ARAP-GS：基于拖动的尽可能刚性三维高斯图编辑与扩散先验

Xiao Han[1,*] Runze Tian[1,*] Yifei Tong[1,*] Fenggen Yu[2] Dingyao Liu[1]
Yan Zhang[1] [1]Nanjing University [2]Simon Fraser University
hanxiao@smail.nju.edu.cn

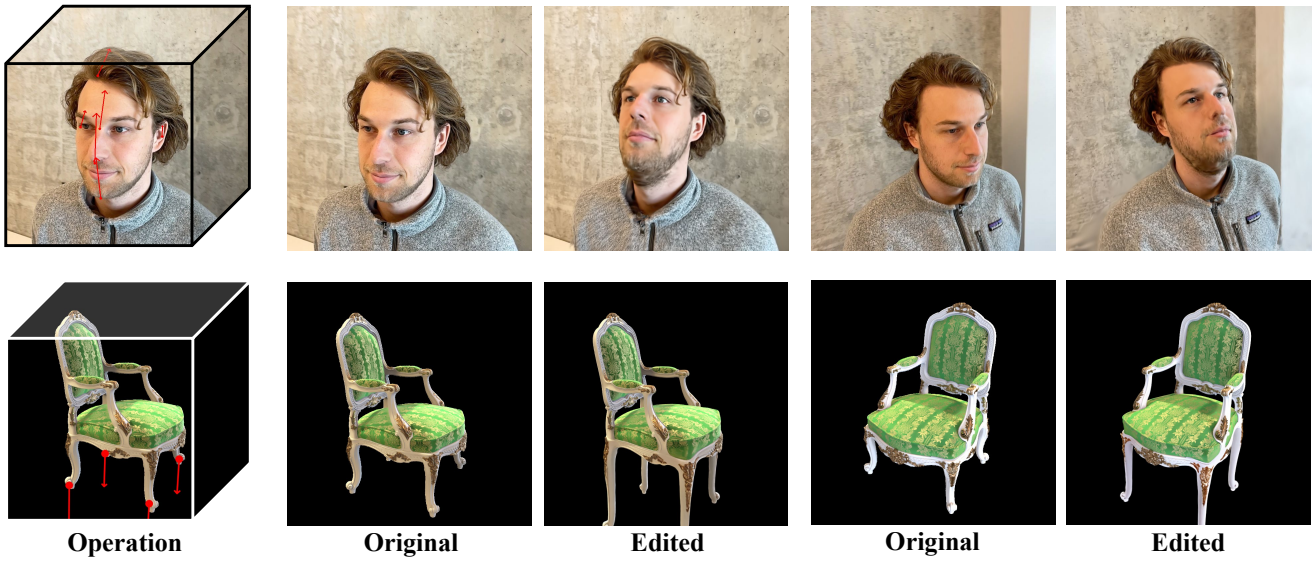

图 1. ARAP-GS **的结果**。给定一组控制点及其变形，ARAP-GS 可以高效地实现拖动驱动的 3DGS 编辑。我们的方法通过对 3DGS 场景进行旋转（上方）或拉伸（下方）来变形几何形状，同时保持原始外观和多视图一致性。第一列说明了拖动操作，红色点表示控制点，箭头表示拖动方向。“原版”和“编辑后”分别表示编辑前后的渲染结果。

Abstract

拖动驱动的编辑因其能够通过简单直观的操作修改复杂的几何结构而受到设计师的欢迎，使用户无需具备高级技术技能即可调整和重塑内容。这种拖动操作已被整合到多种方法中，以促进设计中的 2D 图像和 3D 网格编辑。然而，很少有研究探讨广泛使用的 3D 高斯点云（3DGS）表示形式的拖动驱动编辑，因为在保持形状一致性和视觉连续性的同时变形 3DGS 仍然具有挑战性。在本文中，我们介绍了一种基于尽可能刚性的（ARAP）变形的拖动驱动 3DGS 编辑框架——ARAP-GS。与之前的 3DGS 编辑方法不同，我们是第一个将 ARAP 变形直接应用于 3D 高斯的方法，从而实现了灵活的、拖动驱动的几何变换。为了在变形后保持场景外观，我们在迭代优化过程中加入了先进的扩散先验用于图像超分辨率。这种方法不仅增强了视觉质量，还确保了编辑结果中的多视图一致性。实验表明，ARAP-GS 在各种 3D 场景中均优于现有方法，证明其拖动驱动的 3DGS 编辑的有效性和优越性。此外，我们的方法非常高效，在单个 RTX 3090 GPU 上对一个场景进行编辑只需 10 到 20 分钟。

*对该工作有同等贡献。

1. 介绍

最近，三维高斯散斑（3DGS）[24] 在三维视觉领域引起了广泛关注，由于其显式表示和实时渲染能力，在新型视图合成、基于图像的三维重建及其他视觉任务方面取得了突破性进展。随着技术的发展，它已扩展到各种任务中，包括动态场景建模 [27, 36]、三维场景理解 [2, 52] 和三维场景编辑 [12, 45]。

现有的 3D 场景编辑方法旨在为用户提供一个精确且直观的编辑框架，使他们能够自由调整场景中的每一个细节。这不仅包括纹理编辑，还包括灵活的几何变形。尽管在 3D 网格上的 3D 场景编辑已经得到了广泛的研究，但很少有方法可以直接应用于 3DGS 表示，因为在保持形状一致性和视觉连续性的同时变形 3DGS 仍然是一个挑战。最近，研究人员探索了各种针对 3DGS 的编辑操作，例如文本驱动的 3DGS 编辑 [62, 63, 72]，单目视频驱动的 3DGS 编辑 [7, 21] 和笼子驱动的 3DGS 编辑 [19]。与这些工作不同，我们的目标是开发一种高效的拖动驱动 3DGS 编辑方法，该方法能够实现灵活的几何变换应用于 3DGS，如图 Fig. 1 所示。

拖动驱动的编辑因其直观且用户友好的方法在修改复杂几何结构方面受到设计师的欢迎，允许用户轻松调整和重塑内容而无需太多技术专业知识，并已在三维网格编辑中得到广泛应用 [9, 11, 28, 55]。随着二维生成模型的快速发展 [15, 47]，也出现了一些拖动驱动的图像编辑方法，如 DragGAN [46] 和 DragDiffusion [53]。这些方法提出了一种革命性的交互式图像操作框架，使用户能够“拖动”任意点到目标位置，提供对姿势、形状、表情和布局的精确控制。然而，合成多视角图像以保持几何一致性效果不佳，因为直接编辑高斯属性更有效地利用了 3DGS 的显式表示。

在本文中，我们提出了 ARAP-GS，一种基于 As-Rigid-As-Possible (ARAP) 变形 [55] 且具有扩散先验的直观拖动驱动 3DGS 编辑方法。我们的方法包括两个阶段：一个 ARAP 3DGS 变形阶段和一个带有扩散先验的 3DGS 微调阶段，如 Fig. 2 所示。在第一阶段，我们将初始 3D 高斯转化为一组代表性的子集。然后直接对这个子集应用用户给定的拖动操作进行 ARAP 变形。其余的 3DGS 基于转换后的子集的插值进行变形。然而，诸如颜色和不透明度等属性保持不变，这可能导致渲染图像中出现伪影。为了解决这个问题，在第二阶段，我们使用迭代优化策略对遮罩下的 3D 高斯属性进行微调，并利用预训练的 2D 扩散基础超分辨率模型的能力来修正编辑内容并提高视觉质量。实验表明了我们的拖动驱动 3DGS 编辑方法在各种场景和任务中的有效性。与以前的方法相比，我们的方法显著提高了结果的视觉质量和几何完整性。总之，我们的贡献包括：

- 我们提出了一种基于 ARAP 变形的拖动驱动三维高斯形状 (3DGS) 编辑方法，命名为 ARAP-GS。据我们所知，这是首次将 ARAP 直接应用于三维高斯分布的方法，实现了无需额外监督数据的自由形式 3DGS 变形。
- 我们引入了一种有效的迭代优化策略，该策略利用 2D 扩散先验来细化 ARAP 变形后的遮罩 3DGS，从而提高多视图一致性和视觉质量。
- 大量实验表明，我们的方法在各种 3DGS 场景中比现有方法更易于使用且具有更高的视觉质量。

2. 相关工作

2.1. 3D 高斯图元编辑

NeRF [39] 和 3DGS [24] 都是新型视图合成的流行表示方法，而 NeRF 隐式地在 MLPs 中表示场景，3DGS 使用显式表示。尽管有许多针对 NeRF 的编辑方法 [16, 35, 58, 71]，但由于表示方式的不同，它们无法应用于 3DGS。随着 3DGS 在计算机视觉中的日益流行，也有一些工作出现以支持对 3DGS 进行编辑。文本引导的 3D 编辑方法 [12, 45, 56, 59, 62, 63, 66, 72] 因其在修改 3DGS 方面易于使用而脱颖而出。然而，它们仅通过文本提示来编辑 3DGS，并且面临一个常见问题：编辑结果往往导致语义变化，限制了它们对几何和纹理细节进行编辑的能力。

一些其他的方法 [7, 21, 37] 允许基于从单目视频中提取的充分先验知识进行动态几何变形。但它们需要额外的视频监督，并且对数据集提供的相机视角敏感，限制了它们编辑更广泛类别的能力。还有一些同期工作集中在拖拽驱动的 3DGS 编辑 [10, 51] 上，但是它们都把拖拽操作投影到 2D 图像上并使用高级模型 [46, 53] 来编辑这些图像，然后再根据编辑后的 2D 图像优化 3D 场景。由于先进的 2D 扩散模型，这些方法显示出了令人印象深刻的生成能力。然而，[51] 使用不同编辑图像之间的不一致性，这可能会导致优化后的 3D 场景中出现模糊区域。此外，每个视角的 2D 编辑结果可能不会

严格遵循用户的指令 [10]。与这些工作不同的是，我们的方法直接将编辑操作应用于 3DGS，然后使用 2D 扩散先验来微调蒙版区域，展示出更好的 3D 一致性和编辑后的视觉质量。

2.2. 传统 3D 变形

传统的三维变形方法在计算机图形学中贡献了各种优化技术，如自由形式变形 [50] 、多分辨率变形 [25] 和弹性变形 [5] 。然而，它们缺乏精确控制的能力，在变形过程中导致局部细节的丢失或扭曲。基于笼子的变形 (CBD) [22, 23, 32, 48] 已被提出以支持高质量变形。然而，创建和操作笼子顶点是复杂且繁琐的，需要较高的计算成本。

另一条研究路线基于物理启发的能量，提供了更有前景的结果和直观的操作 [54, 57, 65] 。ARAP [55] 是该领域最受欢迎的算法之一，它提出了一种基于最小化局部刚性偏差原理的迭代网格变形框架，使直观且保留细节的变形成为可能。其扩展版本 [9, 11, 28] 通过使用替代能量公式提高了鲁棒性。由于 ARAP 具有直接解释和满足句柄约束的同时易于优化的特点，它已被应用于各种任务 [3, 20, 26, 42] 。据我们所知，我们是第一个将 ARAP 变形直接应用于三维高斯，并实现拖动驱动的 3DGS 编辑的研究。

2.3. 二维图像编辑

扩散模型 [18, 43, 47, 49] 在图像生成方面展现了令人印象深刻的结果。嵌入到像稳定扩散 [47] 这样的模型中的扩散先验可以应用于各种下游任务，例如图像修复 [13, 30, 61, 68] 、图像超分辨率 [31, 60, 67] 和图像编辑 [1, 6, 17, 29, 38] 。虽然基于扩散模型的文本驱动图像编辑对于用户来说较为友好，但完全通过文本传达编辑意图可能会有难度，导致结果不符合预期。

一些方法 [14, 33, 40, 44] 识别了这一限制，并开始专注于使用自由形式操作编辑图像细节。DragGAN [46] 引入了一种基于生成对抗网络（GANs）[15] 的拖动驱动图像编辑方法，该方法利用点操控实现。给定一组把手点和目标点，通过将把手点移动到各自的目标点来编辑图像。DragDiffusion [53] 将编辑框架扩展到了扩散模型，使得在图像编辑中可以实现更精确的空间控制，并且生成能力显著提高。需要注意的是，由于二维视图的歧义性和多视角的一致性问题，直接将这些拖动驱动的编辑方法从 2D 扩展到 3D 是具有挑战性的。

3. 方法

ARAP-GS 将预训练的 3DGS 场景作为输入，以及 N 个把手点 $h \in \mathbb{R}^{N\times 3}$ 和相应的拖动操作。然后从这些操作中获得目标位置 $h' \in \mathbb{R}^{N\times 3}$。我们的目标是在保持原始场景刚性的同时，尽可能维持内容相似度，将围绕把手点的 3D 高斯重新定位到其目标位置。这一过程涉及 ARAP 3DGS 变形以及带有扩散先验的外观微调。在以下内容中，我们首先回顾 Sec. 3.1 中的 ARAP 3D 网格变形和 3DGS 的初步知识，接着介绍我们提出的方法，包括 Sec. 3.2 中基于 ARAP 的 3DGS 变形，以及在 Sec. 3.3 中使用扩散先验进行微调。

3.1. 初步的

ARAP 3D **网格变形**。ARAP 变形 [55] 是一种用于操纵三维网格的技术，旨在在整个变形过程中保持局部刚性。此目标通过最小化以下能量函数实现：

$$E = \sum_{i=1}^{n} w_i \sum_{j\in\mathcal{N}(i)} w_{ij} \|(p'_i - p'_j) - R_i(p_i - p_j)\|^2, \quad (1)$$

其中 $\mathcal{N}(i)$ 表示顶点 i 的邻居索引，p 和 p' 分别表示变换前后的顶点坐标，R 是旋转矩阵，而 w_i 和 w_{ij} 是一些固定的单元和边权重。能量通过迭代过程被最小化。初始时，$p' = p$ 和 R 设置为单位矩阵。给定从拖动操作中更新的位置集合 $\{p'_0, p'_1, \ldots, p'_N\}$，相应的旋转矩阵 $\{R_0, R_1, \ldots, R_N\}$ 通过边长的协方差矩阵的奇异值分解（SVD）进行更新。随后，其他点的位置通过求解以下方程进行更新：

$$\sum_{j\in\mathcal{N}(i)} w_{ij}(p'_i - p'_j) = \sum_{j\in\mathcal{N}(i)} \frac{w_{ij}}{2}(R_i + R_j)(p_i - p_j), \quad (2)$$

并且其他点的旋转矩阵将相应地进行更新。这个过程迭代重复直到收敛。

3D **高斯图元法**。3D 高斯点阵 [24] 是一种显式表示方法，它使用一组各向异性的 3D 高斯函数来表示场景，记为 $G = \{g_1, g_2, \ldots, g_N\}$，其中 $g_i = \{\mu, \Sigma, c, \alpha\}$，$i \in \{1, \ldots, N\}$。在此公式中，$\mu$ 表示高斯的中心位置，Σ 是 3D 协方差矩阵，c 是用球谐系数编码的颜色，而 α 表示不透明度。为了确保 Σ 是半正定的并允许独立优化，将其分解为 $\Sigma = RSS^TR^T$，其中 R 是一个可以

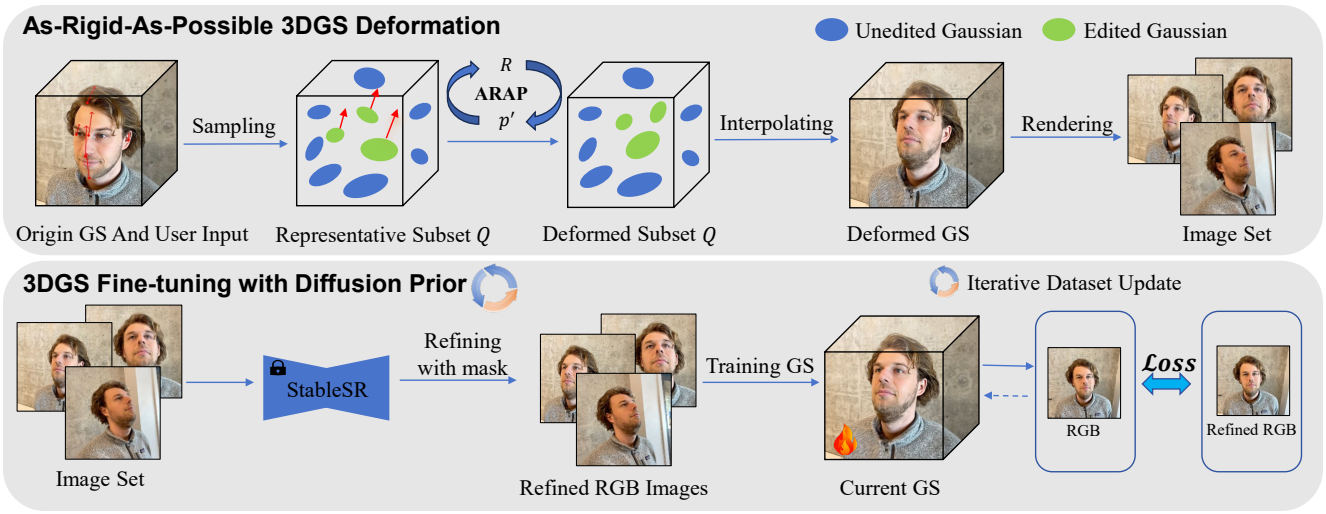

图 2. **方法概述**。我们的方法分为两个阶段实施。在第一阶段，对于编辑过程中的几何变形，我们利用了三维高斯的显式表示，建立了一个代表子集 Q。然后我们将传统的 ARAP 变形直接应用于 Q 中的 3D 高斯，并为每个 3D 高斯获得旋转矩阵 R 和新位置 p'。剩余 3D 高斯的旋转矩阵和新位置则是通过插值那些最近邻在 Q（Sec. 3.2）中的来进行推断。变形改变了三维高斯的位置和协方差。为了进一步更新诸如颜色和尺度值等属性，我们在第二阶段对变形后的 3D 高斯进行了微调。具体来说，我们利用二维扩散先验来去除渲染图像中的伪影，并基于增强的图像迭代优化三维高斯分布（Sec. 3.3）。

轻松转换为四元数 q 的旋转矩阵，而 S 是一个对角缩放矩阵。

3.2. 尽可能刚性的 3DGS 变形

为了实现用 3DGS 表示的场景的几何变形，有两个主要挑战：(1) 提供用户友好的操作，如文本提示或稀疏控制点，来驱动 3DGS 变形；(2) 在 3DGS 几何变形过程中有效保持对象刚性的同时支持灵活编辑操作，例如平移、拉伸和旋转。本文重点讨论拖动驱动的变形，这种变形方式对于修改复杂的几何结构对用户友好，允许用户轻松调整和重塑内容而无需太多技术专业知识。为了在变形过程中维持 3DGS 的刚性，我们采用了传统 ARAP 3D 网格变形 [55] 的尽可能刚性的原则，并直接将其应用于 3DGS 进行变形。详细的方法如下。

如 Eq. (1) 所示，能量最小化过程旨在尽可能保持顶点之间的边长为常数。基于这一见解，我们认为在变形后保留局部细节应尽量减少三维高斯分布之间距离的变化。因此，将三维高斯的中心视为 ARAP 变形中的顶点以进行能量最小化是很自然的，从而将 ARAP 的概念从 3D 网格扩展到 3DGS。

另一方面，由于场景中存在大量的 3D 高斯分布，直接求解所有 3D 高斯分布的变换在计算上非常耗费资源，并且受到时间和内存限制。为了解决这个问题，我们对原始的 3D 高斯分布集合进行随机采样，创建一个具有代表性的子集 Q。然后我们将 K-最近邻（KNN）应用于 Q，以建立相邻关系并将其保留给 Q 中的 3D 高斯分布。我们初始化 p' 和 R 如 Sec. 3.1 所示。随后，我们按照 Sec. 3.1 中所述的方法迭代更新旋转矩阵 R_i 和目标中心 p'_i，对于每个 3D 高斯分布。求解 R_i 时，令 $\mathcal{N}(i)$ 表示 p_i 的邻居的索引，我们得到协方差矩阵 $S_i = P_i D_i P_i'^T$，其中 $D_i \in \mathbb{R}^{|\mathcal{N}(i)|\times|\mathcal{N}(i)|}$ 是由 w_{ij} 组成的对角矩阵，即 $j \in \mathcal{N}(i)$。$P_i \in \mathbb{R}^{3\times|\mathcal{N}(i)|}$ 是一个包含 3D 高斯中心距离 $p_i - p_j$ 作为其列的矩阵，P'_i 类似定义。然后我们对 S_i 执行 SVD：$S_i = U_i \Sigma_i V_i^T$，并获得 $R_i = V_i U_i^T$。R_i 用于更新其他 3D 高斯分布的中心和旋转四元数，这是通过使用与 Eq. (2) 相同的方程实现的。数学的全面推导见于 [55]。

在转换了 Q 中的三维高斯分布后，我们专注于通过插值来变换剩余的三维高斯分布。对于每个高斯分布，我们从 Q 中选择 k 个最近的三维高斯分布，并应用线性插值以获得相应的旋转 q'_l 和变换 p'_l：

$$p'_l = \sum_{i=1}^{k} w_{il} p'_i + p_l, \tag{3}$$

$$q'_l = \sum_{i=1}^{k} w_{il} q'_i \otimes q_l, \tag{4}$$

$$w_{il} = \frac{\exp(-\|(p'_i - p_l)\|_2)}{\sum_{i=1}^{k} \exp(-\|(p'_i - p_l)\|_2)}, \tag{5}$$

其中，p_l 和 q_l 表示原始中心和四元数，l、p'_i 和 q'_i 是索引剩余的 3D 高斯分布，这些在 Q 中发生了变形，而 $\otimes$ 表示四元数乘法。

3.3. 使用扩散先验的 3DGS 微调

感谢 3DGS 上的 ARAP 变形，场景的几何形状得到了很好的保留，但不透明度和颜色等属性仍未得到优化，这可能会导致渲染图像中出现轻微的伪影。为了解决这个问题，我们利用基于扩散的 2D 图像增强技术的最新进展，应用现成的 StableSR [60] 模型来细化渲染图像集。借助第一阶段建立的几何完整性，StableSR 可以有效去除光线伪影并生成高分辨率图像。然而，我们也观察到它可能会带来不同视角之间的不一致区域。因此，我们设计了几种策略来减少不一致视图的影响。

迭代数据集更新。 为了减少来自 StableSR 的增强视图对不一致性的影响，我们采用了来自 Instruct-NeRF2NeRF (I-N2N) [16] 的迭代数据集更新 (IDU) 策略。具体来说，在微调过程开始时，我们将每个视角的原始捕获图像单独存储。每 $t = 10$ 次迭代，我们选择一个包含 n 个视图的子集，并使用 StableSR 来增强它们。这些增强后的图像替代原始图像作为那些视角的监督。这种简单的策略有效地改善了 3DGS 外观，同时保持场景的几何属性，与一次性更新所有视图进行微调相比。

蒙版引导的 3DGS 微调。 在 3DGS 微调过程中，我们实现额外机制以进一步避免 StableSR 在多个视图中产生的不一致区域。想法是通过屏蔽编辑区域来减少校正范围，从而尽可能保留原始视角的一致性。该掩码通过计算 GS 位置的位移、识别超过指定阈值的点并将它们投影到摄像机平面上生成。因此，更新后的视图图像表示如下：

$$I_{merge} = M \odot I_{sr} + (1 - M) \odot I_{gt}, \tag{6}$$

其中 M 表示掩码，I_{sr} 表示增强图像，I_{gt} 表示真实值，而 $\odot$ 表示逐元素乘积。然后使用 I_{merge} 来计算 vanilla 3DGS 的损失。

4. 实验

4.1. 实验设置

数据集。我们使用了 10 个场景进行评估，从真实场景到单一物体不等。这些场景收集自 [4, 16, 39, 41]。**实现细节**。我们在单个 RTX 3090 上进行实验。原始场景的大小范围从 500k 到 3000k 高斯分布。考虑到某些场景中的高斯分布远离需要编辑的部分，我们手动设置了一个掩码来过滤它们。代表性的高斯子集的大小是 $N = 16384$，用于构建邻域关系的 KNN 被设定为 $k = 32$，插值中使用的邻居数量为 8。ARAP 迭代的最大次数为 16。在迭代数据集更新过程中，我们遵循 I-N2N [16] 设置，每隔 10 次迭代更新一个视点。3DGS 优化迭代的次数范围从 800 到 2000，这取决于初始高斯分布的数量。更多关于敏感参数的细节可以在补充材料中找到。

基线。据我们所知，目前没有公开可用的拖动驱动的 3D 图形编辑方法。因此，我们选择了两种具有代表性的文本驱动的 3D 图形编辑方法进行比较，分别是 I-N2N [16] 和 GaussianEditor (GSEditor) [12]。此外，我们还选择了两种具有代表性的拖动驱动的 2D 图像编辑方法作为基线，即 DragDiffusion (DragDiff) [53] 和 SDEDrag [44]。具体来说，我们将 2D 编辑方法应用于 3DGS 生成的渲染视图上，然后利用编辑后的结果来微调 3DGS 的参数。

运行时间。我们的编辑时间取决于优化迭代次数和场景大小。对于一个大约有 50 万个高斯分布需要编辑的场景，ARAP 大约需要 6 分钟，而 1000 次精细调整优化则大约需要 5 分钟。相比之下，I-N2N 需要大约一个小时，GSEditor 需要大约 20 分钟。对于每个视角，DragDiff 平均需要 5 分钟，而 SDEDrag 平均需要 15 分钟。它们的总时间取决于视角的数量，在最快的情况下至少需要 2 小时。

评估指标。我们采用拖动准确率指数（DAI）[70] 进行定量评估。DAI 定义如下：

$$DAI = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{m} \frac{\|\Omega(I_{o,i}, p_{i,j}, \gamma) - \Omega(I_{e,i}, q_{i,j}, \gamma)\|_2^2}{(1 + 2\gamma)^2}, \tag{7}$$

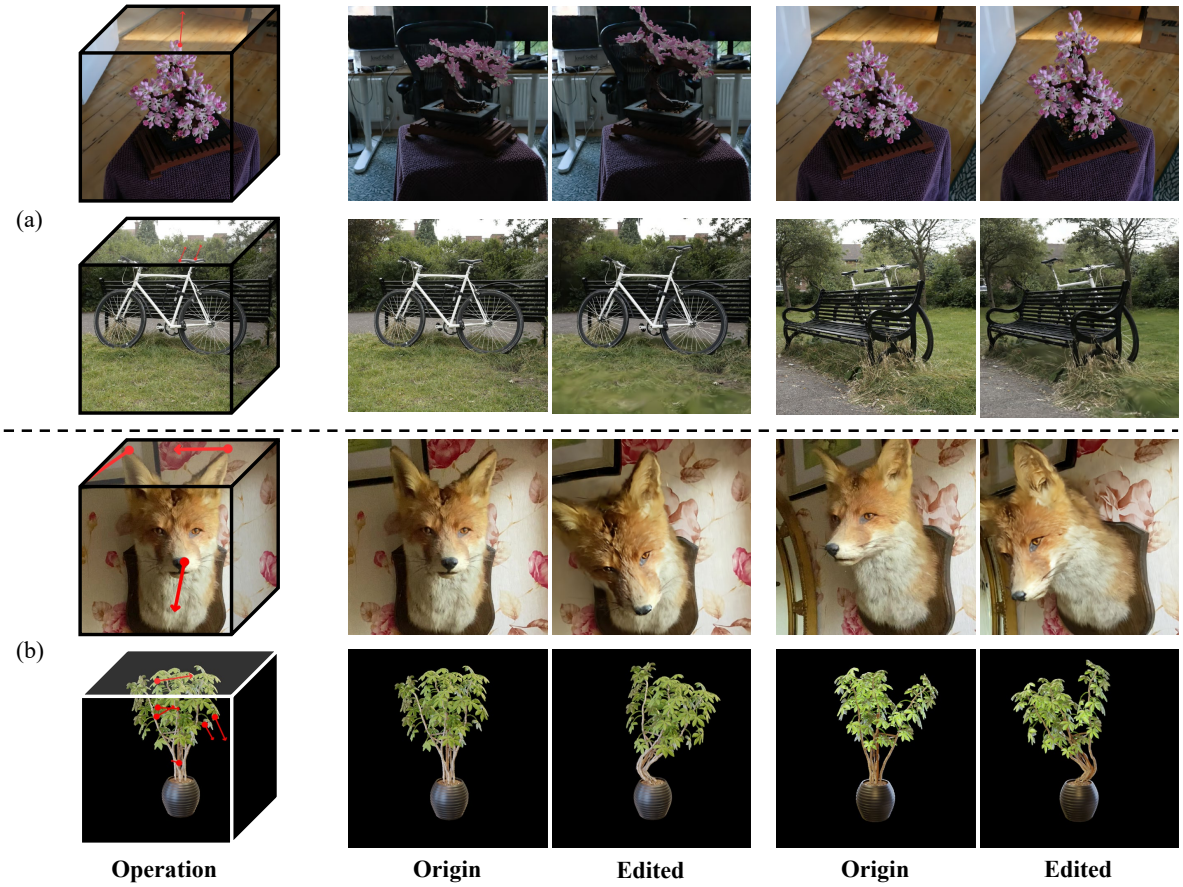

图 3. **我们的方法的可视化结果**。前两行显示了拉伸的结果，后两行显示了旋转的结果。第一列显示了拖动操作，红色点表示把手点，箭头表示拖动方向。“原始”和“编辑后”分别表示编辑前后的渲染结果。

其中，n 表示视图数量，m 表示手柄点数量，$I_{o,i}$ 和 $I_{e,i}$ 分别表示在 i-th 视图上的编辑前和编辑后图像，$p_{i,j}$ 和 $q_{i,j}$ 分别表示投影到相机平面的手柄点和目标点，$\Omega(I_{o,i}, p_{i,j}, \gamma)$ 表示以 $p_{i,j}$ 为中心、大小为 γ 的 $I_{o,i}$ 面积的正方形补丁。我们随机选取 $n = 10$ 个视角进行评估，并将 $\gamma = \{1, 5, 10, 20\}$ 设为平均不同层次的上下文。DAI 衡量了一种方法在将源内容转移到目标点的有效性。考虑到编辑质量与用户感知密切相关，我们在编辑结果上进一步进行了用户研究和 GPT 评估。我们从用户研究中的 60 名参与者那里收集投票，并且不同的方法编辑结果随机出现在每个问题中。由于 [64] 提出 GPT 的图像评价能力与人类感知一致，我们将图像呈现给 GPT-4o 并要求它评估编辑的有效性，得到一个从 1 到 10 的评分。每种方法的平均分也在 Tab. 1 中报告。

4.2. 定性结果

我们展示了来自不同拖动操作的 Fig. 3 方法编辑结果。如前两行所示，红色箭头指示了拉伸方向。很明显，我们的方法实现了符合用户意图的拉伸，并在不同的视角下保持视觉一致性。最后两行在 Fig. 3 中展示了旋转变换，红色箭头指示了旋转方向。通过标注几个控制点并指定旋转角度，3D 场景可以实现视觉上合理的旋转变形。

如 Fig. 4 所示，虽然 I-N2N 和 GSEditor 可以执行基于文本的 3D 编辑，但由于其有限的文字理解能力，它们难以处理复杂的动作命令，例如“向上看”或“拉伸”。因此，它们经常无法进行准确的几何变换。这一限制凸显了我们几何 3DGS 变形策略的优势。虽然 DragDiff 和 SDEDrag 在 2D 中实现了令人满意的几何编辑，但由于缺乏跨视点的空间一致性，在应用于

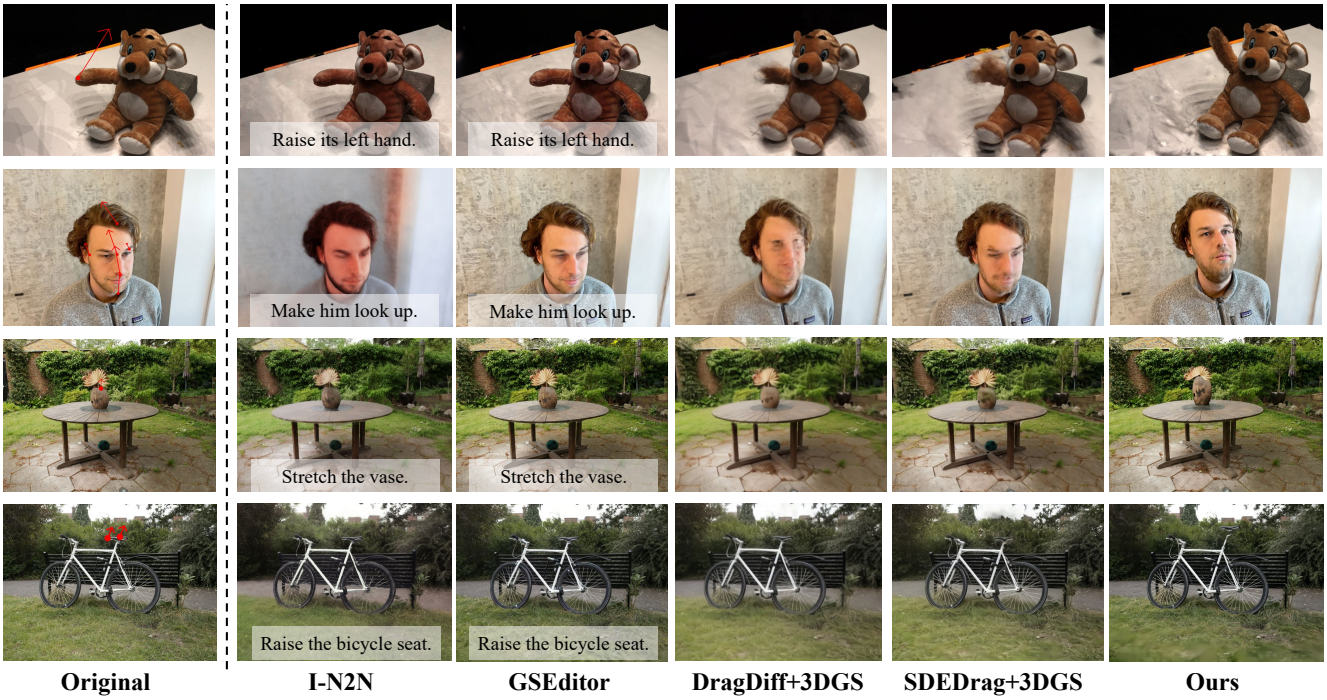

图 4. **定性比较结果**。我们将我们的方法与包括 I-N2N [16]、GaussianEditor [12]、DragDiffusion [53] 和 SDEDrag [44] 在内的最先进的方法进行了比较。相比这些方法，我们的方法实现了更精确的几何变形和更好的多视图一致性，提供了更高的视觉质量结果。

3D 时它们未能产生一致的编辑效果，这导致优化后的 3DGS 出现伪影。相比之下，由于直接进行 ARAP 3DGS 变形和基于扩散先验的视角一致图像增强，我们的方法在变形后能保持更好的外观。更多定性结果请参见补充材料。

4.3. 定量结果

定量比较结果如 Tab. 1 所示。我们的方法在所有度量标准上都获得了最高分数，与其他方法相比，在拖动驱动的 3DGS 编辑任务中表现更优。DAI 得分表明，我们的方法在变形后有效地保持了源位置和目标位置之间的内容相似性。用户研究表明，参与者明显偏好我们的方法，而其他方法在拖动驱动的 3DGS 编辑任务上表现不足。此外，GTP-4o 得分进一步验证了我们的结果取得了更好的效果，并且与输入变形的一致性更高。

4.4. 消融研究

我们主要讨论了我们的方法中两种主要设计的影响：ARAP 变形和扩散先验。更多关于其他设计的消融研究包含在补充材料中。

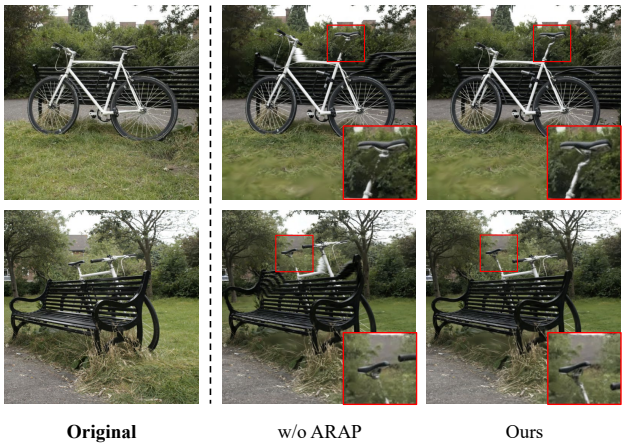

图 5. ARAP **变形的有效性**。我们选择两个视图来比较有无 ARAP 变形的结果。我们将放大变形区域，以突出我们的方法在变形后保持场景几何结构的能力。

ARAP **变形的有效性**。为了验证 ARAP 变形的有效性，我们将其与直接插值进行了比较，在直接插值中，其他高斯函数根据最近的输入控制点和变形直接进行线性

	I-N2N	GSEditor	DragDiff+3DGS	SDEDrag+3DGS	Ours
DAI(↓)	0.3249	0.2994	0.2811	0.3037	0.0968
User Votes(↑)	5.1%	7.8%	5.1%	4.2%	77.8%
GTP-4o(↑)	4.600	5.467	4.733	5.867	8.067

表 1. **定量评估**。我们将我们的方法与包括 Instruct-NeRF2NeRF [16], GaussianEditor [12], DragDiffusion [53] 和 SDE-Drag [44] 在内的最先进的方法进行了比较。相比这些方法，我们的方法在所有评估指标上都表现更优。

插值，使用在 Eq. (5) 中定义的权重。如 Fig. 5 所示，我们的方法实现了更精确的修改并更好地保留了原始模型的几何信息。相比之下，使用直接插值不仅无法正确地变形高斯函数，导致断裂的部分，还会在其他区域引起意外的变化。此外，直接拖动和插值策略难以对高斯函数执行旋转。

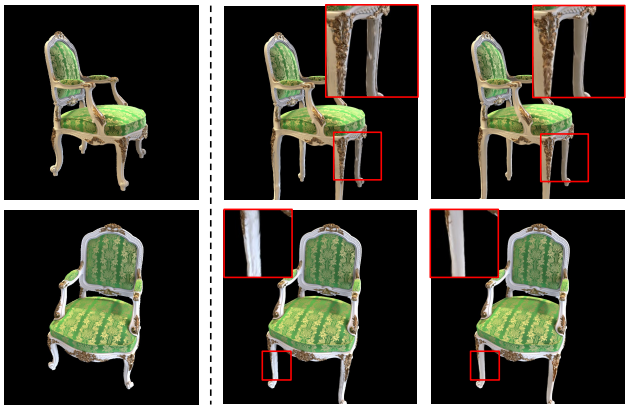

图 6. **扩散先验的有效性**。我们选择两个视图来比较有无扩散先验的结果。我们将变形区域放大，以突出我们方法在变形后保持高保真渲染质量的能力。

扩散先验的有效性。我们展示了没有扩散先验的结果在 Fig. 6 中。很明显，除了几何 3DGS 变形外，结合超分辨率模型进一步通过使用增强图像微调 3DGS 参数来提升高斯的外观，从而减少了伪影并提高了质量。%**运行时间**。我们的编辑时间取决于优化迭代次数和场景大小。对于大约 500k 需要被编辑的高斯分布，ARAP 大约需要 6 分钟，而 1000 次微调优化迭代大约需要 5 分钟。相比之下，现有的基于文本的 3D 编辑方法，如 Instruct-NeRF2NeRF，需要大约一小时，而 GaussianEditor 需要大约 20 分钟。

5. 结论、限制与未来工作

我们提出了一种名为 ARAP-GS 的拖动驱动型 ARAP 三维高斯模型编辑方法，该方法具有扩散先验。我们将 ARAP 变形直接应用于 3D 高斯分布上，使得 3DGS 能够进行自由形式几何变形而不需额外的监督数据。此外，我们利用扩散先验的内容增强能力来优化渲染图像，并调整 3DGS 的外观。为了在微调过程中保持视角一致性和几何完整性，我们实施了几种策略，如迭代数据集更新和基于掩码引导的 3DGS 微调。广泛的实验表明，我们的方法实现了高质量的拖动驱动型 3DGS 编辑，并在各种场景中明显优于其他方法。

限制。我们的方法以最小干扰保持了场景的拓扑结构，因此如 Fig. 7 所示的一些期望编辑结果不被我们的方法所支持。

图 7. **方法失败的情况**。在试图让这个人张开嘴巴时，嘴部区域的高斯分布保持其原始结构而不撕裂以形成孔洞，导致在这种情况下结果不太令人满意。

未来工作。尽管我们的方法为拖曳驱动的 3D 高斯编辑任务做出了开创性的贡献，但我们尚未直接修改 ARAP 变形过程中 3D 高斯的其他属性，例如缩放和颜色。未来，我们计划探索在 ARAP 变形过程中的额外 3DGS 属性转换。此外，虽然我们使用扩散先验来提高图像质量，但并未充分利用扩散模型的生成能力。未来的工作可以结合其他扩散模型，如 ControlNet [69]，以增强所提出方法的生成能力。探索其他传统的网格操作用于

3DGS 变形也将会很有趣，例如基于手柄 [34] 和基于骨架 [8] 的变形。

参考文献

[1] Omri Avrahami, Dani Lischinski, and Ohad Fried. Blended diffusion for text-driven editing of natural images. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pages 18208–18218, 2022. 3

[2] Jiayang Bai, Letian Huang, Jie Guo, Wen Gong, Yuanqi Li, and Yanwen Guo. 360-gs: Layout-guided panoramic gaussian splatting for indoor roaming. arXiv preprint arXiv:2402.00763, 2024. 2

[3] Daniele Baieri, Filippo Maggioli, Zorah Lähner, Simone Melzi, and Emanuele Rodolà. Implicit-arap: Efficient handle-guided deformation of high-resolution meshes and neural fields via local patch meshing. arXiv preprint arXiv:2405.12895, 2024. 3

[4] Jonathan T Barron, Ben Mildenhall, Dor Verbin, Pratul P Srinivasan, and Peter Hedman. Mip-nerf 360: Unbounded anti-aliased neural radiance fields. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pages 5470–5479, 2022. 5

[5] Asker M Bazen and Sabih H Gerez. Thin-plate spline modelling of elastic deformations in fingerprints. In Proceedings of 3rd IEEE Benelux Signal Processing Symposium. Citeseer, 2002. 3

[6] Tim Brooks, Aleksander Holynski, and Alexei A Efros. Instructpix2pix: Learning to follow image editing instructions. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 18392–18402, 2023. 3

[7] Weiwei Cai, Weicai Ye, Peng Ye, Tong He, and Tao Chen. Dynasurfgs: Dynamic surface reconstruction with planar-based gaussian splatting. arXiv preprint arXiv:2408.13972, 2024. 2

[8] Steve Capell, Seth Green, Brian Curless, Tom Duchamp, and Zoran Popović. Interactive skeleton-driven dynamic deformations. ACM transactions on graphics (TOG), 21(3):586–593, 2002. 9

[9] Isaac Chao, Ulrich Pinkall, Patrick Sanan, and Peter Schröder. A simple geometric model for elastic deformations. ACM transactions on graphics (TOG), 29(4): 1–6, 2010. 2, 3

[10] Honghua Chen, Yushi Lan, Yongwei Chen, Yifan Zhou, and Xingang Pan. Mvdrag3d: Drag-based creative 3d editing via multi-view generation-reconstruction priors. arXiv preprint arXiv:2410.16272, 2024. 2, 3

[11] Shu-Yu Chen, Lin Gao, Yu-Kun Lai, and Shihong Xia. Rigidity controllable as-rigid-as-possible shape deformation. Graphical Models, 91:13–21, 2017. 2, 3

[12] Yiwen Chen, Zilong Chen, Chi Zhang, Feng Wang, Xiaofeng Yang, Yikai Wang, Zhongang Cai, Lei Yang, Huaping Liu, and Guosheng Lin. Gaussianeditor: Swift and controllable 3d editing with gaussian splatting. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 21476–21485, 2024. 2, 5, 7, 8

[13] Ciprian Corneanu, Raghudeep Gadde, and Aleix M Martinez. Latentpaint: Image inpainting in latent space with diffusion models. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pages 4334–4343, 2024. 3

[14] Yutao Cui, Xiaotong Zhao, Guozhen Zhang, Shengming Cao, Kai Ma, and Limin Wang. Stabledrag: Stable dragging for point-based image editing. arXiv preprint arXiv:2403.04437, 2024. 3

[15] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. Advances in neural information processing systems, 27, 2014. 2, 3

[16] Ayaan Haque, Matthew Tancik, Alexei A Efros, Aleksander Holynski, and Angjoo Kanazawa. Instruct-nerf2nerf: Editing 3d scenes with instructions. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pages 19740–19750, 2023. 2, 5, 7, 8

[17] Amir Hertz, Ron Mokady, Jay Tenenbaum, Kfir Aberman, Yael Pritch, and Daniel Cohen-Or. Prompt-to-prompt image editing with cross attention control. arXiv preprint arXiv:2208.01626, 2022. 3

[18] Jonathan Ho, Chitwan Saharia, William Chan, David J Fleet, Mohammad Norouzi, and Tim Salimans. Cascaded diffusion models for high fidelity im-

age generation. Journal of Machine Learning Research, 23(47):1–33, 2022. 3

[19] Jiajun Huang and Hongchuan Yu. Gsdeformer: Direct cage-based deformation for 3d gaussian splatting. arXiv preprint arXiv:2405.15491, 2024. 2

[20] Qixing Huang, Xiangru Huang, Bo Sun, Zaiwei Zhang, Junfeng Jiang, and Chandrajit Bajaj. Arapreg: An as-rigid-as possible regularization loss for learning deformable shape generators. In Proceedings of the IEEE/CVF international conference on computer vision, pages 5815–5825, 2021. 3

[21] Yi-Hua Huang, Yang-Tian Sun, Ziyi Yang, Xiaoyang Lyu, Yan-Pei Cao, and Xiaojuan Qi. Sc-gs: Sparse-controlled gaussian splatting for editable dynamic scenes. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 4220–4230, 2024. 2

[22] Pushkar Joshi, Mark Meyer, Tony DeRose, Brian Green, and Tom Sanocki. Harmonic coordinates for character articulation. ACM transactions on graphics (TOG), 26(3):71–es, 2007. 3

[23] Tao Ju, Scott Schaefer, and Joe Warren. Mean value coordinates for closed triangular meshes. In Seminal Graphics Papers: Pushing the Boundaries, Volume 2, pages 223–228. 2023. 3

[24] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3d gaussian splatting for real-time radiance field rendering. ACM Transactions on Graphics, 42(4):1–14, 2023. 2, 3

[25] Leif P Kobbelt, Thilo Bareuther, and Hans-Peter Seidel. Multiresolution shape deformations for meshes with dynamic vertex connectivity. In Computer Graphics Forum, pages 249–260. Wiley Online Library, 2000. 3

[26] Youna Le Vaou, Jean-Claude Léon, Stefanie Hahmann, Stéphane Masfrand, and Matthieu Mika. As-stiff-as-needed surface deformationcombining arap energy with an anisotropic material. Computer-Aided Design, 121:102803, 2020. 3

[27] Junoh Lee, Chang-Yeon Won, Hyunjun Jung, Inhwan Bae, and Hae-Gon Jeon. Fully explicit dynamic gaussian splatting. arXiv preprint arXiv:2410.15629, 2024. 2

[28] Zohar Levi and Craig Gotsman. Smooth rotation enhanced as-rigid-as-possible mesh animation. IEEE transactions on visualization and computer graphics, 21(2):264–277, 2014. 2, 3

[29] Dongxu Li, Junnan Li, and Steven Hoi. Blip-diffusion: Pre-trained subject representation for controllable text-to-image generation and editing. Advances in Neural Information Processing Systems, 36, 2024. 3

[30] Wenbo Li, Zhe Lin, Kun Zhou, Lu Qi, Yi Wang, and Jiaya Jia. Mat: Mask-aware transformer for large hole image inpainting. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pages 10758–10768, 2022. 3

[31] Xinqi Lin, Jingwen He, Ziyan Chen, Zhaoyang Lyu, Bo Dai, Fanghua Yu, Wanli Ouyang, Yu Qiao, and Chao Dong. Diffbir: Towards blind image restoration with generative diffusion prior. arXiv preprint arXiv:2308.15070, 2023. 3

[32] Yaron Lipman, David Levin, and Daniel Cohen-Or. Green coordinates. ACM transactions on graphics (TOG), 27(3):1–10, 2008. 3

[33] Haofeng Liu, Chenshu Xu, Yifei Yang, Lihua Zeng, and Shengfeng He. Drag your noise: Interactive point-based editing via diffusion semantic propagation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 6743–6752, 2024. 3

[34] Minghua Liu, Minhyuk Sung, Radomir Mech, and Hao Su. Deepmetahandles: Learning deformation metahandles of 3d meshes with biharmonic coordinates. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 12–21, 2021. 9

[35] Steven Liu, Xiuming Zhang, Zhoutong Zhang, Richard Zhang, Jun-Yan Zhu, and Bryan Russell. Editing conditional radiance fields. In Proceedings of the IEEE/CVF international conference on computer vision, pages 5773–5783, 2021. 2

[36] Zhicheng Lu, Xiang Guo, Le Hui, Tianrui Chen, Min Yang, Xiao Tang, Feng Zhu, and Yuchao Dai. 3d geometry-aware deformable gaussian splatting for dynamic view synthesis. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 8900–8910, 2024. 2

[37] Shaojie Ma, Yawei Luo, and Yi Yang. Reconstructing and simulating dynamic 3d objects with

mesh-adsorbed gaussian splatting. arXiv preprint arXiv:2406.01593, 2024. 2

[38] Chenlin Meng, Yutong He, Yang Song, Jiaming Song, Jiajun Wu, Jun-Yan Zhu, and Stefano Ermon. Sdedit: Guided image synthesis and editing with stochastic differential equations. arXiv preprint arXiv:2108.01073, 2021. 3

[39] 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. 2, 5

[40] Chong Mou, Xintao Wang, Jiechong Song, Ying Shan, and Jian Zhang. Dragondiffusion: Enabling drag-style manipulation on diffusion models. arXiv preprint arXiv:2307.02421, 2023. 3

[41] Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. Instant neural graphics primitives with a multiresolution hash encoding. ACM transactions on graphics (TOG), 41(4):1–15, 2022. 5

[42] Yuichi Nagata and Shinji Imahori. Creation of dihedral escher-like tilings based on as-rigid-as-possible deformation. ACM Transactions on Graphics, 43(2):1–18, 2024. 3

[43] Alexander Quinn Nichol and Prafulla Dhariwal. Improved denoising diffusion probabilistic models. In International conference on machine learning, pages 8162–8171. PMLR, 2021. 3

[44] Shen Nie, Hanzhong Allan Guo, Cheng Lu, Yuhao Zhou, Chenyu Zheng, and Chongxuan Li. The blessing of randomness: Sde beats ode in general diffusion-based image editing. arXiv preprint arXiv:2311.01410, 2023. 3, 5, 7, 8

[45] Francesco Palandra, Andrea Sanchietti, Daniele Baieri, and Emanuele Rodolà. Gsedit: Efficient text-guided editing of 3d objects via gaussian splatting. arXiv preprint arXiv:2403.05154, 2024. 2

[46] Xingang Pan, Ayush Tewari, Thomas Leimkühler, Lingjie Liu, Abhimitra Meka, and Christian Theobalt. Drag your gan: Interactive point-based manipulation on the generative image manifold. In ACM SIGGRAPH 2023 Conference Proceedings, pages 1–11, 2023. 2, 3

[47] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pages 10684–10695, 2022. 2, 3

[48] Leonardo Sacht, Etienne Vouga, and Alec Jacobson. Nested cages. ACM Transactions on Graphics (TOG), 34(6):1–14, 2015. 3

[49] Chitwan Saharia, William Chan, Saurabh Saxena, Lala Li, Jay Whang, Emily L Denton, Kamyar Ghasemipour, Raphael Gontijo Lopes, Burcu Karagol Ayan, Tim Salimans, et al. Photorealistic text-to-image diffusion models with deep language understanding. Advances in neural information processing systems, 35:36479–36494, 2022. 3

[50] Thomas W Sederberg and Scott R Parry. Free-form deformation of solid geometric models. In Proceedings of the 13th annual conference on Computer graphics and interactive techniques, pages 151–160, 1986. 3

[51] Sitian Shen, Jing Xu, Yuheng Yuan, Xingyi Yang, Qiuhong Shen, and Xinchao Wang. Draggaussian: Enabling drag-style manipulation on 3d gaussian representation. arXiv preprint arXiv:2405.05800, 2024. 2

[52] Yahao Shi, Yanmin Wu, Chenming Wu, Xing Liu, Chen Zhao, Haocheng Feng, Jingtuo Liu, Liangjun Zhang, Jian Zhang, Bin Zhou, et al. Gir: 3d gaussian inverse rendering for relightable scene factorization. arXiv preprint arXiv:2312.05133, 2023. 2

[53] Yujun Shi, Chuhui Xue, Jun Hao Liew, Jiachun Pan, Hanshu Yan, Wenqing Zhang, Vincent YF Tan, and Song Bai. Dragdiffusion: Harnessing diffusion models for interactive point-based image editing. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 8839–8849, 2024. 2, 3, 5, 7, 8

[54] Daniel Sieger, Stefan Menzel, and Mario Botsch. Constrained space deformation for design optimization. Procedia Engineering, 82:114–126, 2014. 3

[55] Olga Sorkine and Marc Alexa. As-rigid-as-possible surface modeling. In Symposium on Geometry processing, pages 109–116. Citeseer, 2007. 2, 3, 4

[56] Yanhao Sun, Runze Tian, Xiao Han, Xinyao Liu, Yan Zhang, and Kai Xu. Gseditpro: 3d gaussian splatting editing with attention-based progressive localization. In Computer Graphics Forum, page e15215. Wiley Online Library, 2024. 2

[57] Michael Wand, Philipp Jenke, Qixing Huang, Martin Bokeloh, Leonidas Guibas, and Andreas Schilling. Reconstruction of deforming geometry from time-varying point clouds. In Symposium on Geometry processing, pages 49–58, 2007. 3

[58] Can Wang, Menglei Chai, Mingming He, Dongdong Chen, and Jing Liao. Clip-nerf: Text-and-image driven manipulation of neural radiance fields. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 3835–3844, 2022. 2

[59] Junjie Wang, Jiemin Fang, Xiaopeng Zhang, Lingxi Xie, and Qi Tian. Gaussianeditor: Editing 3d gaussians delicately with text instructions. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 20902–20911, 2024. 2

[60] Jianyi Wang, Zongsheng Yue, Shangchen Zhou, Kelvin C.K. Chan, and Chen Change Loy. Exploiting diffusion prior for real-world image super-resolution. 2024. 3, 5

[61] Su Wang, Chitwan Saharia, Ceslee Montgomery, Jordi Pont-Tuset, Shai Noy, Stefano Pellegrini, Yasumasa Onoe, Sarah Laszlo, David J Fleet, Radu Soricut, et al. Imagen editor and editbench: Advancing and evaluating text-guided image inpainting. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pages 18359–18369, 2023. 3

[62] Yuxuan Wang, Xuanyu Yi, Zike Wu, Na Zhao, Long Chen, and Hanwang Zhang. View-consistent 3d editing with gaussian splatting. In European Conference on Computer Vision, pages 404–420. Springer, 2025. 2

[63] Jing Wu, Jia-Wang Bian, Xinghui Li, Guangrun Wang, Ian Reid, Philip Torr, and Victor Adrian Prisacariu. Gaussctrl: multi-view consistent text-driven 3d gaussian splatting editing. arXiv preprint arXiv:2403.08733, 2024. 2

[64] Tong Wu, Guandao Yang, Zhibing Li, Kai Zhang, Ziwei Liu, Leonidas Guibas, Dahua Lin, and Gordon Wetzstein. Gpt-4v (ision) is a human-aligned evaluator for text-to-3d generation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 22227–22238, 2024. 6

[65] Weiwei Xu, Kun Zhou, Yizhou Yu, Qifeng Tan, Qunsheng Peng, and Baining Guo. Gradient domain editing of deforming mesh sequences. ACM Transactions On Graphics (TOG), 26(3):84–es, 2007. 3

[66] Mingqiao Ye, Martin Danelljan, Fisher Yu, and Lei Ke. Gaussian grouping: Segment and edit anything in 3d scenes. arXiv preprint arXiv:2312.00732, 2023. 2

[67] Fanghua Yu, Jinjin Gu, Zheyuan Li, Jinfan Hu, Xiangtao Kong, Xintao Wang, Jingwen He, Yu Qiao, and Chao Dong. Scaling up to excellence: Practicing model scaling for photo-realistic image restoration in the wild. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 25669–25680, 2024. 3

[68] Tao Yu, Runseng Feng, Ruoyu Feng, Jinming Liu, Xin Jin, Wenjun Zeng, and Zhibo Chen. Inpaint anything: Segment anything meets image inpainting. arXiv preprint arXiv:2304.06790, 2023. 3

[69] Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. Adding conditional control to text-to-image diffusion models. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pages 3836–3847, 2023. 8

[70] Zewei Zhang, Huan Liu, Jun Chen, and Xiangyu Xu. Gooddrag: Towards good practices for drag editing with diffusion models. arXiv preprint arXiv:2404.07206, 2024. 5

[71] Jingyu Zhuang, Chen Wang, Liang Lin, Lingjie Liu, and Guanbin Li. Dreameditor: Text-driven 3d scene editing with neural fields. In SIGGRAPH Asia 2023 Conference Papers, pages 1–10, 2023. 2

[72] Jingyu Zhuang, Di Kang, Yan-Pei Cao, Guanbin Li, Liang Lin, and Ying Shan. Tip-editor: An accurate 3d editor following both text-prompts and image-prompts. ACM Transactions on Graphics (TOG), 43 (4):1–12, 2024. 2