Supplementary: Instructive3D: Editing Large Reconstruction Models with Text Instructions

Organization of Appendix

Organization of Appendix

A. Introduction

We present additional results and other details related to our proposed method : Instructive3D. We present implementation details in Appendix B. We present additional experimental results in Appendix C.

B. Implementation Details

For Tri-VAE, we use 3 DownEncoderBlock2D for the encoder and 3 UpDecoderBlock2D for the decoder, with 3 layers per block. The number of in channels and out channels is 40 for each VAE and the number of channels in latent space is 4 per plane of the triplane. The sample size used is 64.

For Latent TriPlane Diffusion model, we use 3 CrossAttnDownBlock2D along with 1 DownBlock2D for the encoder part of the model and 3 CrossAttnUpBlock2D along with 1 UpBlock2D for the decoder part, also we use 1 UNetMidBlock2DCrossAttn in the middle, the text embedding obtained from the CLIP [3] transformer is fed to all the cross attention blocks. We use 2 layers per block, the number of in channels is 24 and out channels is 12, with a sample size of 16. In the background the UNetMidBlock2DCrossAttn, CrossAttnDownBlock2D and CrossAttnUpBlock2D uses BasicTransformerBlock2D, in which both self attention and cross attention is enabled.

C. Experimental Results

We compare our method with Text2Mesh [2], Paint3D [5] and TEXTure [4], which takes a 3D mesh and a text prompt as input and generates an output mesh with the given text conditioning. We provide the mesh generated by Real3D [1] to these models with an edit prompt and compare the output with our generated mesh. We show additional results in Fig. 7- 33. These results show that our method preserves geometry and performs edits consistent with the input edit prompts.

References

[1] Hanwen Jiang, Qixing Huang, and Georgios Pavlakos. Real3d: Scaling up large reconstruction models with real-world images. *arXiv preprint arXiv:2406.08479*, 2024. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28

[2] Oscar Michel, Roi Bar-On, Richard Liu, Sagie Benaim, and Rana Hanocka. Text2mesh: Text-driven neural stylization for meshes. *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 13482–13492, 2021. 1

[3] 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. 1

[4] Elad Richardson, Gal Metzer, Yuval Alaluf, Raja Giryes, and Daniel Cohen-Or. Texture: Text-guided texturing of 3d shapes, 2023. 1

[5] Xianfang Zeng, Xin Chen, Zhongqi Qi, Wen Liu, Zibo Zhao, Zhibin Wang, Bin Fu, Yong Liu, and Gang Yu. Paint3d: Paint anything 3d with lighting-less texture diffusion models, 2023. 1

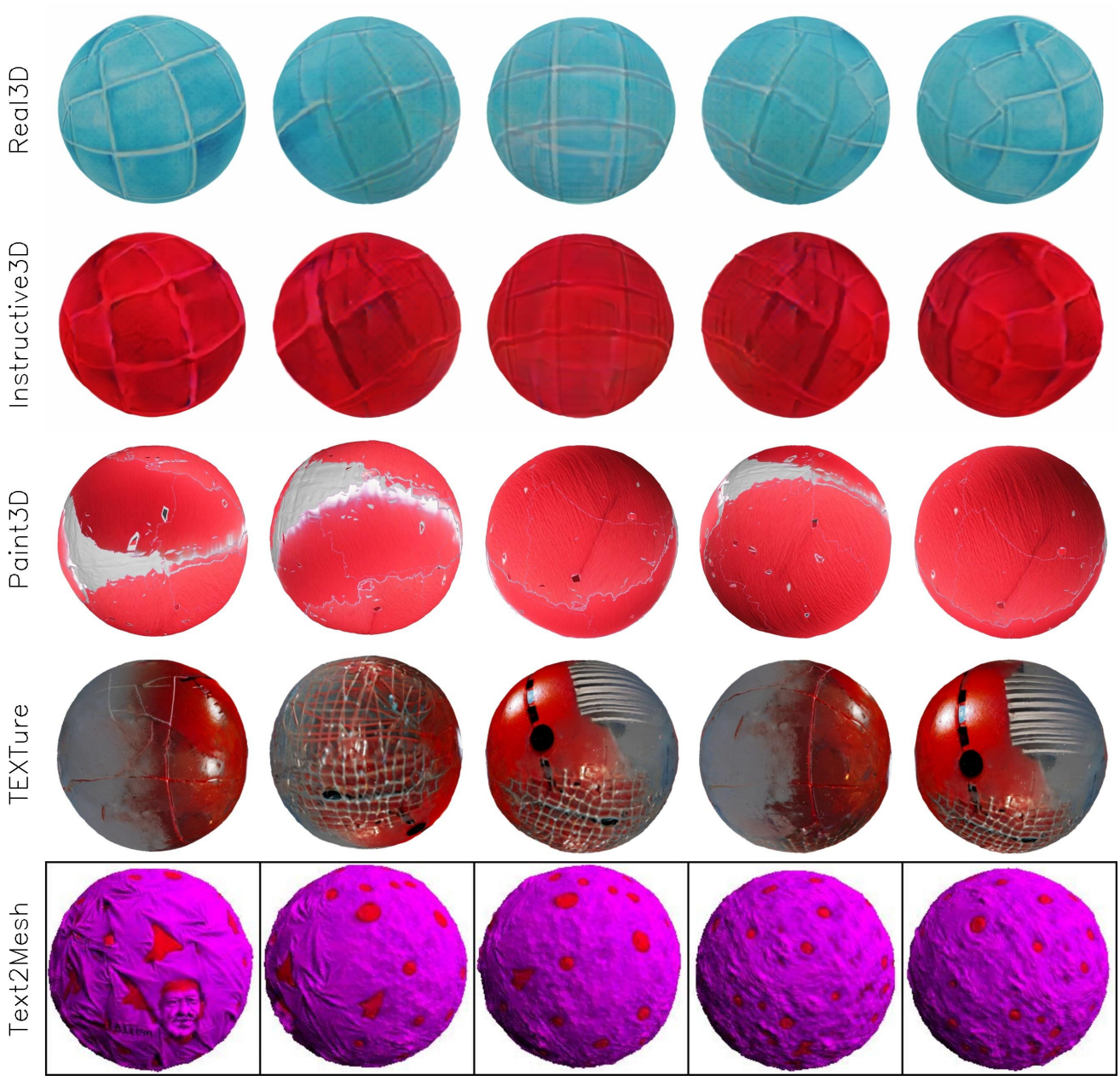

Figure 7. **Baseline comparison results.** Top row shows the rendered images from the mesh obtained from Real3D [1]. Second row shows results from our method. Caption used for editing is: *"change color to red"*.

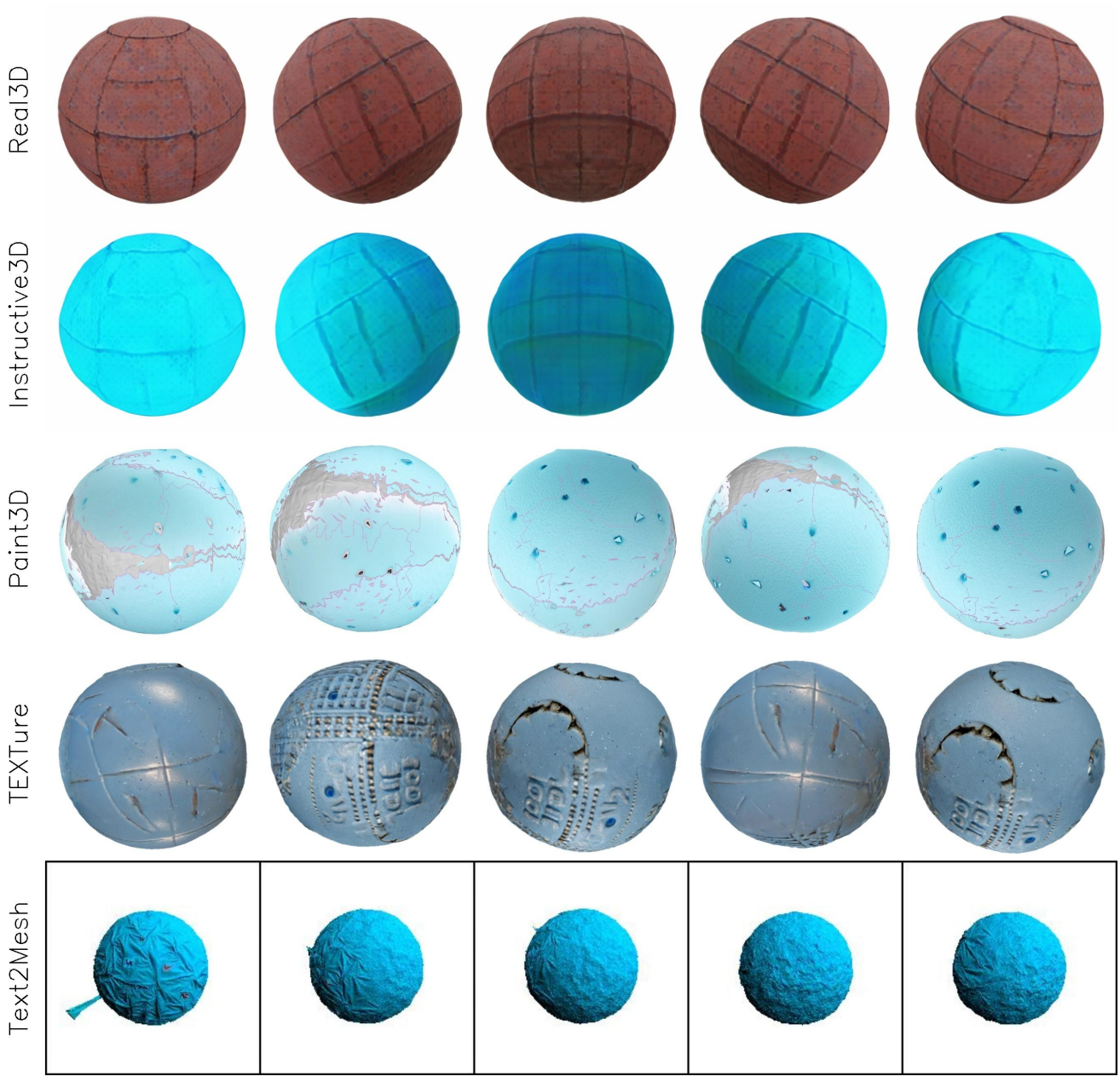

Figure 8. **Baseline comparison results.** Top row shows the rendered images from the mesh obtained from Real3D [1]. Second row shows results from our method. Caption used for editing is: *'change color to powder blue"*.

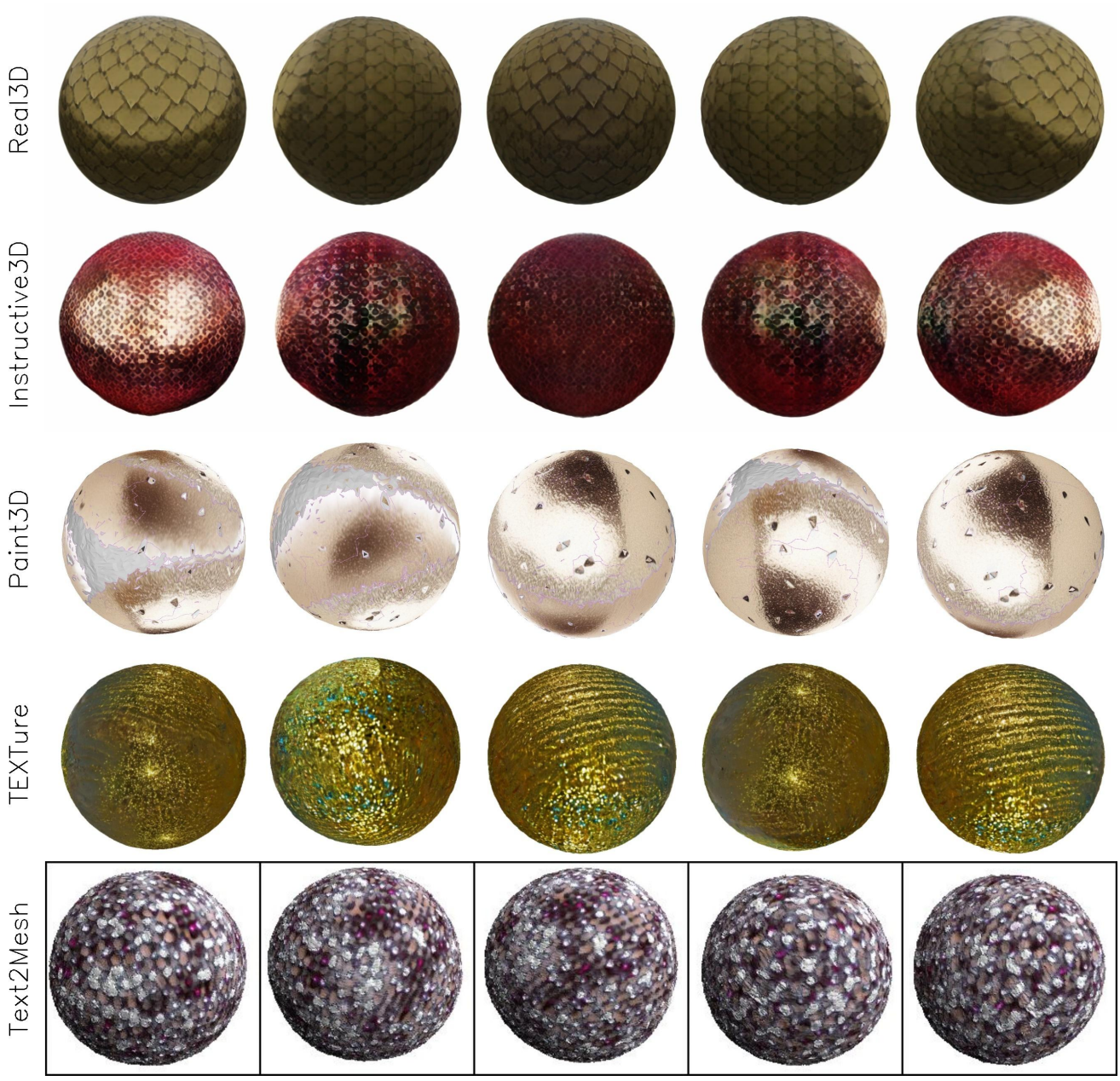

Figure 9. **Baseline comparison results.** Top row shows the rendered images from the mesh obtained from Real3D [1]. Second row shows results from our method. Caption used for editing is: *'add a glittery look to the ball"*.

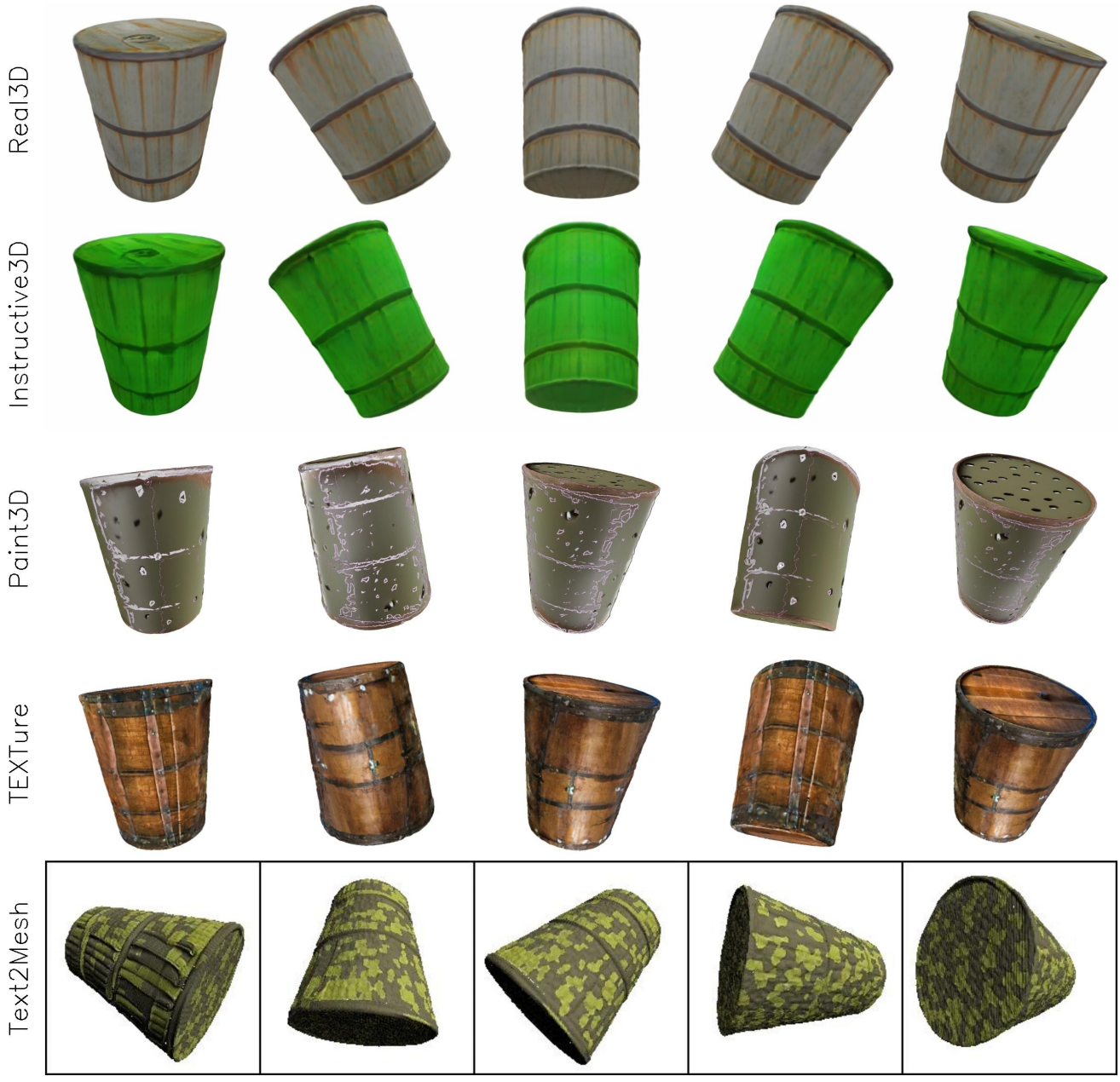

Figure 10. **Baseline comparison results.** Top row shows the rendered images from the mesh obtained from Real3D [1]. Second row shows results from our method. Caption used for editing is: *'change color of barrel to bamboo green"*.

Figure 11. **Baseline comparison results.** Top row shows the rendered images from the mesh obtained from Real3D [1]. Second row shows results from our method. Caption used for editing is: *'change color of barrel to cream"*.

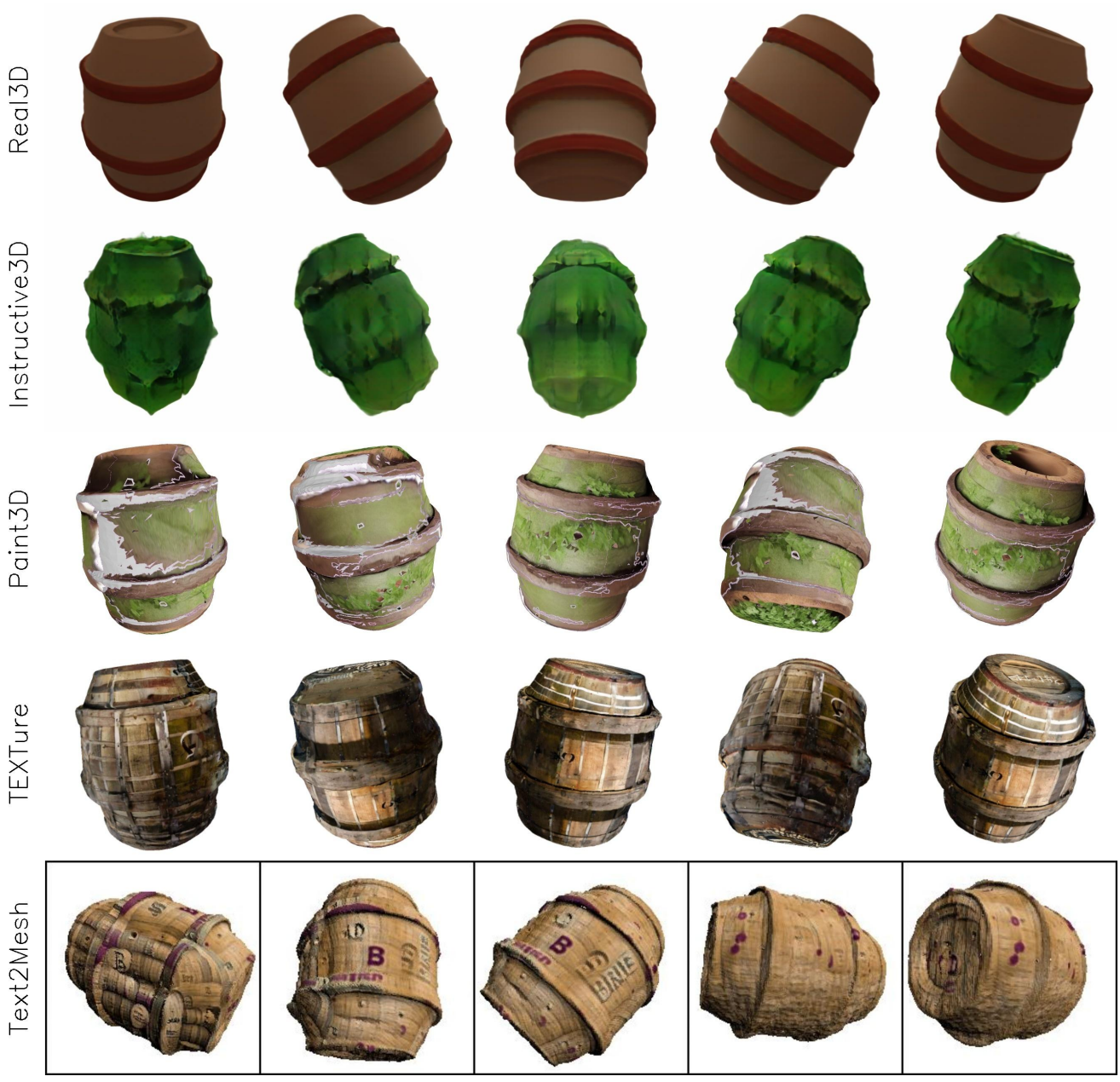

Figure 12. **Baseline comparison results.** Top row shows the rendered images from the mesh obtained from Real3D [1]. Second row shows results from our method. Caption used for editing is: *'apply leaves on the barrel"*.

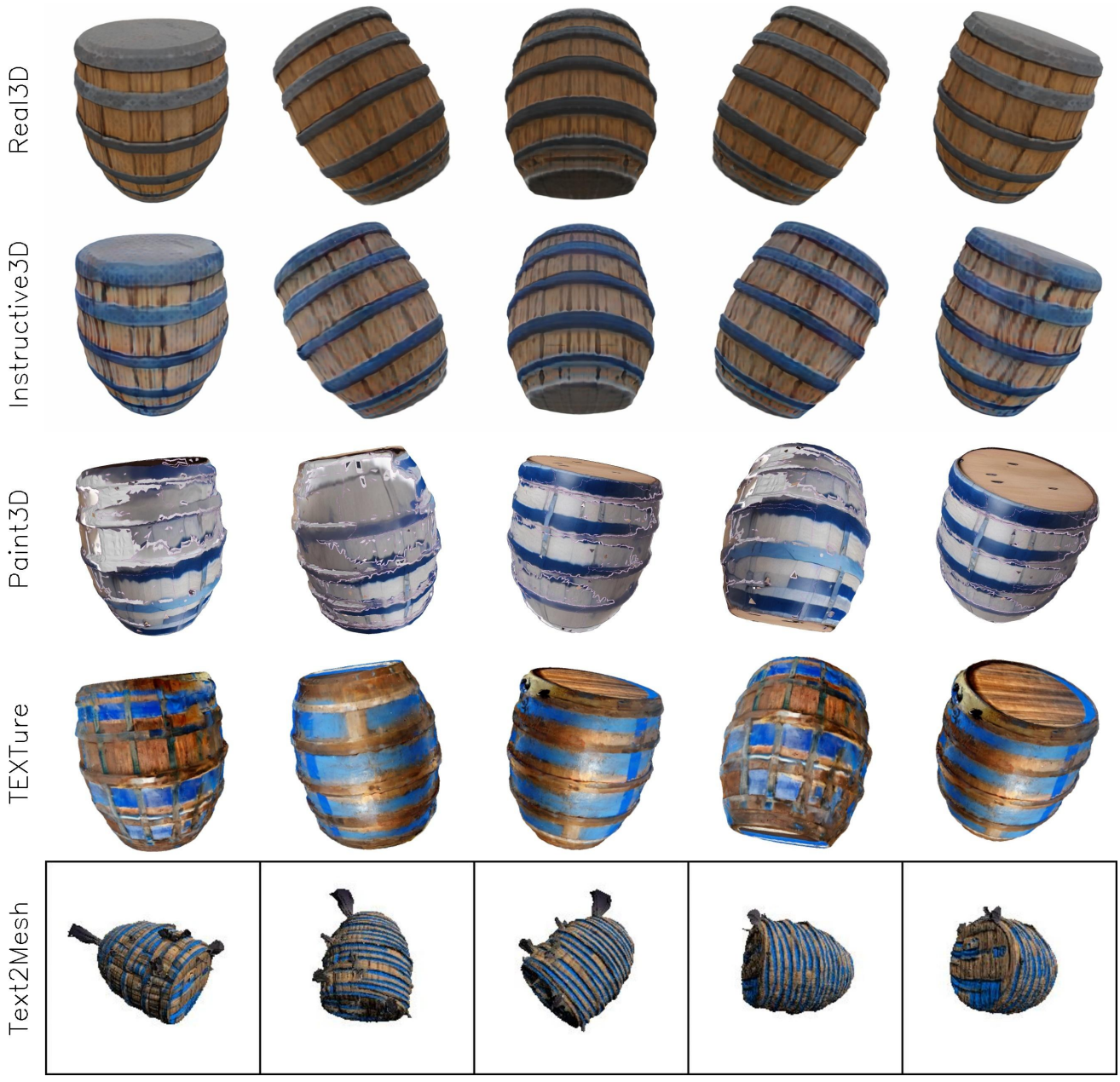

Figure 13. **Baseline comparison results.** Top row shows the rendered images from the mesh obtained from Real3D [1]. Second row shows results from our method. Caption used for editing is: *'add blue stripes to the barrel"*.

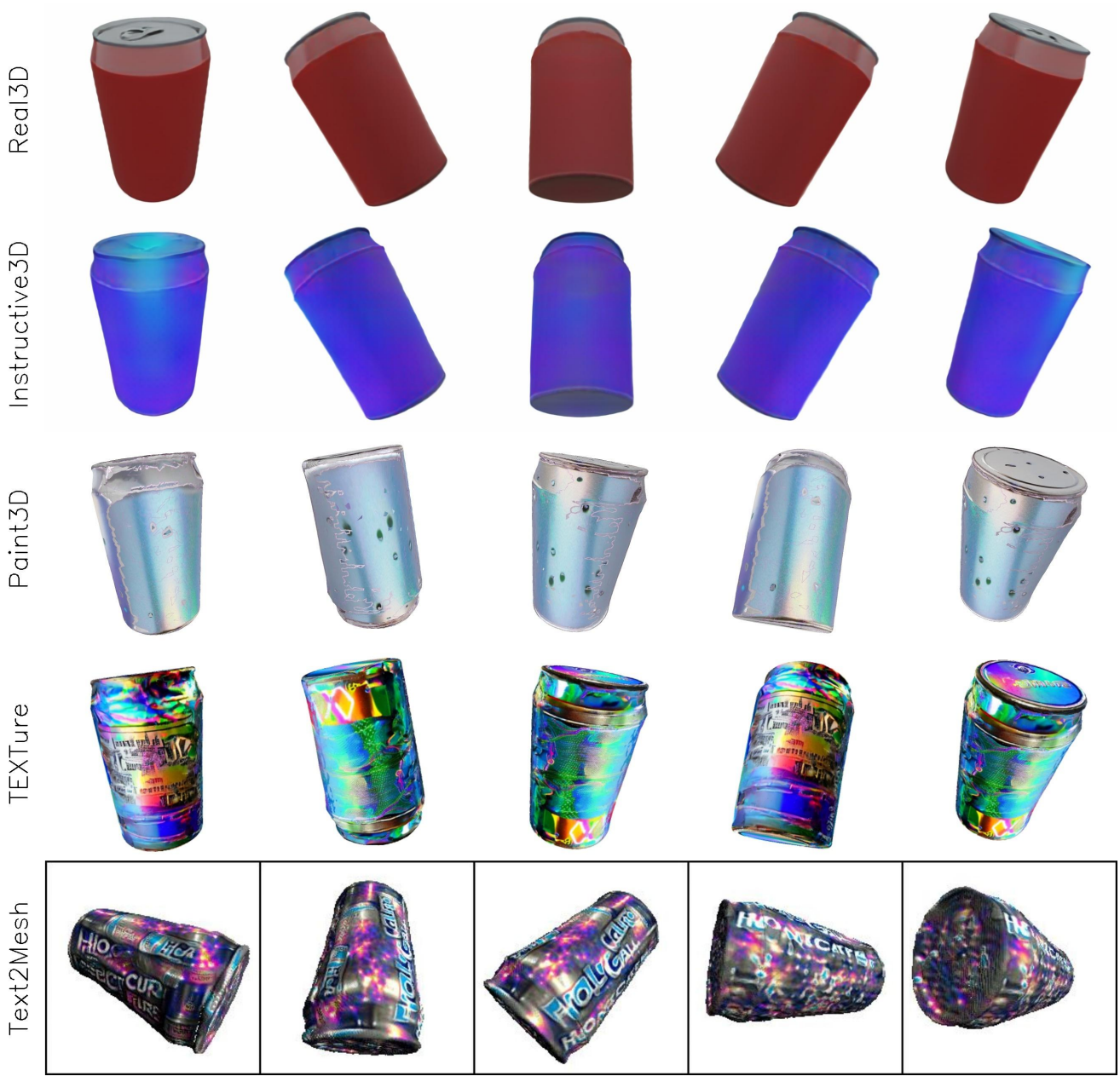

Figure 14. **Baseline comparison results.** Top row shows the rendered images from the mesh obtained from Real3D [1]. Second row shows results from our method. Caption used for editing is: *'apply a purple gradient color to can"*.

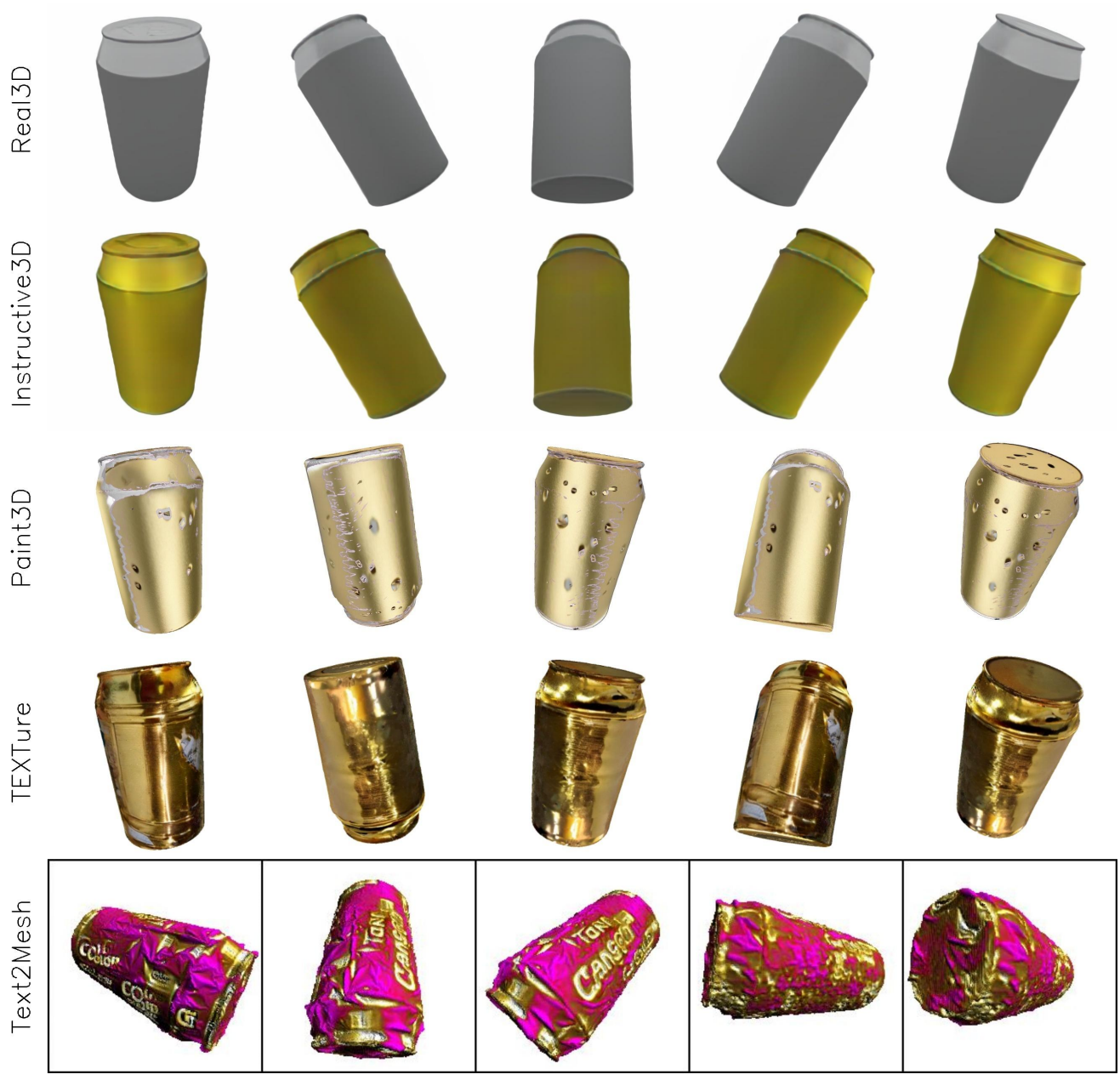

Figure 15. **Baseline comparison results.** Top row shows the rendered images from the mesh obtained from Real3D [1]. Second row shows results from our method. Caption used for editing is: *'change color of can to gold"*.

Figure 16. **Baseline comparison results.** Top row shows the rendered images from the mesh obtained from Real3D [1]. Second row shows results from our method. Caption used for editing is: *'add a marble effect to the can"*.

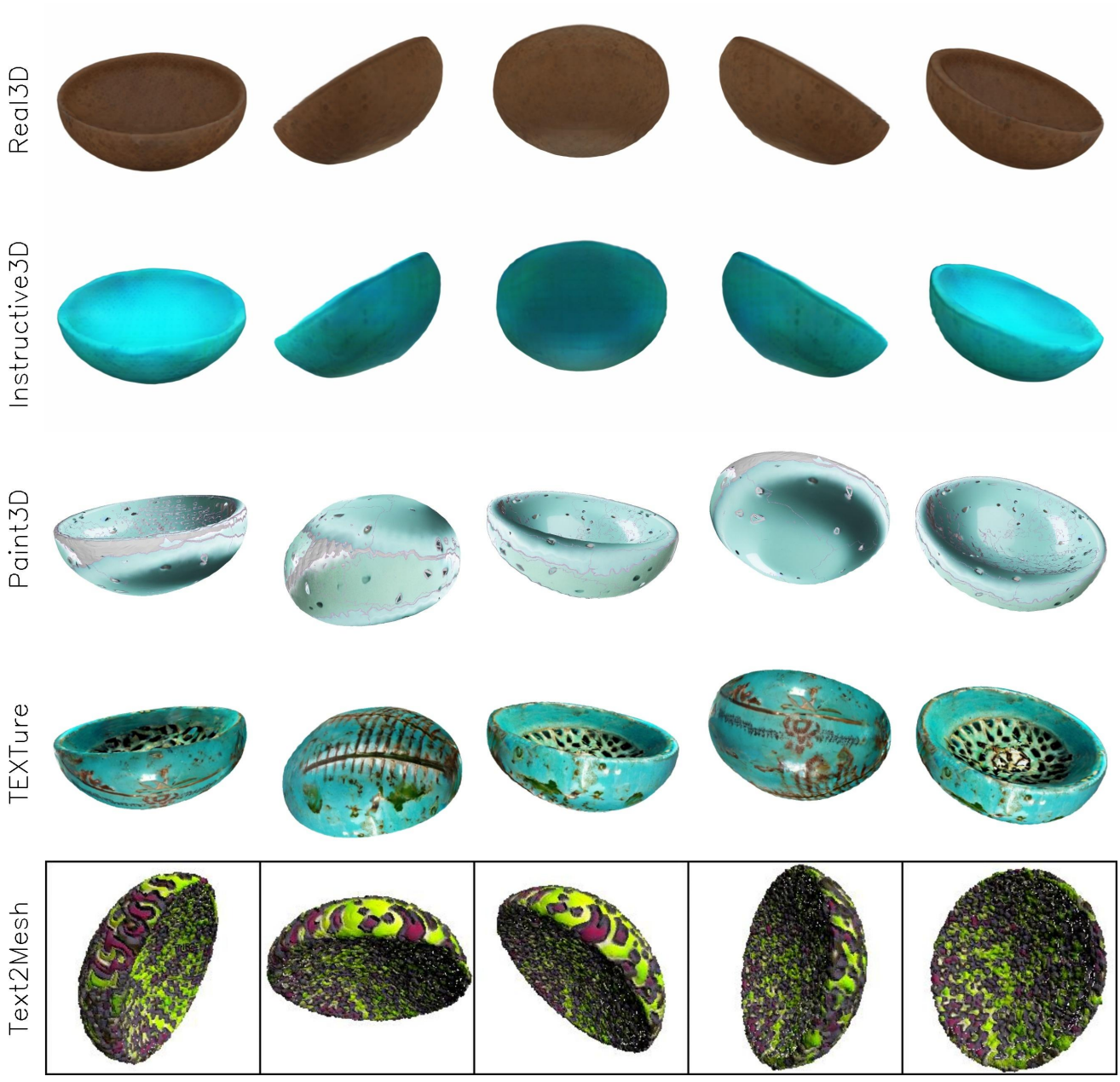

Figure 17. **Baseline comparison results.** Top row shows the rendered images from the mesh obtained from Real3D [1]. Second row shows results from our method. Caption used for editing is: *'change color of bowl to turqoise"*.

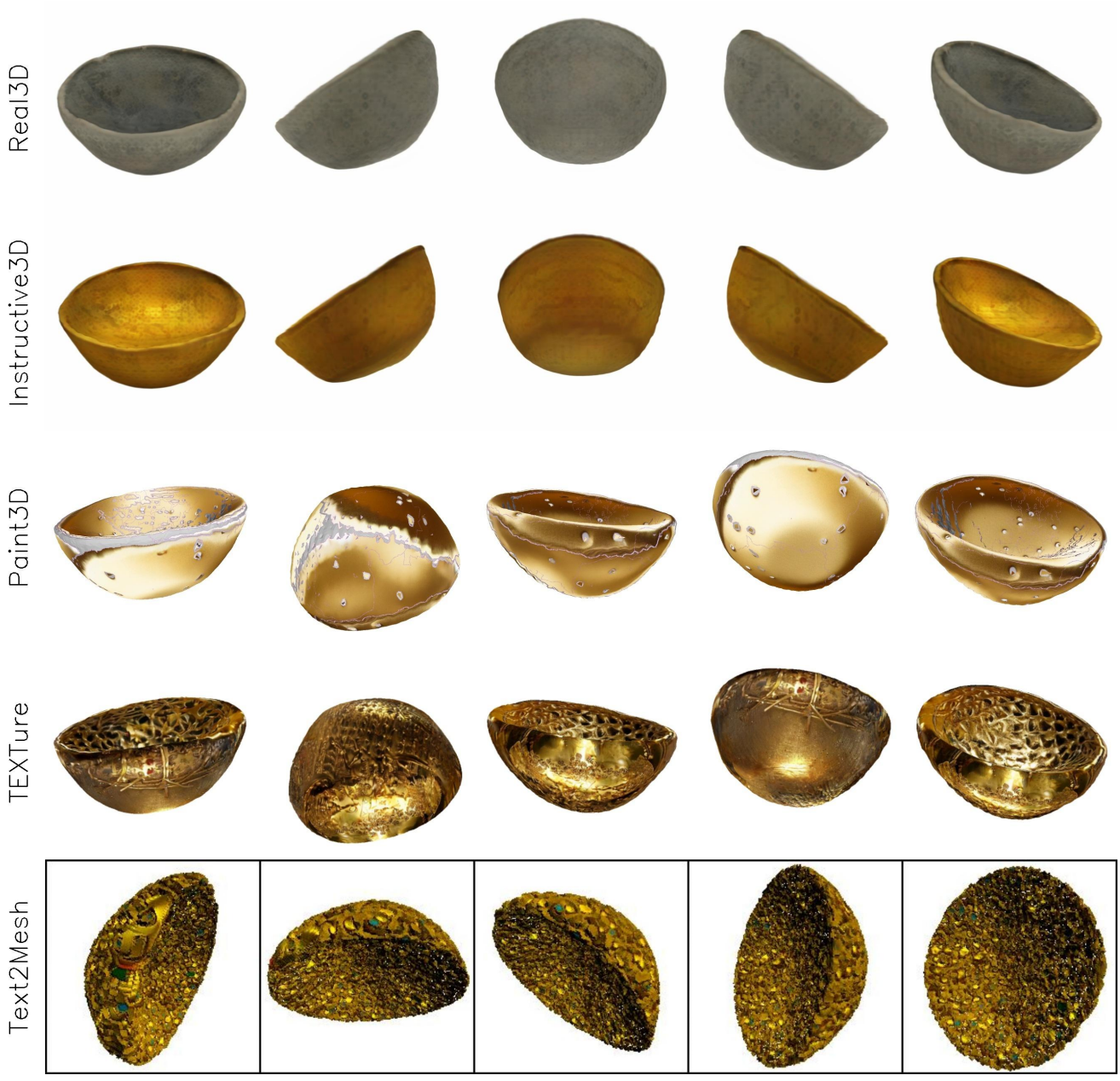

Figure 18. **Baseline comparison results.** Top row shows the rendered images from the mesh obtained from Real3D [1]. Second row shows results from our method. Caption used for editing is: *'change color of bowl to gold"*.

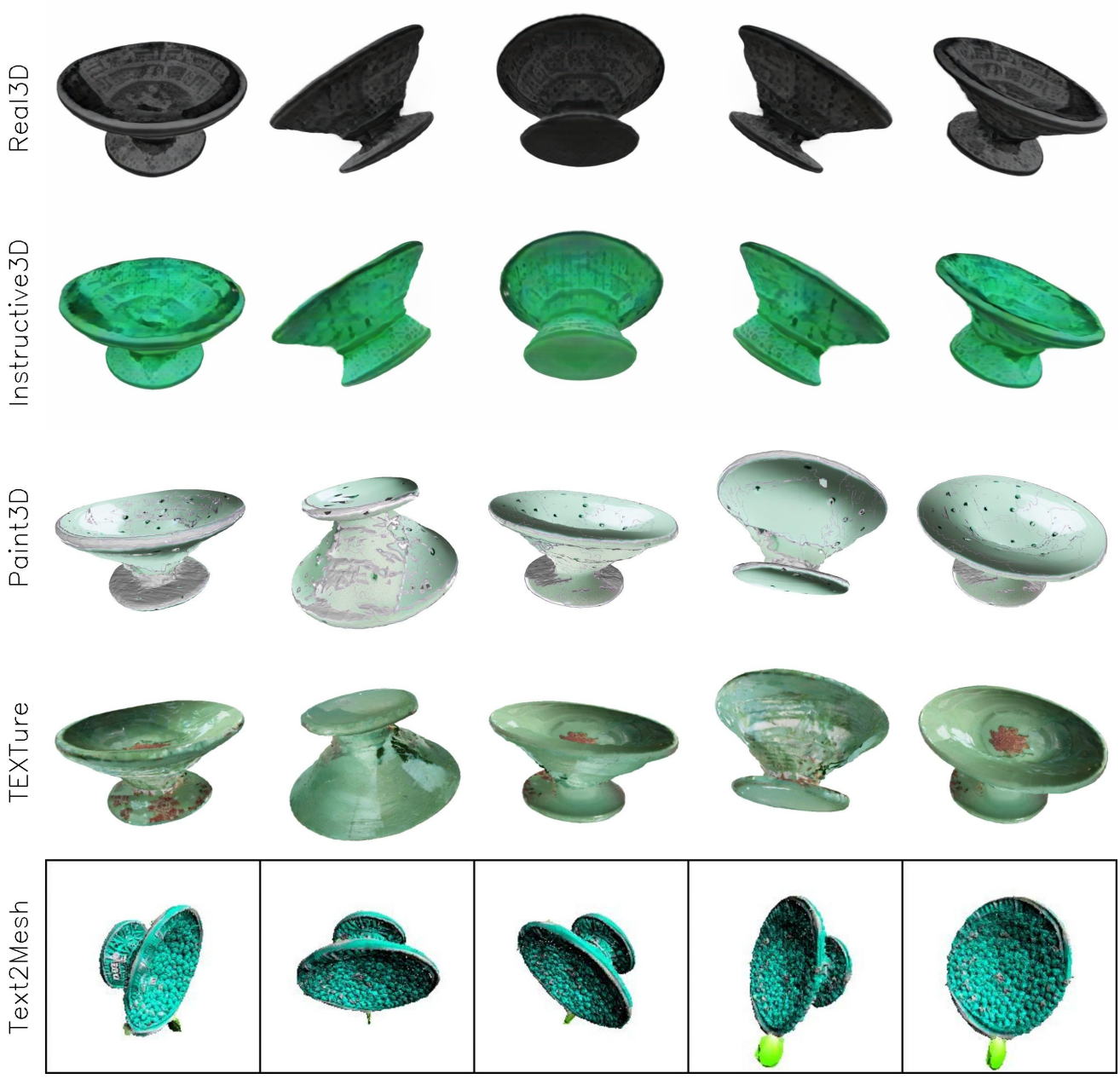

Figure 19. **Baseline comparison results.** Top row shows the rendered images from the mesh obtained from Real3D [1]. Second row shows results from our method. Caption used for editing is: *'change color of bowl to mint green"*.

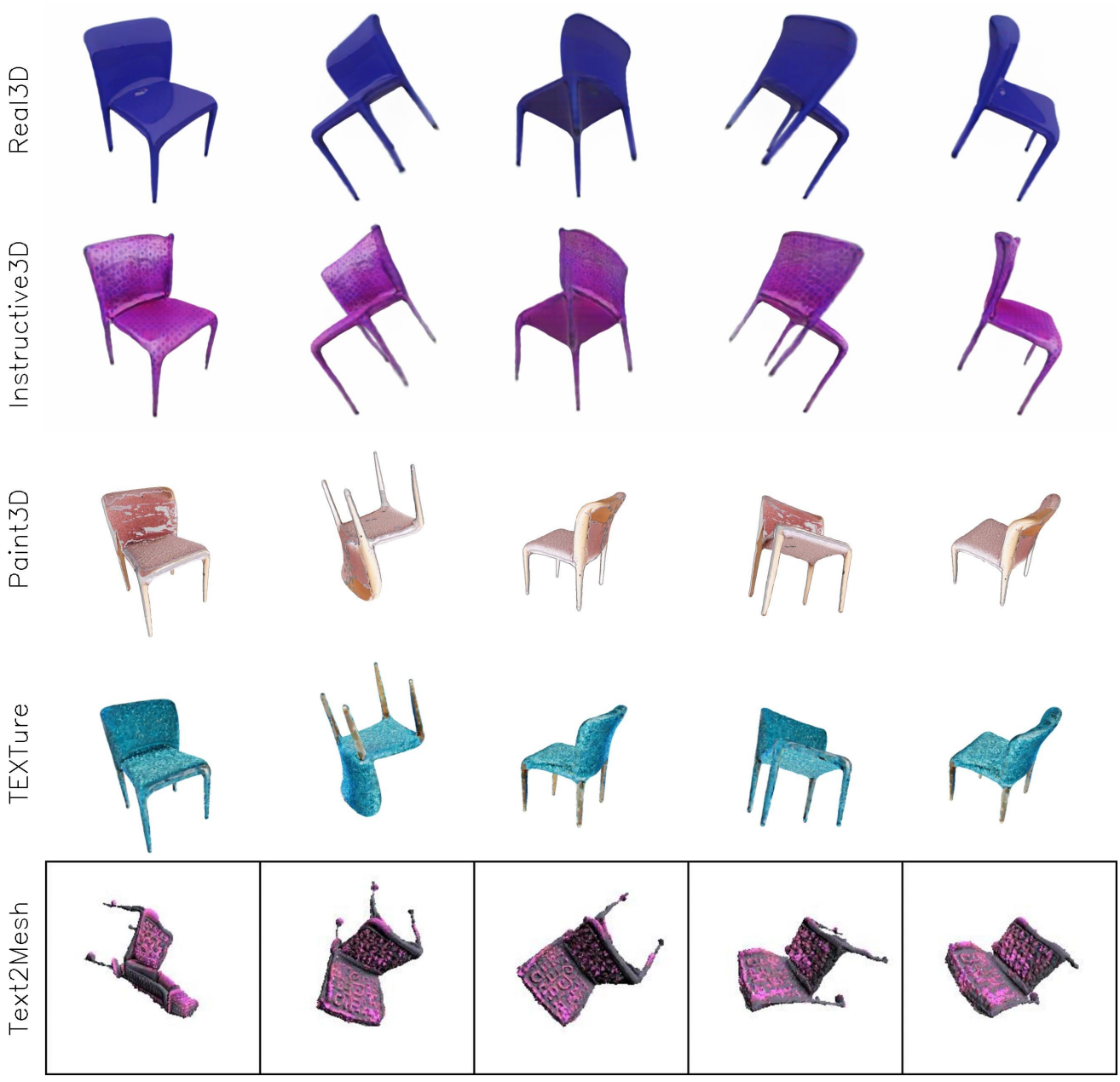

Figure 20. **Baseline comparison results.** Top row shows the rendered images from the mesh obtained from Real3D [1]. Second row shows results from our method. Caption used for editing is: *"add a purple glittery look to chair"*.

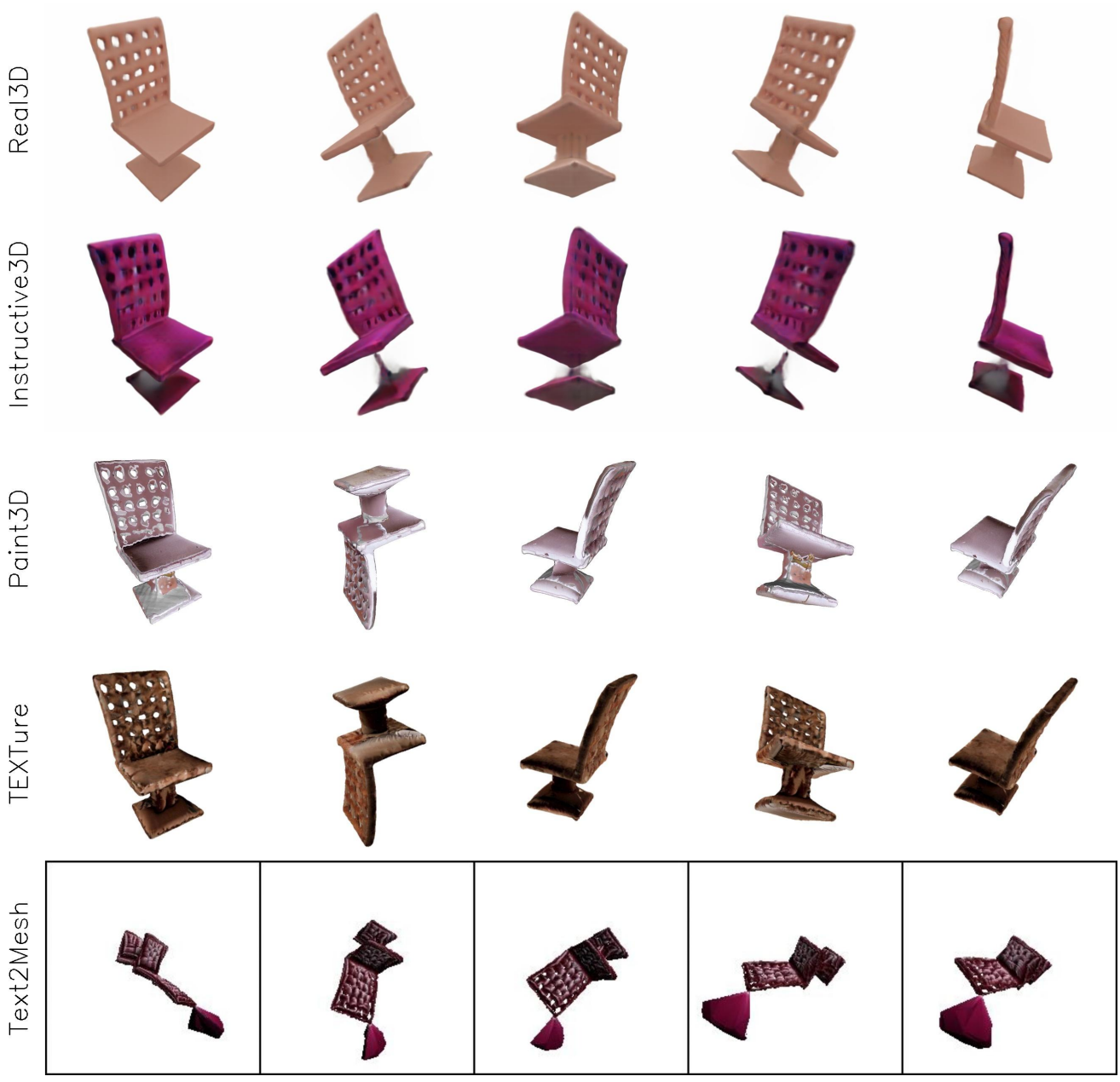

Figure 21. **Baseline comparison results.** Top row shows the rendered images from the mesh obtained from Real3D [1]. Second row shows results from our method. Caption used for editing is: *"add a velvet texture to the chair"*.

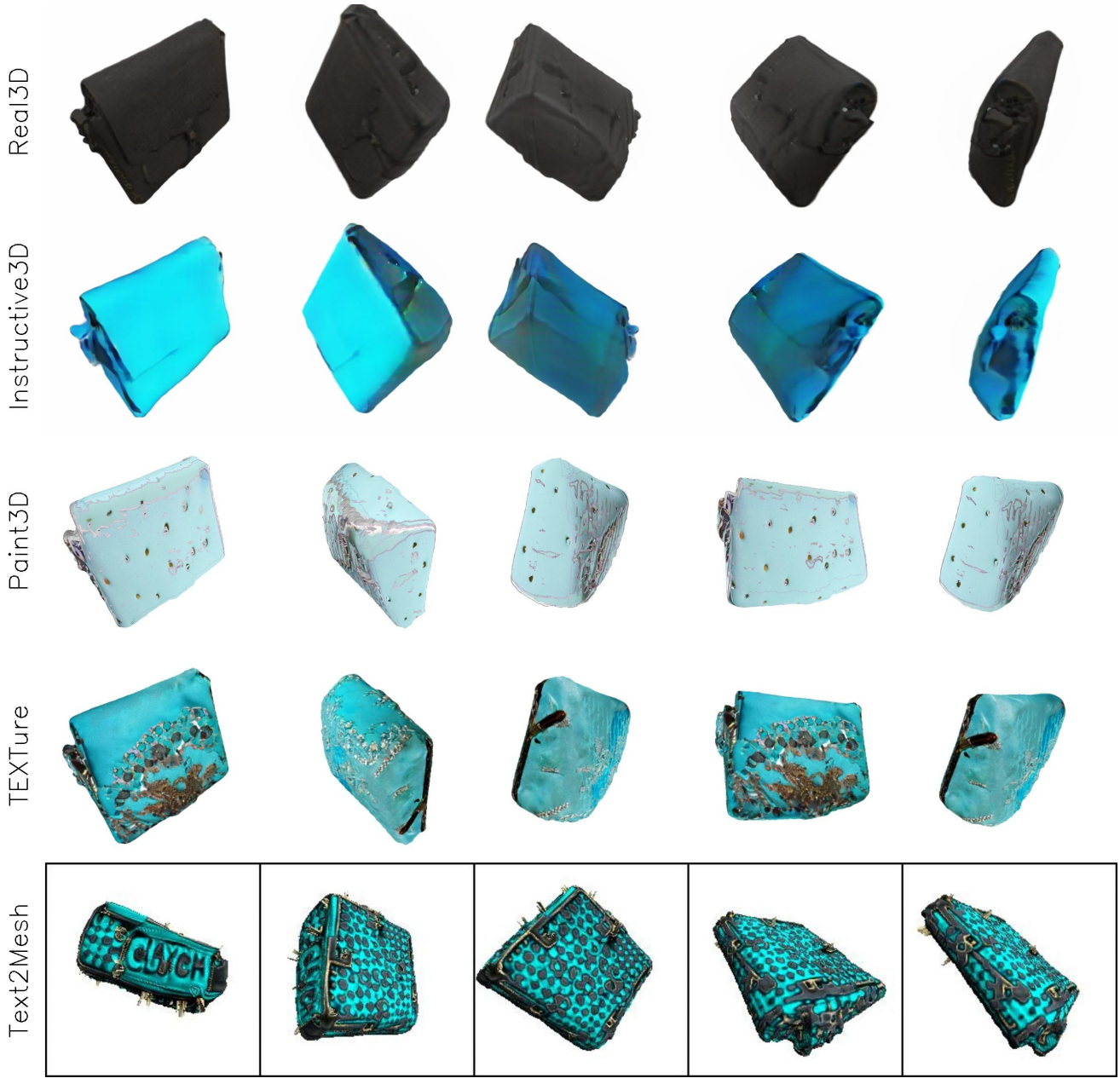

Figure 22. **Baseline comparison results.** Top row shows the rendered images from the mesh obtained from Real3D [1]. Second row shows results from our method. Caption used for editing is: *"change color of clutch bag to cyan"*.

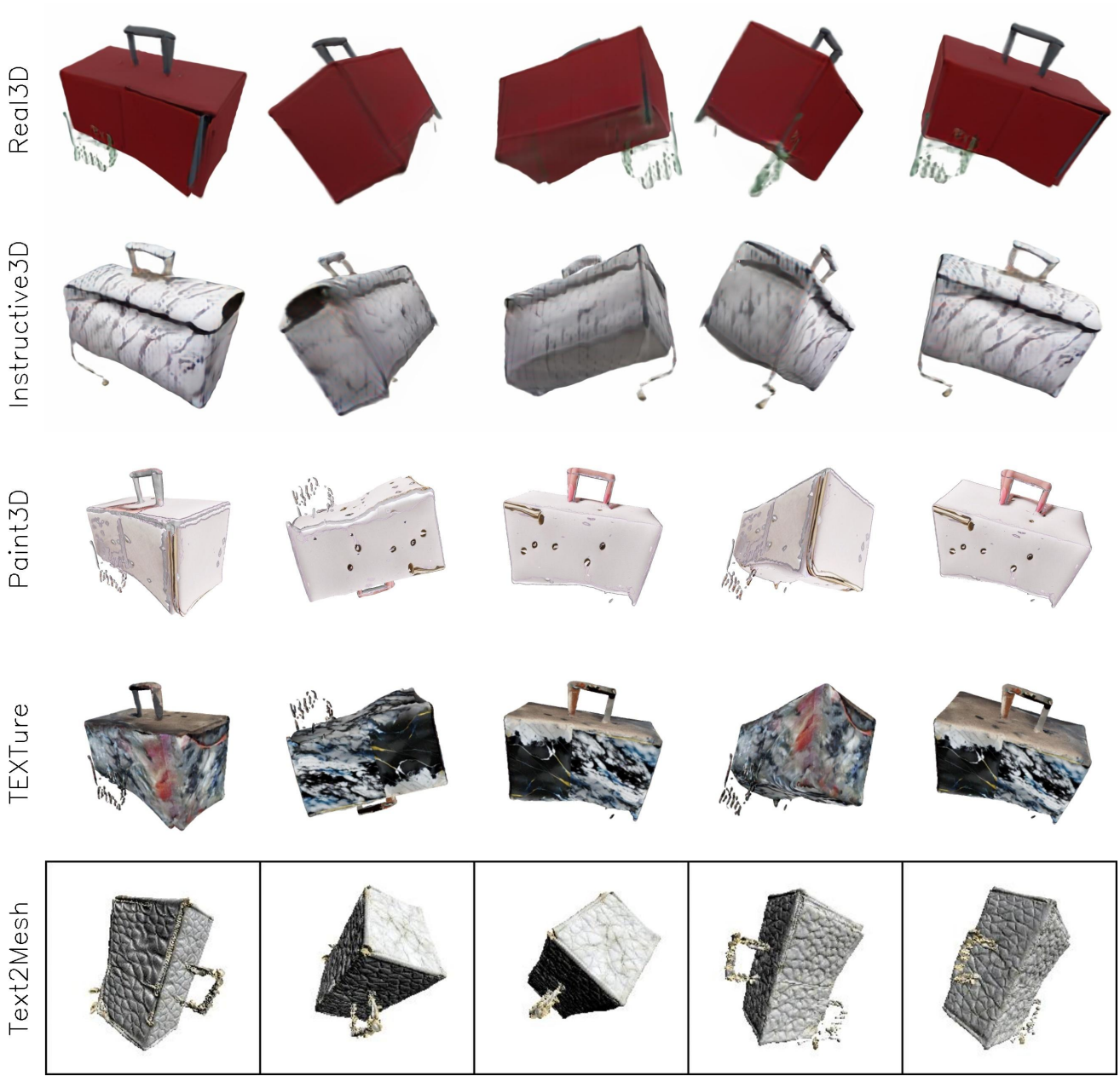

Figure 23. **Baseline comparison results.** Top row shows the rendered images from the mesh obtained from Real3D [1]. Second row shows results from our method. Caption used for editing is: *"apply marble texture to the clutch bag"*.

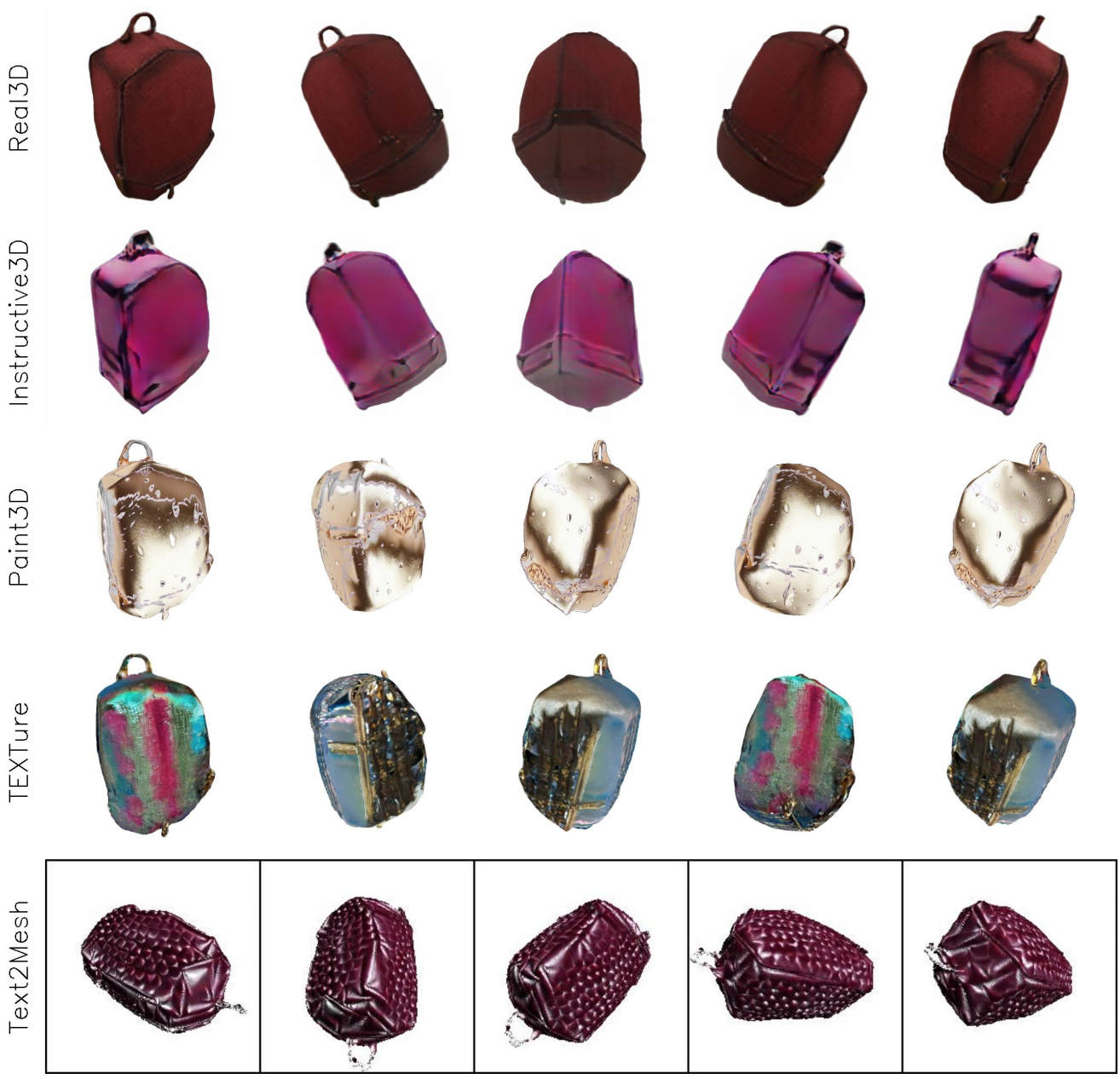

Figure 24. **Baseline comparison results.** Top row shows the rendered images from the mesh obtained from Real3D [1]. Second row shows results from our method. Caption used for editing is: *"add a glossy texture to the clutch bag"*.

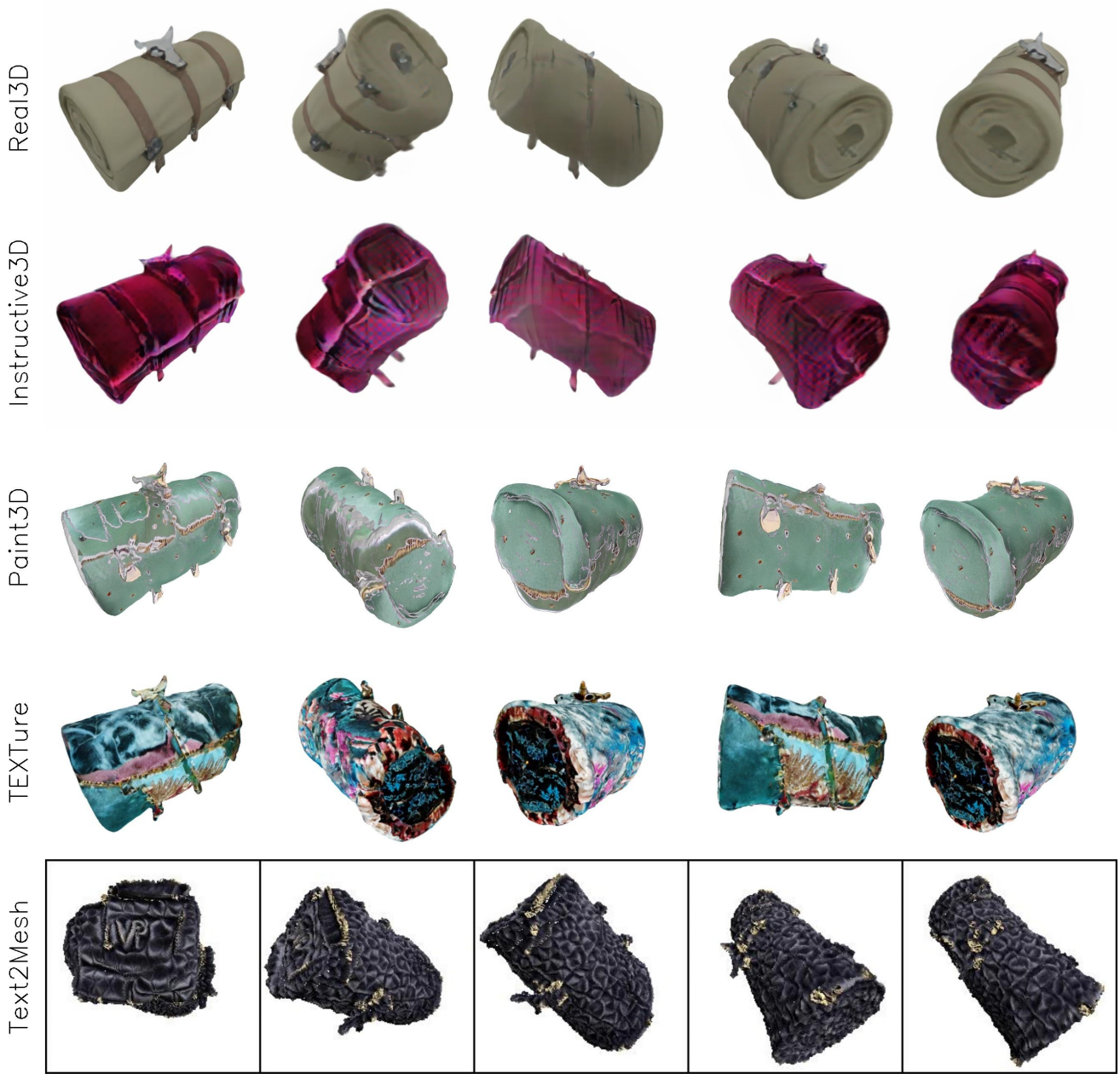

Figure 25. **Baseline comparison results.** Top row shows the rendered images from the mesh obtained from Real3D [1]. Second row shows results from our method. Caption used for editing is: *"add a velvet texture to the clutch bag"*.

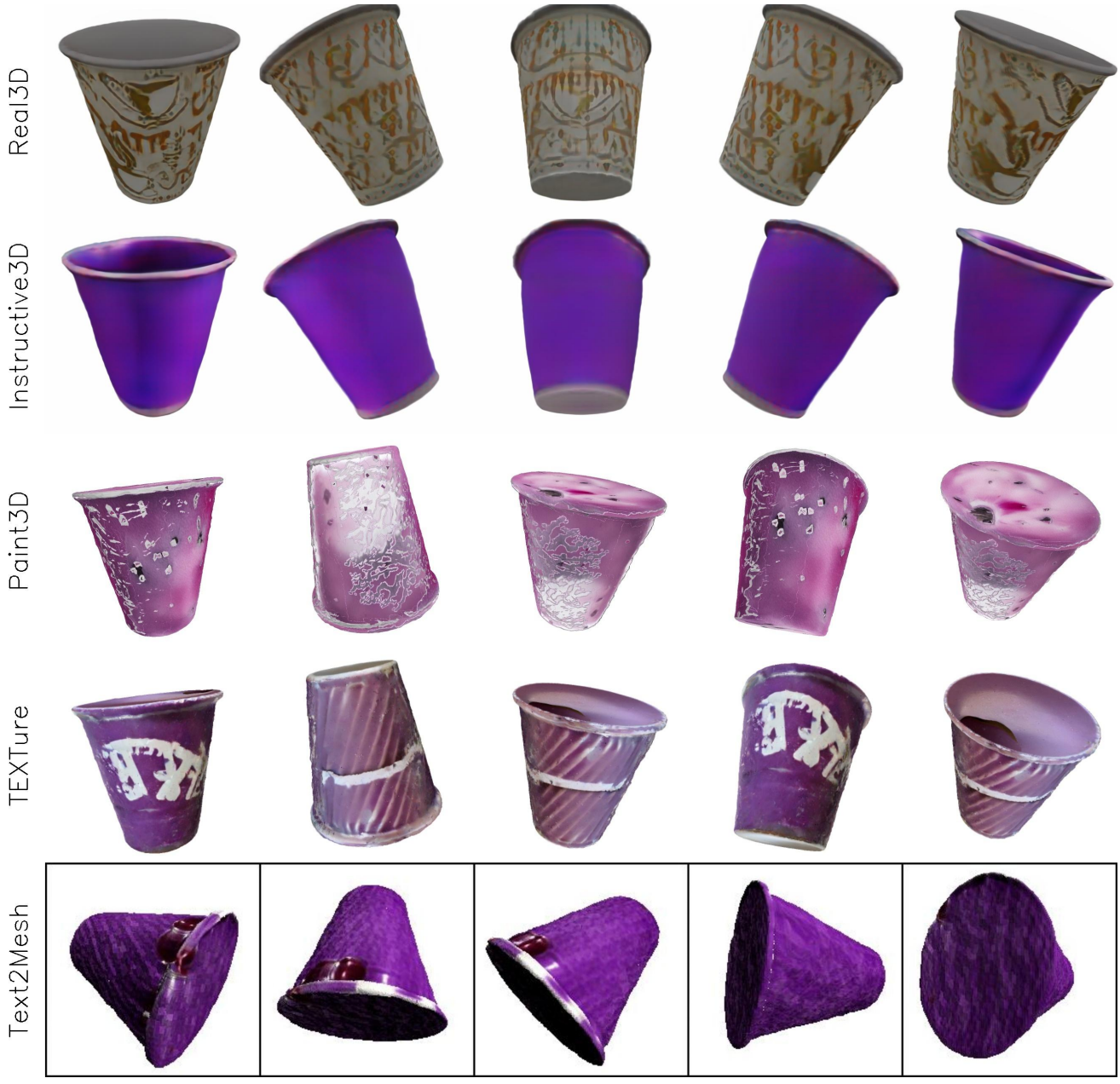

Figure 26. **Baseline comparison results.** Top row shows the rendered images from the mesh obtained from Real3D [1]. Second row shows results from our method. Caption used for editing is: *"change color to purple"*.

Figure 27. **Baseline comparison results.** Top row shows the rendered images from the mesh obtained from Real3D [1]. Second row shows results from our method. Caption used for editing is: *"add a glittery pink overlay to the cup"*.

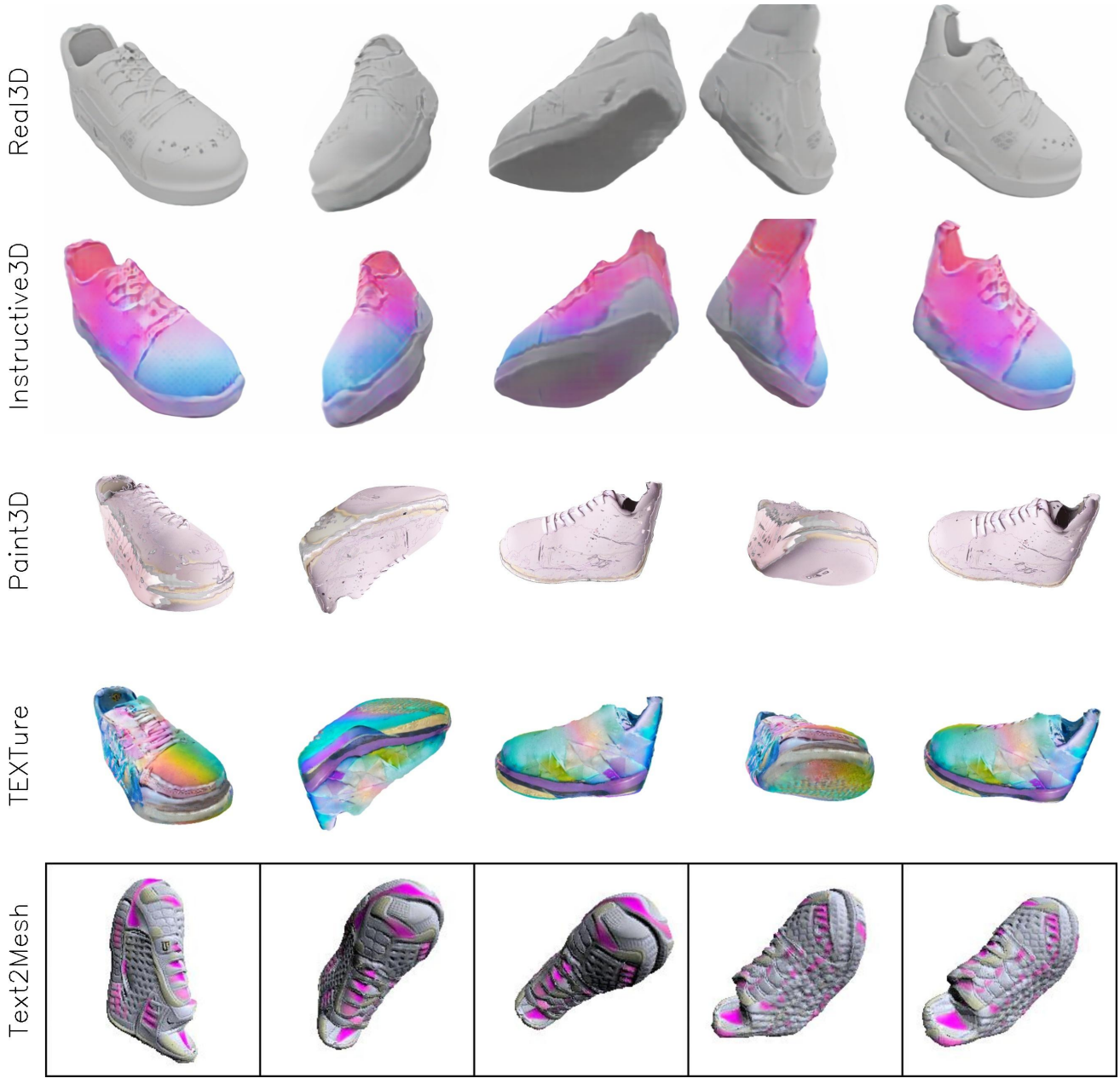

Figure 28. **Baseline comparison results.** Top row shows the rendered images from the mesh obtained from Real3D [1]. Second row shows results from our method. Caption used for editing is: *"add a pastel gradient to the shoe"*.

Figure 29. **Baseline comparison results.** Top row shows the rendered images from the mesh obtained from Real3D [1]. Second row shows results from our method. Caption used for editing is: *"add a flame design to the shoe"*.

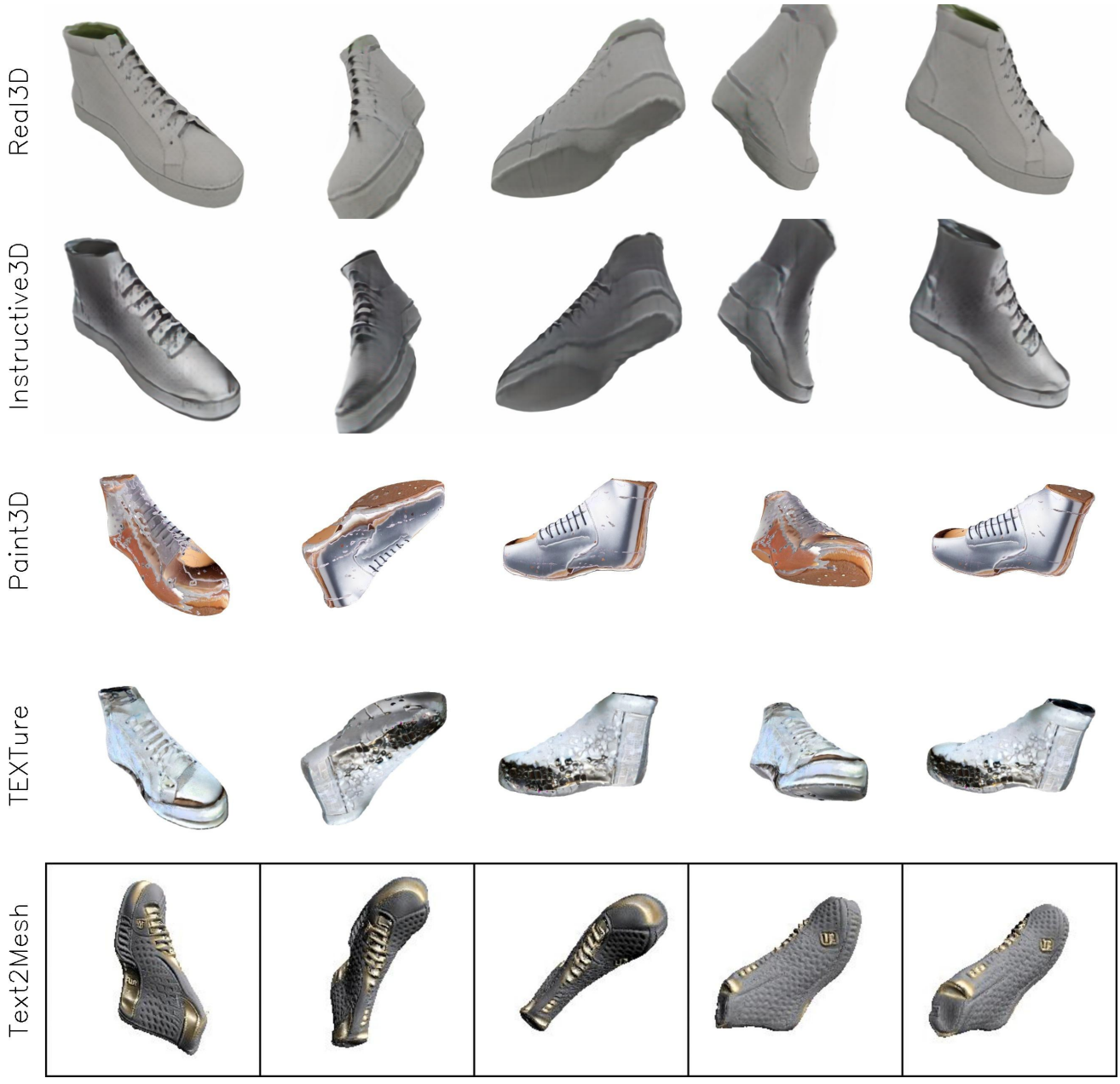

Figure 30. **Baseline comparison results.** Top row shows the rendered images from the mesh obtained from Real3D [1]. Second row shows results from our method. Caption used for editing is: *"add a brushed metal finish to the shoe"*.

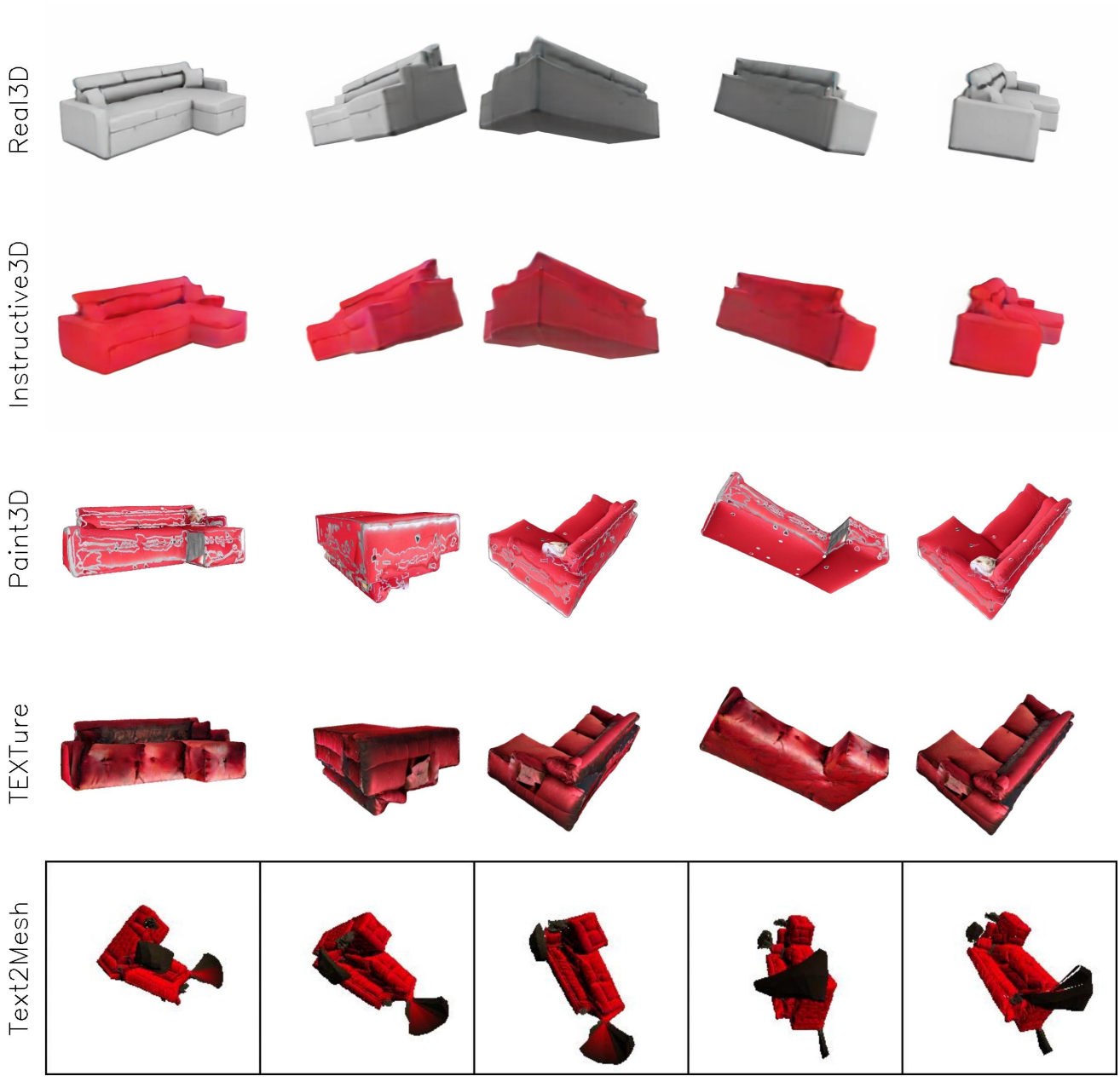

Figure 31. **Baseline comparison results.** Top row shows the rendered images from the mesh obtained from Real3D [1]. Second row shows results from our method. Caption used for editing is: *"change the color of sofa to red"*.

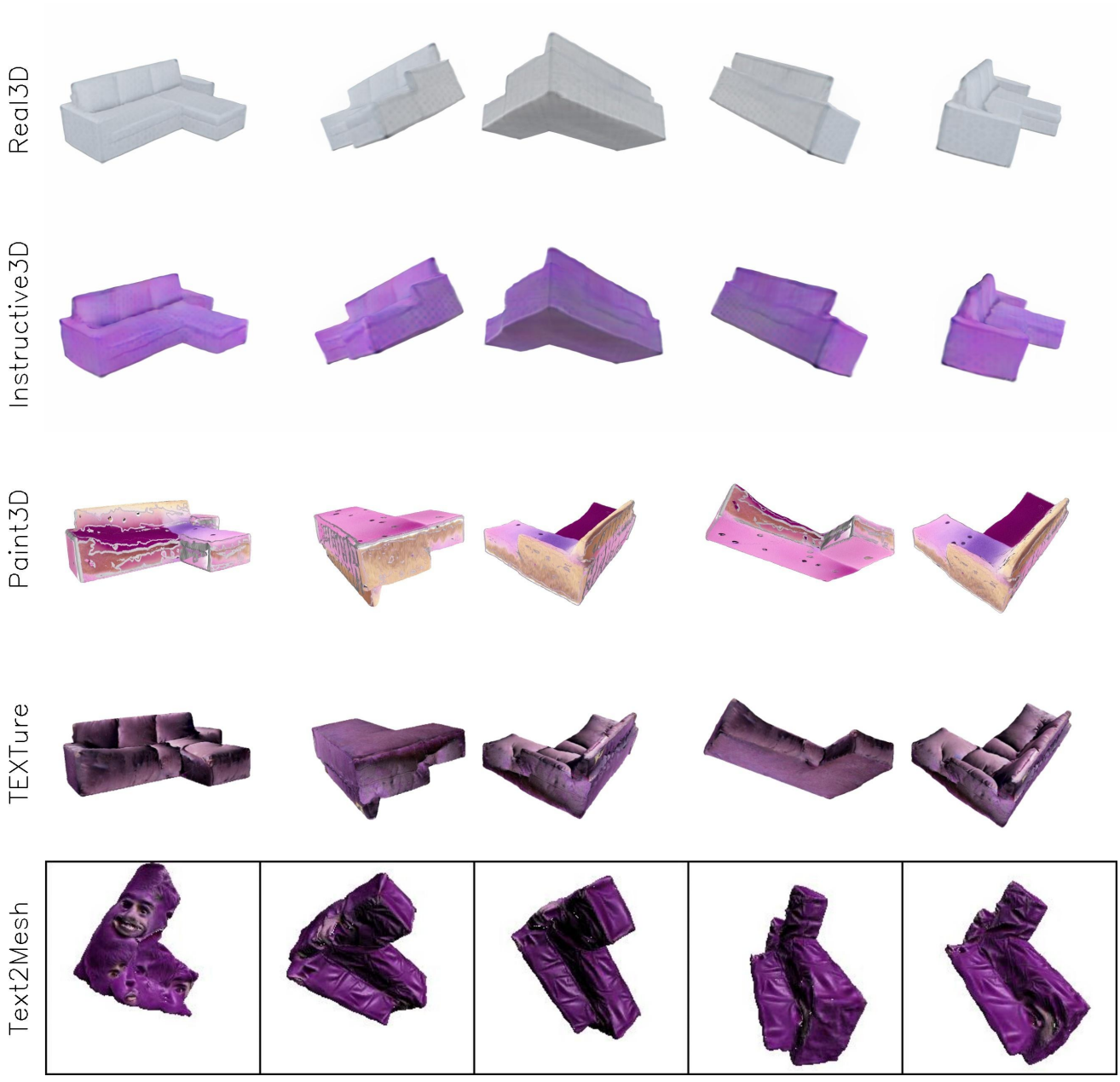

Figure 32. **Baseline comparison results.** Top row shows the rendered images from the mesh obtained from Real3D [1]. Second row shows results from our method. Caption used for editing is: *"change color of sofa to purple"*.

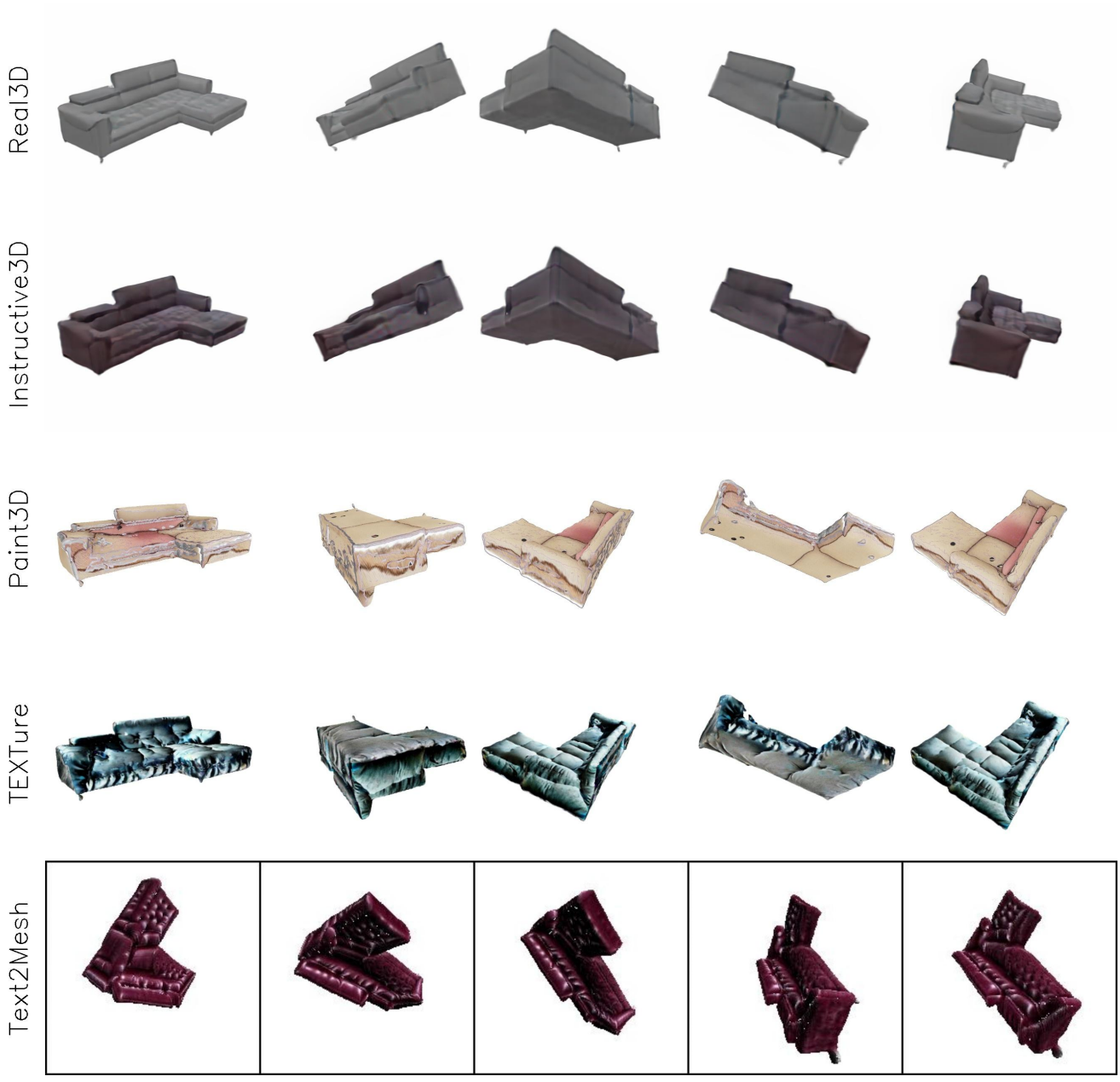

Figure 33. **Baseline comparison results.** Top row shows the rendered images from the mesh obtained from Real3D [1]. Second row shows results from our method. Caption used for editing is: *"darken the color of the sofa"*.