

基于点云数据的三维目标识别和模型分割方法

牛辰庚¹, 刘玉杰¹, 李宗民¹, 李 华^{2,3}

(1. 中国石油大学(华东)计算机与通信工程学院, 山东 青岛 266580;
2. 中国科学院计算技术研究所智能信息处理重点实验室, 北京 100190;
3. 中国科学院大学, 北京 100190)

摘 要: 三维模型的深度特征表示是三维目标识别和三维模型语义分割的关键和前提, 在机器人、自动驾驶、虚拟现实、遥感测绘等领域有着广泛的应用前景。然而传统的卷积神经网络需要以规则化的数据作为输入, 对于点云数据需要转换为视图或体素网格来处理, 过程复杂且损失了三维模型的几何结构信息。借助已有的可以直接处理点云数据的深度网络, 针对产生的特征缺少局部拓扑信息问题进行改进, 提出一种利用双对称函数和空间转换网络获得更鲁棒、鉴别力更强的特征。实验表明, 通过端到端的方式很好地解决缺少局部信息问题, 在三维目标识别、三维场景语义分割任务上取得了更好的实验效果, 并且相比于 PointNet++在相同精度的情况下训练时间减少了 20%。

关 键 词: 点云; 深度学习; 原始数据; 三维目标识别; 三维模型分割

中图分类号: TP 391

DOI: 10.11996/JGj.2095-302X.2019020274

文献标识码: A

文章编号: 2095-302X(2019)02-0274-08

3D Object Recognition and Model Segmentation Based on Point Cloud Data

NIU Chen-geng¹, LIU Yu-jie¹, LI Zong-min¹, LI Hua^{2,3}

(1. College of Computer and Communication Engineering, China University of Petroleum, Qingdao Shandong 266580, China;

2. Key Laboratory of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China;

3. University of Chinese Academy of Sciences, Beijing 100190, China)

Abstract: Deep representation of 3D model is the key and prerequisite for 3D object recognition and 3D model semantic segmentation, providing a wide range of applications ranging from robotics, automatic driving, virtual reality, to remote sensing and other fields. However, convolutional architectures require highly regular input data formats and most researchers typically transform point cloud data to regular 3D voxel grids or sets of images before feeding them to a deep net architecture. The process is complex and the 3D geometric structure information will be lost. In this paper, we make full use of the existing deep network which can deal with point cloud data directly, and propose a new algorithm that uses double symmetry function and space transformation network to obtain more robust and discriminating features. The local topology information is also incorporated into the final features. Experiments show that the proposed method solves the problem of lacking local information in an end-to-end way and achieves ideal results in the task of 3D object recognition and 3D scene semantic segmentation. Meanwhile, the method can save 20% training time compared to PointNet++

收稿日期: 2018-09-03; 定稿日期: 2018-10-10

基金项目: 中央高校基本科研业务费专项资金项目(18CX06049A); 国家自然科学基金项目(61379106, 61379082, 61227802); 山东省自然科学基金项目(ZR2015FM011, ZR2013FM036)

第一作者: 牛辰庚(1993-), 男, 河北衡水人, 硕士研究生。主要研究方向为三维目标识别。E-mail: niuchengeng@foxmail.com

通信作者: 刘玉杰(1971-), 男, 辽宁沈阳人, 副教授, 博士。主要研究方向为计算机图形图像处理。E-mail: 782716197@qq.com

with the same precision.

Keywords: point cloud; deep learning; raw data; 3D object recognition; 3D model segmentation

近年来, 随着三维成像技术的快速发展, 像微软 Kinect, 英特尔的 RealSense 和谷歌的 Tango 等低成本小型化三维传感器都可以很好的捕获场景的三维信息, 帮助智能设备更好的感知、理解世界的同时很大程度上也降低了人们以三维的方式获取真实世界信息的门槛。另一方面, 伴随着 GPU 计算能力的迭代更新和大型三维模型数据的出现, 深度学习的思想在三维模型分类、检索等任务范围逐渐占据了绝对主导地位。这就使高效、准确并且直接处理三维数据的技术成为广泛的需求, 并且成为自动驾驶、虚拟现实以及遥感测绘发展的关键。

然而, 通过便携式三维扫描设备获取的原始三维数据通常是点云的形式, 区别于传统的图像和体素结构, 属于不规则的三维形状数据结构。深度学习中传统卷积结构为了实现权值共享和核函数优化需要以规则化的数据结构作为输入, 所以之前对于点云数据的处理通常转换为多视图或者体素的形式再输入到深度网络中去。但该数据处理形式的转换往往会带来几何结构损失、分辨率下降等问题, 由此产生识别精度低, 模型错误分割的实验结果。

之前利用深度学习在点云上提取特征的工作有 PointNet^[1]和 PointNet++^[2]。PointNet 以记录空间坐标的原始点云数据直接作为网络的输入, 学习点云模型的空间编码后转换为全局特征描述子用于目标分类和模型分割任务。PointNet++为了学习到模型更多的局部结构信息, 首先通过最远点采样和球查询的方式提取包含模型局部结构的点集, 并利用 PointNet 学习带有局部特征的特征串为全局特征用于模型分割任务。

本文方法在 PointNet 直接处理原始点云模型的深度网络基础上, 以端到端的方式完成输入到高层特征表示的映射。且利用多层感知机网络单独地提取每个点的深度特征, 然后引入与二维图像上处理仿射变换不变性的空间转换网络(spatial transformer networks, STN)^[3]相似的结构学习模型的拓扑结构信息, 同时利用双对称函数对点集特征进行编码, 消除点序对全局特征的影响并且进一步产生更有鉴别力和稳健性更强的深度特征。相比于 PointNet, 本文通过构建端到端的深度网络模型学习带有模

型拓扑结构的全局信息, 以更小的时间代价达到了更高的目标识别精度, 网络结构更加简单并且易于训练。

本文直接处理点云数据的深度网络关键在于转换网络和对称函数的设计, 理论分析及实验证明本文方法产生的特征蕴涵更多的模型信息以及具有更好的稳健性, 其充分解释了网络对于存在缺失和扰动的点云模型具有一定鲁棒性的原因。从函数逼近的角度看, 由于避免了最远点采样等提取局部结构信息的模型预处理模块, 本文的网络可以对任意连续的集合进行函数逼近。

1 相关工作

基于点云模型的特征提取自上世纪 90 年代开始至今已有 20 余年的发展, 以 2012 年为分界线总体可以分为 2 个阶段: 手工设计特征阶段和基于深度学习的特征阶段。而基于深度学习的三维模型特征提取依据不同三维形状数据的表示方法又可以分为: 基于手工特征预处理的方法, 基于投影图的方法, 基于体素的方法和基于原始数据的方法。

手工设计特征阶段通常通过提取三维形状的空间分布或直方图统计等方法得到, 典型代表如 Spin Image、FPFH、HKS (heat kernel signature)、MshHOG、RoPS 等^[4]。这类模型驱动的方法在前一阶段中占据着主导地位, 但是依赖于研究者的领域知识, 并且获取的特征在不同属性数据集中的区分力、稳定性和不变性都不容易得到保证, 可拓展性差。

2012 年普林斯顿大学建立了大型三维 CAD 模型库项目 ModelNet^[5], 伴随着深度学习算法在图像领域取得了巨大的成功, 三维形状数据结合深度网络提取特征并应用于目标分类、模型分割、场景语义解析任务也取得了很好的结果, 三维领域中数据驱动方法开始发展起来并逐渐在各项三维领域的任务中取得重要地位。

由于初期三维模型库较小以及深度网络由二维到三维的复杂性, 最先发展起来的是基于手工特征预处理的方法。该方法首先在三维模型上提取手工特征, 然后将手工特征作为深度网络的输入从而提取高层特征。典型的工作有 BU 等^[6]首先在三维模型上提取热核和平均测地距离特征, 利用词包转

换为中级特征输入到深度置信网络中提取高层特征。XIE 等^[7]首先提取三维模型的热核特征构建出多尺度直方图,然后在每个尺度上训练一个自编码器并将多个尺度隐含层的输出串联得到三维模型描述子,并在多个数据集上测试了该方法用于形状检索的有效性。KUANG 等^[8]利用嵌入空间下局部特征和全局特征融合的方法得到三维模型描述子,用来解决非刚体三维模型检索任务。这类方法可以充分利用之前的领域内知识作为先验指导并且能够很好的展现手工特征和深度网络各自的优势。但是依赖于手工特征的选择和对模型参数的调整、优化,一定程度上削弱了深度网络高层特征的表达能力。

文献[9]首先尝试通过多视图表示三维模型,然后输入到深度网络提取高层特征,在三维模型分类、检索任务上取得了很好的表现。即给定视点和视距将三维模型投影为 12 或 20 幅视图输入到卷积神经网络提取每幅图像的特征,然后经过相同位置最大池化处理输入到第二个卷积神经网络提取模型特征。文献[10]充分考虑多视图之间的关系,在输入网络前按照视点重要程度对相应视图进行预排序。之后文献[11]又提出沿三维模型主轴投影为全景图并通过卷积神经网络提取深度特征的方法。该方法的优点为可以充分利用二维图像领域中成熟的深度网络架构及充足的图像数据完成深度网络的训练、调整。但是投影的方式损失了三维模型的几何结构,一定程度上损失了特征的鉴别力,并且往往需要三维模型沿 Z 轴对齐。

基于体素的方法通过将三维模型体素化,仿照二维图像上的卷积操作利用深度网络直接在体素上提取深度特征。WU 等^[5]将三维模型用 $32 \times 32 \times 32$ 的二值化体素表示,采用深度卷积置信网络学习三维数据和标签之间的联合概率分布。QI 等^[12]发现基于体素的深度网络存在过拟合问题,通过在网络中加入用局部三维形状信息预测类别标签的辅助任务很大程度上避免了过拟合问题。但是文献[13]指出,随着体素分辨率的提高,数据稀疏性和计算复杂度问题难以处理。由于传统深度网络只对模型边缘体素信息敏感,文献[14]提出用八叉树的方式组织体素模型然后提取深度特征的 O-CNN,在保证模型精度的前提下提高了数据利用率。

丢失几何结构信息和数据稀疏性问题限制了基于多视图和基于体素的深度网络的发展,那么最优的发展方向就是调整深度网络适应原始三维数据作为输入^[15]。文献[16]提出基于面片的卷积受限

波兹曼机(MCRBM),实现了三维形状的无监督特征学习。文献[17]提出使用 KD 树组织点云数据,规则化深度网络输入结构。文献[18]将激光点云数据划分为若干体素块,然后利用体素特征编码模块(voxel feature encoding, VFE)进行局部特征提取,并通过三维卷积实现高层特征的抽象。QI 等^[1-2]提出基于点云数据的深度网络 PointNet,通过多层感知机提取每个点的深度特征并利用对称函数转换为对点序不变的全局特征向量,在三维模型分类、语义分割任务上取得了很好的效果。之后借鉴于文献[19]和文献[20]中的方法,为了学习点云模型局部拓扑结构特征提出 PointNet++,首先通过最远点采样和球查询聚集的方法对点云模型进行处理,然后通过 PointNet 映射成带有局部信息的高层特征。PointNet++进一步拓展了 PointNet 获得了更精细的模型局部特征,在模型分割和场景语义分割任务上取得了更高的精度。文献[21]通过自组织映射(self-organizing map, SOM)聚类算法得到点云模型关键点并建立与周围模型点的联系,并使用 PointNet 模块得到带有空间信息的模型描述子。文献[22]利用图像目标检测方法将三维模型检测范围缩小到视锥中,然后利用 PointNet 提取深度特征用于目标分类和包围盒估计。

2 本文工作

本文通过设计直接处理点云数据的深度网络,提取三维模型深度特征应用于三维模型识别和三维模型分割任务。提取模型全局特征的网络结构如图 1 所示(卷积核大小除第 1 层为 1×3 ,其余均为 1×1 ,且步长均为 1,同层卷积权值共享。对于目标识别任务,输入点云序列只记录空间坐标信息,大小为 $N \times 3$;对于三维模型语义分割任务,输入点云序列记录空间坐标、颜色、法向信息,大小为 $N \times 9$)。网络以点云数据为输入,经过 5 个卷积层,差异性对称函数和姿态变换子网络处理,将池化特征和姿态特征串联得到最终的全局特征。对于 k 类三维目标识别任务,深度网络以记录空间信息 $\{x,y,z\}$ 的点云数据直接作为输入,对每个模型上的点做单独处理,输出对应所属类别概率的 k 维向量。对于有 m 个语义标签的三维模型语义分割任务,深度网络以从每个场景目标模型中采样得到的 n 个点的点云模型作为输入,输出对应每个点语义标签的 $N \times m$ 维特征矩阵。本文所使用的深度网络可以分为 3 个部

分: 深度卷积神经网络单独提取每个点的深度信息, 差异性双对称函数提取模型不同显著性特征,

空间转换网络预测出姿态变换矩阵融合为带有局部信息的全局特征。

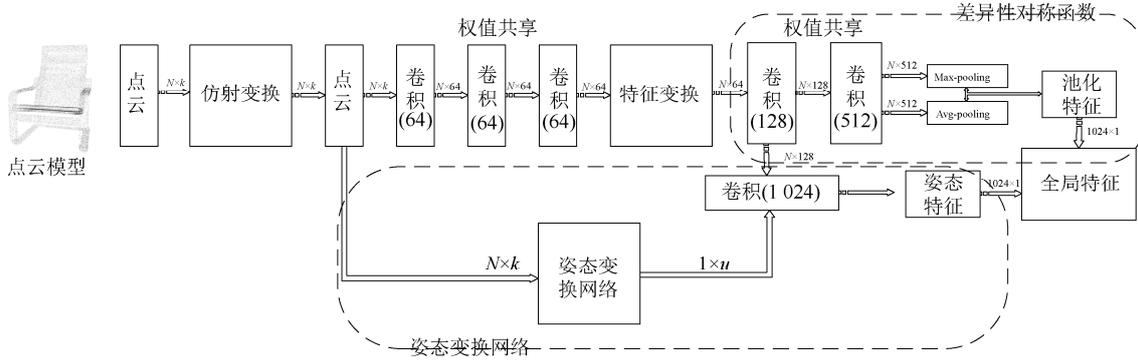


图 1 本文网络结构图

2.1 深度卷积网络

由于集合中的点以记录空间坐标信息、颜色和法向信息的点集形式存在, 所以是一种不规则形式的三维数据, 不能直接输入到传统卷积深度神经网络。此外, 在点云上提取模型特征时还需要考虑到点序对最终特征的影响, 避免模型在仿射变换之后产生错误识别, 或者模型上的点对应的语义标签发生改变的情况。这里通过调整深度卷积网络适应点云数据的输入形式, 先对模型上的每个点进行处理, 然后在得到的特征层面进行点序的处理。

深度卷积神经网络结构的设置类似于传统的多层感知器网络, 本文通过设置卷积核大小为 1×1 来实现对表示模型的点集的特征提取, 即对于点云模型上记录空间坐标等信息的 n 个点 $\{p_1, p_2, \dots, p_n \mid p_i = (x_i, y_i, z_i)\}$, 深度卷积网络单独将其每个点映射为中层特征 $\{K(p_1), K(p_2), \dots, K(p_n)\}$, 为接下来局部拓扑信息处理和全局特征提取做准备。实验表明, 相比于模型表示形式的转换和先对点集进行排序预处理的方式, 本文方法可以充分发挥点云数据本身的优势, 同时避免了排序预处理情况下需要考虑 $n!$ 种不同的组合情况。

2.2 差异性对称函数

在经过深度卷积神经网络的处理之后, 可以得到模型上每个点对应的中层特征向量 $\{K(p_1), K(p_2), \dots, K(p_n)\}$, 通过训练支持向量机 (support vector machine, SVM) 分类器或者多层感知器网络可以实现对模型的分类。但是对于点云数据上的语

义分割任务, 需要模型局部信息和全局信息的融合, 也就是需要增加模型局部拓扑信息。在得到的中层点集特征上进行卷积操作可用获得局部拓扑信息, 同时容易实现不同尺度的特征融合。但是由于点集的无序性导致需要考虑大量的排列组合情况, 多种同类别卷积特征的融合会损失模型识别精度。

对称函数具有输出对输入顺序不变的性质, 因此可以用来解决点云问题上特征对点序不变性的问题。并且不同的对称函数代表着不同的输入到输出映射关系, 因此可以通过差异性对称函数的结合获得模型不同显著性的特征, 增强模型语义分割任务特征的表达能力。本文结合 3 种不同的对称函数进行对比实验, 最终选择最大池化和平均池化特征相结合作为模型最终的全局特征。从函数逼近的角度看, 对于从深度卷积网络中得到的中层特征 $\{K(p_1), K(p_2), \dots, K(p_n)\}$, 假设 ϕ, φ 为 2 种差异性对称函数, 那么可以得到

$$f(\{p_1, p_2, \dots, p_n\}) \approx \text{reduce_mean}\{\phi(\{K(p_1), K(p_2), \dots, K(p_n)\}) + \dots \varphi(\{K(p_1), K(p_2), \dots, K(p_n)\})\} \quad (1)$$

其中, $\{f_1, f_2, \dots, f_n\}$ 为点集映射得到的高层全局特征。用于分类任务的网络结构如图 2 所示, 通过 3 个全连接层将得到的模型深度特征转换为 k 维概率矩阵。其结果表明融入了更多不同显著性信息的全局特征, 在模型分类任务精度上相较于 PointNet 有一定的提高。

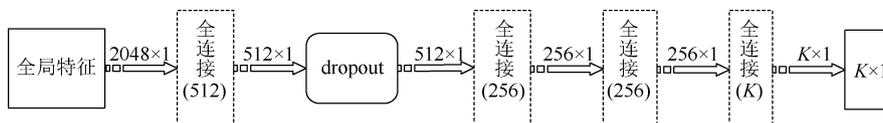


图 2 全局特征用于目标识别任务网络结构图

2.3 姿态对齐网络

PointNet 中使用模型全局特征和网络中间层的点特征进行串联用于进行后续的分割任务,但是由于特征不够精细且缺少局部上下文信息,容易产生失真的情况,并且在细粒度模式识别和复杂场景的识别问题上效果不佳。后续的工作增加模型输入到深度网络前的预处理步骤来解决缺少局部上下文信息的问题。但是重复进行最远点采样、采样点聚集和调用 PointNet 网络提取特征,一定程度上增大了问题复杂度,同时在不同尺度、不同密度下非端到端地使用 PointNet 网络提取高层特征也增加了时间开销。

JADERBERG 等^[3]于 2015 年提出了空间转换网络,通过采样和插值操作将图像对齐到标准空间中。本文采取类似的策略构建出姿态对齐网络,结构如图 3 所示(其中除第一层卷积核大小为 1×9 ,其余卷基层卷积核大小均为 1×1 ,步长为 1。同层

卷积层、同层全连接层权值共享),3 个卷积层和 2 个全连接层用来从点云数据中预测出变换矩阵,将点集映射到对姿态不变的空间同时获得点之间局部拓扑联系。其中输入点集大小为 $B \times N \times k$, B 为输入模型个数, N 为每个模型采样点个数, k 为输入特征维度,在模型识别任务中取 $k=3$,语义分割任务中取 $k=9$ 。用来处理模型语义分割任务的网络结构如图 4 所示。由于增加了新的局部网络模块,且姿态变换矩阵较于网络输入层的仿射变换矩阵维度较高($3 \times 3 \sim 32 \times 32 / 64 \times 64$),不易进行优化调整,所以在网络损失定义中增加正则项,使得对齐矩阵趋向于正交阵,即

$$L = \|I - PP^T\|_2^2 \quad (2)$$

其中, P 为姿态对齐网络输出对齐矩阵。正交阵不会损失输入信息,并且损失中增加正则项提高优化速度的同时也带来了一定程度上精度的提升。

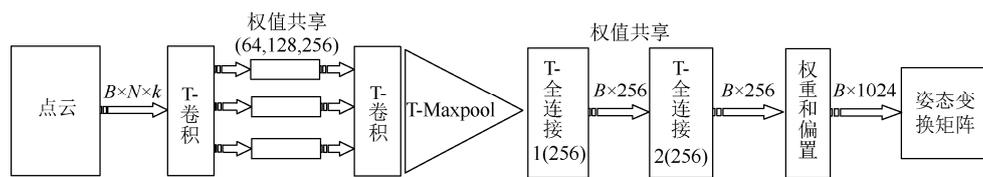


图 3 姿态变换网络结构图

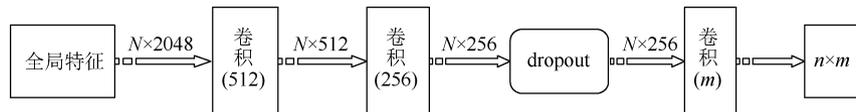


图 4 全局特征用于模型语义分割任务网络结构图

3 实验

实验部分分为 4 个模块,①融入双对称函数的深度网络在三维目标识别任务上测试;②加入姿态对齐网络和双对称函数模块的深度网络在三维模型语义分割任务上进行的测试;③网络鲁棒性测试以及相关对比实验;④对实验结果的可视化和相关分析。本文实验所用数据集通过在三维面片模型表面采样取点的方式得到,将三维模型表示为记录空间坐标信息和表面颜色、法向信息的一系列三维点集 $\{P_i | i=1,2,\dots,n\}$ 的集合。本文实验环境配置为英特尔 Xeon E5-2675 处理器,128 G 运行内存,Titan Xp 12 G 显存显卡。

3.1 三维目标识别任务实验

对于三维目标识别任务,本文方法充分利用端

到端的深度网络学习到具有不同显著性的模型特征,并在 ModelNet40 模型分类数据集上进行测试。ModelNet40 模型库包含 40 类 12 311 个 CAD 模型,其中训练集有 9 843 个模型,测试集有 2 468 个模型。2017 年之前的大部分工作是转换模型表达方式,以视图或者体素化的三维模型作为处理对象,本文是基于点云数据进行实验。

本文对于所有模型在表面按照面积的不同均匀地采集 1 024 个点,每个点记录空间坐标信息,且为了便于训练将所有点的坐标标准化到单元球中。在训练阶段,为了增强网络对模型仿射变换的特征不变性以及增加训练数据,对训练集模型进行随机角度的沿 Z 轴旋转以及添加均值为零,方差 0.03 的高斯噪声。实验中设置 dropout 参数为 0.7,实验结果对比见表 1。

表 1 ModelNet40 数据集目标识别任务实验对比

方法	输入类型	平均类精度	精度	发表时间(年)
LFD	视图(10)	75.5	-	2015 CVPR
MVCNN	视图(80)	90.1	-	2015 ICCV
3DShapeNets	体素	77.3	84.7	2015 CVPR
Subvolume	体素	86.0	89.2	2016 CVPR
MLP	点云	72.6	77.4	-
PointNet	点云	86.2	88.1	2017 CVPR
PointNet++	点云	-	90.7	2017 NIPS
本文方法	点云	87.4	89.2	-

本文方法同之前基于体素的基准方法有了 4.5% 的精度提升, 取得了最佳的结果。并且由于本文采用端到端的方式对模型进行处理, 网络主要结构为处理点云空间坐标信息的卷积, 双对称函数映射模块和全连接, 可以通过 GPU 进行高效的并行计算。相比于在点云数据上提取手工特征(点云密度, 测地线距离等)再利用多层感知器提取深度特征的方式(表 1 MLP 方法)以及通过 PointNet 提取模型全局特征的方法, 本文的方法取得了最佳的效果。

3.2 三维模型语义分割任务实验

相比于模型分类任务, 三维模型语义分割需要输入更为精细的点特征, 因此是一项更具有挑战性的细粒度任务。本文方法中结合姿态估计网络和多层感知器网络对原始点云数据进行处理, 同目标识别任务采用相似的方法在每个三维模型表面均匀地采集 4 096 个点, 并且将每个点对应的 RGB 值和法向信息同空间坐标统一作为本文深度网络的输入。

本文在斯坦福大学三维语义分割标准数据集上进行实验。该数据集包含了 6 个区域 271 个房间的 Matterport 扫描数据, 其中所有的点标注为桌子、地板、墙壁等 13 个类别。在网络训练阶段, 将所有点按照房间编号分开, 并且将每个房间划分为棱长 1 m 的小区域。语义分割网络将整个区域的点云数据作为输入, 输出每个区域中点的类别信息。

将本文语义分割结果与其余 3 种方法分割结果通过平均交并比和整体精度的评价指标进行比较, 见表 2。其中 MLP 方法为首先在点云数据上提取手工特征, 然后通过多层感知器网络获得语义分割特征。本文方法相比于 MLP 方法在平均交并比和整体分类精度指标上产生了巨大的提升。并且相比于 PointNet, 由于更好的融入了局部拓扑信息, 精度提高了 6.64%。同 PointNet++ 相比, 由于本文采用

端到端方式的处理, 在训练时间上缩短了 20%。

表 2 语义分割任务结果比较

方法	mean IoU	overall accuracy
MLP	20.12	53.19
PointNet ^[1]	47.71	78.62
PointNet++ ^[2]	-	83.34
本文方法	50.11	82.26

3.3 对比实验以及鲁棒性测试

三维目标识别任务中, 差异性对称函数的组合会影响最终全局特征的识别精度。为了达到深度网络的最佳的性能, 本文结合 3 种对称函数进行对比实验, 实验结果见表 3。

表 3 差异性对称函数组合比较

对称函数	识别精度
Attention sum	82.8
Avg pooling	83.0
Max pooling	87.1
Attention sum + Max pooling	87.4
Attention sum + Avg pooling	83.7
Max pooling + Avg pooling	89.2

为了验证本文深度网络对于模型采样点个数的鲁棒性, 随机丢弃测试集 50%, 75%, 87.5% 的采样点, 最终在 ModelNet40 上测试结果如图 5 所示。即在只保留 256 个采样点的条件下本文深度网络依然可以达到 85.3% 的识别精度。

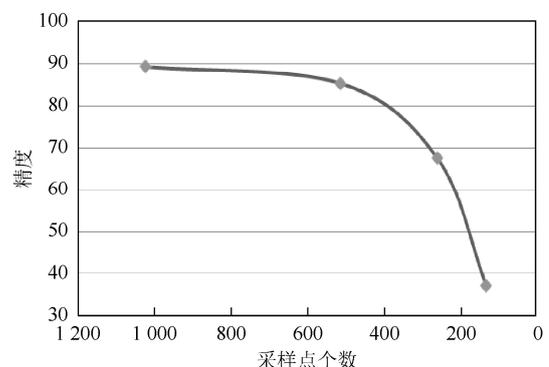
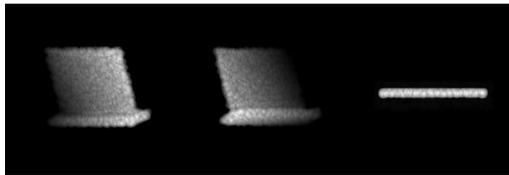


图 5 深度网络对采样点个数鲁棒性测试

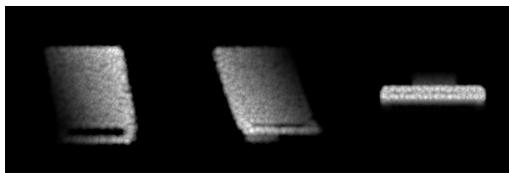
3.4 可视化结果以及实验分析

为了定性分析实验结果以及通过实验效果分析本文方法, 本文给出几种典型的误分类模型的可视化结果, 以及部分空间语义分割结果。图 6 模型

为 2 种镜子, 网络分类结果为书架。图 7 中 2 种错误情况为将沙发分类为床, XBOX 模型分类为书架。由此可知, 点云的稀疏性导致网络单纯的依靠空间坐标信息不能很好的区分出几何相似的模型。通过增加模型采样点的法向信息以及 RGB 信息可以一定程度上解决此问题。

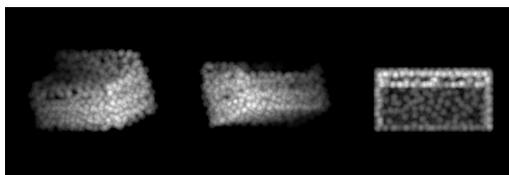


(a) 镜子模型 1(误分类为书架)

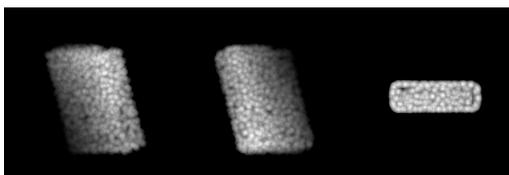


(b) 镜子模型 2(误分类为书架)

图 6 误分类情况模型可视化(镜子模型)



(a) 沙发模型(误分类为床)



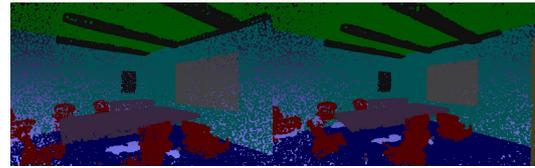
(b) XBOX 模型(误分类为书架)

图 7 误分类情况模型可视化

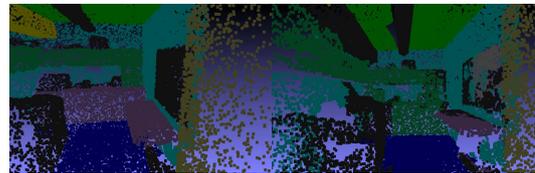
图 8 为 4 组空间语义分割结果可视化效果图, 不同颜色代表不同类别信息, 左栏为人工标注结果, 右栏为网络预测结果。

4 结束语

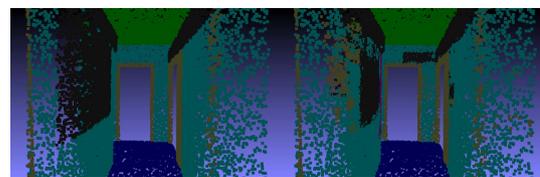
本文借助已有的可以直接处理点云数据的深度网络进行改进, 针对产生的特征缺少局部拓扑信息问题, 提出一种利用差异性双对称函数和空间转换网络来获得更鲁棒、鉴别力更强的特征。在 ModelNet40 数据集上分类任务以及在 S3DIS 数据



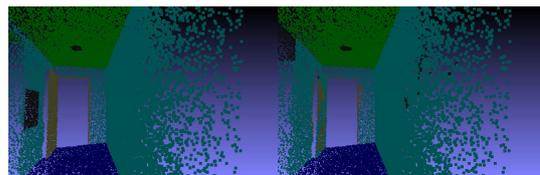
(a) 会议室



(b) 复印室



(c) 走廊 1



(d) 走廊 2

图 8 语义分割模型可视化

集语义分割任务上实验表明, 本文设计的网络和对应的特征有更好的表现。下一步工作重点是在点云模型全局特征中融入更多局部拓扑信息, 进一步提升语义分割精度以及提高模型识别精度。

参考文献

- [1] QI C R, SU H, MO K, et al. Pointnet: Deep learning on point sets for 3D classification and segmentation [C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). New York: IEEE Press, 2017: 77-85.
- [2] QI C R, YI L, SU H, et al. Pointnet++: Deep hierarchical feature learning on point sets in a metric space [C]//The 24th Annual Conference on Neural Information Processing Systems. Cambridge: MIT Press, 2017: 5105-5114.
- [3] JADERBERG M, SIMONYAN K, ZISSERMAN A. Spatial transformer networks [C]//The 22th Annual Conference on Neural Information Processing Systems. Cambridge: MIT Press, 2015: 2017-2025.
- [4] GUO Y, BENAMOUN M, SOHEL F, et al. 3D object recognition in cluttered scenes with local surface

- features: a survey [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2014, 36(11): 2270-2287.
- [5] WU Z, SONG S, KHOSLA A, et al. 3D shapenets: A deep representation for volumetric shapes [C]//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). New York: IEEE Press, 2015: 1912-1920.
- [6] BU S H, LIU Z, HAN J, et al. Learning high-level feature by deep belief networks for 3-D model retrieval and recognition [J]. *IEEE Transactions on Multimedia*, 2014, 16(8): 2154-2167.
- [7] XIE J, DAI G, ZHU F, et al. Deepshape: Deep-learned shape descriptor for 3D shape retrieval [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(7): 1335-1345.
- [8] KUANG Z, LI Z, JIANG X, et al. Retrieval of non-rigid 3D shapes from multiple aspects [J]. *Computer-Aided Design*, 2015, 58: 13-23.
- [9] SU H, MAJI S, KALOGERAKIS E, et al. Multi-view convolutional neural networks for 3D shape recognition [C]//2015 IEEE International Conference on Computer Vision. New York: IEEE Press, 2015: 945-953.
- [10] LENG B, LIU Y, YU K, et al. 3D object understanding with 3D convolutional neural networks [J]. *Information Sciences*, 2016, 366: 188-201.
- [11] SHI B G, BAI S, ZHOU Z, et al. DeepPano: Deep panoramic representation for 3D shape recognition [J]. *IEEE Signal Processing Letters*, 2015, 22(12): 2339-2343.
- [12] QI C R, SU H, NIEßNER M, et al. Volumetric and multi-view cnns for object classification on 3D data [C]//2016 IEEE conference on Computer Vision and Pattern Recognition. New York: IEEE Press, 2016: 5648-5656.
- [13] LI Y, PIRK S, SU H, et al. FPNN: Field probing neural networks for 3D data [C]//The 23th Annual Conference on Neural Information Processing Systems. Cambridge: MIT Press, 2016: 307-315.
- [14] WANG P S, LIU Y, GUO Y X, et al. O-cnn: Octree-based convolutional neural networks for 3D shape analysis [J]. *ACM Transactions on Graphics (TOG)*, 2017, 36(4): 72.
- [15] VINYALS O. Order matters: Sequence to sequence for sets. (2015-05-05). [2019-03-21]. <https://arxiv.org/abs/1506.02025>.
- [16] HAN Z, LIU Z, HAN J, et al. Mesh convolutional restricted Boltzmann machines for unsupervised learning of features with structure preservation on 3D meshes [J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2017, 28(10): 2268-2281.
- [17] KLOKOV R, LEMPITSKY V. Escape from cells: Deep kd-networks for the recognition of 3D point cloud models [C]//2017 IEEE International Conference on Computer Vision (ICCV). New York: IEEE Press, 2017: 863-872.
- [18] 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). New York: IEEE Press, 2015: 922-928.
- [19] KAMOUSHI P, LAZARD S, MAHESHWARI A, et al. Analysis of farthest point sampling for approximating geodesics in a graph [J]. *Computational Geometry*, 2016, 57: 1-7.
- [20] RODOLÀ E, ALBARELLI A, CREMERS D, et al. A simple and effective relevance-based point sampling for 3D shapes [J]. *Pattern Recognition Letters*, 2015, 59: 41-47.
- [21] LI J, CHEN B M, HEE LEE G. So-net: Self-organizing network for point cloud analysis [C]//2018 IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE Press, 2018: 9397-9406.
- [22] QI C R, LIU W, WU C, et al. Frustum pointnets for 3D object detection from rgb-d data [C]//2018 IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE Press, 2018: 918-927.