

DiffLocks: Generating 3D Hair from a Single Image using Diffusion Models

Supplementary Material

1. Implementation details.

1.1. Strand VAE

We train the strandVAE using a batch size of 200 strands, each consisting of 256 points. For the loss we set the directional weight term to $\lambda_1 = 2e-3$, the curvature weight term to $\lambda_2 = 7.8e-2$ and the KL term to $\lambda_{KL} = 6e-4$. We train the StrandVAE for a total of 3M iterations using AdamW at a learning rate of $3e-3$ with cosine decay schedule.

1.2. Diffusion model

We follow similar optimizer parameters as HDiT [1] and train the diffusion model for $\approx 400K$ iterations at an effective batch size of 128 using a constant learning rate of $5e-4$.

2. Dataset.

The synthetic hair dataset was created by launching 12 Blender processes in parallel, each creating a chunk of the 40K samples. The process is particularly CPU-intensive since the Blender geometry nodes don't tend to take advantage of GPU acceleration. As a result, the generation process took approximately 1 week on a cluster with 4 H100 GPUs and a Intel Xeon Platinum 8480+ CPU (56 Cores).

3. Discussions

3.1. Density Map

Compared to GroomGen's binary map, we treat the proposed density map as an alternative hair representation with different properties. GroomGen's binary mask marks exact scalp pixels where strands grow, while our focus is on local hair density rather than precise root positions. This makes the density map smoother and easier to edit—since strand latent codes exist for every pixel, we can modify hair density by simply sampling more strands (Fig. 1).

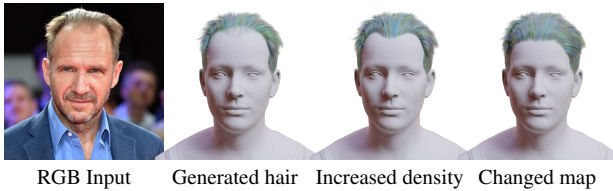


Figure 1. The density map allows the density of hair or hairline to be easily controlled by directly modifying its values.

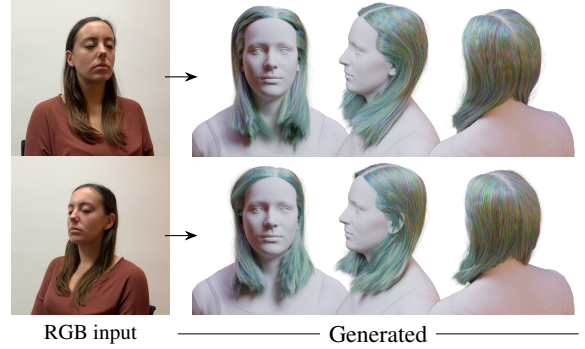


Figure 2. **Robustness.** We train our method using synthetic images with diverse camera positions and focal lengths. Testing on real images, we observe that the method is robust and consistently generates the same hairstyle regardless of changes in viewpoint.

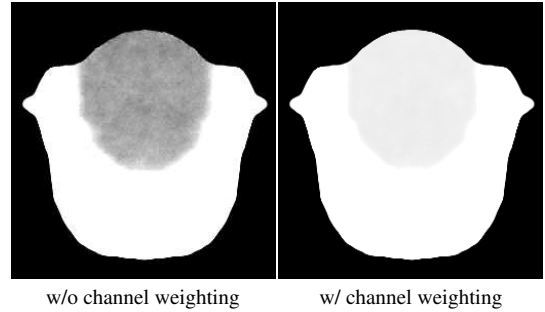


Figure 3. **Ablation channel weighting.** Without channel weighting, the diffusion model tends to generate noisy density maps. Applying our proposed weighing, the network focuses on the important information from the scalp textures and allows it to create smoother density maps.

4. Additional Evaluation and Results

4.1. Additional studies

Camera pose robustness. To further demonstrate the robustness of our method, we evaluate it on images captured from varying camera poses from the H3DS dataset [4]. As shown in Fig. 2, our approach can reconstruct consistent hairstyles despite large changes in camera position and distance to the subject.

Ablation on curvature loss. To evaluate the effectiveness of curvature loss while training the strandVAE, we perform an ablation study as shown in Fig. 4 in which we encode the ground-truth strand and decode it with models trained with

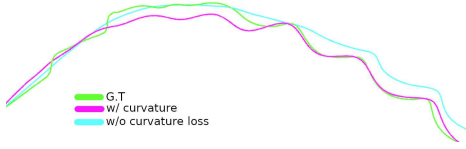


Figure 4. **StrandVAE ablation.** We observe that without the curvature loss, the predicted strand (blue) tends to be smoother than the ground-truth (green). However, with the curvature loss (purple), the curly pattern is more accurately recovered, improving the visual quality of the whole hairstyle.

and without the curvature loss. We observe that the curvature loss is crucial in encouraging the network to reconstruct curly and wavy hair.

Ablation on weighting scheme. We also ablate our proposed per-channel weighting scheme in Fig. 3. It is clear that without the weighting scheme, the output of the diffusion model exhibits more noise which is especially evident when viewing the density map. Our proposed weighting scheme allows the network to focus on the important information stored in the latent code of the strands and ignore noisy dimensions.

4.2. Generalization ability

We test reconstructing a hairstyle outside the training set and find the network generalizes well (Fig. 5). The significant gap between the generated hair and the top-3 samples from the training set confirms that our method generates new hairstyles rather than merely retrieving from training data.



Figure 5. DiffLocks generalizes beyond the training set: Generated hair and the closest top3 hairstyles from the training set.

4.3. DINOv2 vs orientation map

We ran an experiment where we replace DINO features with an orientation map together with hair segmentation mask. We find that DINO features are more robust especially for short or dark hair (Fig. 6) where the orientation map can be noisy.

4.4. Extended comparison with baselines

We provide extended quantitative comparison and qualitative comparison with HairStep [9] and NeuralHDHair [6] on the DiffLocks evaluation dataset and Yuksel dataset [8]. We



Figure 6. DINOv2 features are a more robust and richer conditioning signal than orientation maps.

note that the DiffLocks evaluation set is created using our proposed Blender pipeline and contains samples that were not used during training.

Qualitative. The qualitative comparison with HairStep and NeuralHDHair is shown in Fig. 7. Their method performs well on straight hairstyles but struggle with curly and wavy ones, especially when their predicted 2D orientation map [3] represent an incorrect parting or growth direction. The reliance on intermediate representations like orientations maps is a long-standing limitation of hair reconstruction. In contrast, our method effectively reconstructs both straight and curly hairstyles, remaining unaffected by 2D orientation ambiguities, as it directly utilizes RGB images as input.

An additional limitation of previous methods becomes evident when viewing the backside of the head where the baselines methods tend to create balding areas or overly smooth strands. In contrast, our approach, powered by robust data priors, generates more reasonable and realistic results for occluded or invisible regions.

Quantitative. We provide extended quantitative comparisons of HairStep and NeuralHDHair on both DiffLocks evaluation dataset and Yuksel dataset [8]. We calculate precision, recall and F-score using 3D ground-truth strands similar to previous methods [5, 7] as shown in Tab. 1 and Fig. 8. We provide per-example results to complement the aggregate quantitative metrics presented in the main paper. Since HairStep and NeuralHDHair are trained primarily on frontal views, and their training hairstyles lack diversity, they tend to perform poorly on the DiffLocks dataset which contains a range of hairstyles (curly, balding, combed-back, and afro-like) together with images that deviate slightly from the frontal view.

Considering that the baselines models were trained on the USC-HairSalon dataset [2], while our training data aligns more closely with the distribution of the evaluation dataset (rendering manner and strands distribution on scalp), our results will be better aligned with the distribution of ground truth, leading to higher evaluation metrics. Thus, we also perform a quantitative evaluation on Yuksel synthetic dataset [8]. We show the metrics for

the different hairstyles separately for a more comprehensive analysis. HairStep and NeuralHdHair both achieved high F-score in straight hairstyle, but for curly hairstyle, their method struggle to reconstruct it accurately and both precision, recall and F-score are greatly reduced. In contract, our method still performs well on curly hairstyle, recovering plausible geometry even on the backside of the head.

4.5. Additional in-the-wild results

Lastly, to evaluate the robustness and effectiveness of our method, we provide additional reconstructions from in-the-wild single images. As shown in Fig. 9, our method can robustly reconstruct a large variety of hairstyles, achieving high-quality and realistic results.

References

- [1] Katherine Crowson, Stefan Andreas Baumann, Alex Birch, Tanishq Mathew Abraham, Daniel Z Kaplan, and Enrico Shippole. Scalable high-resolution pixel-space image synthesis with hourglass diffusion transformers. In *Forty-first International Conference on Machine Learning*, 2024. 1
- [2] Liwen Hu, Chongyang Ma, Linjie Luo, and Hao Li. Single-view hair modeling using a hairstyle database. *ACM Transactions on Graphics (ToG)*, 34(4):1–9, 2015. 2
- [3] Sylvain Paris, Hector M Briceno, and François X Sillion. Capture of hair geometry from multiple images. *ACM transactions on graphics (TOG)*, 23(3):712–719, 2004. 2
- [4] Eduard Ramon, Gil Triginer, Janna Escur, Albert Pumarola, Jaime Garcia, Xavier Giro-i Nieto, and Francesc Moreno-Noguer. H3d-net: Few-shot high-fidelity 3d head reconstruction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5620–5629, 2021. 1
- [5] Vanessa Sklyarova, Jenya Chelishev, Andreea Dogaru, Igor Medvedev, Victor Lempitsky, and Egor Zakharov. Neural haircut: Prior-guided strand-based hair reconstruction. In *ICCV*, 2023. 2
- [6] Keyu Wu, Yifan Ye, Lingchen Yang, Hongbo Fu, Kun Zhou, and Youyi Zheng. Neuralhdhair: Automatic high-fidelity hair modeling from a single image using implicit neural representations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1526–1535, 2022. 2, 4, 5
- [7] Keyu Wu, Lingchen Yang, Zhiyi Kuang, Yao Feng, Xutao Han, Yuefan Shen, Hongbo Fu, Kun Zhou, and Youyi Zheng. Monohair: High-fidelity hair modeling from a monocular video. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 24164–24173, 2024. 2
- [8] Cem Yuksel, Scott Schaefer, and John Keyser. Hair meshes. *ACM Transactions on Graphics (TOG)*, 28(5):1–7, 2009. 2, 4
- [9] Yujian Zheng, Zirong Jin, Moran Li, Haibin Huang, Chongyang Ma, Shuguang Cui, and Xiaoguang Han. 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*, pages 12726–12735, 2023. 2, 4, 5

Method	Straight									Curly								
	2/20	3/30	4/40	2/20	3/30	4/40	2/20	3/30	4/40	2/20	3/30	4/40	2/20	3/30	4/40			
	Precision			Recall			F-score			Precision			Recall			F-score		
NeuralHDHair [6]	72.58	86.85	92.07	45.28	63.57	74.55	55.77	73.41	82.39	23.24	43.29	55.45	17.92	38.65	54.73	20.23	40.84	55.09
HairStep [9]	64.05	78.95	84.52	56.22	73.37	82.17	59.88	76.06	83.33	17.71	33.02	43.67	8.11	16.55	24.14	11.13	22.05	31.08
Ours	57.38	76.15	84.49	38.54	54.84	65.74	46.11	63.77	73.94	29.83	60.46	79.09	31.41	61.69	77.84	30.60	61.07	78.46

Table 1. Quantitative comparison with [6, 9] on Yuksel dataset [8]. We evaluate our method on straight and curly hair separately. Our method achieves superior results on curly hair. Our method performs slightly worse on straight hair due to the diffusion model, which introduces perturbations to enhance the realism of the generated hairstyles.

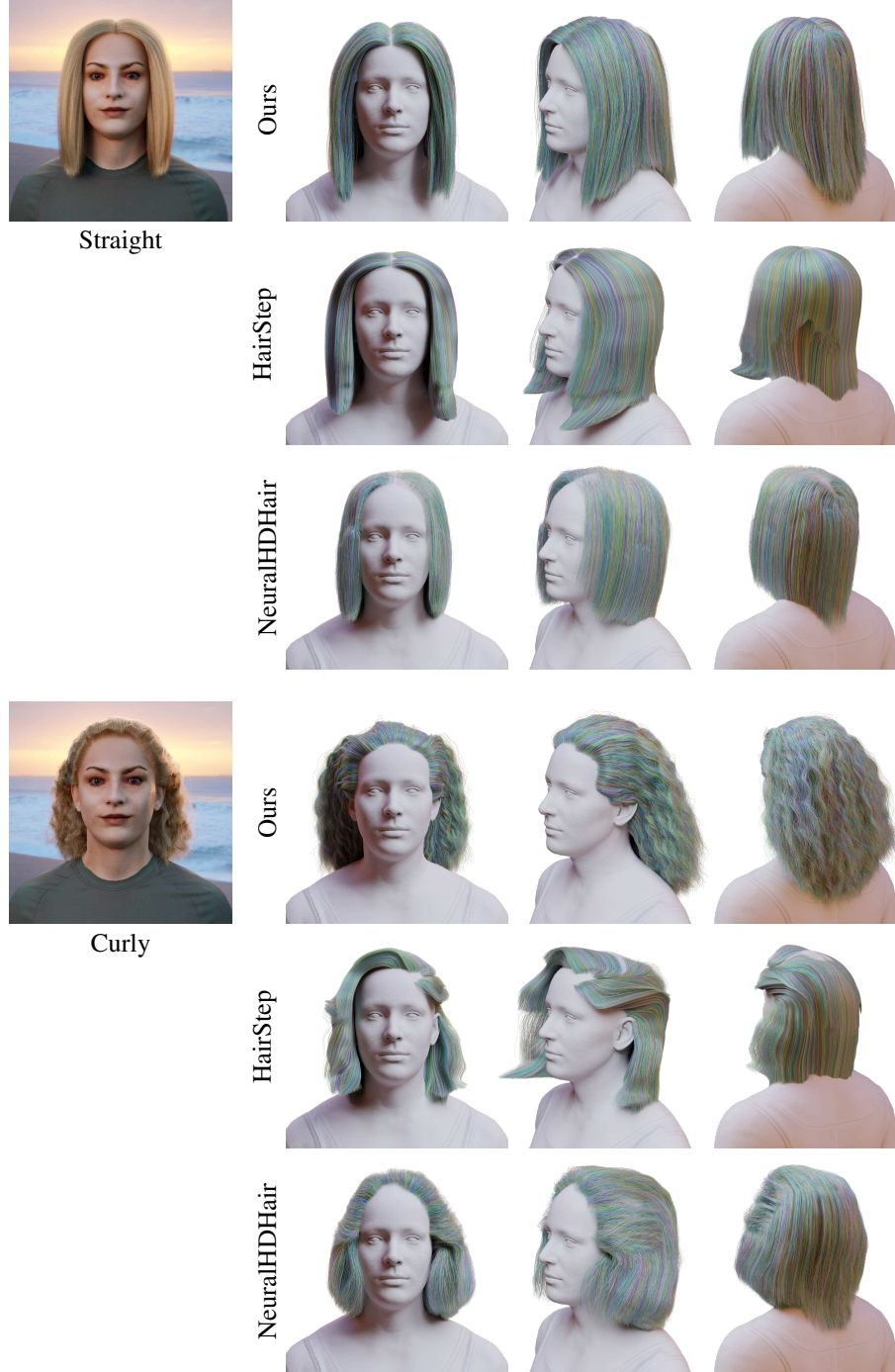


Figure 7. Results on the synthetic dataset of [8]



Method	Thresholds: mm / degrees								
	2/20	3/30	4/40	2/20	3/30	4/40	2/20	3/30	4/40
	Precision (↑)			Recall (↑)			F-score (↑)		
NeuralHDSHair [6]	27.57	47.30	57.12	22.67	44.90	63.06	24.88	46.07	59.94
HairStep [9]	32.10	52.78	65.90	20.04	35.55	47.71	24.67	42.48	55.35
Ours	49.93	73.46	84.42	52.20	76.88	88.90	51.04	75.13	86.60

Method	Thresholds: mm / degrees								
	2/20	3/30	4/40	2/20	3/30	4/40	2/20	3/30	4/40
	Precision (↑)			Recall (↑)			F-score (↑)		
NeuralHDSHair [6]	38.30	54.20	61.76	54.50	80.77	89.65	44.99	64.87	73.14
HairStep [9]	33.40	54.09	63.08	36.42	60.63	71.23	34.84	57.17	66.91
Ours	89.68	96.70	98.24	86.47	93.82	96.44	88.05	95.24	97.33

Method	Thresholds: mm / degrees								
	2/20	3/30	4/40	2/20	3/30	4/40	2/20	3/30	4/40
	Precision (↑)			Recall (↑)			F-score (↑)		
NeuralHDSHair [6]	31.27	62.32	80.14	32.20	60.04	77.37	31.73	61.16	78.73
HairStep [9]	41.17	61.83	71.11	38.43	66.04	81.06	39.76	63.87	75.76
Ours	63.71	84.26	92.67	62.97	84.77	93.34	63.34	84.51	93.00

Method	Thresholds: mm / degrees								
	2/20	3/30	4/40	2/20	3/30	4/40	2/20	3/30	4/40
	Precision (↑)			Recall (↑)			F-score (↑)		
NeuralHDSHair [6]	16.75	43.33	62.75	9.33	22.73	37.59	11.99	29.82	47.01
HairStep [9]	26.50	60.19	79.40	10.14	23.01	37.34	14.67	33.29	50.80
Ours	29.53	62.11	81.67	34.39	69.56	87.98	31.77	65.62	84.71

Method	Thresholds: mm / degrees								
	2/20	3/30	4/40	2/20	3/30	4/40	2/20	3/30	4/40
	Precision (↑)			Recall (↑)			F-score (↑)		
NeuralHDSHair [6]	16.83	31.60	43.07	20.15	42.45	62.18	18.34	36.23	50.89
HairStep [9]	28.87	42.79	50.31	38.85	64.23	78.16	33.13	51.37	61.21
Ours	85.73	96.97	99.05	84.34	96.37	98.70	85.03	96.67	98.87

Method	Thresholds: mm / degrees								
	2/20	3/30	4/40	2/20	3/30	4/40	2/20	3/30	4/40
	Precision (↑)			Recall (↑)			F-score (↑)		
NeuralHDSHair [6]	10.60	31.98	52.94	5.45	17.77	32.98	7.19	22.84	40.64
HairStep [9]	6.56	21.95	40.96	3.83	10.15	19.50	4.84	13.88	26.42
Ours	38.33	71.01	88.46	34.37	65.80	83.83	36.24	68.30	86.08

Method	Thresholds: mm / degrees								
	2/20	3/30	4/40	2/20	3/30	4/40	2/20	3/30	4/40
	Precision (↑)			Recall (↑)			F-score (↑)		
NeuralHDSHair [6]	9.90	24.95	40.50	10.70	24.12	38.74	10.29	24.53	39.60
HairStep [9]	9.78	18.98	26.92	14.08	29.42	45.88	11.54	23.07	33.93
Ours	54.17	81.83	92.75	51.87	78.94	90.60	53.00	80.36	91.67

Method	Thresholds: mm / degrees								
	2/20	3/30	4/40	2/20	3/30	4/40	2/20	3/30	4/40
	Precision (↑)			Recall (↑)			F-score (↑)		
NeuralHDSHair [6]	17.64	31.10	41.55	21.04	43.40	59.30	19.19	36.24	48.87
HairStep [9]	11.90	18.39	24.29	17.26	33.64	50.13	14.09	23.78	32.7
Ours	85.82	97.15	85.82	73.10	86.27	90.94	78.95	91.39	94.85

Method	Thresholds: mm / degrees								
	2/20	3/30	4/40	2/20	3/30	4/40	2/20	3/30	4/40
	Precision (↑)			Recall (↑)			F-score (↑)		
NeuralHDSHair [6]	11.34	21.98	34.32	8.76	19.74	27.98	9.89	20.80	30.82
HairStep [9]	7.69	15.41	21.67	21.96	44.94	65.67	11.39	22.95	32.58
Ours	70.03	91.91	97.29	55.16	77.37	87.19	61.71	84.02	91.96

Method	Thresholds: mm / degrees								
	2/20	3/30	4/40	2/20	3/30	4/40	2/20	3/30	4/40
	Precision (↑)			Recall (↑)			F-score (↑)		
NeuralHDSHair [6]	6.38	13.38	20.47	7.85	17.39	27.23	7.04	15.12	23.37
HairStep [9]	4.29	9.02	13.72	7.02	17.73	31.59	5.32	11.96	19.13
Ours	63.73	86.23	94.16	44.22	67.34	80.17	52.21	75.63	86.60

Figure 8. **Quantitative comparison.** We provide the quantitative comparison for each example in DiffLocks evaluation set. When the hairstyle is curly or balding or the image is not in frontal view, our method achieves significant improvement.



RGB input

Front

Side

Back

Figure 9. More results of in-the-wild reconstruction of hairstyles.