

Head Avatars with Dynamic Explicit Hair

Vanessa Sklyarova^{1,2*}, Haonan Chen^{1*}, Berna Kabadayi^{2,4}, Tobias Kirschstein⁵, Zicong Fan¹, Xi Wang^{1,5,8}, Gerard Pons-Moll^{3,4}, Matthias Nießner⁵, Marc Pollefeys^{1,7}, Michael J. Black^{2†}, and Justus Thies^{6†}

¹ ETH Zürich, Switzerland

² Max Planck Institute for Intelligent Systems, Tübingen, Germany

³ Max Planck Institute for Informatics, Saarland Informatics Campus, Germany

⁴ University of Tübingen, Germany

⁵ Technical University of Munich, Germany

⁶ Technical University of Darmstadt, Germany

⁷ Microsoft

⁸ MCML

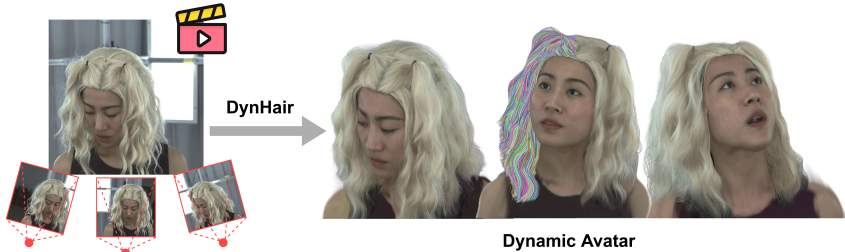


Fig. 1: DynHair is a novel method for learning hair dynamics of 3D strands from data. It leverages multi-view video input to learn a dynamic hair strand deformation model based on the velocity and acceleration of the human head.

Abstract. We present DynHair, a novel method for tracking and modeling dynamic hair for human head avatars. From video input, we reconstruct a dynamic head avatar with an explicit strand-based hair representation using structured 3D Gaussian Splatting. In contrast to the face region of human head avatars, which can be modeled with 3D Gaussians that are attached or generated with respect to some expressive 3D head model, hair is particularly challenging as it exhibits dynamic motion effects. Therefore, we present a novel method that models the dynamic deformations of the hair strands using a temporal network that is conditioned on angular velocity and acceleration of the head, as well as relative gravity. Specifically, an LSTM encodes the motion history and modulates per-point strand features via FiLM conditioning which is then used by an MLP to produce physically plausible displacements to the canonical hairstyle. We jointly optimize this motion and appearance representation of the hair, with a 3DGS-based representation of the face-region, via differentiable Gaussian splatting with photometric, geometric, and physics-based supervision. As a result, we obtain hair

* Equal contribution. † Equal supervision.

tracking of the training video data and an animatable head avatar with controllable hair dynamics. In our experiments, we demonstrate state-of-the-art performance in terms of hair dynamics, temporal consistency, and generalization across subjects. Our data and code are available at <https://dynhair.is.tue.mpg.de/>.

Keywords: Digital Humans · Dynamic Hair Modeling · Hair Tracking

1 Introduction

In the past few years, there has been tremendous progress in modeling photo-realistic 3D digital human appearance for applications in e-commerce, entertainment, and telepresence in AR/VR [73]. Especially, the introduction of neural radiance fields [39] and 3D Gaussian splatting [30] has led to a series of works [9, 24, 29, 52, 55] focusing on visual quality [18, 31, 47, 62], runtime [26, 70, 71], few-shot reconstruction [10, 32, 72], and disentanglement [14, 54]. In addition to photorealism, realistic motion [3, 12, 53] is a critical aspect for immersive avatars. Avatars should behave like their physical counterparts. However, current avatar methods concentrate mostly on the face region, where they use 3D morphable model-based [33, 43] conditioning to handle expression-dependent deformations, ignoring dynamic hair motion. The lack of dynamic hair breaks the sense of realism as the avatars do not follow physical principles. When the head rotates to the side, the hair should follow the direction of gravity. When the head shakes, the hair should lag behind, swing with inertia, and exhibit physically plausible follow-through and settling behavior.

Such behavior can be modeled with physics-based simulation [1, 11], requiring a strand-based reconstruction of the hair and appropriate physics parameters for mass, stiffness, damping, friction and more. Recovering such a strand-based reconstruction is highly challenging as most of the hair strands are occluded and only the outer hair layer is visible. Existing works that aim at reconstructing strand-based hair [4, 21, 46, 50, 60, 65] from images and video assume static hair geometry and exploit learned hairstyle priors that help reconstruct the inner structure of the hair. They demonstrate animation with such reconstructed hairstyles, but the physics parameters for simulation are hand-tuned and not recovered from the observations. Automatically finding realistic physics parameters for simulation is currently an unsolved problem, as it would require efficient differentiable hair simulation. In contrast, we propose to learn hair dynamics directly from data with physics-based losses.

Specifically, we propose *DynHair*, a method to learn dynamic strand-based hair directly from multi-view video. It captures appearance through strand-aligned 3D Gaussians that follow the underlying hair geometry. The canonical hairstyle is initialized from a strand-based hair prior [50] and deformed at each timestep based on the head motion history. The key component is the temporal motion model for the hair that is conditioned on gravity, and the velocity and acceleration of the head. Specifically, we extract angular velocity, acceleration, and relative gravity from 3D head tracking and encode a sliding window of

these to condition an LSTM to produce a temporal motion embedding. Based on this motion embedding, we use FiLM [44] conditioning to modulate per-point strand features before a strand MLP predicts displacements of the 3D Gaussians. A learnable root attenuation factor encodes the physical prior that strand roots anchored to the scalp move less than free-moving tips. To train the method end-to-end, we combine photometric losses (RGB, SSIM, perceptual) with hair-geometric supervision (silhouette, orientation, penetration), physics-based regularization (elastic stretch preservation), and color regularization for spatially and temporally coherent hair appearance. In addition to the hair, we jointly render and train the upper body region as an unstructured set of 3D Gaussians, whose positions and attributes are conditioned on facial expression and pose via deformation MLPs following [62].

Our approach combines the rendering quality of Gaussian-based avatars with the geometric fidelity and physical plausibility of explicit strand representations. The experiments demonstrate state-of-the-art results on dynamic head avatars, with improved hair dynamics, temporal consistency, and generalization across subjects and motion patterns compared to existing methods. In summary, our main contributions are:

- We propose a dynamic hair motion model based on structured hair strands. An LSTM encoder processes head-local angular velocity, acceleration, and relative gravity as temporal conditioning signals. The resulting pose features modulate canonical hair point features via FiLM layers, which are then passed to an MLP to predict non-rigid hair motion.
- We propose an avatar reconstruction approach that disentangles hair and upper body regions, with region-dependent motion modeling and appearance representations learned from multi-view video.

2 Related Work

Head avatars. Early head avatar approaches rely on explicit geometric representations, including meshes [6, 7, 19, 27, 59] and volumetric primitives [17, 36]. Subsequently, following NeRF [39], implicit representations [38, 39, 64] have become predominant [15, 16, 29, 31, 40–42, 67, 71]. Initial efforts focus on deformable NeRFs; for instance, Park *et al.* [42] learn a hyperspace that maps observation space to a canonical space via a learnable latent code. Parametric face models such as FLAME [33] and BFM [43] are widely used to guide the deformation from observation space to canonical space [15, 16, 71]. For example, Kirschstein *et al.* [31] introduce a deformation field and an ensemble of 3D multi-resolution hash encodings to construct high-fidelity human heads from multi-view videos. IMavatar [67] learns an expression-conditioned occupancy network from monocular video similar to SNARF [8].

Although implicit representations achieve high realism, they require dense sampling in 3D space, which limits rendering efficiency. To overcome this, recent works use Gaussian Splatting [30] for head avatar modeling [9, 18, 32, 45, 48, 54,

55,61,62], offering real time rendering while maintaining visual fidelity. Gaussian Head Avatar (GHA) [62] conditions Gaussians on BFM parameters and incorporates a super resolution module for high quality rendering. GaussianAvatars [45] attaches Gaussians to a rigged FLAME mesh to enable controllable animation. SplattingAvatar [48] embeds Gaussians on a triangular mesh using barycentric coordinates. Similarly, FlashAvatar [61] embeds Gaussians in the UV space of a FLAME mesh with a learnable offset for fine details. GaussianHead [55] uses a parametric face mesh to describe large scale motion and stores view-dependent appearance in tri-planes. Unlike our approach, all the above methods treat hair implicitly as part of the head representation, limiting strand-level detail and hair motion. We disentangle the hair and the face, introducing a new hair model, and represent the head following GHA [62].

Hair Animation. Hair motion plays a crucial role in character animation, games, and virtual avatars, yet remains challenging due to the large number of strands, complex physical behavior, and frequent self- and body-collisions. Methods that model hair dynamics can be broadly categorized into physics-based simulators [5, 13, 22, 25], approaches that approximate simulator behavior with learned models [51, 56, 68], and methods that learn dynamics from image observations [35, 57, 58, 63].

Physics-based simulators generate hair dynamics using numerical solvers and are widely integrated in production [1, 2, 11, 20]. While stable and interpretable, they are computationally expensive and require manual parameter tuning for each hairstyle type.

Learned approximations of simulation. Recent methods, such as GroomGen [68] and DGH [56], approximate the simulator’s behavior by training a neural network on 3D data generated in Houdini. Quaffure [51] extends this direction with self-supervised learning of hair dynamics conditioned on body pose using physics-based constraints. However, none of these methods are personalized, nor do they model full avatar appearance.

Capture-based methods learn dynamics directly from an observed video. Yang et al. [63] reconstruct dynamic hair from a monocular video using separate networks for spatial occupancy and temporal motion. HVH [58] represents hair volumetrically and learns dynamics via guide-strand tracking with RGB and optical flow supervision. NeuWigs [57] learns a compressed appearance representation followed by a temporal transfer module for latent-space dynamics. HHAvatar [35] models hair with unstructured Gaussians and predicts deformations via an MLP conditioned on past hair states and head poses. PhysHead [28] first reconstructs a static strand-based hairstyle using Neural Haircut [49], then animates the resulting Gaussians with dynamics from the physics-based simulator Maya. While PhysHead optimizes hair appearance from a static time step, we optimize from a dynamic sequence. Concurrent work HADES [34] also uses strand-based geometry and learns dynamics conditioned on head pose for full-body avatars, but conditions on a history of past hair states and models only a sparse set of strands, interpolating the rest via nearest neighbors.

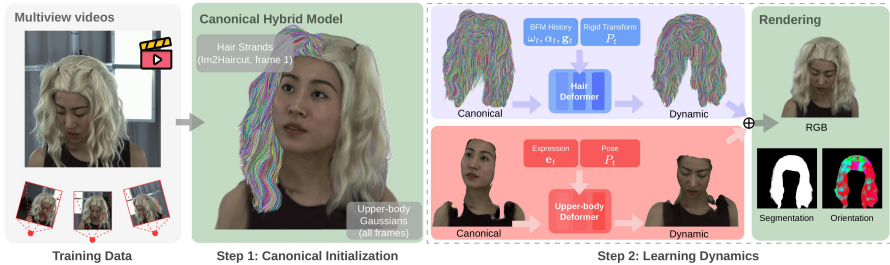


Fig. 2: Method overview. Given multi-view video input, we first initialize a canonical hybrid model for the upper body and hair. Hair strands are initialized using the Im2Haircut [50] prior model, and the upper body is represented with unstructured 3D Gaussians [62]. We then jointly optimize hair and head deformations with two separate motion models. The hair deformation is conditioned on BFM motion history with an additional rigid transformation, while the head deformation is conditioned on facial expressions and pose parameters. Finally, the hair and upper body Gaussians are composited and rendered via differentiable splatting to produce the output images.

In contrast to these capture-based methods, we model hair motion in a structured manner using a strand-based representation. We predict deformations for all strands relative to canonical hair, rather than from a history of past states, and optimize appearance directly from a dynamic sequence rather than a static one. We further incorporate acceleration and gravity as explicit inputs and introduce a physics-based elastic term to improve the realism of the resulting motion.

3 Method

Our method models a dynamic head avatar with explicit strand-based hair from multi-view videos. As shown in Figure 2, our proposed method disentangles hair and the upper body. Both regions, and especially their deformations, are handled differently. Following [62], the upper body is represented with an unstructured set of 3D Gaussians whose positions and attributes are conditioned on facial expression and pose. In contrast to [62], which does not disentangle body and hair, we model the hairstyle separately as a set of explicit polyline strands with strand-aligned Gaussian primitives and a learned dynamic deformation module that is conditioned on the head motion history and gravity. At each timestep, we predict deformations of both the upper body and the hair, concatenate all their Gaussians, and render them jointly via differentiable Gaussian Splatting [30] to obtain RGB color, segmentation labels, and strand orientation maps enabling joint supervision through photometric losses, hair-geometric constraints, and physics-based regularization. In the following, we introduce our novel hair modeling, while the upper body model is detailed in the suppl. material.

3.1 Modeling Dynamic Hair

Our dynamic hair modeling consists of two stages: static hairstyle reconstruction at the initial timestep, followed by learning a head motion-conditioned deformation network for dynamic animation of hair.

Canonical representation. We model the hairstyle $\mathcal{H}$ as a set of N polylines originating from a scalp mesh: $\mathcal{H} = \{S_i\}_{i=1}^N$, where each strand $S_i = \{p_{ij}\}_{j=1}^L$ consists of L points. To obtain the canonical hairstyle, we leverage Im2Haircut [50], a PCA-based hair prior with learned coarse-to-fine encoders. Given multi-view images at the initial timestep ($t=0$), the encoder predicts PCA coefficients that are decoded into strand geometry through a learned shape basis. The resulting strands are upsampled to the target resolution, filtered using a scalp silhouette mask to remove spurious strands, and then resampled to $L = 40$ points per strand. We jointly optimize the strand geometry using the Im2Haircut reconstruction losses and additional regularization for geometric consistency:

$$\mathcal{L}_{\text{static}} = \mathcal{L}_{\text{im2haircut}} + \lambda_{\text{pca}} \mathcal{L}_{\text{pca}} + \lambda_{\text{length}} \mathcal{L}_{\text{length}} + \lambda_{\text{smooth}} \mathcal{L}_{\text{smooth}} + \lambda_{\text{consist}} \mathcal{L}_{\text{consist}}, \quad (1)$$

where $\mathcal{L}_{\text{pca}}$ regularizes PCA coefficients, $\mathcal{L}_{\text{length}}$ and $\mathcal{L}_{\text{smooth}}$ encourage nearby strands to have similar lengths and shapes (measured via root-based KNN), and $\mathcal{L}_{\text{consist}}$ enforces that the PCA distribution of back-view strands matches that of front-view strands, improving geometry in occluded regions. This yields the canonical hairstyle $\mathcal{H}_0$ used as input for the dynamic stage.

Motion conditioning. Hair dynamics result from both rigid head motion and non-rigid deformations caused by inertia and gravity. To model these dynamics, we extract the global pose $P_t = \{R_t, T_t, S_t\}$ from the parametric face model [43] at each timestep t and compute the *angular velocity*, *acceleration*, and *relative gravity direction*. The *angular velocity and acceleration* of the head are computed using central differences. At each timestep t , they are given by:

$$\omega_t^{\text{world}} = \frac{\text{Log}(R_{t+1} R_{t-1}^\top)}{2 \Delta t}, \quad \alpha_t^{\text{world}} = \frac{\omega_{t+1}^{\text{world}} - \omega_{t-1}^{\text{world}}}{2 \Delta t}, \quad (2)$$

where $\text{Log}(\cdot)$ converts a rotation from $\text{SO}(3)$ to $\mathbb{R}^3$. Both signals are then rotated into the head-local frame to obtain pose-invariant conditioning: $\omega_t = R_t^\top \omega_t^{\text{world}}$ and $\alpha_t = R_t^\top \alpha_t^{\text{world}}$. The *relative gravity direction*, defined as the gravity vector in the head-local frame, is obtained by rotating the canonical gravity vector:

$$\mathbf{g}_t = \text{normalize}(R_t^\top [0, -1, 0]^\top). \quad (3)$$

For training and inference stability, we normalize ω_τ and α_τ per-axis by precomputed maximum absolute values from the training set and clamped to $[-1, 1]$.

Hair deformation network. We train a hair deformer network $\mathcal{D}_{\text{hair}}$ that predicts non-rigid per-point displacements for the canonical hairstyle, conditioned on the head motion history of the past T frames. Specifically, for each

frame τ in the sliding window $\{t-T, \dots, t\}$, we form the conditioning vector $\mathbf{c}_\tau = [\boldsymbol{\omega}_\tau, \boldsymbol{\alpha}_\tau, \mathbf{g}_\tau] \in \mathbb{R}^9$. Each $\mathbf{c}_\tau$ is mapped through positional encoding $\gamma(\mathbf{c}_\tau)$ and then processed by an LSTM with learnable initial hidden states (h_0, c_0) to produce a temporal motion embedding $\mathbf{z}_t$:

$$\mathbf{z}_t = \text{LSTM}(\{\gamma(\mathbf{c}_\tau)\}_{\tau=t-T}^t; h_0, c_0). \quad (4)$$

The motion embedding $\mathbf{z}_t$ modulates per-point strand features via Feature-wise Linear Modulation (FiLM) [44]. A linear layer maps $\mathbf{z}_t$ to scale and shift parameters $(\boldsymbol{\gamma}_{\text{film}}, \boldsymbol{\beta}_{\text{film}})$, which modulate the positional encoding of strand point p_{ij} :

$$(\boldsymbol{\gamma}_{\text{film}}, \boldsymbol{\beta}_{\text{film}}) = \text{Linear}(\mathbf{z}_t), \quad \text{FiLM}(p_{ij}) = \gamma(p_{ij}) \odot (1 + \boldsymbol{\gamma}_{\text{film}}) + \boldsymbol{\beta}_{\text{film}}. \quad (5)$$

The modulated features are concatenated with a strand position embedding (encoding the normalized position j/L along the strand) and passed through a strand-wise MLP to predict per-point displacements:

$$\Delta p_{ij} = \text{MLP}\left(\text{FiLM}(p_{ij}), \gamma(j/L)\right) \cdot \rho_j, \quad (6)$$

where ρ_j is a learnable root attenuation factor initialized as a linear ramp from small values near the root to 1.0 at the tip. This encodes the physical prior that strand roots anchored to the scalp exhibit less displacement than free-moving tips. The final MLP layer is initialized near zero to ensure small initial deformations and stable early training. The deformed hairstyle is:

$$\mathcal{H}_t = \mathcal{H}_0 + \Delta \mathcal{H}_t, \quad (7)$$

which is rigidly transformed to world coordinates using the head’s global pose.

Strand-aligned Gaussian rendering. Following [37, 65, 69], we constrain the 3D Gaussian primitives to lie on the strand polylines S_i . The mean of each Gaussian is placed at the midpoint of a line segment, its primary scale axis is aligned with the segment direction (with magnitude proportional to the segment length), and the two cross-sectional scale axes are set to a fixed strand width. The Gaussian rotation quaternion is computed via parallel transport from the reference axis to the segment direction. We taper the Gaussians near strand roots to reduce visual artifacts at the scalp boundary.

3.2 Joint Training and Losses

We concatenate the 3D Gaussians for the hair and upper body and jointly render them via differentiable Gaussian Splatting [30] to obtain RGB images, segmentation masks for hair and body, and orientation maps for hair. Training is performed in two stages: (i) the *static initialization* stage, where we optimize for the canonical hairstyle $\mathcal{H}_0$ using the first timestep across all views; and (ii) the *dynamic training* stage, where we jointly train the hair deformation network

$\mathcal{D}_{\text{hair}}$, $\mathcal{H}_0$ with small learning rate, head deformation networks, and hair appearance. The objective combines photometric, hair-geometric, physics-based, and color-regularization losses:

$$\mathcal{L}_{\text{dynamic}} = \mathcal{L}_{\text{photo}} + \mathcal{L}_{\text{hair}} + \mathcal{L}_{\text{color_reg}}. \quad (8)$$

Photometric losses. We supervise the rendered images using a combination of pixel-wise, structural similarity, and perceptual losses:

$$\mathcal{L}_{\text{photo}} = \lambda_{\text{rgb}} \mathcal{L}_{\text{rgb}} + \lambda_{\text{ssim}} \mathcal{L}_{\text{ssim}} + \lambda_{\text{vgg}} \mathcal{L}_{\text{vgg}}, \quad (9)$$

where $\mathcal{L}_{\text{rgb}}$ is the L1 distance between the rendered and ground-truth images. The structural dissimilarity loss $\mathcal{L}_{\text{ssim}} = 1 - \text{SSIM}(\hat{I}, I)$ penalizes deviations in local image structure using a Gaussian-windowed SSIM computation. The perceptual loss $\mathcal{L}_{\text{vgg}}$ is the LPIPS distance [66] computed with a pre-trained VGG backbone, encouraging realistic texture and high-frequency detail.

Geometric losses. To supervise geometry, we introduce losses on the silhouette, strand orientation, head-hair penetration and inextensibility constraints:

$$\mathcal{L}_{\text{hair}} = \lambda_{\text{seg}} \mathcal{L}_{\text{seg}} + \lambda_{\text{orient}} \mathcal{L}_{\text{orient}} + \lambda_{\text{penetr}} \mathcal{L}_{\text{penetr}} + \lambda_{\text{elastic}} \mathcal{L}_{\text{elastic}}. \quad (10)$$

The *segmentation loss* $\mathcal{L}_{\text{seg}}$ encourages the rendered hair silhouette to match the ground-truth hair mask $\hat{M}_h$. We use an asymmetric recall formulation that penalizes missing ground-truth coverage more strongly than over-coverage:

$$\mathcal{L}_{\text{seg}} = \frac{1}{|\Omega|} \sum_{\Omega} \max(\hat{M}_h - M_h, 0) + \frac{0.5}{|\Omega|} \sum_{\Omega} \max(M_h - \hat{M}_h, 0), \quad (11)$$

where M_h is the predicted hair silhouette and Ω denotes the set of visible pixels. An exponential decay schedule is applied after a warm-up period to avoid over-constraining the silhouette. The *orientation loss* $\mathcal{L}_{\text{orient}}$ matches predicted 2D strand orientations with ground-truth orientation maps. Since undirected orientations have 180° ambiguity, we use a wrap-aware formulation:

$$\mathcal{L}_{\text{orient}} = \frac{\pi}{|\Omega_h|} \sum_{\Omega_h} \min(|\theta_{\text{pred}} - \theta_{\text{gt}}|, |\theta_{\text{pred}} - \theta_{\text{gt}} - 1|, |\theta_{\text{pred}} - \theta_{\text{gt}} + 1|), \quad (12)$$

where Ω_h is the intersection of predicted and ground-truth hair masks. The *penetration loss* $\mathcal{L}_{\text{penetr}}$ prevents hair strands from penetrating the head mesh. We randomly sample up to 10,000 strand points and use signed queries to identify points inside the head mesh $\mathcal{M}$. For inside points, we penalize the squared distance to the nearest mesh triangle:

$$\mathcal{L}_{\text{penetr}} = \frac{1}{|\mathcal{P}_{\text{in}}|} \sum_{p \in \mathcal{P}_{\text{in}}} d^2(p, \mathcal{M}), \quad \mathcal{P}_{\text{in}} = \{p_{ij} : \text{inside}(p_{ij}, \mathcal{M})\}, \quad (13)$$

where $d^2(p, \mathcal{M})$ denotes the squared point-to-mesh distance. The *elastic regularization term* penalizes stretching and compression of hair segments during

deformation. For consecutive strand points p_j and p_{j+1} , let $d^{\text{rest}} = p_{j+1}^0 - p_j^0$ and $d^{\text{posed}} = p_{j+1}^t - p_j^t$ denote the segment vectors in the rest and posed configurations, respectively. We apply this loss only after static initialization:

$$\mathcal{L}_{\text{elastic}} = \frac{1}{NL} \sum_{i=1}^N \sum_{j=1}^{L-1} \left(\|d_{ij}^{\text{posed}}\| - \|d_{ij}^{\text{rest}}\| \right)^2. \quad (14)$$

Hair color regularization. We introduce three complementary color regularization terms to ensure spatially and temporally coherent hair appearance:

$$\mathcal{L}_{\text{color_reg}} = \lambda_{\text{cg}} \mathcal{L}_{\text{cg}} + \lambda_{\text{cs}} \mathcal{L}_{\text{cs}} + \lambda_{\text{cc}} \mathcal{L}_{\text{cc}}. \quad (15)$$

The *color gradient loss* $\mathcal{L}_{\text{cg}}$ penalizes abrupt color changes between consecutive points along each strand: $\mathcal{L}_{\text{cg}} = \frac{1}{NL} \sum_{i,j} \|c_{i,j+1} - c_{i,j}\|^2$. The *spatial color smoothness loss* $\mathcal{L}_{\text{cs}}$ encourages spatially nearby Gaussians (across different strands) to share similar colors, computed via k-nearest-neighbor ($k=8$) matching: $\mathcal{L}_{\text{cs}} = \frac{1}{|\mathcal{N}|} \sum_{(a,b) \in \mathcal{N}} \|c_a - c_b\|^2$. The *color consistency loss* $\mathcal{L}_{\text{cc}}$ encourages uniform color along each strand by penalizing the deviation from the per-strand mean color: $\mathcal{L}_{\text{cc}} = \frac{1}{NL} \sum_{i,j} \|c_{i,j} - \bar{c}_i\|^2$. See suppl. material for loss weights.

4 Experiments

Dataset details. We train DynHair on two multiview datasets: three scenes from the HHAvatar dataset [35] and two from a newly captured dataset. HHAvatar has 4 synchronized cameras, about 3 minutes per scene at 2048×2048 resolution, with diverse hair motion and facial expressions. Our new dataset captures actors with 15 synchronized cameras following a setup similar to [31]. It includes a predefined set of diverse hair motions, spanning slow to rapid accelerations, nodding, tilting, rotations, lateral moves, sudden stops, with varied facial expressions. All participants performed the same 22 actions, 10 targeting hair dynamics. Sequences are at 4k and 72 FPS. We provide an additional 360° coverage with and without hairnet for static geometry estimation.

Training details. We train DynHair in three stages. In Stage 1, we train a mesh-based head model to obtain canonical head geometry. In Stage 2, we first reconstruct the canonical hairstyle using Im2Haircut [50] from all views and then we train the hair deformation and optimize the head and hair Gaussian primitives jointly. We train our full model for 320,000 iterations on a single A100, increasing the VGG perceptual loss weight for the last 80,000 iterations to enhance fine-grained appearance details. In all experiments, we use a resolution 1024×1024 , window history of 5 frames, 40 points per strand, and around 11,000 strands depending on hairstyle. See details in suppl. material.

4.1 Comparison and Results

Baselines. To our knowledge, there is no publicly available method that learns hair dynamics from data for direct comparison. Therefore, we evaluate our ap-

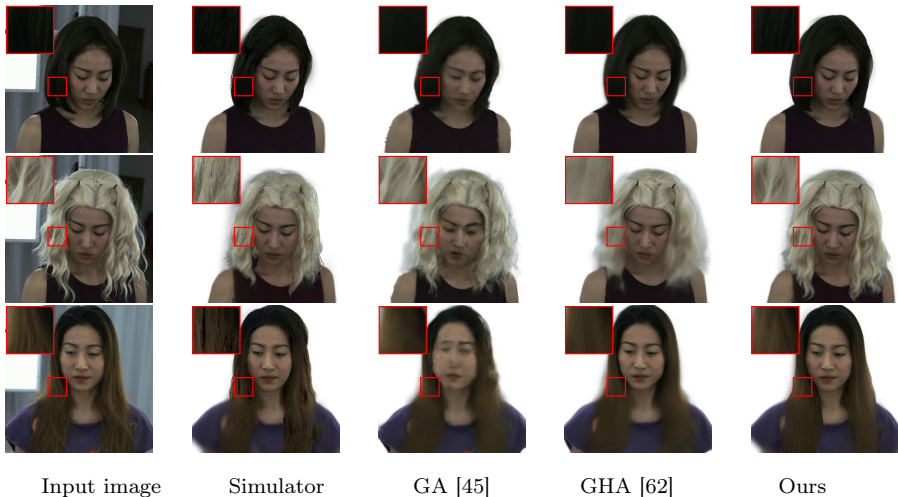


Fig. 3: Self-reenactment comparison. Each row shows a different test frame or subject. Columns from left to right: ground-truth image, physics-based simulator from Maya [1], GaussianAvatars [45], Gaussian Head Avatar [62], and our method. Our approach produces sharper strand-level detail and more temporally coherent hair motion.

proach against state-of-the-art multi-view dynamic head avatar methods, Gaussian Head Avatar [62] (GHA) and GaussianAvatars (GA) [45] as well as a physics simulator (Maya [1]) that computes hair dynamics via numerical integration of Newtonian forces. Since it outputs geometry only, we apply our estimated colors and drive Gaussians using the simulated motion.

Qualitative evaluation. We compare with the baselines under both self-reenactment and cross-reenactment settings by evaluating appearance fidelity and motion realism in each scenario. Figure 3 shows self-reenactment results on held-out test sequences. Our method produces sharper hair details and more temporally coherent strand motion than baselines, which produce over-smoothed and flickering hair regions. The physics-based simulator deviates from the true hair motion due to manually obtained parameters and the bounded accuracy of the physical model. Moreover, directly attaching learned appearance to simulator-driven strand positions results in degraded rendering quality (see Figure 3, middle row). This further motivates our joint learning approach.

Figure 4 presents cross-reenactment results, where the driving signal comes from a different subject. Our explicit strand-based representation, conditioned on relative motion, produces geometry-consistent hair animation for novel driving sequences, while baselines struggle to preserve identity and realistic dynamics.

Quantitative evaluation. For quantitative comparison, we report pixel-aligned metrics (PSNR, SSIM) alongside perceptual LPIPS [66] and measure distribution realism (FID) [23]. We evaluate on both full images and the hair region, computed on a test set of three subjects. Our method achieves slightly lower

Table 1: Quantitative comparison. Self-reenactment results averaged across test sets of 3 subjects. As pixel-wise alignment metrics are insufficient to fully capture hair motion quality, we additionally report $tLPIPS_{\text{ex}}$ as a dedicated temporal measure. $tIoU_{\text{hair}}^{gt}$ is equal to 0.946, and $tLPIPS_{\text{ex}}$ should be close to 0.

Method	Full Image				Hair Region					Temporal $tLPIPS_{\text{ex}}$	Motion ($\times 10^{-3}$)	
	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$	FID $\downarrow$	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$	$IoU_{\text{hair}} \uparrow$	$tIoU_{\text{hair}} \uparrow$		vel. $\uparrow$	accel. $\downarrow$
GA [45]	20.17	0.7787	0.2536	45.73	19.94	0.7335	0.1362	—	—	126.93	-0.0230	0.40 0.08
GHA [62]	22.33	0.7925	0.2056	36.25	22.15	0.7446	0.1086	—	—	81.07	-0.0127	2.17 0.24
Maya*	19.39	0.6610	0.2465	62.64	17.89	0.5182	0.1484	0.776	0.925	121.97	0.0183	2.38 0.11
Ours	21.60	0.7638	0.2010	30.06	21.01	0.6885	0.1101	0.878	0.936	37.51	0.0045	2.41 0.19

PSNR and SSIM compared to GHA [62]. This behavior is expected: enforcing strand-level geometric consistency prevents the model from arbitrarily smoothing high-frequency structures. While this introduces small pixel-wise deviations that negatively affect reconstruction metrics, it better preserves fine strand-level details and overall geometric fidelity. The marginally higher hair LPIPS similarly reflects a fundamental metric mismatch, as LPIPS penalizes any deviation from a single GT strand configuration regardless of physical plausibility, and thus favors appearance memorization over dynamic correctness. As a method that explicitly models hair dynamics, ours generates temporally coherent, physically plausible motion rather than reproducing one specific GT realization. We show superior performance in terms of the FID metric, which measures perceptual realism and better reflects the goals of hair dynamics learning than pixel-aligned reconstruction scores. To measure temporal flickering, we compute $tLPIPS_{\text{ex}} =$

$$\frac{1}{T-1} \sum_{t=1}^{T-1} \left[LPIPS(I_t \odot \hat{M}_t, I_{t+1} \odot \hat{M}_t) - LPIPS(\hat{I}_t \odot \hat{M}_t, \hat{I}_{t+1} \odot \hat{M}_t) \right],$$

where $\hat{M}_t = \text{clamp}(\hat{H}_t + \hat{H}_{t+1}, 0, 1)$. Here, I_t and $\hat{I}_t$ denote the predicted and ground-truth images at frame t , respectively, and $\hat{H}_t$ is the ground-truth hair segmentation mask at frame t . LPIPS between consecutive frames measures perceptual temporal variation. Subtracting the corresponding ground-truth temporal LPIPS isolates excess perceptual change introduced by the model, yielding a measure of temporal fidelity beyond natural motion. Values near zero indicate alignment with the ground truth, positive values indicate excess variation, and negative values indicate over-smoothing.

Temporal evaluation highlights the benefit of explicit strand-based modeling. Baselines produce negative $tLPIPS_{\text{ex}}$, indicating over-smoothed, stiff motion, while our method better matches ground-truth hair dynamics. These results demonstrate a principled trade-off: unconstrained Gaussians optimize pixel alignment at the cost of sharp high-frequency detail and realistic temporal behavior, while our structured representation prioritizes geometric and temporal consistency, producing sharper and more dynamically faithful hair reconstructions while maintaining competitive global image quality.

We also compute the non-rigid hair velocity and acceleration by unposing the head to a fixed canonical pose. Then, we define velocity as the mean nearest-

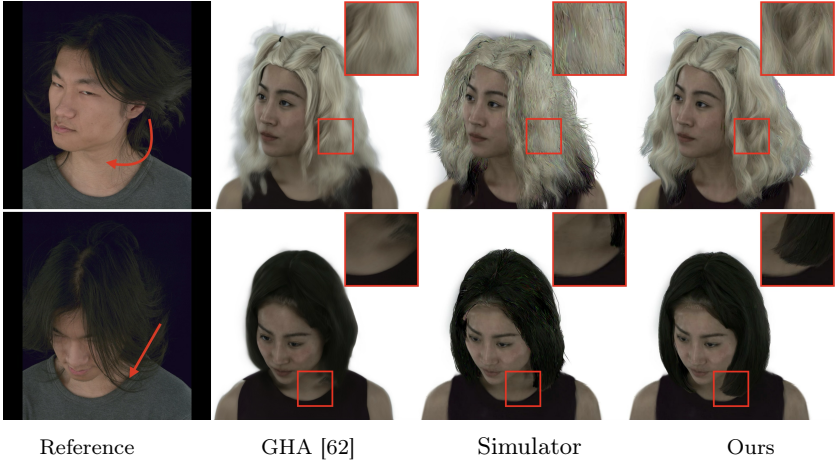


Fig. 4: Cross-reenactment comparison. The 1st row shows left-right head rotation, while the second row corresponds to nodding. Our hairstyle better follows physics constraints and gravity. The image in the first column serves as a reference for human motion. While the hairstyle should follow general dynamics, it is not an exact match.

neighbor displacement between consecutive hair point clouds in this head-local frame, and acceleration as its finite difference. For our method and simulator, we use the structured strand vertices; for baselines with unstructured Gaussians, we select those whose projections fall within the hair silhouette mask at the first frame and use the same fixed indices across all subsequent frames. Both representations are subsampled to the same point count for fair comparison. Our method produces higher hair velocity, indicating more expressive hair motion (see Tab. 1), compared to baselines with unstructured representation. The Maya simulator achieves lower acceleration, demonstrating more temporally coherent hair dynamics; however, there is increased error in IoU_{hair} and $\text{tIoU}_{\text{hair}}$.

4.2 Ablation Study

Tab. 2 presents an ablation of our key design choices. Hair dynamics are evaluated using (1) image-space temporal metrics and (2) motion-based metrics, including velocity explosion ratio, curvature temporal smoothness, displacement spike, and drift.

Visual metrics. For visual 2D metrics, we compute IoU_{hair} , which measures the overlap between predicted and ground-truth hair silhouettes, and $\text{tIoU}_{\text{hair}}$, which measures temporal stability by computing the overlap between predicted hair masks in consecutive frames. The ground-truth $\text{tIoU}_{\text{hair}}$ averaged across all scenes is 0.951. We also include LPIPS in hair region and $\text{tLPIPS}_{\text{ex}}$.

Physical realism. To evaluate the realism of hair dynamics beyond standard image-quality metrics, we introduce four motion-based metrics. Let $p_{ij}^t \in \mathbb{R}^3$ be

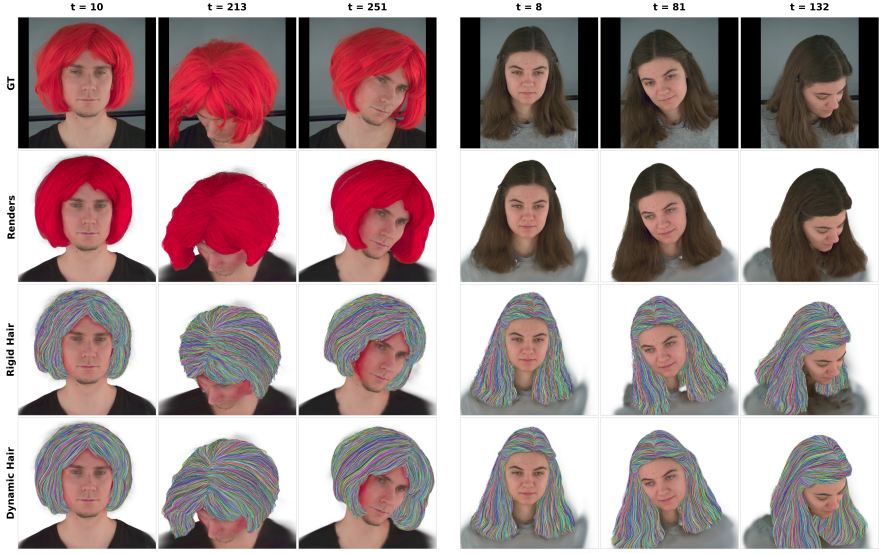


Fig. 5: We show more results on newly captured sequences with diverse hair motions. Notably, ours recovers per-strand temporal hair motion. We include an ablation without the non-rigid deformation module, where hair moves rigidly with the head transform.

the position of the j -th point on the i -th strand at frame t , and Δt the time step. Then *velocity explosion ratio* (VER) is defined as the ratio of maximum to median speed across all points and frames:

$$VER = \frac{\max_{t,i,j} \|v_{ij}^t\|}{\text{median}_{t,i,j} \|v_{ij}^t\| + \epsilon}, \quad v_{ij}^t = \frac{p_{ij}^{t+1} - p_{ij}^t}{\Delta t}. \quad (16)$$

We also include *Curvature Temporal Smoothness* (CTS), which quantifies temporal jitter in strand curvature by averaging the frame-to-frame change in curvature vectors across all frames, strands, and strand points. Lower values indicate smoother, more temporally consistent hair deformation.

To detect sudden, physically implausible positional jumps, we introduce the *Displacement Spike* ($P95\text{-Disp}$) metric, which is defined as the 95th percentile of per-frame maximum strand displacements:

$$P95\text{-Disp} = P_{95} \left(\left\{ \max_{i,j} \|p_{ij}^{t+1} - p_{ij}^t\| \right\}_{t=1}^{T-1} \right). \quad (17)$$

Finally, *Angular Momentum Drift* (AMD) measures violations of angular momentum conservation. In an isolated system, angular momentum should remain constant over time:

$$AMD = \frac{1}{T-1} \sum_{t=1}^{T-1} \|L^t - L^1\|, \quad L^t = \sum_{i,j} (p_{ij}^t - \bar{p}^t) \times v_{ij}^t, \quad (18)$$



Fig. 6: Hairstyle manipulation. The reconstructed dynamic hairstyle can be further manipulated by changing the brightness of Gaussians or trimming them.

Table 2: Quantitative comparison on physical realism of hair motion. Results are averaged across 2 scenes and 3 motion types, estimated at the same number of iterations.

Method	Silhouette & Appearance				Physics			
	IoU _{hair} ↑	tIoU _{hair} ↑	LPIPS _{hair} ↓	tLPIPS _{ex}	VER ↓	P95 ↓	CTS ↓	AMD ↓
Ours	0.883	0.945	0.101	0.0024	26.88	0.071	0.00022	7787
Maya (Reference)	0.789	0.941	0.131	0.0080	33.73	0.057	0.00010	7120
w/ pose	0.885	0.945	0.101	0.0028	36.27	0.071	0.00019	7816
w/o acc.	0.878	0.944	0.101	0.0026	29.05	0.077	0.00021	8332
w/o g	0.852	0.932	0.104	0.0058	23.13	0.070	0.00022	8159
w/o $\mathcal{L}_{elastic}$	0.873	0.943	0.103	0.0047	283.75	1.159	0.00156	196993
w/o film	0.877	0.945	0.102	0.0009	40.46	0.104	0.00038	8277
w/ mlp enc.	0.879	0.949	0.102	-0.0010	32.18	0.065	0.00020	7377

where $\bar{p}^t = \frac{1}{NM} \sum_{i,j} p_{ij}^t$ is the center of mass. For all metrics, lower values indicate more physically plausible motion.

In Tab. 2, we compute all metrics for the same number of iterations (240k), averaged across 2 scenes with 3 motion types: nodding, left-right motion, and head rotation. Maya-simulated strands are included for reference. While Maya simulations expectedly produce better physics metrics (except VER), we observe degradation in coverage metrics due to suboptimal parameters. We study the effect of each conditioning signal in the hair deformation model. Conditioning on absolute pose worsens generalization, while removing gravity prevents strands from falling properly (e.g., during nodding). Removing acceleration (“w/o acc”) slightly degrades the metrics and reduces motion expressiveness. We find that the deformation model plays a crucial role. Without it, hair motion appears rigid and unrealistic (see “Rigid Hair” vs. “Dynamic Hair” in Fig. 5). Replacing the LSTM head pose encoder with an MLP results in stiffer motion (see negative tLPIPS_{ex}). Modulating strand features with head pose features improves performance across all metrics (compare ours with “w/o film”). Removing $\mathcal{L}_{elastic}$ significantly degrades results. Strands tend to stretch, reducing physical plausibility. See more ablations in suppl. material.

5 Discussion

Applications. Disentangled modeling of head and hairstyle provides rich possibilities for direct hair manipulation. For example, we can edit the brightness of Gaussians or trim the hairstyle length while preserving realistic hair motion

dynamics (see Fig. 6). Furthermore, the explicit strand-based representation allows us to derive physically meaningful motion signals from the reconstructed hair geometry, see Fig. 5, which can be used for transfer applications.

Limitations. Our method models only hair-head interactions; hair-body collisions are handled implicitly through data-driven priors and are not explicitly resolved. The approach depends on hair silhouette estimates from pre-trained segmentation models, which can be inaccurate for complex hairstyles or under occlusion. We also observe that rendering quality at hair-head boundaries can exhibit artifacts due to the discrete transition between the two Gaussian representations. Additionally, reliance on the landmark detector and tracker can lead to incorrect reconstructions when hair occludes the face. Our formulation provides a foundation for future large-scale learning of dynamic hair, where richer datasets and explicit collision modeling can be naturally integrated.

Conclusion. In this work, we presented a novel dynamic head avatar framework with explicit strand-based hair, learned end-to-end from multi-view video. Our method represents hair as strand-aligned 3D Gaussians whose deformations are predicted by a temporal motion model conditioned on head angular velocity, acceleration, and relative gravity. By combining an LSTM head pose encoder with FiLM modulation and physics-based regularization, the approach produces temporally coherent hair dynamics with strand-level geometric fidelity. Experiments demonstrate high visual quality and improved motion plausibility compared with state-of-the-art head avatar methods, both quantitatively and qualitatively. We believe that explicitly modeling hair structure and dynamics within a unified Gaussian avatar framework is a step toward more realistic and controllable digital humans.

6 Acknowledgements

Vanessa Sklyarova is supported by the Max Planck ETH Center for Learning Systems. Justus Thies is supported by the ERC Starting Grant 101162081 “LeMo” and the DFG Excellence Strategy - EXC-3057. This work is supported by the ERC Consolidator Grant Gen3D (101171131). Berna Kabadayi is supported by the IMPRS-IS. Gerard Pons-Moll is a member of the ML Cluster of Excellence, EXC-2064/1 – Project No. 390727645, and is endowed by the Carl Zeiss Foundation. This work is supported by the BMBF: Tübingen AI Center, FKZ: 01IS18039A, and by the DFG – 409792180. This work is supported by the Swiss National Science Foundation Advanced Grant 216260: “Beyond Frozen Worlds: Capturing Functional 3D Digital Twins from the Real World.”

The authors would like to thank Zhanfeng Liao for providing data from HHAvatar, Egor Zakharov for supervising Haonan during his time at ETH, and Benjamin Pellkofer for IT support.

Disclosure. While MJB is employed by Epic Games, this work was performed solely at, and funded solely by, the Max Planck Society.

References

1. Autodesk, INC. maya. Maya, accessed: 2025-10-10
2. Epic Games unreal engine. Unreal Engine, accessed: 2025-10-10
3. Aneja, S., Sevastopolsky, A., Kirschstein, T., Thies, J., Dai, A., Nießner, M.: Gaussianspeech: Audio-driven personalized 3d gaussian avatars. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 13065–13075 (October 2025)
4. Ben Ayed, F., Becherini, G., Thies, J., Sklyarova, V.: Vid2Haircut: 3D Strand-Based Hairstyle Reconstruction from Video. In: Musialski, P., Lim, I. (eds.) Eurographics 2026 - Short Papers. The Eurographics Association (2026)
5. Bertails, F., Audoly, B., Cani, M.P., Querleux, B., Leroy, F., Lévêque, J.L.: Superhelices for predicting the dynamics of natural hair. Transactions on Graphics, (Proc. SIGGRAPH) **25**(3), 1180–1187 (2006)
6. Blanz, V., Basso, C., Poggio, T., Vetter, T.: Reanimating faces in images and video. In: Computer graphics forum. vol. 22, pp. 641–650. Wiley Online Library (2003)
7. Bouaziz, S., Wang, Y., Pauly, M.: Online modeling for realtime facial animation **32**(4) (Jul 2013)
8. Chen, X., Zheng, Y., Black, M.J., Hilliges, O., Geiger, A.: SNARF: Differentiable forward skinning for animating non-rigid neural implicit shapes. In: International Conference on Computer Vision (ICCV) (2021)
9. Chen, Y., Wang, L., Li, Q., Xiao, H., Zhang, S., Yao, H., Liu, Y.: Monogaussianavatar: Monocular gaussian point-based head avatar. In: SIGGRAPH Conference Papers (SA). pp. 1–9 (2024)
10. Chu, X., Harada, T.: Generalizable and animatable gaussian head avatar. In: The Thirty-eighth Annual Conference on Neural Information Processing Systems (2024)
11. Community, B.O.: Blender - a 3d modelling and rendering package (2018)
12. Daněček, R., Schmitt, C., Polikovsky, S., Black, M.J.: Supervising 3D talking head avatars with analysis-by-audio-synthesis. In: International Conference on 3D Vision (3DV) (2026)
13. Daviet, G.: Interactive hair simulation on the gpu using admm. In: SIGGRAPH Conference Papers (SA). SIGGRAPH '23, Association for Computing Machinery, New York, NY, USA (2023)
14. Feng, Y., Liu, W., Bolkart, T., Yang, J., Pollefeys, M., Black, M.J.: Learning disentangled avatars with hybrid 3d representations. arXiv preprint arXiv:2309.06441 (2023)
15. Gafni, G., Thies, J., Zollhofer, M., Nießner, M.: Dynamic neural radiance fields for monocular 4d facial avatar reconstruction. In: Conference on Computer Vision and Pattern Recognition (CVPR). pp. 8649–8658 (2021)
16. Gao, X., Zhong, C., Xiang, J., Hong, Y., Guo, Y., Zhang, J.: Reconstructing personalized semantic facial nerf models from monocular video. ACM Transactions on Graphics (TOG) **41**(6), 1–12 (2022)
17. Garbin, S.J., Kowalski, M., Estellers, V., Szymanowicz, S., Rezaeifar, S., Shen, J., Johnson, M.A., Valentin, J.: Voltemorph: Real-time, controllable and generalizable animation of volumetric representations. In: Computer graphics forum. vol. 43, p. e15117. Wiley Online Library (2024)
18. Giebenhain, S., Kirschstein, T., Rünz, M., Agapito, L., Nießner, M.: Npga: Neural parametric gaussian avatars. In: SIGGRAPH Asia 2024 Conference Papers. pp. 1–11 (2024)

19. Grassal, P.W., Prinzler, M., Leistner, T., Rother, C., Nießner, M., Thies, J.: Neural head avatars from monocular rgb videos. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. pp. 18653–18664 (2022)
20. He, C., Amador Herrera, J.A., Zhou, Y., Shu, Z., Sun, X., Feng, Y., Pirk, S., Michels, D.L., Zhang, M., Wang, T.Y., Rushmeier, H.: Digital salon: An ai and physics-driven tool for 3d hair grooming and simulation. In: *SIGGRAPH Asia 2024 Real-Time Live! SA Real-Time Live! '24* (2024)
21. He, C., Sun, X., Shu, Z., Luan, F., Pirk, S., Herrera, J.A.A., Michels, D.L., Wang, T.Y., Zhang, M., Rushmeier, H., Zhou, Y.: Perm: A parametric representation for multi-style 3d hair modeling. In: *International Conference on Learning Representations (ICLR)* (2025)
22. Herrera, J.A.A., Zhou, Y., Sun, X., Shu, Z., He, C., Pirk, S., Michels, D.L.: Augmented mass-spring model for real-time dense hair simulation. In: *International Conference on Computer Vision (ICCV)* (2024)
23. Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., Hochreiter, S.: Gans trained by a two time-scale update rule converge to a local nash equilibrium (2018)
24. Hong, Y., Peng, B., Xiao, H., Liu, L., Zhang, J.: Headnerf: A real-time nerf-based parametric head model. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 20374–20384 (2022)
25. Huang, L., Yang, F., Wei, C., Chen, Y.J.E., Yuan, C., Gao, M.: Towards realtime: A hybrid physics-based method for hair animation on gpu. *Computer Graphics and Interactive Techniques (PACMCGIT)* **6**(3) (2023)
26. Iandola, F., Pidhorskyi, S., Santesteban, I., Gupta, D., Pahuja, A., Bartolovic, N., Yu, F., Garbin, E., Simon, T., Saito, S.: SqueezeMe: Mobile-ready distillation of gaussian full-body avatars. In: *SIGGRAPH. ACM* (2025)
27. Ichim, A.E., Bouaziz, S., Pauly, M.: Dynamic 3d avatar creation from hand-held video input. *ACM Trans. Graph.* **34**(4) (Jul 2015)
28. Kabadayi, B., Sklyarova, V., Zielonka, W., Thies, J., Pons-Moll, G.: PhysHead: Simulation-ready gaussian head avatars. In: *Proc. IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)* (June 2026)
29. Kabadayi, B., Zielonka, W., Bhatnagar, B.L., Pons-Moll, G., Thies, J.: Gan-avatar: Controllable personalized gan-based human head avatar. In: *International Conference on 3D Vision (3DV)* (March 2024)
30. Kerbl, B., Kopanas, G., Leimkühler, T., Drettakis, G.: 3d gaussian splatting for real-time radiance field rendering. *Transactions on Graphics, (Proc. SIGGRAPH)* **42**(4), 139–1 (2023)
31. Kirschstein, T., Qian, S., Giebenhain, S., Walter, T., Nießner, M.: Nersemble: Multi-view radiance field reconstruction of human heads. *ACM Transactions on Graphics (TOG)* **42**(4), 1–14 (2023)
32. Kirschstein, T., Romero, J., Sevastopolsky, A., Nießner, M., Saito, S.: Avat3r: Large animatable gaussian reconstruction model for high-fidelity 3d head avatars. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. pp. 12089–12100 (2025)
33. Li, T., Bolkart, T., Black, M.J., Li, H., Romero, J.: Learning a model of facial shape and expression from 4d scans. *Transactions on Graphics, (Proc. SIGGRAPH Asia)* **36**(6), 194–1 (2017)
34. Liao, Z., Tu, H., Peng, C., Zhang, H., Zhou, B., Liu, Y.: Hades: Human avatar with dynamic explicit hair strands. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)* (October 2025)

35. Liao, Z., Xu, Y., Li, Z., Li, Q., Zhou, B., Bai, R., Xu, D., Zhang, H., Liu, Y.: Hhavatar: Gaussian head avatar with dynamic hairs. *Transactions on Pattern Analysis and Machine Intelligence (TPAMI)* **PP** (2025)
36. Lombardi, S., Simon, T., Schwartz, G., Zollhoefer, M., Sheikh, Y., Saragih, J.: Mixture of volumetric primitives for efficient neural rendering. *Transactions on Graphics, (Proc. SIGGRAPH)* **40**(4) (2021)
37. Luo, H., Ouyang, M., Zhao, Z., Jiang, S., Zhang, L., Zhang, Q., Yang, W., Xu, L., Yu, J.: GaussianHair: Hair modeling and rendering with light-aware gaussians. *arXiv* (2024)
38. Mescheder, L., Oechsle, M., Niemeyer, M., Nowozin, S., Geiger, A.: Occupancy networks: Learning 3d reconstruction in function space. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. pp. 4460–4470 (2019)
39. Mildenhall, B., Srinivasan, P.P., Tancik, M., Barron, J.T., Ramamoorthi, R., Ng, R.: Nerf: Representing scenes as neural radiance fields for view synthesis. In: *European Conference on Computer Vision (ECCV)*. Springer (2020)
40. Nehvi, J., Kabadayi, B., Valentin, J., Thies, J.: Volumetric portrait avatar. In: *DAGM German Conference on Pattern Recognition*. pp. 3–19. Springer (2024)
41. Park, K., Sinha, U., Barron, J.T., Bouaziz, S., Goldman, D.B., Seitz, S.M., Martin-Brualla, R.: Nerfies: Deformable neural radiance fields. In: *Proceedings of the IEEE/CVF international conference on computer vision*. pp. 5865–5874 (2021)
42. Park, K., Sinha, U., Hedman, P., Barron, J.T., Bouaziz, S., Goldman, D.B., Martin-Brualla, R., Seitz, S.M.: Hypernerf: A higher-dimensional representation for topologically varying neural radiance fields. *ACM Trans. Graph.* **40**(6) (dec 2021)
43. Paysan, P., Knothe, R., Amberg, B., Romdhani, S., Vetter, T.: A 3d face model for pose and illumination invariant face recognition. In: *2009 sixth IEEE international conference on advanced video and signal based surveillance*. pp. 296–301. Ieee (2009)
44. Perez, E., Strub, F., de Vries, H., Dumoulin, V., Courville, A.C.: Film: Visual reasoning with a general conditioning layer. In: *AAAI* (2018)
45. Qian, S., Kirschstein, T., Schoneveld, L., Davoli, D., Giebenhain, S., Niefkner, M.: Gaussianavatars: Photorealistic head avatars with rigged 3d gaussians. In: *Conference on Computer Vision and Pattern Recognition (CVPR)*. pp. 20299–20309 (2024)
46. Rosu, R.A., Wu, K., Feng, Y., Zheng, Y., Black, M.J.: DiffLocks: Generating 3d hair from a single image using diffusion models. In: *Conference on Computer Vision and Pattern Recognition (CVPR)* (2025)
47. Saito, S., Schwartz, G., Simon, T., Li, J., Nam, G.: Relightable gaussian codec avatars. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. pp. 130–141 (2024)
48. Shao, Z., Wang, Z., Li, Z., Wang, D., Lin, X., Zhang, Y., Fan, M., Wang, Z.: Splattingavatar: Realistic real-time human avatars with mesh-embedded gaussian splatting. In: *Conference on Computer Vision and Pattern Recognition (CVPR)*. pp. 1606–1616 (2024)
49. Sklyarova, V., Chelishev, J., Dogaru, A., Medvedev, I., Lempitsky, V., Zakharov, E.: Neural haircut: Prior-guided strand-based hair reconstruction. *2023 IEEE/CVF International Conference on Computer Vision (ICCV)* pp. 19705–19716 (2023)
50. Sklyarova, V., Zakharov, E., Prinzler, M., Becherini, G., Black, M., Thies, J.: Im2haircut: Single-view strand-based hair reconstruction for human avatars. In: *International Conference on Computer Vision (ICCV)* (2025)

51. Stuyck, T., Lin, G.W.C., Larionov, E., Chen, H.y., Bozic, A., Sarafianos, N., Roble, D.: Quaffure: Real-time quasi-static neural hair simulation. In: Conference on Computer Vision and Pattern Recognition (CVPR) (2024)
52. Teotia, K., Rhodin, H., Mendiratta, M., Kim, H., Habermann, M., Theobalt, C.: Audio driven universal gaussian head avatars. In: Proceedings of the SIGGRAPH Asia 2025 Conference Papers. SA Conference Papers '25, Association for Computing Machinery, New York, NY, USA (2025)
53. Thambiraja, B., Prinzler, M., Aliakbarian, S., Cosker, D., Thies, J.: 3diFACE: Synthesizing and editing holistic 3d facial animation. In: International Conference on 3D Vision 2025 (2025)
54. Wang, C., Kang, D., Sun, H., Qian, S., Wang, Z., Bao, L., Zhang, S.H.: 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)
55. Wang, J., Xie, J.C., Li, X., Xu, F., Pun, C.M., Gao, H.: Gaussianhead: High-fidelity head avatars with learnable gaussian derivation. IEEE Transactions on Visualization and Computer Graphics (2025)
56. Wang, J., Xu, Y., Tretschk, E., Wang, Z., Ianina, A., Bozic, A., Neumann, U., Tung, T.: Dgh: Dynamic gaussian hair. In: Advances in Neural Information Processing Systems (NeurIPS) (2025)
57. Wang, Z., Nam, G., Stuyck, T., Lombardi, S., Cao, C., Saragih, J., Zollhoefer, M., Hodgins, J., Lassner, C.: Neuwigs: A neural dynamic model for volumetric hair capture and animation. In: Conference on Computer Vision and Pattern Recognition (CVPR) (2023)
58. Wang, Z., Nam, G., Stuyck, T., Lombardi, S., Zollhoefer, M., Hodgins, J., Lassner, C.: Hvh: Learning a hybrid neural volumetric representation for dynamic hair performance capture. In: Conference on Computer Vision and Pattern Recognition (CVPR) (2022)
59. Weise, T., Bouaziz, S., Li, H., Pauly, M.: Realtime performance-based facial animation. ACM transactions on graphics (TOG) **30**(4), 1–10 (2011)
60. Wu, K., Yang, L., Kuang, Z., Feng, Y., Han, X., Shen, Y., Fu, H., Zhou, K., Zheng, Y.: Monohair: High-fidelity hair modeling from a monocular video. In: Conference on Computer Vision and Pattern Recognition (CVPR) (2024)
61. Xiang, J., Gao, X., Guo, Y., Zhang, J.: Flashavatar: High-fidelity head avatar with efficient gaussian embedding. In: Conference on Computer Vision and Pattern Recognition (CVPR). pp. 1802–1812 (2024)
62. Xu, Y., Chen, B., Li, Z., Zhang, H., Wang, L., Zheng, Z., Liu, Y.: Gaussian head avatar: Ultra high-fidelity head avatar via dynamic gaussians. In: Conference on Computer Vision and Pattern Recognition (CVPR). pp. 1931–1941 (2024)
63. Yang, L., Shi, Z., Zheng, Y., Zhou, K.: Dynamic hair modeling from monocular videos using deep neural networks. Transactions on Graphics, (Proc. SIGGRAPH) **38**(6) (2019)
64. Yariv, L., Gu, J., Kasten, Y., Lipman, Y.: Volume rendering of neural implicit surfaces. In: Thirty-Fifth Conference on Neural Information Processing Systems (2021)
65. Zakharov, E., Sklyarova, V., Black, M.J., Nam, G., Thies, J., Hilliges, O.: Human hair reconstruction with strand-aligned 3d gaussians. In: European Conference of Computer Vision (ECCV) (2024)
66. Zhang, R., Isola, P., Efros, A.A., Shechtman, E., Wang, O.: The unreasonable effectiveness of deep features as a perceptual metric. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 586–595 (2018)

67. Zheng, Y., Abrevaya, V.F., Bühler, M.C., Chen, X., Black, M.J., Hilliges, O.: Im avatar: Implicit morphable head avatars from videos. In: Conference on Computer Vision and Pattern Recognition (CVPR). pp. 13545–13555 (2022)
68. Zhou, Y., Chai, M., Pepe, A., Gross, M., Beeler, T.: Groomgen: A high-quality generative hair model using hierarchical latent representations. Transactions on Graphics, (Proc. SIGGRAPH) (2023)
69. Zhou, Y., Chai, M., Wang, D., Winberg, S., Wood, E., Sarkar, K., Gross, M., Beeler, T.: Groomcap: High-fidelity prior-free hair capture. Transactions on Graphics, (Proc. SIGGRAPH) (2024)
70. Zielonka, W., Bolkart, T., Beeler, T., Thies, J.: Gaussian eigen models for human heads. In: CVPR (June 2025)
71. Zielonka, W., Bolkart, T., Thies, J.: Instant volumetric head avatars. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 4574–4584 (2023)
72. Zielonka, W., Garbin, S.J., Lattas, A., Kopanas, G., Gotardo, P., Beeler, T., Thies, J., Bolkart, T.: Synthetic prior for few-shot drivable head avatar inversion. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (June 2025)
73. Zielonka, W., Kirschstein, T., Bolkart, T., Giebenhain, S., Sklyarova, V., Deng, X., Xiang, D., Saito, S., Liu, Y., Nießner, M., Thies, J.: How to build digital humans? from priors to photorealistic avatars. Computer Graphics Forum (Eurographics State-of-the-Art Report) **45**(2) (2026)