

刘诚辉, 李光平. 基于可分离Transformer的点云分类方法[J]. 广东工业大学学报, 2025, 42(2): 29–36. doi: 10.12052/gdutxb.240003.
Liu Chenghui, Li Guangping. Point cloud classification based on separable transformer[J]. Journal of Guangdong University of Technology, 2025, 42(2): 29–36. doi: 10.12052/gdutxb.240003.

基于可分离Transformer的点云分类方法

刘诚辉, 李光平

(广东工业大学 信息工程学院, 广东 广州 510006)

摘要: Transformer往往利用捕获远程依赖项的优势来提取点云远程点的关系交互, 忽略重要的局部结构细节, 并且依赖于大量的计算成本来实现高性能, 导致计算负担增加。为了缓解这个问题, 本文借鉴可分离视觉Transformer的思想, 提出了一种可分离的Transformer点云分类方法, 简称Sep-point。Sep-point通过深度可分离的自注意, 帮助在点云组内和组间按顺序进行局部-全局关系交互。采用新的位置令牌嵌入和分组自注意方法, 分别以可忽略的代价计算组间的注意关系, 并建立跨多个区域的远程信息交互。提取局部-全局特征的同时, 大大减少了计算负担。实验结果表明提出的Sep-point在数据集ModelNet40上比现有的PCT(Point Cloud Transformer)分类精度提升了0.2%, 在真实数据集ScanObjectNN上的分类精度提升了6.3%。同时网络参数量和FLOPS指标分别降低了0.72M和0.18G, 充分验证了本文方法的有效性。

关键词: 点云分类; 可分离Transformer; 位置令牌嵌入; 局部-全局关系交互

中图分类号: TP302.7

文献标志码: A

文章编号: 1007-7162(2025)02-0029-08

Point Cloud Classification Based on Separable Transformer

Liu Chenghui, Li Guangping

(School of Information Engineering, Guangdong University of Technology, Guangzhou 510006, China)

Abstract: Transformer tends to take advantage of capturing remote dependencies to extract relational interactions at remote points of the point cloud, ignoring important local structural details, and achieves high performance by significantly increasing the computational burden. To alleviate this problem, we propose a separable Transformer point cloud classification method, named Sep-point, based on the idea of separable visual Transformer. The proposed Sep-point facilitates sequential local-global relational interactions within and between groups of point clouds through depth-separable self-attention. New location token embedding and group self-attention methods are used to compute inter-group attentional relationships with negligible computational cost and to establish telematic interactions across multiple regions, respectively. In this way, the local-global features are extracted while the computational burden is significantly reduced. Experimental results show that the proposed Sep-point improves the classification accuracy by 0.2% on the ModelNet40 dataset over the existing PCT (Point Cloud Transformer) and by 6.3% on the real ScanObjectNN dataset, respectively. Moreover, the number of network parameters and FLOPS metrics are reduced by 0.72M and 0.18G, respectively. These experimental results clearly demonstrate the promising effectiveness of our proposed method.

Key words: point cloud classification; separable Transformer; location token embedding; local-global relational interactions

点云学习作为一种经典的处理技术, 在各领域受到广泛的关注。在自动驾驶领域PointRCNN^[1]提出了用于原始点云三维目标检测, 在机器人领域提

出了一种基于编码器-解码器配置中的复杂循环或卷积神经网络^[2]。此外, 它还被先后应用于各种任务, 并取得相应进展。在点云识别中提出了一种基于

收稿日期: 2024-01-03 录用日期: 2024-03-15 网络首发日期: 2024-08-30

基金项目: 国家自然科学基金资助项目(61601130); 大亚湾科技规划项目(2020010203)

作者简介: 刘诚辉(1997-), 男, 硕士研究生, 主要研究方向为3D点云处理、特征提取、点云分类, E-mail: 1152278768@qq.com

通信作者: 李光平(1981-), 男, 副教授, 博士, 主要研究方向为计算机视觉、人工智能、雷达信号处理, E-mail: gpli@gdut.edu.cn

球对称性的通用、灵活的三维旋转不变框架^[3];在零件分割中提出了一种采用光谱卷积神经网络(Convolutional Neural Networks, CNN)的方法来预测顶点函数^[4];在场景分割中提出了一种能够有效提取点云局部结构信息的“坐标卷积”,以解决传统卷积神经网络结构在三维点云上的不适用性问题^[5];在目标检测中提出了一种新颖且强大的基于点云的3D目标检测框架^[6];在单目图像中提出了一个简单有效的框架,能够降低注释成本或显著提高检测精度,而不会引入额外的注释成本^[7];在点云配准中提出了一种快速且可认证的算法,用于在存在大量异常对应关系的情况下配准两组三维^[8]。与具有均匀像素网络的二维图像不同,三维(3D)点云具有排列恒定、空间不规则和密度变化的特点。因此,如何快速准确地获取三维点云的语义信息就变得更加具有挑战性。

Transformer作为自然语言处理的主流框架,在图像视觉任务中得到了广泛的应用。由于Transformer在捕获远程依赖项方面表现出优异的性能,因此它通常用于提取各种点云学习任务中远程点的关系交互。例如PCT(Point Cloud Transformer)^[9]和Point Transformer^[10]利用Transformer模块作为骨干网络,专注于点云的全局特征学习。由于Transformer机制学习的是整个欧几里得空间中所有点之间的相互关系,因此无法捕捉到特定的局部细粒度结构,这间接表明了仅使用该机制提取点云特征的局限性。此外,Transformer机制的计算成本较高,如何最大限度地提高其使用效率至关重要。

为了解决以上问题,本文借鉴图像视觉领域SepVit^[11]的深度可分离自注意的思想,将其运用到点云分类领域,提出了一种Sep-point方法。该方法重新设计了一个自注意模块,它由组内自注意(Local Self-Attention, LSA)和组间自注意(Global Self-Attention, GSA)模块组成,其中组内自注意用于捕获每个分组内的局部结构关系,组间自注意用于建立远程点之间的依赖关系,显著提高了表达能力,同时几乎可以忽略组间自注意的计算成本。此外,本文还提出了一种全新获取组间代表点的方法,通过嵌入特殊的位置令牌和分组自注意去获取极具代表性的初始组间信息,使特征信息挖掘更加全面,进一步提升网络性能。

1 相关工作

现有的点云表示方法有两种:基于体素的模型

和基于点的模型。基于体素的模型通常对规则网格的点云进行栅格化处理;基于点的方法则直接处理三维空间中不规则分布的原始点,大致分为基于卷积的方法和基于Transformer的方法。

1.1 基于体素的方法

基于体素的模型通常对规则网格的点云进行栅格化,并应用3D卷积进行特征学习^[12]。由于具有良好的内存局部性,这些模型具有较高的计算效率。但它们不仅丢失了局部细粒度的形状信息^[13],而且需要大量的内存和计算成本。而基于点的模型直接对点云进行操作,既保留了原始点云的语义信息,又能更直观地提取出细粒度结构。

1.2 基于卷积的方法

基于卷积的方法直接处理不规则的点云,旨在设计复杂的局部特征提取器。PointNet^[14]是直接针对稀疏、非结构化的点云,利用多层感知机(Multilayer Perceptron, MLP)和最大池化操作对全局信息进行汇总的开创性工作。PointNet++^[15]采用分组策略和带max-pooling的分层结构捕获局部信息,弥补了PointNet只关注全局信息的缺点。然而,PointNet++只考虑了局部特征聚合,忽略了另一个重要的区域间信息。在后续的研究中,重点设计复杂的局部特征聚合模块,以充分利用局部细粒度特征。PointCNN^[16]将相邻点转换为规范阶,使传统的卷积能够正常工作。Point2Sequence^[17]利用注意力机制聚合来自不同局部区域的信息。为了提取更丰富的局部细节,许多方法利用邻域与局部中心点的关系来学习卷积的动态权重。DGCNN^[18]提出了EdgeConv模块来生成边缘特征并搜索邻近的特征空间。PointConv^[19]将两个点的相对位置投影到卷积权值。KPConv^[20]和PAConv^[21]通过设计复杂卷积核自适应学习卷积权值,实现了比2D卷积更大的灵活性。这些工作虽然效果显著,但也带来了复杂的卷积计算和更多的内存消耗。

1.3 基于Transformer的方法

受自然语言处理(Natural Language Processing, NLP)和成像领域的启发,研究人员也在点云的学习架构中使用了Transformer,旨在建立跨区域间的相互联系。Mobilenets^[22]通过引入标准集合抽象层(Set Abstraction layer, SA)机制提取全局特征。PCT^[9]提出了Offset-Attention模块,通过元素减法计算集合抽象层特征与输入特征之间的偏移差,学习点云的全局特征作为主要学习模型,有一定效果,但代价是参数和计算复杂度的增大。局部云嵌入器^[23]提出了在

邻近点集中构造局部向量注意力的方法,并取得了显著进展。但现有的方法都存在忽略局部细节的同时增加计算负担的缺陷。

2 Sep-point

在本节中,首先说明Sep-point的设计概述,然后讨论Sep-point中的一些关键模块。最后,介绍由Sep-point作为单元模块构成的点云分类网络模型。

2.1 Sep-point模块

在以往的研究中,往往采用两个连续分离的模

块来分别完成点云的局部-全局建模,借鉴图像视觉领域 SepVit 的深度可分离自注意的思想,将其运用到点云分类领域,开发出能一步完成局部-全局建模的Sep-point 模块,模块框架如图1所示。局部上下文概念和全局抽象都可以在单个 Sep-point 块中捕获,大大提升网络建模能力的同时较大程度地减少计算复杂度。值得注意的是 SepVit 方法无法直接运用到点云领域,本文在其思想基础上加以创新,成功开发出适用于点云局部-全局关系建模的可分离自注意模块Sep-point。

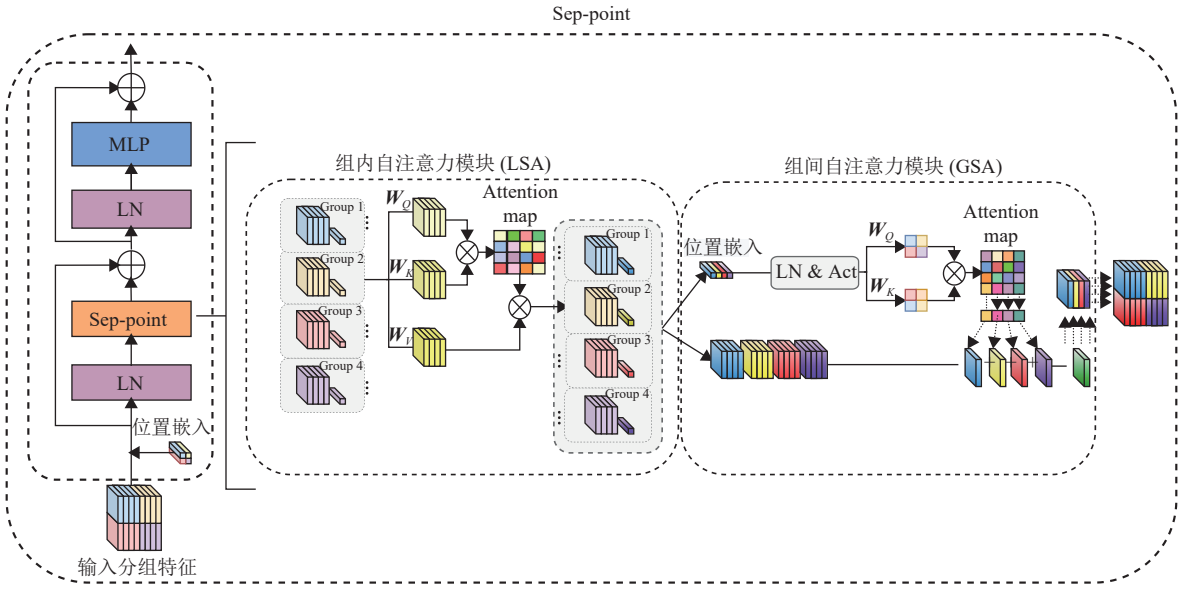


图1 Sep-point:深度可分离自注意力模块

Fig.1 Sep-point: deeply separable self-attention module

在 Sep-point 模块中,每个特征分组内部的局部信息通信通过组内自注意力(LSA)实现,特征分组之间的全局信息交换通过组间自注意力(GSA)实现。具体地,采用最远点采样(Furthest Point Sampling, FPS)获取局部分组中心点坐标信息 $\omega \in \mathbf{R}^{N \times c}$ 以及利用KNN分组方法对每个中心点邻域进行分组,并将每个邻域点与中心点之间的欧氏距离作为初始化边缘特征 $F_l \in \mathbf{R}^{N \times d \times s}$,其中 N 表示采样点个数、 d 和 c 表示特征维度、 s 表示每个分组的点个数,具体的分组实现过程请参考文献[9]。这种利用邻域点与中心点之间的欧氏距离代替传统原始点特征的方法,有利于捕捉分组内部特征点之间的边缘细节。首先将初始化特征及位置令牌进行拼接操作,形成融合特征向量,具体如下:

$$F_l^1 = \text{Concat}(F_l, \omega) \quad (1)$$

式中:Concat表示特征拼接操作。

随后利用设计的LSA模块在每个分组内部执行自注意操作,建模出每个局部内核的注意力权重矩阵,从而捕捉点云每个局部分组点的细节相关信息,弥补注意力机制只关注全局语义的不足。

$$F_l^2 = \text{LSA}(\text{LN}(F_l^1)) \quad (2)$$

式中:LN表示层归一化(Layer Normalization)、LSA表示特征进入LSA模块。

随后利用切片操作分离出更新后的分组特征以及新的位置令牌,此时的特征输出包含丰富的组内关系交互信息,每个局部令牌对应一个分组,蕴含了相应内核的特征代表点。

$$F_l^3, \omega^1 = \text{Slice}(F_l^2) \quad (3)$$

式中:Slice表示切片操作。

GSA模块的目的是执行全局自注意,旨在建模特征的全局轮廓,与传统自注意机制针对全局特征建模远程关系不同的是,GSA模块利用新的位置令牌形成注意力权重矩阵,减少了较多的计算负担的同时捕捉丰富的语义信息。此外,本文设计了一个残差连接用来减少计算过程中的信息损失。

$$F_l^4 = \text{GSA}(F_l^3, \omega^1) + F_l^3 \quad (4)$$

式中:GSA表示特征进入GSA模块。

最后,利用一个MLP层来更新特征向量,目的是将隐性特征转变为显性表达。

$$F_a = \text{MLP}(\text{LN}(F_l^4)) + F_l^4 \quad (5)$$

2.2 LSA模块

Sep-point是建立在图像视觉中基于窗口的自注意方案之上的。首先,对输入特征映射进行局部分组。每个独立分组都可以看作是特征向量的一个输入通道,不同的分组包含不同的信息。与以往的工作不同,本文为每个分组创建了一个位置嵌入令牌,作为一个全局表示,并用于在组内自注意力模块(LSA)中来逐点建模注意关系。然后,对每个组内的所有逐点令牌及其对应的位置嵌入令牌执行深度自注意。这种基于分组的操作非常类似于深度逐点卷积层,旨在融合每个分组通道内的空间信息。此外,为了防止信息丢失,LSA采用了残差连接。假设由逐点特征和位置特征组成的融合特征为 $F_l^1 \in \mathbf{R}^{N \times d \times s}$ 、 $W_Q, W_K, W_V \in \mathbf{R}^{d \times d}$ 是自注意权重矩阵,用于常规自注意中的查询、键和值计算。LSA的学习过程可以概括如下:

$$Q, K, V = F_l^1(W_Q, W_K, W_V) \quad (6)$$

$$\text{LSA} = \text{Attention}(Q, K, V) + F_l^1 \quad (7)$$

式中: $Q, K, V \in \mathbf{R}^{N \times d \times s}$ 分别表示注意力向量中的查询、键、值矩阵, $\text{LSA} \in \mathbf{R}^{N \times d \times s}$ 为每组的组内自注意输出特征, Attention 是一个标准的自注意算子,表示为

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{d_k}\right)V \quad (8)$$

式中: $d_k = \sqrt{d}$ 表示比例因子,作用于每个局部分组。softmax是注意力机制中标准的归一化函数,用于归一化自注意权重矩阵,具体表达式为

$$\text{softmax}(a_{i,j}) = \frac{\exp(a_{i,j})}{\sum_k \exp(a_{i,k})} \quad (9)$$

2.3 位置令牌嵌入

为了建立远程分组之间的注意力关系,一个简

单的解决方案是使用每个分组的所有点标记。然而,这将带来巨大的计算成本,并导致整个模型变得复杂。为了缓解这个问题,本文提出了一种相对位置令牌嵌入方案,该方案利用单个位置令牌封装每个组内的核心信息,每个局部分组都有一个对应的令牌。每个令牌初始化为相应中心点位置信息。值得注意的是,由于点云数据的稀疏性和不确定性,本文没有采用SepVit中随机生成可学习的位置令牌的方式,而是将中心点作为初始的位置令牌。因为每个局部分组的中心点包含了一定的原始点云准确的位置信息,具有较强的代表性,有效地防止了随机生成所带来的不确定性和干扰信息。在通过LSA时,每个分组内的位置令牌和组内各点之间存在信息交互。因此,每个局部点与令牌之间的相对位置可以学习该分组的全局表示,并通过切片操作更新出包含每个分组特征的位置令牌用于接下来的GSA模块中新的位置令牌。由于高效的位置令牌,本文在计算成本可以忽略不计的情况下建立远程分组间的注意关系。

2.4 GSA模块

模仿点向卷积融合来自不同信道信息的方法来开发组间自注意力(GSA)模块,并建立远程点之间的相互联系,以获得远程依赖关系。GSA主要用于融合跨区域的信息,并获得输入特征图的最终表示。

更具体地说,从LSA的输出中提取特征映射和位置标记。然后,使用位置令牌计算远程分组之间的注意关系,并经过LN层和GELU激活函数Act生成注意图。同时,直接将特征映射作为GSA的值分支,不做任何额外的操作。

利用以点云局部分组形式呈现的注意力图和特征图,在分组之间进行注意力计算,实现全局信息交换。同样,GSA也采用残差连接。GSA的正式实施可以描述如下:

$$\text{GSA}(\lambda, \omega^1) = \text{Attention}(\text{Act}(\text{LN}(\omega^1))W_Q, \text{Act}(\text{LN}(\omega^1))W_K, \lambda) + \lambda \quad (10)$$

式中: ω^1 表示位置令牌。这里, Attention 也是一个标准的自注意操作符,但它适用于 λ 特征形式的整个功能映射。

2.5 由Sep-point模块组成的点云分类网络

为了验证本文的Sep-point模块的先进性能,本文将其嵌入到现有的点云分类基线网络中。本文使用了类似于PointNet++的分层结构设计,具体点云分类网络结构如图2所示。不同之处在于首先将点云坐标嵌入到高维特征空间中,学习高维空间信

息。将本文的Sep-point模块替换掉基准网络中的特征提取模块。每个层次结构分别由采样与分组层、Sep-point模块以及池化层构成,逐步扩大感知域。高维特征和采样的中心点坐标作为点云特征提取的输入。在每个层次结构中,采用与PCT相同的抽

样和分组策略。对于两个采样与分组层,下采样的中心点个数分别为512和128,输出特征通道维度分别为128和256。随后,使用两个MLP层将分层特征通道维度映射到1 024,并由分类解码器输出分类分数。

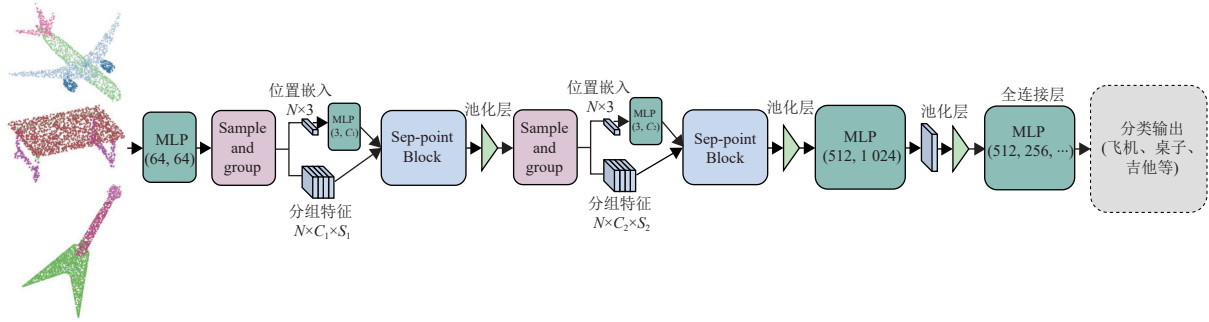


图2 由Sep-point模块组成的点云分类网络

Fig.2 Point cloud classification network with sep-point modules

3 实验结果及讨论

3.1 数据集

本文对ModelNet40^[24]数据集进行对象分类评估。该数据集包含40个类别的12 311个CAD模型,其中9 843个用于训练,2 468个用于测试。本文采用对于每个模型从网格平面均匀采样1 024个点并被缩放以适应单位球体的数据预处理方法。在训练过程中,使用[0.2,0.2]范围内的随机平移、[0.67,1.5]范围内的随机各向异性缩放和随机输入Dropout来增强输入数据。对于分类解码器部分,采用全局最大池化方法获取点云的全局特征,然后采用两个全连接层(512、256)来转换全局特征。在最后两个完全连接的层中使用保持概率为0.5的Dropout。输出层都包括LeakyReLU和批处理规范化。

此外本文还在ScanObjectNN^[25]基准测试上进行了实验。该数据集包含15 000个对象,这些对象被分为15类,在现实世界中有2 902个独特的对象实例。ScanObjectNN上的实验设置与ModelNet40上的分类实验相同。

3.2 实验细节

在每种情况下,使用学习率为0.1的SGD优化函数进行训练,本文使用余弦退火将学习率降低到0.001。批归一化的动量为0.9,不使用批量归一化衰减。训练批次大小为32,测试批次大小为16,动量为0.9,丢弃率为0.5,并使用交叉熵损失函数以及值为1的随机种子。所有点云分类实验只使用坐标作为输入,不包括法向量、投票策略和任何数据增强方法。

所有实验均在NVIDIA RTX 4 090 GPU、python 3.8、pytorch 2.0.1 和 CUDA 11.8上进行。

3.3 实验结果及分析

3.3.1 在ModelNet40数据集上的分类实验

为了验证本文方法的有效性,ModelNet40数据集上的实验结果如表1所示。本文方法总体准确率为93.4%,平均准确率为90.9%。与基线模型PointNet++相比,总体准确率提高了2.7个百分点,这充分说明本文提出的方法在捕捉局部细节与全局语义方面占据较大优势,弥补了PointNet++在建模全局信息上的空缺。与基于全局自注意机制的PCT相比,本文方法整体准确率和平均准确率分别提高了0.2和0.7个百分点,这表明组内自注意方法成功建模出局部细粒度结构,具有良好的语义特征学习能力。另外,与其他基于卷积的方法相比,本文的方法也具有较大优势,例如,与DGCNN相比本文的方法分别提升了0.5和0.7个百分点。总体实验结果表明,由Sep-point组成的点云分类网络达到了良好的性能,对特征提取内容的完整性做出较大贡献。

3.3.2 在ScanObjectNN数据集上的分类实验

为了验证本文方法在真实场景中的分类性能,本文在包含噪音干扰的ScanObjectNN真实数据集上进行实验,该实验结果如表2所示,表中所有实验的输入点云个数均为1 024。结果显示,本文方法在ScanObjectNN数据集上仍然显示出较先进的性能。最高整体准确率为84.2%,平均准确率为82.5%,与其他方法相比性能显著。具体地,与PointNet++相比,本文方法整体准确率和平均准确率分别提高了6.3和

7.36个百分点。与PCT相比,本文方法整体准确率和平均准确率分别提高了4.8和5.8个百分点。实验结果表明,本文方法非常适合于具有真实数据的点云分类任务,并取得了较好的分类效果,同时也表明本文的方法具有一定的抗干扰效果,在解决点云数据的无序性和稀疏性方面具备较强的适应能力。

表 1 ModelNet40数据集上不同方法的分类性能
Table 1 Classification performance of different methods on the ModelNet40 dataset

模型	输入点类型	输入点数量/K	总体准确率/%	平均准确率/%
PointNet ^[14]	P	1	89.2	86.0
PointCNN ^[16]	P	1	92.5	88.1
SpiderCNN ^[26]	P	1	92.4	
PointConv ^[19]	P+N	1	92.5	
KPConv ^[20]	P	7	92.9	
DGCNN ^[18]	P	1	92.9	90.2
RS-CNN ^[27]	P	1	92.9	
PointASNL ^[28]	P+N	1	93.2	
Point-TnT ^[29]	P	1	92.6	
PointStack ^[30]	P	1	93.3	89.6
PointNet++ ^[15]	P	1	90.7	
PCT ^[9]	P	1	93.2	90.2
Sep-point	P	1	93.4	90.9

表 2 ScanObjectNN数据集上不同方法的分类性能
Table 2 Classification performance of different methods on the ScanObjectNN dataset

模型	输入点数量/K	总体准确率/%	平均准确率/%
PointNet ^[14]	1	68.2	63.4
PointCNN ^[16]	1	78.5	75.1
DGCNN ^[18]	1	78.1	73.6
DRNet ^[31]	1	80.3	78.0
GBNet ^[32]	7	80.5	77.8
Simpleview ^[33]	1	80.5	
MVTN ^[34]	1	82.8	
PointBERT ^[35]	1	83.1	
PRA-Net ^[36]	1	82.1	79.1
Point-TnT ^[29]	1	83.5	81.0
PointNet++ ^[15]	1	77.9	75.4
PCT ^[9]	1	79.4	76.7
Sep-point	1	84.2	82.5

3.3.3 计算需求分析

为了说明由Sep-point组成分类网络的性能和复杂性,分类性能和计算需求分析表如表3所示,本文通过比较所需的浮点运算数(FLOPs,单位为M)、参数数量(Params,单位为G)以及时延(Latency,单位为ms),分析了本文的分类网络与其他几个基线网络的

计算需求与性能方面的优势。结果显示,Sep-point具有较低的计算需求和复杂度,只有2.16 M参数、2.10 G FLOPs和173 ms的时延,但却提供了93.4%的高精度分类结果。与基线模型PointNet++相比,本文方法只增加了0.68 M的网络参数和1.33 G的浮点运算数(FLOPs),时延减少了11 ms,这表明本文方法利用较低的计算负担,却提取了极具代表性的特征,在性能效率与需求轻量级中达到了较好的平衡。与PCT相比,Sep-point不仅参数减少了0.72 M、FLOPs减少了0.18 G以及Latency减少了40 ms,而且精度方面也存在较大优势,解决了注意力机制提供性能优势但存在较高计算负担的弊端。总的来说,Sep-point具有较好的准确性和较低的计算和内存要求,实现了高性能与轻量级的双重优势。

表 3 ModelNet40上不同分类网络的性能和计算资源需求
Table 3 The performance and computational resource requirements of different classification networks on the ModelNet40 dataset

模型	Params /M	FLOPs/G	总体准确率/%	Latency /ms
PointNet ^[14]	3.47	0.45	89.2	13.6
PointNet++ ^[15]	1.48	0.77	90.7	184
DGCNN ^[18]	1.81	2.43	92.9	85.8
PCT ^[9]	2.88	2.28	93.2	213
Sep-point	2.16	2.10	93.4	173

3.3.4 消融实验

为了量化模型组件的有效性,本文分别在ModelNet40、ScanObjectNN数据集上对模型进行了消融研究,采用与上述分类实验相同的实验设置。分别评估了Sep-point中组内自注意力(LSA)、组间自注意力(GSA)和位置令牌嵌入(LTE)模块对模型性能的影响,消融实验结果如表4所示。

表 4 Sep-point消融实验结果
Table 4 Ablation experiment results of the Sep-point %

模型	LSA	GSA	LTE	ModelNet40		ScanObjectNN	
				OA	ACC	OA	ACC
Sep-point	√		√	93.1	90.1	82.2	79.7
		√	√	93.3	90.7	83.1	80.9
	√	√		93.0	89.8	80.8	78.0
	√	√	√	93.4	90.9	84.2	82.5

对于ModelNet40数据集来说,LSA模块分别提升了0.1个百分点的总体准确率和0.2个百分点的平均准确率;GSA模块分别提升了0.3个百分点的总体准确率和0.8个百分点的平均准确率;LTE模块分别提升了0.4个百分点的总体准确率和1.1个百分点的

平均准确率。这说明本文提出的Sep-point中的各模块都各自发挥着相应的功能,让所提取的特征信息更加全面。对于ScanObjectNN数据集来说,LSA模块分别提升了1.1个百分点的总体准确率和1.6个百分点的平均准确率;GSA模块分别提升了2.0个百分点的总体准确率和2.8个百分点的平均准确率;LTE模块分别提升了3.4个百分点的总体准确率和4.5个百分点的平均准确率。每个模块的效果非常显著,这将对点云真实场景的实际应用提供有价值的参考。总体而言,本文提出的Sep-point中的所有模块都有助于提升网络性能,重点突出在GSA和LTE组件。这验证了可分离自注意思想以及新的位置令牌嵌入方法在点云分类领域应用的必要性。

此外,本文还评估了单独模块所呈现出的性能价值,仅含单个模块的消融实验结果如表5所示。从表中可以分析得出,若只使用单个模块作为网络特征提取模块,两种数据集的结果均呈现出较低的性能,特别是只使用LSA模块或GSA模块,这说明每个模块对于网络性能的贡献是相辅相成的,配套使用可呈现出良好的性能优势。结果表明,本文提出的模块发挥着各自的功能,且这些功能存在互补的优势,使提取的特征更完整和更具代表性。

表 5 使用单个模块的实验结果

Table 5 Experimental results using a single module

%

模型	LSA	GSA	LTE	ModelNet40		ScanObjectNN	
				OA	ACC	OA	ACC
Sep-point	√			92.7	89.6	80.7	78.6
		√		92.7	89.2	80.5	77.5
			√	93.2	90.2	82.4	80.5

为了进一步直观验证本文核心模块对网络性能的贡献,对LSA以及GSA模块所提取的特征进行可视化分析,特征可视化图如图3所示。可以发现对于LSA模块,负责重点捕捉点云的局部细节结构,例如:飞机机翼与机舱的连接处、机舱与机尾的连接处;杯子的四周边缘以及杯子与把手的连接处等。而GSA模块从全局的角度出发,重点关注远程点之间的联系,建模点云整体结构轮廓,例如:飞机左右机翼以及整个机舱;杯子的整个杯身等。

3.3.5 附加分类指标性能评估

为了进一步解释本文方法的优势和局限性,额外在ModelNet40数据集中评估了分类网络的Precision、Recall和F1-score指标,额外性能指标实验结果如表6所示。实验结果表明,该方法的Precision、Recall和F1-score指标均高于其他方法。在

ModelNet40数据集中,本文方法的准确率、召回率和F1-score指标分别比基线网络PointNet++提高0.8个百分点、0.7个百分点和0.7个百分点。实验结果表明,本文方法具有较好的实验结果,进一步增强了Sep-point优势的可解释性。



图 3 LSA模块(左边)与GSA模块(右边)的可视化效果,红色点是各自提取的特征信息

Fig.3 Visualization of the LSA module (left) and the GSA module (right) , where the red points are the corresponding feature information

表 6 ModelNet40数据集上的Precision、Recall和F1-score分数的实验结果

Table 6 Experimental results of the Precision, Recall and F1-score scores on the ModelNet40 dataset

%

模型	Precision	Recall	F1-score
PointNet	86.3	85.6	85.8
PointNet++	90.0	89.7	89.7
DGCNN	90.2	89.5	89.7
PCT	90.3	89.2	89.6
Ours	90.5	90.1	90.1

4 结论

本文提出Sep-point模块用于3D点云分类任务,通过借鉴视觉领域深度可分离Transformer的思想,将组内自注意力和组间自注意力机制有机结合,以相互嵌套运用的形式提取点云局部-全局依赖关系,有效提取局部细节的同时关注远程点之间的交互信息,可以更全面地捕获多方面深层关键特征信息。另外,通过引入新的位置令牌嵌入方法以替代使用每个分组所有点标记的方法,以可忽略的计算代价来实现高效的网络性能。尽管在点云分类任务中取得了一些进展,但还存在网络稳定性和任务差异性带来的挑战。未来将继续研究更先进的方法,特别是在扩展可分离自注意方法上,以提升点云分类效果,为点云领域提供更多可靠的支持。

参考文献:

[1] SHI S, WANG X, LI H. PointRCNN: 3D object proposal

- generation and detection from point cloud[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019: 770-779.
- [2] VASWANI A, SHAZEER N, PARMAR N, *et al.* Attention is all you need[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach: ACM, 2017: 6000-6010.
- [3] RAO Y, LU J, ZHOU J. Spherical fractal convolutional neural networks for point cloud recognition[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019: 452-460.
- [4] YI L, SU H, GUO X, *et al.* Synspecnn: synchronized spectral CNN for 3D shape segmentation[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017: 6584-6592.
- [5] ZHANG Z, LI K, YIN X, *et al.* Point cloud semantic scene segmentation based on coordinate convolution [J]. *Computer Animation and Virtual Worlds*, 2020, 31(4-5): e1948.
- [6] SHI S, WANG Z, SHI J, *et al.* From points to parts: 3D object detection from point cloud with part-aware and part-aggregation network [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020, 43(8): 2647-2664.
- [7] PENG L, LIU F, YU Z, *et al.* Lidar point cloud guided monocular 3D object detection[C]//2022 European Conference on Computer Vision. Tel Aviv: Springer, 2022: 123-139.
- [8] YANG H, SHI J, CARLONE L. Teaser: fast and certifiable point cloud registration [J]. *IEEE Transactions on Robotics*, 2020, 37(2): 314-333.
- [9] GUO M, CAI J, LIU Z, *et al.* PCT: point cloud transformer [J]. *Computational Visual Media*, 2021, 7(2): 187-199.
- [10] ZHAO H, JIANG L, JIA J, *et al.* Point transformer [J]. *IEEE Access*, 2021, 9(3): 16259-16268.
- [11] LI W, WANG X, XIA X, *et al.* Sepvit: separable vision transformer[EB/OL]. arXiv: 2203.15380(2022-06-15) [2024-04-16]. <https://doi.org/10.48550/arXiv.2203.15380>.
- [12] WANG Z, LU F. VoxSegNet: volumetric CNNs for semantic part segmentation of 3D shapes [J]. *IEEE Transactions on Visualization and Computer Graphics*, 2019, 26(9): 2919-2930.
- [13] SHI S, GUO C, JIANG L, *et al.* PV-RCNN: point-voxel feature set abstraction for 3D object detection[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020: 10526-10535.
- [14] QI C, SU H, MO K, *et al.* PointNet: deep learning on point sets for 3D classification and segmentation[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017: 77-85.
- [15] QI C, YI L, SU H, *et al.* PointNet++: deep hierarchical feature learning on point sets in a metric space[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach: ACM, 2017: 5105-5114.
- [16] LI Y, BU R, SUN M, *et al.* PointCNN: convolution on x-transformed points[C]//Proceedings of the 32nd International Conference on Neural Information Processing Systems. Montréal: ACM, 2018: 828-838.
- [17] LIU X, HAN Z, LIU Y, *et al.* Point2sequence: learning the shape representation of 3D point clouds with an attention-based sequence to sequence network[C]//Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence and Thirty-First Innovative Applications of Artificial Intelligence Conference and Ninth AAAI Symposium on Educational Advances in Artificial Intelligence. Honolulu: AAAI, 2019, 33: 8778-8785.
- [18] 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.
- [19] WU W, QI Z, LI F. PointConv: deep convolutional networks on 3D point clouds[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019: 9613-9622.
- [20] THOMAS H, QI C, DESCHAUD J, *et al.* KPConv: flexible and deformable convolution for point clouds[C]//2019 IEEE/CVF International Conference on Computer Vision. Seoul: IEEE, 2019: 6410-6419.
- [21] XU M, DING R, ZHAO H, *et al.* PAConv: position adaptive convolution with dynamic kernel assembling on point clouds[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE, 2021: 3172-3181.
- [22] HOWARD A, ZHU M, CHEN B, *et al.* Mobilenets: efficient convolutional neural networks for mobile vision applications[EB/OL]. arXiv: 1704.04861(2017-04-17) [2024-04-16]. <https://doi.org/10.48550/arXiv.1704.04861>.
- [23] LANDRIEU L, BOUSSAHA M. Point cloud oversegmentation with graph-structured deep metric learning[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019: 7432-7441.
- [24] WU Z, SONG S, KHOSLA A, *et al.* 3D ShapeNets: a deep representation for volumetric shapes[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston: IEEE, 2015: 1912-1920.
- [25] UY M, PHAM Q, HUA B, *et al.* Revisiting point cloud classification: a new benchmark dataset and classification model on real-world data[C]//2019 IEEE/CVF International Conference on Computer Vision. Seoul: IEEE, 2019: 1588-1597.
- [26] XU Y, FAN T, XU M, *et al.* SpiderCNN: deep learning on point sets with parameterized convolutional filters[C]//Computer Vision-ECCV 2018. Munich: Springer, 2018: 99-105.
- [27] LIU Y, FAN B, XIANG S, *et al.* Relation-shape convolutional neural network for point cloud analysis[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019: 8887-8896.
- [28] YAN X, ZHENG C, LI Z, *et al.* PointASNL: robust point clouds processing using nonlocal neural networks with adaptive sampling[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020: 5588-5597.

