






Learning 3D Geometry and Feature Consistent Gaussian Splatting for Object Removal -Supplementary Material-

Yuxin Wang¹, Qianyi Wu², Guofeng Zhang³, and Dan Xu¹

¹ Hong Kong University of Science and Technology

² Monash University ³ Zhejiang University

ywangom@cse.ust.hk, qianyi.wu@monash.edu,

zhangguofeng@zju.edu.cn, danxu@cse.ust.hk

In this supplementary material, we provide additional experimental results in Sec. 1. We also provide a video demo for more qualitative state-of-the-art comparisons and ablation studies as discussed in Sec. 2.


1 Additional Experiments

1.1 Comparison with GaussianEditor

In addition to the comparisons in the main text with the NeRF-based method [4, 5, 8], we also compared our approach **GScream** with the GaussianEditor [2], which is a general editing pipeline also based on Gaussian Splatting representation. They utilize 2D prior from the diffusion model to guide the updates of the Hierarchical Gaussian splatting (HGS) in order to achieve stabilized editing. The qualitative comparison of object removal with GaussianEditor is illustrated in Fig. 1. From the Fig. 1, our method can fill the mask region with more realistic and plausible textures.

1.2 Ablations on using different Depth Estimation Models

In our method GScream, we use a depth estimation model to obtain depth maps for each viewpoint independently. In this subsection, we compare two different single image depth estimation methods: Midas [1] and Marigold [3] for obtaining depth prior, which is utilized to supervise our GScream. As shown in Fig. 2, the depth predicted by Midas at the red fence is not particularly continuous, resulting in the texture and depth of the learned GScream at the red fence being less continuous as well. On the other hand, Marigold can predict more continuous depth, thereby guiding a relatively more continuous GScream. This also demonstrates that accurate depth guidance is crucial for learning geometric continuity in the representation of 3D Gaussian Splatting.

 Corresponding author.

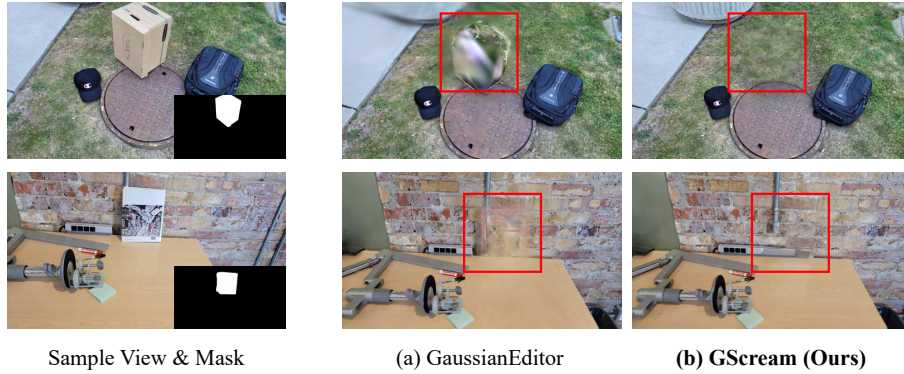


Fig. 1: Qualitative comparison of object removal with the 3D Gaussian Splatting-based method GaussianEditor [2].

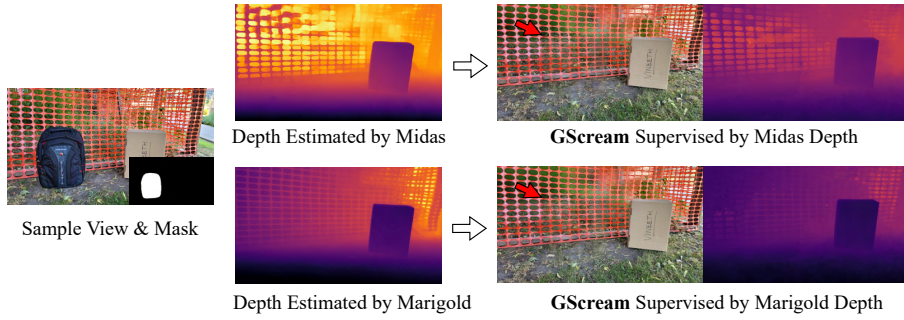


Fig. 2: Qualitative results of using different monocular depth estimation models: Midas [1] and Marigold [3].

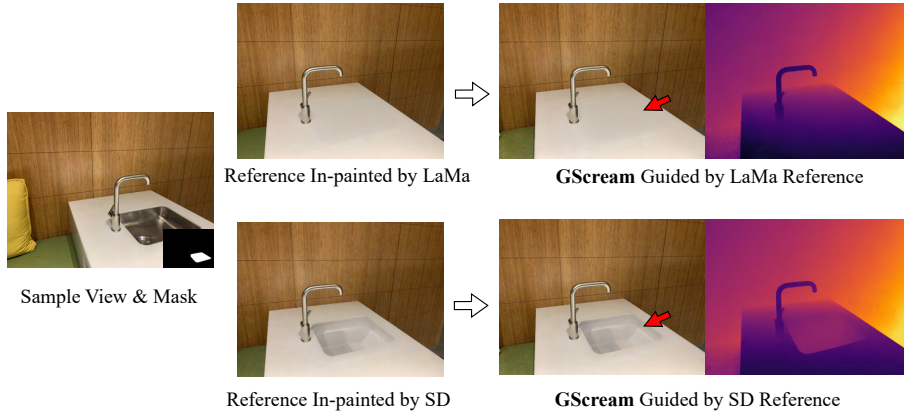


Fig. 3: Qualitative results of using different 2D in-painting models: LaMa [7] and Stable Diffusion (SD) [6]

1.3 Ablations on using different 2D In-Painting Models

In our method, we use a 2D in-painting model to obtain a reference image from a certain viewpoint as guidance. In this subsection, we compare two different methods using different 2D in-painting models: LaMa [7] and Stable Diffusion [6] to obtain guidance for GScream. As shown in Fig. 3, LaMa removes both the metal sink and the indentation, while Stable Diffusion removes the metal sink but retains the indentation. Both results from these two in-painting methods can serve as guidance for obtaining reasonable GScream results. This indicates that the choice of the in-painting method for obtaining the reference is not crucial; what matters more is obtaining a reasonable reference. As long as there is a reasonable reference, GScream can generate 3D geometry and texture continuous results.

1.4 GScream on harder cases

We conducted experiments on two challenging scenes, as shown in Fig. 4: removing a truck and removing a valve disc with large parallax, both with limited training views. Rendered images and depth maps were produced under small and large camera angle variations. With small angle variations, the novel view synthesis results were good, as observed in the tree leaves and brass bolt (highlighted by red arrows). However, with significant angle variations, undesirable outcomes appeared at the image edges, and the regions where objects were removed became blurry, such as the tree trunk and valve protrusion with larger parallax. This is due to the insufficient extrapolation capability of the 3DGS when novel views fall outside the domain of training views, a common issue in implicit representation.

2 Video Demo

We also provide a demo video for more qualitative state-of-the-art comparisons and ablation studies. The video demo can be found from the project page: <https://w-ted.github.io/publications/gstream>

3 Limitation

GScream aims to provide a comprehensive solution for scene editing within the 3DGS framework, specifically targeting the removal of 3D objects. While our quantitative and qualitative experiments underscore the effectiveness and efficiency of our proposed method, it is crucial to acknowledge certain limitations. Firstly, our approach relies on multi-view object masks. Although this necessitates additional pre-processing, we would like to highlight that there are currently strategies to obtain multi-view consistent masks with minimal manual involvement. For instance, one can annotate a 3D mask or a single frame and then utilize video segmentation techniques. These methods largely automate the

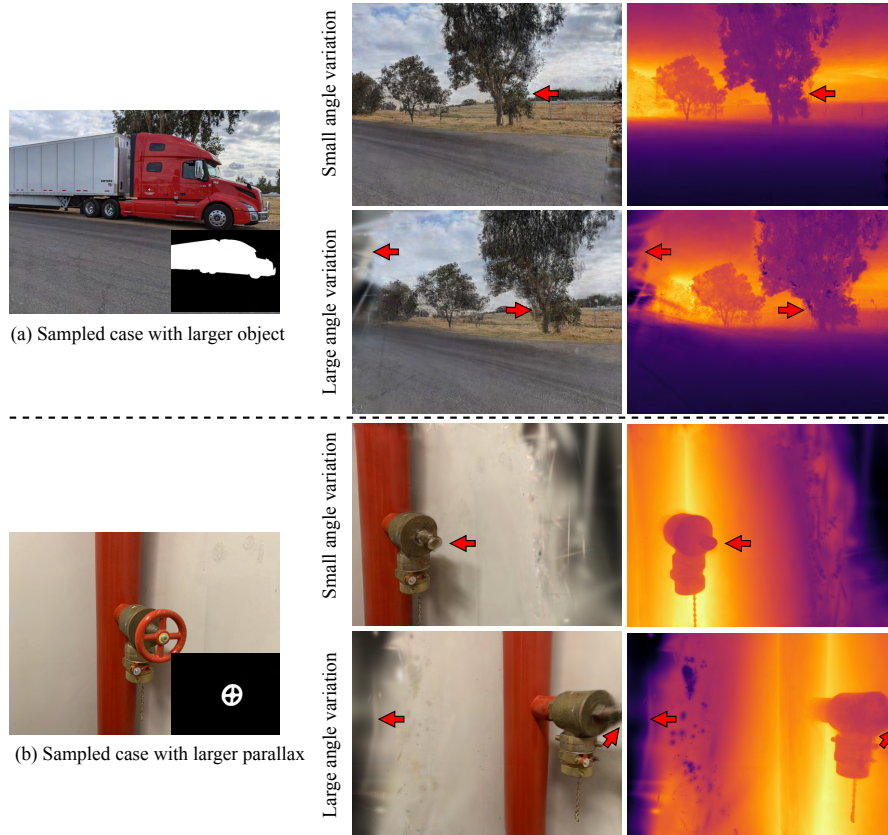


Fig. 4: Qualitative results on harder cases with a larger object and better parallax.

process, significantly reducing the need for manual intervention. Secondly, our method depends on external models for performing the in-painting of the reference view and for estimating depth from various viewpoints, as discussed in Sec. 1.2 and 1.3. These dependencies are presently essential because we believe these external models provide a preliminary pseudo-reference for the geometry and texture of 3DGS in the absence of ground truth. With future advancements in 3DGS representation techniques, it may become feasible to eliminate these dependencies.

References

1. Birkl, R., Wofk, D., Müller, M.: Midas v3.1 – a model zoo for robust monocular relative depth estimation. arXiv preprint arXiv:2307.14460 (2023)
2. Chen, Y., Chen, Z., Zhang, C., Wang, F., Yang, X., Wang, Y., Cai, Z., Yang, L., Liu, H., Lin, G.: Gaussianeditor: Swift and controllable 3d editing with gaussian splatting. In: CVPR (2024)

3. Ke, B., Obukhov, A., Huang, S., Metzger, N., Dautt, R.C., Schindler, K.: Repurposing diffusion-based image generators for monocular depth estimation. In: CVPR (2024)
4. Mirzaei, A., Aumentado-Armstrong, T., Brubaker, M.A., Kelly, J., Levinshtein, A., Derpanis, K.G., Gilitschenski, I.: Reference-guided controllable inpainting of neural radiance fields. In: ICCV (2023)
5. Mirzaei, A., Aumentado-Armstrong, T., Derpanis, K.G., Kelly, J., Brubaker, M.A., Gilitschenski, I., Levinshtein, A.: Spin-nerf: Multiview segmentation and perceptual inpainting with neural radiance fields. In: CVPR (2023)
6. RunwayML: Stable diffusion. <https://huggingface.co/runwayml/stable-diffusion-inpainting> (2021)
7. Suvorov, R., Logacheva, E., Mashikhin, A., Remizova, A., Ashukha, A., Silvestrov, A., Kong, N., Goka, H., Park, K., Lempitsky, V.: Resolution-robust large mask inpainting with fourier convolutions. In: WACV (2022)
8. Yin, Y., Fu, Z., Yang, F., Lin, G.: Or-nerf: Object removing from 3d scenes guided by multiview segmentation with neural radiance fields. arXiv preprint arXiv:2305.10503 (2023)