

文章编号:1007-2780(2022)07-0880-11

基于 YOLOv5 和重识别的行人 多目标跟踪方法

贺愉婷^{1,2}, 车进^{1,2*}, 吴金蔓^{1,2}

(1. 宁夏大学 物理与电子电气工程学院, 宁夏 银川 750021;

2. 宁夏沙漠信息智能感知重点实验室, 宁夏 银川 750021)

摘要:针对目前遵循基于检测的多目标跟踪范式存在的不足,本文以 DeepSort 为基础算法展开研究,以解决跟踪过程中因遮挡导致的目标 ID 频繁切换的问题。首先改进外观模型,将原始的宽残差网络更换为 ResNeXt 网络,在主干网络上引入卷积注意力机制,构造新的行人重识别网络,使模型更关注目标关键信息,提取更有效的特征;然后采用 YOLOv5 作为检测算法,加入检测层使得模型适应不同尺寸的目标,并在主干网络加入坐标注意力机制,进一步提升检测模型精度。在 MOT16 数据集上进行多目标跟踪实验,多目标跟踪准确率达到 66.2%,多目标跟踪精确率达到 80.8%,并满足实时跟踪的要求。

关键词:多目标跟踪;行人重识别;YOLOv5;注意力机制;深度学习

中图分类号:TP391 **文献标识码:**A **doi:**10.37188/CJLCD.2022-0025

Pedestrian multi-target tracking method based on YOLOv5 and person re-identification

HE Yu-ting^{1,2}, CHE Jin^{1,2*}, WU Jin-man^{1,2}

(1. School of Physics and Electronic-Electrical Engineering, Ningxia University, Yinchuan 750021, China;

2. Ningxia Key Laboratory of Intelligent Sensing for Desert Information, Yinchuan 750021, China)

Abstract: Aiming at the shortcomings of current detection-based multi-target tracking paradigm, a research is conducted based on the algorithm of DeepSort to address the issue of frequent switching of targeted ID resulting from occlusion in tracking process. Firstly, focus should be placed on improving appearance model. Efforts should be made in replacing broadband and residual networks with ResNeXt networks, which introduces the mechanism for convolution attention into the backbone network and establish a new person re-identification network. In doing so, the model can pay more attention to critical information of targets and obtain effective features. Then, YOLOv5 serves as a detection algorithm. Adding detection layer enables the model to respond to targets of different sizes. Moreover, the mechanism for coordinate attention is introduced into the backbone networks. These efforts can further

收稿日期:2022-01-24;修订日期:2022-02-11.

基金项目:国家自然科学基金(No. 61861037)

Supported by National Natural Science Foundation of China(No. 61861037)

*通信联系人, E-mail:koalache@126.com

improve the accuracy of detection model. The multi-target tracking experiment is carried out on data sets of MOT16, the multi-target tracking accuracy rate is up to 66.2%, and the multi-target tracking precision ratio is up to 80.8%. All these can meet the needs of real-time tracking.

Key words: multi-target tracking; person re-identification; YOLOv5 network; attention mechanism; deep learning

1 引言

多目标跟踪(Multiple Target Tracking, MTT)主要任务是在给定视频中同时对多个特定目标进行定位,同时保持目标的ID稳定,最后跟踪记录他们的轨迹^[1]。本文主要关注对多行人跟踪的研究。目前主流的行人跟踪算法大多是基于检测的跟踪范式(Tracking-by-Detection, TBD),即在检测结果的基础上进行目标跟踪。它先通过目标检测算法检测出视频帧中目标对象可能出现的区域,然后通过关联模型将属于同一运动目标的检测框关联到一起,得到目标的关联轨迹,完成目标对象的跟踪。这就导致了基于检测的跟踪范式的跟踪效果很大程度上取决于行人特征的质量,因此,如何获取有效的行人特征是本文研究的重点。

行人重识别(Person re-identification, ReID)被认为是图像检索的子问题^[2],可以依据行人的外观特征实现跨摄像头无重叠视域下的目标行人的检索。行人重识别中的特征提取和度量学习可以为目标跟踪提供强有力的支撑。现在许多研究将行人重识别技术与检测/跟踪技术相结合,并广泛应用于智能安防系统。

基于TBD范式,Bewley等人^[3]提出SORT算法,重点关注帧间的预测和关联,结合卡尔曼滤波(Kalman Filter)和匈牙利算法(Hungarian Algorithm),提出一种简单有效的在线跟踪框架。Wojke等人^[4]提出DeepSORT算法。考虑到SORT算法并未过多关注长时间跟踪过程中由于遮挡导致的目标ID频繁切换问题,于是在SORT算法的基础上引入行人重识别技术作为外观模型。通过在重识别模型的学习增强网络对不同目标对象的鉴别能力。同时提出级联匹配策略提高目标匹配准确度。Chen等人^[5]提出MOTDT算法。针对不可靠检测结果对跟踪造成的误导,考虑检测任务和跟踪任务的互补

性,设计了一种新的多目标跟踪框架,并提出一种分层数据关联策略,充分利用重识别特征和空间信息提升跟踪性能。Wang等人^[6]提出JDE算法。从实时性方面考虑,融合一阶段检测与行人重识别,同时输出检测和ReID信息,加快推理速度。Zhang等人^[7]提出FairMOT算法。探究了目标检测和行人重识别任务的集成问题,采用Anchor-free范式的目标检测算法CenterNet^[8]作为检测分支,在此基础上增加一个平行分支输出ReID特征区分不同目标,将目标检测和重识别很好地统一起来。

综上所述,行人重识别技术提取出的有效特征为目标跟踪任务提供了强有力的支撑,且与目标检测算法的有效结合使目标跟踪的速度也有了很大提升,满足实时性的要求。因此本文提出一种基于YOLOv5与重识别的行人多目标跟踪方法,同时对检测算法与特征提取网络部分进行改进。

2 行人重识别特征提取网络

2.1 深度特征提取网络

为了使多目标跟踪算法具有更有鉴别力的目标特征,本文采用ResNeXt50网络作为主干网络。ResNeXt网络^[9]的提出是由于传统的提高模型准确率方法都是选择加深网络,这样会导致模型越来越复杂,超参数的数量随之增加,加大计算成本。因此Xie等人^[10]设计用一种平行堆叠相同拓扑结构的blocks替换了ResNet^[10]中的block,并且对ResNet进行了基数扩充。实验表明:在相同的模型大小和计算复杂度的条件下ResNeXt网络相比较原残差网络有更高的精度。因此为了获得更优的外观模型,本文采用ResNeXt50作为新的特征提取网络。

2.2 GAN网络进行数据增强

本文在行人再识别算法的公开数据集 Mar-

ket1501上进行实验。为了防止模型过拟合,基于Pytorch深度框架搭建了生成式对抗网络^[11](Generative adversarial network, GAN)模型,参见文献[12]中的相同的参数训练网络。使用GAN网络对Market1501数据集的训练集进行了扩充,由原来的12 936张图片扩充至77 616张图片,使用新的训练集对网络进行训练。如图1所示,原始图像经过GAN网络之后生成了其他5个相机风格的伪图像。

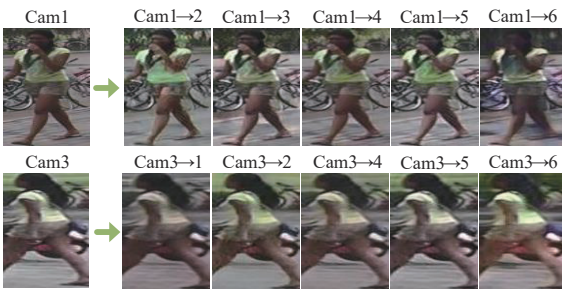


图1 生成图示例

Fig. 1 Generate image examples

将GAN网络扩增后的Market1501数据集送入ResNeXt50中进行特征提取,实验结果如表1所示。

表1 特征提取实验

Tab. 1 Feature extraction network experiments (Top-1)		
Model	Market1501	GAN(Market1501)
ResNeXt50	98.2%	98.8%

2.3 CBAM注意力机制

在多目标跟踪过程中,往往会面临遮挡问题,这种遮挡可能来自不同行人之间的遮挡,也可能来自固定物体的遮挡。遮挡极大可能导致目标在跟踪过程中前一帧的目标ID与后一帧的目标ID发生切换,这样网络就会认为这时出现了新的目标,导致跟踪中断,影响跟踪精度。因此考虑采用添加卷积注意力机制^[13](Convolutional Block Attention Module, CBAM)的方法增加目标对象的表现力,关注图中重要特征并抑制无用特征。该注意力机制是一个轻量级的通用模块,它将注意力过程分为两个独立的部分:通道注意力模块和空间注意力模块(图2(a),(b))。通道注意力关注什么样的特征是有意义的,采用了全局平均池

化和最大池化两种方式来分别利用不同的信息。之后再引入空间注意力模块来关注哪里特征是有效的,使得到的特征图更显著。

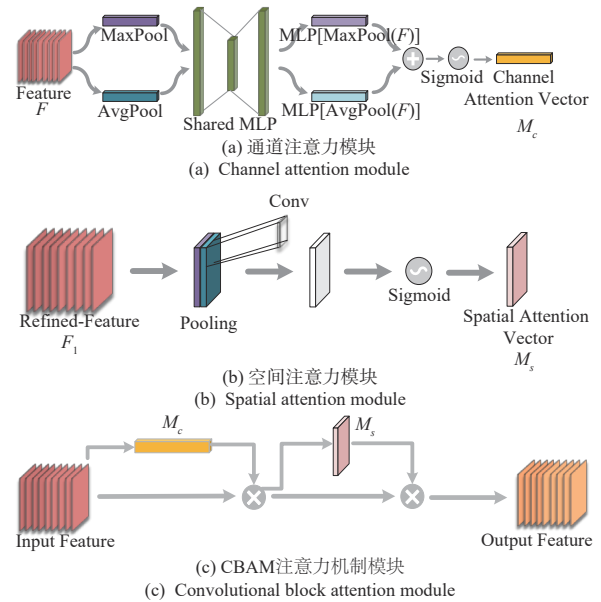


图2 CBAM注意力机制

Fig. 2 Convolutional block attention

2.3.1 通道注意力机制

通道注意力模块对每个输入通道的权重进行重新标定,使得包含目标对象的关键区域特征通道对最终卷积特征有更大的贡献^[14]。核心思想就是增大有效通道权重,减少无效通道权重。具体实现如式(1)所示:

$$M_c(F) = \sigma \left(\text{MLP} \left(\text{AvgPool}(F) \right) + \text{MLP} \left(\text{MaxPool}(F) \right) \right), \quad (1)$$

式中的输入特征 F 在空间维度上经过平均池化以及最大池化,再分别经过共享网络主元素进行求和来合并输出特征向量。

2.3.2 空间注意力机制

通道注意力机制告诉网络需要注意的部分,空间注意力机制给出关键特征的位置。具体实现如公式(2)所示:

$$M_s(F) = \sigma \left(f^{7 \times 7} \left(\left[\text{AvgPool}(F); \text{MaxPool}(F) \right] \right) \right), \quad (2)$$

由式(2)可知,特征 F 同样经过平均池化和最大池化对信息进行评估,然后采用 7×7 卷积进行提取,最后使用Sigmoid进行归一化。

2.4 基于CBAM注意力机制的行人重识别特征提取网络

本文设计的行人重识别特征提取网络结构如图3所示。首先是对输入的图像集进行预处理,使用GAN网络进行扩增。然后将处理好的图像送入主干网络ResNeXt中进行特征提取,同时在不同位置加入注意力机制,得到有鉴别力的特征,对行人进行分类识别。

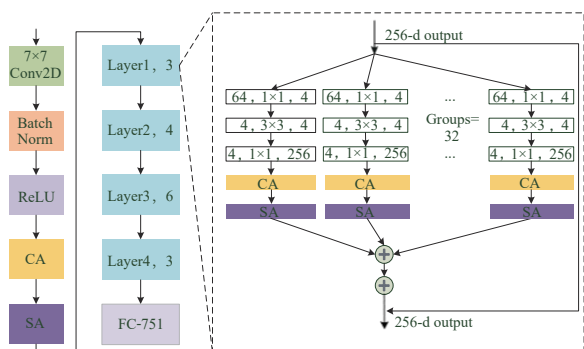


图3 特征提取网络结构

Fig. 3 Feature extraction network structure

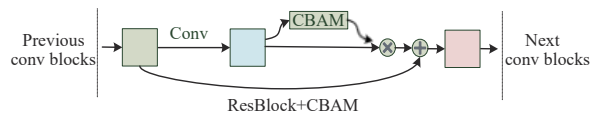


图4 CBAM模块嵌入到ResNeXt

Fig. 4 CBAM module embedded to ResNeXt

本文采用GAN技术扩充大型行人重识别公开数据集Market1501,重构新的数据集,以ResNeXt50为主干网络。分析对比之后,将注意力模块添加在ResNeXt50网络的第一个卷积层以及Layer1的残差块之后(图4),并使用交叉熵损失训练网络,损失函数如式(3)所示:

$$L_{\text{cross}} = - \sum_{j=1}^M \lg(p(n))q(n), \quad (3)$$

式(3)中, n 为行人ID分类数, $p(n)$ 为标签的预测概率, $q(n)$ 为输入真实分布。离线训练深度特征提取网络,将输入图像的尺寸统一为 128×64 ,在新的Market1501数据集上训练40个epoch,最终得到128维的特征向量作为外观特征。改进后模型的可视化结果如图5所示,可以看到前一帧中的ID为149的目标对象,由于ID为18的行人遮挡,导致在后帧出现ID切换的问题,由原来的149切换为184。本文改进算法之后,导入新的模

型权重,重新得到跟踪结果。从图5可以看出,在相同帧处,ID切换问题得以解决。ID为172的行人被ID为17的行人遮挡后仍能保持ID,直观反映了改进网络的有效性。

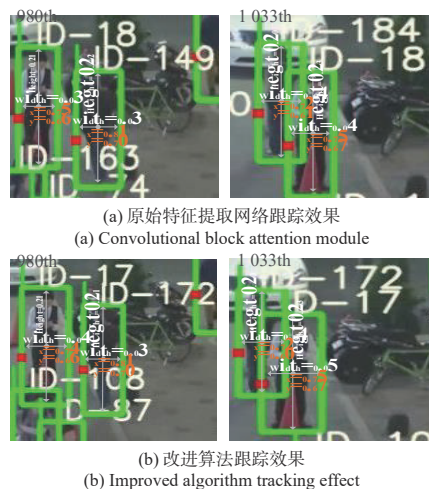


图5 改进前算法与改进后算法跟踪过程前后帧目标ID对比

Fig. 5 Contrast of the target ID before and after frame in the tracking process using the prealgorithm and improved algorithm

3 YOLOv5检测算法的改进

本文采用基于检测的多目标跟踪范式,因此检测作为第一步十分关键。目标检测旨在判断给定的视频帧中除背景信息以外的目标的尺寸和位置。本文研究的目标对象是行人,因此需要检测出行人的位置并用矩形框进行标注。2020年,Glenn Jocher发布了YOLOv5算法,该目标检测模型运行准确度较高,且运行速度快。YOLOv5基于PyTorch实现,可以有效地应用于嵌入式设备和移动端。根据模型的深度以及卷积核的个数,模型共有4个版本:YOLOv5s、YOLOv5m、YOLOv5l、YOLOv5x,模型体积越大,相应的检测精度越高,但是复杂的模型也导致检测速度变慢。

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YOLOv5s^[15]模型为基础的目标检测算法。YOLOv5s模型主要由4部分组成。输入端首先对原始图像进行预处理,预处理操作包括Mosaic数据增强、自适应图

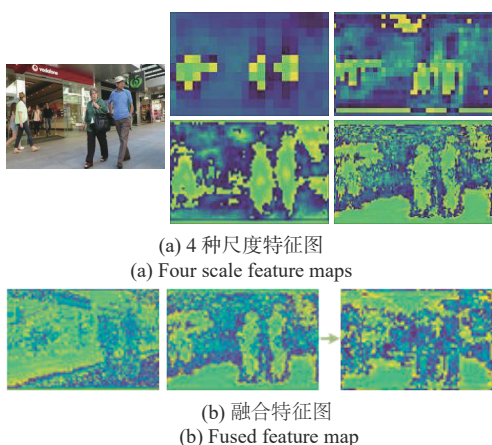


图 7 特征图可视化

Fig. 7 Feature map visualization

表 2 YOLOv5s 消融实验

Tab. 2 Ablation of YOLOv5s

Model	Precision/ %	Recall/ %	mAP/ %	mAP@0.5: 0.95/%
YOLOv5	93.5	92.1	95.8	76.4
YOLOv5+ 上采样	98.6	95.7	99.4	87.1

由表 2 可知,本文在 YOLOv5s 基础算法中 添加上采样操作,构成小目标检测层。改进模型 的各项指标相较于 YOLOv5s 均有明显提高。

3.2 CA 注意力机制

在 CV 领域,注意力机制被广泛应用于图像 分类、目标检测、分割任务^[17]等。注意力机制本

质上是找出需要重点关注的候选区域,并通过一 系列权重参数对图中焦点区域的信息进行增强, 以提取目标更多的细节信息,同时抑制一些无用 信息。文献[18]中提出的 SENet 通道域注意力 机制通过对图像特征通道域的相关性进行建模, 优化特定类别的特征信息,但是 SENet 只考虑通 道间信息的编码,而忽略了位置信息的重要性。 文献[19]中提出的 ECANet 对 SENet 模块进行 了一些改进,提出了一种不降维的局部跨信道交 互策略和自适应选择一维卷积核大小的方法,从 而实现了性能上的优化,但是 ECANet 也并未将 位置信息考虑在内。2021 年最新提出的坐标注 意力机制^[20](Coordinate Attention, CA)将位置信息 嵌入到通道注意力中,它不仅捕获跨通道的关键 信息,还捕获方向感知和位置信息,这使得模型能 够更为精确地定位和识别目标对象。因此,本文 考虑采用 CA 注意力机制,其结构如图 8 所示。

坐标注意力机制首先对输入沿水平方向和垂 直方向对通道进行编码,沿两个方向聚合特征获 得一对具有方向感知能力的特征图。将两个方向 感知特征图拼接在一起,再送入卷积核为 1×1 的 卷积模块进行降维操作,然后经过批量归一化处 理送入 Sigmoid 激活函数得到中间特征图。接着 沿空间维度将中间特征图分成沿水平方向和垂直 方向的独立的张量。最后利用另外两个 1×1 卷 积变换函数对两个张量进行处理得到通道数相同 的输出,分别展开并用作注意力权重。

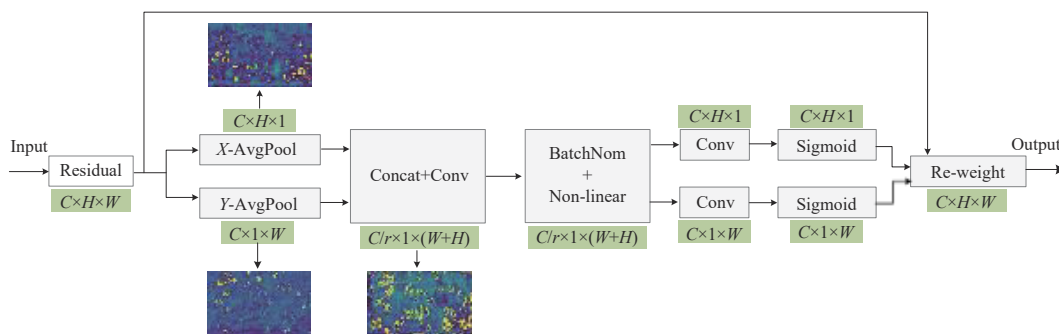


图 8 CA 模块

Fig. 8 CA module

本文将 CA 模块集成到改进后的 YOLOv5s 网络中,经过实验验证,考虑将 CA 模块添加在 Neck 网络中卷积层之间,添加注意力机制赋予目 标信息更大的权重,使得网络对目标信息更加关

注,对检测结果有一定的修正作用。

本文随机选取在 MOT16 训练集中的图像验 证改进模型的效果,原始模型与改进模型的检测 效果对比如图 9 所示。



图9 可视化结果对比

Fig. 9 Visual results comparison

不难看出,原始的网络检测有很多漏检的情况(为了更加直观,我们使用绿色框进行了标

注),发现主要有两个问题:遮挡和远距离导致的目标比例小的问题。因此,本文对原始网络进行改进,从图9可以看出,改进后的模型检测效果更好,尤其是检测到了原来未检测出的远距离行人目标,验证了改进算法的正确性,为后续多目标跟踪任务奠定了基础。

4 基于改进的YOLOv5和重识别的多目标跟踪方法

本文采用的多目标跟踪算法是基于检测的跟踪范式,且为了满足实时性的要求,对目标行人进行在线跟踪。多目标跟踪任务基本是由目标检测、运动预测、外观模型、数据关联模块组成。上文已经获得了精确的目标检测框以及外观模型,然后利用卡尔曼滤波和匈牙利算法完成关联和匹配。本文整体网络架构如图10所示。

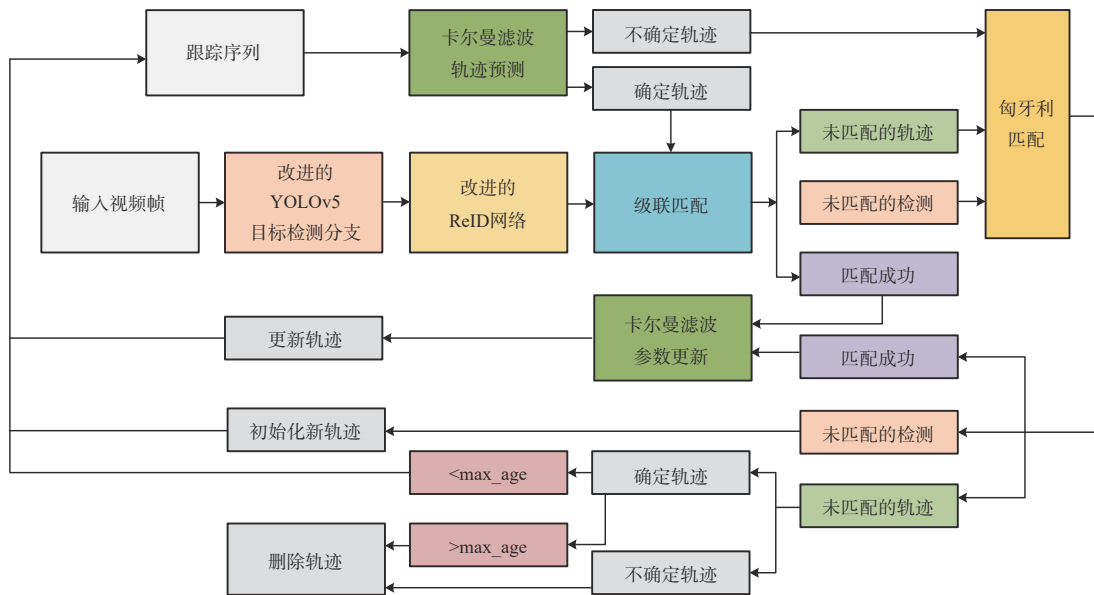


图10 整体网络构架

Fig. 10 Overall network architecture

具体工作流程如下。第一步:输入的视频序列首先经过YOLOv5s检测出前景对象,得到行人的位置响应,并通过检测响应裁剪出当前帧图像中目标对象位置所在的图像块。为了使得检测结果更加精确,本文在此处加入检测层以满足多尺度检测任务,同时加入CA注意力机制使得模型更加优化。第二步:将裁剪出来的

包含目标的图像块送入线下训练好的行人重识别模型来提取深度外观特征信息,本文在此处也做出改进,将主干网络更换为ResNeXt50,以采用相对少的参数达到较好的效果。另外,考虑到实际情况中的遮挡问题给特征提取带来的困难,添加注意力机制CBAM,再用GAN网络扩增后的行人重识别公开数据集Market1501训

练模型,生成的权重文件作为新的外观模型。第三步:将上一步提取到的ReID特征与经过卡尔曼预测的确定轨迹上的ReID信息进行级联匹配建立数据关联,得到3种结果:匹配成功、未匹配的轨迹、未匹配的检测。针对未匹配的检测、未匹配的轨迹以及卡尔曼预测的不确定轨迹再采用匈牙利算法进行二次匹配,同样得到前面提到的3种结果。第四步:通过级联匹配和匈牙利匹配,对于匹配成功的轨迹,将其送入卡尔曼滤波中更新轨迹信息。对于未匹配的检测,将其初始化作为新的轨迹,并且该轨迹必须满足连续3帧都能匹配到目标对象才能被定为确定轨迹。对于未匹配的轨迹,若其是确定轨迹且连续失配帧数少于max_age,则将其加入跟踪序列中;若连续失配帧数大于max_age,则删除轨迹。若未匹配且是未确定轨迹,则直接删除轨迹。第五步:将得到的所有跟踪送入卡尔曼滤波进行预测,得到确定轨迹和不确定轨迹,重复上述步骤。

5 实验结果与分析

5.1 数据集与实验环境

本文提取大型目标检测公开数据集COCO中的行人数据子集作为训练集,用MOT16的训练集^[21]作为验证集验证算法性能。实验环境基于Ubuntu 16.04操作系统,Nvidia GeForce RTX 2080Ti显卡,运行内存为64G,采用Pytorch1.6.0深度学习框架,在Python3.7的服务器下实现。选择多目标跟踪公开数据MOT16测

试集测试本文算法,并将测试结果提交MOT Challenge官网进行评估,与其他先进算法进行对比,并分析模型性能。

5.2 评价指标

为了使模型的评价更加客观准确,并与其他算法进行合理比较,本文采用多目标跟踪领域通用的评估指标进行评估:多目标跟踪准确度(Multi-object Tracking Accuracy, MOTA)、多目标跟踪精度(Multi-object Tracking Precision, MOTP)、多目标跟踪器ID维持能力(Identification F1 Score, IDF1)、行人ID切换次数(ID Switch, IDs)、大多数跟踪目标百分比(Mostly Tracked, MT)、大多数丢失目标百分比(Mostly Lost, ML)。部分评价指标的公式如公式(4)和(5)所示:

$$MOTA = 1 - \frac{N_{FN} + N_{FP} + N_{IDs}}{N_{GT}}, \quad (4)$$

$$MOTP = \frac{\sum_{i,t} d_i^i}{\sum_t c_t}, \quad (5)$$

式中的 t 表示当前帧为第 t 帧, $t \in [1, N]$; N_{GT} 表示真实边界框数量; N_{FN} 表示整个视频的漏检数; N_{FP} 表示整个视频的误检数; N_{IDs} 表示总的行人ID切换次数; d_i^i 表示第 t 帧中第 i 个预测框与真实框之间的重叠率,即IoU(Intersection-Over-Union)距离; c 表示目标成功匹配数量。

5.3 结果分析

对于多目标跟踪,本文选择MOT16数据集进行实验,与几种先进多目标跟踪算法结果进行对比。受文献[22]启发,本文考虑在计算速率时加入检测器需要的时间,结果如表3所示。

表3 本文算法与其他先进算法在MOT16数据集上的对比结果

Tab. 3 Comparative results on this grade algorithm and other advanced algorithm on MOT16 data set

Algorithm	MOTA	MOTP	IDF1	MT/%	ML/%	IDs ↓	帧率/Hz
SORT ^[3]	59.8	79.6	53.8	25.4	22.7	1 423	8.6
DeepSORT ^[4]	61.4	79.1	62.2	32.8	18.2	781	6.4
JDE ^[6]	64.4	—	55.8	35.4	20.0	1 544	18.5
POI ^[23]	66.1	79.5	65.1	34.0	20.8	805	5.0
TubeTK ^[24]	64.9	59.4	59.4	33.5	19.4	1 117	1.0
Ours	66.2	80.8	65.8	35.3	17.6	760	5.8

综合分析指标可知,本文算法与部分先进算法相比有相对优势,但是与JDE算法相比,在实时性方面优势并不突出。JDE算法属于一

阶段跟踪算法,实时性相对较高,但ID切换相对频繁,这是在大量密集行人场景下,由于目标相互遮挡导致的。而本文着重考虑遮挡情况

下跟踪效果差的问题,同时考虑到数据集中行人密集且动态变化,增强了对小目标对象的检测能力,因此改进算法在 MOTA、MOTP 等指标上都有所提升。MOTA 相较于 JDE 算法提升了 1.8%,MOTP 相较于 SORT 算法提升了

1.2%,IDF1 相较于 DeepSort 算法提升了 3.6%,MT 提升了 2.5%,ID 切换频次下降。综上,可以验证本文算法的良好性能,且在实际跟踪场景中具有一定的优势。多目标跟踪效果如图 11 所示。

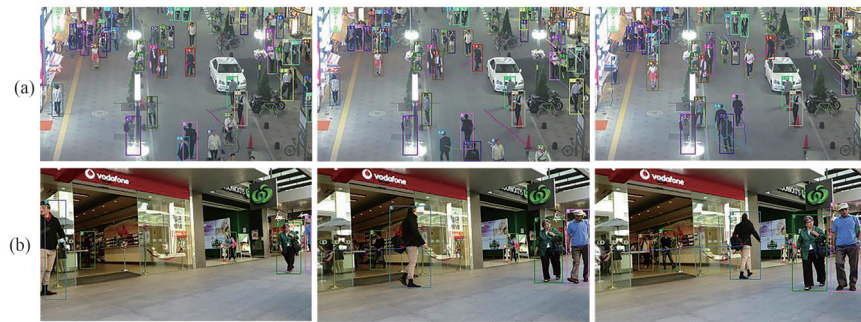


图 11 多目标跟踪效果。(a) MOT16-03; (b) MOT16-09。

Fig. 11 Multi-target tracking effect. (a) MOT16-03; (b) MOT16-09.

6 结 论

本文针对行人多目标跟踪过程中因为遮挡导致的行人 ID 频繁切换,跟踪效果差的问题,提出一种改进算法。该算法选用 TBD 跟踪范式,设计新的特征提取网络,并且通过引入卷积注意力机制,提升了模型对目标对象的表征能力。同

时改进了 YOLOv5 目标检测算法,考虑加入小目标检测层和坐标注意力机制使得检测更加准确,进而提升了跟踪的精度和准确度。实验表明,本文算法有效缓解了遮挡导致的行人 ID 频繁切换的问题,相较于 DeepSort 算法,ID 切换减少了 21,MOTA 提升了 4.8%,并且跟踪速度也达到了实时性的要求。

参 考 文 献:

- [1] LUO W H, XING J L, MILAN A, *et al.* Multiple object tracking: a literature review [J]. *Artificial Intelligence*, 2021, 293: 103448.
- [2] 罗浩,姜伟,范星,等. 基于深度学习的行人重识别研究进展[J]. *自动化学报*, 2019, 45(11): 2032-2049.
LUO H, JIANG W, FAN X, *et al.* A survey on deep learning based person re-identification [J]. *Acta Automatica Sinica*, 2019, 45(11): 2032-2049. (in Chinese)
- [3] BEWLEY A, GE Z Y, OTT L, *et al.* Simple online and realtime tracking [C]//*Proceedings of 2016 IEEE International Conference on Image Processing (ICIP)*. Phoenix, AZ, USA: IEEE, 2016: 3464-3468.
- [4] WOJKE N, BEWLEY A, PAULUS D. Simple online and realtime tracking with a deep association metric [C]//*Proceedings of 2017 IEEE International Conference on Image Processing (ICIP)*. Beijing, China: IEEE, 2017: 3645-3649.
- [5] CHEN L, AI H Z, ZHUANG Z J, *et al.* Real-time multiple people tracking with deeply learned candidate selection and person re-identification [C]//*Proceedings of 2018 IEEE International Conference on Multimedia and Expo (ICME)*. San Diego, CA, USA: IEEE, 2018: 1-6.
- [6] WANG Z D, ZHENG L, LIU Y X, *et al.* Towards real-time multi-object tracking [M]//VEDALDI A, BISCHOF H, BROX T, *et al.* *Computer Vision—ECCV 2020*, Cham: Springer, 2020.
- [7] ZHANG Y F, WANG C Y, WANG X G, *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.

- [8] DUAN K W, SONG B, XIE L X, *et al.* CenterNet: keypoint triplets for object detection [C]//*Proceedings of 2019 IEEE/CVF International Conference on Computer Vision (ICCV)*. Seoul, Korea (South): IEEE, 2019: 6568-6577.
- [9] XIE S N, GIRSHICK R, DOLLÁR P, *et al.* Aggregated residual transformations for deep neural networks [C]//*Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Honolulu, HI, USA: IEEE, 2017: 5987-5995.
- [10] HE K M, ZHANG X Y, REN S Q, *et al.* Deep residual learning for image recognition [C]//*Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Las Vegas, NV, USA: IEEE, 2016: 770-778.
- [11] ZHENG Z D, LIANG Z, YI Y. Unlabeled samples generated by GAN improve the person re-identification baseline *in vitro* [C]//*Proceedings of 2017 IEEE International Conference on Computer Vision (ICCV)*. Venice, Italy: IEEE, 2017: 3774-3782.
- [12] ZHONG Z, ZHENG L, ZHENG Z D, *et al.* Camera style adaptation for person re-identification [C]//*Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City, UT, USA: IEEE, 2018: 5157-5166.
- [13] WOO S, PARK J, LEE J Y, *et al.* CBAM: convolutional block attention module [C]//*Proceedings of the 15th European Conference on Computer Vision*. Munich. Germany: Springer, 2018: 3-19.
- [14] 李天宇,李栋,陈明举,等.一种高精度的卷积神经网络安全帽检测方法[J].液晶与显示,2021,36(7):1018-1026.
LI T Y, LI D, CHEN M J, *et al.* High precision detection method of safety helmet based on convolution neural network [J]. *Chinese Journal of Liquid Crystals and Displays*, 2021, 36(7): 1018-1026. (in Chinese)
- [15] 赵睿,刘辉,刘沛霖,等.基于改进YOLOv5s的安全帽检测算法[J/OL].北京航空航天大学学报:1-16[2022-01-12].<https://kns.cnki.net/kcms/detail/detail.aspx?FileName=BJHK20211120004&DbName=CAPJ2021>.
ZHAO R, LIU H, LIU P L, *et al.* Research on safety helmet detection algorithm based on improved YOLOv5s [J/OL]. *Journal of Beijing University of Aeronautics and Astronautics*: 1-16[2022-01-12]. <https://kns.cnki.net/kcms/detail/detail.aspx?FileName=BJHK20211120004&DbName=CAPJ2021>. (in Chinese)
- [16] 李永上,马荣贵,张美月.改进YOLOv5s+DeepSORT的监控视频车流量统计[J].计算机工程与应用,2020,58(5):271-279.
LI Y S, MA R G, ZHANG M Y. Traffic monitoring video vehicle volume statistics method based on improved YOLOv5s+DeepSORT [J]. *Computer Engineering and Applications*, 2020, 58(5): 271-279. (in Chinese)
- [17] GUO M H, XU T X, LIU J J, *et al.* Attention mechanisms in computer vision: a survey [EB/OL]. (2021-11-15)[2022-01-12]. <https://arxiv.org/abs/2111.07624>.
- [18] HU J, SHEN L, ALBANIE S, *et al.* Squeeze-and-excitation networks [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020, 42(8): 2011-2023.
- [19] WANG Q L, BANG G W, ZHU P F, *et al.* ECA-Net: efficient channel attention for deep convolutional neural networks [C]//*Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Seattle, WA, USA: IEEE, 2020: 11531-11539.
- [20] HOU Q B, ZHOU D Q, FENG J S. Coordinate attention for efficient mobile network design [C]//*Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Nashville, TN, USA: IEEE, 2021: 13708-13717.
- [21] MILAN A, LEAL-TAIXE L, REID I, *et al.* MOT16: a benchmark for multi-object tracking [EB/OL]. (2016-03-02)[2022-01-12]. <https://arxiv.org/abs/1603.00831v2>.
- [22] 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 B J, LI B Z, LIU S. 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. (in Chinese)
- [23] YU F W, LI W B, LI Q Q, *et al.* POI: multiple object tracking with high performance detection and appearance feature [M]//HUA G, JÉGOU H. *Computer Vision—ECCV 2016 Workshops*, Cham: Springer, 2016.

- [24] PANG B, LI Y Z, ZHANG Y F, *et al.* TubeTK: adopting tubes to track multi-object in a one-step training model [C]//*Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Seattle, WA, USA: IEEE, 2020: 6307-6317.

作者简介:



贺愉婷(1998—),女,陕西榆林人,硕士研究生,2020年于西安邮电大学获得学士学位,主要从事基于深度学习的行人再识别及跟踪研究。E-mail: 2356854359@qq.com



车进(1973—),男,宁夏银川人,博士,教授,2014年于天津大学获得博士学位,主要从事图像处理、智能视频方面的研究。E-mail:koalache@126.com