DOI：10.13890/j.issn.1000-128X.2024.03.105

基于三维点云与图像融合的轨道交通场景行人检测方法

贺　佳
（国能包神铁路集团有限责任公司，内蒙古 包头 014010）

摘　要：在轨道交通场合由于行人非法侵限造成的安全事故时有发生，严重影响列车的安全运行。利用单一传感器数据进行行人检测存在检测结果召回率低、结果缺乏类别信息或方位信息等问题，无法满足实际现场需求。针对这些问题，文章提出一种基于三维点云与图像融合的轨道交通场景行人检测方法。该方法首先利用构建的轨道交通场景行人数据集训练的深度学习模型分别对三维点云与图像进行行人检测，在此基础上根据靶标在三维点云与图像中的空间位置一致性原理求解出激光雷达与相机之间的旋转平移矩阵，最后将三维点云目标检测结果投影至图像坐标系。为解决多个相邻目标之间误匹配、检测结果之间一对多的问题，通过计算2种检测结果之间的交并比和中心点距离作为融合的约束条件，进而更精确地检测行人。在现场采集的数据集上的试验结果表明，相比于三维点云与图像检测结果，在满足时效性的同时，该方法在检测精度相差不大的情况下，召回率分别提高了4.5%和5.5%，能够有效减少由于行人漏检可能造成的安全事故，满足在列车实际运营过程中对行人检测的需求。

关键词：三维点云；图像检测；深度学习；多传感器融合；目标检测

中图分类号：TP18；U2　　**文献标志码：**A　　**文章编号：**1000-128X(2024)03-0146-10

0　引言

随着我国铁路事业的高速发展，在出行更加便捷的同时，对列车安全运行也提出了更大的挑战。近年，在铁路沿线、站台、道口等场合由于行人非法侵限造成的轨道交通事故时有发生，严重影响列车的安全运营，因此对列车运行过程中沿线的行人进行自动准确的检测，识别侵入物体的具体位置、大小和类别，对保障列车安全可靠运营具有重要意义。

现有的目标检测算法可分为基于图像的目标检测、基于三维点云的目标检测和基于多种传感器数据融合的目标检测。文献[1-2]通过基于深度学习的图像目标检测方法对列车前方的障碍物进行检测，但是该方法忽略了以下问题：相机成像质量受天气和光照影响极大，在典型恶劣天气或者夜间成像质量较差；当轨道交通场景为开放环境下的大场景时，其中的行人目标物体较小，特征不明显；图像检测结果不具备尺寸信息与在三维空间中的方位信息。为解决上述难点，现行轨道交通场景目标检测围绕小目标检测[3-4]、典型恶劣天气条件下的图像增强[5-9]和双目立体视觉测距[10]等关键理论方法进行了大量研究，虽然取得了一定的成果，但是仍然无法有效解决上述原因导致的检测精度低的问题。因此，基于三维点云的目标检测方法逐渐运用于轨道交通场景[11-12]，文献[13]通过对三维点云进行分割，并利用视点特征直方图对分割后的目标点云进行特征提取，最后利用最近点搜索方法完成单物体点云分类识别；文献[14]首先利用栅格地图进行轨道点云检测，并通过轨道拟合进行地面分割，然后对分割后的点云进行特征提取与匹配，从而完成目标检测。与图像目标检测结果相比，该方法能够获取目标的具体位置和目标尺寸等信息，对于轨道交通场景的行人

通信作者：贺　佳，男，主要从事智能驾驶相关研究；Email: 81498144@qq.com

引用格式：贺佳. 基于三维点云与图像融合的轨道交通场景行人检测方法[J]. 机车电传动, 2024(3): 146-155.DOI:10.13890/j.issn.1000-128X.2024.03.105.

Citation：HE Jia. Pedestrian detection method in rail transit scenes based on fusion of 3D point clouds and images[J]. Electric drive for locomotives, 2024(3): 146-155.DOI:10.13890/j.issn.1000-128X.2024.03.105.

侵限检测具有重要意义。但是基于聚类分割的三维点云目标检测方法无法提供目标类别信息，相比于利用图像进行目标检测，三维点云目标检测结果精度较低，文献[15-17]通过利用深度学习的方法进行三维点云目标检测，有效解决了聚类分割方法检测结果缺乏类别信息的问题，但是相较于图像目标检测法，三维点云检测距离较近，而且数据采集与标注难度大，因此有学者研究基于三维点云与图像融合的目标方法[18-20]。文献[21]通过图像目标检测方法提取出目标候选区域，然后将该区域投影至三维点云，进一步对三维点云候选区域进行目标检测，该方法提高了三维点云阶段检测的精度，但是最终检测精度受限于图像检测阶段的检测精度；MV3D 模型[22]将点云的前视图、鸟瞰图和图像同时作为网络输入，从鸟瞰图生成目标的提议框，并投影到 3 个视图中进行特征提取，通过对特征进行融合来回归出 3D 检测框，取得了较好的效果。该方法能够利用不同特征之间的互补性，提升检测结果的精度，但是特征融合导致计算量提升，难以在边缘端部署应用；文献[23]将三维点云投影至图像坐标系，并进行插值生成相应的深度图，然后分别利用深度学习算法对深度图像与 RGB 图像进行目标检测，并利用 D-S 算法对 2 种检测结果进行融合，该方法独立使用点云和图像进行目标检测，可以有效减少计算量，但是点云投影插值计算会损失三维点云的一部分信息，从而造成精度上的损失，并且在融合过程中，没有考虑多个相邻目标之间误匹配、检测结果之间一对多的问题。

针对上述问题，在保证实时性的前提下，为了使目标检测结果在包含类别、尺寸和方位等信息的同时，尽可能降低漏检率，提高目标检测的召回率，本文设计一种基于三维点云与图像融合的轨道交通场景行人检测方法。该方法分别选取 YOLOv8 与 SE-SSD 进行图像和三维点云的目标检测，在此基础上，通过激光雷达与相机之间的时空对齐，将同一时间戳的三维点云目标检测结果投影至图像坐标系，并设计基于投影面积与投影框中心点距离的融合算法，将三维点云目标检测结果与图像目标检测结果进行决策层次的融合，有效解决了多个相邻检测框之间误匹配、一对多的问题，降低了行人漏检率，达到了提高算法可靠性的目的。

1　三维点云与图像融合行人检测框架

本文设计的基于三维点云与图像融合的轨道交通场景行人检测方法的总体结构框架如图 1 所示，主要包含 3 个部分，以下进行详细分析。

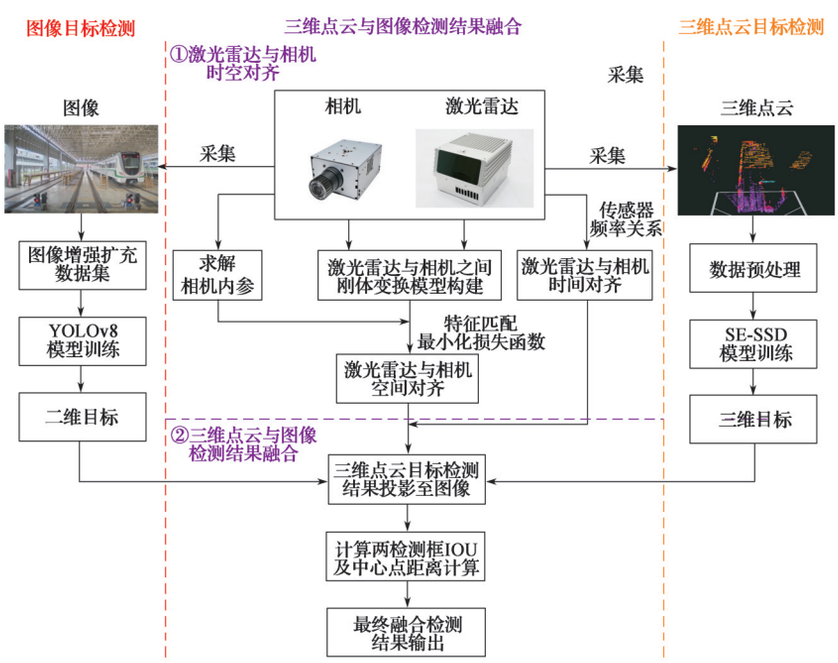


图 1　基于三维点云与图像融合的轨道交通场景行人检测方法

Fig. 1　Pedestrian detection method in rail transit scene based on fusion of 3D point clouds and images

①三维点云与图像目标检测。本文所提的三维点云目标检测采用 SE-SSD 算法获取行人 3D 检测框，其信息包含类别信息、行人在三维空间中的坐标位置和行人的尺寸信息；图像目标检测采用 YOLOv8 算法获取行人检测框，其信息包含类别信息、中心点像素坐标及其长宽尺寸信息等。

②激光雷达与相机时空对齐。通过张正友标定法对相机内参进行求解，然后通过建立激光雷达与相机之间的空间变换模型，对靶标在三维点云与图像中所提取的特征进行匹配，从而求解出二者之间的变换参数，完成激光雷达与相机之间的空间对齐；利用最近时间匹配，完成激光雷达与相机之间的时间对齐。

③三维点云与图像检测结果融合。利用激光雷达与相机之间的旋转平移矩阵与相机内参矩阵，将三维

点云目标检测结果投影至图像坐标系，通过将三维点云检测结果投影值与图像检测结果进行交并比（Intersection Over Union, IOU）计算与中心点距离匹配，将匹配成功的检测框全部输出，而对于未匹配成功的检测框，分别输出其检测结果，最终得到同时具有类别信息与方位信息的检测结果。

2 三维点云与图像融合行人检测方法

2.1 图像及三维点云目标检测算法

2.1.1 图像目标检测算法

本文使用的图像基础检测算法框架为YOLOv8模型，框架分为数据预处理、Backbone、Neck、Head 4个部分。算法首先对输入的图像进行Mosaic数据增强，然后通过随机缩放、随机裁剪和随机排布的方式对图像进行拼接，以增强图像的多样性；Backbone部分主要由C2F模块组成，如图2所示，相较于传统CSP模块，C2F增加1个split对更多的梯度流分支进行并行处理，在保证轻量化的同时获得更加丰富的梯度流信息；在Neck部分利用SPPF结构替代原有的SPP结构，如图3所示，相较于SPP模块，SPPF采用多个小尺寸池化核级联代替SPP模块中单个大尺寸池化核，进一步融合不同感受野的特征图，在丰富特征图的表达能力的情况下，进一步提高了运行速度；最后在Head层将分类和回归进行解耦，分别对目标的类别和位置进行预测，大幅度提高模型的收敛速度。

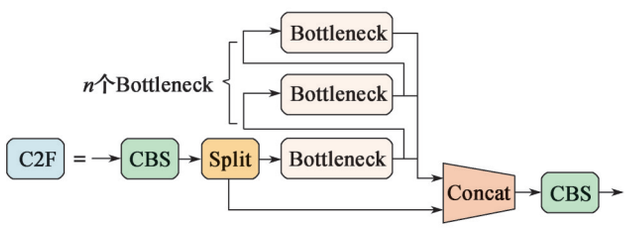


图2 C2F模块示意图

Fig. 2 Schematic diagram of C2F module

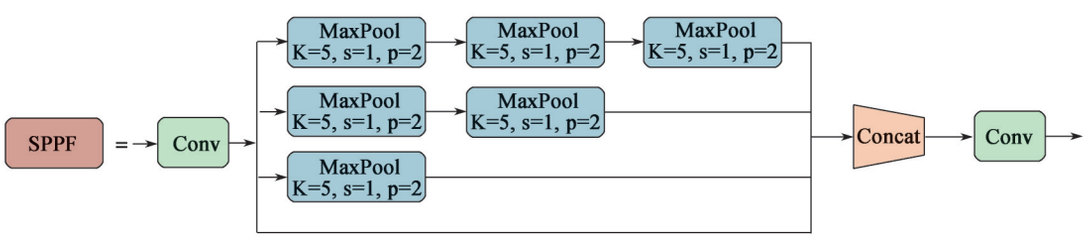


图3 SPPF模块示意图

Fig. 3 Schematic diagram of SPPF module

2.1.2 三维点云目标检测算法

本文使用的三维点云基础检测算法框架为SE-SSD模型。由于遮挡、距离变化和物体形状等因素，同一目标物体的三维点云差异极大，SE-SSD模型设计了一种数据增强方法，对于点云中的每个目标，通过将真值边界框中心与顶点连接起来，将目标点分成6个子集，然后对6个子集进行随机交换、稀疏和丢失等操作，以模拟点云的遮挡、稀疏和缺失等问题，进而提高原始训练数据的鲁棒性。此外，为避免三维点云由于遮挡或距离导致的同一物体的样本点云特征差异大的问题，SE-SSD算法通过同时训练student和teacher这2个网络模型，并将相对精确的teacher预测结果当作soft目标，利用它们来监督student模型，从而使得student网络具有更好的检测精度。在进行监督的过程中，通常采用Smooth-L1损失来约束边界框回归，本文利用SE-SSD算法进行三维点云目标检测的流程如图4所示。

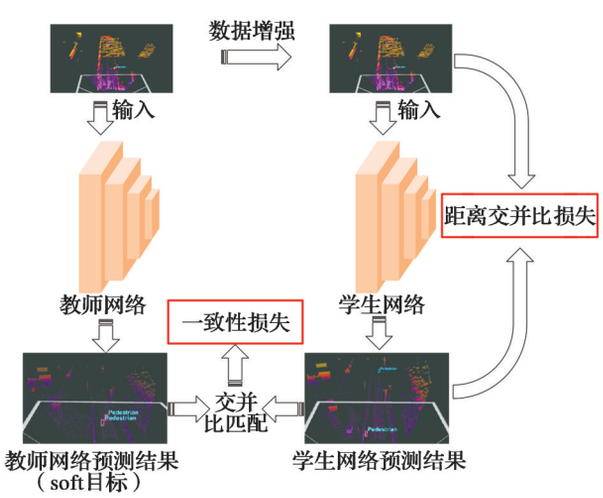


图4 SE-SSD目标检测流程示意图

Fig. 4 Flowchart of SE-SSD object detection

2.2 激光雷达与相机时空对齐

2.2.1 激光雷达与相机空间对齐

激光雷达与相机分别获取目标物的三维点云与图像信息，本文采用基于靶标的激光雷达与相机之间的联合标定。激光雷达、相机和靶标组成的联合标定系统如图5所示，共有激光雷达坐标系、相机坐标系、世界坐标系、图像坐标系4种坐标系。激光雷达坐标系可以描述物体与激光雷达的相对位置，表示为$[X_L, Y_L, Z_L]$；相机坐标系可以描述物体与相机的相对位置，表示为$[X_C, Y_C, Z_C]$；像素坐标系表示为$[u, v]$，其中原点为图像左上角，u轴水平向右，v轴竖直向下。

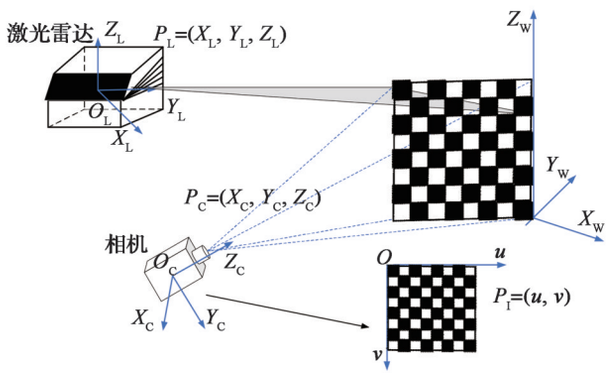


图 5　基于靶标的联合标定

Fig. 5　Joint calibration based on target

在不考虑相机成像畸变的情况下，激光雷达坐标系下点P_L与其像素投影点p_I之间的转换关系为

$$s\begin{bmatrix}u\\v\\1\end{bmatrix}=\begin{bmatrix}\frac{1}{\mathrm{d}X}&0&u_0\\0&\frac{1}{\mathrm{d}Y}&v_0\\0&0&1\end{bmatrix}\begin{bmatrix}f&0&0&0\\0&f&0&0\\0&0&1&0\end{bmatrix}\begin{bmatrix}\boldsymbol{R}&\boldsymbol{T}\\\boldsymbol{0}&1\end{bmatrix}\begin{bmatrix}X_L\\Y_L\\Z_L\\1\end{bmatrix}=$$

$$\begin{bmatrix}f_x&0&u_0&0\\0&f_y&v_0&0\\0&0&1&0\end{bmatrix}\begin{bmatrix}\boldsymbol{R}&\boldsymbol{T}\\\boldsymbol{0}&1\end{bmatrix}\begin{bmatrix}X_L\\Y_L\\Z_L\\1\end{bmatrix}=\boldsymbol{M_1M_2}P_L \tag{1}$$

式中：(u_0, v_0)为相机中心点在图像坐标系中的位置坐标；dX,dY分别为单个像素的物理尺寸；s为比例因子；f为相机焦距的长度；$\boldsymbol{R}$为由激光雷达坐标系到相机坐标系的旋转参数矩阵，是1个3×3正交单位矩阵；$\boldsymbol{T}$为由激光雷达坐标系到相机坐标系的平移向量矩阵，是1个3×1矩阵。

记上述由激光雷达坐标系下点到图像坐标系下点的坐标变换为

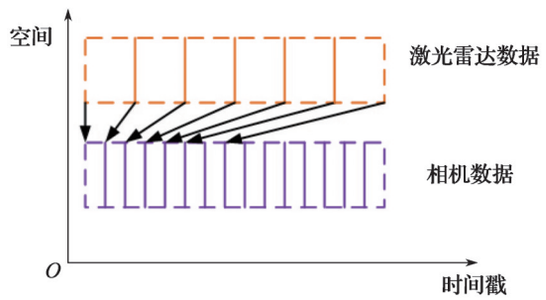
$$sp_I=\varphi(P_L:\boldsymbol{M_1},\boldsymbol{M_2}) \tag{2}$$

在式(2)中，$\boldsymbol{M_1}$只与相机内部参数有关，可采用张正友标定法求解相机内参矩阵，故此处只需通过提取靶标特征来建立相应的约束函数即可求解出激光雷达与相机之间的外参矩阵$\boldsymbol{M_2}$。

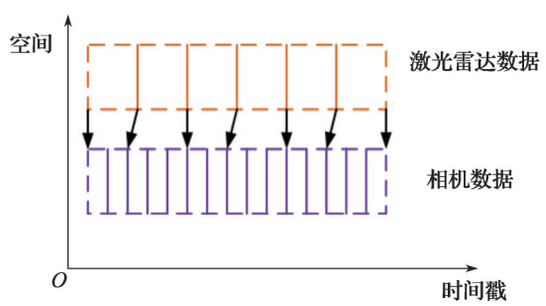
2.2.2　激光雷达与相机时间对齐

在实际使用过程中，不同传感器之间除空间坐标系不同外，两者之间的工作频率也不同，若不将两者进行时间对齐，将会出现在同一时间节点所处理的三维点云数据与图像数据并非同一场景，从而导致融合失败。

本文所使用的激光雷达工作频率为10 Hz，相机工作频率为25 Hz，由于激光雷达工作频率小于相机工作频率，若采用逐帧对齐，如图6(a)所示，将会在每2帧对齐的过程中产生固定的时间差，并且随着匹配帧数的增加时间差会越来越大，从而导致错误的结果。因此，本文采用最近时间匹配的方法进行时间对齐操作，如图6(b)所示，在对齐的过程中以激光雷达时间戳为基准，当处理器接收到三维点云时，向前匹配与该时间戳最近的图像数据，从而保证处理的三维点云与图像是近似处于同一时间戳下的数据。



(a) 逐帧对齐



(b) 最近时间对齐

图 6　时间对齐原理示意图

Fig. 6　Schematic diagram of time alignment principle

2.3　三维点云与图像检测结果融合

本文在三维点云与图像检测结果的基础上，利用三维点云与图像之间的坐标转换关系，设计相应的融合算法，使得最终输出为两者之间的融合，且在任一传感器数据失效的情况下，仍能输出其中一种检测结果，以提升算法的鲁棒性，算法流程图如图7所示。

利用SE-SSD模型进行三维点云目标检测，输出为目标在三维空间中的位置(x_1, y_1, z_1)、目标的物理尺寸$(\Delta x, \Delta y, \Delta z)$。因此，可得目标物在三维空间中的范围为

$$\left.\begin{aligned}x_{1\min}&=x_1-\frac{1}{2}\Delta x,\quad x_{1\max}=x_1+\frac{1}{2}\Delta x\\y_{1\min}&=y_1-\frac{1}{2}\Delta y,\quad y_{1\max}=y_1+\frac{1}{2}\Delta y\\z_{1\min}&=z_1-\frac{1}{2}\Delta z,\quad z_{1\max}=z_1+\frac{1}{2}\Delta z\end{aligned}\right\} \tag{3}$$

式中：$x_{1\min}, y_{1\min}, z_{1\min}$和$x_{1\max}, y_{1\max}, z_{1\max}$分别表示目标点云在三维空间3个方位上的最小值和最大值。

由此得到三维点云目标检测框为$\boldsymbol{p}_{1i}=(x_{1i}, y_{1i}, z_{1i})$（$i=1,2,\cdots,8$），根据激光雷达与相机之间的外参矩阵$\boldsymbol{M_2}$和相机内参矩阵$\boldsymbol{M_1}$，将三维点云检测结果投影至图像坐标系，可得三维点云检测结果在图像坐标系上的投影$\boldsymbol{p}_{ci}=(u_i, v_i)$（$i=1,2,\cdots,8$）为

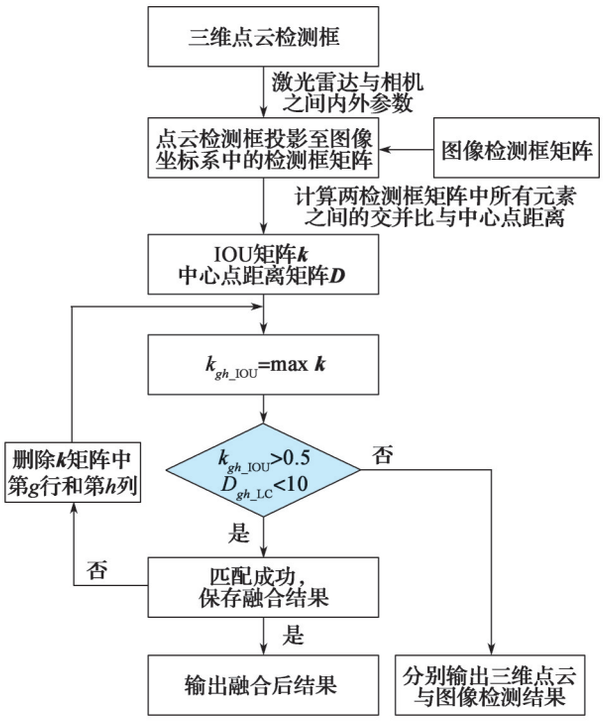


图 7　三维点云与图像检测结果融合算法流程图

Fig. 7　Flowchart of fusion algorithm for 3D point cloud and image detection results

$$\boldsymbol{p}_{ci}=\begin{bmatrix}u_i\\v_i\\1\end{bmatrix}=\frac{1}{s}\begin{bmatrix}f_x&0&u_o&0\\0&f_y&v_o&0\\0&0&1&0\end{bmatrix}\begin{bmatrix}\boldsymbol{R}&\boldsymbol{T}\\\boldsymbol{0}&1\end{bmatrix}\begin{bmatrix}x_{li}\\y_{li}\\z_{li}\\1\end{bmatrix}\tag{4}$$

将上述三维点云投影结果所形成面积最大的矩形记为 $L_{_BBox}$，记 YOLOv8 模型检测的目标框为 $C_{_BBox}$，由三维点云检测结果投影框与图像检测框之间进行融合输出最终结果过程步骤如下：

①记三维点云检测结果的投影检测框为 $\{L_{i_BBox}\}$（$i=1,2,\cdots,m$），图像检测目标框为 $\{C_{j_BBox}\}$（$j=1,2,\cdots,n$）。

②计算 $\{L_{i_BBox}\}$ 与 $\{C_{j_BBox}\}$ 里每一个检测框之间的交并比和中心点之间的像素距离，其中 2 个检测框之间的交并比计算如下：

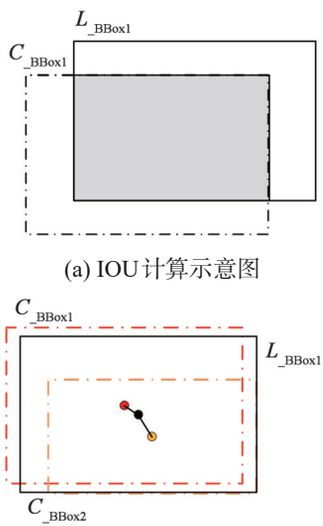
$$k_{ij_IOU}=\frac{L_{i_BBox}\cap C_{j_BBox}}{L_{i_BBox}\cup C_{j_BBox}}\tag{5}$$

设图像目标检测框中心点坐标为 (u_c,v_c)，三维点云检测结果投影值的中心点坐标为 (u_l,v_l)，定义中心点距离为

$$D_{LC}=|u_c-u_l|+|v_c-v_l|\tag{6}$$

由式(5)和式(6)可得 2 个检测框集合之间的交并比矩阵 $\boldsymbol{k}=\begin{bmatrix}k_{11_IOU}&k_{12_IOU}&\cdots&k_{1n_IOU}\\k_{21_IOU}&k_{22_IOU}&\cdots&k_{2n_IOU}\\\vdots&\vdots&&\vdots\\k_{m1_IOU}&k_{m2_IOU}&\cdots&k_{mn_IOU}\end{bmatrix}$ 和中心点距离矩阵 $\boldsymbol{D}=\begin{bmatrix}D_{11_LC}&D_{12_LC}&\cdots&D_{1n_LC}\\D_{21_LC}&D_{22_LC}&\cdots&D_{2n_LC}\\\vdots&\vdots&&\vdots\\D_{m1_LC}&D_{m2_LC}&\cdots&D_{mn_LC}\end{bmatrix}$。

交并比和中心点距离计算过程示意图如图 8 所示，其中实线框表示 $L_{_BBox}$，虚线框表示 $C_{_BBox}$。



(a) IOU 计算示意图

(b) 中心点距离计算示意图

图 8　二维检测框与三维检测框投影值交并比和中心点距离计算示意图

Fig. 8　Schematic diagram for calculation of intersection over union and center point distance between 2D and 3D detection frameworks

③设 IOU 阈值 $\alpha=0.5$，当图像检测结果与三维点云检测结果投影值之间的交并比 $k_{ij_IOU}\in[0,0.5)$ 时，判断其属于 2 个单独的目标物。

④令 $k_{gh_IOU}=\max\boldsymbol{k}$，当 $k_{gh_IOU}>0.5$ 且 $D_{gh_LC}<10$ 时，则认为融合成功。取三维点云投影框与图像检测框之间并集作为新的目标检测框，记为 $K_{i_LC_BBox_gh}$；同时，记本次融合成功的三维点云投影框序号为 g_i，融合成功的图像检测框序号为 h_i，删除矩阵 $\boldsymbol{k}$ 中第 g 行与第 h 列。

⑤重复步骤④，直至矩阵中没有满足条件的值，输出融合检测结果 $\boldsymbol{K}_{LC_BBox}=\{K_{i_LC_BBox_gh}\}$（$i=1,2,\cdots,k$）；

⑥令 $\boldsymbol{g}=\{g_i\},\boldsymbol{h}=\{h_i\}$（$i=1,2,\cdots,k$）。对尚未匹配成功的三维点云投影框 $\forall i\in\{1,2,\cdots,m\}$ and $i\notin\boldsymbol{g}$，输出 $\{L_{i_BBox}\}$；对于尚未匹配成功的二维图像检测框 $\forall i\in\{1,2,\cdots,n\}$ and $i\notin\boldsymbol{h}$，输出 $\{C_{j_BBox}\}$。

3　试验结果及分析

3.1　试验数据集构建

本文数据采集系统如图 9 所示。其中，三维点云数据采集所用硬件设备安装于车头处；图像数据采集所用硬件设备为眼擎相机，安装于司机室车窗位置；数

据采集场景为铁路站点、段内和部分沿线运行场景。除去该套数据采集系统采集的数据外，为进一步提升数据集中样本的多样性，原始数据组成还包括其他轨道交通场景数据若干。最终数据集包含三维点云与对应的图像数据各 1 288 帧，主要用于模型的训练与测试。由于本文所采集的轨道交通场景行人数据集包含段内、库内和正线等场景，如图 10 所示，均匀选取不同场景的图像与三维点云数据各 270 帧组成测试集；考虑本文三维点云检测行人的有效距离为 80 m，图像检测行人有效距离为 200 m，为验证本文所提融合检测算法的有效性，本文选取的三维点云与图像数据均在三维点云的有效检测距离内，即满足行人在距激光雷达原点 80 m 范围内，270 帧共有行人真实标签 400 个。剩余的三维点云与图像数据集各 1 018 帧组成训练集，表 1 为轨道交通场景数据集各类标签的分布情况。

表 1　数据集标签分布

Table 1　Distribution of dataset labels

数据集	数量	行人标签数量
测试集	270	400
训练集	1 018	1 452

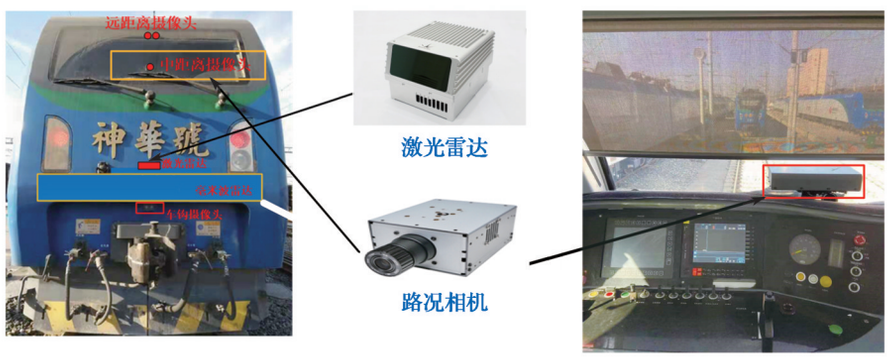


图 9　数据采集系统

Fig. 9　Data acquisition system



(a) 段内场景　　(b) 挂车场景

(c) 正线场景　　(d) 库内场景

图 10　典型轨道交通场景行人图像数据

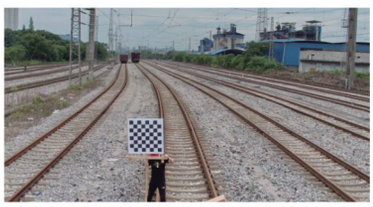
Fig. 10　Typical image data of pedestrians in rail transit scenes

3.2　激光雷达与相机联合标定试验

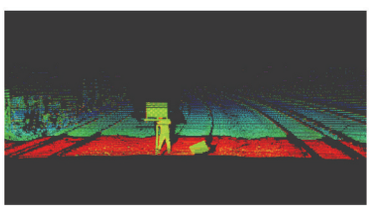
本文所使用的激光雷达与相机之间的联合标定采用基于靶标的标定方法，激光雷达与相机现场采集的带有靶标的三维点云与图像数据如图 11 所示。

提取靶标在三维点云与图像中的角点与中心点法向量作为特征，匹配靶标角点与中心点法向量在 2 种坐标系中的位置，建立并最小化损失函数，进而求解出激光雷达与相机之间的旋转平移矩阵。为了对联合标定结果进行验证，本文首先进行重投影试验，将标定完成后采集的现场三维点云数据投影至对应的图像坐标系，其效果图如图 12 所示。

由图 12 可以看出，联合标定结果具有很高的准确率。为了对联合标定结果进行进一步验证，在重投影



(a) 图像



(b) 三维点云

图 11　原始数据采集

Fig. 11　Raw data acquisition

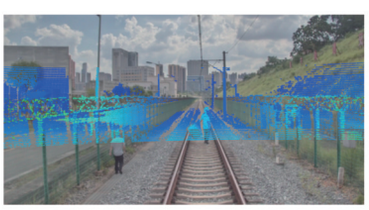




图 12　三维点云在图像上的投影

Fig. 12　Projection of 3D point clouds onto images

定性分析的基础上，通过计算靶标角点的平均像素误差与标准差对标定结果进行定量分析，结合式(2)，定义角点平均像素误差为

$$\varepsilon_{\mathrm{err}}=\sqrt{\frac{1}{4N}\sum_{i=1}^{4N}\left(\frac{1}{s}\varphi\left(\boldsymbol{P}_{\mathrm{Le}_i}:\boldsymbol{M}_1,\boldsymbol{R},\boldsymbol{T}\right)-\boldsymbol{p}_{\mathrm{le}_i}\right)^2} \tag{7}$$

式中：$\varepsilon_{\mathrm{err}}$为靶标角点平均像素误差；$\boldsymbol{P}_{\mathrm{Le}_i}$为靶标角点在三维点云中的坐标；$\boldsymbol{p}_{\mathrm{le}_i}$为靶标角点在图像坐标系中的位置；$N$为参与角点提取的靶标的数量。

使用不同数量的靶标对标定结果进行验证的结果如图 13 所示。

由图 13 可知，激光雷达与相机之间联合标定结果角点平均像素误差小于 2 个像素点，标准像素误差小于 1 个像素点，即标定结果具有较高的精度。

3.3　行人检测融合试验

为了对本文所提的融合算法进行验证，使用人工统计的数据作为参考依据，基于本文算法检测出的行人数量对所提方法的召回率和精确率进行计算，其中召回率和精确率计算如下：

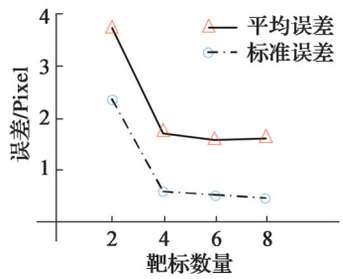


图 13　采用不同靶标数量进行联合标定时角点平均误差与标准误差

Fig. 13　Average error and standard error of corner points during joint calibration with different number of targets

$$R=T_{\mathrm{P}}/\left(T_{\mathrm{P}}+F_{\mathrm{N}}\right) \tag{8}$$

$$P=T_{\mathrm{P}}/\left(T_{\mathrm{P}}+F_{\mathrm{P}}\right) \tag{9}$$

式中：R为召回率；P为精确率；T_{P}为检测到的行人与真实类别一致的样本数量；F_{P}为检测到的行人与真实类别不一致的数量；F_{N}为未被检测的行人数量。

本文算法测试环境软件平台基于 Linux 系统 Ubuntu 发行版本，其中基于图像与三维点云的目标检测具体的环境配置如表 2 所示。

表 2　图像与三维点云的目标检测环境配置

Table 2　Configuration of object detection environment for images and 3D point clouds

项目	操作系统	GPU	深度学习框架	编程语言
说明	Ubuntu18.04	RTX 3090	Pytorch	Python

分别利用 YOLOv8 模型和 SE-SSD 模型对图像与三维点云进行目标检测，结果分别如图 14 和图 15 所示。

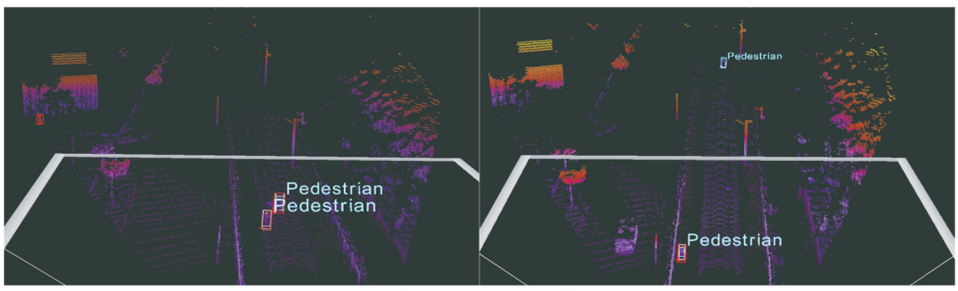


—— 真值；　—— 检测值。

图 14　基于图像的目标检测结果

Fig. 14　Image-based object detection results

从上述检测结果可以看出，利用单一传感器数据进行行人检测时会存在不同程度的误检或者漏检。如图 14 所示，在行人前方出现遮挡或行人与背景相似的情况下，利用图像进行检测会存在漏检；如图 15 所示，三维点云将轨道旁的灌木林、立柱等物体误检为行人，



—— 真值；　—— 检测值。

图 15　基于三维点云的目标检测结果

Fig. 15　Object detection results based on 3D point clouds

并且对相对较远距离的行人也会漏检。

利用上文得到的激光雷达与相机之间的旋转平移矩阵，将三维点云目标检测结果投影至图像坐标系，并取其最大投影矩阵为新的检测框，具体如图 16 所示。

通过将三维点云检测结果与图像检测结果进行投影匹配，同时将未匹配成功的目标单独输出，其过程如图 17 所示。

图 17(a)为三维点云与图像分别存在漏检、图像漏检、三维点云漏检的场景，图 17(b)为由 2 种结果融合后的输出。由融合结果可知，通过本文所提融合方法对 2 种检测结果进行融合，可以有效解决单一传感器数据检测结果漏检多和召回率低的问题。利用测试集数据对本文所提融合算法进一步验证，结果如表 3 所示。

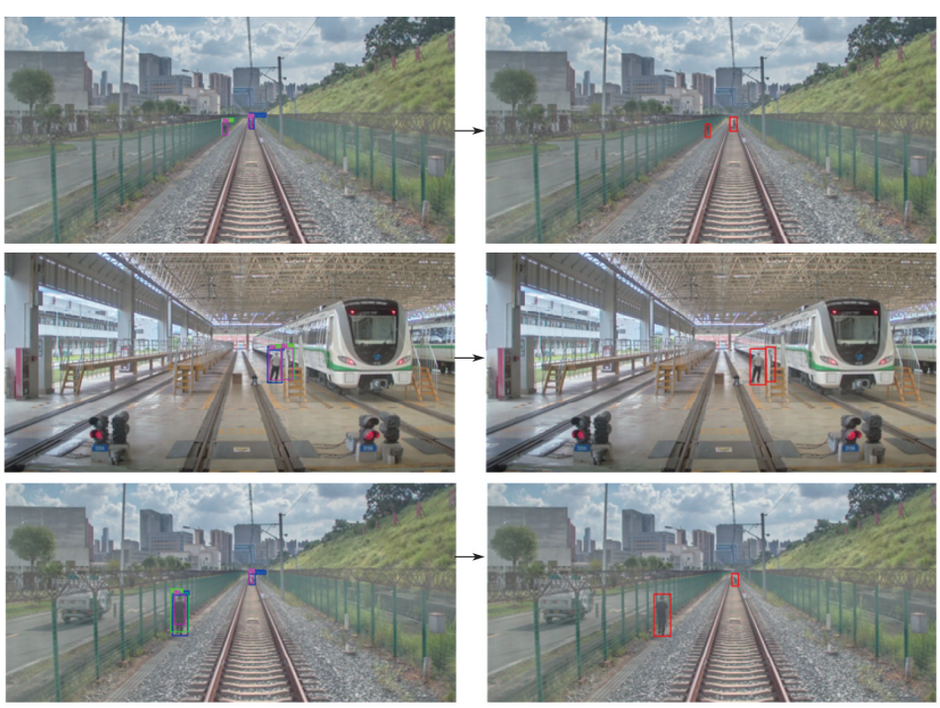
由表 3 可知，对于行人目标检测而言，本文所提融合算法相比于单一传感器数据检测，大幅提高了目标



—— 真值；　—— 投影值。

图 16　三维点云目标检测结果在图像上的投影

Fig. 16　Projection of 3D point cloud object detection results onto images



(a)融合前　　(b)融合后

—— 真值；　—— 图像检测值；　—— 点云检测投影值；　—— 输出值。

图 17　检测结果融合

Fig. 17　Fusion of detection results

表3　基于三维点云与图像融合的行人检测与单一传感器数据检测结果对比

Table 3　Comparison of pedestrian detection results between fusion of 3D point clouds and images and detection with single sensor data

指标	YOLOv8模型	SE-SSD模型	本文融合方法
真实标签数量	400	400	400
T_P	377	380	398
F_P	3	8	11
P/%	99.21	97.94	97.31
R/%	94.49	95.00	99.50
推理时间/ms	16	22	24

检测召回率，对于降低现场由于行人漏检而发生的安全事故具有重要意义。

4　结束语

在列车运营过程中，站点、铁路沿线等场景的行人侵限时有发生，往往带来严重的安全隐患，对轨道交通场景的行人进行自动准确的检测对保障列车安全运营具有重要意义。利用单一传感器数据进行行人检测存在着各种不足，因此本文在三维点云与图像检测结果的基础上，通过激光雷达与相机之间的相对位置关系，设计基于检测结果的融合算法，将三维点云目标检测结果与图像目标检测结果进行融合进而输出最终的行人检测结果。试验结果表明，与单一传感器数据检测结果相比，本文所提算法大幅降低了行人的漏检率，在精度相差不大的情况下召回率显著提升，能够为轨道交通场景的行人自动准确检测提供较高的可靠性，从而减少事故发生率。

本文对三维点云与图像检测结果进行融合，虽然大幅提高了检测结果的召回率，但是检测精度略有下降，后期将围绕图像增强（图像去雾、去雨和夜间图像增强等）、点云增强（目标区域点云缺失补齐等）以及融合网络的设计等方面开展工作，以提高单一传感器数据的检测精度和最后的融合精度。

参考文献：

[1] 于晓英, 苏宏升, 姜泽, 等. 基于YOLO的铁路侵限异物检测方法[J]. 兰州交通大学学报, 2020, 39(2): 37-42.
YU Xiaoying, SU Hongsheng, JIANG Ze, et al. Detection method of railway intruding obstacle based on YOLO algorithm[J]. Journal of Lanzhou jiaotong university, 2020, 39(2): 37-42.

[2] 刘力, 苟军年. 基于YOLO v4的铁道侵限障碍物检测方法研究[J]. 铁道科学与工程学报, 2022, 19(2): 528-536.
LIU Li, GOU Junnian. Research on detection method of railway intrusion obstacles based on YOLO v4[J]. Journal of railway science and engineering, 2022, 19(2): 528-536.

[3] 王瑞, 李霄峰, 史天运, 等. 基于视频深度学习的铁路周界入侵检测算法研究[J]. 交通运输系统工程与信息, 2020, 20(2): 61-68.
WANG Rui, LI Xiaofeng, SHI Tianyun, et al. Railway perimeter intrusion detection algorithms based on video deep learning[J]. Journal of transportation systems engineering and information technology, 2020, 20(2): 61-68.

[4] 徐岩, 陶慧青, 虎丽丽. 基于Faster R-CNN网络模型的铁路异物侵限检测算法研究[J]. 铁道学报, 2020, 42(5): 91-98.
XU Yan, TAO Huiqing, HU Lili. Railway foreign body intrusion detection based on faster R-CNN network model[J]. Journal of the China railway society, 2020, 42(5): 91-98.

[5] SCHECHNER Y Y, NARASIMHAN S G, NAYAR S K. Instant dehazing of images using polarization[C]//IEEE. Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Kauai, HI: IEEE, 2001.

[6] SHEN Linghao, ZHAO Yongqiang, PENG Qunnie, et al. An iterative image dehazing method with polarization[J]. IEEE transactions on multimedia, 2019, 21(5): 1093-1107.

[7] YU Long, GAO Shibin, ZHANG Dongkai, et al. A survey on automatic inspections of overhead contact lines by computer vision[J]. IEEE transactions on intelligent transportation systems, 2022, 23(8): 10104-10125.

[8] XU Jing, ZHAO Wei, LIU Peng, et al. An improved guidance image based method to remove rain and snow in a single image[J]. Computer and information science, 2012, 5(3): 49-55.

[9] XU Jing, ZHAO Wei, LIU Peng, et al. Removing rain and snow in a single image using guided filter[C]//IEEE. 2012 IEEE International Conference on Computer Science and Automation Engineering (CSAE). Zhangjiajie: IEEE, 2012: 304-307.

[10] SARNES B. Inspection of contact wire position and thickness at any speed[J]. Eb-elektrische bahnen, 2001, 99(12): 490-495.

[11] 刘邦繁, 张慧源, 李晨, 等. 基于车载激光雷达的机车可通行区域检测[J]. 控制与信息技术, 2021(4): 78-82.
LIU Bangfan, ZHANG Huiyuan, LI Chen, et al. Locomotive passable area detection based on onboard LiDAR[J]. Control and information technology, 2021(4): 78-82.

[12] 李志南, 刘玮, 韩丽娟, 等. 基于二维LiDAR的轨道侵入异物检测系统与方法研究[J]. 光电子·激光, 2020, 31(3): 262-268.
LI Zhinan, LIU Wei, HAN Lijuan, et al. Research on track invasion foreign object detection system and method based on two-dimensional LiDAR[J]. Journal of optoelectronics·laser, 2020, 31(3): 262-268.

[13] 郭保青, 余祖俊, 张楠, 等. 铁路场景三维点云分割与分类识别算法[J]. 仪器仪表学报, 2017, 38(9): 2103-2111.
GUO Baoqing, YU Zujun, ZHANG Nan, et al. 3D point

cloud segmentation, classification and recognition algorithm of railway scene[J]. Chinese journal of scientific instrument, 2017, 38(9): 2103-2111.

[14] 曾祥, 蒋国涛, 鲍纪宇, 等. 基于车载激光雷达的隧道内障碍物检测[J]. 控制与信息技术, 2021(1): 1-8.
ZENG Xiang, JIANG Guotao, BAO Jiyu, et al. Detection of obstacles in tunnel based on vehicle-borne LiDAR[J]. Control and information technology, 2021(1): 1-8.

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

[16] YAN Yan, MAO Yuxing, LI Bo. Second: sparsely embedded convolutional detection[J]. Sensors, 2018, 18(10): 3337.

[17] CHARLES R Q, SU Hao, KAICHUN Mo, et al. PointNet: deep learning on point sets for 3D classification and segmentation[C]//IEEE. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, HI: IEEE, 2017: 77-85.

[18] 何友, 彭应宁, 陆大. 多传感器数据融合模型综述[J]. 清华大学学报(自然科学版), 1996(9): 14-20.
HE You, PENG Yingning, LU Da. A review: multi-sensor data fusion model[J]. Journal of Tsinghua university: natural science edition, 1996(9): 14-20.

[19] 李梦洁. 基于激光点云与图像结合的行人检测与跟踪技术研究[D]. 长沙: 国防科技大学, 2017.
LI Mengjie. The research of pedestrian detecting and tracking technology based on the combination of laser points and image[D]. Changsha: National Defense University, 2017.

[20] 郑少武, 李巍华, 胡坚耀. 基于激光点云与图像信息融合的交通环境车辆检测[J]. 仪器仪表学报, 2019, 40(12): 143-151.
ZHENG Shaowu, LI Weihua, HU Jianyao. Vehicle detection in the traffic environment based on the fusion of laser point cloud and image information[J]. Chinese journal of scientific instrument, 2019, 40(12): 143-151.

[21] YIN Xuanyu, SASAKI Y, WANG Weimin, et al. 3D object detection method based on YOLO and K-Means for image and point clouds[DB/OL]. (2020-04-21) [2024-01-16]. https://arxiv.org/abs/2005.02132.

[22] LIANG Ming, YANG Bin, CHEN Yun, et al. Multi-task multi-sensor fusion for 3D object detection[C]//IEEE. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, CA: IEEE, 2019: 7345-7353.

[23] 陈毅, 张帅, 汪贵平. 基于激光雷达和摄像头信息融合的车辆检测算法[J]. 机械与电子, 2020, 38(1): 52-56.
CHEN Yi, ZHANG Shuai, WANG Guiping. Vehicle detection algorithm based on information fusion of LIDAR and camera[J]. Mechanical and electronics, 2020, 38(1): 52-56.

Pedestrian detection method in rail transit scenes based on fusion of 3D point clouds and images

HE Jia

(CHN ENERGY Baoshen Railway Co., Ltd., Baotou，Inner Mongolia 041010, China)

Abstract: Safety incidents caused by pedestrians illegally intruding onto railway tracks occur frequently in rail transit scenes, significantly affecting the safe operation of trains. Utilizing single sensor data for pedestrian detection often leads to low recall rates of detection results and lack of category or orientation information in the results, which cannot meet practical field requirements. To address these issues, this paper proposed a pedestrian detection method based on the fusion of 3D point clouds and images in rail transit scenes. This method first employed a deep learning model trained on a constructed dataset of pedestrian data in rail transit scenes to detect pedestrians separately in 3D point clouds and images. Subsequently, based on the principle of spatial position consistency of the targets in 3D point clouds and images, the rotation and translation matrix between the LiDAR and camera was solved. Finally, the 3D point cloud object detection results were projected onto the image coordinate system. To solve the issues of misalignment between multiple adjacent targets and the one-to-many relationship between detection results, the intersection over union ratio and center point distance between the two detection results were calculated as fusion constraints, enabling more accurate pedestrian detection. Experimental results using data acquired from the field demonstrate that, compared to detection results from separate data of 3D point clouds and images, while maintaining timeliness, this method improves the recall rate by 4.5% and 5.5%, respectively, effectively reducing the risk of safety accidents caused by missed pedestrian detections, meeting the demand for pedestrian detection during actual train operations.

Keywords: 3D point clouds; image detection; deep learning; multi-sensor fusion; object detection

收稿日期：2023-05-05；修回日期：2024-02-06