

# 一种基于空间特征注意力机制的点云分析方法

曲彦霖, 王悦, 张倩, 韩绍坤\*

北京理工大学光电学院精密光电测试仪器及技术北京市重点实验室, 北京 100081

**摘要** 针对现有的基于点的网络平等地对待所有的点从而无法有效关注重要特征的问题,在激光雷达点云处理领域引入注意力机制,即 CSA 模块,其中 CA 表示通道注意力,SA 表示空间注意力。两个模块以数据驱动的方式自动学习不同特征通道信息和不同空间位置信息的重要性,从而提升网络在点云分类和分割任务上的表现。在基于点的网络中引入了上述两个模块,提出了 CSA-PointNet++ 结构。实验结果表明:所提方法在 ModelNet40 数据集上的分类准确率达 93.20%,在 ShapeNetPart 数据集上的部件分割实验的平均交并比(mIoU)为 82.62%,优于其他对比方法,验证了所提网络的有效性;同时,在真实世界自建数据集上,所提方法的分类准确率达 92.14%,证明了网络在真实世界的的数据上具有良好的泛化能力。

**关键词** 深度学习; 激光点云; 点云处理; 特征提取; 注意力机制

中图分类号 TP391.4

文献标志码 A

DOI: 10.3788/LOP230840

## Point Cloud Analysis Method Based on Spatial Feature Attention Mechanism

Qu Yanlin, Wang Yue, Zhang Qian, Han Shaokun\*

Beijing Key Lab for Precision Optoelectronic Measurement Instrument and Technology, School of Optics and Photonics, Beijing Institute of Technology, Beijing 100081, China

**Abstract** To address the limitations of existing point-based networks, which treat all points with equal emphasis, thereby overlooking crucial features, this paper introduces an attention mechanism to lidar point cloud processing. This mechanism, referred to as the CSA module, integrates the channel attention and spatial attention elements. In a data-driven approach, the two proposed modules autonomously learn the importance of different feature channel information and spatial location information, thereby enhancing the performance of the network on point cloud classification and segmentation tasks. This paper introduces the two modules stated above in a point-based network and proposes a CSA-PointNet++ architecture. The results reveal that the proposed method achieves an accuracy of 93.20% for classification experiments on the ModelNet40 dataset and a mean intersection over union (mIoU) of 82.62% for part segmentation experiments on the ShapeNetPart dataset. This performance is better than that of other comparative methods, indicating the effectiveness of the proposed network. Moreover, classification experiments of the proposed method on a real-world self-constructed dataset yield an accuracy of 92.14%, demonstrating the excellent generalization capability of the proposed network on real-world data.

**Key words** deep learning; laser point cloud; point cloud processing; feature extraction; attention mechanism

## 1 引言

近年来,随着计算机视觉、人工智能等前沿信息科学技术的飞速发展,3D 传感器日益普及,如深度相机、双目相机、激光雷达等,获取点云数据的方式越来越多,3D 点云数据的质量、采集效率和有效性在不断提高。点云数据广泛应用于无人驾驶汽车、自动导航、智

慧城市、测绘、遥感等领域。与在现实世界中失去深度信息和两个或多个物体之间的相对位置的 2D 图像相比,3D 点云为机器提供了更好地了解周围环境的机会,这样机器可以更加准确地捕捉到物体的形状、大小、运动轨迹等细节。因此如何有效地使用 3D 点云数据正逐渐成为一个研究热点<sup>[1-4]</sup>。

目前,在二维图像领域,已经有了大量比较成熟的

收稿日期: 2023-03-13; 修回日期: 2023-04-17; 录用日期: 2023-05-15; 网络首发日期: 2023-05-27

通信作者: \*skhan@bit.edu.cn

深度学习网络架构和技术手段,受其影响,许多研究者也希望将这些技术应用于对三维点云数据的研究。激光雷达点云是近些年逐渐兴起的一种数据表达方式,相比二维图像数据,具有不均匀性和无序性的特点。由于卷积神经网络(CNN)要求输入数据的格式是高度规则的,以便执行权重共享和其他相关操作,因此,难以将传统的2D卷积神经网络直接应用于点云特征学习。

一些方法提出先将无序的点云转换为规则的数据格式,再对其进行卷积操作。例如,基于多视图的方法首先将点云数据从不同的角度投影到图像上,再利用传统的视觉CNN对不同视图的特征信息进行处理和融合<sup>[5-9]</sup>。基于多视图的方法的缺点是会损失点云的三维信息,因此会影响后续特征提取的效果,并且计算成本大。基于体素的方法首先将点云转换为三维体素,再对体素进行卷积操作来学习点云的特征<sup>[10-13]</sup>。然而基于体素的方法想取得更好的效果就必须对网格的分辨率进行提高,这会造成网络规模膨胀。

研究者提出一种名为PointNet<sup>[14]</sup>的直接处理原始点云数据的深度网络架构,它是关于点云的深度学习的开山之作。PointNet<sup>[14]</sup>的结构简单而且有效,它的主要创新点是使用一个对称函数和一个多层感知器(MLP)直接处理原始的无序点云。PointNet<sup>[14]</sup>通过MLP将每一个点嵌入到潜在空间,它存在不能提取局部特征的局限性,这就限制了它对精细特征的学习和对复杂场景的通用性。为了解决这个缺点,作者进一步提出了PointNet++<sup>[15]</sup>,它是一个基于PointNet<sup>[14]</sup>的分层神经网络,可以看作是PointNet的扩展。该网络通过对分组的邻居点进行卷积运算来提取局部特征,解决了PointNet的不足,并且在点云的分类和分割上都表现出良好的性能。然而无论是PointNet<sup>[14]</sup>还是PointNet++<sup>[15]</sup>,都具有局限性,对所有的点和所有的特征通道一视同仁,这是不合理的。一般来说,现有的方法忽略了每个点的特征的重要性对分类和分割的任务是不同的这个特点,这会影响识别物体的准确度。在处理点云数据时,应仔细考虑这个问题。在2D图像领域已经证明,总会有一些关键点和关键特征通道是识别物体的关键,需要引起更多的注意。2018年,Hu等<sup>[16]</sup>提出了SE-Net,将它引入到CNN中,该网络在二维图像的分类上取得良好的效果。后来,研究者又提出了一些注意力机制,如ECA-Net<sup>[17]</sup>、SK-Net<sup>[18]</sup>、ResNeSt<sup>[19]</sup>、CBAM<sup>[20]</sup>等,分别从通道维度和空间维度这两个方面增强了注意力机制的性能。注意力机制能够学习更有用的特征,从而大幅度地提升模型的训练速度和任务效果。因此在点云数据处理中引入注意力机制关注点云的关键点和关键特征是有意义的。

由于PointNet<sup>[14]</sup>和PointNet++<sup>[15]</sup>把所有的点看得同样重要,缺失了对识别物体的关键信息的关注。注意力机制能有效关注目标场景中所需要的特征,忽

略其他不重要的特征。为了提高对重点信息的关注,本文采用PointNet++模型<sup>[15]</sup>作为基准模型,提出多种注意力机制CSA模块,将其引入到PointNet++中,名为CSA-PointNet++。具体来说,CSA模块由两部分组成,其中CA表示通道注意力,SA表示空间注意力。为了发掘点云不同特征通道维度信息的重要性,提出了CA模块<sup>[20]</sup>。同时,为了发掘点云不同空间位置信息的重要性,提出了SA模块。两个模块发掘通道和空间维度的有用特征,过滤掉冗余的特征信息,从而提升了模型在点云分类和分割任务上的表现。本文的主要工作和创新点如下。

1)基于PointNet++<sup>[15]</sup>,提出多种注意力机制模型,即CSA模块,把通道注意力与空间注意力模型串行引入到PointNet++<sup>[15]</sup>中,用于三维点云的特征提取。网络采用编码器-解码器的结构来探索相邻点与上下文之间的关系,并且自动关注重要信息,在分类和分割任务上得到了很好的性能。

2)使用Kinect V2深度相机对真实场景的数据集进行采集,对4类军事车辆模型进行采集。通过3D刚体变换得到深度图像,再对采集到的深度图像进行处理,得到三维点云数据,构成自建数据集。

3)使用公开数据集ModelNet40<sup>[21]</sup>进行分类实验,使用ShapeNetPart<sup>[22]</sup>进行部件分割实验,结果表明,相比原始的点云网络,所提网络的性能得到了提升。同时,在真实世界的自建数据集上的分类实验验证了所提网络具有良好的泛化能力和广泛的应用场景。

## 2 点云特征提取方法

通常,利用PointNet++<sup>[15]</sup>等深度学习方法对点云数据进行特征提取,这些方法把所有的点看得同样重要,从而忽略了对识别物体的关键信息的关注。因此,本文提出了一种改进的网络结构,名为CSA-PointNet++。所提方法使用PointNet++<sup>[15]</sup>作为主干网络,引入一个基于注意力机制的局部关系学习模块(CSA模块)来收集局部特征,选择对任务贡献较大的特征点,即突出重要通道和空间维度的信息,抑制无效的通道信息,使所选择的点集能够代表整个采样空间。

### 2.1 CSA 模块

输入的点云数据的特征维度为 $B \times C \times S \times N$ ,其中 $B$ 、 $C$ 、 $S$ 、 $N$ 分别是batch size、特征通道数、采样点数量、点云数目。为了克服PointNet++<sup>[15]</sup>平等地对待每一个点的问题,受应用于图像领域的注意力机制<sup>[20]</sup>的启发,提出了一个能应用于点云数据的注意力模块(CSA模块)来细化提取特征。CSA模块由两个子模块组成:沿特征通道数 $C$ 的通道注意力(CA)模块和沿点云数目 $N$ ,即点云空间位置的空间注意力(SA)模块。特征图将依次通过这两个注意力模块。CSA模块结构如图1所示。

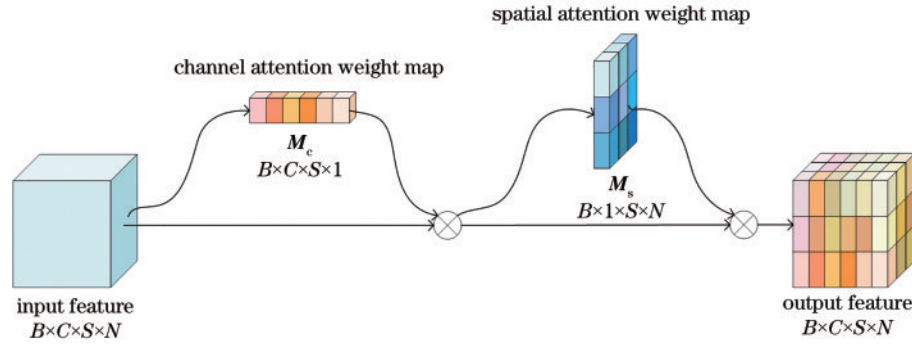


图 1 CSA 模块结构

Fig. 1 CSA module structure

2.1.1 CA 模块

通道注意力机制<sup>[20]</sup>通过自动学习的方式来获得特征图中每个通道的重要性,从而为每个特征的权重赋值,使神经网络聚焦于重要的特征通道,增强特征图中对当前任务有用的特征通道,抑制无用的特征通道。CA 模块的结构如图 2 所示,特征图中不同的权重用不同的颜色表示,在输入通道注意力模块之前,特征图各通道的颜色是相同的,即权重值相同;通过通道注意力模块后,各特征通道的颜色变得不同,即权重值不同,神经网络会关注权重值较大的特定通道。

输入的特征图经过通道注意力模块,分别进行最大池化和平均池化操作,生成  $F_{avg}^C$  和  $F_{max}^C$ 。不同于二维图像数据的通道注意力,由于在点云领域无法直接使用 CNN,本文使用共享的包含多个隐藏层的 MLP 网

络对生成的向量进行处理,并对得到的结果进行元素求和。MLP 网络通过利用两个全连接层对通道之间的特征关联关系进行建模,并输出每个特征的权重,权重的数量与输入特征的个数相同。在经过这两个全连接层时,特征维度首先降到输入的  $1/r$ ,接着通过 ReLU 函数进行激活,最后特征维度升回与输入维度相同。再经过 Sigmoid 函数进行映射处理,生成最终的通道注意力权重图,对该通道注意力权重图和输入特征图进行逐元素乘法操作,生成的输出特征就是空间注意力模块需要的输入特征。其中  $r$  是一个缩放参数,它的作用是减少通道的个数从而降低计算量。通道注意力机制可以表示为

$$M_c(F) = \sigma \left[ M(F_{max}^C) + M(F_{avg}^C) \right], \quad (1)$$

式中:  $M$  是一个具有多隐藏层的共享 MLP;  $\sigma$  表示 Sigmoid 函数。

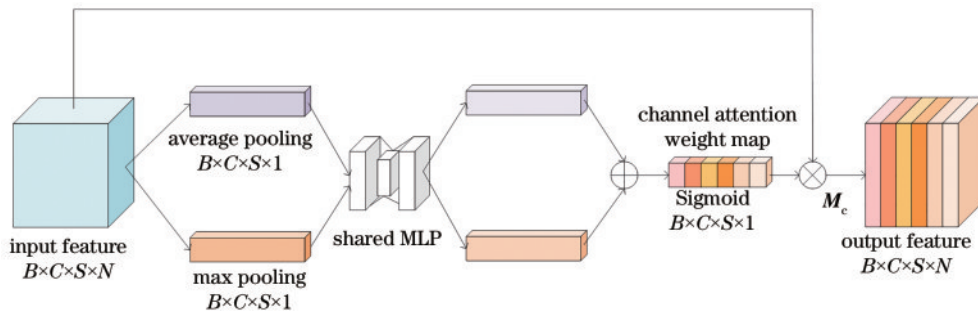


图 2 CA 模块结构

Fig. 2 CA module structure

总的来说,所提 CA 模块以数据驱动的方式根据输入特征的特点学习特征通道的重要程度,其中权重图中的权重越大证明该通道的特征越重要,在不同的任务中根据不同的任务要求增强或抑制不同的通道。

2.1.2 SA 模块

利用空间注意力机制<sup>[20]</sup>来区分点云数据中不同区域对任务的贡献程度。SA 模块的结构如图 3 所示。在经过 SA 模块之前,输入的特征图在点云空间位置的维度上颜色是相同的,即权重值相同;通过空间注意力模块后,点云空间位置维度的颜色变得不同,即权重

值变得不同。神经网络在对特征图进行处理时会关注权重值较大的特定位置。

输入的特征图经过空间注意力模块,在点云空间位置的维度上分别进行平均池化和最大池化操作,生成  $F_{avg}^S$  和  $F_{max}^S$ ,然后对平均池化和最大池化的结果按照通道进行拼接,得到特征图维度  $B \times 2 \times S \times N$ 。与 CA 模块相同,不同于二维图像数据的空间注意力,由于 CNN 不能直接在点云领域应用,本文没有使用卷积操作来处理拼接后的特征,而是使用有多个隐藏层的共享 MLP。MLP 网络通过利用两个全连接层对点云空

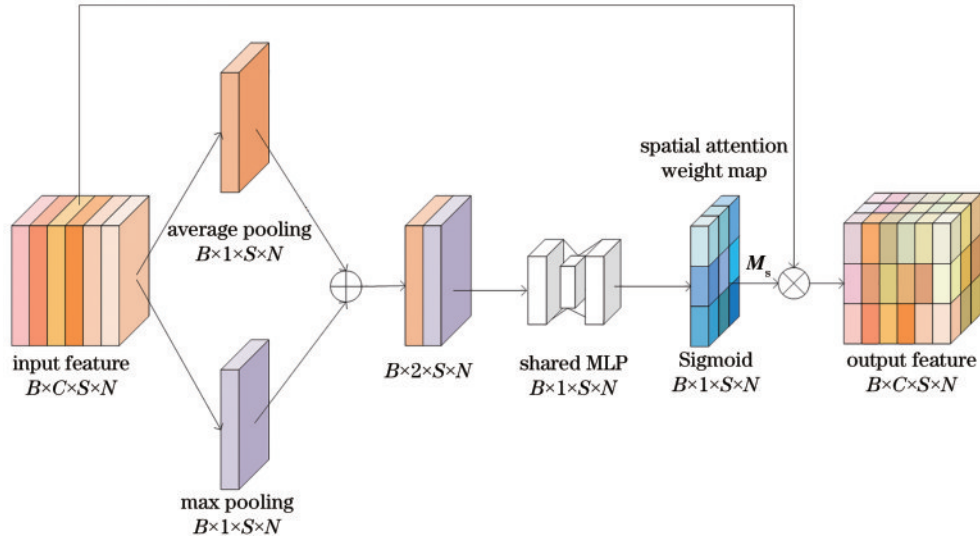


图3 SA模块结构

Fig. 3 SA module structure

间位置的重要性进行建模,并输出每个空间位置的权重。在经过两个全连接层时,点云的空间维度首先降到输入的 $1/r$ ,接着通过ReLU函数进行激活,最后将空间维度升回与输入维度相同。接着对经过Sigmoid函数的信息进行归一化,得到包含各通道权值信息的空间注意力图。最后,对空间注意力权重图与输入特征图进行逐元素乘法操作,得到输出特征。同样地, $r$ 为缩放参数,它的作用是通过减少点云空间维度的参数量从而减少计算量。简而言之,空间注意力机制可以表示为

$$M_s(\mathbf{F}) = \sigma \left[ M(\mathbf{F}_{\text{avg}}^s) \otimes M(\mathbf{F}_{\text{max}}^s) \right]. \quad (2)$$

总的来说,所提SA模块以数据驱动的方式根据输入特征的特点学习点云空间位置的重要程度,其中权重图中的权重越大证明该空间位置的特征越重要,然后针对不同的任务增强或抑制不同的空间位置。最后,将CA模块和SA模块的结果串联起来,对权值较高的通道进行激励。两个注意力模块都学习关注“什么”,关注“哪里”。通过学习哪些是加强的,哪些是抑制的,有效地帮助信息在网络中流动。CSA模块简单有效,能很好地用于前馈式CNN。

## 2.2 CSA-PointNet++

为了解决PointNet++<sup>[15]</sup>对所有点一视同仁,缺少对重点信息的关注的问题,本文在其中引入了CSA模块。

### 2.2.1 整体网络结构

图4显示了所提方法的概述,所提CSA-PointNet++主要通过PointNet++<sup>[15]</sup>中加入注意力机制,实现对点云数据的分类和分割性能的提升。具体而言,将网络设计为分层的编码器-解码器架构,包括编码器网络、解码器网络和跳过连接。编码器提取更多的点云多尺度特征,是自下而上的过程;解码器

恢复更强的语义特征表示,生成高精度的分类点云,是自上而下的过程;跳过连接是一种能提高深度神经网络性能和收敛性的技术,有条件地连接不同层中的局部点特性,通过在神经网络层中传播线性分量来缓解非线性引起的优化困难。

### 2.2.2 编码器网络

编码器网络<sup>[15]</sup>由4个集合抽象模块组成,通过多个集合抽象结构实现多层次降采样。对点数为 $N$ 的输入点云提取尺度为 $\{1/4, 1/16, 1/64, 1/256\}$ 的多尺度特征,得到不同规模的逐点的特征,最后一个集合抽象结构的输出可以看作是全局特征。图5显示了第一个改进后的集合抽象模块结构<sup>[23]</sup>,该结构将CSA模块插入到原始PointNet++的集合抽象模块的PointNet层进行特征提取,输入一个规模为 $N \times 6$ 的点集,其中 $N$ 表示点集中点的数量,6是点的维度。

点云采样层采用最远点采样法(FPS算法)对点集进行降采样,将输入点的规模从 $N$ 降到 $N/4$ ,即在 $N$ 个点中选取 $N/4$ 个中心点,输出规模为 $N/4 \times 6$ 。FPS算法是随机选取一个点作为起始点,选取点集中距离起始点最远的点作为下一个采样点,一直到采样到 $N/4$ 个目标采样点为止的算法。

然后利用分组层(grouping layer)构造相应的局部区域,分组层采用球查询的方法来获取采样点的邻域信息。球查询的过程:以这 $N/4$ 个中心点为中心,首先确定球的半径,然后在这个半径的球内找到其固定规模( $K$ )的邻点,共同组成一个局部区域,如果在这个区域内点数大于 $K$ ,则只采样 $K$ 个点;小于 $K$ ,则对点进行重复采样。本文中 $K$ 设置为32,所以输出规模为 $N/4 \times 32 \times 6$ 。

接下来将采样层和分组层的结果输入到CSA模块进行特征提取,将这些局部区域编码为特征向量。对每个输入的点采用多层感知器提取特征,并通过

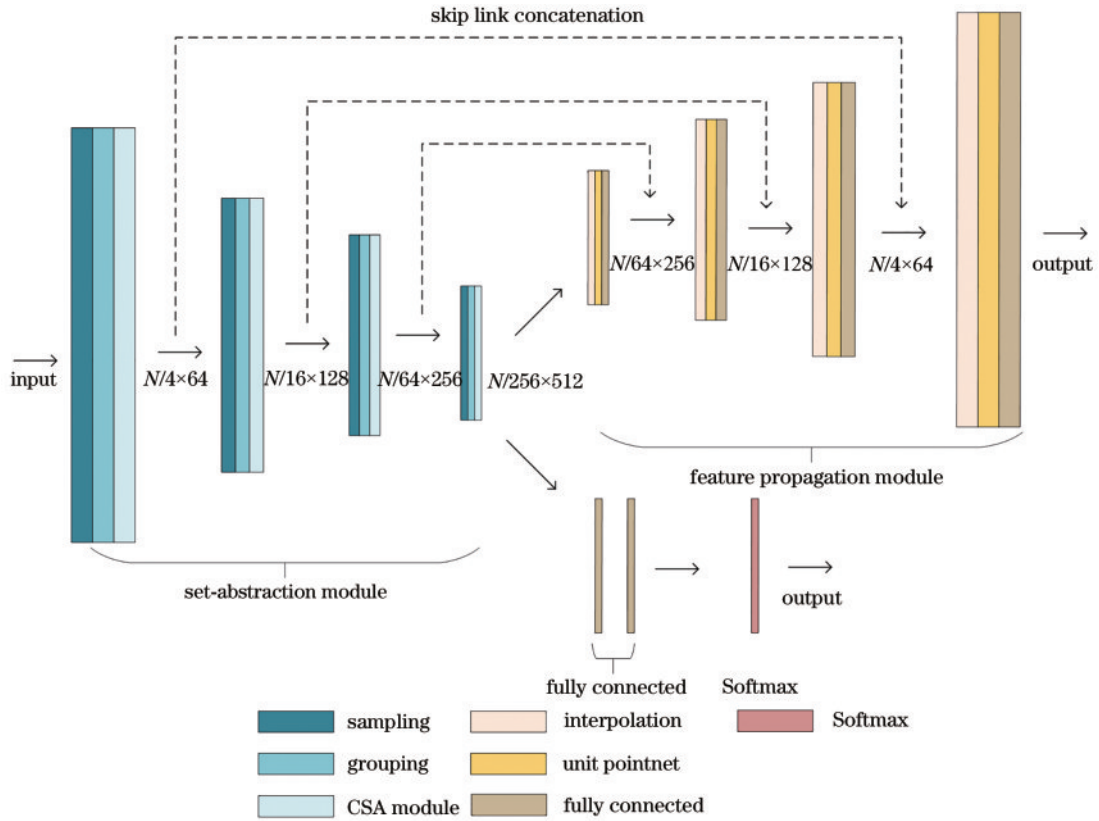


图 4 CSA-PointNet++ 结构概述

Fig. 4 Overview of the CSA-PointNet++ structure

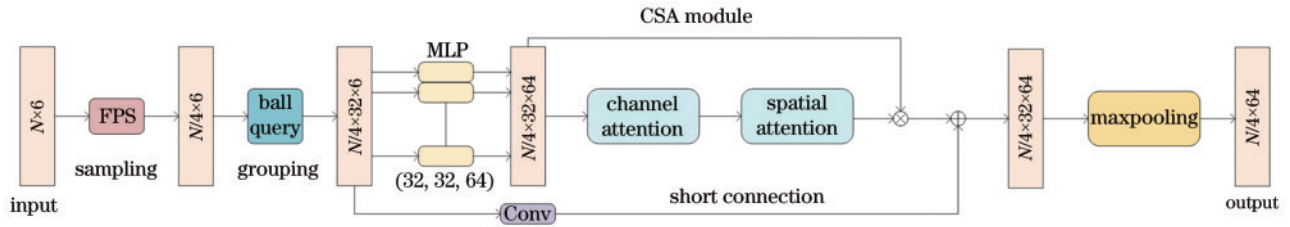


图 5 改进后的集合抽象模块结构

Fig. 5 Structure of the improved set abstraction module

CSA 模块发掘点云在不同特征维度信息和空间位置信息的重要性,获得不同特征通道-空间维度的权重,从而提升网络在分类和分割任务上的表现。为了避免权值接近零时出现特征缺失的情况,采用短连接的方式将 CSA 层前后的特征连接起来。由图 5 可以看出,点云数据在输入到 CSA 层前后的特征维度不同,因此在短连接中使用卷积操作对其维度进行匹配。

最后利用最大池化操作提取每个局部区域的局部特征,这些提取到的局部特征在网络的下一级被分组到更大的单元,然后将它们作为输入的一部分来提取更高一级的特征。如图 4 所示,对于更深层次的集合抽象模块,卷积核的数量比前一层多,经过 4 个集合抽象层后编码器的输出大小为  $N/256 \times 512$  (其中 512 为特征维数),包含了更多有用的通道-空间特征信息。对于分类任务,将编码器降采样得到的全局特征送入多层全连接网络,最后通过一个 Softmax 网络进行

分类。

### 2.2.3 解码器网络

PointNet++<sup>[15]</sup>设计了一种反向插值的方法来实现上采样的解码器结构,解码器网络包括 4 个特征传播模块,是一个上采样的过程。为了将学习到的特征从采样点传播到原始点,首先在特征传播模块内利用反距离加权进行插值。经过第一个特征传播模块,点的特征从  $N/256 \times 512$  个点传播到  $N/64 \times 256$ ,其中  $N/256$  是第四个集合抽象模块的输出点集的数量。对于分割任务,将编码器的输出送入解码器网络进行操作。

由于得到的逐点特征是从解码器的上一层得到的,是全局的信息,还需要得到局部信息。为了获取全局上下文信息,捕获点与点之间的长期依赖关系,将每个特征传播模块上的插值特征通过跳过连接与集合抽象模块中对应层的点特征连接,对二者进行拼接来获

取局部信息。然后,通过类似于CNN中的 $1 \times 1$ 卷积的‘unit pointnet’,从粗级信息中获取特征。几个共享的全连接和ReLU层更新每个点的特征向量,这个过程不断重复,直到特征被传播到原始点集。通过反向插值和跳过连接在解码器网络中进行逐级上采样,得到了局部和全局的逐点特征,将这些特征应用于分割任务。

与原始的 PointNet++ 相比,改进的 CSA-PointNet++ 对通道域和空间域的对象进行了权值分配,更关注局部特征的重要性,这会提高实验的准确度。

### 3 实验

#### 3.1 数据集

采用几个公开的数据集和真实世界的自建的数据集进行对比实验,公开的数据集是 ModelNet40<sup>[21]</sup>和 ShapeNetPart<sup>[22]</sup>。ModelNet40 共有 40 个大类,是用三角形网格表示的 12311 个 CAD 模型,数据分为 9843 个

训练样本和 2468 个测试样本。ShapeNetPart 是一个由对象的三维 CAD 模型表示的形状存储库,注释丰富。数据集包括 16 个形状类别,每个模型标注了 2 到 6 个部分,总共标注了 50 个部分。该数据集包括 16880 个模型,数据分为 14006 个训练样本和 2874 个测试样本。

真实世界的自建数据集由比激光雷达成本更低的可以近距离采集目标深度图像的 Kinect V2 深度相机采集。Kinect V2 是 Kinect 的第二代产品,主要由 RGB 相机、深度相机和红外激光发射器组成,它的外形和内部结构如图 6 所示。深度相机的测距方式和激光雷达相同,都是采用直接飞行时间测距法,通过测量向目标发射的红外光的返回时间来获取距离信息。通过对采集到的深度信息进行变换,即对方位角( $0 \sim 360^\circ$ )和自旋角( $0 \sim 360^\circ$ )的改变得到三维点云信息,构成数据集,数据集共有 4 个大类,1188 个样本,分为 960 个训练样本和 228 个测试样本。

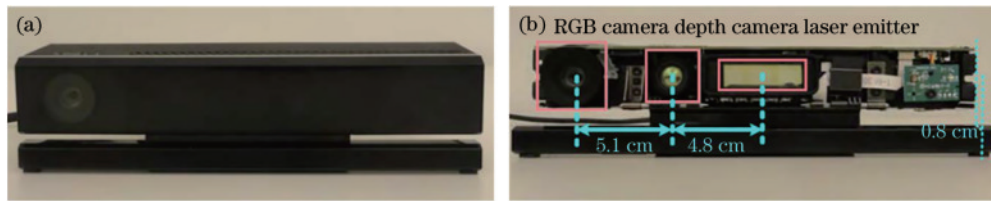


图 6 Kinect V2 相机。(a)外形图;(b)内部结构

Fig. 6 Kinect V2 camera. (a) Appearance; (b) internal structure

得到如图 7(a)所示的点云数据样本,由于原始的样本是在平台上采集的,不同的类的点云数据都会有很大的重合部分,这极大地降低了实验的准确性。因此,对点云进行了一个截取操作,得到如图 7(b)所示的点云数据。并且为了验证所提算法适用于具有平移

不变性和旋转不变性的点云数据,又对点云进行了平移和旋转操作,其中平移变换是分别在  $x$ 、 $y$ 、 $z$  和同时在  $xyz$  三个维度进行 1 个单位长度平移的过程,如图 8 所示。旋转变换是对点云数据在  $x$  轴上( $0 \sim 360^\circ$ , 间隔  $30^\circ$ )进行旋转的过程,如图 9 所示。

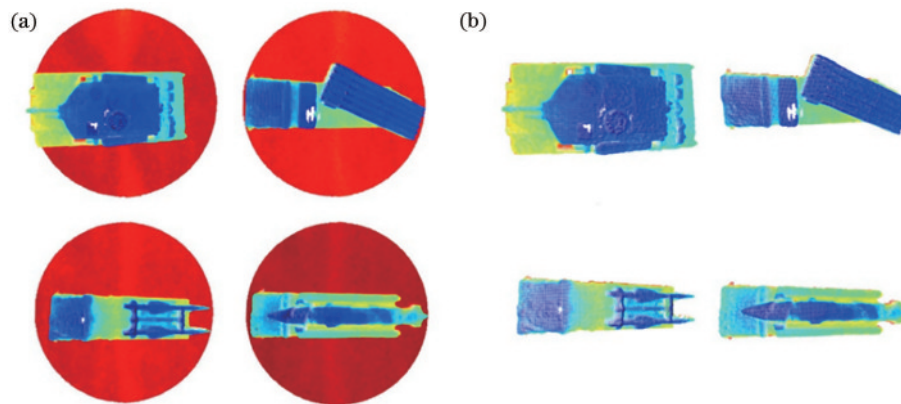


图 7 真实世界的自建点云数据。(a)原始点云数据;(b)截取后的点云数据

Fig. 7 Real-world self-constructed data. (a) Original point cloud data; (b) intercepted point cloud data

#### 3.2 实验设置

在 Ubuntu18.04 系统、PyTorch1.8.1 框架下完成所有测试,并且使用 GTX 2070 SUPER GPU 进行训练。在改进的网络和原 PointNet++ 主干集合抽象层

重合的部分和后续的全连接层使用完全一致的参数和设置,保证在实验过程中不会因为参数设置的不同而对结果造成影响。自建数据集中每个点包括其坐标  $(x, y, z)$ 。参数设置如下:batch\_size 为 24、衰减率为

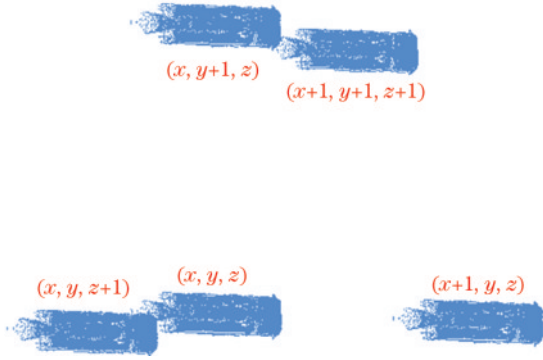


图 8 平移变换后的点云数据

Fig. 8 Point cloud data after translational transformation

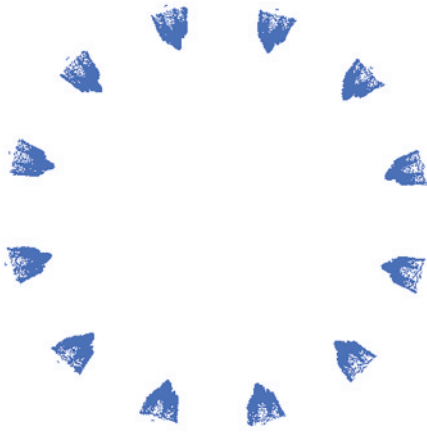


图 9 旋转变换后的点云数据

Fig. 9 Point cloud data after rotational transformation

0.0001、epoch 为 200、学习率为 0.001、样本种类为 40/4、采样点数为 1024。

使用全局准确率(global accuracy)、平均准确率(mean accuracy)、交并比(IoU)和平均交并比(mIoU)来评价 CSA-PointNet++ 的性能。平均准确率和 mIoU 分别是数据中所有类别的准确率和 IoU 的平均值。评价指标的公式分别为

$$A_g = \frac{N_{TP} + N_{TN}}{N_{TP} + N_{TN} + N_{FP} + N_{FN}}, \quad (3)$$

$$A_m = \frac{\sum_k A_k}{K}, \quad (4)$$

$$R_{IoU} = \frac{N_{TP}}{N_{TP} + N_{FP} + N_{FN}}, \quad (5)$$

$$R_{mIoU} = \frac{\sum_k R_{IoU_k}}{K}, \quad (6)$$

式中:  $N_{TP}$ 、 $N_{TN}$ 、 $N_{FP}$ 、 $N_{FN}$  分别为真阳性、真阴性、假阳性、假阴性的点数;  $K$  是类别的数量。

### 3.3 对比实验及结果分析

为了证明引入注意力机制的 CSA-PointNet++ 比原来的网络在分类和分割任务上具有更好的性能,进行如下几个实验:比较原始 PointNet++<sup>[15]</sup> 和 CSA-

PointNet++ 的分类结果,评估了所提网络在分类性能上的提升;比较原始 PointNet++<sup>[15]</sup> 和 CSA-PointNet++ 的部件分割结果,评估了所提网络在部件分割性能上的提升;通过消融实验评估 CA 模块和 SA 模块的有效性;比较 CSA-PointNet++ 在带平台和不带平台的真实世界自建数据集上的分类结果,验证了所提网络的泛化性;比较 CSA-PointNet++ 在经过旋转和平移变换后的自建数据集上的分类结果,结果表明该网络能够正确识别旋转和平移后的数据,进一步证明所提网络在真实数据上具有良好的泛化能力。

#### 3.3.1 公开数据集下的分类实验

为了证明引入注意力机制的 CSA-PointNet++ 比原来的网络在分类任务上具有更好的性能,采用 ModelNet40 数据集<sup>[21]</sup> 进行实验。在单位球内对模型进行归一化,并且根据 PointNet++<sup>[15]</sup> 的设置对模型的表面均匀采样 1024 个点,以生成点云目标。使用 3 个全连接层进行分类实验,对比实验如下:首先,对原始的 PointNet<sup>[14]</sup>、PointNet++<sup>[15]</sup> 和 PCT<sup>[24]</sup> 网络进行复现实验;其次,对 CSA-PointNet++ 进行实验,验证改进网络的性能提升。为了找出不同参数设置下实验准确度最高的情况,利用控制变量的方法进行实验:分别在通道注意力和空间注意力引入一个缩放参数  $r$ ,把缩放参数  $r$  设置为 (4, 4)、(4, 16)、(16, 16)、(16, 4);不使用 ReLU 函数。分类实验准确率(ACC)如表 1 所示。

表 1 ModelNet40 数据集下的分类结果(ACC)

Table 1 Classification results under ModelNet40 dataset (ACC)

Method	ACC / %
MVCNN	90.10
VoxNet	85.90
PointNet	90.32
PointNet++	91.89
DGCNN	92.20
PCT	93.07
Ours( $r=4,4$ )	93.20
Ours( $r=4,16$ )	93.16
Ours( $r=16,16$ )	93.10
Ours( $r=16,4$ )	92.55
Ours(without ReLU)	92.56

由表 1 可以看到,当输入的数据大小(1024 点)和特征相同时,所提网络对点云的分类效果相比 PointNet++<sup>[15]</sup> 有明显的提升,当  $r$  为 (4, 4)、(4, 16)、(16, 16) 和 (16, 4) 时,准确率分别提升了 1.31 个百分点、1.27 个百分点、1.21 个百分点和 0.66 个百分点,不使用 ReLU 函数时,准确率提升 0.67 个百分点。

实验结果表明:引入 CSA 模块会提升 PointNet++<sup>[15]</sup> 的特征表示能力,从而提升了网络在

分类任务上的表现性能;其中,当 $r$ 为(4,4)时,实验的准确率最高,比原始 PointNet++<sup>[15]</sup>提高 1.31 个百分点,所提网络在点云分类任务上有优秀的性能。

### 3.3.2 公开数据集下的分割实验

为了证明引入注意力机制的 CSA-PointNet++ 比原来的网络在分割任务上具有更好的性能,使用 ShapeNetPart 数据集<sup>[22]</sup>进行实验。根据 PointNet++<sup>[15]</sup>

的设置,在形状类别已知的情况下,将部件分割任务视为对每个点的分类,在实验中对每个形状采样 2048 个点,对比实验如下:首先,对原始的 PointNet<sup>[14]</sup>、PointNet++<sup>[15]</sup>和 PCT<sup>[24]</sup>等网络进行复现实验;其次,对 CSA-PointNet++ 进行实验,验证改进网络的性能提升。由于在分类实验中已经验证当缩放参数 $r$ 为(4,4)时,网络性能最好,所以将 $r$ 设置为(4,4)。分割结果如表 2 所示。

表 2 ShapeNetPart 数据集下的部件分割实验结果(IoU)

Table 2 Part segmentation results under ShapeNetPart dataset (IoU)

unit: %

Method	mIoU	airplane	bag	cap	car	chair	earphone	guitar	knife
PointNet	79.87	81.73	77.28	<b>87.18</b>	75.01	90.24	73.54	90.81	86.05
PointNet++	81.75	82.21	<b>81.91</b>	83.17	78.41	<b>90.62</b>	73.90	91.15	86.38
DGCNN	82.20	82.01	81.45	84.70	78.83	90.56	73.89	<b>91.47</b>	86.58
PCT	82.31								
Ours	<b>82.62</b>	<b>82.64</b>	80.67	85.35	<b>78.94</b>	90.52	<b>75.90</b>	91.36	<b>87.18</b>
Method	lamp	laptop	motorbike	mug	pistol	rocket	skateboard	table	
PointNet	81.85	95.15	61.27	93.26	79.98	48.17	74.22	82.25	
PointNet++	83.06	95.64	70.91	<b>95.87</b>	80.65	55.19	<b>76.84</b>	82.04	
DGCNN	83.81	<b>95.70</b>	70.37	94.97	81.15	61.53	76.56	81.65	
PCT									
Ours	<b>84.08</b>	95.63	<b>71.33</b>	95.43	<b>82.06</b>	<b>61.91</b>	76.61	<b>82.29</b>	

表 2 列出了点云网络在部件分割任务中的 IoU,通过比较 PointNet++<sup>[15]</sup>和 CSA-PointNet++ 在各个类别上的 IoU,可以看到在大多数类别上 CSA-PointNet++ 的 IoU 都高于 PointNet++<sup>[15]</sup>,特别是在 car、earphone、rocket 等类别上,CSA-PointNet++ 的 IoU 比 PointNet++<sup>[15]</sup>高出 2 个百分点到 6 个百分点。CSA-PointNet++ 的 mIoU 比 PointNet++<sup>[15]</sup>提升 0.87 个百分点。同时,实验结果表明,所提网络在部

件分割任务中有优秀的性能。

图 10 显示了 PointNet++<sup>[15]</sup>和所提网络在 ShapeNetPart 模型<sup>[22]</sup>上的部件分割结果。从图 10 可以发现,所提网络对细节的预测更加清晰,如椅子的腿和面、汽车的轮胎和骨架、刀和刀柄等,更接近真值。实验结果表明,所提网络的部件分割性能优于原始的 PointNet++<sup>[15]</sup>。

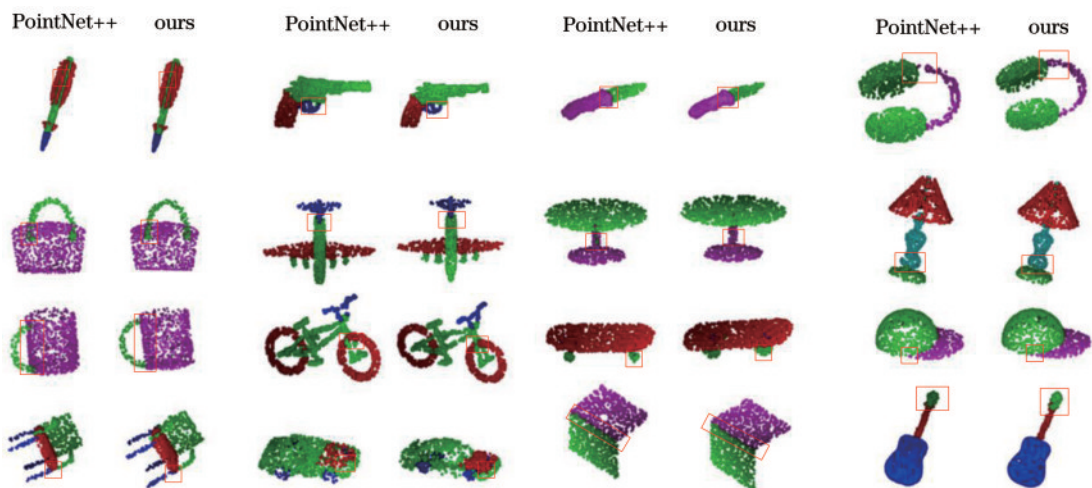


图 10 部件分割结果

Fig. 10 Part segmentation results

### 3.3.3 消融实验

通过消融实验来评估通道注意力和空间注意力模块在点云分类和部件分割任务上的有效性,实验结果

如表 3 和表 4 所示。

实验结果表明,相比 PointNet++, 单独增加通道注意力、单独增加空间注意力、同时增加通道注意力和



表 3 消融实验的 ACC

Table 3 ACC of the ablation experiment

Method	ACC / %
PointNet++	91.89
CA+PointNet++	92.93
SA+PointNet++	92.99
CSA+PointNet++	93.20
SCA+PointNet++	92.20

表 4 消融实验的部件分割结果(mIoU)

Table 4 Part segmentation results of the ablation experiment (mIoU)

Method	mIoU / %
PointNet++	81.75
CA+PointNet++	82.15
SA+PointNet++	82.42
CSA+PointNet++	82.62
SCA+PointNet++	82.13

空间注意力、交换通道注意力和空间注意力的位置后,网络的分类准确率分别提升了1.04个百分点、1.10个百分点、1.31个百分点、0.31个百分点,mIoU分别提升了0.40个百分点、0.67个百分点、0.87个百分点和0.38个百分点。即单独增加空间注意力或通道注意力模块都可以显著提高PointNet++<sup>[15]</sup>的性能,而同时使用这两种模块对性能的提升最多。因此可以得出结论,通道注意力和空间注意力模块都是有效的,它们可以使特征更独特,并且二者顺序的改变会对实验结果有较大的影响。

### 3.3.4 真实世界的自建数据集下的分类实验

为了验证CSA-PointNet++的泛化性,对真实世界的自建数据集进行分类实验。分别对自建数据集的带平台数据和不带平台数据进行实验,实验结果如表5所示。

表 5 自建数据集下的 ACC

Table 5 ACC under the real-world self-constructed dataset

Type	ACC / %
With ground	52.81
Without ground	92.14

由表5可知,在真实世界的自建数据集中,带平台的分类准确率远远低于不带平台的,这是因为带平台的不同种类的数据重合部分很多,难以提取有用特征信息,从而影响实验的准确性。对截取后的数据集进行实验,实验结果表明所提网络可以达到较高的精度,验证了所提网络的泛化能力。

### 3.3.5 变换后的真实世界的自建数据集下的分类实验

为了验证CSA-PointNet++适用于具有平移不

变性的点云数据,对真实世界的自建数据集进行平移变换,分别在 $x$ 轴、 $y$ 轴、 $z$ 轴的正方向上平移1个单位长度,又同时在 $xyz$ 轴的正方向上分别平移1个单位长度,进行对比实验,实验结果如表6所示。

表 6 平移后的真实世界的自建数据集下的 ACC

Table 6 ACC of the translated real-world self-constructed dataset

Translation	ACC / %
$(x+1, y, z)$	90.96
$(x, y+1, z)$	91.85
$(x, y, z+1)$	90.85
$(x+1, y+1, z+1)$	91.70

同样地,为了验证所提网络适用于具有旋转不变性的点云数据,对真实世界的自建数据集进行旋转变换,旋转角度从0到360°,间隔为30°,实验结果如表7所示。

表 7 旋转后的真实世界的自建数据集下的 ACC

Table 7 ACC of the rotated real-world self-constructed dataset

Rotation	ACC / %
Rot_30°	91.75
Rot_60°	89.59
Rot_90°	92.91
Rot_120°	90.99
Rot_150°	91.40
Rot_180°	90.59
Rot_210°	91.00
Rot_240°	90.01
Rot_270°	90.93
Rot_300°	91.37
Rot_330°	92.55

由表6和表7可知,在经过平移变换和旋转变换的真实世界的自建数据集上的分类实验结果与原始自建数据集接近,证明所提网络能够正确识别旋转和平移后的数据,进一步证明所提网络在真实数据上具有良好的泛化能力。

## 4 结 论

提出了一种基于注意力机制的点云特征提取方法CSA-PointNet++,将通道-空间注意力机制集成到PointNet++中,用于点云数据的特征提取与分类分割任务。该网络是编码器-解码器网络,既能从非均匀的采样数据中提取点之间的局部几何关系,同时又能增强重要特征通道,提高了点云分类和分割精度。在几个公开数据集和真实世界的自建数据集上测试了CSA-PointNet++,在公开的数据集上的实验结果表明,该方法的分类准确率和分割mIoU比原始的PointNet++得到了提升。在真实世界的自建数据集

上的实验结果表明该网络在点云分类任务中有良好的性能,并且通过对经过平移和旋转变换后的点云数据进行实验,验证了该网络适用于具有平移不变性和旋转不变性的点云数据,具有良好的泛化能力和广泛的应用前景。

综上,与原始的 PointNet++ 相比,引入了 CA 和 SA 机制的 CSA-PointNet++ 能够学习通道和空间维度的重要性,从而提升网络在点云分类和分割任务上的表现,达到了最优性能,并且对真实世界的点云数据具有泛化能力。

### 参 考 文 献

- [1] Jiang M Y, Wu Y R, Zhao T Q, et al. PointSIFT: a SIFT-like network module for 3D point cloud semantic segmentation[EB/OL]. (2018-07-02)[2023-02-04]. <https://arxiv.org/abs/1807.00652>.
- [2] 陈西江, 林嘉颖, 韩贤权, 等. 基于指数函数密度聚类模型的室内物体提取[J]. 中国激光, 2022, 49(11): 1104001.  
Chen X J, Lin J Y, Han X Q, et al. Indoor object extraction based on exponential function density clustering model[J]. Chinese Journal of Lasers, 2022, 49(11): 1104001.
- [3] 李立刚, 郭玉杰, 李林, 等. 基于变尺寸栅格地图的船载激光雷达目标检测[J]. 激光与光电子学进展, 2022, 59(8): 0828002.  
Li L G, Guo Y J, Li L, et al. Target detection of shipborne lidar based on variable size grid map[J]. Laser & Optoelectronics Progress, 2022, 59(8): 0828002.
- [4] 梅圣明, 黄妙华, 柳子晗, 等. 基于三维激光雷达的复杂场景中地面分割方法[J]. 激光与光电子学进展, 2022, 59(10): 1028003.  
Mei S M, Huang M H, Liu Z H, et al. Ground segmentation method in complex scenes based on three-dimensional lidar[J]. Laser & Optoelectronics Progress, 2022, 59(10): 1028003.
- [5] Simon M, Amende K, Kraus A, et al. Complexer-YOLO: real-time 3D object detection and tracking on semantic point clouds[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), June 16-17, 2019, Long Beach, CA, USA. New York: IEEE Press, 2020: 1190-1199.
- [6] Zhou Y, Tuzel O. VoxelNet: end-to-end learning for point cloud based 3D object detection[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 4490-4499.
- [7] Yan Y, Mao Y X, Li B. SECOND: sparsely embedded convolutional detection[J]. Sensors, 2018, 18(10): 3337.
- [8] Zia M Z, Stark M, Schindler K. Are cars just 3D boxes? jointly estimating the 3D shape of multiple objects[C]//2014 IEEE Conference on Computer Vision and Pattern Recognition, June 23-28, 2014, Columbus, OH, USA. New York: IEEE Press, 2014: 3678-3685.
- [9] Wu B C, Wan A, Yue X Y, et al. SqueezeSeg: convolutional neural nets with recurrent CRF for real-time road-object segmentation from 3D LiDAR point cloud[C]//2018 IEEE International Conference on Robotics and Automation (ICRA), May 21-25, 2018, Brisbane, QLD, Australia. New York: IEEE Press, 2018: 1887-1893.
- [10] Engelcke M, Rao D, Wang D Z, et al. Vote3Deep: fast object detection in 3D point clouds using efficient convolutional neural networks[C]//2017 IEEE International Conference on Robotics and Automation (ICRA), May 29-June 3, 2017, Singapore. New York: IEEE Press, 2017: 1355-1361.
- [11] 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), September 28-October 2, 2015, Hamburg, Germany. New York: IEEE Press, 2015: 922-928.
- [12] Graham B, Engelcke M, van der Maaten L. 3D semantic segmentation with submanifold sparse convolutional networks[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 9224-9232.
- [13] Guo Y L, Wang H Y, Hu Q Y, et al. Deep learning for 3D point clouds: a survey[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 43(12): 4338-4364.
- [14] Charles R Q, Hao S, Mo K C, et al. PointNet: deep learning on point sets for 3D classification and segmentation[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 77-85.
- [15] Qi C R, Yi L, Su H, et al. PointNet++: deep hierarchical feature learning on point sets in a metric space [EB/OL]. (2017-06-07)[2023-02-05]. <https://arxiv.org/abs/1706.02413>.
- [16] Hu J, Shen L, Sun G. Squeeze-and-excitation networks [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 7132-7141.
- [17] Wang Q L, Wu B G, Zhu P F, et al. ECA-net: efficient channel attention for deep convolutional neural networks [C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 11531-11539.
- [18] Li X, Wang W H, Hu X L, et al. Selective kernel networks[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2020: 510-519.
- [19] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York:

- IEEE Press, 2016: 770-778.
- [20] Woo S, Park J, Lee J Y, et al. CBAM: convolutional block attention module[M]//Ferrari V, Hebert M, Sminchisescu C, et al. Computer vision-ECCV 2018. Lecture notes in computer science. Cham: Springer, 2018, 11211: 3-19.
- [21] Wu Z R, Song S R, Khosla A, et al. 3D ShapeNets: a deep representation for volumetric shapes[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 7-12, 2015, Boston, MA, USA. New York: IEEE Press, 2015: 1912-1920.
- [22] 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.
- [23] Jing Z W, Guan H Y, Zhao P R, et al. Multispectral LiDAR point cloud classification using SE-PointNet++ [J]. Remote Sensing, 2021, 13(13): 2516.
- [24] Guo M H, Cai J X, Liu Z N, et al. PCT: point cloud transformer[J]. Computational Visual Media, 2021, 7(2): 187-199.