

文章编号: 2095-2163(2023)12-0019-04

中图分类号: TP391.4

文献标志码: A

# 基于改进 PV-RCNN++算法的三维点云聚焦式特征研究

段界余, 宁媛, 黎玉成

(贵州大学 电气工程学院, 贵阳 550025)

**摘要:** 为增强 RoI 网格局部特征的表征能力、加强细节特征的表达效果, 以进一步提高点云语义分割精度, 针对 PV-RCNN++网络采用的 RoI 网格池化模块进行了研究。在 PV-RCNN++网络中, RoI 网格池化模块只是将网格点周围的体素特征进行简单的空间位置排序, 导致局部特征表达效果欠佳。为加强 RoI 网格池化模块对局部特征的表征能力, 引入 CBAM 注意力机制, 从通道和空间两个作用域出发, 一方面处理特征集通道的分配关系, 另一方面可使神经网络更加关注特征集中对分类起决定性作用的体素区域, 以强化重要信息在网络的有效传递并提高点云语义分割结果的鲁棒性。对自动驾驶领域公开数据集 Kitti 的语义分割实验表明, 所提出的改进 PV-RCNN++的聚焦式特征的算法训练出的模型, 较基准模型提升效果显著, 有效增强了 RoI 网格池化模块对局部特征的表征能力, 强化了细节特征的表达效果, 提高了点云语义分割精度。

**关键词:** 激光雷达; 深度学习; 卷积神经网络; PV-RCNN++; CBAM 注意力机制

## Research on 3D point cloud focusing features based on improved PV-RCNN++ algorithm

DUAN Jieyu, NING Yuan, LI Yucheng

(College of Electrical Engineering, Guizhou University, Guiyang 550025, China)

**Abstract:** In order to enhance the representation ability of local features of RoI grid, enhance the expression effect of detailed features, and further improve the accuracy of cloud semantic segmentation, the RoI grid pooling module used in PV-RCNN++ network was studied. In PV-RCNN++ network, RoI grid pooling module only performs simple spatial position ordering of voxel features around grid points, resulting in poor local feature expression effect. In order to enhance the representation ability of RoI grid pooling module for local features, CBAM attention mechanism was introduced. The CBAM attention mechanism starts from two domains of channel and space. On the one hand, it deals with the distribution relationship of feature set channels; on the other hand, it enables the neural network to pay more attention to the voxel region where feature set plays a decisive role in classification, so as to enhance the effective transmission of important information in the network and improve the robustness of the semantic segmentation results of the high point cloud. Semantic segmentation experiments on Kitti, an open data set in the field of automatic driving, show that the model trained by the proposed improved PV-RCNN++ focusing feature algorithm has a significant improvement effect compared with the benchmark model, which effectively enhances the representation ability of RoI grid pool module for local features, strengthens the expression effect of detailed features, and improves the semantic segmentation accuracy of point cloud.

**Key words:** Lidar; deep learning; CNN; PV-RCNN++; CBAM-attention mechanism

## 0 引言

近年来,深度传感器技术快速发展,其中激光雷达设备获取 3D 点云数据的质量和效率不断提升,被广泛应用于自动驾驶、智能交通系统、机器人等领域和行业。迄今为止,以卷积神经网络(Convolutional Neural Network, CNN)为代表的深度

学习技术已在图像分割、分类识别等多类视觉任务中大获成功<sup>[1]</sup>。与 2D 图像不同,3D 点云不规则、非均匀和无序性的数据特点,使快速、精确的点云语义分割处理成为一项极具挑战性的任务。设计针对大规模、复杂场景 3D 点云的语义分割网络模型,成为当前研究重点<sup>[2]</sup>。

目前,为了从稀疏和不规则的点云中学习识别

**基金项目:** 国家自然科学基金(61663005)。

**作者简介:** 段界余(1997-),男,硕士研究生,主要研究方向:计算机视觉、三维点云处理。

**通讯作者:** 宁媛(1968-),女,硕士,教授,硕士生导师,主要研究方向:检测技术与自动化装置、人工智能。Email: ee.yning@gzu.edu.cn

收稿日期: 2022-11-29

特征,然后对学习到的特征进行点云语义分割,出现了两种主流的语义分割方法,主要分为直接点云语义分割和间接点云语义分割。

间接的点云语义分割方法,主要是基于 2D 投影的多视图卷积神经网络 (Multi-View CNN, MVCNN)<sup>[3]</sup> 和基于 3D 体素的 VoxNet<sup>[4]</sup>。Su 等<sup>[3]</sup> 提出,先获取点云数据在不同视角下的 2D 投影图像,再利用传统 CNN 提取投影图像特征,并通过池化层、全连接层聚合不同投影图像特征,得到最终语义分割结果。与多视角融合图像特征相比,体积离散意义下的点云体素化处理较好地维护了其数据内在几何关系。以 Maturana 等<sup>[4]</sup> 提出的 VoxNet 为代表的 3D CNN,可以充分学习 3D 空间上下文信息用于场景分割。

直接点云语义分割方法,对原始的点云数据直接进行卷积神经网络操作,通过学习全局特征来进行点级的语义分割。Qi 等<sup>[5]</sup> 构建的 PointNet 成功解决了点云无序性、置换不变性等问题,但每个点的操作过于独立,不能很好的捕捉局部特征。针对这一问题,Qi 等<sup>[6]</sup> 在 PointNet 的基础上结合分层结构、最远距离点采样,提出通过融合不同层次区域特征,来捕获不同空间感知域的点云上下文信息的 PointNet++。

为实现更高效、更准确的三维目标检测,Shaoshuai 等<sup>[7]</sup> 先将输入的原始点通过 3D 体素 CNN 划分为分辨率为  $L \times H \times W$  的小体素,再利用一系列的  $3 \times 3 \times 3$  的三维稀疏卷积,将点云转化为下采样大小为  $1 \times 2 \times 4 \times 8 \times$  的特征集,之后再转换为 2D 鸟瞰图以生成 3D 建议框。但是,体素量过大会导致内存不足,实际应用中计算效率低下。为了解决这个问题,Shaoshuai 等<sup>[8]</sup> 将不同层级的体素编码为少量的关键点特征,然后将关键点特征聚合到 ROI 网格中,用于进一步的置信度预测和包围盒回归。在此基础上又提出 PV-RCNN++ 框架,引入两个新模块来提高 PV-RCNN 框架的准确性和效率。一个是以建议框为中心的扇区并行计算的关键点采样算法,另一个是 RoI 网格池化聚合模块,用于从大规模的点云中聚合有效的局部特征。

为更好地学习点云的细粒度局部特征,在 PV-RCNN++ 网络的基础上,本文将 CBAM 注意力机制嵌入到 RoI 网格池化模块,通过调试和优化模型参数,可构建出改进的 PV-RCNN++ 算法,不仅增强了信息在网络间的有效传递,也得到了细粒度更饱满的局部特征。

## 1 PV-RCNN++ 目标检测算法

PV-RCNN++ 目标检测算法主要由两部分组成:

第一部分为体素到关键点场景特征编码。先通过带有扇区并行计算结构的,并以建议框为中心扩展半径区域内的最远距离点采样 (Sectorized Proposal-Centric Farthest Point Sampling, SPC-FPS) 算法,从原始输入点云  $P$  中采样少量关键点  $K = \{p_1, \dots, p_n\}$ ,其中  $n$  为超参数(本文  $n = 2048$ );再通过体素集抽象模块 (Voxel Set Abstract, VSA) 编码,从 3D 体素 CNN 到关键点多尺度语义特征,即

$$f_i^{(pv)} = [f_i^{(pv_1)}, f_i^{(pv_2)}, f_i^{(pv_3)}, f_i^{(pv_4)}], i = 1, \dots, n \quad (1)$$

其中,  $pv_1, pv_2, pv_3, pv_4$  分别表示不同层级的神经网络。

第二部分为关键点到网格点的 RoI 网格池化模块。首先在每个由 2D 鸟瞰特征图生成的 3D 建议框中,统一采样  $6 \times 6 \times 6$  个网格点,记为  $G = \{g_1, \dots, g_{216}\}$ ;再在网格点周围设置  $3 \times 3 \times 3$  的小体素;采用 3 个最近关键点特征插值的方法,可以得到体素局部特征。计算公式为

$$f_{i_x, i_y, i_z}^{(v)} = \frac{\sum_{j=1}^3 (\omega_{\sigma(j)} f_{\sigma(j)})}{\sum_{j=1}^3 \omega_{\sigma(j)}} \quad (2)$$

其中,  $\omega_{\sigma(j)} = (\|p_{\sigma(j)} - v_{i_x, i_y, i_z}\|)^{-1}$ ,  $v_{i_x, i_y, i_z}$  为该体素的中心位置。考虑到不同局部体素的特征存在较大的差异性,对这些体素局部使用不同的局部内核权重进行编码,公式为

$$\hat{U}(i_x, i_y, i_z) = E(\wedge_{r_{i_x, i_y, i_z}} f_{i_x, i_y, i_z}^{(v)}) \times W_v \quad (3)$$

其中,  $\hat{U}(i_x, i_y, i_z) \in R^{C_{r2}}$ ,  $\hat{r}_{i_x, i_y, i_z} \in R^{(3 \times 3)}$  为 3 个最邻近关键点的相对位置,  $E(\cdot)$  是融合相对位置和特征的拼接操作,  $W_v \in R^{(9+C_{r1}) \times C_{r2}}$  是对局部体素进行特征编码的可学习核权值矩阵,不同位置的局部体素的核权值不同。最后网格点周围的这些局部体素特征将通过一个多层感知机 (MLP) 进行聚合,得到  $C_{out}$  特征通道的网格点特征用于后续处理。

## 2 CBAM 注意力机制

为了更好地学习点云的细粒度局部特征,在 PV-RCNN++ 目标检测框架中引入注意力机制,通过忽略无关信息、关注重点信息,进一步提升语义分

割精度。不同于 SENet<sup>[9]</sup> (Squeeze-and-Excitation Networks) 和 ECANet<sup>[10]</sup> (Efficient Channel Attention Networks), CBAM<sup>[11]</sup> (Convolutional Block Attention Module) 注意力机制从通道和空间两个维度出发, 实现从通道到空间的顺序注意力结构。通道注意力机制模块用于处理特征集通道的分配关系, 空间注意力机制模块可以使神经网络更加关注点云中点对分类起决定作用的体素区域而忽略无关紧要的区域。

CBAM 中的通道注意力机制模块首先通过平均池化和最大池化操作聚合特征集的空间信息, 生成两个富含上下文信息的特征矢量  $F_{avg}^c$  和  $F_{max}^c$ , 再经过一个共享的多层感知机 (MLP), 得到通道注意力映射  $M_c \in R^{C \times 1 \times 1 \times 1}$ , 公式为

$$M_c(F) = \sigma(MLP(F_{avg}^c) + MLP(F_{max}^c)) \quad (4)$$

其中,  $\sigma$  表示 sigmoid 函数。

CBAM 中的空间注意力机制模块首先沿通道轴通过平均池化和最大池化操作聚合特征集的通道信息, 生成两个 3D 的特征集  $F_{avg}^s \in R^{1 \times L \times W \times H}$  和  $F_{max}^s \in R^{1 \times L \times W \times H}$  (本文  $L, W, H$  均设置为 3), 再经过一个  $3 \times 3 \times 3$  的卷积核操作, 生成三维空间注意力映射  $M_s(F) \in R^{1 \times L \times W \times H}$ , 公式为

$$M_s(F) = \sigma(f^{3 \times 3 \times 3}([F_{avg}^s; F_{max}^s])) \quad (5)$$

其中,  $\sigma$  表示 sigmoid 函数。

### 3 改进的 PV-RCNN++ 目标检测算法

改进的 PV-RCNN++ 目标检测算法如图 1 所示。

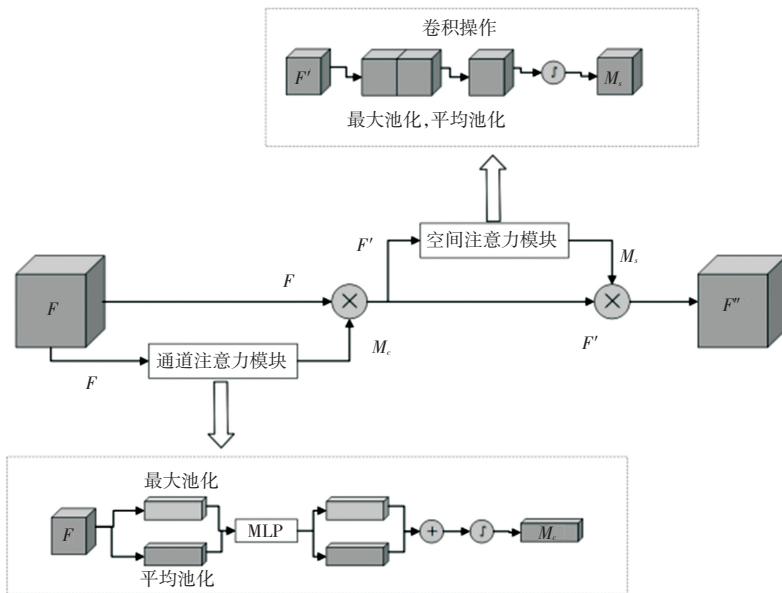


图 1 嵌入 CBAM 注意力机制的改进 PV-RCNN++

Fig. 1 Improved PV-RCNN++ with CBAM

改进的方法是在网格点周围的局部体素特征被编码后, 加入 CBAM 注意力机制模块。在前向网络训练过程中, 添加 CBAM 模块, 会在通道和空间双维度上得到局部体素的聚焦式特征, 相较于其他注意力机制只在单通道或单空间上聚焦, 此改进方提高了网格点周围局部体素的细节特征提取效率和点云的分割精度。

## 4 实验与分析

### 4.1 数据集准备和平台搭建

Kitti<sup>[12]</sup> 数据集是目前自动驾驶视觉算法领域应用最广泛的数据集, 为度量和测试算法性能提供了可靠的评估基准。实验所采用的数据集样本数量共有 15 000 个, 其中包括 7 482 个训练样本数据和

7 518 个测试样本数据。

实验使用的开发平台为 Ubuntu18.04, 内核是 Linux, 平台配置参数见表 1。

表 1 实验平台参数

Table 1 Experimental platform parameters

软硬件	版本或型号
操作系统	Ubuntu18.04
CPU	i7-12800HX
GPU	GTX 3080Ti
内存	32 GB
CUDA	11.4.0
Torch	1.10.0
Torchvision	0.11.0
Python	3.7

### 4.2 实验结果分析

训练得到 C-PV-RCNN++ 的模型后, 在 Kitti 测

试集上对模型进行了评估,其中汽车的  $IoU$  阈值设置为 0.7,以 40 个召回位置的平均精度作为评估结果,不同模型对不同难度目标的  $mAP$  (%) 对比结果详见表 2。其中,简单、中等、困难分别对应目标完全可见、部分遮挡、大部分遮挡。

表 2 不同模型对不同难度汽车的准确率对比

Table 2 Comparison of accuracy of different models for cars of different difficulty %

模型	简单	中等	困难
PointPillars <sup>[13]</sup>	82.58	74.31	68.99
PointRCNN <sup>[14]</sup>	86.96	75.64	70.70
PV-RCNN <sup>[7]</sup>	90.25	81.43	76.82
PV-RCNN++ <sup>[8]</sup>	90.14	81.88	77.15
C-PV-RCNN++	90.15	81.98	77.40

实验结果表明,嵌入 CBAM 注意力机制的 PV-RCNN++ 算法,面对中等、困难的汽车的识别精度得到了有效的提升,证明本文改进的关于局部体素的聚焦式特征,在识别中等、困难目标时提升效果显著。

## 5 结束语

自动驾驶是未来道路交通发展的必然趋势,运用更高精度的目标检测算法,将大大保证汽车道路行驶的安全性,减少交通事故的发生频次。改进的 PV-RCNN++ 检测算法通过在 RoI 网格池化模块嵌入 CBAM 注意力机制,确保网格点周围的重要信息可以有效传递,模型检测精度得到了有效的提升。

## 参考文献

[1] SHELHAMER E, LONG J, DARRELL T. Fully convolutional networks for semantic segmentation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 39(4): 640-651.

[2] GUO Y L, 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, 2021, 43(12): 4338-4364.

[3] SU H, MAJI S, KALOGERAKIS E, et al. Multi-view convolutional

neural networks for 3D shape recognition [C] // Proceedings of the IEEE International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2015: 945-953.

[4] MATURANA D, SCHERER S. VoxNet: a 3D convolutional neural network for real-time object recognition [C] // Proceedings of the IEEE/RISJ International Conference on Intelligent Robots and Systems. Los Alamitos: IEEE Computer Society Press, 2015: 922-928.

[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. Los Alamitos: IEEE Computer Society Press, 2017: 77-85.

[6] QI C R, YI L, SU H, et al. PointNet++: deep hierarchical feature learning on point sets in a metric space [C] // Proceedings of the 24<sup>th</sup> Annual Conference on Neural Information Processing Systems. Cambridge: MIT Press, 2017: 5099-5108.

[7] 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.

[8] SHI S, JIANG L, DENG J, et al. PV-RCNN++: Point-voxel feature set abstraction with local vector representation for 3D object detection [J]. International Journal of Computer Vision, 2023, 131(2): 531-551.

[9] HU J, SHEN L, SUN G. Squeeze-and-excitation networks [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 7132-7141.

[10] WANG Q, WU B, ZHU P, et al. ECA-Net: Efficient channel attention for deep convolutional neural networks [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 11534-11542.

[11] WOO S, PARK J, LEE J Y, et al. Cbam: Convolutional block attention module [C] // Proceedings of the European Conference on Computer Vision (ECCV). 2018: 3-19.

[12] GEIGER A, LENZ P, URTASUN R. Are we ready for autonomous driving? the kitti vision benchmark suite [C] // Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2012: 3354-3361.

[13] 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.

[14] SHI S, WANG X, LI H. Pointrcnn: 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.