3DGH: 3D Head Generation with Composable Hair and Face

CHENGAN HE, Yale University, USA
JUNXUAN LI, Meta Codec Avatars Lab, USA
TOBIAS KIRSCHSTEIN, Technical University of Munich, Germany
ARTEM SEVASTOPOLSKY, Technical University of Munich, Germany
SHUNSUKE SAITO, Meta Codec Avatars Lab, USA
QINGYANG TAN, Meta Codec Avatars Lab, USA
JAVIER ROMERO, Meta Codec Avatars Lab, USA
CHEN CAO, Meta Codec Avatars Lab, USA
HOLLY RUSHMEIER, Yale University, USA
GILJOO NAM, Meta Codec Avatars Lab, USA

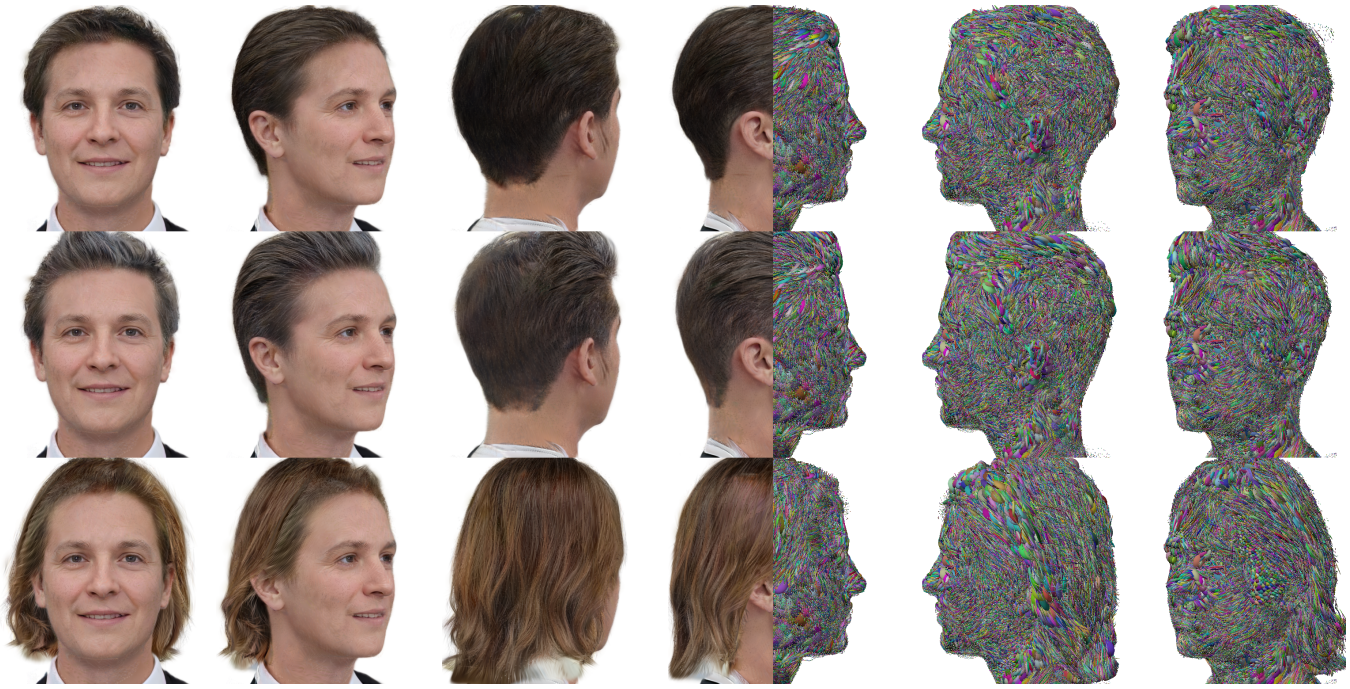


Fig. 1. 3DGH: Our method generates 3D head representations that can be rendered at 360° with photorealistic rendering quality and high-fidelity geometry based on 3D Gaussian Splatting. The generated heads include composable hair and face components, enabling 3D hairstyle editing with multi-view consistency.

Authors' Contact Information: Chengan He, Yale University, USA, chengan.he@yale.edu; Junxuan Li, Meta Codec Avatars Lab, USA, junxuanli@meta.com; Tobias Kirschstein, Technical University of Munich, Germany, tobias.kirschstein@tum.de; Artem Sevastopolsky, Technical University of Munich, Germany, artem.sevastopolsky@gmail.com; Shunsuke Saito, Meta Codec Avatars Lab, USA, shunsukesaito@meta.com; Qingyang Tan, Meta Codec Avatars Lab, USA, qytan@meta.com; Javier Romero, Meta Codec Avatars Lab, USA, javierromero1@meta.com; Chen Cao, Meta Codec Avatars Lab, USA, chencao@meta.com; Holly Rushmeier, Yale University, USA, holly.rushmeier@yale.edu; Giljoo Nam, Meta Codec Avatars Lab, USA, giljoonam@meta.com.

ACM 1557-7368/2025/8-ART59
https://doi.org/10.1145/3731211

We present 3DGH, an unconditional generative model for 3D human heads with composable hair and face components. Unlike previous work that entangles the modeling of hair and face, we propose to separate them using a novel data representation with template-based 3D Gaussian Splatting, in which deformable hair geometry is introduced to capture the geometric variations across different hairstyles. Based on this data representation, we design a 3D GAN-based architecture with dual generators and employ a cross-attention mechanism to model the inherent correlation between hair and face. The model is trained on synthetic renderings using carefully designed objectives to stabilize training and facilitate hair-face separation. We conduct extensive experiments to validate the design choice of 3DGH, and evaluate it both qualitatively and quantitatively by comparing with several state-of-the-art 3D GAN methods, demonstrating its effectiveness in unconditional

full-head image synthesis and composable 3D hairstyle editing. More details will be available on our project page: https://c-he.github.io/projects/3dgh/.

CCS Concepts: • **Computing methodologies** → **Shape representations**; **Appearance and texture representations**.

Additional Key Words and Phrases: Facial Modeling, Hair Modeling, Generative 3D Modeling

ACM Reference Format:
Chengan He, Junxuan Li, Tobias Kirschstein, Artem Sevastopolsky, Shunsuke Saito, Qingyang Tan, Javier Romero, Chen Cao, Holly Rushmeier, and Giljoo Nam. 2025. 3DGH: 3D Head Generation with Composable Hair and Face. *ACM Trans. Graph.* 44, 4, Article 59 (August 2025), 12 pages. https://doi.org/10.1145/3731211

1 Introduction

The generation of high-quality 3D human heads has broad applications in digital avatars, telepresence, immersive gaming, and so on. Recently, many generative models have been proposed to facilitate 3D head generation by integrating geometry-aware representations with well-studied 2D image generative models [An et al. 2023; Chan et al. 2022; Kirschstein et al. 2024; Li et al. 2024].

Despite their advancements, existing methods often overlook the inherent diversity difference between hair and face, where faces of different identities still share a similarity in facial features while hairstyles are significantly more diverse. As a result, most prior 3D head generative models, which entangle the modeling of hair and face, are unsuitable for finer-grained editing tasks such as hair transfer. Although some approaches enable editability for 2D images [Nikolaev et al. 2024; Richardson et al. 2021] or geometry-aware representations like tri-plane [Gao et al. 2023; Sun et al. 2022], they either suffer from view inconsistency in 3D applications or are inadequate for modeling 3D hair, particularly in back-of-head regions. To address these limitations, we introduce **3DGH**, a generative model for 3D full-head synthesis that supports composable hair and face components.

To train a generative model that supports compositionality, there are two key issues we need to address: (1) first, we need to ensure a clear **separation** between hair and face to disentangle these components, and meanwhile, (2) we need to respect their inherent **correlation**, as observed in real-world patterns where male faces are predominantly associated with short hairstyles, while female faces typically feature medium to long hairstyles. To tackle these challenges, we propose a novel data representation with template-based 3D Gaussian Splatting (3DGS), in which two separate mesh templates are involved to model the overall structure of hair and face with 3D Gaussians spawned on their 2D *uv* texture maps. To capture the geometric variations among different hairstyles, we make the hair geometry itself deformable through PCA-based linear blend shapes fitted from multi-view facial capture data. This data representation then drives the design of our network architecture, which employs dual branches of StyleGAN2 [Karras et al. 2020] generators to independently generate hair and face Gaussians. A cross-attention mechanism [Vaswani et al. 2017] is introduced to carefully model the inherent correlation between hair and face, ensuring coherent and realistic outputs. The model is trained using a comprehensive objective that combines adversarial loss, reconstruction terms, and regularization terms, all carefully designed to stabilize training and facilitate effective hair-face separation.

We train our model using synthetic renderings from PanoHead [An et al. 2023]. After training on $25m$ images, we obtain a 3D generative model capable of producing diverse 3D heads with composable hair and face components. We evaluate the model both qualitatively and quantitatively, demonstrating its effectiveness in unconditional full-head image synthesis and composable 3D hairstyle editing through comparisons with several state-of-the-art 3D GAN methods.

In summary, our contributions are as follows:

- We propose 3DGH, a Gaussian-based 3D GAN for human heads that supports composable hair and face components.
- We introduce a novel data representation that utilizes separate template meshes for hair and face, with deformable hair geometry to capture diverse hairstyles.
- We design network architectures and training objectives to model hair-face separation and correlation, demonstrating their effectiveness through qualitative and quantitative comparisons with state-of-the-art 3D GAN methods.

2 Related Work

In this section, we review prior work in 3D head generative models, conditional image editing methods, and 3D hair modeling approaches.

2.1 3D Generative Adversarial Networks

3D Generative Adversarial Networks (GANs) are able to leverage adversarial training to develop generative models for 3D representations from 2D image collections. Early approaches primarily employed implicit 3D representations, such as NeRF [Mildenhall et al. 2020], to render either raw pixels [Chan et al. 2021; Schwarz et al. 2020] or features subsequently decoded by a CNN-based neural renderer [Niemeyer and Geiger 2021; Xue et al. 2022]. However, the high computational cost for volume rendering posed challenges for training high-resolution GANs. To address these limitations, more recent works have adapted successful 2D GAN architectures to generate compact intermediate representations, which can then be lifted to 3D [Chan et al. 2022; Gu et al. 2021; Or-El et al. 2022]. Among these approaches, the tri-plane representation introduced in EG3D [Chan et al. 2022] has proven to be the most effective for generating diverse and realistic geometry-aware portrait images. Building on this foundation, subsequent works, such as PanoHead [An et al. 2023] and SphereHead [Li et al. 2024], extend the tri-plane representation for full-head synthesis including back-view head images.

Despite the use of compact intermediate representations, the aforementioned methods typically rely on a 2D super-resolution network to enhance efficiency during training and inference, which may introduce unwanted artifacts in the form of 3D inconsistencies as well as low-resolution geometry. 3D Gaussian Splatting (3DGS) by Kerbl et al. [2023] provided an alternative direction by utilizing an explicit representation, where a set of 3D Gaussians is optimized from multi-view images using volume splatting [Zwicker et al. 2001]. Although the original 3DGS representation is unstructured, subsequent works such as Gaussian Shell Maps [Abdal et al. 2024] and

GGHead [Kirschstein et al. 2024] have demonstrated strategies to rig Gaussians in an organized manner relative to an underlying template mesh, thereby enabling the training of Gaussian-based 3D GANs for human body and head generations. In this work, we aim at training a similar Gaussian-based 3D GAN for human head generation, while focusing on disentangling the generation of hair and face to enable a composable model through the introduction of novel data representations, network architectures, and training strategies.

2.2 Conditional Image Editing

Beyond unconditional generation, GANs are extensively employed to learn mappings from a reference in a source domain to a target domain. Examples include translating semantic masks [Park et al. 2019; Zhu et al. 2020] or hand-drawn sketches [Chen et al. 2020] into photorealistic images. A common approach involves using a pre-trained StyleGAN [Karras et al. 2019] generator as a decoder while training customized encoders for different input modalities [Richardson et al. 2021]. Specifically for hairstyle editing, works like [Nikolaev et al. 2024; Wei et al. 2022, 2023; Zhu et al. 2021] take this approach to train customized encoders to map the input conditions to the latent space of StyleGAN, thereby achieving hairstyle editing with conditions such as reference images and text. While these methods achieve impressive results in 2D image editing, they often struggle to edit geometry-aware content due to the lack of mechanisms for preserving multi-view consistency in the synthesized outputs. To address this limitation, geometry-aware editing approaches such as IDE-3D [Sun et al. 2022] and SketchFaceNeRF [Gao et al. 2023] train 3D GANs from scratch with additional intermediate 3D representations for conditions such as semantic masks or sketches, thereby ensuring multi-view consistency during editing. A concurrent work by Bilecen et al. [2024] solves a similar 3D hairstyle editing problem through tri-plane editing. In this work, we adopt a similar approach by rendering hair-face segmentation as additional supervision. Furthermore, we carefully model the correlation between hair and face to enhance editing fidelity while respecting plausible conditional hairstyle distributions observed in the real world.

2.3 3D Hair Modeling

Given the complexity and variability of hair, high-quality 3D hair modeling has remained a persistent challenge for decades, as summarized in the comprehensive survey by Ward et al. [2007]. To capture high-quality 3D hair models, existing methods typically require high-end capture systems [Jakob et al. 2009; Paris et al. 2004; Xu et al. 2014] or even CT scanners [Shen et al. 2023] to fully recover strand-level details, leaving them inaccessible to most users. With the availability of synthetic 3D hair datasets [Hu et al. 2015], deep learning-based methods have emerged to regress 3D hair models from single-view [Chai et al. 2016; Saito et al. 2018; Wu et al. 2022; Zheng et al. 2023; Zhou et al. 2018] or sparse-view [Zhang et al. 2017] image input, thereby reducing the hardware requirement for 3D hair reconstruction. However, the performance of these data-driven methods is inherently constrained by the quality of their synthetic training datasets, which often lack realism and fail to represent intricate hairstyles, such as afro-textured hair. Recently, more advanced hair capture techniques have been proposed to jointly reconstruct hair geometry and appearance by incorporating neural volumetric primitives [Wang et al. 2022] or strand-aligned Gaussians [Luo et al. 2024; Zakharov et al. 2024]. While these methods produce impressive results, creating a large-scale 3D hair dataset with them remains tedious. Consequently, current hair generative models [He et al. 2025; Sklyarova et al. 2024; Zhou et al. 2023] continue to rely on synthetic data with augmentations and primarily focus on modeling hair geometry without appearance. In this work, we introduce a hair generative model trained on 2D image collections, capable of modeling both hair geometry and appearance with our deformable hair geometry representation and 3DGS-based rendering framework.

3 Methodology

An overview of our method is presented in Fig. 2, which integrates 3D Gaussian Splatting with the well-studied 3D GAN formulation. Our approach comprises three key components: a novel data representation that incorporates 3DGS and deformable hair geometry (Section 3.1), a newly designed network architecture that simultaneously models hair-face separation and correlation (Section 3.2), and training objectives specifically crafted to stabilize GAN training and enhance hair-face separation (Section 3.3).

3.1 Data Representation

3.1.1 Template-Based 3D Gaussian Splatting. Since the emergence of 3D Gaussian Splatting [Kerbl et al. 2023], it has shown outstanding expressivity in 3D scene representation, in which the scene is represented as a collection of 3D Gaussians, with each Gaussian denoted as $\mathbf{g}_i = \{\mathbf{p}_i, \mathbf{q}_i, \mathbf{s}_i, \mathbf{c}_i, o_i\} \in \mathbb{R}^{14}$, characterized by a set of parameters. These parameters include its center position $\mathbf{p}_i \in \mathbb{R}^3$, rotation parameterized by a unit quaternion $\mathbf{q}_i \in \mathbb{R}^4$, scale factor $\mathbf{s}_i \in \mathbb{R}^3$ along each axis, color $\mathbf{c}_i \in \mathbb{R}^3$, and opacity value $o_i \in \mathbb{R}$. This collection of 3D Gaussians can then be efficiently rendered through its differentiable tile-based rasterizer given the camera pose Π.

Considering the highly unstructured nature of the original 3DGS, we follow previous works [Abdal et al. 2024; Kirschstein et al. 2024; Saito et al. 2024] and associate each 3D Gaussian with a template mesh with corresponding *uv* layout. In this way, 3D Gaussians are represented as a 2D texture map $\mathbf{T} \in \mathbb{R}^{256\times256\times14}$, where each texel stores the parameters for a single 3D Gaussian primitive. In our 3D head representation, we adopt two different meshes and texture maps for hair and face, respectively, resulting in $\sim 131K$ Gaussians in total for the final rendering.

3.1.2 Deformable Hair Geometry. Even with our template-based 3DGS representation, there is an inherent difference between hair and face remaining unsolved, that is, hair contains much more geometric variation than different faces, which always share some commonalities among facial features such as eyes and mouth. For hair, there are many different hairstyles in the real world, ranging from short to long, fluffy to flat, making it hard to find a single template mesh to cover all these variations. To solve this problem, we learn a hair geometry prior that allows to produce deformable hair geometry to fit different hairstyles.

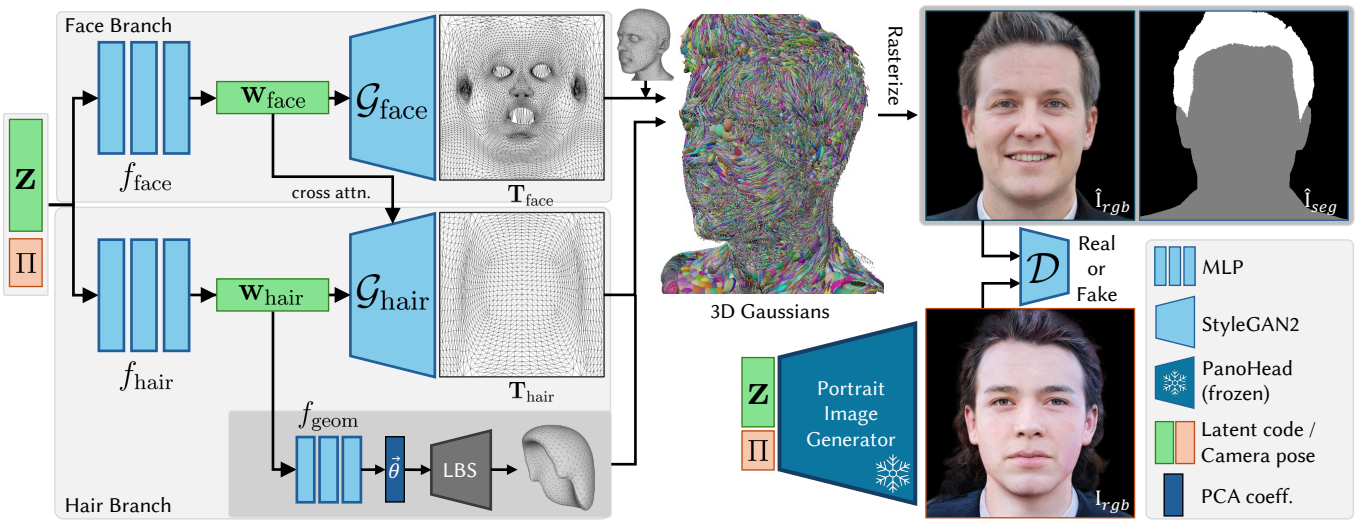


Fig. 2. Overview of 3DGH, which takes a randomly sampled Gaussian noise vector $\mathbf{z}$ and camera pose Π as input and outputs disentangled 3D head and hair representations, which are modeled as 3D Gaussians spawned on the 2D uv textures of the underlying 3D meshes. The generated 3D Gaussians are rasterized and supervised by the discriminator, with additional reconstruction supervision provided by inferring PanoHead [An et al. 2023] using the same noise $\mathbf{z}$ and camera pose Π. In 3DGH, the hair geometric variation is modeled with a separate geometry mapping network f_{geom} that produces PCA coefficients for our pre-computed linear blend shapes, and the hair-face correlation is modeled using cross-attention layers to integrate the information from $\mathbf{w}_{\text{face}}$.

To learn the prior, we first fit hair geometries from multi-view facial capture data that are similar to the Multiface dataset [Wuu et al. 2022], which provides calibrated camera parameters and semantic segmentations for each captured image. Given a template hair mesh, we formulate its deformation as an optimization problem, in which we differentiably render the segmentation of the deformed hair mesh using DRTK [Pidhorskyi et al. 2024], and compute the L_1 loss over the provided calibration. To suppress artifacts such as flipped and folded faces during deformation, we adopt an idea similar to Neural Jacobian Fields [Aigerman et al. 2022], where we optimize for Jacobians $\mathbf{J} \in \mathbb{R}^{F\times 3\times 3}$ (F refers to the number of faces of the template mesh) and centroid translation $\mathbf{t} \in \mathbb{R}^3$ of the template mesh, rather than vertex offsets directly. Vertex positions can be computed efficiently with a differentiable Poisson solver, and this process can thus be formally defined as:

$$\mathbf{J}^*, \mathbf{t}^* \coloneqq \arg\min_{\mathbf{J},\mathbf{t}} \|\mathcal{R}(\text{PoissonSolve}(\mathbf{J}) + \mathbf{t}; \Pi) - \mathbf{I}_{\text{seg}}\|_1, \tag{1}$$

where $\mathcal{R}(\cdot;\Pi)$ is the differentiable rendering operator of DRTK given the camera pose Π, PoissonSolve$(\cdot)$ is the differentiable Poisson solver in Neural Jacobian Fields, and $\mathbf{I}_{\text{seg}}$ is the pre-calibrated ground truth segmentation. Our experiments show that this optimization process converges within 500 iterations, and in Fig. 3 we visualize 10 hair meshes fitted from this process.

In total, we collect 283 different hair meshes, and then learn their prior using the conventional PCA-based methods in digital humans [Blanz and Vetter 1999; Loper et al. 2015]. Specifically, we solve a set of linear blend shapes by performing PCA on the normalized hair meshes, and different deformed hair meshes can thus be obtained from the linear function $\mathcal{M}(\vec{\theta})$:

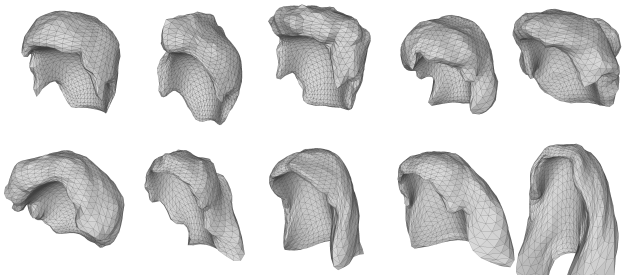


Fig. 3. Examples of hair meshes fitted by our algorithm. Since all examples are deformed from the same template, they share a consistent topology, as illustrated by their wireframe visualization.

$$\mathcal{M}(\vec{\theta}; \mathbf{X}) = \bar{\mathbf{M}} + \sigma \sum_{n=1}^{|\vec{\theta}|} \vec{\theta}_n \mathbf{X}_n, \tag{2}$$

where $\vec{\theta} = [\theta_1, \ldots, \theta_{|\vec{\theta}|}]^\top$ is a vector of blend shape coefficients, and $\mathbf{X} = [\mathbf{X}_1, \ldots, \mathbf{X}_{|\vec{\theta}|}]^\top \in \mathbb{R}^{|\vec{\theta}|\times 3V}$ forms a matrix of orthogonal principal components of shape displacements, with V referring to the number of vertices of the template mesh, and $\bar{\mathbf{M}} \in \mathbb{R}^{3V}$ and $\sigma \in \mathbb{R}$ denote the mean shape and standard variation computed from the original fitted hair meshes. We set the number of blend shape coefficients $|\vec{\theta}| = 32$, which ensures that the deformed mesh is smooth while still covering enough variations.

3.2 Network Architecture

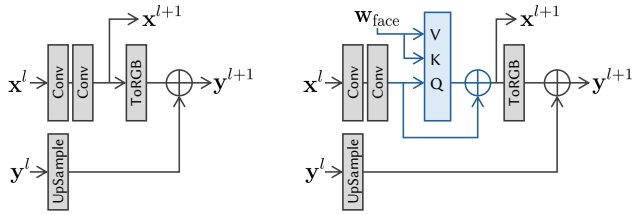
To obtain enough training data of frontal and back-of-head images with accurate camera poses, we adopt PanoHead [An et al. 2023] as our training data generator and train our generative model following the scheme of StyleGAN2 [Karras et al. 2020]. Formally, given a randomly sampled latent code $\mathbf{z} \in \mathbb{R}^{512}$ and camera pose $\Pi \in \mathbb{R}^{25}$, they are first passed to PanoHead to obtain a rendered RGB image $\mathbf{I}_{\text{rgb}} \in \mathbb{R}^{512\times512\times3}$, which is later segmented and parsed to obtain the foreground mask $\mathbf{I}_{\text{mask}} \in \mathbb{R}^{512\times512}$ and the hair-face segmentation map $\mathbf{I}_{\text{seg}} \in \mathbb{R}^{512\times512}$ using [Lin et al. 2021; Zheng et al. 2022]. These images, i.e., $\mathbf{I}_{\text{rgb}}$, $\mathbf{I}_{\text{mask}}$, $\mathbf{I}_{\text{seg}}$, serve as supervision signals to train our model.

3.2.1 Dual-Branch 3D GAN. As illustrated in Fig. 2, we design two separate branches to handle the generation of hair and face respectively. Given the same latent code $\mathbf{z}$ and camera pose Π, the mapping network $f : \mathcal{Z} \mapsto \mathcal{W}$ is first introduced to map them to the intermediate latent space, denoted as $\mathbf{w}_{\text{hair}}$ and $\mathbf{w}_{\text{face}}$. For hair, an additional geometry mapping network $f_{\text{geom}} : \mathcal{W} \mapsto \vec{\theta}$ is designed, which maps the latent code $\mathbf{w}_{\text{hair}}$ to proper blend shape coefficients that can represent the global shape of the hairstyle. These intermediate latent codes are then fed into two separate StyleGAN generators $\mathcal{G}_{\text{hair}}$ and $\mathcal{G}_{\text{face}}$, yielding two textures $\mathbf{T}_{\text{hair}}$ and $\mathbf{T}_{\text{face}}$ that store Gaussian parameters on each texel. We spawn 3D Gaussians from these textures and associate them with the underlying template mesh, combining and rendering them together to obtain the rendered RGB image $\hat{\mathbf{I}}_{\text{rgb}}$ and mask $\hat{\mathbf{I}}_{\text{mask}}$ from the provided camera pose. We use a similar dual discrimination method as EG3D [Chan et al. 2022], where we concatenate the rendered RGB and mask images and feed them into the discriminator with the camera pose Π. Aligning with the findings of Mimic3D [Chen et al. 2023], this adversarial training scheme helps increase diversity and maintain high-frequency details in our generated outputs.

3.2.2 Hair-Face Correlation. In reality, the distributions of plausible faces and hairstyles are correlated. For instance, hairstyles often correlate with gender and ethnicity. To encourage our model to learn these correlations, which are commonly observed in the real world, we use cross-attention layers [Vaswani et al. 2017] to inject $\mathbf{w}_{\text{face}}$ into each synthesis block of $\mathcal{G}_{\text{hair}}$, thereby influencing the hair generation process at different scales. Specifically, the intermediate feature map $\mathbf{y}^{l+1}$ generated at layer $l+1$ is computed as:

$$\begin{aligned} \mathbf{x}^{l} &= \text{Conv}(\mathbf{x}^{l}) \\ \mathbf{x}^{l+1} &= \mathbf{x}^{l} + \text{CrossAttention}(\mathbf{Q} = \mathbf{x}^{l}, \mathbf{K} = \mathbf{V} = \mathbf{w}_{\text{face}}) \\ \mathbf{y}^{l+1} &= \text{Upsample}(\mathbf{y}^{l}) + \text{ToRGB}(\mathbf{x}^{l+1}) \end{aligned} \tag{3}$$

In Eq. (3), $\mathbf{x}^{l}$ is the input feature map for convolution from layer l, $\text{Conv}(\cdot)$ is the modulated convolution layers inside synthesis blocks, $\text{Upsample}(\cdot)$ is the spatial upsampling operator to upscale the previous feature map $\mathbf{y}^{l}$ to match the spatial resolution of $\mathbf{x}^{l+1}$, $\text{ToRGB}(\cdot)$ is the convolution layer that adjusts the number of channels in the convolved feature map $\mathbf{x}^{l+1}$ to match $\mathbf{y}^{l}$, and $\text{CrossAttention}(\cdot,\cdot)$ is the cross-attention layers we newly introduced compared to the original synthesis blocks in StyleGAN2. The diagram for our hair-face correlation module is provided in Fig. 4.



(a) Original architecture (b) Ours with cross-attention

Fig. 4. Diagram of our hair-face correlation module, which utilizes cross-attention layers to inject $\mathbf{w}_{\text{face}}$ into each synthesis block of StyleGAN2.

Inspired by classifier-free guidance [Ho and Salimans 2022], we design a similar technique by randomly dropping the condition $\mathbf{w}_{\text{face}}$ (replacing it with all-zero vectors $\varnothing$) during training with a probability of 10%. Then in the inference stage, we blend the conditional feature map $\mathbf{x}^{l}$ and unconditional feature map $\mathbf{x}^{l}_{\varnothing}$ with the CFG factor ω:

$$\tilde{\mathbf{x}}^{l} = \omega\mathbf{x}^{l} + (1-\omega)\mathbf{x}^{l}_{\varnothing}, \tag{4}$$

thereby allowing for further control over the hair-face correlation with the CFG factor ω.

3.3 3D GAN Training

As PanoHead gives us direct supervision during training, our final training objective consists of adversarial loss, reconstruction losses on rendered images, and several regularization terms to stabilize the training and improve the generation quality.

Adversarial Loss. Following EG3D [Chan et al. 2022], we incorporate the standard non-saturating GAN loss $\mathcal{L}_{\text{adv}}$ [Goodfellow et al. 2014] with R_1 gradient regularization [Mescheder et al. 2018] on both the RGB and mask images, where the regularization strengths are set to 1 for both of them.

RGB and Mask Loss. From Gaussian Splatting we can obtain the rendered RGB image $\hat{\mathbf{I}}_{\text{rgb}}$ and mask $\hat{\mathbf{I}}_{\text{mask}}$, on which we compute the L_1 loss to measure their reconstruction quality. These loss terms can be expressed as:

$$\mathcal{L}_{\text{rgb}} = \|\hat{\mathbf{I}}_{\text{rgb}} - \mathbf{I}_{\text{rgb}}\|_1, \quad \mathcal{L}_{\text{mask}} = \|\hat{\mathbf{I}}_{\text{mask}} - \mathbf{I}_{\text{mask}}\|_1. \tag{5}$$

Segmentation Loss. To encourage a clear separation between hair and face, we further assign different one-hot labels to Gaussians spawned from $\mathbf{T}_{\text{hair}}$ and $\mathbf{T}_{\text{face}}$ ($[0,0,1]$ for hair and $[0,1,0]$ for face, $[1,0,0]$ is left for background), which will be used to render an additional segmentation map $\hat{\mathbf{I}}_{\text{seg}} \in \mathbb{R}^{512\times512\times3}$ with other Gaussian parameters. To ensure that the generated hair mesh faithfully represents the hairstyle in the image, we additionally render the hair mesh segmentation $\hat{\mathbf{I}}^{\text{mesh}}_{\text{seg}} \in \mathbb{R}^{512\times512}$ with DRTK [Pidhorskyi et al. 2024] by assigning different scalar values to vertices of the hair and face meshes (2 for hair and 1 for face, the default value 0 is for background). These segmentation maps are supervised with the segmentation map $\mathbf{I}_{\text{seg}}$ parsed from $\mathbf{I}_{\text{rgb}}$, where the loss terms are defined as:

$$\mathcal{L}_{\text{seg}} = \text{CrossEntropy}(\hat{\mathbf{I}}_{\text{seg}}, \mathbf{I}_{\text{seg}}), \quad \mathcal{L}^{\text{mesh}}_{\text{seg}} = \|\hat{\mathbf{I}}^{\text{mesh}}_{\text{seg}} - \mathbf{I}_{\text{seg}}\|_1. \tag{6}$$

Note that we use the cross-entropy loss rather than L_1 loss for the segmentation map $\hat{\mathbf{I}}_{\text{seg}}$ rendered from Gaussian Splatting, as its α-blending nature will inevitably change values around the boundary, thus causing mislabeling issues for pixels on the boundary if their labels are scalar.

Regularization Terms. As our adversarial training is weakly supervised and 3D Gaussians are quite sensitive to gradient updates during early training stages, unconstrained training will quickly lead to divergence or mode collapse with overly large or extremely small Gaussians in the early stage. Therefore, we apply some regularization terms to stabilize the training. First, based on our hybrid 3DGS representation, the center position $\mathbf{p}_i$ of each Gaussian is defined as the sum of the 3D position $\mathbf{v}_i$ on the mesh surface and the delta position $\Delta\mathbf{p}_i$ decoded from the generated texture maps. To ensure that all Gaussians stay within a thin layer around the mesh surface, we first clamp the absolute value of the decoded delta position $\Delta\mathbf{p}_i$ with a threshold γ. We use $\gamma = 40$ for face Gaussians, meaning that they can move at most 40mm away from the surface. As the hair mesh itself is deformable, we reduce γ to 20 for hair Gaussians to make sure that different hairstyles are generated with different geometries, rather than similar geometries with largely deviated Gaussians. A regularization term for delta positions is applied:

$$\mathcal{L}_{\text{reg}}^{\text{pos}} = \sum_i \|\Delta\mathbf{p}_i\|_2, \tag{7}$$

which encourages all predicted Gaussians to stay close to the mesh surface. To constrain the scale of Gaussians to stay within a reasonable range, the most effective regularization term we experimented with can be defined as:

$$\mathcal{L}_{\text{reg}}^{\text{scale}} = \begin{cases} 10 \times |s_i - s_{\min}| & s_i < s_{\min} \\ (s_i - s_{\max})^2 & s_i > s_{\max} \end{cases} \tag{8}$$

It applies different penalties to constrain the Gaussian scaling along all axes if they are outside of a reasonable range $[s_{\min}, s_{\max}]$, where $s_{\min} = 0.2$ and $s_{\max} = 5$. Finally, we also apply the *uv* total variation loss $\mathcal{L}_{\text{reg}}^{\text{uv}}$ proposed in GGHead [Kirschstein et al. 2024] to prevent Gaussians in the back from shining through to the front.

Combining all the terms discussed above, the final training objective is defined as their weighted sum, expressed as:

$$\begin{aligned}\mathcal{L} = \mathcal{L}_{\text{adv}} + \lambda_{\text{rgb}}\mathcal{L}_{\text{rgb}} + \lambda_{\text{mask}}\mathcal{L}_{\text{mask}} + \lambda_{\text{seg}}\mathcal{L}_{\text{seg}} + \lambda_{\text{seg}}^{\text{mesh}}\mathcal{L}_{\text{seg}}^{\text{mesh}} \\ + \lambda_{\text{reg}}^{\text{pos}}\mathcal{L}_{\text{reg}}^{\text{pos}} + \lambda_{\text{reg}}^{\text{scale}}\mathcal{L}_{\text{reg}}^{\text{scale}} + \lambda_{\text{reg}}^{\text{uv}}\mathcal{L}_{\text{reg}}^{\text{uv}}\end{aligned} \tag{9}$$

where we set the weighting factors $\lambda_{\text{rgb}} = 10$, $\lambda_{\text{mask}} = 10$, $\lambda_{\text{seg}} = 1$, $\lambda_{\text{seg}}^{\text{mesh}} = 100$, $\lambda_{\text{reg}}^{\text{pos}} = 0.1$, $\lambda_{\text{reg}}^{\text{scale}} = 1$, and $\lambda_{\text{reg}}^{\text{uv}} = 1$ to balance the influence of different terms.

4 Experiments

We use multi-view capture data [Saito et al. 2024; Wuu et al. 2022] to solve the linear blend shapes for our deformable hair geometry, and we use PanoHead [An et al. 2023] as the portrait image generator to train our generative model. For details about these datasets, please refer to Sec. A in supplemental.

4.1 Comparisons

We compare against several competitive baseline methods from the 3D GAN literature. Unless otherwise indicated, all baselines are their official checkpoints to maintain their original quality. We evaluate the quality of the generated multi-view images, both quantitatively and qualitatively.

4.1.1 Qualitative Comparisons. Fig. 5 visually compares the image quality against baselines including EG3D [Chan et al. 2022], PanoHead [An et al. 2023], SphereHead [Li et al. 2024], and GGHead [Kirschstein et al. 2024], where EG3D is the pioneering work that synthesizes high-quality portrait images with the tri-plane representation and a 2D super-resolution network, PanoHead and SphereHead are two subsequent works that achieve full-head synthesis by improving the tri-plane representation, and GGHead utilizes a similar hybrid representation with 3DGS and a template mesh as ours. All synthesized images contain 5 different views, with yaw angles ranging from 0° to 180°. While all methods successfully synthesize realistic frontal-view images, the rendering quality of EG3D and GGHead deteriorates significantly when rendering from large camera poses or back-view areas, as these methods are not specifically designed for full-head image synthesis. Compared to SphereHead and PanoHead, our method achieves comparable visual quality while providing the additional advantage of compositionality.

To prove the compositionality of our method, we first present samples generated by our model alongside the rendered hair-face segmentation maps and mesh normal maps in Fig. 6. Leveraging our deformable hair geometry, we achieve smooth deformations of the hair mesh, effectively capturing the overall structure of various hairstyles. Additionally, the 3D Gaussians associated with the hair mesh exhibit a clear separation from the face Gaussians and are capable of representing some strand-level details, thus yielding hair-face segmentation maps with finer-grained details that are difficult to obtain from common image segmentation models. More uncurated samples can be found in Sec. D of our supplemental.

We then evaluate our compositionality through 3D hairstyle editing, where the 3D hairstyle in the reference sample is transferred to another one. The qualitative results are provided in Fig. 7, demonstrating that our approach preserves the reference hairstyle with high fidelity while producing a natural blending around the hairline. Both the hair geometry and appearance are transferred through a simple latent code swap. Note that in 2D hairstyle editing methods, such as HairFastGAN [Nikolaev et al. 2024], such hairstyle editing typically requires multiple processing steps and network modules. Furthermore, our editing results inherently maintain multi-view consistency, attributed to the 3D nature of our representations.

We finally investigate the impact of our hair-face correlation modeling technique by generating hair-face compositions with varying levels of the CFG scale factor ω. As illustrated in Fig. 8, when the reference face is male, it favors short-length hairstyles to align with plausible hairstyle distributions observed for this face condition in real life. Consequently, as ω increases, the transferred hairstyle gradually becomes shorter while preserving the overall style of the reference. This experiment provides strong evidence for the effectiveness of our hair-face correlation module, which introduces an extra dimension for editing transferred hairstyles while maintaining

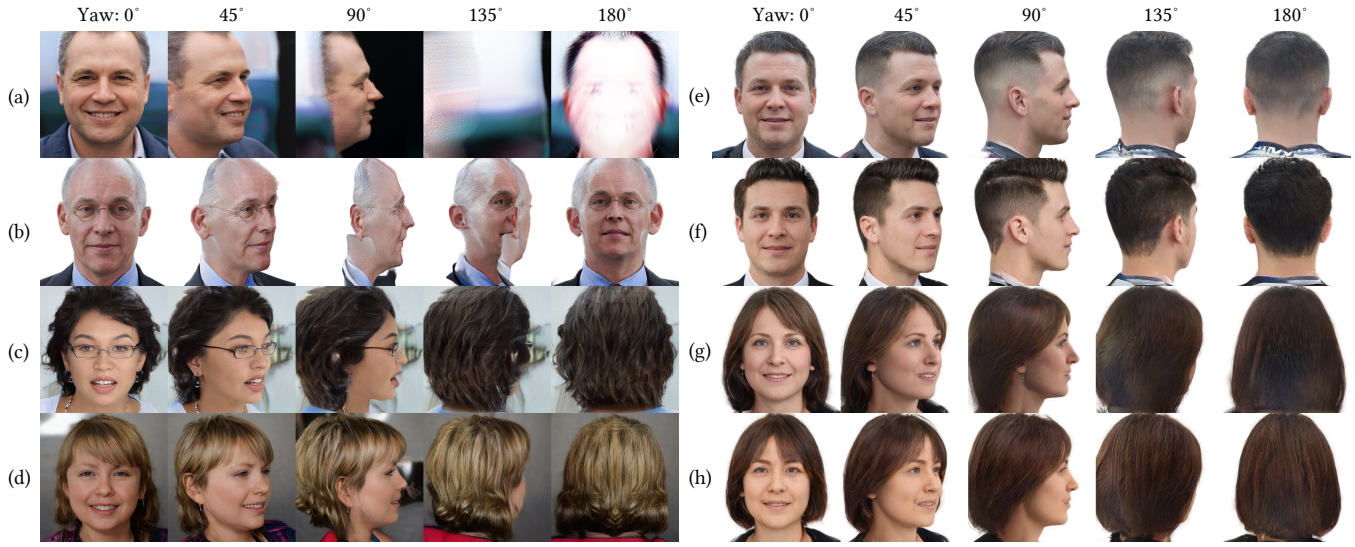


Fig. 5. Qualitative comparison with various 3D GANs. (a) EG3D [Chan et al. 2022] and (b) GGHead [Kirschstein et al. 2024] are not specifically designed for full-head image synthesis, resulting in significant quality degradation when rendering large-pose or back-view images. (c) SphereHead [Li et al. 2024] and (d) PanoHead [An et al. 2023] are two 3D GANs tailored for full-head image synthesis. Compared with these methods, our results (e-h) demonstrate comparable quality while offering the additional advantage of compositionality.

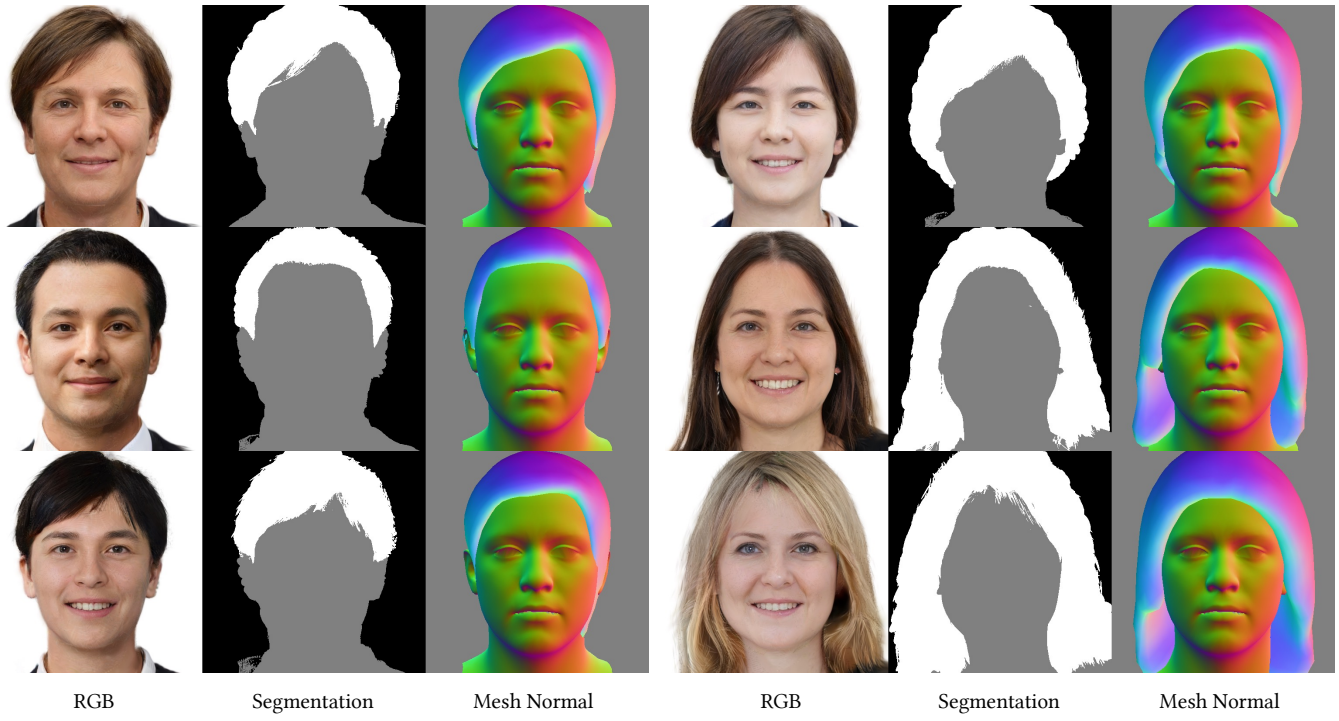


Fig. 6. Generated samples with corresponding hair-face segmentation and deformed hair geometry. Our method enables smooth deformation of the hair mesh to represent various hairstyles, while assigning 3D Gaussians to capture strand-level details and appearance.

plausibility – a critical point often overlooked by previous hairstyle editing methods.

4.1.2 Quantitative Comparisons. We first quantitatively evaluate our generation quality by measuring the Fréchet Inception Distance

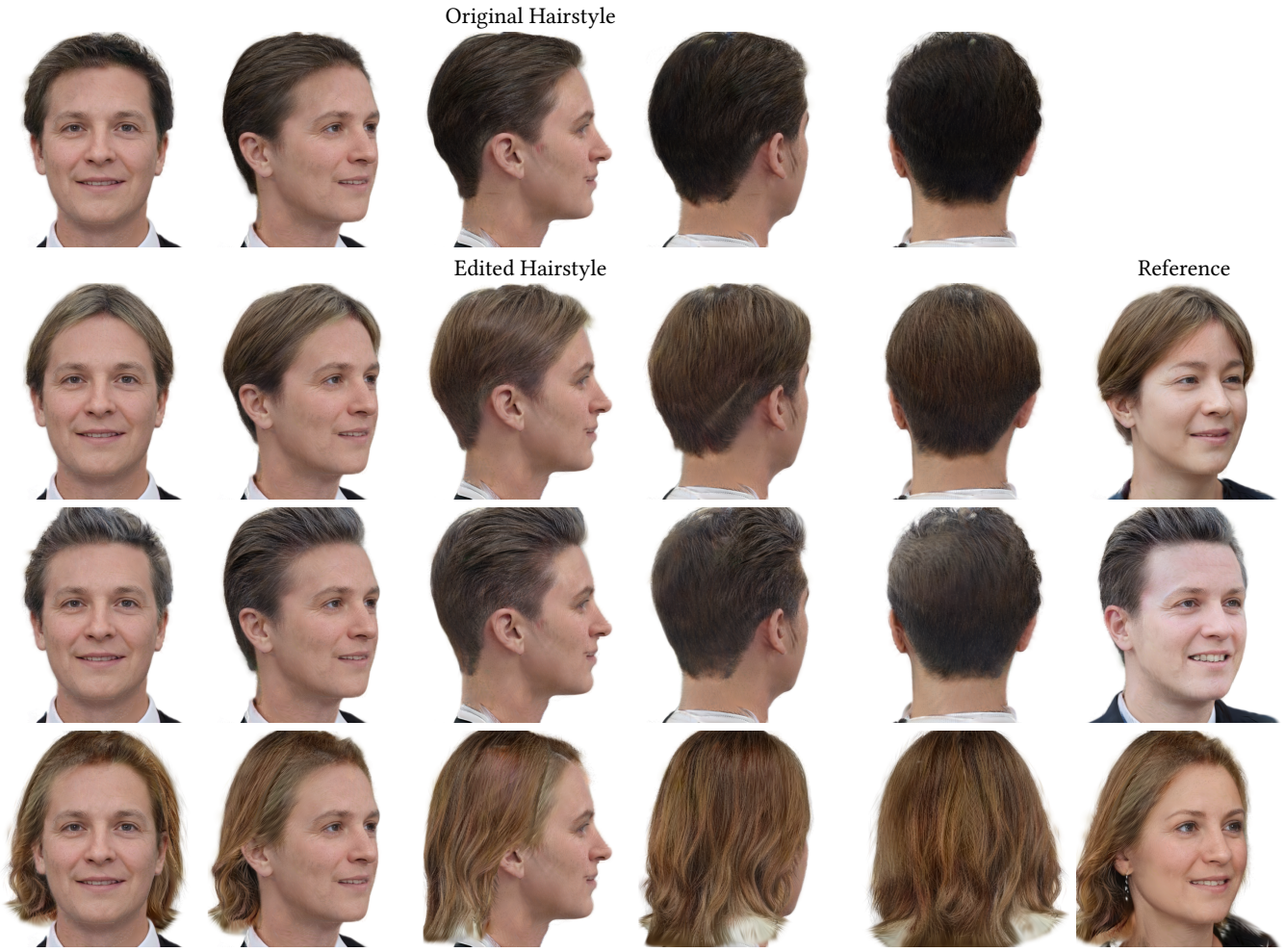


Fig. 7. Our method supports 3D hairstyle editing by swapping the hair latent code $\mathbf{w}_{\text{hair}}$ from the reference samples. This editing process transfers both the hair geometry and appearance, while ensuring multi-view consistency thanks to the inherently 3D nature of our representations.

Fig. 8. Analysis of the CFG scale factor ω, where we present hair-face compositions generated using varying levels of ω. When ω is small, the hair-face correlation has weak influence on the final output, resulting in hairstyles more similar to the reference. As ω increases, the composition process becomes more biased toward the face distribution, producing hairstyles that are more contextually appropriate for the given face.

(FID) [Heusel et al. 2017] computed over $50k$ real and fake image samples. Since PanoHead [An et al. 2023] serves as our training data generator, its renderings are treated as real image samples for the FID evaluation. In addition to the overall FID (FID-all) computed on randomly sampled poses, we evaluate generation quality at a finer granularity by sampling camera poses from different regions. Specifically, FID-front evaluates facial details using images synthesized from frontal views ($|yaw| < 90°$), while FID-back assesses hair

Table 1. Quantitative comparison on different FID metrics.

	FID ↓		
	back	front	all
Ours	9.86	5.47	6.55

Table 2. Quantitative comparison between our method and other 3D GANs on multi-view consistency.

	EG3D [Chan et al. 2022]	GGHead [Kirschstein et al. 2024]	SphereHead [Li et al. 2024]	Ours
ID ↑	0.678	0.683	0.581	**0.690**

details using images synthesized from back views ($|yaw| \geq 90°$). As shown in Table 1, all three metrics yield values less than 10, indicating that the quality of our generated results is still comparable to that of PanoHead.

We then conduct quantitative comparisons with EG3D [Chan et al. 2022], SphereHead [Li et al. 2024], and GGHead [Kirschstein et al. 2024] to evaluate multi-view consistency, in which we measure the identity similarity score (ID) by calculating the average Adaface [Kim et al. 2022] cosine similarity between paired images rendered from different camera poses. As reported in Table 2, our method achieves the best multi-view consistency, since EG3D and GGHead struggle for large-pose images and SphereHead involves a 2D super-resolution network that may introduce artifacts.

4.2 Ablation Study

In Table 3, we analyze key design decisions in our method, including the choice of supervision on segmentation maps, the use of deformable hair geometry, and variations in hair-face correlation modules. To assess the compositionality of our approach, we generate $50k$ samples by randomly swapping the intermediate latent codes $\mathbf{w}_{\text{hair}}$ and $\mathbf{w}_{\text{face}}$, thereby creating novel samples with mismatched hair and face combinations. We then compute FID for these swapped samples against the real samples, denoted as FID-swap, to evaluate the realism of the randomly combined hair and face outputs produced by our model.

In Table 3, the first 3 rows evaluate the impact of our choice of segmentation supervision. Specifically, *Seg. in* $\mathcal{D}$ refers to concatenate the rendered segmentation maps to the input of discriminator and let it determine whether the segmentation is realistic or not in an adversarial manner. Meanwhile, *w/o Seg. loss* denotes setting λ_{seg} to 0. In these experiments, we observed that passing segmentation maps to the discriminator often caused mode collapse during the early stages of GAN training, resulting in meaningless outputs and high FID scores. We hypothesize that it is mainly due to the mismatch in value representations: the rendered segmentation maps contain continuous floating-point values, whereas the ground-truth segmentations parsed from RGB images are discrete labels. This is a fundamental difference between segmentation maps and RGB/mask images, since both ground truth RGB and mask images can contain continuous values between 0 and 1. Therefore, the discriminator

Table 3. Ablation studies on segmentation supervision, deformable hair geometry, and hair-face correlation module. The last row refers to our final architecture design.

Seg.	Hair Geom.	Hair-Face Corr.	FID ↓	FID-swap ↓
Seg. in $\mathcal{D}$	✗	cross attn.	296.97	296.89
w/o Seg. loss	✗	cross attn.	10.56	31.97
w/ Seg. loss	✗	cross attn.	12.15	34.18
w/ Seg. loss	✓	–	12.11	29.99
w/ Seg. loss	✓	concat.	10.30	**14.95**
w/ Seg. loss	✓	cross attn.	**7.67**	20.56

can easily distinguish real and generated samples based on this quantization discrepancy, breaking the training process at an early stage. Surprisingly, removing the segmentation loss still produced reasonable generation results with acceptable segmentation quality. Adding the segmentation loss encouraged a cleaner separation between hair and face Gaussians, though it slightly increased the FID score due to the additional constraints imposed on Gaussians. We provide qualitative comparisons to discuss these observations in Fig. 1 of our supplemental. The 3rd row further illustrates the importance of our deformable hair geometry. In this experiment, we replaced the deformable hair geometry with the average hair mesh fitted from studio capture data and trained our model using this fixed geometry. Quantitative results demonstrate that incorporating a deformable hair geometry improves overall generation quality. Visually, we observed that when the hair geometry is fixed, hair Gaussians need larger deviations to represent varying hairstyles, resulting in floating Gaussians appearing in random positions. Fig. 2 in supplemental includes a qualitative comparison of this artifact. The last 3 rows examine different hair-face correlation modules. In the 4th row, we remove this module entirely, while the 5th row replaces it with a mechanism that concatenates $\mathbf{w}_{\text{hair}}$ and $\mathbf{w}_{\text{face}}$ with a lightweight MLP to fuse them and model their correlation. The results indicate that our cross-attention mechanism achieves the lowest FID, signifying better generation quality. Although the concatenation mechanism achieves a lower FID-swap, qualitative analysis reveals that it introduces a strong dependency on $\mathbf{w}_{\text{face}}$, which reduces the diversity when $\mathbf{w}_{\text{face}}$ is fixed and $\mathbf{w}_{\text{hair}}$ is swapped. Fig. 3 in supplemental shows this artifact. Overall, these experiments demonstrate that our final architecture design (as shown in the last row) achieves the best balance of generation quality and diversity, enabling 3D hairstyle editing in the generated results with a certain guarantee of realism.

4.3 Latent Space Interpolation

Fig. 9 illustrates how variations in the generated 3D head correspond to interpolations in the latent space of 3DGH. We begin by randomly sampling two pairs of latent codes, which are linearly interpolated to produce intermediate representations. The resulting renderings, arranged from left to right, show a smooth semantic transition between two identities. Leveraging our composable design, we independently interpolate between $\mathbf{w}_{\text{hair}}$ and

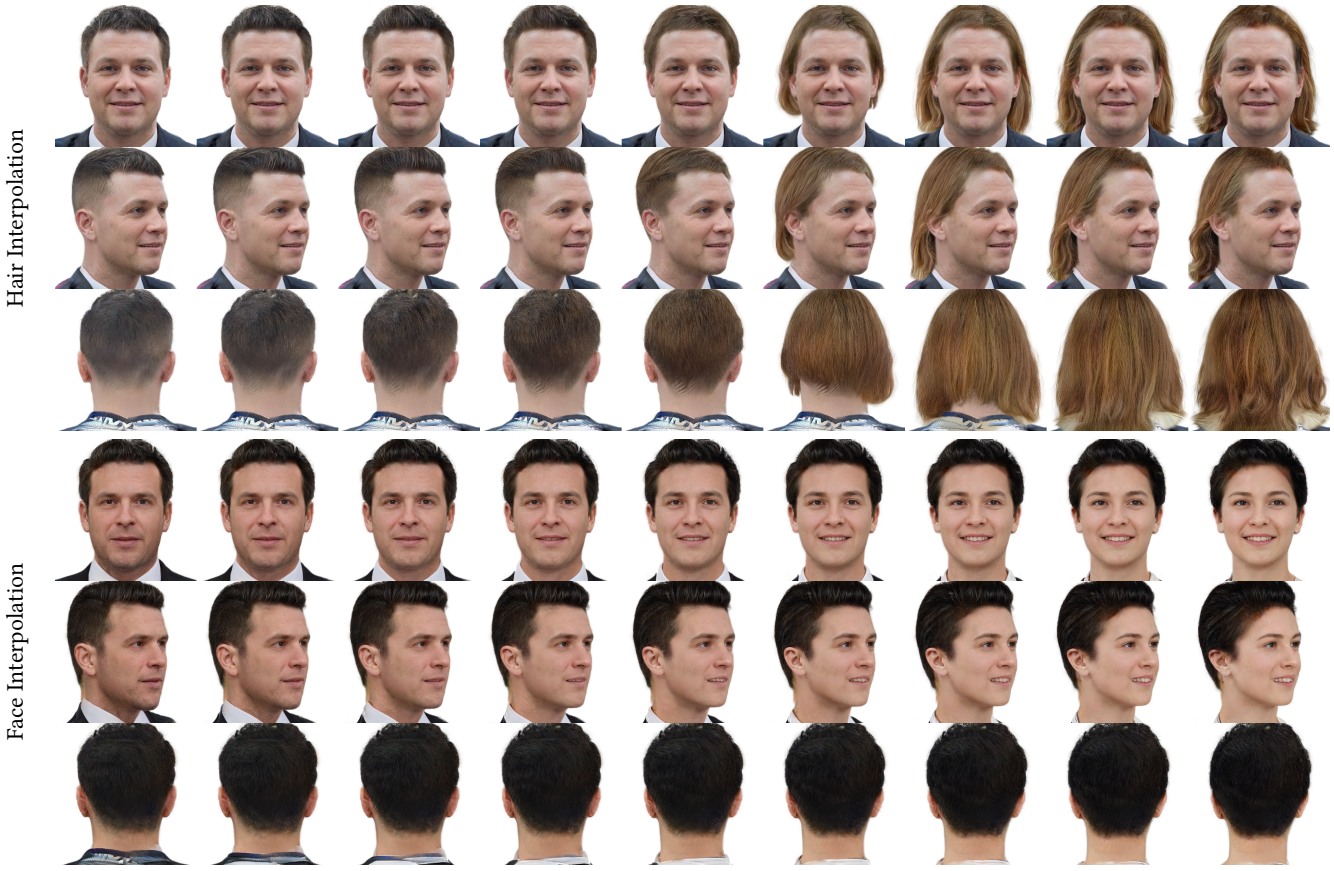


Fig. 9. Linear interpolation in the latent space of 3DGH. Top: Interpolation of $\mathbf{w}_{\text{hair}}$ with the facial identity fixed, demonstrating smooth transitions between hairstyles. Bottom: Interpolation of $\mathbf{w}_{\text{face}}$ while keeping the hairstyle constant, illustrating gradual changes in facial features.

$\mathbf{w}_{\text{face}}$, enabling disentangled control over hair and facial features. The seamless transitions and consistent semantic structure across interpolations highlight the continuity and expressiveness of the latent space learned by our model.

5 Discussion

Limitations and Future Work. While our method demonstrates strong performance, it still faces several limitations. First, the expressiveness of our model is constrained by the quality and diversity of the training data, which in our case are the images generated by PanoHead [An et al. 2023]. A clear domain gap exists between these synthetic images and in-the-wild images, making certain hairstyles, such as buns and braids, difficult to generate. Addressing this issue necessitates a large-scale dataset of in-the-wild images with comprehensive coverage of frontal and back views, accurate camera calibration, and reliable image alignment. Although works like PanoHead [An et al. 2023] and SphereHead [Li et al. 2024] have made progress in this direction, their in-house training data are not publicly available at this time, and specific processing steps are still needed to calibrate and align back views in the absence of facial landmarks. Therefore, combining real-world multi-view datasets such as RenderMe-360 [Pan et al. 2023] with synthetic data would be an interesting future work to explore that may alleviate these training data issues. Second, our model occasionally produces back-view artifacts for long hairstyles that occupy a large portion of the frontal view, as illustrated in Fig. 10. We attribute this to our generator's conditioning on both the latent code and camera pose, following the design choice of EG3D [Chan et al. 2022] and PanoHead [An et al. 2023]. Despite using an 80% pose-swapping probability during training, rendering quality degrades when rendering poses differ significantly from the conditioning poses. This limitation is also observed in PanoHead [An et al. 2023] and SphereHead [Li et al. 2024] because full-head image synthesis requires 360-degree rendering. Better conditioning on camera pose may require an advanced architecture as a future work. Third, our method can produce hollow artifacts, particularly in hair regions, due to the stretching and scaling of Gaussian primitives to represent thin strand structures. While increasing the number of Gaussians may alleviate this problem by providing denser coverage, it would also lead to higher computational cost in terms of model size and training time. Fourth, while our design choice of the cross-attention layer in Eq. (3) is motivated by the real-world observation that there exist correlations between ethnic facial features and culturally associated

Fig. 10. Failure case illustrating artifacts in the back-view rendering of a long hairstyle.

hairstyles, our analysis in Fig. 8 does not fully cover various correlations other than gender. This is mainly because gender is the dominant factor in hair-face correlations in the dataset we used. Lastly, extending our framework toward animatable 3D avatars is an exciting future direction. By incorporating parametric models such as FLAME [Li et al. 2017], our method could be adapted to support disentangled control over head pose, facial expression, and hairstyle, paving the way for more versatile and customizable 3D avatar generation.

Ethical considerations. As with other generative models for digital avatars, our method carries potential risks related to misuse (e.g., identity manipulation) and biases. These concerns are partly due to the use of synthetic training data, which may lack sufficient diversity in demographics and hairstyles, limiting the representation of hair-face correlations across different ethnicities (see Fig. 4 in supplemental). To mitigate such issues, we strongly advocate for the responsible use of 3DGH, transparency in its deployment, and the continued development of diverse, representative datasets. We explicitly oppose any use of our work for malicious purposes, including the spread of misinformation or the violation of individual rights.

6 Conclusion

We introduce 3DGH, a Gaussian-based 3D GAN framework that supports composable hair and face generation. Leveraging multi-view studio capture data, we propose a novel data representation with template-based 3DGS, in which hair Gaussians are rigged to a deformable hair geometry constructed using PCA-based linear blend shapes. This data representation drives the design of our network architecture, which incorporates dual branches to independently generate hair and face Gaussians. A cross-attention mechanism is introduced to model the inherent correlation between hair and face, ensuring coherent and realistic outputs. Our model is then trained using a comprehensive objective that includes adversarial loss, reconstruction terms, and regularization terms designed to stabilize training and facilitate hair-face separation. We evaluate the trained model both qualitatively and quantitatively, demonstrating its superior performance in unconditional full-head image synthesis and composable 3D hairstyle editing.

References

Rameen Abdal, Wang Yifan, Zifan Shi, Yinghao Xu, Ryan Po, Zhengfei Kuang, Qifeng Chen, Dit-Yan Yeung, and Gordon Wetzstein. 2024. Gaussian Shell Maps for Efficient 3D Human Generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 9441–9451.

Noam Aigerman, Kunal Gupta, Vladimir G. Kim, Siddhartha Chaudhuri, Jun Saito, and Thibault Groueix. 2022. Neural Jacobian Fields: Learning Intrinsic Mappings of Arbitrary Meshes. *ACM Trans. Graph.* 41, 4, Article 109 (July 2022), 17 pages.

Sizhe An, Hongyi Xu, Yichun Shi, Guoxian Song, Umit Y. Ogras, and Linjie Luo. 2023. PanoHead: Geometry-Aware 3D Full-Head Synthesis in 360°. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 20950–20959.

Bahri Batuhan Bilecen, Yigit Yalin, Ning Yu, and Aysegul Dundar. 2024. Reference-Based 3D-Aware Image Editing with Triplane. *arXiv preprint arXiv:2404.03632* (2024).

Volker Blanz and Thomas Vetter. 1999. A Morphable Model for the Synthesis of 3D Faces. In *Proceedings of the 26th annual conference on Computer graphics and interactive techniques*. 187–194.

Menglei Chai, Tianjia Shao, Hongzhi Wu, Yanlin Weng, and Kun Zhou. 2016. AutoHair: Fully Automatic Hair Modeling from a Single Image. *ACM Transactions on Graphics* 35, 4, Article 116 (jul 2016), 12 pages.

Eric R Chan, Connor Z Lin, Matthew A Chan, Koki Nagano, Boxiao Pan, Shalini De Mello, Orazio Gallo, Leonidas J Guibas, Jonathan Tremblay, Sameh Khamis, et al. 2022. Efficient Geometry-aware 3D Generative Adversarial Networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 16123–16133.

Eric R Chan, Marco Monteiro, Petr Kellnhofer, Jiajun Wu, and Gordon Wetzstein. 2021. pi-GAN: Periodic Implicit Generative Adversarial Networks for 3D-Aware Image Synthesis. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 5799–5809.

Shu-Yu Chen, Wanchao Su, Lin Gao, Shihong Xia, and Hongbo Fu. 2020. Deep Generation of Face Images from Sketches. *arXiv preprint arXiv:2006.01047* (2020).

Xingyu Chen, Yu Deng, and Baoyuan Wang. 2023. Mimic3D: Thriving 3D-Aware GANs Via 3D-to-2D Imitation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*.

Lin Gao, Feng-Lin Liu, Shu-Yu Chen, Kaiwen Jiang, Chun-Peng Li, Yu-Kun Lai, and Hongbo Fu. 2023. SketchFaceNeRF: Sketch-Based Facial Generation and Editing in Neural Radiance Fields. *ACM Transactions on Graphics (Proceedings of ACM SIGGRAPH 2023)* 42, 4 (2023), 159:1–159:17.

Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. Generative Adversarial Nets. *Advances in neural information processing systems* 27 (2014).

Jiatao Gu, Lingjie Liu, Peng Wang, and Christian Theobalt. 2021. StyleNeRF: A Style-based 3D-Aware Generator for High-resolution Image Synthesis. *arXiv preprint arXiv:2110.08985* (2021).

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. 2025. Perm: A Parametric Representation for Multi-Style 3D Hair Modeling. In *International Conference on Learning Representations*.

Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. 2017. GANs Trained by a Two Time-scale Update Rule Converge to a Local Nash Equilibrium. *Advances in neural information processing systems* 30 (2017).

Jonathan Ho and Tim Salimans. 2022. Classifier-Free Diffusion Guidance. *arXiv preprint arXiv:2207.12598* (2022).

Liwen Hu, Chongyang Ma, Linjie Luo, and Hao Li. 2015. Single-View Hair Modeling Using a Hairstyle Database. *ACM Transactions on Graphics (Proceedings SIGGRAPH 2015)* 34, 4 (July 2015).

Wenzel Jakob, Jonathan T. Moon, and Steve Marschner. 2009. Capturing Hair Assemblies Fiber by Fiber. *ACM Trans. Graph.* 28, 5 (dec 2009), 1–9.

Tero Karras, Samuli Laine, and Timo Aila. 2019. A Style-Based Generator Architecture for Generative Adversarial Networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. 2020. Analyzing and Improving the Image Quality of StyleGAN. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 8110–8119.

Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 2023. 3D Gaussian Splatting for Real-Time Radiance Field Rendering. *ACM Transactions on Graphics* 42, 4 (July 2023).

Minchul Kim, Anil K Jain, and Xiaoming Liu. 2022. AdaFace: Quality Adaptive Margin for Face Recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.

Tobias Kirschstein, Simon Giebenhain, Jiapeng Tang, Markos Georgopoulos, and Matthias Nießner. 2024. GGHead: Fast and Generalizable 3D Gaussian Heads. In *SIGGRAPH Asia 2024 Conference Papers (SA '24)*. Article 126, 11 pages.

Heyuan Li, Ce Chen, Tianhao Shi, Yuda Qiu, Sizhe An, Guanying Chen, and Xiaoguang Han. 2024. SphereHead: Stable 3D Full-head Synthesis with Spherical Tri-plane Representation. *arXiv preprint arXiv:2404.05680* (2024).

Tianye Li, Timo Bolkart, Michael. J. Black, Hao Li, and Javier Romero. 2017. Learning a Model of Facial Shape and Expression from 4D Scans. *ACM Transactions on Graphics, (Proc. SIGGRAPH Asia)* 36, 6 (2017), 194:1–194:17.

Shanchuan Lin, Andrey Ryabtsev, Soumyadip Sengupta, Brian L Curless, Steven M Seitz, and Ira Kemelmacher-Shlizerman. 2021. Real-Time High-Resolution Background Matting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 8762–8771.

Matthew Loper, Naureen Mahmood, Javier Romero, Gerard Pons-Moll, and Michael J. Black. 2015. SMPL: a Skinned Multi-Person Linear Model. *ACM Trans. Graphics (Proc. SIGGRAPH Asia)* 34, 6 (Oct. 2015), 248:1–248:16.

Haimin Luo, Min Ouyang, Zijun Zhao, Suyi Jiang, Longwen Zhang, Qixuan Zhang, Wei Yang, Lan Xu, and Jingyi Yu. 2024. GaussianHair: Hair Modeling and Rendering with Light-aware Gaussians. *arXiv preprint arXiv:2402.10483* (2024).

Lars Mescheder, Andreas Geiger, and Sebastian Nowozin. 2018. Which Training Methods for GANs Do Actually Converge?. In *International conference on machine learning*. PMLR, 3481–3490.

Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. 2020. NeRF: Representing Scenes As Neural Radiance Fields for View Synthesis. In *ECCV*.

Michael Niemeyer and Andreas Geiger. 2021. GIRAFFE: Representing Scenes As Compositional Generative Neural Feature Fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 11453–11464.

Maxim Nikolaev, Mikhail Kuznetsov, Dmitry Vetrov, and Aibek Alanov. 2024. HairFastGAN: Realistic and Robust Hair Transfer with a Fast Encoder-Based Approach. *arXiv preprint arXiv:2404.01094* (2024).

Roy Or-El, Xuan Luo, Mengyi Shan, Eli Shechtman, Jeong Joon Park, and Ira Kemelmacher-Shlizerman. 2022. StyleSDF: High-resolution 3D-Consistent Image and Geometry Generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 13503–13513.

Dongwei Pan, Long Zhuo, Jingtan Piao, Huiwen Luo, Wei Cheng, Yuxin Wang, Siming Fan, Shengqi Liu, Lei Yang, Bo Dai, et al. 2023. RenderMe-360: a Large Digital Asset Library and Benchmarks Towards High-fidelity Head Avatars. *Advances in Neural Information Processing Systems* 36 (2023), 7993–8005.

Sylvain Paris, Hector M. Briceño, and François X. Sillion. 2004. Capture of Hair Geometry from Multiple Images. *ACM Trans. Graph.* 23, 3 (aug 2004), 712–719.

Taesung Park, Ming-Yu Liu, Ting-Chun Wang, and Jun-Yan Zhu. 2019. Semantic Image Synthesis with Spatially-Adaptive Normalization. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2337–2346.

Stanislav Pidhorskyi, Tomas Simon, Gabriel Schwartz, He Wen, Yaser Sheikh, and Jason Saragih. 2024. Rasterized Edge Gradients: Handling Discontinuities Differentiably. arXiv preprint. arXiv:2405.02508.

Elad Richardson, Yuval Alaluf, Or Patashnik, Yotam Nitzan, Yaniv Azar, Stav Shapiro, and Daniel Cohen-Or. 2021. Encoding in Style: a StyleGAN Encoder for Image-to-Image Translation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

Shunsuke Saito, Liwen Hu, Chongyang Ma, Hikaru Ibayashi, Linjie Luo, and Hao Li. 2018. 3D Hair Synthesis Using Volumetric Variational Autoencoders. *ACM Trans. Graph.* 37, 6, Article 208 (dec 2018), 12 pages.

Shunsuke Saito, Gabriel Schwartz, Tomas Simon, Junxuan Li, and Giljoo Nam. 2024. Relightable Gaussian Codec Avatars. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 130–141.

Katja Schwarz, Yiyi Liao, Michael Niemeyer, and Andreas Geiger. 2020. GRAF: Generative Radiance Fields for 3D-Aware Image Synthesis. *Advances in Neural Information Processing Systems* 33 (2020), 20154–20166.

Yuefan Shen, Shunsuke Saito, Ziyan Wang, Olivier Maury, Chenglei Wu, Jessica Hodgins, Youyi Zheng, and Giljoo Nam. 2023. CT2Hair: High-Fidelity 3D Hair Modeling Using Computed Tomography. *ACM Transactions on Graphics* 42, 4 (2023), 1–13.

Vanessa Sklyarova, Egor Zakharov, Otmar Hilliges, Michael J. Black, and Justus Thies. 2024. Text-Conditioned Generative Model of 3D Strand-based Human Hairstyles. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 4703–4712.

Jingxiang Sun, Xuan Wang, Yichun Shi, Lizhen Wang, Jue Wang, and Yebin Liu. 2022. IDE-3D: Interactive Disentangled Editing for High-Resolution 3D-aware Portrait Synthesis. *ACM Transactions on Graphics (TOG)* 41, 6, Article 270 (2022), 10 pages.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Ł ukasz Kaiser, and Illia Polosukhin. 2017. Attention Is All You Need. In *Advances in Neural Information Processing Systems*, Vol. 30.

Ziyan Wang, Giljoo Nam, Tuur Stuyck, Stephen Lombardi, Michael Zollhöfer, Jessica Hodgins, and Christoph Lassner. 2022. 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*. 6143–6154.

Kelly Ward, Florence Bertails, Tae-Yong Kim, Stephen R Marschner, Marie-Paule Cani, and Ming C Lin. 2007. A Survey on Hair Modeling: Styling, Simulation, and Rendering. *IEEE transactions on visualization and computer graphics* 13, 2 (2007), 213–234.

Tianyi Wei, Dongdong Chen, Wenbo Zhou, Jing Liao, Zhentao Tan, Lu Yuan, Weiming Zhang, and Nenghai Yu. 2022. HairCLIP: Design Your Hair by Text and Reference Image. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 18072–18081.

Tianyi Wei, Dongdong Chen, Wenbo Zhou, Jing Liao, Weiming Zhang, Gang Hua, and Nenghai Yu. 2023. HairCLIPv2: Unifying Hair Editing Via Proxy Feature Blending. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 23589–23599.

Keyu Wu, Yifan Ye, Lingchen Yang, Hongbo Fu, Kun Zhou, and Youyi Zheng. 2022. 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*. 1526–1535.

Cheng-hsin Wuu, Ningyuan Zheng, Scott Ardisson, Rohan Bali, Danielle Belko, Eric Brockmeyer, Lucas Evans, Timothy Godisart, Hyowon Ha, Xuhua Huang, et al. 2022. Multiface: A Dataset for Neural Face Rendering. *arXiv preprint arXiv:2207.11243* (2022).

Zexiang Xu, Hsiang-Tao Wu, Lvdi Wang, Changxi Zheng, Xin Tong, and Yue Qi. 2014. Dynamic Hair Capture Using Spacetime Optimization. *ACM Transactions on Graphics (TOG)* 33, 6, Article 224 (nov 2014), 11 pages.

Yang Xue, Yuheng Li, Krishna Kumar Singh, and Yong Jae Lee. 2022. GIRAFFE HD: A High-Resolution 3D-aware Generative Model. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 18440–18449.

Egor Zakharov, Vanessa Sklyarova, Michael J Black, Giljoo Nam, Justus Thies, and Otmar Hilliges. 2024. Human Hair Reconstruction with Strand-Aligned 3D Gaussians. In *European Conference of Computer Vision (ECCV)*.

Meng Zhang, Menglei Chai, Hongzhi Wu, Hao Yang, and Kun Zhou. 2017. A Data-driven Approach to Four-view Image-based Hair Modeling. *ACM Trans. Graph.* 36, 4 (2017), 156–1.

Yujian Zheng, Zirong Jin, Moran Li, Haibin Huang, Chongyang Ma, Shuguang Cui, and Xiaoguang Han. 2023. 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*.

Yinglin Zheng, Hao Yang, Ting Zhang, Jianmin Bao, Dongdong Chen, Yangyu Huang, Lu Yuan, Dong Chen, Ming Zeng, and Fang Wen. 2022. General Facial Representation Learning in a Visual-linguistic Manner. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 18697–18709.

Yuxiao Zhou, Menglei Chai, Alessandro Pepe, Markus Gross, and Thabo Beeler. 2023. GroomGen: A High-Quality Generative Hair Model Using Hierarchical Latent Representations. *ACM Trans. Graph.* 42, 6, Article 270 (Dec. 2023), 16 pages.

Yi Zhou, Liwen Hu, Jun Xing, Weikai Chen, Han-Wei Kung, Xin Tong, and Hao Li. 2018. HairNet: Single-View Hair Reconstruction Using Convolutional Neural Networks. In *Proceedings of the European Conference on Computer Vision (ECCV)*. 235–251.

Peihao Zhu, Rameen Abdal, John Femiani, and Peter Wonka. 2021. Barbershop: GAN-based Image Compositing Using Segmentation Masks. *arXiv preprint arXiv:2106.01505* (2021).

Peihao Zhu, Rameen Abdal, Yipeng Qin, and Peter Wonka. 2020. Sean: Image Synthesis with Semantic Region-Adaptive Normalization. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 5104–5113.

Matthias Zwicker, Hanspeter Pfister, Jeroen Van Baar, and Markus Gross. 2001. EWA Volume Splatting. In *Proceedings Visualization, 2001. VIS'01.* IEEE, 29–538.