

U-RED: Unsupervised 3D Shape Retrieval and Deformation for Partial Point Clouds

摘要

U-RED 是一种无监督的 3D 形状检索与变形方法，该方法可以将物体观测作为输入，比如 RGB 图像或者点云输入，在预先建立数据库的几何上检索相似的 CAD 模型，再分别对该模型的每个部件进行形变，最后紧密地拟合目标物体。现有方法通常无法处理嘈杂的部分输入，U-RED 从两个方面解决这一问题。首先，由于一个部分形状可能对应多个潜在的完整形状，因此检索方法必须能够处理这种模糊的多对一关系。为此，U-RED 将所有可能的完整形状投影到单位球面上，在推理过程中，球面上的每个采样都将产生一个可行的检索结果。其次，由于现实世界的部分观测通常包含明显的噪声，因此需要一个可靠的学习度量来衡量形状之间的相似性，以确保稳定的检索效果。U-RED 设计了一种新颖的逐点残差引导度量，能够在有噪声的情况下进行有效的度量。然而，该方法在检索时忽略了部件级别的信息，检索整个物体，忽视了局部的几何细节。因此，我们对该方法进行了改进，提出了一种新型的 3D 形状部件检索与变形框架。

关键词：检索与形变；三维重建；部件检索

1 引言

3D 语义场景感知 [24] 涉及将场景分解为其组成的物体，理解并重建所有检测到的物体，并将它们放置在适当的位置以形成整体场景表示。近年来，在多物体几何分析 [18, 26]、场景和物体的动态重建 [5, 6]、结构感知场景补全 [3] 等方面取得了显著进展。这些方法在整体重建质量上表现出良好的结果，但通常难以保持细粒度的几何结构。为了解决这个问题，研究者提出了检索与变形 (R&D) 方法 [20–23, 25]。这些方法利用预先准备的 3D 形状数据库（通常是 CAD 模型）作为先验知识，采用一种两阶段的方案来生成干净紧凑的场景表示。首先，根据选择的度量方式进行检索，从数据库中选择与目标最相似的形状。然后将检索到的形状进行缩放、对齐和旋转，以拟合目标形状。

然而，这些方法面临两个挑战，使得它们对噪声和部分观测不够鲁棒。首先，部分形状可能对应多个完整形状。例如，如果仅观察到一个平面，它可能是椅子的背部或座椅。没有额外的先验信息时，这种对应关系是完全模糊的。直接使用单一的真值进行监督可能导致错误或不期望的结果。因此，检索网络应支持一种一对多 (OTM) 的检索方式。其次，由于光照条件的复杂性和传感器的固有限制，现实场景中的观测通常伴随噪声。因此，设计一个鲁棒的学习度量来衡量形状之间的相似性，对于检索最相似的源形状至关重要。

为了解决这些挑战，U-RED 提出了一种新型的无监督联合 3D 形状检索与变形框架，能够有效处理嘈杂、部分和未见过的物体观测。U-RED 模拟了现实世界中的遮挡、传感器噪声和扫描误差，以生成每个形状的部分点云作为网络训练的输入，随后直接将该方法应用于具有挑战性的现实场景，无需细化调优。

为了实现一对多检索，U-RED 提出将目标部分观测的所有可能完整形状封装在高维单位球面上。具体来说，在联合训练过程中，U-RED 为处理完整形状添加了一个附加分支，以提取归一化的全局特征，该特征对应球面上的一个点。完整形状的特征与目标部分形状的特征进行拼接，作为单个检索的指示符。通过这种方式，检索网络学会在球面上插值不同的完整形状。在推理过程中，U-RED 在球面上均匀采样，以产生多个检索结果，并将唯一的结果作为最终输出。此外，还可以基于联合学习方案建立跨分支的几何一致性，帮助部分分支学习结构感知特征，并提高抗噪声能力。在相似性度量方面，U-RED 设计了一种新颖的逐点残差引导度量。对于目标部分观测中的每个点，预测一个残差向量，描述其坐标与源形状中最近邻点的差异。通过聚合所有残差向量并去除异常值，U-RED 计算剩余向量的平均范数作为最终的度量。残差引导度量对噪声具有鲁棒性，并且可以直接应用于现实场景，而且仅需使用合成数据进行训练。

2 相关工作

2.1 三维形状生成与表示

许多基于深度学习的方法都在潜在空间中构建 3D 形状的紧凑表示。[\[10, 13, 14\]](#) 尝试通过神经网络构建隐式函数。[\[1, 15, 19\]](#) 采用生成模型生成高质量的点云。一些先前的研究也提出了 3D 形状的分解表示技术，其中结构变异在每个几何部件内被单独建模。另一个常见的 3D 形状表示学习方向 [\[4, 12\]](#) 利用编码器-解码器网络生成包含几何和语义信息的高维潜在编码。形状检索任务的一个简单解决方案是直接在编码器-解码器网络生成的潜在空间中比较源形状和目标形状的相似性。然而，这种方法对目标形状的质量极为敏感，且难以处理部分和噪声点云。

2.2 CAD 模型检索

检索与真实物体扫描紧密匹配的高质量 CAD 模型一直是 3D 场景理解中的一个重要问题。先前的静态检索方法将 CAD 与扫描匹配任务视为在描述符空间 [\[2\]](#) 或由深度神经网络编码的潜在嵌入空间中衡量相似性 [\[8\]](#)。相比之下，其他方法 [\[17\]](#) 将此任务建模为一个显式的分类问题。由于数据库中的形状在经过变形步骤后可能具有最佳拟合细节，[\[20\]](#) 提出提取具有变形感知的嵌入，而 [\[9\]](#) 设计了一种新颖的优化目标。然而，这些检索方法忽略了检索与变形之间的内在联系，导致误差累积和性能下降。

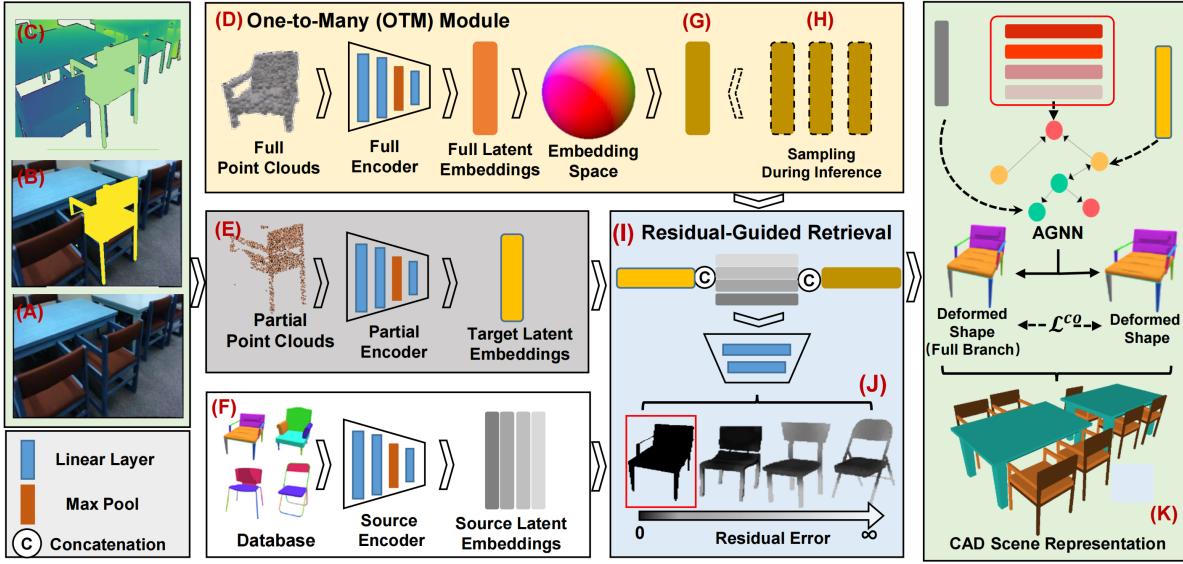


图 1. 方法示意图

2.3 3D 形状变形

计算机图形学领域传统的方法直接优化变形后的形状以适应输入目标。然而，这些方法在处理真实的噪声和部分扫描时常常表现不佳。基于神经网络的技术则尝试从形状集合中学习变形先验，将变形表示为体积扭曲、笼式变形 [23]、基于顶点的偏移 [22] 或流动 [11]。这些变形技术通常需要对网格控制点、笼式网格或顶点数量施加约束，这使得它们不太适用于具有几何异构部分的数据库。此外，这些假设在噪声较大和严重遮挡的实际扫描中往往难以满足。最近，[21] 提出了一种结合损失的新颖训练策略，以同时优化检索与变形。尽管在合成数据集上取得了不错的性能，但其在处理现实世界中部分观测的噪声扫描时的泛化能力有限。

3 论文方法

3.1 论文方法概述

U-RED 提出了一种无监督三维形状检索与变形框架，专注于从部分点云中获取几何匹配的完整形状并进行精确变形以贴合目标。为解决部分形状与完整形状的一对多模糊映射问题，作者设计了一个将完整形状投影到高维单位球表面的模块，通过球面采样生成多个可能的检索结果；同时，为提高对真实场景中噪声的鲁棒性，提出了一种基于点残差的检索技术，用于量化目标与候选形状间的几何差异。在变形阶段，使用基于图注意力的网络实现部件级特征更新，并预测边界框以生成变形后的形状。此外，该框架采用协作式无监督训练策略，通过完整形状指导部分点云分支学习几何特征，并引入跨分支一致性约束以提升鲁棒性。该方法的整体框架如图 1 所示。

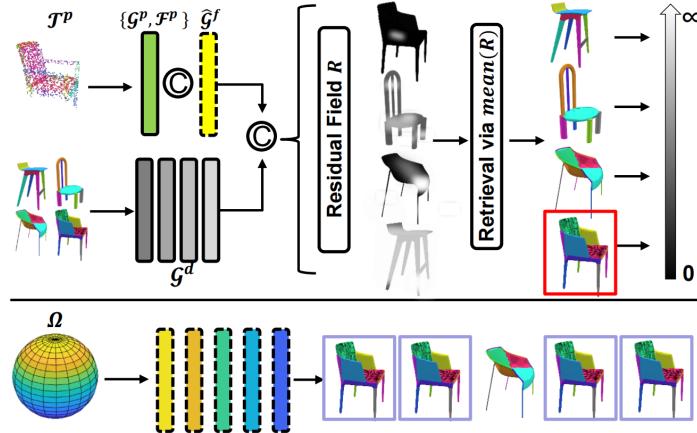


图 2. 部分点云检索

3.2 3D 形状检索

为了实现一对多检索 (One-to-Many Retrieval)，U-RED 设计了一个新颖的 OTM 模块，该模块能够将目标部分形状 T^p 对应的所有可能完整形状 T^f 映射到高维单位球 Ω 的表面。具体来说，在训练过程中，该模块引入了一个额外的分支用于处理完整形状 T^f ，并提取其全局特征 $\{\mathcal{F}^f, \mathcal{G}^f\}$ 。通过归一化操作， \mathcal{G}^f 被映射为单位球 Ω 上的一个点，即 $\hat{\mathcal{G}}^f$ 。对于目标部分形状的特征 $\{\mathcal{F}^p, \mathcal{G}^p\}$ ，U-RED 将其与 $\hat{\mathcal{G}}^f$ 拼接作为检索网络的输入，其中 $\hat{\mathcal{G}}^f$ 作为指示器，最终实现一对一检索 (One-to-One Retrieval)。检索网络通过学习在 Ω 上的插值，使得 Ω 上的每个点都对应一个可能的完整形状。在推理过程中，OTM 在 Ω 的表面均匀采样点作为指示器，每次采样都会生成一个检索结果。最后将所有唯一的检索结果收集为可能的结果集合，并将被选择次数最多的形状视为最佳检索结果。

为了增强对噪声观测的鲁棒性，U-RED 设计了一种基于点残差的检索相似性度量 (Residual-Guided Similarity Metric for Retrieval)。如图 2 所示，U-RED 的检索网络预测一个残差场 $R = \{R_i \in \mathbb{R}^3, i = 1, \dots, M\}$ ，其中每个点 $P_i \in T^p$ 的残差向量 R_i 描述了从 P_i 到其在变形后源形状 \tilde{O}^c 中的最近邻 O_i 的位移矢量。这里采用 L_2 损失函数来监督 R 的训练，公式如下：

$$\mathcal{L}^{re} = \frac{1}{M} \sum_{i=1}^M \|P_i + R_i - Q_i\|^2, \quad (1)$$

其中， Q_i 是通过最近邻搜索获得的点。利用 R ，可以推导出检索的度量 $mean(R)$ 。为增强对噪声的鲁棒性，U-RED 对 R 按范数进行排序，去除残差范数最大的 10% 点，然后计算剩余点的平均残差范数 $mean(R)$ ，作为最终的检索相似性度量。

3.3 基于图注意力的形变

U-RED 的变形网络由一个用于部件级特征聚合的图注意力神经网络 (AGNN) 和一个用于预测每个部件边界框的回归器组成。AGNN 接受三类节点的输入：1. 从检索到的源形状 O_c 中提取的全局特征 G^d ；2. 从目标对象 T^p 中提取的全局特征 G^p ；3. 源形状 O_c 的部件特征 $\{\mathcal{P}_i^f, i = 1, 2, \dots, N\}$ 。

该模块堆叠了两个交替的自注意力模块和交叉注意力模块。在自注意力中，不同的部件之间交换信息；而在交叉注意力中，全局节点传播全局结构信息以指导部件节点的特征更新。总体更新过程定义为：

$$\mathcal{F}' = \mathcal{F} + MHA(Q, K, V), \quad (2)$$

其中 MHA 表示多头注意力机制。在自注意力模块中， $Q, K, V = \mathcal{F} = \mathcal{P}^f$ ；而在交叉注意力模块中， $Q = \mathcal{F} = \mathcal{P}^f$ ， $K, V = \{\mathcal{G}^p, \mathcal{G}^d\}$ ，即全局特征的拼接。

最后， \mathcal{P}^f 被输入到回归器中，回归器预测部件的中心位移 C_d 和轴对齐缩放参数 $\{s_w, s_h, s_l\}$ 。每个部件的最终边界框通过以下公式恢复：

$$C = C_d + C_0, \quad (3)$$

$$\{H, W, L\} = \{s_w W_0, s_h H_0, s_l L_0\}, \quad (4)$$

其中 $\{C_0, W_0, H_0, L_0\}$ 分别是初始边界框的中心、宽度、高度和长度。

3.4 损失函数定义

如图 1(D) 所述，U-RED 引入了一个额外的分支来处理完整形状，从而实现一对多检索。除了这一作用之外，完整形状分支还可以用来引导部分形状分支理解部分观测点云的几何特征 1(E)。通过这种方式，U-RED 的联合学习策略在两个分支之间建立了显式的几何一致性。

在变形网络中，检索到的源形状 O_c 被部分形状分支变形为 \tilde{O}_p^c ，同时也被完整形状分支变形为 \tilde{O}_f^c 。 O_p^c 和 O_f^c 应该保持一致。此外，给定由部分形状分支生成的 \tilde{O}_p^c 和由完整形状分支生成的 \tilde{O}_f^c ，检索网络生成残差 R_p 和 R_f 。通过在两个分支之间强制保持一致性，部分形状分支被引导生成与完整形状分支相似的结果，并被迫从部分输入中利用完整形状的几何特征。

对于部分形状中的每个点 P_i ，其残差 $\{R_{pi} \in R_p\}$ 应与完整形状中对应点 Q_i 的残差 $\{R_{qi} \in R_f\}$ 保持一致。定义 $R'_f = \{R_{qi}\}$ ，因此跨分支一致性损失定义如下：

$$\mathcal{L}^{co} = \mathcal{L}_1^{co} + \mathcal{L}_2^{co} = \|O_p^c - O_f^c\|^2 + \|R_p - R'_f\|^2, \quad (5)$$

其中使用点集之间的欧几里得距离作为损失函数。

U-RED 在 R&D 框架中聚合了所有损失，最终的无监督训练目标函数可以总结为：

$$\mathcal{L} = \lambda_0 \mathcal{L}^b + \lambda_1 \mathcal{L}^{re} + \lambda_2 \mathcal{L}^{co}, \quad (6)$$

其中 $\{\lambda_0, \lambda_1, \lambda_2\}$ 是权重参数。 \mathcal{L}^b 包含多个基本损失，包括 Chamfer 距离损失 \mathcal{L}^{cd} 和重建损失 \mathcal{L}^r 。

4 复现细节

4.1 与已有开源代码对比

我们提出了 PA-Red (Part-aware 3D Shape Retrieval and Deformation)，一种新型的部件感知 3D 形状检索与变形框架。该框架以点云为输入，从数据库中检索并变形几何上相似的

部件，以准确重建目标物体。与以往专注于整体物体检索的方法不同，PA-Red 引入了精细化的部件级别检索，使系统能够准确捕获单个部件的局部几何细节。我们的方法在 U-RED 的代码基础上，进行了较大的改动：

- 进行数据预处理，生成数据集，对数据进行划分，计算部件样本之间的相似度并定义正负样本对
- 对检索和形变任务进行解耦，通过对比学习的方法来训练检索网络
- 通过部件的连接性来进行形变的学习，重新设计了部件连接的损失函数

4.2 改进方法概述

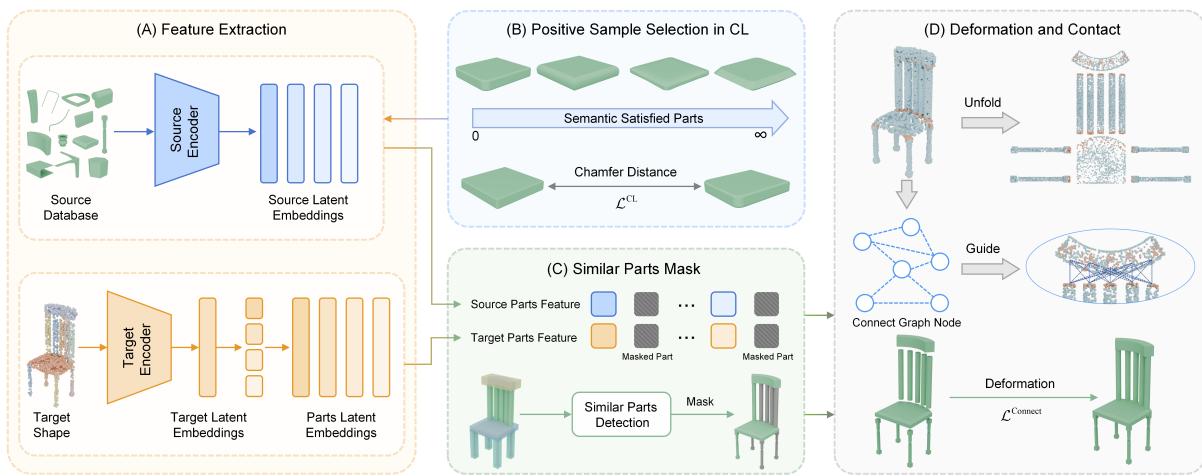


图 3. PA-Red 方法框架

现有的 R&D 方法通常面临两大挑战。首先，大多数方法在检索时忽略了部件级别的信息，而是检索整个物体。由于复杂物体通常由多个异质部件组成，整体检索往往忽视局部几何细节，从而降低了检索精度。为了提升检索性能，方法需要能够关注部件细节并实现部件级别的检索。其次，这些方法的检索性能往往不尽如人意。许多现有方法仅依赖于简单的全局特征或基本的相似性度量，而在处理形状变体和局部差异时，常常难以提供准确且有区分度的结果。这种粗略的相似性评估无法有效区分具有细微几何变化的对象，尤其是在对相似形状进行精确匹配时表现不足。

为了解决上述挑战，我们提出了 PA-Red (Part-aware 3D Shape Retrieval and Deformation)，一种新型的部件感知 3D 形状检索与变形框架。该框架以点云为输入，从数据库中检索并变形几何上相似的部件，以准确重建目标物体。与以往专注于整体物体检索的方法不同，PA-Red 引入了精细化的部件级别检索，使系统能够准确捕获单个部件的局部几何细节。具体而言，为了在广泛的部件空间中进行检索，我们使用一种对比学习方法，其中正样本对的定义同时考虑了几何相似性和语义一致性。然而，当同一对象的多个部件具有相似的几何特性时，会引发正样本选择不一致的问题。为此，我们引入了一种掩码机制。该机制通过比较部件的边界框来检测对象内相同或几乎相同的部件。被认为重复的部件会被掩码遮盖，从而确保在对比学习过程中每组相似部件仅使用一个实例。这有助于保持正样本选择的一致性，减少

Method	Chair	Table	Cabinet	Average
Uy et al. [21]	0.7152	1.1765	0.7774	0.9614
U-Red [7]	1.2012	0.8478	1.4357	1.0258
Ours	0.5150	0.9804	1.001	0.7876

表 1. 在 PartNet 数据集上 Chamfer Distance 的实验结果.

歧义并提高模型的学习效率。此外，当部件被单独检索时，由于部件之间的关系并未被保留，相邻部件之间的连通性可能会丢失。为了解决这一问题，我们提出了一种基于接触点的连通性监督方法。在训练期间，我们利用目标点云的连通图来引导部件之间接触点的对齐。通过比较部件之间的接触区域，我们让检索到的部件根据连通关系对齐，从而减少间隙并确保部件之间的紧密契合。将这些基于连通性的损失项纳入训练目标，确保模型不仅关注部件检索，还优化部件之间的连通性，从而增强模型在重建完整物体时的整体性能。

4.3 创新点

我们的主要贡献如下

- 我们提出了 PA-Red，这是一种新颖的部件感知 3D 形状检索与变形框架，实现了部件级别的检索与变形，从而捕获细粒度的几何细节
- 我们在对比学习中引入了一种掩膜机制，以减少相似部件之间的歧义，确保一致的特征表示并提升检索精度
- 我们开发了一种基于接触点的连通性监督方法，在变形过程中保持部件连通性和对齐性

5 实验结果分析

为验证 PA-Red 的有效性，我们在 PartNet [16] 数据集的三个类别（椅子、桌子、柜子）上进行了实验。我们遵循 [21] 中的划分方法，将数据分为可变形源数据库、训练集和测试集。PartNet 数据集来源于 ShapeNet，包含 1,419 个源模型、11,433 个训练模型和 2,861 个测试模型。我们通过均匀采样点来表示所有形状。对于完整形状，我们采样 2048 个点，而对于源数据库中的单个部件，我们采样 1024 个点。我们采用 Point Transformer V3 作为骨干网络。优化器使用 Adam，初始学习率设为 $1e^{-3}$ ，在单块 NVIDIA 4090 GPU 上以 16 的批量大小训练 300 个 epochs。在实验中，我们设置默认损失权重为 $\{\lambda_0, \lambda_1, \lambda_2\} = \{1.0, 0.5, 3.0\}$ 。无论是点级、全局还是部件级特征表示，我们都将其维度统一为 512。

我们为每个类别分别计算了平均 Chamfer Distance（以 $1e^{-2}$ 为单位），并进一步计算了整个数据集的实例平均 Chamfer Distance。该指标通过比较变形后的形状与其对应的真实形状得出。我们选择 U-RED [21] 和 Uy [7] 作为基线方法进行比较，在所有实验中使用相同的数据划分。每种基线方法均按照其原始论文中指定的参数进行配置，以公平评估各方法的性能。

我们在合成数据集 PartNet 上进行了多项实验。这些实验结果表 1 表明，PA-Red 在处理合成场景中的点云时显著优于现有的最先进方法。在 PartNet 数据集中，PA-Red 在三个类别上的平均性能比 U-Red 提高了 30.2%，并相较于 Uy 的方案也表现出显著优势，平均 Chamfer Distance 减少了 22.1%。图 4 提供了定性对比，显示 PA-Red 在检索结果的准确性和变形能力上具有更强的表现。由于 Uy 和 U-Red 检索的是完整物体，导致生成的模型与点云的拟合较差，而 PA-Red 能够实现更好的拟合。

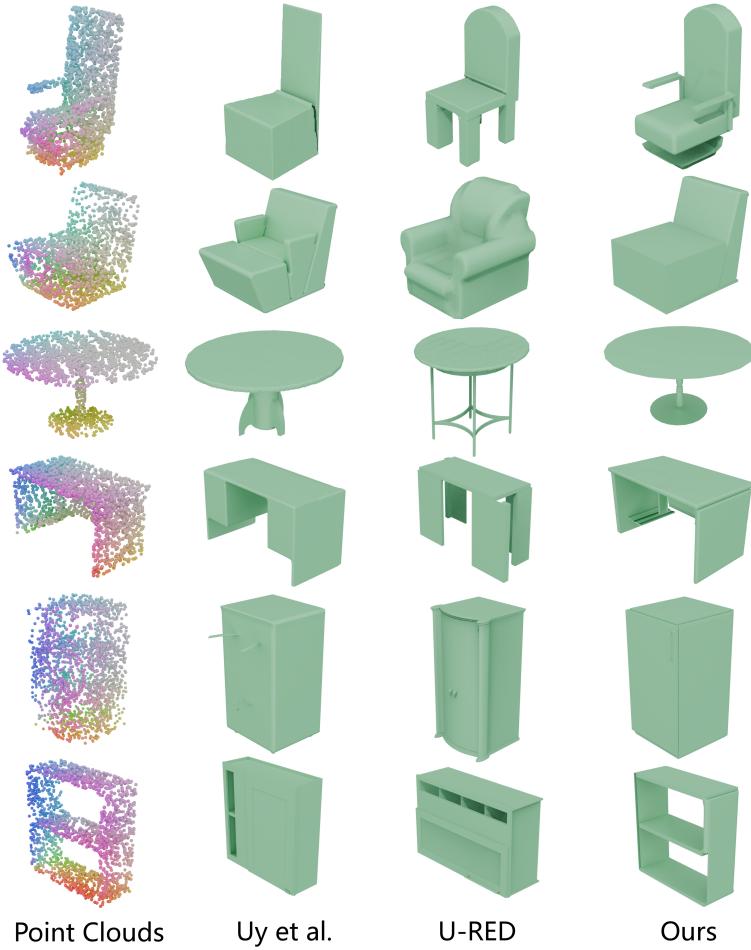


图 4. PartNet 实验结果展示

6 总结与展望

在本文中，我们提出了一种新颖的部分感知框架——PA-Red，用于 3D 形状的检索和变形。PA-Red 通过关注部件级细节而非整体对象检索，提高了检索和重建的准确性，使模型能够捕捉复杂形状所需的细粒度几何细节。通过对比学习，PA-Red 通过基于几何相似性和语义一致性定义正样本，并结合解决相似部件引起歧义的掩码机制，确保了稳健的检索性能。为保持结构的一致性，我们还引入了基于部件接触点的连接性监督，以确保检索的部件能够正确对齐并保持连接性。实验结果表明，PA-Red 在处理部分和噪声输入方面显著提升了对象重建的质量和准确性，并在合成和真实场景中均优于传统方法。

参考文献

- [1] Panos Achlioptas, Olga Diamanti, Ioannis Mitliagkas, and Leonidas Guibas. Learning representations and generative models for 3d point clouds. In *International conference on machine learning*, pages 40–49. PMLR, 2018.
- [2] Frederic Bosche and Carl T Haas. Automated retrieval of 3d cad model objects in construction range images. *Automation in Construction*, 17(4):499–512, 2008.
- [3] Anh-Quan Cao and Raoul De Charette. Monoscene: Monocular 3d semantic scene completion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3991–4001, 2022.
- [4] Angela Dai, Charles Ruizhongtai Qi, and Matthias Nießner. Shape completion using 3d-encoder-predictor cnns and shape synthesis. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5868–5877, 2017.
- [5] Yan Di, Henrique Morimitsu, Shan Gao, and Xiangyang Ji. Monocular piecewise depth estimation in dynamic scenes by exploiting superpixel relations. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4363–4372, 2019.
- [6] Yan Di, Henrique Morimitsu, Zhiqiang Lou, and Xiangyang Ji. A unified framework for piecewise semantic reconstruction in dynamic scenes via exploiting superpixel relations. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pages 10737–10743. IEEE, 2020.
- [7] Yan Di, Chenyangguang Zhang, Ruida Zhang, Fabian Manhardt, Yongzhi Su, Jason Rambach, Didier Stricker, Xiangyang Ji, and Federico Tombari. U-red: Unsupervised 3d shape retrieval and deformation for partial point clouds. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 8884–8895, 2023.
- [8] Can Gümeli, Angela Dai, and Matthias Nießner. Roca: Robust cad model retrieval and alignment from a single image. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4022–4031, 2022.
- [9] Vladislav Ishimtsev, Alexey Bokhovkin, Alexey Artemov, Savva Ignatyev, Matthias Niessner, Denis Zorin, and Evgeny Burnaev. Cad-deform: Deformable fitting of cad models to 3d scans. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XIII 16*, pages 599–628. Springer, 2020.
- [10] Wonbong Jang and Lourdes Agapito. Codenerf: Disentangled neural radiance fields for object categories. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 12949–12958, 2021.

- [11] Chiyu Jiang, Jingwei Huang, Andrea Tagliasacchi, and Leonidas J Guibas. Shapeflow: Learnable deformation flows among 3d shapes. *Advances in Neural Information Processing Systems*, 33:9745–9757, 2020.
- [12] Jincen Jiang, Xuequan Lu, Lizhi Zhao, Richard Dazaley, and Meili Wang. Masked autoencoders in 3d point cloud representation learning. *IEEE Transactions on Multimedia*, 2023.
- [13] Lars Mescheder, Michael Oechsle, Michael Niemeyer, Sebastian Nowozin, and Andreas Geiger. Occupancy networks: Learning 3d reconstruction in function space. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4460–4470, 2019.
- [14] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1):99–106, 2021.
- [15] Kaichun Mo, Paul Guerrero, Li Yi, Hao Su, Peter Wonka, Niloy Mitra, and Leonidas J Guibas. Structurenet: Hierarchical graph networks for 3d shape generation. *arXiv preprint arXiv:1908.00575*, 2019.
- [16] Kaichun Mo, Shilin Zhu, Angel X Chang, Li Yi, Subarna Tripathi, Leonidas J Guibas, and Hao Su. Partnet: A large-scale benchmark for fine-grained and hierarchical part-level 3d object understanding. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 909–918, 2019.
- [17] Liangliang Nan, Ke Xie, and Andrei Sharf. A search-classify approach for cluttered indoor scene understanding. *ACM Transactions on Graphics (TOG)*, 31(6):1–10, 2012.
- [18] Jiaming Sun, Yiming Xie, Linghao Chen, Xiaowei Zhou, and Hujun Bao. Neuralrecon: Real-time coherent 3d reconstruction from monocular video. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 15598–15607, 2021.
- [19] Yongbin Sun, Yue Wang, Ziwei Liu, Joshua Siegel, and Sanjay Sarma. Pointgrow: Autoregressively learned point cloud generation with self-attention. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 61–70, 2020.
- [20] Mikaela Angelina Uy, Jingwei Huang, Minhyuk Sung, Tolga Birdal, and Leonidas Guibas. Deformation-aware 3d model embedding and retrieval. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part VII 16*, pages 397–413. Springer, 2020.
- [21] Mikaela Angelina Uy, Vladimir G Kim, Minhyuk Sung, Noam Aigerman, Siddhartha Chaudhuri, and Leonidas J Guibas. Joint learning of 3d shape retrieval and deformation.

In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11713–11722, 2021.

- [22] Weiyue Wang, Duygu Ceylan, Radomir Mech, and Ulrich Neumann. 3dn: 3d deformation network. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1038–1046, 2019.
- [23] Wang Yifan, Noam Aigerman, Vladimir G Kim, Siddhartha Chaudhuri, and Olga Sorkine-Hornung. Neural cages for detail-preserving 3d deformations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 75–83, 2020.
- [24] Cheng Zhang, Zhaopeng Cui, Yinda Zhang, Bing Zeng, Marc Pollefeys, and Shuaicheng Liu. Holistic 3d scene understanding from a single image with implicit representation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8833–8842, 2021.
- [25] Ruida Zhang, Chenyangguang Zhang, Yan Di, Fabian Manhardt, Xingyu Liu, Federico Tombari, and Xiangyang Ji. Kp-red: Exploiting semantic keypoints for joint 3d shape retrieval and deformation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 20540–20550, 2024.
- [26] Zihan Zhu, Songyou Peng, Viktor Larsson, Weiwei Xu, Hujun Bao, Zhaopeng Cui, Martin R Oswald, and Marc Pollefeys. Nice-slam: Neural implicit scalable encoding for slam. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 12786–12796, 2022.