DOI:10. 3969/j. issn. 1672-0032. 2025. 03. 001

基于传感器数据融合的地铁车站隔栏递物检测方法

班魁国,高佼*,阮久宏,沈本兰
山东交通学院轨道交通学院,山东 济南　250357

摘要:为解决地铁车站隔栏递物异常行为检测中场景复杂度高、识别难度大和误检率高等问题,本文提出一种基于激光雷达和相机传感器数据融合的检测方法。采用体素差分算法处理激光雷达点云数据,划分检测区域体素单元并建立递物触发机制。采用目标锁定算法融合激光雷达与相机采集的数据,补充人体关键点的深度信息,锁定隔栏递物行为目标物。对时空图卷积网络(spatial-temporal graph convolutional network, STGCN)进行轻量化改造,降低模型复杂度,减少计算时间;引入时间趋势注意力(temporal trend attention, TTA)模型,增强隔栏递物行为姿态时空变化特征的提取能力,形成 TTA-STGCN 模型,计算隔栏递物行为发生的置信度。通过实验室模拟和地铁车站现场采集隔栏递物数据并制定检测效果评价指标,进行 STGCN 模型、STGCN-MIN 模型、TTA-STGCN 模型的训练、验证和测试,在训练阶段,TTA-STGCN 模型的准确率比前二者均增大 3. 73%,整体损失比前二者均减小 66. 00%;在验证阶段,TTA-STGCN 模型的准确率比前二者分别增大 3. 89%、0. 68%,整体损失比前二者分别减小 58. 95%、58. 48%;在测试阶段,TTA-STGCN 模型的准确率比前两者均增大 3. 15%,整体损失比前二者分别减小 42. 85%、44. 40%。进行现场试验,相比 STGCN 模型、STGCN-MIN 模型,TTA-STGCN 模型的准确率分别增大 2. 99%、3. 49%;精确率分别增大 2. 28%、1. 31%;召回率分别增大 4. 30%、6. 45%;F1 分数分别增大 0. 033 5、0. 040 4,表明 TTA-STGCN 模型显著提高地铁车站特定场景下隔栏递物行为的检测精度。

关键词:隔栏递物;体素差分算法;目标锁定算法;数据融合;STGCN;TTA
中图分类号:U298. 2　　**文献标志码**:A　　**文章编号**:1672-0032(2025)03-0001-11
引用格式:班魁国,高佼,阮久宏,等. 基于传感器数据融合的地铁车站隔栏递物检测方法[J]. 山东交通学院学报,2025,33(3):1-11.
BAN Kuiguo, GAO Jiao, RUAN Jiuhong, et al. Detection method for object passing through subway station barriers based on sensor data fusion[J]. Journal of Shandong Jiaotong University,2025,33(3):1-11.

0　引言

在轨道交通运营过程中,乘客非正常途径进入车站、携带危险品、隔栏递物等异常行为可能扰乱运营秩序,威胁交通安全[1]。地铁车站隔栏递物行为可能导致管制刀具、爆炸物、危险化学品等危险物品进入车站或列车,严重威胁地铁车站、列车运营安全,影响乘客安全出行。及时发现并妥善处理隔栏递物行为是保障地铁车站、列车安全运营和乘客安全出行的有效途径[2]。目前,地铁车站采用人工监控方式检测隔栏递物行为,检测效率较低且漏检率较高,无法实现全时段全场景覆盖。因此,研究能适应复杂场景的自动化检测方法,有助于提高地铁车站隔栏递物检测效率,降低漏检率。

收稿日期:2025-01-21
基金项目:山东省自然科学基金项目(ZR2022QF107)
第一作者简介:班魁国(1980—),男,山东曹县人,硕士研究生,主要研究方向为轨道交通运营安全,E-mail:bkg2006@163. com。
*通信作者简介:高佼(1988—),男,济南人,工学硕士,主要研究方向为机器人环境感知,E-mail:gaojiao@ sdjtu. edu. cn。

通过检测人体姿态判定隔栏递物行为,主要采用光流算法和图像识别算法在复杂场景下分析人体姿态。Rao 等[3]采用光流算法检测人体姿态,通过分析光流矢量长度变化区分不同的姿态特征,提高目标图像的监测准确率。Maji 等[4]采用 YOLO(you only look once)目标检测算法开展多人姿态二维图像检测,通过捕捉和分析视频或静态图像提取运动信息特征。也可采用骨骼模型[5]、红外成像[6]、点云数据[7]等开展人体姿态检测研究。基于整个递物行为场景分析检测隔栏递物,需解析递物行为全时空动态特征,并非直接检测递物本体;检测时传感器受光照变化、复杂场景及目标遮挡影响,检测的准确性和鲁棒性受限。

在地铁车站检测隔栏递物面临检测区域大、场景复杂度高、人员流动密集、检测设备安装受限较多等问题。采用单源传感设备不能精准定位目标检测区域,无法全面捕获目标特征信息,不能精准识别递物目标,如红外传感器无法获取目标颜色信息,激光雷达传感器无法获取目标纹理信息[8]等。多传感数据融合技术可整合不同数据源的多模态信息,提高目标特征信息的识别能力和检测精确度,主要包含特征层级融合[9]、决策层级融合[10]、混合层级融合[11]、模型层级融合[12]等 4 种融合方式。多传感数据融合技术与人工智能、机器学习等技术结合,可提高人体姿态检测精确度和效率[13]。Qiu 等[14]采用多传感器融合技术识别人体姿态,提取融合姿态变化的三个重要特征信息,识别精确度达 96.67%,验证多传感器数据融合算法的有效性。采用激光雷达与相机融合方法在铁路站场内检测障碍物[15]、智能汽车环境感知[16]等实时动态多目标检测取得较好应用效果。

针对地铁车站隔栏递物异常行为检测中场景复杂度高、识别难度大和误检率高等问题,本文提出一种基于激光雷达和相机传感器数据融合的检测方法。采用体素差分算法处理激光雷达点云数据,以期解决隔栏递物行为识别难度较大的问题;采用目标锁定算法融合激光雷达与相机采集的数据,以期实现递物目标的精准锁定,解决隔栏递物行为误检率较高的问题;改进时空图卷积网络(spatial-temporal graph convolutional network, STGCN),引入时间趋势注意力(temporal trend attention, TTA)模型,计算隔栏递物行为发生的置信度,以期解决隔栏递物行为检测精确度较低的问题,为复杂场景下的隔栏递物行为检测提供有效解决方案。

1 检测方案

隔栏递物行为检测方案包括递物事件触发检测、递物目标锁定、递物行为置信度计算等。

1)采用体素差分算法处理激光点云数据,建立隔栏递物触发机制

设置隔栏递物行为检测区域,采用体素差分算法将检测区域三维空间划分为均匀的体素单元,建立基准模型。若当前体素单元点云数据与基准模型的差异值超过设定阈值,则检测区域有目标物入侵,触发隔栏递物行为检测动作。

2)人体姿态关键点坐标信息与激光雷达点云数据融合

采用实时多人姿态估计技术(OpenPose)对相机采集的视频帧图像提取人体骨架平面关键点坐标信息[17-18],采用目标锁定算法融合激光雷达与相机采集的数据,补充人体关键点的深度信息,锁定隔栏递物行为目标物。

3)采用目标锁定算法实现对隔栏递物目标物的精确定位

根据平面深度信息判断检测区域内围栏两侧是否存在隔栏递物目标物,忽略围栏同侧的递物行为,采用目标锁定算法判断隔栏递物行为的深度差值是否在设定阈值范围内,精准定位隔栏递物目标物。

4)采用 TTA-STGCN 模型计算隔栏递物行为发生的置信度

采用 TTA-STGCN 模型提取隔栏递物目标的时空特征时序信息,预测人体动作的变化趋势,计算隔栏递物行为发生的置信度。

在隔栏递物行为检测过程中,需重点关注 2 个检测条件:1)确保激光雷达与相机的时钟同步,在相同时间基准下获取不同传感器采集的特征数据;2)明确检测区域,即标定围栏中线及有效检测空间。

2　检测原理

2.1　体素差分算法

进行目标物入侵检测需设置检测区域。以围栏为中心向两侧扩展，标定长方体空间作为隔栏递物检测区域。将无目标物入侵的检测区域作为基准模型。采用体素差分算法对检测区域划分体素单元并编码[19]，通过激光雷达实时采集检测区域的点云数据，比较体素单元内当前点云数据与基准模型点云数据是否存在差异并统计差异个数，当差异个数超过阈值时，判定有目标物入侵，并作为检测系统启动隔栏递物行为识别的触发条件。检测区域划分体素单元过程示意图如图 1 所示。

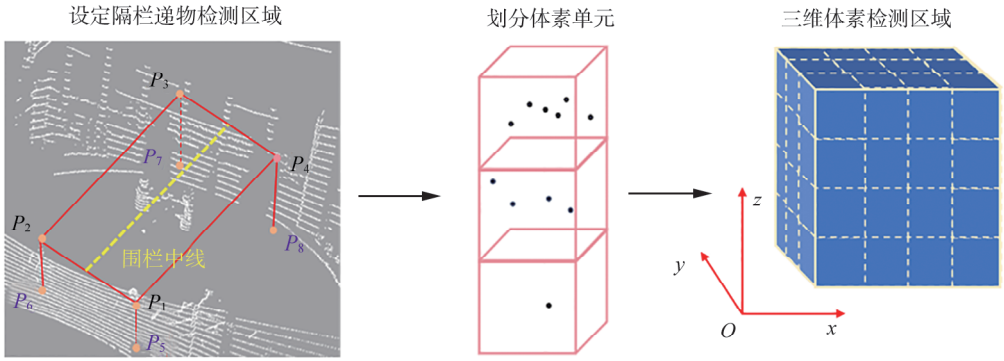


图 1　检测区域划分体素单元过程示意图

由图 1 可知：体素单元为隔栏递物检测区域内均匀的小立方体网格离散单元，体素单元存储激光雷达采集的点云数据，体素单元内点云分布随检测区域状态的变化而变化。通过检测和分析体素单元，能快速高效处理三维体素检测区域的变化情况，降低数据处理的复杂度。

建立激光雷达检测区域三维坐标系，以隔栏递物检测区域左下角为原点，设检测区域的长为 X_{max}、宽为 Y_{max}、高为 Z_{max}，将检测区域划分为等体积的立方体体素单元，设体素单元的边长为 L，N_x、N_y、N_z 分别为 x、y、z 轴方向的体素单元个数，则 $N_x=X_{max}/L$，$N_y=Y_{max}/L$，$N_z=Z_{max}/L$。检测区域内体素单元的总个数

$$N_{total}=N_xN_yN_z \text{ 。}$$

检测系统工作前需初始化基准模型中体素单元点云占用状态，生成体素单元点云占用初始矩阵。检测过程中实时采集点云数据生成体素单元点云占用状态矩阵。分两步计算当前状态下点云占用状态矩阵与基准模型点云占用初始矩阵间的差异值，判断是否有目标物入侵检测区域：1）计算单个体素单元内点云数据变化是否超过阈值 W；2）统计检测区域内点云数据变化超过阈值 W 的体素单元总数 N_C，并判断是否超过阈值 W_P。

单个体素单元内点云数据变化差异值

$$V(i,j,k)=\sum_{i,j,k}\begin{cases}1, \mid \boldsymbol{D}_r(i,j,k)-\boldsymbol{D}_b(i,j,k)\mid > W\\ 0, \mid \boldsymbol{D}_r(i,j,k)-\boldsymbol{D}_b(i,j,k)\mid < W\end{cases},$$

式中：$\boldsymbol{D}_r(i,j,k)$ 为单个体素单元点云占用状态矩阵，$\boldsymbol{D}_b(i,j,k)$ 为单个体素单元点云占用初始矩阵，i、j、k 分别为体素单元在 x、y、z 轴方向上的索引。

检测区域内点云数据变化超过阈值 $\boldsymbol{W}$ 的体素单元总数

$$N_C=\sum_{i=1}^{N_x}\sum_{j=1}^{N_y}\sum_{k=1}^{N_z}V(i,j,k) \text{ 。}$$

当 N_C 个体素单元超过 W_P 时，判定有目标物入侵检测区域，启动隔栏递物行为识别。

采用 VLP16 激光雷达采集检测区域点云数据，考虑到激光雷达点云分布较稀疏，选取体素单元边长 $L=20$ cm，满足检测精度和计算效率要求。受限于激光雷达检测空间范围能力，设置 $X_{max}=400$ cm，$Y_{max}=$

400 cm，$Z_{max}=200$ cm，则 $N_{total}=4\ 000$。

2.2 目标锁定算法

通过融合激光雷达与相机的数据，实现在检测区域内精确定位隔栏递物行为目标物和捕捉行为姿态。出现入侵目标物时，采用目标锁定算法融合激光雷达采集的三维点云数据和相机采集的人体图像信息，增强隔栏递物行为目标物的姿态数据，实现对目标物的精确空间定位。目标锁定算法具体步骤为：1）提取相机视频帧图像，采用实时多人姿态估计库 OpenPose 获取 18 个人体关键点二维坐标；2）对激光雷达点云数据进行筛选、滤波等预处理，去除墙壁、地面等无效点和噪声；3）通过映射、归一化处理将激光雷达采集的点云数据投影到图像二维坐标系；4）将 18 个人体关键点二维坐标与点云数据投影后得到的二维坐标融合，补充人体关键点的深度信息。

设 OpenPose 提取 18 个人体关键点的二维坐标为 $P_m=(u_m,\ v_m)$，其中 $m=1,2,\cdots,18$；激光雷达点云数据的三维坐标为 $\boldsymbol{P}_n=(x_n\quad y_n\quad z_n)^{\mathrm{T}}$；相机内参矩阵为 $\boldsymbol{K}$，将三维相机坐标系的点映射到二维图像坐标系，外参矩阵为 $\boldsymbol{L}$，用于激光雷达坐标系转换为相机坐标系。将激光雷达采集的点云数据 $\boldsymbol{P}_n$ 坐标转换为相机坐标系下的坐标

$$\boldsymbol{P}_n^{\mathrm{cam}}=\boldsymbol{L}\boldsymbol{P}_n\ ,$$

式中 $\boldsymbol{P}_n^{\mathrm{cam}}=(x_n^{\mathrm{cam}}\quad y_n^{\mathrm{cam}}\quad z_n^{\mathrm{cam}})^{\mathrm{T}}$。

将相机坐标系下的点云数据 P_j^{cam} 坐标投影为二维图像坐标系下的齐次坐标

$$\boldsymbol{P}_n^{\mathrm{hom}}=\boldsymbol{K}\boldsymbol{P}_n^{\mathrm{cam}}\ ,$$

式中 $\boldsymbol{P}_n^{\mathrm{hom}}=(u_n^{\mathrm{hom}}\quad v_n^{\mathrm{hom}}\quad w_n)^{\mathrm{T}}$，其中，$w_n$ 为 z_n^{cam} 的深度值。

通过归一化计算得最终的二维图像坐标系坐标

$$P_n^{\mathrm{img}}=(u_n^{\mathrm{hom}}/w_n,v_n^{\mathrm{hom}}/w_n)=(u_n,v_n)\ 。$$

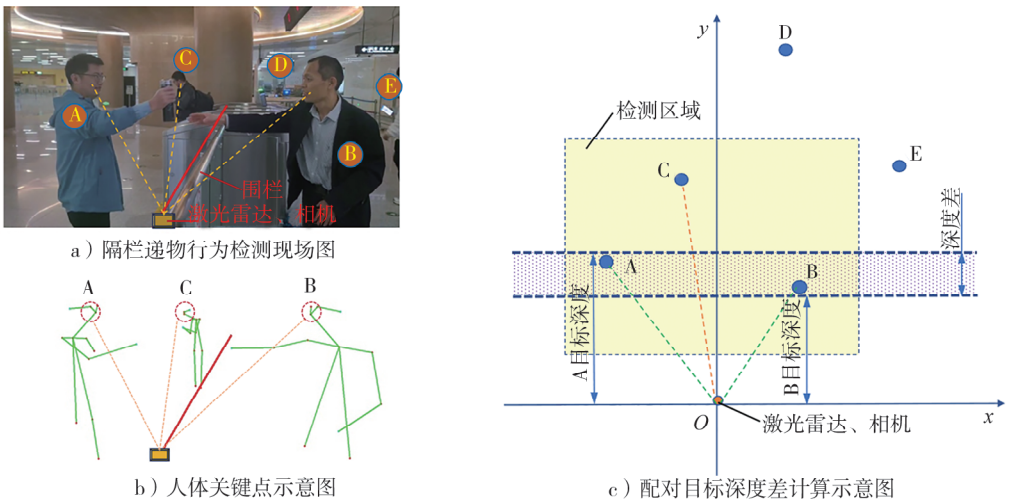
补充人体关键点的深度信息，以 P_m 为圆心，r 为半径开展领域搜索，搜索邻近点云数据。r 不小于 P_m 与邻近点云数据坐标 P_n^{img} 的欧式距离 d_{mn}，即 $r\geqslant d_{mn}$，$d_{mn}=\sqrt{(u_n-u_m)^2+(v_n-v_m)^2}$。

统计搜索到的邻近点云数据的数量为 N_{D}，则平均深度信息

$$w_{\mathrm{avg}}=\sum_{n=1}^{N_{\mathrm{D}}}w_n/N_{\mathrm{D}}\ 。$$

补充深度信息后的人体关键点坐标为 $P_m=(u_m,\ v_m,\ w_{\mathrm{avg}})$。

摄像机的安装位置及角度受车站的复杂场景影响，实际拍摄范围远大于实际要求的检测区域，围栏及两侧部分区域为检测区域，其他区域为非检测区域，采用激光雷达与相机数据特征优势互补进行目标锁定。隔栏递物行为目标锁定算法原理如图 2 所示。



a）隔栏递物行为检测现场图　b）人体关键点示意图　c）配对目标深度差计算示意图

图 2 隔栏递物行为目标锁定算法原理

由图 2 可知:根据激光雷达设定的基准模型检测区域,判定 A、B、C、D、E 五个目标物的特征关键点坐标 P_m 是否在检测区域内,为简化计算复杂度,选择目标物头部区域鼻子部位关键点作为特征坐标用于目标判断,通过融合后的特征坐标深度信息排除 D、E 非监控区域目标物。根据预先标定的围栏中线空间坐标判断检测区域内递物目标物的相对位置并设置左、右标志,目标 A、C 为左检测点,目标 B 为右检测点。设置围栏两侧递物目标物(配对目标)深度差阈值为 Δd,依次判断 A 与 B、C 与 B 的深度差 ε,当 $\varepsilon<\Delta d$ 时,该配对目标为隔栏递物行为的锁定目标。

2.3　TTA-STGCN 模型算法

2.3.1　STGCN 模型

STGCN 是一种结合图卷积网络(graph convolutional network, GCN)与时间卷积网络(temporal convolutional network, TCN)的神经网络深度学习模型,专门用于处理图形结构数据的时空依赖关系,通过图卷积操作捕捉空间元素相关性,并结合时间卷积建模。计算时,时空图卷积网络将输入数据(如人体骨骼关节)映射为图结构,节点映射为空间实体,边映射为空间关系,采用图卷积层和时间卷积层联合提取时空特征[20-21]。

2.3.2　TTA 模型

TTA 是一种针对时间序列或时序数据设计的深度学习模型,采用多头注意力机制增强对关键时间趋势的建模能力,动态捕捉不同时间尺度下的上下文依赖关系。假定输入向量 $\boldsymbol{X}\in\mathbf{R}^{B\times N\times C\times T}$,其中,$B$ 为样本数,N 为关键点数,C 为特征维度,T 为时间序列步数。在处理隔栏递物行为的连续特征时,重点关注递物目标手臂位姿及距离的变化,采用因果卷积操作保证时间序列前、后同步,通过填充操作使输出结构与输入结构一致。TTA 网络通过 $1\times k$ 卷积核操作生成注意力头查询向量 $\boldsymbol{Q}$ 和键向量 $\boldsymbol{K}$,通过 1×1 的卷积核线性变换生成值向量 $\boldsymbol{V}$。采用多头注意力机制,将扩展时间窗口分解为若干固定长度的子窗口,使各注意力头在参数隔离的条件下独立处理局部时序片段,为不同时序片段分配不同权重系数,使网络聚焦于关键时间点,忽略噪声或次要信息[22-23]。

注意力权重矩阵

$$\boldsymbol{A}=\mathrm{softmax}\left(\frac{\boldsymbol{Q}\boldsymbol{K}^{\mathrm{T}}}{\sqrt{d}}\right)\boldsymbol{V},$$

式中:softmax 为归一化指数函数,将实数向量转化为概率分布;d 为注意力头键向量 $\boldsymbol{K}$ 的维度。

2.3.3　TTA-STGCN 模型

对 STGCN 模型进行轻量化改造,将原有的 10 层结构压缩至 6 层,降低模型复杂度,减少计算时间;引入 TTA 网络,增强隔栏递物行为姿态时空变化特征的提取能力,形成 TTA-STGCN 模型。该模型以隔栏递物场景中人体姿态的时序变化特征为学习重点,重点捕捉递物动作过程中人体姿态在时间维度上的连续变化趋势,提升模型对动作特征的动态建模能力,构建适用于隔栏递物场景的时空网络架构。TTA-STGCN 模型工作原理如图 3 所示。

由图 3 可知:TTA-STGCN 模型包含 6 层网络结构,每层网络结构均包含 GCN 模型、TCN 模型和 TTA 模型。TTA-STGCN 模型的输入向量 $\boldsymbol{X}=(T\quad N\quad P_m)$,其中,$T$ 为视频帧数(时间序列步数),N 为人体关节数(关键点数),P_m 为人体关节坐标(人体关键点坐标)。在 TTA-STGCN 模型计算过程中,从第 1 层至第 6 层通道数递增,时间维度减小,有效增加图像特征信息的关注视野。GCN 模型采用人体关节的拓扑关系提取空间特征,捕捉不同关节间的相对位置和连接结构。TCN 模型在各时间维度对各关节的三维坐标进行卷积操作,提取时间特征中关节位置随时间变化的动态信息。TTA 模型采用多头注意力机制在时间维度局部信息变化的时序趋势,通过动态调整感知窗口处理连续帧间关系,有效采集时空特征间的关联性。每一层网络结构输出的卷积结果均采用残差连接方式,有助于缓解深层网络中的梯度消失问题,并加速训练过程。隔栏递物场景中人体姿态动作大部分约为 135 帧,帧率为 20 帧/s。传统的池化层操作在连续帧处理中会导致丢失重要时序信息,TTA-STGCN 网络模型未采用池化层操作,采用全连接层(fully connected layer,FC Layer)将特征信息映射转化为输出结果,并通过 Softmax 函数将实数向量转换为

概率分布，计算隔栏递物行为发生的置信度。

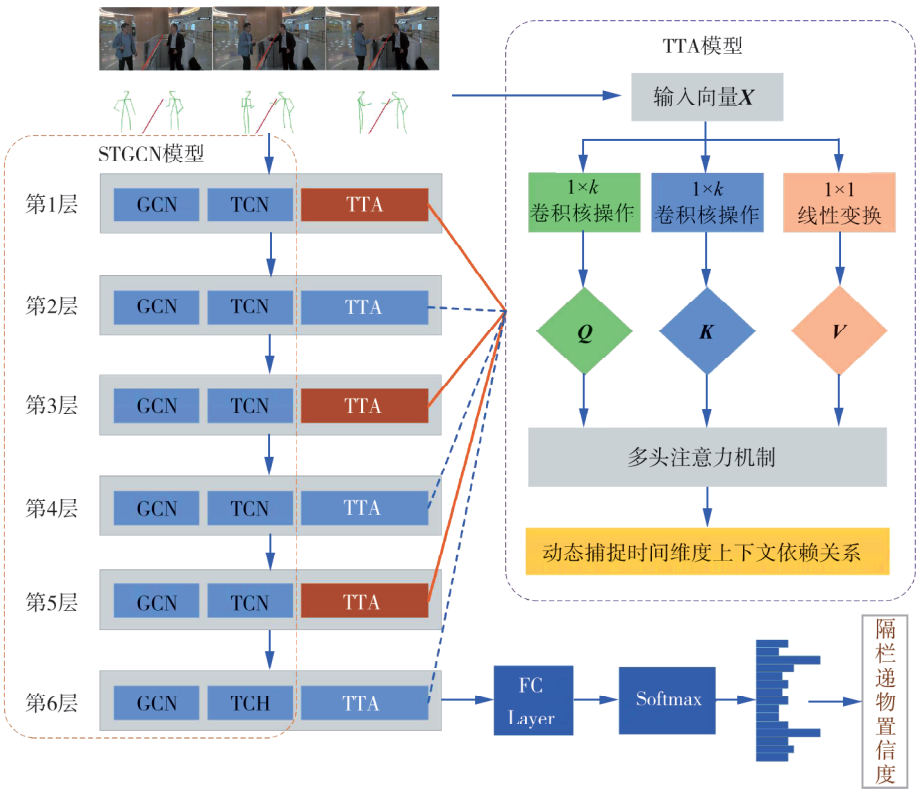


图 3 TTA-STGCN 模型工作原理

3 试验结果与分析

3.1 数据采集

数据采集包含实验室模拟采集和地铁车站现场采集两类样本，并分别采集正、负样本。在本试验中隔栏递物为正样本，非隔栏递物为负样本。共采集视频 1 009 段，其中正样本 533 段，负样本 476 段。对视频进行取帧，获得 137 104 张图像，其中正样本 71 705 张，负样本 65 399 张。

采集过程中通过变化不同递物位置和姿态拍摄视频，包含远、近、上、中、下不同位置，双向递物、远距离投掷、多人不同位置递物等不同姿态。递物的种类包括水杯、雨伞、打火机、衣服等二十余种常见递送物品。对正、负图像样本进行标签编码，以确保数据标签分类清晰。正、负图像样本按比例 7∶1∶2 分为训练集、验证集和测试集。

3.2 试验环境及参数

数据采集环境为 Ubuntu 18.04.6 LTS 版本，操作系统 ROS(robot operating system)采用 Melodic 版本。基于 Windows 10 操作系统训练及测试 TTA-STGCN 模型，配置 NVIDIA GeForce RTX 4070 Ti 显卡，采用 Python 3.7.2 为深度学习语言，采用 Cuda 12.6 编程。设置训练周期数为 100，批次为 64，权重衰减系数为 0.000 4。

3.3 评价指标

试验结果的评价指标包括准确率 P_{acc}、精确率 P_{pre}、召回率 P_{rec}、F1 分数 F。

准确率 P_{acc} 直观评估隔栏递物模型的整体正确性，衡量模型在所有预测中正确预测的样本所占比例，反映模型整体预测能力[24-25]。准确率

$$P_{acc}=(N_{TP}+N_{TN})/(N_{TP}+N_{TN}+N_{FP}+N_{FN})\times 100\%,$$

式中：N_{TP} 为正样本被正确预测为正样本的数量，N_{TN} 为负样本被正确预测为负样本的数量，N_{FP} 为负样本被错误预测为正样本的数量，N_{FN} 为正样本被错误预测为负样本的数量。

精确率 P_{pre} 反映模型预测结果为正样本中真实正样本所占比例，P_{pre} 越大表明模型将非递物行为错误预测为递物行为所占比例越小，则检测系统误报率越低。精确率

$$P_{pre}=N_{TP}/(N_{TP}+N_{FP})\times 100\%。$$

召回率 P_{rec} 反映所有正样本中正确预测的正样本所占比例，P_{rec} 越大表明模型对递物行为正确检测的置信度越大，召回率

$$P_{rec}=N_{TP}/(N_{TP}+N_{FN})\times 100\%。$$

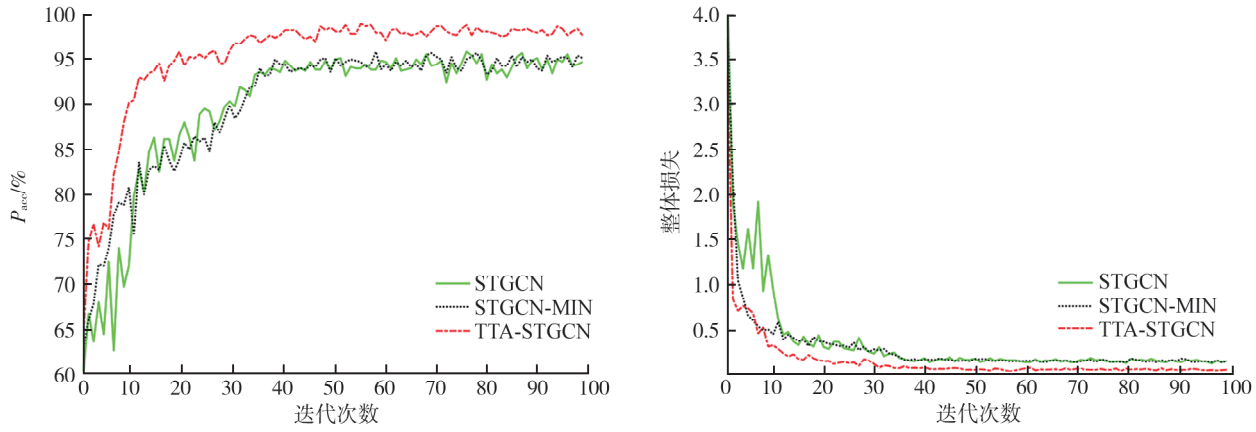
F1 分数 F 为精确率 P_{pre} 和召回率 P_{rec} 的调和平均数，反映递物行为检测性能的综合评价，F 越大表明模型检测性能越好，即负样本被错误预测为正样本的数量较少，且正样本被错误预测为负样本的数量较少。F1 分数

$$F=2(P_{pre}P_{rec})/(P_{pre}+P_{rec})。$$

3.4　模型训练、验证和测试

为验证本文设计模型的有效性，设置对比试验，对 STGCN 模型、STGCN-MIN 模型和 TTA-STGCN 模型进行训练、验证和测试，对比 3 个模型检测隔栏递物行为的准确率和整体损失值。STGCN-MIN 模型为 STGCN 模型的轻量化改造，将原有的 10 层结构压缩至 6 层。

STGCN 模型、STGCN-MIN 模型和 TTA-STGCN 模型在训练阶段的准确率和整体损失曲线如图 4 所示。由图 4a）可知：STGCN 模型的准确率在迭代前 10 次增大较慢且波动较大，迭代 11～37 次准确率逐渐增大，迭代 38～100 次趋于稳定，准确率稳定在 94.40%；STGCN-MIN 模型的准确率在迭代前 10 次增大速度比 STGCN 模型快，迭代 11～100 次变化规律与 STGCN 模型基本一致；TTA-STGCN 模型的准确率在迭代前 10 次增速最快，迭代 11～32 次准确率逐渐增大，迭代 33～100 次趋于稳定，准确率稳定在 98.13%，比 STGCN 模型、STGCN-MIN 模型均提高 3.73%。由图 4b）可知：STGCN 模型的整体损失在迭代前 10 次减小较快且波动较大，迭代 11～37 次整体损失逐渐减小，迭代 38～100 次趋于稳定，整体损失稳定在 0.150 9；STGCN-MIN 模型的整体损失在迭代前 10 次减小速度比 STGCN 模型快，迭代 11～100 次变化规律与 STGCN 模型基本一致；TTA-STGCN 模型的整体损失在迭代前 10 次减小速度最快，迭代 11～37 次逐渐减小，迭代 38～100 次趋于稳定，整体损失稳定在 0.051 3，比 STGCN 模型、STGCN-MIN 模型均减小 66.00%。

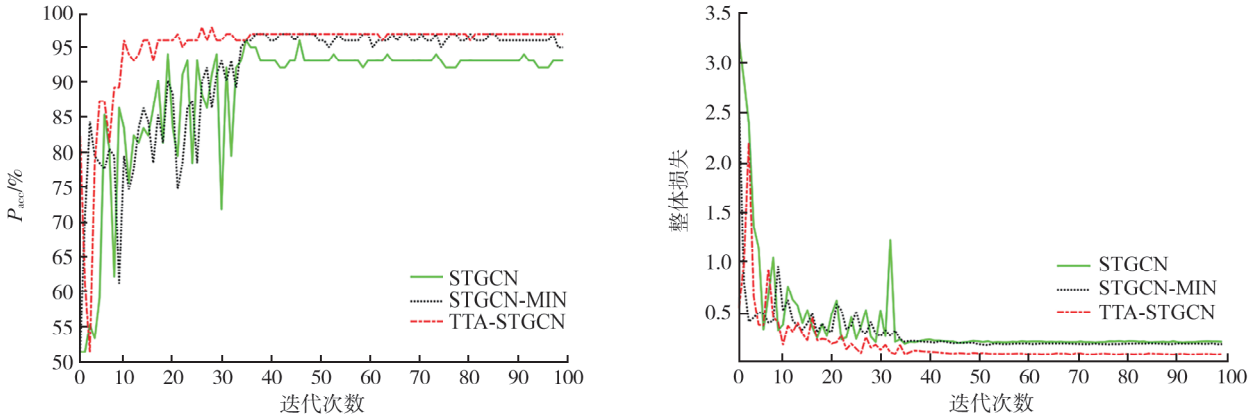


a）准确率曲线　　b）整体损失曲线

图 4　STGCN 模型、STGCN-MIN 模型和 TTA-STGCN 模型在训练阶段的准确率和整体损失曲线

STGCN 模型、STGCN-MIN 模型和 TTA-STGCN 模型在验证阶段的准确率和整体损失曲线如图 5 所示。由图 5a）可知：STGCN 模型的准确率在迭代前 10 次增大较慢，迭代 11～37 次准确率逐渐增大且波动较大，迭代 38～100 次趋于稳定，准确率稳定在 93.17%；STGCN-MIN 模型的准确率在迭代前 10 次增大速度比 STGCN 模型快，迭代 11～100 次变化规律与 STGCN 模型基本一致，准确率稳定在 96.38%；TTA-

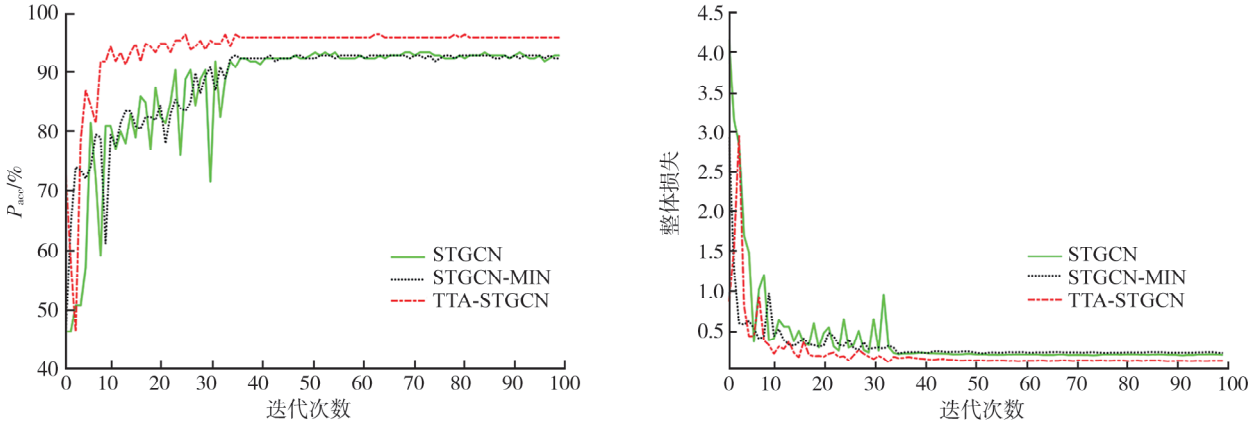
STGCN 模型的准确率在迭代前 10 次增速最快，迭代 11～27 次逐渐增大，迭代 28～100 次趋于稳定，准确率稳定在 97.06%，比 STGCN 模型、STGCN-MIN 模型分别提高 3.89%、0.68%。由图 5b）可知：STGCN 模型的整体损失在迭代前 10 次减小较慢且波动较大，迭代 11～34 次整体损失逐渐减小，迭代 35～100 次趋于稳定，整体损失稳定在 0.201 7；STGCN-MIN 模型的整体损失在迭代前 10 次减小速度比 STGCN 模型快，迭代 11～100 次变化规律与 STGCN 模型基本一致，整体损失稳定在 0.199 4；TTA-STGCN 模型的整体损失在迭代前 10 次减小速度最快，迭代 11～34 次整体损失逐渐减小，迭代 35～100 次趋于稳定，整体损失稳定在 0.082 8，比 STGCN 模型、STGCN-MIN 模型分别减小 58.95%、58.48%。



a）准确率曲线　　b）整体损失曲线

图 5　STGCN 模型、STGCN-MIN 模型和 TTA-STGCN 模型在验证阶段的准确率和整体损失曲线

STGCN 模型、STGCN-MIN 模型和 TTA-STGCN 模型在测试阶段的准确率和整体损失曲线如图 6 所示。由图 6a）可知：STGCN 模型的准确率在迭代前 10 次增大较慢，迭代 11～33 次准确率逐渐增大且波动较大，迭代 34～100 次趋于稳定，准确率稳定在 92.90%；STGCN-MIN 模型的准确率在迭代前 10 次增大速度比 STGCN 模型快，迭代 11～100 次变化规律与 STGCN 模型基本一致；TTA-STGCN 模型的准确率在迭代前 10 次增速最快，迭代 11～33 次逐渐增大，迭代 34～100 次趋于稳定，准确率稳定在 96.05%，比 STGCN 模型、STGCN-MIN 模型均提高 3.15%。由图 6b）可知：STGCN 模型的整体损失在迭代前 10 次减小较慢，迭代 11～33 次整体损失逐渐减小，迭代 34～100 次趋于稳定，整体损失稳定在 0.182 5；STGCN-MIN 模型的整体损失在迭代前 10 次减小速度比 STGCN 模型快，迭代 11～100 次变化规律与 STGCN 模型基本一致，整体损失稳定在 0.187 6；TTA-STGCN 模型的整体损失在迭代前 10 次减小速度最快，迭代 11～33 次整体损失逐渐减小，迭代 34～100 次趋于稳定，整体损失稳定在 0.104 3，比 STGCN 模型、STGCN-MIN 模型分别减小 42.85%、44.40%。



a）准确率曲线　　b）整体损失曲线

图 6　STGCN 模型、STGCN-MIN 模型和 TTA-STGCN 模型在测试阶段的准确率和整体损失曲线

3.5　现场试验

为验证本文设计在隔栏递物行为检测中的实际应用效果,设置现场试验。采用 STGCN 模型、STGCN-MIN 模型和 TTA-STGCN 模型分别检测特定时间段内隔栏递物行为,并计算准确率 P_{acc}、精确率 P_{pre}、召回率 P_{rec} 和 F1 分数 F 四项评价指标,结果如表 1 所示。

表 1　STGCN 模型、STGCN-MIN 模型和 TTA-STGCN 模型隔栏递物行为检测评价指标

模型	P_{acc}/%	P_{pre}/%	P_{rec}/%	F
STGCN	93.53	95.45	90.32	0.928 2
STGCN-MIN	93.03	96.47	88.17	0.921 3
TTA-STGCN	96.52	97.78	94.62	0.961 7

由表 1 可知:TTA-STGCN 模型隔栏递物行为检测的准确率 P_{acc} 比 STGCN 模型、STGCN-MIN 模型分别增大 2.99%、3.49%;精确率 P_{pre} 比 STGCN 模型、STGCN-MIN 模型分别增大 2.28%、1.31%;召回率 P_{rec} 比 STGCN 模型、STGCN-MIN 模型分别增大 4.30%、6.45%;F1 分数 F 比 STGCN 模型、STGCN-MIN 模型分别增大 0.033 5、0.040 4。TTA-STGCN 模型显著提高地铁车站特定场景下对隔栏递物行为的检测精度。

在本文所述测试条件及现场环境下,TTA-STGCN 模型提取单帧图像关键点约需 32 ms,点云处理及深度信息融合时间约为 110 ms,加载网络模型并识别 135 帧时序文件约需 2 s。考虑 TTA-STGCN 模型其他结构的耗时,整体系统响应时间应小于 3 s。

4　结论

针对地铁车站隔栏递物异常行为检测中场景复杂度高、识别难度大和误检率高等问题,从递物事件触发检测、递物目标锁定、递物行为置信度计算 3 个阶段制定隔栏递物行为检测方案。

1)采用体素差分算法处理激光雷达点云数据,分析目标物是否入侵设定区域,作为启动隔栏递物行为检测的触发条件。采用目标锁定算法融合激光雷达与相机采集的数据,补充人体关键点的深度信息,计算围栏两侧递物目标物(配对目标)深度差,锁定隔栏递物行为目标物。

2)对 STGCN 模型进行轻量化改造,降低模型复杂度,减少计算时间;引入 TTA 模型,增强隔栏递物行为姿态时空变化特征的提取能力,形成 TTA-STGCN 模型。该模型以隔栏递物场景中人体姿态的时序变化特征为学习重点,重点捕捉递物动作过程中人体姿态在时间维度上的连续变化趋势,提高模型对动作特征的动态建模能力,适用于计算隔栏递物行为发生的置信度。

3)对 STGCN 模型、STGCN-MIN 模型、TTA-STGCN 模型进行训练、验证和测试,在训练阶段,TTA-STGCN 模型的准确率比前二者均增大 3.73%,整体损失比前二者均减小 66.00%;在验证阶段,TTA-STGCN 模型的准确率比前二者分别增大 3.89%、0.68%,整体损失比前二者分别减小 58.95%、58.48%;在测试阶段,TTA-STGCN 模型的准确率比前二者均增大 3.15%,整体损失比前二者分别减小 42.85%、44.40%。

4)开展现场试验,TTA-STGCN 模型隔栏递物行为检测的准确率 P_{acc} 比 STGCN 模型、STGCN-MIN 模型分别增大 2.99%、3.49%;精确率 P_{pre} 比 STGCN 模型、STGCN-MIN 模型分别增大 2.28%、1.31%;召回率 P_{rec} 比 STGCN 模型、STGCN-MIN 模型分别增大 4.30%、6.45%;F1 分数 F 比 STGCN 模型、STGCN-MIN 模型分别增大 0.033 5、0.040 4,表明 TTA-STGCN 模型显著提高地铁车站特定场景下隔栏递物行为的检测精度。

参考文献:

[1]　吴洁. 城市轨道交通运营安全管理的有效措施[J]. 人民公交, 2024(20): 82-84.

[2]　安俊峰, 刘吉强, 卢萌萌, 等. 基于改进 YOLOv8 的地铁站内乘客异常行为感知[J]. 北京交通大学学报, 2024, 48(2): 76-89.

[3] RAO A S, GUBBI J, MARUSIC S, et al. Crowd event detection on optical flow manifolds[J]. IEEE Transactions on Cybernetics, 2016, 46(7): 1524-1537.

[4] MAJI D, NAGORI S, MATHEW M, et al. YOLO-pose: enhancing YOLO for multi person pose estimation using object keypoint similarity loss [C]//Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. New Orleans, LA, USA: IEEE, 2022: 2636-2645.

[5] YAN S J, XIONG Y J, LIN D H. Spatial temporal graph convolutional networks for skeleton-based action recognition[C]// Proceedings of the AAAI Conference on Artificial Intelligence. New Orleans, Louisiana, USA: AAAI, 2018, 32(1):7444-7452.

[6] BREHAR R D, MURESAN M P, MARIŢA T, et al. Pedestrian street-cross action recognition in monocular far infrared sequences[J]. IEEE Access, 2021, 9: 74302-74324.

[7] DU Y, HOU Z J, LI X, et al. PointDMIG: a dynamic motion-informed graph neural network for 3D action recognition[J]. Multimedia Systems, 2024, 30(4): 192.

[8] SUN Z H, KE Q H, RAHMANI H, et al. Human action recognition from various data modalities: a review[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(3): 3200-3225.

[9] SHI Z S, LIANG J, LI Q Q, et al. Multi-modal multi-action video recognition [C]//Proceedings of IEEE/CVF International Conference on Computer Vision. Montreal, QC, Canada: IEEE, 2021: 13658-13667.

[10] GUO W Z, WANG J W, WANG S P. Deep multimodal representation learning: a survey[J]. IEEE Access, 2019, 7: 63373-63394.

[11] ZHANG D, JU X C, ZHANG W, et al. Multi-modal multi-label emotion recognition with heterogeneous hierarchical message passing[C]//Proceedings of the AAAI Conference on Artificial Intelligence. Pennsylvania, USA: AAAI, 2021, 35(16): 14338-14346.

[12] NIE W Z, YAN Y, SONG D, et al. Multi-modal feature fusion based on multi-layers LSTM for video emotion recognition [J]. Multimedia Tools and Applications, 2021, 80(11):16205-16214.

[13] BLASCH E, PHAM T, CHONG C Y, et al. Machine learning/artificial intelligence for sensor data fusion-opportunities and challenges[J]. IEEE Aerospace and Electronic Systems Magazine, 2021, 36(7): 80-93.

[14] QIU S, ZHAO H K, JIANG N, et al. Multi-sensor information fusion based on machine learning for real applications in human activity recognition: state-of-the-art and research challenges[J]. Information Fusion, 2022, 80: 241-265.

[15] 董文波. 基于激光雷达与相机融合的铁路站场内障碍物检测研究[D]. 成都: 西南交通大学, 2022.

[16] LIU H B, WU C, WANG H J. Real time object detection using LiDAR and camera fusion for autonomous driving[J]. Scientific Reports, 2023, 13(1): 8056.

[17] OSOKIN D. Real-time 2D multi-person pose estimation on CPU: lightweight OpenPose[EB/OL]. (2018-11-29)[2024-12-21]. https://arxiv.org/abs/1811.12004v1.

[18] CAO Z, SIMON T, WEI S H, et al. Realtime multi-person 2D pose estimation using part affinity fields[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA: IEEE, 2017: 1302-1310.

[19] ZHOU Y, TUZEL O. VoxelNet: end-to-end learning for point cloud based 3D object detection[C]//Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA: IEEE, 2018: 4490-4499.

[20] YU B, YIN H T, ZHU Z X. Spatio-temporal graph convolutional networks: a deep learning framework for traffic forecasting[EB/OL]. (2018-07-12)[2024-12-21]. https://arxiv.org/abs/1709.04875v4.

[21] LI S Y, JIN X Y, XUAN Y, et al. Enhancing the locality and breaking the memory bottleneck of transformer on time series forecasting[EB/OL]. (2020-01-03)[2024-12-21]. https://arxiv.org/abs/1907.00235v3.

[22] ZHOU H Y, LIU Q J, WANG Y H. Learning discriminative representations for skeleton based action recognition[C]// Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, BC, Canada: IEEE, 2023: 10608-10617.

[23] 李磊磊. 基于多模态关键语义学习的驾驶环境交通事故预测[D]. 西安:长安大学,2024.

[24] 张浩晨,张竹林,史瑞岩,等. 基于 YOLO-NPDL 的复杂交通场景检测方法[J]. 山东交通学院学报,2025,33(2): 34-37.

[25] 蒋仕新,邹小雪,杨建喜,等. 复杂背景下基于改进 YOLOv8s 的混凝土桥梁裂缝检测方法[J]. 交通运输工程学报,

2024,24(6):135-147.

Detection method for object passing through subway station barriers based on sensor data fusion

BAN Kuiguo, *GAO Jiao**, *RUAN Jiuhong*, *SHEN Benlan*

School of Rail Transportation, Shandong Jiaotong University, Jinan 250357, *China*

Abstract: To address the problems of high scene complexity, difficult recognition, and high false detection rate in detecting abnormal behavior of object passing through subway station barriers, a detection method is proposed that data fusion based on light detection and ranging(LiDAR) and camera sensors. A voxel difference algorithm is used to process LiDAR point cloud data, divide the detection area into voxel units, and establish a trigger mechanism for object passing. An object locking algorithm is employed to fuse data collected by LiDAR and cameras, supplementing depth information of human key points and locking onto target of object passing through barriers. The spatial-temporal graph convolutional network(STGCN) is lightweight-modified to reduce model complexity and computation time. A temporal trend attention(TTA) model is introduced to enhance the extraction of spatial-temporal feature changes in postures of object passing through barriers, forming the TTA-STGCN model to calculate the confidence of behavior occurrence of object passing through barriers. Collecting data of object passing through barriers through laboratory simulation and on-site in subway stations. Detection performance evaluation metrics are established. Training, validation, and testing of STGCN, STGCN-MIN, and TTA-STGCN models are conducted. In the training phase, the accuracy of the TTA-STGCN model improved by 3.73% compared to the first two, the overall loss decreased by 66.00%. In the validation phase, the accuracy of the TTA-STGCN model improved by 3.89% and 0.68% compared to the first two respectively, the overall loss decreasing by 58.95% and 58.48% respectively. In the testing phase, the accuracy of the TTA-STGCN model improved by 3.15% compared to the first two, the overall loss decreasing by 42.85% and 44.40% respectively. Field experiments show that the TTA-STGCN model's accuracy improved by 2.99% and 3.49% compared to STGCN-MIN and STGCN models respectively, precision improved by 2.28% and 1.31% respectively, recall improved by 4.30% and 6.45% respectively, and F1 score improved by 0.033 5 and 0.040 4 respectively, demonstrating that the TTA-STGCN model significantly enhances the detection accuracy of behavior of object passing through barriers in specific subway station scenarios.

Keywords: object passing through barrier; voxel difference algorithm; object locking algorithm; data fusion; STGCN; TTA

(责任编辑:边文超)