

Direction-Aware Hybrid Representation Learning for 3D Hand Pose and Shape Estimation

Supplementary Material

In this supplementary material, we report the Percentage of Correct Keypoints (PCK) metric for 3D keypoints on the FreiHAND and HO3Dv2 datasets, provide qualitative results on two additional datasets (COCO [1] and RHD [2]), along with three demo videos, and finally show some failure cases.

A. PCK on FreiHAND and HO3Dv2

To quantitatively assess the performance of the proposed DaHyF, we conduct an extensive evaluation using the PCK metric. The PCK metric is determined by normalizing the Euclidean distances between the predicted and Ground Truth (GT) keypoints by the length of the head segment. A keypoint is deemed correct if its normalized distance falls below a predefined threshold. We plot the PCK curves for thresholds ranging from 0cm to 5cm.

As illustrated in Figures 1 and 2, our DaHyF exhibits superior performance in the 3D PCK metric for procrustes aligned keypoints.

B. Quantitative Results on RHD and COCO

In Figures 3 and 4, we compare our method with Mesh Graphormer (one of the state-of-the-art methods) on the RHD and COCO datasets, respectively. Both use HRNet-W64 as the backbone. As shown in the figures, DaHyF is effective at handling challenging cases such as small hands, occlusion, and appearance variance, thanks to our proposed 2D+3D end-to-end joint optimization framework with direction-aware hybrid features.

C. Video Results on Temporal Filtering with the Predicted Confidence

We demonstrate the effectiveness of our confidence prediction module on three videos with small hands and severe motion blur. Figure 5 shows one extracted frame from each of the three videos. Our method provides more robust and smoother results with less jittering and flipping, thanks to the predicted confidence in temporal filtering. The three videos are also provided separately in the supplementary material:

- Video_Comparison_with_Mesh_Graphormer_1.mp4
- Video_Comparison_with_Mesh_Graphormer_2.mp4
- Video_Comparison_with_Mesh_Graphormer_3.mp4

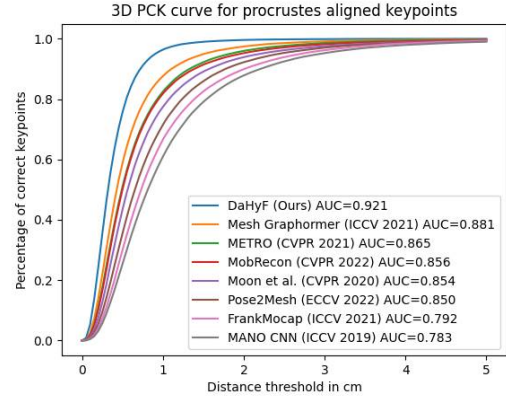


Figure 1. 3D PCK on FreiHAND.

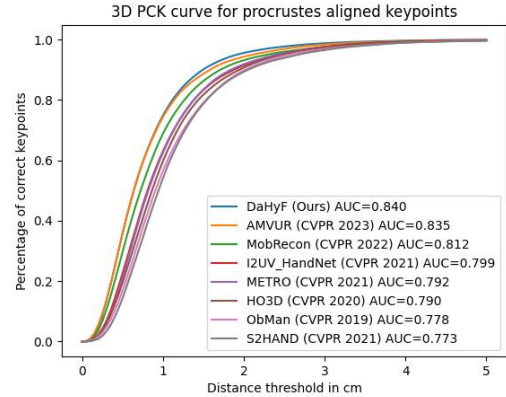


Figure 2. 3D PCK on HO3Dv2.

D. Failure Cases

We show some examples in Figure 6 where our method fails to estimate the poses correctly. When there is severe occlusion or motion blur across multiple consecutive frames, our method faces challenges in accurately estimating the hand poses. Note that Mesh Graphormer also fails to do so.

References

- [1] L. Tsung-Yi, M. Michael, B. Serge, and et al. Microsoft coco: Common objects in context. In *ECCV*, 2014. 1
- [2] C. Zimmermann and T. Brox. Learning to estimate 3d hand pose from single rgb images. In *ICCV*, 2017. 1

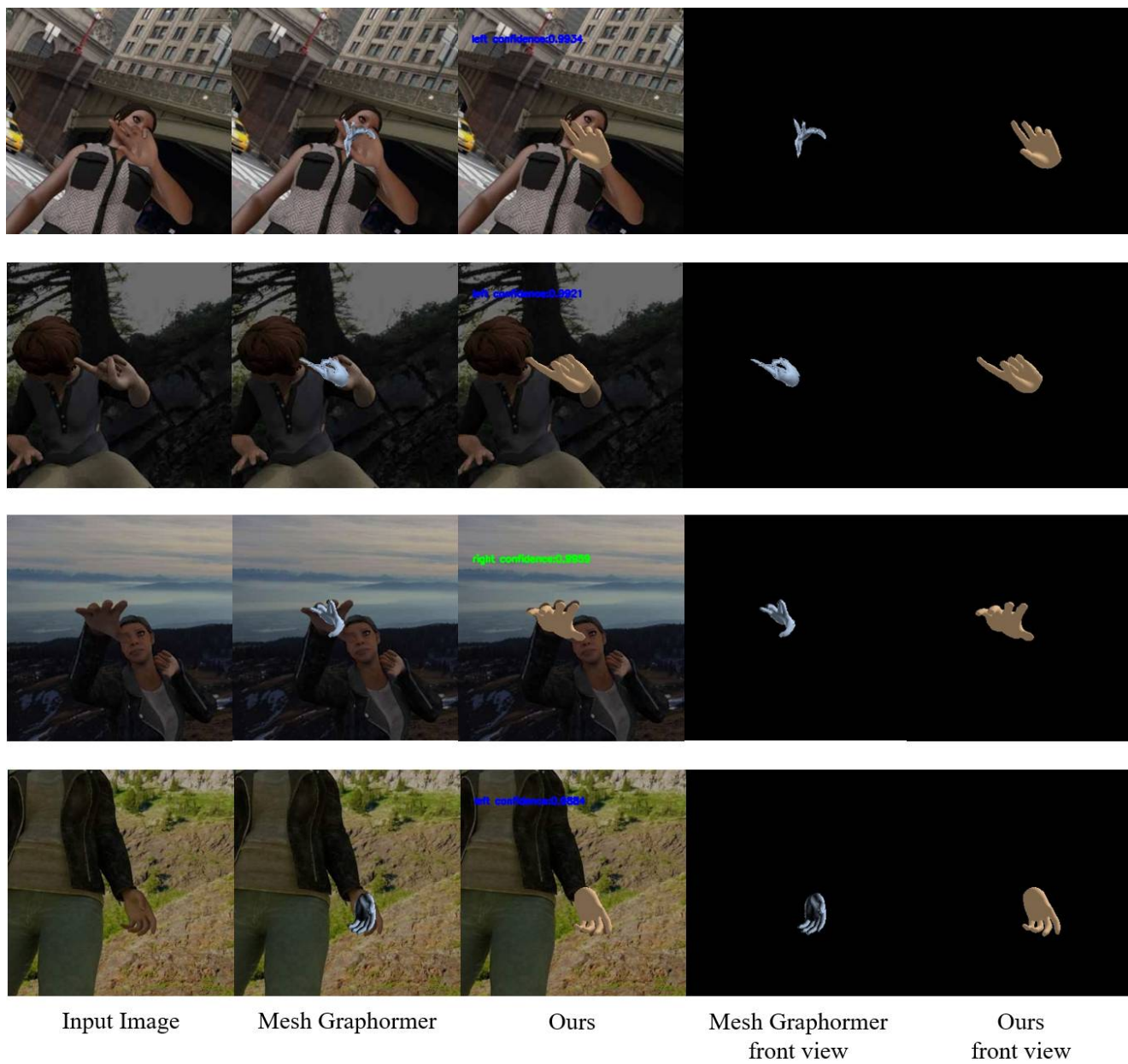


Figure 3. Qualitative comparison between DaHyF and Mesh Graphormer on the RHD dataset.

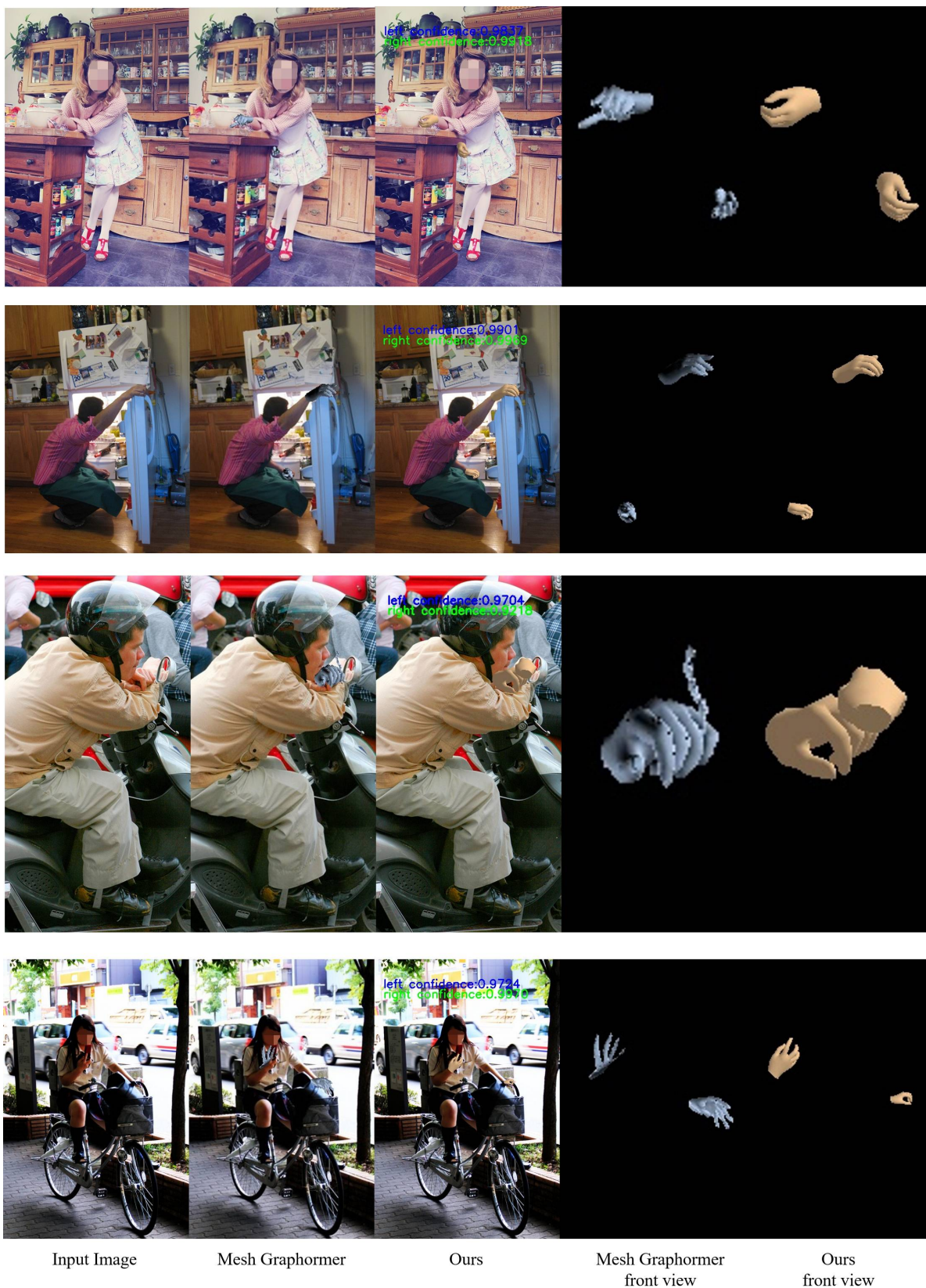


Figure 4. Qualitative comparison between DaHyF and Mesh Graphormer on the COCO dataset .



Figure 5. Qualitative results on three test videos.

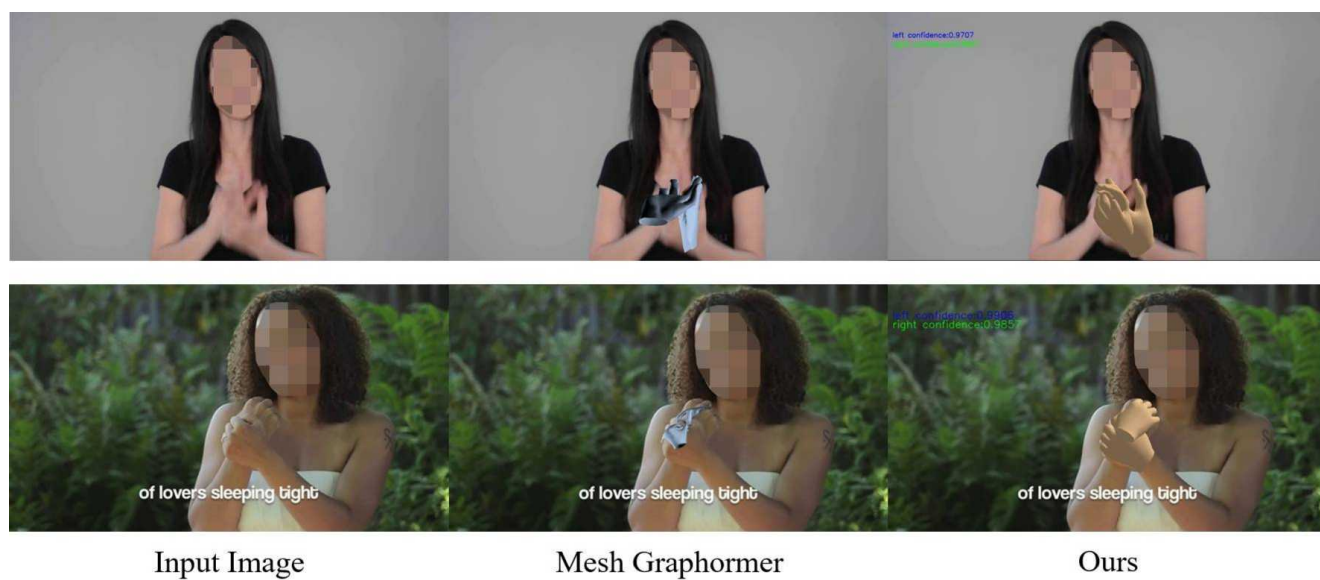


Figure 6. Failure cases.