doi:10.19562/j.chinasae.qcgc.2024.02.004

基于虚拟点云的二阶段多模态融合网络*

程 腾[1,2,3],倪 昊[1,2,3],张 强[1,2,3,4],王文冲[4],石 琴[1,2,3]
(1. 合肥工业大学,自动驾驶汽车安全技术安徽省重点实验室,合肥 230009; 2. 安徽省智慧交通车路协同工程研究中心,合肥 230000; 3. 合肥工业大学汽车与交通工程学院,合肥 230000; 4. 奇瑞汽车股份有限公司,芜湖 241000)

[摘要] 针对点云的稀疏性和无序性对目标检测准确率的影响,本文提出了一种基于虚拟点云的二阶段多模态融合网络VPC-VoxelNet。首先,利用图像检测目标信息构造虚拟点云,增加点云的密集程度,从而提高目标特征的表现;其次,增加点云特征维度以区分真实和虚拟点云,并使用含置信度编码的体素,增强点云的相关性;最后,采用虚拟点云的比例系数设计损失函数,增加图像检测有监督训练,提高二阶段网络训练效率,避免二阶段端到端网络模型存在的模型误差累计问题。该目标检测网络VPC-VoxelNet在KITTI数据集上进行了测试,检测精度优于经典三维点云检测网络和某些多传感器信息融合网络,车辆检测精度达到了86.9%。

关键词:目标检测;多模态感知;虚拟点云;损失函数

Two-Stage Multimodal Fusion Networks Based on Virtual Point Clouds

Cheng Teng[1,2,3], Ni Hao[1,2,3], Zhang Qiang[1,2,3,4], Wang Wenchong[4] & Shi Qin[1,2,3]
1. *Hefei University of Technology, Key Laboratory for Automated Vehicle Safety Technology of Anhui Province, Hefei* 230009;
2. *Engineering Research Center for Intelligent Transportation and Cooperative Vehicle-Infrastructure of Anhui Province, Hefei* 230000;
3. *School of Automotive and Transportation Engineering, Hefei University of Technology, Hefei* 230000;
4. *Chery Automobile Co., Ltd., Wuhu* 241000

[**Abstract**] To address the impact of sparsity and disorder of point clouds on target detection accuracy, a two-stage multimodal fusion network VPC-VoxelNet based on virtual point clouds is proposed in this paper. Firstly, virtual point clouds are constructed using image detection target information to increase the density of point clouds, thus improving the performance of target features. Secondly, the dimensionality of point cloud features is increased to distinguish real and virtual point clouds, and a voxel with confidence encoding is used to enhance the correlation of point clouds. Finally, the scale factor of the virtual point clouds is adopted to design the loss function to increase the supervised training of image detection and improve the training efficiency of the two-stage network, and avoid the cumulative model error problem of the two-stage end-to-end network model. The target detection network, VPC-VoxelNet, is tested on the KITTI dataset, and the detection accuracy is better than that of the classical 3-dimensional point cloud detection network and certain multi-sensor information fusion networks, with a vehicle detection accuracy of 86.9%.

Keywords: target detection; multimodal perception; virtual point cloud; loss function

* 国家自然科学基金(82171012)、安徽省自然科学基金(2208085MF171)、安徽省新能源汽车暨智能网联汽车创新工程项目(JZ2021AFKJ0002)和汽车标准化公益性开放课题(CATARC-Z-2022-01350)资助。
原稿收到日期为 2023 年 05 月 10 日,修改稿收到日期为 2023 年 07 月 30 日。
通信作者:程腾,副教授,博士,E-mail:cht616@hfut.edu.cn。

前言

近年来,环境感知已经成为机器人和自动驾驶研究的重要领域,其中三维目标检测在智能交通系统中起着不可或缺的作用,因此越来越受到关注[1]。近年来,基于激光雷达点云的三维目标检测取得了快速发展,但是由于部分目标在扫描过程中采样密度较低,检测性能明显下降。因此,融合图像数据和激光雷达点云数据的目标检测方法使数据源更加丰富,可以实现更加准确的三维目标检测[2-3]。

基于激光雷达点云的目标检测是指利用三维点云数据,通过深度学习算法预测场景中每个目标的位置和类别,相较于传统使用支持向量机、随机森林[4]等方法具有更高的精度[5]。由斯坦福大学Charles等提出的网络PointNet[6]是最早的纯点云目标检测方法,它是一种全连接的神经网络,可以对点云数据进行分类和语义分割;后来,改进版本PointNet++被提出,它在PointNet的基础上使用了多层次聚合和特征提取,进一步提高了点云数据的表示能力;基于PointNet的一些变种网络,例如PointCNN[7]、DGCNN[8]等,它们通过改进点云上的卷积操作、局部聚合方式和特征提取等方面来提高点云目标检测的性能。此外,将点云数据转化为体素形式表达的方法,即将点云数据在三维空间中划分为一系列大小相等的体素也一定程度上克服点云数据的稀疏性和无序性问题,提高检测精度。最早的点云体素化方法SubCNN由斯坦福大学的邹亮提出,VoxelNet[9]、Second[10]等方法都是应用最为广泛的点云体素化方法。还有利用不同模态之间的互补性,通过多模态目标检测来提高检测的准确性和鲁棒性[11]。MV3D是最早的多模态目标检测网络,MV3D网络将图像和点云分别输入到两个独立的CNN和VoxelNet网络中进行处理,最后通过简单的算法将两个结果融合起来。

然而,基于点云的目标检测仍存在两个问题:(1)纯点云方案由于点云数据是非常稀疏的,并且点的数量和位置是无序的,容易出现漏检和误检的问题[12];(2)多模态方案需要将不同类型的数据进行特征提取并融合,复杂的转化和对齐导致模型复杂度高,训练时间长。

对此,本文提出了一种基于虚拟点云的二阶段多模态融合网络VPC-VoxelNet,可以有效改善点云数据的稀疏性和无序性,提高目标检测的精度。本文主要工作和贡献如下:

(1)提出基于图像检测结果的虚拟点云构造方法,维持点云数据完整性的同时,使点云更加密集,增强目标检测的性能;

(2)使用含目标置信度的点云体素进行卷积,增强点云有序性,同时设计基于虚拟点云的比例系数优化损失函数,避免模型误差累积,加快网络训练;

(3)设计图像检测结果与点云数据的多模态融合方案,避免复杂的图像和点云特征对齐融合,降低模型复杂度。

1　网络

网络VPC-VoxelNet的整体结构如图1所示。首先将图像送入主干网络DLA-34[13]进行特征提取,再通过回归网络获得一定数量目标3D关键点的坐

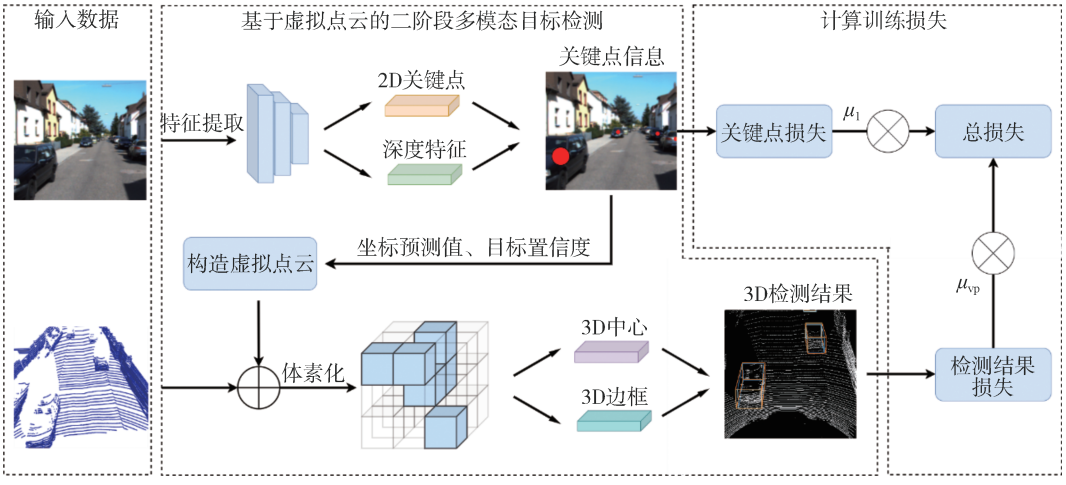

图1　VPC-VoxelNet网络整体结构

标预测值和目标置信度;再根据生成的关键点信息,在激光雷达点云中构造相应的虚拟点云,并增加点云特征维度区别虚拟和真实点云,把输出的目标置信度同时纳入特征编码,再与真实点云一起送入基于体素的3D目标检测;同时,为了避免二阶段端到端串联网络[14]存在的模型误差累计问题,采用虚拟点云的比例关系设计损失函数,增加图像检测有监督训练,从而提高了图像处理模块的训练效率。具体细节如下。

1.1 基于图像的关键点检测

将单目图像经过特征提取网络之后得到对应大小特征图,基于CenterNet[15]的思想,直接预测出目标3D关键点在2D图像上的投影点。将点云数据中3D关键点真值通过相机公式转换到相机平面投影,再编码成一张张2D高斯图。高斯图是一个二维概率分布函数,它将较高的概率值分配给物体中心附近的像素,将较低的概率值分配给离中心较远的像素。对于每个关键点,通过计算以关键点位置为中心的二维高斯概率分布生成高斯图。高斯的标准差通常被设定为一个固定值,这决定了关键点周围的概率值的分布。然后将所有关键点的高斯图相加,生成最终的热图,它代表了每个像素属于某个特定物体类别的可能性。每个热图中概率值最高的像素被视为相应关键点的位置。二维高斯函数公式如下:

$$G(x,y)=\frac{1}{2\pi\sigma^2}\mathrm{e}^{-\frac{x^2+y^2}{2\sigma^2}} \tag{1}$$

式中:x和y为关键点的像素点坐标;σ为关键点所代表目标的大小自适应标准偏差,往往与对象的尺寸有关。

图像特征图经过一系列网络,预测头输出对高斯图偏移量的预测$[\delta_{x_c},\ \delta_{y_c}]$。之后,借鉴SMOKE[20]的思想,先采用数理统计的方式获得转换矩阵中的常量参数即3D关键点深度的均值μ_z与方差σ_z,再结合预测头预测出深度的偏移量δ_z,通过矩阵运算最终得到关键点的3D坐标预测值$[x_p,\ y_p,\ z_p]$。

$$\begin{bmatrix}x_t\\y_t\\z_t\end{bmatrix}=\boldsymbol{K}_{3\times3}^{-1}\begin{bmatrix}z_t\cdot x_{ct}\\z_t\cdot y_{ct}\\z_t\end{bmatrix} \tag{2}$$

$$z_t=\mu_z+\delta_z\sigma_z \tag{3}$$

式中:x_t、y_t、z_t为关键点在相机坐标系下的位置;x_{ct}、y_{ct}为像素点坐标;$\boldsymbol{K}$是相机内参矩阵。

1.2 构造虚拟点云

对于相机而言,目标的3D中心在投影到图像平面上之后,与2D包围框的中心差距较大,甚至有可能超出图像的界限;激光雷达无法扫描物体内部,而目标3D中心位于物体内部,点云只能扫描到物体的表面,所以单目网络预测出来的3D关键点周边找不到点,因此无法一一对应。

由此本文提出了一种构造虚拟点云的融合方式,如图2所示。具体处理流程如下:在每次训练中,单目网络的最终特征图选取出N个置信度最高的关键点;取预测出的深度z,结合相机内参变换矩阵,得到在点云空间中的N个虚拟3D点$[x_{vp},\ y_{vp},\ z_{vp}]$;为了防止超出真实点云的前视图范围,对这些虚拟点云进行筛选过滤后得到N'个点,并添加到点云数据中,其反射强度采用整体点云的平均值替代。

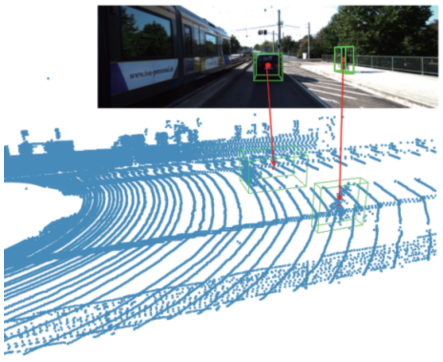

图2 虚拟点云构造示意图

1.3 基于点云体素化的目标检测

随后的点云3D目标检测网络采用基于体素特征的形式。其主要思想是将整个3D空间沿着x、y、z3个轴分割成大小相同的体素块。对每个体素块中的点云进行特征编码,充分考虑其全局和局部特征后得到体素特征,再采用3D卷积的方式进行目标检测。

此外,本文提出的检测网络还对虚拟点云所在的体素块进行数据扩充。方法如下:根据虚拟点云的位置信息,可以确定其对应的体素块位置,通过验证该位置是否存在体素,进行标记。假如存在就对该位置体素进行标记,方式为在$[x,y,z,r]$的回波后增加一个置信度数值$conf$。如果该位置不存在真实点云,则考虑整个体素块的空间分布,按照均匀分布的策略进行添加。在设计单个体素时,规定了最多取5个点云,由于体素3D目标检测方法在高度方向上的考量不太大,所以选择矩形的切面,并均匀构造4个点,将其与虚拟点一起添加到整个点云数据中,如图3所示。

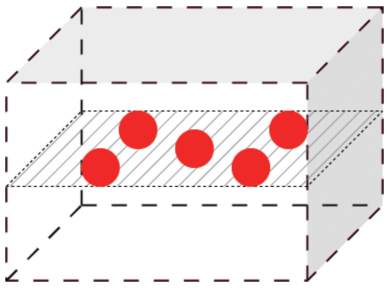

图3　体素内虚拟点云构造位置

除了将目标置信度纳入点云的特征编码,还需要增加点云的特征维度,以区分激光雷达扫描到的真实点云与构造的虚拟点云,因此在$[x, y, z, r, conf]$之后,添加了额外的维度,以1/0区分是否为虚拟点云。这样在点云体素化后得到的体素特征中,便可以包含3D目标关键点信息,本文借鉴了SECOND网络作为后续的检测头,在对特征进行3D稀疏卷积和压缩后,对目标的种类和边框回归分量进行预测。

1.4　训练损失

本文提出的检测网络的损失分为两部分。对于单目图像的检测增加有监督训练,主要关注3D关键点的定位偏差,采用focal loss。对于体素3D目标检测,则主要关注目标分类以及最终预测到整体3D包围框的回归偏差,采用smoothL1损失。

此外,基于本文设计的网络结构,还针对损失函数提出了优化方案。首先,由于整体网络属于二阶段网络的一次训练,设置一个损失波动系数。该系数是前若干次训练中单目网络损失的偏差加权。由于在训练后期单目检测网络部分的损失波动较小,不需要过多的偏重,因此可以将更多的资源用于对体素网络部分的优化。

由于图像检测阶段在网络训练前期检测效果差,关键点坐标偏移严重,再通过坐标转换到点云空间后,该位置可能已经存在了真实点云。而在训练后期,单目网络可以较好地预测出3D关键点所在位置,该位置基本上不存在真实点云。综上,可以记录需要扩充虚拟均匀点云的3D关键点数量,间接反映出单目网络的准确程度,当该数量较少时,给予大的损失权重μ_{up},从而进一步提高第一部分单目网络的训练效率。

$$\mu_1 = \frac{\sum_{i=1}^{n} \left| \Delta Loss_i - \Delta Loss_{i-1} \right|}{Loss_i} + \beta \tag{4}$$

$$\mu_2 = \frac{N}{N_{\max}} + \beta \tag{5}$$

式中:$\Delta Loss_i$和$\Delta Loss_{i-1}$为本轮和上一轮的损失值;n为已参与训练的训练轮次;N为本轮训练构造的符合3D空间范围的虚拟点云数量;$N_{\max}$为关键点网络设定选取3D关键点的数量;β为可调的极小值。

总损失为两部分损失之和:

$$Loss = \mu_1 L_1 + (1 - \mu_2) L_2 \tag{6}$$

式中:L_1为3D关键点的定位损失;L_2为最终预测结果的损失。

2　实验

2.1　实验设置

为了验证算法的性能,本文选取KITTI数据集作为训练和验证数据集。城市景观数据集KITTI包含从城市地区、村庄和高速公路等场景收集的真实图像数据。每张图像包含最多15辆汽车和30名行人,具有不同程度的遮挡和截断。根据目标的数量和类型的不同而定遮挡程度,测试集将检测任务分为3个级别:容易、中等和困难。在本文中,3 712张图片用于训练,3 769张图片用于验证。

模型训练平台为ubuntu18.04 + tesla V100S显卡,模型搭建框架采用Pytorch1.3.1+Cuda11.0。本文使用KITTI数据集的官方评价标准,包括常用的鸟瞰平均精度(AP BEV)和3D平均精度(AP3D)[16]。同时,本文也根据公开的方法对数据集进行了划分。KITTI数据集的探测对象根据它们在前视图中的像素大小、是否可见以及是否被遮挡,分为简单、中等和困难。通常使用交集/联合(IOU)为0.5或0.7来衡量预测值和标签是否是目标。为了检测新网络的性能和表现,本文最终选择IOU=0.7作为衡量标准。

2.2　实验分析

为了验证提出的多模态检测模型的效果,本文进行了如下的相关试验。

2.2.1　关键点

单目图像的关键点检测结果将决定虚拟点云的构建,这些关键点的位置代表目标中心点的位置,因此,必须确保它们的准确性,以便通过构建虚拟点云来精确地表示目标。激光点云空间中,目标中心位置没有点云分布。然而,如果关键点的位置位于物体边缘,那么该位置已经存在少量点云,这将导致在进行点云体素化后只能对体素进行置信度编码,而不能构建更多的虚拟点云,从而导致融合效果的大幅降低。

通过输出关键点检测的特征图,并将其投影到图像上,如图4所示,以确认所得到的关键点坐标均位于目标物体内部。这个步骤对于确保实现所需功能非常关键,因此本文中详细描述了该步骤的实施与结果。

图4 关键点位置验证示意图

2.2.2 虚拟点数目

本文提出的网络关键在于利用关键点信息构造虚拟点云。然而,关键点的输出数量和虚拟点云的构造数量对检测效果和效率具有重要影响。因此进行了基于不同虚拟点云数量的对比实验。

结果如表1所示,随着虚拟点云数量的减少,模型的检测效果有所下降。这说明增加虚拟点云构造数量可以提供更多信息,有助于提高基于点云的目标检测性能。但过多的点云数量也会带来一些问题,例如消耗大量内存和训练时间过长等。因此,在实际应用中,需要根据实际情况确定虚拟点云数量,以实现最佳的检测性能和效率。

表1 单体素虚拟点云数目对比图

虚拟点云数量	汽车AP/%	推理时间/s
1	84.64	0.08
2	85.32	0.13
5	86.92	0.20

2.2.3 损失函数

本文提出的网络中,损失函数由关键点的定位偏差和整体三维包围框的偏差构成。然而,将这两个部分的损失直接相加,既无法体现各个模态检测的重要性,也不能体现两阶段之间的前后关系。此外,这种方法可能会增加迭代次数,导致收敛速度变慢。

因此,在网络的训练过程中尝试了不同的方法,包括增加损失偏差权重和直接将两部分损失相加,通过训练收敛过程进行了比较。实验结果如图5所示,将偏差权重引入损失函数后,模型的收敛速度明显加快,检测效果也得到了部分提升,橙色线为加入偏差权重后的损失变化。这种方法不仅可以更好地平衡两部分损失,还可以更好地表达不同检测模态的重要性,提高了模型性能的表现。

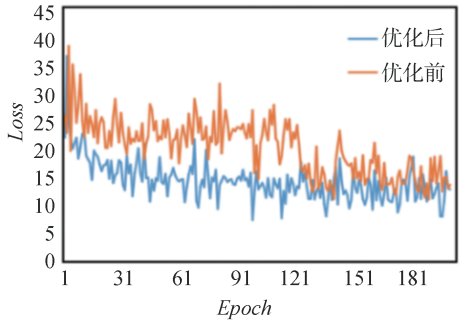

图5 损失变化前后对比图

2.2.4 多模态融合

图6为VPC-VoxelNet在KITTI验证集上点云空间中的检测结果,同时将检测结果投影到图像上,其中,蓝色包围框表示汽车的标签值,红色包围框表示非机动车的标签值,绿色包围框表示人的标签值,而橙色包围框表示检测出的目标是汽车,褐色包围框表示检测出的目标是非机动车,紫色包围框表示检测出的目标是人。

由图6(f)可见,当目标与目标之间出现严重遮挡的情况下,网络依然可以很好地检测出三维目标,这无疑是虚拟点云为这些目标增加了大量的关键信息。本文基于图像的关键点检测并在点云空间中构造虚拟点云,将目标置信度纳入点云体素编码,使激光点云获得了更多的信息和维度,充分处理了两种模态的数据,优化了检测的结果。

2.2.5 消融实验

本文进行了基于对象距离范围的不同子集的检测结果的对比[21],使用的数据集是KITTI验证集,结果如表2所示,其中方法1为SECOND网络,方法2为VPC-VoxelNet(Ours)。本文提出的方法是使用单目3D检测算法预测目标的中心点热力图,热力图目标中心区域会形成高斯核,该区域的特征值比无目标区域的特征值更高,后续构造虚拟点云所依据的关键点是选取热力图中特征值靠前的一些点,这样一个目标中心区域就会有多个点入选。

为探究本文所提出方法的优势,通过针对不同模态的检测也进行了消融实验,结果如表3所示。

本文在提出方法的基础上增加了点云检测算法的对比实验,消融实验结果表明在高精度激光雷达

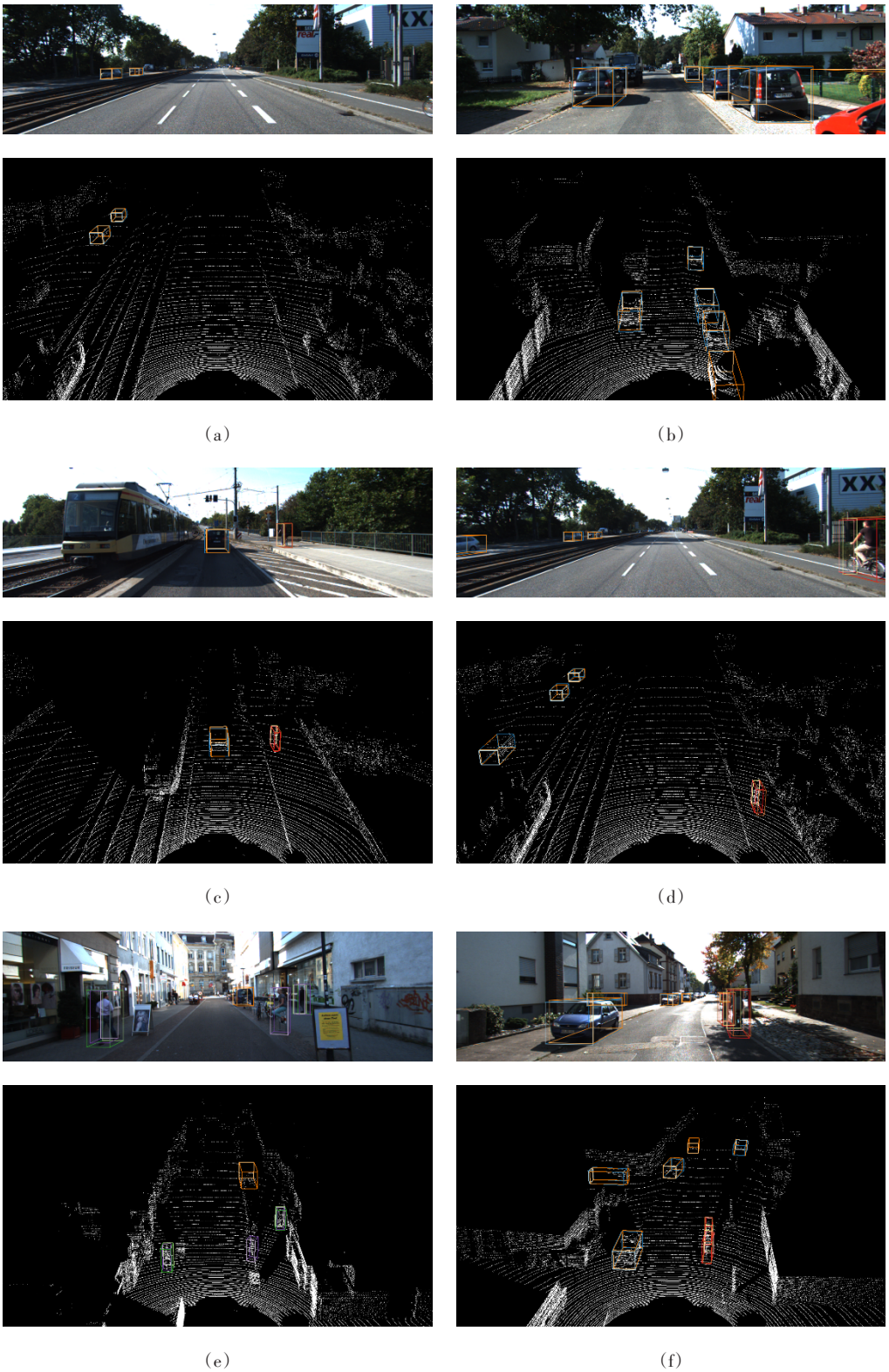

图6　在KITTI验证集中的部分检测结果

点云算法上依旧有略微提升,结果如表4所示,其中方法1为PointRCNN模型,方法2为基于PointRCNN改进的VPC-VoxelNet(Ours)模型。

2.3　实验结果

本文使用KITTI数据集对提出的多模态检测模型进行了实验,并将结果与几种仅激光雷达和多模

表2 不同距离检测结果比较

距离/m	方法	检测精度(汽车)/%		
		简单	中等	困难
0-10	1	85.47	88.78	89.13
	2	87.20	90.01	89.90
10-20	1	89.40	89.50	88.83
	2	89.54	90.42	87.63
20-30	1	85.52	86.11	77.19
	2	86.01	86.98	80.21
>30	1	27.32	53.69	51.33
	2	24.68	49.22	50.54

表3 不同模态检测结果比较

模态		检测精度(汽车)/%		
相机	雷达	简单	中等	困难
	√	83.34	72.55	65.82
√		82.33		
√	√	86.92	74.47	69.97

态3D物体检测方法进行了比较,对比结果如表5所示。对于车辆检测,本文提出的网络表现优良,与经典网络PointPillars[17]相比提高了4%,与多模态检测网络F-PointNet[18]相比提高了4%。

表4 不同点云算法检测结果比较

方法	检测精度(汽车)/%		
	简单	中等	困难
1	86.96	75.64	70.70
2	88.33	74.72	71.23

本文提出的3D检测网络在无障碍的目标检测中发挥优良,即使它们被遮挡,也可以很好地被检测到。并且远距离目标检测中也具有良好效果,准确性的提高主要由于同时处理了图像和激光点云信息,通过对图像获取关键点来构造虚拟点云,使点云空间中远距离目标的点云不再稀疏,因此对远距离和小物体具有更好的检测效果。但是,本文方法只使用了体素的方法,后续研究会针对如何将基于点的方法融入其中。

表5 KITTI测试集上不同网络目标检测精度对比

方法	传感器类型	汽车(3D)/%			汽车(BEV)/%		
		简单	中等	困难	简单	中等	困难
SECOND	激光雷达	83.34	72.55	65.82	89.39	83.77	78.59
PointPillars	激光雷达	82.58	74.31	68.99	90.07	86.56	**82.81**
PC-CNN-V2	激光雷达	85.57	73.79	65.65	90.87	**87.84**	80.52
BirdNet+	激光雷达	76.15	64.04	59.79	87.43	81.85	75.36
PV-RCNN	激光雷达	90.25	81.43	76.82			
PointRGBNet[19]	相机+激光雷达	83.99	73.49	68.56			
MV3D	相机+激光雷达	74.97	63.63	63.63	86.62	78.93	69.80
DMF	相机+激光雷达	77.55	67.33	62.44	84.64	80.29	76.05
AVOD-FPN	相机+激光雷达	83.07	71.76	65.73	90.99	84.82	79.62
F-PointNet	相机+激光雷达	82.19	69.79	60.59	91.17	84.67	74.77
PVF-DectNet	相机+激光雷达	84.60	75.00	69.70			
VPC-VoxelNet(ours)	相机+激光雷达	**86.92**	**74.47**	**69.97**	**92.24**	84.13	81.50

3 结论

本文提出了一种基于虚拟点云的二阶段多模态融合网络VPC-VoxelNet。该网络利用图像提取关键点信息,并根据关键点信息在点云空间中构造虚拟点云,并对所有点云增加维度和使用含置信度编码的体素,实现了RGB图像和激光雷达点云的有效融合。基于KITTI数据集的实验表明,本文提出的融合方法提高了检测结果的准确性,虚拟点云可以有效提高其检测效果。在接下来的研究中将考虑如何将更多的图像关键点信息加入到点云中,同时保证3D目标检测的精度和速度。

参考文献

[1] ALABA S Y, BALL J E. A survey on deep-learning-based LiDAR 3D object detection for autonomous driving[J]. Sensors, 2022,22(24).

[2] JISEN W. A study on target recognition algorithm based on 3D point cloud and feature fusion[C]. 2021 IEEE 4th International Conference on Automation, Electronics and Electrical Engineering (AUTEEE). IEEE, 2021: 630-633.

[3] MENDEZ J, MOLINA M, RODRIGUEZ N, et al. Camera-Li-

DAR multi-level sensor fusion for target detection at the network edge[J]. Sensors, 2021, 21(12): 3992.

[4] PESSACH D, SHMUELI E. A review on fairness in machine learning[J]. ACM Computing Surveys (CSUR), 2022, 55(3): 1-44.

[5] HUANG Siyuan, LIU Limin, FU Xiongjun, et al. Overview of LiDAR point cloud target detection methods based on deep learning [J]. Sensor Review, 2022, 42(5).

[6] QI C R, SU H, MO K, et al. PointNet: deep learning on point sets for 3D classification and segmentation[J]. CoRR, 2016, abs/1612.00593.

[7] LI Y, BU R, SUN M, et al. PointCNN: convolution on X-transformed points[J]. Advances in Neural Information Processing Systems, 2018, 31.

[8] ZHANG M, CUI Z, NEUMANN M, et al. An end-to-end deep learning architecture for graph classification[C]. Proceedings of the AAAI Conference on Artificial Intelligence, 2018, 32(1).

[9] ZHOU Y, TUZEL O. VoxelNet: end-to-end learning for point cloud based 3D object detection[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 4490-4499.

[10] YAN Y, MAO Y, LI B. Second: sparsely embedded convolutional detection[J]. Sensors, 2018, 18(10): 3337.

[11] 王麒. 基于深度学习的自动驾驶感知算法[D]. 杭州:浙江大学,2022.DOI:10.27461/d.cnki.gzjdx.2022.001691.
WANG Qi. Deep learning based autonomous driving perception algorithm [D]. Hangzhou: Zhejiang University, 2022. DOI: 10.27461/d.cnki.gzjdx.2022.001691.

[12] ZHANG X, WANG L, ZHANG G, et al. RI-Fusion: 3D object detection using enhanced point features with range-image fusion for autonomous driving[J]. IEEE Transactions on Instrumentation and Measurement, 2022.

[13] YU F, WANG D, SHELHAMER E, et al. Deep layer aggregation [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 2403-2412.

[14] 王宏任,陈世峰. 基于关键点检测二阶段目标检测方法研究[J].集成技术,2021,10(5):34-42.
WANG Hongren, CHEN Shifeng. Research on two-stage object detection method based on key point detection[J]. Journal of Integration Technology, 2021, 10(5): 34-42.

[15] DUAN K, BAI S, XIE L, et al. CenterNet: keypoint triplets for object detection[C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 6569-6578.

[16] WENG X, KITANI K. Monocular 3D object detection with pseudo-lidar point cloud[C]. Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops, 2019.

[17] LANG A H, VORA S, CAESAR H, et al. PointPillars: fast encoders for object detection from point clouds[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 12697-12705.

[18] QI C R, LIU W, WU C, et al. Frustum pointnets for 3D object detection from RGB-D data[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 918-927.

[19] 谢德胜,徐友春,陆峰,等. 基于多传感器信息融合的3维目标实时检测[J]. 汽车工程, 2022, 44(3): 340-349.
XIE Desheng, XU Youchun, LU Feng, et al. Real-time detection of 3D objects based on multi-sensor information fusion[J]. Automotive Engineering, 2022, 44(3): 340-349.

[20] LIU Z, WU Z, TÓTH R. Smoke: single-stage monocular 3D object detection via keypoint estimation [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2020: 996-997.

[21] GUO X, ZHANG Y, GONG L, et al. 3D object detection on voxels in spherical coordinate system [C]. 2021 7th International Conference on Big Data Computing and Communications (BigCom). IEEE, 2021: 286-293.