

孙文, 赵柞旭, 谭凌番, 等. 基于行人重识别的目标跟踪的系统设计与实现[J]. 智能计算机与应用, 2026, 16(2): 90-94.  
DOI: 10. 20169/j. issn. 2095-2163. 24042902

# 基于行人重识别的目标跟踪的系统设计与实现

孙文, 赵柞旭, 谭凌番, 王远, 张剑书

(南京工程学院 计算机工程学院, 南京 211167)

**摘要:** 本文设计了一套基于行人重识别的行人检测系统, 主要采用 YOLO 算法检测并跟踪行人目标。系统利用三维重建技术将二维轨迹转换成三维形式, 通过 Web 界面实现实时监控和轨迹展示。该系统在行人检测和跟踪方面表现出良好的性能, 适用于人流量大、环境复杂的场景。系统通过检测行人特征并进行三维映射, 能够直观地展示场景内行人运动轨迹。通过资源监控保证系统正常运行, 同时可以保存历史数据进行分析。该系统具有实时性、便捷性和高效性, 在行人检测和跟踪领域具有广阔的应用前景。

**关键词:** 行人检测; 计算机视觉; 目标跟踪; 行人重识别; 卷积神经网络

**中图分类号:** TP391.4 **文献标志码:** A **文章编号:** 2095-2163(2026)02-0090-05

## Design and implementation of target tracking system based on pedestrian re-identification

SUN Wen, ZHAO Zuoxu, TAN Lingfan, WANG Yuan, ZHANG Jianshu

(School of Computer Engineering, Nanjing Institute of Technology, Nanjing 211167, China)

**Abstract:** This paper designs a pedestrian detection system based on pedestrian re-identification, which mainly uses YOLO algorithm to detect and track pedestrian targets. The system uses 3D reconstruction technology to convert 2D trajectory into 3D form, and realizes real-time monitoring and trajectory display through Web interface. The system shows good performance in pedestrian detection and tracking, and is suitable for the scene with large pedestrian flow and complex environment. By detecting pedestrian features and mapping them in three dimensions, the system can intuitively display pedestrian trajectories in the scene. Resource monitoring ensures normal system running and saves historical data for analysis. The system has the advantages of real-time, convenience and high efficiency, and has broad application prospects in the field of pedestrian detection and tracking.

**Key words:** pedestrian detection; computer vision; target tracking; pedestrian re-identification; convolutional neural network

## 0 引言

基于行人重识别的行人检测和跟踪技术是通过识别画面中的行人数据, 提取所有的行人目标, 并将其标记出来, 使出现在画面中的所有行人在每一帧都保持一个稳定唯一不变的 ID<sup>[1]</sup>, 进而能够生成运动轨迹。外部摄像头采集到视频数据后, 经过视频预处理, 用 YOLO 的 CNN 网络对图片进行分割并将行人目标检测出来<sup>[2]</sup>, 再利用卷积层的矩阵卷积, 通过卷积核(也称为滤波器)与输入数据进行卷积操作来提取特征<sup>[3]</sup>, 使用 Mish 激活函数增加网络的表达能力<sup>[4]</sup>。同时在识别时可以得到目标在视频

中的坐标信息<sup>[5]</sup>, 再进行三维映射, 映射到 3D 模型中, 可以得到行人单位在 3D 模型中的运动轨迹。通过这样的方法, 可以快速得到指定环境中, 行人单位的移动数据, 以供后续进行数据分析, 或安全防控等。基于行人重识别的目标跟踪具有高效性、及时性、便捷性等特点, 使用方法灵活, 学校、医院等人流量大、环境复杂的场所都可以使用, 在数据分析、安全防控方面有着广阔的应用场景<sup>[6]</sup>。

## 1 算法框架及实现原理

### 1.1 目标识别

目标识别是实现行人跟踪与重识别算法的前

**基金项目:** 江苏省高等学校大学生创新创业训练计划(202311276086Y); 南京工程学院校级科研基金(QKJ202301)。

**作者简介:** 孙文(2003—), 男, 本科生, 主要研究方向: 目标检测; 赵柞旭(2002—), 男, 硕士研究生, 主要研究方向: 机器学习; 谭凌番(2003—), 男, 硕士研究生, 主要研究方向: 重识别技术; 王远(1980—), 男, 实验师, 主要研究方向: Web 开发技术。

**通信作者:** 张剑书(1992—), 男, 博士, 高级实验师, 主要研究方向: 视频图像处理。Email: Z790447010@outlook.com。

**收稿日期:** 2024-04-29

哈尔滨工业大学主办 ◆ 学术研究与应用

提,本文采用 YOLO 算法实现目标识别<sup>[7-8]</sup>。YOLO 的 CNN 网络将输入的视频的每一帧图片分割成  $S \times S$  网格<sup>[9]</sup>,并由每个单元格去检测那些中心点落在该格子内的目标,如图 1 所示,可以看到这个目标的中心落在右上角一个单元格里,那么该单元格负责预测这个人。



图 1 检测目标

Fig. 1 Detect objects

每个单元格会预测  $B$  个边界框 (Bounding Box) 以及边界框的置信度 (Confidence Score)。置信度包

含 2 个方面:边界框含有目标的可能性大小和边界框的准确度。将边界框含有目标的可能性大小记为  $\Pr(\text{object})$ ,当边界框是背景时(即不包含目标), $\Pr(\text{object}) = 0$ ;而当边界框包含目标时, $\Pr(\text{object}) = 1$ 。边界框的准确度可以用预测框与实际框的交并比 (Intersection over Union, IoU) 来表征,记为  $\text{IoU}$ 。因此可将置信度定义为  $\Pr(\text{object}) \times \text{IoU}$ 。边界框的大小与位置可以用 4 个值来表征:  $(x, y, w, h)$ , 其中  $(x, y)$  是边界框的中心坐标,  $w$  和  $h$  是边界框的宽与高。中心坐标的预测值  $(x, y)$  是相对于每个单元格左上角坐标点的偏移值,单位是相对于单元格大小的比例,而边界框的  $w$  和  $h$  预测值是相对于整个图片的宽与高的比例,从而上述 4 个元素的大小应该在  $[0, 1]$  范围。综上分析可知,每个边界框的预测值实际上包含 5 个元素:  $(x, y, w, h, c)$ , 其中  $c$  是置信度。此时就可以将一段视频中可能存在的目标识别出来,如图 2 所示。

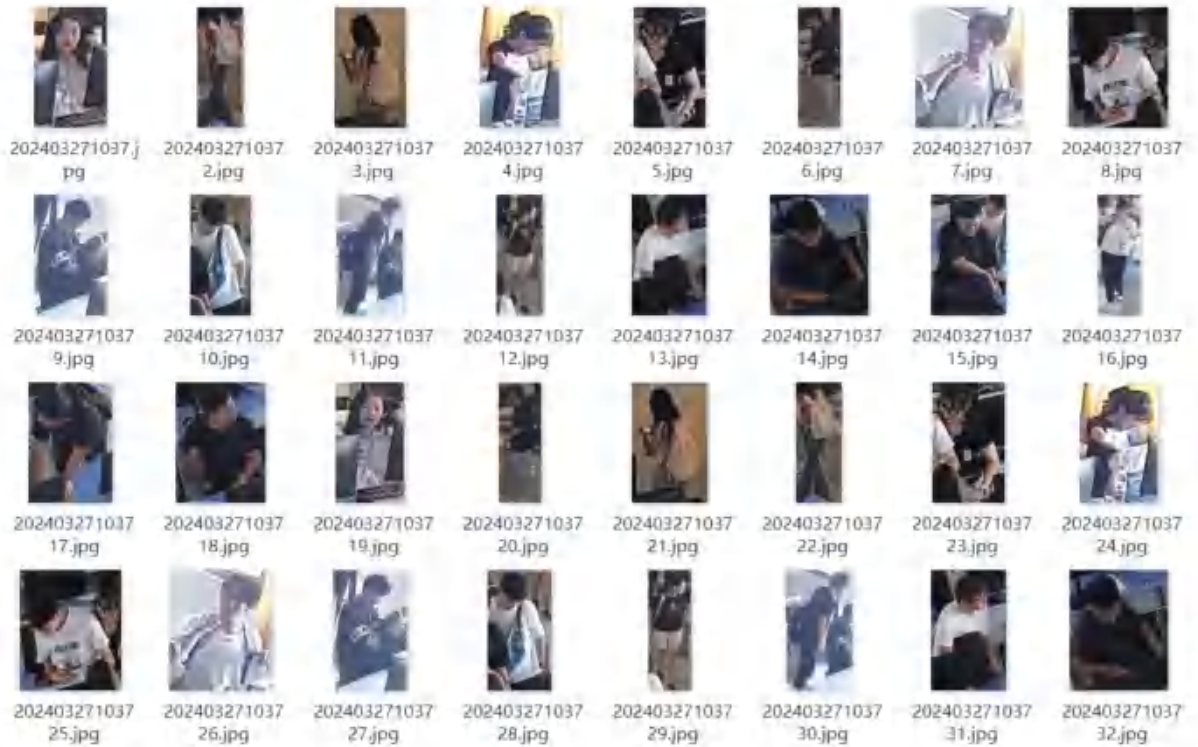


图 2 提取的目标

Fig. 2 Extracted target

### 1.2 行人重识别

在本系统中,基于特征提取和特征比较进行了行人重识别。

特征提取主要使用卷积神经网络 (CNN) 中卷积层的矩阵卷积<sup>[3,10]</sup>和激活函数。在卷积层中,通过卷积核(也称为滤波器)与输入数据进行卷积操作来提取特征<sup>[11]</sup>。

输入数据为一个三维矩阵(图像),用  $3 \times 3$  的卷积核进行特征提取<sup>[12]</sup>。卷积操作的数学表达式如下:

$$\text{Conv}(\mathbf{I}, \mathbf{K})_{i,j} = \sum_m \sum_n I_{i+m, j+n} \cdot K_{m,n} \quad (1)$$

其中,  $\mathbf{I}$  表示输入数据矩阵;  $\mathbf{K}$  表示卷积核矩阵;  $i, j$  表示输出矩阵中的位置;  $m, n$  表示卷积核的索

引;  $\text{Conv}(\mathbf{I}, \mathbf{K})_{i,j}$  表示输出矩阵中位置  $(i, j)$  的值。

激活函数用于在神经网络的每个神经元上引入非线性,以增加网络的表达能力。常用的激活函数包括 ReLU、Sigmoid 和 Tanh 等。本文使用的是 Mish 激活函数,于 2019 年由 Misra<sup>[13]</sup> 提出。该函数结合了线性函数和非线性函数,有助于提高模型的性能,其定义如下:

$$f(x) = x \times \tanh(\log(1 + e^x)) \quad (2)$$

通过特征提取,可以得到行人的特征向量。在特征空间中,行人图像之间的相似性可以通过计算对应的特征向量之间的距离来衡量,并以此来进行特征的比较。通过特征比较来寻找目标结果如图 3 所示。一般来说,距离越小表示 2 个行人图像越相似。本文通过计算马氏距离<sup>[14]</sup>,度量行人的相似性,马氏距离是一种考虑了数据的协方差矩阵的距离度量方法。对于给定的 2 个向量  $\mathbf{x}$  和  $\mathbf{y}$ , 其马氏距离计算方式为:

$$D(\mathbf{x}, \mathbf{y}) = \sqrt{(\mathbf{x} - \mathbf{y})^T \mathbf{S}^{-1} (\mathbf{x} - \mathbf{y})} \quad (3)$$

其中,  $\mathbf{S}$  表示样本集合的协方差矩阵,  $\mathbf{S}^{-1}$  表示其逆矩阵。

马氏距离不仅考虑了特征之间的距离,还考虑了特征之间的相关性<sup>[15]</sup>。这使其在数据具有相关性的情况下更加适用,同时马氏距离会根据数据的协方差矩阵对特征进行标准化,因此对于不同尺度的特征具有更好的鲁棒性。由于考虑了数据的协方差矩阵,马氏距离可以更好地适应数据的分布情况,对非球形的数据分布也具有较好的适应性。



图 3 通过特征比较寻找目标

Fig. 3 Find targets by comparing features

### 1.3 目标跟踪

行人目标跟踪是指在输入视频中,使出现在画面中的所有行人在每一帧都保持一个稳定唯一不变的 ID,识别和跟踪特定目标的位置和运动信息,并且能够描绘出所有行人在画面中行走的轨迹<sup>[16]</sup>。

本系统是基于行人重识别的跟踪<sup>[17]</sup>,所以文中设计通过存储某一时间节点下所有的行人目标特征

信息,和前一相邻时间节点下所有的行人目标特征信息通过本文所述的特征比较算法进行比较<sup>[18]</sup>,即完成目标匹配,如匹配成功,则继续沿用上时间节点的行人序列 id。其中包括目标之间坐标的距离测算、特征数据的提取和比较。

关于坐标计算,使用 YOLO 算法会输出一个或多个边界框<sup>[19]</sup>,每一个边界框都会有一个坐标值,通常表示为  $(x, y, \text{width}, \text{height})$ 。其中,  $(x, y)$  是边界框相对于图像左上角的偏移量, width 和 height 是边界框的宽度和高度。在此基础上,为了得到行人的中心点坐标,可以使用以下公式:

$$x_{\text{center}} = x + \frac{\text{width}}{2} \quad (4)$$

$$y_{\text{center}} = y + \frac{\text{height}}{2} \quad (5)$$

通过以上步骤,就可以从 YOLO 算法的输出中计算求得行人的坐标。

对于每个匹配成功的行人,研究需要更新其在视频中的轨迹信息,包括其历史位置和当前位置。通过对同一行人目标坐标的记录进行连线,就可以得到某一行人目标的运动轨迹。

## 2 目标跟踪系统的设计与实现

### 2.1 系统架构设计

系统基于 Flask+Python 搭建<sup>[20]</sup>,本系统的架构主要包括设备监控和信息查看。具体的系统架构如图 4 所示。

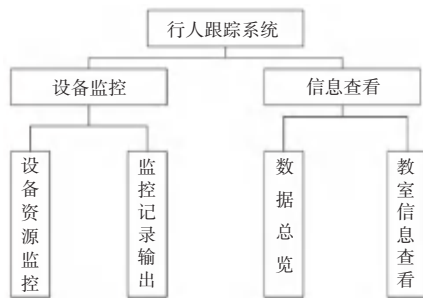


图 4 目标跟踪系统架构

Fig. 4 Architecture of target tracking system

其中,设备监控模块可以监控当前系统运行设备的资源占用情况,当资源占用过高的时候发出警报;信息查看模块可以查看设备运行情况,以及教室详细信息。

### 2.2 系统模块设计与实现

本系统的组成模块主要有设备监控模块和信息查看模块。用户可以在登录系统之后,通过进入相

应的模块使用对应的功能。

在信息查看模块中,可以查看当前加载成功的监控设备数量及运行情况,如图5所示。

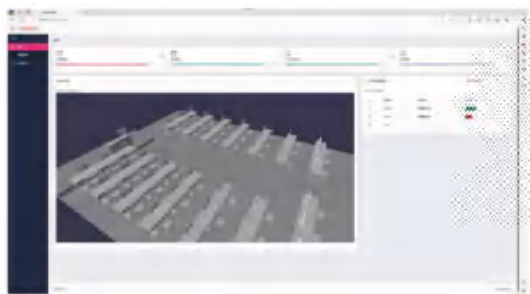


图5 系统主界面

Fig. 5 Main interface of the system

通过进入详情页面可以查看环境具体信息,可以实时观看监控视频,观察环境情况,系统将自动标记画面中的行人目标,并描绘其运动轨迹,同时将运动轨迹映射到三维场景<sup>[21]</sup>,如图6所示。



图6 详细界面

Fig. 6 Detailed interface

设备监控模块中,用户可以看到系统资源占用折线图,包括:CPU使用情况,显存使用情况等。通过对系统资源使用情况的监控,保证系统的正常运行,以防资源占用过高造成系统卡死导致的数据丢失。用户可以设置系统资源占用上限值,当占用情况超出阈值的时候,系统发出警报。资源监控界面如图7所示。

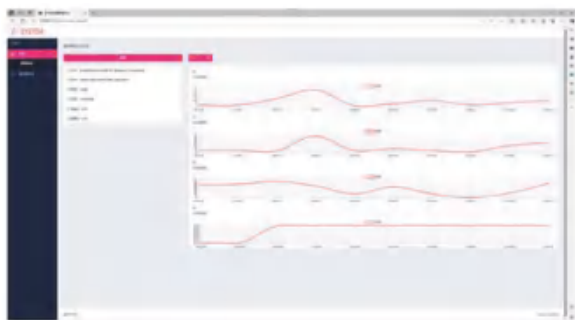


图7 资源监控界面

Fig. 7 Resource monitoring interface

## 2.3 系统部署

本系统的服务与图形用户界面的实现采用B/S (Browser/Service) 结构。在服务器中需要进行Python环境的配置,安装Flask、OpenCV、PyMysql等软件库及其插件。启动服务器后,在客户端的浏览器中输入服务地址即可访问系统。使用预先由超级管理员写入的账号即可登录。

## 3 结束语

本文介绍了基于行人重识别的行人检测和跟踪系统的设计与实现。系统使用YOLOv5算法进行行人检测,通过计算马氏距离进行行人特征提取和重识别。利用神经辐射场(NeRF)技术进行三维重建,将二维轨迹映射到三维空间。系统采用B/S架构,通过Web界面实现实时监控和轨迹展示。该系统在行人检测和跟踪方面表现出良好的性能,适用于人流量大、环境复杂的场景。

## 参考文献

- [1] 张嘉辉,赵威,王子琛,等. 基于检测和重识别的无人机行人跟踪算法[J]. 北京航空航天大学学报,2024,50(8):2538-2546.
- [2] 戴巨超,王洪元,曹亮,等. 一种多目标跨摄像头跟踪技术研究与应用[J]. 南京大学学报(自然科学),2021,57(2):227-236.
- [3] 李擎,胡伟阳,李江昀,等. 基于深度学习的行人重识别方法综述[J]. 工程科学学报,2022,44(5):920-932.
- [4] 王黎明,孙俊,陈祺东. 加强重识别的行人多目标跟踪算法[J]. 计算机工程与应用,2022,58(21):213-222.
- [5] 邹北骥,李伯洲,刘妹. 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.
- [6] 张广学. 基于深度学习的行人多目标跟踪算法研究[D]. 济南:山东大学,2024.
- [7] 周科宇,李军. 基于深度学习的目标检测研究进展[J]. 单片机与嵌入式系统应用,2023,23(7):38-40.
- [8] 史涛,刘祖林,朱文旭,等. 基于改进YOLOv5s的车辆行人检测[J]. 国外电子测量技术,2023,42(12):195-200.
- [9] 曹自强,赛斌,吕欣. 行人跟踪算法及应用综述[J]. 物理学报,2020,69(8):084203.
- [10] 项俊,张金城,江小平,等. Transformer-CNN特征跨注意力融合学习的行人重识别[J]. 计算机工程与应用,2024,60(16):94-104.
- [11] 王敬余. 基于深度学习的行人重识别技术研究[D]. 南京:南京邮电大学,2023.
- [12] 黄凯文. 基于卷积神经网络的行人检测与跟踪算法研究[D]. 淮南:安徽理工大学,2023.
- [13] MISRA D. Mish: A self regularized non-monotonic neural activation function[J]. arXiv preprint arXiv, 1908.08681,2019.
- [14] 姚足,龚勋,陈锐,等. 面向行人重识别的局部特征研究进展、挑战与展望[J]. 自动化学报,2021,47(12):2742-2760.
- [15] 王珂瑶,王惠文,赵青,等. 一种修正的马氏距离判别法[J].

- 北京航空航天大学学报, 2022, 48(5):824-830.
- [16]刘光明, 李媛. 基于深度学习的目标跟踪与行为识别技术[J]. 科技创新与应用, 2024, 14(12):164-167.
- [17]李雪松. 基于轨迹预测和行人重识别模型的多目标跟踪方法研究[D]. 北京:中国科学院大学, 2020.
- [18]冯星辰, 阮秋琦. 行人跟踪的多特征融合算法研究[J]. 信号处理, 2016, 32(11): 1308-1317.
- [19]REDMON J, DIVVALA S K, GIRSHICK B, et al. You only look once: Unified, real-time object detection[C]// Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ:IEEE, 2015:779-788.
- [20]陈嘉发, 黄宇靖. Flask 框架在数据可视化的应用[J]. 福建电脑, 2022, 38(12):44-48.
- [21]顾冷曦, 徐鹏飞, 张厚江, 等. 基于 NeRF 的树木数字化重建[J]. 苏州科技大学学报(工程技术版), 2024, 37(1):69-80.