Supplementary Materials for Human Hair Reconstruction with Strand-Aligned 3D Gaussians

Egor Zakharov[1], Vanessa Sklyarova[1,2], Michael Black[2], Giljoo Nam[3], Justus Thies[2,4], and Otmar Hilliges[1]

[1] ETH Zürich, Switzerland
[2] Max Planck Institute for Intelligent Systems, Tübingen, Germany
[3] Meta, Pittsburgh, USA
[4] Technical University of Darmstadt, Germany

1 Implementation Details

Below we describe each processing step of our proposed method in more detail.

1.1 Preprocessing

We experiment with video-based captures but can employ a similar pipeline to process image-based captures. First, we extract the frames from the video with the framerate of 3 FPS. To reduce the motion blur, among the frames spanning each 1/3 of a second, we pick the one with the highest image quality score calculated via a HyperIQA [7] network. We run COLMAP SfM [4] algorithm for the extracted frames to estimate the initial point cloud and the camera parameters. To obtain the segmentation masks, we rely on Matte-Anything [8] prompt-based image matting system. To produce the corresponding segmentation masks, we use the following prompts: "hair", "face", and "human". We then filter the cases where segmentation masking fails. For that, we have designed a heuristic in which we calculate the intersection between the hair and face and prune the frame if it is higher than 10% of the body mask area. Lastly, we calculate image quality scores again for the masked hair crops. We split the filtered images into consecutive 128 bins and picked the image with the highest score from each bin for training, resulting in 128 training views. These pre-processing steps closely follow the ones used in Neural Haircut [5]. We rely on the pipeline proposed in [1] for orientation map calculation and use the code of Neural Haircut [5].

1.2 Guiding Strands Upsampling

We follow HAAR [6] in the procedure that we use for the interpolation of the hair strands. In this work, they perform upsampling using a regular grid of latent maps and blend the strands produced using nearest and bilinear interpolation methods. In our case, we use irregular grids and thus use K nearest neighbors for both, replacing bilinear interpolation with the one that blends four nearest neighbors. The KNN-based interpolation weights are inversely proportional to the distance between the interpolated strands' origins and the query origin in the space of texture coordinates. Another change compared to HAAR is that we conduct the interpolation in the space of hair maps instead of latent maps. Thus, we directly interpolate the 3D coordinates of the strands, each being defined within its own TBN (tangent-bitangent-normal) basis, which is defined using the scalp face where the strand originates [3].

1.3 Diffusion-Based Prior

Our diffusion-based prior follows Neural Haircut [5]. We rely on Score Distillation Sampling (SDS) [2] to enforce this prior at the level of latent hair maps. Thus, during the coarse strand fitting stage, we directly follow Neural Haircut [5] in applying it. To utilize diffusion-based prior at the level of strand-based hair maps H, we use a pre-trained encoder $\mathcal{E}$ that maps a hair map H into its latent version Z. Specifically, we randomly sample a set of 1,000 guiding strands and map them into a regular grid of origins using an interpolation method described above. As a result, we produce a guiding hair map H', which we can encode into a guiding latent hair map Z'. This map is then directly fed into a diffusion-based prior in order to calculate an SDS loss.

1.4 Postprocessing

After training, to achieve simulatable hairstyles, we post-process the resulting strands to ensure they have no intersections with the FLAME mesh. To achieve that, we calculate a signed distance for all the points on the strands towards the head mesh. Then, for the strands that intersect the mesh, we prune the intersection regions and then connect the beginning of each of the consecutive strand segment to the nearest vertex on the scalp.

2 Experiments

Real-world evaluation. We extend the comparison with Neural Haircut [5] in Fig. 1. We also provide the reconstruction results for additional scenes in Fig. 2-4. An extra comparison with the one-shot hair reconstruction method named HairStep [9] is provided in Fig. 5. We also expand the ablation study of the strands fitting in Figs 7-8. Besides the experiments presented in the main paper, we provide the close-ups and three more experiments: we remove color rendering loss $\mathcal{L}_{\text{rgb}}$, orientation rendering loss $\mathcal{L}_{\text{dir}}$ and diffusion-based loss $\mathcal{L}_{\text{sds}}$. We observe that each of the components of our method meaningfully contribute to the final performance, as illustrated on the close-up images.

Limitations. The main limitation of our method follows Neural Haircut [5] and is related to the modeling of curly hairstyles, see Fig. 6. We attribute this behavior to the root-to-tip design of the strand-based prior. In our work, we use the same variant of the prior as the one proposed in Neural Haircut. Improving the performance for the curly hairstyles thus remains future work.

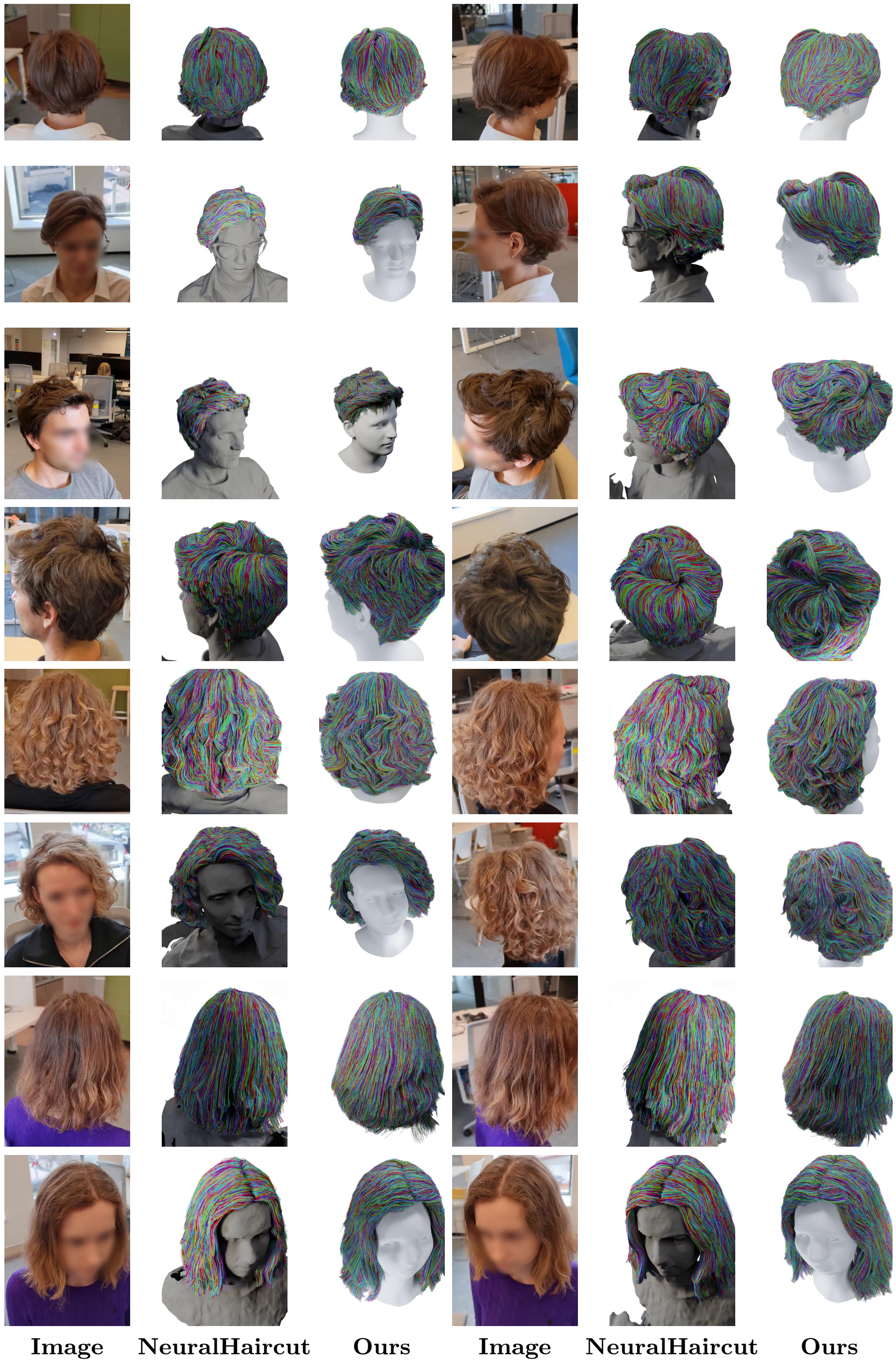

Fig. 1: Additional qualitative comparison of the reconstructed strand-based geometry against Neural Haircut [5]. Slight misalignment appears due to test-view cameras interpolation.

Fig. 2: Qualitative results of our method on additional scenes. Slight misalignment appears due to test-view cameras interpolation.

Image **Ours** **Image** **Ours**

Fig. 3: Qualitative results of our method on additional scenes. Slight misalignment appears due to test-view cameras interpolation.

Fig. 4: Qualitative results of our method on additional scenes. Slight misalignment appears due to test-view cameras interpolation.

Reference **HairStep** **Ours**

Fig. 5: Comparison with the HairStep one-shot hair reconstruction method. We use a frontal image to infer the HairStep result.

GT Reconstructions GT Reconstructions

Fig. 6: The limitations of our method include the problem of curly hairstyles modeling.

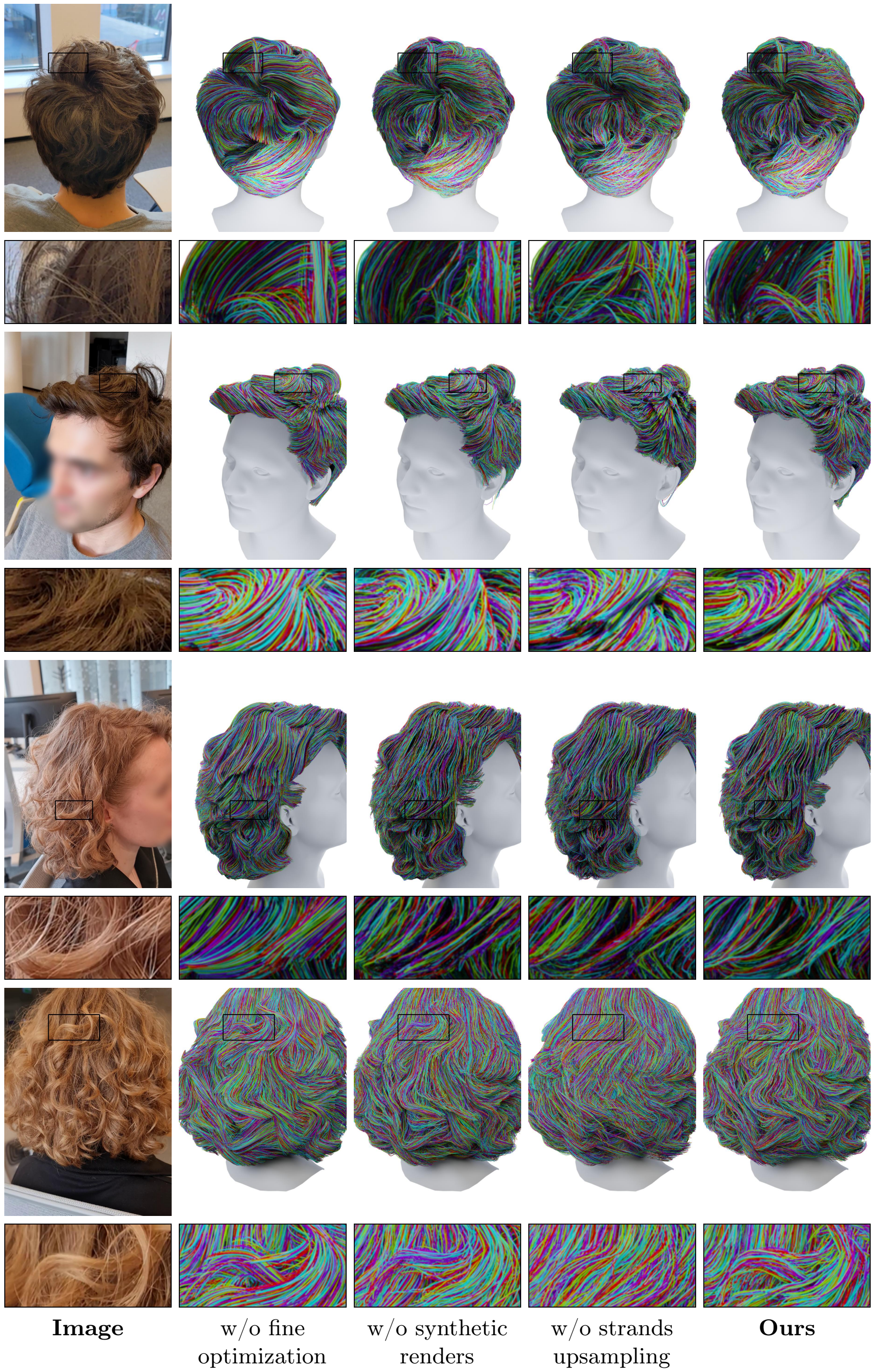

Fig. 7: Extended ablation study of the strands fitting stage (Part 1).

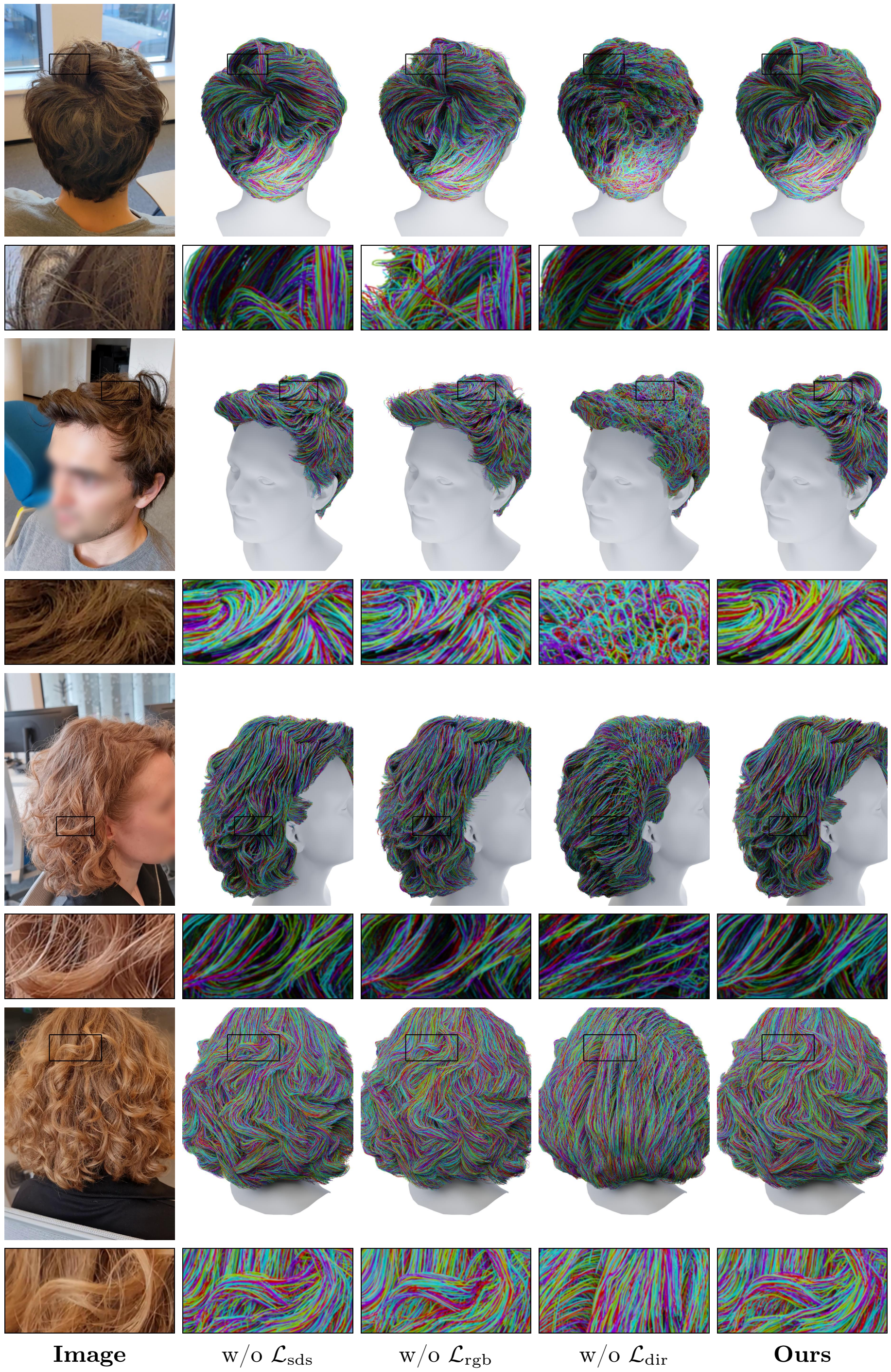

Fig. 8: Extended ablation study of the strands fitting stage (Part 2).

References

1. Paris, S., Briceño, H.M., Sillion, F.X.: Capture of hair geometry from multiple images. ACM SIGGRAPH 2004 Papers (2004) 1
2. Poole, B., Jain, A., Barron, J.T., Mildenhall, B.: Dreamfusion: Text-to-3d using 2d diffusion. ICLR **abs/2209.14988** (2023) 2
3. Rosu, R.A., Saito, S., Wang, Z., Wu, C., Behnke, S., Nam, G.: Neural strands: Learning hair geometry and appearance from multi-view images. In: European Conference on Computer Vision (2022) 2
4. Schönberger, J.L., Frahm, J.M.: Structure-from-motion revisited. In: Conference on Computer Vision and Pattern Recognition (CVPR) (2016) 1
5. Sklyarova, V., Chelishev, J., Dogaru, A., Medvedev, I., Lempitsky, V., Zakharov, E.: Neural haircut: Prior-guided strand-based hair reconstruction. In: Proceedings of IEEE International Conference on Computer Vision (ICCV) (2023) 1, 2, 3
6. Sklyarova, V., Zakharov, E., Hilliges, O., Black, M.J., Thies, J.: Haar: Text-conditioned generative model of 3d strand-based human hairstyles. ArXiv (Dec 2023) 1
7. Su, S., Yan, Q., Zhu, Y., Zhang, C., Ge, X., Sun, J., Zhang, Y.: Blindly assess image quality in the wild guided by a self-adaptive hyper network. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) pp. 3664–3673 (2020) 1
8. Yao, J., Wang, X., Ye, L., Liu, W.: Matte anything: Interactive natural image matting with segment anything models. arXiv preprint arXiv:2306.04121 (2023) 1
9. Zheng, Y., Jin, Z., Li, M., Huang, H., Ma, C., Cui, S., Han, X.: Hairstep: Transfer synthetic to real using strand and depth maps for single-view 3d hair modeling. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 12726–12735 (2023) 2