

基于中心点检测和重识别的多行人跟踪算法

邹北骥, 李伯洲, 刘姝

引用本文:

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.
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.

相似文章推荐 (请使用火狐或IE浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

复杂建筑平面图纸的空间布局融合对偶识别方法

Dual Recognition Method of Spatial Layout Fusion for Complex Architectural Plan Drawings

武汉大学学报·信息科学版. 2021, 46(9): 1354–1361 <https://doi.org/10.13203/j.whugis20210323>

一种基于常梯度模板插值的声线跟踪算法

A Sound Ray Tracking Algorithm Based on Template-Interpolation of Constant-Gradient Sound Velocity

武汉大学学报·信息科学版. 2021, 46(1): 71–78 <https://doi.org/10.13203/j.whugis20180405>

空天网格化星间通视及路由路径规划算法

Aerospace Grid-Based Algorithm of Inter-satellite Visibility and Route Path Planning for Satellite Constellation

武汉大学学报·信息科学版. 2021, 46(1): 50–57 <https://doi.org/10.13203/j.whugis20200234>

一种综合利用图像和光谱信息的物体真假模式识别方法

A Method of True and Fake Objects Pattern Recognition Integrating Image Information and Spectral Information

武汉大学学报·信息科学版. 2019, 44(8): 1174–1181 <https://doi.org/10.13203/j.whugis20190139>

决策树弱分类支持的卫星视频运动检测

Satellite Video Motion Detection Supported by Decision Tree Weak Classification

武汉大学学报·信息科学版. 2019, 44(8): 1182–1190 <https://doi.org/10.13203/j.whugis20180094>

DOI:10.13203/j.whugis20210328



文章编号:1671-8860(2021)09-1345-09

基于中心点检测和重识别的多行人跟踪算法

邹北骥^{1,2} 李伯洲^{1,2} 刘 姝^{1,2}

1 中南大学计算机学院,湖南 长沙,410083

2 湖南省机器视觉与智慧医疗工程技术研究中心,湖南 长沙,410083

摘要:在基于视频的多目标运动跟踪中,目标检测和重识别具有很强的相关性。目前常将目标检测和重识别网络分别进行训练和使用,因此实时跟踪速度不能达到要求。针对多目标跟踪(multiple object tracking, MOT)中行人身份切换和跟踪丢失问题,将行人重识别模块进行遮挡优化并嵌入行人检测网络,由此提出了一种基于中心点检测和重识别的多行人跟踪算法。首先建立了行人运动模型,通过中心点检测得到行人最优状态估计;然后根据深层特征融合的行人重识别模型,利用马氏距离和余弦距离增强行人身份辨别能力;最后利用匈牙利算法进行在线数据关联,同时利用卡尔曼滤波剔除不准确的结果,对未关联的丢失目标做运动预测。利用所提算法和其他跟踪算法分别在MOT15、MOT16、MOT17数据集上进行多行人跟踪对比实验,结果表明,所提算法的多目标跟踪精度(multiple object tracking accuracy, MOTA)分别为63.5、72.4、70.9,正确识别的检测和计算的检测数的比值(identity F1-measure, IDF1)最优,且保证了实时跟踪速率,验证了所提跟踪算法的有效性。

关键词:多行人跟踪;中心点检测;行人重识别;深层特征融合

中图分类号:P231;TP391

文献标志码:A

多行人跟踪一直是计算机视觉领域的研究热点,在监控安防、自动驾驶、场景解析、动作识别等方面具有重要的应用价值^[1-4]。真实场景中光照和尺寸变化,以及行人间频繁遮挡等问题,给多行人跟踪研究带来很大挑战。

传统的多行人跟踪算法通常采用先检测再跟踪的两步法,如Faster RCNN算法^[5]和YOLOv3算法^[6]。两步法通常利用基于卷积神经网络(convolutional neural network, CNN)的行人检测器进行定位,首先需要输入视频序列中的所有行人框,然后将框内裁剪图像输入至下一个行人重识别网络提取特征,通过重识别特征和交并比(intersection over union, IOU)计算距离代价矩阵,最后利用卡尔曼滤波和匈牙利算法将所有行人框关联成轨迹。文献[1]提出SORT(simple online and real-time tracking)算法,使用Faster RCNN进行检测,并利用卡尔曼滤波器对状态进行预测。匈牙利算法基于检测帧位置和IOU进行轨迹跟踪,计算速度快,但未考虑框内的目标特征,因此易发生身份变换。行人重识别可以增加网络对行人消失和遮

挡的鲁棒性,在多目标跟踪(multiple object tracking, MOT)任务中使用CNN在大规模行人数据集上进行训练和提取行人重识别特征,增强了模型对行人身份的辨别能力。文献[2]在多行人跟踪任务中引入行人重识别模块,提出了DeepSORT算法,使用更可靠的深度关联度量来代替距离关联度量。对于短时预测和匹配,DeepSORT引入了有效的距离度量;对于长时丢失的轨迹,行人重识别模块保留了行人外观信息。文献[7]提出的(joint detecting and embedding, JDE)算法将行人重识别模型合并到检测器网络中,摒弃了两步法的通用范式。多行人跟踪被当作一个多任务学习问题^[8],同时输出目标在图像中的边框位置和检测帧中目标的表征嵌入,可加快多行人跟踪速度。但JDE算法在行人相互遮挡情况下检测器效果较差,行人重识别过于依赖检测器的检测结果。

随着多任务学习的发展,已有许多研究提出了基于JDE一体化框架的优良算法,其中多行人跟踪单步法通过检测器加重识别模块嵌入的框架解决了跟踪准确度和实时性的问题^[9]。文献[10]

收稿日期:2021-06-21

项目资助:国家自然科学基金(61902435);湖南省自然科学基金(2019JJ50808)。

第一作者:邹北骥,博士,教授,博士生导师,主要研究方向为计算机视觉、图像处理。bjzhou@csu.edu.cn

通讯作者:刘姝,博士,讲师。slu35@csu.edu.cn

提出了一种基于管道的跟踪 TubeTK 算法,能够实现端到端的训练,将过去单帧图像先检测后跟踪的框架改为多帧图像连成三维的管道数据,包含了帧内、帧间的空间和时间信息,能有效应对行人尺度变化,同时对于行人运动也更具鲁棒性。这种基于管道的模型在遮挡和低可见度条件下表现了良好的跟踪性能,但是 TubeTK 在 MOT15、MOT16、MOT17 数据集上的跟踪速率分别只有 5.8 Hz、1.0 Hz、3.0 Hz,完全无法达到视频实时跟踪要求^[10]。

针对多行人跟踪两步法中的低实时性及行人身份切换频繁的问题,本文在单步法的框架上融合了运动信息和外观信息,将行人重识别模块进行遮挡优化并嵌入行人检测网络中,由此提出了一种基于中心点检测和重识别的多行人跟踪算法。该算法分为中心点检测和行人重识别两个模块,整体框架如图 1 所示,两个模块使用同一个网络共享了部分参数,以提高跟踪速度。同时针对过去的行人重识别相似性距离进行改进,提取到的行人外观信息联合运动信息可提高跟踪的精度。

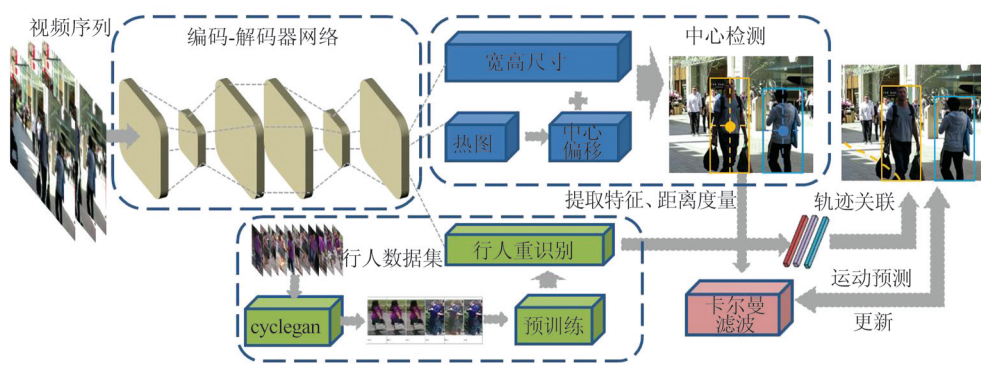


图 1 本文算法流程图

Fig.1 Flowchart of Our Proposed Algorithm

1 基于中心关键点的检测

多行人跟踪的关键在于准确描述行人运动状态,本文基于中心点进行检测,能够精准定位行人,为行人重识别模块提供高质量的行人图像,并且解决了行人重识别特征不对齐的问题。当 S_t 帧进入检测模块,首先通过步长为 2 的卷积和残差层,将特征图的宽高压缩至原来的图像的 1/4;然后通过编码-解码器网络得到 4 个不同维度的分支 $\{\text{heatmap}, \text{boxsize}, \text{offset}, \text{Re-id}\}$,其中检测模块的三大分支如图 2 所示。这 4 个分支分别输出所有行人中心点的位置 $\text{heatmap}_t = \{C_t^1, C_t^2 \dots C_t^N\}$; 所有行人框宽和高 $\text{boxsize}_t = \{z_t^1, z_t^2 \dots z_t^N\}$; offset 用于精炼 heatmap 的行人中心点,提高检测的准确度; $\text{Re-id} = [128, H, W]$ 则是用 128 维特征向量来表征行人。



图 2 中心点检测的三大分支

Fig.2 Three Branches of Center Detection

本文将多行人跟踪看作一个多变量估计问题。给定一个视频序列,将 S_t^i 定义为第 i 号目标行人在第 t 帧时刻的状态,包含中心点位置 $C_t^i = \{x_t^i, y_t^i\}$, 框尺寸 $z_t^i = \{H_t^i, W_t^i\}$, 速度矢量 $V_t^i = [\mu_t^i, \nu_t^i]$ 和加速度矢量 $A_t^i = [\alpha_t^i, \beta_t^i]$, $S_t = \{S_t^1, S_t^2 \dots S_t^N\}$ 表示在第 t 帧所有目标 N 的状态序列, $S^i = \{S_j^i, S_{j+1}^i \dots S_k^i\}$ 表示第 i 号目标在不同帧的状态序列轨迹, S_j^i 和 S_k^i 表示其出现的初始帧和终止帧。某些情况下因遮挡或者目标离开视界又重新进入导致目标的身份切换,例如 S^i 在经过某个遮挡物后未被行人重识别出为 i 号目标,就可能将其新的轨迹定义为 S^{i+1} ,但实际是同一行人,这就是多目标跟踪性能指标中的身份切换(identity definition switch, IDs),错误的身份会导致跟踪错误,本文将行人状态估计与外观模型关联融合,优化了行人身份切换问题,使 IDs 指标降低。

1.1 热图

目标中心点在热图中响应为峰值 1,其他位置响应随着热图与中心点距离的增大呈指数衰减。假定目标中心点个数 $i = N$,真实的框 $G^i = (C^i, z^i)$, 经过网络输出在热图上位置为 $\bar{C}^i = [C^i/4] = (\bar{C}_x^i, \bar{C}_y^i)$, 距离中心点距离 $\Delta d_{(x,y)}^2 = (x -$

$\bar{C}_x^i)^2 + (y - \bar{C}_y^i)^2$, 中心点热图响应 $M_{c^i} = 1$, 其余位置热图响应 $M_{(x,y)} = \sum_{i=1}^N \exp\left(-\frac{\Delta d_{(x,y)}^2}{2\sigma_c^2}\right)$ 。用聚焦损失 focal loss^[11] 的像素级逻辑回归对预测的 heatmap 和真实的 heatmap 损失函数求解, 计算如下:

$$L_{\text{heatmap}} = -\frac{1}{N} \sum_{(x,y)} \begin{cases} (1 - \hat{M}_{(x,y)})^\alpha \log(\hat{M}_{(x,y)}), M_{(x,y)} = 1 \\ (1 - M_{(x,y)})^\beta (\hat{M}_{(x,y)})^\alpha \log(1 - \hat{M}_{(x,y)}), \\ M_{(x,y)} \neq 1 \end{cases} \quad (1)$$

式中, $\hat{M}_{(x,y)}$ 为预测的 heatmap 响应; α 和 β 是 focal loss 已设定的参数。

1.2 中心偏移和行人框尺寸

在行人重识别和检测研究中相互遮挡是最大的难点之一, 检测器的图像并不能达到行人重识别数据集所要求的精确程度, 检测行人不对齐会对后续行人重识别问题造成困难。经过采样得到的特征图会引起量化误差, 少量的量化误差对于行人检测来说并不影响检测性能, 但是对于行人重识别模块, 精准的中心对齐对提取有辨识力的特征非常重要, 需要通过中心点偏移降低量化误差。中心点偏移分支的作用是估计每个像素相对于热图输出中心点的连续偏移量, 消除下采样的误差以更精确地定位对象, 为重识别模块对齐行人提供真实的中心位置。

宽高尺寸分支是对每个中心点行人的尺寸进行估计, 得到包围框的宽高信息。

假定两个分支输出为 $\hat{z}, \hat{o} \in R^{w \times h \times 2}$, $o^i = (C^i/4) - \lfloor \lfloor C^i/4 \rfloor \rfloor = (C_x^i/4 - \bar{C}_x^i, C_y^i/4 - \bar{C}_y^i)$ 表示真实的中心点的映射和热图的中心点位置之间的偏差, 用两个 $L1$ 损失来求解这两个分支特征图, 计算如下:

$$L_{\text{box}} = \sum_{i=1}^N \|o^i - \hat{o}\|_1 + \|z^i - \hat{z}\|_1 \quad (2)$$

2 行人重识别

行人重识别是指从剪裁的框内行人中提取具有辨识性的特征, 当同一行人再次出现时能够准确识别身份。在重识别特征距离可视化图中, 相同身份的行人图像应该是聚在一起的, 理想情况下, 每一个聚类就代表唯一身份的行人。

为了加强相同行人的关联性, 使其在向量空间紧密聚在一起, 同时降低不同行人的关联性,

本文提出一方面加强行人重识别模块对于行人特征的提取能力, 另一方面通过混合马氏距离和余弦距离的相似性度量函数计算代价距离矩阵来增强不同行人的辨识性。

2.1 重识别特征提取

本文在主流的行人重识别数据集 Market1501^[12]、CUHK03^[13] 以及密集人群数据集 CrowdHuman^[14] 上进行训练。由于各个数据集存在域差异, 在真实场景下的测试性能严重下降, 可利用 cycleGAN 风格迁移降低数据域差异^[15]。

行人重识别的输入图像往往是全局图片的一部分, 不同视角距离下同一行人的尺度变化很大, 本文通过有效融合高维特征和低维特征来提升行人重识别特征提取能力。本文选取深层聚合 (deep layer aggregation, DLA) 网络^[16] 作为骨干网络来构建编码解码器网络形式, 如图 3 所示, 采用树状结构融合块内特征和迭代传递相邻块间特征, 逐步加深特征融合, 在有效解决行人尺度和姿态变化的同时也缓解了重识别特征对齐问题。

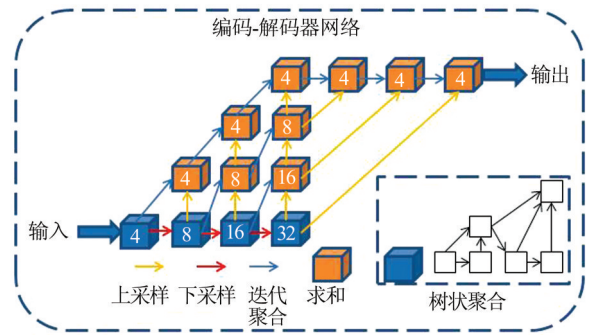


图 3 编码-解码器网络

Fig.3 Encoder-Decoder Network

2.2 相似性度量

行人外观相似性度量是多行人跟踪关联模型的核心, 外观模型的设计要求提取具有鉴别能力的视觉特征, 且满足正样本 (相同目标) 之间外观关联代价小、负样本 (不同目标、或者目标与背景) 之间具有较大的外观关联代价。

不同行人目标之间的相似性通常采用欧氏距离进行度量, 但是欧氏距离忽略了空间域的不同分布, 不能准确反映两者之间的距离。在特征空间中计算某个样本身份类别时, 即便样本与某个身份总体的欧氏距离最近, 也未必属于该身份类别。比如某个身份类别的方差很小, 说明需要非常近才能归为该身份, 故本文选取马氏距离代替欧氏距离进行判别。

马氏距离是标准化后的欧氏距离, 将数据归

一化映射到 $N(0, 1)$ 区间,再求其欧氏距离^[17]。马氏距离更加符合数据特征和实际意义,但仍然不能很好地解决物体长时间遮挡后关联轨迹失败导致的轨迹中断和身份切换问题。余弦距离属于相似性度量方式,便于区分特征向量在特征空间中的角度^[18],可以更好地进行 softmax 损失训练。然而,单一使用余弦距离进行度量学习的问题在于完全忽视了特征模长的影响。因此,本文将马氏距离和余弦距离混合来计算代价矩阵。使用余弦距离衡量不同个体在维度之间的差异,而同一个体中维度间差异由马氏距离进行补偿。两者的混合距离达到一个相对全面的差异性度量,同时考虑了外观信息和运动信息。

2.3 分类器训练

本文通过分类任务来学习行人重识别特征,身份类别数目为 K ,同一个身份的行人被认为是同一个类别,使用 softmax 损失进行训练,损失函数表示如下:

$$L_{id} = - \sum_{i=1}^N \sum_{k=1}^K L^i(k) \log(P(k)) \quad (3)$$

式中, $P(k)$ 为 k 号行人的身份可能性分布; $L(k)$ 为 k 号行人真实的 one-hot 编码。

目标行人通过卷积神经网络计算得到的 128 维特征向量 Re-id 成为外观表征。对于第 i 号行人,跟踪器轨迹保留过去每一次成功跟踪后的所有 Re-id 集合,然后再对第 i 号行人的 Re-id 集合分别与新一帧的每一个行人检测计算最小混合距离。如果判断为同一行人,将新一帧的特征向量 Re-id 加入轨迹的集合中,如图 4 所示,这样就可以产生聚类效果。

本文将检测部分 3 个分支合在一起与重识别分支共同训练,并且用不确定性损失^[19]来自动平衡检测和重识别任务,根据式(1)、(2)、(3)得:

$$L_{total} = \frac{1}{2} \left(\frac{1}{e^{\omega_1}} (L_{heatmap} + L_{box}) + \frac{1}{e^{\omega_2}} L_{id} + \omega_1 + \omega_2 \right) \quad (4)$$

式中, ω_1 和 ω_2 是自动平衡检测和重识别任务权重的参数。

3 行人运动跟踪

视频序列是具有时序信息的连续帧图像,建立行人运动模型,定义行人状态为 x ,包括行人位置 C 、速度 v 、加速度 a 。

假定行人运动为匀速直线运动,即 v 恒定, a

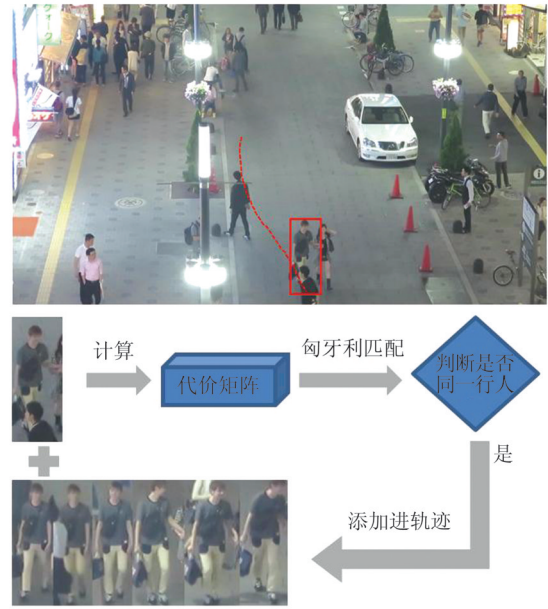


图4 重识别匹配成功流程 判断是否同一行人

Fig.4 Matching Process of Successful Re-identification

为零,则行人在第 t 帧下的状态为:

$$x_t = \begin{bmatrix} C \\ v \\ a \end{bmatrix}_t = \begin{bmatrix} 1 & \Delta t & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} C \\ v \\ a \end{bmatrix}_{t-1} + \omega(t) \quad (5)$$

现实中行人可能因为自身原因而停下脚步或加速,或者因为外部环境影响阻碍绕路,因此行人匀速直线的运动模型不可取。本文通过中心点检测得 C ,通过相邻帧间距离差与时间关联得到速度 v ,从而得到行人观测状态 Z ,而卡尔曼滤波^[20]根据上一时刻的估计状态 $\{C, v, a\}_{t-1}$ 和当前时刻观测状态 Z_t 加权得到行人状态的最优估计。为达到实时的速率要求,在保证准确度情况下采用牛顿运动定律得到 C 和 v, a 之间的变化关系,计算如下:

$$\begin{cases} \begin{bmatrix} C \\ v \\ a \end{bmatrix}_t = \begin{bmatrix} 1 & \Delta t & \frac{\Delta t^2}{2} \\ 0 & 1 & \Delta t \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} C \\ v \\ a \end{bmatrix}_{t-1} + \omega(t) \\ Z_t = [1 \ 0 \ 0] \begin{bmatrix} C \\ v \\ a \end{bmatrix}_t + \nu(t) \\ v(t) = \begin{bmatrix} \frac{\Delta t^5}{20} & \frac{\Delta t^4}{8} & \frac{\Delta t^3}{6} \\ \frac{\Delta t^4}{8} & \frac{\Delta t^3}{3} & \frac{\Delta t^2}{2} \\ \frac{\Delta t^3}{6} & \frac{\Delta t^2}{2} & \Delta t \end{bmatrix} \sigma^2 \end{cases} \quad (6)$$

式中, $\omega(t)$ 为系统噪声; $\nu(t)$ 为观测噪声; σ 为干扰噪声的均方根误差。

将式(6)用卡尔曼滤波一般形式表述,得到:

$$\begin{cases} X'_i = FX_{i-1} + v_{i-1} \\ P'_i = FP_{i-1}F^T + Q \end{cases} \quad (7)$$

式中, F 为行人状态转移矩阵; H 为观测矩阵; Q 、 P 分别为预测、误差的方差矩阵。

卡尔曼滤波迭代更新部分为:

$$\begin{cases} K_i = P'_i H^T (HP'_i H^T + R)^{-1} \\ X_i = X'_i + K_i (Z_i - HX'_i) \\ P_i = (I - K_i H) P'_i \end{cases} \quad (8)$$

式中, Z 为观测值状态矩阵; R 为观测方差矩阵; I 为单位矩阵; X_i 为结合预测状态和观测状态得到的最优估计。基于观测值来不断迭代更新下一时刻状态估计, 达到更精确的跟踪效果。

4 实验结果与分析

4.1 数据集和预训练

MOTChallenge^[21] 是一个公开的 MOT 基准平台, 也是多目标跟踪最常用的评估工具。本文在 MOT15、MOT16、MOT17^[22] 数据集上对提出的跟踪算法进行评估实验。

对于卷积神经网络, 良好的性能需要大量良好的数据支撑。行人重识别数据难以标注身份, 缺少遮挡情况下行人图像, 本文采用多个主流行人数据集混合, 并加入遮挡情况的数据集, 训练得到鲁棒性强的行人重识别模型。在 Market1501^[12]、CUHK03^[13]、CrowdHuman^[14] 数据集上, 用 cyclegan^[15] 风格迁移消除了数据域差异, 得到了高质量的行人图片, 一定程度上弥补了每个行人类别样本较少不均匀的情况。实验模型在经风格迁移处理的行人数据集上用自监督的方法进行了 60 epochs 的预训练行人外观模型, 然后

在 Caltech Pedestrian^[23]、CityPersons^[24]、CUHK-SYSU^[25]、PRW^[26]、ETHZ^[27]、MOT16、MOT17 多目标跟踪数据集进行了 30 epochs 的训练。

4.2 评价指标

多行人跟踪由于涉及大量的行人身份, 无法使用目标跟踪的评价指标, 也难以使用单一指标来评估其性能。本文选用 MOTChallenge 定义的评价指标^[21]: (1) 多目标跟踪精度 (multiple object tracking accuracy, MOTA), 综合考虑了轨迹中的虚检、漏检、身份切换的现象; (2) 正确识别的检测和计算的检测数的比值 (identity F1-measure, IDF1), 综合考虑了身份准确率和身份召回率; (3) 主要跟踪目标百分比 (mostly tracked, MT), 超过 80% 被成功跟踪的轨迹与所有轨迹的比值; (4) 主要丢失目标百分比 (mostly lost, ML), 低于 20% 被成功跟踪的轨迹与所有轨迹的比值; (5) IDs, 属于轨迹中断的一种情况, 跟踪器出现了误判, 偏移到其他轨迹上然后身份重识别错误。其中, MOTA、IDF1、MT 越高越好, ML 和 IDs 则越低越好。MOTA 和 IDF1 这两个综合指标尤为重要, 前者可以反映 IDs 的现象, 后者对轨迹身份信息更敏感。

4.3 结果与分析

本文将平均任意两个行人框交并比大于 0.5 的人群定义为拥挤人群, 图 5 为从第 1 帧至 800 帧的跟踪结果, 拥挤人群 91 号至 96 号行人身份没有发生切换。在整幅图片左上角有大量拥挤人群, 存在严重的行人互相遮挡问题, 基于候选框回归的检测器会产生大量误报, 导致后续跟踪性能不好。由图 5 可知, 当行人相互遮挡时, 本文方法仍然可以通过高质量的行人重识别分配正确的行人身份, 在拥挤场景下具有良好的鲁棒性。

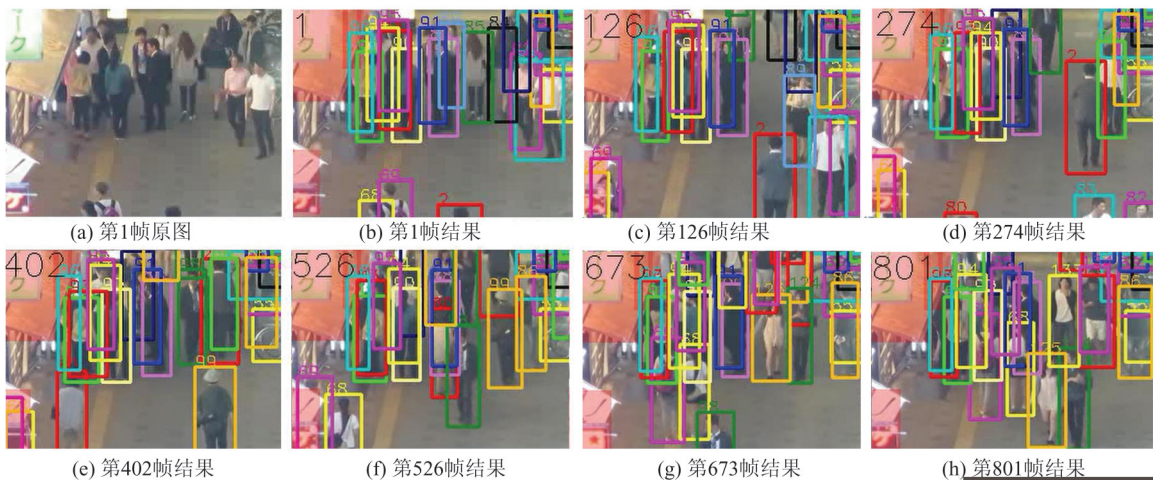


图 5 MOT17 的局部拥挤人群跟踪结果

Fig.5 Results of Crowded Human Tracking of MOT17

表1列出了本文算法和另外3种传统的多行人跟踪算法的性能指标,以及进行多行人跟踪实验得到的结果评价。其中离线算法表示可以利用整个视频序列的信息,包括未来帧;接近在线算法表示可以使用部分未来帧;在线算法则是在不依赖未来帧的信息下进行跟踪,更符合应用需求。由表1可知,相较于各种类型的两步法,本文算法跟踪结果的MOTA和IDF1均有显著提升。此外,在添加了辅助运动跟踪部分和针对遮挡人群的行人重识别特别训练后,相较于单独核心部分,本文算法跟踪结果的MOTA、IDF1分别有7.3%和7.6%的提升,说明本文算法能有效提升网络对于身份的辨识能力。

表1 不同类型算法的跟踪结果对比

Tab.1 Comparison of Results of Different Algorithms

算法	类型	MOTA	IDF1
JCC算法 ^[28]	离线	51.1	54.4
NT算法 ^[29]	接近在线	47.5	43.6
DeepSORT算法 ^[2]	在线	61.4	62.2
单独核心部分	在线	67.5	69.8
本文	在线	72.4	75.1

此前,多目标跟踪在性能上的提升往往依赖于目标检测算法的进步,而忽视了目标运动在空间和时间上的连续性。基于两步法的多行人跟踪同样依赖于检测器效果,如图6所示,18帧时左下角第80号女生检测失败,进而导致了跟踪轨迹中断。本文算法通过辅助部分卡尔曼滤波利用运动信息联合检测加行人重识别,跟踪框一直都能够紧紧地跟踪成功,也没有出现身份切换。同样的,本文算法也成功跟踪了图6正下方马路上第5号半身男子。

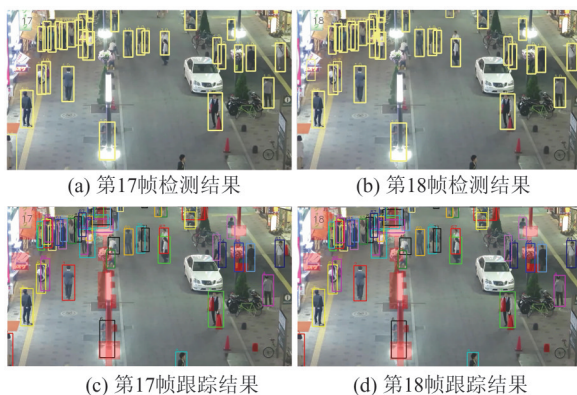


图6 检测失败情况下跟踪成功

Fig.6 Successful Tracking in the Case of Detection Failure

行人重识别可以迅速地判断一个行人的身份,当出现新的目标行人,也能识别出与已有行人库不匹配。图7(a)、7(b)分别为171、212帧新进入的第122和123号行人。当从未出现过的行人进入视界,跟踪器会迅速判断其与已有轨迹之间的差异,并为其分配新的轨迹。



图7 迅速分配新的ID轨迹

Fig.7 Quick Assignment for a New ID

表2、3、4为本文算法与最先进的几种在线跟踪方法分别在MOT15、MOT16、MOT17测试集上的跟踪结果比较。对于传统的两步法,本文计算速率时加入了检测器需要的时间,其中加粗字体表示最佳指标。

表2 MOT15测试集上的跟踪结果

Tab.2 Results of MOT15 Test Set

算法	类型	MOTA	IDF1	MT /%	ML /%	IDs	帧率 /Hz
EAMTT ^[30]	两步法	53.0	54.0	35.9	19.6	7 538	4.0
AP_HWD-PL ^[31]	两步法	53.0	52.2	29.1	20.2	708	6.7
RAR ^[32]	两步法	56.5	61.3	45.1	14.6	428	3.4
TubeTK ^[10]	单步法	58.4	53.1	39.3	18.0	854	5.8
FairMOT ^[33]	单步法	60.6	64.7	47.6	11.0	591	30.5
本文算法	单步法	63.5	65.5	50.1	10.4	504	32.7

表3 MOT16测试集上的跟踪结果

Tab.3 Results of MOT16 Test Set

算法	类型	MOTA	IDF1	MT /%	ML /%	IDs	帧率 /Hz
SORT ^[1]	两步法	59.8	53.8	25.4	22.7	1 423	8.6
DeepSORT ^[2]	两步法	61.4	62.2	32.8	18.2	781	6.4
RAR ^[32]	两步法	63.0	63.8	39.9	22.1	482	1.4
POI ^[34]	两步法	66.1	65.1	34.0	20.8	805	5.0
JDE ^[7]	单步法	64.4	55.8	35.4	20.0	1 544	18.5
TubeTK ^[10]	单步法	64.0	59.4	33.5	19.4	1 117	1.0
FairMOT ^[33]	单步法	74.9	72.8	44.7	15.9	1 074	25.9
本文算法	单步法	72.4	75.1	45.8	17.2	890	30.6

表 4 MOT17 测试集上的跟踪结果

Tab.4 Results of MOT17 Test Set

算法	类型	MOTA	IDF1	MT	ML	IDs	帧率 /Hz
				/%	/%		
SST ^[35]	两步法	52.4	49.5	21.4	30.7	8 431	3.9
TubeTK ^[10]	单步法	63.0	58.6	31.2	19.9	4 137	3.0
Center-Track ^[36]	单步法	67.3	59.9	34.9	24.8	2 898	22.0
Fair-MOT ^[33]	单步法	73.7	72.3	43.2	17.3	3 303	25.9
本文算法	单步法	70.9	78.3	40.6	16.4	1 978	29.7

从表 2~4 中可以看出,本文算法在 MOT15 数据集上达到了最佳跟踪模型。在 MOT15、MOT16 和 MOT17 3 个数据集上, IDF1 和帧率都达到了最佳水平,表明本文算法对于行人身份具有较高的辨别能力,能够快速分配正确的行人身份。热图的输入使跟踪器更好地保存来自前一帧的轨迹,这提高了跟踪的命中率;中心偏移的提炼使命中的轨迹更接近真实轨迹,使得 MT 指标有所提高。

5 结 语

本文提出的多行人跟踪算法属于单步法在线跟踪,不使用未来帧的信息即可预测行人轨迹。实验结果表明,本文对行人重识别模块的优化使得行人身份切换大大减少,并且两个网络一体化使得跟踪速度达到了视频实时速率。然而,本文在建立运动模型时忽视了行人的碰撞体积,没有建立行人交互模型和排斥模型,对行人突然变向加速预测不准,容易导致轨迹丢失。未来的工作将围绕拥挤人群遮挡和强光照变化等难点展开。

参 考 文 献

- [1] Bewley A, Ge Z, Ott L, et al. Simple Online and Realtime Tracking[C]. International Conference on Image Processing, Phoenix, Arizona, USA, 2016
- [2] Wojke N, Bewley A, Paulus D. Simple Online and Realtime Tracking with a Deep Association Metric [C]. International Conference on Image Processing, Beijing, China, 2017
- [3] Chen Enjia, Tang Xianghong, Fu Bowen. Pedestrian Search Method Based on Faster R-CNN with the Integration of Pedestrian Detection and Re-identification[J]. *Journal of Computer-Aided Design & Computer Graphics*, 2019, 31(2): 332-339 (陈恩加, 唐向宏, 傅博文. Faster R-CNN 行人检测与再识别为一体的行人检索算法[J]. *计算机辅助设计与图形学学报*, 2019, 31(2): 332-339)
- [4] Hou Jianhua, Zhang Guoshuai, Xiang Jun. Designing Affinity Model for Multiple Object Tracking Based on Deep Learning [J]. *Acta Automatica Sinica*, 2020, 46(12): 2 690-2 700 (侯建华, 张国帅, 项俊. 基于深度学习的多目标跟踪关联模型设计[J]. *自动化学报*, 2020, 46(12): 2 690-2 700)
- [5] Ren S, He K, Girshick R, et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2016, 39(6): 1 137-1 149
- [6] Redmon J, Divvala S, Girshick R, et al. You Only Look Once: Unified, Real-Time Object Detection [C]. Computer Vision and Pattern Recognition, Las Vegas, USA, 2016
- [7] Wang Z, Zheng L, Liu Y, et al. Towards Real-Time Multi-Object Tracking [C]. European Conference on Computer Vision, Glasgow, Scotland, UK, 2020
- [8] Kokkinos I. UberNet: Training a Universal Convolutional Neural Network for Low-, Mid-, and High-Level Vision Using Diverse Datasets and Limited Memory [C]. Computer Vision and Pattern Recognition, Hawaii, USA, 2017
- [9] Voigtlaender P, Krause M, Osep A, et al. MOTs: Multi-Object Tracking and Segmentation [C]. Computer Vision and Pattern Recognition, Los Angeles, USA, 2019
- [10] Pang B, Li Y, Zhang Y, et al. TubeTK: Adopting Tubes to Track Multi-Object in a One-Step Training Model [C]. Computer Vision and Pattern Recognition, Seattle, Washington, USA, 2020
- [11] Lin T Y, Goyal P, Girshick R, et al. Focal Loss for Dense Object Detection [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020, 42(2): 318-327
- [12] Zheng L, Shen L, Tian L, et al. Scalable Person Re-identification: A Benchmark [C]. International Conference on Computer Vision, Santiago, Chile, 2015
- [13] Li W, Zhao R, Xiao T, et al. DeepReID: Deep Filter Pairing Neural Network for Person Re-identification [C]. Computer Vision and Pattern Recognition, Columbus, USA, 2014
- [14] Shao S, Zhao Z, Li B, et al. CrowdHuman: A Benchmark for Detecting Human in a Crowd [J]. *Arxiv Preprint Arxiv*, 2018, DOI:1805.00123
- [15] Zhu J Y, Park T, Isola P, et al. Unpaired Image-to-

- Image Translation Using Cycle-Consistent Adversarial Networks [C]. International Conference on Computer Vision, Venice, Italy, 2017
- [16] Yu F, Wang D, Shelhamer E, et al. Deep Layer Aggregation [C]. Computer Vision and Pattern Recognition, Salt Lake City, Utah, USA, 2018
- [17] Roth P M, Hirzer M, Köstinger M, et al. Mahalanobis Distance Learning for Person Re-identification [M]. London: Springer Press, 2014
- [18] Wojke N, Bewley A. Deep Cosine Metric Learning for Person Re-identification [C]. Winter Conference on Applications of Computer Vision, Lake Tahoe, USA, 2018
- [19] Kendall A, Gal Y, Cipolla R. Multi-task Learning Using Uncertainty to Weigh Losses for Scene Geometry and Semantics [C]. Computer Vision and Pattern Recognition, Salt Lake City, Utah, USA, 2018
- [20] Welch G, Bishop G. An Introduction to the Kalman Filter [J]. *Proc of Siggraph*, 1995, 41 (8): 127-132
- [21] Leal-Taixé L, Milan A, Reid I, et al. MOTChallenge 2015: Towards a Benchmark for Multi-Target Tracking [J]. *Arxiv Preprint Arxiv*, 2015, DOI: 1504.01942
- [22] Milan A, Leal-Taixé L, Reid I, et al. MOT16: A Benchmark for Multi-Object Tracking [J]. *Arxiv Preprint Arxiv*, 2016, DOI: 1603.00831
- [23] Dollár P, Wojek C, Schiele B, et al. Pedestrian Detection: A Benchmark [C]. Computer Vision and Pattern Recognition, Miami, Florida, USA, 2009
- [24] Zhang S, Benenson R, Schiele B. CityPersons: A Diverse Dataset for Pedestrian Detection [C]. Computer Vision and Pattern Recognition, Honolulu, Hawaii, USA, 2017
- [25] Xiao T, Li S, Wang B, et al. Joint Detection and Identification Feature Learning for Person Search [C]. Computer Vision and Pattern Recognition, Honolulu, Hawaii, USA, 2017
- [26] Zheng L, Zhang H, Sun S, et al. Person Re-identification in the Wild [C]. Computer Vision and Pattern Recognition, Honolulu, Hawaii, USA, 2017
- [27] Ess A, Leibe B, Schindler K, et al. A Mobile Vision System for Robust Multi-Person Tracking [C]. Computer Vision and Pattern Recognition, Anchorage, Alaska, USA, 2008
- [28] Keuper M, Tang S, Zhongjie Y, et al. A Multi-cut Formulation for Joint Segmentation and Tracking of Multiple Objects [J]. *Arxiv Preprint Arxiv*, 2016, DOI: 1607.06317
- [29] Wen L, Du D, Li S, et al. Learning Non-Uniform Hypergraph for Multi-Object Tracking [C]. the AAAI Conference on Artificial Intelligence, Hawaii, USA, 2019
- [30] Sanchez-Matilla R, Poiesi F, Cavallaro A. Online Multi-target Tracking with Strong and Weak Detections [C]. European Conference on Computer Vision, Amsterdam, Netherlands, 2016
- [31] Chen L, Ai H, Shang C, et al. Online Multi-Object Tracking with Convolutional Neural Networks [C]. International Conference on Image Processing, Beijing, China, 2017
- [32] Fang K, Xiang Y, Li X, et al. Recurrent Autoregressive Networks for Online Multi-object Tracking [C]. Winter Conference on Applications of Computer Vision, Lake Tahoe, USA, 2018
- [33] Zhang Y, Wang C, Wang X, et al. FairMOT: On the Fairness of Detection and Re-identification in Multiple Object Tracking [J]. *Arxiv Preprint Arxiv*, 2020, DOI: 2004.01888
- [34] Yu F, Li W, Li Q, et al. POI: Multiple Object Tracking with High Performance Detection and Appearance Feature [C]. European Conference on Computer Vision, Amsterdam, Netherlands, 2016
- [35] Sun S J, Akhtar N, Song H S, et al. Deep Affinity Network for Multiple Object Tracking [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019, 43(1): 104-119
- [36] Zhou X, Koltun V, Krähenbühl P. Tracking Objects as Points [C]. European Conference on Computer Vision, Glasgow, Scotland, UK, 2020

A Multi-Pedestrian Tracking Algorithm Based on Center Point Detection and Person Re-identification

ZOU Beiji^{1,2} LI Bozhou^{1,2} LIU Shu^{1,2}

¹ School of Computer Science and Engineering, Central South University, Changsha 410083, China

² Hunan Engineering Research Center of Machine Vision and Intelligent Medicine, Changsha 410083, China

Abstract: Objectives: In video-based multiple object tracking, the object detection and re-identification have a strong correlation. The existing methods generally train the object detection and re-identification networks separately, which makes the tracking speed fail to meet the real-time requirements. In this paper, we integrate the detection and re-identification into one network to accelerate the tracking process, and also solve the problems of identity switching and failure tracking. **Methods:** This paper develops a pedestrian motion model and obtains the optimal state estimation of pedestrians using center point detection. The person re-identification model with deep layer features uses the Mahalanobis distance and cosine distance to enhance the ability of person identification. And the Hungary algorithm is used for data online association, where the state estimation results become more accurate using Kalman filtering, and the unrelated lost objects are predicted by motion. **Results:** Experiments are conducted on MOT15, MOT16 and MOT17 datasets using the proposed algorithm and other multi-pedestrian tracking algorithms, and the multiple object tracking accuracy of tracking results using our proposed algorithm is 63.5, 72.4 and 70.9, respectively, and the identity F1-measure is optimal, with the real-time rate. **Conclusions:** The proposed algorithm can accelerate the tracking speed by network parameter sharing, and improve the recognition accuracy by person re-identification training.

Key words: multi-pedestrian tracking; center point detection; person re-identification; deep feature aggregation

First author: ZOU Beiji, PhD, professor, specializes in computer vision and image processing. E-mail: bjzou@csu.edu.cn

Corresponding author: LIU Shu, PhD, lecturer. E-mail: sliu35@csu.edu.cn

Foundation support: The National Natural Science Foundation of China(61902435); the Natural Science Foundation of Hunan Province (2019JJ50808).

引文格式: 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. DOI:10.13203/j.whugis20210328(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. DOI:10.13203/j.whugis20210328)