

doi: 10.3969/j.issn.1672-6073.2025.01.014

# 基于骨架识别的城轨车站监控视频 乘客行为特征辨识研究

管洋<sup>1,2</sup>, 贾利民<sup>1</sup>, 陶思涵<sup>1</sup>, 豆飞<sup>3</sup>

(1. 北京交通大学交通运输学院, 北京 100044; 2. 北京全路通信信号研究设计院集团有限公司, 北京 100070; 3. 北京联合大学, 北京 100101)

**摘要:** 城市轨道交通领域传统监控分析方法对视频监控图像(如摔倒、晕倒和打斗等异常行为识别)漏识率高、参数调整复杂,且难以高效地应用于现实城轨车站监控场景,针对此问题,采用基于骨架模式识别的人体姿态特征辨识框架,引入基于人体骨架的姿态估计技术,采用 Alpha Pose 模型对乘客姿态进行精确估计,并结合时空图卷积网络(spatial temporal graph convolutional networks, ST-GCN)模型的方法,实现对城轨车站监控场景中异常行为的辨识。在 COCO 数据集和 MPII 数据集上分别达到了 72.3 mAP 和 82.1 mAP 的效果,相比较于 OpenPose 模型提升高达 17%,验证了模型的有效性和实用性。结果表明,本文所提出的方法不仅提高了乘客行为的识别速度,同时具备对复杂场景的适应能力,为城轨安全监控提供一种新的技术方案。

**关键词:** 轨道交通; 骨架识别; 模式识别; 城轨车站安全; 乘客行为特征辨识; ST-GCN

中图分类号: TP391.41 文献标志码: A 文章编号: 1672-6073(2025)01-0106-06

## Passenger Behavior Feature Identification in Urban Rail Station Surveillance Videos Using Skeleton Recognition Techniques

GUAN Yang<sup>1,2</sup>, JIA Limin<sup>1</sup>, TAO Sihan<sup>1</sup>, DOU Fei<sup>3</sup>

(1. School of Traffic and Transportation, Beijing Jiaotong University, Beijing 100044; 2. CRSC Research &amp; Design Institute Group Co., Ltd., Beijing 100070; 3. Beijing Union University, Beijing 100101)

**Abstract:** In order to solve the problem that the traditional monitoring analysts in the field of urban rail transit have a high false-negative rates and complex parameter adjustment of abnormal behaviors such as falling, fainting and fighting, making them difficult to apply efficiently to actual urban rail station monitoring scenarios, this paper proposes a human posture feature recognition framework based on skeleton pattern recognition, introducing the attitude estimation technology based on human skeleton. The Alpha Pose model is used to accurately estimate the posture of passengers, and combined with the Spatial Temporal Graph Convolutional Networks model, it achieves abnormal behavior recognition in the monitoring scenario of urban rail stations. By achieving 72.3 mAP on the COCO dataset and 82.1 mAP on the MPII dataset, the performance is improved by up to 17% compared to the OpenPose model, verifying the effectiveness and practicality of the model. The results show that the method proposed in this paper not only improves the recognition speed of passenger behavior but also has the ability to adapt to complex scenarios, providing a new technical solution for urban rail safety monitoring.

**Keywords:** rail transit; skeleton recognition; pattern recognition; urban rail station safety; passenger behavior identification; spatial

收稿日期: 2024-03-19 修回日期: 2024-08-13

第一作者: 管洋, 男, 博士研究生, 高级工程师, 从事轨道交通运营安全与管理方面研究, 20114083@bjtu.edu.cn

通信作者: 贾利民, 男, 教授, 从事轨道交通运营安全、交通智能系统等方面研究, lmjia@bjtu.edu.cn

基金项目: 北京市自然科学基金-丰台轨道交通前沿研究联合基金资助(L221006)

引用格式: 管洋, 贾利民, 陶思涵, 等. 基于骨架识别的城轨车站监控视频乘客行为特征辨识研究[J]. 都市轨道交通, 2025, 38(1): 106-111.

GUAN Yang, JIA Limin, TAO Sihan, et al. Passenger behavior feature identification in urban rail station surveillance videos using skeleton recognition techniques[J]. Urban rapid rail transit, 2025, 38(1): 106-111.

temporal graph convolutional networks (ST-GCN)

## 1 研究背景

随着城市轨道交通的快速发展, 确保乘客安全已成为公共安全管理的重要部分。当前, 视频监控系统作为安全管理的关键技术, 在实时监控和异常行为识别方面发挥着核心作用<sup>[1-2]</sup>。尽管如此, 目标检测与异常行为识别在复杂场景下仍面临挑战, 特别是在高峰时段和多种遮挡条件下的效果不佳, 这主要是由于传统方法高度依赖于视频质量和监控角度<sup>[3]</sup>。传统方法在视频监控图像异常行为识别应用中效果不佳, 存在漏识率较高, 且识别用时较长等问题<sup>[4]</sup>。

目标检测技术是视频监控系统中的基础, 它涉及在图像中自动识别和定位人体等目标物。尽管传统算法如方向梯度直方图(histogram of oriented gradient, HOG)和基于部件的可变型模型(deformable part model, DPM)在目标检测方面取得了进展, 但它们在遮挡和动态背景下的应用效果有限<sup>[5]</sup>。近年来, 深度学习的兴起带来了新的解决方案, 如基于区域的卷积神经网络算法(regions with CNN features, R-CNN)<sup>[6]</sup>和YOLO算法(you only look once)<sup>[7]</sup>, 它们通过学习大量数据提高检测的准确性和鲁棒性<sup>[5]</sup>。这些算法已广泛应用于不同场景的目标检测任务中, 但在城市轨道交通复杂环境中的表现仍有待提升<sup>[1]</sup>。

在异常行为识别领域, 现有技术多依赖于复杂的前处理和后处理技术。虽然基于深度学习的方法在一些标准数据集上表现良好, 但它们在实际应用中往往因视角变化、遮挡和光照变化等问题而受限<sup>[8]</sup>。例如, 刘雨萌等<sup>[3]</sup>提出基于关键帧定位的人体异常行为识别模型, 通过筛选和提取视频中的关键帧来提高识别效率。然而, 这些方法在处理高密度客流和复杂背景时, 仍存在一定的局限性<sup>[9]</sup>。

人体骨架识别技术, 特别是Alpha Pose<sup>[10]</sup>和OpenPose<sup>[11]</sup>等模型, 通过提取人体关键点估计姿态, 显示了在复杂环境下识别人体动作的巨大潜力。Alpha Pose<sup>[10]</sup>模型采用区域多人姿态估计技术(regional multi-person pose estimation, RMPE), 显著提高了在复杂背景下的识别准确率<sup>[5]</sup>。这些技术能有效地从遮挡和动态背景中识别出人体, 为进一步的行为分析提供了基础<sup>[11-12]</sup>。此外, 通过结合时空图卷积网络(spatial-temporal graph convolutional network, ST-GCN)模型, 进一步分析人体动作的时空序列变化, 进而提高异常行为的检测精

度和效率。

现有的视频分析技术中, 行为识别算法已取得了一定进展。例如, 吴田等<sup>[9]</sup>提出基于改进ST-GCN的10 kV带电作业人员视频异常行为识别方法, 通过引入通道注意力模块提升了识别准确率。此外, 章东平等<sup>[8]</sup>提出的基于多通道耦合的时空增强异常行为检测方法, 通过引入时间增强模块和空间增强模块, 从而提高了特征提取的效果。

本研究基于城市轨道交通视频监控系统, 整合了前沿的目标检测与人体骨架识别技术, 开发了一种高效的异常行为识别系统。系统采用Alpha Pose模型提取区域多人行为特征, 结合时空图卷积网络(ST-GCN)深度分析人体动作时空序列变化, 能有效识别诸如摔倒、晕倒及打斗等复杂监控场景中的异常行为。

## 2 基于Alpha Pose模型的乘客行为特征提取

### 2.1 Alpha Pose 人体姿态估计模型

人体姿态估计是目前计算机视觉领域的一个重点研究方向, 其通过算法来对视频或图像中的人体进行关键点位置的识别, 从而进行姿态估计, 被广泛应用于智能监控、动画制作等领域。在多人姿态估计领域, 目前主要存在两种框架, 即自顶向下(top-down)和自底向上(bottom-up)。

本文面向的场景是城轨车站, 这类场景面临着客流聚集、遮挡严重等问题, 因此应部署多视角监控, 覆盖车站不同的区域, 避免漏检目标, 同时使用高分辨率设备, 提升高密度人流中人体特征识别质量。综合考虑硬件设施资源和视频分析框架对自底向上和自顶向下方法的处理, 可以得出以Alpha Pose人体姿态估计模型为代表的自顶向下方法更适用于这类复杂场景。因此, 本文选择Alpha Pose模型进行城轨车站内的乘客行为特征提取。不同于其他模型, Alpha Pose采用区域多人姿态估计框架(RMPE)代替单人姿态估计框架(single-person pose estimation, SPPE), 解决了检测框定位错误和姿态冗余的问题, 有效提升了算法的性能。RMPE框架主要由对称空间变化网络(symmetric spatiotemporal transformer networks, SSTN)、姿态引导区域框生成器(pose-guided proposals generator, PGPG)和参数化非极大值抑制(parametric pose NMS, P-NMS)组成, 如图1所示。

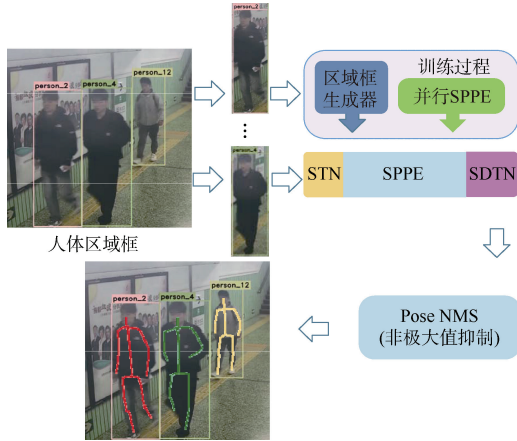


图 1 区域姿态估计框架 RMPE  
Figure 1 RMPE framework

第一个组成部分对称空间变化网络(SSTN)主要用于解决检测框定位错误的问题, 由空间变换网络(spatial transformer networks, STN)和反向空间变换网络(spatial de-transformer networks, SDTN)两部分组成, 如图 2 所示, 当检测框质量较差时, 利用空间变换重新调整检测框使目标人体位于检测框的中心, 优化检测效果。

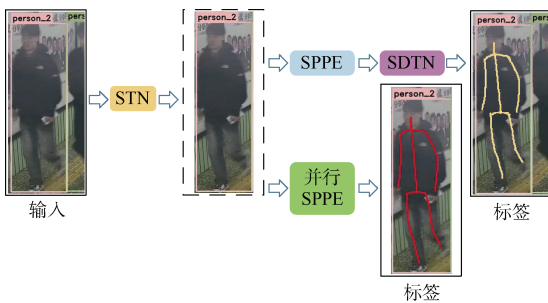


图 2 SSTN 对称空间变换网络  
Figure 2 SSTN Symmetric Spatial Transformer Network

PGPG 考虑应用环境以及人体在监控中被截断的可能性, 根据不同人体姿态检测器的分布生成额外的检测框用于 SSTN+SPPE 模块的训练, 以此进行数据增强, 这种做法能够大大提升姿态估计的准确度。

P-NMS 则使用新的姿态距离度量方法消除多余的姿态, 提高人体姿态估计的精度。P-NMS 由置信度消除和距离消除两种消除标准组成, 置信度消除是指置信度相似的关节点, 而距离消除则是指位置相近的关节点, 只要满足二者之一即会被消除。消除过程如下: 以得分最高的姿态作为基准, 消除与之相近的其他姿态, 直至剩下单一姿态为止。

基于上述改进点, Alpha Pose 作为一种自顶向下的多人姿态估计模型, 首次在 COCO 数据集上达到

72.3 mAP, 相对于运用自底向上框架的 OpenPose 模型提高 17%, 如表 1 所示。

表 1 各开源系统在 COCO 数据集上的效果对比  
Table 1 Comparison of the performance of various open-source systems on the COCO dataset

开源系统	AP	AP <sup>50</sup>	AP <sup>75</sup>	AP <sup>M</sup>	AP <sup>L</sup>
OpenPose (CMU-Pose)	61.8	84.9	67.5	57.1	68.2
Detectron (Mask R-CNN)	67.0	88.0	73.1	62.2	75.6
Alpha Pose	72.3	89.2	79.1	69.0	78.6

此外, Alpha Pose 在 MPII 数据集上也表现优异, 达到 82.1 mAP, 具有较高的精度和跟踪速度, 如表 2 所示。

表 2 各开源系统在 MPII 数据集上的效果对比  
Table 2 Comparison of the performance of various open-source systems on the MPII dataset

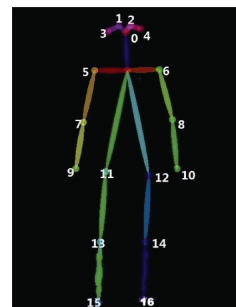
开源系统	头	肩	肘	手腕
OpenPose (CMU-Pose)	91.2	87.6	77.7	66.8
Newell & Deng	92.1	89.3	78.9	69.8
Alpha Pose	91.3	90.5	84.0	76.4

开源系统	臀部	膝	踝	平均
OpenPose (CMU-Pose)	75.4	68.9	61.7	75.6
Newell & Deng	76.2	71.6	64.7	77.5
Alpha Pose	80.3	79.9	72.4	82.1

## 2.2 数据集及关键点标注方案

本文选用 COCO 数据集进行 Alpha Pose 人体姿态估计模型的训练。COCO 数据集是一个包含大规模图像、字幕和对象类别的数据集, 主要用于对象检测、分割等任务<sup>[13]</sup>。其中, COCO 数据集中的关键点标注方案如图 3 所示, 对于每一个人体, 都有 17 个关键点数量, 各个关键点代表的含义如表 3 所示。



注: 17 个关键点含义依次为(0-16): 鼻子、左眼、右眼、左耳、右耳、左肩、右肩、左肘、右肘、左腕、右腕、左臀、右臀、左膝、右膝、左踝、右踝

图 3 COCO 数据集关键点标注方案  
Figure 3 Annotation scheme for keypoints in the COCO dataset

### 3 基于 ST-GCN 的乘客行为特征识别

在运用 Alpha Pose 人体姿态估计模型提取出乘客行为特征后,使用时空图卷积网络模型(ST-GCN)对乘客行为特征进行识别,判断其属于哪一类行为动作。图 4 所示为 ST-GCN 的骨架序列时空图,蓝点表示身体的关节,人体关节之间的连接根据人体自然构造来定义,帧间边根据视频连续帧之间的相同关节来连接。

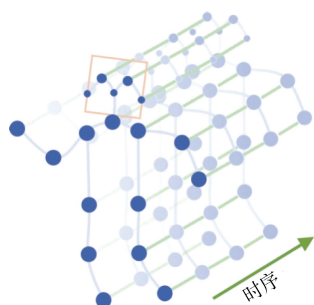


图 4 骨架序列时空图

Figure 4 Spatio-temporal graph of skeleton sequences

ST-GCN 结构为堆叠时空块的层次结构,其内部由空间卷积(GCN)和时间卷积(TCN)组成。空间卷积模块具体实现公式为

$$V_G = \sum_k^{K_s} (V_{in} A_k) W_k \quad (1)$$

$$A_k = D_k^{-2} (\tilde{A}_k + I) D_k^{-2}, D_k^{ij} = \sum_k^{K_s} (\tilde{A}_k^{ij} + I_{ij}) \quad (2)$$

式中,  $V_G$  为空间卷积的特征结果;  $V_{in}$  为输入的节点特征向量矩阵;  $K_s$  为在空间维度上的卷积核;  $A_k$  为卷积核内连接无向图的相邻矩阵;  $W_k$  为在训练过程中可学习权重矩阵;  $D_k$  是度矩阵;  $\tilde{A}_k$  是对原始邻接矩阵  $A_k$  的变换处理,  $I$  为单位矩阵;  $D_k^{ij}$  是度矩阵  $D_k$  中对应元素的计算结果。

时间卷积模型(TCN)是基于空间卷积模块的输出,引入时间维度的卷积( $V, T$ ),具体实现公式为

$$V_T = (C_{in}, V_G, T) \quad (3)$$

式中,  $V_T$  为时空卷积输出特征,  $C_{in}$  为输入通道数;  $T$  为时间维度。

图 5 所示为 ST-GCN 网络。基于 ST-GCN 网络,使用人体姿态估计算法对输入的视频序列进行姿态估计,获得关节的坐标,其次以人体自然构造和时间作为连接构建骨架序列时空图,之后应用多层时空图卷积对其进行卷积,逐步生成更高层次的特征图,最

后运用标准 Softmax 分类器将其分类到相应的动作类别,实现人体姿态的识别。

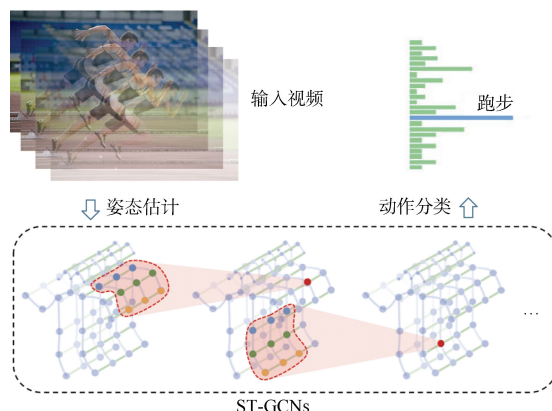


图 5 ST-GCN 网络

Figure 5 ST-GCN network

### 4 乘客行为特征辨识

#### 4.1 数据集构建

常见的需要识别的乘客行为包括行走、坐立、站立等正常行为,以及摔倒、晕倒、打架斗殴等对自身及车站客流状态危害较大的异常行为。

本文所使用的 ST-GCN 模型提供了预训练权重,为提高数据集的质量,使最终训练得到的模型能够更好地应用于城轨车站场景下,本文以北京地铁西直门枢纽站为场景面向城轨车站的乘客行为收集数据集,并将其运用于预训练模型的增量训练中,以期得到更好的模型效果。图 6 展示了收集的城轨车站乘客行为数据集的部分视频图像。经过筛选,本数据集共有 2 000 余张视频图像投入最终训练,数据集类别包含摔倒、晕倒、

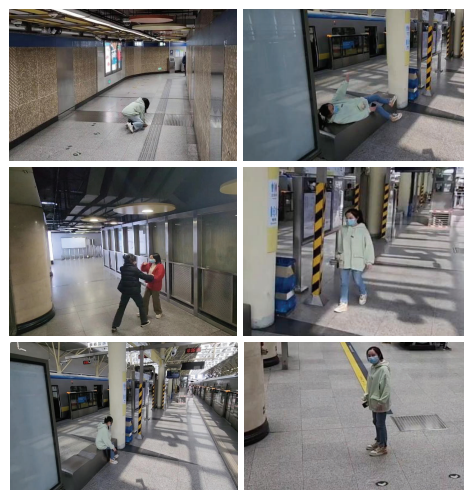


图 6 城轨车站乘客行为数据集示例

Figure 6 Passenger behavior dataset for urban rail stations

打架斗殴、行走、坐立、站立六种不同行为类别图像各 400 余张，其中摔倒、晕倒、打架斗殴为异常行为数据，行走、坐立、站立为正常行为数据。在样本预处理阶段，使用尺度不变特征变换(SIFT)特征匹配方法，评估视频图像质量，过滤掉畸变的样本数据。

#### 4.2 模型训练

本实验通过 AutoDL 提供的环境进行训练，CPU 为 12 vCPU Intel(R) Xeon(R) Platinum 8255C CPU @ 2.50 GHz, GPU 为 RTX 3090(24 GB)\*1, 内存为 43 GB, 操作系统为 ubuntu 20.04, 开发语言为 Python 3.8, 深度学习框架为 Pytorch 1.11.0。

基于深度学习训练平台，将收集的城轨车站乘客行为数据集投入模型的训练中。随机抽取数据集的 20%作为测试集，其余的 80%则作为训练集。

模型训练需要对数据预处理，首先基于 Excel 表格逐帧对城轨车站乘客行为视频图像进行行为类别标注，其次利用预先训练好的 Alpha Pose 模型提取每帧视频图像的人体骨骼关节的坐标，然后对所得人体骨骼关节的坐标数据进行归一化处理等，得到最终输入模型训练的数据。

基于上述所得数据进行 ST-GCN 时空图卷积网络模型的训练，训练参数的设置为批处理大小取值 32, 迭代轮数为 50, 类目数为 6, 学习率为 0.001。

在训练过程中，选取 Adam 优化器进行神经网络模型参数的更新及优化，同时选取交叉熵作为损失函数，判断模型在样本上的表现。Adam 优化器的更新方法为

$$\Delta w_t = \alpha \cdot \frac{m_t}{\sqrt{v_t + \epsilon}} \quad (4)$$

式中， $\Delta w_t$ 为第  $t$  步参数的更新量； $\alpha$  为学习率； $\epsilon$  为能够使分母稳定的系数； $m_t$  为一阶矩； $v_t$  为二阶矩，交叉熵损失函数计算式为

$$L = -\frac{1}{N} \sum_{i=1}^N \sum_{c=1}^M y_{ic} \ln(p_{ic}) \quad (5)$$

式中， $N$  为样本量； $M$  为类别数； $y_{ic}$  为符号函数，当样本  $i$  的真实类别为  $c$  时则取 1，否则为 0； $p_{ic}$  为样本  $i$  属于类别  $c$  的概率，机器学习函数处理中默认是自然对数为底，因此常用  $\log$  函数代表  $\ln$  函数。

#### 4.3 训练结果分析

利用 plt 函数绘制出训练集和测试集的损失及精度变化曲线，如图 7 所示，训练集的训练效果较好，训练曲线能够在短时间内较好地收敛；测试集的训练曲线

有所波动，但在训练约 10 轮后其损失及精度曲线也均趋于平稳。最终在经过 50 轮训练后，训练集的精度达到 99.66%，测试集的精度达到 99.49%，平均精度达到 99.59%。

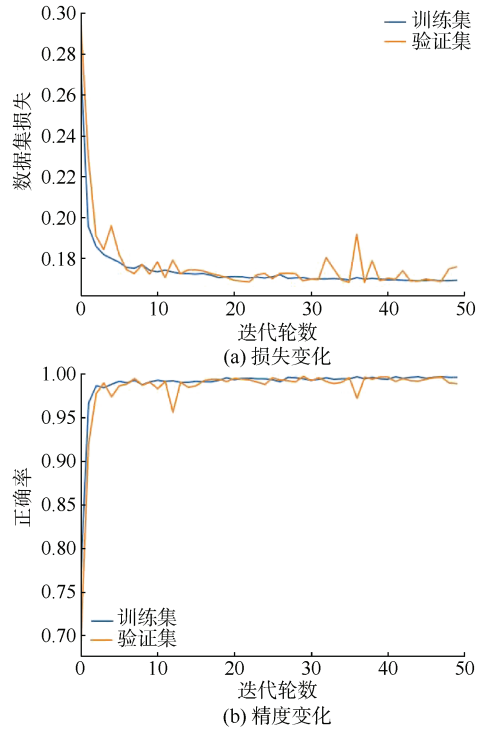


图 7 训练过程损失及精度变化曲线

Figure 7 Chart of loss and accuracy changes during training

#### 4.4 乘客行为特征辨识过程

在得到训练好的模型后开始乘客行为特征的辨识，具体过程如下：

- 1) 输入城轨车站视频采集装备采集到的监控视频图像。
- 2) 利用 Alpha Pose 人体姿态估计模型提取目标乘客的人体骨架序列。
- 3) 利用 ST-GCN 时空图卷积网络模型对提取的人体骨架序列进行分析，识别目标乘客的行为特征。
- 4) 输出乘客行为特征辨识结果。

#### 4.5 乘客行为特征辨识结果展示

运用 Alpha Pose 人体姿态估计模型提取目标乘客人体骨架序列，再利用 ST-GCN 分析人体骨架序列，从而识别出目标乘客的行为特征。基于视频图像的车站乘客行为特征辨识结果如图 8 所示，对于车站内乘客的行为类别(如打架、躺倒、逗留和坐座等)均能较为准确地识别。

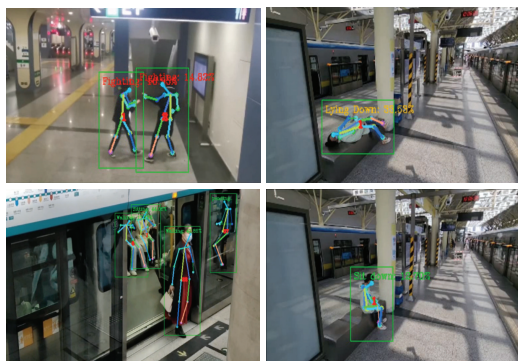


图8 基于城轨车站视频图像的乘客行为特征辨识结果

Figure 8 Identification results of passenger behavior characteristics utilizing urban rail transit station video images

## 5 结论

本研究利用人体骨架技术,通过 Alpha Pose 模型精准地估计乘客的人体姿态,并结合 ST-GCN 模型深入识别其行为特征,实现了视频图像中乘客行为类别的准确判定。主要结论如下:

1) 通过对 Alpha Pose 多人姿态估计模型的若干改进,在 COCO 数据集上具有 72.3 mAP,在 MPII 数据集上具有 82.1 mAP 的较高精度和跟踪速度。

2) 基于深度学习训练平台进行实验,在 50 轮训练后,训练集和测试集分别达到了 99.66%和 99.49%的高精度,平均精度达到 99.59%。

3) 利用 ST-GCN 时空图卷积网络模型对提取出来的人体骨架序列进行分析,识别目标乘客的行为特征,能有效识别诸如摔倒、晕倒及打斗等复杂监控场景中的异常行为,对于提升城轨车站的安全管理水平具有显著意义。未来,该技术有望在城市交通安全管理领域发挥更加重要的作用。

## 参考文献

- [1] 米保伟,王术明,张强.城市轨道交通工程安全监控平台建设及应用[J].都市轨道交通,2023,36(3):10-15.  
MI Baowei, WANG (S/Z)M, ZHANG Qiang. Construction and application of safety risk monitoring platform for urban rail transit projects[J]. Urban rapid rail transit, 2023, 36(3): 10-15.
- [2] 陆海亭,郭伟,郑灵敏.城市轨道交通视频监控系统的回顾与展望[J].都市轨道交通,2014,27(5):93-96.  
LU Haiting, GUO Wei, ZHENG Lingmin. Review and prospect of urban rail transit video surveillance system[J]. Urban rapid rail transit, 2014, 27(5): 93-96.
- [3] 刘雨萌,桑海峰.基于关键帧定位的人体异常行为识别[J].电子测量与仪器学报,2024,38(3):104-111.  
LIU Yumeng, SANG Haifeng. Human abnormal behavior recognition based on keyframes localization[J]. Journal of

electronic measurement and instrumentation, 2024, 38(3): 104-111.

- [4] 刘云萍.基于神经网络的视频监控图像异常行为识别方法[J].数字通信世界,2023(7):71-73.  
LIU Yunping. Recognition of abnormal behavior in video surveillance image based on neural network[J]. Digital communication world, 2023(7): 71-73.
- [5] FANG Haoshu, XIE Shuqin, TAI Y W, et al. RMPE: regional multi-person pose estimation[C]//2017 IEEE International Conference on Computer Vision (ICCV). Venice, Italy. IEEE, 2017: 2353-2362.
- [6] HE Kaiming, GKIOXARI G, DOLLÁR P, et al. Mask R-CNN[C]//2017 IEEE International Conference on Computer Vision (ICCV). Venice, Italy. IEEE, 2017: 2980-2988.
- [7] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV, USA. IEEE, 2016: 779-788.
- [8] 章东平,潘鑫,马道滨,等.基于多通道耦合的时空增强异常行为检测[J/OL].北京航空航天大学学报,1-10 [2024-12-19]. <https://doi.org/10.13700/j.bh.1001-5965.2023.0752>.  
ZHANG Dongping, PAN Xin, MA Daobin, et al. Multi-channel coupled spatial-temporal enhanced abnormal behavior detection[J/OL]. journal of Beijing University of Aeronautics and Astronautics, 1-10 [2024-12-19]. <https://doi.org/10.13700/j.bh.1001-5965.2023.0752>.
- [9] 吴田,万亚旭,王中华,等.基于改进 ST-GCN 的 10 kV 带电作业人员视频异常行为识别方法[J].南方电网技术,2024,18(10):31-39.  
WU Tian, WAN Yaxu, WANG Shenhua, et al. Video Abnormal Behavior Recognition Method for 10 kV Live Workers Based on Improved ST-GCN[J]. Southern power system technology, 2024, 18(10): 31-39.
- [10] FANG Haoshu, LI Jiefeng, TANG Hongyang, et al. Alpha-Pose: whole-body regional multi-person pose estimation and tracking in real-time[J]. IEEE transactions on pattern analysis and machine intelligence, 2023, 45(6): 7157-7173.
- [11] CAO Zhe, SIMON T, WEI Shiheng, et al. Realtime multi-person 2D pose estimation using part affinity fields[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, HI, USA. IEEE, 2017: 1302-1310.
- [12] TOSHEV A, SZEGEDY C. DeepPose: human pose estimation via deep neural networks[C]//2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus, OH, USA. IEEE, 2014: 1653-1660.
- [13] LIN T, MAIRE M, BELONGIE S, et al. Microsoft coco: Common objects in context [C]//13th European Conference. Zurich, Switzerland, 2014: 740-755.

(编辑:王艳菊)