

基于邻域信息编码的点云语义分割

安金都, 周美丽*

延安大学物理与电子信息学院, 陕西 延安

收稿日期: 2023年3月20日; 录用日期: 2023年4月20日; 发布日期: 2023年4月28日

摘要

由于点云这种数据在各个领域的广泛应用, 如何对其进行充分的利用就成了亟需解决的问题。良好的点云的语义分割结果可以为点云下游应用任务提供良好的基础。对于语义分割任务而言, 全局信息和局部信息都是必要的, 都发挥着重要的作用。基于局部特征越好最终的分割效果也会越好这样一个基本事实设计了一个名为邻域信息编码的模块, 这个模块在进行局部特征提取的时候将中心点与中心点周围的邻域点之间的方向信息充分利用。原始的输入特征被认为是空间特征和附加特征(例如: 颜色和深度)的组合, 附加特征通过共享多层感知机进行处理, 空间特征将会通过所提出的邻域信息编码模块来提取每一个中心点的邻域特征, 处理之后的两部分特征将会被拼接起来作为最终的局部输出特征。从在公开数据集SS3DIS上的实验结果来看, 所提出的模型在分割任务上有良好的表现, 它的分割结果比当前最好的基线模型高了1.8%。

关键词

点云, 语义分割, 局部特征编码, 邻域信息编码

Point Cloud Segmentation Based on Neighboring Information Encoding

Jindu An, Meili Zhou*

School of Physics and Electronic Information, Yan'an University, Yan'an Shaanxi

Received: Mar. 20th, 2023; accepted: Apr. 20th, 2023; published: Apr. 28th, 2023

Abstract

How to take full advantages of point cloud is an urgent task due to the universal application of point cloud data format. Superb semantic segmentation of point cloud can be the solid basis of subsequent task of point cloud application. For semantic segmentation task, both the global and

*通讯作者。

文章引用: 安金都, 周美丽. 基于邻域信息编码的点云语义分割[J]. 数据挖掘, 2023, 13(2): 194-201.

DOI: 10.12677/hjdm.2023.132019

local feature play significant role. Based on the basic truth that the better the local feature, the better the segmentation performance, the neighboring feature encoding module is elaborated, this module aims to make the most of direction feature between center point and its neighboring points during the local feature extraction process. Original input feature is considered as concatenation of spatial information and plus feature (e.g.: RGB and depth), the plus feature will be handled by shared Multi-Layer Perceptron, the spatial information will be sent to proposed neighboring feature encoding module to capture neighboring feature of every center point, output of two part will be concatenated as final local feature. From the perspective of result of experiment on public dataset S3DIS, the proposed model has excellent performance on feature extraction, its performance excels SOTA baseline model with 1.8% mIoU.

Keywords

Point Cloud, Semantic Segmentation, Local Feature Encoding, Neighboring Feature Encoding

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

随着自动驾驶和机器人技术的快速发展,准确、快速的感知系统也越发凸显重要。感知系统在自动驾驶以及机器人系统中是基础的一环,负责为后续的规划控制系统提供稳定可靠的感知结果,帮助其做出合理的决策。感知系统的性能直接影响自动驾驶系统或者机器人系统的安全性和功能性,因此,一个精准、快速的感知系统对于当前的自动驾驶和机器人而言是不可或缺的。自动驾驶系统和机器人系统通过布置在系统周围的传感器系统来实时地采集周围环境的数据,智能系统需要根据对当前环境的感知和理解结果做出最合理的决策,但是单纯的检测或者分类方法难以为智能系统提供对当前环境全面的感知和理解,而相比于检测,语义分割是像素级的,可以为智能系统提供一个全面、信息丰富、层次清晰且易于理解的语义环境,从而使得控制系统可以立足于这种全面、清晰的理解而做出合理的规划决策。

作为一种常用的 3D 数据的形式,点云是刻画现实世界目标物表面形状结构的点的集合,是机器认识和理解现实世界的重要基础,是一种通过激光雷达以及深度相机等主动传感器对现实世界进行扫描所获得的可以反映目标物表面特点的数据形式。相比于图像数据,点云有一些显著的特点:1) 三维表达、高密度、非结构化,点云是目标表面形状结构三维点的集合,由一个个散乱的三维点组成,每个点有准确的三维空间信息,并且其分布密度很高。点云的点与点之间没有明显的位置关系,导致对点云的组织管理、浏览查询等操作的时间度非常高;2) 点云的点具有一定的属性信息,由于可能由不同的设备进行采集,点云还具有一定的属性信息,例如激光雷达所采集的点云具有强度信息可以反映目标物的反射率等信息;3) 分布不均匀,由于目标物之间的遮挡,会导致遮挡目标难以被扫描到,从而导致被遮物体的数据缺失,形成数据空洞。另一方面由于与传感器的距离不同,点云的点的密度也有很大不同,距离传感器较近的密度较大,距离较远的密度较小。由于点云的这些特点,导致难以对其利用标准卷积进行特征提取,同时,由于不均匀的数据分布导致局部信息的提取是非常困难的事情。

点云的语义分割可以分为三种方法:基于投影的方法、基于体素的方法和基于点的方法。其主要的发展历程如图 1 所示。

基于投影的方法主要思想是在不同视角下将点云投影为 2D 的图像,然后利用成熟的图像分割方法

来对图像进行分割, 以综合多个视角下的图像分割结果来作为最终点云的分割结果。Lawin *et al.* 首先在工作中应用了这种思想[1], 他们从多个机位角度将点云投影成图片, 然后使用 FCN 网络来对图片进行分割[2]。Boulchetal 使用多机位从点云中生成深度图像[3], 然后使用 2D 图像分割网络来对图像进行分割, 对于图像的分割结果使用残差修正来进行融。

合得到最终点云的语义标签。为了保留更多的信息, Wu *et al.* 在 SqueezeNet [4] 和条件随机场的基础上提出了一个端到端的网络[5], 同时引入了 SqueezeSegV2 来解决域转换问题[6]。Molioto *et al.* 提出了 RangeNet++ 实现雷达点云的实时语义分割[7]。基于投影的方法没有充分利用到点云的几何结构信息并且这种基于数据转换的方法不可避免会带来信息损失和很高的内存和计算消耗。

基于体素的方法主要是通过对点云进行离散化操作, 将点云中的点划分到一个个体素当中, 然后使用 3D 标准卷积来进行特征的提取以及分类工作, 最终, 在同一个体素内的点被赋予相同的语义标签。Huang *et al.* 首次将点云分割成体素集合[8], 然后使用 3D 全卷积来进行体素级别的语义分割。Tchapmi *et al.* 提出了 SEG Cloud [9] 来实现对点云全局连续的分割并且使用了 3D-FCNN 来进行体素的划分[10]。Meng *et al.* 提出了基于核的方法来实现对点云局部特征的编码[11]。基于体素的方法可以处理不同尺度的点云, 这得益于全卷积点网络分级的从点云中抽取不同层次的几何关系[10]。Dai *et al.* 提出 ScanComplete 来完成 3D 补全任务和语义分割任务[12]。基于体素的方法在一定程度上保留了点云的几何结构信息, 但是其存在离散噪声和信息损失, 分辨率低有较大信息损失, 分辨率高则会导致极高的内存和计算消耗。

基于点的方法则不对原始点云进行任何转换处理, 其直接在原始的点云上进行处理。代表性的工作是 PointNet [13], 其首先提出了这种方法, 创造性地使用最大池化解决点云的无序性问题, 使用多层感知机(Multi-Layer Perceptron, MLP)来解决特征提取的问题, 使得可以在不进行数据预处理的情况下进行语义分割工作。后续的基于点的工作大多也是基于这种思想[14] [15] [16] [17], PointNet++ 在前人工作的基础上添加了用于局部信息提取的模块[14], RandLA-Net 设计了专用的局部信息编码块[15], SCF-Net 添加了坐标极化方法并且将点云的体积作为可靠的辅助信息来减弱点云的方向敏感性[16], CGA-Net 设计了用于处理边界点的模块以提升局部特征表达能力[17]。Jiang *et al.* 将 SIFT 算子迁移至点云领域应用[18], 挖掘邻域信息。Engelmann *et al.* 使用了 KNN 聚类对点云进行聚类分析[19]。Zhao *et al.* 提出了 PointWeb 来提取局部区域点对之间的关系[20]。基于点的方法不需要对点云数据进行任何的转换工作, 加速了整个处理流程, 逐渐成为点云语义分割的重要方法。

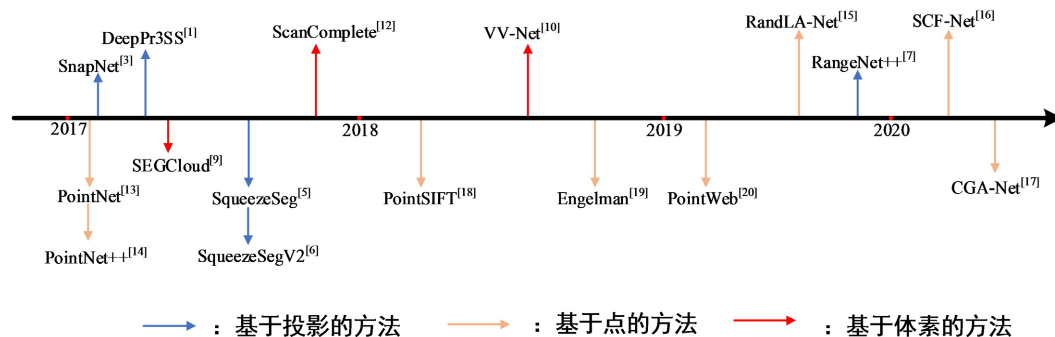


Figure 1. Chronological overview of point cloud semantic segmentation

图 1. 点云语义分割方法的发展历程

2. 方法

给定一个包含 n 个点 $p_1, p_2, \dots, p_n \in R^{3+d}$ 的点云 P , 每一个点的特征可以看作是 3 维的空间特征和 d

维的附加特征组成, 因此其特征向量由其空间坐标 (x_i, y_i, z_i) 和附加特征(例如 RGB, 深度信息等)组成。点云的标签集合表示为 L 。语义分割的目标是找到一个函数 Ω , 这个函数表示语义分割的整个过程, 包含神经网络的所有层以及所有的参数以及设置等, 以公式表示为:

$$\Omega: P^n \rightarrow L^n \tag{1}$$

2.1. 邻域信息编码

邻域信息编码模块是局部信息编码模块的核心, 邻域信息编码模块负责抽取中心点的邻域特征, 这一部分特征会和另外一部分特征拼接起来作为最终整个局部信息编码模块的输出。邻域信息编码模块主要基于点云的空间特征进行处理, 它的结构如图 2 所示, 处理过程表述如下:

1) 对于在中心点 $p_0 = (x_0, y_0, z_0)$ 周围一定范围内的每一个邻域点 $p_i = (x_i, y_i, z_i)$ 将它的坐标转换到以中心点为原点的坐标系下, 坐标转换的目的在于方便进行后续的空间信息抽取过程。公式描述为:

$$(x_{new}^i, y_{new}^i, z_{new}^i) = (x_i - x_0, y_i - y_0, z_i - z_0) \tag{2}$$

$$p_{new}^0 = (0, 0, 0) \tag{3}$$

$(x_{new}^i, y_{new}^i, z_{new}^i)$ 表示第 i 个点转换后的新坐标, p_{new}^0 表示中心点作为原点的坐标。

2) 首先, 我们需要定义一个标准方向向量 $v_s = (0, 0, 1)$ 。其次, 在完成坐标转换之后, 获取每一个邻域点在当前坐标系下的方向向量 v_i 。每一个方向向量都需要和标准向量进行计算从而获取当前邻域点和中心点之间的方向信息, 这个方向信息可以反映邻域点和中心点之间的空间位置关系。

$$v_i = (x_{new}^i, y_{new}^i, z_{new}^i) - p_{new}^0 \tag{4}$$

$$\cos \theta = v_s \cdot v_i / |v_i| \tag{5}$$

$$\langle p_0, p_i \rangle = \arccos(\cos \theta) \tag{6}$$

\cdot 表示向量之间点乘, $|v_i|$ 表示对向量进行取模运算, $\langle p_0, p_i \rangle$ 表示中心点 p_0 与 p_i 之间的夹角。夹角的集合表示为 J 。

3) 基于局部区域点特征距离越近越相似的假设, 距离特征同样是需要关注的特征。邻域信息编码模块的另一个分支负责提取每一个邻域点 p_i 与中心点 p_0 之间的距离特征 d_i 。中心点和邻域点之间的距离信息主要通过以下公式进行计算, 距离特征的集合表示为 D :

$$d_i = \sqrt{x_{new}^i + y_{new}^i + z_{new}^i} \tag{7}$$

中心点特征 f_0 , 邻域点特征 f_i 以及他们之间对应的角度信息集合 J 和距离特征 D 在特征融合后作为当前中心点的邻域特征。

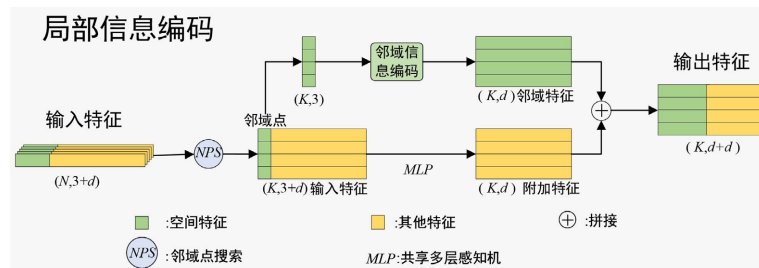


Figure 2. The architecture of neighborhood information encoding module
图 2. 邻域信息编码模块结构

2.2. 局部信息编码

本模块主要负责提取点云的局部特征, 本模块的具体细节如图 3 中所示。对于输入的特征 F , 本文将其看作是空间特征 S 和附加特征 E 的组合。经过邻域点选取(Neighboring Point Search, NPS)过程之后, 选取过程如图 4 所示。经过邻域点选取过程位于中心点周围范围 ε 内的 N 个邻域点 N_p 将会被选取出来并被送进邻域信息提取模块进行邻域信息的抽取, 得到邻域信息 N_k^i ; 附加特征则会经过共享多层感知机进行进一步的特征提取和维度变换得到变换后的特征 E_k^i 。邻域信息 N_k^i 和附加特征 E_k^i 会在通道维度上进行融合得到最终的局部特征 F_{local} 。整体过程用公式描述为:

$$F = Concat(S, E) \tag{8}$$

$$N_p = NPS(p_0, \varepsilon) \tag{9}$$

$$N_k^i = NFE(S_{N_p}) \tag{10}$$

$$E_k^i = MLP(E_{N_p}) \tag{11}$$

$$F_{local} = Concat(N_k^i, E_k^i) \tag{12}$$

所得到的局部特征 F_{local} 将会被送进下一个模块中与全局特征融合进行最终的语义分割过程。整个网络的结构如图 5 所示。

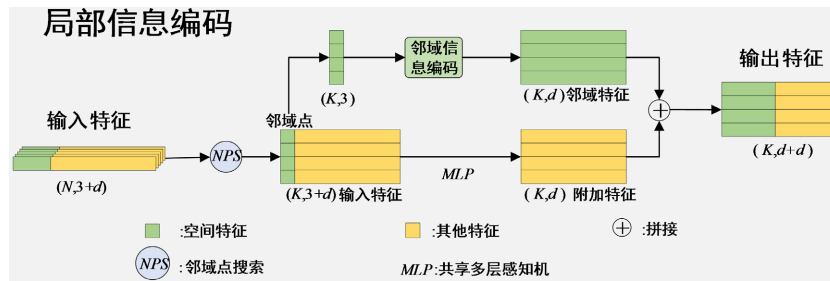


Figure 3. The basic architecture of local information encoding module. $(N, 3 + d)$ represents the dimension of current feature is $N*(3 + d)$, other same representation forms represent same meaning.

图 3. 局部信息编码模块基本结构。 $(N, 3 + d)$ 表示当前特征维度为 $N*(3 + d)$, 其他相同形式的表示也表示同一含义

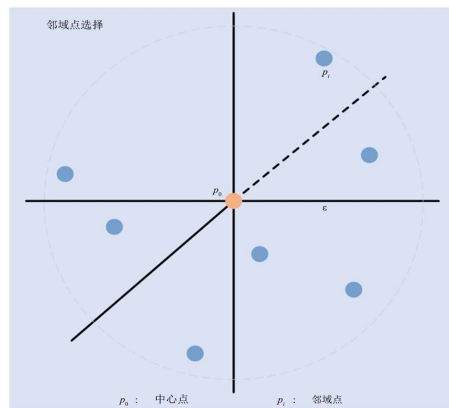


Figure 4. The schematic diagram of neighboring point search
图 4. 邻域点选取示意图

3. 实验及结果讨论

3.1. 实验环境设置

所有的实验的硬件平台都是基于 Tesla T4GPU, 软件环境基于 ubuntu18.04, python = 3.6, Tensorflow = 1.14, cuda = 10.2 cudnn = 7.6.5。实验开始之前需要首先进行编译以使用 scikit-learn 科学计算库。

3.2. 数据集介绍

本文实验基于公开数据集 S3DIS, S3DIS 是斯坦福大学提供的大场景室内 3D 点云数据集, 包含 6 个教学和办公区域, 总共有 695, 878, 620 个带有色彩信息以及语义标签的 3D 点。S3DIS 数据集是规模较大的数据集, 因为其在在一个场景内包含百万以上的点, 这对模型处理大场景的能力提出了较高的要求。

3.3. 实验

在数据集划分上, 本实验采用 6 折交叉验证, 即将数据集随机分成 6 份, 其中 5 份用作训练集, 1 份用作测试集。具体的参数设置如下表 1 所示。在 S3DIS 数据集上的实验结果如表 2 所示。从表中数据可以看出, 改进后的模型的表现比 baseline 方法表现更为优越。从整体的效果来看, 本文所提出的方法比 baseline 方法提高了 1.8%, 在个别难以区分的类别上比其他方法表现更为突出, 如 Beam 横梁类, 证明了所提出方法的有效性。

Table 1. Model training hyperparameter setting
表 1. 模型训练超参数设置

参数名称	默认值
迭代次数	100
学习率	0.01
下采样率	0.04

Table 2. The result of segmentation on S3DIS dataset
表 2. S3DIS 数据集分割结果

方法	mIoU (%)	Ceil	Floor	Wall	Beam	Col.	Wind.	Door	Table	Chair	Sofa	Book.	Board	Clut.
PointNet [13]	47.6	88.0	88.7	69.3	0.0	23.1	47.5	51.6	54.1	42.0	9.6	38.2	29.4	35.2
PointNet++ [14]	53.2	89.6	98.0	75.5	0.0	7.5	61.1	18.7	69.7	78.5	35.6	64.1	49.7	43.9
RandLA-Net [15]	63.4	91.6	97.2	80.9	0.0	21.9	59.5	47.9	77.4	86.1	73.6	67.7	69.6	51.1
SCF-Net [16]	63.6	90.1	96.1	79.5	0.0	29.2	61.6	28.8	76.6	87.6	78.4	72.5	73.0	50.3
CGA-Net [17]	63.1	92.1	96.7	80.7	0.0	18.8	62.8	40.9	76.8	87.6	68.0	69.7	73.4	53.3
Ours	65.2	91.5	96.8	80.8	2.1	30.5	61.1	46.2	74.9	87.2	77.6	72.1	75.1	51.8

3.4. 结果分析

本文以 mIoU 为测试标准分别对前人所提出的方法和本文所改进的方法在相同条件下进行了测试, 证明了所提出方法的有效性。从 PointNet 的提出到现在各种基于点的方法的蓬勃发展[13], 各种方法性能的提升大多是对点云的局部信息进行了重点关注, 提升了局部信息的表达能力。基于这种思路, 本文

从提升点云的邻域特征表达能力出发, 提升点云局部特征的表达能力, 同时我们特别关注点云的局部方向信息, 提升了点云邻域特征的表达能力。

从实验结果可以看出, 本文所提出的方法相比较于其他方法, 在性能上有了显著的提升, 证明了更好的局部特征可以提升分割效果这一假设, 并且与所使用的基线方法比较的话可以看出方向信息对于点云而言提升较为显著, 进一步验证了本文思路的正确性。

4. 结论

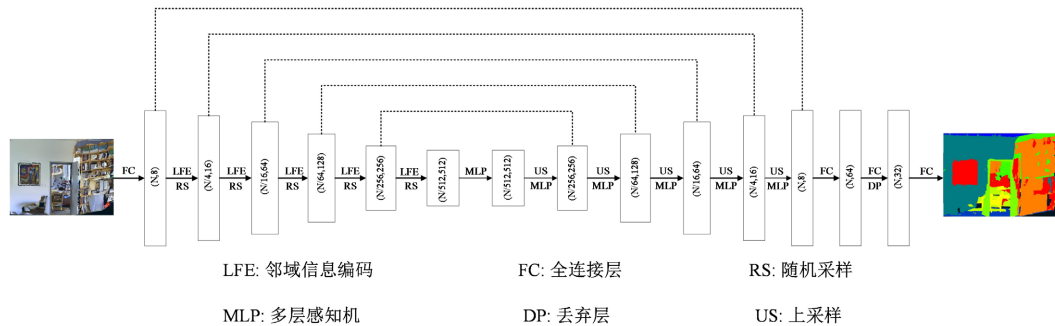


Figure 5. The architecture of network

图 5. 网络整体结构

本文的目的是对现有的网络进行改进从而提升点云分割网络的性能, 得到更好的分割结果, 最终网络的整体结构如图 5 所示。通过添加设计的邻域信息编码模块来提升邻域特征的表达能力从而进一步提升局部特征的表达能力的提高分割的精度。经过在公开数据集上的实验, 可以得出: 在局部特征的提取过程中更加注重中心点周围邻域点的空间信息以及方向信息可以进一步提高局部特征的表达能力的提高最终的分割精度。

基金项目

工业人工智能中跨媒体协同深度安全态势感知理论应用和研究(62266045)。

参考文献

- [1] Lawin, F.J., Danelljan, M., Tosteberg, P., *et al.* (2017) Deep Projective 3D Semantic Segmentation. In: Felsberg, M., Heyden, A. and Kruger, N., Eds., *Computer Analysis of Images and Patterns*, Springer International Publishing, Cham, 95-107. https://doi.org/10.1007/978-3-319-64689-3_8
- [2] Long, J., *et al.* (2015) Fully Convolutional Networks for Semantic Segmentation. 2015 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, 7-12 June 2015, 3431-3440.
- [3] Boulch, A., Saux, B.L. and Audebert, N. (2017) Unstructured Point Cloud Semantic Labeling Using Deep Segmentation Networks. *Proceedings of the Workshop on 3D Object Retrieval*, **3**, 17-24.
- [4] Iandola, F.N., Han, S., Moskewicz, M.W., *et al.* (2016) SqueezeNet: AlexNet-Level Accuracy with 50x Fewer Parameters and < 0.5 MB Model Size.
- [5] Wu, B., Wan, A., Yue, X., *et al.* (2017) Squeezeseg: Convolutional Neural Nets with Recurrent crf for Real-Time 290 Road-Object Segmentation from 3d LiDAR Point Cloud. 2018 *IEEE International Conference on Robotics and Automation (ICRA)*, Brisbane, 21-25 May 2018, 1887-1893. <https://doi.org/10.1109/ICRA.2018.8462926>
- [6] Wu, B., Zhou, X., Zhao, S., *et al.* (2018) SqueezeSegV2: Improved Model Structure and Unsupervised Domain Adaptation for Road-Object Segmentation from a LiDAR Point Cloud. 2019 *International Conference on Robotics and Automation (ICRA)*, Montreal, 20-24 May 2019, 4376-4382. <https://doi.org/10.1109/ICRA.2019.8793495>
- [7] Milioto, A., Vizzo, I., Behley, J., *et al.* (2019) RangeNet ++: Fast and Accurate LiDAR Semantic Segmentation. 2019 *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Macau, 3-8 November 2019, 4213-4220.

- <https://doi.org/10.1109/TROS40897.2019.8967762>
- [8] Huang, J. and You, S. (2016) Point Cloud Labeling Using 3D Convolutional Neural Network. 2016 *23rd International Conference on Pattern Recognition (ICPR)*, Cancun, 4-8 December 2016, 2670-2675.
- [9] Tchapmi, L., Choy, C., Armeni, I., *et al.* (2017) Segcloud: Semantic Segmentation of 3D Point Clouds. 2017 *International Conference on 3D Vision (3DV)*, Qingdao, 10-12 October 2017, 537-547. <https://doi.org/10.1109/3DV.2017.00067>
- [10] Rethage, D., Wald, J., Sturm, J., *et al.* (2018) Fully-Convolutional Point Networks for Large-Scale Point Clouds. *15th European Conference*, Munich, 8-14 September 2018, 625-640. https://doi.org/10.1007/978-3-030-01225-0_37
- [11] Meng, H.-Y., Gao, L., Lai, Y.-K., *et al.* (2019) Vv-net: Voxel Vae Net with Group Convolutions for Point Cloud Segmentation. 2019 *IEEE/CVF International Conference on Computer Vision (ICCV)*, Seoul, 27 October-2 November 2019, 8499-8507. <https://doi.org/10.1109/ICCV.2019.00859>
- [12] Dai, A., Ritchie, D., Bokeloh, M., *et al.* (2018) ScanComplete: Large-Scale Scene Completion and Semantic Segmentation for 3D Scans. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-23 June 2018, 4578-4587. <https://doi.org/10.1109/CVPR.2018.00481>
- [13] Qi, C.R., Su, H., Mo, K., *et al.* (2017) PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation. 2017 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, 21-26 July 2017, 77-85.
- [14] Qi, C.R., Li, Y., Hao, S., *et al.* (2017) PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space. *Proceedings of the 31st International Conference on Neural Information Processing Systems*, Long Beach, 4-9 December 2017, 5105-5114.
- [15] Hu, Q., Yang, B., Xie, L., Rosa, S., *et al.* (2020) RandLA-Net: Efficient Semantic Segmentation of Large-Scale Point Clouds. 2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, 13-19 June 2020, 11105-11114. <https://doi.org/10.1109/CVPR42600.2020.01112>
- [16] Fan, S., Dong, Q., Zhu, F., *et al.* (2021) SCF-Net: Learning Spatial Contextual Features for Large Scale Point Cloud Segmentation. 2021 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Nashville, 20-25 June 2021, 14499-14508. <https://doi.org/10.1109/CVPR46437.2021.01427>
- [17] Lu, T., Wang, L.M. and Wu, G.S. (2021) CGA-Net: Category Guided Aggregation for Point Cloud Semantic Segmentation. 2021 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Nashville, 20-25 June 2021, 11688-11697. <https://doi.org/10.1109/CVPR46437.2021.01152>
- [18] Jiang, M.Y., Wu, Y.R., Zhao, T.Q., *et al.* (2018) PointSIFT: A SIFT-Like Network Module for 3D Point Cloud Semantic Segmentation. arXiv:1807.00652.
- [19] Engelmann, F., Kontogianni, T., Schult, J., *et al.* (2019) Know What Your Neighbors Do: 3D Semantic Segmentation of Point Clouds. *Computer Vision—ECCV 2018 Workshops*, Munich, 8-14 September 2018, 395-409. https://doi.org/10.1007/978-3-030-11015-4_29
- [20] Zhao, H., Jiang, L., Fu, C.W., *et al.* (2019) PointWeb: Enhancing Local Neighborhood Features for Point Cloud Processing. 2019 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, 15-20 June 2019, 5560-5568. <https://doi.org/10.1109/CVPR.2019.00571>