

# 基于图神经网络和注意力机制的点云分类模型

徐海涛<sup>1,2,3,4</sup>, 郝晓萍<sup>5</sup>, 晁欣<sup>5</sup>, 董少锋<sup>5</sup>, 李祥<sup>1,2,4</sup>

(1. 中国科学院沈阳自动化研究所, 辽宁 沈阳 110016; 2. 中国科学院机器人与智能制造创新研究院, 辽宁 沈阳 110169;  
3. 中国科学院大学, 北京 100049; 4. 辽宁省智能检测与装备技术重点实验室, 辽宁 沈阳 110179; 5. 中国航发动力股份有限公司, 陕西 西安 710021)

**摘要:**为了增强基于深度学习的三维点云分类模型对全局特征的建模能力,提高模型的泛化性能,在 PointNet 的基础上,提出了基于图神经网络和注意力机制融合的点云分类模型。首先,将提取的特征分别通过增加通道注意力模块和空间注意力模块,使模型更加关注全局上下文信息,抑制噪声信息,减少冗余参数,增强对全局特征的建模能力;其次,通过在不同尺度球半径内进行不同 K 值最近邻搜索对编码的输入特征进行构图,既减小了图的规模,降低训练开销,又使模型学习不同层级的特征表示;最后,通过图卷积神经网络汇聚邻域信息,更新节点特征,并将不同图卷积神经网络层输出特征进行相加,融合多层次特征,提高分类准确率。本文在公用数据集 ModelNet40 上进行训练与测试,其总体分类准确率为 88.6%, 优于通用的 3DShapeNets、VoxNet、ECC、PointNet 模型,证明了模型在点云分类上的优越性。

**关键词:**三维点云;注意力机制;图神经网络;多尺度特征融合

中图分类号:TP391;TN958.98 文献标识码:A DOI:10.3969/j.issn.1001-5078.2024.08.006

## Point cloud classification model based on graph neural network and attention mechanism

XU Hai-tao<sup>1,2,3,4</sup>, HAO Xiao-ping<sup>5</sup>, CHAO Xin<sup>5</sup>, DONG Shao-feng<sup>5</sup>, LI Xiang<sup>1,2,4</sup>

(1. Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang 110016, China;  
2. Institutes for Robotics and Intelligent Manufacturing, Chinese Academy of Sciences, Shenyang 110169, China;  
3. University of Chinese Academy of Sciences, Beijing 100049, China; 4. Key Laboratory on Intelligent Detection and Equipment Technology of Liaoning Province, Shenyang 110179, China; 5. China Aviation Engine Power Co., Ltd, Xi'an 710021, China)

**Abstract:** In order to enhance the modeling capability of global features in deep learning-based 3D point cloud classification models and improve their generalization performance, a point cloud classification model based on the fusion of graph neural network and attention mechanism is proposed on the basis of PointNet. Firstly, the extracted features are used to make the model pay more attention to the global context information, suppress the noise information, reduce the redundant parameters, and enhance the modelling ability of the global features by increasing the channel attention module and the spatial attention module, respectively. Secondly, different K-values nearest neighbor searches are performed within multiple scales of sphere radius to construct the input features for encoding, which not only reduces the scale of the graph and training overhead but also enables the model to learn features at different levels. Finally, neighborhood information is aggregated and node features are updated through graph convolutional neural networks. The output features of different graph convolutional neural network layers are summed up to fuse multi-level features and improve classification accuracy. The proposed model is trained and tested on the public dataset ModelNet40, achieving an overall classification accuracy of 88.6%, which outperforms the commonly used 3DShapeNets, VoxNet, ECC, and

**基金项目:**辽宁省应用基础研究计划项目(No. 2022JH2/101300203; No. 2023JH2/101300148)资助。

**作者简介:**徐海涛(2000-),男,硕士,主要从事工业故障诊断的研究。E-mail: xuhaitao@sia.cn

**收稿日期:**2023-11-16; **修订日期:**2023-12-29

PointNet models, demonstrating its superiority in point cloud classification.

**Keywords:** 3D point cloud; attention mechanism; graph neural network; multi-scale feature fusion

## 1 引言

三维点云分类是计算机视觉领域的一个具有挑战性任务,它涉及到从三维空间中获取并处理大量离散数据,并按需进行分类。然而,由于三维点云数据具有无序性、非结构性和点云旋转不变性<sup>[1]</sup>等特征,且点云数据量极大,使得依赖于几何约束和统计规则人工设计特征的传统方法<sup>[2]</sup>,如高斯混合模型<sup>[3]</sup>、支持向量机<sup>[4]</sup>、随机森林<sup>[5]</sup>等,在处理数据时容易忽略邻域点间的联系,无法有效提取环境变化时的特征,泛化能力弱,存在诸多限制。

近年来,随着深度学习理论的发展,卷积神经网络(CNN)<sup>[6]</sup>、图神经网络(GNN)<sup>[7]</sup>和循环神经网络(RNN)<sup>[8]</sup>等不断应用于实际场景并取得了不错的效果,研究者们尝试将这些神经网络迁移到点云分类问题上,使基于深度学习的方法逐渐成为处理点云分类问题的新途径。Su等<sup>[9]</sup>提出了多视图卷积神经网络(MVCNN),它从不同视角捕获3D物体的2D视图,再通过传统的CNN提取特征,最后使用支持向量机进行分类。Maturam等<sup>[10]</sup>提出了VoxNet,将输入点云体素化,再通过三维卷积神经网络提取特征图,用最大池化层处理不重叠的体素块。Wu等<sup>[11]</sup>提出了3DShapeNets,将输入点云体素化后表示为二进制变量的概率分布,继而使用卷积神经网络来提取特征。Zhang等<sup>[12]</sup>引入Transformer架构,并提出了稀疏窗口注意模块,减少空体素的计算,使用不同的注意力机制的变种来提取物体的细粒度特征。Qi等<sup>[13]</sup>提出了PointNet,使用T-Net模块对输入点矩阵进行仿射变换,再逐点通过多层感知器(MLP)提取特征,最后使用最大值池化聚合全局特征,有效地处理了点云数据的三大难题。Qi等<sup>[14]</sup>为了解决PointNet不捕获度量空间局部结构特征的问题,提出了PointNet++,通过采样分组层构建不同邻域大小的局部子集,提取不同尺度的局部特征。相对于PointNet++使用分组采样的方法来弥补PointNet网络丢失的点云数据中的局部结构信息,本文采用GNN来提取度量空间中的局部结构信息,通过多尺度球半径内进行不同K值最近邻搜索构图,来捕获不同层级的局部特征信息,解决PointNet网络的局限性。

GNN作为一种可以对不规则场景建模的神经网络结构,在处理具有复杂拓扑结构、不规则形状等非欧几里得结构化数据方面表现出色。近几年来,研究者们提出了许多基于GNN的点云分类模型。Mohammadi等<sup>[15]</sup>提出了基于图卷积神经网络(GCN)的多视角点云分类方法,利用图卷积网络编码物体和它多视图之间的关,聚合多视图的形状特征。Simonovsky等<sup>[16]</sup>提出了边缘条件卷积(ECC),在空域中对图信号进行操作,滤波器权重依赖于边标签和每个特定样本动态生成。Wang等<sup>[17]</sup>利用GNN去恢复点云本身缺乏的拓扑信息,提出了EdgeConv模块,更好地捕获点云的局部几何特征,动态更新层与层之间的关系图。Li等<sup>[18]</sup>提出了DeepGCNs,将残差连接和扩张卷积应用到GNN。Shi等<sup>[19]</sup>提出了Point-GNN,通过图神经网络预测每个顶点所属类别,并设计了一种自动配准机制来减少平移方差。

GNN通过汇聚局部特征信息,提高模型对全局特征的建模能力,而注意力机制<sup>[20]</sup>让模型更加关注输入数据中的有用信息,进一步提升模型性能。因此,本文提出了一种基于图神经网络和注意力机制融合的点云分类模型。该模型将原始点云数据作为输入,将提取的特征使用图神经网络进行编码,表示点云局部邻域内关系,通过设置不同邻域大小和不同数目K最近邻,提取更加丰富的局部邻域特征信息,增强模型对全局特征的表征能力。然后采用注意力机制对不同节点的权重进行调整,关注重要信息,而注意力机制参考CBAM模块<sup>[21]</sup>结合使用通道注意力机制和空间注意力机制,在通道和空间维度上分别生成注意力特征信息,与输入特征图相乘后进行自适应调整,提高了点云分类的准确性。

## 2 算法描述

本文在PointNet模型基础上进行改进,在PointNet网络提取特征阶段上融入注意力机制和图卷积神经网络,如图1所示,在共享MLP层后面连接CBAM注意力模块,对重要位置信息和特征信息加以强调;通过多尺度邻域大小内选择不同K值构图,使用GCN聚合局部邻域内信息,并融合不同图卷积网络的输出,既增强了网络对点云数据中局部结构信息的提取能力,又改善了网络对全局特征信

息的建模能力,提高分类精度。

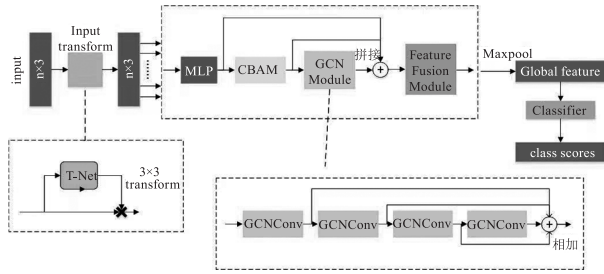


图1 点云分类网络总体框架图

Fig. 1 Overall framework diagram of point cloud classification network

## 2.1 CBAM 模块

CBAM 模块作为注意力机制的一种,如图 2 所示,它分别在通道和空间上通过神经网络生成一个掩码,对当前所需关注点进行评估,使模型关注重要特征并抑制不重要特征。

通道注意力机制关注特征通道间的关系,使用最大池化和平均池化聚合特征图空间信息,其公式:

$$M_c(F) = \sigma(\text{MLP}(\text{AvgPool}(F)) + \text{MLP}(\text{MaxPool}(F)))$$

$$= \sigma(W_1(W_0(F_{\text{Avg}}^c)) + W_1(W_0(F_{\text{Max}}^c))) \quad (1)$$

其中,  $F_{\text{Avg}}^c$  和  $F_{\text{Max}}^c$  分别表示平均池化特征和最大池化特征;  $\sigma$  为 sigmoid 函数;  $W_0$ 、 $W_1$  作为共享 MLP 的权重参数。

空间注意力机制关注特征中的空间关系,在通道维度上使用平均池化和最大池化来聚合特征图的通道信息,其公式:

$$M_s(F) = \sigma(f^{7 \times 7}(\text{[ AvgPool}(F); \text{MaxPool}(F) ]))$$

$$= \sigma(f^{7 \times 7}(\text{[ } F_{\text{Avg}}^s; F_{\text{Max}}^s ])) \quad (2)$$

其中,  $\sigma$  为 sigmoid 函数;  $f^{7 \times 7}$  表示卷积核大小为  $7 \times 7$  的卷积运算。

## 2.2 图卷积神经网络

图卷积神经网络将卷积操作引入到图数据中,其通过聚合节点本身特征和邻接节点的特征来生成该节点的特征表示,不断堆叠图卷积层、激活函数和池化层来提取节点特征的最终表示。

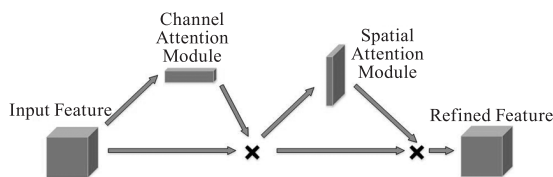


图2 CBAM 模块

Fig. 2 CBAM module

图卷积神经网络中节点更新准则如式:

$$H^{(l+1)} = \sigma(\tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} H^{(l)} W^{(l)}) \quad (3)$$

其中,  $H^{(l)}$  为第  $l$  层节点矩阵表示;  $H^{(l+1)}$  为第  $l+1$  层节点矩阵表示;  $\sigma$  为非线性激活函数;带自连接的图邻接矩阵  $\tilde{A} = A + I_N$ ;  $I_N$  为单位矩阵;  $\tilde{D}_i = \sum_j \tilde{A}_{ij}$ ;  $W^{(l)}$  是需要学习的权重参数矩阵。

## 2.3 基于图神经网络和注意力机制融合点云分类模型设计

将输入的原点云数据依次通过 MLP 模块和 CBAM 模块提取特征图,用于后续图构建。受到 PointNet++ 的启发,构图时在多尺度球半径内进行不同  $K$  值最近邻搜索,若半径内搜索不到足够的邻域点,则使用最近邻点填充至  $K$ ,扩充局部信息,构图过程如图 3 所示。将特征图中每点初始化为图中的节点,与最邻近的  $K$  个节点相连,构建无向图,将无向图送入多层 GCN,汇聚节点局部特征信息,更新节点特征向量。然后将不同图卷积神经网络层的输出相加后取平均,保证网络稳定的情况下进行特征融合,增强模型对特征信息的学习能力。模型最后将 MLP 模块输出、CBAM 模块输出与图卷积神经网络输出进行拼接,通过  $1 \times 1$  卷积层将三者特征信息进行融合,增强模型对点云全局特征信息的建模能力,并输出最终的特征向量,最后经过分类器得到该点云的输出类别。

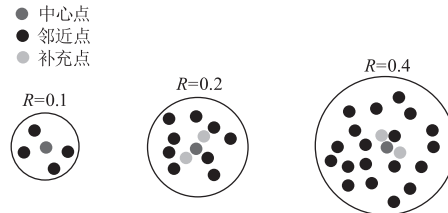


图3 图构建

Fig. 3 Graph construction

## 3 实验结果与分析

为了验证该模型的分类效果,本研究在 ModelNet40<sup>[22]</sup> 数据集上进行训练和评估。ModelNet40 有 40 个类别共 12311 个 CAD 模型,其中有 9843 个物体用于训练模型,2468 个物体用于模型测试,对 ModelNet40 中部分点云数据进行可视化处理,如图 4 所示。

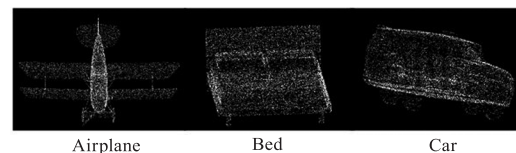


图4 点云数据

Fig. 4 Point cloud data

分类实验的系统环境为 window10, GPU 配置为 NVIDIA GeForce GTX4090, 实验环境为 Python3.9, Pytorch2.0.1。在训练时,将所有原始三维点云数据进行预处理,以增强神经网络的鲁棒性。

实验中,构图时所进行的多尺度半径与不同 K 值最近邻搜索的参数为(0.1,16)、(0.2,32)、(0.4,64)和(0.8,128), epoch 设置为 600, batch\_size 设置为 16,使用 adam 优化器对模型参数进行优化,初始学习率设置为 0.001,权重衰减率设置为 0.0001,模型中的激活函数为 ReLU 函数,使用 NLLloss 函数作为分类网络的损失函数,网络参数与 PointNet 参数设置相同,其余参数设置如表 1 所示。此外,对训练数据进行了一定程度的随机平移和放缩,以实现数据增强。

本文提出模型在 ModelNet40 数据集上分类精度与现有点云分类模型 3DShapeNets、VoxNet、ECC、PointNet 进行对比,使用平均分类精度(mA)和总体分类精度(OA)来评估不同模型的性能,其中,mA为测试集中每个类别分类正确结果的平均值,OA为测试集中分类正确的平均值,点云分类结果对比如表 2 所示,模型在不同训练阶段的平均分类精度和总体分类精度如图 5 所示。

由图 4 可知,模型在前 200 轮训练时,分类精度不断提高,200 轮训练后,分类精度基本稳定,网络趋于收敛,并在最终 OA 达到 88.6%,mA 达到 86.0%。由表 2 的数据可知,本文所提出的模型相比于 3DShapeNets、VoxNet、ECC 和 PointNet 模型在点云分类任务中的准确率均有了提高。其中,相较于 PointNet 模型,所提出模型的 OA 提高了 1%,mA 提高了 2.3%,这表明所提出模型能够有效提取点云的局部特征信息,提高分类精度。

表 1 模型详细参数

Tab. 1 Detailed model parameters

Type	Operation
GCN Module	GCNConv(input_dim,512)
	GCNConv(512,128)
	GCNConv(128,128)
	GCNConv(128,128)
Feature FusionModule	Conv(in_dim,out_dim,1),BN,ReLU
Classifier	FC(512),BN,ReLU,Dropout(0.4)
	FC(256),BN,ReLU,Dropout(0.5)
	FC(40)

表 2 点云分类结果

Tab. 2 Point cloud classification results

算法	平均分类精度/%	总体分类精度/%
3DShapeNets <sup>[11]</sup>	77.3	84.7
VoxNet <sup>[10]</sup>	83.0	85.9
ECC <sup>[16]</sup>	83.2	87.4
PointNet <sup>[23]</sup>	83.7	87.6
本文算法	86.0	88.6

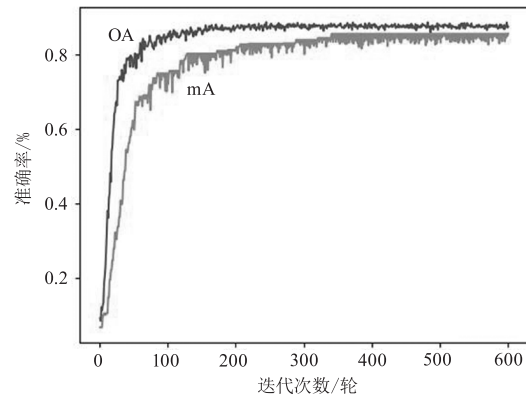


图 5 模型分类结果

Fig. 5 Model classification results

## 4 结论

为提高模型对局部点云特征信息的利用率,加强模型对全局特征的建模能力,本文提出了基于神经网络和注意力机制点云分类模型。该方法融合 CBAM 模块和 GCN,利用 CBAM 模块选取更重要的特征信息和位置信息,提升模型性能;使用 GCN 和多尺度半径内不同 K 值构图,聚合点云局部邻域内特征信息,提高模型对点云数据内局部特征的提取能力,利于后续全局特征的学习,在 ModelNet40 数据集上取得了不错的结果。但网络整体结构较为复杂,对算力要求较高,训练时间较长;后续将在这些方面进行优化,减少模型的计算量和复杂度,增加模型鲁棒性,使模型在实际工程中得到应用。

## 参考文献:

- [1] Wang Zhenyan, Sun Hongyan, Sun Xiaopeng. Overview of large scale 3D point cloud processing based on deep learning [J]. Computer System Applications, 2023, 32(2): 1-12. (in Chinese)  
王振燕,孙红岩,孙晓鹏.基于深度学习的大规模三维点云处理综述[J].计算机系统应用,2023,32(2):1-12.
- [2] Wen Pei, Cheng Yinglei, Yu Wangsheng. Review of point cloud classification methods based on deep learning [J].

- Progress in Laser and Optoelectronics, 2021, 58 (16): 49–75. (in Chinese)  
文沛,程英蕾,余旺盛. 基于深度学习的点云分类方法综述[J]. 激光与光电子学进展, 2021, 58(16): 49–75.
- [3] Lalonde J F, Unnikrishnan R, Vandapel N, et al. Scale selection for classification of point-sampled 3D surfaces [C]//Fifth International Conference on 3-D Digital Imaging and Modeling(3DIM'05). IEEE, 2005: 285–292.
- [4] Gao Z, Liu X. Support vector machine and object-oriented classification for urban impervious surface extraction from satellite imagery [C]//2014 The Third International Conference on Agro-Geoinformatics. IEEE, 2014: 1–5.
- [5] Wang C, Shu Q, Wang X, et al. A random forest classifier based on pixel comparison features for urban LiDAR data [J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2019, 148: 75–86.
- [6] Chua L O, Roska T. The CNN paradigm [J]. IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications, 1993, 40(3): 147–156.
- [7] Scarselli F, Gori M, Tsoi A C, et al. The graph neural network model [J]. IEEE Transactions on Neural Networks, 2008, 20(1): 61–80.
- [8] Zaremba W, Sutskever I, Vinyals O. Recurrent neural network regularization [J/OL]. arXiv Preprint arXiv: 1409.2329, 2014.
- [9] 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, 2015: 945–953.
- [10] 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), IEEE, 2015: 922–928.
- [11] 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, 2015: 1912–1920.
- [12] Zhang C, Wan H, Shen X, et al. PVT: Point-voxel transformer for point cloud learning [J]. International Journal of Intelligent Systems, 2022, 37(12): 11985–12008.
- [13] 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 Recognition, 2017: 652–660.
- [14] Qi C R, Yi L, Su H, et al. PointNet++: deep hierarchical feature learning on point sets in a metric space [J/OL]. <https://arxiv.org/abs/1706.02413>.
- [15] Mohammadi S S, Wang Y, Del Bue A. Pointview-gcn: 3d shape classification with multi-view point clouds [C]//2021 IEEE International Conference on Image Processing (ICIP), IEEE, 2021: 3103–3107.
- [16] 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, 2017: 3693–3702.
- [17] Wang Y, Sun Y, Liu Z, et al. Dynamic graph cnn for learning on point clouds [J]. Acm Transactions On Graphics (tog), 2019, 38(5): 1–12.
- [18] Li G, Müller, Matthias, et al. DeepGCNs: making GCNs Go as deep as CNNs [J/OL]. <https://arxiv.org/abs/1910.06849>.
- [19] Shi W, Rajkumar R. Point-gnn: graph neural network for 3D object detection in a point cloud [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 1711–1719.
- [20] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need [J]. Advances in Neural Information Processing Systems, 2017, 1706: 03762.
- [21] Woo S, Park J, Lee J Y, et al. Cbam: convolutional block attention module [C]//Proceedings of the European Conference on Computer Vision (ECCV), 2018: 3–19.
- [22] 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, 2015: 1912–1920.
- [23] Hou Xiangdan, Yu Xixin, Liu Hongpu. A 3D point cloud classification and segmentation model based on graph convolutional networks [J]. Progress in Laser and Optoelectronics, 2020, 57(18): 204–211. (in Chinese)  
侯向丹,于习欣,刘洪普. 基于图卷积网络的三维点云分类分割模型 [J]. 激光与光电子学进展, 2020, 57(18): 204–211.