

文章编号: 2095-2163(2024)01-0152-04

中图分类号: TP391

文献标志码: A

基于 YOLO 目标检测算法的人群多目标识别跟踪方法

张四平

(湖南信息职业技术学院 经济管理学院, 长沙 410200)

摘要: 目前对人群跟踪方法主要是建立跟踪模型, 实现人群群体检查和跟踪, 但是无法实现人群中多行人的识别和个体跟踪, 造成人群跟踪与多目标识别存在效率低下和不准确。本文提出基于 YOLO 目标检测算法的人群多目标识别跟踪方法, 通过对人群多目标的可见特征提取人群的行人轨迹和外观特征, 实现人群多目标识别的跟踪。实验结果表明, 该方法提高了人群多目标的识别效率, 具有一定的实用性。

关键词: YOLO 目标检测算法; 人群多目标跟踪; 识别方法

A crowd multiple target recognition and tracking method based on YOLO object detection algorithm

ZHANG Siping

(School of Economics and Management, Hunan College of Information, Changsha 410200, China)

Abstract: The current tracking method for crowd mainly establishes a tracking model to realize crowd group checking and tracking, but it cannot realize the identification of multiple pedestrians in the crowd and individual tracking, resulting in inefficiency and inaccuracy of crowd tracking and multi-target recognition. This paper proposes a crowd multi-target recognition tracking method based on YOLO target detection algorithm, which realizes the tracking of crowd multi-target recognition by extracting the pedestrian trajectory and appearance features of the crowd through the visible features of the crowd multi-target. The experimental results show that the method improves the efficiency of crowd multi-target recognition and has certain practicality.

Key words: YOLO object detection algorithm; multi target tracking of crowds; identification method

0 引言

为适应社会对安全的需要, 保障人民的平安出行, 设置了许多摄像头进行监控。针对摄像头拍摄的海量视频, 若要在密集的人流中对多个对象进行监视, 则需要由看守人来完成, 需要耗费巨大的人力物力; 此外, 在实际时间中很难找到准确的对象, 数据处理效率的下降^[1]。从广泛的监控摄像机数据中快速准确地确定目标是当前计算机视觉领域的主要研究内容。目前人脸识别技术广泛应用于智慧城市的建设, 但城市网络监控的人脸识别存在问题, 如超过监控范围画面模糊难以辨别、监控盲区、监控相机依赖光照强度、拍摄死角多等。通过对人群目标行人的特征以及后续行人行动的重新确认, 可以弥补在复杂场景中的具体目标识别的不足。跟踪目标可以通过对视频序列中特定目标信息预测未来可

能存在的图像位置, 在同一摄像机中找到目标特征图像进行识别并恢复, 最后可以基于行人的衣着、身体形态和其他人脸特征来进行人脸识别^[2]。对于一些比较复杂的情况, 如远距离摄影、不同场景图像等, 利用多个视频序列来追踪特定的行人。当监控目标从一个视频中短暂失踪, 可以利用多个监控视频序列进行算法演算, 形成一个完整的目标行人识别, 最终实现识别跟踪。一个完整的识别系统由3个主要模块构成: 数据采集模块、特征提取与表示模块和模型训练与推理模块^[3]。该系统能够实现对特定行人的追踪, 但是对人群中多目标识别跟踪, 则可以利用行人识别技术、融合行人运动轨迹, 从而实现多目标的准确识别。本文提出的基于 YOLO 目标检测算法的人群多目标识别跟踪方法, 可以很好实现多目标识别跟踪。

基金项目: 湖南省教育厅科研项目(22C1171)。

作者简介: 张四平(1979-), 男, 硕士, 副教授, 主要研究方向: 图像处理、人工智能技术。

收稿日期: 2023-09-24

哈尔滨工业大学主办 ◆ 专题设计与应用

1 基于YOLO目标检测算法的密集人群多目标跟踪识别方法设计

1.1 行人目标位置识别

为了提高模型的准确性,采用特征工程、增加数据量,调整模型参数和对数据进行清洗。构建基于图像的状态转换和监控模式的图像分类体系,并利用目标的运动轨迹和不同场景监控图像来获取目标图像的信息,获得更为精确、完整的图像信息^[4]。采用基于ResNeXt50的多个目标追踪技术,并将其应用于基于YOLO的多个目标追踪中。通过多场景摄像头监控,对目标动态位置进行确定,最后得到目标图像的特征。假设 x_k 是 k 周期内系统动态位置的一个分量, z_k 是监控系统的一个组建,则目标系统的过渡模型和监测模型可表示为式(1)和式(2):

$$x_{k+1} = f(x_k, u_k) \quad (1)$$

$$z_k = h(x_k, v_k) \quad (2)$$

其中, u_k 表示系统噪声; v_k 表示监测噪声; $f(x_k, u_k)$ 和 $h(x_k, v_k)$ 分别是系统的状态过渡模型和系统监控模型。

如果两个模型的监控图像概率和监控几率发生改变,动态评估可以通过监控动态进行预期判断和更新,采用最佳贝叶斯评估结果。图像生成概率 $p(x | x_{k-1})$,式(3):

$$p(x | x_{k-1}) = p(x_{k-1} | z_{1:k-1}) dx_{k-1} \quad (3)$$

其中, dx_{k-1} 为图像不同场景获取率。

检测几率的变化 $p(x | z_{1:k-1})$,式(4):

$$p(x | z_{1:k-1}) = \frac{P(z_k | x_k) p(x_k | z_{1:k-1})}{p(z_k | z_{1:k-1})} \quad (4)$$

其中, k 为检测动态位置权值。

该方法使用部分加权样本量来反映所需的最大图像可能密度。模型图像中的图像监控关键点应扣除每个输入目标图像的权重,对部分图像信号进行加权,反映所需的目标图像的有效信道,使该监控区域内的信号通道能够增强,减少低效信道。图像信号 $M_c(F)$,式(5):

$$M_c(F) = \sigma z_k (MLP(\text{AvgPool}(F)) + MLP(MLP(\text{MaxPool}(F)))) \quad (5)$$

其中, σ 为行人的不同场景出现的次数; F 为采用最大合并值获取图像信号在空间维度上的值; $MLP(\text{MaxPool}(F))$ 为获取合并最大值。

用 w_k^i 表示归一化权重,对图像经过颗粒过滤之后,在有效信道 ζ 中列出 x_{k-1} 场景内包含行人的图像 ρ ,式(6):

$$\rho = \sum w_k^i + \zeta(x_k - x_{k-1}) \times C(g^0) + F(g^1) \quad (6)$$

其中, C 为密度函数; g^0 为颗粒密度; g^1 为颗粒阈值。

为确定行人在图像中的实际位置,通过最大比合并来判断目标所在区域,使用Sigmoid函数解决多标签问题,实现行人目标位置识别,实现公式(7):

$$\rho_r = \frac{1}{Q} + \sum w_k^i + F[\zeta(x_k - x_{k-1})] \times \text{Sigmoid}(v) \quad (7)$$

其中, Q 为概率分布, v 为原始输出值。

1.2 行人多目标识别特征提取

设包含行人目标的位置所在区域为低维子区域,基于卷积神经网络从包含行人目标位置的图像中提取出可能包含行人多目标的低维子区域。使用区域提取器从图像中提取出多个候选区域。 $T(\rho_r)$ 为包含行人目标位置的候选区域,式(8):

$$T(\rho_r) = \frac{F \times T(a_n t_n)}{a_n} + \varepsilon \quad (8)$$

其中, $T(a_n t_n)$ 为 a_n 范围内包含 t_n 个行人目标的总区域; F 为模型图像; ε 为区域维度。

包含行人目标的数据样本可用 z 来表示, D 为目标类型,由于每个图像中包含行人目标位置的大小不同,因此需要对图像中行人目标位置的大小进行归一化处理。以图像标准的最小化为依据, γ 表示处理后行人目标图像,式(9):

$$\gamma = T(\rho_r) [\min \| Da - z \|_2 + \lambda \| \alpha \|] \quad (9)$$

其中, λ 表示图像标准化参数, α 表示图像转换系数。

图像归一化后出现残差的情况可能会导致图像的质量下降,从而影响后续的图像处理。为了避免这种情况,可以通过测量图像相似性来度量图像的尺度。

灰度直方图是一个用于描述图像灰度级分布的统计量,反映了图像中各个灰度级的像素点个数。将图像转换为灰度图像,并计算其灰度直方图 E ,式(10):

$$E = T(\rho_r) \times F(\gamma) \frac{F \times F(V)}{a_n(1-X)} \quad (10)$$

其中, X 表示像素点个数, V 表示灰度级。

根据灰度直方图构建直方图矩阵 \ddot{E} ,对于每对图像,将其灰度直方图分别作为矩阵的行和列,计算直方图矩阵中的元素值 χ ,式(11):

$$\chi = T(\rho_r) \times \sum_{\max V} F(v) \times \frac{\ddot{E}(y, z)}{(b_m)} \quad (11)$$

其中, y, z 表示直方图矩阵 \vec{E} 的行和列; b 表示直方图矩阵中元素; m 表示图像空间维度。

在直方图矩阵中, 每个元素 b 可表示两个包含行人目标位置图像在该灰度级 V 上的相似程度, 因此可以将 b 值视为该灰度级上的坐标。使用欧几里得距离计算公式(12), 计算元素值之间的差异, 得到每对图像之间的相似度。

$$D = \text{sqrt}[(y_1 - z_1)^2 + (y_2 - z_2)^2 + \dots + (y_m - z_m)^2] \quad (12)$$

其中, y_1, y_2, \dots, y_m 和 z_1, z_2, \dots, z_m 表示两个点在 m 维空间中的坐标。

在研究多目标跟踪时, 观测矩阵用于描述目标的位置和速度的测量值, 以及监控设备测量结果之间的关系。在实际应用中, 观测矩阵通常是由一系列测量值组成的矩阵, 其中每一行表示一次测量, 每一列表示一个特定的监控设备的输出。

通过观测矩阵, 可以将包含行人目标位置图像中多目标进行整合, 以便更好地描述目标的位置, 实现特征提取。

将每个行人的观测结果组成一个观测向量, 并将这些向量组合成观测矩阵。 $G_{(i,j)}$ 表示行人 i 在第 j 时刻下的观测值, 设图像通道数为 k , 通道包括颜色、边缘和形状等多个通道。在这种情况下, 观测矩阵中的每个元素可以表示为一个观测向量中的一个通道的值。 $K(i,j)$ 表示行人 i 在第 j 时刻所处通道值, 式(13):

$$K(i,j) = \frac{1}{k} A(i) \times I \quad (13)$$

其中, A 表示行人移动速度, I 表示观测数据。

利用观测矩阵中的数据, 识别行人轨迹的边缘信息 $H(i,j)$, 及行人多目标识别特征 θ 提取, 式(14)和式(15)。

$$H(i,j) = \frac{A(i) \times I}{k(\varphi)} \times (1 - g) \quad (14)$$

$$\theta = \frac{k(\varphi)}{k \times H(i,j)} A(i) \times I \quad (15)$$

其中, φ 表示像素低点梯度方向, g 表示梯度幅值。

1.3 密集人群检测和跟踪模型

YOLO 跟踪检测是将图像划分为固定大小的网格单元, 每个单元负责检测图像中的对象, YOLO 预测多个边界框(Bounding Box)和每个边界框对应的类别概率。首先, 对图像进行特征详细提取; 其次, 在浅层特征中输入详细特征进行深度提取; 最后, 在

预测特征中输入深度特征, 获得最终测试结果。

基于改进 YOLO 的密集人群目标跟踪方法对图像进行处理, 然后利用改进的 YOLO 方法提取图像中各帧图像中的物体的位置及特性; 在此基础上, 采用卡尔曼滤波器对其进行估计, 并根据本帧与上一幅图像中的运动目标的距离, 求出运动目标间的相对距离; 利用目前探测到的图像与轨迹中的多条弧线, 利用余弦距离求出最短的一段, 以此来确定运动目标与轨迹间的特征距离; 通过对多幅图像间的距离信息进行比对, 以判定不同图像中的物体是不是相同的。根据运动目标的定位及运动物体的特性, 如果符合, 就将其添加至运动轨迹中, 添加一个跟踪, 如果不符合, 再次进行运动, 建立一个新的跟踪。在多个目标追踪方法中, 除非有特殊规定, 否则都是采用多个目标追踪过程来追踪。

1.4 人群多目标跟踪识别

将视频序列引入到 YOLO 检测, 获得具有行人外观特征的检查框, 通过基于 YOLO 跟踪识别算法完成多目标特征的关联和兼容。本文的密集人群多目标跟踪识别步骤如下:

步骤 1 通过 YOLO 检测器输入视频序列, 以提取和深化目标行人特征;

步骤 2 提取行人识别特征, 增加行人检测, 包括目标坐标、所在不同场景判断, 构成新的公开检查数据, 连接行人轨迹的特征, 以完成行人的目标检测;

步骤 3 基于 YOLO 算法, 将当前主要目标发现框划分为聚合物公开框和低公开框;

步骤 4 执行聚合物公开框对应的中间通联和预测框的结果, 计算公开框与预测框的相似度;

步骤 5 将提取的行人特征与相似度相结合, 使用 YOLO 算法获得结果即成功协调行人、多场景识别和目标轨迹。对新轨迹执行下次关联, 对公开框和预测框之间进行初始关联;

步骤 6 添加步骤 5 中未匹配的目标轨迹, 第二次使用公开框预测, 进行 3 次检索;

步骤 7 将步骤 6 中检索结果放回 YOLO 算法中, 以更新与场景不匹配的少数不匹配的行人特征信息, 若连续 30 帧速率与打开的框不一致, 则删除行人特征信息;

步骤 8 使用深度学习算法预测跟踪序列, 再依次与高分检测框、低分检测框进行关联匹配, 重复上述步骤。

2 实验论证

2.1 数据集

数据集是由部署在 7.5 m×11 m 长方形区域内的 4 个摄像头记录下的关于 9 个人活动轨迹的数据构成,场景非常拥挤,大约 5 000 帧视频。数据集的参数设置如下:基于外观结构的最近邻目标数量 $k = 25$,颜色直方图类似度的阈值 $\tau = 0.85$,搜索基于外观的最近邻的图像边框时间为 $T = 8$ s,行人的最快行走速度为 $v = 3$ m/s,最大定位误差为 $\delta = 125$ cm。另外,设行人步长 $P = 1.05$,即 $p^{(m+1)} \leftarrow p^{(m)}/1.05$,一共有 $M = 94$ 个不同的 p 值, $p^{(m)} = 0.01$ 。

2.2 评价标准

本文所有实验都位于同一环境中,采用多目标调查领域常用的评价指标多目标跟踪精度 (MOTA) 和多目标识别点 (IDF) 进行评价,式 (16) 和式 (17):

$$MOTA = 1 - \frac{N_{FN} + N_{FP} + N_{IDs}}{MN_{GT}} \quad (16)$$

其中, N_{GT} 表示搜索时间 T 内的边界的框数量; N_{FN} 表示多个视频中的目标遗漏数; N_{FP} 表示整个视频的误检数; N_{IDs} 表示行人 ID 的切换次数。

$$MOTP = \frac{\sum d_i^t}{\sum c_i} \quad (17)$$

其中, t 表示当前帧为第 t 帧, $t \in [1, N]$; d_i^t 表示第 t 帧中第 i 个预测框与真实框之间的重叠率; c_i 是成功识别目标的数量。

2.3 对比实验

在数据集上获得的跟踪结果快照如图 1 所示。

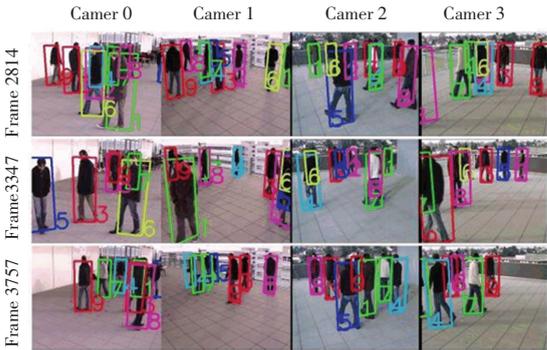


图 1 在数据集上获得的跟踪结果快照

Fig. 1 Snapshot of the tracking results obtained on the dataset

将本文中提出的算法与传统的算法 CenterTrack 和 fair MOT 进行对比实验,结果见表 1。

表 1 不同方法对比实验结果

Table 1 Comparative experimental results with different algorithm

跟踪算法	MOTA (跟踪准确度)	MOTP (跟踪精度)	FPS (帧/s)
CenterTrack	67.8	64.7	17.5
Fair MOT	73.7	72.3	25.9
本文方法	80.3	77.3	29.6

相比于传统方法,模型参数量不变的情况下,本文基于 YOLO 算法密集人群多目标跟踪识别方法跟踪更准确、精度更高、速度更快,识别效率更高。

3 结束语

本文基于 YOLO 目标检测算法,设计一种人群多目标识别跟踪方法。通过对人群多目标的可见特征提取,采用目标检测算法构建跟踪模型,获取人群的行人轨迹和外观特征,通过检测对比最终实现人群多目标识别的跟踪方法。仿真对白实验结果证明,此算法提高了人群多目标的识别效率,具有一定的实用性。

参考文献

- [1] 赵安新,杨金桥,杨浩波,等. 基于改进 DeepSORT 和 FastReID 的室内多目标人员跨镜识别与跟踪[J]. 西安科技大学学报, 2023, 43(3): 622-630.
- [2] 杜觉晓,赵留学,莽修伟,等. 基于边缘端目标识别与跟踪算法的输电线路航拍视频流诊断技术研究[J]. 智慧电力, 2023, 51(6): 106-113.
- [3] 陈光,乔梁,黄晓明,等. 基于目标跟踪的行为识别方法研究——以安全帽佩戴识别检测为例[J]. 黑龙江科学, 2023, 14(8): 50-52.
- [4] 查文文,潘伟豪,陈成鹏,等. 基于边缘计算与改进 YOLOv5 的群养生猪姿态识别及跟踪研究[J]. 东北农业大学学报, 2023, 54(3): 83-96.