光学 精密工程
OPTICS AND PRECISION ENGINEERING
核心期刊
主管：中国科学院
主办：中国科学院长春光学精密机械与物理研究所
中国仪器仪表学会
主编：曹健林

采用深度级联卷积神经网络的三维点云识别与分割

杨军，党吉圣

引用本文:
杨军, 党吉圣. 采用深度级联卷积神经网络的三维点云识别与分割[J]. 光学 精密工程, 2020, 28(5): 1187-1199.
YANG Jun, DANG Ji-sheng. Recognition and segmentation of three-dimensional point cloud based on deep cascade convolutional neural network[J]. *Optics and Precision Engineering*, 2020, 28(5): 1187-1199.

在线阅读 View online: https://doi.org/10.3788/OPE.20202805.1187

文章编号 1004-924X(2020)05-1187-13

采用深度级联卷积神经网络的三维点云识别与分割

杨 军*,党吉圣
(兰州交通大学 电子与信息工程学院,甘肃 兰州 730070)

摘要:三维目标识别和模型语义分割在自动驾驶、机器人导航、3D 打印和智能交通等领域均有着广泛应用。针对 PointNet++未能结合三维模型的上下文几何结构信息的问题,提出一种采用深度级联卷积神经网络的三维点云识别与分割方法。首先,通过构建深度动态图卷积神经网络捕捉点云的深层语义几何特征;其次,通过将深度动态图卷积神经网络作为深度级联卷积神经网络的子网络递归地应用于输入点集的嵌套分区,以充分挖掘三维模型的深层细粒度几何特征;最后,针对点集特征学习中的点云采样不均匀问题,构建一种密度自适应层,利用循环神经网络编码每个采样点的多尺度邻域特征以捕捉上下文细粒度几何特征。实验结果表明,本算法在三维目标识别数据集 ModelNet40 和 MoelNet10 上的识别准确率分别为 91.9%和 94.3%,在语义分割数据集 ShapeNet Part,S3DIS 和 vKITTI 上的平均交并比分别为 85.6%,58.3%和 38.6%。该算法能够提高三维点云目标识别和模型语义分割的准确率,且具有较高的鲁棒性。

关 键 词:三维点云;目标识别;语义分割;卷积神经网络;循环神经网络
中图分类号:TP391 **文献标识码**:A **doi**:10.3788/OPE.20202805.1187

Recognition and segmentation of three-dimensional point cloud based on deep cascade convolutional neural network

YANG Jun*, DANG Ji-sheng

(*School of Electronic and Information Engineering, Lanzhou Jiaotong University, Lanzhou* 730070, *China*)
* *Corresponding author, E-mail: yangj@mail.lzjtu.cn*

Abstract: Three-dimensional (3D) object recognition and model semantic segmentation are widely appliedin fields such as automatic driving, robot navigation, 3D printing, and intelligent transportation. With a focuson the inability of PointNet++ to integrate contextual geometric structure information, a method for recognition and segmentation of 3D point cloud modes based on a deep cascade Convolutional Neural Network (CNN) was proposed herein. The deep semantic geometric features of the point cloud could be captured via construction of a deep dynamic graph CNN. Subsequently, the deep dynamic graph CNN was applied recursively as a subnetwork of a deep cascade CNN for nested partition of the input point set for full exploration of the fine-grained geometric features of the 3D model. Finally, to address the point cloud sampling nonuniformity problem in point set feature learning, a density adaptive layer was constructed. A recurrent neural network was used to

收稿日期:2019-12-02;**修订日期**:2020-03-13.
基金项目:国家自然科学基金资助项目(No. 61862039)

encode the multiscale neighborhood features of each sample point to capture the contextual fine-grained geometric features. The experimental results showed that the recognition accuracy of this algorithm on ModelNet40 and ModelNet10 were 91.9% and 94.3%, respectively. The mean intersection-over-union on the ShapeNet Part, S3DIS, and vKITTI datasets was 85.6%, 58.3%, and 38.6%, respectively. This algorithm can improve the accuracy of 3D point cloud recognition and model semantic segmentation, and it shows high robustness.

Key words: three-dimensional (3D) point cloud; object recognition; semantic segmentation; convolutional neural network; recurrent neural network

1 引 言

随着三维建模技术以及深度传感器的广泛应用,三维模型的数量呈现出爆炸式增长,三维模型的目标识别和语义分割作为三维模型分析处理的前提和基础,已成为机器视觉领域的一个重要研究课题。三维目标识别和模型语义分割是通过比较各模型特征描述符之间的相似性和差异性来完成的,因此其关键问题是如何提取准确而鲁棒的三维特征描述符。传统方法利用手工设计形状描述符来提取三维模型的特征,如几何形状描述符[1]和热核签名描述符[2]等,但是手工设计的特征描述符良莠不齐,严重依赖专家经验,而且泛化能力较差。

近年来,深度学习[3-6]方法在机器视觉领域取得了一定的阶段性成果,越来越多的学者开始尝试采用深度学习方法来进行三维目标识别和模型语义分割,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。

基于多视图的方法。由于三维点云的不规则性,直接从三维点云数据中提取特征有一定的困难。文献[7]首先对三维模型进行多方位渲染得到二维投影视图,然后把二维多视图作为训练数据输入到经典的 VGG (Visual Geometry Group)[8]中训练并提取特征,最后通过视图池化层把视图特征聚合得到一维的全局特征描述符。该方法虽提高了三维模型识别的准确率,但存在视图特征冗余和三维模型几何信息丢失的问题。

基于体素的方法。文献[9]提出把不规则的点云数据规则化为 3D 体素网格的形式,然后使用三维卷积神经网络直接作用于 3D 体素数据提取特征描述符。文献[10]将点云数据转化为二值 3D 体素矩阵,通过附加正则化项的随机梯度下降算法提取体素矩阵的特征,以此对模型类别进行预测。文献[11]把不规则的点云数据体素化为规则的体素数据并进行旋转扩充以增强网络的泛化能力,并通过堆叠小卷积核构建深度卷积神经网络挖掘模型内部隐含信息,提取体素矩阵深层特征。上述算法虽然有效保留了模型的几何结构信息,但是体素化操作内存消耗严重,使捕获高分辨率信息和细粒度特征变得困难。由于对于低分辨率的模型识别精度不高,文献[12]提出了空间划分方法,但仍然缺乏捕捉局部几何特征的能力。

基于点云表示的方法。该方法可直接利用矩阵运算对点云模型进行仿射变换,避免了把点云转化为其他规则数据形式的繁杂操作,已广泛应用于计算机图形学和机器视觉领域,如室内导航[13]、自动驾驶[14]、机器人[15]以及车载激光雷达[16]等。对于三维目标识别和语义分割,文献[17]提出的 PointNet 网络模型成为把深度学习框架直接作用于原始点云数据的先驱,但 PointNet 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。文献[18]提出了 PointNet++网络,通过划分局部点云分层提取细粒度特征信息,对三维点云模型识别和语义分割展现出良好的性能。该网络虽然有效捕获了点云局部邻域信息,但是没有考虑局部邻域内点与点之间的距离度量,缺乏捕捉上下文细粒度局部几何信息的能力,导致识别效果不佳。为此,本文提出了基于深度级联卷积神经网络(Deep Cascade Convolutional Neural Network, DCCNN)的三维点云识别与分割方法,能够有效捕捉点云模型的上下文深层细粒度局部几何特征,提高了三维目标识别和模型语义分割的精度。主要创新点和贡献有:(1)通过在 DGCNN[19](Dynamic Graph Convolutional Neural Network)中引入残差学习加深网络深度,构建深

度动态图卷积神经网络以充分挖掘点云的深层语义几何特征。(2)构建深度级联卷积神经网络。将深度动态图卷积神经网络作为 PointNet++[18]的子网络递归地应用于输入点集的嵌套分区以提取点云模型的深层细粒度几何特征。(3)针对点云的采样密度不均匀导致的网络学习性能下降的问题，提出一种多尺度分组循环神经网络(Multi Scale Grouping-Recurrent Neural Network，MSG-RNN)编码策略。通过编码采样点的不同尺度的邻域几何特征，来提取采样点的上下文细粒度几何特征以增强网络的鲁棒性。

2　深度级联卷积神经网络

2.1　深度动态图卷积神经网络

为了捕捉三维点云的局部几何特征，DGCNN[19]通过度量相邻点之间的距离关系，提出了边缘卷积层(Edge Convolution，EdgeConv)操作，一定程度上提高了网络识别性能，但网络深度较浅，无法捕捉更抽象的深层语义特征信息。受文献[20]启发，本文在 DGCNN 的基础上构建深度动态图卷积神经网络(Deep Dynamic Graph Convolutional Neural Network，DDGCNN)，以充分挖掘点云的深层语义几何特征，网络结构如图 1 所示。DDGCNN 由 6 个 EdgeConv 层、1 个 MLP 层和 1 个最大池化层构成，EdgeConv 层结构如图 1 下方子图所示。DDGCNN 的输入为特征维度为 F 的 $k+1$ 个点构成的局部点云 $\boldsymbol{X}=\{x_1,x_2,\ldots,x_{(k+1)}\mid x_{(k+1)}\in\mathbb{R}^g\}$，采用 7 个卷积层把点云中的每个点的原始特征映射到高维特征空间，卷积层的各层参数如表 1 所示。本网络把前层动态图的低级特征连接到后层动态图的高级特征中，避免了梯度消失问题的同时，加深了网络深度，有助于提取更具有代表性的深层语义特征信息。DDGCNN 与 DGCNN[19]的不同之处在于：(1)通过残差学习[21]将来自不同动态图的不同层次的特征相互连接，避免了梯度消失问题。(2)增加了卷积层的数目，以充分挖掘深层语义几何特征。(3)去除了空间转换网络，减少了网络参数，降低了过拟合风险。

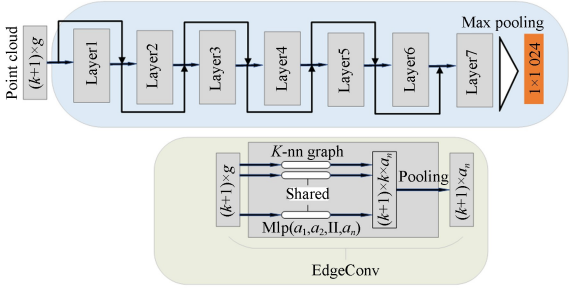


图 1　深度动态图卷积神经网络结构

Fig. 1　Network structure of deep dynamic graph convolutional neural network

表 1　卷积层各层参数

Tab. 1　Parameters of each convolution layer

Convolution layer	Convolution type	Input channels	Output channels	Convolution kernel size	Step	Batch normalization	Activation function
Layer1	EdgeConv	3	64	3×1	1	是	Selu
Layer2	EdgeConv	67	64	1×1	1	是	Selu
Layer3	EdgeConv	128	64	1×1	1	是	Selu
Layer4	EdgeConv	128	64	1×1	1	是	Selu
Layer5	EdgeConv	128	128	1×1	1	是	Selu
Layer6	EdgeConv	192	128	1×1	1	是	Selu
Layer7	MLP	256	1 024	1×1	1	否	Selu

2.2　深度级联卷积神经网络

在 PointNet++[18]网络中，集合抽象层中采用 PointNet 提取分组层的局部特征，然而，PointNet 缺乏捕捉局部几何结构信息的能力。本文将 DDGCNN 作为 PointNet++的子网络以构建深度级联卷积神经网络(Deep Cascade

Convolutional Neural Network, DCCNN),该网络包含了 3 个集合抽象层,网络结构如图 2 所示。网络的输入是大小为 $N\times(C+d)$的点云矩阵,其中 N 为点的数目,d 为点的 x,y,z 3 个坐标维度,C 为点的特征维度。第 1 个集合抽象层首先对整个输入点云采用迭代最远点采样算法采样 N_1 个点,对每个采样点采用 k 最近邻算法搜索距离采样点最近的 k 个点构建每个采样点的 k 邻域分组,即得到大小为 $N_1\times(k+1)\times(C+d)$的点云矩阵,然后采用 DDGCNN 提取每个分组的深层语义几何特征,得到 N_1 个特征维度为 C_1 的点构成的新点云,再次输入第 2 个集合抽象层经过采样分组得到大小为 $N_2\times(k+1)\times(C_1+d)$)的点云矩阵,采用 DDGCNN 提取特征后得到大小为 $N_2\times(C_2+d)$的点云矩阵。对于分类(Classification)任务,将该点云矩阵输入第 3 个集合抽象层,以此递归抽象整个点云,得到能表示整个点云的一维特征向量 C_3。然后采用 3 个全连接层 MLP(512,256,R)对全局特征向量进行降维转换,最后采用 Softmax 分类器计算分类分数。对于分割(Segmentation)任务,为了获取每个点的点级别的特征,在网络中引入两个插值层[18],通过上采样将特征从形状级别传播到点级别,并采用 MLP 和 Selu 促进点特征的提取,最后网络输出每个点的预测标签。

本文采用三维空间中点与点之间的欧氏距离来实现特征传播,由点 o 与其 k 最近邻点 o_i 的欧几里得距离插值而成。计算公式如式(1)所示:

$$\chi(o)=\sum_{i=1}^{k}u(o_i)\chi(o_i)/\sum_{i=1}^{k}u(o_i),\tag{1}$$

其中:

$$u(o_i)=1/(o-o_i)^2.\tag{2}$$

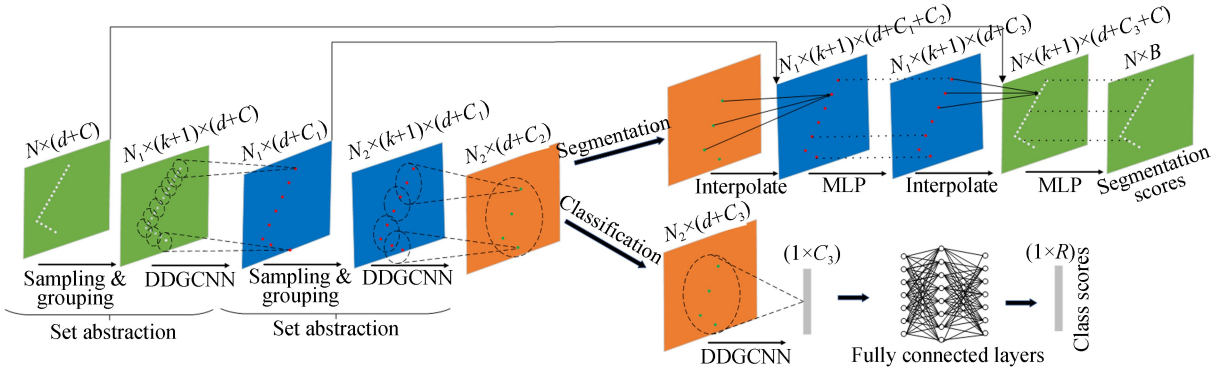


图 2 深度级联卷积神经网络结构

Fig. 2 Structure of deep cascade convolutional neural network

2.3 密度自适应层

现实生活中,在 3D 扫描仪生成点云数据时,由于透视效应、径向密度变化等因素的干扰,采集到的点云的密度在不同区域往往是不均匀的,这种不均匀性增加了点集特征学习的难度。本文构建的 DCCNN 在采样分组时是采用单尺度分组(Single Scale Grouping, SSG),在密度均匀的点云数据集上表现良好,而对于密度不均匀的采样点集的特征学习效果并不理想。为此,本文构建多尺度分组循环神经网络(Multi Scale Grouping-Recurrent Neural Network, MSG-RNN)编码策略,在输入点集密度不均匀时能够自动结合每个采样点的多个不同尺度的上下文邻域特征以增强网络的鲁棒性。本文将采用 MSG-RNN 编码策略的 DCCNN 命名为上下文深度级联卷积神经网络(Contextual-Deep Cascade Convolutional Neural Network, C-DCCNN)。

MSG-RNN 示意图如图 3 所示,通过设置不同的邻域点数目 k 以获得每个采样点的多个不同尺度的 k 邻域分组,并采用 DDGCNN 提取每个分组的几何特征向量,然后把每个采样点的所有邻域的几何特征向量组成一个特征向量序列 $\boldsymbol{S}_k=\{\boldsymbol{s}_k^1,\boldsymbol{s}_k^2,\ldots,\boldsymbol{s}_k^t,\ldots,\boldsymbol{s}_k^M\}$,其中,$\boldsymbol{s}_k^t$ 表示采样点的第 t 个邻域的几何特征向量,M 为邻域的个数。把采样点的几何特征序列 $\boldsymbol{S}_k=\{\boldsymbol{s}_k^1,\boldsymbol{s}_k^2,\ldots,\boldsymbol{s}_k^t,\ldots,\boldsymbol{s}_k^M\}$输入 RNN 编码器,用一个隐藏层依次编码采样点的不同尺度的邻域特征向量来学习上下文高级几何特征。原因在于 MLP 和

EdgeConv 倾向于捕捉点云中级几何特征，而 RNN 对高级几何特征更为敏感。RNN 编码器由一个隐藏层 $\boldsymbol{h}$ 和一个输出层 $\boldsymbol{v}$ 组成，当 RNN 编码采样点的每一个邻域几何特征时，RNN 编码器的隐藏层状态 $\boldsymbol{h}_t$ 都要被更新，如公式(3)所示：

$$\boldsymbol{h}_t = f(\boldsymbol{h}_{t-1}, \boldsymbol{s}_k^t), \tag{3}$$

其中：f 为一个非线性激活函数，实验中采用 LSTM 单元。$\boldsymbol{h}_{t-1}$ 为编码上一个邻域的几何特征时的隐藏层状态。在 RNN 编码采样点的第 t 个邻域的特征向量时，编码器的输出 $\boldsymbol{v}_t$ 如公式(4)所示：

$$\boldsymbol{v}_t = \boldsymbol{W}_a \boldsymbol{h}_t, \tag{4}$$

式中 $\boldsymbol{W}_a$ 是一个可学习的权重矩阵。当 RNN 编码采样点的最后一个邻域特征向量 $\boldsymbol{s}_k^T$ 时，网络已学习完全部输入特征向量，得到编码器隐藏层最后一个状态 $\boldsymbol{h}_T$，$\boldsymbol{h}_T$ 和 $\boldsymbol{W}_a$ 相乘得到采样点的上下文高层几何特征 $\boldsymbol{v}_T$，其包含了采样点的整个特征序列的上下文高层细粒度几何信息。密度自适应层 MSG-RNN 编码策略通过把 DDGCNN 提取到的采样点不同尺度的邻域特征向量依次输入 RNN 进行编码，可以获取采样点的不同邻域之间的上下文隐含关联信息，不仅解决了密度不均匀点集特征学习困难的问题，而且增强了对于密度均匀点集特征学习的能力。

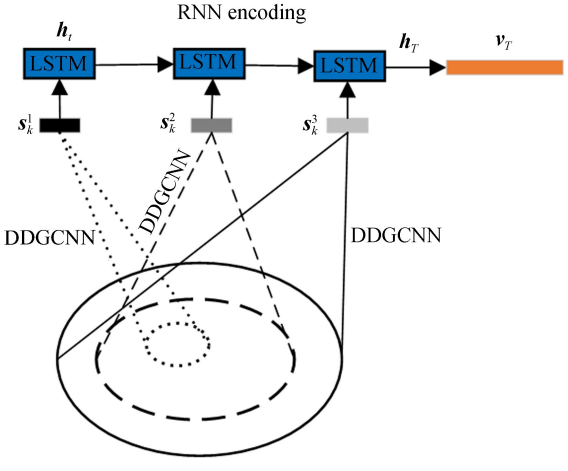


图 3　MSG-RNN 策略

Fig. 3　MSG-RNN strategy

3　实验结果与分析

3.1　实验数据集

对于三维目标识别任务，选择 ModelNet40[22] 和 ModelNet10[22] 两个标准数据集进行实验。ModelNet40 共有 40 个类别的 12 311 个 CAD 模型，其中 9 843 个模型用于网络训练，2 468 个模型用于网络测试。ModelNet10 共有 10 个类别的 4 899 个 CAD 模型，3 991 个用于训练，908 个用于测试。对于三维模型语义分割任务，分别采用部件语义分割数据集 ShapeNet Part[23]、室内场景语义分割数据集 S3DIS[24] 和户外自动驾驶场景语义分割数据集 vKITTI[27] 进行实验。ShapeNet Part 数据集包含 16 个类别的 16 881 个 CAD 模型，共有 50 个部件语义标签。S3DIS 是一个室内大规模点云数据集，包含 6 个室内区域，共 272 个房间，其中所有点标注为木板(Board)、书柜(Bookcase)、椅子(Chair)、天花板(Ceiling)和横梁(Beam)等 13 个语义类别。vKITTI 是一个自动驾驶实际场景的户外大规模点云数据集，分为 6 个不同的城市场景，其中所有点标注为自动驾驶场景中的汽车(Car)、树木(Tree)、建筑物(Building)、马路(Road)、交通灯(Traffic Light)和行人(Pole)等 13 个语义类别。

3.2　参数设置

实验采用基于动量的随机梯度下降(Stochastic Gradient Descent，SGD)优化算法，动量因子为 0.9，初始学习率为 0.001，学习率衰减指数为 0.7，衰减速度为 200 000。采用 Adam 算法来更新 SGD 的步长，网络参数初始化采用 Xavier 优化器，批处理归一化的衰减率初始值为 0.5，最终值为 0.99。激活函数采用 Selu 以缓解梯度消失，增加网络非线性拟合能力。为了防止过拟合，在全连接层采用 Dropout_Selu 函数[26]，除最后一层外，所有层都包含批处理规范化。

3.3　三维目标识别实验结果分析

为了探究本文构建的 DDGCNN 和 C-DCCNN 的有效性，分别在 ModelNet40 数据集上对 DGCNN(BASELINE)[19]，DDGCNN 和 C-DCCNN 三个网络进行训练并测试，实验结果如表 2 所示。DGCNN(BASELINE)为去除空间转换网络的 DGCNN。可以看出，在 DGCNN(BASELINE)中引入残差学习构建的 DDGCNN 的识别准确率比 DGCNN(BASELINE)提高了 0.2%，验证了 DDGCNN 能够有效捕获深层语义几何特征的能力。在 PointNet++网络中嵌入 DDGCNN 构建的 C-DCCNN 的识别准确率比

DDGCNN 高出 0.5%，因为 C-DCCNN 采用分层特征学习策略能够捕捉细粒度局部几何特征，同时 MSG-RNN 在编码多尺度特征向量时可以有效结合上下文信息。

表 2 不同算法的三维模型识别准确率比较

Tab. 2 Comparison of the accuracy of 3D models recognition among different algorithms (%)

Algorithm	Accuracy
DGCNN(BASELINE)[19]	91.2
Ours(DDGCNN)	91.4
Ours(C-DCCNN)	91.9

为了验证本文算法的优越性，在 ModelNet40 和 ModelNet10 数据集上分别与其他先进方法进行了对比实验，结果如表 3 所示。可以看出，本文算法的识别准确率明显优于其他主流算法。原因在于本文算法通过构建 DDGCNN 能够有效提取点云模型的深层语义几何特征，并采用分层特征学习策略充分挖掘了三维模型的上下文细粒度深层几何特征。此外，表 4 比较了本文算法与 PointNet 算法在 ModelNet40 数据集上各类别模型的识别准确率。对于测试集中的 40 类点云模型，其中有 27 类本文算法的识别准确率高于 PointNet 算法，有 11 类本文算法与 PointNet 算法的识别准确率相同，只有 2 类本文算法的识别准确率低于 PointNet 算法，充分证明了本文算法的优越性。从表中还可以看出，本文算法以及 PointNet 算法对花盆(Flower pot)这一类别的模型识别准确率最低，而且远低于其他类模型，原因在于花盆(Flower pot)类部分模型只包含花盆(Flower pot)，而部分模型同时包含了花盆(Flower pot)和植物(Plant)，因此与植物(Plant)类造成了混淆，所以难以识别。图 4 给出了在 ModelNet40 测试集上测试得到的几种典型的误分类模型实例。图中从第 1 列到第 4 列分别为真实值、预测值、真实值与预测值的共同结构、标签信息。可以看出，错误预测的模型和真实的模型之间均具有相同的局部结构。例如在图 4 第 1 行中，真实的标签是花盆(Flower pot)，而本文算法预测为花瓶(Vase)，预测错误的原因在于它们的共同局部结构瓶嘴。在图 4 第 2 行中，真实的标签是花盆(Flower pot)，而本文算法预测为植物(Plant)，造成预测错误的原因在于花盆(Flower pot)类部分模型里有植物(Plant)。所以，本文算法对于如何排除干扰的局部特征，只关注显著结构特征，还需要进一步提高。

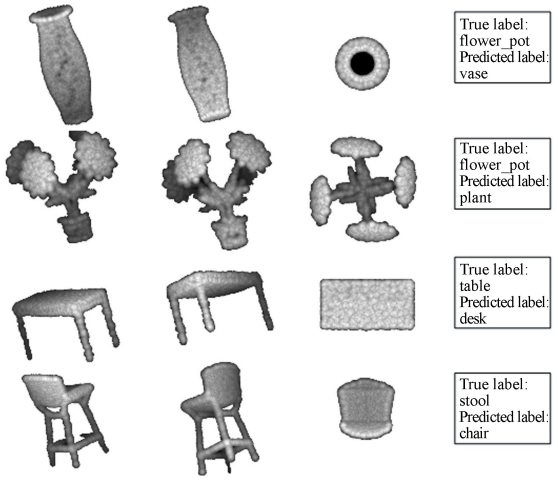


图 4 错误预测的点云模型实例

Fig. 4 Examples of mispredicted point cloud models

表 3 不同方法的三维模型识别准确率比较

Tab. 3 Comparison of recognition accuracy rate of 3D models using different methods (%)

Method	Input	Model Net40	Model Net10
3D ShapeNets[22]	Voxel	84.7	83.5
VoxNet[9]	Voxel	85.9	92.0
MVCNN[7]	Images	90.1	—
PointNet[17]	Point cloud	89.2	—
PointNet++[18]	Point cloud	90.2	—
Kd-Net[12]	Point cloud	90.6	94.0
DGCNN(BASELINE)[19]	Point cloud	91.2	—
Ours	Point cloud	91.9	94.3

图 5 和图 6 分别给出了本文算法在 ModelNet40 数据集上模型识别准确率、训练误差与迭代次数的统计结果，其中，横坐标均为训练迭代次数，图 5 纵坐标为识别准确率，图 6 纵坐标为训练误差(彩图见期刊电子版)。阴影线表示原始迭代数据，橙色曲线表示经过平滑后的迭代结果。由图可见，在训练初期，随着迭代次数的增加，识别准确率逐渐提高，训练误差呈下降趋势，因为网络训练过程中不断优化参数，由卷积层学习到的特

征对数据集中模型的描述准确度不断提高。当迭代次数达到 40 000 次时,识别准确率和训练误差趋于稳定,网络趋于收敛,说明网络中的参数已达到最优。图 5 和图 6 充分验证了本文网络具有在训练过程中能够不断提取三维模型的有效特征的能力。

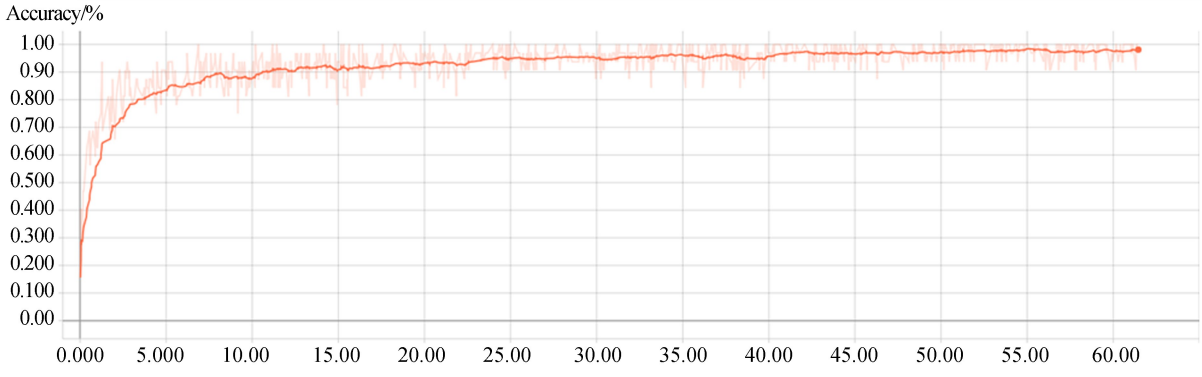


图 5　迭代次数与模型识别准确率的统计结果

Fig. 5　Statistical results of iteration times and model recognition accuracy

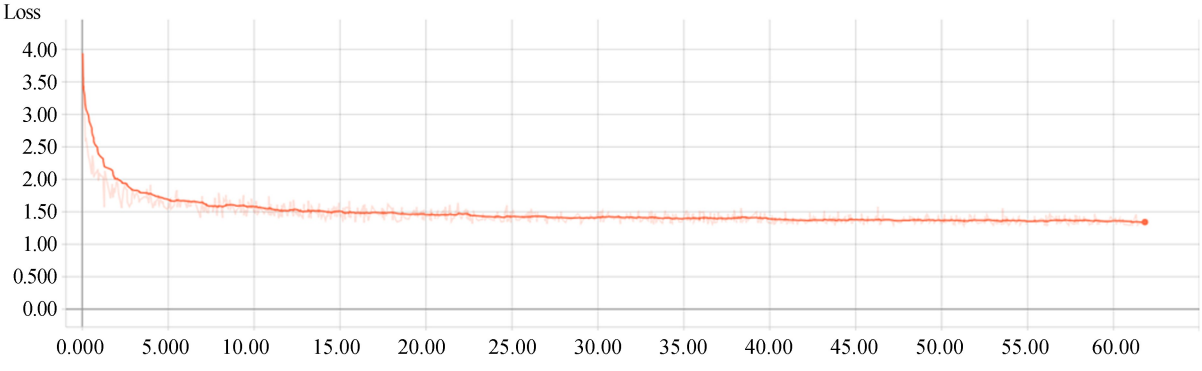


图 6　迭代次数与训练误差的统计结果

Fig. 6　Statistical results of iteration times and training error

表 4　ModelNet**40** 数据集上每一类识别准确率的对比

Tab. 4　Comparison of per-class accuracy of object recognition on ModelNet40 dataset　(%)

Network	Airplane	Bathtub	Bed	Bench	Bookshelf	Bottle	Bowl	Car	Chair	Range_hood
PointNet	99.0	86.0	98.0	75.0	89.0	95.0	95.0	98.0	97.0	91.0
Ours	100	92.0	99.0	75.0	98.0	95.0	100	100	98.0	97.0
Network	Cup	Curtain	Desk	Door	Dresser	Flower pot	Glass_box	Guitar	Cone	Sink
PointNet	55.0	85.0	83.7	85.0	67.4	25.0	98.0	100	100	75.0
Ours	75.0	95.0	84.9	90.0	68.6	15.0	98.0	99.0	100	90.0
Network	Laptop	Mantel	Monitor	Night_stand	Person	Piano	Plant	Radio	Keyboard	Wardrobe
PointNet	100	95.0	97.0	72.1	95.0	91.0	75.0	70.0	100	55.0
Ours	100	97.0	100.0	86.0	95.0	96.0	85.0	90.0	100	60.0
Network	Sofa	Stairs	Stool	Table	Tent	Toilet	Tv_stand	Vase	Lamp	Xbox
PointNet	97.0	85.0	70.0	85.0	95.0	97.0	84.0	79.0	90.0	80.0
Ours	98.0	95.0	70.0	89.0	95.0	98.0	89.0	83.0	90.0	80.0

与此同时，为了继续探究本算法对于稀疏点云的鲁棒性，采用不同密度的数据集进行实验。由于 ModelNet40 数据集中的三维模型都是密度均匀的，为了得到密度不均匀的数据集，对数据集中的三维点云模型做以下预处理：首先采用随机输入丢弃策略以随机概率对输入点进行随机丢弃，即对输入的点云模型，以 p（$p\leqslant 1$）的比例选择待丢弃点集，对于待丢弃点集中的每个点以概率 q 进行丢弃，为了避免空集，设置 $p=0.90$，以此得到具有不同密度的点云模型，如图 7 左侧所示。分别将训练好的网络模型在密度不同的数据集进行测试，实验结果如图 7 右侧所示。其中，DP 表示训练期间的输入点随机丢弃策略，SSG 为每层集合抽象层中使用单一尺度分组的 DCCNN 网络。可以看出，随着点数的减少，SSG 的识别准确率明显下降，原因在于 SSG 采用 DDGCNN 提取点云的局部深层几何特征，点数的减少破坏了局部几何结构。PointNet 在点数减少时网络稳健性强于 SSG，因为它专注于全局特征而不是精细局部细节，然而点数的减少也使其识别准确率明显下降。PointNet＋DP（在训练期间采用输入点随机丢弃策略的 PointNet）网络鲁棒性明显优于 PointNet，因为在训练期间随机输入丢弃策略可以增强网络学习稀疏点云特征的能力。SSG＋DP（在训练期间采用输入点随机丢弃策略的 SSG）在测试期间点数从 1 024 减少到 256 时，识别准确率下降不到 3%，原因在于随机输入丢弃策略增强了网络的鲁棒性，但随着点数减少到 128 时识别准确率明显下降。本文提出的密度自适应层 MSG-RNN＋DP（在训练期间采用输入点随机丢弃策略和多尺度分组 RNN 编码策略）对于点云密度变化非常稳健，从 1 024 个点减少到 256 个点时，MSG-RNN＋DP 的识别准确率下降不到 1%。与其他方法相比，MSG-RNN＋DP 几乎在所有点云采样密度上都实现了最佳性能，展现了最好的鲁棒性。

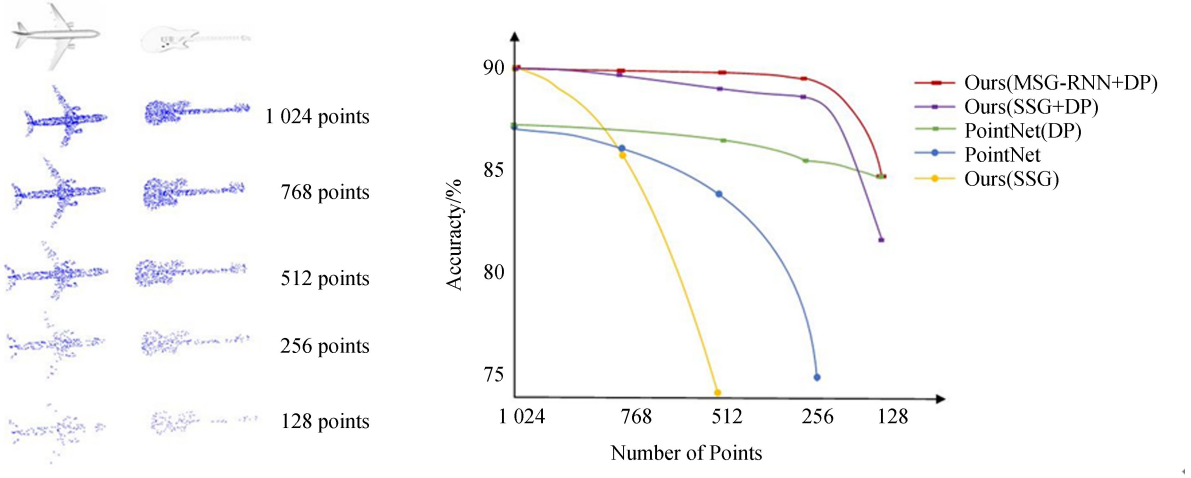


图 7 点云密度变化时不同网络的识别准确率的统计结果

Fig. 7 Statistical results of recognition accuracy for different networks when point cloud density changes

3.4 三维模型语义分割实验结果分析

与三维目标识别相比，三维模型语义分割需要更精细地识别每个点的语义类别，所以是一项更具挑战性的任务。为了进一步分析本文算法处理三维点云细粒度任务的能力，在 ShapeNet Part 数据集上进行了语义分割实验，并与其他主流算法进行了对比，评价指标为文献[17]中采用的交并比（Intersection-over-Union，IoU），实验结果如表 5 和表 6 所示。

表 5 不同算法在 ShapeNet Part 数据集上平均交并比的比较

Tab. 5 Comparison of mIoU of different algorithms on ShapeNet Part dataset (%)

Algorithms	Accuracy
Kd-Net[12]	82.3
PointNet[17]	83.7
PointNet＋＋[18]	85.1
DGCNN [19]	85.1
Ours	85.6

表 6　不同算法在 ShapeNet Part 数据集上的各类别的交并比的比较

Tab. 6　Comparison of IoU of each category of different algorithms on ShapeNet Part dataset　　(%)

Algorithm	Airplane	Bag	Cap	Car	Chair	Earphone	Guitar	Knife
Kd-Net[12]	80.1	74.6	74.3	70.3	88.6	73.5	90.2	87.2
PointNet[17]	83.4	78.7	82.5	74.9	89.6	73.0	91.5	85.9
PointNet++[18]	82.4	79.0	87.7	77.3	90.8	71.8	91.0	85.9
DGCNN[19]	84.2	83.7	84.4	77.1	90.9	78.5	91.5	87.3
Ours	84.7	86.7	84.4	78.8	91.6	76.4	92.0	88.4
Algorithm	Lamp	Laptop	Motorbike	Mug	Pistol	Rocket	Skateboard	Table
Kd-Net[12]	81.0	94.9	57.4	86.7	78.1	51.8	69.9	80.3
PointNet[17]	80.8	95.3	65.2	93.0	81.2	57.9	72.8	80.6
PN++[18]	83.7	95.3	71.6	94.1	81.3	58.7	76.4	82.6
DGCNN[19]	82.9	96.0	67.8	93.3	82.6	59.7	75.5	82.0
Ours	83.3	97.1	66.9	95.4	82.4	60.4	76.9	81.1

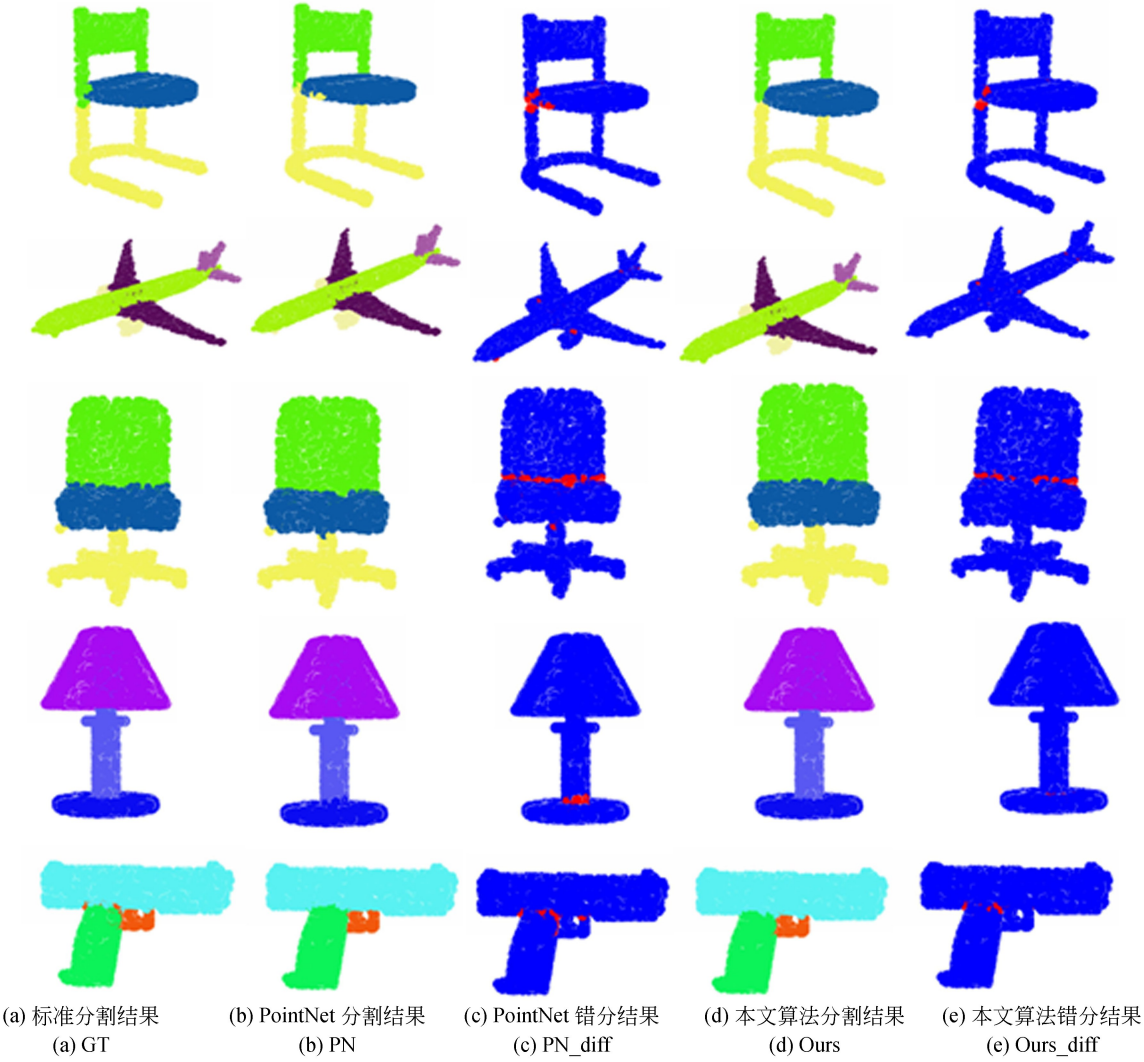


图 8　ShapeNet Part 数据集上语义分割模型可视化

Fig. 8　Visualization of semantic segmentation models on ShapeNet Part dataset

可以看出，本文算法以85.6%的 mIoU 获得了最好的语义分割性能。图 8 所示为 C-DCCNN 和 PointNet 在 ShapeNet Part 数据集上的语义分割可视化效果图，在第 3 列和第 5 列的错分结果可视化中，蓝色表示预测正确，红色表示预测错误（彩图见期刊电子版）。与 PointNet 相比，本算法的语义分割结果与标准分割结果高度一致，尤其细粒度细节处的分割准确率明显提升，如台灯（Lamp）柱身的底端、手枪（Pistol）的握柄处等，进一步验证了本文算法具有能够捕获点云深层细粒度几何特征的能力。

为了验证本文算法同样适用于大规模点云场景分析，在三维室内场景语义分割数据集 S3DIS 和户外自动驾驶实际场景的语义分割数据集 vKITTI 上分别对 C-DCCNN 进行了训练和测试，并与主流算法进行了对比，实验结果如表 7 和表 8 所示。可以看出，本文算法的分割准确率均优于其他主流算法。除了定量分析外，图 9 和图 10 分别展示了定性的语义分割模型可视化效果图。从图 9 中可以看出，C-DCCNN 能够纠正 PointNet 预测错误的点，获得更准确的分割结果，并且挖掘了 PointNet 所遗漏的细粒度细节信息。例如，椅子（Chair）的腿在很大程度上得到了保留，门（Door）的预测也比 PointNet 更准确。事实上，门（Door）和墙（Wall）在几何形状上极其相似，但是本文算法有效结合了门的上下文位置信息（门框的特征），可以更好地预测门（Door）这一类别，进一步证明了 MSG-RNN 编码策略能够有效结合上下文几何信息的能力。从图 10 中可以看出，本文算法整体分割错误率相比于 PointNet 有所减少，尤其对于马路（Road）和地带（Terrain）这两类语义的分割准确性提高最为明显。原因在于地带（Terrain）和马路（Road）在几何形状上极其相似，区别在于地带（Terrain）中有树木（Tree），马路（Road）中没有树木（Tree），单纯提取马路（Road）和地带（Terrain）的几何特征很难区分这两类语义，需结合其上下文信息。由此进一步验证了本文算法具有提取上下文细粒度局部几何特征的能力。然而，本文算法对于同时存在上下文信息车（Car）的地带（Terrain）和马路（Road）识别混淆，可见本文算法对邻域上下文信息缺乏自适应筛选能力。

表 7　S3DIS 数据集上不同算法的分割准确率对比

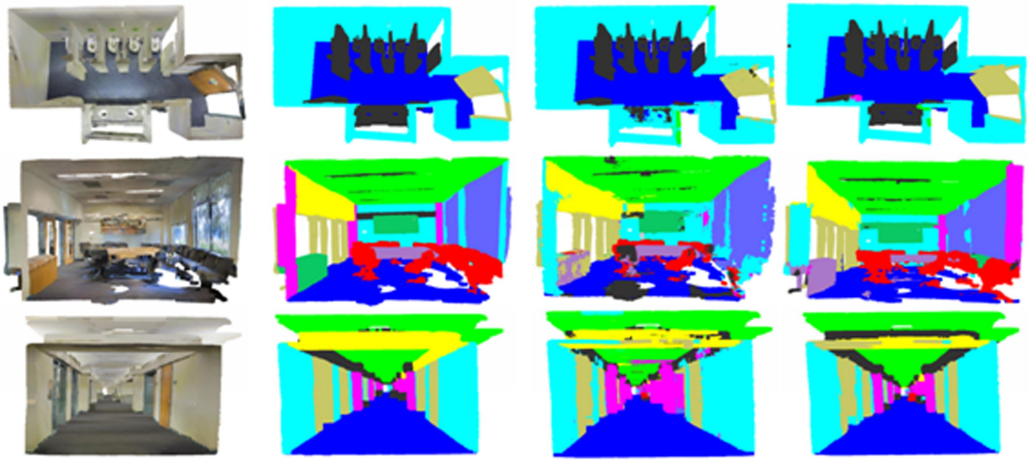
Tab. 7　Comparison of segmentation accuracy of different algorithms on S3DIS dataset　(%)

Algorithm	mIoU	OA
PointNet [17]	47.6	78.5
MS+CU [25]	47.8	79.2
G+RCU [25]	49.7	81.1
PointNet++ [18]	54.5	81.0
DGCNN [19]	56.1	84.1
Ours	58.3	86.0

表 8　vKITTI 数据集上不同算法的分割准确率对比

Tab. 8　Comparison of segmentation accuracy of different algorithms on vKITTI dataset　(%)

Algorithms	OA	mAcc	mIoU
PointNet [17]	79.7	47.0	34.4
G+RCU [25]	80.6	49.7	36.2
Ours	82.5	51.8	38.6



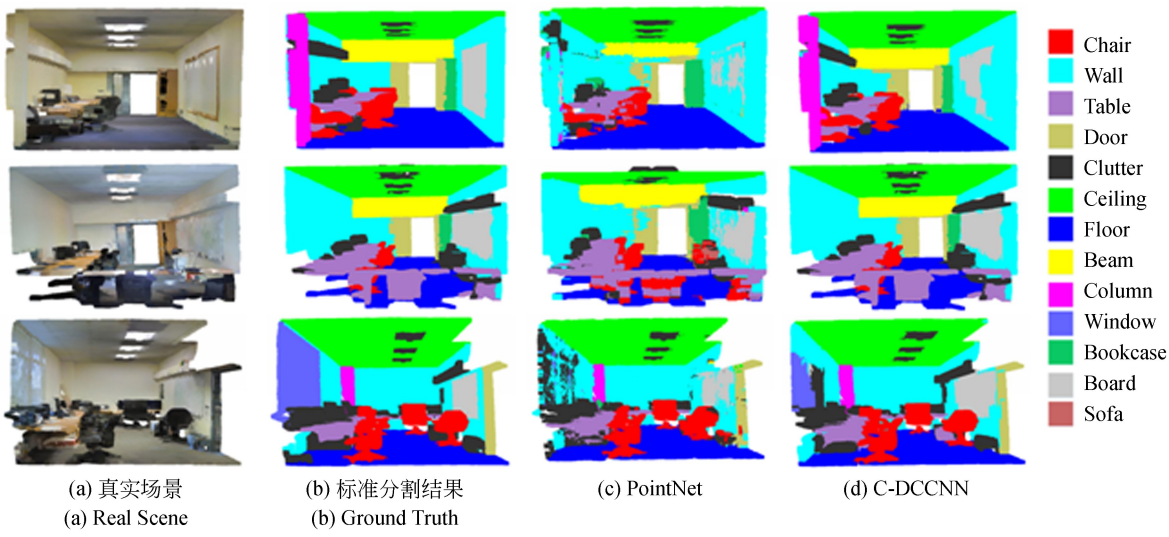


图 9　S3DIS 数据集上语义分割模型可视化

Fig. 9　Visualization of semantic segmentation models on S3DIS dataset

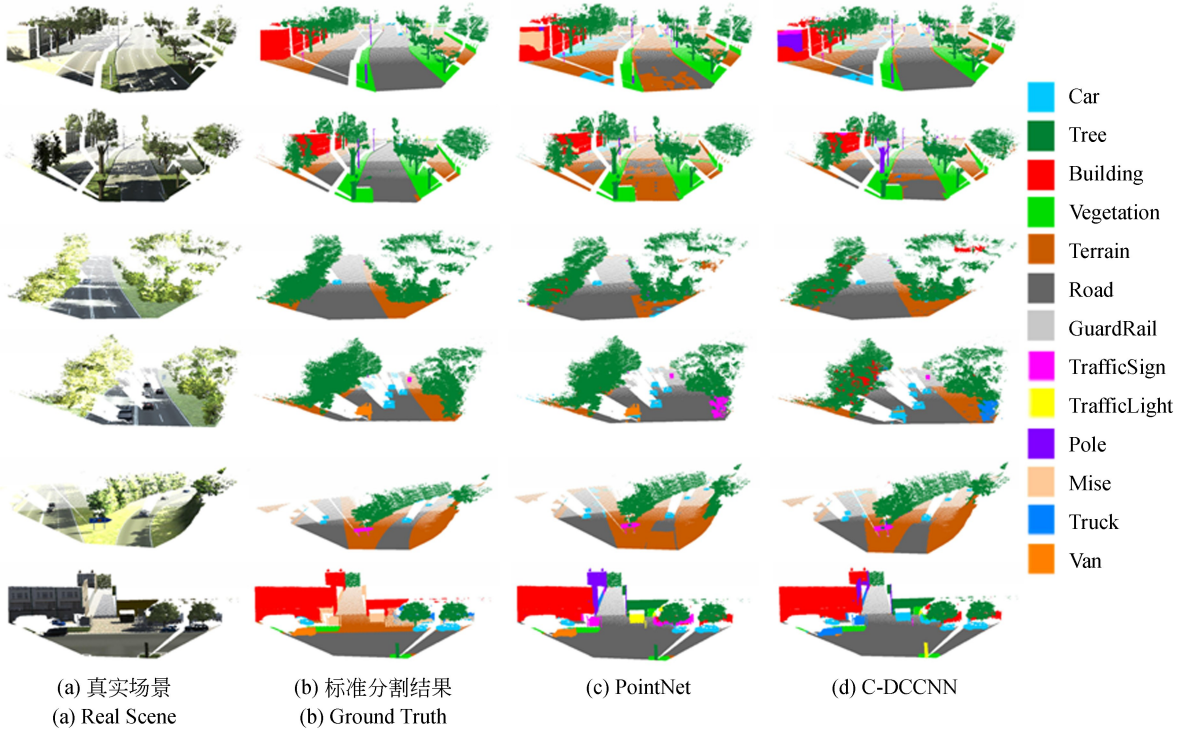


图 10　vKITTI 数据集上语义分割模型可视化

Fig. 10　Visualization of semantic segmentation models on vKITTI dataset

4　结　论

本文提出了一种基于深度级联卷积神经网络的三维目标识别和模型语义分割方法。通过构建深度动态图卷积神经网络作为深度级联卷积神经网络的子网络,对输入点集进行分层学习以捕捉点云的深层隐含细粒度几何特征。为了提高在非均匀采样点云上的特征学习能力,构建了 MSG-RNN 密度自适应层编码策略,可以根据局部点云密度利用 RNN 编码器自适应地聚合不同尺度的上下文几何信息,增强了网络的鲁棒性。实验结

果表明，本文算法在三维目标识别数据集ModelNet40和ModelNet10上的识别准确率分别为91.9%，94.3%，在模型语义分割数据集ShapeNet Part，S3DIS，vKITTI上的mIoU分别为85.6%，58.3%，38.6%。在三维点云目标识别准确率、语义分割准确率和网络鲁棒性上都优于其他主流算法。实验中发现，本文算法对如何忽略次要干扰局部特征，只关注显著局部特征还存在改进的空间，这也是今后要继续研究的方向。

参考文献：

[1] OSADA R，FUNKHOUSER T，CHAZELLE B，*et al*.. Shape distributions [J]. *ACM Transactions on Graphics*，2002，21(4)：807-832.

[2] SUN J，OVSJANIKOV M，GUIBAS L. A concise and provably informative multi-scale signature based on heat diffusion [J]. *Computer Graphics Forum*，2009，28(5)：1383-1392.

[3] KRIZHEVSKY A，SUTSKEVER I，HINTON G E. Imagenet classification with deep convolutional neural networks [C]. *Advances in Neural Information Processing Systems*，*Long Beach*，*USA*：*NIPS*，2012：1097-1105.

[4] 吕晓琪，吴凉，谷宇，等. 基于三维卷积神经网络的低剂量CT肺结节检测 [J]. 光学 精密工程，2018，26(5)：1211-1218.
LV X Q，WU L，GU Y，*et al*.. Detection of low dose CT pulmonary nodules based on 3D convolution neural network [J]. *Opt. Precision Eng.*，2018，26(5)：1211-1218. (in Chinese)

[5] 潘仙张，张石清，郭文平. 多模深度卷积神经网络应用于视频表情识别 [J]. 光学 精密工程，2019，27(4)：963-970.
PAN X ZH，ZHANG SH Q，GUO W P. Video-based facial expression recognition using multimodal deep convolutional neural networks [J]. *Opt. Precision Eng.*，2019，27(4)：963-970. (in Chinese)

[6] 郑斌琪，李宝清，刘华巍，等. 采用自适应一致性UKF的分布式目标跟踪 [J]. 光学 精密工程，2019，27(1)：260-270.
ZHENG B Q，LI B Q，LIU H W，*et al*.. Distributed target tracking based on adaptive consensus UKF [J]. *Opt. Precision Eng.*，2019，27(1)：260-270. (in Chinese)

[7] SU H，MAJI S，KALOGERAKIS E，*et al*.. Multi-view convolutional neural networks for 3D shape recognition [C]. *IEEE International Conference on Computer Vision*，*New York*，*USA*：*IEEE*，2015：945-953.

[8] SIMONYAN K，ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [J]. *ArXiv Preprint ArXiv*：1409.1556，2014.

[9] MATURANA D，SCHERER S，FRANKA A. VoxNet：a 3D convolutional neural network for real-time object recognition [C]. *IEEE International Conference on Intelligent Robots and Systems*，*New York*，*USA*：*IEEE*，2015：922-928.

[10] 杨军，王亦民. 基于深度级联卷积神经网络的三维模型识别 [J]. 重庆邮电大学学报，2019，31(2)：253-260. YANG J，WANG Y M. 3D model recognition and classification based on deep convolution neural network [J]. *Journal of Chongqing University*，2019，31(2)：253-260. (in Chinese)

[11] 杨军，王顺，周鹏. 基于深度体素卷积神经网络的三维模型识别分类 [J]. 光学学报，2019，39(4)：0415007.
YANG J，WANG SH，ZHOU P. 3D model recognition and classification based on deep voxel convolution neural network [J]. *Acta Optica Sinica*，2019，39(4)：0415007. (in Chinese)

[12] 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*，*New York*，*USA*：*IEEE*，2017：863-872.

[13] ZHU Y，MOTTAGHI R，KOLVE E，*et al*.. Target-driven visual navigation in indoor scenes using deep reinforcement learning [C]. *IEEE International Conference on Robotics and Automation*，*New York*，*USA*：*IEEE*，2017：3357-3364.

[14] QI C，LIU W，WU C，*et al*.. Frustum pointnets for 3D object detection from RGB-D data [C]. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*，*New York*，*USA*：*IEEE*，2018：918-927.

[15] RUSU R，MARTON Z，BLODOW，*et al*.. Towards 3D point cloud based object maps for household environments [J]. *Robotics and Autonomous Systems*，2008，56(11)：927-941.

[16] 赵传，张保明，余东行，等. 利用迁移学习的机载激光雷达点云分类 [J]. 光学 精密工程，2019，27

(7):1601-1612.
ZHAO CH, ZHANG B M, ZHANG D X, *et al*.. Airborne LiDAR point cloud classification using transfer learning [J]. *Opt. Precision Eng.*, 2019, 27(7):1601-1612. (in Chinese)

[17] QI C, 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*, *New York*, *USA*: *IEEE*, 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 [C]. *Advances in neural information processing systems*, *Long Beach*, *USA*: *NIPS*, 2017: 5099-5108.

[19] WANG Y, SUN Y, LIU Z, *et al*.. Dynamic graph cnn for learning on point clouds [J]. *ACM Transactions on Graphics*, 2019, 38(5): 146.

[20] ZHANG K, HAO M, WANG J, *et al*.. Linked dynamic graph CNN: learning on point cloud via linking hierarchical features [J]. *ArXiv Preprint ArXiv*: 1904.10014, 2019.

[21] 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*. *New York*, *USA*: *IEEE*, 2016: 770-778.

[22] 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*, *New York*, *USA*: *IEEE*, 2015: 1912-1920.

[23] 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): 210.

[24] 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*. *New York*, *USA*: *IEEE*, 2016: 1534-1543.

[25] 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*. *New York*, *USA*: *IEEE*, 2017: 716-724.

[26] KLAMBAUER G, UNTERHINER T, MAYR A, *et al*.. Self-normalizing neural networks [C]. *Advances in Neural Information Processing Systems*, *Long Beach*, *USA*: *NIPS*, 2017: 971-980.

[27] GAIDON A, WANG Q, CABON Y, *et al*.. Virtual worlds as proxy for multi-object tracking analysis[C]. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. *New York*, *USA*: *IEEE*, 2016.

作者简介:

杨　军(1973-),男,宁夏吴忠人,博士后,教授,博士生导师,1995 年于西北师范大学获得学士学位,2002 年于兰州交通大学获硕士学位,2007 年于西南交通大学计算机科学专业获博士学位,主要从事三维模型的空间分析、模式识别等方面的研究。E-mail: yangj@mail.lzjtu.cn

党吉圣(1991-),男,甘肃武威人,硕士研究生,2016 年于兰州交通大学获得学士学位,主要从事模式识别、计算机视觉方面的研究。E-mail: 1442342449@qq.com