

基于语义信息补偿全局特征的物体点云分类分割

林森 赵振禹 任晓奎 陶志勇

Object point cloud classification and segmentation based on semantic information compensating global features

Lin Sen, Zhao Zhenyu, Ren Xiaokui, Tao Zhiyong

在线阅读 View online: <https://doi.org/10.3788/IRLA20210702>

您可能感兴趣的其他文章

Articles you may be interested in

基于深度学习的3D打印球形粉末颗粒自动统计与测量

Particle auto-statistics and measurement of the spherical powder for 3D printing based on deep learning

红外与激光工程. 2021, 50(10): 2021G004 <https://doi.org/10.3788/IRLA2021G004>

合并分割块的点云语义分割方法

Point cloud semantic segmentation method based on segmented blocks merging

红外与激光工程. 2021, 50(10): 20200482 <https://doi.org/10.3788/IRLA20200482>

深度学习算法及其在光学的应用

Deep learning algorithm and its application in optics

红外与激光工程. 2019, 48(12): 1226004 <https://doi.org/10.3788/IRLA201948.1226004>

深度学习技术在条纹投影三维成像中的应用

Application of deep learning technology to fringe projection 3D imaging

红外与激光工程. 2020, 49(3): 0303018 <https://doi.org/10.3788/IRLA202049.0303018>

基于深度学习的机载激光海洋测深海陆波形分类

Classification of sea and land waveforms based on deep learning for airborne laser bathymetry

红外与激光工程. 2019, 48(11): 1113004 <https://doi.org/10.3788/IRLA201948.1113004>

多阶段深度学习单帧条纹投影三维测量方法

Multi-stage deep learning based single-frame fringe projection 3D measurement method

红外与激光工程. 2020, 49(6): 20200023 <https://doi.org/10.3788/IRLA20200023>

基于语义信息补偿全局特征的物体点云分类分割

林 森¹, 赵振禹^{2*}, 任晓奎², 陶志勇²

- (1. 沈阳理工大学 自动化与电气工程学院, 辽宁 沈阳 110159;
2. 辽宁工程技术大学 电子与信息工程学院, 辽宁 葫芦岛 125105)

摘要: 3D 点云数据处理在物体分割、医学图像分割和虚拟现实等领域起到了重要作用。然而现有 3D 点云学习网络全局特征提取范围小, 难以描述局部高级语义信息, 进而导致点云特征表述不完整。针对这些问题, 提出一种基于语义信息补偿全局特征的物体点云分类分割网络。首先, 将输入的点云数据对齐到规范空间, 进行数据的输入转换预处理。然后, 利用扩张边缘卷积模块提取转换后数据的每一层特征, 并叠加生成全局特征。而在局部特征提取时, 利用提取到的低级语义信息来描述高级语义信息和有效几何特征, 用于补偿全局特征中遗漏的点云特征。最后, 融合全局特征和局部高级语义信息得到点云的整体特征。实验结果表明, 文中方法在分类和分割性能上优于目前经典和新颖的算法。

关键词: 语义信息; 3D 模型分类、分割; 特征提取; 深度学习

中图分类号: TP391.41 **文献标志码:** A **DOI:** 10.3788/IRLA20210702

Object point cloud classification and segmentation based on semantic information compensating global features

Lin Sen¹, Zhao Zhenyu^{2*}, Ren Xiaokui², Tao Zhiyong²

- (1. School of Automation and Electrical Engineering, Shenyang Ligong University, Shenyang 110159, China;
2. School of Electronic and Information Engineering, Liaoning Technical University, Huludao 125105, China)

Abstract: 3D point cloud data processing has played an essential role in object segmentation, medical image segmentation, and virtual reality. However, the existing 3D point cloud learning network has a small global feature extraction range and cannot obtain local high-level semantic information, which leads to incomplete point cloud feature representation. Aiming at these problems, a classification, and segmentation network of object point cloud based on semantic information compensating global features was proposed. Firstly, align the input point cloud data to the specification space, and perform the preprocessing of the input conversion of the data. Then, the expanded edge convolution module was used to extract the features of each layer of the converted data and superimpose them to generate global features. In the local feature extraction, the extracted low-level semantic information was used to describe the high-level semantic features and effective geometric information, which was used to compensate for the missing point cloud features in the global features. Finally, the global feature and local

收稿日期: 2022-01-20; 修订日期: 2022-02-25

基金项目: 国家重点研发计划 (2018 YFB1403303)

作者简介: 林森, 男, 副教授, 硕士生导师, 博士后, 主要从事图像处理、人工智能等方面的研究。

通讯作者: 赵振禹, 男, 硕士生, 主要从事点云分类分割和图像处理方面的研究。

high-level semantic information were combined to obtain the overall feature of the point cloud. The experimental results show that the method in this paper is superior to the current classic and novel algorithms in classification and segmentation performance.

Key words: semantic information; 3D model classification, segmentation; feature extraction; deep learning

0 引言

近年来,3D 模型成像技术发展非常迅速,数据获取更加便捷。点云是 3D 模型数据中一种重要的表现形式,包含着大量的空间几何信息,使模型的描述更加直观和准确。3D 点云处理技术已广泛应用于智能车辆、模型重构、医学成像和遥感测绘等领域,成为计算机视觉和图形学领域的重要研究项目。因此,对 3D 点云数据处理显得至关重要。

传统点云处理方法主要通过手工设计几何形状^[1]描述符或签名描述符^[2]对特征进行提取。手工设计的描述符效果较差,导致提取到的特征数量变少,进而影响最终的实验结果。近年来,深度学习 3D 点云算法^[3-4]以处理数据量大等优势受到学者们的广泛关注。由于点云数据的稀疏性、非结构性和无序性,传统的卷积神经网络不适合直接应用于点云领域。针对此问题,Qi 等人^[5]提出 PointNet 网络,通过多层感知器直接学习点的特征,并结合基于对称函数思想的最大池化方法处理点云数据的无序性难题。该算法较好地解决了卷积神经网络直接应用的问题,但点云特征得不到完整地提取。为此,Qi 等人^[6]针对 PointNet 的不足之处进行了修改,设计出 PointNet++,首次提出基于最远点采样的局部特征提取。相较于 PointNet 有显著提升,但是仍缺乏点与点之间相互关系的表示。Li 等人^[7]提出点卷积网络(Point Convolutional Neural Networks, PointCNN)以解决点云的无序排列问题,该网络没有采用最大池化作为对称函数,而是训练了一个“X”型的变换网络,但是局部几何信息仍有大量的丢失。Wang 等人^[8]提出动态图卷积神经网络(Dynamic Graph Convolutional Neural Networks, DGCNN),通过最近 K 阶邻点采样(K Nearest Neighbor, KNN)获取局部信息,构造局部图结构并提取该部分特征,效果比 PointCNN^[7]更好,且未造成有效信息的丢失。尽管 DGCNN^[8]能够很好地采集低级语义信

息,仍无法描述大部分的高级语义信息和隐式高级语义特征。

针对上述研究方法的不足,文中提出一种局部语义信息补偿全局特征的点云学习网络,通过扩张卷积来增大特征提取的范围,在保持点云序列不变的情况下,利用边缘卷积提取几何特征。目的是尽可能考虑点的坐标与邻点的距离,避免部分几何信息提取不完整的问题。在局部特征提取时,利用 KNN 模块来提取低级语义信息,并利用局部特征融合(Vector of Locally Aggregated Descriptors, VLAD)模块将所提取到的低级语义信息来描述高级语义信息和隐式高级语义特征,进一步补偿全局提取时遗漏的特征及有效信息。实验及分析表明,文中网络对点云的全局及局部特征提取效果具有显著的提高,低级语义信息描述高级语义特征也得到了进一步的完善。

1 点云分类、分割网络

1.1 网络体系结构

图 1 为网络的总体架构,由点云特征提取学习、分类处理和分割处理三部分组成。其中点云特征提取学习为核心部分,首先将所输入的点云数据通过规范空间(Spatial Transformer Network, STN)模块进行输入转换处理;其次利用扩张边缘卷积(Dilated-Edge conv)模块提取各层点云特征,并汇总成全局特征;最后利用 KNN-VLAD 模块提取点云局部特征,与全局特征进行融合生成更完整的点云特征。

1.2 空间转换

3D 点云数据是由一组排列无序的 3D 点 $P = \{p_1, p_2, \dots, p_N | p_n \in \mathbf{R}^3\}$ 组成,其中各点坐标为 (x, y, z) ,将这组点进行空间转换处理,定义如下:

$$\tilde{P} = PH_{\theta} \quad (1)$$

式中: \tilde{P} 为空间中转换后的 3D 点云数据; H_{θ} 为转换后的矩阵; H_{θ} 可以表示为:

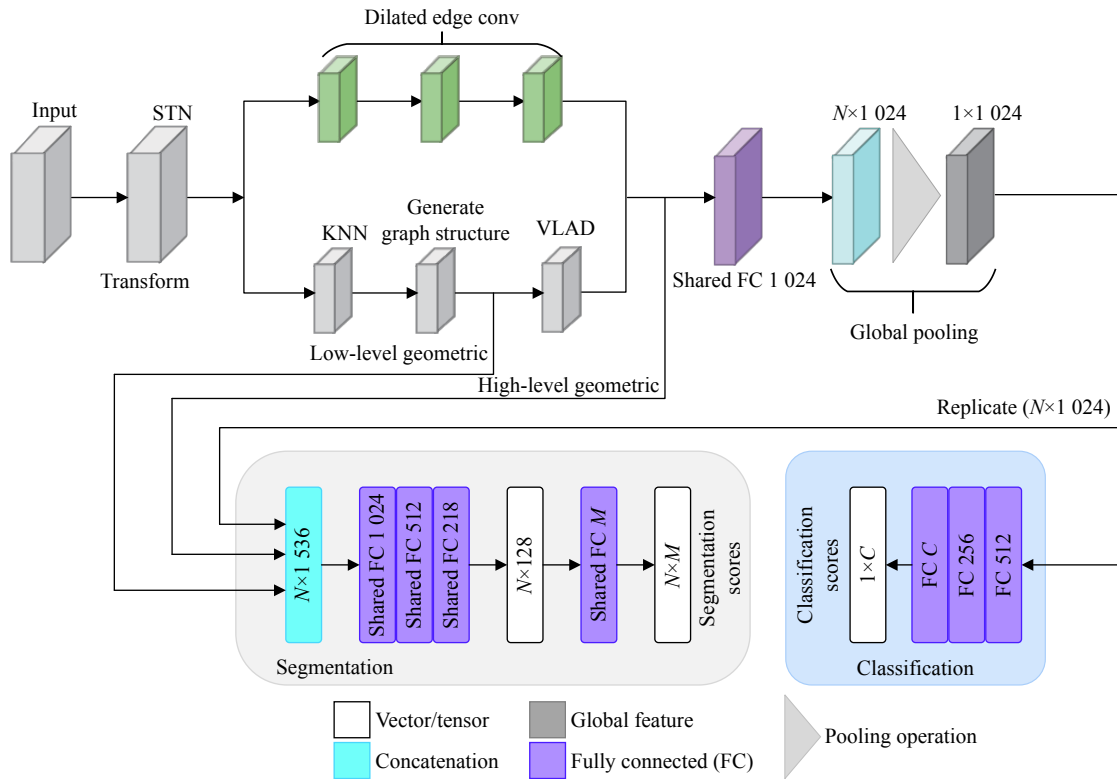


图 1 网络结构及流程示意图

Fig.1 Schematic diagram of network structure and process

$$H_{\theta} = \begin{bmatrix} \theta_{11}, & \theta_{12}, & \theta_{13} \\ \theta_{21}, & \theta_{22}, & \theta_{23} \\ \theta_{31}, & \theta_{32}, & \theta_{33} \end{bmatrix} \quad (2)$$

Jaderberg 等人^[9]提出利用深度学习对齐 2D 图像的空间转换,文中对此进行创新改进并运用至 3D 点云数据上,通过 STN 模块对点云数据进行输入转换处理。该模块通过端到端的方式将输入点云进行空间规范,保证其输入后依然保持空间序列的不变性。图 2 为经 STN 模块输入转换后的矩阵。

图 2 中, T-Net 表示一个回归网络,将输入的点云数据转化为含有 9 个参数的矩阵 H_{θ} 输出。STN 模块将输入的点云数据编排至一个规范的矩阵,不改变点云数据排列方式并传到下一模块进行处理,通过规范空间的排列解决了点云数据的无序性。

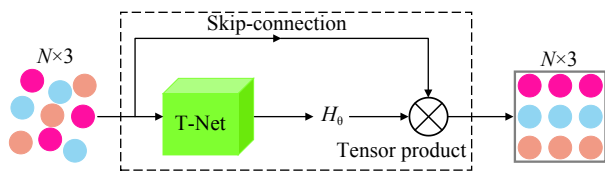


图 2 STN 输入转换结构图

Fig.2 Diagram of STN input conversion structure

1.3 最近 K 阶邻点采样-KNN

KNN(K Nearest Neighbor)^[10]表示:该算法通过采用测量不同特征值之间的距离方法进行分类。其优点在于精度高,对于异常值不敏感以及无需数据输入假定。

经 STN 模块转换后,原始 3D 点集 $P = \{p_1, p_2, \dots, p_N | p_n \in \mathbf{R}^3\}$ 转换为规范空间的 3D 点集 $\tilde{P} = \{\tilde{p}_1, \tilde{p}_2, \dots, \tilde{p}_N | \tilde{p}_n \in \mathbf{R}^3\}$ 。将规范好的 3D 点集输入到 KNN 模块,提取低级几何特征, KNN 模块结构如图 3 所示。图中,在 3D 点集 $\tilde{P} - \tilde{p}_n$ 中的每个点采集 $K(K = 16)$ 阶最近邻点。 k 阶采集可表示为:

$$\hat{P}_{n,k} = kNN(\tilde{p}_n | (\tilde{P} - \tilde{p}_n)) \quad (3)$$

式中: $k \in K$ 且 $k \leq K = 16$; $\hat{P}_{n,k}$ 表示点 \tilde{p}_n 的第 k 个最近邻点。因此, \tilde{p}_n 的 K 个最近邻集可以表示为:

$$\{\tilde{p}_{n,1}, \tilde{p}_{n,2}, \dots, \tilde{p}_{n,K} | \tilde{p}_{n,k} \in \mathbf{R}^3\} \quad (4)$$

在该网络中 KNN 模块有两个作用:(1) VLAD 模块利用 KNN 所采集到的低级语义信息来描述高级语义信息和隐式高级语义特征。(2) 扩张边缘卷积模块

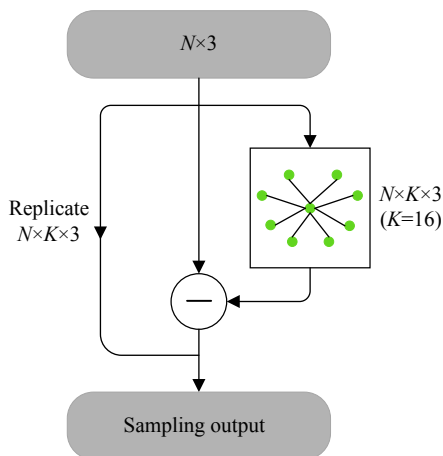


图 3 KNN 模块的结构

Fig.3 The structure of KNN module

利用 KNN 采集得到的点集,并提取点的有效信息来构造全局特征图。

1.4 局部高级语义特征描述

VLAD^[11]是一种经典的实例检索和图像分类的描述符池化方法。直接描述点云高级语义特征的隐式表达式是极为困难的,为解决这一问题文中利用 KNN 所提取到的低级语义信息描述符与一些视觉词之间的关系,来间接描述高级语义特征。VLAD 结构如图 4 所示,改进的 VLAD 模块主要有两个功能:(1) Top-K 特征的选择;(2) 特征的变化与融合。

Top-K 特征选择: 选择 N 个低级语义信息描述符 $\{v_1, \dots, v_N | v_n \in \mathbf{R}^D\}$ (其中 $D = 128$) 作为 VLAD 的输入, M 个视觉词 (视觉词可以为“集群中心”) 初始化, 表示

为 $\{c_1, \dots, c_M | c_m \in \mathbf{R}^D\}$, 将可学习的参数反向传播。把各点的低级语义信息描述符 v_n 分配给各个 c_m , 可以使用残差向量 $(v_m - c_m)$ 表示, 并记录二者之间的差异。点的低级语义信息描述符 v_n 与视觉词之间的关系 r 可以表示为:

$$r_{n,d} = \sum_{m=1}^M a_n(c_m)(v_{n,d} - c_{m,d}), d \in [1, D] \quad (5)$$

式中: $a_n(c_m)$ 是注意系数; $c_{m,d}$ 和 $v_{n,d}$ 分别是 d 维中的第 m 视觉词和第 n 个点的低级语义信息描述符。权重可以利用注意系数 $a_n(c_m)$ 表示, 突出第 m 个视觉词和第 n 个点的低级语义信息描述符的重要性。 c_m 和 v_n 的距离最近时 $a_n(c_m) = 1$, 反之则为 0。

文中对视觉词和低级语义信息描述符进行 softmax 处理, 目的是让 VLAD 模块的注意系数更容易在不同视觉词之间比较, 注意系数的计算表达式为:

$$a_n(c_m) = \frac{e^{W_m^T v_n + b_m}}{\sum_{m=1}^M e^{W_m^T v_n + b_m}} \quad (6)$$

式中: w_m 表示权重; b_m 表示偏差; $a_n(c_m)$ 将最高权重分配给距离最近的视觉词, 取值范围为 $[0, 1]$ 。公式 (5) 是每个视觉词的残差加权和, 由于 3D 模型组各点可能与多个视觉词相关联, 需要分析较高注意系数的视觉词对高级语义特征的影响。则公式 (6) 应设定一个低级语义信息描述符 v_n 和一个视觉词集合 $C = \{c_1, \dots, c_M | c_m \in \mathbf{R}^D\}$, Top-K 的注意力数值返回一个子集 $\tilde{C} = \{\tilde{c}_1, \dots, \tilde{c}_K | \tilde{c}_k \in \mathbf{R}^D\}$, $\tilde{C} \in C$, 使得任意一个视觉词都

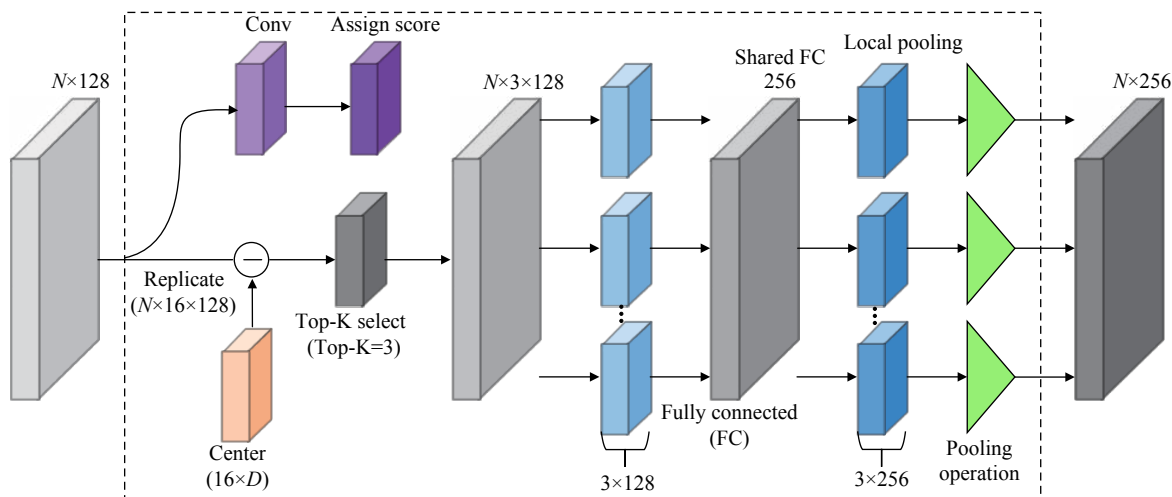


图 4 VLAD 结构图

Fig.4 VLAD structure diagram

有 $\tilde{c} \in \tilde{C}$ 并且 $\tilde{c} \text{ simin} C - \tilde{C}, a_n(\tilde{c}) \geq a_n(\hat{c})$ 。将公式 (6) 改进可得:

$$r_{n,d} = \sum_{k=1}^{Top-K} \frac{e^{W_k^T v_n + b_k}}{\sum_{k=1}^{Top-K} e^{W_k^T v_n + b_k}} (v_{n,d} - C_{k,d}) \quad (7)$$

式中: $k \in [1, top - K], d \in [1, D]$ 。Top-K 既可以控制剩余向量的数量, 又可以表示不同视觉词之间的重叠。引入共享 FC 层以便改善网络的非线性变化, 最后融合 Top-K 变换后的特征。

图 4 中, 在网络初始化过程中, 通过 $[-0.01, 0.01]$ 的统一初始化得到视觉单词。然后, 在网络训练过程中, 通过优化损失函数对其进行不断调整。将视觉词 $c_m \text{ softmax}$ (软分配) 给低级几何特征描述符 v_n 可以视为一个两步过程: (1) 与一组 M 个滤波器 $\{w_m\}$ 卷积, 具有空间支持 1×1 和偏差 $\{b_m\}$, 产生输出 $w_m^T v_n + b_m$ 。(2) 卷积输出通过 softmax 传输函数获得最终的软分配 $a_n(c_m)$ 。Top-K 的特征将根据该定义来执行选择。

文中将 KNN 和 VLAD 模块进行融合, KNN 模块负责提取点云数据中的低级语义信息, VLAD 模块将 KNN 提取到的信息进行整合并描述点云特征中难以提取到的高级语义信息和隐式高级语义特征。两个模块相辅相成, 共同描述全局特征中遗漏的有效信息, 为整体特征完整性的表示起到了决定性的作用。

如图 5 所示, 第一列为真实物体分割结果, 第二列为无局部补偿全局特征的分割结果, 第三列是含有局部补偿全局特征的分割结果, 依据图中红色方框标记处可知。当不含有局部补偿全局特征时, 虽然能够大致分割标记出物体各部分组织的形状, 但是由于缺少局部高级语义信息和一些隐式高级语义特征地补

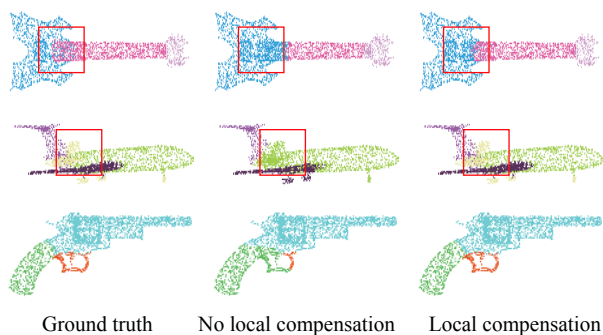


图 5 局部补偿的对比

Fig.5 Comparison of partial compensation

偿, 导致物体的一些细微地方出现分割标记错误的情况。例如, 吉他的琴弦处、飞机的尾翼处以及手枪的扳机处。由此可见, 局部高级语义信息以及隐式高级语义特征补偿全局特征能够更好地表示物体各部分特征的细节, 充实整体特征的表示。

1.5 全局特征的提取

$X_i = (x_i, y_i, z_i) \subseteq \mathbf{R}^3$ 表示经 STN 模块处理后点云数据中各个点的 3D 坐标, 还应该包含相应的颜色、曲面法线等附加坐标。点集经 KNN 处理后, 会生成一个点集结构的有向图。

文中建立一个点集结构的有向图 $G = (V, E)$, 其中 $V = \{1, \dots, n\}, \varepsilon \subseteq V \times V$ 分别表示顶点和边。在 \mathbf{R}^3 中 G 为 X_i 的自循环 KNN 特征图, 目的是让每个节点指向自己。 $e_{ij} = h_{\Theta}(X_i, X_j)$ 定义为边缘特征, h_{Θ} 为一组可学习非线性函数。最后, 通过对每个点发出的所有与边相关联的边缘特征用于信道对称聚合操作来定义边缘卷积。第 i 个顶点处输出的边缘卷积为:

$$X'_i = \sum_{j:(i,j) \in \varepsilon}^{\max} h_{\Theta}(X_i, X_j) \quad (8)$$

式中: $\Theta = (\theta_1, \dots, \theta_M, \Phi_1, \dots, \Phi_M)$ 。把 X_i 作为中心像素, $\{X_j : (i, j) \in \varepsilon\}$ 表示其周围补偿, 边缘卷积特征提取详见图 6, 边缘卷积提取流程详见图 7。

在图 6 左边计算出 X_i 和 X_j 边缘特征 e_{ij} , $h_{\Theta}()$ 表示为全连接层实例化中的相关权重。在图 7 右边是边缘卷积层提取特征并汇总成全局特征操作, 通过聚合每个相连的顶点所发出的全部相关联的边来计算边缘特征并将其作为输出。

结合全局几何形状结构, 以 X_i 中心点坐标作为补偿, 局部信息由 $X_j - X_i$ 获取, 可表示为:

$$e'_{ijm} = \text{ReLU}(\theta_m \cdot (X_j - X_i) + \Phi_m \cdot X_i) \quad (9)$$

利用公式 (11) 可以实现多层感知器的共享:

$$X'_{im} = \max_{j:(i,j) \in \varepsilon} e'_{ijm} \quad (10)$$

如图 8 所示为边缘卷积结构图, 其中 $n \times f$ 表示输入数据的张量, 通过应用多层感知器 (Multi-Layer Perceptron, MLP) 来计算每个点的边缘特征, 其中 $\{a_1, a_2, \dots, a_n\}$ 表示每层的神经元数量, $n \times a_n$ 表示经相邻边缘特征之间池化后所生成的形状张量。

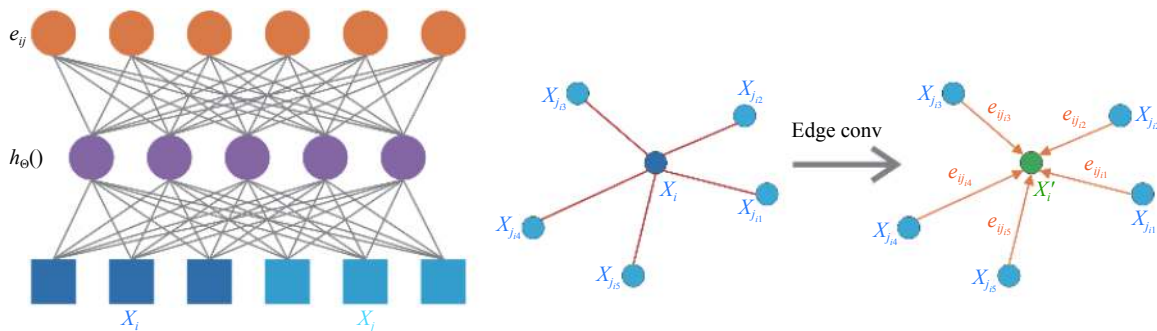


图 6 边缘卷积特征提取示意图

Fig.6 Schematic diagram of edge conv feature extraction

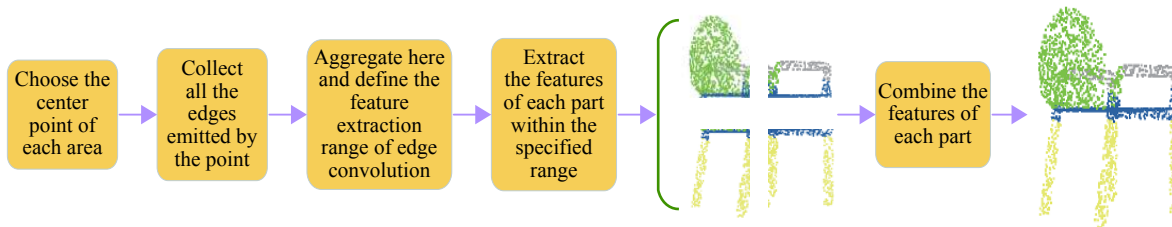


图 7 边缘卷积特征提取流程图

Fig.7 Edge conv feature extraction flowchart

扩张卷积^[12]是在标准的卷积核中注入空洞,以此来增加模型的感受野。相比原正常卷积操作,扩张卷积多了一个参数:扩张率,指的是卷积核的点的间隔数量,比如常规的卷积操作扩张率为 1。扩张卷积不仅能够减轻池化层,还能更好地解决连续池化所造成信息丢失的问题。图 9 所示为扩张卷积的原理示意图,在网格中的结构化图形中内核为 2,从左至右扩张率为 1、2(底部扩张率相同)展开二维动态卷积。

在文中的算法中,利用扩张卷积来扩大 KNN 采样的感受野,便于增大边缘卷积提取的范围。每提取一层全局特征,就要寻找其邻点并构造图结构。将扩张后的 KNN 和步长 d 作为图 $G = (V, E)$ 的离散率的输入, KNN 越过 d 个邻点,之后在返回区域为 $k \times d$ 的大

小内,以 12 为步长确定最近的 k 个邻点。

图 10 中,扩张卷积扩大了特征提取的范围,并不增加额外多余输入,它的内部结构与边缘卷积的结构基本一致。

$$N^{(d)}(v) = \{u_1, u_{1+d}, u_{1+2d}, \dots, u_{1+(k-1)d}\} \quad (11)$$

设 $N^{(d)}(v)$ 为顶点的 d 扩张邻域,如果 $(u_1, u_2, \dots, u_{k \times d})$ 为第一个序列的 $k \times d$ 最近相邻区域,则顶点

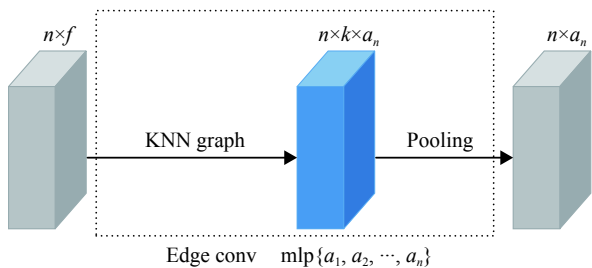


图 8 边缘卷积结构图

Fig.8 Edge Conv structure diagram

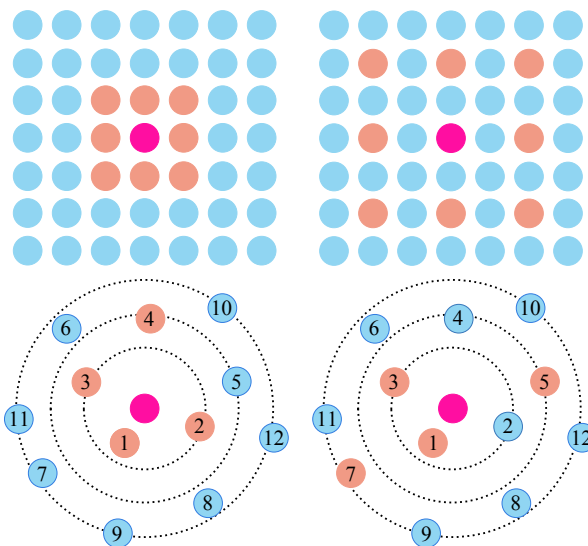


图 9 动态图卷积中的扩张卷积

Fig.9 Dilated conv in dynamic graph convolution

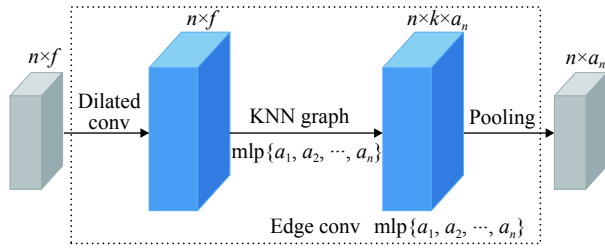


图 10 扩张边缘卷积结构图

Fig.10 Structure diagram of expanded edge conv

$(u_1, u_{1+d}, u_{1+2d}, \dots, u_{1+(k-1)d})$ 是顶点 v 的 d 扩张相邻序列。

扩张边缘卷积改善了 KNN 提取视野范围受限的问题,大大提高了全局特征的提取范围,增大了点云特征信息的数量。如图 11 所示,三组图像分别是经过扩张边缘卷积和边缘卷积特征提取后的物体可视化图像。

根据图 11 中红色方框处标记可知,经扩张边缘

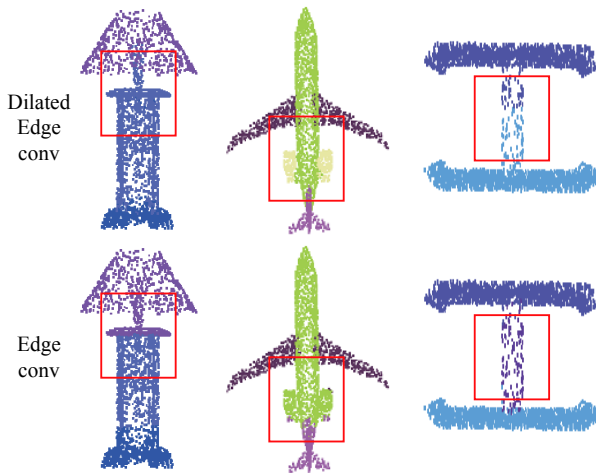


图 11 特征提取的对比

Fig.11 Comparison of feature extraction

卷积处理后的可视化图像,更能准确地标记分辨出物体的各部分组织特征,在灯台、飞机涡轮扇和桌子支杆处更准确的进行识别。而经边缘卷积处理后的可视化图像并不能准确的进行标识,这是由于边缘卷积的特征提取范围较小,在物体各部分组织提取时丢失了大量的几何信息,进而导致最终融合后丢失了大量的全局特征。物体各部分的几何信息给其特征提取提供了重要的信息描述,所以为尽可能避免几何信息的丢失,文中将扩张卷积和边缘卷积进行融合更完整地提取点云的整体特征。

2 实验与分析

2.1 实验数据集

在物体点云分类实验中,文中采用的是普林斯顿大学制作的用于 3D 模型识别的标准数据集 Model Net40^[13],该数据集是一系列网格化的 3D-CAD 模型,包含 40 个类别的 12311 个 CAD 模型,其中有 9843 个训练模型,2468 个测试模型。

在物体点云分割实验中,文中采用的是 ShapeNet part^[14]数据集。其中数据集包含 16881 个点云模型,共 16 个形状类别;其中 14006 个模型用于训练,2875 个模型用于测试,不同类别的物体被分割成各个部分并且没有重叠,将每个模型中随机采样出 2048 个点进行物体分割实验。

2.2 实验平台参数设置

为了保证实验的统一性和真实性,文中采用相同的实验平台、系统以及编译环境。基于 Ubuntu 和 Cuda11.1 建立的深度学习环境下进行,实验中所使用的深度学习网络框架和相关环境配置以及各部分模型参数详如表 1 所示。

表 1 实验平台配置

Tab.1 Experimental platform configuration

Environment configuration		Model parameters	
Name	Configuration	Name	Value
CPU	Intel i7-10700 F	Batch size	32
GPU	RTX3090	Number point	1024
RAM	32 G	Max epoch	250
Operation system	Ubuntu18.04	Optimizer	Adam
Language	Python 3.7	Learning rate	0.001
Learning framework	TensorFlow GPU 1.15.0	Momentum	0.9

2.3 分类实验

为证明文中算法在处理点云分类任务上的优势,在 ModelNet40^[29]数据集上进行实验对比,其中对比算法包括: PointNet^[5]、PointNet++^[6]、DGCNN^[8]、PointCNN^[7]、Pointwise^[15]、StructuralRelationalReasoning PointNet (SRN- PointNet)^[16]、Graph Geometric Moments

Net (GGM)^[17]、Multi-scale Dynamic Graph CNN (MSDGCNN)^[18]和 EllipsoidNet (EllNet)^[19]。通过对比评估准确度 (Eval accuracy) 和平均分类精确度 (Avg class acc),来定量分析各算法的优越性。各算法分类实验结果详如表 2 所示。

表 2 ModelNet40^[13] 各分类算法实验对比

Tab.2 Experimental comparison of various classification algorithms of ModelNet40^[13]

Method	Representation	Input	Eval accuracy	Avg class acc
PointNet ^[5]	Points	1 024×3	89.2%	86.2%
PointNet++ ^[6]	Points(+normal)	1 024×(3+3)	90.7%	87.8%
DGCNN ^[8]	Points	1 024×3	92.2%	88.9%
PointCNN ^[7]	Points	1 024×3	91.7%	88.5%
Pointwise ^[15]	Points	1 024×3	91.6%	89.1%
SRN-PointNet ^[16]	Points	1 024×3	91.5%	88.6%
GGM ^[17]	Points	1 024×3	92.5%	89.0%
MSDGCNN ^[18]	Points	1 024×3	91.8%	88.3%
EllNet ^[19]	Points	1 024×2	92.6%	89.0%
Ours	Points	1 024×3	92.7%	89.3%

表 2 中,除了 PointNet++的输入为 Points+normal 的 1024×(3+3)数据、EllNet 输入的是 1024×2 二维点云数据,其余算法采用的均是标准 Points 的 1024×3 数据。文中算法在评估准确度和平均分类精确度数值上均优于其他 9 种算法,其中评估准确率为 92.7%、平均分类精度为 89.3%。PointNet 和 PointNet++ 虽然尽可能的提取点云的全局特征,但是只能考虑到全局和单个点的特征,缺少局部几何信息以及高级语义特征的描述,所以其分类的实验效果较差,最高的评估准确度仅为 90.7%。尽管 PointCNN、Pointwise 以及 SRN-PoinNet 将 PointNet 算法进行了提升,分类实验的效果也大幅提升但是相较于文中的算法还是有一定的差距,最高的评估准确度也只有 91.7%。在 DGCNN 和 MSDGCNN 这两个动态图卷积算法中,该实验结果有所提高,但局部有效信息仍有所丢失,最高的评估准确度仅为 92.2%。在 EllNet 和 GGM 算法中, EllNet 算法为提取潜在的局部特征,将点云数据的特征投影到一个二维椭圆空间上,但是仍会造成部分隐式的高级语义特征丢失,而且这两个算法所提取的全局特征也会因为投影的过程造成丢失,最终导致

评估准确度比文中算法低了 0.1%。GGM 算法是从局部点集入手,学习其几何特征,利用显式编码点构造几个特征进而生成表面几何特征,这种以局部特征构造全局特征的方法会对点云的真实几何信息造成丢失,而且点云隐式特征依然没能提取到,从而导致评估准确度仅有 91.8%。相比而言,文中算法不仅提高了全局特征的提取范围,还通过局部提取到的高级语义信息和有效几何特征进一步完善了整体特征。通过分类实验对比,文中算法更具优越性。

文中选取最具代表性的 PointNet、PointNet++、DGCNN、GGM 网络作为对比,通过测试点数为 64、128、256、512、1024 的稀疏点集数据,进一步测验该网络对点云数据集密度的稳定性如图 12 所示。

图 12 中, PointNet 和 DGCNN 网络随着点集密度的增加,分类准确度提升的幅度极大,在处理稀疏点云数据时分类准确度极低。PointNet++网络在密度为 64~128 的点集之间分类准确度变化较大,在密度为 128~512 之间增长较为缓慢,当密度超过 512 时逐渐趋于稳定。GGM 网络和文中方法的分类准确度整体增长幅度较小,当密度大于 512 之后分类准确度基

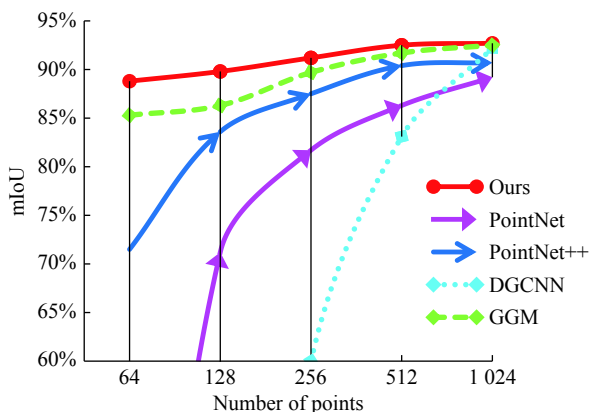


图 12 点数与分类精度的关系

Fig.12 Relationship between points and classification accuracy

本保持不变,但各点集密度的分类准确度都要高于 GGM。文中方法不随点集的密度变化而过度影响分类准确度,在点集密度呈倍数增减的状态下分类准确度的整体波动范围不超过 4%,而且不同密度的点集经本方法处理后的分类准确度仍然高于其他方法。

为了更好地测评本网络的性能,选取更具代表性的 PointNet、PointNet++、DGCNN、PointCNN 作为对比方法,通过测评模型大小、测试耗时及精度进行定量分析。从表 3 可以清晰地看出该网络复杂度低且模型轻量化、耗时较短,在分类准确度上更是优于其他算法。

表 3 各算法测试时间对比

Tab.3 Comparison of algorithm test time

Method	Size/MB	Time/s	Accuracy
PointNet ^[5]	40	78.9	89.2%
PointNet++ ^[6]	12	163.2	90.7%
DGCNN ^[8]	21	89.7	92.2%
PointCNN ^[7]	94	117.0	91.7%
Ours	21	86.4	92.7%

2.4 分割实验

在分割实验中,采用 ShapeNet part^[14] 数据集来评估各算法的物体分割效果,与分类对比算法保持一致。

利用各个类别的 IoU(%) 和总体分割精度 mIoU (%) 作为定量分析的评价度量。其中 IoU 的求值表达式为:

$$I_{IoU} = \frac{c_{ii}}{c_{ii} + \sum_{j \neq i} c_{ij} + \sum_{k \neq i} c_{ki}} \quad (12)$$

各算法分割测试结果如表 4 所示,通过实验数据可知,经文中算法处理后的数据集物体分割总体分割精度 mIoU(%) 的值是 85.5%,而且在 16 个类别中,文中算法有一半以上类别的精度达到了最优效果。虽

表 4 各算法在 ShapeNet Part^[14] 数据集上局部分割测试结果

Tab.4 Local segmentation test results of each algorithm on the ShapeNet Part^[14] data set

Method	mIoU	Shapes IoU															
		Plane	Bag	Cap	Car	Chair	Earcup	Guitar	Knife	Lamp	Laptop	Motor	Mug	Pistol	Rocket	Skate	Table
PointNet ^[5]	83.7%	83.4%	78.7%	82.5%	74.9%	89.6%	73.0%	91.5%	85.9%	80.8%	95.3%	65.2%	93.0%	81.2%	57.9%	72.8%	80.6%
PointNet ^[6]	85.1%	82.4%	79.0%	87.7%	77.3%	90.8%	71.8%	91.0%	85.9%	83.7%	95.3%	71.6%	94.1%	81.3%	58.7%	76.4%	82.6%
DGCNN ^[8]	85.2%	84.0%	83.4%	86.7%	77.8%	90.6%	74.7%	91.2%	87.5%	82.8%	95.7%	66.3%	94.9%	81.1%	63.5%	74.5%	82.6%
PointCNN ^[7]	86.1%	84.1%	86.4%	86.0%	80.8%	90.6%	79.7%	92.3%	88.4%	85.3%	96.1%	77.2%	95.3%	84.8%	64.2%	80.0%	83.0%
Pointwise ^[15]	85.1%	82.9%	80.7%	87.8%	76.6%	90.8%	79.2%	91.0%	86.6%	83.3%	95.3%	71.9%	94.4%	80.9%	62.0%	75.1%	82.5%
SRNPNet ^[16]	85.3%	82.4%	79.8%	88.1%	77.9%	90.7%	69.6%	90.9%	86.3%	84.0%	95.4%	72.2%	94.9%	81.3%	62.1%	75.9%	83.2%
GMM ^[17]	85.2%	83.9%	82.8%	88.0%	79.8%	90.7%	76.8%	91.3%	87.6%	82.6%	95.5%	66.6%	94.8%	81.8%	62.6%	73.8%	82.6%
MSDGCN ^[18]	85.4%	83.7%	84.7%	87.5%	77.0%	90.8%	68.2%	91.5%	86.5%	96.0%	95.5%	72.0%	95.1%	83.4%	61.9%	77.4%	82.9%
EIL-Net ^[19]	85.0%	82.8%	81.5%	87.6%	76.8%	90.6%	78.8%	90.8%	86.8%	86.9%	95.1%	71.8%	94.2%	80.8%	61.8%	75.0%	82.2%
Ours	85.5%	84.3%	85.8%	88.1%	80.0%	90.8%	79.5%	91.6%	88.2%	91.6%	95.8%	76.7%	96.1%	82.6%	65.6%	81.2%	82.8%

然在 mIoU 数值上文中相比较 PointCNN 相差了 0.6%，但其算法需要增加信息量来提高精度。相比而言文中算法处理点云数据更简洁，计算成本也比较低。

18 个类别里随机抽取一个例子进行可视化展示，并与 PointNet^[5] 算法及标准实例图 (Ground Truth) 进行比较，详情如图 13 所示。

ShapeNet Part 数据集共分为 18 个类别，依次从

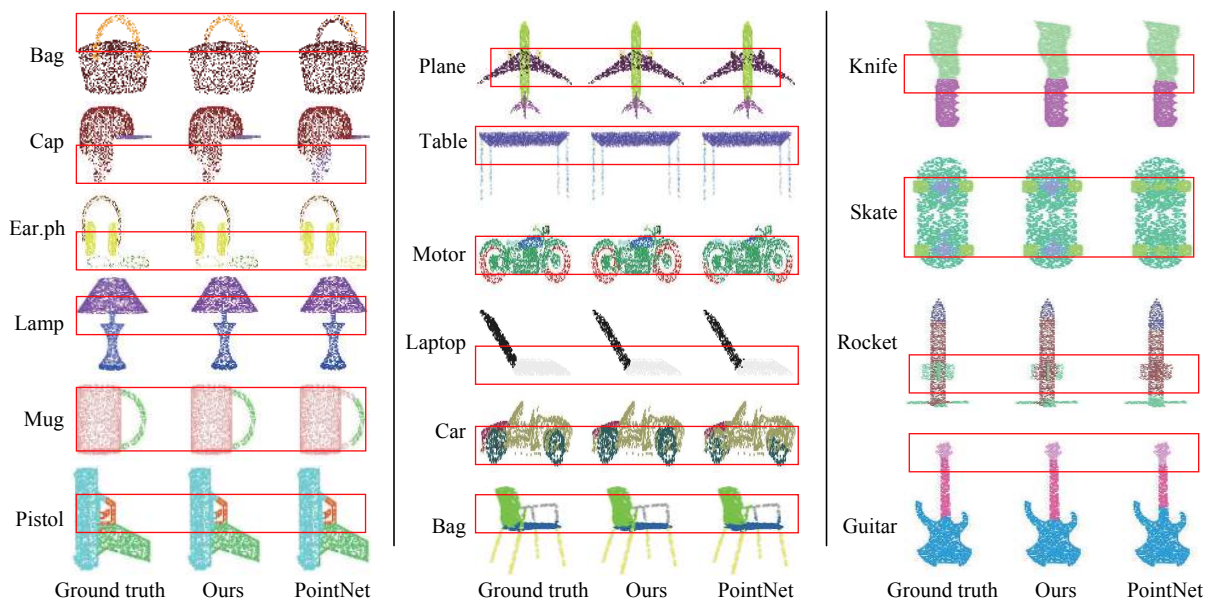


图 13 物体分割测试结果可视化

Fig.13 Visualization of object segmentation test results

如图 13 所示，依据红色方框标记处的各类别分割可视化对比展示。文中算法的可视化结果基本接近于标准分割图例，更接近于真实分割效果。然而经 PointNet 算法分割后的可视图出现了较为严重的错误现象，例如 bag 中的手袋部分不能进行准确的分割标记，这是由于 PointNet 算法不能完整地提取点云全部特征所导致的。根据主观可视化视图对比和客观数据对比，文中网络能大幅度提高物体分割的准确率。

为更好地验证文中算法能够应用于更多复杂的点云数据，对稀疏点集物体，大型复杂场景 (其中大型复杂场景包含遮挡场景和遮挡物体数据) 进行分割测试。如图 14、15 所示，分别表示稀疏点集物体和大型复杂场景的分割可视化结果图。文中算法不仅能够权威的 ShapeNet part^[14] 标准数据集下进行准确分割，还能够精准地分割稀疏点集和大型复杂场景，分割的效果更具优越性。

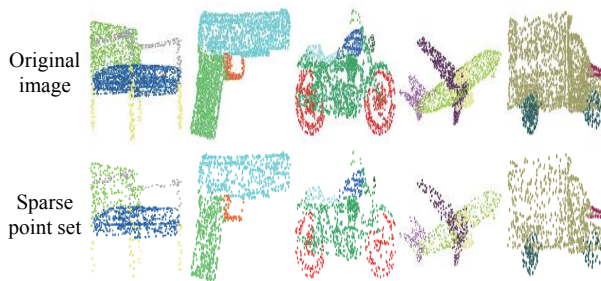


图 14 物体稀疏点图分割结果

Fig.14 Object sparse point map segmentation results

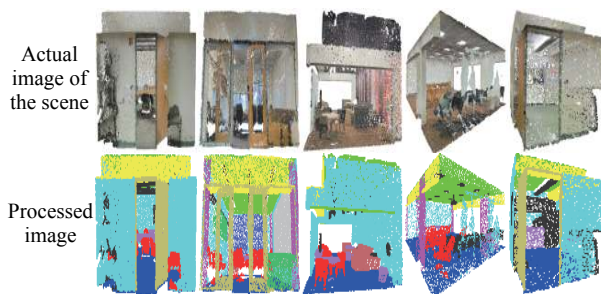


图 15 复杂场景分割结果

Fig.15 Complex scene segmentation results

图 16 所示为点集密度与分割精度的关系，图 17 为不同点集密度下的分割可视图。文中分别从点集的不同位置减少点的数量并依次进行分割精度的测

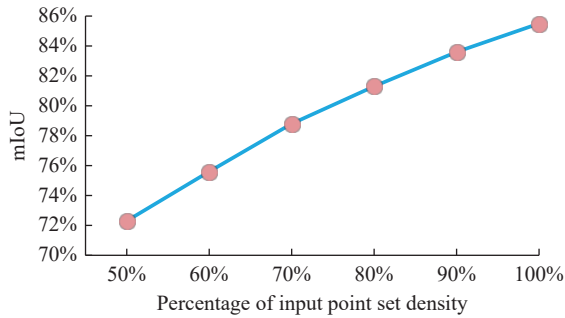


图 16 密度与精度的关系

Fig.16 Relationship between density and accuracy



图 17 不同点集密度下的分割可视图

Fig.17 Segmentation viewable under different point set densities

试,实验结果表明,随着点数的增多分割的精度也逐渐增大,当物体点集数量为整体密度的 100% 时分割精度达到最大。该算法在分割点集密度不足 100% 的数据时仍能将物体的各部分信息表示出来,由此可见在处理残缺或遮挡点云数据时算法依然可以达到较好的准确度。

2.5 消融实验

为了优化网络,寻求最佳参数组合,进行消融实验分析,依然采用标准的 ModelNet40^[13] 数据集进行实验操作。

局部采样的关键在于感受野的大小,最佳的感受野能更好的对局部特征进行感知。感受野过大会造成特征提取的不细致,而感受野过小则会造成细节信

息的丢失。感受野的大小取决于邻近点的数量,通过测试不同数量的最近邻点对结果的影响来进行实验分析,进而得到最佳的 K 值。

由表 5 可知,当 K 值为 20 时分类评估准确度最高,表明当前感受野的大小合适,局部特征提取最佳。当 K 值大于 20 时,分类评估准确度开始下降。实验结果表明,当最近邻点数为 20 时,局部感受野最佳,能够尽可能的将局部信息提取完整。

表 5 最佳 K 值的分析

Tab.5 Analysis of the best K value

KNN point number	Avg class acc	Eval accuracy
16	89.5%	91.4%
20	89.8%	92.7%
25	89.1%	91.3%
30	88.6%	91.1%
32	88.5%	90.8%

在文中网络中,扩张边缘卷积模块和 KNN-VLAD 模块是提取点云全局几何信息特征和局部高级语义特征的模块,可以大幅度提升网络的性能。扩张边缘卷积模块通过扩大特征提取的范围,进一步增加提取点云各处的几何信息数量,最终汇总成全局特征。KNN-VLAD 模块通过 KNN 采样获取点与点之间的联系,VLAD 模块利用已经提取的低级语义信息来描述高级语义信息,补偿全局提取特征时遗漏的有效信息。

如表 6 所示,若网络中仅含扩张边缘卷积模块时,此时只能全局特征的提取,缺乏重要的局部特征的补偿,由于整体特征不完整导致分类评估准确度仅有 92%。如果网络中仅含 KNN-VLAD 模块时,由于只有对局部特征进行细致提取而缺乏全局特征的提取及汇总,其分类评估准确度只有 90.8%。而当 2 个模块同时工作时才能提高网络的性能,分类评估准确

表 6 模块影响的分析

Tab.6 Analysis of module impact

Dilated-Edge conv	KNN-VLAD	Avg class acc
√	×	92.0%
×	√	90.8%
×	√	92.7%

度才能达到最高,其数值为 92.7%。经消融实验分析,证明了上述模块的重要性及有效性。

3 结 论

针对目前物体点云分类分割网络提取点云特征不完整的问题,文中提出了基于语义信息补偿全局特征的物体点云分类和分割的网络。首先将待处理的数据输入至 STN 模块,使数据转换到对齐的规范空间中,保证输入数据的排列方式不变。其次,通过扩张边缘卷积模块层层提取点云数据的各部分特征,并汇总成全局特征。在局部特征的提取中,采用 KNN-VLAD 模块将提取到的低级几何信息来尽可能地描述高级语义特征,这些特征和有效信息用于补偿全局遗漏的特征,这两个模块保证了点云特征的完整性。最后,将局部特征和全局特征融合生成整体特征。实验结果表明,文中算法不仅能够有效地提升物体点云数据的分类和分割的准确度,而且在处理稀疏、残缺及复杂场景数据时仍能保持准确度的稳定;通过实验对比,文中算法的网络结构复杂度低、耗时短,相较于其他算法更具优越性。

参考文献:

- [1] Osada R, Funkhouser T, Chazelle B, et al. Shape distributions [J]. *ACM Transactions on Graphics (TOG)*, 2002, 21(4): 807-832.
- [2] Sun J, Ovsjanikov M, Guibas L. A concise and provably informative multiscale signature based on heat diffusion[C]//Computer graphics forum. Oxford, UK: Blackwell Publishing Ltd, 2009, 28(5): 1383-1392.
- [3] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep conv neural networks [J]. *Communications of the ACM*, 2017, 60(6): 84-90.
- [4] Pan X Z, Zhang S Q, Guo W P. Application of multi-mode deep convolutional neural network to video expression recognition [J]. *Optics and Precision Engineering*, 2019, 27(4): 230-237. (in Chinese)
- [5] Qi C R, 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 (CVPR), 2017: 77-85.
- [6] Qi C R, Li Y, Hao S, et al. PointNet++: Deep hierarchical feature learning on point sets in a metric space[C]//Advances in Neural Information Processing Systems, 2017, 30: 5099-5108.
- [7] Li Y, Bu R, Sun M, et al. Pointnet: Convolution on X-transformed points[C]//Advances in Neural Information Processing Systems, 2018, 31: 820-830.
- [8] Wang Y, Sun Y, Liu Z, et al. Dynamic graph CNN for learning on point clouds [J]. *ACM Transactions on Graphics (TOG)*, 2019, 38(5): 1-12.
- [9] Jaderberg M, Simonyan K, Zisserman A. Spatial transformer networks[C]//Advances in Neural Information Processing Systems, 2015, 28: 2017-2025.
- [10] Hastie T, Tibshirani R. Discriminant adaptive nearest neighbor classification and regression[C]//Advances in Neural Information Processing Systems, 1996, 9: 409-415.
- [11] Jégou H, Douze M, Schmid C, et al. Aggregating local descriptors into a compact image representation[C]//2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE, 2010: 3304-3311.
- [12] Yu F, Koltun V. Multi-scale context aggregation by dilated convolutions [J]. *arXiv preprint arXiv*, 2015: 1511.07122.
- [13] Wu Z, Song S, Khosla A, et al. 3 D ShapeNets: A deep representation for volumetric shapes[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015: 1912-1920.
- [14] Yi L, Guibas L, Hertzmann A, et al. Learning hierarchical shape segmentation and labeling from online repositories [J]. *ACM Transactions on Graphics*, 2017, 36(4): 1-12.
- [15] Zhang D, He F, Tu Z, et al. Pointwise geometric and semantic learning network on 3 D point clouds [J]. *Integrated Computer-Aided Engineering*, 2020, 27(1): 57-75.
- [16] Duan Y, Zheng Y, Lu J, et al. Structural relational reasoning of point clouds[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 949-958.
- [17] Li D, Shen X, Yu Y, et al. GGM-net: Graph geometric moments convolution neural network for point cloud shape classification[C]//IEEE Access, 2020, 8: 124989-124998.
- [18] Zhai Z, Zhang X, Yao L. Multi-scale dynamic graph convolution network for point clouds classification[C]//IEEE Access, 2020, 8: 65591-65598.
- [19] Lyu Y, Huang X, Zhang Z. EllipsoidNet: Ellipsoid representation for point cloud classification and segmentation [J]. *arXiv preprint arXiv*, 2021: 2103.02517.