RFNet: 用于三维点云分类的卷积神经网络

单铉洋 [1] 孙战里 [2] 曾志刚 [3]

摘 要 由于点云的非结构性和无序性, 目前已有的点云分类网络在精度上仍然需要进一步提高. 通过考虑局部结构的构建、全局特征聚合和损失函数改进三个方面, 构造一个有效的点云分类网络. 首先, 针对点云的非结构性, 通过学习中心点特征与近邻点特征之间的关系, 为不规则的近邻点分配不同的权重, 以此构建局部结构; 然后, 使用注意力思想, 提出加权平均池化 (Weighted average pooling, WAP), 通过自注意力方式, 学习每个高维特征的注意力分数, 在应对点云无序性的同时, 可以有效地聚合冗余的高维特征; 最后, 利用交叉熵损失与中心损失之间的互补关系, 提出联合损失函数 (Joint loss function, JL), 在增大类间距离的同时, 减小类内距离, 进一步提高了网络的分类能力. 在合成数据集 ModelNet40、ShapeNetCore 和真实世界数据集 ScanObjectNN 上进行实验, 与目前性能最好的多个网络相比较, 验证了该整体网络结构的优越性.

关键词 深度学习, 三维点云, 点云分类, 注意力机制, 损失函数

引用格式 单铉洋, 孙战里, 曾志刚. RFNet: 用于三维点云分类的卷积神经网络. 自动化学报, 2023, **49**(11): 2350−2359

DOI 10.16383/j.aas.c210532

RFNet: Convolutional Neural Network for 3D Point Cloud Classification

SHAN Xuan-Yang[1] SUN Zhan-Li[2] ZENG Zhi-Gang[3]

Abstract Due to the unstructured and disordered nature of point cloud, the classification accuracy is still needed to improve for the currently existed point cloud classification approaches. In this paper, an effective point cloud classification network is proposed by considering local structure construction, global feature aggregation, and loss function improvement. Firstly, in view of the unstructured nature of point cloud, the local structure is constructed via assigning different weights to those irregular neighborhood points, which are computed via the relationship between the center point feature and its neighbor feature. Moreover, a strategy of weighted average pooling (WAP) is designed to learn the attention score of each high-dimensional feature through self-attention mechanization, which can effectively aggregate redundant high-dimensional features while coping with point cloud disorder. In addition, a joint loss function (JL) is presented by utilizing the complementary relationship between the cross-entropy loss and the center loss, which increases the inter-class distance while reduces the intro-class distance, and further improves the classification ability of the network. Compared to several state-of-the-art networks, the experimental results on the synthetic dataset ModelNet40 and ShapeNetCore and the real-world dataset ScanObjectNN demonstrate the superiority of the overall network structure.

Key words Deep learning, 3D point cloud, point cloud classification, attention mechanism, loss function

Citation Shan Xuan-Yang, Sun Zhan-Li, Zeng Zhi-Gang. RFNet: Convolutional neural network for 3D point cloud classification. *Acta Automatica Sinica*, 2023, **49**(11): 2350−2359

近年来, 随着自动驾驶仪、高精度地图、智慧城市等新概念的出现, 许多应用场景都需要基于点云的 3D 环境感知和交互. 三维扫描技术和深度摄像机的快速发展, 使得点云的获取更加简单方便, 在三维数据深度学习领域, 点云逐渐成为一种流行的三维数据表达方式[1]. 但与二维图像不同, 由于采样不均匀、传感器精度等因素, 点云在空间上是高度稀疏无序的. 因此, 如何用深度学习的方式处理点云数据, 成为一个关键的问题[2].

由于卷积神经网络 (Convolutional neural network, CNN) 方法在图片处理上有着广泛且成熟的应用, 而点云并不是像图片一样规则排列的, 因此许多工作都是先对点云进行预处理, 将点云数据转化为规则的数据, 然后再使用传统的 CNN 方法[3].

收稿日期 2021-06-15 录用日期 2022-02-10
Manuscript received June 15, 2021; accepted February 10, 2022
国家自然科学基金 (61972002) 资助
Supported by National Natural Science Foundation of China (61972002)
本文责任编委 陈胜勇
Recommended by Associate Editor CHEN Sheng-Yong
1. 安徽大学电气工程与自动化学院 合肥 230601 2. 安徽大学人工智能学院 合肥 230601 3. 华中科技大学人工智能与自动化学院 武汉 430000
1. School of Electrical Engineering and Automation, Anhui University, Hefei 230601 2. School of Artificial Intelligence, Anhui University, Hefei 230601 3. School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, Wuhan 430000

根据神经网络输入数据类型的不同, 现有的点云分类方法可以分为多视图法、体素法和基于点的方法 3 种[4]. 多视图法是将三维点云数据从不同角度投影到二维平面上, 然后再对二维平面图做二维卷积[5−7]. 但由于存在遮挡问题, 在三维向二维投影的过程中, 会导致三维点云中固有信息的大量丢失. 体素法是通过将点云转化为规则的三维网格, 在三维网格上使用卷积神经网络[8−9]. 但这种方法受到分辨率的限制, 低分辨率训练速度快, 但会造成点云信息的大量丢失; 高分辨率形状描述能力强, 但同时也会让空间复杂度大幅度提高.

基于点的方法是将直接点云的三维坐标 (x, y, z) 作为输入, 对坐标进行卷积. PointNet[3] 是这条路线的先驱. 为克服点云无序性而直接处理点云数据, PointNet 首先对输入点云使用多个共享的多层感知器 (Multi-layer perceptron, MLP), 从而将低维的坐标信息映射为高维特征, 然后使用对称聚合函数来提取最终的全局特征, 最后使用全连接层 (Fully connected layer, FC) 进行最终分类[3]. 然而在这种设计中, 每个点的特征都是独立学习的, 忽略了点与点之间的结构关系. 为此, PointNet++[10] 把点云划分为多个局部, 在每个局部中使用 PointNet 提取局部特征, 然后对提取到的局部特征进行局部划分并使用 PointNet, 通过多次叠加, 最终获取全局特征. PointNet++ 在一定程度上解决了 PointNet 缺少局部特征的问题, 但这种方法只是简单地划分局部, 并没有进一步说明局部点之间的关系. 谱图卷积 (Spectral graph convolution, SpecGCN)[11] 使用递归聚类和池化策略替换标准的最大池化过程, 用于聚合来自其谱坐标中彼此接近的节点簇内信息, 从而产生更丰富的整体特征描述符. 但这种方法需要将三维点坐标转换到谱坐标中, 增大了网络的复杂度. 动态图卷积神经网络 (Dynamic graph convolutional neural network, DGCNN)[12] 是一种边缘卷积方法, 通过与近邻点构建局部图结构, 然后使用共享的 MLP 层, 对图的每一个边缘进行卷积获取邻边特征, 最后使用最大池化方式对每一个邻边特征进行聚合, 获取局部特征, 并在网络的每一层后进行动态更新. 这种方法增强了局部结构中点与点之间的关联, 但由于点云并不是均匀分布的, 近邻点对于中心点的影响也是不同的, 而边缘卷积平等地处理每一个邻边, 这种提取方式是不准确的. RSCNN (Relation-shape convolutional neural network)[13] 通过从三维点云之间的几何关系, 学习隐含的形状信息, 从而构建局部. 这种利用局部结构三维几何特征的方法可以方便地提取低维特征, 但对高维特征的提取有局限性.

近年来, 注意力网络在神经网络学习领域已成为一个重要概念, 并在不同领域有着广泛应用. 注意力机制通过加权方式来提高有用信息, 抑制无用信息. 在点云分析上注意力网络也有了一些应用. 例如图注意力卷积网络可以根据动态学习的特征, 有选择地关注其中最相关部分[14]. 基于点网络的图卷积[15] 通过学习每个点的注意力特征, 然后利用多头机制获取特征, 最后利用注意力池化层来捕捉重要信号. 点云变压器 (Point cloud transformer, PCT)[16] 网络的主要思想是通过使用 Transformer 固有的顺序不变性避免定义点云数据的顺序和通过注意力机制进行特征学习. 但是, 相较于卷积神经网络, 注意力模块有更多的参数, 大量使用注意力模块会使整体网络的训练难度增加.

为了更好地应对不规则点云带来的问题和现有网络存在的问题, 本文提出基于特征关系的神经网络 (Network based on relationships between features, RFNet). 在构建点云局部结构上, RFNet 使用基于特征关系的卷积 (Convolution based on relationships between features, RFConv) 方法, 来增强局部结构中近邻点特征与中心点特征之间的联系. 首先, 通过描述局部结构中近邻点与中心点之间的多种关系学习近邻点权重; 然后, 将加权后的近邻点与中心点进行组合构成局部; 最后, 再对组合后的局部进行学习聚合为高维的局部特征. 另外, 本文针对对称函数和联合损失函数 (Joint loss function, JL) 提出改进方法. 在现有的网络中, 对称函数多是使用最大池化和平均池化这样硬性的聚合高维特征方法, 会导致信息的损失, 为此, 本文基于注意力思想, 提出加权平均池化 (Weighted average pooling, WAP). 另外, 本文利用交叉熵损失函数和中心损失函数的互补优势, 组成联合损失函数, 在增大类间距离的同时, 减小类内距离, 从而增强网络的分类性能[17]. 本文主要贡献如下:

1) 提出一个全新的 CNN 模块, 可以有效聚合点云局部和全局特征;

2) 引入联合损失函数, 在减小点云类内距离的同时, 增大点云类间距离, 增强网络分类性能;

3) 在合成数据集和真实世界数据集上的实验结果表明, 本文网络与现有网络相比, 有着更为明显的优势.

1 算法介绍

RFNet 是基于点的方法实现的. 整体网络结构如图 1 所示, 其中 C 为物体的类别数. 经典的基于点的方法的网络一般以 MLP 网络和对称聚合函数为基础. MLP 网络的作用是将低维点特征映射到

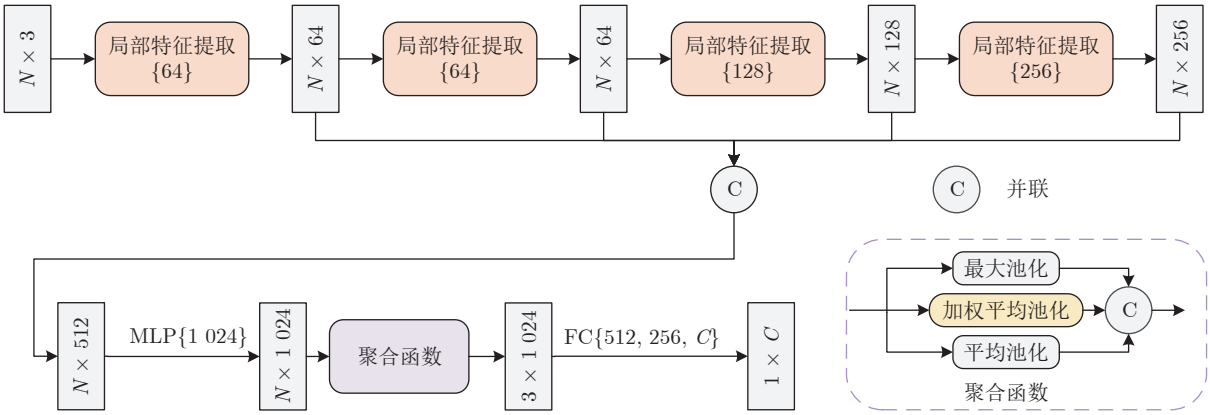


图 1 整体网络结构图

Fig. 1 Diagram of overall network structure

高维空间中来提取高维点特征. 在此过程中, MLP 网络首先对点特征进行 1×1 的卷积, 然后进行批标准化, 最后通过激活函数为网络增加非线性因素, 增强网络表达能力.

$$\mathcal{M}(\cdot) = \sigma\left(bn\left(c_{1\times1}(\cdot)\right)\right) \tag{1}$$

式中, σ 是激活函数, bn 是批标准化, $c_{1\times1}$ 是 1×1 的卷积操作. 对称聚合函数可以保证不论网络输入特征的顺序如何改变, 都能使整体网络得到一个不变的全局特征, 从而应对点云的无序性:

$$F_o = \mathcal{A}\left(f_1, f_2, f_3, \cdots, f_l\right) \tag{2}$$

式中, $\mathcal{A}$ 为对称聚合函数, F_o 为全局特征, f_l 为第 l 层的高维点特征. 在现有网络中, 对称聚合函数一般选用最大池化和平均池化.

1.1 基于特征局部结构的卷积操作

现有的网络已经证明点云局部结构的构建对于点云高维特征的提取是至关重要的, 为此本文提出基于特征关系的局部构建方法.

1) 局部结构构建. 对于输入的点云集合有 $\boldsymbol{P} = \{X_i | i = 1, 2, 3, \cdots, N\} \in \mathbf{R}^{N\times d}$, 其中 X_i 表示第 i 个点的特征向量, N 为点的个数, d 为 X_i 的特征维度. 对于第 1 层特征向量来说, 维度 $d = 3$ 即每个点的三维坐标 $X_i = (x_i, y_i, z_i)$, 随着网络深度的增加, 特征维度 d 会不断增大. 中心点和近邻点的局部结构如图 2 所示. 首先, 在本文网络中, 对于输入的每一个点云特征 X_i 都视为点云的中心点; 然后, 利用 K 近邻 (K-nearest neighbors, KNN) 算法, 通过计算中心点 X_i 到其余点之间的欧氏距离, 选取距离最近的 K 个点作为中心的近邻点, X_{ij} $(j = 1, 2, 3, \cdots, K) \in \mathbf{R}^d$; 最后, 将中心点与近邻点组合构成一个局部. 由于近邻点并不是均匀分布在中心点周围, 因此对于不同位置的近邻点, 其特征对中心点特征的影响也是不同的. 所以, 本文根据近邻点与中心点之间的不同特征关系, 对近邻点进行权重分配:

$$\mathcal{L}_\theta = (X_i, \theta_{ij} \odot X_{ij}) \tag{3}$$

式中, $\theta_{ij} \in \mathbf{R}^d$ 是通过特征关系学习到的权重, $\odot$ 为点乘操作.

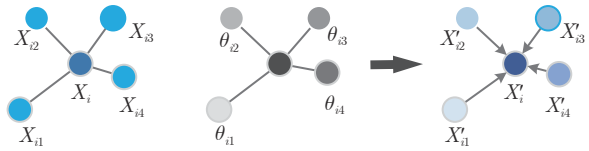


图 2 中心点和近邻点的局部结构图

Fig. 2 Diagram of local structure between the central point and its neighbors

2) 特征关系描述. 本文通过对中心点特征与近邻点特征之间的关系进行描述, 对不同的特征关系进行组合, 获取近邻点对于中心点影响的权重. 具体学习方法如下:

$$\theta_{ij} = \mathcal{M}\left(e_{ij} \oplus l_{ij} \oplus s_{ij}\right) \tag{4}$$

其中

$$\begin{cases} e_{ij} = X_i - X_{ij} \\ l_{ij} = ||X_i - X_{ij}|| \\ s_{ij} = \dfrac{\displaystyle\sum_{m=1}^{d}(X_{i_m} - \bar{X}_i)(X_{ij_m} - \bar{X}_{ij})}{\sqrt{\displaystyle\sum_{m=1}^{d}(X_{i_m} - \bar{X}_i)^2}\sqrt{\displaystyle\sum_{m=1}^{d}(X_{ij_m} - \bar{X}_{ij})^2}} \end{cases} \tag{5}$$

式中, $\oplus$ 为并联操作, 总体并联后的维度为 $\mathbf{R}^{d+2}$; $e_{ij} \in \mathbf{R}^d$ 为中心点与近邻点之间的向量信息; $l_{ij} \in \mathbf{R}$ 为中心点特征与近邻点特征的欧氏距离; $s_{ij} \in \mathbf{R}$ 为中心点与近邻点之间的余弦相似度, 其中 $\bar{X}_i$、

$\bar{X}_{ij}$ 分别表示向量 X_i、X_{ij} 的均值; $\theta_{ij} \in \mathbf{R}^d$ 为近邻特征的权重; $\mathcal{M}$ 为式 (1) 的 MLP 操作, 用来提取权重信息并将权重的维度由 $d+2$ 调整为 d.

3) 局部特征提取. 首先将中心点特征 X_i 分别与加权后的 K 个近邻点特征并联构建局部结构, 然后对构建好的局部结构使用共享的 MLP 提取 K 个高维的邻边特征, 最后使用最大池化函数将提取的 K 个邻边特征聚合为 X_i 为中心的局部特征 L_i:

$$L_i = \mathrm{Max}(\mathcal{M}(X_i \oplus \theta_{ij} \odot X_{ij})) \tag{6}$$

聚合后的局部特征 L_i 即为下一层的特征输入, 其中 $L_i \in \mathbf{R}^{d'}$, d' 表示通过 MLP 升维后的维度. 局部特征提取的网络结构如图 3 所示, 其中 $\boldsymbol{P} = \{X_i | i = 1, 2, 3, \cdots, N\}$为中心点特征 X_i 的集合, $\boldsymbol{\theta} = \{\theta_{ij} | i = 1, 2, 3, \cdots, N, j = 1, 2, 3, \cdots, K\}$为权重 θ_{ij}的集合, $\boldsymbol{L} = \{L_i | i = 1, 2, 3, \cdots, N\}$ 为局部特征 L_i 的集合.

1.2　聚合函数

在全局特征的聚合过程中, 现有的网络算法大多是以最大池化函数或平均池化函数硬性聚合高维特征, 这样可以保持网络的对称性, 但并不能准确地学习到有用信息, 同时还造成信息的大量丢失. 为了有选择地聚合高维特征, 同时还要保持整体网络的对称性, 本文基于注意力机制, 提出加权平均池化, 通过使用最大池化、平均池化和加权平均池化对特征进行聚合.

1) 特征输入. 如图 1 所示, 经过 4 次局部特征提取, 将每一层提取到的局部特征 L_i 做并联操作, 再经过 MLP 网络提取特征并将维度升级到 1 024 维, 得到最终的高维特征输入 $f_i \in \mathbf{R}^{1\,024}$:

$$f_i = \mathcal{M}\left(L_i^{64} \oplus L_i^{64} \oplus L_i^{128} \oplus L_i^{256}\right) \tag{7}$$

2) 加权平均池化. 如图 4 所示, 对于 N 个高维特征输入 f_i, 首先经过 MLP 学习特征分数, 然后经过 softmax 来将特征分数归一化:

$$s_i = \mathrm{softmax}(\mathcal{M}(f_i)) = \frac{\mathrm{e}^{f_i}}{\sum_{j=1}^{N} \mathrm{e}^{f_j}} \tag{8}$$

最后, 将学习到的特征分数与相对应的特征相乘并聚合, 以获取加权池化后的全局特征:

$$F_o = \mathcal{A}(f_i) = \sum_{i=1}^{N} s_i f_i \tag{9}$$

式中, f_i 为第 i 个高维特征, 其集合为 $\boldsymbol{F} = \{f_i | i = 1, 2, 3, \cdots, N\}$; s_i 为第 i 个高维特征的特征分数, 其集合为 $\boldsymbol{S} = \{s_i | i = 1, 2, 3, \cdots, N\}$.

1.3　联合损失函数

现有点云网络的损失函数大多是采用交叉熵损失函数. 交叉熵损失函数可以使类间距离增大[17], 但交叉熵损失函数对类内距离的影响较小, 这样学习到的特征包含较大类内变化. 为此, 本文引用中心损失函数, 采用交叉熵损失和中心损失联合监督的方法, 进一步减小类内距离.

1) 中心损失函数. 中心损失函数通过对每一个类随机生成一个特征中心, 然后在训练的过程中对每一个特征中心进行更新学习, 使深层特征与其对应的类特征中心之间的距离最小化, 具体的中心损失函数为:

$$L_C = \frac{1}{2} \sum_{i=1}^{m} ||F_i - c_{y_i}||_2^2 \tag{10}$$

式中, c_{y_i} 表示深度特征 y_i 类的特征中心, $F_i \in \mathbf{R}^d$ 表示深层次的特征, m 表示输入批次的大小. 因此, 中心损失就是希望一个批次中的每个样本的特征离特征中心的距离的平方和越小越好, 也就是类内距离越小越好.

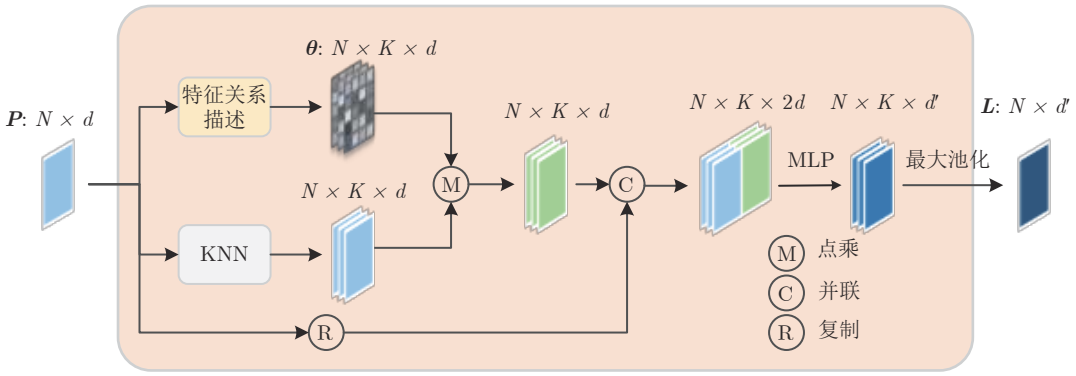


图 3　局部特征提取结构图

Fig. 3　Diagram of local feature extraction structure

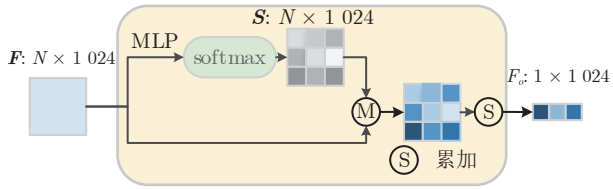


图 4 加权平均池化结构图

Fig. 4 Diagram of weighted average pooling structure

2) 联合损失函数. 联合损失结构如图 5 所示. 通过将交叉熵损失与中心损失进行联合监督, 在增大类间距离的同时, 减小类内距离, 使其获得的特征具有更强识别能力, 具体表达式如下:

$$Loss = L_{CE} + \lambda L_C \tag{11}$$

$$L_{CE} = -\sum_{i=1}^{m}\sum_{j=1}^{n} P(y_{ij})\log_2(Q(y_{ij})) \tag{12}$$

式中, L_{CE} 为交叉熵损失函数; n 为分类的类别数; $P(y_{ij})$ 代表 y_i 类对应的真实标签, 也就是 n 个类别的预测概率; $Q(y_{ij})$ 代表 y_i 类的预测值; L_C 为中心损失函数; λ 参数用以调节两个损失的比重, 当 $\lambda = 0$ 时, 该函数表示为仅有交叉熵损失情况.

2 实验结果与分析

首先, 介绍实验所使用的数据集以及网络的基本设置; 然后, 将本文网络与现有典型网络在不同数据集上进行对比, 并对实验结果做进一步的分析; 最后, 为了说明本文网络模块的有效性和网络的抗干扰能力, 进行了消融实验和鲁棒性实验. 此外, 为了说明网络的资源占用, 还进行了复杂度实验.

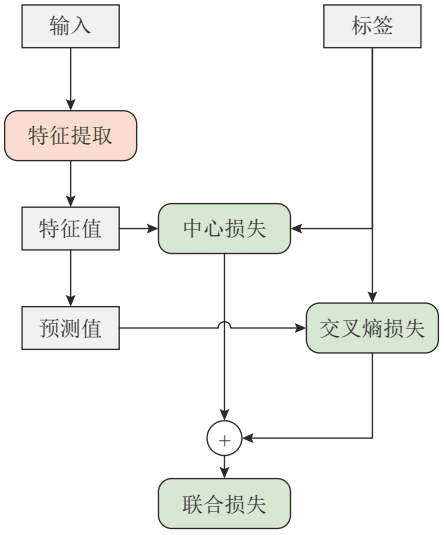


图 5 联合损失结构图

Fig. 5 Diagram of joint loss structure

2.1 数据集介绍

为了评估本文提出的点云分类算法的有效性, 选用标准的合成数据集 ModelNet40[18]、ShapeNetCore[19] 以及由真实世界对象点云组成的 ScanObjectNN[20] 进行实验结果对比.

1) ModelNet40 数据集. 作为点云分析中应用最广泛的基准, ModelNet40 因其简洁的形状和构造良好等特点而广受欢迎. 该数据集共有 12 311 个物体, 包括 40 个类别, 其中 9 843 个模型用作训练数据集, 其余 2 468 个模型为测试数据集[18].

2) ShapeNetCore 数据集是一个丰富标注的大规模点云数据集, 相较于 ModelNet40, ShapeNetCore 包含更为丰富的模型, 其中包含 55 种常见的物品类别和 513 000 个三维模型. 因此在训练时, 也更具挑战性[19].

3) ScanObjectNN 数据集. 为了进一步说明本文分类网络的有效性和鲁棒性, 本文还在以真实世界对象为模型的数据集 ScanObjectNN 上进行实验. 该数据集有 15 个类别, 包含约 15 000 个对象[20]. 由于复杂的背景、缺少的部件和各种变形, 使得 ScanObjectNN 在训练时, 有更高的挑战性.

2.2 实验设置

本文实验使用的硬件配置和软件版本如表 1 所示.

表 1 实验配置

Table 1 Experimental configuration

项目	详情
操作系统	CentOS Linux7
GPU	Tesla V100 32 GB
CUDA	10.0
CUDNN	7.0
Python	3.6.0
Pytorch	1.3.0

网络参数的设置. 网络的优化器选择了随机梯度下降法, 将动量设置为 0.9, 并使用了余弦退火法, 将学习速率从 0.1 降低到 0.001. 训练批次设置为 32, 测试批次设置为 16, 并最终进行了 500 次训练. 激活函数使用负向斜率为 0.2 的 Leaky-ReLU 激活函数. 在每个 MLP 层后, 都加入比例为 0.5 的暂退率, 以抑制过拟合.

2.3 分类实验

本文实验的输入数据为只含有 1 024 个三维坐标的点云数据, 在表 2 中表示为 1 k, 2 k 表示输入

的是含有 2048 个三维坐标的点云数据, 5 k 表示输入的是含有 5120 个三维坐标的点云数据, nor 表示输入额外的法向量信息, voting 表示实验使用了多次投票验证策略. 本文使用最常用的整体精度 (Overall accuracy, OA) 和平均精度 (Mean accuracy, mAcc) 性能标准, 进行分类性能评价. 其中, 整体精度代表所有测试实例的平均精度, 平均精度代表所有类别的平均精度.

表 2 给出了 RFNet 与目前主流网络在 ModelNet40 数据集上的实验结果. 由表 2 实验结果可以看出, RFNet 整体精度为 93.6%, 平均精度为 91.2%, 在与同样输入为 1 k的网络情况下, 取得了最优的结果. RFNet 甚至超过了一些输入额外信息的网络, 例如 PointNet++ 和 SpiderCNN[21] 在输入了 5 k 个点和法向量信息后, 整体精度仍低于本文中的 FRNet. 另外, RFNet 超越了输入同样是 1 k 的 RSCNN, 与使用了 10 次投票评估方法后的 RSCNN 网络相同. 本文在 ModelNet40 数据集上进行输入为 2 k 的实验, 最终取得了整体精度 94.0% 和平均精度 91.4% 的好成绩, 超越了所有输入额外信息的网络.

表 2　在 ModelNet40 数据集上的实验结果
Table 2　Experimental results on ModelNet40

方法	输入	mAcc (%)	OA (%)
PointNet	1 k	86.0	89.2
PointNet++	1 k	—	90.7
Spec-GCN	1 k	—	91.5
DGCNN	1 k	90.2	92.9
RSCNN	1 k	—	92.9
PCT	1 k	—	93.2
ECC[22]	1 k	83.2	87.4
RMFP-DNN[23]	1 k	88.9	92.6
PointCNN[24]	1 k	88.1	92.2
Point2sequence[25]	1 k	90.4	92.6
Point Transformer [26]	1 k	—	92.8
Octant-CNN[27]	1 k	88.7	91.9
DRNet[28]	1 k	—	93.1
AdaptConv[29]	1 k	90.7	93.4
RFNet	**1 k**	**91.2**	**93.6**
PointNet++	5 k + nor	—	91.9
DGCNN	2 k	90.7	93.5
RSCNN	1 k + voting	—	93.6
SpiderCNN	5 k + nor	—	92.4
RFNet	**2 k**	**91.4**	**94.0**

此外, 本文在 ModelNet40 数据集上进行了真实标签和预测标签的混淆实验, 实验结果如图 6 所示. 由图 6 可以看出, 网络最易混淆的是标签为 15 的花盆类, 这与数据集的标定有关. 为了进一步探究预测结果, 进行了易错模型的分类对比实验, 实验结果如表 3 所示. 由表 3 可以看出, RFNet 相较于 PointNet 和 DGCNN 有着更强的细节分辨能力, 能较好地区分易混模型. 由表 3 也可以看出, ModelNet40 数据集中部分花盆类的标定有较大的歧义.

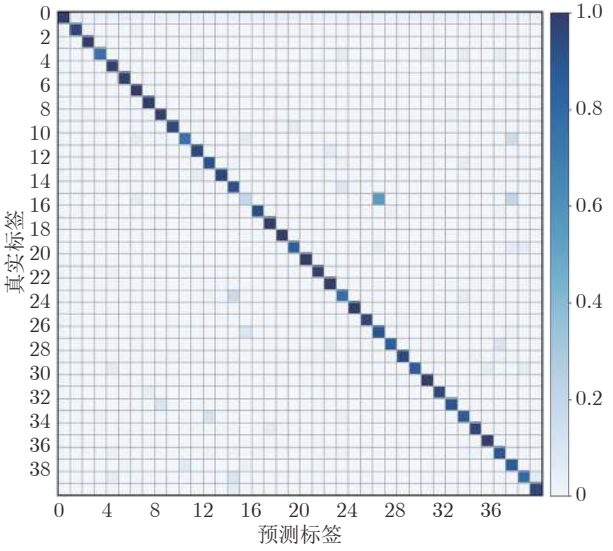


图 6　混淆矩阵
Fig. 6　Confusion matrix

表 3　不同网络的易错模型分类对比
Table 3　Classification comparison of error-prone models for different networks

易错模型	真实标签	PointNet	DGCNN	RFNet
	植物	植物	植物	**植物**
	花盆	植物	植物	植物
	花瓶	花盆	花瓶	**花瓶**
	花瓶	瓶子	花瓶	**花瓶**
	杯子	花盆	花盆	**杯子**

为了验证整体网络应对更丰富分类数据时的能力, 本文在 ShapeNetCore 数据集上进行实验, 实验结果见表 4. 在输入点数 1 k 情况下, RFNet 整体精度为 88.3%. 可以看出, ShapeNetCore 数据集更为丰富的分类确实给网络带来了不小的挑战. 但对比其他网络, RFNet 依旧领先, 甚至高于输入额外的法向量信息的 SpiderCNN.

表 4 在 ShapeNetCore 数据集上的实验结果
Table 4 Experimental results on ShapeNetCore

方法	输入	OA (%)
PointNet	1 k	83.7
PointNet++	1 k	85.1
DGCNN	1 k	84.7
Point2sequence	1 k	85.2
SpiderCNN	1 k + nor	85.3
KPConv[30]	1 k	86.2
RFNet	**1 k**	**88.3**

为了说明 RFNet 在面对真实世界数据时的表现, 本文在 ScanObjectNN 数据集上进行实验, 实验结果见表 5. 实验的输入依旧是 1 k 点云坐标, 最终取得了整体精度 79.6% 和平均精度 76.3% 的成绩. 与其他网络对比可以看出, RFNet 无论是在整体精度还是平均精度上, 都取得了最优异成绩; 另外, 表中还对 ScanObjectNN 数据集中的 15 个类别 (Bag, Bin, Box, ···) 的分类精度进行了对比, 其中有 6 个分类都取得了最高精度, 表明了 RFNet 的优越性.

2.4 不同模块的消融实验

为了验证网络中不同组件的功能和有效性, 本文对 RFNet 中的网络模块进行了消融实验. 实验在数据集 ModelNet40 上进行. 表 6 显示了不同模块的消融实验结果, 表中“×”表示不使用对应模块, “√”表示使用对应模块.

网络 0 表示在只使用普通的动态图网络时的分类结果. 在网络 1 中, 将图结构模块换成了 RFConv 模块, 网络的整体分类精度也提高到了 93.3%. 这归因于 RFConv 模块相较于图结构, 可以通过特征加权的方式更有效地构建局部; 同时可以看出, 局部结构的构建对特征提取有着很大的影响. 网络 2 在网络 1 基础上, 增加加权平均池化模块, 其整体的分类精度提高了 0.2%. 加权平均池化模块可以通过自注意力方式, 对不同层次的输入特征进行特征聚合, 并且注意力机制也可以通过学习特征分数, 在一定程度上弱化无效特征, 增强有效特征. 网络 3 在网络 2 基础上, 增加了联合损失模块. 联合损失模块可以在增大类间距离的基础上, 通过迫使同类向特征中心靠近, 将不同类进一步分离. 可以看出, 网络 3 在网络 2 基础上进一步提高. 在网络 4 中, 同时使用了所有模块, 其整体精度达到了 93.6%, 平均精度达到了 91.2%, 相较于网络 0, 整体精度提高了 0.7%.

2.5 鲁棒性实验

本文在数据集 ModelNet40 上进行了采样密度的鲁棒性实验和高斯噪声鲁棒性实验.

1) 采样密度鲁棒性实验. 实验中, 分别将点云的采样数设置为 1 024、512、256、128, 然后将网络 PointNet、PointNet++、PCNN、DGCNN 和 RSCNN 作为对比, 对比结果如图 7 所示. 由图 7 可以看出, 随着采样点数的降低, RFNet 有着更为稳健的结果. RSCNN 在使用投票机制后, 有着很小的差距, 但随着采样数降为 128, 差距进一步拉开, 而 RFNet 的准确率依旧保持在 91% 以上. 表明本文网络具有较强的采样密度鲁棒性.

2) 高斯噪声鲁棒性实验. 本文在原始数据集上加入了高斯噪声, 噪声均值设置为 0.02, 方差设置为 0.01, 可视化模型如图 8 所示, 实验结果见表 7, 其中 PointNet++ 和 RSCNN 的平均精度数据缺失, 本文不做对比. 由其余数据的对比结果可以看出, 在加入高斯噪声后, 各种分类网络的结果都有明显下降, 但 RFNet 影响较低. 这是由于在构建局部结构时, RFNet 可以通过对不同近邻点分配不同权重来降低干扰点的影响, 也使得 RFNet 在面对高斯噪声的影响时, 更具有鲁棒性.

2.6 特征关系的消融实验

不同特征关系组合的选取对局部结构构造有很大影响, 过少的特征关系无法准确描述近邻点对中心点的影响; 而过多的特征关系则会给网络带来冗余信息, 导致网络过拟合. 本节对不同组合进行探讨. 除式 (5) 提到的特征外, 还加入了新的特征关系:

表 5 在 ScanObjectNN 数据集上的实验结果 (%)
Table 5 Experimental results on ScanObjectNN (%)

方法	mAcc	OA	Bag	Bin	Box	Sofa	Desk	Shelf	Table	Door	Bed	Cabinet	Chair	Display	Pillow	Sink	Toilet
PointNet	63.4	68.2	36.1	69.8	10.5	76.7	50.0	72.6	67.8	**93.8**	61.8	62.6	89.0	73.0	67.6	64.2	55.3
PointNet++	75.4	77.9	49.4	**84.4**	31.6	90.5	74.0	72.6	72.6	85.2	75.5	77.4	91.3	79.4	81.0	**80.8**	**85.9**
DGCNN	73.6	78.1	49.4	82.4	33.1	91.4	63.3	79.3	77.4	89.0	64.5	**83.9**	91.8	77.0	77.1	75.0	69.4
PointCNN	75.1	78.5	**57.8**	82.9	33.1	91.9	65.3	**84.2**	67.4	84.8	80.0	83.6	**92.6**	78.4	80.0	72.5	71.8
SpiderCNN	69.8	73.7	43.4	75.9	12.8	90.5	65.3	78.0	65.9	91.4	69.1	74.2	89.0	74.5	80.0	65.8	70.6
RFNet	**76.3**	**79.6**	54.0	81.9	**34.1**	**92.2**	**74.1**	79.9	72.1	91.5	**81.3**	83.5	91.3	**80.2**	**82.6**	78.9	67.0

表 6　不同模块的消融实验 (%)

Table 6　Ablation experiment of different modules (%)

方法	RFConv	WAP	JL	mAcc	OA
网络 0	×	×	×	90.2	92.9
网络 1	√	×	×	91.0	93.3
网络 2	√	√	×	91.1	93.5
网络 3	√	×	√	91.1	93.4
网络 4	**√**	**√**	**√**	**91.2**	**93.6**

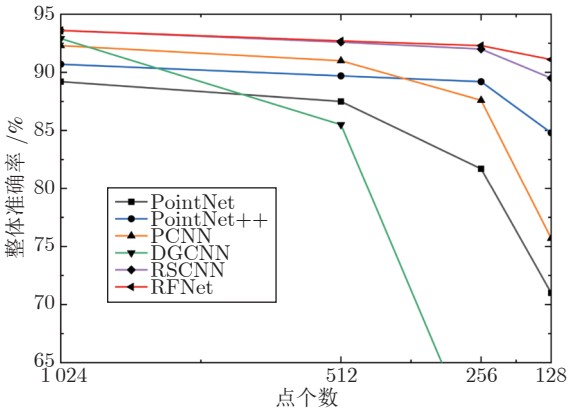


图 7　采样密度实验结果对比

Fig. 7　Comparison of experiment results with different sampling densities

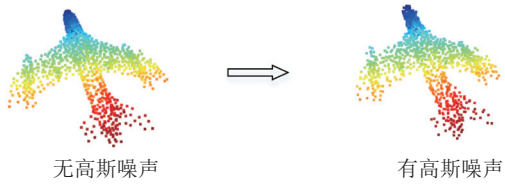


图 8　添加高斯噪声前/后点云样本的对照图

Fig. 8　Comparison of one sample with and without Gaussian noise

表 7　高斯噪声鲁棒性实验 (%)

Table 7　Robustness experiment of Gaussian noise (%)

方法	无高斯噪声		有高斯噪声	
	mAcc	OA	mAcc	OA
PointNet	86.0	89.2	83.4	87.6
PointNet++	—	90.7	—	89.6
DGCNN	90.2	92.9	89.7	92.6
RSCNN	—	92.9	—	92.3
RFNet	**91.2**	**93.6**	**91.0**	**93.5**

$$dis_{ij} = \frac{X_i - X_{ij}}{||X_i - X_{ij}||} \tag{13}$$

式中, $dis_{ij} \in \mathbf{R}^d$ 为中心点与近邻点之间的方向信息. 最终实验结果见表 8, 表中“×”表示不使用对应特征关系, “√”表示使用对应特征关系.

表 8　不同特征关系的消融实验 (%)

Table 8　Ablation studies about different relationships between features (%)

方法	e_{ij}	dis_{ij}	l_{ij}	s_{ij}	OA
网络 1	√	×	×	×	93.3
网络 2	√	×	√	×	93.5
网络 3	√	×	√	√	93.6
网络 4	√	√	×	×	93.1
网络 5	×	√	√	√	93.5
网络 6	√	√	√	√	93.2

由表 8 可知, 网络 1 在仅使用向量信息 e_{ij} 时, 仍能取得 93.3% 的整体分类精度, 可见 e_{ij} 中含有丰富的特征信息. 网络 2 和网络 3 在网络 1 的基础上增加了距离信息 l_{ij} 和余弦相似度信息 s_{ij} 后, 整体分类精度也都有进一步的提高. 需要注意的是, 虽然 e_{ij} 中内在包含了 l_{ij}, 但从分类结果可以看出, 一个准确的距离关系信息对局部特征的提取是很有必要的. 网络 4 在网络 1 的基础上增加了方向信息 dis_{ij}, 最终的整体分类精度却降低了 0.2%, 这是因为二者在相同的维度 d 下存在着冗余信息, 而且冗余信息会随着维度 d 的增大而不断增加. 这点通过网络 3 与网络 6、网络 5 与网络 6 的对比也可以看出. 虽然二者有着相同的维度, 但 e_{ij} 中却内在包含了 dis_{ij}, 由网络 3 与网络 5 的对比也能看出, 使用 e_{ij} 能够取得更好效果.

2.7　网络复杂度实验

本节通过对比网络的参数量和浮点运算数来说明网络的复杂度. 网络的复杂度对比见表 9.

表 9　网络的复杂度对比

Table 9　Comparison of network complexity

方法	参数量 (M)	浮点运算数 (G)	OA (%)
PointNet	3.47	0.45	89.2
PointNet++	1.74	4.09	91.9
DGCNN	1.81	2.43	92.9
PCT	2.88	2.32	93.2
KPConv	14.30	—	92.9
RFNet	**2.36**	**2.95**	**93.6**

对比 DGCNN 可以看出, RFNet 在参数量和浮点运算数上都有一定提高; 同时, 在整体精度上也有很大程度的提高. 这说明, 网络中增添的模块对网络精度的提高具有正向意义. 对比其他网络也可以看出, 网络在参数量和浮点运算数上做了折中

方案, 但整体精度却最高.

3 结束语

本文提出一种用于点云分类的卷积神经网络. 首先, 该网络通过卷积模块提取近邻点与中心点之间的局部特征; 然后, 通过使用加权平均池化, 从冗余的高维特征中聚合全局特征; 最后, 使用联合损失函数增强整体网络的分类能力. 从合成数据集和真实世界数据集上的实验结果可以看出, 该网络比现有网络更具优势. 同时, 本文还进行了消融实验、采样密度的鲁棒性实验、高斯噪声鲁棒性实验和网络复杂度实验, 实验结果验证了本文网络模块的有效性和整体网络的稳定性.

References

1 Guo Y, Bennamoun M, Sohel F, Min L, Wan J. 3D object recognition in cluttered scenes with local surface features: A survey. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2014, **36**(11): 2270−2287

2 Tian Yong-Lin, Shen Yu, Li Qiang, Wang Fei-Yue. Parallel point clouds: Point clouds generation and 3D model evolution via virtual-real interaction. *Acta Automatica Sinica*, 2020, **46**(12): 2572−2582
(田永林, 沈宇, 李强, 王飞跃. 平行点云: 虚实互动的点云生成与三维模型进化方法. 自动化学报, 2020, **46**(12): 2572−2582)

3 Qi C R, Su H, Mo K, Guibas L J. PointNet: Deep learning on point sets for 3D classification and segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE, 2017. 77−85

4 Guo Y, Wang H, Hu Q, Liu H, Bennamoun M. Deep learning for 3D point clouds: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020, **43**(12): 4438−4364

5 Su H, Maji S, Kalogerakis E, Learned-Miller E. Multi-view convolutional neural networks for 3D shape recognition. In: Proceedings of the IEEE International Conference on Computer Vision. Santiago, Chile: IEEE, 2015. 945−953

6 Yang Z, Wang L. Learning relationships for multi-view 3D object recognition. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. Seoul, South Korea: IEEE, 2019. 7504−7513

7 Xu Si-Yu, Zhu Ji-Hua, Tian Zhi-Qiang, Li Yao-Chen, Pang Shan-Min. Stepwise refinement approach for registration of multi-view point sets. *Acta Automatica Sinica*, 2019, **45**(8): 1486−1494
(徐思雨, 祝继华, 田智强, 李垚辰, 庞善民. 逐步求精的多视角点云配准方法. 自动化学报, 2019, **45**(8): 1486−1494)

8 Maturana D, Scherer S. VoxNet: A 3D convolutional neural network for real-time object recognition. In: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems. Hamburg, Germany: IEEE, 2015. 922−928

9 Riegler G, Osman U A, Geiger A. OctNet: Learning deep 3D representations at high resolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE, 2017. 6620−6629

10 Qi C R, Yi L, Su H, Guibas L J. PointNet++: Deep hierarchical feature learning on point sets in a metric space. In: Proceedings of the Advances in Neural Information Processing Systems. Long Beach, USA: 2017. 5099−5108

11 Wang C, Samari B, Siddiqi K. Local spectral graph convolution for point set feature learning. In: Proceedings of the European Conference on Computer Vision. Munich, Germany: 2018. 56−71

12 Wang Y, Sun Y, Liu Z, Sarma S E, Bronstein M M, Solomon J M. Dynamic graph CNN for learning on point clouds. *ACM Transactions on Graphics*, 2019, **38**(5): 1−12

13 Liu Y, Fan B, Xiang S, Pan C. Relation-shape convolutional neural network for point cloud analysis. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE, 2019. 8895−8904

14 Wang L, Huang Y, Hou Y, Zhang S, Shan J. Graph attention convolution for point cloud semantic segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE, 2019. 10288−10297

15 Chen C, Fragonara L Z, Tsourdos A. GAPointNet: Graph attention based point neural network for exploiting local feature of point cloud. *Neurocomputing*, 2021, **438**: 122−132

16 Guo M H, Cai J X, Liu Z N, Mu T J, Martin R R, Hu S M. PCT: Point cloud transformer. *Computational Visual Media*, 2021, **7**(2): 187−299

17 Wen Y, Zhang K, Li Z, Qiao Y. A discriminative feature learning approach for deep face recognition. In: Proceedings of the European Conference on Computer Vision. Amsterdam, Netherland: 2016. 499−515

18 Wu Z, Song S, Khosla A, Yu F, Zhang L, Tang X, et al. 3D ShapeNets: A deep representation for volumetric shapes. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Boston, USA: IEEE, 2015. 1912−1920

19 Yi L, Kim V G, Ceylan D, Shen I C, Yan M, Su H, et al. A scalable active framework for region annotation in 3D shape collections. *ACM Transactions on Graphics*, 2016, **35**(6): 1−12

20 Uy M A, Pham Q H, Hua B S, Nguyen T, Yeung S K. Revisiting point cloud classification: A new benchmark data-set and classification model on real-world data. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. Seoul, South Korea: IEEE, 2019. 1588−1597

21 Xu Y, Fan T, Xu M, Long Z, Yu Q. SpiderCNN: Deep learning on point sets with parameterized convolutional filters. In: Proceedings of the European Conference on Computer Vision. Munich, Germany: 2018. 90−105

22 Simonovsky M, Komodakis N. Dynamic edge-conditioned filters in convolutional neural networks on graphs. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE, 2017. 3693−3702

23 Tian Yu-Jie, Guan You-Qing, Gong Rui. A robust deep neural network for multi-feature point cloud classification and segmentation. *Computer Engineering*, 2021, **47**(11): 234−240
(田钰杰, 管有庆, 龚锐. 一种鲁棒的多特征点云分类分割深度神经网络. 计算机工程, 2021, **47**(11): 234−240)

24 Li Y, Bu R, Sun M, Chen B. PointCNN: Convolution on x-transformed points. In: Proceedings of the Advances in Neural Information Processing System. Montreal, Canada: 2018. 828−838

25 Liu X, Han Z, Liu Y, Zwicker M. Point2sequence: Learning the shape representation of 3D point clouds with an attention-based sequence to sequence network. In: Proceedings of the AAAI Conference on Artificial Intelligence. Honolulu, USA: AAAI, 2019. 8778−8785

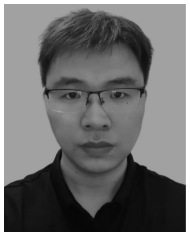
26 Engel N, Belagiannis V, Dietmayer K. Point transformer. *IEEE Access*, 2021, **9**: 134826−134840

27 Xu Xiang, Shuai Hui, Liu Qing-Shan. Octant convolutional neural network for 3D point cloud analysis. *Acta Automatica Sinica*, 2021, **47**(12): 2791−2800
(许翔, 帅惠, 刘青山. 基于卦限卷积神经网络的 3D 点云分析. 自动化学报, 2021, **47**(12): 2791−2800)

28 Qiu S, Anwar S, Barnes N. Dense-resolution network for point cloud classification and segmentation. In: Proceedings of the IEEE Winter Conference on Applications of Computer Vision. Virtual Event: IEEE, 2021. 3812−3821

29 Zhou H, Feng Y, Fang M, Wei M, Qin J, Lu T. Adaptive graph convolution for point cloud analysis. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE, 2021. 4965−4974

30 Thomas H, Qi C R, Deschaud J E, Marcotegui B, Goulette F, Guibas L. KPConv: Flexible and deformable convolution for point clouds. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. Seoul, South Korea: IEEE, 2020. 6420−6429



单铉洋 安徽大学电气工程与自动化学院硕士研究生. 主要研究方向为模式识别, 计算机视觉和深度学习.
E-mail: shanxy128@163.com
(**SHAN Xuan-Yang** Master student at the School of Electrical Engineering and Automation, Anhui University. His research interest covers pattern recognition, computer vision, and deep learning.)

孙战里 安徽大学人工智能学院教授. 主要研究方向为模式识别, 机器学习和图像信号处理. 本文通信作者.
E-mail: zhlsun2006@126.com
(**SUN Zhan-Li** Professor at the School of Artificial Intelligence, Anhui University. His research interest covers pattern recognition, machine learning, and image & signal processing. Corresponding author of this paper.)

曾志刚 华中科技大学人工智能与自动化学院教授. 主要研究方向为神经网络, 智能计算和模式识别.
E-mail: zgzeng@hust.edu.cn
(**ZENG Zhi-Gang** Professor at the School of Artificial Intelligence and Automation, Huazhong University of Science and Technology. His research interest covers neural networks, computational intelligence, and pattern recognition.)