

用于三维点云语义分割的边增强图卷积网络

张鲁建, 毕远伟*, 刘耀文, 黄延森

烟台大学计算机与控制工程学院, 山东 烟台 264000

摘要 现在大多数基于图卷积的点云语义分割方法忽略了边构建的重要性,不能充分地表示局部区域的特征。基于此,提出一种结合注意力机制的边增强的图卷积网络(AE-GCN)。首先,将邻居点特征加入边中,而不仅仅是中心点与邻居点的特征差异;其次,加入注意力机制保证点云的局部信息得到更充分的利用;最后,采用U-Shape的分割结构,确保网络更好适应点云的语义分割这一任务。在两个公开数据集 Toronto_3D 和 S3DIS 上的实验结果表明,与目前的大多数方法相比,AE-GCN 取得了具有竞争力的结果:在 Toronto_3D 数据集的平均交并比为 80.3%,总体准确度为 97.1%;在 S3DIS 数据集的平均交并比为 68.0%,总体准确度为 87.2%。

关键词 三维图像处理; 点云语义分割; 注意力机制; 边增强; 图卷积

中图分类号 TP391

文献标志码 A

DOI: 10.3788/LOP231522

Augmented Edge Graph Convolutional Networks for Semantic Segmentation of 3D Point Clouds

Zhang Lujian, Bi Yuanwei*, Liu Yaowen, Huang Yansen

School of Computer and Control Engineering, Yantai University, Yantai 264000, Shandong, China

Abstract Currently, most point cloud semantic segmentation methods based on graph convolution overlook the critical aspect of edge construction, resulting in an incomplete representation of the features of local regions. To address this limitation, we propose a novel graph convolutional network AE-GCN that integrates edge enhancement with an attention mechanism. First, we incorporate neighboring point features into the edges rather than solely considering feature differences between the central point and its neighboring points. Second, introducing an attention mechanism ensures a more comprehensive utilization of local information within the point cloud. Finally, we employ a U-Shape segmentation structure to improve the network's semantic point cloud segmentation adaptability. Our experiments on two public datasets, Toronto_3D and S3DIS, demonstrate that AE-GCN outperforms most current methods. Specifically, on the Toronto_3D dataset, AE-GCN achieves a competitive average intersection-to-union ratio of 80.3% and an overall accuracy of 97.1%. Furthermore, on the S3DIS dataset, the model attains an average intersection-to-union ratio of 68.0% and an overall accuracy of 87.2%.

Key words three-dimensional image processing; point cloud semantic segmentation; attention mechanism; augmented edge; graph convolutional

1 引言

三维(3D)点云作为3D数据的一种,有着丰富的空间结构信息和几何信息。近些年来,随着3D传感器^[1-2]的不断发展,点云数据的获取成本也在不断降低。因此点云数据得以在自动驾驶、机器人控制、虚拟现实(VR)等领域^[3-5]得到广泛应用。但是点云数据与

1D的自然语言和2D图片等结构化数据不同,其无序性、不均匀性与非结构化性限制了常规深度学习方法在点云语义分割上的应用。

在早期的深度学习方法中,研究人员致力于将点云这一非结构化数据转换为结构化数据来处理,其中,代表性的方法有体素法^[6-7]和投影法^[8-9]。体素法是将点云中的点分配到不同的体素栅格中,再利用深度学

收稿日期: 2023-06-12; 修回日期: 2023-08-01; 录用日期: 2023-08-08; 网络首发日期: 2023-08-21

基金项目: 国家自然科学基金(62272405)、山东省自然科学基金(ZR2022MF238)、山东省青年创新科技支持计划(2021KJ080)

通信作者: byw@ytu.edu.cn

习的方法对结构化的体素进行处理以获得 3D 对象的一般特征。投影法是使用特定算法将点云从不同方向进行投影,得到多幅 2D 图像,再使用成熟的 2D 图像算法对其进行分析和处理。虽然目前这两种方法都取得了值得肯定的成果,但在点云与其他格式的数据转换过程中仍然会不可避免地丢失部分信息。

近年来,研究人员致力于直接使用原始点云作为输入的研究,以最大程度保留点云信息的丰富性。PointNet^[10]作为开创性成果,首先采用多层感知器(MLP)来直接处理点云,虽然该项工作取得了值得肯定的贡献,但却忽略了局部特征的重要性。Qi 等^[11]在 PointNet 的基础上提出 PointNet++, 该项工作通过设立分层结构来提取局部特征。Hu 等^[12]提出的大规模点云语义分割网络 RandLA-Net 采用随机下采样策略,大大提升了计算效率。虽然这些方法都取得了令人满意的成绩,但面对点云这一不规则的数据,应用深度学习的方法来揭示点云的内在结构信息仍具有挑战^[13]。

而图卷积神经网络(GCN)^[14-16]可以有效揭示不规则数据的内在结构。针对点云这一不规则数据,基于 GCN 的方法^[17-22]也取得了良好的表现。Wang 等^[17]提出动态图卷积神经网络(DGCNN),采用不同的参数提取中心点与其邻居的边缘特征来表示局部特征。Zeng 等^[20]提出一种随机图的图卷积网络(RG-GCN),在图的生成过程中随机丢弃图结构中的部分边,一定程度上达到了增强数据的目的。张坤等^[21]提出一种空间图卷积的方法(PCGCN),采用局部区域内不同尺度的特征融合策略。Li 等^[22]提出一种深层图卷积网络(DeepGCNs),解决深层神经网络中 GCN 节点特征难以区分的问题。

以上图卷积网络方法在构建局部图结构中的边时,更多考虑的是中心点以及中心点与邻近点的差异,弱化了局部区域内邻近点的特征表示,这可能会使局

部特征提取不充分,从而导致模型的性能下降。同时这些方法都采用一般深层神经网络来完成点云的语义分割任务,但是对于语义分割任务来说,一般的深层神经网络可能不是最合适的。因为语义分割任务需要考虑更多综合的信息,即不同层级的特征,而一般的深层神经网络很难做到这一点。

为解决上述问题,本文提出一种采用 U-Shape 分割结构的边增强图卷积神经网络,称之为 AE-GCN。首先,在 AE-GCN 图结构的构建中,重新考虑局部图结构中边的构建方式,提出一种新的边卷积算子 AE-Conv: 先利用 K 最近邻算法(KNN)构建局部区域,再将局部区域内的邻居点特征加入局部图结构边的表示中,从而使得局部图结构对局部特征的提取更加充分,进而提升模型的性能。然后,将注意力机制加入网络,使得模型可以自主确定需要的特征,加强特征表示。除此之外,受 U-Net^[23]启发,设计 U-Shape 分割结构,实现点云不同层级之间的特征融合: U-Shape 分割结构中的上层结构保留了点云较多精细的局部特征, U-Shape 分割结构的下层结构包含较多点云的长距离依赖特征。最后融合不同层级的特征得到多尺度的特征,得到最终分割结果。

2 所提方法

受 U-Net^[23]与 DGCNN^[17]的启发,结合 U-Shape 分割结构与图卷积神经网络,构建边增强的局部图结构并应用 AE-Conv 来有效提取局部特征。下面介绍网络的主要结构,即 U-Shape 分割结构以及边增强的图卷积算子 AE-Conv。

2.1 U-Shape 分割结构

U-Shape 分割结构如图 1 所示,该结构的输入是原始点云。 $(N \times d_{in})$ 为原始点云的尺寸, N 为输入点云的大小,即点的数量, d_{in} 为输入点云信息的维度,本实验中, $d_{in} = 6$,包括 (x, y, z, r, g, b) ,表示点云的三维空

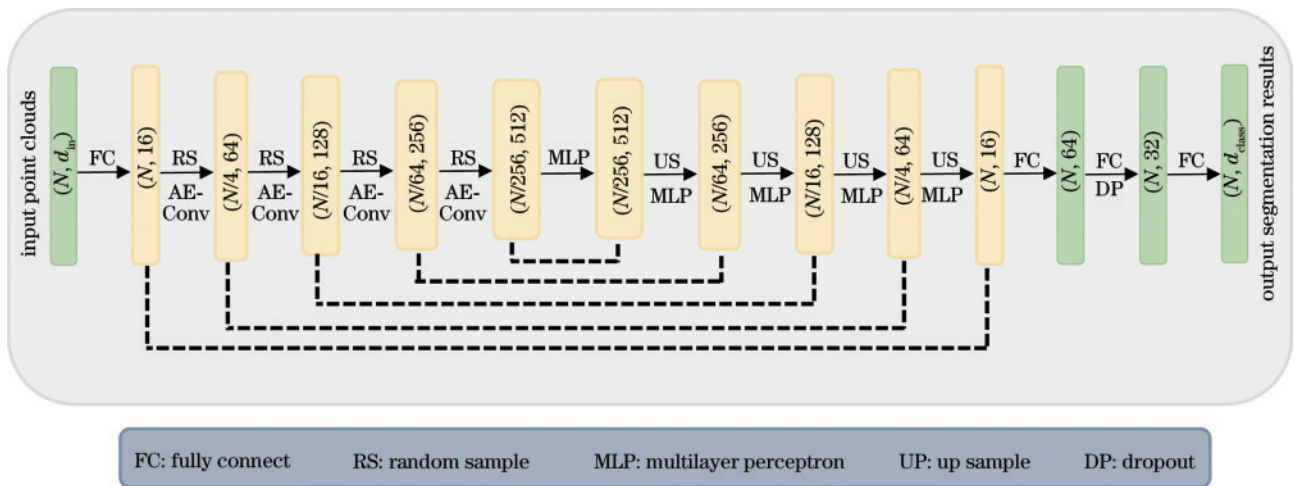


图 1 AE-GCN 的 U-Shape 分割结构

Fig. 1 The U-Shape segmentation architecture of AE-GCN

间坐标和颜色信息。

原始点云经过全连接层进行逐点特征提取,再送入经过对称的编码-解码层,从解码层最后一层输出之后,送入 1 个丢弃率为 0.5 的随机丢弃层(dropout)和 3 个全连接层得到输出结果。输出结果($N \times d_{\text{class}}$)中, N 为点的数量, d_{class} 为数据集标签的类别数量。

网络设计了多个编码层和解码层。在编码层的设计中,通过随机采样策略(RS)^[12]进行降采样操作,AE-Conv 则用于聚合局部特征。另外设计了与编码层对应的解码层,通过最近邻插值和 MLP 来恢复点云的数量和尺度。在 U-Shape 分割结构中,低层结构包含更多的点,有着较多的长依赖特征。在高层结构中,虽然由于多层级的下采样只保留较少的点,但是 MLP 操作和 AE-Conv 使得相应的特征维度得以增加,能够保留点云更加精细的局部特征。编码层与解码层对应的层级之间通过跳跃连接(skip connection)融合,并作为解码层中下一层的输入。通过融合多层之间的信息得到多尺度的信息,综合考虑点云的长依赖特征和精细的局部特征,大大提升了网络模型的性能。

2.2 AE-Conv

一般来说,一组包含了 N 个点的点云可以用 $X = \{x_i\}_{i=1}^N$ 来表示,可以使用一个简单的有向图 $G = (V, E)$ 来表示点云的局部特征, x_i 是节点 i 的特征表示, $i \in V, V$ 为图节点。将 G 构建成为以 x_i 为中心点的 KNN 领域图,即 x_i 有 k 个邻近点。对于图中边 $e_{ij} \in E$,可由一个具有可学习参数 Θ 的非线性函数 $h_{\Theta}(x_i, x_j)$ 得到,其中, $j \in k$ 。

局部图结构中的边对分割性能有着极大的影响,

因为在目前的边缘图卷积方法中,分析的主要对象是局部图结构中的边。目前大多数方法在定义 $h_{\Theta}(x_i, x_j)$ 时主要使用以下方式:

$$h_{\Theta}(x_i, x_j) = \text{Concat}(x_i, x_i - x_j), \quad (1)$$

式中,Concat 表示通道层面的连接操作,即将中心点 x_i 以及 x_i 与邻居点 x_j 的差异 $x_i - x_j$ 在通道层面进行特征的合并。这种定义方式不仅使得局部图结构中包括中心点的特征,也包括中心点和邻居点差异,可以较完整地表示局部区域的特征。但是这种定义仍然存在一些不足,因为在边缘图卷积中,边表示的是整个局部区域。而针对于整个局部区域来说,不仅包括中心点,也包括邻居点。显而易见的是,中心点和邻居点的差异并不足以完整表示邻居点的特征,因此这种方式不能很好地表示整个局部区域的特征。另外一个原因是在构建局部区域时,不仅仅要考虑局部区域内点与点之间的联系,更要考虑局部区域本身,如果忽略邻居点原本的特征,则会影响局部区域的特征表示。因此,提出一种新的构建边的方式——增强边(AE),用于更好地表示局部区域的特征:

$$h_{\Theta}(x_i, x_j) = \text{Concat}(x_i, x_i - x_j, x_j), \quad (2)$$

式中: x_i 表示局部区域的中心点; x_j 为 x_i 的邻居点。在构建边时,加入邻居点的特征表示,使中心点、邻居点,以及中心点和邻居点的差异在通道层面上进行连接。通过这样的方式,使得局部区域关注的特征变得更加丰富,进而构建出更完善的局部区域特征表示。通过式(2)的增强边,结合注意力机制与残差连接(Res-Connection)^[24]设计 AE-Conv,如图 2 所示。

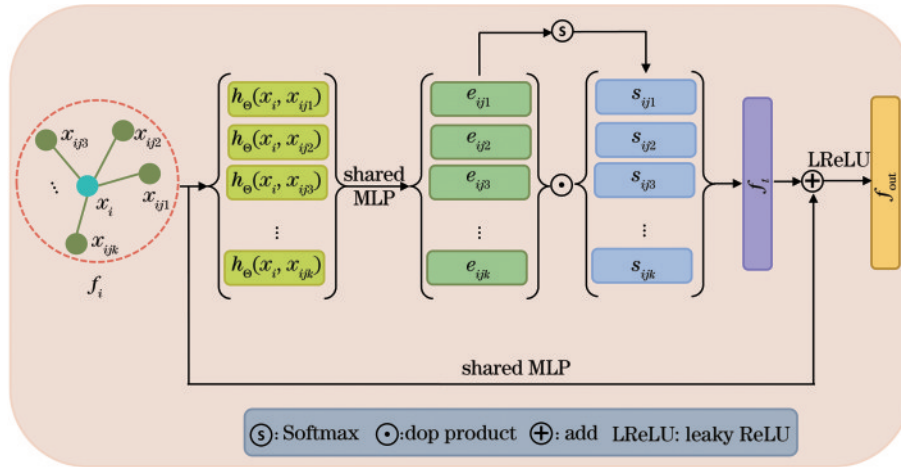


图 2 AE-Conv 结构

Fig. 2 The structure of AE-Conv

令 f_i 为以 x_i 为中心点的局部区域的原始特征,经过非线性函数 $h_{\Theta}(x_i, x_j)$ 之后,得到局部图结构中的增强边 $E = \{e_{ij}\}_{j=1}^k$ 。受 Hu 等^[25] 的启发,越重要的特征,注意力分数越高,所以首先计算出每一条边的注意力

分数:

$$S = \text{Softmax}(E). \quad (3)$$

首先通过 Softmax 处理得到注意力分数权重 $S = \{s_{ij}\}_{j=1}^k$,然后将 E 点乘 S 得到 f_i :

$$f_i = \sum_{j=1}^k e_{ij} \cdot s_{ij} \quad (4)$$

得到 f_i 之后,加入 Res-Connection^[24]使其与原始的局部特征 f_i 相结合得到最终输出的特征 f_o ,增强网络的特征传递,如式(5)所示。残差连接使得深层网络泛化得到的特征得到较好的利用,同时原始特征的加入使得局部特征提取得到增强,解决了深层网络退化的问题,从而提高语义分割的效果。

$$f_o = f_i + f_i \quad (5)$$

3 实验

3.1 实验数据与环境

为了评估所提 AE-GCN 的性能,在室内数据集 S3DIS^[26]和室外数据集 Toronto_3D^[27]上进行实验,以验证其在不同场景下的性能。

S3DIS 数据集是由斯坦福大学提出的大规模室内 3D 点云数据集,包括 6 个区域,271 个房间,11 个场景。S3DIS 数据集共包含约 2.74 亿个点,13 个类别,每个点由坐标信息、颜色信息和标签信息组成。在 S3DIS 数据集上采用六轮交叉验证的方式来评估 AE-GCN

的性能。

Toronto_3D 数据集是在加拿大多伦多收集的大型室外城市点云数据集,该数据集覆盖约 1 km 的真实路况场景,包括约 7830 万个点,共 4 个区域和 8 个类别。其中,每个点由坐标信息、颜色信息、法向量、GPS 时间、扫描角度和标签信息组成。遵循 Toronto_3D 数据集原始论文的策略,将 L001、L003 和 L004 用于训练,L002 用来测试。

在带有 Tesla V100 GPU 的服务器上使用 TensorFlow 实现 AE-GCN 的部署,使用 Adam^[28]优化器,初始学习率设置为 0.01,每个 epoch 之后减少 5%,且 epoch 数量设置为 100。在 S3DIS 数据集^[26]上的 batch size 设置为 6,在 Toronto_3D 数据集^[27]上的 batch size 设置为 4。使用总体准确率(OA)、交并比(IoU)和平均交并比(mIoU)等 3 个评价指标来评估模型的性能。

3.2 S3DIS 数据集实验结果及分析

表 1 为 AE-GCN 在 S3DIS 数据集^[26]上的定量分割结果,并与近些年来经典的深度学习的方法进行了对比,给出了 OA、mIoU 和各个类的 IoU,最优的结果已加粗显示,次优的结果已通过下划线标出。

表 1 S3DIS 数据集实验结果对比

Table 1 Comparison of experimental results of S3DIS dataset

unit: %

Method	OA	mIoU	ceiling	floor	wall	beam	column	window	door	table	chair	sofa	book	board	clut
PointNet ^[10]	78.6	47.6	88.0	88.7	69.3	42.4	23.1	47.5	51.6	54.1	42.0	9.6	38.2	29.4	35.2
PointNet++ ^[11]		53.2	90.2	91.7	73.1	42.7	21.2	49.7	42.3	62.7	59.0	19.6	45.8	48.2	45.6
DeepGCNs ^[22]	85.9	60.0	93.1	95.3	78.2	33.9	37.4	56.1	68.2	64.9	61.0	34.6	51.5	51.1	54.4
SPGraph ^[29]	86.4	62.1	89.9	95.1	76.4	62.8	47.1	55.3	68.4	73.5	69.2	<u>63.2</u>	45.9	8.7	52.9
PCGCN ^[21]	81.3	49.5													
AMFF-DGCNN ^[30]	85.6	59.9	92.9	94.7	78.3	54.0	43.3	59.1	62.6	67.4	59.1	18.3	51.3	47.4	50.5
RG-GCN ^[20]	<u>88.1</u>	63.7	94.0	96.2	79.1	<u>60.4</u>	44.3	60.1	65.9	<u>70.8</u>	64.9	30.8	51.9	52.6	56.4
DPFA ^[31]	89.2	61.7	94.6	98.0	79.2	40.7	36.6	52.2	70.8	65.9	74.7	27.7	49.8	51.6	60.6
SC-CNN ^[32]		<u>67.4</u>	<u>93.6</u>	94.1	83.1	50.5	33.0	<u>60.3</u>	<u>69.6</u>	63.0	<u>72.3</u>	63.8	64.3	<u>60.6</u>	57.1
DGCNN ^[17]	84.5	55.5	93.2	95.9	72.8	54.6	32.2	56.2	50.7	62.8	63.4	22.7	38.2	32.5	46.8
AE-GCN (ours)	87.2	68.0	93.3	<u>96.8</u>	<u>79.8</u>	56.6	<u>45.9</u>	63.9	67.8	68.3	70.2	59.6	<u>61.9</u>	62.6	<u>57.7</u>

可以看出,尽管 AE-GCN 在 OA 上没有达到最优,但是其综合表现却更为出色,达到 87.2% 的 OA 和 68.0% 的 mIoU,相较于基线 DGCNN^[17]来说,分别提升 2.7 个百分点和 12.5 个百分点。DPFA^[31]通过动态池和注意机制选择性地执行邻域特征聚合,达到 89.2% 的 OA,但其强调聚合不同层级的特征导致局部特征分散,这可能是 DPFA 在 mIoU 指标上表现并不突出的主要原因。而 AE-GCN 所采用的增强边的图结构使得其在局部区域内取得较好的表现,U-Shpae 分割结构在不同层级之间的作用使其可以融合丰富的上下文特征,进而使得 AE-GCN 在 OA 上也取得良好的成绩。可以看出,AE-GCN 的总体表现是值得肯定的,虽然

在 OA 上没有达到最优,但在 mIoU 上表现最好,同时在一半左右的类别上都取得较好的表现,比如窗户、黑板等。

图 3 列举了 AE-GCN 在 S3DIS^[26]数据集上的可视化结果,从左到右依次是 S3DIS 数据集真实场景可视化、S3DIS 数据集原始数据和标签可视化、AE-GCN 的结果可视化、DGCNN^[17]的结果可视化。从图 3 第 1 行和第 3 行可以看出,相较于 DGCNN,AE-GCN 在边缘的表现上更加平滑,这主要是因为增加了邻居点的信息表示。另外,从图 3 第 2 行可以看出,AE-GCN 在复杂的局部区域也取得了与真实可视化更接近的分割结果。

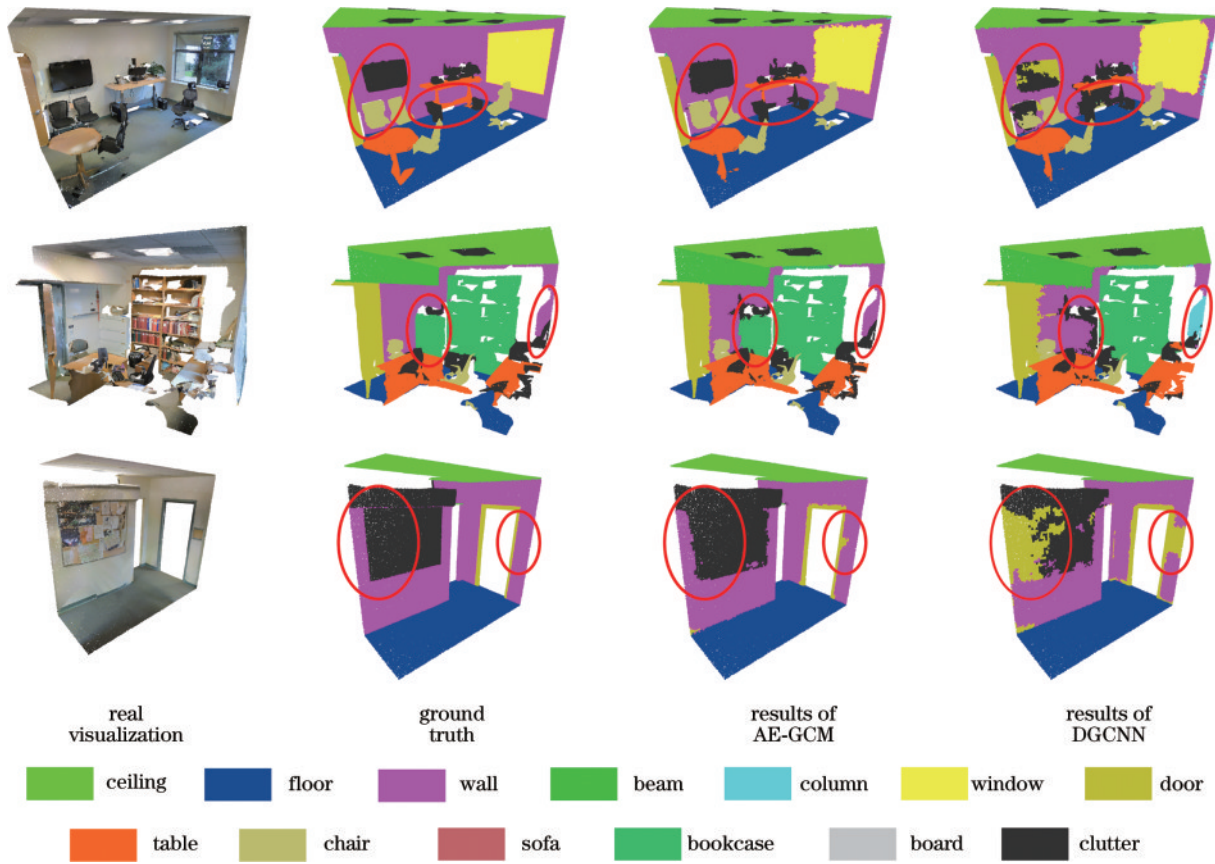


图 3 S3DIS 数据集可视化结果
Fig. 3 The visualization results of S3DIS dataset

3.3 Toronto_3D 数据集实验结果及分析

表 2 给出了 AE-GCN 在 Toronto_3D 数据集^[27]上的定量语义分割结果,并与近年来经典的深度学习的方法进行了对比,给出了 OA、mIoU 和各个类别的 IoU,最优的结果已加粗显示,次优的结果已通过下划线标出。

可以看出,AE-GCN 在 Toronto_3D 数据集上取得 97.1% 的 OA 和 80.3% 的 mIoU,表现最优,在各个类别上的 IoU 虽然没有取得最好的表现,但综合来看

也取得了有竞争力的结果。相较于基线 DGCNN^[17],AE-GCN 在 OA 和 mIoU 上分别提升 2.9 个百分点和 18.5 个百分点,足以体现所提方法的优越性。KPFCNN^[33]提出的可变卷积核使其在诸如路灯、公用事业管线等形状大小各异的物体类别上取得优异的成绩。MFA^[36]通过自适应聚合局部区域内不同层级之间的特征使得其取得优异的表现,但不得不考虑的是,高维空间内的特征聚合对于计算机资源要求较高,在实际中应用难度大。

表 2 Toronto_3D 数据集实验结果对比
Table 2 Comparison of experimental results of Toronto_3D dataset

Method	OA	mIoU	road	road marking	natural	building	utility line	pole	car	fence
PointNet++ ^[11]	84.9	41.8	89.3	0.0	69.0	54.1	43.7	23.3	52.0	3.0
KPFCNN ^[33]	95.4	69.1	94.6	0.1	96.1	91.5	87.7	81.6	85.7	15.7
TGNet ^[34]	94.1	61.3	93.5	0.0	90.8	81.6	65.3	63.0	88.7	7.9
MS-TGNet ^[27]	95.7	70.5	94.4	17.2	95.7	88.8	76.0	74.0	94.2	23.6
Method of reference ^[35]	83.6	71.0	92.8	27.4	89.9	95.2	85.6	74.5	44.4	58.3
RG-GCN ^[20]	96.5	74.5	98.2	79.4	91.8	86.1	72.4	69.9	82.1	16.0
MFA ^[36]	<u>97.0</u>	<u>79.9</u>	<u>96.8</u>	<u>70.0</u>	96.1	92.3	86.3	<u>80.4</u>	91.5	29.4
Method of reference ^[37]	95.3	73.9	95.7	25.9	94.0	86.3	81.5	71.8	78.1	<u>58.1</u>
RandLA-Net ^[12]	93.0	77.7	94.6	42.6	96.9	93.0	<u>86.5</u>	78.1	<u>92.9</u>	37.1
DGCNN ^[17]	94.2	61.8	93.9	0.0	91.3	80.4	62.4	62.3	88.3	15.8
AE-GCN	97.1	80.3	96.7	65.2	<u>96.6</u>	<u>93.5</u>	85.3	77.3	92.6	35.3

图 4 列举了 AE-GCN 在 Toronto_3D 数据集^[27]的可视化结果,从左到右的顺序与 S3DIS 数据集^[26]可视化类似,从上到下分别展示了 L002 区域的整体结果、部分道路标记的可视化,以及驾驶视角的可视化。可以看出,相较于 DGCNN^[17],AE-GCN 在复杂的局部区域内表现更好,可以更好地区分道路标志和道路之间的边界,这主要得益于 U-Shape 分割网络的加入,得以综合上下文不同层级的信息,最大限度保留信息的完整性。

3.4 消融实验

为了更好地分析网络模型的性能,还在 S3DIS 数据集^[26]上设计了一系列的消融实验来验证 AE-GCN 中各个模块的性能:主要模块的消融实验,包括边的构建方式、注意力机制和 U-Shape 分割结构;AE-Conv 中池化方式的消融实验;局部区域内邻居点数量的消融实验。

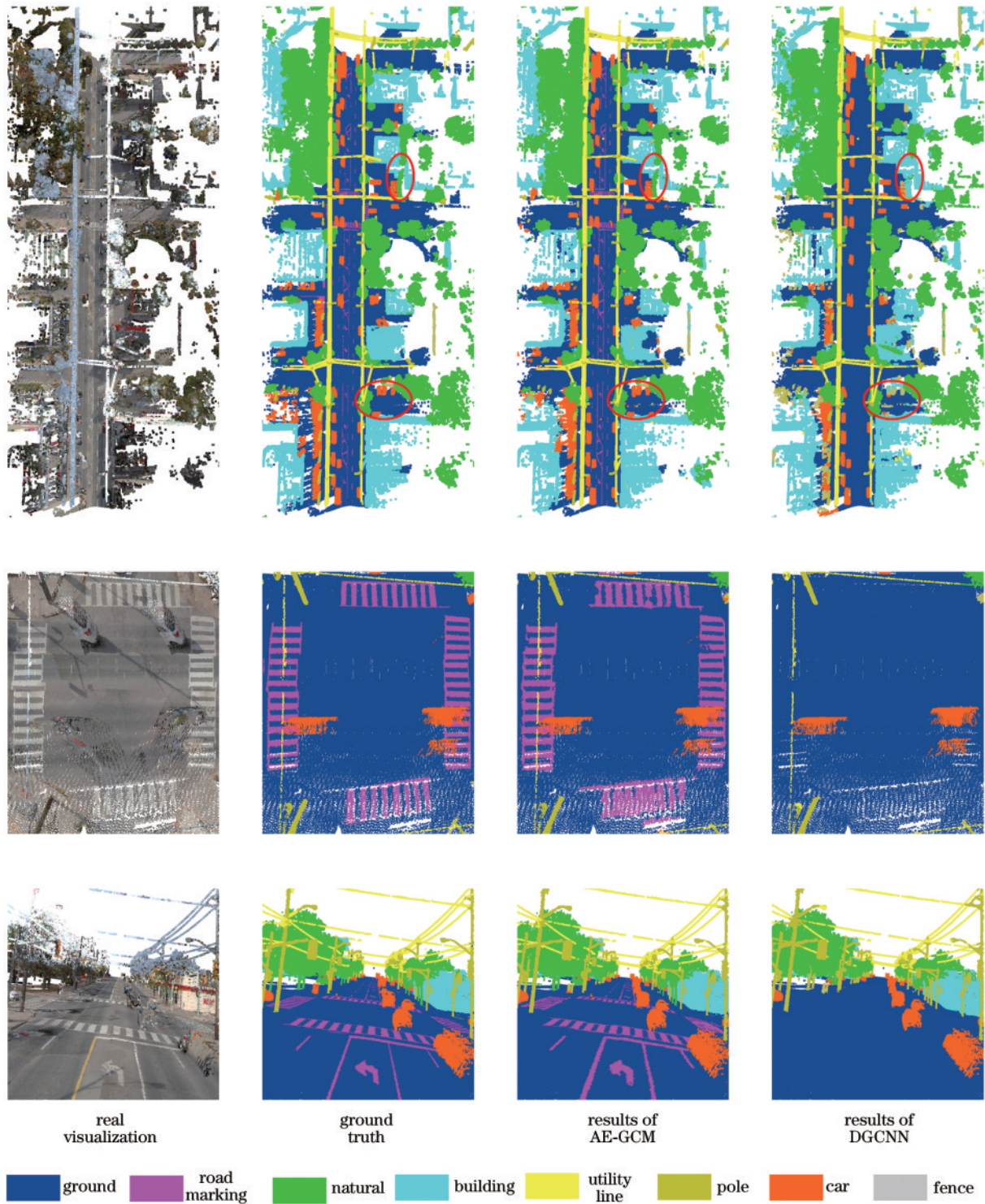


图 4 Toronto_3D 数据集可视化结果
Fig. 4 The visualization results of Toronto_3D dataset

3.4.1 主要模块的消融实验

表 3 给出了主要模块的消融实验的结果。为了方便展示,这里省略了边 $h_{\theta}(x_i, x_j)$ 构造的表达式,在表 3 中的边选项中,各个模型的元素都在通道层面进行连接。由模型 A 与 B、C 与 E、D 与 F 之间的对比可以看出,无论是否加入 U-Shape 结构或者注意力机制,增强边都可比原始边取得更好的效果,充分说明了所提增强边的有效性。由模型 B 与 C 之间的对比可以看出,U-Shape 结构对于分割效果的提升是十分令人满意的,这主要源于 U-Shape 分割结构不同层级之间特征的融合,细粒度的特征和长距离依赖特征的结合使得模型可以捕捉到更全面的特征。由模型 C 与 D、E 与 F 之间的对比可以看出,注意力池化对于模型性能的提升也是令人惊讶的,因为注意力机制是一项柔性选择特征的方法,不同于基线方法使用的最大池化,注意力池化在选择明显特征的时候不会抛弃其他特征,保留了更多的特征丰富度。

表 3 主要模块的消融实验
Table 3 The ablation experiment of main modules

Model	Edge	U-Shpae structure	Attention	mIoU / %
A	$x_i, x_j - x_i$			49.1
B	$x_i, x_j - x_i, x_j$			50.6
C	$x_i, x_j - x_i$	✓		56.3
D	$x_i, x_j - x_i$	✓	✓	61.8
E	$x_i, x_j - x_i, x_j$	✓		58.5
F	$x_i, x_j - x_i, x_j$	✓	✓	62.6
G	$x_j - x_i, x_j$	✓	✓	55.1

此外,还加入了模型 G 进行对比实验,在该组实验中,只考虑中心点和邻居点的差异以及邻居点的特征。可以看出,模型 G 的性能表现并不优异,这主要是因为这种构造方式只有邻居点信息,将局部区域视为小块的集合,失去了局部结构。而从模型 D 与 F、F 与 G 对比结果也可以看出,中心点以及中心点与邻居点的差异并不能代表邻居点的特征,邻居点以及中心点与邻居点的差异也不能完整表示中心点的特征。综上所述,在局部图结构中的边的构造中,加入完整的局部特征表示是很有必要的,不仅保证了局部区域内点与点之间的联系,更保证了局部特征的完整表达。

3.4.2 AE-Conv 池化方式的消融实验

表 4 给出了关于 AE-Conv 中池化方式的消融实验的结果,可以看出,注意力池化表现最好,这得益于注意力机制对特征的柔性选择。平均池化的表现最差,这主要是因为平均池化考虑局部区域内所有点的特征,减少了特征之间的区分度。和池化与平均池化类似,虽然也是聚合所有点的特征,但是和池化与平均池化的不同之处在于和池化保留了特征之间的原始差异,因而和池化表现更好。最大池化只保留局部区域

表 4 AE-Conv 池化方式的消融实验

Model	Pooling method	mIoU / %
A	Mean	57.0
B	Max	58.2
C	Sum	58.5
D	Attention	62.6

内最明显的特征,忽略了其他特征的影响,不能完整表示局部特征,不能达到较为理想的效果。注意力池化兼顾了和池化与最大池化的优点,考虑了局部区域内所有点的特征,在保留突出特征的同时也考虑了其他特征的影响,如此双重保障使得注意力池化的表现更为突出。

3.4.3 局部区域内邻居点数量的消融实验

表 5 给出了局部区域内邻居点数量的消融实验结果。邻居点数量对于局部特征表示有着极为重要的影响,数量太少可能会导致局部区域的特征表达不足,如模型 A 与 B、B 与 C 所示,随着邻居点数量的增加,mIoU 指标持续提升。但是这并不意味着数量越多越好,较多的点虽然可以丰富局部区域的特征表达,但是数量增加,引入的无关特征也会增加,因此过多的点也会导致模型性能下降,如模型 C 与 D 所示,当邻居点数量为 16 时效果最好,继续增加到 20,引入了较多无关点的特征,导致性能下降。

表 5 局部区域内邻居点数量的消融实验
Table 5 The ablation experiment of the number of neighbor points in the local area

Model	Number	mIoU / %
A	8	59.4
B	12	61.3
C	16	62.6
D	20	61.4

4 结 论

提出一种用于点云语义分割的图卷积方法,设计了一种新型图卷积中边的构造方式,加强了局部区域内的特征表示,同时注意力机制的加入增强了网络获取局部特征的能力。U-Shape 分割结构的加入也使得网络更多地关注到不同层级之间的特征,细粒度的特征和长距离的上下文特征相互融合,使模型的性能表现更加优异。该模型在室内场景和室外场景数据集上都达到了较好的表现,相较于基线模型来说,有了较大的提升。即使与最近的方法相比,也基本上能达到不相上下的水平。同时设计的一系列消融实验也阐述了所提各个模块的有效性以及网络中一些细节选择的合理性。

但是目前点云分析中图卷积的方法严重依赖于图

结构,而图结构只考虑边的局限性可能会导致局部区域特征利用不足,仍然存在改进的空间。在未来的工作中,我们将会继续探究图结构的表示,包括边的构造。一种可行的方式是弱化边对点特征的包含度,重新思考边的构造方式,突出图结构中边与点的差异,且保证边与点共同表示局部区域,进一步提升网络的性能。

参 考 文 献

- [1] Zhang L Q, Zhou Z Q, Hu X S, et al. The recent progress of laser-induced graphene based device applications [J]. *Journal of Semiconductors*, 2023, 44(3): 031701.
- [2] Zeng X, Cui S Z, Jiang H W, et al. Single-frequency upconverted laser generation by phase summation[J]. *High Power Laser Science and Engineering*, 2023, 11(2): e18.
- [3] Wang L, Song Z Y, Zhang X Y, et al. SAT-GCN: self-attention graph convolutional network-based 3D object detection for autonomous driving[J]. *Knowledge-Based Systems*, 2023, 259: 110080.
- [4] Wakchaure M, Patle B K, Mahindrakar A K. Application of AI techniques and robotics in agriculture: a review[J]. *Artificial Intelligence in the Life Sciences*, 2023, 3: 100057.
- [5] Parmaxi A. Virtual reality in language learning: a systematic review and implications for research and practice[J]. *Interactive Learning Environments*, 2023, 31(1): 172-184.
- [6] Maturana D, Scherer S. VoxNet: a 3D Convolutional Neural Network for real-time object recognition[C]//2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), September 28-October 2, 2015, Hamburg, Germany. New York: IEEE Press, 2015: 922-928.
- [7] Riegler G, Ulusoy A O, Geiger A. OctNet: learning deep 3D representations at high resolutions[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 6620-6629.
- [8] Wei X, Yu R X, Sun J. View-GCN: view-based graph convolutional network for 3D shape analysis[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 1847-1856.
- [9] Boulch A, Guerry J, Le Saux B, et al. SnapNet: 3D point cloud semantic labeling with 2D deep segmentation networks[J]. *Computers & Graphics*, 2018, 71: 189-198.
- [10] Charles R Q, Hao S, Mo K C, et al. PointNet: deep learning on point sets for 3D classification and segmentation [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 77-85.
- [11] Qi C R, Yi L, Su H, et al. PointNet++: deep hierarchical feature learning on point sets in a metric space [EB/OL]. (2017-06-07)[2023-03-05]. <https://arxiv.org/abs/1706.02413>.
- [12] Hu Q Y, Yang B, Xie L H, et al. Learning semantic segmentation of large-scale point clouds with random sampling[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022, 44(11): 8338-8354.
- [13] Du Z J, Ye H L, Cao F L. A novel local-global graph convolutional method for point cloud semantic segmentation[J/OL]. *IEEE Transactions on Neural Networks and Learning Systems*, 2022: 1-15[2023-03-05]. <https://ieeexplore.ieee.org/document/9733414>.
- [14] Scarselli F, Gori M, Tsoi A C, et al. The graph neural network model[J]. *IEEE Transactions on Neural Networks*, 2009, 20(1): 61-80.
- [15] Zhang X K, Xu C, Tian X M, et al. Graph edge convolutional neural networks for skeleton-based action recognition[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2020, 31(8): 3047-3060.
- [16] Bhatti U A, Tang H, Wu G L, et al. Deep learning with graph convolutional networks: an overview and latest applications in computational intelligence[J]. *International Journal of Intelligent Systems*, 2023, 2023: 1-28.
- [17] Wang Y, Sun Y B, Liu Z W, et al. Dynamic graph CNN for learning on point clouds[J]. *ACM Transactions on Graphics*, 38(5): 146.
- [18] 刘友群, 敖建锋, 潘仲泰. DGPoint:用于三维点云语义分割的动态图卷积网络[J]. *激光与光电子学进展*, 2022, 59(16): 1610014.
Liu Y Q, Ao J F, Pan Z T. DGPoint: a dynamic graph convolution network for point cloud semantic segmentation [J]. *Laser & Optoelectronics Progress*, 2022, 59(16): 1610014.
- [19] Wang L, Huang Y C, Hou Y L, et al. Graph attention convolution for point cloud semantic segmentation[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2020: 10288-10297.
- [20] Zeng Z Y, Xu Y Y, Xie Z, et al. RG-GCN: a random graph based on graph convolution network for point cloud semantic segmentation[J]. *Remote Sensing*, 2022, 14(16): 4055.
- [21] 张坤, 朱亚薇, 王晓红, 等. 基于空间图卷积的三维点云语义分割[J]. *激光与光电子学进展*, 2023, 60(2): 0228007.
Zhang K, Zhu Y W, Wang X H, et al. Three-dimensional point cloud semantic segmentation network based on spatial graph convolution network[J]. *Laser & Optoelectronics Progress*, 2023, 60(2): 0228007.
- [22] Li G H, Müller M, Thabet A, et al. DeepGCNs: can GCNs go as deep as CNNs? [C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV), October 27-November 2, 2019, Seoul, Korea (South). New York: IEEE Press, 2020: 9266-9275.
- [23] Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation[M]//Navab N, Hornegger J, Wells W M, et al. Medical image computing and computer-assisted intervention-MICCAI

2015. Lecture notes in computer science. Cham: Springer, 2015, 9351: 234-241.
- [24] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 770-778.
- [25] Hu J, Shen L, Sun G. Squeeze-and-excitation networks [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 7132-7141.
- [26] Armeni I, Sax S, Zamir A R, et al. Joint 2D-3D-semantic data for indoor scene understanding[EB/OL]. (2017-02-03)[2023-03-05]. <https://arxiv.org/abs/1702.01105>.
- [27] Tan W K, Qin N N, Ma L F, et al. Toronto-3D: a large-scale mobile LiDAR dataset for semantic segmentation of urban roadways[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), June 14-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 797-806.
- [28] Kingma D P, Ba J. Adam: a method for stochastic optimization[EB/OL]. (2014-12-22) [2023-03-05]. <https://arxiv.org/abs/1412.6980>.
- [29] Landrieu L, Simonovsky M. Large-scale point cloud semantic segmentation with superpoint graphs[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 4558-4567.
- [30] 郝雯, 王红霄, 汪洋. 结合空间注意力与形状特征的三维点云语义分割[J]. 激光与光电子学进展, 2022, 59(8): 0828004.
- Hao W, Wang H X, Wang Y. Semantic segmentation of three-dimensional point cloud based on spatial attention and shape feature[J]. Laser & Optoelectronics Progress, 2022, 59(8): 0828004.
- [31] Chen J J, Kakillioglu B, Velipasalar S. Background-aware 3-D point cloud segmentation with dynamic point feature aggregation[J]. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: 5703112.
- [32] Wang C S, Ning X, Sun L J, et al. Learning discriminative features by covering local geometric space for point cloud analysis[J]. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60: 5703215.
- [33] Thomas H, Qi C R, Deschaud J E, et al. KPConv: flexible and deformable convolution for point clouds[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV), October 27-November 2, 2019, Seoul, Korea (South). New York: IEEE Press, 2020: 6410-6419.
- [34] Li Y, Ma L F, Zhong Z L, et al. TGNet: geometric graph CNN on 3-D point cloud segmentation[J]. IEEE Transactions on Geoscience and Remote Sensing, 2020, 58(5): 3588-3600.
- [35] Rim B, Lee A, Hong M. Semantic segmentation of large-scale outdoor point clouds by encoder - decoder shared MLPs with multiple losses[J]. Remote Sensing, 2021, 13(16): 3121.
- [36] Chen J Q, Zhao Y D, Meng C T, et al. Multi-feature aggregation for semantic segmentation of an urban scene point cloud[J]. Remote Sensing, 2022, 14(20): 5134.
- [37] Wang Y C, Wang W C, Liu J Z, et al. Framework for geometric information extraction and digital modeling from LiDAR data of road scenarios[J]. Remote Sensing, 2023, 15(3): 576.