

基于局部区域动态覆盖的 3D 点云分类方法*

王昌硕^{1,2,3}, 王含^{4,5}, 宁欣^{1,3}, 田生伟⁶, 李卫军^{1,2,7}



¹(中国科学院 半导体研究所, 北京 100083)

²(中国科学院大学 材料科学与光电工程中心&微电子学院, 北京 100049)

³(威富集团形象认知计算联合实验室, 北京 100083)

⁴(广州南方学院, 广东 广州 510970)

⁵(珠海复旦创新研究院, 广东 珠海 518057)

⁶(新疆大学 软件学院, 新疆 乌鲁木齐 830091)

⁷(半导体神经网络智能感知与计算技术北京市重点实验室, 北京 100083)

通信作者: 宁欣, E-mail: ningxin@semi.ac.cn

摘要: 局部几何形状的描述能力, 对不规则的点云形状表示是十分重要的. 然而, 现有的网络仍然很难有效地捕捉准确的局部形状信息. 在点云中模拟深度可分离卷积计算方式, 提出一种新型的动态覆盖卷积(dynamic cover convolution, DC-Conv), 以聚合局部特征. DC-Conv 的核心是空间覆盖算子(space cover operator, SCOP), 该算子通过在局部区域中构建各向异性的空间几何体覆盖局部特征空间, 以加强局部特征的紧凑性. DC-Conv 通过在局部邻域中动态组合多个 SCOP, 实现局部形状的捕捉. 其中, SCOP 的注意力系数通过数据驱动的方式由点位置自适应地学习得到. 在 3D 点云形状识别基准数据集 ModelNet40, ModelNet10 和 ScanObjectNN 上的实验结果表明, 该方法能有效提高 3D 点云形状识别的性能和对稀疏点云的鲁棒性. 最后, 也提供了充分的消融实验验证该方法的有效性. 源代码发布在 <https://github.com/changshuwang/DC-CNN>.

关键词: 点云分类; 动态覆盖卷积; 空间覆盖算子; 局部邻域; 注意力系数

中图法分类号: TP391

中文引用格式: 王昌硕, 王含, 宁欣, 田生伟, 李卫军. 基于局部区域动态覆盖的 3D 点云分类方法. 软件学报, 2023, 34(4): 1962–1976. <http://www.jos.org.cn/1000-9825/6683.htm>

英文引用格式: Wang CS, Wang H, Ning X, Tian SW, Li WJ. 3D Point Cloud Classification Method Based on Dynamic Coverage of Local Area. Ruan Jian Xue Bao/Journal of Software, 2023, 34(4): 1962–1976 (in Chinese). <http://www.jos.org.cn/1000-9825/6683.htm>

3D Point Cloud Classification Method Based on Dynamic Coverage of Local Area

WANG Chang-Shuo^{1,2,3}, WANG Han^{4,5}, NING Xin^{1,3}, TIAN Sheng-Wei⁶, LI Wei-Jun^{1,2,7}

¹(Institute of Semiconductors, Chinese Academy of Science, Beijing 100083, China)

²(Center of Materials Science and Optoelectronics Engineering & School of Microelectronics, University of Chinese Academy of Sciences, Beijing 100049, China)

³(Cognitive Computing Technology Joint Laboratory, Wave Group, Beijing 100083, China)

⁴(Nanfang College Guangzhou, Guangzhou 510970, China)

⁵(Zhuhai Fudan Innovation Institute, Zhuhai 518057, China)

⁶(School of Software, Urumqi 830091, China)

* 基金项目: 国家自然科学基金(61901436); 广东省重点领域研发计划(2019B010107001); 中国科学院 B 类先导科技专项培育项目(XDPB22)

收稿时间: 2021-11-24; 修改时间: 2022-01-07, 2022-02-10; 采用时间: 2022-03-17; jos 在线出版时间: 2022-09-20

⁷(Beijing Key Laboratory of Semiconductor Neural Network Intelligent Sensing and Computing Technology, Beijing 100083, China)

Abstract: The ability to describe local geometric shapes is very important for the representation of irregular point cloud. However, the existing network is still difficult to effectively capture accurate local shape information. This study simulates depthwise separable convolution calculation method in the point cloud and proposes a new type of convolution, namely dynamic cover convolution (DC-Conv), to aggregate local features. The core of DC-Conv is the space cover operator (SCOP), which constructs anisotropic spatial geometry in a local area to cover the local feature space to enhance the compactness of local features. DC-Conv achieves the capture of local shapes by dynamically combining multiple SCOPs in the local neighborhood. Among them, the attention coefficients of the SCOPs are adaptively learned from the point position in a data-driven way. Experiments on the 3D point cloud shape recognition benchmark dataset ModelNet40, ModelNet10, and ScanObjectNN show that this method can effectively improve the performance of 3D point cloud shape recognition and robustness to sparse point clouds even in the case of a single scale. Finally, sufficient ablation experiments are also provided to verify the effectiveness of the method. The open-source code is published at <https://github.com/changshuwang/DC-CNN>.

Key words: point cloud classification; dynamic cover convolution; space cover operator; local neighborhood; attention coefficient

当前, 计算机感知外界事物通常处理的是由相机投影而来的 2D 图像。然而, 2D 图像丢失了 3D 世界的相对位置和几何深度信息, 容易造成背景干扰等问题。近年来, 随着激光雷达等传感设备的可用性不断提高, 3D 视觉技术已经广泛应用于众多领域, 例如机器人技术^[1,2]、自动驾驶^[3,4]、遥感^[5,6]等。其中, 3D 点云是描述物体三维结构的主要数据形式, 但是点云的无序性、不规则性, 导致目前的许多技术无法有效地捕捉准确的局部形状信息。

点云是一组在 3D 物体表面离散采样的无序点的集合。这些点通常用 3D 坐标 (x,y,z) 表示, 有时也会携带其他附加信息, 比如 RGB 信息、表面法向量和强度等信息。改变点的顺序不影响对物体的表示。因此, 2D 卷积神经网络不能直接用来处理这种不规则的点云信息。随着目前深度学习在 2D 领域取得显著的成功, 一些研究者^[7-10]试图将 3D 点云映射为 2D 图像, 然后使用 2D 卷积神经网络处理映射后的 2D 图像, 或者利用体素化技术^[11,12]将非结构的点云转换为结构化的网格数据, 然后使用 3DCNN 进行处理。这些间接处理点云的方法不仅需要占用较大的内存和计算复杂度, 而且也忽略了大量的几何信息。

目前的主流方法是在点云上直接使用深度学习技术。PointNet^[13]是该类方法的开山之作。该工作通过共享的多层感知器(multi-layer perceptrons, MLP), 直接将每个点映射到高维空间, 然后使用最大值池化聚合物体的全局特征。但是该方法仅仅是孤立地处理每个点, 忽略了点对之间隐含的形状信息。为了解决该问题, PointNet++^[14]通过堆叠集合抽象层(set abstraction layer, SA-Layer)提取物体全局特征。在每个 SA-Layer 中, 通过最远点采样(farthest point sampling, FPS)的方法确定每个采样点, 然后在每个采样点, 通过 k 近邻(k -nearest neighbor, KNN)或者球形查询得到局部邻域。最后, 在每个局部邻域使用 PointNet 作为局部特征提取器。但是这种方法在每个局部邻域依然没有提取到有效的局部信息。后续的一些方法^[15-17]基本都是沿着这种局部特征提取方法研究。尽管这种方式有利于提取局部形状信息, 但局部邻域的点与点之间通过简单的仿射基函数(affine basis function, ABF)^[18]进行映射, 如图 1(a)所示, 其输出实质上是局部邻域点在特征空间中到一个超平面的距离, 这种方式表达能力有限, 而且容易忽略局部特征在空间的分布关系。另一方面, 现有的网络为了增加网络的表达能力, 通常采用增加多尺度、增加通道数量和网络深度等策略, 这会使用更多的算力。因此, 局部特征表示仍然存在很大的挑战。

为了解决上述问题, 如图 1(b)所示, 考虑到局部邻域的空间性和相似性^[19,20], 我们试图在局部邻域构建多个各向异性的空间几何体感知潜在的形状信息。因此, 我们借鉴深度可分离卷积的运算形式, 提出一种新型的 DC-Conv。首先, 我们提出一种 SCOP 作为深度卷积。该算子的权重由邻域点和采样点的坐标通过 MLP 学习得到, 邻域点的边缘特征作为该算子的输入特征。该算子的实质就是在局部邻域的特征空间中建立各项异性的空间几何体, 从而从坐标空间和特征空间中挖掘局部邻域中隐含的高阶形状信息。然后, 我们在局部邻域中使用多个 SCOP, 并使用注意力系数将这些算子动态组合在一起。其中, SCOP 的注意力系数由邻域点的坐标自适应学习得到。最后, 通过 MLP 作为逐点卷积, 将 SCOP 的输出映射到高维空间, 实现更强大的形状感知表示。通过分层堆叠 DC-Conv, 构建了用于点云形状识别的动态覆盖卷积神经网络(dynamic cover

convolutional neural network, DC-CNN). 通过实验证明: 即使在单尺度的情况下, DC-Conv 也可以挖掘到局部邻域更丰富的形状信息.

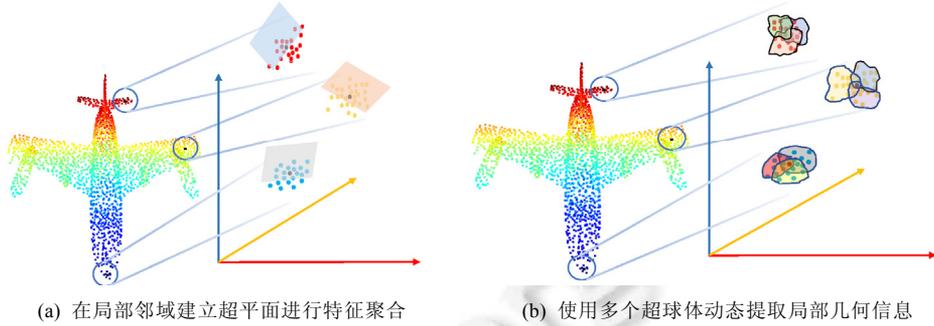


图 1

因此, 本文的主要贡献有以下 3 个方面.

- (1) 在本文中, 我们提出一种新型的 SCOP. 该算子通过在局部邻域中构建各向异性的空间几何体, 把握局部邻域坐标空间和特征空间的亲和力;
- (2) 为了提高局部邻域表示能力, 我们提出一种 DC-Conv. 该卷积的权重由点坐标自适应学习得到, 邻域点的边缘特征坐标作为该卷积的输入特征, 并从点坐标中自适应学习注意力系数, 将多个 SCOP 进行动态组装, 从而能够更加灵活地挖掘局部邻域中隐含的丰富的形状信息;
- (3) 通过分层堆叠 DC-Conv, 构建了 DC-CNN. 实验结果表明: 在单尺度的情况下, 该网络在点云分类中就能取得优越的分类效果.

1 相关研究工作

1.1 基于投影的方法

早期的一些学者^[8-10,21]从多个视角对点云进行投影, 然后使用卷积神经网络对投影的 2D 图像进行处理. MVCNN^[21]是多视角学习的开山之作, 它将每个视图通过共享的卷积层和池化层整合成描述 3D 物体的全局特征, 然后将聚合后的特征输入到全连接层, 得到分类结果. 后续的一些作品主要解决视角选择和视角信息融合问题. 但是基于多视图的方法仍然需要人为设定视角, 对遮挡也比较敏感. 同时, 这类方法在一定程度上导致了几何结构信息的丢失.

另一种基于投影的方法^[11,12,22]是利用体素的方法将 3D 物体转换为 3D 网格, 然后应用 3DCNN 对 3D 网格进行进一步处理. VoxNet^[11]是最早基于体素方法的 3D CNN 模型, 该模型展示了 3D CNN 从体素化网格中学习特征的潜力. 虽然体素模型的提出解决了点云无序性和非结构化的问题, 但三维数据的稀疏性与空间信息的不完整性, 导致体素化方法计算开销更大, 并且限制了体素模型的分辨率.

1.2 直接处理点云的方法

1.2.1 非局部特征聚合

该类方法通过构建非局部网络, 直接计算任意两个位置之间的交互关系来捕获远程依赖性. PointNet^[13]是直接处理点云的开山之作, 它直接通过共享的 MLP 将每个点映射到高维空间, 然后使用最大值池化聚合所有点的信息, 从而得到描述 3D 物体的全局特征. 但是该方法仅仅孤立地处理每个点, 忽略了点间的依赖关系. Guo 等人^[23]提出一种 PCT (point cloud transformer)模型, 将 Transformer 应用到 3D 点云分析. 该模型的核心思想是: 利用 Transformer 固有的排列不变性, 避免需要定义点云数据的顺序, 并通过注意力机制进行特征学习. CurveNet^[24]在点云曲面上生成连续的点线段序列, 来学习点云的非局部特征聚合.

1.2.2 局部特征聚合

目前, 许多工作主要是以解决局部特征聚合问题而提出的. 该类方法通过在采样点周围按照某种规则确

定局部邻域, 然后计算采样点和局部点的位置和特征关系, 最后将其传递到局部特征聚合模块中提取局部特征. 在基于逐点 MLP 的方法中, PointNet++^[14]为了解决 PointNet 忽略局部几何结构的问题, 通过 FPS 将点云划分为有重叠的局部邻域, 并在每个邻域中使用 PointNet 作为特征提取器进行局部特征聚合. 该方法在局部特征聚合方面, 仍然没有充分学习局部点间的几何关系. 结构关系网络(SRN-PointNet++)^[25]通过学习局部区域的几何关系和位置关系, 建模局部区域之间的结构依赖关系. RandLA-Net^[26]使用随机采样法替代 FPS, 然后通过局部特征聚集模块捕获局部几何特征, 该方法在存储和计算方面实现了显著的提高. PointWeb^[27]通过自适应特征调整模块学习点之间的局部特征. PointASNL^[28]通过自适应模块减弱噪声和异常值, 然后通过 local-nonlocal 模块聚合局部特征和全局特征. 在基于卷积的方法中, PointCNN^[29]提出了 X-Conv 算子探索点云的排列顺序, 然后用常规卷积处理局部邻域. PointConv^[30]根据 3D 连续卷积的蒙特卡洛估计, 在局部邻域上实现了卷积操作. Thomas 等人^[31]通过一组可学习的内核点, 设计了一种可形变的 KPConv. RS-CNN^[32]通过将邻域点和采样点的低层关系映射到高维空间, 然后通过 MLP 实现卷积操作. 还有一些工作基于图的方法^[15,33,34]聚合局部图形的特征.

基于局部特征聚合的方法大都采用共享的 MLP 来转换点的特征, MLP 在表示能力上是有限的, 而且多尺度架构会引起感受野的重叠, 导致计算冗余. 因此, 在本文中, 我们根据局部邻域的几何关系提出一种 DC-Conv, 以提高网络对点云复杂空间变化的拟合能力

1.3 动态卷积学习

由于小型网络深度和宽度的限制, 在局部区域使用单一的卷积核, 限制了网络的表达能力. 因此, 在 2D 图像分析领域, 动态卷积已经被广泛研究. DY-CNN^[35]在一个卷积层中用到了多个卷积核, 并且用注意力机制去结合不同卷积核的输出信息. CondConv^[36]通过一组组合系数将多个卷积核组合一起, 再处理输入信息. DRConv^[37]主要是为具有特征相似表示的相应空间区域自动地分配定制的卷积核. 在 3D 点云中, 仅有 PAConv^[38]提出了一种即插即用的动态卷积核. 但是该方式使用共享的 MLP 变换函数, 表达能力有限. 另一方面, 权重矩阵忽略了点的空间关系. 为此, 我们提出了 DC-Conv, 该卷积的卷积核和组合系数均由局部领域点的几何先验中自适应学习得到, 可以获得潜在的局部几何关系, 即使在单尺度的情况下, 也实现了卓越的效果.

2 基础知识

在局部特征聚合中, 通常在某一层中使用 FPS 得到采样点集, 然后通过球形查询或者最近邻确定每个采样点的局部邻域. 假设在某一层有 N_s 个采样点, 该采样点集可以表示为 $P_s = \{N_1, N_2, \dots, N_s\}$, d 通常为采样点的坐标维度, 即 $d=3$. 该点集的特征集合可表示为 $F_s = \{f_i | i=1, 2, \dots, N_s\} \in R^{N_s \times c}$, c 为特征通道数. 假设每个局部邻域有 k 个邻居点, 则采样点 p_i 的局部邻域可以表示为 $p_{i,1}, \dots, p_{i,k} \in \mathcal{N}(p_i)$, 对应的特征为 $f_{i,1}, \dots, f_{i,k}$.

2D 连续卷积可以表述为

$$s(x, y) = f * w = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x+u, y+v)w(u, v)du dv \quad (1)$$

图像处理中常用的是离散卷积, 并且基本是通过互相关函数实现的, 即:

$$s(x, y) = \sum_u \sum_v F(x+u, y+v)W(u, v) \quad (2)$$

借鉴 2D 的卷积计算方式, 在 3D 点云中使用的卷积计算可以表述为

$$s_i = A(\{L(\varphi(p_i, p_{ij}), \theta(f_i, f_{ij})) | \forall p_{ij} \in \mathcal{N}(p_i)\}) \quad (3)$$

其中, $\varphi(p_i, p_{ij})$ 是表示邻域点几何先验的函数, 它可以为 p_i 和 p_{ij} 的位置差、3D 距离或者拼接; $\theta(f_i, f_{ij})$ 表示 f_i 和 f_{ij} 之间的边缘特征或者单独的 f_i 或 f_{ij} ; L 为邻域点的特征变换函数, 通常情况下, $L = W^T f_{ij}$ 或者 $L = w_{ij} \odot f_{ij}$; A 为对称函数(例如平均池化、最大值池化或者求和池化), 其作用是将局部邻域点的特征 f_{ij} 在通道方向上进行聚合, 从而将整个局部邻域特征归纳为 s_i .

在 2D 图像分析中, 假设卷积层的输入特征图大小为 $H_i \times W_i \times C_i$, 输出特征图大小为 $H_o \times W_o \times C_o$, 卷积核大小为 $H_k \times W_k \times C_i \times C_o$, 且卷积层使用填充方法使输入和输出具有相同的空间大小, 则常规卷积^[39]的参数量为 $H_k \times W_k \times C_i \times C_o \times H_i \times W_i$. 深度可分离卷积^[40](如图 2(b)所示)由深度卷积和逐点卷积构成. 类似地, 假设卷积层的输入特征图大小为 $H_i \times W_i \times C_i$, 输出特征图大小为 $H_o \times W_o \times C_o$, 深度卷积的卷积核大小为 $H_k \times W_k \times C_i \times 1$, 逐点卷积的卷积核大小为 $1 \times 1 \times C_i \times C_o$, 则深度卷积的计算量为 $H_k \times W_k \times C_i \times H_i \times W_i$, 逐点卷积的计算量为 $C_i \times C_o \times H_i \times W_i$. 因此, 深度可分离卷积和常规卷积的参数量之比为 $\frac{1}{C_o} + \frac{1}{H_k + W_k}$, 并且深度可分离卷积可以以更少的参数学习到丰富的特征. 如上所述: 当 $L=W^T f_{ij}$ 时, 公式(3)的本质是常规的卷积操作; 当 $L=w_{ij} \odot f_{ij}$ 时, 公式(3)的本质为深度卷积. 但是深度卷积的特征通道没有变化, 这限制了特征的表达. 因此, 我们参照深度可分离卷积的运算形式, 介绍用于点云分类的 DG-Conv 的构建过程.

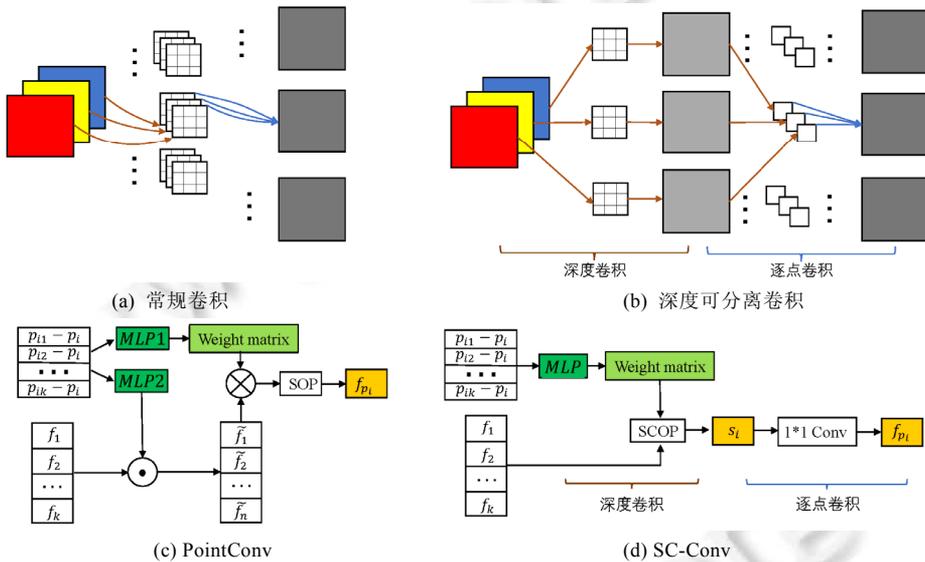


图 2 图像和点云中的卷积

图 2 中, 图 2(a)为 2D 常规卷积; 图 2(b)为 2D 深度可分离卷积; 它由深度卷积和逐点卷积组成; 图 2(c)为 PointConv, 它是图 2(a)在 3D 点云中的扩展, SOP 表示对称操作; 图 2(d)为空间覆盖卷积(space cover convolution, SC-Conv), 它是图 2(b)在 3D 点云中的扩展, 并且只是用一个 SCOP, SCOP 实现的是深度卷积操作, 然后通过 1×1 卷积实现逐点卷积操作.

3 动态覆盖卷积

在本节, 我们主要介绍 DC-Conv 的生成(如图 3 所示), 并分析了该卷积的特性, 最后, 基于 PointNet++^[14] 架构提出 DC-CNN.

首先使局部邻域中的邻居点和采样点在位置空间中产生信息交互, 然后利用这种几何先验知识生成多组权重 w_{ij} . 通过 w_{ij} 对邻域点的边缘特征 $f_{ij}-f_i$ 进行线性变换, 可以得到邻域点的多组特征. 同理, 从邻域点的几何先验中学习注意力系数, 将邻域点的多组特征进行动态组装. 最后, 通过对称函数得到局部邻域的特征 f_{p_j} . 上述这种操作称为动态空间覆盖算子. 但是这种操作在减少点云分辨率的情况下, 通道数没有发生改变, 限制了卷积的表达能. 因此, 我们通过 1×1 卷积实现逐点卷积, 从而得到局部特征的抽象表达.

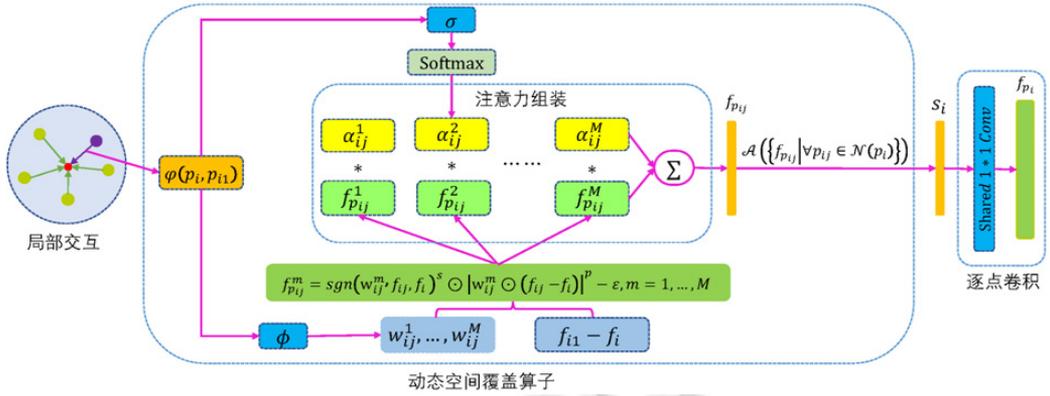


图 3 DC-Conv 的结构

3.1 权重生成

为了能使 DC-Conv 具有更强的感知局部几何结构的能力, 我们借鉴 PointConv^[30]和 RS-CNN^[32]的方式, 从采样点和邻域点的位置信息中, 自适应生成卷积权重. 对于每个采样点, 邻域点的几何先验可以编码为

$$\varphi(p_i, p_{ij}) = p_i \oplus p_{ij} \oplus (p_{ij} - p_i) \oplus \|p_{ij} - p_i\| \quad (4)$$

其中, \oplus 表示拼接操作, $\|\cdot\|$ 表示计算 3D 欧式距离.

因此, 局部邻域点的卷积权重可以表述为

$$w_{ij} = \phi(\varphi(p_i, p_{ij})) \quad (5)$$

其中, ϕ 是一个通过 MLP 实现的非线性函数. 公式(5)通过数据驱动的方式将局部邻域的几何先验知识编码在卷积权重中, 增强了适应复杂的空间几何变化的能力.

3.2 空间覆盖算子

从上述分析可知: 当仅仅使用简单的共享的 MLP 转换邻域点的特征, 限制了模型适应空间变化的能力. 因此在本小节中, 我们提出一种表达能力更强的 SCOP, 它可以利用点云的坐标和特征信息的相关性, 在局部邻域中建立复杂的几何形状学习局部结构中隐含的形状信息.

在该算子中, 我们采用局部邻域点相对于采样点的边缘特征作为输入, 即:

$$\mathcal{A}(f_i, f_{ij}) = f_{ij} - f_i \quad (6)$$

SCOP 可表示为

$$s_i = \mathcal{A}(\{f_{p_{ij}} \mid \forall p_{ij} \in \mathcal{N}(p_i)\}) = \mathcal{A}(\{\text{sgn}(w_{ij}, f_{ij}, f_i)^s \odot |w_{ij} \odot (f_{ij} - f_i)|^p - \varepsilon \mid \forall p_{ij} \in \mathcal{N}(p_i)\}) \quad (7)$$

其中, $s \in \{0, 1\}$, 符号函数 $\text{sgn}(w_{ij}, f_{ij}, f_i) = \frac{w_{ij} \odot (f_{ij} - f_i)}{|w_{ij} \odot (f_{ij} - f_i)|}$ 用于决定特征向量的元素的正负性, \odot 表示逐元素相乘, $|\cdot|$ 表示逐元素取绝对值, p 是 SCOP 的可学习的阶数, ε 为阈值.

当 \mathcal{A} 为求和操作时, 图 4 可视化了 SCOP 的几何形状. 其中, s 控制 SCOP 的封闭性, p 控制 SCOP 的胖瘦, w_{ij} 控制 SCOP 是否各向异性, f_i 控制 SCOP 在空间中的位置. 由图 4 可知:

- 当 $s=1, p=1, f_i$ 均为 0 时, 则 SCOP 就是 $s_i = \mathcal{A}(w_{ij} \odot f_{ij} - \varepsilon)$, 我们将该函数称为仿射基函数(ABF)^[18], 其本质是 高维空间中的超平面(如图 4(g)所示);
- 当 $s=0, p=2, w_{ij}$ 均为 1 时, 则 SCOP 就是径向基函数(radial basis function, RBF)^[41]: $s_i = \mathcal{A}(w_{ij} \odot (f_{ij} - f_i)^2 - \varepsilon)$, 其本质是各向同性的封闭超球体(如图 4(d)所示).

s 决定输出特征的方向性和 SCOP 的封闭性: 当 s 为 0 时, 改变 w_{ij} 和 p, f_{ij} 的轨迹就封闭超曲面; 当 s 为 1 时, f_{ij} 的轨迹是不封闭超曲面. w_{ij} 决定了神经元的体态, 而且 w_{ij} 由局部邻域的几何先验自适应学习得到. 因此, SCOP 对复杂的几何空间具有更加灵活的拟合能力.

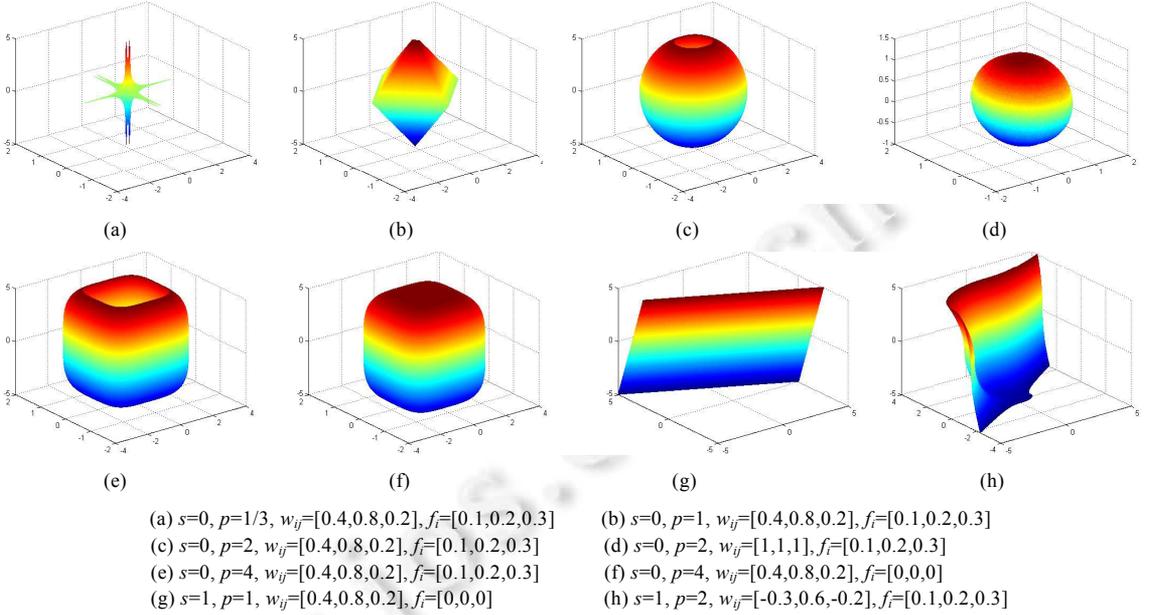


图 4 SCOP 的不同形状可视化

3.3 逐点卷积

SCOP 通过逐点相乘对局部邻域的点特征进行线性变换,然后在通道方向对局部特征进行聚合,这种操作可以视为深度卷积.但是这种操作没有增加通道数量,从而限制了模型的表达能力.因此,我们提出逐点卷积实现不同通道间的交互,从而得到局部几何结构的抽象表达.该卷积可以表示为

$$f_{p_i} = \gamma(s_i) \quad (8)$$

其中, f_{p_i} 是采样点 p_i 的更新后的特征, γ 是共享的 1D 卷积 $Conv:R^c \rightarrow R^{c'}$. 在这里, c 和 c' 分别为输入和输出的特征通道数.

因此,我们将 SCOP 和逐点卷积的组合称为空间覆盖卷积(space cover convolution, SC-Conv).

3.4 动态覆盖卷积的注意力组装

我们发现:在局部邻域使用单一的 SC-Conv,将会限制模型的表达能力.虽然可以通过提高网络的宽度、深度或者更复杂的设计解决该问题,但是这会增加模型的复杂度.因此,我们通过在局部邻域使用多个卷积权重解决该问题.

我们定义 $\phi = \{\phi_m | m=1, \dots, M\}$, 其中, ϕ_m 表示通过 MLP 实现的非线性函数, M 表示由公式(5)生成的卷积权重的个数, 即 $w_{ij} = \{w_{ij}^m | m=1, \dots, M\}$, $w_{ij}^m = \phi_m(\varphi(p_i, p_{ij}))$. 由公式(7)可知 $f_{p_{ij}} = \{f_{p_{ij}}^m | m=1, \dots, M\}$, 其中,

$$f_{p_{ij}}^m = \text{sgn}(w_{ij}^m, f_{ij}, f_i) \odot |w_{ij} \odot (f_{ij} - f_i)|^p - \varepsilon \quad (9)$$

由上述分析可知,每个局部邻域点将会有 M 个特征输出,我们通过注意力系数将这些特征进行动态组装.考虑到局部点特征与采样点和邻域点的特定关系,我们参考卷积权重 w_{ij} 的生成过程,从点位置中自适应生成注意力系数.该系数矩阵可表示为

$$\alpha_{ij} = \text{softmax}(\sigma(\varphi(p_i, p_{ij}))) \quad (10)$$

其中, $\alpha_{ij} = \{\alpha_{ij}^m | 1, \dots, M\}$, σ 是 MLP 实现的非线性函数, softmax 函数将注意力系数限制在(0,1)内.

于是,我们通过注意力系数将邻域点的特征动态组装起来,即:

$$f_{p_{ij}} = \sum_{m=1}^M \alpha_{ij}^m f_{p_{ij}}^m \quad (11)$$

因此, 公式(11)和公式(7)、公式(8)结合, 就构成了基于注意力系数的 DC-Conv. 其计算过程如下所示.

算法 1. DC-Conv.

输入: 采样点 p_i , 采样点特征 f_i , 邻居点 p_{ij} , 邻居点特征 f_{ij} ;

输出: 局部邻域特征 $f_{p_{ij}}$.

1. 计算邻居点与采样点的三维欧氏距离 $\|p_i, p_{ij}\|$
2. 将邻居点坐标转化为相对于采样点的相对坐标 $p'_{ij} = p_{ij} - p_i$
3. 计算每个邻居点的几何先验知识 $\varphi(p_i, p_{ij})$
4. 使用 ϕ_m 将 $\varphi(p_i, p_{ij})$ 映射到高维空间, 得到卷积权重 $w'_{ij} = \phi_m(\varphi(p_i, p_{ij}))$
5. 计算邻居点特征相对于采样点特征的边缘特征 $f'_{ij} = f_{ij} - f_i$
6. 通过公式(9)得到变换后的邻域点特征 $f_{p_{ij}}^m$
7. 利用 σ 和 *Softmax* 生成注意力系数 $\alpha_{ij} = \text{softmax}(\sigma(\varphi(p_i, p_{ij})))$
8. 然后得到加权后的邻域点特征 $f_{p_{ij}} = \sum_{m=1}^M \alpha_{ij}^m f_{p_{ij}}^m$
9. 计算 SCOP 的输出 $s_i = A(\{f_{p_{ij}} \mid \forall p_{ij} \in \mathcal{N}(p_i)\})$
10. 最终, 得到 DC-Conv 在局部邻域的输出特征 $f_{p_i} = \gamma(s_i)$

4 动态覆盖卷积神经网络

为了更好地与其他局部特征聚合方法进行比较, 在不需要任何附加操作的情况下, 我们沿用了 PointNet++^[14] 提取特征的网络结构, 并采用统一的采样策略. 利用 DC-Conv 作为局部特征提取器, 通过叠加 SA-Layer 构成了分层学习的 DC-CNN, 图 5 显示了用于点云分类的网络结构.

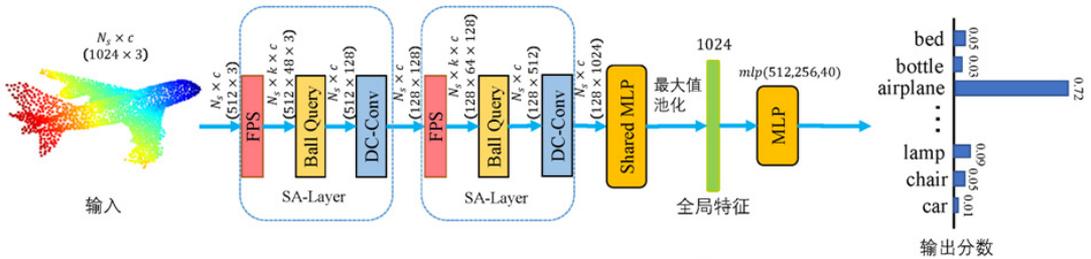


图 5 用于点云分类的 DC-CNN 的架构

从图 5 可以看出: 动态覆盖卷积神经网络的输入是由 1024 个点组成的点云, 通过两层的集合抽象层和共享的全连接层得到描述物体的全局特征, 最后通过全连接层得到物体的分类分数. 其中, SA-Layer 由采样层 (FPS)、分组层 (球形查询) 和局部特征提取层 (DC-Conv) 组成. 该网络最大的特性是, 在每一层的 SCOP 的阶数都是可学习的. 因此, DC-CNN 可以学习到不同阶次的特征. 另一个特性是在每个卷积层, 动态特性使得 DC-CNN 充分提取局部几何信息, 从而形成更准确的形状感知表示.

对于模型优化, 在本文中, 我们使用分类任务中最常使用的交叉熵损失函数作为模型的目标函数. 此外, 为了加速模型的训练, 我们期望空间覆盖算子的指数 p 不要太大. 因此, 通过反向传播算法更新参数时, 我们通过截断函数限制了指数 p 的范围, 具体公式如下所示:

$$\hat{p} = \begin{cases} p, & c \leq p \leq d \\ c, & p \leq c \\ d, & d \leq p \end{cases} \quad (12)$$

其中, c 和 d 为截断区间的两个端点, \hat{p} 为截断后的 p 值.

5 实验分析

在本节中,我们首先进行了点云分类实验,并与一些目前先进的算法进行了比较;然后,我们研究了 SCOP 的参数影响;最后,我们对 DC-CNN 的池化函数选择、鲁棒性和模型复杂度等进行了分析。

5.1 形状分类

5.1.1 数据集

- 合成数据集
 - ModelNet40. ModelNet40 数据集^[42]是 3D 点云分类的基准数据集,该数据集有 40 个类别的物体,共 12 311 个 CAD 模型. 整个数据集被分割为训练集和测试集,其中,9 843 个模型用于训练,2 468 个模型用于测试;
 - ModelNet10. 它是 ModelNet40 数据集的子集,该数据集共包含 10 个类别的 4 900 个 3DCAD 模型,其中,3 991 个模型作为训练集,909 个模型作为测试集. 我们在该实验中也保存了上述两个数据集的划分;
- 真实世界数据集
 - ScanObjectNN. ScanObjectNN^[56]是最近发布的用于 3D 分类的点云数据集,它从现实世界采集的物体包括了背景并考虑了遮挡,比 ModelNet40 更现实和更具挑战性. 该数据集由 2 902 个点云对象组成,分为 15 个对象类别,我们在它最难扰动的变体(PB_T50_RS 变体)上进行了实验.

5.1.2 实施

我们对 ModelNet40 数据集和 ModelNet10 数据集的每个 CAD 模型表面进行均匀采样 1 024 个点作为网络的输入,这些点只保留了 (x,y,z) 坐标信息,并被归一化到单位球内,并在全连接层使用丢弃率为 0.5 的 dropout 层. 在训练阶段,我们遵从 RS-CNN^[32]的数据增强设置,将点云在 $[0.66,1.5]$ 范围内进行随机缩放,在坐标的 3 个方向上进行 $[-0.2,0.2]$ 范围内的随机平移. 在参数优化过程中,我们使用 Adam 优化函数,并且初始学习率设置为 0.002,每经过 25 个迭代次数,学习率衰减为原来的 0.7 倍. 模型总共被训练 200 个迭代次数. 在测试阶段,我们也对待测试点云在 $[0.66,1.5]$ 范围内随机缩放,并使用投票机制进行了预测,投票结果的平均值作为预测的最终结果. 训练阶段和测试阶段的批次大小均设置为 16.

同样,我们对 ScanObjectNN 数据集的每个物体均匀采样 1 024 个点作为网络的输入. 在该实验中,我们使用的是多尺度 DC-CNN,即在每一层中,通过球形邻域查询确定每个尺度的邻居点,然后进行特征提取并将每个尺度的特征进行拼接. 具体地,第 1 个 SA-Layer 的球形邻域半径分别为 0.1, 0.2 和 0.4,对应的邻居点数分别为 16, 32, 128;第 2 个 SA-Layer 的邻域半径分别为 0.1, 0.2 和 0.4,对应的邻居点数分别为 16, 32 和 128. 训练和测试阶段的批次分别为 32 和 16. 在该实验中,我们对输入没有使用任何的数据增强技术,其他设置与 ModelNet40 上的相同.

5.1.3 结果

表 1 报告了我们的 DC-CNN 和 SC-CNN 与最近的一些工作的比较结果(✓表示在单尺度的情况下使用球形查询建立局部邻域). 从表中可以看出:即使仅仅使用位置信息的情况下,我们的方法与其他模型在 ModelNet40 数据集和 ModelNet10 数据集上取得了相当甚至更好的结果. 与使用 ABF 的 RS-CNN^[32]相比,在单尺度的情况下,SC-CNN 比 RS-CNN 的准确率提高了 0.8%. 这说明 SCOP 能够覆盖局部几何结构,从而实现更复杂的形状感知表达. 特别是在单尺度的情况下,DC-CNN 在 ModelNet40 数据集和 ModelNet10 数据集上的 10 次投票的平均准确率分别达到了 93.6%和 96.2%,特别是在 ModelNet10 数据集上的准确率比 SC-CNN(只使用一个空间覆盖算子)提高了 0.4%. 这表明 DC-Conv 在很大程度上增强了局部特征在通道上的多样性.

表 2 报告了 DC-CNN 和 SC-CNN 在现实 ScanObjectNN 基准测试上的分类准确度. 从实验结果看:我们的 DC-CNN 在分类准确度上有显著的提高,超过了所有方法. 例如,我们的准确率比 PRANet 高出 1.1%. 即使与多视图投影方法 SimpleView 相比,我们的 DC-CNN 仍然表现得更好(高出 1.6%). 同时,DC-CNN 比 SC-

CNN 高出 0.7%, 说明我们提出的动态覆盖卷积比空间覆盖卷积表达能力更强. 此外, ScanObjectNN 数据集的物体是从真实世界中获取的, 在具有遮挡、背景和几何扭曲等干扰因素的情况下, DC-CNN 仍能达到 82.1% 的准确率, 说明我们的方法具有良好的稳健性和一定的泛化能力.

表 1 DC-CNN 的点云分类结果

方法	输入	点的个数	单尺度	ModelNet10	ModelNet40
PointwiseCNN ^[43]	xyz	1k	✓	-	86.1
PointNet ^[13]	xyz	1k	-	91.2	89.2
Kd-Net t(depth=10) ^[22]	xyz	1k	-	93.3	90.6
PointNet++ ^[14]	xyz	1k	✓	92.5	90.7
SO-Net ^[44]	xyz	2k	-	94.1	90.9
SRN-PointNet++ ^[25]	xyz	1k	✓	-	91.5
Kd-Net (depth=15) ^[22]	xyz	32k	-	94.0	91.8
DGCNN ^[15]	xyz	1k	-	93.3	92.2
PCNN ^[45]	xyz	1k	-	94.9	92.3
FP-conv ^[46]	xyz, nor	1k	-	-	92.5
RS-CNN ^[32]	xyz	1k	✓	-	92.7
KPCConv (rigid) ^[31]	xyz	1k	-	-	92.9
DTNet ^[47]	xyz	1k	-	-	92.9
InterpCNN ^[48]	xyz	1k	-	-	93.0
PosPool ^[49]	xyz	5k	✓	-	93.2
PointASNL ^[28]	xyz, nor	1k	-	95.9	93.2
PCT ^[23]	xyz	1k	-	-	93.2
DensePoint ^[50]	xyz	1k	-	-	93.2
GS-Net ^[51]	xyz	2k	✓	-	93.3
PointCutmix-S (PointNet++) ^[52]	xyz	1k	-	96.3	93.4
RSMix (DGCNN) ^[53]	xyz	1k	-	95.9	93.5
SC-CNN (our)	xyz	1k	✓	95.8	93.2
DC-CNN (our)	xyz	1k	✓	96.2	93.6

表 2 DC-CNN 在 ScanObjectNN 数据集上的分类结果

方法	准确率(%)
3DmFV ^[54]	63.0
PointNet ^[13]	68.2
SpiderCNN ^[55]	73.7
PointNet++ ^[14]	77.9
DGCNN ^[15]	78.1
PointCNN ^[29]	78.5
BGA-DGCNN ^[56]	79.7
BGA-PN++ ^[56]	80.2
DRNet ^[57]	80.3
GBNet ^[58]	80.5
SimpleView ^[59]	80.5
PRANet ^[60]	81.0
SC-CNN (our)	81.4
DC-CNN (our)	82.1

5.2 动态覆盖卷积有效性分析

为了更好地理解 DC-Conv, 我们在 ModelNet40 数据集上进行了 DC-Conv 的有效性分析: 首先, 我们对 SCOP 的阶数 P 进行了研究; 其次, 探究了 SCOP 的个数对点云分类效果的影响; 然后, 讨论了池化操作的选择和 DC-CNN 的鲁棒性; 最后, 对模型复杂度进行了比较. 特别地, 除非特殊声明, 其他细节均与上述其他设置相同.

5.2.1 阶数 P 研究

根据上述分析, 阶数 s 和 p 影响 SCOP 的几何形状的表达. 在这里, 我们探索了当 s 分别为 0 或 1、 p 可学习时, 网络性能的表达.

结果见表 3 (“-”表示结果不可知): 当 s 为 1 时, 对 p 值的范围不进行限制, 即 $p \in (-\infty, +\infty)$, 模型在训练时出现崩溃的情况, 因而在这里我们使用了截断函数对 p 的范围进行了限制. 当 s 为 1 时, DC-CNN 在 ModelNet40 数据集上的准确率高于当 s 为 0 时的准确率. 特别是当 $p \in [1, 2]$ 时, 分类准确率达到 93.6%. 当 p

的范围增大时, 准确率有稍微的下降. 这主要是因为 p 作为 SCOP 的阶数, 当该值增大时, 在学习过程中增大了模型训练的难度.

表 3 SCOP 的参数对网络性能的影响

参数设置	准确率(%)
$s=0, p \in [0.5, 2]$	92.8
$s=0, p \in [1, 2]$	93.2
$s=1, p \in [0.5, 2]$	93.0
$s=1, p \in [1, 2]$	93.6
$s=1, p \in [1, 3]$	93.4
$s=1, p \in [1, 5]$	92.9
$s=1, p \in (-\infty, +\infty)$	-

5.2.2 局部特征提取器的选择

为了比较局部特征提取器的性能, 本组实验对 MLP、常规卷积、深度可分离卷积(空间覆盖卷积)和动态覆盖卷积分别作为局部特征提取器时, 对最终结果的影响.

实验结果见表 4, 可以观察到, 直接使用 MLP 聚合局部点云特征的效果最差. 这是因为该操作是对每个邻域点分别操作, 通过最大值池化只保留了每个通道中最重要的信息, 从而导致细节信息的丢失. 而使用深度可分离卷积优于常规卷积的结果, 从实验层面验证了深度可分离卷积能够通过相对较少的参数实现优越的性能. 最终, 使用动态覆盖卷积的效果最好. 这表明动态覆盖卷积考虑到了局部点云的分布空间, 充分捕捉到更加全面的局部信息.

表 4 不同局部特征操作方式在 Modelnet40 上的分类结果

局部特征运算	准确率(%)
MLP	90.7
常规卷积	92.5
深度可分离卷积	93.2
动态覆盖卷积	93.6

5.2.3 卷积权重的个数

虽然卷积权重个数的增加有利于局部特征的表达, 但是当增加到一定程度时, 也会带来模型训练困难的问题. 因此在本小节, 我们探究了卷积权重个数对网络性能的影响.

从表 5 可知: 当增加少量的参数量的情况下, 权重个数为 4 的时候, 模型的性能达到了 93.6%. 但是随着卷积权重个数的增加, 模型复杂度也在增加, 由于优化的问题, 模型性能并不能更进一步得到提升.

表 5 拥有不同的 SCOP 个数的 DC-Conv 在 Modelnet40 上分类结果

个数	准确率(%)
1	93.2
2	93.4
4	93.6
8	93.1
16	92.8

5.2.4 池化方法的选择

我们也探究了 3 种对称函数(平均池化、最大值池化和求和池化)对 DC-Conv 的影响. 实验结果见表 6 所示. 最大池化的性能最好, 其次是求和池化, 平均池化的效果最差. 该结论与 PointNet^[13]是一致的.

表 6 动态覆盖卷积神经网络基于不同池化策略在 ModelNet40 数据集上的分类结果

平均池化	最大值池化	求和池化	准确率(%)
✓	-	-	93.1
-	✓	-	93.6
-	-	✓	92.9

5.2.5 鲁棒性分析

DC-Conv 通过在局部特征空间中建立空间几何体和对称函数聚合局部邻域特征, 具有一定的鲁棒性. 因此, 我们在无数据增强的训练中得到测试模型, 分别测试 DC-Conv 对置换、旋转和平移的鲁棒性.

由表 7 可知, 上述方法均具有置换不变性. 但是 PointNet 和 PointNet++ 的稳定性较差, RS-CNN 具有一定的鲁棒性, 但是 DC-Conv 在面对平移和旋转时具有更优的鲁棒性.

表 7 DC-Conv 在 ModelNet40 数据集上的鲁棒性结果(%)

方法	-	置换	+0.2	-0.2	90°	180°
PointNet ^[13]	88.7	88.7	70.8	70.6	42.5	38.6
PointNet++ ^[14]	90.3	90.3	90.1	89.0	55.8	47.6
RS-CNN ^[32]	92.3	92.3	92.0	91.5	57.3	74.4
SC-CNN(our)	93.0	93.0	92.7	92.2	59.4	77.9
DC-CNN(our)	93.3	93.3	93.2	92.6	60.4	78.8

5.2.6 模型复杂度

复杂度是对模型性能检测的另一个有价值且重要的度量指标, 包括时间复杂度和空间复杂度. 因此, 表 8 展示了我们的方法与其他先进方法的模型复杂度比较结果. 可以看出, SC-CNN 在参数量和计算量比其他方法均得到大幅下降. 特别是, 在 SC-CNN 的基础上设计的 DC-CNN 在增加少量的计算量的情况下, 准确率提高了 0.4%. 虽然比使用 ABF 的 RS-CNN 在计算量上有所增加, 但是在参数量上减少了 7.1%. 这表明 DC-CNN 能够在较少的参数和时间下, 收敛到先进的精度, 因此有助于未来在实时任务中的应用.

表 8 动态覆盖卷积神经网络在 ModelNet40 数据集上的复杂度

方法	参数量	计算量/样本
PointNet ^[13]	3.50M	440M
PointNet++ ^[14]	1.48M	1684M
PCNN ^[45]	8.20M	294M
RS-CNN ^[32]	1.41M	295M
SC-CNN (our)	1.29M	276M
DC-CNN (our)	1.31M	390M

6 总结

本文提出一种动态覆盖卷积, 这是一种在点云中动态学习局部几何的新方法. 动态覆盖卷积的核心是, 在局部邻域的特征空间中使用多个空间覆盖算子捕捉丰富的局部特征. 空间覆盖算子是在局部特征邻域构建各向异性的几何体学习局部特征, 该算子的权重从局部邻域点中自适应学习得到. 通过从点位置中自适应学习注意力系数, 将多个空间覆盖算子组装在一起. 基于动态覆盖卷积设计的动态覆盖卷积神经网络, 在点云分类中实现了显著的效果. 大量的实验也证明了动态覆盖卷积的有效性.

此外, 在本文中, 动态覆盖卷积的阶数 P 的学习范围是受限的, 且相比于传统的神经元, 参数量上只增加了阶数 P , 但是增加了一些计算量, 这就会影响动态覆盖卷积在大规模点云中的实际应用. 因此, 如何解决阶数 P 的学习问题、简化动态覆盖卷积的计算方式, 使模型更加鲁棒和轻量化, 是我们下一步研究的重点. 同时, 作为一种新型的点云分析方法, 我们将在点云分割、检测等复杂任务以及将主流图像卷积网络体系结构引入点云数据中, 进一步探索动态覆盖卷积的实用性和优越性.

References:

- [1] Deng W, Huang K, Chen X, *et al.* Semantic RGB-D SLAM for rescue robot navigation. *IEEE Access*, 2020, 8: 221320–221329. [doi: 10.1109/ACCESS.2020.3031867]
- [2] Gul F, Rahiman W, NazliAlhady SS. A comprehensive study for robot navigation techniques. *Cogent Engineering*, 2019, 6(1): 1632046. [doi: 10.1080/23311916.2019.1632046]
- [3] Yurtsever E, Lambert J, Carballo A, *et al.* A survey of autonomous driving: Common practices and emerging technologies. *IEEE Access*, 2020, 8: 58443–58469. [doi: 10.1109/ACCESS.2020.2983149]

- [4] Grigorescu S, Trasnea B, Cocias T, *et al.* A survey of deep learning techniques for autonomous driving. *Journal of Field Robotics*, 2020, 37(3): 362–386. [doi: 10.1002/rob.21918]
- [5] Zhang J, Lin X. Advances in fusion of optical imagery and LiDAR point cloud applied to photogrammetry and remote sensing. *Int'l Journal of Image and Data Fusion*, 2017, 8(1): 1–31. [doi: 10.1080/19479832.2016.1160960]
- [6] Li Y, Ma J, Zhang Y. Image retrieval from remote sensing big data: A survey. *Information Fusion*, 2021, 67: 94–115. [doi: 10.1016/j.inffus.2020.10.008]
- [7] Zhu XX, Shahzad M. Facade reconstruction using multiview spaceborne TomoSAR point clouds. *IEEE Trans. on Geoscience and Remote Sensing*, 2013, 52(6): 3541–3552. [doi: 10.1109/TGRS.2013.2273619]
- [8] Li L, Zhu S, Fu H, *et al.* End-to-end learning local multi-view descriptors for 3D point clouds. In: *Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition*. 2020. 1919–1928. [doi: 10.1109/cvpr42600.2020.00199]
- [9] Wei X, Yu R, Sun J. View-GCN: View-based graph convolutional network for 3D shape analysis. In: *Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition*. 2020. 1850–1859. [doi: 10.1109/cvpr42600.2020.00192]
- [10] Yang Z, Wang L. Learning relationships for multi-view 3D object recognition. In: *Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition*. 2019. 7505–7514. [doi: 10.1109/ICCV.2019.00760]
- [11] Maturana D, Scherer S. VoxNet: A 3D convolutional neural network for real-time object recognition. In: *Proc. of the IEEE/RSJ Int'l Conf. on Intelligent Robots and Systems*. 2015. 922–928. [doi: 10.1109/IROS.2015.7353481]
- [12] Riegler G, Ulusoy AO, Geiger A. Octnet: Learning deep 3D representations at high resolutions. In: *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*. 2017. 3577–3586. [doi: 10.1109/CVPR.2017.701]
- [13] Qi CR, Su H, Mo K, *et al.* Pointnet: Deep learning on point sets for 3D classification and segmentation. In: *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*. 2017. 652–660. [doi: 10.1109/CVPR.2017.16]
- [14] Qi CR, Yi L, Su H, *et al.* PointNet++: Deep hierarchical feature learning on point sets in a metric space. *Advances in Neural Information Processing Systems*, 2017, 30: 5099–5108.
- [15] Wang Y, Sun Y, Liu Z, *et al.* Dynamic graph CNN for learning on point clouds. *ACM Trans. on Graphics*, 2019, 38(5): 1–12. [doi: 10.1145/3326362]
- [16] Wang C, Samari B, Siddiqi K. Local spectral graph convolution for point set feature learning. In: *Proc. of the European Conf. on Computer Vision*. 2018. 52–66. [doi: 10.1007/978-3-030-01225-0_4]
- [17] Komarichev A, Zhong Z, Hua J. A-CNN: Annularly convolutional neural networks on point clouds. In: *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*. 2019. 7421–7430. [doi: 10.1109/CVPR.2019.00760]
- [18] Rosenblatt F. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 1958, 65(6): 386. [doi: 10.1037/H0042519]
- [19] Ning X, Li WJ, Li HG, *et al.* Uncorrelated locality preserving discriminant analysis based on bionics. *Journal of Computer Research and Development*, 2016, 53(11): 2623–2629 (in Chinese with English abstract). [doi: 10.7544/issn1000-1239.2016.20150630]
- [20] Ning X, Li WJ, Tang B, *et al.* BULDP: Biomimetic uncorrelated locality discriminant projection for feature extraction in face recognition. *IEEE Trans. on Image Processing*, 2018, 27(5): 2575–2586. [doi: 10.1109/TIP.2018.2806229]
- [21] Su H, Maji S, Kalogerakis E, *et al.* Multi-view convolutional neural networks for 3D shape recognition. In: *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*. 2015. 945–953. [doi: 10.1109/ICCV.2015.114]
- [22] Klovov R, Lempitsky V. Escape from cells: Deep kd-networks for the recognition of 3D point cloud models. In: *Proc. of the IEEE Int'l Conf. on Computer Vision*. 2017. 863–872. [doi: 10.1109/ICCV.2017.99]
- [23] Guo MH, Cai JX, Liu ZN, *et al.* PCT: Point cloud transformer. *Computational Visual Media*, 2021, 7(2): 187–199. [doi: 10.1007/s41095-021-0229-5]
- [24] Xiang T, Zhang C, Song Y, *et al.* Walk in the cloud: Learning curves for point clouds shape analysis. In: *Proc. of the IEEE/ CVF Int'l Conf. on Computer Vision*. 2021. 915–924.
- [25] Duan Y, Zheng Y, Lu J, *et al.* Structural relational reasoning of point clouds. In: *Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition*. 2019. 949–958. [doi: 10.1109/CVPR.2019.00104]
- [26] Hu Q, Yang B, Xie L, *et al.* Randla-net: Efficient semantic segmentation of large-scale point clouds. In: *Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition*. 2020. 11108–11117. [doi: 10.1109/CVPR42600.2020.01112]
- [27] Zhao H, Jiang L, Fu CW, *et al.* PointWeb: Enhancing local neighborhood features for point cloud processing. In: *Proc. of the IEEE/ CVF Conf. on Computer Vision and Pattern Recognition*. 2019. 5565–5573. [doi: 10.1109/CVPR.2019.00571]

- [28] Yan X, Zheng C, Li Z, *et al.* Pointasnl: Robust point clouds processing using nonlocal neural networks with adaptive sampling. In: Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition. 2020. 5589–5598. [doi: 10.1109/cvpr42600.2020.00563]
- [29] Li Y, Bu R, Sun M, *et al.* Pointcnn: Convolution on x-transformed points. *Advances in Neural Information Processing Systems*, 2018, 31: 820–830.
- [30] Wu W, Qi Z, Fuxin L. Pointconv: Deep convolutional networks on 3D point clouds. In: Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition. 2019. 9621–9630. [doi: 10.1109/CVPR.2019.00985]
- [31] Thomas H, Qi CR, Deschaud JE, *et al.* KPConv: Flexible and deformable convolution for point clouds. In: Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition. 2019. 6411–6420. [doi: 10.1109/ICCV.2019.00651]
- [32] Liu Y, Fan B, Xiang S, *et al.* Relation-shape convolutional neural network for point cloud analysis. In: Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition. 2019. 8895–8904. [doi: 10.1109/CVPR.2019.00910]
- [33] Qi X, Liao R, Jia J, *et al.* 3D graph neural networks for RGBD semantic segmentation. In Proc. of the IEEE Int'l Conf. on Computer Vision. 2017. 5199–5208. [doi: 10.1109/ICCV.2017.556]
- [34] Simonovsky M, Komodakis N. Dynamic edge-conditioned filters in convolutional neural networks on graphs. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. 2017. 3693–3702. [doi: 10.1109/CVPR.2017.11]
- [35] Chen Y, Dai X, Liu M, *et al.* Dynamic convolution: Attention over convolution kernels. In: Proc. of the IEEE/ CVF Conf. on Computer Vision and Pattern Recognition. 2020. 11030–11039. [doi: 10.1109/CVPR42600.2020.01104]
- [36] Yang B, Bender G, Le QV, *et al.* CondConv: Conditionally parameterized convolutions for efficient inference. *Advances in Neural Information Processing Systems*, 2019, 32: 1307–1318.
- [37] Chen J, Wang X, Guo Z, *et al.* Dynamic region-aware convolution. In: Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition. 2021. 8064–8073. [doi: 10.1109/CVPR46437.2021.00797]
- [38] Xu M, Ding R, Zhao H, *et al.* PACConv: Position adaptive convolution with dynamic kernel assembling on point clouds. In: Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition. 2021. 3173–3182. [doi: 10.1109/CVPR46437.2021.00319]
- [39] Zeiler MD, Fergus R. Visualizing and understanding convolutional networks. In: Proc. of the European Conf. on Computer Vision. Cham: Springer, 2014. 818–833. [doi: 10.1007/978-3-319-10590-1_53]
- [40] Kaiser L, Gomez AN, Chollet F. Depthwise separable convolutions for neural machine translation. In: Proc. of the Int'l Conf. on Learning Representations. 2018.
- [41] Moody J, Darken CJ. Fast learning in networks of locally-tuned processing units. *Neural Computation*, 1989, 1(2): 281–294. [doi: 10.1162/neco.1989.1.2.281]
- [42] Wu Z, Song S, Khosla A, *et al.* 3D shapenets: A deep representation for volumetric shapes. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. 2015. 1912–1920. [doi: 10.1109/CVPR.2015.7298801]
- [43] Hua BS, Tran MK, Yeung SK. Pointwise convolutional neural networks. In: Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition. 2018. 984–993. [doi: 10.1109/CVPR.2018.00109]
- [44] Li J, Chen BM, Lee GH. So-net: Self-organizing network for point cloud analysis. In: Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition. 2018. 9397–9406. [doi: 10.1109/CVPR.2018.00979]
- [45] Atzmon M, Maron H, Lipman Y. Point convolutional neural networks by extension operators. *ACM Trans. on Graphics*, 2018, 37: 1–12. [doi: 10.1145/3197517.3201301]
- [46] Lin Y, Yan Z, Huang H, *et al.* FPConv: Learning local flattening for point convolution. In: Proc. of the IEEE/ CVF Conf. on Computer Vision and Pattern Recognition. 2020. 4293–4302. [doi: 10.1109/cvpr42600.2020.00435]
- [47] Han XF, Jin YF, Cheng HX, *et al.* Dual transformer for point cloud analysis. arXiv:2104.13044, 2021.
- [48] Mao J, Wang X, Li H. Interpolated convolutional networks for 3D point cloud understanding. In: Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition. 2019. 1578–1587. [doi: 10.1109/ICCV.2019.00166]
- [49] Liu Z, Hu H, Cao Y, *et al.* A closer look at local aggregation operators in point cloud analysis. In: Proc. of the European Conf. on Computer Vision. Cham: Springer, 2020. 326–342. [doi: 10.1007/978-3-030-58592-1_20]
- [50] Liu Y, Fan B, Meng G, *et al.* Densepoint: Learning densely contextual representation for efficient point cloud processing. In: Proc. of the IEEE/CVF Int'l Conf. on Computer Vision. 2019. 5239–5248. [doi: 10.1109/ICCV.2019.00534]
- [51] Xu M, Zhou Z, Qiao Y. Geometry sharing network for 3D point cloud classification and segmentation. In: Proc. of the AAAI Conf. on Artificial Intelligence. 2020, 34(7): 12500–12507. [doi: 10.1609/AAAI.V34I07.6938]
- [52] Zhang J, Chen L, Ouyang B, *et al.* PointCutMix: Regularization strategy for point cloud classification. arXiv:2101.01461, 2021.

- [53] Lee D, Lee J, Lee J, *et al.* Regularization strategy for point cloud via rigidly mixed sample. In: Proc. of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition. 2021. 15900–15909. [doi: 10.1109/CVPR46437.2021.01564]
- [54] Ben-Shabat Y, Lindenbaum M, Fischer A. 3DMFV: Three-dimensional point cloud classification in real-time using convolutional neural networks. IEEE Robotics and Automation Letters, 2018, 3(4): 3145–3152. [doi: 10.1109/LRA.2018.2850061]
- [55] Xu Y, Fan T, Xu M, *et al.* Spidercnn: Deep learning on point sets with parameterized convolutional filters. In: Proc. of the European Conf. on Computer Vision. 2018. 87–102. [doi: 10.1007/978-3-030-01237-3_6]
- [56] Uy MA, Pham QH, Hua BS, *et al.* Revisiting point cloud classification: A new benchmark dataset and classification model on real-world data. In: Proc. of the IEEE/CVF Int'l Conf. on Computer Vision and Pattern Recognition. 2019. 1588–1597. [doi: 10.1109/ICCV.2019.00167]
- [57] Qiu S, Anwar S, Barnes N. Dense-resolution network for point cloud classification and segmentation. In: Proc. of the IEEE/CVF Winter Conf. on Applications of Computer Vision. 2021. 3813–3822. [doi: 10.1109/WACV48630.2021.00386]
- [58] Qiu S, Anwar S, Barnes N. Geometric back-projection network for point cloud classification. IEEE Trans. on Multimedia, 2021. [doi: 10.1109/TMM.2021.3074240]
- [59] Goyal A, Law H, Liu B, *et al.* Revisiting point cloud shape classification with a simple and effective baseline. In: Proc. of the Int'l Conf. on Machine Learning. PMLR, 2021.
- [60] Cheng S, Chen X, He X, *et al.* Pra-net: Point relation-aware network for 3D point cloud analysis. IEEE Trans. on Image Processing, 2021, 30: 4436–4448. [doi: 10.1109/TIP.2021.3072214]

附中文参考文献:

- [19] 宁欣, 李卫军, 李浩光, 等. 基于仿生学的不相关局部保持鉴别分析. 计算机研究与发展, 2016, 53(11): 2623–2629. [doi: 10.7544/issn1000-1239.2016.20150630]



王昌硕(1995—), 男, 博士生, 主要研究领域为三维点云分析, 行人再识别, 拓扑数据分析.



田生伟(1973—), 男, 博士, 博士生导师, 主要研究领域为人工智能, 软件工程.



王含(1981—), 男, 博士, 高级工程师, 主要研究领域为人工智能应用规范.



李卫军(1975—), 男, 博士, 研究员, 博士生导师, CCF 高级会员, 主要研究领域为神经网络与深度学习, 符号回归与深度建模, 视觉计算模型与算法, 3D 主动视觉芯片与系统.



宁欣(1989—), 男, 博士, 副研究员, CCF 高级会员, 主要研究领域为低功耗计算机视觉技术, 多视角学习及三维视觉中的应用, 形象认知计算理论与方法.