

# 融合细粒度特征编码的点云分类分割网络

陶志勇<sup>1</sup>, 豆淼森<sup>1</sup>, 李衡<sup>1</sup>, 林森<sup>2</sup>

(1. 辽宁工程技术大学电子与信息工程学院, 葫芦岛 125105; 2. 沈阳理工大学自动化与电气工程学院, 沈阳 110159)

**摘要:** 有效获取点云特征是分析和处理三维点云场景的关键。针对目前深度学习方法特征信息提取不充分, 难以捕捉深层次语义信息的问题, 提出了一种融合细粒度特征编码的网络来提高点云分类与分割任务的准确率。首先, 特征提取模块包含2个子模块: 一个是扩张图卷积模块, 相比图卷积能够提取更丰富的几何信息; 另一个是细粒度特征编码模块, 能够获取局部区域的细节特征。其次, 通过可学习参数将二者动态融合, 有效地学习每个点的上下文信息。最后, 将提取的所有特征相加, 通过通道亲和注意力模块来强调不同通道, 协助特征图来避免可能的冗余。在ModelNet40及ScanObjectNN数据集上进行点云分类实验, 总体分类精度分别为93.3%和80.0%。在ShapeNet Part数据集上进行点云部件分割实验, 平均交并比为85.6%。实验结果表明, 与目前主流方法相比, 该网络具有较优的性能。

**关键词:** 深度学习; 局部特征提取; 点云分类; 部件分割; 细粒度特征

**中图分类号:** TP391 **文献标志码:** A

## Fusion Fine-Grained Feature Encoding for Point Cloud Classification and Segmentation

TAO Zhiyong<sup>1</sup>, DOU Miaosen<sup>1</sup>, LI Heng<sup>1</sup>, LIN Sen<sup>2</sup>

(1. School of Electronic and Information Engineering, Liaoning Technical University, Huludao 125105, China; 2. School of Automation and Electrical Engineering, Shenyang Ligong University, Shenyang 110159, China)

**Abstract:** Effective acquisition of point cloud features is the key to analyzing and processing 3D point cloud scenes. To address the problem that current deep learning methods have inadequate feature information extraction and difficulty in capturing deep semantic information, a fusion fine-grained feature encoding network is proposed to improve the accuracy of point cloud classification and segmentation tasks. First, the feature extraction module contains two sub-modules, one is the dilation graph convolution module, which can extract richer geometric information than graph convolution; and the other is the fine-grained feature encoding module, which can capture detailed features of local regions. Second, the two modules are dynamically fused by learnable parameters to efficiently learn the contextual information of each point. Finally, all the extracted features are summed and pass the channel-wise affinity attention module, assisting the feature map to avoid redundancy by emphasizing its distinct channels. Point cloud classification experiment is performed on the ModelNet40 and ScanObjectNN datasets, and the overall accuracy is 93.3% and 80.0%, respectively. The mean intersection over union (mIoU) is 85.6% for part

segmentation experiments on the ShapeNet Part dataset. Experimental results show that the proposed method performs better than the current mainstream methods.

**Key words:** deep learning; local feature extraction; point cloud classification; part segmentation; fine-grained feature

## 引 言

近些年,随着三维点云采集技术的快速发展,点云数据的获取变得更加简单方便。点云不仅可以表示最基本的三维几何信息,还可以表示颜色、灰度值等信息,是三维对象最简单且有效的形状表示,能够准确直观地显示真实世界,在三维重建<sup>[1]</sup>及语义分割<sup>[2]</sup>等方面具有较好的应用前景。随着三维传感设备的不断发展,三维点云处理技术在无人驾驶<sup>[3]</sup>、城市规划、医学影像和机器人技术<sup>[4]</sup>等众多领域得到了广泛的应用。点云分类与分割任务作为点云数据处理的一个分支,具有重要的研究价值。

卷积神经网络(Convolutional neural network, CNN)在二维图像处理领域取得了巨大的成功<sup>[5]</sup>,但是点云数据并不像二维图像那样规则排列,而是非结构化和无序的,并离散分布在采样对象表面,因此传统的卷积神经网络不适合直接应用到点云上。在图像处理中,二维图像在处理速度与精度均高于三维图像,因此一些研究员尝试将三维数据转换为二维图像数据,以便能够对其应用传统的卷积神经网络<sup>[6]</sup>。

基于多视图的方法是将点云从不同视角投影至二维平面,收集三维对象的一系列二维图像,进而应用卷积神经网络来提取每幅图像的关键特征,最后将不同视角获取到的特征进行聚合,得到点云的全局特征。Su等<sup>[7]</sup>提出多视图卷积神经网络(Multi-view CNN, MVCNN),是该领域开创性的工作。虽然基于多视图的方法能够使用二维卷积神经网络实现三维点云的识别和分割,但该方法未能考虑几何结构信息,并且在投影时会不可避免地带来信息丢失。

基于体素的方法是将无序的点云数据转换为有规律的三维网格形式,再通过使用三维卷积神经网络来提取特征。三维卷积神经网络VoxNet的提出被认为是最著名的基于体素的方法<sup>[8]</sup>。但基于体素的方法会引入过多的内存成本,增加繁重的计算量,丢失有价值的信息,并且很难捕获细粒度的几何细节。同时,该方法分辨率较差,难以在大规模点云中使用。

基于点的方法是直接对原始点云数据进行处理。Qi等<sup>[9]</sup>设计了基于点方法的深度学习网络PointNet来直接处理原始的点云数据,没有使用传统的卷积算子,而是在每个点上独立应用多层感知器(Multi-layer perceptron, MLP)来提取特征,并使用对称函数聚合全局特征,该方法是直接处理稀疏点云的先驱工作之一,解决了点云的置换不变性问题。在点云处理中,点之间的连接关系比独立点提供的信息更加可靠,然而PointNet只考虑了全局信息并不能很好地处理点周围的邻域信息,导致点与点之间缺乏良好的交互。为了弥补PointNet的不足,Qi等<sup>[10]</sup>对PointNet进行了改进并提出了PointNet++,采用最远点采样(Farthest point sampling, FPS)的方法从点云中找到一些重要的点作为中心点,使用分层特征提取的思想,在中心点的一定范围内使用K近邻(K-nearest neighbor, KNN)算法选取一些邻近点并通过逐级降采样划分为不同层次的局部区域,使用PointNet进行特征提取,下一层继续相同的操作,随着层数的加深中心点包含的信息就越丰富。虽然该方法提高了准确率,但也没有考虑点之间的关系。

基于图的方法利用邻近点和中心点之间的连接关系来构建局部图,然后对其应用卷积和池化操作。图卷积在二维图像处理领域获得了较好的效果,本质上比CNN模型更方便。Wang等<sup>[11]</sup>提出一种动态图卷积神经网络(Dynamic graph CNN, DGCNN),使用K近邻算法构建局部邻域图来捕获中心点

及其邻近点之间的局部几何关系,提取边缘特征,获得了较好的分类分割结果。Zhang等<sup>[12]</sup>受DGCNN的启发提出了LDGCNN(Linked DGCNN),使用跳跃连接的方式提取不同层次的局部特征,使得网络性能得到改善。然而邻近点 $k$ 值的不同会使网络的性能受到影响,如果 $k$ 值较小则不能覆盖足够的局部区域, $k$ 值较大则会提取较多的冗余信息增加计算量。Engelmann等<sup>[13]</sup>提出了一种基于扩张点卷积(Dilated point convolution, DPC)的网络,引入扩张率 $d$ 来增大感受野。

针对点云处理中局部特征信息提取不充分的问题,本文提出一种融合细粒度特征编码的点云分类分割网络。首先为输入的点构建一个局部扩张图增加感受野的大小,较大的感受野能够帮助提取更丰富的几何信息。其次通过可学习参数将扩张图卷积与细粒度特征编码动态融合,运用该模块来进行点云的局部特征提取,更好地捕获局部细节信息。最后引入了通道亲和注意力模块(Channel-wise affinity attention, CAA)细化点云的特征,减少信息冗余<sup>[14]</sup>。在多个数据集上的实验结果表明,本文方法有效增强了点云局部特征信息的提取能力,在点云分类与分割任务上表现出较好的性能。

## 1 点云分类分割网络

### 1.1 网络模型设计

本文提出的融合细粒度特征编码的点云分类分割网络结构如图1所示。网络的输入为 $N \times 3$ 的原始点云数据, $N$ 为输入点云的点数, $3$ 为特征的维度,通过融合细粒度特征编码的扩张图卷积模块(Fusion fine-grained feature encoding with dilated graph convolution, FFDG)来提取点云局部特征信息,在每个FFDG模块中构建局部扩张图,从而提取不同维度的局部信息。将提取到不同维度的特征进行拼接融合,通过通道亲和注意力模块细化点云的特征,减少信息冗余。分类网络如图1(a)所示,首先将FFDG模块提取到的所有局部特征信息相加,使用MLP提取为一维全局特征,其次通过CAA模块减少信息冗余,然后使用最大池化和平均池化(Max-avg pool)后得到高维特征向量,最后通过全连接层(Fully connected layer, FCL)输出 $C$ 个类别的分类得分。分割网络如图1(b)所示,包含3个FFDG模块,最终输出 $p$ 个语义标签的分割得分。

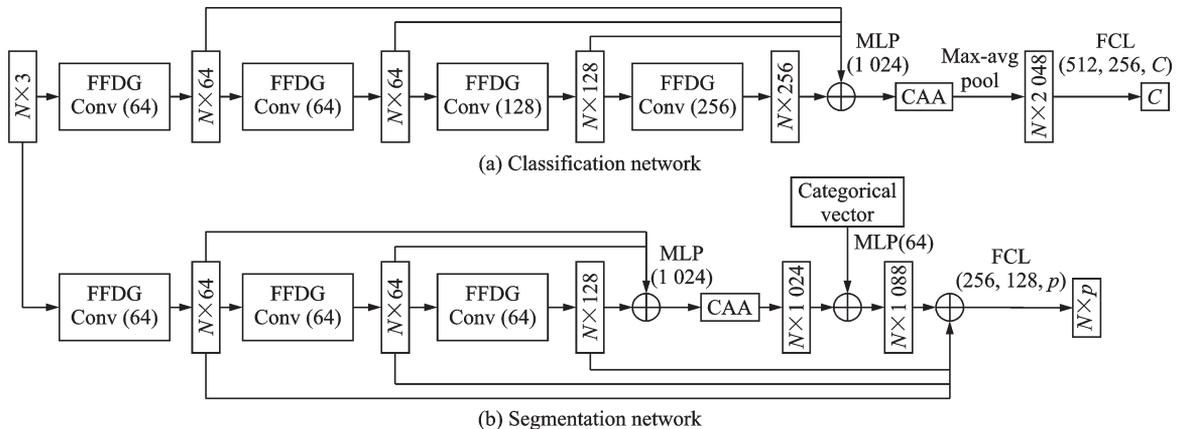


图1 本文提出的点云分类分割网络框架

Fig.1 Framework of the proposed point cloud classification and segmentation networks

### 1.2 局部扩张图构建

点云通常为 $N$ 个点的集合,维度为 $D$ ,将输入的点云表示为: $X = \{x_1, x_2, \dots, x_i, \dots, x_N\} \in \mathbf{R}^3$ 。其中,每个点 $x_i = \{u_i, w_i, z_i\}$ 表示坐标信息,并且在某些情况下可以与法线和颜色等向量信息结合起来改变

$D$ 的大小。

在三维点云中构建局部扩张图是一种有效的几何表示,引入扩张率  $d$ ,以允许中心点与邻近点相连接。局部扩张图构建的原理图如图 2 所示,当  $d=1$  为普通的图卷积,而当  $d=2$  为扩张率为 2 的扩张图卷积。对于输入的点云,使用  $K$  近邻算法搜索中心点  $x_i$  附近的  $k$  个邻近点。在特征空间中点与点之间的欧氏距离能够衡量点密度分布的标准,首先计算邻近点与中心点  $x_i$  之间的欧氏距离,将其按升序排列。计算公式为

$$e_{ij}^d = (x_{ijd} - x_i) \| x_i \quad (1)$$

式中:  $e_{ij}^d$  为第  $j$  个邻近点到中心点  $x_i$  之间的距离;  $\|$  为连接操作;  $x_{ijd}$  为扩张率为  $d$  范围内的第  $j$  个邻近点;  $k$  表示所选择扩张邻近点的个数。根据不同的扩张率  $d$  选出  $k$  个邻近点  $x_{ijd} (j = 1, 2, \dots, k)$ , 最终输出的点集为  $m_i = \{x_{id}, x_{i2d}, \dots, x_{ikd}\}$ 。

每个中心点  $x_i$  与其扩张邻近点的集合  $m_i$  构成局部扩张图  $G_d = (V_d, E_d)$ , 其中  $V_d = \{x_i, m_i\}$  和  $E_d = \{e_{i1}^d, e_{i2}^d, \dots, e_{ik}^d\}$  分别表示扩张图结构的顶点和边。

### 1.3 融合细粒度特征的扩张图卷积模块

使用最大池化函数提取突出特征来总结整体轮廓的特征,能够有效地编码点云的几何信息。当处理较大的点集时,这一操作是有效的,但是当点的数量较少时,则需要一种合理的方法来考虑所有的邻近点。为了进行更加全面的特征表示,通过共享的本地全连接层(Local fully-connected, LFC)来学习权重。LFC对局部区域进行  $1 \times k$  卷积操作,  $k$  的大小与邻近点个数相同,学习区域内  $k$  个邻近点的权重,最后将特征聚合作为详细的局部上下文信息。LFC通过可学习权重聚合本地信息和映射的几何特征,通过这种方式能够保留局部区域的细节特征。细粒度特征编码输出为

$$F_2 = \text{EdgeLFC}(\bullet) = \tau(\text{BN}(c_{1 \times k}(\bullet))) \quad (2)$$

式中  $\tau$  为激活函数。

本文设计的 FFDG 模块主要由扩张图卷积模块和细粒度特征编码模块组成,如图 3 所示。为了更好地利用提取到的特征信息,通过注意力向量  $\alpha$  将二者自适应结合起来。最终输出的特征为

$$F = \alpha \odot F_1 \oplus (1 - \alpha) \odot F_2 \quad (3)$$

式中:  $\alpha \in \mathbb{R}^{1 \times c}$ ,  $c$  为通道数;  $\odot$  为 Hadamard 乘积;  $\oplus$  为元素相加;  $F_1$  为扩张图卷积提取到的特征;  $F_2$  为细粒度特征编码。

### 1.4 通道亲和注意力模块

目前点云中大多数注意力结构都在点空间上操作,效果并不明显。因此本文引入一种针对高级点云特征图的 CAA 模块来避免通道冗余。其主要包括紧凑的通道比较器模块(Compact channel-wise comparator, CCC)、通道亲和估计模块(Channel affinity estimator, CAE)。

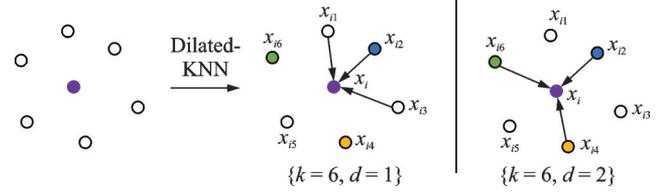


图 2 局部扩张图构建

Fig.2 Construction of the local dilated graph

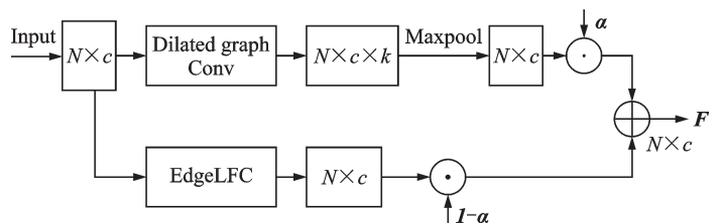


图 3 融合细粒度特征编码的扩张图卷积模块

Fig.3 Fusion fine-grained feature encoding with dilated graph Conv

将  $N \times d$  维的特征通过通道比较器模块,对通道向量  $c_i$  进行 MLP 操作来简化每个通道,并将通道数缩小,以此来减少尺寸,保留每个通道的信息。其中  $c_i \in \mathbf{R}^N$ ,  $F_{N \times d} = [c_1, c_2, \dots, c_d]$ 。

通道比较器模块:将输入的  $N \times d$  维特征进行转置,得到  $d \times N$  维特征,随后通过 MLP 操作计算查询矩阵与关键矩阵。

查询矩阵为

$$Q_{N' \times d} = [q_1, q_2, \dots, q_d] \quad (4)$$

关键矩阵为

$$K_{N' \times d} = [k_1, k_2, \dots, k_d] \quad (5)$$

式中:  $q_i = \text{MLP}_q(c_i)$ ,  $q_i \in \mathbf{R}^{N'}$ ;  $k_i = \text{MLP}_k(c_i)$ ,  $k_i \in \mathbf{R}^{N'}$ 。

将查询矩阵进行转置操作,然后与关键矩阵进行乘积得到通道相似矩阵  $S$ ,即

$$S_{d \times d} = Q^T \cdot K \quad (6)$$

沿着通道相似矩阵  $S$  的列选出相似度最大的通道,首先将其扩展为与矩阵  $S$  阶数大小相同的扩展矩阵,其次将扩展矩阵减去原始矩阵,加入 softmax 函数进行归一化处理,通过这种方式可以对不同的通道施加更高的权重,从而减少冗余信息。通道亲和矩阵的特征图为

$$A_{d \times d} = \text{softmax}(\text{expand}(\max_{d \rightarrow 1}(S)) - S) \quad (7)$$

对输入的  $N \times d$  维的特征进行 MLP 操作得到矩阵  $V$ ,将其与通道亲和矩阵  $A_{d \times d}$  相乘可以实现细化过程。最后使用一个可学习权值  $\beta$  来简化训练,  $F_{\text{out}}$  为 FFDG 模块提取到的局部信息之和,最终 CAA 模块输出为

$$F' = F_{\text{out}} \oplus \beta \cdot V \cdot A \quad (8)$$

## 2 实验与分析

### 2.1 实验环境与配置

本文实验环境基于 Linux Ubuntu 操作系统,使用 Python3.7 和 Pytorch1.7 深度学习框架,网络的模型训练和测试都在 GeForce RTX 3090 GPU 上进行。分类网络的输入为 1 024 个均匀采样点,邻近点  $k$  设置为 20,训练周期为 250,采用动量为 0.9 的 SGD 优化器,初始学习率为 0.1,使用余弦退火将学习率衰减至 0.001。分割网络的输入为 2 048 个均匀采样点,邻近点  $k$  设置为 20,训练周期为 250,使用 Adam 优化器,初始学习速率为 0.003。所有训练模型的训练批次为 32,测试批次为 16。

### 2.2 数据集描述

为验证本文所提出点云分类分割算法的性能,分别在人工合成 ModelNet40 数据集<sup>[15]</sup>、真实世界对象组成的 ScanObjectNN 数据集<sup>[16]</sup>和 ShapeNet Part 数据集<sup>[17]</sup>进行实验。ModelNet40 数据集中包含 40 个类别,12 311 个 CAD 模型,其中 9 843 个训练样本,2 468 个测试样本,对每个样本进行均匀采样。在训练过程,通过点的移动、缩放和抖动来增强数据,同时在测试期间没有执行预处理或后处理。

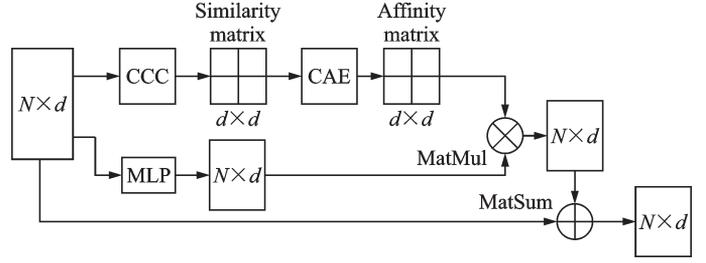


图4 通道亲和注意力模块

Fig.4 Channel-wise affinity attention module

ScanObjectNN是一个用于点云分类的真实世界数据集,包含有15 000个样本,并将对象分为15个类别,其中2 902个唯一的对象实例,但由于样本背景复杂、缺失部分和不同的形变,使得分类更具有挑战性。ShapeNet Part数据集包含16个类别,16 881个点云模型,其中14 006个模型为训练样本,2 875个模型为测试样本,将样本分为2~6个不同的部分进行手工标注处理,共50个部件标签,在训练时从每个形状中采样2 048个点进行部件分割实验。

### 2.3 实验结果及分析

#### 2.3.1 ModelNet40分类实验

为验证本文方法的有效性,在ModelNet40数据集上进行分类实验。采用总体分类精度(Overall accuracy, OA)和平均分类精度(Mean accuracy, mAcc)来评估本文方法,如式(11, 12)所示。

$$OA = \frac{V_T}{V_T + V_F} \quad (11)$$

$$mAcc = \frac{1}{C} \sum_{i=1}^C \frac{V_{Ti}}{V_{Ti} + V_{Fi}} \quad (12)$$

式中:  $V_T$  为正确预测个数;  $V_F$  为错误预测个数;  $V_T + V_F$  为测试样本总数;  $C = 40$  为数据集中的类别数;  $V_{Ti}$  和  $V_{Fi}$  为第  $i$  个样本中预测正确和错误的概率。

选用基于多视图、体素和原始点云3种类型的方法进行对比,结果如表1所示。从表1中实验对比结果可知本文方法在ModelNet40数据集上总体分类精度达到了93.3%,平均分类精度达到了90.7%,均优于表中其他主流方法。在总体分类精度上与基于体素方法的VoxNet和基于多视图方法的MVCNN相比分别提高了7.4%和3.2%。同经典算法PointNet、PointNet++和DGCNN相比,分别提高了4.1%、2.6%和0.4%。与多点输入SpiderCNN和SO-Net相比仍具有优势。与同样使用扩张卷积的DGANet网络相比,总体分类精度与平均分类精度分别提高了1.0%和1.3%,证明了本文设计的FFDG模块及通道亲和注意力模块能够有效增强模型表达能力,提高模型的分类准确率。

#### 2.3.2 ScanObjectNN分类实验

ModelNet40数据集上采集的物体没有缺损且没有任何背景噪声,属于理想状态,因此在真实世界中3D扫描获得的对象可能与之不同。为验证本文方法对真实世界中点云数据的表现性能,在ScanObjectNN数据集上进行实验,评价指标与ModelNet40数据集相同,但类别数  $C = 15$ , 结果如表2所示。同表2中其他对比方法一致,网络输

表1 在ModelNet40数据集上的分类结果

Table 1 Classification results on ModelNet40 dataset

方法	输入	点数	mAcc/%	OA/%
VoxNet <sup>[8]</sup>	体素	—	83.0	85.9
Subvolume <sup>[18]</sup>	体素	—	86.0	89.2
MVCNN <sup>[7]</sup>	多视图	—	—	90.1
PointNet <sup>[9]</sup>	坐标	1 000	86.0	89.2
PointNet++ <sup>[10]</sup>	坐标	1 000	—	90.7
DGCNN <sup>[11]</sup>	坐标	1 000	90.2	92.9
LDGCNN <sup>[12]</sup>	坐标	1 000	90.3	92.9
PointCNN <sup>[19]</sup>	坐标	1 000	88.1	92.2
SO-Net <sup>[20]</sup>	坐标	2 000	—	90.9
P2Sequence <sup>[21]</sup>	坐标	1 000	—	92.6
DGANet <sup>[22]</sup>	坐标	1 000	89.4	92.3
SpiderCNN <sup>[23]</sup>	坐标+法线	5 000	89.2	92.4
Our method	坐标	1 000	90.7	93.3

表2 在ScanObjectNN数据集上的分类结果

Table 2 Classification results on ScanObjectNN dataset

方法	点数	mAcc/%	OA/%
PointNet <sup>[9]</sup>	1 000	63.4	68.2
SpiderCNN <sup>[23]</sup>	1 000	63.4	68.2
PointNet++ <sup>[10]</sup>	1 000	75.4	77.9
DGCNN <sup>[11]</sup>	1 000	73.6	78.1
PointCNN <sup>[19]</sup>	1 000	75.1	78.7
Our method	1 000	77.5	80.0

入均为坐标形式。当输入点数为1 000时,总体精度为80.0%,平均分类精度为77.5%。与其他主流方法相比,在复杂情况下对点云分类任务性能更好,分类准确率较高。

### 2.3.3 ShapeNet Part 分割实验

为验证本文方法对分割任务的性能,在ShapeNet Part数据集上进行点云部件分割实验。采用平均交并比(mean Intersection over union, mIoU)作为评价指标,表示所有类的真实值与预测值两个集合的交集与并集之比,其表达式为

$$mIoU = \frac{1}{p} \sum_{i=1}^p \left( \frac{T \cap P}{T \cup P - T \cap P} \right)_i \quad (13)$$

式中: $T$ 为真实值; $P$ 为预测值; $p$ 为语义标签的个数, $p = 50$ 。

与经典和新颖的深度学习方法进行对比,结果如表3所示。本文方法在ShapeNet Part数据集上的平均交并比达到了85.6%。与经典网络PointNet和PointNet++相比,分别提高了1.9%和0.5%,与DGCNN和DGANet网络相比,均提高了0.4%。同时在帽子、汽车、耳机、吉他、笔记本电脑、杯子、滑板 and 桌子上的分割准确率均优于其他方法。为了更加直观地展示点云分割效果,图5给出了部分零件的分割可视化结果。图6为未改进方法与本文方法的结果,将未准确分割的部分与本文方法对比,可以更直观地反映本文方法的优势。

表3 ShapeNetPar数据集上零件分割结果

Table 3 Part segmentation results on ShapeNet Part dataset

方法	分割精度															mIoU	
	Air	Bag	Cap	Car	Chair	Ear	Gui	Kni	Lam	Lap	Mot	Mug	Pis	Roc	Ska		Tab
PointNet <sup>[9]</sup>	83.4	78.7	82.5	74.9	89.6	73.0	91.5	85.9	80.8	95.3	65.2	93.0	81.2	57.9	72.8	80.6	83.7
PointNet++ <sup>[10]</sup>	82.4	79.0	87.7	77.3	90.8	71.8	91.0	85.9	83.7	95.3	71.6	94.1	81.3	58.7	76.4	82.6	85.1
DGCNN <sup>[11]</sup>	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	85.2
SO-Net <sup>[20]</sup>	82.8	77.8	88.0	77.3	90.6	73.5	90.7	83.9	82.8	94.8	69.1	94.2	90.9	53.1	72.9	83.0	84.9
SpiderCNN <sup>[23]</sup>	83.5	81.0	87.2	77.5	90.7	76.8	91.1	87.3	83.3	95.8	70.2	93.5	82.7	59.7	75.8	82.8	85.3
LDGCNN <sup>[12]</sup>	84.0	83.0	84.9	78.4	90.6	74.4	91.0	88.1	83.4	95.8	67.4	94.9	82.3	59.2	76.0	81.9	85.1
P2Sequence <sup>[21]</sup>	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	85.2
DGANet <sup>[22]</sup>	84.6	85.7	87.8	78.5	91.0	77.3	91.2	87.9	82.4	95.8	67.8	94.2	81.1	59.7	75.7	82.0	85.2
Our method	83.7	83.6	<b>90.1</b>	<b>78.5</b>	90.2	<b>77.7</b>	<b>91.7</b>	87.7	82.1	<b>96.1</b>	71.3	<b>95.4</b>	82.2	61.4	<b>82.0</b>	<b>83.7</b>	<b>85.6</b>

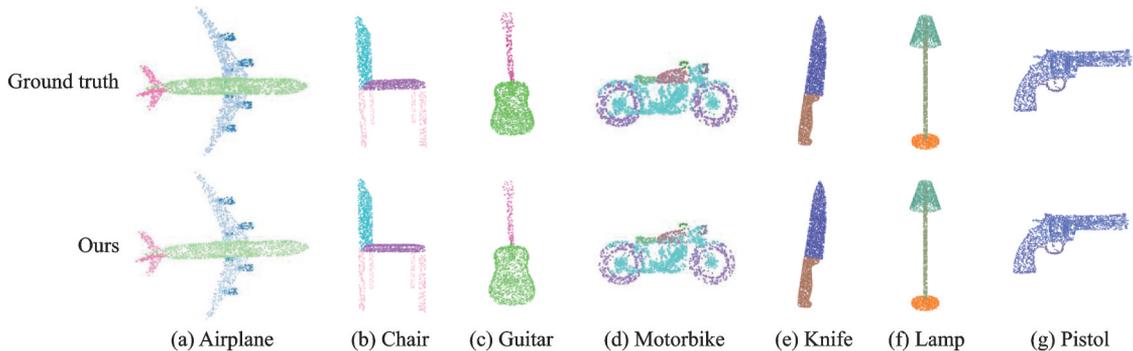


图5 ShapeNet Part数据集上点云分割示例

Fig.5 Examples of the point cloud segmentation results on ShapeNet Part dataset

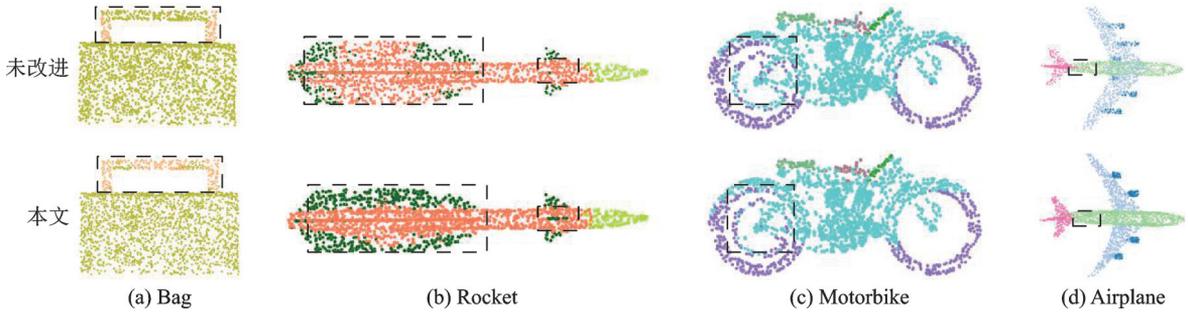


图6 分割效果对比图

Fig.6 Comparison of segmentation results

## 2.4 不同邻近点 $k$ 对网络性能的影响

本节讨论不同邻近点  $k$  对分类模型性能的影响。实验选取了几个典型的  $k$  值在 ModelNet40 数据集上进行分类测试,结果如表 4 所示。不同  $k$  值代表着不同大小的局部区域感受野,当  $k$  值较小时,虽然能够有效地减少计算量,但由于接受域太小,不能有效捕捉中心点周围的细节信息,造成准确率较低。而  $k$  值过大则会包含较多的冗余信息,并增加训练时间,因此合适的  $k$  值对提高网络性能至关重要。通过实验结果可以看出,网络的性能并不会随着  $k$  值的增大而一直增大,而随着  $k$  值的增大训练时间也会随之增加。为了平衡精度和训练时间,取  $k=20$ 。

## 2.5 不同输入特征对网络性能的影响

输入特征的不同会直接影响点与点之间的关系以及局部几何图形的表示,通过中心点与最近的两个邻近点之间的几何关系来扩展输入的几何信息。为了寻找合适的特征组合,在坐标信息基础上加入不同类型的低级几何信息,主要包含 4 种类型的低级几何特征:坐标  $x_i$ 、法线  $\mathbf{n} \in \mathbf{R}^3$ 、边  $\mathbf{e} \in \mathbf{R}^3$  和边的长度  $l \in \mathbf{R}^1$ , 4 种几何特征的 Channels 分别为 3、3、3、1,不同的组合方式会导致输入特征通道数发生改变。本文在 ModelNet40 数据集上对 6 种不同的设置分别进行实验,实验结果如表 5 所示。从表 5 结果可见,方法 1 仅有坐标位置信息时仍能取得 92.9% 准确率,方法 2 加入边的长度信息,能更好地估计点邻域的密度分布,准确率有所提升。方法 3 中加入法线信息,但由于点云采样密度不均匀存在一定的信息冗余,从而可能导致准确率降低。方法 4 是在方法 1 上同时加入法线和边的信息,取得了最优的准确率,可以看出结合法线和边的信息更有利于提取局部特征信息。方法 5 和方法 6 结果可知,特征信息越多可能包含越多的冗余信息,会对分类准确率造成一定的影响。根据上述实验结果,本文选用方法 4 中的特征作为网络的输入。

表 4  $k$  值对网络性能的影响Table 4 Impact of  $k$  on network performance

$k$	mAcc/%	OA/%	Train speed/(epoch $\cdot$ s $^{-1}$ )
5	90.2	92.7	48
10	90.2	93.2	52
15	89.1	92.9	57
20	90.7	93.3	65
25	90.2	92.9	71

表 5 不同输入特征的消融实验

Table 5 Ablation experiment about different input features

方法	输入特征	Channel	OA/%
1	$x_i$	3	92.9
2	$x_i, l_1, l_2$	5	93.1
3	$x_i, \mathbf{n}_i$	6	92.8
4	$x_i, \mathbf{n}_i, l_1, l_2$	8	93.3
5	$x_i, \mathbf{e}_1, \mathbf{e}_2, l_1, l_2$	11	93.2
6	$x_i, \mathbf{n}_i, \mathbf{e}_1, \mathbf{e}_2, l_1, l_2$	14	93.1

## 2.6 消融实验

为了验证 FFDG 模块及 CAA 模块的有效性,在 ModelNet40 数据集上进行消融实验,结果如表 6 所示。从表 6 可以看出,方法 1 为未改进的原始网络。方法 2 将特征提取模块替换成 FFDG 模块,分类准确率提升为 92.9%,能够有效证明本文所设计的特征提取模块能够更好地捕捉丰富的几何信息。由方法 2、方法 3 的结果可知,FFDG 模块与 CAA 模块都能在一定程度上提高网络的性能。因此在方法 4 中同时结合两个模块进一步提高网络的分类性能。

## 3 结束语

为进一步提取点云深层次特征信息,提高网络对点云分类与分割任务的准确率,本文提出了一种融合细粒度特征编码的点云分类分割网络。在 FFDG 模块中,通过可学习参数将扩张图卷积提取到的特征与细粒度特征编码进行动态融合,增强了点云局部特征提取能力。并在网络中引入通道亲和注意模块,减少特征空间信息冗余,从而提升模型性能。在多个基准数据集上的实验证明,本文网络能够实现较高的分类精度和分割准确性,在处理复杂点云数据时仍能保持较好的性能,并通过消融实验,证明了所提出模块的有效性。下一步计划将本文方法推广到点云场景语义分割领域进行研究。

## 参考文献:

- [1] 李红卫,魏泽勇,曹轩铭,等.基于深度相机的飞机中央翼油箱高精度三维重建方法[J].南京航空航天大学学报, 2022, 54(3): 457-465.  
LI Hongwei, WEI Zeyong, CAO Xuanming, et al. 3D reconstruction of aircraft central wing fuel tank based on depth camera [J]. Journal of Nanjing University of Aeronautics & Astronautics, 2022, 54(3): 457-465.
- [2] 李嘉祥,宣士斌,刘丽霞,等.学习几何结构特征的真实点云场景语义分割[J].数据采集与处理, 2023, 38(2): 336-349.  
LI Jiaxiang, XUAN Shibin, LIU Lixia, et al. Semantic segmentation for real point cloud scenes via geometric features[J]. Journal of Data Acquisition and Processing, 2023, 38(2): 336-349.
- [3] LIANG M, YANG B, WANG S, et al. Deep continuous fusion for multi-sensor 3D object detection[C]//Proceedings of the European Conference on Computer Cision (ECCV). Munich, Germany: [s.n.], 2018: 641-656.
- [4] RUSU R B, MARTON Z C, BLODOW N, et al. Towards 3D point cloud based object maps for household environments[J]. Robotics and Autonomous Systems, 2008, 56(11): 927-941.
- [5] RAHMAN M M, TAN Y, XUE J, et al. Recent advances in 3D object detection in the era of deep neural networks: A survey [J]. IEEE Transactions on Image Processing, 2019, 29: 2947-2962.
- [6] CHEN S, ZHENG L, ZHANG Y, et al. VERAM: View-enhanced recurrent attention model for 3D shape classification[J]. IEEE Transactions on Visualization and Computer Graphics, 2018, 25(12): 3244-3257.
- [7] SU H, MAJI S, KALOGERAKIS E, et al. Multi-view convolutional neural networks for 3D shape recognition[C]// Proceedings of the IEEE International Conference on Computer Vision. Boston, MA, USA: IEEE, 2015: 945-953.
- [8] MATURANA D, SCHERER S. VoxNet: A 3D convolutional neural network for real-time object recognition[C]// Proceedings of 2015 IEEE/RSJ International Conference on Intelligent Robosts and Systems (IROS). Hamburg, Germany: IEEE, 2015: 922-928.
- [9] QI C R, SU H, MO K, et al. Pointnet: Deep learning on point sets for 3D classification and segmentation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recongnition. Honolulu, USA: IEEE, 2017: 77-85.
- [10] QI C R, YI L, SU H, et al. PointNet++: Deep hierarchical feature learning on point sets in a metric space[C]//Proceedings

表 6 不同模块的消融实验

Table 6 Ablation experiment about different modules

方法	FFDG 模块	CAA 模块	OA/%
1	×	×	92.7
2	✓	×	92.9
3	×	✓	93.0
4	✓	✓	93.3

- of the 31st International Conference on Neural Information Processing Systems. Red Hook, NY, USA: NIPS, 2017: 5105-5114.
- [11] WANG Y, SUN Y, LIU Z, et al. Dynamic graph CNN for learning on point clouds[J]. ACM Transactions on Graphics, 2019, 38(5): 1-12.
- [12] ZHANG K, HAO M, WANG J, et al. Linked dynamic graph CNN: Learning on point cloud via linking hierarchical features [C]//Proceedings of 2021 27th International Conference on Mechatronics and Machine Vision in Practice (M2VIP). Shanghai, China: [s.n.], 2021: 7-12.
- [13] ENGELMANN F, KONTOGIANNI T, LEIBE B. Dilated point convolutions: On the receptive field size of point convolutions on 3D point clouds[C]//Proceedings of 2020 IEEE International Conference on Robotics and Automation (ICRA). Paris, France: IEEE, 2020: 9463-9469.
- [14] QIU S, ANWAR S, BARNES N. Geometric back-projection network for point cloud classification[J]. IEEE Transactions on Multimedia, 2021, 24: 1943-1955.
- [15] 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, MA, USA: IEEE, 2015: 1912-1920.
- [16] UY M A, PHAM Q H, HUA B S, et al. Revisiting point cloud classification: A new benchmark dataset and classification model on real-world data[C]//Proceedings of the IEEE/CVF International Conference on Computer vision. Seoul, Korea (South): IEEE, 2019: 1588-1597.
- [17] YI L, 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): 1-12.
- [18] QI C R, SU H, NIEßNER 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, NV, USA: IEEE, 2016: 5648-5656.
- [19] LI Y, BU R, SUN M, et al. PointCNN: Convolution on  $\chi$ -transformed points[C]//Proceedings of the 32nd International Conference on Neural Information Processing Systems. Montreal, Canada: NIPS, 2018: 828-838.
- [20] LI J, CHEN B M, LEE G H. SO-Net: Self-organizing network for point cloud analysis[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA: IEEE, 2018: 9397-9406.
- [21] LIU X, HAN Z, LIU Y S, et al. Point2sequence: 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. Hawaii, USA: AAAI, 2019: 8778-8785.
- [22] WAN J, XIE Z, XU Y, et al. DGANet: A dilated graph attention-based network for local feature extraction on 3D point clouds [J]. Remote Sensing, 2021, 13(17): 3484.
- [23] 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, Germany: [s.n.], 2018: 87-102.

## 作者简介:



陶志勇(1978-),男,博士,教授,研究方向:机器学习、图像处理、智能信息处理等, E-mail:xyzmail@126.com。



豆淼森(2001-),通信作者,男,硕士研究生,研究方向:计算机视觉、点云处理, E-mail: doumiaosen@126.com。



李衡(1998-),男,硕士研究生,研究方向:机器学习、点云处理, E-mail: PaperLH@163.com。



林森(1980-),男,副教授,研究方向:图像处理与模式识别, E-mail: lin\_sen@126.com。