参赛学生姓名：James Jincheng Hu

中学：合肥安生学校

省份：安徽省

国家/地区：中国北方赛区

指导老师姓名：刘利刚 蔡有城

指导老师单位：中国科学技术大学数学科学学院

论文题目：CraftMesh: High-Fidelity Generative Mesh Manipulation via Poisson Seamless Fusion

高保真生成式 3D 模型编辑

CraftMesh: High-Fidelity Generative Mesh Manipulation via Poisson Seamless Fusion

James Jincheng Hu

注：本部分为中文的论文总览。对于可视化展示和技术细节参考实验视频和英文部分。

1 论文总览

近几年生成式人工智能（Generative AI）进展飞快。以文本-图像的扩散模型为代表（比如当前流行的图像编辑/生成模型），用户只需用一句话或一张示意图，就能得到风格统一、语义精确的高质量图片。类似地，近来的文本/图像 → 3D 大模型（如腾讯混元 3D、CraftsMan3D 等）能在分钟级把文字或图片“变成”带几何与纹理的 3D 模型，这对动画、游戏、VR/AR、虚拟试衣等产业意义重大，因为它极大降低了 3D 内容制作的门槛。

但生成式 3D 模型在可控编辑（editing）上仍面临挑战。用户常常希望在已有模型上做局部改动——比如给鹿加上翅膀、替换雕像头部、把一只狐狸扩成九尾狐等。现有一些顶会工作（如 FocalDreamer、MagicClay、Instant3dit 等）虽然能自动编辑，但面对复杂局部结构时，常出现几何扭曲、边界不自然、颜色风格不匹配等问题。

CraftMesh 的出发点是：将 2D 图像编辑和 3D 生成的优势结合起来，再用“泊松（Poisson）”思想做几何与颜色的无缝融合，从而实现高保真、局部精细且整体和谐的网格（mesh）编辑。

1.1 方法总览

CraftMesh 把一次编辑分成三步，思路直观：

(1) 先做 2D 图像编辑 + 生成局部 3D 网格（初始编辑）

把原始模型渲染成参考图片，用强大的图像编辑模型（文本或拖拽驱动）在图片上完成期望修改（例如给鹿加翅膀）。然后把修改后的局部图像单独送入 3D 生成模型，得到只包含新部件（如翅膀）的高细节局部网格。最后把该局部网格粗拼回原始模型，得到初始融合结果。这个过程利用了 2D 模型在语义与风格控制上的优势，以及 3D 生成模型在局部细节上的优势。

(2) 泊松几何融合（Poisson Geometric Fusion）——让交接处“长得像一个整体”

粗拼的结果往往在交界处出现突兀、法向不连续或接缝明显。为了解决这个问题，CraftMesh 引入一个混合 SDF/网格（SDF/Mesh）表示，并用“法线图的泊松融合”作为指导来优化过渡区的几何形状。通俗说法：

SDF（Signed Distance Field）是一个“高度场”，任一点的值表示到最近表面的距离（正负号告诉点在表面内外）。把网格绑定到一个可学习的 SDF 上，修改 SDF 就可以稳定而连续地改变网格表面。

我们从三处来源得到法线图（normal map）：原始拼接网格的法线、编辑后参考网格的法线、以及当前 SDF 渲染出的法线。先用泊松图像融合在二维法线图上把“细节”（来自局部编辑网格）和“整体平滑结构”（来自参考网格）融合起来，得到一个既细致又过渡自然的“混合法线”。再把这个混合法线作为监督，用 SDF 优化网格，使几何真实地遵循这个平滑且富细节的目标法线场，从而得到边界平滑、细节保留的几何融合结果。

(3) 泊松颜色调和（Poisson Texture Harmonization）——让外观“融为一色”

仅靠生成模型直接给新几何着色，常出现亮度/色调偏移与边界断裂。为此，我们提出一个分布对齐 + 梯度保真 + 边界平滑的一体化调和框架（可无缝扩展到 PBR 多通道）：**分布感知的颜色对齐**：将保留模型区域的颜色视作目标分布、将新区域的颜色视作待对齐分布。颜色通过隐式神经颜色场在 $\mathbb{R}^3$ 中表示，利用核密度估计（KDE）构建两者在 RGB 空间的概率密度，并最小化分布差异，从全局统计上消除色调/亮度偏移。**梯度保真的泊松融合**：在颜色场梯度层面，对保留模型区域与新区域的梯度进行一致性约束，以稳健保留高频纹理细节，可视作经典泊松融合“保细节、调低频”的 3D 颜色场形式。**边界平滑的过渡细化**：在新旧区域交界处，引入距离加权的最近邻颜色匹配损失，近边界权重大、远离边界权重小，从而在局部几何邻域上实现自然衔接与风格一致。

1.2 举例说明（“给鹿加翅膀”——飞廉）

(1) 把鹿的模型渲染出一张参考图片，用图像编辑模型把鹿“加上翅膀”（图像里看起来很自然）。

(2) 把图片中只包含翅膀的那一块提取出来，送入 3D 生成模型得到翅膀的高细节网格。

(3) 将翅膀网格拼回鹿身，得到初始合并模型（通常接缝处较生硬）。

(4) 用泊松几何融合：提取接缝区域、渲染法线图并做法线域的泊松融合，用 SDF 优化使几何平滑地过渡；

(5) 用泊松颜色调和：在颜色分布（全局统计）与颜色梯度（局部细节）双层面实现一致性，辅以距离加权的边界细化，得到风格统一、细节保真、边界无缝的外观融合；并天然支持 PBR 多通道的联合调和。

1.3 方法优势

(1) 利用 2D 模型的语义/风格控制 + 3D 模型的局部几何细节，规避了直接在 3D 空间用扩散先验（SDS/MVD）去“逐顶点优化”所带来的不稳定与模糊问题。

(2) SDF + 法线域泊松监督使得几何优化更稳定、易收敛，同时保留细节；直接在顶点空间做同类优化更容易受噪声和离散化影响。

(3) 在纹理空间做泊松融合，既能保留源区域的高频细节（细纹、羽毛纹理等），又能在边界处做到肉眼难辨的平滑过渡，支持 PBR 通道扩展。

1.4 实验与结论

论文对比了 FocalDreamer、MagicClay、Instant3dit 等方法，在复杂编辑任务上（插入、替换、拖拽式精细编辑）取得了更好的定性与定量结果。定量上使用 CLIP 相似度（CLIPsim），方向性 CLIP（CLIPdir），质量指标 NIQE，和 NIMA 作为评测指标，CraftMesh 在这两类指标上均优于基线；消融实验也验证了“泊松几何融合”和“泊松颜色调和”对最终质量的关键作用。实现上，“泊松几何融合”在单块 4090 GPU、约 1000 次迭代（示例约 5 分钟）即可得到良好结果。CraftMesh 的核心思想：先在 2D 上把想法“画好”，再把局部高质量地“变成立体”，最后用泊松思想把新旧部分在几何和颜色上无缝拼接在一起。这种“2D 编辑 → 局部 3D 生成 → 泊松融合”策略利用了各类生成式大模型的长处，解决了直接在 3D 空间编辑时常见的扭曲和不自然问题。

CraftMesh: High-Fidelity Generative Mesh Manipulation via Poisson Seamless Fusion

Abstract

Controllable, high-fidelity mesh editing remains a significant challenge in 3D content creation. Existing generative methods often struggle with complex geometries and fail to produce detailed results. We propose CraftMesh, a novel framework for high-fidelity generative mesh manipulation via Poisson Seamless Fusion. Our key insight is to decompose mesh editing into a pipeline that leverages the strengths of 2D editing and 3D generation models: we edit a 2D reference image, then generate a region-specific 3D mesh, and seamlessly fuse it into the original model. We introduce two core techniques: Poisson Geometric Fusion, which utilizes a hybrid SDF/Mesh representation with normal blending to achieve seamless geometric integration, and Poisson Texture Harmonization for visually harmonious texture blending. Experimental results demonstrate that CraftMesh outperforms state-of-the-art methods, delivering superior global consistency and local detail in complex editing tasks.

Keywords: 3D Mesh Editing, Generative Models, Poisson Fusion, Texture Harmonization

CraftMesh

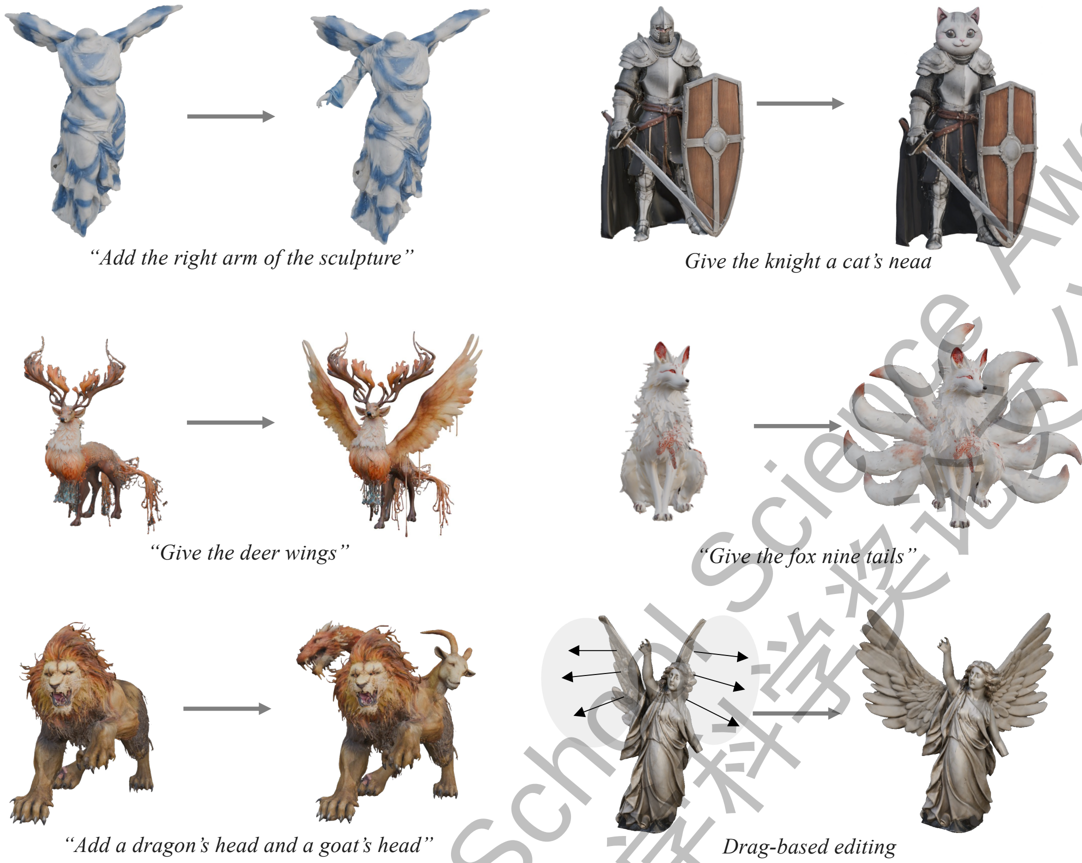

Fig. 1: Mesh editing results produced by CraftMesh. CraftMesh is a versatile 3D mesh editing framework that enables users to perform text-based and drag-based editing for insertion, deletion and replacement, while producing high-quality results.

1 Introduction

In recent years, the rapid advancement of 3D generation technologies [1–4] has enabled the synthesis of high-quality 3D content directly from text prompts or images through diffusion-based generative models. These advances have substantially accelerated downstream applications in video games, augmented and virtual reality (AR/VR) [5], robotics [6], and digital manufacturing [7].

Despite these notable achievements in 3D generation, the challenge of controllable 3D editing remains largely unresolved. Most current 3D generation frameworks are designed to reconstruct 3D models from 2D images and provide limited flexibility for localized modifications. Neural field-based representations, such as Neural Radiance Fields (NeRF) [8] and 3D Gaussian Splatting (3DGS) [9], have demonstrated strong capability in capturing fine-grained details while leveraging differentiable rendering. Consequently, research has focused on neural field editing, encompassing appearance-guided and text-

or image-driven approaches [10–12], which remain restricted to appearance-level modifications and cannot inherently support geometric manipulations on explicitly surfaced meshes.

In contrast to the rapidly expanding body of work on neural field editing, mesh-based generative editing has received considerably less attention, even though meshes remain the most widely adopted representation in professional 3D content creation pipelines. In practical design workflows, artists and engineers often need to iteratively refine meshes with precise part-level control to satisfy both aesthetic and functional requirements, while avoiding unintended alterations to unrelated geometry. This demand underscores the necessity for editing methods that enable fine-grained controllability while faithfully preserving the geometry of the original model.

Existing generative mesh editing methodologies can be broadly categorized into two principal paradigms: score distillation sampling (SDS) based approaches and multi-view diffusion (MVD) based approaches. SDS-based methods [13, 14] enhance 3D awareness by directly optimizing the mesh through an SDS loss. MVD-based approaches [15–17] pair multi-view consistent editing with a reconstruction step. However, these methods exhibit several limitations: (1) they are not well-suited for editing complex models; (2) the quality of the generated edits is frequently suboptimal, failing to satisfy the requirements for high-fidelity mesh manipulation.

To address these challenges, we propose an novel methodology that harnesses the capabilities of generative models by reframing editing tasks as generative processes. We introduce an **image editing–mesh generation–seamless fusion** framework that fully capitalizes on the strengths of 2D models for image editing and 3D models for high-quality mesh generation. Specifically, we edit the image, generate 3D content for the edited region, and integrate the generated mesh into the original model. The principal challenge lies in ensuring both geometric and textural consistency between the generated mesh and the original model.

In this paper, we present a **High-Fidelity Generative Mesh Manipulation** framework, coined CraftMesh, which harnesses the capabilities of generative large models to accomplish complex mesh editing tasks (see Fig. 1). First, we employ a 2D image editing model to edit reference images derived from the original mesh, extract the modified regions, and generate region-specific meshes for these edited regions. Second, we propose a **Poisson Geometric Fusion** strategy, employing a robust SDF/Mesh representation with a Poisson normal blending technique to achieve seamless geometric fusion of the edited region mesh with the original mesh. Finally, we introduce a **Poisson Texture Harmonization** strategy to facilitate seamless texture fusion between the edited region mesh and the original mesh within texture space. Experimental results demonstrate the superiority of our approach in achieving high-fidelity mesh editing. Additionally, we conduct further experiments utilizing a drag-based method for fine-grained image editing, demonstrating ours framework's versatility and (see Fig. 6).

Our contributions are summarized as follows:

- A novel framework that reformulates mesh editing as an **image editing–mesh generation–seamless fusion** pipeline integrating 2D and 3D generative models.
- Seamless geometric fusion, introducing a Global and Local Consistency Poisson Geometric Fusion strategy for integrating the edited region mesh into the original mesh.
- Seamless texture harmonization, proposing a Poisson Texture Harmonization strategy that enables coherent blending of edited textures with the original appearance.

2 Related Work

3D Generation Models. Recent advances in 2D diffusion models [18–20] have profoundly accelerated 3D content creation.

SDS-based Approaches. Score Distillation Sampling (SDS) bridges 2D diffusion priors and 3D optimization. DreamFusion [21] first optimized NeRF under text-to-image diffusion guidance, followed by Magic3D [22], which introduced a two-stage low-to-high resolution refinement. Later, LucidDreamer [23] further enhanced stability and fidelity through interval score matching, whereas ProlificDreamer [24] incorporated a variational SDS formulation to improve diversity and quality. These methods successfully bridge 2D diffusion priors and 3D optimization, although they frequently remain computationally intensive.

MVD-based Approaches. Multi-view diffusion (MVD) enforces view consistency during image synthesis to reconstruct 3D assets. SyncDreamer [25] and MVDream [26] exploit multi-view diffusion for geometrically coherent text-to-3D generation. Wonder3D [27] and One-2-3-45++ [28] further extend this paradigm to single-image 3D generation. Recent large-scale methods such as SV3D [29] and Instant3D [30] achieve high-quality reconstructions from sparse views.

3D Native Generation Approaches. More recently, researchers have shifted toward training generative models directly on large-scale 3D datasets, thereby overcoming the inherent limitations of 2D priors. Foundational resources such as Objaverse [31], Objaverse-XL [32], and OmniObject3D [33] provide millions of diverse, well-annotated 3D objects, enabling scalable learning of both geometry and appearance. Clay [34] demonstrates controllable large-scale text-to-3D generation by training on millions of objects. Trellis [35] proposes structured 3D latent representations that improve scalability and versatility, making generative models more efficient at capturing complex shapes. Hunyuan3D 2.0 [1] pushes diffusion-based 3D generation to high-resolution textured assets, significantly enhancing realism. 3DTopia-XL [36] scales primitive-based diffusion approaches, achieving improved generalization across diverse categories. These native 3D models mark a shift toward more direct, efficient, and realistic 3D generation.

Generative Mesh Editing. Most existing generative editing approaches primarily focus on implicit representations [10–12, 37, 38]. While these methods achieve promising results, they are constrained by implicit representations and thus cannot be applied to mesh-level editing. In this paper, we focus on generative mesh editing, which can be broadly categorized into two paradigms: SDS-based editing and MVD-based editing.

SDS-based Editing. SDS-based editing approaches extend the concept of Score Distillation Sampling (SDS) loss to editing tasks by guiding mesh optimization using pretrained diffusion priors. FocalDreamer [13] introduces focal-fusion assembly for localized text-driven 3D editing, thereby enabling controllable, region-specific modifications. MagicClay [14] bridges generative neural fields with mesh sculpting, allowing users to refine or modify mesh geometry under the guidance of Score Distillation Sampling.

MVD-based Editing. MVD-based Editing approaches employ multi-view diffusion to ensure multi-view consistency during editing, thus bridging 2D image generation and 3D mesh manipulation. MVEdit [15] adapts generic 3D diffusion priors for controlled multi-view editing. CMD [17] purposed CondMV, which takes a target image and multi-view conditions and generates multi-view consistent edits. Instant3dit [16] introduces fast multi-view inpainting to accelerate editing workflows, while MaskedLRM [39] leverages large reconstruction models with masked conditioning for efficient mesh editing.

However, these methods fail to edit highly complex models or achieve high-quality mesh manipulation. In this paper, our method fully capitalizes on the complementary strengths of 2D and 3D generative models. By employing a Poisson seamless fusion strategy, our approach merges generated region-specific meshes with the original mesh, thereby achieving high-fidelity and structurally consistent mesh manipulation.

Seamless Editing. Seamless editing is a fundamental topic in computer graphics and digital image processing. The primary goal is to achieve smooth and imperceptible transitions in images or textures, thus maintaining visual consistency [40–43]. Liao et al. [44] develop Deep Image Analogy, which leverages convolutional neural networks to establish semantically meaningful dense correspondences between two images, thus advancing seamless editing capabilities. Yu et al.[45] apply the Poisson equation to mesh editing, enabling smooth geometric merging via gradient field manipulation, although this method does not address appearance blending.

Recently, SeamlessNeRF [46] achieves seamless stitching of neural radiance fields through gradient propagation, focusing on radiance field merging without considering explicit mesh geometry. GS-Stitching [47] advances example-based 3D modeling by introducing 3D Gaussian stitching. While these works offer smooth merging in radiance fields or 3D Gaussians, explicit mesh geometry is not considered. In this paper, we consider both geometry and appearance, ensuring seamless fusion between the edited region mesh and the original mesh.

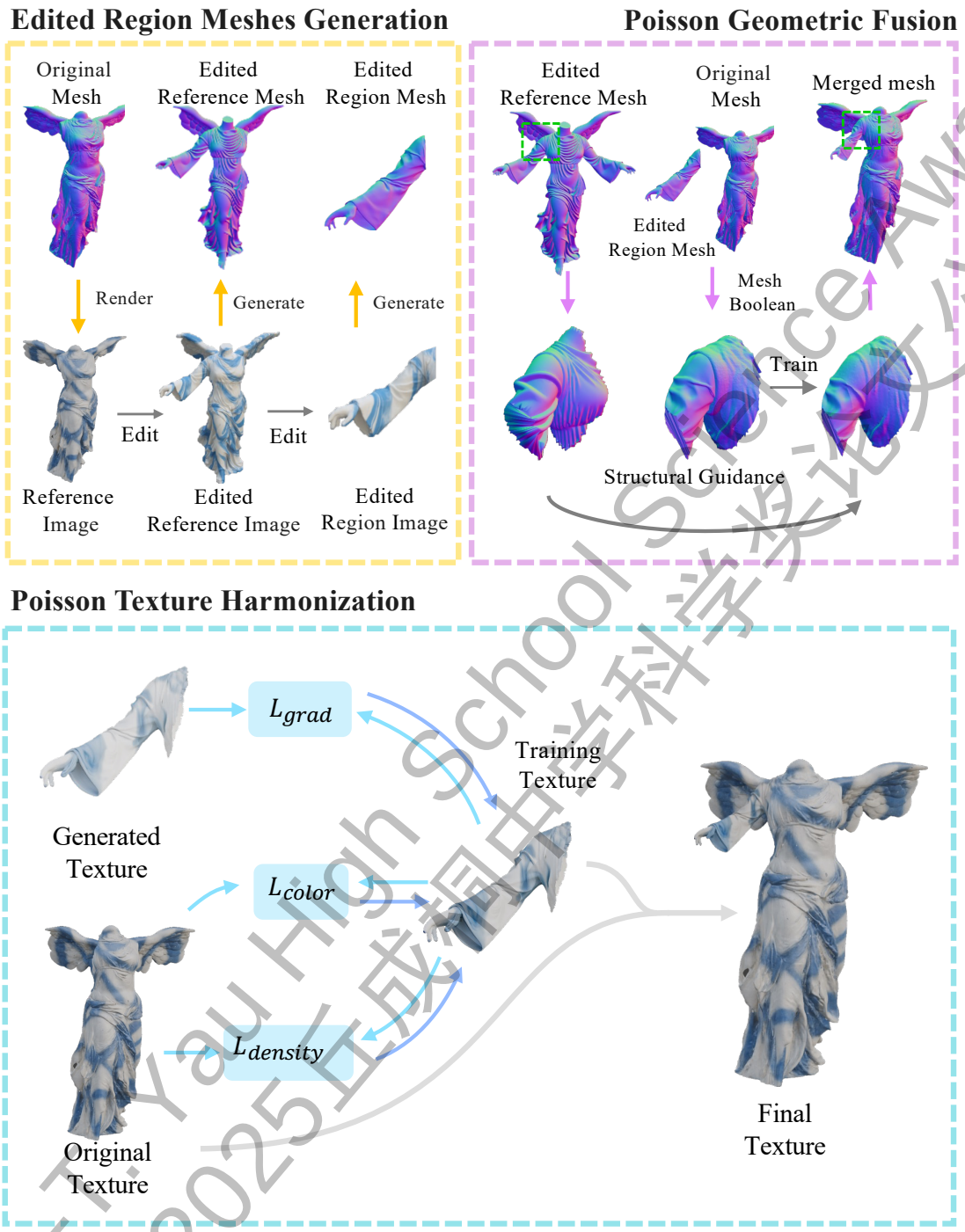

Fig. 2: The overview of CraftMesh's architecture. First, Edited Region-Specific Meshes Generation is done as the basis of editing. Then, Poisson Geometric Fusion harmonizes a rough geometric transition. Last, Poisson Texture Harmonization colors the edited parts in a seamless manner.

3 Method

We propose CraftMesh, a high-fidelity generative mesh manipulation framework that integrates 2D image editing, 3D mesh generation, and Poisson-based

fusion. Fig. 2 illustrates the overall workflow. Our framework is designed to address the limitations of existing 3D editing approaches, which are often not well-suited for editing highly complex models and achieving high-fidelity mesh manipulation. Specifically, we first edit reference images using 2D image editing models to achieve user-intent-consistent modifications, followed by generating edited region meshes with 3D generative models. Second, we propose a Poisson Geometric Fusion strategy that employs global and local consistency constraints to achieve seamless geometric fusion of the edited region mesh with the original mesh. Finally, we introduce a Poisson Texture Harmonization strategy to ensure appearance consistency and facilitate seamless texture fusion between the edited region mesh and the original mesh. This design enables controllable editing while maintaining both the structural integrity and high visual quality of the final mesh.

3.1 Edited Region Meshes Generation

Text-to-image models have demonstrated remarkable performance in controllable image editing, producing semantically aligned and globally consistent results. Representative examples include FLUX Kontext [20], Qwen3 [48], and Gemini 2.5 [49], which can effectively preserve content structure while introducing new details. Compared with direct 3D editing, these 2D approaches are lightweight, controllable, and well-suited for generating high-quality edited reference images. On the other hand, recent progress in 3D generative modeling, such as CraftsMan3D [3] and Hunyuan3D [1], has enabled the synthesis of meshes with unprecedented geometric fidelity and textural realism. However, existing 3D mesh editing methods lag significantly behind. For instance, Instant3dit [16] fine-tunes multi-view diffusion models to regenerate 3D content, but often struggles with consistency. Similarly, FocalDreamer [13] and MagicClay [14] are limited to simple objects and frequently yield low-quality results in the edited region.

To bridge this gap, we propose jointly leveraging the complementary advantages of 2D image editing models and 3D mesh generation models. Specifically, we generate **Edited Region Meshes** as intermediate assets, which are later fused with the original mesh. The generation proceeds in two steps:

2D Editing. We begin by editing the reference image rendered from the original mesh, guided by the user's intent. Users can leverage a variety of tools, such as image editing models [20], software, or other instruments, providing flexibility and creative control. Next, we use FLUX Kontext [20], a state-of-the-art image editing model, to extract the *edited region image* from the edited reference image, highlighting only the modified areas, thereby localizing the mesh editing scope. FLUX Kontext excels at fine-grained text-guided edits while maintaining structural consistency and handling occlusions.

3D Generation. We then use CraftsMan3D [3] to generate meshes from both the edited reference image and the edited region image, producing the *edited reference mesh* and the *edited region mesh*, respectively. The edited reference mesh provides a global structure but typically lacks fine detail. In contrast,

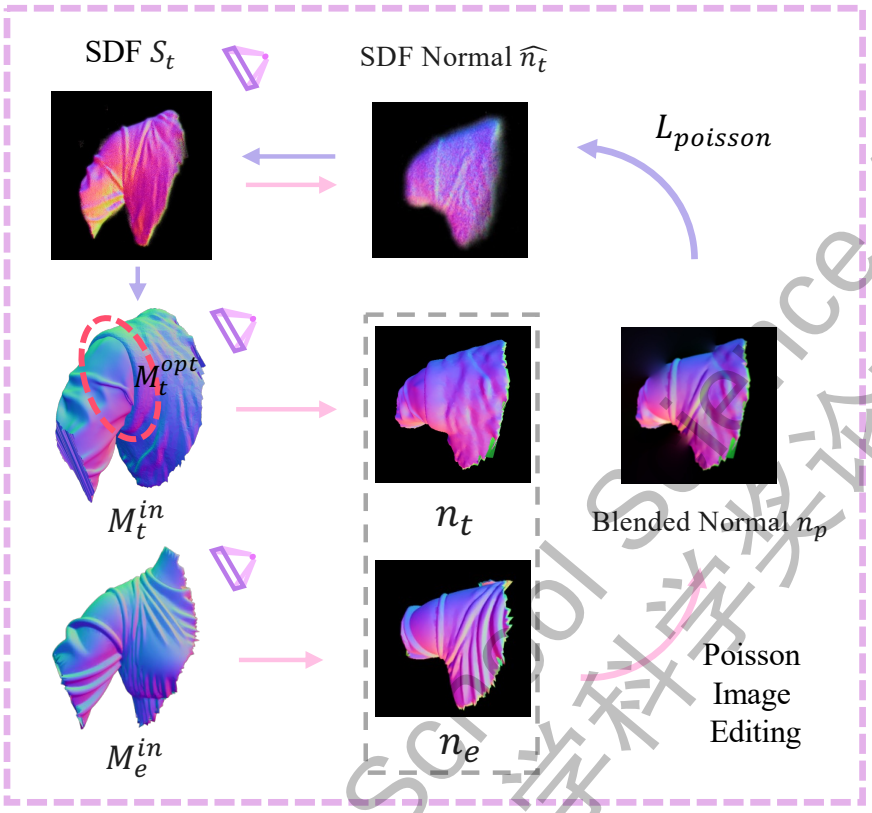

Fig. 3: Details of Poisson Geometric Fusion.

the edited region mesh offers higher local fidelity but isn't seamlessly integrated with the original mesh. This discrepancy arises from the inherent generative trade-off: holistic reconstructions emphasize plausibility over accuracy, whereas localized generation prioritizes detail at the expense of alignment.

Our central idea is to fuse the edited region mesh into the original mesh while using the edited reference mesh as guidance. This ensures that the final model inherits the global smoothness of the edited reference mesh and the fine-grained quality of the edited region mesh. Compared with prior methods, CraftMesh offers: (1) no requirement for manual specification of precise 3D editing locations, unlike FocalDreamer [13], MagicClay [14], and Instant3dit [16], making editing more controllable and user-friendly; (2) effective integration of 2D editing capabilities with 3D mesh generation, ensuring high-quality edited regions with global coherence.

3.2 Poisson Geometric Fusion

Naively integrated the edited region mesh into original mesh using mesh Boolean can introduce noticeable artifacts, such as surface normal discontinuities and inharmonious geometric details. Our objective is to seamlessly

integrate the edited region into the original mesh while simultaneously preserving local fine-grained details and maintaining global structure. To this end, we propose a Poisson Geometric Fusion strategy, which leverages the edited reference mesh as structural guidance. This ensures that the final reconstructed mesh inherits the harmonious global structure of the reference mesh while retaining the local details of the edited region.

Fig. 3 gives an overview of the workflow. We first employ a mesh Boolean operation [50] to obtain a coarse merged mesh from the original mesh and the edited region mesh. We then adopt a hybrid SDF/Mesh representation, which enables flexible refinement of mesh geometry by optimizing vertex positions, splitting triangles and collapsing edges. The refinement is guided by normal maps rendered from both the edited reference mesh and the edited region mesh, which are blended using a Poisson-based approach. This fusion strategy allows the edited region to be naturally incorporated into the original mesh with smooth boundary transitions.

Intersection Region Extraction Given the original mesh M_o and the edited region mesh M_r, we first apply a mesh Boolean operation to obtain a merged mesh M_t. For insertion tasks, we use mesh Boolean union, and for deletion tasks, we use mesh Boolean difference. Since geometric discontinuities mainly occur at the transition boundary, we explicitly refine this region using a hybrid SDF/Mesh representation.

The Boolean operation produces a set of vertices V_{in} at the intersection between M_o and M_r. We align the edited reference mesh M_e with M_t, and define the corresponding intersection regions as:

$$M_t^{in} = \left\{ v \in M_t \mid \min_{u \in V_{in}} \|u - v\|_2 < \epsilon_0 \right\}, \tag{1}$$

$$M_e^{in} = \left\{ v \in M_e \mid \min_{u \in V_{in}} \|u - v\|_2 < \epsilon_0 \right\}, \tag{2}$$

where ϵ_0 controls the extent of the intersection. We further define the optimization region as a smaller subset within the intersection:

$$M_t^{opt} = \left\{ v \in M_t^{in} \mid \min_{u \in V_{in}} \|u - v\|_2 < \epsilon_1 \right\}, \quad \epsilon_1 < \epsilon_0. \tag{3}$$

This ensures that the optimization is restricted to M_t^{opt}, focusing refinements on the transition area, while M_e^{in} provides structural guidance for achieving smooth and coherent fusion.

Poisson Normal Blending Guidance To refine the optimization region M_t^{opt}, we bind a neural SDF S_t to the mesh, changes in the SDF will be propagated to the mesh through vertex optimization. SDF-based vertex optimization offers stable convergence and robustness against noise, achieving more precise and natural geometry than voxel-based methods like DMTet [51].

During optimization, we render multiple supervision images from random viewpoints: (1) a normal map of M_t^{in}, denoted n_t; (2) a binary mask of M_t^{opt}, denoted $mask^{opt}$; (3) a normal map rendered from the SDF S_t, denoted $\hat{n}_t$; (4) a normal map of M_e^{in}, denoted n_e. To enforce consistency, we apply the classical Poisson Image Editing (PIE) algorithm [40] to blend n_t and n_e under $mask^{opt}$:

$$n_p = \Gamma(n_t, n_e, mask^{opt}), \tag{4}$$

where $\Gamma(\cdot)$ denotes the Poisson blending operator. This blended normal map n_p preserves fine-grained details from n_e inside the mask while achieving a smooth transition into n_t at the mask's boundary.

We then minimize the discrepancy between the rendered normal $\hat{n}_t$ and the blended normal n_p:

$$\mathcal{L}_{\text{poisson}} = \sum_i \|\hat{n}_t^i - n_p^i\|_F^2, \tag{5}$$

where $\|\cdot\|_F$ denotes the Frobenius norm and i indexes different camera viewpoints. Although the blended normal maps n_p^i are not strictly multi-view consistent, the implicit SDF effectively resolves inconsistencies and learns a coherent transition geometry. Following MagicClay, we further incorporate additional regularization terms, such as a smoothness loss $\mathcal{L}_{\text{smooth}}$ and an Eikonal loss $\mathcal{L}_{\text{eik}}$, to improve geometric fidelity and to enforce implicit surface constraints. The final loss is formulated as:

$$\mathcal{L}_{\text{geo}} = \mathcal{L}_{\text{poisson}} + \lambda_1 \mathcal{L}_{\text{smooth}} + \lambda_2 \mathcal{L}_{\text{eik}}, \tag{6}$$

where λ_1 and λ_2 are hyperparameters.

3.3 Poisson Texture Harmonization

After geometric editing, the newly synthesized regions of the mesh M_t lack texture information. A straightforward solution is to employ texture generation models for color synthesis; however, the resulting textures often exhibit noticeable color shifts from the original mesh and discontinuities along region boundaries. Although recent work [46, 47] has explored seamless texture fusion in NeRF and 3DGS frameworks, no existing method directly addresses this challenge for explicit mesh representations.

Distribution-Aware Color Alignment. Let M_t^{new} denote the newly synthesized geometry and M_t^{pr} the preserved geometry. The preserved mesh M_t^{pr} inherits textures directly from the original mesh M_o, while M_t^{new} is textured using a generative model (MeshyAI [52]). Following [14], colors are encoded using an implicit neural color field defined over $\mathbb{R}^3$. The predicted colors form a probability density distribution in RGB space, which we regularize using kernel density estimation (KDE):

$$\rho(q) = \frac{1}{N}\sum_{i=1}^{N}\exp\left(-\frac{\|q-r_i\|^2}{2\sigma^2}\right), \tag{7}$$

where $r_i{}_{i=1}^{N}$ are sampled mesh colors and σ is the standard deviation of the Gaussian kernel.

We denote the color distributions of M_t^{new} and M_t^{pr} as ρ^{new} and ρ^{pr}, respectively. Distribution-aware alignment is achieved by minimizing the discrepancy between these two distributions:

$$\mathcal{L}_{\text{density}} = \frac{1}{N}\sum_{i=1}^{N}\left\|\rho^{new}(q_i) - \rho^{pr}(q_i)\right\|_2, \tag{8}$$

where $q_i i = 1^N$ are color samples from M_t^{new}.

Gradient-Preserving Poisson Fusion. For each 3D point x on the mesh, let $C(x)$ denote its color, and C_{pr} and C_{new} represent color fields on M_t^{pr} and M_t^{new}, respectively. To preserve fine-grained appearance details, we enforce gradient consistency across regions:

$$\mathcal{L}_{\text{grad}} = \text{MSE}\left(\sigma\left(\frac{\nabla C_{pr}}{\gamma}\right), \sigma\left(\frac{\nabla C_{new}}{\gamma}\right)\right), \tag{9}$$

where ∇ denotes numerical color gradients, $\sigma(\cdot)$ is a sigmoid function, and γ is a gradient scaling constant.

Smooth Transition Refinement. To ensure smooth transitions at the intersection boundary, we introduce a distance-weighted color matching loss:

$$\mathcal{L}_{\text{color}} = \sum_{p_i^{new}\in M_t^{new}} w_i \|C_{new}(p_i^{new}) - C_{pr}(p_i^{pr})\|_2^2, \tag{10}$$

where p_i^{pr} is the nearest point on M_t^{pr} to p_i^{new}, and

$$w_i = \left(1 - \frac{\delta}{\|p_i^{new} - p_i^{pr}\|_2}\right)^2, \tag{11}$$

attenuates the influence with distance. The parameter δ controls the effective boundary width.

The overall optimization objective for Poisson Texture Harmonization is:

$$\mathcal{L}_{\text{tex}} = \mathcal{L}_{\text{density}} + \theta_1\mathcal{L}_{\text{grad}} + \theta_2\mathcal{L}_{\text{color}}, \tag{12}$$

where θ_1 and θ_2 balance the gradient and boundary consistency terms. Unlike prior mesh editing pipelines that synthesize only textures, our formulation directly extends to physically based rendering (PBR) materials, as texture generation models inherently support multi-channel PBR texture maps.

4 Experiments

4.1 Experiment Setup

Implementation. We use FLUX Kontext [20] as the generative image-editing method, and CraftsMan3D [3] as the image-to-mesh method. It is worth noting that our framework is agnostic to these choices. As more powerful models come out, they should be used instead when conducting experiments. We use MagicClay [14] as the hybrid SDF/Mesh representation backbone and implicit neural color field backbone. On a single 4090 GPU, Poisson Geometric Fusion takes 5 minutes and 1000 iterations, Poisson Texture Harmonization takes 1 minute and 2000 iterations

Mesh Dataset The evaluation dataset consists of meshes with intricate detail and complex geometry. We test these meshes with complex editing tasks to best showcase our method's capabilities for insertion, deletion, and drag-based mesh editing, and demonstrate our method's achievements in global geometry consistency and local high-quality detail.

Baselines We compare our method against recent mesh editing approaches, specifically FocalDreamer [13], MagicClay [14], and Instant3dit [16]. The official open-source implementations of these baselines are used.

4.2 Qualitative Results

Fig. 4 presents a qualitative comparison with baseline methods. As illustrated, the baselines struggle with complex examples, resulting in coarse geometry and a lack of detail. The generated colors are often simple, flat, and inharmonious. In contrast, our method produces intricate geometry with a harmonious global structure, rich local details, and high-fidelity colors. For the fourth task, where mesh removal is applied on the volcano, MagicClay replaces the volcano with a rock of a distorted color style; Instant3dit substitutes the volcano with a bland patch of grass, but fails to preserve the original part's geometry and quality; our method seamlessly removes the volcano and fills the space with rocks similar to those in adjacent regions, thereby achieving both visual and geometric harmony.

Method	CLIP_{sim} ↑	CLIP_{dir} ↑	NIQE ↓	NIMA ↑
FocalDreamer	20.831	8.224	3.377	4.834
MagicClay	20.350	2.201	3.797	4.886
Instant3dit	19.530	2.079	3.790	4.765
Ours	**22.768**	**13.594**	**3.203**	**5.071**

Table 1: Quantitative comparison with other methods using CLIP similarity (CLIP_{sim}), directional CLIP similarity (CLIP_{dir}), NIQE, and NIMA scores.

Fig. 4: Qualitative comparisons show that our method produces intricate geometry with a harmonious global structure, rich local details, and high-fidelity colors.

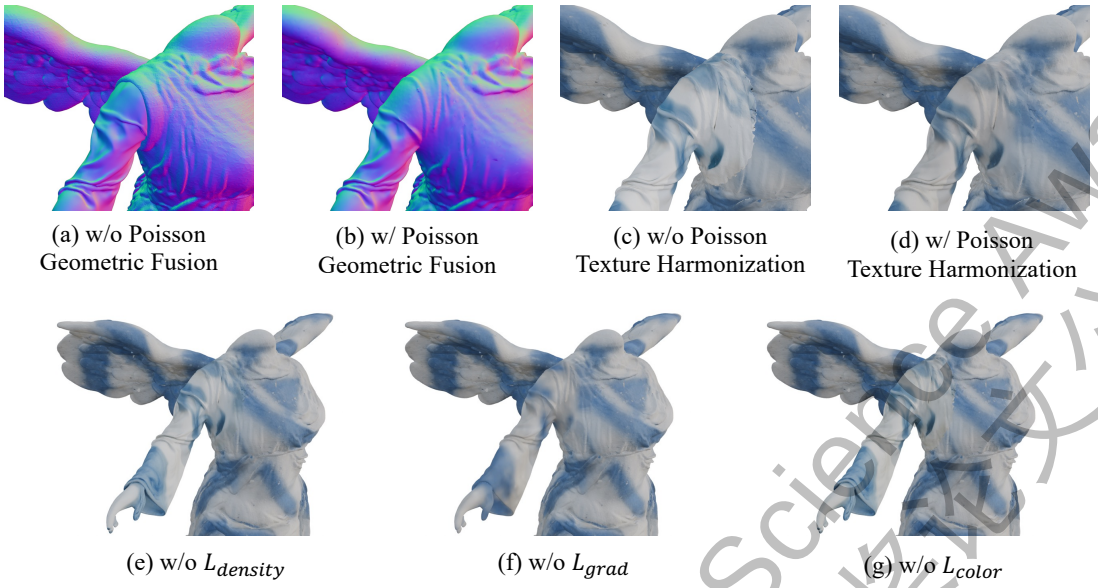

Fig. 5: Ablation study. (a,b) Poisson Geometric Fusion; (c,d) Poisson Texture Harmonization; (e–g) Poisson Texture Harmonization losses.

4.3 Quantitative Results

Following prior work [13, 53], we use CLIP-based metrics for quantitative evaluation: (1) CLIP_{sim}, which measures the alignment between a rendered view of the edited mesh and the target text description; and (2) CLIP_{dir}, which evaluates editing effectiveness by computing the directional CLIP similarity [54] between the initial and edited meshes, based on their respective text prompts. In addition, we report NIQE [55]and NIMA [56], two no-reference image quality metrics that assess perceptual fidelity and better correlate with human visual judgment.

Table 1 present the results of the four metrics. Our method achieves the highest scores across all of them, demonstrating its strong ability to produce edits that are both semantically faithful and visually consistent with the desired objectives.

4.4 Ablation

As seen in Fig. 5, Poisson Geometric Fusion resolves the harsh geometric transition with natural details, while Poisson Texture Harmonization corrects the shifted colors of the hand, rectifying the bright whiteness to the body's darker gray. The details of the hand are retained, and texture continuity is achieve at the boundary.

We conduct ablations for each loss of Poisson Texture Harmonization. Without properly learning the color distribution of the original mesh (Fig. 5e), the resulting hand has a darker white and a greener blue, breaking harmony. Without retaining the original gradients (Fig. 5f), the original details

are blurred. When colors aren't learned at the boundary (Fig. 5g), the discontinuity of colors at the boundary is noticeable.

(a) Original (b) Edited

Fig. 6: Drag-based mesh editing. (a) shows the original mesh, with arrows drawn indicating the desired drag to apply. (b) shows the results.

4.5 Drag-based Mesh Editing

Beyond mesh insertion and deletion, our approach can be extended to more sophisticated mesh editing tasks. To showcase this versatility, we apply our framework to enable **drag-based mesh editing** via **drag-based image editing**.

Unlike prompt-based image editing, drag-based image editing empowers users to specify edits by drawing arrows that encode the desired drag deformations, providing precise and intuitive control over the editing process. For this operation, we leverage LightningDrag [57] as the drag-based image editor.

The workflow for drag-based mesh editing involves three steps: First, drag-based image editing is performed; Then, mesh deletion is applied to the corresponding region of the mesh; Last, mesh insertion is conducted using the Edited Region meshes derived from the edited images.

Fig. 6a depicts the original meshes, with arrow annotations drawn, signifying the intention to open the angle's wings and raising the cat's hands. Fig. 6b are the successful results of drag-based mesh editing. The effectiveness in drag-based mesh editing validates the adaptability of our approach, demonstrating the feasibility of extending our ideas to other advanced mesh editing operations.

5 Conclusion

We present CraftMesh, a framework for high-fidelity mesh manipulation. Our approach addresses the limitations of current methods by combining 2D image editing and 3D generation models. We further purpose a Poisson Seamless Fusion strategy, which ensures both geometric and textural consistency when integrating new content. The proposed Poisson Geometric Fusion and Poisson Texture Harmonization techniques enable complex, detailed edits that

are seamlessly blended into the original mesh. Experimental results demonstrate that CraftMesh achieves superior performance over existing baselines, achieving harmonious global geometric structure, intricate local detail, and high-fidelity colors. The framework is also designed to be extensible, enabling seamless integration with future advances in generative AI driven by the rapid development of image editing and mesh generation models. Future work can be done to apply our ideas to more advanced mesh editing operations, or ensure robustness against edge cases.

References

[1] Lai, Z., Zhao, Y., Liu, H., Zhao, Z., Lin, Q., Shi, H., Yang, X., Yang, M., Yang, S., Feng, Y., et al.: Hunyuan3d 2.5: Towards high-fidelity 3d assets generation with ultimate details. arXiv preprint arXiv:2506.16504 (2025)

[2] Xu, J., Cheng, W., Gao, Y., Wang, X., Gao, S., Shan, Y.: Instantmesh: Efficient 3d mesh generation from a single image with sparse-view large reconstruction models. arXiv preprint arXiv:2404.07191 (2024)

[3] Li, W., Liu, J., Yan, H., Chen, R., Liang, Y., Chen, X., Tan, P., Long, X.: Craftsman3d: High-fidelity mesh generation with 3d native generation and interactive geometry refiner. arXiv preprint arXiv:2405.14979 (2024)

[4] Siddiqui, Y., Monnier, T., Kokkinos, F., Kariya, M., Kleiman, Y., Garreau, E., Gafni, O., Neverova, N., Vedaldi, A., Shapovalov, R., *et al.*: Meta 3d assetgen: Text-to-mesh generation with high-quality geometry, texture, and pbr materials. Advances in Neural Information Processing Systems **37**, 9532–9564 (2024)

[5] Thi Vo, K.H.: Augmented reality, virtual reality, and mixed reality: A pragmatic view from diffusion of innovation. International Journal of Architectural Computing **23**(1), 27–45 (2025)

[6] Liang, J.E.: Diffusion models for robotics. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 39, pp. 29587–29589 (2025)

[7] Li, X., Tao, F., Ye, W., Nassehi, A., Sutherland, J.W.: Generative manufacturing systems using diffusion models and chatgpt. arXiv preprint arXiv:2405.00958 (2024)

[8] Mildenhall, B., Srinivasan, P.P., Tancik, M., Barron, J.T., Ramamoorthi, R., Ng, R.: Nerf: Representing scenes as neural radiance fields for view synthesis. Communications of the ACM **65**(1), 99–106 (2021)

[9] Kerbl, B., Kopanas, G., Leimkühler, T., Drettakis, G.: 3d gaussian splatting for real-time radiance field rendering. ACM Trans. Graph. **42**(4), 139–1 (2023)

[10] Haque, A., Tancik, M., Efros, A.A., Holynski, A., Kanazawa, A.: Instruct-nerf2nerf: Editing 3d scenes with instructions. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 19740–19750 (2023)

[11] Wang, J., Fang, J., Zhang, X., Xie, L., Tian, Q.: Gaussianeditor: Editing 3d gaussians delicately with text instructions. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 20902–20911 (2024)

[12] Zhuang, J., Kang, D., Cao, Y.-P., Li, G., Lin, L., Shan, Y.: Tip-editor: An accurate 3d editor following both text-prompts and image-prompts. ACM Transactions on Graphics (TOG) **43**(4), 1–12 (2024)

[13] Li, Y., Dou, Y., Shi, Y., Lei, Y., Chen, X., Zhang, Y., Zhou, P., Ni, B.: Focaldreamer: Text-driven 3d editing via focal-fusion assembly. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 38, pp. 3279–3287 (2024)

[14] Barda, A., Kim, V., Aigerman, N., Bermano, A.H., Groueix, T.: Magicclay: Sculpting meshes with generative neural fields. In: SIGGRAPH Asia 2024 Conference Papers, pp. 1–10 (2024)

[15] Chen, H., Shi, R., Liu, Y., Shen, B., Gu, J., Wetzstein, G., Su, H., Guibas, L.: Generic 3d diffusion adapter using controlled multi-view editing. arXiv preprint arXiv:2403.12032 (2024)

[16] Barda, A., Gadelha, M., Kim, V.G., Aigerman, N., Bermano, A.H., Groueix, T.: Instant3dit: Multiview inpainting for fast editing of 3d objects. In: Proceedings of the Computer Vision and Pattern Recognition Conference, pp. 16273–16282 (2025)

[17] Li, P., Ma, S., Chen, J., Liu, Y., Zhang, C., Xue, W., Luo, W., Sheffer, A., Wang, W., Guo, Y.: Cmd: Controllable multiview diffusion for 3d editing and progressive generation. In: Proceedings of the Special Interest Group on Computer Graphics and Interactive Techniques Conference Conference Papers, pp. 1–10 (2025)

[18] Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B.: High-resolution image synthesis with latent diffusion models. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 10684–10695 (2022)

[19] Oppenlaender, J.: The creativity of text-to-image generation. In: Proceedings of the 25th International Academic Mindtrek Conference, pp. 192–202 (2022)

[20] Labs, B.F., Batifol, S., Blattmann, A., Boesel, F., Consul, S., Diagne, C., Dockhorn, T., English, J., English, Z., Esser, P., et al.: Flux. 1 kontext: Flow matching for in-context image generation and editing in latent space. arXiv preprint arXiv:2506.15742 (2025)

[21] Poole, B., Jain, A., Barron, J.T., Mildenhall, B.: Dreamfusion: Text-to-3d using 2d diffusion. arXiv preprint arXiv:2209.14988 (2022)

[22] Lin, C.-H., Gao, J., Tang, L., Takikawa, T., Zeng, X., Huang, X., Kreis, K., Fidler, S., Liu, M.-Y., Lin, T.-Y.: Magic3d: High-resolution text-to-3d content creation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 300–309 (2023)

[23] Liang, Y., Yang, X., Lin, J., Li, H., Xu, X., Chen, Y.: Luciddreamer: Towards high-fidelity text-to-3d generation via interval score matching. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 6517–6526 (2024)

[24] Wang, Z., Lu, C., Wang, Y., Bao, F., Li, C., Su, H., Zhu, J.: Prolificdreamer: High-fidelity and diverse text-to-3d generation with variational score distillation. Advances in neural information processing systems **36**, 8406–8441 (2023)

[25] Liu, Y., Lin, C., Zeng, Z., Long, X., Liu, L., Komura, T., Wang, W.: Syncdreamer: Generating multiview-consistent images from a single-view image. arXiv preprint arXiv:2309.03453 (2023)

[26] Shi, Y., Wang, P., Ye, J., Long, M., Li, K., Yang, X.: Mvdream: Multi-view diffusion for 3d generation. arXiv preprint arXiv:2308.16512 (2023)

[27] Long, X., Guo, Y.-C., Lin, C., Liu, Y., Dou, Z., Liu, L., Ma, Y., Zhang, S.-H., Habermann, M., Theobalt, C., *et al.*: Wonder3d: Single image to 3d using cross-domain diffusion. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 9970–9980 (2024)

[28] Liu, M., Shi, R., Chen, L., Zhang, Z., Xu, C., Wei, X., Chen, H., Zeng, C., Gu, J., Su, H.: One-2-3-45++: Fast single image to 3d objects with consistent multi-view generation and 3d diffusion. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 10072–10083 (2024)

[29] Voleti, V., Yao, C.-H., Boss, M., Letts, A., Pankratz, D., Tochilkin, D., Laforte, C., Rombach, R., Jampani, V.: Sv3d: Novel multi-view synthesis and 3d generation from a single image using latent video diffusion. In: European Conference on Computer Vision, pp. 439–457 (2024). Springer

[30] Li, J., Tan, H., Zhang, K., Xu, Z., Luan, F., Xu, Y., Hong, Y., Sunkavalli, K., Shakhnarovich, G., Bi, S.: Instant3d: Fast text-to-3d with sparse-view generation and large reconstruction model. arXiv preprint arXiv:2311.06214 (2023)

[31] Deitke, M., Schwenk, D., Salvador, J., Weihs, L., Michel, O., VanderBilt, E., Schmidt, L., Ehsani, K., Kembhavi, A., Farhadi, A.: Objaverse: A universe of annotated 3d objects. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 13142–13153 (2023)

[32] Deitke, M., Liu, R., Wallingford, M., Ngo, H., Michel, O., Kusupati, A., Fan, A., Laforte, C., Voleti, V., Gadre, S.Y., VanderBilt, E., Kembhavi, A., Vondrick, C., Gkioxari, G., Ehsani, K., Schmidt, L., Farhadi, A.: Objaverse-xl: A universe of 10m+ 3d objects. NeurIPS (2023)

[33] Wu, T., Zhang, J., Fu, X., Wang, Y., Ren, J., Pan, L., Wu, W., Yang, L., Wang, J., Qian, C., *et al.*: Omniobject3d: Large-vocabulary 3d object dataset for realistic perception, reconstruction and generation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 803–814 (2023)

[34] Zhang, L., Wang, Z., Zhang, Q., Qiu, Q., Pang, A., Jiang, H., Yang, W., Xu, L., Yu, J.: Clay: A controllable large-scale generative model for creating high-quality 3d assets. ACM Transactions on Graphics (TOG) **43**(4), 1–20 (2024)

[35] Xiang, J., Lv, Z., Xu, S., Deng, Y., Wang, R., Zhang, B., Chen, D., Tong, X., Yang, J.: Structured 3d latents for scalable and versatile 3d generation. In: Proceedings of the Computer Vision and Pattern Recognition Conference, pp. 21469–21480 (2025)

[36] Chen, Z., Tang, J., Dong, Y., Cao, Z., Hong, F., Lan, Y., Wang, T., Xie, H., Wu, T., Saito, S., *et al.*: 3dtopia-xl: Scaling high-quality 3d asset generation via primitive diffusion. In: Proceedings of the Computer Vision and Pattern Recognition Conference, pp. 26576–26586 (2025)

[37] Cheng, X., Yang, T., Wang, J., Li, Y., Zhang, L., Zhang, J., Yuan, L.: Progressive3d: Progressively local editing for text-to-3d content creation with complex semantic prompts. arXiv preprint arXiv:2310.11784 (2023)

[38] Sabat, B.O., Achille, A., Trager, M., Soatto, S.: Nerf-insert: 3d local editing with multimodal control signals. arXiv preprint arXiv:2404.19204 (2024)

[39] Gao, W., Wang, D., Fan, Y., Bozic, A., Stuyck, T., Li, Z., Dong, Z., Ranjan, R., Sarafianos, N.: 3d mesh editing using masked lrms. arXiv

preprint arXiv:2412.08641 (2024)

[40] Pérez, P., Gangnet, M., Blake, A.: Poisson image editing. In: Seminal Graphics Papers: Pushing the Boundaries, Volume 2, pp. 577–582 (2003)

[41] Agarwala, A., Dontcheva, M., Agrawala, M., Drucker, S., Colburn, A., Curless, B., Salesin, D., Cohen, M.: Interactive digital photomontage. In: ACM SIGGRAPH 2004 Papers, pp. 294–302 (2004)

[42] Kwatra, V., Essa, I., Bobick, A., Kwatra, N.: Texture optimization for example-based synthesis. In: ACM Siggraph 2005 Papers, pp. 795–802 (2005)

[43] Barnes, C., Shechtman, E., Finkelstein, A., Goldman, D.B.: Patchmatch: A randomized correspondence algorithm for structural image editing. ACM Trans. Graph. **28**(3), 24 (2009)

[44] Liao, J., Yao, Y., Yuan, L., Hua, G., Kang, S.B.: Visual attribute transfer through deep image analogy. arXiv preprint arXiv:1705.01088 (2017)

[45] Yu, Y., Zhou, K., Xu, D., Shi, X., Bao, H., Guo, B., Shum, H.-Y.: Mesh editing with poisson-based gradient field manipulation. In: ACM SIGGRAPH 2004 Papers, pp. 644–651 (2004)

[46] Gong, B., Wang, Y., Han, X., Dou, Q.: Seamlessnerf: Stitching part nerfs with gradient propagation. In: SIGGRAPH Asia 2023 Conference Papers, pp. 1–10 (2023)

[47] Gao, X., Yang, Z., Gong, B., Han, X., Yang, S., Jin, X.: Towards realistic example-based modeling via 3d gaussian stitching. In: Proceedings of the Computer Vision and Pattern Recognition Conference, pp. 26597–26607 (2025)

[48] Yang, A., Li, A., Yang, B., Zhang, B., Hui, B., Zheng, B., Yu, B., Gao, C., Huang, C., Lv, C., et al.: Qwen3 technical report. arXiv preprint arXiv:2505.09388 (2025)

[49] Comanici, G., Bieber, E., Schaekermann, M., Pasupat, I., Sachdeva, N., Dhillon, I., Blistein, M., Ram, O., Zhang, D., Rosen, E., et al.: Gemini 2.5: Pushing the frontier with advanced reasoning, multimodality, long context, and next generation agentic capabilities. arXiv preprint arXiv:2507.06261 (2025)

[50] Cherchi, G., Pellacini, F., Attene, M., Livesu, M.: Interactive and robust mesh booleans. arXiv preprint arXiv:2205.14151 (2022)

[51] Munkberg, J., Hasselgren, J., Shen, T., Gao, J., Chen, W., Evans, A.,

Müller, T., Fidler, S.: Extracting triangular 3d models, materials, and lighting from images. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 8280–8290 (2022)

[52] Meshy Inc.: Meshy.ai: AI-powered 3D Generation Platform (2025). https://www.meshy.ai/

[53] Sella, E., Fiebelman, G., Hedman, P., Averbuch-Elor, H.: Vox-e: Text-guided voxel editing of 3d objects. arXiv preprint arXiv:2303.12048 (2023)

[54] Gal, R., Patashnik, O., Maron, H., Chechik, G., Cohen-Or, D.: Stylegan-nada: Clip-guided domain adaptation of image generators. arXiv preprint arXiv:2108.00946 (2021)

[55] Mittal, A., Soundararajan, R., Bovik, A.C.: Making a "Completely Blind" Image Quality Analyzer. IEEE Signal Processing Letters **20**(3), 209–212 (2013). https://doi.org/10.1109/LSP.2012.2227726

[56] Talebi, H., Milanfar, P.: Nima: Neural image assessment. IEEE transactions on image processing **27**(8), 3998–4011 (2018)

[57] Shi, Y., Liew, J.H., Yan, H., Tan, V.Y.F., Feng, J.: Lightningdrag: Lightning fast and accurate drag-based image editing emerging from videos. arXiv preprint arXiv:2405.13722 (2024)

致谢

非常感谢刘老师和蔡老师能给我提供这次做科研的机会。我从最初用编程做游戏到现在已经有快 8 年了。这一路上，积累的图形学方面的经验，计算机和数学的知识都在这次的项目中得到了运用。无论是对社区与生态的了解，编程的能力，建模的能力，linux，ssh，还是自学微积分，参加数学竞赛，其他高等数学方面的学习。本次项目也让我找到了自己真正热爱与擅长的东西，即，图形学方面的科研。

本次项目具体的过程如下：

3 月中旬开始，进行了 3 个月的学习。从最初的复现 Poisson image editing，到复现 Gaussian Image，学会使用 PyTorch，到后来复现了 “Mesh Editing with Poisson-Based Gradient Field Manipulation”。这期间慢慢开始对科研有所了解，学会怎读论文。

6 月中旬，正式开始了科研的项目。最初的方向和刘老师、蔡老师一起商定，大致想法是：把 Mesh Editing LRM 的结果无缝拼回原来的 mesh。接下来便进行了两个月的尝试，这期间一直有开线上会议进行指导。这期间想过的想法有，利用 mesh parametrization 将 source 和 target mesh 建立对应，利用这个对应采样，让 target 学习 source 的几何，几何信息用散度储存；在 source 和 target 上找两个一一对应的环，然后优化它的距离，直到距离为 0，便可以直接拼接 source 和 target，成为一个完整的 mesh，而找环的过程虽然是 np-hard 的（可以严谨证明），可以通过 heuristics 和聪明的暴力求解在 1 秒内找到环。但是实际效果一直都很差，没有结果。我认为是因为过多在意理论与思考而缺乏实践，缺乏对实际效果的认知。

8 月中旬左右，蔡老师提到了 Flux Kontext，以及用它来实现我们想要的效果。按照实践的思想，我跑了一个马的例子，马 $\rightarrow$ 飞马 $\rightarrow$ 翅膀，三张图片。然后意识到了把翅膀拼到马身上的可行性，意识到效果出来了。于是，我们框架的第一部分确定了。用前面对于泊松的经验和理解，确定了后面两部分。

框架是 8 月中旬确定的，几何的泊松和材质的泊松分别都用了 4-5 天，每天 10 个小时以上写代码，然后晚上想问题。在 8 月底实验结果出了。

接下来的两个星期，在蔡老师的协助下完成了论文的写作。

所有代码的工作均由我一人独立完成。几何的泊松和材质的泊松，具体实施的方案是我想的，但是所有的工作都离不开两位导师宝贵的指导，对我的想法给予建议，给我正确的方向。导师的指导都是无偿的，在此，再次感谢两位导师的帮助。