



基于超点图网络的三维点云室内场景分割模型

霍占强, 王勇杰, 雒 芬, 乔应旭

(河南理工大学 计算机科学与技术学院, 河南 焦作 454000)

摘 要: 针对点云数据集样本不均衡及PointNet网络无法充分利用点云邻域信息的问题, 提出一种三维点云场景分割模型。根据几何信息将原始点云块同质分割为超点, 利用小型PointNet网络将点云原始特征映射到高维空间中, 并挖掘场景中深层语义信息。在此基础上, 构建自归一化属性门控单元优化点云上下文语义分割效果, 采用二维图像领域中的Focal Loss损失函数实现点云场景分割。实验结果表明, 该模型在S3DIS数据集上的平均交并比、总体精度、平均精度分别达到63.8%、86.4%、74.3%, 较SPG模型分别提升1.7、0.9、1.3个百分点。

关键词: 场景分割; 三维点云; 上下文信息; 同质分割; 深度学习

开放科学(资源服务)标志码(OSID):



中文引用格式: 霍占强, 王勇杰, 雒芬, 等. 基于超点图网络的三维点云室内场景分割模型[J]. 计算机工程, 2021, 47(12): 308-315.

英文引用格式: HUO Z Q, WANG Y J, LUO F, et al. Indoor scene segmentation model using three-dimensional point cloud based on super point graph network[J]. Computer Engineering, 2021, 47(12): 308-315.

Indoor Scene Segmentation Model Using Three-Dimensional Point Cloud Based on Super Point Graph Network

HUO Zhanqiang, WANG Yongjie, LUO Fen, QIAO Yingxu

(School of Computer Science and Technology, Henan Polytechnic University, Jiaozuo, Henan 454000, China)

[Abstract] The samples of the existing current point cloud datasets are unbalanced, and PointNet can not effectively utilize the neighborhood information of the point cloud. To address the problem, a three-dimensional scene segmentation method using point cloud is proposed based on super point graph network. Based on geometric information, the original point cloud is divided into several blocks homogeneously. Utilizing small-sized PointNet, the original features of point cloud are mapped to high-dimensional space to mine the deep fine-grained semantic information in the scene. By constructing a kind of self-normalization gate recurrent units, the contextual semantic segmentation performance of the point cloud is improved, and the Focal Loss function in the two-dimensional image field is used to segment the scenes of the point cloud. The experimental results on the S3DIS dataset show that the proposed model exhibits a MIOU of 63.8%, OA of 86.4% and mAcc of 74.3%, which are respectively 1.7%, 0.9% and 1.3% higher than the SPG model. The proposed model significantly improves the semantic segmentation performance of point cloud for three-dimensional scenes.

[Key words] scene segmentation; 3D point cloud; contextual information; homogeneous segmentation; deep learning

DOI: 10.19678/j.issn.1000-3428.0059789

0 概述

随着三维立体成像技术的快速发展及深度传感器的广泛应用, 研究人员能快速且准确地获取物体表面的三维坐标和深度信息, 从而获取大量的场景点云数据。这些数据能够帮助智能机器对所处

空间位置进行准确的判断, 进而理解场景并更好地感知周围环境。场景感知有利于智能机器产生和人类一样对周边环境的分辨功能, 提高机器的智能水平^[1]。由于点云数据对复杂场景及物体的三维空间结构表达具有独特的优势, 加上获取数据的高速性和便捷性, 被广泛应用在自动驾驶^[2]、自动室

基金项目: 国家自然科学基金(61872311); 河南省高校科技创新团队支持计划(19IRTSTHN012)。

作者简介: 霍占强(1979—), 男, 副教授、博士, 主研方向为计算机视觉、图像处理、深度学习; 王勇杰, 硕士; 雒芬、乔应旭, 讲师。

收稿日期: 2020-10-21 **修回日期:** 2020-12-09 **E-mail:** qiaoyingxu@hpu.edu.cn

内导航^[3]、无人机和空天一体化^[1]等领域。

近年来,随着数据量的指数级增长和计算力不断增强,深度学习的方法得到了飞速发展和广泛应用,尤其在2D图像分类、分割和检测领域^[4]。在物体识别、目标检测等领域上其识别率和人类相当甚至超过人类。针对三维点云数据的无序性、不规则性、稀疏性等特点,ZENG等^[5]提出一种融合2D和3D多特征的近邻传播聚类集成分割方法,虽然该方法比传统的点云分割算法能更准确地划分点云场景,但传统手工特征提取无法完全利用三维点云的全部信息,尤其是点云的深度信息,导致点云分割、分类的精度难以提升。越来越多的学者尝试将深度学习和点云分割相结合以进行特征提取,但传统卷积神经网络结构只能处理规则数据如图片、文本、音频等。若将传统卷积神经网络应用到点云数据,需先将其处理为多视图或体素化网格形式,然后利用深度学习网络对其进行训练。然而这些数据形式转换将造成空间特征信息的丢失,导致点云分类精度低和效果差等问题。

本文针对超点图网络模型无法充分捕获点云上下文的深层细粒度局部几何特征,以及目前主流点云分割方法对困难样本分割效果稍差的问题,提出基于超点图的三维点云分割方法。通过在SPG网络模型中构造一种特殊的门控循环单元(Gated Recurrent Unit, GRU)^[6],实现隐藏层状态变量和输出变量的自归一化过程。为改善三维场景数据集样本不平衡问题,将焦点损失函数Focal Loss^[7]引入到3D点云领域,并提高网络训练过程中困难样本的权重。此外,在斯坦福大规模室内点云数据集(Stanford large-scale 3D Indoor Spaces Dataset, S3DIS)^[8]上进行训练测试,并与现有方法的常规指标进行对比,以验证本文方法的有效性。

1 相关研究

在深度学习领域,处理点云分割的主要方法分为基于多视图的方法、基于体素的方法和基于点云表示的方法。

1) 基于多视图的方法。由于三维点云数据的无序性,传统的卷积网络无法直接提取其空间特征。2015年SU等^[9]受到2D图像的启发,提出了二维多视角下的3D目标识别(MVCNN),该项工作成功将传统卷积神经网络应用到点云这类不规则化的数据中。其主要思想是对3D目标进行多方位渲染,得到不同视角下的投影图,并通过经典的VGG(Visual Geometry Group)^[10]提取特征,将不同视角的视图特征融合,得到全局描述符来实现点云的分类、分割。该方法与传统手工设计特征的方法相比,虽在一定程度上提高了三维模型识别的准确率,但采用投影提取多视角下的视图特征将导致一定程度上的特征冗余和大量关键几何特征的丢失,从而影响点云分类和分割结果。针对该问题,BOULCH等^[11]提出了SnapNet方法,主要通过二维神经网络分别对深度图和RGB图处理以生成三维场景的图像。虽然该方

法改善了分割效果,但是物体边界的分割精度仍有待提高。由此可见,此类方法在解决点云空间信息丢失问题上仍然面临着巨大的挑战。

2) 基于体素的方法。体素化处理法是先先将点云进行体素填充,变为规则的网格结构,然后利用三维卷积网络进行特征提取以实现点云的语义分割,如:VoxNet^[12]、PointGrid^[13]等方法。基于体素的多数网络性能表现优于多视图方法,说明体素能够有效描述点云数据的空间结构,同时具有规则的数据结构拟合标准的卷积操作,上述因素使体素在深度学习领域得到一定范围的应用。但体素化操作使内存消耗严重,而且该类方法目前对体素化导致的量化伪影仍无具体改进措施,导致提取高分辨率特征存在一定的困难。KLOKOV等^[14]提出空间划分方法以解决模型捕获低分辨率特征精度不高的问题,但该方法缺乏提取点云局部特征的能力。由上可知,体素化虽然能够规范三维点云,并适用于3D卷积操作,但是体素化将不可避免地导致空间信息的离散化,且在实操过程中选择合适的体素网格半径具有一定难度。

3) 基于点云表示的方法。该方法是目前研究点云的主流方法,已广泛应用于计算机图形学、机器视觉领域和工业技术中^[15]。例如在自动驾驶技术中,点云场景分割算法可通过对车辆、道路、红绿灯和行人的分割帮助自动驾驶系统判断道路情况;在医学自动诊断技术中,点云分割算法可以辅助医生提取和分析医学影像中人工难以获取的信息,从而更加精准的找到病灶关键^[16];在虚拟现实和增强现实技术中,点云场景分割算法可以通过分割场景中的前景和背景帮助虚拟和增强多种现实效果^[17],以达到身临其境的效果。QI等^[18]提出一种PointNet网络模型,该模型可直接作用于原始点云数据,而不需要将点云转化为其他数据形式,从而最大程度保留了点云的空间信息。该模型通过多层感知器(Multi-Layer Perception, MLP)提取点云的特征信息,并使用对称池函数提取全局特征解决点云输入的无序性问题。但该模型仅关注单个独立点的特征,没有考虑局部邻域信息的重要性。针对这些问题,QI等^[19]在PointNet模型的基础上引入分层神经网络,提出一种新的网络模型PointNet++,该模型可以提取不同尺度下的局部特征,提升了分类分割性能,但该模型在提取点云上下文局部几何特征上效果不佳。为了更好地获取不规则的点云数据几何结构信息,LOIC等^[20]利用最小割理论,将大规模输入点云进行几何同构分割,利用有向图建模构建超点图(Super Point Graph, SPG)^[21],较好地改善了场景分割问题。

2 算法描述

2.1 深度超点图网络模型

为更有效地挖掘点云上下文的空间特征,本文以超点图为基础,构建深度超点图网络模型以进行点云语义

分割。超点分割类似于二维图像领域的超像素分割,两者都属于过度分割(Over Segmentation)^[22]。过度分割能够更好地感知样本相似颜色、空间和纹理分布等信息,从而更有利于网络模型确定分割的边界。

本文的网络结构如图1所示,把输入的大规模点云记作 X , X 中的每一个样本是区域中的一个室内场景,记为 X_i , $X = \{X_1, X_2, \dots, X_i | X_i \in R^d, i \in n\}$, d 代表超点网络的输入维度。首先,对室内场景进行几何同质分割,所获得的每一个独立块即为超点。本文超点网络模型的输入维度为 $s \times n_p \times d_p$, s 代表点云室内数据集中每个室内空间被分割的超点个数, n_p 代表每个超点包含的三维点个数, d_p 代表每个采样点的原始特征维度。其次,为了更高效地提取超点的特征,选取 PointNet 模型做为前期点云特征提取器,因为 PointNet 模型对局部点云有着很高的鲁棒性和效率。点云经过只带有单个 STN(Spatial Transform Network)的简化版 PointNet 模型^[15]作用后,生成 32 维的嵌入张量,其中 STN 的作用是实现输入点云

的旋转不变性。原始三维点经过 PointNet 模型中多层 MLP 和最大池化的作用后,使每个点的原始特征能够映射到高维特征空间。在网络结构图中,MLP 括号中数字代表每层结构的宽度,每层结构均应用了 Batch Normalization。最后,为了能有效地捕获场景中点云块之间的上下文关系,将生成的高维特征嵌入到循环神经网络(Recurrent Neural Network, RNN)^[23]中,并与边条件卷积(Edge-Conditioned Convolution, ECC)^[24]作用后,得到一组 13 维特征向量(13 代表场景中待分割目标的种类)。本文受自归一化神经网络(Self-Normalizing Neural Network, SNN)启发,构造了一种特殊的 RNN 神经单元,即自归一化门控循环单元(Self-Normalizing Gated Recurrent Unit, SN-GRU),相对于原始 SPG 网络结构中的 GRU 神经单元,该单元可以进一步增强网络模型对高级几何特征的捕获能力,从而实现点云的场景分割。

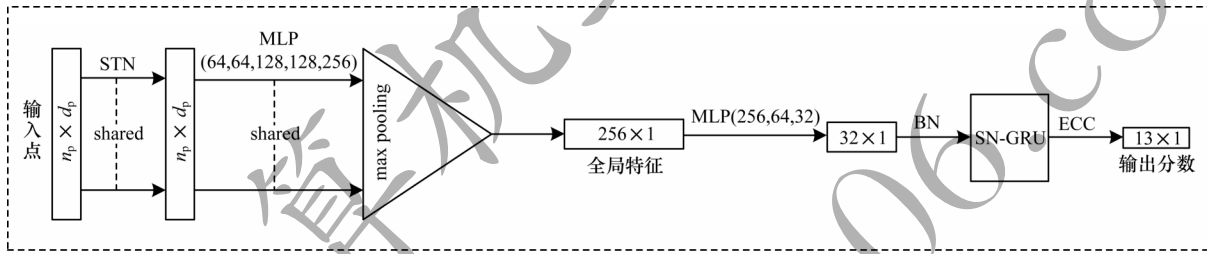


图1 网络结构

Fig.1 Network structure

2.2 几何同质分割

本文基于超点图网络提出新型网络模型,超点的构造由法国国家地理研究所计算机系的 LANDRIEU 等^[17]提出,利用最小割理论^[17]进行几何划分,实现对单个目标物体的同质划分,最小割理论如式(1)所示:

$$\arg \min \sum_{i \in G} (g_i - f_i)^2 + \mu \sum_{(i,j) \in E} w_{i,j} [g_i - g_j \neq 0] \quad (1)$$

其中: G 为 n_p 个三维点组成的超点; E 表示边组成的集合; f_i 表示第 i 个三维点局部邻域 D 形状; g_i 表示待求的解,相同的 g 值可视为一个超点; μ 表示正则化强度,用来区分分区的粗糙度; $w_{i,j}$ 表示三维点 i,j 之间的边权重; $[\cdot]$ 表示艾佛森括号。该理论可根据颜色、空间几何等点云特征将点云组成的单个物体分为几个有意义的块,也就是同质分割块,构成超点。恒定连通分量 $S = \{S_1, S_2, \dots, S_k\}$ 构成超点的集合即为式(1)的解。这些超点 S 除了具有原始三维空间特征和颜色特征外,还具有线性度 L_λ 、平面度 P_λ 和散射度 S_λ 这3种几何特征,它们分别由式(2)~式(4)确定:

$$S_\lambda = \frac{\lambda_3}{\lambda_1} \quad (2)$$

$$P_\lambda = \frac{\lambda_2 - \lambda_3}{\lambda_1} \quad (3)$$

$$L_\lambda = \frac{\lambda_1 - \lambda_2}{\lambda_1} \quad (4)$$

其中: $\lambda_1, \lambda_2, \lambda_3$ 均表示点云块协方差矩阵的特征值,这些几何特征^[25]可以表现点云块的分布模式。

2.3 自归一化门控循环神经单元

为更好地改进网络模型的上下文分割效果,增强网络模型中超点对周边邻域的感知能力,受 SNN 中缩放指数型线性单元(Scaled Exponential Linear Unit, SELU)^[26]能使神经单元具备自归一化属性特点的启发,构造了一种特殊的 RNN 神经单元,即 SN-GRU。该单元以传统门控循环单元(Gated Recurrent Unit, GRU)^[19]为基础,结合 SELU 线性单元而构造。这使得 SN-GRU 即继承了传统 GRU 防止梯度弥散的属性,又能使神经单元具有自归一化的特点。该特性能使网络在整个训练过程表现得更加鲁棒,从而有助于充分挖掘深层次的空间信息表征。SELU 单元主要使用函数 y 映射前后 2 层神经网络的均值和方差以达到归一化的效果: $y: (\mu, v) \mapsto (\tilde{\mu}, \tilde{v})$ 。SELU 单元如式(5)所示:

$$\text{SELU}(z) = \begin{cases} \beta \times z, & z > 0 \\ \beta \times \alpha (e^z - 1), & z \leq 0 \end{cases} \quad (5)$$

其中:参数 α, β 可根据 Banach 不动点定理(fixed-point theorem)和中央极限定理(central limit theorem)证得^[26],

$\alpha=1.673\ 263\ 242\ 354\ 377, \beta=1.050\ 700\ 987\ 355\ 480$ 。

SN-GRU能通过对同质点云块之间的远程语义依赖性建模,从而可以更加有效地捕获点云间的上下文信息。该神经网络结构如图2所示,由3个门组成:重置门(Reset Gate),更新门(Update Gate)和归一门(Normalize Gate)。

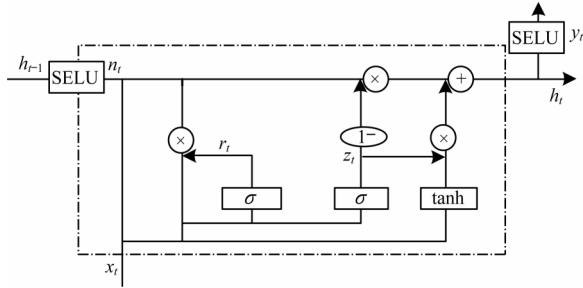


图2 SN-GRU单元结构

Fig.2 SN-GRU unit structure

重置和更新门控向量能够保存点云序列中空间信息,不会因为预测的弱相关性而被移除,可通过控制2个门控向量的开启和关闭,保存和更新有用的语义信息。归一门控向量能使神经元激励进行自动地转移(shift)和重缩放(rescale),在没有明确的归一化情况下实现零均值和单位方差。SN-GRU神经网络内部的计算流程如式(6)~式(11)所示:

$$n_t = \text{SELU}(h_{t-1}) \quad (6)$$

$$r_t = \sigma(W_r \cdot [n_t, x_t]) \quad (7)$$

$$z_t = \sigma(W_z \cdot [n_t, x_t]) \quad (8)$$

$$\tilde{h}_t = \tanh(W \cdot [r_t \odot n_t, x_t]) \quad (9)$$

$$h_t = (1 - z_t) \odot n_t + z_t \odot \tilde{h}_t \quad (10)$$

$$y_t = \text{SELU}(h_t) \quad (11)$$

其中: x_t 表示输入单元; y_t 表示输出单元; h_{t-1} 表示上个时刻的隐藏门; h_t 表示此时此刻的隐藏门; \tilde{h}_t 表示隐藏候选态; n_t, r_t, z_t 分别表示门控单元的归一门、重置门和更新门; W_r, W_z, W 表示权重矩阵; σ, \tanh 分别表示sigmoid函数激活和双曲正切函数激活;SELU表示隐藏门的激活函数; \cdot 表示矩阵内积; \odot 表示哈达玛积。SN-GRU主要通过门控实时调整激活函数的输入、输出,来捕获点云空间信息的中长期依赖关系,从而有效地实现场景分割。

2.4 损失函数设计

在网络模型中,损失函数是度量预测值与真实值之间差异的指标。在某些深度学习任务数据集中,往往会出现某些类别样本远大于其他类别样本的情况,所造成的样本不平衡性将影响模型学习的效果。在点云数据中,由于深度传感器和实际场景的影响,也会造成不同程度的样本不平衡。如本文数据集场景中地板和墙的数据量远多于书架和沙发

的数据量。针对该问题,本文将在二维图像领域中针对数据样本不平衡的焦点损失函数Focal Loss^[20]引入3D点云领域,以替代常规的交叉熵损失函数(Cross Entropy Loss),从而减少样本不平衡的影响。Focal Loss定义公式如式(12)和式(13)所示:

$$L_{\text{focal_loss}}(p_{\text{target}}) = -(1 - p_{\text{target}})^{\gamma} \cdot L_{\text{cross_entropy}} \quad (12)$$

$$L_{\text{cross_entropy}} = - \sum_{i=1}^N \sum_{c=1}^C \log_a((p_{\text{target}})(q_{\text{target}})) \quad (13)$$

其中: p_{target} 表示网络输入在经过softmax计算处理后得到的预测概率; q_{target} 表示样本真实类别标签的分布; N 代表网络训练批次的大小; C 代表点云数据集中心分割的类别数; $(1 - p_{\text{target}})^{\gamma}$ 表示调制系数; γ 表示调制系数中的超参数, $\gamma \geq 0$ 。

在网络训练过程中,普通交叉熵损失的正样本预测概率越大,损失越小,负样本则反之,这会使损失函数在大量简单样本的迭代中无法优化到最优。焦点损失函数在其基础上增加了一个调制因子 $(1 - p_{\text{target}})^{\gamma}$,能够使网络模型关注容易错分割的困难样本,减少易分割样本的损失。如待分割的目标预测概率为0.90时,则 $(1 - p_{\text{target}})^{\gamma}$ 的值将很小,而目标预测概率为0.21时, $(1 - p_{\text{target}})^{\gamma}$ 将很大,其损失也会相对较大。相对标准交叉熵,预测概率为0.5的目标,其损失的减少对于预测概率为0.90的目标将小很多倍。这可以增大困难样本如S3DIS数据集中心的书架等在损失函数的贡献度,降低简单样本的影响力。模型也会将计算力更多集中在困难样本中,提高模型的分割能力。

3 实验结果与分析

为验证所提模型的有效性,本文在被广泛使用的S3DIS数据集上进行了测试。实验环境为2块RTX 2080Ti GPU显卡,每块显卡有11 GB显存,CPU为Intel Core i9-9900K,服务器系统为Ubuntu16.04,语言为python 3.7版本,pytorch1.2版本。CUDA为10.0,Boost版本至少高于1.63.0。

3.1 实验数据集

为测试本文网络模型的泛化能力,本文采用斯坦福大学大规模3D室内场景数据集(S3DIS)^[21]。该数据集是目前最大的室内点云数据集,场景相对比较复杂,主要由6个大型室内区域组成,每个区域分别覆盖约为965 m²、1 100 m²、450 m²、1 700 m²、870 m²和935 m²,总共6 020 m²。这些区域在建筑风格和外观上属性不同,主要包括办公室、会议室、大堂、卫生间、教室、楼梯和走廊等建筑,共包含272个房间的3D扫描数据,共计超过2.15亿个点。扫描中的每个点都对应椅子(chair)、桌子(table)、地板(floor)、墙壁(wall)、天花板(ceiling)、柱子(column)、窗户(window)、门(door)、书架(bookcase)、沙发(sofa)、板(board)等13个类别语义标签。

3.2 网络模型参数设置

在同构分割之前,为提高分割效率,本文采用体素网格进行点云下采样,网格(bin)为3 cm。超点包含的三维点数量 n_p 设为128。网络训练中优化器采用ADAM,初始学习率为0.01,样本批处理大小(batchsize)为8,共训练250个周期(epoch),在周期为150和200时,学习率衰减指数为0.7,动量(momentum)为0.9。

3.3 场景语义分割评价指标

与点云分类和识别相比,三维点云语义分割需要精细地识别每个点的语义类别,具备一定挑战性。为进一步评估本文网络模型处理三维点云细粒度任务的能力,在三维点云领域最常用的主流数据集S3DIS上进行了语义分割实验,并与其他主流算法进行了指标对比。本文采用常用的语义分割评价指标,包括平均交并比MIOU、总体精度(Overall Accuracy, OA)和平均精度(mean Accuracy, mAcc)来衡量分割结果,分别由式(14)~式(16)定义:

$$M_{\text{IOU}} = \frac{1}{k+1} \sum_{i=0}^k \frac{T_p}{T_p + F_p + F_N} \quad (14)$$

$$O_{\text{OA}} = \sum_{i=0}^k \frac{T_p^i}{N_p^i} \quad (15)$$

$$m_{\text{mAcc}} = \frac{1}{k+1} \sum_{i=0}^k \frac{T_p^i}{N_p^i} \quad (16)$$

其中: k 代表分割的类别数; T_p 代表分类正确的正样本数量; F_p 代表分类错误的正样本数量; F_N 代表分类错误的负样本数量; T_p^i 表示第 i 类分割的正确的个数; N_p^i 表示总的点云的点数; N_p^i 表示第 i 类点云的总点数。

3.4 上下文语义分割的有效性

为验证SN-GRU神经单元对点云空间上下文分割的有效性,本文分别采用3种不同的RNN进行了对比实验,它们分别是长短期记忆模型(Long Short-Term Memory, LSTM)^[27]、GRU和SN-GRU。表1所示为在S3DIS数据集上进行6折交叉验证之后的实验结果,表中黑色粗体表示实验过程中最优结果。从表1中可以看出,本文所提方法取得了63.7%的MIOU和85.7%的OA准确率,优于其他2种RNN神经单元的结果。本文所提方法的准确率为73.8%,仅次于LSTM神经单元74.2%的准确率,原因可能是LSTM神经单元在提取较大面积的点云块特征如天花板、地板等的性能略优于本文的算法。

表1 不同循环神经网络单元分割精度对比

Table 1 Comparison of segmentation accuracy with different recurrent neural unit %

RNN 神经单元	平均交并比	总体精度	平均精度
LSTM	62.9	85.2	74.2
GRU	62.1	85.5	73.0
SN-GRU	63.7	85.7	73.8

3.5 Focal Loss 损失函数的有效性

为选取和验证Focal Loss损失函数中的超级参数,本文对调制系数 $(1-p_{\text{target}})^{\gamma}$ 中的 γ 取值进行了对比实验。表2中给出了网络模型在不同 γ 值下的MIOU、OA和mAcc的值,表中黑色粗体表示实验过程中最优结果。本文选取的 γ 值范围为0~4,从式(12)可以看出,当 $\gamma=0$ 时,Focal Loss损失函数相当于交叉熵损失函数,即本文所采用的baseline。从表2可以看出,当 $\gamma=1.5$ 时,场景语义分割达到了最优效果,MIOU、OA和mAcc在基准的水平上分别提高了1.4、0.3、1.1个百分点。当 $\gamma>1.5$ 时,场景语义评价指标的准确率开始下降并趋于稳定,说明当 $\gamma=1.5$ 时,已经能够有效提高较难分割样本如椅子、书架等的权重,降低了容易分割样本如墙面、地板等的权重,从而更有效地实现了语义分割。

表2 不同调制因子 γ 下网络模型分割精度对比

Table 2 Comparison of segmentation accuracy under different values of modulation γ %

调制因子 γ	MIOU	OA	mAcc
0.0	62.1	85.5	73.0
0.5	62.0	85.2	72.8
1.0	63.1	85.4	73.3
1.5	63.5	85.8	74.1
2.0	62.9	85.5	73.2
2.5	62.7	85.1	73.3
3.0	62.5	85.2	73.1
3.5	62.5	85.2	73.1
4.0	62.5	85.2	73.2

3.6 本文网络模型与现有方法比较

为从整体上验证本文网络模型在场景语义分割上的有效性,本文与目前主流网络模型进行了对比实验。实验结果如表3所示,表中“—”表示原文献中没有该项实验结果,黑色粗体表示实验过程中最优结果。从表中可以看出,本文网络模型的MIOU、OA和mAcc均达到最优值。与基于网络SPGragh相比,本文模型的MIOU、OA和mAcc分别提高1.7、0.9、1.3个百分点,与经典网络模型PointNet相比也分别提高了16.2、7.9、8.1个百分点,分割效果有了明显的提升。从表3中还可以看到本文网络模型在相对不易分割的目标上,MIOU有了较高的提升,如椅子的MIOU与基于网络相比提高了7.1个百分点,由此可见,本文网络模型在细粒度细节处的分割能力也有了较明显提升,从而验证了本文网络模型具有能够捕获点云深层细粒度几何特征的能力。

表3 S3DIS数据集上与现有主流方法的性能比较

Table 3 Performance comparison with the existing methods on S3DIS dataset

%

模型	MIOU	OA	mAcc	ceiling	floor	wall	beam	column
PointNet ^[18]	47.6	78.5	66.2	88.0	88.7	69.3	42.4	23.1
MS+CU ^[28]	47.8	79.2	59.7	88.6	95.8	67.3	36.9	24.9
G+RCU ^[28]	49.7	81.1	66.4	90.3	92.1	67.9	44.7	24.2
RSNet ^[29]	56.4	—	66.4	92.4	92.8	78.5	32.7	34.3
SPGraph ^[21]	62.1	85.5	73.0	89.9	95.1	76.4	62.8	47.1
本文模型	63.8	86.4	74.3	92.0	97.2	79.9	60.1	48.2
模型	window	door	table	chair	sofa	bookcase	board	clutter
PointNet ^[18]	47.5	51.6	54.1	42.0	9.6	38.2	29.4	35.2
MS+CU ^[28]	48.6	52.3	51.9	45.1	10.6	36.8	24.7	37.5
G+RCU ^[28]	52.3	51.2	58.1	47.4	6.9	39.0	30.0	41.9
RSNet ^[29]	51.6	68.1	60.1	59.7	50.2	16.4	44.8	52.0
SPGraph ^[21]	55.3	68.4	73.5	69.2	63.2	45.9	8.7	52.9
本文模型	56.3	70.2	75.1	76.3	61.1	49.2	13.5	50.2

3.7 点云分割结果可视化与分析

为更直观地展示本文网络模型场景语义分割的过程,本文将网络的输入、中间的同质点云块分割、标准分割结果(Ground Truth)和模型预测结果分别进行了可视化处理。如图3所示,在第1组办公室场景中,本文网络模型对墙、书架、门等简单几何物体

边缘的分割效果良好。在第3组的会议室场景中,由于白板(board)嵌在白色墙体(wall)内(彩色效果见《计算机工程》官网HTML版本),故在同质分割过程中将部分墙体和板面划分到一起,导致分割效果不理想。因此,如何提高相似颜色的同质划分为今后继续深入研究的方向。

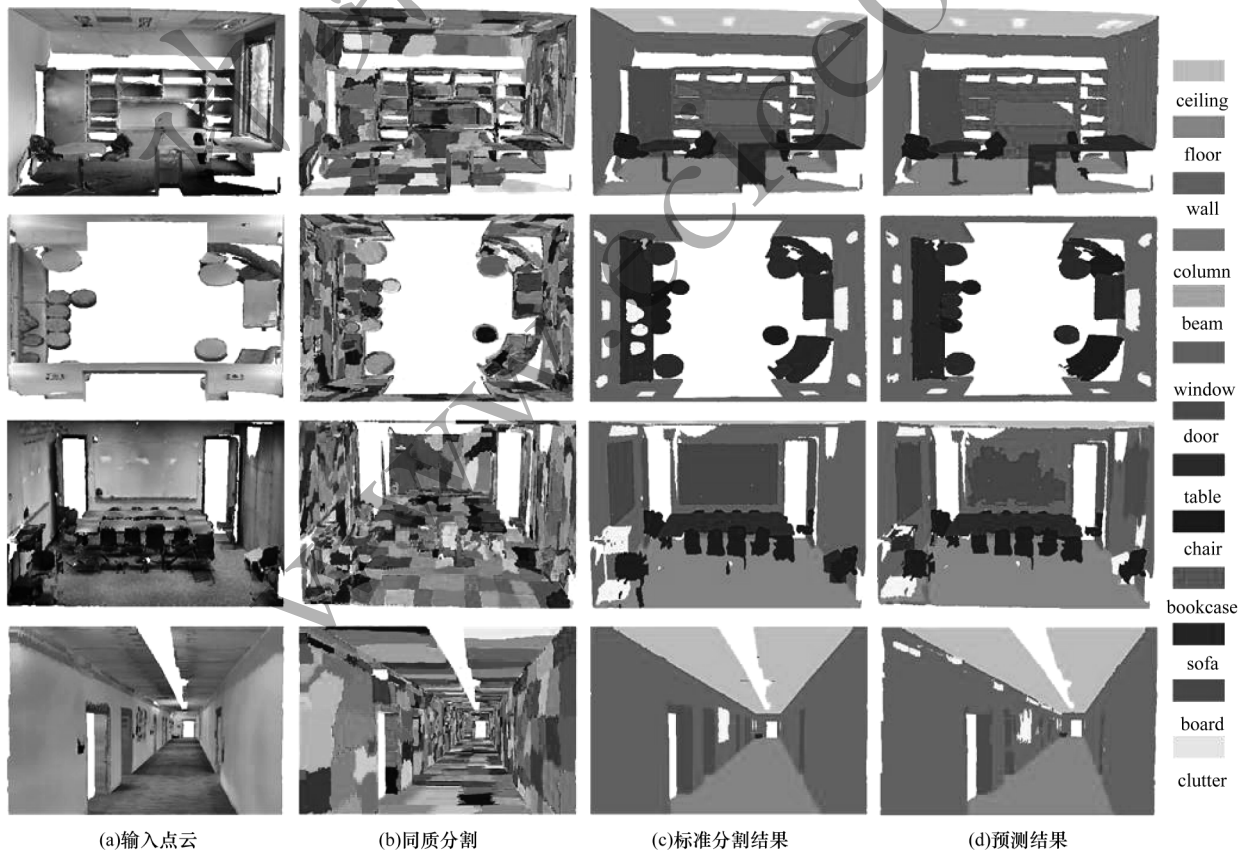


图3 S3DIS数据集上的部分结果可视化

Fig.3 Visualization of some results on S3DIS dataset

本文使用的 Focal Loss 损失函数能够提高困难样本的关注程度,与基于网络相比(如表3结果所示),其分割结果有了明显改善。在其他场景中,如第2组中的休息室场景,可以看到本文网络模型在桌子、椅子等相对独立的目标中获得了准确的分割结果,进一步证明本文的网络模型在物体细粒度分割上有明显的提升。总体来看,本文模型在对天花板、墙体和地板等面积相对大的物体,分割效果更好,尤其在其分割边界上的效果提高显著,这反映了本文提出的 SN-GRU 在上下文语义分割上有更加出色的表现,可以更有效地捕获点云块之间的联系,从而能准确地分割边缘。由于本文所采用的数据集(S3DIS)中大部分室内场景均为复杂场景,使本文的网络模型具有较高的泛化能力,在相对简单的场景中能够取得和标准分割高度一致的结果,如第4组的走廊场景所示。

4 结束语

为提高场景分割算法对点云块上下文信息的提取能力并有效利用点云数据集的样本均衡性,本文提出一种整合上下文空间几何特征且可调节难易样本权重的场景分割网络模型。通过融合多层感知机和循环神经单元提取场景中的局部特征和全局特征,实现较好的语义分割效果。同时通过采用大规模场景数据集和焦点损失函数的方法提高模型的泛化能力,实现复杂空间特征的提取。实验结果表明,与 PointNet 和 G+RCU 等传统模型相比,本文模型在大型室内场景数据集 S3DIS 上获得了较高的分割准确率。下一步将采用球形卷积核与八叉树空间划分相结合的方法,改善模型针对空间几何特征相似物体的分割效果。

参考文献

- [1] 姜名扬. 三维点云图形的分类和分割[D]. 上海: 上海交通大学, 2019.
JIANG M Y. Classification and segmentation of 3D point clouds[D]. Shanghai: Shanghai Jiao Tong University, 2019. (in Chinese)
- [2] CHEN X Z, MA H M, WAN J, et al. Multi-view 3D object detection network for autonomous driving[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2017: 1907-1915.
- [3] ZHU Y K, MOTTAGHI R, KOLVE E, et al. Target-driven visual navigation in indoor scenes using deep reinforcement learning[C]//Proceedings of IEEE International Conference on Robotics and Automation. Washington D. C., USA: IEEE Press, 2017: 3357-3364.
- [4] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6): 84-90.
- [5] 曾碧, 黄文. 一种融合多特征聚类集成的室内点云分割方法[J]. 计算机工程, 2018, 44(3): 281-286.
ZENG B, HUANG W. An indoor point cloud segmentation method fusing with multi-feature cluster ensemble[J]. Computer Engineering, 2018, 44(3): 281-286. (in Chinese)
- [6] CHO K, VAN MERRIËNBOER B, GULCEHRE C, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation[EB/OL]. [2020-09-02]. <https://arxiv.org/abs/1406.1078v3>.
- [7] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[C]//Proceedings of IEEE International Conference on Computer Vision. Washington D. C., USA: IEEE Press, 2017: 2980-2988.
- [8] ARMENI I, SAX S, ZAMIR A R, et al. Joint 2D-3D semantic data for indoor scene understanding[EB/OL]. [2020-09-02]. <https://arxiv.org/abs/1406.1078v3>.
- [9] SU H, MAJI S, KALOGERAKIS E, et al. Multi-view convolutional neural networks for 3D shape recognition[C]//Proceedings of IEEE International Conference on Computer Vision. Santiago, Chile: IEEE Press, 2015: 945-953.
- [10] 李浩, 霍雯, 裴春营, 等. 融合 VGG 与 FCN 的智能出租车订单预测模型[J]. 计算机工程, 2020, 46(12): 276-282.
LI HAO, HUO W, PEI C Y, et al. Intelligent taxi order forecasting model fusing VGG with FCN[J]. Computer Engineering, 2020, 46(12): 276-282. (in Chinese)
- [11] BOULCH A, GUERRY J, LE S B, et al. SnapNet: 3D point cloud semantic labeling with 2D deep segmentation networks[J]. Computers and Graphics, 2017, 71: 189-198.
- [12] MATURANA D, SCHERER S. VoxNet: a 3D convolutional neural network for real-time object recognition[C]//Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems. Washington D. C., USA: IEEE Press, 2015: 922-928.
- [13] LE T, DUAN Y. PointGrid: a deep network for 3D shape understanding[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2018: 9204-9214.
- [14] KLOKOV R, LEMPITSKY V. Escape from cells: deep KD-networks for the recognition of 3D point cloud models[C]//Proceedings of IEEE International Conference on Computer Vision. Washington D. C., USA: IEEE Press, 2017: 863-872.
- [15] 顾砾, 季怡, 刘纯平. 基于多模态特征融合的三维点云分类方法[J]. 计算机工程, 2021, 47(2): 279-284.
GU L, JI Y, LIU C P. A classification model of 3D point cloud based on multimodal feature fusion[J]. Computer Engineering, 2021, 47(2): 279-284. (in Chinese)

- [16] 李美佳,于泽宽,刘晓,等. 点云算法在医学领域的研究进展[J]. 中国图象图形学报, 2020, 25(10): 2013-2023. LI M J, YU Z K, LIU X, et al. Progress of point cloud algorithm in medical field [J]. Journal of Image and Graphics, 2020, 25(10): 2013-2023. (in Chinese)
- [17] 张蕊,李锦涛. 基于深度学习的场景分割算法研究综述[J]. 计算机研究与发展, 2020, 57(4): 859-875. ZHANG R, LI J T. A survey on algorithm research of scene parsing based on deep learning[J]. Journal of Computer Research and Development, 2020, 57(4): 859-875. (in Chinese)
- [18] QI C R, SU H, MO K, et al. Pointnet: deep learning on point sets for 3D classification and segmentation [C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2017: 652-660.
- [19] QI C R, YI L, SU H, et al. PointNet++: deep hierarchical feature learning on point sets in a metric space[EB/OL]. [2020-09-02]. https://www.researchgate.net/publication/317426798_PointNet_Deep_Hierarchical_Feature_Learning_on_Point_Sets_in_a_Metric_Space.
- [20] LANDRIEU L, OBOZINSKI G. Cut pursuit: fast algorithms to learn piecewise constant functions on general weighted graphs[J]. SIAM Journal on Imaging Sciences, 2017, 10(4): 1724-1766.
- [21] LANDRIEU L, SIMONOVSKY M. Large-scale point cloud semantic segmentation with superpoint graphs [C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2018: 4558-4567.
- [22] ACHANTA R, SHAJI A, SMITH K, et al. SLIC superpixels compared to state-of-the-art superpixel methods[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 34(11): 2274-2282.
- [23] SUTSKEVER I, VINYALS O, LE Q V. Sequence to sequence learning with neural networks[EB/OL]. [2020-09-02]. <https://wenku.baidu.com/view/9dad28f2c281e53a5902ffe9.html>.
- [24] SIMONOVSKY M, KOMODAKIS N. Dynamic edge-conditioned filters in convolutional neural networks on graphs[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2017: 3693-3702.
- [25] DEMANTKÉ J, MALLET C, DAVID N, et al. Dimensionality based scale selection in 3D lidar point clouds [J]. ISPRS-International Archives of the Photogrammetry Remote Sensing and Spatial Information Sciences, 2011, 38(5): 97-102.
- [26] KLAMBAUER G, UNTERTHINER T, MAYR A, et al. Self-normalizing neural networks[EB/OL]. [2020-09-02]. <https://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE07457319>.
- [27] GRAVES A. Long short-term memory: supervised sequence labelling with recurrent neural networks[J]. Springer, 2012, 9(8): 1735-1780.
- [28] ENGELMANN F, KONTOGIANNI T, HERMANS A, et al. Exploring spatial context for 3D semantic segmentation of point clouds [C]//Proceedings of IEEE International Conference on Computer Vision Workshops. Washington D. C., USA: IEEE Press, 2017: 716-724.
- [29] HUANG Q G, WANG W Y, NEUMANN U. Recurrent slice networks for 3D segmentation of point clouds[C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2018: 2626-2635.