

结合局部增强与反向残差的点云语义分割网络

温智成,王 蕾,冯锦梁,叶森辉
(东华理工大学信息工程学院,江西 南昌 330013)

摘要:激光点云是3D传感器的输出,且对它的语义分割任务是理解真实世界的基础。基于图卷积的点云分割网络在许多场景下都展现了优异的性能。然而,现有的图卷积方法存在部分问题:点云局部表示的能力未得到加强,忽略了全局几何信息,并且聚合操作只保留局部最大响应值信息,而次最大值信息丢失。为了处理这些问题,本文提出GRes-Net网络。利用局部几何增强(Local Geometry Augment, LGA)模块,使网络对Z轴具有旋转不变性,以便加强点云局部信息表示;采用全局几何特征(Global Geometry Feature, GGF)模块,计算局部与全局的球体体积比,将其与坐标特征X进行连接,使全局几何信息特征得以保留;通过多个对称聚合操作将局部信息多方面地保留;网络中每层都使用残差操作,将上一层信息传递到下一层,以及利用反向残差模块(Reversed Residual MLP, RevResMLP)挖掘更深层次的语义信息。本文在S3DIS数据集上进行语义场景分割实验,验证网络分割的性能。实验结果表明该方法在分割精度上达到61%,相比于基准网络DGCNN提高14%,有效地提高了模型性能。

关键词:点云语义分割;动态图卷积;局部几何增强;全局几何特征;反向残差

中图分类号:TP391;TN249 **文献标识码:**A **DOI:**10.3969/j.issn.1001-5078.2023.07.022

Combining local augmentation and reverse residual for point cloud semantic segmentation network

WEN Zhi-cheng, WANG Lei, FENG Jin-liang, YE Sen-hui

(School of Information Engineering, East China University of Technology, Nanchang 330013, China)

Abstract: The laser point cloud is the output of a 3D sensor, and its semantic segmentation task is fundamental to understanding the real world. The point cloud segmentation network based on graph convolution shows excellent performance in many scenarios. However, existing graph convolution methods are partially problematic: the local representation of the point cloud is not augmented, global geometric information is ignored, and the aggregation operation only keeps the local maximum response value information, while the sub-maximum information is lost. In order to address these problems, the GRes-Net network is proposed in this paper. The Local Geometry Augment (LGA) module is used to make the network rotationally invariant to the Z-axis, so as to strengthen the local information representation of the point cloud. The Global Geometry Feature (GGF) module is used to calculate the volume ratio between local and global spheres, and connect it with coordinate feature X, so that the global geometry information feature can be retained. Multiple symmetric aggregation operations are used to preserve the local information in many ways. Residual operations are used in each layer of the network to transfer information from one layer to the next as well as to mine

基金项目:江西省核地学数据科学与系统工程技术研究中心基金项目(No. JELRGT202202);江西省放射性地学大数据技术工程实验室开放基金项目(No. JELRGT202103)资助。

作者简介:温智成(1998-),男,硕士研究生,主要从事三维点云分割方面的研究。E-mail:1326855218@qq.com

通讯作者:王 蕾(1979-),女,博士,副教授,硕士生导师,主要从事计算机视觉方面的研究。E-mail:wlei598@163.com

收稿日期:2022-12-06;**修订日期:**2022-12-29

deeper semantic information using the Reversed Residual MLP (RevResMLP) module. In this paper, semantic scene segmentation is conducted on S3DIS data set to verify the performance of network segmentation. The experimental results show that the segmentation accuracy of this method is up to 61 % ,which is 14 % higher than that of the benchmark network DGCNN, and the performance of the model is effectively improved.

Keywords: point cloud semantic segmentation; dynamic graph convolution; local geometric augment; global geometric feature; reverted residual network

1 引言

随着 3D 传感器的快速发展,激光点云的获取越来越便利。利用激光点云感知真实世界,在家庭智能机器人、自动驾驶与 AR/VR 等领域发挥着重要的作用。然而,点云与图像不同,图像为规则化的数据,而点云为非结构化且无序的数据,这使得基于网格的卷积神经网络(CNN)难以应用。

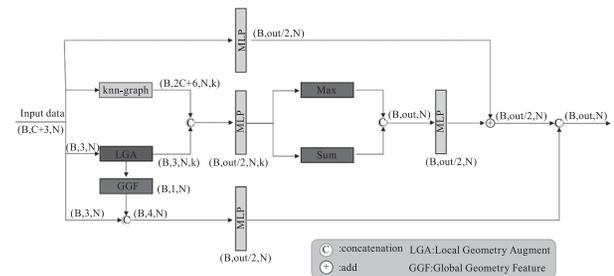
在最近几年,许多基于深度学习的方法被提出用于处理这一问题。这些方法大致可以分为 3 类:基于投影化的方法、基于体素化的方法和基于点的方法。然而,基于投影的方法^[1-3]会导致大量的信息丢失,基于体素化的方法^[4-5]会消耗大量的计算资源,而基于点的方法^[6-20],无需对点云数据进行预处理操作,从而不会存在信息的丢失以及计算成本开销大的问题。但是在基于点的方法中,基于图卷积的方法^[16-22]对上下文特征有很强的信息挖掘能力。它们根据点特征之间的相似性将点构造为图数据,并将二维卷积应用到激光点云上,其比基于逐点的方法取得了更高的精度。然而,该方法对点云局部表示能力未进行加强以及忽略了全局几何信息,并且使用最大池化操作聚合局部信息,导致仅保留了局部内的最大响应值信息。

针对上述问题,本文提出 GRes-Net 网络,由 Graph Residual(GRes)(如图 1 所示)模块完成对点云局部的加强与全局几何信息的保留,GRes 模块包含了 Local Geometry Augment(LGA)与 Global Geometry Feature(GGF)模块。LGA 模块利用局部点与质心间的相对球坐标,使网络对点云的 Z 轴有旋转不变性,从而加强点云局部表示;GGF 模块获取点云局部与全局的体积比,将其与点云坐标特征相连接,以保留全局几何信息;通过多个聚合操作以保留除最大响应值以外的特征信息;添加反向残差模块,挖掘更深层次的特征信息。

本文贡献如下:

- 使用 LGA 模块加强点云局部的特征表示;

- 利用 GGF 模块保留全局几何信息;
- 多个聚合操作将局部最大响应值与其他信息保留;
- 通过反向残差模块挖掘更深层次的语义信息;
- GRes-Net 网络在 S3DIS 数据集上的结果,超越了多个基线。



B 为批量大小, N 为点数, C 为点的 RGB 等特征通道大小, 3 为点的坐标特征通道大小, k 为邻近点数, out 为输出通道大小

图 1 Graph Residual 模块

Fig. 1 Graph Residual module

2 基于深度学习的点云语义分割网络

虽然深度学习在 NLP 与图像领域取得了重大的进展,但是在激光点云领域还未得到很好的探索。本文将回顾此前的研究,分别为基于投影的方法、基于体素的方法、基于点的方法与基于图的方法。

基于投影与体素的方法。这两种方法需将点云转换为规则化的数据,对其进行卷积操作。前后两者分别将点云转换为图像与体素。MVCNN^[1]提取每个视图的特征,并利用最大池化将它们聚合到全局描述符中。GVCNN^[2]将每个视图提取到的视觉描述子进行建模分组。VoxNet^[4]将点云转换为体素,利用三维卷积提取特征。OctNet^[5]采用不平衡的八叉树将空间分区,以减少空体素的数量,提高内存与计算效率。

基于点的方法。PointNet^[6]是第一个将深度学习用于点云处理的网络,其提出以逐点的方式应用 MLP,随后利用最大池化聚合全局特征,但是未对点

云局部信息进行挖掘。PointNet++^[7]采用 Encoder-Decoder 的结构,关注点云局部的特征信息。KP-Conv^[8]利用核函数对局部每个点计算权重矩阵。Point2Sequence^[10]通过自注意力机制从不同的局部区域学习信息。PCT^[11]在全局上利用自注意力机制使每个点之间都相互关注彼此信息。

基于图的方法。将点作为图的节点,并根据点之间的特征关系建立边。DGCNN^[16]在特征空间上获取每个点相邻点,通过 MLP 与聚合操作进行预测。3D-GCN^[17]提出了可变形核,通过在多个尺度上进行聚合以提高模型性能。GAC^[18]为不同的图节点分配不同的权重。

综上所述,直接对点云进行操作解决了需对点云进行预处理的问题。然而,在上述基于图的方法中,点云局部表示的能力未得到加强,忽略了全局几何信息,以及聚合操作仅保留局部最大响应值信息。为此,本文提出了 GRes-Net 网络,可以有效地加强局部表示与保留全局几何信息,提高语义分割精度。

3 本文方法

本文所提出 GRes-Net 网络(如图 2 所示)受到动态图卷积 DGCNN 启发,采用 Encoder-Decoder 架构。DGCNN 通过将点云坐标、RGB 等信息作为特征,并利用该特征对每个点进行 k 近邻搜索,随后计算每个点与其相邻点特征的差值,将中心点特征与该差值进行拼接,从而构建了新的特征图,如图 3 所示。利用 MLP 与最大池化学习新的特征图局部特征信息。

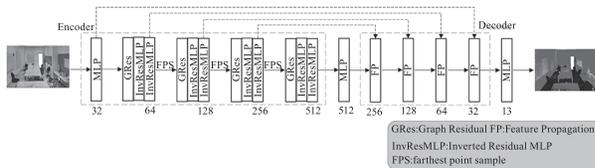


图 2 GRes-Net 网络结构

Fig.2 GRes-Net network architecture

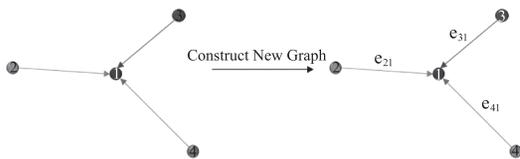


图 3 构建新的特征图

Fig.3 Constructa new feature graph

该网络由 3 个关键模块组成。第一个是 Local

Geometry Augment (LGA) 模块,其可加强点云局部信息表示。第二个是 Global Geometry Feature (GGF) 模块,目的是保留全局几何信息。第三个是 Reverted Residual MLP (RevResMLP) 模块,为了挖掘点云深层次的语义信息。

3.1 LGA 模块

在点云大场景语义分割中,同类别物体的不同朝向,会导致该物体的语义分割结果出现差别。这会在一定程度上影响整个场景语义分割的性能。为了增强对方向的鲁棒性,且受到 SCF-Net^[15]的启发,本文利用球坐标加强局部几何表示。通过计算局部内每个点与该局部质心点之间的相对角度,其对 Z 轴具有旋转不变性,如图 4 所示。其中(a)计算 p_i 与 p_i^k 之间的角度 θ_i^k, ϕ_i^k ; (b)计算 p_i 与 p_i^c 之间的角度 θ_i^c, ϕ_i^c ; (c)、(d)、(e)计算 θ_i^k 与 θ_i^c , 以及 ϕ_i^k 与 ϕ_i^c 之间的差值。

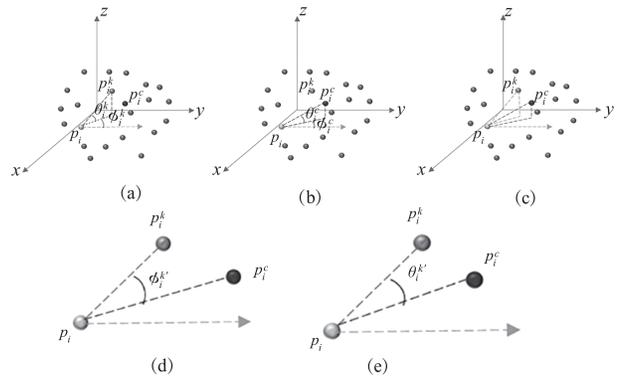


图 4 局部几何加强模块(LGA)

Fig.4 Local Geometry Augment module(LGA)

计算距离与角度。通过将笛卡尔坐标系的点坐标应用至式(1)、式(2)、式(3)以计算距离 (dis_i^k) 与角度 (θ_i^k, ϕ_i^k)。 dis_i^k 为点 p_i 与点 p_i^k 之间的距离, θ_i^k, ϕ_i^k 为点 p_i 与点 p_i^k 之间的角度。定义: $\{p_i^1, p_i^2, \dots, p_i^k\}$ 为点 p_i 的 k 个近邻点, i 表示某个局部区域。

$$dis_i^k = \sqrt{x_i^k{}^2 + y_i^k{}^2 + z_i^k{}^2} \quad (1)$$

$$\theta_i^k = \arctan\left(\frac{z_i^k}{\sqrt{x_i^k{}^2 + y_i^k{}^2}}\right) \quad (2)$$

$$\phi_i^k = \arctan\left(\frac{y_i^k}{x_i^k}\right) \quad (3)$$

其中, (x_i^k, y_i^k, z_i^k) 是点 p_i 在笛卡尔坐标系中的坐标。

计算质心的角度。通过局部区域的点坐标计算出质心 p_i^c 。本文定义从 p_i 到 p_i^c 的方向为局部方向。

此种定义法存在有 2 种益处:1)质心能够反映出局部区域的整体特性;2)通过使用质心,能够减少下采样所产生的随机性。根据式(2)与式(3)计算出质心的角度 θ_i^c, ϕ_i^c 。

计算相对角度。将区域内每个点的与质心点的角度相减,可以获得相对角度 $\theta_i^{k'}, \phi_i^{k'}$ 。

$$\theta_i^{k'} = \theta_i^k - \theta_i^c \quad (4)$$

$$\phi_i^{k'} = \phi_i^k - \phi_i^c \quad (5)$$

3.2 GGF 模块

点云密度作为点云的内在几何属性特征,反映的是点云内部间隔的变化。密度的剧烈变化,通常表示该点所属物体的变化。这种特性可以在点云特征中额外添加几何信息,以提高分割的性能。

将 $r_i = \frac{v_k}{v_g}$ 定义为局部与全局的球体体积比,如图 5 所示。其中, v_k 表示点 p_i 的 k 个近邻点所组成局部的球体体积, v_g 是输入点云的球体体积。将在局部内距点 p_i 最远点的距离作为局部球体半径,而计算全局球体半径,需要计算输入点云的质心点,将在输入点云内距质心点最远点的距离作为全局球体半径。通过式(7)可以获得局部球体体积与全局球体体积。

$$V = \frac{4}{3}\pi R^3 \quad (6)$$

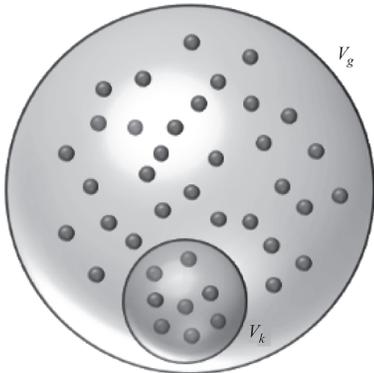


图 5 全局几何特征模块(GGF)

Fig. 5 Global Geometry Feature module(GGF)

3.3 RevResMLP 模块

受到 PointNeXt^[12] 启发,本文利用 RevResMLP 模块挖掘深层次的语义信息。传统残差结构是先将输入特征降维再升维,而反向残差结构将输入特征升维再降维。通过将输入特征升维,学习更丰富的特征信息。该模块添加在 GRes 模块后,以挖掘深层次语义信息。

4 实验

4.1 数据集

本文在 S3DIS 数据集上验证 GRes-Net 网络的有效性。S3DIS 是一个高质量的室内场景数据集,大小约为 14.2 GB,它包含了 6 个不同室内区域,总共占地约 6020 m²,包含 272 个房间。该数据集有 13 个类别(墙,地板,椅子等)。以 X 轴与 Y 轴所组成的平面为基准,随机性地选取 1 m × 1 m 的区域。

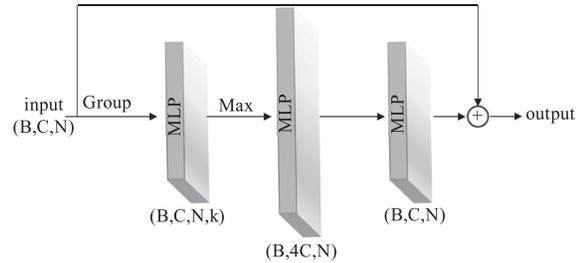


图 6 反向残差模块(RevResMLP)

Fig. 6 Reversed Residual MLP module(RevResMLP)

从该区域内随机性地选取 4096 个点为输入数据。在一个装有 NVIDIA GeForce RTX 3080GPU 的服务器上进行实验,并且使用 pytorch 框架。

4.2 实验参数设置

该网络总共训练 100 轮,本文使用初始学习率为 0.01 的 AdamW 优化器,以及利用余弦退火算法使学习率逐渐衰减。邻居点数设置为 16,即 $k = 16$ 。训练时,batch size 设置为 16;测试时,batch size 也同样设置为 16。

4.3 实验结果

通过表 1 可知,GRes-Net 网络在 S3DIS Area5 上分割指标中,mIOU 为 61%,OA 为 86.7%,mACC 为 68.7%。与此前的方法相比,特别是与图卷积算法 DGCNN^[16] 和 DeepGCN^[20] 相比,mIOU 指标分别提高 14% 和 8.5%,验证了本文方法的有效性。此外,为了直观地呈现分割结果,本文将 S3DIS Area5 数据集内的部分房间可视化,如图 7 所示,图 7 从左至右每列分别为:输入点云数据、真实语义、DGCNN 预测语义、GRes-Net 预测语义。

4.4 消融实验

在本节中,为了验证网络中各个模块的有效性,本文在 S3DIS Area5 上进行消融实验。如表 2 所示,移除所有模块,分割性能下降 7.6%。第一行至第三行说明,添加 LGA 与 GGF 模块可以提高模型的性能。添加 RevResMLP 模块与未添加该模块的

结果相比,即第三行与第四行相比,分割性能上升 5.4%,表明该模块能够显著地提高模型性能。

表 1 不同方法在 S3DIS Area5 上的结果

Tab.1 Results by different methods on S3DIS Area5

Method	mIOU/%	OA/%	mAcc/%
PointNet ^[6]	41.1	/	49.0
DGCNN ^[16]	47.0	/	/
SegCloud ^[19]	48.9	/	57.4
DeepGCN ^[20]	52.5	/	/
PointCNN ^[13]	57.3	85.9	64.9
PointWeb ^[9]	60.3	87.0	66.6
GRes-Net	61.0	86.7	68.7

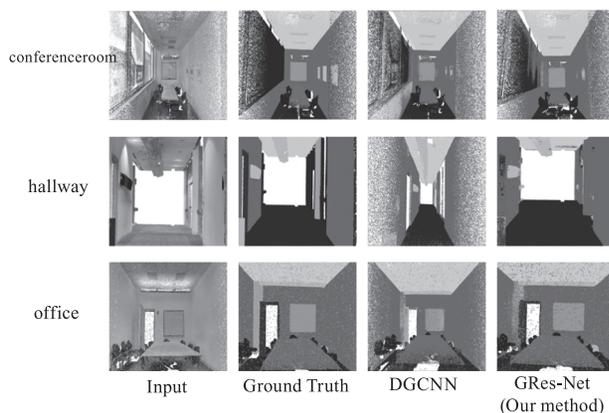


图 7 在 S3DIS Area5 上语义分割可视化

Fig.7 Visualization of semantic segmentation on S3DIS Area5

表 2 在 S3DIS Area5 上消融实验结果

Tab.2 Results of ablation experiments on S3DIS Area5

LGA	GGF	RevResMLP	结果 mIOU/%
/	/	/	53.4
√	/	/	54.9
√	√	/	55.6
√	√	√	61.0

5 总结

针对现有图卷积点云分割网络的局部表示能力弱、全局几何信息缺失与局部聚合仅保留最大响应值等问题,本文提出 GRes-Net 网络,通过引进 LGA 模块与 GGF 模块,在图卷积过程中对点云局部信息加强与保留全局几何信息,采用多个聚合操作保留点云局部内最大响应值与其他信息,同时利用 RevResMLP 模块挖掘深层次的语义信息。

在 S3DIS Area5 上评估所提出网络,该网络在

分割精度上达到 61%,与 DGCNN、DeepGCN 以及 PointWeb 网络相比,语义分割性能分别提高了 14%、8.5%与 0.7%。针对 LGA 模块、GGF 模块与 RevResMLP 模块的消融实验表明,这三个模块能够有效地提高模型的分割性能。

虽然该网络提高了分割性能,但在点云分割任务中,收集并标注大规模点云的成本相当昂贵,所以使用无监督学习来处理未标注的点云将成为下一步工作的重点。

参考文献:

- [1] Su H, Maji S, Kalogerakis E, et al. Multi-view convolutional neural networks for 3D shape recognition [C]// IEEE International Conference on Computer Vision. IEEE, 2015.
- [2] Feng Y, Zhang Z, Zhao X, et al. Gvcnn: group-view convolutional neural networks for 3d shape recognition [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 264 - 272.
- [3] Hamdi A, Giancola S, Ghanem B. Mvtn: multi-view transformation network for 3D shape recognition [C]// Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021: 1 - 11.
- [4] Maturana D, Scherer S. Voxnet: a 3D convolutional neural network for real-time object recognition [C]// 2015 IEEE/RSJ international conference on intelligent robots and systems (IROS). IEEE, 2015: 922 - 928.
- [5] Riegler, G., A. O. Ulusoy, A. Geiger. OctNet: learning deep 3D representations at high resolutions [C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) IEEE, 2017.
- [6] 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.
- [7] Qi C R, Yi L, Su H, et al. Pointnet ++: deep hierarchical feature learning on point sets in a metric space [J]. Advances in Neural Information Processing Systems, 2017.
- [8] Thomas H, Qi C R, Deschaud J E, et al. Kpconv: flexible and deformable convolution for point clouds [C]// Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 6411 - 6420.
- [9] Zhao H, Jiang L, Fu C W, et al. Pointweb: enhancing local neighborhood features for point cloud processing [C]//

- Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 5565 – 5573.
- [10] Liu X, Han Z, Liu Y S, et al. Point2sequence: Learning the shape representation of 3d point clouds with an attention-based sequence to sequence network [C]//Proceedings of the AAAI Conference on Artificial Intelligence, 2019, 33(1): 8778 – 8785.
- [11] Guo M H, Cai J X, Liu Z N, et al. Pct: point cloud transformer [J]. Computational Visual Media, 2021, 7(2): 187 – 199.
- [12] Qian G, Li Y, Peng H, et al. PointNeXt: revisiting pointNet + with improved training and scaling strategies [J]. arXiv Preprint arXiv:2206.04670, 2022.
- [13] Li Y, Bu R, Sun M, et al. Pointenn: convolution on x-transformed points [C]//32nd Conference on Neural Information Processing Systems (Near Ips 2018), Montréal, Canada, 2018.
- [14] Zhao H, Jiang L, Jia J, et al. Point transformer [C]//Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021: 16259 – 16268.
- [15] Fan S, Dong Q, Zhu F, et al. SCF-net: learning spatial contextual features for large-scale point cloud segmentation [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 14504 – 14513.
- [16] Wang Y, Sun Y, Liu Z, et al. Dynamic graph cnn for learning on point clouds [J]. Acm Transactions On Graphics (tog), 2019, 38(5): 1 – 12.
- [17] Lin Z H, Huang S Y, Wang Y C F. Convolution in the cloud: Learning deformable kernels in 3d graph convolution networks for point cloud analysis [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 1800 – 1809.
- [18] Wang L, Huang Y, Hou Y, et al. Graph attention convolution for point cloud semantic segmentation [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 10296 – 10305.
- [19] Tchapmi L, Choy C, Armeni I, et al. Segcloud: semantic segmentation of 3d point clouds [C]//2017 International Conference on 3D Vision (3DV). IEEE, 2017: 537 – 547.
- [20] Li G, Müller M, Qian G, et al. Deepgcn: making gcns go as deep as cnns [J]. arXiv Preprint arXiv:1904.03751.
- [21] Liu Youqun, Ao Jianfeng. 3D point cloud semantic segmentation based on multi-information deep learning [J]. Laser & Infrared, 2021, 51(5): 675 – 680. (in Chinese) 刘友群, 敖建锋. 基于多信息深度学习的 3D 点云语义分割 [J]. 激光与红外, 2021, 51(5): 675 – 680.
- [22] Ao Jianfeng, Su Zekai, Liu Chuanli, et al. 3D object detection based on point cloud deep learning [J]. Laser & Infrared, 2020, 50(10): 1276 – 1282. (in Chinese) 敖建锋, 苏泽锴, 刘传立, 等. 基于点云深度学习的 3D 目标检测 [J]. 激光与红外, 2020, 50(10): 1276 – 1282.