的实验中,我们观察到我们的方法在处理目标遮挡时能够

更好地维持目标的连续性和准确性。我们的方法能够正确地预测遮挡目标的位置,及时识别出目标的重新出现,并将其与之前的轨迹正确关联起来。这为复杂场景下的目标跟踪提供了一种强大的解决方案,展示了我们的方法在应对遮挡问题方面的优势。

表 2 跟踪效果对比

算法	视频	IDF1	IDFP	IDFN	HOTA	MOTA	FP	FN	帧速
GSDT_v2 ^[23]	MOT20-01	81.41	2 417	4 570	67.47	86.46	233	2 386	8.0
	MOT20-02	72.32	32 653	43 665	62.40	86.54	2 113	18 068	7.8
	MOT20-03	89.59	19 100	125 419	70.62	88.60	5 288	29 853	6.5
	MOT20-05	83.42	81 571	222 334	67.73	87.95	16 221	60 141	4.9
FairMOT ^[6]	MOT20-01	84.13	2 404	3 696	70.79	88.42	445	1 737	10.2
	MOT20-02	75.15	32 994	41 728	65.19	87.45	4 923	13 657	9.3
	MOT20-03	89.68	27 983	35 961	71.19	89.03	12 761	20 739	7.1
	MOT20-05	85.48	84 745	100 612	69.67	88.49	28 235	44 102	6.5
Mine	MOT20-01	84.52	2 358	3 599	70.32	88.41	372	1 712	9.0
	MOT20-02	73.16	31 997	41186	68.52	87.83	4923	13 299	8.9
	MOT20-03	90.31	19002	33 129	71.24	89.86	6 984	19 498	6.7
	MOT20-05	86.04	83256	82516	70.02	88.13	18 286	34 247	5.4

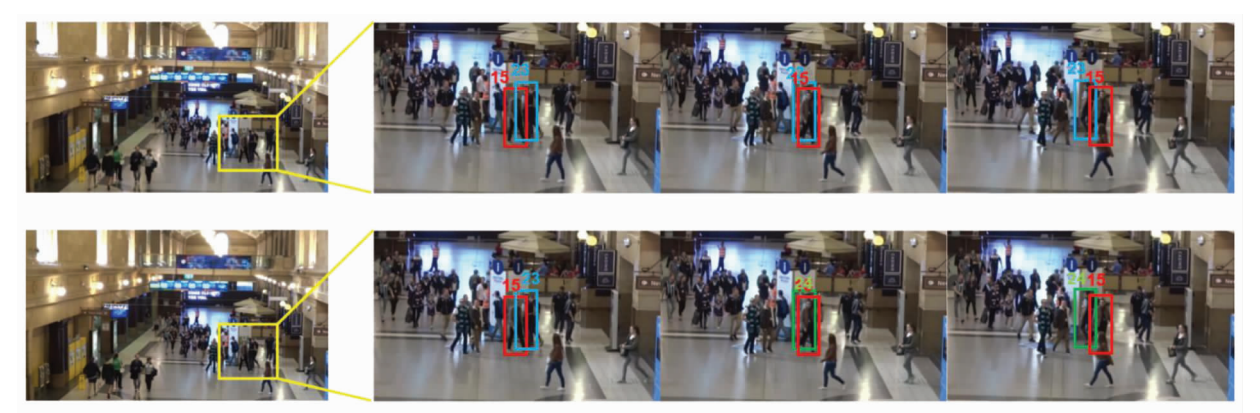


图 5 跟踪效果可视化

4 结 论

为了解决复杂场景下目标之间频繁的遮挡问题,本文提出了一种基于数据和群体运动仿真模型迭代优化的多目标跟踪算法。不同于以往的基于外观特征的 MOT 算法,本文通过引入群体运动仿真模型来提取目标的运动特征,从而在发生遮挡时可以利用更加稳健的运动特征为数据关联提供更加可靠的特征,进而提升了最终的目标跟踪准确度。但是我们的方法也存在一定的问题,因为本文主要基于社会力模型,由于该模型是一个比较基础

的群体仿真模型,在一些场景中可能无法通过简单的参数优化来很好地建模该场景中目标的运动特征。因此在下一步的工作中我们考虑使用更加复杂的群体仿真模型来更好的建模目标运动特征。另外,我们还考虑使用基于大模型的方法辅助 MOT 模型的跟踪。大模型基于大量的训练数据,同时包含千亿级别的参数量,其拥有的知识是基于规则的模型和普通的小型深度学习网络难以比拟的。因此我们下一步将会考虑将大模型的知识引入我们的方法中,用于提升多目标跟踪的效果。

参考文献

- [1] 李睿,商家赫.改进的CamShift无人机目标跟踪算法[J].计算机与现代化,2020(11):83—88.
- [2] ZHANG Y, WANG C, WANG X, et al. FairMOT: on the fairness of detection and re-identification in multiple object tracking [J]. International Journal of Computer Vision, 2021, 129(11): 3069—3087.
- [3] WANG Y, KITANI K, WENG X. Joint object detection and multi-object tracking with graph neural networks[C]//2021 IEEE International Conference on Robotics and Automation (ICRA). Los Alamitos: IEEE, 2021: 13708—13715.
- [4] 杨轲,董兵,吴悦,等.基于轻量化YOLOv4机场场面遥感图像目标检测方法[J].计算机与现代化,2024(2):93—99.
- [5] OKUMA K, TALEGHANI A, DE FREITAS N, et al. A boosted particle filter; multitarget detection and tracking [C]//Computer Vision-ECCV 2004: 8th European Conference on Computer Vision, Prague, Czech Republic, May 11—14, 2004. Proceedings, Part I 8. Springer, 2004: 28—39.
- [6] WU B, NEVATIA R. Detection and tracking of multiple, partially occluded humans by bayesian combination of edgelet based part detectors [J]. International Journal of Computer Vision, 2007, 75: 247—266.
- [7] XING J, AI H, LAO S. Multi-object tracking through occlusions by local tracklets filtering and global tracklets association with detection responses[C]//2009 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2009: 1200—1207.
- [8] XIANG Y, ALAHI A, SAVARESE S. Learning to track; online multi-object tracking by decision making [C]//Proceedings of the IEEE International Conference on Computer Vision, 2015: 4705—4713.
- [9] BREITENSTEIN M D, REICHLIN F, LEIBE B, et al. On-line multiperson tracking-by-detection from a single, uncalibrated camera [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010, 33(9): 1820—1833.
- [10] SHU G, DEGHAN A, OREIFEJ O, et al. Part-based multiple-person tracking with partial occlusion handling [C]//2012 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2012: 1815—1821.
- [11] MILAN A, ROTH S, SCHINDLER K. Continuous energy minimization for multitarget tracking [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 36(1): 58—72.
- [12] HUANG C, WU B, NEVATIA R. Robust object tracking by hierarchical association of detection responses[C]//Computer Vision-ECCV 2008: 10th European Conference on Computer Vision, Marseille, France, October 12—18, 2008. Proceedings, Part II 10. Springer, 2008: 788—801.
- [13] ZHANG L, LI Y, NEVATIA R. Global data association for multi-object tracking using network flows[C]//2008 IEEE conference on computer vision and pattern recognition. IEEE, 2008: 1—8.
- [14] BRENDEN W, AMER M, TODOROVIC S. Multiobject tracking as maximum weight independent set [C]// 2011 IEEE conference on computer vision and pattern recognition. IEEE, 2011: 1273—1280.
- [15] YANG B, NEVATIA R. An online learned CRF model for multi-target tracking[C]//2012 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2012: 2034—2041.
- [16] KIM C, LI F, CIPTADI A, et al. Multiple hypothesis tracking revisited [C]//Proceedings of the IEEE International Conference on Computer Vision, 2015: 4696—4704.
- [17] PELECHANO N, ALLBECK J M, KAPADIA M, et al. Simulating heterogeneous crowds with interactive behaviors [M]. CRC Press, 2016.
- [18] GUY S J, CHHUGANI J, KIM C, et al. Clearpath: highly parallel collision avoidance for multi-agent simulation[C]//Proceedings of the 2009 ACM SIGGRAPH/Eurographics Symposium on Computer Nimation, 2009: 177—187.
- [19] GUY S J, CURTIS S, LIN M C, et al. Least-effort trajectories lead to emergent crowd behaviors [J]. Physical Review E, 2012, 85(1): 016110.
- [20] WANG H, HE Y, WANG Z, et al. Data-driven online traffic reconstructions; interactively optimizing in virtual reality [J]. Computers & Graphics, 2022, 105: 85—93.
- [21] LIAO X C, CHEN W N, GUO X Q, et al. Crowd management through optimal layout of fences; an ant colony approach based on crowd simulation[J]. IEEE Transactions on Intelligent Transportation Systems, 2023, 24(9): 9137—9149.
- [22] LI Y, CHRISTIE M, SIRET O, et al. Cloning crowd motions[C]//Proceedings of the ACM SIGGRAPH/Eurographics Symposium on Computer Animation, 2012: 201—210.
- [23] KIM M, HWANG Y, HYUN K, et al. Tiling motion patches[C]//Proceedings of the ACM SIGGRAPH/Eurographics Symposium on Computer Animation, 2012: 117—126.
- [24] JORDAO K, PETTRE J, CHRISTIE M, et al. Crowd sculpting: a space-time sculpting method for populating virtual environments [C]//Computer Graphics Forum. Wiley Online Library, 2014: 351—360.
- [25] LAI Y C, CHENNEY S, FAN S. Group motion graphs [C]//Proceedings of the 2005 ACM SIGGRAPH/Eurographics Symposium on Computer Animation, 2005: 281—290.
- [26] LERNER A, CHRYSANTHOU Y, LISCHINSKI D. Crowds by example [C]//Computer Graphics Forum. Wiley Online Library, 2007: 655—664.
- [27] CHAO Q, DENG Z, REN J, et al. Realistic data-driven traffic flow animation using texture synthesis [J]. IEEE Transactions on Visualization and Computer Graphics, 2017, 24(2): 1167—1178.
- [28] ZHOU X, KOLTUN V, KRAHENBUHL P. Tracking objects as points [C]//European Conference on Computer Vision, Cham: Springer International Publishing, 2020: 474—490.
- [29] LIN T-Y, DOLLAR P, GIRSHICK R, et al. Feature pyramid networks for object detection [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 2117—2125.
- [30] ZANLUNGO F, IKEDA T, KANDA T. Social force model with explicit collision prediction [J]. Europhysics Letters, 2011, 93(6): 68005.
- [31] DENDORFER P. Mot20: A benchmark for multi object tracking in crowded scenes[J]. arXiv preprint arXiv:2003.09003, 2020.
- [32] BERNARDIN K, STIEFELHAGEN R. Evaluating multiple object tracking performance: the clear mot metrics [J]. EURASIP Journal on Image and Video Processing, 2008, 2008: 1—10.
- [33] LUITEN J, OSEP A, DENDORFER P, et al. Hota: a higher order metric for evaluating multi-object tracking [J]. International Journal of Computer Vision, 2021, 129(2): 548—578.