结合位置关系卷积与深度残差网络的三维点云识别与分割

杨军[1,2]，王连甲[1]

（1. 兰州交通大学电子与信息工程学院，730070，兰州；2. 兰州交通大学测绘与地理信息学院，730070，兰州）

摘要：针对现有点云识别与分割算法因忽视点的位置特征和局部几何特征关系而导致难以捕获具有鉴别力的局部几何信息的问题，提出基于位置关系深度残差神经网络的三维点云识别与分割算法。将原始点云嵌入到高维空间并获取其高维特征；将点云的高维特征输入位置关系卷积实现局部邻域内当前点特征与位置几何特征的信息交流，并通过深度残差模块强化提取到的深层语义特征，分层重复以上步骤可逐步得到点云的高级上下文语义特征；通过全连接层与解码器，得到点云的识别与分割结果。实验结果表明，所提算法在ModelNet40点云分类数据集的识别精度达到了93.9%，在ShapeNet Part点云部件语义分割数据集的平均交并比达到了86.0%。所提算法能够提取三维点云的关键特征信息，具有较好的三维点云识别与分割能力。

关键词：三维点云；位置关系卷积；残差网络；模型识别；语义分割

中图分类号：TP391 **文献标志码**：A

DOI：10.7652/xjtuxb202305018 **文章编号**：0253-987X(2023)05-0182-12

Recognition and Segmentation of 3D Point Cloud Through Positional Relation Convolution in Combination with Deep Residual Network

YANG Jun[1,2], WANG Lianjia[1]

(1. School of Electronic and Information Engineering, Lanzhou Jiaotong University, Lanzhou 730070, China;
2. Faculty of Geomatics, Lanzhou Jiaotong University, Lanzhou 730070, China)

Abstract: The current point cloud recognition and segmentation algorithms ignore the relationship between positional features and local geometric features of points, resulting in difficulties in capturing the discriminative local geometric information. To tackle this problem, this paper proposes a 3D point cloud recognition and segmentation algorithm based on positional relation convolution and deep residual neural network. Firstly, the original point cloud is embedded into a high-dimensional space to obtain its high-dimensional features. Then, the high-dimensional features of the point cloud are input into the positional relation convolution to realize the information exchange between the current point features and the positional geometric features in the local region. The deep residual module is also used to enhance the extracted deep semantic features. The high-level contextual semantic features of the point cloud are gradually obtained by repeating the above steps hierarchically. Finally, the recognition and segmentation results of the point cloud are obtained through the fully connected layer and the decoder respectively. The experimental re-

收稿日期：2022-09-28。 作者简介：杨军(1973—)，男，教授，博士生导师。 基金项目：国家自然科学基金资助项目(42261067)；兰州市人才创新创业项目(2020-RC-22)；兰州交通大学天佑创新团队资助项目(TY202002)。
网络出版时间：2023-01-03 网络出版地址：https://kns.cnki.net/kcms/detail/61.1069.T.20230103.1358.004.html

sults show that the recognition accuracy of the algorithm proposed in this paper reaches 93.9% in the ModelNet40 point cloud classification dataset, and the average intersection over union reaches 86.0% in the ShapeNet Part point cloud part semantic segmentation dataset. The algorithm in this paper can extract the critical feature information of 3D point clouds and is capable of 3D point cloud recognition and segmentation.

Keywords: 3D point cloud; positional relation convolution; residual network; model recognition; semantic segmentation

近年来，深度学习已经在二维图像处理如图像分类[1-2]、目标检测[3-4]和语义分割[5-6]等方面取得了显著的成绩。越来越多的研究开始对三维空间进行探索。三维点云由于其包含的信息丰富且数据结构简单，成为了三维空间的主要数据表达方式。与此同时，三维(3D)激光扫描技术的兴起使得点云数据的获取成本快速降低，促进了许多依赖 3D 点云数据应用的发展，例如自动驾驶[7]、机器人[8]和三维重建[9]等。三维点云的识别与分割算法作为以上应用开发的技术基石具有重要研究意义。三维点云识别就是将输入的三维点云模型进行分析并输出模型所属类别，而三维点云分割则是将输入的点云场景中每个点都预测其所属的物体或部件类别。然而，由于点云由三维空间中无序和不规则的点集组成，不能直接将处理二维图像的深度学习方法直接用于三维点云处理，这使得该领域的研究备受关注。

早期的研究将三维点云投影至二维平面，再用成熟的二维图像的卷积神经网络模型提取特征。但是，多视图方法会不可避免地丢失某些具有鉴别力的几何信息，并且投影视角的选择也需要丰富的先验知识。基于体素的方法将三维点云规则化至三维体素块，然后用三维卷积操作对体素块进行处理。该类方法一定程度上保留了点云的几何信息，然而将三维点云映射至三维体素的过程会消耗额外的内存与计算空间，并且无法细分物体边界的几何信息。以 PointNet[10]为代表的基于点云的方法直接将原始点云输入深度学习网络进行处理，这样能够利用点云原始信息且不增加额外操作，可以充分获取点云所有信息，但由于点云具有不规则、无序等特点，当前此类方法存在局部特征提取不充分，点云识别与分割精度不足等问题。

针对上述问题，本文提出了基于位置关系卷积与深度残差网络(positional relation convolution and deep residual network，PRCDRNet)的三维点云识别与分割算法。位置关系卷积可以对局部邻域内点的语义特征与几何位置关系特征进行编码融合，增强局部特征的上下文信息。深度残差模块通过基本残差单元的堆叠能够对位置关系卷积得到的融合特征进行深层强化，充分提取深度语义信息，在获得更精细的点云特征的同时，解决了因网络层数加深而产生的网络退化问题。主要贡献与创新点有两点：一是构建位置关系卷积，充分提取并融合局部邻域内位置关系特征与几何结构特征，得到更丰富的局部邻域特征；二是引入深度残差模块，缓解因网络层数加深产生的网络退化问题，提升点云识别与分割准确率。

1　相关研究

目前，关于三维点云的分类与分割算法主要分为基于多视图、基于体素与基于原始点云三类方法。

1.1　基于多视图的方法

由于点云的不规则与无序性，早期的许多研究先将三维点云投影至二维平面，再使用现有的二维卷积神经网络模型分析处理。Su 等[11]首先将三维点云按照不同视角投影至二维平面得到二维投影视图，然后将得到的二维投影视图分别送入用于二维图像分类的 VGG[12]网络进行特征提取，最后将提取到的特征进行池化操作得到点云的全局特征描述符。冯元力等[13]通过对三维模型进行球面深度投影得到球面全景图，并将每个模型的球面全景图从多个角度展开，创建多幅平面图像作为多分支的卷积神经网络的输入，对多幅全景图进行整合分析后可获得三维点云的识别结果。Wei 等[14]将二维投影视图看作节点构造有向图，然后将得到的节点特征最大池化后形成全局特征描述符，最后通过全连接层得到点云的识别结果。

基于多视图的方法可以利用现有成熟有效的二维卷积神经网络实现三维点云的识别与分割，但这种方法忽略了三维点云的空间内在几何关系，且在投影时三维点云的部分信息会因为被遮挡而损失，从而影响识别精度。

1.2 基于体素的方法

基于体素的方法通常将三维点云映射至三维体素块，然后使用标准的三维卷积处理。Maturana 等[15]提出一种体素占用网络 VoxNet，把不规则的点云数据规则化为三维体素网格的形式，然后用三维卷积神经网络处理三维体素网格获得特征描述符。Choy 等[16]提出一种用于时空三维点云数据的 4D 稀疏卷积网络，并创建了稀疏张量自动微分的开源库。所提出的广义稀疏卷积能够有效处理高维数据，显著降低传统 3D 卷积核计算成本，且该卷积核对于立方体结构的物体识别能力更强。文献[17]在模型特征的提取过程中，使用堆叠小卷积核的深度卷积神经网络以便更加有效地挖掘模型内部的隐含信息并提取体素矩阵的深层特征，提高了复杂 3D 模型识别准确率。

基于体素的方法在一定程度上保留了点云的三维几何结构，然而体素化过程中体素块大小的选择会影响网络性能，体素块太大会丢失几何细节，太小会增加计算成本与内存消耗。设计基于体素的点云处理算法需要进行计算效率和体素比例的权衡。

1.3 基于原始点云的方法

Charles 等[10]首次提出了直接使用原始点云作为输入的网络 PointNet，充分考虑了直接处理点云所需的置换不变性与平移旋转不变性，利用多层感知机(multi-layer perceptron, MLP)学习每个独立点的高维表征，然后采用最大池化层对所有点的高维特征进行聚合得到全局特征描述符。PointNet 是直接处理原始点云方法的先驱工作，但是其只关注到每个点对于全局特征的影响，没有考虑到局部特征。Qi 等[18]在 PointNet 基础上提出 PointNet++网络，通过划分局部点云分层提取多尺度特征，将提取到的局部特征与全局特征充分融合得到点云语义特征。周鹏等[19]提出一种基于资源平衡的神经网络架构搜索方法来实现三维点云模型的识别。Li 等[20]利用点的规范顺序，通过 X-Conv 操作符对输入点和特征进行排列和加权，不同的输入顺序对应不同的 X 变换矩阵，并用常规卷积对重组后的点进行处理，解决了点云输入无序的问题。Dang 等[21]提出分层并行组卷积，可以同时捕捉点云的区分性独立单点特征和局部几何特征，以较少的冗余信息增强特征的丰富性，提高了网络识别复杂类别的能力。PointWeb 不仅考虑了局部邻域中邻近点与中心点的关系，还考虑到邻近点的结构，通过连接局部区域中的每个点对，以获得更具代表性的区域特征[22]。Wang 等[23]提出边缘卷积 EdgeConv 来生成描述点与其相邻点之间关系的边特征，将每个中心点与其 k 个最近邻点连接起来构成局部邻域来聚合本地上下文信息；成对的特征由 MLP 独立编码构成图结构，并且在每层 MLP 编码后动态更新图结构，使得对不同规模与类型的点云得到更好的识别与分割结果。文献[24]提出了一种轻量级的代理点图卷积，能够简化边缘卷积操作且降低内存消耗，获得深层细粒度的几何特征，并对语义特征和局部几何特征进行编码，增强了特征局部的上下文信息。PointConv 通过使用 MLP 逼近卷积滤波器中的连续权重和密度函数，将动态滤波器扩展到新的卷积运算[25]。PointASNL 使用自注意力机制构建自适应采样(adaptive sampling, AS)模块与局部-非局部(local-non local, L-NL)模块提取特征，AS 模块可以自适应地调整最远点采样后的点坐标与特征，L-NL 模块用于捕获采样点的局部与全局依赖关系[26]。

基于原始点云的方法不需要将点云投影至二维平面或体素化，节省了计算消耗与内存空间，但现有的方法没有充分利用点云最重要的几何位置特征，缺乏对局部特征的精细提取。

2 位置关系卷积与深度残差网络

三维点云可以表示为空间中的一组无序点集。不同点云数据集包含的特征通道数并不统一。一般情况下，点云的输入特征通道数等于 3，表示每个点由三维坐标表示，不同的点云数据可能包含其他额外的特征信息，如颜色、深度或表面法向量等。在各种特征中最具代表性与决定性的特征是点云的位置坐标特征，因此本文重点关注点云的位置信息特征。在位置关系深度残差神经网络中，通过位置关系卷积(positional relation convolution, PRConv)将局部邻域内当前点的语义特征、位置关系特征与几何特征充分融合，得到点云局部区域的高级特征；将融合后的特征送入深度残差模块(deep residual block, DRB)进行深层语义特征提取，将得到的深层语义特征按顺序分层送入位置关系卷积与深度残差模块即可得到包含丰富位置与几何特征的点云全局特征。

2.1 位置关系卷积

二维(2D)图像处理中的卷积操作是卷积神经网络中的核心步骤，如图 1 所示，大小固定的卷积核自左往右，从上而下的扫描整个图像，经过反向传播得到学习后的权重共享的核参数。二维卷积操作可以抽象表示为

$$\mathrm{Conv}_{2\mathrm{D}}(W,\ C)=\iint_{v\in\mathbf{R}^2}W(v)C_v\mathrm{d}\boldsymbol{v}\qquad(1)$$

式中：Conv(·)表示卷积操作；v 表示卷积核大小，例如 1×1、3×3、5×5 等；$W(\cdot)$ 表示权重函数；C_v 表示 $\boldsymbol{v}$ 范围像素上的通道特征。

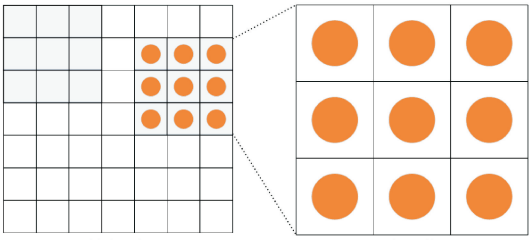

图 1　二维图像上的卷积操作

Fig. 1　Convolution on 2D image

由于点云(point cloud)的无序性与不规则性，每个点并不固定在如二维图像的规则网格中，因此不能直接使用二维卷积处理点云。当前主流处理方法如图 2 所示，首先使用最远点采样(farthest point sampling，FPS)算法对输入点云进行下采样，确定局部区域的中心点，然后用 k 最近邻(k-nearest neighbor，k-NN)分类算法确定中心点的 k 个邻域点，再对局部区域内处理。点云的卷积操作可以抽象表示为

$$\mathrm{Conv}_{\mathrm{pc}}(W,F)=\iiint_{(l_\alpha,l_\beta,l_\gamma)\in L}W(l_\alpha,l_\beta,l_\gamma)\cdot F_{(\alpha+l_\alpha,\beta+l_\beta,\gamma+l_\gamma)}\mathrm{d}l_\alpha\mathrm{d}l_\beta\mathrm{d}l_\gamma\qquad(2)$$

式中：(α,β,γ) 为点云通过 FPS 下采样后任意点的坐标；L 表示以 (α,β,γ) 为中心点以 k 个点为邻域的局部区域；$(l_\alpha,l_\beta,l_\gamma)$ 为局部邻域中的任意点坐标；$F_{(\alpha+l_\alpha,\beta+l_\beta,\gamma+l_\gamma)}$ 表示局部区域 L 内的通道特征。

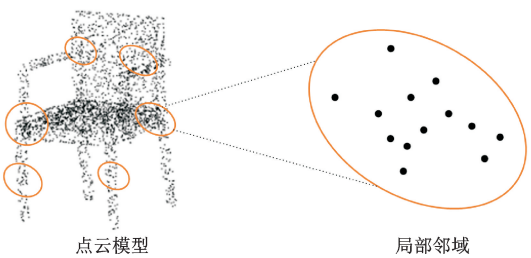

图 2　三维点云上的卷积操作

Fig. 2　Convolution on 3D point cloud

现有的基于原始点云三维模型识别与分割方法大多数都采用如式(2)所示的方式进行点云的特征提取，不同的方法区别在于对权重函数 $W(\cdot)$ 的设计不同。如图 3(a)所示，图中每个圆表示点云中的一个点，其中红色表示点云下采样后的中心点，蓝色表示局部邻域内的点，点与点间的连线表示点与点的不同关系，PointNet++首次考虑了局部邻域内中心点与邻域点各自独立的特征，其卷积函数可表示为

$$W=M(f(x_j)),\ x_j\in L\qquad(3)$$

式中：$M(\cdot)$ 表示权重共享的多层感知机；$f(\cdot)$ 表示点的当前特征；x_j 为局部邻域内任意点。DGCNN 通过学习高维特征空间中的点关系来捕捉相似的局部形状，如图 3(b)所示，其卷积函数可表示为

$$W=M[f(x_j),f(x_j)-f(x_i)],\ x_j\in L\qquad(4)$$

式中：x_i 为局部邻域中心点。PointWeb 使用自适应特征调整模块自适应地学习局部邻域内每个点对之间的影响指标，如图 3(c)所示，其卷积函数可表示为

$$W=M[f(x_j),F_{\mathrm{rel}}(x_j,x_h)],\quad x_h,x_j\in L\qquad(5)$$

$$F_{\mathrm{rel}}=\begin{cases}f(x_j)-f(x_h), & x_h\neq x_j\\ f(x_j), & x_h=x_j\end{cases}\qquad(6)$$

式中：x_h 为局部邻域内的任意点；F_{rel} 为自适应学习函数。

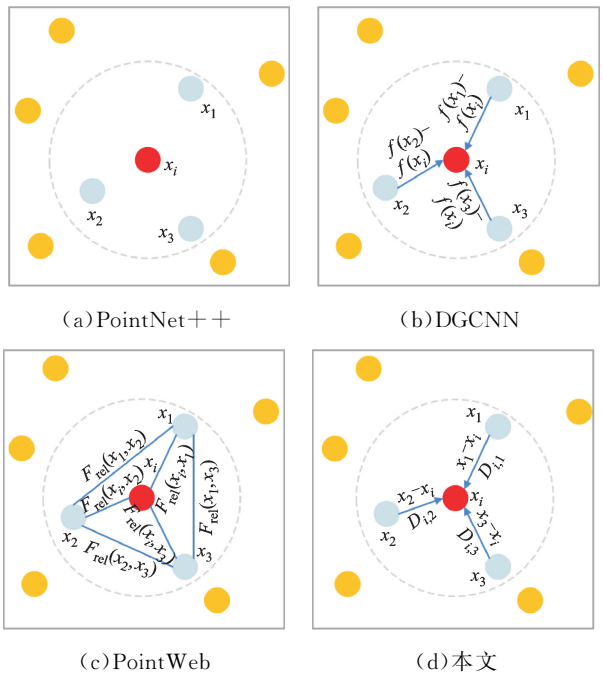

图 3　不同网络模型邻域内点的结构关系

Fig. 3　The structural relationships of points in local region of different network models

PointNet++考虑了局部邻域内中心点与邻域点各自独立的特征，但是缺乏对邻域内中心点与邻域点关系的探索，DGCNN 考虑了邻域内中心点与邻域点之间的关系，但是邻域点与中心点的关系表示为点高维特征之间的运算，没有重点关注点云最重要的位置与几何特征，且随着网络层数加深，特征

提取至高维时，网络的运算量变得很大，计算效率大大降低。PointWeb 不仅考虑了邻域点与中心点的关系，还尝试将邻域内每个点的特征关联起来，但是这样会极大增加网络的参数，造成特征冗余，增加运算成本。为了挖掘点云邻域内的位置关系与几何拓扑结构，更充分地提取点云的高级语义特征，本文提出了位置关系卷积，其中邻域内点的关系如图 3(d)所示，可以表示为

$$W = M[f(x_j), x_i, (x_j - x_i), D_{i,j}], x_j \in L \tag{7}$$

$$D_{i,j} = \| x_j - x_i \|^2 \tag{8}$$

式中：$D_{i,j}$ 为邻域中心点 x_i 与邻域点 x_j 的欧氏距离。式(7)中位置关系卷积函数将局部邻域特征 $f(x_j)$、局部邻域中心点位置特征 x_i、邻域内点与中心点的相对位置特征 $x_j - x_i$ 和邻域内点与中心点的欧氏距离特征拼接后送入 MLP 进行特征融合，融合过程中既保留了当前点的高级特征，增强了当前邻域内中心点与邻域点位置关系特征，又引入欧氏距离强化了局部邻域内的几何拓扑特征。

综上，本文提出的位置关系卷积(positional relation convolution，PRConv)结构如图 4 所示。图中不同立方体表示位置关系卷积中产生的不同特征，矩形表示位置关系卷积中的运算操作，主要由两条分支组成。上路分支通过对原始点云的位置信息进行运算得到局部邻域内的强化位置特征与几何特征；下路分支通过 FPS、k-NN 操作得到当前局部邻域的高维语义特征。然后将两条分支获取的特征进行拼接后送入权重共享的 MLP 进行特征融合，最后对融合后的特征采用最大池化操作，输出表征整个局部邻域的高级语义特征。

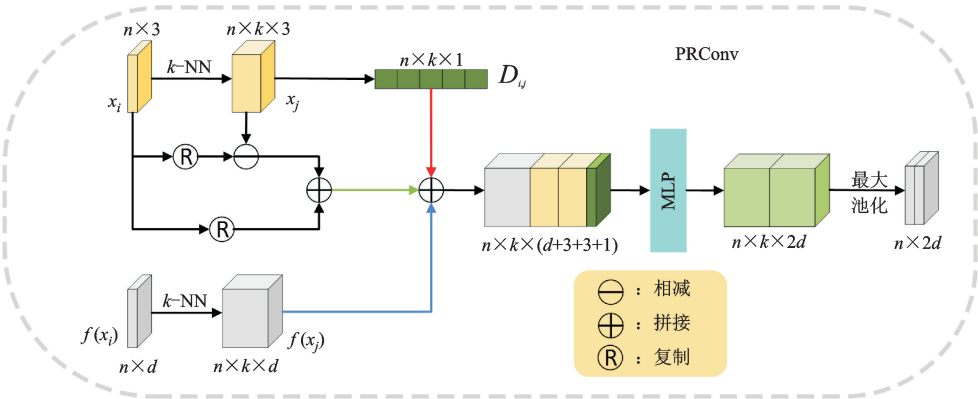

图 4 位置关系卷积结构示意图

Fig. 4 The structure of positional relation convolution

2.2 深度残差模块

通过位置关系卷积，得到了整个局部邻域的高级语义特征，但是仅通过位置关系卷积中的 MLP 不能充分提取局部邻域中的深层特征。受 ResNet[27]的启发，本文引入残差结构，构建了新的深度残差模块，在增加网络深度，充分提取局部邻域中深层特征的同时，降低梯度消失，缓解因网络层数加深产生的网络退化问题。

神经网络需要使用 MLP 来提取三维点云的特征，普通 MLP 结构对通道特征的处理可以简单表示为

$$C_{\text{out}} = f(C_{\text{in}}) \tag{9}$$

式中：C_{in} 为输入特征；C_{out} 为输出特征；$f(\cdot)$ 为映射函数即 MLP 操作。MLP 通过反向传播更新网络的参数，反向传播过程中需要对参数求导数

$$\frac{\mathrm{d}C_{\text{out}}}{\mathrm{d}C_{\text{in}}} = f'(C_{\text{in}}) \tag{10}$$

由于当前网络由多层 MLP 堆叠组成，当 MLP 层数过多时，反向传播时会产生梯度消失的问题，即 $f'(C_{\text{in}})$ 接近于 0。这会导致反向传播过程难以进行，最终使 MLP 无法准确地提取通道特征。因此，ResNet 提出一种残差结构，可以简单表示为

$$C_{\text{out}} = f(C_{\text{in}}) + C_{\text{in}} \tag{11}$$

此时对参数进行求导数可得到

$$\frac{\mathrm{d}C_{\text{out}}}{\mathrm{d}C_{\text{in}}} = f'(C_{\text{in}}) + 1 \tag{12}$$

当网络包含的 MLP 层数过多，在反向传播时会出现梯度消失问题，即 $f'(C_{\text{in}})$ 接近 0 时，MLP 的梯度将大于等于 1，在一定程度上解决了反向传播中梯度消失问题。据此，本文设计了一个基本残差单元(basic residual unit，BRU)，如图 5 所示，其中立方

体表示特征，矩形表示批归一化、激活函数等操作，其过程可以表示为

$$C_{\text{out}}=\sigma(B(M_2(\sigma(B(M_1(C_{\text{in}}))))))+C_{\text{in}}) \tag{13}$$

式中：$\sigma(\cdot)$ 为激活函数；$B(\cdot)$ 为批归一化操作；$M_1(\cdot)$ 与 $M_2(\cdot)$ 均为 MLP 操作，并且 $M_1(\cdot)$ 与 $M_2(\cdot)$ 输入与输出的通道数均等于 C_{in} 的通道数。

深度残差模块可由一个或多个基本残差单元串联组成。经过消融实验，本文构建的深度残差模块由两个基本残差单元组成，其结构如图 6 所示，其中立方体表示特征，矩形为基本残差单元(BRU)。

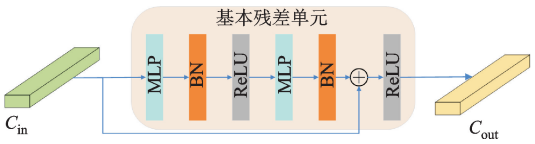

图 5　基本残差单元结构示意图

Fig. 5　The structure of basic residual unit

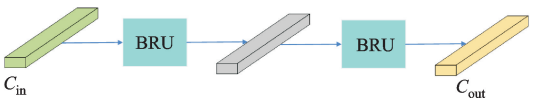

图 6　深度残差模块结构示意图

Fig. 6　The structure of deep residual block

2.3　基于位置关系卷积与深度残差网络的模型架构

本文构建的基于位置关系卷积与深度残差网络的模型架构采用编码器-解码器结构，如图 7 所示，其中矩形为位置关系卷积、插值等操作，不同颜色矩形块的堆叠表示网络各层输出的特征。在将原始点云输入编码器之前，先将原始点云的初始特征映射至高维空间，使编码器可以更充分地提取原始点云特征。为了充分挖掘点云的细粒度特征和多尺度上下文信息，编码器采用 4 层结构设计，每一层包含下采样、位置关系卷积模块和深度残差模块。在第一层中，首先将原始点云嵌入得到的高维点云特征 F_{e} 采用 FPS 算法下采样，并通过位置关系卷积模块对邻域内高维点云特征、位置关系特征与几何结构特征进行拼接融合，然后将融合后的特征输入深度残差模块充分提取局部邻域的深层语义特征，最后输出第一层结构提取到的浅层局部特征 F_1。通过相同的方法，可以得到更深层的局部特征 F_2 与 F_3，直至最后在第四层输出提取到的最高级局部特征 F_4。通过最大池化操作得到包含丰富上下文信息的点云全局特征 F_{g}，送入设计的全连接层(fully connected layer，FCL)得到点云分类的概率预测得分 $1\times C$，C 为点云类别数目。

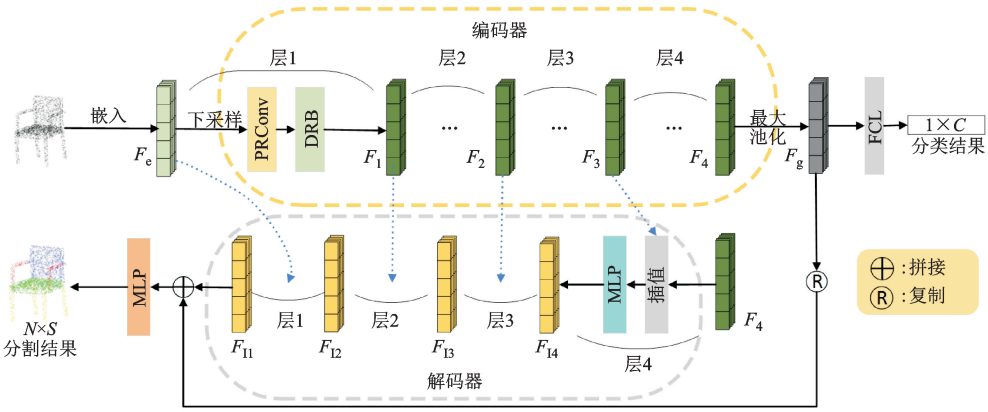

图 7　基于位置关系卷积与深度残差网络的模型架构

Fig. 7　Model architecture based on positional relation convolution and deep residual network

解码器同样采用分层结构设计，每一层包含插值操作与 MLP，插值操作用来实现分割所需的上采样过程，MLP 融合深层语义特征与浅层语义特征并对齐插值特征中的通道数。首先对局部深层特征 F_3 与 F_4 用距离插值方法[18]进行合并插值，并将得到的结果输入 MLP 与局部深层特征 F_3 的特征通道数对齐得到插值特征即解码器的首层输出 F_{I4}，通过相同的步骤可以分别得到插值特征 F_{I3}、F_{I2}、F_{I1}，插值特征 F_{I1} 为编码器的最终输出。最后将全局特征与解码器的最终输出拼接后送入 MLP 得到点云分割的概率预测结果 $N\times S$，N 为原始点云点的数目，S 为每个点的语义类别数目。

3　实验结果与分析

为了探究本文构建的基于位置关系卷积与深度残差网络(PRCDRNet)的模型对点云识别与分割的

有效性，分别在三维点云分类数据集 ModelNet40[28]与三维点云部件语义分割数据集 ShapeNet Part[29]上评估了网络模型的识别与分割性能。训练和测试的实验硬件为英特尔 i9 10900k CPU 和 GeForce RTX3090 GPU(24 GB 显存)显卡，实验环境与主要参数如表 1 所示。

表 1 实验环境与主要参数

Table 1 Environment and main parameters of experiment

实验环境	分类实验	分割实验
操作系统	Ubuntu 18.04	Ubuntu 18.04
Pytorch 版本	1.11	1.11
CUDA 版本	11.3	11.3
优化器	SGD	Adam
学习率	0.1	0.003
训练轮次	300	350

3.1 三维点云识别实验

本文选择在 ModelNet40 数据集上进行三维点云识别实验。ModelNet40 数据集是由普林斯顿大学用 CAD 软件构建的三维模型数据集，其包含来自 40 个类别的 12 311 个三维形状，如飞机、汽车、键盘等。其中 9 843 个模型用于训练，2 468 个模型用于测试。采用总体精度作为识别的性能评估指标。

文中选取了基于多视图、体素与原始点云 3 种方法中性能突出的网络模型进行对比实验，其中以基于原始点云输入方法为主，并对较为经典且创新性较强的网络模型与本文 PRCDRNet 模型进行了性能对比分析。在 ModelNet40 数据集上的实验结果对比如表 2 所示。可以看出：本文算法在 ModelNet40 数据集上的总体精度达到了 93.9%，由于本文采用的原始点云输入的方法不需要将点云转换为二维图像与体素块，减少了转换过程中信息的丢失，大幅领先基于体素方法的 VoxNet[15] 10.9%，相比基于多视图方法的 MVCNN[11]，总体精度超出 3.8%；与基于原始点云输入方法对比，比 1 000 个点作为输入的 PointNet++[18]的精度提升了 3.2%，主要由于 PRCDRNet 模型中的位置关系卷积充分提取了局部邻域内的位置关系特征与局部几何特征；与网络层数较深、参数量较大的 KPConv[32]模型相比，PRCDRNet 模型中的深度残差模块在增加网络深度的同时降低梯度消失，缓解因网络层数加深产生的网络退化问题，使得总体精度提升了 1%。此外，PRCDRNet 模型仅采用 1 024 个点作为输入，总体精度超过了以更多点或更多信息作为输入的网络模型，如 SpiderCNN[30]、PointWeb[22]等。所以，本文提出的基于位置关系卷积与深度残差网络的模型具有较强的点云识别能力。

表 2 不同网络模型在 ModelNet40 数据集上的总体精度对比

Table 2 Comparison of recognition accuracy of different network models on ModelNet40 dataset

网络模型	输入	总体精度/%
MVCNN[11]	image	90.1
3DShapeNets[28]	voxel	77.3
VoxNet[15]	voxel	83.0
PointNet[10]	1 000 points	89.2
PointNet++[18]	1 000 points	90.7
PointNet++[18]	5 000 points+normal	91.9
DGCNN[23]	1 000 points	92.9
SpiderCNN[31]	1 000 points+normal	92.4
PointCNN[20]	1 000 points	92.5
PointWeb[22]	1 000 points+normal	92.3
PointConv[25]	1 000 points+normal	92.5
RS-CNN[32]	1 000 points	93.6
KPConv[30]	1 000 points	92.9
PCNN[33]	1 000 points	92.3
DensePoint[34]	1 000 points	93.2
PointASNL[26]	1 000 points	92.9
FPConv[35]	1 000 points	92.5
PAConv[36]	1 000 points	93.2
Point Transformer[37]	1 000 points	93.7
PRCDRNet(本文)	1 000 points	93.9

PRCDRNet 模型在 ModelNet40 数据集上的训练轮次与总体精度的关系如图 8 所示，可以看出训练在 50 轮内快速收敛，在 269 轮训练时达到最高识别精度。

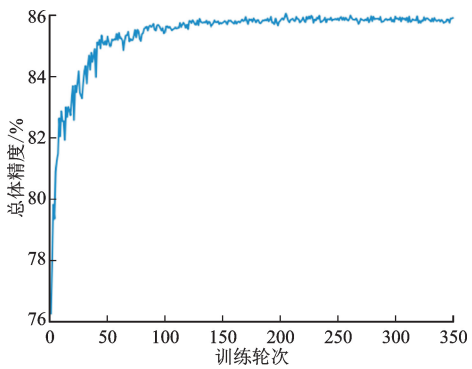

图 8 训练轮次与识别的总体精度

Fig. 8 Epoch and overall accuracy of recognition

3.2　三维点云分割实验

本文选择在 ShapeNet Part 数据集上进行三维点云分割实验。ShapeNet Part 数据集包含 16 个类别的 15 881 个 CAD 模型，每个类别包含 2～6 个不同部件，共有 50 个部件语义标签。在 ShapeNet Part 数据集上采用实例平均交并比 M_{I} 与类别平均交并比 M_{C} 作为网络分割性能的评价指标，公式如下

$$M_{\mathrm{I}}=\frac{\sum_{i=1}^{z}\sum_{i=1}^{p}T_{\mathrm{P}}}{\sum_{i=1}^{z}\sum_{i=1}^{p}T_{\mathrm{P}}+\sum_{i=1}^{z}\sum_{i=1}^{p}F_{\mathrm{P}}+\sum_{i=1}^{z}\sum_{i=1}^{p}T_{\mathrm{N}}} \tag{14}$$

$$M_{\mathrm{C}}=\frac{1}{z}\sum_{i=1}^{z}\frac{\sum_{i=1}^{p}T_{\mathrm{P}}}{\sum_{i=1}^{p}T_{\mathrm{P}}+\sum_{i=1}^{p}F_{\mathrm{P}}+\sum_{i=1}^{p}T_{\mathrm{N}}} \tag{15}$$

式中：z 为类别数；p 为每个类别中点的个数；T_{P} 为预测结果是正类，真实标签也为正类的点；F_{P} 为预测结果是正类，真实标签为负类的点；T_{N} 为预测结果是负类，真实标签为正类的点。

本文 PRCDRNet 模型和其他经典的网络模型在 ShapeNet Part 数据集上的实验结果对比如表 3 所示。可以看出，本文算法在 ShapeNet Part 数据集上的 M_{C} 达到了最高的 84.7%，M_{I} 达到了 86.0%，高于大多数主流网络模型。与 DGCNN[23] 模型相比，本文算法的 PRConv 在 EdgeConv 的基础上融合了几何关系特征丰富了提取到的语义特征，并且将高维空间中的邻域空间搜索改为位置坐标邻域搜索，降低了高维特征信息冗余，使得 M_{I} 与 M_{C} 分别提升了 0.8%与 2.4%。虽然 M_{I} 略逊于 PointASNL[26] 0.1%，但在手提包、车辆、耳机、匕首等 11 个类别实现了最近分割性能。实验结果充分证明，本文方法在点云语义分割任务中具有较强的竞争力。

表 3　不同网络模型在 ShapeNet Part 数据集上的分割结果对比

Table 3　Comparison of segmentation results of different network models on ShapeNet Part dataset

网络模型	类别平均交并比/%	实例平均交并比/%	交并比/%															
			飞机	包	帽子	汽车	椅子	耳机	吉他	匕首	台灯	笔记本	摩托车	杯子	手枪	火箭	滑板	桌子
PointNet[10]	80.4	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++[18]	81.9	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[38]	—	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
SO-Net[39]	—	84.9	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
PCNN[33]	81.8	85.1	82.4	80.1	85.5	79.5	90.8	73.2	91.3	86.0	85.0	95.7	73.2	94.8	83.3	51.0	75.0	81.8
DGCNN[23]	82.3	85.2	84.0	83.4	86.7	77.8	90.6	74.7	91.2	87.5	82.8	95.7	66.3	94.9	81.1	63.5	74.5	82.6
P2Sequence[40]	—	85.2	82.6	81.8	87.5	77.3	90.8	77.1	91.1	86.9	83.9	95.7	70.8	94.6	79.3	58.1	75.2	82.8
PointASNL[26]	—	86.1	84.1	84.7	87.9	79.7	92.2	73.7	91.0	87.2	84.2	95.8	74.4	95.2	81.0	63.0	76.3	83.2
SynSpecCNN[41]	82.0	84.7	81.6	81.7	81.9	75.2	90.2	74.9	93.0	86.1	84.7	95.6	66.7	92.7	81.6	60.6	82.9	82.1
SpiderCNN[31]	82.4	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
PRCDRNet(本文)	84.7	86.0	83.8	84.8	89.1	79.9	90.0	82.4	92.0	87.8	83.0	96.1	77.1	95.2	83.2	64.8	83.5	83.2

PRCDRNet 模型在 ShapeNet Part 数据集上的训练轮次与平均交并比的关系如图 9 所示，可以看出训练大概在 120 轮时收敛，在 205 轮时得到最佳分割结果。

PRCDRNet 模型在 ShapeNet Part 数据集上的分割结果可视化如图 10 所示，第一行为真实标签，第二行为 PRCDRNet 模型分割结果。可以看出，本文算法分割结果已接近于真实标签，可以准确地提取到点云特征。

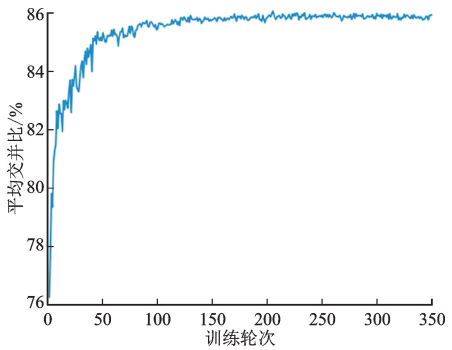

图 9　训练轮次与平均交并比

Fig. 9　Epoch and mean intersection over union

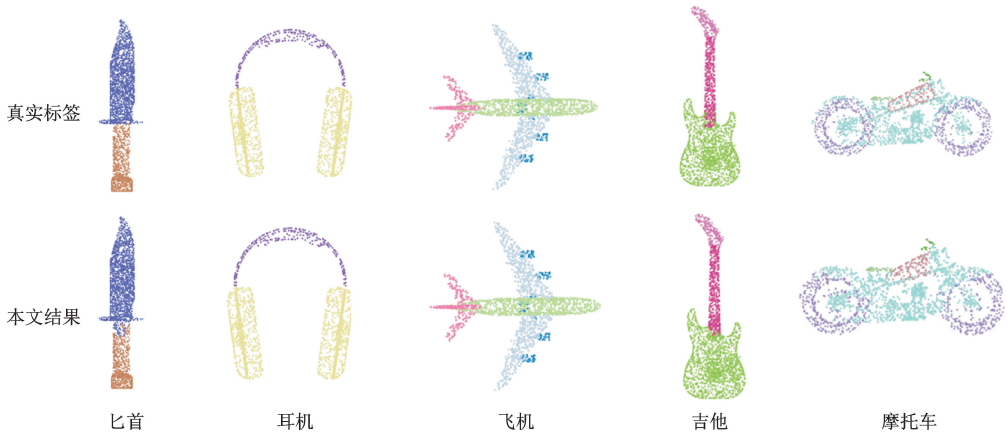

图 10 ShapeNet Part 数据集分割结果可视化

Fig. 10 Visualization of segmentation results on ShapeNet Part dataset

3.3 消融实验

3.3.1 k 近邻点

本文提出的位置关系卷积在提取点云局部特征的过程中采用 k-NN 算法构建局部邻域，邻域点的数目 k 在一定程度上会影响网络提取到的局部特征。较小的邻域点数目构建局部邻域速度相对较快，网络计算量低，但是构建的邻域范围过小无法学习到全面的几何特征，导致精度较低，而邻域点数目过大会导致局部邻域范围重叠，会引入更多的噪声，影响网络对几何特征的学习。当前主流网络模型根据实验将 k 设置为 16、24、32 等数值。此外，不同的数据集点云的规模大小也不统一，不同近邻数目的选择对不同的数据集的影响存在差异。本文分别在 ModelNet40 数据集与 ShapeNet Part 数据集上进行了消融实验，测试了不同近邻数目 k 对识别总体精度与分割精度的影响。从表 4 可以看出，当 k 为 16 时，网络的总体精度与分割精度均较低，网络的总体精度在 k 为 24 时达到最高，当 k 为 32 时，近邻点过多反而导致了网络在 ModelNet40 数据集上总体精度下降。由于点云分割数据集 ShapeNet Part 的输入点是点云识别数据集 ModelNet40 的 2 倍，因此当近邻点较大时，网络能更充分地提取较大邻域内的几何特征，但如果近邻点数目过大也会导致不同邻域内的特征信息重叠冗余，影响分割结果。实验表明，网络的分割性能在 k 为 32 时表现最佳。

3.3.2 基本残差单元

本文提出了深度残差模块旨在加强提取局部邻域中深层特征，缓解因网络层数加深产生的网络退化问题。深度残差模块由如图 4 所示的一个或多个基本残差单元(BRU)串连堆叠组成。为了验证深度残差模块对点云识别与分割精度的影响，分别在 ModelNet40 数据集与 ShapeNet Part 数据集上进行消融实验，实验结果如表 5 所示。可以看出，当深度残差模块由 2 个基本残差单元串联组成时，点云的识别与分割效果最好，相较于没有深度残差模块的网络，总体精度提高 0.5%，分割平均交并比提高 0.2%。在一定配置下，深度残差模块能够加强局部邻域中深层特征的提取，提高点云的识别与分割精度。

表 4 近邻数目对点云识别与分割结果的影响对比

Table 4 Comparison of influence of number of neighbor-hood points on recognition and segmentation results

k	总体精度/%	分割平均交并比/%
16	93.4	85.7
24	93.9	85.9
32	93.7	86.0
40	93.4	85.8

表 5 基本残差单元数目对点云识别与分割精度的影响对比

Table 5 Comparison of influence of number of BRU on recognition and segmentation results

基本残差单元数量	总体精度/%	分割平均交并比/%
0	93.4	85.8
1	93.5	85.9
2	93.9	86.0
3	93.6	85.8

4 结　论

本文提出了一种基于位置关系深度残差神经网络的三维点云识别与分割算法。首先，对原始点云做嵌入操作，获得点云的高维特征表示。其次，通过FPS下采样并用 k-NN 算法确定局部邻域。然后，将局部邻域分别输入位置关系卷积对局部邻域当前点的特征与局部邻域内的几何特征做融合并强化局部邻域的中心特征，通过深度残差模块强化提取深层局部特征。分层重复以上步骤可以获得点云的深层上下文特征，实现了点云的识别与分割处理。本文算法在 ModelNet40 点云分类数据集和 ShapeNet Part 点云分割数据集分别达到了 93.9%总体精度与 86.0%的平均交并比，具有较好的点云识别与分割能力。

本文算法在点云规模较小的识别任务取得了较好的成绩，然而对于大规模点云分割任务的结果还不是很理想，仍然存在较大的改进空间。如何提高算法对大规模点云的分割能力并提升网络的泛化性能，是未来继续研究的方向。

参考文献：

[1] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks [J]. Communications of the ACM, 2017, 60(6): 84-90.

[2] SZEGEDY C, LIU Wei, JIA Yangqing, et al. Going deeper with convolutions [C]//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2015: 1-9.

[3] REN Shaoqing, HE Kaiming, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137-1149.

[4] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2016: 779-788.

[5] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation [C]//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2015: 3431-3440.

[6] RONNEBERGER O, FISCHER P, BROX T. U-Net: convolutional networks for biomedical image segmentation [C]//Medical Image Computing and Computer-Assisted Intervention-MICCAI 2015. Cham, Germany: Springer International Publishing, 2015: 234-241.

[7] SHI Shaoshuai, WANG Xiaogang, LI Hongsheng. PointRCNN: 3D object proposal generation and detection from point cloud [C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2019: 770-779.

[8] WANG D Z, POSNER I. Voting for voting in online point cloud object detection [C]//Robotics: Science and Systems. Rome, Italy: The MIT Press, 2015: 10-15.

[9] CHOY C B, XU Danfei, GWAK J Y, et al. 3D-R2N2: a unified approach for single and multi-view 3D object reconstruction [C]//Computer Vision-ECCV 2016. Cham, Germany: Springer International Publishing, 2016: 628-644.

[10] CHARLES R Q, SU Hao, KAICHUN Mo, et al. PointNet: deep learning on point sets for 3D classification and segmentation [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2017: 77-85.

[11] SU Hang, MAJI S, KALOGERAKIS E, et al. Multi-view convolutional neural networks for 3D shape recognition [C]//2015 IEEE International Conference on Computer Vision (ICCV). Piscataway, NJ, USA: IEEE, 2015: 945-953.

[12] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [EB/OL]. [2022-08-01]. https://doi.org/10.48550/arXiv.1409.1556.

[13] 冯元力，夏梦，季鹏磊，等. 球面深度全景图表示下的三维形状识别 [J]. 计算机辅助设计与图形学学报，2017，29(9)：1689-1695.
FENG Yuanli, XIA Meng, JI Penglei, et al. Deep spherical panoramic representation for 3D shape recognition [J]. Journal of Computer-Aided Design & Computer Graphics, 2017, 29(9): 1689-1695.

[14] WEI Xin, YU Ruixuan, SUN Jian. View-GCN: view-based graph convolutional network for 3D shape analysis [C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2020: 1847-1856.

[15] 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). Piscataway,

NJ, USA: IEEE, 2015: 922-928.

[16] CHOY C, GWAK J Y, SAVARESE S. 4D spatio-temporal ConvNets: Minkowski convolutional neural networks [C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2019: 3070-3079.

[17] 杨军, 王顺, 周鹏. 基于深度体素卷积神经网络的三维模型识别分类 [J]. 光学学报, 2019, 39(4): 306-316.
YANG Jun, WANG Shun, ZHOU Peng. Recognition and classification for three-dimensional model based on deep voxel convolution neural network [J]. Acta Optica Sinica, 2019, 39(4): 306-316.

[18] QI C R, YI Li, SU Hao, et al. PointNet++: deep hierarchical feature learning on point sets in a metric space [C]//Advances in Neural Information Processing Systems. New York, NJ, USA: Curran Associates, Inc., 2018: 828-838.

[19] 周鹏, 杨军. 采用神经网络架构搜索的三维模型分类 [J]. 计算机辅助设计与图形学学报, 2022, 34(5): 722-733.
ZHOU Peng, YANG Jun. 3D model classification based on neural architecture search [J]. Journal of Computer-Aided Design & Computer Graphics, 2022, 34(5): 722-733.

[20] LI Yangyan, BU Rui, SUN Mingchao, et al. PointCNN: convolution on *X*-transformed points [C]//Advances in Neural Information Processing Systems. New York, NJ, USA: Curran Associates, Inc., 2018: 828-838.

[21] DANG Jisheng, YANG Jun. HPGCNN: hierarchical parallel group convolutional neural networks for point clouds processing [C]//Computer Vision-ACCV 2020. Cham, Germany: Springer International Publishing, 2021: 20-37.

[22] ZHAO Hengshuang, JIANG Li, FU C W, et al. PointWeb: enhancing local neighborhood features for point cloud processing [C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2019: 5560-5568.

[23] WANG Yue, SUN Yongbin, LIU Ziwei, et al. Dynamic graph CNN for learning on point clouds [J]. ACM Transactions on Graphics, 2019, 38(5): 146.

[24] 杨军, 李博赞. 基于自注意力特征融合组卷积神经网络的三维点云语义分割 [J]. 光学精密工程, 2022, 30(7): 840-853.
YANG Jun, LI Bozan. Semantic segmentation of 3D point cloud based on self-attention feature fusion group convolutional neural network [J]. Optics and Precision Engineering, 2022, 30(7): 840-853.

[25] WU Wenxuan, QI Zhongang, LI Fuxin. PointConv: deep convolutional networks on 3D point clouds [C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2019: 9613-9622.

[26] YAN Xu, ZHENG Chaoda, LI Zhen, et al. PointASNL: robust point clouds processing using nonlocal neural networks with adaptive sampling [C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2020: 5588-5597.

[27] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Deep residual learning for image recognition [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2016: 770-778.

[28] WU Zhirong, SONG Shuran, KHOSLA A, et al. 3D ShapeNets: a deep representation for volumetric shapes [C]//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2015: 1912-1920.

[29] YI Li, KIM V G, CEYLAN D, et al. A scalable active framework for region annotation in 3D shape collections [J]. ACM Transactions on Graphics, 2016, 35(6): 210.

[30] THOMAS H, QI C R, DESCHAUD J E, et al. KPConv: flexible and deformable convolution for point clouds [C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway, NJ, USA: IEEE, 2019: 6410-6419.

[31] XU Yifan, FAN Tianqi, XU Mingye, et al. SpiderCNN: deep learning on point sets with parameterized convolutional filters [C]//Computer Vision-ECCV 2018. Cham, Germany: Springer International Publishing, 2018: 90-105.

[32] LIU Yongcheng, FAN Bin, XIANG Shiming, et al. Relation-shape convolutional neural network for point cloud analysis [C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2019: 8887-8896.

[33] ATZMON M, MARON H, LIPMAN Y. Point convolutional neural networks by extension operators [J]. ACM Transactions on Graphics, 2018, 37(4): 71.

[34] LIU Yongcheng, FAN Bin, MENG Gaofeng, et al. DensePoint: learning densely contextual representation

for efficient point cloud processing [C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway, NJ, USA: IEEE, 2019: 5238-5247.

[35] LIN Yiqun, YAN Zizheng, HUANG Haibin, et al. FPConv: learning local flattening for point convolution [C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2020: 4292-4301.

[36] XU Mutian, DING Runyu, ZHAO Hengshuang, et al. PAConv: position adaptive convolution with dynamic kernel assembling on point clouds [C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2021: 3172-3181.

[37] ZHAO Hengshuang, JIANG Li, JIA Jiaya, et al. Point transformer [C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway, NJ, USA: IEEE, 2021: 16239-16248.

[38] KLOKOV R, LEMPITSKY V. Escape from cells: deep Kd-networks for the recognition of 3D point cloud models [C]//2017 IEEE International Conference on Computer Vision (ICCV). Piscataway, NJ, USA: IEEE, 2017: 863-872.

[39] LI Jiaxin, CHEN B M, LEE G H. SO-Net: self-organizing network for point cloud analysis [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ, USA: IEEE, 2018: 9397-9406.

[40] LIU Xinhai, HAN Zhizhong, LIU Yushen, et al. Point2 Sequence: learning the shape representation of 3D point clouds with an attention-based sequence to sequence network [C]//Proceedings of the AAAI Conference on Artificial Intelligence. Palo Alto, CA, USA: AAAI Press, 2019: 8778-8785.

[41] YI Li, SU Hao, GUO Xingwen, et al. SyncSpecCNN: Synchronized spectral CNN for 3D shape segmentation [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2017: 6584-6592.

（编辑　亢列梅）