

DOI:10.19651/j.cnki.emt.2005560

# 基于聚类方法的自动驾驶场景下的三维目标检测

毕雪婷<sup>1,2,3</sup> 刘小军<sup>1,2</sup> 邵文远<sup>1,2,3</sup>(1.中国科学院空天信息创新研究院 北京 100094; 2.中国科学院电磁辐射与探测技术重点实验室 北京 100190;  
3.中国科学院大学电子电气与通信工程学院 北京 100049)

**摘要:** KITTI 数据集是自动驾驶场景下融合多个传感器的大型数据集,它的数据采集平台包括 2 个灰度摄像头、2 个彩色摄像头、1 个 velodyne 64 线激光雷达、4 个光学镜头和 1 个 GPS 导航系统。KITTI 3D Object Detection Evaluation 可为各种 3D 目标检测算法验证准确性和有效性,是自动驾驶领域最重要的数据集。此文的重点是 KITTI 数据集的数据重构和数据清洗:首先对 KITTI 数据集中的每一帧激光雷达数据使用 RANSAC 算法进行地面去除,并用 DBSCAN 算法对地面上的目标进行聚类,然后根据标签文件对聚类后的目标使用最近邻搜索赋予每个目标类别标签以完成数据重构,基于此,再对数据进行重采样以均衡类别完成数据清洗。针对重构和清洗后的 KITTI 数据使用 PointNet 算法完成分类任务,准确率高达 95.13%,最后完成了 KITTI 数据集上 3D 目标检测与评估的总体框架。结果表明重构和清洗后的新数据集质量高,分类算法鲁棒性强,3D 目标检测过程清晰完整。

**关键词:** KITTI;自动驾驶;聚类;点云分类;3D 目标检测评估

中图分类号: TN919.8 文献标识码: A 国家标准学科分类代码: 510.4050

## 3D object detection in automatic driving scene clustering

Bi Xueting<sup>1,2,3</sup> Liu Xiaojun<sup>1,2</sup> Shao Wenyuan<sup>1,2,3</sup>

(1.Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China;

2.Key Laboratory of Electromagnetic Radiation and Sensing Technology, Chinese Academy of Sciences, Beijing 100190, China;

3.School of Electronic, Electrical and Communication Engineering, University of Chinese Academy of Sciences, Beijing 100049, China)

**Abstract:** KITTI is a large data set fused with multiple sensors in automatic driving scene, its data acquisition platform includes two gray-scale cameras, two color cameras, a velodyne 64 line lidar, four optical lenses and a GPS navigation system, KITTI 3D Object Detection Evaluation can verify the accuracy and effectiveness of various 3D object detection algorithms. It is the most important data set in the field of autonomous driving. The focus of this article is the data reconstruction and data cleaning of the KITTI data set; first, use the RANSAC algorithm to remove the ground from each frame of lidar data in the KITTI data set, and use the DBSCAN algorithm to cluster the targets on the ground, and then according to the label the file uses the nearest neighbor search to assign tags to each target category to complete the data reconstruction. Based on this, the data is resampled to balance the categories to complete the data cleaning. For the reconstructed and cleaned KITTI data, the PointNet algorithm is used to complete the classification task, and the accuracy rate is as high as 95.13%. Finally, the overall framework of 3D target detection and evaluation on the KITTI data set is completed. The results show that the quality of the reconstructed and cleaned new data set is high, the classification algorithm is robust, and the 3D target detection process is clear and complete.

**Keywords:** KITTI; autonomous driving; clustering; point cloud classification; 3D object detection evaluation

## 0 引言

点云可以表达物体在空间中的位置信息,且表达形式简单。点云可以来源于激光雷达和深度相机等设备,一些算法比如 slam 也可以产生一些点云。近两年,点云

在机器人、自动驾驶、人脸识别和土木工程等领域应用广泛。

PointNet<sup>[1]</sup>是一个用于点云分类和分割的网络,也是第一个直接在点云上使用深度学习的工作。它利用多层感知机把点云从低维映射到高维,然后使用 max pooling 最

收稿日期:2020-12-19

• 103 •

大池化方法来提取点云的全局特征。Max pool 是对称函数,可以适应点云的无序性和旋转不变性,但其模型过于简单,缺乏对多层信息的表达。PointNet++<sup>[2]</sup>是 PointNet 的升级版,它能够提取多层次,提高了语义信息的丰富性和模型的表现力。它首先使用 FPS (farthest point sampling)方法对原始点云进行降采样,降采样后,以每个点为中心进行最近邻搜索建立 groups,对每个分组使用 PointNet 进行特征提取。在插值阶段,恢复点云原始的密集分布并使用双线性差值方法得到每个点在当前层的多层级特征表示。这两个网络是现在绝大部分网络的骨干网络。SECOND<sup>[3-5]</sup>是基于点云的三维目标检测网络,为了解决传统 3D CNN 计算资源需求大和内存消耗严重的问题,提出了一种新型的 3D 稀疏卷积加快模型的训练和推理速度。PointPillars 提出了一种柱状特征提取方法,将三维的点云转成二维伪图像,然后在其上使用二维目标检测网络,省去 3D CNN 操作很大程度的加快了算法的运行速度<sup>[6]</sup>。PointRCNN<sup>[7]</sup>是第一个两阶段的 3D 目标检测网络,在 KITTI 3D 目标检测排行榜上曾获得第一名。TANet<sup>[8]</sup>在 PointPillars 的基础上,增加了体素、空间和点的三元注意力机制,效果有大幅度提升。Pointpainting 是目前融合图像和点云的三维目标检测方法中效果最好的算法<sup>[9-10]</sup>,它将图像中检测到的目标通过坐标变换映射到点云场景中,作为下一阶段三维目标检测网络的初始解,然后应用 PointPillars、PointRCNN 等网络进行细化。Part-A<sup>2</sup> Net<sup>[11]</sup>使用了部分感知和部分聚合的方法构建三维目标检测网络。PV-RCNN<sup>[12]</sup>第一次将基于体素的特征提取更加有效和基于点的方法更加灵活和更大的感受野的优势结合起来,也曾在 KITTI 排行榜上排名第一并遥遥领先。随后 deformable PV-RCNN<sup>[13]</sup>对这项工作又进行了改进,使用可变形卷积和门控机制提高了网络的鲁棒性。

这项工作中主要有两个亮点:

- 1) 使用聚类算法和标定参数将原始 KITTI 数据集进行重构,并进行数据清洗以达到类别均衡。
- 2) 完成深度学习在重构数据上的分类任务。实现基于 KITTI 三维目标检测的评估流程。

## 1 重构 KITTI 分类数据集

KITTI 数据集来源于真实三维空间中的道路场景。为了在其上应用分类网络,本文将其重构为类似 modelnet40 数据集的 numpy 格式的点云分类数据集。

### 1.1 平面估计与地面去除

采用 RANdom SAMpling Consensus 算法对每帧点云进行平面拟合。主要思想遵循随机性和假设性,即假设随机抽样出的数据是正确的,然后用所选数据拟合模型来解决问题,如图 1 所示。

假设内点在整个数据集所占的比例是  $t$ , 它的计算方式可以表示为:

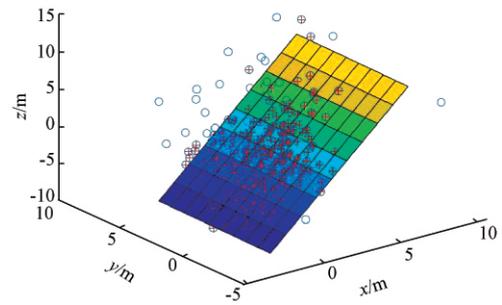


图 1 RANSAC 算法拟合示意图

$$t = \frac{n_{inliers}}{n_{inliers} + n_{outliers}} \quad (1)$$

使用该模型解决问题需要的点数是  $n$ , 即样本数,  $n$  的选取是根据所解决问题的具体情况而定的。在解决点云平面拟合问题中,所需的点数为 3。  $k$  是迭代次数,每迭代一次随机选取的  $n$  个点的数值来计算模型参数。  $P$  是所希望的在得到的参数下求得正确解的概率。  $n$  个点都属于内点的概率是  $t^n$ , 那么  $n$  个点中至少有一个点是外点的概率是  $1-t^n$ ,  $(1-t^n)^k$  表示  $k$  次随机抽样中没有一个都是内点,这种错误的概率计算如下:

$$P = 1 - (1 - t^n)^k \quad (2)$$

迭代次数由我们要求的正确概率  $P$  计算:

$$k = \frac{\log(1 - P)}{\log(1 - t^n)} \quad (3)$$

### 1.2 目标提取与聚类

KITTI 数据集中主要有 3 种目标类别:行人、汽车和自行车。针对每帧点云中聚类后<sup>[14]</sup>的目标,以这一帧 Ground Truth 信息与基准,采用最近邻搜索算法给每个聚类后的点云标定类别,不是这 3 种类别以外的点被归为第 4 类。

### 1.3 数据清洗

KITTI 数据集中每个类别的样本数量非常不均匀。当用于分类的样本的每一类的数量差异很大时,会影响深度学习的训练过程,使模型更多的去输出数量最多的那一类,从而使网络学习到没有价值的参数模型。

在模型训练之前,通过平移、绕  $Z$  轴进行随机旋转、建立真值样本库随机采样等数据增广方法增加部分类别的样本数量。如图 2 所示,图 2(a)和(b)分别代表进行数据清洗前后各别所占的比重,可以看到经过数据增广后,类别的分布已经非常均匀。

## 2 搭建点云分类网络

如图 3 所示,在模型训练阶段,网络的输入是一帧点云中  $n$  个点的三维坐标,即一个  $n \times 3$  的矩阵。然后,通过 MLP 多层感知机将每个点的维度从 3 维映射到 64 维,使矩阵成为  $n \times 64$ 。在此之前,需要通过一个用来学习标准朝向的  $3 \times 3$  的小网络。接下来,使用  $n \times 64$  矩阵作为输入,通过一个可以在特征空间中学习标准方向的  $64 \times 64$  网

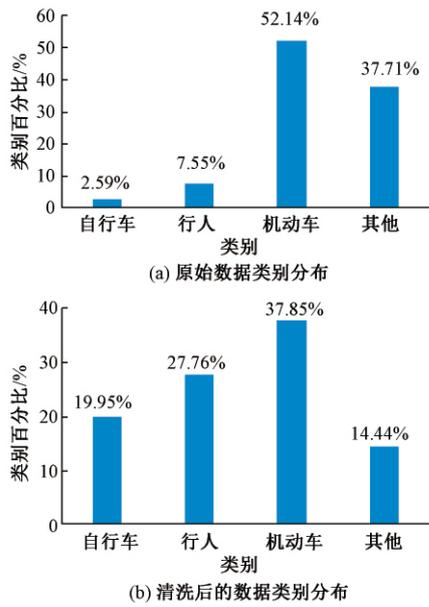


图 2 数据清洗前后的分布

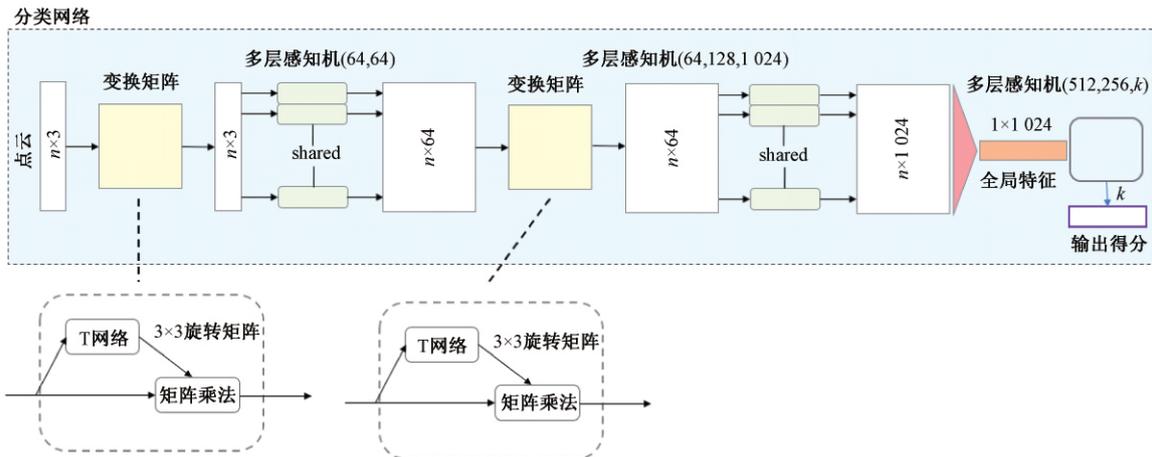


图 3 点云分类网络结构

用 DBSCAN 漫水算法进行聚类, 根据标签文件对行人、汽车和自行车进行分类标定, 最后得到以 numpy 格式存储的重构后 KITTI 分类数据集。如图 4 所示, 图 4(b) 给出了



(a) 相机采集的真实道路场景

在原始点云场景中进行地面去除和聚类的可视化效果。在这里使用 Open3D 库进行可视化。

如图 4(a) 为摄像机采集到的原始图像, 图 4(b) 为原始



(b) 点云场景下去除地面的聚类效果

图 4 真实场景下的聚类效果

点云场景中去除地面并对目标进行聚类后的可视化效果。在这一阶段, 虽然没有使用 GPU 资源, 但也取得了很好的效果。如图 5 所示, 图 5(a) 和 (b) 分别对提取后的行人和汽车单个点云目标进行可视化。在本例中, 上述给出了行

人和车辆。可以看到这项工作的提取效果是非常不错的, 即使用肉眼观察, 对行人和汽车识别率也是很高的。

### 3.2 3D 目标检测模型训练和评估

如表 1 所示, 显示了点云分类网络下对重构的 KITTI



图 5 数据提取可视化

数据集进行训练和测试的结果<sup>[15]</sup>。从表 1 中可以看出,这 4 类分类都获得了相当高的精确率和召回率,即使是行人、自行车等识别率较低类别,也得到了很好的检测结果,并对 4 种类别的分类结果求取均值得到更客观的 AP 值和 mAP 值,同时预测准确率也高达 95.13%。

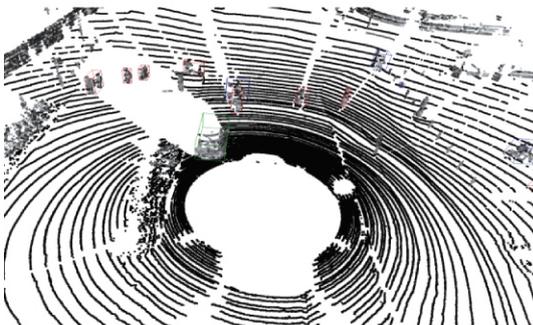
表 1 搭建的点云分类网络对重构后的 numpy 格式的 KITTI 数据集的分类结果

类别	精确度	召回率	平均精确度	F 值	支持集
自行车	0.97	0.97	0.94	0.97	1 600
行人	0.90	0.80	0.72	0.84	1 157
机动车	0.96	0.97	0.93	0.96	2 225
其他	0.95	0.98	0.93	0.96	3 034
平均精确度均值			0.88		
准确率			0.95		

如图 6 所示,图 6(a)是 KITTI 数据集中的图像,图 6(b)是与(a)相对应的同时同地的用 velodyne 64 线激光雷达采集的点云场景的,同时对基于点云的目标检测结果进行可视化。框架的不同颜色代表不同类型的对象。可以看出,即使是在角落里的行人这样的小目标,也能被很好地探测到。



(a) 视觉传感器下的真实场景



(b) 3D目标检测结果可视化

图 6 目标检测效果可视化

## 4 结 论

总的来说,这项工作完成了在 KITTI 数据集上进行三维物体检测的整体过程。不同的是,这项工作重建了 KITTI 数据集,从原始场景中提取出目标,将其分为 4 类,并将格式修改为 numpy,然后对数据进行重采样以缓解不同类别数量不均衡的问题。根据观察提取出的目标的可视化结果以及 PointNet 点云分类网络高达 95.13%的准确率可以看出,重建和数据清洗后的 KITTI 分类数据集质量很高,具有标准化意义。最后完成 KITTI 数据集下 3D 目标检测任务的评估流程。

## 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 IEEE Conference on Computer Vision and Pattern Recognition, 2017: 652-660.
- [2] QI C R, YI L, SU H, et al. Pointnet++: Deep hierarchical feature learning on point sets in a metric space[J]. ArXiv Preprint, 2017, ArXiv:1706.02413.
- [3] YAN Y, MAO Y, LI B. Second: Sparsely embedded convolutional detection[J]. Sensors, 2018, 18(10): 3337.
- [4] 闫岩. 基于深度学习的点云三维目标检测算法研究[D]. 重庆:重庆大学, 2019.
- [5] LI X, GUIVANT J, KWOK N, et al. Three-dimensional backbone network for 3D object detection in traffic scenes [J]. ArXiv Preprint, 2019, ArXiv: 1901.08373.
- [6] 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.
- [7] SHI S, WANG X, LI H. Pointcnn: 3D object proposal generation and detection from point cloud[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 770-779.
- [8] LIU Z, ZHAO X, HUANG T, et al. Tanet: Robust 3d object detection from point clouds with triple attention[C]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(7): 11677-11684.

- [9] 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.
- [10] VORA S, LANG A H, HELOU B, et al. Pointpainting: Sequential fusion for 3d object detection[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 4604-4612.
- [11] SHI S, WANG Z, WANG X, et al. Part-A<sup>2</sup> Net: 3D part-aware and aggregation neural network for object detection from point cloud[J]. ArXiv Preprint, 2019, ArXiv:1907.03670.
- [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/CVF Conference on Computer Vision and Pattern Recognition, 2020: 10529-10538.
- [13] BHATTACHARYYA P, CZARNECKI K. Deformable PV-RCNN: Improving 3D object detection with learned deformations [J]. ArXiv Preprint, 2020, ArXiv:2008.08766.
- [14] 王盛杰, 刘博, 李和平, 等. 基于面阵三维成像激光雷达的目标点云分割技术[J]. 半导体光电, 2020, 41(5): 749.
- [15] 杨永光. 基于点云的目标检测方法研究[D]. 南京: 南京邮电大学, 2020.

#### 作者简介

毕雪婷, 硕士研究生, 主要研究方向为机器学习、深度学习、三维点云处理、基于点云的分类、分割与三维目标检测。  
E-mail: bixueting18@mails.ucas.edu.cn

刘小军, 工学博士, 研究员, 主要研究方向为信号与信息处理、超宽带雷达技术。  
E-mail: lxjdr@mail.ie.ac.cn

邵文远, 博士研究生, 主要研究方向为雷达信号处理、毫米波与太赫兹成像。  
E-mail: shaowenyuan18@mails.ucas.edu.cn