

基于深度神经网络的自动驾驶场景三维目标检测算法

陆慧敏, 杨 朔

(九州工业大学工学部, 福岡 804-8550, 日本)

摘要: 在自动驾驶场景中,使用激光雷达相机获取精确度较高、可感知距离的点云数据,因此,有效利用点云数据,实现目标检测是完成自动驾驶任务的关键技术. 点云数据本身具有稀疏性、无序性和数据量较大的问题,传统的深度学习目标检测方法难以有效处理提取点云特征和满足准确度要求. 针对这一现状,提出一种体素化卷积网络与多层感知机模型融合的三维目标检测方法,利用体素化卷积网络提取点云数据的全局特征,结合多层感知机所提取点云数据的局部特征与距离关系,再利用候选三维区域检测方法,可以改进三维目标分类与位置预测的精度和速度. 采用德国卡尔斯鲁厄理工学院提供的 KITTI 自动驾驶数据集,对所提出的方法与经典的方法进行对比实验. 结果表明,本研究所提出的方法比以往的方法在精度上有较大提升.

关键词: 自动驾驶; 点云数据; 三维目标表征; 三维目标检测; 深度神经网络; 模型融合

中图分类号: U 461; TP 308

文献标志码: A

文章编号: 0254-0037(2022)06-0589-09

doi: 10.11936/bjntxb2021100027

Three-dimensional Object Detection Algorithm Based on Deep Neural Networks for Automatic Driving

LU Huimin, YANG Shuo

(Graduate School of Engineering, Kyushu Institute of Technology, Fukuoka 804-8550, Japan)

Abstract: In the automatic driving scene, LiDAR camera is usually used to obtain the point cloud data with high accuracy and perceptible distance. Therefore, achieving object detection by effectively using point cloud data is the key technology to complete the automatic driving task. Point cloud has the problems of sparsity, disorder, and large amount. Traditional deep learning object detection method is difficult to effectively extract features map and meet the accuracy requirements. This paper proposed a 3D object detection method based on the fusion of voxel convolution network and multi-layer perception model. The voxel convolution network was used to extract the global features of point cloud, combined with the local features and distance relationship of point cloud extracted by multi-layer perception. It can improve the accuracy and speed of 3D object classification and position prediction. In this paper, the KITTI dataset was used to compare the proposed method with the classical method. According to the experimental results, the accuracy of the proposed method is significantly improved than the previous methods.

Key words: autonomous driving; point cloud; 3D object feature representation; 3D object detection; deep neural networks; model fusion

收稿日期: 2021-10-27; 修回日期: 2021-12-06

作者简介: 陆慧敏(1985—), 男, 副教授, 日本九州工业大学, 主要从事人工智能、深海观测和机器人技术方面的研究,

E-mail: riku@cntl.kyutech.ac.jp

通信作者: 杨 朔(1994—), 男, 博士研究生, 日本九州工业大学, 主要从事计算机视觉、机器人智能控制方面的研究,

E-mail: yang.shuo264@mail.kyutech.jp

近年来,随着对人工智能技术的深入研究和激光雷达相机的广泛应用,针对点云数据的三维目标检测方法成为机器人控制技术和自动驾驶领域的研究热点问题之一.与普通相机拍摄的二维图像相比,点云数据包含物体的深度与几何信息,不仅可以帮助检测类别和定位,还可以有效给出物体的三维空间信息.因此,在自动驾驶和机器人抓取等任务中,针对点云数据的三维目标检测方法是解决问题的关键^[1-2].图1展示了自动驾驶场景与场景点云数据.

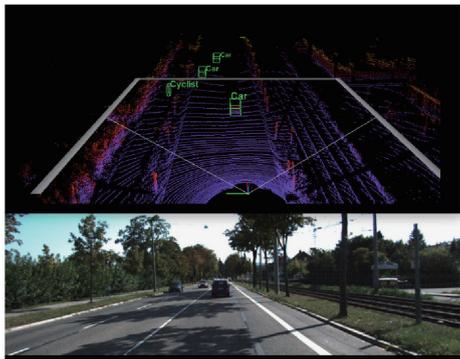


图1 自动驾驶场景与三维点云数据示例

Fig.1 Example for 3D point cloud and autonomous driving

在实际的应用中,点云数据处理也面临着大量问题,例如点云的稀疏性、无序性和数据处理量较大等.传统的检测方法,如时间空间聚类 and 分类方法,在数据量和场景复杂度增加条件下,模型处理的点云数据量有限,且无法完成正常的运行,也难以实现准确的定位和分类.因此,处理实际场景的点云数据实现物体检测是一个具有挑战性的问题.

随着深度学习理论在二维检测方法精度与速度上的突破,许多基于二维图像的检测方法被应用在点云数据处理上,以实现三维目标检测.这类方法采用基于学习的方式,构造深度神经网络从原始点云中提取三维目标特征,并采用区域生成方法实现三维目标的分类与位置框定位.典型的方法有基于体素转换的方法^[3]和基于点云学习的方法^[4].

基于体素转换的方法是将离散点云在三维空间中进行体素化处理,构造出规则的空间矩阵向量,经过转换后的数据可以应用三维卷积神经网络进行特征提取.其优点在于构造出全局特征可以有效提高分类和位置定位精度,问题在于转换过程中会产生局部特征细节的损失.基于点云学习的方法是通过构造多层感知神经网络,实现将点云直接输入进行学习,并构造出目标特征.其优点在于点云之间的特

征关系更加细化,对于局部特征可以有效抽出,问题在于输入数据量会影响网络处理速度,难以应用到实际计算设备.

所以,本文针对上述方法问题,提出基于深度神经网络的自动驾驶场景三维目标检测算法.通过将基于体素化卷积的特征提取网络与点云学习网络模型进行特征融合,提高对三维点云目标的表征能力,并采用基于候选区域生成的检测框架完成对三维目标的检测任务.本文总体内容可以总结如下.

1) 建立一个全新的基于三维点云的特征提取网络模型,利用体素化卷积网络完成全局特征的提取与表征,并使用点云学习网络完成局部特征提取,融合2类特征以实现建立更优的三维目标特征模型.

2) 建立基于候选区域生成的检测框架,利用K-means聚类方法从数据中计算候选框尺寸和数量,通过两阶段方式对候选区域进行筛选和位置回归,完成最终的分类与定位任务.通过实验验证,这种方法可以有效减少无效候选区域数量,并提高训练效率.

3) 更加优化的数据训练方法,利用预训练模型策略和数据增广方法(包括几何变换与加噪),优化三维目标检测的网络训练过程.

1 相关研究

由于软件和硬件发展瓶颈的问题,自动驾驶场景中的三维目标检测算法很大程度上借鉴于二维目标检测算法.传统的检测方法受限于模型泛化性和数据处理量有限的问题.基于深度学习的二维目标检测方法研究已经非常成熟,例如Faster R-CNN^[5]、YOLO算法^[6]等,均已被应用到工业级的检测中.三维目标检测算法则借鉴了这2类经典的二维检测框架,随着研究的深入,从最初利用多幅二维图像进行三维物体预测,到利用二维图像和深度图,再到利用点云数据直接完成物体位置和类别预测.相比于二维图像和深度图像,点云对于三维物体的形状和空间位置信息具有更好的刻画.因此,点云数据的处理与表征,是解决自动驾驶场景三维目标检测问题的关键研究内容.

基于点云数据的三维目标检测任务,在自动驾驶场景中的具体实现为:算法需要在三维点云数据上,建立目标的特征模型,进而确定目标在场景中的位置(通常为三维矩形框区域)和语义类别.根据不同方法所使用的特征区域不同,可以将

目标检测任务划分为2类:基于全局特征的目标分类与姿态估计^[7-9]和基于局部特征的目标分类与姿态估计^[10-12]。两者最大的区别在于,全局特征来自于点云对应二维映射图像或三维空间深度图像,局部特征来自于局部点云与点云之间的距离特征关系。

针对全局特征的三维目标检测算法研究,是早期深度学习三维目标检测研究的重点方向,其解决了点云处理和直观二维图像的映射问题,例如基于体素化的方法和基于多视图的方法。VoxelNet^[13]是一个基于点云体素化的卷积神经网络。该模型利用体素化方法,使点云可以利用卷积神经网络进行特征提取,并结合检测框架进行位置预测和类别估计。对于多视图的方法,如MV3D^[14]使用点云和鸟瞰图作为输入。

在三维目标检测中,鸟瞰图比前视图/图像平面有以下优势。首先,物体在投射到鸟瞰图时保持物理尺寸,因此有小的尺寸变化,这在前视图/图像平面不是这样的情况。其次,鸟瞰图中的物体占据不同的空间,从而避免遮挡问题。该方法用紧凑的多视图表示对稀疏的三维点云进行编码,该网络从三维点云的鸟瞰图表示中生成三维候选框从而进行目标检测。但是,这2类模型的问题在于点云体素化或鸟瞰图生成过程中,存在特征信息的丢失,影响网络对尺寸较小物体的特征提取。同时,模型依赖于人工设定的体素化和二维转换采样的体积和尺寸大小,易造成细节信息的损失,影响物体位置的估计^[15]。

针对局部特征的分类与姿态估计算法研究,是近年来三维目标检测任务探索的热点方向^[16]。该类方法通过更加细化的局部点特征,估计目标在三维空间中的位置,并且其不再使用复杂的多次数据处理操作,整体方法更加简洁。如PointNet++网络^[17]先将点云在空间上对齐,再通过多层感知机将其映射到高维的空间上。这时对于每一个点,都有一个1024维的向量表征,而这样的向量表征对于一个三维的点云明显是冗余的,因此这个时候引入最大池化操作(即对称函数,不改变点云的排列不变性),得到点云的局部关系,最后使用全连接层输出预测结果。但是,该类方法主要构造局部点之间距离关系特征,忽略了目标在空间中表现的纹理等特征信息,因此,总体的预测精度仍然难以满足当前的自动驾驶场景三维目标检测任务需求。

所以,本文将当前的2类特征提取方法进行有

效结合,构造一个统一的端到端检测网络实现自动驾驶场景目标检测任务。通过相关研究分析,这也是未来三维目标检测算法研究的重要方向。

2 基于深度神经网络三维目标检测算法

2.1 模型结构

主要介绍本文所提出的基于深度神经网络的三维目标检测算法模型结构。本文为解决三维点云数据的表征问题,设计了一个将全局特征与局部特征融合的多模型表征网络。该模型总体结构可以分为3个部分:全局特征抽取与表征模型、局部特征抽取与表征模型和检测模型。本文算法的总体模型结构如图2所示。

1) 全局特征抽取模型,其主体结构为体素化处理部分与三维卷积神经网络。为了提取三维目标的语义特征以及上下文信息,三维卷积网络可以提供更好的语义信息表达,并且可用于候选区域框的生成。但是,由于点云数据的离散特点,造成卷积网络无法直接进行计算。因此,通过对点云进行体素化处理,生成有规则的矩阵张量,再利用卷积神经网络进行全局特征抽取。

在分析当前的三维卷积神经网络时,发现仅使用下采样卷积,所获得的特征图对后续的检测精度和分类精度都较低,特别是对于远处的小物体车辆难以检出。因此,本文提出基于跨层特征融合的三维卷积神经网络,以提高特征提取网络的表征稳定性。

网络均使用 3×3 卷积核为主体,由5层下采样卷积与4层上采样卷积构成,其中上采样卷积将浅层特征图分别以2、4和8倍的尺寸进行特征融合,融合方式为同一感受野特征图上进行通道组合的方式完成。本文的全局特征提取网络不仅通过加深网络来改善语义分类错误,而且通过特征融合来提取细粒度特征改善对远处小物体车辆的检测能力。同时,最后大尺度的特征图输出可以给出更多候选区域框。

2) 局部特征抽取模型,其主体结构为最远点采样模块和点云学习网络。通常输入点云的数量较为庞大,需要进行一定的预处理,减少后续网络冗余计算,因此,本文采用最远点采样方法,对原始点云数据进行有效的点数量筛减,降低网络计算的内存与运算量。点云学习网络由多层感知机构成,用于计算点云之间距离等关系的特征计算与抽取。相比于卷积神经网络,其在提取点云之间细节信息具有更

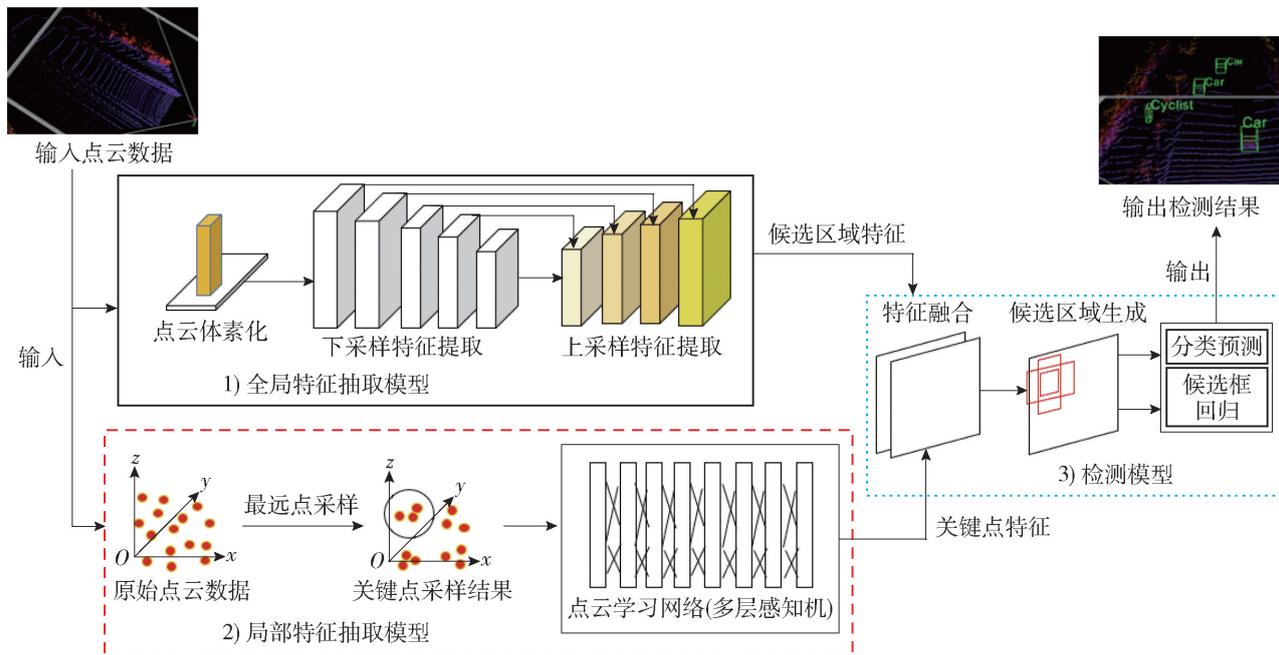


图2 基于深度神经网络的自动驾驶场景三维目标检测算法模型结构

Fig. 2 Module of 3D object detection algorithm for automatic driving based on deep neural networks

好的表征能力。因此,本文使用该模型对三维目标的局部特征进行抽取,并将其作为全局特征的细节信息补充。

为了与全局特征抽取模型进行特征区域的对应,本文实现了卷积区域与点云学习网络之间的对应映射,以在后续进行特征融合时,实现关键点特征可以有效放入正确的感受野区域,保证局部关键点特征有效补充全局特征。

3) 检测模型,本文基于经典两阶段检测算法 Faster R-CNN 思想^[5],构造检测框架。但是,本文模型仅完成一次分类和候选框回归计算,以提高整体模型的速度。在检测模型中,其主要完成特征模型融合、候选区域生成和分类与候选框回归计算。首先,对于特征模型融合,主要以第1部分全局特征抽取模型提供的特征图为主体,将第2部分局部特征抽取模型的局部关键点特征与其进行结合,通过映射相同感受野位置,将局部关键点特征融入全局特征中。其次,通过前期对数据真值标注框进行统计,用 K-means 聚类方法设计候选框的宽高,并在融合的特征图上进行候选框生成。通过设置真值与预测值的重叠比率,进行有效的包围框筛选。最后,在得到初步筛选的候选框上,进行分类与位置框回归计算,得到精确的目标类别与包围框。

2.2 数据增广

在对检测模型的训练策略中,数据增广是提高

目标分类和定位精度的有效方法。对于二维检测算法研究,数据增广技术相对成熟,例如图像旋转、几何变换、噪声与图像对比度变换等方法,这些方法均在一定程度上对深度学习模型的训练有效帮助,提高模型的鲁棒性、训练效率和模型质量。数据增广训练方法也被认为是基于深度学习检测算法训练的基本方式。

相比于二维图像的数据增广,基于点云的三维目标检测方法的数据增广训练则具有一定的技术复杂性和难度。其原因如下。

1) 点云的数据格式和特征与二维图像在空间结构、颜色特征等方面均存在显著差异。

2) 点云的稀疏性和无序性对使用数据增广训练方法有影响,几何变换或加噪方法均会改变点云位置和结构,反而影响正确的网络训练。

3) 可视化点云的方式多样,不同的可视化方式对于点云 z 轴计算方式也不同,难以给出坐标统一的数据增广方法。

通过对上述问题的分析,本文提出使用在线数据增广的几何位置变换方法,来实现深度模型训练的数据增广。在线数据增广,是在模型训练过程中,对输入数据进行调整,实现模型增广训练。

该方式的优点在于不需要占用大量的内存,生成的数据不会存储,仅在训练阶段中使用,训练后内存就会被释放。同时,在训练过程中进行几何位置

变换,能够实现统一的坐标系下实现三维点云目标的位置变换,不会产生坐标系转换问题,更容易完成三维目标检测模型的训练操作.此外,本文的数据增广策略为端到端方式,不额外占用存储空间,训练效率提升,也提高深度模型的应用价值.

2.3 多任务损失函数

在三维目标检测中共需要完成2个任务,目标语义分类和三维包围框定位.因此,基于深度学习理论,使用多任务损失函数作为训练目标函数.其中,对于语义分类部分,仍然采样分类精度较好的 Softmax 交叉熵损失函数 L_{cls} ;对于包围框的预测部分,为了减少计算量和维度,使用平滑的 $L1$ 损失函数作为三维包围框的回归计算函数 L_{reg} .所以,本文多任务损失函数 L_{loss} 可以表示为

$$L_{loss} = \alpha L_{cls} + (1 - \alpha) \frac{1}{N_{pos}} \sum_{y \in (x, y, z, l, w, h, \theta)} L_{reg}(y^*, y) \quad (1)$$

式中: y^* 、 y 分别为预测结果与真实值; N_{pos} 为正样本数据量;三维包围框的参数为 $(x, y, z, l, w, h, \theta)$,其中, x, y, z 为三维包围框的中心坐标, l, w, h 分别为三维包围框的长度、宽度和高度, θ 为三维包围框在 x, y 平面中的旋转角度.

此外,为了更加平衡回归与分类任务学习偏差,本文使用 α 参数去自适应调整模型训练过程中对于分类和回归任务的学习比率.

3 实验验证

3.1 数据库与实验环境简介

本文分别在已公开的数据库 KITTI 3D 目标检测数据库^[18]和 Waymo 3D 数据库^[19]上进行了模型验证与方法对比实验.

KITTI 3D 目标检测数据库,由德国卡尔斯鲁厄理工学院在自动驾驶场景中采集完成,数据集包含二维图片和雷达点云三维数据,该数据库的三维目标检测部分共包含 7 481 个训练数据和 7 518 个测试数据,分别属于 10 个类别.本文共对 2 个类别进行了更加详细的模型分析,分别为汽车和自行车,主要原因该 2 个类别在数据集的占比超过 80%.

Waymo 3D 数据库由自动驾驶公司 Waymo 公布,整个数据集包含 1 150 个场景,每个场景均包含雷达点云数据和二维数据同步采集,整个数据库包含约 1 200 万个三维包围框和二维包围框.根据其数据划分,共分为 1 000 个场景为训练集、150 个场景为测试集.

本文将在这 2 个公开数据库上完成模型验证,并选择同样使用点云数据作为唯一输入的相关深度学习检测方法作为对比方法.由于许多检测模型并未同时给出在 2 个数据集上的验证结果,因此分别选取 5 个方法进行模型对比实验.在 KITTI 数据库上,本文选取的 5 个对比方法包括 SECOND 方法^[20]、PointPillars 方法^[21]、Fast Point-RCNN 方法^[22]、Part-A² 方法^[23]和 PV-RCNN 方法^[16].在 Waymo 数据库上,本文选取的 5 个对比方法包括 PointPillars 方法^[21]、MVF 方法^[24]、Pallar-OD 方法^[25]、PV-RCNN 方法^[16]、CenterPoint-Voxel 方法^[26].所有对比方法均为目前精度较好的方法.

在实验环境中,本文所有实验均在内存 125 GB、处理器为 Intel Core i9-9940X 的计算服务器中完成.该服务器包含 1 块型号为 NVIDIA RTX2080 12GB 的 GPU.

3.2 评价指标函数

本文按照二维目标检测算法评价精度的方法多类均值精度 (mean average precision, mAP),将其应用在评价三维目标检测算法. mAP 方法为多类均值精度,对于单类均值精度 (average precision, AP),其计算为 precision 查准率和 recall 查全率构成的曲线下面积值.因此,查准率和查全率分别为

$$\text{precision} = \frac{TP}{TP + FP} \quad (2)$$

$$\text{recall} = \frac{TP}{TP + FN} \quad (3)$$

式中: TP (true positive) 为真阳性,即对真值预测正确的结果; FP (false positive) 为假阳性,对真值预测错误的结果; FN (false negative) 为假阴性,对真值未能预测出来的结果.

相比于二维目标检测评价,在三维目标检测评价时,需要计算真值与预测包围框的重叠比 (intersection-over-union, IOU),即在三维空间中的包围框进行交并比计算.通过设置该阈值结果,来调整 AP 与 mAP 的计算.通常 IOU 阈值设置为 0.5.

3.3 KITTI 数据集模型对比与验证实验

首先,本文模型与其他 5 个对比模型在 KITTI 测试集上的精度对比结果如表 1 所示.本文在 7 518 幅测试数据集上进行了定量结果输出.

从表 1 中可知,在仅使用 LiDAR 点云数据作为输入的条件下,本文模型的总体精度高于当前主流的 5 类检测模型, mAP 值达到 94.33%,相比于最优精度模型 PV-RCNN 提升 4.15%.

表1 KITTI 3D 检测数据库模型对比实验结果(1)

Table 1 Results on the KITTI 3D dataset (1)

方法	数据输入	评价 mAP/%
SECOND ^[20]	LiDAR 点云	79.58
PointPillars ^[21]	LiDAR 点云	82.33
Fast Point-RCNN ^[22]	LiDAR 点云	85.39
Part-A ² ^[23]	LiDAR 点云	88.36
PV-RCNN ^[16]	LiDAR 点云	90.18
本文	LiDAR 点云	94.33
提升精度		+4.15

为了更加详细地对比模型精度与模型速度,本文在汽车和自行车2个类别上进行了AP与模型检测速度的对比实验,并选用数据库提供的不同难度的数据进行了模型对比实验,如表2所示。

从表2中可知,本文模型方法在三维目标检测

表2 KITTI 3D 检测数据库模型对比实验结果(2)

Table 2 Results on the KITTI 3D dataset (2)

方法	数据输入	单张处理 时间/s	汽车类样本			自行车类样本		
			简单	中等	困难	简单	中等	困难
SECOND ^[20]	LiDAR 点云	6.0	86.55	77.42	68.38	66.58	52.74	49.24
PointPillars ^[21]	LiDAR 点云	4.3	88.32	79.19	70.35	68.55	54.09	50.18
Fast Point-RCNN ^[22]	LiDAR 点云	4.5	89.72	81.84	74.28	70.31	53.28	50.33
Part-A ² ^[23]	LiDAR 点云	4.0	90.05	86.48	76.53	78.33	59.35	52.24
PV-RCNN ^[16]	LiDAR 点云	3.5	93.07	90.59	81.47	80.17	66.41	60.18
本文	LiDAR 点云	2.0	97.31	91.26	82.28	81.55	66.84	60.77
提升		+1.5	+4.24	+0.67	+0.19	+1.38	+0.43	+0.59

3.4 Waymo 数据集模型对比与验证实验

首先验证本文模型与对比方法在总体4个类别(车辆、行人、骑行者和指示牌)上的mAP。本文依据数据集提供的150个测试场景,分帧得到测试集数据库。本文模型与对比方法在Waymo测试数据库所有类别上的定量精度结果如表3所示。

从表3中可知,在仍然仅使用LiDAR点云数据为唯一输入条件下,本文模型的总体mAP为58.9%,相比于其他模型,本文模型精度高于其他对比方法,总体提升0.3%。由于Waymo数据集的场景复杂度较高,其中包含雨、雪等恶劣天气条件,造成整体三维目标检测精度相比于KITTI数据集较低。因此,在后续的研究中,仍然可以在模型上进行改进,以应对恶劣天气条件的三维目标检测任务。为了进一步对比模型检测效果,仍然在Waymo数据库上选取了2个类别(车辆和行人)进行AP精度与

的精度与速度方面均优于当前主流的检测模型,同时,本文模型更容易应对场景更为简单的目标检测任务。对于汽车类别,本文模型在3个不同的测试难度(简单、中等和困难难度)上,相比于当前最优检测模型分别提升4.24%、0.67%和0.19%的精度。同时,在运行速度上也取得一定的改进,单张图像处理时间提升1s。对于尺寸更小的自行车类别,本文模型在3个不同测试难度上也取得一定的提升,总体模型在简单、中等和困难难度的提升精度分别为1.38%、0.43%和0.59%。但是,模型总体精度提升有限,特别是在中等和困难场景提升较小。主要原因在于自行车类相比于汽车类,其图占比更小,总体尺寸较小,且存在类别不平衡问题。因此,在后续的小尺寸三维目标研究中仍需对模型进行进一步改进。

检测速度的对比实验,如表4所示。

表3 Waymo 检测数据库模型对比实验结果(1)

Table 3 Results on the Waymo dataset (1)

方法	数据输入	评价 mAP/%
PointPillars ^[21]	LiDAR 点云	41.5
MVF ^[24]	LiDAR 点云	47.3
Pallar-OD ^[25]	LiDAR 点云	56.4
PV-RCNN ^[16]	LiDAR 点云	57.9
CenterPoint-Voxel ^[26]	LiDAR 点云	58.6
本文	LiDAR 点云	58.9
提升精度		+0.3

从表4的结果可知,在车辆和行人2个类目标的检测精度上,相比于主流检测框架,本文模型方法的总体精度更高,并且在速度上具有一定的优势。对于车辆类别,相比于5个对比方法,本文模型精度

表4 Waymo 检测数据库模型对比实验结果(2)

Table 4 Results on the Waymo dataset (2)

难度	方法	单张处理时间/s	车辆评价 AP/%	行人评价 AP/%
Level 1	PointPillars ^[21]	6.0	57.4	59.7
	MVF ^[24]	8.0	62.7	67.4
	Pallar-OD ^[25]	7.0	66.5	70.3
	PV-RCNN ^[16]	4.7	76.3	65.7
	CenterPoint-Voxel ^[26]	5.7	78.1	79.7
	本文	4.4	79.3	80.1
	提升	+0.3	+1.2	+0.4

提升1.2%。可见,在更为复杂和规模更大的数据集上,对于尺寸较大车辆目标,本文模型仍然具有一定的鲁棒性。同时,在模型检测速度上,也取得0.3s的速度提升,可见模型融合并未带来运行速度降低,并且特征提取的并行方式,也带来计算速度的提升。对于行人类别,其依然具有小尺寸目标的特点,总体图占比较小,且行为变化更加多样,在恶劣天气环境下更难以有效检测。在检测精度上,本文对行人类别实现0.4%的精度提升。

通过在2个数据库上的检测结果,本文模型总体检测精度具有一定的优势,并且在数量占比较大的类别上取得更好的检测结果。对于数量占比较小和尺寸较小的物体类别,本文模型可以保持一定的精度,但是仍然进一步提升的空间。因此,定量实验结果证明本文检测模型的鲁棒性、精度与速度的平衡性。

3.5 模型消融实验

为了进一步验证本文模型4个改进方法的效果,探究所使用的不同改进方案对检测精度的提升效果,决定进行模型消融实验。由于本文提出多个针对检测模型的改进方案,通过在数据集上的结果显示,已证明精度提升效果。但是,无法确认不同改进方案对模型精度的提升效果。因此,利用控制变量方式,改进方案逐步增加的验证方法,探究改进方法对模型精度的提升效果。

本文重要的改进方法为:全局特征抽取模型、局部特征抽取模型、数据增广与多任务损失函数。由于全局特征抽取模型是本文主干网络模型,因此以该模型作为基础,逐步将其他3个改进方案放入到模型中,进行检测精度验证,进而完成整个消融实验过程。本文使用mAP指标作为精度衡量标准,同时,使用KITTI 3D数据库的测试集作为消融实验的验证数据集。

对于这4种改进方法对检测精度的影响效果,

具体结果如表5所示。

表5 改进方法的消融实验

Table 5 Results for the ablation experiment

改进方案	本文模型			
全局特征抽取模型?	✓	✓	✓	✓
局部特征抽取模型?		✓	✓	✓
数据增广?			✓	✓
多任务损失函数?				✓
评价 mAP/%	91.58	93.37	94.05	94.33
提升		+1.79	+0.68	+0.28

从表5中可以看出,增加局部特征抽取模型、数据增广方案与多任务损失函数均可以对模型精度进行有效提升。根据实验结果可知,融合全局特征抽取模型与局部特征抽取模型,更有助于对于三维点云目标的检测效果,其提升精度为1.79%;使用数据增广与多任务损失函数对于模型精度提升效果有限,分别提升0.68%与0.28%。通过消融实验,可以看出数据增广与损失函数修改对于三维目标检测方案提升效果有限,因此在之后的探究中,可以进行进一步的拓展与改进。目前,更好地建立并抽取三维目标特征,仍然是未来提升模型检测精度的重要研究方向。

4 结论

针对自动驾驶场景的三维目标检测任务,本文提出基于深度神经网络的三维目标检测算法,以三维点云数据作为研究对象,构造全局特征与局部点云特征相结合的特征提取网络,并利用候选框区域生成方式,完成对目标语义类别和位置定位的任务。通过在KITTI 3D数据库和Waymo数据库上的验证实验,可以得到如下结论。

1) 与主流的基于深度神经网络的三维目标检测框架相比,本文模型在检测精度与检测速度上均

具有一定的优势.

2) 在复杂、恶劣的天气环境中,对于三维目标检测任务,相比于其他方法,本文模型仍然具有一定的精度和速度优势.

3) 对于三维点云数据,构造一个将全局特征与局部特征相结合的特征提取网络,更能有效地提升检测模型的分类与定位能力.

当然,本文仍然存在一定的问题,面对物体尺寸较小的类别时,模型的精度呈现下降.同时,本文模型在检测速度上仍有一定的提升空间,当前模型检测速度仍未能达到可以进行实时检测的目标.因此,本文将在未来对小尺寸物体的三维实时检测任务进行更加深入的研究.

参考文献:

- [1] 李宇杰, 李焯鹏, 张为公. 基于视觉的三维目标检测算法研究综述[J]. 计算机工程与应用, 2020, 56(1): 11-24.
- LI Y J, LI X P, ZHANG W G. Survey on vision-based 3D object detection methods[J]. Computer Engineering and Applications, 2020, 56(1): 11-24. (in Chinese)
- [2] LIANG M, YANG B, WANG S L, et al. Deep continuous fusion for multi-sensor 3D object detection[C]//European Conference on Computer Vision. Berlin: Springer, 2018: 641-656. (2018).
- [3] 王刚, 王沛. 基于深度学习的三维目标检测方法研究[J]. 计算机应用与软件, 2020, 12: 164-168.
- WANG G, WANG P. 3D object detection method based on deep learning[J]. Computer Applications and Software, 2020, 12: 164-168. (in Chinese)
- [4] LIU Z, TANG H T, LIN Y J, et al. Point-Voxel CNN for efficient 3D deep learning[C]//Conference and Workshop on Neural Information Processing Systems. New York: ACM, 2019: 1-11.
- [5] REN S Q, HE K M, ROSS G, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 11-37.
- [6] JOSEPH R, SANTOSH D, ROSS G, et al. You only look once: unified, real-time object detection [C] // IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2016: 779-788.
- [7] LI P L, CHEN X Z, SHEN S J. Stereo R-CNN based 3D object detection for autonomous driving [C] // IEEE International Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2019: 7636-7644.
- [8] LI Y Y, BU R, SUN M C, et al. PointCNN: convolution

on X-transformed points [C] // Conference and Workshop on Neural Information Processing Systems. New York: ACM, 2018: 828-838.

- [9] TEKIN B, SINHA S, FUA P. Real-time seamless single shot 6D object pose prediction [C] // IEEE International Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 292-301.
- [10] QI CR, SU H, MO KCa, et al. PointNet: deep learning on point sets for 3D classification and segmentation [C] // IEEE International Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2017: 77-85.
- [11] JARITZ M, GU JY, SU H. Multi-view pointnet for 3D scene understanding [C] // IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE, 2019: 3995-4003.
- [12] QI CR, CHEN XL, LI O, et al. ImVoteNet: boosting 3D object detection in point clouds with image votes [C] // IEEE International Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2020: 4403-4412.
- [13] ZHOU Y, TUZEI O. VoxelNet: end-to-end learning for point cloud based 3D object detection [C] // IEEE International Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 4490-4499.
- [14] CHEN X Z, MA H M, WANG J, et al. Multi-view 3D object detection network for autonomous driving [C] // IEEE International Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2017: 6526-6534.
- [15] GUO Y, WANG H Y, HU Q Y, et al. Deep learning for 3D point clouds: a survey [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence. Piscataway: IEEE, 2020: 1-27.
- [16] SHI S H, GUO C X, LI J, et al. PV-RCNN: Point-Voxel feature set abstraction for 3D object detection [C] // IEEE International Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2020: 10526-10535.
- [17] QI C R, LI Y, SU H, et al. PointNet ++: deep hierarchical feature learning on point sets in a metric space [C] // Conference and Workshop on Neural Information Processing Systems. New York: ACM, 2017: 5105-5114.
- [18] ANDREAS G, PHILIP L, RAQUEL U. Are we ready for autonomous driving? the kitti vision benchmark suite [C] // IEEE International Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2012: 3354-3361.
- [19] SUN P, HENRIK K, XERXES D, et al. Scalability in perception for autonomous driving: Waymo open dataset [C] // IEEE International Conference on Computer Vision

- and Pattern Recognition. Piscataway: IEEE, 2020: 2443-2451.
- [20] YAN Y, MAO YX, LI B. Second: sparsely embedded convolutional detection [J]. *Sensors*, 2018, 18(10): 3337.
- [21] LANG AH, VORA S, HOLGER C, et al. Pointpillars: fast encoders for object detection from point clouds[C]// IEEE International Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2019: 12689-12697.
- [22] CHEN YL, LIU S, SHEN XY, et al. Fast point R-CNN [C]// IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE, 2019: 9774-9783.
- [23] LEHNER J, ANDREAS M, ADLER T, et al. Patch refinement—localized 3D object detection [C]// IEEE International Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2019: 3812-3822.
- [24] ZHOU Y, SUN P, ZHANG Y, et al. End-to-end multi-view fusion for 3D object detection in LiDAR point clouds [C]// Conference on Robot Learning. Piscataway: IEEE, 2019: 923-932.
- [25] WANG Y, FATHI A, KUNDU Ab, et al. Pillar-based object detection for autonomous driving [C]// European Conference on Computer Vision. Berlin: Springer, 2020: 18-34.
- [26] YIN TW, ZHOU XY, PHILIPP K. Center-based 3D object detection and tracking [C]// IEEE International Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2021: 11784-11793.

(责任编辑 杨开英)

(上接第 588 页)

- [23] 陈登峰, 耿建勤, 张温, 等. 无尾扑翼无人机动力学建模及仿真[J]. *计算机测量与控制*, 2020, 28(6): 202-206.
- CHEN D F, GENG J Q, ZHANG W, et al. Dynamic modeling and simulation of a tailless flapping wing robot [J]. *Computer Measurement & Control*, 2020, 28(6): 202-206. (in Chinese)
- [24] SCHONBERGER J L, FRAHM J M. Structure-from-motion revisited [C]// IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2016: 4104-4113.
- [25] LOW D G. Distinctive image features from scale-invariant keypoints[J]. *International Journal of Computer Vision*, 2004, 60(2): 91-110.
- [26] FISCHLER M A, BOLLES R C. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography [J]. *Communications of the ACM*, 1981, 24(6): 381-395.
- [27] TRIGGS B, MCLAUCHLAN P F, HARTLEY R I, et al. Bundle adjustment—a modern synthesis [C]// International Workshop on Vision Algorithms: Theory & Practice. Berlin: Springer-Verlag, 1999: 298-372.
- [28] GOESELE M, SNAVELY N, CURLESS B, et al. Multi-view stereo for community photo collections [C]// IEEE 11th International Conference on Computer Vision. Piscataway: IEEE, 2007: 1-8.
- [29] CARR J C, BEATSON R K, CHERRIE J B, et al. Reconstruction and representation of 3D objects with radial basis functions [C]// Proceedings of ACM SIGGRAPH. New York: ACM, 2001: 67-76.
- [30] WAECHTER M, MOEHRLE N, GOESELE M. Let there be color! large-scale texturing of 3D reconstructions [C]// European Conference on Computer Vision. Berlin: Springer International Publishing, 2014: 836-850.
- [31] 王伟, 段虎荣, 张德成, 等. 一种求解三维激光扫描后不规则体的体积及表面积方法[J]. *测绘与空间地理信息*, 2019, 42(8): 172-175.
- WANG W, DUAN H R, ZHANG D C, et al. A method for solving the volume and surface area of irregular body after three-dimensional laser scanning [J]. *Geomatics & Spatial Information Technology*, 2019, 42(8): 172-175. (in Chinese)

(责任编辑 张蕾)