

# 监控视频中行人安全距离评估

罗 昊

(中国刑事警察学院, 辽宁 沈阳 110035)

**摘要:**在后疫情阶段, 针对疫情零星爆发的地区, 防疫工作依旧紧迫。为了配合这些地区的治安防疫措施快速有效地开展, 该文融合了行人识别深度学习算法、欧式距离测距、安全距离分类评估的想法, 提出一种行人检测模型。对不同的行人运动状态、不同视角视频中行人距离差别进行评估, 并通过行人之间不同的距离对识别出的行人进行分类标定, 能有效减少治安工作人员的劳动付出, 保障了防疫措施有效开展。

**关键词:**安全距离评估; 监控视频; 行人; 疫情

**中图分类号:**TP311

**文献标识码:**A

**文章编号:**2096-9759(2022)04-0004-03

## 0 引言

视频监控已经广泛应用于公安日常过程中, 但是公安人员的主要办案手段依旧是靠人工对视频中的特定目标进行识别与搜索。长时间观看监控会造成工作人员视觉疲劳, 导致对关键目标的误判、漏判, 而智能计算机技术能有效降低监控视频识别的误判率。人物识别是计算机视觉中十分重要的一类问题, 通过行人识别再进一步对行人目标进行处理, 有助于快速了解行人的运动状态, 为日常治安工作稳定开展提供帮助。在新冠疫情背景下, 通过行人运动状态到群体的运动状态的识别, 能有效规范行人的安全距离, 简化维护秩序人员的工作, 当前图像测距的方式有单目<sup>[1]</sup>和双目<sup>[2]</sup>测距两种, 多用于道路汽车自动驾驶中测量车辆距离和预估障碍物距离<sup>[3-5]</sup>。

在当今新冠疫情大背景下, 人们外出格外注意安全距离, 但是人的注意力并不能一直集中于保持安全距离。例如, 在聚集场所(商场, 车站进出口等), 行人之间不能长时间地保持安全距离。尽管这些场所通常在地上标注安全线来规范行人之间的安全距离, 但是行人的数量远高于安全线数量。没有安全线规范的行人失去了直观的规范标准, 此时若能够快速检测出行人间距离的变化, 对非安全距离进行标注, 就能有效减少防疫人员的工作量, 达到科学防疫的效果。从上述需求出发, 本文提出的行人安全距离评估模型对聚集场所视频中行人距离进行评估。首先, 使用基于 YOLOv3 算法<sup>[6]</sup>的行人检测模块对视频中的行人进行检测, 并用滑动窗口标定。然后, 选取滑动窗口中心点作为目标点, 依据滑动窗口相关参数计算行人目标点间的欧式距离, 将行人间距离计算问题转换为目标点间的距离计算问题, 得到不同的三组距离。最后, 对所得的目标点距离进行分类评估, 利用不同颜色的滑动窗口对三种距离评估结果进行区分。

## 1 行人识别

在卷积神经网络的参与下, 行人识别技术识别准确度和识别速度获得了空前提升, 当下, 由 Fast-RCNN 为主的卷积神经网络识别技术已经在各大领域取得一定的运用成果。在广大学者的不断推进下, 卷积神经网络算法逐渐简化、深化, 催生了残差网络、批量标准化等方法, 让卷积神经网络的越来越小, 精度越来越高, 目标识别技术的速度也越来越快, 其中较为突出的就是 YOLO 系列算法。一直追求极速的 YOLO 系列算法, 在逐次迭代的过程中结合了残差网络的思想, 产生了 YO-

LOv3 网络算法, 为单一设备进行目标识别提供了良好的基础保障。

在新冠疫情背景下, 最短时间检测出行人并对其进行相关指标的评估能有效保障防疫工作的开展, 达到在一定程度上遏制病毒传播的效果。在我们的行人检测模块中, YOLOv3 算法在训练数据集的选用时, 使用了单一分类实例最多的 MS COCO<sub>17</sub>数据集。在行人实例方面, 数据集包含 250000 个进行标注的人物目标, 这样得到的训练网络参数对不同场景和不同视角下行人检测具有很强的鲁棒性。基于 Darknet53 这一精简的网络骨架, YOLOv3 算法对单帧图像处理速度达到了 22ms 的优异成果, 满足了聚集场所下实时检测的要求。

行人识别网络模型如图 1 所示, 模型建立在 Darknet53 网络框架上, 包含 53 个卷积层。YOLOv3 行人检测算法是通过 53 个卷积层在  $13 \times 13$ 、 $26 \times 26$  和  $52 \times 52$  三个尺度范围内进行特征采样, 算法框架由 5 个降采样层, 卷积层提取视频帧  $V(X, Y, k)$  (其中  $X, Y, T$  分别代表视频帧的长、宽及序列) 中的行人特征。之后, 通过两个上采样层和 3 个全连接层, 将三个尺度的特征依次进行融合, 得到三个尺度的行人预测滑动窗口和其输出维度:  $\sum S \times S \times (C, x_i, y_i, w_i, h_i, p_i)$ , 其中:  $S$  取值为 13、26 和 52,  $C$  指 MS COCO 数据集中的 80 个类别的概率,  $x_i, y_i, w_i, h_i, p_i$  分别是指当前预测的第  $i$  个滑动窗口的坐标、宽、高、置信度。在损失函数方面, YOLOv3 网络保留了位置部分的平方差损失计算的方法, 在置信度和类别预测方面采用了交叉熵损失及算法。在标签分类的问题上, YOLOv3 使用逻辑分类器中的 sigmoid 函数, 将输出约束在 0 到 1 之间, 大大提高了网络的检测性能。最后, 使用置信度和 NMS<sup>[8]</sup> (Non Maximum Suppression) 联合的方式进行最终的筛选, 确定行人检测滑动窗口, 得到行人检测标注结果, 以此对监控视频序列  $V(X, Y, T)$  中目标进行独立检测<sup>[9]</sup>。

## 2 行人安全距离评估

现阶段大多数的图像测距算法都依赖着前期的场景标定或者世界坐标系、相机坐标系和图像坐标系三者的转换来计算距离, 这种场景标定模型和一一对应的坐标转换模型针对固定场景具有一定的效果。在新冠疫情背景下, 聚集场所的行人距离会相对疏松, 人群遮挡情况有所缓解, 这就为行人距离评估提供了有利条件。为了提高算法的泛化性, 本文在视频序列中建立二维图像坐标系, 并在此基础上进行行人目标处理。

收稿日期: 2022-02-15

作者简介: 罗昊(1996-), 男, 安徽省安庆人, 硕士研究生。研究方向: 声像资料检验技术。

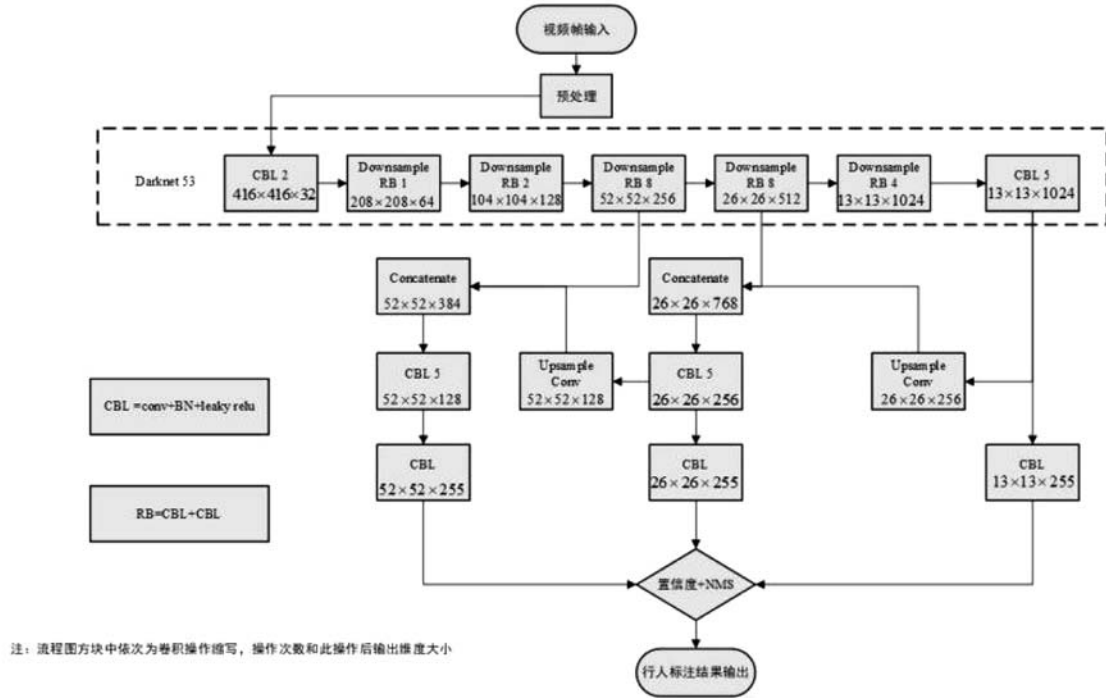


图 1 网络结构图

实现行人距离评估, 首先需要将行人转化为图像坐标系中的目标点进行测距。行人作为像素集合  $Q(q_j), j \in X \times Y$  含于视频  $V(X, Y, T)$  之中, 而在像素集合中又很难取得具有代表性的目标点  $q_i(x, y)$ 。虽然通过肢体框架标定的神经网络算法可以在一定程度上标定出头部、四肢目标点, 但是在不同场景中, 行人的运动具有不确定性, 视频不能一直记录所有行人的完整身体框架, 这就导致肢体框架标定的目标点具有一定的偏差。而通过滑动窗口标定行人的方式只关注于视频帧  $V(X, Y, k)$  中行人像素集合  $Q(q_j)$  的边界, 因此能够更好实现行人与背景的分隔效果和行人与行人的区分。本文中, 为了能够统一对行人进行距离计算, 我们将滑动窗口的中心点  $c_i(x, y)$  作为行人的目标点, 如图 2 所示。



图 2 关键点确定计算

本文使用了欧式距离对视频帧中两关键点间的距离  $d_i$  进行测算, 如式 (1) 所示。为了得到合理的距离评价指标, 我们选取行人坐标的横纵坐标之和作为标尺, 由此判断人与人之间的距离是否安全。本文将安全距离分为三个等级, 分别为安全距离、低风险距离 (如式 (2) 所示), 高风险距离 (如式 (3) 所示)。式 (2) (3) 中比例参数是根据现实聚集场所地面标线测得。

$$d_i = \sqrt{(x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2} \quad (1)$$

$$0 \leq d_i \leq 0.1 \times (x_i + y_i) \quad (2)$$

$$0 \leq d_i \leq 0.01 \times (x_i + y_i) \quad (3)$$

### 3 实验

#### 3.1 实验设置

本文的实验环境为 Windows 10 系统, 使用 pycharm 平台搭建 Anaconda3.0 环境, 在 Darknet、TensorFlow 深度学习框架下完成实验。导入的数据库包括 opencv、numpy、matplotlib、imutils、os、json 等。设置网络模型参数时, 加载 YOLOv3 的权值文件、模型配置文件和 MS COCO 数据集类别标签文件。将目标检测的置信度阈值设为 0.5, 在视频左下角建立文本块, 对三种距离分类的人数进行统计, 通过本文算法对处于安全距离、低风险和高风险距离的行人进行区分, 并用计数的方法直接在输出视频的左下角进行分类统计, 增加直观效果。

#### 3.2 数据集介绍

我们的实验数据采用 PETS2009 公用数据库和自建数据集机场安检口、商店门口等三个场景。其中 PETS2009 数据集包含了相同场景下不同角度的人群活动 (约 20 人), 本次实验采用了 Crowd\_PETS09/S1/L3/Time\_14-33/View\_001 这一场景。在视频场景中, 行人向视频中心汇聚, 相互距离也随着各自运动不断改变。在机场安检口, 这是一个必然需要排队的场景, 人们在有标志线的地方有序前进。商场门口的队列由于没有全面的疏导, 又受制于场地大小, 队列中的行人会相对无序。

#### 3.3 实验展示

我们首先将待检测视频输入行人安全距离评估模型, 通过行人检测得到滑框, 如图 3 所示。然后通过几何转换得到滑框中心点, 利用欧式距离计算行人间距离。最后依据我们选定的距离标尺对距离进行评估, 得到相对距离分类, 标为三

个等级,最后通过分等级计数的方式直观展示。



图3 YOLOv3 算法行人识别展示

在使用 PETS2009 公用数据库得出的实验结果图 4 中我们可以看出,视频中被识别出的行人分别被区分标定,在视频左下角进行直观计数,有效地区分出了实验设定中的三种行人距离。从视频截图中我们可以看出,因为视频像素影响,少部分人不能很好地被模型识别,当然也有地面标志物被识别为行人的情况,但是总体不影响整体人数安全距离的检测。



图4 PETS2009 数据集检测结果

从机场场景的实验结果如图 5,可以看出,视频中的行人主要是成侧身或者背对拍摄设备站立的姿势,行人安全距离评估模型能有效的检测评估视频中行人距离。虽然人群在有地面警示线时能很好地保持安全距离,但是熟人同行等情况让行人无法维持安全距离,此时通过行人安全距离检测模型的标注,防疫工作者能快速有效地发现行人聚集现象,从而进一步展开应对措施。



图5 机场安检口场景检测结果

从商店门口行人队列场景视频评估结果如图 6,我们可以看出,此场景处于闹市街道,除排队行人外,过路行人的加入让场景中人员构成变得复杂,且队列并没有指挥人员,也没有受到标志物的制约,大家相对松散,缺乏有效的规范标准。在这一复杂场景,行人安全距离评估模型很好地地将视频帧中的行人进行检测评估,按照不同的距离进行标注,有效简化了防疫工作者的日常工作。



图6 商场门口队列场景检测结果

在最后的算法识别率统计中,依据不同场景,我们统计了视频序列关键帧(在视频序列中每间隔 10 帧选取一帧)中安全距离明显判断失误的行人和识别准确的行人,并根据识别率=判断失误行人/总人数<sup>[10]</sup>对最后的算法结果进行统计,如表 1 所示。

表 1 模型检测结果统计

场景类别	识别率(%)	关键帧数
PETS2009 数据集	87	47
机场安检口场景	90	52
商店门口队列场景	82	108

#### 4 结语

本文建立的行人安全距离评估模型,使用 YOLOv3 网络算法对视频中目标行人进行识别,定位,跟踪,得到不同行人滑框,再利用行人滑框的中点测算出人与人之间的距离,判断其是否达到安全距离,并将视频帧中的行人标注为安全距离、低风险距离和高风险距离三类,有效地在人群聚集场所规范行人间距,简化治安工作人员的工作。整体实验结果较为理想,符合当下治安工作人员多变的工作环境,有利于简化聚集场所的治安工作,为防疫工作的有效开展提供助力的工作环境。不过,本文的行人检测算法还存在误检的情况,这也是我们后期需要解决的问题。

#### 参考文献:

- [1] 罗静云.一种基于人眼定位的单目测距方法[D].深圳大学,2015.
- [2] 王浩,许志闻,谢坤,等.基于 OpenCV 的双目测距系统[J].吉林大学学报:信息科学版,2014,(2):188-194.
- [3] 王士明.基于计算机视觉的车辆单目测距系统研究[D].天津大学,2012.
- [4] 罗时光.基于单目测距技术的道路障碍物检测方法[J].软件,2013,(9):4.
- [5] 罗思杰,邹复民,郭峰,等.一种基于单目测距的相邻车辆加速度检测方法及其装置:,CN111103437A[P].2020.
- [6] Redmon J, Farhadi A. YOLOv3: An Incremental Improvement[M]. arXiv e-prints, 2018.
- [7] Lin T Y, Maire M, Belongie S, et al. Microsoft COCO: Common Objects in Context[M]. Springer International Publishing, 2014.
- [8] Sun C, Vallotton P. Fast linear feature detection using multiple directional non-maximum suppression [J]. Journal of Microscopy, 2009, (2): 147-157.
- [9] 岳晓新,贾君霞,陈喜东,李广安.改进 YOLO V3 的道路小目标检测[J].计算机工程与应用, 2020, (21): 224-229.
- [10] 刘翔.监控视频中融合携带物判断的步态识别研究[J].广东公安科技, 2021, (03): 52-54.