

基于注意力的毫米波-激光雷达融合目标检测

李朝^{1,2}, 兰海^{1*}, 魏宪¹

(1. 中国科学院海西研究院 泉州装备制造研究所, 福建 泉州 362216; 2. 中北大学 电气与控制工程学院, 太原 036005)

(* 通信作者电子邮箱 lanhai09@fjirsm.ac.cn)

摘要:针对自动驾驶中使用激光雷达进行目标检测时漏检被遮挡目标、远距离目标和复杂天气场景下目标的问题,提出一种基于注意力机制的毫米波-激光雷达特征融合的目标检测方法。首先,将毫米波和激光雷达各自的扫描帧数据分别聚合到它们的标注帧上,并将毫米波和激光雷达的数据点进行空间对齐;其次,对两者进行聚合和空间对齐后的数据分别进行PointPillar点云柱快速编码,转换成伪图像;最后,通过中间卷积层提取两者的传感器特征,并利用注意力机制对两者的特征图进行融合,融合后的特征图通过单阶段检测器得到检测结果。实验结果显示,该融合算法在nuScenes数据集中的平均精度均值(mAP)高于PointPillar基础网络,而且注意力融合的检测方法的性能表现优于利用拼接融合、相乘融合、相加融合的检测方法。可视化结果显示所提方法是有效的,能提高网络对被遮挡目标、远处目标和雨雾天气下目标检测的鲁棒性。

关键词:自动驾驶;目标检测;传感器融合;注意力机制;激光雷达;毫米波雷达

中图分类号:TP391.4 **文献标志码:**A

Attention-based object detection with millimeter wave radar-lidar fusion

LI Chao^{1,2}, LAN Hai^{1*}, WEI Xian¹

(1. Quanzhou Institute of Equipment Manufacturing, Haixi Institute, Chinese Academy of Sciences, Quanzhou Fujian 362216, China;

2. School of Electrical and Control Engineering, North University of China, Taiyuan Shanxi 036005, China)

Abstract: To address problems of missing occluded objects, distant objects and objects in extreme weather scenarios when using lidar for object detection in autonomous driving, an attention-based object detection method with millimeter wave radar-lidar feature fusion was proposed. Firstly, the scan frame data of millimeter wave radar and lidar were aggregated into their respective labeled frames, and the points of millimeter wave radar and lidar were spatially aligned, then PointPillar was employed to encode both the millimeter wave radar and lidar data into pseudo images. Finally, the features of both millimeter wave radar and lidar sensors were extracted by the middle convolution layer, and the features maps of them were fused by attention mechanism, and the fused feature map was passed through a single-stage detector to obtain detection results. Experimental results on nuScenes dataset show that compared to the basic PointPillar network, the mean Average Precision (mAP) of the proposed attention fusion algorithm is higher, which performs better than concatenation fusion, multiply fusion and add fusion methods. The visualization results show that the proposed method is effective and can improve the robustness of the network for detecting occluded objects, distant objects and objects surrounded by rain and fog.

Key words: autonomous driving; object detection; sensor fusion; attention mechanism; lidar; millimeter wave radar

0 引言

车辆自动驾驶的安全性依赖于对周围环境的准确感知。目前车辆采用的主要感知器有激光雷达、摄像头和毫米波雷达。其中:激光雷达精度高,探测距离较远,受天气影响小,但数据较稀疏;摄像头图像则具有丰富的颜色信息,但受天气和光照影响较大;毫米波雷达精度较低,但探测距离远,受天气影响极小,也存在数据稀疏的特点。目前有基于单个传感器或多个传感器融合的目标检测,其中不同传感器数据进行融合能提高无人驾驶系统的鲁棒性和冗余性。

在光照条件不友好的环境下,摄像头难以发挥作用,激光雷达和毫米波雷达是车辆感知环境的主要手段。激光雷达与毫米波雷达所产生的传感数据均以三维点云数据为主,两者在数据形式上有着很高的相似性。基于激光雷达点云数据的目标检测基本上还是解决数据的无序性和稀疏性问题。文献[1]中提出的PointNet是具有开创性的工作,真正地实现了无序点云的端到端学习。PointNet通过池化操作解决点的无序性问题,通过数据对齐操作保证旋转不变性。除了直接将无序点云输入进网络,还可以通过将点云数据先离散化处理再输入检测网络,例如文献[2-3]中通过将无序的点云划分到有

收稿日期:2020-09-01;修回日期:2020-11-28;录用日期:2020-12-08。

基金项目:国家自然科学基金青年科学基金资助项目(61806186);福建省智能物流产业技术研究院建设项目(2018H2001);机器人技术与系统国家重点实验室(HIT)资助项目(SKLRIS-2019-KF-15);泉州市科技计划项目(2019C112, 2019STS08)。

作者简介:李朝(1994—),男,江西萍乡人,硕士研究生,主要研究方向:三维目标检测、传感器融合;兰海(1988—),男,福建莆田人,助理研究员,硕士,主要研究方向:机器学习与模式识别及其在医疗影像中的应用;魏宪(1986—),男,河南泌阳人,研究员,博士,CCF会员,主要研究方向:机器学习与模式识别及其在无人系统中的应用。

序的空间体素的方法解决点云数据的无序性问题,之后再通过3D卷积提取特征,但3D卷积计算量太大。AVOD网络^[4]、MV3D网络^[5]使用2D卷积对点云鸟瞰图进行特征提取,提高了检测速度。

毫米波雷达数据比激光雷达数据更稀疏,但信息比较丰富。文献[6]中基于调频连续波(Frequency Modulated Continuous Wave, FMCW)算法利用毫米波雷达检测目标的方位角、速度、距离,但是误差较大,且无法检测出目标的属性。文献[7]中提出对毫米波雷达数据利用随机森林分类器和长短期记忆(Long Short-Term Memory, LSTM)网络对目标进行分类。文献[8]中则将整个原始雷达数据作为输入,采用PointNet++^[9]的基础架构,得到了每一次毫米波雷达反射的各个类概率,不需要进行聚类 and 人为地选择特征。文献[10]中认为虽然毫米波雷达数据比激光雷达数据更稀疏,但与激光雷达单一坐标和强度数据相比,还拥有多普勒速度和雷达截面积数据,能检测到激光检测不到的弱目标或遮挡目标,因此开创性地使用雷达数据的位置、速度和雷达截面积信息在PointNet框架上实现了车辆2D边界框的检测。

无论是基于激光雷达还是毫米波雷达的目标检测方法,单一传感器的感知能力都是有限的,因此传感器融合已经成为目标检测的主要方法。传感器融合主要分为数据级融合、特征级融合和目标级融合。文献[11-13]中结合激光雷达精度高、毫米波雷达能够检测车辆速度的优点进行车辆的检测和跟踪,提高了检测范围和跟踪精度。文献[14]中提出的RRPN(Radar Region Proposal Network)利用投影到图像坐标系中的毫米波雷达点生成预设置大小的锚框作为目标感兴趣区域,再通过检测网络进行检测,减少了90%的锚框数量,提高了运算速度。文献[15]中将毫米波雷达投影到图像坐标系后变成二维图像,使用卷积神经网络提取毫米波雷达和摄像头图像特征图,并对特征图对应元素进行相加融合,再对融合后的特征图使用SSD(Single Shot multibox Detector)^[16]框架进行目标检测。与采取投影方法不同,文献[17]中将毫米波雷达的距离、横向速度和纵向速度分别转换为图像R、G、B通道的真实像素值,再将转换后的毫米波雷达和图像相乘融合。文献[18]中则提出了毫米波雷达和图像融合网络RVNet,该网络是基于YOLO^[19]检测框架的特征图拼接融合网络,并且为大目标和小目标分别设有两个输入分支和输出分支以提高检测精度。文献[20]中提出了毫米波雷达和图像融合的CRF-Net,在各个卷积网络层进行特征图拼接融合,以学习在哪个层的融合目标检测效果更优,并提出了一种叫作BlackIn的训练策略以确保融合网络收敛。

除了传感器融合方法以外,注意力机制也被应用到图像领域并取得了巨大的进展。注意力机制最早从人类的视觉原理中获取灵感,并在自然语言处理中取得了很好的效果^[21-22]。注意力机制通过捕捉数据点之间的相互影响,获取数据间的上下文信息并以此作为权重输出结果,是对深度学习模型的有力补充。文献[23]中提出的两级注意力模型应用物体级和部位级两种注意力,使用卷积网络得到物体级信息,再使用聚类的方法得到重点局部区域,从而能更精确地利用多层次信息。文献[24]中提出了通道注意力机制,认为特征图的不同通道的重要程度不同,网络通过全局平均池化获取特征图每个通道的数值分布情况,增大有效特征图通道的权重,利用激励操作来获取通道之间的依赖性,并以此作为权重输出结果。

除了利用通道注意力机制判断不同通道之间的权重关系,另外就是像素点之间的注意力机制。文献[25]中认为卷积神经网络只能关注卷积核感受野内的像素点信息,无法学习全局信息对当前区域的影响,因此通过特征图之间矩阵相乘的方法确定每个像素和其他像素间的关系。

本文针对激光雷达进行目标检测时对遮挡目标、远距离目标和复杂天气场景中的目标检测能力弱的问题,提出基于注意力机制的毫米波-激光雷达数据融合的目标检测方法。原因如下:1)毫米波雷达不受天气光照影响,并且对车辆等金属敏感,能够穿透树木草丛检测出车辆,弥补激光雷达受到的干扰^[10];2)激光雷达对远处的物体探测结果较为稀疏,难以实现远处物体的类别检测,而毫米波雷达探测距离远,原理上探测距离的四次方与雷达散射面积成正比,兼具多普勒效应,能够检测速度,极大地增强了远处物体的检测精度;3)注意力机制能够有效提取数据间的上下文信息,利用数据点之间的权重关系输出结果,十分适合毫米波-激光点云数据之间的融合,能够充分发挥毫米波雷达和激光雷达各自的优点。本文通过点云柱快速编码网络PointPillar^[26]提取经过空间对齐的激光雷达和毫米波雷达特征,然后将毫米波-激光雷达特征图进行融合,弥补单一雷达传感器检测上存在的不足,提高了算法模型对物体目标的检测精度,亦提高了恶劣天气下算法表现的鲁棒性。本文代码公开在<https://github.com/MVPR-Group/radar-lidar-fusion>。

1 注意力融合方法

本章主要介绍激光雷达和毫米波雷达融合的方法,通过利用不同传感器各自的优势,弥补激光雷达存在的缺陷,提高网络性能。文献[10]的研究发现,激光雷达在探测目标时,目标距离越远,返回的激光雷达点越少,强度越弱,易受雨雾、树木遮挡;其次,毫米波雷达发送信号所使用波长远大于激光雷达,能够穿透塑料、墙板和衣服等特定的材料,并且不受雨、雾、灰尘和雪等环境条件的干扰;另外毫米波雷达数据相对于激光雷达数据更稀疏,但毫米波雷达数据在目标速度和雷达截面积(Radar Cross-Section, RCS)信息上具有很强的特征。例如,移动的车辆具有较高的相对速度以及车身能够产生高RCS值。所有这些特征对于目标检测都非常有用。本文设计了基于注意力机制的毫米波-激光雷达数据融合目标检测网络,如图1所示。该网络包含四个模块:点云柱快速编码模块、卷积特征提取模块、注意力融合模块和SSD检测模块。

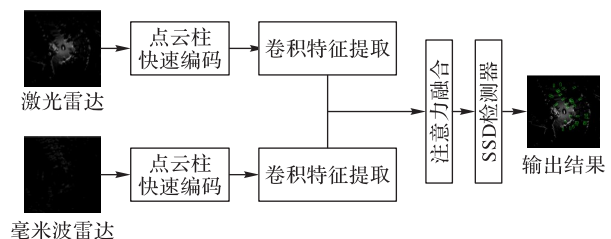


图1 传感器注意力机制融合网络框架

Fig. 1 Sensor attention mechanism fusion network framework

1.1 点云柱快速编码

激光雷达和毫米波雷达都是无序的稀疏点云数据。为了使激光雷达和毫米波雷达能够良好地融合,本文对激光雷达和毫米波雷达采取了点云柱快速编码^[26]方法。如图2所示,点云柱快速编码方法首先以自身为中心,在100 m×100 m的

地平面范围中均匀生成 400×400 个立方柱体,即每个柱体的底面大小为 $0.25 \text{ m} \times 0.25 \text{ m}$,高度限制为 10 m ;每个点云柱中的点数约束为 N ,多则采样,少则补0,并对每个点进行维度扩展。将激光雷达点云原始数据的三维坐标 (x_l, y_l, z_l) 和强度 I 加入 $(x_c, y_c, z_c, x_p, y_p)$ 5 个额外维度。其中, (x_c, y_c, z_c) 为该点云柱中所有点的坐标平均值,即所有点的聚类中心; (x_p, y_p) 为各点到点云柱中心的 x - y 坐标偏移量,此时点云柱中的每个点有 9 个维度。考虑到点云数据的稀疏性,因此在单次训练样本中的非空点云柱数目约束为 P ,并根据实际数量随机采样或补0。整个点云数据被编码为形状 (D, P, N) 的张量, D 是点云柱特征维度, P 是非空点云柱数量, N 为单个点云柱中数据点的个数。 (D, P, N) 利用 1×1 卷积操作进行线性变换后得到张量 (C, P, N) ,对每个点云柱中的所有点进行最大化池化操作得到特征矩阵 (C, P) 。最后将 P 个非空点云柱内的点映射回检测范围内的原始位置得到大小为 (C, W, H) 的二维点云伪图像。

毫米波雷达点云数据共有 18 维(具体见 3.1 章),与激光雷达只利用位置信息和强度信息不同,为了弥补激光雷达数据的不足,毫米波雷达保留其中的坐标 (x_m, y_m, z_m) 、补偿速度 (V_{x_comp}, V_{y_comp}) 及目标雷达截面积 (I_{RCS}) 共 6 个维度。相比激光雷达点的位置信息,毫米波雷达点的位置信息正样本比例高,受距离因素、天气因素影响小;相比激光雷达的反射强度信息,毫米波雷达 RCS 信息能够直接反映出目标的体积大小,尤其卡车、汽车和行人 RCS 特征差别明显,起到了信息互补作用。除此之外毫米波雷达还能检测出目标的矢量速度信息来辅助检测任务。

为更好地提取毫米波雷达的特征,本文对点云柱快速编码方法做了改进。由于所有毫米波雷达点云数据中的 z_r 都为 0,在对毫米波雷达点云柱快速编码过程中,去除了激光点云数据 (x_c, y_c, z_c) 中的 z_c 项及 (x_p, y_p) 两项。改进后的毫米波雷达点云柱快速编码网络提取 8 个特征 $(x_m, y_m, z_m, V_{x_comp}, V_{y_comp}, x_c, y_c, I_{RCS})$ 。编码后的毫米波雷达为形状 (D_r, P_r, N_r) 的张量,之后根据柱体坐标映射得到与激光雷达相同维度的二维点云伪图像。

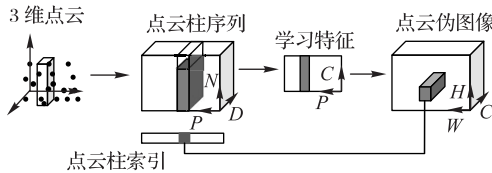


图2 点云柱快速编码

Fig. 2 Fast encode for PointPillar

1.2 注意力融合

本文提出的基于注意力机制的毫米波-激光雷达数据融合方法如图3所示:首先采用注意力机制对卷积特征提取模块输出的毫米波与激光雷达特征图进行融合,如式(1)所示,其中: $X_l \in \mathbf{R}^{C \times N}$ 表示激光雷达特征图, $X_r \in \mathbf{R}^{C \times N}$ 表示毫米波雷达特征图, O 表示注意力融合后的激光雷达特征图。

$$\text{attention}(X_l, X_r) = O \quad (1)$$

注意力融合过程中,定义式(2):

$$\begin{cases} Q = W_q \cdot X_l \\ K = W_k \cdot X_r \\ V = W_v \cdot X_l \end{cases} \quad (2)$$

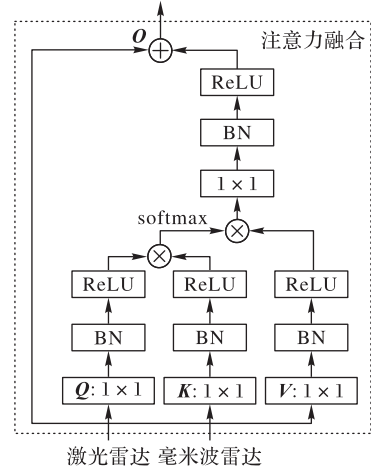


图3 激光雷达和毫米波雷达注意力融合

Fig. 3 Attention fusion of lidar and millimeter wave radar

如图3所示:激光雷达特征图经过 1×1 卷积层和 BatchNorm 层、ReLU 激活层后得到 Q 和 V ;毫米波雷达特征图经过 1×1 卷积层和 BatchNorm 层、整流线性单位函数 (Rectified Linear Unit, ReLU) 激活层后得到 K 。使用点乘作为 Q 与 K 的内积形式,并将结果利用 Softmax 进行归一化,可计算出激光雷达特征图所对应的 Q 与毫米波雷达特征图所对应的 K 之间的关系权重矩阵 A , A 中各项 a_{ij} 计算如式(3)所示。

$$a_{ij} = \frac{\exp(q_i^T \cdot k_j)}{\sum_{j=1}^M \exp(q_i^T \cdot k_j)} \quad (3)$$

在得到毫米波-激光雷达点云数据间的关系权重矩阵 A 后,如式(4)所示,将优化后的权重矩阵和激光雷达特征图所对应的 V 相乘,即得到融合结果 O 。

$$O_j = \sum_{i=1}^M A_{ij} \cdot V_j \quad (4)$$

另外,借鉴残差模块的概念^[27],如式(5)所示,将融合结果 O 乘上比例系数 λ 并加上激光雷达特征图 X_l ,得到最终输出结果 y 。 λ 初始值设为 0,通过训练学习增大该权重系数。其物理含义可视为一开始注意力机制的影响为 0,随着训练的逐渐增大注意力在输出中的影响。

$$y = X_l + \lambda O \quad (5)$$

除了传感器注意力融合方法,本文还与拼接融合、相乘融合和相加融合进行了对比,各个方法在网络中的融合位置相同。文献[20]中进行了传感器特征图拼接融合实验。本文对激光雷达和毫米波雷达特征图通道维度进行叠加,得到维度 $(2C \times W \times H)$ 融合特征图。融合后的特征图通过 1×1 卷积降维到原来的维度。

文献[15]中进行了特征图相加融合实验。本文对激光雷达特征图和毫米波雷达特征图对应元素相加融合。

文献[17]中采用了特征图相乘的融合方式。由于毫米波雷达的稀疏性毫米波雷达特征图 X_r 比激光雷达特征图 X_l 有更多的元素为 0,因此对毫米波雷达特征图 X_r 为 0 的元素进行加 1 操作,如式(6)所示,得到毫米波特征图 X_r' ;将毫米波-激光

雷达特征图相乘,如式(7)所示。加1操作保证相乘融合时不会丢失激光雷达特征图中包含的信息,但又能通过毫米波雷达强化相同位置激光雷达特征图信息流。

$$X_r' = \begin{cases} X_r + 1, & X_r = 0 \\ X_r, & \text{其他} \end{cases} \quad (6)$$

$$O = X_r' \cdot X_l \quad (7)$$

2 融合检测网络结构

本文使用 PointPillar 点云快速编码网络框架作为基础网络,并在此网络模型上加入融合模块进行改进。PointPillar 采用类似文献[2]的主干网络结构。输入数据在经过点云柱快速编码之后,生成点云伪图像后进入主干网络,主干网采用空间金字塔池化结构,包含两个子网络:一个是自上向下的下采样卷积网络产生空间分辨率越来越小的特征;另一个卷积网络分支将前面3个卷积块的输出卷积成相同大小的特征图,如图4所示。提取出毫米波-激光雷达点云数据的特征之后,将两者送入融合模块,最终将融合结果送入检测模块,输出结果。

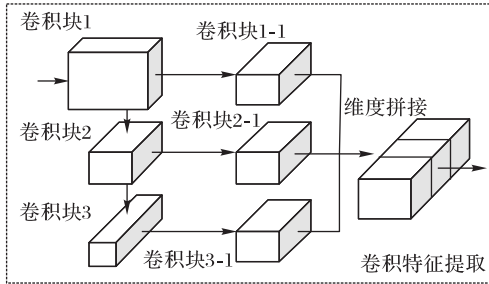


图4 卷积特征提取主干网络

Fig. 4 Convolutional feature extraction backbone network

在激光雷达和毫米波雷达的点云柱快速编码模块中,每个点云柱中包含点数量 N 设置为 60,非空点云柱数量 P 设置为 30 000。编码后得到维度为 (C, W, H) 的伪图像,其中 W 和 H 等于 400, C 等于 64。

通过点云柱编码得到维度 (C, W, H) 的伪图像后,为了检测不同尺寸的目标,在卷积特征提取层设置了两个子网络,它们的连接方式如图4所示。前子网络的每个卷积块第一层下采样步长为 2,每个卷积后面都接一个 BatchNorm 层和 ReLU 层。前子网络卷积块输出作为后子网络卷积块和后子网络卷积块的输入。每个卷积后的特征图经过 1-1、2-1、3-1 子网络卷积块得到相同的维度为 $(2C, W/8, H/8)$ 的特征图,三个模块拼接成维度 $(6C, W/8, H/8)$ 的特征图。

分别提取了激光雷达和毫米波雷达特征图后,本文尝试了注意力融合方法和另外三种融合方法:拼接融合、相加融合和相乘融合。通过上述的点云柱快速编码模块和卷积特征提取模块后,激光雷达和毫米波雷达从无序的点云转化为有序的伪图像。两者在空间上具有良好的对应性,这对传感器融合十分重要。nuScenes^[28]数据集标签注释的各类目标物框内的激光雷达点数量是毫米波雷达的 4~10 倍,这就意味着一个目标物上有很少的毫米波雷达点。例如,一辆车的长大约 4.5 m,宽 2 m,使用 $(0.25, 0.25)$ 的点云柱的条件下,车辆所占的激光雷达点云柱约有 100 个,而毫米波雷达只有几个。如图1所示,通过将融合模块放置在卷积特征提取层后,利用卷积特征提取操作来扩大毫米波雷达感受野,增强网络整体性

能。将扩大了感受野的毫米波雷达特征图使用上述介绍的注意力融合方法进行实验,并在相同位置进行另外三种融合方法对比。

经过传感器注意力后的特征图使用 SSD 检测器进行 3D 检测。通过匹配设置的先验框和真实框的 2D 平面重叠度交并比 (Intersection over Union, IoU) 进行筛选。框的高度和距离地面的高度作为额外的回归目标。

本文通过 3 个 1×1 的卷积层实现分类、位置回归和方向回归。根据先验知识设置 9 种大小的 3D 框,每个类都设置不同的匹配和非匹配 IoU 阈值。每个框有 7 个维度 $(x, y, z, w, h, l, \theta)$,分别代表着框的长、宽、高、中心坐标和方向。使用文献[26]的损失函数计算损失。真实框和生成框之间的位置回归残差定义为式(8),尺寸回归残差定义为式(9),方向回归定义为式(10)。

$$\begin{cases} \Delta x = (x^{\text{gt}} - x^{\text{a}}) / \sqrt{(w^{\text{a}})^2 + (l^{\text{a}})^2} \\ \Delta y = (y^{\text{gt}} - y^{\text{a}}) / \sqrt{(w^{\text{a}})^2 + (l^{\text{a}})^2} \\ \Delta z = (z^{\text{gt}} - z^{\text{a}}) / \sqrt{h^{\text{a}}} \end{cases} \quad (8)$$

$$\begin{cases} \Delta w = \log(w^{\text{gt}} / w^{\text{a}}) \\ \Delta l = \log(l^{\text{gt}} / l^{\text{a}}) \\ \Delta h = \log(h^{\text{gt}} / h^{\text{a}}) \end{cases} \quad (9)$$

$$\Delta \theta = \sin(\theta^{\text{gt}} - \theta^{\text{a}}) \quad (10)$$

其中:上标 gt 表示真实值,上标为 a 表示预测值。总位置损失函数的定义为式(11):

$$L_{\text{loc}} = \sum_{b \in \{x, y, z, w, h, l, \theta\}} \text{smoothL1}(\Delta b) \quad (11)$$

其中 SmoothL1 定义如下:

$$\text{smoothL1}(x) = \begin{cases} 0.5x^2, & |x| < 1 \\ |x| - 0.5, & \text{其他} \end{cases} \quad (12)$$

由于文献[2]中定义的方向损失函数不能区分 0° 和 180° 旋转的框,本文使用文献[26]中的方向损失函数,定义为式(13):

$$L_{\text{dir}} = \text{smoothL1}(\Delta \theta) \quad (13)$$

分类函数 L_{cls} 使用的是 Focal loss^[29] 损失函数,如式(14)所示。其中 p^{a} 代表框的分类概率, $\alpha = 0.25$, $\gamma = 2$ 。

$$L_{\text{cls}} = -\alpha(1 - p^{\text{a}})^{\gamma} \log p^{\text{a}} \quad (14)$$

总的损失函数定义为:

$$L = \frac{1}{N_{\text{pos}}} (\lambda_1 L_{\text{loc}} + \lambda_2 L_{\text{cls}} + \lambda_3 L_{\text{dir}}) \quad (15)$$

其中: N_{pos} 代表正样本框的数量,即大于设定 IoU 阈值的框的数量;设置 $\lambda_1 = 2$, $\lambda_2 = 1$, $\lambda_3 = 0.2$ 。

3 实验与结果分析

3.1 数据预处理

本文采用的是 nuScenes 数据集,该数据集包含 1 个 32 线激光雷达、5 个毫米波雷达、5 个摄像头的所有传感数据。数据集提供的毫米波雷达数据是经过聚类处理的雷达点,每个雷达点有 18 个维度,包含坐标、速度、雷达散射面积、雷达动态特性、多普勒模糊解状态、有效性状态等,如表1提供的部分信息所示。可以通过雷达状态通道对雷达点进行筛选的方法来滤除不相关雷达点。本文实验中对毫米波雷达滤波设置

是保留多普勒模糊解:3表示清楚的,以及点有效性状态:0表示有效的和所有雷达动态特性下的毫米波雷达点。滤波前和滤波后的毫米波图像如图5所示,图左上角为安装在车头处毫米波雷达数据,雷达点上的线条表示速度方向和大小。

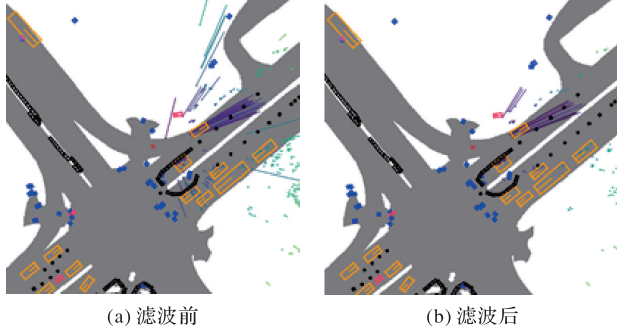


图5 毫米波雷达点滤波前后图像对比

Fig. 5 Image comparison of millimeter wave radar points before and after filtering

表1 毫米波雷达数据各个通道及其说明

Tab. 1 Channels and channel descriptions of radar data

雷达通道	描述说明
x, y, z	雷达点坐标
dyn_prop	雷达点动态特性:0表示移动,1表示固定,2表示接近,3表示静止的候选目标,4表示未知,5表示穿过静止的,6表示穿过移动的,7表示停止
I_{RCS}	雷达截面积
V_x, V_y	x 和 y 方向的速度,单位:m/s
V_{x_comp}, V_{y_comp}	x 和 y 方向的补偿速度,单位:m/s
$ambig_state$	多普勒模糊解的状态:0表示有效的,1表示不明确,2表示交错,3表示清楚,4表示等待候选
$invalid_state$	点有效性状态:0表示有效的,1表示低RCS无效等18个状态

在本文中使用的激光雷达的坐标和强度信息,毫米波雷达的坐标、RCS和速度信息。激光雷达和毫米波雷达安装在车辆的不同位置并使用不同坐标系。以车辆的惯性测量单元(Inertial Measurement Unit, IMU)作为参考点;激光雷达平移矩阵 T_l ,旋转矩阵 R_l ,毫米波雷达雷达平移矩阵 T_r ,旋转矩阵 R_r ,其中毫米波雷达转换到激光雷达坐标系的旋转矩阵 $R = R_l \cdot R_r$,转换到激光雷达安装位置的平移矩阵 $T = T_l - T_r$ 。通过式(16)可将毫米波雷达点云数据中的坐标转换到激光雷达空间,转换后的毫米波雷达坐标记为 (x_d, y_d, z_d) 。如图6所示,毫米波雷达的速度方向并不能反映物体的绝对速度 V ,而是表示与自身车辆的相对径向速度 V_r 。该速度在 $x-y$ 方向上的分量为 $(V_x, V_y) = (V_r \cdot \cos \alpha, V_r \cdot \sin \alpha)$,车辆自身速度 (V_{ex}, V_{ey}) ,补偿速度 $(V_{x_comp}, V_{y_comp}) = (V_x, V_y) - (V_{ex}, V_{ey})$,利用式(17)将毫米波雷达坐标系下的速度转化为激光雷达坐标系的速度 $(V_{x_comp_l}, V_{y_comp_l})$ 。

虽然毫米波雷达数据缺乏相对切向速度,不能完全反映出物体的真实运动速度,但是通过结合其他信息对物体的运动状态进行粗略判断也能够在对障碍物检测中发挥积极作用。

$$\begin{bmatrix} x_d \\ y_d \\ z_d \end{bmatrix} = R \begin{bmatrix} x_r \\ y_r \\ z_r \end{bmatrix} - T \quad (16)$$

$$\begin{bmatrix} V_{x_comp_l} \\ V_{y_comp_l} \\ 1 \end{bmatrix} = R \begin{bmatrix} V_{x_comp} \\ V_{y_comp} \\ 1 \end{bmatrix} \quad (17)$$

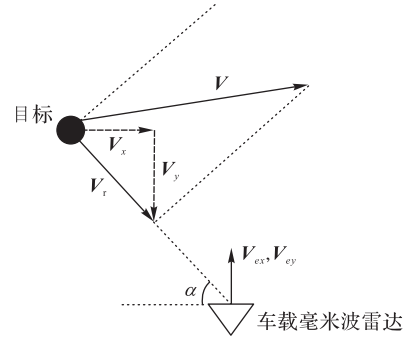


图6 毫米波雷达点速度示意图

Fig. 6 Schematic diagram of radar point velocity

3.2 实验设置

nuScenes数据集包含了28130个训练样本和6019个测试样本。数据集的激光雷达扫描频率是20帧/秒(Frame Per Second, FPS),32线束,探测距离100 m,精度 ± 0.02 m,每帧大约3万个点。毫米波雷达是77 Hz的调频连续波(Frequency Modulated Continuous Wave, FMCW)雷达,扫描频率13 FPS,探测距离250 m,近距离精度 ± 0.1 m,远距离精度 ± 0.4 m,每帧扫描聚类后的点数最多125个。因为标注样本所占的比例是每秒2帧,所以将全部扫描帧中连续10帧激光雷达和连续5帧毫米波雷达聚合到样本帧进行数据增强。本文中目标检测包含9个目标分类:汽车、卡车、客车、拖车、工程车辆、行人、摩托车、交通锥和栅栏。各个类在整个数据集所占比例如图7所示,以下实验均使用单个GPU完成,由于数据集较大,训练完整数据集进行耗时较长,所以使用1/2数据集即14065个训练样本进行训练,测试样本6019个。

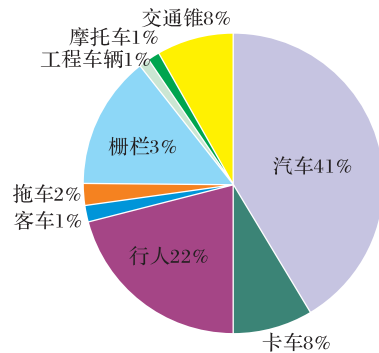


图7 数据集各类实例所占百分比

Fig. 7 Percentages of various classes of instances in dataset

与训练一个网络仅识别一类目标不同,训练一个网络同时进行9类目标的检测。训练时批量大小设置为3,测试时为1,训练次数为30个epoch(140000次迭代)。本文总共设置2500个锚点,每个点上18个3D框,即每个点上每个类两个框方向分别设置为 0° 和 90° 。

在实验过程中进行了多组对比实验。在毫米波-激光雷达融合方法上使用了注意力融合、拼接融合、相加融合和相乘融合,并和激光雷达单一传感器的自注意力^[25]进行对比。实验平台的操作系统为Centos7,并带有型号为NVIDIA RTX Titan XP的GPU和Intel Xeon Silver 4210的CPU。

3.3 毫米波-激光雷达数据融合实验对比

首先,使用PointPillar点云快速编码网络框架作为基础网络,并在基础网络上加入基于注意力机制的毫米波-激光雷达点云数据融合模块进行实验对比,为证明实验结果的提升并非因为网络参数的增加而导致,额外加入了拥有相同参数数量的激光雷达点云数据的自注意力模块作为参考。实验结果如表2所示,基于注意力机制的毫米波-激光雷达点云数据融合方法的目标检测准确率与基础网络以及激光雷达的自注意力方法相比,取得了一定提升,基于注意力机制的数据融合方法的平均精度均值(mean Average Precision, mAP)高出基础网

络0.62个百分点,证明了本文中所提算法的有效性。

另外,从实验中可以看出,激光雷达自注意方法实验准确率比基础网络性能要低,初步推测是由于在点云柱快速编码过程中,其中的最大池化操作将点云柱内大量高相关性数据进行了简化,之后的注意力机制仅能够捕捉到点云柱间的上下文信息,因此,对于体积较大的目标,其所占点云柱数目较多,注意力机制能够对其检测性能加以提升,而体积较小的物体,所占点云柱数目较小,注意力机制无法捕捉该目标的上下文信息从而影响了该类目标的检测结果。在未来的工作中,将考虑这一因素对点云柱的快速编码模块进行优化。

表2 基准网络、自注意网络和注意力融合方法的AP与mAP对比

单位: %

Tab. 2 AP and mAP comparison of baseline network, self-attention network and attention fusion method

unit: %

方法	AP									mAP
	汽车	客车	卡车	行人	摩托车	拖车	工程车辆	栅栏	交通锥	
PointPillar基础网络	74.61	38.47	21.85	40.57	11.39	18.58	0.03	30.90	15.91	28.02
PointPillar+自注意力	74.03	38.35	20.02	38.99	10.30	19.97	0.20	26.82	15.11	27.09
传感器注意力融合	74.92	39.55	22.51	40.67	10.03	19.56	0.73	32.84	16.95	28.64

3.4 不同数据融合方法的对比实验

本节将基于注意力机制的融合方法与拼接、相加、相乘三种常见融合方法进行相比,实验结果如图8所示,可见基于注意力机制的融合方法的性能明显优于其他方法。

根据实验结果,基于注意力融合的目标检测方法性能优于拼接、加和、相乘融合目标检测方法。通过分析可知:一方面聚类后的毫米波雷达点位置误差较大,Nusence数据集中使用的ARS408型号毫米波雷达数据30 m外误差为0.4 m,因此部分与目标关联的毫米波雷达点并不在该目标上,而可能在目标周围;另一方面一个目标可能与多个毫米波雷达点相关联,使用拼接、加和、相乘融合只能融合对应的局部位置信息,而注意力融合能够通过全图的来学习毫米波雷达目标和激光雷达目标之间的关联。

网络的检测效果不佳,出现漏检及方向检测错误,而本文所提出的融合网络能够正确检测被树木遮挡的车辆,这是由于毫米波信号对树木草丛等的穿透性增强了融合网络对这类遮挡目标检测的性能。如第四行图片所示,基础网络在雨雾天气下由于空气水滴反射干扰更容易出现错检和漏检,而由于毫米波雷达对极端天气的鲁棒性更强,融合网络在雨雾天气下比基础网络也更为稳定。通过实验结果图对比可以发现,传感器注意力融合方法充分发挥了毫米波雷达可以穿透树木草丛、不受天气影响和探测距离远等特点,能有效提高网络检测性能。

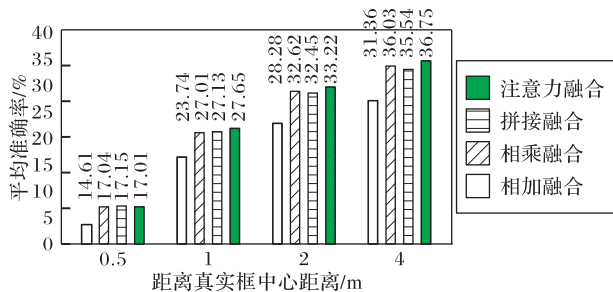


图8 注意力融合和拼接融合、相乘融合、相加融合的平均准确率对比
Fig. 8 Average accuracy comparison of attention fusion, concatenation fusion, multiply fusion and add fusion

3.5 实验结果可视化分析

本文对基础网络模型和注意力融合网络模型的检测效果进行鸟瞰图可视化,可视化范围为前后左右各50 m的x-y平面。如图9所示,图中框表示检测的目标物,框的闭合方向表示目标的方向。通过对比第一行第三列左上角,第二行第三列左上角,第三行第三列图片右下角可以发现,基础网络遗漏了部分的远处目标,而融合了毫米波雷达数据的网络模型能够很好地将检测出来,说明融合网络成功地将毫米波对远处目标的感知优势融入激光雷达特征图中,弥补了激光雷达对远处目标检测点数稀疏而造成的漏检。另外,在对比第一行和第二行图片右下角可以发现,当目标被树木遮挡后基础

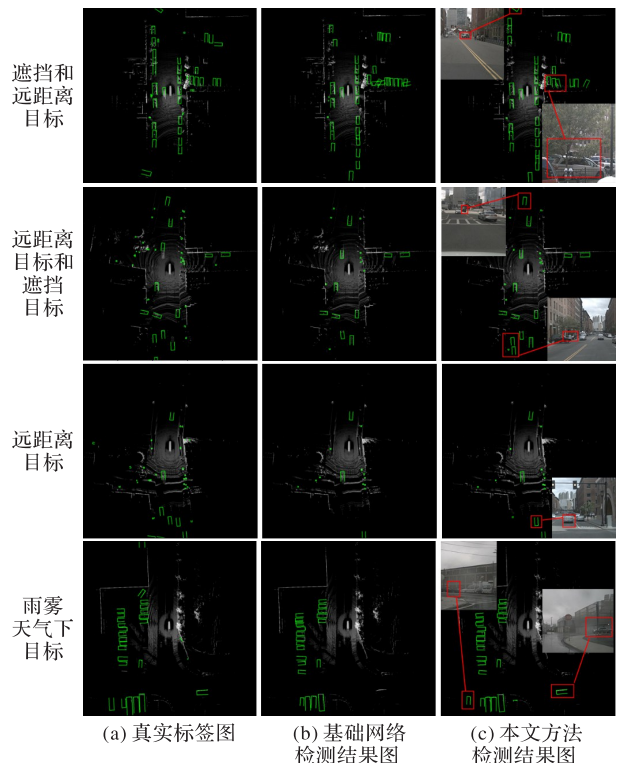


图9 基础网络与毫米波-激光雷达注意力融合检测结果对比
Fig. 9 Detection result comparison of basic network and millimeter wave radar-lidar attention fusion

3.6 与 nuScenes 上最先进方法的对比

本文在完整数据集下进行训练后对大型车辆的检测结果和目前数据集上公开的现有最先进算法 SARPNET^[30]、MonoDIS^[31]进行比较。其中SARPNET是基于激光雷达的目标检测,MonoDIS是基于摄像头的目标检测。通过实验结果表3可以发现,本文融合方法对车辆的检测准确率高于其他两种方法,在nuScenes数据集上取得了优异的表现。

表3 nuScenes数据集上本文融合方法和

SARPNET、MonoDIS的mAP对比 单位: %

Tab. 3 mAP comparison of the proposed fusion method, SARPNET and MonoDIS on nuScenes dataset unit: %

方法	AP				mAP
	汽车	客车	卡车	拖车	
SARPNET	59.9	19.40	18.7	18.0	29.0
MonoDIS	47.8	18.8	22.0	17.6	26.6
本文融合方法	77.7	44.9	28.8	37.6	47.3

4 结语

本文在点云快速编码网络PointPillar的基础上,提出了一种基于注意力机制的毫米波-激光雷达数据融合的目标检测方法,充分利用了毫米波雷达探测距离远、不受天气影响、可穿透树木和具有径向速度探测等特点,弥补了激光雷达的不足。本文的实验结果验证了所提方法的有效性,而且该方法也优于其他融合方法和自注意力方法。

考虑到本文使用的nuScenes数据集目标类的分布极不均匀,使得在一些类的检测结果准确率很低;另外本文毫米波雷达进行滤波只根据单通道数值进行过滤,而在毫米波雷达特征提取方法上借鉴的激光雷达特征提取方法,未充分考虑到毫米波雷达的稀疏性问题;以及点云柱快速编码过程中造成的小体积目标上下文信息丢失等问题,在未来的工作中将考虑利用数据增强及半监督学习等方法解决类数量不平衡问题,重新设计端对端的点云数据编码-特征提取-检测网络,从而进一步提升算法性能。

参考文献 (References)

- [1] QI C R, SU H, MO K, et al. PointNet: deep learning on point sets for 3D classification and segmentation[C]// Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2017: 652-660.
- [2] ZHOU Y, TUZEL O. VoxelNet: end-to-end learning for point cloud based 3D object detection[C]// Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 4490-4499.
- [3] 彭育辉,郑玮鸿,张剑锋. 基于深度学习的道路障碍物检测方法[J]. 计算机应用, 2020, 40(8): 2428-2433. (PENG Y H, ZHENG W H, ZHANG J F. Deep learning-based on-road obstacle detection method[J]. Journal of Computer Applications, 2020, 40(8): 2428-2433)
- [4] KU J, MOZIFIAN M, LEE J, et al. Joint 3D proposal generation and object detection from view aggregation[C]// Proceedings of the 2018 IEEE/RSJ Conference on Intelligent Robots and Systems. Piscataway: IEEE, 2018: 1-8.
- [5] CHEN X, MA H, WAN J, et al. Multi-view 3D object detection network for autonomous driving[C]// Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2017: 6526-6534.
- [6] 杜佳,宋春林. 一种改进的毫米波雷达多目标检测算法[J]. 通信技术, 2015, 48(7): 808-813. (DU J, SONG C L. A modified millimeter-wave radar multi-target detection algorithm [J]. Communications Technology, 2015, 48(7): 808-813)
- [7] SCHUMANN O, WÖHLER C, HAHN M, et al. Comparison of random forest and long short-term memory network performances in classification tasks using radar [C]// Proceedings of the 2017 Workshop on Sensor Data Fusion: Trends, Solutions, Applications. Piscataway: IEEE, 2017: 1-6.
- [8] SCHUMANN O, HAHN M, DICKMANN J, et al. Semantic segmentation on radar point clouds [C]// Proceedings of the 21st International Conference on Information Fusion. Piscataway: IEEE, 2018: 2179-2186.
- [9] 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. Red Hook, NY: Curran Associates Inc., 2017: 5005-5114.
- [10] DANZER A, GRIEBEL T, BACH M, et al. 2D car detection in radar data with PointNets [C]// Proceedings of the 2019 IEEE Intelligent Transportation Systems Conference. Piscataway: IEEE, 2019: 61-66.
- [11] GÖHRING D, WANG M, SCHNÜRMACHER M, et al. Radar/lidar sensor fusion for car-following on highways [C]// Proceedings of the 5th International Conference on Automation, Robotics and Applications. Piscataway: IEEE, 2011: 407-412.
- [12] HAJRI H, RAHAL M C. Real time lidar and radar high-level fusion for obstacle detection and tracking with evaluation on a ground truth [J]. International Journal of Mechanical and Mechatronics Engineering, 2018, 129(8): 821-827.
- [13] LEE H, CHAE H, YI K. A geometric model based 2D LiDAR/Radar sensor fusion for tracking surrounding vehicles [J]. IFAC-PapersOnLine, 2019, 52(8): 130-135.
- [14] NABATI R, QI H. RRPNet: radar region proposal network for object detection in autonomous vehicles [C]// Proceedings of the 2019 IEEE International Conference on Image Processing. Piscataway: IEEE, 2019: 3093-3097.
- [15] CHADWICK S, MADDERN W, NEWMAN P. Distant vehicle detection using radar and vision [C]// Proceedings of the 2019 International Conference on Robotics and Automation. Piscataway: IEEE, 2019: 8311-8317.
- [16] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot MultiBox detector [C]// Proceedings of the 2016 European Conference on Computer Vision, LNCS 9905. Cham: Springer, 2016: 21-37.
- [17] CHANG S, ZHANG Y, ZHANG F, et al. Spatial attention fusion for obstacle detection using mmWave radar and vision sensor[J]. Sensors, 2020, 20(4): No. 956.
- [18] JOHN V, MITA S. RVNet: deep sensor fusion of monocular camera and radar for image-based obstacle detection in challenging environments [C]// Proceedings of the 2019 Pacific-Rim Symposium on Image and Video Technology, LNCS 11854. Cham: Springer, 2019: 351-364.
- [19] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection [C]// Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2016: 779-788.
- [20] NOBIS F, GEISSLINGER M, WEBER M, et al. A deep learning-

- based radar and camera sensor fusion architecture for object detection[C]// Proceedings of the 2019 Workshop on Sensor Data Fusion: Trends, Solutions, Applications. Piscataway: IEEE, 2019: 1-7.
- [21] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]// Proceedings of the 31st International Conference on Neural Information Processing Systems. Red Hook, NY: Curran Associates Inc., 2017: 6000-6010.
- [22] 卢玲, 杨武, 王远伦, 等. 结合注意力机制的长文本分类方法[J]. 计算机应用, 2018, 38(5): 1272-1277. (LU L, YANG W, WANG Y L, et al. Long text classification combined with attention mechanism[J]. Journal of Computer Applications, 2018, 38(5): 1272-1277.)
- [23] XIAO T, XU Y, YANG K, et al. The application of two-level attention models in deep convolutional neural network for fine-grained image classification[C]// Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2015: 842-850.
- [24] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]// Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 7132-7141.
- [25] WANG X, GIRSHICK R, GUPTA A, et al. Non-local neural networks[C]// Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 7794-7803.
- [26] LANG A H, VORA S, CAESAR H, et al. PointPillars: fast encoders for object detection from point clouds[C]// Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2019: 12689-12697.
- [27] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]// Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE 2016: 770-778.
- [28] CAESAR H, BANKITI V, LANG A H, et al. nuScenes: a multimodal dataset for autonomous driving[C]// Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2020: 11618-11628.
- [29] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[C]// Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2017: 2999-3007.
- [30] YE Y, CHEN H, ZHANG C, et al. SARPNET: shape attention regional proposal network for LiDAR-based 3D object detection[J]. Neurocomputing, 2020, 379: 53-63.
- [31] SIMONELLI A, BULÒ S R, PORZI L, et al. Disentangling monocular 3D object detection[C]// Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE, 2019: 1991-1999.

This work is partially supported by the Youth Program of the National Science Foundation of China (61806186), the Construction Program of Fujian Intelligent Logistics Industry Technology Research Institute (2018H2001), the Programme of State Key Laboratory of Robotics and System (Harbin Institute of Technology) (SKLRS-2019-KF-15), the Quanzhou Science and Technology Program (2019C112, 2019STS08).

LI Chao, born in 1994, M. S. candidate. His research interests include 3D object detection, sensor fusion.

LAN Hai, born in 1988, M. S., research assistant. His research interests include machine learning and pattern recognition and their applications in medical imaging

WEI Xian, born in 1986, Ph. D., research fellow. His research interests machine learning and pattern recognition and their applications in unmanned system.