

三维形状智能生成方法综述

肖懿, 乔锦浩, 季铁
(湖南大学, 长沙 410082)

摘要: **目的** 通过对近年来基于深度学习的三维形状智能生成方法进行梳理和分析, 总结该方向存在的问题与挑战, 预测未来发展趋势, 为相关研究者提供参考。**方法** 首先从数据表示角度分析三维形状智能生成中常用的体素、点云、网格、隐函数等表示方法的优缺点并总结了常用的数据集。其次从三维形状生成内容的可控性出发, 综述了目前研究中整体生成、结构感知生成和交互式生成方法。最后从生成结果的相似性、多样性和结构一致性方面总结了在三维形状智能生成中常用的评价指标。**结论** 三维形状智能生成领域虽然已经取得明显发展, 但是在数据集的规模、生成方法的有效性和评价指标的全面性方面仍然面临较大挑战。

关键词: 三维形状生成; 深度学习; 形状表示; 智能生成; 形状评价

中图分类号: TB472 **文献标识码:** A **文章编号:** 1001-3563(2021)22-0078-16

DOI: 10.19554/j.cnki.1001-3563.2021.22.011

Summary of Approaches of Intelligent 3D Shape Generation

XIAO Yi, QIAO Jin-hao, JI Tie
(Hunan University, Changsha 410082, China)

ABSTRACT: By investigating and analyzing the deep learning based approaches in 3D shape intelligent generation in recent years, this paper summarizes the problems and challenges of this field and predict the tendency of the future work, to provide a technical reference for researchers in related fields. Firstly, the advantages and disadvantages of the commonly used 3D shape representation methods and datasets are analyzed, such as voxels, point clouds, meshes, implicit functions, etc. Secondly, from the perspective of controllability of 3D shape generation objects, holistic generative methods, structure-aware generation methods and interactive generation methods are reviewed. After that, the evaluation metrics commonly used in 3D shape intelligent generation, such as similarity, diversity and structural consistency of the generation results are summarized. Although the field of 3D shape intelligent generation has made significant development, it still faces great challenges in terms of the size of data sets, the effectiveness of generation methods and the comprehensiveness of evaluation metrics.

KEY WORDS: 3D shape generation; deep learning; shape representation; intelligent generation; shape evaluation

作为设计与制造业中表达产品的基本形式, 三维形状(3D Shape)被广泛应用于各行各业。随着近年来信息技术和人工智能的飞速发展, 工业互联网和工业4.0中的关键技术数字孪生更是对三维形状的智能生成提出了迫切的需求。理想的三维形状包含比较精细的几何细节和复杂的结构, 同时理想的三维形状生成方法以结构感知的方式表达3D形状和场景, 且允

许用户对形状的拓扑结构和几何细节进行操作和编辑以生成符合目标的3D内容。不同于2D图像可以由规则的像素网格统一表示, 3D形状根据不同的应用需求存在多种表示方式, 如深度图像、多视角图像、体素、点云、网格、隐函数等。在不同应用中, 三维形状的表现形式很大程度上决定了最终实现的效果和性能, 没有一种表示可以很好地适用于所有的应

收稿日期: 2021-06-09

基金项目: 国家自然科学基金(61872137); 湖南省科技创新计划(2021RC3064); 湖南省自然科学基金(2020JJ4009)

作者简介: 肖懿(1984—), 男, 湖南人, 博士, 湖南大学教授, 主要研究方向为智能设计方法、机器学习、图形图像处理。

通信作者: 乔锦浩(1995—), 男, 河南人, 湖南大学博士生, 主要研究方向为人工智能、图形图像处理、智能设计方法。

用。三维形状生成方法大致可分为传统三维生成方法和基于深度学习的三维生成方法。传统的三维模型生成一般利用深度信息通过几何约束进行特征立体匹配实现。如 2011 年, Newcombe 等人^[1]首次实现仅使用手持式 Kinect 和 GPU 硬件准确实时地重建静态室内场景。基于传统的三维重建方法在对目标对象进行重建时易受外部信息和噪声干扰, 且这类方法主要依赖于纯粹的数学运算, 对输入信息的全面性、准确性要求非常苛刻, 因此一部分传统方法依靠昂贵的信息采集设备来满足以上要求, 如 3D 扫描仪等, 这给三维重建工作带来了许多挑战^[2]。

近年来, 随着计算能力增强和数据规模增加, 基于深度学习的三维形状的生成方法得到了广泛的研究并取得了长足的进步。这类方法以数据驱动的方式, 利用神经网络学习形状的数据分布, 直接生成合理的三维形状。根据生成三维形状的可控性, 基于深度学习的形状生成方法大致可以分为 3 类: 整体生成方法^[3-6]、基于结构感知的生成方法^[7-10], 以及交互式的生成方法^[11-12]。整体生成方法利用神经网络以一种整体的且结构无关的方式简化生成三维形状的难度, 但这类方法通常在生成复杂形状时对形状的结构和细节表现不足, 且难以控制生成结果。基于结构感知的生成方法以简单的部件为基本单元, 通过部件与部件间的结构关系组成复杂和精细的三维形状。这类方法通常将三维形状解耦为可以独立控制的几何和结构, 实现对生成结果的结构和几何一定程度上的单独控制和编辑。交互的生成方法则通过草图或某些操作工具利用输入和生成三维形状之间的密集对应关系, 实现对生成结果进行局部控制和更改。综上所述, 相较于传统方法, 基于深度学习的三维形状生成方法能够以数据驱动的方式生成形状具有更高的质量。因此, 本文将更多地关注最近基于深度学习的形状智能生成方法。在下面的章节中, 首先介绍三维形状表示方法(第 1 节)和常用数据集(第 2 节), 之后根据在形状生成中的可控性对其进行分类归纳(第 3 节), 并对三维形状生成研究中常用的评价指标进行总结(第 4 节), 最后对该领域所面临的挑战和发展

趋势进行展望(第 5 节)。术语对照见表 1。

1 数据表示

基于深度学习的三维形状生成方法的研究与三维形状表示息息相关, 合适的三维形状表示能简化生成的难度, 提高生成的精度和质量。常用的三维模型的表示类型有: 多视角图像(Multi-views image)、体素(Voxel)、点云(Point cloud)、网格(Mesh)、部件表示(Part-based representation)、隐函数(Implicit function)。本节将分别介绍这些表示类型的优缺点及其相关数据集。

多视角图像表示是利用图像的规则结构在二维域中表示三维模型, 数据处理更加简单且能避免使用体素表示所耗费的大量空间资源。根据是否包含深度信息, 多视角图像可以分为 RGB(颜色)或 RGB-D(颜色和深度)图像。由于多视角图像包含的 3D 形状信息不完整, 通常需要额外的后处理来生成合理的三维形状^[6]。Wang 等人^[13]将多视角重建定义为一个序列到序列的预测问题, 并提出 VOT 的新框架, 使用 Transformer 网络统一特征提取和视图融合, 取得较好的重建效果。

体素是像素概念的三维扩展。类似 2D 中的像素, 体素表示方法在 3D 空间中使用具有规则结构的体积网格表示形状。因此, 部分在二维图像领域被证明非常有效的处理方法或神经网络框架^[14-16]可以很容易地扩展到体素形式, 如将二维卷积推广到三维卷积^[4-5,17]。但是随着生成形状分辨率的提高, 体素表达所需的内存成立方增长, 同时所处理信息的维数也变得巨大。因此, 这类方法通常将分辨率限制在 64 个网格^[3]以内, 但这也导致三维形状表面大量的几何信息丢失, 影响生成形状的细节表达。为了改善这种情况, 研究者提出了可微移动立方体操作^[18], 以及八叉树策略^[19-23]对这类方法进行优化, 以期实现在不占用过多内存的情况下生成更精细的结果。

点云表示方法使用一组采样的三维点来表示形状的表面。点云可以表示为 $n \times d$ 的矩阵形式, n 代表三维形状表面的 n 个点, d 代表每个点的特征维数,

表 1 术语对照
Tab.1 Glossary of terms

中文	英文	释义
形状	Shape	整个三维模型
几何	Geometry	组成三维形状的基本部件的几何细节
结构	Structure	基本部件之间的关系如父子、对称、平移、旋转等
基元	3D primitives	对三维部件的简化表示, 如使用有方向的包围盒代表该部件
隐函数	Implicit function	由神经网络拟合函数, 无法使用公式显示表达。
潜空间	Latent space	由神经网络学习得到特征表示
潜向量	Latent vector	由潜空间采样得到的向量

$d=3$ 表示该点的三维坐标 (x,y,z) , 同时还可以为该点指定其他的属性(如 RGB 颜色、深度等)。点云表示的优势在于便于获取和简单灵活, 但是由于点之间缺乏顺序和连接信息, 且点云数据多存在噪声和排列无序等问题, 点云数据处理仍有较大挑战。在文献[24-25]中, 研究人员使用顺序不变算子(如深度神经网络中的最大池化算子)来缓解点之间缺乏顺序问题。文献[26]使用神经网络学习一个点采样器, 以一幅图像和一个随机向量作为输入, 输出一个 $n \times 3$ 矩阵来表示预测点集(n 个点的 x,y,z 坐标)。并提出用倒角距离(Chamfer Distance, CD)和绝对移动距离(Earth Mover's Distance, EMD)^[27]作为损失函数对网络进行训练。

多边形网格表示使用一组带公共顶点的凸多边形表面表示形状, 可以用更少的内存和计算成本描述更高质量的3D形状。常用的多边形网格有三角形网格、四边形网格等。网格可以表示为点、边、面的集合 $\langle \mathbf{V}, \mathbf{E}, \mathbf{F} \rangle$, 其中 \mathbf{V} 为网格的 n 个顶点, \mathbf{E} 表示网格中顶点之间的连接关系, \mathbf{F} 表示网格中的三角面片^[28]。与点云表示相比, 基于网格的表示方法中包含相邻点之间的连通性, 因此更适合于描述形状的局部区域。网格表示作为非欧几里得空间的一种典型表示形式, 在空间域和光谱域都可以通过深度学习模型进行处理^[29]。Wang等人^[30]利用输入图像中提取的感知特征, 通过将椭球体逐步变形以生成合理的几何形状, 实现从RGB图像重建基于网格的3D形状。Wen等人^[31]提出GCN框架使用从多个图像构建的感知特征统计来搜索最优变形来改进粗糙网格, 从而生成准确的网格表示3D形状。

基于部件的表示方法将复杂三维形状视为简单的基本部件集合, 并将这些基本部件以线性^[32]或层次^[10,33-34]的形式进行组织, 从而得到结构化的表示。通常, 三维形状的每个基本部件几何细节通过非结构化模型进行处理, 部件关系由循环神经网络(RNNs)^[32]、递归神经网络(RvNNs)^[9]或其他网络进行结构化建模。与其他表示方式相比, 基于部件表示方法在表现复杂形状及形状中部件之间的关系(如对称、支持、被支持等)方面具有更强的能力。Li等人^[9]首次使用神经网络对三维形状结构进行编码并提出了GRASS模型。他们利用层次二叉树(Hierarchical binary tree)来描述形状结构, 其中子节点通过邻接关系或对称关系合并到父节点中。该结构树中的叶子表示各部分的有向包围盒(Oriented Bounding Box, OBB)和几何特征, 中间节点表示子节点的几何特征和子节点之间的结构关系。

与体素、点云、网格方法这种对形状进行采样的方法不同, 隐函数(Implicit function)表示方法将形状表面建模为可学习的隐函数^[1,35]。这类方法使用深度神经网络学习的隐函数来定义空间中三维点和形

状表面之间的关系。原则上, 这种表示方式能够提供具有合理内存消耗的无限分辨率的并能够表示拓扑变化的三维形状。然而, 隐函数不能直观反映三维形状的几何特征, 通常需要额外的处理转换为体素、点云或网格等显式表达^[18]。Mescheder等人^[36]提出Occupancy Networks, 通过神经网络学习隐式地将三维表面表示为深度神经网络分类器的连续决策边界, 使用占用函数反映三维点相对于三维形状表面的状态, 函数值为1表示三维点在形状内, 值为0表示三维点不在形状内。最后他们对生成的隐式场进行多分辨率等值面提取(MISE)和移动立方体算法处理, 渲染成高质量的网格。

2 数据集

随着三维扫描仪的发展, 三维模型的获取变得更加容易, 不同表示形式的三维形状数据集不断涌现, 促进了深度学习在不同3D表示上的发展。选择合适的数据集有利于加大深度学习生成模型的质量和泛化性能。三维形状生成数据集示例见图1。NYU Depth^[37-38]数据集是公开的大型室内场景数据集, 由Microsoft Kinect的RGB和Depth摄像头采集的视频序列组成, 其样例见图1a。数据集包含标注数据、原始数据、数据处理工具包以及训练集测试集划分。标注数据集由成对的RGB和Depth同步帧组成, 每个图像中添加了密集的标签, 深度图使用Levin等人^[39]的着色方案对其缺失值进行了填充。原始数据集包含原始图像和来自Kinect的加速度计采样, RGB和Depth相机采样率介于20~30 fps。其V1版本^[37]收集64个类别, V2版本^[38]收集464个场景类别, 1449个标注的RGB和深度图像对, 以及407024个新的原始数据, 对每个对象使用类别和实例编号进行标注(如Cup1, Cup2)。ScanNet^[40]也是一个大型的带注释的RGB-D视频数据集, 见图1b。该数据集共搜集1513个场景的数据、250万个视图、21个类型对象, 1201个场景用于训练, 312个场景用于测试。同时该数据集使用3D相机姿态、表面重建和实例级语义分割进行注释。

ModelNet^[41]是用于三维形状分析的著名CAD模型数据集之一, 包括662个类别的127915个3D CAD模型, 覆盖了最常见的3D形状类别, 见图1c。该数据集通过互联网收集每个类别的3D CAD模型, 然后由人工进行分类。特别地, 他们选择10个对象类别, 并手动删除不属于这些类别的模型, 同时还手动对齐了这10类CAD模型的方向以得到非常干净的数据集ModelNet 10。同样地, ModelNet提供了包括40个类别的子集ModelNet 40。

ShapeNet^[42]是一个大规模的点云数据集, 包含超过300万个模型, 超过4000个类别, 见图1d。ShapeNet

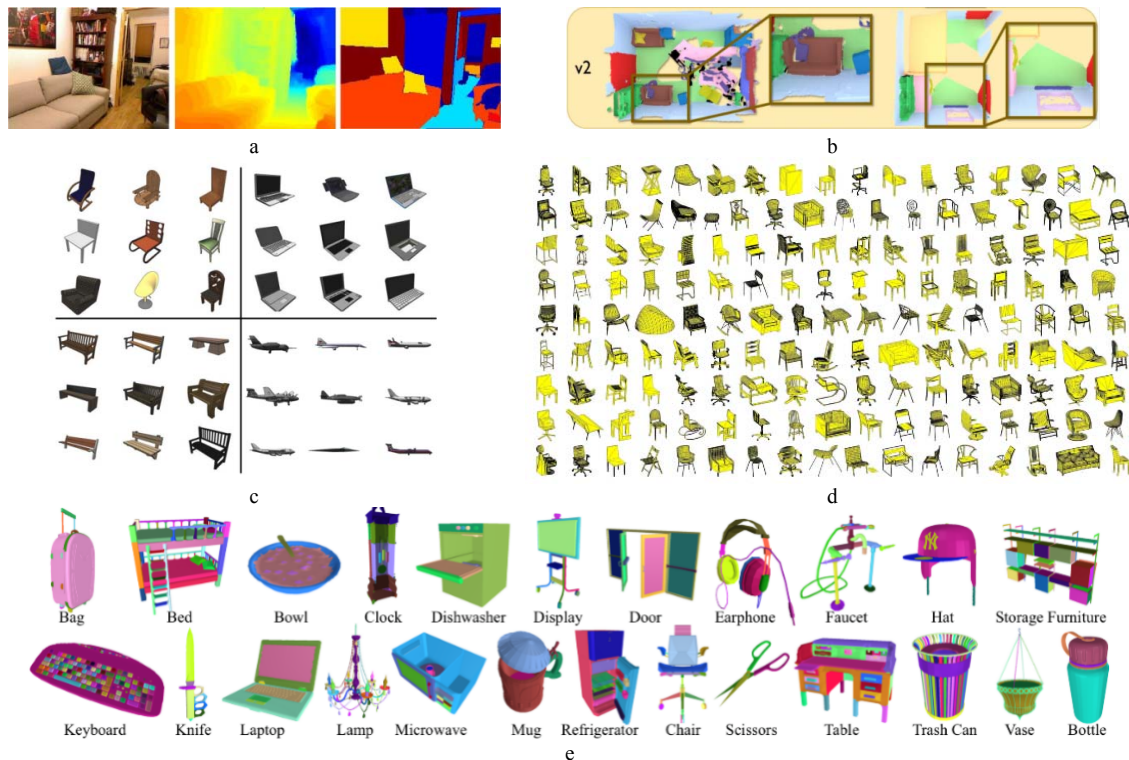


图 1 三维形状生成数据集示例
Fig.1 The samples of 3D shape generation datasets

为数据集中的 3D 对象提供了丰富的注释，包括形状所属类别标签、部件标签、部件对称信息等。ShapeNet 还包含两个较小的子集：ShapeNetCore 和 ShapeNetSem。ShapeNetCore 包含干净的 3D 模型和手工验证的类别和对齐注释。它涵盖了 55 个常见的物体类别，约 51 300 个不同的 3D 模型，涵盖了 3D 基准数据集 PASCAL 3D+ 的 12 个对象类别。ShapeNetSem 是一个更小、注释更密集的子集，包含 270 个类别共 12 000 个模型。除了手工验证的类别标签和一致的对齐之外，这些模型还使用真实世界的特征进行注释，在类别级别上估计它们的材料组成，以及它们的总体积和重量。

PartNet^[43]是一个更详细的 CAD 模型数据集，提供细粒度层次结构的部件注释。数据集包含 573 585 个零件实例，超过 26 671 个 3D 模型，涵盖 24 个对象类别如桌子、椅子、橱柜、台灯等，见图 1e。该数据集极大地促进了语义分割、形状编辑和形状生成等 3D 对象应用发展，并能支持和促进形状分析、动态 3D 场景建模和仿真、功能可见性分析等任务的研究。PartNet 建立了细粒度语义分割、层次语义分割和实例分割三维零件识别基准任务。

3 基于深度学习的三维形状方法

在 2D 图像生成领域，基于深度学习的深度生成模型取得了巨大成功，大致有着生成图像质量越来越高，生成图像逐步可控的趋势。如 BIGGAN^[44]能够

通过学习图像数据分布，生成以假乱真的高清图像，但生成内容不可控。StyleGAN^[45]将图像生成解耦为风格和内容，以实现一定程度上单独控制生成图像中的风格。更进一步地，Deep Face Drawing^[46]实现了交互式生成方法，通过绘制草图实时控制生成图像内容。同 2D 生成模型的研究类似，三维深度生成模型的研究也存在类似的趋势，本文根据对生成三维形状的可控性，将基于深度学习的三维形状生成方法分为整体生成方法、结构感知的生成方法和交互生成方法。

3.1 整体生成方法

整体生成方法一般从潜空间中采样随机向量，然后直接生成三维形状，生成过程具有完全的自由度，生成的形状不受部件注释或者任何预先确定的结构关系的影响和制约。这类方法虽然已经取得了实质性的进展，但是由于三维空间中形状的多样性和高度复杂性，精细形状的生成仍然非常具有挑战性。此外，由于这类方法是直接学习从潜空间到三维模型的映射，仅能描述采样随机向量与整体形状的变化，缺乏对部件和结构的感知，所以无法有效地对生成形状进行控制。接下来本文将对不同表示形式的三维形状整体生成方法进行概述。三维形状整体生成模型比较见表 2。

3.1.1 体素

基于体素表示的整体生成结果见图 2。Sharma

表2 三维形状整体生成模型比较

Tab.2 The comparison of holistic 3D shape generative models

文献	三维表示类型	可控性	评价指标
Sharma 等人	Voxel	整体生成	Average error
Girdhar 等人	Voxel	整体生成	Average Precision
Wu 等人	Voxel	整体生成	Average precision
Achlioptas 等人	Point clouds	整体生成	JSD, MMD-CD, MMD-EMD, COV-EMD, COV-CD
Cai 等人	Point clouds	整体生成	MMD-CD, MMD-EMD, COV-EMD, COV-CD
Sun 等人	Point clouds	整体生成	PointNet Distance, mIoU
Yang 等人	Point clouds	整体生成	1-Nearest Neighbor Accuracy, MMD-CD, MMD-EMD, COV-EMD, COV-CD
Ramasinghe 等人	Point clouds	整体生成	FID, MMD-CD, MMD-EMD, COV
Wang 等人	Mesh	整体生成	CD, EMD
Wen 等人	Mesh	整体生成	CD, F-score
Dominic 等人	Mesh	整体生成	IoU
Groueix 等人	Mesh	整体生成	CD, Euclidean distance
Nash 等人	Mesh	整体生成	CD, Accuracy
Chen 等人	Mesh	整体生成	CD, Normal Consistency, Light Field Distance
Chen 等人	Implicit function	整体生成	COV-LFD, MMD-LFD
Park 等人	Implicit function	整体生成	CD, EMD, Mesh Acc

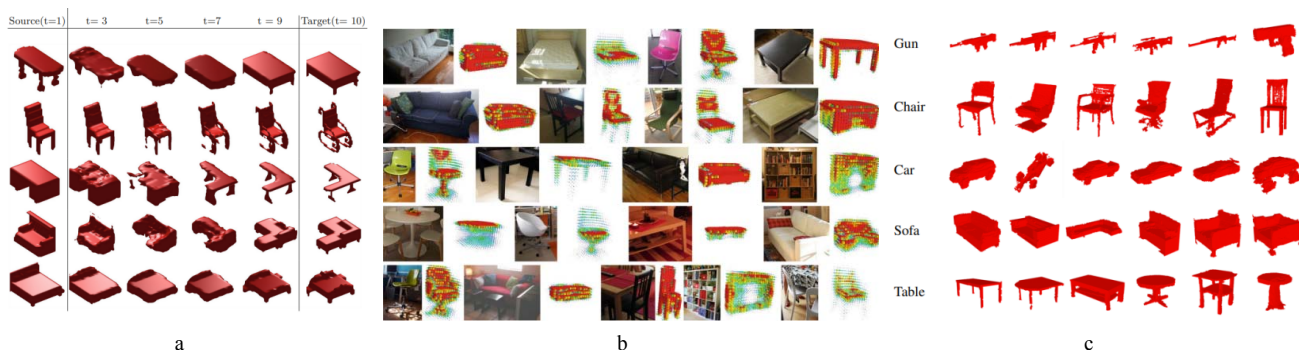


图2 基于体素表示的整体生成结果

Fig.2 The results of holistic generative models based on voxel representation

等人^[47]提出了一种全卷积自动编码器 VConv-DAE, 学习从噪声数据估计三维体素网格。该方法在去噪和形状补全等具有挑战性的任务上取得较好效果, 其学习得到的三维模型嵌入表示在分类和形状插值任务中都表现较好, 其形状插值效果见图 2a。Girdhar 等人^[5]认为二维图像应包含三维模型的部分表示, 为此他们结合 CNN 和自动编码器构建了从 2D 图像生成三维体素的 TL-embedding 网络。该网络包括两个组成部分: (1) 自动编码器, 它将一个 3D 体素网格映射到 64 维嵌入空间, 并将其解码回体素网格以确保三维表示是可生成的, 能够从嵌入的潜在空间生成出三维形状; (2) 卷积网络, 它将二维图像映射到 64 维嵌入空间, 确保三维表示是能从图像预测的。自动编码器的目标是重构体素网格, 卷积网络的目标是预测中间嵌入, 其 3D 重建结果见图 2b。Wu 等人^[4]提出基于三维的生成对抗网络 3D-GAN 直接生成三维

对象。该模型建立了从低维概率空间到三维物体空间的映射, 能够实现三维形状的随机生成、三维模型进行插值和形状运算等任务, 且 3D-GAN 还可以同变分自编码器组合实现从 2D 输入图像重建三维物体, 其随机生成结果见图 2c。然而, 受制于体素表示的弊端, 他们的方法产生的结果只能产生 $64^{[3]}$ 个网格的低分辨率的模型。

3.1.2 点云

基于点云表示的整体生成结果见图 3。Achlioptas 等人^[48]提出了基于点云表示的 GAN。他们首先训练了一个具有较好重构质量和泛化能力的深度自动编码器 (AE) 网络, 使用该自动编码器来学习三维点云的密集表示, 实现如语义部件编辑、形状插值以及形状补全等任务。然后他们提出了在原始点云上训练生成模型 r-GAN 和在 AE 潜空间中训练的生成模型 l-GAN, 后者生成质量有显著改进, 生成结果

见图 3a。Cai 等人^[49]将三维形状生成看作将随机三维点移动到形状表面区域的过程。他们通过在非归一化概率密度上执行随机梯度上升将采样点移动到高似然区域，学习对数密度场的梯度来表示三维形状。他们的方法是直接建模梯度，使用预测和真实梯度场的 L2 损失进行训练，从而能够利用更具表现力的网络来模拟复杂的分布且更容易优化。除此之外，他们的模型能够提供形状的隐式表达，在三维形状生成和超分辨率应用方面均表现良好，其形状生成结果见图 3b。

Sun 等人^[50]以逐点的方式建模三维空间分布，提出自回归的点云生成模型 PointGrow，实现多样化且逼真的 3D 形状生成。PointGrow 充分利用点云之间的相关性，以已生成的点的分布为条件生成新的三维点，使 3D 形状生成过程具有更好的可解释性。同时，PointGrow 使用自注意模块来增强模型捕捉点云间远程依赖关系的能力，其在无条件和有条件的点云生成、形状补全等任务上都取得了令人满意的性能，生成结果见图 3c。PointFlow^[51]采用一种两层层次分布来生成三维点云，第一层为形状的分布，第二层为给

定形状下点云的分布，这种表示能够在某一种类中采样形状且从形状中采样任意数量的点云。PointFlow 不仅可以忠实地重建点云，而且能够以无监督的方式学习三维形状的生成，其形状生成结果见图 3d。不同于以上在空间域的方法，Ramasinghe 等人^[52]提出了一种谱域生成对抗网络 (Spectral-GAN) 的三维点云合成方法，在谱域中以最小的计算开销生成高分辨率点云。同时他们提出了一个可微的方案实现从谱域到空间域之间的转换。Spectral-GAN 在点云生成任务中表现良好并且能够以无监督的方式准确地重建三维物体，其形状生成结果见图 3e。

3.1.3 网格

基于网格表示的整体生成方法结果见图 4。Wang 等人^[30]提出通过变形先验球体上网格从 RGB 图像生成 3D 形状的 Pixel2Mesh 模型。该模型采用基于图的卷积神经网络表示 3D 网格，并利用从输入图像中提取的感知特征，将先验的球体逐步变形生成合理的几何形状。同时，他们采用从粗到细的策略使整个变形过程稳定收敛，并定义各种网格相关的损失来捕捉形状不同层次的属性，以保证视觉上的吸引力和物理

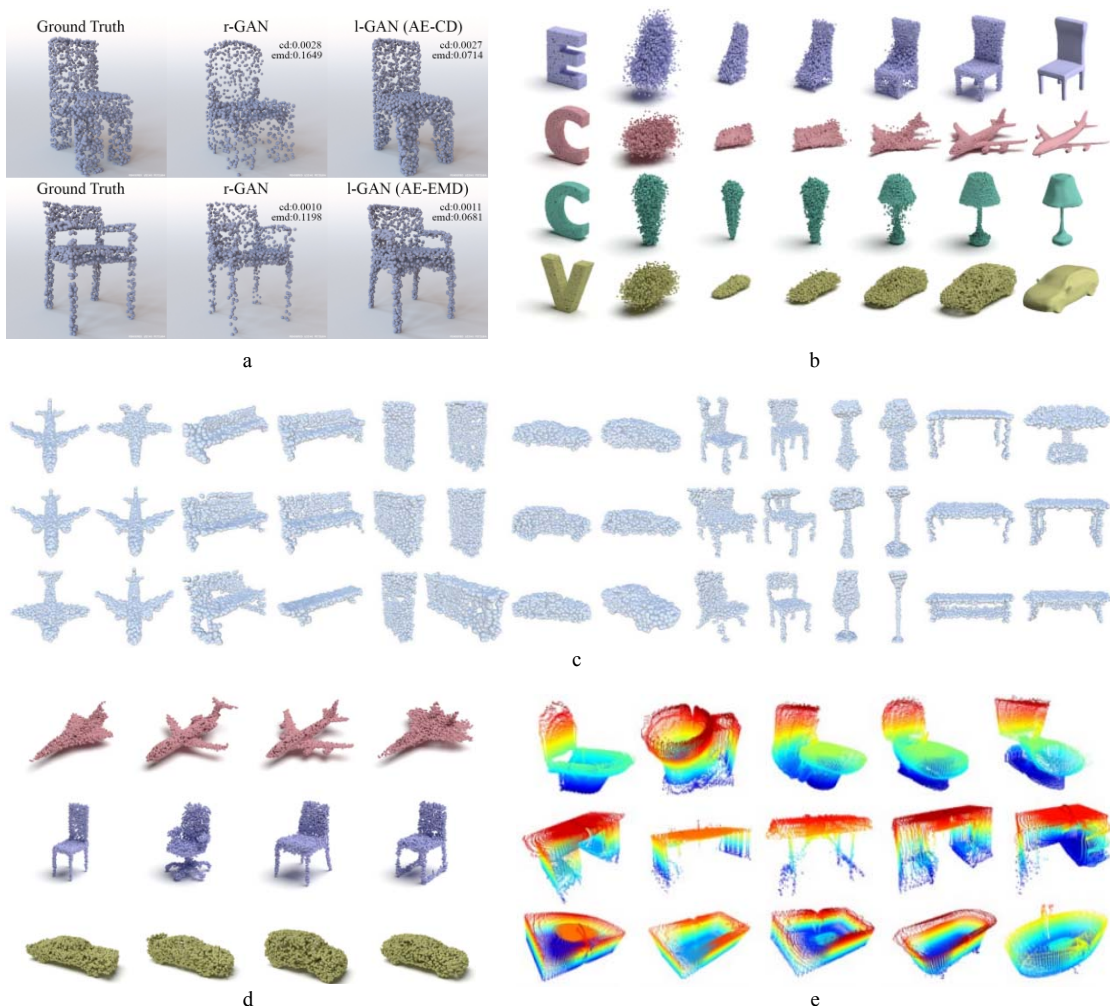


图 3 基于点云表示的整体生成结果

Fig.3 The results of holistic generative models based on point cloud representation

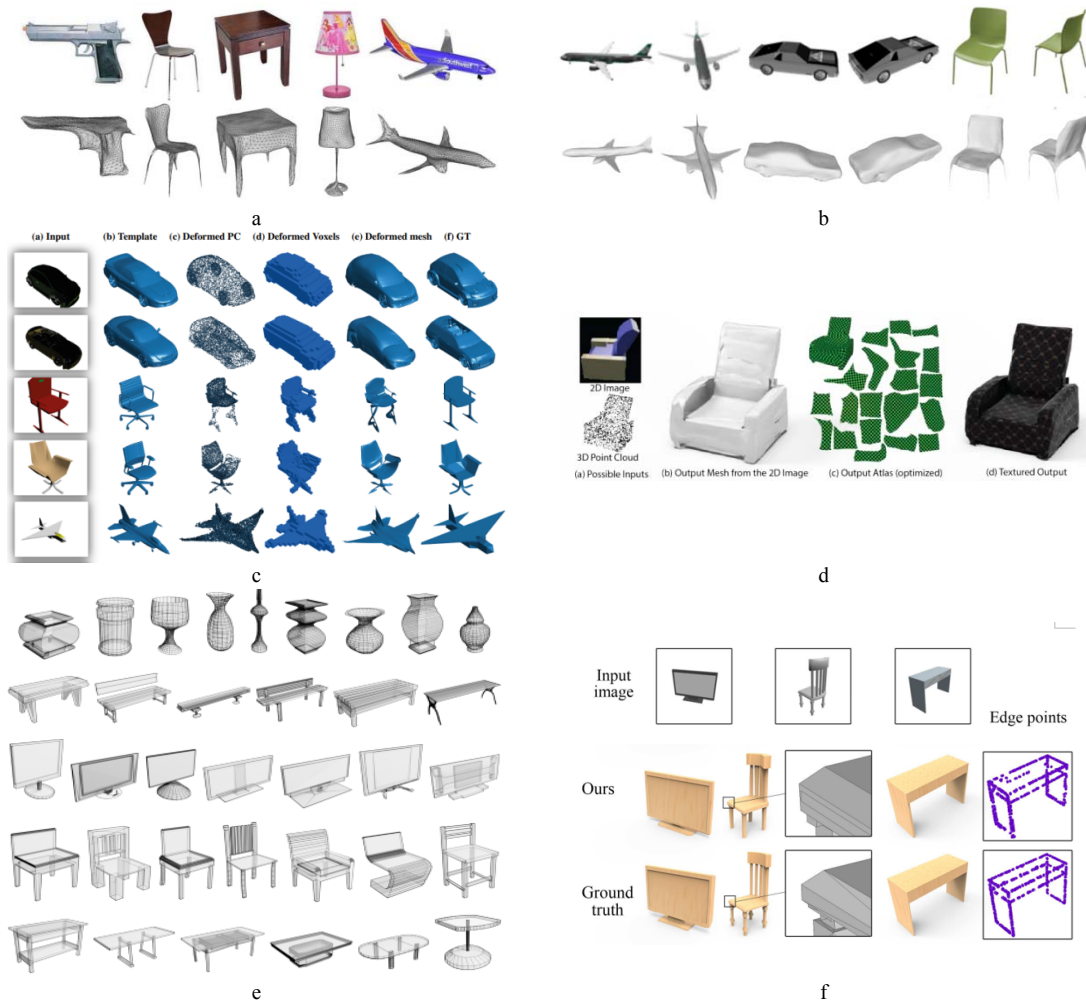


图4 基于网格表示的整体生成方法结果

Fig.4 The results of holistic generative models based on mesh representation

上的精确性。大量的实验表明,该方法不仅定性地生成了具有更好细节的网格模型,而且获得了更高的三维形状精度,其生成结果见图4a。Wen等人^[31]对Pixel2Mesh进行了改进,提出了Pixel2Mesh++,引入了多视角变形网络,将多视角信息融入网格生成过程中以提供更多约束实现更好地重建效果,其生成结果见图4b。

Dominic等人^[53]使用CNN提取目标RGB图像的特征,预测模板Mesh的自由变形(FFD)参数从而生成合理的三维形状。他们使用轻量级CNN来推断多个模板的低维FFD参数,并对拓扑形式不同的模板采用大变形约束来产生合理的形状,其生成结果见图4c。Groueix等人^[54]将一个3D形状看作参数化表面的集合,提出AtlasNet。这种表示方式提高了3D生成模型的精度和泛化能力,并有可能生成任意分辨率的形状,而不会出现内存问题。AtlasNet在自动编码形状、单视角重建,以及超分辨率等方面展示了巨大潜力,但这种方法通常会产生可见的接缝,其生成结果见图4d。

Nash等人^[55]提出了三维网格深度生成模型PolyGen。他们将网格生成看作自回归序列建模问题,并结合Transformer和Pointer网络的优点灵活地建模可变长度的网格序列。PolyGen由两部分组成:一为顶点模型,使用掩码Transformer解码器建模网格顶点;另一为面模型,它在输入顶点的条件下使用带有Pointer网络的Transformer网络建模网格面。这两部分都利用了Transformer架构^[56]以有效地捕捉网格数据中的长期依赖关系。实验证明这种方法能够生成连贯和多样化的网格三维模型,可以直接用于图形学的下游应用中,其生成结果见图4e。Chen等人^[57]基于计算机图形学的经典空间数据结构——二分空间分割,提出了BSP-Net。该模型对空间进行递归细分获得凸集,通过凸分解学习表示三维形状的网络。BSP-Net是无监督的,网络用来学习使用从平面上建立的BSP tree获得的一组凸面来重建形状。由BSP-Net推导出的面集可以很容易地提取出来,形成一个具有任意拓扑的且具有水密性的多边形网格,其生成结果见图4f。

3.1.4 隐函数

相比于以上基于采样的形状整体生成方法而言，使用隐式表示的整体生成形状不受分辨率的影响，引起研究人员关注。隐函数为三维点预测一个有符号的数值，表示该点到形状表面的距离，因此，3D 形状由某一等值面表示。最后，通过后处理如等值面提取进行可视化。基于隐函数表示的整体生成结果见图 5。Chen 等人^[35]提出隐式场解码器 IM-NET。该模型以一个三维点的坐标和形状的特征向量作为输入，通过二分类器输出一个值判断该点是否在形状内。他们将 IM-NET 嵌入到自动编码器 (AEs)、变分自动编码器 (VAEs) 和生成对抗网络 (GAN)，取代了其解码器，得到 IM-AEs 和 IM-GANs。实验结果表明他们所提出的解码器在形状表示学习、3D 形状生成、形状插值以及单视图 3D 形状重建等任务中取得较好的效果，其生成结果见图 5a。Park 等人^[1]引入一种连续的带符号的距离函数 DeepSDF，从有噪声的 3D 输入数据中实现高质量的形状表示、插值和补全。他们用连续的体积场来代表形状的表面，场中点数值的大小表示到表面边界的距离，符号表示该区域是在形状的内部 (-) 还是外部 (+)，从而，隐式地将形状边界表示为所学函数值为零的集合，明确地将空间分为形状内部或非内部，其生成结果见图 5b。

3.2 结构感知生成方法

在实际建模中，复杂的三维形状往往由简单的部件装配得到。由此，越来越多的研究人员模仿这种建模过程，以部件为基本单元，将部件的几何和结构信息集成到深度学习模型中探索基于部件的生成方法。三维形状结构感知生成模型对比见表 3。这类方法将三维形状分解成基本部件或基元 (有方向的 3D 包围盒)，并利用神经网络学习部件之间联系，能够实现部件结构关系和部件几何细节的单独操控和编辑，生成更多细节的真实的 3D 形状。结构感知形状生成方法其粒度是零件级别，生成的形状取决于零件的几

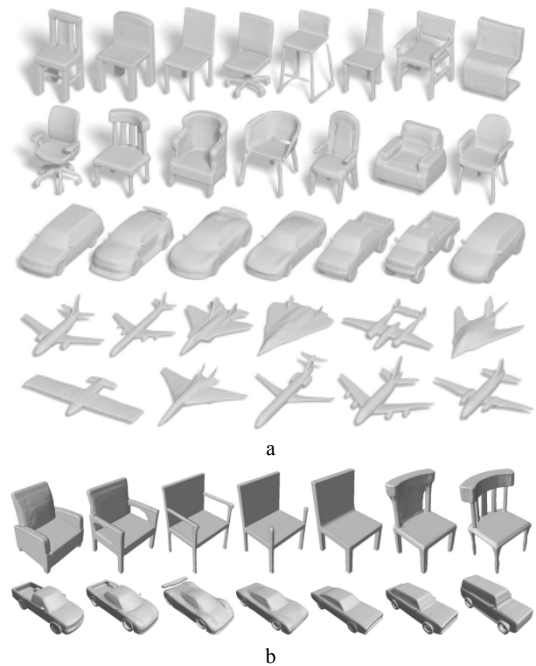


图 5 基于隐函数表示的整体生成结果
Fig.5 The results of holistic generative models based on implicit function representation

何细节、部件间结构关系及零件标注的可用性和质量。对结构的表示大致可以分为线性结构、层次结构和隐式结构。

3.2.1 线性结构

使用线性结构表示部件间关系的研究受自然语言处理和形状生成之间相似性启发。Noam Chomsky^[58]在短语结构规则领域中首先提出句子既是一个线性的单词串，也是一个层次结构，如短语嵌套在短语中。类似单词组成句子，线性结构表示方法将部件类比为单词，将部件间关系类比为单词之间关系，将三维形状的生成看作零件序列装配成“句子”的过程。这类方法使用序列到序列的模型，以部件序列为输入，一次输出一个部件，最终生成完整的形状。序列三维形状生成过程^[59]见图 6。

表 3 三维形状结构感知生成模型对比

Tab.3 The comparison of structure-aware 3D shape generative models

文献	三维表示类型	可控性	评价指标
Zou 等人	3D Primitives	结构感知	Surface-to-Surface Distance, IOU
Wu 等人	Implicit functions	结构感知	COV, MMD, JSD
Li 等人	3D Primitives	结构感知	
Mo 等人	3D Primitives Point cloud	结构感知	Quality, Coverage
Gao 等人	Mesh	结构感知	JSD, MMD-CD, MMD-EMD, COV-CD, COV-EMD
Yang 等人	Mesh	结构感知	Coverage, Quality
Wang 等人	Voxel	结构感知	3D Inception Score, Symmetry Score, Distribution Distance
Wu 等人	Voxel	结构感知	MMD-CD, COV-CD, MMD-EMD, COV-EMD, Overlapping Score, Inception Score
Li 等人	Voxel	结构感知	Symmetry Measures, Inception Score, IOU

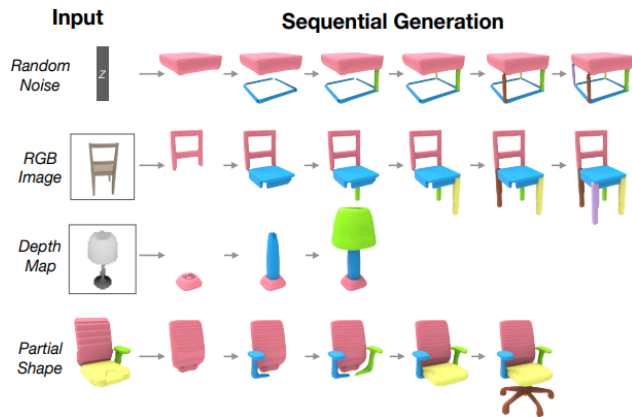


图6 序列三维形状生成过程

Fig.6 The process of sequence 3D shape generation

Li 等人^[60]提出了 PAGENet。该模型将部件几何和结构分别交由不同网络处理,大大简化构建 3D 形状的困难。部件由 VAE-GAN 生成部件的几何细节,然后通过组装模块来预测每个部件的关系,并将它们组装成一个合理的形状。线性结构形状生成结果见图 7。PAGENet 在语义形状分割及基于部件的形状编辑应用中有较好效果,其生成结果见图 7a。Zou 等人^[32]将三维形状采用基元表示(有方向的 3D 包围盒)使得三维形状的表达更加紧凑并且大大简化了三维形状的稳定性和连接性及其他重要属性。他们提出了一种基于高斯场和能量最小化的拟合基元方法得到点云的基元表示,之后将 LSTM 应用于部件生成器中,在以对象为中心的坐标系下,依次生成基元序列,其

生成结果见图 7b。

Wu 等人^[59]进一步提出了一种基于 GRU 的 PQNET 方法,PQNET 的输入是一个分割成多个部件的 3D 形状,其中每个部件首先使用 PartVAE 编码成一个特征表示。这些零件特征序列通过序列到序列自动编码器将其编码成固定大小的潜在向量,之后,使用解码器一次重构一个零件的三维形状,最终实现顺序装配。在该模型中,编码器学习的潜空间既对部件间结构关系又对部件的几何细节进行编码。解码器可用于执行多个生成任务,包括形状插值、新形状生成和单视图 3D 重建,其生成结果见图 7c。

3.2.2 层次结构

层次结构的生成方法认为三维形状的基础部件以树状分层组织,部件的几何和结构一一对应,形状的层次结构表示^[61]见图 8,左侧为部件树表示形状结构,右侧为部件表示三维形状。部件树根节点代表整个 3D 形状,内部节点为部件构成的单元,叶子节点表示构成 3D 形状的基本部件。部件树的父子关系表示该节点的部件组成,同一层次兄弟节点关系表示部件间存在的几何约束如对称、旋转等。这种层次结构的表示方法能够在较大程度上解决纠缠三维形状的几何和结构,能够对生成形状的几何和结构单独控制。

层次结构形状生成结果见图 9。2017 年, Li 等人^[9]将结构关系建模为二叉树提出 GRASS 模型。他们提出一个基于递归神经网络的自动编码器,将任意

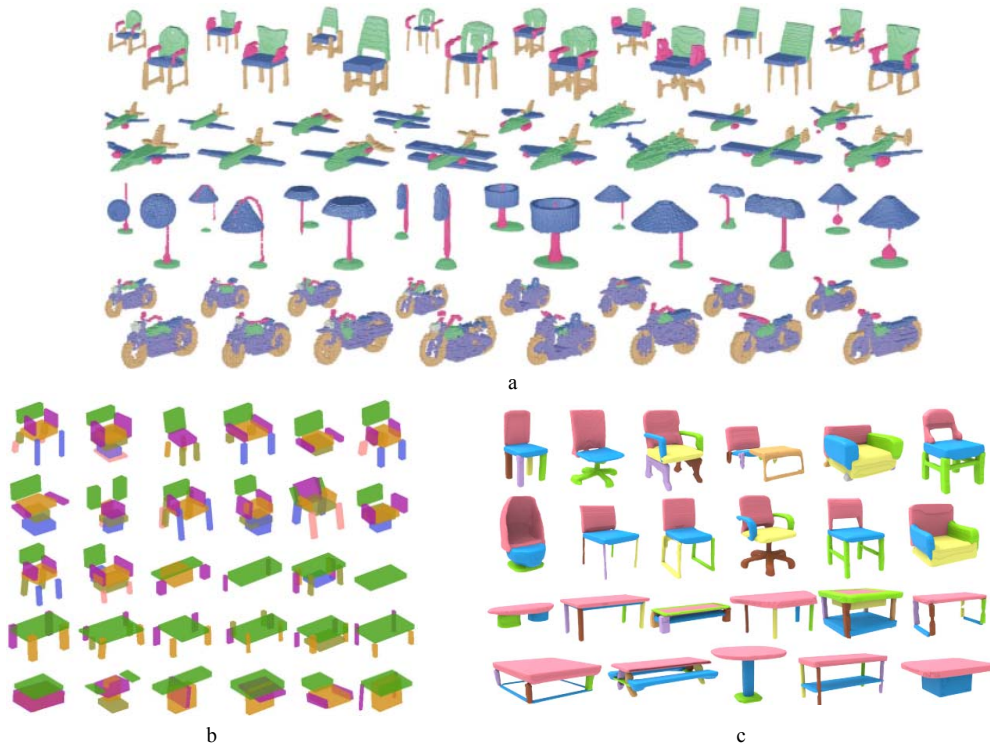


图7 线性结构形状生成结果

Fig.7 The results of 3D shape generation based on sequence structure

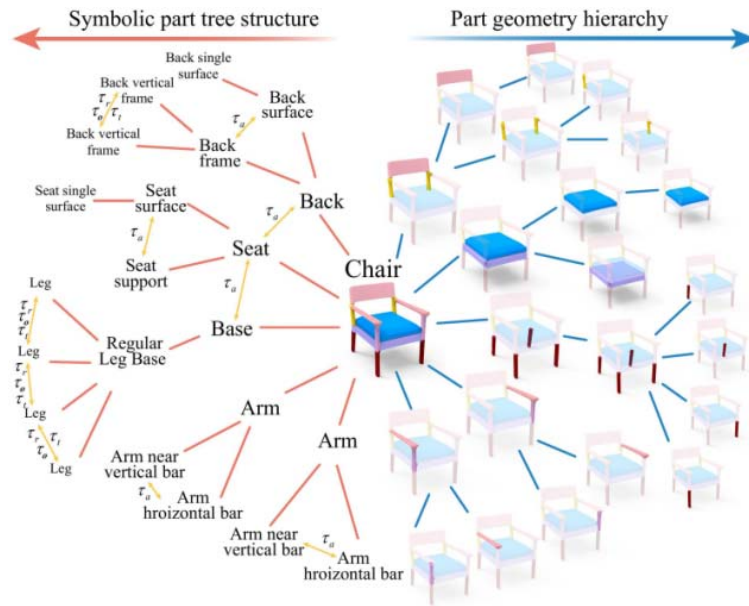
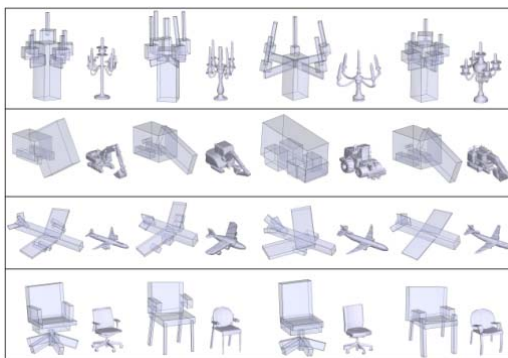


图 8 形状的层次结构表示
Fig.8 The hierarchical representation of 3D shape



a



b



c



d

无标记的部件层次结构映射到潜空间，并利用解码器映射回完整的层次结构，以学习结构的嵌入表示。其生成结果见图 9a。在此之后，Mo 等人^[10]从图神经网络的最新进展中获得灵感，以 n 元层次图结构来组织形状结构，提出 StructureNet。他们通过在编码过程中使用对称函数，实现编码和解码时部件节点排序的不变性。另外，StructureNet 利用部件兄弟节点关系，提出了一个新的基于图的消息传递方法使图神经网络更好地传递部件几何和结构信息。该模型能够实现 3D 形状的结构和几何解耦，完成几何形状生成和插值、形状编辑等结构感知的几何处理应用，其生成结果见图 9b。

Gao 等人^[62]提出了基于两级变分自编码器的深度生成模型 SDM-NET。在零件层面，Part VAE 学习零件几何形状的可变形模型。在结构层面，他们训练 Structured Parts VAE (SP-VAE) 来联合学习形状的零件结构和零件几何形状，确保整体形状结构和表面细节之间的一致性。SP-VAE 中每个部件都以精心设计编码形式进行表示，包括结构信息(是否存在对称、对称平面、支撑关系)和几何特征(部件的特征向量、中心)，其生成结果见图 9c。Yang 等人^[61]认为三维形状中基础部件的结构关系和几何细节应该以一种既解耦又协同的方式进行编码，结构用于指导几何，反之亦然。为此他们提出 DSM-Net。该模型通过变分自动编码器 (VAEs) 以分层的方式同时学习结构关系和几何细节，每一层几何和结构都存在双向映射。通过这种方式，他们有效地在独立的潜在空间中编码几何细节和结构关系，同时确保两者的兼容性。另外，在叶子层，部件几何形状使用有条件的 partVAE 来表示，以结构关系作为条件编码高质量的部件几何细节。DSM-Net 可以有效地表示和生成具有复杂结构和

图 9 层次结构形状生成结果
Fig.9 The results of 3D shape generation based on hierarchical structure

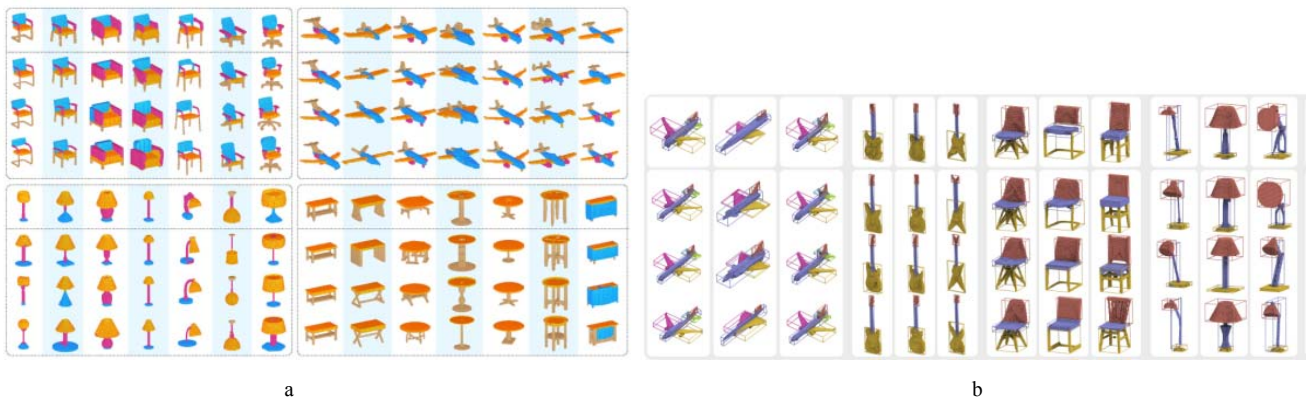


图 10 隐式结构形状生成结果

Fig.10 The results of 3D shape generation based on implicit structure

注：图中第一行为生成结果，第 2—4 行为数据集中与生成结果最接近的 3D 形状

精细几何细节的高质量形状，其生成结果见图 9d。

3.2.3 隐式结构

不同于以上两种明确将部件结构建模为线性或者层次结构的做法，隐式结构生成方法虽然基于结构感知的生成，但并没有直接给出部件间结构的形式，而是通过神经网络学习建模实现。隐式结构形状生成结果见图 10。Wang 等人^[63]根据结构信息由粗到细地生成 3D 形状，提出了 Global-to-Local (G2L) 生成模型。G2L 首先应用生成对抗网络生成带有语义标签的粗糙体素网格，并获得全局尺度上部件的分割和语义标签，然后将部件语义标签和全局尺度分割对应体素一同输入到 Part Refiner 的自动编码器中，利用结构信息的指导逐步优化零件几何细节，这种全局到局部生成模型无论是在形状多样性方面，还是在训练数据分布方面比普通 3DGAN 产生了明显更好的结果，其生成结果见图 10a。Wu 等人^[64]提出 SAGnet。该模型通过自动编码器联合学习部件的几何形状和它们之间的结构关系将其嵌入到潜空间中。其编码器采用两分支的 GRU 编码器分别将结构和几何嵌入到潜在空间中。在编码过程中两个分支之间交换信息，从而学习结构和几何之间的依赖关系，然后通过解码器将潜在空间的特征进行分解并重建为三维模型的几何和结构，其生成结果见图 10b。以上两种方法利用部件的结构信息指导形状生成的过程，以期实现更好的生成结果，这种对结构的表示形式可以分为线性结构和层次结构。

3.3 交互生成方法

交互生成方法主要利用生成三维形状和输入条件之间的对应关系，通过改变输入条件达到交互生成和编辑形状的目的。输入条件可以是草图、文本或者图像等形式。三维形状交互生成方法对比见表 4，交互生成方法结果见图 11。Mo 等人^[65]以端到端方式学习从部件树生成三维点云，提出了基于条件 GAN 的

PT2PC 模型。通过沿着部件树层次结构自上而下和自下而上传递消息，部件树的结构信息条件指导着形状生成过程，最终得到真实多样的点云生成。PT2PC 模型以部件树为条件，改变部件树可以控制生成形状中各部件的结构关系，得到形状各异的三维形状，其生成结果见图 11a。Li 等人^[12]提出了一种无监督球面引导的生成模型 SP-GAN，在不需要任何部件标注的条件下，直接生成复杂、高质量的点云。在 SP-GAN 中，他们使用分布着均匀点的球体作为全局先验在空间上指导生成过程，并在每个球面点上附加作为局部先验的随机潜码来提供局部细节信息。最为重要的是，该模型在球面上的点和生成形状中的点之间建立了隐式的对应关系，改变球面上点的局部先验，相应生成的形状中会进行局部改变，从而实现零件编辑、局部形状插值和多形状零件合成等应用。视觉效果和定量评估都表明 SP-GAN 能够合成具有更多细节和更少的噪声的高质量点云，其生成结果见图 11b。

另外，Lun 等人^[66]提出以单个草图或多个草图作为输入条件以生成三维点云，他们首先通过编码器对草图中的形状信息进行编码，将草图转换为特征向量，之后使用解码器将此特征转换为依赖于三维形状不同视角的深度图和法线图，通过求解融合所有视点深度和法线的优化问题，将多视点的深度和法线图合并到同一个三维形状中。实验结果表明，该方法具有重构可靠性高、输出表面分辨率高、拓扑结构和形状结构保存良好等优点，其生成结果见图 11c。Zong 等人^[11]也尝试通过深度学习解决从专业的 2D 草图生成 3D 形状的问题。他们收集了第一个大型专业草图数据集，其中每个草图都与一个参考 3D 形状配对，总共收集了 1 500 个专业草图，500 个 3D 形状。他们在深度对抗网络中引入注意力损失和全局非线性几何变换以解决草图到 3D 形状的不对称性，从而解决基于专业草图的 3D 模型的难题，使得基于草图的三维形状编辑成为可能，生成结果见图 11d。

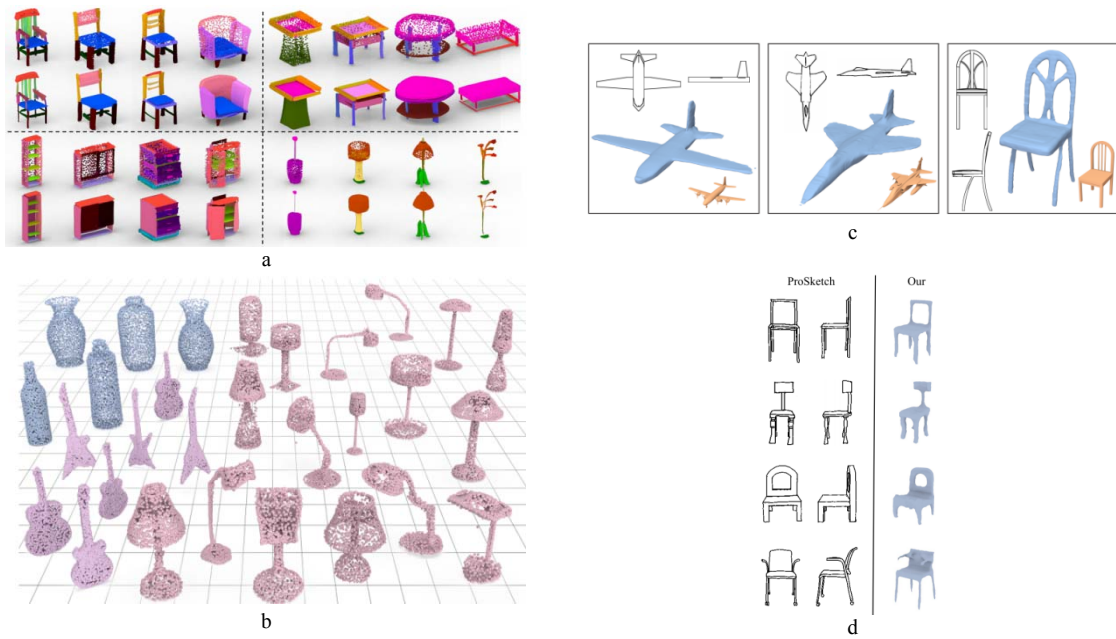


图 11 交互生成方法结果
Fig.11 The results of interactive generation methods

表 4 三维形状交互生成方法对比

Tab.4 The comparison of interactive 3D shape interactive generative methods

文献	三维表示类型	可控性	评价指标
Mo 等人	Point cloud	交互	FPD, Cov, HIS
Li 等人	Point cloud	交互	MMD, COV, FPD
Lun 等人	Mesh	交互	CD, Hausdorff Distance, Surface Normal Distance, Depth Map Error, Volumetric Jaccard Distance
Zong 等人	Mesh	交互	CD, Normal Angle Distance, Hausdorff Distance

4 评价指标

基于深度学习的三维形状生成领域发展离不开评价指标的进步。在训练阶段，评价指标为深度学习模型阶段提供梯度，促进模型收敛；在评价阶段，它能在一定程度上客观量化生成模型的质量。本节总结了常用的三维形状评价指标，将其分为相似性指标、多样性指标和结构一致性指标。相似性指标用于量化生成形状与数据集中已有形状之间的相似程度，多样性用以衡量生成形状与数据集中形状的重叠程度，结构一致性指标用以衡量生成形状结构在结构上的约束。

4.1 相似性指标

4.1.1 Intersection over Union (IoU)

IoU 利用生成体素模型坐标点集合 **A** 与原始的三维模型坐标点集合 **B** 的交集与并集的比值来评价每一个生成模型的精度。由于网络生成的体素模型一般都是概率体素模型，即体素坐标的每一点的值是概率，需要一个阈值 t 判定该点为实体还是空，因此 IoU 值一般还依赖于 t 。IoU 值越大，代表生成体素模型的精度越高。IoU 计算如公式(1)所示：

$$IoU = \frac{A \cap B}{A \cup B} \quad (1)$$

4.1.2 Earth Mover's Distance (EMD)

EMD^[27]是比较无序点集的置换不变性的度量。EMD 越小表示点集之间越接近。对于两个大小相等的子集 $S_1 \subseteq R^3, S_2 \subseteq R^3$ ，EMD 定义如公式(2)所示：

$$d_{EMD}(S_1, S_2) = \min_{\phi: S_1 \rightarrow S_2} \sum_{x \in S_1} \|x - \phi(x)\|_2 \quad (2)$$

其中 ϕ 是一个双向映射。

4.1.3 Chamfer Distance (CD)

CD^[26]度量一个集合中每一点与另一个集合中最近的相邻点之间的距离的平方，CD 是可微的，比 EMD 计算效率更高。CD 值越小表示两个集合中点越接近。CD 计算如公式(3)所示：

$$d_{CH}(S_1, S_2) = \sum_{x \in S_1} \min_{y \in S_2} \|x - y\|_2^2 + \sum_{y \in S_2} \min_{x \in S_1} \|x - y\|_2^2 \quad (3)$$

4.1.4 Jensen-Shannon Divergence (JSD)

JSD^[48]是欧几里得三维空间中定义的边缘分布之间的 Jensen-Shannon 散度。假设点云数据是轴向对齐的，并且在环境空间中有一个典型的体素网格。**A**

的点云与 **B** 的点云的位置相似性计算了 **A** 和 **B** 的每个体素内点云中的点的数量, 并获得经验分布。JSD 越小表示 **A**、**B** 的点云分布距离越近。JSD 计算如公式(4)所示:

$$JSD(P_A \parallel P_B) = \frac{1}{2}D(P_A \parallel M) + \frac{1}{2}D(P_B \parallel M) \quad (4)$$

这里 $M = \frac{1}{2}(P_A + P_B)$, $D(\cdot \parallel \cdot)$ 表示两个分布之间的 KL 散度。

4.1.5 Minimum Matching Distance (MMD)

MMD^[48] 衡量 **A** 和 **B** 之间的保真度, MMD 越小代表 **A** 和 **B** 越接近。计算时将 **B** 的每个点云与 **A** 中的最小距离 (MMD) 点云进行匹配, 并计算匹配中距离的平均值。任意点集间的距离都可以使用, 如使用 CD 和 EMD 可得 MMD-CD 和 MMD-EMD。由于 MMD 直接依赖于匹配的距离, 所以它与 **A** 元素相对于 **B** 的忠实程度密切相关。

4.2 多样性评价指标

4.2.1 Coverage (COV)

Coverage 通过 **B** 中的点云与 **A** 中的点云匹配比例来计算。对于 **A** 中的每一个点, 首先找到它在 **B** 中距离最近的邻居。距离可以用 CD 或 EMD 点集距离来计算。从而产生了两种不同的指标, COV-CD 和 COV-EMD。Coverage 得分高表明 **B** 中的大部分点在 **A** 中有表示。

4.2.2 3D Inception Score (IS)

3D Inception Score^[60] 用来衡量生成形状的多样性。在 2D 中, IS 评分基于在 ImageNet 数据集上预先训练的初始分类网络中分类器的可信度和生成类的方差来衡量生成内容的多样性及其质量。G2L^[63] 中将来自同一类别的训练数据聚类为 10 个聚类, 然后将这些聚类视为不同的类, 训练相应的 3D Inception 分类网络来计算 Inception 得分。

4.2.3 Frechét Point-cloud Distance (FPD)

FPD 由 Shu 等人引入^[67], 其值同时反映生成质量、多样性和覆盖率。其灵感来自用于评价 2D 图像生成质量的常用的 FID。与 2D 不同的是, FPD 首先在 ModelNet40^[41] 上训练一个 PointNet^[24] 进行三维形状分类, 然后利用从 PointNet 中提取的点云全局特征计算真假特征分布之间距离。其计算如公式(5)所示: Frechet Distance = $\mu_r - \mu_g^2 + \text{Tr}(\Sigma_r + \Sigma_g - 2(\Sigma_r \Sigma_g)^{1/2})$ (5)

其中, μ 和 Σ 为真实数据分布 r 和生成的 g 的特征的均值向量和协方差矩阵, Tr 为矩阵迹。

4.3 结构一致性评价指标

4.3.1 HierInsSeg Score (HIS)

HIS 衡量生成的点云满足部件树条件的程度^[65]。

HIS 算法首先利用预先训练的层次实例分割网络 Seg(x) 对输入形状点云 x 进行分层零件实例分割, 并输出描述其生成形状的结构树, 然后计算该结构树与真实结构树之间的距离。

4.3.2 Symmetry Score

由于生成的三维形状中存在对称的性质, G2L^[63] 中采用 Symmetry Score 来衡量生成质量。G2L 只评估镜像对称性, 并将对称性评分定义为给定一个对称平面后, 经过微分变换得到匹配的体素的百分比。因为每个类别中的所有形状位置上都是对齐的, 所以他们为所有训练生成的类别设置了一个固定的对称平面。然后计算每个语义部分和全局形状对应的对称分数。

4.4 其他评价指标

Zhong 等人^[11] 还提出法向角距离 (Normal Angle Distance, NAD) 衡量生成的 3D 形状和真实形状之间每对最近的点的平均法向角距离。Hausdorff 距离 (Hausdorff Distance, HD) 衡量从重建形状上的每个表面点到参考形状上最近的表面点的最大距离。深度误差 (Depth Error, DE) 表示真实和重建的深度图之间的欧氏像素距离。掩码误差 (Mask Error, ME) 是通过预测掩码和真实掩码之间的像素平均误差来计算。

5 面临的挑战

虽然在三维形状方面取得很大的发展, 尤其是深度学习在这个领域逐步发力, 生成高质量形状的算法不断涌现, 但是领域发展还面临着如下限制:

1) 数据集。目前 3D 形状数据集规模仍然较小, 数据集与真实数据之间仍然存在较大差异, 且数据集中形状结构关系和几何细节较为简单。建立复杂三维形状、大场景和多目标的三维生成场景数据集仍然有较大挑战。

2) 形状表示方法。现有研究方法中, 三维形状通常采用体素、点云、网格等基于采样的方法表达三维形状, 但这些方法受分辨率的限制, 导致生成形状对细节表现不足。隐函数的表示方法虽然不受分辨率的影响, 但是表示复杂、不够直观。因此探索更合适的形状的表达方法, 以更简单、更高效的数据结构表示形状仍需要继续探索。

3) 智能生成方法。基于深度学习的三维形状生成除了依赖高质量的大规模数据之外, 损失函数和网络结构也直接关系着生成的三维形状的质量。较好的网络结构能够更好地利用有效特征, 学习形状的几何和结构解耦, 促进生成结果可控。良好的损失函数在学习中提供梯度和优化的方向使模型更稳定快速收敛。现有模型依然存在表现力不足、生成三维形状质量低、无法全面控制生成结果和训练时不收敛或收敛慢的问题。因此探索更先进的智能生成方法是未来研

究的重要方向之一。

4) 评价指标。虽然研究者已经在评价指标方面精心设计出几种客观的相似性度量, 但是指标结果与直观的感受仍然相差较大。另外, 设计师一般是先根据功能来构思设计的, 同时设计的成果尽可能地符合人因、美学等标准。现有评价指标大都关注生成形状的精确度, 而对生成形状的人因、美学或者功能方面关注较少。因此探索更加符合人类判断标准的尤其是人因、美学、功能等因素的评价指标仍然面临着挑战。

6 结语

本文基于深度学习在三维形状生成方向的研究上进行了分析和总结。首先总结了现有三维形状生成方法中三维模型的表现形式, 并分析其优缺点。其次根据智能生成的可控性, 从整体生成、结构感知和交互 3 个方面总结了目前所取得的研究成果。然后, 总结了现有方法中对生成的三维形状的评价指标。最后对目前该方向所面临的挑战和未来的研究方向进行了展望。

参考文献:

- [1] PARK J J, FLORENCE P, STRAUB J, et al. DeepSDF: Learning Continuous Signed Distance Functions for Shape Representation[C]. Long Beach: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019.
- [2] 万群. 基于生成对抗网络的 3D 目标重建研究[D]. 北京: 北京交通大学, 2020.
WAN Qun. 3D Object Reconstruction with Generative Adversarial Networks[D]. Beijing: Beijing Jiaotong University, 2020.
- [3] RIEGLER G, OSMAN ULUSOY A, GEIGER A. Octnet: Learning Deep 3D Representations at High Resolutions[C]. Honolulu: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017.
- [4] WU J, ZHANG C, XUE T, et al. Learning a Probabilistic Latent Space of Object Shapes Via 3D Generative-adversarial Modeling[C]. Barcelona: Proceedings of the 30th International Conference on Neural Information Processing Systems, 2016.
- [5] GIRDHAR R, FOUHEY D F, RODRIGUEZ M, et al. Learning a Predictable and Generative Vector Representation for Objects[C]. Amsterdam: European Conference on Computer Vision, 2016.
- [6] ARSALAN SOLTANI A, HUANG H, WU J, et al. Synthesizing 3D Shapes Via Modeling Multi-view Depth Maps and Silhouettes with Deep Generative Networks[C]. Honolulu: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017.
- [7] HUANG H, KALOGERAKIS E, MARLIN B. Analysis and Synthesis of 3D Shape Families Via Deep-learned Generative Models of Surfaces[J]. Computer Graphics Forum, 2015(1): 25-38.
- [8] [NASH C, WILLIAMS C K. The Shape Variational Autoencoder: a Deep Generative Model of Part-segmented 3D Objects[J]. Computer Graphics Forum, 2017(1): 1-12.
- [9] LI J, XU K, CHAUDHURI S, et al. Grass: Generative Recursive Autoencoders for Shape Structures[J]. ACM Transactions on Graphics (TOG), 2017, 36(4): 1-14.
- [10] MO K, GUERRERO P, YI L, et al. Structrnet: Hierarchical Graph Networks for 3D Shape Generation [J]. arXiv Preprint arXiv, 2019, 8: 575.
- [11] ZHONG Y, QI Y, GRYADITSKAYA Y, et al. Towards Practical Sketch-based 3D Shape Generation: the Role of Professional Sketches[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2020, 31(9): 3518-3528.
- [12] LI R, LI X, HUI K-H, et al. SP-GAN: Sphere-guided 3D Shape Generation and Manipulation[J]. ACM Transactions on Graphics (TOG), 2021, 40(4): 1-12.
- [13] WANG D, CUI X, CHEN X, et al. Multi-view 3D Reconstruction with Transformer[J]. arXiv Preprint arXiv, 2021, 3(12): 957.
- [14] HE K, ZHANG X, REN S, et al. Deep Residual Learning for Image Recognition[C]. Las Vegas: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016.
- [15] SIMONYAN K, ZISSERMAN A. Very Deep Convolutional Networks for Large-scale Image Recognition[J]. arXiv Preprint arXiv, 2014, 9: 1556.
- [16] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative Adversarial Nets[J]. Advances in Neural Information Processing Systems, 2014, 27: 10.
- [17] LIAO Y, DONNE S, GEIGER A. Deep Marching Cubes: Learning Explicit Surface Representations[C]. Salt Lake City: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018.
- [18] LORENSEN W E, CLINE H E. Marching Cubes: a High Resolution 3D Surface Construction Algorithm[J]. ACM Siggraph Computer Graphics, 1987, 21(4): 163-169.
- [19] HÄNE C, TULSIANI S, MALIK J. Hierarchical Surface Prediction for 3D Object Reconstruction[C]. Qingdao: 2017 International Conference on 3D Vision (3DV), 2017.
- [20] TATARCHENKO M, DOSOVITSKIY A, BROX T. Octree Generating Networks: Efficient Convolutional Architectures for High-resolution 3d Outputs[C]. Venice: Proceedings of the IEEE International Conference on Computer Vision, 2017.
- [21] WANG P-S, LIU Y, GUO Y-X, et al. O-cnn: Octree-based Convolutional Neural Networks for 3D Shape Analysis[J]. ACM Transactions On Graphics (TOG), 2017, 36(4): 1-11.
- [22] WANG P-S, SUN C-Y, LIU Y, et al. Adaptive O-CNN: a

- Patch-based Deep Representation of 3D Shapes[J]. *ACM Transactions on Graphics (TOG)*, 2018, 37(6): 1-11.
- [23] RIEGLER G, ULUSOY A O, BISCHOF H, et al. Octnetfusion: Learning Depth Fusion from Data[C]. Qingdao: 2017 International Conference on 3D Vision (3DV), 2017.
- [24] QI C R, SU H, MO K, et al. Pointnet: Deep Learning on Point Sets for 3D Classification and Segmentation[C]. Honolulu: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017.
- [25] QI C R, YI L, SU H, et al. Pointnet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space[J]. *arXiv Preprint arXiv*, 2017, 6(2): 413.
- [26] FAN H, SU H, GUIBAS L J. A Point Set Generation Network for 3D Object Reconstruction from a Single Image[C]. Honolulu: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017.
- [27] RUBNER Y, TOMASI C, GUIBAS L J. The Earth Mover's Distance as a Metric for Image Retrieval[J]. *International Journal of Computer Vision*, 2000, 40(2): 99-121.
- [28] 黄骅. 多视角的三维形状风格转换[D]. 大连: 大连理工大学, 2021.
HUANG Ye. Multi-view 3D Shape Style Transformation[D]. Dalian: Dalian University of Technology, 2021.
- [29] BRONSTEIN M M, BRUNA J, LECUN Y, et al. Geometric Deep Learning: Going Beyond Euclidean Data[J]. *IEEE Signal Processing Magazine*, 2017, 34(4): 18-42.
- [30] WANG N, ZHANG Y, LI Z, et al. Pixel2mesh: Generating 3D Mesh Models from Single RGB Images[C]. Munich: Proceedings of the European Conference on Computer Vision (ECCV), 2018.
- [31] WEN C, ZHANG Y, LI Z, et al. Pixel2mesh++: Multi-view 3D Mesh Generation Via Deformation[C]. Seoul: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019.
- [32] ZOU C, YUMER E, YANG J, et al. 3D-prnn: Generating Shape Primitives with Recurrent Neural Networks[C]. Venice: Proceedings of the IEEE International Conference on Computer Vision, 2017.
- [33] NIU C, LI J, XU K. Im2struct: Recovering 3D Shape Structure from a Single RGB Image[C]. Salt Lake City: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018.
- [34] ZHU C, XU K, CHAUDHURI S, et al. SCORES: Shape Composition with Recursive Substructure Priors[J]. *ACM Transactions on Graphics (TOG)*, 2018, 37(6): 1-14.
- [35] CHEN Z, ZHANG H. Learning Implicit Fields for Generative Shape Modeling[C]. Long Beach: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019.
- [36] MESCHEDER L, OECHSLE M, NIEMEYER M, et al. Occupancy Networks: Learning 3D Reconstruction in Function Space[C]. Long Beach: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019.
- [37] SILBERMAN N, FERGUS R. Indoor Scene Segmentation Using a Structured Light Sensor[C]. Barcelona: 2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops), 2011.
- [38] SILBERMAN N, HOIEM D, KOHLI P, et al. Indoor Segmentation and Support Inference from RGB-D Images[C]. Florence: European Conference on Computer Vision, 2012.
- [39] LEVIN A, LISCHINSKI D, WEISS Y. Colorization Using Optimization[M]. *ACM SIGGRAPH 2004 Papers*, 2004.
- [40] DAI A, CHANG A X, SAVVA M, et al. Scannet: Richly-annotated 3d Reconstructions of Indoor Scenes[C]. Honolulu: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017.
- [41] WU Z, SONG S, KHOSLA A, et al. 3D Shapenets: a Deep Representation for Volumetric Shapes[C]. Boston: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015.
- [42] CHANG A X, FUNKHOUSER T, GUIBAS L, et al. Shapenet: an Information-rich 3D Model Repository[J]. *arXiv Preprint arXiv*, 2015, 12(3): 12.
- [43] MO K, ZHU S, CHANG A X, et al. Partnet: a Large-scale Benchmark for Fine-grained and Hierarchical Part-level 3D Object Understanding[C]. Long Beach: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019.
- [44] BROCK A, DONAHUE J, SIMONYAN K. Large Scale GAN Training for High Fidelity Natural Image Synthesis[J]. *arXiv Preprint arXiv*, 2018, 9(11): 96.
- [45] KARRAS T, LAINE S, AILA T. A Style-based Generator Architecture for Generative Adversarial Networks[C]. Long Beach: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019.
- [46] CHEN S-Y, SU W, GAO L, et al. DeepFaceDrawing: Deep Generation of Face Images from Sketches[J]. *ACM Transactions on Graphics (TOG)*, 2020, 39(4): 72: 1-16.
- [47] SHARMA A, GRAU O, FRITZ M. Vconv-dae: Deep Volumetric Shape Learning without Object Labels[C]. Amsterdam: European Conference on Computer Vision, 2016.
- [48] ACHLIOPTAS P, DIAMANTI O, MITLIAGKAS I, et al. Learning Representations and Generative Models for 3D Point Clouds[C]. Jinan: International Conference on Machine Learning, 2018.
- [49] CAI R, YANG G, AVERBUCH-ELOR H, et al. Learning Gradient Fields for Shape Generation[C]. Glasgow: 16th European Conference, 2020.

- [50] SUN Y, WANG Y, LIU Z, et al. Pointgrow: Autoregressively Learned Point Cloud Generation with Self-attention[C]. Snowmass Village: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2020.
- [51] YANG G, HUANG X, HAO Z, et al. Pointflow: 3D Point Cloud Generation with Continuous Normalizing Flows[C]. Seoul: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019.
- [52] RAMASINGHE S, KHAN S, BARNES N, et al. Spectral-gans for High-resolution 3D Point-cloud Generation[C]. Las Vegas: 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2020.
- [53] JACK D, PONTES J K, SRIDHARAN S, et al. Learning Free-form Deformations for 3D Object Reconstruction[C]. Perth: Asian Conference on Computer Vision, 2018.
- [54] GROUEIX T, FISHER M, KIM V, et al. AtlasNet: a Papier-Mâché Approach to Learning 3D Surface Generation[J]. arXiv Preprint arXiv, 2018, 2(5): 384.
- [55] NASH C, GANIN Y, ESLAMI S A, et al. Polygen: an Autoregressive Generative Model of 3D Meshes[C]. Vienna: International Conference on Machine Learning, 2020.
- [56] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is All You Need[C]. Long Beach: Advances in Neural Information Processing Systems, 2017.
- [57] CHEN Z, TAGLIASACCHI A, ZHANG H. Bsp-net: Generating Compact Meshes Via Binary Space Partitioning[C]. Seattle: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020.
- [58] GAZDAR G. Phrase Structure Grammar[M]. Berlin: Springer, 1982.
- [59] WU R, ZHUANG Y, XU K, et al. Pq-net: a Generative Part Seq2seq Network for 3D Shapes[C]. Seattle: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020.
- [60] LI J, NIU C, XU K. Learning Part Generation and Assembly for Structure-aware Shape Synthesis[C]. New York: Proceedings of the AAAI Conference on Artificial Intelligence, 2020.
- [61] YANG J, MO K, LAI Y-K, et al. DSM-Net: Disentangled Structured Mesh Net for Controllable Generation of Fine Geometry[J]. CoRR, 2020, 20(8): 544.
- [62] GAO L, YANG J, WU T, et al. SDM-NET: Deep Generative Network for Structured Deformable Mesh[J]. ACM Transactions on Graphics (TOG), 2019, 38(6): 1-15.
- [63] WANG H, SCHOR N, HU R, et al. Global-to-local Generative Model for 3d Shapes[J]. ACM Transactions on Graphics (TOG), 2018, 37(6): 1-10.
- [64] WU Z, WANG X, LIN D, et al. Sagnet: Structure-aware Generative Network for 3D-shape Modeling[J]. ACM Transactions on Graphics (TOG), 2019, 38(4): 1-14.
- [65] MO K, WANG H, YAN X, et al. PT2PC: Learning to Generate 3D Point Cloud Shapes from Part Tree Conditions[C]. Glasgow: European Conference on Computer Vision, 2020.
- [66] LUN Z, GADELHA M, KALOGERAKIS E, et al. 3D Shape Reconstruction from Sketches Via Multi-view Convolutional Networks[C]. Qingdao: 2017 International Conference on 3D Vision (3DV), 2017.
- [67] SHU D W, PARK S W, KWON J. 3D Point Cloud Generative Adversarial Network Based on Tree Structured Graph Convolutions[C]. Seoul: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019.