

文章编号:1001-5078(2024)01-0017-07

· 激光应用技术 ·

# 基于 Transformer 的道路场景点云分类与分割方法

马庆禄<sup>1</sup>, 孙 泉<sup>1</sup>, 黄筱潇<sup>1</sup>, 王江华<sup>2</sup>

(1. 重庆交通大学交通运输学院, 重庆 400074; 2. 重庆奉建高速公路有限公司, 重庆 401120)

**摘要:**针对多目标识别过程中点云分类和分割精度不高的问题,提出了一种基于改进 Transformer 模型的点云分类与分割方法 DRPT(Double randomness Point Transformer),该方法在 Transformer 模型卷积投影层创建新的点嵌入,利用局部邻域的动态处理在数据特征向量中持续增加全局特征属性,从而提高多目标识别中点云分类和分割的精度。实验中采用了标准基准数据集(ModelNet40、ShapeNet 部分分割和 SemanticKITTI 场景语义分割数据集)以验证模型的性能,实验结果表明:DRPT 模型的 pIoU 值为 85.9%,比其他模型平均高出 3.5%,有效提高了多目标识别检测时点云分类与分割精度,是对智能网联技术发展的有效支撑。

**关键词:**点云识别;道路场景;点云分割;点集嵌入;Transformer

**中图分类号:**TP391.4;TN249 **文献标识码:**A **DOI:**10.3969/j.issn.1001-5078.2024.01.003

## A Transformer-based classification and segmentation approach for classifying and segmenting road field attraction clouds

MA Qing-lu<sup>1</sup>, SUN Xiao<sup>1</sup>, HUANG Xiao-xiao<sup>1</sup>, WANG Jiang-hua<sup>2</sup>

(1. School of Traffic and Transportation, Chongqing Jiaotong University, Chongqing 400074, China;

2. Chongqing Fengjian Expressway Co. LTD, Chongqing 401120, China)

**Abstract:**To address the problem of low accuracy of point cloud classification and segmentation in the process of multi-target recognition, a point cloud classification and segmentation method DRPT(Double randomness Point Transformer) based on the improved Transformer model is proposed in this paper. The approach creates new point embeddings in the convolutional projection layer of the Transformer model and uses local dynamic processing of local neighborhoods to continuously add global feature attributes in the data feature vector, thus improving the accuracy of point cloud classification and segmentation in multi-target recognition. Standard benchmark datasets (ModelNet40, ShapeNet partial segmentation and SemanticKITTI scene semantic segmentation datasets) are used in the experiments to validate the performance of the model. The experimental results show that the pIoU value of the DRPT model is 85.9%, which is 3.5% higher than other models on average, and effectively improves the accuracy of point cloud classification and segmentation during multi-target recognition detection, which is an effective support for the development of intelligent network technology.

**Keywords:** pointcloud recognition; road scenes; point cloud segmentation; point set embedding; Transformer

**基金项目:**国家重点研发计划项目(No. 2018YFB1600200);重庆市教育委员会教委科学技术研究计划重点项目(No. KJZD-K202000704)资助。

**作者简介:**马庆禄(1980-),男,教授,研究方向为智能交通与安全、大数据与智慧城市、智慧公路感知与安全等。  
E-mail: mql360@qq.com

**通讯作者:**孙 泉,男,硕士研究生,研究方向为智能交通技术等。E-mail: 1504632042@qq.com

**收稿日期:**2023-02-27; **修订日期:**2023-04-29

## 1 引言

三维道路场景点云数据的精准分类与分割技术,是智能网联汽车进行实时控制决策的前提。然而,与传统 2D 图像相比,3D 点云图像具有更复杂的结构以及更多的检测信息,在三维空间中以不规则和无序的方式分布<sup>[1]</sup>。因此,为了更好地分类与分割,国内外学者围绕优化深度学习网络以实现高效的全局特征提取方面做了大量的研究。康自祥等<sup>[2]</sup>提出一种基于 Transformer 的体素化激光点云目标检测算法,提高了运算效率,目标检测的准确率达到了 88.5%。梁志强等<sup>[3]</sup>利用图卷积提取点云基于位置信息的低维局部特征,并通过注意力机制提取点云基于语义高维局部特征,同时使用多层感知机提权全局特征,通过多个特征融合来提高精度。Qi 等<sup>[4]</sup>提出了 PointNet 方法,以点的方式实现点云特征学习。PointNet 由几个核心模块组成:刚性变换(T-Net)、共享多层感知器(MLPs)和最大池(max-pooling),这些模块保证了网络对点排列和形状旋转的不变性<sup>[5]</sup>,之后通过引入局部特征提取,提出了几个变体来提高 PointNet 的性能。Li<sup>[6]</sup>等提出的 PointCNN,类似于图片的卷积操作,充分考虑了局部信息。Wu<sup>[7]</sup>等提出将 PointConv 扩展到反卷积,以获得更好的分割结果。Wang 等<sup>[8]</sup>提出了 DGCNN 设计了 EdgeConv,可以很好地集成到多个已经存在的点云处理的 pipeline 中。

谷歌公司所提出的 Transformer<sup>[9]</sup>具有显著的全局特征学习能力,来代替网络中的所有卷积运算,以便更好地表达全局特征。尽管 Transformer 架构目前在机器视觉研究的各个领域都得到了应用,但它们在 3D 点云处理中的应用仍然非常有限。在目前基于 Transformer 结构的点云研究中留下了一个空白。首先 Transformer 已被证明对输入数据中的排列具有本质上的不变性,这使其成为集处理任务的理想选择;其次,现有大多数深度学习方法处理的 3D 点利用局部特征来提高性能。这些技术在局部尺度上处理点,以保持它们对于输入排列不变,导致忽略点表示之间的几何关系,并且所述模型不能很好地捕捉全局概念<sup>[10]</sup>。

为解决上述问题,提出了一种基于双随机 Transformer 的点云分类与分割模型,并将其命名为

DRPT,利用 Teansformer 模型获取全局变量,并将双随机代替 Transformer 的行随机以获取更多点特征。在多个 3D 点云数据集上评估 DRPT,用于分类、部分分割和语义场景分割任务。通过实验结果表明,DRPT 可以作为各种点云处理任务的准确有效的主干。

## 2 点云注意机制研究

### 2.1 Transformer 网络函数定义

Transformer 网络依赖于多头注意力与使用单一注意函数不同,多头注意首先使用单独的前馈网络从不同的子空间学习关系,多头注意首先将  $Q$ 、 $K$ 、 $V$  分别投影  $h$  次到  $d_v$ 、 $d_k$ 、 $d_m$  维。然后,注意力平行应用于每个投影。使用前馈网络将输出连接起来并再次投影。因此,多头注意力可以定义如公式(1)所示:

$$\text{Multihead}(Q, K, V) = (\sigma[QW_1^Q \cdot (K)^T]VW_1^V \oplus \dots \oplus \sigma[QW_h^Q \cdot (K)^T]VW_h^V)W^o \quad (1)$$

式中,  $W$  为对应维数的矩阵,  $W_i^Q \in \mathbb{R}^{d_m \times d_k}$ ,  $W_i^K \in \mathbb{R}^{d_m \times d_v}$ ,  $W_i^V \in \mathbb{R}^{d_m \times d_v}$ , 其中  $i \in \{1, 2, 3, \dots, h\}$ ;  $\oplus$  运算表示矩阵串联并且  $W^o \in \mathbb{R}^{hd_m \times d_k}$  是一个可学习的参数矩阵。为了实现与传统注意力类似的计算复杂度,每个头部  $d_k$ 、 $d_v$  的尺寸被减小,使得  $d_k = d_v = d_m/h$ 。对于 Transformer 架构定义了相同层的编码器和解码器堆栈,这些层由多个头部注意力和一个逐点完全连接的层组成,每个都有一个剩余连接,然后是层标准化。将这一层称为“多头注意力”,并将其定义如公式(2)、(3):

$$A^{MH}(X, Y) = \text{LayerNorm}(S + \text{rFF}(S)) \quad (2)$$

$$S = \text{LayerNorm}(X + \text{Multihead}(X, Y, Z)) \quad (3)$$

公式(2)中  $A^{MH}$  表示多头注意力机制的输出,取决于查询矩阵  $X$  的顺序。LayerNorm 是层归一化, rFF 为向前反馈层因此其不具备置换不变性。  $S$  表示多头注意力的中间结果,公式(3)是  $S$  的具体计算过程。公式(3)中,  $X$  是查询矩阵、 $Y$  是键矩阵、 $Z$  是值矩阵, Multihead 是多头注意力函数,它将查询、键和值作为输入,并输出它们的加权和。无论输出顺序,每个点相对应的输出值始终相同,因为  $A^{MH}$  仅由矩阵乘法和求和组成。

### 2.2 Transformer 点云中的应用

Transformer 通过 ViT 来提高学习全局特征的能力,所以当前的点云处理倾向于使用变换器来替换

网络中的所有卷积运算,以获得更好的特征表达。基于 Transformer 的点云分类方法,是为通过 Transformer 获取点云的全局特性。PCT (Point Cloud Transformer) 采用了 PointNet 架构,其中共享的 MLP 层被替换为标准 Transformer 块。通过利用偏移注意机制和邻域信息嵌入,PCT 在点云分类中实现了最先进的性能。Han 等人提出了另一种逐点学习全局特征的方法并将其命名为 PT(PointTransformer)。具体来说,他们设计了一个多级 Transformer 来提取具有不同分辨率的目标点云的全局特征,然后将这些特征串联起来,并将它们输入到一个多尺度 Transformer 中,以获得最终的全局特征。Point Transformer 不是以逐点方式提取全局特征,而是将 Transformer 层应用于点云的局部邻域,并通过向下转换模块分层提取局部特征。最后,通过全局平均池操作可以获得全局特征。然而,由于 Point Transformer 块应用于每层的所有输入点,此方法可能会产生信息冗余。此外,纯 Transformer 体系结构(无 CNN),由于变压器中存在大量线性层,可能会导致较高的计算和内存成本。现有的模型包括 PointTransformer 和 Point Cloud Transformer, PointTransformer 在点上使用单个头部的自我注意,以创建通过 K-NN 和最远点采样获得的局部补片的排列不变表示,Point Cloud Transformer 将偏移注意块应用于 3D 点云。两种模型创建了一种新颖的 Transformer 模块,该模块对输入执行特征方式以及点积关注,以实现精确有效的处理。

### 3 DRPT 网络构建

将  $D$  维的  $N$  个点嵌入网络第  $l$  层卷积层,表示为  $X^l = \{x_1^l, \dots, x_N^l\}, x_i^l \in \mathbb{R}^D$ 。K 邻域内的点  $i \in \{1 \dots N\}$ , 将邻域内点的集合定义为  $\Delta X_i^l = \{x_{j_1}^l - x_i^l, \dots, x_{j_k}^l - x_i^l\}, j_k \in \{1, \dots, N\}, j_k \neq i$ 。DRPT 的整体架构如图 1 所示。

DRPT 的总体架构如图 1 所示。本文主要改进为在点云 Transformer 中加入点嵌入模块和具有卷积注意投影的点间注意模块并将其应用于 3D 点云识别与分割。具体表现为当一个 size 为  $N \times 3$  的输入点云通过该体系结构时,通过基于欧几里德距离寻找其 K-NN 来计算点近邻图。然后,将计算结果将通过点嵌入层传递,该层将输入数据映射到隐式包含点的最近邻域表示中。这其通过 2D 卷积运算

完成的,注意力机制选取的每个点的重叠程度可以通过步幅的长度来控制。将点积注意力操作应用于该嵌入表示,然后插入式注意模块。点积注意力可以看作是学习嵌入点的相关特征,作为其 K 近邻的函数。这种注意机制学习关注点的特征,而不是点本身,也就是说,对于批次中的一组点,它学习加权单个特征变换。

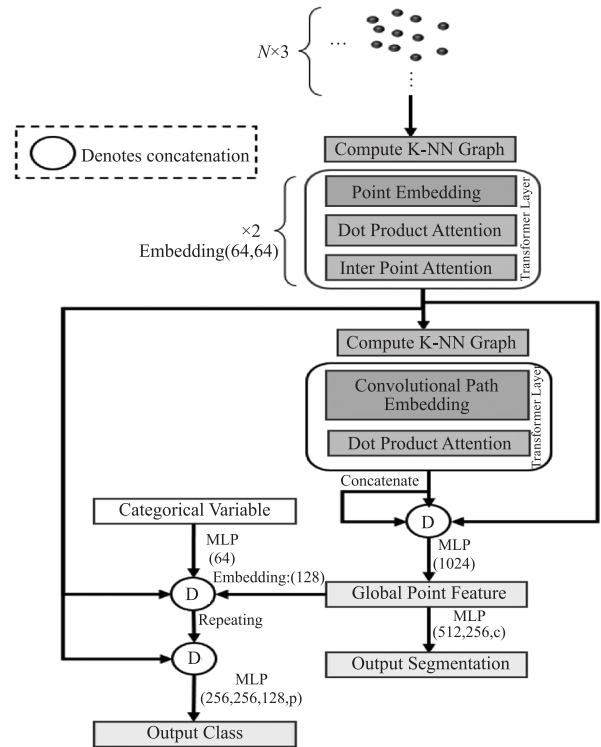


图 1 DRPT 模型结构

Fig. 1 DRPT model structure

另一方面,插入点注意可以解释为在一个批次内学习不同点之间的关系(按点嵌入而不是按点的单个特征操作的行矩阵注意)。这形成了 DRPT 的一层。我们在 InterPoint Attention 操作之后更新近邻图,然后将其传递到下一个 DRPT 层使用特征空间嵌入维数较高的第三层也是最后一层 DRPT 层来学习点云的表示,无需插入注意。深层的插入注意并不能提供很多继续你,因为 Transformer 最终可以学习超越其点集嵌入维度的位置的关系。这也意味着,与共享加权 MLP 相比,Transformer 能够在更深的层中可视化超出其有限感受野的输入,使其成为特征空间嵌入的更好选择。将此注意模块添加到更深的层只会略微提高性能,但会显著增加计算量。然后将每个变压器层的输出串联起来,并通过最终的共享加权 MLP。

然后将其合并为全局最大值,以获得形成点云表示的全局特征向量。然后通过一系列 MLP 进一步处理该表示,以对点云进行分类,或为点云中的每个点获取语义点标签。

### 3.1 点积及双随机点间注意力机制嵌入

具有点嵌入的 PCT 是提取全局特征的有效网络。然而,它忽略了在点云学习中同样重要的局部邻域信息。点积注意力操作被应用于嵌入后,点积注意力可以被看作是学习作为其  $M$  个最近邻函数点嵌入的相关特征。这种注意机制学习关注点的特征不是点本身,即对于一批中的一组点,它学习的式加权各个特征变换。之后是点间注意模块。点间注意力可以被解释为在邻域中学习不同点本身之间的关系,点间注意力采用逐行矩阵注意力针对每个点嵌入操作,本文将行随机矩阵注意力改为双随机矩阵注意力,其能够增加点之间的交互,增强特征提取。将 Softmax 被 Sinkhorn 的输出所代替,然后将它传递给下一个 DRPT 层。使深层的点间注意力能够提供信息,Transformer 最终可以学习跨越其嵌入维度的局部性的关系。这也意味着 Transformer 能够在更深的层中可视化超出其有限感受野的输入,与共享加权的 MLPs 相比,这使它们成为特征空间嵌入的更好选择。将这种关注块添加到更深的层仅略微提高了性能,但是显著增加了计算如图 2 所示。

每个 Transformer 层的输出然后被连接并通过最终的共享加权 MLP。然后对其进行全局最大汇集,以获得形成点云表示的全局特征向量。然后,通

过一系列 MLP 进一步处理该表示,以对点云进行分类,或者获得点云中每个点的语义点标签。如在传统的 Transformer 模型结构中一样,在关注层之后使用剩余连接层标准化以及 MLP 层。

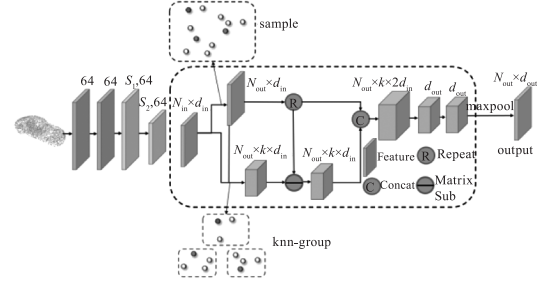


图 2 邻域嵌入架构

Fig. 2 Neighborhood embedding architecture

此模块输入的一般结构由  $(B, f, N, \Delta X_i^l) \in \mathbb{R}^{(B \times f \times N \times \Delta X_i^l)}$  表示,其中  $B$  和  $f$  分别是批量大小和输入特征。这一层本质上是一个映射函数  $F_\theta$ ,它将输入映射到  $E_\theta$ ,表达式为  $y = F_\theta(I(B, f, N, \Delta X_i^l))$ ,  $y \in \mathbb{R}^{(B \times f \times E_\theta)}$ 。映射函数  $F_\theta$  使用具有固定大小填充和步幅的卷积运算。这能够在端到端设置中训练该模块以及网络的其余部分,因为它是完全不同的可以插入到架构中的任何位置。

### 3.2 Point Transformer 卷积层构建

使用点积注意力和自注意力的组合,通过网络传播输入中最重要的特征。DRPT 通过使用卷积层对注意力矩阵进行采样。利用卷积的能力来学习 3D 点的相关特征集。因此用形成卷积投影的深度方向卷积层替换关注块中的原始全连接层,以获得注意矩阵,如图 3 所示。

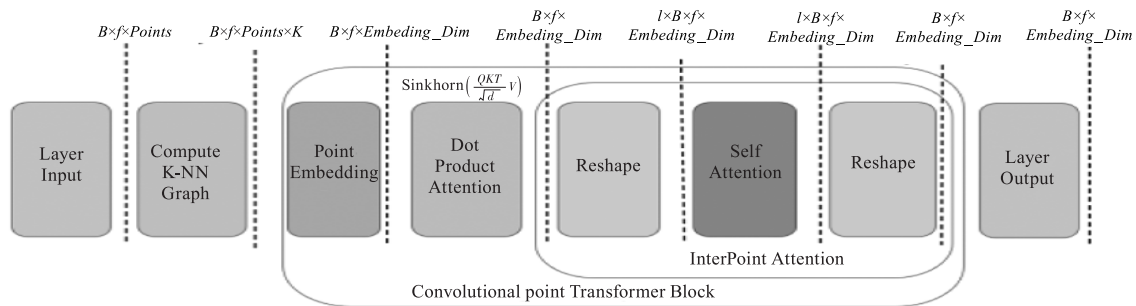


图 3 点积注意力和点间注意力的 Transformer 块

Fig. 3 Transformer blocks for dot product attention and interpoint attention

Transformer 卷积块的其余部分的功能类似于 ViT 的每个注意力块之后添加归一化和前馈层,并且使用残差映射来增强特征学习。形式上,DRPT 层的投影操作由  $z_i^{q/k/v} = \text{Convolution}(z_i, p, s)$  表示,

其中,  $z_i^{q/k/v}$  是第  $i$  层的  $Q/K/V$  注意力矩阵的输入,而  $z_i$  是第  $i$  层的卷积块的输入。在点间注意模块中的 Softmax 替换为能够实现更多点之间交互的 Sinkhorn 算法, Sinkhorn 为将 Softmax 的行随机矩阵变化

内双随机矩阵。直观地说,这种规范化依赖于投票原则,即所有点都以不同的强度相互匹配,因此比 SoftMax 规范化考虑更多的交互如式:

$$\begin{aligned} \text{Sinkhorn}(C) &= \text{argmin} \mathbf{K} \mathbf{1}_n \\ &= \mathbf{K}^\top \mathbf{1}_n = \mathbf{1}_n \sum_{i,j} K_{i,j} \log\left(\frac{K_{i,j}}{K_{i,j}^0}\right) \end{aligned} \quad (4)$$

卷积运算是一种深度可分离的卷积运算,其元组为内核大小和步长。现在,将通过整个 DRPT 模块的数据流形式化为批量大小  $B$  的,可得点积注意力块输出如式:

$$\text{out}_i^a = \beta\left(\text{Sinkhorn}\left(\frac{QK^\top}{\sqrt{d}}V\right)\right) + z_i^{q/k/v} \quad (5)$$

其中,  $\text{out}_i^a$  是点积注意力块的输出;  $d$  用于缩放注意力权重以避免梯度不稳定性;  $\beta$  是卷积层归一化操作。加法表示剩余连接。对输出采用以如下式在 transformer 中被进一步处理,如公式:

$$\begin{cases} \text{out}_i^b = \beta(F(\text{out}_i^a)) + \text{out}_i^a \\ \text{out}_i^c = \beta(D(\text{out}_i^b)_{i=1}^B) + \text{out}_i^b \\ \text{out}_i^d = \beta(F(\text{out}_i^c)) + \text{out}_i^c \end{cases} \quad (6)$$

这里  $F$  表示前馈层,  $\text{out}_i^b$  表示对于输出的进一步处理,  $\text{out}_i^c$  表示点间注意力输出,然后由归一化层和前馈层以类似的方式进行处理输出  $\text{out}_i^d$ , 在此步骤之后,重新计算图形,并对第  $l+1$  层重新开始该过程。按照官方的提出的采用 mIoU (mean Intersection over Union) 作为评价方法效率进行实验。mIoU 是一种常用的用于评估语义分割算法性能的指标,是对 IoU (Intersection over Union) 的平均值进行计算。其中, IoU 是指预测结果和真实标签之间的交集与并集之比,即:  $\text{IoU} = \text{TP} / (\text{TP} + \text{FP} + \text{FN})$ , 其中, TP 表示真实标签和预测结果都为正例的像素数, FP 表示预测结果为正例,但真实标签为负例的像素数, FN 表示预测结果为负例,但真实标签为正例的像素数。而 mIoU 则是对所有类别的 IoU 求平均,即:  $\text{mIoU} = (\text{IoU}_{_1} + \text{IoU}_{_2} + \dots + \text{IoU}_{_n}) / n$  其中,  $n$  表示类别数。mIoU 越高,则代表模型的性能越好。

#### 4 实验和结果

点云分类、零件分割和语义场景分割三个不同的数据集上对模型进行了评估。对于分类,使用基准 ModelNet40 数据集,对于对象部分分割,使用

ShapeNet 部分数据集。SemanticKITTI 数据集进行语义场景分割。使用基于余弦退火的学习率调度器,不使用批量标准化衰减。在初始实验之后,定制的学习率计划被用于分段任务。ModelNet40 数据集包含 3D CAD 模型的网格。共有 311 个模型可用于 40 个类别,分成 468 个训练测试集。使用所有的实验和数据集提供的官方分割,以保持公平的比较。在数据预处理方面,遵循与相同的步骤。从网格表面均匀采样 1024 个点,并重新调整点云以适应单位球,在训练过程中使用数据扩充。在增强过程中,用随机抖动和缩放来扰动这些点最终的检测结果如表 1。其中将 Class Accuracy, Instance Accuracy 作为两项评价指标进行模型对比,将 DRPT 与其他先进算法进行对比。

表 1 ModelNet40 精度对比

Tab. 1 Progress comparison of ModelNet40

Method	Class Accuracy/%	Instance Accuracy/%
3D ShapeNets	77.3	84.7
VoxNet	83.0	85.9
PointNet	86.0	89.2
PointNet++	/	90.7
DGCNN	90.2	92.2
PointTransformer	90.6	93.7
DRPT	<b>90.6</b>	<b>93.9</b>

从表 1 中可以看出, DRPT 优于现有的分类方法,甚至略好于 PointTransformer。当模型中采用图形计算方法时,与处理静态图形以及动态图形的其他方法进行比较,以进行点云分类。在这里,即使没有在一层重新计算图形, DRPT 也优于动态边缘条件滤波器 (ECC), 并且与 DGCNN 性能相当。每层之前的动态图形计算有助于 DRPT 超越现有的基于图形和非图形的方法的准确性其中包括略微超过现有的基于 Transformer 的方法。

#### 4.1 ShapeNet Part 分割结果

ShapeNetPart 数据集包含来自 16 个对象类别, 881 个 3D 形状。总共有 50 个对象部分可用于分割。仅选取部分在数据集上的结果总结在表 2 中。

表 2 ShapeNet Part 分割对比

Tab. 2 ShapeNet Part segmentation comparison

Method	air-plane/%	Motor/%	Car/%	mIoU/%
Kd-Net	80.1	57.4	70.3	82.3
PointNet++	82.4	71.6	77.3	85.1
PointNet	83.4	65.2	74.9	83.7
DGCNN	84.0	66.3	77.8	85.1
PCT	85.0	64.6	81.2	86.4
PT	85.3	65.3	82.1	86.6
DRPT	85.2	75.6	85.8	<b>87.0</b>

表 2 显示了按类别划分的结果,所使用的评估度量是部分精度和平均交集/并集,并针对整体和每个对象类别给出。结果表明,本文的 DRPT 比 PointNet++ 和 PCT 分别提高 1.9 % 和 0.6 %。DRPT 以 87.0 % 的 mIoU 取得了最好的结果。在测试过程中,使用了多尺度测试策略,对于所有三个模型,批量大小,训练时期和学习率被设置为与正常估计任务的训练相同,最终通过对模型在 ShapeNet 数据集的可视化如图 4 所示。

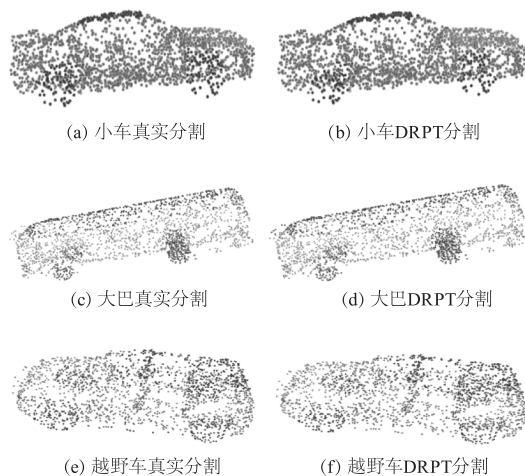


图 4 ShapeNet 上的 DRPT 分割结果

Fig. 4 DRPT segmentation results on ShapeNet

实验在完整形状和部分扫描上训练模型。结果显示,DRPT 能够有最高 87.0 的 mIoU。在图 4 中,给出了完整和部分数据的定性结果。可以看到,尽管部分数据区分复杂,但模型展示出的预测是合理的。

#### 4.2 SemanticKITTI 点云分割

SemanticKITTI 基于 KITTI Vision Benchmark,为所有序列提供语义注释,涵盖了所用汽车 LiDAR 完

整的 360 度视野。数据集包含 28 个类,包括区分非移动和移动对象。类别涵盖所有交通参与者,也包括地面功能类别,如停车场、人行道。用不同的类别注释了移动和非移动的交通参与者,包括汽车、卡车、摩托车、行人和骑自行车的人如图 5 所示,这可以推理场景中的动态对象。

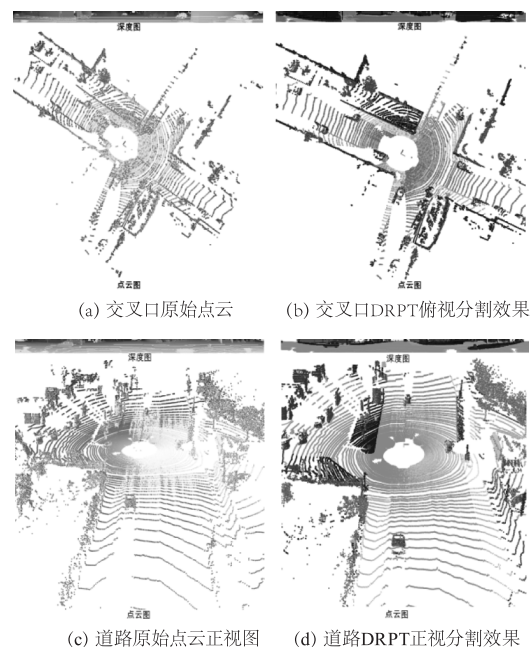


图 5 SemanticKITTI 分割图

Fig. 5 SemanticKITTI segmentation diagram

由图 5 可以看出 DRPT 模型在道路场景中对于道路、车辆、植物等检测效果明显,能够有效将这些物体在点云图中进行有效分割。另外,从图 5 中 DRPT 算法处理后的深度图与原始深度图的对比能够看出,DRPT 算法能进行高精度的分割,并且从点云图的效果能够看出在点云分割中对于道路以及道路周围环境和参与者的分割具有良好的效果,在分割完成后能够使得检测目标更为简单。按照官方提出的采用 pIoU (Point Intersection over Union) 作为评价指标进行实验。pIoU 是点云语义分割任务中常用的评价指标,含义是所有真实标签和预测标签的交集与并集的比值,包括完全匹配和部分匹配的样本,其值越高代表模型处理点云能力越强。实验通过与基于雷达数据的基准网络 PointNet, Voxel RCNN, VoxelNet, PointNet++, PointTransformer 等经典网络进行对比,使用数据集对比结果如表 3 所示。

表 3 Semantickitti 精度对比  
Tab.3 Semantickitti precision comparison

Method	pIoU/%	Pole/%	Bicycyle/%	Car/%	Road/%	Person/%	Truck/%	Motor/%	Building/%	Traffic sign/%
PointNet	80.4	83.4	78.7	74.9	89.6	73.0	91.5	65.2	81.2	80.6
Voxel RCNN	81.9	80.1	79.0	77.3	90.8	71.8	91.0	71.6	81.3	82.6
VoxelNet	82.4	81.6	74.6	70.3	88.6	73.5	90.2	57.4	78.1	80.3
PointNet +	82.7	83.1	76.7	73.8	89.1	67.6	90.6	64.4	79.7	80.6
PointTransformer	84.6	85.5	81.7	75.2	90.2	74.9	93.0	66.7	81.6	82.1
DRPT	<b>85.9</b>	84.0	86.4	78.2	90.4	69.3	91.4	66.0	81.8	82.2

从表3中可以看出,DRPT的pIoU值为85.9%,明显优于现有的分类方法,甚至略好于PointTransformer。DRPT加强了网络对于全局特征的提取,提高了模型对于整体检测精度,但对于小目标类的检测能力并未有明显提升。当模型中采用图形计算方法时,与处理静态图形以及动态图形的其他方法进行比较,以进行点云分类,并在距离检测方面展示优越性。

## 5 结论

本文提出了DRPT,一种使用Transformers处理3D点云的方法。DRPT通过在每个中间网络层进行动态图形计算来优化点云处理。它使用点嵌入并将其转换为转换器友好的数据表示,然后使用双随机点间注意模块来促进点之间的串扰。DRPT优于现有的卷积方法,并在各种基准任务上与现有的基于Transformer的方法相比具有更好的性能。未来研究方向是学习沿着三维点云的流形均匀采样点以提高模型准确性。虽然DRPT可以作为点云处理任务的有效骨干,并扩展到各种应用,但是由于Transformer框架的计算量较大,处理速度较慢,需要进一步提高处理速度。

## 参考文献:

- [1] J W Ma, F Leite. Performance boosting of conventional deep learning-based semantic segmentation leveraging unsupervised clustering[J]. Automation in Construction, 2022, 136: 104167 - 104178.
- [2] Kang Zixiang, Wang Shengzhe, Cui Yuyong, et al. Transformer-based algorithm for voxelized laser point cloud target detection[J]. Laser & Infrared, 2023, 53(2): 202 - 207. (in Chinese)  
康自祥,王升哲,崔雨勇,等. 基于Transformer的体素化激光点云目标检测算法[J]. 激光与红外, 2023, 53(2): 202 - 207.
- [3] Liang Zhiqiang, Chen Chunmei, Chen Yanjie, et al. Point cloud classification and segmentation network based on double local features fusion of high-dimensional and low-dimensional[J]. Laser & Infrared, 2022, 52(10): 1557 - 1564. (in Chinese)  
梁志强,陈春梅,陈妍洁,等. 高低维双局部特征融合的点云分类分割网络[J]. 激光与红外, 2022, 52(10): 1557 - 1564.
- [4] T M Khan, A Robles-Kelly, S Naqvi. T-Net: a resource-constrained tiny convolutional neural network for medical image segmentation[C]//Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 2022: 644 - 653.
- [5] J Ma, X Jiang, A Fan, et al. Image matching from hand-crafted to deep features: a survey[J]. International Journal of Computer Vision, 2021, 129(1): 23 - 79.
- [6] Y Wang, Y Sun, Z Liu, et al. Dynamic graph CNN for learning on point clouds[J]. Association for Computing Machinery (ACM), 2019, 38(5): 1 - 12.
- [7] Y Guo, H Wang, Q Hu, et al. Deep learning for 3D point clouds: a survey[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 43(12): 4338 - 4364.
- [8] J Stanisz, K Lis, M Gorgon. Implementation of the point pillars network for 3D object detection in reprogrammable heterogeneous devices using FINN[J]. Journal of Signal Processing Systems, 2021: 1 - 16.
- [9] S K Roy, S R Dubey, S Chatterjee, et al. FuSENet: fused squeeze-and-excitation network for spectral-spatial hyperspectral image classification[J]. IET Image Processing, 2020, 14(8): 1653 - 1661.
- [10] R Huang, Y Xu, U Stilla. GraNet: global relation-aware attentional network for semantic segmentation of ALS point clouds[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2021, 177: 1 - 20.
- [11] Y Feng, J Chen, T Zhang, et al. Semi-supervised meta-learning networks with squeeze-and-excitation attention for few-shot fault diagnosis[J]. ISA Transactions, 2022, 120: 383 - 401.