文章编号:2096-1472(2023)-06-24-06

基于局部邻域 Transformer 的点云特征提取方法

张海博¹,沈 洋^{1,2},许 浩²,包艳霞^{2,3},刘 江²

(1.浙江理工大学计算机科学与技术学院,浙江 杭州 310018;
 2.丽水学院工学院,浙江 丽水 323000;
 3.浙江掌信传媒科技有限公司,浙江 丽水 323020)



⊠ linhai992@163.com; 1178409849@gg.com; oah ux@126.com; 82240849@gg.com; elecliu@lsu.edu.cn

摘 要: 受点云非结构化、无序性等特性的影响,一些现有的自注意力方法不能充分提取上下文语意特征, 基于此提出了一种用于点云特征提取的局部邻域 Transformer(Local Neighborhood Transformer, LNT)。首先, 通过最远点采样(FPS)和 K 最近邻算法(KNN)对点云进行邻域划分。其次,结合相对位置编码,在各个邻域内计 算局部自注意力,达到线性计算复杂度。最后,通过连接操作以及线性层捕获点云局部特征。此外,设计了点云 多特征融合方法对各层特征信息进行聚合,以提高模型的性能。实验结果表明,该方法在 ModelNet40 数据集中 分类的总体精度可达到 93.3%,比 PCT 提升了 0.1%;类平均精度可达到 92.0%, 比 PointMLP 提升了 0.6%。 同时,在 ShapeNet 数据集中的点云分割结果也是有效的。

关键词:点云处理,Transformer,特征融合,神经网络中图分类号:TP183 文献标识码:A

Point Cloud Feature Extraction Method Based on Local Neighborhood Transformer

ZHANG Haibo¹, SHEN Yang^{1,2}, XU Hao², BAO Yanxia^{2,3}, LIU Jiang²

(1.School of Information Science and Technology, Zhejiang Sci-Tech University, Hangzhou 310018, China; 2.School of Engineering, Lishui University, Lishui 323000, China;

3.Zhejiang Zhangxin Media Technology Co., Ltd., Lishui 323020, China)

🖂 linhai992@163.com; 1178409849@qq.com; oah_ux@126.com; 82240849@qq.com; elecliu@lsu.edu.cn

Abstract: Affected by the unstructured and disordered nature of point clouds, some existing self-attentive methods cannot fully extract contextual semantic features, based on which a Local Neighborhood Transformer (LNT) for point cloud feature extraction is proposed. Firstly, neighborhood partitioning of point clouds is achieved through Farthest Point Sampling (FPS) and K-nearest Neighbor (KNN) algorithm. Secondly, local self-attention is computed within each neighborhood by combining relative position encoding to achieve linear computational complexity. Finally, the local features of the point cloud are captured by the join operation as well as the linear layer. In addition, a point cloud multi-featured fusion method is designed to aggregate the feature information of each layer to improve the performance of the model. The experimental results show that the overall accuracy of the method can reach 93.3% in the ModelNet40 dataset for classification, which is 0.1% higher than PCT; the class average accuracy can reach 92.0%, which is 0.6% higher than PointMLP. Meanwhile, the point cloud segmentation results are also valid in the ShapeNet dataset.

Keywords: point cloud processing; Transformer; feature fusion; neural network

1 引言(Introduction)

三维点云近年来在不同领域的应用越来越广泛,如人脸 识别、机器人技术和生成技术等^[1-5],点云特征提取方法作为

基金项目:浙江省自然科学基金项目(LY21F02004).

以上应用开发的技术基石具有重要的研究意义。点云是表示 三维对象的点集,具有稀疏性、非结构化、无序性等特性, 在数据结构上与结构化的二维图像不同,这使得在计算机视 觉中已经成熟的深度学习方法^[6]无法直接适配到点云处理中。 针对该问题,近年来出现了各种关于点云的深度学习方法。 基于投影的方法^[7,8]使用点云的二维投影进行推理预测,这类 方法会导致信息丢失,并且其性能因投影视点变化而产生差 异。基于体素的方法^[9]将点云转换为体素数据结构,应用卷 积进行处理,存在由于体素化而失去几何细节的问题。基于 点的方法^[10-11]在点云原始数据上进行研究,减少了信息的丢 失。基于 Transformer(一种采用自注意力机制的深度学习模 型)的方法^[12-15]缺乏对点云局部结构信息的利用,不能充分提 取上下文语意特征且存在计算复杂度高的问题。

Transformer 作为集合运算符,适合处理非结构化的、无 序的点集,具有很好地适配点云数据结构的特点。因此,本 文提出了一种用于点云处理的局部邻域 Transformer(LNT), 该方法是对点云进行邻域划分后,在局部邻域内计算自注意 力,达到线性复杂度。此外,通过在自注意力计算过程中加入 有效的相对位置编码,在网络构建中使用多特征融合方法,进 一步提高模型性能。实验结果表明:LNT 模型在 ModelNet40 数据集和 ShapeNet 数据集中表现出较好的性能优势。

2 方法设计(Method design)

2.1 设计思路

Transformer 能较好地提取全局特征,但缺乏对局部特征 的学习,因此研究人员针对点云数据结构进行改进,提出了 一个用于三维点云处理的局部邻域 Transformer。在此基础 上,构建了一种新的用于点云学习的局部邻域 Transformer 网络。LNT 的关键思想是利用 Transformer 固有的置换不变 性,避免需要定义点云数据的顺序,引入局部邻域化的思想, 通过局部自注意力机制进行特征学习。

局部邻域 Transformer 网络实现流程图如图 1 所示。首 先通过最远点采样算法和 KNN 算法对点 云进行邻域分组, 然后在每个局部邻域内使用自注意力机制得到特征向量。使 用金字塔式的网络结构,逐步形成全局特征表示。





2.2 局部邻域 Transformer

给定输入为一个由 N 个点组成的点云 $P \in \mathbb{R}^{N \times D}$,其中 D 表示特征维度。使用最远点采样和 KNN 对点云 P 进行邻 域划分 $P = \{P_s \in \mathbb{R}^{K \times D}, s = 1, 2, 3, \dots, S\}$, S 表示最远点采样 的采样点数, K 表示 KNN 算法中邻域内点的个数,此时 P 被分为S个不同的邻域(每个邻域内有 K 个点)。我们需要设 置合理的 K、S,使得不同的邻域之间能够有相互重叠的部 分,进一步加强邻域间的关系。接下来,需要对邻域内的点 进行自注意力的计算,由于点本身是三维坐标,减法更能表 现出点与点之间的关系,并且研究人员通过实验测试了乘法 和减法的效果(见本文"3.4.2"部分),最终选择使用减法计 算自注意力权重:

 $A_s = \sigma(Q(P_s) - K(P_s) + \delta) \odot(VP_s) + \delta)$ (1) 其中, s 表示第 s 个邻域, Q(•)、K(•)、V(•)是线性变 换函数,其计算方式如公式(2)至公式(4)所示。研究人员添 加了相对位置编码 δ ,用于补足模型对于点云结构信息的学 习, σ 为 Softmax 函数对权重进行归一化处理。

$$\mathbf{Q}(\mathbf{P}_s) = \mathbf{P}_s \mathbf{W}_q \in \mathbb{R}^{K \times D'}$$
(2)

 $K(\boldsymbol{P}_{s}) = \boldsymbol{P}_{s} \boldsymbol{W}_{k} \in \mathbb{R}^{K \times D'}$ (3)

$$V(\boldsymbol{P}_s) = \boldsymbol{P}_s \boldsymbol{W}_v \in \mathbb{R}^{K \times D'}$$
(4)

其中, W_q 、 W_k 、 W_v 为可学习的权重参数。

$$\boldsymbol{F}_{s} = \boldsymbol{A}_{s} + \boldsymbol{P}_{s} \tag{5}$$

$$\mathbf{F} = MLP(Stack(\mathbf{F}_1, \mathbf{F}_2, \cdots, \mathbf{F}_S)) \in \mathbb{R}^{S \times D'}$$
(6)

随后,每个邻域内的注意力特征与输入进行残差连接后, 对所有邻域的输出特征(F_1 , F_2 ,..., F_s)进行堆叠,再经过多 层感知器得到输出特征 F。局部邻域 Transformer 结构如图 2 <u>所</u>示,其中, h 表示对头注意力的头数。



图 2 局部邻域 Transformer Fig. 2 Local neighborhood Transformer

局部邻域 Transformer 促进了局部邻域内特征向量之间 的信息交换,同时邻域间的重叠区域也促进了邻域间的信息 交换。最终,所有数据点产生新的特征向量作为输出。其中, 研究人员在注意力计算环节使用了多头自注意力机制^[12],这 让模型同时关注不同表示子空间的信息。总的来说,为了使 Transformer适合点云的处理,研究人员对它进行了改进,构 造了一个核心为 Transformer 的局部邻域自注意力层。该层 集成了采样分组、局部自注意力、线性变换和一个残差连接 (缓解网络退化问题)。该层能够减少特征数量,提升特征维 度,为构建金字塔式的网络结构做铺垫。多层感知器中的 Max Pooling运算符解决了点云置换不变性的问题。

在计算复杂度方面,由于注意力的计算集中在注意力权 重分支,因此本文采用减法而非乘法,在单个邻域内的时间 复杂度为 O(KD),那么整体的时间复杂度则为 O(SKD), 优于传统注意力的二次时间复杂度,基本达到线性时间复 杂度。

2.3 多特征融合

随着网络的深入,模型底层细节信息逐渐丢失,为了进 一步提高模型的性能,研究人员加入了多特征融合模块。由 于 LNT 能够减少特征个数 N,提升特征维度 D,因此每层 输出的特征尺度都不相同,为了能够进行融合,首先需要将 特征个数 N 进行统一,然后对特征进行合并。

第一方案是使用随机下采样的方式,使得所有底层(前两 层)输出特征个数与第三层的输出特征个数统一,然后对这些 特征进行合并,但模型精度并没有提升,研究人员分析原因 后发现,精度与随机下采样和局部特征提取的方式有关,这 导致特征之间的关系并不对应,使合并后特征变得混乱,导 致最终模型效果不好。

为了解决这一问题,研究人员修改了下采样的方式,将 其改为对应位置的下采样,具体来说就是由于 LNT 的注意力 计算是将邻域内的特征聚合到中心点上的,保留了中心点的 索引,因此可以通过高层的中心点索引在底层中找到对应的 点,并在底层只保留这些点的特征,用于合并,多特征融合 模块如图 3 所示。



Fig. 3 Multi-featured fusion module

点云通过两层 LNT 后,点的个数逐渐减少,特征维度逐渐增加,如图 3 中右指向的箭头(步骤 1)。为了实现特征融合,通过已知高层中心点的索引,可以找到底层特征中对应的点特征,并只保留白色的点特征(白色点),如图 3 左指向箭头(步骤 2)。最终将每层的对应点进行合并,完成融合(步骤 3)。

2.4 相对位置编码

Transformer 本质上是集合算子,计算全局特征,缺乏对 点云局部相对位置信息的利用。在使用自注意力处理图像的 任务中,位置编码可以使得模型更好地利用数据的局部结构 信息。位置编码方案很多是人工制定的,例如基于正弦和余 弦函数或归一化范围值^[12]。还有一些是通过学习得到的,例 如在 NAT^[16]中,对窗口内像素位置编号后,通过可学习的 偏置将位置信息加入注意力权重中,其位置偏置用于规则二 维网格图像,不适用于处理离散的三维点云。

在点云处理中加入相对位置编码,可以加强模型对于点 云局部结构的学习,但编码器太简单,不能充分学习结构特 征,也不可设计复杂造成过拟合。

本文提出的相对位置编码,通过将两点坐标做差,引入 可训练、参数化的多层感知器学习点之间的相对位置信息, 使模型达到更好的效果,弥补了自注意力机制在点云处理中 的缺点,解决点云的旋转不变性问题。研究人员将位置编码 δ定义如下:

 $\delta = MLP(P_i - P_j)$ (7) 其中, P_i 和 P_j 为两个点的三维坐标。编码器是线性层和激 活层组成的多层感知器。相对位置编码可以强调模型更加注 重提取点云三维结构上的特征,这在实验中(表 4)也得到了 验证:添加相对位置编码的模型与不添加的相比精度更高。 因此、研究人员在公式(1)的两个分支中都添加了可训练的相 对位置编码。

5 局部邻域 Transformer 网络

基于 LNT,研究人员构造了完整的用于三维点云处理的 网络,如图 4 所示。整个网络模型根据任务不同(分类任务和 分割任务),网络后半段结构有所不同,但编码器的网络结构 是一致的。

编码器:LNT 能将输入点编码为一个新的高维特征向 量,同时降低特征的基数。我们将它作为编码器的基础,以 金字塔式的结构构建编码器。K 表示 KNN 算法中找到邻近 点的个数。将点云输入 3 个堆叠的LNT中,在局部邻域内学 习每个点丰富的语义特征表示,然后将 3 层LNT 的输出进行 多特征融合,再经过一个全局 Transformer 层,最终生成输 出特征。在全局 Transformer 层中不进行邻域的划分,因此 K 值设定为1。在局部邻域 Transformer 网络框架中,以输入 2 048 个点为例,编码阶段每层降采样目标点数为(512,128, 64,1),每层中 K 值设定为(16,16,32,1)。全局自注意力是 N 个点与其他 N-1 个点计算自注意力权重,因此其空间复 杂度为 O(N²),而本文的局部邻域自注意力是 N 个点与其 邻域内 K 个点计算自注意力权重,所以模型的参数量大幅减 少,达到 O(NK)。

分类: 网络的细节如图 4 中的下分支所示。将点云数据 分类为 N_e 个对象类别(如飞机、桌子、椅子)。将编码器输 出的特征向量输入两个前馈神经网络 LBRD(线性层、批归一 化层和 Dropout 层结合的模块,其中 Dropout 率设定为 0.5), 最终由线性层预测最终分类分数,得分最高的类确定为类 标签。

分割: 网络的细节如图 4 中的上分支所示。将点云分割 成 N_p 个部件(如将飞机分割为机体、两翼、尾翼等部件), 研究人员需要预测每个点的部件标签。在分割任务中采用 U-Net 编码器-解码器网络结构,分割网络编码器的架构与分 类网络的架构相同,编码器和解码器的基本结构(网络基数和 特征维度相同)是对称的,编码层的特征通过桥接到解码层辅 助上采样。

图 4 中 ULBR 模块包含上采样层、线性层、归一化层和 激活层。上采样层的主要功能是将特征从低分辨率的输入点 集映射到高分辨率的点集上。输入特征经过三线性插值将特 征映射到更高分辨率的点集上,然后将插值后的特征,与桥 接的编码器阶段的特征进行结合,最终通过线性层、归一化 层和激活层得到输出特征。

经由 4 个 ULBR 模块,再通过一个线性层,预测输入点 云的逐点分割分数,每个点的部件标签由具有最大分数的部 件标签确定,最终得到分割后的点云。



图 4 局部邻域 Transformer 网络架构

Fig. 4 Local neighborhood Transformer network architecture

3 实验和结果(Experiment and results)

研究人员在点云分类和分割任务中评估了局部邻域 Transformer 网络设计的有效性,并与其他方法进行了比较。 针对点云形状分类任务,在 ModelNot40 数据集^[17]上进行测试。 针对点云分割任务,在 ShapeNetPart 数据集^[18]进行测试。

研究人员使用 PyTorch 实现了局部邻域 Transformer 网络。在实验过程中,均使用负对数似然损失函数和 Adam 优化器进行训练,初始学习率设置为 0.001,权重衰减设置为 0.000 1,训练了 200 个周期。

3.1 ModelNet40 数据集分类实验

ModelNet40 数据集包含 40 个类别共 12 311 个 CAD 模型, 它被广泛应用于点云形状分类模型的基准测试。为保证公平性,研究人员使用官方设定的 9 843 个模型用于训练, 2 468 个模型用于测试,使用类平均精度(*mAcc*)和总体精度(*OA*)进行评估。模型在两张 RTX3090 显卡(24 GB 显存)上运行, *mini-batch size* 设定为 32。实验中,除了使用重采样的方式增强数据,没有使用别的方式增强数据。原则上如果附加其他训练技巧,训练精度还可以进一步提升。

实验结果如表 1 所示。本文提出的局部邻域 Transformer 在 ModelNet40 数据集上的总体精度(OA)为 93.3%,类平均 精度(mAcc)为 92.0%,优于其他模型。

本文所提模型的类平均精度(mAcc)是最高的,达到 92.0%,远超其他模型。这说明对于个别训练集数量较小的 类别,本文所提模型的训练效果也很好,模型泛化性强,能 够提取到丰富的特征。

表1 在 ModelNet40 数据集上的形状分类结果

Tab.1 Shape classification results on ModelNet40 dataset

方法	输入数据类型	类平均精度/%	总体精度/%
3DshapeNets ^[17]	voxel	77.3	84.7
VoxNet ^[9]	voxel	83.0	85.9
Subvolume ^[17]	voxel	86.0	89.2
MVCNN ^[7]	image	—	90.1
PointNet ^[10]	point	86.2	89.2
A-SCN ^[19]	point	87.6	90.0
Set Transformer ^[20]	point	_	90.4
PAT ^[21]	point	—	91.7
PointNet++ ^[U]	point	—	91.9
PointCNN ^[22]	point	88.1	92.2
DGCNN ^[23]	point	90.2	92.2
PointWeb ^[24]	point	89.4	92.3
PointConv ^[25]	point	—	92.5
Point2Sequence ^[26]	point	90.4	92.6
KPConv ^[27]	point	_	92.9
PCT ^[14]	point	_	93.2
PointMLP ^[28]	point	91.4	94.5
LNT(本文)	point	92.0	93.3

3.2 ShapeNetPart 数据集部件分割实验

点云部件分割是一项具有挑战性的任务,旨在将三维模型分割为多个不同的部件。本文所提模型在 ShapeNetPart 数据集进行了实验评估,该数据集由 2 048 个点的 16 880 个实例对象组成,共有 16 个类别和 50 个部件标签。训练中按照 官方划分,14 006 个对象用于训练,另外 2 874 个对象用于 测试,其中每个实例对象分割为 2—6 个部件。分割任务的模型中,批处理大小、训练周期和学习率均与分类任务中的训 练设置相同。

研究人员使用总体平均交并比(*mIoU*)作为评估指标,并 列举出了每个类平均交并比。实验中,将本文所提模型与其 他模型(如 PointNet、PointNet++、M-Conv等)进行了比 较,每个类别中的最高精度用加粗字体标识出来,结果如表2所示。结果表明,LNT模型的总体分割精度实现了超越, 类别精度在包、汽车、刀等7个类别上取得了最佳结果。图5 展示了分割效果可视化实例,第一行为未进行分割处理的对象,第二行为LNT模型分割预测的结果。

表2 在 ShapeNetPart 数据集上的部件分割结果

Tab.2 Part segmentation results on ShapeNetPart dataset

方法	mIoU	飞机	包	帽子	汽车	椅子	耳机	吉他	Л
PointNet ^[10]	83.7	83. 4	78.7	82.5	74.9	89.6	73.0	91.5	85.9
SO-Net ^[29]	84.9	82.8	77.8	88	77.3	90.6	73.5	90.7	83.9
PointNet++[11]	85.1	82.4	79.0	87.7	77.3	90.8	71.8	91.0	85.9
3DContextNet ^[30]	84.3	83.3	78.0	84.2	77.2	90.1	73.1	91.6	85.9
M-Conv ^[31]	84.3	82.3	81.1	88.3	76.2	89.5	79.5	90.5	86.3
LNT(本文)	85.2	82.5	83.8	87.4	77.9	90.7	73.5	90.7	86. 9
方法	灯	电脑	摩托	杯子	手枪	火箭	滑板	桌子	
PointNet ^[10]	80.8	95.3	65.2	93.0	81.2	57.9	72.8	80.6	
SO-Net ^[29]	82.8	94.8	69.1	94.2	80.9	53.1	72.9	80.3	
PointNet++[11]	83.7	95.3	71.6	94.1	81.3	58.7	76.4	82.6	
3DContextNet ^[30]	81.4	95.4	69.1	92.3	81.7	60.8	71.8	81.4	
M-Conv ^[31]	82.5	95.7	70.8	95.0	81.1	62.5	80.4	81.8	
LNT(本文)	84. 3	95.5	68.8	95.0	82.1	60.4	75.5	82.7	



图 5 ShapeNet 数据集分割可视化实例

Fig. 5 Example of ShapeNet dataset segmentation visualization

3.3 模型空间复杂度

当前,使用自注意力机制的其他模型大多使用的全局自注意力,导致这些模型在注意力部分的空间复杂度为 O(N²)。相比之下,本文所提的LNT模型在局部邻域内计算 自注意力,可以让模型的参数量大幅减少,达到O(NK²), 其中N为点的个数,K为邻域内点的个数(K<32)。

表 3 以分割任务为例,展示了 LNT 模型的参数量,并与 其他模型进行了对比,表明 LNT 在空间复杂度上的优势。

表3 模型参数量对比

Tab.3 Comparison of model parameters

方法	参数量/MB
KPConv ^[27]	15.0
Point Transformer ^[15]	7.8
CBL ^[32]	18.6
PointNeXt-L ^[33]	7.1
LNT(本文)	6.0

3.4 消融性实验

3.4.1 相对位置编码的影响

通过引入可训练的、参数化的位置编码器来学习点与点 之间的相对位置信息,可以使得局部自注意力机制更好地利 用点云的结构信息;并以 ModelNet40 数据集上的分类任务为 例,具体研究了相对位置编码添加位置对模型的影响,结果 如表 4 所示。_____

表4 添加相对位置编码的实验对比

Tab.4 Experimental comparison of adding relative position coding

位置编码	总体精度/%
None	92.3
relative for feature	92.5
relative for attention	92.6
Relative	93.3

人表4中可以看到:如果没有位置编码,模型的精度就会显著下降。采用相对位置编码时的模型精度高于无位置编码。可见,相对位置编码可以提高网络模型的性能。当相对位置编码只添加到注意力权重计算分支或仅添加到特征转换分支时,模型精度会不如整体都添加时的模型精度。可见,在两个分支都添加相对位置编码是很重要的。

3.4.2 自注意力计算方式的影响

Transformer 中自注意力权重计算的计算方式有多种,如 减法、乘法等,其中乘法($W_q^T W_k$)是 Transformer 通用的计 算方式。但是,研究人员认为减法的计算方式($W_q - W_k$)要更 适合点云自注意力的计算,因为点云本身是三维空间中的位 置信息,减法能更好地提取点与点之间的位置关系,同时更 能突出点与点之间的差异性,更有利于特征的学习。研究人 员通过实验证明了这一点,使用乘法计算注意力权重的总体 精度为 92.5%,与使用减法计算注意力权重的总体精度 93.3%相比,低了 0.8%。

3.4.3 K 值的设定

超参数 K 表示最近邻采样点的数量即邻域的范围大小, K 值的合理设定是决定模型性能的重要一环,研究人员需要 设定合适的 K 值保证邻域之间能够有相互重叠的部分,以此 进一步加强邻域间的关系。在网络的底层,研究人员希望模 型专注于低级特征的学习,因此在前两层设定的 K 值较小; 在网络的高层,设定较大的 K 值对特征进行聚合,形成全局 特征。

可以通过实验寻找每层合适的 *K* 值设定,结果见表 5 所示。最合理的每层 *K* 值设定为(16,16,32,1)。

表5 K 值的选择

Tab.5 Selection of K value

每层 K 值的设定	总体精度/%
(16,16,16,1)	92.5
(32,32,32,1)	92.3
(16,16,48,1)	92.8
(8,16,32,1)	91.9
(16,16,32,1)	93.3

4 结论(Conclusion)

本文提出了一种用于点云特征提取的局部邻域 Transformer模型,该模型从点云数据结构的特点出发,利用 Transformer自注意力机制在点云局部几何结构中学习特征, 能够充分提取上下文语意信息,提升特征提取能力,并达到线 性计算复杂度。同时,多特征融合方法能有效地利用模型底层 特征,进一步提升模型的性能。实验结果表明,LNT模型在 ModelNet40数据集形状分类任务中的总体精度可达到 93.3%, 类平均精度可达到 92.0%,优于现有其他模型,并且在分割低 务中点云分割结果也是有效的。同时,实验证明了模型中的相 对位置编码、局部自注意力计算方式及 K 值设定的合理性, 未来,研究人员会对模型的网络架构进行改进,在点云分割、 语义分割、点云生成等任务中做进一步的扩展应用。

参考文献(References)

- 潘海鵰,郝慧,苏雯.基于注意力机制与多尺度特征融合的人 脸表情识别[J].浙江理工大学学报(自然科学版),2022,47(3): 382-388.
- [2] 高工,杨红雨,刘洪.基于特征融合的三维人脸点云质量判断[]]. 计算机应用,2022,42(3):968-973.
- [3] 张磊,徐孝彬,曹晨飞,等.基于动态特征剔除的图像与点云融合的机器人位姿估计方法[J].中国激光,2022,49(6):58-69.
- [4] 崔扬.移动 LiDAR 点云室内三维结构化重建方法和关键技术研究[]].测绘学报,2021,50(7):990.
- [5] 卢林鹏,关柏良,林淑金.面向点云三维重建的空间感知对抗 神经网络[J].计算机工程与科学,2022,44(7):1247-1255.
- [6] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6):84-90.
- [7] SU H, MAJI S, KALOGERAKIS E, et al. Multi-view convolutional neural networks for 3d shape recognition[C]// MORTENSEN E.

Proceedings of the IEEE International Conference on Computer Vision. Santiago: IEEE, 2015:945-953.

- [8] WEI X, YU R, SUN J. View-gcn: View-based graph convolutional network for 3d shape analysis[C]// MORTENSEN E. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2020:1850-1859.
- [9] MATURANA D, SCHERER S. Voxnet A 3d convolutional neural network for real-time object recognition[C]// YAMANE K. Proceedings of the IEEE International Conference on Intelligent Robots and Systems. Hamburg: IEEE, 2015:922-928.
- [10] QI C R, SU H, MO K, et al. Pointnet: Deep learning on point sets for 3d classification and segmentation[C]// MORTENSEN E. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017:652-660.
- [11] QI C R, YI L, SU H, et al. Pointnet++ : Deep hierarchical feature learning on point sets in a metric space[J]. Advances in Neural Information Processing Systems, 2017, 30:5100-5109.
- [12] VASWANI A, SHAZPER N, PARMAR N, et al. Attention is all you need[J]. Advances in Neural Information Processing Systems, 2017, 40:5999-6009.
- [13] HAN X F, RUANG Y J, XIAO G Q. Point cloud learning with transformer[DB/OL]. (2021-04-28) [2022-07-12]. https://arxiv. 53yd.com/abs/2202.07123.
- [14] GUO M H, CAI J X, LIU Z N, et al. Pct: Point cloud transformer[J]. Computational Visual Media, 2021, 7(2):187-199.
- [15] ZHAO H, JIANG L, JIA J, et al. Point transformer[C]// MORTENS-EN E. Proceedings of the IEEE/CVF International Conference on Computer Vision. New York: IEEE, 2021:16259-16268.
- [16] HASSANI A, WALTON S, LI J, et al. Neighborhood attention transformer[DB/OL]. (2022-04-14) [2022-07-12]. https://arxiv. 53yu.com/abs/2204.07143.
- [17] WU Z, SONG S, KHOSLA A, et al. 3d shapenets: A deep representation for volumetric shapes[C]// LEOTTA M. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Boston: IEEE, 2015:1912-1920.
- [18] YI L, KIM V G, CEYLAN D, et al. A scalable active framework for region annotation in 3d shape collections[J]. ACM Transactions on Graphics, 2016, 35(6):1-12.
- [19] XIE S, LIU S, CHEN Z, et al. Attentional shapecontextnet for point cloud recognition [C] // MORTENSEN E. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018:4606-4615.
- [20] LEE J, LEE Y, KIM J, et al. Set transformer: A framework for attention-based permutation-invariant neural networks [DB/OL].
 (2018-10-01) [2022-07-12]. https://arxiv.53yu.com/abs/1810.00825.
 (下接第 35 页)