# 基于上下文注意力 CNN 的三维点云语义分割

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**：针对三维点云语义分割中缺乏结合点云的上下文细粒度信息导致的欠分割问题，提出一种基于上下文注意力卷积神经网络的三维点云语义分割算法。首先，通过注意力编码机制挖掘点云的局部区域内细粒度特征；然后，通过上下文循环神经网络编码机制捕捉多尺度局部区域之间的上下文特征，且与细粒度局部特征相互补偿；最后，采用多头部机制增强网络的泛化能力。实验结果表明，所提算法在 ShapeNet Parts、S3DIS 和 vKITTI 标准数据集上的平均交并比分别为 85.4%、56.7%和 38.1%，分割性能良好，且具有较好的泛化能力。

**关键词**：三维点云；语义分割；上下文注意力卷积层；卷积神经网络；深度学习
**中图分类号**：TP391
**文献标识码**：A
**doi:** 10.11959/j.issn.1000−436x.2020128

# Semantic segmentation of 3D point cloud based on contextual attention CNN

YANG Jun, DANG Jisheng
School of Electronic and Information Engineering, Lanzhou Jiaotong University, Lanzhou 730070, China

**Abstract:** Aiming at the under-segmentation of 3D point cloud semantic segmentation caused by the lack of contextual fine-grained information of the point cloud, an algorithm based on contextual attention CNN was proposed for 3D point cloud semantic segmentation. Firstly, the fine-grained features in local area of the point cloud were mined through the attention coding mechanism. Secondly, the contextual features between multi-scale local areas were captured by the contextual recurrent neural network coding mechanism and compensated with the fine-grained local features. Finally, the multi-head mechanism was used to enhance the generalization ability of the network. Experiments show that the mIoU of the proposed algorithm on the three standard datasets of ShapeNet Parts, S3DIS and vKITTI are 85.4%, 56.7% and 38.1% respectively, which has good segmentation performance and good generalization ability.

**Key words:** 3D point cloud, semantic segmentation, contextual attention convolution layer, convolutional neural network, deep learning

## 1 引言

点云是三维模型最重要的数据表示形式之一，其能够准确、直观地描述三维模型。随着三维成像技术的飞速发展，三维点云数据呈海量增长趋势，对其进行分析和处理显得尤其重要。语义分割作为三维点云数据分析处理的前提与基础，已广泛应用于医学成像、自动驾驶、机器人导航、虚拟现实、遥感测绘等领域，成为计算机视觉和计算机图形学领域的一个重要研究课题。由于卷积神经网络[1]的飞速发展以及 GPU（graphics processing unit）计算能力的显著提高，传统的手工设计描述符[2-4]的方法已渐渐被基于深度学习的方法所取代，一些研究者开始设计针对大规模、多种类的复杂三维点云模型语义分割的深度学习框架。目前，基于深度学习的三维点云模型语义

**收稿日期**：2020−01−06；**修回日期**：2020−04−20
**基金项目**：国家自然科学基金资助项目（No.61862039）
**Foundation Item:** The National Natural Science Foundation of China (No.61862039)

分割方法主要有：基于投影的方法、基于多模态的方法、基于体素化的方法和基于点云表示的方法。

基于投影的方法。受二维图像语义分割方法的启发，Ku 等[5]将三维点云投影为二维鸟瞰图，然后使用经典的二维深度学习网络提取特征用于三维点云语义分割。Yang 等[6]设计了一个单级检测器，从像素级神经网络输出点云分割结果。Kalogerakis 等[7]提出了投影卷积神经网络，首先对三维模型进行多方位拍照，然后将二维视图输入 VGG（visual geometry group）[8]中提取特征，最后将特征反投影到三维点云表面预测每个点的语义类别。Huang 等[9]提出多视图卷积神经网络，从对应关系中学习局部描述符，增强了网络的泛化能力。然而，基于投影的方法的关键问题是在生成 2D 投影图时，丢失了具有鉴别力的几何结构信息。

基于多模态的方法。Chen 等[10]提出 MV3D（multi-view 3D）目标检测网络，把鸟瞰图和点云同时作为输入获得多模态特征，进一步融合用于点云语义分割任务。Qi 等[11]利用一个二维检测网络构建截锥体点云。然而，基于多模态的方法通常计算效率较低。

基于体素化的方法。Wu 等[12]将不规则的三维点云模型转化为规则的体素网格，这样便可以使用三维卷积神经网络处理体素模型提取特征，但是体素方块大小的选择会影响网络性能，太大会丢失细节，太小会增加计算负担。改进算法提出了空间划分方法，如 Kd 树[13]或者八叉树[14]，解决了一些分辨率低的问题，但仍然依赖于边界体的细分，没有考虑局部几何结构。

基于点云表示的方法。为了避免多方位投影、体素化等方法的烦琐操作，Qi 等[15]提出了可以直接作用于无序点云数据的 PointNet 模型，利用多层感知机（MLP, multi-layer perceptron）学习每个独立点的高维表征，然后采用一个最大池化层对所有点的高维特征进行聚合得到全局特征描述符。PointNet 是深度学习框架可以直接作用于不规则三维点云数据的先驱性工作，然而，PointNet 只关注每个点的全局特征，缺乏捕捉局部特征的能力。Qi 等[16]提出了 PointNet++，通过划分局部点云分层提取多尺度特征。该网络虽然考虑了点云局部特征，但是没有考虑点对之间的关联信息，缺乏捕捉几何特征的能力。Wang 等[17]通过建立和更新动态图，在保证置换不变性的同时捕获局部几何特征，取得了较先进的分割结果。Chen 等[18]建立局部区域加强对相邻点的关注，以充分提取点云的局部几何特征。Liu 等[19]提出一种点云序列学习模型，采用注意力机制来突出不同尺度区域的重要性。Wang 等[20]提出了相似分组提议网络（SGPN, similarity group proposal network）模型，引入相似矩阵作为输出，表征嵌入特征空间中每对点之间的相似度，从而预测每个点的语义标签。Jiang 等[21]设计了 PointSIFT 模块，该模块可以对不同方向的信息进行编码，并且通过堆叠几个方向编码单元来实现多尺度表示，最后将解码器的输出连接到全连接层，用于预测每个点的语义类别。Ye 等[22]提出利用超点图来有效捕捉点云的组织结构，然后通过图卷积神经网络从超点图中提取特征来完成语义分割任务。Landrieu 等[23]提出了一种新的端到端语义分割方法，首先研究了利用多尺度邻域捕获不同密度局部结构特征的金字塔池化模型，然后利用双向循环神经网络（RNN, recurrent neural network）学习点云的空间相关性来捕捉空间结构信息，最后输出每个点的语义标签。

相对于三维目标识别任务，三维点云语义分割需要预测每个点的语义类别，提取更精细的点特征，因此是一项需要结合上下文信息的更具挑战性的细粒度点云分析任务，但由于点云存储在不规则和无序的结构中，提取点云的上下文细粒度特征信息仍然具有很大的挑战性。现有方法在三维点云语义分割过程中对于整体几何形状极其相似、局部细节结构略有不同的语义类不能进行有效区分，造成的欠分割问题一直没有得到很好的解决。本文提出一种基于上下文注意力卷积神经网络（CACNN, contextual attention convolutional neural network）的三维点云语义分割算法，充分挖掘三维模型的多尺度上下文细粒度特征信息，改善了三维点云语义分割的过分割问题，提高了三维点云语义分割的准确率。主要贡献和创新点如下。1) 构建上下文注意力卷积层。通过在局部点云中引入图注意力机制来对邻域特征进行自适应筛选，更好地学习点云的细粒度局部特征。2) 通过上下文 RNN 编码每个采样点的不同尺度邻域特征来学习每个采样点的多尺度上下文几何特征，并与细粒度局部特征相互补偿增强特征描述符的语义丰富性。3) 采用多头部机制聚合不同的单头部上下文注意力卷积层的特征，使网络具有良好的泛化能力，同时在网络中引入残差学

习以充分挖掘三维点云的深层隐含特征信息，进一步提高网络特征学习的能力。

## 2　上下文注意力卷积神经网络

### 2.1　上下文注意力卷积层

在实际应用（如自动驾驶）中，点云的数目非常大，为了减少计算成本，需要构建一个 $k$ 近邻图 $G=(V, E)$来表示点云的一个局部区域。其中，$V=\{1, 2, \cdots, N\}$为点的集合，$E \subseteq V \times \alpha_i$ 表示连接相邻点对的边，$\alpha_i$ 为点 $x_i$ 的邻域点的集合。为了使点集的特征学习不受旋转、平移等变换的影响，将每个局部区域点的坐标 $x_{ij}$ 转换为中心点 $x_i$ 的相对坐标，即得到边的特征为

$$y_{ij}=(x_i, x_{ij}-x_i), x_i \in \mathbb{R}^F, \forall x_{ij} \in \text{Neighbors}(x_i) \quad (1)$$

其中，$x_i \in V$，$x_{ij} \in \alpha_i$。

为了充分挖掘点云的细粒度细节和多尺度上下文信息，在 PointNet[15]的基础上，本文构建上下文注意力卷积（CAC, contextual attention convolutional）层，采用注意力编码和上下文 RNN 编码 2 个并行编码机制分别学习局部区域内细粒度特征和局部区域之间的多尺度上下文几何特征，上下文注意力卷积层网络结构如图 1 所示。其中，MLP{} 表示多层感知机操作，{ }中的数字表示卷积核的数目。

注意力编码机制首先采用输出通道为 $F_1$ 的 MLP 将原始点特征和边特征映射到高维特征空间，如式(2)和式(3)所示。

$$u_i'=\lambda_{\Theta}(\text{BN}(c_{F_1\times 1}(x_i))) \quad (2)$$

$$h_{ij}'=\lambda_{\Theta}(\text{BN}(c_{F_1\times 1}(y_{ij}))) \quad (3)$$

其中，$\lambda$ 为参数化的非线性激活函数，$\Theta$ 为卷积核中可学习的参数集合，BN 为批归一化处理，$c$ 为卷积操作，其下标 $F_1\times 1$ 表示卷积核大小。实验中 $F_1$ 取 16，即特征通道数为 16。对 $u_i'$ 和 $h_{ij}'$ 分别采用一个 MLP 生成描述点 $x_i$ 自注意力系数和邻域注意力系数，并将两者进行融合得到描述点 $x_i$ 到其邻域内 $k$ 个邻近点的注意力系数 $b_{ij}$，如式(4)所示。

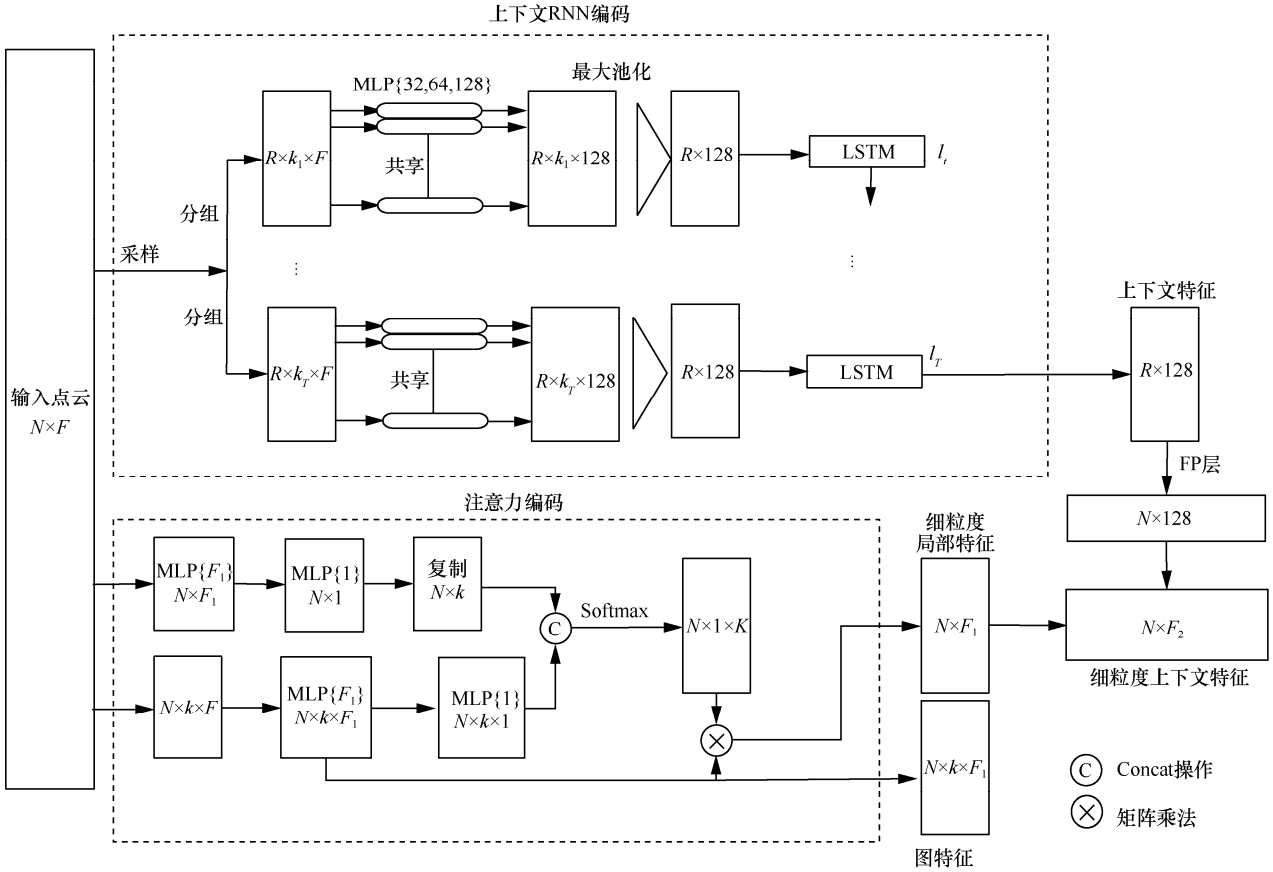



图 1　上下文注意力卷积层网络结构

$$b_{ij}=\mathrm{Selu}(\lambda_{\Theta}(\mathrm{BN}(c_{1\times1}(u_i')))+\lambda_{\Theta}(\mathrm{BN}(c_{1\times1}(h_{ij}')))) \quad (4)$$

其中，Selu()为非线性激活函数。为了提高模型收敛速度，采用 Softmax 函数对注意力系数进行归一化处理，如式(5)所示。

$$a_{ij}=\frac{\exp(b_{ij})}{\sum_{j=1}^{k}\exp(b_{ij})} \quad (5)$$

为了挖掘细粒度局部特征，将注意力系数 $a_{ij}$ 与局部图特征 $h_{ij}'$ 相乘，则细粒度局部特征 $l_i$ 为

$$l_i=\mathrm{Selu}\left(\sum_{j=1}^{k}a_{ij}h_{ij}'\right) \quad (6)$$

此时，注意力系数作为一个特征选择器，自适应地对描述点 $x_i$ 具有鉴别力的邻域特征进行增强，抑制无意义的邻域特征（如噪声）充分挖掘点云局部区域内的细粒度细节信息。

上下文 RNN 编码机制首先采用迭代最远点采样算法从输入点云中选取 $R$（$R$<$N$）个点作为 $R$ 个局部区域的中心点。对于每个采样点，采用 $k$ 最近邻算法分别搜索距离采样点最近的[$K_1$, ···, $K_t$, ···, $K_T$]个点来构建 $T$ 个不同尺度的局部邻域。然后，分别采用卷积核数目为 32、64、128 的 3 个 MLP 提取每个局部邻域的几何特征，第一个 MLP 卷积核大小为 $F$×1，由于 1×1 卷积核[24]可以增强网络的非线性拟合能力，减少参数的同时可以聚合各通道信息，因此，网络中其余 MLP 的卷积核大小均采用 1×1。最后，采用一个最大池化层分别将不同尺度的邻域特征聚合到每个采样点上，得到每个采样点的不同尺度的特征序列 $\boldsymbol{S}_k=\{\boldsymbol{s}_k^1,\cdots,\boldsymbol{s}_k^t,\cdots,\boldsymbol{s}_k^T\}$。其中，$\boldsymbol{s}_k^t$ 为采样点的 $k_t$ 邻域的几何特征向量。

为了获得采样点的不同尺度邻域之间的相关性，把采样点的特征序列 $\boldsymbol{S}_k=\{\boldsymbol{s}_k^1,\cdots,\boldsymbol{s}_k^t,\cdots,\boldsymbol{s}_k^T\}$ 输入 RNN 编码器，并用一个隐藏层 $d$ 依次编码采样点不同尺度的邻域特征向量来充分挖掘上下文几何信息。RNN 在编码采样点 $x_i$ 的不同尺度邻域的特征向量时，依次更新隐藏层状态，如式(7)所示。

$$d_t=p(d_{t-1},\boldsymbol{s}_k^t),t\in[1,T] \quad (7)$$

其中，$p$ 为非线性激活函数，实验中采用 LSTM 单元；$d_{t-1}$ 为编码上一个邻域特征向量 $\boldsymbol{s}_{k-1}$ 的隐藏层状态。在 RNN 编码采样点的第 $t$ 个邻域的特征向量 $\boldsymbol{s}_k$ 时，编码器的输出 $o_t$ 为

$$o_t=\boldsymbol{W}_a d_t \quad (8)$$

其中，$\boldsymbol{W}_a$ 是一个可学习的权重矩阵。当网络学习完成整个特征序列后，得到隐藏层状态 $d_T$，和 $\boldsymbol{W}_a$ 相乘得到采样点的多尺度上下文几何特征 $o_T$。

注意力编码虽然引入注意力机制增强了网络捕捉局部区域内细粒度细节的能力，但是忽略了对于点云语义分割至关重要的局部区域之间的上下文几何信息。上下文 RNN 编码机制充分挖掘了点云的多尺度上下文高级特征，因此低级别的细粒度局部特征和高级别的多尺度上下文几何特征可以相互补偿。采用 Selu 非线性激活函数将采样点的不同层次的细粒度局部特征和上下文几何特征融合，可以得到采样点的大小为 $N$×$F_2$ 的上下文细粒度几何特征。在特征融合前，采用插值操作[16]在 $R$×128 的点云上采样 $N$×128 的点云。特征融合计算式为

$$q_i=\mathrm{Selu}(o_T+l_i) \quad (9)$$

## 2.2 多头部上下文注意力卷积层

为了获得丰富的特征信息以进一步增强网络的泛化能力，本文引入多头部机制。在计算 CAC 层的上下文细粒度特征和图特征时引入随机丢弃（dropout）算法，通过随机丢弃一些权重得到 $M$ 个不同的单头部（single-head）CAC 层。然后，把 $M$ 个单头部 CAC 层连接到一起得到特征信息更加丰富的多头部上下文注意力卷积（M-CAC, multi-heads contextual attention convolutional）层。多头部 CAC 层网络结构如图 2 所示，计算式如式(10)所示。

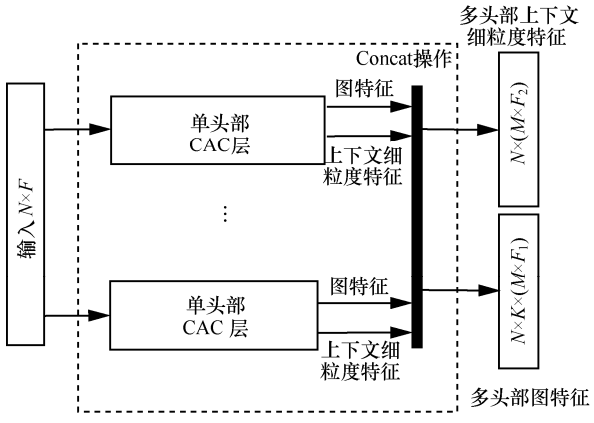



图 2 多头部 CAC 层网络结构

$$l_i'=\mathrm{Selu}(l_i^1+l_i^m+\cdots+l_i^M) \quad (10)$$

其中，$l_i^m$ 为描述点 $x_i$ 的第 $m$ 个头部的上下文细粒度特征，$M$ 为头部数，“+”为特征连接操作。由于不同的单头部 CAC 层提取到的特征差异性是有限的，为了增强特征丰富性，同时避免造成特征的冗余，

实验中头部数 $M$ 设为 3。为了有效保持不同单头部 CAC 层的差异性和原有特征的完整性，设置权重保留率 dropout 为 0.7。

### 2.3　上下文注意力卷积神经网络

为了进一步挖掘点云的深层隐含语义特征信息，本文在构建的上下文注意力卷积神经网络中引入残差学习，网络结构如图 3 所示。其中，空间转换网络为一个 3×3 的矩阵。

对于网络输入的 $N\times F$ 点云矩阵，首先采用一个空间转换网络对其进行规范化，实现点云矩阵的空间变换不变性。然后，采用 M-CAC 层提取输入点云的上下文细粒度几何特征和图特征，将得到的特征维度为 $N\times176$ 的上下文细粒度几何特征与区域中心点的三维坐标特征相结合，获得维度为 $N\times179$ 的点云矩阵输入堆叠的MLP层进行二次特征提取。此外，为了挖掘深层隐含语义特征，引入残差连接，在避免梯度消失的同时加深了网络深度，各层卷积层的具体参数设置如表 1 所示。最后一层卷积层 $\text{Layer}_6$ 输出的 $N\times1\ 024$ 特征矩阵通过一个最大池化层进行特征聚合得到 1×1 024 的全局特征描述符。为了从全局形状特征描述符中获取点级别的特征，在网络中引入 2 个插值层[16]，通过上采样将特征从形状级别传播到点级别。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过程 $\chi$，由点 $q$ 与其 $k$ 最近邻点 $q_i$ 的欧氏距离插值而成，计算式如下

$$\chi(q)=\frac{\sum_{i=1}^{k}w(q_i)\chi(q_i)}{\sum_{i=1}^{k}w(q_i)} \tag{11}$$

其中，$w(q_i)$ 为

$$w(q_i)=\frac{1}{(q-q_i)^2} \tag{12}$$

其中，$q-q_i$ 表示点 $q$ 与 $q_i$ 的欧氏距离。为了引导插值过程，将插值后的特征与对应的点特征连接起来，并在网络中引入多个 MLP 层和 Selu 层促进点级别特征的提取。最后网络输出分割结果 $N\times S$ 点云矩阵，表示每个点的语义类别。

## 3　实验结果与分析

### 3.1　实验数据集

为了验证本文算法的语义分割性能和泛化性，实验选用 3 个标准公开数据集，分别为部件语义分割数据集 ShapeNet Parts[25]、室内场景语义分割数

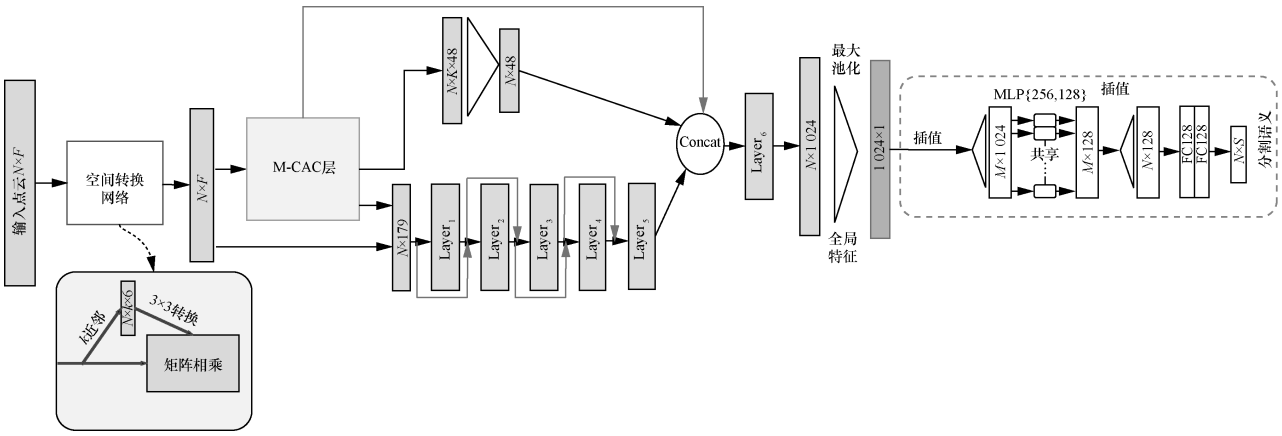



图 3　CACNN 的网络结构

**表 1**　　**卷积层参数设置**

| 层名称 | 输入通道 | 输出通道 | 卷积核大小 | 卷积步长 | 批规范化 | 激活函数 |
|---|---|---|---|---|---|---|
| $\text{Layer}_1$ | 179 | 64 | 1×1 | 1 | 是 | Selu |
| $\text{Layer}_2$ | 243 | 64 | 1×1 | 1 | 是 | Selu |
| $\text{Layer}_3$ | 128 | 64 | 1×1 | 1 | 是 | Selu |
| $\text{Layer}_4$ | 128 | 64 | 1×1 | 1 | 是 | Selu |
| $\text{Layer}_5$ | 128 | 128 | 1×1 | 1 | 是 | Selu |
| $\text{Layer}_6$ | 352 | 1 024 | 1×1 | 1 | 否 | Selu |

据集 S3DIS[26]和户外场景语义分割数据集 vKITTI[27]。ShapeNet Parts 数据集包含 16 个类别的 16 881 个 CAD 模型，其中 9 843 个模型用于训练，2 468 个模型用于测试，定义了 50 个部件语义标签。S3DIS 数据集是一个大规模室内 RGB-D 数据集，共有 6 个区域 271 个房间，定义了地板、窗户、门、横梁等 13 个语义类别。实验设置和文献[15]一样，采用 6 个区域交叉验证。vKITTI 数据集是一个模拟现实世界的可应用于自动驾驶的户外大规模点云数据集，将 5 个不同城市场景的视频序列分为互不重叠的 6 个区域，定义了车、树木、马路、建筑物等 13 个语义类别。

### 3.2 网络参数设置

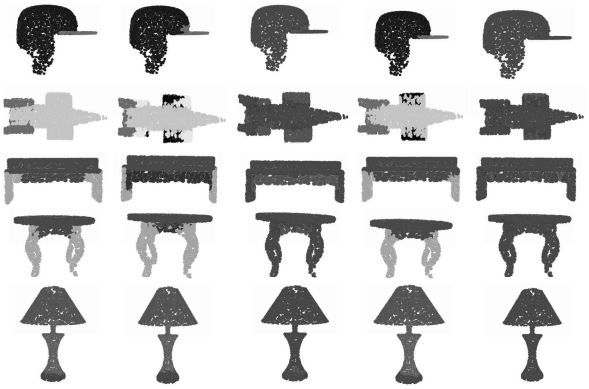
本文算法 CACNN 的训练和测试过程的实验环境基于 Linux Ubuntu 16.04 操作系统、Intel i7 8700k CPU、内存 32 GB、GeForce RTX 2080 GPU，运算平台为 CUDA-Toolkit 9.0，采用 Cudnn 7.13 作为网络的 GPU 加速库，深度学习框架为 Tensorflow-GPU，版本号为 1.9.0。在实验中，CACNN 的训练过程采用基于动量的随机梯度下降（SGD, stochastic gradient descent）优化算法，设置动量为 0.9，权重衰减为 0.000 5，初始学习率为 0.001，学习率衰减系数为 0.5，衰减速度为 300 000，全连接层中 dropout 的参数保留率为 0.5。优化器采用 Adam，网络参数初始化采用 Xavier 优化器。

### 3.3 实验结果与分析

为了验证本文算法在处理三维点云语义分割任务上的优越性，在 ShapeNet Parts 数据集上与其他先进算法在识别准确率和效率两方面进行了对比，评估准则采用前向传播时间和平均交并比 mIoU（mean intersection-over-union），实验结果如表 2 所示。可以看出，本文算法以 85.4%的 mIoU 和 39.0 ms 的前向传播时间获得了较好的分割性能。本文算法相比于当前主流算法 DGCNN（dynamic graph convolutional neural network），在分割准确率和计算效率方面都具有一定优势。图 4 为本文算法 CACNN 与 PointNet[15]在 ShapeNet Parts 数据集上几个类别的模型分割结果对比，其中 PointNet_diff 和 CACNN_diff 分别标出了 PointNet 和 CACNN 的预测结果与真实体的不同之处。与 PointNet 相比，本文算法总体分割错误率明显减少，纠正了 PointNet 在细粒度边界处的欠分割问题，如桌子的底部、台灯的底端等，进一步验证了本文算法通过构建 CAC 层能够捕捉对于点云语义分割至关重要的上下文细粒度信息。

**表 2 不同算法在 ShapeNet Parts 数据集上的网络性能比较**

| 算法 | mIoU | 前向传播时间/ms |
|---|---|---|
| Kd-Net[13] | 82.3% | — |
| PointNet[15] | 83.7% | 16.9 |
| PointNet++[16] | 85.1% | 30.7 |
| DGCNN[17] | 85.1% | 40.4 |
| 本文算法 | 85.4% | 39.0 |



(a) 真实值 (b) PointNet (c) PointNet_diff (d) CACNN (e) CACNN_diff

图 4 本文算法与 PointNet 在 ShapeNet Parts 数据集上的模型分割结果对比

此外，为了探究本文构建的 CAC 层中采用注意力编码和上下文 RNN 编码机制的有效性，在实验中依次引入这 2 种编码机制以验证每一种编码机制的作用，实验结果如表 3 所示。PointNet 算法采用自编码机制对独立的点提取特征，本文算法通过引入注意力编码方式对三维点云语义分割的 mIoU 比 PointNet 提高了 0.7%，原因在于注意力编码机制通过对每个点的邻域特征进行区分，能够充分挖掘局部点云的细粒度局部特征信息。注意力编码和上下文 RNN 编码相结合的方式对三维点云语义分割的 mIoU 比 PointNet 提高了 1.7%，因为上下文编码可以结合每个采样点的多尺度上下文几何信息，和注意力编码提取到的细粒度特征进行相互补偿，得到特征信息更加丰富的细粒度多尺度上下文特征。

**表 3 不同编码机制的有效性分析**

| 算法 | mIoU |
|---|---|
| PointNet[15]（自编码） | 83.7% |
| 本文算法（注意力编码） | 84.4% |
| 本文算法（注意力编码+上下文 RNN 编码） | 85.4% |

本文继续探究了 CAC 层、M-CAC 层和残差学习对网络性能的影响。通过构造不同的网络进行训练并测试，对比实验结果如表 4 所示。可以看出，在 PointNet 基础上加入 CAC 层后 mIoU 提高了 1.4%，因为 CAC 层中聚合了多尺度上下文细粒度特征。采用 M-CAC（$M$=3，dropout=0.7）后 mIoU 提高了 0.2%，原因在于多头部机制提高了特征的丰富性，增强了网络的泛化能力。引入残差学习后 mIoU 又提高了 0.1%，因为残差学习在避免梯度消失问题的同时加深了网络容量，能够充分挖掘点云的深层语义特征信息。

**表 4　不同组件的有效性分析**

| 组件 | mIoU |
|---|---|
| PointNet 无组件[16] | 83.7% |
| CAC | 85.1% |
| M-CAC ($M$=3，dropout=0.7) | 85.3% |
| M-CAC ($M$=4，dropout=0.6) | 84.9% |
| M-CAC+残差学习 | 85.4% |

### 3.4　室内和室外场景分割

为了验证本文算法对于大规模点云分析的有效性，在室内数据集 S3DIS 和户外场景分割数据集 vKITTI 上进行了训练和测试，并与主流算法进行了对比，实验结果如表 5 和表 6 所示。可以看出，本文算法语义分割的总体准确率（OA, overall accuracy）和 mIoU 都优于其他主流方法，取得了较理想的分割准确率。除了定量分析外，图 5 和图 6 展示了定性的分割可视化结果。

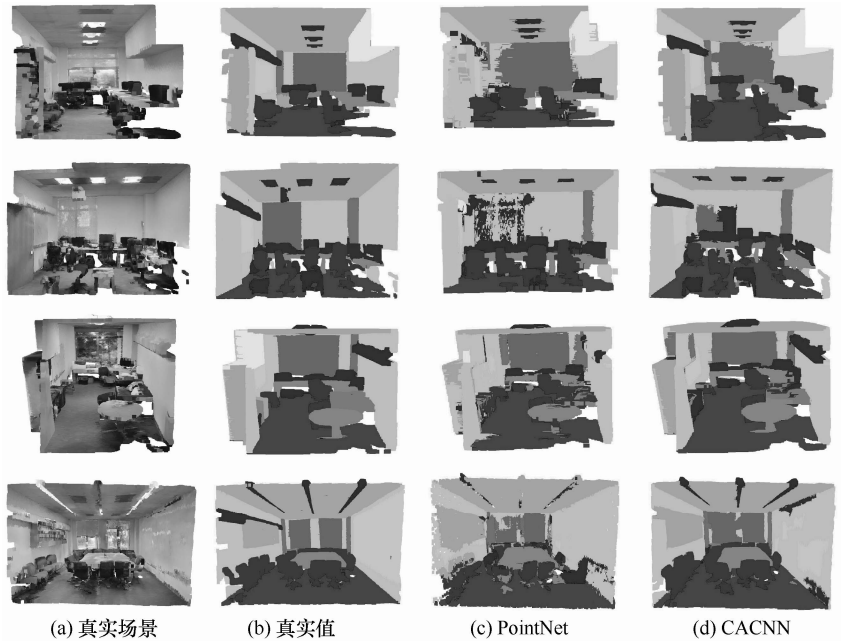
**表 5　S3DIS 数据集上不同方法的分割准确率对比**

| 算法 | mIoU | OA |
|---|---|---|
| PointNet[15] | 47.6% | 78.5% |
| MS+CU[28] | 47.8% | 79.2% |
| G+RCU[28] | 49.7% | 81.1% |
| SGPN[20] | 50.4% | 80.8% |
| PointNet++ [16] | 54.5% | 81.0% |
| DGCNN [17] | 56.1% | 84.1% |
| 本文算法 | 57.6% | 85.2% |

**表 6　vKITTI 数据集上不同算法的分割准确率对比**

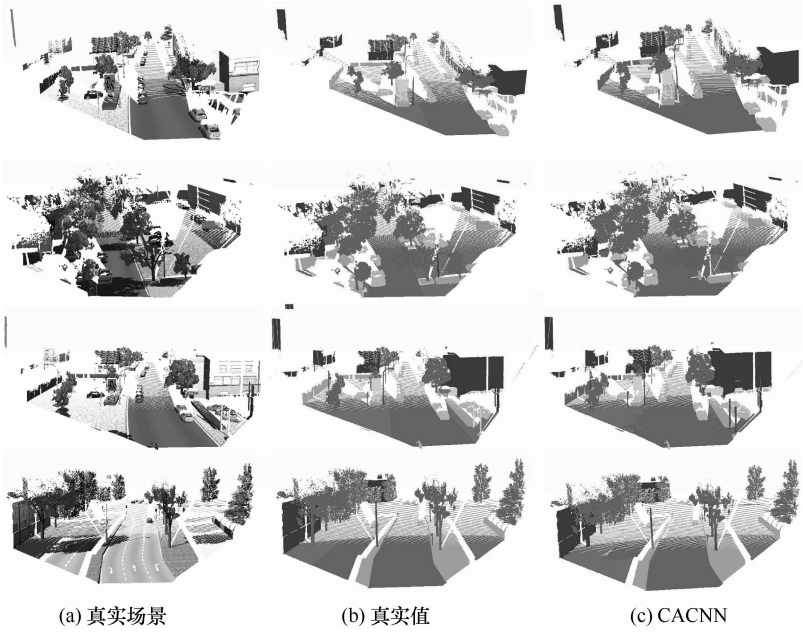
| 算法 | OA | mIoU |
|---|---|---|
| PointNet[15] | 79.7% | 34.4% |
| G+RCU[28] | 80.6% | 36.2% |
| 本文算法 | 82.0% | 38.1% |

从图 5 中可以看出，CAC 层能够改善 PointNet 存在的欠分割问题，获得更准确的分割结果。例如，PointNet 对于椅子腿这类细粒度语义类识别能力有限，而本文算法能够很好地分割出椅子腿的边界腿梢，总体上以更少的错误分割整个场景，证明了本文算法的注意力编码机制能够挖掘局部点云的细粒度细节信息的能力。此外，相比于 PointNet 的粗预测，本文算法对 Board 的预测准



(a) 真实场景　(b) 真实值　(c) PointNet　(d) CACNN

图 5　S3DIS 数据集上模型语义分割可视化



(a) 真实场景　　(b) 真实值　　(c) CACNN

图 6　vKITTI 数据集上模型语义分割可视化

确率也有所提高。原因在于 Board 和 Wall 的几何形状十分相似，只有上下文细节信息有所差异，本文算法能够结合 Board 的上下文信息和 Wall 的特征，可以更好地确定其边界，改善了欠分割问题，进一步证明了本文算法的上下文 RNN 编码机制能够有效结合点云上下文几何信息的能力。在图 6 中，本文算法对于几何形状极其相似的 Terrain 和 Road 这两类语义的识别能力明显提高，总体分割错误率也相对较少，原因在于本文算法能够结合每个点的上下文细粒度信息，对于识别 Terrain，能够结合其上下文 Tree 的细粒度细节信息以便于确定其边界，可以减少与 Road 的混淆。

## 4　结束语

本文提出了一种基于上下文注意力卷积神经网络的三维点云语义分割算法。首先将注意力机制引入 CAC 层来挖掘点云的局部细粒度特征，其次通过 RNN 编码不同尺度邻域特征以捕捉点云的多尺度上下文特征，与局部细粒度特征进行优势互补，并采用多头部机制增强了网络的泛化能力。同时在网络中引入残差学习进一步充分挖掘点云的深层隐含语义特征。定性和定量实验结果表明，本文算法有效改善了三维点云语义分割中存在的欠分割问题，总体分割准确率得到了提升，且本文算法在 3 个标准公开数据集上都表现优异，充分证明了其具有良好的泛化性。然而，本文网络结构复杂，训练参数较多，难以适用于实时点云分割任务，如何构建一个可部署到嵌入式设备中的轻量级实时点云分割网络是需要进一步研究的问题。

## 参考文献：


[1] KRIZHEVSKY A, SUTSKEVER I, HINTON G. ImageNet classification with deep convolutional neural networks[C]//Advances in Neural Information Processing Systems. Piscataway: IEEE Press, 2012: 1097-1105.

[2] YI L, SU H, GUO X, et al. SyncSpecCNN: synchronized spectral CNN for 3D shape segmentation[C]//IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2017: 2282-2290.

[3] ZHANG J, ZHENG J, WU C, et al. Variational mesh decomposition[J]. ACM Transactions on Graphics (TOG), 2012, 31(21): 1-14.

[4] XIAO D, LIN H, XIAN C, et al. CAD mesh model segmentation by clustering[J]. Computers & Graphics, 2011, 35(3): 685-691.

[5] KU J, MOZIFIAN M, LEE J, et al. Joint 3D proposal generation and object detection from view aggregation[C]//2018 IEEE/RSJ International Conference on Intelligent Robots and Systems. Piscataway: IEEE Press, 2018: 1-8.

[6] YANG B, LUO W, URTASUN R. PIXOR: real-time 3D object detection from point clouds[C]//IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 7652-7660.

[7] KALOGERAKIS E, AVERKIOU M, MAJI S, et al. 3D shape segmentation with projective convolutional networks[C]//IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2017: 3779-3788.

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

[9] HUANG H, KALOGERAKIS E, CHAUDHURI S, et al. Learning local shape descriptors from part correspondences with multi view convolutional networks[J]. ACM Transactions on Graphics (TOG),

2018, 37(1): 1-14.

[10] CHEN X, MA H, WAN J, et al. Multi-view 3D object detection network for autonomous driving[C]//IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2017: 1907-1915.

[11] QI C, LIU W, WU C, et al. Frustum PointNets for 3D object detection from RGB-D data[C]//IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 918-927.

[12] WU Z, SONG S, KHOSLA A, et al. 3D ShapeNets: a deep representation for volumetric shapes[C]//IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2015: 1912-1920.

[13] KLOKOV R, LEMPITSKY V. Escape from cells: deep Kd-networks for the recognition of 3D point cloud models[C]//IEEE International Conference on Computer Vision. Piscataway: IEEE Press, 2017: 863-872.

[14] TATARCHENKO M, DOSOVITSKIY A, BROX T. Octree generating networks: efficient convolutional architectures for high-resolution 3D outputs[C]//IEEE International Conference on Computer Vision. Piscataway: IEEE Press, 2017: 2088-2096.

[15] QI C, SU H, MO K, et al. PointNet: deep learning on point sets for 3D classification and segmentation[C]//IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2017: 652-660.

[16] QI C, 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. Piscataway: IEEE Press, 2017: 5099-5108.

[17] WANG Y, SUN Y, LIU Z, et al. Dynamic graph CNN for learning on point clouds[J]. ACM Transactions on Graphics, 2019, 38(5): 146-160.

[18] CHEN C, FRAGONARA L, TSOURDOS A. GAPNet: graph attention based point neural network for exploiting local feature of point cloud[J]. arXiv Preprint, arXiv: 1905.08705, 2019.

[19] LIU X, HAN Z, LIU Y, et al. Point2sequence: learning the shape representation of 3D point clouds with an attention-based sequence to sequence network[C]//Proceedings of the AAAI Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2019, 33: 8778-8785.

[20] WANG W, YU R, HUANG Q, et al. SGPN: similarity group proposal network for 3D point cloud instance segmentation[C]//IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 2569-2578.

[21] JIANG M, WU Y, ZHAO T, et al. PointSIFT: a SIFT-like network module for 3D point cloud semantic segmentation[J]. arXiv Preprint, arXiv: 1807.00652, 2018.

[22] YE X, LI J, HUANGU H, et al. 3D recurrent neural networks with context fusion for point cloud semantic segmentation[C]//Proceedings of the European Conference on Computer Vision (ECCV). Piscataway: IEEE Press, 2018: 403-417.

[23] LANDRIEU L, SIMONOVSKY M. Large-scale point cloud semantic segmentation with superpoint graphs[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 4558-4567.

[24] LIN M, CHEN Q, YAN S. Network in network[J]. arXiv Preprint, arXiv: 1312.4400, 2013.

[25] YI L, GUIBAS L, KIM V, et al. A scalable active framework for region annotation in 3D shape collections[J]. ACM Transactions on Graphics, 2016, 35(6):1-12.

[26] ARMENI I, SENER O, ZAMIR A, et al. 3D semantic parsing of large-scale indoor spaces[C]// IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2016: 1534-1543.

[27] GAIDON A, WANG Q, CABON Y, et al. Virtual worlds as proxy for multi-object tracking analysis[C]// IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2016: 4340-4349.

[28] ENGELMANN F, KONTOGIANNI T, HERMANS A, et al. Exploring spatial context for 3D semantic segmentation of point clouds[C]//IEEE International Conference on Computer Vision. Piscataway: IEEE Press, 2017: 716-724.

[作者简介]

**杨军**（1973− ），男，回族，宁夏吴忠人，博士，兰州交通大学教授、博士生导师，主要研究方向为计算机图形学、数字图像处理等。

**党吉圣**（1991− ），男，甘肃武威人，兰州交通大学硕士生，主要研究方向为机器视觉、模式识别。