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

中国图象图形学报

2021
11
VOL.26

ISSN1006-8961
CN11-3758/TB

电力视觉前沿技术

中国图象图形学报

刊名题字：宋健 | 月刊（1996年创刊）

第26卷第11期（总第307期）

2021年11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

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

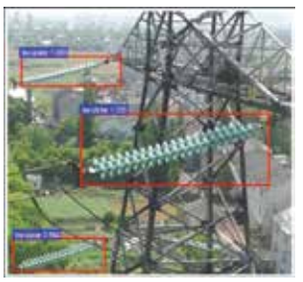
Distributed by Beijing Bureau for Distribution of Newspapers and Journals
Domestic All Local Post Offices in China
Overseas China International Book Trading Corporation
(P.O.Box 399, Beijing 100048,P.R.China))
Printed by Beijing Kexin Printing Co., Ltd.

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

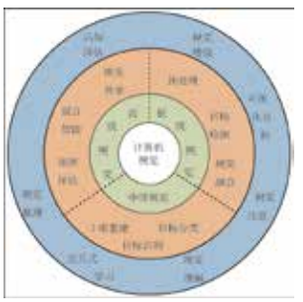
国外发行代号 M1406
国内邮发代号 82-831
国内定价 60.00元

2021年11月 第26卷 第11期（总第307期）
中国图象图形学报
Zhongguo Tuxiang Tuxing Xuebao

目次



面向改进尺度缩放网络的绝缘子识别(第2561页)

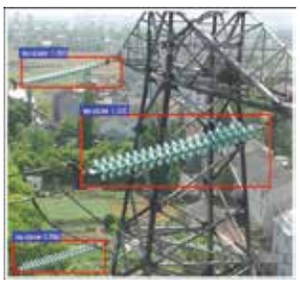


混合增强视觉认知架构及其关键技术进展(第2619页)

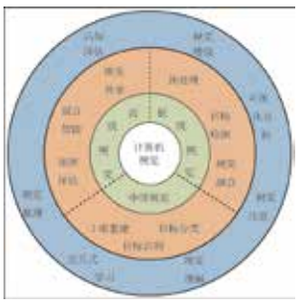
激光点云的稀疏体素金字塔邻域构建与分类(第2703页)

Volume 26, Number 11
Published November 16, 2021

CONTENTS
JOURNAL OF IMAGE AND GRAPHICS



Insulator recognition based on an improved scale-transferrable network(P2561)



Hybrid enhanced visual cognition framework and its key technologies(P2619)

Sparse voxel pyramid neighborhood construction and classification of LiDAR point cloud(P2703)

中图法分类号:TP391　　**文献标识码**: A　　**文章编号**: 1006-8961(2021)11-2691-12

论文引用格式: Song W, Cai W Y, He S Q and Li W J. 2021. Dynamic graph convolution with spatial attention for point cloud classification and segmentation. Journal of Image and Graphics,26(11):2691-2702(宋巍,蔡万源,何盛琪,李文俊. 2021. 结合动态图卷积和空间注意力的点云分类与分割. 中国图象图形学报,26(11):2691-2702)[DOI:10.11834/jig.200550]

结合动态图卷积和空间注意力的点云分类与分割

宋巍[1],蔡万源[1],何盛琪[1],李文俊[2*]

1. 上海海洋大学信息学院, 上海　201306; 2. 中国极地研究中心, 上海　200136

摘　要: **目的**　随着3维采集技术的飞速发展,点云在计算机视觉、自动驾驶和机器人等领域有着广泛的应用前景。深度学习作为人工智能领域的主流技术,在解决各种3维视觉问题上已表现出巨大潜力。现有基于深度学习的3维点云分类分割方法通常在聚合局部邻域特征的过程中选择邻域特征中的最大值特征,忽略了其他邻域特征中的有用信息。**方法**　本文提出一种结合动态图卷积和空间注意力的点云分类分割方法(dynamic graph convolution spatial attention neural networks,DGCSA)。通过将动态图卷积模块与空间注意力模块相结合,实现更精确的点云分类分割效果。使用动态图卷积对点云数据进行K近邻构图并提取其边特征。在此基础上,针对局部邻域聚合过程中容易产生信息丢失的问题,设计了一种基于点的空间注意力(spatial attention,SA)模块,通过使用注意力机制自动学习出比最大值特征更具有代表性的局部特征,从而提高模型的分类分割精度。**结果**　本文分别在ModelNet40、ShapeNetPart和S3DIS(Stanford Large-scale 3D Indoor Spaces Dataset)数据集上进行分类、实例分割和语义场景分割实验,验证模型的分类分割性能。实验结果表明,该方法在分类任务上整体分类精度达到93.4%;实例分割的平均交并比达到85.3%;在室内场景分割的6折交叉检验平均交并比达到59.1%,相比基准网络动态图卷积网络分别提高0.8%、0.2%和3.0%,有效改善了模型性能。**结论**　使用动态图卷积模块提取点云特征,在聚合局部邻域特征中引入空间注意力机制,相较于使用最大值特征池化,可以更好地聚合邻域特征,有效提高了模型在点云上的分类、实例分割与室内场景语义分割的精度。

关键词:点云;动态图卷积;空间注意力(SA);分类;分割

Dynamic graph convolution with spatial attention for point cloud classification and segmentation

Song Wei[1], Cai Wanyuan[1], He Shengqi[1], Li Wenjun[2*]

1. College of Information Technology, Shanghai Ocean University, Shanghai 201306, China;
2. Polar Research Institute of China, Shanghai 200136, China

Abstract: **Objective** With the rapid development of 3D acquisition technologies, point cloud has wide applications in many areas, such as medicine, autonomous driving, and robotics. As a dominant technique in artificial intelligence (AI), deep learning has been successfully used to solve various 2D vision problems and has shown great potential in solving 3D vision

收稿日期:2020-09-15;**修回日期**:2020-12-24;**预印本日期**:2020-12-31

*通信作者:李文俊　liwenjun@pric.org.cn

基金项目:国家重点研发计划项目(2016YFC1400304);国家自然科学基金项目(61972240);上海科委部分地方高校能力建设项目(20050501900)

Supported by: National Key Research and Development Program of China (2016YFC1400304); National Natural Science Foundation of China (61972240); Shanghai Science and Technology Commission part of the Local University Capacity Building Projects (20050501900)

problems. Using regular grid convolutional neural networks (CNN) for non-Euclidian space of point cloud data and capturing the hidden shapes from irregular points remains challenging. In recent years, deep learning-based methods have been more effective in point cloud classification and segmentation than traditional methods. Deep learning-based methods can be divided into three groups: pointwise methods, convolutional-based methods, and graph convolutional-based methods. These methods include two important processes: feature extraction and feature aggregation. Most of the methods focus on the design of feature extraction and pay less attention to feature aggregation. At present, most point cloud classification and segmentation methods based on deep learning use max pooling for feature aggregation. However, using the maximum value features of neighborhood features in local neighborhood features has the problem of information loss caused by ignoring other neighborhood information. **Method** This paper proposes a dynamic graph convolution with spatial attention for point cloud classification and segmentation method based on deep learning—dynamic graph convolution spatial attention (DGCSA) neural networks. The key of the network is to learn from the relationship between the neighbor points and the center point, which avoid the information loss caused by feature aggregation using max pool layers in feature aggregation. This network is composed of a dynamic graph convolution module and a spatial attention (SA) module. The dynamic graph convolution module mainly performs K-nearest neighbor (KNN) search algorithm and multiple-layer perception. For each point cloud, it first uses the KNN algorithm to search its neighbor points. Then, it extracts the features of the neighbor points and center points by convolutional layers. The K-nearest neighbors of each point vary in different network layers, leading to a dynamic graph structure updated with layers. After feature extraction, it applies a point-based SA module to learn the local features that are more representative than the maximum feature automatically. The key of the SA module is to use the attention mechanism to calculate the weight of K-neighbor points of the center point. It consists of four units: 1) attention activation unit, 2) attention scores unit, 3) weighted features unit, and 4) multilayer perceptron unit. First, the attention activation of each potential feature is learned through the fully connected layer. Second, the attention score of the corresponding feature is calculated by applying the SoftMax function on the attention activation value. The learned attention score can be regarded as a mask for automatically selecting useful potential features. Third, the attention score is multiplied by the corresponding elements of the local neighborhood features to generate a set of weighted features. Finally, the sum of the weighted features is determined to obtain the locally representative local features, followed by another fully connected convolutional layer to control the output dimension of the SA module. The SA module has strong learning ability, thereby improving the classification and segmentation accuracy of the model. DGCSA implements a high-performance classification and segmentation of point clouds by stacking several dynamic graph convolution modules and SA modules. Moreover, feature fusion is used to fuse the output features of different spatial attention layers that can effectively obtain the global and local characteristics of point cloud data, achieving better classification and segmentation results. **Result** To evaluate the performance of the proposed DGCSA model, experiments are carried out in classification, instance segmentation, and semantic scene segmentation on the datasets of ModelNet40, ShapeNetPart, and Stanford large-scale 3D Indoor spaces dataset, respectively. Experiment results show that the overall accuracy (OA) of our method reaches 93.4%, which is 0.8% higher than the baseline network dynamic graph CNN (DGCNN). The mean intersection-to-union (mIoU) of instance segmentation reaches 85.3%, which is 0.2% higher than DGCNN; for indoor scene segmentation, the mIoU of the six-fold cross-validation reaches 59.1%, which is 3.0% higher than DGCNN. Overall, the classification accuracy of our method on the ModelNet40 dataset surpasses that of most existing point cloud classification methods, such as PointNet, PointNet++, and PointCNN. The accuracy of DGCSA in instance segmentation and indoor scene segmentation reaches the segmentation accuracy of the current excellent point cloud segmentation network. Furthermore, the validity of the SA module is verified by an ablation study, where the max pooling operations in PointNet and linked dynamic graph CNN (LDGCNN) are replaced by the SA module. The classification results on the ModelNet40 dataset show that the SA module contributes to a more than 0.5% increase of classification accuracy for PointNet and LDGCNN. **Conclusion** DGCSA can effectively aggregate local features of point cloud data and achieve better classification and segmentation results. Through the design of SA module, this network solves the problem of partial information loss in the aggregation local neighborhood information. The SA module fully considers all neighborhood contributions, selectively strengthens the features containing useful information, and suppresses useless features. Combining the spatial attention module with the dynamic graph convolution module, our network

can improve the accuracy of classification, instance segmentation, and indoor scene segmentation. In addition, the spatial attention module can integrate with other point cloud classification model and substantially improve the model performance. Our future work will improve the accuracy of DGCSA in segmentation task in the condition of an unbalanced dataset.

Key words: point cloud; dynamic graph convolution; spatial attention (SA); classification; segmentation

0 引 言

随着激光雷达扫描仪、RGBD 相机等 3D 采集技术的快速发展,3 维点云数据的种类和数量显著增长。基于 3 维点云的分类、分割和识别等技术在自动驾驶(Chen 等,2017)、文化遗产保护、医疗、机器人和城市规划等领域的应用需求也大大提升。传统的点云分类或分割方法,如基于区域分割的方法(Besl 和 Jain,1988)、基于属性的方法(Filin 和 Pfeifer,2006)和基于图优化(Fischler 和 Bolles,1981)的方法等,往往需要手工设计特征。因此,数据驱动的方法被提出用于点云特征的自动学习,特别是深度学习方法(LeCun 等,2015)受到了越来越多的关注。然而,3 维点云数据固有的无序性、排列不变性和稀疏性给现有的点云分类与分割方法带来了巨大的挑战。现有深度学习方法在点云分类分割上具有效率低、局部细节识别易混淆等缺点(Guo 等,2020)。如何有效地对 3 维点云模型进行高精度识别成为一个亟待解决的问题。

一些工作将不规则的点云转化为规则的数据形式以便于特征学习,例如多视图方法(Su 等,2015)和体素化方法(Zhou 和 Tuzel,2018)。Qi 等人(2017a)提出了直接处理原始 3 维点云数据的点云分类分割深度学习框架,避免了冗余和费力的多视图和立体像素化数据预处理操作,缺点是只考虑了点云的全局特征而忽略了局部特征。随后,学者们开始重视点云局部特征的提取。例如,PointNet ++(Qi 等,2017b)使用球半径查询算法或 K 近邻算法(K-nearest neighbor, KNN)构造局部邻域,并提取局部邻域特征。PointWeb(Zhao 等,2019)构造局部邻域内的稠密连接网,并通过一个自适应特征调整(adaptive feature adjustment, AFA)模块学习各点之间的相互关联。这些方法都使用最大值对局部邻域进行特征聚合,忽略了局部邻域内的其他邻域特征。

针对现有网络使用邻域特征中的最大值特征代表局部特征而忽略其他特征信息的问题,本文提出了空间注意力(spatial attention, SA)模块。空间注意力模块通过权值计算获取到更具有代表性的局部特征,避免了最大值操作所带来的信息丢失问题,提高了网络的分类与分割精度。同时,本文使用图卷积与空间注意力模块搭建分类分割网络。利用动态构图法表达 3 维点云局部拓扑结构,提取其边的特征,并使用 SA 对边的特征进行特征聚合。此方法在 ModelNet40(Wu 等,2015)分类任务、ShapeNet-Part(Yi 等,2016)和斯坦福大学大型 3D 室内空间数据集(Stanford Large-scale 3D Indoor Spaces Dataset, S3DIS)(Armeni 等,2016)分割任务中都取得了显著的效果,超过或达到了现有的大多数方法。

1 深度学习点云分类分割网络

根据网络特点,可以将基于深度学习的 3 维点云分类或分割大体分为 3 种:基于点态多层感知机的方法、基于卷积的方法和基于图的方法(Guo 等,2020)。

1.1 基于点态多层感知机的方法

最早的 3D 点云分类/分割网络是 PointNet(Qi 等,2017a)和 PointNet ++(Qi 等,2017b)。点态多层感知方法使用多层感知(multi-layer perception, MLP)层独立提取每个点的特征,然后通过最大值池化操作将其聚合为全局特征(Guo 等,2020)。PointNet ++ 与 PointNet 的不同之处在于,PointNet ++ 以 PointNet 作为其层次结构的核心,在特征提取之前添加了采样层和分组层,虽然考虑到了点云的局部特征,但是忽略了点云的上下文关联信息,缺乏捕获 3 维点云的集合特征的能力。在 PointNet 和 PointNet ++ 网络的基础上,还提出了其他方法,例如 PointWeb(Zhao 等,2019)。这种点态 MLP 网络的主要缺点是没有充分考虑 3 维点之间的几何关系。

1.2 基于卷积的方法

从设计适合于 3 维点云不规则性的卷积核的角

度,发展了许多基于卷积的方法。Wu 等人(2019)在 PointConv 中,将卷积核定义为 MLP 学习的加权函数和用于补偿 3 维空间中非均匀采样的学习权重的密度函数。Xu 等人(2018)在 SpiderCNN 中,将卷积设计为捕获局部测地距离信息的简单阶跃函数与保证特征提取可表达性的 KNN 的泰勒多项式的乘积。Thomas 等人(2019)在核点卷积(kernel point convolution,KPConv)中,提出了一种点卷积设计,不同于网格卷积,该卷积核的参数是一系列带有权重的核点组成的,核点的数目不固定,对规则和可变形卷积都具有良好的灵活性。PointCNN(Li 等,2018)使用了一个χ-Conv 操作来实现点云的置换不变性,但计算复杂度较高。

1.3 基于图卷积的方法

由于图卷积网络(graph convolutional networks,GCN)在处理非结构化和无序数据(Thomas 等,2019)方面的成功,基于图卷积网络的方法在点云应用中蓬勃发展。一般而言,基于图的神经网络将点云中的每个点视为图的一个顶点,并对每个点的所有邻居产生有向边。为了从顶点和边学习特征,可以在空间域和频谱域(Thomas 等,2019)中设计卷积算子。2017 年,边缘卷积神经网(edge-conditioned convolution,ECC)(Simonovsky 和 Komodakis,2017)提出了边条件卷积,该方法使用一个过滤生成网络对每个点的空间邻域进行过滤生成,并通过最大池化操作来聚合邻域信息。DGCNN(Wang 等,2019)提出边卷积(edge convolution,EdgeConv)层,采用边卷积操作捕获中心点与其相邻点之间的局部关系,可以动态地逐层更新关系图。EdgeConv 的核心层采用 KNN 算法构造图,并通过 MLP 学习边特征,最后通过最大值池化操作对特征进行聚合。LDGCNN(linked dynamic graph CNN)(Zhang 等,2019)通过链接不同层次的特征提高其性能。

图在谱域的卷积是为了找到相应的傅里叶性质。Bruna 等人(2013)利用图拉普拉斯算子的谱来生成卷积操作,提取图的全局特征。Defferrard 等人(2016)将图上的卷积定义为切比雪夫多项式近似,并通过 K 阶局部滤波操作实视每个点的快速特征映射。RGCNN(regularized graph CNN)(Te 等,2018)逐层更新图拉普拉斯矩阵,使特征学习具有可扩展性。这些方法作用于全图,忽略了相邻点的相对位置和特征。为了获取更丰富的局部结构信息,LocalSpecGCN(local spectral graph convolution)(Wang 等,2018)对一个由 k 个最近邻构造的局部图进行频谱图卷积。

总结目前针对不同点云识别任务(如分类、分割和目标检测)的深度学习网络,可以认为,无论局部学习、全局学习还是分层学习,深度学习网络通常都包括两个重要的步骤:特征学习和特征聚合。如前文所述,大部分工作集中在特征学习的卷积算子的设计上,较少关注特征聚合操作的作用。Wang 等人(2018)指出,在网络层中对信息聚合时采用了赢者通吃的最大值池化策略是目前点云深度学习方法的一个局限。因此,他们用递归聚类池化代替标准的最大值池化操作,该策略对点云的分类精度略有提升,但计算代价较高。Graph-CNN(Zhang 和 Rabbat,2018)使用全局和多分辨率池化来捕获点云的全局和局部特征,但是这种策略是针对特定网络的。

2 方 法

本文提出的点云深度学习网络结构如图 1 所示。图中上半部分的分支为点云分类任务的网络,下半部分的分支为分割任务的网络。输入点集 $\boldsymbol{P}=\{\boldsymbol{p}_1,\boldsymbol{p}_2,\cdots,\boldsymbol{p}_i,\cdots,\boldsymbol{p}_N\}\subseteq\mathbf{R}^3$,$\boldsymbol{p}_i$ 代表第 i 个点的空间坐标。N 为采样点的个数,D 为每个点的特征维度,$D=3$ 意味着输入点云数据的 3 维坐标(x,y,z)。输入 $N\times D$ 维的点云到可训练的空间转换网络 STN(spatial transform networks),此网络训练得到的空间转换矩阵可以对输入的点云数据进行坐标对齐;对齐后的数据使用动态图卷积(dynamic graph convolutional,DGC)模块操作得到局部邻域特征。分类网络中包含 4 层图卷积,从左到右卷积核个数依次为 64,64,128 和 256,每层图卷积提取完特性信息之后使用空间注意力(SA)模块对局部邻域特征进行特征聚合。各层输出的特征进行特征融合后通过 MLP 进行全局特征提取。如图 1 中所示,分类网络与分割网络的区别主要有两点:1)每层图卷积中 MLP 的数量不同,2)分割网络做了两次特征信息融合。分类网络输出整个点云属于 m 类的得分,分割网络则输出点云中每一个点所属 m 个类别的得分情况。

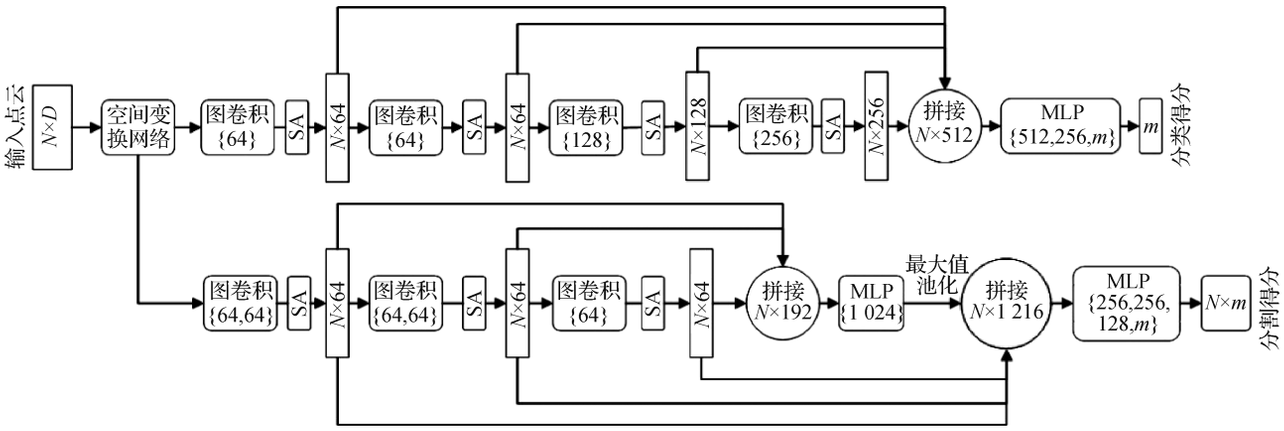


图 1　点云分类分割网络框架

Fig. 1　Point cloud classification and segmentation network architecture

如图 2 所示，点云分类分割网络的特征学习过程可以分为两个步骤：1）使用图卷积学习局部图结构中的边特征 $\boldsymbol{l}_i$；2）使用空间注意力模块对局部邻域特征 $\boldsymbol{l}_i$ 进行特征聚合，得到局部特征 $\boldsymbol{L}_i$。图 2 中 N 代表输入点的个数，f 代表每个输入点的特征维度，k 表示点云中心点的邻居点的个数，$\{32,64,\cdots,D\}$ 表示 MLP 各感知层神经元的数量，最后一层神经元数量为 D，则经过空间注意模块特征聚合后得到的特征维度是 $N \times D$。

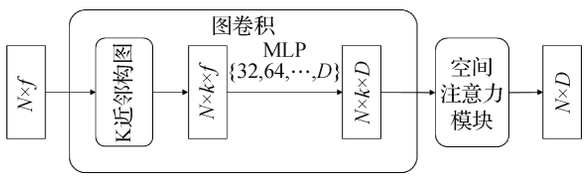


图 2　点云分类分割网络核心模块

Fig. 2　Point cloud classification segmentation network core module

2.1　动态图卷积

把数据构造成带有顶点和边的图，在图数据上进行卷积的操作称为图卷积。普通的图卷积在网络中的每一层都应用一个固定的图，并在图结构上应用卷积操作。本文采用了动态的图卷积，每个点的 K 近邻在不同的网络层中是不相同的。K 近邻点的变化使得每一层所构建的图结构随着层的不同而进行动态更新，因此称为动态图卷积。图卷积（Kipf 和 Welling，2016）可以分为两种类型：基于频谱域和基于空间域，本文所使用的图卷积属于后者。使用动态图卷积提取特征分为两个步骤：1）使用 K 近邻算法构造图结构；2）使用 MLP 提取图结构中的边特征。

2.1.1　构建图结构—K 近邻

由于点云的数目非常大，为了减少计算成本，需要构建一个 K 近邻图结构来表示点云的一个局部区域。使用 K 近邻算法是在非欧氏空间中聚集点云局部邻居信息的常用方法。K 近邻构图法不是构造具有完全连接的边的图，而是构造局部有向图 $\boldsymbol{G}$，从而减轻计算内存和计算资源的负担。如图 3 所示。

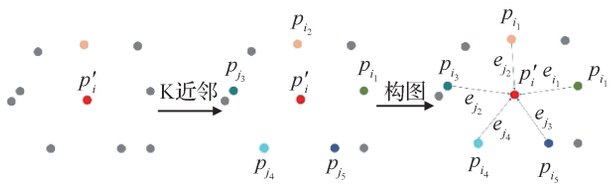


图 3　KNN 算法构造图结构

Fig. 3　Using KNN to construct the graph structure

由于特征提取操作对于每个中心点都是相同的，因此本文使用一个中心点 $\boldsymbol{p}_i$ 及其 k 个最近邻点 $\boldsymbol{p}_{i_j}, j = 1,2,\cdots,k$ 来说明此操作。其中，$\boldsymbol{p}_i \subseteq \mathbf{R}^f$（在网络第 1 层输入时，中心点的特征仅包括 3 维坐标，则其特征维度 f 的值为 3）；$\forall\ \boldsymbol{p}_{i_j} \subseteq \boldsymbol{Nei}(\boldsymbol{p}_i)$，$\boldsymbol{Nei}()$ 表示邻域。首先对中心点 $\boldsymbol{p}_i$（图 3 中红色的点）使用 K 近邻算法，所选取到的邻居点 $\boldsymbol{p}_{i_j}$ 在图 3 中以彩色表示，然后将中心点 $\boldsymbol{p}_i$ 与每个邻居点 $\boldsymbol{p}_{i_j}$ 使用有向边连接起来。使用 K 近邻算法所构建的图结构 $\boldsymbol{G}$ 的定义为

$$\boldsymbol{G} = (\boldsymbol{V},\boldsymbol{E}) \tag{1}$$

$$\boldsymbol{V} = \{\boldsymbol{p}_i \mid i = 1,2,\cdots,N\} \tag{2}$$

$$\boldsymbol{E} = \{\boldsymbol{e}_i = (\boldsymbol{e}_{i_1}, \boldsymbol{e}_{i_2}, \cdots, \boldsymbol{e}_{i_j}) \mid i = 1, 2, \cdots, N\} \tag{3}$$

在图 $\boldsymbol{G}$ 中，$\boldsymbol{V}$ 表示 N 个局部顶点集合，$\boldsymbol{E}$ 表示顶点之间的边的集合。

2.1.2　MLP 提取特征

在本文网络中，对图结构中边特征的提取是由一个共享的多层感知机实现的，其中包含卷积层，批处理归一化层 BN（batch normalization）和激活层 ReLU（rectified linear unit）。多层感知机定义为 h_{Θ}，它是带有一组可学习的权重参数的非线性函数，其参数集合定义为 $\boldsymbol{\Theta} = (\theta_1, \cdots, \theta_M, \phi_1, \cdots, \phi_M)$，$M$ 是滤波器的数量。$\boldsymbol{e}_{i_j}$ 表示中心点 $\boldsymbol{p}_i$ 及其相邻点 $\boldsymbol{p}_{i_j}$ 之间的边特征。边函数的选择对图卷积生成的边特征具有重要影响。Wang 等人（2018）对每个局部子图边特征的学习既考虑了中心点与邻居点的相对特征，也考虑了中心点 $\boldsymbol{p}_i$ 本身的特征，这样能够有效结合全局几何信息和局部几何信息。本文采取了这种边特征 $\boldsymbol{e}_{i_j}$ 定义

$$\boldsymbol{e}_{i_j} = h_{\Theta}(\boldsymbol{p}_i, \boldsymbol{p}_i - \boldsymbol{p}_{i_j}) \tag{4}$$

式中，h_{Θ} 是用来实现 $\mathbf{R}^f \times \mathbf{R}^f \rightarrow \mathbf{R}^{f'}$ 的特征学习。

由于该网络多次使用 K 近邻算法构建图结构，每一层所构造的图结构都不同。所以，使用 $\boldsymbol{l}_i$ 表示第 i 个顶点在第 l 个残差通道注意力层中多层感知机的输出。通过式（4）计算输出边特征 $\boldsymbol{e}_{i_j}$。局部子图 $\boldsymbol{G}$ 通过多层感知机得到局部邻域特征 $\boldsymbol{l}_i$，其中 $\boldsymbol{l}_i$ 的具体定义为

$$\boldsymbol{l}_i = ReLU(BN(\sum_{j:(i,j)\in\boldsymbol{E}} \boldsymbol{e}_{i_j})) \tag{5}$$

式中，*ReLU* 表示激活函数，*BN* 表示批处理归一化操作。

输入的点云数据经过动态图卷积层提取到局部邻域特征 $\boldsymbol{l}_i$，现有方法通常选择局部邻域特征中的最大值特征来代表该局部特征，定义为

$$\boldsymbol{L}_i = \max(\boldsymbol{l}_i) \tag{6}$$

式中，$\boldsymbol{L}_i$ 表示聚合之后的局部特征。

但是，简单地对局部邻域进行最大值池化操作难免会丢失邻域中的其他特征。因此，本文在对局部邻域特征的聚合过程中设计了针对点云的空间注意力模块，该模块顾及局部邻域内所有元素的特征，并且使用注意力机制增强局部邻域内有用的特征，弱化无用的特征。其定义为

$$\boldsymbol{L}_i = SA(\boldsymbol{l}_i) \tag{7}$$

式中，*SA* 表示空间注意力模块的操作。

2.2　空间注意力模块

注意力模块的关键是使用注意力机制计算 k 个邻居点对中心点的注意力权重。本文所设计的空间注意力模块 SA 的结构如图 4 所示。它由 4 个单元组成：1）注意力激活；2）注意力得分；3）加权特征；4）多层感知机（MLP）。首先通过全连接层学习每个潜在特征的注意力激活，然后通过 Softmax 函数计算相应特征的注意力得分。所学习的注意力分数可以视为自动选择有用的潜在特征的掩膜。随后，将注意力得分与局部邻域特征进行对应元素相乘，生成一组加权特征。最后，对加权特征进行求和运算得到有局部代表性的局部特征。

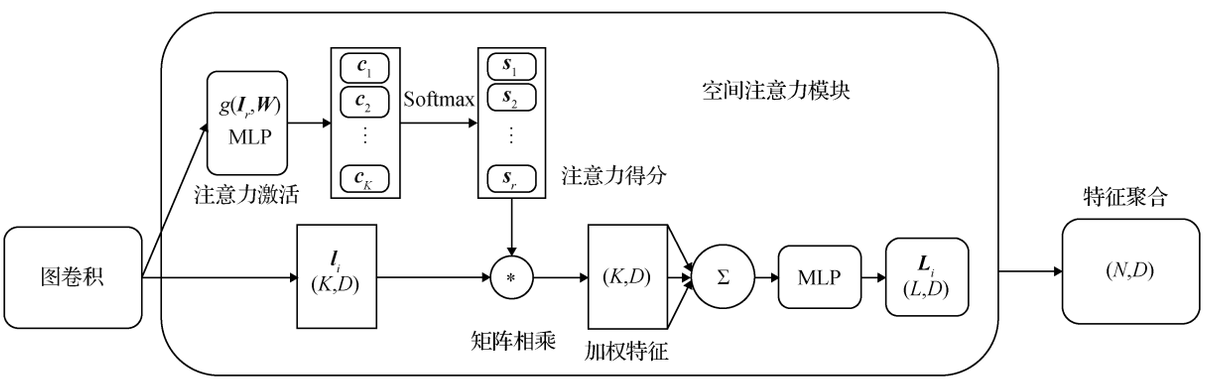


图 4　空间注意力模块

Fig. 4　Spatial attention module

1）注意力激活。如图 4 所示，给定一组局部特征集合 $\boldsymbol{l}_i = \{\boldsymbol{l}_i^1, \cdots, \boldsymbol{l}_i^k, \cdots, \boldsymbol{l}_i^K\}$，$\boldsymbol{l}_i \in \mathbf{R}^{K\times D}$，其中 K 表示 K 近邻图中的 K 个顶点，D 表示局部特征的维数。将局部特征 $\boldsymbol{l}_i$ 输入 MLP 函数 $g(\cdot)$，即全连接

层,输出为一组学习得到的注意力激活参数 $\boldsymbol{C} = \{\boldsymbol{c}_1, \boldsymbol{c}_2, \cdots, \boldsymbol{c}_K\} \in \mathbf{R}^{K\times D}$。

$$\boldsymbol{C} = g(\boldsymbol{l}_i, \boldsymbol{W}) \tag{8}$$

式中,$\boldsymbol{W}$ 表示共享的全连接层所学习的权重参数。

2)注意力得分。本文使用 Softmax 函数作为注意力激活的归一化操作,计算出一组注意力得分参数 $\boldsymbol{s} = \{\boldsymbol{s}_1, \boldsymbol{s}_2, \cdots, \boldsymbol{s}_K\} \in \mathbf{R}^{K\times D}$。第 k 个特征向量的注意力得分为

$$\boldsymbol{s}_k = \frac{\exp(\boldsymbol{c}_k)}{\sum_{j=1}^{K} \exp(\boldsymbol{c}_j)} \tag{9}$$

3)加权特征。本文使用注意力分数 $\boldsymbol{s}$ 和局部邻域特征 $\boldsymbol{l}_i$ 之间进行矩阵乘法以生成加权邻域特征,然后沿着 K 个顶点对加权邻域特征求和以获得聚合之后的局部特征 $\boldsymbol{L}_i \in \mathbf{R}^{1\times D}$,计算过程为

$$\boldsymbol{L}_i = \sum_{k=1}^{K} (\boldsymbol{l}_i^k * \boldsymbol{s}_k) \tag{10}$$

其中,* 符号表示矩阵相乘,即对应元素相乘。

4)多层感知机模块。最后,本文将 $\boldsymbol{L}_i$ 输入到 MLP 中以控制局部特征向量的维度。该 MLP 包括一个全连接层(fully connected layer, FC),归一化层(BN)和 ReLU 激活层。该层使注意力池化模块具有更大的灵活性来处理特征尺寸的减小。

通过式(8)–(10)及图 5 演示了空间注意力模块如何将 $\boldsymbol{l}_i$ 的 K 个局部邻域特征聚合到单个向量局部特征 $\boldsymbol{L}_i$ 中。

2.3 空间转换网络

PointNet(Qi 等,2017a)中提出的空间转换网络(STN)的作用是训练得到一组空间旋转矩阵,该旋转矩阵可以对输入的点云数据进行坐标对齐。空间转换网络可以直接将输入点云数据旋转到一个更好的角度,这更有利于网络对点云数据进行分类和分割。具体实现如图 5 所示,输入的点云数据通过多个 MLP 与最大值池化预测 3×3 旋转矩阵 $\boldsymbol{R}$ 并将该矩阵 $\boldsymbol{R}$ 直接与输入点云数据进行矩阵相乘来实现坐标对齐。

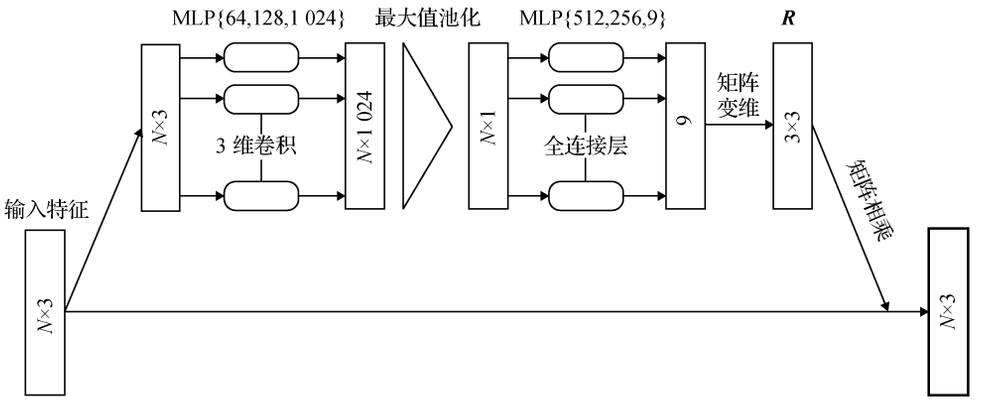


图 5 空间转换网络

Fig. 5 Spatial transformation network

3 实验验证

本文方法在分类、实例分割和室内场景语义分割任务上与其他方法进行实验对比分析。实验运行环境和参数设置如表 1 所示;模型训练的参数设置如表 2 所示。

评估指标:对于点云分类,使用两个指标来定量测试分类性能:总体分类精度(overall accuracy, OA)和每个类别的平均分类精度(mean accuracy,

表 1 实验配置

Table 1 Experimental configuration

操作系统	GPU	运算加速库	框架	语言
Linux Centos7	RTX 2080Ti	CUDA 10.1 + cuDNN7.5	Pytorch1.3	Python 3.5.2

表 2　实验参数设置

Table 2　Experimental parameters setting

数据集	点的数量	K 近邻点	优化器	学习率	批处理量	训练次数
ModelNet40	1 024	20	SGD	0.001	32	250
ShapeNetPart	2 048	40	SGD	0.001	14	200
S3DIS	4 096	20	SGD	0.001	14	100

mAcc)。对于分割任务,性能指标为:总体准确率(OA)和平均交并比(mean intersection-over-union, mIoU)。为了计算某个特定类别的 mIoU,需要对此类别中所有 3 维模型的 IoU 求平均值。其中 IoU 定义为

$$IoU = \frac{TP}{TP + FP + FN} \tag{11}$$

式中,TP 表示真阳性数,FP 表示假阳性数,FN 表示假阴性数。

3.1　在 ModelNet40 上的分类任务

3.1.1　数据集

在 ModelNet40(Wu 等,2015)数据集上评估分类网络的分类性能。ModelNet40 数据集包含来自 40 个类别的 12 311 个网格 CAD(computer aided design)模型,其中 9 843 个模型用于训练,2 468 个模型用于测试。本文参考 Qi 等人(2017b)的实验设置,对训练数据随机旋转、移动和缩放,并在每个点上添加随机噪声,从而达到对数据进行增强的效果。对于每个数据模型,从网格面采样 1 024 个点,本文使用采样点的 3 维坐标(x,y,z)作为网络的输入。

3.1.2　实验结果

为验证网络的分类性能,将 DGCSA 与 3 种基于点态 MLP 的方法、5 种基于卷积的方法和 4 种基于图的方法进行比较,如表 3 所示。这些对比网络的分类精度主要来自其原始论文实验结果,DGCNN 网络(Wang 等,2019)的结果为本地复现的实验结果。

表 3 列出了本文网络和现有网络在 ModelNet40 测试集上的总体分类精度和平均分类精度。从表 3 中可以明显看出,无论是总体分类精度还是平均分类精度,本文网络都达到了最好的性能,OA 和 mAcc 分别达到 93.4% 和 90.6% 。在总体分类精度方面,本文网络比 DGCNN 高出了 0.8% ,比 PointNet 高出了 4.2% ,比 PointnNet ++ 高出了 1.5% 。同时,本文网络在获得最高分类精度的情况下,网络的输入仅为 1 024 个点的 3 维坐标。通过对比实验结果,证明了本文网络在点云分类任务上具有较强的性能。

表 3　在 ModelNet40 数据集上的分类结果实验对比

Table 3　Comparison of classification results on ModelNet40 dataset

/%

模型	OA	mAcc
PointNet(Qi 等,2017a)	89.2	86.2
PointNet ++ (Qi 等,2017b)	91.9	–
PointWeb(Zhao 等,2019)	92.3	89.4
PointConv (Wu 等,2019)	92.5	–
PointCNN(Li 等,2018)	92.2	88.1
KPConv Rigid(Thomas 等, 2019)	92.9	–
KPConv deform(Thomas 等, 2019)	92.7	–
SpiderCNN(Xu 等,2018)	92.4	–
ECC(Simonovsky 等,2017)	87.4	83.2
RGCNN(Te 等,2018)	90.5	87.2
LDGCNN(Zhang 等,2019)	92.9	90.3
DGCNN(Wang 等,2019)	92.6	90.1
DGCSA (本文)	**93.4**	**90.6**

注:加粗字体为每列最优值。

3.2　实例分割

3.2.1　数据集描述

与点云分类任务相比,实例分割是一项非常具有挑战性的 3D 细粒度分割任务,其目的是给定一个已知类别的模型,得出模型中每一个点所属零部件的类别标签得分。例如,输入一个飞机模型,将其分为机身、机尾和机翼等。本文网络在 ShapeNetPart 数据集(Yi 等,2016)上验证了其优越的实例分割性能。ShapeNetPart 数据集包含 16 个对象类的 16 881 个 3 维模型,注释了 50 个部件标签。大多数 3 维模型

被标记为 2 ～5 个标签。在该实验中，从每个 3 维模型中采样 2 048 个点。为确保公平，本文采用与 Qi 等人(2017b)方法同样的训练、验证和测试数据集拆分方案。

3.2.2　实验结果

将本文方法与 PointNet、PointNet ++ 和 DGCNN 进行了比较。ShapeNetPart 实例分割数据集上的分割结果的性能对比如表 4 第 1 列所示，其中列出了 16 个类别的交并比和平均交并比 mIoU。在表 4 中，第 1 行类别后的数字表示该类别在数据集中的数量。通过对比实验结果，可以发现本文网络的 mIoU 高出 PointNet 方法 1.6%，高出 PointNet ++ 和 DGCNN 方法 0.2%。本文网络在 8 个类别的 IoU 取得 4 种方法的最高值，实例分割总体性能较好。

为了直观展示实例分割在各个类别上的分割效果，本文可视化展示了一些实例分割效果，如图 6 所示，其中不同颜色代表不同的部件类别。以真值为基准，通过对比部件边缘和部件连接处的分割细节，可以看出本文网络在实例分割的易混淆细节上的分割效果要优于 DGCNN。

表 4　ShapeNetPart 数据集上的实例分割 IoU 结果对比

Table 4　IoU results of instance segmentation on ShapeNetPart dateset

/%

模型	mIoU	IoU														
		aero (2 690)	bag (76)	cap (55)	car (898)	chair (3 758)	ear phone (69)	guitar (787)	knife (392)	lamp (1 547)	laptop (451)	motor (202)	mug (184)	pistol (283)	rocket (66)	skate board (152)	table (5 271)
PointNet	83.7	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
PointNet ++	85.1	82.4	79.0	**87.7**	77.3	90.8	71.8	91.0	85.9	83.7	95.3	**71.6**	91.4	81.3	58.7	**76.4**	82.6
DGCNN	85.1	84.0	**83.7**	84.4	**77.8**	90.6	74.4	91.0	88.1	83.4	95.8	67.8	93.3	**82.3**	**59.2**	76.0	81.9
DGCSA (本文)	**85.3**	**84.2**	73.3	82.3	77.7	**91.0**	**75.3**	91.2	**88.6**	**85.3**	**95.9**	58.9	**94.3**	81.8	56.9	75.4	**82.7**

注：加粗字体为每列最优值。

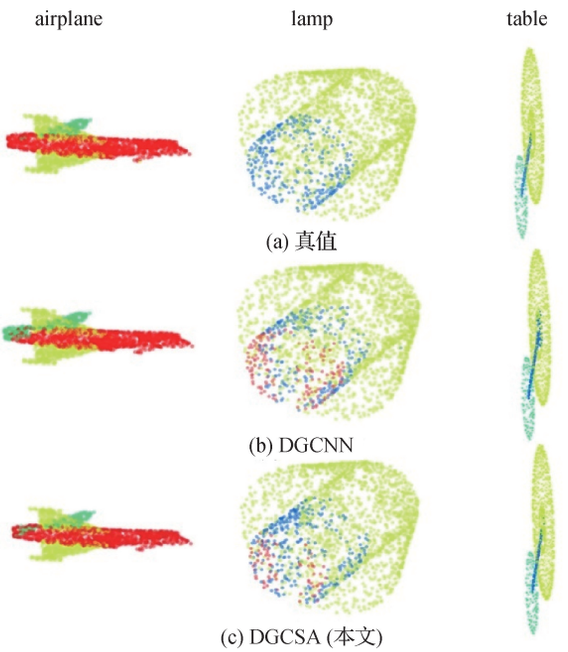


图 6　实例分割的效果对比

Fig. 6　Comparison of instance segmentation results ((a) ground truths; (b) DGCNN; (c) DGCSA (ours))

3.3　语义场景分割

3.3.1　数据集

针对场景分割任务，采用了斯坦福大学大型 3D 室内空间数据集 S3DIS(Armeni 等，2016)验证本文网络对室内场景的分割性能。S3DIS 数据集是由 Matterport 扫描仪获得的室内场景 3 维模型，其中包括来自 6 个区域(Area1，Area2，Area3，Area4，Area5，Area6)的 271 个房间的 3 维点云数据。共 13 个类别，包括 ceiling、floor、door、chair 等。为了确保测试结果的可比性，使用与 PointNet ++ 相同的设置：将每个房间划分为 1 m × 1 m 的块，每个点都由 9 维向量(XYZ，RGB 和归一化的空间坐标)表示，在训练过程中对每个块区域随机采样 4 096 个点作为输入。本文做了两组实验，第 1 组实验：在 S3DIS 数据集的 Area1—Area6 上进行训练，在 Area5 上进行测试。第 2 组实验：所有的数据都用于训练，并在 Area1—Area6 所有室内场景中进行 6 折交叉验证。

3.3.2 实验结果

表5中展示了Area5的评估结果。从表5中可以看出本文网络在Area5中mIoU达到50.1%,比DGCNN高出3.0%,比PointNet高出9.01%。PointNet++也获得了很好的结果(50.04%),这是由于PointNet++可以从局部几何结构中学习特征,并逐层抽象局部特征。

表6为6折交叉验证结果,展现了不同方法在S3DIS数据集中的总体分类精度与平均交并比。其中本文网络的分割性能优势也非常显著,在平均交并比(mIoU)方面比DGCNN高3.0%,比PointNet++高4.6%,比PointNet高11.5%。在总体分类精度方面,本文网络也是最高的,达到了85.1%。

表5 S3DIS数据集上的3D场景分割结果(在Area5上测试)

Table 5 Results of semantic segmentation of 3D indoor scenes on S3DIS (test on Area5)

/%

模型	mIoU	IoU												
		ceiling	floor	wall	bean	column	window	door	chair	table	bookcase	sofa	board	clutter
PointNet	41.09	88.80	97.33	69.80	0.05	3.92	46.26	10.76	52.61	58.93	**40.28**	5.85	26.38	33.22
PointNet++	50.04	90.79	96.45	74.12	0.02	5.77	43.59	25.39	69.22	**76.94**	21.45	55.61	**49.34**	41.88
DGCNN (baseline)	47.08	92.42	97.46	76.03	**0.37**	12.00	**51.59**	27.01	64.85	68.58	7.67	43.76	29.44	40.83
DGCSA (本文)	**50.10**	**93.21**	**97.70**	**77.04**	0.29	**15.13**	50.70	**27.90**	**69.74**	69.00	13.90	**56.38**	44.29	**45.00**

注:加粗字体为每列最优值。

表6 S3DIS室内场景分割实验结果(6折交叉检验)

Table 6 Results of semantic segmentation of 3D indoor scenes on S3DIS (6-fold cross validation)

/%

模型	mIoU	OA
PointNet	47.6	78.6
PointNet++	54.5	81.0
DGCNN(baseline)	56.1	84.1
DGCSA (本文)	**59.1**	**85.1**

注:加粗字体为每列最优值。

表7 消融实验(ModelNet40分类任务)

Table 7 Ablation study (classification on ModelNet40)

/%

模型	OA
PointNet	89.2
PointNet使用SA模块进行优化	89.7(↑0.5)
LDGCNN	92.7
LDGCNN使用SA模块进行优化	93.3(↑0.6)

注:"↑"表示比未优化方法的结果提升。

3.4 消融实验

为了验证空间注意力模块的有效性以及相对于最大值池化的优越性,本文将空间注意力模块应用于PointNet和LDGCNN网络中替换其中的最大值池化操作,并在ModelNet40数据集上对比分类精度。实验结果如表7所示。LDGCNN在ModelNet40的分类任务中分类精度为92.7%,使用空间注意力SA模块替换最大值池化使得LDGCNN的分类精度提高了0.6%,达到93.3%。使用空间注意力模块优化的PointNet网络在ModelNet40数据集上的分类精度相对于PointNet提高了0.5%。消融实验的结果对比表明,空间注意力模块相对于最大值池化而言具有更好的特征聚合能力。

4 结 论

针对现有点云分类分割网络简单地使用局部邻域中的最大值特征代表局部特征所带来的部分信息丢失问题,本文通过设计空间注意力模块,在聚合局部邻域信息的过程中,充分考虑了所有邻域特征,并且有选择性地加强包含有用信息的特征同时抑制无用特征,避免部分信息丢失。将空间注意力模块与动态图卷积的特征提取相结合,不仅可以获取更加强的表征,而且通过增强特征和减少信息丢失来提

高模型的分类分割准确性。从实验结果来看，DGC-SA 可以显著提高分类、实例分割和室内场景分割的准确性。通过对实验结果的可视化结果来分析，本文方法对分割细节的处理要优于现有方法。针对注意力模块的消融实验也表明，该模块与其他点云分类模型相结合也能有效提高模型性能。

然而，在实例分割与室内场景分割任务的实验结果中存在某些类别的分割精度相对较低的现象。经过分析总结出两个原因：1）实例分割数据集中存在严重的样本不均衡现象，使得训练出的模型对特定的类别存在一定偏好；2）本文方法尚未针对点云数据集中分割样本不均衡的问题进行优化。

综上所述，下一步研究工作的重点是提升样本不均衡情况下的分割模型精度，具体的实现过程为优化点云分类与分割网络的损失函数并采用基于多模型网络结合不同方法中的特征，从而提高网络的特征提取能力。与此同时，采用最远点采样的方法减少模型的内存消耗并提高模型的推理速度。

参考文献（References）

Armeni I, Sener O, Zamir A R, Jiang H L, Brilakis I, Fischer M and Savarese S. 2016. 3D semantic parsing of large-scale indoor spaces//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE: 1534-1543 [DOI: 10.1109/CVPR.2016.170]

Besl P J and Jain R C. 1988. Segmentation through variable-order surface fitting. IEEE Transactions on Pattern Analysis and Machine Intelligence, 10(2): 167-192 [DOI: 10.1109/34.3881]

Bruna J, Zaremba W, Szlam A and LeCun Y. 2013. Spectral networks and locally connected networks on graphs [EB/OL]. [2020-08-15]. https://arxiv.org/pdf/1312.6203.pdf

Chen X Z, Ma H M, Wan J, Li B and Xia T. 2017. Multi-view 3D object detection network for autonomous driving//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE: 6526-6534 [DOI: 10.1109/CVPR.2017.691]

Defferrard M, Bresson X and Vandergheynst P. 2016. Convolutional neural networks on graphs with fast localized spectral filtering//Proceedings of the 30th International Conference on Neural Information Processing Systems. Barcelona, Spain: NIPS: 3844-3852 [DOI: 10.5555/3157382.3157527]

Filin S and Pfeifer N. 2006. Segmentation of airborne laser scanning data using a slope adaptive neighborhood. ISPRS Journal of Photogrammetry and Remote Sensing, 60(2): 71-80 [DOI: 10.1016/j.isprsjprs.2005.10.005]

Fischler M A and Bolles R C. 1981. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Communications of the ACM, 24(6): 381-395 [DOI: 10.1145/358669.358692]

Guo Y L, Wang H Y, Hu Q Y, Liu H, Liu L and Bennamoun M. 2020. Deep learning for 3D point clouds: a survey. IEEE Transactions on Pattern Analysis and Machine Intelligence [DOI: 10.1109/TPAMI.2020.3005434]

Kipf T N and Welling M. 2016. Semi-supervised classification with graph convolutional networks [EB/OL]. [2020-08-15]. https://arxiv.org/pdf/1609.02907.pdf

LeCun Y, Bengio Y and Hinton G. 2015 Deep learning. Nature, 2015, 521(7553): 436-444 [DOI: 10.1038/nature14539]

Li Y Y, Bu R, Sun M C, Wu W, Di X H and Chen B Q. 2018. PointCNN: convolution on *X*-transformed points//Proceedings of the 32nd International Conference on Neural Information Processing Systems. Montréal, Canada: NeurIPS: 828-838

Qi C R, Su H, Mo K C 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: NIPS: 5105-5114 [DOI: 10.5555/3295222.3295263]

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]

Su H, Maji S, Kalogerakis E and Learned-Miller E. 2015. Multi-view convolutional neural networks for 3D shape recognition//Proceedings of 2015 IEEE International Conference on Computer Vision. Santiago, Chile: IEEE: 945-953 [DOI: 10.1109/ICCV.2015.114]

Te G S, Hu W, Zheng A M and Guo Z M. 2018. RGCNN: regularized graph CNN for point cloud segmentation//Proceedings of the 26th ACM International Conference on Multimedia. Seoul, Korea (South): ACM: 746-754 [DOI: 10.1145/3240508.3240621]

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 C, Samari B and Siddiqi K. 2018. Local spectral graph convolution for point set feature learning//Proceedings of the 15th European Conference on Computer Vision. Munich, Germany: Springer: 56-71 [DOI: 10.1007/978-3-030-01225-0_4]

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, 1-12 [DOI: 10.1145/3326362]

Wu W X, Qi Z G and Li F X. 2019. 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]

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]

Yi L, Kim V G, Ceylan D, Shen I C, Yan M Y, Su H, Lu C W, Huang Q X, Sheffer A and Guibas L. 2016. A scalable active framework for region annotation in 3D shape collections. ACM Transactions on Graphics, 35(6): 210, 1-12 [DOI: 10.1145/2980179.2980238]

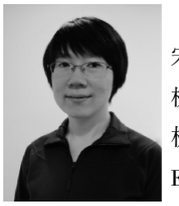
Zhang K G, Hao M, Wang J, de Silva C W and Fu C L. 2019. Linked dynamic graph CNN: learning on point cloud via linking hierarchical features. [EB/OL]. [2020-08-15]. https://arxiv.org/pdf/1904.10014.pdf

Zhang Y X and Rabbat M. 2018. A graph-CNN for 3D point cloud classification//Proceedings of 2018 IEEE International Conference on Acoustics, Speech and Signal Processing. Calgary, Canada: IEEE: 6279-6283 [DOI: 10.1109/ICASSP.2018.8462291]

Zhao H S, Jiang L, Fu C W and Jia J Y. 2019. PointWeb: enhancing local neighborhood features for point cloud processing//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE: 5560-5568 [DOI: 10.1109/CVPR.2019.00571]

Zhou Y and Tuzel O. 2018. Voxelnet: end-to-end learning for point cloud based 3D object detection//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 4490-4499 [DOI: 10.1109/CVPR.2018.00472]

作者简介



宋巍,1977 年生,女,教授,主要研究方向为机器视觉、图像/视频处理、海洋大数据分析。
E-mail: wsong@shou.edu.cn

李文俊,通信作者,男,博士,工程师,主要研究方向为遥感、视通数据分析。
E-mail: liwenjun@pric.org.cn

蔡万源,男,硕士研究生,主要研究方向为计算机视觉、3 维点云。E-mail:865645064@qq.com
何盛琪,男,博士,工程师,主要研究方向为 3 维建模、海洋信息工程。E-mail: sqhe@shou.edu.cn