

DOI:10.19651/j.cnki.emt.2211484

基于几何仿射与注意力的三维激光点云分类*

胡平安¹ 黎浩民¹ 李光平¹ 林映生²

(1. 广东工业大学信息工程学院 广州 510006; 2. 深圳市金百泽电子科技股份有限公司 深圳 518000)

摘要: 三维点云分类和分割对于三维重建和自动驾驶等技术的发展具有积极的推动作用。三维点云数据具有无序、不规则和稀疏等特点,因此三维点云分类和分割的研究面临诸多挑战。PCT分类网络采用标量注意力机制提取三维点云局部特征,具有良好的三维点云特征学习能力,在三维点云分类和分割任务中表现出先进的分类精度。然而PCT在对三维点云数据进行下采样时忽视了其稀疏性对几何结构所产生的影响,从而无法充分地提取局部特征致使三维点云分类和分割精度下降。针对该问题,本文提出一种基于注意力机制的三维点云分类分割网络GAM-PCT,具体地,GAM-PCT网络采用了向量注意力机制对单通道特征的权重进行调节,利用减法关系和邻域位置编码对三维点云邻域求取注意力特征,同时对整体点云下采样时插入即插即用的几何形状仿射(GAM)模块来解决三维点云局部区域的稀疏性问题,进而提升网络的分类准确率。实验结果表明,与PCT三维点云分类和分割网络相比,所提出GAM-PCT网络在数据集ModelNet40上的分类精度提升了0.3%,而在ScanObjectNN数据集上的分类精度提升了1.9%,在ShapeNet数据集上的分割平均交并比值提升了0.2%。同时在网络参数数量和FLOPs指标上分别降低了0.31 G和0.69 M。实验结果表明改进后网络的复杂度得到了简化,充分验证了改进方法的有效性。

关键词: 三维点云;机器视觉;注意力机制;深度学习;三维重建

中图分类号: TP391 **文献标识码:** A **国家标准学科分类代码:** 510.403

Point cloud classification based on geometric affine and attention mechanism

Hu Ping'an¹ Li Haomin¹ Li Guangping¹ Lin Yingsheng²

(1. Faculty of Information Engineering, Guangdong University of Technology, Guangzhou 510006, China;

2. Shenzhen jinbaize Electronic Technology Co., Ltd., Shenzhen 518000, China)

Abstract: The classification and segmentation of 3D laser point cloud have a positive role in promoting the development of 3D reconstruction and automatic driving technology. The 3D laser point cloud data has the characteristics of disorder, irregularity, and sparsity, so the research of 3D laser point cloud classification and segmentation faces many challenges. The point cloud transformer (PCT) classification network uses the scalar attention mechanism to extract the local features of 3D laser point clouds. It has a good 3D laser point cloud feature learning ability and shows advanced classification accuracy in 3D laser point cloud classification and segmentation tasks. However, when PCT downsamples the 3D laser point cloud data, it ignores the influence of its sparsity on the geometric structure, so it cannot fully extract the local features, resulting in the degradation of the classification and segmentation accuracy of the 3D laser point cloud. To solve this problem, this paper proposes a three-dimensional laser point cloud classification and segmentation network GAM-PCT based on the attention mechanism. Specifically, the GAM-PCT network uses the vector attention mechanism to adjust the weight of the single channel features and uses the subtraction relationship and neighborhood location coding to obtain the attention features of the three-dimensional laser point cloud neighborhood. At the same time, a plug and play geometric affine (GAM) module is inserted to solve the sparsity problem of the local area of the three-dimensional laser point cloud when downsampling the whole point cloud, thereby improving the classification accuracy of the network. The experimental results show that, compared with the PCT three-dimensional laser point cloud classification and segmentation network, the classification accuracy of the proposed GAM-PCT network on the data set modelnet40 is increased by 0.3%, while the classification accuracy on the ScanObjectNN data set is increased by 1.9%, and the average intersection ratio of segmentation on the shipment data set is increased by 0.2%. At the same time, the network parameters and the flops index are reduced by 0.31 g and 0.69 m respectively. The experimental results show that the complexity of the improved network is simplified, which fully verifies the effectiveness of the improved method.

Keywords: point cloud; machine vision; attention mechanism; deep learning; three-dimensional reconstruction

0 引言

深度学习技术的迅速发展为三维点云处理提供了丰富

多样的研究方法,使得三维点云分类和分割研究得到了长足的进步。近年来,卷积神经网络(convolutional neural network, CNN)在图像处理领域取得了突破性进展,三维

收稿日期:2022-09-21

* 基金项目:国家自然科学基金(61601130)、大亚湾科技计划项目(2020010203)资助

点云数据不同于图像数据,图像中每个像素点具有固定位置,三维点云内点的不同位置排序不会改变三维点云的几何形状结构,于是 Su 等^[1]提出了 MVCNN 网络,将三维点云转化为不同视角下的投影图像,用 CNN 学习图像特征,然后通过视觉几何图组^[2](visual geometry group, VGG)网络融合三维点云图像特征实现三维点云分类和分割,但在生成投影图过程中会丢失三维点云的有效特征。Maturana 等^[3]提出了 Voxnet 网络,该网络将无规则的三维点云转换为体素形式,再利用三维卷积对体素进行分类,但是该类方法会增加计算开销。丢失三维点云原始的几何形状特征。Qi^[4]将三维点云数据输入到共享感知机(share MLP)中,获取三维点云每个点的特征。但是该方法忽略了三维点云的局部特征。Pointnet++^[5]是在 Pointnet 的基础上做出改进,该方法将三维点云划分为不同的局部区域,有效提取局部区域特征,但该网络体系复杂,训练参数量大。DGCNN^[6]是基于图卷积神经网络(graph convolutional networks, GCN)设计的,该网络在三维点云局部区域内采用边缘卷积模块,融合了局部邻居信息,通过堆叠多层模块用于学习全局特征。但忽略了点之间的向量方向导致点云局部信息丢失。

注意力机制具有参数少,处理速度快等优点,使网络关注重要特征,忽略次要信息,在自然语言处理领域得到了广泛应用。文献[7]在处理中文评论的比较句识别任务时,提出了端到端的层次多注意力网络模型进行文本的分类。Luong 等^[8]将注意力机制应用于改进神经机器翻译中,验证了关注全局和局部源词两种方法在 WMT 翻译任务中的有效性。采用注意力机制来学习输入点间存在特殊关联性,可以充分捕获点云局部特征。Guo 等^[9]提出了一种参数少、精度高的 PCT 网络,该网络对点云进行局部下采样,将输入点云映射到高维特征空间,采用自注意力模块捕获点云中的局部特征,结合全局特征实现点云分类,但是该方法存在缺陷是没有考虑点云的稀疏性对分类精度造成的影响。于是本文在 PCT 点云分类网络的基础上做出改进,添加几何仿射模块解决点云稀疏性问题,在点云的局部邻域内应用向量注意,运用减法关系和增加位置编码,不关注孤立点的特征,以学习点云的点间的联系为目标,提高点云分类精度。

为了解决三维点云无序的问题,Liu 等^[10]提出了几何关系编码的卷积算子 RSCov,设计基于几何关系来学习 3D 形状关系的卷积神经网络 RSCNN,该网络可以显式地对点的几何关系进行编码,从而在很大程度上提高了对形状的感知能力和鲁棒性。Wang 等^[11]引入图注意力卷积构建 GACNet,通过建立每个点与周围点的图结构,引入注意力机制给中心点的不同邻近点分配适当的注意权重,GACNet 网络结构简单,能够捕获三维点云的结构化特征,进行细粒度分割,避免了对象之间的特征污染。Yang 等^[12]提出基于三维点云推理的点注意力变压器(PAT),采

用组混洗注意力机制(group shuffle attention, GSA)代替并行注意力机制建模点之间的关系,GSA 模块可以更好地挖掘点之间的特征关系,但将 GSS 应用于真实环境三维点云的泛化性不够理想。

以往的工作将全局注意力^[13]应用于整个三维点云,这会引入大量计算,并使这些方法不适用于大规模 3D 场景理解。PCT 网络考虑到自注意力机制可以有效地捕获三维点云全局范围内的特征信息,但是却忽略了对局部特征的关注。基于此,本文提出 GAM-PCT 网络用于点云分类和分割,其一,PCT 网络在局部下采样中使得三维点云密度变得稀疏,使得网络分类精度降低,GAM-PCT 网络采用几何仿射模块对局部采样点进行特征处理,改善网络的鲁棒性。其二,GAM-PCT 网络采用向量注意力对单个通道的特征向量的权重进行调节,向量注意力在 3D 感知任务上处理速度更快,增强了点云分类和分割特征的表达力,提升了分类准确率。

1 GAM-PCT 分类和分割网络

1.1 PCT 三维点云分类和分割网络

PCT 三维点云分类和分割网络^[14]首先采用卷积操作将点云的三维坐标映射到高维的特征空间,为了学习三维点云的高维嵌入特征用以区分不同类别点云。接着采用 SG 模块对整体三维点云进行不同规模的局部点采样和区域划分,将不同尺度下的局部区域的特征进行提取和聚合,SG 模块结构如图 1 所示。SG 模块由两个相同的 sg 层组成,sg1 和 sg2 层内部参数是相互独立的。高维的三维点云特征由 sg1 层的最远点采样算法(Farthest Point Sampling, FPS)进行下采样获取 512 个中心点的特征,sg1 层采用卷积核为 128×1 的一维卷积和池化层对特征进行压缩,增强特征的表达力。sg2 层对 sg1 层采样得到的局部三维点云进行 FPS 下采样获取 256 个中心点的特征,接着连接一个卷积核为 256×1 的一维卷积和池化层处理三维点云特征。

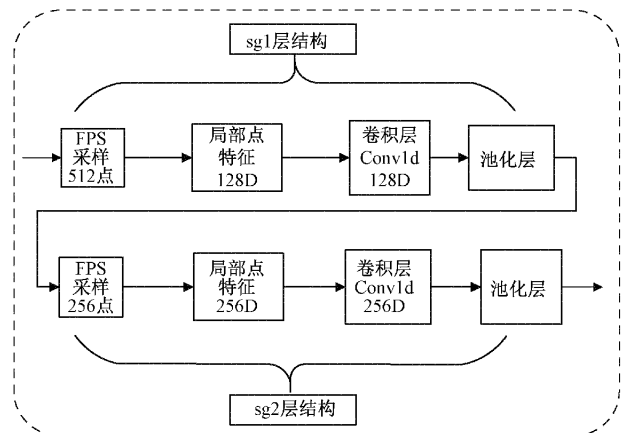


图 1 SG 模块结构

由于三维点云具有良好的三维形状表征能力,三维点云的局部点之间具有特殊的关联性。PCT 用偏移注意力模块(offset-attention,OA)学习三维点云局部特征以捕获三维点云形状表示,PCT 三维点云分类和分割网络结构如图 2 所示。每个 OA 模块输出 256 维注意力特征,由 4 个 OA 模块级联输出的特征经过拼接操作后输出 1 024 维的高维特征。再和 SG 模块输出的 256 维特征进行拼接得到 1 280 维全局

特征,全局特征由最大池化层对特征向量进行降维输出 1 024 维特征,保留最关键的特征信息。全局特征输入到共享的 MLP(全连接网络)层,MLP 中的 3 层全连接层的输出特征维度依次是 512、256、40。全连接网络的每一层使用 ReLU 函数(rectified linear unit)作为激活函数,并进行批标准化预处理(batch normalization),最后一层全连接网络输出 K 个类别得分,用于预测最终的分类结果。

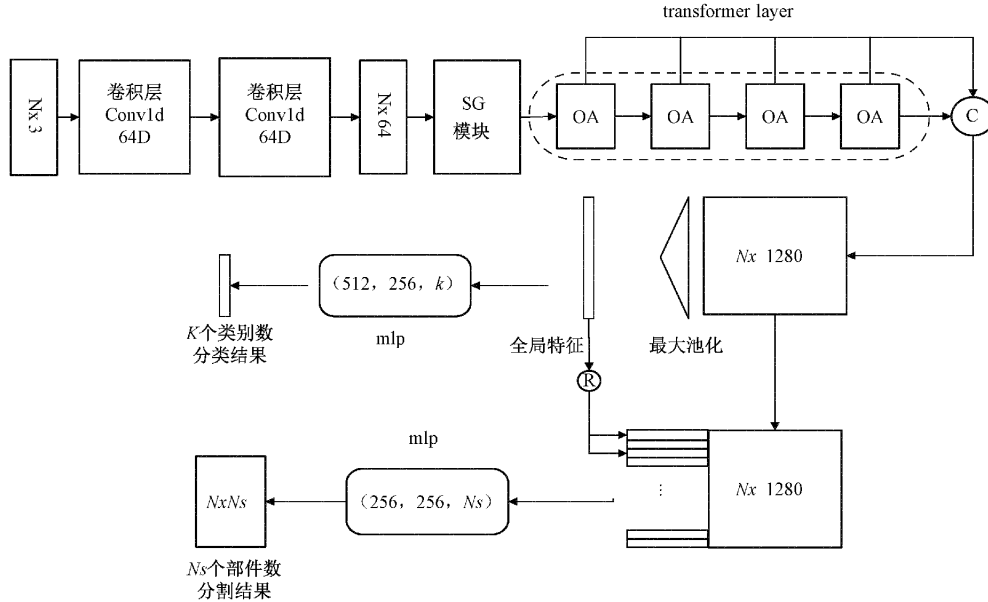


图 2 PCT 三维点云分类和分割网络结构

三维点云的分割网络的具体操作是将 1 024 维全局特征复制了 N 份,形成 $N \times 1\ 024$ 的矩阵,再和 1 280 维全局特征进行拼接,拼接后的特征既包括每个点的特征,也包括点云全局特征。将拼接后的特征输入到 3 层共享的 MLP 中,MLP 的层数分别是 256、256、Ns,其中 Ns 表示输出的分割部件的类别数。

1.2 改进的 SG 模块结构

PCT 三维点云形状分类网络旨在关注三维点云局部细粒度特征,用于区分不同形状差异性,因此对三维点云整体进行下采样,可以减少模型训练计算量,降低计算复杂度。PCT 网络中 SG 模块下采样操作降低了三维点云的分辨率,同时也使得三维点云密度变得稀疏,从而影响三维点云分类和分割精度。本文在 SG 模块中加入了 GAM 模块,旨在解决原来 SG 模块下采样过程导致的三维点云稀疏问题。GAM 是一个轻量级即插即用的嵌入模块^[13],GAM 模块的整体如图 3 所示。GAM 模块对中心点特征采用 k 最近邻算法(K-nearest neighbor,KNN)获取邻居点的特征, x_{ij} 表示中心点周围邻域内第 j 个邻居点的特征向量。再将特征 x_i 进行复制得到和邻居点相同的特征维度,两者特征相减,计算出中心点与邻居点之间的相对距离,再乘以系数矩阵 A,加上偏置矩阵 B,将局部点特征映射到一个相对规则的分布,保持三维点云的几何形状稳定,

改善了三维点云分类和分割网络的鲁棒性,几何仿射模块对邻居点特征的变换由式(1)所示。

$$x_{ij} = A \frac{x_{ij} - x_i}{\sigma + \epsilon} + B, = \sqrt{\frac{1}{k \times n \times d} \sum_{i=1}^n \sum_{j=1}^k (x_{ij} - x_i)^2} \quad (1)$$

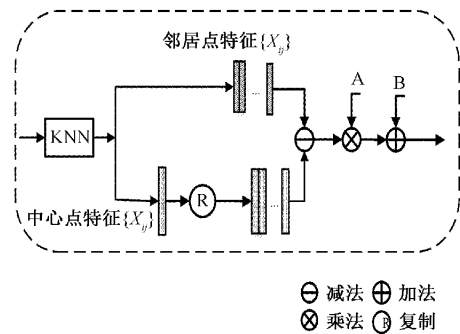


图 3 GAM 模块结构

本文对 SG 模块改进的具体操作如下:在 sg1 层和 sg2 层的 FPS 下采样输出端分别插入 GAM 模块,输入三维点云经过 sg1 层的 FPS 进行下采样获取 512 个中心点,得到的 64 维的中心点特征,GAM 对每个中心点的邻居点提取相同维度的特征,所有邻居点特征与中心点特征相减,得到了局部点的相对距离,再经过线性变换矩阵进行处理,

与中心点特征进行拼接,得到了变换后的 128 维特征。sg2 层再次通过 FPS 下采样获取 256 个点,得到 128 维的中心点特征,GAM 模块对具有 128 维特征的中心点进行线性变换,和邻居点进行特征拼接,再经过线性变换得到 256 维特征。接着连接一个卷积核为 256×1 的一维卷积和池化层处理三维点云特征,改进后的 SG 模块的结构如图 4 所示。

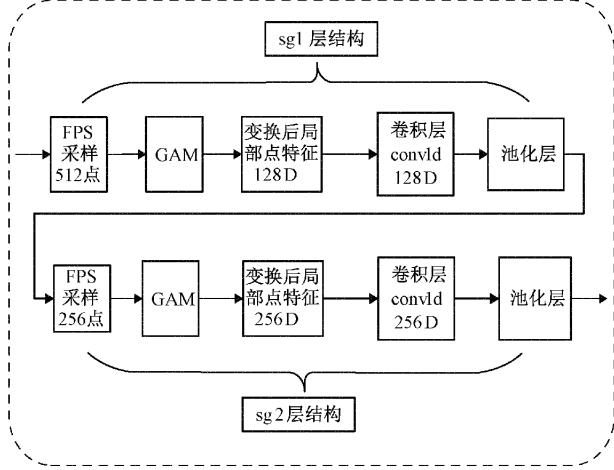


图 4 SG 模块结构

1.3 改进的自注意力模块

在文本分类任务中,自注意力机制将文本映射到 Q 、 K 、 V 3 个矩阵,并对 Q 与 K 作点积为每个单词分配注意力权重,用注意力权重对输入特征进行加权求和,即可得到带有单词之间特殊关联性的特征输出,整体计算流程如式(2)~(4)所示。

$$f(Q, K) = Q \cdot K^T \quad (2)$$

$$w_i = \text{softmax}(f(Q, K)) \quad (3)$$

$$\text{Attention}(Q, K, V) = \sum w_i V \quad (4)$$

w_i 表示经过 softmax 激活函数归一化处理得到的注意力权重值。

PCT 网络将自注意力模型中采用 QKV 特征矩阵引入到三维点云处理任务中,假定 PCT 网络的注意力层的输入特征为 F_m , 注意力层的输出特征为 F_{sa} , W_q 、 W_k 、 W_v 是 3 个参数可学习的点共享的线性变换, A 表示注意力权重矩阵,注意力特征的计算流程如式(5)~(7)所示。

$$(Q, K, V) = F_m \cdot (W_q, W_k, W_v) \quad (5)$$

$$A = Q \cdot K^T \quad (6)$$

$$F_{sa} = A \cdot V \quad (7)$$

本文采用向量注意力机制替换了 PCT 网络的自注意力模块,本文中向量注意力模块称为 PT block,PT block 模块的内部注意力机制流程如图 5 所示。在模块中添加位置编码用于网络获取三维点云局部特征,同时 GAM-PCT 网络改用减法关系的向量注意计算注意力权重,向量注意更具表现力,因为它支持对单个特征通道的自适应调

制,而不仅仅是整个特征向量,这种表现力在 3D 数据处理中效果明显。本文定义的位置编码函数为 δ 对三维点云局部位置进行编码^[16],如式(8)所示。

$$\delta = \theta(p_i - p_j) \quad (8)$$

其中, $p_i (i=1, 2, \dots, j, \dots, n)$ 表示三维点云单个点坐标, θ 函数用 MLP 处理点特征。

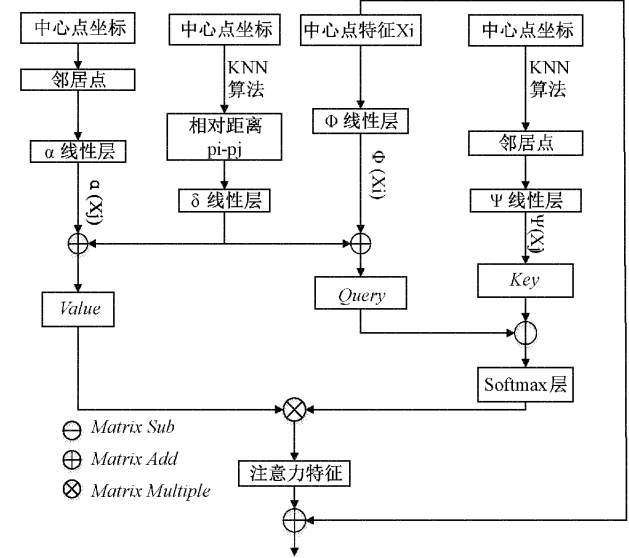


图 5 PT block 模块流程

其中, ϕ 、 α 、 φ 分别表示线性层,将特征向量映射到更高维度特征空间,向量注意力权重由 ψ 和 φ 层进行特征变换得到。 α 变换用于聚集邻居点特征,与向量注意力权重进行加权输出,PT block 模块计算特征算法如式(9)~(12)所示。

$$Q = \varphi(x_i) \quad K = \psi(x_j) \quad V = \alpha(x_j) \quad (9)$$

$$H_{x_j} = \text{softmax}(\gamma(Q - K + \delta)) \quad (10)$$

$$\text{Attention}(Q, K, V) = \sum H_{x_j} (V + \delta) \quad (11)$$

$$y_i = \sum_{x_j \in \mathcal{N}(x_i)} \rho(\gamma(\varphi(x_i) - \psi(x_j) + \delta)) \odot (\alpha(x_j) + \delta) \quad (12)$$

其中, x_i 表示最远点采样得到的中心点的特征向量, x_j 表示中心点周围的邻居点的特征向量。 γ 是具有两个线性层和一个 ReLU 非线性函数的 MLP, softmax 激活函数对关联性矩阵进行归一化,可以得到注意力权重矩阵 H_{x_j} , \odot 表示两个特征矩阵作哈达玛积运算。位置编码在自我注意中起着重要作用,GAM-PCT 网络将位置编码函数 δ 同步添加到了查询矩阵 Q 和值矩阵 V 中,让局部特征参与注意力特征的计算。

1.4 GAM-PCT 网络

本文提出 GAM-PCT 网络将 GAM 模块插入到 SG 模块中用于解决三维点云局部区域的稀疏和不规则的几何结构问题,针对 PCT 网络采用的标量注意力模块提取点间

局部特征,本文提出向量注意力模块 PT block,该模块采用减法关系和邻域位置编码对三维点云局部求取注意力

特征,替换了 PCT 网络中的标量注意力模块,GAM-PCT 网络整体框架如图 6 所示。

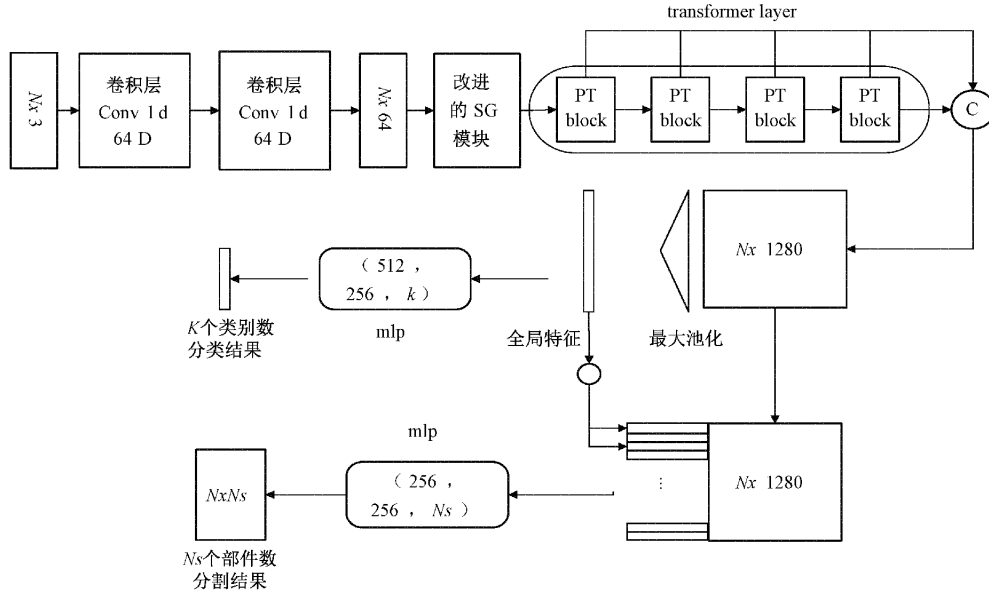


图 6 改进后的 GAM-PCT 三维点云分类和分割网络结构

GAM-PCT 网络的输入为原始三维点云, N 为三维点云单点个数, D 为每个点的特征维度, $D=3$ 表示输入三维点云数据只包含点的空间坐标 (x, y, z) 特征。三维点云数据是由 CAD 物体模型表面均匀采样获取 N 个点, 先通过两个一维卷积 Conv1d 对 3 维的点云特征转换为 64 维特征。改进后的 SG 模块对三维点云进行局部区域采样和分组, 对每个中心点采用 KNN 最近邻算法聚集中心点周围点的局部特征, 算法设置的邻居点的数量为 32。SG 模块的输出连接着 PT block 模块的输入, SG 模块输出了 256 维局部特征, 由于受到 DGCNN 网络框架的启发, 本文在 Transformer Layer 层采用 4 个 PT block 级联操作来提取局部特征, 假设: GAM-PCT 网络的 SG 模块输出的特征向量用 F_c 表示, 经过 transformer Layer 层输出的特征向量 F_o 可以用式 (13)~(15) 所示。

$$F_1 = \text{PT}^1(F_c) \quad (13)$$

$$F_i = \text{PT}^i(F_{i-1}), i = 2, 3, 4 \quad (14)$$

$$F_o = \text{concat}(F_1, F_2, F_3, F_4) \quad (15)$$

其中, PT^i 表示第 i 个 PT block 注意力模块, F_i 表示第 i 个 PT block 模块输出的三维点云特征向量, 每一个 PT block 输出的特征向量维度相同, 将所有的注意力模块的输出特征进行拼接输出高维特征。

2 实验验证

2.1 三维点云形状分类

本文所做实验采用的数据集是 Modelnet40^[17] 和 ScanObjectNN^[18]。Modelnet 系列数据集是用于三维点云处理任务研究的合成数据集。ModelNet40 数据集里面一

共包含有 12 311 个 CAD 物体模型, 其中包括 40 个物体类别。9 843 个模型为训练集用于训练, 2 468 个模型为测试集用于测试。Modelnet40 数据集是 CAD 模型, 在 CAD 模型的表面上均匀采样 1 024 个点得到物体的三维点云数据, 对于每个点只保留它的三维坐标 (x, y, z) 作为三维点云分类和分割网络的输入。不同于 Modelnet40, ScanObjectNN 数据集是通过激光雷达传感器采集的现实世界的真实数据, 物体三维点云中还含有背景信息和噪声干扰, 更适用于检验三维点云分类和分割模型的泛化能力。

2.2 实验环境设置

本文的三维点云分类和分割模型训练和测试的实验环境是基于 Linux ubuntu16.04 操作系统、NVIDIA GeForce RTX 2080Ti GPU、Intel Core (TM) i5-9600K CPU、内存 32 GB, 采用的深度学习框架为 Pytorch1.2 版本, 编程语言采用 python3.6, 模型运算平台选用 CUDA-Toolkit 10.1, 采用 cuDNN7.5 作为 CUDA 的运算加速库。模型训练的过程中采用 SGD 优化器, 其调整动量设置为 0.9, 初始学习率设置为 0.001, 并设置学习率动态衰减, 训练批次为 32, 训练轮次为 250。

2.3 实验结果分析

本文方法和一些成熟的三维点云分类和分割模型做出分类性能对比, 所有实验的输入为原始三维点云的三维坐标, 点的数量均为 1 k, 分类结果如表 1 所示。本文方法在 ModelNet40 数据集上的分类准确度达到 93.5%, 比 PCT 分类网络的精度提升了 0.3%, 本文方法采用了和 DGCNN 分类网络类似的 4 层相同模块堆叠级联结构, 由

实验结果可以表明,多层向量注意力机制级联输出,可以充分提取三维点云细粒度局部特征,对于提高分类精度有重要作用。在总体分类精度方面,GAM-PCT 网络比 DGCNN 高出了 0.6%,比 PointNet 高出了 4.3%,比 PointNet++ 高出了 2.8%,本文采用 Pytorch 深度学习框架对模型进行训练,模型在 ModelNet40 数据集上测试的分类准确率曲线如图 7 所示,模型训练到 250 个轮次后,基本趋向收敛。本文选用投票策略对训练好的模型进行投票测试,训练轮次为 300 轮,投票数设置为 10,最后经过投票得到模型的综合分类精度为 93.5%,投票过程曲线如图 8 所示。

表 1 不同算法在 ModelNet40 数据集上的分类结果

| 方法 | 输入点数 | 分类准确度/% |
|--------------------------------|------|-------------|
| PointNet | 1k | 89.2 |
| PointNet++ | 1k | 90.7 |
| DGCNN | 1k | 92.9 |
| PointCNN ^[19] | 1k | 92.2 |
| Point2Sequence ^[20] | 1k | 92.6 |
| RS-CNN | 1k | 92.9 |
| PCT | 1k | 93.2 |
| GAM-PCT | 1k | 93.5 |

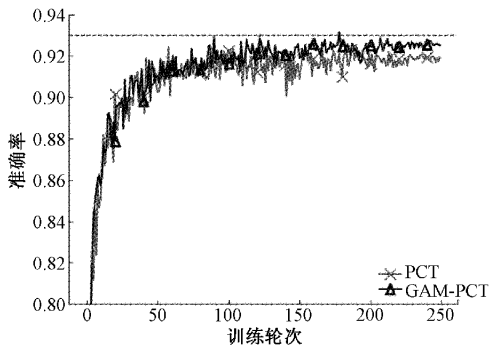


图 7 ModelNet40 数据集训练结果

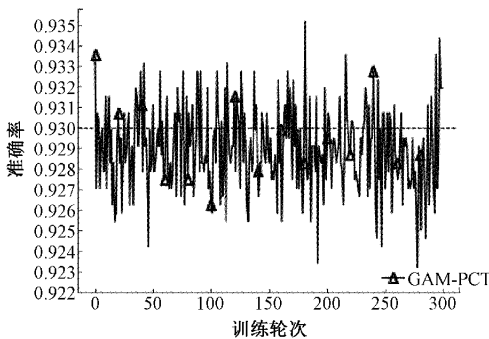


图 8 模型投票结果

本文在 ScanObjectNN 数据集上验证模型的综合性能,分别用 PCT 和 GAM-PCT 网络在该数据集上进行模

型训练,两个方法的训练对比结果如图 9 所示。本文选择如下算法作为对比方法: PointNet、PointNet++、DGCNN、PCT、SpiderCNN^[21]、PointCNN、PointConv^[22]。其中 OA (overall accuracy) 表示整体分类准确度, mAcc (Mean Accuracy) 表示类平均分类精度。各方法在 ScanObjectNN 数据集上的分类性能对比如表 2 所示。实验结果表明,本文方法在总体分类精度上面,比 PCT 网络提升了 1.9%,相比其他的方法在精度上面都有一定的提升。

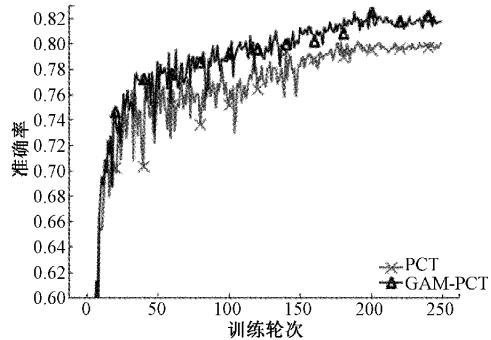


图 9 ScanObjectNN 数据集训练结果

表 2 不同算法在 ScanObjectNN 数据集上的分类结果

| 方法 | 输入点数 | mAcc/% | OA/% |
|------------|------|-------------|-------------|
| PointNet | 1 k | 63.4 | 68.2 |
| PointNet++ | 1 k | 75.4 | 77.9 |
| SpiderCNN | 1 k | 69.8 | 73.7 |
| DGCNN | 1 k | 73.6 | 78.1 |
| PointCNN | 1 k | 75.1 | 78.5 |
| PointConv | 1 k | 74.1 | 77.4 |
| PCT | 1 k | 76.3 | 80.1 |
| GAM-PCT | 1 k | 76.9 | 82.0 |

为了验证改进方法的有效性,本文继续比较 GAM-PCT 和 PCT 分类网络的稳定性问题。具体地,本文分别绘制了 GAM-PCT 和 PCT 网络的训练损失函数值随训练轮次的变化曲线,如图 10 所示。

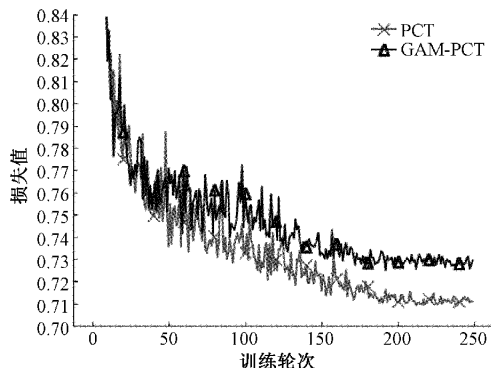


图 10 ScanObjectNN 数据集训练损失曲线

由图 10 可知,PCT 分类网络和本文方法的损失曲线在第 200 个 epoch 左右基本趋向收敛,PCT 网络的损失曲线的震荡幅度变化更大,可见 GAM-PCT 网络的分类稳定性更强。可以发现,在网络接近收敛时,GAM-PCT 的损失值大于 PCT 网络,说明了其拟合训练数据的能力更低,这一点正是由于本文分类网络的参数量少导致的。同时,在 ScanObjectNN 数据集上的分类实验结果可以发现,GAM-PCT 网络的分类准确率高于 PCT 网络,这说明了本文分类网络具有更强的泛化能力。

2.4 部件分割

三维点云的分割本质上是更细粒度的分类任务。本文提出的 GAM-PCT 网络在 ShapeNet^[23]数据集上进行实验,并和现有的分割网络做了性能对比。ShapeNet 一共包含有 16 881 个 CAD 物体模型,总共含有 16 个物体类别,

包含 50 种部件种类。在 CAD 模型表面随机选取 2 048 个点形成三维点云作为分割网络的输入数据。

三维点云分割实验通常采用平均交并比(mIoU)作为衡量分割任务性能,表 3 给出了不同网络在 ShapeNet 数据集上的部件分割的实验结果。由表 3 可以看出,提出的 GAM-PCT 网络得到的 mIoU 值比 DGCNN 高出了 1.4%,比 PCT 高出了 0.2%,在特定的类别如飞机、包和吉他等类别取得了更好的分割效果,图 11 为 GAM-PCT 和 PCT 网络在 ShapeNet 数据集上的分割效果对比图。图 11(a)表示分割真值图,图 11(b)表示 PCT 分割效果图,图 11(c)表示 GAM-PCT 分割效果图,从图中看出,GAM-PCT 网络在物体部件种类分界处的效果更好,例如椅子的扶手、汽车和摩托车的车轮和车轮轴、桌子的支架、飞机的机尾等,由此说明了本文网络改进策略的有效性。

表 3 不同算法在 ShapeNet Parts 数据集上的零件分割实验结果

| 方法 | mIoU | Air-plane | bag | cap | car | chair | Ear-phone | guitar | knife |
|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| PointNet | 83.7 | 83.4 | 78.7 | 82.5 | 74.9 | 89.6 | 73.0 | 91.5 | 85.9 |
| Kd-Net | 82.3 | 80.1 | 74.6 | 74.3 | 70.3 | 88.6 | 73.5 | 90.2 | 87.2 |
| SO-Net | 84.9 | 82.8 | 77.8 | 88.0 | 77.3 | 90.6 | 73.5 | 90.7 | 83.9 |
| PointNet+++ | 85.1 | 82.4 | 79.0 | 87.7 | 77.3 | 90.8 | 71.8 | 91.0 | 85.9 |
| DGCNN | 85.2 | 84.0 | 83.4 | 86.7 | 77.8 | 90.6 | 74.7 | 91.2 | 87.5 |
| P2Sequence | 85.2 | 82.6 | 81.8 | 87.5 | 77.3 | 90.8 | 77.1 | 91.1 | 86.9 |
| RS-CNN | 86.2 | 83.5 | 84.8 | 88.8 | 79.6 | 91.2 | 81.1 | 91.6 | 88.4 |
| PCT | 86.4 | 85.0 | 82.4 | 89.0 | 81.2 | 91.9 | 71.5 | 91.3 | 88.1 |
| GAM-PCT | 86.6 | 85.4 | 83.5 | 90.7 | 78.1 | 91.5 | 77.5 | 92.4 | 87.1 |
| 方法 | lamp | laptop | Motorbike | mug | pistol | rocket | Skateboard | table | |
| PointNet | 80.8 | 95.3 | 65.2 | 93.0 | 81.2 | 57.9 | 72.8 | 80.6 | |
| Kd-Net | 81.0 | 94.9 | 57.4 | 86.7 | 78.1 | 51.8 | 69.9 | 80.3 | |
| SO-Net | 82.8 | 94.8 | 69.1 | 94.2 | 80.9 | 53.1 | 72.9 | 83.0 | |
| PointNet+++ | 83.7 | 95.3 | 71.6 | 94.1 | 81.3 | 58.7 | 76.4 | 82.6 | |
| DGCNN | 82.8 | 95.7 | 66.3 | 94.9 | 81.1 | 63.5 | 74.5 | 82.6 | |
| P2Sequence | 83.9 | 95.7 | 70.8 | 94.6 | 79.3 | 58.1 | 75.2 | 82.8 | |
| RS-CNN | 86.0 | 96.0 | 73.7 | 94.1 | 83.4 | 60.5 | 77.7 | 83.6 | |
| PCT | 86.3 | 95.8 | 64.6 | 95.8 | 83.6 | 62.2 | 77.6 | 83.7 | |
| GAM-PCT | 84.2 | 94.3 | 68.5 | 93.4 | 81.7 | 60.4 | 78.2 | 82.5 | |

2.5 模型计算需求

模型的整体复杂度和计算开销取决于所需的 FLOPs(浮点操作)数量和模型参数量大小,本文方法和基于注意力的 PCT 模型的复杂度在 ModelNet40 数

据集上的计算结果如表 4 所示。本文模型的参数量大小相比于 PCT 减少了 0.69 M,处理器上只有 2.01 G FLOPs 的低负载,但模型的分类准确度提升了 0.3%。

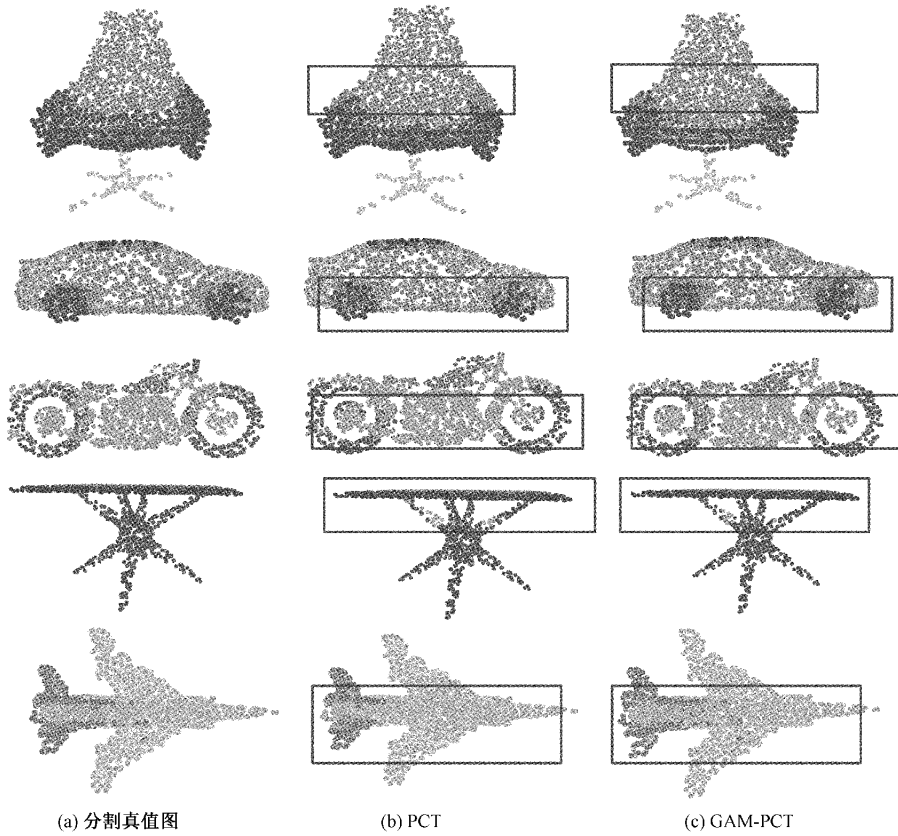


图 11 ShapeNet 数据集上 PCT、GAM-PCT 的部件分割效果

表 4 模型参数量和分类准确率对比表

| 方法 | 模型参数量/ FLOPs/ 分类准确度/ | | |
|-----------------|----------------------|------|------|
| | M | G | % |
| PointNet | 3.47 | 0.45 | 89.2 |
| PointNet++(SSG) | 1.48 | 1.68 | 90.7 |
| PointNet++(MSG) | 1.74 | 4.09 | 91.9 |
| DGCNN | 1.81 | 2.43 | 92.6 |
| PCT | 2.88 | 2.32 | 93.2 |
| GAM-PCT | 2.19 | 2.01 | 93.5 |

3 结 论

本文提出了一种基于向量注意力机制的点云分类网络 GAM-PCT,该网络改进了 PCT 网络中的 SG 模块,采用两种不同尺度的采样点数对整体点云进行最远点采样,插入几何仿射模块克服稀疏点云特征提取问题,采用向量注意力机制提取局部注意力特征,并通过多个 PT block 模块级联的方式对注意力特征进行拼接得到细粒度局部特征。通过实验表明,本文提出的网络与 PCT 相比,模型自身的复杂度更低,降低了计算开销,达到了更高的分类和分割准确度。

参考文献

[1] SU H, MAJI S, KALOGERAKIS E, et al. Multi-

view convolutional neural networks for 3d shape recognition [C]. IEEE International Conference on Computer Vision, 2015: 945-953.

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

[3] MATURANA D, SCHERER S. VoxNet: A 3D convolutional neural network for real-time object recognition [C]. IEEE International Conference on Intelligent Robots and Systems (IROS), 2015: 922-928.

[4] QI C R, 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, 2017: 77-85.

[5] QI C R, YI L, SU H, et al. Pointnet++: Deep hierarchical feature learning on point sets in a metric space [C]. Annual Conference on Neural Information Processing Systems (NIPS), 2017: 5105-5114.

[6] 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.

[7] 朱茂然, 王奕磊, 高松, 等. 中文比较关系的识别: 基

- 于注意力机制的深度学习模型[J]. 情报学报, 2019, 38(6):612-621.
- [8] LUONG M T, PHAM H, MANNING C D. Effective approaches to attention-based neural machine translation[J]. Computer ence, 2015, DOI:10.48550/arXiv.1508.04025.
- [9] GUO M H, CAI J X, LIU Z N, et al. Pct: Point cloud transformer[J]. Computational Visual Media, 2021, 7(2):187-199.
- [10] LIU Y, FAN B, XIANG S, et al. Relation-shape convolutional neural network for point cloud analysis[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2019: 8887-8896.
- [11] WANG L, HUANG Y, HOU Y, et al. Graph attention convolution for point cloud semantic segmentation[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2019:10288-10297.
- [12] YANG J, ZHANG Q, NI B, et al. Modeling point clouds with self-attention and gumbel subset sampling[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2019: 3318-3327.
- [13] SAINING X, LIU S, CHEN Z, et al. Attentional ShapeContextNet for point cloud recognition [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2018: 4606-4615.
- [14] 于魁梧, 宋玉琴, 徐轩. 基于双注意力融合和残差优化的点云语义分割[J]. 国外电子测量技术, 2022, 41(8): 12-18.
- [15] MA X, QIN C, YOU H, et al. Rethinking network design and local geometry in point cloud: A simple residual MLP framework[J]. Arxiv Preprint, 2022, ArXiv:2202.07123.
- [16] ZHAO H, JIANG L, JIA J, et al. Point transformer[C]. IEEE International Conference on Computer Vision, 2021: 16239-16248.
- [17] 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, 2015: 1912-1920.
- [18] UY M A, PHAM Q H, HUA B S, et al. Revisiting point cloud classification: A new benchmark dataset and classification model on real-world data[C]. IEEE International Conference on Computer Vision, 2019: 1588-1597.
- [19] ATZMON M, MARON H, LIPMAN Y. Point convolutional neural networks by extension operators [J]. ACM Transactions on Graphics, 2018, DOI: 10.1145/3197517.3201301.
- [20] 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]. AAAI Conference on Artificial Intelligence, 2019: 8778-8785.
- [21] XU Y F, FAN T Q, XU M Y, et al. SpiderCNN: Deep learning on point sets with parameterized convolutional filters[C]. European Conference on Computer Vision (ECCV), 2018, 11212: 90-105.
- [22] WU W X, QI Z G, LI F X. PointConv: Deep convolutional networks on 3D point clouds[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2019: 9613-9622.
- [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): 1-12.

作者简介

胡平安, 硕士研究生, 主要研究方向为模式识别、机器学习、三维点云。

E-mail: 2337780019@qq.com

李光平(通信作者), 博士后, 副教授, 主要研究方向为计算机视觉、毫米波雷达信号处理、无线能量传输。

E-mail: gpli@gdut.edu.cn