

全局结构引导的人造物体参数化基元检测

陈柱瀚, 黄惠*

(深圳大学可视计算研究中心 深圳 518060)
(huihuang@szu.edu.cn)

摘要: 点云是表达三维数据的常见形式, 点云数据提取出的几何基元能够帮助人们快速理解并处理场景信息, 也方便后续其他任务的开展. 为了更好地利用人造物体中普遍存在的全局结构关系, 增强基元检测过程中全局结构的正向引导, 提出参数化基元检测网络——RelationNet, 包括 2 个子模块. 首先, 为了更好地编码三维点与其所在基元的结构关系, 通过空间偏移预测模块预测三维点所在基元中心的偏移向量, 提升点对其所在基元的位置感知能力, 为后续分割任务提供更多的特征依据; 其次, 人造物体的基元与基元之间常常具有如平行、垂直、轴对齐等结构关系, 为了更好地利用这些关系实现对几何基元检测结果的改进, 还包含全局结构关系提取模块, 利用基元拟合后获得的参数判断各个基元之间的结构关系, 并通过设置相应的损失函数对提取到的结果进行引导监督. 在大型 ABC 数据集与基元监督拟合(SPFN), ParseNet 等主流算法进行对比的实验结果表明, RelationNet 在基元分割和基元分类任务上的 MIoU 分别达到 85.32%和 90.10%, 与当前先进方法相比有明显的效果提升.

关键词: 几何基元检测; 全局结构关系; 参数化

中图法分类号: TP391.41 **DOI:** 10.3724/SP.J.1089.2023.19754

Global Structure Guided Parametric Primitive Detection of Man-Made Objects

Chen Zhuhan and Huang Hui*

(Visual Computing Research Center, Shenzhen University, Shenzhen 518060)

Abstract: Point cloud is a common form of expressing 3D data. Extracting geometric primitives from point clouds can help people quickly understand and process scene information, and also facilitate other subsequent tasks. In order to better utilize the global structural relationship prevalent in man-made objects, this paper propose RelationNet to enhance its positive guidance in the process of primitive detection. RelationNet consists of two sub-modules. First, in order to better encode the structural relationship between 3D points and their primitives, RelationNet includes a spatial offset prediction module, which is used to predict the offset vector of 3D points to the center of the primitives where they are located. The location awareness of primitives also provides more valid features for subsequent segmentation tasks. Second, the primitives of man-made objects often have structural relationships, such as parallel, vertical, and axis alignment. In order to use these relationships to improve the detection results of geometric primitives, RelationNet also includes a global structural relationship extraction module, which uses the parameters obtained after primitive fitting to determine the structural relationship between each primitive, and sets the corresponding loss function to supervise the extracted results. This method is tested on the large scale ABC dataset and compared with mainstream methods, such as SPFN (supervised primitive fitting network) and ParseNet. The experimental results show that the accuracy of RelationNet in primitive segmentation

收稿日期: 2022-06-08; 修回日期: 2022-09-28. 基金项目: 国家自然科学基金(U21B2023, 62161146005); 广东省高等学校创新团队项目(2022KCXTD025); 深圳市科技创新项目(KQTD20210811090044003, RCJC20200714114435012, JCYJ20210324120213036); 深圳大学研究生教育改革项目(SZUGS2022JG01). 陈柱瀚(1996—), 男, 硕士研究生, 主要研究方向为计算机图形学; 黄惠(1977—), 女, 博士, 教授, 博士生导师, CCF 杰出会员, 论文通信作者, 主要研究方向为计算机图形学.

and primitive classification reaches MIOU of 85.32% and 90.10%, respectively, which significantly outperforms state-of-the-art methods.

Key words: geometric primitive detection; global structure relationship; parameterization

作为人工智能的重要组成部分,深度学习技术已经成功地用于解决各种二维视觉问题.与二维图像相比,三维数据能够更好地为计算机了解周围环境提供机会.三维数据在不同领域有许多应用,并且可以用不同的形式进行表示,包括深度图像、三角网格、点云等.在众多三维数据的表达方式中,点云在没有任何离散化的情况下保留了三维空间中的原始几何信息,是自动驾驶和机器人技术等许多场景理解相关应用的首选表示.

尽管利用三维传感器等技术可以快速获取三维点云数据,但因为缺乏物体的结构信息和连通情况,而且在采集的过程中数据经常受到采集噪声和物体部件遮挡的影响,点云原始数据通常是不完整的、嘈杂的并伴随有稀疏变化的,所以点云原始数据格式很难进行处理.为了在实际应用中充分利用原始捕获的三维数据,人们经常需要对单个物体进行抽象表示,提供类似人脑所获得的且能够理解的视觉总结.相当于只使用简单的几何基元来描述复杂对象,如球体、圆柱体、平面或圆锥等形状^[1].

通常三维传感器扫描获取的点云数据保留了场景的直接信息,但不便于人们进行直接的理解和分析,在点云所表达的低级视觉特征和人们所能处理的高级抽象语义之间存在着理解的鸿沟,通过检测点云中普遍存在的几何基元,人们能够建立跨越这道鸿沟的桥梁.

随着各种智能设备的发展,家用机器人和无人驾驶车辆开始频繁地出现在人们的日常生活中.这些设备大多通过丰富的视觉系统对周围环境形成感知,提取周围环境中的几何基元能够降低设备感知周围环境的难度.

随着数字孪生、室内环境建模技术的发展,人们常常需要依赖于点云数据进行算法开发和展示,不可避免地需要保存众多的点云数据.通常,表示 1 个普通人造物体需要上万个点,而为了表示室外的建筑物组成的大场景,需要数十万个点,这种保存方式会产生巨大的存储成本.若采用参数化基元这种形式简化表示,能够极大地压缩数据的量级,达到节省存储空间的目的^[2].

总的来说,参数化基元检测这种视觉抽象方

法不仅提取了点云数据的几何和拓扑结构,还提供了一种能够替代点云的、方便存储与展示的表达方式.此外,也可以利用真实物体的高阶几何特征抽象特点,快速地为重建出的物体编辑语义信息,这是后续结构化重建的中间结果.

目前,基于深度学习的基元检测方法在不同程度上捕获了全局特征和局部信息,但它们忽略了人造物体中普遍存在的全局结构关系.这种全局结构关系既包含点与单个基元之间的结构关系,也包含基元与基元之间的结构关系.为了编码点与单个基元之间的结构关系,本文首先使用空间偏移预测模块预测三维点与所在基元中心的偏移向量、编码点在基元中的相对位置;然后,提取基元之间的平行、垂直、轴对齐等结构关系,并利用这些结构关系约束检测过程,以有效地提升基元检测的准确率.

1 相关工作

1.1 传统几何基元检测

在深度学习迅速发展之前,研究人员处理三维点云时借鉴二维图像中提取直线等图形的思路,提出了一系列传统算法进行基元检测.传统的参数化基元检测算法有随机采样一致算法(random sample consensus, RANSAC)^[2],区域生长算法和 Hough 变换算法^[3]等.

Schnabel 等^[2]提出的 RANSAC 是一种常用的、用于迭代地估计给定数据集模型参数的算法,已运用在计算机视觉的多种场合;其基本原理是尝试多种可能拟合输入数据的随机模型并评估该模型是否足够好,以便能拟合绝大部分数据. RANSAC 有许多改进版本,但大多保留了随机抽样的环节.例如,Chum 等^[4]改变了随机采样的策略,在部分数据中按置信度排序选择样本,直至找到合适的模型.

区域生长算法首先通过人工指定或者随机的方式在点云当中选取种子点作为区域生长的起点;然后以种子点为中心选取邻域点,根据设定好的区域生长准则判断邻域点与种子点是否足够相似,如计算两者法向量的夹角是否小于阈值.重复这

一过程,直到没有新的邻域点满足生长准则的要求。种子点的选择既可通过随机的方式初始化^[5],也可以在有规律的地方初始化^[6]。

Hough 变换最早由 Hough^[7]提出,用于检测平面中的直线;经过拓展后,也可以用于进行其余二维图像形状^[3]和三维几何形状的检测。Hough 变换的基本思想是借助 2 个参数空间之间的变换,将一个空间中的某种特定形状映射到另一个参数空间中的一个点,通过在对应点上进行投票(或称为累计)形成全局最大值,将基元检测问题转为在高维空间的统计问题。

根据已有的先验信息,适当地引入正则化方法对初步输出的基元检测结果进行调整,能够提升算法整体的准确性。Li 等^[8]使用 RANSAC 从点云中提取基元,提取过程中逐个基元地判断结构关系,并利用这些结构关系构建全局的关系图;然后根据提取的结构关系约束基元参数,将点云到所在基元的距离纳入能量方程优化,寻找最小拟合误差,实现正则化。

1.2 基于深度学习的几何基元检测方法

除了传统的方法,随着深度学习技术的发展以及出现了合适的数据集,研究人员开始利用深度学习的方法对多种基元进行检测。

为了在三维点云重建任务中达到更好的效果,研究人员开始致力于检测点云中的边缘信息,并利用该信息辅助获取结果更真实、边缘更锐利的输出。在此过程中,边缘成为人们提取的一种新的基元类型。

尽管不以直接提取边缘信息为最终目的,但 Yu 等^[9]提出的 EC-Net 是最早利用深度学习技术提取点云中存在的边缘信息的工作。首先标注三角网格物体的边缘信息,并围绕边缘采样出若干块作为输入数据进行训练;然后在训练过程中,网络逐点预测点到最近边缘的距离,并且预测点到最近边缘的偏移向量,通过多个损失函数的共同作用实现点云对于边缘信息的感知;最后,原始输入点云根据偏移向量进行移动,距离边缘较近的点被视为边缘点构成边缘,再根据该边缘信息对点云重建结果进行辅助优化。

除了用曲线表示提取到的边缘外,研究人员还提出利用各种其他的基元类型描述三维点云数据的方法。Tulsiani 等^[10]最早提出使用长方体描述输入形状的方法,首先对输入物体进行体素化,再使用卷积神经网络(convolutional neural network, CNN)预测出一组长方体的形状参数和位置参数,

形成对原始物体的抽象表达;为了增强网络的表达能力,在网络推理的过程中使用自适应数量的基元表达原始模型,获得更加准确的基元拟合结果。

Loizou 等^[11]以动态图 CNN(dynamic graph CNN, DGCNN)^[12]为基础,提出一种新的图卷积网络模型,为每个点预测其处于 2 个或多个基元边缘处的概率,提取点云的边界信息;而这种边界信息自然地适合不同基元之间的分类和分割,因此使用传统的漫水填充方法,从 1 个随机非边缘点出发对整个输入点云进行基元分割。该方法的问题是分割过程过于依赖点云边界信息,导致分割结果很容易因预测错误的边界信息而产生较大的缺陷。

文献[10-11]方法要么使用的基元类型有限,要么缺少了针对不同种类的基元进行针对性的拟合,未完整地实现参数化基元检测。针对这一问题, Li 等^[13]提出一种端到端的深度神经网络(supervised primitive fitting network, SPFN),可以同时检测平面、球体、圆柱和圆锥 4 种参数化基元,神经网络将为每个点预测所属的基元、表面法向量和基元类别,得到分类分割结果后再通过拟合器将点云拟合为参数化基元,整个网络通过可微的方法完成,实现了端到端的训练。在文献[13]的基础上, Sharma 等^[14]增加了对于参数化曲面的表示,用 B 样条曲面表示人造 CAD 物体中的 2 种曲面形状:开样条曲面以及闭样条曲面,并加入后处理模块得到连接更加紧密的拟合结果;但由于无法进行多尺度重建,该方法常丢失物体中体积较小的部分。类似地, Yan 等^[15]提出 HPNet,其注重使用多种方式对点云特征进行编码,增强特征的表示能力进而提升基元检测结果。首先逐点地预测基元类别、基元参数及表面法向量,然后引入 2 个描述符增强对点云的分割:第 1 个描述符由基元类别和基元参数生成,编码了中心点到达邻域点对应参数所表示的基元的几何一致性;第 2 个描述符由表面法向量生成,编码了某一点和周围邻域点的法向量变化情况,针对边缘区域施加更多的关注。

2 本文方法

人造物体几何基元检测的目的是从输入点云中分割出不同的部件,并通过一组参数化几何基元拟合这些部件。本文提出参数化基元检测网络——RelationNet,分为 3 个阶段。首先,点云特征提取模块通过深度神经网络对输入的三维点云进行特征提取;其次,基于提取的特征,继续使用

网络对点云进行逐点分类,同时使用均值偏移算法和空间偏移预测模块对点云进行基元分割,并针对不同的部件使用文献[13-14]提出的拟合模块对分割结果进行拟合;最后,将拟合得到的参数输入全局结构约束模块进行损失计算,通过梯度回传提升网络分类分割性能. RelationNet 结构如图 1 所示.

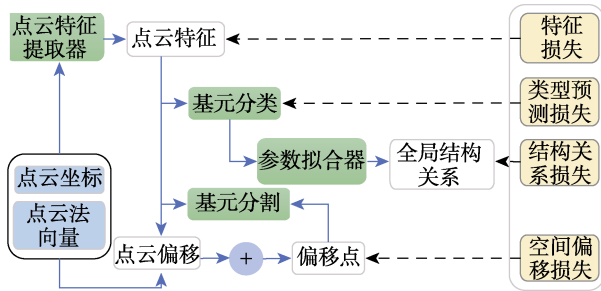


图 1 RelationNet 结构

2.1 点云特征提取模块

为了从无序点云中提取出合适的特征,已有许多不同的深度神经网络. Wang 等^[12]提出的 DGCNN 使用简单的网络结构取得了良好的性能,具有一定的代表性.因此,RelationNet 采用 DGCNN 的关键部分设计点云特征提取模块.

为了更好地提取点云之间的局部特征, DGCNN 里新增了边缘卷积模块.对于具有 n 个点的三维点云 $P = \{p_1, p_2, \dots, p_n\} \subseteq \mathbb{R}^F$, 边缘卷积模块使用 k NN (k -nearest neighbor)算法,为每个点寻找到 K 个邻居 N 并构建无向图 $G = (V, E)$, 其中, $V = \{1, 2, \dots, n\}$ 表示顶点, $E \subseteq V \times K$ 表示边, 顶点 V_i 具有 $(i, j_{i_1}), (i, j_{i_2}), \dots, (i, j_{i_k})$ 共 k 条边, 表示距离点 p_i 最近的点分别为 $\{p_{j_{i_1}}, p_{j_{i_2}}, \dots, p_{j_{i_k}}\}$. 因此, 边缘卷积模块在中心点 p_i 提取到的特征可以表示为

$$p'_i = \max_{j \in N(i)} \text{MLP}(p_i, p_i - p_j).$$

由于点云本身存在着无序性,使用 $\max$ 对称函数对多层感知机(multi-layer perceptron, MLP)的输出特征进行聚集,能够避免因点云的输入顺序不同而带来的输出结果不同.此外,与单独使用点 p_i 的特征或点 p_i 与邻域点 p_j 的差值作为 MLP 的输入相比,堆叠 $(p_i, p_i - p_j)$ 作为输入能够使提取的边缘特征同时保留全局的结构信息和邻域内的局部信息.

在点云特征提取阶段,浅层特征基于三维空间中的邻域点进行特征提取,能够捕获更多的局

部特征并保留更多的细节信息;深层特征基于高维空间中的邻域点进行特征提取,扩大感受野的同时提供了更抽象的全局特征信息.为了缓解深层特征对于点云细节信息丢失的问题, RelationNet 采用跳层连接的特征提取方法,将浅层特征与深层特征进行组合,使得最终提取的特征既包括全局结构信息又含有局部的细节内容.

本文通过堆叠 3 层边缘卷积层组合浅层特征与深层特征,最终为每个点提取 256 维的高维特征描述.通过上述的骨干网络,点云由原本的三维空间映射到了高维空间,携带的高维特征能够为后续空间偏移预测和基元分类分割提供更丰富的上下文信息.

2.2 几何基元分类分割模块

为了实现全局结构关系的提取, RelationNet 中加入几何基元分类分割模块对点云进行分类与分割.该模块以点云的三维坐标和提取的高维特征作为输入,利用 MLP 逐点预测基元类别;同时,利用可微分的均值漂移算法将完整的点云分割为若干个基元.由于每个人造物体都是由不同数量的基元构成的,因此采用均值漂移算法比需要提前指定基元数量的 k -means 算法更具有灵活性.已有工作中,几何基元的分割任务往往只利用深度神经网络提取到的特征,忽略了点云三维坐标本身提供的信息.为此,本文在基元分割任务中引入空间偏移预测模块,能让归属于同一个几何基元的点向模块中心移动,增加该模块后,不仅能额外提供一个显式的约束,还能让网络在分割过程中充分利用高维特征信息和偏移后的三维空间信息,实现信息互补.

基于点云特征提取模块提取的特征,几何基元分类分割模块还将对输入点云进行逐点分类.根据人造物体中常见的基元类型,将输入点云分为 2 种常见基元,这些常见基元又可以分为 2 种类型: (1) 基础基元类型,包括平面、球体、圆柱和圆锥; (2) 样条曲面类型,包括开样条与闭样条 2 种.

2.2.1 空间偏移预测模块

在基元检测任务中,由于人造物体存在不同尺度的部件,在进行部件分割时,与大型部件相邻的细小部件常常会发生错误分割的情况,降低了分割准确度.在分割前,若各个部件向各自的三维中心产生一定偏移,则会在三维空间中产生较明显的区分效果,有助于后续的基元分割任务.因此,本文设计了空间偏移预测模块,通过对象之间的空隙空间来更好地对点进行分组.基于上述提

取的点云特征, 深度神经网络会为每个三维点 $P(x, y, z)$ 预测出 1 个朝向基元中心的偏移量 $\Delta P = (\Delta x, \Delta y, \Delta z)$, 并按照该偏移量更新三维点坐标为 $P' = (x + \Delta x, y + \Delta y, z + \Delta z)$.

但是, 直接预测学习到的偏移不够准确, 导致有少数点偏离物体的中心点. 为了让这些点更好地掌握中心点和邻域点之间点与点的关系, 需要获取更大范围内的全局信息. 对于三维点云 $\{P_i\}_{i=1}^N$, 本文将点 p_i 最终的空间偏移结果调整为 p_i 的 k 邻域点 $\{p_{i_1}, p_{i_2}, \dots, p_{i_k}\}$ 的空间偏移的加和, 并将此过程称为邻域偏移聚合, 表示为

$$V'(P_i) = V(P_i) + \sum_{j=1}^k \frac{V(P_j) - V(P_i)}{\|V(P_j) - V(P_i)\|_2} e^{-\|V(P_i) - V(P_j)\|} \quad (1)$$

其中, $V(\cdot)$ 表示某点经过空间偏移后的位置; 邻域点 p_j 对于中心点 p_i 空间偏移的影响既包含方向上的影响, 也包含距离上的影响, 在受到 p_j 点的影响后, p_i 点最终的空间偏移位置为 $V'(P_i)$. 邻域偏移聚合示意图如图 2 所示. 式(1)中, 利用 $V(P_j) - V(P_i) / \|V(P_j) - V(P_i)\|_2$ 描述 $V(P_i)$ 指向 $V(P_j)$ 的单位方向向量; 同时, 根据 $V(P_i)$ 与 $V(P_j)$ 之间距离的差值作为权重, 衡量邻域点对中心点的影响程度, 即其中 $e^{-\|V(P_i) - V(P_j)\|}$ 项.

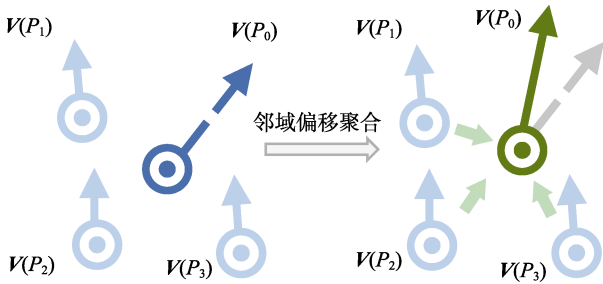


图 2 邻域偏移聚合示意图

在空间偏移的过程中, 该模块显式地要求每个点移向各自基元的中心, 不仅能够将同一类的相邻基元进行分离, 还能使特征主动编码各个部件的偏移信息, 最终偏移点生成过程如图 3 所示.

2.2.2 几何基元分割模块

使用第 2.1 节提出的点云特征提取模块, 网络模型能够获得逐点的特征描述, 但这与获取人造物体中的基元检测结果仍有一定距离, 即网络模

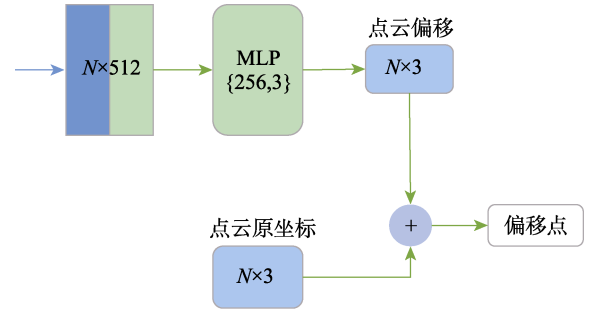


图 3 空间偏移预测过程示意图

型需要依据逐点的高维特征描述对点云进行基元分割. 该步骤一般依赖于聚类算法, 但这种算法都是不可微的, 也无法嵌入网络中进行端到端的训练; 而且, 由于无法提前确定一个物体中存在多少个部件, 因此无法采用 k -means 等需要预先指定聚簇数量 k 的算法. 本文借鉴一种可微的均值偏移方法, 在高维空间根据点云特征为点云进行聚类, 将点云的高维特征向量转换为少量的基元实例标签, 实现基元的分割. 均值偏移及其密切相关的算法使用核密度估计获取一组样本点整体的概率密度, 然后将每个样本往局部密度增加最快的方向进行移动, 最终收敛到同一个点的样本被认为是同一类的成员.

针对点云特征提取模块输入的点云特征 $F = \{F_1, F_2, \dots, F_n\}$, $\forall F_i \in \mathbb{R}^D$, 其中, D 表示特征维度数量. 选择所有输入点作为种子点进行均值偏移, 并在高维空间中计算全体种子点之间的距离矩阵 $M \in \mathbb{R}^{n \times n}$, 然后根据距离矩阵计算种子点的偏移向量. 考虑到每个邻域点对于样本点的影响程度不相同且通常与 2 点之间的距离有关, 因此引入核函数 g 对每个样本点进行加权. 本文采用高

斯核函数, 可以表示为 $g(x) = e^{\frac{-x}{2h^2}}$. 其中, h 表示核函数的带宽. 通过添加高斯核函数, 当 2 个点之间的距离越近时, 核函数值越大, 邻域点对中心点施加的影响也就越大, 最终偏移结果也会更加接近距离较近的点. 添加核函数之后, 每轮迭代中第 i 点的特征偏移可以表示为

$$F'_i = \left(\sum_{j=1, j \neq i}^n F_j g(F_i - F_j) \right) / \left(\sum_{j=1, j \neq i}^n g(F_i - F_j) \right) - F_i.$$

重复迭代多次后, 从均值漂移后的结果中密度最大的区域开始运行非极大值抑制, 将给定距离内的点从全集中移除并不断重复该过程; 最终只保留若干个聚类中心点, 对全体点云寻找距离

其最近的聚类中心点作为部件分割结果。

2.2.3 几何基元分类

基于学习到的点云特征和基元分割结果, 本文使用 2 层全连接层逐点地预测基元类别以进行基元的分类。需要注意的是, 由于基元分类与分割任务由 2 个不同的神经网络分支完成, 因此不同的点对应的基元类型与基元分割结果无直接关系。即使是组成同一基元的点也可能存在不同的基元类型, 最终一个基元的类别由组成该基元的所有点进行投票确定。

2.3 参数拟合模块

几何基元分割之后, 输入的三维点云被分割为各个单独的基元, 每个基元有相应的类别。本文沿用文献[13-14]的方法对分割好的几何基元进行参数化拟合, 根据输入的基元类型, 参数化拟合模块将输出不同种类的参数, 如表 1 所示。

表 1 各几何基元对应参数

基元类型	拟合参数
平面	平面法向量 $\mathbf{n}$, 平面与原点的距离 d
球体	球心 $\mathbf{o}$, 半径 r
圆柱	圆心 $\mathbf{o}$, 半径 r , 轴线方向向量 $\mathbf{c}$
圆锥	顶点 $\mathbf{o}$, 夹角 θ , 轴线方向向量 $\mathbf{c}$
开样条曲面	20×20 控制点网格
闭样条曲面	20×20 控制点网格

需要注意的是, 前 4 种基元的拟合中未使用点云特征提取模块的特征, 而是直接利用了点云的三维坐标、法向量和分类分割结果。因此, 若该阶段的结果不够理想, 原因是分类分割模块的结果输出存在缺陷。这也是第 2.4 节利用基元参数提取点云的全局结构, 进而对基元检测结果施加进一步约束的重要假设。

2.4 全局结构关系提取模块

第 2.1 与第 2.2 节的方法只是基于局部邻域内的点云进行拟合, 未充分考虑几何基元之间的联系, 而基于产品功能性和设计生产的角度考虑, 人造物体的不同基元在物体朝向、摆放位置和大小关系上是有联系的。例如, 相邻的 2 个平面常保持一种相互垂直的朝向; 一个球体的球心常位于相邻圆柱的轴线的延长线上。

本文将组成人造物体的平面、球体、圆柱和圆锥基元之间存在着平行、垂直、轴对齐等包含多个基元的几何关系定义为全局结构关系, 作为几何基元检测过程中的先验知识, 并依此对检测结果进行约束, 帮助网络模型输出一个更加干净、整

洁的几何基元检测结果。为了达到这个目标, 需要对拟合出的几何基元进行全局结构关系提取。本文中使用的基元参数均来自参数化拟合模块的输出, 由于部分基元的全局结构关系难以定义, 如球体之间无法确定平行与垂直关系、圆柱和平面之间不存在轴对齐关系, 因此将这些基元从全局结构关系中排除。

2.4.1 平行与垂直

平行与垂直是人造物体基元中存在最广泛的全局结构关系之一。为了提取基元之间的这种朝向关系, 本文将平面法向量 $\mathbf{n}$ 、圆柱与圆锥的轴线方向向量 $\mathbf{c}$ 作为这 3 种基元的方向向量, 判断基元之间的平行与垂直关系。理想情况下, 对于基元 x_i 和 x_j , 当两者存在平行关系时, 方向向量 $\mathbf{c}$ 应满足 $\mathbf{c}_i \mathbf{c}_j = 1$; 当两者存在垂直关系时, 方向向量 $\mathbf{c}$ 应满足 $\mathbf{c}_i \mathbf{c}_j = 0$ 。

受点云数据的稀疏性和噪声的影响, 从基元拟合结果中提取平行与垂直关系时允许存在 10° 的误差。例如, 当两个基元的方向向量 $\mathbf{e}$ 夹角为 $[85^\circ, 95^\circ]$ 时, 两者属于垂直关系。

2.4.2 轴对齐

对齐是人造物体中广泛存在的, 常使得来自不同基元的中心点和轴线位于同一延长线上, 这也是人造物体具有整齐观感的重要因素。本文将基元之间存在的这种关系称为轴对齐关系, 常见于球体、圆柱和圆锥 3 种几何基元之间, 包含 2 种情况: 一种是球心位于圆柱或圆锥的轴线的延长线上, 另一种是圆柱和圆锥的轴线位于同一延长线上。

针对第 1 种情况, 本文直接计算球心到轴线的距离。对于球体球心 $\mathbf{o}(x_0, y_0, z_0)$ 和另一基元的轴线 $(x-x_1)/l = (y-y_1)/m = (z-z_1)/n$, 球心 $\mathbf{o}$ 到轴线的距离定义为

$$d = \frac{|(x_1 - x_0, y_1 - y_0, z_1 - z_0) \times (l, m, n)|}{\sqrt{(l^2 + m^2 + n^2)}}.$$

当 d 小于给定阈值时, 认为球与圆柱或圆锥存在轴对齐关系, 本文实验中将阈值设置为 0.05 (归一化坐标系下)。针对第 2 种情况, 本文不直接计算 2 条轴线的位置是否属于同一延长线上, 而是选取圆柱的圆心或圆锥的顶点, 利用第 1 种情况的方法计算其到另一基元的轴线延长线的距离。

具体实现中, 本文将参数化拟合模块提取出的基元进行两两组合生成候选集。(1) 对候选集中

的每对基元判断是否满足某种全局结构关系. 例如, 球体和平面的两两组合不适合 3 种全局结构关系, 因此将该组合排除出候选集. (2) 针对保留下来的元素, 根据全局结构关系提取方法利用对应的参数进行结构关系提取. (3) 将提取结果与真实标注进行比较, 判断哪些提取结果属于多余部分, 计算这部分结构关系损失并纳入反向传播的计算中, 实现对结果的进一步约束.

2.5 损失函数

损失函数是深度学习中非常重要的一环, 合适的损失函数能约束网络产生人们期望的结果, 帮助训练过程实现收敛. 下面介绍本文使用的损失函数.

(1) 空间偏移损失函数. 空间偏移预测模块为 N 个点预测其到达对应基元中心点的偏移向量 $\mathbf{O} = \{o_1, \dots, o_N\} \in \mathbb{R}^{N \times 3}$, 通过叠加空间偏移, 基元分割模块能够获得更多的信息以提升分割性能. 为了约束网络产生正确的偏移向量, 使得偏移后的点更加靠近基元的中心, 本文将空间偏移损失函数设计为

$$L_{\text{offset}} = \frac{1}{N} \sum_{i=1}^N \|\mathbf{o}_i - (\hat{\mathbf{c}}_i - \mathbf{p}_i)\|.$$

其中, $\hat{\mathbf{c}}_i$ 表示点 $\mathbf{p}_i$ 所在基元的中心, 读取数据集时, 通过求解基元的三维包围盒的中心点提供.

(2) 特征损失函数. 几何基元分割模块不仅依赖于预测的空间偏移, 也依赖特征提取模块提取到的特征. 与其他点云特征提取模块一样, RelationNet 也希望针对点云特征进行更好的编码, 使得来自同一基元的点在特征空间中相互靠近, 而来自不同基元的点则能够互相远离, 增大不同基元之间的区分度. 本文使用三元组损失函数对提取的特征进行损失计算, 首先从点云中随机选择一个点作为锚点 $\mathbf{P}_A$, 然后从与锚点同一基元的点中选择一点作为类内点 $\mathbf{P}_P$, 最后从其他不同基元当中选择一点作为类外点 $\mathbf{P}_N$, 这 3 个点称为一个三元组 $t = \langle \mathbf{P}_A, \mathbf{P}_P, \mathbf{P}_N \rangle$, 重复该过程, 生成三元组集合 $T = \{t_i\}_{i=1}^{T_s}$, 三元组集合 T 对应的特征损失为

$$\begin{cases} L_{\text{emb}} = \frac{1}{T_s} \sum_{i=1}^{T_s} [\max(0, U - V + \alpha)] \\ U = \|f(\mathbf{P}_A) - f(\mathbf{P}_P)\|^2 \\ V = \|f(\mathbf{P}_A) - f(\mathbf{P}_N)\|^2 \end{cases}.$$

其中, α 表示类内点与类外点之间最少应拉开多大的距离, α 值越大, 表示两者之间需要的距离

也越大, 对于网络提取的特征要求也就越高. 本文将 α 设置为 0.02.

(3) 类型预测损失函数. 为输入点云逐点预测基元类型, 并输出对应每种类型的概率. 本文采用交叉熵损失作为类型预测损失, 其计算公式为

$$L_{\text{type}} = -\frac{1}{N} \sum_{i=1}^N \ln(P_i).$$

(4) 全局结构关系损失函数. 该函数是利用提取到的全局结构关系对基元检测任务进行约束的重要部分. 通过全局结构关系提取模块, 能根据拟合的基元参数将其区分为 R_{para} , R_{orth} 和 R_{align} 这 3 个集合, 分别表示其中存在平行、垂直和轴对齐关系的基元组合. 如 $R_{\text{para}} = \{\dots, \langle i, j \rangle, \dots\}$ 表示基元 X_i 和 X_j 之间存在平行关系. 类似地, 本文利用原始数据集提取出基元之间真实存在的结构关系, 并将其保存在 G_{para} , G_{orth} 和 G_{align} 这 3 个集合中. 对于额外检测出的、不属于原始数据固有的结构关系, 需要引入全局结构关系损失对其进行惩罚. 下面分别介绍本文提取的 3 种结构关系. 若 2 个基元 X_i 和 X_j 误分类为平行但实际上不具有平行关系时, 需要对其进行惩罚. 本文将这 2 个基元平行结构关系损失定义为

$$L_{\text{para}} = \sum_{\substack{(i,j) \in R_{\text{para}} \\ (i,j) \notin G_{\text{para}}}} (\mathbf{e}_i \cdot \mathbf{e}_j - 1)^2.$$

(5) 类似地, 垂直结构关系损失定义为

$$L_{\text{orth}} = \sum_{\substack{(i,j) \in R_{\text{orth}} \\ (i,j) \notin G_{\text{orth}}}} (\mathbf{e}_i \cdot \mathbf{e}_j)^2.$$

对于轴对齐关系, 本文选择轴上一点 $O(x_0, y_0, z_0)$ 到达另一经过点 $O'(x_1, y_1, z_1)$. 且方向向量为 $(\mathbf{l}, \mathbf{m}, \mathbf{n})$ 的轴线的距离作为损失, 即

$$L_{\text{align}} = \frac{|(x_1 - x_0, y_1 - y_0, z_1 - z_0) \times (\mathbf{l}, \mathbf{m}, \mathbf{n})|}{\sqrt{(\mathbf{l}^2 + \mathbf{m}^2 + \mathbf{n}^2)}}.$$

本文用 5 个损失函数进行训练, 最终总的损失函数为

$$\begin{cases} L_{\text{all}} = \lambda_{\text{offset}} L_{\text{offset}} + \lambda_e L_{\text{emb}} + \\ \quad \lambda_{\text{type}} L_{\text{type}} + \lambda_{\text{relation}} L_{\text{relation}} \\ L_{\text{relation}} = L_{\text{para}} + L_{\text{orth}} + L_{\text{align}} \end{cases}.$$

为了平衡损失之间的比例, 避免某些损失过大或过小从而影响网络的正常训练. 本文依照经验设置对应的比例系数 λ , 其中, $\lambda_{\text{offset}} = 10$, $\lambda_e = 1.5$, $\lambda_{\text{type}} = 0.5$, $\lambda_{\text{relation}} = 0.1$.

3 实验结果与分析

3.1 数据集与数据处理

本文实验采用 2019 年 Koch 等^[16]提出的大型 CAD 模型数据集(a big CAD model dataset, ABC), 该数据集包含 100 万个人造 CAD 模型, 其拥有庞大的样本数据, 为基元检测、几何特征检测及形状重建任务提供了充足的数据, 是目前该领域数据量最大的数据集. ABC 数据集中考虑到样本类型之间的平衡性, 保证每类样本都有足够数量的样例. 原始的数据集中, 将基元分成包括基础基元和各种样条曲面在内的 8 个类别, 按照出现频率进行统计, 平面基元的出现频率最高, 约 48%, 其次是圆柱, 占比约 17%, 两者合并占比约 65%. 考虑到平面之间常存在的平行、垂直关系, 以及圆柱和其他基元之间的轴对齐关系, 本文在全局结构关系提取模块中对这 2 种关系进行实验. 由于如此庞大的数据集不适合直接进行训练, 因此本文采用预处理的数据集^[12], 经过处理后, ABC 数据集由 24000 个训练点云、4000 个验证点云和 4000 个测试点云组成, 其中, 每个点云都包含数量不一的平面、圆柱、圆锥、球体和 2 种样条曲面共 6 种基元, 每个点都提供三维坐标、法向量、基元类别和实例标签. 充足的标注能够让本文专注于网络模型的设计, 是验证本文方法的理想数据集.

对于一个输入点云, 这种标注信息依靠数据集中包含的基元类别、实例标签和参数化拟合模块进行生成. (1) 如图 4 所示, 本文首先根据真实的实例标签从点云中提取出 K 个基元实例, 然后根据实例内部多数点的类别确定各个基元对应的类别. 针对每种基础基元类型(平面、圆柱、圆锥和球体), 使用第 2.3 节的方法对点云进行参数化拟合, 获取逐个基元的几何参数. (2) 如图 5 所示, 遍历点云中提取出的各个参数化基元, 使用第 2.4 节的方法对各个基元进行结构关系提取, 并保存在数据集中作为真实标注信息, 实现对训练过程的监督. 最终提取到的 3 种全局结构关系信息如图 6 所示, 其中, 不同颜色分别代表不同的基元.

3.2 对比实验

为了验证 RelationNet 模型的优势, 针对 ABC 数据集, 选取 4 种基准算法进行对比, 包括最近邻(nearest neighbor, NN)及 RANSAC 这 2 种传统的算法, 以及基元监督拟合(supervised primitive fitting, SPFN)^[13]及 ParseNet^[14]这 2 种目前领先的深度学习算法. 将 5 种算法在 ABC 数据集上训练并进行测试, 结果如表 2 所示. 可以看出, RelationNet 得到

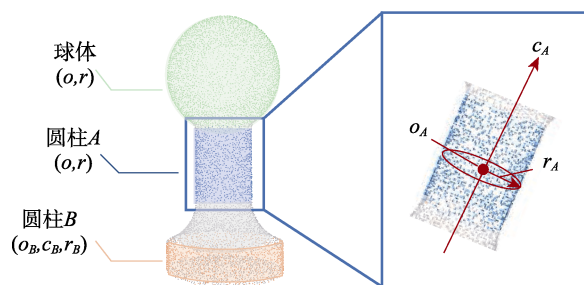


图 4 参数化拟合标注过程



图 5 关系提取标注过程

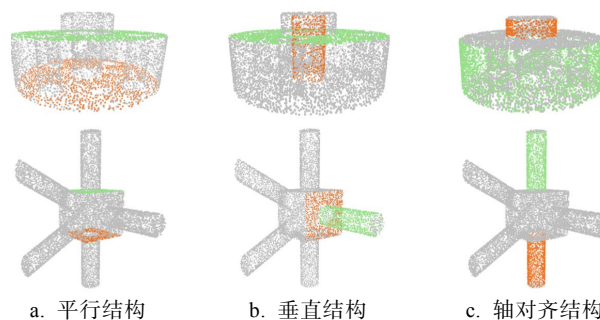


图 6 全局结构关系标注示例

了更好的结果, 通过添加空间偏移预测和全局结构关系约束, 其在分割精度、分类精度上都取得了进步, 其中, 基元分割平均交并比(mean intersection over union, MIoU)从 82.14%提升到 85.32%, 提高了 3.18 个百分点; 而基元分类 MIoU 从 88.6%提升到 90.1%, 提高了 1.5 个百分点.

表 2 ABC 数据集上 5 种算法 MIoU 对比 %

算法	分割 MIoU	分类 MIoU
NN	54.10	61.10
RANSAC ^[2]	67.21	
SPFN ^[13]	69.01	79.94
ParseNet ^[14]	82.14	88.60
RelationNet	85.32	90.10

图 7 所示为 RelationNet 在 ABC 数据集中与 ParseNet 的分割结果对比. 可以看出, 与完全根据提取的点云特征进行基元分割的 ParseNet 相比, 得益于空间偏移预测模块提供的点云偏移, RelationNet 能够在特征提取过程中主动编码单个点与

所在基元之间的相对位置关系,生成比其他算法更加锐利的边缘,如图 7 中第 2 行所示;由于全局结构关系模块将基于整个基元拟合的参数进行损失计算,同时考虑的点数更多,因此提取的损失也具有更强的约束能力,使 RelationNet 不容易出现大面积的分割错误,如图 7 中第 1 行所示.

图 8 所示为本文提出的空间偏移预测模块的结果示例.可以看出,该模块在处理结构较为简单的输入点云时,能够非常准确地将点偏移至基元的中心位置,即使在处理复杂结构的输入点云时,也能在不同基元之间产生一个较为显著间隔,有助于后续的基元分割任务.

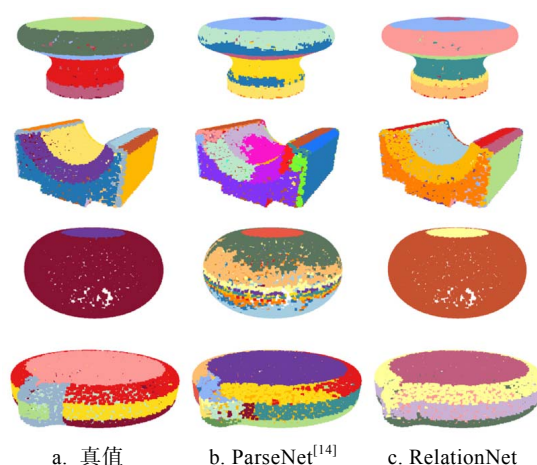


图 7 RelationNet 与 ParseNet 分割结果对比

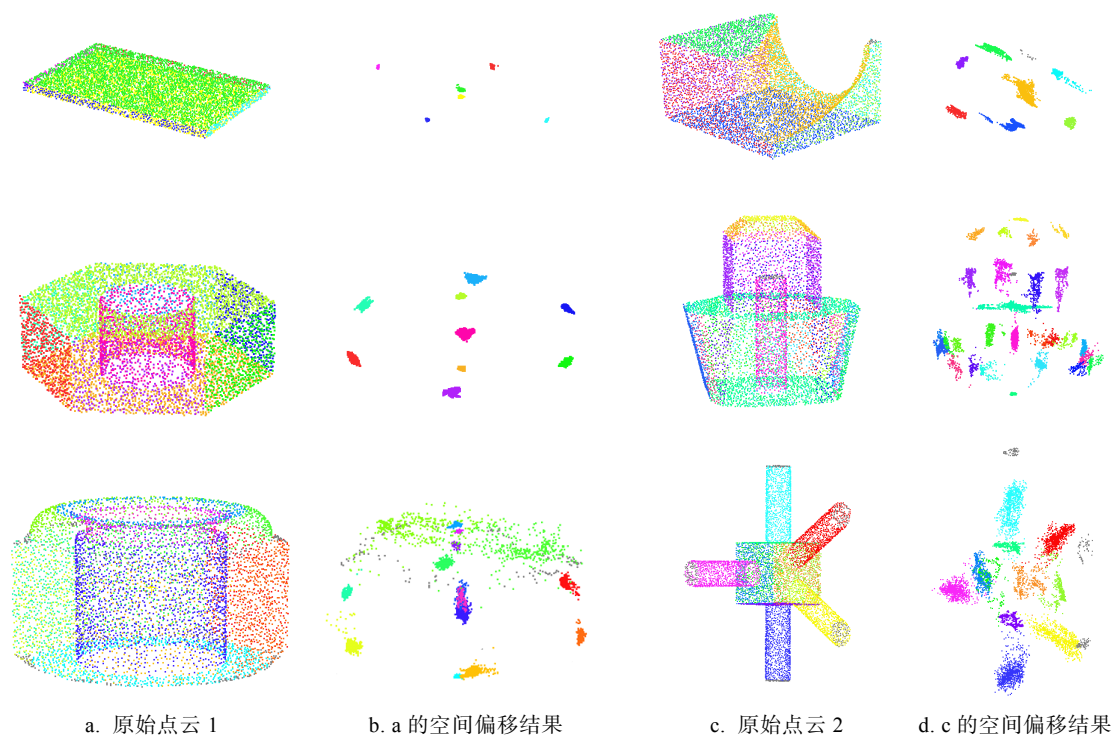


图 8 空间偏移预测模块结果示例

3.3 消融实验

为了验证本文提出的模块对基元检测结果精度提升的有效性,在 ABC 数据集的测试集(简称 ABC 测试集)上进行实验.为了公平地验证每个模块的作用,各组实验迭代相同次数,预训练模型迭代 30 次,端到端训练迭代 15 次.

全局结构引导的人造物体参数化基元检测包含空间偏移预测和全局结构关系约束 2 个模块,围绕这 2 个模块进行消融实验.

空间偏移预测模块显式地要求基元点预测其往对应基元中心点的偏移向量,将该预测结果纳入基元分割模块的输入中作为参考.通过增大不同基元之间的空隙,有效地避免部分基元由于体

积较小且位于相邻基元的接缝处时被忽略的风险.全局结构关系约束模块利用针对不同基元拟合出的几何参数进行全局结构关系的提取,再与数据集中的真实结构关系标注进行比较和计算误差.由于基元的参数是由组成该基元所有点的三维坐标和法向量共同确定的,若存在错误的结构关系,则与空间偏移预测模块相比,这部分损失的反传会影响到更多的点,从而产生更显著的约束力.在 ABC 测试集上的模块消融实验结果如表 3 所示.可以看出,通过增加空间偏移预测模块,RelationNet 的分割 MIoU 由 82.14% 提升到 84.11%,提高了 1.97 个百分点;通过增加全局结构关系约束模块,RelationNet 的分割 MIoU 由 82.14% 提升到 84.67%,

表 3 ABC 测试集上的模块消融实验 %

算法	空间偏移预测	全局结构关系约束	分割 MIoU
ParseNet ^[14]			82.14
RelationNet	√		84.11
RelationNet		√	84.67
RelationNet	√	√	85.32

提高了 2.53 个百分点。

表 4 所示为全局结构关系提取模块提取的 3 种全局结构关系的作用。与未使用全局结构关系对输出结果进行进一步约束的 ParseNet 相比, RelationNet 通过结构关系的提取实现不同基元之间信息的传递, 将基元分割 MIoU 由 82.14% 提升到 84.67%, 证明该模块能够帮助单一基元提升对整体点云结构的感知能力。

表 4 3 种全局结构关系的消融实验 %

算法	平行	垂直	轴对齐	分割 MIoU
ParseNet ^[14]				82.14
RelationNet	√			83.53
RelationNet	√	√		84.36
RelationNet	√	√	√	84.67

4 结 语

针对已有的基于深度学习的基元检测方法在检测过程中过度依赖网络自动提取的特征, 缺乏对全局结构的主动感知的问题, 提出一种人造物体参数化基元检测 RelationNet 来强化网络对于全局结构的主动感知。在 ABC 数据集上进行实验, 通过添加空间偏移预测和全局结构关系约束模块, RelationNet 在分割精度、分类精度上都取得进步, 其中, 基元分割 MIoU 达到 85.32%, 而基元分类 MIoU 达到 90.10%, 超越了现有先进方法, 证明了 RelationNet 的优越性。

虽然 RelationNet 具有良好性能, 但仍有值得探索的改进空间。

(1) RelationNet 提取人造物体中最广泛存在的平行、垂直及轴对齐关系, 并基于这些结果对检测过程进行约束。然而, 人造物体中还存在其他的全局结构关系, 如等夹角平面、轴对称等。通过探索这些新的全局结构关系, 能够将更多的先验知识引入到对基元的检测中, 有望取得更好的结果。

(2) 空间偏移预测模块能够为基元上的点预测一个指向基元中心的偏移向量, 需要网络在特征提取阶段的同时对点的局部邻域和全局结构进

行感知和编码。然而, 目前基元中心的真值是通过计算其包围盒中心点位置而获取的, 对于某些样条曲面基元, 这种方式获取的中心点可能距离其表面距离较远, 或与人们认为的曲面中心不符。因此, 下一步将探索如何根据不同情况定义一个合适的中心点标签, 帮助网络预测一个更合适的偏移位置; 以及如何实现对基元之间存在的嵌套关系等复杂结构的感知, 帮助基元更好地分割。

(3) 深度学习的方法都是由数据驱动的, 导致在一个数据集上训练的网络有时难以直接运用到另一个数据集的测试中。本文通过设置空间偏移预测和全局结构关系提取模块, 增强网络对全局结构的感知能力, 提升对 CAD 模型的基元检测性能。但在常见的人造物体中, 还存在诸如家居家具、车辆, 以及更复杂、更大规模的室外建筑物数据, 它们在基元种类、全局结构关系、场景先验知识方面都有所不同, 需要考虑不同场景下的差异问题。如何寻找不同场景、不同种类下的共同点信息, 提取共同特征来获得一种通用的基元检测方法, 也是一个具有挑战性的问题。

参考文献(References):

- [1] Kaiser A, Zepeda J A Y, Boubekeur T. A survey of simple geometric primitives detection methods for captured 3D data[J]. Computer Graphics Forum, 2019, 38(1): 167-196
- [2] Schnabel R, Wahl R, Klein R. Efficient RANSAC for point-cloud shape detection[J]. Computer Graphics Forum, 2007, 26(2): 214-226
- [3] Duda R O, Hart P E. Use of the Hough transformation to detect lines and curves in pictures[J]. Communications of the ACM, 1972, 15(1): 11-15
- [4] Chum O, Matas J. Matching with PROSAC - progressive sample consensus[C] //Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2005: 220-226
- [5] Oesau S, Lafarge F, Alliez P. Planar shape detection and regularization in tandem[J]. Computer Graphics Forum, 2016, 35(1): 203-215
- [6] Zhou Y, Yin K X, Huang H, et al. Generalized cylinder decomposition[J]. ACM Transactions on Graphics, 2015, 34(6): Article No.171
- [7] Hough V P C. Method and means for recognizing complex patterns: US, 3069654[P]. 1962-12-18
- [8] Li Y Y, Wu X K, Chrysathou Y, et al. GlobFit: consistently fitting primitives by discovering global relations[J]. ACM Transactions on Graphics, 2011, 30(4): Article No.52
- [9] Yu L Q, Li X Z, Fu C W, et al. EC-net: an edge-aware point set consolidation network[C] //Proceedings of the 15th European Conference on Computer Vision. Heidelberg: Springer, 2018: 398-414

- [10] Tulsiani S, Su H, Guibas L J, *et al.* Learning shape abstractions by assembling volumetric primitives[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2017: 1466-1474
- [11] Loizou M, Averkiou M, Kalogerakis E. Learning part boundaries from 3D point clouds[J]. Computer Graphics Forum, 2020, 39(5): 183-195
- [12] Wang Y, Sun Y B, Liu Z W, *et al.* Dynamic graph CNN for learning on point clouds[J]. ACM Transactions on Graphics, 2019, 38(5): Article No.146
- [13] Li L X, Sung M, Dubrovina A, *et al.* Supervised fitting of geometric primitives to 3d point cloud[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2019: 2647-2655
- [14] Sharma G, Liu D F, Maji E S, *et al.* PARSENET: a parametric surface fitting network for 3D point clouds[C] //Proceedings of the 16th European Conference on Computer Vision. Heidelberg: Springer, 2020: 261-276
- [15] Yan S M, Yang Z P, Ma C Y, *et al.* HPNet: deep primitive segmentation using hybrid representations[C] //Proceedings of the IEEE/CVF International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2021: 2733-2742
- [16] Koch S, Matveev A, Jiang Z S, *et al.* ABC: a big CAD model dataset for geometric deep learning[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2019: 9593-9603