



中国图形学报

2021
12
VOL.26

储量利用度80%

储量利用度95%

储量利用度95%

储量利用度98%

储量利用度90%

储量利用度100%

储量利用度82%

储量利用度82%

储量利用度90%

储量利用度75%

储量利用度75%

储量利用度75%

储量利用度90%

储量利用度80%

储量利用度90%

遥感油罐监测 P2953



第26卷第12期（总第308期）
2021年12月16日

中国精品科技期刊
中国国际影响力优秀学术期刊
中国科技核心期刊
中文核心期刊

版权声明

凡向《中国图象图形学报》投稿，均视为同意在本刊网站及CNKI等全文数据库出版，所刊载论文已获得著作权人的授权。本刊所有图片均为非商业目的使用，所有内容，未经许可，不得转载或以其他方式使用。

Copyright

All rights reserved by Journal of Image and Graphics, Institute of Remote Sensing and Digital Earth, CAS. The content (including but not limited text, photo, etc) published in this journal is for non-commercial use.

主管单位 中国科学院
主办单位 中国科学院空天信息创新研究院
中国图象图形学学会
北京应用物理与计算数学研究所

主 编 吴一戎
编辑出版 《中国图象图形学报》编辑出版委员会
通信地址 北京市海淀区北四环西路19号
邮 编 100190
电子信箱 jig@aircas.ac.cn
电 话 010-58887035
网 址 www.cjig.cn

广告发布登记号 京朝工商广登字20170218号
总 发 行 北京报刊发行局
订 购 全国各地邮局
海外发行 中国国际图书贸易集团有限公司
(邮政信箱: 北京399信箱 邮编: 100048)
印刷装订 北京科信印刷有限公司

Journal of Image and Graphics

Title inscription: Song Jian | Monthly, Started in 1996

Superintended by Chinese Academy of Sciences
Sponsored by Aerospace Information Research Institute, CAS
China Society of Image and Graphics
Institute of Applied Physics and Computational Mathematics

Editor-in-Chief Wu Yirong
Editor, Publisher Editorial and Publishing Board of Journal of Image and Graphics
Address No. 19, North 4th Ring Road West, Haidian District, Beijing, P. R. China
Zip code 100190
E-mail jig@aircas.ac.cn
Telephone 010-58887035
Website www.cjig.cn

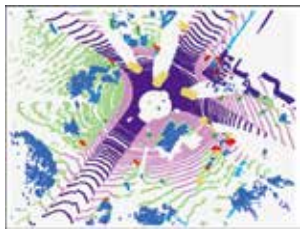
Distributed by Beijing Bureau for Distribution of Newspapers and Journals
Domestic All Local Post Offices in China
Overseas China International Book Trading Corporation
(P.O.Box 399, Beijing 100048, P.R.China)
Printed by Beijing Kexin Printing Co., Ltd.

CN 11-3758/TB
ISSN 1006-8961
CODEN ZTTXFZ

国外发行代号 M1406
国内邮发代号 82-831
国内定价 60.00元



环视鱼眼图像处理深度学习
研究进展(第2778页)



结合空间深度卷积和残差
的大尺度点云场景分割
(第2848页)



曼哈顿世界环境下消失点非线
性优化估计算法(第2931页)

综述

深度对抗视觉生成综述

谭明奎, 许守恺, 张书海, 陈奇 2751

人机交互中的智能体情感计算研究

刘婷婷, 刘箴, 柴艳杰, 王瑾, 王媛怡 2767

环视鱼眼图像处理深度学习研究进展

包俊, 刘宏哲, 褚文博 2778

图像处理和编码

图像分解与色彩先验下的多曝光图像融合

李嫛源, 王琴, 朱智勤, 齐观秋 2800

双尺度分解和显著性分析相结合的红外与可见光图像融合

霍星, 邹韵, 陈影, 檀结庆 2813

轻量级注意力特征选择循环网络的超分辨率重建

徐雯捷, 宋慧慧, 袁晓彤, 刘青山 2826

区域级通道注意力融合高频损失的图像超分辨率重建

周波, 李成华, 陈伟 2836

图像分析和识别

结合空间深度卷积和残差的大尺度点云场景分割

刘盛, 黄圣跃, 程豪豪, 沈家瑜, 陈胜勇 2848

细粒度图像分类的互补注意力方法

赵勋, 王家宝, 李阳, 王亚鹏, 苗壮 2860

联合分块谱带优选和深度特征的高光谱人脸识别

谢志华, 李毅, 牛杰一 2870

时空双仿射微分不变量及骨架动作识别

李琪, 墨瀚林, 赵婧涵, 郝宏翔, 李华 2879

多尺度分割先验迭代重加权低秩恢复显著性检测

张荣国, 郑维佳, 赵建, 胡静, 刘小君 2892

融合环境特征与改进YOLOv4的安全帽佩戴检测

葛青青, 张智杰, 袁琰, 李秀梅, 孙军梅 2904

图像理解和计算机视觉

关节轴角先验对3维人体重建结果的影响

姚砾, 张幼安, 张梦雪, 万燕 2918

曼哈顿世界环境下消失点非线性优化估计算法

韩冰心, 卢惠民, 于清华, 张礼廉 2931

双目机器视觉及RetinaNet模型的路侧行人感知定位

连丽容, 罗文婷, 秦勇, 李林 2941

遥感图像处理

遥感图像中油罐目标精确定位与参数提取

江哈, 张月婷, 郭嘉逸, 赵鑫, 李芳芳, 黄丽佳, 胡玉新, 雷斌, 丁赤飏 2953

参数合成空间变换网络的遥感图像一致性配准

陈颖, 张祺, 李文举, 石艳娇, 陈磊 2964

改进Mask R-CNN模型的海洋锋检测

徐慧芳, 黄冬梅, 贺琪, 杜艳玲, 覃学标 2981

噪声鲁棒的轻量级深度遥感场景图像分类检索

王亚鹏, 李阳, 王家宝, 赵勋, 苗壮 2991

改进U-Net型网络的遥感图像道路提取

杨佳林, 郭学俊, 陈泽华 3005

弱监督深度语义分割网络的多源遥感影像水体检测

李鑫伟, 李彦胜, 张永军 3015

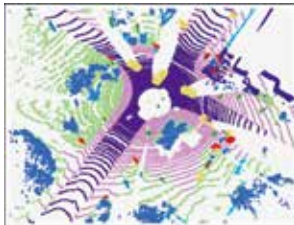
总目次 I

CONTENTS

JOURNAL OF IMAGE AND GRAPHICS



Research progress of fisheye image processing based on deep learning(P2778)



A deep residual network with spatial depthwise convolution for large-scale point cloud semantic segmentation(P2848)



Vanishing point estimation based on non-linear optimization in manhattan world environments (P2931)

Review

A review on deep adversarial visual generation	
Tan Mingkui, Xu Shoukai, Zhang Shuhai, Chen Qi	2751
Agent affective computing in human-computer interaction	
Liu Tingting, Liu Zhen, Chai Yanjie, Wang Jin, Wang Yuanyi	2767
Research progress of fisheye image processing based on deep learning	
Bao Jun, Liu Hongzhe, Chu Wenbo	2778

Image Processing and Coding

Multi-exposure image fusion based on image decomposition and color prior	
Li Yuanyuan, Wang Qin, Zhu Zhiqin, Qi Guanqiu	2800
Dual-scale decomposition and saliency analysis based infrared and visible image fusion	
Huo Xing, Zou Yun, Chen Ying, Tan Jieqing	2813
Lightweight attention feature selection recursive network for super-resolution	
Xu Wenjie, Song Huihui, Yuan Xiaotong, Liu Qingshan	2826
Region-level channel attention for single image super-resolution combining high frequency loss	
Zhou Bo, Li Chenghua, Chen Wei	2836

Image Analysis and Recognition

A deep residual network with spatial depthwise convolution for large-scale point cloud semantic segmentation	
Liu Sheng, Huang Shengyue, Cheng Haohao, Shen Jiayu, Chen Shengyong	2848
Complemented attention method for fine-grained image classification	
Zhao Xun, Wang Jiabao, Li Yang, Wang Yapeng, Miao Zhuang	2860
Hyperspectral face recognition based on partition bands optimal selection and deep features	
Xie Zhihua, Li Yi, Niu Jieyi	2870
Spatio-temporal dual affine differential invariants for skeleton-based action recognition	
Li Qi, Mo Hanlin, Zhao Jinghan, Hao Hongxiang, Li Hua	2879
Iterative reweighted low-rank recovery salient object detection with multiscale segmentation prior	
Zhang Rongguo, Zheng Weijia, Zhao Jian, Hu Jing, Liu Xiaojun	2892
Safety helmet wearing detection method of fusing environmental features and improved YOLOv4	
Ge Qingqing, Zhang Zhijie, Yuan Long, Li Xiumei, Sun Junmei	2904

Image Understanding and Computer Vision

The impact of joint axis angle prior on the results of 3D human body reconstruction	
Yao Li, Zhang Youan, Zhang Mengxue, Wan Yan	2918
Vanishing point estimation based on non-linear optimization in manhattan world environments	
Han Bingxin, Lu Huimin, Yu Qinghua, Zhang Lilian	2931
Roadside pedestrian detection and location based on binocular machine vision and RetinaNet	
Lian Lirong, Luo Wenting, Qin Yong, Li Lin	2941

Remote Sensing Image Processing

Accurate localization and parameter extraction of oil tank in remote sensing images	
Jiang Han, Zhang Yueting, Guo Jiayi, Zhao Xin, Li Fangfang, Huang Lijia, Hu Yuxin, Lei Bin, Ding Chibiao	2953
Consistent registration of remote sensing images in parametric synthesized spatial transformation network	
Chen Ying, Zhang Qi, Li Wenju, Shi Yanjiao, Chen Lei	2964
Ocean front detection method based on improved Mask R-CNN	
Xu Huifang, Huang Dongmei, He Qi, Du Yanling, Qin Xuebiao	2981
A robust lightweight deep neural network for remote sensing scene image classification and retrieval under label noise	
Wang Yapeng, Li Yang, Wang Jiabao, Zhao Xun, Miao Zhuang	2991
Road extraction method from remote sensing images based on improved U-Net network	
Yang Jialin, Guo Xuejun, Chen Zehua	3005
Weakly supervised deep semantic segmentation network for water body extraction based on multi-source remote sensing imagery	
Li Xinwei, Li Yansheng, Zhang Yongjun	3015

中图法分类号: TP391 文献标识码: A 文章编号: 1006-8961(2021)12-2848-00

论文引用格式: Liu S, Huang S Y, Cheng H H, Shen J Y and Chen S Y. 2021. A deep residual network with spatial depthwise convolution for large-scale point cloud semantic segmentation. Journal of Image and Graphics, 26(12): 2848-2859 [刘盛, 黄圣跃, 程豪豪, 沈家瑜, 陈胜勇. 2021. 结合空间深度卷积和残差的大尺度点云场景分割. 中国图象图形学报, 26(12): 2848-2859] [DOI: 10. 11834/jig. 200477]

结合空间深度卷积和残差的大尺度点云场景分割

刘盛, 黄圣跃, 程豪豪, 沈家瑜, 陈胜勇*

浙江工业大学计算机科学与技术学院, 杭州 310023

摘要: 目的 在点云场景中, 语义分割对场景理解来说是至关重要的视觉任务。由于图像是结构化的, 而点云是非结构化的, 点云上的卷积通常比图像上的卷积更加困难, 会消耗更多的计算和内存资源。在这种情况下, 大尺度场景的分割往往需要分块进行, 导致效率不足并且无法捕捉足够的场景信息。为了解决这个问题, 本文设计了一种计算高效且内存高效的网络结构, 可以用于端到端的大尺度场景语义分割。方法 结合空间深度卷积和残差结构设计空间深度残差(spatial depthwise residual, SDR)块, 其具有高效的计算效率和内存效率, 并且可以有效地从点云中学习到几何特征。另外, 设计一种扩张特征整合(dilated feature aggregation, DFA)模块, 可以有效地增加感受野而仅增加少量的计算量。结合 SDR 块和 DFA 模块, 本文构建 SDRNet(spatial depthwise residual network), 这是一种 encoder-decoder 深度网络结构, 可以用于大尺度点云场景语义分割。同时, 针对空间卷积核输入数据的分布不利于训练问题, 提出层级标准化来减小参数学习的难度。特别地, 针对稀疏雷达点云的旋转不变性, 提出一种特殊的 SDR 块, 可以消除雷达数据绕 Z 轴旋转的影响, 显著提高网络处理激光雷达点云时的性能。结果 在 S3DIS(Stanford large-scale 3D indoor space)和 SemanticKITTI(Karlsruhe Institute of Technology and Toyota Technological Institute)数据集上对提出的方法进行测试, 并分析点数与帧率的关系。本文方法在 S3DIS 数据集上的平均交并比(mean intersection over union, mIoU)为 71.7%, 在 SemanticKITTI 上的 mIoU 在线单次扫描评估中达到 59.1%。结论 实验结果表明, 本文提出的 SDRNet 能够直接在大尺度场景下进行语义分割。在 S3DIS 和 SemanticKITTI 数据集上的实验结果证明本文方法在精度上有较好表现。通过分析点数量与帧率之间的关系, 得到的数据表明本文提出的 SDRNet 能保持较高精度和较快的推理速率。

关键词: 深度学习; 语义分割; 标准化; 点云; 残差神经网络; 感受野

A deep residual network with spatial depthwise convolution for large-scale point cloud semantic segmentation

Liu Sheng, Huang Shengyue, Cheng Haohao, Shen Jiayu, Chen Shengyong*

College of Computer Science and Technology, Zhejiang University of Technology, Hangzhou 310023, China

Abstract: **Objective** Point cloud semantic segmentation has been an essential visual task for scene understanding from two-dimensional vision to three-dimensional vision. Deep learning processing point cloud has been divided into three methods as following: point-based method, projection-based method and voxel-based method. Projection-based methods have obtained a two-dimensional image from the point cloud based on spherical projection. The semantic segmentation on the point cloud

收稿日期: 2020-08-20; 修回日期: 2020-09-18; 预印本日期: 2020-09-25

* 通信作者: 陈胜勇 csy@zjut.edu.cn

基金项目: 国家重点研发计划资助(2018YFB1305200)

Supported by: National Key R&D Program of China (2018YFB1305200)

has been conducted via a two-dimensional convolution neural network method. The original point cloud has been restored via some post-processing. However, those methods have usually only been used for LiDAR point clouds. Voxel-based methods have often consumed a lot of memory due to voxel representation. The above two methods have both represented the unstructured point cloud into a structured form and processed it via a two-dimensional convolutional neural network or a three-dimensional convolutional neural network. However, this method will lose geometric details. Point-based methods have often consumed more memory subjected to additional neighborhood information storage. Some existing methods have usually divided the entire point cloud into blocks for processing. However, this method will destroy the geometric structure of the scene to cause incomplete information capture from the scene. In addition, some point-based methods in large-scale scenes have the problem of insufficient receptive fields caused by shallow network structures due to excessive memory consumption. A computation-based and memory-efficient network structure has been presented that can be used for end-to-end large-scale scene semantic segmentation. **Method** The spatial depthwise residual (SDR) block has been designed via combining the spatial depthwise convolution and residual structure to learn geometric features from the point cloud effectively. The receptive field has been regarded as one of the key factors in semantic segmentation. In order to increase the receptive field, a dilated feature aggregation (DFA) module, which has a larger receptive field than the SDR block, but with less calculation. The core idea of this module has reduced computational consumption and memory consumption via down sampling. Combining SDR block and DFA module, SDRNet, a deeper encoder-decoder network structure has been constructed, which can be applied to large-scale scenes semantic segmentation. The data distribution of the input data has affected the training process of the network. Data distribution is not conducive to network learning based on the analysis of input data of the convolution kernel. Hierarchical normalization (HN) can reduce the learning difficulty of the convolution kernel. A special SDR block has been used for a kind of rotation invariance of sparse LiDAR point clouds. Before convolution, the point and its neighborhood have been first rotated to a fixed angle. The influence of the rotation of the radar data around the Z-axis can be eliminated. The prediction result has not be changed via the rotated point cloud around the Z-axis. This special SDR block can significantly improve the performance of the network when processing LiDAR point clouds. **Result** The stanford large-scale 3D indoor space (S3DIS) dataset and the Karlsruhe Institute of Technology and Toyota Technological Institute (SemanticKITTI) dataset have been used. Different parameters for different tasks to adapt to the application scenarios of the task have been setup. A larger model for higher accuracy has been constructed because the S3DIS task has been focused on accuracy. The SemanticKITTI scene has required more speed. A lighter hyperparameter has been chosen. The designed model has been compared with several state-of-the-art models on the S3DIS datasets by using 6-fold cross validation. Mean intersection over union (mIoU), mean accuracy (mAcc) and overall accuracy (OA) have been evaluated on the S3DIS dataset. The method has achieved 88.9% OA, 82.4% mAcc and 71.7% mIoU each. These methods have presented well on different metrics. The online single scan evaluation has been conducted on the SemanticKITTI dataset. 59.1% mIoU has been obtained. The method has achieved better results in mIoU and several accuracy of several classes in comparison with point-based methods and projection-based methods. In an unmanned driving scenario like SemanticKITTI, the inference speed of the mode is a crucial factor. In addition, the inference speed of SDRNet with the different number of points has been tested. When the number of points is 50 K, the network processing point cloud speed can reach 11 frames per second (fps) by using a machine with NVIDIA RTX and i7-8700K. Moreover, this paper has constructed experiment of ablation study to explain the performance of each part of the model further. **Conclusion** The experiments on the S3DIS dataset SemanticKITTI dataset have shown that the research method can directly perform semantic segmentation in large-scale point cloud scenes. It can extract information from the scene effectively and achieve high accuracy. The experiment of ablation study on S3DIS area-5 has demonstrated that both DFA and HN can improve performance. The experiment of ablation study on SemanticKITTI validation set has presented that eliminating the influence of rotation by using the special SDR block can effectively improve the performance of the network. The higher accuracy and a relatively fast speed have been achieved via analyzing the relationship between the number of points and the frame rate.

Key words: deep learning; semantic segmentation; normalization; point cloud; residual neural network; receptive field

0 引言

深度学习在计算机视觉领域发展迅速,在视觉任务中取得了优异成绩。许多经典的网络结构(He等,2016;Krizhevsky等,2017;Simonyan和Zisserman,2015)被用于处理2维图像,解决2维视觉任务,例如检测、分割、分类和跟踪等,并逐步发展到3维视觉上。3维点云是典型的3维数据,通常通过TOF(time of flight)相机和激光雷达扫描获取。然而,3维点云是一种非结构化数据,无序性和密度不一致性导致其不能像2维图像一样使用结构化的卷积核融合周围点云的特征信息。

语义分割是计算机视觉的重要任务之一。在2维图像中,语义分割的任务是为每个像素分配一个标签,即逐像素分类。而在3维点云中,需要为点云中的每个点分配标签。语义分割可以帮助计算机进行场景理解,从而更好地完成工业任务。尤其是在无人驾驶场景中,语义分割是一项核心任务。有效且高效地理解场景信息,能帮助无人驾驶汽车提高效率 and 安全性,加速产业落地,推动产业发展。目前,在2维图像上进行语义分割已经取得了显著成果(Long等,2015;Yu等,2020)。在3维点云上也不断有新的研究成果(Yan等,2020;Zhang等,2020)提出,但不像2维语义分割那样成熟。

点云语义分割方法按数据表示可以分为基于投影的方法、基于体素的方法和基于点的方法。1)基于投影的方法先将点云通过球面投影得到2维图像,然后通过2维卷积神经网络的方法对点云进行语义分割。再通过一些后处理还原到原始点云上。SqueezeSeg系列(Wu等,2018,2019a;Xu等,2020)将点云数据投影之后,像SqueezeNet(Iandola等,2016)一样能够快速地进行语义分割。3D-MiniNet(Alonso等,2020a)完成了更高的速度和精度,其结构主要基于MiniNet(Alonso等,2020b)。RangeNet++(Milioto等,2019)在DarkNet(Redmon和Farhadi,2018)的基础上进行修改,使网络能进行语义分割,并提出了一种快速的、支持GPU的、基于KNN(K-nearest neighbor)的算法用于后处理。然而,这类方法通常只能应用于雷达点云,并不适用于所有类型的点云,且在投影过程中常常伴随几何结构损失。2)基于体素的方法(Liu等,2019;Rethage等,

2018;Tchapmi等,2017)首先将点云转化为体素,然后在体素上进行3维卷积。这种方法通常占用大量的内存,计算消耗巨大,并常常伴随几何细节损失,且由于点云的稀疏性导致体素表示中含有大量冗余数据。3)基于点的方法。Qi等人(2017a)提出PointNet首次用于逐点学习点云特征,用maxpooling整合点云全局特征,设计了一个T-net来缓解旋转对特征学习的影响,并用于点云分类和分割任务,但无法有效提取局部特征信息。PointNet++(Qi等,2017b)是一种改进结构,在PointNet的基础上加强了局部信息的提取。PointConv(Wu等,2019b)提出了一种空间连续卷积方法,并提出了一种高效的实现方式,能够有效降低内存消耗。HDGCNN(hierarchical depthwise graph convolutional neural network)(Liang等,2019)将Depthwise卷积(Howard等,2017)拓展到图卷积上并用于3D点云,可以有效减少内存和计算消耗,并依此设计了一种转置卷积方法用于上采样。这些方法都可以帮助完成语义分割,但由于内存消耗过大,导致无法直接在大尺度场景下有效应用,需要分块后进行处理。在大尺度点云场景下进行语义分割,需要更轻量、计算效率更高的网络结构。SPG(superpoint graphs)(Landrieu和Simonovsky,2018)将大尺度点云场景表示为互相相连的超点图,然后在超点图上使用PointNet学习特征。RandLA-Net(Hu等,2020)分析了大尺度场景下采样方式对内存消耗和计算消耗的影响,并设计了一种基于注意力机制的模块用于整合局部特征,可以在大尺度点云场景下进行分割。

虽然这些方法都能提取局部特征并在相关领域取得了不错的性能,但部分方法存在内存消耗过大或者计算消耗过大问题,或者因为前面两种问题而无法设计更深的网络结构导致语义信息提取不足。尤其是在大尺度点云场景中,这些问题更加明显。

为了设计更深和特征表达能力更强且能够用于大尺度点云场景的网络结构,本文提出了空间深度残差(spatial depthwise residual,SDR)块,并基于此设计了更深的网络结构——空间深度残差网络(spatial depthwise residual network,SDRNet)。本文的主要贡献包括:1)提出一种SDR块,可以有效融合局部信息,并具有内存高效性和计算高效性;2)为了扩大感受野,设计了扩张特征整合(dilate feature aggregation,DFA)模块,引入少量的计算量,但

可以有效增大感受野;3)通过堆叠 SDR 块和结合 DFA 模块,设计了一种 encoder-decoder 的网络结构——SDRNet,并用于大尺度点云场景的语义分割;4)对点云数据特性进行分析,针对数据分布过于集中问题,提出层级标准化来缓解输入数据分布造成的影响。针对雷达点云的旋转不变性质,在网络中使用一种特殊的 SDR 块。两者都可以有效降低参数学习的难度,提高网络性能。

1 SDRNet 模型总览

本文提出的 SDRNet 结构如图 1 所示。这是一个 4 阶段的 SDRNet,整体为 encoder-decoder 结构,直接输入点云和每个点的特征来预测语义标签。网络结构的各部分如下:

1) encoder-decoder 结构。类似于 U-Net (Ronneberger 等, 2015) 的设计,SDRNet 设计为 encoder-decoder 结构。网络结构分为多个阶段,每个阶段包含一个编码器和解码器。

2) 编码器。编码器由 DFA 模块和数个 SDR 块组成。特别地,在最后一阶段不使用 DFA 模块。在编码器中,可以通过调整 SDR 块的数量来平衡性能和速度。很显然,更多数量的 SDR 块会使编码层有更大的感受野,同时引入更多的非线性因素,并会有更大的计算量和内存消耗。值得注意的是,过大的感受野和过多的非线性因素不一定能有更好的

性能。

3) 解码器。解码器的输入来自两部分。一部分来自同一阶段的编码器,另一部分来自下一阶段解码器输出的特征。在解码器中,将下一阶段解码器输出的特征进行上采样,然后上采样的特征与编码器输出的特征并联。之后使用一个线性层和一个激活函数输出特征。

4) 下采样。RandLA-Net (Hu 等, 2020) 中讨论了不同采样方法对网络的计算消耗影响。在大尺度场景下,随机采样作为最合适的方式在计算消耗方面有明显优势。受 RandLA-Net 启发,SDRNet 在下采样时采用随机采样的方式。每次进行下采样时,随机选取 1/4 部分的点作为下一阶段的点云,并结合用 KD-Tree 算法找到的邻居点,使用池化操作对每个点与其邻居点的特征信息进行融合,并抑制噪声。

5) 上采样。对于每个上采样层,直接使用最近邻插值将点云扩展到先前的点云。最近邻插值具有高计算效率和高内存效率的特点,符合网络结构的设计思想。

6) SegHead。为了获得语义标签,第 1 阶段解码器的输出特征输入到一个 SegHead 中。SegHead 由 3 层线性层和 1 个 softmax 层组成。为了提高模型的泛化能力,在第 2 个线性层后使用 dropout 层。softmax 层输出每个点属于每个类别的概率。最后,选择每个点概率最大的类别标签作为预测结果。

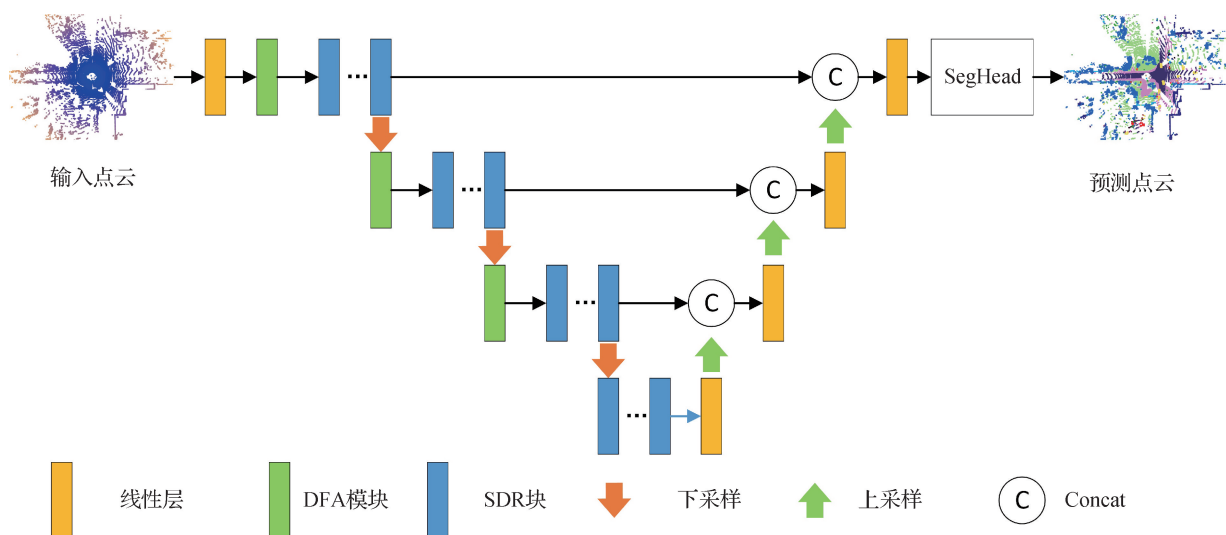


图 1 SDRNet 结构总览

Fig. 1 The overview of SDRNet

7) 多尺度监督。为了强化隐层中的语义信息,使用多尺度监督。在第 2 层后的每个阶段,将编码器输出的特征通过一个辅助的 SegHead 来预测当前阶段的每个点的标签。

总体而言,本文提出的 SDRNet 是一种深层的网络结构,具有足够大的感受野。并且由于其内存高效性和计算高效性,网络可以一次性处理大尺度的场景点云。

2 算法细节

2.1 空间深度残差块

本文使用空间深度残差(spatial depthwise residual, SDR)块在点云上提取特征。2D 图像和 3D 体素是结构化的,且其邻域之间的关系是固定的,可以使用结构化的离散卷积核进行处理。而点云是非结构化的,每个点的邻居的位置通常不固定。在这种情况下,使用连续卷积是一种在点云上进行卷积的方法。

给定点集 $P = \{p_1, p_2, \dots, p_i, \dots, p_N\} \subseteq \mathbf{R}^3$, 以及对应的特征集合 $F = \{f_1, f_2, \dots, f_i, \dots, f_N\}$, 其中, N 是点云中点的数量, p_i 和 f_i 分别代表第 i 个点的空间坐标和特征。对于每个点 p_i , 有其邻域点的索引集合 $\Omega_i = \{a_{i,1}, a_{i,2}, a_{i,3}, \dots, a_{i,k}, \dots, a_{i,K}\}$, 其中 K 是邻居点的数量, $a_{i,k}$ 代表第 i 个点的第 k 个邻居点的索引, Ω_i 可以通过 KD-tree 算法 (Bentley, 1975) 获取。

为了减少计算量和内存消耗,使用空间深度卷积(spatial depthwise convolution, SDC)在点云上融合局部信息从而提取新特征。与 Liang 等人(2019)的方法相同,假定存在一个连续函数 $g(\cdot): \mathbf{R}^3 \rightarrow \mathbf{R}^{IN}$, 可以获取卷积核在空间任意位置的参数,其中 IN 是输入特征的维度。通过函数 $g(\cdot)$, 可以像 2D 图像上使用深度卷积(depthwise convolution)一样 (Howard 等, 2017), 在 3 维点云上进行卷积。由此可以整合 p_i 的特征 f_i 和其邻居点的特征, 提取出新的特征 f'_i , SDC 的定义为

$$f'_i = \sum_{j \in \Omega_i} f_j \odot g(\Delta p_{i,j}) \quad (1)$$

式中, $\Delta p_{i,j}$ 是 p_i 与 p_j 的坐标差, 虽然这种卷积方式无法改变维度, 但可以显著降低内存消耗。SDC 的实现方法如图 2 所示。

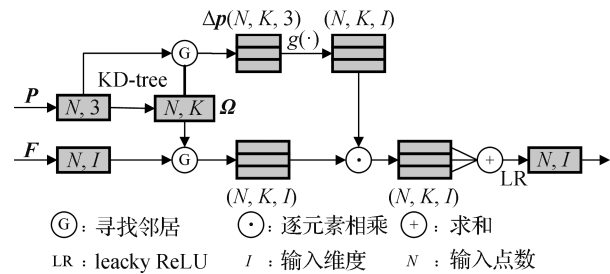


图 2 空间深度卷积的实现方法

Fig. 2 The implementation of the spatial depthwise convolution

特别地,处理雷达点云时,雷达点云数据具有一种旋转不变性质。如图 3 所示,实线点和虚线点分别代表原始点云和绕 Z 轴旋转后的点云。显然,无论点云绕 Z 轴如何旋转,图中相同颜色的点具有相同的性质,并且经过卷积后的特征也不因绕 Z 轴旋转而改变。

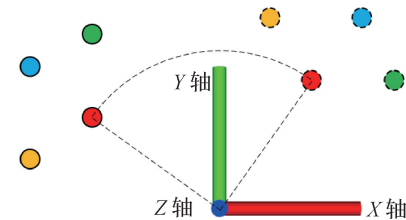


图 3 原始点云和旋转后的点云

Fig. 3 Original point cloud and rotated point cloud

为消除旋转对结果的影响,将所有的 $\Delta p_{i,j}$ 进行旋转,得到对应的 $\Delta p'_{i,j}$, 具体为

$$\Delta p'_{i,j} = \begin{bmatrix} \cos \theta_i & \sin \theta_i & 0 \\ -\sin \theta_i & \cos \theta_i & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \Delta x_{i,j} \\ \Delta y_{i,j} \\ \Delta z_{i,j} \end{bmatrix} \quad (2)$$

式中, $\theta_i = \tan^{-1}(y_i/x_i)$ 。由此可以得到用于雷达点云的卷积公式, 具体为

$$f'_i = \sum_{j \in \Omega_i} f_j \odot g(\Delta p'_{i,j}) \quad (3)$$

式中, $g(\cdot)$ 可以用一个线性层和 ReLU 激活函数替换, 其参数在训练过程中学习。结合这种卷积方式, 本文设计了 SDR 块, 进一步减少计算消耗和内存消耗, 并引入更多的非线性因素。具体结构如图 4 所示。先将输入的点云特征用线性层压缩通道数, 进一步降低卷积操作的计算消耗和内存消耗。再根据数据来源选择式(1)或式(3)在点云上进行 SDC 操作融合局部特征信息。之后用一个线性层对卷积后的特征进行重组。最后用线性层拓展特征维度到输

出维度,与 shortcut 相加并经过激活函数输出特征。shortcut 有两种情况,输入与输出维度相等时,直接将输入特征作为 shortcut;输入特征与输出特征维度不相等时,通过线性层拓展维度得到 shortcut。

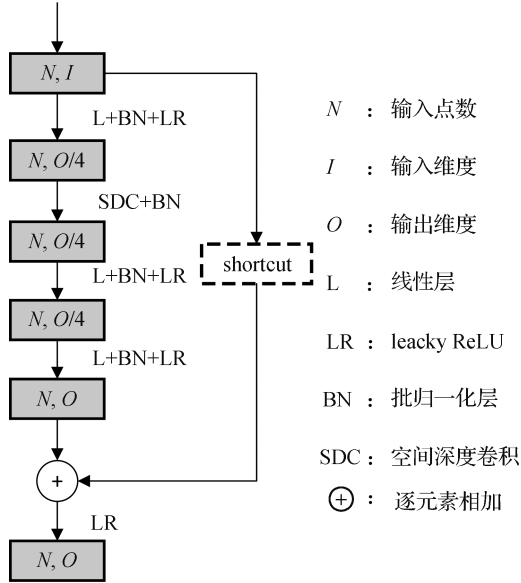


图4 空间深度残差块

Fig.4 Spatial depthwise residual block

SDR 块可以有效整合点与其邻域的特征信息。而且由于其内存高效性,可以通过堆叠 SDR 块构建足够深的网络并用于大尺度点云场景分割。

2.2 扩张特征整合模块

对语义分割来说,感受野是关键因素之一。DPC (dilated point convolutions) (Engelmann 等, 2020) 先通过增大寻找邻域点时 K 的大小,再按照距离大小等间隔选择部分点来获取更大的感受野。然而,这种方式会产生冗余的中间结果以及引入更大的计算量。受 U-Net (Ronneberger 等, 2015) 和 ResNet (He 等, 2016) 等层级网络结构启发,本文设计了扩张特征整合(dilated feature aggregation, DFA)模块,如图5所示。

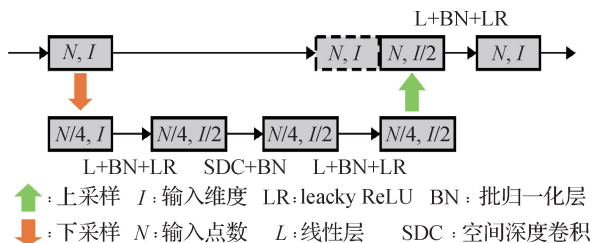


图5 扩张特征整合模块

Fig.5 Dilated feature aggregation module

DFA 模块由4部分组成,分别是线性层、下采样层、SDC 和上采样层。首先通过最大池化和随机采样进行下采样。如同 SDR 块,首先用线性层减少输入特征的维度,从而进一步减少卷积时的计算消耗和内存消耗。之后在下采样后的点云上进行卷积并用线性层对特征进行重组。接着使用最近邻插值还原到原来的尺度,并将得到的特征与输入特征并联。最后,使用一个线性层还原到输入特征的维度。

DFA 模块的核心思想是通过下采样降低卷积时产生的计算消耗和内存消耗。对比 SDR 块,DFA 模块有更大的感受野和更少的计算量。但由于上采样时可能会引入边缘噪声,该模块仅在编码器的最前端使用一次。另外,为了不引入更多的采样操作,网络在最后一阶段不使用该模块(如图1所示)。

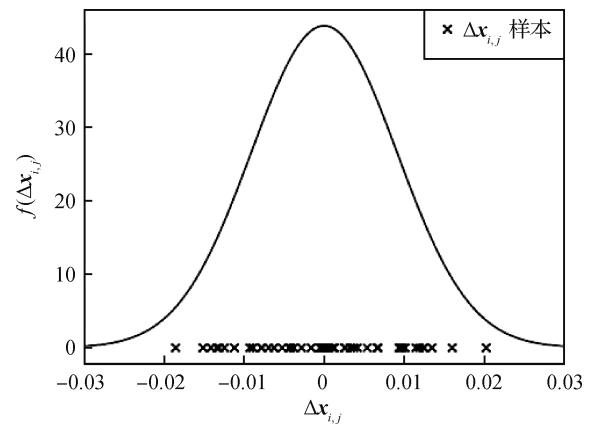
2.3 层级标准化

输入数据的数据分布会影响网络的训练。卷积核映射函数 $g(\cdot)$ 的输入数据 $\Delta p_{i,j}$ 之间的差异非常小,不利于网络学习。为了增大数据之间的差异,本文将数据在输入 $g(\cdot)$ 前进行标准化。以 X 轴为例,假设 $\Delta x_{i,j} \sim N(0, \sigma^2)$,结合 σ 修改式(1)和式(3),可以得到新的卷积公式,具体为

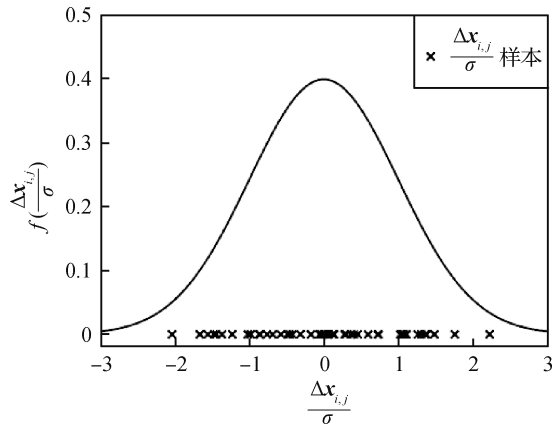
$$f'_i = \sum_{j \in \Omega_i} f_j \odot g\left(\frac{\Delta p_{i,j}}{\sigma}\right) \quad (4)$$

$$f'_i = \sum_{j \in \Omega_i} f_j \odot g\left(\frac{\Delta p'_{i,j}}{\sigma}\right) \quad (5)$$

式中, σ 通过统计数据集的数据,使用极大似然估计获得。 $\Delta x_{i,j}$ 的部分样本数据及其分布如图6所示,标准化之后的数据及分布如图7所示。

图6 $\Delta x_{i,j}$ 的分布Fig.6 The distribution of $\Delta x_{i,j}$

由于每次点云下采样后,数据方差都会产生变

图7 标准化后的 $\Delta x_{i,j}$ 分布Fig. 7 The normalized distribution of $\Delta x_{i,j}$

化,因此提出层级标准化(hierarchical normalization, HN)。即用同样的方式计算每一阶段 $\Delta p_{i,j}$ 的方差 σ , 并对其进行标准化。确保每一阶段空间卷积核的学习不受数据分布的影响。为了减小网络参数在训练过程中的学习难度,使用 HN 缓解数据分布对卷积核映射函数 $g(\cdot)$ 学习的影响。使用 HN 仅带来非常少的计算量。

3 实验及结果分析

在 S3DIS (Stanford large-scale 3D indoor space) (Armeni 等, 2017) 和 SemanticKITTI (Behley 等, 2019) 数据集上对本文方法的性能进行评估, 并与一些前沿方法进行比较。

对于图 1 中所示的第 1 个线性层, 其将原始输入特征的维度拓展到 32 维。第 1 阶段中的 SDR 块输出维度为 64, 之后每一阶段的 SDR 块的输出维度是上一阶段的两倍。为了寻找到邻域点的索引集合, 本文使用 KD-tree 算法, 为点云中每一个点找到 $K=16$ 个邻居。训练过程中, 为了能够在 GPU 上进行批处理, 在训练时采样固定数量的点进行训练。首先随机选择场景中的一个点, 然后寻找该点的 N 个邻居, N 是采样数量。所有的实验使用 Adam (Kingma 和 Ba, 2015) 优化器, 初始学习率为 0.01, 并在每个 epoch 之后衰减到先前的 0.95。为取得更好的结果, 为不同的任务设置了不同的参数。所有实验在相同机器上运行, 机器上包含 Tensorflow 1.15, i7-8700K, NVIDIA RTX。

3.1 大尺度室内场景分割实验

S3DIS 是一个大尺度室内点云场景数据集, 包

含 6 个区域的 271 个房间, 有多种不同建筑风格和不同属性的室内区域。场景中的点云分为 13 类, 每个点包含 XYZ 坐标、RGB 以及在空间中的归一化坐标。整个数据集按照区域分成 6 部分, 评估时采用 6-fold 交叉验证。本文将归一化坐标和 RGB 作为原始特征输入到网络中。使用式 (4) 所述的 SDC 操作。由于数据比较稠密, 低阶段的编码器难以捕捉到场景信息且计算和内存成本较大, 选择使用 5 阶段的 SDRNet, 每个阶段分别使用 1、3、4、6、3 个 SDR 块对特征进行编码。在训练时, 每一次随机选择 4 个场景, 采样 40 960 个点进行训练。每训练 500 次记做一个 epoch, 总共训练 100 个 epoch。使用平均交并比 (mean intersection over union, mIoU)、平均精度 (mean accuracy, mAcc) 以及总体精度 (overall accuracy, OA) 作为评估指标进行评估。

图 8 和表 1 分别展示了本文方法在 S3DIS 数据集区域 5 上的分割效果和本文方法与 PointNet (Qi 等, 2017a)、PointNet++ (Qi 等, 2017b)、DGCNN (dynamic graph CNN) (Wang 等, 2019)、PointCNN (Li 等, 2018)、ShellNet (Zhang 等, 2019)、HDGCNN (Liang 等, 2019)、PointASNL (Yan 等, 2020)、LCP (LightConvPoint) (Boulch 等, 2020)、KPCNN (Kernel point convolution) (Thomas 等, 2019)、RandLA-Net (Hu 等, 2020) 等方法的定量比较结果。可以看出, 本文方法在 mIoU、mAcc 和 OA 评估指标上都取得了较好结果。PointNet++、HDGCN 和 KPCNN 等

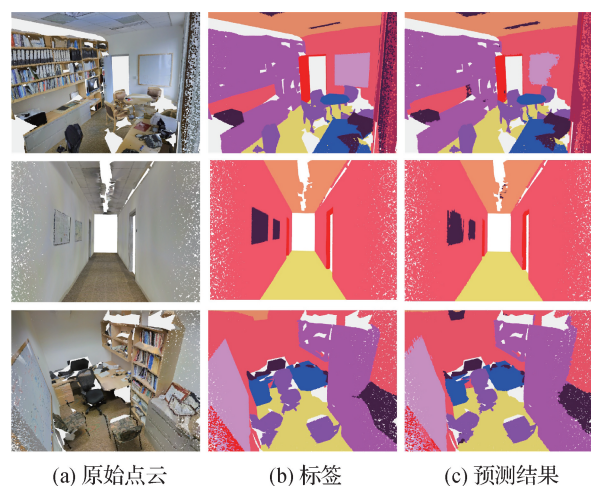


图8 本文方法在 S3DIS 数据集上的语义分割结果

Fig. 8 Semantic segmentation results on S3DIS dataset by our method ((a) original point cloud; (b) ground truth; (c) prediction)

表1 SDRNet在S3DIS上的实验结果(6-fold交叉验证)

Table 1 The results of SDRNet on S3DIS dataset by using 6-fold cross validation

方法	OA	mAcc	mIoU
PointNet	78.6	66.2	47.6
PointNet++	81.0	67.1	54.5
DGCNN	84.1	—	56.1
PointCNN	88.1	75.6	65.4
ShellNet	87.1	—	66.8
HDGCN	—	76.1	66.9
PointASNL	—	—	68.7
LCP	—	—	68.4
KPConv	—	79.0	70.6
RandLA-Net	88.0	82.0	70.0
SDRNet(本文)	88.9	82.4	71.7

注:加粗字体表示每列最优结果。“—”表示原文献未公开相关数据。

方法将场景切成小块再进行分割容易丢失场景信息,而本文方法能够一次性处理大尺度场景的点云,从而能够更好地获取场景信息。对比RandLA-Net,本文的网络结构设计得更深,具有更大的感受野,可以得到更加理想的结果。

3.2 大尺度雷达点云场景实验

SemanticKITTI (Karlsruhe Institute of Technology

and Toyota Technological Institute)是一个无人驾驶场景下的雷达点云数据集。数据通过Velodyne-64获取,每一帧都包含了完整的原始雷达点云。数据集共22组序列,仅提供在线测评。其中序列00—07和09—10为训练集,08为验证集,11—21为测试集。首先将 $p_i \in P$ 编码为 $\{\sqrt{x^2 + y^2}, z\}$,作为输入点云的特征以确保数据绕Z轴旋转特征不变。在SemanticKITTI数据集中,由于雷达点云较为稀疏,细节信息相对于稠密点云更为重要,本文选择使用4阶段的SDRNet,在每个阶段中使用2、3、6、3个SDR块对特征进行编码。相比较3.1节的设置,能够更好地捕捉细节信息。在训练过程中,每次选择6个场景,每个场景采样45 056个点进行训练。完整地在序列00—07和序列09—10上训练一轮,记做一个epoch。

通过SemanticKITTI在线单次扫描评估SDRNet的性能,并与基于点的方法(Hu等,2020;Landrieu和Simonovsky,2018;Qi等,2017a,b;Su等,2018)和基于投影的方法(Alonso等,2020a;Milioto等,2019;Wu等,2018,2019a;Xu等,2020;Zhang等,2020)分别比较mIoU、参数量以及每一类的IoU,对比结果如表2所示。其中,PointNet、PointNet++、SPG、SPLATNet(sparse lattice network)和RandLA-Net为基于点的方法,SqueezeSeg、SqueezeSegV2、SqueezeSegV3、RangeNet53++、3D-MiniNet和PolarNet为基于投影的方法。

表2 不同方法在SemanticKITTI数据集上在线单次扫描评估的定量结果

Table 2 Quantitative results of different methods on SemanticKITTI dataset from online single scan evaluation

方法	mIoU	参数量/M	IoU																		
			1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
PointNet	14.6	3.0	61.6	35.7	15.8	1.4	41.4	46.3	0.1	1.3	0.3	0.8	31.0	4.6	17.6	0.2	0.2	0.0	12.9	2.4	3.7
SPG	17.4	0.3	45.0	28.5	0.6	0.6	64.3	49.3	0.1	0.2	0.2	0.8	48.9	27.2	24.6	0.3	2.7	0.1	20.8	15.9	0.8
SPLATNet	18.4	0.8	64.6	39.1	0.4	0.0	58.3	58.2	0.0	0.0	0.0	0.0	71.1	9.9	19.3	0.0	0.0	0.0	23.1	5.6	0.0
PointNet++	20.1	6.0	72.0	41.8	18.7	5.6	62.3	53.7	0.9	1.9	0.2	0.2	46.5	13.8	30.0	0.9	1.0	0.0	16.9	6.0	8.9
RandLA-Net	53.9	1.2	90.7	73.7	60.3	20.4	86.9	94.2	40.1	26.0	25.8	38.9	81.4	61.3	66.8	49.2	48.2	7.2	56.3	49.2	47.7
SqueezeSeg	29.5	1.0	85.4	54.3	26.9	4.5	57.4	68.8	3.3	16.0	4.1	3.6	60.0	24.3	53.7	12.9	13.1	0.9	29.0	17.5	24.5
SqueezeSegV2	39.7	1.0	88.6	67.6	45.8	17.7	73.7	81.8	13.4	18.5	17.9	14.0	71.8	35.8	60.2	20.1	25.1	3.9	41.1	20.2	36.3
SqueezeSegV3	55.9	—	91.7	74.8	63.4	26.4	89.0	92.5	29.6	38.7	36.5	33.0	82.0	58.7	65.4	45.6	46.2	20.1	59.4	49.6	58.9
RangeNet53++	52.2	50.0	91.8	75.2	65.0	27.8	87.4	91.4	25.7	25.7	34.4	23.0	80.5	55.1	64.6	38.3	38.8	4.8	58.6	47.9	55.9
3D-MiniNet	55.8	4.0	91.6	74.5	64.2	25.4	89.4	90.5	28.5	42.3	42.1	29.4	82.8	60.8	66.7	47.8	44.1	14.5	60.8	48.0	56.6
PolarNet	54.3	14.0	90.8	74.4	61.7	21.7	90.0	93.8	22.9	40.3	30.1	28.5	84.0	65.5	67.8	43.2	40.2	5.6	61.3	51.8	57.5
SDRNet(本文)	59.1	1.4	91.1	75.7	64.5	23.8	90.7	95.4	43.2	42.3	46.0	41.0	80.8	63.8	65.0	61.4	55.5	11.5	61.6	54.7	54.6

注:加粗字体表示每列最优结果。“—”表示原文献未公开相关数据。各图像序号含义:1—马路;2—人行道;3—停车处;4—其他地面;5—建筑;6—轿车;7—卡车;8—自行车;9—摩托车;10—其他车辆;11—植物;12—树干;13—草地;14—行人;15—骑自行车者;16—骑摩托车者;17—围栏;18—柱子;19—交通标志。

另外,本文评估了 SDRNet 在点数量不同的情况下的推理速度。在测试时,每次仅输入一个场景的点云。通过分别在 10 K,20 K,⋯,80 K 数量下的点云进行测试。帧率与场景点数的关系如图 9 所示。数据显示,SDRNet 相比其他方法有更高的精度,而且在多个类别中达到了较优的结果。速度方面,在 50 K 点数量的情况下,可以达到实时的推理速度。结果表明,SDRNet 在雷达点云场景中可以达到较高精度,同时也能保持较快的推理速度,证明了本文方法的计算高效性。图 10 为本文方法在 SemanticKITTI 验证集上的语义分割结果。

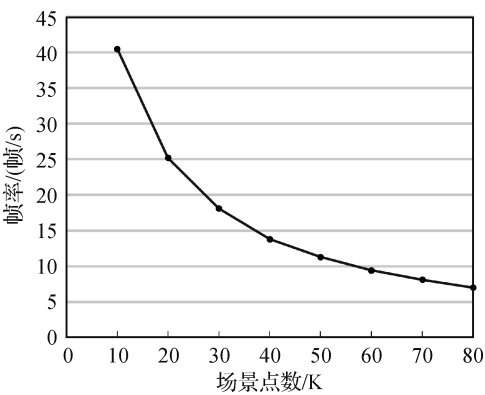


图 9 帧率和场景点数的关系

Fig. 9 The relation between FPS and number of points

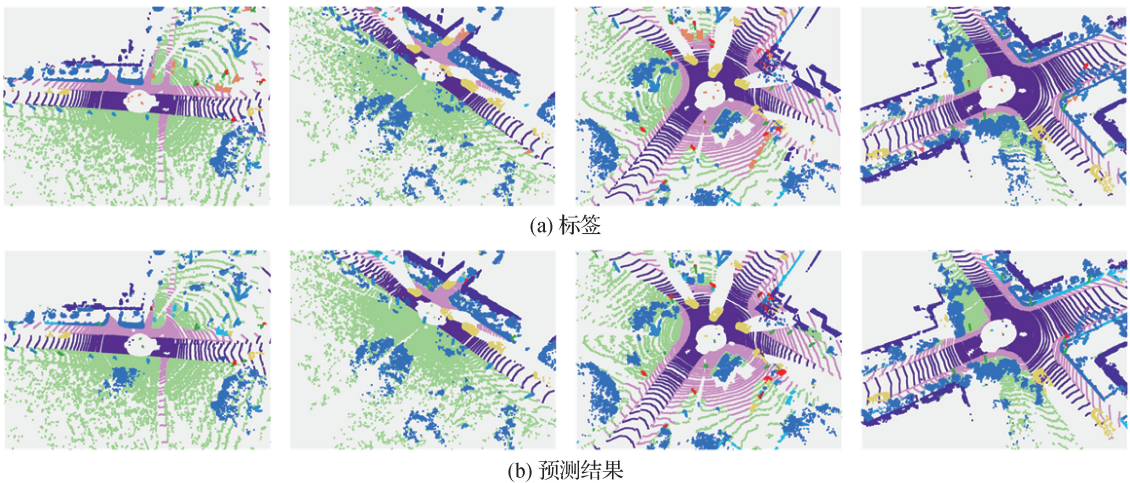


图 10 在 SemanticKITTI 验证集上的语义分割结果

Fig. 10 Semantic segmentation results on the validation set of SemanticKITTI ((a) ground truth;(b) prediction)

3.3 消融实验

为了全面验证 SDRNet 及其各部分的性能,在 S3DIS 数据集区域 5 上进行消融实验。首先移除多尺度监督(multiscale supervision, MS)、DFA 模块和层级标准化(HN)作为基准。然后逐步添加上述模块,测试不同条件下的性能。实验结果如表 3 所示。

表 3 SDRNet 在 S3DIS 区域 5 上的消融实验结果
Table 3 Ablation study of SDRNet test on S3DIS area-5

方法	mIoU /%
SDRNet	63.8
SDRNet + MS	64.3
SDRNet + MS + DFA	66.4
SDRNet + MS + DFA + HN	66.8

注:加粗字体表示最优结果。

表 3 中数据显示,多尺度监督使网络在 S3DIS 数据集区域 5 上的 mIoU 提升 0.5%。DFA 模块增大了感受野,可以提供更多的场景信息,使 mIoU 提高了 2.1%。HN 能够缓解数据分布对训练过程的影响,使 mIoU 提高了 0.4%。实验结果表明,本文提出的 SDRNet 可以有效处理点云分割任务,MS、DFA 模块和 HN 皆有助于网络提高性能。

同时,本文分别使用式(4)和式(5)作为卷积操作,在 SemanticKITTI 上进行语义分割,对比不同卷积方式下的精度。实验仅在 SemanticKITTI 验证集进行测试。采用 mIoU 和 OA 作为评估指标评估分割精度,具体结果如表 4 所示。可以看出,与式(4)相比,式(5)消除了点云绕 Z 轴旋转的影响,mIoU 和 OA 分别提高了 3.5% 和 0.7%。说明使用式(5)进行卷积操作,可以有效降低特征学习的难度,提高网络性能。

表4 在雷达点云上使用不同卷积方式的结果

Table 4 Results of using different convolution methods on LiDAR point clouds

卷积方式	mIoU	OA
式(4)	58.0	91.2
式(5)	61.5	91.9

注:加粗字体表示每列最优结果。

4 结 论

提出了一种可以在端到端的大型场景中完成语义分割任务的深度网络结构 SDRNet,可以有效地从点云中提取语义信息。同时,提出的 DFA 模块可以有效增加感受野,HN 可以缓解输入数据分布过于集中造成的影响。两者都可以提高 SDRNet 的性能,并只引入少量的计算量。本文在公开数据集上进行实验,提出的网络结构在 S3DIS 大尺度场景数据集集中的 mIoU 为 71.7%,在 SemanticKITTI 在线测试中的 mIoU 为 59.1%。与现有其他流行方法相比,本文算法在 S3DIS 和 SemanticKITTI 数据集上都有更好的表现,且在 SemanticKITTI 中,本文方法在多个类别上达到了较优结果。结果表明,本文方法在室内密集点云和稀疏激光雷达点云场景的语义分割任务中都可以达到较高的精度和相对较快的处理速度,可以有效处理大尺度 3 维点云,并有助于无人驾驶场景中的场景理解。

然而,从实验数据可以看出,本文方法在某些类别中的精度较低。其中一部分原因是数据集对应的样本较少,另一部分原因是本文方法尚未针对样本不平衡问题进行进一步优化。

在未来工作中,将针对以下内容展开研究:1) 点云语义分割的样本不平衡问题;2) 结合基于点的方法和基于投影的方法,融合不同方法中的特征,增强网络的特征提取能力;3) 改进现有模型的推理速度,减少计算消耗。

参考文献 (References)

Alonso I, Riazuelo L, Montesano L and Murillo A C. 2020a. 3D-mininet: learning a 2D representation from point clouds for fast and efficient 3D LIDAR semantic segmentation [EB/OL]. [2020-08-

12]. <https://arxiv.org/pdf/2002.10893.pdf>

Alonso I, Riazuelo L and Murillo A C. 2020b. MiniNet: an efficient semantic segmentation ConvNet for real-time robotic applications. IEEE Transactions on Robotics, 36 (4): 1340-1347 [DOI: 10.1109/TRO.2020.2974099]

Armeni I, Sax S, Zamir A R and Savarese S. 2017. Joint 2D-3D-semantic data for indoor scene understanding [EB/OL]. [2020-08-12]. <https://arxiv.org/pdf/1702.01105.pdf>

Behley J, Garbade M, Milioto A, Quenzel J, Behnke S, Stachniss C and Gall J. 2019. SemanticKITTI: a dataset for semantic scene understanding of LiDAR sequences//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision. Seoul, Korea (South): IEEE: 9297-9306 [DOI: 10.1109/ICCV.2019.00939]

Bentley J L. 1975. Multidimensional binary search trees used for associative searching. Communications of the ACM, 18 (9): 509-517 [DOI: 10.1145/361002.361007]

Boulch A, Puy G and Marlet R. 2020. FKConv: feature-kernel alignment for point cloud convolution [EB/OL]. [2020-08-12]. <https://arxiv.org/pdf/2004.04462.pdf>

Engelmann F, Kontogianni T and Leibe B. 2020. Dilated point convolutions: on the receptive field size of point convolutions on 3D point clouds [EB/OL]. [2020-08-12]. <https://arxiv.org/pdf/1907.12046.pdf>

He K M, Zhang X Y, Ren S Q and Sun J. 2016. Deep residual learning for image recognition//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE: 770-778 [DOI: 10.1109/CVPR.2016.90]

Howard A G, Zhu M L, Chen B, Kalenichenko D, Wang W J, Weyand T, Andreetto M and Adam H. 2017. MobileNets: efficient convolutional neural networks for mobile vision applications [EB/OL]. [2020-08-12]. <https://arxiv.org/pdf/1704.04861.pdf>

Hu Q Y, Yang B, Xie L H, Rosa S, Guo Y L, Wang Z H, Trigoni N and Markham A. 2020. RandLA-Net: efficient semantic segmentation of large-scale point clouds//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 11105-11114 [DOI: 10.1109/CVPR42600.2020.01112]

Iandola F N, Han S, Moskewicz M W, Ashraf K, Dally W J and Keutzer K. 2016. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5 mb model size [EB/OL]. [2020-08-12]. <https://arxiv.org/pdf/1602.07360.pdf>

Kingma D P and Ba J. 2015. Adam: a method for stochastic optimization [EB/OL]. [2020-08-12]. <https://arxiv.org/pdf/1412.6980v5.pdf>

Krizhevsky A, Sutskever I and Hinton G E. 2017. ImageNet classification with deep convolutional neural networks. Communications of the ACM, 60 (6): 84-90 [DOI: 10.1145/3065386]

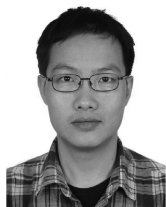
Landrieu L and Simonovsky M. 2018. Large-scale point cloud semantic segmentation with superpoint graphs//Proceedings of 2018 IEEE/

- CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA; IEEE: 4558-4567 [DOI: 10.1109/CVPR.2018.00479]
- Li Y Y, Bu R, Sun M C, Wu W, Di X H and Chen B Q. 2018. PointC-NN: convolution on X -transformed points//Proceedings of the 32nd International Conference on Neural Information Processing Systems. Montréal, Canada; Curran Associates Inc.: 828-838
- Liang Z D, Yang M, Deng L Y, Wang C X and Wang B. 2019. Hierarchical depthwise graph convolutional neural network for 3D semantic segmentation of point clouds//Proceedings of 2019 International Conference on Robotics and Automation (ICRA). Montreal, Canada; IEEE: 8152-8158 [DOI: 10.1109/ICRA.2019.8794052]
- Liu Z J, Tang H T, Lin Y J and Han S. 2019. Point-voxel CNN for efficient 3D deep learning//Proceedings of the 33rd Conference on Neural Information Processing Systems. Vancouver, Canada; Curran Associates Inc.: 965-975
- Long J, Shelhamer E and Darrell T. 2015. Fully convolutional networks for semantic segmentation//Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, USA; IEEE: 3431-3440 [DOI: 10.1109/CVPR.2015.7298965]
- Milioto A, Vizzo I, Behley J and Stachniss C. 2019. RangeNet ++: fast and accurate LiDAR semantic segmentation//Proceedings of 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Macau, China; IEEE: 4213-4220 [DOI: 10.1109/IROS40897.2019.8967762]
- Qi C R, Su H, Mo K and Guibas L J. 2017a. PointNet: deep learning on point sets for 3D classification and segmentation//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA; IEEE: 77-85 [DOI: 10.1109/CVPR.2017.16]
- Qi C R, Yi L, Su H and Guibas L J. 2017b. 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, USA; Curran Associates Inc.: 5105-5114
- Redmon J and Farhadi A. 2018. YOLOv3: an incremental improvement [EB/OL]. [2020-08-12]. <https://arxiv.org/pdf/1804.02767.pdf>
- Rethage D, Wald J, Sturm J, Navab N and Tombari F. 2018. Fully-convolutional point networks for large-scale point clouds//Proceedings of 2018 European Conference on Computer Vision. Munich, Germany; Springer: 625-640 [DOI: 10.1007/978-3-030-01225-0_37]
- Ronneberger O, Fischer P and Brox T. 2015. U-Net: convolutional networks for biomedical image segmentation//Proceedings of 2015 International Conference on Medical Image Computing and Computer-Assisted Intervention. Munich, Germany; Springer: 234-241 [DOI: 10.1007/978-3-319-24574-4_28]
- Simonyan K and Zisserman A. 2015. Very deep convolutional networks for large-scale image recognition [EB/OL]. [2020-08-12]. <https://arxiv.org/pdf/1409.1556.pdf>
- Su H, Jampani V, Sun D Q, Maji S, Kalogerakis E, Yang M H and Kautz J. 2018. SPLATNet: sparse lattice networks for point cloud processing//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA; IEEE: 2530-2539 [DOI: 10.1109/CVPR.2018.00268]
- Tchapmi L, Choy C, Armeni I, Gwak J and Savarese S. 2017. SEG-Cloud: semantic segmentation of 3D point clouds//Proceedings of 2017 International Conference on 3D Vision (3DV). Qingdao, China; IEEE: 537-547 [DOI: 10.1109/3DV.2017.00067]
- Thomas H, Qi C R, Deschaud J E, Marcotegui B, Goulette F and Guibas L. 2019. KPConv: flexible and deformable convolution for point clouds//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision. Seoul, Korea (South); IEEE: 6410-6419 [DOI: 10.1109/ICCV.2019.00651]
- Wang Y, Sun Y B, Liu Z W, Sarma S E, Bronstein M M and Solomon J M. 2019. Dynamic graph CNN for learning on point clouds. ACM Transactions on Graphics, 38 (5): #146 [DOI: 10.1145/3326362]
- Wu B C, Wan A, Yue X Y and Keutler K. 2018. SqueezeSeg: Convolutional neural nets with recurrent CRF for real-time road-object segmentation from 3D LiDAR point cloud//Proceedings of 2018 IEEE International Conference on Robotics and Automation (ICRA). Brisbane, Australia; IEEE: 1887-1893 [DOI: 10.1109/ICRA.2018.8462926]
- Wu B C, Zhou X Y, Zhao S C, Yue X Y and Keutler K. 2019a. SqueezeSegV2: improved model structure and unsupervised domain adaptation for road-object segmentation from a LiDAR point cloud//Proceedings of 2019 International Conference on Robotics and Automation (ICRA). Montreal, Canada; IEEE: 4376-4382 [DOI: 10.1109/ICRA.2019.8793495]
- Wu W X, Qi Z G and Fuxin L. 2019b. PointConv: deep convolutional networks on 3D point clouds//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA; IEEE: 9613-9622 [DOI: 10.1109/CVPR.2019.00985]
- Xu C F, Wu B C, Wang Z N, Zhan W, Vajda P, Keutler K and Tomizuka M. 2020. SqueezeSegV3: spatially-adaptive convolution for efficient point-cloud segmentation [EB/OL]. [2020-08-12]. <https://arxiv.org/pdf/2004.01803.pdf>
- Yan X, Zheng C D, Li Z, Wang S and Cui S G. 2020. PointASNL: robust point clouds processing using nonlocal neural networks with adaptive sampling//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA; IEEE: 5588-5597 [DOI: 10.1109/CVPR42600.2020.00563]
- Yu C Q, Gao C X, Wang J B, Yu G, Shen C H and Sang N. 2020. BiSeNet V2: bilateral network with guided aggregation for real-time semantic segmentation [EB/OL]. [2020-08-12]. <https://arxiv.org/pdf/2004.02147.pdf>
- Zhang Y, Zhou Z X, David P, Yue X Y, Xi Z R, Gong B Q and Forosh H. 2020. PolarNet: an improved grid representation for online LiDAR point clouds semantic segmentation//Proceedings of 2020

IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA; IEEE; 9598-9607 [DOI: 10.1109/CVPR42600.2020.00962]

Zhang Z Y, Hua B S and Yeung S K. 2019. ShellNet: efficient point cloud convolutional neural networks using concentric shells statistics//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision. Seoul, Korea (South); IEEE; 1607-1616 [DOI: 10.1109/ICCV.2019.00169]

作者简介



刘盛, 1975 年生, 男, 副教授, 主要研究方向为数字图像处理和计算机视觉。

E-mail: edliu@zjut.edu.cn



陈胜勇, 通信作者, 男, 教授, 主要研究方向为计算机视觉、图像处理和机器人智能技术。

E-mail: csy@zjut.edu.cn

黄圣跃, 男, 硕士研究生, 主要研究方向为计算机视觉和点云分割。E-mail: 2111812012@zjut.edu.cn

程豪豪, 男, 硕士研究生, 主要研究方向为计算机视觉和人体姿态估计。E-mail: 2111812002@zjut.edu.cn

沈家瑜, 男, 硕士研究生, 主要研究方向为计算机视觉和目标检测。E-mail: 2267536171@qq.com