结合行人检测与单应性变换的安全社交距离估计

张建贺,陶杭宇,王亚名,陈积泽,姜晓燕

(上海工程技术大学 电子电气工程学院,上海 201620)

摘 要:为解决未标定摄像头监控视频中行人安全社交距离的估计问题,提出将行人检测、单应性与尺度估计相结合的方法,对单目相机中行人是否处于安全社交距离进行二分类。首先基于YOLOv5s框架,采用MSCOCO数据集中只含有行人的数据训练得到鲁棒性较好的行人检测器;然后根据相机成像模型假设,推导出从场景地面到图像平面的单应性矩阵,再通过人类平均身高和图像行人检测框高度估计从场景地面到图像中行人局部区域的尺度信息,从而在图像上投影出行人椭圆形安全区域;最后通过计算图像中行人安全区域的重叠情况判断其是否违反安全社交距离。实验结果表明,当IoU=0.5时,该方法在MSCOCO只含行人数据的验证集上的行人检测准确率、召回率和AP分别达到81.39%、82.39%和76.52%;在OTC数据集上行人安全社交距离二分类的准确率、召回率和F1值分别达到98.99%、89.12%和93.79%。所提方法在一般监控场景下对行人安全社交距离违反情况的检测性能较佳。

关键词:行人社交距离估计;目标检测;单应性矩阵;单目视觉

DOI:10.11907/rjdk.211622 **开放科学(资源服务)标识码(OSID)**:

中图分类号:TP751 **文献标识码**:A **文章编号**:1672-7800(2022)003-0089-06

Safety Social Distance Estimation Based on Person Detection and Homography Transformation

ZHANG Jian-he, TAO Hang-yu, WANG Ya-ming, CHEN Ji-ze, JIANG Xiao-yan

(*School of Electronic and Electrical Engineering*, *Shanghai University of Engineering Science*, *Shanghai* 201620, *China*)

Abstract: In order to solve the problem of estimating the safe social distance of pedestrians in surveillance videos with uncalibrated cameras, a method combining pedestrian detection, homography and scale estimation was proposed to classify whether pedestrians were in the safe social distance in monocular cameras. Firstly, based on the YOLOv5s network, the pedestrian detector with better robustness was trained by using the data of MSCOCO containing only pedestrians. Secondly, according to the assumptions of the camera imaging model, the homography matrix from the scene ground to the image plane was deduced, and the scale information from the scene ground to the local area of the image was estimated by the average height of human and the height of the image pedestrian detection box, so as to project the elliptical pedestrian safety area on the image. Finally, the overlap of the pedestrian safety area in the image is calculated to determine whether the safe social distance is violated. Experimental results show that when IoU=0.5, pedestrian detection precision, recall rate and AP of MSCOCO dataset reach 81.39%, 82.39% and 76.52%, respectively. The precision, recall rate and F1-score of pedestrian safety social distance classification of OTC dataset reached 98.99%, 89.12% and 93.79%, respectively. The proposed method has high performance in pedestrian detection and safe social distance violation detection under normal security camera circumstances.

Key Words: pedestrians social distance estimation; object detection; homography matrix; monocular vision

0 引言

新冠肺炎的全球大流行严重影响了公共卫生安全。世界卫生组织提出了减少病毒传播的指导方针,其中最重要的措施之一是外出时与他人保持安全的社交距离。社交距离是一种公共卫生实践,旨在防止患病人群与健康人群密切接触,以减少疾病传播机会[1]。最新研究表明[2],与

收稿日期:2021-04-23

基金项目:国家自然科学基金项目(61772328);国家基金委联合基金重点项目(U2033218)

作者简介:张建贺(1991-),男,CCF会员,上海工程技术大学电子电气工程学院硕士研究生,研究方向为数字图像处理、计算机视觉和机器学习;姜晓燕(1985-),女,博士,上海工程技术大学电子电气工程学院副教授,研究方向为多目标跟踪、语义分割和视觉SLAM。本文通讯作者:姜晓燕。

感染者距离大于1m时被传染几率为2.6%，1m内则可能高达12.8%。因此，监测和调节人与人之间的社交距离在抑制病毒传播方面具有至关重要的作用。

从图像或视频中自动估计现实场景中行人之间的物理距离被称为视觉社交距离估计[3]，其中基于单张图像的视觉社交距离估计又称为单目视觉社交距离估计，既可以用于基于安全原因的社交距离监测，也可以利用监测结果分析其各种影响。虽然目前也有利用移动电子设备、Wi-Fi或蓝牙技术的测距方法[4]，但视觉社交距离估计方法具有非侵入性、安全性、易部署等优点，应用较为广泛。此外，视频监控设备遍布主要公共场所，为单目视觉社交距离估计提供了广泛基础。通过单目视觉社交距离估计方法提醒人们及时保持安全社交距离，不仅能有效抑制病毒传播，而且通过统计分析社交距离违规情况可以识别出拥挤的危险区域，为公共场所的安全性改进提供建议。

1 相关研究

基于单目视觉的社交距离估计涉及到两个任务[3]，分别为行人检测和距离估计，具体分为4种方法：第一种是基于行人检测框中心的像素距离计算。例如姚博等[5]利用YOLOv4行人检测模型得到行人检测框，然后直接计算每个行人检测框中心点像素坐标之间的欧氏距离，最后通过设定阈值判断行人是否违反安全社交距离。该种方法只能粗略地估计像素距离，并没有直接估计现实场景中人与人之间的物理距离；第二种是基于行人检测和已知相机内外参数，将像素行人距离转换为现实场景的物理距离。例如Rezaei等[6]同时利用行人检测器和相机内外参数计算行人社交距离，但现实场景中无法直接获取相机的内外参数，因此该方法有一定的局限性；第三种是基于行人检测和手动校准估算社交距离。例如Shorfuzzaman等[7]首先利用行人检测模型检测行人，再通过手动选取4个点并校准尺度信息得到透视变换矩阵，从而得到鸟瞰图，最后计算鸟瞰图中的行人距离，即行人物理社交距离。该种方法因需要手动校准而无法直接应用于一般场景；第四种仅基于单张图像而无需相机校准信息。例如Gupta等[8]利用行人检测和追踪技术完成行人定位与计数，然后估计每个行人到成像平面的距离，再由此计算行人之间的距离；Aghaei等[9]提出利用人体姿态估计算法检测未校准图像中行人之间安全距离的违反情况，具体方法为通过两个预定义的比例参数估计地面到图像平面的单应性矩阵，然后使用人体关节点长度先验信息及其对应的图像关节点像素长度得到尺度信息，最后估计得到图像中行人椭圆形安全区域，通过判断其是否重叠来检测行人违反安全社交距离的情况。

综上可知，如果已知监控摄像头的内外参数，可以准确将像素距离转换为现实场景的物理距离。然而对于大多数已安装在公开场所的摄像头来说，即使通过摄像头厂商得到相机内参，但外参由摄像头姿态决定，无法直接获取。为使算法不局限于摄像头内外参数必须已知的条件，对于未知摄像头校准信息的单张图像，本文引入Aghaei等[9]提出的通过识别图像中行人局部近邻安全区域是否重叠以检测行人安全社交距离违反情况的方法。然而，Aghaei等[9]采用姿态关键点检测估计尺度信息，由于行人一直在移动，各个关节部位在帧间产生的变化幅度较大，加之姿态关键点算法的估计误差，即使利用近似刚体的躯干高度估计尺度信息，也会引起较大误差，使得前后帧的行人安全区域估计不符合实际情况。为了弥补该误差，本文直接使用行人检测框的高度作为计算尺度的基础，同时训练了一个鲁棒性较好的行人检测器生成行人检测框，其不受人体关节局部运动的影响，能提供帧间稳定的行人检测框高度，提高了尺度估计的稳定性。

2 YOLOv5s行人检测网络

一个鲁棒性好的行人检测模型需要克服行人形态变化、行人尺度变化、场景变化等困难。深度学习技术中的卷积神经网络（Convolutional Neural Network，CNN）能较好地提取高层语义特征与多尺度特征。基于深度学习的目标检测方法一类是以R-CNN[10]为代表的两阶段算法，另一类是以YOLO[11]为代表的基于回归的单阶段算法。理论上，两阶段算法比单阶段算法精度高而速度慢。以Faster-RCNN[12]为代表的两阶段目标检测算法模型由卷积层、区域候选网络、感兴趣区域池化层和预测层组成，需要先提取特征产生候选框，再对候选区域进行分类和候选框调整，因此推理时间较长。而以YOLO系列为代表的单阶段目标检测算法模型仅由卷积层和预测层组成，一次性完成目标定位与分类。目前新提出的YOLOv5[13]目标检测算法达到了速度和精度的先进水平。

YOLOv5包括YOLOv5s、YOLOv5m、YOLOv5l、YOLOv5x，它们的网络深度和宽度依次增大，可适用于不同速度和精度要求的场合，其中YOLOv5s体积最小，权重仅14.4MB。本文在行人检测模块中利用推理速度最快的轻量型YOLOv5s网络，其具有与其他领先方法（例如SSD[14]或Faster-RCNN[12]）相当的性能。YOLOv5s的原始工作是基于MSCOCO数据集训练80类目标检测器[15]，最简单的方法是使用多类别检测器，只提取检测出的行人目标而忽略其他类别，这意味着如果图片中还有其他类别的物体，检测器也会检测到它们。为了提高检测器的有效性，本文利用MSCOCO数据集中只包含行人的数据训练一个只检测行人的单类别目标检测器。

YOLOv5s模型结构如图1所示，分别由Input、Backbone、Neck和Prediction组成。Input：输入图像大小为640×640×3，数据增强方法包括Mosaic、随机翻转，使用自适应锚

框计算。Backbone：包含 Focus、CSPNet[16]（Cross Stage Partial Networks）和 SPP（Space Pyramid Pooling）结构，其中 Focus 结构包含 4 次切片操作和 1 次 32 个卷积核的卷积操作，可最大程度地减少信息损失，从而进行下采样操作；CSPNet 利用跳层连接的思想，进行局部跨层融合获得更为丰富的特征图；SPP 模块先采用不同大小的核进行最大池化操作，再通过拼接融合多尺度的池化特征。Neck：包含 PANet[17]（Path Aggregation）结构，其是对 FPN[18]（Feature Pyramid Network）的改进，首先自顶向下将高层特征信息与不同层 CSP 模块的输出特征进行聚合，再通过自底向上路径聚合结构浅层特征，从而充分融合不同层的图像特征。Prediction：使用 3 个尺度的特征图分别预测与之对应的预定义锚框的偏移量，每个特征点最多预测 3 个行人，因此最后预测的总通道数为（1class+1object+4coordinates）*3anchors=18。采用 CIOU[19]（Complete-IoU）代替 Smooth L1 范数作为损失函数，即使预测框与真实框交并比为零时，仍然可以反向传播优化网络权值。分类损失直接使用 BCE（Binary Cross Entropy）损失，后处理采用 NMS（Non-maximum Suppression）对行人目标检测框进行筛选，只保留最接近真实矩形框的检测结果。

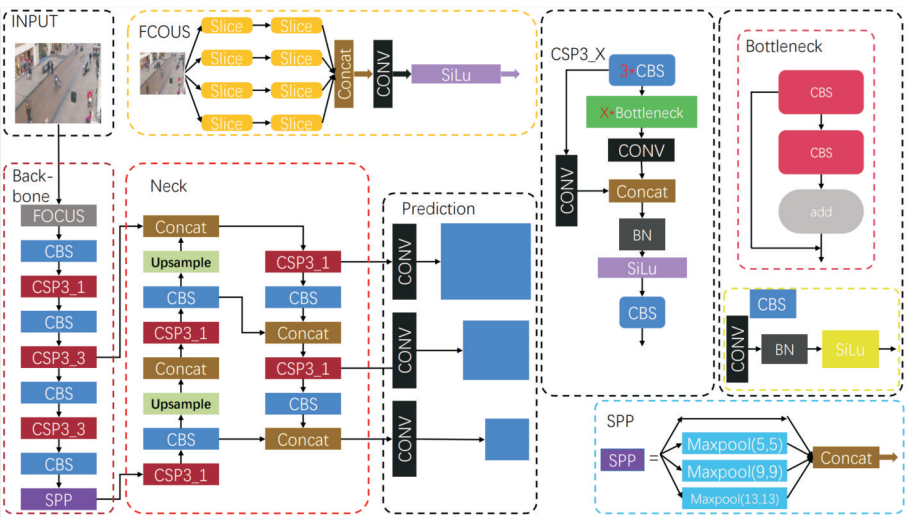

Fig. 1　YOLOv5s model structure

图 1　YOLOv5s 模型结构

3　单应性与尺度估计

相机的成像过程可以简化为小孔成像模型，式（1）为投影过程的数学模型，其中 s 为比例系数；$[u\ v\ 1]^T$ 为图像坐标的齐次坐标；K 为相机固有参数，称为相机内参矩阵；R、T 分别为世界坐标系相对于相机坐标系的旋转矩阵和位移向量；$[X\ Y\ Z\ 1]^T$ 为真实世界中三维坐标的齐次坐标。在投影过程中，空间中的任意三维点先经过 R、T 从世界坐标系转换到相机坐标系，再通过相机内参矩阵 K 投影到图像平面上。

$$s\begin{bmatrix}u\\v\\1\end{bmatrix}=K\begin{bmatrix}R & T\\0 & 1\end{bmatrix}\begin{bmatrix}X\\Y\\Z\\1\end{bmatrix} \tag{1}$$

假设地面是三维世界坐标系中的一个平面，世界坐标系的原点建立其上，则将地面上的三维点投影到图像平面上时，它们的坐标 Z 总是为 0。因此，上述投影方程可以简化，利用 Z 坐标始终为 0 的条件消去相机旋转矩阵的第三列，表示为：

$$s\begin{bmatrix}u\\v\\1\end{bmatrix}=K\begin{bmatrix}r_{11} & r_{12} & r_{13} & T_x\\r_{21} & r_{22} & r_{23} & T_y\\r_{31} & r_{32} & r_{33} & T_z\\0 & 0 & 0 & 1\end{bmatrix}\begin{bmatrix}X\\Y\\0\\1\end{bmatrix}=K\begin{bmatrix}r_{11} & r_{12} & T_x\\r_{21} & r_{22} & T_y\\r_{31} & r_{32} & T_z\end{bmatrix}\begin{bmatrix}X\\Y\\1\end{bmatrix} \tag{2}$$

通过式（2）可将三维空间中地面上的任一点 $P=[X\ Y\ 1]^T$ 投影到图像平面点 $p=[u\ v\ 1]^T$，并用 H 表示该映射矩阵，则 H 称为从地面到图像平面的单应性矩阵。

$$s\begin{bmatrix}u\\v\\1\end{bmatrix}=\begin{bmatrix}h_1 & h_2 & h_3\\h_4 & h_5 & h_6\\h_7 & h_8 & h_9\end{bmatrix}\begin{bmatrix}X\\Y\\1\end{bmatrix} \tag{3}$$

由于 K、R、T 内外参数未知，本文通过摄像头安装的通用场景进行合理假设来估计 H 矩阵。首先假设世界坐标系到相机坐标系的翻滚角为 0°（绕 Y 轴的旋转角为 0°），因为通过观察可以发现图像中的行人身体都是竖直的；然后假设摄像机的偏移角为 0°（绕 Z 轴的旋转角为 0°）。如图 2 所示，定义摄像机主轴在地平面上的交点处为世界坐标系原点，摄像机安装高度为 h，相机俯仰角为 θ（绕 X 轴的旋转角 $90°+\theta$），则可得到：

$$s\begin{bmatrix}u\\v\\1\end{bmatrix}=K\begin{bmatrix}1 & 0 & 0\\0 & -\cos\theta & -\dfrac{h}{\tan\theta}\\0 & -\sin\theta & h\end{bmatrix}\begin{bmatrix}X\\Y\\1\end{bmatrix} \tag{4}$$

基于以上假设，地面上平行于X轴的直线投影到图像平面上仍保持平行，而平行于Y轴的直线在图像平面上则交汇于灭点。因此，地面上宽高为W、H的矩形投影到图像平面后会变成宽高为W'、H'的等腰梯形。如式(5)所示，定义水平边的投影比例为ρ_v，竖直边的投影比例为ρ_h，这两个参数只与相机俯仰角θ和相机安装高度h相关。

$$\rho_h = \frac{W'}{W}, \rho_v = \frac{H'}{H} \tag{5}$$

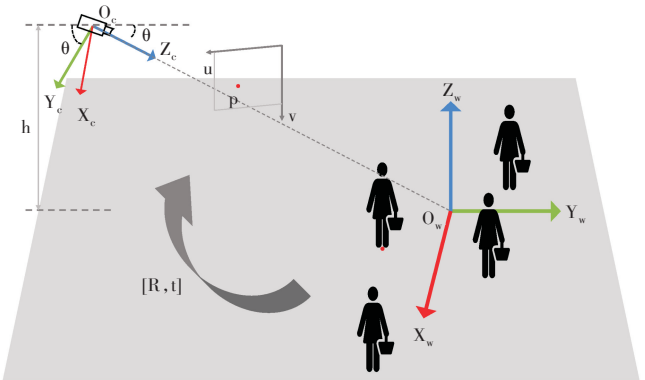

Fig. 2 Camera imaging model

图2 相机成像模型

给定$\rho_h, \rho_v \in [0,1]$，可以找到地平面矩形与图像平面等腰梯形之间相对应的4个角点，从而估计出单应矩阵H。这里H并不反映像素与真实长度之间的度量映射，因为K是未知的。这里估计的H可以将每个人周围的圆形安全区域经过透视变换投影在图像平面上，形成一个近似的椭圆。为了计算地面投影到图像平面上的尺度信息，本文选取人类平均身高。研究表明，人类平均身高为1.7m[20]，在不同种族之间只有很小差异。直接利用垂直于地面的人体平均身高和图像行人检测框像素高度计算地平面到图像平面的比例是不合理的，但是对于大多数监控摄像头安装场景来讲，安装高度大于行人高度很多，而且朝地面的倾角θ是接近45°的锐角，所以仍然可以利用人体身高近似地估计地面上行人局部区域投影到图像平面的尺度。设定行人安全距离为2m，则行人周围的圆形安全区域半径为1m。首先利用行人检测器得到图像中的行人检测框，通过检测框像素高度与人类平均身高的比例计算出行人对应地面圆形安全区域半径为1m的图像平面像素半径；然后根据估计出的H将圆形投影为椭圆，以行人检测框的底边中点像素坐标为中心，在图像上绘制椭圆形行人安全区域(见图3)；最后通过判断图像中所有行人椭圆形安全区域是否重叠检测其安全社交距离的违反情况。

4 实验方法与结果分析

实验运行环境：操作系统为Ubuntu18.04，开发语言为Python，深度学习框架为PyTorch，CPU为Intel Xeon w-2123，GPU为GeForce GTX 1080Ti，内存为16G。

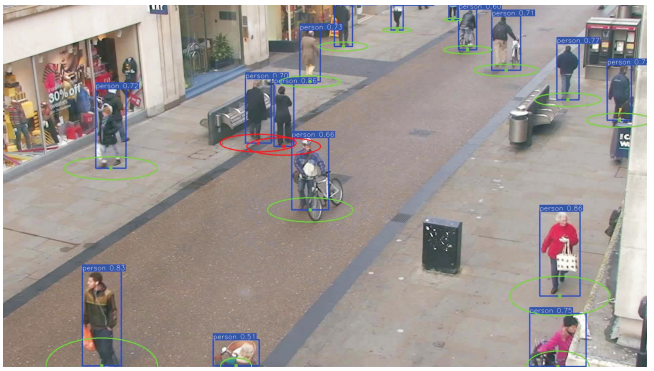

Fig. 3 Pedestrian safety social distance violation detection results

图3 行人安全社交距离违反检测结果示意图

Yolov5s行人检测模型训练数据使用MSCOCO中只包含行人的数据，其中训练数据图像共58 783张，验证数据图像2 693张。使用在MSCOCO数据集上多目标检测训练过的模型对网络进行初始化，优化算法使用随机梯度下降算法。主要训练参数：Batch大小为32，最大迭代次数为100，初始学习率设置为0.01，并采用OneCycleLR[21]策略周期性调整学习率。

分别使用YOLOv5s-Mul-Cls和YOLOv5s-Person模型在MSCOCO验证数据集上对行人检测效果进行评估。其中YOLOv5s-Mul-Cls是对MSCOCO数据集进行训练的80类原始多目标检测器，YOLOv5s-Person是只利用MSCOCO数据集中行人数据训练的单目标检测器。准确率、召回率和AP值的计算是在交并比IoU=0.5的条件下得到的，mAP0.5∶0.9是指IoU从0.50~0.95间隔0.05取值对应AP的平均值。从表1中可以看出，本文训练的单目标行人检测器具有更高的召回率和准确率，对场景中的行人检测数量更多，识别更准确。

Tabel 1 Pedestrian detector evaluation results in MSCOCO verification set

表1 行人检测器在MSCOCO只含行人数据验证集上的评价结果

(%)

行人检测器	准确率	召回率	AP@0.5	mAP@0.5:0.95
YOLOv5s-Mul-Cls	79.2	69.6	77.1	48.4
YOLOv5s-Person	81.1	72.4	80.7	53.9

在OTC数据集上[22]进行行人安全社交距离违反检测的验证与行人检测性能的评价。OTC数据集是一段包含7 498帧的视频，每帧平均包含16个行人，其中行人是否违反安全社交距离的真实标签是利用数据集提供的相机内外参数计算得到的。行人检测评价指标使用IoU=0.5时检测的准确率、召回率和AP，其中姿态关键点检测算法预测的行人检测框是由姿态关键点的竖直外接矩形计算得到。行人安全社交距离检测评价指标采用二分类的准确率(Precision)、召回率(Recall)和F1值，表示为：

$$\text{Precision} = \frac{TP}{TP + FP}, \text{Recall} = \frac{TP}{TP + FP}, F1 - score = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{6}$$

式中，*TP* 表示算法检测出违反安全社交距离的行人数量，*FP* 表示未违反安全社交距离但是被检测为违反安全社交距离的行人数量，*FN* 表示违反了安全社交距离行人的漏检数量。为与改进前 Aghaei 等[9]提出的基于姿态关键点估计的方法进行比较，设置 $\rho_v = 0.5$，$\rho_h = 0.8$，保持与改进前算法设置相同。具体评价结果如表 2 所示，其中 OpenPose_BBX_VSD[9]是指利用 OpenPose 算法[23]检测的关键点竖直外接矩形高度估计尺度信息的检测方法；OpenPose_Torso_VSD[9]是指利用 OpenPose 算法得到躯干关节点间的长度以估计尺度信息的检测方法；YOLOv5s_VSD 是本文利用 YOLOv5s 行人检测算法得到行人检测框的高度进而估计尺度信息的检测方法。

Tabel 2　Evaluation results of pedestrian detection and safe social distance violation detection on OTC

表 2　OTC 数据集上行人检测及安全社交距离违反检测评价结果

单位：%

算法	行人检测			行人安全社交距离违反检测		
	准确率	召回率	AP@0.5	准确率	召回率	F1-score
OpenPose_BBX_VSD[9]	35.61	32.65	–	63.18	90.12	72.38
OpenPose_Torso_VSD[9]	35.61	32.65	–	82.98	82.30	81.04
YOLOv5s_VSD	81.39	82.39	76.52	98.99	89.12	93.79

从表 2 可以看出，本文训练的 YOLOv5s 行人检测模型性能具有很大提升，与 OpenPose 算法相比，其行人检测准确率和召回率分别提升 45.8% 和 49.7%。从图 4 可视化对比结果来看，由于 OpenPose 姿态关键点检测不能很好地克服行人遮挡、尺度变化、姿态变化等问题，无法作出稳定检测。如图 4(a)所示，OpenPose 没有检测出小尺度的半身行人，对于骑行姿态的行人关键点检测存在误差，导致得到的矩形框与真实标签的 IoU 小于 0.5，无法判断为正样本。相比图 4(b)用 YOLOv5s 对同一图像区域的检测结果来看，YOLOv5s 能更好地解决上述问题，对小尺度半身行人、骑行姿态的行人均作出了正确检测。

关于行人安全社交距离违反检测结果，YOLOv5s_VSD 的准确率、召回率和 F1 值分别比 OpenPose_Torso_VSD 提升了 16.01%、6.82%、12.75%，说明利用本文训练的 YOLOv5s 模型得到的行人检测框高度比躯干高度计算尺度信息更具有鲁棒性，降低了由于人体关节局部运动产生的尺度估计误差。此外，YOLOv5s_VSD 和 OpenPose_BBX_VSD 均利用检测框高度估计尺度信息，但从对比结果来看，YOLOv5s_VSD 性能具有更大提升，原因是 OpenPose 算法专注于整张图中人体姿态关键点的回归检测[23]，在对每个行人的关键点进行预测时可能会错误连接其他近邻行人关键点，或漏检头部、脚部等关键点(如图 4(c)所示)，导致由此计算的行人检测框高度及安全区域中心点可能存在误差。相比之下，使用 YOLOv5s 模型直接对行人整体语义特征进行矩形框检测(如图 4(d)所示)，能够预测出更稳定的接近真实标签的行人检测框，有利于图像上行人安全区域的估计。

(a)　(b)　(c)　(d)

Fig. 4　Detection results comparison of OpenPose (a,c) and YOLOv5s (b,d) on OTC

图 4　OpenPose(a,c)和 YOLOv5s(b,d)在 OTC 数据集上的检测结果比较

5　结语

本文结合行人检测与单应性变换估计对监控场景作出一般性假设，提出一种用于未校准摄像头下行人安全社交距离违反检测的方法。行人社交距离检测任务的前提是行人检测漏检率要低、准确率要高，且要保证在现代 GPU 算力下具有良好的实时性。为此，本文提供了一个基于 YOLOv5s 的轻量型行人检测模型，可对监控视频实现实时高效的行人检测，从而为行人社交安全区域的估计提供稳定、准确的行人位置和检测框高度信息。本文并没有对行人社交距离进行直接计算，而是将任务简化为图像中行人局部安全区域的重叠判断，比直接估计全局单应性矩阵与尺度信息，再通过计算行人之间的距离判断是否违反安全社交距离的方法更具有准确性。此外，本文还通过对监控场景中摄像机的安装情况作出合理假设，简化了从场景地面到图像平面单应性矩阵估计的任务，保证了一般监控场景下行人安全社交距离违反检测的良好性能。然而，本文提出的方法仍然受遮挡问题影响而导致图像中行人安全区域估计不够准确，后续将结合监控视频的帧间信息对行人遮挡问题进行改进，从而优化尺度估计，使其适用于更多场景。

参考文献：

[1]　PEARCE K. What is social distancing and how can it slow the spread of COVID-19[EB/OL]. https://hub.jhu.edu/2020/03/13/what-is-social-distancing/.

[2]　CHU D, DUDA S, SOLO K, et al. Physical distancing, face masks, and eye protection to prevent person-to-person transmission of SARS-

CoV-2 and COVID-19: a systematic review and meta-analysis[J]. The Lancet, 2020, 395(10242):1973-1987.

[3] CRISTANI M, BUE A D, MURINO V. The visual social distancing problem[J]. IEEE Access, 2020, 8: 126876-126886.

[4] NGUYEN C, SAPUTRA Y, VANHUYNH N, et al. A comprehensive survey of enabling and emerging technologies for social distancing part I: fundamentals and enabling technologies[J]. IEEE Access, 2020, 8: 153479-153507.

[5] YAO B, WEN C L, LIN Z P. A pedestrian social distance monitoring based on YOLOv4[P]. China,202010879084.6. 2020-8-27.
姚博,文成林,林志鹏.一种基于YOLOv4的行人社交距离实时监测[P].中国,202010879084.6. 2020-8-27.

[6] REZAEI M, AZARMI M. DeepSOCIAL: social distancing monitoring and infection risk assessment in COVID-19 pandemic [DB/OL]. https://arxiv.org/abs/2008.11672? context=eess.

[7] SHORFUZZAMAN M, HOSSAIN M, ALHAMID M. Towards the sustainable development of smart cities through mass video surveillance: a response to the COVID-19 pandemic[J]. Sustainable Cities and Society, 2021, 64:102582.

[8] GUPTA S, KAPIL R, KANAHASABAI G, et al. SD-Measure: a social distancing detector [C]//Bhimtal: International Conference on Computational Intelligence and Communication Networks, 2020.

[9] AGHAEI M, BUSTREO M, WANG Y, et al. Single image human proxemics estimation for visual social distancing[C]//Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2021: 2785-2795.

[10] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]// Columbus: 2014 IEEE Conference on Computer Vision and Pattern Recognition, 2014.

[11] REDMON J,DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection[C]//IEEE Conference on Computer Vision and Pattern Recognition, 2016: 779-788.

[12] REN S Q, HE K M, 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,2017,39(6):1137-1149.

[13] JOCHER G, STOKEN A, BOROVEC J, et al. Ultralytics YOLOv5: v5.0-YOLOv5-P6 1280 models, AWS, supervisely and YouTube integrations[EB/OL]. https://doi.org/10.5281/zenodo.4679653. 10.5281/zenodo.4679653.

[14] LIU W,ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector[C]//European Conference on Computer Vision, 2016: 21-37.

[15] LIN T, MAIRE M,BELONGIE S,et al.Microsoft COCO: common objects in context [C]//European Conference on Computer Vision, 2014:740-755.

[16] BOCHKOVSKIY A,WANG C Y, LIAO H. YOLOv4: optimal speed and accuracy of object detection [DB/OL]. https://arxiv. org/abs/2004.10934.

[17] TRINH H C, LE D H, KWON Y K, et al. PANET: a GPU-based tool for fast parallel analysis of robustness dynamics and feed-forward/feedback loop structures in large-scale biological networks[J]. PLoS One, 2014, 9(7): e103010.

[18] KIRILLOV A, GIRSHICK R, HE K.Panoptic feature pyramid networks[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 6392-6401.

[19] ZHENG Z, WANG P, LIU W, et al. Distance-IoU loss: faster and better learning for bounding box regression[C]//AAAI Conference on Artificial Intelligence, 2020:12993-13000.

[20] BATEN J, BLUM M. Human height since 1820[M]. UK: Organisation for Economic Cooperation and Development, 2014: 117-137.

[21] SMITH L, TOPIN N. Super-convergence: very fast training of residual networks using large learning rates[EB/OL]. https://openreview.net/forum? id=H1A5ztj3b.

[22] BENFOLD B, REID I. Stable multi-target tracking in real-time surveillance video[C]//Conference on Computer Vision and Pattern Recognition, 2011: 3457-3464.

[23] CAO Z, HIDALGO G, SIMON T, et al. OpenPose: realtime multi-person 2d pose estimation using part affinity fields[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 43(1): 172-186.

(责任编辑:尹晨茹)