

Vinedresser3D: Towards Agentic Text-guided 3D Editing

Supplementary Material



Figure 8. Examples illustrating the importance of our editing-region computation algorithm. We compare our full method with a variant that simply treats all voxels except the preserved parts as the editing region. The latter variant leads to undesired changes near the preserved content (red arrows).

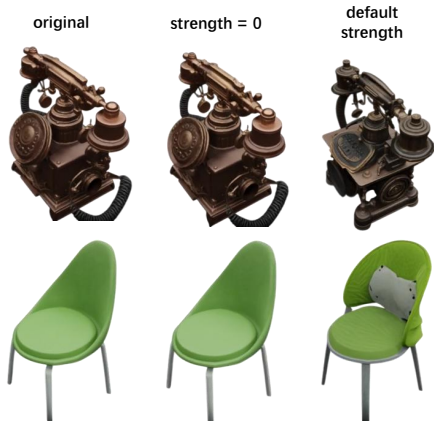


Figure 9. Inversion-reconstruction results obtained with different classifier-free guidance (CFG) strengths. We compare setting the CFG strength to 0 with the default strength (7.5) used in Trellis [59]. Using zero strength yields noticeably more accurate reconstructions.

7. Necessity of the Editing-Region Computation Algorithm

A natural baseline is to treat all voxels except the preserved parts as the editing region ($R_{\text{edit}} = C \setminus P_{\text{pres}}$), instead of using our editing-region computation algorithm. However, this naive strategy allows the model to modify voxels immediately outside the preserved parts and thus alter their appearance. Fig. 8 shows representative examples illustrating this issue.

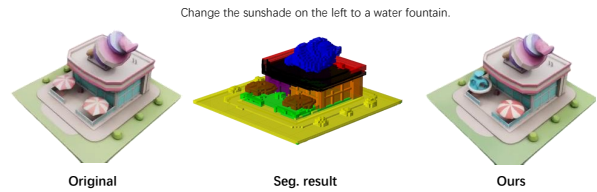


Figure 10. An example demonstrating the robustness of our pipeline.

Module	Failure Rate
Generating Guidance	22.6%
Detecting Editing Region	19.4%
3D Inversion and Generation	58.0%

Table 4. Failure rates of different components in our pipeline.

8. CFG Strength for Inversion

We empirically found through grid search that the inversion accuracy can be greatly improved by setting the classifier-free guidance (CFG) strength [21] to 0, which is critical for our 3D editing module. Compared with using the default strength 7.5 in Trellis [59], this setting yields more accurate inversion-reconstruction results, as shown in Fig. 9.

9. Cascading Errors and Robustness

Although our pipeline contains multiple modules, an early module failure is not unrecoverable. We show a qualitative example in Fig. 10: PartField classifies two sunshades as one semantic part (brown), causing our agent to treat both as the editing region. However, with accurate image guidance from Nano Banana, our agent modifies only the desired sunshade on the left and produces the intended result. We also provide a quantitative failure decomposition in Tab. 4, obtained from 31 failure examples through human inspection. Most failures occurred during 3D inversion and generation.