Supplementary Material for
NeuralHDHair: Automatic High-fidelity Hair Modeling from a Single Image Using Implicit Neural Representations

A. Implementation Details

The image encoder of the coarse module contains 5 downsampling layers and 4 upsampling layers with (32, 64, 128, 256, 256) and (256, 128, 64, 32) feature channels, respectively, where skip connections are added between them to capture more information. Note that our skip connections utilize the implicit toVoxel module to expand the 2D features to 3D (e.g., 8×8 to $6 \times 8 \times 8$, 16×16 to $12 \times 16 \times 16$). Finally, the size of the output voxel-wise latent code is $96 \times 128 \times 128 \times 64$. The MLP for the coarse module has (65, 256, 128, 64, 3) and (65, 256, 128, 64, 1) neurons for the orientation field and the occupancy field, respectively. Here the output of the second layer is concatenated with the local features as well as depth z before being fed into fine module's MLP. Thus, the MLP for the fine module has (289, 512, 256, 128, 64, 3) neurons to refine the orientation field and (289, 512, 256, 128, 64, 1) neurons to refine the occupancy field. The coarse module is pre-trained with the 2D orientation map resized to 256×256 while the fine module is trained with the luminance map resized to 1024×1024. The GrowingNet is composed of an encoder and a decoder. The encoder E contains several downsamplings with output channels (3, 16, 32, 64, 128) to compress the local patch into a latent code, and the decoder D is an MLP with the number of neurons of (131, 128, 64, 32, 3). Our IRHairNet and GrowingNet are implemented using the PyTorch framework and trained with the Adam optimizer for 2-3 days and 1 day, respectively. Our learning rate is 0.0001, and it decays every 20 epochs.

B. More comparisons

To better compare and demonstrate the effectiveness of our method, we compared with Dynamic Hair [2] and PIFuHD [1] on large-scale test data using some quantitative metrics similar to [2] and conducted a user study as shown in Tab. 1. We use precision for occupancy field while the L2 error for orientation field on synthetic data. We calculate the L1 error between the projection of the growth direction of each point on the strand with the 2D orientation to measure the model's performance on the real data. Our user study involved 38 users and 25 test cases, and 65.67% chose our reconstruction as the best results.

Method	**Synthetic data**		**Real data**	
	Precision(%)	L2	L1	User study(%)
PIFuHD	71.08	0.1543	0.2662	14.95
Dynamic Hair	73.14	0.1293	0.2091	19.38
Ours	**76.36**	**0.1040**	**0.1458**	**65.67**

Table 1. Quantitative comparison and a user study.

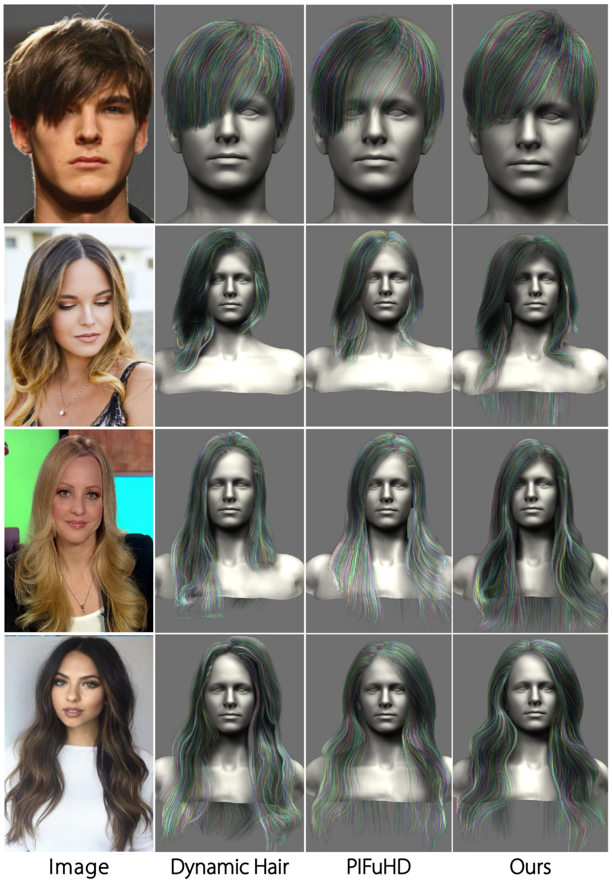

Figure 1. More qualitative comparison with Dynamic Hair [2] and PIFuHD [1].

In addition, as shown in Fig. 1, we also demonstrate more qualitative comparative examples to prove that our method achieves the SOTA.

References

[1] 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/CVF Conference on Computer Vision and Pattern Recognition*, pages 84–93, 2020. 1

[2] Lingchen Yang, Zefeng Shi, Youyi Zheng, and Kun Zhou. Dynamic hair modeling from monocular videos using deep neural networks. *ACM Transactions on Graphics (TOG)*, 38(6):1–12, 2019. 1