International Journal of Computer Vision (2023) 131:531–551
https://doi.org/10.1007/s11263-022-01710-9

Check for updates

PV-RCNN++: Point-Voxel Feature Set Abstraction With Local Vector Representation for 3D Object Detection

Shaoshuai Shi[1,2] · Li Jiang[1,2] · Jiajun Deng[3] · Zhe Wang[4] · Chaoxu Guo[4] · Jianping Shi[4] · Xiaogang Wang[1] · Hongsheng Li[1]

Received: 3 April 2022 / Accepted: 31 October 2022 / Published online: 24 November 2022

Abstract

3D object detection is receiving increasing attention from both industry and academia thanks to its wide applications in various fields. In this paper, we propose Point-Voxel Region-based Convolution Neural Networks (PV-RCNNs) for 3D object detection on point clouds. First, we propose a novel 3D detector, PV-RCNN, which boosts the 3D detection performance by deeply integrating the feature learning of both point-based set abstraction and voxel-based sparse convolution through two novel steps, *i.e.*, the voxel-to-keypoint scene encoding and the keypoint-to-grid RoI feature abstraction. Second, we propose an advanced framework, PV-RCNN++, for more efficient and accurate 3D object detection. It consists of two major improvements: sectorized proposal-centric sampling for efficiently producing more representative keypoints, and VectorPool aggregation for better aggregating local point features with much less resource consumption. With these two strategies, our PV-RCNN++ is about 3× faster than PV-RCNN, while also achieving better performance. The experiments demonstrate that our proposed PV-RCNN++ framework achieves state-of-the-art 3D detection performance on the large-scale and highly-competitive Waymo Open Dataset with 10 FPS inference speed on the detection range of $150m \times 150m$.

Keywords 3D object Detection · Point clouds · LiDAR · Autonomous driving · Sparse convolution

1 Introduction

3D object detection on point clouds aims to localize and recognize 3D objects from a set of 3D points, which is a fundamental task of 3D scene understanding and is widely-adopted in lots of real-world applications like autonomous driving, intelligent traffic system and robotics. Compared to 2D detection methods on images (Girshick, 2015; Ren et al., 2015; Liu et al., 2016; Redmon et al., 2016; Lin et al., 2017, 2018), the sparsity and irregularity of point clouds make it challenging to directly apply 2D detection techniques to 3D detection on point clouds.

To tackle these challenges, most of existing 3D detection methods (Chen et al., 2017; Zhou and Tuzel, 2018; Yang et al., 2018b; Lang et al., 2019; Yan et al., 2018) transform the points into regular voxels that can be processed with conventional 2D/3D convolutional neural networks and well-studied 2D detection heads. But the voxelization operation inevitably brings quantization errors, thus degrading their localization

Communicated by Jean-Sébastien Franco.

✉ Shaoshuai Shi
sshi@mpi-inf.mpg.de

Li Jiang
lijiang@mpi-inf.mpg.de

Jiajun Deng
jiajun.deng@sydney.edu.au

Zhe Wang
wangzhe@sensetime.com

Chaoxu Guo
cxguo@sensetime.com

Jianping Shi
shijianping@sensetime.com

Xiaogang Wang
xgwang@ee.cuhk.com.hk

Hongsheng Li
hsli@ee.cuhk.com.hk

1 The Chinese University of Hong Kong, Hong Kong SAR, China

2 Max Planck Institute for Informatics, Saarbrücken, Germany

3 The University of Sydney, Sydney, Australia

4 SenseTime Research, Shanghai, China