PhysHead: Simulation-Ready Gaussian Head Avatars

Berna Kabadayi[1,3] Vanessa Sklyarova[1,2] Wojciech Zielonka[4*] Justus Thies[1,4] Gerard Pons-Moll[3,5,6]

[1]Max Planck Institute for Intelligent Systems [2]ETH Zürich [3]University of Tübingen
[4]Technical University of Darmstadt [5]Tübingen AI Center [6]Max Planck Institute for Informatics

Figure 1. We introduce PhysHead, a novel method for creating photorealistic head avatars with dynamic hair from multi-view videos. We leverage a vision language model to create a layered representation of the avatar, decomposed into a head and a hair layer (left). The two separate layers enable localized control of facial expressions with a 3DMM and physics-guided simulation of the hair. The strand-based representation of the hair works seamlessly with physics engines and enables hair geometry and appearance editing (right). Reference images are shown in orange.

Abstract

Realistic digital avatars require expressive and dynamic hair motion; however, most existing head avatar methods assume rigid hair movement. These methods often fail to disentangle hair from the head, representing it as a simple outer shell and failing to capture its natural volumetric behavior. In this paper, we address these limitations by introducing ***PhysHead****, a hybrid representation for animatable head avatars with realistic hair dynamics learned from multi-view video. Our approach combines a 3D parametric mesh for the head with strand-based hair, which can be directly simulated using physics engines. For the appearance model, we employ Gaussian primitives attached to both the head mesh and hair segments. This representation enables the creation of photorealistic head avatars with dynamic hair behavior, such as wind-blown motion, overcoming the constraints of rigid hair in existing methods. However, these animation capabilities also require new training schemes. In particular, we propose the use of VLM-based models to generate appearance of regions that are occluded in the dynamic training sequences. In quantitative and qualitative studies, we demonstrate the capabilities of the proposed model and compare it with existing baselines. We show that our method can synthesize physically plausible hair motion besides expression and camera control. For additional results and code, please refer to* `https://phys-head.github.io`.

1. Introduction

Human hair in real life is dynamic. It moves in different directions when we rotate our heads or blow in the wind. Modeling hair dynamics for digital avatars is challenging as hair is complex, composed of thousands of tiny strands. Thus, most data-driven approaches for animatable head avatar reconstruction often assume static hair. Despite being photorealistic [23, 59], there is no clear separation between the head and the hair in these approaches. When the head moves, the hair moves rigidly with it. Recent works have addressed this problem by disentangling the hair from head [17, 31, 82], however, it does not solve the problem of the head avatar having static hair. Concurrent follow-up methods extend this approach by capturing hair dynam-

*This work was conducted while Wojciech Zielonka was at TU Darmstadt.

ics data and learning hair dynamics from the data [42, 43]. While promising, it is simply not feasible to capture all possible hair dynamics and effects such as wind, since changing wind source direction would affect the hair dynamics. Unfortunately, these methods cannot generalize to unseen hair dynamics scenarios.

In contrast, explicit strand-based hair of digital avatars in movies and games is simulated using physical simulation systems [6, 15, 37] and the appearance is carefully crafted by artists. Strand-based representation gives artists full control over hair properties such as stiffness or damping. There are methods to obtain such a representation from multiview [68, 69, 88, 94] or one-shot images [87, 96, 98]. Strand-based hair reconstruction methods typically focus on the geometry of the hair, and when hair strands are animated, a user-defined appearance model is manually incorporated [15]. Since the focus is on hair geometry and appearance [97], facial expressions are ignored.

Instead, we look for a solution with the following properties: (1) a photorealistic head avatar which is driven by a standard 3D morphable model (3DMM), and (2) a hair representation which is compatible with classical physics engines, while providing a photorealistic appearance. To this end, we propose PhysHead, a fully controllable head avatar. Specifically, we use a layered representation for the face and hair based on a 3D Gaussian appearance model. The dynamics of the face are driven by FLAME [40], while the hair is represented by strands, and the hair dynamics are driven by a physics simulator [6, 12]. This layered representation is flexible, however, there are challenges: (i) As our method, uses physics simulator, it generalizes to unseen head poses. However, these unseen head poses also reveal the unobserved head regions, such as ears and neck during animation. This is especially visible for actors with long hair, where the hair often covers the ears or side views. (ii) During animation of the hair, previously occluded hair Gaussian primitives will become visible. To address the first challenge, we propose to use a vision language model (VLM) to complete the head that is occluded by hair. Specifically, we task the VLM [11] to edit the captured images to remove the hair and generate bald images. As the VLM does not guarantee the generation of a consistent appearance across all frames of our training sequence, where the subject shows different facial expressions, we first optimize for a shared texture of the FLAME model using differentiable rendering. We then generate bald training frames by blending the rendered shared texture with visible skin parts of the captured images. We use these additional generated images to train a dynamic head avatar without hair. Then, in a second stage, we optimize hair appearance as an additional layer on top of the bald avatar. To address the second challenge, we design a color consistency loss for unobserved hair regions, ensuring that hair appearance remains consistent, even in areas like inner strands that are not directly observed in the images.

Extensive experiments show the effectiveness of our novel approach. We compare against state-of-the-art dynamic avatar methods, including the concurrent work HairCup [31] which also disentangles hair and skin using SDS-based training schemes. In summary, our main contributions are:

- A two-stage framework for reconstructing 3D head avatars with physics-guided animation of hair.
- A method to generate dynamic bald avatars leveraging a vision language model (VLM) efficiently with differentiable rendering, then combining them with real captures through image-based blending, enabling generalization across diverse skin tones.
- A strand-level regularization that allows occluded strands to transfer appearance from views where the hair is visible, yielding plausible hair appearance during animation.

2. Related Work

This work focuses on generating high-quality Gaussian avatars with realistic and dynamic hair, positions itself within the broader landscape of hair reconstruction, hair simulation, compositional head avatars, and holistic head avatar reconstruction. For a comprehensive overview of neural rendering for photorealistic face and full-body avatars, we refer to the reports on face tracking and reconstruction [105], morphable models [14], and two neural rendering surveys [77, 78].

2.1. Animatable Head Avatar

A common approach to controllable avatar creation is directly optimizing the parameters of a 3DMM [7, 40]. Early works, such as Thies *et al*. [79, 80], employed custom second-order GPU optimizers for real-time performance, inspiring numerous follow-up methods [58, 73, 76, 102]. However, PCA-based appearance models limit reconstruction quality, and typical 3DMMs omit hair geometry, hindering realistic novel-view synthesis.

To improve fidelity, several methods leverage NeRFs to capture fine structures like hair, skin, and eyes [1, 2, 4, 5, 19, 21, 22, 27–29, 38, 44, 51, 83, 101]. Gafni *et al*. [19] introduced the first NeRF-based animatable avatar conditioned on expression coefficients, while others employ triangle-based deformations [101], triplanes [27], mixtures of volumetric primitives [44], tetrahedral meshes [22], or multi-level voxel blendshapes [21]. Despite high-quality facial animation, these methods ignore hair dynamics, producing static hair during motion or lacking explicit strand modeling. PhysHead addresses this by jointly modeling facial expressions and strand-based hair dynamics.

The emergence of 3D Gaussian Splatting (3DGS) [30] led to Gaussian-based avatar methods [3, 16, 18, 23, 33, 34, 41, 42, 50, 53, 59, 62, 65, 75, 82, 90, 92, 103, 104]. Qian *et al*. [59] rigs Gaussians to a 3DMM for real-time performance with high-quality rendering, while others embed Gaussians on the mesh surface with neural corrective fields for wrinkles and self-shadows [34, 41, 65, 74, 90, 92, 103, 104]. Gaussian Head Avatars (GHA) [91] first reconstructs per frame head meshes, and later deformations are learned by MLPs. An additional super-resolution module enables synthesizing high-quality avatars. Advanced methods like RGCA [62] learn relightable primitives for ultra-realistic rendering. However, these methods do not explicitly model hair strands, which limits their applicability for physics-based hair simulation. This limitation extends to multi-person avatar models. GPHM [92] employs a neural parametric Gaussian model controlled by shape and expression coefficients, analogous to PCA-based 3DMMs. SynShot [104] leverages synthetic training data and bridges the domain gap through pivotal fine-tuning. CAP4D [74] adopts a two-stage pipeline similar to [20], first regressing per-subject images using a multi-view diffusion model, then constructing drivable avatars. Avat3r [34] directly regresses Gaussian primitives in a single stage, producing high-quality, multi-view-consistent avatars. Despite these advances, none enforce strand-level hair modeling, making them unsuitable for physics-based hair simulation which is an aspect explicitly addressed by PhysHead.

2.2. Compositional Human Modeling

A growing body of work [8, 9, 17, 25, 31, 52, 82, 93, 95] decomposes human geometry and appearance into semantically distinct components (face, hair, and clothing) for modular and controllable representations. Independent modeling enables flexible editing and cross-person manipulations, such as hairstyle or outfit transfer. PEGASUS [8] and PERSE [9] focus on generating animatable 3D personalized avatars with disentangled and editable facial attributes. DELTA [17] and TECA [95] leverage the parametric SMPL-X [54] model for body and face geometry while integrating volumetric representations [48] to capture hair and clothing. More recently, MeGA [82] represents the face using the FLAME [40] model with additional displacements and models hair with 3D Gaussians [30]. Similarly, 3DGH [25] introduces an unconditional generative model for 3D human heads with composable face and hair components using separate Gaussian primitives. Concurrent work HairCUP [31] builds a universal compositional prior by first modeling the bald head and then representing hair and face as disentangled Gaussian primitives. It uses Score Distillation Sampling (SDS) [57] to obtain bald images. In contrast, our method benefits from recent VLM models and optimizes for shared texture UV maps from sparse views. It is personalized and employs a strand-based hair representation coupled with structured Gaussians, enabling physically plausible hair animation through a physics engine.

2.3. Hair Modeling and Simulation

Strands are a widely used representation for hairstyles in modern computer graphics engines, essential for realistic rendering, editing [24], and simulation [13, 24, 70, 99]. However, creating accurate strand-based hair geometry is a challenging task that typically requires skilled artists. To simplify this process, various methods have been developed to reconstruct realistic strand-based hairstyles from monocular video [45, 60, 67, 72, 88, 94], from multi-view images [35, 49, 100], from single images [26, 61, 81, 87, 96, 98], and even from CT scans [66]. For hairstyle simulation, earlier works such as [68, 94] leverage Unreal Engine [15] to generate realistic hair motion. However, these approaches often derive the appearance model directly from the simulator, which limits their ability to capture person-specific characteristics. Data-driven methods [42, 43, 85, 86] model hair motion by learning from multi-view images. While effective, these approaches are often hairstyle-specific, require large and diverse training datasets, and may fail in out-of-distribution scenarios. In addition, most do not incorporate dynamic physics-based priors, which can result in unrealistic motion and inaccurate collision handling. GroomGen [99] trains a neural hair simulator using synthetic data. In contrast, Quaffure [70] employs physics-based self-supervised losses [63, 64, 71], removing the need for data generation. Although these methods achieve realistic results and generalize well across different poses and hairstyles, they are not personalized and do not explicitly model hair appearance.

3. Method

Our method takes multiview video to learn facial expressions, along with a 360-degree static capture of a human head for hair reconstruction. From this data, we reconstruct a disentangled, animatable head avatar that can change expressions, be rendered from different camera views, and support physics-based animations through an attached physics simulator. Figure 2 shows an overview of our method. As described in Section 3.1, we propose a layered optimization to disentangle the face from the hair. Our representation consists of two key components: a strand-based hair H and a dynamic face module D, where Gaussian primitives are attached to the parametric face model FLAME [40] and move consistently with different expressions, similar to recent work Gaussian Avatars (GA) [59]. As described in Section 3.4, we use photometric and silhouette-based losses and we regularize the appearance of unseen hair Gaussians using nearest neighbor strands. After obtaining the appearance model, we use guiding hair strands and a physics engine to simulate head pose-dependent effects and transfer the motion to the dense 3D

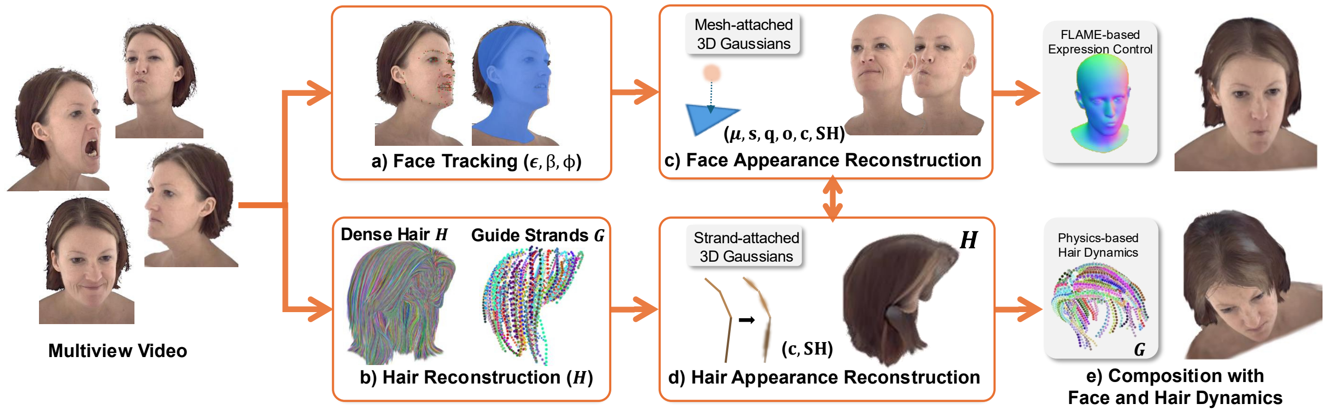

Figure 2. **Overview.** PhysHead reconstructs an animatable 3D human head avatar (e) from a multiview input video. It is based on a 3D Gaussian appearance representation that is split into a face (c) and a hair region (d). The face region uses 3D Gaussians that are attached to a 3DMM-based mesh (FLAME [40]), which allows for parametric facial expression as well as head pose control (a,c). To enable physics-based animation of the hair region, we rely on a strand-based hair model (b). The appearance of the individual hair strands is represented as structured 3D Gaussians attached to each hair strand segment (d).

Gaussian primitives, see Section 3.3.

3.1. Layered Representation of Human Head

An essential aspect of our method is to disentangle the face and hair from real-world captures. This is challenging, as the exact separation between head and hair in real data is unknown, and image captures show head and hair together. Modeling head with a single layer of 3D Gaussian primitives is insufficient for disentanglement and results in artifacts, such as skin peeling off when animating the hair. To achieve a clean layered representation, we propose a two-stage optimization of the Gaussians. In the first stage, only the regions represented by a tracked FLAME [40] mesh are optimized, while regions outside FLAME's representation, such as hair, are optimized in the second stage. In the following, we detail the two layers, namely the expression-dependent head part and the physics-driven hair part.

Expression-dependent Head Part D is represented by a set of 3D Gaussians, denoted as follows:

$$\mathcal{G}_k = \{\mu_k, s_k, q_k, o_k, c_k, \mathrm{SH}_k\}$$

The parameters are mean position μ_k, scale s_k, rotation q_k represented as a quaternion, opacity o_k, color c_k, and spherical harmonics coefficients SH_k. Following Gaussian Avatars (GA) [59], we attach the 3D Gaussian primitives to the FLAME mesh. Instead of optimizing the global positions of the Gaussians, we attach them to the FLAME triangles and move the triangles, allowing them to move with FLAME's expression and motion parameters. This binding strategy is simple but effective for modeling the face part, as it allows us full control of the head avatar using FLAME.

Hair Part H is represented as a collection of strands, initially obtained by NeuralHaircut [67]. This gives us the vertices of the hair strand polylines defined as $H \in \mathbb{R}^{N_d \times N_{\text{seg}} \times 3}$, where H represents the dense set of hair strands, N_d is the number of dense strands, and N_{seg} is the number of segments per strand.

This strand-based reconstruction results in a large number of points. These points are not uniformly distributed along the curves, which causes problems for simulation. To address this, we redistribute the points along the strand and uniformly sample $m = 60,000$ 3D hair strands and $n = 16$ points per strand while preserving the strand shape.

We assign a 3D Gaussian primitive to each hair strand segment following [45, 94, 100] and use Frenet–Serret frames (TNB frames) to compute their rotations ($\mathbf{g}_{\mathbf{R}} = [TNB]$) using the tangent $\mathbf{T}$, normal $\mathbf{N}$, and binormal $\mathbf{B}$. Given two points p_1 and p_2 of a segment, we define an elongated Gaussian with the following properties:

$$\mathbf{g}_{\text{mean}}(p_1, p_2) = (p_1 + p_2)/2, \mathbf{g}_{\text{scale}} = (\|p_2 - p_1\|, k, k)$$

where $k = 0.0001$. These Gaussians are splatted using a differentiable tile-based rasterizer [30, 106] and are supervised by ground truth images. In addition to color and perceptual losses, we use a strand-level color consistency loss to regularize the appearance of unseen hair, see Section 3.4.

3.2. Generation of Bald Training Images

To train the head layer, we remove hair from the training sequences using VLM-based editing [11], shown in Figure 3. Specifically, we use Nano-Banana [11] to automatically remove the hair from the first frame and filter views that are not multiview consistent. Based on these sparse multiview

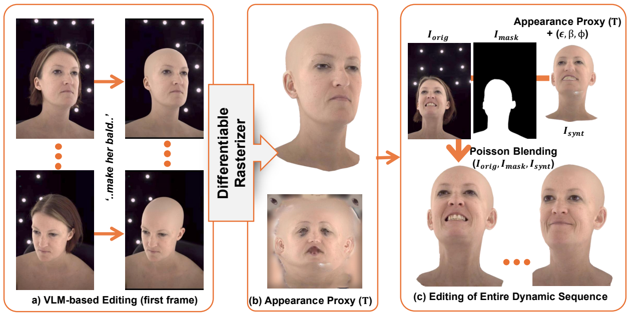

Figure 3. We use a VLM-based editing to remove the hair from the first frame of the multi-view input sequence (a). From this, we construct an appearance proxy (b) based on differentiable rendering of the FLAME head model. We use this appearance proxy to remove the hair part in the entire dynamic sequence (c) by employing Poisson image editing [56]. These new images are used as targets for the Gaussian splatting based appearance optimization of the head layer.

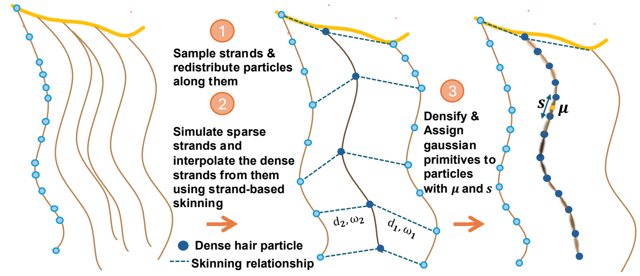

Figure 4. Using strand-skinning [10], the dense strands are affected by $k = 10$ sparse strands. The skinning weights are inversely proportional to the distance from the sparse hair strands. Specifically, if d_1 is smaller than d_2 in the figure, then ω_1 is larger than ω_2.

consistent views, we use a differentiable rasterizer [36] to optimize a shared texture map of the FLAME model, referred to as $T \in \mathbb{R}^{2048\times 2048\times 3}$.

As the FLAME model [40] does not properly represent the geometry of the mouth interior, the textures also might contain artifacts in these regions (e.g., baked-in teeth on the lips, etc.). Despite this, texture map T serves as an appearance proxy of unobserved regions, i.e., regions covered by hair such as the ears or the scalp, as these regions do not deform a lot compared to the facial expressions. To this end, we render the appearance proxy T with the head pose and expression of a particular video frame and blend it with the face region of the original image using dilated and blurred segmentation masks. We employ a blending scheme that is based on Poisson image editing [56, 89]. In contrast to HairCUP [31], this scheme allows optimization for a head layer without strong boundary artifacts and generalizes to diverse skin tones.

3.3. Simulation of Hair Strands

For simulation of the hair, we use sparse guiding strands attached to the scalp, and create a hair particle system. We use tracked FLAME head poses to guide our character. In our framework, we employ a physics engine [12] that integrates rigid-body dynamics using a semi-implicit Euler method. Newton's second law governs the translational dynamics:

$$\frac{d\mathbf{v}}{dt} = \frac{\mathbf{F}}{m}, \quad \frac{d\mathbf{x}}{dt} = \mathbf{v}.$$

with discrete updates given by:

$$\mathbf{v}(t+\Delta t) = \mathbf{v}(t) + \frac{\mathbf{F}(t)}{m}\,\Delta t, \tag{1}$$

$$\mathbf{x}(t+\Delta t) = \mathbf{x}(t) + \mathbf{v}(t+\Delta t)\,\Delta t. \tag{2}$$

An iterative constraint solver is utilized to resolve collisions and to enforce joint or contact constraints, ensuring a stable and realistic simulation. Here, $\mathbf{x}$ is the particle position, $\mathbf{v}$ is the velocity, and $\mathbf{F}$ is the applied force vector.

From sparse simulation to a dense hairstyle. The physics simulator [6] gives us sparse segment locations per frame that require interpolating guide strands to complete the groom [46]. Between the dense hair strands and the sparse guiding hair strands, we find the k-nearest neighbor strands and transfer only the relative displacement instead of the absolute locations of the guiding hair particles, see Figure 4.

3.4. Optimization and Regularization

First, we only optimize the head appearance layer with the generated training sequence (Section 3.2). Then, both the head and the hair appearance layer are rendered jointly, and only the hair part is optimized to match the original training data, see Figure 2. For the head layer, we optimize FLAME-attached 3D Gaussians similar to [59]. However, we incorporate facial masks and ignore the regions which FLAME cannot represent (i.e. hair). Specifically, we adapted the tracker from GA [59] to only track the facial part without hair. The optimization of the 3D Gaussian primitives is supervised with a combination of $\mathcal{L}_1$ and D-SSIM terms.

$$\mathcal{L}_{\text{rgb}} = (1-\lambda)\cdot\mathcal{L}_1 + \lambda\cdot\mathcal{L}_{\text{D-SSIM}}, \quad \text{with} \quad \lambda = 0.2. \tag{3}$$

Similar to GA [59], to reduce spiky Gaussian primitives, we apply the position and scaling regularizations, resulting in the loss function for the bald head part:

$$\mathcal{L} = \mathcal{L}_{\text{rgb}} + \lambda_{\text{pos}}\mathcal{L}_{\text{pos}} + \lambda_{\text{scaling}}\mathcal{L}_{\text{scaling}}. \tag{4}$$

From this optimization stage, we obtain the head layer, which is denoted as D, where all 3D Gaussian parameters $\mathcal{G}_k = \{\mu_k, s_k, q_k, o_k, c_k, \text{SH}_k\}$ are optimized for.

To recover the hair appearance, we initialize the geometry with NeuralHaircut [67]. We begin optimizing hair appearance using the photometric rendering loss for 3000 iterations. As we have strand-based representation and hair is animatable by a physics-based engine, having skin color in the hair region is undesired. To do so, we compute the loss 5 only on the masked hair region. This prevents having fewer artifacts near hairtips, in contrast to [31, 94]. Given M_{hair} is the binary hair mask, I^{rend} is rendered 3D Gaussians, I^{gt} ground truth image, formally the loss is defined as:

$$\mathcal{L}_{\text{rgb}} = \left\| \left(I^{\text{rend}} - I^{\text{gt}} \right) \odot M_{\text{hair}} \right\|_1 . \tag{5}$$

While this recovers the appearance of the outer hair layer, it leads to arbitrary colors in unobserved regions—such as back views or interior strands—due to the lack of supervision. To mitigate random colors of hidden strands, we introduce a regularization that encourages neighboring strands $j \in \mathcal{N}(i)$ of each strand $i \in \mathcal{S}$ to have similar colors. The hair appearance is being diffused into unseen regions with:

$$\mathcal{L}_{\text{consistency}} = \sum_{i \in \mathcal{S}} \sum_{j \in \mathcal{N}(i)} \|\mathbf{c}_i - \mathbf{c}_j\|_2^2 , \tag{6}$$

where $\mathbf{c}_i \in \mathbb{R}^{49 \times 3}$ is the color of strand i. The final objective for hair appearance optimization is:

$$\mathcal{L}_{\text{hair}} = \mathcal{L}_{\text{rgb}} + \lambda_{\text{consistency}} \mathcal{L}_{\text{consistency}} . \tag{7}$$

During the optimization, the 3D Gaussian parameters s_k, and q_k are initialized and kept fixed from the TNB frames [84], $o_k = 1$, and only c_k and SH$_k$ are optimized.

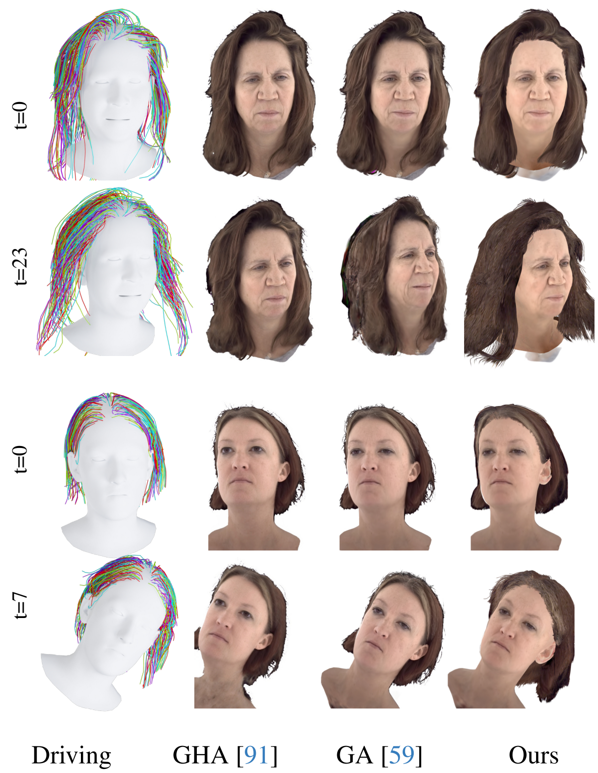

Figure 5. **Qualitative Comparison** All methods synthesize photorealistic avatars. However, both GHA [91] and GA [59] have rigid hair when animated, as Gaussians are unstructured. In contrast, our method, PhysHead, benefitting from strand representation, generalizes under novel driving signals.

4. Results

Dataset. We use the Ava-256 dataset [47] with 16 cameras per subject, selected via the Hungarian algorithm to match the Nersemble dataset [32] distribution. Our subset includes 8 expressions with one sequence held out for expression synthesis and one frontal camera for novel view synthesis. See suppl. doc. for actor details.

Baselines. *Gaussian Avatars (GA) [59]* represents appearance using FLAME-rigged Gaussian primitives, with control driven directly by the FLAME model. It is computationally efficient since no neural network is queried for appearance. However, the hair geometry is modeled through 3DMM offsets, resulting in rigid and non-dynamic hair.

Table 1. Comparison to GaussianHaircut [94]. Green indicates the best and yellow indicates the second.

	PSNR↑	SSIM↑	LPIPS↓
GH [94]	30.55	0.915	0.071
Ours w/$\mathcal{L}_{\text{consistency}}$	28.24	0.893	0.078
Ours	30.28	0.946	0.061

Gaussian Head Avatars (GHA) [91] is a Gaussian-based head avatar approach comprising two stages. First, a mesh head is optimized at each time step from multiview videos. Then, a deformation MLP learns temporal deformations, followed by a super-resolution module for refinement. GHA uses BFM [55] for expression control and produces rigid hair, as hair geometry is not explicitly modeled.

Gaussian Haircut (GH) [94] optimizes hair appearance by attaching Gaussian primitives to the midpoints of hair segments and using photometric losses combined with orientation-based regularization. Our method builds on Neural Haircut by leveraging strand-based hair geometry and optimizing appearance to achieve realistic hair rendering during physical simulations.

HairCUP [31], a concurrent universal compositional avatar method built on UrAvatar [39], also models the head and hair separately, enabling hair transfer in addition to expression control. However, bald avatars synthesized by HairCUP exhibit artifacts in unseen regions. Moreover, although the representation is decompositional, the hair is modeled as unstructured Gaussians, leading to rigid and non-dynamic behavior.

Figure 6. **GaussianHaircut [94] comparison.** As can be seen, GaussianHaircut is not yet suitable for simulation with learned appearance, as some hair strands penetrate the head and are assigned skin colors. In contrast, our reconstructed hair strands exhibit more consistent coloring due to the proposed consistency loss and can be used directly in simulation.

4.1. Comparison

Hair reconstruction evaluation. We evaluate the quality of our obtained hair appearance model in comparison to GaussianHaircut. In Figure 6, we show novel views of the hair region, which is represented with hair strands and attached 3D Gaussian primitives. As can be seen, GaussianHaircut has hair strands inside the head region that learn skin color, which restricts it from being animated through simulation with learned colors. The overall static appearance of the GaussianHaircut reconstruction and ours is similar, see Table 1.

Cross-reenactment evaluation. In Figure 5 we show a comparison to GHA [91] and GA [59]. When there is not much hair motion, the baseline methods show similar quality as our method (t=0). However, the difference gets clear in a motion sequence that results in dynamically moving hair. As our method, Physhead completes missing parts during capture; it generalizes to very diverse motions such as rotating head, nodding, effects like wind while baseline methods remains with rigid hair.

Comparison of layered avatar. As shown in Figure 7, the concurrent work HairCUP is also compositional. Despite being realistic when head and hair layer are combined, the head layer has artifacts due to the underlying SDS-based procedure. In contrast, our method utilizes an appearance proxy generated with a vision language model and differentiable rendering which is then blended with the original image content to faithfully preserve its details and skin color.

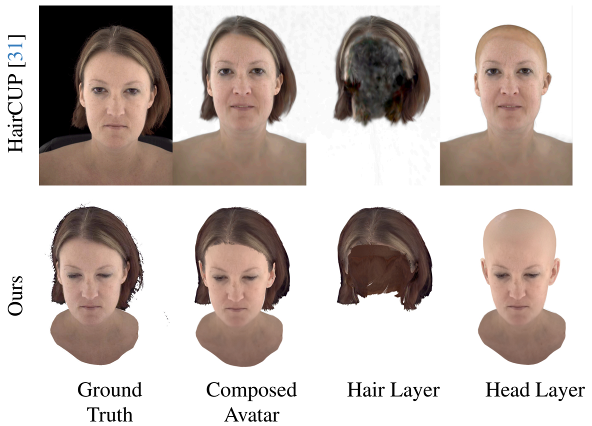

Figure 7. **HairCUP [31] comparison.** Since haircup uses SDS, it exhibits artifacts and has tendency to generate the same type of skin color. In contrast, we use VLM model for image generation which is trained on a lot of data. This helps with generalization to different skin colors. Despite being compositional, it has unstructured gaussians which does not allow for physics simulation. For more results, see supp mat. *HairCUP images taken from their original paper.*

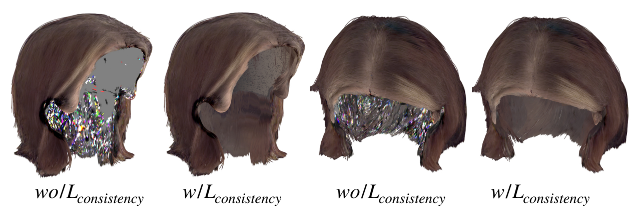

Figure 8. The consistency loss helps to assign meaningful colors to hidden 3D gaussian primitives.

4.2. Ablation studies

In the supplemental document, we show ablation studies with respect to our proposed layered representation. It is key for being able to simulate the hair. A single-layer approach would lead to artifacts such as skin peeling. In addition, hidden regions of skin and hair would not be handled well. Besides the layered representation, we have introduced a color consistency regularization. Without this regularization, the inner hair strands will have random colors, restricting it from being animated. We enable this regularization for the hair strands after 5k iterations. The intuition behind this is that the visible strands are initially optimized based on the RGB images, allowing their colors to propagate naturally to the hidden strands, see Figure 8.

4.3. Applications

Our method uses dense explicit strands, which consist of connected 3D points. Thus, it can be used in several applications, such as hairstyle transfer Figure 10, hair geometry and appearance editing Figure 9.

Figure 9. The strand-based hair layer can be edited by the user to change the geometry and appearance.

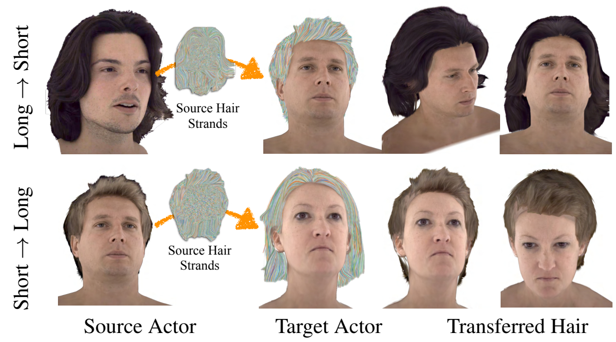

Figure 10. The layered avatar representation allows for editing operations like hair-swapping.

4.4. Limitations

While PhysHead is able to produce physically plausible hair dynamics for animatable 3D head avatars, the appearance quality is dependent on the quality of the foreground and hair masks. Imperfect masks might result in wrong disentanglement between hair and head. In addition, PhysHead relies on the hair geometry quality of NeuralHaircut [67], thus, inheriting its limitations (e.g., curly hair).

5. Conclusion

In this paper, we have presented the first method that combines physics-driven hair dynamics with an animatable 3D Gaussian-splatting-based head avatar. It is enabled by a layered representation that disentangles the face region from hair. The face region is modeled with 3D Gaussian primitives that are attached to a 3DMM which allows for facial expression control. To handle the hair region and its dynamics that are mainly dependent on the head pose, a strand-based hairstyle reconstruction is executed to provide a representation that is compatible to physics simulation. On top of this strand-based representation, 3D Gaussian primitives are attached which are optimized to reproduce the appearance of the hair from the input images. This layered representation allows us to demonstrate physically plausible hair motion, as well as hair editing. We believe that PhysHead is a stepping stone towards highly realistic human head avatars that generalize to novel poses.

6. Acknowledgments

We thank Y. Feng, J. E. Lenssen, D. Eskandar, S. Qian, S. Giebenhain, K. Cetinkaya, N. Kister, G. Tiwari, and S. Bharadwaj for helpful discussions and help during the project. Berna Kabadayi is supported by the International Max Planck Research School for Intelligent Systems (IMPRS-IS). 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 – Project No. 533677015. Gerard Pons-Moll is a member of the Machine Learning Cluster of Excellence, EXC number 2064/1 – Project No. 390727645 and is endowed by the Carl Zeiss Foundation. This work was supported by the German Federal Ministry of Education and Research (BMBF): Tübingen AI Center, FKZ: 01IS18039A and funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – 409792180 (Emmy Noether Programme, project: Real Virtual Humans).

References

[1] *LitNeRF: Intrinsic Radiance Decomposition for High-Quality View Synthesis and Relighting of Faces*, 2023. 2

[2] Sizhe An, Hongyi Xu, Yichun Shi, Guoxian Song, Umit Y. Ogras, and Linjie Luo. PanoHead: Geometry-aware 3D full-head synthesis in 360°. *CVPR*, pages 20950–20959, 2023. 2

[3] Shivangi Aneja, Sebastian Weiss, Irene Baeza, Prashanth Chandran, Gaspard Zoss, Matthias Niessner, and Derek Bradley. Scaffoldavatar: High-fidelity gaussian avatars with patch expressions. In *Proceedings of the Special Interest Group on Computer Graphics and Interactive Techniques Conference Conference Papers*, pages 1–11, 2025. 2

[4] ShahRukh Athar, Zexiang Xu, Kalyan Sunkavalli, Eli Shechtman, and Zhixin Shu. RigNeRF: Fully controllable neural 3d portraits. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 20364–20373, 2022. 2

[5] ShahRukh Athar, Zhixin Shu, and Dimitris Samaras. FLAME-in-NeRF: Neural control of radiance fields for free view face animation. In *IEEE International Conference on Automatic Face and Gesture Recognition (FG)*, 2023. 2

[6] Autodesk, INC. Maya. 2, 5

[7] Volker Blanz and Thomas Vetter. A morphable model for the synthesis of 3D faces. In *SIGGRAPH*, pages 187–194, 1999. 2

[8] Hyunsoo Cha, Byungjun Kim, and Hanbyul Joo. Pegasus: Personalized generative 3d avatars with composable attributes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1072–1081, 2024. 3

[9] Hyunsoo Cha, Inhee Lee, and Hanbyul Joo. Perse: Personalized 3d generative avatars from a single portrait. In *Proceedings of the Computer Vision and Pattern Recognition Conference (CVPR)*, pages 15953–15962, 2025. 3

[10] Menglei Chai, Changxi Zheng, and Kun Zhou. Adaptive skinning for interactive hair-solid simulation. *IEEE transactions on visualization and computer graphics*, 23 (7):1725–1738, 2016. 5

[11] Gheorghe Comanici, Eric Bieber, Mike Schaekermann, Ice Pasupat, Noveen Sachdeva, et al. Gemini 2.5: Pushing the frontier with advanced reasoning, multimodality, long context, and next generation agentic capabilities. *arXiv preprint arXiv:2507.06261*, 2025. 2, 4

[12] Erwin Coumans. Bullet physics simulation. In *ACM SIGGRAPH 2015 Courses*, page 1, 2015. 2, 5

[13] Gilles Daviet. Interactive hair simulation on the gpu using admm. *ACM SIGGRAPH 2023 Conference Proceedings*, 2023. 3

[14] Bernhard Egger, William A. P. Smith, Ayush Tewari, Stefanie Wuhrer, Michael Zollhoefer, Thabo Beeler, Florian Bernard, Timo Bolkart, Adam Kortylewski, Sami Romdhani, Christian Theobalt, Volker Blanz, and Thomas Vetter. 3D morphable face models—past, present, and future. *ACM TOG*, 39(5), 2020. 2

[15] Epic Games. Unreal engine. 2, 3

[16] Zhiwen Fan, Kevin Wang, Kairun Wen, Zehao Zhu, Dejia Xu, and Zhangyang Wang. Lightgaussian: Unbounded 3d gaussian compression with 15x reduction and 200+ fps, 2023. 2

[17] Yao Feng, Weiyang Liu, Timo Bolkart, Jinlong Yang, Marc Pollefeys, and Michael J. Black. Learning disentangled avatars with hybrid 3d representations. *arXiv*, 2023. 1, 3

[18] Yutao Feng, Xiang Feng, Yintong Shang, Ying Jiang, Chang Yu, Zeshun Zong, Tianjia Shao, Hongzhi Wu, Kun Zhou, Chenfanfu Jiang, and Yin Yang. Gaussian splashing: Dynamic fluid synthesis with gaussian splatting, 2024. 2

[19] Guy Gafni, Justus Thies, Michael Zollhofer, and Matthias Nießner. Dynamic neural radiance fields for monocular 4D facial avatar reconstruction. *CVPR*, pages 8645–8654, 2020. 2

[20] Ruiqi Gao*, Aleksander Holynski*, Philipp Henzler, Arthur Brussee, Ricardo Martin-Brualla, Pratul P. Srinivasan, Jonathan T. Barron, and Ben Poole*. Cat3d: Create anything in 3d with multi-view diffusion models. *Advances in Neural Information Processing Systems*, 2024. 3

[21] Xuan Gao, Chenglai Zhong, Jun Xiang, Yang Hong, Yudong Guo, and Juyong Zhang. Reconstructing personalized semantic facial nerf models from monocular video. *ACM Transactions on Graphics (Proceedings of SIGGRAPH Asia)*, 41(6), 2022. 2

[22] Stephan J. Garbin, Marek Kowalski, Virginia Estellers, Stanislaw Szymanowicz, Shideh Rezaeifar, Jingjing Shen, Matthew Johnson, and Julien Valentin. VolTeMorph: Realtime, controllable and generalizable animation of volumetric representations. *Computer Graphics Forum*, 2024. To appear. 2

[23] Simon Giebenhain, Tobias Kirschstein, Martin Rünz, Lourdes Agapito, and Matthias Nießner. Npga: Neural parametric gaussian avatars, 2024. 1, 2

[24] Chengan He, Jorge Alejandro Amador Herrera, Yi Zhou, Zhixin Shu, Xin Sun, Yao Feng, Sören Pirk, Dominik L

Michels, Meng Zhang, Tuanfeng Y Wang, and Holly Rushmeier. Digital salon: An ai and physics-driven tool for 3d hair grooming and simulation. *ACM SIGGRAPH Asia 2024 Real-Time Live!*, 2024. 3

[25] Chengan He, Junxuan Li, Tobias Kirschstein, Artem Sevastopolsky, Shunsuke Saito, Qingyang Tan, Javier Romero, Chen Cao, Holly Rushmeier, and Giljoo Nam. 3dgh: 3d head generation with composable hair and face. *ACM Transactions on Graphics*, 44(4):1–12, 2025. 3

[26] Chengan He, Xin Sun, Zhixin Shu, Fujun Luan, Sören Pirk, Jorge Alejandro Amador Herrera, Dominik L. Michels, Tuanfeng Y. Wang, Meng Zhang, Holly Rushmeier, and Yi Zhou. Perm: A parametric representation for multi-style 3d hair modeling. In *International Conference on Learning Representations*, 2025. 3

[27] Berna Kabadayi, Wojciech Zielonka, Bharat Lal Bhatnagar, Gerard Pons-Moll, and Justus Thies. Gan-avatar: Controllable personalized gan-based human head avatar. In *2024 International Conference on 3D Vision (3DV)*, pages 882–892. IEEE, 2024. 2

[28] Kacper Kania, Kwang Moo Yi, Marek Kowalski, Tomasz Trzciński, and Andrea Tagliasacchi. CoNeRF: Controllable Neural Radiance Fields. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 20314–20323, 2022.

[29] Kacper Kania, Stephan J. Garbin, Andrea Tagliasacchi, Virginia Estellers, Kwang Moo Yi, Julien Valentin, Tomasz Trzciński, and Marek Kowalski. BlendFields: Few-Shot Example-Driven Facial Modeling. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2023. 2

[30] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkuehler, and George Drettakis. 3D gaussian splatting for real-time radiance field rendering. *ACM TOG*, 42:1 – 14, 2023. 2, 3, 4

[31] Byungjun Kim, Shunsuke Saito, Giljoo Nam, Tomas Simon, Jason Saragih, Hanbyul Joo, and Junxuan Li. Haircup: Hair compositional universal prior for 3d gaussian avatars. *arXiv preprint arXiv:2507.19481*, 2025. 1, 2, 3, 5, 6, 8

[32] Tobias Kirschstein, Shenhan Qian, Simon Giebenhain, Tim Walter, and Matthias Nießner. NeRSemble: Multi-view radiance field reconstruction of human heads. *ACM TOG*, 42: 1 – 14, 2023. 6

[33] Tobias Kirschstein, Simon Giebenhain, Jiapeng Tang, Markos Georgopoulos, and Matthias Nießner. Gghead: Fast and generalizable 3d gaussian heads, 2024. 2

[34] Tobias Kirschstein, Javier Romero, Artem Sevastopolsky, Matthias Nießner, and Shunsuke Saito. Avat3r: Large animatable gaussian reconstruction model for high-fidelity 3d head avatars, 2025. 3

[35] Zhiyi Kuang, Yiyang Chen, Hongbo Fu, Kun Zhou, and Youyi Zheng. DeepMVSHair: Deep hair modeling from sparse views. In *SIGGRAPH Asia 2022 Conference Papers*, New York, NY, USA, 2022. Association for Computing Machinery. 3

[36] Samuli Laine, Janne Hellsten, Tero Karras, Yeongho Seol, Jaakko Lehtinen, and Timo Aila. Modular primitives for high-performance differentiable rendering. *ACM Transactions on Graphics*, 39(6), 2020. 5

[37] Steve Lesser, Alexey Stomakhin, Gilles Daviet, Joel Wretborn, John Edholm, Noh-Hoon Lee, Eston Schweickart, Xiao Zhai, Sean Flynn, and Andrew Moffat. Loki: a unified multiphysics simulation framework for production. *ACM Trans. Graph.*, 41(4), 2022. 2

[38] Gengyan Li, Kripasindhu Sarkar, Abhimitra Meka, Marcel Buehler, Franziska Mueller, Paulo Gotardo, Otmar Hilliges, and Thabo Beeler. ShellNeRF: Learning a Controllable High-resolution Model of the Eye and Periocular Region. *Computer Graphics Forum*, 2024. 2

[39] Junxuan Li, Chen Cao, Gabriel Schwartz, Rawal Khirodkar, Christian Richardt, Tomas Simon, Yaser Sheikh, and Shunsuke Saito. Uravatar: Universal relightable gaussian codec avatars. In *ACM SIGGRAPH 2024 Conference Papers*, 2024. 6

[40] Tianye Li, Timo Bolkart, Michael. J. Black, Hao Li, and Javier Romero. Learning a model of facial shape and expression from 4D scans. 36(6):194:1–194:17, 2017. 2, 3, 4, 5

[41] Zhe Li, Zerong Zheng, Lizhen Wang, and Yebin Liu. Animatable gaussians: Learning pose-dependent gaussian maps for high-fidelity human avatar modeling. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024. 3

[42] Zhanfeng Liao, Yuelang Xu, Zhe Li, Qijing Li, Boyao Zhou, Ruifeng Bai, Di Xu, Hongwen Zhang, and Yebin Liu. Hhavatar: Gaussian head avatar with dynamic hairs. *arXiv e-prints*, pages arXiv–2312, 2023. 2, 3

[43] 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*, pages 12318–12327, 2025. 2, 3

[44] Stephen Lombardi, Tomas Simon, Gabriel Schwartz, Michael Zollhoefer, Yaser Sheikh, and Jason M. Saragih. Mixture of volumetric primitives for efficient neural rendering. *ACM TOG*, 40:1 – 13, 2021. 2

[45] 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. 3, 4

[46] Qing Lyu, Menglei Chai, Xiang Chen, and Kun Zhou. Real-time hair simulation with neural interpolation. *IEEE Transactions on Visualization and Computer Graphics*, 28(4): 1894–1905, 2020. 5

[47] Julieta Martinez, Emily Kim, Javier Romero, Timur Bagautdinov, Shunsuke Saito, Shoou-I Yu, Stuart Anderson, Michael Zollhöfer, Te-Li Wang, Shaojie Bai, Chenghui Li, Shih-En Wei, Rohan Joshi, Wyatt Borsos, Tomas Simon, Jason Saragih, Paul Theodosis, Alexander Greene, Anjani Josyula, Silvio Mano Maeta, Andrew I. Jewett, Simon Venshtain, Christopher Heilman, Yueh-Tung Chen, Sidi Fu, Mohamed Ezzeldin A. Elshaer, Tingfang Du, Longhua Wu, Shen-Chi Chen, Kai Kang, Michael Wu,

Youssef Emad, Steven Longay, Ashley Brewer, Hitesh Shah, James Booth, Taylor Koska, Kayla Haidle, Matt Andromalos, Joanna Hsu, Thomas Dauer, Peter Selednik, Tim Godisart, Scott Ardisson, Matthew Cipperly, Ben Humberston, Lon Farr, Bob Hansen, Peihong Guo, Dave Braun, Steven Krenn, He Wen, Lucas Evans, Natalia Fadeeva, Matthew Stewart, Gabriel Schwartz, Divam Gupta, Gyeongsik Moon, Kaiwen Guo, Yuan Dong, Yichen Xu, Takaaki Shiratori, Fabian Prada, Bernardo R. Pires, Bo Peng, Julia Buffalini, Autumn Trimble, Kevyn McPhail, Melissa Schoeller, and Yaser Sheikh. Codec Avatar Studio: Paired Human Captures for Complete, Driveable, and Generalizable Avatars. *NeurIPS Track on Datasets and Benchmarks*, 2024. 6

[48] Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. NeRF: Representing scenes as neural radiance fields for view synthesis. In *ECCV*, pages 405–421, 2020. 3

[49] 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*, pages 155–164, 2019. 3

[50] KL Navaneet, Kossar Pourahmadi Meibodi, Soroush Abbasi Koohpayegani, and Hamed Pirsiavash. Compgs: Smaller and faster gaussian splatting with vector quantization. *ECCV*, 2024. 3

[51] Jalees Nehvi, Berna Kabadayi, Julien Valentin, and Justus Thies. Volumetric portrait avatar. In *Pattern Recognition: 46th DAGM German Conference, DAGM GCPR 2024, Munich, Germany, September 10–13, 2024, Proceedings, Part II*, page 3–19, 2025. 2

[52] Mirela Ostrek, Michael J. Black, and Justus Thies. Hairfree: Compositional 2d head prior for text-driven 360° bald texture synthesis. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025. 3

[53] Haokai Pang, Heming Zhu, Adam Kortylewski, Christian Theobalt, and Marc Habermann. ASH: animatable gaussian splats for efficient and photoreal human rendering. In *CVPR*, pages 1165–1175. IEEE, 2024. 3

[54] Georgios Pavlakos, Vasileios Choutas, Nima Ghorbani, Timo Bolkart, Ahmed A. A. Osman, Dimitrios Tzionas, and Michael J. Black. Expressive body capture: 3D hands, face, and body from a single image. In *CVPR*, pages 10975–10985, 2019. 3

[55] P. Paysan, R. Knothe, B. Amberg, S. Romdhani, and T. Vetter. A 3D face model for pose and illumination invariant face recognition. *IEEE International Conference on Advanced Video and Signal based Surveillance (AVSS)*, 2009. 6

[56] Patrick Pérez, Michel Gangnet, and Andrew Blake. Poisson image editing. In *ACM SIGGRAPH 2003 Papers*, page 313–318, New York, NY, USA, 2003. Association for Computing Machinery. 5

[57] Ben Poole, Ajay Jain, Jonathan T. Barron, and Ben Mildenhall. Dreamfusion: Text-to-3d using 2d diffusion. *arXiv*, 2022. 3

[58] Shenhan Qian. Vhap: Versatile head alignment with adaptive appearance priors, 2024. 2

[59] Shenhan Qian, Tobias Kirschstein, Liam Schoneveld, Davide Davoli, Simon Giebenhain, and Matthias Nießner. GaussianAvatars: Photorealistic head avatars with rigged 3D gaussians. In *CVPR*, pages 20299–20309, 2024. 1, 3, 4, 5, 6, 7

[60] 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 *Computer Vision – ECCV 2018: 15th European Conference*, 2022. 3

[61] Radu Alexandru Rosu, Keyu Wu, Yao Feng, Youyi Zheng, and Michael J. Black. DiffLocks: Generating 3d hair from a single image using diffusion models. In *IEEE/CVF Conf. on Computer Vision and Pattern Recognition(CVPR)*, 2025. 3

[62] Shunsuke Saito, Gabriel Schwartz, Tomas Simon, Junxuan Li, and Giljoo Nam. Relightable gaussian codec avatars. In *CVPR*, pages 130–141, 2024. 3

[63] Igor Santesteban, Nils Thuerey, Miguel A. Otaduy, and Dan Casas. Self-supervised collision handling via generative 3d garment models for virtual try-on. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 11763–11773, 2021. 3

[64] Igor Santesteban, Miguel A. Otaduy, and Dan Casas. Snug: Self-supervised neural dynamic garments. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 8140–8150, 2022. 3

[65] Zhijing Shao, Zhaolong Wang, Zhuang Li, Duotun Wang, Xiangru Lin, Yu Zhang, Mingming Fan, and Zeyu Wang. SplattingAvatar: Realistic Real-Time Human Avatars with Mesh-Embedded Gaussian Splatting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024. 3

[66] Yuefan Shen, Shunsuke Saito, Ziyan Wang, Olivier Maury, Chenglei Wu, Jessica Hodgins, Youyi Zheng, and Giljoo Nam. Ct2hair: High-fidelity 3d hair modeling using computed tomography. *ACM Transactions on Graphics*, 42(4): 1–13, 2023. 3

[67] 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)*, 2023. 3, 4, 6, 8

[68] Vanessa Sklyarova, Egor Zakharov, Otmar Hilliges, Michael J. Black, and Justus Thies. Text-conditioned generative model of 3d strand-based human hairstyles. In *IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2024. 2, 3

[69] Vanessa Sklyarova, Berna Kabadayi, Anastasios Yiannakidis, Giorgio Becherini, Michael J. Black, and Justus Thies. Neuralfur: Neuralfur: Animal fur reconstruction from multi-view images. *ArXiv*, 2026. 2

[70] Tuur Stuyck, Gene Wei-Chin Lin, Egor Larionov, Hsiao-yu Chen, Aljaz Bozic, Nikolaos Sarafianos, and Doug Roble. Quaffure: Real-time quasi-static neural hair simulation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2025. 3

[71] Zhaoqi Su, Liangxiao Hu, Siyou Lin, Hongwen Zhang, Shengping Zhang, Justus Thies, and Yebin Liu. Caphy: Capturing physical properties for animatable human avatars. In *IEEE/CVF International Conference on Computer Vision (ICCV)*, 2023. 3

[72] 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)*, pages 20601–20611, 2024. 3

[73] Felix Taubner, Prashant Raina, Mathieu Tuli, Eu Wern Teh, Chul Lee, and Jinmiao Huang. 3D face tracking from 2D video through iterative dense UV to image flow. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1227–1237, 2024. 2

[74] Felix Taubner, Ruihang Zhang, Mathieu Tuli, and David B. Lindell. Cap4d: Creating animatable 4d portrait avatars with morphable multi-view diffusion models, 2024. 3

[75] Kartik Teotia, Hyeongwoo Kim, Pablo Garrido, Marc Habermann, Mohamed Elgharib, and Christian Theobalt. Gaussianheads: End-to-end learning of drivable gaussian head avatars from coarse-to-fine representations. *ACM Transactions on Graphics (ToG)*, 43(6):1–12, 2024. 3

[76] Ayush Tewari, Florian Bernard, Pablo Garrido, Gaurav Bharaj, Mohamed Elgharib, Hans-Peter Seidel, Patrick Pérez, Michael Zollhofer, and Christian Theobalt. Fml: Face model learning from videos. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10812–10822, 2019. 2

[77] A. Tewari, O. Fried, J. Thies, V. Sitzmann, S. Lombardi, K. Sunkavalli, R. Martin-Brualla, T. Simon, J. Saragih, M. Nießner, R. Pandey, S. Fanello, G. Wetzstein, J.-Y. Zhu, C. Theobalt, M. Agrawala, E. Shechtman, D. B. Goldman, and M. Zollhöfer. State of the art on neural rendering. *EG*, 2020. 2

[78] Ayush Tewari, Justus Thies, Ben Mildenhall, Pratul Srinivasan, Edgar Tretschk, Yifan Wang, Christoph Lassner, Vincent Sitzmann, Ricardo Martin-Brualla, Stephen Lombardi, Tomas Simon, Christian Theobalt, Matthias Niessner, Jonathan T. Barron, Gordon Wetzstein, Michael Zollhoefer, and Vladislav Golyanik. Advances in neural rendering. *Comput. Graph. Forum*, pages 703–735, 2022. 2

[79] J. Thies, M. Zollhöfer, M. Nießner, L. Valgaerts, M. Stamminger, and C. Theobalt. Real-time expression transfer for facial reenactment. *ACM Transactions on Graphics (TOG)*, 34(6), 2015. 2

[80] Justus Thies, Michael Zollhöfer, Marc Stamminger, Christian Theobalt, and Matthias Nießner. Face2Face: Real-time face capture and reenactment of RGB videos. *CVPR*, pages 2387–2395, 2016. 2

[81] Sklyarova Vanessa, Zakharov Egor, Prinzler Malte, Becherini Giorgio, Black Michael, and Thies Justus. Im2Haircut: single-view strand-based hair reconstruction for human avatars. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, Honolulu, USA, 2025. 3

[82] Cong Wang, Di Kang, He-Yi Sun, Shen-Han Qian, Zi-Xuan 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)*, 2025. 1, 3

[83] Daoye Wang, Prashanth Chandran, Gaspard Zoss, Derek Bradley, and Paulo Gotardo. Morf: Morphable radiance fields for multiview neural head modeling. New York, NY, USA, 2022. Association for Computing Machinery. 2

[84] Wenping Wang, Bert Jüttler, Dayue Zheng, and Yang Liu. Computation of rotation minimizing frames. *ACM Transactions on Graphics (TOG)*, 27(1):1–18, 2008. 6

[85] Ziyan Wang, Giljoo Nam, Tuur Stuyck, Stephen Lombardi, Chen Cao, Jason Saragih, Michael Zollhoefer, Jessica Hodgins, and Christoph Lassner. Neuwigs: A neural dynamic model for volumetric hair capture and animation. *arXiv preprint arXiv:2212.00613*, 2022. 3

[86] 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)*, pages 6143–6154, 2022. 3

[87] Keyu Wu, Yifan Ye, Lingchen Yang, Hongbo Fu, Kun Zhou, and Youyi Zheng. Neuralhdhair: Automatic high-fidelity hair modeling from a single image using implicit neural representations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1526–1535, 2022. 2, 3

[88] Keyu Wu, Lingchen Yang, Zhiyi Kuang, Yao Feng, Xutao Han, Yuefan Shen, Hongbo Fu, Kun Zhou, and Youyi Zheng. Monohair: High-fidelity hair modeling from a monocular video. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024. 2, 3

[89] Yiqian Wu, Yong-Liang Yang, and Xiaogang Jin. Hairmapper: Removing hair from portraits using gans. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4227–4236, 2022. 5

[90] Jun Xiang, Xuan Gao, Yudong Guo, and Juyong Zhang. Flashavatar: High-fidelity head avatar with efficient gaussian embedding. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024. 3

[91] 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)*, 2024. 3, 6, 7

[92] Yuelang Xu, Lizhen Wang, Zerong Zheng, Zhaoqi Su, and Yebin Liu. 3d gaussian parametric head model. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2024. 3

[93] Li Xueting, Ye Yuan, Shalini De Mello, Gilles Daviet, Jonathan Leaf, Miles Macklin, Jan Kautz, and Umar Iqbal. Simavatar: Simulation-ready avatars with layered hair and clothing. *Arxiv*, 2024. 3

[94] Egor Zakharov, Vanessa Sklyarova, Michael J Black, Giljoo Nam, Justus Thies, and Otmar Hilliges. Human hair reconstruction with strand-aligned 3d gaussians. *ArXiv*, 2024. 2, 3, 4, 6, 7

[95] Hao Zhang, Yao Feng, Peter Kulits, Yandong Wen, Justus Thies, and Michael J. Black. Teca: Text-guided generation and editing of compositional 3d avatars. *arXiv*, 2023. 3

[96] Yujian Zheng, Zirong Jin, Moran Li, Haibin Huang, Chongyang Ma, Shuguang Cui, and Xiaoguang Han. Hairstep: Transfer synthetic to real using strand and depth maps for single-view 3d hair modeling. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 12726–12735, 2023. 2, 3

[97] Yang Zheng, Menglei Chai, Delio Vicini, Yuxiao Zhou, Yinghao Xu, Leonidas Guibas, Gordon Wetzstein, and Thabo Beeler. Groomlight: Hybrid inverse rendering for relightable human hair appearance modeling, 2025. 2

[98] Yi Zhou, Liwen Hu, Jun Xing, Weikai Chen, Han-Wei Kung, Xin Tong, and Hao Li. Hairnet: Single-view hair reconstruction using convolutional neural networks. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018. 2, 3

[99] Yuxiao Zhou, Menglei Chai, Alessandro Pepe, Markus Gross, and Thabo Beeler. Groomgen: A high-quality generative hair model using hierarchical latent representations. *ACM Trans. Graph.*, 42(6), 2023. 3

[100] Yuxiao Zhou, Menglei Chai, Daoye Wang, Sebastian Winberg, Erroll Wood, Kripasindhu Sarkar, Markus Gross, and Thabo Beeler. GroomCap: High-fidelity prior-free hair capture. *ACM Trans. Graph.*, 43(6), 2024. 3, 4

[101] Wojciech Zielonka, Timo Bolkart, and Justus Thies. Instant volumetric head avatars. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4574–4584, 2022. 2

[102] Wojciech Zielonka, Timo Bolkart, and Justus Thies. Towards metrical reconstruction of human faces. In *European Conference on Computer Vision*, 2022. 2

[103] Wojciech Zielonka, Timo Bolkart, Thabo Beeler, and Justus Thies. Gaussian eigen models for human heads. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2025. 3

[104] Wojciech Zielonka, Stephan J. Garbin, Alexandros Lattas, George Kopanas, Paulo Gotardo, Thabo Beeler, Justus Thies, and Timo Bolkart. Synthetic prior for few-shot drivable head avatar inversion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2025. 3

[105] Michael Zollhöfer, Justus Thies, Darek Bradley, Pablo Garrido, Thabo Beeler, Patrick Péerez, Marc Stamminger, Matthias Nießner, and Christian Theobalt. State of the art on monocular 3D face reconstruction, tracking, and applications. *Comput. Graph. Forum*, 37(2):523–550, 2018. 2

[106] Matthias Zwicker, Hans Rüdiger Pfister, Jeroen van Baar, and Markus H. Gross. Surface splatting. *SIGGRAPH*, pages 371–378, 2001. 4