

浙江大学

硕士研究生读书报告



题目 StructureNet: Hierarchical Graph Networks
for 3D Shape Generation 读书报告

作者姓名 沈佳军

作者学号 22351163

指导教师 李启雷

学科专业 软件工程

所在学院 软件学院

提交日期 二〇二四年一月六日

StructureNet: Hierarchical Graph Networks for 3D Shape
Generation^[1]

A Dissertation Submitted to
Zhejiang University
in partial fulfillment of the requirements for
the degree of
Master of Engineering

Major Subject: Software Engineering

Advisor: Li Qilei

By

Shen Jiajun

Zhejiang University, P.R. China

2024.01.06

摘要

生成创新、多样且真实的三维形状以及相关的部分语义和结构对于许多需要高质量三维资产或是大量真实训练数据的应用十分重要。实现该目标中的一个关键挑战是如何适应多样的形状变化,包括部件的连续变形以及结构或离散的变化,这些变化可以增加、删除或修改形状成分和组成结构。这种物体结构通常可以组织成组成对象部件和关系的层次结构,表示为 n 元图的层次结构。作者引入了 StructureNet,这是一种分层图网络,它(1)能够直接编码由这种 n 元图所表示的形状;(2)可以在大量且复杂的形状类别上稳健地进行训练;(3)能够用于生成多种真实的结构化几何形状。从技术上讲,作者从图神经网络的最新进展中汲取灵感,提出一种 n 元图的顺序不变编码,并在网络训练期间共同考虑部件几何形状和部件间关系来实现这一目标。作者广泛评估了各种形状类别的学习潜在空间的质量,并显示出相对于基线和竞争方法的显著优势。学习到的潜在空间支持各种结构感知的几何处理应用,包括形状生成和插值、形状编辑,或是直接从无标注的图像、点云或部分扫描中发现形状结构。

关键词: 三维形状生成, 层次表示, 图神经网络

Abstract

The ability to generate novel, diverse, and realistic 3D shapes along with associated part semantics and structure is central to many applications requiring high-quality 3D assets or large volumes of realistic training data. A key challenge towards this goal is how to accommodate diverse shape variations, including both continuous deformations of parts as well as structural or discrete alterations which add to, remove from, or modify the shape constituents and compositional structure. Such object structure can typically be organized into a hierarchy of constituent object parts and relationships, represented as a hierarchy of n-ary graphs. This paper introduces StructureNet, a hierarchical graph network which (i) can directly encode shapes represented as such n-ary graphs; (ii) can be robustly trained on large and complex shape families; and (iii) can be used to generate a great diversity of realistic structured shape geometries. Technically, this paper accomplishes this by drawing inspiration from recent advances in graph neural networks to propose an order-invariant encoding of n-ary graphs, considering jointly both part geometry and inter-part relations during network training. This paper extensively evaluates the quality of the learned latent spaces for various shape families and show significant advantages over baseline and competing methods. The learned latent spaces enable several structure-aware geometry processing applications, including shape generation and interpolation, shape editing, or shape structure discovery directly from un-annotated images, point clouds, or partial scans.

Keywords: 3D shape generation, hierarchical representations, graph neural networks

1 引言

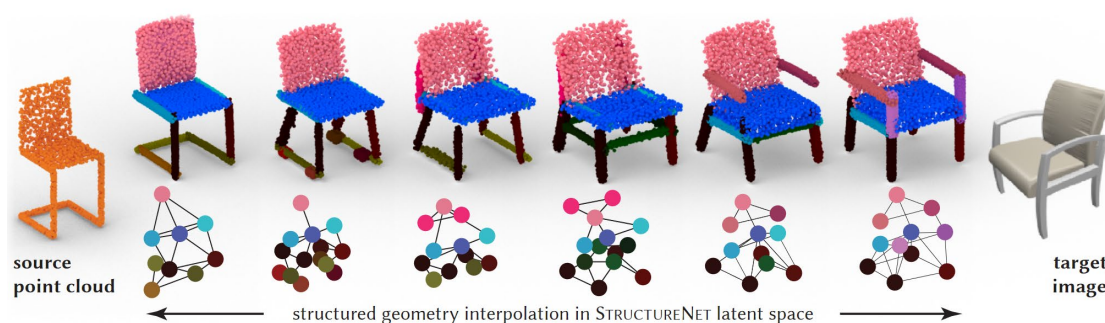


图 1 StructureNet^[1]结果示意图

三维形状分析和生成中长期存在的一个问题是如何构建一个生成模型来支持创建新的、多样的且真实的形状。一个关键的挑战是适应多样的形状变化，包括部件的连续变形以及结构或离散的变化，这些变化可以增加、删除或修改现有形状的子结构。作者寻找一个连续的隐空间来融合所有这些多样性，并能够编码如下形状：有或没有扶手的椅子，有四个腿或旋转底座的椅子，以及高或低的靠背，细或粗的腿等。这种隐空间反过来又使许多重要的应用成为可能，包括生成具有新颖结构和几何的形状，通过将原始未标注的图像或点云投影到学习到的隐空间中来发现物体的结构，以结构感知的方式操纵形状等。

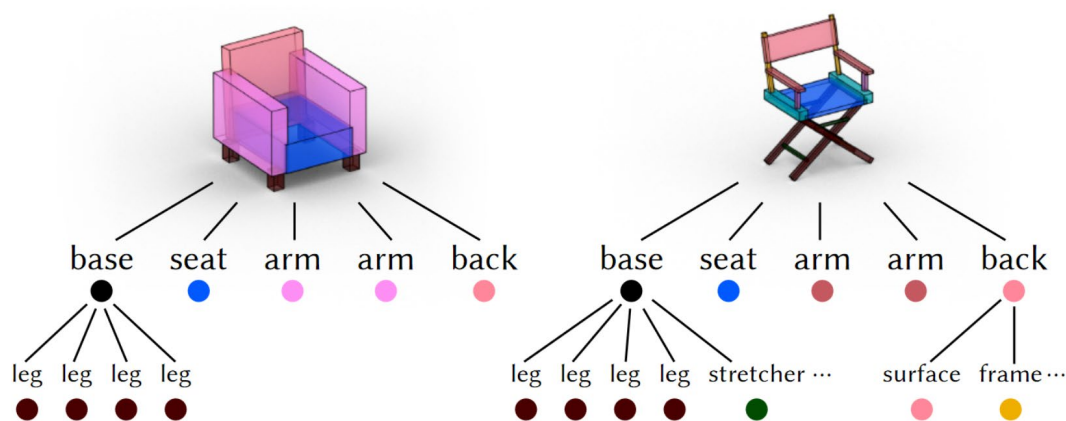


图 2 n 元部件层次结构

实现该目标的一种途径是将形状表示成由一系列部件组成的结构化的物体，这些部件是通过部件层次的连接以及部件内部的联系来组织的。此外，这些部分可以自然地组织成编码为 n 叉树的层次结构，来自相同形状类别的对象共享相似的层次结构（见图 2）。然而，值得注意的是，三维形状的几何中许多语义上重要的关系（例如对称性）可以连接层次结构中可能在空间上分离的遥远节点。这些水平关系对形状编码器造成了额外的限制。作者将这种具有水平连接的分层 n 元

树称为 n 元图的层次结构，或简称为图的层次结构。现在能够访问大量 3D 数据（例如 Turbosquid、3D Warehouse），借助于这些数据集中对部件标注的重要工作，为从数据中学习潜在形状空间提供了可能。以 ShapeNet 为例，粗粒度和细粒度的部件注释都可用。

创建此类生成模型的一项著名工作是 GRASS^[2] 框架。受到自然语言处理中引入的用于编码二叉树的递归神经网络的启发，GRASS 进一步将部件级物体层次结构细化为二叉树，然后递归地利用编码器-解码器网络建立一个潜在空间，可以从中解码叶子结点的层次结构和几何形状。然而，由于二元约束，GRASS 必须另外搜索物体中发现的 n 元层次结构的可能二值化，以便二值化版本在相同形状类别的物体之间保持一致。虽然这在中小型数据集上效果很好，但由于寻找规范二叉树表示的任务变得越来越具有挑战性，因此该设置很难在大型到非常大的形状类别（例如 PartNet）上进行训练。

因此作者提出了 StructureNet，这是一种分层图网络，它直接编码更通用的图，其中父结点拥有可变数量的子节点以及兄弟结点之间的水平关系。StructureNet 依赖于三个主要创新：首先，通过直接使用物体结构的 n 元图，作者从根本上避免了二值化引入的不必要的数据变化，从而可以显著简化学习任务。其次，通过在编码期间使用对称函数（例如最大池化），并解决线性分配问题以在解码期间建立对应关系，作者在编码和解码时实现了部件级同级排序的不变性。最后，作者采用了一种新的基于图的消息传递协议来利用水平部件之间的关系边。这些功能使得作者能够在大型到非常大的形状类别上稳健地训练 StructureNet，从而有效地捕获结构和几何变化。

为结构化形状构建这样的潜在空间具有多种优点，可以被各种应用所利用。首先，结构（即部件层次结构和部件间关系）本身对于下游应用程序很有用。例如，编辑者可以单独编辑、交换或建模部件，并利用对称性等结构约束。其次，形状类别内的结构通常比几何更一致。例如，椅子通常具有粗略层次结构的座椅、靠背和底座，即使这些部件的几何形状存在很大差异（见图 2）。最后，将原始非结构化形状（例如图像、点云、部件扫描）投影到此类潜在空间的能力会自动在原始输入上引入结构（即提供分层部件分割、捕获部件级接触和/或对称性关系），并随后实现一系列不同的结构感知操作。

作者提出的分层图网络支持广泛的应用，可分为以下几类：（1）抽象原始输

入, 包括点云、图像、部件扫描以获得其结构; (2) 基于一组训练形状以参数化形式或点云形式创建新颖的形状; (3) 允许在源形状和目标形状之间进行结构感知插值, 同时显示拓扑和几何变化; 最后 (4) 通过结构感知对象操作来智能地修改部件或替换部件。例如, 在图 1 中, 对输入源点云和目标图像进行独立抽象, 然后直接插值到从 **PartNet** 数据集的 5K 椅子上学习到的潜在空间中。

作者的贡献包括如下: (1) 引入一种用于 n 元部件图的形状层次结构的编码, 该编码足够通用, 足以允许在类别内对形状进行一致的层次表示。(2) 学习编码和解码兄弟结点部件之间在层次图中表示为边的丰富几何关系 (例如, 邻接、对称), 并使用图神经网络来约束现实形状的生成; (3) 开发生成模型 **StructureNet**, 允许对具有多种有效几何和结构变化的盒结构和点云进行形状合成; (4) 说明 **StructureNet** 在形状分析和生成任务中的使用, 包括形状重建、新颖形状生成、结构感知形状插值、未标注点云或图像的抽象以及各种形状编辑应用。

2 概述

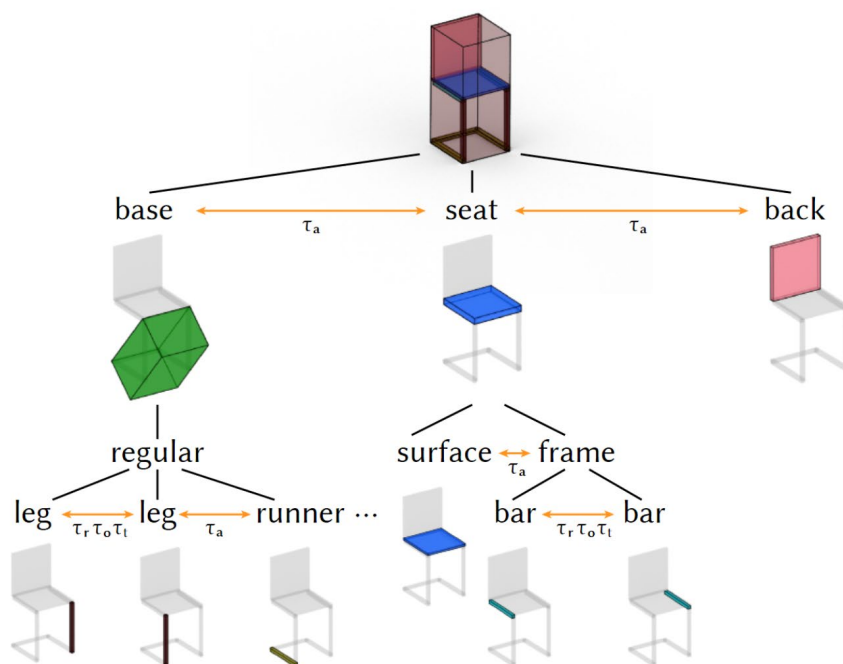


图 3 三维形状的层次图表示

作者将形状部件的层次结构表示为 n 叉树, 其中每个节点都是部件或部件组件。部件的几何关系 (例如对称性和邻接性) 由树的兄弟结点之间的附加边捕获, 从而在兄弟结点之间形成图 (见图 3)。每个节点都包含有关部件几何形状的信息, 捕获形状的总体几何形状。在给定的类别中, 例如椅子, 形状往往具有一致的部件层次结构, 如 **PartNet** 数据集。例如, 大多数椅子在最高层次上自然地分

解为靠背、座椅和底座。作者的形状表示中的 n 叉树可以直接捕获这种层次结构，从而为作者提供在给定类别的形状之间具有高度一致性的形状表示。

这种形状表示（包括结构和几何形状）通过变分自编码器（VAE）映射到潜在空间。作者为编码器和解码器引入了分层图网络，它们是递归神经网络，在每个递归级别执行图卷积。

3 形状结构的层次图

作者引入了一种形状表示，它可以捕获形状的几何形状和结构，并且适合由作者的分层图网络进行处理。作者假设使用的形状可以分解为一组有意义的部分。一个形状 $S = (P, H, R)$ 是由一组部件 $P = \{P_1, \dots, P_N\}$ 表示来描述形状的几何，结构 (H, R) 描述这些部件是如何组织和相互联系的。该结构由两个重叠的图组成：形状分解成部件的层次分解 H 和部件之间的一组几何关系 R 。图 3 给出了示例。

3.1 部件表示

在作者的实验中，部件 P_i 的几何形状支持两种可选的表示。或者，作者使用最小有向包围盒 $B_i = (c_i, q_i, r_i)$ 来表示部件的几何，其中 $c \in R^3$ 是盒子中心的世界坐标， $q \in H$ 是编码为四元数的盒子的方向， $r \in R^3$ 是盒子的尺寸；或者，作者使用对应的点云 $A_i = \{x_1, \dots, x_k\}$ 来表示部件，其中 $x \in R^3$ 是点的世界坐标。除了几何形状之外，每个部件还有一个语义标签 l_i ，例如椅背、座椅或基座，这些标签在同一类别的形状之间是一致的。

3.2 层次分解

分层分解从整个形状作为根结点开始，它被分成一系列组成部分，如座椅、靠背和椅子底座。然后将这些部分递归分解为它们的组成部分，直到到达叶子结点上最细粒度的部分。作者用树 (P, H) 来表示这种分解，其中 P 表示形状部分， $H \in P^2$ 表示父部件结点到所有组成它的子部件结点的有向边。注意一个结点可能有任意数量的子结点，并且该树不需要是平衡树，即从一个结点到所有叶子结点的路径长度可以有很大的不同。

3.3 几何联系

除了层次结构中的垂直组合之外，部件还可以通过邻接和/或对称等关系“水平”关联。这些关系可能是形状的关键特征。例如，椅子通常表现出反射对称性，以及多个部件之间的相邻关系，如果不考虑这些关系，通常会导致生成不切实际

的椅子。另一方面，考虑到层次结构中所有部件对之间的潜在关系，将会需要数量级为 $O(N^2)$ 的关系，其中 N 是部件总数，并且准确地自动编码如此大的一组关系会带来很大的困难。然而，作者在实验中发现，没有必要对所有关系进行编码，因为最重要的关系发生在层次结构的兄弟结点之间（例如，椅子腿、柜子里的抽屉、沙发的扶手），而层次结构的其他部分之间的关系通常不那么重要，或者通过层次结构树后面的关系链来间接暗示。因此，作者选择只捕获层次结构中兄弟结点之间的几何关系，从而显著简化了关系图。

作者用父结点 P_i 的子结点 C_i 中的兄弟结点 $\{P_j, P_k\}$ 之间的附加无向边 R_i 来表示这些关系。作者将这些边表示为 $(\{P_j, P_k\}, \tau)$ 。每条边从可能的关系类型 T 的列表中有一个关联的关系类型 τ 。作者的实验中使用四种关系类型：邻接、反射对称、旋转对称和平移对称。这些边在兄弟结点之间形成一个图 (C_i, R_i) ，该层次结构有效地变成了一个图的层次结构，其中每个形状部分在下一个更低的层次上扩展成一个图。作者称这些图为 **n 元图**，以强调这些图的层次结构的 **n 元** 性质。

4 层次图网络

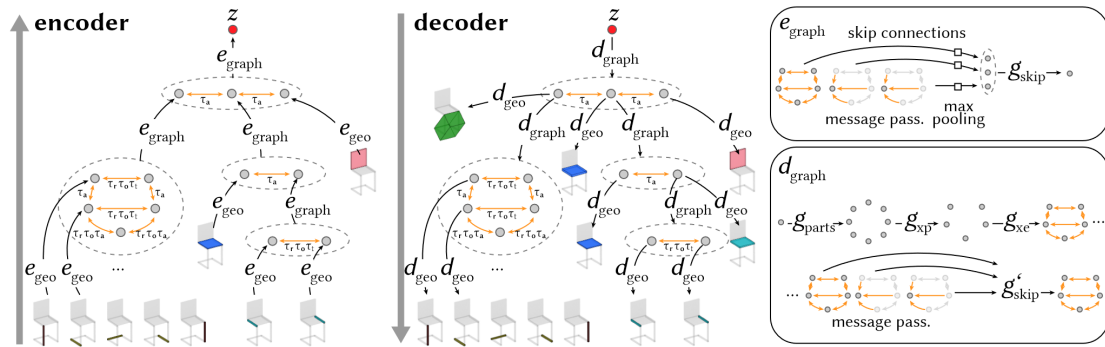


图 4 层次图网络

作者为上述的形状表示提出了一种新颖的变分自编码器（VAE）网络，它可用于生成、插值以及其他应用。作者的 VAE 由一个编码器 e 组成，它将一个形状 S 映射到一个潜在的特征向量 $z = e(S)$ ，以及一个解码器 d ，它将特征向量映射回一个形状，因此 $d(e(S)) \approx S$ 。作者基于几何和结构相似性来测量这个近似的质量。作者引入 **StructureNe** 或分层图网络作为编码器和解码器的一种新的网络架构，它可以有效地编码和解码形状表示的几何形状和结构。图 4 显示了网络体系结构。

4.1 编码器

编码器将表示为 **n 元图** 层次结构的形状映射到潜在特征向量 z 。作者将特征

空间的维数设置为 256，即 $z \in R^{256}$ 。树 (P, H) 中的每个（叶子或中间）结点 i 也被映射到一个特征向量 $f_i \in R^{256}$ 。完整形状的代码 z 就是描述根结点 $z = f_1$ 的特征向量。编码器使用两种类型的编码器以自下而上的方式递归地工作（见图 4）。首先，作者使用几何编码器计算叶节点的特征向量。然后，作者使用图编码器对中间节点进行编码。这种方法的递归性质类似于之前在结构编码方面的工作，又受到了自然语言处理的启发。然而，由于需要编码 n 元图，这可能额外具有可变数量的部分，因此作者的编码器和解码器的体系结构和之前的工作明显不同。

4.1.1 几何编码器

几何编码器 $f_i = e_{geo}(P_i)$ 将叶子结点 P_i 的几何表示编码为特征向量 f_i 。部件的几何表示既可以是部件的包围盒 B_i ，也可以是其点云 A_i 。作者对这些表示中的每一个都采用专门的几何编码器。包围盒编码器由单层感知器（SLP）组成。由于部件编码器后面总是跟着一个或多个图编码器，因此单层就足够了。点云 A_i 首先以其平均值为中心，并均匀缩放以具有单位边界球。然后使用另一个 SLP 对中心和尺度进行编码，并使用 PointNet 对归一化的点云进行编码。这两个输出由另一个 SLP 合并并编码到特征向量 f_i 中。

4.1.2 图编码器

部分 P_i 的子图由 (C_i, R_i) 给出，其中 C_i 是子部分， R_i 是它们的关系边。子图中的每个部分 $P_j \in C_i$ 由其特征向量和标签 $\hat{f}_j = (f_j, l_j)$ 的串联表示，而 R_i 中的关系边只有其类型 τ 作为特征。部件标签和边类型都是独热编码向量。作者不存储关系参数，例如旋转对称的轴。图编码器 $f_i = e_{graph}(\{\hat{f}_j \mid P_j \in C_i\}, R_i)$ 将该子图编码为定长特征向量 f_i 。

图编码器的架构受到最近的图同构网络(GIN)和动态图 CNNs 的启发。为了对子图进行编码，作者对沿着图的边传递的消息执行几次迭代。在每次迭代中，结点聚集其邻居的特征来计算更新的特征向量。作者将边的特征也包含在通过边传递的每条消息中。在每次迭代 t 中，部件的特征向量更新为：

$$f_j^{(t)} = \frac{1}{M} \sum_{(\{P_j, P_k\}, \tau) \in R_i} h^{(t)}(f_j^{(t-1)}, f_k^{(t-1)}, \tau), \quad (1)$$

其中， M 为 P_j 的邻居结点数量， $h^{(t)}$ 为 SLP，每次迭代一个； $h^{(t)}$ 对沿边传递的消息进行编码，该消息由源部件的特征向量 $f_j^{(t-1)}$ 、目标部分的特征向量 $f_k^{(t-1)}$ 和边类型 τ 组成。对所有相邻部分进行平均以获得更新的特征 $f_j^{(t)}$ 。迭代从 $f_j^{(0)} = \hat{f}_j$ 开

始，作者在实验中对每个图执行两次消息传递迭代。消息传递后，对所有子部分 $P_j \in C_i$ 进行最大池化来计算整个图的特征向量：

$$f_i^{(t)} = \max\{f_j^{(t)}\}. \quad (2)$$

最后，作者将每次迭代后计算的图特征向量连接起来，并将它们传递给另一个 SLP g ：

$$f_i = g_{skip}(f_i^{(0)}, f_i^{(1)}, f_i^{(2)}). \quad (3)$$

其中， g_{skip} 类似于迭代的跳跃连接，并允许网络利用所有迭代的特性。

4.2 解码器

解码器将根结点特征向量 z 转换回表示为图层次结构的形状。它以自顶向下的方式扩展节点。在每一步中，它首先执行图编码器的反向操作，使用图解码器 d_{graph} 将潜在代码 f_i 转换为其子图。然后，解码器使用几何解码器 d_{geo} 将每个子结点的结果特征向量转换回子结点的几何表示。与编码器不同，作者解码层次结构中每个部件的几何形状，而不仅仅是叶子结点部件。这为训练期间的监督提供了额外的机会，以解码中间结点几何形状的重建损失的形式。

4.2.1 几何解码器

对于一个部件的包围盒表示 $B_i = d_{geo}(f_i)$ 和点云表示 $A_i = d_{geo}(f_i)$ ，有两个可选的解码器。两者都将部件的特征向量转换回部件的几何表示。包围盒解码器实现为两层的多层感知器 (MLP)，将特征向量 f_i 转换为包围盒 $B_i = (c_i, q_i, r_i)$ 。点云解码器使用三层 MLP 从特征向量中得到归一化的点云，使用 SLP 得到点云的中心和尺度。作者对点云的几何编码器和解码器进行了预训练，作为形状部件的点云几何的独立自编码器。这使得能够以稍微降低重建精度为代价大大提高了训练稳定性。

4.2.2 图解码器

图解码器将父结点特征向量 f_i 转换回子图 $(\{\hat{f}_j \mid P_j \in C_i\}, R_i) = d_{graph}(f_i)$ ，其中每个子部分 $P_j \in C_i$ 由特征向量及其标签 $\hat{f}_j = (f_j, l_j)$ 表示。由于子图具有可变数量的部分和边，因此总是解码固定最大数量 n_p 的子部分和它们之间的所有 n_p^2 边，以及子图中对部分或边预测存在的二元概率。部件及其关系是同时解码的。在作者的实验中，最多使用 10 个部件。预测图中不存在的部件和边将被丢弃。

作者首先使用 SLP g_{parts} 从父结点特征向量解码初始特征向量：

$$(\tilde{f}_1, \dots, \tilde{f}_{n_p}) = g_{parts}(f_i) \quad (4)$$

其中， n_p 是最大的子结点数量。为了预测部件的存在，作者计算：

$$p_j = \sigma(g_{xp}(\tilde{f}_j)), \quad (5)$$

其中 p_j 是预测子部件 j 存在的概率， σ 是 sigmoid 函数， g_{xp} 是单线性层。 $p_j < 0.5$ 的部件会被丢弃。

为了预测边的存在，图编码器在每个部件的特征中积累关于图邻域的信息。因此，可以根据一对部件的特征向量来恢复它们之间的边：

$$(p_{(j_1, j_2, \tau_1)}, \dots, p_{(j_1, j_2, \tau_{|T|})}) = \sigma(g_{xe}(\tilde{f}_{j_1}, \tilde{f}_{j_2})), \quad (6)$$

其中 $p_{(j_1, j_2, \tau_1)}$ 部件 P_{j_1} 和 P_{j_2} 之间有类型 τ 的边存在的概率， $|T|$ 是边类别的数量， g_{xe} 是两层的 MLP。如果任何相邻部分不存在，或者 $p_{(j_1, j_2, \tau)} < 0.5$ ，则丢弃边。

作者沿着预测边对消息传递进行两次迭代，从初始特征向量 \tilde{f}_j 开始，得到最终的子结点特征向量 f_j 。实验中作者发现消息传递使部件能够根据边之间描述的关系（如对称或邻接）来改进和协调其几何形状。作为最后一步，作者解码来自特征向量 f_j 的两个附加值：语义标签 l_j 和 P_j 是叶子结点的概率 p_j^{leaf} 。如果一个部件被预测为叶子节点，则不会尝试预测该结点是否存在子结点，从而使网络更容易停止递归。作者发现这有助于收敛，尤其是在训练的早期阶段。

4.3 训练和损失

作者在给定类别的形状数据集 S 上训练 VAE。假设数据集中的模型不具有超过 $n_p = 10$ 个子节点的部分。每个形状都表示为具有已知结构的图的层次结构。作者的目的是训练 VAE 的编码器和解码器，在潜在空间中执行每个形状 S 到特征向量 z 的可逆映射，在潜在空间中，形状的操作（如生成和插值）更容易。为了学习这种映射，作者使用由三部分组成的损失：

$$L_{total} = E_{S \sim \mathcal{S}}[L_r(S) + L_{sc}(S) + \beta L_v(S)], \quad (7)$$

其中， S 为某类形状的分布， E 为期望值。重建损失 L_r 鼓励映射的可逆性，结构一致性损失 L_{sc} 鼓励重建部分和重建边关系之间的一致性，而具有正则化权值 β 的 VAEs 的传统变分正则化 L_v 鼓励潜在空间中的形状流形光滑和简单。实验经验设置 $\beta = 0.05$ 。

5 实验结果和分析

5.1 数据准备

作者使用 PartNet 作为文中所有实验的主数据集。PartNet 为来自 24 个对象类别的 26,671 个 3D 对象提供了具有一致语义标签的细粒度和分层部件注释。作者在实验中使用了三个最大的类别：橱柜、椅子和桌子。由于每个父结点部件的子部件的最大数量 $n_p = 10$ ，因此作者删除在其任何部件中具有超过 10 个子部件的形状。如果需要也可以增加这个最大值，这会稍微增加内存消耗。此外，作者删除了具有未标记部分的形状。剩下的 4871 把椅子、5099 张桌子和 862 个橱柜使用 PartNet 数据集中发布的数据分割分成训练集、验证集和测试集，其比例为 7:1:2。

在 PartNet 中，形状被表示为网格，网格被分成单独的部分。数据集中的每个形状都被缩放到包含在单位球体中。为了获得每个部分的包围盒 B_i ，作者对每个部分的网格拟合一个定向的最小体积包围盒。点云 A_i 是通过部件表面进行 1000 个点的均匀采样得到的。

5.2 三维形状生成

表 1 三维形状生成结果

	rel. quality	rel. coverage
GRASS Orig.	0.714	0.818
GRASS Partnet	0.788	0.818
STRUCTURENET (no edges)	0.984	0.989
STRUCTURENET	1.0000	1.0000

分层图网络的一个直接应用是形状生成。作者使用变分正则化权值 $\beta = 0.05$ 的 VAE。在潜在空间中提供了密集而平滑的形状分布，我们可以从中绘制出新的形状样本，包括几何形状和结构。作者展示了这个应用与 GRASS 的定性结果和定量比较。

5.2.1 定性分析

作者在图 5 中给出了几个生成形状的示例，它们使用包围盒和点云来表示部件几何形状。作者的结果显示了结构和部件几何形状的巨大变化，单个部件的布局在功能上是合理的。对于每个形状，作者生成完整的形状表示，包括单个部件的几何形状、这些部件的分层分解、兄弟结点之间的对称和邻接的边关系以及部件语义。

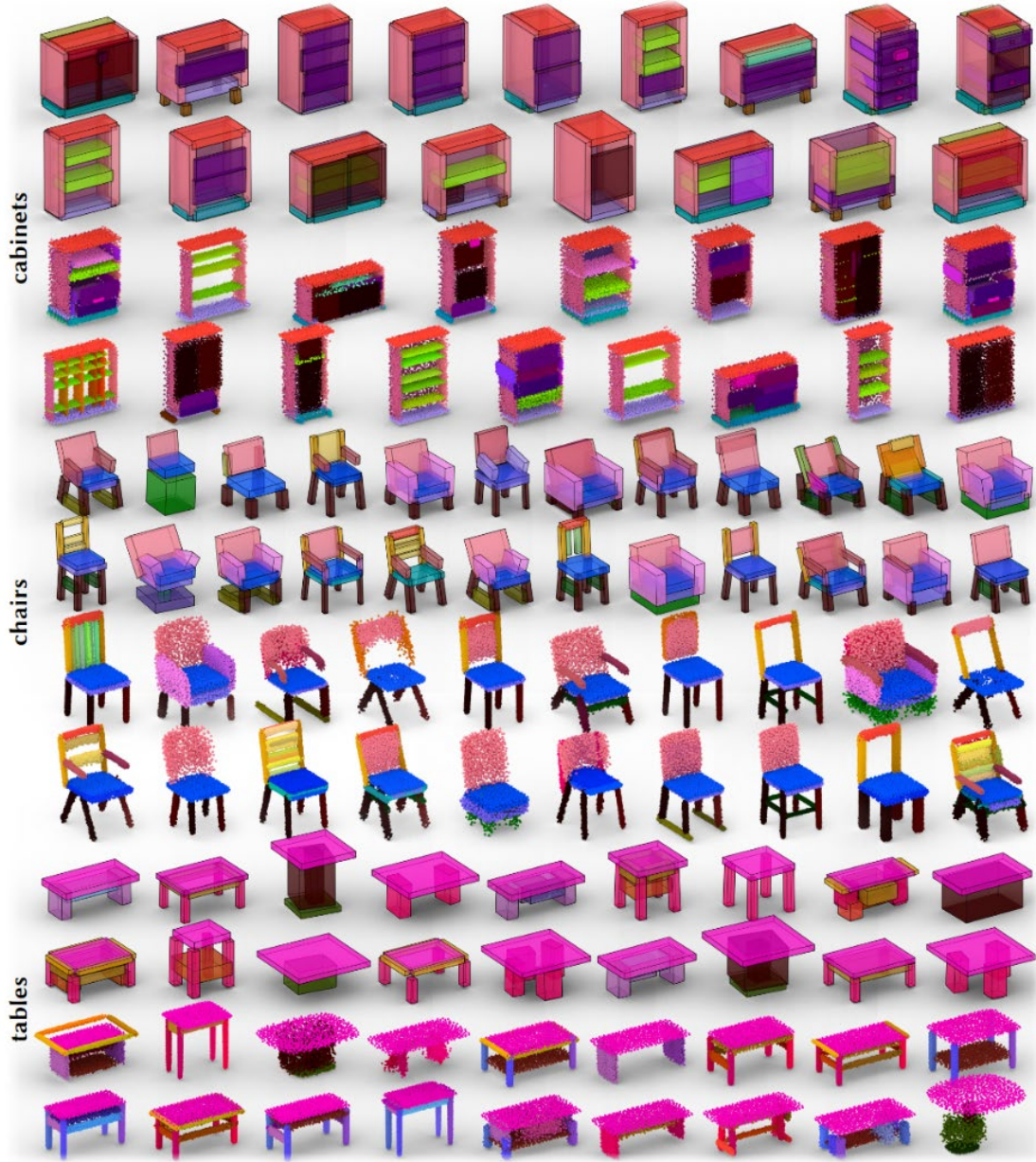


图 5 生成的三维形状

5.2.2 定量分析

从量化分析的角度上讲，此应用的目标是覆盖尽可能多的数据分布，同时避免远离数据分布的主要部分的不切实际的椅子。作者用两个指标来量化这个目标。生成的形状集的质量是通过与任何数据样本的平均最近距离来衡量的，而覆盖率是通过从每个数据样本到生成样本的平均最近距离来衡量的。

$$\begin{aligned}
 quality &:= \sum_{S' \in \mathcal{S}_G} \min_{S \in \mathcal{S}} d_S(S', S) \\
 coverage &:= \sum_{S \in \mathcal{S}} \min_{S' \in \mathcal{S}_G} d_S(S', S)
 \end{aligned} \tag{8}$$

其中 \mathcal{S} 为训练集， \mathcal{S}_G 为生成的形状集合， d_S 为两个形状的点表示之间的倒角距离。

为了与使用这些指标的 GRASS 进行比较，作者使用 StructureNet 和 GRASS 计算了 1000 个形状，并计算了它们的质量和覆盖率。作者在表 1 中显示了与 StructureNet 性能相关的结果（即 StructureNet 分数除以方法分数）。从中可以看到 StructureNet 生成形状的质量和覆盖范围都比 GRASS 有所提高。

5.3 局限性和失败案例

作者讨论了几个限制和失败案例：（1）StructureNet 作为一种数据驱动的方法，自然地继承了数据集中的任何数据采样偏差（例如乒乓球桌这类具有很少示例的形状类别）；（2）尽管作者的经验实验在邻接恢复和对称执行方面表现良好，但推断的潜在空间可能包含具有分离部分或不对称部分的模型，特别是对于包含奇异的，表现不佳的形状变体的数据集；（3）作者将最大兄弟数限制为 $n_p = 10$ ，因此不能在任何给定的部分中编码具有超过 10 个子的形状（完整的形状可以有更多的部分）。内存成本是以 n_p 为单位的二次方，尽管在作者当前的设置下，这仍然远远不是最消耗内存的组件；（4）在整体生成方法中，噪声会使点云更加模糊，而不是影响作者方法中的结构。强噪声可能导致部件缺失、重复部件、分离零件等，尽管结构是离散的，对这种类型的噪声具有相当的鲁棒性；（5）结构感知的点云生成仍然是一个需要进一步研究的新课题。在作者的实验中，为了稳定网络训练，作者预训练和冻结部分点云网络，这以无法恢复细粒度几何细节（例如，椅子腿和椅背上的细节）为代价增加了训练的鲁棒性。

6 全文总结

作者已经将 StructureNet 作为一个 VAE 来呈现，它直接编码形状结构和几何图形，表示为 n 元图的层次结构。作者通过提出递归/分层编码器-解码器架构实现了这一点，该架构同时考虑了部件的几何形状，无论是有向包围盒还是点云，还是捕获邻接和对称关系的部件间结构。作者的关键技术创新是通过（1）明确预测部分或关系边的存在或不存在来处理 n 元图；（2）将编码器和解码器设计为 n 元图实例间的兄弟结点序列的顺序不变；（3）引入新的损失，在所有层次结构中加强几何和结构之间的一致性。学习到的 n 元结构图潜空间，通过对几何和结构的联合捕获，极大地简化了几种应用。例如，可以通过将未标注的数据（例如，部件扫描，点云或图像）投射到潜在空间来“进入”潜在空间；对单个形状执行结构感知编辑；在多个形状之间进行保持层次的插值，或者通过直接采样编码的潜在形状来产生新的和不同的变化。

参考文献

- [1] Mo Kaichun, Guerrero Paul, Yi Li, et al. StructureNet: hierarchical graph networks for 3D shape generation[J]. ACM Transactions on Graphics(TOG), 2019, 38(6): 1-19.
- [2] Li Jun, Xu Kai, Chaudhuri Siddhartha, et al. Grass: Generative recursive autoencoders for shape structures[J]. ACM Transactions on Graphics(TOG), 2017, 36(4): 1-14.