PHOTOREALISTIC DYNAMIC HAIR RECONSTRUCTION WITH STRAND GAUSSIANS

Anonymous authors
Paper under double-blind review

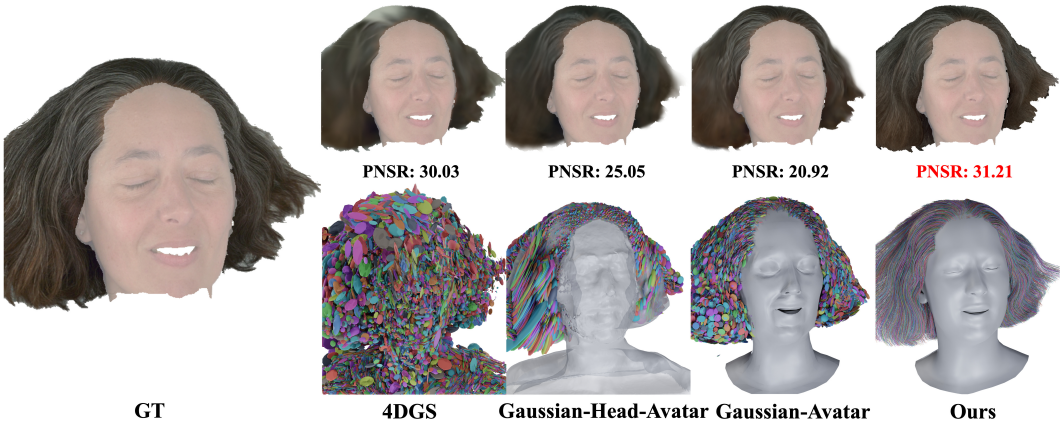

Figure 1: We propose a dynamic hair reconstruction method based on Strand Gaussians that explicitly models time-varying hair geometry and appearance from multi-view videos. By constraining all hair Gaussians within the segmented hair region, our method preserves the strand structure and achieves high photometric fidelity. Since the Gaussians in Gaussian Head Avatar (GHA) are not guaranteed to lie outside the underlying head mesh, we render the mesh with partial transparency for visualization.

ABSTRACT

Reconstructing dynamic human hair from multi-view videos is essential for creating high-quality 3D assets for digital avatars and animation. Existing methods augment static hair reconstruction with learned hair dynamics, facing challenges in balancing frame-wise photometric fidelity and optimization efficiency. We present a Strand Gaussian based framework for reconstructing time-varying hair geometry while maintaining photometric fidelity for the observed sequence and preserving photorealistic appearance. Qualitative and quantitative evaluations on the public NeRSemble dataset suggests that our method achieves reconstruction quality that outperforms state-of-the-art methods by a significant margin.

1 INTRODUCTION

Reconstructing dynamic human hair is a core task for acquiring high-quality and natural hair assets, which is crucial for modeling 3D talking heads or human avatars. Traditional methods model human hairs with geometric primitives (Paris et al., 2004; Wei et al., 2005; Nam et al., 2019), which admit limited rendering quality. More recently, researchers propose using unstructured neural representations (e.g., NeRF and gaussian splattings), which ignores the inherent strand structure of human hair, making it difficult to animate the reconstructed human hair. In light of this, multiple research works (Luo et al., 2024; Zakharov et al., 2024) propose using strand gaussians to represent human hair, which achieves superior rendering quality while maintaining the strand structure.

Dynamic hair reconstruction from multi-view videos must preserve coherent motion and frame-wise detail. GaussianAvatars (Qian et al., 2024) binds Gaussians to a deforming face mesh that drives

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

their motion, so hair largely follows the head rather than deforming independently. DynHair (Sklyarova et al., 2026) reconstructs a canonical hairstyle from the first frame, then learns a model for subsequent motion. This captures non-rigid motion, but training an additional dynamics model is costly and complicates the balance between fidelity and efficiency. This motivates directly fitting strand geometry and appearance.

In this work, we propose a dynamic hair reconstruction framework based on strand Gaussians. It consists of a strand generation module and a refinement module, optimized in a coarse-to-fine manner. The former directly models time-varying hair geometry and appearance from multi-view observations, while the latter refines color details on the generated strands to improve rendering fidelity.

The strand generation module shares a scalp UV-time latent representation across the sequence, allowing multiple frames to be optimized jointly in parallel rather than fitting hair frame by frame. Its features are interpolated across the scalp and decoded into time-varying strand geometry and appearance, with spatial and temporal regularization encouraging coherent motion.

The refinement module predicts time-varying color residuals for individual Strand Gaussians in a strand-aligned space defined by scalp UV location and position along each strand. Compared with Cartesian coordinates, this alignment captures finer appearance variations along the hair and restores local details that may be smoothed out during strand interpolation. Experiments on NeRSemble demonstrate clearer strand-level details and improved photometric and perceptual fidelity over the evaluated baselines.

To sum up, our contributions are:

- We propose a dynamic human hair reconstruction method based on strand Gaussians that directly models time-varying hair geometry and appearance in latent space.
- We introduce a coarse-to-fine reconstruction strategy that first establishes the global strand structure and motion and then refines local appearance to improve hair fidelity.
- Our method achieves better geometric reconstruction and appearance quality than existing state-of-the-art methods.

2 Related works

2.1 Static hair reconstruction

Hair modeling has been recognized as a challenge in human avatar reconstruction, with representations including geometric primitives, neural/volumetric representation. Early approaches (Paris et al., 2004; Wei et al., 2005; Nam et al., 2019) represent hair using line segments reconstructed from multiple views. Additionally, Dr.Hair (Takimoto et al., 2024) uses guiding hair and differentiable rendering to reconstruct internal hair flow. However, hair reconstruction with geometric representations remains challenging when dealing with various hairstyles. Video-based reconstruction (Liang et al., 2018) retrieves and fits a static hairstyle from a database, while a 3D morphable model (Blanz & Vetter, 1999) estimates the face shape.

Neural approaches (Kuang et al., 2022; Wu et al., 2022) use implicit fields to infer hair geometry and generate strands. Neural Strands (Rosu et al., 2022) encodes strand geometry and appearance in neural scalp textures, combining explicit strand decoding with neural rendering. Neural Haircut (Sklyarova et al., 2023) combines volumetric and strand-based reconstruction. 3DGS (Kerbl et al., 2023) provides an explicit Gaussian representation for radiance fields. Building on this representation, Gaussian Haircut (Zakharov et al., 2024) models hair with strand-aligned Gaussians, while GaussianHair (Luo et al., 2024) uses chains of cylindrical Gaussian primitives. However, the complexity of hair motion makes it challenging to directly transfer static models to dynamic scenes.

2.2 Dynamic hair reconstruction

Dynamic hair reconstruction recovers time-varying hair geometry and appearance from videos. Xu et al. (2014) extract motion paths from spacetime video slices and recover temporally coherent strands through spacetime optimization. For facial hair, Winberg et al. (2022) jointly track dense 3D facial hair and the underlying skin throughout facial performances. HVH (Wang et al., 2022)

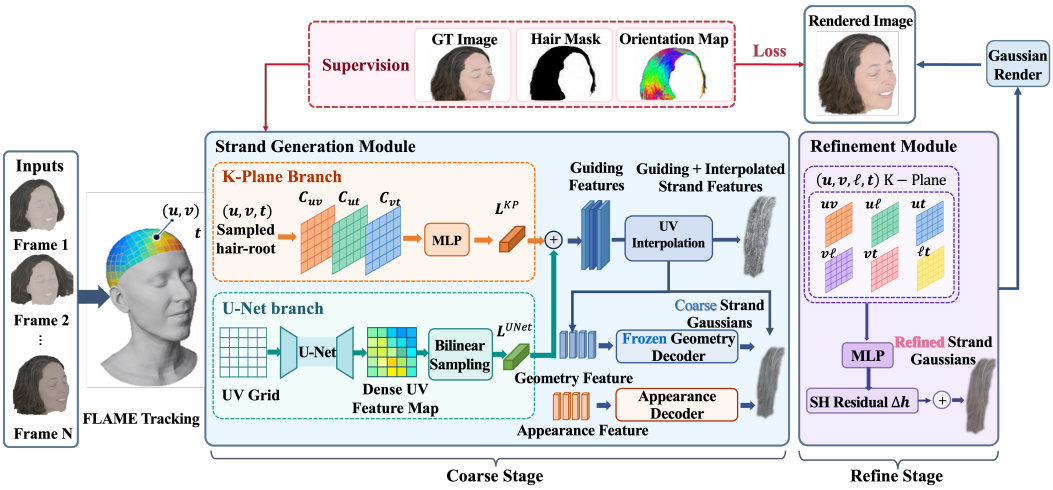

Figure 2: Our dynamic hair reconstruction method takes multi-view videos and models dynamic hair with Strand Gaussians rooted on the FLAME scalp. A strand generation module produces and interpolates guiding strands, followed by a refinement module that enhances their appearance to improve rendering quality.

combines tracked guide strands with volumetric primitives and multi-view optical-flow supervision. NeuWigs (Wang et al., 2023) learns a VAE-based hair state representation and a separate dynamics model for hair animation. MeGA (Wang et al., 2025) represents hair with Gaussians and an MLP deformation field, while GaussianAvatars (Qian et al., 2024) binds Gaussians to a deforming FLAME mesh. However, unstructured Gaussians do not explicitly encode strand connectivity. Recent methods integrate explicit hair strands into Gaussian avatars. HADES (Liao et al., 2025) binds 3D Gaussians to scalp-rooted hair strands and predicts strand deformations from head motion and historical hair states, enabling dynamic hair animation within full-body avatars. PhysHead (Kabadayi et al., 2026) attaches Gaussians to a parametric head mesh and explicit strands, producing simulation-ready avatars with physics-driven hair motion. DynHair (Sklyarova et al., 2026) instead learns temporal strand deformations conditioned on head-motion history and gravity. Our method directly reconstructs time-varying strand geometry and appearance through a shared temporal latent representation.

3 PRELIMINARY

In this section, we introduce the problem setting of dynamic hair reconstruction; the strand gaussian representation and its rendering equation; and the data preparation pipeline.

Problem Setting. Given multi-view videos $\mathcal{F} = \{F_1, ..., F_M\}$ from M views, where each $F_i \in \mathbb{R}^{T\times H\times W\times C}$ is a stack of T image frames from a single-view video, H, W, C are height, width and RGB channels, we aim to reconstruct the geometry and appearance of human hair. Following the strand gaussian representation from Luo et al. (2024), we denote the geometry and appearance of each hair strand as G_i and A_i, respectively. The geometry of a hair strand is represented by a series of gaussians $G_i = \{\boldsymbol{g}_{i,k} | k = 1, \ldots, K\}$, where $\boldsymbol{g}_{i,k} : \mathbb{R}^3 \to \mathbb{R}^+$ is a spatial density function:

$$\boldsymbol{g}_{i,k}(\boldsymbol{x}) = e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}, \tag{1}$$

where $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are the center position and covariance matrix. The covariance matrix $\boldsymbol{\Sigma} = \boldsymbol{R}\boldsymbol{S}\boldsymbol{S}^T\boldsymbol{R}^T$ is derived from the rotation matrix $\boldsymbol{R}$ and diagonal scaling matrix $\boldsymbol{S} = \text{diag}(d, d, s)$, where $d = 10^{-4}$ represents the hair radius and s represents the hair length. To satisfy the geometric constraint so that strand Gaussians are connected, the strand Gaussians satisfy the constraints:

$$\mu_i + \frac{1}{2}s_i\boldsymbol{d}_i = \mu_{i+1} - \frac{1}{2}s_{i+1}\boldsymbol{d}_{i+1}, \tag{2}$$

where $\boldsymbol{d}_i$ is the third column of $\boldsymbol{R}$, encoding the hair orientation direction. For convenience, we define the endpoints of gaussians $\boldsymbol{p}_{i+1} = \boldsymbol{\mu}_i + \frac{1}{2}s_i\boldsymbol{d}_i$. From the root of the hair on the scalp $\boldsymbol{p}_0$, we

can accumulate each strand gaussian to generate the hair strand. The appearance vector A_i includes the opacity $\alpha_i \in \mathbb{R}^+$ and view-dependent RGB color encoded with degree-1 spherical harmonics (SH), $\boldsymbol{h}_i \in \mathbb{R}^{12}$.

Rendering Strand Gaussian. Following the rendering equation of 3DGS (Kerbl et al., 2023), we project the gaussian onto image space. The 2D covariance matrix can be derived by

$$\boldsymbol{\Sigma}' = \boldsymbol{J}\boldsymbol{W}\boldsymbol{\Sigma}\boldsymbol{W}^T\boldsymbol{J}^T, \tag{3}$$

where $\boldsymbol{J}$ is the Jacobian matrix of the affine approximation of the projective transformation, and $\boldsymbol{W}$ is the view transformation matrix. The rendered color of a camera pixel is accumulated as

$$\boldsymbol{C} = \sum_i T_i \alpha_i' \boldsymbol{c}_i, \qquad T_i = \prod_{j=1}^{i-1} \left(1 - \alpha_j'\right), \tag{4}$$

where $\boldsymbol{c}_i \in \mathbb{R}^3$ is the learned color. α_i' is the final opacity which denoted as:

$$\alpha_i' = \alpha_i \cdot \left(-\frac{1}{2} \left(\boldsymbol{x}_i' - \boldsymbol{\mu}_i'\right)^T \boldsymbol{\Sigma}_i'^{-1} \left(\boldsymbol{x}_i' - \boldsymbol{\mu}_i'\right) \right), \tag{5}$$

where $\boldsymbol{x}_i'$ and $\boldsymbol{\mu}_i'$ are projected and center position in 2D space, and α_i is the opacity.

Hair Perception. Dynamic hair reconstruction requires additional perception information to ensure an accurate reconstruction of strand orientations. Therefore, we adopt perception model (Yao et al., 2024) to segment the hair regions for each video frame, resulting in a binary mask $M_h \in [0,1]^{H \times W}$. We also adopt the Gabor filter method from Zakharov et al. (2024) to compute pixel-wise hair orientation and confidence, denoted as $O_d \in \mathbb{R}^{H \times W \times 2}$ and $O_v \in \mathbb{R}^{H \times W}$, respectively.

Dynamic Head Reconstruction. To disentangle the head motion and hair motion, we adopt FLAME (Li et al., 2017) to track the head poses over each time steps, resulting in a template mesh M_t with predefined scalp region at each time step. During the reconstruction procedure, we fix the head pose and grow hair strands from root points on the scalp to ensure an accurate hair reconstruction.

4 Approach

In this section, we first introduce in Sec. 4.1 a novel dynamic hair reconstruction algorithm via strand Gaussians, which is adapted from state-of-the-art static hair reconstruction methods. We then describe its coarse and refinement optimization stages in Sec. 4.2.

4.1 Dynamic hair reconstruction via strand gaussians

Inspired by the graphics industry to reduce the number of strands, we explicitly model a set of coarse-grained optimizeable guiding hairs and densify them via interpolation. This strategy imposes a smoothness prior on the reconstructed hair. Our hair model consists of two components: 1) the strand generation module and 2) the refinement module. We will elaborate each of the modules and describe the optimization procedure of dynamic hair reconstruction.

Strand Generation Module. We initialize a large set of hair strands whose root coordinates $\mathcal{H} = \{(u_i, v_i)\}$ are randomly distributed over the valid scalp UV domain. During each optimization iteration, we randomly sample a subset $\mathcal{B} \subset \mathcal{H}$ and compute its features using two complementary branches.

For the first branch, we retain the K-Plane (Fridovich-Keil et al., 2023) representation over the scalp UV space and T time steps. Instead of constructing a dense 4D tensor of size $|U| \times |V| \times T \times D$, we directly query the learned pairwise feature planes at each sampled coordinate $(u_i, v_i) \in \mathcal{B}$ and time t. The bilinearly interpolated plane features are combined as

$$C_i(t) = C_{uv}(u_i, v_i) \circ C_{vt}(v_i, t) \circ C_{ut}(u_i, t) \in \mathbb{R}^{D_C}, \tag{6}$$

where $\circ$ is element-wise multiplication, $C_{uv} \in \mathbb{R}^{|U| \times |V| \times D_C}$, $C_{vt} \in \mathbb{R}^{|V| \times T \times D_C}$, and $C_{ut} \in \mathbb{R}^{|U| \times T \times D_C}$. An MLP then directly maps each queried K-Plane feature to one feature vector for the corresponding sampled strand:

$$L_i^{\mathrm{KP}}(t) = f_{\mathrm{MLP}}(C_i(t)) \in \mathbb{R}^{D_L}, \qquad (u_i, v_i) \in \mathcal{B}. \tag{7}$$

For the second branch, we define a standard regular UV coordinate grid $P_{\mathrm{UV}} \in \mathbb{R}^{H_{\mathrm{UV}} \times W_{\mathrm{UV}} \times 2}$, whose two channels store the (u, v) coordinates of its grid locations. Inspired by the U-Net-based designs of Zakharov et al. (2024); Rosu et al. (2022), we process this coordinate grid with a U-Net to produce a dense feature map

$$F_{\mathrm{UNet}} = f_{\mathrm{UNet}}(P_{\mathrm{UV}}) \in \mathbb{R}^{H_{\mathrm{UV}} \times W_{\mathrm{UV}} \times D_L}. \tag{8}$$

For each sampled hair root, we bilinearly interpolate this U-Net feature map at its continuous UV coordinate:

$$L_i^{\mathrm{UNet}} = \sum_{(u_m, v_n) \in \mathcal{N}(u_i, v_i)} w_{i,mn} F_{\mathrm{UNet}}(u_m, v_n) \in \mathbb{R}^{D_L}, \tag{9}$$

where $\mathcal{N}(u_i, v_i)$ denotes the four neighboring locations on the standard UV grid and $w_{i,mn}$ are the corresponding interpolation weights. The final strand-wise feature is obtained by adding the directly computed K-Plane–MLP feature and the interpolated U-Net feature:

$$L_i(t) = L_i^{\mathrm{KP}}(t) + L_i^{\mathrm{UNet}} \in \mathbb{R}^{D_L}, \qquad (u_i, v_i) \in \mathcal{B}. \tag{10}$$

The sampled features $\{L_i(t) \mid (u_i, v_i) \in \mathcal{B}\}$ serve as guiding-strand features. We further interpolate these guiding features in the scalp UV domain at additional hair-root coordinates. The guiding features and the resulting interpolated features together constitute the complete set of strand features optimized during the coarse stage.

Each strand feature consists of a geometry component and an appearance component,

$$L_i(t) = [L_i^{\mathrm{geo}}(t); L_i^{\mathrm{app}}(t)], \quad L_i^{\mathrm{geo}}(t) \in \mathbb{R}^{D_G}, \quad L_i^{\mathrm{app}}(t) \in \mathbb{R}^{D_A}, \quad D_G + D_A = D_L. \tag{11}$$

We decode $L_i^{\mathrm{geo}}(t)$ into strand displacements using the pretrained geometry decoder from Zakharov et al. (2024), which remains frozen to provide a data-driven strand prior and stabilize optimization. A similar appearance decoder maps $L_i^{\mathrm{app}}(t)$ to spherical harmonic coefficients $\boldsymbol{h}_j \in \mathbb{R}^{12}$ and confidence scores $\gamma_j \in \mathbb{R}^+$ for the precomputed orientations, for $j = 1, \ldots, K$.

Refinement Module. Interpolating guiding-strand features yields a dense and structurally coherent hairstyle, but may smooth out fine appearance variations. To recover these details, we predict a time-varying spherical-harmonic residual $\Delta \boldsymbol{h}_{i,j}(t)$ for each strand Gaussian, enhancing local color variation in the rendered hair.

We model the color residuals using another K-Plane representation in the strand-aligned space (u, v, ℓ, t) rather than the Cartesian space (x, y, z, t). Here, (u_i, v_i) is the scalp UV coordinate of the i-th hair, $\ell_j = (j-1)/(K-1)$ is the normalized index of the j-th Gaussian segment along that hair, and $\tau_t = (t-1)/(T-1)$ is the normalized time coordinate. For each strand Gaussian, we query all six pairwise feature planes and combine them by element-wise multiplication:

$$\begin{aligned} C_{i,j}^{\mathrm{ref}}(t) = C_{uv}^{\mathrm{ref}}(u_i, v_i) \circ C_{u\ell}^{\mathrm{ref}}(u_i, \ell_j) \circ C_{ut}^{\mathrm{ref}}(u_i, \tau_t) \\ \circ C_{v\ell}^{\mathrm{ref}}(v_i, \ell_j) \circ C_{vt}^{\mathrm{ref}}(v_i, \tau_t) \circ C_{\ell t}^{\mathrm{ref}}(\ell_j, \tau_t). \end{aligned} \tag{12}$$

As in the strand generation module, each plane is queried by bilinear interpolation. A lightweight MLP maps the resulting feature to the spherical harmonics residual

$$\Delta \boldsymbol{h}_{i,j}(t) = f_{\mathrm{SH}}\big(C_{i,j}^{\mathrm{ref}}(t)\big) \in \mathbb{R}^{12}, \tag{13}$$

which is added to the coarse-stage coefficient $\boldsymbol{h}_{i,j}(t)$ to refine the rendered color. Thus, appearance variation is represented in the strand-aligned (u, v, ℓ, t) domain rather than the ambient Cartesian space.

4.2 Optimization

The optimization procedure of our dynamic hair reconstruction algorithm is divided into the coarse and the refine stage. We use the traditional 3DGS (Kerbl et al., 2023) rendering objectives, using only $\mathcal{L}_1$ in the coarse stage and additionally introducing $\mathcal{L}_{\mathrm{SSIM}}$ in the refine stage. To enforce the hair strand structure and appearance, we introduce the silhouette loss $\mathcal{L}_{mask}$ (Zakharov et al., 2024) to constraint that the 2D projection of hair strands matches the predicted hair segmentation mask. We also use the hair orientation direction detected on images (see Hair Perception in Sec. 3) to regularize

the 3D strand orientation direction $\boldsymbol{d}_i$. To this end, we project each 3D orientation direction to pixel P and define the orientation loss

$$\mathcal{L}_{ori} = \sum_{P} \left\| \sum_{i=1}^{N_P} T_i \boldsymbol{d}_i^P \alpha_i' - O_d(P) \right\|, \quad T_i = \prod_{j=1}^{i-1} \left(1 - \alpha_j'\right), \tag{14}$$

where $\boldsymbol{d}_i^P$ is the projected 2D orientation, similar to Eq. (4).

Coarse Optimization Stage. During the coarse stage, we disable the refinement module and focus on optimizing the strand generation module. The coarse optimization stage ends when the descent amount of the objective function at a certain resolution is smaller than a predefined threshold ϵ.

Let N_c denote the number of coarse-stage strands, each with K Gaussian segments and $K+1$ endpoints $\{\boldsymbol{p}_{i,j}(t)\}_{j=0}^{K}$. We normalize segment directions to regularize strand shape independently of segment lengths:

$$\boldsymbol{v}_{i,j}(t) = \boldsymbol{p}_{i,j+1}(t) - \boldsymbol{p}_{i,j}(t), \qquad \widehat{\boldsymbol{v}}_{i,j}(t) = \frac{\boldsymbol{v}_{i,j}(t)}{\|\boldsymbol{v}_{i,j}(t)\|_2 + \varepsilon}. \tag{15}$$

Let $s_t(\boldsymbol{x})$ denote the tracked head's signed-distance field at time t, with positive values outside the head. We penalize adjacent-direction differences, endpoint penetration, and squared endpoint displacements between consecutive frames to encourage smooth strands, prevent scalp penetration, and maintain temporal coherence:

$$\mathcal{L}_{\text{smooth}} = \frac{1}{TN_c(K-1)} \sum_{t=1}^{T} \sum_{i=1}^{N_c} \sum_{j=0}^{K-2} \left\| \widehat{\boldsymbol{v}}_{i,j+1}(t) - \widehat{\boldsymbol{v}}_{i,j}(t) \right\|_2, \tag{16}$$

$$\mathcal{L}_{\text{SDF}} = \frac{1}{TN_c(K+1)} \sum_{t=1}^{T} \sum_{i=1}^{N_c} \sum_{j=0}^{K} \text{ReLU}(-s_t(\boldsymbol{p}_{i,j}(t))), \tag{17}$$

$$\mathcal{L}_{\text{time}} = \frac{1}{(T-1)N_c(K+1)} \sum_{t=1}^{T-1} \sum_{i=1}^{N_c} \sum_{j=0}^{K} \left\| \boldsymbol{p}_{i,j}(t+1) - \boldsymbol{p}_{i,j}(t) \right\|_2^2. \tag{18}$$

The coarse-stage objective is therefore

$$\begin{aligned} \mathcal{L}_{\text{coarse}} = \mathcal{L}_1 + \lambda_{mask}\mathcal{L}_{mask} + \lambda_{ori}\mathcal{L}_{ori} \\ + \lambda_{smooth}\mathcal{L}_{\text{smooth}} + \lambda_{SDF}\mathcal{L}_{\text{SDF}} + \lambda_{time}\mathcal{L}_{\text{time}}. \end{aligned} \tag{19}$$

Refine Optimization Stage. Strand Gaussians from the converged coarse stage exhibit plausible geometry and motion, but their interpolated color may lack fine details. We freeze the strand generation module and optimize the refinement module to recover fine-scale appearance details. The refinement objective is

$$\mathcal{L}_{\text{refine}} = \mathcal{L}_1 + \mathcal{L}_{\text{SSIM}}. \tag{20}$$

5 Experiments

We presents an experimental evaluation of proposed approach, including comparable evaluation by novel view synthesis and the ablation study.

5.1 Experimental setup

Dataset. We select six subjects from the NeRSemble (Kirschstein et al., 2023) dataset for training and evaluation. The recordings in NeRSemble are captured simultaneously from multiple viewpoints and include movements of various parts of face. For each subject, we select 16 viewpoints that cover the hair regions. We randomly hold out one viewpoint for novel-view evaluation and use the remaining 15 viewpoints for training. We downsample the images to a resolution of 550 × 802, which is one-quarter of the original size.

Baselines. We evaluate our method against three dynamic reconstruction baselines: 4DGS (Wu et al., 2024), GaussianAvatars (Qian et al., 2024), and Gaussian Head Avatar (Xu et al., 2024).

All three baselines are point-based Gaussian methods. 4DGS optimizes point-based Gaussians in dynamic scenes, whereas GaussianAvatars and Gaussian Head Avatar combine point-based Gaussians with dynamic head representations. DynHair (Sklyarova et al., 2026) also addresses dynamic hair reconstruction with an explicit strand-based representation. However, its code was not publicly available at the time of our experiments, so we do not include it in the baseline comparison.

Metrics. We adopt the evaluation metrics used in Kerbl et al. (2023), including LPIPS, SSIM, and PSNR, to assess rendering quality. Before computing PSNR, LPIPS, and SSIM, we apply the ground-truth hair mask to both the rendered and ground-truth images, setting all pixels outside the hair region to zero. We then compute all three metrics on the resulting full images, rather than restricting the calculation to pixels inside the mask.

Computational Resources. All experiments are executed on a single NVIDIA A100 GPU with 80 GB GPU memory.

Implementation Details. The detailed implementation settings are described in App. A.

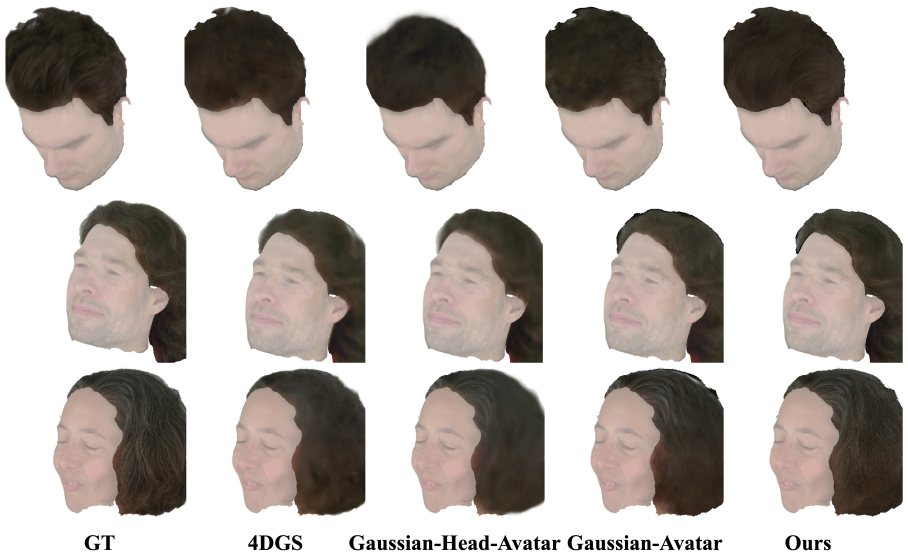

Figure 3: The comparing results of hair rendering on novel view for dynamic methods. From left to right represents the ground truth, the results of 4DGS, Gaussian Head Avatar, GaussianAvatars, and our method.

5.2 Evaluation & comparison

Dynamic Hair Reconstruction. We conduct experiments on six subjects from the NeRSemble dataset (Kirschstein et al., 2023), comparing with three baselines: 4DGS (Wu et al., 2024), GaussianAvatars (Qian et al., 2024), and Gaussian Head Avatar (Xu et al., 2024). We evaluate appearance quality through novel-view synthesis.

4DGS represents both the body and hair using point-based Gaussians, without explicitly encoding the connectivity or directional structure of hair strands. GaussianAvatars and Gaussian Head Avatar incorporate head geometry priors, which provide structural guidance for reconstructing the overall head shape and appearance. These priors help organize the Gaussian representation, but do not explicitly describe the strand structure of hair. Their benefits therefore do not translate into uniform improvements in hair rendering: GaussianAvatars outperforms 4DGS on all three average metrics, whereas Gaussian Head Avatar achieves a slightly lower LPIPS but lower PSNR and SSIM.

Fig. 3 illustrates the differences in reconstructed hair detail. 4DGS produces smooth, blurred hair regions with limited visible strand structure. Gaussian Head Avatar also loses fine-scale texture, particularly in the long-hair example in the bottom row. GaussianAvatars preserves more directional texture, but individual strands remain less distinct in parts of the hair. In contrast, our strand-based representation produces clearer strand directions and more detailed local texture across the illus-

trated hairstyles. This difference is especially visible in the wavy hair in the middle row and the long hair in the bottom row, where our results retain more fine structures visible in the ground truth.

Tab. 1 complements these qualitative observations. Our method achieves the highest average PSNR of 31.59 dB and the lowest average LPIPS of 0.0764 across the six subjects, indicating improved photometric and perceptual fidelity. GaussianAvatars achieves the highest SSIM of 0.9410, while our method obtains 0.9233. Thus, our method improves PSNR and LPIPS and preserves clearer strand-level detail, although it does not achieve the best SSIM. Due to page limitations, additional qualitative reconstruction results are provided in App. A.

Table 1: Quantitative comparison of dynamic hair reconstruction. Metrics are averaged over six NeRSemble subjects. The best result for each metric is highlighted in bold.

Category	Method	PSNR(NVS)↑	SSIM(NVS)↑	LPIPS(NVS)↓
Point	4DGS (Wu et al., 2024)	29.92	0.9401	0.1100
	GaussianAvatars (Qian et al., 2024)	31.34	**0.9410**	0.0896
	Gaussian Head Avatar (Xu et al., 2024)	28.82	0.9357	0.1085
Strand	Ours	**31.59**	0.9233	**0.0764**

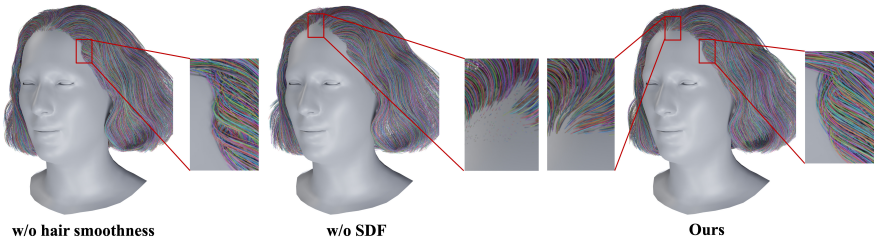

Figure 4: Ablation of the SDF and hair smoothness losses for coarse-stage geometry regularization.

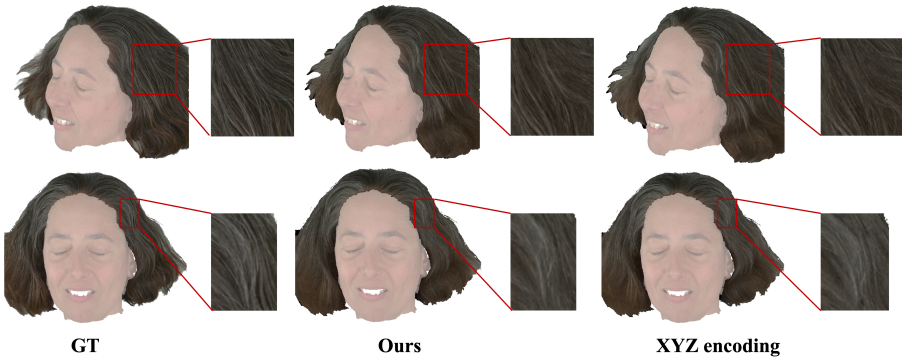

Figure 5: Ablation of coordinate encoding in the fine-stage refinement module: Cartesian XYZ versus strand-aligned UVL.

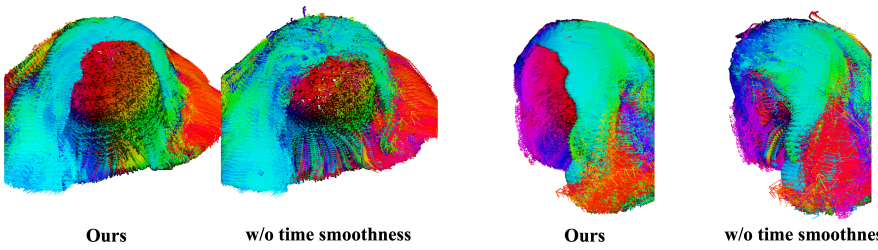

Figure 6: Ablation of the temporal smoothness loss for dynamic hair reconstruction.

5.3 Ablation study

We conduct four ablation experiments to evaluate the geometry regularization in the coarse stage and the coordinate parameterization in the refinement stage. We examine the effect of strand-aligned UVL encoding on appearance refinement in the fine stage. For each experiment, we modify only the component under evaluation while keeping all other settings unchanged. Tab. 2 reports the quantitative results for the SDF loss, hair smoothness loss, and coordinate encoding ablations.

SDF Loss. We remove the signed-distance-field loss $\mathcal{L}_{\text{SDF}}$ from the coarse-stage objective by setting $\lambda_{SDF} = 0$. As shown in Fig. 4, removing this regularization causes hair strands near the forehead to penetrate the scalp. Without the SDF penalty, these intersections are not explicitly penalized, allowing the optimization to fit the observed images with implausible geometry. This demonstrates the role of SDF regularization in preventing scalp penetration and maintaining plausible geometry.

Hair Smoothness Loss. We disable the hair smoothness loss $\mathcal{L}_{\text{smooth}}$ while retaining the other coarse-stage objectives. As shown in Fig. 4, removing this loss leads to more pronounced local bends and zigzag artifacts in the hair along the side of the head. By penalizing differences between adjacent normalized segment directions, the smoothness regularization alleviates these artifacts and promotes smoother, more continuous strand geometry.

Time Smoothness Loss. We remove the temporal smoothness loss $\mathcal{L}_{\text{time}}$ by setting $\lambda_{time} = 0$. Fig. 6 visualizes the optical flow obtained by projecting the reconstructed 3D hair motion between consecutive frames onto the 2D image plane. With temporal regularization, the projected flow is visibly smoother and more coherent, whereas removing this loss produces irregular local variations in the flow field. This comparison demonstrates that temporal smoothness regularization helps suppress frame-to-frame jitter and promotes consistent hair motion.

UVL Encoding in the Fine Stage. We replace the strand-aligned (u, v, ℓ, t) encoding with Cartesian (x, y, z, t) in the fine-stage appearance refinement module, keeping all other settings unchanged. UVL anchors the refinement features to the scalp UV location and the normalized segment index along each strand.

Tab. 2 shows that UVL improves all three metrics over XYZ, with a 14.6% reduction in LPIPS and modest gains in PSNR and SSIM. Consistently, the enlarged regions in Fig. 5 show clearer strand textures and local directional details with UVL, while XYZ produces smoother, less distinct hair textures. These results support the benefit of strand-aligned encoding for fine-scale appearance.

Table 2: Ablation of geometry regularization and fine-stage coordinate encoding. Each variant changes only the indicated component of the full model. The best result for each metric is highlighted in bold.

Variant	PSNR↑	SSIM↑	LPIPS↓
w/o SDF loss	32.26	0.9096	**0.0551**
w/o hair smoothness loss	**32.48**	**0.9097**	0.0553
XYZ encoding	32.17	0.9076	0.0658
Full model (UVL)	32.22	0.9088	0.0562

6 Conclusion and Limitations

Conclusion. In this work, we propose a dynamic hair reconstruction method based on a strand Gaussian representation. By modeling time-varying hair geometry with strand-aware spatial and temporal priors, our method captures flexible hair motion while preserving detailed appearance. Experiments on the NeRSemble dataset demonstrate that our method achieves high-quality dynamic hair reconstruction and produces photorealistic novel-view renderings.

Limitations. Our method relies on hair segmentation, orientation estimation, and head tracking to provide geometric and appearance supervision. Errors in these inputs may propagate to the reconstructed strands, especially in regions with severe occlusion, fine flyaway hairs, or rapid motion. Improving the robustness of hair perception and incorporating stronger hair priors could further enhance reconstruction quality in these challenging cases.

AI USE STATEMENT

In this work, we used generative AI tools to design or provide feedback on research methodology or experiments, help develop theoretical models or conceptual frameworks, propose or refine hypotheses, implement methods, and interpret results. We did not use generative AI tools to generate synthetic datasets, assist with translation, clean or reformat datasets, or support qualitative and thematic data analysis. Formulating mathematical claims, providing critical ingredients for proving mathematical claims, and assisting in the writing of proofs are not applicable to this work. Additionally, we used generative AI tools to summarize or analyze existing literature, discover research topics or identify gaps, brainstorm, and revise the wording and organization of the manuscript. We have reviewed all AI-assisted work. For manuscript preparation, we reviewed AI-assisted drafts through iterative revision, checking that the wording and organization reflected the intended method. We also corrected the presentation of pseudocode, citation formatting, and figure captions to maintain consistency with the manuscript and figures. We take responsibility for the final content of this work, including text, claims, or artifacts produced with the aid of generative AI.

REFERENCES

Volker Blanz and Thomas Vetter. A morphable model for the synthesis of 3d faces. In *Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH)*, SIGGRAPH '99, pp. 187–194. ACM Press/Addison-Wesley Publishing Co., 1999. ISBN 0201485605.

Sara Fridovich-Keil, Giacomo Meanti, Frederik Rahbæk Warburg, Benjamin Recht, and Angjoo Kanazawa. K-planes: Explicit radiance fields in space, time, and appearance. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 12479–12488, 2023.

Berna Kabadayi, Vanessa Sklyarova, Wojciech Zielonka, Justus Thies, and Gerard Pons-Moll. PhysHead: Simulation-ready gaussian head avatars. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4109–4121, June 2026. URL https://arxiv.org/abs/2604.06467.

Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3d gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics (TOG)*, 42(4), July 2023. URL https://repo-sam.inria.fr/fungraph/3d-gaussian-splatting/.

Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *International Conference on Learning Representations (ICLR)*, 2015. URL https://arxiv.org/abs/1412.6980.

Tobias Kirschstein, Shenhan Qian, Simon Giebenhain, Tim Walter, and Matthias Nießner. Nersemble: Multi-view radiance field reconstruction of human heads. *ACM Transactions on Graphics (TOG)*, 42(4), jul 2023. ISSN 0730-0301. doi: 10.1145/3592455. URL https://doi.org/10.1145/3592455.

Zhiyi Kuang, Yiyang Chen, Hongbo Fu, Kun Zhou, and Youyi Zheng. DeepMVSHair: Deep hair modeling from sparse views. In *SIGGRAPH Asia 2022 Conference Papers*, pp. 1–8, 2022.

Tianye Li, Timo Bolkart, Michael J. Black, Hao Li, and Javier Romero. Learning a model of facial shape and expression from 4d scans. *ACM Transactions on Graphics (TOG)*, 36(6), November 2017. ISSN 0730-0301.

Shu Liang, Xiufeng Huang, Xianyu Meng, Kunyao Chen, Linda G Shapiro, and Ira Kemelmacher-Shlizerman. Video to fully automatic 3d hair model. *ACM Transactions on Graphics (TOG)*, 37 (6):206:1–206:14, 2018.

Zhanfeng Liao, Hanzhang Tu, Cheng Peng, Hongwen Zhang, Boyao Zhou, and Yebin Liu. HADES: Human avatar with dynamic explicit hair strands. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 12318–12327, October 2025.

Haimin Luo, Min Ouyang, Zijun Zhao, Suyi Jiang, Longwen Zhang, Qixuan Zhang, Wei Yang, Lan Xu, and Jingyi Yu. GaussianHair: Hair modeling and rendering with light-aware gaussians. *arXiv preprint arXiv:2402.10483*, 2024.

Giljoo Nam, Chenglei Wu, Min H. Kim, and Yaser Sheikh. Strand-accurate multi-view hair capture. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 155–164, June 2019.

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.

Shenhan Qian, Tobias Kirschstein, Liam Schoneveld, Davide Davoli, Simon Giebenhain, and Matthias Nießner. GaussianAvatars: Photorealistic head avatars with rigged 3D gaussians. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 20299–20309, 2024.

Radu Alexandru Rosu, Shunsuke Saito, Ziyan Wang, Chenglei Wu, Sven Behnke, and Giljoo Nam. Neural strands: Learning hair geometry and appearance from multi-view images. In *European Conference on Computer Vision (ECCV)*, pp. 73–89. Springer, 2022.

Vanessa Sklyarova, Jenya Chelishev, Andreea Dogaru, Igor Medvedev, Victor Lempitsky, and Egor Zakharov. Neural haircut: Prior-guided strand-based hair reconstruction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 19762–19773, 2023.

Vanessa Sklyarova, Haonan Chen, Berna Kabadayi, Tobias Kirschstein, Zicong Fan, Xi Wang, Gerard Pons-Moll, Matthias Nießner, Marc Pollefeys, Michael J. Black, and Justus Thies. DynHair: Head avatars with dynamic explicit hair. In *European Conference on Computer Vision (ECCV)*, 2026. URL https://arxiv.org/abs/2607.23861.

Yusuke Takimoto, Hikari Takehara, Hiroyuki Sato, Zihao Zhu, and Bo Zheng. Dr. hair: Reconstructing scalp-connected hair strands without pre-training via differentiable rendering of line segments. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 20601–20611, 2024.

Cong Wang, Di Kang, Heyi Sun, Shenhan Qian, Zixuan Wang, Linchao Bao, and Song-Hai Zhang. MeGA: Hybrid mesh-gaussian head avatar for high-fidelity rendering and head editing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 26274–26284, June 2025.

Ziyan Wang, Giljoo Nam, Tuur Stuyck, Stephen Lombardi, Michael Zollhöfer, Jessica Hodgins, and Christoph Lassner. Hvh: Learning a hybrid neural volumetric representation for dynamic hair performance capture. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 6143–6154, 2022.

Ziyan Wang, Giljoo Nam, Tuur Stuyck, Stephen Lombardi, Chen Cao, Jason Saragih, Michael Zollhöfer, Jessica Hodgins, and Christoph Lassner. Neuwigs: A neural dynamic model for volumetric hair capture and animation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 8641–8651, June 2023.

Yichen Wei, Eyal Ofek, Long Quan, and Heung-Yeung Shum. Modeling hair from multiple views. *ACM Transactions on Graphics (TOG)*, 24(3):816–820, July 2005. ISSN 0730-0301. doi: 10.1145/1073204.1073267. URL https://doi.org/10.1145/1073204.1073267.

Sebastian Winberg, Gaspard Zoss, Prashanth Chandran, Paulo Gotardo, and Derek Bradley. Facial hair tracking for high fidelity performance capture. *ACM Transactions on Graphics (TOG)*, 41 (4):165:1–165:12, 2022.

Guanjun Wu, Taoran Yi, Jiemin Fang, Lingxi Xie, Xiaopeng Zhang, Wei Wei, Wenyu Liu, Qi Tian, and Xinggang Wang. 4d gaussian splatting for real-time dynamic scene rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 20310–20320, June 2024.

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 (CVPR)*, pp. 1526–1535, 2022.

Yuelang Xu, Benwang Chen, Zhe Li, Hongwen Zhang, Lizhen Wang, Zerong Zheng, and Yebin Liu. Gaussian head avatar: Ultra high-fidelity head avatar via dynamic gaussians. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1931–1941, 2024.

Zexiang Xu, Hsiang-Tao Wu, Lvdi Wang, Changxi Zheng, Xin Tong, and Yue Qi. Dynamic hair capture using spacetime optimization. *ACM Transactions on Graphics (TOG)*, 33(6), November 2014. ISSN 0730-0301. doi: 10.1145/2661229.2661284. URL https://doi.org/10.1145/2661229.2661284.

Jingfeng Yao, Xinggang Wang, Lang Ye, and Wenyu Liu. Matte anything: Interactive natural image matting with segment anything model. *Image and Vision Computing*, 147:105067, 2024. doi: 10.1016/j.imavis.2024.105067.

Egor Zakharov, Vanessa Sklyarova, Michael J. Black, Giljoo Nam, Justus Thies, and Otmar Hilliges. Human hair reconstruction with strand-aligned 3D gaussians. In *European Conference on Computer Vision (ECCV)*, pp. 409–425. Springer, 2024.

A APPENDIX

Brief Introduction. The appendix presents experiment settings, supplementary results, and dynamic hairstyle transfer experiments to assess reconstruction quality and asset reusability.

Experiment Settings. We use Adam (Kingma & Ba, 2015) for parameter optimization. We set the learning rate to 10^{-3} for the spatial-temporal feature query network and appearance decoder in coarse stage, 1.6×10^{-6} for the position and keep the same learning rates as 3DGS (Kerbl et al., 2023) in refine stage. We set the coarse-stage loss weights to $\lambda_{SDF} = 30$ for the SDF loss, $\lambda_{smooth} = 1$ for hair smoothness, $\lambda_{time} = 1.5$ for temporal smoothness, and $\lambda_{ori} = \lambda_{mask} = 0.1$ for orientation and mask supervision. We train for 50,000 iterations in the coarse stage and 10,000 iterations in the fine stage. Training across both stages takes approximately 12 hours on a single NVIDIA A100 GPU. In our method, the number of hair strands is set to 20,000, with each strand consisting of 100 nodes, corresponding to 99 Strand Gaussians per strand. The root nodes of our hair strands are generated from the scalp, which is derived from fixed index positions on the FLAME (Li et al., 2017) mesh.

Supplementary Experiments. We select seven subjects from the NeRSemble dataset for qualitative visualization. For each subject, we arbitrarily select two viewpoints and show the ground-truth images, rendered images, and geometry renderings to illustrate both appearance and geometry. Figs. 7 and 8 present these results. Across different hairstyles and viewpoints, our renderings closely match the ground truth in overall hair shape, color, and visible texture, demonstrating good appearance fidelity. The geometry renderings further show coherent strand flow and plausible hair volume, indicating that the reconstructed geometry preserves the structure of each hairstyle.

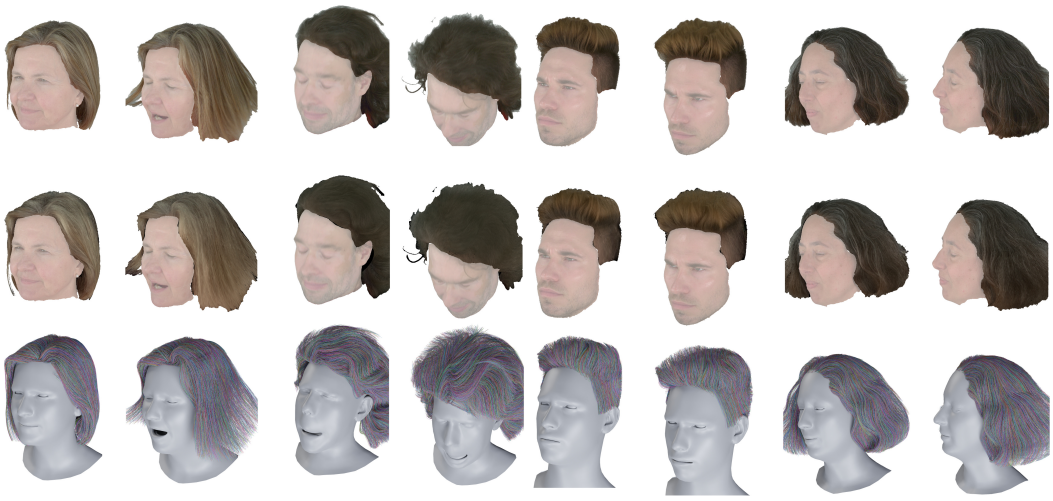

Figure 7: Supplementary results for four subjects, each shown from two viewpoints. The rows show ground-truth images, rendered images, and geometry renderings, respectively.

Dynamic Hairstyle Transfer. We further demonstrate dynamic hairstyle transfer by applying reconstructed hair sequences to target head meshes and replacing the corresponding FLAME coefficients for each sequence. Fig. 9 presents representative transfer results. These results demonstrate that our reconstructed hair can serve as reusable dynamic hair assets, supporting hairstyle transfer across head meshes and the customization of dynamic avatars.



Figure 8: Supplementary results for three additional subjects, each shown from two viewpoints. The rows show ground-truth images, rendered images, and geometry renderings, respectively.



Figure 9: Dynamic hairstyle transfer with corresponding FLAME coefficient replacement. The top row shows reference appearances, the middle row shows reconstructed hairstyles on their original heads, and the bottom row shows transfer results on target head meshes.