InvSculpt: Inverse Sculpting Modeling via Controlled 3D Generation and a Vector Displacement Field

HENGYU MENG, The Hong Kong University of Science and Technology (Guangzhou), China
LANJIONG LI, The Hong Kong University of Science and Technology (Guangzhou), China
ZHIJING SHAO, The Hong Kong University of Science and Technology (Guangzhou), China
YINGDA YIN*, LIGHTSPEED, China
LINGTING ZHU*, LIGHTSPEED, China
ZEYU HU, LIGHTSPEED, China
XIN WANG, LIGHTSPEED, China
LIGANG LIU, Laoshan Laboratory, China
ZEYU WANG†, The Hong Kong University of Science and Technology (Guangzhou), China and The Hong Kong University of Science and Technology, China

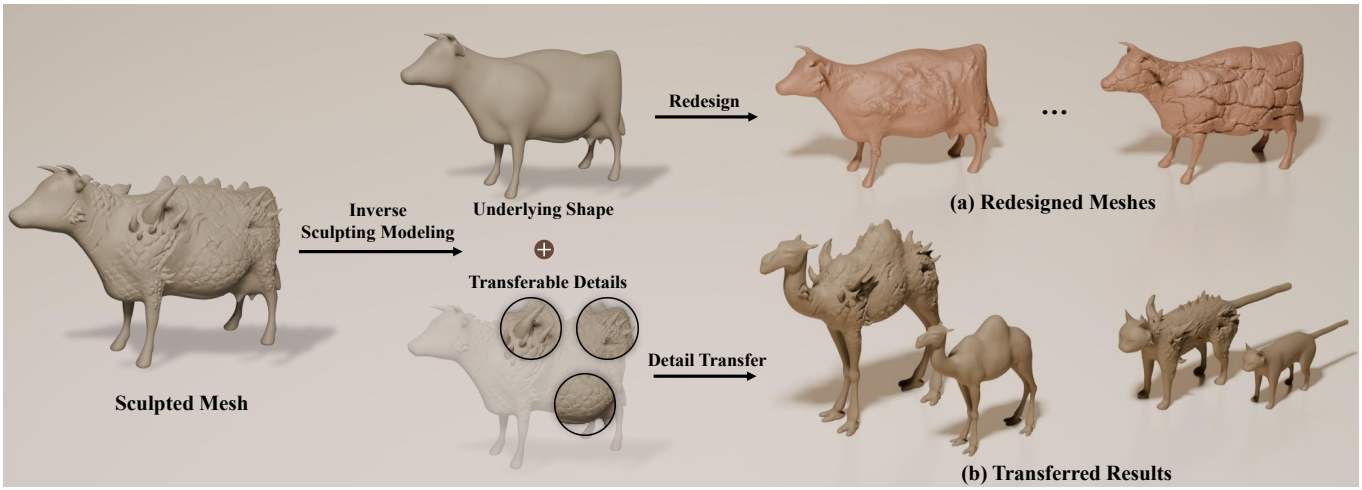

Fig. 1. **Results of inverse sculpting modeling and applications supported by InvSculpt.** InvSculpt decomposes sculpted meshes into an underlying shape and transferable details via a vector displacement field (VDF). Users can further redesign the underlying shape for other variants and transfer the details across models for rapid prototyping of high-quality 3D collections.

Inverse sculpting modeling aims to decompose a sculpted mesh into an underlying base shape and reusable geometric details, enabling non-expert users to inherit professional sculpting effort. We present InvSculpt, a novel inverse sculpting framework that decomposes a sculpted mesh into a high-fidelity underlying shape and reusable geometric details represented as a vector displacement field (VDF). Our approach combines semantic priors from text-guided 2D image editing with a 3D rectified flow model to perform inversion-based, mask-free detail removal, recovering an underlying shape that preserves the identity of the source mesh. To represent sculpted details in a lossless and transferable manner, we extract a VDF defined on the surface of the recovered underlying shape and learn a continuous neural representation for geometry-aware transfer. We observe that standard conditional sampling after inversion often suffers from trajectory drift, leading to identity shift and low-frequency distortion. To address this issue, we introduce a trajectory correction strategy that constrains early sampling steps to follow the inversion path, effectively stabilizing subsequent conditional guidance. This design enables robust detail removal and precise extraction of the VDF. Extensive experiments demonstrate that InvSculpt achieves significantly higher-quality mesh decomposition than prior methods and supports a wide range of applications, including geometry redesign and high-fidelity geometric detail transfer.

*Project leads.
†Corresponding author.

Authors' Contact Information: Hengyu Meng, The Hong Kong University of Science and Technology (Guangzhou), China; Lanjiong Li, The Hong Kong University of Science and Technology (Guangzhou), China; Zhijing Shao, The Hong Kong University of Science and Technology (Guangzhou), China; Yingda Yin, LIGHTSPEED, China; Lingting Zhu, LIGHTSPEED, China; Zeyu Hu, LIGHTSPEED, China; Xin Wang, LIGHTSPEED, China; Ligang Liu, Laoshan Laboratory, China; Zeyu Wang, The Hong Kong University of Science and Technology (Guangzhou), China and The Hong Kong University of Science and Technology, China.

SIGGRAPH Conference Papers '26, Los Angeles, CA, USA

ACM ISBN 979-8-4007-2554-8/2026/07
https://doi.org/10.1145/3799902.3811123

CCS Concepts: • **Computing methodologies → Shape modeling**.

Additional Key Words and Phrases: mesh decomposition, geometric transfer

ACM Reference Format:
Hengyu Meng, Lanjiong Li, Zhijing Shao, Yingda Yin, Lingting Zhu, Zeyu Hu, Xin Wang, Ligang Liu, and Zeyu Wang. 2026. InvSculpt: Inverse Sculpting Modeling via Controlled 3D Generation and a Vector Displacement Field. In *Special Interest Group on Computer Graphics and Interactive Techniques Conference Conference Papers (SIGGRAPH Conference Papers '26), July 19–23, 2026, Los Angeles, CA, USA*. ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3799902.3811123

1 Introduction

Sculpting modeling is a cornerstone of professional 3D content creation pipelines, enabling artists to enrich a base mesh with intricate geometric details using a wide range of sculpting brushes in modeling software [Blender 2025; ZBrush 2025] (Fig. 2). Despite its expressive power, sculpting modeling remains largely inaccessible to general users as producing high-quality sculpted assets requires substantial artistic expertise in both shape design and fine-grained geometric refinement, as well as a significant time investment, making the workflow costly and difficult to scale.

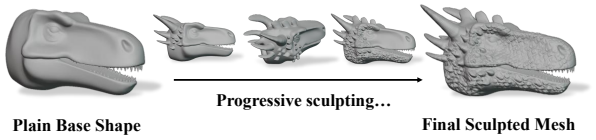

Fig. 2. **Professional sculpting modeling workflow.** Artists sculpt high-fidelity 3D models by adding fine details to a base shape. We aim to investigate the inverse process to accelerate forward 3D modeling.

This gap motivates an inverse sculpting modeling problem: can a sculpted high-resolution mesh be decomposed into an underlying base shape and reusable geometric details? Such a decomposition would allow non-expert users to reuse the sculpting effort embedded in professionally created assets, enabling geometry redesign, rapid prototyping, and large-scale asset creation by transferring sculpted details onto new base shapes.

Unfortunately, this problem remains unsolved. Existing geometry processing [Shen et al. 2022; Zhao et al. 2021] and generative modeling techniques [Barda et al. 2025; Li et al. 2025a; Ye et al. 2025] are not designed for disentangling sculpted details from base geometry, and fail to provide a representation that is both faithful to the original asset and transferable across shapes. The difficulty of inverse sculpting modeling is fundamental rather than incremental, arising from three intrinsic challenges. First, sculpted details and the underlying shape are intrinsically coupled and cannot be easily disentangled based on geometric features. Second, obtaining a decomposed base shape that preserves the identity of the source mesh without distortion is non-trivial. Third, the geometric details cannot be faithfully represented by conventional unidirectional displacement maps or extracted accurately.

To address these challenges, we propose *InvSculpt*, a novel framework that decomposes a sculpted mesh into an underlying shape and corresponding transferable details represented by a vector displacement field (VDF). We leverage the priors of a text-guided 2D image editing model [Google 2025] to identify details on the source mesh according to user intent, and remove them to obtain an edited image. We then employ a pretrained 3D rectified flow model [Xiang et al. 2024] to perform inversion-based, mask-free removal of details guided by the image to obtain the underlying shape. For detail extraction, we deform the underlying shape toward the source while preserving topology using the Chamfer distance on sampled points and normals. The per-vertex difference between the deformed and original shape yields vector displacements defined on the surface of the underlying shape. We then train a neural network [Ling et al. 2025], using normalized coordinates of the underlying shape as input to map the discrete displacements to a continuous transferable vector displacement field.

We observe that directly applying standard sampling after inversion often suffers from trajectory deviations, leading to noticeable identity shift and low-frequency distortion in the resulting mesh. To mitigate this issue, we introduce a trajectory correction strategy during the sampling process: in the early steps, we first perform unconditional sampling to steer the trajectory back toward its inversion path, and only thereafter apply conditional image guidance to obtain an underlying shape that satisfies the intended detail removal while faithfully preserving the identity of the source mesh. Furthermore, we extract multi-scale features from the two-stage generative model to compute both surface masks and geometric structure masks, preventing redundant displacement noise outside the transfer region and enabling precise extraction of the VDF.

Our experiments demonstrate the effectiveness of our trajectory correction for detail removal and its generalization capabilities in structure-consistent and orientation-aligned 3D generation. Compared with previous methods, our framework achieves significantly higher-quality decomposition, enabling various applications such as geometry redesign and geometric detail transfer.

To summarize, this paper makes the following contributions:

- We design a novel framework for inverse sculpting modeling, i.e., decomposing sculpted meshes into an underlying shape and transferable details.
- We propose a trajectory correction strategy in a 3D rectified flow model during post-inversion sampling to perform detail removal, obtaining a high-fidelity underlying shape.
- We introduce a vector displacement field defined on the surface of the underlying shape as the representation of lossless and transferable geometric details.

2 Related Work

Generative 3D Shape Editing. With the rapid progress of generative models [Labs et al. 2025; Lipman et al. 2023; Rombach et al. 2021; Xiang et al. 2024], the workflow of 3D editing has shifted from traditional manual operations [Biermann et al. 2002; Sorkine et al. 2004] to text- or image-guided editing. In the early stage, representative works [Barda et al. 2024; Dinh et al. 2025; Gao et al. 2023; Li et al. 2024; Liu et al. 2024; Meng et al. 2025; Mikaeili et al. 2023; Sella et al. 2023; Wang et al. 2022, 2025b,a; Zhuang et al. 2024] relied on optimization-based techniques, taking textual prompts as input and

leveraging 2D generative priors such as CLIP [Radford et al. 2021] and score distillation sampling (SDS) [Poole et al. 2022] to optimize implicit or explicit 3D representations. However, such methods are typically computationally expensive. Subsequent works [Bar-On et al. 2025; Barda et al. 2025; Li et al. 2025b; Mu et al. 2023; Yang et al. 2025] adopted feed-forward multi-view diffusion models to accelerate editing by modifying rendered views and lifting the edits back to 3D, but inconsistent predictions often degrade edit fidelity. Another line of work focuses on 3D detailization [Chen et al. 2025, 2021] by training 3D convolutional networks. More recently, the emergence of native 3D generative models has enabled both mask-assisted [Li et al. 2025a] and mask-free [Ye et al. 2025; Zhou et al. 2026] localized 3D editing by exploiting their strong generative priors. However, detail removal editing remains particularly challenging: it is inherently a global editing operation, making it difficult for native 3D editing methods to perform removal while keeping identity unchanged.

Mesh Decomposition. Decomposing a mesh into an underlying shape and other geometric signals, such as high-frequency noise or geometric details, has long been a challenging problem in computer graphics. One common formulation treats mesh decomposition as a mesh denoising task, where the goal is to recover an underlying shape from noisy input meshes. Early techniques addressed this by employing robust statistics and local first-order surface predictors [Jones et al. 2003], or by adapting 2D bilateral filtering to the 3D domain [Fleishman et al. 2003]. Subsequent research shifted toward optimization-based frameworks [Diebel et al. 2006; He and Schaefer 2013; Wang et al. 2014], seeking a denoised mesh that approximates the input while satisfying specific priors imposed on the ground-truth geometry or noise distributions. With the advent of deep learning, a variety of learning-based approaches have been proposed [Hattori et al. 2022; Shen et al. 2022; Wang et al. 2016; Zhao et al. 2021]. While these methods can effectively remove surface noise via geometric signals, they lack the semantic awareness to handle large-scale sculpted structures. Another class of methods treats decomposition as a detail extraction and transfer task. Early approaches [Biermann et al. 2002; Sorkine et al. 2004] applied geometric smoothing to the source mesh to obtain a base shape, then computed the residual details and transferred them using manually specified correspondences. Subsequent methods rely on pre-defined 3D shape correspondence and represent details using unidirectional displacement representations, extracting them through multi-scale mesh hierarchies [Berkiten et al. 2017], various geometric features with neural feature extraction [Li and Zhang 2021], or decompositions as high- and low-frequency SIREN networks [Yifan et al. 2022]. These approaches, constrained by the limited expressiveness of unidirectional displacement representations, struggle to accurately extract the details in sculpted meshes.

3 Method

Our framework is to decompose a sculpted mesh into an underlying shape and transferable details represented as a vector displacement field (Fig. 3). Guided by a user-specified prompt, we first apply a text-driven 2D image editing model to remove details from the source detailed mesh in a semantic-aware manner, and use the resulting image to guide inversion-based, mask-free removal of details with a 3D generative model to obtain the base shape and 3D detail mask. We then deform the masked base shape toward the source mesh using Chamfer distance on sampled points and normals, yielding per-vertex vector displacements. Finally, we train a neural network to map discrete displacements to a continuous vector displacement field. To mitigate the original trajectory drift and achieve the intended detail removal while preserving the source mesh identity, we introduce a trajectory correction strategy. Specifically, we constrain early sampling steps to follow the inversion path, thereby guiding subsequent conditional sampling to converge toward a proper mode.

3.1 Preliminaries

3D Rectified Flow Model. Our underlying shape estimation framework is based on a 3D rectified flow model [Xiang et al. 2024] that utilizes multi-view image features to construct a structured latent representation, defined as a set of local latents $\{(z_i, p_i)\}_{i=1}^{L}$ anchored to active voxels p_i that intersect the object surface. Each latent code $z_i \in \mathbb{R}^C$ encodes fine-grained geometry and appearance within its corresponding voxel. During inference, the generation process begins by sampling noise in the voxel-based latent space, followed by a two-stage denoising procedure. In the structure **(ST)** stage, the rectified flow model predicts voxel occupancy over a 64^3 grid, yielding a voxel-based geometry structure, then the structured latents are denoised in the sparse-latent stage **(SLAT)** to recover fine-grained geometry and texture.

Inversion-Based Editing. Inversion-based image editing has been extensively explored in prior works [Huberman-Spiegelglas et al. 2024; Mokady et al. 2023; Rout et al. 2025; Wang et al. 2024a]. However, the quality of such methods is often limited by inaccurate inversion, which typically stems from the accumulation of numerical errors during ODE integration, leading to noticeable deviations in the reconstructed samples. Recently, RF-Solver [Wang et al. 2024a] addresses this issue by improving inversion accuracy through a second-order Taylor expansion. Given a state x_0, RF-Solver integrates the rectified-flow ODE from data to noise using a Taylor-improved Euler scheme:

$$x_{t-\Delta} = x_t + \Delta f_\theta(x_t, t, c) + \frac{1}{2}\Delta^2 \, \partial_t f_\theta(x_t, t, c), \tag{1}$$

where the temporal derivative $\partial_t f_\theta(x_t, t, c)$ is approximated via a finite-difference formulation:

$$\partial_t f_\theta(x_t, t, c) \approx \frac{f_\theta(x_{t-\Delta/2}, t - \Delta/2, c) - f_\theta(x_t, t, c)}{\Delta/2}. \tag{2}$$

Here, x_t denotes the state at timestep t, $x_{t-\Delta}$ is the predicted state at the next integration step with step size Δ, and $f_\theta(x_t, t, c)$ represents the noise-prediction network with an input condition c. VoxHammer [Li et al. 2025a] further extends inversion-based editing to native 3D generative models [Xiang et al. 2024], enabling mask-assisted local geometry editing.

3.2 Underlying Shape Estimation

We estimate the underlying shape via two components: 3D inversion, which incorporates source mesh priors during sampling based on the inverted noise, and trajectory correction, which steers the sampling path toward the intended modes. Given a source detailed mesh S_{det}, we render it into a single-view image I_{det}. We then perform

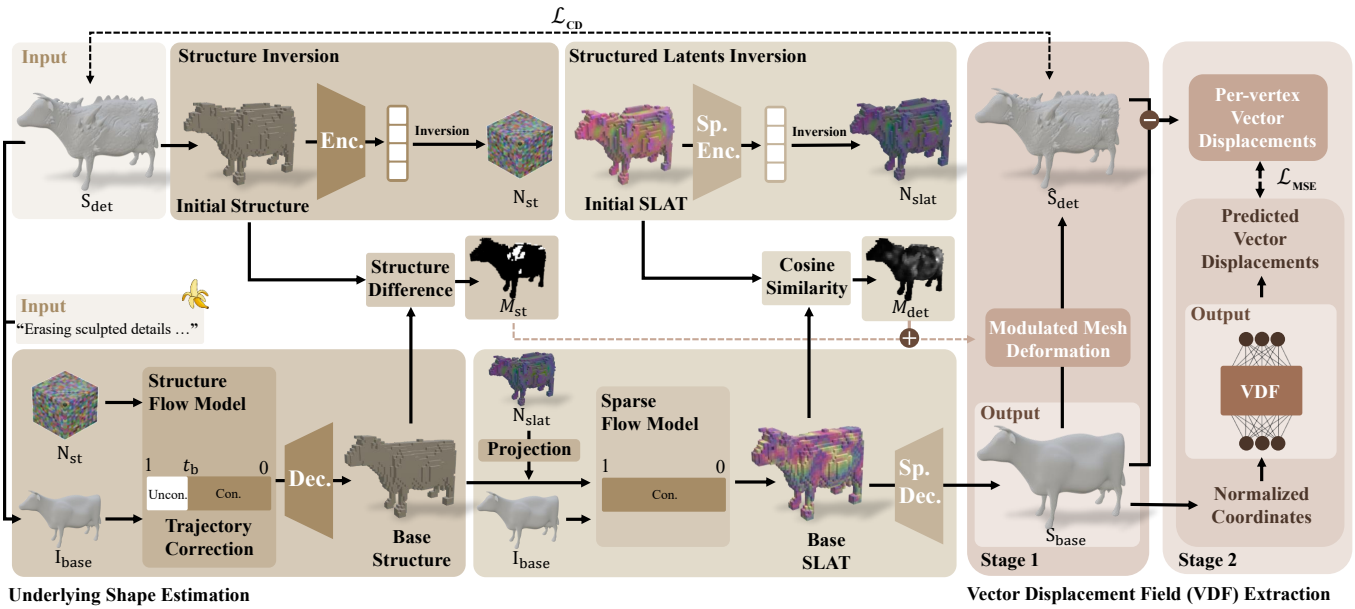

Fig. 3. **Overview of InvSculpt.** Given a source detailed mesh and a user-specified prompt, we first perform 3D inversion on the initial structure and SLAT to obtain the noise representations N_{st} and N_{slat}. We then apply a text-guided 2D image editing model to remove details in the image domain, producing I_{base}, which guides inversion-based 3D detail removal with trajectory correction to recover the underlying shape S_{base} and detail mask M_{st} and M_{det}. Finally, we deform S_{base} under the mask constraint to obtain a detailed mesh $\hat{S}_{det}$ with topology consistent with S_{base}. The per-vertex differences between S_{base} and $\hat{S}_{det}$ yield vector displacements, which are used to train an MLP-based vector displacement field defined on the surface of S_{base}.

a Taylor-improved inversion [Li et al. 2025a; Wang et al. 2024a] within the 3D rectified flow model with I_{det} at both the ST and SLAT stages to map the source detailed mesh to its corresponding noise representation N_{st} and N_{slat}. Specifically, we reverse the forward time schedule $0 = s_0 < s_1 < \cdots < s_T = 1$ by traversing the rectified-flow trajectory backward from timestep s_T to s_0. Throughout both stages, we employ classifier-free guidance ω:

$$f_{\text{cfg}} = (1+\omega)\, f_\theta(\text{cond}) - \omega\, f_\theta(\text{uncond}), \tag{3}$$

where the guidance weight is set to $\omega = 0$ for $t \in [0, 0.5]$ to stabilize the early inversion steps while preserving sufficient semantic sharpness [Li et al. 2025a], and set to $\omega = -1$ for the remaining timesteps. This unconditional inversion in the later stage facilitates subsequent trajectory correction by enabling the sampling process to retrace the inversion path.

During sampling, we apply a text-guided 2D image editing model to remove details from I_{det} by using a user-specified prompt (e.g., erasing sculpted details of horns and scars while preserving the overall smooth structure), producing an image I_{base} that leverages the strong 3D generative prior to semantically guide detail removal over the entire shape. Rectified flow models tend to follow relatively straight trajectories, making the initial velocity direction critical. We observe that directly applying conditional sampling from the beginning causes the trajectory to deviate significantly from the inversion path at early timesteps (Fig. 4), leading the solution to converge toward the mode induced by I_{base} and resulting in identity shift and degraded structures at the structure generation stage.

To address this issue, we adopt a simple yet effective trajectory correction strategy. Specifically, we perform unconditional sampling in the early timesteps t_b by setting $\omega = -1$ for $t \in [t_b, 1]$, forcing the trajectory to follow the inversion path back toward the source mesh mode, and then switch to conditional sampling. In our experiments, we found that using $t_b = 0.8$ can stably correct the convergence behavior to obtain a higher-fidelity base structure. To bridge the structural difference between the base voxels and N_{slat}, we project latent features of N_{slat} to the base structure via n-nearest neighbor query and aggregation process, where we set n to 3 in practice. This yields the base SLAT, which is then fed into the sparse decoder to reconstruct the underlying shape S_{base}. The resulting shape both satisfies the intended removal specified by I_{base} and faithfully preserves the identity of the source mesh.

3.3 Vector Displacement Field Extraction

We first extract two complementary 3D masks at different stages of the detail removal process to enable accurate vector displacement field extraction. Specifically, a structure-level 3D mask is derived from the voxel-based representation in the first stage to capture large-scale geometric changes:

$$M_{\text{st}}(p) = \mathbb{I}(|O_{\text{src}}(p) - O_{\text{base}}(p)| > \tau_{\text{st}}), \tag{4}$$

where p denotes a voxel location, $O_{\text{src}}(p)$ and $O_{\text{base}}(p)$ represent the voxel occupancy values before and after removal, respectively, τ_{st} is a threshold that controls the sensitivity to structural changes, which we set to one voxel difference, and $\mathbb{I}(\cdot)$ is the indicator function. Voxels with $M_{\text{st}}(p) = 1$ (white) correspond to regions exhibiting

significant differences and are identified as removed structures, while voxels with $M_{\text{st}}(p) = 0$ (black) preserve the original geometry. A soft detail-level 3D mask is then computed in the second stage using cosine similarity between latent codes to identify fine-grained surface variations:

$$M_{\text{det}}(p) = 1 - \frac{\langle z_{\text{src}}(p), z_{\text{base}}(p) \rangle}{\|z_{\text{src}}(p)\|_2 \, \|z_{\text{base}}(p)\|_2}, \tag{5}$$

where $z_{\text{src}}(p) \in \mathbb{R}^C$ is the source latent codes obtained by dense multiview visual features on S_{det} extracted from a DINOv2 [Jose et al. 2025] encoder, $z_{\text{base}}(p) \in \mathbb{R}^C$ is latent codes generated by sparse flow model after detail removal, $\langle \cdot, \cdot \rangle$ denotes the inner product, and $\|\cdot\|_2$ is the ℓ_2 norm. Combining these two voxel-based masks as $M_{\text{final}}(p) = M_{\text{st}}(p) + M_{\text{det}}(p)$ yields a more precise localization of geometric details.

Given S_{base}, we follow the framework of a Sobolev preconditioned gradient descent [Nicolet et al. 2021], which achieves smoother deformations without losing detail. In this formulation, the base mesh is reparameterized by the mesh Laplacian L:

$$v^* = (I + \lambda L)v. \tag{6}$$

This preconditioning involves solving a sparse linear system at every iteration, modifying the gradient descent update for each mesh deformation step to:

$$v \leftarrow v - \eta (I + \lambda L)^{-1} \frac{\partial \mathcal{L}_{\text{CD}}}{\partial v}, \tag{7}$$

where $\mathcal{L}_{\text{CD}}$ is Chamfer distance loss on sampled points and normals on deformed mesh and S_{det}, η is the learning rate, I is the identity matrix, and λ is a hyperparameter to control the extent of gradient diffusion over the entire domain which we set to 15 throughout our experiments to obtain high fidelity results [Meng et al. 2025].

We first perform a global mesh deformation to obtain an intermediate source detail mesh $\hat{S}_{\text{det}}$ that is topologically consistent with S_{base}. The voxel-based 3D mask is then projected onto $\hat{S}_{\text{det}}$ to produce a vertex-based 3D mask, which is subsequently used to guide a second deformation step to obtain $\hat{S}^{\text{final}}_{\text{det}}$ so that the base mesh is deformed only in regions containing transferable details. This process effectively suppresses redundant noise (Fig. 6) outside the detail regions when computing per-vertex vector displacements by taking vertex-wise differences between S_{base} and $\hat{S}^{\text{final}}_{\text{det}}$.

We then train a neural network [Ling et al. 2025] that takes as input the coordinates of S_{base} normalized to the range of $[0, 1]$ and predicts the vector displacements, supervised by the per-vertex displacements using MSE loss. This network lifts discrete displacements into a continuous VDF, enabling detail transfer to meshes of arbitrary resolution via 3D shape correspondence obtained by the existing model [Liu et al. 2025], as detailed in Sec. 5.

4 Experiments

We conducted experiments to evaluate the various capabilities of InvSculpt both quantitatively and qualitatively for underlying base mesh estimation and accuracy of detail extraction. We then present an ablation study that validates the significance of our key insight into trajectory correction strategy in the 3D rectified flow model, as well as the effect of the 3D detail mask.

4.1 Qualitative Evaluation

We evaluate our method from two perspectives: fidelity of underlying shape estimation and detail extraction quality. For underlying shape estimation, we compare our approach with several alternatives, including mask-free multiview image editing–based 2D lifting method using EditP23 [Bar-On et al. 2025], direct 3D generation from images using Trellis [Xiang et al. 2024] and Hunyuan3D [Hunyuan3D et al. 2025], and mask-assisted native 3D editing method using VoxHammer [Li et al. 2025a]. Notably, to ensure a fair comparison with VoxHammer, we manually annotate masks on the regions to be removed in the source mesh.

We further evaluate the quality of detail extraction through geometric transfer. Specifically, we adopt a paradigm that first performs transfer in the image domain [Google 2025] and then generates geometry using image-guided 3D generation methods for comparison. We consider three representative image-guided 3D generation approaches: Trellis and Hunyuan3D that directly generate 3D geometry from images, and Phidias [Wang et al. 2024b], which introduces an explicit 3D reference to provide geometric supervision during image-guided generation.

In the comparison of underlying shape estimation (Fig. 11), EditP23 suffers from multi-view editing inconsistencies and limitations of the subsequent mesh reconstruction algorithms, which lead to degraded quality. Trellis and Hunyuan3D benefit from strong generative priors, improving the quality of meshes generated from a single image. However, due to the lack of 3D information from the source mesh, their results often exhibit noticeable identity drift. VoxHammer injects source mesh geometry through inversion, but detail removal remains a global editing that is difficult to specify with masks. Even though we manually annotate multiple masks for each example, the edited results still have incomplete removal and identity shift. In contrast, our method incorporates source mesh information via inversion and further introduces trajectory correction during the sampling process, enabling the final results to both satisfy the intended removal and faithfully preserve the identity.

In geometric transfer comparison (Fig. 8), we observe that image editing models are relatively stable for detail removal of a single image, yet unstable when transferring details between two images. Directly generating transferred results from an image often fails to maintain the target mesh's identity and may add or lose details. Moreover, even with the target mesh as a 3D constraint, Phidias still produces blurry details. In contrast, our method extracts a reusable vector displacement field defined on the surface of the recovered underlying shape and enables applying 3D shape correspondence via a frozen model to transfer these details. This ensures that the target mesh preserves its identity while accurately inheriting high-quality geometric details.

4.2 Quantitative Evaluation

We quantitatively evaluated the underlying shape by regarding geometric structure fidelity and identity consistency with the source mesh. We use 20 sculpted meshes for decomposition.

Geometric Structure Fidelity. We evaluate underlying shape fidelity by computing the average Chamfer distance between two point sets, each consisting of 20,000 samples drawn from the decomposed

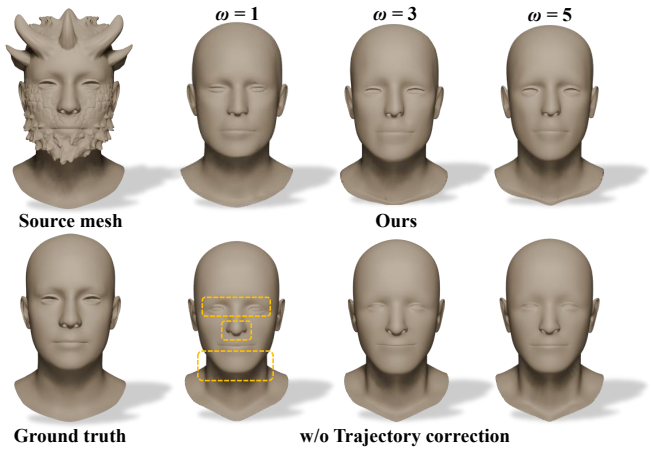

Fig. 4. **Effect of trajectory correction.** Our method effectively mitigates trajectory drift in standard sampling across a reasonable range of CFG values by correcting the sampling trajectory at early stages, thereby preserving the source mesh identity (e.g., eyes, nose, and jaw of the head) and avoiding low-frequency distortions. The ground truth mesh is obtained from flame [Li et al. 2017].

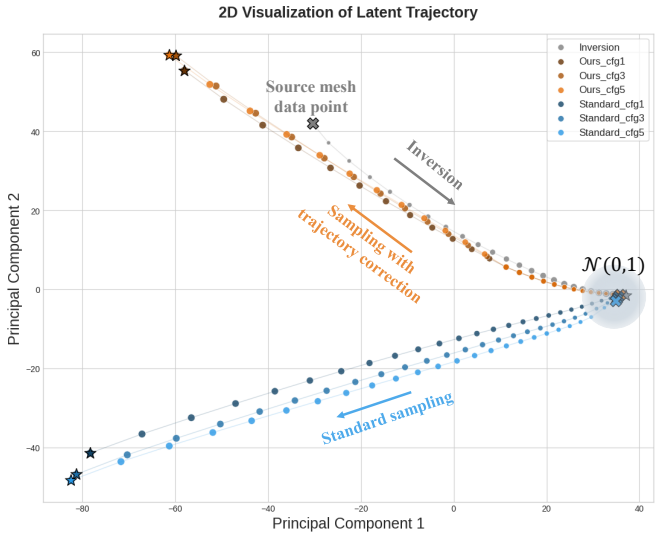

Fig. 5. **Visualization of trajectories in 3D rectified flow model.** We visualize the inversion trajectory in Fig. 4, together with sampling trajectories under three different CFG values, with and without trajectory correction, by projecting them to 2D using PCA. The results show that standard sampling exhibits drift in the early stage, whereas our method effectively corrects the trajectories, yielding outputs that both follow image guidance and preserve the source mesh identity.

underlying shape and the ground truth of the source mesh before sculpting. Compared to alternative methods, our approach achieves the highest fidelity in recovering the underlying shape.

Identity Consistency. We compute CLIP scores between the detail-erased images produced by the 2D editing model and renderings in 24 different views of the underlying shapes decomposed by different methods. Direct generation from a single image typically results in identity drift, whereas our method effectively maintains identity through trajectory correction.

User Study. We further conducted a user study to assess the geometric fidelity and identity consistency of the decomposed underlying shapes, as well as the quality of detail extraction as reflected by subsequent geometric detail transfer. For each part, we invited 25 participants to evaluate 12 decomposed results of different methods, rating them on a scale from 1 to 5, where higher scores indicate better performance. As shown in Table 1, participants preferred our method by a significant margin.

Table 1. **User evaluation of mesh decomposition.**

Method	Fidelity of Underlying Shape ↑
VoxHammer	2.3778
EditP23	1.5722
Trellis	3.2944
Hunyuan3D	2.9444
Ours	**4.7056**
	Quality of Detail Extraction ↑
Phidias	1.7167
Trellis	2.8056
Hunyuan3D	2.4556
Ours	**4.3667**

4.3 Ablation Study

Effect of Trajectory Correction Strategy. Fig. 4 compares the underlying shape estimation results produced by the standard conditional sampling process and by our trajectory correction strategy, evaluated under three classifier-free guidance (CFG) strengths: 1, 3, and 5. Unlike the original sampling process, our approach constrains the early sampling steps to follow the reverse direction of the inversion path, correcting the subsequent conditional sampling process. As a result, across all CFG strengths, the sampling process consistently converges to a mode that both preserves the identity of the source mesh and achieves the desired detail removal effect.

Key Insight. We further visualize and analyze the sampling trajectories (Fig. 5) for the examples shown in Fig. 4. Specifically, we project the latents from the inversion path, the three standard sampling trajectories (blue lines), and the three corrected trajectories (yellow lines) into 2D using PCA. In the visualizations, the rightmost point corresponds to the noise obtained after inversion, and the left gray point denotes the latent encoded from the source mesh.

As the rectified flow model produces the relatively straight trajectories during sampling [Albergo and Vanden-Eijnden 2023; Lipman et al. 2023; Liu et al. 2023], the early trajectory direction plays a critical role in determining the final converged mode. As shown in Fig. 5, standard sampling deviates from the inversion path at early timesteps, drifting prematurely toward the mode induced by the image. This leads to noticeable inconsistencies in the final mesh, including deviations from the source mesh identity and mismatched orientation. In contrast, our method constrains the early sampling

Table 2. **Quantitative comparison of baseline methods.** We evaluate underlying shape fidelity using the average Chamfer distance against the ground truth. Identity preservation is assessed by CLIP scores between detail-erased images and 24-view renderings of decomposed shapes.

Method	Chamfer Distance ↓	CLIP Score ↑
Hunyuan3D	0.0727	0.9433
TRELLIS	0.0557	0.9371
EditP23	0.1262	0.8919
VoxHammer	0.0488	0.9534
Ours	**0.0334**	**0.9901**

steps to follow the inversion path back toward the source mesh mode before applying conditional sampling. This ensures robust behavior across a reasonable range of CFG strengths and yields results that preserve the source mesh's geometric attributes, such as orientation and low-frequency structure, while achieving the intended detail removal.

This observation further motivates us to apply trajectory correction to training-free generation tasks. As illustrated in Fig. 9, given an A-pose human body as the source mesh to be inverted and an input image of a person in an arbitrary pose, applying trajectory correction enables the model to generate a human mesh that aligns with both the pose and orientation of the source mesh while remaining faithful to the appearance specified by the input image. This demonstrates the generalization ability of our approach to structure-consistent and orientation-aligned generation tasks.

Effect of Extracted 3D Detail Mask. Fig. 6 illustrates the comparison among detail transfer results without using the trajectory correction strategy and mask, without using the mask, and our full framework. When distorted results obtained without trajectory correction are used as the underlying shape for detail extraction, the accuracy of the resulting vector displacement field degrades significantly. Moreover, during detail removal, inversion inaccuracies and the limited reconstruction fidelity of the 3D VAE may introduce redundant noise into the extracted vector displacement field outside the detail regions. By using region masks that localize the details during the detail removal process, we are able to obtain a more accurate VDF, leading to higher-quality transfer.

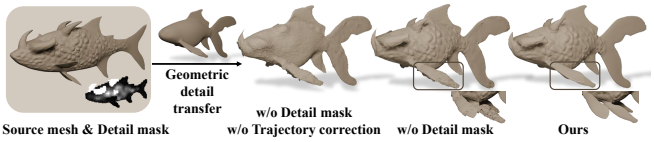

Fig. 6. **Effect of the detail mask.** While extracting details directly from distorted meshes often leads to failure, our approach robustly extracts high-quality geometric details, effectively suppressing noisy displacements.

5 Applications

Geometric Detail Transfer. Our method is capable of extracting transferable vector displacement fields from diverse sculpted meshes. Then, we can accurately transfer details across models within the same category via a pretrained 3D shape correspondence model [Liu et al. 2025]. Specifically, we initialize correspondences between the target and base shapes using nearest-neighbor matching in the extracted semantic feature space, and further refine these matches with Smooth Discrete Optimization [Magnet et al. 2022], which iteratively solves for functional maps in a coarse-to-fine manner to recover a smooth point-to-point correspondence. Based on the resulting 3D correspondence, we query the vector displacement field using the corresponding normalized vertex coordinates on the source base mesh, and transfer the vector displacements with standard local coordinate frame construction [Berkiten et al. 2017] to synthesize geometric details. Various geometric details from source meshes, including fine-grained surface details such as fish scales and stone cracks, as well as large-scale structures like horns, can be seamlessly transferred to diverse target models (Fig. 10).

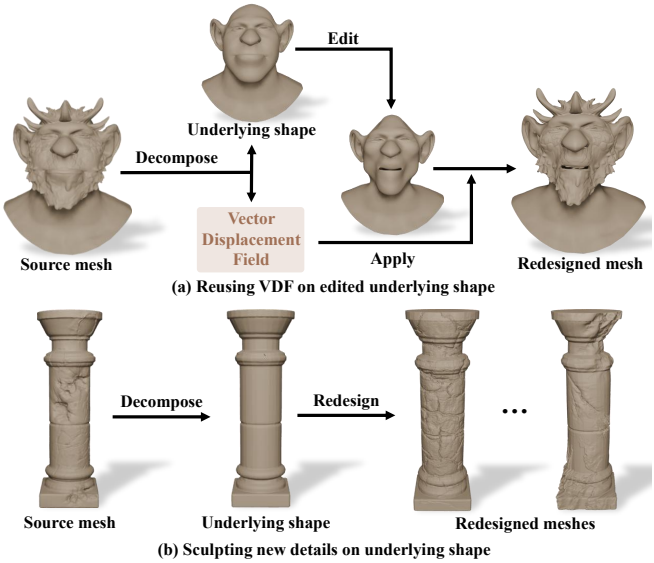

Fig. 7. **Geometry redesign.** After decomposition, our framework enables reusing the extracted VDF on edited underlying shapes, significantly reducing the complexity of direct high-resolution mesh editing. In addition, new details can be sculpted on the underlying shape to efficiently produce diverse shape variations.

Geometry Redesign. Once the underlying shape is obtained, users can freely perform further sculpting on it. As shown in Fig. 7, we decompose the base model of an intact stone pillar from a damaged pillar, enabling users to sculpt alternative expressive variants. Moreover, due to the topology-agnostic property of the vector displacement field, our approach supports large-scale deformations of the base model while preserving fine details. The base model of a monster head can be globally reshaped (e.g., made slimmer), after which the geometric details are automatically reattached by using the normalized coordinates before being reshaped to query the vector displacements, avoiding the detail loss and operational difficulty associated with directly editing high-resolution meshes.

6 Conclusion

We have presented InvSculpt, a novel framework that decomposes a sculpted mesh into an underlying shape and its transferable details represented by a vector displacement field. Identifying and obtaining the underlying shape and the transferable details in a source mesh is inherently ambiguous, making it challenging for traditional and generative geometry processing methods. Therefore, we leverage the strong semantic priors of 2D editing models and integrate them into a native 3D rectified flow model to achieve reliable decomposition. To further mitigate identity drift caused by trajectory deviation during the standard sampling, we introduce a trajectory correction strategy that enables high-fidelity detail removal for obtaining identity-preserved results. Moreover, we compute 3D detail masks from the native 3D generative model to more accurately extract the vector displacement field. The resulting underlying shape and transferable vector displacement field support a wide range of downstream applications, including geometry redesign and geometric detail transfer.

Limitations and Future Work. While our method can achieve high-quality decomposition, the fidelity of the underlying shape remains constrained by the generative priors of the 3D generation models as well as the reconstruction capacity of the 3D VAE. In addition, the quality of detail transfer is limited by the 3D shape correspondences provided by the pretrained semantic model, where inaccurate semantic features may lead to subpar results. In the future, we plan to use our framework to generate datasets of paired sculpted and base meshes, which will be beneficial for training end-to-end models that can achieve detail removal more efficiently.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Nos. 62502410 and U25A20384) and the Laoshan Laboratory (No. LSKJ202300305).

References

Michael S. Albergo and Eric Vanden-Eijnden. 2023. Building Normalizing Flows with Stochastic Interpolants. International Conference on Learning Representations.

Roi Bar-On, Dana Cohen-Bar, and Daniel Cohen-Or. 2025. EditP23: 3D Editing via Propagation of Image Prompts to Multi-View. arXiv:2506.20652

Amir Barda, Matheus Gadelha, Vladimir G. Kim, Noam Aigerman, Amit H. Bermano, and Thibault Groueix. 2025. Instant3dit: Multiview Inpainting for Fast Editing of 3D Objects. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 16273–16282.

Amir Barda, Vladimir G. Kim, Noam Aigerman, Amit H. Bermano, and Thibault Groueix. 2024. MagicClay: Sculpting Meshes with Generative Neural Fields. *SIGGRAPH Asia (Conference track)* (2024).

Sema Berkiten, Maciej Halber, Justin Solomon, Chongyang Ma, Hao Li, and Szymon Rusinkiewicz. 2017. Learning Detail Transfer based on Geometric Features. *Computer Graphics Forum* 36 (05 2017), 361–373.

Henning Biermann, Ioana Martin, Fausto Bernardini, and Denis Zorin. 2002. Cut-and-paste Editing of Multiresolution Surfaces. *ACM Trans. Graph.* 21, 3 (July 2002), 312–321.

Blender. 2025. Blender. https://www.blender.org. Accessed Mar 5, 2025.

Qimin Chen, Zhiqin Chen, Vladimir G Kim, Noam Aigerman, Hao Zhang, and Siddhartha Chaudhuri. 2025. DECOLLAGE: 3D Detailization by Controllable, Localized, and Learned Geometry Enhancement. In *European Conference on Computer Vision*.

Zhiqin Chen, Vladimir G. Kim, Matthew Fisher, Noam Aigerman, Hao Zhang, and Siddhartha Chaudhuri. 2021. DECOR-GAN: 3D Shape Detailization by Conditional Refinement. *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

James R. Diebel, Sebastian Thrun, and Michael Brünig. 2006. A Bayesian method for probable surface reconstruction and decimation. 25, 1 (Jan. 2006), 39–59.

Nam Anh Dinh, Itai Lang, Hyunwoo Kim, Oded Stein, and Rana Hanocka. 2025. Geometry in Style: 3D Stylization via Surface Normal Deformation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 28456–28467.

Shachar Fleishman, Iddo Drori, and Daniel Cohen-Or. 2003. Bilateral mesh denoising. *ACM Trans. Graph.* 22, 3 (July 2003), 950–953.

William Gao, Noam Aigerman, Thibault Groueix, Vova Kim, and Rana Hanocka. 2023. TextDeformer: Geometry Manipulation Using Text Guidance. In *ACM SIGGRAPH 2023 Conference Proceedings* (Los Angeles, CA, USA) *(SIGGRAPH '23)*. Association for Computing Machinery, New York, NY, USA, Article 82, 11 pages.

Google. 2025. Nano Banana. https://www.nano-banana.ai.

Shota Hattori, Tatsuya Yatagawa, Yutaka Ohtake, and Hiromasa Suzuki. 2022. Learning Self-prior for Mesh Denoising using Dual Graph Convolutional Networks. In *Proceedings of the European Conference on Computer Vision (ECCV)*.

Lei He and Scott Schaefer. 2013. Mesh denoising via L0 minimization. 32, 4, Article 64 (July 2013), 8 pages.

Inbar Huberman-Spiegelglas, Vladimir Kulikov, and Tomer Michaeli. 2024. An Edit Friendly DDPM Noise Space: Inversion and Manipulations. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 12469–12478.

Team Hunyuan3D, Shuhui Yang, Mingxin Yang, Yifei Feng, Xin Huang, Sheng Zhang, Zebin He, Di Luo, Haolin Liu, Yunfei Zhao, et al. 2025. Hunyuan3D 2.1: From Images to High-Fidelity 3D Assets with Production-Ready PBR Material. *arXiv* 2506.15442 (2025).

Thouis R. Jones, Frédo Durand, and Mathieu Desbrun. 2003. Non-iterative, feature-preserving mesh smoothing. 22, 3 (July 2003), 943–949.

Cijo Jose, Théo Moutakanni, Dahyun Kang, Federico Baldassarre, Timothée Darcet, Hu Xu, Daniel Li, Marc Szafraniec, Michaël Ramamonjisoa, Maxime Oquab, Oriane Siméoni, Huy V. Vo, Patrick Labatut, and Piotr Bojanowski. 2025. DINOv2 Meets Text: A Unified Framework for Image- and Pixel-Level Vision-Language Alignment. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 24905–24916.

Black Forest Labs, Stephen Batifol, Andreas Blattmann, Frederic Boesel, Saksham Consul, Cyril Diagne, Tim Dockhorn, Jack English, Zion English, Patrick Esser, Sumith Kulal, Kyle Lacey, Yam Levi, Cheng Li, Dominik Lorenz, Jonas Müller, Dustin Podell, Robin Rombach, Harry Saini, Axel Sauer, and Luke Smith. 2025. FLUX.1 Kontext: Flow Matching for In-Context Image Generation and Editing in Latent Space. arXiv:2506.15742

Lin Li, Zehuan Huang, Haoran Feng, Gengxiong Zhuang, Rui Chen, Chunchao Guo, and Lu Sheng. 2025a. VoxHammer: Training-Free Precise and Coherent 3D Editing in Native 3D Space. *arXiv* 2508.19247 (2025).

Manyi Li and Hao Zhang. 2021. d^2im-net: learning detail disentangled implicit fields from single images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 10246–10255.

Peng Li, Suizhi Ma, Jialiang Chen, Yuan Liu, Congyi Zhang, Wei Xue, Wenhan Luo, Alla Sheffer, Wenping Wang, and Yike Guo. 2025b. CMD: Controllable Multiview Diffusion for 3D Editing and Progressive Generation *(SIGGRAPH Conference Papers '25)*. Association for Computing Machinery, New York, NY, USA, Article 81, 10 pages.

Tianye Li, Timo Bolkart, Michael J Black, Hao Li, and Javier Romero. 2017. Learning a Model of Facial Shape and Expression from 4D Scans. *ACM Trans. Graph.* 36, 6 (2017), 194–1.

Yuhan Li, Yishun Dou, Yue Shi, Yu Lei, Xuanhong Chen, Yi Zhang, Peng Zhou, and Bingbing Ni. 2024. FocalDreamer: Text-Driven 3D Editing via Focal-Fusion Assembly. *Proceedings of the AAAI Conference on Artificial Intelligence* 38 (03 2024), 3279–3287.

Selena Ling, Merlin Nimier-David, Alec Jacobson, and Nicholas Sharp. 2025. Stochastic Preconditioning for Neural Field Optimization. *ACM Trans. Graph.* 44, 4, Article 84 (July 2025), 10 pages.

Yaron Lipman, Ricky T. Q. Chen, Heli Ben-Hamu, Maximilian Nickel, and Matt Le. 2023. Flow Matching for Generative Modeling. arXiv:2210.02747

Feng-Lin Liu, Hongbo Fu, Yu-Kun Lai, and Lin Gao. 2024. SketchDream: Sketch-based Text-to-3D Generation and Editing. *ACM Trans. Graph* 43, 4 (2024).

Minghua Liu, Mikaela Angelina Uy, Donglai Xiang, Hao Su, Sanja Fidler, Nicholas Sharp, and Jun Gao. 2025. PartField: Learning 3D Feature Fields for Part Segmentation and Beyond. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. 9704–9715.

Xingchao Liu, Chengyue Gong, and Qiang Liu. 2023. Flow Straight and Fast: Learning to Generate and Transfer Data with Rectified Flow. International Conference on Learning Representations.

Robin Magnet, Jing Ren, Olga Sorkine-Hornung, and Maks Ovsjanikov. 2022. Smooth Non-Rigid Shape Matching via Effective Dirichlet Energy Optimization. In *2022 International Conference on 3D Vision (3DV)*. 495–504.

Hengyu Meng, Duotun Wang, Zhijing Shao, Ligang Liu, and Zeyu Wang. 2025. Text2VDM: Text to Vector Displacement Maps for Expressive and Interactive 3D Sculpting. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. 16882–16892.

Aryan Mikaeili, Or Perel, Mehdi Safaee, Daniel Cohen-Or, and Ali Mahdavi-Amiri. 2023. SKED: Sketch-guided Text-based 3D Editing. *2023 IEEE/CVF International Conference*

on Computer Vision (ICCV) (2023).

Ron Mokady, Amir Hertz, Kfir Aberman, Yael Pritch, and Daniel Cohen-Or. 2023. Null-text Inversion for Editing Real Images using Guided Diffusion Models. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 6038–6047.

Tai-Jiang Mu, Hao-Xiang Chen, Jun-Xiong Cai, and Ning Guo. 2023. Neural 3D reconstruction from sparse views using geometric priors. *Computational Visual Media* 9, 4 (2023), 687–697.

Baptiste Nicolet, Alec Jacobson, and Wenzel Jakob. 2021. Large Steps in Inverse Rendering of Geometry. *ACM Trans. Graph* 40, 6 (Dec 2021).

Ben Poole, Ajay Jain, Jonathan T Barron, and Ben Mildenhall. 2022. DreamFusion: Text-to-3D using 2D Diffusion. In *The Eleventh International Conference on Learning Representations*.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. 2021. Learning Transferable Visual Models From Natural Language Supervision. *CoRR* abs/2103.00020 (2021). arXiv:2103.00020 https://arxiv.org/abs/2103.00020

Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. 2021. High-Resolution Image Synthesis with Latent Diffusion Models. arXiv:2112.10752

Litu Rout, Yujia Chen, Nataniel Ruiz, Constantine Caramanis, Sanjay Shakkottai, and Wen-Sheng Chu. 2025. Semantic Image Inversion and Editing using Rectified Stochastic Differential Equations. In *International Conference on Representation Learning*, Y. Yue, A. Garg, N. Peng, F. Sha, and R. Yu (Eds.), Vol. 2025. 77330–77365.

Etai Sella, Gal Fiebelman, Peter Hedman, and Hadar Averbuch-Elor. 2023. Vox-E: Text-guided Voxel Editing of 3D Objects. *2023 IEEE/CVF International Conference on Computer Vision (ICCV)* (2023), 430–440.

Yuefan Shen, Hongbo Fu, Zhongshuo Du, Xiang Chen, Evgeny Burnaev, Denis Zorin, Kun Zhou, and Youyi Zheng. 2022. GCN-Denoiser: Mesh Denoising with Graph Convolutional Networks. *ACM Trans. Graph.* 41, 1, Article 8 (Feb. 2022), 14 pages.

O. Sorkine, D. Cohen-Or, Y. Lipman, M. Alexa, C. Rössl, and H.-P. Seidel. 2004. Laplacian surface editing. In *Proceedings of the 2004 Eurographics/ACM SIGGRAPH Symposium on Geometry Processing* (Nice, France) *(SGP '04)*. Association for Computing Machinery, New York, NY, USA, 175–184.

Can Wang, Menglei Chai, Mingming He, Dongdong Chen, and Jing Liao. 2022. Clip-nerf: Text-and-image Driven Manipulation of Neural Radiance Fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 3835–3844.

Chen Wang, Hao-Yang Peng, Ying-Tian Liu, Jiatao Gu, and Shi-Min Hu. 2025b. Diffusion Models for 3D Generation: A Survey. *Computational Visual Media* 11, 1 (2025), 1–28.

Duotun Wang, Hengyu Meng, Zeyu Cai, Zhijing Shao, Qianxi Liu, Lin Wang, Mingming Fan, Xiaohang Zhan, and Zeyu Wang. 2025a. HeadEvolver: Text to Head Avatars via Expressive and Attribute-Preserving Mesh Deformation. *International Conference on 3D Vision*.

Jiangshan Wang, Junfu Pu, Zhongang Qi, Jiayi Guo, Yue Ma, Nisha Huang, Yuxin Chen, Xiu Li, and Ying Shan. 2024a. Taming Rectified Flow for Inversion and Editing. *arXiv* 2411.04746 (2024).

Peng-Shuai Wang, Yang Liu, and Xin Tong. 2016. Mesh denoising via Cascaded Normal Regression. 35, 6, Article 232 (Dec. 2016), 12 pages.

Ruimin Wang, Zhouwang Yang, Ligang Liu, Jiansong Deng, and Falai Chen. 2014. Decoupling Noise and Features via Weighted L1-analysis Compressed Sensing. *ACM Trans. Graph.* 33, 2, Article 18 (April 2014), 12 pages.

Zhenwei Wang, Tengfei Wang, Zexin He, Gerhard Hancke, Ziwei Liu, and Rynson Lau. 2024b. Phidias: A Generative Model for Creating 3D Content from Text, Image, and 3D Conditions with Reference-Augmented Diffusion. In *International Conference on Learning Representations*.

Jianfeng Xiang, Zelong Lv, Sicheng Xu, Yu Deng, Ruicheng Wang, Bowen Zhang, Dong Chen, Xin Tong, and Jiaolong Yang. 2024. Structured 3D Latents for Scalable and Versatile 3D Generation. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

Yuezhi Yang, Qimin Chen, Vladimir Kim, Siddhartha Chaudhuri, Qixing Huang, and Zhiqin Chen. 2025. GenVDM: Generating Vector Displacement Maps From a Single Image. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

Junliang Ye, Shenghao Xie, Ruowen Zhao, Zhengyi Wang, Hongyu Yan, Wenqiang Zu, Lei Ma, and Jun Zhu. 2025. NANO3D: A Training-Free Approach for Efficient 3D Editing Without Masks. arXiv:2510.15019

Wang Yifan, Lukas Rahmann, and Olga Sorkine-hornung. 2022. Geometry-Consistent Neural Shape Representation with Implicit Displacement Fields. In *International Conference on Learning Representations*.

ZBrush. 2025. ZBrush. https://www.maxon.net/zbrush. Accessed Oct 18, 2024.

Wenbo Zhao, Xianming Liu, Yongsen Zhao, Xiaopeng Fan, and Debin Zhao. 2021. NormalNet: Learning-Based Mesh Normal Denoising via Local Partition Normalization. *IEEE Transactions on Circuits and Systems for Video Technology* 31, 12 (2021), 4697–4710.

Zhenglin Zhou, Fan Ma, Chengzhuo Gui, Xiaobo Xia, Hehe Fan, Yi Yang, and Tat-Seng Chua. 2026. AnchorFlow: Training-Free 3D Editing via Latent Anchor-Aligned Flows. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

Jingyu Zhuang, Di Kang, Yan-Pei Cao, Guanbin Li, Liang Lin, and Ying Shan. 2024. TIP-Editor: An Accurate 3D Editor Following Both Text-Prompts and Image-Prompts. *ACM Trans. Graph.* 43, 4, Article 121 (July 2024), 12 pages.

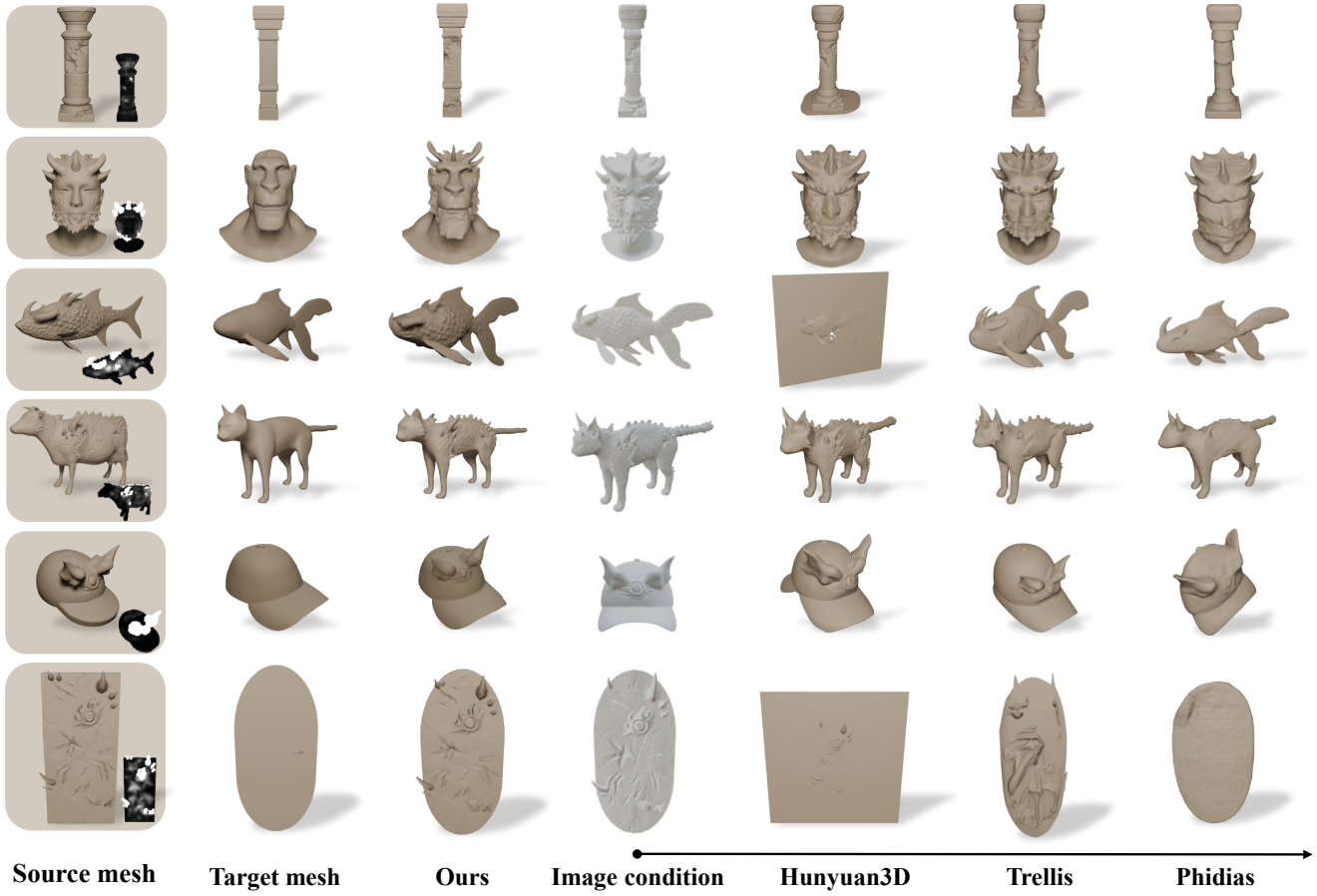

Fig. 8. **Qualitative comparisons of geometric detail transfer.** Our method extracts high-fidelity geometric details from the source mesh, enabling high-quality detail transfer by combining with existing 3D shape correspondence techniques.

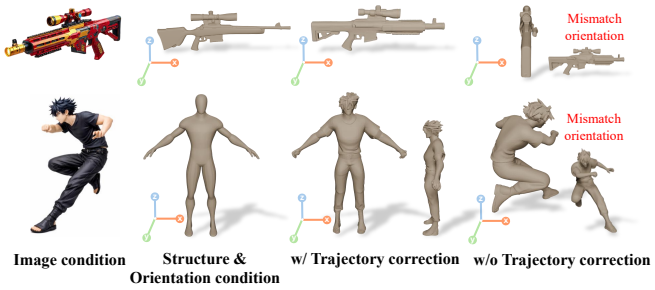

Fig. 9. **Generalization of trajectory correction.** Trajectory correction effectively injects structural and orientation constraints from the source mesh into the generation process, enabling training-free orientation-consistent synthesis and structure-consistent generation.

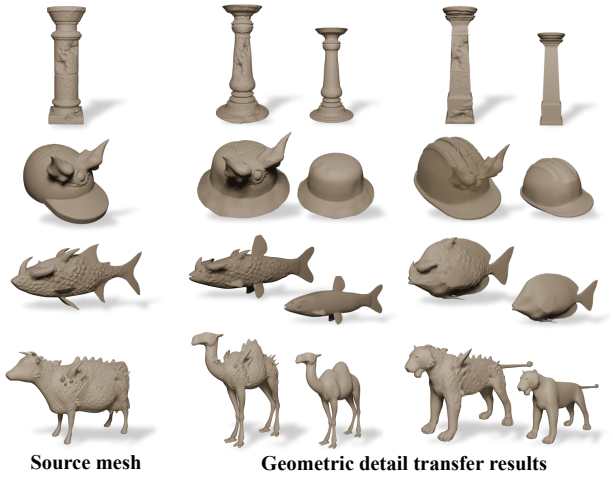

Fig. 10. **Geometric detail transfer results.** Our method extracts high-fidelity geometric details from the source meshes to facilitate high-quality transfer, enabling rapid prototyping and scalable creation of thematically consistent 3D assets.

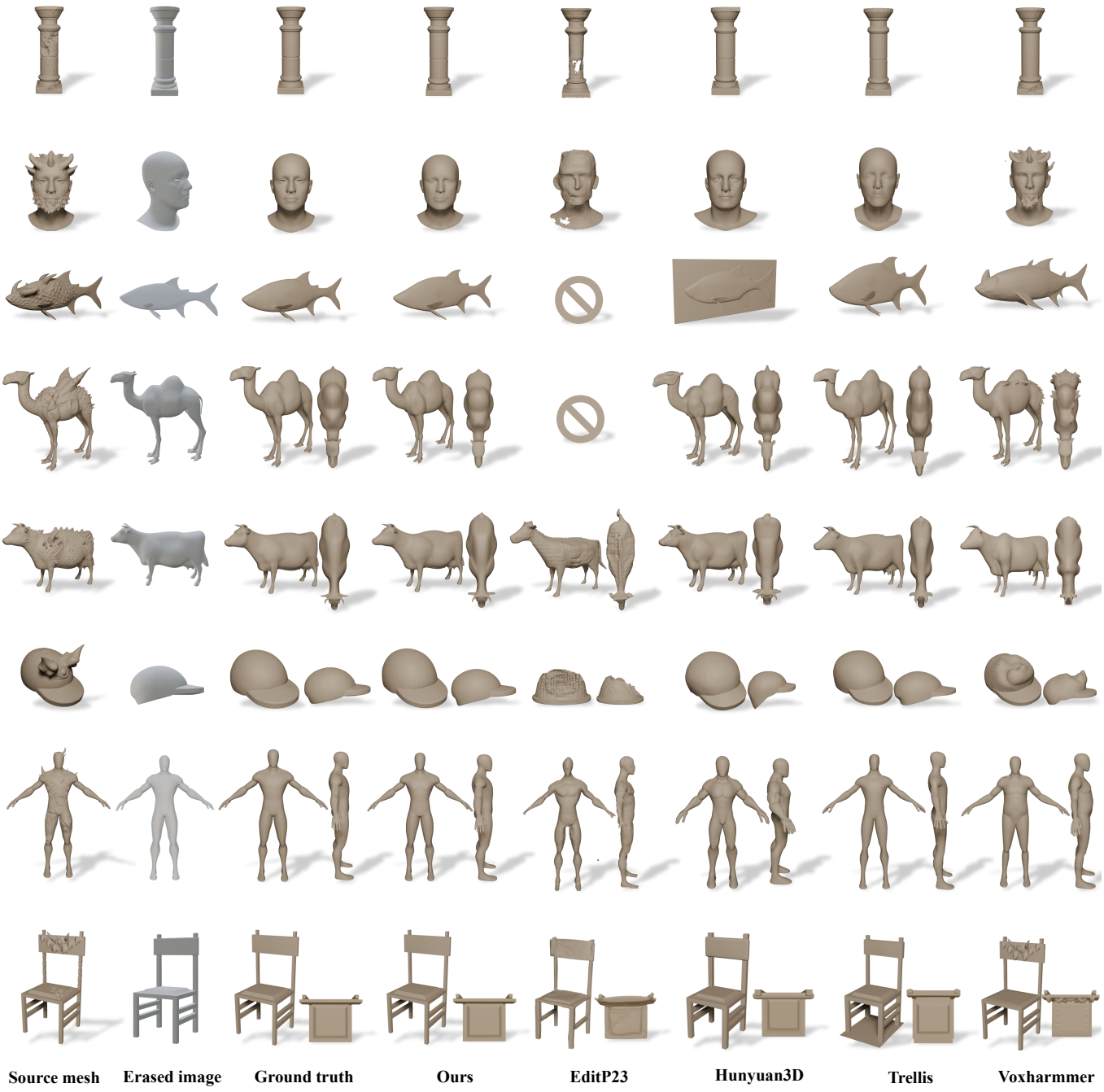

Fig. 11. **Qualitative comparisons of underlying shape estimation.** InvSculpt aims to address the issue of identity drift, which is an inherent limitation of image-to-3D models rather than being caused by the image editing model. The results show that our method effectively preserves both the identity and the geometric structure of the source mesh, enabling more accurate detail extraction and more reliable downstream geometry redesign.