计算机应用
Journal of Computer Applications
ISSN 1001-9081,CN 51-1307/TP

《计算机应用》网络首发论文

题目：基于注意力的毫米波-激光雷达融合目标检测
作者：李朝，兰海，魏宪
收稿日期：2020-09-01
网络首发日期：2020-12-31
引用格式：李朝，兰海，魏宪．基于注意力的毫米波-激光雷达融合目标检测[J/OL]．计算机应用. https://kns.cnki.net/kcms/detail/51.1307.TP.20201229.1700.012.html

Journal of Computer Applications 计算机应用 ISSN 1001-9081 CODEN JYIIDU http://www.joca.cn

doi:10.11772/j.issn.1001-9081.2020081334

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，山西 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遮挡目标、远距离目标和复杂天气场景下目标的问题，提出了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多帧聚合和空间对齐；其次对两者预处理后的数据分别进行点云柱快速编码，转换成伪图像；最后，通过中间卷积层进行两者传感器特征提取，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 Nuscence 数据集中对 9 类目标检测的准确率比基础网络提高 0.3%至 2.9%不等，9 类目标的平均准确率提高 0.62%，同时高于拼接融合、相乘融合、相加融合方法。可视化结果显示该方法提高了网络在遮挡场景、远处目标和雨雾天气下检测的鲁棒性，证明了该方法的有效性。

关键词: 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 radar-lidar fusion

LI Chao[1,2], LAN Hai[1*], WEI Xian[1]

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

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

Abstract: Millimeter-wave Radar is an important tool to complementally enhance the performance of Lidar in self-driving car. Focused on problems of missing occluded targets, distant targets and targets in extreme weather scenarios when using lidar for object detection in self-driving car, an attention-based Radar-Lidar data fusion object detection network was proposed in this work. PointPillar was employed to encode both the Radar and Lidar data into pseudo images, and the attention mechanism is then adopted to fuse the feature map. Experiments on Nuscenes datasets were carried out for comparison, where the detection accuracy of all 9 types objects, with the proposed approach, are increased by 0.3% to 2.9%, and 0.62% on average. In addition, Radar-Lidar attention fusion performs better than concatenation fusion, multiply fusion and add fusion. The visualization results showed that the proposed method could improve the robustness of the network when occluded targets, distant targets and targets surrounded by rain and fog appeared.

Keywords: 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，毫米波雷达。激光雷达精度高，探测距离较远，受天气影响小，数据较稀疏。摄像头图像具有丰富的颜色信息，受天气和光照影响较大。毫米波雷达精度较低，探测距离远，受天气影响极小，数据稀疏。目前有基于单个传感器做目标检测，也有基于多个传感器融合进行目标检测。不同传感器数据进行融合，提高了无人驾驶系统的鲁棒性和冗余性是未来重要的工作。

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

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）；机器人与系统国家重点实验室（HIT）资助项目(SKLRS-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，以及在医疗影像中应用等；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，以及在无人系统中的应用等。

献[1]提出 Pointnet 是具有开创性工作，真正的实现了无序的点云的端到端学习。Pointnet 网络通过池化操作来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送进网络，另外是通过数据预处理的方法。例如文献[2][3]网络通过将无序的点云划分到有序的空间体素的方法解决点云数据的无序性问题，之后通过 3D 卷积特征提取，但是 3D 卷积计算量较为巨大。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更稀疏，但信息比较丰富。文献[6]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，但是误差大，且无法检测出目标的属性。文献[7]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网络(Long Short Term Memory networks, LSTM)对目标进行分类。文献[8]则将整个原始雷达数据作为输入，采用 Pointnet++[9]的基础架构，得到每一次毫米波雷达反射的各个类概率，不需要进行聚类算法和人为选择特征。文献[10]认为虽然毫米波雷达数据比激光雷达稀疏，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，开创性的使用雷达数据的位置、速度和雷达截面积信息在 Pointnet 框架上实现了检测车辆 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，单一传感器感知能力是有限的，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 3 种：数据级融合、特征级融合和目标级融合。文献[11-13]结合激光雷达的精度高和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]提出的 RRPN 网络中，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]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，对特征图对应元素进行相加融合，对融合后的特征图使用 SSD[16]框架进行目标检测。与采取投影方法不同，文献[17]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，然后对转换后的毫米波雷达和图像相乘融合。文献[18]提出毫米波雷达和图像融合网络 RVNet，RVNet 使用是基于 YOLO[19]检测框架特征图拼接融合网络，并且为提高检测精度针对大目标和小目标分别设有两个输入分支和输出分支。文献[20]的工作中，作者提出了毫米波雷达和图像融合的 CRF-Net 网络，作者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，并且提出一种叫作 BlackIn[20]的训练策略确保融合网络收敛。

除了传感器融合方法，注意力机制也被应用到图像领域中。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。注意力机制在图像领域也取得了巨大的进展。文献[23]提出的两级注意力模型应用于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，从而更为精确地利用多层次信息。文献[24]提出通道注意力机制，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]认为卷积神经网络神经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。论文作者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之间关系。

在本文中，针对激光雷达进行目标检测时对树木遮挡目标、雨雪天气中目标和远处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，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，能够极大的增强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，加强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]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稀疏，但毫米波雷达数据在目标速度和雷达散射截面 (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 4 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。

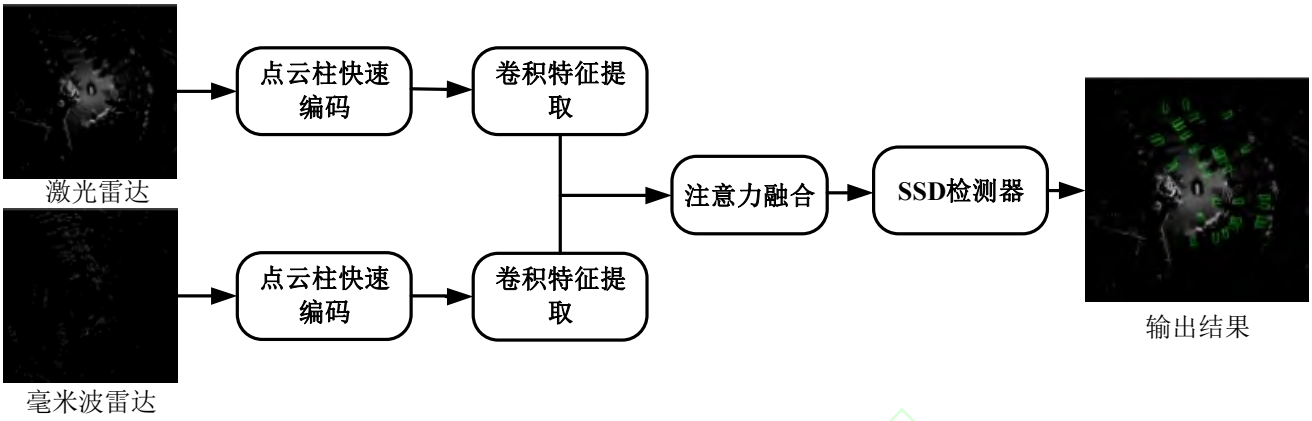


图1 传感器注意力机制融合网络框架

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的融合，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 3 维点云空间中均匀生成 400×400 个立方柱体，即每个柱体的底面大小为 0.25 m ×0.25 m，高度限制为 10 m；每个点云柱中的点数约束为 N 个，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_{rl}, y_{rl}, z_{rl})、补偿速度 (V_{xcomp}, V_{ycomp}) 及目标雷达散射面积 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。相比于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都 z_r 为 0，在对毫米波雷达雷达特征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 $(x_{rl}, y_{rl}, z_{rl}, V_{xcomp}, V_{ycomp}, x_c, y_c, RCS)$。编码后的毫米波雷达为形状 (D_r, P, N) 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。

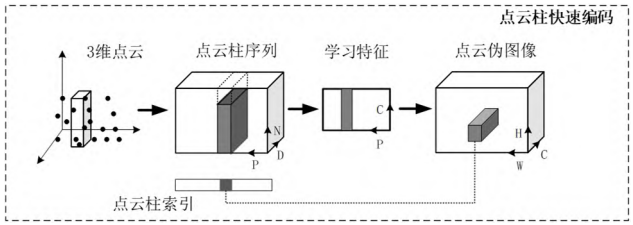


图2 点云柱快速编码

Fig. 2 Point cloud pillar fast encode

1.2 注意力融合

在本文中提出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，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，如公式(1)所示，$\boldsymbol{X}_l \in \boldsymbol{R}^{C\times N}$ 表示激光雷达特征图，$\boldsymbol{X}_r \in \boldsymbol{R}^{C\times N}$ 表示毫米波雷达特征图，$\boldsymbol{O}$ 表示注意力融合后的激光雷达特征图。

$$attention\left(\boldsymbol{X}_{\mathrm{l}},\boldsymbol{X}_{\mathrm{r}}\right)=\boldsymbol{O} \tag{1}$$

注意力融合过程中，定义式(2)：

$$\begin{cases}\boldsymbol{Q}=\boldsymbol{W}_{\mathrm{q}}\cdot\boldsymbol{X}_{\mathrm{l}}\\ \boldsymbol{K}=\boldsymbol{W}_{\mathrm{k}}\cdot\boldsymbol{X}_{\mathrm{r}}\\ \boldsymbol{V}=\boldsymbol{W}_{\mathrm{v}}\cdot\boldsymbol{X}_{\mathrm{l}}\end{cases} \tag{2}$$

其中，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 $\boldsymbol{Q}$ 和 $\boldsymbol{V}$。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 $\boldsymbol{K}$。使用点乘作为 $\boldsymbol{Q}$ 与 $\boldsymbol{K}$ 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 $\boldsymbol{Q}$ 与毫米波雷达特征图所对应的 $\boldsymbol{K}$ 之间的关系权重矩阵 $\boldsymbol{A}$，$\boldsymbol{A}$ 中各项 $\boldsymbol{a}_{ij}$ 计算如公式(3)所示。

$$\boldsymbol{a}_{ij}=\frac{\exp\left(\boldsymbol{q}_i^{\mathrm{T}}\cdot\boldsymbol{k}_j\right)}{\sum_{j=1}^{\mathrm{M}}\exp\left(\boldsymbol{q}_i^{\mathrm{T}}\cdot\boldsymbol{k}_j\right)} \tag{3}$$

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

$$\boldsymbol{O}_j=\sum_{j=1}^{\mathrm{M}}\boldsymbol{A}_{ij}\cdot\boldsymbol{V}_j \tag{4}$$

另外，借鉴残差模块的概念[27]，如公式(5)所示，将融合结果 $\boldsymbol{O}$ 乘上比例系数λ并加上激光雷达特征图 $\boldsymbol{X}_{\mathrm{l}}$，得到最终输出结果 $\boldsymbol{y}$。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，随着训练的进行逐渐增大注意力在输出中的影响。

$$\boldsymbol{y}=\boldsymbol{X}_{\mathrm{l}}+\lambda\boldsymbol{O} \tag{5}$$

除了传感器注意力融合方法，本文进行了拼接融合，相乘融合，相加融合实验进行对比，各个方法在网络中的融合位置相同。参考文献[20]进行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 $\left(2C\times W\times H\right)$ 融合特征图。融合后的特征图通过 1×1 卷积进行降维到原来的维度。

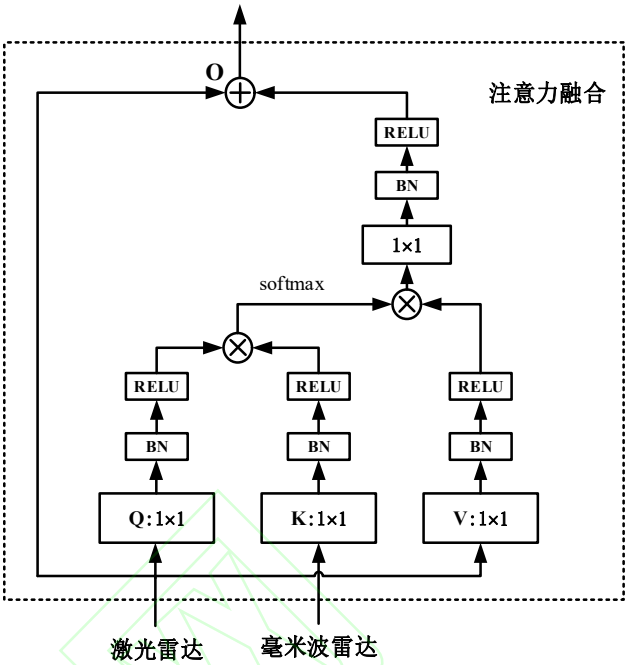


图3 激光雷达和毫米波雷达注意力融合

Fig. 3 Lidar and radar attention fusion

参考文献[15]进行特征图相加融合实验。本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。

参考文献[17]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 $\boldsymbol{X}_{\mathrm{r}}$，比激光雷达特征图 $\boldsymbol{X}_{\mathrm{l}}$ 有更多的元素为 0，因此对毫米波雷达特征图为 0 的元素进行加 1 操作如式(6)所示，得到毫米波特征图 $\boldsymbol{X}_{\mathrm{r}}^{'}$，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。

$$\boldsymbol{X}_{\mathrm{r}}^{'}=\begin{cases}\boldsymbol{X}_{\mathrm{r}}+1 & \text{if } \boldsymbol{X}_{\mathrm{r}}=0\\ \boldsymbol{X}_{\mathrm{r}} & \text{else}\end{cases} \tag{6}$$

$$\boldsymbol{O}=\boldsymbol{X}_{\mathrm{r}}^{'}\cdot\boldsymbol{X}_{\mathrm{l}} \tag{7}$$

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，有包含两个子网络：一个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。

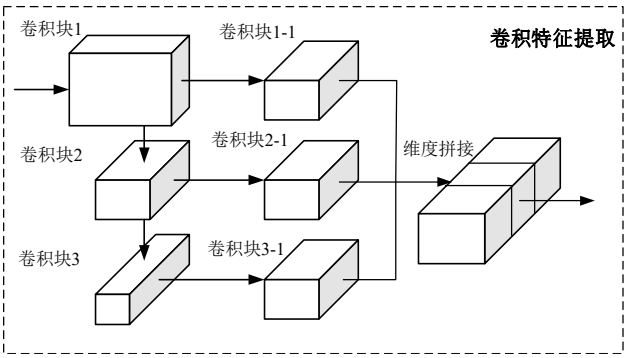


图4 卷积特征提取主干网络

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 BacthNorm 层和 RELU 层。前子网络卷积块输出作为同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 3 种融合方法：拼接融合，相加融合，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 至 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 3 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 IOU(Intersection over Union)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,θ)，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)：

$$\Delta x=\frac{x^{gt}-x^{a}}{\sqrt{(w^{a})^{2}+(l^{a})^{2}}},\Delta y=\frac{y^{gt}-x^{a}}{\sqrt{(w^{a})^{2}+(l^{a})^{2}}},\Delta z=\frac{z^{gt}-z^{a}}{\sqrt{h^{a}}} \tag{8}$$

$$\Delta w=\log\frac{w^{gt}}{w^{a}},\Delta l=\log\frac{l^{gt}}{l^{a}},\Delta h=\log\frac{h^{gt}}{h^{a}} \tag{9}$$

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

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)：

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

其中 SmoothL1 定义为公式(12)：

$$\text{smoothL1}(x)=\begin{cases}0.5x^{2} & \text{if } |x|<1\\ |x|-0.5 & \text{else}\end{cases} \tag{12}$$

由于文献[2]定义的方向损失函数不能区分 0 度和 180 度旋转的框，本文中使用文献[26]的方向损失函数，定义为式(13)：

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

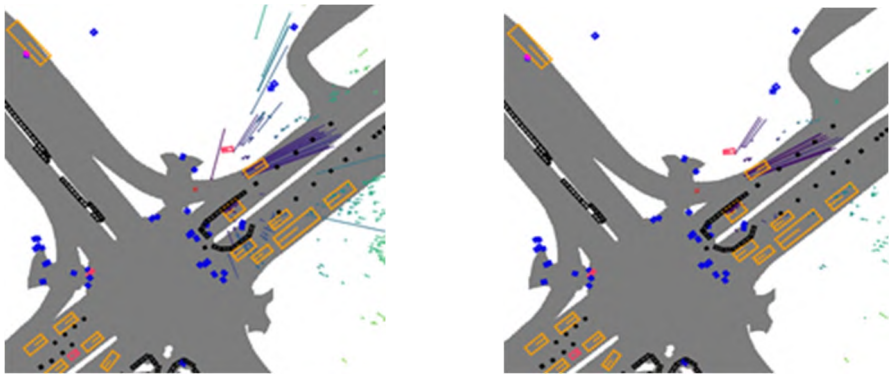
分类函数 L_{cls} 使用的是 Focal loss[29] 损失函数如公式(14)所示。其中 p^{a} 代表框的分类概率，α =0.25，γ =2。

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

总的损失函数定义为公式(15)：

$$L=\frac{1}{N_{\text{pos}}}(\lambda_{1}L_{\text{loc}}+\lambda_{2}L_{\text{cls}}+\lambda_{3}L_{\text{dir}}) \tag{15}$$

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，设置的 λ_{1} =2，λ_{2} =1，λ_{3} =0.2。



(a) 滤波前 (b) 滤波后

图5 毫米波雷达点滤波前后图片对比

Fig. 5 Comparison of unfiltered (left) and filtered (right) images of front radar points

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，本数据集包含 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，多普勒迷糊解状态，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-清楚的，以及点有效性状态：0-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，图右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。

表1 毫米波雷达数据各个通道和通道说明

Tab. 1 Each channel and channel description of radar data

雷达通道	序号	描述说明
x,y,z	0,1,2	雷达点坐标
dyn_prop	3	雷达点动态特性:0-移动，1-固定，2-接近，3-静止的候选目标，4-未知，5-穿过静止的，6-穿过移动的，7-停止
RCS	5	雷达截面积
vx,vy	6,7	x 和 y 方向的速度 m/s
vx_comp, vy_comp	8,9	x,y 方向的补偿速度 m/s
ambig_state	11	多普勒模糊解的状态:0-有效的，1-不明确，2-交错，3-清楚，4-等待候选
invalid_state	14	点有效性状态:0-有效的,1-低 RCS 无效等 18 个状态。

在本文中使用的是激光雷达坐标和强度信息和毫米波雷达信息坐标信息，雷达散射面积 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 $\boldsymbol{T}_{\mathrm{l}}$，旋转矩阵 $\boldsymbol{R}_{\mathrm{l}}$，毫米波雷达雷达平移矩阵 $\boldsymbol{T}_{\mathrm{r}}$，旋转矩阵 $\boldsymbol{R}_{\mathrm{r}}$，其中毫米波雷达转换到激光雷达坐标系的旋转矩阵 $\boldsymbol{R}=\boldsymbol{R}_{\mathrm{l}}\cdot\boldsymbol{R}_{\mathrm{r}}$，转换到激光雷达安装位置的平移矩阵 $\boldsymbol{T}=\boldsymbol{T}_{\mathrm{l}}-\boldsymbol{T}_{\mathrm{r}}$。通过公式(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 $\left(x_{\mathrm{rl}},y_{\mathrm{rl}},z_{\mathrm{rl}}\right)$。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 $\boldsymbol{V}$，而是表示与自身车辆的相对径向速度 $\boldsymbol{V}_{\mathrm{r}}$。该速度在 x-y 方向上的分量为 $\left(\boldsymbol{V}_{\mathrm{x}},\boldsymbol{V}_{\mathrm{y}}\right)=\left(\boldsymbol{V}_{\mathrm{r}}\cdot cos\,\alpha,\boldsymbol{V}_{\mathrm{r}}\cdot sin\,\alpha\right)$，车辆自身速度 $\left(\boldsymbol{V}_{\mathrm{ex}},\boldsymbol{V}_{\mathrm{ey}}\right)$，补偿速度 $\left(\boldsymbol{V}_{\mathrm{xcomp}},\boldsymbol{V}_{\mathrm{ycomp}}\right)=\left(\boldsymbol{V}_{\mathrm{x}},\boldsymbol{V}_{\mathrm{y}}\right)-\left(\boldsymbol{V}_{\mathrm{ex}},\boldsymbol{V}_{\mathrm{ey}}\right)$，利用公式(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 $\left(\boldsymbol{V}_{\mathrm{xcomp\_l}},\boldsymbol{V}_{\mathrm{ycomp\_l}}\right)$。

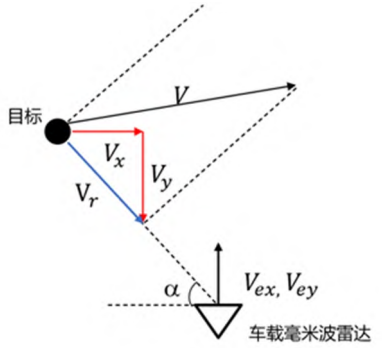


图6 雷达点速度示意图

Fig. 6 Radar point velocity diagram

虽然毫米波雷达数据缺乏相对切向速度，不能完全地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。

$$\begin{bmatrix} x_{\mathrm{rl}} \\ y_{\mathrm{rl}} \\ z_{\mathrm{rl}} \end{bmatrix} = \boldsymbol{R} \begin{bmatrix} x_{\mathrm{r}} \\ y_{\mathrm{r}} \\ z_{\mathrm{r}} \end{bmatrix} - \boldsymbol{T} \tag{16}$$

$$\begin{bmatrix} \boldsymbol{V}_{\mathrm{xcomp\_l}} \\ \boldsymbol{V}_{\mathrm{ycomp\_l}} \\ 1 \end{bmatrix} = \boldsymbol{R} \begin{bmatrix} \boldsymbol{V}_{\mathrm{xcomp}} \\ \boldsymbol{V}_{\mathrm{ycomp}} \\ 1 \end{bmatrix} \tag{17}$$

3.2 实验设置

本文使用 Nuscenes 公开数据集作为实验验证数据，Nuscenes 数据集包含了 28 130 个训练样本和 6 019 个测试样本。数据集的激光雷达扫描频率是 20 帧/秒，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 FMCW(调频连续波)雷达,扫描频率 13 帧/秒,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，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长,所以使用 1/2 数据集即 14 065 个训练样本进行训练，测试样本 6 019 个。

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 epochs(140 000 次迭代)。本文总共设置 2 500 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 度和 90 度。

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、加和融合和相乘融合，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(R) Silver 4210 的 CPU。

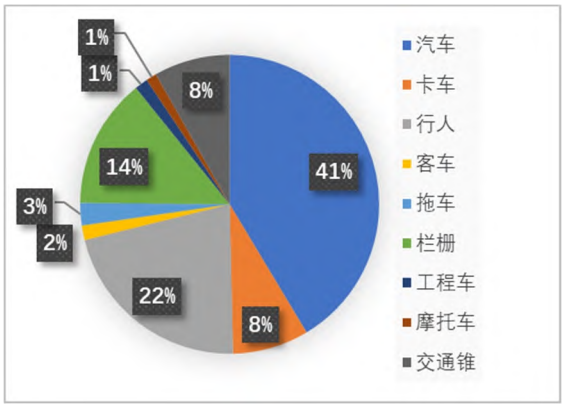


图7 数据集各类实例所占百分比

Fig. 7 Percentage of various instances in the dataset

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比对，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，取得了显著的提升，基于注意力机制的数据融合方法的 mAP 高出基础网络 0.62%，证明了本文中所提算法的有效性。

另外，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，之后的注意力机制进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 mAP 对比 单位：%

Tab. 2 mAP Comparison of baseline network, self-attention network and attention fusion method unit: %

方法	汽车	客车	卡车	行人	摩托车	拖车	工程车辆	栅栏	交通锥	平均
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

证实了毫米波-激光雷达数据融合对性能的提升，本文又将基于注意力机制的融合方法与拼接、加和、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，注意力融合检测性能优于拼接、加和、相乘融合方法。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 米外误差为 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。

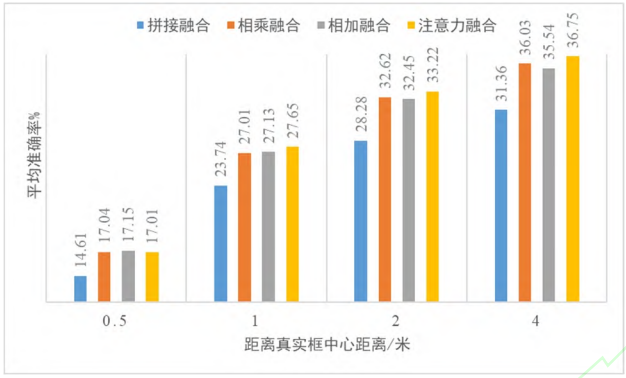


图8 注意力融合和拼接融合、相乘融合、相加融合的平均准确率对比

Fig. 8 Average accuracy comparison of attention fusion, concatetion 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，图中绿色框表示检测的目标物，框的闭合方向表示目标的方向。从左到右依次是真实标签图，基础网络目标检测图和注意力融合网络目标检测图。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其检测出来。证明了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的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。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，有效提高了网络检测性能。

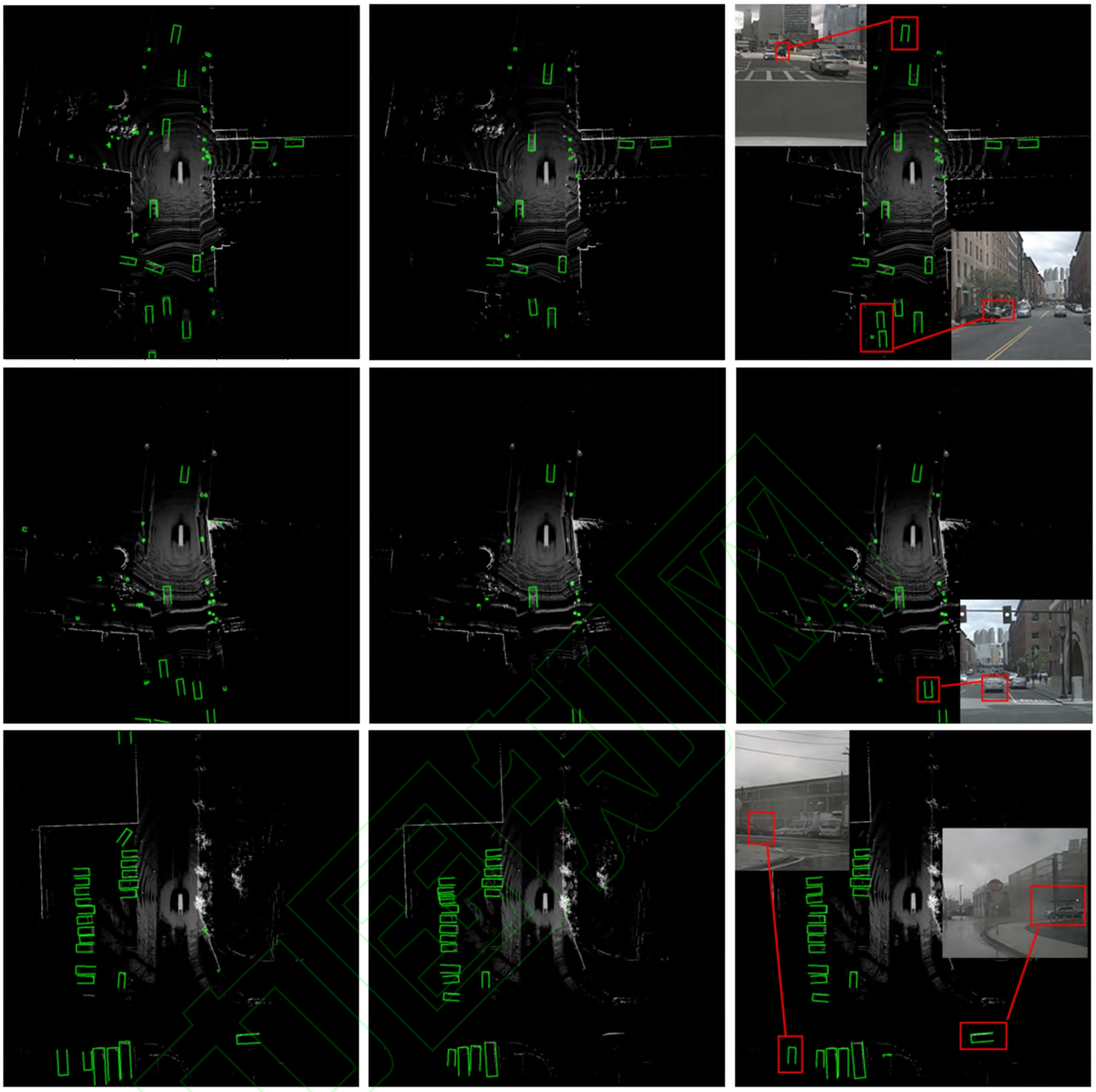
3.6 与 Nuscenes 上最先进方法的对比

本文在完整数据集下进行训练后的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 可以发现本文所提出的融合方法在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 数据集下本文融合方法和 SARPNET、MonoDIS 的 mAP 对比 单位： %

Tab. 3 mAP comparison of the fusion method in this article and the SARPNET and MonoDIS methods use the Nuscenes data set unit: %

方法	汽车	客车	卡车	拖车	平均
SARPNET	59.9	19.40	18.7	18.0	29.0
MonoDIS	47.8	18.8	22.0	17.6	26.6
注意力融合	**77.7**	**44.9**	**28.8**	**37.6**	**47.3**



（a） 真实标签图 (b) 基础网络检测结果图 (c) 毫米波-激光雷达注意力融合检测结果图

图9 基础网络检测结与毫米波-激光雷达注意力融合检测结果对比

Fig. 9 Fusion detection result comparison of basic network detection result image, and the radar-lidar attention

4 结语

本文在点云快速编码网络 PointPillar 的基础上，创新性地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。文中的实验结果证实了所提方法的有效性，并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其不均匀，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以及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

[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 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 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 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 Computer Vision and Pattern Recognition. Piscataway: IEEE, 2017: 1907-1915.

[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. 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 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 2018 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 2017 Neural Information Processing Systems. Cambridge, MA: MIT Press 2017: 5099-5108.

[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 (ITSC). 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 2011 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]. arXiv preprint arXiv:1807.11264, 2018.

[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. RRPN: 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, MADDETN 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. Springer, Cham, 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): 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 Pacific-Rim Symposium on Image and Video Technology. Springer, Cham, 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 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 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 2017 Neural Information Processing Systems. Cambridge: MIT Press 2017: 5998-6008.

[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]// CVPR 2015: Proceedings of the 2015 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 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 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 Computer Vision and Pattern Recognition. Piscataway: IEEE, 2019: 12697-12705.

[27] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]// Proceedings of the 2016 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 Computer Vision and Pattern Recognition. Piscataway: IEEE, 2020: 11621-11631.

[29] LIN T Y, GOYAL P, Girshick R, et al. Focal loss for dense object detection[C]// Proceedings of the 2017 Computer Vision and Pattern Recognition. Piscataway: IEEE, 2017: 2980-2988.

[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, BULO S R, PORZI L, et al. Disentangling monocular 3d object detection[C]// Proceedings of the 2019 Conference on Computer Vision. Piscataway: IEEE, 2019: 1991-1999.

This work was partially supported by National Science Found for Young Scholars under Grant No. 61806186, State Key Laboratory of Robotics and System(HIT) under Grant No. SKLRS-2019-KF-15, the program ‘Fujian Intelligent Logistics Industry Technology Research Institute’under Grant No. 2018H2001, and the program ‘Quanzhou Science and Technology Plan’ under Grant No. 2019C112 and No. 2019STS08.

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

LAN Hai, born in 1988, M.S. His research interests include Machine learning and pattern recognition, and its application in medical imaging

WEI Xian, born in 1986, Ph.D., researcher. His research interests Machine learning and pattern recognition, and applications in autonomous driving.