



JOURNAL OF IMAGE AND GRAPHICS

主办: 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

中国图象学报 中国图形学报

2022
02
VOL.27

ISSN1006-8961
CN11-3758/TB



三维视觉与智能图形



第27卷第2期（总第310期）
2022年2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

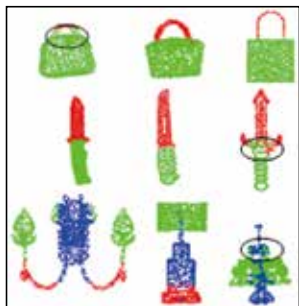
国外发行代号 M1406
国内邮发代号 82-831
国内定价 60.00元



面向本征图像分解的高质量渲染数据集与非局部卷积网络(第0404页)



高效鲁棒三维结构化重建(第0421页)



利用隐式解码器的三维模型簇协同分割(第0550页)

三维视觉与智能图形专刊简介

陈宝权, 虞晶怡, 周昆, 郭裕兰, 黄惠, 刘利刚, 刘焯斌, 徐凯, 章国锋, 周晓巍 0289

综述

基于深度学习的单目深度估计技术综述

宋巍, 朱孟飞, 张明华, 赵丹枫, 贺琪 0292

深度学习刚性点云配准前沿进展

秦红星, 刘镇涛, 谭博元 0329

三维点云配准方法研究进展

李建微, 占家旺 0349

多源融合SLAM的现状与挑战

王金科, 左星星, 赵祥瑞, 吕佳俊, 刘勇 0368

深度学习单目深度估计研究进展

罗会兰, 周逸风 0390

数据集论文

面向本征图像分解的高质量渲染数据集与非局部卷积网络

王玉洁, 樊庆楠, 李坤, 陈冬冬, 杨敬钰, 卢健智, Dani Lischinski, 陈宝权 0404

深度估计与三维重建

高效鲁棒三维结构化重建

潘珊珊, 吕佳辉, 方昊, 黄惠 0421

结合LiDAR与RGB数据构建稠密深度图的多阶段指导网络

贾迪, 王子滔, 李宇扬, 金志杨, 刘泽洋, 吴思 0435

多尺度相似性迭代查找的可靠双目视差估计

晏敏, 王军政, 李静 0447

显微光学模糊图像的景物深度提取及应用

苗国超, 魏阳杰, 侯威翰 0461

融合注意力机制和多层U-Net的多视图立体重建

刘会杰, 柏正尧, 程威, 李俊杰, 许祝 0475

单目相机轨迹的真实尺度恢复

刘思博, 房立金 0486

三维形状分析

基于显著性图的点云替换对抗攻击

刘复昌, 南博, 缪永伟 0500

边收缩池化的网格变分自编码器

袁宇杰, 来煜坤, 杨洁, 段琦, 傅红波, 高林 0511

雅可比矩阵引导的无翻转体映射生成

徐茂峰, 刘利刚 0525

嵌入Transformer结构的多尺度点云补全

刘心溥, 马燕新, 许可, 万建伟, 郭裕兰 0538

三维点云分割

利用隐式解码器的三维模型簇协同分割

杨军, 张敏敏 0550

面向形状特征的多维度多层级点云分析

徐嘉利, 方志军, 伍世虞 0562

多特征融合与几何卷积的机载LiDAR点云地物分类

戴莫凡, 邢帅, 徐青, 李鹏程, 陈坤 0574

图像视频分析

聚合细粒度特征的深度注意力自动截图

方玉明, 钟裕, 鄢杰斌, 刘丽霞 0586

单幅人脸图像的全景纹理图生成方法

刘洋, 樊养余, 郭哲, 吕国云, 刘诗雅 0602

时空联合视差优化的立体视频重定向

金康俊, 柴雄力, 邵枫 0614

形状的全尺度可视化表示与识别

闵睿朋, 李一凡, 黄瑶, 杨剑宇, 钟宝江 0628

结合掩码定位和漏斗网络的6D姿态估计

李冬冬, 郑河荣, 刘复昌, 潘翔 0642

CONTENTS

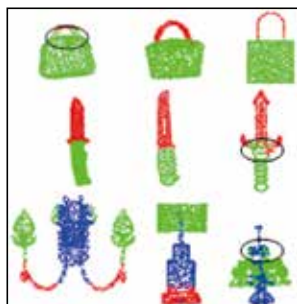
JOURNAL OF IMAGE AND GRAPHICS



High quality rendered dataset and non-local graph convolutional network for intrinsic image decomposition(P0404)



Efficient and robust 3D structure-aware reconstruction(P0421)



Co-segmentation of 3D shape clusters based on implicit decoder(P0550)

Review

A review of monocular depth estimation techniques based on deep learning	
Song Wei, Zhu Mengfei, Zhang Minghua, Zhao Danfeng, He Qi	0292
Review on deep learning rigid point cloud registration	
Qin Hongxing, Liu Zhenhao, Tan Boyuan	0329
Review on 3D point cloud registration method	
Li Jianwei, Zhan Jiawang	0349
Review of multi-source fusion SLAM: current status and challenges	
Wang Jinke, Zuo Xingxing, Zhao Xiangrui, Lyu Jiajun, Liu Yong	0368
Review of monocular depth estimation based on deep learning	
Luo Huilan, Zhou Yifeng	0390

Dataset

High quality rendered dataset and non-local graph convolutional network for intrinsic image decomposition	
Wang Yujie, Fan Qingnan, Li Kun, Chen Dongdong, Yang Jingyu, Lu Jianzhi, Dani Lischinski, Chen Baoquan	0404

Depth Estimation & 3D Reconstruction

Efficient and robust 3D structure-aware reconstruction	
Pan Shanshan, Lyu Jiahui, Fang Hao, Huang Hui	0421
Multi-stage guidance network for constructing dense depth map based on LiDAR and RGB data	
Jia Di, Wang Zitao, Li Yuyang, Jin Zhiyang, Liu Zeyang, Wu Si	0435
Reliable binocular disparity estimation based on multi-scale similarity recursive search	
Yan Min, Wang Junzheng, Li Jing	0447
Scene depth extraction and application of microscopic optical blur image	
Miao Guochao, Wei Yangjie, Hou Weihai	0461
Fusion attention mechanism and multilayer U-Net for multiview stereo	
Liu Huijie, Bai Zhengyao, Cheng Wei, Li Junjie, Xu Zhu	0475
Monocular camera trajectory recovery with real scale	
Liu Sibao, Fang Lijin	0486

3D Shape Analysis

Point cloud replacement adversarial attack based on saliency map	
Liu Fuchang, Nan Bo, Miao Yongwei	0500
Mesh variational auto-encoders with edge contraction pooling	
Yuan Yujie, Lai Yukun, Yang Jie, Duan Qi, Fu Hongbo, Gao Lin	0511
Generation of foldover-free volumetric mapping guided by Jacobian matrix	
Xu Maofeng, Liu Ligang	0525
Multi-scale Transformer based point cloud completion network	
Liu Xinpu, Ma Yanxin, Xu Ke, Wan Jianwei, Guo Yulan	0538

3D Point Cloud Segmentation

Co-segmentation of 3D shape clusters based on implicit decoder	
Yang Jun, Zhang Minmin	0550
Multi-dimensional multi-layer point cloud analysis for shape features	
Xu Jiali, Fang Zhijun, Wu Shiqian	0562
Semantic segmentation of airborne LiDAR point cloud based on multi-feature fusion and geometric convolution	
Dai Mofan, Xing Shuai, Xu Qing, Li Pengcheng, Chen Kun	0574

Image & Video Analysis

Deep attention guided image cropping with fine-grained feature aggregation	
Fang Yuming, Zhong Yu, Yan Jiebin, Liu Lixia	0586
Solo face image-based panoramic texture map generation	
Liu Yang, Fan Yangyu, Guo Zhe, Lyu Guoyun, Liu Shiya	0602
Optimizing spatiotemporal disparities for stereoscopic video retargeting	
Jin Kangjun, Chai Xiongli, Shao Feng	0614
Visualized all-scale shape representation and recognition	
Min Ruipeng, Li Yifan, Huang Yao, Yang Jianyu, Zhong Baojiang	0628
6D pose estimation based on mask location and hourglass network	
Li Dongdong, Zheng Herong, Liu Fuchang, Pan Xiang	0642

中图法分类号: TP391 文献标识码: A 文章编号: 1006-8961(2022)02-0562-12

论文引用格式: Xu J L, Fang Z J and Wu S Q. 2022. Multi-dimensional multi-layer point cloud analysis for shape features. Journal of Image and Graphics, 27(02): 0562-0573 (徐嘉利, 方志军, 伍世虔. 2022. 面向形状特征的多维度多层级点云分析. 中国图象图形学报, 27(02): 0562-0573) [DOI: 10.11834/jig.210592]

面向形状特征的多维度多层级点云分析

徐嘉利¹, 方志军^{1*}, 伍世虔²

1. 上海工程技术大学电子电气工程学院, 上海 201620; 2. 武汉科技大学机械自动化学院, 武汉 430081

摘要: **目的** 3维点云是编码几何信息的主要数据结构,与2维视觉数据不同的是,点云中隐藏了3维物体中重要的形状特征。为更好地从无序的点云中挖掘形状特征,本文提出一种能够端到端且鲁棒地处理点云数据的多维度多层级神经网络(multi-dimensional multi-layer neural network, MM-Net)。**方法** 多维度特征修正与融合(multi-dimensional feature correction and fusion module, MDCF)模块从多个维度自适应地修正局部特征和逐点特征,并将其整合至高维空间以获得丰富的区域形状。另一方面,多层级特征衔接(multi-layer feature articulation module, MLFA)模块利用多个层级间的远程依赖关系,推理得到网络所需的全局形状。此外设计了两端分别应用于点云分类与分割任务的网络结构 MM-Net-C(multi-dimensional multi-layer feature classification network)和 MM-Net-S(multi-dimensional multi-layer feature segmentation network)。**结果** 在公开的 ModelNet40 数据集与 ShapeNet 数据集上进行测试,并与多种方法进行比较。在 ModelNet40 数据集中,MM-Net-C 的分类精度较 PointNet++ 和 DGCNN(dynamic graph convolutional neural network)方法分别提高了 2.2% 和 1.9%;在 ShapeNet 数据集中,MM-Net-S 的分割精度较 ELM(extreme learning machine)和 A-CNN(annularly convolutional neural networks)方法分别提高了 1.2% 和 0.4%。此外,在 ModelNet40 数据集上的消融实验验证了多维度多层级神经网络(MM-Net)架构的可靠性,消融实验的结果也表明了多维度特征修正与融合(MDCF)模块和多层级特征衔接(MLFA)模块设计的必要性。**结论** 本文提出的多维度多层级神经网络(MM-Net)在分类与分割任务中取得了优秀的性能。

关键词: 3维点云;点云分类与分割;深度学习;形状特征;多维度特征;多层级特征

Multi-dimensional multi-layer point cloud analysis for shape features

Xu Jiali¹, Fang Zhijun^{1*}, Wu Shiqian²

1. School of Electronic and Electrical Engineering, Shanghai University of Engineering Science, Shanghai 201620, China;

2. College of Machinery and Automation, Wuhan University of Science and Technology, Wuhan 430081, China

Abstract: **Objective** With the widespread use of depth cameras and 3D scanning equipment, 3D data with point clouds as the main structure have become more readily available to people. As a result, 3D point clouds are widely used in practical applications such as self-driving cars, location recognition, robot localization, and remote sensing. In recent years, the great success of convolutional neural networks (CNNs) has changed the landscape of 2D computer vision. However, CNNs cannot directly process unstructured data such as point clouds due to the disorderly, irregular characteristics of 3D point clouds. Therefore, mine shape features from disordered point clouds have become a viable research direction in point cloud

收稿日期: 2021-07-23; 修回日期: 2021-09-12; 预印本日期: 2021-09-19

* 通信作者: 方志军 zjfang@sues.edu.cn

基金项目: 国家自然科学基金项目(61772328, U2033218)

Supported by: National Natural Science Foundation of China (61772328, U2033218)

analysis. **Method** An end-to-end multidimensional multilayer neural network (MM-Net), which can directly process point cloud data, is presented in this paper. The multi-dimensional feature correction and fusion (MDCF) module can correct local features in different dimensions rationally. First, the local area division unit, using farthest point sampling and ball query, constructs local areas at different radii from which the 10D geometric relations and local features required are obtained for the module. Inspired by related research, the module uses geometric relations to modify the point-wise features, enhance the interaction between points, and encode useful local features, which are supplemented by point-wise features. Finally, the shape features of different region ranges are fused and mapped to a higher dimensional space. At the same time, the multi-layer feature articulation (MLFA) module focuses on integrating the contextual relationships between local regions to extract global features. In particular, these local regions are seen as distinct nodes, and global features are acquired by using convolution and jump fusion. The MLFA module uses the long-range dependencies between multiple layers to reason about the global shape required for the network. Furthermore, two network architectures (multidimensional multi-layer feature classification network (MM-Net-C) and multidimensional multi-layer feature segmentation network (MM-Net-S)) for point cloud classification and segmentation tasks are designed in this paper. In detail, MM-Net-C goes through three tandem MDCF modules with three layers of interlinked local shape features. The global features are then obtained by connecting and integrating the correlations between each local region through the MLFA module. In MM-Net-S, after processing by the MLFA module, the object data are encoded global feature vector with 1 024 dimensions. Then, the features are summed to obtain shapes that fuse local and global information, so that they are linked to the labels of the objects (e. g. , motorbikes, cars). This process is followed by feature propagation, where successive up sampling operations are performed to recover the details in the original object data and to obtain a robust point-wise vector. Finally, the outputs of the different feature propagation layers are integrated and fed into the convolution operation. The features are transformed to obtain an accurate prediction of each point cloud within the object. **Result** The method in this paper is adequately tested on the publicly available ModelNet40 dataset and ShapeNet dataset. The experimental results are compared with various methods. In the ModelNet40 dataset, MM-Net-C is compared with several pnt-based (input point cloud coordinates only), such as dynamic graph convolutional neural network (DGCNN) (92.2%) with 1.9% accuracy improvement and relation-shape convolutional neural network (RS-CNN) (93.6%) with 0.5% accuracy improvement. MM-Net-C is also compared with several pnt-nor (coordinates and normal vectors of the input point cloud) based: point attention transformers (PAT) (91.7%) improves accuracy by 2.4%; PointConv (92.5%) improves accuracy by 1.6%; PointASNL (93.2%) improves accuracy by 0.9%. Even when several studies input more points for training, MM-Net-C still outperforms them. For example, PointNet++ (5 k, 91.9%) improves accuracy by 2.2%, and self-organizing network (SO-Net) (5 k, 93.4%) improves accuracy by 0.7%. In addition, MM-Net-C achieves higher accuracy rates than other studies with less complexity. For example, compared with PointCNN (8.20 M, 91.7%), MM-Net-C has less than one-eighth of the number of parameters while the accuracy rate is increased by 2.4%. Compared with RS-CNN (1.41 M, 93.6%), MM-Net-C has 0.33 M fewer parameters while the accuracy rate is increased by 0.5%. In the ShapeNet dataset, MM-Net-S compared with DGCNN (85.1%), the accuracy is improved by 1.4%; compared with shape-oriented convolutional neural network (SO-CNN) (85.7%), the accuracy is improved by 0.8%; and compared with annularly convolutional neural networks (A-CNN) (86.1%), the accuracy is improved by 0.4%. Ablation experiments are also conducted on the ModelNet40 dataset to confirm the effectiveness of the MM-Net architecture. The ablation experiments results validate the need for the MDCF module and MLFA module design. The results further confirm that MDCF module, which uses rich point-wise features modified and fused with potential local features, can effectively improve the network's mining of shape information within a local region. By contrast, the MLFA module captures contextual information at the global scale and reinforces the long-range dependency links that exist between different layers, effectively enhancing the robustness of the model in dealing with complex shapes. Ablation experiments are conducted on whether the MDCF needs to be designed with different dimensions. The experimental results demonstrate that MM-Net performs better than RS-CNN for the same dimensionality. **Conclusion** In this paper, an MM-Net with MDCF module and MLFA module as core components is proposed. After conducting sufficient experiments, thorough comparisons and verifying MM-Net, a higher correct rate is achieved with the advantage of fewer parameters.

Key words: three-dimensional point cloud; point cloud classification and segmentation; deep learning; shape features; multi-dimensional feature; multi-layer feature

0 引言

随着深度激光、深度相机和 3 维扫描设备的广泛应用,以点云作为主要结构的 3 维数据变得更加容易获取。3 维点云在自动驾驶汽车 (Shi 等, 2019)、位置识别 (Liu 等, 2019d)、机器人定位与导航 (Zhu 等, 2017) 和遥感 (Qi 等, 2018) 等新兴领域得到了广泛应用。近年来,卷积神经网络 (convolutional neural networks, CNN) (Zeiler 和 Fergus, 2014) 的巨大成功改变了 2 维计算机视觉的格局,但是由于 3 维点云具有无序性和不规则性的特点,卷积神经网络无法直接处理点云这样的非结构化数据。如何从点云中鲁棒地提取网络所需的形状特征进行点云分析仍然是一个尚待解决的问题。

PointNet (Qi 等, 2017a) 是首个应用深度学习直接处理不规则点云的方法,但是忽视了对局部区域的描述。PointNet++ (Qi 等, 2017b) 是 PointNet 的升级版,但是仍然未有效地探索点之间的关系。DGCNN (dynamic graph convolutional neural network) (Wang 等, 2019) 使用一种新的卷积 EdgeConv 侧重点云的局部几何特征,但无法兼顾局部邻域之间的全局联系。Point2Sequence (Liu 等, 2019a) 和 PointANSL (Yan 等, 2020) 引入注意力机制专注于各局部区域之间的关系,提高了模型训练的鲁棒性。Dense-Point (Liu 等, 2019b) 引用 2D 图像领域中 DenseNet (Huang 等, 2017) 的思想,在同一层中重复地聚合多级语义输出以捕获该层中的上下文信息。RS-CNN (relation-shape convolutional neural network) (Liu 等, 2019c) 通过球查询构建不同半径的局部邻域,并选择 10 维的几何关系应用于局部形状,但是没有考虑点云的逐点特征和全局特征,而逐点特征和全局特征同样是准确处理点云数据的关键。

为了解决当前研究存在的一些问题,提升神经网络在点云分析中的性能,本文设计了多维度特征修正与融合模块 (multi-dimensional feature correction and fusion module, MDCF) 和多层级特征衔接模块 (multi-layer feature articulation module, MLFA)。MDCF 模块能够在不同维度下合理地修正、融合逐

点特征和局部特征。在局部区域划分单元 (local area division unit, LAD), 利用最远点采样 (farthest point sampling, FPS) (Kamoussi 等, 2016) 和球查询 (ball query) (Qi 等, 2017a) 的方法构建不同半径下的局部区域,从中获得模块所需的 10 维几何关系和局部特征。受注意力机制 (Wang 等, 2018b) 的启发, MDCF 模块利用几何关系对局部特征进行修正,增强点间的交互能力,并编码得到合理的逐点特征,同时辅以局部特征对其进行充实,最后融合得到不同区域范围的形状特征,并映射到高维空间。此外, MLFA 模块侧重于整合局部区域之间的上下文关系来提取全局形状,强化了层次之间的远程依赖联系,能够利用丰富的层间上下文信息提高形状特征的抽象性,同时降低了上游存在的错误对下游点云分析造成的影响,缓解了卷积网络存在的梯度消失缺陷,提高了模型的鲁棒性。特别地,将这些局部信息看做不同的节点,利用卷积和跳跃融合的方式获取并丰富全局特征,以此提高网络学习全局形状的表达能力。本文结合 MDCF 模块和 MLFA 模块的优势,设计了两种不同的网络架构 MM-Net-C (multi-dimensional multi-layer feature classification network) 和 MM-Net-S (multi-dimensional multi-layer feature segmentation network) 端到端地、鲁棒地处理点云分类与分割任务。

1 多维度多层级神经网络 (MM-Net)

1.1 多维度特征修正与融合模块 (MDCF)

卷积运算依赖于规则的网格结构,然而由于点云数据在 3 维空间中分布不规则的特性,无法直接应用卷积运算。

为了充分得到某一局部空间中的点与点之间存在的关联性,首先利用局部区域划分单元 (LAD) 将点云划分为不同半径大小的局部区域,如图 1 所示。

给定输入无序的点云 $P = \{p_1, p_2, p_3, \dots, p_N\}$, 其中 $p_n \in \mathbf{R}^C$ 是输入点的坐标向量, N 是点的个数, C 是每一点的坐标维数。对于输入的 N 个点,选择最远点采样的方法获得 h 个质心点 p_i , 得到质心点集 $E = \{p_\alpha, \dots, p_h\}$ 。最远点采样算法的具体步

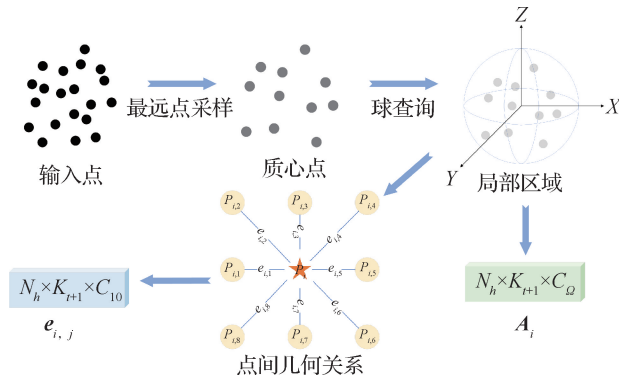


图 1 局部区域划分单元(LAD)细节图

Fig. 1 Details of local area division unit

骤如下:

输入: 无序点集 $P = \{p_1, p_2, p_3, \dots, p_N\}$, 设定质心点个数 h 。

输出: 质心点点集 $E = \{p_\alpha, \dots, p_h\}$ 。

1) 从无序点集 P 中, 随机选取点 p_α 作为初始点, 并移动至质心点点集 E 内。

2) 若质心点点集 $E = \{p_\alpha, \dots, p_h\}$ 内点的数量小于 h , 则首先计算无序点集 P 中剩余的点 p_n 与质心点点集 E 内每一点之间的欧氏距离 $l_n = \{l_\alpha, \dots, l_i\}$, 然后在 l_n 中选取最小值作为该点的距离值, 得到 P 中剩余各点的距离值集合 L , 最后在 L 中选取最大距离值对应的点 p_β , 并移动至质心点点集 E 。

接着, 通过球查询方法找到 t 个点作为半径 r 内每一质心点 p_i 的邻点, 构建球形的局部区域。在不同维度都设定不同的 t 值和 r 值, 得到不同条件下的局部点集 $A_i = \{p_i, p_{i,1}, p_{i,2}, p_{i,3}, \dots, p_{i,t}\}$, $d_{i,j} < 2r$, $A_i \in \mathbf{R}^{N_h \times K_{t+1} \times C_\Omega}$, 其中 p_i 是质心点, $p_{i,t}$ 是质心点的 t 个邻点, $d_{i,j}$ 是质心点与其邻点之间的欧氏距离。球查询算法的具体步骤如下:

输入: 设定球形区域半径 r 、球形区域范围内邻点个数 t ; 无序点集 $P = \{p_1, p_2, \dots, p_N\}$; 质心点点集 $E = \{p_\alpha, \dots, p_h\}$ 。

输出: 局部点集 $A_i = \{p_i, p_{i,1}, p_{i,2}, p_{i,3}, \dots, p_{i,t}\}$ 。

1) 迭代选取质心点点集 $E = \{p_\alpha, \dots, p_h\}$ 内的每一个点 p_i , 并存入以该质心点为中心的局部点集 A_i 。

2) 计算 p_i 与无序点集 P 中所有点之间的欧氏距离, 并标记大于 $2r$ 的距离值。

3) 对未标记的距离值进行升序。

4) 在升序后未标记的距离值中选取 t 个, 并将 t 个距离值对应的点存入局部点集 A_i 。

点的几何关系是通过质心点与邻点的位置信息构成单一各点的初始特征, 同时点与点之间的空间布局联系(点的分布密度及点间相互作用关系)反映了区域范围内的局部特征。本文受 RS-CNN 启发, 编码几何关系(见图 1), 为后续鲁棒地学习局部形状打下坚实基础。

另一方面, 该操作也能够处理好点存在的刚性变换与排列不变性的前提下, 为 MDCF 模块挖掘区域内的形状信息。这里, 点间的几何关系为 $e_{i,j}$, 即

$$e_{i,j} = [p_i, p_{i,j}, |p_i - p_{i,j}|, \|p_i, p_{i,j}\|_{ed}] \quad (1)$$

式中, $p_i \in \mathbf{R}^{N_h \times K_{t+1} \times 3}$ 是质心点的坐标; $p_{i,j} \in \mathbf{R}^{N_h \times K_{t+1} \times 3}$ 是邻点的坐标, $|p_i - p_{i,j}| \in \mathbf{R}^{N_h \times K_{t+1} \times 3}$ 是质心点与邻点之间的相对位置, $\|p_i, p_{i,j}\|_{ed} \in \mathbf{R}^{N_h \times K_{t+1} \times 1}$ 是两点之间的欧氏距离, $[\cdot]$ 是连接函数。

以第 1 维为例, 如图 2 所示, 在 LAD 单元构建的局部区域中, 将局部特征定义为 $A_1 \in \mathbf{R}^{N_\alpha \times K_\beta \times C_\gamma}$, 将点与点之间的几何关系定义为 $e_1 \in \mathbf{R}^{N_\alpha \times K_\beta \times 10}$ 。同时, 将原始的局部特征与 10 维的几何先验知识有机聚合起来, 得到蕴含有丰富点间相关性的逐点特征 $B_1 \in \mathbf{R}^{N_\alpha \times K_\beta \times C_\gamma}$ 。具体操作为

$$B_1 = [MLP_1(e_1)] \cdot A_1 \quad (2)$$

式中, MLP_1 是多层感知机 (multi-layer perceptron, MLP), “ $\cdot$ ” 是点乘函数。

多层感知机用于整理点间的几何向量, 并将其编码至特定维度, 使之能够更方便与局部特征 A_1 进行点乘运算。考虑到在同一维度下, 逐点特征与局部特征对于区域内形状信息的提取, 所做出的贡献是不一样的, 例如, 当质心点与邻点之间分布较为分散, 那么模型应该更侧重于局部特征。倘若部分邻点与质心点之间的空间布局较为紧凑, 则应该更倾向于逐点特征的选择。本文受注意力机制启发, 使网络自适应地选择合适的特征信息, 表示当前局部区域内的形状关系。具体实现为

$$[a_1, b_1] = \sigma(MLP_2[A_1, B_1]) \quad (3)$$

$$H_1 = a_1 \otimes A_1 \quad (4)$$

$$T_1 = b_1 \otimes B_1 \quad (5)$$

式中, $a_1 \in \mathbf{R}^{N_\alpha \times K_\beta \times 1}$ 和 $b_1 \in \mathbf{R}^{N_\alpha \times K_\beta \times 1}$ 分别是局部特征和逐点特征对应的归一化权重, σ 是激活函数

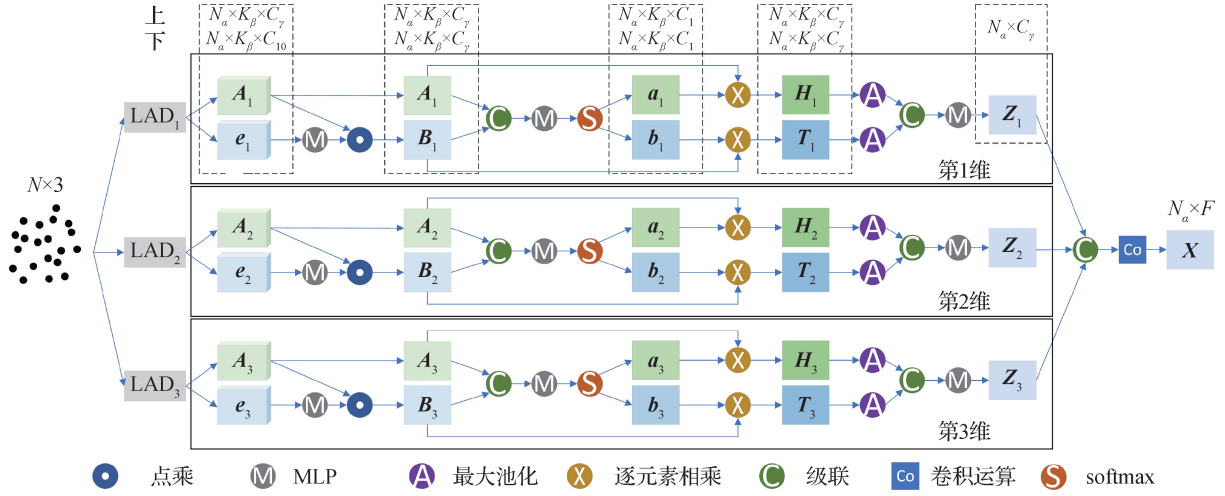


图2 多维度特征修正与融合模块(MDCF)的细节图

Fig. 2 Details of the multi-dimensional feature correction and fusion module (MDCF)

softmax, $H_1 \in \mathbf{R}^{N_a \times K_\beta \times C_\gamma}$ 和 $T_1 \in \mathbf{R}^{N_a \times K_\beta \times C_\gamma}$ 分别是通过自适应修正后所得的局部特征与逐点特征, $\otimes$ 是逐元素相乘。为了使 MDCF 模块能够有效选择出所需的特征强度, 首先将原始的局部特征与逐点特征进行融合, 并使用多层感知机将其通道数处理至特定数。随后利用 softmax 激活函数, 得到这一维度下局部区域内每个点的对应强度, 这个强度可以理解为一质点与其邻点在局部特征与逐点特征两条条件下的自适应权值。利用 a_1 和 b_1 这两个修正因子, 使模块学习到相对于不同特征的重要程度, 抑制无关因素的影响。随后将其逐元素与原始的局部特征和逐点特征相乘, 得到经过网络自我选择同时也是模型所需要的特征信息。上述操作强化了模型自我学习局部区域内形状关系的能力, 也提高了网络对点间关键信息的提取能力。模块将所得修正后的特征通过对称函数融合并处理, 得到该维度下的结果 $Z_1 \in \mathbf{R}^{N_a \times C_\gamma}$ 。具体实现为

$$Z_1 = MLP_3 \left[\text{MaxP} \sum_{j=1}^{\beta} (H_1), \text{MaxP} \sum_{j=1}^{\beta} (T_1) \right] \quad (6)$$

式中, MaxP 是对称函数 MaxPooling, 其分别将逐点特征与局部特征聚合到质点, 获得对局部区域表达能力更强的形状信息。接着将上述两者级联, 并将结果输入到多层感知机中, 使其有机地融合在一起, 输出该维度下、该局部区域范围内最为合适的形状关系。

多维度特征修正与融合模块 (MDCF) 分别从 3 个不同维度, 即 3 个不同尺度半径、邻点个数构建的球形局部区域捕捉不同维度下最有效的形状信息 $X \in \mathbf{R}^{N_a \times F}$ 。具体操作为

$$X = \text{Conv}[Z_1, Z_2, Z_3] \quad (7)$$

式中, Conv 是卷积运算, 其将融合了不同维度下的形状信息进行二次加工, 强化了模型对范围形状的理解能力, 缓解了点云数据中存在的点分布不均匀、密度不均匀对特征学习造成的不良影响, 同时提高了模型对局部差异性的理解能力, 弥补了模型短视的不足。例如, 当网络对飞机数据集进行分类或分割任务时, 倘若使用单一维度, 即单一尺度下的球形区域, 模型的可视范围也许只能够聚焦飞机的某一片面范围如机头或机翼, 因火箭头部形状与飞机头部形状相似、椅子脚部形状与机翼形状相似, 便会将其错误分类为火箭或椅子。而使用多维度设计, 可以将局部视野范围交叉、扩大, 这样提升了模型对形状的甄别能力, 进而提高了网络的准确性和鲁棒性。

1.2 多层次特征衔接模块 (MLFA)

一些经典研究 (Qi 等, 2017a; Wang 等, 2019) 致力于如何在局部区域内捕捉几何关系, 但忽略了各局部区域之间存在的相互依赖联系。本文受 2 维图像领域和其他 3 维点云的研究启发, 建模不同局部区域之间的形状信息, 并通过跳跃连接构建局部邻域之间的上下文联系, 捕捉不同层级之间存在的关联性信息。虽然多维度特征修正与融合模块

(MDCF)侧重于对局部区域形状的理解,但忽视了从整体角度对全局形状分析。而在多层级特征衔接模块(MLFA)中,捕捉了远距离区域之间的全局描述关系,如图3所示。

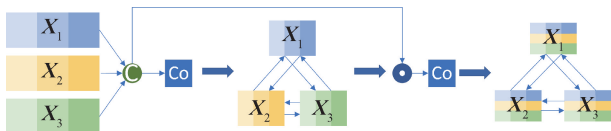


图3 多层级特征衔接模块(MLFA)的细节图

Fig. 3 Detail diagram of the multi-layer feature articulation module (MLFA)

多层级特征衔接模块(MLFA)接收不同层级多维度特征修正与融合模块(MDCF)处理输出的局部形状,这些层级的局部形状之间并不是孤立、没有联系的。例如,对于飞机点云数据,第1层级MDCF模块输出的局部形状也许更专注于机头的理解,而第2层级的MDCF模块却更加重视机翼的学习。倘若单独处理两者,必然会降低整体点云形状的分析能力,因此需要找出两者之间存在的关联性,进而从基于全局视野的角度捕获所有局部区域之间的上下文信息。整个多层级特征衔接模块(MLFA)的具体操作为

$$Y = \text{Conv}_2 \{ (\text{Conv}_1 [X_1, X_2, X_3]) \cdot [X_1, X_2, X_3] \} \quad (8)$$

式中, $Y \in \mathbf{R}^{N_h \times F_4}$ 是多层级特征衔接模块(MLFA)的

输出,是经过推理后具有全局描述的特征。 $X_1 \in \mathbf{R}^{N_h \times F_1}$ 、 $X_2 \in \mathbf{R}^{N_h \times F_2}$ 和 $X_3 \in \mathbf{R}^{N_h \times F_3}$ 分别为不同层级下MDCF模块输出的局部形状,首先将其融合在一起,通过 Conv_1 运算建立不同层级之间存在的上下文联系,并将其加权至原始的局部形状中,以实现局部区域之间的密集连接。随后利用 Conv_2 更新各局部区域的形状信息,实现模型从局部到全局的学习。

1.3 多维度多层级神经网络架构(MM-Net)

1.3.1 分类网络架构(MM-Net-C)

面向点云分类任务,完整的多维度多层级神经网络分类架构由前端与后端组成,如图4所示。

在前端中,输入的点云数据是 $N \times 3$ 格式,其中 N 是数据中点云的个数,每一个点云都具有初始的3维原始特征。随后通过3层串联的多维度特征修正与融合模块(MDCF),得到输出3层相互联系的局部形状特征 $\mathbf{R}^{N_h \times F_i}$,其中 F 是特征数。接着通过多层级特征衔接模块(MLFA)对每个局部区域之间存在的关联性进行连接并整合得到全局形状。

在后端中,将全局形状和局部形状结合在一起,这样既能够保留局部特征,同时也丰富了全局特征。随后使用 MaxPooling 实现维度变换 $\mathbf{R}^{N_h \times (F_1 + F_2 + F_3 + F_4)} \rightarrow \mathbf{R}^{1 \times (F_1 + F_2 + F_3 + F_4)}$, 并得到 1 024 维的点云形状向量。最后应用 3 个全连接层 $1\,024 \rightarrow 512 \rightarrow 256 \rightarrow N_{\text{Class}}$, 转换并得到分类任务的概率图。

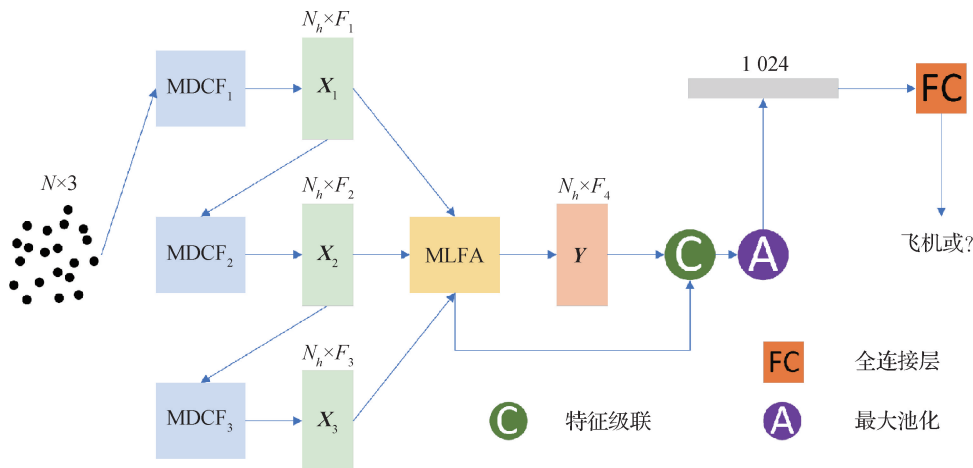


图4 多维度多层级特征分类网络的架构图

Fig. 4 Architecture of a multi-dimensional multi-layer features classification network

1.3.2 分割网络架构(MM-Net-S)

完整的分割网络也由前端和后端两部分组成,如图5所示。在前端中,与分类网络相比,使用单个对象作为分割网络的输入,同时在多维度特征修正与融

合模块(MDCF)中,选择更大的半径以及更多的质心点与邻点来挖掘局部形状。在经过多层级特征衔接模块(MLFA)处理后,对象数据编码为 $\mathbf{R}^{N_h \times F_4}$ 维度的全局特征。在后端部分,经过特征级联,得到融合了

局部和全局信息的形状,维度为 $\mathbf{R}^{N_h \times (F_1 + F_2 + F_3 + F_4)}$,使其与对象(如摩托、轿车等)的标签相连。然后进行特征传播(Qi 等,2017a),对其进行连续的上采样操作,以此

恢复原始对象数据中的细节,同时得到鲁棒的逐点向量。最后将不同特征传播层的输出整合并馈送至卷积运算中,通过变换特征获得对象内每个点云的准确预测。

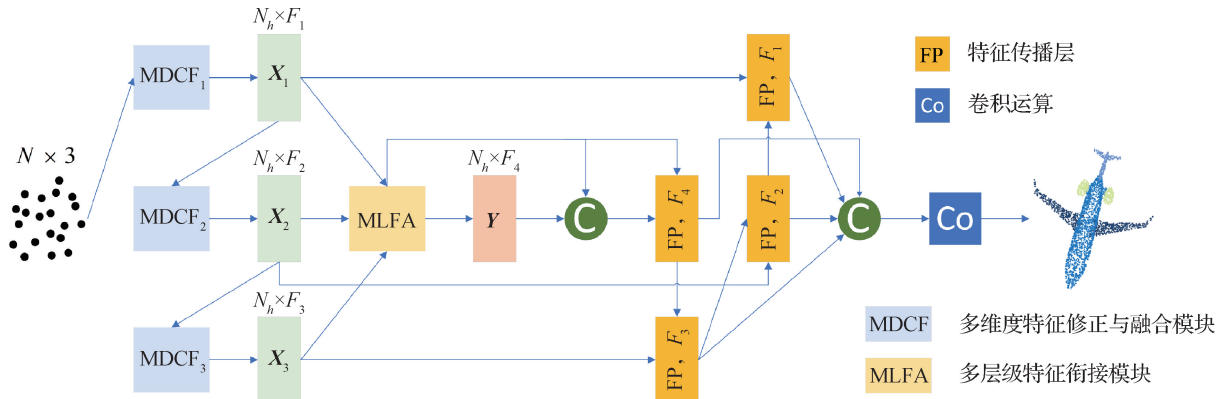


图5 多维度多层次特征分割网络的架构图

Fig. 5 Architecture of a multi-dimensional multi-layer features segmentation network

2 实验

为验证本文网络在点云分类与分割任务中的性能,选择标准数据集 ModelNet40 (Wu 等,2015) 和 ShapeNet (Chang 等,2015) 对多维度多层次神经网络 (MM-Net) 的性能进行评估,并与当前最新的研究进行对比。

在多维度多层次神经网络 (MM-Net) 中,选择 Adam 作为训练中的优化算法。在分类和分割实验中,将 batchsize 分别定为 32 和 24,设定初始学习率为 0.001,每 30 个 epochs 衰减 1/2 学习率,设置 batchnorm 为 0.6,并以每 20 个 epochs 衰减 0.5。所有实验都在 NVIDIA GTX 2080 GPU 上进行,并通过 pytorch 编写训练和测试代码。

2.1 ModelNet40 形状分类任务

选择基准数据集 ModelNet40 测试多维度多层次特征分类网络 (MM-Net-C) 的有效性。该数据集包含 12 311 个计算机辅助设计 (computer aided design, CAD) 模型,分为 40 类。实验时,将 ModelNet40 数据集分为两部分,训练阶段 9 843 个,测试阶段 2 468 个。同时在每一 CAD 表面均匀采样 1 024 个点,并且仅将 3 维坐标作为点的原始属性输入网络。另一方面,在增强策略中,在 $[-0.66, 1.5]$ 范围内设置随机各向异性缩放和 $[-0.2, 0.2]$ 范围内的平移,选定 dropout 为 50%。在测试

过程中,类似于 PointNet 等研究,经由随机缩放做 10 次投票测试,随后对结果进行平均以获得最后预测。

当前点云分类领域中各先进方法在 ModelNet40 数据集的分类准确率 (accuracy, Acc) 如表 1 所示。其中, pnt 表示仅输入点云坐标, pnt 和 nor 表示输入点云的坐标和法向量。

从表 1 可以看出, MM-Net-C 在准确度指标上表现比其他方法更好,较基于 pnt 的 DGCNN 和 RS-CNN 方法提高了 1.9% 和 0.5%; 较基于 pnt 和 nor 的 PAT、PointConv 和 PointASNL 方法提高了 2.4%、1.6% 和 0.9%。即使是一些方法输入了更多的点进行训练, MM-Net-C 的准确率依然较高,比 PointNet++ (5 k) 和 SO-Net (5 k) 方法分别提高了 2.2% 和 0.7%。实验结果表明,本文模型不仅提高了准确性,还避免了较大的计算量。

为了验证 MM-Net-C 的网络复杂度,与一些先进方法进行对比,结果如表 2 所示。可以看出, MM-Net-C 依靠较少的复杂度,取得了更优的效果。与 PointCNN 方法 (8.20 M, 91.7%) 相比,多维度多层次特征分类网络参数量仅约为其 1/8,而正确率却提高了 2.4%; 与 RS-CNN (1.41 M, 93.6%) 相比, MM-Net-C 参数量降低了 0.33 M,而正确率却提高了 0.5%。实验结果表明 MM-Net-C 具有较少的参数,更低的计算成本,实现了更高的分类精度。

表1 不同方法在 ModelNet40 数据集的分类准确率比较

Table 1 Comparison of classification accuracy on ModelNet40 dataset among different methods

方法	输入	训练点/k	Acc/%
PAT(Yang 等,2019)	pnt 和 nor	1	91.7
Spec-GCN(Wang 等,2018a)	pnt 和 nor	1	91.8
PointConv(Wu 等,2019)	pnt 和 nor	1	92.5
A-CNN (Komarichev 等,2019)	pnt 和 nor	1	92.6
PointASNL(Yan 等,2020)	pnt 和 nor	1	93.2
ELM(Fujiwara 和 Hashimoto,2020)	pnt 和 nor	1	93.2
RS-CNN(Liu 等,2019c)	nor	1	92.8
Pointwise-Cnn(Hua 等,2018)	pnt	1	86.1
ECC (Simonovsky 和 Komodakis,2017)	pnt	1	87.4
PointNet(Qi 等,2017a)	pnt	1	89.2
PointCNN(Li 等,2018b)	pnt	1	91.7
DGCNN(Wang 等,2019)	pnt	1	92.2
SO-CNN(Zhang 等,2020)	pnt	1	93.1
Dense-Point(Liu 等,2019b)	pnt	1	93.2
RS-CNN(Liu 等,2019c)	pnt	1	93.6
SO-Net(Li 等,2018a)	pnt 和 nor	2	90.9
PointNet ++ (Qi 等,2017b)	pnt 和 nor	5	91.9
Spider-CNN(Xu 等,2018)	pnt 和 nor	5	92.4
SO-Net(Li 等,2018a)	pnt 和 nor	5	93.4
本文	pnt	1	94.1

注:加粗字体表示最优结果。表中: PAT 为 point attention transformers, Spec-GCN 为 spectral graph convolution, A-CNN 为 annularly convolutional neural network, ELM 为 extreme learning machine, ECC 为 edge-conditioned convolution, SO-CNN 为 shape-oriented convolutional neural network, SO-Net 为 self-organizing network。

表2 ModelNet40 数据集的分类复杂度比较

Table 2 Comparison of the classification complexity on ModelNet40 dataset

方法	参数量/M	Acc/%
PointNet	3.50	89.2
PointCNN	8.20	91.7
PointNet ++	1.48	91.9
RS-CNN	1.41	93.6
本文	1.08	94.1

注:加粗字体表示各列最优结果。

2.2 ShapeNet 形状分割任务

为验证多维度多层级特征分割网络 (MM-Net-S) 的性能,在 ShapeNet 分割数据集上进行实验。ShapeNet 数据集由 16 881 个 3D 对象构成,包括 16 个不同的类别,有 50 个形状被分别注释,每个形状有 2 ~ 6 个部件。在该项分割任务中,预定义 ShapeNet 数据集中的每个点对应的形状类别标签。与 PointNet 策略数据处理操作相同,将 ShapeNet 数据集划分为训练对象(12 137 个)、验证对象(1 870 个)和测试对象(2 874 个),并且单个形状为 2 048 个点。选择平均交并比(mean intersection over union, mIoU)作为所有形状的评估指标,实验结果如表 3 所示。可以看出,在 ShapeNet 分割数据集上,与当前优秀的方法进行对比,MM-Net-S 达到了最佳效果。与 DGCNN、SO-CNN 和 A-CNN 相比,分割准确率分别提高了 1.4%、0.8% 和 0.4%,表明 MM-Net-S 能够有效区分不同形状类别,例如“包”、“帽子”、“轿车”、“椅子”、“摩托车”、“杯子”。在 ShapeNet 分割数据集中,“包”和“帽子”是点云形状构成比较简单的对象,MM-Net-S 相比其他方法表现的更好,因为 MM-Net-S 同时考虑到了局部与全局的关系。对其他比较复杂的形状,例如“轿车”和“摩托车”都拥有更多的类别,同时这些类别更小,类别与类别之间差异较小,MM-Net-S 同样能够很好地找出类别之间存在的微小差异,实现类别的正确分割。这样的优势,归功于本文网络中的多维度特征修正与融合模块(MDCF),该模块能够在多维度探索微小类别内的边缘差异。图 6 展示了 MM-Net-S 在 ShapeNet 分割数据集的可视化结果。从这些结果可以发现,即使每个对象中的类别构成较为复杂,MM-Net-S 依旧能够很好地将它们分割出来。

2.3 消融实验

多维度特征修正与融合模块(MDCF)和多层级特征衔接模块(MLFA)是多维度多层级神经网络(MM-Net)的核心。为了确定每一模块在网络中存在的必要性,在 ModelNet40 基准数据集上评估不同模块的性能。

各模块消融实验的结果如表 4 所示。可以看出,基线模型的准确率为 90.1%。RS-CNN(*)是另一个参考模型,由 RS-CNN 修改而成,仅有双层的共享多层感知机,准确率为 92.0%。基线模型装配多维度特征修正与融合模块(MDCF)后,准确率为

表 3 ShapeNet 数据集的分割准确率比较
Table 3 Comparison of segmentation accuracy of ShapeNet dataset

方法	ShapeNet 中的形状																平均
	飞机	包	帽子	轿车	椅子	耳机	吉他	小刀	灯	手提电脑	摩托车	杯子	手枪	火箭	滑板	桌子	
PointNet	83.4	78.7	82.5	74.9	89.6	73.0	91.5	85.9	80.8	95.3	65.2	93.0	81.2	57.9	72.8	80.6	83.7
SONet	82.8	77.8	88.0	77.3	90.6	73.5	90.7	83.9	82.8	94.8	69.1	94.2	80.9	53.1	72.9	83.0	84.9
PointNet ++	82.4	79.0	87.7	77.3	90.8	71.8	91.0	85.9	83.7	95.3	71.6	94.1	81.3	58.7	76.4	82.6	85.1
DGCNN	84.2	83.7	84.4	77.1	90.9	78.5	91.5	87.3	82.9	96.0	67.8	93.3	82.6	59.7	75.5	80.6	85.1
ELM	84.0	80.4	88.0	80.2	90.7	77.5	91.2	86.4	82.6	95.5	70.0	93.9	84.1	55.6	75.6	82.1	85.3
SpiderCNN	83.5	81.0	87.2	77.5	90.7	76.8	91.1	87.3	83.3	95.8	70.2	93.5	82.7	59.7	75.8	82.8	85.3
SO-CNN	83.9	84.1	85.0	77.4	91.3	78.3	91.7	87.4	83.8	96.4	69.7	93.5	83.1	58.9	76.2	82.9	85.7
A-CNN	84.2	84.0	88.0	79.6	91.3	75.2	91.6	87.1	85.5	95.4	75.3	94.9	82.5	67.8	77.5	83.3	86.1
RS-CNN	83.5	84.8	88.8	79.6	91.2	81.1	91.6	88.4	86.0	96.0	73.7	94.1	83.4	60.5	77.7	83.6	86.2
本文	83.9	85.6	89.5	80.2	91.6	79.1	91.4	87.8	85.8	96.0	75.4	95.0	84.0	63.7	76.0	83.4	86.5

注:加粗字体表示各列最优结果。

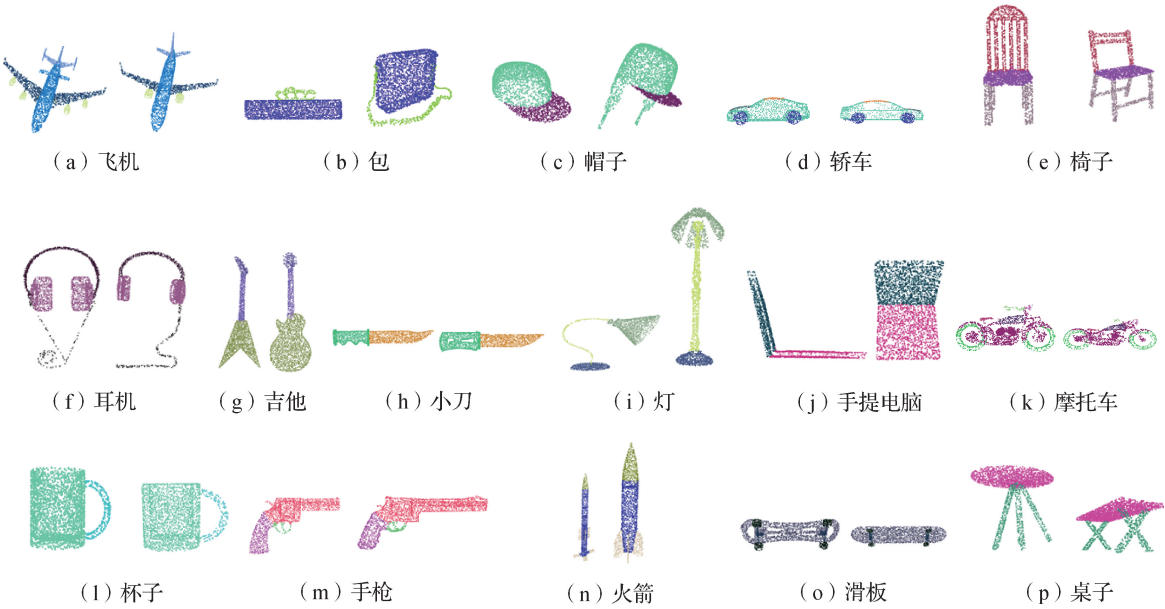


图 6 多维度多层次特征分割网络在 ShapeNet 数据集的可视化结果

Fig. 6 Visualization results of a multi-dimensional multi-layer features segmentation network on ShapeNet dataset
((a) airplane; (b) bag; (c) cap; (d) car; (e) chair; (f) earphone; (g) guitar; (h) knife; (i) lamp; (j) lap; (k) motorbike;
(l) mug; (m) pistol; (n) rocket; (o) skate; (p) table)

93.3%, 相较其基准模型提高了 3.2%。RS-CNN (*) 配备多层次特征衔接模块 (MLFA) 后, 准确率为 93.1%, 相较 RS-CNN (*) 提高了 1.1%。基线模型同时装配 MDCF 和 MLFA 模块后, 正确率为 94.1%, 相较其基准模型提高了 4.0%。实验结果

表明, 添加 MDCF 或 MLFA 模块对点云分析任务都能够大幅提高模型的准确率。同时也说明 MDCF 模块利用丰富的逐点特征与潜在的局部特征相修正并融合, 能够更好地提高网络对局部区域内形状信息的挖掘。而 MLFA 模块能够把握全局范围内的上

表4 模块在 ModelNet40 数据集上的消融实验
Table 4 Ablation experiments with modules on ModelNet40 dataset

模型	Acc /%
基线	90.1
RS-CNN(*)	92.0
基线 + MDCF	93.3
RS-CNN(*) + MLFA	93.1
基线 + MDCF + MLFA	94.1

注:加粗字体表示最优结果。

下文信息,强化不同层级之间存在的远程依赖联系,有效加强了模型处理复杂形状的鲁棒性。同时装配 MDCF 和 MLFA 模块的效果最优,能够考虑局部形状特征与全局上下文联系,验证了二者在 MM-Net 中的重要性。

表5是对多维度特征修正与融合模块(MDCF)是否需要设计不同维度问题进行相关实验的结果。可以看出,在相同维度条件下,MM-Net 比 RS-CNN 表现得更加优秀。同时,在不同维度下,MM-Net 的准确率也不同,原因是3个维度下的 MDCF 模块能够合理地由点云数据集中构建不同范围的学习区域,并自适应地从这些区域范围中挖掘模型所需的形状信息。

表5 在 ModelNet40 数据集上对 MDCF 模块维度的消融实验

Table 5 Ablation experiments on ModelNet40 dataset for dimensions in the multi-dimensional feature correction and fusion module

模型	维度	Acc/%
RS-CNN	2	92.4
RS-CNN	3	93.6
MM-Net	2	92.5
MM-Net	3	94.1

注:加粗字体表示最优结果。

表6是在 ModelNet40 数据集上对多层次特征衔接(MLFA)模块选择不同层级对整体网络准确率的消融实验结果。可以发现,当选择3层层级结构设计时,模型实现了最佳的精度 94.1%,与双层设计(92.8%)相比,提高了 1.3%。但进一步增加

MLFA 模块的层级数后,模型准确率相较于3层设计下降了 0.7%,原因是网络层级的加深导致模型在训练中学习形状的困难,降低了 MLFA 模块构建层间关联信息的能力。

表6 在 ModelNet40 数据集上对 MLFA 模块层级的消融实验

Table 6 Ablation experiments on ModelNet40 dataset for layers in the multi-layer feature articulation module

层级	Acc /%
2	92.8
3	94.1
4	93.4

注:加粗字体表示最优结果。

3 结 论

本文提出一个以多维度特征修正与融合模块(MDCF)和多层级特征衔接模块(MLFA)为核心部件的多维度多层次神经网络(MM-Net),能够端到端且鲁棒地对不规则点云进行分析,弥补了现有方法未有效兼顾局部形状与全局形状的不足,缓解了缺乏逐点特征对模型学习局部形状的消极影响。在多维度特征修正与融合模块(MDCF)中,利用局部区域划分单元(LAD)得到点间几何关系与局部特征,并与逐点特征通过自适应方式进行修正与融合。MDCF 模块整理了可靠的点间几何关系,并输出能够合理描述该区域的局部形状。同时,在多层次特征衔接模块(MLFA)中,利用层级间存在的关联性,识别出能够提高形状感知的层间上下文信息。通过跳跃连接的方式,合理利用了全局特征,减少了梯度消失的影响。另一方面,也加快层级之间的信息交互,进而大幅提高了对错误形状的鲁棒性。最后设计了两种能够分别处理分类与分割任务的网络架构 MM-Net-C 和 MM-Net-S。实验结果表明,本文提出的多维度多层次神经网络(MM-Net),在参数较少的优势下,实现了较高的正确率。

在未来的工作中,将进一步挖掘网络的鲁棒性,在不影响准确率和参数量的情况下,有效处理富有噪声的点云数据,缓解现实环境中存在的点云异常

值难以区分并处理的问题,以实现在遥感和自动驾驶等领域对相关场景进行解析的应用。

参考文献 (References)

- Chang A X, Funkhouser T, Guibas L, Hanrahan P, Huang Q X, Li Z M, Savarese S, Savva M, Song S R, Su H, Xiao J X, Yi L and Yu F. 2015. ShapeNet: an information-rich 3D model repository. [EB/OL]. [2021-06-23]. <https://arxiv.org/pdf/1512.03012.pdf> arXiv
- Fujiwara K and Hashimoto T. 2020. Neural implicit embedding for point cloud analysis//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA; IEEE: 11731-11740 [DOI: 10.1109/CVPR42600.2020.01175]
- Hua B S, Tran M K and Yeung S K. 2018. Pointwise convolutional neural networks//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA; IEEE: 984-993 [DOI: 10.1109/CVPR.2018.00109]
- Huang G, Liu Z, Van Der Maaten L and Weinberger K Q. 2017. Densely connected convolutional networks//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA; IEEE: 2264-2269 [DOI: 10.1109/CVPR.2017.243]
- Kamoussi P, Lazard S, Maheshwari A and Wuhler S. 2016. Analysis of farthest point sampling for approximating geodesics in a graph. Computational Geometry, 57: 1-7 [DOI: 10.1016/j.comgeo.2016.05.005]
- Komarichev A, Zhong Z C and Hua J. 2019. A-CNN: annularly convolutional neural networks on point clouds//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA; IEEE: 7413-7422 [DOI: 10.1109/CVPR.2019.00760]
- Li J X, Chen B M and Lee G H. 2018a. SO-Net: self-organizing network for point cloud analysis//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA; IEEE: 9397-9406 [DOI: 10.1109/CVPR.2018.00979]
- Li Y Y, Bu R, Sun M C, Wu W, Di X H and Chen B Q. 2018b. PointCNN: convolution on X-transformed points//Proceedings of the 32nd International Conference on Neural Information Processing Systems. Montréal, Canada; Curran Associates Inc.: 820-830
- Liu X H, Han Z Z, Liu Y S and Zwicker M. 2019a. Point2Sequence: learning the shape representation of 3D point clouds with an attention-based sequence to sequence network. Proceedings of the AAAI Conference on Artificial Intelligence, 33(1): 8778-8785 [DOI: 10.1609/aaai.v33i01.33018778]
- Liu Y C, Fan B, Meng G F, Lu J W, Xiang S M and Pan C H. 2019b. DensePoint: learning densely contextual representation for efficient point cloud processing//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision. Seoul, Korea (South): IEEE: 5238-5247 [DOI: 10.1109/ICCV.2019.00534]
- Liu Y C, Fan B, Xiang S M and Pan C H. 2019c. Relation-shape convolutional neural network for point cloud analysis//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA; IEEE: 8887-8896 [DOI: 10.1109/CVPR.2019.00910]
- Liu Z, Zhou S B, Suo C Z, Yin P, Chen W, Wang H S, Li H A and Liu Y H. 2019d. LPD-Net: 3D point cloud learning for large-scale place recognition and environment analysis//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision (ICCV). Seoul, Korea (South): IEEE: 2831-2840 [DOI: 10.1109/ICCV.2019.00292]
- Qi C R, Liu W, Wu C X, Su H and Guibas L J. 2018. Frustum PointNets for 3D object detection from RGB-D data//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA; IEEE: 918-927 [DOI: 10.1109/CVPR.2018.00102]
- Qi C R, Su H, Mo K, and Guibas L J. 2017a. Pointnet: Deep learning on point sets for 3 d classification and segmentation//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition Honolulu, USA; IEEE: 77-85
- 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
- Shi S S, Wang X G and Li H S. 2019. PointRCNN: 3D object proposal generation and detection from point cloud//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, USA; IEEE: 770-779 [DOI: 10.1109/CVPR.2019.00086]
- Simonovsky M and Komodakis N. 2017. Dynamic edge-conditioned filters in convolutional neural networks on graphs//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA; IEEE: 29-38 [DOI: 10.1109/CVPR.2017.11]
- Wang C, Samari B and Siddiqi K. 2018a. Local spectral graph convolution for point set feature learning//Proceedings of the 15th European Conference on Computer Vision (ECCV). Munich, Germany: Springer: 56-71 [DOI: 10.1007/978-3-030-01225-0_4]
- Wang X L, Girshick R, Gupta A and He K M. 2018b. Non-local neural networks//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA; IEEE: 7794-7803 [DOI: 10.1109/CVPR.2018.00813]
- 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 W X, Qi Z G and Li F X. 2019. Point Conv: deep convolutional networks on 3D point clouds//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long

- Beach, USA; IEEE: 9613-9622 [DOI: 10. 1109 / CVPR. 2019. 00985]
- Wu Z R, Song S R, Khosla A, Yu F, Zhang L G, Tang X O and Xiao J X. 2015. 3D ShapeNets: a deep representation for volumetric shapes//Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, USA; IEEE: 1912-1920 [DOI: 10.1109/CVPR.2015.7298801]
- Xu Y F, Fan T Q, Xu M Y, Zeng L and Qiao Y. 2018. SpiderCNN: deep learning on point sets with parameterized convolutional filters//Proceedings of the 15th European Conference on Computer Vision. Munich, Germany; Springer: 90-105 [DOI: 10.1007/978-3-030-01237-3_6]
- 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]
- Yang J C, Zhang Q, Ni B B, Li L G, Liu J X, Zhou M D and Tian Q. 2019. Modeling point clouds with self-attention and gumbel subset sampling//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA; IEEE: 3318-3327 [DOI: 10.1109/CVPR.2019.00344]
- Zeiler M D and Fergus R. 2014. Visualizing and understanding convolutional networks//Proceedings of the 13th European Conference on Computer Vision. Zurich, Switzerland; Springer: 818-833 [DOI: 10.1007/978-3-319-10590-1_53]
- Zhang C Y, Song Y, Yao L N and Cai W D. 2020. Shape-oriented convolution neural network for point cloud analysis. Proceedings of the AAAI Conference on Artificial Intelligence, 34(7): 12773-12780 [DOI: 10.1609/aaai.v34i07.6972]

- Zhu Y K, Mottaghi R, Kolve E, Lim J J, Gupta A, Li F F and Farhadi A. 2017. Target-driven visual navigation in indoor scenes using deep reinforcement learning//Proceedings of 2017 IEEE International Conference on Robotics and Automation (ICRA). Singapore, Singapore; IEEE: 3357-3364 [DOI: 10.1109/ICRA.2017.7989381]

作者简介



徐嘉利, 1996年生,男,硕士研究生,主要研究方向为模式识别和人工智能、计算机视觉、点云分析。

E-mail: xuxjlxujiali@gmail.com



方志军,通信作者,男,教授,主要研究方向为图像视频分析、计算机视觉、智能数据分析、模式识别和人工智能、点云分析。

E-mail: zjfang@sues.edu.cn

伍世虔,男,教授,主要研究方向为机器视觉、智能机器人、公共安全与视频技术、人工智能及机器学习。

E-mail: sqwu@hotmail.com