

基于多尺度双分支双注意力的点云分类网络

顾君豪, 张孙杰, 秦辰栋

(上海理工大学光电信息与计算机工程学院, 上海 200093)

摘要: 尽管 Transformers 在三维点云处理中已取得显著进展, 但高效且准确地学习有价值的低频和
高频信息仍然是一个挑战。此外, 现有方法大多侧重于局部空间信息, 忽略了全局空间的信息, 从而导
致信息的丢失。本文提出了一种新的点云学习网络, 称为多尺度双分支双注意力网络。首先, 在点云
的提取过程中, 与在固定的尺度上寻找邻近点的提取方法相比, 利用多尺度 K 近邻 (K-nearest neighbor,
KNN) 的方法不仅保留了局部结构细节, 还更有效地捕获了全局几何信息。其次, 本文引入了双分支双
注意力架构提取不同空间特征, 提出了局部窗口注意力与全局通道内容注意力双注意力机制, 分别提
取网络的低频信息与高频信息。在此基础上, 本文在分类头中引入 GR-KAN (Group-rational
Kolmogorov-Arnold) 层代替传统使用的多层感知器 (Multilayer perceptron, MLP) 层, 其能够更灵活地处
理非线性特征, 使得网络对复杂的数据集更加敏感。最后, 大量实验表明, 本文提出的模型在
ModelNet40 数据集获得了 93.8% 的准确率, 在 ScanObjectNN 数据集上获得了 86.5% 准确率, 显示了其
在三维点云处理中优越的性能和广阔的应用前景。

关键词: 点云分类; 多尺度 KNN; Transformer; GR-KAN; 双分支

中图分类号: TP391 **文献标志码:** A

Multi-scale Dual-Branch Dual-Attention-Based Point Cloud Classification Network

GU Junhao, ZHANG Sunjie, QIN Chendong

(School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China)

Abstract: Although Transformers have made significant progress in 3D point cloud processing, efficiently and accurately learning valuable low-frequency and high-frequency information remains a challenge. Moreover, most existing methods focus primarily on local spatial information, neglecting global spatial information, which leads to information loss. This paper proposes a novel point cloud learning network, referred to as the multi-scale dual-branch dual-attention network. First, in the feature extraction process of the point cloud, compared to methods that search for neighboring points at a fixed scale, the multi-scale K-nearest neighbor (KNN) approach not only preserves local structural details but also more effectively captures global geometric information. Second, this paper introduces a dual-branch dual-attention architecture to extract different spatial features, proposing a dual-attention mechanism combining local window attention and global channel content attention to extract low-frequency and high-frequency information from the network, respectively. Additionally, on this basis, this paper introduces the

基金项目: 国家自然科学基金(61603255); 上海市晨光计划项目(18CG52)。

收稿日期: 2024-10-09; **修订日期:** 2025-01-10

group-rational Kolmogorov-Arnold (GR-KAN) layer into the classification head, replacing the traditionally used multilayer perceptron (MLP) layer, which allows for more flexible handling of nonlinear features and makes the network more sensitive to complex datasets. Finally, extensive experiments demonstrate that the proposed model achieves an accuracy of 93.8% on the ModelNet40 dataset and 86.5% on the ScanObjectNN dataset, showcasing its superior performance and broad application prospects in 3D point cloud processing.

Key words: point cloud classification; multi-scale KNN; Transformer; GR-KAN; dual-branch

引言

随着三维传感技术的发展,点云在多媒体、场景理解、智能驾驶和机器人等领域的应用和研究不断增加。激光雷达作为一种重要的三维数据获取工具,广泛应用于各种场景。与二维图像相比,三维点云可以提供足够的空间和几何信息。然而,由于三维点云数据具有离散分布、非结构化、无序性的特点,对其进行有效处理一直是一个巨大的挑战。近年来,随着深度学习的不断发展,利用深度学习处理三维点云成为了一个重要的研究手段。

基于点的方法是处理点云的常见方式,相比投影和体素方法,它直接利用三维坐标和法线特征作为输入,减少了局部几何信息的损失。早期 Charles 等^[1]提出 PointNet,利用多层感知器(Multilayer perceptron, MLP)和最大池化(Maxpooling)实现特征学习。为了聚合局部特征,PointNet++^[2]分层地应用 PointNet,使用查询球分组来构建局部邻域。随后,基于点的方法被扩展到各种局部聚集算子,主要分为基于卷积和注意力两类方法。

基于卷积的方法可以分成通过卷积核来学习局部区域的特征与通过图形建立局部点之间的连接,并对局部几何信息进行建模。通过改变求和顺序,将其简化为矩阵乘法和 2D 卷积,从而显著提高了内存效率。Zhao 等^[3]提出的 PointWeb 通过密集点对点连接提升局部特征学习。Wu 等^[4]提出的 PointConv 实现了平移和置换不变的卷积操作,并通过反卷积恢复点云的原始分辨率。KPConv^[5]利用预定义核点进行操作,但固定核点在复杂 3D 变化下表现有限。PAConv^[6]引入了动态核,通过 MLP 从相对点位置学习权值。图卷积 DGCNN^[7]构建动态邻域进行图卷积,而 3DGCN^[8]在深度学习网络中引入了平移和尺度不变性,并用图极大池化机制定义了可学习核。DeepGCN^[9]通过残差和膨胀卷积提升深度学习效果。AdaptConv^[10]使用了适应性卷积核捕获点云中的多样关系。Chang 等^[11]提出了一种基于 3D 形状知识图谱的跨领域检索方法,通过图嵌入和相似性度量策略实现跨领域多模态的 3D 形状检索。最近的进展进一步凸显了图卷积网络(Graph convolutional network, GCN)中分层结构和消息传递机制对于点云的有效性。HPGNN^[12]、CoFiNet^[13]和 Point-GNN^[14]等网络展示了这些方法用于配准、对象检测和分类等任务的前景。

2021 年,Zhao 等^[15]提出了 Point Transformer,作者提出点云本质上是嵌入在 3D 空间中的集合,而自注意力算子本质上是一个集合算子,它与输入元素的排列和基数是不变的,所以 Transformer 模型特别适合于点云处理。因此,自注意力的方法运用在 3D 点云中具有天然的适用性。随后有很多基于 Transformer 的框架问世。PCT(Point cloud Transformer)^[16]采用了 PointNet 架构,其中共享 MLP 层被标准的 Transformer 块取代。结合原始位置编码和输入嵌入,为每个点生成独特的特征,通过替换自注意力模块的输出来增强网络性能,PCT 在点云分类中获得了最先进的性能。Zhong 等^[17]提出了一种端到端逐点学习全局特征的方法,直接从点云数据中提取几何和语义特征,再通过多尺度转换器生成全

局特征。Liu等^[18]提出的Pointcont通过聚类高频和低频特征增强特征表达能力,采用两层简单的MLP代替注意力层以降低模型复杂度,但这种方法在处理复杂点云时性能有所下降。Han等^[19]在DPCT (Dual-point cloud Transformer)中提出了双注意力策略,设计了两个注意力模块以充分提取特征。然而,由于将原始特征直接传入两个模块,导致两个分支捕获过多相似信息,因此产生信息冗余。Zhao等^[20]提出了一种新的方式,通过层次化地探索局部信息和多视图之间的关系,有效地提取局部特征,并聚合这些特征以生成区分性强的形状描述符,从而在3D形状分类和检索任务上达到了先进的性能。Zhao等^[21]提出了一种基于层次多尺度变换器网络的3D形状识别方法,通过多级区域变换器模块实现局部特征提取,并结合全局引导变换器整合全局信息。该方法利用跨粒度对比学习机制,在特征空间中实现更有效的对齐与增强,显著提升了分类与检索性能。

作为替代MLP的方案,Kolmogorov-Arnold(KAN)^[22]网络于2024年被提出,其因较高的参数效率而备受关注。KAN通过B样条曲线参数化复杂函数,使用较少参数即可近似复杂函数,因此将其集成到Transformer中成为潜在方向。然而,KAN的成功应用目前还非常有限。原因是B样条在GPU上的并行性能较差,递归计算导致即使网络优化后仍较慢。此外,KAN中每个输入输出对需独立的参数和基函数,隐藏层增加时参数量呈指数增长,导致计算开销大、扩展性差。同时,其权重初始化方式也未能满足收敛要求,影响了训练稳定性和性能。为解决这些问题,Group-rational KAN (GR-KAN)^[23]作为KAN的变体被提出,旨在替代Transformer中的MLP。GR-KAN计算效率高,易于集成到视觉Transformer中,专为GPU架构设计,并且采用有理激活函数和方差保持初始化,确保训练稳定性并减少计算开销。其CUDA(Compute unified device architecture)实现更适配现代GPU,边组共享策略进一步降低了计算负担,在提升计算效率的同时保持性能稳定。

本文提出了一种双分支架构,通过并行结构分别经过Maxpooling和平均池化操作(Avgpooling)提取高频与低频特征信息,避免了注意力在处理原始特征时出现分布不稳定和信息冗余的问题,从而使注意力更聚焦于处理后的特征信息进行精确提取。高频分支注重细节,提取物体的局部精细信息;而低频分支则捕捉全局上下文,保留类别的整体特征。这种设计不仅增强了特征的多样性,也解决了单一分支结构难以同时兼顾细节与全局信息的不足。相较于仅使用MLP代替注意力层来聚类低频特征,本文在此基础上通过局部窗口注意力(Local window attention, LWA)和全局通道内容注意力(Global channel content attention, GCCA)进一步细化了特征的表达层次,使得高频和低频特征在不同注意力模块中融合后,实现了局部细节与全局上下文的联合建模。这种融合方式不仅减少了信息冗余问题,更显著提升了模型对复杂点云场景的特征表达能力。最后,本文将GR-KAN层融入到网络中,作为点云分类头。在点云分类任务中能提供更强的特征提取能力、计算效率和鲁棒性,使其成为优于传统MLP的选择。本文做出的主要贡献概况如下:

(1) 在高频特征提取器中,设计了LWA,在局部窗口内计算注意力权重并进行加权聚合,通过位置编码增强空间感知能力,有效地保留了细粒度特征,降低了计算复杂度。

(2) 在低频特征提取器中,引入GCCA,聚焦特征内容而非空间位置,捕捉全局依赖并动态聚焦相似特征,提升非欧几里得结构处理能力,增强模型判别力和平衡计算开销。

(3) 采用多尺度KNN搜索获取不同尺度下的邻近点,实现局部与全局信息融合,保留结构细节并捕获全局几何,提升特征表示的丰富性与准确性。

(4) 首次在点云分类网络中引入GR-KAN,通过核注意力建模局部依赖性并保持全局上下文,有效地增强了模型的特征提取能力。

1 本文方法

1.1 整体框架

图1展示了本文所提出的完整网络架构。这个架构的核心是由3个阶段的特征提取模块(Feature extraction block, FEB)组成。给定一个输入点云 $P \in \mathbb{R}^{N \times 3}$, N 表示所有点云数量。

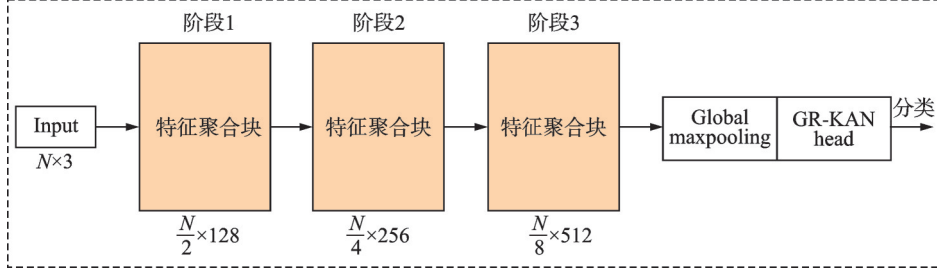


图1 本文网络的整体架构

Fig.1 Overall architecture of the proposed network

在阶段1中,初始特征聚合块首先将点云划分为重叠的块,然后将每个点的坐标嵌入到一个新的特征空间。在每一层中,通过减少点的数量和增加特征的维度,模型可以更高效地处理更具语义的全局结构。减少点的数量可以降低计算复杂度,特别是在注意力机制中,这可以显著减少需要计算的注意力权重的数量。尽管点的数量减半,但通过增加特征维度,可以保持或甚至增加每个点所携带的信息量,从而确保信息不会在层次传递过程中丢失。所以本文采用每一个阶段将点云的数量减半,并将通道数翻倍的策略。

对于分类任务,网络在完成特征提取后会输入到最终的分类器。分类器首先进行全局最大池化,然后连接两个全连接层。通过这种逐步提取和细化特征的方式,网络能够从原始的三维点云中提取出高维且信息量较多的特征,用于后续的分类任务。

1.2 特征聚合块

以阶段1为例,局部特征提取块如图2所示。给定一个输入点云 $P \in \mathbb{R}^{N/2 \times 128}$, 初始特征聚合块首先通过最远点采样(Farthest point sampling, FPS)以两个速率对中心点进行下采样,然后执行本文提出的多尺度KNN算法对局部块进行分组。

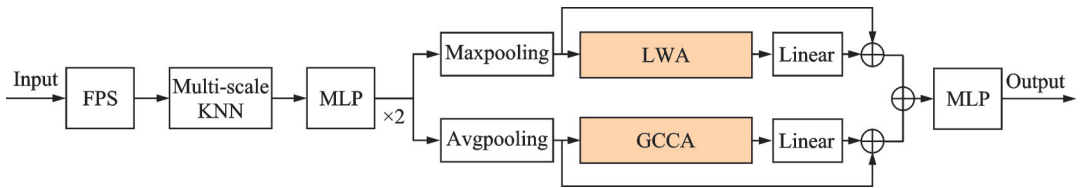


图2 特征聚合块的细节图

Fig.2 Detailed diagram of feature aggregation block

本文采用EdgeConv方法在每一个块中提取周围特征 f_r 与中心点特征 f_c 之间的关系 f_g , 表达式为

$$f_g = \text{MLP}(\|f_c, f_r - f_c\|) \quad (1)$$

式中: $(f_r - f_c)$ 代表周围特征 f_r 减去中心特征 f_c 来获得中心点 i 的相邻特征, $\|\cdot\|$ 代表级联运算, MLP 为包括逐点卷积层、批归一化层和激活函数的简单网络。

本文提出了一种混合池化策略来聚合局部补丁的特征。在以往大多数的工作中,最大值池化被用来提取聚合网络的高频特征信息,但是缺少对低频特征信息的提取。本文使用了由Maxpooling和Avgpooling操作组成的并行结构,而不是单独使用Maxpooling来提取特征信息。通过Maxpooling操作来提取网络的高频信息,通过Avgpooling操作来提取网络的低频信息。

高频信息提取分支可以表示为

$$f_L = \text{LWA}(\text{Maxpool}(f_g)) \quad (2)$$

$$f_m = \text{Linear}(f_L) + f_L \quad (3)$$

式中: f_m 表示Maxpooling分支获得的特征;Linear代表一个线性层,在最后做了一个残差连接来保证不会有信息丢失。

低频信息提取分支可以表示为

$$f_G = \text{GCCA}(\text{Avgpool}(f_g)) \quad (4)$$

$$f_a = \text{Linear}(f_G) + f_G \quad (5)$$

式中 f_a 表示Avgpooling分支获得的特征。

最后,将局部特征提取分支和全局特征提取分支的特征进行特征融合,并输入到1个MLP块中,得到阶段1特征聚合块的输出。

1.3 多尺度 KNN

如图3所示,在给定输入点云的情况下,执行FPS以获得称为采样点集的点云子集,图中 R 为半径。为了保证采样点感受野的多样性,使用查询球分组的方法构建每个采样点的多尺度邻域。通过在不同尺度下进行KNN搜索,获取各个尺度范围内的邻近点,从而兼顾局部细节和全局信息。

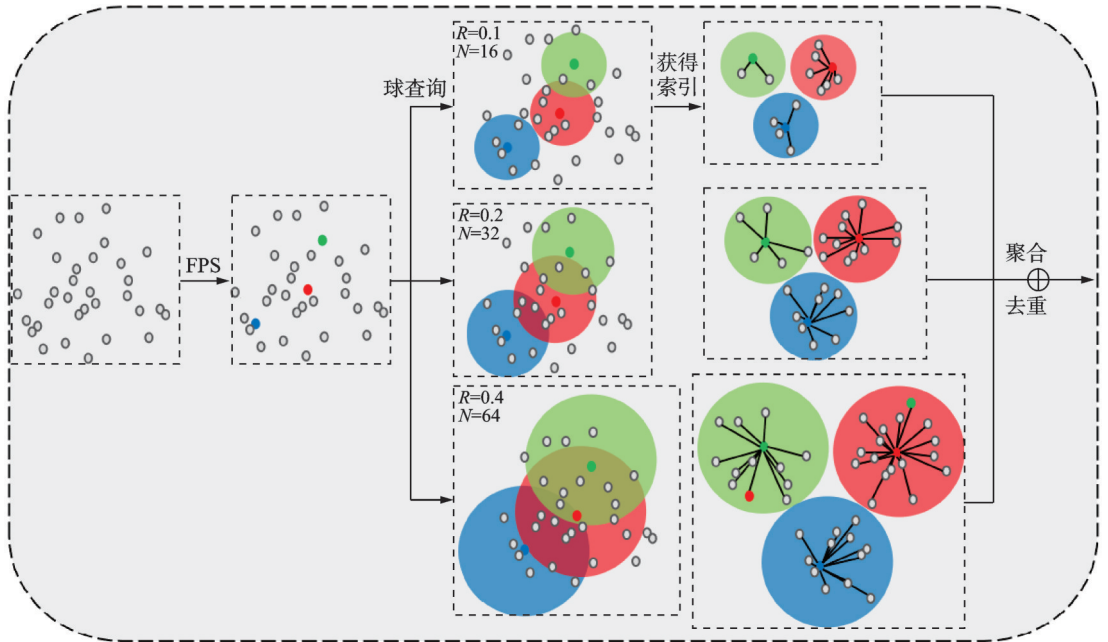


图3 多尺度 KNN 特征聚合细节图

Fig.3 Detailed feature aggregation diagram of multi-scale KNN

在具体实现中,首先选取多个不同的尺度参数,对同一个查询点进行KNN搜索,分别获取不同尺度下的邻近点集合。接着将这些集合中的邻近点进行聚合,以形成一个包含多尺度信息的综合邻近点集合。在聚合过程中,应用去重策略以避免重复点的影响,确保每个邻近点在综合集合中唯一存在。最终,通过这种多尺度的邻近点聚合方法,不仅保留了局部结构细节,而且有效地捕获了全局几何信息,从而提升了特征表示的丰富性和准确性。这一策略在保持计算效率的同时,显著增强了模型对不同尺度特征的感知能力。

1.4 局部窗口注意力

局部窗口注意力用来提取高频特征信息,其框架图如图4所示,将经过Maxpooling的聚集特征 F_m 作为输入。

首先,将 F_m 投影到3个不同的特征空间中,以生成查询矩阵 Q 、键矩阵 K 和价值矩阵 V ,表达式为

$$Q = F_m * W_q \quad (6)$$

$$K = F_m * W_k \quad (7)$$

$$V = F_m * W_v \quad (8)$$

式中: $W_q, W_k, W_v \in \mathbb{R}^{d \times d}$ 为可学习的权重矩阵, d 为特征空间维数。随后计算注意力(Attention $\in \mathbb{R}^d$)为

$$\text{Attention} = \text{softmax}\left(\frac{Q^T K}{\sqrt{d}} + p_r\right) \quad (9)$$

式中: p_r 为可学习位置编码矩阵。通过将邻域点特征 $f_i (i=1, 2, 3)$ 与线性投影后的 W_v 矩阵逐元素相乘,实现了特征的有效融合,得到了高频特征点,即

$$f_{vi} = f_i * \text{Linear}(W_v) \quad i=1, 2, 3 \quad (10)$$

随后,将高频特征点分别与注意力头进行特征融合,表达式为

$$f_{vai} = f_{vi} * \text{Attention} \quad i=1, 2, 3 \quad (11)$$

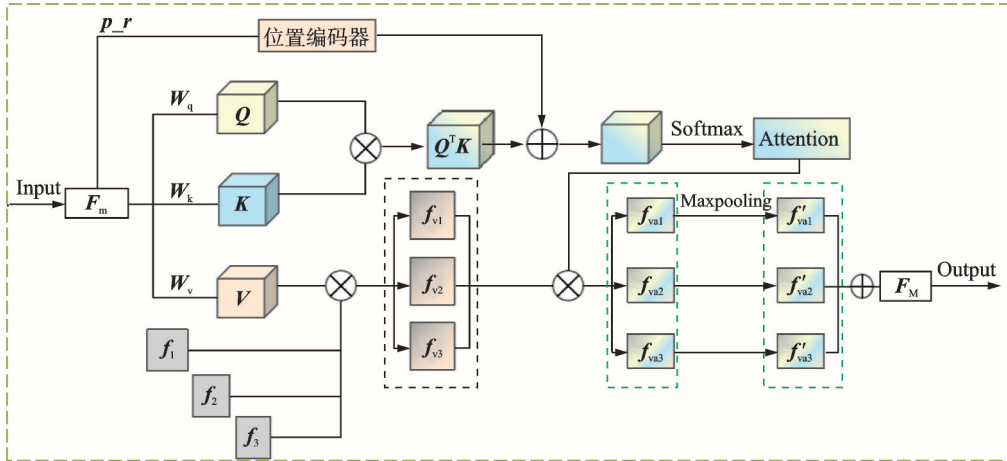


图4 局部窗口注意力框架图

Fig.4 Framework diagram of local window attention

特征融合将高频特征点的表示分配给不同的注意力头,使得每个头可以独立处理其对应的特征。该操作充分利用了多头注意力机制的优势,使模型能够捕捉到不同头部之间的特征差异。随后将捕捉到的特征 $f'_{vai} (i=1, 2, 3)$ 进行融合输出,表达式为

$$F_M = \text{concat}(f'_{vai}) \quad (12)$$

式中 concat 表示特征拼接。本文通过拼接多个高频特征,充分利用了不同高频特征点的信息,增强了特征表示的丰富性。

1.5 全局通道内容注意力

继 Point Transformer 计算局部空间邻居之间的自我注意力之后,本文提出了一种基于通道内容的注意力,如图 5 所示。它根据查询的内容动态地将所有查询划分为多个集群,并选择相应的键和值来计算局部自注意力。图中 $F_a \in \mathbf{R}^{S \times d}$ (S 为特征长度) 为 1 组特征向量, F_A 为经过全局通道内容注意力框架得到的特征向量。

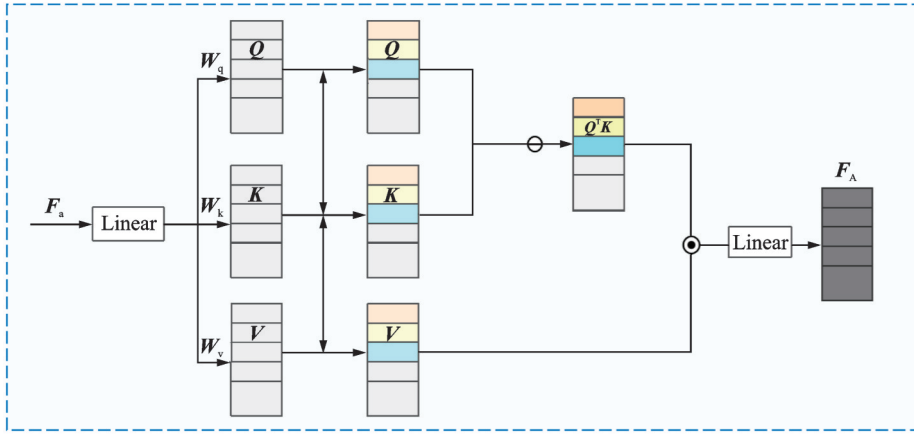


图 5 全局通道内容注意力框架图

Fig.5 Framework diagram of global channel content attention

具体地,首先在输入端设计 1 个线性层调整输入维度,保证数据的一致性和模型的稳定性,表达式为

$$f_a = \text{Linear}(F_a) \quad (13)$$

随后使用 3 个矩阵 W_q 、 W_k 、 W_v 将输入嵌入矩阵 f_a 变换为查询矩阵 Q 、键矩阵 K 和价值矩阵 V ,表达式为

$$Q = f_a * W_q \quad (14)$$

$$K = f_a * W_k \quad (15)$$

$$V = f_a * W_v \quad (16)$$

然后使用聚类算法将查询分散到不同的簇中。K-均值聚类算法是解决聚类问题的经典算法,然而由于 K-均值聚类通常使得每个聚类包含不同数量的查询,因此该算法不能通过使用 GPU 以并行方式实现。为了解决这个问题,本文采用平衡二进制聚类算法,该算法将 1 组查询分层地平均划分为 2 个聚类,表达式为

$$c_1 = \frac{1}{S/2} \sum_{i=1}^{S/2} q_i, c_2 = \frac{1}{S/2} \sum_{i=\frac{S}{2}+1}^S q_i \quad (17)$$

$$r_i = \frac{\text{dist}(q_i, c_1)}{\text{dist}(q_i, c_2)} \quad \forall i \in [1, S] \quad (18)$$

$$[i_1, i_2, \dots, i_S] = \text{argsort}(\{r_i | i = 1, 2, \dots, S\}) \quad (19)$$

$$\mathcal{C}_1 = \left\{ q_i \middle| j = 1, 2, \dots, \frac{S}{2} \right\}, \mathcal{C}_2 = \left\{ q_i \middle| j = \frac{S}{2} + 1, \frac{S}{2} + 2, \dots, S \right\} \quad (20)$$

式中: c_1, c_2 分别表示第1和第2个子集的特征中心,即质心; r_i 表示每个点到质心的距离比值; dist表示特征空间中的欧几里得距离; \mathcal{C}_1 和 \mathcal{C}_2 表示使用 r_i 的排序执行硬分配后,重新划分的两个大小相同的子集。在执行 n 次迭代的二进制聚类,得到 L 个子集具有相同的大小。

由式(17~20)可知:聚类分配依赖于所有聚类质心和每个样本之间的距离,二进制聚类从查询 Q 随机划分为两个聚类开始,计算两个聚类质心,之后计算距离比来执行硬分配。

通过聚类,查询 Q 被分组为 L 个子集 $\{Q_i | i = 1, 2, \dots, L\}$, 其中每个子集具有相等的大小 Q_i 。随后,通过相同的索引将键 K 和值 V 分成 $\{K_i | i = 1, 2, \dots, L\}$ 和 $\{V_i | i = 1, 2, \dots, L\}$ 。每个子集中的自我注意力(Self-attention, SA)表示为

$$SA = \text{Softmax} \left(\frac{Q - K}{\sqrt{d}} \right) \odot V \quad (21)$$

式中“ \odot ”表示 Hadamard 积,即对应元素相乘。首先计算查询和键之间的差异,然后除以特征维度 d 的平方根,用于归一化,确保差异在不同维度上是可比的。最后,Softmax 函数的结果与值矩阵 V 进行逐元素相乘,得到加权的值。

在点云分类任务中,基于内容的注意力机制通过对相似特征点的聚焦,帮助模型学习到更具判别性的特征表示,从而提升分类精度。这些优势使得该机制成为处理点云数据的理想选择。

1.6 GR-KAN 分类头

相比于传统的 MLP, KAN 网络的优势在于它的灵活性和适应性。通过注意力机制、卷积选择、灵活的激活函数以及正则化, KAN 能够更好地处理复杂数据并提高模型的泛化能力。这使得 KAN 网络在处理高维和复杂结构的点云数据时具有明显的优势。但是传统 KAN 网络存在3个主要缺点:(1)其标准 B 样条函数在现代 GPU 的并行计算架构上效率不佳,由于递归计算,导致速度显著降低;(2)参数和计算效率低,每个输入-输出对都需要独立的参数和基函数,随着隐藏层大小增加,参数数量呈指数增长,带来计算开销和扩展性问题;(3)权重初始化与 MLP 相似,但无法满足 KAN 的收敛要求,导致训练过程中不稳定性和性能下降。上述缺点限制了其在大型深度神经网络中的使用。本节将改进其设计,以更好地适应现代变压器,从而能够用 KAN 取代 MLP 层。本文引入 GR-KAN 来代替原有的 KAN 网络。

GR-KAN 使用有理函数作为 KAN 的基函数,并在一组边之间共享参数。首先,从理论角度来看,有理函数能够比多项式更高效且准确地逼近更广泛的函数,尤其是那些具有奇点或剧烈变化的函数。由于 B 样条本质上是局部多项式的求和,因此有理函数在建模复杂行为方面,相较于 B 样条具有理论上的优势。其次,从实践角度来看,有理激活函数已成功应用于神经网络中的激活函数,证明了其在实际应用中的有效性。

如式(22)所示,将每条边上的函数 $\phi(x)$ 参数化为有理式,其形式为两个多项式 $P(x)$ 和 $Q(x)$ 的比值,分别为 m 阶和 n 阶,即

$$\phi(x) = wF(x) = w \frac{P(x)}{Q(x)} = w \frac{a_0 + a_1x + \dots + a_mx^m}{b_0 + b_1x + \dots + b_nx^n} \quad (22)$$

式中: a_m 和 b_n 为有理函数的系数, w 为缩放因子。本文通过端到端的反向传播来学习这些参数。

为了避免由于极点引起的不稳定性(即当 $Q(x) \rightarrow 0$ 时, $\phi(x) \rightarrow \pm\infty$), 本文采用安全的 Padé 激活单元(Padé activation unit, PAU)作为基础,式(23)为标准有理函数的修改版本。

$$F(x) = \frac{a_0 + a_1x + \dots + a_mx^m}{1 + |b_1x + b_2x^2 + \dots + b_nx^n|} \quad (23)$$

本文将式(23)中的有理函数与组内参数相结合,构建了逐组有理函数(Group-wise rational)。在实际应用中,本文为每个组共享有理函数 F 的参数,但每条边仍保留唯一的标量 w ,即

$$\text{Grope-rational}(x) = \Phi \circ x = \left[\sum_{i=1}^{d_{\text{in}}} w_{i,1} F_{\lfloor \frac{i}{d_g} \rfloor} \left(x_i \right) \quad \sum_{i=1}^{d_{\text{in}}} w_{i,2} F_{\lfloor \frac{i}{d_g} \rfloor} \left(x_i \right) \quad \cdots \quad \sum_{i=1}^{d_{\text{in}}} w_{i,d_{\text{out}}} F_{\lfloor \frac{i}{d_g} \rfloor} \left(x_i \right) \right] \quad (24)$$

式中:“ \circ ”表示算子复合运算; i 为输入通道的索引, g 表示组的个数, i/d_g 为组索引, $d_g = d_{\text{in}}/g$ 表示每个组的通道数, d_{in} 为输入的特征维度; d_{out} 为输出的特征维度。

通过简单的重写,可以用矩阵形式表示权重矩阵 $\mathbf{W} \in \mathbb{R}^{d_{\text{in}} \times d_{\text{out}}}$ 与对每个输入通道分别作用的有理函数 F 的矩阵乘积,即

$$\text{Grope-rational}(x) = \mathbf{W}F(x) = \begin{bmatrix} w_{1,1} & \cdots & w_{1,d_{\text{in}}} \\ \vdots & \ddots & \vdots \\ w_{d_{\text{out}},1} & \cdots & w_{d_{\text{out}},d_{\text{in}}} \end{bmatrix} * \begin{bmatrix} F_{\lfloor \frac{1}{d_g} \rfloor}(x_1) & F_{\lfloor \frac{2}{d_g} \rfloor}(x_2) & \cdots & F_{\lfloor \frac{d_{\text{in}}}{d_g} \rfloor}(x_{d_{\text{in}}}) \end{bmatrix}^T \quad (25)$$

因此,本文将该GR-KAN层视为逐组有理函数,在这之后紧接着线性层,即

$$\text{GR-KAN}(x) = \text{Linear}(\text{Group-rational}(x)) \quad (26)$$

在这种形式中,通过在组内共享参数,有理函数能够以逐通道的方式直接作用于输入向量,等效地将其应用于每个分组边缘。以这种方式,GR-KAN作为一个替代MLP网络层,包括可学习的非线性函数、线性层之前的激活函数,以及为每组边缘定制的唯一激活函数。

本文使用两次GR-KAN层替代原来的MLP层,不同分类头对比图如图6所示。GR-KAN采用了自定义的Grove-rational作为激活函数而不是MLP中传统的激活函数ReLU,这样的设计更加灵活,尤其是在处理复杂任务或需要自适应特征提取时,可以通过调整Grove-rational的参数来优化。对比传统的MLP,本文设计的GR-KAN架构更加灵活,适应性更强,能够处理不同类型的数据,具有更强的通用性,可以更有效地提取出全局和局部特征,从而提升分类性能,尤其是在复杂数据集上。

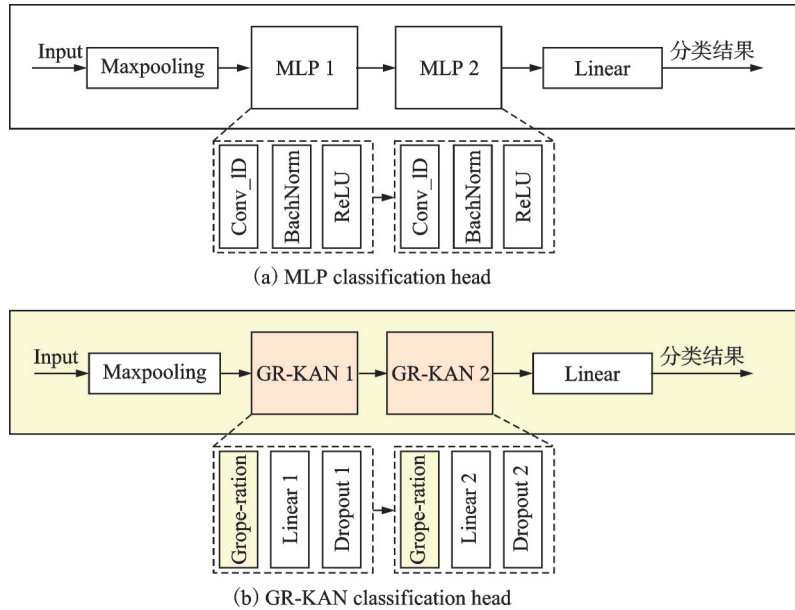


图6 不同分类头对比图

Fig.6 Comparison of different classification heads

2 实验验证

本节首先介绍了硬件配置和超参数设置;其次给出了网络在公共合成和实时扫描数据集 ModelNet40 和 ScanObjectNN 上的分类结果,并与最先进的点云分类工作进行了比较;然后展示了一系列消融研究的结果,以验证框架中每个主要组件的有效性;最后通过热图可视化结果证明了网络的可解释性,表明本文提出的方法能够根据不同的形状特征理解不同的形状。

2.1 实验环境

在设备配置方面,本文所有实验在 1 台搭载 1 块 NVIDIA 4070 Ti Super 显卡(显存 16 GB)的服务器上,操作系统为 Ubuntu 20.04,编程语言为 Python 3.8.19,利用 Pytorch 1.10.1 框架搭建模型,采用 Cuda 12.1、Cudnn 8.9 加速模型训练。在参数设置方面,批次大小为 16;学习率为 0.001,在数据集 ModelNet40 和 ScanObjectNN 上分别训练 300 轮和 400 轮;选用 SGD 优化器。上述设置能够保证网络在收敛后取得较好的分类精度。

2.2 ModelNet40 数据集上的点云实验分类结果分析

ModelNet40 数据集是 CAD 绘制的 3D 模型转换成点云数据,包含 12 308 个点云形状个体,共 40 个类别,其中训练集为 9 840 个点云模型,测试集为 2 468 个点云模型,每个模型都进行了 1 024 个点的采样作为网络的输入。本文采用每个类别内的总体准确度(Overall accuracy, OA)作为评估指标,表达式为

$$OA = \frac{T}{N} \quad (27)$$

式中: T 为所有正确预测的点云的数量, N 为所有点云的数量。

同时采用平均准确度(mean Accuracy, mAcc)作为评估指标,表达式为

$$mAcc = \frac{\sum_{i=1}^K \frac{T_i}{N_i}}{K} \quad (28)$$

式中: T_i 为第 i 类中正确预测的点云数量, $T = \sum_{i=1}^K T_i$; K 表示数据集的种类数目; N_i 为第 i 类中

所有点云的数量, $N = \sum_{i=1}^K N_i$ 。

本文将所提出的方法与最先进的基于 Transformer、CNN 以及其他深度学习的方法进行了比较,结果如表 1 所示。具体而言,在 ModelNet40 数据集上,与使用单注意力机制的 Point-ConvT^[18]相比,本文提出的双注意力机制在 OA 上提升了 0.3%。与同为双注意力机制的 DPCT^[19]相比,本文的双分支架构在 OA 上提升了 0.4%,充分展示了网络架构的优越性。此外,本文方法在 OA 和 mAcc 指标上分别取得了 93.8% 和 91.2% 的最佳成绩,超越了所有基准方法。

表 1 ModelNet40 数据集上的形状分类结果

| Method | Model | Input size | OA/% | mAcc/% |
|----------------------|----------------------------|------------|------|--------|
| Based on CNN | PointWeb ^[3] | 1 024 | 92.3 | 89.4 |
| | PointConv ^[4] | 1 024 | 92.5 | — |
| | RSCNN ^[24] | 1 024 | 92.9 | — |
| | KPConv ^[5] | 1 024 | 92.9 | — |
| | DiffConv ^[25] | 1 024 | 93.1 | 90.4 |
| | FFG ^[26] | 1 024 | 93.3 | 90.7 |
| Based on Transformer | Adapt-Conv ^[10] | 1 024 | 93.4 | 90.7 |
| | PointNet ^[1] | 1 024 | 89.2 | 86.0 |
| | Point-Net++ ^[2] | 1 024 | 91.9 | 88.2 |
| | PointNext ^[27] | 1 024 | 93.2 | 90.6 |
| | PointPN ^[28] | 1 024 | 93.3 | — |
| Ours | PAT ^[29] | 1 024 | 91.7 | — |
| | PointTrans ^[15] | 1 024 | 92.8 | 89.0 |
| | 3DCTN ^[30] | 1 024 | 93.2 | 90.7 |
| | PCT ^[19] | 1 024 | 93.2 | — |
| | DPCT ^[18] | 1 024 | 93.4 | — |
| | PointCont ^[18] | 1 024 | 93.5 | — |
| | Ours | 1 024 | 93.8 | 91.2 |

2.3 ScanObjectNN数据集上的点云实验分类结果分析

为了进一步评估本文提出的网络对LiDAR扫描获取的真实世界数据的性能,在实验中测试了ScanObjectNN数据集的分类性能。扫描大约有15 000个对象,分为15个类别,其中有2 902个唯一对象实例。由于每个物体都是从扫描的场景点云中分割出来的,物体点云通常包含大量背景点形式的离群点,并受到遮挡和噪声的干扰。因此,对该数据集进行形状分类更具挑战性。

图7展示了在ScanObjectNN数据集上生成的混淆矩阵,其中横轴代表点云的预测类别,纵轴代表实际的类别。由于对列进行了归一化处理,矩阵中的单元格数值范围在0~1之间。混淆矩阵的对角线部分表示各类别的召回率,而精确率则可以通过计算该类别召回率与该类别所在行总和的比值来推导得出。

在ScanObjectNN数据集中,运用了与ModelNet40相同的总体准确度与平均准确度作为评价指标,结果如表2所示。对比PointCont^[18]的单注意力网络相比,本文采用的双注意力机制将准确率提升了0.5%,证明了双注意力设计的有效性。相比同样采用双注意力机制的Point-TNT^[36]网络,本文提出的并行架构在准确率上提升了3%,进一步展现了其在泛化能力和鲁棒性上的优势。最终本文方法以86.5%的整体准确率超越了现有所有模型,并在mAcc上以85.0%取得了优异结果,验证了本文所提出的网络在处理真实场景点云时的卓越表现。

2.4 消融实验

本文对形状分类任务中方法的关键设计进行了消融研究,所有实验都是在相同的训练设置下进行。模块消融实验结果如表3所示。通过比较表3中实验1、2和5可以发现,双池化策略相较于单池化的Maxpooling或Avgpooling策略,在ModelNet40数据集上的效果分别提升了0.3%和0.8%,而在ScanObjectNN数据集上则

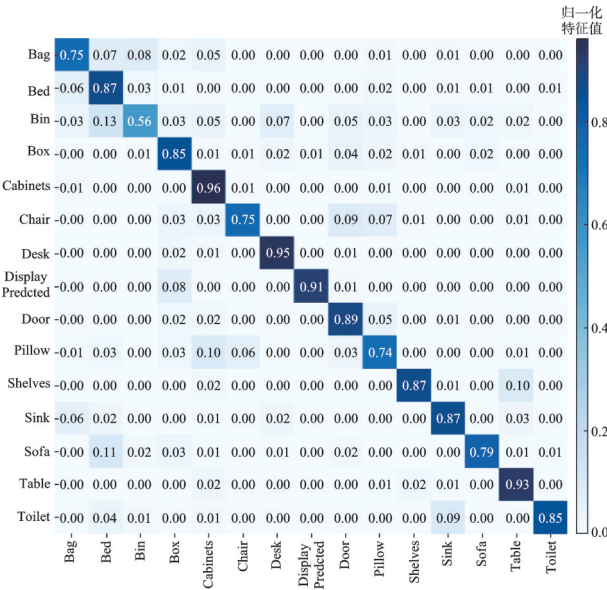


图7 ScanObjectNN数据集混淆矩阵

Fig.7 Confusion matrix on the ScanObjectNN dataset

表2 ScanObjectNN数据集上的形状分类结果

Table 2 Shape classification results on the ScanObjectNN dataset

| Method | Model | Input size | OA/% | mAcc/% |
|----------------------|-----------------------------------|------------|------|--------|
| Based on CNN | DGCNN ^[7] | 1 024 | 78.1 | 73.6 |
| | FFG ^[26] | 1 024 | 80.0 | 77.5 |
| | PRANet ^[31] | 1 024 | 82.1 | 79.1 |
| Others | PointNet ^[11] | 1 024 | 68.2 | 63.4 |
| | PointNet++ ^[2] | 1 024 | 77.9 | 75.4 |
| | PointMLP ^[32] | 1 024 | 85.7 | 84.4 |
| | Point Transformer ^[33] | 1 024 | 77.6 | 75.3 |
| Based on Transformer | GBNet ^[34] | 1 024 | 80.5 | 77.8 |
| | PointBert ^[35] | 1 024 | 83.1 | — |
| | Point-TNT ^[36] | 1 024 | 83.5 | — |
| | PointMAE ^[37] | 1 024 | 85.2 | — |
| | PointCont ^[18] | 1 024 | 86.0 | — |
| | Ours | 1 024 | 86.5 | 85.0 |

信息交互、抑制冗余和噪声,以及优化梯度流动和训练稳定性等方面。通过结合不同的注意力模块,双注意力机制能够更加全面地捕捉局部和全局的特征信息,促进特征之间的有效交互,减少无关噪声,确保梯度在深层网络中的平稳传递,从而提升模型的鲁棒性与整体性能。在实验中,双注意力机制带来了0.3%~0.7%的准确率提升。对比表3中实验1、3和实验2、4进一步验证了本文设计的注意力机制的有效性,无论是在Maxpooling分支还是Avgpooling分支中添加注意力模块,都能显著提升网络性能。对比表3中实验8和9可以看出,本文设计的GR-KAN分类头相较于传统的MLP分类头,在两个数据集上的精度均有所提升,尤其是在ScanObjectNN数据集上,精度提升了0.5%。这说明GR-KAN网络在复杂数据集中对特征的提取能力优于MLP网络。在表3最终的实验9中,本文所设计的网络架构及注意力机制在ModelNet40和ScanObjectNN数据集上都取得了最佳效果。

表 3 不同模块的消融实验结果

Table 3 Ablation study results of different modules

| 实验 序号 | Maxpooling | LWA | Avgpooling | GCCA | MLP 分类头 | GR-KAN 分类头 | ModelNet40 | | ScanObjectNN | |
|----------|------------|-----|------------|------|------------|---------------|------------|--------|--------------|--------|
| | | | | | | | OA/% | mAcc/% | OA/% | mAcc/% |
| 1 | ✓ | | | | | ✓ | 93.0 | 90.8 | 85.7 | 83.9 |
| 2 | | | ✓ | | | ✓ | 92.5 | 89.7 | 80.3 | 76.7 |
| 3 | ✓ | ✓ | | | | ✓ | 93.3 | 90.3 | 85.9 | 84.1 |
| 4 | | | ✓ | ✓ | | ✓ | 92.9 | 90.3 | 81.5 | 78.2 |
| 5 | ✓ | | ✓ | | | ✓ | 93.3 | 90.5 | 86.0 | 84.3 |
| 6 | ✓ | ✓ | ✓ | | | ✓ | 93.5 | 91.0 | 86.3 | 84.5 |
| 7 | ✓ | | ✓ | ✓ | | ✓ | 93.1 | 90.6 | 86.2 | 84.3 |
| 8 | ✓ | ✓ | ✓ | ✓ | ✓ | | 93.5 | 90.9 | 86.0 | 84.1 |
| 9 | ✓ | ✓ | ✓ | ✓ | | ✓ | 93.8 | 91.2 | 86.5 | 85.0 |

本文在添加所有模块的情况下,做了不同尺度的消融实验,结果如表4所示。由表4可以发现,在除尺度不同、其他条件完全相同的情况下,多尺度KNN比单尺度KNN的OA指标在ModelNet40数据集上提高了0.9%,在ScanObjectNN数据集上提高1.1%。这可以证明,本文提出的多尺度提取特征可以有效地保留局部信息和全局信息。

本文针对多尺度不同参数做了消融实验,以选择多尺度策略的最佳参数设置,结果如表5所示。由表5可以看出参数设置为2时,在ModelNet40数据集上实现了最高的分类精度。由于多尺度策略的目的是聚集局部特征,大尺度邻域往往包含许多无关点,降低了分类性能。因此,本文选择2设置作为多尺度策略的最佳设置。

表 4 不同尺度间的消融实验结果

Table 4 Ablation study result of different scales

| 实验 序号 | 单尺度 | 多尺度 | ModelNet40 | | ScanObjectNN | |
|----------|-----|-----|------------|------------|--------------|------------|
| | | | OA/ % | mAcc/ % | OA/ % | mAcc/ % |
| 1 | ✓ | | 92.9 | 90.2 | 85.4 | 83.8 |
| 2 | | ✓ | 93.8 | 91.2 | 86.5 | 85.0 |

表 5 多尺度策略的不同参数设置的ModelNet40数据集分类结果

Table 5 Classification results of different parameter settings for the multi-scale strategy on ModelNet40 dataset

| 参数 设置 | 阶段1 | | 阶段2 | | 阶段3 | | ModelNet40 | |
|----------|---------------|------------|---------------|-------------|---------------|--------------|------------|--------|
| | 尺度 | K 邻近 | 尺度 | K 邻近 | 尺度 | K 邻近 | OA/% | mAcc/% |
| 1 | [0.1,0.2,0.4] | [4,8,16] | [0.2,0.4,0.8] | [8,16,32] | [0.4,0.8,1.6] | [16,32,64] | 93.4 | 90.5 |
| 2 | [0.1,0.2,0.4] | [8,16,32] | [0.2,0.4,0.8] | [16,32,64] | [0.4,0.8,1.6] | [32,64,128] | 93.8 | 91.2 |
| 3 | [0.1,0.2,0.4] | [16,32,64] | [0.2,0.4,0.8] | [32,64,128] | [0.4,0.8,1.6] | [64,128,256] | 93.3 | 90.3 |

表6展示了不同阶段数下的消融实验结果。为了评估更深网络的有效性,本文在ModelNet40和ScanObjectNN数据集上逐步增加了网络的深度。从实验结果可以看出,使用3个阶段数便可实现最佳性能,而进一步增加阶段数会引入冗余信息,反而导致性能下降。从参数量的角度来看,增加阶段数会显著提升参数量,对复杂度较低的数据集而言,使用较少的阶段数即可满足需求。而在处理复杂度较高的数据集时,较浅的网络难以有效提取特征,因此适当增加网络深度是必要的。因此,本文综合考虑参数量以及准确度,最终选择了3个阶段的网络结构。

2.5 可视化实验验证

为了进一步说明本文提出的注意力机制如何工作,并直观地理解模型如何关注点云中的不同特征,本文对应用注意力前后的特征分布进行了可视化,如图8、9所示。图中颜色由深至浅表示注意力权重的大小,颜色越深代表注意力关注度越高,颜色越浅则代表关注度越低。

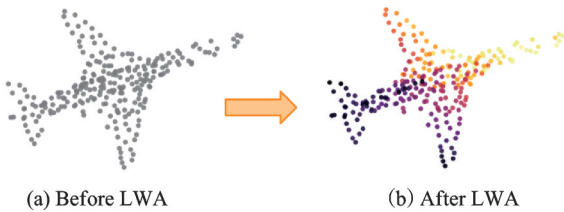


图8 局部窗口注意力前后对比图
Fig.8 Comparison before and after LWA

表6 不同阶段数的消融实验结果
Table 6 Ablation study results on different numbers of stages

| 阶段数 | 参数量/ 10 ⁶ | ModelNet40 | | ScanObjectNN | |
|-----|-------------------------|------------|--------|--------------|--------|
| | | OA/% | mAcc/% | OA/% | mAcc/% |
| 2 | 4.99 | 93.1 | 89.7 | 84.6 | 82.8 |
| 3 | 9.24 | 93.8 | 91.2 | 86.5 | 85.0 |
| 4 | 18.14 | 92.9 | 90.4 | 86.8 | 85.3 |

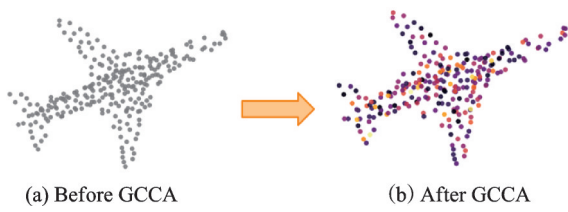


图9 全局通道内容注意力前后对比图
Fig.9 Comparison before and after GCCA

在可视化图像中,注意力分布可视化图清楚地显示了模型在关注不同区域的权重变化。对于最大池化分支,可以看到高权重区域集中在具有显著结构的边缘,而平均池化分支的权重则相对均匀地分布在几何细节区域。

在双分支结构中,特征提取分别通过平均池化和最大池化操作,以更全面地捕获点云数据的全局和局部信息。具体而言,平均池化通过均匀分布特征值,能够平滑地提取点云整体的几何形状和结构特征。这种方法对全局信息进行综合,使模型对整体轮廓、形态以及分布有更深入的理解,从而有助于捕捉点云的宏观结构模式。另一方面,最大池化则专注于提取局部显著特征,将点云中最重要几何细节,例如边缘和角落,进行强化。这种方法可以突出局部关键信息,使模型更敏锐地检测和关注复杂局部区域或重要几何细节。通过结合平均池化和最大池化,这种双重注意力机制为模型提供了一个多尺度特征表示,能够平衡全局和局部特征的表达,增强模型在识别点云细微差异与整体模式时的表现和适应性。

本文采用传统MLP分类头和GR-KAN分类头进行可视化实验。为了直观地比较这些模型的特征表示,使用t-分布式随机邻域嵌入(t-Stochastic neighbor embedding, t-SNE)将高维编码特征投影到二维平面上,这有助于直接比较它们的分布特性。本文在ModelNet40和ScanObjectNN两个数据集进行可视化的对比实验,结果如图10、11所示,t-SNE可视化展示了使用GR-KAN和MLP分类头的两种模型在特征表示上的显著差异。在ModelNet40数据集中,GR-KAN分类头使同类别的样本在二维空间

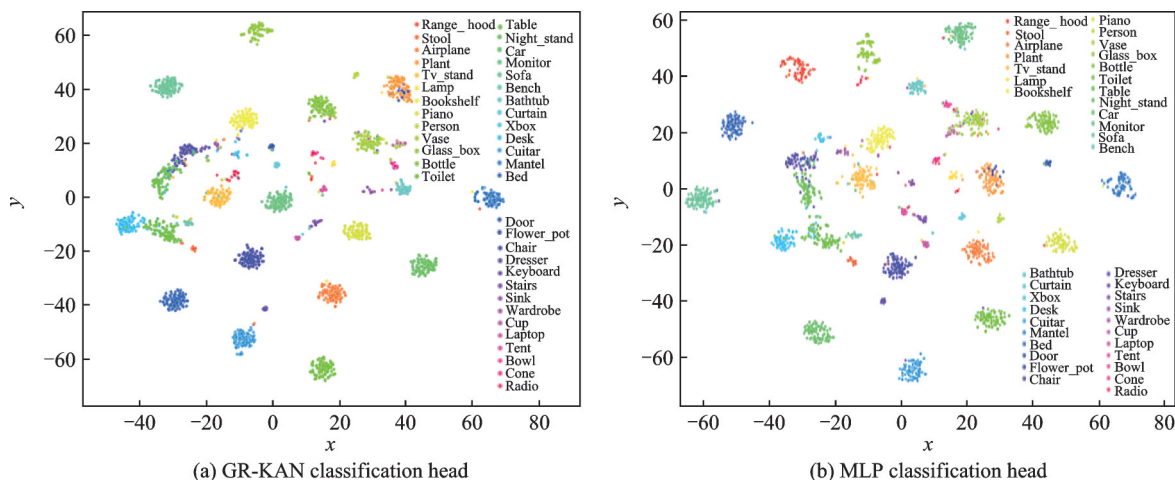


图 10 使用 MLP 分类头与 GR-KAN 分类头在 ModelNet40 数据集中编码器功能可视化

Fig.10 Visualization of encoder functions using MLP and GR-KAN classification heads on the ModelNet40 dataset

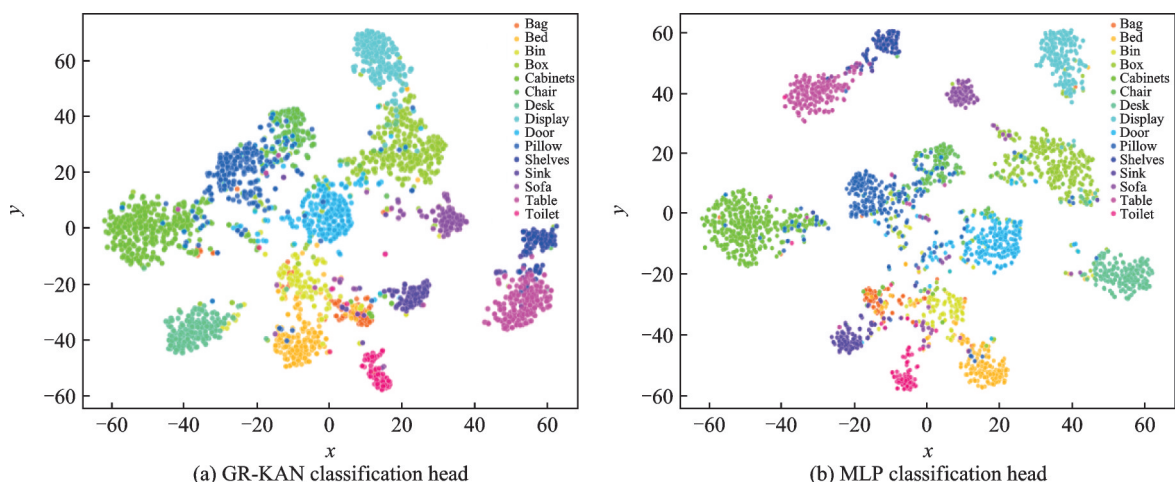


图 11 使用 MLP 分类头与 GR-KAN 分类头在 ScanObjectNN 数据集中编码器功能可视化

Fig.11 Visualization of encoder functions using MLP and GR-KAN classification heads on the ScanObjectNN dataset

中聚集得更加紧密,尤其是样本量较大的类别,其特征分布呈现更紧密的圆形聚类,这表明 GR-KAN 在捕捉类内相似性和增强特征一致性方面具有更强的优势。在 ScanObjectNN 数据集中,GR-KAN 分类头的特征聚类分布之间的重叠区域较少,各类特征分布之间的界限更加清晰,显示出其在复杂数据集上对类别间区分能力的提升。相比之下,传统的 MLP 分类头在特征分布中存在较多的重叠区域,并且聚类结构较为松散,难以有效区分某些类别。实验结果表明,GR-KAN 分类头在点云分类任务中表现出显著优势。通过有效融合局部点特征与全局集合特征,GR-KAN 分类头不仅增强了模型对点云数据的特征表达能力,还提升了类区分的清晰度。该融合方法使模型能够捕捉点云数据中的细微特征,从而在特征空间中形成更结构化的分布,最终实现更高的分类精度。

每个阶段的聚类结果可视化如图 12 所示。图 12 可视化结果表明这些聚类能够覆盖长期依赖关系。图中每个点的颜色深度代表每个点经过一系列变换后的数值。颜色较深的点表示特征值比较大,颜色较浅的值表示特征值比较小。

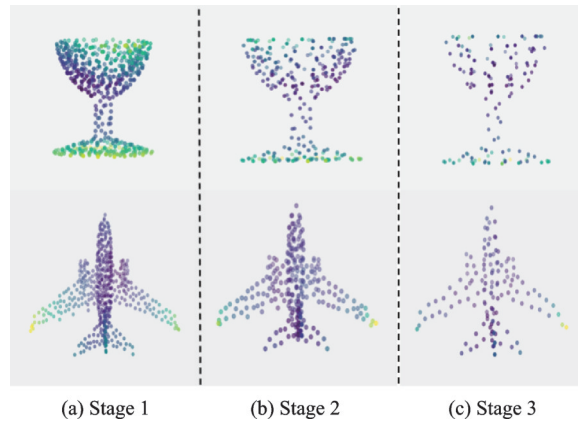


图 12 每个阶段的聚类结果可视化

Fig.12 Visualization of clustering results at each stage

3 结束语

本文提出了一种多尺度双分支双注意力网络用于点云分类。通过引入多尺度KNN的特征提取方法,本文设计的编码器能够有效地保留局部细节和全局信息。在信息提取过程中,本文采用了全局最大池化和全局平均池化的双分支操作,分别捕提高频和低频特征。高频特征的提取通过LWA模块进行,该模块利用多头注意力机制,有效捕捉各头部的特征差异,从而可以提取丰富的细节信息。低频特征的提取则依赖于GCCA模块,通过点积运算实现全局特征交互,捕捉低频信息。这种双注意力机制互相补充,避免了单一特征提取方式的局限。在分类头中,本文引入了GR-KAN层替代传统的MLP层。GR-KAN不仅保留了局部特征之间的依赖性,还能够有效整合全局上下文信息,进一步提升了分类性能。大量实验证明,本文提出的方法在ModelNet40和ScanObjectNN等合成数据和真实场景LiDAR数据集上表现出色,验证了该方法在点云分类任务中的优越性和实用性。

本文主要探讨了双分支注意力机制在点云特征提取中的有效性,但在计算效率、内存占用和推理速度方面的分析仍显不足。未来的工作将关注优化模型的计算效率,特别是在资源有限的情况下,对于一些不太复杂的数据集,可以考虑降低网络的复杂度,力求在保持精度的同时提升实用性。同时考虑改进网络架构利用跨域的方式^[11,38],在不同领域间有效地对齐和共享特征,提高3D模型检索的鲁棒性和准确性。

参考文献:

- [1] CHARLES R Q, HAO S, MO K, et al. PointNet: Deep learning on point sets for 3D classification and segmentation[C]//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, HI, USA: IEEE, 2017: 77-85.
- [2] QI C R, YI L, SU H, et al. PointNet++: Deep hierarchical feature learning on point sets in a metric space[EB/OL]. (2017-06-10). <https://arxiv.org/abs/1706.02413>.
- [3] ZHAO H, JIANG L, FU C W, et al. PointWeb: Enhancing local neighborhood features for point cloud processing[C]//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, CA, USA: IEEE, 2020: 5560-5568.
- [4] WU W, QI Z, LI F. PointConv: Deep convolutional networks on 3D point clouds[C]//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, CA, USA: IEEE, 2020: 9613-9622.
- [5] THOMAS H, QI C R, DESCHAUD J E, et al. KPConv: Flexible and deformable convolution for point clouds[C]//

- Proceedings of 2019 IEEE/CVF International Conference on Computer Vision (ICCV). Seoul, Korea: IEEE, 2019: 6410-6419.
- [6] XU M, DING R, ZHAO H, et al. PAConv: Position adaptive convolution with dynamic kernel assembling on point clouds [C]//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville, TN, USA: IEEE, 2021: 3172-3181.
- [7] PHAN A V, LE NGUYEN M, NGUYEN Y L H, et al. DGCNN: A convolutional neural network over large-scale labeled graphs[J]. Neural Networks, 2018, 108: 533-543.
- [8] CHO H, CHOI I S. Three-dimensionally embedded graph convolutional network (3DGCN) for molecule interpretation[EB/OL]. (2018-11-11). <https://arxiv.org/abs/1811.09794>.
- [9] LI G, XIONG C, THABET A, et al. DeeperGCN: All you need to train deeper GCNs[EB/OL]. (2020-06-30). <https://arxiv.org/abs/2006.07739>.
- [10] ZHOU H, FENG Y, FANG M, et al. Adaptive graph convolution for point cloud analysis[C]//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal, QC, Canada: IEEE, 2022: 4945-4954.
- [11] CHANG R, MA Y, HAO T, et al. 3D shape knowledge graph for cross-domain 3D shape retrieval[J]. CAAI Transactions on Intelligence Technology, 2024, 9(5): 1199-1216.
- [12] THIESHANTHAN A, NIWARTHANA A, SOMARATHNE P, et al. HPGNN: Using hierarchical graph neural networks for outdoor point cloud processing[C]//Proceedings of 2022 26th International Conference on Pattern Recognition (ICPR). Montreal, QC, Canada: IEEE, 2022: 2700-2706.
- [13] YU H, LI F, SALEH M, et al. CofNet: Reliable coarse-to-fine correspondences for robust point cloud registration[J]. Advances in Neural Information Processing Systems, 2021, 34: 23872-23884.
- [14] SHI W, RAJKUMAR R. Point-GNN: Graph neural network for 3D object detection in a point cloud[C]//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, WA, USA: IEEE, 2020: 1708-1716.
- [15] ZHAO H, JIANG L, JIA J, et al. Point transformer[C]//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal, QC, Canada: IEEE, 2022: 16239-16248.
- [16] GUO M H, CAI J X, LIU Z N, et al. PCT: Point cloud transformer[J]. Computational Visual Media, 2021, 7(2): 187-199.
- [17] ZHONG Q, HAN X F. Point cloud learning with transformer[EB/OL]. (2021-04-01). <https://arxiv.org/abs/2104.13636>.
- [18] LIU Y, TIAN B, LV Y, et al. Point cloud classification using content-based transformer via clustering in feature space[J]. IEEE/CAA Journal of Automatica Sinica, 2024, 11(1): 231-239.
- [19] HAN X F, JIN Y F, CHENG H X, et al. Dual transformer for point cloud analysis[J]. IEEE Transactions on Multimedia, 2023, 25: 5638-5648.
- [20] ZHAO Y, NIE W, NIE J, et al. RJAN: Region-based joint attention network for 3D shape recognition[J]. CAAI Transactions on Intelligence Technology, 2025, 10(2): 460-473.
- [21] ZHAO Y, ZHAO Y, NIE W, et al. HMTN: Hierarchical multi-scale transformer network for 3D shape recognition[C]//Proceedings of the 30th ACM International Conference on Multimedia. Lisboa, Portugal: ACM, 2022: 316-324.
- [22] LIU Z, WANG Y, VAIDYA S, et al. KAN: Kolmogorov-Arnold networks[EB/OL]. (2024-04-15). <https://arxiv.org/abs/2404.19756>.
- [23] YANG X, WANG X. Kolmogorov-Arnold transformer[EB/OL]. (2024-09-10). <https://arxiv.org/abs/2409.10594>.
- [24] LIU Y, FAN B, XIANG S, et al. Relation-shape convolutional neural network for point cloud analysis[C]//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, CA, USA: IEEE, 2020: 8887-8896.
- [25] LIN M, FERAGEN A. DiffConv: Analyzing irregular point clouds with an irregular view[C]//Proceedings of Computer Vision—ECCV 2022. Cham: Springer, 2022: 380-397.
- [26] 陶志勇, 豆淼森, 李衡, 等. 融合细粒度特征编码的点云分类分割网络[J]. 数据采集与处理, 2024, 39(4): 944-953.
TAO Zhiyong, DOU Miaosen, LI Heng, et al. Fusion fine-grained feature encoding for point cloud classification and segmentation[J]. Journal of Data Acquisition and Processing, 2024, 39(4): 944-953.
- [27] QIAN G, LI Y, PENG H, et al. PointNeXt: Revisiting PointNet++ with improved training and scaling strategies[EB/OL].

(2022-06-28). <https://arxiv.org/abs/2206.04670>.

- [28] ZHANG R, WANG L, WANG Y, et al. Starting from non-parametric networks for 3D point cloud analysis[C]//Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Vancouver, BC, Canada: IEEE, 2023: 5344-5353.
- [29] YANG J, ZHANG Q, NI B, et al. Modeling point clouds with self-attention and gumbel subset sampling[C]//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, CA, USA: IEEE, 2020: 3318-3327.
- [30] LU D, XIE Q, GAO K, et al. 3DCTN: 3D convolution-transformer network for point cloud classification[J]. IEEE Transactions on Intelligent Transportation Systems, 2022, 23(12): 24854-24865.
- [31] CHENG S, CHEN X, HE X, et al. PRA-Net: Point relation-aware network for 3D point cloud analysis[J]. IEEE Transactions on Image Processing, 2021, 30: 4436-4448.
- [32] MA X, QIN C, YOU H, et al. Rethinking network design and local geometry in point cloud: A simple residual MLP framework[EB/OL]. (2022-02-25). <https://arxiv.org/abs/2202.07123>.
- [33] ENGEL N, BELAGIANNIS V, DIETMAYER K. Point transformer[J]. IEEE Access, 2021, 9: 134826-134840.
- [34] QIU S, ANWAR S, BARNES N. Geometric back-projection network for point cloud classification[J]. IEEE Transactions on Multimedia, 2022, 24: 1943-1955.
- [35] YU X, TANG L, RAO Y, et al. Point-BERT: Pre-training 3D point cloud transformers with masked point modeling[C]//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans, LA, USA: IEEE, 2022: 19291-19300.
- [36] BERG A, OSKARSSON M, O'CONNOR M. Points to patches: Enabling the use of self-attention for 3D shape recognition [C]//Proceedings of 2022 26th International Conference on Pattern Recognition (ICPR). Montreal, QC, Canada: IEEE, 2022: 528-534.
- [37] PANG Y, WANG W, TAY F E H, et al. Masked autoencoders for point cloud self-supervised learning[C]//Proceedings of Computer Vision—ECCV 2022. Cham: Springer, 2022: 604-621.
- [38] LI W, ZHOU H, ZHANG C, et al. Dual-stage uncertainty modeling for unsupervised cross-domain 3D model retrieval[J]. IEEE Transactions on Multimedia, 2024, 26: 8996-9007.

作者简介:



顾君豪(2001-),男,硕士研究生,研究方向:深度学习、3D点云分类、图像处理等, E-mail:1078373482@qq.com。



张孙杰(1988-),通信作者,男,博士,副教授,研究方向:智能图像处理、模糊系统控制与滤波等, E-mail: zhang_sunjie@126.com。



秦辰栋(2000-),男,硕士研究生,研究方向:医学图像分割。

(编辑:张黄群)