

# HandDiffuser: Text-to-Image Generation With Realistic Hand Appearances

## –Supplementary Material–

### A. Additional Results

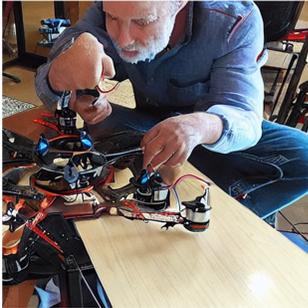
We show additional qualitative comparisons with the Stable Diffusion [2] baseline and the proposed HandDiffuser in Fig. 1 and Fig. 2.

We also show additional qualitative results and intermediate outputs from the two components of HandDiffuser in Fig. 3 and Fig. 4.

### References

- [1] Matthew Loper, Naureen Mahmood, Javier Romero, Gerard Pons-Moll, and Michael J. Black. SMPL: A skinned multi-person linear model. In *ACM Trans. Graphics (Proc. SIGGRAPH Asia)*, 2015. 4, 5
- [2] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 10684–10695, 2022. 1
- [3] Javier Romero, Dimitrios Tzionas, and Michael J. Black. Embodied hands: Modeling and capturing hands and bodies together. *ACM Transactions on Graphics, (Proc. SIGGRAPH Asia)*, 2017. 4, 5

A man sitting a table and repairing a quadcopter.



A woman showing thumbs-up gesture.



A college student holding a file in her left hand.



A male paleontologist examining ancient animals.



Stable Diffusion



HanDiffuser (Ours)

Asian old woman talking on a phone.



A happy African man smiling and being positive.



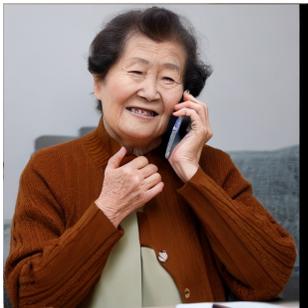
A girl meditating with closed eyes.



A happy bearded man pointing to the side.



Stable Diffusion



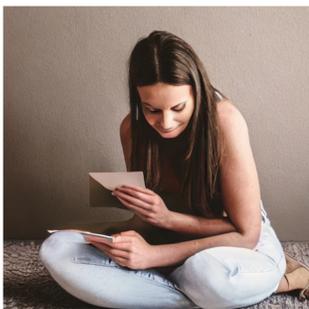
HanDiffuser (Ours)

Figure 1. **Qualitative Comparison.** We compare the quality of hands in images generated from Stable Diffusion and the proposed HanDiffuser.

A girl holding a plant in a blue background.



A woman sitting on the floor reading a card.



A disappointed man wearing a camera.



An accountant working with a computer.



Stable Diffusion

HanDiffuser (Ours)



A man sitting holding a water bottle.



A man in a suit holding a card.



A woman drinking a lemonade outside.



A small girl playing with a black purse.



Stable Diffusion

HanDiffuser (Ours)

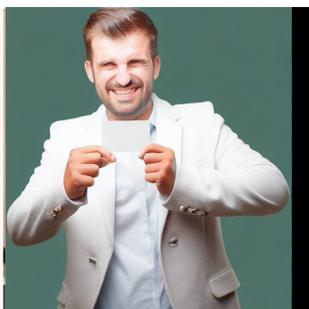


Figure 2. **Qualitative Comparison.** We compare the quality of hands in images generated from Stable Diffusion and the proposed HanDiffuser.

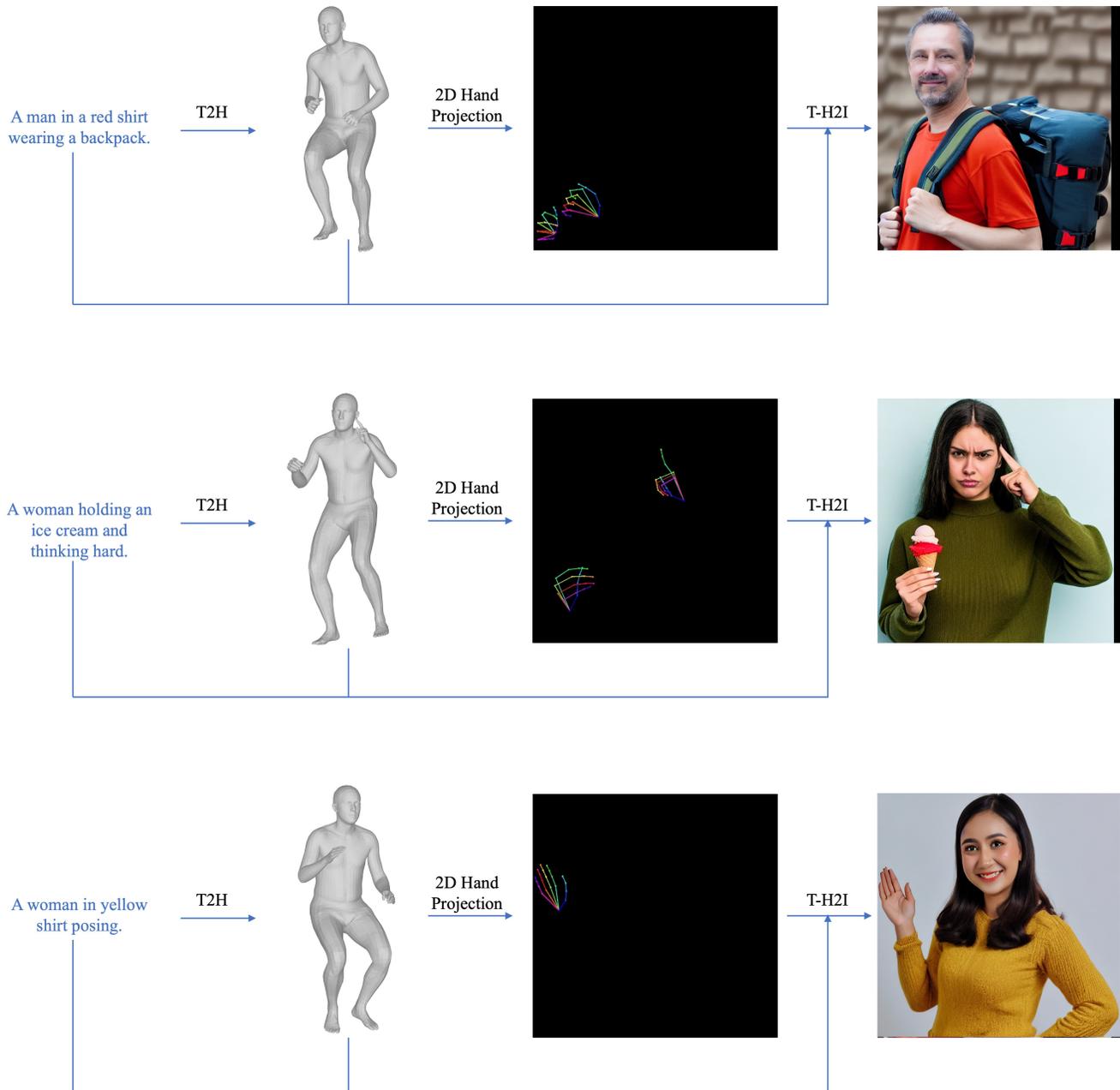


Figure 3. **HanDiffuser Qualitative Results.** Given a text input, Text-to-Hand-Params (T2H) generates SMPL-H [1, 3] parameters. We extract the MANO-Hand from SMPL-H and use some camera parameters to obtain 2D hand poses. The text, MANO-Hand and 2D hand poses are used to generate the final image using Text-Guided Hand-Params-to-Image .

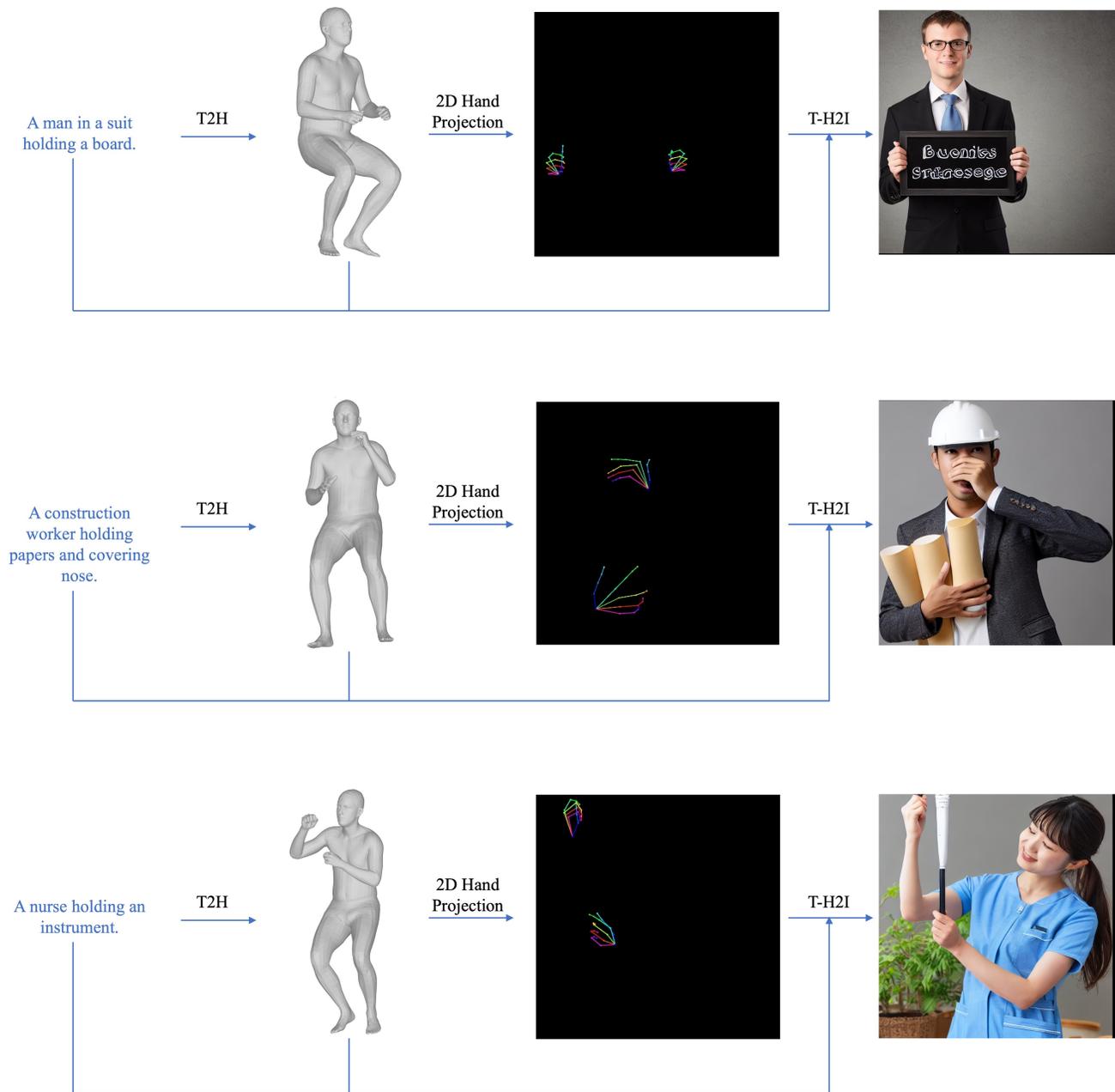


Figure 4. **HanDiffuser Qualitative Results.** Given a text input, Text-to-Hand-Params (T2H) generates SMPL-H [1, 3] parameters. We extract the MANO-Hand from SMPL-H and use some camera parameters to obtain 2D hand poses. The text, MANO-Hand and 2D hand poses are used to generate the final image using Text-Guided Hand-Params-to-Image .