

# Supplementary Material to “InterGSEdit: Interactive 3D Gaussian Splatting Editing with 3D Geometry-Consistent Attention Prior”

Minghao Wen<sup>1\*</sup> Shengjie Wu<sup>1\*</sup> Kangkan Wang<sup>2†</sup> Dong Liang<sup>1,3†</sup>

<sup>1</sup>MIIT Key Laboratory of Pattern Analysis and Machine Intelligence,  
College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics

<sup>2</sup>The Key Lab of Intelligent Perception and Systems for  
High-Dimensional Information of Ministry of Education,

School of Computer Science and Engineering, Nanjing University of Science and Technology

<sup>3</sup>Shenzhen Research Institute, Nanjing University of Aeronautics and Astronautics

{lanche, wushengjie, liangdong}@nuaa.edu.cn, wangkangkan@njjust.edu.cn

The following materials are provided in this supplementary file:

- User study;
- Compatibility with InstructPix2Pix [1];

## 1. User study

To further validate the effectiveness of 3DGS [6] scene editing method introduced in Sec. 5.2, we conducted a user study based on 20 sets of multi-view editing results. Each set comprises the editing outputs of different methods applied to the same 3D scene under identical text prompts. The outputs within each set consist of multiple views to reflect better the spatial consistency and overall visual quality of the edits. A total of 24 volunteers participated in the study, evaluated with Overall Quality and 3D Consistency. As shown in Fig. 1, 67.71% of users preferred our method for overall quality, while 72.50% favored it for 3D consistency. These results indicate a strong user preference for our approach and further validate its superiority in spatially consistent 3D edits.

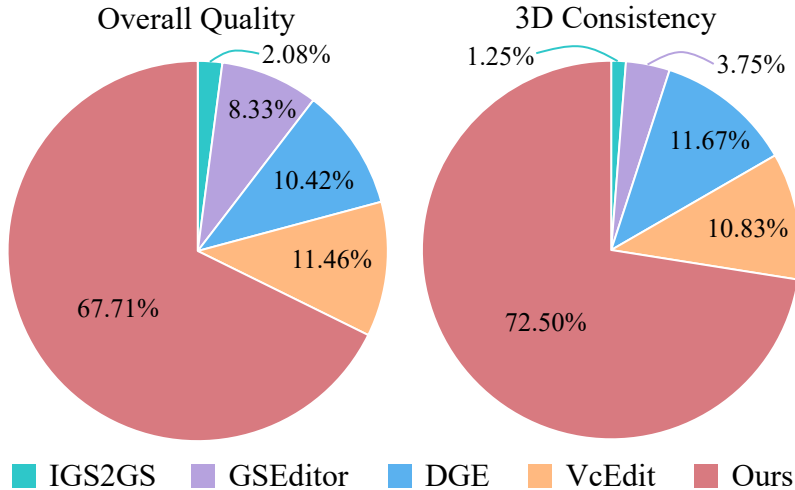


Figure 1. User study results.

\*These authors contributed equally to this work.

†Corresponding authors.

## 2. Compatibility with InstructPix2Pix

Note that IGS2GS [8], GSEditor [4], and DGE [3] adopt the InstructPix2Pix [1] as their editing backbone, while VcEdit [9] employs the InfEdit [10]. To ensure a fair comparison and demonstrate the robustness of our approach across different diffusion frameworks, we conduct experiments using both InstructPix2Pix and InfEdit for our method. We experiment by replacing InfEdit with InstructPix2Pix in our method. Using InstructPix2Pix, the accuracy is 0.2265 in CLIP Similarity [7], 0.1416 in CTIDS [2], and 0.8424 in CDC [5], showing our method outperforms these baselines in the three metrics. Qualitative results in Fig. 2 show the effectiveness of our method on InstructPix2Pix.



Figure 2. Results based on InstructPix2Pix.

## References

- [1] Tim Brooks, Aleksander Holynski, and Alexei A Efros. Instructpix2pix: Learning to follow image editing instructions. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 18392–18402, 2023. 1, 2
- [2] Jun-Kun Chen and Yu-Xiong Wang. Proedit: Simple progression is all you need for high-quality 3d scene editing. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024. 2
- [3] Minghao Chen, Iro Laina, and Andrea Vedaldi. Dge: Direct gaussian 3d editing by consistent multi-view editing. In *European Conference on Computer Vision*, pages 74–92. Springer, 2024. 2
- [4] Yiwen Chen, Zilong Chen, Chi Zhang, Feng Wang, Xiaofeng Yang, Yikai Wang, Zhongang Cai, Lei Yang, Huaping Liu, and Guosheng Lin. Gaussianeditor: Swift and controllable 3d editing with gaussian splatting. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 21476–21485, 2024. 2
- [5] Ayaan Haque, Matthew Tancik, Alexei A Efros, Aleksander Holynski, and Angjoo Kanazawa. Instruct-nerf2nerf: Editing 3d scenes with instructions. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 19740–19750, 2023. 2
- [6] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkuehler, and George Drettakis. 3d gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics (TOG)*, 42(4):1–14, 2023. 1
- [7] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pages 8748–8763. PmLR, 2021. 2

- [8] Cyrus Vachha and Ayaan Haque. Instruct-gs2gs: Editing 3d gaussian splats with instructions, 2024. [2](#)
- [9] Yuxuan Wang, Xuanyu Yi, Zike Wu, Na Zhao, Long Chen, and Hanwang Zhang. View-consistent 3d editing with gaussian splatting. In *European Conference on Computer Vision*, pages 404–420. Springer, 2024. [2](#)
- [10] Sihan Xu, Yidong Huang, Jiayi Pan, Ziqiao Ma, and Joyce Chai. Inversion-free image editing with language-guided diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9452–9461, 2024. [2](#)