

Geometry-Guided Latent Diffusion for Localized 2D Surface Carving

Jung Yang
National Cheng Kung University
Tainan, Taiwan
andyjungyang@gmail.com

Dong-Yi Wu
National Cheng Kung University
Tainan, Taiwan
cutechubbit@gmail.com

Sheng-Yi Yao
National Cheng Kung University
Tainan, Taiwan
ND8081018@gs.ncku.edu.tw

Wei Chen
National Cheng Kung University
Tainan, Taiwan
vivian1357910@gmail.com

Thi-Ngoc-Hanh Le
International University, VNU-HCM
Ho Chi Minh, Vietnam
ltnhanh@hcmiu.edu.vn

Tong-Yee Lee
National Cheng Kung University
Tainan, Taiwan
tonylee@mail.ncku.edu.tw

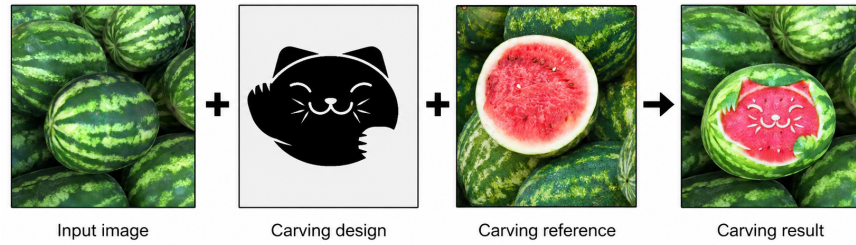


Figure 1: Our method takes an input image, a carving design, and a reference style image to synthesize a carved surface result. The design pattern dictates the shape and layout of the carving, while the style reference guides the local texture and material details.

CCS Concepts

• Computing methodologies → Image manipulation.

Keywords

Diffusion Model, Image Editing, Carving Art

ACM Reference Format:

Jung Yang, Dong-Yi Wu, Sheng-Yi Yao, Wei Chen, Thi-Ngoc-Hanh Le, and Tong-Yee Lee. 2026. Geometry-Guided Latent Diffusion for Localized 2D Surface Carving. In *SIGGRAPH Asia 2026 Posters (SA Posters '26)*, December 01–04, 2026, Kuala Lumpur, Malaysia. ACM, New York, NY, USA, 3 pages. <https://doi.org/10.1145/3829333.3847894>

1 Introduction

Generative image editing models have made significant progress in semantic manipulation, yet performing precise, geometrically constrained edits—such as 2D surface carving or engraving—remains challenging. This capability is essential for digital artists and crafters who need to visualize intricate designs—like jack-o'-lantern carvings on varied materials—before procuring physical supplies or making irreversible cuts. Creating realistic physical edits requires

accurately synthesizing surface depth, local material shifts, and ambient lighting transitions. While some computer graphics pipelines attempt this in 3D, programmatic mesh reconstruction incurs heavy computational overhead, and manually carving into 3D geometry requires specialized technical skills that are impractical for average users.

Conversely, standard 2D image editing frameworks and zero-shot material transfer methods, such as SDEdit [Meng et al. 2022], ZeST [Cheng et al. 2024], and DealMaTe [Huang et al. 2026], struggle with localized structural cuts. Because these methods apply global material swaps or basic spatial masks, they fail to accurately replicate the internal carving material and fine details from the reference image. Furthermore, they frequently shift global illumination or disrupt the natural textures of the unedited background surface.

To address these limitations, we present a reference-guided 2D surface sculpting framework that operates without explicit 3D meshes or text prompts. By combining explicit depth conditioning with multi-stage feature injection and localized latent blending, our pipeline ensures that synthesized carving details conform naturally to the target object's surface contours and lighting, while keeping the unedited background intact.

2 Method

An overview of our system pipeline can be seen in Figure 2. Our framework builds upon a pre-trained latent diffusion model and



This work is licensed under a Creative Commons Attribution 4.0 International License. *SA Posters '26, Kuala Lumpur, Malaysia*
© 2026 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-2839-6/2026/12
<https://doi.org/10.1145/3829333.3847894>

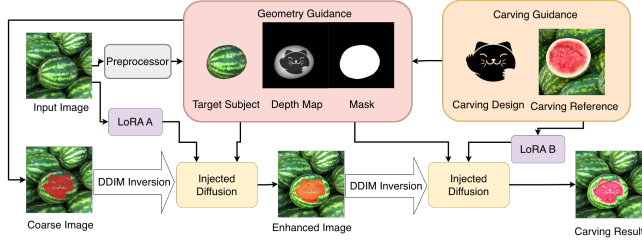


Figure 2: System architecture of our proposed framework. The pipeline processes a segmented canvas, carving design map, and style reference through texture adaptation, depth-conditioned geometry guidance, and localized feature injection to generate a seamlessly carved target image.

introduces three key modules to control surface texture, geometric depth, and boundary alignment across the edited regions.

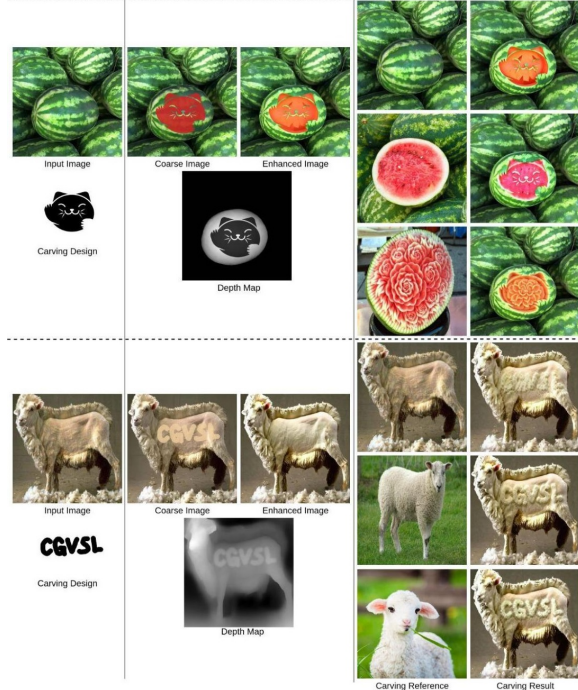


Figure 3: Surface carving results. (Top) A cat design carved onto a watermelon canvas. (Bottom) Geometric letter carving synthesized onto a sheep canvas. Both demonstrate variations across different texture references.

Dual-Exemplar Texture Adaptation. Physical surface carving involves two distinct visual properties: the native surface material (e.g., smooth watermelon skin) and the newly exposed internal cut (e.g., rough inner pulp). To model both behaviors without feature bleeding, we utilize subject-driven fine-tuning techniques (Dream-Booth [Ruiz et al. 2023]) to train two separate Low-Rank Adaptation (LoRA) modules. **LoRA A** is fine-tuned on the unedited target



Figure 4: Visual comparisons between our method and priors.

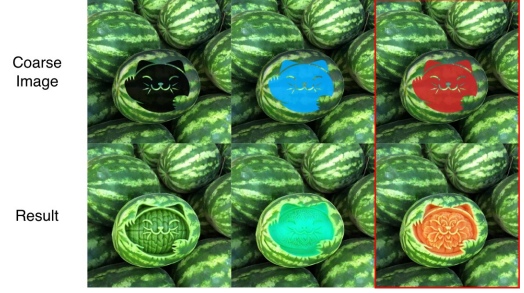


Figure 5: Limitation: Output carved color depends on the coarse image mask.

canvas image to capture its surface grain and ambient lighting. Concurrently, **LoRA B** is trained on a carving style reference image to learn internal cut textures, depth shadows, and edge behaviors.

Depth-Controlled Geometry Guidance. To simulate physical incisions entirely within 2D space, we generate a composite geometric depth representation. We extract a base 2.5D depth map from the target canvas and fuse it with the user’s 2D carving pattern. A scalar depth parameter $\lambda \in [0.0, 1.0]$ scales the intensity of the cut. This composite depth map is fed into a depth-conditioned ControlNet branch, providing explicit conditional boundaries that force the model to generate realistic self-shadowing and surface curvature along the carved edges.

Localized Feature-Injected Diffusion. To preserve background fidelity and prevent harsh boundary artifacts, we introduce a feature-injected denoising pipeline utilizing the Plug-and-Play framework [Tumanyan et al. 2023]. First, we perform deterministic DDIM inversion on the original canvas image to project it into latent noise space. During denoising, intermediate spatial features and self-attention keys from the source inversion path are injected into the UNet decoder layers to anchor background structure. Finally, a soft spatial mask restricts generative edits strictly to the carving coordinates, blending the modified latent values smoothly into the unedited background at each timestep.

3 Results and Conclusion

Figure 3 presents qualitative surface carving results across different canvas materials and design types. The top half demonstrates intaglio carving, where the design itself (a cat pattern) is incised directly into the watermelon surface. The bottom half illustrates relief carving, where the background is carved away to leave the design raised against the sheep canvas. In both cases, our approach

faithfully transfers the reference style into the carved regions while accurately maintaining the target object’s background geometry.

As shown in Figure 4, compared to SDEdit and ZeST’s global adapters and basic masking, our approach offers two key advantages: **1) Accurate Local Texture Control:** Our feature injection mechanism places texture details precisely inside the carved region without leaking unwanted style onto the canvas. **2) Smooth Boundary Transitions:** Our latent blending technique seamlessly joins the carved interior to the surrounding material, eliminating harsh edges and preserving natural background shading.

In summary, our framework offers a flexible solution for localized 2D surface carving by combining depth guidance, dual-material adaptation, and latent feature injection. Our approach preserves the surrounding object’s structure and texture while generating detailed carvings inside the target area. However, as shown in Figure 5, our current method relies on the color applied to the masked region in the coarse image, which directly influences the final carved material’s color regardless of the reference texture. In future work, we plan to address this color dependency, optimize generation speed, and extend our framework to handle complex materials.

Acknowledgments

This work is supported by the National Science and Technology Council, Taiwan under Grant 113-2221-E006-161-MY3 and Grant 114-2221-E-006-114-MY3.

References

- Ta-Ying Cheng, Prafull Sharma, Andrew Markham, Niki Trigoni, and Varun Jampani. 2024. ZeST: Zero-Shot Material Transfer from a Single Image. In *Proceedings of the European Conference on Computer Vision (ECCV)*.
- Nisha Huang, Yizhou Lin, Jie Guo, Xiu Li, Tong-Yee Lee, and Zitong Yu. 2026. DealMaTe: Multi-Dimensional Material Transfer via Diffusion Transformer. *ACM Transactions on Graphics* (2026).
- Chenlin Meng, Yutong He, Yang Song, Jiaming Song, Jiajun Wu, Jun-Yan Zhu, and Stefano Ermon. 2022. SDEdit: Guided Image Synthesis and Editing with Stochastic Differential Equations. In *International Conference on Learning Representations*.
- Nataniel Ruiz, Yuanzhen Li, Varun Jampani, Yael Pritch, Michael Rubinstein, and Kfir Aberman. 2023. Dreambooth: Fine tuning text-to-image diffusion models for subject-driven generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.
- Narek Tumanyan, Michal Geyer, Shai Bagon, and Tali Dekel. 2023. Plug-and-Play Diffusion Features for Text-Driven Image-to-Image Translation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 1921–1930.