DOI: 10.13382/j.jemi.B2407694

基于 Transformer 的逐通道点云分析网络*

冯凯浩[1] 陶志勇[1] 李 衡[1] 李铭朗[1] 林 森[2]
(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:三维点云能够充分描述目标对象的几何信息,在自动驾驶、医学影像和机器人等领域有着广泛的应用前景。然而,现有方法在处理不同通道间的特征时缺乏差异化,同时对低级空间坐标和高级语义特征采用统一的编码策略,进而导致点云特征提取不全面。因此,提出了基于 Transformer 的逐通道点云分析网络。首先,为了克服传统图卷积在混合通道中难以区分有效信息的挑战,设计了一种深度可分离边缘卷积,可以在逐通道特征提取时保留局部几何信息的同时,显著提升通道间的区分能力。其次,针对 Transformer 在低级空间坐标和高级语义特征中采用统一编码方式,导致信息提取不足的问题,提出了两种特征编码策略,自适应位置编码和空间上下文编码,分别用于探索低级空间中的隐式几何结构和高级空间中的复杂上下文关系。最后,提出了一种有效的融合策略,可以形成更具区分性的特征表示。为了充分证明所提出模型的有效性,在公开数据集 ModelNet40 和 ScanObjectNN 上进行点云分类实验,总体分类精度分别达到 93.7%和 83.2%,在公开数据集 ShapeNet Part 上,整体部件分割的平均交并比达到 86.0%。因而,研究方法在分类和分割任务中均具有先进的性能。

关键词:点云分类;分割;深度可分离卷积;Transfomer;融合算法;ModelNet40
中图分类号:TP391.4;TN911.7 **文献标识码**:A **国家标准学科分类代码**:520.2

Transformer-based channel-by-channel point cloud analysis network

Feng Kaihao[1] Tao Zhiyong[1] Li Heng[1] Li Minglang[1] Lin Sen[2]
(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: 3D point clouds can fully describe the geometric information of target objects and have a wide range of applications in fields such as autonomous driving, medical imaging and robotics. However, existing methods lack differentiation when dealing with features between different channels, and at the same time adopt a unified coding strategy for low-level spatial coordinates and high-level semantic features, which in turn leads to incomplete point cloud feature extraction. Therefore, this paper proposes a channel-by-channel point cloud analysis network based on Transformer. First, in order to overcome the challenge of traditional graph convolution that is difficult to distinguish effective information in mixed channels, a depth-separable edge convolution is designed, which can significantly improve the inter-channel differentiation ability while preserving local geometric information during channel-by-channel feature extraction. Secondly, to address the problem that Transformer adopts a uniform coding approach in low-level spatial coordinates and high-level semantic features, which leads to insufficient information extraction, two feature coding strategies are proposed adaptive positional coding and spatial context coding, which are used to explore implicit geometric structures in low-level space and complex contextual relationships in high-level space, respectively. Finally, an effective fusion strategy is proposed, which can result in a more discriminative feature representation. In order to fully demonstrate the effectiveness of the proposed model, point cloud classification experiments are conducted on the public datasets ModelNet40 and ScanObjectNN, where the overall classification accuracies reach 93.7% and 83.2%, respectively, and the average intersection and merger ratio of overall part segmentation reaches 86.0% on the public dataset ShapeNet Part. Thus, the method in this paper has advanced performance in both classification and segmentation tasks.

收稿日期: 2024-07-17 Received Date: 2024-07-17
*基金项目:辽宁省科技厅应用基础研究项目(2022JH2/101300274)、辽宁省高等学校基本科研项目(LJKMZ20220679)资助

Keywords: point cloud classification; segmentation; deep separable convolution; Transfomer; fusion algorithm; ModelNet40

0 引 言

三维点云是一种代表几何对象表面形状的数据结构,通常广泛应用于自动驾驶[1]、医学影像[2]和机器人视觉[3]等领域。与规则密集像素组成的二维图像不同,点云具有无序和非结构化的特性,因此获取点云的语义信息是一项具有挑战性的任务。

为了解决这一挑战并将神经网络应用于点云分析,传统的基于深度学习的方法是将无结构的点云数据进行结构化表示。其中,一些方法将点云进行体素化处理(如 VoxNet[4]和 OctNet[5]),以便利用三维离散卷积进行特征学习。然而,体素化处理将所有点等同看待,忽略了不同区域内点集的稀疏性,且随着体素分辨率的增加,三维离散卷积的计算和内存成本会呈指数级增长,因此,在语义分割或处理复杂场景时会显著增加计算复杂度。另外一些方法将点云从多个视角投影到二维平面(如 MVCNN[6]和 GVCNN[7]),形成多视图,再应用卷积神经网络在图像领域成熟的技术进行处理。然而,该方法在投影过程中会造成点云信息丢失。

近年来,研究趋势转向直接处理点云数据而不引入中间表示。PointNet[8]作为直接处理点云的开创性工作,采用多层感知机(multi-layer perceptron, MLP)对无序点云进行特征提取,并通过最大值池化来确保点云的置换不变性。然而,该方法未考虑点云邻域的几何结构信息,导致特征提取不充分。针对直接使用 MLP 进行点云分析时,所有点特征被等同对待的问题,Thomas 等[9]提出了一种新的点卷积算子,通过不同核心点与中心点的线性关系设计不同的核点卷积权重。Wu 等[10]提出了密度重加权卷积,可以根据不同区域的密度自适应地学习区域特征。Xu 等[11]提出动态组装权重矩阵来构造新的内核,权重矩阵的组合系数由自适应地学习点的相对位置关系得到,以数据驱动的方式对三维点云进行建模。此外,针对不同的排列顺序导致相同形状的点云数据输出特征存在差异的问题。Li 等[12]提出了 X-Conv 来确保点云的排序不变性,根据相对位置提取的局部特征和逐点的全局特征来形成变换方阵。以上的改进方法主要通过改善卷积核的设计来实现自适应特征提取,但未深入挖掘区域内点的几何关系,仍存在局部特征提取不充分的问题。

为了增强对局部区域内信息的提取,一些研究尝试探索适用于点云分析的分组策略。PointNet++[13]在最远点采样的基础上引入了多分辨率分组和多尺度分组的方法,用于对点云进行局部区域划分。Ran 等[14]通过构建以采样点为中心的三角结构和伞形结构来捕获点云的邻域几何结构信息。而 Xiang 等[15]利用曲线分组算子和曲线聚合算子来增强逐点特征。这些方法显著提升了点云分类分割任务的精度,但由于复杂的分组策略,计算复杂度也随之提升。

上述研究显示构建有效的邻域几何关系是点云邻域特征提取的关键。在点云分析任务中,由于点云特征仅包含三维坐标,直接利用普通卷积提取隐含的几何结构信息存在困难。图结构因其能有效表示非欧几里德数据而备受研究人员关注。DGCNN[16]使用图卷积构造点云邻域内的结构信息,采用边缘卷积算子(edge convolution, EdgeConv)线性聚合中心点特征和边缘特征来捕获局部上下文信息,该结构有助于将点云中的隐式特征显式化。受此启发,LDGCNN[17]通过连接不同动态图的分层特征,实现不同层次特征的信息传递和融合。Zhao 等[18]则通过计算邻域内点的特征差异,加权处理有向边以实现区域内特征聚合。尽管以上方法在利用图结构方面取得了显著进展,但密集连接图结构会产生大量计算和额外冗余信息。此外,普通卷积在图结构中不能区别对待不同的通道,这也给特征学习带来了一定的限制。

在自然语言处理领域,Transformer 系列模型得到了广泛应用[19-20]。Transformer 模型通过注意力机制来动态调整注意力权重,从而捕捉序列中不同位置之间的相关性。在点云数据中,不同邻域点对中心点的相关性也是不同的。基于这一特性,自注意力机制也可以有效地应用于点云在局部区域内的特征学习。Zhao 等[21]利用注意力机制构建了一种新颖的点转换器,以点对为注意力单位实现特征提取。Guo 等[22]设计了 PCT(point cloud transformer)网络,通过增加坐标的输入嵌入和优化的自注意力机制,实现了更具区分性的特征表示方法。于喜俊等[23]在 PCT 的基础上使用集成模型对点云进行分类。Yu 等[24]则在注意力机制的基础上引入掩码建模策略,实现了点云在弱监督条件下的特征训练。然而,以往的工作使用单一的编码方式处理低级坐标和高级特征,导致信息提取不充分。

受到上述工作的启发,本研究基于图结构设计了深度可分离边缘卷积,以实现逐通道特征提取。这一设计不仅修正了普通卷积中卷积核无差别融合所有通道的不足,还简化了网络的计算复杂度,实现了更有效的特征学习。为了解决 Transformer 架构无法区分处理低级坐标和高级特征的问题,设计了两种基于 Transformer 的特征提取方法作为深度可分离边缘卷积的后置模块,位置自适应编码和空间上下文编码。前者在低级坐标空间中引入

多种特征关系,从而增强点云的局部几何学习;后者则通过计算空间相似性度量,在高级特征空间中对区域间上下文关系进行编码。最后,本文提出了自适应融合模块,用于调整权重系数以实现深度可分离边缘卷积和空间上下文编码的输出结果,从而实现紧凑的特征表示。

1　算法介绍

本研究的总体网络架构如图 1 所示。主要包括嵌入层、编码层和解码层。嵌入层通过低级几何构造模块拓展点云输入特征,利用深度可分离边缘卷积和位置自适应编码在低级空间中捕获几何结构关系,实现对低级坐标信息的编码。编码层采用层级结构,主要由深度可分离边缘卷积、空间上下文编码和融合策略 3 个模块组成,实现在高级特征空间中的编码,进而形成更全面和更具区分性的特征表征。在解码层中,针对不同任务使用了全连接层实现高级特征的解码。

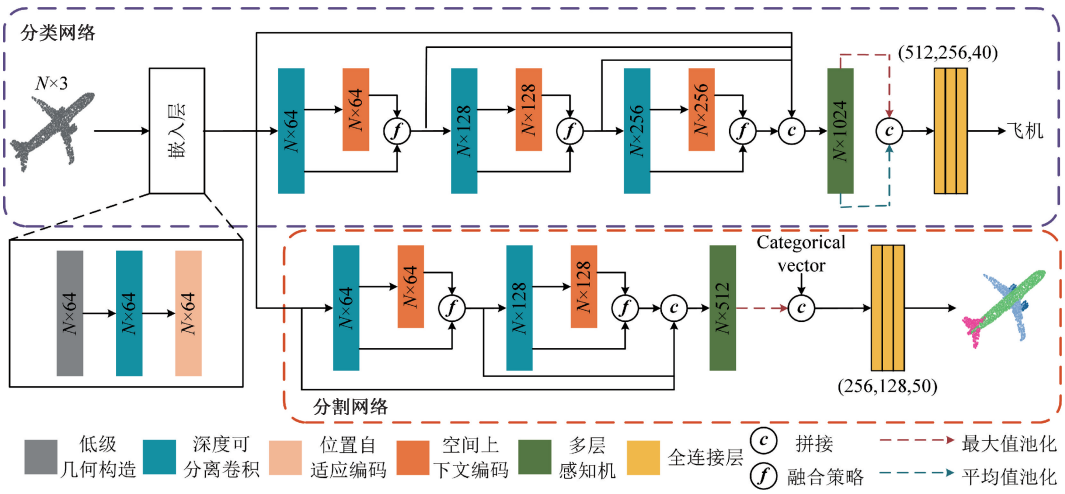


图 1　总体网络架构

Fig. 1　Overall network architecture

在分类网络中,设计了 3 层的编码结构用于特征提取。编码后的特征通过拼接操作后,分别使用最大值池化和平均值池化得到聚合向量,最终通过维度为(512,256,40)的全连接层输出 40 个对象类别的分类得分。在分割网络中,对点云进行两次特征编码,然后将其与 Categorical vector 进行拼接,其中 Categorical vector 是指物体类别标签的 One-hot 编码结果,可以提供每个点的类别信息。

1.1　低级几何构造

为增强低级几何关系表示,在嵌入层中设计低级几何构造模块,该模块可以扩展点云输入,进而丰富低级几何关系。定义输入点云为包含 N 个点的点集 $\boldsymbol{P}=(\boldsymbol{p}_i \mid i=1,2,\cdots,N) \subseteq \boldsymbol{R}^{N\times 3}$,对于每个点 $\boldsymbol{p}_i$,通过索引距离该点最近的两个点构成一个平面,以该平面的法向量作为该点的法向量特征 $\boldsymbol{l}_i$,计算公式为:

$$\begin{cases} \boldsymbol{e}_{i,1}=\boldsymbol{p}_{i,1}-\boldsymbol{p}_i \\ \boldsymbol{e}_{i,2}=\boldsymbol{p}_{i,2}-\boldsymbol{p}_i \\ \boldsymbol{l}_i=\boldsymbol{e}_{i,1}\times\boldsymbol{e}_{i,2} \end{cases} \tag{1}$$

式中:$\boldsymbol{e}_{i,j}$ 表示中心点与邻点的有向边。将原始坐标与经过低级几何构造的法向量特征拼接,输出为 $\boldsymbol{P}' \in \boldsymbol{R}^{N\times 6}$,可以实现邻域几何结构的特征嵌入,为后续的编码提供丰富的低级几何信息。

1.2　深度可分离边缘卷积

受 DGCNN 的启发,将图结构应用于点云特征学习可以更有效地处理点云这类非结构化数据。在以往的基于图的方法中,首先利用采样点和邻域点构建图结构,再使用 3D 卷积对图结构中的有向边进行编码,在这一过程中,点云各个通道的特征信息被无区分对待。因而,本研究在图结构的基础上,使用深度可分离卷积代替普通卷积提取点云特征,该模块被命名为深度可分离边缘卷积。

首先,从点云中选择图的节点和有向边,进而构建有向图 $\boldsymbol{G}=(\boldsymbol{V},\boldsymbol{E})$,其中 $\boldsymbol{V}=(\boldsymbol{p}'_1,\boldsymbol{p}'_2,\cdots,\boldsymbol{p}'_n)$ 代表图节点集合,$\boldsymbol{E} \subseteq \boldsymbol{V}\times\boldsymbol{V}$ 代表节点与邻域点构造出的有向边集合,如图 2 所示。节点代表了点在全局中的绝对位置,有向边代表包含邻域信息的相对位置,通过绝对位置与相对位置的拼接可以构建有效的点云特征表示,计算公式为:

$$\boldsymbol{g}_{i,j}=[\boldsymbol{p}'_i-\boldsymbol{p}'_{i,j},\boldsymbol{p}'_i] \tag{2}$$

式中:[·]代表拼接操作。

深度可分离卷积由两步卷积组成,分别是深度卷积和逐点卷积,如图 3 所示。在深度卷积中,不同的卷积核

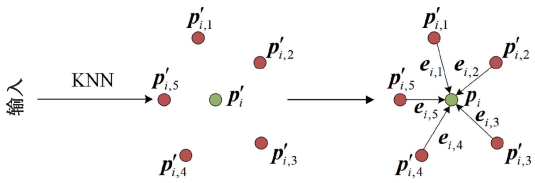


图 2 构建图结构

Fig. 2 Constructing graph structures

分别作用于不同的输入通道,生成相应通道的特征图,再将特征图进行拼接。具体的计算公式为:

$$\boldsymbol{y}'_{i,j,c'} = \sum \boldsymbol{g}_{i,j,c'} \cdot \boldsymbol{\omega}'_{i,j,c'} \tag{3}$$

式中:c' 为深度卷积的通道索引,$\boldsymbol{\omega}'$ 是深度卷积中卷积核的权重。

逐点卷积将深度卷积得到的特征图进行组合,实现不同通道之间的信息融合,新特征图 $\boldsymbol{Y}$ 计算公式为:

$$\begin{aligned} \boldsymbol{y}_{i,j} &= (\boldsymbol{y}_{i,j,c}) = \left(\sum_{c'} \sum \boldsymbol{y}'_{i,j,c'} \cdot \boldsymbol{\omega}_{i,j,c}\right) \\ \boldsymbol{Y} &= (\boldsymbol{y}_i) = (\max(\boldsymbol{y}_{i,j})) \end{aligned} \tag{4}$$

式中:c 是逐点卷积核的通道索引;$\boldsymbol{\omega}$ 是该卷积核的权重;$\boldsymbol{Y} \in \boldsymbol{R}^{N \times F}$ 为经过深度可分离卷积后输出的特征图,F 代表特征通道数。

深度可分离边缘卷积可以对点云数据不同通道间语义相关性较弱的特征进行逐通道特征提取,从而可以在嵌入层中丰富低级几何关系,在编码层中实现高级特征融合和动态特征图更新。同时,相比于普通边缘卷积,深度可分离边缘卷积可以减少参数量,从而降低计算复杂度。

在编码层中,本研究采用 kNN 算法在每一层的特征空间中重新计算图结构。在第 l 层构建的有向图定义为 $\boldsymbol{G}^{(l)} = (\boldsymbol{V}^{(l)}, \boldsymbol{E}^{(l)})$,每一层的邻域图大小可以通过邻域点个数 k_l 自适应调整,通过动态图更新可以将网络的感受野扩展到全局点云。研究中,各层的 k_l 采用相同的数值 k。

1.3 Transfomer 模块

为了增强低级空间坐标中的几何结构信息和高级语义特征中的显著信息,设计位置自适应编码模块和空间

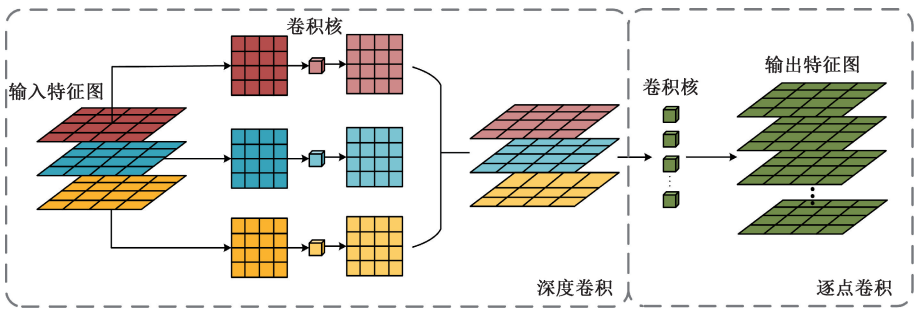


图 3 深度可分离卷积

Fig. 3 Depthwise separable convolution

上下文编码模块分别作用于嵌入层和编码层,作为深度可分离边缘卷积的后置模块,如图 4 所示。

1) 位置自适应编码,为增强嵌入层中的几何关系并丰富邻域特征,采用位置自适应编码对特征图和原始坐标进行嵌入。在位置自适应编码中,$\boldsymbol{Q}$(query)和 $\boldsymbol{K}$(key)分别代表特征图中顶点坐标与有向边做线性操作后得到的特征,被称为查询矩阵和键矩阵,而 $\boldsymbol{V}$(value)代表经过卷积层后的特征,被称为值矩阵。

由于位置自适应编码包含原始坐标信息,因此可以实现位置编码,具体计算公式为:

$$\begin{cases} Q = \boldsymbol{Y} \cdot \boldsymbol{W}_q, & \boldsymbol{q}_i = Linear(\boldsymbol{y}_i) \\ K = \boldsymbol{E} \cdot \boldsymbol{W}_k, & \boldsymbol{k}_i = Linear(\boldsymbol{e}_i) \\ V = \boldsymbol{P}' \cdot \boldsymbol{W}_v, & \boldsymbol{v}_i = MLP(\boldsymbol{p}'_i) \end{cases} \tag{5}$$

式中:Linear 代表线性操作,$\boldsymbol{W}_q$,$\boldsymbol{W}_k$ 和 $\boldsymbol{W}_v$ 为不同特征的权重矩阵。

在注意力权重的计算中,首先计算 $\boldsymbol{Q}$ 矩阵和 $\boldsymbol{K}$ 矩阵之间的差异特征,再将包含原始特征的 $\boldsymbol{V}$ 矩阵结合,从而增强嵌入层中误差的学习能力,达到在低级空间中限制冗余信息传播的目的,实现网络在训练过程中的前向纠错功能。注意力权重 $\boldsymbol{W}_{\mathrm{T}}$ 的计算公式为:

$$\boldsymbol{W}_{\mathrm{T}} = \mathrm{Softmax}(\mathrm{MLP}(\boldsymbol{Q} - \boldsymbol{K} + \boldsymbol{V})) \tag{6}$$

K 矩阵由有向边特征组成,其本质为拉普拉斯矩阵,包含了由绝对位置组成的 $\boldsymbol{Q}$ 矩阵的部分信息,因而,选取 $\boldsymbol{K}$ 矩阵和 $\boldsymbol{V}$ 矩阵作为注意力对象来提取低级空间中的几何结构信息,输出特征 $\boldsymbol{F}_{\mathrm{T}}$ 为:

$$\boldsymbol{F}_T = (\boldsymbol{K} + \boldsymbol{V}) \cdot \boldsymbol{W}_{\mathrm{T}} \tag{7}$$

2) 空间上下文编码,在高级空间中,点的特征包含局部和全局信息,为实现有效的信息融合并限制冗余信息,采用自注意力机制的思想进行特征编码。首先,通过对高级特征进行线性变换,生成 $\boldsymbol{Q}$、$\boldsymbol{K}$ 和 $\boldsymbol{V}$ 矩阵。具体实

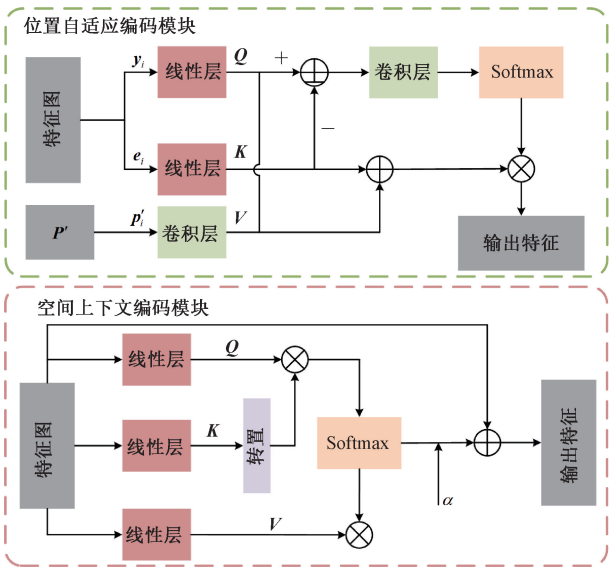


图 4　基于 Transformer 的两种编码方式

Fig. 4　Two encoding schemes based on Transformer

现为：

$$(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V}) = \boldsymbol{Y}\cdot(\boldsymbol{W}_q,\boldsymbol{W}_k,\boldsymbol{W}_v) \tag{8}$$

式中：$\boldsymbol{W}_q$、$\boldsymbol{W}_k$ 和 $\boldsymbol{W}_v$ 为共享的可学习线性变换的权重矩阵。为了在编码过程中不造成信息丢失，摒弃了 PCT 对 $\boldsymbol{Q}$ 和 $\boldsymbol{K}$ 矩阵降通道的编码方案，即 $\boldsymbol{Q}$、$\boldsymbol{K}$ 和 $\boldsymbol{V}$ 矩阵的特征通道数相同。注意力权重计算方法为：

$$\boldsymbol{W'}_{\mathrm{T}} = \mathrm{Softmax}\left(\frac{\boldsymbol{Q}\cdot\boldsymbol{K}^{\mathrm{T}}}{\sqrt{d_a}}\right) \tag{9}$$

式中：d_a 代表 $\boldsymbol{Q}$、$\boldsymbol{K}$ 和 $\boldsymbol{V}$ 矩阵的特征维度，该注意力权重 $\boldsymbol{W'}_{\mathrm{T}}$ 代表高级特征的空间相似度。

在位置自适应编码中，采用了先融合再加权的特征提取策略，但在空间上下文编码中采用先加权再融合的方式。输出特征 $\boldsymbol{F'}_{\mathrm{T}}$ 计算为：

$$\boldsymbol{F'}_{\mathrm{T}} = (V\cdot\boldsymbol{W'}_{\mathrm{T}})\odot\alpha + \boldsymbol{Y} \tag{10}$$

式中：α 为可学习参数，用于自适应特征学习；$\odot$ 为 Hadamard 乘法。

1.4　自适应融合模块

为了有效融合编码层中深度可分离边缘卷积和空间上下文编码的输出特征，从而获得更丰富全面的特征表征，设计了自适应融合模块，如图 5 所示。该模块可以减少网络对单一特征的过渡依赖，从而提高模型的鲁棒性和泛化能力。

在融合模块中，两个分支的特征首先逐元素相加，再经过线性层和归一化操作计算权重 $\boldsymbol{\Omega}$：

$$\boldsymbol{\Omega} = \mathrm{Sigmoid}(\psi(\boldsymbol{F'}_{\mathrm{T}} + \boldsymbol{Y})) \tag{11}$$

式中：ψ 表示线性函数。自适应融合模块输出为：

$$\boldsymbol{F}_{out} = \boldsymbol{\Omega}\odot\boldsymbol{Y} + (1-\boldsymbol{\Omega})\odot\boldsymbol{F'}_{\mathrm{T}} \tag{12}$$

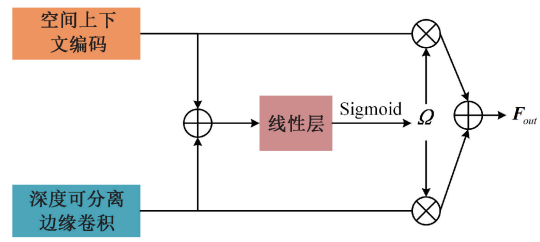


图 5　自适应融合模块

Fig. 5　Adaptive fusion module

2　实验结果与分析

2.1　实验环境配置

本文实验均在 Ubuntu5.15.0 系统下，使用 NVIDIA GeForce RTX 3090 GPU 显卡和 Pytorch 框架进行模型参数训练。表 1 显示了实验中的相关参数设置。

表 1　实验参数设置

Table 1　Experimental parameter settings

参数项	分类	分割
输入点云数量	1 024	2 048
k 值	20	20
训练轮次	300	300
训练批次	32	32
测试批次	16	16
初始学习率	0.01	0.003
优化器	SGD	Adam

2.2　数据集与评价指标

为评估本文所提出方法的有效性，在基准数据集 ModelNet40[25]、ScanObjectNN[26] 上进行形状分类任务，在 ShapeNet Part[27] 数据集上进行部件分割任务。

1）ModelNet40 数据集是一个常用的基准数据集，用于 3D 形状分类任务，包含 40 个类别的 12 311 个计算机辅助设计(CAD)模型，其中 9 843 个用于训练，2 468 个用于测试。

2）ScanObjectNN 数据集包含了 15 个类别的真实世界对象模型，总共包括 15 000 个对象，由于数据集中存在复杂背景、噪声和遮挡，该数据集对于现有点云分类任务更具有挑战性。

3）ShapeNet Part 数据集是由 16 个类别的 16 881 个模型组成，分为 50 个不同的部件标签，每个物体有 2~6 类部件标签，其中每个点均被标签标注。

在分类任务中使用总体准确率(overall accuracy, OA)和平均类别准确率(mean class accuracy, mAcc)作为评价指标，在分割任务中使用平均交并比(mean intersection over union, mIoU)。

$$\begin{cases} OA = \dfrac{TP}{TP + FN} \\ mAcc = \dfrac{1}{c}\sum_{i=1}^{c}\dfrac{TP_i}{TP_i + FN_i} \\ mIoU = \dfrac{TP}{TP + FP + FN} \end{cases} \tag{13}$$

式中：TP 代表正确预测的正样本数量；FN 代表错误预测的正样本数量；FP 代表错误地预测的负样本数量。而 TP_i 和 FN_i 分别代表每个类别中被正确和错误预测为正样本的数量。

2.3 分类结果与分析

为测试模型在分类任务中的有效性，在 Modelnet40 数据集上的分类结果如表 2 所示，其中加粗表示性能最优项。本文算法在 *OA* 和 *mAcc* 指标上分别达到 93.7% 和 91.0%，远超过了传统的基于体素和多视图的方法，甚至优于以大量点云数据作为输入的方法（PointNet++和 SO-Net[28]）。为了充分证明本研究算法的有效性，选取基于图卷积的算法（DGCNN、LDGCNN 和 3D-GCN[29]）、利用融合机制的算法（PRA-Net[30] 和 AdaptiveConv[31]）和基于 Transformer 的算法（PCT）进行对比分析，本算法在精度上均超越了以往方法。与新颖的算法（Test-Time[32]、IBT[33]、PointCAT[34]、文献[35]、LSPConv[36]、文献[37]）相比，本研究算法仍表现出一定的优越性。实验结果表明，改进的 Transformer 的两种特征提取方式与逐通道特征提取相结合的网络更有利于点云的特征学习。

为进一步验证模型在真实世界中的形状分类性能，在 ScanObjectNN 数据集上实验结果如表 3 所示，本文算法在 *OA* 和 *mAcc* 指标上达到最先进性能，分别为 83.2% 和 81.3%，证明本方法在有遮挡和背景噪声的情况下具有较强的鲁棒性。

2.4 部件分割结果与分析

为验证本文方法在点云分割任务中的有效性，在 ShapeNet Part 数据集上进行了部件分割实验，结果如表 4 所示，本文的平均交并比达到 86.0%，显著高于大多数主流方法。在 16 种不同物体中，帽子、吉他、电脑、杯子、火箭、滑板和桌子 7 种类别的分割精度在对比方法中实现最优。

表 2 Modelnet40 数据集上分类对比

Table 2 Comparison of classification on the Modelnet40

方法	输入	点数/($\times10^3$)	*mAcc*/%	*OA*/%
VoxNet	体素	–	83.0	85.9
MVCNN	多视图	–	–	90.1
PointNet	坐标	1	86.2	89.2
PointNet++	坐标+法线	5	–	91.9
SO-Net	坐标+法线	2	–	90.9
DGCNN	坐标	1	90.2	92.9
LDGCNN	坐标	1	90.3	92.3
3D-GCN	坐标	1	–	92.1
PRA-Net	坐标	1	90.6	93.2
AdaptiveConv	坐标	1	90.7	93.4
PCT	坐标	1	–	93.2
Test-Time	坐标	1	89.7	92.6
IBT	坐标	1	**91.0**	93.6
PointCAT	坐标	1	90.9	93.5
文献[35]	坐标	1	**91.0**	93.4
LSPConv	坐标	1	90.5	93.2
文献[37]	坐标	1	89.7	93.0
本文	坐标	1	**91.0**	**93.7**

表 3 ScanObjectNN 数据集上分类对比

Table 3 Comparison of classification on the ScanObjectNN dataset

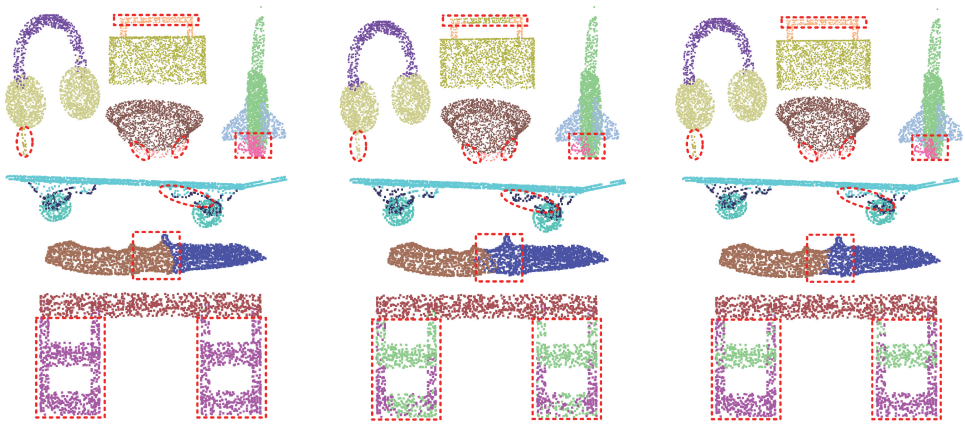
方法	输入	点数/($\times10^3$)	*mAcc*/%	*OA*/%
PointNet	坐标	1	63.4	68.2
PointNet++	坐标	1	75.4	77.9
DGCNN	坐标	1	73.6	78.2
PRA-Net	坐标	2	79.1	82.1
文献[35]	坐标	1	78.8	81.6
IBT	坐标	1	80.0	82.8
LSPConv	坐标	1	76.7	80.0
本文	坐标	1	**81.3**	**83.2**

表 4 ShapeNet Part 数据集上各类别及整体的部件分割对比

Table 4 Comparison of part segmentation for different categories and overall on the ShapeNet Part dataset (%)

方法	飞机	书包	帽子	汽车	椅子	耳机	吉他	刀	灯	电脑	摩托	杯子	手枪	火箭	滑板	桌子	*mIoU*
PointNet	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++	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	59.7	76.4	82.6	85.1
DGCNN	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
3D-GCN	83.1	**84.0**	86.6	77.5	90.3	74.1	90.9	86.4	83.8	95.6	66.8	94.8	81.3	59.6	75.7	82.8	85.1
PointBERT	**84.3**	84.8	88.0	**79.8**	**91.0**	**81.7**	91.1	**87.9**	**85.2**	95.6	**75.6**	94.7	**84.3**	63.4	76.3	81.5	85.6
本文	84.2	83.1	**88.6**	79.7	90.9	79.4	**92.0**	86.7	82.1	**96.0**	75.0	**95.5**	82.0	**64.3**	**81.9**	**84.2**	**86.0**

部件分割结果的可视化效果如图 6 所示，分别使用 DGCNN 算法和本文算法的结果可视化与真实标签的结果可视化进行对比，为突显可视化结果，使用红色虚线标注分割差异显著的部分。通过观察可以得出，不同的部件衔接区域的点云更具识别难度，通过对 7 个不同物体的部件分割可视化对比，本研究可以在物体不同部件的



(a) 真实值
(a) Groud truth

(b) DGCNN结果可视化
(b) Visualization of DGCNN method

(c) 本文算法结果可视化
(c) Visualization of ours method

图 6　部件分割对比可视化

Fig. 6　Visual comparison of parts segmentation

边界部分实现更准确的区分,相比于 DGCNN 模型达到了更精准的分割效果。通过以上定量和定性分析,验证了本文方法在分割任务中的优越性。

2.5　消融实验

基于 ModelNet40 数据集进行消融实验,选择合适的超参数并验证模型各个模块的有效性。

1)图结构中邻域点数 k 的选取

如表 5 所示,针对不同的 k 值分别测试 *mAcc* 和 *OA* 指标,当 k 为 15 时,*mAcc* 指标达到最大值,当 k 为 20 时,*OA* 指标达到最大值。总体来看,随着 k 值的增大,实验指标呈现先增后减的变化趋势。主要原因是当 k 值过低时,深度可分离边缘卷积获得的邻域结构信息不完整,片面的构造图结构导致局部特征提取不充分;当 k 值过高时,存在于点云形状边界的点会学习大量的整体形状信息,进而忽略边界邻域信息,且随着 k 值增长,计算复杂度增加的同时冗余信息也会随之增长。本文将 k 设置为 20 作为最佳超参数设置。

表 5　不同 k 值对模型的影响

Table 5　Effect of different k values on model performance

k	*mAcc*/%	*OA*/%
5	90.1	93.0
10	90.7	93.2
15	91.1	93.5
20	91.0	93.7
25	89.6	93.1

2)不同模块的消融实验

为了测试各个模块的有效性,分别对嵌入层、编码层中的深度可分离边缘卷积和空间上下文编码模块进行消融实验,结果如表 6 所示。

表 6　不同模块对网络性能的影响

Table 6　Effect of different modules on network performance

实验	嵌入层	深度可分离边缘卷积	空间上下文编码	自适应融合	*mAcc*/%	*OA*/%
1	×	√	√	√	89.4	92.9
2	√	×	√	×	89.6	93.1
3	√	√	×	×	89.5	93.2
4	√	√	√	+	90.1	93.3
5	√	√	√	√	**91.0**	**93.7**

实验 1,仅使用编码层进行特征学习,包括深度可分离边缘卷积和空间上下文编码学习两种特征,并对这两种不同的特征使用自适应融合模块进行融合。

实验 2,使用嵌入层和编码层中的空间上下文编码进行特征学习。

实验 3,使用嵌入层和编码层中的深度可分离边缘卷积进行特征学习。

实验 4,使用嵌入层和编码层中的深度可分离边缘卷积和空间上下文编码学习两种特征,采用特征相加的方式替代自适应融合模块进行特征融合(‘+’表示矩阵相加操作)。

实验 5,使用嵌入层和编码层中的两种编码策略和融合策略进行特征学习。

通过实验 1 与 5 对比,当去掉嵌入层时,*mAcc* 和 *OA* 分别下降 1.6%和 0.8%,证明了嵌入层对于提高模型性能的有效性。由于自适应融合模块依赖于深度可分离边缘卷积和空间上下文编码的两种输出特征,实验 2 和 3 使用单独的特征编码策略时均无法采用自适应融合模块。实验 2、3 与实验 5 对比可知,当分别去除深度可分离边缘卷积和空间上下文编码时,总体精度 *OA* 分别下

降 0.6%和 0.5%。其主要原因是深度可分离边缘卷积在图结构的基础上,通过深度卷积和逐点卷积两个阶段进行特征学习,相比于一般的卷积操作能更有效地捕获特征信息;空间上下文编码通过聚焦不同区域的图结构而捕获有效的上下文信息。因而,编码层提出的这两种编码方式对提升了网络精度是有效的。为了测试自适应融合模块的有效性,使用两种特征相加的操作替代自适应融合模块。对比实验 4 与 5 可知,由于特征相加时没有可学习参数,无法自适应调整特征权重,实验总体精度 *OA* 下降了 0.4%,证明了本研究自适应融合模块的有效性。

2.6 鲁棒性测试

目前点云分析的大部分方法在点数较少和存在噪声干扰的情况下表现不佳,为了测试本研究的鲁棒性和泛化能力,在 ModelNet40 数据集上分别针对点的密度和噪声进行测试。

1) 鉴于真实世界中点云数据常常受到密度不均匀的挑战,本研究采用随机下采样的方法,以不同程度地丢弃点数进行实验。图 7 为飞机、吉他和花瓶 3 个类别的物体在不同点数下可视化结果的展示。通过与 DGCNN 进行对比,本研究方法在任意点数下,总体精度 OA 均处于领先,究其原因,低密度条件下点云识别需要网络具备更强大的学习能力,本研究编码层结合了局部几何关系和全局上下文关系,获得全面且更具区分性的特征表示,凸显出较强的密度鲁棒性。

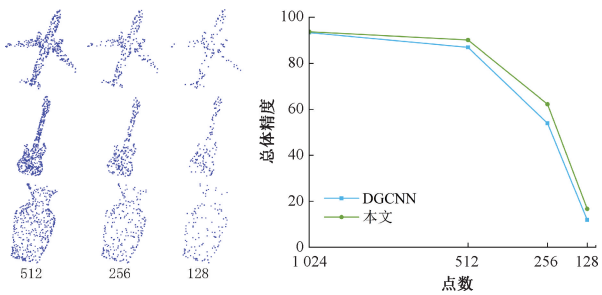


图 7 密度鲁棒性对比

Fig. 7 Density robustness comparison

2) 为模拟现实场景中的噪声干扰条件,本研究在原始点云数据添加不同幅度的噪声扰动来实现。这些噪声服从均值为 0 的高斯分布,其方差分别设置为 0.02、0.04、0.06、0.08 和 0.1。随着方差的增加,点云数据的扰动程度也随之增强,因此噪声的影响也随之加剧。实验结果如表 7 所示,以噪声干扰下的下降程度作为实验指标,具体计算公式为:

$$\text{下降程度} = \frac{OA(\text{噪声下}) - OA(\text{无噪声})}{OA(\text{无噪声})} \tag{14}$$

由表 7 实验数据对比可知,添加噪声的方差为 0.02 时,3D-GCN 算法的 OA 下降程度最小,在其他幅度噪声干扰下本文所提方法的抗噪性能最优。主要原因是研究模型在嵌入层中构造了丰富的低级几何关系,并在编码层中对高级特征实现有效学习,同时自适应融合模块提高了在噪声条件下显著信息的权重,因而,本文的方法具有良好的噪声鲁棒性。

表 7 噪声鲁棒性对比

Table 7 Comparison of the noise robustness

噪声方差	下降程度/%			
	3D-GCN	AdaptiveConv	DGCNN	本文
0.02	**0.7**↓	1.8↓	1.4↓	1.0↓
0.04	2.2↓	2.2↓	2.2↓	**1.8**↓
0.06	4.6↓	3.3↓	3.2↓	**2.9**↓
0.08	8.4↓	6.5↓	5.7↓	**5.5**↓
0.1	14.9↓	10.8↓	13.1↓	**10.2**↓

2.7 模型复杂度测试

模型复杂度也是衡量网络性能的重要指标,在本节实验中以参数量和模型大小对模型复杂度进行评估,并与 PointNet、PointNet++等算法进行对比,在 ModelNet40 数据集上的对比结果如表 8 所示。

表 8 模型复杂度比较

Table 8 Comparison of network complexity

方法	参数量/($\times 10^6$)	模型大小/MB	*OA*/%
PointNet	3.5	25.6	89.2
PointNet++	1.7	20.1	91.9
DGCNN	**1.8**	**6.9**	92.9
AdaptiveConv	1.9	7.5	93.0
PCT	2.9	11	93.2
本文	2.9	12.5	**93.7**

通过对比可知,本研究所提方法取得最优精度,但模型参数量和模型大小并未达到足够的轻量化,这是由于 Transfomer 网络需要对每个点分配相应的权重值,导致计算量增加。但相比于 PCT 算法,本文在参数量相同,模型大小增加 1.5 MB 的条件下,总体精度提高了 0.5%,证明了本文算法可以以较小的内存占用为代价实现更高的识别性能。同时,本文算法也为后续的模型优化提供方向,将致力于更高效的模型设计。

3 结 论

本研究提出了一种基于 Transformer 的逐通道点云分析网络。在图结构的基础上设计了深度可分离边缘卷积,相较于 DGCNN 中的边缘卷积更具灵活性,可以区别对待不同通道的信息。在 Transformer 的启发下设计了位置自适应编码和空间上下文编码模块,分别应用于低级坐标和高级特征的信息提取。同时,在高级空间中,自适应融合模块通过学习融合参数,以加强不同特征间的交互。所提网络实现了从低级坐标到高级特征的有序学

习,既可以在低级空间中捕获有效的几何关系,也可以在高级空间中融合关键上下文信息。最后通过大量实验证明所提方法的有效性和鲁棒性。

综上,本研究模型在点云分类分割任务中展现出良好性能,为点云分析提供一种新的思路,但模型的复杂度仍有改进空间,未来工作将对参数量和模型大小进一步优化,以实现更轻量化的模型设计。

参考文献

[1] WANG Y, CHAO W L, GARG D, et al. Pseudo-lidar from visual depth estimation: Bridging the gap in 3D object detection for autonomous driving[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 8445-8453.

[2] PEREZ-GONZALEZ J, AZAMAR C, PIÑA-RAMIREZ O. Automatic classification of Alzheimer's disease based on CPD brain point cloud registration for feature extraction[C]. 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). Berlin, Germany: IEEE, 2019: 812-815.

[3] AHNED S M, TAN Y ZH, CHEW C M, et al. Edge and corner detection for unorganized 3D point clouds with application to robotic welding[C]. 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 2018: 7350-7355.

[4] 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.

[5] RIEGLER G, OSMAN U A, GEIGER A. Octnet: Learning deep 3D representations at high resolutions[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 3577-3586.

[6] 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.

[7] FENG Y F, ZHANG Z ZH, ZHAO X B, et al. GvCNN: Group-view convolutional neural networks for 3D shape recognition[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 264-272.

[8] QI C R, SU H, MO K CH, 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.

[9] THOMAS H, QI C R, DESCHAUD J E, et al. KpConv: Flexible and deformable convolution for point clouds[C]. Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019: 6411-6420.

[10] WU W X, QI ZH G, LI F X. Pointconv: Deep convolutional networks on 3D point clouds[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 9621-9630.

[11] XU M T, DING R Y, ZHAO H SH, et al. PaConv: Position adaptive convolution with dynamic kernel assembling on point clouds[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 3173-3182.

[12] LI Y Y, BU R, SUN M CH, et al. PointCNN: Convolution on x-transformed points[J]. ArXive-prints, 2018.

[13] QI C R, YI L, SU H, et al. PointNET++: Deep hierarchical feature learning on point sets in a metric space[C]. Advances in Neural Information Processing Systems, 2017: 5099-5108.

[14] RAN H X, LIU J, WANG CH J. Surface representation for point clouds[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022: 18942-18952.

[15] XIANG T G, ZHANG CH Y, SONG Y, et al. Walk in the cloud: Learning curves for point clouds shape analysis[C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021: 915-924.

[16] WANG Y, SUN Y B, LIU Z W, et al. Dynamic graph CNN for learning on point clouds[J]. ACM Transactions on Graphics, 2018, 38(5): 1-12.

[17] ZHANG K G, HAO M, WANG J, et al. Linked dynamic graph CNN: Learning on point cloud via linking hierarchical Features[J]. 2019, DOI: 10.48550/arXiv.1904.10014.

[18] ZHAO H SH, JIANG L, FU C W, et al. Pointweb: Enhancing local neighborhood features for point cloud processing[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 5565-5573.

[19] HAN K, WANG Y H, CHEN H T, et al. A survey on vision transformer[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 45(1): 87-110.

[20] BROWN T, MANN B, RYDER N, et al. Language models are few-shot learners[J]. Advances in Neural Information Processing Systems, 2020, 33: 1877-1901.

[21] ZHAO H SH, JIANG L, JIA J Y, et al. Point transformer[C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021:

16259-16268.

[22] GUO M H, CAI J X, LIU ZH N, et al. Pct: Point cloud transformer[J]. Computational Visual Media, 2021, 7: 187-199.

[23] 于喜俊, 段勇. 基于 PointCloudTransformer 和优化集成学习的三维点云分类[J]. 电子测量与仪器学报, 2024, 38(6):143-153.
YU X J, DUAN Y. 3D point cloud classification based on PointCloudTransformer and optimized ensemble learning [J]. Journal of Electronic Measurement and Instrumentation, 2024, 38(6):143-153.

[24] YU X M, TANG L L, RAO Y M, et al. Point-bert: Pre-training 3D point cloud transformers with masked point modeling[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022: 19313-19322.

[25] WU ZH R, SONG SH R, 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.

[26] 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,2019: 1588-1597.

[27] 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 (ToG), 2016, 35(6): 1-12.

[28] 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,2018: 9397-9406.

[29] LIN ZH H, HUANG SH Y, WANG Y C F. Convolution in the cloud: Learning deformable kernels in 3D graph convolution networks for point cloud analysis [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition,2020: 1800-1809.

[30] 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.

[31] ZHOU H, FENG Y, FANG M, et al. Adaptive graph convolution for point cloud analysis[C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021: 4965-4974.

[32] VU T A, SARKAR S, ZHANG Z, et al. Test-time augmentation for 3D point cloud classification and segmentation[C]. 2024 International Conference on 3D Vision (3DV). IEEE, 2024: 1543-1553.

[33] LI Z H, GAO P, YUAN H, et al. Exploiting inductive bias in transformer for point cloud classification and segmentation[C]. 2023 IEEE International Conference on Multimedia and Expo Workshops (ICMEW). IEEE, 2023: 140-145.

[34] YANG X CH, JIN M Z, HE W J, et al. PointCat: Cross-attention transformer for point cloud[J]. ArXiv preprint arXiv:2304.03012, 2023.

[35] 芦新宇, 杨冰, 叶海良, 等. 基于局部-非局部交互卷积的 3D 点云分类[J]. 模式识别与人工智能, 2022, 35(2): 141-149.
LU X Y, YANG B, YE H L, et al. 3D point cloud classification based on local-nonlocal interactive convolution [J]. Pattern Recognition and Artificial Intelligence, 2022, 35(2): 141-149.

[36] ZHNAG H M, WANG K, ZHONG CH, et al. LSPConv: Local spatial projection convolution for point cloud analysis[J]. PeerJ Computer Science, 2024, 10: e1738.

[37] 武斌,刘溢安,赵洁. 结合空间结构卷积和注意力机制的三维点云分类网络[J]. 中国图象图形学报,2024, 29(2):0520-0532.
WU B, LIU Y AN, ZHAO J. A 3D point cloud classification network combining spatial structural convolution and attention mechanism[J]. Journal of Image and Graphics, 2024, 29(2): 520-532.

作者简介

冯凯浩,2022 年于辽宁工程技术大学获得学士学位,现为辽宁工程技术大学硕士研究生,主要研究方向为基于深度学习的目标检测和三维点云分类分割。
E-mail: fengkaihao_666@163.com

Feng Kaihao received his B. Sc. degree from Liaoning Technical University in 2022. He is currently a M. Sc. candidate at Liaoning Technical University. His main research interests include object detection based on deep learning and the classification and segmentation of 3D point clouds.

陶志勇(通信作者),分别在 2000 年、2005 年和 2015 年于辽宁工程技术大学分别获得学士学位,硕士学位和博士学位,现为辽宁工程技术大学教授,主要研究方向为智能信息处理。
E-mail: xyzmail@126.com

Tao Zhiyong (Corresponding author) received his B. Sc. degree, M. Sc. degree and Ph. D. degree all from Liaoning Technical University in 2000, 2005 and 2015, respectively. He is currently a professor at Liaoning Technical University. His main

research interests include intelligent information processing.

李衡,2021 年于辽宁工程技术大学获得学士学位,现为辽宁工程技术大学硕士研究生,主要研究方向为基于深度学习的点云分类分割算法研究。

E-mail: PaperLH@ 163. com

Li Heng received his B. Sc. degree from Liaoning Technical University in 2021. He is currently a M. Sc. candidate at Liaoning Technical University. His main research interests include the classification and segmentation of 3D point clouds.

李铭朗,2020 年于湖南理工学院获得学士学位,现为辽宁工程技术大学硕士研究生,主要研究方向为基于深度学习的行人重识别和目标检测。

E-mail: 472220498@ stu. lntu. edu. cn

Li Minglang received his B. Sc. degree from Hunan Institute of Science and Technology in 2020. He is currently a M. Sc. candidate at Liaoning Technical University. His main research interests include pedestrian re-identification and object detection based on deep learning.

林森,分别在 2003 年和 2006 年于辽宁工程技术大学分别获得学士学位和硕士学位,2013 年在沈阳工业大学获得博士学位,2019 年于中国科学院沈阳自动化研究所博士后出站,现为沈阳理工大学副教授,主要研究方向为图像信息处理、机器视觉、模式识别、生物特征识别。

E-mail: lin_sen6@ 126. com

Lin Sen received his B. Sc. degree and M. Sc. degree from Liaoning Technical University in 2003 and 2006, Ph. D. degree from Shenyang University of Technology in 2013, post-doctorate in Shenyang Institute of Automation, Chinese Academy of Sciences in 2019. He is currently an associate professor at Shenyang Ligong University. His main research interests include image information processing, machine vision, pattern recognition, biometric recognition.