

# 基于多目标跟踪算法的车站人流聚集监测研究

马建民<sup>†</sup>, 朱磊<sup>2</sup>, 骆友曾<sup>1</sup>

(1. 中铁第一勘察设计院集团有限公司, 陕西 西安 710043; 2. 皖赣铁路安徽有限责任公司, 安徽 合肥 230011)

**摘 要:** 车站运营过程中实时获取人流信息, 对聚集行为与重点区域进行监测、预警、分析在车站安全保障与管理中具有重要意义。目前车站的监控硬件设备数量不断增加、配备复杂, 单独依靠监控系统的预警功能实现车站实时、全方位、多角度的人流密度与聚集分析存在困难。为此, 提出了一种基于 YOLOv5+DeepSort 并结合 HRNet 模型的方法, 实现了人员速度、方向、密度、异常聚集、异常侵入、隔栏递物和逃票检测。在车站实时监控数据集和收集的相关互联网数据集上进行了实验, 实验结果显示, 提出的算法能够实现检测任务, 具有实时高效分析监控视频与实际应用的能力。在此基础上开发了一套车站视频行为分析系统后端服务与前端管理平台, 对车站旅客出行保障与安全运营具有重要意义, 有发展前景。

**关键词:** 多目标跟踪; 车站监测; YOLOv5; DeepSort

**中图分类号:** TP39

**文献标识码:** A

## Research on the Crowd Gathering Monitoring of Station Based on Multi-Object Tracking Algorithm

MA Jianmin<sup>1†</sup>, ZHU Lei<sup>2</sup>, LUO Youzen<sup>1</sup>

(1. China Railway First Survey and Design Institute Group Co., Ltd., Xi'an, Shaanxi 710043, China;

2. Wan&gan Railway Anhui Co., Ltd., Hefei, Anhui 230011, China)

**Abstract:** During the operation of the station, it is of great significance to obtain real-time passenger flow information, monitor, warn and analyze the gathering behavior and key areas in the safety assurance and management of the station. At present, the number of monitoring hardware equipment in the station is increasing and the equipment is complex. It is difficult to achieve real-time, all-round and multi angle passenger flow density and aggregation analysis of the station solely relying on the early warning function of the monitoring system. To solve this problem, this paper proposes a method based on the combination of YOLOv5+DeepSort and HRNet model, which realizes the detection of personnel speed, direction, density, anomaly aggregation, anomaly intrusion, barrier delivery and ticket evasion detection. We have carried out experiments on the real-time monitoring data set of the station and the relevant internet data set collected. The experimental results show that the algorithm proposed in this paper can achieve the detection task, and has the ability to analyze the monitoring video and practical applications in a real-time and efficient manner. On this basis, we have developed a set of back-end service and front-end management platform of the station video behavior analysis system, which is of great significance and development prospects for passenger travel security and safe operation of the station.

**Key words:** multi-object tracking; station monitoring; YOLOv5; DeepSort

多目标跟踪是目标感知领域的重要内容, 其广泛应用于自动驾驶、机器人导航、机器视觉和智能

视频监控等领域<sup>[1-5]</sup>。按照任务划分,多目标跟踪任务可以分为目标检测和目标跟踪两个部分<sup>[6]</sup>。在目标检测方面,近年来基于深度学习的目标检测算法形成了基于候选区域和基于回归两大类<sup>[7]</sup>。

以候选区域为基础的目标检测算法又被称作二阶段方法,把目标检测问题划分为以下两大阶段。

首先是候选区域的生成,其次是在分类器中放入候选区域,并实施针对性的分类,而且要做好位置修正。

该类检测算法可以提供更好的检测精度,其能充分提取图像的某些细节特征。然而,由于该类算法分为两个步骤,容易使得模型的运行速度变慢。典型的两阶段目标检测算法有 R-CNN 算法<sup>[8]</sup>、SPP-Net R-CNN 算法<sup>[9]</sup>、Fast R-CNN 算法<sup>[10]</sup>、Faster R-CNN 算法<sup>[11]</sup>等。

对以回归为基础的目标检测算法而言,它仅存在一个阶段,也就是直接回归相关的预测目标物体<sup>[7]</sup>。该类算法的模型构建更为简单,计算速度快于第一类算法,其检测实时性更强,更利于工程实现。典型的基于回归的检测算法包括 YOLO 系列算法<sup>[12]</sup>、SDD 算法<sup>[13]</sup>、Retina Net<sup>[14]</sup>等算法。

HRNet(High-Resolution Network)是一种用于人体姿态估计的深度学习模型。HRNet 的设计目标是在保持高分辨率特征的同时,提供多尺度的信息。传统的卷积神经网络(CNN)通常会在网络中的不同层级上进行下采样,导致特征图的分辨率逐渐降低。而 HRNet 通过构建并行的多分支网络来解决这个问题,每个分支都负责处理不同分辨率的特征图,并且通过高分辨率特征的融合来生成最终的预测结果。

在目标跟踪方面,有基于检测跟踪(tracking by detection, TBD)的算法和基于初始框跟踪(detection tree tracking, DFT)的算法<sup>[15]</sup>。前者是在实现跟踪前对给定图像或视频内的目标进行检测得到信息,再在跟踪过程中将后续的信息链接到已有轨迹中;后者则是通过人工标定第一帧的目标,边检测边跟踪,需要人为干预<sup>[16]</sup>。由于人工标定对目标的完整性要求较高,目标标定的不完整容易导致跟踪结果不稳定,甚至产生错误,因此,本文研究 TBD 跟踪方法中比较经典的 DeepSort 算法。

综上所述,考虑到模型整体的性能、实时性和封装的程度,提出了一种基于 YOLOv5+DeepSort 并结合 HRNet 模型的方法实现各个检测任务。

## 1 人员速度、方向和异常侵入检测

本节主要介绍了人员速度检测、人员方向检测和人员异常侵入检测的算法实现。算法整体框架采用了 YOLOv5 目标检测技术,搭配 DeepSort 目标追踪技术对多目标进行检测和追踪,并利用识别的位置坐标的标注框进行计算。

### 1.1 人员速度检测

由于摄像头的高度和角度问题,直接计算人员在画面中移动的速度不可行,要开展针对性的透视变换(perspective transformation)。透视变换是一种在计算机视觉和图形学中常用的技术,用于将一个二维平面上的图像转换为另一个视角或投影平面上的图像。它可以改变图像的尺度、旋转角度和倾斜程度,从而实现图像的透视效果。透视变换意味着,采取目标点、像点、透视中心三点共线的必要条件,依据透视旋转定律,让承影面(透视面)围绕着透视轴(绩线)旋转某个适当的角度,将原本存在的投影光线束予以破坏,这个过程中依旧可以保持承影面上的投影几何图形的原来面貌,即借助投影矩阵,把平面投影到本文指定的相关平面上。通过这一变换就能够获得较为真实的人员速度。

采用透视变化进行人员检测的步骤为如下。首先使用摄像机或其他传感器来捕捉场景中的视频数据。确保摄像机的位置和角度适合进行透视变换。其次,使用透视变换技术校正图像,以消除由于相机角度和位置造成的透视畸变。通过选择适当的关键点或线段,在图像中标记出固定位置的物体或场景元素。再次,使用目标检测和跟踪算法来检测和跟踪场景中的人员。然后,通过跟踪算法得到的目标轨迹,计算出目标在图像上的运动轨迹。最后,通过分析目标在不同帧之间的位置变化,计算出目标的速度。可以使用目标在图像上的像素坐标和时间间隔来估计目标的位移,并进一步计算出速度,达到实现图像的校正、视角变换、投影映射、图像纠正和图像修复等目的。

然而,传统的透视变换会导致画面边缘受标定点的影响很大,往往很小的误差就会对画面边缘人员的速度计算产生巨大的影响。由于直接按照画面速度计算的方法在画面边缘处不会产生巨大的误差,因此我们先利用前后帧坐标框中心点移动的距离来计算人员在画面中的速度和人员经过透视变换后的速度,然后进行加权平均得到最终的速度。限制经过透视变换后的速度上限,以减少

误差。

此外,为了减少噪声的影响,利用 10 帧的速度进行平均,每 10 帧更新一次标注的画面。当速度大于跑动阈值,且大于画面平均速度的 2 倍时标注为“run 运行”;当速度小于阈值,且小于画面平均速度的 0.1 倍时标注为“stop 停止”。

综上,人员速度、方向和异常侵入检测流程如图 1 所示。视频画面中的移动速度先经过用前后帧坐标框中心点移动的距离来计算人员在画面中的速度和人员经过改进的透视变换后的速度;其次,两者进行加权平均得到最终的速度;最后,求取 10 帧的平均速度并将此作为阈值基数,当最终画面速度小于平均速度的 0.1 倍时,判定为“stop”,当最终画面速度大于平均速度的 2 倍时,判定为“run”,否则判定为“walk”。

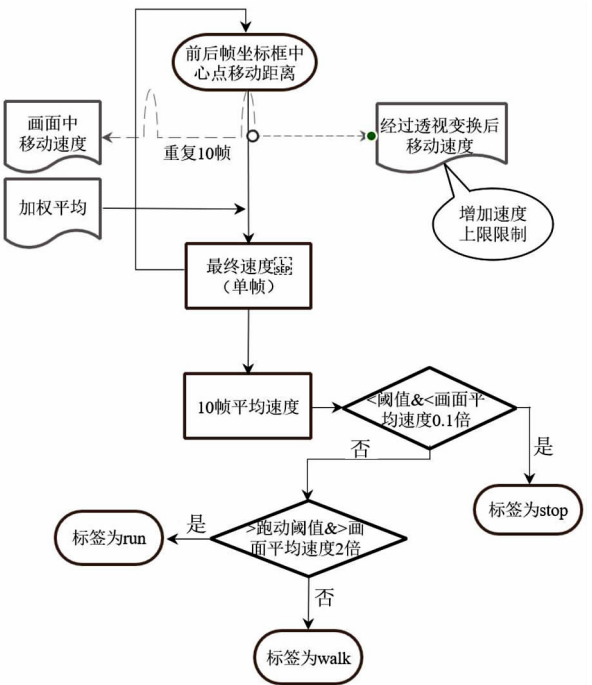


图 1 人员速度、方向和异常侵入检测流程

1.2 人员方向检测

对于人员方向的检测,我们不关心人员的具体方向,只关心人员是流入还是流出,因此,可以用前后帧坐标框中心点计算。

同理,为了减少噪声的影响,每 10 帧检测一次,如果人员向外走标注为“out”,如果人员向内走标注为“in”,并且统计标注为“out”和标注为“in”的人员个数。

1.3 人员异常侵入检测

图 2 是人员方向检测流程图。首先,将模型分为“person”“black”“worker”(即一般旅客、穿黑色

衣服旅客和车站工作人员,因车站工作人员制服也是黑色,所以需要单独标注训练穿黑色衣服旅客以示区分)三类进行训练;其次,用射线检测法计算射线与画面中交叉点的个数,判断是否为“worker”,若是则不预警,反之预警。

判断是否异常侵入需要排除工作人员,利用先前训练好的模型识别画面中的工作人员,由于工作人员都是黑色衣服,为了更好地区分出工作人员,模型设置了“black”标签来识别身着黑衣服的游客,从而能够加大游客与工作人员的区分,除黑衣服及工作人员以外的游客设置为“person”。因此,模型的标签分为三类,分别为“person”“black”“worker”。

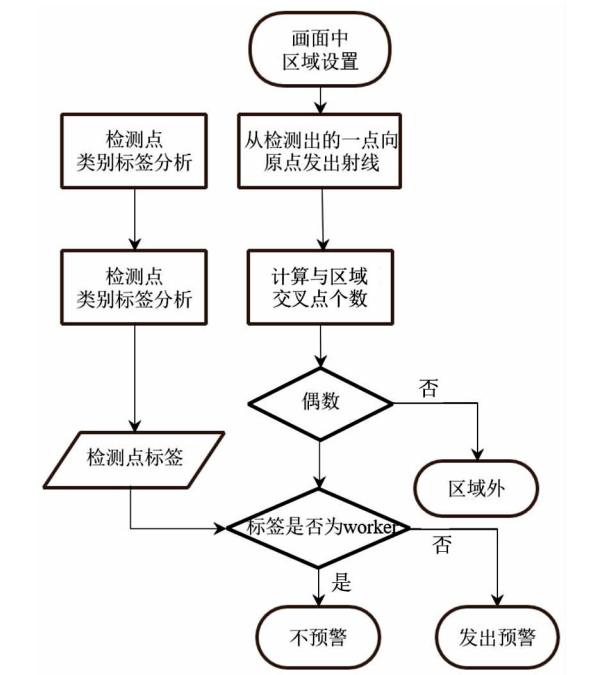


图 2 人员方向检测流程

人员异常侵入需要绘制不允许侵入的区域,利用坐标框底部中心点,采用射线检测的方法判断人员是否进入到区域内。射线检测的主要思想是从一点往外发射出一条射线,如果经过区域的边数为偶数,则该点在区域外,奇数则在区域内。若在区域内且标注不为“worker”,则标注为“Enter”。

此外,考虑到会有部分工作人员未被识别,进入区域后被错误标注为“Enter”,而未被识别的工作人员大概率被标注为“black”。若标注为“black”的人员连续 7200 帧(允许有 20 帧以内的间断)均在区域内,则认为是先前识别失败的工作人员,标注为“worker”,此时如果发生间断识别,不会重新计数标注。

1.4 实验分析

本节采用自建铁路车站环境数据集来验证本文所提方法对人员速度、方向和异常侵入检测的有效性。高铁车站内安装的摄像头数量众多,为保证数据采集最大限度地对全部重点监控场景区域进行全覆盖,在采集视频图像的过程中,要结合旅客进站、到站、出站、乘车等活动的差异化属性,在大厅、站台、候车室、安检区域、进站口等多种区域开展均衡化的数据采样工作。借助视频图像采集、预处理、清晰操作等,其中的 2000 张图像成了训练样本,从理论上来讲,应确保诸多重点监控区域能够获得大致相当的场景图像样本数量。

采用 Labelme 标注工具对图像进行标注工作,数据表述格式使用 YOLO 模式。结合自我解析图像内容的要求,标注全部样本中的六类目标,涵盖工作人员、旅客、列车、出站口、座椅、检票区等多个目标区域。而且,每张图像样本要标注 10 条以上位置信息,全部训练样本中要标注成功 2 万的信息。进行了样本标注的相关信息要以 Json 文件的形式保存下来,保存的主要信息内容涵盖对应的标签以及目标的位置。完成了训练样本标注后,要和样本图像文件名相匹配,可以当作输入数据传输到网络模型中,并开展针对性的训练。

实验过程中网络模型运行在一台配有 4 块 NVIDIA 3090 显卡的 Ubuntu Server 20.04 服务器上,其内存达到了 256G,采取 Intel Xeon 10 核的处理器和 16T 的硬盘存储。模型运行中采用 Python 3.8.13 的编译环境,使用 pytorch1.10.2+cu113 深度学习框架。并使用 Flask web 框架将文中人员运动方向与速度监测、入侵监测能力进行了统一封装。

在测试集上进行的实时监测实验,其检测结果如图 3 所示,图中绿色框线是“person”,蓝色框线是“worker”。从图 3 可以看出,本文提出方法可以准确检测游客与工作人员速度、方向及异常入侵。

2 人员密度和异常聚集检测

本节主要介绍人员密度检测、人员异常聚集检测的算法实现。由于人员处于高密区域时,会发生大量的遮挡,利用 YOLOv5+DeepSort 的方法效果很差,所以将对身体躯干的识别转化为对头部点的识别。利用 HRNet 对人体的头部点进行识别,

并将识别得到的热力图进行算法设计。



图 3 人员运动速度及方向检测结果可视化

2.1 人员密度检测

图 4 是人员密度检测流程图,需要从 HRNet 生成的热力图中提取人头点的位置,利用大小为  $3 \times 3$  的最大池化层筛选那些经过池化和原来值还一样的像素点,就是局部最大的点。然后设定一个阈值将值较低的点排除在外,阈值设定为  $100/255$  倍的最大值,如果这个最大值小于 0.1,则认为这个图中没有人。

当阈值小于 0.1 时,先排除阈值较低点,再在得到人头坐标后,利用人员异常侵入检测中的方法判断人员头部是否进入区域内。若进入,程序会统计识别的人数和区域内人数,从而计算区域内人员的密度。

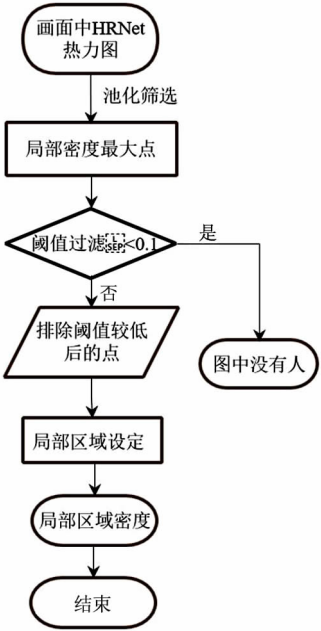


图 4 人员密度检测流程

## 2.2 人员异常聚集检测

图 5 是人员异常聚集检测流程图。首先,利用人员密度检测中的方法获得人员头部坐标,再利用 KDtree 算法获得与其最近的 3 个邻居的坐标和距离,通过其三个邻居的距离值相加可以得到密度参数以定性地衡量其密度大小,并且为了防止部分稀疏点的密度参数过大,需要根据画面大小将其限制在一定范围以内。若密度参数低于平均值的一定程度,则认为该点处于高密区域。

其次,人为设定每个区域内高密点的个数,利用 K-means 算法自动标注高密区域(最少 3 类),并且去除掉只有少数高密点形成的区域。

最后,由于摄像头角度和实际情况的影响,部分地区可能一直处于高密状态,此时不应该算作异常聚集,而是一种正常状态,所以需要引入历史区域数据。如果新产生的高密区域和历史数据中的高密区域的多个数据距离过近,则消除这部分区域的判定,从而检测异常聚集人群。

在该方法中需要人为设定以下 3 个检测参数:

1)midua 设置高密区域的距离判定时所用的参数,控制密度识别灵敏度,越大灵敏度越低,越不容易被识别为高密。

2)midub 把区域内个数大于 midua 的区域加入历史区域,控制历史密度区域识别灵敏度,越大灵敏度越低,越不容易被选为历史密度区域。

3)rulla 设置取消判定区域时所用的参数,控制历史区域的距离参数,越大灵敏度越低,越不容易被取消判定。

## 2.3 实验分析

本节实验中的一部分数据来自我国西部某铁路沿线车站站台以及站内的相关视频监控数据,一些视频源自网络视频中火车站台上的相关旅客的视频数据。视频数据的分辨率大致达到了  $1920 \times 1080$  像素。所选取的一些样本视频中涵盖车站值班员和旅客在站台、大厅、站内行走的相关行为片段;随机选取的另外一些样本来自视频图像样本,对视频进行了大量截图,从中选取 1000 张图像成为所需要的数据集。将其中的 750 张当作训练集,借助 YOLO 数据集的相关标注方式同时开展样本分类标注工作;将剩余的 250 张当作验证集。这个实验模型是以 YOLOv5 为基础的一种预训练模型,开展针对性的样本元学习,图像输入尺寸采取了  $512 \times 215$  像素。设定训练的整个过程涵盖三个子训练任务,所以应该随机抽取确定以上的查询集与支撑集三次,以便适用于各个差异化的子任务。

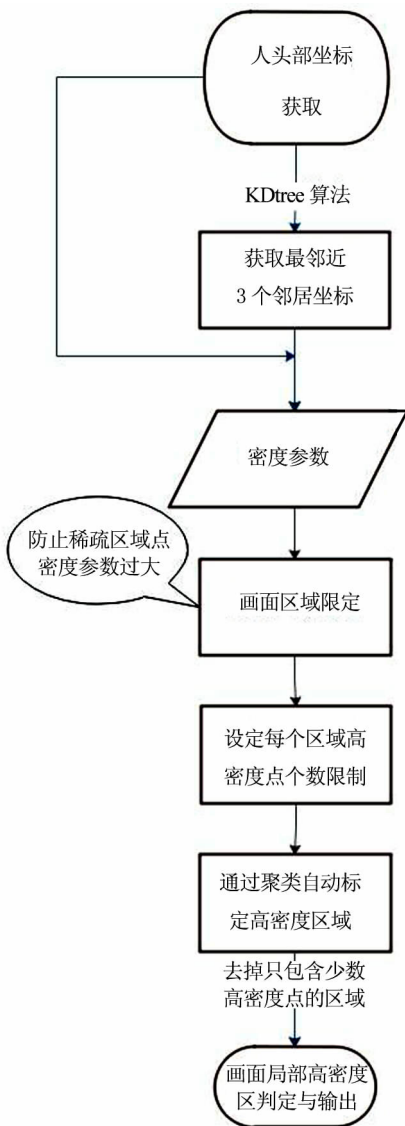


图 5 人员异常聚集检测流程

本节实验电脑配置和 Python 软件条件与 2.4 节相同。在测试集上进行的实时监测实验的检测结果显示如图 6 所示。图中绿色框线是实验检测出的人头部的的位置坐标框,红色框线是检测出的疑似局部聚集的位置区域。从图 6 可以看出,本文提出方法可以实现人员密度和异常聚集检测。

## 3 人员隔栏递物和逃票检测

本节主要介绍了人员隔栏递物检测和人员逃票检测的算法实现。算法仍处于完善阶段,整体框架上采用了 YOLOv5 目标检测技术,搭配 Deep-Sort 目标追踪技术对多目标进行检测和追踪利用识别得到的位置(位置矩形框左上右下的点的)坐标  $(x_1, y_1, x_2, y_2)$  标注框进行算法设计。



图6 人流密度与聚集检测结果可视化

3.1 人员隔栏递物检测

首先,利用异常侵入检测的方法判断人员是否进入栏杆区域。其次,利用两人底部中点坐标的连线是否与栏杆中线有交点,来判断两人是否位于栏杆的两侧,如果位于栏杆两侧则判断为隔栏递物。

3.2 人员逃票检测

首先,利用异常侵入检测的方法判断人员是否进入检票区域。其次,对人员进行计数,若连续一段时间均在区域内,则认为人员没有逃票,而如果人员提前退出区域,则认为人员逃票。

4 实验结果和分析

文中对单摄像头监控区域的人流数量和密度进行了监测能力的封装,结合车站监控管理平台的总体运营管理能力,设计并实验了实时监测车站总体与分区域(一个区域有多个摄像头覆盖的情况)密度估算的方法,并将功能集成到了开发的视频监控管理平台下。图7是区域摄像头配置列表,将视频监控管理平台分为多个子界面,每个界面都呈现相应的管理模式。



图7 区域摄像头配置列表

图8是区域内摄像头重叠率配置界面。对于区域内的密度评估,采用带重叠率的计算方法,即同一区域有多个摄像头覆盖时,重叠区域不重复计算人数。此重叠率由车站运营管理人员根据实际情况估算填写,虽然估算有些误差,但总体的密度估算精度在可以接受的范围之内。

面车站总体的密度估计,则使用各区域的密度值进行加权平均(以各区域的面积作为权重)计算所得。此外,使用互联网搜集的开放数据集对文中提到的视频中人员异常行为监测能力进行了8次集成测试,分别测试其客流逆行、隔栏递物与区域入侵行为。测试结果如表1,由表1可知,从开放数据集的测试结果看,无论是客流逆行、隔栏递物,还是区域入侵行为,本文提出方法都能检测准确。

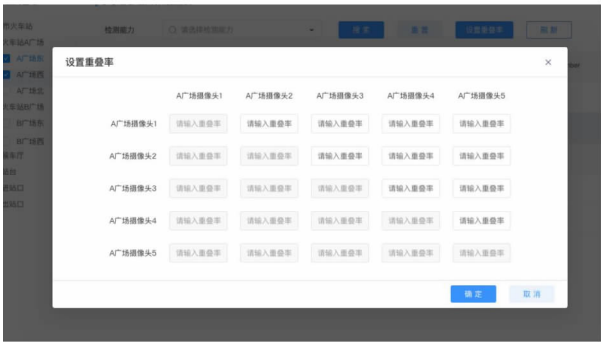


图8 区域内摄像头重叠率配置界面

5 结 论

针对车站单独依靠监控系统预警功能实现车站实时、全方位、多角度的人流密度与聚集分析存在困难的问题,提出了一种 YOLOv5 + DeepSort 和 HRNet 模型结合的方法,实现了人员速度、方向、密度、异常聚集、异常侵入、隔栏递物和逃票检测。本研究基于人工智能视觉分析技术的车站视频中人员异常行为监测,在此基础上开发了一套车站视频行为分析系统后端服务与前端管理平台,并在车站实时监控数据集和收集的相关互联网数据集上进行了实验。实测实验结果证实了研究成果的可用性与适用性。对于铁路车站内及周边的多种场景下的视频中人的异常行为可以进行实时的监测分析,当监测结果出现了异常行为时,可以快速定位摄像头 id 和关联的监控范围,并通过前端监控管理平台进行及时预警,可以提高应急响应时间,提高客户服务质量。

系统试点作为铁路客运安全出行的一个成功

实践案例,为保障铁路车站社会公共安全、提升车站运输经营管理能力提供了有力的支撑。若在实际应用中进行推广应用,将产生较好的社会效益和经济效益。

表 1 测试结果

| 序号 | 测试用例                 | 目标值                                        | 实际值              | 结果   |
|----|----------------------|--------------------------------------------|------------------|------|
| 1  | 客流逆行目录内视频文件 1. mp4   | 应识别以总动能差超过阈值 30 焦耳的逆行事件,发生时间 02 分 26 秒     | 标定时间点检测出逆行       | 检测准确 |
| 2  | 客流逆行目录内视频文件 1. mp4   | 不应识别以总动能差超过阈值 20 焦耳的逆行事件,发生时间 03 分 12 秒    | 标定时间点未误检逆行       | 检测准确 |
| 3  | 隔栏递物行为目录内视频文件 1. mp4 | 应识别出隔栏递物行为,发生时间 00 分 03 秒                  | 标定时间检测出隔栏递物      | 检测准确 |
| 4  | 隔栏递物行为目录内视频文件 1. mp4 | 应识别出隔栏递物行为,发生时间 00 分 13 秒                  | 标定时间检测出隔栏递物      | 检测准确 |
| 5  | 隔栏递物行为目录内视频文件 1. mp4 | 不应识别出隔栏递物行为,发生时间 00 分 01 秒                 | 标定时间点及前后未误检为隔栏递物 | 检测准确 |
| 6  | 区域入侵行为目录内视频文件 1. mp4 | 应识别出入侵行为,发生时间 00 分 04 秒并进行跟踪(左侧地面画一个多边形范围) | 划定区域内检测到入侵行为并跟踪  | 检测准确 |
| 7  | 区域入侵行为目录内视频文件 2. mp4 | 应识别出入侵行为,发生时间 00 分 04 秒并进行跟踪(大街中央画一个多边形范围) | 划定区域内检测到入侵行为并跟踪  | 检测准确 |
| 8  | 区域入侵行为目录内视频文件 2. mp4 | 不应识别出入侵行为(垃圾桶旁边的小片检测区域)                    | 划定区域未误检区域入侵      | 检测准确 |

参考文献

[1] NGUYEN H V, EZATOFIGHI H R, VO B N, et al. Distributed multi-object tracking under limited field of view sensors[J]. IEEE Transactions on Signal Processing, 2021, 69: 5329—5344.

[2] DAI Peng, WANG Xue, ZHANG Weihang, et al. Instance segmentation enabled hybrid data association and discriminative hashing for online multi-object tracking [J]. IEEE Transactions on Multimedia, 2019, 21(7): 1709—1723.

[3] CHEN Y, HUANG J, LIU H, et al. Appearance guidance attention for multi-object tracking[J]. IEEE Access, 2021, 9: 103184—103193.

[4] GAO Y, XU H, ZHENG Y, et al. An object point set inductive tracker for multi-object tracking and segmentation [J]. IEEE Transactions on Image Processing, 2022, 31: 6083—6096.

[5] 朱守刚. 智能视频监控技术研究及应用[J]. 数字技术与应用, 2012(1): 26—28.

[6] 罗志升, 王黎, 高晓蓉, 等. 序列图像中运动目标检测与跟踪方法分析[J]. 现代电子技术, 2009, 32(11): 4.

[7] 罗会兰, 陈鸿坤. 基于深度学习的目标检测研究综述[J]. 电子学报, 2020, 48(6): 1230—1239 .

[8] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic

segmentation[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2014, 81: 580—587.

[9] HE K, ZHANG X, REN S, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1904—1916.

[10] GIRSHICK R. Fast R-CNN[C]. IEEE International Conference on Computer Vision, 2015: 1440— 1448.

[11] REN S, HE K, GIRSHICH R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. IEEE Transaction on Pattern Analysis & Machine Intelligence, 2017, 39(6): 1137—1149.

[12] 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.

[13] LIU W, AAGUELOV D, ERHANR D, et al. SSD: single shot multibox detector[C]. European Conference on Computer VIsion, 2016: 21—37.

[14] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[C]. IEEE International Conference on Computer Vision, 2017: 2980—2988.

[15] 张铎, 肖坦. 视频分析技术在青藏铁路视频监控中的应用[J]. 中国铁路, 2010, 2010(3): 68—69.

[16] 王嘉琳. 基于 YOLOv5 和 DeepSORT 的多目标跟踪算法研究与应用[D]. 济南: 山东大学, 2021.