

基于毫米波雷达和视觉的目标检测方法

赵越坤¹, 罗素云¹, 魏丹¹, 王琦²

(1. 上海工程技术大学 机械与汽车工程学院, 上海 201620;

2. 滨海县科技馆, 江苏 盐城 224500)

摘要:为了提高目标检测网络对远距离目标的检测能力,以及改善由单一视觉传感器的感知系统抗环境干扰能力差的问题,提出了一种基于毫米波雷达和视觉传感器的多源目标检测方法,视觉图像经由多个毫米波雷达获取的点云信息增强后进行检测。首先,对增强同一视觉图像的多个雷达点云进行数据拼接,通过坐标转换将雷达点云投影至视觉图像平面,并对超出雷达探测距离的异常点和经过坐标转换后位于视野外部的无效点进行剔除,生成雷达点云图像。然后,根据雷达点云图像中各雷达点的位置与深度信息,形成对应的感兴趣区域,生成雷达特征图像。最后,将雷达特征图像与视觉图像在 YOLOv4 网络中的主干特征提取部分进行多级融合,并使用通道注意力机制分配通道权重。实验结果表明,基于雷达增强的目标检测网络的平均检测精度提高了 10.93%,并提高了远距离目标和弱光照条件下的检测精度,具有比传统机器视觉的目标检测方法更好的可靠性和鲁棒性。

关键词:深度学习;目标检测;传感器融合;视觉增强;毫米波雷达

中图分类号:TP391

文献标识码:A

文章编号:1673-629X(2023)06-0035-06

doi:10.3969/j.issn.1673-629X.2023.06.006

Target Detection Method Based on Millimeter-wave Radar and Vision

ZHAO Yue-kun¹, LUO Su-yun¹, WEI Dan¹, WANG Qi²

(1. School of Mechanical and Automotive Engineering, Shanghai University of Engineering Science,

Shanghai 201620, China;

2. Binhai County Science and Technology Museum, Yancheng 224500, China)

Abstract: In order to improve the ability of target detection network to detect long-range targets and to improve the poor anti-environmental interference ability of the perception system by a single visual sensor, a multi-source target detection method based on millimeter-wave radar and visual sensor is proposed. The visual image is detected after the point cloud information obtained by multiple millimeter-wave radar is enhanced. First, several radar point clouds that enhance the same visual image are stitched, projected to the plane of the visual image through coordinate transformation, and the abnormal points beyond the radar detection distance and invalid points outside the field of view are eliminated to generate the radar point cloud image. Then, according to the position and depth information of each radar point in the radar point cloud image, the region of interest is formed, and the radar feature image is generated. Finally, the backbone of YOLOv4 network is fused with the radar feature image and the visual image, and the channel weight is allocated using the channel attention mechanism. The experiment shows that the average detection accuracy of target detection network based on radar enhancement is improved by 10.93%, and the detection accuracy of long-range target and weak illumination condition is improved. It has better reliability and robustness than traditional machine vision target detection method.

Key words: deep learning; target detection; sensor fusion; vision enhancement; millimeter-wave radar

0 引言

环境感知作为自动驾驶领域的关键技术之一,主要是通过传感器快速获取车辆周围的环境,是实现自动驾驶规划与控制的基础^[1]。

在自动驾驶车辆的环境感知任务中常用的传感器有摄像头、激光雷达、毫米波雷达、超声波雷达等^[1]。视觉传感器是目前环境感知领域中最广泛使用的传感器,但在暴雨雪天气以及其他恶劣路况条件下能见度

收稿日期:2022-08-10

修回日期:2022-12-11

基金项目:国家自然科学基金项目(62101314)

作者简介:赵越坤(1999-),男,硕士,研究方向为传感器融合;通讯作者:罗素云(1975-),女,副教授,研究方向为自动驾驶、汽车环境感知及控制。

降低,甚至会被雨雪泥泞遮挡而无法使用^[2]。只依靠视觉传感器的感知系统不能够完全保障车辆的安全行驶。

激光雷达测量分辨率高,但在极端天气下性能较差,价格昂贵,目前未广泛使用。毫米波雷达抗干扰能力强,但点云稀疏,无法对目标进行精准建模,且没有目标高度的探测能力^[3]。

在环境感知系统中,通常使用两种或多种传感器同时进行数据采集,再进行信息融合,以实现优势互补,提高抗干扰能力和可靠性^[4]。其中毫米波雷达与视觉结合的方式对环境的鲁棒性较好,并且成本可控。

毫米波与视觉融合算法根据融合的阶段不同可分为前融合和后融合,前融合在初始层将数据融合,只使用一种感知算法,对融合后的综合数据进行感知;后融合中传感器各自有一套独立的感知算法,在主处理器将感知结果进行融合^[4]。文献[5]通过帧差法实现了动态障碍物检测,并根据 ROI(感兴趣区域)大小进行危险级别分类与识别。文献[3]对图像进行形态学开闭运算,增强了融合网络对车辆的夜间检测能力。文献[6-8]提出了先通过雷达点云在图像上形成感兴趣区域,再对区域内进行验证的特征级融合策略,其中文献[8]针对毫米波雷达横向误差较高,通过求和归一化互相关匹配算法和对称检验,对点云的横向位置进行修正。文献[9]提出了中间层融合的概念,通过视锥匹配的方法将点云与一次预测目标进行匹配,并对匹配结果进行二次预测以获得最终结果。

该文提出了一种基于毫米波雷达点云和图像的目标检测方法,该方法首先将毫米波雷达点云信息通过坐标转换投影至图像平面,形成雷达特征图像,再将雷达特征图像作为辅助通道与视觉图像进行分级融合,并通过注意力机制分配通道权重,以提高网络的检测精度。通过实验对该融合算法的有效性进行了验证。

1 毫米波雷达特征图像模型

1.1 点云信息预处理

该文所使用的图像信息由单个视觉传感器获取,点云信息由多个雷达传感器共同获取。为增强图像信息的丰富度,对雷达传感器与视觉传感器使用多重对应关系,使一个视觉传感器所获取的图像信息,由多个雷达传感器整合后的点云数据进行增强。根据传感器配置和安装位置,正前视觉传感器获取到的图像信息由左前、正前和右前毫米波雷达的点云进行增强;正后视觉传感器获取到的图像信息由左后、右后毫米波雷达的点云进行增强。

雷达点云的预处理过程如图 1 所示,确定传感器间的对应关系后,点云需经过坐标转换等步骤,首先获

取位于雷达坐标系下的点云,再通过外参转换至车辆坐标系下,最后再从车辆坐标系根据投影关系转移到像素坐标系中^[10]。

该文对雷达点云进行了两种滤波,分别过滤掉不在雷达有效距离内的异常点和不在视觉传感器感受野范围内的无效点,并根据深度对雷达点按照深度重新排序,以保证增强区域的灰度值递进。

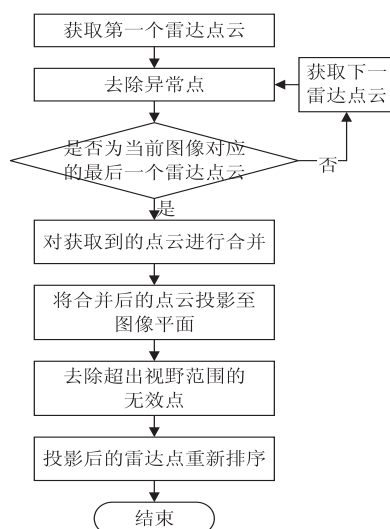


图 1 毫米波雷达点云预处理过程

1.2 雷达 ROI 模型

由于毫米波雷达无法准确获得物体的高度信息,实际得到的点云纵坐标位置偏低,会出现近处物体未被增强区域完全覆盖的问题,因此,首先对预处理后的点云进行了归一化。距离越远的目标点云,归一化产生的误差越小,由于远处的雷达点深度增加,对应的目标投影在像素平面的面积变小,因此受归一化的影响较小。ROI 模型中心的纵坐标归一化计算如公式(1)所示:

$$Y = \frac{\sum y}{n} \quad (1)$$

式中, Y 为归一化后的 ROI 中心点的纵坐标, y 为图像中每个雷达点的纵坐标, n 为雷达点个数。

完成雷达点的纵坐标修正后,以归一化后的雷达点为中心构建 ROI 模型。ROI 形状为矩形,包括灰度值与尺寸两个参数,灰度值与雷达点的深度呈正相关,雷达点深度越大,其在像平面内灰度值越高;ROI 的尺寸与雷达点深度呈负相关,雷达点深度越大,其在像平面内所形成的 ROI 越小。ROI 模型的灰度值计算如公式(2)所示:

$$D = \frac{d}{\max(d)} \times i \quad (2)$$

式中, D 为当前雷达点对应 ROI 的灰度值, d 为当前雷达点深度, $\max(d)$ 为当前图像中所包含雷达点的深度最大值, i 为比例系数,该文取 10。ROI 模型的尺

寸参数计算如公式(3)和公式(4)所示:

$$W = \left[2 \times \frac{\min(d) + \max(d)}{\max(d) - \min(d)} \right] \times d - \left[\frac{\max(d)^2 + \min(d)^2}{\min(d) - \max(d)} \right] + \max(d) + b \quad (3)$$

$$H = 1.2 \times W \quad (4)$$

式中, W 为当前雷达点对应 ROI 的宽度值, H 为当前雷达点对应 ROI 的高度值。 d 为当前雷达点深度, $\max(d)$ 为当前图像中所包含雷达点的深度最大值, $\min(d)$ 为当前图像中所包含雷达点的深度最小值, b 为偏置系数, 文中 b 取 180。

现有的其他用于特征增强的 ROI 模型主要是线增强模型, 这种模型以投影在图像平面上的雷达点为中心, 以在世界坐标系下预设高度的投影作为增强线的长度, 雷达点深度越大, 增强线长度越短。线增强模型参数量较大, 为了在网络上获得更好的训练效果, 需要使用复杂的训练方式, 并且由于毫米波雷达点云稀疏, 需要整合多个时间戳的点云数据, 增加了增强效果的不确定性。文中的 ROI 模型为矩形增强模型, 参数量更少, 增强区域更大, 实现更加简单, 且无需使用额外的数据弥补单帧点云数量不足。

ROI 模型宽高比是 1:1.2, 正方形 ROI 对近距离的高大物体覆盖效果并不理想, 这种矩形 ROI 与正方形的 ROI 相比对目标的覆盖效果更好。使用细长型的 ROI 也可改善 ROI 重叠的问题, 使每一个雷达点对应的 ROI 更关注相应的目标。

如图 2 所示, (a) 为原始图像, (b) 为雷达点云转换到像素坐标下的图像, (c) 为雷达 ROI 模型图像, (d) 为雷达 ROI 模型图像叠加到原始图像上的示意图。

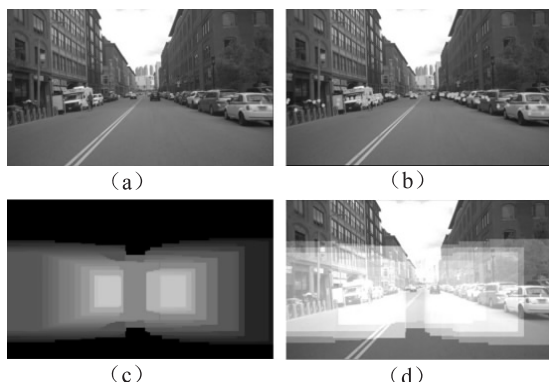


图 2 ROI 模型

1.3 雷达特征图像生成过程

该文所使用的雷达增强特征保存形式为图像格式, 首先遍历数据集中的关键帧, 得到关键帧中有对应关系的雷达点云文件和图像文件, 随后分组遍历关键帧中的传感器数据, 对雷达点云数据进行预处理, 预处理环节包括点云的坐标转换, 点云过滤, 雷达点重新排

序等步骤, 然后根据 1.2 节中的 ROI 模型得到雷达特征图像。生成雷达特征图像的流程如图 3 所示。

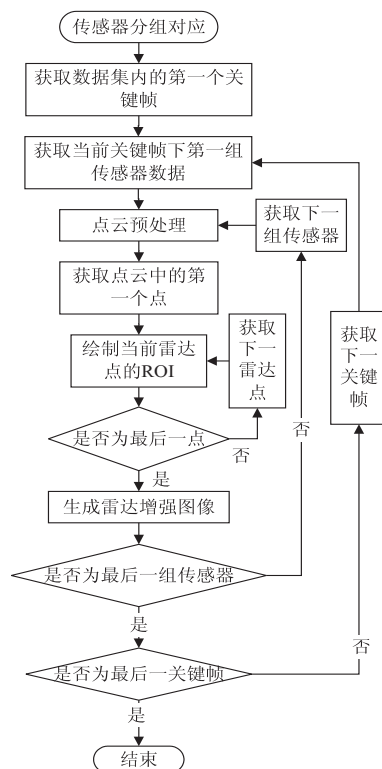


图 3 雷达图像生成过程

2 基于点云和图像的 YOLOv4

基于 YOLOv4^[11] 目标检测网络来验证毫米波雷达特征图像模型的有效性。YOLOv4 网络的主干部分主要由多个残差结构块堆叠而成, 基于 YOLOv4 目标检测网络构建包含附加雷达通道的增强网络更易于验证雷达增强图像模型的有效性。

该文使用的网络结构主要分为四个部分, 分别是主干特征提取模块、雷达特征模块、增强特征提取模块以及预测模块。基于点云和图像的网络结构如图 4 所示。网络的输入尺寸为 $416 \times 416 \times 4$, 其中前三个通道为视觉特征, 第四个通道为雷达特征。在预处理阶段将雷达特征作为图像特征的第四通道输入网络, 一方面易于在批次训练时完成雷达特征与图像特征的匹配, 网络读取数据效率高; 另一方面雷达特征图与图像特征尺寸相同, 易实现主干特征提取模块在原 CSPDarknet53 的基础上进行改进, 保留了由卷积、正则化和 Mish 激活函数组成的基本单元。CSPnet 是该主干网络的主要组成部分, 该结构将残差块拆分为两个部分, 其中的主干部分继续进行残差块堆叠, 另一部分则以残差边的形式连接到结构块的输出部分。

图像完成第一次卷积正则化激活操作后, 形成长宽相同但通道数变为 32 的图像特征, 此时将形成的图像特征与第一级雷达特征进行初次融合, 形成 $416 \times$

416×33 的初级融合特征。在形成初级融合之后加入了通道注意力^[12]模块,将 33 个通道进行特征压缩,作为融合特征层的评价分数,对原始的融合特征层逐通道进行加权。通过通道注意力机制可以使网络对图像特征与融合特征的重要性进行调节,将重要的特征增强,不重要的特征减弱。

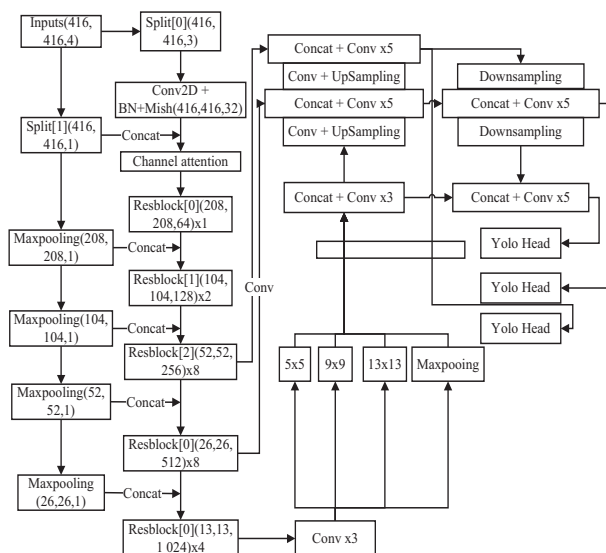


图 4 基于雷达特征图像的 YOLOv4 网络结构

在增强特征提取模块和预测模块保留了 YOLOv4 的基本结构,在增强特征提取部分使用了 SPP^[13]结构和 PANet^[14]结构,其中 SPP 结构以不同尺寸对特征图进行池化形成新的特征,扩大了网络的感受野;PANet 通过多次的上采样和下采样,实现了对特征的重复利用。在预测模块中保留了三个尺度的预测模式,用于对不同尺度的目标进行检测。

3 实验结果与分析

3.1 实验环境

该模型基于 Pytorch 深度学习框架实现,训练以及测试均基于 Linux 系统, Ubuntu16.04, CUDA11.1, Python3.10, Pytorch1.10。硬件条件为 Intel(R) XEON(R) W-2150B CPU @ 3.00 GHz, GPU 为 GeForce RTX 3070×2, 16 GB 显存。

训练过程中不载入预训练权重,训练阶段采用冻结训练策略,先冻结主干网络训练,随后解冻主干网络,最后调整学习率使网络收敛。最大学习率为 0.001,最小学习率为 0.0001,冻结阶段的训练批次为 16,解冻阶段的训练批次为 4。模型训练完成后对特征增强的网络进行测试,测试阶段非极大值抑制(NMS)阈值为 0.5。

3.2 实验数据

为验证基于点云和图像的目标检测方法的可行性,基于 Nuscenes 数据集^[15]进行了实验验证,并与其

他目标检测网络进行比较。数据采集车辆配备 6 台采样频率为 12 Hz 的彩色周视相机,5 台 77 GHz 的毫米波雷达等多种传感器,数据集内目标种类多,数据量充足,满足实验要求。数据采集车辆传感器安装位置及设定的增强关系如图 5 所示。

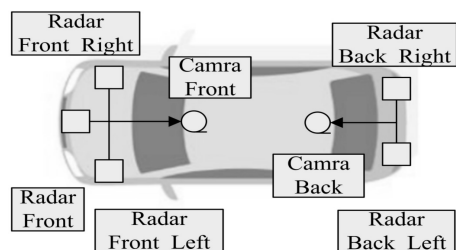


图 5 传感器位置及对应关系

该文选取正前和正后的视觉传感器获取的图像作为数据集,左前、右前、正前毫米波雷达对应正前视觉传感器,左后、右后毫米波雷达对应正后视觉传感器。去除前 170 个场景中的无效图片和雷达点云文件后,共计获取到 12 420 个有效雷达-视觉特征对,并按照 8:1:1 的比例划分训练集、验证集和测试集。完整数据集中共有 23 种类别,在训练过程中保留了日常行车路况中出现频率较高的 6 个类别,包括行人、车辆、货车、救援车辆、自行车和巴士。

3.3 模型训练

基于点云和图像的 YOLOv4 网络和基于图像的 YOLOv4 网络在训练过程中的损失函数如图 6 和图 7 所示。两个网络在起始阶段都可实现损失的快速下降,由于学习率设置较大,损失有一定波动,但总体趋

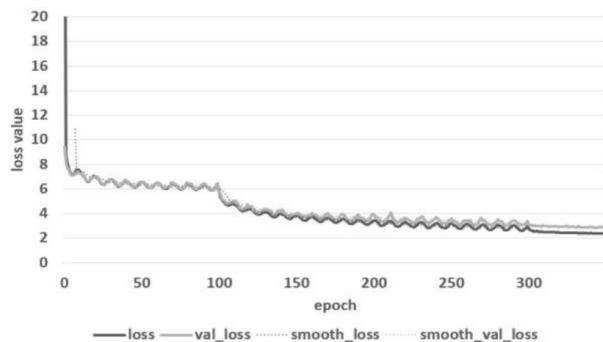


图 6 基于点云和图像的 YOLOv4 损失函数

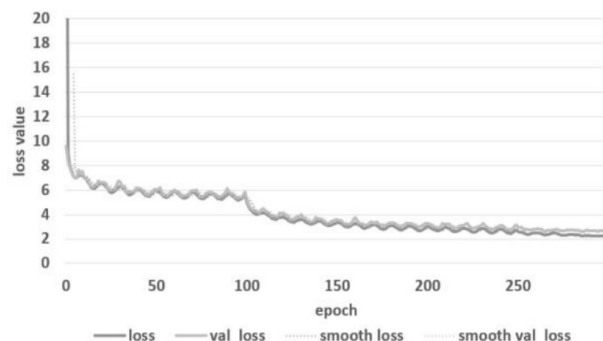


图 7 基于图像的 YOLOv4 损失函数

势下降平稳。在 100 次迭代后网络解冻,两个网络的损失函数都发生了较大幅度的下降。再次降低学习率后,两个网络都进入了收敛状态。

基于点云和图像的 YOLOv4 网络在经过 350 次迭代后收敛,基于图像的 YOLOv4 网络在经过 300 次迭代后收敛。基于点云和图像的 YOLOv4 损失函数下降趋势与基于图像的 YOLOv4 相似,但由于雷达特征的引入,需要更多的批次才能进入收敛状态。最终基于点云和图像的 YOLOv4 网络的训练损失值为 2.37,在验证集上的损失值为 2.87,基于图像的 YOLOv4 网络训练损失值为 2.42,在验证集上的损失值为 2.75。

3.4 结果分析

该文的研究重点在于包含雷达通道的目标检测模型相对于纯视觉的目标检测模型的检测精度提升,主要采用 AP(平均精度)和 mAP(整体平均精度)作为评价指标。AP 为单一类别在 0 到 1 之间的所有召回值上平均的精度,mAP 为各类别 AP 值的平均值。

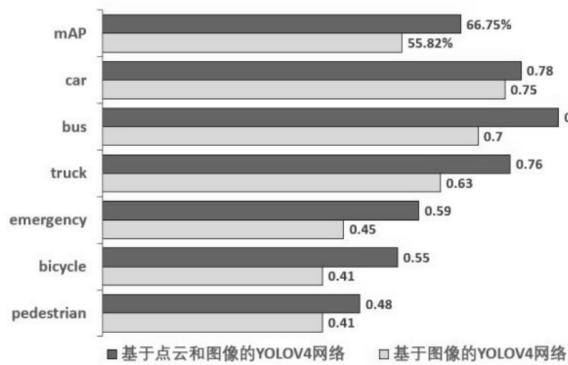


图8 基于点云和图像的 YOLOv4 网络与基于图像的 YOLOv4 网络各类别 AP 值和 mAP 值对比

从图 8 中可以看出,基于点云和图像的 YOLOv4 网络整体平均精度与基于图像的 YOLOv4 网络相比有较大提升,由 55.82% 提升到了 66.75%。具体到每个类的 AP 情况,对 car 类提升效果不明显,而对 bus 和 truck 类提升效果显著。其中 car 类提升约 3%,bus 和 truck 分别提升了 15% 和 13%。这种现象可能是因

为大部分客车和货车体积较大,且为大面积的金属材质,雷达特征强。而 car 类的精度提升不大,可能是因为数据集内该类目标数量较多,而基于图像的 YOLOv4 网络对这类目标的检测性能足够优秀。

为进一步验证所提出的基于点云和图像的目标检测算法的有效性,与文献[16-17]使用检测准确率作为评价指标进行对比分析。由表 1 可见,文中方法的检测效果与其他方法相比有一定提升。

表1 不同算法的检测结果统计

算法	检测种类	检测数目	检测率/%
YOLOv4	车辆	4 615	91.47
文中方法	车辆	4 615	93.42
文献[16]	车辆	1 636	93.2
文献[17]	车辆/人	563	92.5

文献[16-17]中的融合策略为决策级融合策略,这种融合方法需要雷达模块和视觉模块分别进行独立检测,再进行加权融合,对于多个不同分布的数据较难找到合适的权重。相比决策级融合,文中方法更侧重使用雷达特征来增强视觉特征的丰富性,只使用视觉作为最终检测头,算法的规模更小,且无需因环境的变化进行权重调整。

基于点云和图像的 YOLOv4 网络与基于图像的 YOLOv4 网络实际检测结果如图 9 所示,第一组场景为远距离目标检测效果对比。从中可以看出,基于点云和图像的 YOLOv4 网络和基于图像的 YOLOv4 网络都可以检测到近距离的行人目标,但基于图像的 YOLOv4 网络对距离较远的车辆目标发生了漏检。第二组场景为光照条件较差场景中检测效果对比,从中可以看出,在逆光条件下,基于图像的 YOLOv4 网络发生了漏检和误检,将左侧的行人识别为了自行车。第三组场景为置信度对比,从中可以看出,两组网络都完成了正确识别,由于雷达特征的引入,基于点云和图像的网络提高了视野边缘目标的置信度。



图9 检测效果(上:基于点云和图像的 YOLOv4,下:YOLOv4)

测试结果表明,所提出的基于点云和图像的 YOLOv4 算法提升了平均检测精度,在各个类别上均有一定提升。在具体场景中,该算法对远距离目标的检测能力也有提升,并提高了在弱光照场景中的检测能力,验证了雷达特征模型的有效性。

4 结束语

为改善目标检测任务中单一传感器对环境的适应能力,提高目标检测网络的检测性能,提出了一种毫米波雷达特征图像模型和一种基于该模型和视觉传感器的多源目标检测方法,提升了检测精度。实验表明:

(1)基于点云和图像的 YOLOv4 目标检测网络采用逐级下采样的方式融合雷达特征,并使用通道注意力机制能够有效提升网络的检测性能。在 Nuscenes 数据集中总体精度提升为 10.93%。

(2)对比各类别的提升效果,基于点云和图像的 YOLOv4 目标检测网络对卡车、巴士这种雷达特征明显的目标提升效果最为显著,其中对卡车的检测精度提升约 13%,对巴士提升约 15%。

参考文献:

- [1] 张炳力,詹叶辉,潘大巍,等. 基于毫米波雷达和机器视觉融合的车辆检测[J]. 汽车工程,2021,43(4):478-484.
- [2] NOBIS F, GEISLINGER M, WEBER M, et al. A deep learning-based radar and camera sensor fusion architecture for object detection[C]//2019 sensor data fusion:trends,solutions,applications (SDF). Bonn:IEEE,2019:1-7.
- [3] 金立生,程 蕾,成 波. 基于毫米波雷达和机器视觉的夜间前方车辆检测[J]. 汽车安全与节能学报,2016,7(2):167-174.
- [4] 宋维堂,张 鸽. 基于智能车辆的多传感器数据融合算法研究与分析综述[J]. 现代交通技术,2012,9(3):82-85.
- [5] 翟光耀,陈 蓉,张剑锋,等. 基于毫米波雷达和机器视觉信息融合的障碍物检测[J]. 物联网学报,2017,1(2):76-83.
- [6] 刘 雷. 基于毫米波雷达与视觉融合的前方车辆检测[D]. 天津:天津工业大学,2019.
- [7] 高继东,焦 鑫,刘全周,等. 机器视觉与毫米波雷达信息融合的车辆检测技术[J]. 中国测试,2021,47(10):33-40.
- [8] 胡延平,刘 菲,魏振亚,等. 毫米波雷达与视觉传感器信息融合的车辆跟踪[J]. 中国机械工程,2021,32(18):2181-2188.
- [9] NABATI R, QI H. CenterFusion:center-based radar and camera fusion for 3D object detection[C]//2021 IEEE winter conference on applications of computer vision (WACV). Waikoloa:IEEE,2021:1526-1535.
- [10] ZHANG Zhengyou. Flexible camera calibration by viewing a plane from unknown orientations[C]//Proceedings of the seventh IEEE international conference on computer vision. Corfu:IEEE,1999:666-673.
- [11] BOCHKOVSKIY A, WANG C Y, LIAO H M. YOLOv4:optimal speed and accuracy of object detection[J]. arXiv:2004.10934,2004.
- [12] JIE H, LI S, GANG S, et al. Squeeze-and-excitation networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2020,42(8):2011-2023.
- [13] HE K, ZHANG X, REN S, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence,2014,37(9):1904-1916.
- [14] LIU S, QI L, QIN H, et al. Path aggregation network for instance segmentation[C]//2018 IEEE/CVF conference on computer vision and pattern recognition. Salt Lake City:IEEE,2018:8759-8768.
- [15] CAESAR H, BANKITI V, LANG A H, et al. nuScenes:a multimodal dataset for autonomous driving[C]//2020 IEEE/CVF conference on computer vision and pattern recognition (CVPR). New York:IEEE,2020:11618-11628.
- [16] 甘耀东,郑 玲,张志达,等. 融合毫米波雷达与深度视觉的多目标检测与跟踪[J]. 汽车工程,2021,43(7):1022-1029.
- [17] 陈文龙. 基于毫米波雷达与视觉融合的目标检测与识别[D]. 上海:同济大学,2019.