

# Supplementary Materials: S2TD-Face: Reconstruct a Detailed 3D Face with Controllable Texture from a Single Sketch

Anonymous Authors

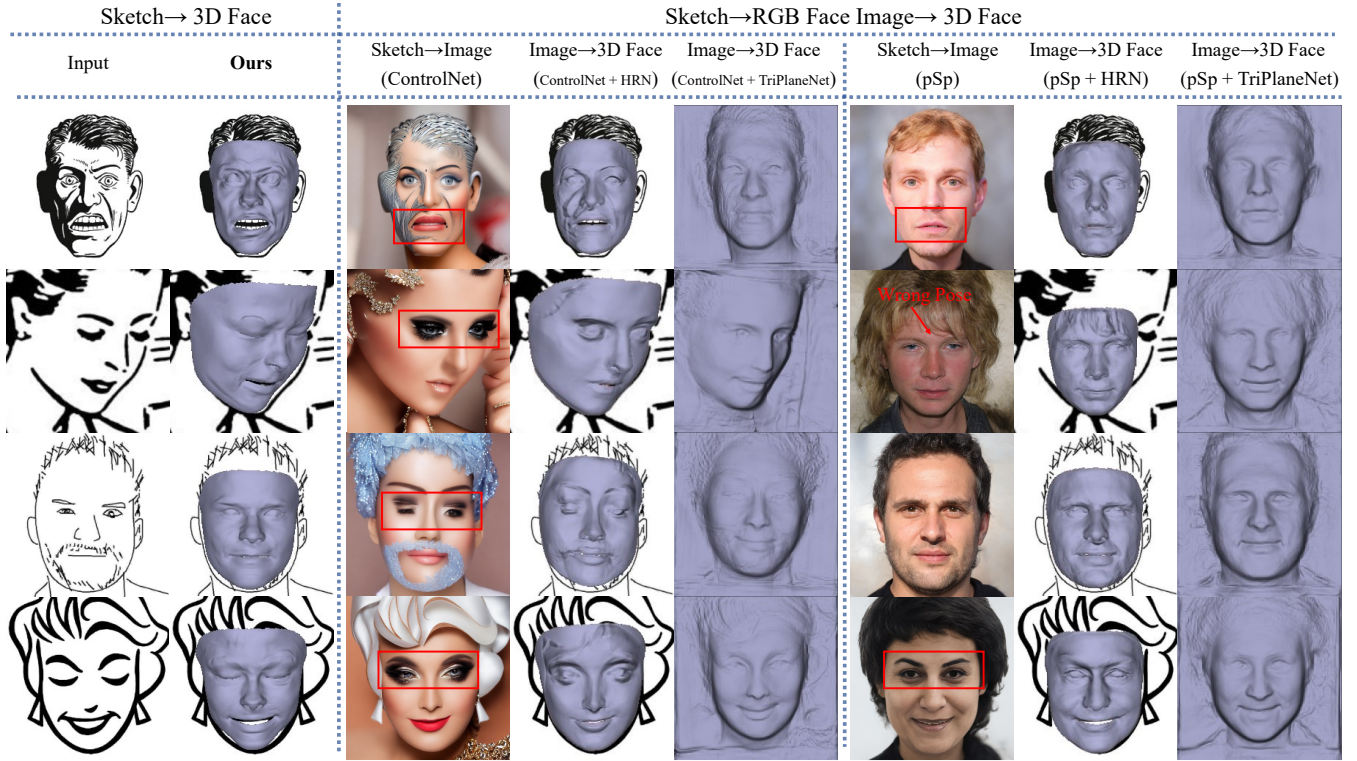


Figure 1: More analysis of the reconstruction approaches ‘Sketch  $\rightarrow$  3D Face’ (S2TD-Face) and ‘Sketch  $\rightarrow$  RGB Face Image  $\rightarrow$  3D Face’. Red boxes indicate areas inconsistent with the input sketch. Direct reconstruction from the sketch optimally preserves geometry (Ours).

## 1 MORE ANALYSIS ABOUT RECONSTRUCTION FRAMEWORK

When considering reconstructing 3D faces from sketches, one option is our proposed S2TD-Face, which directly reconstructs 3D geometry from input sketches. Alternatively, a trivial approach involves initially translating 2D sketches into 2D facial images [8, 10] and then applying existing 3D face reconstruction methods [1, 7] to obtain 3D geometry. In Tab. 1 of the main paper, we have already illustrated that S2TD-Face outperforms the latter significantly in quantitative comparison. To further analyze these two approaches, we implement the latter using state-of-the-art sketch-to-image (ControlNet [10] and pSp [8]) and image-to-3D-face (HRN [7] and TriPlaneNet [1]) methods, and compare the results with those of S2TD-Face. As shown in Fig. 1, these sketch-to-image methods [8, 10] exhibit limited robustness across various sketch styles, failing to translate sketches into face images that maintain consistency with the identity, expression, and pose in input. This indicates that critical geometric information is often lost during the transformation process of ‘Sketch  $\rightarrow$  RGB Face Image  $\rightarrow$  3D

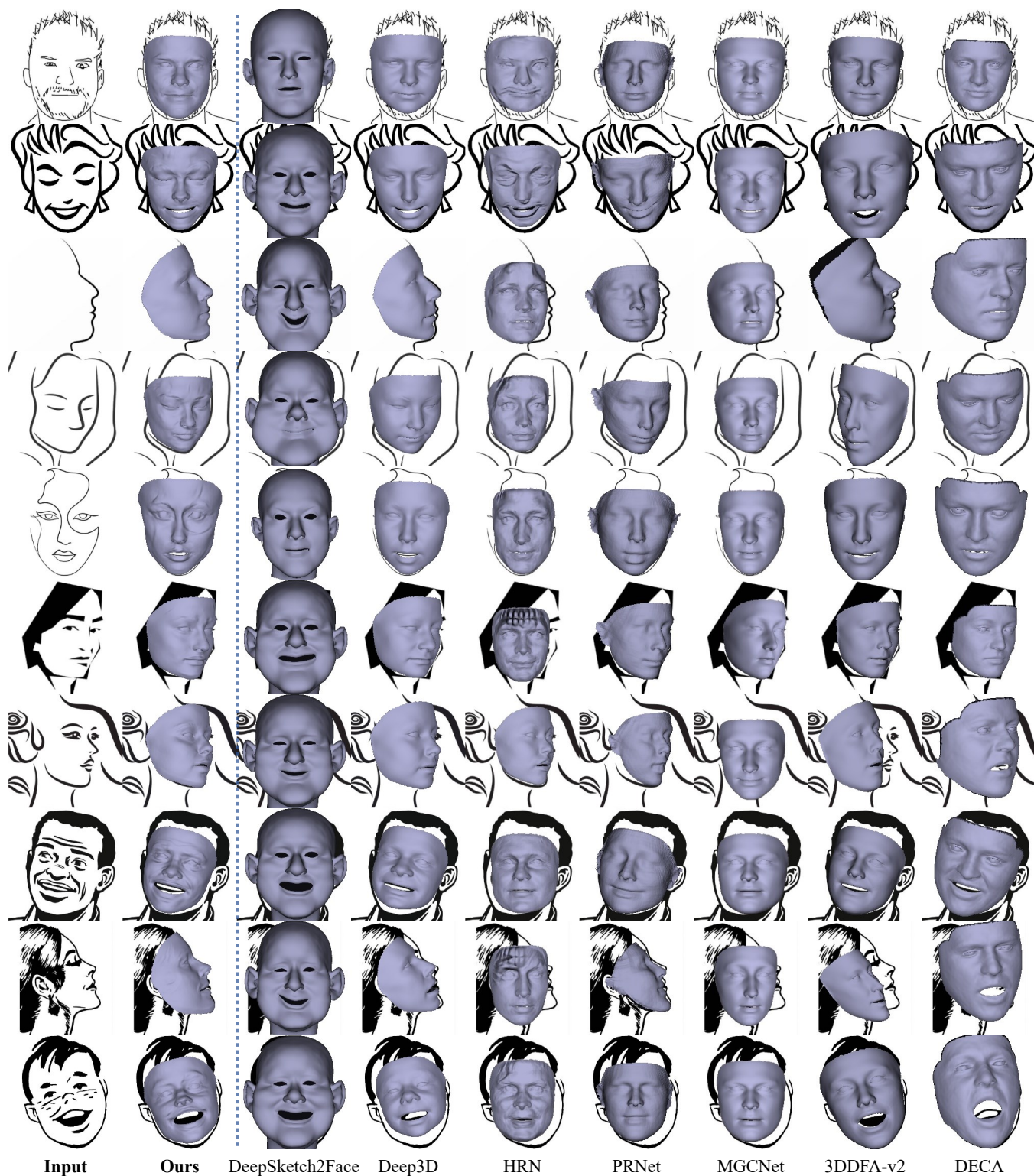
Face’. Note that TriPlaneNet [1] lacks the capability to reconstruct topology-consistent geometry. On the contrary, S2TD-Face is able to reconstruct high-fidelity, topology-consistent detailed geometry from face sketches of diverse styles.

## 2 MORE COMPARISON WITH OTHER METHODS

We further compare the reconstruction results of our S2TD-Face with those of DeepSketch2Face [6], Deep3D [2], HRN [7], PRNet [4], MGCNet [9], 3DDFA-V2 [5], and DECA [3], as shown in the Fig. 2, which indicates that S2TD-Face is capable of handling various styles and poses of facial sketches and achieves the best results consistent with the input sketch details.

## REFERENCES

- [1] Ananta R. Bhattarai, Matthias Nießner, and Artem Sevastopolsky. 2024. TriPlaneNet: An Encoder for EG3D Inversion. (2024).
- [2] Yu Deng, Jiaolong Yang, Sicheng Xu, Dong Chen, Yunde Jia, and Xin Tong. 2019. Accurate 3d face reconstruction with weakly-supervised learning: From single



**Figure 2: Qualitative comparison with the other methods. Our method (S2TD-Face) achieves the best results that consistent with the input sketch details.**

- image to image set. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*. 0–0.
- [3] Yao Feng, Haiwen Feng, Michael J. Black, and Timo Bolkart. 2021. Learning an Animatable Detailed 3D Face Model from In-The-Wild Images. *ACM Transactions on Graphics, (Proc. SIGGRAPH)* 40, 8. <https://doi.org/10.1145/3450626.3459936>
- [4] Yao Feng, Fan Wu, Xiaohu Shao, Yanfeng Wang, and Xi Zhou. 2018. Joint 3d face reconstruction and dense alignment with position map regression network. In *Proceedings of the European conference on computer vision (ECCV)*. 534–551.
- [5] Jianzhu Guo, Xiangyu Zhu, Yang Yang, Fan Yang, Zhen Lei, and Stan Z Li. 2020. Towards fast, accurate and stable 3d dense face alignment. (2020), 152–168.
- [6] Xiaoguang Han, Chang Gao, and Yizhou Yu. 2017. DeepSketch2Face: a deep learning based sketching system for 3D face and caricature modeling. *ACM Transactions on graphics (TOG)* 36, 4 (2017), 1–12.
- [7] Biwen Lei, Jianqiang Ren, Mengyang Feng, Miaomiao Cui, and Xuansong Xie. 2023. A Hierarchical Representation Network for Accurate and Detailed Face Reconstruction from In-The-Wild Images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 394–403.
- [8] Elad Richardson, Yuval Alaluf, Or Patashnik, Yotam Nitzan, Yaniv Azar, Stav Shapiro, and Daniel Cohen-Or. 2021. Encoding in style: a stylegan encoder for image-to-image translation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2287–2296.
- [9] Jiayang Shang, Tianwei Shen, Shiwei Li, Lei Zhou, Mingmin Zhen, Tian Fang, and Long Quan. 2020. Self-supervised monocular 3d face reconstruction by occlusion-aware multi-view geometry consistency. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XV*. Springer, 53–70.
- [10] Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. 2023. Adding Conditional Control to Text-to-Image Diffusion Models.