

DOI:10.13232/j.cnki.jnju.2021.02.007

## 一种多目标跨摄像头跟踪技术研究与实践

戴臣超, 王洪元\*, 曹 亮, 殷雨昌, 张 继

(常州大学计算机与人工智能学院, 阿里云大数据学院, 常州, 213164)

**摘 要:**将行人检测算法和行人重识别算法相结合, 提出一种多目标跨摄像头跟踪算法, 该算法由行人检测、行人重识别和行人数据关联三大模块组成. 首先利用基于 YOLOv3 的行人检测改进算法检测视频中出现的行人, 并保存视频号、帧号和行人的全身位置信息; 其次, 利用基于生成对抗网络和重排序的行人重识别改进算法, 为已检测的行人图片赋予一个标签; 最后整合前两步得到的行人信息, 生成包含视频中所有行人信息的 JSON(JavaScript Object Notation)文件. 该算法可以快速、高效地完成多目标跨摄像头跟踪任务, 有一定实用价值, 在南京市举办的全球人工智能应用大赛中获得了单项奖.

**关键词:**多目标跨摄像头跟踪, 行人重识别, 行人检测, 重排序, 生成对抗网络

中图分类号: TP391.41

文献标识码: A

## Research and implementation of a multi-target cross camera tracking technology

Dai Chenchao, Wang Hongyuan\*, Cao Liang, Yin Yuchang, Zhang Ji

(School of Computer Science and Artificial Intelligence Aliyun School of Big Data,  
Changzhou University, Changzhou, 213164, China)

**Abstract:** Combining the person detection algorithm and the person re-identification algorithm, a multi-target cross camera tracking algorithm is proposed. The algorithm consists of three modules: person detection, person re-identification and person data association. Firstly, the improved person detection algorithm based on YOLOv3 is used to detect the person appearing in the video and to save the video number, frame number and person body position information. Secondly, the improved person re-identification algorithm based on generative adversarial network and reranking is used to assign a label to the detected person images. Finally, the person information obtained in the first two steps is integrated to generate a JavaScript Object Notation (JSON) file containing all information of the person in the video. The algorithm can complete the multi-target cross camera tracking task quickly and efficiently, and has certain practical value, which won a single award in the Global (Nanjing) Artificial Intelligence Application competition.

**Key words:** multi-target cross camera tracking, person re-identification, person detection, reranking, generative adversarial network

随着经济的发展和人们对安全需求的日益加强, 越来越多的监控摄像机被安装在医院、火车站、机场和学校等大型公共场所. 传统的监控摄

像机仅有录制视频、存储视频、播放视频的功能, 若要对异常状况进行实时监控和报警往往需要监控人员不间断地注视屏幕, 这极其耗时耗力. 最

基金项目: 国家自然科学基金(61976028, 61572085, 61806026, 61502058), 江苏省自然科学基金(BK20180956)

收稿日期: 2020-12-09

\* 通讯联系人, E-mail: hywang@cczu.edu.cn

重要的是,各地部署的摄像机每天会产生海量的监控视频,仅仅依靠人工难以处理.为提高监控效率、减轻人工劳动负荷,可以借助计算机视觉、深度学习和模式识别等相关技术<sup>[1-4]</sup>实现对视频中无关信息的自动过滤、特定目标的自动识别和异常状况的及时报警等功能,因此多目标跨摄像头跟踪技术应运而生并迅速成为一大研究热点,同时涌现出一批有价值的多目标跨摄像头跟踪算法.例如,Chen et al<sup>[5]</sup>提出一种基于图模型的联合优化方法.该方法使用全局图模型来匹配局部轨迹片段,在图模型中每个节点代表一个轨迹片段,每条边代表轨迹片段的连接,通过寻找从源节点到目标节点的最小代价流,得到每个跟踪对象的完整运动轨迹.Zhang et al<sup>[6]</sup>使用卷积神经网络为每个追踪对象提取鲁棒的深度特征,并引入一个特征重排序机制匹配各个追踪对象的轨迹片段.Lee et al<sup>[7]</sup>提出一种增量相机链接模型,利用追踪对象的线索信息来自适应匹配轨迹片段,并且在线更新追踪对象的线索信息,以消除视角变化的影响.Xu et al<sup>[8]</sup>使用贝叶斯公式在非重叠摄像机网络中进行路径重建,以便在不同摄像机视图之间创建轨迹片段,同时提出一种时空解析结构,即利用追踪目标的语义属性对匹配候选对象进行剪枝.本文在前人研究的基础上,提出一种新的多目标跨摄像头跟踪算法.

本文的研究对象为不同场景下不同摄像机录制的监控视频.给定 200 张待查找的人脸图像以及 20 段室内视频和 4 段室外视频.视频总计 30 多小时,收集自四个室内摄像机和四个室外摄像机,要求判断 200 张待查找的人脸图像是否出现在以上这些视频中,如果出现则需要找出这个人的全身位置信息、视频号、帧号以及对应的身份信息,并将找到的所有行人的信息合成一个 JSON 文件.问题的关键是跨摄像头跟踪,即根据目标在某段视频中的时刻和在该时刻的行人信息,关联该目标在其他视频中出现的行人信息.针对上述要求,本文提出的多目标跨摄像头跟踪算法旨在通过行人检测<sup>[9]</sup>和行人重识别技术<sup>[10-12]</sup>来关联不同场景下不同监控摄像头录制视频中出现的行人,进而实现多目标跨摄像头跟踪任务.

本文的主要贡献:

(1)针对比赛官方给定的任务,将行人检测算法和行人重识别算法相结合,提出一种多目标跨摄像头跟踪算法.

(2)提出一种基于生成对抗网络和重排序的行人重识别改进算法.首先,调整 ResNet50 模型结构以提取更具有判别力的特征;其次,利用生成对抗网络生成图片做数据增强,以防止模型过拟合;最后,将原始的欧式距离、杰卡德距离和拓展近邻距离相结合,提出一种新的重排序方法,以提高行人检索精度.

(3)构建一个新的行人重识别数据集(记为 MTCCT2018),对日常生活中收集的监控视频数据(多目标跨摄像头跟踪赛题公布的视频数据)进行处理,为视频中出现的行人赋予一个准确的标签.

## 1 算法整体框架

针对上述待解决的任务,本文构建了多目标跨摄像头跟踪技术的整体框架,该框架可以分为三个部分:行人检测、行人重识别和行人数据关联.主要思路如图 1 所示:首先利用行人检测算法检测给定视频中出现的行人,然后将已检测行人的图片与给定的人脸图库相关联构建行人图库,再利用行人重识别算法为已检测图片分配一个标签,最后整合行人检测和行人重识别算法得到的行人信息,生成出最终包含所有行人信息的 JSON 文件.其中行人数据关联部分的主要作用是对不同摄像头中出现的同一目标进行关联.

## 2 算法实现步骤

**2.1 行人检测** 行人检测旨在判断给定的图片或视频帧中的行人位置,并用矩形框标出.视频中行人的检测不仅需要保证准确性,还需要保证其实时性.针对当前行人检测方法精度和实时性不能同时兼顾的问题,本文采用以 YOLOv3 模型为基础的改进算法<sup>[13]</sup>,结合行人检测任务的特点进行优化与改进,准确性与实时性都比较好.

由于视频中的行人距离摄像机远近不同,导致视频帧中的行人尺度不同,小目标的行人检测精度较差.针对这个问题,采用 YOLOv3 改进算法运用多尺度检测,增加两个预测层,利用更多的特征图提高小目标行人检测精度.

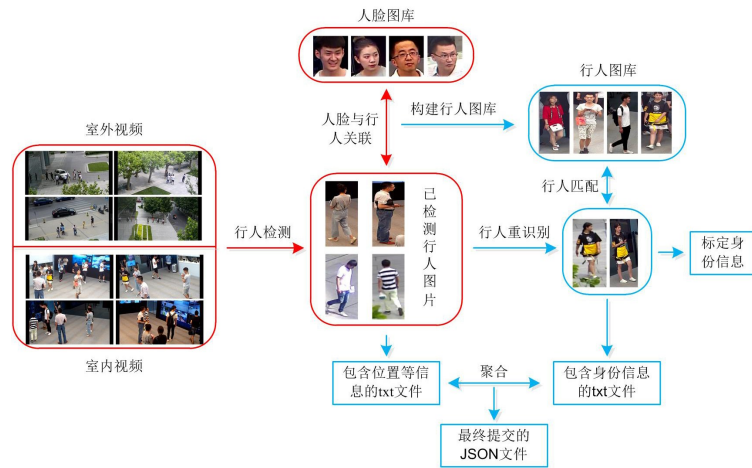


图1 本文算法的整体框架

Fig. 1 The overall framework of our algorithm

针对YOLOv3模型在COCO数据集上训练得到的anchor值不适用于行人检测任务的问题,本文采用的YOLOv3改进算法根据行人检测数据集Caltech的特点以及五个尺度检测层的网络结构,通过 $k$ -means聚类得到anchor的个数与值,给五个尺度预测层分配更加准确的anchor值,进一步提高行人检测的精度。

**2.2 人脸与行人关联** 根据给定的200张已标注的人脸图像,与行人检测部分检测出来的完整行人图像一一对应,构建一个用于行人重识别任务的行人图库。再将检测的类别设置为人脸和行人,根据检测到的人脸和行人的位置关系来自动匹配人脸和行人图像<sup>[14]</sup>,进而得到行人图库。

**2.3 行人重识别** 行人重识别指匹配同一个行人在不相交摄像机视图下的行人图片<sup>[15-18]</sup>(如图2所示)。将视频中已经检测的行人图片作为查询图片,与之前和人脸库关联好的行人图库进行匹

配,进而为已检测行人图片分配一个标签。

针对传统非深度的行人重识别方法检索精度较低的问题,采用以ResNet50为基础的改进算法,并选取当前行人重识别领域的公开数据集为训练集。由于现有的公开数据集中行人图片数量是有限的,采用生成对抗网络生成行人图片,和原始训练集一起构成增强的数据集,并采用标签平滑正则化训练所有的行人图片。此外,为了缓解行人匹配过程中图像匹配错误的困扰,提出一种新的重排序方法对原始排序列表进行排序,提高行人检索精度。

**2.3.1 基于生成对抗网络的数据增强** 近年来,生成对抗网络在图像生成、图像编辑和特征学习等方面取得了令人瞩目的成果,应用于图像修复、文字转图片以及3D数据生成等领域的最新方法均利用了生成对抗网络技术。生成对抗网络成功的关键在于对抗损失的思想,使生成网络生成的



图2 行人重识别任务

Fig. 2 Person re-identification task



数据在理论上和真实数据难以区分. 这种损失对于图像生成任务尤其重要, 这也是许多计算机图形学任务优化的目标.

图 3 展示了深度卷积生成对抗网络<sup>[19]</sup>(Deep Convolutional Generative Adversarial Network, DCGAN)的网络结构示意图. 和普通的生成对抗网络相比, DCGAN 在收敛速度、训练稳定程度以及生成图片质量等方面有显著优势. 采用 DCGAN 生成行人图片, 扩充数据集. 图 4 展示了 DCGAN 在目前行人重识别领域公开数据集上生成的行人图片.

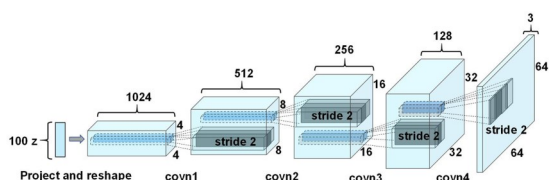


图 3 深度卷积生成对抗网络

Fig. 3 Deep convolutional generative adversarial network



图 4 DCGAN 生成的行人图片

Fig. 4 Person images generated by DCGAN

**2.3.2 改进的网络结构** 选取 ResNet50 为基础网络结构并进行调整. 给定一个基于 ResNet50 训练好用于分类任务的网络模型, 最后一层全连接层中的权重向量往往有较高的相关性. 据文献检索可知<sup>[20]</sup>, 在行人重识别任务中欧式度量通常被用来计算特征空间中不同行人之间的距离, 然而检索分类任务中的欧式度量要发挥最理想的作用, 通常要求参与计算的特征向量中的各项都相互独立. 此时全连接层中这些高相关性的权重向量会明显地损害参与计算的行人特征.

为了缓解全连接层相关性较高的问题, 在原

有的网络框架基础上进行修改, 去掉原来网络中最后的 1000 维分类层, 新增一个 1024 维的全连接层(记为 FC-1024), 并借鉴 Sun et al<sup>[20]</sup> 的训练思想, 即在网络训练过程中始终保持该全连接层中权重向量的正交性, 通过这种方式来降低全连接层中权重向量的相关性.

在此基础上, 在上述全连接层后面再加上一层全连接层, 该层的神经元个数根据参与训练行人重识别数据集中的类别数来定(称为 FC-T). 简单来说, 最后一层全连接层中的神经元实际上就是将前面卷积层提取的特征整合到一起输出为一个值. 在行人重识别任务中, 这个值可以理解置信度, 即网络模型默认输入图片属于某一类别的概率值. 将神经元的个数调整为行人类别数可以有效地减少网络的误判, 例如最后全连接层的神经元个数为 1000, 行人类别有 100 类, 那么在这种情况下就可能出现大量不必要的置信度分配. 上述修改方式可以有效地避免该问题.

此外, 在最后一层全连接层之前加入批归一化层(记为 BN)、ReLU 激活函数和 Dropout 层. 批归一化层对送往激活函数的输入作归一化处理, 缓解输入数据发生偏移和增大的影响, 还能加快网络的训练速度. ReLU 激活函数可以增加网络的非线性, 进而提高网络的判别能力. Dropout 层使每一个神经元的激活值在每一个批次的训练过程中以一定的概率停止更新, 防止陷入局部最优解, 提高模型的泛化性. 图 5 展示了结合生成图片的网络训练框架.

**2.3.3 标签平滑正则化** 标签平滑正则化<sup>[21]</sup>旨在减少真实类别的置信度, 给非真实类别分配较少的置信度, 防止模型过拟合, 一般和交叉熵损失配合使用. 在传统的分类任务中, 通常采用 one-shot 编码设置输入图片的标签分布. 因此, 输入图片的标签分布  $q(k)$  可以定义为:

$$q(k) = \begin{cases} 1 & k=y \\ 0 & k \neq y \end{cases} \quad (1)$$

其中,  $y$  指输入图片对应的真实标签. 网络模型预测输入图片属于每一个标签  $k \in \{1, \dots, K\}$  的概率, 并对预测概率值  $p(k)$  做归一化处理, 此时交叉熵损失可以写为:

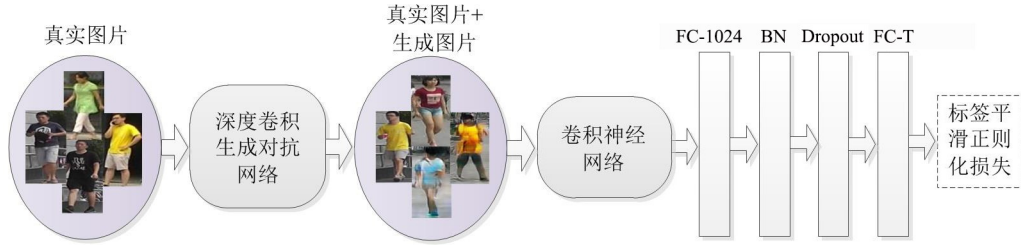


图5 结合生成图片的网络训练框架

Fig. 5 Network training framework combined with generated images

$$L_{\text{cross}} = - \sum_{k=1}^K \log p(k) q(k) = - \log p(y) \quad (2)$$

原始训练集中通常会出现标注错误和不同类别之间样本数量差异过大的问题,式(2)很容易受这些问题的影响. 标签平滑正则化就是针对上述问题设计的一种训练策略,输入图片的标签分布可以重写为:

$$q_{\text{LSR}}(k) = \begin{cases} 1 - \epsilon + \frac{\epsilon}{K} & k = y \\ \frac{\epsilon}{K} & k \neq y \end{cases} \quad (3)$$

其中  $\epsilon \in [0, 1]$ . 此时真实图片交叉熵损失可写为:

$$L_R = -(1 - \epsilon) \log p(y) - \frac{\epsilon}{K} \sum_{k=1}^K \log p(k) \quad (4)$$

虽然 DCGAN 生成的行人图片与原始训练集中的真实图片有某种共性,但无法判断生成图片到底属于哪个类别,无法直接用于全监督的行人重识别训练. 为了将生成图片加入训练过程,引用了 Zheng et al<sup>[3]</sup> 的训练策略,为生成图片分配一个统一的标签. 生成图片的标签分布可定义为:

$$q_{\text{LSR}}(k) = \frac{1}{K} \quad (5)$$

此时,生成图片的交叉熵损失为:

$$L_G = - \frac{1}{K} \sum_{k=1}^K \log p(k) \quad (6)$$

结合式(4)与式(6),本文的损失函数为:

$$L_{\text{Loss}} = (1 - \beta) \left[ -(1 - \epsilon) \log p(y) - \frac{\epsilon}{K} \sum_{k=1}^K \log p(k) \right] - \beta \left[ \frac{1}{K} \sum_{k=1}^K \log p(k) \right] \quad (7)$$

其中,  $\beta$  的取值为 0 或者 1.

**2.3.4 重排序** 给定一张查询图片  $p$  和一个包含  $N$  张行人图片的行人图库  $G = \{g_i | i = 1, 2, \dots, N\}$ .

通过计算查询图片和行人图库之间的欧式距离,获得一个初始排序列表. 重排序<sup>[22-23]</sup>是指根据初始排序列表的排序结果,寻找一种对行人外貌差异鲁棒的距离度量. 查询图片  $p$  的  $k$  近邻集合  $N(p, k)$  可以定义为:

$$N(p, k) = \{g_1, g_2, \dots, g_k\} \quad (8)$$

$N(p, k)$  中的元素是指初始排序列表中查询图片  $p$  最相似的  $k$  张图库图片. 在此基础上,查询图片的拓展近邻集合  $R(p, k)$  可以定义为:

$$\{R(p, M), R(M, N)\} \rightarrow R(p, k) \quad (9)$$

其中,  $R(p, M)$  是与查询图片  $p$  最相似的  $M$  张近邻图片,称为直接近邻;  $R(M, N)$  指与  $R(p, M)$  中每个元素最相似的  $N$  张近邻图片,称为间接近邻. 图6展示了拓展近邻集合计算的过程.

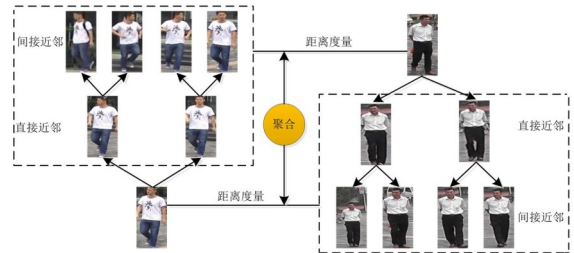


图6 拓展近邻计算

Fig. 6 Extended nearest neighbor computing

拓展近邻距离可以定义为:

$$d_E(p, g_i) = \frac{1}{2K} \sum_{j=1}^K (d(p_j, g_i) + d(g_{ij}, p)) \quad (10)$$

其中,  $p_j$  指的是查询图片的拓展近邻集合  $R(p, K)$  中的第  $j$  个近邻,  $g_{ij}$  指的是图库图片  $g_i$  的拓展近邻集合  $R(g_i, K)$  中的第  $j$  个近邻,  $d(\cdot)$  代表图片对之间的距离.

查询图片  $p$  的  $k$  互邻集合可以定义为:

$$R(p, k) = \{(g_i \in N(p, k)) \cap (p \in N(g_i, k))\} \quad (11)$$

和  $k$  近邻集合相比,  $k$  互邻集合与查询图片  $p$  的相关性更高. 由于行人重识别任务的挑战性, 会导致与查询图片  $p$  身份一致的图库图片可能既不在  $k$  近邻集合中也不在  $k$  互邻集合中. 为了缓解这个问题, 在原有  $k$  互邻集合的基础上再加上  $k$  互邻集合中每个元素的  $k/2$  互近邻, 构成增强型  $k$  互近邻集合:

$$R^*(p, k) \leftarrow R(p, k) \cup R\left(q, \frac{k}{2}\right), \forall q \in R(p, k) \quad (12)$$

杰卡德距离通过计算两个集合的相似性来确定查询图片  $p$  和图库图片  $g_i$  之间的距离. 因此, 根据这两张图片各自的  $k$  互近邻集合, 查询图片与图库图片之间的杰卡德距离可以定义为:

$$d_J(p, g_i) = 1 - \frac{|R^*(p, k) \cap R^*(g_i, k)|}{|R^*(p, k) \cup R^*(g_i, k)|} \quad (13)$$

其中,  $|\cdot|$  表示集合中元素的数量.

将初始的欧式距离、基于  $k$  互邻的杰卡德距离和拓展近邻距离融合在一起修正初始的排序列表, 最终的距离的计算式如下:

$$d^*(p, g_i) = \alpha[(1 - \mu)d_J(p, g_i) + \mu d(p, g_i)] + (1 - \alpha)d_E(p, g_i) \quad (14)$$

其中,  $\alpha$  与  $\mu$  均为权衡因子, 取值范围为 0 到 1.

**2.4 行人的数据关联** 行人检测部分利用改进的 YOLOv3 算法检测给定的 24 段视频中所有的行人图片, 将检测到的行人图片分别存储在 24 个以视频名命名的文件夹中, 同时生成与 24 个文件夹对应的 txt 文件来保存所检测行人图片所在的视频编号、所在的帧号、目标归一化后的全身位置信息和图片编号.

行人重识别部分利用改进的行人重识别算法给已检测出来的行人图片赋予一个身份标签, 同时生成 24 个与行人检测部分行人图片文件夹对应的 txt 文件来保存已标定的行人图片编号和已标定的行人身份标签.

将行人检测部分和行人重识别部分生成的 txt 文件进行组合, 生成最终需要的 JSON 文件 (如图 7 所示). 在 JSON 文件中行人标签为 key 值, 该行人出现的所有视频号、所在帧号和归一化后的位置信息为 value 值. 即, 待查找的 200 个行

人在所有视频中出现的信息在 JSON 文件中以键值对的形式关联起来. 若要查看某个行人的运动轨迹, 只需在 JSON 文件中找出该行人标签对应的 value 值即可.



图 7 行人数据关联

Fig. 7 Person data association

### 3 实验

实验在 2018 全球 (南京) 人工智能应用大赛多目标跨摄像头跟踪赛题公布的测试视频上进行, 测评指标为  $F1$ -score. 该算法的优劣性很大程度上依赖行人检测和行人重识别算法的性能, 因此在目前公开的数据集和自制的 MTCCT2018 数据集上测试本文采用的改进后的行人检测和行人重识别算法, 并用累积匹配特征曲线以及平均查准率 (mean Average Precision,  $mAP$ ) 作为测评指标. 所有实验均在 Ubuntu16.04 环境下进行, 详细配置: 显卡 NVIDIA TITAN XP, Cuda8.0, Cudnn5.1, Caffe, Opencv3.2.0, Python2.7.12, Nvidia 驱动 384 和 MatlabR2015b.

**3.1 自制行人重识别数据集** 基于深度学习算法的行人重识别方法往往需要大量的行人图片参与训练, 然而目前行人重识别领域公开数据集中的行人图片数量较少. 为了获得更多的行人图片参与训练, 收集现实生活中的监控视频 (多目标跨摄像头跟踪赛题公布的视频数据), 并对收集的視頻数据进行处理, 构建一个新的行人重识别数据集, 即 MTCCT2018 数据集, 总计 19757 张行人图片. 图 8 展示了数据集的部分行人图片. 原始



的视频数据收集自四个室内摄像头和四个室外摄像头,总计24段视频,有相对复杂的背景与场景转换.另外,这些视频数据收集自一天的多个时段,能够很好地模拟真实场景,还有较为明显的光照变化.首先利用改进的YOLOv3算法检测视频帧中的行人图片,然后利用行人重识别技术为已检测行人图片赋予一个标签.这期间人为去除那些检测效果不好的行人图片,同时人为检查已检测行人图片标签的正确性.



图8 自制数据集

Fig. 8 Self-made dataset

**3.2 行人检测实验** 本文在Caltech数据集上训练改进后YOLOv3算法(记为YOLOv3\_improve),同时为了测试该算法的泛化性能,在Daimler和INRIA数据集上测试在Caltech数据集上训练好的模型,实验结果如表1所示.可以看出,在Caltech, Daimler和INRIA数据集上,改进的YOLOv3算法的 $mAP$ 明显超过原始的YOLOv3算法,虽然每一帧的检测时间略长于原始的YOLOv3算法,但仍可满足实时检测的要求.和现有的检测算法Faster R-CNN<sup>[24]</sup>相比,改进后的YOLOv3算法在精度上不太弱于Faster R-CNN,但检测速度远胜Faster R-CNN.把在Caltech数据集上训练好的模型直接用于Daimler和INRIA数据集上,改进的YOLOv3算法仍然可以获得较高的 $mAP$ ,证明该算法可以适应不同的场景,有较好的泛化能力.

**3.3 行人重识别实验** 在Market-1501, DukeMTMC-reID和CUHK03数据集<sup>[25-27]</sup>上测试本文提出的基于生成对抗网络和重排序的行人重识别改进算法(记为Ours)的性能,实验结果如表2所

表1 行人检测算法的实验结果对比

Table 1 Experimental results of person detection algorithms

数据集	方法	$mAP$	检测速度(s)
Caltech	Faster R-CNN	0.9271	0.039
	YOLOv3	0.8832	0.015
	YOLOv3_improve	0.9168	0.021
INRIA	Faster R-CNN	0.8798	0.039
	YOLOv3	0.8245	0.014
	YOLOv3_improve	0.8367	0.019
Daimler	Faster R-CNN	0.8489	0.035
	YOLOv3	0.8512	0.015
	YOLOv3_improve	0.8738	0.021

示.可以看出,和基准方法ResNet50相比,改进算法有显著的性能提升:在Market-1501数据集上Rank1提升5.99%, $mAP$ 提升14.84%;在DukeMTMC-reID数据集上Rank1提升6.91%, $mAP$ 提升15.96%;在CUHK03数据集上Rank1提升9.77%, $mAP$ 提升7.04%.和本文的改进算法相比较的还有SOMAnet<sup>[28]</sup>, Verif.-Identif<sup>[29]</sup>, Pose Transfer<sup>[30]</sup>, TriNet<sup>[31]</sup>, VGG16<sup>[32]</sup>和DSA-reID<sup>[33]</sup>等先进算法,从表2可以看出,本文提出的行人重识别改进算法无论是Rank1指标还是 $mAP$ 指标均优于上述对比算法效果.

通常情况下,在源场景下表现很好的行人重

表2 行人重识别算法的实验结果对比

Table 2 Experimental results of person re-identification algorithms

数据集	方法	Rank1	$mAP$
Market-1501	SOMAnet	0.7387	0.4789
	Verif.-Identif	0.7951	0.5987
	ResNet50	0.7887	0.5693
	Ours	0.8486	0.7177
DukeMTMC-reID	Pose Transfer	0.6861	0.4812
	TriNet	0.7244	0.5351
	ResNet50	0.6876	0.4772
	Ours	0.7567	0.6368
CUHK03	VGG16	0.4912	0.5571
	DSA-reID	0.7821	0.7312
	ResNet50	0.6873	0.7458
	Ours	0.7850	0.8162

识别模型在新场景下不一定有好的表现. 为了验证本文提出的行人重识别改进算法的泛化能力, 在自己制作的数据集 MTCCT2018 上分别测试在 Market-1501, DukeMTMC-reID 和 CUHK03 数据集上训练好的模型的性能, 实验结果如表 3 所示. 可以看出, 虽然将改进算法在其他数据集上训练好的模型直接用到 MTCCT2018 数据集上会导致性能下降, 但和现有其他算法相比, 改进算法的识别能力还是较强, 证明改进算法不仅可以有效改善行人重识别性能, 还具有较好的泛化能力.

**3.4 多目标跨摄像头跟踪实验** 经实验测试, 最终完成的多目标跨摄像头跟踪算法基本上可以准确地检测视频帧中出现的行人, 并给已检测行人赋予一个较为准确的标签 (如图 9 所示); 能够根

表 3 行人重识别算法在跨数据集上的实验结果对比

Table 3 Experimental results of person re-identification algorithms across datasets

测试数据集 方法(训练数据集)	MTCCT2018	
	Rank1	mAP
ResNet50 (Market-1501)	0.5147	0.4029
Ours (Market-1501)	0.5578	0.4866
ResNet50 (DukeMTMC-reID)	0.5468	0.4021
Ours (DukeMTMC-reID)	0.6197	0.4894
ResNet50 (CUHK03)	0.3126	0.1885
Ours (CUHK03)	0.3358	0.2334

据室内视频中出现的行人信息关联该目标在其他视频中出现的行人信息, 最终生成较为完整的 JSON 文件. 经比赛官方的测评, 该算法最终取得初赛第一、复赛第二的成绩.

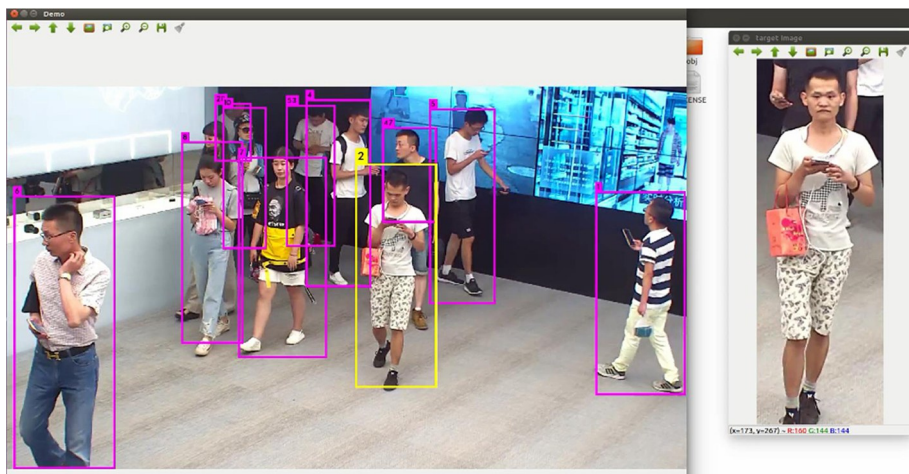


图 9 实验结果示意图

Fig. 9 Schematic diagram of experimental result

## 4 结 论

多目标跨摄像头跟踪技术在智能城市、智能安防和刑事侦查等领域有广泛的应用前景, 可以补充人脸识别的局限, 更准确地确认行人身份信息. 该技术依赖多种信息和技术, 如行人检测和行人重识别, 可以对不同摄像头中出现的同一目标进行关联. 本文提出一种多目标跨摄像头跟踪算法, 将行人检测算法和本文提出的行人重识别改进算法相结合, 通过 JSON 文件来关联不同场景下不同监控摄像头录制视频中出现的行人, 达到多目标跨摄像头跟踪的效果. 根据比赛官方的

测评, 该算法可以快速高效地检索和识别视频中的行人, 并能生成包含视频中所有行人信息的 JSON 文件, 在智能生活领域有巨大的应用价值.

### 参考文献

- [1] 周志华. 机器学习. 北京: 清华大学出版社, 2016, 425. (Zhou Z H. Machine learning. Beijing: Tsinghua University Press, 2016, 425.)
- [2] Zheng L, Yang Y, Hauptmann A G. Person re-identification: past, present and future. 2016, arXiv: 1610.02984.
- [3] Zheng Z D, Zheng L, Yang Y. Unlabeled samples



- generated by gan improve the person re-identification baseline in vitro//Proceedings of 2017 IEEE International Conference on Computer Vision. Venice:Italy:IEEE,2017:3754—3762.
- [4] Zhong Z, Zheng L, Zheng Z D, et al. Camera style adaptation for person re-identification//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA: IEEE,2018:5157—5166.
- [5] Chen W H, Cao L J, Chen X T, et al. An equalized global graph model - based approach for multicamera object tracking. IEEE Transactions on Circuits and Systems for Video Technology,2017,27(11):2367—2381.
- [6] Zhang Z M, Wu J N, Zhang X, et al. Multi-target, multi - camera tracking by hierarchical clustering: recent progress on DukeMTMC Project. 2017, arXiv:1712.09531.
- [7] Lee Y G, Tang Z, Hwang J N. Online-learning-based human tracking across non - overlapping cameras. IEEE Transactions on Circuits and Systems for Video Technology,2018,28(10):2870—2883.
- [8] Xu Y L, Liu X B, Qin L, et al. Cross-view people tracking by scene-centered spatio-temporal parsing// Proceedings of the 31<sup>st</sup> AAAI Conference on Artificial Intelligence. San Francisco, CA, USA: AAAI,2017:4299—4305.
- [9] Redmon J, Farhadi A. Yolov3: An incremental improvement. 2018, arXiv:1804.02767.
- [10] Zheng Z D, Yang Y. Parameter-efficient person Re-identification in the 3D Space. 2020, arXiv: 2006.04569.
- [11] 丁宗元,王洪元,陈付华等. 基于距离中心化与投影向量学习的行人重识别. 计算机研究与发展,2017, 54(8): 1785—1794. (Ding Z Y, Wang H Y, Chen F H, et al. Person re-identification based on distance centralization and projection vectors learning. Journal of Computer Research and Development, 2017, 54 (8):1785—1794.)
- [12] Ni T G, Gu X Y, Wang H B, et al. Discriminative deep transfer metric learning for cross - scenario person re - identification. Journal of Electronic Imaging,2018,27(4):043026.
- [13] 范丽,苏兵,王洪元. 基于YOLOv3模型的实时行人检测改进算法. 山西大学学报(自然科学版),2019, 42(4): 709—717. (Fan L, Su B, Wang H Y. Improved real-time pedestrian detection algorithm based on YOLOv3. Journal of Shanxi University (Natural Science Edition),2019,42(4):709—717.)
- [14] 焦珊珊,李云波,陈佳林等. 多目标跨摄像头跟踪技术. 国防科技,2019,40(6):33—41. (Jiao S S, Li Y B, Chen J L, et al. Research on multi-target multi-camera tracking technology. National Defense Science & Technology,2019,40(6):33—41.)
- [15] Zheng Z D, Yang X D, Yu Z D, et al. Joint discriminative and generative learning for person re-identification//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, CA, USA: IEEE, 2019: 2138—2147.
- [16] Wang H Y, Ding Z Y, Zhang J, et al. Person reidentification by semisupervised dictionary rectification learning with retraining module. Journal of Electronic Imaging,2018,27(4):043043.
- [17] 戴臣超,王洪元,倪彤光等. 基于深度卷积生成对抗网络和拓展近邻重排序的行人重识别. 计算机研究与发展,2019,56(8):1632—1641. (Dai C C, Wang H Y, Ni T G, et al. Person re-identification based on deep convolutional generative adversarial network and expanded neighbor reranking. Journal of Computer Research and Development, 2019, 56(8): 1632—1641.)
- [18] Ni T G, Ding Z Y, Chen F H, et al. Relative distance metric leaning based on clustering centralization and projection vectors learning for person re - identification. IEEE Access,2018,6:11405—11411.
- [19] Radford A, Metz L, Chintala S. Unsupervised representation learning with deep convolutional generative adversarial networks. 2016, arXiv: 1511.06434.
- [20] Sun Y F, Zheng L, Deng W J, et al. Svdnet for pedestrian retrieval//Proceedings of 2017 IEEE Conference on Computer Vision. Venice, Italy: IEEE,2017:3800—3808.
- [21] Szegedy C, Vanhoucke V, Ioffe S, et al. Rethinking the inception architecture for computer vision// Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA:IEEE,2016:2818—2826.
- [22] Zhong Z, Zheng L, Cao D L, et al. Re-ranking person re - identification with k - reciprocal encoding//

- Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA: IEEE, 2017: 1318–1327.
- [23] Sarfraz M S, Schumann A, Eberle A, et al. A pose-sensitive embedding for person re-identification with expanded cross neighborhood re-ranking//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA: IEEE, 2018: 420–429.
- [24] Girshick R. Fast R-CNN//Proceedings of 2015 IEEE International Conference on Computer Vision. Santiago, Chile: IEEE, 2015: 1440–1448.
- [25] Zheng L, Shen L Y, Tian L, et al. Scalable person re-identification: a benchmark//Proceedings of 2015 IEEE International Conference on Computer Vision. Santiago, Chile: IEEE, 2015: 1116–1124.
- [26] Ristani E, Solera F, Zou R, et al. Performance measures and a data set for multi-target, multi-camera tracking//Proceedings of 2016 European Conference on Computer Vision. Springer Berlin Heidelberg, 2016: 17–35.
- [27] Li W, Zhao R, Xiao T, et al. DeepReID: deep filter pairing neural network for person re-identification//Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus, OH, USA: IEEE, 2014: 152–159.
- [28] Barbosa I B, Cristani M, Caputo B, et al. Looking beyond appearances: synthetic training data for deep CNNs in re-identification. Computer Vision and Image Understanding, 2018, 167: 50–62.
- [29] Zheng Z D, Zheng L, Yang Y. A discriminatively learned CNN embedding for person reidentification. ACM Transactions on Multimedia Computing, Communications, and Applications, 2017, 14(1): 1–20.
- [30] Liu J X, Ni B B, Yan Y C, et al. Pose transferrable person re-identification//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA: IEEE, 2018: 4099–4108.
- [31] Hermans A, Beyer L, Leibe B. In defense of the triplet loss for person re-identification. 2017, arXiv: 1703.07737.
- [32] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. 2014, arXiv: 1409.1556.
- [33] Zhang Z Z, Lan C L, Zeng W J, et al. Densely semantically aligned person re-identification//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, CA, USA: IEEE, 2019: 667–676.

(责任编辑 杨可盛)