Supplemental Materials: Learning Anchored Unsigned Distance Functions with Gradient Direction Alignment for Single-view Garment Reconstruction

Fang Zhao[1] Wenhao Wang[2] Shengcai Liao[1*] Ling Shao[1,3]
[1]Inception Institute of Artificial Intelligence [2]ReLER, University of Technology Sydney
[3]Mohamed bin Zayed University of Artificial Intelligence

In the following, we provide more details on model training and inference procedures (Sec. 1) and more reconstruction results on input images with various views and garment images from real world data (Sec. 2).

1. Model Training and Inference

To generate training point sets, we randomly sample points on the ground-truth surface and displace them with Gaussian distribution $\mathcal{N}(0, \sigma)$ along x, y and z axis, where 1% of points are sampled from $\sigma = 0.08$, 49% of points from $\sigma = 0.02$ and 50% of points from $\sigma = 0.003$ as suggested by [1]. For training, we use the RMSprop optimizer with a learning rate of 5e-5. The batch size is 4 and the number of epochs is 35 for MGN and 60 for Deep Fashion3D. The learning rate is decayed by the factor of 0.1 in the last 20 epochs. We jointly optimize the UDF and anchor point regression losses from the beginning of training, and add the gradient direction loss to fine-tune the decoder in the last 10 epochs while fixing the encoders for training efficiency. All compared models are trained by the codes provided their authors. For PIFu [4] and PIFuHD [5], we use the same backbone structure with our method. For BCNet [2], we use the trained model provided by its authors because its training code is not released. At inference, the step number of projecting points is set to 5 and the valid distance to the surface is set to 0.007, which produce robust reconstruction results. Please refer to [1] for detailed algorithm steps of dense point cloud extraction.

We illustrate the flow chart of our AnchorUDF-HD which incorporates the HD module [5] into AnchorUDF in Fig. 1. Here a 3D embedding extracted from the decoder of AnchorUDF and local image features of high-resolution input (1024×1024) are fed simultaneously into the decoder of the HD module to predict UDF. To learn AnchorUDF-HD, we first train AnchorUDF as the training procedure described before. Then, we add the HD module and continue to train the entire model for 15 epochs.

*Corresponding author.

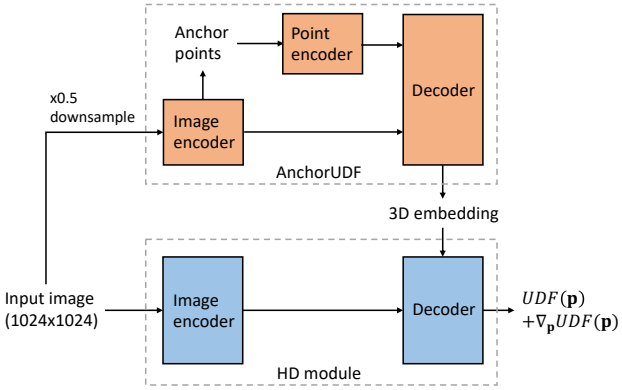

Figure 1. Flow chart of our AnchorUDF-HD. It incorporates the HD module into AnchorUDF, which takes high-resolution images as input.

2. More Results

We visualize more reconstruction results of our method on MGN (Fig. 2) and Deep Fashion3D (Fig. 3) datasets. As one can see, our method can faithfully recover detailed surfaces for inputs with various views and infer plausible shapes for self-occluded regions.

We also test our model trained with Deep Fashion3D on real garment images from DeepFashion dataset [3]. Here we use semantic segmentation annotations provided by the dataset to obtain input garment images. As shown in Fig. 4, our method produces promising reconstruction results for different garment categories, which capture multiple topologies and retain local details present in input images. Note that we do not use real garment images during training and there are some noises in ground truth point clouds which affect the genuineness of rendered training images.

References

[1] Julian Chibane, Aymen Mir, and Gerard Pons-Moll. Neural unsigned distance fields for implicit function learning. In *Ad-*

Figure 2. More garment reconstruction results on MGN dataset.

Figure 3. More garment reconstruction results on Deep Fashion3D dataset.

Figure 4. Reconstruction results on real garment images from DeepFashion dataset [3]. Our method can capture topologies of different garment categories and retain local details present in input images. Note that some artifacts are caused by hand occlusion from the original input images.

vances in neural information processing systems (NeurIPS), 2020. 1

[2] Boyi Jiang, Juyong Zhang, Yang Hong, Jinhao Luo, Ligang Liu, and Hujun Bao. Bcnet: Learning body and cloth shape from a single image. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2020. 1

[3] Ziwei Liu, Ping Luo, Shi Qiu, Xiaogang Wang, and Xiaoou Tang. Deepfashion: Powering robust clothes recognition and retrieval with rich annotations. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, 2016. 1, 4

[4] Shunsuke Saito, Zeng Huang, Ryota Natsume, Shigeo Morishima, Angjoo Kanazawa, and Hao Li. Pifu: Pixel-aligned implicit function for high-resolution clothed human digitization. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2019. 1

[5] Shunsuke Saito, Tomas Simon, Jason Saragih, and Hanbyul Joo. Pifuhd: Multi-level pixel-aligned implicit function for high-resolution 3d human digitization. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020. 1