End-to-end Hand Mesh Recovery from a Monocular RGB Image Supplementary Material

Xiong Zhang[1*], Qiang Li[1*], Hong Mo[2], Wenbo Zhang[1], Wen Zheng[1]
[1]Y-tech, Kwai, [2]State Key Laboratory of VR, Beihang University
[1]{zhangxiong,liqiang03,zhangwenbo,zhengwen}@kuaishou.com, [2]mandymo@buaa.edu.cn

We have exploited the segmentation mask to train the network, as space is limited, we did not show the quantitative experimental result of hand segmentation in our paper. Regard this concern, this additional document quantitative analysis the accuracy of hand segmentation, by projecting the reconstructed hand mesh, which also reflects the precision of the reconstructed mesh from one aspect. Following conventional, we use the mIoU (mean Intersection over Union) as the evaluation metric. We shall point out here, since we can not deduce the ground-truth mask on the STB and Dexter datasets, as a result, we conduct the following experiment on RHD dataset.

The Fig. 1 presents some examples of segmentation mask and Tab. 1 gives the quantitative comparison result.

Figure 1: **Segmentation Examples.** The graph comprising two rows, the first row presents the RGB images drawn from the RHD testing part, and the second row demonstrates the segmentation mask of each example.

Table 1: Performance of segmentation on RHD dataset.

Method	**mIoU**
HMR [3]	0.750
BodyNet [5]	0.852
SMPLify [1]	0.739
Georgios *et.al* [4]	0.806
DeepLab [2]	0.924
HAMR (Ours)	0.931

Quantitative experimental result illustrated in Tab. 1 reveals that our methods outperforms all those state-of-the art methods. We are not astonished by the outstanding performance, since the introducing of parametric hand model can solve the ambiguity, in addition the silhouette consistent loss can refine the hand shape and pose prediction.

References

[1] Federica Bogo, Angjoo Kanazawa, Christoph Lassner, Peter Gehler, Javier Romero, and Michael J Black. Keep it smpl: Automatic estimation of 3d human pose and shape from a single image. In *European Conference on Computer Vision*, pages 561–578. Springer, 2016. 1

[2] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence*, 40(4):834–848, 2018. 1

[3] Angjoo Kanazawa, Michael J Black, David W Jacobs, and Jitendra Malik. End-to-end recovery of human shape and pose. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 7122–7131, 2018. 1

[4] Georgios Pavlakos, Luyang Zhu, Xiaowei Zhou, and Kostas Daniilidis. Learning to estimate 3d human pose and shape from a single color image. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 459–468, 2018. 1

[5] Gul Varol, Duygu Ceylan, Bryan Russell, Jimei Yang, Ersin Yumer, Ivan Laptev, and Cordelia Schmid. Bodynet: Volumetric inference of 3d human body shapes. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 20–36, 2018. 1