Supplementary Material: Object-Occluded Human Shape and Pose Estimation from a Single Color Image

Tianshu Zhang* Buzhen Huang* Yangang Wang†
{tszhang,hbz,yangangwang}@seu.edu.cn
Southeast University, China

1. Additional results

In our video, we show the results which are generated in a frame by frame manner and directly resampled from the output UV maps **without SMPL fitting** or **smooth operation**.

In Fig.1, we show more qualitative results on 3DPW [2], 3DOH50K and Human3.6M with synthetic occlusion [1]. The third column of every example is the mesh directly resampled from the predicted UV map without any extra operations.

On 3DPW dataset, occluded sequences are used to evaluate. Details are provided in Tab.1.

2. Comparison of different structures

In this section, we compare the performance of different model structures to further illustrate the superiority of two-branch structure. Fig.2 illustrates the pipeline of each model structure. The comparing qualitative results are shown in Fig.3 and the quantitative results can be referred in main Table 2.

end-to-end. To verify the validity of the partial UV map on occlusion representation, we compare our method with End-to-End in Fig.2 (a). The network is designed in the end-to-end manner and occluded color images are directly feed to an encoder-decoder network to get a full UV map. The results in Fig.3 (b) shows it is quite difficult to directly reconstruct the mesh occluded by objects by end-to-end network.

cascade. As shown in Fig.2 (b), cascade pipeline first estimates the partial UV map from the occluded color image, then uses the fixed inpainting network to get the full UV map. The difference between our method and this pipeline is latent space supervision. Experiments show that our method has better performance and latent space supervision works well in this problem.

(w/o) saliency map. The structure of Fig.2 (c) removes the saliency map estimation module of our pipeline which is used to illustrate the effectiveness of the saliency map.

Sequence Name
courtyard_backpack
courtyard_basketball
courtyard_bodyScannerMotions
courtyard_box
courtyard_golf
courtyard_jacket
courtyard_laceShoe
downtown_stairs
flat_guitar
flat_packBags
outdoors_climbing
outdoors_crosscountry
outdoors_fencing
outdoors_freestyle
outdoors_golf
outdoors_parcours
outdoors_slalom

Table 1. **Occluded sequences in 3DPW dataset.**

References

[1] Catalin Ionescu, Dragos Papava, Vlad Olaru, and Cristian Sminchisescu. Human3.6m: Large scale datasets and predictive methods for 3d human sensing in natural environments. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(7):1325–1339, jul 2014.

[2] Timo von Marcard, Roberto Henschel, Michael J. Black, Bodo Rosenhahn, and Gerard Pons-Moll. Recovering accurate 3d human pose in the wild using imus and a moving camera. In *ECCV*, 2018.

†Corresponding author.
*Contribute equally. The authorship was determined by a coin toss.

Figure 1. **More results on 3DPW (rows 1-2), 3DOH50K (rows 3-4) and Synthetic Human3.6M dataset (rows 5-6).**

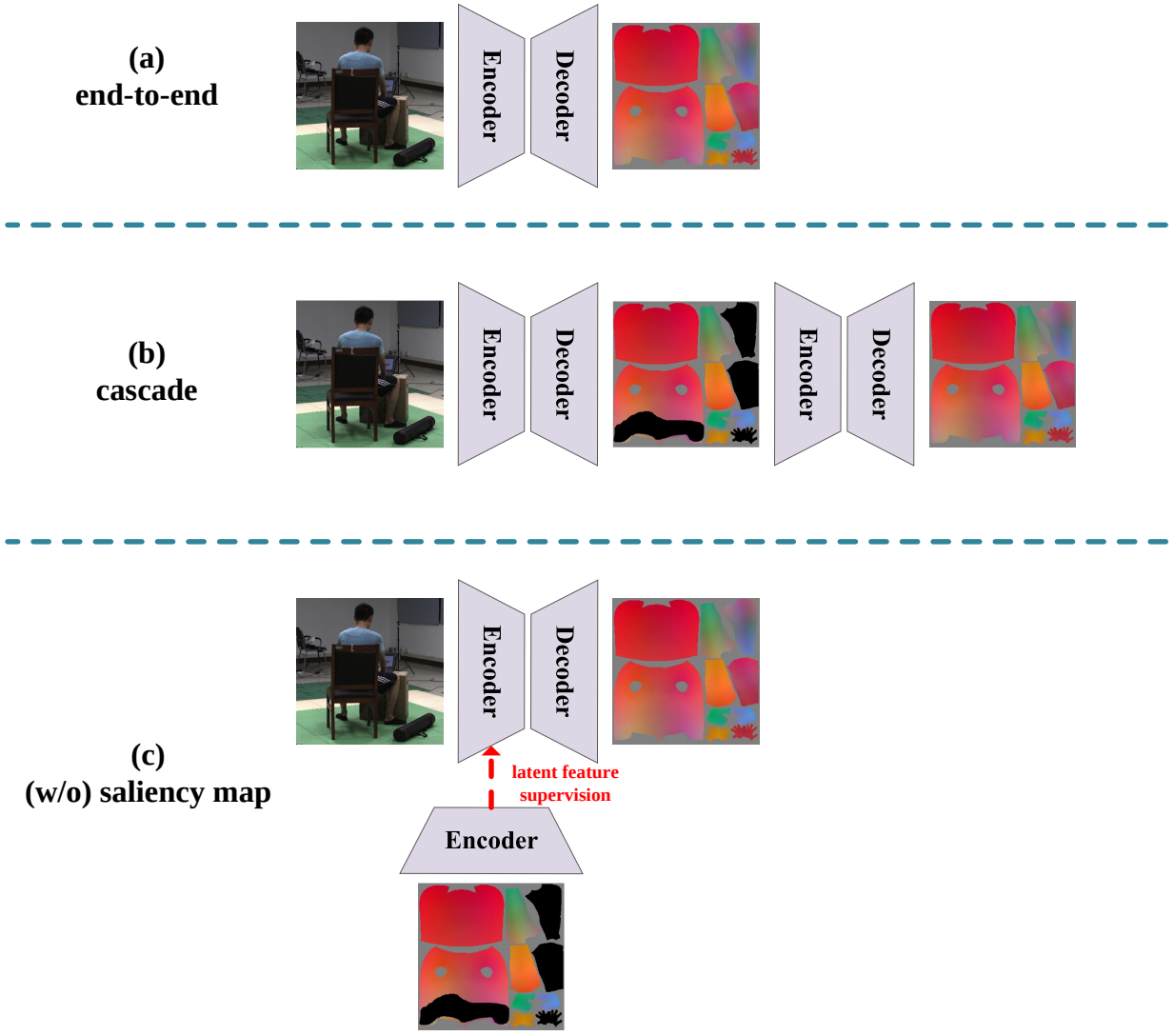

Figure 2. **Detailed architecture of different structures.**

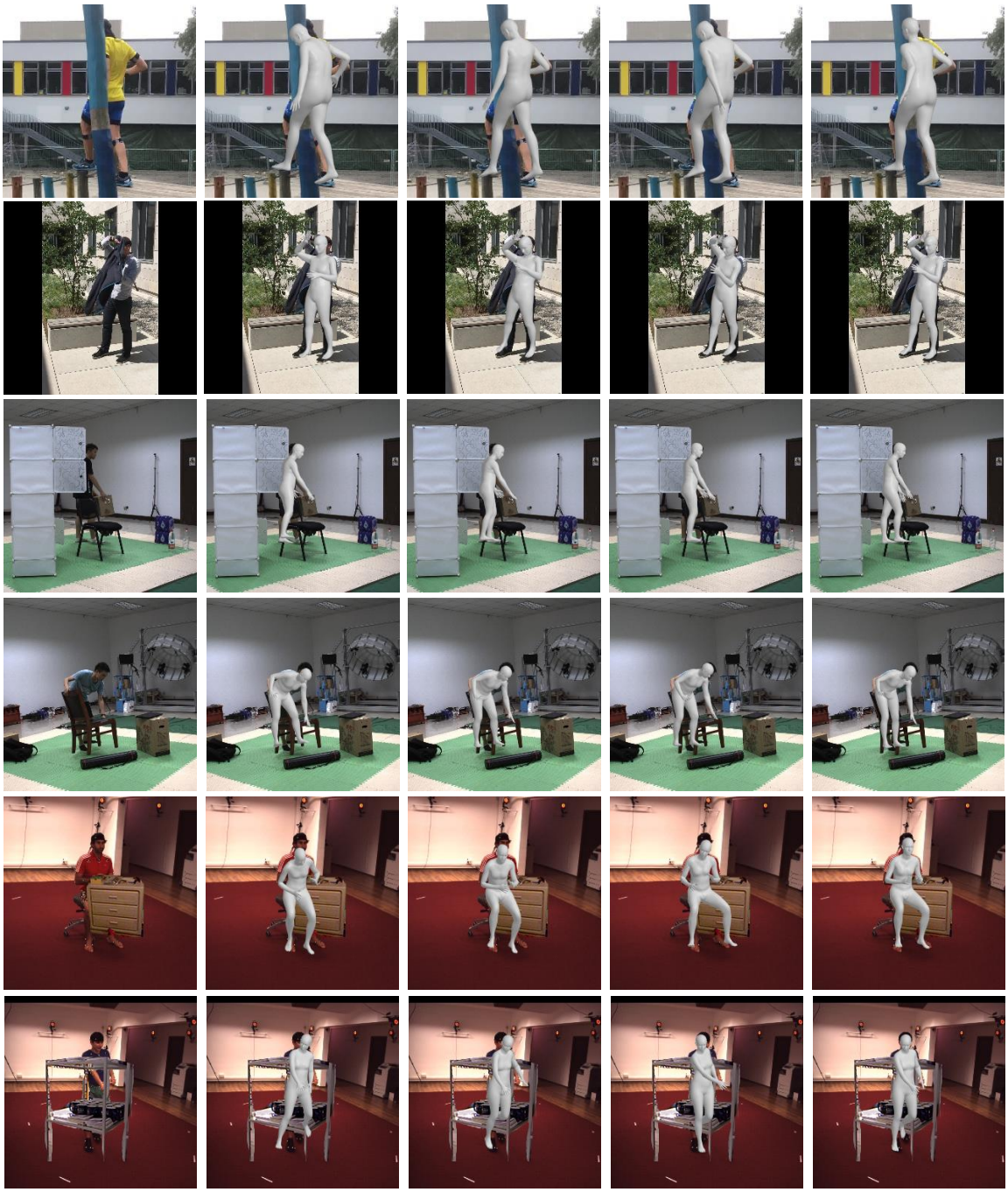

Figure 3. **Comparison of different structures.**