PanoHair: Detailed Hair Strand Synthesis on Volumetric Heads

Shashikant Verma
shashikant.verma@iitgn.ac.in
Shanmuganathan Raman
shanmuga@iitgn.ac.in

CVIG Lab
Indian Institute of Technology
Gandhinagar
India

Abstract

Achieving realistic hair strand synthesis is essential for creating lifelike digital humans, but producing high-fidelity hair strand geometry remains a significant challenge. Existing methods require a complex setup for data acquisition, involving multi-view images captured in constrained studio environments. Additionally, these methods have longer hair volume estimation and strand synthesis times, which hinder efficiency. We introduce PanoHair, a model that estimates head geometry as signed distance fields using knowledge distillation from a pre-trained generative teacher model for head synthesis. Our approach enables the prediction of semantic segmentation masks and 3D orientations specifically for the hair region of the estimated geometry. Our method is generative and can generate diverse hairstyles with latent space manipulations. For real images, our approach involves an inversion process to infer latent codes and produces visually appealing hair strands, offering a streamlined alternative to complex multi-view data acquisition setups. Given the latent code, PanoHair generates a clean manifold mesh for the hair region in under 5 seconds, along with semantic and orientation maps, marking a significant improvement over existing methods, as demonstrated in our experiments.

1 Introduction

High-fidelity 3D hair modeling is essential for creating realistic digital humans. A hairstyle often comprises tens of thousands of individual strands, which adds significant complexity to the process. Despite advancements in high-end capture systems, obtaining high-quality 3D hair models remains challenging, as accurately reconstructing the intricate geometry of hair is still difficult.

The hairstyle generation has traditionally depended on the intricate manual modeling process by expert 3D artists, with strand-based hair geometry being the most commonly adopted approach. Earlier methods [3, 4] involved manipulating hairstyles based on images and strokes as input cues, facilitating faster modeling. Building on recent progress in learning-based shape reconstruction, neural networks are now capable of generating 3D strand models using explicit point sequences [46], volumetric orientation fields [31, 44, 45], and implicit orientation fields [33, 39]. Recent approaches [33, 39, 40] first estimate the volume enclosed by the hair region and 3D orientations in the initial stage, followed by a hair-growing procedure in the second stage to generate the final strand-based hairstyle.

These methods often rely on complex data-processing pipelines that use multi-view images to reconstruct sparse point clouds, estimate camera positions, and perform bust fitting, which provides a shape before learning the fine geometry of the hair region.

Recent 3D-aware generative models [5, 7, 35] have demonstrated impressive results in generating highly realistic novel views. These models also offer straightforward inference by manipulating latent codes, eliminating the need for complex data acquisition and preprocessing pipelines. This latent space manipulation also enables indefinite data generation with novel views. Nevertheless, to fully harness the potential of these generative models for strand-based hair geometry, it is essential to precisely segment the hair region from the full-head geometry and estimate 3D surface orientations that guide strand growth for realistic hairstyle generation.

In this work, we introduce PanoHair, which leverages knowledge distillation from the pre-trained model, PanoHead [1], to generate the geometry of the complete human head. PanoHead serves as the teacher model, guiding PanoHair, the student model, to reconstruct a complete head model. Through knowledge distillation, PanoHair efficiently learns to generate high-quality signed distance fields (SDF) for the full-head geometry. Additionally, for each point on the iso-surface in volume, our approach predicts binary semantic information to classify whether it belongs to the hair region, along with a 3D orientation vector. However, this representation captures only the "outer shell" and does not fully account for the enclosed hair volume. To address this limitation, we propose a simple yet effective method that accurately closes the outer shell to represent the full hair volume. This straightforward approach ensures a fast estimation of the fully enclosed hair volume, an essential requirement for the hair-growing stage and realistic strand synthesis. Compared to NeuralHaircut [33], which requires approximately one day for hair geometry estimation, our method performs pivotal inversion given a real image to obtain the latent code, after which the geometry estimation takes just five seconds. The substantial reduction in processing time and the generation of high-quality hairstyles make PanoHair a significant advancement in 3D hair modeling.

Contributions. In summary, our contributions are as follows: **1.** We introduce PanoHair, a generative framework that distills knowledge from PanoHead to model human heads as SDFs while predicting semantic labels and 3D orientations. **2.** We propose a multi-view orientation projection loss to ensure view-consistent 3D orientation learning and mitigate ambiguities from single-view 2D projections. **3.** We employ a fast and robust hair volume extraction method using boolean operations, leveraging the structured and bounded nature of the learned implicit representation.

2 Related Work

We bridge the gap between generative volumetric head modeling and hair strand synthesis. This section reviews existing methods relevant to our approach.

Implicit Representations for Head and Hairs. A line of research focuses on parametric head models [9, 19], representing the head as a rigged mesh with a fixed topology. More recently, implicit representations, such as NeRF [23, 24] and implicit surfaces [36, 42] have emerged for modeling human heads, enabling applications like novel view synthesis and generative face modeling [5, 7, 35]. However, like parametric head models, volumetric representations inherently lack support for strand-based hair synthesis. A common approach involves estimating the hair region's occupancy from the head geometry and then using a data-driven method to achieve realistic hair growth within the bounded volume. For hair

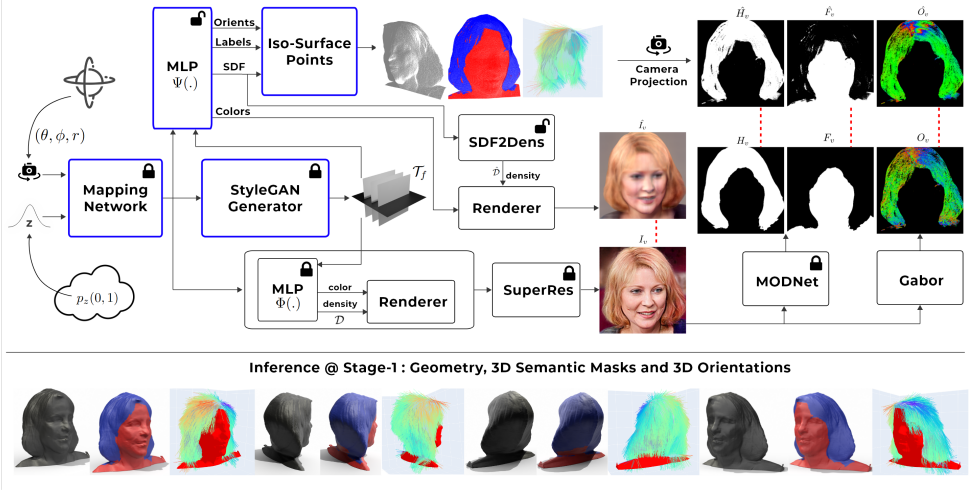

Figure 1: **PanoHair Overview.** Given latent code z and camera parameters $C_{\text{cam}} = (\theta, \phi, r)$, we first obtain tri-grid features $\mathcal{T}_f$ using the pre-trained PanoHead [1]. A set of MLPs $\Psi(.)$ then predicts semantic labels, 3D orientations, and SDF values. Training is supervised by distilling knowledge from PanoHair image renderings, Gabor orientations, and semantic segmentation. During inference, only the Blue-boxed components are required. The bottom row illustrates PanoHair's outputs—geometry, semantic labels, and 3D orientations.

modeling, earlier methods investigated different structured geometric primitives, such as wisp-based models [33], parametric surfaces [20, 26], and cylindrical primitives [6, 18, 37]. Strand-based representations for hair are now widely recognized as the standard for high-fidelity hair modeling [12, 28, 32]. Recent works encode surface strand geometry into latent codes controlling strand length, curvature, and orientation [33, 34].

Generative Modeling of Human Head. With significant progress in generating synthetic facial images using Generative Adversarial Networks (GANs) [15, 47] and Denoising Diffusion Probabilistic Models (DDPMs) [8, 13, 17], recent approaches leverage 3D-aware supervision to produce volumetric face models as view-consistent implicit representations. Models trained on large datasets of frontal 2D human faces [5, 7, 14, 35] are limited to generating geometry and novel view renderings of frontal faces, restricting their ability to estimate the complete geometry of a human head. This limitation restricts their applicability in generative hair modeling, as a substantial portion of hairstyle geometry remains inadequately captured. Recent approaches [1, 48] overcome this limitation by capturing the full volumetric geometry of the head, paving the way for strand-based hair modeling, a challenge we address in this work.

Strand based Hair Reconstruction. Image-based hair modeling [22, 25, 40, 41, 44] has primarily relied on orientation maps estimated from multiple-views [10, 27]. However, orientation maps can only supervise the outermost hair surface, resulting in ambiguity when reconstructing the internal hair strand structure. Furthermore, these methods typically demand intricate data preprocessing, including camera localization, sparse point-cloud reconstruction, and bust fitting. Data-driven methods address this limitation by incorporating hair structure priors. HairStep [45] constructs a database of manually annotated orientations to resolve directional ambiguities present in Gabor filter-based orientation estimations. Mean-

while, [34, 43] utilize diffusion priors to enhance the accuracy and fidelity of hair geometry refinement. NeuralHaircut [33] first extracts the hair volume by estimating geometry employing [36] and 3D orientation maps, guided by multi-view supervision, and then utilizes diffusion-based priors to generate hair strands within the bounded volume.

3 Method

Overview. Given a latent code, our network learns a signed distance function (SDF) to represent the volumetric head. Hair generation proceeds in three stages: (1) predict the SDF, segmentation labels, and 3D orientations; (2) estimate the hair-specific volume; and (3) grow strand-based hair within this volume. We leverage knowledge distillation from PanoHead [1], which predicts radiance field densities, guiding our SDF-based surface modeling. Our network also predicts binary segmentation labels on the SDF iso-surface to identify hair regions. Supervision for segmentation and orientation comes from Gabor maps and segmentation masks rendered from PanoHead. To extract the full hair volume, we fit a FLAME model [19] and apply boolean operations. The FLAME scalp provides structured strand roots, guided by the predicted 3D orientations.

3.1 Semantic Surface Extraction with Orientations

We represent the head as an implicit function modeled by $\Psi(.)$, where Ψ takes a 3D point as input and predicts its signed distance, a binary segmentation mask, and 3D orientation. We now describe how knowledge distillation from PanoHead [1] is used to train PanoHair.

Network Architecture. In Figure 1, we illustrate the complete pipeline for training $\Psi(.)$, a key module of Panohair for predicting color c, signed distance s, orientations o, and semantic labels m. Our proposed network, PanoHair, comprises several sub-modules: Mapping Network, Style-GAN Generator, MLP Network $\Psi(.)$, and SDF2Dens Network. Among these, the Mapping Network and Style-GAN Generator are pre-trained sub-modules adopted from PanoHead [1]. In [1], the MLP Network $\Phi(.)$ decodes tri-grid features into color and density values. Our network $\Psi(.)$ follows a similar architecture; however, in contrast, it predicts signed distances along with orientations and semantic labels essential for hair modeling. Following recent works [36, 42], we aim to model implicit surfaces with image-based reconstruction loss using differentiable volume rendering. To achieve this, we transform SDF values (s) to density values for volume rendering employing the SDF2Dens Module, as illustrated in Figure 1. The conversion is defined in Equation 1, where β is a learnable parameter.

$$\text{SDF2Dens}_\beta(s) = \begin{cases} \frac{1}{2\beta}\exp(\frac{s}{\beta}), & \text{if } s \leq 0 \\ \frac{1}{\beta}(1-\frac{1}{2}\exp(-\frac{s}{\beta}), & \text{if } s > 0 \end{cases} \tag{1}$$

Training requirements from PanoHead. Including details on why PanoHead is preferred over existing methods in the supplementary material, we now outline the training requirements for knowledge distillation using PanoHead. As shown in Figure 1, given a latent code $z \sim p_z(0,1)$ and a desired camera pose C_{cam}, the mapping network maps them into an intermediate latent code w. The Style-GAN generator takes w as input and produces tri-grid representation features $\mathcal{T}_f$. We sample features for a 3D point $\mathbf{x}$ along a casted ray by projecting its coordinates onto each tri-grid and retrieving the corresponding feature vector, t_x, using trilinear interpolation. The MLP Network $\Phi(.)$ takes t_x as input and predicts appearance color

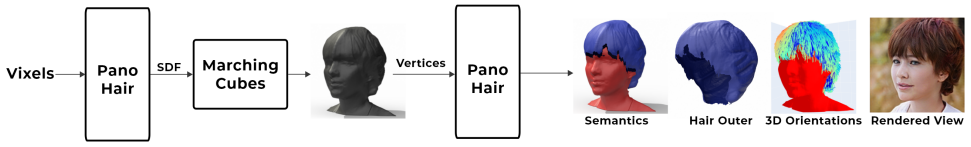

Figure 2: Inferencing hair's outer surface and orientations. A $512 \times 512 \times 512$ volumetric grid is sampled within $[-0.75, 0.75]$ to estimate SDF values. The geometry is extracted using marching cubes, and semantic labels and 3D orientations are computed for the mesh vertices.

and neural radiance density $\mathcal{D}$. Finally, volumetric rendering of this radiance field from the given camera views C_{cam} generates the image I, with a super-resolved version represented as I^+. We utilize the pre-trained weights from [1] and estimate the required components for training PanoHair, as outlined in Equation 2. Here, PH(.) refers to the PanoHead.

$$\text{PH}(z, C_{cam}) \longmapsto (\mathcal{T}_f, I^+, \mathcal{D}) \tag{2}$$

Let the sampled points along rays cast from camera C_{cam} in the volumetric grid be denoted as $\mathcal{X} \in \text{R}^{N\times 3}$. We train $\Psi(.)$ for all $\mathcal{X}$ to predict signed distance from contained surface $\mathcal{S} \in \text{R}^{N\times 1}$, color values $\mathcal{C} \in \text{R}^{N\times 3}$, Semantic labels $\mathcal{M} \in \text{R}^{N\times 1}$, and orientations $\mathcal{O} \in \text{R}^{N\times 3}$, i.e. $\Psi(\mathcal{X}) \longmapsto (\mathcal{S}, \mathcal{C}, \mathcal{M}, \mathcal{O})$. More precisely, given latent code w and tri-grid features t_f for each point $\mathbf{x} \in \mathcal{X}$ in the volume, we predict signed distance $s \in \mathcal{S}$, color $c \in \mathcal{C}$, binary semantic label $m \in \mathcal{M}$ and orientation $o \in \mathcal{O}$.

Image Synthesis. To facilitate volumetric rendering, we convert the signed distance values $s \in \mathcal{S}$ into density values $\sigma \in \hat{\mathcal{D}}$ using the SDF2Dens module, as defined in Equation 1. To synthesize an image $\hat{I}$, we employ a ray-marching approach [23] where rays $\mathbf{r} \in \mathcal{R}$ are cast from the camera center through each pixel of the image plane into the 3D space. Here, $\mathcal{R}$ represents the set of all rays corresponding to the pixels of an image with resolution $H \times W$. Along each ray $\mathbf{r}$, we sample K discrete points $\mathbf{x}_r^i$, where i denotes the index of the sampled point along the ray. For each sampled point $\mathbf{x}_r^i$, we retrieve its corresponding density $\sigma_r^i \in \hat{\mathcal{D}}$ and color $c_r^i \in \mathcal{C}$ from the predicted outputs of $\Psi(.)$. The final pixel color $C(r)$ for a given ray $\mathbf{r}$ is computed using volume rendering, where each sampled color c_r^i is accumulated with an opacity-weighted contribution based on the density values σ_r^i. This accumulation is performed using the volumetric rendering equation defined in Equation 3.

$$\mathcal{C}_r = \sum_{i=1}^{K} T_r^i \alpha_r^i c_r^i, \quad T_r^i = \prod_{j=1}^{i-1} (1 - \alpha_r^j), \quad \alpha_r^i = 1 - e^{-\sigma_r^i \delta_r^i} \tag{3}$$

Surface Points Attributes. We identify the visible surface points with the predicted SDF values at sampled points by detecting the first sign change along each ray. Since the rays are cast from the camera, the first sign change ensures that the detected point lies on the surface and is visible from the given viewpoint. Let the semantic labels of these identified surface points $x_S \in \mathcal{X}$ be denoted as $m_S \in \mathcal{M}$ and their 3D orientations as $o_S \in \mathcal{O}$. To obtain the semantic map $\hat{M}_{2D}$, we directly project m_S to image space using the camera matrix of C_{cam}. In contrast, to compute the 2D orientation map $\hat{\mathcal{O}}_{2D}$, we project the 3D orientations o_S into the camera space using Plücker line coordinates [11], similar to [30, 40]. An illustration of x_S, m_S, and o_S is shown in Figure 1 along with their projections onto image plane, namely $\hat{M}_{2D}$ and $\hat{\mathcal{O}}_{2D}$. We include qualitative results of projections in the *supp. material*.

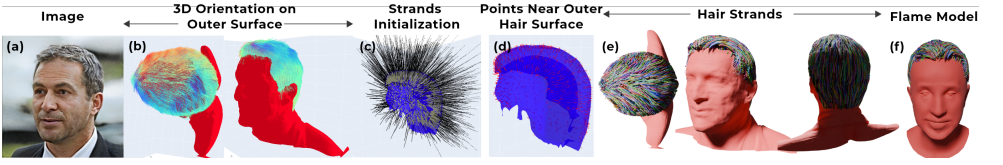

Figure 4: Strand-based hair growth. Given a latent code, (a) shows the generated image, and (b) visualizes 3D orientations on the outer hair surface. Strands are initialized using shape textures from [33]. Unlike their rasterization-based optimization, we directly optimize boundary strand points (d) to align with predicted orientations (b), producing the final hair (e). (f) shows the FLAME model with strands rooted at the scalp.

Training Pairs. Our training pairs for learning accurate SDF through volumetric rendering consist of the super-resolved image I^+ from [1] and the synthesized image $\hat{I}$. To train for semantic masks, we utilize [16] to estimate I_M, which contains hair and facial masks for I^+, and use them as ground truth for supervision. The training pairs for binary semantic classification of the hair region are I_M and $\hat{M}_{2D}$. Furthermore, for I^+, we estimate a 2D orientation map using a set of Gabor filters to obtain I_O, which serves as the ground truth for the predicted orientation map $\hat{\mathcal{O}}_{2D}$.

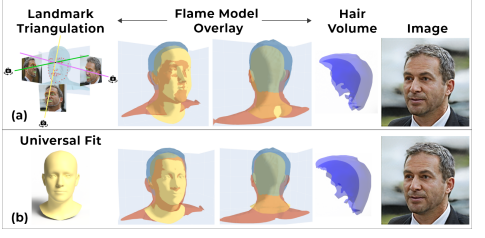

Figure 3: Estimating hair volume. (a) FLAME [19] model fitted to estimated landmarks. (b) Empirically computed universal fit, leveraging the structured representation of volumetric faces within a unit volume by design during training. Subsequently, we employ Boolean subtraction between the fitted FLAME model and the extracted mesh to compute volume.

Loss Functions. To train PanoHair, we enforce image reconstruction by optimizing a Mean Squared Error (MSE) loss between the super-resolved image I^+ and the synthesized image $\hat{I}$. For semantic mask prediction, we utilize a Binary Cross-Entropy (BCE) loss between the ground-truth mask I_M and the predicted mask $\hat{M}_{2D}$. Additionally, to accelerate convergence during the initial training phase, we impose an L1 loss between the density distribution $\mathcal{D}$ from PanoHead and the estimated densities $\hat{\mathcal{D}}$, which are derived by transforming the signed distance values predicted by $\Psi(.)$ using Equation 1. This auxiliary loss is applied only for the initial training iterations and is gradually reduced to zero afterward using a scheduler $\delta(t)$, which decays to zero after iteration t. We define our loss function $\mathcal{L}$ in Equation 4.

$$\mathcal{L} = \underbrace{\lambda_1 \cdot \delta(t) \cdot \|\mathcal{D} - \hat{\mathcal{D}}\|_1}_{\textbf{Density Loss}} + \underbrace{\lambda_2 \cdot \|I^+ - I\|^2}_{\text{Image Reconstruction Loss}} + \underbrace{\lambda_3 \cdot \text{BCE}(I_M, \hat{M}_{2D})}_{\textbf{Semantic Loss}} + \mathcal{L}_{\text{orient}} \quad (4)$$

Here, λ represents the weights assigned to each loss term, and $\mathcal{L}_{orient}$ denotes the orientation loss, which we now describe in detail. Since ground-truth 3D orientations are unavailable for direct supervision, we instead learn to predict them by enforcing losses on their 2D projections, specifically between $\hat{\mathcal{O}_{2D}}$ and the Gabor orientation map I_O. However, this projection-based orientation loss is inherently ill-posed and ambiguous, as multiple 3D ori-

Figure 5: Latent space interpolation between two face codes and hair growth. The second row presents multi-view renderings, followed by semantic masks and computed hair volume in the third and fourth rows. The final rows illustrate the resulting strand-based hair.

entation vectors can result in the same 2D projection on the image plane. Unlike prior works such as [30, 33], which iteratively optimize orientations across multiple views of the same subject, our method is generative, where the identity of the synthesized head is controlled by the latent code z. Consequently, the subject changes in each iteration, preventing us from employing such an optimization strategy. Moreover, our predicted orientations must remain consistent across novel views and align with the geometry implicitly represented through the signed distance field (SDF) for each z. Hence, we impose $\mathcal{L}_{\text{orient}}$ with two key constraints. First, the predicted orientation vectors must be tangential to the surface. Second, to mitigate the ambiguity introduced by projection, the learning process should enforce multi-view consistency in each iteration for every latent code z. Mathematical formulations, training weights, and additional details on $\mathcal{L}_{\text{orient}}$ are provided in the *supp material*.

Inference. During inference for novel-view generation, only the Blue boxes in Figure 1 are required. SDF-based iso-surfaces yield point cloud geometry suited for gradient propagation via point-based rendering in training. To extract a full head mesh, we sample voxels in a volumetric grid, apply marching cubes [21] on SDF values, and evaluate semantic masks and orientations at mesh vertices. By selecting points classified as hair, we obtain the hair's outer surface and 3D orientations, as shown in Figure 2.

Model	Input	Hair Seg. (IoU)	Orient. Error at Hair Region ($\mathcal{L}_{\text{proj}}$)	Hair Volume Est. time	Strand Reconstruction time
NeuralHaircut [34]	Multi-View	0.815	0.391	10-12 Hrs	15-18 Hrs
HairStep [35]	Single View	0.682	0.526	~**3 sec**	~**5 mins.**
PanoHair (Ours)	Latent code	**0.846**	**0.368**	~5 sec	4-5 Hrs.

Table 1: We report IoU between the 2D projection of the hair surface and MODNet [16] output as ground truth. Orientation error is measured using $\mathcal{L}_{proj}$ between the projected orientation $\hat{\mathcal{O}}_{2D}$ and high-confidence Gabor orientations. Additionally, we report the total runtime on an NVIDIA RTX-4090 GPU for a single head.

3.2 Hair Volume Estimation and Growing

We require a closed, bounded volume where hair strands can reside to facilitate hair growth. While we have estimated the hair's outer surface, we must also determine the scalp location within the generated head. **Hair Volume.** One approach is to fit a FLAME head model [19] onto the extracted geometry, which requires identifying known landmark locations on the mesh. Let $\mathcal{G}$ represent the extracted full head geometry from PanoHair, and let ${}^{i}I^{+}$ be a novel view rendered from ${}^{i}C_{\text{cam}}$. We estimate facial landmarks for each view using [2]. We use triangulation to compute landmark positions on $\mathcal{G}$ given the known camera parameters. We then align the FLAME model landmarks with the estimated ones through iterative fitting, optimizing shape, expression, and pose parameters to minimize landmark distance [19], as illustrated in Figure 3(a). Since the learned implicit representation produces structured geometries $\mathcal{G}$ in terms of pose and bounds, we empirically determine a translation $\mathbf{t_f}$ and scale $\mathbf{s}_f$ for a neutral FLAME model to position the FLAME scalp appropriately, as shown in Figure 3(b). We call this fitting Universal Fit, which robustly fits all possible generated heads. While this fit may be less accurate for facial regions, it is sufficient for scalp localization, which is our primary focus. Let the fitted FLAME model geometry be represented as $\mathcal{G}_F$ With the approximate locations of both the scalp and the hair's outer surface, we extract a closed, bounded volume for the hair region. To achieve this, we first extrude the hair's outer surface toward the scalp, ensuring that the extrusion is sufficiently scaled so that the resulting geometry remains entirely within the FLAME model. However, this extrusion process can introduce a non-manifold mesh topology unsuitable for direct Boolean operations. To address this, we convert both the FLAME head geometry ($\mathcal{G}_F$) and the extruded hair mesh into a volumetric representation and perform a Boolean subtraction, yielding a closed, bounded hair volume $\mathcal{G}_h$. Figure 3 qualitatively compares two approaches for obtaining $\mathcal{G}_h$.

Hair Strand Growing. Given the 3D orientations $\mathcal{O}$ at surface points $x_S \in \mathcal{X}$ labeled with hair semantics $\mathcal{M}$, and the hair volume $\mathcal{G}_h$, we adopt a strand-based geometry reconstruction approach similar to [33]. Figure 4 illustrates the optimization steps involved. For more details on strand reconstruction, we refer readers to [33]. In contrast to [33], we exclude silhouette and RGB reconstruction losses on multi-view inputs and instead guide strand growth using 3D orientations and the hair volume $\mathcal{G}_h$ predicted by PanoHair.

4 Discussion

Latent Space Exploration and Inversion. The generative design of PanoHair ensures 3D orientations consistent with geometry, enabling diverse hairstyle synthesis without relying on limited and hard-to-obtain 3D synthetic data. Figure 5 showcases semantic labels and hair

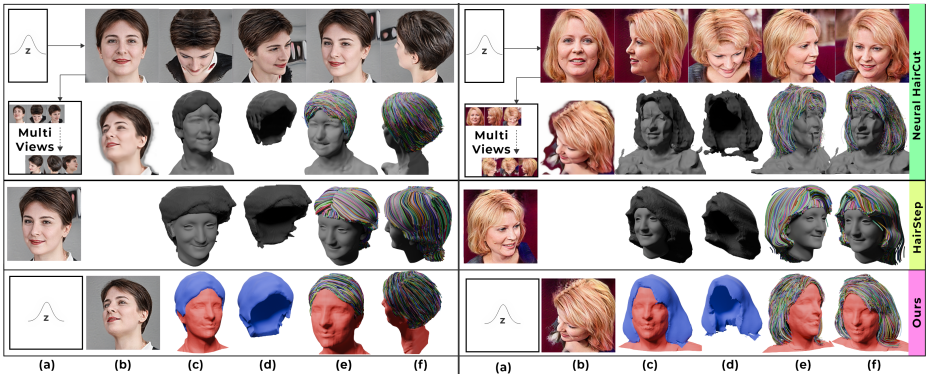

Figure 6: Qualitative comparison of the proposed approach with state-of-the-art methods, NeuralHaircut [33] and HairStep [45]. NeuralHaircut is trained using multi-view renderings to estimate hair volume and orientations, while HairStep takes a single generated image as input. (a) depicts the input for each method, (b) presents a novel view, (c) and (d) show the full and hair volume meshes, and (e) and (f) illustrate strand-based hair from two views.

volume geometry extracted by our approach. Smooth transitions in the latent code result in consistent orientation maps, enabling smooth transitions between strand-based hairstyles in linearly interpolated latent space. To generate hairstyles from a single real-world image, we first optimize the latent code z using an inversion process [29]. With the obtained latent code, the PanoHair framework predicts hair volume and 3D orientations, enabling novel-view synthesis or hair strand generation, as shown in Figure 7.

Comparisons. While no existing method directly addresses hair growth in a generative volumetric setting for direct comparison, we evaluate our approach against NeuralHaircut [33] and HairStep [45] under the following assumptions. NeuralHaircut requires multi-view data to first estimate hair and bust geometry employing [36], while HairStep operates on a single image for input. We randomly sample five latent codes and generate 40 novel views for each, storing camera parameters, Gabor orientations, and masks required by [33]. Of these, 35 views are used for training, while five per code are reserved for evaluation. For HairStep, a single frontal view from the 40 training images serves as input, with the same evaluation data. In contrast, our approach directly utilizes the latent code z, which generates the corresponding renderings. The data comprises five different heads, totaling 200 multi-view images, with 25 instances allocated for validation. Table 1 presents the quantitative results of different methods and their runtime. Hair segmentation is reported using IoU on validation set camera views, while orientation error is evaluated by computing $\mathcal{L}_{\text{orient}}$. Our method achieves superior performance in both hair-region estimation and orientation accuracy. Regarding runtime, [33] requires per-head training with multi-view inputs, taking

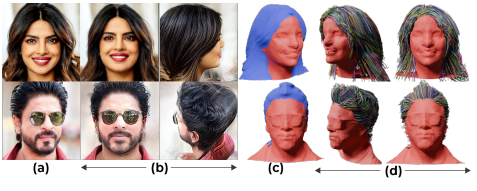

Figure 7: Latent code optimization using Pivotal Inversion [29] to match the given real image (a). The optimized latent code generates novel views (b), hair masks (c), and 3D orientations. The final strand-based hair reconstruction is shown in (d).

approximately 12 hours, whereas our approach completes in around 5 seconds. HairStep is faster but produces poor and inconsistent results across views.

5 Conclusion

In this study, we present PanoHair, a 3D-aware generative framework that capitalizes on the capabilities of tri-grid representations for full-head modeling. Our methodology represents geometry as implicit signed distance functions (SDFs) while concurrently predicting semantic labels and view-consistent 3D orientations on iso-surfaces. The inference process for estimating hair volume and orientation is notably efficient, requiring approximately five seconds. The rapid and generative characteristics of PanoHair facilitate the synthesis of a wide variety of hairstyles, and we demonstrate that our predicted three-dimensional orientations significantly enhance the generation of detailed strand-based hair structures. We believe this work will excite many researchers to enhance the framework to perform hair strand synthesis for full-head models.

Acknowledgements. We would like to acknowledge the support of the Jibaben Patel Chair in AI for this work.

References

[1] Sizhe An, Hongyi Xu, Yichun Shi, Guoxian Song, Umit Y. Ogras, and Linjie Luo. Panohead: Geometry-aware 3d full-head synthesis in 360deg. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 20950–20959, June 2023.

[2] Adrian Bulat and Georgios Tzimiropoulos. How far are we from solving the 2d & 3d face alignment problem?(and a dataset of 230,000 3d facial landmarks). In *Proceedings of the IEEE international conference on computer vision*, pages 1021–1030, 2017.

[3] Menglei Chai, Lvdi Wang, Yanlin Weng, Yizhou Yu, Baining Guo, and Kun Zhou. Single-view hair modeling for portrait manipulation. *ACM Transactions on Graphics (TOG)*, 31(4):1–8, 2012.

[4] Menglei Chai, Lvdi Wang, Yanlin Weng, Xiaogang Jin, and Kun Zhou. Dynamic hair manipulation in images and videos. *ACM Transactions on Graphics (TOG)*, 32(4):1–8, 2013.

[5] 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. Efficient geometry-aware 3d generative adversarial networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 16123–16133, 2022.

[6] Byoungwon Choe and Hyeong-Seok Ko. A statistical wisp model and pseudophysical approaches for interactive hairstyle generation. *IEEE Transactions on Visualization and Computer Graphics*, 11(2):160–170, 2005.

[7] Yu Deng, Jiaolong Yang, Jianfeng Xiang, and Xin Tong. Gram: Generative radiance manifolds for 3d-aware image generation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10673–10683, 2022.

[8] Zheng Ding, Xuaner Zhang, Zhihao Xia, Lars Jebe, Zhuowen Tu, and Xiuming Zhang. Diffusionrig: Learning personalized priors for facial appearance editing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12736–12746, 2023.

[9] Simon Giebenhain, Tobias Kirschstein, Markos Georgopoulos, Martin Rünz, Lourdes Agapito, and Matthias Nießner. Learning neural parametric head models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 21003–21012, 2023.

[10] Goesta H Granlund. In search of a general picture processing operator. *Computer Graphics and Image Processing*, 8(2):155–173, 1978.

[11] Richard Hartley and Andrew Zisserman. *Multiple view geometry in computer vision*. Cambridge university press, 2003.

[12] Jerry Hsu, Tongtong Wang, Zherong Pan, Xifeng Gao, Cem Yuksel, and Kui Wu. Sag-free initialization for strand-based hybrid hair simulation. *ACM Transactions on Graphics (TOG)*, 42(4):1–14, 2023.

[13] Ziqi Huang, Kelvin CK Chan, Yuming Jiang, and Ziwei Liu. Collaborative diffusion for multi-modal face generation and editing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6080–6090, 2023.

[14] Kaiwen Jiang, Shu-Yu Chen, Feng-Lin Liu, Hongbo Fu, and Lin Gao. Nerffaceediting: Disentangled face editing in neural radiance fields. In *SIGGRAPH Asia 2022 Conference Papers*, pages 1–9, 2022.

[15] Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. Analyzing and improving the image quality of stylegan. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8110–8119, 2020.

[16] Zhanghan Ke, Jiayu Sun, Kaican Li, Qiong Yan, and Rynson WH Lau. Modnet: Real-time trimap-free portrait matting via objective decomposition. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 1140–1147, 2022.

[17] Gwanghyun Kim, Taesung Kwon, and Jong Chul Ye. Diffusionclip: Text-guided diffusion models for robust image manipulation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2426–2435, 2022.

[18] Tae-Yong Kim and Ulrich Neumann. Interactive multiresolution hair modeling and editing. *ACM Transactions on Graphics (TOG)*, 21(3):620–629, 2002.

[19] Tianye Li, Timo Bolkart, Michael J Black, Hao Li, and Javier Romero. Learning a model of facial shape and expression from 4d scans. *ACM Trans. Graph.*, 36(6):194–1, 2017.

[20] Wenqi Liang and Zhiyong Huang. An enhanced framework for real-time hair animation. In *11th Pacific Conference onComputer Graphics and Applications, 2003. Proceedings.*, pages 467–471. IEEE, 2003.

[21] William E Lorensen and Harvey E Cline. Marching cubes: A high resolution 3d surface construction algorithm. In *Seminal graphics: pioneering efforts that shaped the field*, pages 347–353. 1998.

[22] Linjie Luo, Cha Zhang, Zhengyou Zhang, and Szymon Rusinkiewicz. Wide-baseline hair capture using strand-based refinement. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 265–272, 2013.

[23] 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. *Communications of the ACM*, 65(1):99–106, 2021.

[24] Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. Instant neural graphics primitives with a multiresolution hash encoding. *ACM transactions on graphics (TOG)*, 41(4):1–15, 2022.

[25] 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.

[26] Paul Noble and Wen Tang. Modelling and animating cartoon hair with nurbs surfaces. In *Proceedings Computer Graphics International, 2004.*, pages 60–67. IEEE, 2004.

[27] Sylvain Paris, Hector M Briceno, and François X Sillion. Capture of hair geometry from multiple images. *ACM transactions on graphics (TOG)*, 23(3):712–719, 2004.

[28] Emmanuel Piuze, Paul G Kry, and Kaleem Siddiqi. Generalized helicoids for modeling hair geometry. In *Computer Graphics Forum*, volume 30, pages 247–256. Wiley Online Library, 2011.

[29] Daniel Roich, Ron Mokady, Amit H Bermano, and Daniel Cohen-Or. Pivotal tuning for latent-based editing of real images. *ACM Transactions on graphics (TOG)*, 42(1): 1–13, 2022.

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

[31] Shunsuke Saito, Liwen Hu, Chongyang Ma, Hikaru Ibayashi, Linjie Luo, and Hao Li. 3d hair synthesis using volumetric variational autoencoders. *ACM Transactions on Graphics (TOG)*, 37(6):1–12, 2018.

[32] 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 (TOG)*, 42(4):1–13, 2023.

[33] 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*, pages 19762–19773, 2023.

[34] Vanessa Sklyarova, Egor Zakharov, Otmar Hilliges, Michael J Black, and Justus Thies. Text-conditioned generative model of 3d strand-based human hairstyles. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4703–4712, 2024.

[35] Shashikant Verma and Shanmuganathan Raman. Semfaceedit: Semantic face editing on generative radiance manifolds. In *International Conference on Pattern Recognition*, pages 46–62. Springer, 2025.

[36] Peng Wang, Lingjie Liu, Yuan Liu, Christian Theobalt, Taku Komura, and Wenping Wang. Neus: Learning neural implicit surfaces by volume rendering for multi-view reconstruction. *arXiv preprint arXiv:2106.10689*, 2021.

[37] Tao Wang and Xue Dong Yang. Hair design based on the hierarchical cluster hair model. *Geometric modeling: techniques, applications, systems and tools*, pages 329–359, 2004.

[38] Yasuhiko Watanabe and Yasuhito Suenaga. A trigonal prism-based method for hair image generation. *IEEE Computer Graphics and applications*, 12(01):47–53, 1992.

[39] 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.

[40] Keyu Wu, Lingchen Yang, Zhiyi Kuang, Yao Feng, Xutao Han, Yuefan Shen, Hongbo Fu, Kun Zhou, and Youyi Zheng. Monohair: High-fidelity hair modeling from a monocular video. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 24164–24173, 2024.

[41] Lingchen Yang, Zefeng Shi, Youyi Zheng, and Kun Zhou. Dynamic hair modeling from monocular videos using deep neural networks. *ACM Transactions on Graphics (TOG)*, 38(6):1–12, 2019.

[42] Lior Yariv, Yoni Kasten, Dror Moran, Meirav Galun, Matan Atzmon, Basri Ronen, and Yaron Lipman. Multiview neural surface reconstruction by disentangling geometry and appearance. *Advances in Neural Information Processing Systems*, 33:2492–2502, 2020.

[43] Egor Zakharov, Vanessa Sklyarova, Michael Black, Giljoo Nam, Justus Thies, and Otmar Hilliges. Human hair reconstruction with strand-aligned 3d gaussians. In *European Conference on Computer Vision*, pages 409–425. Springer, 2024.

[44] Meng Zhang and Youyi Zheng. Hair-gan: Recovering 3d hair structure from a single image using generative adversarial networks. *Visual Informatics*, 3(2):102–112, 2019.

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

[46] 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)*, pages 235–251, 2018.

[47] Peihao Zhu, Rameen Abdal, Yipeng Qin, and Peter Wonka. Sean: Image synthesis with semantic region-adaptive normalization. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5104–5113, 2020.

[48] Wojciech Zielonka, Timo Bolkart, and Justus Thies. Instant volumetric head avatars. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4574–4584, 2023.