

A WHY DUAL-VIEWS SETTING IS IMPORTANT

Here, we demonstrate why the dual-view setting is crucial. We compare it with InstantMesh [Xu et al. \(2024\)](#), a popular method for multi-view generation. Figure 1 shows its editing process: given an image from the front view, Zero-123++ [Shi et al. \(2023\)](#) first generates six fixed viewpoints. Afterward, the back view is edited, and the edited images are used for reconstruction. In contrast, our pipeline inputs a front image, uses Zero-123 [Liu et al. \(2023\)](#) to generate the back view, edits the back view, and then performs reconstruction. We observe that InstantMesh has high demands for viewpoint consistency, leading to poor 3D meshes after editing and reconstruction. In contrast, Tailor3D does not encounter this issue.

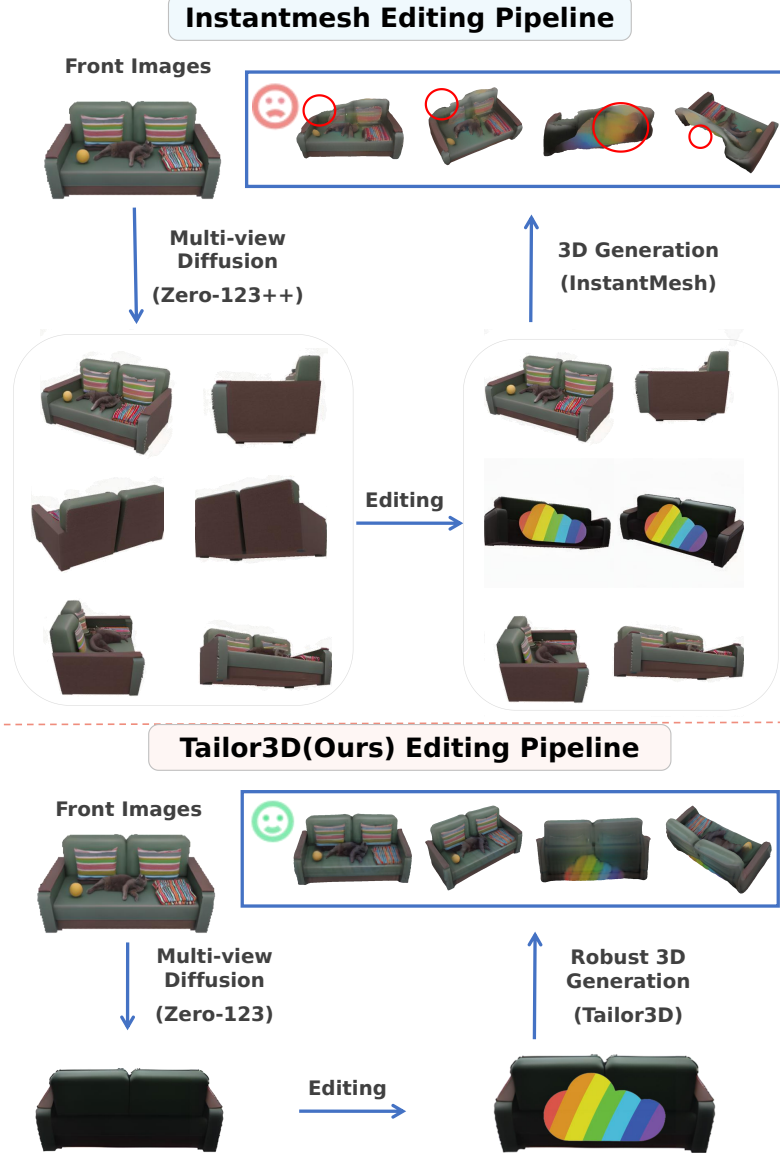


Figure 1: Editing Pipeline of InstantMesh and Tailor3D(Ours).

B COMPARE TO INSTANTMESH

We show more examples in Figure 2, such as Captain America and our small bird. It should be noted that none of our examples are cherry-picked; all the examples are from the original paper. It is clear that using the InstantMesh framework Xu et al. (2024), which has high consistency requirements for viewpoints, leads to inconsistencies in the edited images, and the reconstructed 3D objects fail. In our example, the front and back views of the bird have different styles. However, InstantMesh not only fails to perform style fusion, but also introduces defects and errors in the shape of the bird itself.

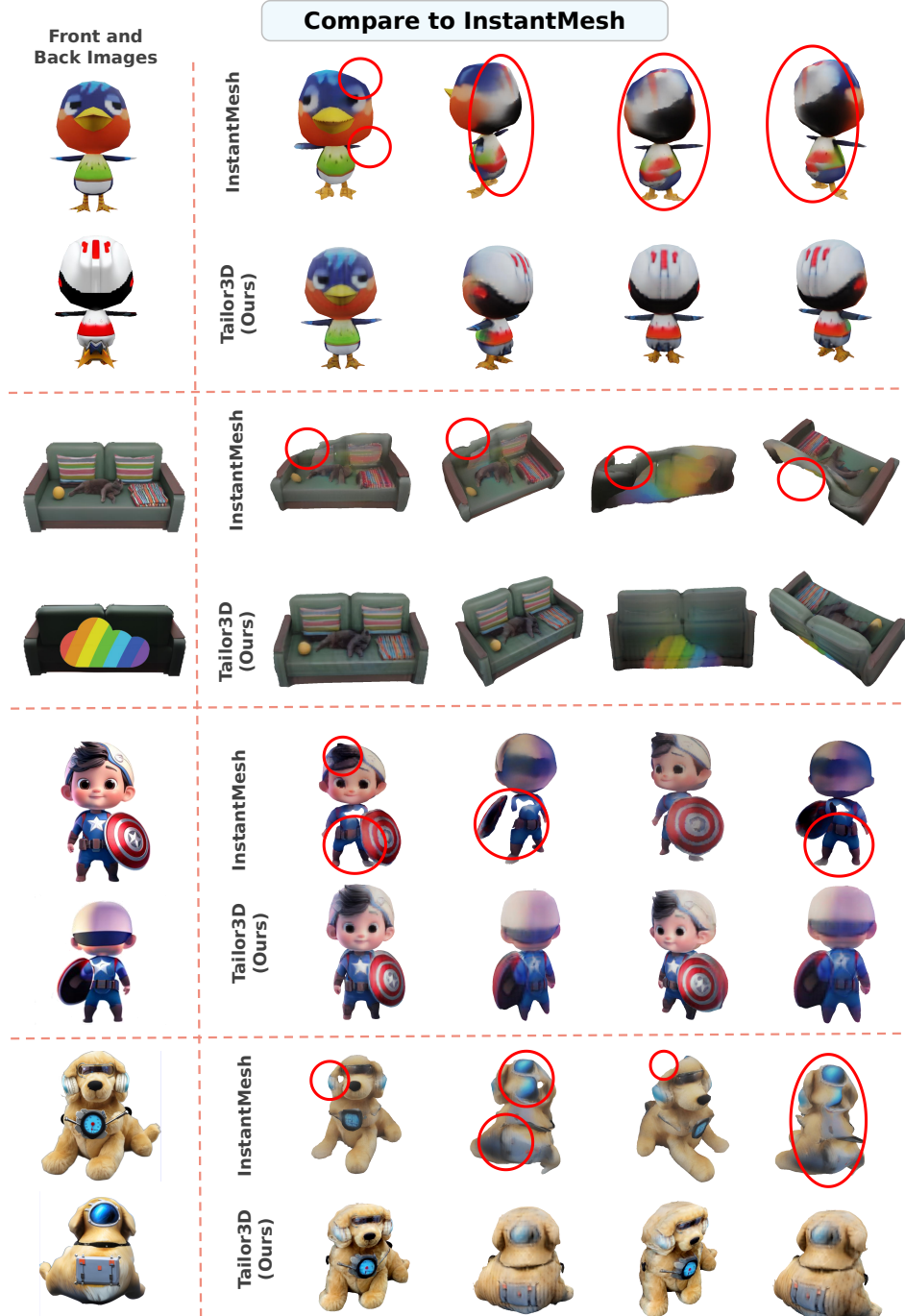


Figure 2: Compared to InstantMesh.

REFERENCES

- Ruoshi Liu, Rundi Wu, Basile Van Hoorick, Pavel Tokmakov, Sergey Zakharov, and Carl Vondrick. Zero-1-to-3: Zero-shot one image to 3d object. In *ICCV*, 2023. [1](#)
- Ruoxi Shi, Hansheng Chen, Zhuoyang Zhang, Minghua Liu, Chao Xu, Xinyue Wei, Linghao Chen, Chong Zeng, and Hao Su. Zero123++: a single image to consistent multi-view diffusion base model. *arxiv: 2310.15110*, 2023. [1](#)
- Jiale Xu, Weihao Cheng, Yiming Gao, Xintao Wang, Shenghua Gao, and Ying Shan. Instantmesh: Efficient 3d mesh generation from a single image with sparse-view large reconstruction models. *arXiv:2404.07191*, 2024. [1](#), [2](#)