

# 基于密度自适应深度网络的点云场景语义分割算法

卫刚, 赵安铭, 王志成

(同济大学电子与信息工程学院, 上海 201804)

**摘要:** 由于传感器噪声干扰, 点云密度不均匀, 场景复杂多样以及物体之间存在遮挡现象等问题, 使得三维点云场景语义分割问题的研究工作极具挑战性。针对三维点云数据采样密度不均匀以及图卷积网络深度有限的问题, 提出一种密度自适应的方法。该方法通过多层感知器学习一个权重函数, 利用核密度估计学习一个密度函数, 对非均匀采样的点云数据进行卷积操作。同时, 受深度学习在图像领域的启发, 引入残差连接、空洞卷积等结构, 训练更深层的点云分割网络。该算法在多个点云分割的标准数据集上取得了优秀的性能。

**关键词:** 点云分割; 图卷积; 三维点云; 深度学习

中图分类号: TP391.4

文献标志码: A

## Point Cloud Scene Semantic Segmentation Algorithm Based on Density Adaptive Deep Network

WEI Gang, ZHAO Anming, WANG Zhicheng

(College of Electronics and Information Engineering, Tongji University, Shanghai 201804, China)

**Abstract:** Sensor noise interference, uneven point cloud density, complex and diverse scenes, and the occlusion between objects pose great challenges to the research of the semantic segmentation of 3D point cloud scenes. In the view of the uneven sampling density of 3D point cloud data and limited depth of graph convolutional networks, this paper proposes a density adaptive method, which uses a multi-layer perceptron to learn a weight function, and uses kernel density estimation to learn a density function. The convolution operation is performed on the evenly sampled point cloud data. At the same time, inspired by deep learning in the image field, residual connection, hole convolution, and other structures were introduced to train deeper point cloud segmentation

networks. The algorithm achieves an excellent performance on standard data sets of multiple point cloud segmentation.

**Key words:** point cloud segmentation; graph convolution; 3D point cloud; deep learning

近年来,随着三维扫描技术的快速发展,三维点云广泛应用于无人驾驶、建筑设计、遥感测绘及虚拟现实等计算机视觉领域。作为三维场景和分析的重要课题之一,点云语义分割一直是三维视觉和计算机图形学范畴的重要研究问题,也是目标识别、场景理解和三维重建等任务的基础。由于采集过程中的传感器噪声干扰、点云密度不均匀、场景复杂多样以及物体之间存在遮挡现象等问题,三维点云场景语义分割问题研究工作极具挑战性。现有传统的点云分割方法的识别准确率还存在较大提升空间。目前,深度学习方法已广泛应用于二维图像数据,但将深度学习方法应用于三维场景仍面临巨大的挑战,有大量的基础性工作需要完成。由于深度学习方法表现出较好的高层语言理解能力,基于深度学习的点云语义分割已成为当前研究的热点。

卷积神经网络(convolutional neural networks, CNN)成功应用的一个关键因素是能够训练和应用深层神经网络。但是,最近的工作<sup>[1-3]</sup>表明,图卷积神经网络(graph convolutional network, GCN)很难扩展到深层网络。原因是叠加多个图卷积层会提高反向传播的复杂性和出现常见的梯度消失问题。由于上述限制,大多数最先进的GCN不超过4层。ResNet<sup>[4]</sup>的提出有效地改善了CNN上的梯度消失问题。ResNet通过在输入层和输出层之间引入残差连接,大大减轻了梯度消失问题。Yu等<sup>[5]</sup>提出空

收稿日期: 2020-10-13

基金项目: 国防基础研究计划(JCKY2020206B037)

第一作者: 卫刚(1973—),男,副研究员,工学博士,主要研究方向为计算机应用、人工智能、计算机辅助设计。

E-mail: weigang@tongji.edu.cn

通信作者: 赵安铭(1991—),男,硕士生,主要研究方向为计算机视觉。E-mail: vistart@tongji.edu.cn



论文  
拓展  
介绍

洞卷积(Dilated convolutions),在不增加更多计算量的同时能够增大感受野。受上述关键概念的巨大成功所启发,将上述想法应用到GCN中,使得更深层次的GCN能够很好地聚合并获得优异的性能。

为了成功训练更深层的GCN、减轻梯度消失问题,借鉴了训练深层卷积神经网络的相关概念,引入残差连接、空洞卷积等结构,训练更深层的点云分割网络。同时,由于上述工作的贡献未考虑到三维点云数据无序性、稀疏性和不规则性的特点,传统的卷积神经网络因其计算卷积时需要规则输入和输出,其不适合直接处理点云数据。为了对非均匀采样的点云数据进行卷积操作,提出了一种密度自适应方法。

## 1 研究背景

深度学习的发展推动了计算机视觉的进步。目前,深度学习已广泛应用于二维图像的处理<sup>[4,6-7]</sup>,但是将其运用到三维数据仍面临巨大的挑战。由于点云具有不规则且无序性的问题,为了使得点云适用于卷积神经网络,研究者们将其转换成规则的结构,再输入到网络中进行处理。目前,根据输入到网络中的数据格式,基于深度学习的点云分割算法主要分为以下几类。

### 1.1 基于多视图的方法

受深度学习在二维图像上取得较好效果的启发,基于多视图的方法将3D点云转化为一系列的多视角拍摄的2D渲染图,将这些产生的二维图像作为训练数据,然后用成熟的2D卷积神经网络来提取特征,从而进行识别或分割任务。在这方面比较具有代表性的方法就是Multi-view CNN<sup>[8]</sup>。Multi-view CNN使用ImageNet预训练好的视觉几何组(visual geometry group, VGG)网络来提取特征,将这些特征组合在一起,再进一步输入到可训练的CNN网络中进一步进行特征学习,然后输出分类结果。Qi等<sup>[9]</sup>研究了基于视图和体积方法的三维形状分类组合。2017年,Boulch等提出SnapNet<sup>[10]</sup>,其思想是使用2D神经网络分别对一系列成对的RGB图和深度图进行处理以生成3D场景的图像。为了改善分割效果,在SnapNet的基础上,Guerry等<sup>[11]</sup>提出SnapNet-R对多个视图直接处理以获取密集的3D点标记。但是SnapNet-R对物体边界的分割精度仍有待加强。由此可见,基于多视图的点云分割方法没有充分利用点云数据的空间信息,同时对于物体

的不同投影角度也会对结果产生影响。虽然这种方法取得了非常好的分类识别结果,但是这类方法容易造成三维结构信息的丢失。另一方面,投影角度的选取与同一角度的投影对物体的表征能力也不同,对网络的泛化能力有一定的影响。

### 1.2 基于体素的方法

基于体素的方法,通过把不规则和无序的点云数据进行体素化操作,使点云数据变成规则化的三维栅格结构,然后使用三维的卷积神经网络架构进行训练。Wu等<sup>[12]</sup>于2015年提出一种基于体素的三维深度信念网络3D ShapeNets。该算法将点云数据表示为一系列的三维栅格,然后用卷积神经网络进行训练。Maturana等提出一种面向三维体素的三维卷积神经网络VoxNet<sup>[13]</sup>。基于体素的方法通过网络化点云数据提供了点云结构信息。但是这种方法只使用了点云结构信息,丢失了点云的颜色信息、法向量信息等其他重要特征。而且,三维卷积运算的计算时间复杂度高,造成网络效率低下。因此,为了减小网络的计算量,通常会减小体素的分辨率,如同图像中减小图像输入尺寸,但是这样操作会造成量化误差的问题。针对上述问题,一些研究人员提出了改进的方案。Klokov等<sup>[14]</sup>提出采用KD树(k-dimensional tree)来规则化点云数据的深度网络模型,称为Kd-Net。Gernot等<sup>[15]</sup>利用非平衡八叉树来表征点云结构,提出OctNet网络模型。Tchapmi等提出一个3D点级分割的端到端框架SEGCloud<sup>[16]</sup>,该框架将点云细分为体素网格,结合了卷积神经网络、三线性插值和全连接条件随机场的优点。这些方法虽然对基于体素的方法做了不同方面的改进,但是对于体素化造成的量化误差没能解决。

### 1.3 基于点云的方法

点云数据本身具有很多特征信息,如果能直接利用这些信息,不仅能减少预处理的过程,而且能更充分地挖掘点云特征信息。因此,近几年基于原始点云的深度网络模型逐渐被提出。Qi等在2017年提出第一个直接对点云数据进行处理而不做预处理的深度学习网络架构PointNet<sup>[17]</sup>。PointNet通过对称函数来处理点云的无序性,使用空间变换网络解决物体姿态变换不变性。虽然PointNet在点云特征提取方面取得了突破性的进展,但是PointNet没有充分利用点云数据的局部特征。为了解决上述问题,PointNet++<sup>[18]</sup>提出一种局部特征提取方案。它根据距离度量将点集划分为重叠的局部区域,然后利用PointNet网络结构提取局部区域的点云特

征,然后把局部区域进一步分组为更大的单元,并进行处理以生成更高级别的特征。虽然该模型改善了局部特征提取的问题,但是点特征是以独立和孤立的方式学习的,忽略了与相邻点及其特征的相对关系。Engelmann等<sup>[19]</sup>在PointNet的基础上,为了合并更大范围的空间上下文信息,提出了输入级上下文和输出级上下文2个扩展,成功应用于室内和室外点云语义分割任务当中。为了使CNN更好地处理无序的点云数据,PointCNN<sup>[20]</sup>提出学习一种X变换算子,即通过深度网络学习一个置换矩阵,对输入数据进行排序和加权。针对处理大场景三维点云语义分割的需求, Landrieu等<sup>[21]</sup>提出超点图(superpoint graph, SPG),具有丰富的边缘特征,可对3D点云中对象部分之间的上下文关系进行编码,可以处理大规模的点云场景。Wang等<sup>[23]</sup>提出动态图卷积神经网络(dynamic graph CNN, DGCNN),首次使用了图卷积神经网络来改进点云中的特征提取工作。该算法构建了EdgeConv算子,用于提取中心点的特征和中心点与 $K$ 近邻域点的边缘向量。Su等<sup>[24]</sup>于2018年提出针对点云处理的稀疏点阵网络SPLATNet (SParse LATtice Networks)。SPLATNet直接将点云数据作为输入,并使用稀疏而高效的点阵滤波器来计算分层和空间感知的特征。此外,SPLATNet允许将2D信息轻松映射到3D,用于点云和多视图图像的联合处理。Jiang等<sup>[25]</sup>受到传统图像中的尺度不变特征变换(SIFT)的启发,设计了一种可用于点云特征提取的模块PointSIFT。该模块用于对三维点云在不同方向上的信息进行编码,具有尺度不变性。

## 2 算法思想

设图 $G=(V,E)$ 表示点云的局部结构,其中 $V$ 是无序顶点的集合, $E$ 是表示顶点 $v \in V$ 之间的连通性的边的集合。如果 $e_{i,j} \in E$ ,则顶点 $v_i$ 和 $v_j$ 通过边 $e_{i,j}$ 彼此连接。

### 2.1 局部特征

PointNet网络<sup>[17]</sup>仅对点进行孤立的特征提取,从而缺乏对局部点云特征的学习。PointNet++<sup>[18]</sup>在PointNet基础上对点云进行区域采样,并使用PointNet作为区域特征提取器,逐层提取并整合局部特征至全局特征。然而PointNet++中仍然用到了PointNet,意味着在采样的区域内,点的特征是单独提取的,对于局部特征的学习仍然不够充分。

EdgeConv<sup>[23]</sup>提出获取点云局部特征的一种方法。不是直接聚合邻域特征,而是建议首先通过从中心特征中减去中心顶点的特征来获取每个邻域的局部邻域信息。然而网络的深度限制了对于更细粒度特征的提取。为了训练更深的GCN,向EdgeConv添加残差连接和膨胀图卷积。将中心点的特征与两点的特征差串联后输入多层感知机(multi-layer perception, MLP),这样边特征就融合了点之间的局部关系与点的全局信息。

### 2.2 残差连接

传统的卷积神经网络存在致梯度消失问题,导致无法训练很深的神经网络。He等<sup>[4]</sup>提出的ResNet在一定程度上解决了上述问题,它设计了一种跳跃连接(skip connection)结构,使得网络具有更强的恒等映射(identity mapping)的能力,从而拓展了网络的深度,同时也提升了网络的性能。借鉴残差学习在CNN的成功应用,将该结构用于图卷积网络当中,解决图卷积网络层数不能过深、梯度消失的问题。

考虑通过卷积网络由 $L$ 层组成,每层实现一个非线性变换 $H_l(\cdot)$ ,其中 $l$ 表示网络第 $l$ 层。定义 $H_l(\cdot)$ 可以是诸如卷积(Conv)、池化(Pooling)、校正线性单元(ReLU)<sup>[26]</sup>或批处理规范化(BN)<sup>[27]</sup>等操作的函数。分别用 $G_l=(V_l, E_l)$ 和 $G_{l+1}=(V_{l+1}, E_{l+1})$ 表示网络第 $l$ 层的输入和输出。网络第 $l$ 层的输出作为第 $l+1$ 层的输入,然后经过第 $l+1$ 层得到 $G_{l+1}=H_l(G_l, W_l)$ ,其中 $W_l$ 是第 $l$ 层可学习的权重参数集合。在这里,提出了一种图残差学习框架,该框架通过拟合另一个残差映射 $F$ 来学习所需的底层映射 $H$ 。在 $G_l$ 通过残差映射 $F$ 转换后,执行逐点加法得到 $G_{l+1}$ 。

$$G_{l+1} = H(G_l, W_l) = F(G_l, W_l) + G_l = G_{l+1}^{\text{res}} + G_l \quad (1)$$

如图1所示,残差映射 $F$ 学习将图作为输入,并为下一层输出残差图形 $G_{l+1}^{\text{res}}$ 。

### 2.3 空洞卷积

传统卷积神经网络通过池化层进行下采样来减少图像大小,同时增大感受野,然后再通过上采样操作将图像恢复成原大小。但是,下采样会降低图像分辨率,造成空间信息损失。Yu等<sup>[5]</sup>提出的空洞卷积(dilated convolutions)可以在一定程度上避免这个问题。空洞卷积是在普通的卷积核内加入空洞,相当于在卷积核相邻2个元素直接加入零元素。空洞

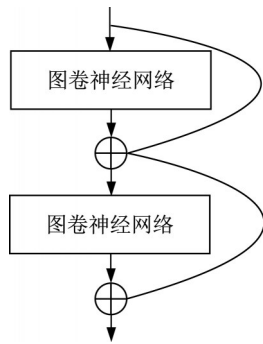


图1 残差连接示意

Fig. 1 Residual connection

卷积网络本质上实现了去掉下采样操作的同时不降低网络的感受野。将空洞卷积的思想应用于图卷积神经网络当中,以减轻由池化聚集操作造成的空间信息损失。

首先需要构建一个扩张领域,在每一个GCN层后使用扩张K近邻(dilated K-NN)去寻找扩张领域,并构建了一个扩张图(dilated graph)。具体来说,对于一个扩张K近邻的输入图,用 $d$ 表示扩张率,扩张K近邻通过跳过每一个 $d$ 邻域,返回 $k \times d$ 邻域内 $k$ 个最近邻居。最近邻是根据预先定义的距离度量确定的,在实验中,使用 $L_2$ 距离,即欧几里得距离(Euclidean Distance)。

让 $N(v)$ 定义为顶点 $v$ 的 $d$ -dilated的邻居,如果 $(u_1, u_2, \dots, u_{k \times d})$ 是排序了的前 $k \times d$ 近的邻居,那么顶点 $(u_1, u_{1+d}, u_{1+2d}, \dots, u_{1+(k-1)d})$ 是顶点 $v$ 的 $d$ -dilated的邻居。因此,输出图的边定义在一组 $d$ -dilated顶点邻域上,如图2a所示。具体地说,存在一个从顶点 $v$ 到每个顶点的有向边,其中GCN聚合和更新函数是基于Dilated k-NN创建的,应用于获得所有顶点的输出顶点的特征,如图2b所示。将更新后的顶点表示为 $v'$ ,然后由具有扩张率为 $d$ 的扩张图卷积(dilated graph convolution)生成的输出图中均匀抽样邻域以很小的概率执行扩张聚合,从而进行随机聚合。

#### 2.4 密度自适应

当前传感器采集到的点云密度分布差异非常大。考虑到传感器采集到的3D点云的不均匀性,提出使用密度函数对学到的权重进行加权。点云表示为一组3D点 $\{P_i | i=1, \dots, n\}$ ,其中每个点 $P_i$ 是其 $(x, y, z)$ 坐标加上额外特征通道(例如颜色、法线等)的向量。为简单起见,除非另有说明,这里仅使用 $(x, y, z)$ 坐标作为点的通道。在邻域 $G$ 中,3D点

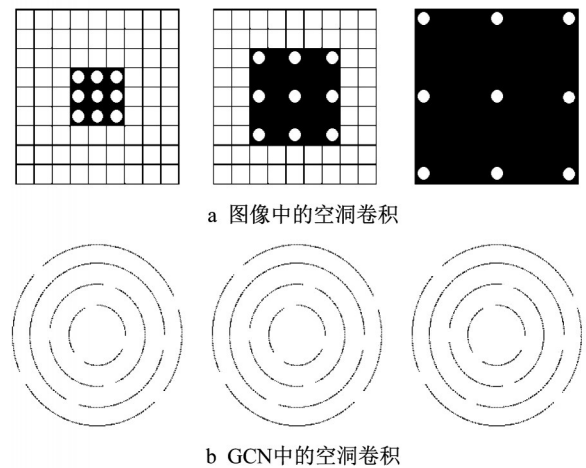


图2 图像中的空洞卷积和GCN中的空洞卷积

Fig. 2 Dilated convolutions of image and dilated convolutions of GCN

的相对坐标表示为 $(x_\mu, y_\mu, z_\mu)$ 。函数 $F(x + x_\mu, y + y_\mu, z + z_\mu)$ 是以点 $p = (x, y, z)$ 为中心的局部区域 $G$ 中一个点的特征。在每个局部区域, $(x_\mu, y_\mu, z_\mu)$ 可以是局部区域中的任意一点位置。连续函数 $W(x_\mu, y_\mu, z_\mu)$ 表示每个点对应的特征 $F$ 的权重,其输入是以 $(x, y, z)$ 为中心的邻域内的点的相对坐标。函数 $W$ 可以使用多层感知器(MLP)来近似。定义密度系数函数 $D(x_\mu, y_\mu, z_\mu)$ ,它的输入是点的密度,它的输出是每个点对应的密度系数。

$$F_{\text{out}} = \sum_{(x_\mu, y_\mu, z_\mu) \in N} D(x_\mu, y_\mu, z_\mu) \cdot$$

$$W(x_\mu, y_\mu, z_\mu) F_{\text{in}}(x + x_\mu, y + y_\mu, z + z_\mu) \quad (2)$$

文献[28]考虑了权重函数的近似,但没有考虑对采样密度的近似,因此不是连续卷积算子的完全近似。本文提出的权重函数通过多层感知机从三维坐标中近似,密度函数利用一个核密度估计<sup>[29]</sup>以及一个非线性变换(MLP)近似实现。MLP在所有点之间共享权重,可以保持排列不变性。为了计算密度尺度估计函数 $D(x_\mu, y_\mu, z_\mu)$ ,首先使用核密度估计离线估计点云中每个点的密度,然后将密度输入MLP以进行一维非线性变换。

#### 2.5 网络架构与分析

提出的网络模型(如图3所示)由GCN模块、融合模块和MLP预测模块组成。GCN主干块将具有4 096个点的点云作为输入,堆叠28个具有动态k-NN的EdgeConv层,每个层都与DGCNN<sup>[22]</sup>中使用的层相似。通过应用连续的GCN层以聚合局部信

息来提取特征,并输出具有4 096个顶点的学习图形表示。然后,在每一个GCN块中添加动态膨胀k-NN和残差图连接。融合和MLP预测块遵循与PointNet++<sup>[18]</sup>和DGCNN<sup>[23]</sup>类似的架构。融合块用于融合全局和多尺度局部特征。它从每个GCN层的GCN主干块中提取出的顶点特征作为输入,并将这些特征连接起来,然后将它们传递给 $1 \times 1$ 卷积

层,然后进行最大池化。后一层将整个图形的顶点特征聚合到单个全局特征向量中,而后者又与来自所有先前GCN层的每个顶点的特征(全局信息和局部信息的融合)串联在一起。MLP预测块将3个MLP层应用于每个点的融合特征,以预测其类别。实际上,这些层是 $1 \times 1$ 卷积。

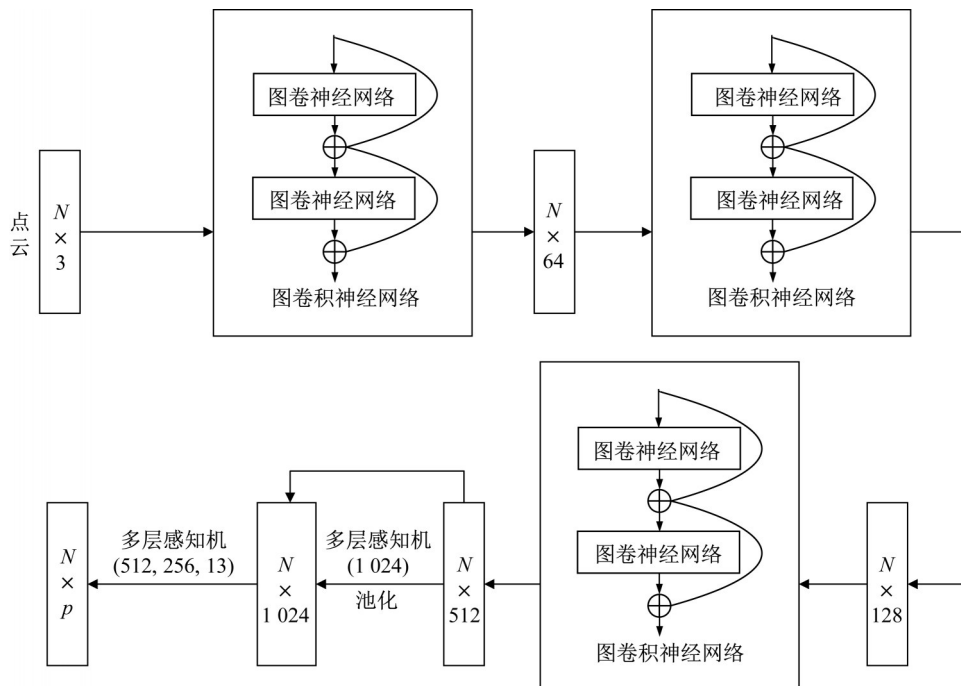


图3 网络架构

Fig. 3 Network architecture

### 3 实验

为了评估本文提出的网络结构的有效性,对点云分割任务进行了广泛的实验。从实验结果可知,本文方法可以显著提高分割准确度。此外,还进行了全面的消融研究,以显示框架中不同结构的效果。

#### 3.1 三维模型零件分割实验

零件分割是一项具有挑战性的点云识别和分割任务。同样在ShapeNet数据集上进行零件分割实验。ShapeNet数据集包含来自16个类别的50个零件的16 881个形状。该任务的输入是点云数据,任务的目标是为点云中的每个点分配一个零件类别标签。给出了每种形状的分类标签。通常,通过使用已知的输入3D对象类别,将可能的零件标签缩小到特定于给定对象类别的部分标签。而且,还将每个点的法线方向计算为输入特征,以更好地描述基础形状。

使用点“交并比(IoU)”来评估的网络,与

PointNet<sup>[18]</sup>、PointNet++<sup>[19]</sup>和其他一些点云分割算法相同。结果显示在表1中。从表格数据可以看出,网络架构获得的实例平均交并比为86.0%;最差的结果出现在对“火箭”的分割结果上,为60.2%;最好的结果出现在“笔记本电脑”,为95.8%。若以单独某个分类的零件分割结果的IoU作为评判标准,表1中列出的分隔结果中,16个分类中有5个分类取得了最好的效果。在其他未取得最好效果的分类中,其分割结果与最佳结果的差距小于3个百分点。网络架构在ShapeNet零件分割任务中取得了较好的结果。

将ShapeNet数据集上零件分割的部分结果可视化,如图4所示。从可视化结果可以看出,模型的分割结果与真实标注接近,对于零件分割的任务表现得较好。

#### 3.2 三维模型场景分割实验

斯坦福大学大型3D室内空间S3DIS数据集<sup>[31]</sup>

表 1 ShapeNet 零件数据集<sup>[30]</sup>上的零件分割结果

Tab. 1 Result of part segmentation in ShapeNet dataset<sup>[30]</sup>

| 类别    | 形状数量  | PointNet <sup>[18]</sup> | PointNet++ <sup>[19]</sup> | Kd-Net <sup>[14]</sup> | SynSpecCNN <sup>[22]</sup> | DGCNN <sup>[23]</sup> | 本文方法 |
|-------|-------|--------------------------|----------------------------|------------------------|----------------------------|-----------------------|------|
| 飞机    | 2 690 | 83.4                     | 82.4                       | 80.2                   | 81.6                       | 84.2                  | 84.8 |
| 袋子    | 76    | 78.7                     | 79.0                       | 74.6                   | 81.7                       | 83.7                  | 84.0 |
| 帽子    | 55    | 82.5                     | 87.7                       | 74.3                   | 81.9                       | 84.4                  | 85.5 |
| 汽车    | 898   | 74.9                     | 77.3                       | 70.3                   | 75.2                       | 77.1                  | 78.2 |
| 椅子    | 3 758 | 89.6                     | 90.8                       | 88.6                   | 90.2                       | 90.9                  | 90.2 |
| 耳机    | 69    | 73.0                     | 71.8                       | 73.5                   | 74.9                       | 78.5                  | 79.2 |
| 吉他    | 787   | 91.5                     | 91.0                       | 90.2                   | 93.0                       | 91.5                  | 91.3 |
| 刀     | 392   | 85.9                     | 85.9                       | 87.2                   | 86.1                       | 87.3                  | 87.1 |
| 灯     | 1 547 | 80.8                     | 83.7                       | 81.0                   | 84.7                       | 82.9                  | 84.2 |
| 笔记本电脑 | 451   | 95.3                     | 95.3                       | 94.9                   | 95.6                       | 96.0                  | 95.8 |
| 摩托车   | 202   | 65.2                     | 71.6                       | 57.4                   | 66.7                       | 67.8                  | 68.9 |
| 杯子    | 184   | 93.0                     | 94.1                       | 86.7                   | 92.7                       | 93.3                  | 93.8 |
| 手枪    | 283   | 81.2                     | 81.3                       | 78.1                   | 81.6                       | 82.6                  | 82.4 |
| 火箭    | 66    | 57.9                     | 58.7                       | 51.8                   | 60.6                       | 59.7                  | 60.2 |
| 滑板    | 152   | 72.8                     | 76.4                       | 69.9                   | 82.9                       | 75.5                  | 76.2 |
| 桌子    | 5 271 | 80.6                     | 82.6                       | 80.3                   | 82.1                       | 82.0                  | 83.5 |
| 均值    |       | 83.7                     | 85.1                       | 82.3                   | 84.7                       | 85.1                  | 86.0 |

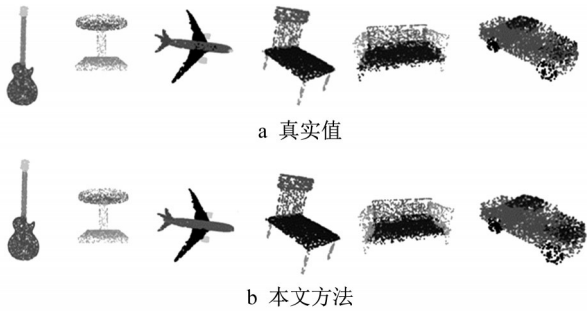


图 4 ShapeNet 零件分割结果

Fig. 4 Result of part segmentation in ShapeNet dataset

从 3 个不同的建筑物(包括 271 个房间)中提取了 6 个文件夹的 RGB-D 点云数据。每个点都有来自 13 个类别的标签注释。按照文献[17]的 k 折策略将其

分为训练和测试集。表 2 中显示了总体精度和平均交并比。可以看到,本文方法获得了 64.6% 的平均交并比、75.7% 的平均准确率。实验效果比 PointAGCN 效果更好。使用本文方法得出的预测值与真实值对比如图 5 所示。

表 2 S3DIS 数据集<sup>[31]</sup>上的三维语义分割结果

Tab. 2 Result of 3D-semantic segmentation in S3DIS dataset<sup>[31]</sup>

| 方法                       | 平均交并比 | 平均精确度 |
|--------------------------|-------|-------|
| PointNet <sup>[17]</sup> | 41.09 | 48.98 |
| MS+CU(2) <sup>[19]</sup> | 47.80 | 59.70 |
| G+RCU <sup>[19]</sup>    | 49.70 | 66.40 |
| SEGCLOUD <sup>[16]</sup> | 48.92 | 57.35 |
| SPG <sup>[21]</sup>      | 62.10 | 73.00 |
| 本文方法                     | 64.60 | 75.70 |

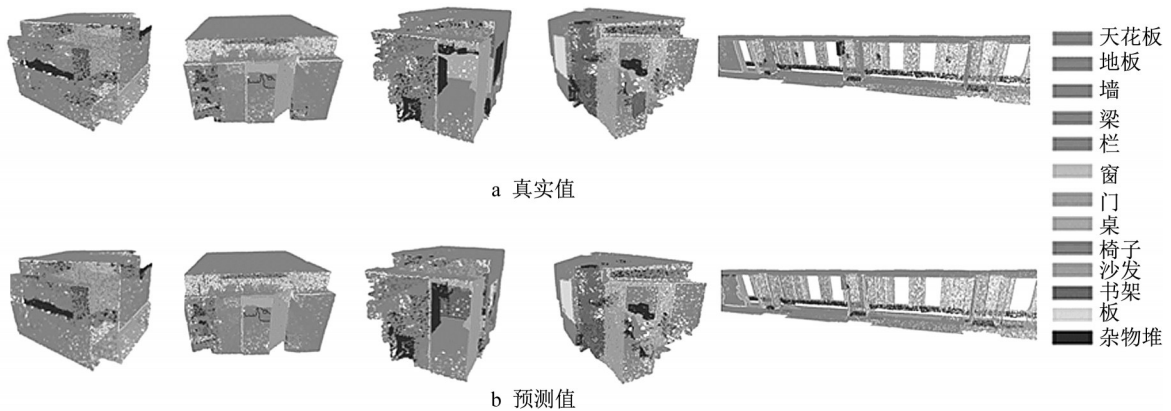


图 5 S3DIS 数据集<sup>[31]</sup>上语义分割的可视化结果

Fig. 5 Visual result of semantic segmentation in S3DIS dataset<sup>[31]</sup>

### 3.3 对比实验

调查了不同的结构的性能,例如残差连接、空洞卷积、密度自适应;研究了不同参数的影响,例如k-NN邻居数(4、8、16),滤波器数(16、32、64)和网络层

数(7、14、28、56);进行了20个实验,并将其结果显示在表3中。其中,模型1是基准模型,没有添加残差图连接和空洞卷积,k-NN邻居数为16,滤波器值为64,网络层数为28,作为对比实验的基线。

表3 S3DIS 斯坦福数据集<sup>[31]</sup>上不同变量的对比实验  
Tab. 3 Comparison of different variables in S3DIS dataset<sup>[31]</sup>

| 模型序号 | 对比变量  | 残差连接 | 空洞卷积 | 密度自适应 | K值 | 滤波器值 | 网络层数 | 平均交并比 | 平均交并比增量 |
|------|-------|------|------|-------|----|------|------|-------|---------|
| 1    | 基准    |      |      |       | 16 | 64   | 28   | 42.73 | 0       |
| 2    | 残差连接  | ✓    |      |       | 16 | 64   | 28   | 48.22 | 5.49    |
| 3    | 空洞卷积  |      | ✓    |       | 16 | 64   | 28   | 49.10 | 6.37    |
| 4    |       | ✓    | ✓    |       | 16 | 64   | 28   | 55.49 | 12.76   |
| 5    | 密度自适应 |      |      | ✓     | 16 | 64   | 28   | 47.42 | 4.69    |
| 6    |       | ✓    |      | ✓     | 16 | 64   | 28   | 56.38 | 3.65    |
| 7    |       |      | ✓    | ✓     | 16 | 64   | 28   | 57.13 | 14.40   |
| 8    |       | ✓    | ✓    | ✓     | 16 | 64   | 28   | 64.60 | 21.87   |
| 9    | K值    | ✓    | ✓    | ✓     | 4  | 64   | 28   | 60.81 | 18.08   |
| 10   |       | ✓    | ✓    | ✓     | 8  | 64   | 28   | 62.45 | 19.72   |
| 11   |       | ✓    | ✓    | ✓     | 16 | 64   | 7    | 59.94 | 17.21   |
| 12   | 网络层数  | ✓    | ✓    | ✓     | 16 | 64   | 14   | 62.19 | 19.46   |
| 13   |       | ✓    | ✓    | ✓     | 16 | 64   | 56   | 65.23 | 22.50   |
| 14   |       | ✓    | ✓    | ✓     | 16 | 16   | 28   | 55.67 | 12.94   |
| 15   | 滤波器值  | ✓    | ✓    | ✓     | 16 | 32   | 28   | 60.34 | 17.61   |

(1)残差连接:从表3中的实验1和2可以看到,在添加残差连接之后,平均IoU提升了5.49%。实验表明,残差图连接在训练更深的网络方面起着至关重要的作用,因为它们往往会导致更稳定的梯度。

(2)空洞卷积:表3中的实验1和3结果表明,加入空洞卷积之后,平均IoU中占比提高了6.37%,这主要是由网络感受野的扩展引起的。当同时添加残差连接和空洞卷积时,性能会进一步得到提高。

(3)K值:表3(K值)中的结果表明,更多的邻居通常对网络有帮助。随着邻居数量减少到原来的1/2或1/4,性能分别下降2.15%和3.79%。

(4)网络层数:按照表3的网络层数一栏所示,将网络层数设为7、14和56进行对比实验。实验结果表明,增加网络层数可以提高网络性能,网络层数越大意味着网络深度增大。当网络层数减小时,实验结果会受到影响。

(5)滤波器数:表3中的结果表明,滤波器数量增多会获得与增加层数类似的性能提升。通常,较高的网络容量可以让网络学习到更深层次的特征。

## 4 结语

为了成功训练更深层的GCN网络,减轻梯度消失问题,借用了成功训练深层CNN的相关概念,引入残差连接、空洞卷积等结构,训练更深层的点云分割网络。同时,由于3D点云是一种不规则且无序的

数据类型,传统的卷积神经网络难以处理点云数据。提出一种密度自适应的方法,可以高效地对非均匀采样的3D点云数据进行卷积操作,该方法在多个数据集上实现了优秀的性能。

### 作者贡献说明:

卫刚:论文撰写,深度神经网络设计。

赵安铭:论文撰写,深度神经网络设计与程序设计。

王志成:深度神经网络设计与数据分析。

### 参考文献:

- [1] LI Q, HAN Z, WU X M. Deeper insights into graph convolutional networks for semi-supervised learning [C]// Proceedings of the AAAI Conference on Artificial Intelligence. New Orleans: [s. n.], 2018:1-3.
- [2] WU Z, PAN S, CHEN F, *et al.* A comprehensive survey on graph neural networks [J]. IEEE transactions on neural networks and learning systems, 2020, 32(1):4.
- [3] ZHOU J, CUI G, HU S, *et al.* Graph neural networks: A review of methods and applications[J]. AI Open, 2020, 1: 57.
- [4] HE K, ZHANG X, REN S, *et al.* Deep residual learning for image recognition [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: [s. n.], 2016: 770-778.
- [5] YU F, KOLTUN V. Multi-scale context aggregation by dilated convolutions[J]. arXiv preprint, 2015:1511.07122.
- [6] REN S, HE K, GIRSHICK R, *et al.* Faster R-CNN: towards real-time object detection with region proposal networks [J].

- IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016, 39(6): 1137.
- [7] CHEN L C, PAPANDREOU G, KOKKINOS I, *et al.* Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs [J]. IEEE Transactions on pattern Analysis and Machine Intelligence, 2017, 40(4): 834.
- [8] 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. Santiago: [s.n.], 2015: 945-953.
- [9] QI C R, SU H, NIEßNER M, *et al.* Volumetric and multi-view cnns for object classification on 3d data [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: [s.n.], 2016: 5648-5656.
- [10] BOULCH A, LE SAUX B, AUDEBERT N. Unstructured point cloud semantic labeling using deep segmentation networks [J]. 3DOR: Eurographics Workshop on 3D Object Retrieval, 2017, 2: 7.
- [11] GUERRY J, BOULCH A, LE SAUX B, *et al.* Snapnet-r: Consistent 3d multi-view semantic labeling for robotics [C]// Proceedings of the IEEE International Conference on Computer Vision Workshops. Venice: [s.n.], 2017: 669-678.
- [12] WU Z, SONG S, KHOSLA A, *et al.* 3d shapenets: A deep representation for volumetric shapes [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Boston: [s.n.], 2015: 1912-1920.
- [13] 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). [S.l.]: IEEE, 2015: 922-928.
- [14] KLOKOV R, LEMPITSKY V. Escape from cells: Deep kd-networks for the recognition of 3d point cloud models [C]// Proceedings of the IEEE International Conference on Computer Vision. Venice: [s.n.], 2017: 863-872.
- [15] RIEGLER G, OSMAN ULUSOY A, GEIGER A. Octnet: Learning deep 3d representations at high resolutions [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: [s.n.], 2017: 3577-3586.
- [16] TCHAPMI L, CHOY C, ARMENI I, *et al.* Segcloud: Semantic segmentation of 3d point clouds [C]// 2017 International Conference on 3D Vision (3DV). [S.l.]: IEEE, 2017: 537-547.
- [17] QI C R, SU H, MO K, *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. Honolulu: [s.n.], 2017: 652-660.
- [18] QI C R, YI L, SU H, *et al.* Pointnet++: Deep hierarchical feature learning on point sets in a metric space [J]. arXiv preprint, 2017: 1706.02413.
- [19] ENGELMANN F, KONTOGIANNI T, HERMANS A, *et al.* Exploring spatial context for 3D semantic segmentation of point clouds [C]// Proceedings of the IEEE International Conference on Computer Vision Workshops. Venice: [s.n.], 2017: 716-724.
- [20] LI Y, BU R, SUN M, *et al.* PointCNN: Convolution On  $\mathcal{X}$ -Transformed Points [J]. arXiv preprint, 2018: 1801.07791.
- [21] LANDRIEU L, SIMONOVSKY M. Large-scale point cloud semantic segmentation with superpoint graphs [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City: [s.n.], 2018: 4558-4567.
- [22] YI L, SU H, GUO X, *et al.* Syncspecnn: Synchronized spectral cnn for 3d shape segmentation [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: [s.n.], 2017: 2282-2290.
- [23] 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.
- [24] SU H, JAMPANI V, SUN D, *et al.* Splatnet: Sparse lattice networks for point cloud processing [C]// Proceedings of the IEEE conference on Computer Vision and Pattern Recognition. Salt Lake City: [s.n.], 2018: 2530-2539.
- [25] JIANG M, WU Y, ZHAO T, *et al.* Pointsift: A sift-like network module for 3d point cloud semantic segmentation [J]. arXiv preprint arXiv: 1807.00652, 2018.
- [26] GLOROT X, BORDES A, BENGIO Y. Deep sparse rectifier neural networks [C]// Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics. Ft. Lauderdale: [s.n.], 2011: 315-323.
- [27] IOFFE S, SZEGEDY C. Batch normalization: Accelerating deep network training by reducing internal covariate shift [C]// International Conference on Machine Learning. Lille: [s.n.], 2015: 448-456.
- [28] HERMOSILLA P, RITSCHER T, VÁZQUEZ P P, *et al.* Monte carlo convolution for learning on non-uniformly sampled point clouds [J]. ACM Transactions on Graphics, 2018, 37(6): 1.
- [29] TURLACH B A. Bandwidth selection in kernel density estimation: A review [C]// Université Catholique de Louvain, Belgium, C.O.R.E. Institut de Statistique, 1993.
- [30] YI L, SHAO L, SAVVA M, *et al.* Large-scale 3d shape reconstruction and segmentation from shapenet core55 [J]. arXiv preprint arXiv: 1710.06104, 2017.
- [31] ARMENI I, SENER O, ZAMIR A R, *et al.* 3d semantic parsing of large-scale indoor spaces [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: [s.n.], 2016: 1534-1543.