一种面向散乱点云语义分割的深度残差–特征金字塔网络框架

彭秀平 [1] 仝其胜 [1] 林洪彬 [2] 冯 超 [1] 郑 武 [1]

摘 要 针对当前基于深度学习的散乱点云语义特征提取方法通用性差以及特征提取不足导致的分割精度和可靠性差的难题, 提出了一种散乱点云语义分割深度残差–特征金字塔网络框架. 首先, 针对当前残差网络在卷积方式上的局限性, 定义一种立方体卷积运算, 不仅可以通过二维卷积运算实现三维表示点的高层特征的抽取, 还可以解决现有的参数化卷积设计通用性差的问题;其次, 将定义的立方体卷积计算与残差网络相结合, 构建面向散乱点云语义分割的深度残差特征学习网络框架; 进一步, 将深度残差网络与特征金字塔网络相结合, 实现三维表示点高层特征多尺度学习与散乱点云场景语义分割. 实验结果表明, 本文提出的立方体卷积运算具有良好的适用性, 且本文提出的深度残差–特征金字塔网络框架在分割精度方面优于现存同类方法.

关键词 散乱点云, 语义分割, 立方体卷积, 残差网络, 特征金字塔网络

引用格式 彭秀平, 仝其胜, 林洪彬, 冯超, 郑武. 一种面向散乱点云语义分割的深度残差–特征金字塔网络框架. 自动化学报, 2021, **47**(12): 2831−2840

DOI 10.16383/j.aas.c190063

A Deep Residual − Feature Pyramid Network Framework for Scattered Point Cloud Semantic Segmentation

PENG Xiu-Ping[1] TONG Qi-Sheng[1] LIN Hong-Bin[2] FENG Chao[1] ZHENG Wu[1]

Abstract Aiming at the problem that the current scattered point cloud semantic segmentation based on the deep learning method is poor in generality and the problem of poor segmentation accuracy and reliability caused by insufficient feature extraction, a scattered point cloud depth residual-feature pyramid network framework is proposed. Firstly, for the limitation of the current residual network in the convolution mode, a cube convolution operation is defined, which can not only extract the high-level features of the three-dimensional representation point through the two-dimensional convolution operation, but also solve the problem of poor generality of current parameterized convolution design. Secondly, a deep residual feature learning network framework is constructed for scattered point cloud semantic segmentation through integrating the defined cube convolution operation into the residual network. Additionally, the proposed deep residual network is combined with the feature pyramid network to enable multi-scale learning of high-level features of three-dimensional representation points and semantic segmentation of scattered point cloud. The experimental results show that the cube convolution proposed in this paper has good applicability and the depth residual-feature pyramid network framework is superior to the existing similar methods in terms of segmentation accuracy.

Key words Scattered point cloud, semantic segmentation, cube convolution, residual network, feature pyramid network

Citation Peng Xiu-Ping, Tong Qi-Sheng, Lin Hong-Bin, Feng Chao, Zheng Wu. A deep residual — feature pyramid network framework for scattered point cloud semantic segmentation. *Acta Automatica Sinica*, 2021, **47**(12): 2831−2840

三维点云数据理解在计算机视觉和模式识别领域是一项非常重要的任务, 该任务包括物体分类, 目标检测和语义分割等. 其中, 语义分割任务最具

收稿日期 2019-01-26 录用日期 2019-07-30
Manuscript received January 26, 2019; accepted July 30, 2019
国家重点研发计划 (2017YFB0306402), 国家自然科学基金 (51305390, 61601401), 河北省自然科学基金 (F2016203312, E2020303188), 河北省高等学校青年拔尖人才计划项目 (BJ2018018), 河北省教育厅高等学校科技计划重点项目 (ZD2019039) 资助
Supported by National Key Research and Development Program of China (2017YFB0306402), National Natural Science Foundation of China (51305390, 61601401), Natural Science Foundation of Hebei Province (F2016203312, E2020303188), Young Talent Program of Colleges in Hebei Province (BJ2018018), and Key Foundation of Hebei Educational Committee (ZD2019039)
本文责任编委 黄庆明
Recommended by Associate Editor HUANG Qing-Ming
1. 燕山大学信息科学与工程学院 秦皇岛 066004 2. 燕山大学电气工程学院 秦皇岛 066004
1. School of Information Science and Engineering, Yanshan University, Qinhuangdao 066004 2. School of Electrical Engineering, Yanshan University, Qinhuangdao 066004

有挑战性, 传统的方法大多是在对点云数据进行必要特征提取的基础上, 应用支持向量机 (Support vector machine, SVM) 一类的分类算法, 通过训练一组特征分类器来完成散乱点云数据的语义分割任务[1]. 显然, 这类方法的性能很大程度上依赖点云特征的设计、特征提取的精度以及特征分类器的性能. 虽然国内外学者提出了几十种点云特征和大量的分类算法, 但是依然没有一种或几种特征能完全适用于所有语义分割场景, 算法适用性、精度和可靠性都得不到保障.

近年来, 随着深度学习技术的发展, 涌现出许多端对端的学习算法, 这种端对端的学习方式不依赖于手工设计的特征, 只需给定输入数据和对应的数据标签, 将其输入神经网络, 即可通过反向传播算法自动学习一组可以抽象高级特征的权重矩阵, 最后再由全连接层 (Fully connected layer, FC) 对高级特征进行分类, 从而完成分割任务. 现有的基于深度学习的点云分割研究方法大体可分为如下两类:

一类是基于散乱点云数据结构规则化的深度学习方法, 这类方法通常是先通过体素化处理或八叉树、KD 树等树形结构, 将无序和不规则的散乱点云处理成规则的结构化数据, 再将结构化数据输入三维卷积神经网络 (Convolutional neural network, CNN) 进行训练. 基于体素化方法[2−3] 的提出首次将深度学习技术应用于三维点云数据理解任务, 通过将三维点云体素化为规则的结构化数据, 解决了其无序性和不规则性问题. 但是, 使用体素占用表示三维点云带来了量化误差问题, 为了减小这种误差必须使用更高的体素分辨率表示, 高分辨率的体素表示又会在训练神经网络过程中带来大内存占用问题. 因此, 受限于目前计算机硬件的发展水平, 基于体素化的散乱点云深度学习方法往往难以完成诸如大规模室内三维场景一类的复杂、需要细粒度的三维场景的语义分割和场景理解任务. 基于八叉树[4−5] 和 KD树[6] 结构方法的提出解决了直接体素化点云所带来的大内存占用问题, 但是这种树形结构对三维点云旋转和噪声敏感, 从而导致卷积内核的可训练权重矩阵学习困难, 算法的鲁棒性往往不够理想.

另一类是基于参数化卷积设计的深度学习方法, 这类方法以原始三维点云作为输入, 通过设计一种能够有效抽象高层次特征的参数化卷积, 再使用堆叠卷积架构来完成点云分割任务. PointNet[7] 是这类方法的代表, 其首先使用共享参数的多层感知机 (Multi-layer perception, MLP) 将三维点云坐标映射到高维空间, 再通过全局最大池化 (Global max pooling, GMP) 得到点云全局特征, 解决了点云的无序性问题; 此外, 文献 [7] 还提出了一种 T-net 网络, 通过学习采样点变换矩阵和特征变换矩阵解决了散乱点云的旋转一致性问题, 但是由于缺乏点云局部特征信息局限了其在点云分割任务中的性能. 随后 PointNet++[8] 提出一种分层网络, 通过在每一图层递归使用采样、分组、PointNet 网络来抽象低层次特征和高层次特征, 再经过特征反向传播得到融合特征, 最终使用全连接层预测点语义标签, 解决了文献 [7] 方法对点云局部特征信息提取不足的问题. RSNet[9] 通过将无序点云特征映射为有序点云特征, 再结合循环神经网络 (Recurrent neural network, RNN) 来提取更丰富的语义信息, 从而进行语义标签预测. PointCNN[10] 则定义了一种卷积, 通过学习特征变换矩阵将无序点云特征变换成潜在的有序特征, 再使用堆叠卷积架构来完成点云分割任务.

总体而言, 基于参数化卷积设计的深度学习为散乱三维点云场景的理解提供了具有广阔前景的新方案. 然而, 目前该领域的研究尚处于萌芽阶段, 许多切实问题尚待解决, 如: 由于卷积方式的局限性, 用于二维图像处理的主流深度神经网络构架 (如: U-Net[11], ResNet[12], Inception V2/V3[13], DenseNet[14] 等) 无法直接用于三维散乱点云数据的处理; 由于针对点云特征提取设计的参数化卷积的局限性, 现有的方法普遍存在特征抽象能力不足、无法将用于二维图像处理的一些主流神经网络框架适用于三维点云分割任务等问题.

基于此, 本文设计了一种立方体卷积运算, 不仅可以通过二维卷积实现三维表示点的高层特征的抽取, 还可以解决当前参数化卷积设计通用性差的问题; 其次, 将定义的立方体卷积计算和残差网络相结合, 构建面向散乱点云语义分割的深度残差特征学习网络框架; 进一步, 将深度残差网络与特征金字塔网络相结合, 以实现三维表示点高层特征多尺度学习和语义分割.

1 提出的方法

本文方法以原始三维点云作为输入, 首先, 将定义的立方体卷积运算和残差网络 (ResNet) 相结合, 构建面向散乱点云语义分割的深度残差特征学习网络框架; 其次, 将深度残差网络与特征金字塔网络 (Feature pyramid network, FPN)[15] 相结合, 以实现三维表示点高层特征多尺度学习; 最后, 通过全连接层对融合特征进行分类得到语义标签输出, 整体分割网络框架如图 1 所示.

1.1 立方体卷积

残差网络自 2015 年提出以来, 一经出世, 便在

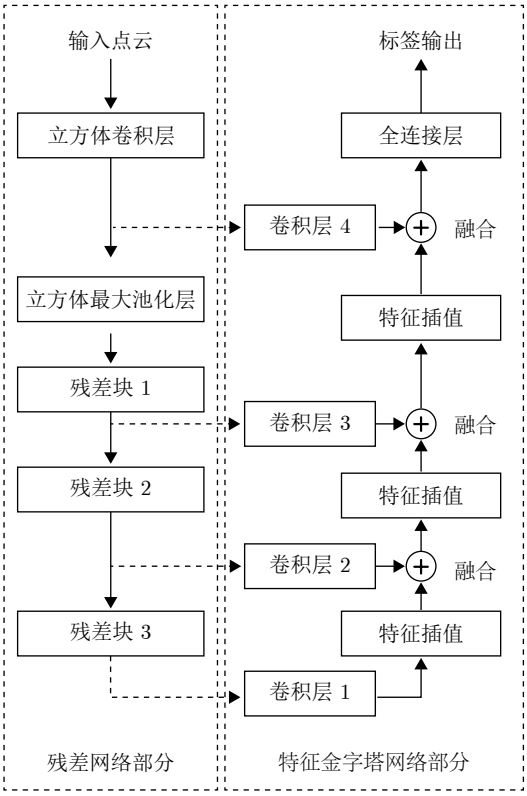


图 1　深度残差–特征金字塔网络框架

Fig. 1　Depth residual – feature pyramid network framework

ImageNet 竞赛中斩获图像分类、检测、定位三项的冠军. 随后许多基于残差网络的研究在图像分割领域也取得了巨大的成功[16–20]. 然而, 残差网络是专门为二维图像类的规则化数据设计的深度网格结构, 在处理类似散乱三维点云等非规则、无序化散乱数据时遇到困难; 另一方面, 现有的基于深度学习的用于点云特征提取的参数化卷积设计普遍存在卷积计算通用性差、无法拓展至现有二维图像处理的深度学习框架的问题. 为此, 本文在深入研究现有二维图像卷积计算的基础上, 基于局部点云结构规则化思想, 提出一种新的适用于散乱三维点云的立方体卷积计算模型, 旨在通过二维卷积运算实现散乱三维点云数据高层次特征的抽象; 同时, 该立方体卷积计算模型具有良好网络框架适用能力, 使现有大多数二维图像处理深度神经网络可用于散乱三维点云分割中.

本文提出的立方体卷积计算模型设计思路如下:

考虑一幅二维图像, 集合表示为 $S=\{s_{x,y}\in\mathbf{R}^3|x=0,1,2,\cdots,h;y=0,1,2,\cdots,w\}$. 对于任意像素点 $s_{x,y}$ $(1\le x\le h-1;1\le y\le w-1)$, 当大小为 3×3 的卷积核作用于该点时, 将其局部感受野表示为集合 $N_{x,y}=\{s_{a,b}\in\mathbf{R}^3|a=x-1,x,x+1;b=y-1,y,y+1\}$, 将学习权重矩阵表示为

$$W=\begin{bmatrix}w_{1,1} & w_{1,2} & w_{1,3}\\ w_{2,1} & w_{2,2} & w_{2,3}\\ w_{3,1} & w_{3,2} & w_{3,3}\end{bmatrix}\tag{1}$$

其中, $w\in\mathbf{R}^3$. 在二维图像分割任务中, 卷积神经网络可通过端对端的方式学习得到权重矩阵 W, 实现图像高层次特征抽取, 其最重要的原因之一在于: 在固定视角下, 感受野所包围的像素点是关于给定表示像素点的欧氏距离邻近点, 且权重值和像素点具有位置对应关系, 如图 2 所示. 然而, 由于散乱三维点云具有无序性和不规则性, 同一点云模型可用多种不同的集合表示, 当直接把二维卷积神经网络应用到三维点云上时, 并不能保证感受野所包含的点与表示点是这种欧氏距离邻近关系, 也无法保证权重值和点是位置对应的.

为此, 基于局部点云结构规则化的思想, 本文提出一种适用于三维点云特征提取的立方体卷积运算, 具体过程描述如下: 设三维点云表示为集合 $P=\{p_i\in\mathbf{R}^3|i=0,1,2,\cdots,n-1\}$, n 为点的个数, 其对应的特征集合为 $F=\{f_i\in\mathbf{R}^c|i=0,1,2,\cdots,n-1\}$, c 为点的特征维度. 对于表示点 p_i, 定义一个边长为 s 的立方体. 以 p_i 中心, 将立方体按网格划分为 27 个子立方体, 每个子立方体以固定顺序进行索引, 如图 3 所示.

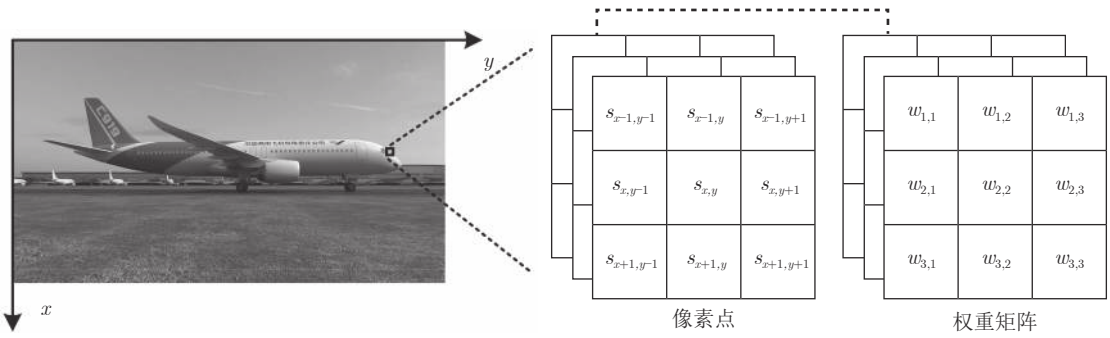


图 2　二维卷积

Fig. 2　The 2D convolution

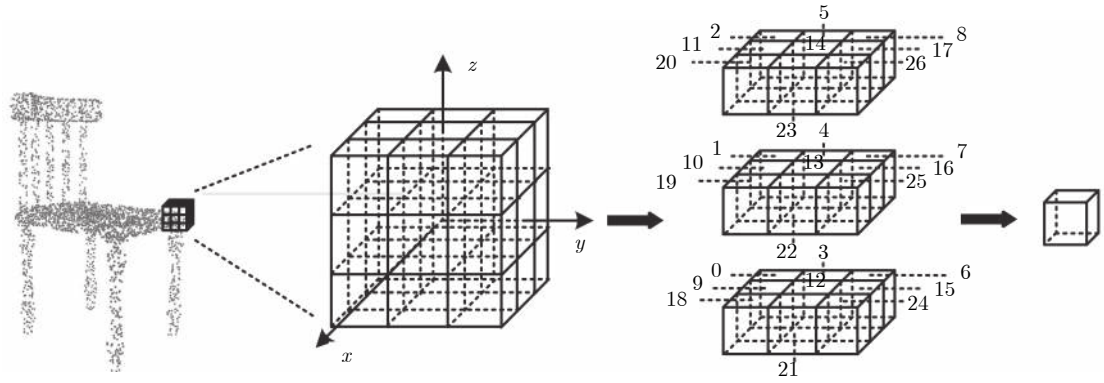


图 3 立方体卷积

Fig. 3 The cube convolution

首先, 定义点云局部特征集合 $V=\{v_i\in\mathbf{R}^{27\times c}|i=0,1,2,\cdots,n-1\}$, 其中, $v_i=\{v_{i,j}\in\mathbf{R}^c|j=0,1,2,\cdots,26\}$. 对于第 j 个子立方体, 选取离子立方体中心欧氏距离最近的一个点作为表示点 p_i 的一个邻近点, $v_{i,j}$ 设置为该邻近点的特征. 当集合 F 等于 P (即输入点云的特征为其三维坐标) 时, $v_{i,j}$ 设置为该邻近点相对表示点的相对坐标值, 如果子立方体内没有点, 则设置为 0. 遍历所有表示点得到特征集合 V, 然后, 再通过二维卷积对特征集合 V 进行卷积来抽象输入三维点云的高层特征. 那么卷积输出即可表示为: $F'=Conv\ (V,1\times 27,c')$, 其中 V 为卷积输入, 1×27 为卷积核和移动步长大小, c' 为输出特征通道数. 对于表示点的邻近点选取, 另一可行方案是: 首先计算子立方体内包围点到其中心的距离, 再对包围点的特征进行反距离加权平均得到的值, 但是随着网络层数的加深以及特征维度的升高将会带来计算量的大幅增加. 因此, 为平衡性能本文选取距离子立方体中心最近的一个点作为表示点的一个邻近点这一近似方案.

本文提出的立方体卷积运算主要有两个关键点: 立方体网格划分和子立方体固定索引排序. 我们给出分析如下: 在二维图像理解领域中, 许多主流网络框架 (如 U-Net[11]、ResNet[12]、Inception V2/V3[13] 和 DenseNet[14] 等) 都是采用大小为 3×3 的卷积核来提取特征. 文献 [13] 中指出大卷积核可以通过小卷积核的叠加获得相同大小的感受野, 并且小卷积核的叠加引入了二次非线性, 其实验结果证明了精确率会得到提升. 其次, 相比大卷积核, 小卷积核具有更小的参数量. 鉴于此, 本文采用将空间划分为网格的方式来感知表示点的局部三维空间结构, 从而可以以较小的参数量获得较高的精确率; 另外, 由于三维点云具有无序性, 直接使用二维卷积进行卷积运算会导致神经网络在训练过程中无法学习到有效的权重矩阵来抽象高层特征. 因此, 本文也是基于常用于二维图像理解任务中的二维卷积特点, 通过将子立方体以固定索引进行排序的方式来保证在固定坐标系下二维卷积核的可学习权重矩阵和无序点云有一一对应的关系, 以此通过二维卷积有效地抽象三维点云的高层特征, 使得用于二维图像处理的主流神经网络框架可以适用于三维点云理解任务中.

1.2 立方体最大池化

在二维图像中, 最大池化是对邻域内特征点取最大值运算, 可以学习某种不变性 (旋转、平移、尺度缩放等). 为了使残差网络完全适用于三维点云分割任务从而可以学习三维点云的旋转、平移、尺度缩放不变等特性, 本文提出一种基于局部点云结构规则化思想的立方体最大池化方法, 具体过程描述如下:

给定三维点云集合 $P=\{p_i\in\mathbf{R}^3|i=0,1,2,\cdots,n-1\}$, n 为点的个数, 以及对应的特征集合 $F=\{f_i\in\mathbf{R}^c|i=0,1,2,\cdots,n-1\}$, c 为点的特征维度. 首先用文献 [7] 中迭代最远点采样算法得到采样点集合 $P'=\{p'_l\in\mathbf{R}^3|\ l=0,1,2,\cdots,m-1\}$, m 为采样点的个数, 定义点云局部特征集合 $V=\{v_l\in\mathbf{R}^{27\times c}|l=0,1,2,\cdots,m-1\}$, 其中 $v_l=\{v_{l,j}\in\mathbf{R}^c|j=0,1,2,\cdots,26\}$. 对于采样点 p'_l, 使用本文提出的立方体卷积运算中邻近点搜索方法在 P 中搜索其邻近点, 得到邻近点特征集合 v_l. 遍历所有采样点得到特征集合 V. 然后, 再通过二维最大池化对特征 V 进行最大池化处理. 最大池化输出即可表示为 $F'=\max(V,1\times 27)$, 其中, V 为最大池化输入, 1×27 为卷积核和移动步长大小.

与本文提出的立方体卷积运算相似, 立方体最大池化的关键点也在于 $3\times 3\times 3$ 网格划分, 所获取

的邻近点对表示点的局部点云几何结构表示完整性将直接影响最大池化的输出. 因此, 我们同样基于局部点云结构规则化的思想, 通过将局部空间划分为网格来获取表示点更加合理的邻近点, 以此再通过常用于二维图像处理的最大池化操作对输入特征进行最大池化处理.

1.3　三维点云特征残差学习结构

残差网络由文献 [12] 提出, 其核心思想是一种特殊的残差结构. 这种结构通过将神经网络中特征映射近似问题转化为残差学习问题, 不仅解决了随着神经网络层数的加深出现的梯度退化问题, 而且能够以更少的模型参数实现更高的准确率. 为了将这种结构适用于三维点云分割任务, 本文提出一种三维点云特征残差学习结构, 如图 4 所示.

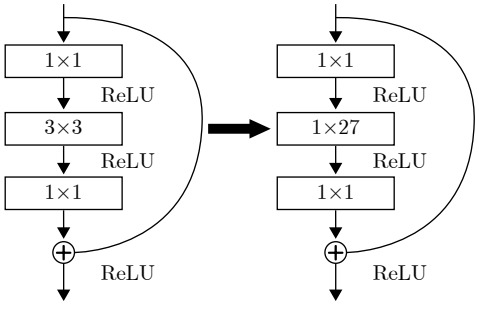


图 4　三维点云特征残差学习结构

Fig. 4　The residual learning structure for 3D point cloud feature

本文提出的三维点云特征残差学习结构和文献 [12] 提出的残差结构都由三层卷积和一次跳跃连接组成, 第 1 层和第 3 层同为卷积核大小为 1×1 的二维卷积. 不同之处在于: 对于二维图像, 当输入输出维度不匹配时, 文献 [12] 通过第 1 层卷积核大小为 1×1、移动步长为 2×2 的卷积层进行下采样. 针对三维点云, 我们则先使用文献 [7] 中迭代最远点采样算法进行下采样, 再通过卷积核大小为 1×1、移动步长为 1×1 的卷积层进行卷积; 另外, 在文献 [12] 中第 2 层为卷积核大小为 3×3、移动步长为 1×1 的卷积层, 我们将其替换为本文提出的立方体卷积运算层.

1.4　三维点云特征金字塔网络

在文献 [19] 中, 特征金字塔网络的组成结构是首先使用最邻近上采样法把高层特征做 2 倍上采样, 然后与对应的前一层特征相加融合. 考虑到特征金字塔网络是为具有规则像素网格结构排列的二维图像设计的, 而三维点云数据具有不规则性, 当表示点的局部点云分布密度不均时, 最邻近点的特征并不能够准确近似表示点的特征. 因此, 我们采用文献 [8] 中的基于 K 邻近的反距离加权插值法进行特征上采样, 如图 5 所示. 反距离加权插值可表示为

$$f^{(j)}(x)=\frac{\sum\limits_{i=1}^{k}w_i(x)f_i^{(j)}}{\sum\limits_{i=1}^{k}w_i(x)} \tag{2}$$

其中, $w_i(x)=\frac{1}{d^p(x,x_i)}$, $j=1,\cdots,C$, C 为待插值点的特征维度, $d(x,x_i)$ 表示待插值点 x 和其邻近点 x_i 的平方欧氏距离, $f^{(j)}(x)$ 和 $f_i^{(j)}$ 分别表示点 x 和 x_i 的特征向量, 与文献 [8] 相同 p 设置为 2, k 设置为 3.

2　实验结果与分析

本文实验环境为 Intel(R) Core(TM) i7-7800X CPU @ 3.50 GHz, 16 GB×4 内存, NVIDIA GTX 1080Ti×2 GPU, 系统为 Ubuntu 16.04.

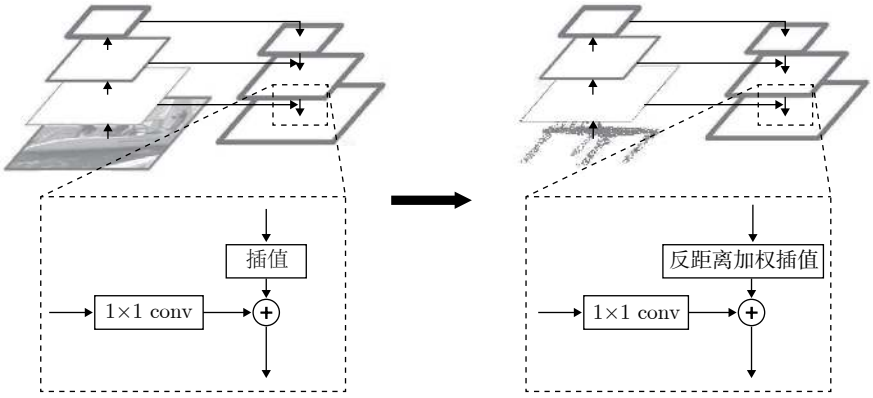


图 5　三维点云特征金字塔网络

Fig. 5　The feature pyramid network for 3D point cloud

2.1 实验数据集

实验数据集为 S3DIS 数据集[21] 和 ScanNet 数据集[22].

S3DIS 数据集总共包含 271 个由 Matterport 扫描仪从真实室内场景扫描得到的场景数据, 包含在 6 个文件夹中. 本文采用与 S3DIS 官方相同的 K 折交叉验证策略进行数据集划分. 采用与文献 [10] 相同的训练方法, 将原始场景沿着 x 轴和 y 轴分成大小为 1.5 m×1.5 m 的小块, 使用点的位置和颜色信息用于训练和测试, 并在训练期间将块点云沿 Z 轴随机旋转一定角度进行数据增强处理.

ScanNet 数据集总共包含 1 513 个从真实室内环境扫描并重建得到的场景数据. 本文按照 ScanNet 官方划分标准将数据集分为训练集和测试集两部分. 采用和 S3DIS 数据集相同的训练方法. 另外, 由于其他方法没有使用颜色信息用于训练, 因此, 为了公平比较, 本文也不使用颜色信息.

2.2 参数设计

为了验证本文提出的立方体卷积运算的有效性和本文方法的可行性, 我们基于文献 [19] 提出的用于二维图像分割的残差网络–特征金字塔网络结合本文提出的立方体卷积运算构建一种面向散乱点云语义分割的残差网络–特征金字塔网络框架 (下文中以 ResNet-FPN_C 表示), 网络结构参数设计如表 1 所示. 所有程序由开源框架 TensorFlow 及其 Python 接口实现, 采用 ADAM (Adaptive moment estimation) 方法进行训练.

2.3 评价指标

本文采用总体精确率 ($oAcc$)、类别平均精确率 ($mAcc$) 和类别平均交并比 ($mIoU$) 评价指标对试验结果进行评估, 并与其他方法进行对比. 假设共有 k 个类别, 定义 p_{ii} 表示类别 i 的预测标签等于真实标签的个数, p_{ij} 表示类别 i 的标签预测为类别 j 的个数. 则 $oAcc$ 可表示为

$$oAcc = \frac{\sum\limits_{i=0}^{k-1} p_{ii}}{\sum\limits_{i=0}^{k-1}\sum\limits_{j=0}^{k-1} p_{ij}} \tag{3}$$

$mAcc$ 表示为

$$mAcc = \frac{1}{k}\sum_{i=0}^{k-1}\frac{p_{ii}}{\sum\limits_{j=0}^{k-1} p_{ij}} \tag{4}$$

$mIoU$ 表示为

$$mIoU = \frac{1}{k}\sum_{i=0}^{k-1}\frac{p_{ii}}{\sum\limits_{j=0}^{k-1} p_{ij} + \sum\limits_{j=0}^{k-1} p_{ji} - P_{ii}} \tag{5}$$

另外, 由于 ScanNet[22] 提供的全卷积神经网络基线方法以体素化数据作为输入, 其预测标签是基于体素统计的, 因此我们采用与文献 [8] 相同的方法将点预测标签转化为体素预测标签来统计预测结果的 $oAcc$、$mAcc$ 和 $mIoU$ 指标.

2.4 实验结果分析

为验证本文方法的有效性, 我们在 S3DIS 数据

表 1 参数设计

Table 1 The parameter design

层	卷积核大小	立方体边长 (m)	输出点个数	输出特征通道数
立方体卷积	$1\times 27, 64$	0.1	8 192	64
立方体最大池化	$1\times 27, 64$	0.1	2 048	64
残差块1	$\begin{bmatrix}1\times 1, 64\\ 1\times 27, 64\\ 1\times 1, 256\end{bmatrix}\times 3$	0.2	2 048	256
残差块2	$\begin{bmatrix}1\times 1, 12\,8\\ 1\times 27, 128\\ 1\times 1, 512\end{bmatrix}\times 4$	0.4	512	512
残差块3	$\begin{bmatrix}1\times 1, 256\\ 1\times 27, 256\\ 1\times 1, 1024\end{bmatrix}\times 6$	0.8	128	1 024
卷积层1	$1\times 1, 256$	—	128	256
卷积层2	$1\times 1, 256$	—	512	256
卷积层3	$1\times 1, 256$	—	2 048	256
卷积层4	$1\times 1, 256$	—	8 192	256
全连接层	—	—	8 192	20

集和 ScanNet 数据集上进行了测试. 同时, 为进一步证明本文提出的立方体卷积运算的通用性, 我们基于文献 [11] 提出的用于医学图像分割的 U-Net 网络搭建了适用于三维点云语义分割的 U-Net 网络框架 (下文中以 U-Net_C 表示).

我们统计了在 S3DIS 数据集上的 6 折交叉验证结果, 如表 2 所示, 本文 ResNet-FPN_C 方法和 U-Net_C 方法在 $oAcc$、$mAcc$ 和 $mIoU$ 评价指标上均优于其他方法. 各类别 IoU 统计结果如表 3 所示, 其中, 本文 ResNet-FPN_C 方法有 7 个优于其他方法, U-Net_C 方法有 8 个优于其他方法.

另外, 我们在 ScanNet 数据集上也进行了测试, 结果如表 4 所示, 从中可以看出, 本文 ResNet-FPN_C 方法和 U-Net_C 方法在 $oAcc$、$mAcc$ 和 $mIoU$ 评价指标上均优于其他方法. 各类别的 IoU 统计结果如表 5 所示, 其中, 本文 ResNet-FPN_C 方法有 8 个优于其他方法, U-Net_C 方法有 10 个优于其他方法.

由于 ScanNet 数据集是由便携设备从真实室内场景扫描重建得到的, 其重建场景中存在大量缺失、未标注、杂乱信息, 且相比 S3DIS 数据集包含更多标注类别, 因此其语义分割任务更具挑战性. 从 S3DIS 数据集和 ScanNet 数据集测试结果可以看出, 本文方法相比其他方法在难以识别的小物体 (如 picture) 和复杂结构物体 (如 chair、sofa) 类别上具有更好的分割性能. 值得注意的是 door 和 window 这两种类别, 它们在空间位置和几何结构上和 wall 很相近, 相比于其他类别这两种类别的分割难度更大, 而本文方法较其他方法有较大的分割精度提升. 我们分析如下: PointCNN 方法采用的是基于 KNN (K-nearest neighbor) 的邻近点搜索算法, 由于这种算法对点云分布密度比较敏感, 当

表 2　S3DIS 数据集分割结果比较 (%)

Table 2　Segmentation result comparisons on the S3DIS dataset (%)

评价指标	PointNet[7]	RSNet[9]	PointCNN[10]	ResNet-FPN_C (本文)	U-Net_C (本文)
$oAcc$	78.5	—	88.14	**89.6**	**88.53**
$mAcc$	66.2	66.45	75.61	**76.83**	**76.98**
$mIoU$	47.6	56.47	65.39	**67.05**	**67.37**

表 3　S3DIS 数据集各类别 IoU 分割结果比较 (%)

Table 3　Comparison of IoU for all categories on the S3DIS dataset (%)

类别	PointNet[7]	RSNet[9]	PointCNN[10]	ResNet-FPN_C (本文)	U-Net_C (本文)
ceiling	88.0	92.48	**94.78**	91.99	91.46
floor	88.7	92.83	**97.30**	94.99	94.12
wall	69.3	**78.56**	75.82	77.04	**79.00**
beam	42.4	32.75	**63.25**	50.29	51.92
column	23.1	34.37	**51.71**	39.40	40.35
window	47.5	51.62	58.38	**65.57**	**65.63**
door	51.6	68.11	57.18	**72.38**	**72.60**
table	54.1	60.13	71.63	**72.20**	**71.57**
chair	42.0	59.72	69.12	**77.10**	**77.56**
sofa	9.6	50.22	39.08	**54.87**	**55.89**
bookcase	38.2	16.42	**61.15**	59.24	59.10
board	29.4	44.85	52.19	**53.44**	**54.54**
clutter	35.2	52.03	58.59	**63.11**	**62.11**

表 4　ScanNet 数据集分割结果比较 (%)

Table 4　Segmentation result comparisons on the ScanNet dataset (%)

评价指标	ScanNet[22]	PointNet[7]	PointNet++[8]	RSNet[9]	PointCNN[10]	ResNet-FPN_C (本文)	U-Net_C (本文)
$oAcc$	73.0	73.90	84.50	—	85.1	**85.5**	**85.3**
$mAcc$	—	19.90	43.77	48.37	57.9	**63.1**	**62.8**
$mIoU$	—	14.69	34.26	39.35	43.7	**45.0**	**46.5**

点云分布密度不均时, 所获取的邻近点可能全部来自表示点的同一个方向, 此时邻近点不能准确反映表示点的局部特征, 且由于其卷积设计的局限性对局部空间几何结构的微小变化也不够敏感; PointNet++提出的 MSG (Muti-scale grouping) 和 MRG (Muti-resolution grouping) 方法虽然能够更加合理地获取邻近点, 但是由于采用的是 PointNet 中共享参数的多层感知机结合全局最大池化的特征提取方法, 而全局最大池化会丢失信息. 因此, 同样不能准确抽象表示点的高层特征; RSNet 则是先将点云数据分别沿着 x, y, z 方向进行切片, 再将切片后的点云所对应的特征输入循环神经网络提取特征. 其试验结果表明这种方法对平面结构物体 (如 wall、floor、desk 等) 有较高的分割精度, 但是将点云切片会严重丢失点的空间邻域关系, 从而导致循环神经网络很难学习非平面复杂结构物体的特征. 而本文提出的立方体卷积运算通过将局部空间划分为 $3\times3\times3$ 网格来获取表示点的邻近点, 能对点云分布密度不均具有更好的鲁棒性, 且能感知空间几何结构的微小变化. 另外, 通过对所获取的邻近点进行排序可以使得二维卷积能够感知视角信息, 从而准确地抽象表示点的高层特征. 因此, 相比其他方法本文方法具有更好分割性能.

同时我们也做了耗时统计实验, 所有方法均在相同实验环境下以在 1080Ti 单 GPU 上所发挥的最大性能统计. 如表 6 所示, 当输入点云个数为 8192 时, 本文 ResNet-FPN_C 方法单 batch 平均训练时间和前向传播时间分别为 0.060 s 和 0.042 s, 略慢于其他方法. 虽然 U-Net_C 方法在 mIoU 指标上可以取得更好的结果, 但是其速度也明显降低. 因此, 本文提出的 ResNet-FPN 网络具有更为平衡的运行效率和分割精度. 由于本文提出的立方体卷

表 5 ScanNet 数据集各类别 *IoU* 分割结果比较 (%)

Table 5 Comparison of *IoU* for all categories on the ScanNet dataset (%)

类别	PointNet[7]	PointNet++[8]	RSNet[9]	PointCNN[10]	ResNet-FPN_C (本文)	U-Net_C (本文)
wall	69.44	77.48	**79.23**	74.5	77.1	77.8
floor	88.59	92.50	**94.10**	90.7	90.6	90.8
chair	35.93	64.55	64.99	68.8	**76.4**	**76.6**
table	32.78	46.60	51.04	**55.3**	52.5	50.8
desk	2.63	12.69	**34.53**	28.8	29.1	25.8
bed	17.96	51.32	55.95	56.1	**57.0**	**57.4**
bookshelf	3.18	52.93	**53.02**	38.9	42.7	42.3
sofa	32.79	52.27	55.41	60.1	**61.5**	**60.9**
sink	0.00	30.23	34.84	**41.9**	41.8	41.7
bathtub	0.17	42.72	49.38	**73.5**	61.0	68.6
toilet	0.00	31.37	54.16	**73.4**	69.8	73.2
curtain	0.00	32.97	6.78	**36.1**	35.4	**41.9**
counter	5.09	20.04	**22.72**	22.5	20.0	19.7
door	0.00	2.02	3.00	7.5	**26.0**	**28.4**
window	0.00	3.56	8.75	11.0	**15.9**	**18.8**
shower curtain	0.00	27.43	29.92	**40.6**	38.1	**42.6**
refrigerator	0.00	18.51	37.90	43.4	**47.3**	**52.3**
picture	0.00	0.00	0.95	1.3	**5.8**	**8.6**
cabinet	4.99	23.81	**31.29**	26.4	27.5	27.5
other furniture	0.13	2.20	18.98	23.6	**25.8**	**24.5**

表 6 耗时比较

Table 6 Comparison of running time

	PointNet[7]	PointCNN[10]	ResNet-FPN_C (本文)	U-Net_C (本文)
输入点个数	8 192	2 048	8 192	8 192
单batch训练时间 (s)	0.035	0.047	0.060	0.140
单batch前向传播时间 (s)	0.023	0.016	0.042	0.068

积运算具有简单、通用性强等特点, 可以将用于二维图像处理的一些主流神经网络适用于三维点云分割任务, 因此后续我们将尝试更多主流的神经网络框架或针对三维点云分割任务对网络结构进行改进来提高精确率和减少耗时. 另外值得一提的是, 由于 PointCNN 方法中参数化卷积设计的局限性, 限制了其输入点的个数, 而本文方法当输入点个数 4 倍于 PointCNN方法时, 在训练时间和前向传播时间方面依然取得了不错的表现. 验证了本文方法的有效性和可行性.

3 结语

本文分析了二维卷积的特点和现有参数化卷积设计的局限性, 提出了一种通用立方体卷积运算, 以通过二维卷积实现三维表示点的高层特征的抽取; 基于此, 提出了一种面向散乱点云语义分割的深度残差–特征金字塔网络, 将用于二维图像处理的神经网络框架适用到了三维点云分割任务中. 实验结果表明, 本文提出的立方体卷积运算具有良好的适用性, 且本文提出的深度残差–特征金字塔网络框架在分割精度方面优于现存同类方法. 在后续工作中, 作者将结合特征可视化分析, 进一步发现本文方法的不足并做出改进. 此外, 结合本文提出的立方体卷积运算, 将更多主流的二维卷积神经网络框架用于三维点云分割任务也是我们下一步的工作.

References

1 Anguelov D, Taskarf B, Chatalbashev V, Koller D, Gupta D, Heitz G, et al. Discriminative learning of Markov random fields for segmentation of 3D scan data. In: Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Diego, USA: IEEE, 2005. 169–176

2 Wu Z R, Song S R, Khosla A, Yu F, Zhang L G, Tang X O, et al. 3D ShapeNets: A deep representation for volumetric shapes. In: Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, USA: IEEE, 2015. 1912–1920

3 Maturana D, Scherer S. VoxNet: A 3D convolutional neural network for real-time object recognition. In: Proceedings of the 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems. Hamburg, Germany: IEEE, 2015. 922–928

4 Riegler G, Ulusoys A, Geiger A. OctNet: Learning deep 3D representations at high resolutions. In: Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE, 2017. 6620–6629

5 Wang P S, Liu Y, Guo Y X, Sun C Y, Tong X. O-CNN: Octree-based convolutional neural networks for 3D shape analysis. *ACM Transactions on Graphics*, 2017, **36**(4): Article No. 72

6 Klokov R, Lempitsky V. Escape from cell: Deep kd-networks for the recognition of 3D point cloud models. In: Proceedings of the 2017 International Conference on Computer Vision. Venice, Italy: IEEE, 2017. 863–872

7 Qi C R, Su H, Mo K C, Guibas L J. Pointnet: Deep learning on point sets for 3D classification and segmentation. In: Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE, 2017. 77–85

8 Qi C R, Yi L, Su H, Guibas L J. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. In: Proceedings of the 2017 Advances in Neural Information Processing Systems. Long Beach, USA: Curran Associates, 2017. 5100–5109

9 Huang Q G, Wang W Y, Neumann U. Recurrent slice networks for 3D segmentation of point clouds. In: Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE, 2018. 262–2635

10 Li Y Y, Bu R, Sun M C, Wu W, Di X H, Chen B Q. PointCNN: Convolution on $\mathcal{X}$-transformed points. In: Proceedings of the 2018 Advances in Neural Information Processing Systems. Montreal, Canada: Curran Associates, 2018. 828–838

11 Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation. In: Proceedings of the 18th International Conference on Medical Image Computing and Computer-Assisted Intervention. Munich, Germany: MICCAI, 2015. 234–241

12 He K M, Zhang X Y, Ren S Q, Sun J. Deep residual learning for image recognition. In: Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE, 2016. 770–778

13 Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the inception architecture for computer vision. In: Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE, 2016. 2818–2826

14 Huang G, Liu Z, Maaten L V D, Weinberger K Q. Densely connected convolutional networks. In: Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE, 2017. 2261–2269

15 Lin T Y, Dollar P, Girshick R, He K M, Hariharan B, Belongie S. Feature pyramid networks for object detection. In: Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE, 2017. 936–944

16 Girshick R, Donahue J, Darrell T, Malik J. Rich feature hierarchies for accurate object detection and semantic segmentation. In: Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus, USA: IEEE, 2014. 580–587

17 Girshick R. Fast R-CNN. In: Proceedings of the 2015 IEEE International Conference on Computer Vision. Santiago, Chile: IEEE, 2015. 1440–1448

18 Ren S Q, He K M, Girshick R, Sun J. Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, **39**(6): 1137–1149

19 He K M, Gkioxari G, Dollar P, Girshick R. Mask R-CNN. In: Proceedings of the 2017 IEEE International Conference on Computer Vision. Venice, Italy: IEEE, 2017. 2980–2988

20 Liu S, Qi L, Qin H F, Shi J P, Jia J. Path aggregation network for instance segmentation. In: Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE, 2018. 8759–8768

21 Armeni I, Sener O, Zamir A R, Jiang H L, Brilakis L, Fischer M, Savarese S. 3D semantic parsing of large-scale indoor spaces. In: Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE, 2016.

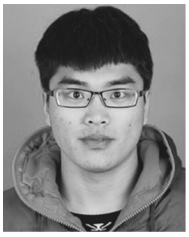
1534−1543

22 Dai A, Chang A X, Savva M, Halber M, Funkhouser T, Nießner M. ScanNet: Richly-annotated 3D reconstructions of indoor scenes. In: Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE, 2017. 2432−2443

彭秀平 燕山大学信息科学与工程学院副教授. 主要研究方向为扩频序列设计, 组合设计编码, 智能信息处理.
E-mail: pengxp@ysu.edu.cn
(**PENG Xiu-Ping** Associate professor at the School of Information Science and Engineering, Yanshan University. Her research interest covers spread spectrum sequence design, combination design coding, and intelligent information processing.)

仝其胜 燕山大学信息科学与工程学院硕士研究生. 主要研究方向为计算机视觉, 点云感知.
E-mail: tsisen@outlook.com
(**TONG Qi-Sheng** Master student at the School of Information Science and Engineering, Yanshan University. His research interest covers computer vision and point cloud perception.)

林洪彬 燕山大学电气工程学院副教授. 主要研究方向为点云处理, 模式识别与计算机视觉. 本文通信作者.
E-mail: honphin@ysu.edu.cn
(**LIN Hong-Bin** Associate professor at the School of Electrical Engineering, Yanshan University. His research interest covers point cloud processing, pattern recognition, and computer vision. Corresponding author of this paper.)



冯 超 燕山大学信息科学与工程学院硕士研究生. 主要研究方向为计算机视觉, 同步定位与建图.
E-mail: chaofenggo@163.com
(**FENG Chao** Master student at the School of Information Science and Engineering, Yanshan University. His research interest covers computer vision and simultaneous localization and mapping (SLAM).)

郑 武 燕山大学信息科学与工程学院硕士研究生. 主要研究方向为计算机视觉, 点云感知.
E-mail: zweducn@163.com
(**ZHENG Wu** Master student at the School of Information Science and Engineering, Yanshan University. His research interest covers computer vision and point cloud perception.)