

一种基于空间特征的三维点云语义分割模型

陈立宜, 赵艮平

广东工业大学计算机学院, 广东 广州

收稿日期: 2022年1月15日; 录用日期: 2022年2月11日; 发布日期: 2022年2月21日

摘要

真实的点云场景通常包含非常复杂的环境, 加上三维点云数据所具有的离散性、无序且分布不均匀的属性, 针对点云的详细信息分析往往非常具有挑战性。针对以上问题, 提出一种基于编码器 - 解码器的分割网络。首先, 通过关键点提取模块获取点云的形状特征并通过全连接层提取形状关键点; 然后, 通过空间细节提取模块获取基于关键点的空间细节特征; 最后, 结合形状特征和空间细节特征, 获取点云丰富空间信息及上下文联系, 提高网络的泛化能力。实验结果表明, 我们的网络在公共数据集Urban Semantic 3D (US3D) 和 ISPRS Vaihingen 3D semantic labelling benchmark (ISPRS) 上的平均交并比分别为 93.44% 和 81.45%。我们的网络分割性能良好, 且具有较好的泛化能力。

关键词

点云, 形状特征, 空间特征, 关键点

A Semantic Segmentation Model of 3D Point Clouds Based on Spatial Features

Liyi Chen, Genping Zhao

School of Computer Science and Technology, Guangdong University of Technology, Guangzhou Guangdong

Received: Jan. 15th, 2022; accepted: Feb. 11th, 2022; published: Feb. 21st, 2022

Abstract

Real point cloud scenes usually contain very complex environments. Coupled with the discrete, disordered and uneven distribution properties of 3D point cloud data, detailed analysis of point clouds is often very challenging. Aiming at the above problems, an encoder-decoder-based segmentation network is proposed. First, the shape features of the point cloud are obtained through the key point extraction module and the shape key points are extracted through the fully connected layer; then, the spatial detail features based on key points are obtained through the spatial

detail extraction module; finally, the shape features and spatial detail features are combined to obtain the rich spatial information and contextual connections of point clouds improve the generalization ability of the network. Experimental results show that our network achieves an average intersection ratio of 93.44% and 81.45% on public datasets Urban Semantic 3D (US3D) and ISPRS Vaihingen 3D semantic labelling benchmark (ISPRS), respectively. Our network has good segmentation performance and good generalization ability.

Keywords

Point Cloud, Shape Feature, Spatial Feature, Key Point

Copyright © 2022 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

随着三维传感硬件能力的不断增强，现在可以很容易地在许多场景中捕获三维数据。与二维图像相比，三维数据提供了更丰富的环境信息。三维数据通常可以用不同的格式表示，包括深度图像、点云、网格和体积网格。点云作为一种常用的表示格式，在三维空间中保留了原始的几何信息，不需要任何离散化。因此，它是许多场景理解相关应用(如自动驾驶和机器人)的首选格式。

点云解译是点云处理的重要任务，包括点云对象分类[1]，点云对象检测[2]和点云语义分割[3] [4]等。在这些任务中，将语义标签分配给点的点云语义分割相对具有挑战性。尽管在深度学习策略的帮助下语义分割领域取得了巨大进展，但大多数方法都是针对二维图像[5] [6] [7]的。与二维图像不同的是，在二维图像中，像素网格与对象颜色信息是规则对应的，而三维对象数据是分散的，大多数空间实际上并未被占用，因此直接将深度神经网络操作从二维扩展到三维是低效的。

点云是在同一空间参考系下表达目标空间分布和目标表面特性的海量点集合，其独立描述每个点的相关属性信息，能提供丰富的信息，但点与点之间没有显著的联系，而点云语义分割需要预测每个点的语义类别，提取更精细的点特征，因此需要结合上下文信息。由于点云离散性、无序且分布不均匀的特点，这使得三维点云语义分割任务更具挑战性。

针对三维点云数据的特点，本文提出了一种适用于三维点云数据的语义分割模型。本文的主要贡献分为两个部分：

1) 我们提出了一个关键点提取模块，通过提取多个方向信息学习形状特征，并以此推算形状关键点，生成的关键点跟点云的位置一致性联系不大，但能对形状有敏锐的感知，适应点云无序且分布不均匀的特点。

2) 我们设计了一个空间细节提取模块，通过提取局部细节特征和空间特征，将其聚合形成空间细节特征。它促进了点云不同区域之间的相关性，并整合了点云的上下文信息，提取更精细的特征。

本文结构如下：第二章介绍点云语义分割处理的相关工作，第三章介绍提出的方法，第四章展示实验，第五章总结本文。

2. 相关工作

目前，基于深度学习的三维点云模型语义分割方法主要有：基于投影的方法、基于多视图的方法、

基于体素化的方法和基于点云表示的方法。

基于多视图的方法。受二维图像语义分割方法的启发, Ku 等[8]将三维点云投影为二维鸟瞰图, 然后使用经典的二维深度学习网络提取特征用于三维点云语义分割。Huang 等[9]提出多视图卷积神经网络, 从对应关系中学习局部描述符, 增强了网络的泛化能力。然而, 基于多视图的方法图像只是三维场景的近似值, 因此它们会导致许多局限性和几何结构损失, 丢失了具有鉴别力的几何结构信息。

基于体素化的方法。Wu 等[10]将不规则的三维点云模型转化为规则的体素网格, 这样便可以使用三维卷积神经网络处理体素模型提取特征, 但是该方法存在由点云稀疏性所导致的体素网格排列效率低、计算过程中占用的内存大、训练时间长、信息丢失等缺点。改进算法提出了空间划分方法, 如 Kd 树或者八叉树, 解决了一些分辨率低的问题, 但仍然依赖于边界体的细分, 没有考虑局部几何结构。

基于点云表示的方法。Qi 等在 2017 年提出的 PointNet [3]是第一个直接处理三维点坐标的深度神经网络, 并使用多层感知器和最大池来提取特征。由于最大池化操作是对所有点的全局操作, 因此 PointNet 缺乏对局部区域的理解。PointNet++ [4]网络进一步采用分层编码器 - 解码器结构来考虑局部区域。该网络首先逐层对点云进行下采样, 然后逐步内插到原始分辨率。由于点特征是由多层感知器独立提取的, 因此该框架虽然考虑了点云局部特征, 但仅利用了每个点与其局部上下文之间的弱联系。Jiang 等[11]设计了 PointSIFT 网络, 该网络利用特征检测算法设计一个方向编码单元, 通过在多个方向上堆叠方向编码单元, 学习各个不同尺度的信息。该方法虽然从多个方向获取空间信息, 但任未充分利用点云之间的联系, 获取详细信息的能力有限。

3. 本文方法

为了在三维点云分割中重点探索点与其局部领域上下文之间的语义关系, 本文基于 PointNet++的编码器 - 解码器结构进行改进。首先在编码器阶段使用空间点提取模块生成形状特征并提取关键点, 随后使用空间细节提取模块聚合形成空间细节特征, 将生成的空间细节特征与形状特征结合, 使最后学习到的特征不仅充分学习空间信息, 也充分利用点云之间的联系。在解码器阶段, 通过特征传播层将点数恢复到原始数量。最后, 学习到的特征由全连接层预测。整体网络结构如图 1 所示。

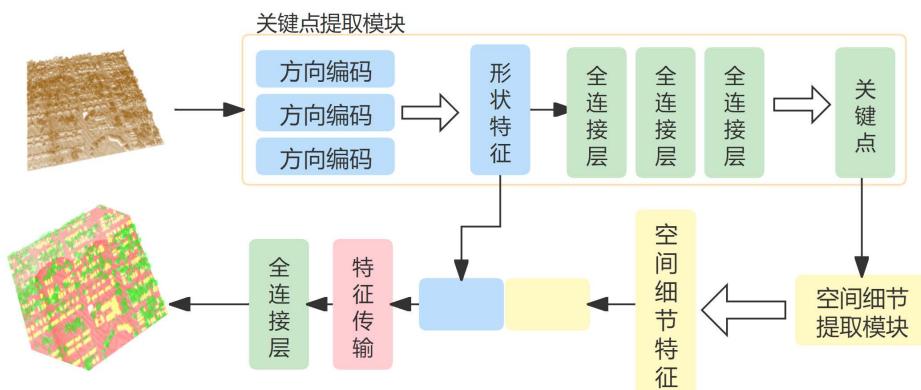


Figure 1. Overall network architecture
图 1. 整体网络架构

3.1. 关键点提取模块

Pointnet++网络虽然采用分层编码器 - 解码器结构来考虑局部区域, 但其每个点操作过于独立, 无法高效刻画相关区域的语义结构, 因此需要有一个结构可以捕获点云的形状特征。在本文中, 我们考虑了

形状表示的方向编码和尺度感知两个基本特征，设计空间点提取模块，针对所有的点使用方向编码卷积，通过学习点的形状特征信息以捕获相关区域的语义信息，并由此推理得到形状关键点。具体地，对于给定特征为 d 维的点 p_0 ，其对应的特征为 f_0 。以特征点 f_0 为中心，将空间分割为八个方向，每个方向中离 f_0 最近的点代表该方向，若在半径 r 内未找到特征点，则用 f_0 表示。此时，每个特征带有周围八个方向的信息。随后，为了使卷积操作能感知方向上的信息，对每个特征点分别在 X、Y、Z 轴上进行三阶段的卷积，将特征点放入张量 $V \in R^{2 \times 2 \times d}$ ，其中前 3 个维度表示 8 个八分圆，结束后每个特征点带有 1 个维度的特征。具体公式如下所示：

$$V_x = g[Conv_x(W_x, V)] \in R^{2 \times 2 \times d}$$

$$V_y = g[Conv_y(W_y, V_x)] \in R^{2 \times d}$$

$$V_z = g[Conv_z(W_z, V_y)] \in R^{1 \times d}$$

其中， W_x 、 W_y 、 W_z 为待优化的卷积权值， $Conv_x$ 、 $Conv_y$ 、 $Conv_z$ 分别表示在 X、Y、Z 轴上的卷积操作， $g[*]$ 表示激活函数 ReLU，并进行批归一化。获得形状特征后将学习到的特征通过三个全连接层输出 K 个点 $Kp = \{kp_u \subseteq \mathbb{R}^3, u = 0, 1, \dots, K-1\}$ 作为形状关键点，其中 kp_j 表示第 j 个关键点。

3.2. 空间细节提取模块

为了充分学习点云之间的联系，我们提取关键点后引入空间细节提取模块。该模块主要包括三部分：

- 1) 提取局部细节特征，2) 提取局部空间特征，3) 特征整合。

提取局部细节特征。对于关键点提取模块输出的 K 个关键点，我们使用一个简单的 knn 搜索来对局部区域进行采样。为了提取与关键点数量相同的局部信息区域，我们将由关键点捕获的 K 个局部邻域分组以形成一个 $K \times M \times 3$ 的局部细节张量。我们使用 (x, y, z) 坐标作为张量的通道，在关键点生成的领域内捕获 M 个点，使用 C_i ， $i = 0, 1, \dots, K-1$ 表示。局部邻域中捕获的每个点由 C_j ， $j = 0, 1, \dots, M-1$ 表示。获得的 K 组点如公式所示：

$$C_i = knn(C_j | kp_u, j = 0, 1, \dots, M-1)$$

随后，我们在局部细节张量上应用一系列多层次感知器以提取 $K \times 256$ 的细节特征。

提取局部空间特征。为了有效学习点云的局部空间特征，我们参考关键点归一化的概念，对每个关键点生成的局部区域内点的位置进行平均，以获得 K 个归一化的关键点

$$Kp^{norm} = \{kp_u^{norm} \subseteq \mathbb{R}^3, u = 0, 1, \dots, K-1\}，其中 kp_u^{norm} 如公式所示：$$

$$kp_u^{norm} = \text{average}(C_i)$$

归一化后的关键点分布在点云表面的判别区域中，有利于点云的空间特征提取。接下来，我们对归一化关键点本身使用 knn 搜索来捕捉点云的局部空间特征。然后对搜索结果进行分组，得到 $K \times M \times 3$ 的局部空间模式张量，其中 M 是每个局部空间区域内归一化关键点的个数。使用与提取细节特征相同的方法提取 $K \times 256$ 的空间特征。

特征整合。在这部分，空间细节提取模块将空间特征与细节特征连接起来，形成 $K \times 512$ 的中间结果。然后该结果通过一堆共享的全连接层进一步聚合，以获取空间细节特征。空间细节特征建立点云不同区域之间的连接，并整合点云的上下文信息。

4. 实验与分析

我们在公共数据集 Urban Semantic 3D (US3D) 和 ISPRS Vaihingen 3D semantic labelling benchmark

(ISPRS)评估了我们的方法，并且本研究采用平均交并比(MIoU)作为评价指标，该评价指标基于算两个集合的交并比，在语义分割的问题中，这两个集合为真实值和预测值进行比较衡量，计算公式如下：

$$mIoU = \frac{1}{a+1} \sum_{i=0}^k \frac{TP}{FN + FP + TP}$$

其中， a 是类别个数(包括空类)， TP 表示预测结果为正例，实际值也为正例，预测正确； FN 表示预测结果为负例，实际值为正例，预测错误； FP 表示预测结果为正例，实际值为负例，预测错误。

4.1. US3D

首先，我们使用 US3D 数据集来评估我们方法的性能。该数据集是一个大型公共数据集，包括多视图和多波段卫星图像、机载激光雷达点云数据和语义标签。在实验中，使用机载激光雷达采集的 8192 个点作为分类输入。数据集中的语义标签包括地面、高处植被、建筑物、水和高架道路，共五类。

为了评估我们方法的性能，我们使用其他两个相关网络 PointNet++、PointSIFT 来进行比较。根据不同值参数设置的分析结果，学习率、batch size 和 epoch size 的超参数分别设置为 0.001、5 和 200。在所有实验中，对于不同的网络，这些参数保持相同。

表 1 给出了不同方法在 MIoU 和 OA 方面 US3D 数据集的分类精度。表 1 显示我们的方法优于三种比较方法的性能。在 MIoU 方面，我们的方法比 PointNet++ 的分类性能提高了 10%，在 OA 方面则提高了 4%。除表 1 外，表 2 总结了每个类别的分类精度。根据表 2，我们提议网络在每个类别中都优于所有其余网络。并且，我们的网络在水上和高架道路类别中取得了特别出色的成绩，IoU 分别大幅增加了 25.22% 和 30.2% (基于最坏情况)。

Table 1. Classification results on the US3D test dataset

表 1. US3D 测试数据集的分类结果

Network	MIoU	OA
PointNet++	0.8358	0.9437
PointSIFT	0.8401	0.9651
Ours	0.9334	0.9870

Table 2. Classification accuracy for each category of the US3D test dataset

表 2. US3D 测试数据集的每个类别的分类准确性

	Ground	High vegetation	Buildings	Water	Elevated roads
PointNet++	0.9695	0.9268	0.8675	0.6714	0.6941
PointSIFT	0.9666	0.9447	0.8420	0.8918	0.5556
Ours	0.9910	0.9826	0.9235	0.9236	0.8578

4.2. ISPRS

我们还在公共 ISPRS 3D 语义标注基准上对我们的方法进行了评估。该数据集是通过德国的一个小村庄上空的 Leica ALS50 系统获取的。该数据集已在 ISPRS 城市分类和三维建筑物重建测试项目的范围内提出，并同时作为 2D 和 3D 语义标注的基准数据集。对于 3D 语义标注，已经定义了九个类别，并且数

据集中的每个点都被相应地标注。

给定的区域被细分为两个部分。第一个区域的点云数据被视为训练集，具有 753,876 个标签点。数据集定义了 9 个类别，分别是电力线、低植被、不透水表面、汽车、栅栏/树篱、屋顶、立面、灌木和树木。第二个点云数据区域被视为测试集，具有 411,722 个未标记的点。

实验的数据处理和操作步骤与在 US3D 数据集上的数据处理和操作步骤相同。**表 3** 对 ISPR 数据的定量分类精度进行了总结，结果表明我们的方法性能优于其他三种竞争方法，在 MIoU 方面，我们的方法比 PointNet++ 的分类性能提高了 17%，在 OA 方面则提高了 5%。这一结论也可以从**表 4** 中得出，**表 4** 中显示了每个类别的分类精度，可以清楚地看到我们的方法能够检测到更好的围栏区域。立面也观察到了同样的结果。

Table 3. Classification results on the ISPRS test dataset

表 3. ISPRS 测试数据集的分类结果

Method	MIoU	OA
PointNet++	0.6403	0.8582
PointSIFT	0.6535	0.8561
Ours	0.8145	0.9104

Table 4. Classification accuracy for each category of the ISPRS test dataset

表 4. ISPRS 测试数据集的每个类别的分类准确性

	PointNet++	PointSIFT	Ours
Powerline	0.205	0.312	0.532
Low vegetation	0.820	0.800	0.888
Impervious surfaces	0.887	0.925	0.943
Car	0.475	0.521	0.813
Fence/Hedge	0.159	0.269	0.542
Roof	0.913	0.916	0.963
Facade	0.464	0.486	0.615
Shrub	0.439	0.484	0.545
Tree	0.769	0.775	0.851

5. 结语

本文提出了一种基于编码器 - 解码器结构的三维点云语义分割算法。首先将关键点提取模块引入特征提取层来挖掘点云的形状特征，并通过形状特征推算形状关键点。随后将关键点送入空间细节提取模块获得基于关键点的空间细节特征，捕捉点云的上下文特征。最后将该特征与形状特征进行结合与形状特征进行优势互补，不仅充分学习点云空间信息，也充分利用点云之间的联系，增强了网络的泛化能力。为了验证我们方法的性能，我们在两个公共数据集上进行了实验，并与其他两个方法进行比较。实验结果表明，我们的方法在两个数据集上都表现优异，总体分割准确率得到了提升，充分证明了其具有良好

的泛化性。

参考文献

- [1] Qi, C.R., Su, H., Niessner, M., Dai, A., Yan, M.Y. and Guibas, L. (2016) Volumetric and Multi-View CNNs for Object Classification on 3D Data. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, 5648-5656.
- [2] Lahoud, J. and Ghanem, B. (2017) 2D-Driven 3D Object Detection in RGB-D Images. 2017 *IEEE International Conference on Computer Vision (ICCV)*, Venice, 22-29 October 2017, 4632-4640. <https://doi.org/10.1109/ICCV.2017.495>
- [3] Qi, C.R., Su, H., Mo, K.C. and Guibas, L.J. (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. <https://doi.org/10.1109/CVPR.2017.16>
- [4] Qi, C.R., Yi, L., Su, H. and Guibas, L.J. (2017) Pointnet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space. *Advances in Neural Information Processing Systems*, **30**.
- [5] Chen, L., Papandreou, G., Kokkinos, I., Murphy, K. and Yuille, A.L. (2017) DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **40**, 834-848. <https://doi.org/10.1109/TPAMI.2017.2699184>
- [6] Long, J., Shelhamer, E. and Darrell, T. (2015) Fully Convolutional Networks for Semantic Segmentation. 2015 *IEEE Conference on Computer Vision and Pattern Recognition*, Boston, 7-12 June 2015, 3431-3440. <https://doi.org/10.1109/CVPR.2015.7298965>
- [7] Garcia-Garcia, A., Orts-Escolano, S., Oprea, S., Villena-Martinez, V. and Rodriguez, J.G. (2017) A Review on Deep Learning Techniques Applied to Semantic Segmentation.
- [8] Ku, J., Mozifian, M., Lee, J., et al. (2018) Joint 3D Proposal Generation and Object Detection from View Aggregation. 2018 *IEEE/RSJ International Conference on Intelligent Robots and Systems*, Madrid, 1-5 October 2018, 1-8. <https://doi.org/10.1109/IROS.2018.8594049>
- [9] Huang, H.B., Kalogerakis, E., Chaudhuri, S., et al. (2018) Learning Local Shape Descriptors from Part Correspondences with Multiview Convolutional Networks. *ACM Transactions on Graphics (TOG)*, **37**, 1-14. <https://doi.org/10.1145/3137609>
- [10] Wu, Z., Song, S., Khosla, A., et al. (2015) 3D ShapeNets: A Deep Representation for Volumetric Shapes. *Proceedings of the IEEE Conference on Computer vision and Pattern Recognition*, 2015, 1912-1920.
- [11] Jiang, M., Wu, Y., Zhao, T., et al. (2018) PointSIFT: A SIFT-Like Network Module for 3D Point Cloud Semantic Segmentation.