

基于点云数据的交通环境下单阶段三维目标检测方法

车爱博, 张辉, 李晨, 王耀南

引用本文

车爱博, 张辉, 李晨, 王耀南. 基于点云数据的交通环境下单阶段三维目标检测方法[J]. 计算机科学, 2022, 49(11A): 210900079-6.

CHE Ai-bo, ZHANG Hui, LI Chen, WANG Yao-nan. [Single-stage 3D Object Detector in Traffic Environment Based on Point Cloud Data](#) [J]. Computer Science, 2022, 49(11A): 210900079-6.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[基于多尺度特征融合和双重注意力机制的肝脏CT图像分割](#)

Liver CT Images Segmentation Based on Multi-scale Feature Fusion and Dual Attention Mechanism
计算机科学, 2022, 49(11A): 210800162-9. <https://doi.org/10.11896/jsjcx.210800162>

[基于大型场景下的多相机标定方法](#)

Multi-camera Calibration Method Based on Large-scale Scene
计算机科学, 2022, 49(11A): 211200054-6. <https://doi.org/10.11896/jsjcx.211200054>

[基于注意力机制与混合监督学习的钢轨表面缺陷检测模型](#)

Rail Surface Defect Detection Model Based on Attention Module and Hybrid-supervised Learning
计算机科学, 2022, 49(11A): 210800241-6. <https://doi.org/10.11896/jsjcx.210800241>

[基于数据增强和监督均衡的鱼眼图像行人检测优化方法](#)

Pedestrian Detection Optimization Method Based on Data Enhancement and Supervised Equalization in Fisheye Image
计算机科学, 2022, 49(11A): 210900070-6. <https://doi.org/10.11896/jsjcx.210900070>

[融合词性与声调特征的越南语语法错误检测](#)

Incorporating Part of Speech and Tonal Features for Vietnamese Grammatical Error Detection
计算机科学, 2022, 49(11): 221-227. <https://doi.org/10.11896/jsjcx.210900247>

基于点云数据的交通环境下单阶段三维目标检测方法

车爱博¹ 张 辉² 李 晨¹ 王耀南²

1 长沙理工大学电气与信息工程学院 长沙 410114

2 湖南大学机器人学院 长沙 410082

(392765762@qq.com)

摘 要 文中在 CIA-SSD 单阶段三维目标检测模型的基础上,将模型中空间语义特征融合方式进行改进,通过一种基于注意力机制的多通道融合模块对两特征进行融合,提出了单阶段检测方法 TFAF-SSD(Two-Feature Attentional Fusion Single-Stage object Detector),该方法主要由流形稀疏卷积网络提取点云的稀疏特征后,再由空间语义卷积层分别提取检测对象的空间语义特征,对融合后的输出特征进行预测,最后通过检测头输出最终的检测框。同时,文中还运用了不同于以往方法的数据增强方法,增强了模型的泛化性能,达到了提升检测精度的效果。在 KITTI 3D 公开数据集上进行了验证,在测试集中汽车检测方面得到了中等检测难度 AP 值为 83.77% 的检测结果,相比 CIA-SSD 模型的 80.28%,所提方法提升了 3.49%。

关键词: 点云数据;三维目标检测;稀疏卷积;特征融合;数据增强;KITTI 数据集

中图法分类号 TP183

Single-stage 3D Object Detector in Traffic Environment Based on Point Cloud Data

CHE Ai-bo¹, ZHANG Hui², LI Chen¹ and WANG Yao-nan²

1 School of Electrical and Information Engineering, Changsha University of Science and Technology, Changsha 410114, China

2 School of Robotics, Hunan University, Changsha 410082, China

Abstract Based on CIA-SSD single-stage 3D object detection model, this paper improves the spatial semantic feature fusion method of the model. A multi-channel fusion module based on attention mechanism is used to fuse the two features. The single-stage detection method TFAF-SSD(two-feature attentional fusion single-stage object detector) is proposed. After extracting the sparse features of point cloud by manifold sparse convolution network, the spatial semantic features of the detected objects are extracted by the spatial semantic convolution layer. After fusion, the output features are predicted. Finally, the final detection frame is output by the detection head. At the same time, different from the previous methods, this paper also uses the data enhancement method to enhance the generalization performance of the model to improve the detection accuracy. The proposed method is verified on KITTI 3D open data set, and the test result of vehicle detection in test set is 83.77%, which is 3.49 % higher than the 80.28% of CIA-SSD model.

Keywords Point cloud data, 3D object detection, Sparse convolution, Feature fusion, Data enhancement, KITTI dataset

1 引言

智能巡检系统是对交通巡检技术一次具有突破性创新的发展,它意味着告别存在严重安全隐患的人工巡检模式,利用机器人等智能化系统进行交通环境下的辅助巡检。交通环境中的智能巡检系统的核心内容是对特定目标进行检测、跟踪等,针对目标检测问题,作为计算机视觉的基础任务之一,其主要目的是在点云或图像序列中精确得出各种目标的类别和位置信息。目前,基于图像的二维目标检测工作已经取得显著进展,但由于二维目标检测对复杂场景的描述度不够,缺乏目标尺寸、姿态等物理参数信息,在实际应用中受到了一定限制,因此结合深度信息的三维目标检测任务引起了广大学者及研究机构的关注。

多传感器融合的三维目标检测方法可以利用不同种类的数据拟合方式来实现检测,其中的点云投影方法首先通过平面、柱面或球面投影将三维点转换为二维图像,再使用二维

卷积等对其进行处理,然后可以使用位置和维度回归恢复三维边界框。如 MV3D^[1]将 LiDAR 点云投影到鸟瞰图(BEV)以生成候选区域,AVOD^[2]对 MV3D^[1]进行了扩展,通过在候选区域生成阶段引入图像特征,来生成更高质量的候选框。MMF^[3]融合深度地图、激光雷达点云、图像和地图的信息,完成了深度完成、2D 目标检测和 3D 目标检测等多项任务。而仅利用激光雷达传感器进行点云数据处理是三维目标检测方法中检测效率较高、准确度也更具竞争力的方法。目前其主要有两种表现形式,一种是基于体素的,它将体素化应用于整个点云。VoxelNet^[4]利用 PointNet^[5]提取每个体素的特征,使用原始点集来生成体素特征,创建统一的点云表示。还有一种直接基于点的方法,以原始点云为输入来处理不规则情况,并根据每个点生成预测,试图减小 3D 空间中投影或量化造成的信息损失。PointNet^[5]介绍了这一领域的开创性工作,将分割后的三维点云数据作为输入进行对象分类和部分分割。该模型在 PointNet++^[6]中得到了进一步的扩展,

其中每一层以层次结构逐步编码更复杂的特性。而且仅利用点云数据的三维目标检测方法通常有两种类型:1)两阶段三维目标检测法通常在第一阶段生成区域建议,然后将这些感兴趣的区域输入第二阶段网络进行细化;2)单阶段的网络相比两阶段简单,输入点云数据后可直接回归一个阶段的 class 分数和边界框。PointRCNN^[7]就是典型的两阶段检测算法,它使用 PointNet++^[6]作为两个阶段的骨干,并设计了无锚策略来生成 3D 提案。在 PointRCNN^[7]的基础上,Part-A²^[8]将 PointNet++^[6]骨干网络替换为稀疏卷积网络,并在细化中提出感知 Roi 的点云特征池化。STD^[9]利用球形锚点和细分支生成方案,来减少正向锚点的数量。FPointNet^[10]采用检测或分割等 2D 机制过滤大部分无用点,并根据这些保留的有用点产生预测。StarNet^[11]直接在点云中进行最远点采样(Farthest Point Sampling, FPS),确保采样点在空间上分布均匀;然后在每个采样点周围形成若干个候选框,以图卷积网络提取点云特征,并产生最终的检测结果。PV-RCNN^[12]在第一阶段使用集合抽象模块,从多尺度体素特征中提取点特征来细化区域提案。

单阶段检测方法如 TANet^[13]提出了嵌入体素特征提取中的三重注意模块,并将其与提出的级联细化网络相结合。PointPillar^[14]将点云划分为柱子而不是体素进行特征提取,然后利用 2D 卷积检测架构进行目标定位。SECOND^[15]利用稀疏卷积和子流形稀疏卷积来取代传统的三维卷积。3DSSD^[16]提出了一种融合采样策略,将特征采样和基于点的最远点采样相结合,以获得更好的分类性能。SASSD^[17]提出了一种辅助网络与稀疏卷积网络并行的算法,通过插值体素特征回归每个点的 Box 中心和语义类。图 1 给出了不同三维目标检测方法在 KITTI 三维检测基准上的结果对比。

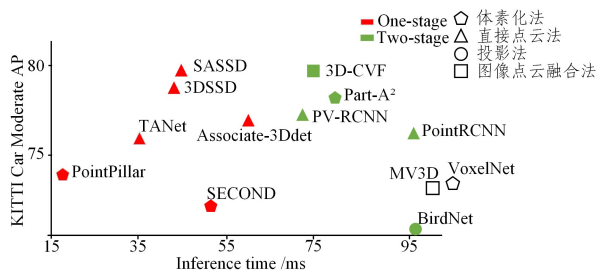


图 1 KITTI 三维检测基准上的三维目标检测结果对比

Fig. 1 Comparison of 3D object detection results on KITTI 3D detection datum

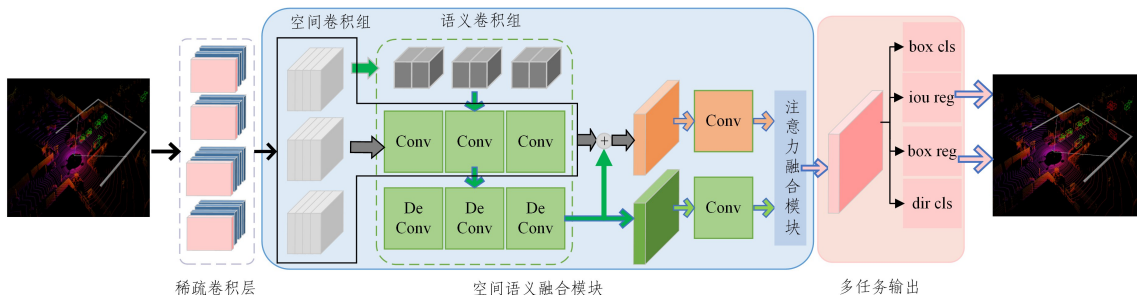


图 2 TFAF-SSD 网络结构

Fig. 2 TFAF-SSD network structure

2.3 子框架结构

2.3.1 流形稀疏卷积

特征提取主干网络使用了稀疏卷积层提取三维特征,

2 本文方法

2.1 三维目标检测问题描述

当巡检机器人处于交通场景中对周围环境进行感知时,需要知道周围物体的三维位置、姿态、结构等影响机器人操作和行进方向的重要信息,以此作为下一步动作的依据,或为周围其他物体发出提醒和警示,机器人可用激光雷达传感器作为感知工具进行环境感知,将得到的点云数据进行处理即可获得大量有用信息。本文以激光雷达传感器扫描出的点云数据为研究基础,主要解决以下 3 个问题:1)回归检测物体位置时必须同时考虑低层空间特征和高层抽象语义特征,然而往往通过堆叠的卷积层来提取特征映射中的高级抽象语义,这样导致了低级空间信息的质量下降,不能有效地获得具有丰富空间信息的鲁棒特征;2)初始的特征融合不能让模型得到充分表达,甚至会严重影响最终特征融合的权重,以致检测精度无法提升;3)对点云数据进行训练时,训练中遇到的主要问题是 Ground-Truth 太少,这大大限制了网络的收敛速度和最终性能。

2.2 整体框架设计

CIA-SSD^[18]模型提出了空间语义特征聚合模块,是一个轻量的 BEV 特征提取模块,可以提取鲁棒的 BEV 特征,提升了检测的器精度。本文在 CIA-SSD 模型的基础上,提出了一种改进的单阶段目标检测模型 Two-Feature Attentional Fusion Single-Stage Object Detector (TFAF-SSD),用于直接在点云数据中进行三维目标检测。如图 2 所示,首先,特征提取主干网络使用了稀疏卷积层提取三维特征,接着为了解决目标遮挡问题以提高检测精度,使用了一个轻量级的可自适应地融合高级抽象语义特征和低级空间特征的双特征注意力融合,利用深层和浅层卷积组提取不同特征信息的差异性,从而提取语义特征和空间特征,并利用反卷积进行特征恢复再进行融合处理。然后,CIA-SSD 采用注意力融合模块,本文在此基础上,探究融合机制,提出使用基于注意力机制的多尺度特征融合模块,将多尺度通道注意力模块中尺度不同的两个分支用于提取通道注意力权重。其中一个分支使用全局平均池化来提取全局特征的注意力,另一个分支直接使用 point-wise 卷积来提取局部特征的通道注意力,以进行语义特征和空间特征的融合。输出融合后的特征,再将其输入最后的多任务检测头,最后输出预测结果。

稀疏卷积层主要由 4 部分组成,每部分包括若干个子流形稀疏卷积(SSC)层和一个稀疏卷积(SC)层。具体来说,这四部分分别具有 2 层、2 层、3 层、3 层 SSC 卷积层,而在每一部分

子流形稀疏卷积(SSC)层的末尾附加一个稀疏卷积(SC)层,用于在3D特征图上进行2倍的降采样,最后将稀疏的体素特征转化为密集的特征映射,并将 z 中的特征连接,生成BEV特征映射作为TFAF模块的输入。

2.3.2 空间语义特征提取模块

为了检测交通环境中的汽车,必须精确回归汽车的位置,在确定汽车位置的过程中,必须同时考虑低层空间特征和高层抽象语义特征。在特征映射中提取高级抽象语义时,有学者提出可以通过堆叠卷积层进行提取,如通过BEV特征来提取模块采用的卷积层,然而这样操作往往导致低级空间信息的质量下降,难以提取高质量的空间特征。为了克服无法同时提取两个高质量的特征的缺点,本文在CIA-SSD模型的基础上通过基于注意力机制的多尺度特征融合模块将空间特征和语义特征进行融合,并将它们分割成两个BEV注意力图,在注意力图中多尺度融合模块建立了两个特征之间的依赖关系,从而实现了自适应特征融合。空间语义特征提取模块如图3所示。下文将介绍多尺度注意力融合模块是如何将低层空间特征和高层抽象语义特征进行融合的。

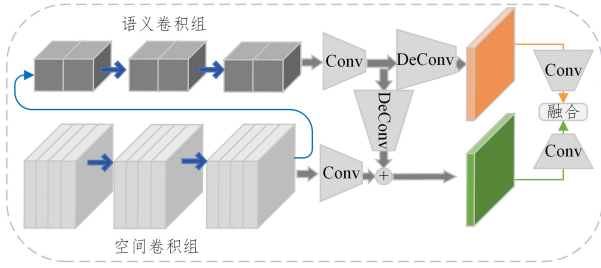


图3 空间语义特征提取模块

Fig.3 Spatial semantic feature extraction module

2.3.3 基于注意力机制的多尺度特征融合模块

在很多工作中,融合不同尺度的特征是提高检测性能的一个重要手段。低层特征分辨率更高,包含更多位置、细节信息,但是由于经过的卷积更少,其语义性更低,噪声更多。高层特征具有更强的语义信息,但是分辨率很低,对细节的感知能力较差。按照融合与预测的先后顺序,分类为早融合(Early fusion)和晚融合(Late fusion)。

(1)早融合(Early fusion)。先融合多层的特征,然后在融合后的特征上训练预测器(只在完全融合之后,才统一进行检测)。这类方法也被称为 skip connection,即采用 concat, add 操作。

(2)晚融合(Late fusion)。通过结合不同层的检测结果来改进检测性能(尚未完成最终的融合之前,在部分融合的层上就开始进行检测,会有多层的检测,最终将多个检测结果进行融合);1)feature 不融合,多尺度的 feature 分别进行预测,然后对预测结果进行融合;2)feature 进行金字塔融合,融合后进行预测。

本文在 concat 特征融合基础上提出使用基于注意力机制的多尺度特征融合模块,将学习到的 BEV 注意力图作为各自特征上的权重,并进行元素相加(见图4)来融合加权特征,使用尺度不同的两个分支来提取通道注意力权重。

首先利用多尺度通道注意力模块,使用尺度不同的两个分支来提取通道注意力权重。其中一个分支使用全局平均池化来提取全局特征的注意力,另一个分支直接使用 point-wise 卷积来提取局部特征的通道注意力。为了自适应融合丰富的

空间特征和上采样的语义特征,本文采用图4所示的注意力模块来解决由于不同尺度特征融合所带来的问题。

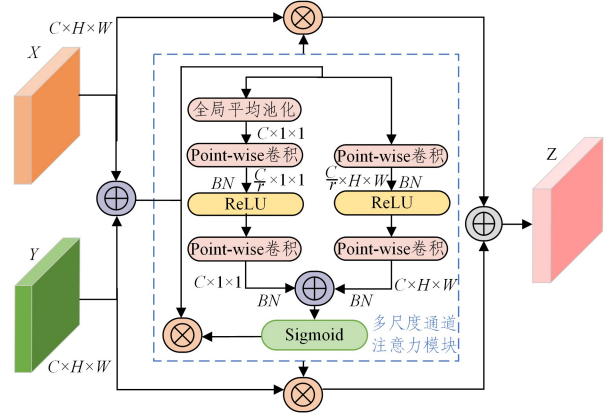


图4 基于注意力机制的多尺度特征融合模块(电子版为彩图)

Fig.4 Multi-scale feature fusion module based on attention mechanism

首先使用不同的空间池化尺寸获得多尺度特征,并在不同尺度特征上实施通道注意力操作。为了维持轻量模型,在注意力模块中将局部信息融入全局信息。每个分支选择使用的卷积是 1×1 卷积或称之为 Point-Wise 卷积(PWConv),将其作为局部通道信息整合器,其仅仅利用了不同通道的对应像素点做信息交互。为了不增加参数量,在计算局部通道信息时使用了一个瓶颈结构。局部通道信息的计算式如下:

$$L(X) = B(PWConv v_2(\delta(B(PWConv v_1(X))))))$$

$PWConv v_1$ 通过 1×1 卷积将输入的特征 X 通道数减少为原先的 $\frac{1}{r}$, B 表示 BatchNorm 层, δ 表示 ReLU 激活函数,

$PWConv v_2$ 又通过 1×1 的卷积将通道数目恢复成与原输入通道数目相同,因此局部通道信息 $L(X)$ 的形状和输入 X 的形状保持一致。如图4中蓝色虚线框内右侧分支即为局部通道信息的计算。

之后通过左侧分支计算得到全局通道信息 $g(X)$,使用全局平均池化的目的是得到的特征包含了全局信息。进而将计算得到的局部通道信息和全局通道信息进行融合,输出一个权重值,用于对输入特征做注意力操作后得到输出,具体操作如下:

$$X' = X \otimes M(X) = X \otimes \sigma(L(X) \oplus g(X))$$

其中, $\oplus$ 表示广播加法操作,这里是因为全局通道信息使用了全局平均池化操作,所以得到的特征高宽形状为 1×1 ,而局部通道信息与输入特征保持着相同的高宽尺寸,因此两者相加需要采取广播操作; $\otimes$ 表示两个 feature map 对应元素相乘。注意, $M(X) = \sigma(L(X) \oplus g(X))$ 表示多尺度通道注意力模块操作,对应流程图4中内测蓝色框部分。

本文场景为低层空间特征和高层语义特征的特征融合,因此在基于多尺度通道注意力模块的基础上可以将注意力特征融合模块表示为:

$$Z = M(X+Y) \otimes X + (1 - M(X+Y)) \otimes Y$$

其中, $Z \in R^{C \times H \times W}$ 是两个特征融合后的输出特征, $+$ 表示初始特征的简单集成,并且本文选择逐元素求和作为初始积分,融合权重 $M(X+Y)$ 由 0 到 1 之间的实数组成, $1 - M(X+Y)$ 也是如此,这使得网络可以在 X 和 Y 之间进行软选择或加权平均,本文选择的是元素相加来融合加权特征。

2.3.4 损失函数

将 Smooth-L1 损失应用于 IoU 回归计算 IoU 预测损失 L_{iou} 。类似于边界框回归损失 L_{bxx} , 本文只计算阳性样本的 IoU 预测损失 L_{iou} , L_{cls} 是分类损失, L_{dir} 是方向分类损失, 总损失 L 定义为:

$$L=L_{cls}+\omega L_{bxx}+\mu L_{dir}+\lambda L_{iou}$$

本文设置 $\omega=2.0, \mu=0.2, \lambda=1.0$ 。

3 实验与分析

本文在 KITTI 3D 数据集^[19]上评估了本文提出的单阶段检测方法。KITTI 数据集是自动驾驶领域三维物体检测的经典数据集, 目前它的检测精度仍然有较大的提升空间, 数据集包含 7481 个训练样本和 7518 个测试样本。在此基础上, 将训练数据进一步划分为 3712 个样本的训练集和 3769 个样本的验证集。汽车类别是 KITTI 中最重要的检测类别, 也是样本数量最大的类别, 因此在汽车类别上的检测精度是评估模型的关键尺度, 故本文主要是对交通场景下最常说的汽车类别进行实验, 并使用平均精度 (AP) IoU 阈值为 0.7 作为评价指标。基准考虑了 3 个难度级别: 基于不同对象大小、遮挡状态和截断级别所对应的简单、中等和困难。

3.1 实验细节

本文实验所用到的开发环境为: 1) Ubuntu 16.04 操作系统; 2) CPU 为 Inter core i7-6700K; 3) GPU 为 NVIDIA GeForce GTX 1008 8GB; 4) 模型和对比实验均在开源框架 Pytorch==1.4.4 下完成。本文考虑 4 种类型的数据扩充来增强 AFTF

模型的泛化能力。第一种类型是对整个点云的全局增强, 包括随机旋转、缩放和翻转。第二种类型是在地面真实物体周围点云的一部分上进行局部增强, 包括随机旋转和平移。第三种引入了一种新的数据增强方法, 来提高收敛速度和性能, 训练中遇到的主要问题是 GT 太少, 这大大限制了网络的收敛速度和最终性能。为了解决这一问题, 本文引入了一种数据增强方法。首先, 从训练数据集中生成一个数据库, 包含所有 GT 的标签及其相关的点云数据 (GT 的 3D 包围框内的点)。然后, 在训练过程中, 从这个数据库中随机选取几个 ground truth, 通过拼接的方式将其引入到当前的训练点云中。使用这种方法可以大大增加每个点云的 GT 的数量, 并模拟存在于不同环境中的物体。最后, 为了提高正样本的质量, 本文将难度等级不归属于容易、中等和困难的对象过滤掉, 并将相似类别的对象作为目标, 如用 van 代替 car, 以缓解训练时的模型混淆。

3.2 与当前先进方法比较

首先本文将本文提出的 TFAF-SSD 3D 点云目标检测算法与其他最先进的方法进行了比较。如表 1 所列, 本文将方法应用在 KITTI 测试集上, 可以看到本文方法在 3D 和 BEV 检测任务中都取得了较好的成绩。具体来说, 本文的骨干网络是建立在基础模型 CIA-SSD 上的, 本文在 3D 检测方面领先本文基础模型 CIA-SSD (2.55%, 3.49%, 3.69%)。此外, 两阶段检测方法由于额外的第二级细化, 通常比单阶段检测方法性能更好, 但本文提出的单阶段检测方法在中等难度的 AP 上仍比大多数最近的两阶段检测方法 (如 PV-RCNN, STD 和 Part-A2) 的性能好约 2.6%~5.5%。

表 1 在 KITTI 测试集上本文方法与其他三维目标检测方法的性能比较

Table 1 Performance comparison of the proposed method and other 3D object detection methods on KITTI test set

检测方法	数据方式	AP _{BEV} /%			AP _{3D} /%		
		简单	中等	困难	简单	中等	困难
1. MV3D	LiDAR+RGB	86.49	78.98	72.23	74.97	63.63	54.00
2. F-PointNet	LiDAR+RGB	91.17	84.67	74.77	82.19	69.79	60.59
3. AVOD	LiDAR+RGB	89.75	84.95	78.32	76.39	66.47	60.23
4. PointRCNN	LiDAR	92.13	87.39	82.72	86.96	75.64	70.70
Two-stage	5. F-ConvNet	91.51	85.84	76.11	87.36	76.39	66.69
	6. Part-A2	91.70	87.97	84.61	87.81	78.49	73.51
	7. MMF	93.67	88.21	81.99	88.40	77.43	70.22
	8. STD	94.74	89.19	86.42	87.95	79.71	75.09
	9. PV-RCNN	94.98	90.65	86.14	90.25	81.43	76.82
One-stage	10. VoxelNet	87.95	78.39	71.29	77.82	64.17	57.51
	11. PointPillars	90.07	86.56	82.81	82.58	74.31	68.99
	12. SECOND	89.39	83.77	78.59	83.34	72.55	65.82
	13. CIA-SSD	—	—	—	89.59	80.28	72.87
TFAF (ours)	LiDAR	94.30	89.25	87.08	92.14	83.77	76.56

不同方法在 KITTI 测试集上的评价结果如图 5 所示, 对于每种方法绘制了 P-R 曲线, 并标注了中等检测难度的 AP 值, 单阶段方法如图 5 中的虚线所示。

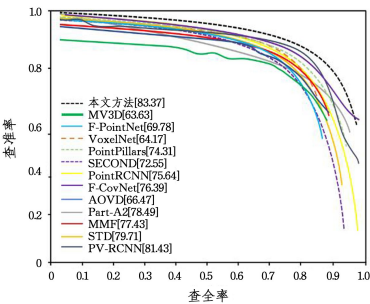


图 5 不同方法在 KITTI 测试集上的 P-R 曲线

Fig. 5 P-R curves of different methods on KITTI test set

从图 5 可以进一步看出, 在不同的查全率设置下, 本文方法优于先进的方法, 说明本文方法获得了更好的检测覆盖率和准确率。

3.3 消融实验

本节对本文方法的有效性进行了全面的分析。本文首先在 KITTI 验证集上评估本文方法, 并与以前的技术的结果进行比较, 如表 2 所列, 本文也在图 6 中展示了一些预测结果, 红色框为预测框, 绿色框为真实框, 本文将从 LiDAR 检测到的 3D 边界框投影到 RGB 图像中, 以便更好地可视化, 可视化结果显示本文方法可以在不同场景下生成高质量的三维边界框。接着本文研究了辅助任务和 TFAF 模块对模型的影响, 结果如表 3 和表 4 所列。

表 2 在 KITTI 验证集上和其他三维目标检测方法的性能比较
Table 2 Performance comparison of the proposed method and other
3D target detection methods on KITTI validation set

方法	AP _{3D}		
	简单	中等	困难
VoxelNet	81.97	65.46	62.85
SECOND	87.43	76.48	69.1
TANet	88.21	77.85	75.62
PointRCNN	88.88	78.63	77.38
PointPillars	82.58	74.31	68.99
3DSSD	89.71	79.45	78.67
7. CIA-SSD	90.04	79.81	78.80
TFAF(ours)	93.23	84.01	81.16

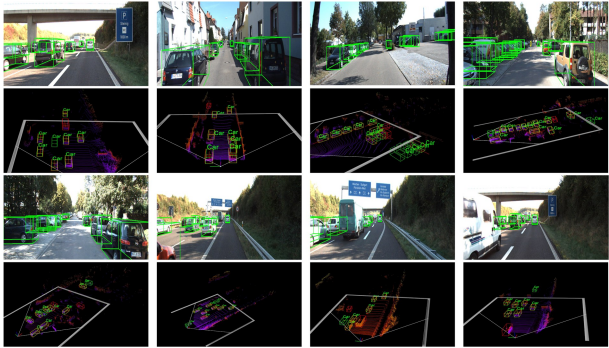


图 6 KITTI 验证集的可视化结果(电子版为彩图)
Fig. 6 Visualization results of KITTI validation set

表 3 不同数据处理方式对检测性能的效果比较
Table 3 Comparison of effects of different data processing methods
on detection performance

glo.	loc.	gt aug.	diff.	简单/%	中等/%	困难/%
				85.12	79.49	76.09
✓				90.36	81.49	78.02
✓	✓			91.46	82.89	71.45
✓	✓	✓		92.31	83.01	79.56
✓	✓	✓	✓	93.11	83.84	80.77

表 4 不同融合方式对检测性能的效果比较
Table 4 Comparison of effects of different fusion methods on
detection performance

融合方法	AP _{3D}		
	简单	中等	困难
直接跳跃连接融合	71.26	65.49	59.30
不同层检测结果融合	65.44	58.26	52.11
多通道注意力融合	86.52	77.40	70.81

3.3.1 验证集上的结果分析

本文方法在 KITTI 测试集的简单和中等难度上有着比较优秀的实验结果,但是在验证集上的效果得到了进一步提升,如表 2 所列。此外,本文方法在测试集的困难程度上低于最先进的方法,而在验证集上效果是最好的。本文认为,这种不一致性可能是由 part- A2 中提到的 KITTI 验证集和测试集分割之间不匹配的分布造成的。

3.3.2 数据处理方式对检测结果的影响

表 3 列出了各种数据处理技术对 baseline 模块的影响结果。glo, loc, gt aug, diff 分别表示全局增强、局部增强、Ground Truth(GT)增强,以及过滤掉难度不属于容易、中等

和困难级别的物体进行 GT 增强。全局和局部的增强都有效地改善了简单和中等难度的 AP 值。此外,GT 增强,过滤掉难度不属于容易、中等和困难等级的物体进行 GT 增强,可以有效地提高所有难度级别的 AP 值。由此可见,本文采用的数据处理技术有助于更好地实现三维目标检测技术。

3.3.3 多通道注意力融合模块对检测结果的影响

本文将点云数据通过流形稀疏卷积层和空间语义特征提取层输出的特征应用不同的融合特征方式进行融合后输出,融合方式包括直接跳跃连接融合、不同层检测结果融合以及本文提出的基于注意力机制的多通道特征融合方式,融合语义特征和空间特征,如表 4 所列,并通过对比不同融合方式对检测精度的影响表现出本文融合方式的优势。

结束语 本文在三维检测模型 CIA-SSD 基础上提出了一种基于点云数据的交通环境下单阶段三维目标检测方法,此方法在流形稀疏卷积网络提取点云的稀疏特征后,再由空间语义卷积层分别提取检测对象的空间语义特征,然后通过一种基于注意力机制的多通道融合模块对两特征进行融合,进而输出融合后的特征进行特征预测,最后通过检测头输出最终的检测框。本文方法提出的基于注意力机制的双特征融合模块对于解决遮挡程度高检测难度大的目标具有明显的检测精度方面的提升。同时,本文还运用了不同于以往方法的数据增强方法,增强了模型的泛化性能。实验结果表明,TFAF-SSD 在 KITTI 数据上的检测精度与现有的先进三维目标检测方法相比不分伯仲,在测试集中汽车检测方面得到了中等检测难度 AP 值为 83.77% 的检测结果。

参 考 文 献

[1] CHEN X,MA H,WAN J,et al. Multi-view 3d object detection network for autonomous driving[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 1907-1915.

[2] KU J,MOZIFIAN M,LEE J,et al. Joint 3D proposal generation and object detection from view aggregation[C]//Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems. Los Alamitos: IEEE Computer Society Press, 2018:1-8.

[3] LIANG M,YANG B,CHEN Y,et al. Multi-task multi-sensor fusion for 3d object detection[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019: 7345-7353.

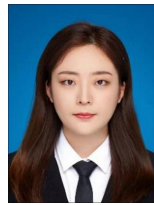
[4] ZHOU Y,TUZEL O. Voxelnet:End-to-end learning for point cloud based 3d object detection[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 4490-4499.

[5] QI C R,SU H,MO K,et al. PointNet:Deep learning on point-sets for 3d classification and segmentation[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017:652-660.

[6] RUI C,QI Z T,YI L,et al. Pointnet++ :Deep hierarchical feature learning on point sets in a metric space[C]// Conference and Workshop on Neural Information Processing Systems. 2017.

[7] SHI S,WANG X,LI H. PointRcnn:3d object proposal genera-

- tion and detection from point cloud[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019:770-779.
- [8] SHI S S, WANG Z, WANG X G, et al. Part-a² net: 3d part-aware and aggregation neural network for object detection from point cloud[J]. arXiv:1907.03670, 2019:6.
- [9] YANG Z T, SUN Y N, LIU S, et al. STD: Sparse-to-Dense 3D Object Detector for Point Cloud[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). 2019:1951-1960.
- [10] 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.
- [11] NGIAM J, CAINE B, HAN W, et al. StarNet: targeted computation for object detection in point clouds[J]. arXiv:1908.11069, 2019.
- [12] SHI S, GUO C, JIANG L, et al. PV-RCNN: Point-Voxel Feature Set Abstraction for 3D Object Detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2020:10529-10538.
- [13] LIU Z, ZHAO X, HUANG T, et al. TANET: Robust 3D Object Detection from Point Clouds with Triple Attention[C]//AAAI. 2020:11677-11684.
- [14] LANG A H, VORA S, CAESAR H, et al. PointPillars: Fast Encoders for Object Detection from Point Clouds[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019:12697-12705.
- [15] YAN Y, MAO Y, LI B. SECOND: Sparsely Embedded Convolutional Detection[J]. Sensors, 2018, 18(10):3337.
- [16] YANG Z, SUN Y, LIU S, et al. 3DSSD: Point-based 3D Single Stage Object Detector[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2020:11040-11048.
- [17] HE C H, ZENG H, HUANG J Q, et al. Structure Aware Single-stage 3D Object Detection from Point Cloud[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2020:11870-11879.
- [18] ZHENG W, TANG W L, CHEN S J, et al. CIA-SSD: Confident IoU-Aware Single-Stage Object Detector From Point Cloud [C]//AAAI. 2021.
- [19] GEIGER A, LENZ P, STILLER C, et al. Vision Meets Robotics: The KITTI Dataset[J]. The International Journal of Robotics Research, 2013, 32(11):1231-1237.



CHE Ai-bo, born in 1995, master. Her main research interests include machine vision and deep learning.



ZHANG Hui, born in 1983, Ph.D. His main research interests include intelligent robot vision detection, deep learning image recognition, and robot intelligent control.