DOI:10.16136/j.joel.2023.01.0145

基于毫米波雷达点云和视觉信息差异性特征注意力融合的3D目标检测

李 艳,沈 韬*,曾 凯

(昆明理工大学 信息工程与自动化学院 云南省计算机技术应用重点实验室,云南 昆明 650500)

摘要:自动驾驶中传感器融合是感知系统的重要组成部分,雷达点云信息和视觉信息融合可以提高车辆的感知能力。然而现有的研究将雷达点投影到图像上时只是对雷达点简单的增加高度,无法提供更加准确的横向信息,缺乏空间信息。同时对两个模态只是进行简单的融合,虽然产生了一个联合表征,但不足以充分捕捉两种模态之间的复杂联系。文中同时增加了雷达点云的宽度来进行空间信息增强,另外设计了一种利用差异性特征注意力融合的方法,使两个模态进行跨模态交互融合。本文在具有挑战性的nuScenes数据集上对模型进行了评估,提出的模型的*NDS*评分和*mAP*分别达到了46.3%和33.9%,体现了优秀的性能。

关键词:差异性特征注意力;空间信息增强;跨模态融合;3D目标检测
中图分类号:O436 **文献标识码**:A **文章编号**:1005-0086(2023)01-026-08

3D object detection based on attention fusion of millimeter wave radar point cloud and visual information disparity features

LI Yan, SHEN Tao*, ZENG Kai

(Yunnan Key Laboratory of Computer Technologies Application, Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming, Yunnan 650500, China)

Abstract: Sensor fusion in autonomous driving is an important part of the perception system, and the fusion of radar point cloud information and visual information can improve vehicle perception. However, existing studies projecting radar points onto images simply add height to the radar points, which does not provide more accurate lateral information and lacks spatial information. Simultaneous fusion of the two modalities is only simple, which produces a joint representation but is not sufficient to fully capture the complex connection between the two modalities. In this paper, we simultaneously increase the width of the radar point cloud for spatial information enhancement, and additionally design a method for cross-modal interaction fusion of the two modalities using differential feature attention fusion. In this paper, the model is evaluated on the challenging nuScenes dataset, and the proposed model achieves 46.3% and 33.9% in *NDS* score and *mAP*, respectively, reflecting excellent performance.

Key words: differential feature attention; spatial information enhancement; cross-modal fusion; 3D object detection

0 引 言

随着经济的快速发展,人民生活水平不断提高,汽车成为人们日常出行的主要交通工具。由于需要处理各种可能的场景、对象和天气条件,汽车交通场景是复杂的。自动感知系统不能适应狭

* **E-mail**: shentao@kmust.edu.cn
收稿日期:2022-03-09 **修订日期**:2022-04-18
基金项目:国家自然科学基金(61671225,61971208,61702128)、云南省应用基础研究计划项目重点项目(2018FA043)、云南省中青年学术技术带头人后备人才项目(Shen Tao,2018)和云南省万人计划青年拔尖人才项目(云南省人社厅(2018 73)资助项目

窄的特定任务领域,而且必须应对不断变化的环境和不可预见的事件,因此,自动驾驶汽车通常配备了一套传感器,如毫米波雷达、视觉传感器、激光雷达等[1]。与视觉传感器和激光雷达相比,毫米波雷达的探测性能受极端天气的影响较小,因为其基本设计和长波长的特性,即使在恶劣的天气条件下也能稳健地工作。现阶段利用毫米波雷达点云和视觉信息融合进行 3D 目标检测还存在以下两个问题:

1)目前主流方法对毫米波雷达投影到图像上的点只是在纵向上进行垂直拉伸[2,3],缺失横向信息。对于高度丢失问题,虽然现有的方法在纵向上对雷达点增加了高度,但是不同种类目标的高度不同,在不同的场景下难以确定最佳的直线高度。因此,接收到的物体的纵向空间信息与物体的实际空间信息不匹配,无法反映物体的真实空间范围。而且在横向上,毫米波雷达发射的横向电磁波对非金属物体的反应较弱,同时横向的分辨率较低,对行人、摩托、自行车等较小的目标效果差,因此需要在横向上对信息进行增强。

2)目前毫米波雷达信息和视觉信息融合方法一般为拼接法[2]和元素加法[4]。元素加法直接将两个矩阵合并成一个矩阵,拼接法将雷达特征矩阵和图像特征矩阵连接成一个多通道矩阵,通常将来自多个输入的特征图叠放到一起,从而达到融合多种特征的目的。这些融合方法都不是最适合特征融合的,因为雷达特征和视觉特征不均匀,忽略了雷达信号的特征。

为了克服上述的两个问题,本文提出空间注意力融合模型。本研究提出将投影到图像上的雷达点在垂直方向拉伸,纵向上增加高度;在水平方向拉伸,横向上增加宽度,进行毫米波雷达点云投影点空间信息增强,能够包含图像上的所有目标。本文进一步提出了差异性特征注意力融合方法,当雷达特征和视觉特征不均匀,加强雷达点云特征的权重关系,有效引导视觉传感器的信息流,使雷达信息和图像信息跨模态融合。

1　相关工作

在本节中,对毫米波雷达点云和视觉信息融合相关研究进行简要回顾。对于雷达高度缺失问题,NOBIS 等[2]提出 CRF-Net 将雷达探测投影到图像平面上,将投影的雷达点在垂直方向拉伸,以便更好地与图像融合,用垂直线表示高度信息,其中像素值对应于每个测点的深度。LO 等[3]提出使给定的多帧雷达点高度从单一的 0.5 m 固定高度扩展到 0.25—2 m 的高度范围,用于后续毫米波雷达和图像融合的深度估计问题。本研究提出将投影到图像上的雷达点在垂直方向拉伸,纵向上增加高度;在水平方向拉伸,横向上增加宽度,进行毫米波雷达点云投影点空间信息增强。

在 3D 目标检测中,对于单目 RGB 图像,最近 ZHOU 等[5]提出的 CenterNet 使用关键点检测网络来寻找图像上的目标中心点。仅利用目标中心点处的图像特征进行回归,即可获得目标的三维尺寸和位置等其他属性。对于点云的处理,可分为基于体素和基于点的两种方法。基于体素的方法[6],将不规则点云网格化为规则体素,然后进行稀疏的 3D 卷积以学习高维特征。基于点的方法[7],直接对图卷积的深度学习点云进行分类。

对于融合问题,NABATI 等[8]首先使用雷达检测生成三维目标提案,然后将它们投射到图像平面上进行联合二维目标检测和深度估计。CHANG 等[9]提出了一种基于毫米波雷达和视觉传感器的空间融合方法,考虑到雷达点的稀疏可以嵌入到特征提取阶段,有效地利用了毫米波雷达和视觉传感器的特征。NABATI 等[10]提出 RRPN 模型将雷达检测映射到图像坐标系并为每个映射的雷达检测点生成预定义的锚框来生成目标建议,根据目标距车辆的距离对这些锚框进行转换和缩放,为检测目标提供更准确的建议。NABATI 等[11]提出 Centerfusion 模型使用基于截锥的关联方法将雷达探测与图像上的目标精确关联,并创建雷达的特征图,以补充图像特征。PRAKASH 等[12]提出了一种新的多模态融合变压器 TransFuser,将三维场景的全局上下文融合到不同模式的特征提取层中。这些融合方法都不是最适合特征融合的,因为雷达特征和视觉特征不均匀,忽略了雷达信号的特征。本文提出的差异性特征注意力融合方法不同于级联融合和元素叠加融合,基于自适应神经网络生成关注权矩阵来融合视觉特征。

针对上述方法可以得出毫米波雷达和视觉信息融合在空间信息缺失和融合过程中忽略雷达信息的问题。本文基于上述两个问题开展研究。

2　模　型

2.1　网络架构

网络整体架构如图 1 所示,有两个支路,一个雷达支路,一个图像支路。对于图像支路,先利用骨干网络进行图像特征提取。对于雷达支路,先进行雷达数据预处理和空间信息增强,然后加上利用图像特征处理的雷达特征。将雷达特征和图像特征共同输入差异性特征注意力进行融合,最后输入检测头和 3D 框解码器进行分类和回归。

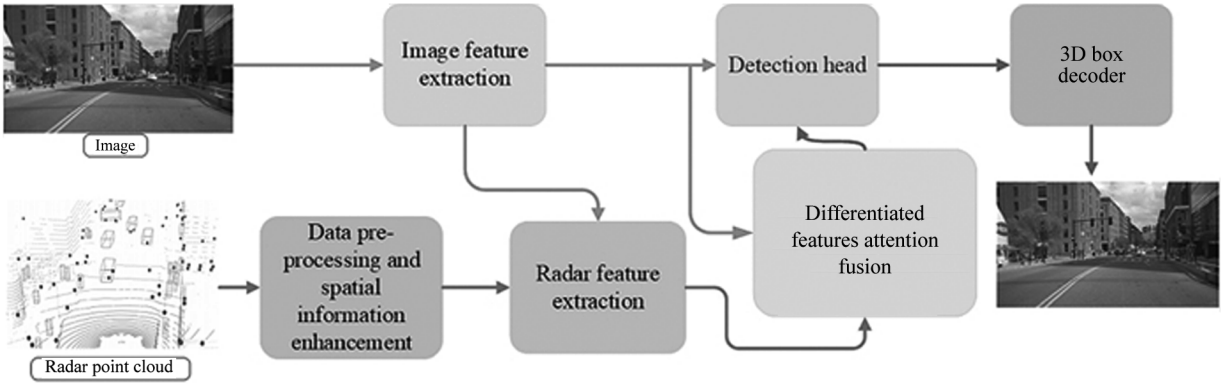

图 1　模型结构图

Fig. 1　Overall structure of the model

2.2　雷达数据预处理和空间信息增强

为了解决雷达数据的稀疏性，本文使用与文献[2]、[11]相同方法用 13 个最临近时间戳(大约 1 s)联合表示增加雷达数据的密度。对于有限的垂直视场(如图 2 所示)，在纵向上与文献[2,3]相同将给定的雷达点高度扩展到 0.25—2.5 m 的高度范围；在横向上增加宽度 0.25—0.5 m，进行空间信息增强。实验效果图如 3 所示，对于空间信息增强的有效性，将在 4.5 节消融实验中给出。第一列为雷达点投影到图像上；第二列表示对雷达点增加了高度和宽度的效果图；第三列为经过处理的空间信息增强效果图；第四列为空间信息增强效果图 2D 展现图，可以直观体现对雷达点的处理效果。经过信息增强，基本可以覆盖所有目标，便于目标信息提取。

对于雷达点的表示，具体地，将每个雷达探测表示为自我中心坐标系中的一个 3D 点，并将其参数化为 $P(x,y,z,v_x,v_y)$，其中(x,y,z)是物体位置，(v_x,v_y)是物体在 x 和 y 方向上的径向速度，并使用毫米波雷达的自身位置进行补偿。

图 2　雷达测量垂直视场有限示意图

Fig. 2　Limited schematic of radar measurement of vertical field of view

2.3　骨干网络

CenterNet[5] 网络直接使用关键点检测网络来寻找图像上的目标中心点，并回归到其他对象属性。在本文中，输入图像 $I_m\in\mathscr{R}^{3\times H\times W}$($H$,$W$ 为图像的高和宽)，经过特征提取器生成 $F_m\in\mathscr{R}^{C\times H\times W}$($C$ 是图像通道数)，然后生成关键点热图 $\hat{Y}\in[0,1]^{c_l\times\frac{H}{R}\times\frac{W}{R}}$，$R$ 是下采样率，c_l 是目标类别的数量。

对于生成的热图，使用 focal loss 函数[5]，如式(1)所示：

$$L_k=-\frac{1}{N}\sum_{xyc}\begin{cases}(1-\hat{Y}_{xyc})^{\alpha}\log(\hat{Y}_{xyc})\\(1-Y_{xyc})^{\beta}(\hat{Y}_{xyc})^{\alpha}\log(1-\hat{Y}_{xyc})\end{cases},\tag{1}$$

式中，N 是图像中的关键点数，α,β 是超参数，取 $\alpha=2$,$\beta=4$，$Y\in[0,1]^{C\times\frac{H}{R}\times\frac{W}{R}}$是目标生成的 ground truth 热图，$\hat{Y}\in[0,1]^{C\times\frac{H}{R}\times\frac{W}{R}}$为预测的输出。

对于图像处理，利用 CenterNet[5] 进行 3D 检测，对每个中心点需要回归 3 个额外的属性为深度、3D 维度和方向。3D 维度用一个单独的头回归到他们的绝对值 $\hat{\Gamma}\in[0,1]^{3\times\frac{H}{R}\times\frac{W}{R}}$，并且使用 L1 损失。对于方向，遵循提出的具体维度为 $Rot=\mathscr{R}\in^{8\times\frac{H}{R}\times\frac{W}{R}}$。为了恢复由输出步幅引起的离散误差，对于每一个点预测了一个局部偏移 $\hat{O}\in^{2\times\frac{H}{R}\times\frac{W}{R}}$。对于雷达点云处理，先进行雷达数据预处理，并且利用图像提供的 2Dbbox、初步的深度信息和大小，得到雷达初步特征 $F_r\in\mathscr{R}^{3\times H_0\times W_0}$(3 是通道数)。最后，将雷达图像最后输出的特征输入到 detection head，回归对象的属性，输出进行分类与回归。该 detection head 同文献[11]设置一样，由 3×3 卷积核和 1×1 卷积层组成，输出的属性信息均使用 L1 损失。

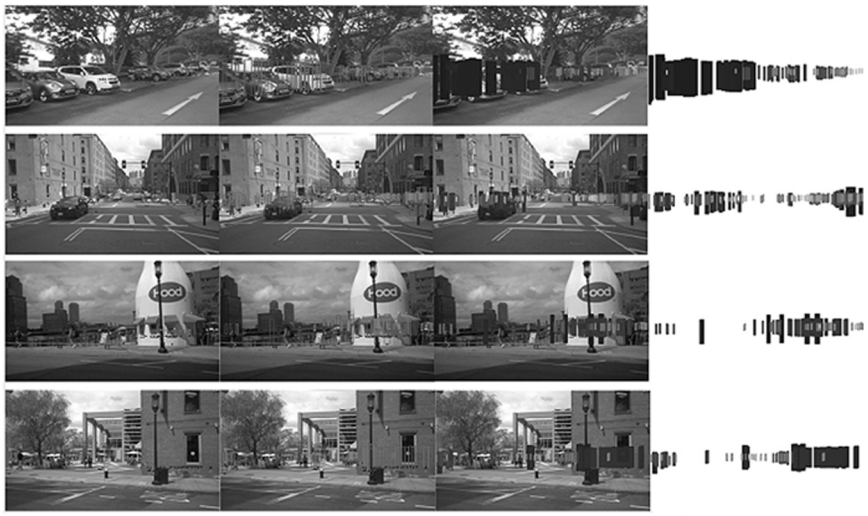

图 3　雷达高度和宽度扩展的信息增强效果图

Fig. 3　Information enhancement effect diagram of radar height and width extension

2.4　差异性特征注意力

在本文中,提出的差异性特征注意力可以对雷达特征和图像特征的不同位置进行加权,当雷达特征和视觉特征不均匀,加强雷达点云特征的权重关系有效地利用了毫米波雷达和视觉传感器的特征,模型图如图 4 所示。其中具有两个输入,分别是雷达特征 $F_r \in \mathcal{R}^{3\times H_0\times W_0}$(3 是通道数)和图像特征 $F_m \in \mathcal{R}^{C\times H_0\times W_0}$。

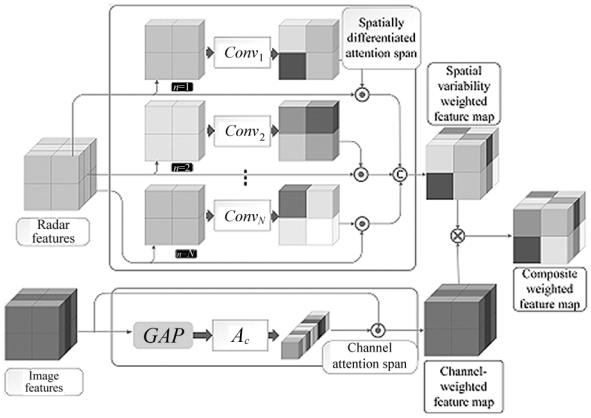

图 4　差异性特征注意力模型图

Fig. 4　Diagram of differential feature attention model

对于雷达支路,利用空间差异性注意力支路,对雷达信息进行加权,如式(2)所示:

$$A_i^s = Conv_s(F_r), \tag{2}$$

式中,$Conv_s$ 为卷积核大小为 1 的卷积层,代表对应于 F_r 的注意力映射。在式(3)中将该权重 A_i^s 与对应的子特征图 F_r^n 相结合,从而获得一张新的特征图。

$$F_R = A_i^s \odot F_r^n, \tag{3}$$

式中,$\odot$ 代表哈达玛积,F_R 蕴含了经过强化的雷达信息。

本文提出的注意力机制,不仅利用空间上的重要性,同时也对不同通道的权重予以调节。此处利用通道注意力对图像特征 F_m 进行加权。

$$A_i^C = A_c(GAP(F_m)), \tag{4}$$

式中,$A_i^C \in \mathcal{R}^C$ 代表 F_m 通过注意力计算层 $A_c(\cdot)$ 计算后得到的通道注意力向量。

将该通道注意力向量同原特征图 F_i 依式(5)结合便可以在通道层面强化特征。

$$F_M = A_i^C \odot (F_m)。 \tag{5}$$

为了使得网络既能在空间层面注意更多的信息,又可以在通道层面对关键的信息进行强化,依据式(6)将两者连接起来。

$$F = F_R + F_M。 \tag{6}$$

输出的信息输入 detection head,与文献[11]设置相同,以产生所需的输出,有助于从雷达特征图中学习更高层次的特征,如 3Dsize、热图、方向、旋转、速度等。

3　实验和分析

3.1　nuScenes 数据集

nuTonomy 建立的 nuScenes 数据集[13]是现有最大的自动驾驶数据集。该数据集不仅提供相机和激光雷达数据,还包含毫米波雷达数据。其中包括 6 个摄像头、5 个雷达和 1 个激光雷达。数据集是按场

景组织的，整个数据集包含1 000个场景，可分割为训练集和测试集，但只有trainval split的注释是公开可用的(850个场景)，其中700个训练场景，150个验证场景。在每个样本中，本文有6张图像和5个不同方向的雷达扫描。

3.2 评价指标

对于模型的评估，采用与文献[5]、[11]、[14]相同的评价指标。其中，mAP是根据平均精度度量(AP)评估不同类别的检测以得到检测模型的综合结果，可反映模型检测的准确性，如式(7)所示：

$$mAP=\frac{1}{|C||D|}\sum_{c\in C}\sum_{d\in D}AP_{c,d}, \tag{7}$$

式中，AP为召回率和精度超过10%的精确召回曲线下的归一化面积，C为类别集合，$D=\{0.5,1,2,4\}$为匹配阈值的平均值。

在3D目标检测中，mAP指标不能捕捉nuScenes检测任务的所有方面，如速度和属性估计。nuScenes的官方提出一个新的评价指标，建议将不同的误差类型合并到一个标量分数中，称为nuScenes检测分数(NDS)，如式(8)、(9)所示：

$$NDS=\frac{1}{10}\left[5mAP+\sum_{mTP\in TP}(1-\min(1,mTP))\right], \tag{8}$$

$$mTP=\frac{1}{|C|}\sum_{c\in C}TP_c, \tag{9}$$

式中，NDS是mAP与盒位置(ATE，平均平移误差)、大小(ASE，平均尺度误差)、方向(AOE，平均方向误差)、属性(AAE，平均属性误差)和速度(AVE，平均速度误差)的加权平均值。

3.3 实验细节设置

本文使用预先训练的CenterNet[5]网络中DLA[5]骨干作为目标检测网络，其中CenterNet在nuScenes数据集上训练了140个epoch。其中设置10个类是所有在nuScenes中注释的23个类的子集。nuScenes数据集[13]中的相机数据的分辨率为1 600×900 pixel，其中通过相机参数调整输入RGB图像x_{RGB}为448×800，可以加快训练收敛速度。另外，实验部分使用的所有模型都是在PyTorch中实现的，本文的模型在两个Nvidia V100 GPU上进行额外的60个epoch(26批大小)的训练，学习率为0.000 025。

3.4 对比实验分析

3.4.1 不同模态3D目标检测方法对比实验

将本文提出的模型与在nuScenes[13]数据集上的不同模态进行3D目标检测方法比较，如表1所示。其中CenterNet[5]和MonoDIS[14]是相机模型；PointPillars[7]是基于激光雷达的模型；SPRCNN[15]是激光雷达和相机融合模型；CenterFusion[11]是毫米波雷达和图像融合的模型。表1是nuScenes测试集评估指标性能比较，表2是3D对象的检测分类结果。其中L代表激光雷达，C代表摄像头，R代表毫米波雷达。

分析表1可得出在分割测试集上，与基于激光的模型PointPillars[7]和基于其他传感器的模型相比，本文的模型的NDS得分和mAP均得到提高。特别是与CenterNet[5]、CenterFusion[11]相比，本文模型在NDS上提高了6.3%和1.4%，mAP上提高了0.1%和1.3%。另外，和其他的误差指标相比，本文模型误差相对减少，体现了模型的优越性和有效性。分析表2在测试集对象分类检测结果，与MonoDIS[14]、CenterFusion[11]相比，性能提高明显。对拖车、行人、摩托、自行车等目标，与MonoDIS[14]相比，本文的模型分别提高了8.1%、3.8%、7.4%和1.8%；与CenterNet[5]相比，分别提高了0.6%、3.3%、7.3%和1.8%；与CenterFusion[11]相比，分别提高了2.2%、3.8%、5%和2.4%。可以得出本文的模型达到了优秀的性能，体现了模型性能的优越性。

3.4.2 与现有的毫米波雷达点云信息和视觉信息融合方法对比

在自动驾驶领域，对于毫米波雷达和视觉信息融合进行目标检测，许多学者提出了很多方法，体现了优秀的性能。目前，主要有RRPN[10]、CRF-Net[2]和CenterFusion[11]。为了体现文中模型在这个领域的优势，比较本文的模型与这些模型，如表3所示。与RRPN[10]模型比较，对于汽车、行人、摩托，分别提高了11.8%、23.7%和5.9%；对于CRF-Net[2]和CenterFusion[11]，对汽车、行人、摩托和自行车，分别提高4.5%、6.1%、15.4%和8.5%；2.7%、3.8%、5%和2.4%。对于其他的类别，本文提出的模型均有所提高，表明了在毫米波雷达点云和摄像头视觉信息融合方法中，文中模型的有效性。

3.5 消融实验分析

在nuScenes验证集进行消融研究来验证本文模型各个模块的合理性，其中以改进的CenterNet[5]作为baseline，分析在测试集nuScenes上的性能指标(表4)，可以得到与baseline比较，加入空间信息增强模块后，NDS得分和mAP分别提高了2.1%和1.3%，表明加入了空间信息模块后学习到更多相关的雷达信息，更好地引导视觉传感器的信息流。加

入差异性特征注意力融合模块以后,*NDS* 得分和 *mAP* 分别提高了 1.5%和 0.9%,表明本文提出的融合方法的有效性。除误差 *mAAE* 外,其他的误差均减少。当所有模块相互作用时,本文的模型达到了最好的性能,体现了模型的优越性和优秀性。分析模型在 nuScenes 测试集对象的分类检测结果(表 5),可以得到相比于 baseline,加入空间信息增强模块后,对各个类别的提升最明显,对于拖车,提升了 10.9%;对行人、摩托、自行车、交通锥等一些小目标,分别提升了 1.9%、8.4%、0.2%和 3.1%。加入差异性特征注意力模块以后,对行人、摩托、自行车和交通锥,分别提高了 2.5%、8.8%、0.8%和 4.1%,表明本文提出的融合方法对分类检测结果的有效性。另外,可以得到当所有模块相互作用时,本文模

表 1　3D 目标检测模型在 nuScenes 测试集的性能比较结果

Tab. 1　Performance comparison results of 3D target detection models in the nuScenes test set

Methods	Modality	*NDS*	*mAP*	*mATE*	*mASE*	*mAOE*	*mAVE*	*mAAE*
PointPillars[7]	L	0.453	0.305	0.517	0.290	0.500	0.316	0.319
MonoDIS[14]	C	—	0.304	0.738	0.263	0.546	1.553	0.134
CenterNet[5]	C	0.400	0.338	0.658	0.255	0.629	1.629	0.142
SPRCNN[15]	L+C	—	0.361	0.751	0.231	0.571	1.672	0.112
CenterFusion[11]	R+C	0.449	0.326	0.631	0.261	0.516	0.614	0.115
Our model	R+C	0.463	0.339	0.632	0.258	0.531	0.539	0.132

表 2　nuScenes 测试集的对象检测结果

Tab. 2　Object detection results of nuScenes test set

Methods	Modality	Car	Truck	Bus	Trailer	Const	Pedest	Motor	Bicycle	Traffic cone	Barrier
PointPillar[7]	L	0.705	0.250	0.334	0.167	0.045	0.599	0.200	0.016	0.296	0.332
MonoDIS[14]	C	0.478	0.220	0.188	0.176	0.074	0.370	0.290	0.207	0.583	0.533
CenterNet[5]	C	0.536	0.270	0.248	0.251	0.086	0.375	0.291	0.207	0.583	0.533
CenterFusion[11]	R+C	0.509	0.258	0.234	0.235	0.077	0.370	0.314	0.201	0.575	0.484
Our model	R+C	0.536	0.276	0.336	0.257	0.058	0.408	0.364	0.225	0.576	0.468

表 3　毫米波雷达和图像融合在 nuScenes 测试集的对象分类检测结果

Tab. 3　Object classification detection results of millimeter wave radar and image fusion in nuScenes test set

Methods	Car	Truck	Bus	Trailer	Const	Pedest	Motor	Bicycle	Traffic cone	Barrier
RRPN[10]	0.418	0.447	0.572	—	—	0.171	0.305	0.214	—	—
CRF-Net[2]	0.491	0.267	0.431	—	—	0.347	0.210	0.140	—	—
CenterFusion[11]	0.509	0.258	0.234	0.235	0.077	0.370	0.314	0.201	0.575	0.484

表 4　本文提出模型在 nuScenes 测试集的性能比较结果

Tab. 4　The performance comparison results of the proposed model in nuScenes test set

Baseline	Spatial information enhancement	Differentiated features attention fusion	*NDS*	*mATE*	*mASE*	*mAOE*	*mAVE*	*mAAE*
√	—	—	0.438	0.654	0.289	0.566	0.573	0.120
√	√	—	0.442	0.646	0.271	0.552	0.550	0.126
√	—	√	0.448	0.640	0.263	0.546	0.542	0.128
√	√	√	0.463	0.632	0.258	0.531	0.539	0.132

表5 本文提出的模型 nuScenes 测试集的对象检测结果

Tab. 5 Object detection results of the proposed model innuScenes test set

Baseline	Spatial information enhancement	Differentiated features attention fusion	Car	Truck	Bus	Trailer	Const	Pedest	Motor	Bicycle	Traffic cone	Barrier
√			0.508	0.232	0.310	0.132	0.046	0.373	0.273	0.213	0.535	0.459
√	√		0.521	0.258	0.324	0.241	0.051	0.392	0.357	0.215	0.566	0.461
√		√	0.526	0.263	0.327	0.250	0.054	0.398	0.361	0.221	0.570	0.463
√	√	√	0.536	0.276	0.336	0.257	0.058	0.408	0.364	0.225	0.576	0.468

型达到了最好的分类检测结果。

3.6 可视化

另外，本文可视化了最后的实验检测效果图，使模型与 baseline 检测模型进行比较，如图5、图6所示。其中，图5为漏检的效果图，漏检目标由矩形框标出；图6为错检的效果图，错检目标由圆框标出。在可视化图片中，第一列为预测的目标深度值，直观表现图像中的目标；第二列为 baseline 模型检测目标的3D框图；第三列为 baseline 模型检测目标的具体分类和检测效果图，可以直接体现目标检测效果；第四列为提出的模型目标的3D框图；第五列为本文模型得到的具体目标分类检测效果图。从图5中可以看出，对于卡车、行人、自行车、交通锥等目标，模型很大程度上减少了漏检现象。同理在图6中可以得到，对于摩托、行人、自行车等目标存在的错检现象，本文模型得到了有效的提高，展现了模型的有效

图5 与 baseline 模型相比漏检实验效果图

Fig. 5 Experimental effect diagram of leakage detection compared with the baseline model

图6 与 baseline 模型相比错检实验效果图

Fig. 6 Experimental effect diagram of error detection compared with the baseline model

性和优越性。

4　结　论

文中提出的毫米波雷达点云和视觉信息差异性特征注意力融合的模型,考虑了雷达空间信息缺失的问题和现有的融合方法大部分只是简单的叠加忽略了雷达特征的问题,设计了在雷达投影到图像上的点纵向增加高度和横向增加宽度,并且设计了差异性特征注意力加强雷达信息权重,使雷达信息和图像信息跨模态融合。在 nuScenes 验证集对本文提出的模态进行了评估,验证了模型的优越性和有效性。今后,团队将继续在恶劣的天气情况下进行研究。

参考文献:

[1] REN K,WANG Q,WANG C,et al. The security of autonomous driving:Threats,defenses,and future directions[J]. Proceedings of the IEEE,2019,3(2): 357-372.

[2] NOBIS F, GEISSLINGER 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), October 15-17,2019,Bonn,Germany. New York:IEEE,2019:1-7.

[3] LO C C,VANDEWALLE P. Depth estimation from monocular images and sparse radar using deep ordinal regression network[C]//2021 IEEE International Conference on Image Processing (ICIP), September 19-22, 2021, Anchorage,AK,USA. New York:IEEE,2021:3343-3347.

[4] KOWOL K,ROTTMANN M,BRACKE S,et al. YOdar:uncertainty-based sensor fusion for vehicle detection with camera and Radar Sensors[C]//13th International Conference on Agents and Artificial Intelligence(ICAART), February 4-6, 2021, Virtual Event. Setúbal, Portugal: Science and Technology Publications,2021,2:177-186.

[5] ZHOU X Y,WANG D Q,KRÄHENBÜHL P. Object as as-points[EB/OL]. (2019-04-16)[2022-03-09]. https://arxiv.org/abs/1904.07850.

[6] HOU J,DAI A,NIESSNER M. 3D－SIS:3D semantic instance segmentation of RGB-D scans[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 15-20, 2019, Long Beach, CA, USA. New York:IEEE,2019:4421-4430.

[7] LANG A H,VORA S,CAESAR H,et al. Pointpillars:Fast encoders for object detection from point clouds[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 15-20, 2019, Long Beach, CA, USA. New York:IEEE,2019:12697-12705.

[8] NABATI R,QI H R. Radar-camera sensor fusion for joint object detection and distance estimation in autonomous vehicles. (2020-09-17) [2022-03-09]. https://arxiv.org/abs/2009.08428.

[9] 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.

[10] NABATI R,QI H. RRPN:radar region proposal network for object detection in autonomous vehicles[C]//2019 IEEE International Conference on Image Processing (ICIP), September 22-25, 2019, Taipei, Taiwan, China. New York:IEEE,2019:3093-3097.

[11] NABATI R, QI H. Centerfusion: center-based radar and camera fusion for 3D object detection[C]//IEEE/CVF Winter Conference on Applications of Computer Vision, January 3-8,2021,Waikoloa,HI,USA. New York: IEEE, 2021:1527-1536.

[12] PRAKASH A, CHITTA K, GEIGER A. Ulti-modal fusion transformer for end-to-end autonomous driving[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 20-25,2021,Nashville,TN,USA. New York:IEEE,2021:7077-7087.

[13] CAESAR H, BANKITI V, LANG A H, et al. Nuscenes: A multimodal dataset for autonomous driving. (2019-03-26)[2022-03-09]. https://arxiv.org/abs/1903.11027.

[14] SIMONELLI A,BULO S R,PORZI L, et al. Disentangling monocular 3D object detection[C]//2019 IEEE/CVF International Conference on Computer Vision, October 27-November 2, 2019, Seoul, Korea (South). New York: IEEE,2019:1991-1999.

[15] LI P,CHEN X,SHEN S. Stereo R-CNN based 3D object detection for autonomous driving[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 15-20, 2019, Long Beach, CA, USA. IEEE, New York:IEEE,2019:7644-7652.

作者简介:

沈　韬　(1984－),男,教授,博士生导师,从事太赫兹技术和智能感知与计算等方面的研究.