

文章编号: 2095-2163(2022)08-0027-07

中图分类号: TP391

文献标志码: A

# 基于跨尺度特征聚合网络的行人检测与跟踪

刘康安, 肖永超

(上海工程技术大学 机械与汽车工程学院, 上海 201620)

**摘要:** 针对行人检测跟踪过程中目标空间尺度差异大、小目标漏检率高, 以及跟踪稳定性差的问题, 提出一种基于跨尺度特征聚合网络(TFACN)的行人多目标检测跟踪方法。该方法选取 CenterNet 网络作为目标检测器, 并设计跨尺度特征聚合模块(TFA), 对高、低层特征信息进行有效融合, 实现对卷积层特征的增强表示, 有效地降低虚警率、小目标的漏检率和计算复杂度; 其次在高精度识别的基础上, 结合自适应卡尔曼滤波, 通过实时计算残差信息在线更新观测噪声协方差, 对行人运动状态进行自适应估计, 提高行人跟踪的位置估计精度。试验结果表明, 本文算法相较于现有的算法, 可以在较低的计算成本下获得更高的精度。

**关键词:** 行人检测; 特征聚合; 多目标跟踪; 自适应滤波

## Pedestrians detection and tracking based on Trans-scale Feature Aggregation Network

LIU Kang'an, XIAO Yongchao

(School of Mechanical and Automotive Engineering, Shanghai University of Engineering Science, Shanghai 201620, China)

**[Abstract]** Aiming at the problems of large spatial scale difference of targets, high missing rate of small target pedestrians and poor tracking stability in pedestrians detection and tracking process, a pedestrian multi-target detection and tracking method based on Trans-scale Feature Aggregation CenterNet (TFACN) is proposed. Firstly, CenterNet network is selected as the target detector, and the Trans-scale Feature Aggregation module (TFA) is designed to effectively fuse the high and low level feature information, realize the enhanced representation of the feature of the convolutional layer, and effectively reduce the false alarm rate, the missing rate of small targets and the computational complexity. Secondly, on the basis of high precision identification, combined with adaptive Kalman filter, the pedestrians state is estimated adaptively through real-time residual information computation to update observation noise covariance, and the accuracy of pedestrians tracking position estimation is improved. Experimental results show that compared with the existing algorithms, the proposed detection and tracking algorithm can achieve higher accuracy at lower computational cost when tested on Caltech pedestrian dataset.

**[Key words]** pedestrians detection; feature aggregation; multiple object tracking; adaptive filtering

## 0 引言

在复杂的交通环境下, 通过对车辆前方行人进行有效的检测和跟踪, 从而使自动驾驶系统及时做出判断和预警, 对降低交通安全隐患具有重要意义<sup>[1]</sup>。而检测算法模型精度与计算速度的优化, 是保证稳定跟踪的前提与手段<sup>[2]</sup>。

传统的行人检测方法主要是基于特征学习, 通过特征提取和机器学习来实现图像或视频中的行人检测。其中, 方向梯度直方图(HOG)是最常用的特征提取方法, 通过 HOG 行人特征描述子结合支持向量机(SVM)进行行人检测, 具有较强的鲁棒性。与传统的检测方法相比, 基于深度学习的目标检测算法, 采用神经网络替代人为设定的特征提取器, 具有强大的特征表示能力, 显著提高了视频图像中目标

检测的性能。在之前的应用中, 最流行的检测网络是基于锚点的方法<sup>[3]</sup>。这类方法设计了具有预定义尺寸和高宽比的先验框, 并对先验框进行分类和回归, 得到目标的边界框。然而, 边界框的最小和最大尺寸会受到一定限制。在利用锚点检测目标时, 由于不同范围内目标的大小不变性, 会导致小目标行人的漏检, 降低了检测精度; 同时由于后处理需要从候选边界框中过滤出最终的边界框, 增加了计算复杂度。近年来, 人们提出了基于关键点的目标检测方法, 此类方法无需设计先验框, 避免了大量先验框造成的复杂计算。众多方法中, 最具代表性的方法有 ConerNet<sup>[4]</sup>、ExtremeNet 以及 CenterNet。其中, ConerNet 通过检测边界框的左上角和右下角, 以确定目标的位置和大小, 但这些角的语义信息相对较弱, 难以检测, 且该方法还需要对属于同一目标的 2

**作者简介:** 刘康安(1997-), 男, 硕士研究生, 主要研究方向: 视觉测控与智能导航。

**收稿日期:** 2021-11-02

个点进行后处理,点的匹配难度较大。ExtremeNet使用最佳关键点估计框架,通过预测目标中心找到极值点。然而,由于所涉及的点数较多,其计算成本较高、且检测精度也未见显著提高。此外,由于网络过于关注边缘,容易导致误检。为此,Duan等人<sup>[5]</sup>提出了CenterNet,该算法只需要提取每个对象的中心点,就能回归出宽高尺寸、偏移量等目标,具有计算成本低、检测精度高且速度快等优点。

一般情况下,目标跟踪算法是根据前一帧的位置来预测下一帧的目标位置。现有行人跟踪算法包括均值漂移<sup>[6]</sup>、粒子滤波<sup>[7]</sup>、核相关滤波和深度学习等。Li等人<sup>[8]</sup>通过特征融合和均值漂移跟踪行人,并引入尺度不变特征变换(SIFT),增强跟踪方法的鲁棒性。但行人被遮挡时(即特征点数量非常有限),该方法则无法精确定行人位置。Wu等人<sup>[9]</sup>基于粒子滤波算法,对粒子进行均值移位重采样,并在局部区域实现快速收敛。Henriques等人<sup>[10]</sup>提出了基于KCF(kernelized correlation filters)的跟踪算法,该算法只需指定目标的初始位置,就可以预测目标在下一时刻的位置,速度较快。然而,当形状发生较大变化时,会对预测结果造成影响;当目标被遮挡时,极易导致目标信息丢失。近年来,随着深度学习技术在计算机视觉领域的兴起,深度学习算法也被引入到目标跟踪研究领域。该方法可以同时检测目标,并学习连接特征。如re-ID特征、外观特征或运动特征,从而进行数据关联。Gordon等人<sup>[11]</sup>提出利用长短时记忆(Long-Short Term Memory, LSTM)构建一种实时递归回归网络,从而对一般对象进行视觉跟踪。虽然该方法具有较好的实时性,但其跟踪结果不适用于杂乱区域内的小目标,会产生虚警,在快速运动时也会导致跟踪失效。

为了解决复杂道路场景下行人检测与跟踪所面临的问题,本文从优化目标检测器和跟踪算法两个方面考虑,提出了一种基于跨尺度特征聚合网络与自适应卡尔曼滤波的多目标行人检测跟踪算法。论文主要完成任务可表述如下。

(1)提出了从视频数据中检测目标行人的TFACN和ROI-TFACN方法。该方法利用跨尺度特征聚合(TFA),有效估计目标行人的中心关键点、尺寸和回归偏移量。在估计过程中,TFA使用双向信息流,在不同的特征层之间上下流动,以提高准确性和效率。归功于该反馈流,TFA帮助网络了解上层对下层的影响,反之亦然。

(2)提出RSCR(region of interest-scale-crop-

resize)方法,结合连续帧的上下文信息,进一步合并TFACN和ROI-TFACN。该算法有助于信息在2个被提出的网络之间流动,通过去除背景效应,放大小目标,帮助检测器聚焦于ROI,不仅提高了精度,而且进一步降低了算法的计算复杂度。

(3)提出了一种优化的自适应卡尔曼滤波算法,通过实时调整观测噪声,提高行人跟踪准确度。

## 1 算法流程

本文提出的整体方法主要由TFACN、RSCR、ROI-TFACN和动态估计组成,用于检测和跟踪行人。行人多目标检测跟踪流程如图1所示。在检测部分,使用TFACN获取上一帧目标的边界框。根据最终候选边界框的大小或边界框预测得分,选择最终候选边界框用于初始化跟踪器,并将其设置为RSCR输入值。在当前帧中,根据RSCR获得的ROI,将 $512 \times 512$ 像素的输入帧裁剪为 $128 \times 128$ 像素,并用于ROI-CFACN中预测目标边界框,从而最小化目标搜索区域,扩大小目标。将每一帧中获得的最终候选边界框保存为观测量。在跟踪部分,采用卡尔曼滤波算法来预测行人在当前帧中的状态,并采用匈牙利算法视频,对相邻帧中的行人进行关联,实现对行人多目标的稳定跟踪。

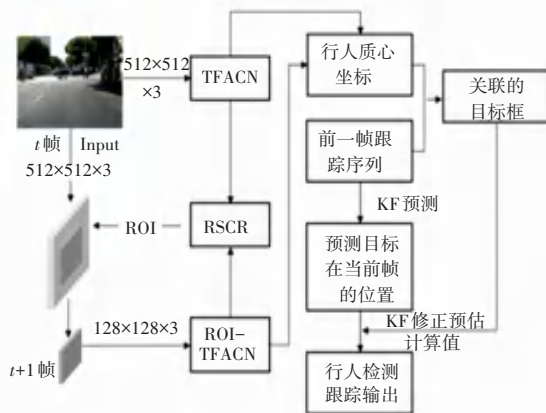


图1 行人多目标检测跟踪流程图

Fig. 1 Flow chart of pedestrians multi-target detection and tracking

本文选取CenterNet网络作为目标检测器,将视频中的连续帧作为全卷积CNN的输入,生成热力图(获取中心点)、大小(边界框的高度和宽度)和偏移量,具有计算成本低、检测精度高、且速度快等优点。在CNN模型中,每个卷积层只能处理特定的尺度。然而,不同行人目标的大小差别很大,基于CenterNet的模型只检测中心点,获取的目标信息较少,导致小目标行人的误检和虚警较多。因此,为保

证模型对尺度变化的鲁棒性, 必须提取多尺度特征。为了提高对小目标和遮挡目标的检测效果, 本文提出基于中心点的跨尺度特征聚合网络 (TFACN)。TFACN 以  $512 \times 512$  像素帧作为输入, 对行人进行识别和定位, 并提出候选的 ROI。第二个网络 ROI-TFACN 使用裁剪的  $128 \times 128$  像素帧作为输入, 通过降低背景复杂度来定位较低像素尺度的行人。这种策略不仅减少了计算成本, 还限制了目标搜索时容易出现的误报次数。为了对多级特征图进行融合, 本文设计了一个基于 ResNet-18 的跨尺度特征聚合模块 (TFA), 以便通过其反馈连接提取更多的重要信息。

传统的多尺度特征聚合是针对不同的特征层进行聚合。在使用骨干网络提取特征后, 给出一个多尺度特征列表  $C^{in} = (C_{11}^{in}, C_{12}^{in}, C_{13}^{in}, C_{14}^{in}, C_{15}^{in})$ 。其中,  $C_{li}^{in}$  表示  $li$  级特征输出, 本文的目标是找到能够有效聚合不同特征并输出新特征列表的变换  $f: C^{out} = f(C^{in})$ 。图 2(a) 显示了传统的自顶向下的 FPN 方法。取 3~7 层的卷积特征层作为输入特征  $C^{in} = (C_3^in, C_4^in, C_5^in, C_6^in, C_7^in)$ 。其中,  $C_i^in$  表示输入图像中分辨率为  $1/2^i$  的特征层。

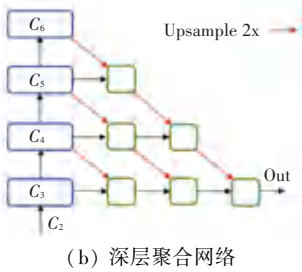
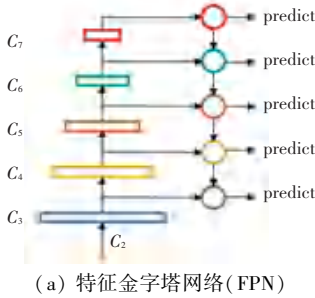


图 2 不同特征网络模型示意图

Fig. 2 Schematic diagram of different characteristic network models

例如, 如果输入分辨率是  $512 \times 512$ , 那么  $C_3^in$  代表特征级别为 3 ( $512/23 = 64$ ), 分辨率为  $64 \times 64$ 。FPN 以自上而下的方式聚合多尺度特征:

$$\begin{aligned} C_7^{out} &= Conv(C_7^{in}) \\ C_6^{out} &= Conv(C_6^{in} + Resize(C_7^{out})) \\ &\dots \\ C_3^{out} &= Conv(C_3^{in} + Resize(C_4^{out})) \end{aligned} \quad (1)$$

其中, *Resize* 通常是特征匹配的上采样或下采样操作, *Conv* 通常是特征处理的卷积操作。

由于层间单向信息流动, 这种传统方法通常会抑制结果的准确性。为改进结果, 本文提出了深度多尺度特征聚合 (deep Multi-scale Feature Aggregation, MFA), 如图 2(b) 所示。然而, 这种网络也需要更多的参数, 使得其计算复杂度很高。令 CenterNet 使用 2 种不同的骨干网来比较得到的准确性和效率。结果表明, MFA 具有较好的精度, 但效率较低。为了在有限的计算资源下提高精度和效率, 本文提出了一种新的基于 MFA 和 TFA 的网络。

在 CenterNet 中, 网络被设计成只有单一信息流通过的网络。为了提高精度, 采用 MFA 并行信息流网络, 设计了多尺度特征聚合中心网 (MFACN)。同时, 为了进一步改进 MFACN, 设计了带有反馈架构的 TFA 双向平行流网络, 如图 3 所示。使用 ResNet 作为主干, 提取不同层的特征; 将 ResNet 第 2 层到第 5 层的输出作为 TFA 的输入。在 TFA 的设计中, 首先最小化上下节点, 通过最小化不同层之间的上节点聚合并以交叉三角形的形式连接, 使计算代价最小化, 提高检测精度。其次, 通过将 TFA 中最后一个中间层 ( $C_4, C_3, C_2$ ) 与每个阶段最后一个卷积层 ( $C_5, C_4, C_3$ ) 连接, 增加下节点聚合 (如图 3 中给出的红色连接), 以提高检测精度。此外, 通过由下向上节点获得改进的检测器。当网络层较小时, 使用下采样或上采样层 (如池化或上采样层) 是无效的。为了解决这个问题, 使用一个 *stride* 为 2 的卷积层作为下采样层, 一个转置卷积层作为上采样层, 有助于更好地理解不同大小的输出特征层之间的相关性。TFA 的计算公式如下:

$$\begin{aligned} C_5^{out} &= \sigma(Conv(C_5) + TransConv(C_4^{td})) \\ C_4^{out} &= \sigma(\sigma(Conv(C_4^{td}) + TransConv(C_5^{out})) + \\ &\quad ConvSt(C_3^{td})) \\ C_3^{out} &= \sigma(\sigma(Conv(C_3^{td}) + TransConv(C_4^{out})) + \\ &\quad ConvSt(C_2^{td})) \\ C_3^{out} &= \sigma(Conv(C_2^{td}) + TransConv(C_3^{out})) \end{aligned} \quad (2)$$

其中,  $\sigma$  是 ReLu 函数; *ConvSt* 是用于特征处理和特征匹配的 *stride* 为 2 的卷积和下采样操作; *TransConv* 是用于特征处理和特征匹配的转置卷积和上采样操作;  $C_4^{td}$  是自上而下路径上第 4 级的中间特征, 显示了不同特征的信息流在一个向上向下的方向。





图3 具有更高精确度和计算效率的TFA网络结构

Fig. 3 TFA network structure diagram with higher accuracy and computational efficiency

TFA 简单而有效,允许更高分辨率的输出(步幅4)。同时,将TFA的4个转置卷积层的通道分别改为512、256、128和64,上卷积滤波器使用均值为零的正态高斯分布初始化。

TFACN与ROI-TFACN的网络结构如图4所示。由图4可知,TFACN检测器主要由ResNet18骨干网络、TFA跨尺度特征聚合模块和预测模块三部分组成。TFACN和ROI-TFACN框架皆包括 $3\times 3$ 卷积核、 $1\times 1$ 卷积核、Keypoint热图预测头、回归预测头和尺寸预测头。ROI-TFACN是TFACN的轻量化版本,不同之处在于,聚合网络中TFA层的通道深度分别为256、128、64和32。2种算法中的主干网络具有相同的层数和通道深度。由图4可看到,对于每个框架,主干网络的特征都通过TFA进行传递。

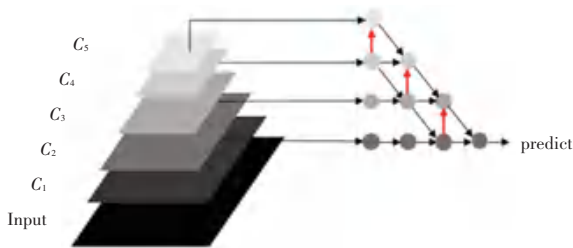


图4 TFACN与ROI-TFACN的网络结构图

Fig. 4 Network structure of TFACN and ROI-TFACN

## 2 基于卡尔曼滤波的行人多目标跟踪

### 2.1 行人目标位置预测

卡尔曼滤波预测目标位置分为预测和更新两个阶段。其中,预测阶段主要根据上一时刻行人的位置来预测行人在当前帧的位置;更新阶段主要根据当前帧检测到的行人位置来更新预测阶段的行人位置。

以视频第一帧检测到的目标行人位置 $X_0 = [c_x, c_y, \dot{c}_x, \dot{c}_y]$ 为初始状态, $(c_x, c_y)$ 表示各目标包围框的中心, $\dot{c}_x, \dot{c}_y$ 分别为目标中心点在 $x$ 轴和 $y$ 轴上的速度分量。假设 $X_{t-1}$ 为上一时刻目标行人的状态, $A$ 为状态转移矩阵, $P_{t-1}$ 为上一时刻目标位置的误差协方差矩阵, $Q$ 为过程噪声协方差矩阵,则预测阶段方程为:

$$X_{t/t-1} = AX_{t-1} \quad (3)$$

$$P_{t/t-1} = AP_{t-1}A^T + Q \quad (4)$$

其中, $X_{t/t-1}$ 为上一时刻预测的当前时刻的行人位置, $P_{t/t-1}$ 为预测的当前状态误差协方差矩阵。

假设 $H$ 为观测矩阵, $K_t$ 为当前时刻的卡尔曼增益, $R$ 为观测噪声协方差矩阵, $X_t$ 表示当前时刻的行人位置, $Z_t$ 为当前时刻行人位置的观测量, $P_t$ 为当前时刻估计误差协方差矩阵。更新阶段方程为:

$$K_t = P_{t/t-1}H^T(HP_{t/t-1}H^T + R)^{-1} \quad (5)$$

$$X_t = X_{t/t-1} + K_t(Z_t - HX_{t/t-1}) \quad (6)$$

$$P_t = P_{t/t-1} - K_tHP_{t/t-1} \quad (7)$$

### 2.2 基于残差信息的自适应卡尔曼滤波

卡尔曼滤波器在迭代更新时的残差为:

$$n_t = Z_t - HX_{t/t-1} \quad (8)$$

由上式可得残差的协方差矩阵为:

$$E[n_t n_t^T] = C_{n_t} = HP_{t/t-1}H^T + R \quad (9)$$

此时滤波增益可表示为:

$$K_t = P_{t/t-1}H^T(HP_{t/t-1}H^T + R)^{-1} = P_{t/t-1}H^TC_{n_t}^{-1} \quad (10)$$

为降低复杂交通环境对观测的干扰,需要通过计算 $n_t$ ,从而调节 $K_t$ 对行人进行准确跟踪。将观测受到的干扰反映到噪声 $R$ 上,由式(9)得:

$$C_{n_t} = HP_{t/t-1}H^T + r_t R \quad (11)$$

其中, $r_t$ 是 $R$ 的比例因子矩阵。由式(11)可知:

$$r_t = (C_{n_t} - HP_{t/t-1}H^T)R^{-1} \quad (12)$$

在 $[0, t]$ 时刻,有:

$$C_{n_t} = \eta_t = \frac{1}{t} \sum_{i=1}^t n_{t,i} n_{t,i}^T \quad (13)$$

由式(12)、(13)可知:

$$r_t = (\eta_t - HP_{t/t-1}H^T)R^{-1} \quad (14)$$

为使比例因子矩阵 $r_t$ 的对角元素 $r_{t,ii} \geq 1$ ,定义该矩阵的第 $ii$ 个对角元素为:

$$r_{t,ii}^* = \max\{1, r_{t,ii}\} \quad (15)$$

在线更新后的观测噪声为:

$$R^* = r_t^* R \quad (16)$$

### 2.3 目标关联

匈牙利算法主要用作求解二分图的指派问题,并具有时间、空间复杂度较低等优点。本文采用匈牙利算法解决跟踪目标和检测目标的关联问题。

假设:上一时刻跟踪到的行位置坐标集合为 $T = \{(x_1^{t-1}, y_1^{t-1}), (x_2^{t-1}, y_2^{t-1}), \dots, (x_n^{t-1}, y_n^{t-1})\}$ ,当前时刻检测到的行人的位置坐标的集合为 $D = \{(x_1^t, y_1^t), (x_2^t, y_2^t), \dots, (x_n^t, y_n^t)\}$ 。以欧氏距离对检测集 $D$ 和跟踪集 $T$ 进行关联。欧氏距离的计算公式为:

$$d_n = \sqrt{(x_n^{t-1} - x_n^t)^2 + (y_n^{t-1} - y_n^t)^2} \quad (17)$$

若存在某一检测目标无法与已有的任何跟踪目标相关联,则将其作为新的跟踪目标并添加对应的 ID。在实际跟踪过程中,当计算的欧氏距离大于一定的阈值  $d_{max}$  或连续  $f_{max}$  帧没有对应的检测行人与当前跟踪集中的某个目标相匹配,则判断此跟踪目标消失。

### 3 实验验证

本文实验基于 Python 编译环境和 Pytorch 框架实现。硬件平台主要配置包括: Intel (R) Core (TM) i5-9400f CPU、16 G RAM、NVIDIA GeForce GTX 2080 GPU。本研究实验分为检测算法实验与跟踪算法实验。拟展开研究分述如下。

#### 3.1 行人检测实验

本次实验选用 Caltech 行人数据集进行模型训练和测试。训练时网络参数设置为: 学习率从 0.000 1 至 0.000 001 递减, 衰减系数为 0.999 5。对算法进行 48 000 次训练迭代, *batch* 设置为 16, 并且对原数据进行数据增强, 以提高模型的泛化性能。

为了评估本文算法的检测性能、检测速度以及先进性, 采用精确率 - 召回率 (*PR*) 曲线, 以及均值平均精度 (*mAP*) 和传输速率进行评价, 并与其它通用目标检测算法进行对比实验。各检测器 *P-R* 对比曲线如图 5 所示。由图 5 可知, 本文提出的跨尺度特征聚合中心网络, 相对于其它检测算法具有更高查全率和召回率。

实验中不同检测算法在数据集上的测试结果见表 1。由表 1 可见, 本文算法检测精度达到 96.56%, 优于其它算法, 并以 46 f/s 的检测速度达到了实时检测的要求。

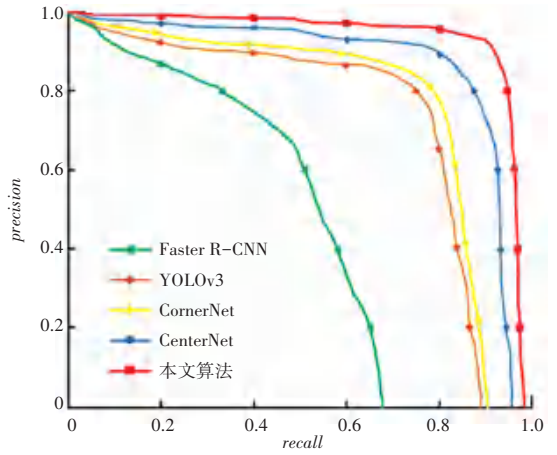


图 5 各检测器 *P-R* 曲线图

Fig. 5 *P-R* curve of each detector

表 1 不同检测方法在 Caltech 数据集上的检测结果对比

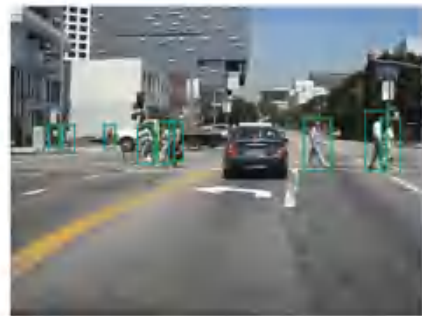
Tab. 1 Comparison of detection results of different detection methods on Caltech dataset

算法	<i>mAP</i> / %	检测帧率 / ( $f \cdot s^{-1}$ )
Faster R-CNN	84.71	7
YOLOv3	94.25	34
CornerNet	94.37	11
CenterNet	95.14	31
本文算法	96.56	46

图 6 为本文算法在测试集上的检测结果。其中, 图 6(a) 为前方行人较多且遮挡率较高的交通路口场景; 图 6(b) 是目标距离较远的城市道路场景, 该场景下小目标行人较多。从效果图中可以看出, 本文方法在多类别行人检测中, 检测框与目标行人重合程度大, 对小目标行人以及受遮挡行人具有较好的检测效果, 误检、漏检和重复检测现象较少。



(a) 拥挤场景下检测效果



(b) 远距离目标检测效果

图 6 行人多目标检测效果图

Fig. 6 Pedestrians multi-target detection results diagram

#### 3.2 行人跟踪实验

为验证改进的卡尔曼滤波行人跟踪算法的有效性, 采用多目标跟踪准确度 (*MOTA*) 和多目标跟踪精

确度 (*MOTP*) 两个指标衡量算法的总体性能。在数据集上进行测试, 并将其与粒子滤波、KCF、Sort 等代表性多目标跟踪算法进行比较。对比结果见表 2。

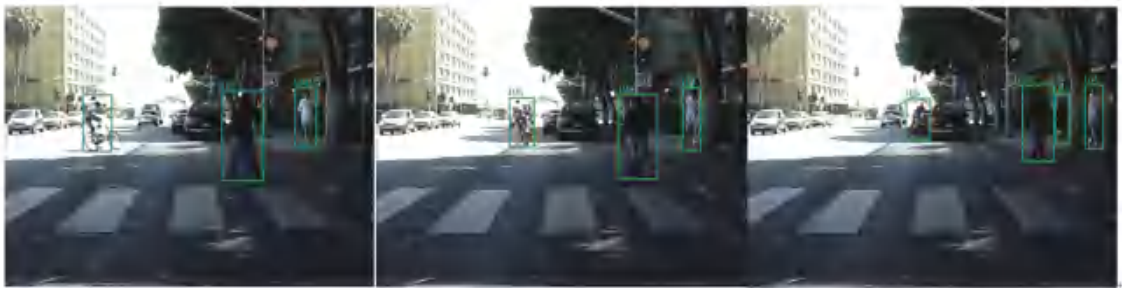
由表2可知,本文算法在  $MOTA$  和  $MOTP$  两个指标上相比于其它算法均具有明显的提升。在测试集上的多目标行人跟踪准确度达到 72.43%,跟踪精确度为 84.26%,满足复杂道路场景下稳定跟踪的要求。

表2 多目标跟踪算法的  $MOTA$  与  $MOTP$  对比

Tab. 2 Comparison of  $MOTA$  and  $MOTP$  for multi-objective tracking algorithms

算法	$MOTA$ / %	$MOTP$ / %	ID_Switch
粒子滤波	65.24	77.61	306
KCF	67.59	83.12	217
Sort	69.51	82.95	239
本文算法	72.43	84.26	152

实际行人多目标跟踪效果如图7所示。在城市道路上对多目标行人进行跟踪实验,跟踪目标在图像上较小且距离较远,视频中部分行人出现不同程度的遮挡,甚至消失。可以看到,本文算法的跟踪鲁棒性较好,无ID跳变,无明显漏检、误检,在行人不断进入、消失的情况下,仍能够准确跟踪到远距离小目标行人,且具有较好的抗遮挡性。图8为本文算法以及 Sort、KCF 和 PF 等算法跟踪到的目标质心在图像中的像素误差,4种方法得到的单帧平均像素误差分别为 2.85、4.67、5.35、6.96。由此可见,本文算法在复杂交通环境下具有良好的性能。



(a) 第399帧

(b) 第409帧

(c) 第419帧

图7 行人多目标跟踪效果图

Fig. 7 Pedestrians multi-target tracking results diagram

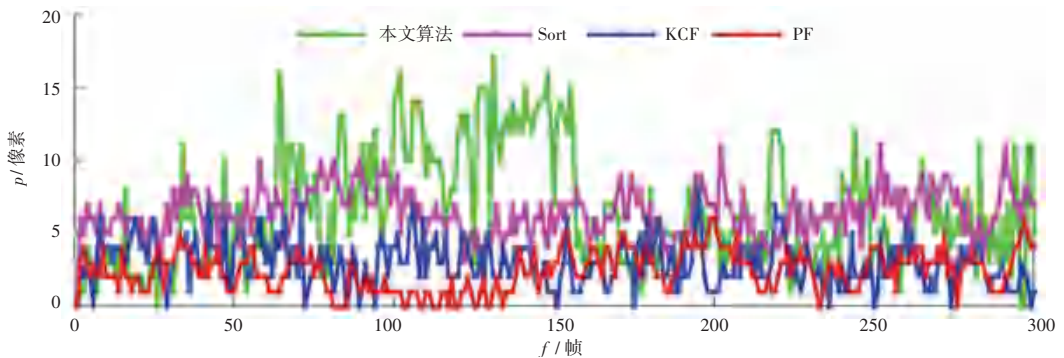


图8 各跟踪算法的目标质心误差对比

Fig. 8 Comparison of target centroid error of each tracking algorithm

## 4 结束语

本文针对传统的基于检测的目标跟踪算法对环境干扰适应能力差、漏检较多、且跟踪准确率不高的问题,提出了一种基于跨尺度特征聚合中心网络与自适应卡尔曼滤波相结合的行人多目标检测与跟踪方法。在原始的特征提取模块中加入特征聚合模块,有效地聚合具有语义鲁棒性的高层特征和具有精确定位信息的低层特征,生成一个更有效的特征

图,提高识别检测精度。并结合自适应卡尔曼滤波算法,通过对受扰观测进行权重在线更新,从而对运动行人进行鲁棒跟踪。实验结果表明,本文提出的跨尺度特征聚合网络相对其它类似方法可以获得更好的性能,对行人多目标的跟踪,提出的算法在  $MOTA$  和  $MOTP$  两个指标上相较于其它方法有较好的提升, ID 跳变数得到了大幅改善,可以满足复杂道路场景下行人检测与跟踪应用的需求。

(下转第42页)