

# 光电工程

## Opto-Electronic Engineering

中文核心期刊 中国科技核心期刊  
Scopus CSCD

### 图像引导和点云空间约束的公路洒落物检测定位方法

蔡怀宇, 杨朝乾, 崔子扬, 汪毅, 陈晓冬

#### 引用本文:

蔡怀宇, 杨朝乾, 崔子扬, 等. 图像引导和点云空间约束的公路洒落物检测定位方法[J]. 光电工程, 2024, 51(3): 230317.

Cai H Y, Yang Z Q, Cui Z Y, et al. Image-guided and point cloud space-constrained method for detection and localization of abandoned objects on the road[J]. *Opto-Electron Eng*, 2024, 51(3): 230317.

<https://doi.org/10.12086/oe.2024.230317>

收稿日期: 2023-12-27; 修改日期: 2024-02-29; 录用日期: 2024-03-01

### 相关论文

#### 基于激光雷达数据的火星表面障碍物识别

陈海平, 李萌阳, 曹庭分, 严寒, 张亮, 张尽力, 王成程

光电工程 2023, 50(2): 220240 doi: [10.12086/oe.2023.220240](https://doi.org/10.12086/oe.2023.220240)

#### 面向道路场景语义分割的移动窗口变换神经网络设计

杭昊, 黄影平, 张栩瑞, 罗鑫

光电工程 2024, 51(1): 230304 doi: [10.12086/oe.2024.230304](https://doi.org/10.12086/oe.2024.230304)

#### 轴注意力引导的锚点分类车道线检测

罗鑫, 黄影平, 梁振明

光电工程 2023, 50(7): 230079 doi: [10.12086/oe.2023.230079](https://doi.org/10.12086/oe.2023.230079)

#### 融合空间掩膜预测与点云投影的多目标跟踪

陆康亮, 薛俊, 陶重桦

光电工程 2022, 49(9): 220024 doi: [10.12086/oe.2022.220024](https://doi.org/10.12086/oe.2022.220024)

更多相关论文见光电期刊集群网站 

 **光电工程**  
Opto-Electronic Engineering

<http://cn.ojournal.org/oe>



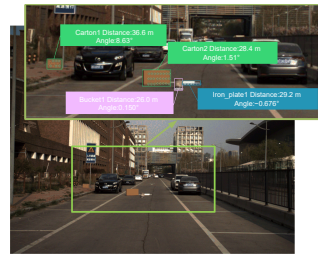
 OE\_Journal



Website

DOI: 10.12086/oe.2024.230317

## 图像引导和点云空间约束的公路洒落物检测定位方法

蔡怀宇<sup>1,2</sup>, 杨朝乾<sup>1,2</sup>, 崔子扬<sup>1,2</sup>, 汪毅<sup>1,2</sup>, 陈晓冬<sup>1,2\*</sup><sup>1</sup>天津大学精密仪器与光电子工程学院, 天津 300072;<sup>2</sup>天津大学光电信息技术教育部重点实验室, 天津 300072

**摘要:** 公路洒落物是影响交通安全的重要因素之一,为了解决中小尺度公路洒落物检测中的漏检、误检以及难以定位等问题,本文提出了一种图像引导和点云空间约束的公路洒落物检测定位方法。该方法使用改进的 YOLOv7-OD 网络处理图像数据获取二维目标预测框信息,将目标预测框投影到激光雷达坐标系下得到锥形感兴趣区域 (region of interest, ROI)。在 ROI 区域内的点云空间约束下,联合点云聚类算法获得不同尺度的洒落物在三维空间中的检测定位结果。实验表明:改进的 YOLOv7-OD 网络在中尺度目标上的召回率和平均精度分别为 85.4% 和 82.0%,相比 YOLOv7 网络分别提升 6.6 和 8.0 个百分点;在小尺度目标上的召回率和平均精度分别为 66.8% 和 57.3%,均提升 5.3 个百分点;洒落物定位方面,对于距离检测车辆 30~40 m 处的目标,深度定位误差为 0.19 m,角度定位误差为 0.082°,实现了多尺度公路洒落物的检测和定位。

**关键词:** 公路洒落物; 图像; 激光雷达点云; 目标检测

中图分类号: TP277

文献标志码: A

蔡怀宇, 杨朝乾, 崔子扬, 等. 图像引导和点云空间约束的公路洒落物检测定位方法 [J]. 光电工程, 2024, 51(3): 230317

Cai H Y, Yang Z Q, Cui Z Y, et al. Image-guided and point cloud space-constrained method for detection and localization of abandoned objects on the road[J]. *Opto-Electron Eng*, 2024, 51(3): 230317

## Image-guided and point cloud space-constrained method for detection and localization of abandoned objects on the road

Cai Huaiyu<sup>1,2</sup>, Yang Zhaoqian<sup>1,2</sup>, Cui Ziyang<sup>1,2</sup>, Wang Yi<sup>1,2</sup>, Chen Xiaodong<sup>1,2\*</sup><sup>1</sup>School of Precision Instrument and Opto-electronics Engineering, Tianjin University, Tianjin 300072, China;<sup>2</sup>Key Laboratory of Optoelectronic Information Technology Ministry of Education, Tianjin University, Tianjin 300072, China

**Abstract:** Abandoned objects on the road significantly impact traffic safety. To address issues such as missed detections, false alarms, and localization difficulties encountered in detecting of small-to-medium-sized abandoned objects, this paper proposes a method for detecting and locating abandoned objects on the road using image guidance and point cloud spatial constraints. The method employs an improved YOLOv7-OD network to process image data, extracting information about two-dimensional target bounding boxes. Subsequently, these bounding boxes are projected onto the coordinate system of the LiDAR sensor to get a pyramidal region of interest (ROI).

收稿日期: 2023-12-27; 修回日期: 2024-02-29; 录用日期: 2024-03-01

\*通信作者: 陈晓冬, xdchen@tju.edu.cn。

版权所有©2024 中国科学院光电技术研究所

Under the spatial constraints of the point cloud within the ROI, the detection and localization results of abandoned objects on the road in three-dimensional space are obtained through a combination of point cloud clustering and point cloud generation algorithms. The experimental results show that the improved YOLOv7-OD network achieves recall and average precision rates of 85.4% and 82.0%, respectively, for medium-sized objects, representing an improvement of 6.6% and 8.0% compared to the YOLOv7. The recall and average precision rates for small-sized objects are 66.8% and 57.3%, respectively, with an increase of 5.3%. Regarding localization, for targets located 30-40 m away from the detecting vehicle, the depth localization error is 0.19 m, and the angular localization error is 0.082°, enabling the detection and localization of multi-scale abandoned objects on the road.

**Keywords:** abandoned objects on the road; image; LiDAR point cloud; object detection

## 1 引言

公路是国家重要的经济命脉。随着公路里程和车流量的不断提高,公路日常养护工作的重要性愈发凸显。公路洒落物清理是公路养护的主要任务之一。如果对洒落物没有及时清理,则容易引发交通拥堵甚至造成事故。目前洒落物检测主要采用人工巡检的方式进行但工作效率低且存在安全隐患。因此,智能化公路洒落物检测技术的研发对保障交通安全,提高公路养护效率具有重要意义。

公路洒落物的检测定位属于特定对象的目标检测任务。相机和激光雷达是目标检测任务中常用的两类传感器。根据使用的传感器不同可以将常见的目标检测任务分为基于图像、点云和多传感器融合等三类。基于图像的算法根据原理不同划分为一阶段和二阶段算法。其中,一阶段算法的典型代表有 SSD<sup>[1]</sup>、RetinaNet<sup>[2]</sup>、DETR<sup>[3]</sup>、YOLO 系列<sup>[4-6]</sup>等目标检测网络;二阶段算法的典型代表有 Faster R-CNN<sup>[7]</sup>、Mask RCNN<sup>[8]</sup>等。金瑶等人<sup>[9]</sup>就是使用基于 YOLOv3 改进的目标检测网络实现了对城市道路视频中洒落物目标的检测。图像算法可以利用丰富的图像纹理、色彩等信息提取到更多特征从而提高检测精度,但仅依赖图像算法难以对检测目标实现三维空间中的精确定位。基于点云的算法主要根据点云团簇的空间分布实现三维目标检测,典型的算法有 PointNet<sup>[10]</sup>、PointNet++<sup>[11]</sup>、PointNeXt<sup>[12]</sup>等。此类算法能够直接在三维空间中实现对目标的检测定位,但是由于缺乏纹理信息,检测精度不如图像算法,并且点云数据较为稀疏,当目标较远或尺度较小时,检测效果将显著降低。为了充分结合图像和点云的优势,基于多传感器融合的解决方案成为目前的研究热点。Qi 等人<sup>[13]</sup>先使用 mask-RCNN 网络在图像上找到目标的预测框,并将其投影

到激光雷达坐标系下获得 ROI 区域,再使用 PointNet++ 进行三维预测框的回归,实现对空间中目标的检测定位。Vora 等人<sup>[14]</sup>使用语义分割网络获取图像不同区域的语义信息,然后将激光雷达点云投影到图像上进行染色,为点云数据增加颜色信息,然后通过点云检测网络实现目标检测。梁浩林等人<sup>[15]</sup>使用激光雷达点云聚类得到目标点云团簇,并将其投影到图像平面生成 ROI 区域,再通过分类网络进行类别预测来实现对公路洒落物的检测和定位。以上方法虽然融合了图像和点云的优势,在检测精度上获得了一定的提升,但是并没有克服点云的稀疏性所带来的小目标漏检问题。针对点云稀疏问题,郑欣悦等人<sup>[16]</sup>将单目深度估计的结果与激光雷达点云数据融合获得稠密深度图,提高了对小目标的定位能力,但是该方法十分依赖单目深度估计网络的准确性。范婉舒等人<sup>[17]</sup>使用多帧点云数据来推算通过图像数据检出的小尺度目标的空间位置,但是该方法的实时性较差。

综上所述,基于图像和点云数据融合的算法是目标检测定位任务的常用手段,但考虑到公路洒落物大多数属于中小尺度目标,如何克服点云稀疏性的影响仍需进一步探究。为了解决以上问题,本文提出一种图像引导和点云空间约束的公路洒落物检测定位方法:首先在 YOLOv7 目标检测网络的基础上增加小目标检测层并引入通道注意力机制,增强网络对图像内中小尺度洒落物的二维预测框的提取能力。然后在预测框的引导下,联合点云聚类和点云生成的定位算法在点云空间约束下对洒落物进行定位。通过实验验证,所提方法可有效克服激光雷达点云的稀疏性,实现对多尺度洒落物的检测和定位功能。

## 2 方法原理

与通用的目标检测任务不同,公路洒落物检测任

务的目标尺度一般小于道路上的车辆、行人等, 多属于中小尺度目标, 在点云稀疏性的影响下, 直接使用点云数据进行检测定位难免出现漏检。因此, 本文所提方法的主要思想是利用目标检测网络对中小尺度目标更强的检测能力将其从场景中检测出来, 然后根据坐标系之间的旋转变换关系指导点云数据进行定位。主要操作如下: 首先利用目标检测网络对图像数据进行检测获得二维目标预测框, 然后将预测框投影到激光雷达坐标系下生成 ROI 区域。对于较大目标而言, ROI 区域内有足够的点云数据, 通过聚类即可得到目标的三维空间位置; 对于较小目标而言, ROI 区域内没有足够的点云进行聚类, 则通过地面点云提供的高度约束, 根据目标框投影变换关系生成对应的点云团簇数据来获取目标的空间位置信息, 进而实现在三维空间中对多尺度公路洒落物的检测定位。系统整体框架如图 1 所示。

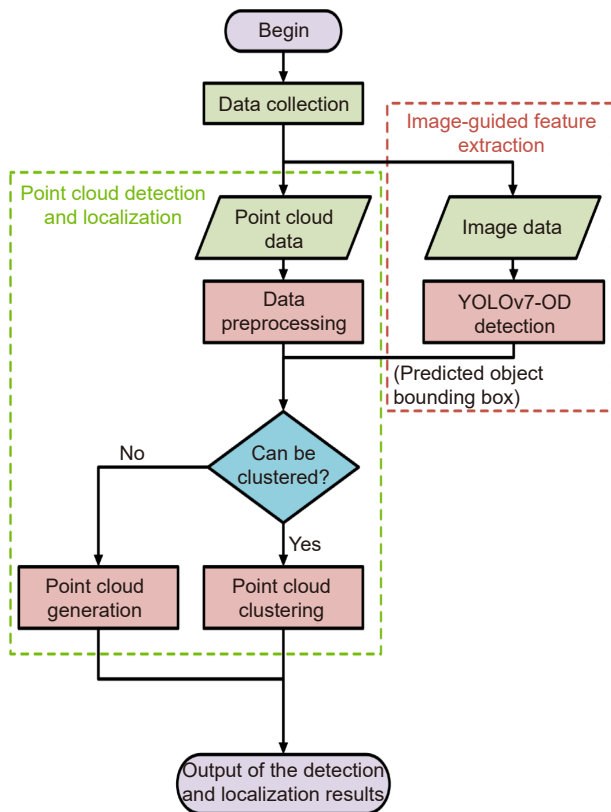


图 1 所提方法的总体框架

Fig. 1 The overall framework of the proposed method

### 2.1 通过图像获取二维目标预测框

公路洒落物在采集到的图像数据中像素占比通常较小, 属于中小目标。因此, 为了强化网络对公路洒落物的检测能力, 本文在 YOLOv7 目标检测网络<sup>[18]</sup>

基础上进行改进, 增加了小目标检测层 (small object detection layer, SOD Layer) 和基于空洞卷积的感受野自适应注意力机制 (selective dilated kernel attention, SDK Attention), 提出了 YOLOv7-OD (obstacle detection) 目标检测网络, 增强对图像中小目标特征的提取能力以获得更加精准的二维目标预测框作为图像引导信息, 网络结构如图 2 所示。

SOD Layer 网络结构: 原 YOLOv7 网络 Neck 部分采用 FPN (Feature Pyramid Network) + PAN (Path Aggregation Network) 结构, 将 Backbone 部分提取到的不同分辨率特征进行融合, 提高检测结果的召回率和目标框精度。一般而言, 随着特征提取的网络深度增加, 特征的语义信息将更加丰富, 但空间信息会由于分辨率的下降而丢失。网络模型下采样次数越多, 特征图分辨率越低, 则小目标特征信息丢失就越多。要想提高网络对小目标的检测能力, 就必须加强对浅层特征利用。但浅层特征信息量大, 单一的引入容易导致网络的臃肿。因此, 本文提出了 SOD Layer 结构, 如图 3 所示。在该结构中, 将 Backbone 部分下采样四倍得到的特征层添加到 FPN + PAN 结构中进行特征融合, 在保持网络深度不变的同时, 引入更多浅层特征信息用于增强小目标检测。并且, 由于检测任务以中小目标为主, 在正负样本匹配策略影响下, 网络结构中分辨率最小的预测层对于总体预测结果的贡献相对较小。本文在浅层特征处加入小目标预测头并替代原有深层特征的预测头, 在提高模型对中小尺度目标识别能力的同时, 进一步精简网络结构, 降低参数量。

增加 SDK Attention 模块: 浅层特征的引入给网络带来了丰富的位置、细节信息, 但也导致了噪声数据增加和语义信息缺乏等问题。为了减少噪声数据的干扰, 增强语义信息, 本文提出 SDK Attention 模块, 网络结构见图 4。设  $F_m$  为网络不同阶段得到的特征数据,  $\oplus$  为特征数据在维度上的直接拼接,  $\otimes$  为权重与各层特征数据相乘。设空洞卷积操作为  $DC(\cdot)$ , 全连接层操作为  $FC(\cdot)$ 。对于输入特征  $F$ , SDK Attention 模块为特征  $F$  添加不同感受野注意力权重, 从而获得加权特征  $F_w$  的过程如式 (1)~(2):

$$F_{fuse} = DC_{dilation=1}(F) \oplus DC_{dilation=2}(F), \quad (1)$$

$$F_w = (FC_1(GlobalPool(F_{fuse})) \otimes F_1) + (FC_2(GlobalPool(F_{fuse})) \otimes F_2). \quad (2)$$

SDK Attention 模块使用不同空洞率的空洞卷积分别对输入的特征进行提取, 为网络引入不同感受野

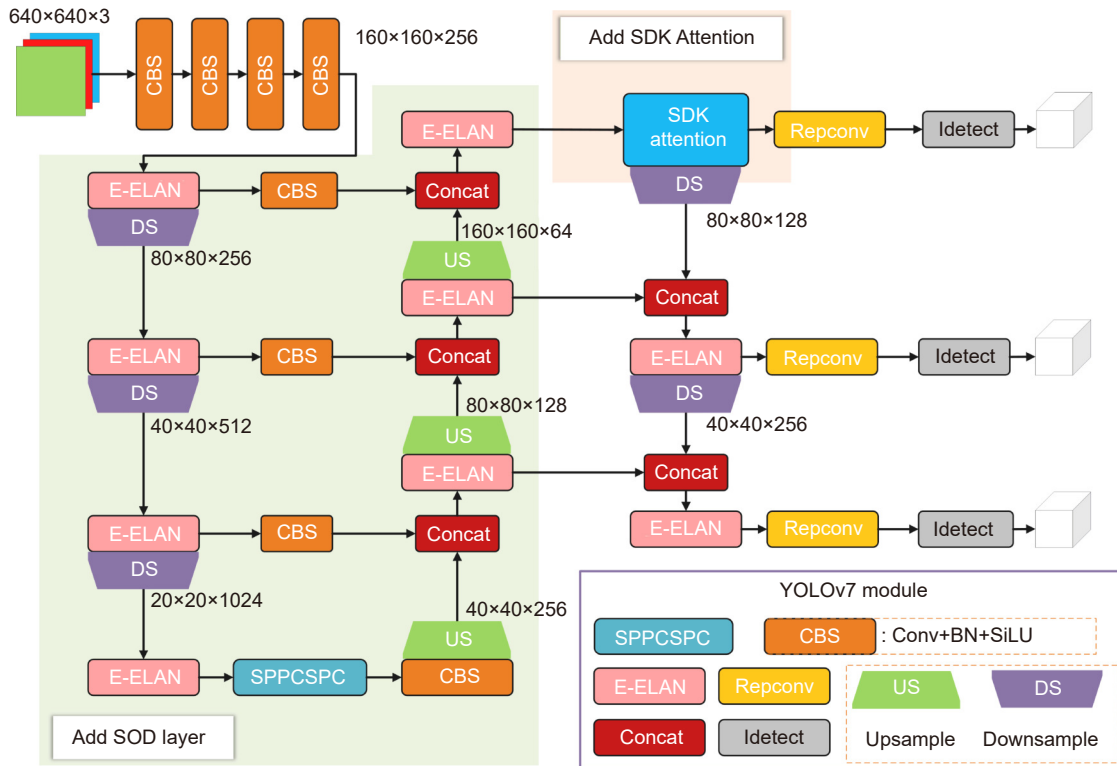


图 2 YOLOv7-OD 网络结构  
Fig. 2 The network architecture of YOLOv7-OD

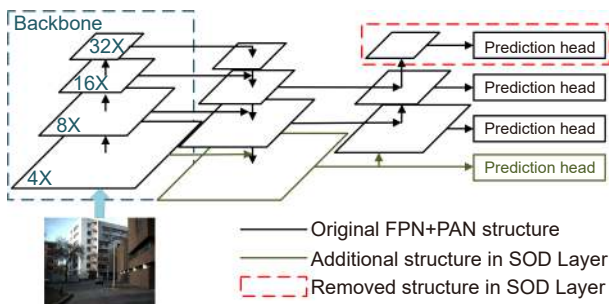


图 3 SOD Layer 网络结构  
Fig. 3 The network structure of SOD Layer

特征。将获得的不同感受野特征送入通道注意力模块中, 分别计算其各通道注意力权重。然后通过训练, 不断反馈优化通道自主选择方案, 最终实现网络自主选择加强有效特征或降低噪声特征通道权重。最后, 通过加权求和的方式将通道权重作用到对应特征上, 使网络更加关注对当前任务有用的通道信息, 进而实现动态选择机制, 提升模型性能。SDK Attention 模块位于 FPN 和 PAN 结构转换处, 主要目的是在浅层特征数据引入的第一时间就对其进行处理, 加强语义信息并降低噪声干扰, 为后续检测提供更加高质量的特征信息。

通过 YOLOv7-OD 网络对图像数据进行处理, 得到图像中的公路洒落物目标框信息, 将其作为图像引

导信息用以辅助点云进行定位。

### 2.2 预测框引导下点云定位算法

由于激光雷达点云存在稀疏性, 对于远距离中小尺度目标而言, 没有充足的点云数据描述其空间特征。因此单一点云检测方法无法对此类目标进行检测而出现漏检、误检, 也就难以应用到中小尺度洒落物的检测定位中。为了解决这一问题, 本文在图像预测框的引导下, 采用点云聚类 and 点云生成方法在点云数据提供的空间约束下进行联合定位, 有效提升了对中小尺度洒落物的检测定位效果。

首先, 对点云数据进行预处理。将激光雷达和相机进行视场匹配, 二者重叠的视场区域为算法的工作区域, 剔除工作区域外的数据以降低数据量。再通过布料模拟滤波算法 (cloth simulation filter, CSF)<sup>[19]</sup> 进行点云滤波, 提取地面和非地面点云作为空间约束信息。然后根据相机小孔成像模型, 将目标预测框角点与相机光心  $O$  点分别相连并向外延伸, 构建出一个以  $O$  点为顶点的锥形 ROI 区域, 如图 5 所示。随后, 利用图像坐标系和激光雷达坐标系之间的旋转变换矩阵, 将该锥形 ROI 区域映射到激光雷达坐标系中。由于二者之间为刚性的旋转平移变换, ROI 区域在激光雷

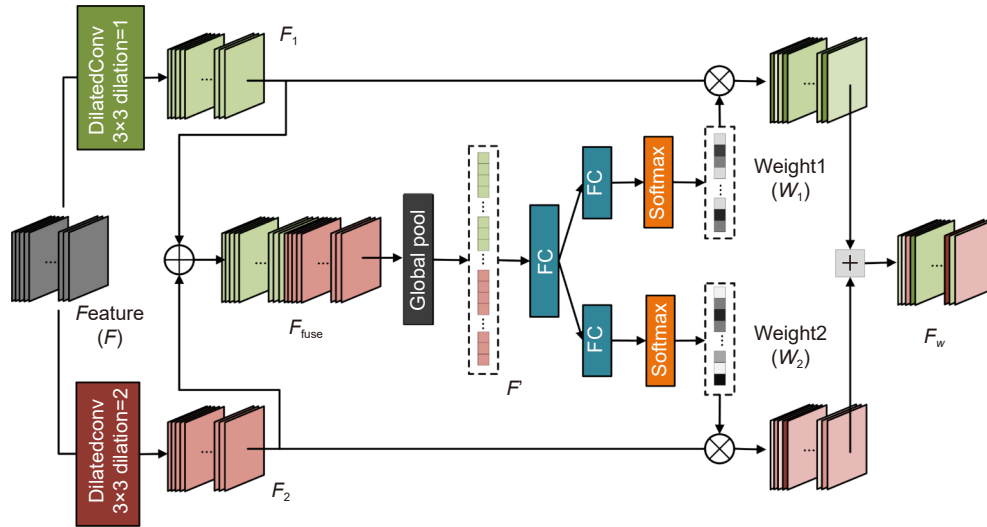


图 4 SDK Attention 模块  
Fig. 4 SDK Attention module

达坐标系下依然为锥形结构, 可视作目标检测框在激光雷达坐标系下的投影区域。

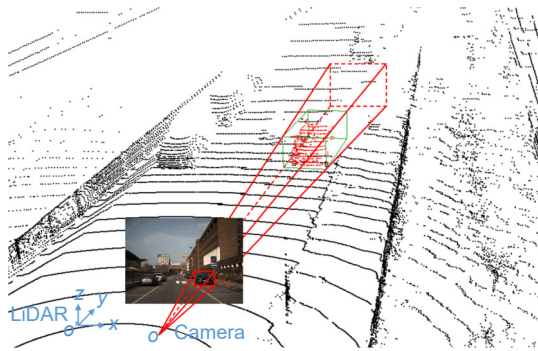


图 5 根据图像目标检测框生成激光雷达坐标系下的 ROI 区域

Fig. 5 Generation of ROI areas in the LiDAR coordinate system based on image object detection bounding boxes

设处于 ROI 区域内的点云集合为  $PC_{ROI}$ , 待求目标点云集合为  $PC_{Obstacle}$ 。考虑激光雷达和相机探测维度的差异以及目标遮挡等情况, 三维空间中目标周围的其他点云数据也可能落入相同区域内。此类数据则被视为噪声点云。因此,  $PC_{Obstacle} \subseteq PC_{ROI}$ 。使用 DBSCAN 聚类算法对 ROI 区域内的点云进行聚类。如果聚类结果中有满足点云数量大于等于 3 的点云团簇时, 聚类有效; 如果聚类结果中存在多个点云团簇, 则将聚类中心最靠近 ROI 区域中心轴线且离检测车辆更近的点云团簇视作目标点云数据  $PC_{Obstacle}$ , 赋予预测框的类别信息。该方法不但提高了聚类速度, 还降低了参数和噪声点对聚类效果的影响, 得到的点云团簇在空间中的位置信息即为该洒落物的真实位置数据。

如果聚类结果中没有满足条件的点云团簇, 说明待测洒落物尺度较小或距离激光雷达较远, 此时需要采用联合点云生成算法在点云空间约束下进行补充定位。在公路洒落物检测定位任务中, 已知待检测的公路洒落物均位于路面上, 且公路路面的变化较为平缓。洒落物底部距激光雷达的高度差和其周围地面距激光雷达的高度差基本相同。以此为先验条件, 点云生成算法借助目标框底部地面点云提供的高程信息对目标框底边的三维数据进行恢复。设待测洒落物的目标框底边上的点在图像上坐标为  $[U, V]$ , 在激光雷达坐标系中的坐标为  $[X_L, Y_L, Z_L]$ ; 空间中该点到相机光心的深度为  $d$ , 则根据激光雷达坐标系到像素坐标系下的转换关系<sup>[20]</sup>可得:

$$d \begin{bmatrix} U \\ V \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{f}{dx} & 0 & u_0 & 0 \\ 0 & \frac{f}{dy} & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} R_{11} & R_{12} & R_{13} & T_{14} \\ R_{21} & R_{22} & R_{23} & T_{24} \\ R_{31} & R_{32} & R_{33} & T_{34} \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} X_L \\ Y_L \\ Z_L \\ 1 \end{bmatrix}, \quad (3)$$

其中,  $R_{mn}$ 代表旋转分量,  $T_{mn}$ 代表平移分量,  $m, n$ 分别为元素在矩阵中的坐标序号,  $f$ 为相机焦距,  $dx, dy$ 分别为像素平面  $x$  轴、 $y$  轴方向单位像素宽度,  $u_0, v_0$ 为相机光心在像素坐标系中的坐标。以上参数均可通过传感器标定得到。对目标框底边周围的地面点云求高度平均值和深度平均值, 分别设为  $Z_1$ 和  $X_1$ 。考虑装配误差, 激光雷达安装完成后, 其坐标系  $xoy$  平面

与现实水平平面夹角为 $\theta_1$ 。得到目标框底边高度的修正方程为

$$Z_L = Z_1 - (X_1 - X_L) \tan \theta_1. \quad (4)$$

联立方程式 (3) 和 (4) 求解即可得到图像上二维点  $[U, V]$  在激光雷达坐标系下的三维数据  $[X_L, Y_L, Z_L]$ 。根据目标检测框底边的像素坐标, 对洒落物底部三维信息逐一恢复, 生成对应目标在激光雷达坐标系下的点云团簇。

点云生成算法利用目标框周围地面点云的空间约束, 根据目标底部的二维坐标恢复其空间三维信息, 最后生成目标底部的点云团簇来实现定位, 改善了稀疏点云难以对中小尺度洒落物进行定位的问题。综上, 在图像目标框的引导和点云空间约束下, 通过联用点云聚类 and 点云生成算法实现了对多尺度公路洒落物目标的检测定位。

### 3 实验结果与分析

#### 3.1 实验系统设计

本文以电动代步车为载体, 安装 Velodyne VLP-32C 激光雷达、大华 A5131CU210 相机、工控机和 GPS 等设备搭建公路洒落物检测实验系统, 如图 6 所示。系统搭建完成后, 对激光雷达和相机进行联合标定, 得到激光雷达坐标系到相机坐标系的旋转变换矩阵  $RT$  和相机的内参矩阵  $M$ , 如式 (5) 所示。

$$RT = \begin{bmatrix} 0.0110 & -1.00 & -0.00262 & -0.211 \\ 0.137 & 0.00410 & -0.991 & -0.404 \\ 0.991 & 0.0105 & 0.137 & -0.553 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

$$M = \begin{bmatrix} 1655 & 0 & 590.7 \\ 0 & 1655 & 504.1 \\ 0 & 0 & 1 \end{bmatrix}. \quad (5)$$

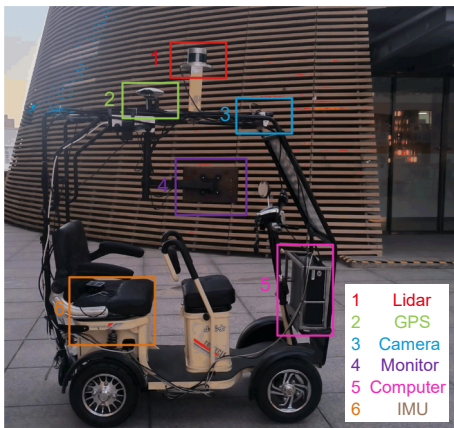


图 6 实验装置结构图

Fig. 6 Structural diagram of the experimental device

#### 3.2 实验数据集

由于缺乏丰富的公路洒落物开源数据集, 本文采用迁移学习的方法训练目标检测网络, 其中源域数据集采用无人驾驶公司 Waymo 的开源数据集 (Waymo open dataset, WOD)<sup>[21]</sup>, 目标域数据集采用自建的公路洒落物数据集。

WOD 数据集大部分为复杂的公路环境且含有大量的中小目标数据, 和公路洒落物检测任务的工作场景相似, 有助于进行迁移学习。该数据集覆盖雨天、黄昏、傍晚等多种光照条件, 对网络的鲁棒性提出了更高的要求。

自建数据集是在校园内模拟的公路洒落物场景中, 通过实验平台以每小时 20 km 的速度采集获得。获得原始数据后, 经过传感器时间同步、数据清洗和运动补偿等预处理, 共得到激光雷达点云和对应的图像数据 1218 份。根据实际公路环境下常见的纸箱、金属物、塑料等洒落物材质, 自制数据集共设置 car、people 等 2 类常见目标和 carton、bucket、handcart、iron-plate 和 extinguisher 等 5 类不同尺度的洒落物目标并进行标注。

本文采用 WOD 数据集对深度学习网络进行预训练, 然后使用迁移学习的方法, 将预训练好的网络模型迁移到公路洒落物检测任务上。

#### 3.3 基于图像的目标预测框提取实验

本实验在装备 Intel(R) Core(TM) i7-11700K @ 3.60 GHz, 32 G 内存, NVIDIA GeForce RTX 3090 显卡的计算机平台运行, 使用 Pytorch 深度学习框架搭建 YOLOv7-OD 网络进行训练, 通过消融实验和对比实验分别验证了网络结构的改进效果和整体的提升效果。

##### 3.3.1 网络训练

在训练部分, 网络输入的图像尺寸为  $640 \times 640$ , 批量 batchsize 为 16, 最高迭代轮次 epoch 为 300 次, 使用 Adam 优化器, 初始学习率 lr 为 0.001, 动量 momentum 为 0.937, 权重衰减 weight\_decay 为 0.0005, 热身训练设置为 3 轮。采用图像翻转、饱和度、亮度、色调调整和 mosaic 等方法进行数据增强。

训练过程中首先将 YOLOv7-OD 网络在 WOD 数据集下进行预训练, 得到表现最优的模型作为预训练模型。然后使用自制数据集以预训练模型中的参数作为初始参数对网络进行迁移学习, 最终得到适用于公路洒落物检测任务的网络模型。

### 3.3.2 实验结果

为了验证 YOLOv7-OD 网络中提出的 SOD Layer 和 SDK Attention 模块的有效性, 分别在 WOD 和自制数据集上进行消融实验。实验结果见表 1~3: 其中表 1、表 2 中 AP 为平均精度, AR 为平均召回率, 括号内为目标尺度。设预测框和真值框交并比 (intersection over union, IOU) 的阈值为 THR。mAP0.5 表示所有目标在 THR=50 时的平均精度; mAP0.5:0.95 表示所有目标在不同 THR(从 0.5 到 0.95, 步长为 0.05) 上的平均精度均值; 表中加粗字体为本列最优指标。表 3 中 Params、GFLOPs 和 FPS 分别表示网络的参数量、计算量和每秒处理图片的帧数。

由表 1 和表 2 中的数据可以看出: 单独加入 SOD Layer 模块后, 网络在两个数据集中对中小目标的平均精度 AP 和平均召回率 AR 都得到了明显提升。说明该模块能够更加合理地融合不同尺度的特征, 充分利用浅层特征, 实现网络对中小目标检测能力的提升。单独加入 SDK Attention 模块后, 网络通过自主选择不同感受野的特征, 获得了更加丰富的特征数据, 在 mAP0.5 和 mAP0.5:0.95 上相较于原 YOLOv7 网络

也获得了较大的提升。YOLOv7-OD 网络同时包含了 SOD Layer 和 SDK Attention 两个模块, 在所有对比模型中对中小尺度目标检测的平均精度和平均召回率均取得了最优的检测效果。从整体评价指标 mAP0.5 和 mAP0.5:0.95 来看, 相较于原网络, YOLOv7-OD 在 WOD 数据集上分别提升了 2.31 个百分点和 2.03 个百分点; 在自制数据集上分别提升了 1.1 和 7.1 个百分点, 均为最优检测效果, 证明该网络在对中小尺度目标检测任务上具有突出的检测能力。

由表 3 可以看出, 与原网络相比, YOLOv7-OD 网络在时间复杂度上增加了 19.36%, 帧率下降 19.95%, 但依然保持了每秒 66 帧的处理速度, 满足算法实时性的要求。此外, SOD Layer 网络结构的引入不仅提升了网络小目标检测的效果, 也使得 YOLOv7-OD 网络的参数量下降 27.07%, 降低了算法训练、部署时对硬件设备的要求。

在 Waymo 数据集上, 本文将 YOLOv7-OD 与其他主流目标检测算法的性能进行比较, 实验结果见表 4。表中粗体用以突出显示各性能指标的最佳值; 下划线用以标注各性能指标的次优值。从对比实验中

表 1 WOD 数据集上的消融实验

Table 1 Ablation experiments on the WOD dataset

| YOLOv7 | SOD Layer | SDK Attention | AP/%         |              |              | AP/%         |              |              | mAP0.5 /%    | mAP0.5:0.95 /% |
|--------|-----------|---------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|----------------|
|        |           |               | small        | medium       | large        | small        | medium       | large        |              |                |
| √      |           |               | 10.00        | 35.80        | <b>70.40</b> | 23.20        | 47.20        | <b>75.60</b> | 57.15        | 32.30          |
| √      | √         |               | 11.90        | 38.90        | 66.90        | 26.00        | 49.70        | 73.10        | 59.28        | 34.18          |
| √      |           | √             | 11.20        | 37.60        | 66.80        | 24.30        | 48.70        | 73.00        | 58.12        | 33.21          |
| √      | √         | √             | <b>12.00</b> | <b>39.00</b> | 67.70        | <b>26.20</b> | <b>50.80</b> | 73.20        | <b>59.46</b> | <b>34.33</b>   |

表 2 自制数据集上的消融实验

Table 2 Ablation experiments on custom dataset

| YOLOv7 | SOD Layer | SDK Attention | AP/%         |              |              | AP/%         |              |              | mAP0.5 /%    | mAP0.5:0.95 /% |
|--------|-----------|---------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|----------------|
|        |           |               | small        | medium       | large        | small        | medium       | large        |              |                |
| √      |           |               | 52.00        | 74.00        | 85.30        | 61.50        | 78.80        | 89.00        | 94.20        | 64.80          |
| √      | √         |               | 55.00        | 79.80        | 92.70        | 65.40        | 84.00        | <b>95.50</b> | 94.30        | 69.80          |
| √      |           | √             | 54.10        | 81.50        | <b>93.40</b> | 63.20        | 85.30        | 95.10        | 93.80        | 70.00          |
| √      | √         | √             | <b>57.30</b> | <b>82.00</b> | 92.00        | <b>66.80</b> | <b>85.40</b> | 93.90        | <b>95.30</b> | <b>71.90</b>   |

表 3 不同网络模型的其他评价指标

Table 3 Additional evaluation metrics for different network models

| YOLOv7 | SOD Layer | SDK Attention | Params/MB | GFLOPs | FPS  |
|--------|-----------|---------------|-----------|--------|------|
| √      |           |               | 71.3      | 103.3  | 82.2 |
| √      | √         |               | 51.4      | 108.2  | 73.2 |
| √      |           | √             | 73.7      | 118.3  | 75.4 |
| √      | √         | √             | 52        | 123.3  | 65.8 |



表 4 YOLOv7-OD 与其他目标检测算法的对比实验

Table 4 Comparative experiments of YOLOv7-OD with other object detection algorithms

| Model                      | Params /MB | AP/%  |        |       | AP/%  |        |       | mAP0.5 | mAP0.5:0.95 |
|----------------------------|------------|-------|--------|-------|-------|--------|-------|--------|-------------|
|                            |            | small | medium | large | small | medium | large | /%     | /%          |
| Faster-RCNN <sup>[2]</sup> | 79.0       | 3.70% | 19.0%  | 41.7% | 9.00% | 26.2%  | 46.6% | 29.7%  | 15.6%       |
| RetinaNet <sup>[7]</sup>   | 61.7       | 3.00% | 29.8%  | 64.7% | 10.6% | 40.4%  | 71.4% | 43.9%  | 24.7%       |
| YOLOX <sup>[22]</sup>      | 104        | 7.40% | 31.3%  | 68.1% | 13.9% | 39.2%  | 72.8% | 48.1%  | 28.0%       |
| DETR <sup>[8]</sup>        | 79.0       | 4.83% | 26.4%  | 62.2% | 11.3% | 35.7%  | 68.7% | 45.4%  | 24.0%       |
| YOLOv3 <sup>[5]</sup>      | 119        | 3.00% | 26.2%  | 64.1% | 5.80% | 35.0%  | 69.8% | 41.9%  | 22.9%       |
| YOLOv5 <sup>[23]</sup>     | 88.5       | 6.60% | 31.4%  | 64.1% | 13.3% | 45.5%  | 71.1% | 48.6%  | 27.2%       |
| YOLOv6 <sup>[6]</sup>      | 72.4       | 7.50% | 37.5%  | 71.7% | 16.7% | 46.4%  | 78.0% | 51.9%  | 31.0%       |
| YOLOv7 <sup>[18]</sup>     | 71.3       | 10.0% | 35.8%  | 70.4% | 23.2% | 47.2%  | 75.6% | 57.2%  | 32.3%       |
| YOLOv8 <sup>[24]</sup>     | 83.6       | 8.30% | 38.9%  | 71.3% | 17.7% | 46.7%  | 76.5% | 53.0%  | 32.0%       |
| YOLOv7-OD                  | 52.1       | 12.0% | 39.0%  | 67.7% | 26.2% | 50.8%  | 73.2% | 59.5%  | 34.3%       |

可以看出, YOLOv7 在中小尺寸目标的检测性能上超越了其他主流目标检测网络, 并且在综合评价指标 mAP0.5 以及 mAP0.5:0.95 上显示出明显的优势。基于此, 本文选择将 YOLOv7 作为基准模型, 并在此基础上改进提出 YOLOv7-OD 网络。YOLOv7-OD 在中小目标检测性能上均优于对比算法, 综合评价指标 mAP0.5 以及 mAP0.5:0.95 相较于次优值分别提升 2.3 和 2 个百分点, 并且模型参数量比次优值降低 15.6%。

### 3.4 预测框引导下的点云定位算法实验

#### 3.4.1 点云预处理

点云预处理主要是为了降低算法数据量, 获取地面点云和非地面点云。首先根据点云深度信息进行滤波, 去除所有深度值为负的点云数据。然后根据相机视场和激光雷达视场相匹配, 删除匹配区域外的无关点云数据。最后使用 CSF 滤波算法对点云数据进行处理, 分割为地面点云和非地面点云。滤波结果如图 7 所示。

#### 3.4.2 点云定位结果

为了验证点云定位算法的有效性, 本文从自制数据集集中随机抽取数据样本进行洒落物的定位实验。根据目标洒落物的召回率 Recall、在空间中距离车辆中心的深度  $D$ 、在水平方向的宽度  $W$  和洒落物中心到车辆中心连线与车辆行进方向的夹角  $\theta$  等指标对定位结果进行评价, 部分参数定义如图 8 所示。目标定位的真实值通过人工手动选点的方式从点云数据中进行筛选和计算获得。DBSCAN 聚类算法中领域半径  $\epsilon$  设为 1, 核心对象在领域半径内的最少点数  $\text{MinPts}$  设

为 3。根据 YOLOv7-OD 提取的目标检测框划定 ROI 区域, 在 ROI 区域内使用 DBSCAN 算法对点云进行聚类。如果没有有效聚类结果, 说明目标较小或距离较远导致没有充足的点云数据, 改用点云生成算法进行补充定位。

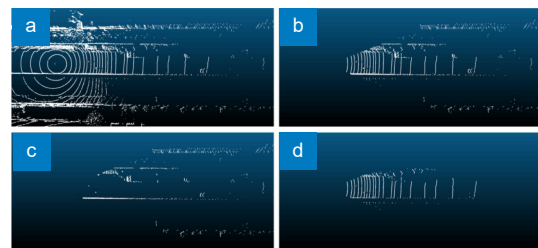


图 7 滤波结果图。(a) 原始点云数据; (b) 视场匹配获得有效点云数据; (c) CSF 滤波后的非地面点云; (d) CSF 滤波后的地面点云

Fig. 7 Filtering results. (a) Original point cloud data; (b) Effective point cloud data obtained by field-of-view matching; (c) Non-ground point cloud after CSF filtering; (d) Ground point cloud after CSF filtering

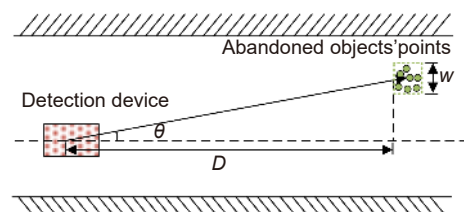


图 8 各评价指标在空间中的具体含义

Fig. 8 The specific meanings of various evaluation metrics in space

本文所提出的预测框引导下点云聚类 and 点云生成联合定位算法与梁浩林等人通过点云聚类实现洒落物检测定位的方法 (Method A)<sup>[15]</sup> 相比较, 结果见表 5。

表 5 两种点云定位方法的实验结果

Table 5 Experimental results of two point cloud localization methods

| Method   | $N$ (road objects) | $N$ (abandoned object) | $N$ (predicted) | $N$ (true) | $N$ (false) | Precision/% | Recall/% |
|----------|--------------------|------------------------|-----------------|------------|-------------|-------------|----------|
| Method A | 920                | 270                    | 150             | 147        | 3           | 98.00       | 55.56    |
| Ours     | 920                | 270                    | 258             | 250        | 8           | 96.90       | 95.56    |

由于远距离处中小尺度目标上点云数量较少, 无法直接通过点云聚类得到目标信息。因此, Method A 在自制数据集上只能达到 55.56% 的召回率。本文所提出的预测框引导下点云定位算法, 联合点云聚类 and 点云生成, 平均精度达 96.90%, 召回率达 95.56%, 额外提升 40% 的召回率, 实现了对中小尺度目标的补充定位。

图 9、10 中 Distance 和 Angle 分别为深度  $D$  和夹角  $\theta$  的测量结果, 根据可视化定位结果显示: 对于较大尺度目标, 两种方法检测结果基本相同; 但对于中

小尺度目标, 与 Method A 相比, 本文所提方法能够对更多的洒落物实现检测定位, 有效避免漏检。图 9(c)、10(c) 中放大区域的红色点云代表洒落物在空间中的真实位置, 其他颜色的点云代表由算法预测生成的结果。二者在点云空间中的位置越重合则说明点云生成方法对目标位置的预测越准确。从图 9 和图 10 中 6 份放大区域可以看出, 本文算法所生成的洒落物点云与其在空间中的真实位置接近重合, 证明其对公路洒落物具有良好的定位效果。

公路洒落物检测任务需要提前发现并定位洒落物。

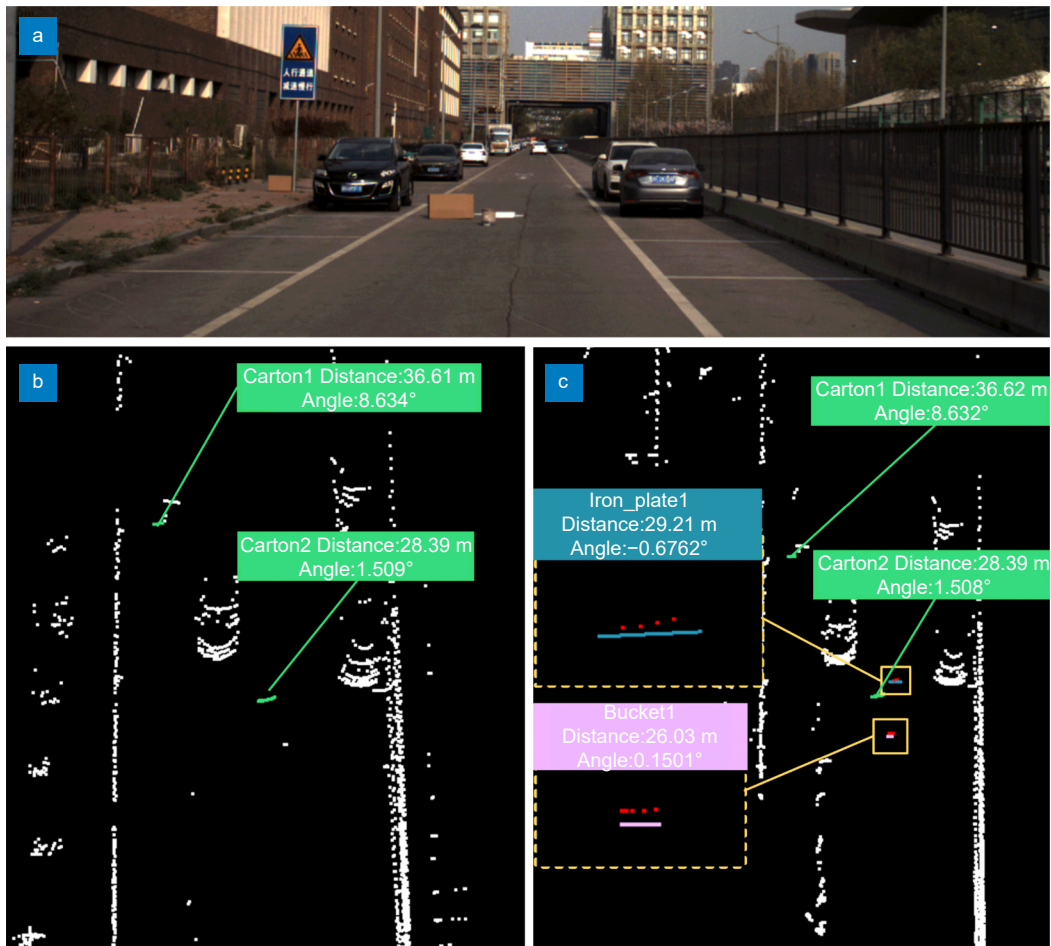


图 9 公路洒落物检测定位结果 (场景一, 截取部分区域)。 (a) 待测场景图像; (b) Method A 的检测定位结果; (c) 本文所提算法的检测定位结果

Fig. 9 Detection and localization results for abandoned objects on the road (Scene one, selected partial area). (a) Image of the scene to be tested; (b) Detection and localization results by method A; (c) Detection and localization results by our method



图 10 公路洒落物检测定位结果 (场景二, 截取部分区域)。 (a) 待测场景图像; (b) Method A 的检测定位结果; (c) 本文所提算法的检测定位结果

Fig. 10 Detection and localization results for abandoned objects on the road (Scene two, selected partial area). (a) Image of the scene to be tested; (b) Detection and localization results by method A; (c) Detection and localization results by our method

受制于相机像素和远距离处激光雷达点云的稀疏性, 检测难度会随着检测距离的增加而增大。不同距离下通过点云生成方法进行洒落物定位的误差实验结果见表 6。考虑到一般车辆在高速公路最低行驶速度为 60 km/h, 制动距离为 20~30 m。因此, 为了给检测系统留出预处理时间, 要求检测算法必须在 30 m 以外实现对目标的检出和定位。本文所提算法在 30~40 m 处可实现深度定位误差为 0.19 m, 水平宽度估计误差为 0.016 m, 角度定位误差为 0.082°, 有效实现了对公路上洒落物的定位功能。

### 3.4.3 公路洒落物检测与定位总体可视化效果

图像引导和点云空间约束的公路撒落物检测定位方法可以实现对场景中多尺度洒落物目标进行检测定

位, 并获取其二维预测框、目标类别、相对距离和角度等结果, 总体可视化效果如图 11 所示。

表 6 不同距离下通过点云生成方法进行洒落物定位的误差

Table 6 Error in abandoned objects localization using point cloud generation method at different distances

| Distance/m | MAE-error    |              |                           |
|------------|--------------|--------------|---------------------------|
|            | $\Delta D/m$ | $\Delta W/m$ | $\Delta\theta/(\text{°})$ |
| 0~20       | 0.132        | 0.0099       | 0.195                     |
| 20~30      | 0.156        | 0.0121       | 0.115                     |
| 30~40      | 0.188        | 0.0162       | 0.0819                    |
| Over 40    | 0.223        | 0.0271       | 0.0541                    |
| Total      | 0.181        | 0.0218       | 0.122                     |



图 11 公路洒落物检测与定位实验结果。(a) 场景一; (b) 场景二

Fig. 11 Experimental results for detecting and locating abandoned objects on the road. (a) Scene one; (b) Scene two

对于图中平放的铁板、灭火器等高程变化不明显的洒落物目标而言, 如果仅依赖点云进行检测则容易被分类为地面点云而产生漏检。本文所提方法能够在图像数据的辅助下解决这一问题, 实现对难例的检测和定位, 充分利用相机和激光雷达各自的优势, 提升了在复杂应用场景下的鲁棒性。使用测试集数据进行统计, 目标预测框提取部分处理每帧数据平均耗时 15.2 ms, 点云定位算法平均耗时 12.5 ms。所提方法每秒可处理数据 36 帧, 满足实时处理要求。

## 4 结论

本文提出了一种图像引导和点云空间约束的公路洒落物检测定位方法, 并围绕该方法搭建了检测系统进行实验验证。该方法主要利用改进的 YOLOv7-OD 网络从图像数据中识别目标并获取预测框信息, 将预测框投影到激光雷达坐标系下获得 ROI 区域。在 ROI 区域内, 通过点云聚类 and 点云生成算法在点云空

间约束下实现对目标的三维空间定位, 充分利用图像和点云的优势, 有效解决点云稀疏性导致中小尺度或远距离目标无法检测定位等问题。考虑实际检测环境, 30 ~ 40 m 为最佳检测距离。在此距离上, 该算法生成的洒落物点云数据在三维空间中深度定位误差为 0.19 m, 角度定位误差为 0.082°, 水平宽度估计误差为 0.016 m。除了可应用于公路洒落物检测定位任务, 本文所提方法也为缓解点云稀疏性和实现三维空间中中小目标定位等问题提供了一种解决思路。在未来的研究工作中, 我们将进一步考虑如何判断洒落物的质量和体积, 以实现后续自动拾取等工作。

## 参考文献

- [1] Liu W, Anguelov D, Erhan D, et al. SSD: single shot multibox detector[C]//14th European Conference on Computer Vision, Amsterdam, 2016: 21–37. [https://doi.org/10.1007/978-3-319-46448-0\\_2](https://doi.org/10.1007/978-3-319-46448-0_2).
- [2] Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection[C]//2017 IEEE International Conference on Computer Vision (ICCV), Venice, 2017: 2999–3007.

- <https://doi.org/10.1109/ICCV.2017.324>.
- [3] Carion N, Massa F, Synnaeve G, et al. End-to-end object detection with transformers[C]//16th European Conference on Computer Vision, Glasgow, 2020: 213–229. [https://doi.org/10.1007/978-3-030-58452-8\\_13](https://doi.org/10.1007/978-3-030-58452-8_13).
- [4] Bochkovskiy A, Wang C Y, Liao H Y M. YOLOv4: optimal speed and accuracy of object detection[Z]. arXiv: 2004.10934, 2020. <https://doi.org/10.48550/arXiv.2004.10934>.
- [5] Redmon J, Farhadi A. YOLOv3: an Incremental Improvement[Z]. arXiv: 1804.02767, 2018. <https://doi.org/10.48550/arXiv.1804.02767>.
- [6] Li C Y, Li L L, Jiang H L, et al. YOLOv6: a single-stage object detection framework for industrial applications[Z]. arXiv: 2209.02976, 2022. <https://doi.org/10.48550/arXiv.2209.02976>.
- [7] Ren S Q, He K M, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. *IEEE Trans Pattern Anal Mach Intell*, 2017, 39(6): 1137–1149.
- [8] He K M, Gkioxari G, Dollár P, et al. Mask R-CNN[C]//2017 IEEE International Conference on Computer Vision (ICCV), Venice, 2017: 2980–2988. <https://doi.org/10.1109/ICCV.2017.322>.
- [9] Jin Y, Zhang R, Yin D. Object detection for small pixel in urban roads videos[J]. *Opto-Electron Eng*, 2019, 46(9): 190053. 金瑶, 张锐, 尹东. 城市道路视频中像素小目标检测[J]. *光电工程*, 2019, 46(9): 190053.
- [10] Charles R Q, Hao S, Mo K C, et al. PointNet: deep learning on point sets for 3D classification and segmentation[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, 2017: 77–85. <https://doi.org/10.1109/CVPR.2017.16>.
- [11] Qi C R, Yi L, Su H, et al. PointNet++: deep hierarchical feature learning on point sets in a metric space[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems, Long Beach, 2017: 5105–5114. <https://doi.org/10.5555/3295222.3295263>.
- [12] Qian G C, Li Y C, Peng H W, et al. PointNeXt: Revisiting pointnet++ with improved training and scaling strategies[C]//36th Conference on Neural Information Processing Systems, New Orleans, 2022: 23192–23204.
- [13] Qi C R, Liu W, Wu C X, et al. Frustum PointNets for 3D object detection from RGB-D data[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, 2018: 918–927. <https://doi.org/10.1109/CVPR.2018.00102>.
- [14] Vora S, Lang A H, Helou B, et al. PointPainting: sequential fusion for 3D object detection[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, 2020: 4603–4611. <https://doi.org/10.1109/CVPR42600.2020.00466>.
- [15] Liang H L, Cai H Y, Liu B C, et al. Road falling objects detection algorithm based on image and point cloud fusion[J]. *Laser Optoelectron Prog*, 2023, 60(10): 1010001. 梁浩林, 蔡怀宇, 刘博翀, 等. 基于图像与点云融合的公路撒落物检测算法[J]. *激光与光电子学进展*, 2023, 60(10): 1010001.
- [16] Zheng X Y, Lai J Z, Lv P, et al. Object detection and positioning method based on infrared vision/lidar fusion[J]. *Navig Position Timing*, 2021, 8(3): 34–41. 郑欣悦, 赖际舟, 吕品, 等. 基于红外视觉/激光雷达融合的目标识别与定位方法[J]. *导航定位与授时*, 2021, 8(3): 34–41.
- [17] Fan W S, Lai J Z, Lv P, et al. Vision/lidar object tracking and localization method based on improved DeepSORT[J]. *Navig Position Timing*, 2022, 9(4): 77–84. 范婉舒, 赖际舟, 吕品, 等. 基于改进DeepSORT的视觉/激光雷达目标跟踪与定位方法[J]. *导航定位与授时*, 2022, 9(4): 77–84.
- [18] Wang C Y, Bochkovskiy A, Liao H Y M. YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, 2023: 7464–7475. <https://doi.org/10.1109/CVPR52729.2023.00721>.
- [19] Zhang W M, Qi J B, Wan P, et al. An easy-to-use airborne LIDAR data filtering method based on cloth simulation[J]. *Remote Sens*, 2016, 8(6): 501.
- [20] Huang J K, Grizzle J W. Improvements to target-based 3D LIDAR to camera calibration[J]. *IEEE Access*, 2020, 8: 134101–134110.
- [21] Sun P, Kretschmar H, Dotiwalla X, et al. Scalability in perception for autonomous driving: waymo open dataset[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, 2020: 2443–2451. <https://doi.org/10.1109/CVPR42600.2020.00252>.
- [22] Ge Z, Liu S T, Wang F, et al. YOLOX: exceeding YOLO series in 2021[Z]. arXiv: 2107.08430, 2021. <https://doi.org/10.48550/arXiv.2107.08430>.
- [23] Jocher G, Chaurasia A, Stoken A, et al. ultralytics/yolov5: v7.0 - YOLOv5 SOTA Realtime Instance Segmentation[EB/OL]. (2022-11-22). <https://github.com/ultralytics/yolov5>.
- [24] Jocher G, Chaurasia A, Qiu J. Ultralytics YOLO[EB/OL]. (2023-01-10). <https://github.com/ultralytics/ultralytics>.

## 作者简介



蔡怀宇 (1965-), 女, 教授, 博士生导师, 1988年、1991年、2001年于天津大学分别获得学士、硕士、博士学位, 现为天津大学精密仪器与光电子工程学院教授, 主要从事光电成像与检测技术、信息光学和图像处理技术方面的研究。

E-mail: [hycail@tju.edu.cn](mailto:hycail@tju.edu.cn)



【通信作者】陈晓冬 (1975-), 男, 教授, 博士生导师, 1996年、2002年于天津大学分别获得学士学位和博士学位, 现为天津大学精密仪器与光电子工程学院教授, 主要从事光电成像与检测技术方向的研究。

E-mail: [xdchen@tju.edu.cn](mailto:xdchen@tju.edu.cn)



扫描二维码, 获取PDF全文

# Image-guided and point cloud space-constrained method for detection and localization of abandoned objects on the road

Cai Huaiyu<sup>1,2</sup>, Yang Zhaoqian<sup>1,2</sup>, Cui Ziyang<sup>1,2</sup>, Wang Yi<sup>1,2</sup>, Chen Xiaodong<sup>1,2\*</sup>



Experimental results for detecting and locating abandoned objects on the road

**Overview:** Highways constitute a vital economic lifeline for a nation. With the continuous increase in highway mileage and traffic volume, the significance of daily maintenance work on the road has become more pronounced. The detection and localization of abandoned objects on the road are among the primary tasks in highway maintenance. Because if abandoned objects are not promptly cleared, they can easily lead to traffic congestion or even cause accidents. Detecting and locating abandoned objects on the road is a specific object detection task. In order to fully leverage the advantages of both image and point cloud data, solutions based on multisensor fusion have become a research hotspot. However, due to the sparse nature of the LiDAR point clouds, existing multisensor fusion methods usually encounter challenges such as missed detection, false alarms, and difficulties in localization when detecting small-to-medium-sized abandoned objects. To address the aforementioned issues, this paper proposes a method for detecting and locating abandoned objects on the road using image guidance and point cloud spatial constraints. Firstly, on the foundation of the YOLOv7, a small object detection layer has been added, and a channel attention mechanism has been introduced to enhance the network's ability to extract two-dimensional bounding boxes for small-to-medium-sized targets within the image. Subsequently, the predicted bounding boxes are projected onto the LiDAR coordinate system to generate a pyramidal region of interest (ROI). For larger targets, sufficient point cloud data allows for three-dimensional spatial position estimation through point cloud clustering within the ROI. For smaller targets, which have insufficient point cloud data for clustering within the ROI, spatial constraints from surrounding ground point cloud data are used. Using projection transformation relationships, point cloud data is generated to obtain spatial position information for the smaller targets, achieving the detection and localization of multiscale abandoned objects on the road in three-dimensional space. The experimental results show that the improved YOLOv7-OD network achieves recall and average precision rates of 85.4% and 82.0%, respectively, for medium-sized objects, representing improvements of 6.6% and 8% compared to the YOLOv7. The recall and average precision rates for small-sized objects are 66.8% and 57.3%, respectively, with an increase of 5.3%. In terms of localization, for abandoned objects located 30~40 m away from the detecting vehicle, the depth localization error is 0.19 m, and the angular localization error is 0.082°. The proposed algorithm can process 36 frames of data per second, effectively achieving real-time detection and localization of abandoned objects on the road.

Cai H Y, Yang Z Q, Cui Z Y, et al. Image-guided and point cloud space-constrained method for detection and localization of abandoned objects on the road[J]. *Opto-Electron Eng*, 2024, 51(3): 230317; DOI: [10.12086/oe.2024.230317](https://doi.org/10.12086/oe.2024.230317)

<sup>1</sup>School of Precision Instrument and Opto-electronics Engineering, Tianjin University, Tianjin 300072, China; <sup>2</sup>Key Laboratory of Optoelectronic Information Technology Ministry of Education, Tianjin University, Tianjin 300072, China

\* E-mail: [xdchen@tju.edu.cn](mailto:xdchen@tju.edu.cn)