计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.190700180

基于动态图卷积和空间金字塔池化的点云深度学习网络

朱　威[1,2]　绳荣金[1]　汤　如[1]　何德峰[1,2]
1 浙江工业大学信息工程学院　杭州 310023
2 浙江省嵌入式系统联合重点实验室　杭州 310023

摘　要　点云数据的分类和语义分割在自动驾驶、智能机器人、全息投影等领域中有着重要应用。传统手工提取点云特征的方式,以及将三维点云数据转化为多视图、体素网格等数据形式后再进行特征学习的方式,都存在处理环节多、三维特征损失大等问题,分类和分割的精度较低。目前可以直接处理点云数据的深度神经网络 PointNet 忽略了点云的局部细粒度特征,对复杂点云场景的处理能力较弱。针对上述问题,提出了一种基于动态图卷积和空间金字塔池化的点云深度学习网络。该网络在 PointNet 的基础上使用动态图卷积模块来替换 PointNet 中的特征学习模块,增强了网络对局部拓扑结构信息的学习能力;同时设计了一种基于点的空间金字塔池化结构来捕获多尺度局部特征,该方式比 PointNet++的多尺度采样点云、重复分组进行多尺度局部特征学习的方法更加简洁高效。实验结果表明,在点云分类和语义分割任务的 3 个基准数据集上,所提网络相较于现有网络具有更高的分类和分割精度。

关键词:点云;PointNet;动态图卷积;空间金字塔池化;局部特征
中图法分类号　TP391

Point Cloud Deep Learning Network Based on Dynamic Graph Convolution and Spatial Pyramid Pooling

ZHU Wei[1,2],SHENG Rong-jin[1],TANG Ru[1] and HE De-feng[1,2]
1 College of Information Engineering,Zhejiang University of Technology,Hangzhou 310023,China
2 United Key Laboratory of Embedded System of Zhejiang Province,Hangzhou 310023,China

Abstract　The classification and semantic segmentation of point cloud data have important applications in automatic driving,intelligent robot and holographic projection. While using the traditional method of manually extracting point cloud features or the feature learning method of firstly transforming three-dimensional point cloud data into data forms of multi-view and volumetric grid,there exist problems such as many processing links and great loss of three-dimensional features,resulting in low accuracy of classification and segmentation. The existing deep neural network PointNet,which can directly process point cloud data,ignores the local fine-grained features of point cloud and is weak in processing complex point cloud scenarios. To solve the above problems,this paper proposes a point cloud deep learning network based on dynamic graph convolution and spatial pyramid pooling. On the basis of PointNet,the dynamic graph convolution module GraphConv is used to replace the feature learning module in PointNet,which enhances the network's ability to learn local topological structure information. At the same time,a point-based spatial pyramid pooling structure PSPP is designed to capture multi-scale local features. Compared with the multi-scale sampling point cloud of PointNet++ and the repeated grouping method for multi-scale local features learning,it is simpler and more efficient. Experimental results show that,on the three benchmark data sets of point cloud classification and semantic segmentation task,the proposed network has higher classification and segmentation accuracy than the existing network.

Keywords　Point cloud,PointNet,Dynamic graph convolution,Spatial pyramid pooling,Local features

到稿日期:2019-07-26　返修日期:2019-12-30　本文已加入开放科学计划(OSID),请扫描上方二维码获取补充信息。
基金项目:浙江省自然科学基金(LY17F010013);国家自然科学基金(61401398)
This work was supported by the Natural Science Foundation of Zhejiang Province (LY17F010013) and National Natural Science Foundation of China (61401398).
通信作者:朱威(weizhu@zjut.edu.cn)

1　引言

近年来，点云数据在自动驾驶、智能机器人、全息投影等领域中被广泛应用并发挥着重要作用。点云数据的处理任务主要包括分类和语义分割等。由于点云数据具有数据量大、形状不规则、密度不均匀等特点，因此点云数据的分类和语义分割一直是一个极具挑战性的难题[1-2]，也是目前的一个研究热点[2-7]。

在对点云数据进行分类和语义分割时，首要任务是提取点云特征。通过手工提取特征[1]的传统方式依赖于手工制作和精心设计的优化方法，同时手工提取特征会丢失点云的大量有用信息，导致分类和分割的精度不高，因此该类方法的效果已达到了瓶颈。虽然深度学习方法在图像特征提取上取得了巨大成功，但是，由于点云具有形状不规则和密度不均匀等特性，传统的卷积神经网络(Convolutional Neural Networks，CNN)无法直接处理输入的非结构化原始点云数据，因此近年来出现了一系列基于深度卷积神经网络的非直接学习的点云特征提取方案。一种典型的方案是先将三维(3D)数据表示为多视图的形式[3,8]，然后对各个视图进行二维特征学习，再将获得的视图特征拼接起来，以得到近似完整的三维点云特征。然而，多视图的方法实质上是将三维感知转化为二维感知，并没有直接对三维数据进行学习；另外，将通过二维数据形式学习到的特征重投影到三维空间会引发高计算复杂度和拼接误差等新问题，导致三维数据的完整性表达存在较大的缺陷。另一种直观的方法是将非结构化的点云数据转化为规则化的体素网格数据[9-10]，并使用三维卷积神经网络(3D-Convolutional Neural Networks，3D-CNN)学习体素级的特征信息。由于点云具有密度不均匀性和稀疏性，因此在稀疏的点云空间中进行体素化的效率较低，且忽略了细节特征；在点云稠密的区域，由于相同体素内的所有点都被无区别地赋予相同的语义标签，导致语义分割的精度有限。此外，通过先体素化点云再学习的方法在本质上不具有旋转不变性，点云体素化时存在内存占用率高和计算量大的缺点，大大提升了算法的空间复杂度和时间复杂度，因此这类方法不太适合高数据量、大场景点云数据的特征学习。

直到2017年，Qi等提出了点云神经网络PointNet[2]，其开创性地直接在非结构化的原始点云数据上进行特征学习。该网络首先训练T-Net转换网络[11]，生成一个空间转换矩阵，以解决点云的旋转性问题，并用多层感知器(Multi-Layer Perceptron，MLP)分别学习每个点的高维度特征；然后将最大池化函数作为一个简单的对称函数，提取点云的全局特征，解决了点云的无序性问题。尽管PointNet在点云分类和语义分割上表现出了不错的性能，但是，其在处理局部区域点时对点的处理过于独立，忽视了点与点之间的几何关系，因此不能获取由相邻点构成的局部细粒度特征信息。因此，PointNet在点云局部特征提取能力上的缺失导致其在复杂点云场景上的性能表现不佳。为解决此问题，Qi等提出了基于PointNet的改进版——PointNet++[12]。该网络首先随机选取一定数目的点作为每个局部区域的中心，然后在这些中心点的多尺度球半径内选取k个欧氏距离的近邻点进行最远点采样，将这些采样点进行分组后通过mini-PointNet网络进行特征学习与提取。通过不断地重复采样和分组特征学习，构建了一个有层次结构的点云特征学习网络来学习点云的局部特征，并将此方法推广到非均匀密度的点云分割中。然而，PointNet++计算复杂，运行耗时极长；其本质上与PointNet一样，独立地处理局部点集中的各个点，并且不考虑点对之间以及具有相关性的点之间的深层次特征关系。最近，由于图卷积在二维图像处理方面的应用[13]取得了较好的效果，基于图的动态边缘卷积神经网络ECC(Dynamic edge-conditioned filters in convolutional neural networks on graphs)[14]首次尝试将图卷积用于三维点云的分割；然而，其权重是由相邻点间的边权决定，而不是图结构，因此效果并不是很理想。Wang等提出的DGCNN(Dynamic Graph CNN for Learning on Point Clouds)[15]在ECC动态边缘卷积方法的基础上考虑了点的坐标以及邻域点间的距离信息，但由于忽略了相邻点之间的向量方向，其最终还是损失了部分局部几何信息。GAPNet[6]通过在多层感知器层中嵌入图注意机制来获得局部几何表示，以增强网络对局部几何结构的提取能力，但实验结果显示其性能并无明显提高。

为了解决上述问题，尤其是提升网络在局部邻域特征学习方面的性能，本文在PointNet网络结构和图卷积理论的基础上提出了一种基于动态图卷积和空间金字塔池化的点云深度学习网络。与PointNet相比，本文网络对特征提取方式进行了改进，将PointNet中直接通过MLP对全局点云进行特征提取的方式改为通过把点云数据构造为有顶点和边的图再进行卷积的图卷积方式，这主要是由于图卷积可在保持点云置换不变性的同时捕获点云的局部几何结构，从而弥补了PointNet在局部点云特征提取上的不足，增强了网络对点云局部区域细粒度特征的提取能力。为了继续强化网络对局部特征的提取能力，本文在提出的网络中设计了空间金字塔池化结构，对学习的特征分空间按层次地进行多尺度池化，从而得到全局特征和不同尺度的局部特征信息。实验表明，这种网络设计提高了分类和分割的精度，效果也优于普通的图卷积网络。

2　点云深度学习网络

2.1　网络结构

本文提出的点云深度学习网络首先采用动态图卷积结构(GraphConv)来替换PointNet网络中用MLP构造的特征提取结构。图卷积由于是通过学习所选取顶点与其相邻点之间边的关系来获得局部几何特征，因此对相邻点的学习具有顺序不变性，GraphConv可以在保持点云置换不变性的同时捕获点云的局部几何特征，弥补了PointNet对局部点云特征提取的不足。此外，所提网络中设计了一个多窗口和窗口步长

的点云空间金字塔池化(Point Cloud Spatial Pyramid Pooling,PSPP)结构,该结构弥补了 PointNet 对称函数模块单一窗口的 Max pooling 的不足。通过 PSPP 可以同时得到点云的全局特征和局部特征,再次增强了网络对局部几何特征信息的捕获能力。

点云深度学习网络的总体网络结构如图 1 所示,其中 N 为采样点的个数,D 为每个点的数据维度,$D=3$ 代表输入三维的点云数据。输入 $N\times D$ 维的点云经过可训练的空间转换网络 Spatial T-Net,其处理流程如图 1 左上方虚线框所示。首先将输入点云经 Spatial T-Net 网络训练得到的空间转换矩阵进行坐标对齐,然后输入到 GraphConv 中提取特征。GraphConv 的详细结构如图 1 左下方虚线框所示,其中 k-NN graph 表示用 k 近邻点选取图的范围,$mlp(L_1,L_2,\cdots,L_n)$ 代表具有 n 个共享权值的 MLP 层,用来提取经过 k-NN 图结构的边缘信息,经过多层图卷积进行特征提取后得到 $N\times k\times L_n$ 维度的特征信息。第一组图卷积 GraphConv 包含 3 个 MLP 层,$mlp(64,64,64)$ 最终输出 $N\times 64$ 维度的特征信息;经过第二组 GraphConv 后输出 $N\times 1024$ 维度的特征信息;再将此特征信息经过 PSPP 进行多尺度池化,使其既包含多尺度的局部区域特征,又包含全局特征,以用于后续的分类和分割网络。如图 1 所示,中间的分类网络结构和下方的分割网络结构的最大区别是,分割网络需要将经金字塔池化得到的特征与低维度的特征进行融合后再输入全连接层,以得到每个点的标签分类,实现语义分割。其中,分类网络输出整个点云属于 k 类的得分,分割网络则输出点云中每一点所属 m 个类别的得分情况。

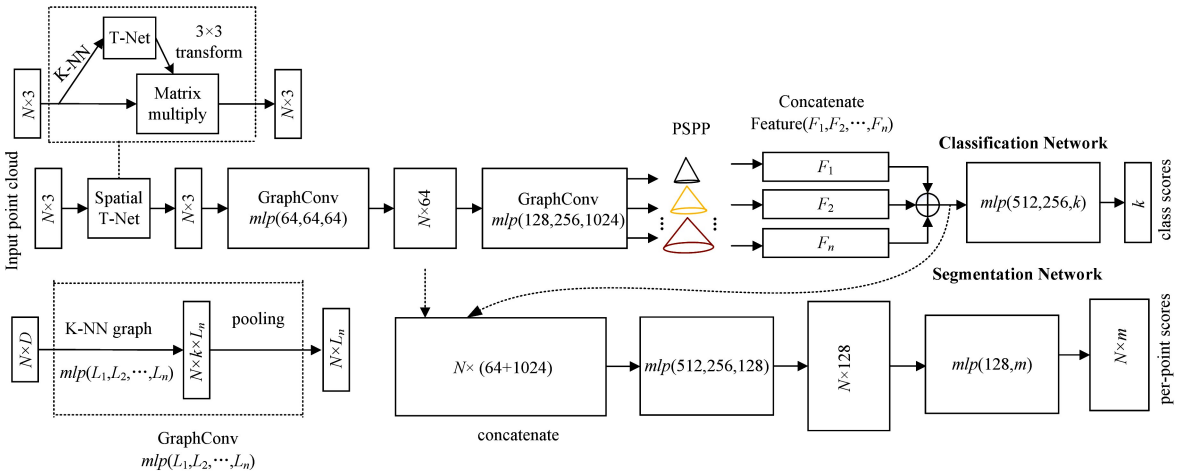

图 1 基于动态图卷积和空间金字塔池化的点云深度学习网络总体结构

Fig. 1 Overall structure of point cloud deep learning network based on dynamic graph convolution and spatial pyramid pooling

2.2 动态图卷积

把数据构造成有边和顶点的图,然后在图数据上进行卷积的操作被称为图卷积。图卷积分为基于谱的图卷积和基于空间的图卷积,而本文中的图卷积是基于空间的图卷积。

首先,将一个点数为 n、维度为 D 的点云表示为:

$$X=\{x_1,\cdots,x_n\}\subseteq R^D \tag{1}$$

当 $D=3$ 时,表示输入具有 u,v,w 3 个坐标维度的点云,用式(2)表示;若输入带有 r,g,b 信息的彩色点云,则 $D=6$,用式(3)表示;若输入带有曲面法向量 α,β,γ,则 $D=9$,可表示为式(4)。式(2)—式(4)中,i 对应点云中的第 i 个点。

$$x_i=(u_i,v_i,w_i) \tag{2}$$

$$x_i=(u_i,v_i,w_i,r_i,g_i,b_i) \tag{3}$$

$$x_i=(u_i,v_i,w_i,r_i,g_i,b_i,\alpha_i,\beta_i,\gamma_i) \tag{4}$$

本文用一个有向图来代表局部点云结构。有向图用式(5)表示,其中 $V=\{1,\cdots,n\}$ 代表图结构的顶点,$E\subseteq V\times V$ 表示图结构的边。

$$G=(V,E) \tag{5}$$

首先,通过 k-NN 图的形式在每一层的点云 R^D 中构建图结构 G。对于点 x_i,找到其 k 个近邻点(用 $x_{j_{i1}},\cdots,x_{j_{ik}}$ 表示),其中 x_i 与邻近点之间的定向边为 $(i,j_{i1}),\cdots,(i,j_{ik})$。定义边的特征为式(6),其中 h_Θ 是被可学习的参数 Θ 的集合参数化的一系列非线性函数,用来完成 $R^D\times R^D\rightarrow R^{D'}$ 的特征学习。类比图像中的卷积运算,可以把顶点 x_i 看作中心像素,只不过此处的 x_i 是三维的,捕获所有的 x_i 来构成点云全局信息。$\{x_j:(i,j)\in E\}$ 是环绕在 x_i 周围的点云块,x_j-x_i 代表 x_i 邻近点构成的局部块信息。简言之,输入具有 N 个点的 D 维点云,通过图卷积后,输出具有 D' 维点云特征的 N 个点。

$$e_{ij}=h_\Theta(x_i,x_j) \tag{6}$$

边缘函数和聚合操作的选择对 GraphConv 层生成的属性具有重要影响。PointNet 网络可被看作 GraphConv 最简单的特殊形式,即点与点之间为空边信息的图卷积模式,边缘函数如式(7)所示,则第 i 个点输出的特征如式(8)所示。

$$e_{ij}=h_\Theta(x_i) \tag{7}$$

$$x_i'=\sum_{i=1,\cdots,n}h_\Theta(x_i) \tag{8}$$

式(7)中边缘函数的形式只考虑所有点构成的全局信息,忽略了局部几何信息。聚合特征的操作函数为最大池化函数固定单窗口尺度模式。本文同时考虑点云的全局信息和局部信息,边缘函数的定义如式(9)所示,第 i 个点输出的特征如式(10)所示,其中 $\sum$ 表示对学习的特征进行基于空间金字塔池化的聚合运算。

$$e_{ij}=h_\Theta(x_i,x_j-x_i) \tag{9}$$

$$x_i'=\sum_{j:(i,j)\in\varepsilon} h_\Theta(x_i,x_j-x_i) \tag{10}$$

如图 2 所示，首先，对于选定顶点 x_i，以 k 近邻的方式选择其 k 个邻近点。以 $k=5$ 为例，所选取的邻近点在图中以黄色标识，分别为 $x_{j_{i1}}$，$x_{j_{i2}}$，$x_{j_{i3}}$，$x_{j_{i4}}$，$x_{j_{i5}}$。其次，通过 GraphConv 学习得到顶点和邻近点之间的边特征，边特征用 e_{ij} 表示。最后，将得到的所有边特征聚合起来。

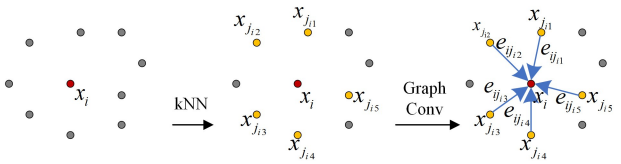

图 2　基于 k-NN 的 GraphConv 示意图(电子版为彩色)

Fig. 2　GraphConv diagram based on k-NN

普通的静态图卷积在网络中的每一层都应用一个固定图，并在图上应用卷积操作；而本文的图卷积是动态不固定的，一个点的 k 近邻点在不同的网络层中随着特征情况而变化，k 近邻点的变化使得每一层的图的参数也随着层的不同而进行动态更新，从而构成了动态的图卷积网络。第 l 层的输出可简单表示为：

$$X^{(l)}=\{x_1^{(l)},\cdots,x_n^{(l)}\}\subseteq R^{D_l} \tag{11}$$

原始的点云输入可以表示为 $X^{(0)}$。每一层的点云及其特征都对应一个不同的图，用式(12)表示，因此 D 维度的第 l 层的顶点和 k_l 个近邻点经过 GraphConv 学习到边缘信息，并且经过聚合操作后得到的 D 维度的第($l+1$)层的特征输出可以表示为式(13)，其中 $h^{(l)}$ 表示特征学习过程 $R^{Dl}\times R^{Dl}\rightarrow R^{D(l+1)}$。如此一来，GraphConv 不仅学会了如何提取局部几何特征，而且学会了如何在点云中对点进行分组。

$$G=(V^{(l)},E^{(l)}) \tag{12}$$

$$x_i^{(l+1)}=\sum_{j:(i,j)\in\varepsilon^{(l)}} h_\Theta^{(l)}(x_i^{(l)},x_j^{(l)}) \tag{13}$$

2.3　点云的空间金字塔池化

在 PointNet 结构中，用 Max pooling 对称函数对(N，D)($D=1\,024$)维度的特征进行池化，得到(1，D)维度的全局特征，然后将此全局特征复制到 N 个点中的每个点，再经过多个全连接层后得到每个点的语义得分。PointNet 的操作流程可用式(14)概括表示，从中可以看出 PointNet 通过多层感知器对点云的操作是逐点进行的。其中 g 代表对称函数，用 Max pooling 函数操作实现；c 代表高维度特征和低维度特征的融合，融合后再通过多个 MLP 实现每个点的语义信息的输出。

$$F(x_1,\cdots,x_N)=\underset{i=1,\cdots,N}{mlp}\left(c\left\{g\left\{\underset{i=1,\cdots,N}{mlp}(x_i)\right\},\ \underset{i=1,\cdots,N}{mlp}(x_i)\right\}\right) \tag{14}$$

由于 PointNet 池化的窗口大小是 N，仅可以归纳出全局特征，缺少对点云局部细粒度特征的描述，因此，受 SSP[16] 像素域的二维空间金字塔池化的启发，本文设计了一种适用于三维点云的空间金字塔池化模型 PSPP。

PointNet 用固定最大窗口的池化，只能得到固定维度大小的全局特征信息；而 PSPP 采用多窗口、特定步长的池化形式，可得到任意维度、兼具全局信息和局部信息的特征，如图 3 所示。PSPP 也可以用式(15)表示：

$$G(x_1,\cdots,x_N)=\underset{x=x_1,\cdots,x_N}{mlp}\left[c\{g(f,s_1),\cdots,g(f,s_n)\}\right] \tag{15}$$

其中，s_n 表示池化的窗口尺寸；f 代表通过由多层 MLP 构建的图卷积网络学习得到的高层次特征信息；g 代表最大池化操作；c 代表将空间金字塔池化得到的多尺度特征进行合并。

在图 3 中，当输入 $N\times D$ 的点云特征时，图中右侧蓝色条形块表示点的 D 维度特征信息，不同大小和颜色的圆锥体则表示大小为 N/s_1，N/s_2，…，N/s_n 的金字塔池化窗口，池化步长与窗口的大小相同。PSPP 可聚合不同尺寸的池化特征，得到 $(s_1+s_2+\cdots+s_n)\times D$ 维度特征，其中 $s_n\times D$ 代表 N/s_n 窗口池化得到的特征，同时对应图 1 网络结构中表示的特征 F_n。因此，经 PSPP 得到的特征信息既包含多尺度的局部区域特征，又包含全局特征。

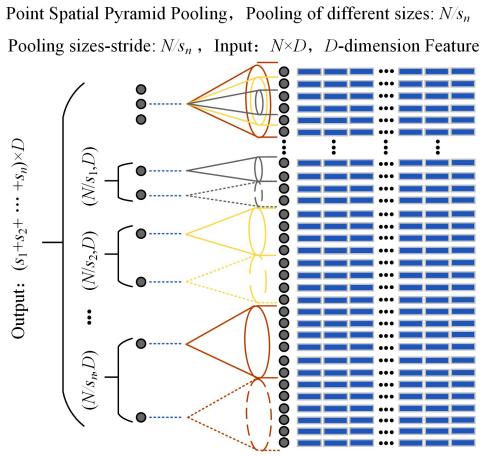

图 3　PSPP 模型示意图(电子版为彩色)

Fig. 3　Schematic diagram of PSPP model

PSPP 通过在多尺度邻域内对不同尺度规模的点进行金字塔池化汇聚，得到不同邻域的局部特征。每个点由局部特定的金字塔池化特征表示，与 PointNet 复制全局池化特征作为所有点的特征不同，增强了网络对点云局部特征的学习能力，并且 PSPP 还具有非参数化的优势，比 PointNet++的效率更高。

3　实验结果与分析

3.1　实验设置

实验分为目标分类、部件分割和场景语义分割 3 部分。实验配置以及实验参数设置如表 1、表 2 所列。表 2 中 ModelNet40[10] 数据集及网络参数设置对应 3.2 节的 3D 目标分类实验；ShapeNet[17] 数据集及网络参数设置对应 3.3 节的 3D 目标分类实验；S3DIS[18] 数据集及网络参数设置对应 3.4 节的场景语义分割实验。

表 1　实验配置

Table 1　Experimental configuration

CPU	GPU	CUDA	cuDNN	Ubuntu	Tensorflow	Python
i7-8700K	GTX 1080 Ti	9.0	7.0	16.04 LTS	1.9	2.7

表 2 实验参数设置

Table 2 Experimental parameters setting

Datasets	Num points	K	Pyramid Pooling	Optimizer	Learning Rate	Weight Decay	Momentum	Batch Size	Epoch
ModelNet40[10]	1024	20	$N, N/4, N/8, N/16$	Adam	0.001	1×10^{-4}	0.9	32	300
ShapeNet[17]	2048	30	$N, N/4, N/8, N/16$	Adam	0.001	1×10^{-4}	0.9	16	200
S3DIS[18]	4096	30	$N, N/4, N/8, N/16$	Adam	0.001	1×10^{-5}	0.9	24	60

3.2 3D 目标分类

本节在 ModelNet40[10] 标准形状分类数据集上进行实验，并评估分类网络的性能。ModelNet40 有 40 个人造物体类别，共计 12311 个 CAD 模型，其中 9843 个模型用于训练，2468 个模型用于测试。表 3 列出了本文网络和现有网络在测试集上的总体分类精度、平均分类精度以及推理时间。

表 3 ModelNet40 数据集上的分类实验结果对比

Table 3 Comparison of classification experimental results on ModelNet40 data set

Model	Accuracy avg class /%	Accuracy Overall/%	Model Size/MB	Forward Time/ms
VoxNet[9]	83.0	85.9	—	—
Subvolume[8]	86.0	89.2	—	—
MVCNN[3]	90.1	—	—	—
ECC[14]	83.2	87.4	—	—
PointNet[2]	86.2	89.2	40.0	25.3
LightPointNet[7]	—	89.5	—	—
Kd-Net[4]	—	91.8	—	—
PointNet++[12]	—	90.7	12.0	163.2
SpiderCNN[5]	—	92.4	—	—
DGCNN[15]	90.2	92.2	21.0	94.6
GAPNet[6]	89.7	92.4	23.0	44.8
Ours	90.7	93.1	44.9	28.3

从表 3 可以看出，无论是总体分类精度还是平均分类精度，本文提出的网络都达到了较好的性能：在总体分类精度方面，本文网络比 PointNet 高出 3.9%，比 PointNet++高出 2.4%，比 DGCNN 高出 0.9%，比 SpiderCNN 和 GAPNet 均高出 0.7%；在平均分类精度方面，本文网络比 PointNet 高出 4.5%，比 GAPNet 高出 1.0%，比 DGCNN 高出 0.5%。虽然本文网络的推理时间比 PointNet 稍长，但是其分类精度有显著提升。另外，本文网络的推理时间仅为 PointNet++的 17.3%，因此其推理效率和分类精度都要优于 PointNet++。

为了验证空间金字塔池化结构对网络性能的影响，表 4 列出了设置不同 PSPP 参数以及无 PSPP 的对比结果。实验表明，选择设置 PSPP 比不设置 PSPP 能获得明显的效果提升；同时，PSPP 窗口类型为 $N, N/4, N/8, N/16$ 时性能最优，在数据集上的总体分类精度和平均分类精度都达到较高值。

表 4 不同 PSPP 参数设置的分类效果对比

Table 4 Comparison of classification effects with different PSPP parameter settings

(单位：%)

PSPP	Accuracy avg class	Accuracy overall
无	89.7	92.3
$N, N/4, N/16$	89.8	92.9
$N, N/4, N/8, N/16$	**90.7**	93.1
$N, N/2, N/4, N/8, N/16$	90.2	**93.2**

为了进一步验证图的 k 近邻点数量设置对网络性能的影响，对 $k=10$，$k=20$，$k=30$，$k=40$ 时的网络分类精度进行了对比，结果如表 5 所列。

表 5 选择不同 k 值的分类效果对比

Table 5 Comparison of classification effects with different k values

(单位：%)

k	Accuracy avg class	Accuracy overall
10	90.2	93.0
20	90.7	93.1
30	89.5	92.7
40	88.9	92.6

从表 5 可以看出，当 $k=10$ 时，网络的平均分类精度为 90.2%，总体分类精度为 93.0%；当 $k=20$ 时，网络性能最好，平均分类精度为 90.7%，总体分类精度达到 93.1%；$k=30$ 时，网络性能开始下降，平均分类精度为 89.5%，总体分类精度为 92.7%；$k=40$ 时，网络性能再次下降。实验表明，k 值过高或过低都会导致网络的分类性能下降。因此，本文选择 $k=20$ 作为分类网络的近邻点数量。

3.3 部件分割

相比目标分类，部件分割是一项对点云细粒度特征要求更高、更具难度的点云处理任务，其目的是给定一个已知类别的模型，如飞机，然后得出模型中每一个点所属零部件的类别标签得分，如机身、机翼和机尾。因此，需要在网络中添加表示输入形状的 16 个类别的 one-hot 向量信息，并与经 PSPP 得到的特征信息以及经 GraphConv 得到的 $N\times64$ 和 $N\times128$ 的低维度特征进行融合，从而得到一个 $N\times1232$ 维度的特征，然后通过共享的 3 个全连接层 $mlp(256,256,128)$ 得到每个点的特征信息，最终输出 $N\times50$ 的部件类别结果。

本节实验采用常用的部件分割数据集 ShapeNet[17] 对模型进行评估，此数据集包含 16881 个形状，这些形状分属 16 个类别，并注释了 50 个部件标签，每个类别包含 2～5 个部件标签，还提供了真值标签。实验严格遵循官方对训练集、验证集和测试集的划分。为了更严格地进行实验对比，本节中的实验采用与 PointNet 同样的基于点 mIoU 的评价方案。对 C 个类别中的每个形状 S 计算 $mIoU$，对 C 个类别中的每个部件类型在真实值和预测值之间求 IoU。若点的真实值和预测值的交集为空，则 IoU 为 1。计算类别中某一种形状的 $mIoU$，是通过对这个类别中此形状所有部件类型求平均 IoU 而得到的。为了计算某一类别的 $mIoU$，对此类别中所有形状的 $mIoU$ 求平均值。

ShapeNet 部件数据集上的分割效果（$mIoU$）如表 6 所列。可以看出，本文网络的 $mIoU$ 高出 PointNet 1.6%，高出 GAPNet 0.6%，高出 DGCNN 0.2%，高出 PontNet++0.2%，具有更好的部件分割性能；与 SpiderCNN 相比，本文网络在 7 个类别上的 IoU 更高，在 2 个类别上的 IoU 相同，总体表现极佳。

表 6　ShapeNet 数据集上的部件分割实验效果对比

Table 6　Experimental effects of comparison of component segmentation on ShapeNet data set

（单位：%）

model	mean	aero	bag	cap	car	chair	ear phone	guitar	knife	lamp	laptop	motor	mug	pistol	rocket	skate board	table
PointNet[2]	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++[12]	85.1	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
Kd-Net[4]	82.3	80.1	74.6	74.3	70.3	88.6	73.5	90.2	87.2	81.0	94.9	57.4	86.7	78.1	51.8	69.9	80.3
DGCNN[15]	85.1	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	82.0
GAPNet[16]	84.7	84.2	84.1	88.8	78.1	90.7	70.1	91.0	87.3	83.1	96.2	65.9	95.0	81.7	60.7	74.9	80.8
SpiderCNN[5]	85.3	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
Ours	85.3	84.1	83.4	86.4	78.0	90.7	74.7	91.2	87.6	82.7	95.8	66.4	94.8	81.1	63.5	75.6	82.7

为了直观展示部件分割网络在各个类别上的分割效果，对于预测的 16 个类别，从每个类别随机选取 3 个分割图进行实验，结果如图 4 所示，其中不同的颜色表示不同的部件类别。

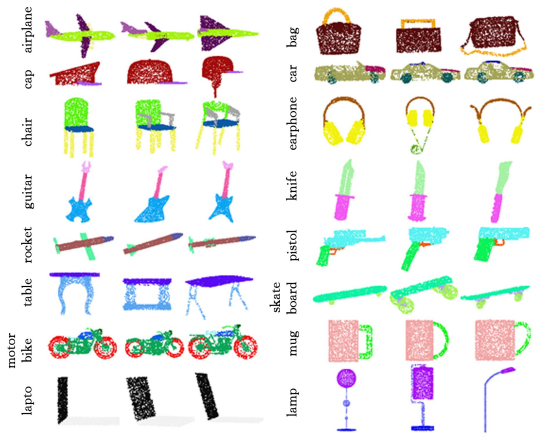

图 4　ShapeNet 数据集上 16 个类别的部件分割效果（电子版为彩色）

Fig. 4　Component segmentation effects of 16 categories on ShapeNet data set

图 5 进一步给出了部件分割网络对椅子的分割效果图。从图中可以看出，尽管椅子的形状不同，但是本文网络依然清晰地分割出了靠背、扶手、椅面、椅腿等重要部位。

图 5　椅子部件的分割效果

Fig. 5　Segmentation effects of chair parts

如图 6 所示，上、中、下 3 行分别是 PointNet 的分割效果、本文网络的分割效果以及真值。可以看到，第 1 列的桌子腿部，PointNet 将大量应为浅蓝色的桌腿误识别为蓝紫色的桌面；第 2 列主要为手提袋提手部位的识别效果对比；第 3 列为摩托车后轮红色和车体绿色的区分效果；第 4 列为耳机耳罩黄色部位中的绿色杂点问题；第 5 列为火箭头蓝色部分的完整性对比；第 6 列为帽沿处紫色部分和帽体红色点之间的分割区分。可以看出，本文网络在这些细粒度部位的分割效果都优于 PointNet 网络。

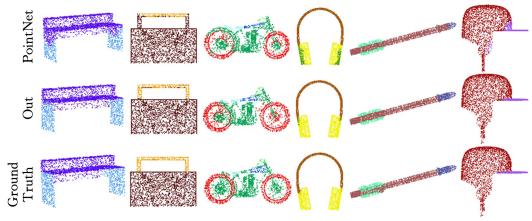

图 6　PointNet、本文网络对部件分割的效果以及真实值的对比（电子版为彩色）

Fig. 6　Comparison of actual value and part segmentation effects of PointNet and the proposed network

3.4　场景语义分割

不同于部件的语义分割，3D 场景语义分割无须在网络中输入训练的 16 个类的 one-hot 向量，网络结构与图 1 所示分割网络一致。把经 PSPP 聚合得到的高维度特征与经 Graph-Conv 得到的 $N\times 64$ 的低维度特征进行融合，得到不同层次的特征，形成 $N\times(64+1\,024)$ 维度的点特征，再经过两组共享参数的全连接层 $mlp(512,256,12)$ 和 $mlp(128,m)$ 输出 N 个点在 m 个类别上的得分，从而达到语义分割的目的。

实验采用了斯坦福大学的语义分割数据集 S3DIS[18]，该数据集包含来自 6 个区域的 271 个房间的 3D 扫描数据。场景中的每个点都进行了语义标注，标注类别有桌子、椅子、地板、墙壁、天花板等 13 个类别。在数据集的处理上，依然遵循 PointNet 中的规范，按房间对点进行划分，将房间分成 1m×1m 的块，用本文的分割网络在块区域上训练预测每一个点的所属类别。每个点都由一个包含 XYZ、RGB 和归一化空间坐标的 9 维向量表示。在训练过程中，从每个块区域随机采集 4 096 个点作为输入；在测试过程中，所有点都被用于测试。在 Area1，Area2，Area3，Area4，Area5，Area6 这 6 个场景中进行交叉验证，得到最终的平均评估结果。

表 7 对比了本文网络模型与 PointNet（baseline）和 PointNet 的语义分割实验结果。度量标准是 13 个类别的平均 IoU 和点分类精度。不论是在平均 IoU 还是点分类精度上，本文网络的实验效果都优于 PointNet。同时，图 7 给出了本文网络与 PointNet 网络的场景语义分割效果对比图，上、中、下 3 行分别是 PointNet 效果、本文效果和真值标签，4 列结果分别为取自不同场景的效果图。第 1 列为从会议室场景的门口观看时门口右小角红色部分的区别；第 2 列为桌椅部分的分割效果对比；第 3 列为窗户以及门右上角的横梁；第 4 列为场景左下角的横梁以及红色椅子部分。从主观效果图也可以

看出，本文网络在复杂场景解析上的表现效果优于 PointNet。

表 7 S3DIS 数据集上的 3D 场景语义分割效果

Table 7 Semantic segmentation effects of 3D scene on S3DIS data set

（单位：%）

Model	Mean *IoU*	Overall accuracy
PointNet(baseline)	20.1	53.2
PointNet	47.6	78.5
Ours	56.2	84.3

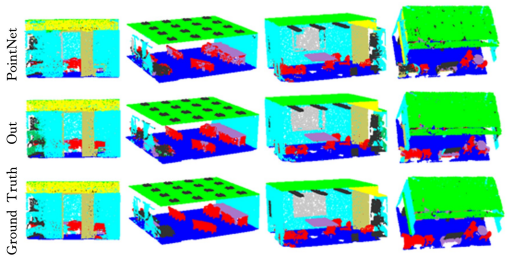

图 7 场景语义分割效果的对比(电子版为彩色)

Fig. 7 Comparison of semantic segmentation effects in different scenes

结束语 针对直接对点云数据进行特征学习的现有卷积神经网络存在局部特征缺失、处理环节多、计算量大等问题，本文提出了基于动态图卷积和空间金字塔池化的点云深度学习网络。通过设计动态图卷积的特征学习机制和针对点云的空间金字塔池化结构，显著提升了网络对点云局部特征的提取能力；同时通过实验分析对网络参数和结构进行优化，使得网络更加简洁、高效。在点云分类、部件分割和场景解析 3 个基准数据集上的评估结果表明，相较于其他网络，本文提出的网络能取得更高的分类和分割精度，对复杂点云场景分割的泛化能力也更强。当然，本文提出的网络依然有许多地方需要改善，比如如何增强对更高复杂度、极细粒度处点云的局部拓扑信息的捕获能力，这是我们下一步研究工作的重点。

参 考 文 献

[1] LIU J, WU Z K, ZHOU M Q. Overview of point cloud model segmentation and application technology[J]. Computer Science, 2011, 38(4): 21-24.

[2] QI C R, SU H, KAICHUN M, et al. PointNet: deep learning on point sets for 3D classification and segmentation[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE Computer Society Press, 2017: 77-85.

[3] SU H, MAJI S, KALOGERAKIS E, et al. Multi-view convolutional neural networks for 3D shape recognition [C]// 2015 IEEE International Conference on Computer Vision. New York: IEEE Press, 2015: 945-953.

[4] KLOKOV R, LEMPITSKY V. Escape from cells: Deep kd-networks for the recognition of 3d point cloud models[C]// Proceedings of the IEEE International Conference on Computer Vision. Honolulu: IEEE Computer Society Press, 2017: 863-872.

[5] XU Y, FAN T, XU M, et al. Spidercnn: Deep learning on point sets with parameterized convolutional filters[C]// Proceedings of the European Conference on Computer Vision (ECCV). Munich: IEEE Press, 2018: 87-102.

[6] CHEN C, LUCA Z F, ANTONIOS T. GAPNet: Graph Attention based Point Neural Network for Exploiting Local Feature of Point Cloud[J]. arXiv: 1905.08705.

[7] BAI J, SI Q L, QIN F Y. LightPointNet, a lightweight real-time point cloud classification network[J]. Journal of Computer-Aided Design and Graphics, 2019, 31(4): 612-621.

[8] QI C R, SU H, NIEβNER M, et al. Volumetric and multi-view cnns for object classification on 3d data[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE Computer Society Press, 2016: 5648-5656.

[9] MATURANA D, SCHERER S. Voxnet: A 3d convolutional neural network for real-time object recognition [C] // 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). New York: IEEE Press, 2015: 922-928.

[10] WU Z, SONG S, KHOSLA A, et al. 3d shapenets: A deep representation for volumetric shapes[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Boston: IEEE Computer Society Press, 2015: 1912-1920.

[11] JADERBERG M, SIMONYAN K, ZISSERMAN A. Spatial transformer networks[C] // The 24th Annual Conference on Neural Information Processing Systems. Cambridge: MIT Press, 2015: 2017-2025.

[12] QI C R, YI L, SU H, et al. PointNet++: Deep hierarchical feature learning on point sets in a metric space[C]// The 24th Annual Conference on Neural Information Processing Systems. Cambridge: MIT Press, 2017: 5105-5114.

[13] DEFFERRARD M, BRESSON X, VANDERGHEYNST P. Convolutional neural networks on graphs with fast localized spectral filtering[C]// Advances in Neural Information Processing Systems. New York: IEEE Press, 2016: 3844-3852.

[14] SIMONOVSKY M, KOMODAKIS N. Dynamic edge-conditioned filters in convolutional neural networks on graphs[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE Computer Society Press, 2017: 3693-3702.

[15] WANG Y, SUN Y, LIU Z, et al. Dynamic Graph CNN for Learning on Point Clouds[J]. arXiv: 1801.07829.

[16] HE K, ZHANG X, REN S, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1904-1916.

[17] YI L, GUIBAS L, KIM V G, et al. A scalable active framework for region annotation in 3D shape collections[J]. ACM Transactions on Graphics, 2016, 35(6): 1-12.

[18] ARMENI I, SENER O, ZAMIR A R, et al. 3d semantic parsing of large-scale indoor spaces[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE Computer Society Press, 2016: 1534-1543.

ZHU Wei, born in 1982, Ph.D, associate professor. His main research interests include video processing, machine learning and intelligent robot.