arXiv:2303.04341v2 [cs.CV] 3 Jun 2023

神经向量场：通过显式学习实现隐式表示

Xianghui Yang[1], GuoshengLin[2],ZhenghaoChen[1],Luping Zhou[1]悉尼大学, [2]南洋理工大学{xianghui.yang,zhenghao.chen,luping.zhou}@sydney.edu.au, gslin@ntu.edu.sg

摘要

深度神经网络（*DNN*）广泛应用于当今的三维表面重建任务。此类方法可进一步分为两类：一类通过移动顶点显式变形模板，另一类则将三维曲面隐式表示为有符号或无符号距离函数。结合先进的显式学习过程与隐式函数强大的表示能力，我们提出了一种新颖的三维表示方法——神经向量场（*NVF*）。该方法不仅采用显式学习过程直接操作网格，还利用无符号距离函数（*UDF*）的隐式表示，突破分辨率和拓扑方面的限制。具体而言，我们的方法首先预测从查询点指向曲面的位移，并将形状建模为向量场。与大多数现有的基于无符号距离场的方法依赖网络微分来获取方向场不同，生成的向量场同时编码距离场和方向场，并消除了“脊”点处的歧义，使方向场的计算简单直接且免微分。免微分这一特性使我们能够进一步通过向量量化学习形状码本，以编码跨对象先验、加速训练过程，并增强模型在跨类别重建中的泛化能力。在表面重建基准数据集上开展的大量实验表明，在多种评估场景下，我们的方法均优于最先进的方法，包括水密与非水密形状、类别特定与类别无关重建、未见类别重建以及跨域重建。我们的代码已发布于 https://github.com/Wi-sc/NVF。

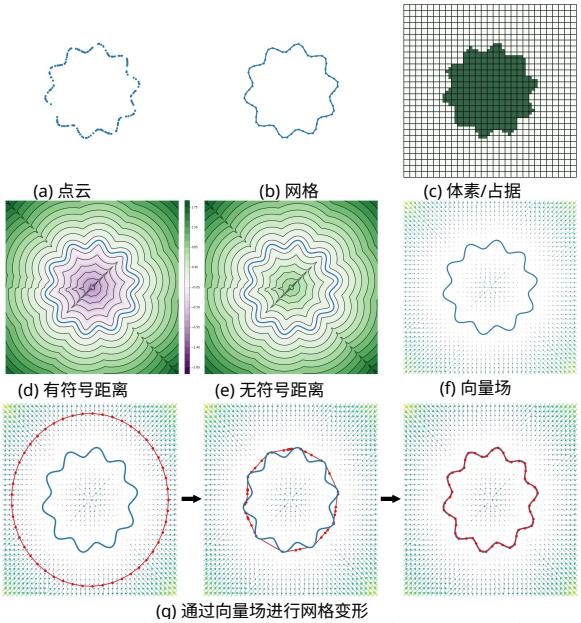


图 1. 常见的三维表示。显式表示：(a) 点云，(b) 网格，(c) 体素。隐式表示：(c) 占据，(d) 从有符号距离函数重建，以及 (e) 从无符号距离函数重建。我们的方法通过 (f) 向量场表示连续曲面。(g) 向量场可以像显式表示方法那样，使网格（红色）发生形变。

1. 引言

从无结构、离散且稀疏的点云重建连续曲面，是当今机器人学、视觉和图形学应用中一项新兴但并非易事的任务，因为点云难以直接用于下游应用，除非将其还原为高分辨率曲面 [5, 7, 38, 42]。

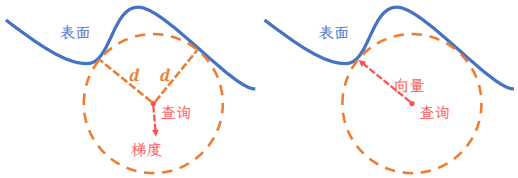
随着深度神经网络（DNN）取得巨大成功，已有少数基于 DNN 的表面重建方法展现出良好的重建性能。根据输出表示是显式还是隐式，这些方法大致可分为两类。如图 1所示，包括 网格 和 体素 方法在内的显式表示方法能够准确表示曲面的位置，它们通过学习变形模板 [3, 4, 19, 26, 29, 68] 或预测体素网格 [10, 30, 59]来实现这一点。显式表示便于下游应用，但通常受分辨率和拓扑的限制。另一方面，诸如.....的隐式表示

占据 和 有符号距离函数（SDF） 将曲面表示为标量函数的等值线。由于能够表示拓扑结构更复杂、分辨率任意的曲面，这类方法日益受到关注 [12, 14, 22, 35, 44, 48]。然而，大多数隐式表示方法通常需要进行特定的预处理，以封闭非水密网格并移除内部结构。为摆脱隐式表示所需的上述预处理，Chibane 等人 [15]提出了 神经无符号距离场（*NDF*），它采用 无符号距离函数（*UDF*）作为神经隐式函数（NIF），并通过预测查询位置与目标表面之间的正标量来对连续曲面建模。尽管 UDF 具有一定优势，但与其他隐式表示方法（例如有符号距离函数）相比，其表面提取过程更为复杂。使用滚球算法 [5] 或基于梯度的移动立方体算法 [28, 83] 进行此类提取时，推理过程中需要对模型求导（即依赖微分）。此外，在“脊”点处，UDF 存在梯度歧义；如图 [1] 所示，用于表面提取的梯度无法准确指向目标点。2a。

在这项工作中，我们提出了一种新颖的三维表示方法，神经向量场（*NVF*），它结合了直接操作网格的显式学习过程和无符号距离函数的隐式表示，兼具两种方法的优势。也就是说，NVF 可以像图1g所示的显式表示方法那样直接操作网格，同时也能像隐式表示方法那样，以任意分辨率和拓扑表示形状。具体而言，NVF 将三维形状建模为向量场，并通过学习得到的函数$\mathbf{q} \in \mathbb{R}^3$ 计算点与其在表面上的最近邻点之间的位移$\hat{}\mathbf{q} \in \mathbb{R}^3$ $f(\mathbf{q}) = \mathbf{\Delta q} = \hat{}\mathbf{q} - \mathbf{q} : \mathbb{R}^3 \Rightarrow \mathbb{R}^3$。因此，NVF 既可用作隐式函数，也可用作显式形变函数，因为该函数输出的位移可直接用于形变源网格（即，图1g）。总体而言，它在向量场中编码了距离场和方向场，两者均可从向量场中直接获得。

与现有的基于无符号距离场的方法不同，我们的 NVF 表示避免了全面的推理过程
通过在表面提取过程中跳过梯度计算[1]，，并如图所示直接学习位移，从而缓解歧义。2b这种单次前向传播的特性使 NVF 摆脱了对微分的依赖，显著缩短了推理时间并降低了内存开销，还使我们的模型能够在嵌入特征空间中学习由不可微的离散形状代码组成的形状码本。该

[1]基于学习的方法通过模型微分计算距离场的梯度。梯度的反方向应指向目标表面上的最近邻点。



(a) NDF [15]。 (b) NVF。

图 2. 梯度歧义。(a) NDF [15] 无法保证将点拉到曲面上（即，梯度歧义）；而 (b) 我们的 NVF 通过直接学习位移解决了这一问题。

学习到的形状码本进一步提供了跨对象先验，从而提升模型在跨类别重建任务中的泛化能力；同时，它在训练过程中作为正则化项，加快了训练进程。**本文以向量量化为例，说明 NVF 的无微分特性为模型设计提供了更大的灵活性。**

我们在两个表面重建基准数据集上进行了大量实验：合成数据集 ShapeNet [8] 和真实扫描数据集 MGN [6]。除大多数重建方法所展示的类别特定重建 [15, 76] 外，我们还通过类别无关重建、未见类别重建和跨域重建任务评估了我们的方法，以考察模型泛化能力。实验结果表明，与其他基于无符号距离场的方法相比，我们的 NVF 能显著缩短推理时间，因为我们省去了全面的表面提取步骤，也无需在查询位置计算梯度。此外，借助形状码本，我们观察到性能显著提升，并且模型在不同类别间具有更好的泛化能力。

我们的贡献总结如下。

- 我们提出了一种三维表示*NVF* ，用于更好地实现三维场表示。它连接了显式学习与隐式表示，并兼具二者的优势。我们的方法能够以免微分的方式获取查询位置的位移，因此显著降低了推理复杂度，并为网络结构设计提供了更大的灵活性，使其能够包含不可微分组件。

- 得益于我们的无微分设计，我们进一步提出了一种在特征空间中学习形状码本的方法，并采用向量量化策略提供跨对象先验。通过这种方式，每个查询位置都被编码为特征空间中离散编码的组合，进而用于学习 NVF。

- 我们进行了大量实验，以评估所提出方法的有效性。该方法在两个基准数据集上始终展现出良好的性能，

涵盖多种评估场景：水密形状与非水密形状、类别特定与类别无关的重建、未见类别重建，以及跨域重建。

2. 相关工作

2.1. 显式三维表示

早期的形状表示方法建立在体素之上 [17, 43, 58]。由于三维空间被离散化为网格，因此可以采用基于学习的图像处理技术进行处理 [34, 37, 53, 69, 70]。然而，体素通常会随着分辨率的提高而产生立方级增长的内存占用。因此，近期的基于体素的方法引入了自适应离散化 [10, 30, 54, 59] 和体素块哈希 [57]，以缓解更高分辨率所带来的更高计算需求。

相比之下，基于点云的方法能够以更低的计算成本和更高的易用性，对曲面进行紧凑且稀疏的编码。近年来，已有多种基于学习的点云处理方法 [9, 49, 60, 67, 82] 被提出，并在三维形状分析 [16, 32, 33, 41, 66, 79, 84] 和合成 [1, 26, 75, 78]方面取得了巨大成功。然而，点云通常缺乏丰富的几何信息（例如，曲面和拓扑），因此在下游任务中的应用受到限制。

为了补充这类几何信息，多边形网格 [36, 40, 50, 65, 74] 通过由顶点和边组成的图来定义三维形状。然而，它们通常受限于固定拓扑，因为大多数方法都是通过移动顶点，将模板（例如，椭球体 [65], 球体 [36, 46, 74]）变形为目标形状。近期研究通过从图表 [26], 体素 [25], 变形，以及剔除不必要的面 [46]，解决了这些问题。

我们观察到，这些方法促使网络在三维空间中模拟向量场。受此启发，我们使用显式学习来学习模型，以向量场而非网格间距离作为学习向量场的依据。如此一来，我们便能突破模板在固定拓扑和分辨率方面的限制。

2.2. 隐式三维表示

为克服大多数显式表示所面临的有限分辨率和固定网格拓扑结构障碍，隐式表示方法通过隐式函数表示连续曲面。此类表示通常由多层感知器（MLP）实现，根据作为查询点的位置生成二值占据 [14, 18, 24, 44, 55, 56] 和有符号距离函数 [12, 35, 47] 。另一方面，由于占据和有符号距离函数的特性，这些方法需要在预处理阶段通过人为闭合形状来获得水密网格。首先，对于开放曲面，定义其内部和外部同样并非易事。其次，预处理会导致大量几何信息丢失，并且在非水密网格上缺乏泛化能力。

为对一般的非水密形状进行建模，Chibane 等人。[15] 引入了无符号距离函数，以学习无符号距离。近期，其他方法提升了无符号距离函数的性能和泛化能力 [62, 64, 76, 81, 83]。例如，GIFS [76]采用交分类分支，而 CAP-UDF [83]直接在原始点云上优化模型。尽管这类基于无符号距离场的方法能够直观地提升隐式表示的泛化能力，但它们需要复杂的表面提取处理，例如滚球算法 [15]和基于梯度的移动立方体算法 [28, 83]。

与上述基于无符号距离场的方法不同，我们的 NVF 无需通过微分来获取方向场的梯度，只需一次前向传播即可获得距离场和方向场。这种高效的免微分策略能够显著减轻推理负担。

2.3. 向量量化

向量量化（VQ）策略将提取的特征转换为量化且紧凑的潜在表示。Oord 等人 首先提出了 VQ-VAE [61]，将 VQ策略与变分自编码器结合，用于图像和语音生成。随后，Yu等人提出 VQ-GAN [77]，将码本与对抗学习结合，用于高质量图像合成。总体而言，已有多种方法采用这一策略进行图像生成 [21, 27, 39, 51], 、语音合成 [20, 73], 、视频 [23, 31, 52, 63], 、形状生成 [13, 45, 72, 80] 以及压缩 [2, 11]。在本研究中，我们同样采用 VQ，并在特征空间中使用多头码本，以去噪、加速训练并提升形状重建的泛化能力。

3. 方法

给定一个三维查询位置 $\mathbf{q}$，它来自查询集 $Q \in \mathbb{R}^{M\times 3}$，以及一个稀疏点云 $P \in \mathbb{R}^{N\times 3}$，我们的 NVF框架会预测位移向量 $\Delta\mathbf{q} \in \mathbb{R}^3$，将 $\mathbf{q}$ 移动到 P 的底层表面。这通过三个主要模块实现，即，特征提取、多头码本和场预测，如图 3 所示。获得预测的向量场后，我们采用移动立方体算法生成表面网格。以下将详细介绍各个模块。

3.1. 神经向量场

如上所述，我们的 NVF 是一个神经网络函数，它通过预测给定查询点 $\mathbf{q}$ 的位移场 $\Delta\mathbf{q} \in Q$ 到其最近的

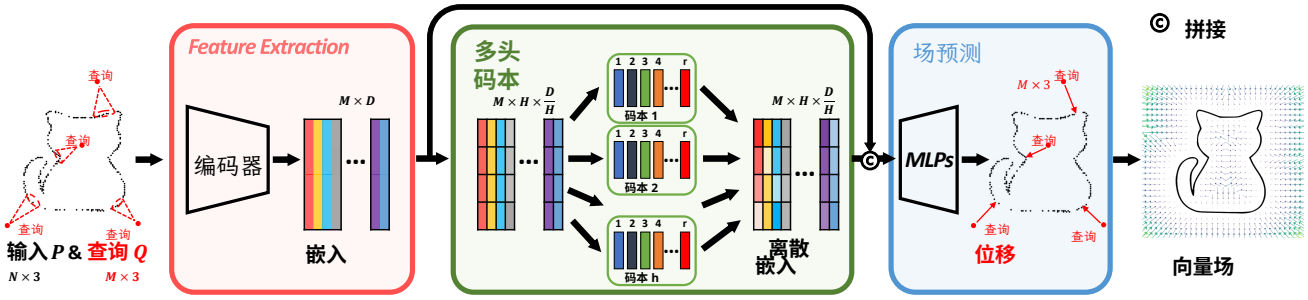


图3. 我们的 NVF 框架概览。该框架对输入点云（黑点）进行编码，为每个查询点（红点）采样特征，并通过向量量化离散化查询嵌入，从而预测查询位置到目标表面的位移（红色箭头）。

位于表面上的点 $\hat{\mathbf{q}} \in \mathcal{S}$ 来对形状 进行建模。其公式为

$$NVF(\mathbf{q}) = \Delta\mathbf{q} = \min_{\hat{q}\in\mathcal{S}} \hat{\mathbf{q}} - \mathbf{q}. \tag{1}$$

请注意，NVF 是有向距离场，能够以直接的方式同时编码距离场 $d = ||\hat{\mathbf{q}} - \mathbf{q}||_2$ 和方向场 $\mathbf{g} = (\hat{\mathbf{q}} - \mathbf{q})/||\hat{\mathbf{q}} - \mathbf{q}||_2$。因此，与距离场不同，我们的 NVF 无需对距离场求导即可获得方向信息。本文将这一特性称为“免微分”，它能显著缩短推理时间。另一方面，我们的 NVF 兼具无符号距离函数的优点，能够表示范围广得多的曲面和流形类别，甚至不必是闭合的。

3.2. 特征提取

首先，将输入点云 P 送入特征编码器，为每个点提取三维形状特征，从而得到输出特征图 $F \in \mathbb{R}^{N\times C}$。可以使用多种点云网络 [9, 49, 82] 作为特征编码器，实验部分还对骨干网络进行了消融研究。对于查询点 $\mathbf{q}$，我们基于欧几里得距离搜索，在 $P \in \mathbb{R}^{N\times 3}$ 中找到其 K 个最近点 $\mathbf{p}_i (i = 1, \cdots, k)$，并通过拼接这些最近点 $\mathbf{p}_i$ 的特征签名 z_{qi}，构造 $\mathbf{q}$ 的特征嵌入 z_q。每个特征签名 z_{qi} 都是位置特征（即查询点 $\mathbf{q}$、最近点 $\mathbf{p}_i$ 的位置以及相对位置 $\mathbf{q} - \mathbf{p}_i$）和形状特征（即 $\mathbf{p}_i$ 对应的特征 $f_i \in F$）的非线性映射，由多层感知机实现。从数学上说，查询嵌入 $z_q \in \mathbb{R}^D$ 可表示为

$$\begin{aligned} z_q &= z_{q1} \oplus z_{q2} \oplus ... z_{qi} ... \oplus z_{qk}, \\ z_{qi} &= MLP(\mathbf{q}, \mathbf{p_i}, \mathbf{p_i} - \mathbf{q}, f_i), \quad i = 1, 2, ..., k. \end{aligned} \tag{2}$$

符号 $\oplus$ 表示拼接操作。可以看出，查询嵌入 z_q 包含了查询点 $\mathbf{q}$ 及其在 P 上的 K 最近点的信息。

3.3. 多头码本

获得维度为 D 的嵌入 z_q 后，我们学习一个多头码本 $\mathcal{C} \in \mathbb{R}^{H\times R\times \frac{D}{H}}$，以生成跨对象先验。我们的码本包含 H 个子码本（即多个头）。每个子码本 $\mathcal{C}^h$ （$h = 1, \cdots, H$）包含 R 个离散编码 c_r^h（$r = 1, \cdots, R$），且每个编码的维度为 $\frac{D}{H}$。码本采用随机初始化。

因此，我们沿通道维度将连续嵌入 z_q 划分为 H 个片段。对于每个嵌入片段 z_q^h，我们根据欧几里得距离，在相应的子码本 $\mathcal{C}^h$ 中搜索与其距离最近的编码。这样，我们就可以将连续嵌入 z_q 离散化为由离散片段编码组成的 $\hat{z}_q$。这一过程可表示如下：

$$\begin{aligned} z_q &= z_q^1 \oplus z_q^2 \oplus ... \oplus z_q^h ... \oplus z_q^H, \\ \hat{z}_q &= c_{r_1^*}^1 \oplus c_{r_2^*}^2 \oplus ... \oplus c_{r_h^*}^h ... \oplus c_{r_H^*}^H, \\ r_h^* &= \arg\min_{r\in\{1,...,R\}} ||z_q^h - c_r^h||_2. \end{aligned} \tag{3}$$

与仅使用一个码本 $\mathcal{C} \in \mathbb{R}^{(H*R)\times D}$ 相比，多头码本能够实现 R^H 种排列，从而将特征空间从 $\mathbb{R}^{H*R}$ 扩展至 $\mathbb{R}^{R^H}$，并增强码本表示能力。请注意，最近邻搜索在此处停止梯度。

通过最小化连续嵌入 z 与其离散化结果 $\hat{z}_q$ 之间的距离，可以在训练过程中联合学习多头码本，如下所示：

$$\mathcal{L}_{code} = \sum_{h\in\{1,...,H\}} ||sg(c_{r_h^*}^h) - z_q^h||_2^2 + \beta||sg(z_q^h) - c_{r_h^*}^h||_2^2, \tag{4}$$

其中 sg 表示停止梯度操作，β 是第二项中承诺损失的权重。另外，我们可以通过指数移动平均 [61],更新离散化嵌入 $\hat{z}_q$，进而更新码本中对应的编码（$\hat{z}_q^h \equiv c_{r_h^*}^h$）

$$c_{r_h^*}^h := \gamma c_{r_h^*}^h + (1 - \gamma) z_q^h, \tag{5}$$

其中 γ 是一个介于 0 和 1 之间的值。

3.4. 场预测

最后，我们利用连续嵌入 z_q 及其离散化表示 $\hat{z}_q$ 预测一个向量，作为

查询点 $\mathbf{q}$ 移动到真实表面 $\mathcal{S}$的位移。通过组合嵌入 z_q 和 $\hat{}z_q$，使用多层感知机预测最终的向量场，即 $NVF(\mathbf{q}) \approx \Delta\mathbf{q} = MLP(z_q, \hat{}z_q)$。

3.5. 优化

我们通过最小化预测位移 $\Delta\mathbf{q}$ 与真实位移 $\hat{}\mathbf{q} - \mathbf{q}$ 之间的差异，以及将连续嵌入 z_q 离散化为 $\hat{}z_q$ 所产生的误差，来优化 NVF。向量量化后，我们使用停止梯度操作（即., sg）截断梯度。总而言之，我们的总体目标函数为：

$$\mathcal{L} = \sum_{\mathbf{q}\in Q} ||\hat{\mathbf{q}} - \mathbf{q} - \Delta\mathbf{q}||_1 + \lambda \sum ||z_q - sg(\hat{z}_q)||_2^2. \quad (6)$$

超参数 λ 平衡了两个损失项。承诺损失项 $||sg(z) - z^q||_2^2$ （用于公式 4 ）被 指数移动平均 算法取代，用于更新多头码本中的相应码。

3.6. 表面提取

有符号距离函数和无符号距离函数都将曲面定义为零水平集。有符号距离函数通过移动立方体算法提取曲面 [42]，这是因为通过符号翻转可以轻松捕捉交点。近期，MeshUDF [28]和 CAP-UDF [83] 提出检测相反的梯度方向，以替代有符号距离函数中的符号，并成功将移动立方体算法应用 [42] 于无符号距离函数。需要注意的是，这两种方法学习距离场，并且需要对距离场求导以获得梯度方向。相比之下，我们的 NVF 同样可以使用移动立方体算法提取曲面，同时避免对距离场求导。这是因为我们的 NVF 直接编码了距离场和方向场。表面提取算法将空间划分为网格，并判断相邻角点位于同一侧还是两侧。NVF 预测所有格点的向量 $\Delta\mathbf{q}$，并将其归一化为法向量 $\mathbf{g_i} = \Delta\mathbf{q_i}/||\Delta\mathbf{q_i}||_2$。给定一个格点梯度 $\mathbf{g_i}$ 及其相邻格点的梯度 $\mathbf{g_j}$，算法检查它们的梯度是否具有相反的方向，并为该格点赋予伪符号 s_i。有关更多细节，请参阅 MeshUDF [28]。

4. 实验

4.1. 实验方案

任务。 我们通过四项任务评估框架的有效性：1）类别特定重建，2）类别无关重建，3）未见类别重建，以及 4）跨域重建。首先，我们在第 4.2 节通过类别特定重建，展示 NVF 重建非水密网格的能力。接下来，为评估泛化能力，

方法	CD↓	EMD↓	$F1_{1\times10^{-5}}$	$F1_{2\times10^{-5}}$
输入	0.363	0.707	23.735	41.588
NDF [15]	0.197	1.248	64.116	84.902
GIFS [76]	0.146	0.970	54.867	79.722
Ours	**0.114**	**0.945**	**64.261**	**85.290**

表 1. ShapeNet Cars 上的定量评估。我们使用 ShapeNet“Car”类别的原始数据训练并评估我们的方法。我们的方法优于最先进的基于无符号距离场的方法。

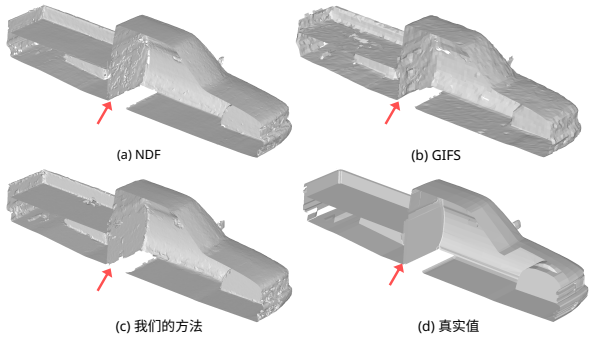


图 4. ShapeNet Cars 上类别特定重建的定性可视化。我们切开形状的部分区域，以便更好地展示内部结构。

我们在第 4.3 节将 NVF 与现有方法在类别无关重建和未见类别重建任务上进行比较。我们还在第 4.4 节通过不经训练或微调直接重建真实扫描数据，测试跨域重建能力。

实现细节。 我们采用 PointTransformer [82] 作为特征编码器，并将 $k = 16$ 设为最近点的数量。多头码本由 4 个子码本组成，每个子码本包含 128 个码和 64 个维度通道。我们将表面提取的网格分辨率设为 256。训练网络时，学习率设为 $l_r = 0.001$，并在之后的 30、70 和 120 步中按 0.3 的因子衰减。损失权重 λ 设为 0.001。

数据集。 大多数实验都在大规模合成数据集 ShapeNet [8] 和真实扫描数据集 MGN [6]上进行。我们在 ShapeNet [8] 上进行类别特定、类别无关和未见类别重建实验，同时在 MGN [6] 上进行跨域重建实验，以展示模型在真实环境中的泛化能力。所有网格都归一化到范围为 $[-0.5, 0.5]$的单位立方体内。

ShapeNet 是一个包含 55 个物体类别的合成数据集。我们遵循 UDF 文献，选择 汽车 进行类别特定重建，并选择 汽车、椅子、飞机和桌子 进行类别无关重建。对于未见类别重建，我们将 音箱、长椅、灯具和水上交通工具 作为测试集。上述用于类别无关重建的 4 个类别称为基类，用于未见类别重建的类别称为新类别，见表 2 和表 3。

方法	Base				新颖			
	CD↓	EMD↓	$F1_{2.5\times10^{-5}}$ ↑	$F1_{1\times10^{-4}}$ ↑	CD↓	EMD↓	$F1_{1\times10^{-5}}$ ↑	$F1_{2\times10^{-5}}$ ↑
输入	0.840	1.045	14.148	25.111	0.800	1.024	17.576	29.815
OccNet [44]	2.766	1.694	30.877	46.644	44.762	4.013	15.943	24.433
IF-Net [14]	0.190	1.120	65.975	85.421	0.596	1.608	61.670	81.106
NDF [15]	0.169	1.538	66.802	84.809	0.169	1.741	65.622	84.069
GIFS [76]	0.179	1.280	56.188	78.458	0.194	1.534	56.644	78.016
Ours	**0.091**	**1.079**	**78.503**	**91.408**	**0.144**	**1.145**	**80.883**	**91.836**

表2。ShapeNet 水密形状上的类别无关重建和未见类别重建的定量结果。我们训练所有在基类上训练模型，分别在基类和新类别上对其进行评估。

方法	Base				新颖			
	CD↓	EMD↓	$F1_{2.5\times10^{-5}}$ ↑	$F1_{1\times10^{-4}}$ ↑	CD↓	EMD↓	$F1_{1\times10^{-5}}$ ↑	$F1_{2\times10^{-5}}$ ↑
输入	0.317	0.867	32.875	51.105	0.289	0.843	39.902	58.092
NDF [15]	0.099	1.372	72.425	88.754	0.093	1.532	76.162	89.977
GIFS [76]	0.118	1.260	64.915	85.115	0.296	1.499	69.252	86.518
Ours	**0.085**	**1.197**	**75.372**	**90.266**	**0.078**	**1.340**	**79.723**	**91.576**

表 3. ShapeNet 非水密形状上的类别无关重建和未见类别重建的定量结果。我们在基类上训练所有模型，并分别在基类和新类别上对其进行评估。

方法	CD↓	EMD↓	$F1_{1\times10^{-5}}$	$F1_{2\times10^{-5}}$
输入	0.124	0.157	52.189	72.969
NDF [15]	0.025	0.216	96.338	98.687
GIFS [76]	0.039	0.192	93.330	97.295
Ours	**0.014**	**0.184**	**98.499**	**99.498**

表 4. 在 MGN 上进行跨域重建的定量结果 [6]。我们基于包含基类的 ShapeNet 训练模型，并在来自 MGN 的原始数据上对其进行评估。

MGN 是一个真实扫描数据集，包含 5 个服装类别（即长外套、裤子、衬衫、短裤和 T 恤），这些类别均为开放曲面。我们在 ShapeNet 上训练模型，并在 MGN 上进行评估，以展示模型在真实场景中的重建能力。结果见第 4.4 节。

指标。 我们使用倒角距离（CD）、地球移动距离（EMD）和 F 分数作为评估指标。具体而言，对于倒角距离（CD $\times10^{-4}$），我们从重建表面采样 100k 个点；对于地球移动距离（EMD $\times10^{-2}$），采样 2048 个点；对于 F 分数（$\times10^{-2}$），采样 100k 个点，阈值为 1×10^{-5} 和2×10^{-5}。

4.2. 类别特定案例

我们首先报告*ShapeNet* 汽车类别特定重建的结果，以展示我们的 NVF 对单一类别的表示能力。我们采用与 NDF 相同的训练/测试划分，并采样 10k 个点作为输入 [15]，与 GIFS [76]相同。表1 中的定量比较表明，与包括 NDF 和 GIFS 在内的先前最先进方法相比，我们的方法取得了更好的结果。具体而言，在倒角距离和地球移动距离方面，我们的 NVF 优于排名第二的 GIFS；在 F分数方面，我们的方法优于排名第二的 NDF。此外，我们在图4中给出了定性比较。可视化结果表明，尽管 NDF、GIFS 和我们的 NVF 都能够恢复内部结构，但我们的 NVF 可以生成更平滑的曲面，并减少精细部位的孔洞（例如，驾驶座椅的靠背）。

4.3. 类别无关和未见类别案例

水密形状上的基准与新颖类别。 我们开展类别无关和未见类别重建实验，以探究 NVF 的泛化能力。为与无法将非水密形状作为输入的先前方法（例如，OccNet [44] 和 IF-Net [14]）进行公平比较，我们报告了经 DISN [71]预处理的水密形状上的定量实验结果，见表 2，并在图5中给出定性可视化结果。我们采用与 IF-Net [14]相同的训练/测试划分。本实验中所有方法的输入均为 3k 个点。

表 2 表明，基于无符号距离场的方法生成的表面重建结果远优于占据表示，尤其是在新类别上。其中，无论是在基类还是新类别上，我们的 NVF 都以显著优势胜过其他方法。图 5 中的定性比较显示，OccNet 和 IF-Net 无法构建许多细节。NDF 和 GIFS 的表现优于它们，但仍无法捕捉某些细节，例如飞机的起落架，或船只上的小凸起。相比之下，我们的方法能够重建平坦且忠实的表面，从而实现更好的视觉效果。

基准类别与新颖类别在非水密形状上的对比。我们还在与上述相同的设置下测试了非水密形状重建，结果见表3和图6。对于训练/测试划分，我们从每个基类中随机抽取 3000 个形状作为训练集，并从每个基类和新类别中各抽取 200 个形状作为验证集和测试集。输入是从原始数据中采样的 10k 个点

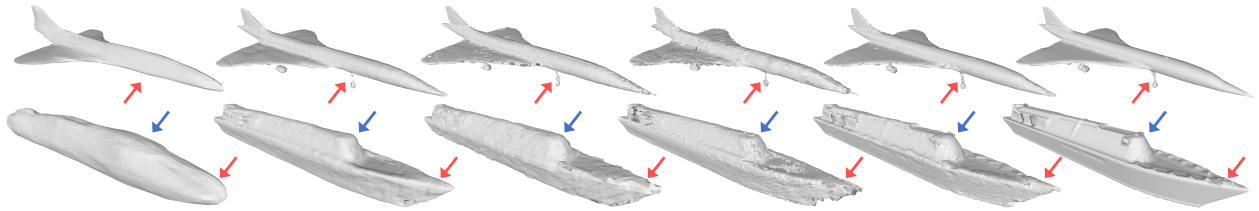


图 5. 类别无关和未见类别形状的重建可视化，展示了来自 ShapeNet 数据集的水密形状。第 1 st 行中的飞机来自基准类别。第 2rd 行中的船只来自新类别。

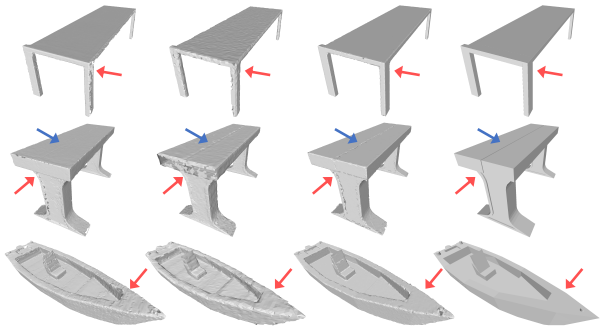


图 6. ShapeNet 数据集中的非水密形状的类别无关与未见类别重建可视化。 1st 行桌子来自基类。 2^{nd}和 3rd 行的长椅与船只来自新类别。

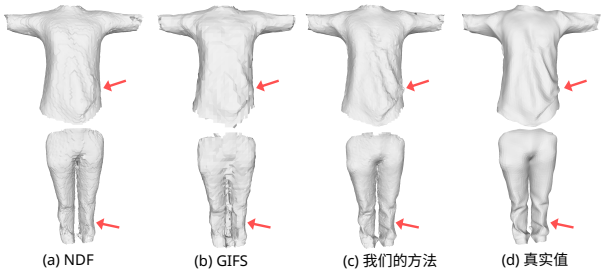


图 7. MGN 数据集上的跨域重建可视化。所有模型均基于包含基类的 ShapeNet 进行训练，并直接在 MGN 上评估。

这些形状的数据。

表3 中的定量比较表明，无论是在类别无关重建还是未见类别重建方面，我们的 NVF 都优于 NDF 和 GIFS。例如，与 NDF 相比，我们在基类上的 $\sim$ 10% 倒角距离降低了，在新类别上的 $\sim$ 15%倒角距离降低了。图6 中的定性可视化进一步支持了这一结果：NDF 和 GIFS 遗漏了一些重要细节（例如，长椅上的接缝），并留下了不少孔洞（例如，桌腿、长椅侧面），而且其曲面不如我们的方法平滑（例如，船只的板面）。

	K	码本	CD↓	EMD↓	$F1_{1\times10^{-5}}$	$F1_{2\times10^{-5}}$
Base	k=8	✗	0.089	1.195	74.139	89.392
	k=8	✓	0.087	**1.172**	75.104	90.057
	k=16	✗	0.121	1.212	73.642	88.962
	k=16	✓	**0.085**	1.197	**75.372**	**90.266**
新颖	k=8	✗	0.080	1.334	78.701	90.972
	k=8	✓	0.083	1.354	79.522	91.434
	k=16	✗	0.081	**1.329**	78.800	91.084
	k=16	✓	**0.078**	1.340	**79.723**	**91.576**

表5. 特征数量 K 和多头码本的影响。多头码本能够提升性能，并在 $K = 16$上取得最佳表现。

4.4 跨域案例

与以往大多数重建方法仅关注域内性能不同，我们还探索了本方法的跨域能力，并设计了跨域重建实验，分别在表中展示定量结果和定性可视化。4和图7。具体而言，我们将基于 ShapeNet 基类训练的类别无关模型直接应用于 MGN [6]以评估其性能。我们从 MGN 的 5 个类别中随机抽取 20个形状，并从原始数据中抽取 3,000个点作为输入。

表4 表明，在所有指标上，我们的 NVF 都大幅优于其他方法。例如，与 NDF 和 GIFS 相比，我们的方法分别将倒角距离降低了近 40% 和 60% 。图 7 中的定性可视化结果也印证了这一观察：只有我们的 NVF 能够忠实地重建衬衫和裤子上的褶皱。

5. 消融研究

我们针对非水密的类别无关和未见类别重建开展了全面的消融实验（设置与第4.3节相同），以评估所提模块（包括多头码本）的有效性，以及超参数（如特征数量 K （即第 3.2节中的最近点数量）对结果的影响。此外，我们还提供了复杂度分析，以证明我们模型的效率。

特征数量。 查询点的特征数量

	骨干网络	CD↓	EMD↓	$F1_{1\times10^{-5}}$	$F1_{2\times10^{-5}}$
Base	NDF [15]	0.099	1.372	72.425	88.754
	三维卷积	0.092	**1.189**	72.415	89.365
	PointNet++	0.088	1.198	74.156	89.418
	PointTransformer	**0.085**	1.197	**75.372**	**90.266**
新颖	NDF [15]	0.093	1.532	76.162	89.977
	三维卷积	0.090	1.342	76.155	90.462
	PointNet++	0.080	**1.339**	78.444	90.850
	PointTransformer	**0.078**	1.340	**79.723**	**91.576**

表 6. 非水密ShapeNet上不同骨干网络的比较。基于点云的骨干网络优于3D卷积骨干网络。PointTransformer表现最佳。

方法	骨干网络	码本	运行时间	内存
NDF [15]	三维卷积	✗	0.75秒	13.44G
Ours	三维卷积	✗	0.27秒	9.21G
	三维卷积	✓	0.34秒	9.21G
	PointTransformer	✗	0.29秒	9.28G
	PointTransformer	✓	0.35秒	9.28G

表 7. 推理分析。运行时间和内存分别是在推理 20 万个查询点期间的时间开销和峰值内存。NVF 在推理运行时间和内存方面效率更高。

（由点云上最近点的数量决定 P）是模型的一个重要因素。我们使用 $k = 8,\ 16$ 在表 5中报告网络的性能，并观察到更大的特征数量 k 能够取得更好的结果，但代价是更高的内存占用。

码本。 我们还在表 5 中展示了所引入的码本机制的有效性。结果表明，使用码本通常可以提升所有指标的整体表现。

训练有效性。我们想强调的是，所提出的多头码本还可作为正则化手段，以缩短训练时间。如图8所示，带有（w/）码本的模型（即，实线）的损失曲线通常比相应不带（w/o）码本的模型（即，虚线）收敛得更快。

3D特征提取。我们还在不同的3D特征提取骨干网络（即，三维卷积 [14, 15, 76] ，PointNet++ [49]和PointTransformer [82]）下报告了我们的方法，见表6。结果表明，在使用相同特征提取骨干网络（例如，三维卷积）时，我们的NVF已能取得与NDF相当（甚至更好）的结果，同时我们的方法在运行时间和内存开销方面更高效。使用PointTransformer可持续提升整体性能。

复杂度。 我们在表格中报告了使用一块 RTX 3090Ti 的机器上，获取 20 万个查询点的距离和方向所需的推理时间及 GPU 内存峰值。 7
结果表明，在使用相同的特征提取骨干网络时，与 NDF 相比，我们的方法仅需 50% 的推理时间和 70% 的内存。这是因为 NDF 中距离场求导会带来额外开销，而我们的方法避免了这一问题，并引入了多

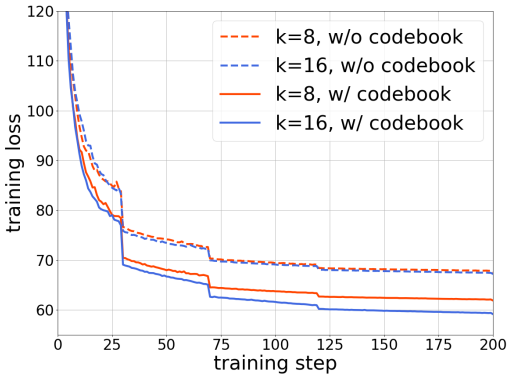


图8. 使用和不使用码本的模型训练曲线。使用码本的模型比不使用码本的模型收敛更快。

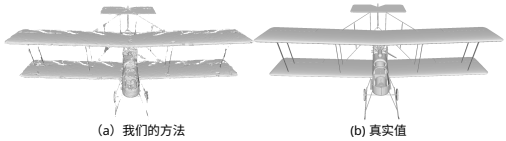


图9. 展示了模型在重建非常薄且复杂的三维结构时的一个失败案例。

头部码本仅带来可忽略的内存和时间开销，如表 7 所示。

局限性。我们的方法能够处理具有任意分辨率和拓扑的水密或非水密形状。然而，如图9所示，对于非常薄或复杂结构的重建仍远不完美。

6. 结论

在这项工作中，我们提出了一种新颖的表面重建表示方法 NVF，它兼具隐式表示和显式学习过程的优势。该方法能够重建高质量的一般物体形状，包括水密形状、非水密形状和多层形状。得益于我们的无微分设计，我们还引入了向量量化并构建了多头码本，以提升跨类别重建的泛化能力。实验表明，我们的方法不仅在不同拓扑结构的重建任务中达到了最先进的性能，还提高了训练和推理效率。

7. 资助致谢

本研究部分由澳大利亚研究委员会（ARC）DP200103223，以及新加坡教育部通过其学术研究基金第二级（MOE-T2EP20220-0007）和第一级（RG14/22）资助。

参考文献

[1] Panos Achlioptas、Olga Diamanti、Ioannis Mitliagkas 和 Leonidas J. Guibas。点云的表征学习与对抗生成。ArXiv，abs/1707.02392，2017年。3[2] Eirikur Agustsson、Fabian Mentzer、Michael Tschannen、Lukas Cavigelli、Radu Timofte、Luca Benini 和 Luc V Gool。用于端到端学习可压缩表征的软到硬向量量化。载于 I. Guyon、U. Von Luxburg、S. Bengio、H. Wallach、R. Fergus、S. Vishwanathan 和 R. Garnett 编，《神经信息处理系统进展》，第30卷。Curran Associates, Inc.，2017年。3[3] Abhishek Badki、Orazio Gallo、Jan Kautz 和 Pradeep Sen。用于三维网格重建的网格簇先验。2020年IEEE/CVF计算机视觉与模式识别会议（CVPR），2020年6月。1[4] Jan Bednarik、Shaifali Parashar、Erhan Gundogdu、Mathieu Salzmann 和 Pascal Fua。通过学习可微表面表示进行形状重建。2020年IEEE/CVF计算机视觉与模式识别会议（CVPR），2020年6月。1[5] F. Bernardini、J. Mittleman、H. Rushmeier、C. Silva 和 G. Taubin。用于表面重建的滚球算法。IEEE可视化与计算机图形学汇刊，5(4)：349–359，1999年。1，2[6] Bharat Lal Bhatnagar、Garvita Tiwari、Christian Theobalt 和 Gerard Pons-Moll。多服装网络：从图像中学习为三维人物穿衣。载于 IEEE国际计算机视觉会议（ICCV）。IEEE，2019年10月。2，5，6，7[7] Edwin Catmull 和 James Clark。任意拓扑网格上的递归生成 B 样条曲面。《计算机辅助设计》，10(6)：350–355，1978年。1[8] Angel X Chang、Thomas Funkhouser、Leonidas Guibas、Pat Hanrahan、Qixing Huang、Zimo Li、Silvio Savarese、Manolis Savva、Shuran Song、Hao Su 等。ShapeNet：内容丰富的三维模型库。arXiv 预印本 arXiv:1512.03012，2015年。2，5[9] R. Qi Charles、Hao Su、Mo Kaichun 和 Leonidas J. Guibas。PointNet：用于三维分类和分割的点集深度学习。2017年IEEE计算机视觉与模式识别会议（CVPR），2017年7月。3，4[10] Jiawen Chen、Dennis Bautembach 和 Shahram Izadi。可扩展的实时体积表面重建。《ACM图形学汇刊》，32(4)，2013年7月。1，3[11] Ting Chen 和 Yizhou Sun。用于端到端嵌入压缩的可微乘积量化。载于 ICML，2020年。3[12] Zhiqin Chen 和 Hao Zhang。为生成式形状建模学习隐式场。2019年IEEE/CVF计算机视觉与模式识别会议（CVPR），2019年6月。2，3[13] An-Chieh Cheng、Xueting Li、Sifei Liu、Min Sun 和 Ming-Hsuan Yang。通过规范映射进行自回归三维形状生成。载于 ECCV，2022年。3[14] Julian Chibane、Thiemo Alldieck 和 Gerard Pons-Moll。特征空间中的隐式函数，用于三维形状重建

……及补全。2020年IEEE/CVF计算机视觉与模式识别会议（CVPR），2020年6月。2, 3, 6, 8[15] Julian Chibane、Aymen Mir 和 Gerard Pons-Moll。用于隐式函数学习的神经无符号距离场。载于《神经信息处理系统进展》（NeurIPS），2020年12月。2, 3, 5, 6, 8[16] Christopher Choy、JunYoung Gwak 和 Silvio Savarese。4D时空卷积网络：Minkowski卷积神经网络。载于2019年IEEE/CVF计算机视觉与模式识别会议（CVPR），第3070–3079页，2019年。3[17] Christopher B Choy、Danfei Xu、JunYoung Gwak、Kevin Chen 和 Silvio Savarese。3d-r2n2：一种用于单视图与多视图三维物体重建的统一方法。载于《欧洲计算机视觉会议》，第628–644页。Springer，2016年。3[18] Boyang Deng、J. P. Lewis、Timothy Jeruzalski、Gerard Pons-Moll、Geoffrey Hinton、Mohammad Norouzi 和 Andrea Tagliasacchi。NASA神经可动形状近似。计算机科学讲义，第612–628页，2020年。3[19] Theo Deprelle、Thibault Groueix、Matthew Fisher、Vladimir Kim、Bryan Russell 和 Mathieu Aubry。学习用于三维形状生成与匹配的基本结构。载于《神经信息处理系统进展》，第7433–7443页，2019年。1[20] Prafulla Dhariwal、Heewoo Jun、Christine Payne、Jong Wook Kim、Alec Radford 和 Ilya Sutskever。Jukebox：一种音乐生成模型。ArXiv预印本 arXiv:2005 00341，2020年。3[21] Ming Ding、Zhuoyi Yang、Wenyi Hong、Wendi Zheng、Chang Zhou、Da Yin、Junyang Lin、Xu Zou、Zhou Shao、Hongxia Yang 和 Jie Tang。CogView：通过Transformer掌握文本到图像生成。载于NeurIPS，2021年。3[22] Philipp Erler、Paul Guerrero、Stefan Ohrhallinger、Niloy J. Mitra 和 Michael Wimmer。Points2Surf：从点云学习隐式曲面。载于 Andrea Vedaldi、Horst Bischof、Thomas Brox 和 Jan-Michael Frahm（编），《计算机视觉——ECCV 2020》，第108–124页，Cham，2020年。Springer国际出版公司。2[23] Songwei Ge、Thomas Hayes、Harry Yang、Xiaoyue Yin、Guan Pang、David Jacobs、Jia-Bin Huang 和 Devi Parikh。使用时间无关VQGAN和时间敏感Transformer生成长视频。载于ECCV，2022年。3[24] Kyle Genova、Forrester Cole、Avneesh Sud、Aaron Sarna 和 Thomas Funkhouser。用于三维形状的局部深度隐式函数。载于20202 IEEE计算机视觉与模式识别会议（CVPR），2019年。3[25] Georgia Gkioxari、Jitendra Malik 和 Justin Johnson。Mesh R-CNN。载于IEEE/CVF国际计算机视觉会议论文集，第9785–9795页，2019年。3[26]Thibault Groueix、Matthew Fisher、Vladimir G Kim、Bryan C Russell 和 Mathieu Aubry。纸浆塑形法：一种学习三维曲面生成的方法。载于IEEE计算机视觉与模式识别会议论文集，第216–224页，2018年。1, 3

[27] Shuyang Gu、Dong Chen、Jianmin Bao、Fang Wen、Bo Zhang、Dongdong Chen、Lu Yuan 和 Baining Guo。用于文本到图像合成的向量量化扩散模型。2022年IEEE/CVF计算机视觉与模式识别会议（CVPR），第10686–10696页，2022年。3[28] Benoit Guillard、Federico Stella 和 Pascal Fua。MeshUDF：无符号距离场网络的快速可微网格化。ArXiv预印本 arXiv:2111.14549，2021年。2, 3, 5[29] Kunal Gupta 和 Manmohan Chandraker。神经网格流：通过微分同胚流进行三维流形网格生成。载于《神经信息处理系统进展》，第abs/2007.10973卷，2020年。1[30] Christian Hane、Shubham Tulsiani 和 Jitendra Malik。用于三维物体重建的层次化曲面预测。2017年国际三维视觉会议（3DV），2017年10月。1, 3[31] Wenyi Hong、Ming Ding、Wendi Zheng、Xinghan Liu 和 Jie Tang。CogVideo：通过Transformer进行文本到视频生成的大规模预训练，2022年。3[32] Qingyong Hu、Bo Yang、Linhai Xie、Stefano Rosa、Yulan Guo、Zhihua Wang、Niki Trigoni 和 Andrew Markham。RandLA-Net：大规模点云的高效语义分割。2020年IEEE/CVF计算机视觉与模式识别会议（CVPR），2020年6月。3[33] Binh-Son Hua、Minh-Khoi Tran 和 Sai-Kit Yeung。逐点卷积神经网络。载于IEEE计算机视觉与模式识别会议论文集，第984–993页，2018年。3[34] Mengqi Ji、Juergen Gall、Haitian Zheng、Yebin Liu 和 Lu Fang。SurfaceNet：用于多视图立体视觉的端到端三维神经网络。载于IEEE国际计算机视觉会议（ICCV）论文集，第2307–2315页，2017年。3[35] Chiyu Jiang、Avneesh Sud、Ameesh Makadia、Jingwei Huang、Matthias Nießner 和 Thomas Funkhouser。用于三维场景的局部隐式网格表示。2020年IEEE/CVF计算机视觉与模式识别会议（CVPR），2020年6月。2, 3[36] Angjoo Kanazawa、Shubham Tulsiani、Alexei A. Efros 和 Jitendra Malik。从图像集合中学习类别特定的网格重建。计算机科学讲义，第386–402页，2018年。3[37] Abhishek Kar、Christian Hane 和 Jitendra Malik。学习多视图立体视觉系统。载于NeurIPS，2017年。3¨ Michael Kazhdan、Matthew Bolitho 和 Hugues Hoppe。泊松表面重建。载于第四届Eurographics几何处理研讨会论文集，第7卷，2006年。1[38] Doyup Lee、Chiheon Kim、Saehoon Kim、Minsu Cho 和 Wook-Shin Han。使用残差量化进行自回归图像生成。2022年IEEE/CVF计算机视觉与模式识别会议（CVPR），2022年6月。3[39] Chen-Hsuan Lin、Oliver Wang、Bryan C. Russell、Eli Shechtman、Vladimir G. Kim、Matthew Fisher 和 Simon Lucey。用于视频对齐的三维物体重建的光度网格优化。2019年IEEE/CVF计算机视觉与模式识别会议（CVPR），2019年6月。3

[41] Yongcheng Liu、Bin Fan、Shiming Xiang 和 Chunhong Pan。用于点云分析的关系形状卷积神经网络。2019年IEEE/CVF计算机视觉与模式识别会议（CVPR），2019年6月。3[42] William E. Lorensen 和 Harvey E. Cline。移动立方体算法：一种高分辨率三维表面构建算法。《SIGGRAPH计算机图形学》，21(4):163–169，1987年8月。1, 5[43] Daniel Maturana 和 Sebastian Scherer。Voxnet：一种用于实时目标识别的三维卷积神经网络。载于 IEEE/RSJ智能机器人与系统国际会议，第922–928页，2015年。3[44] Lars Mescheder、Michael Oechsle、Michael Niemeyer、Sebastian Nowozin 和 Andreas Geiger。占用网络：在函数空间中学习三维重建。载于 IEEE计算机视觉与模式识别会议论文集（CVPR），2019年。2, 3, 6[45] Paritosh Mittal、Yen-Chi Cheng、Maneesh Singh 和 Shubham Tulsiani。AutoSDF：用于三维补全、重建和生成的形状先验。2022年IEEE/CVF计算机视觉与模式识别会议（CVPR），2022年6月。3[46] Junyi Pan、Xiaoguang Han、Weikai Chen、Jiapeng Tang 和 Kui Jia。通过拓扑修改网络从单张RGB图像进行深度网格重建。2019年IEEE/CVF国际计算机视觉会议（ICCV），2019年10月。3[47] Jeong Joon Park、Peter Florence、Julian Straub、Richard Newcombe 和 Steven Lovegrove。DeepSDF：学习用于形状表示的连续有符号距离函数。2019年IEEE/CVF计算机视觉与模式识别会议（CVPR），2019年6月。3[48] Songyou Peng、Michael Niemeyer、Lars Mescheder、Marc Pollefeys 和 Andreas Geiger。卷积占用网络。载于欧洲计算机视觉会议（ECCV），2020年。2[49] Charles Ruizhongtai Qi、Li Yi、Hao Su 和 Leonidas J Guibas。PointNet++：在度量空间中对点集进行深度层次特征学习。《神经信息处理系统进展》，第30卷，2017年。3, 4, 8[50] Anurag Ranjan、Timo Bolkart、Soubhik Sanyal 和 Michael J. Black。使用卷积网格自编码器生成三维人脸。《计算机科学讲义》，第725–741页，2018年。3[51] Ali Razavi、A¨aron van den Oord 和 Oriol Vinyals。使用 VQ-VAE-2 生成多样化的高保真图像。Curran Associates Inc.，美国纽约州雷德胡克，2019年。3[52] Jingjing Ren、Qingqing Zheng、Yuan yu Zhao、Xue miao Xu 和 Chen Li。DLFormer：用于视频修复的离散潜变量变换器。2022年IEEE/CVF计算机视觉与模式识别会议（CVPR），第

3501 3510, 2022 3 –页。

[53] Danilo Jimenez Rezende、S. M. Ali Eslami、Shakir Mohamed、Peter W. Battaglia、Max Jaderberg 和 Nicolas Manfred Otto Heess。从图像中无监督地学习三维结构。载于 NIPS，2016年。3[54] Gernot Riegler、Ali Os man Ulusoy 和 Andreas Geiger。OctNet：以高分辨率学习深度三维表示。2017年IEEE计算机视觉与模式识别会议（CVPR），2017年7月。3

[55] Shunsuke Saito、Zeng Huang、Ryota Natsume、Shigeo Morishima、Hao Li 和 Angjoo Kanazawa。Pifu：用于高分辨率着衣人体数字化的像素对齐隐式函数。2019年IEEE/CVF国际计算机视觉会议（ICCV），2019年10月。3
[56] Shunsuke Saito、Tomas Simon、Jason Saragih 和 Hanbyul Joo。Pifuhd：用于高分辨率三维人体数字化的多层级像素对齐隐式函数。载于 IEEE计算机视觉与模式识别会议论文集，2020年6月。3[57] Frank Steinbr¨ucker、J¨urgen Sturm 和 Daniel Cremers。在中央处理器上实时进行体积三维映射。载于2014年IEEE机器人与自动化国际会议（ICRA），第2021–2028页，2014年。3[58]Xingyuan Sun、Jiajun Wu、Xiuming Zhang、Zhoutong Zhang、Chengkai Zhang、Tianfan Xue、Joshua B Tenenbaum 和 William T Freeman。Pix3D：用于单图像三维形状建模的数据集和方法。载于 IEEE计算机视觉与模式识别会议（CVPR），2018年。3[59] Maxim Tatarchenko、Alexey Dosovitskiy 和 Thomas Brox。八叉树生成网络：用于高分辨率三维输出的高效卷积架构。2017年IEEE国际计算机视觉会议（ICCV），2017年10月。1, 3[60] Hugues Thomas、Charles R. Qi、Jean-Emmanuel Deschaud、Beatriz Marcotegui、Francois Goulette 和 Leonidas Guibas。KPConv：面向点云的灵活可变形卷积。2019年IEEE/CVF国际计算机视觉会议（ICCV），2019年10月。3[61] A¨aron van den Oord、Oriol Vinyals 和 Koray Kavukcuoglu。神经离散表示学习。载于 NIPS，2017年。3, 4[62] Rahul Venkatesh、Tejan Karmali、Sarthak Sharma、Aurobrata Ghosh、R. Venkatesh Babu、Laszlo A. Jeni 和 Maneesh Singh。深度隐式表面点预测网络。2021年IEEE/CVF国际计算机视觉会议（ICCV），2021年10月。3
[63] Jacob Walker、Ali Razavi 和 A¨aron van den Oord。使用 VQ-VAE 预测视频。ArXiv，abs/2103.01950，2021年。3[64] Bing Wang、Zhengdi Yu、Bo Yang、Jie Qin、Toby Breckon、Ling Shao、Niki Trigoni 和 Andrew Markham。RangeUDF：从三维点云进行语义表面重建，2022年。3[65] Nanyang Wang、Yinda Zhang、Zhuwen Li、Yanwei Fu、Wei Liu 和 Yu-Gang Jiang。Pixel2Mesh：从单张RGB图像生成三维网格模型。载于欧洲计算机视觉会议（ECCV）论文集，第52 67页，2018年 3 – 。
[66] Shenlong Wang、Simon Suo、Wei-Chiu Ma、Andrei Pokrovsky 和 Raquel Urtasun。深度参数化连续卷积神经网络。2018年IEEE/CVF计算机视觉与模式识别会议，2018年6月。3[67] Yue Wang、Yongbin Sun、Ziwei Liu、Sanjay E. Sarma、Michael M. Bronstein 和 Justin M. Solomon。用于点云学习的动态图卷积神经网络。《ACM图形学汇刊》，38(5):1–12，2019年11月。3
[68] Francis Williams、Teseo Schneider、Claudio Silva、Denis Zorin、Joan Bruna 和 Daniele Panozzo。《用于表面重建的深度几何先验》。2019年IEEE/CVF计算机视觉与模式识别大会（CVPR），2019年6月。1[69] Jiajun Wu、Yifan Wang、Tianfan Xue、Xingyuan Sun、William T Freeman 和 Joshua B Tenenbaum。《MarrNet：通过2.5D草图进行三维形状重建》。载于《神经信息处理系统进展》，2017年。3[70] Jiajun Wu、Chengkai Zhang、Tianfan Xue、Bill Freeman 和 Joshua B. Tenenbaum。《通过三维生成对抗建模学习物体形状的概率潜在空间》。载于NIPS，2016年。3
[71] Qiangeng Xu、Weiyue Wang、Duygu Ceylan、Radomir Mech 和 Ulrich Neumann。《DISN：用于高质量单视图三维重建的深度隐式表面网络》。载于H. Wallach、H. Larochelle、A. Beygelzimer、F. d'Alch´e-Buc、E. Fox 和 R. Garnett 编，《神经信息处理系统进展》第32卷，第492–502页。Curran Associates, Inc.，2019年。6[72] Xingguang Yan、Liqiang Lin、Niloy J. Mitra、Dani Lischinski、Danny Cohen-Or 和 Hui Huang。《ShapeFormer：通过稀疏表示进行基于Transformer的形状补全》，2022年。3[73] Karren Yang、Dejan Markovic、Steven Krenn、Vasu Agrawal 和 Alexander Richard。《视听语音编解码器：通过重新合成重新思考视听语音增强》。2022年IEEE/CVF计算机视觉与模式识别大会（CVPR），2022年6月。
2022 3 Xianghui Yang、
[74] Guosheng Lin 和 Luping Zhou。《ZeroMesh：零样本单视图三维网格重建》，2022年。3[75] Yaoqing Yang、Chen Feng、Yiru Shen 和 Dong Tian。《FoldingNet：对三维点云进行可解释的无监督学习》。2018年IEEE/CVF计算机视觉与模式识别大会，abs/1712.07262，2017年。3[76] Jianglong Ye、Yuntao Chen、Naiyan Wang 和 Xiaolong Wang。《GIFS：用于通用形状表示的神经隐式函数》。载于IEEE/CVF计算机视觉与模式识别大会论文集，2022年。2、3、5、6、8[77] Jiahui Yu、Xin Li、Jing Yu Koh、Han Zhang、Ruoming Pang、James Qin、Alexander Ku、Yuanzhong Xu、Jason Baldridge 和 Yonghui Wu。《改进版VQGAN的向量量化图像建模》。国际学习表征会议，abs/2110.04627，2022年。3[78] Wentao Yuan、Tejas Khot、David Held、Christoph Mertz 和 Martial Hebert。《PCN：点云补全网络》。2018年国际三维视觉大会（3DV），2018年9月。3[79]Manzil Zaheer、Satwik Kottur、Siamak Ravanbakhsh、Barnab´as P´oczos、Ruslan Salakhutdinov 和 Alex Smola。《深度集合》。神经信息处理系统进展，abs/1703.06114，2017年。3[80] Biao Zhang、Matthias Nießner 和 Peter Wonka。《3DILG：用于三维生成建模的非规则潜在网格》。NIPS，abs/2205.13914，2022年。3[81] Fang Zhao、Wenhao Wang、Shengcai Liao 和 Ling Shao。《通过梯度方向对齐学习锚定无符号距离函数，以实现单视图服装重建》。载于IEEE/CVF国际计算机视觉会议论文集，第12674–12683页，2021年。3

[82] Hengshuang Zhao、Li Jiang、Jiaya Jia、Philip Torr 和 Vladlen Koltun。《点云 Transformer》。*2021*年*IEEE/CVF*国际计算机视觉会议（*ICCV*），2021年10月。3，4，5，8[83] Junsheng Zhou、Baorui Ma、Liu Yu-Shen、Fang Yi 和 Han Zhizhong。《从原始点云逐步学习一致性感知无符号距离函数》。载于《神经信息处理系统进展》（*NeurIPS*），2022年。2，3，5[84] Yin Zhou 和 Oncel Tuzel。《VoxelNet：用于基于点云的三维目标检测的端到端学习。*2018*年*IEEE/CVF*计算机视觉与模式识别大会，2018年6月。3