

GeomHair: Reconstruction of Hair Strands from Colorless 3D Scans

RACHMADIO NOVAL LAZUARDI*, Technical University of Munich, Germany

ARTEM SEVASTOPOLSKY*, Technical University of Munich, Germany

EGOR ZAKHAROV, ETH Zürich, Switzerland

MATTHIAS NIESSNER, Technical University of Munich, Germany

VANESSA SKLYAROVA, Max Planck Institute for Intelligent Systems, Germany & ETH Zürich, Switzerland

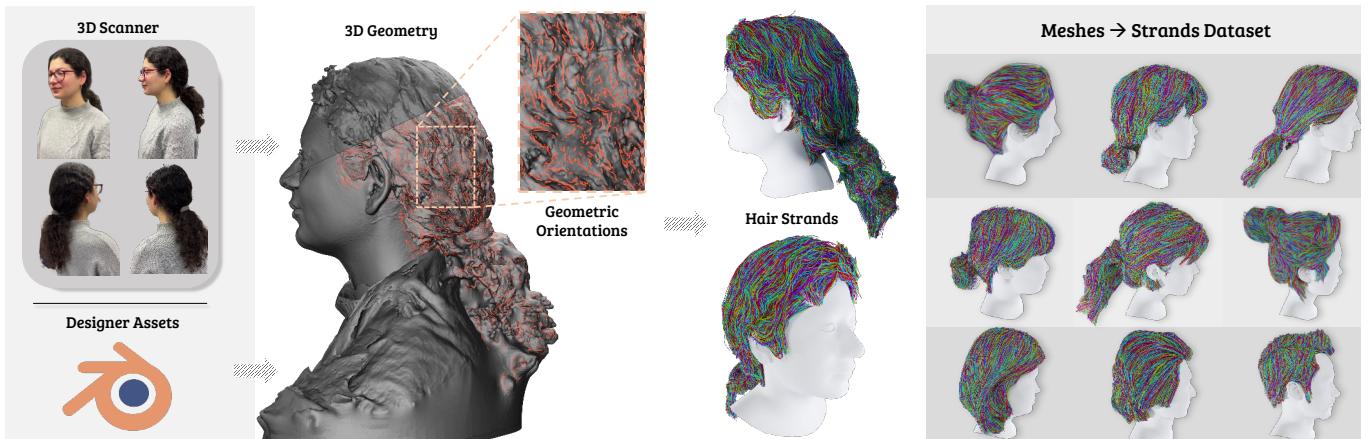


Fig. 1. We present *GeomHair* – a method for reconstructing complete hair strands representations from 3D scans that can be obtained from various sources, such as handheld 3D scanners, designer assets, and others. Our method extracts information about guiding sharp features directly from the scan geometry by employing a combination of 3D and 2D orientation detectors and fits strands with a diffusion prior conditioned on the scan-specific prompt embedding. We also provide a dataset of reconstructed 3D assets obtained using our method from the meshes produced by a structured light scanner.

We propose a novel method that reconstructs hair strands directly from colorless 3D scans by leveraging multi-modal hair orientation extraction. Hair strand reconstruction is a fundamental problem in computer vision and graphics that can be used for high-fidelity digital avatar synthesis, animation, and AR/VR applications. However, accurately recovering hair strands from raw scan data remains challenging due to human hair’s complex and fine-grained structure. Existing methods typically rely on RGB captures, which can be sensitive to the environment and can be a challenging domain for extracting the orientation of guiding strands, especially in the case of challenging hairstyles. To reconstruct the hair purely from the observed geometry, our method finds sharp surface features directly on the scan and estimates strand orientation through a neural 2D line detector applied to the renderings of scan shading. Additionally, we incorporate a diffusion prior trained on a diverse set of synthetic hair scans, refined with an improved noise schedule, and adapted to the reconstructed contents via a scan-specific text prompt. We demonstrate that this combination of supervision signals enables accurate reconstruction of both simple and intricate hairstyles without relying on color information. To facilitate further research, we introduce Strands400, the largest publicly available dataset of hair strands with detailed surface geometry extracted from real-world data, which contains reconstructed hair strands from the scans of 400 subjects.

Additional Key Words and Phrases: 3D Computer Vision, Human Head, Hair Reconstruction, Mesh-Based Graphics, Hair Strands

Authors’ addresses: Rachmadio Noval Lazuardi*, Technical University of Munich, Germany; Artem Sevastopolsky*, Technical University of Munich, Germany; Egor Zakharov, ETH Zürich, Switzerland; Matthias Nießner, Technical University of Munich, Germany; Vanessa Sklyarova, Max Planck Institute for Intelligent Systems, Germany & ETH Zürich, Switzerland.

1 INTRODUCTION

We propose a new method of modeling strand-based hair from a single colorless 3D scan of a human head. Reconstructing hair as 3D curves is a highly challenging problem at the intersection of computer vision and graphics. However, it is an essential component for synthesizing high-fidelity digital avatars, as reconstructing not only the rigid surface of the avatar but also full-length strands allows for realistic animation under the physical parameters of the environment. In addition, modeling the interior of the hair volume via strands is essential to achieving a realistic and physically based rendering of hair dynamics, which is crucial for telepresence, motion capture, gaming, and other applications of human head modeling.

Recovering hair strands is problematic and often highly ill-posed due to human hair’s complex and fine-grained structure since only a portion of the hair strands form the observed surface. To solve this ill-posedness, different approaches have been proposed to infer the location and orientation of the coarse strand structure (often referred to as guide strands) of real-life subjects from various modalities.

Earlier works attempted to estimate a simplified hair surface via proxy geometry [Fu et al. 2022; Gafni et al. 2020; Grassal et al. 2021; Khakhulin et al. 2022; McGuire et al. 2021; Park et al. 2020; Wang et al. 2022; Zheng et al. 2021, 2022], making the result not always suitable for plausible animation. Later attempts required rather tedious capture with setups such as light stages to obtain

* Denotes equal contribution

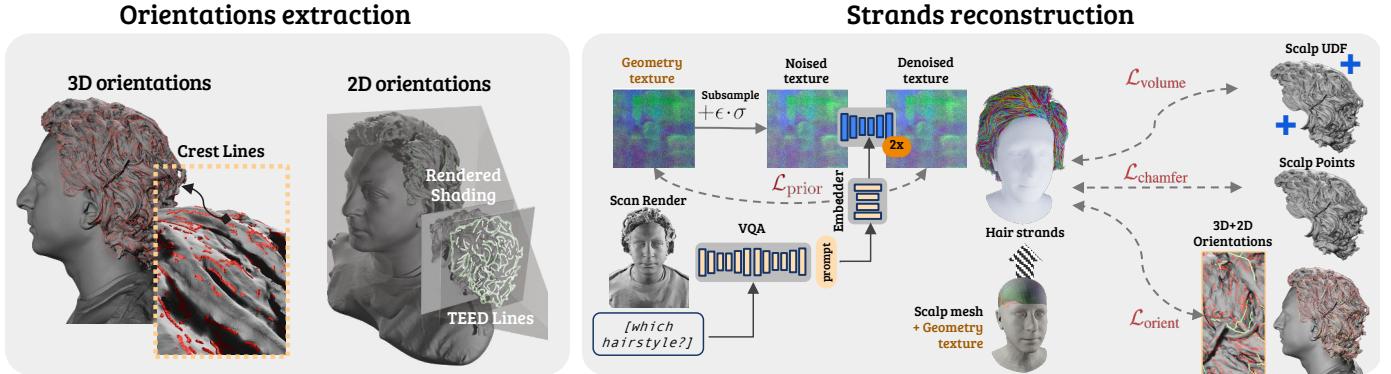


Fig. 2. Overview of our GeomHair framework. Our method consists of two main stages: orientations extraction (left) and strands reconstruction (right). In the orientation extraction stage, we extract complementary orientation signals by combining 3D orientations from crest lines with 2D orientations obtained from TEED features applied to rendered shading of the scans. During reconstruction, we optimize a geometry texture to generate hair strands while enforcing multiple constraints: the orientation loss ($\mathcal{L}_{\text{orient}}$) ensures strand growth aligns with our extracted 3D+2D orientation field, volume loss ($\mathcal{L}_{\text{volume}}$) keeps strands near the surface, and Chamfer distance ($\mathcal{L}_{\text{chamfer}}$) promotes uniform coverage of the hair volume. To enhance realism, we additionally incorporate a diffusion prior ($\mathcal{L}_{\text{prior}}$) conditioned on hairstyle descriptions generated by a VQA model analyzing the input scan.

RGB information about hair surface under various lighting conditions [McGuire et al. 2021; Nam et al. 2019; Rosu et al. 2022; Wang et al. 2022, 2021]. Modern approaches [Luo et al. 2024; Sklyarova et al. 2023a; Wu et al. 2024; Zakharov et al. 2024; Zhou et al. 2024] lift the requirement of employing a capture setup by drawing information about the guide strands directly from RGB video frames, such as a 360° handheld smartphone capture under static conditions. While these approaches are the most affordable in terms of the capture effort, relying on RGB input alone can introduce sensitivity to lighting conditions, shadows, and occlusions, making the respective methods less effective for reconstructing intricate hairstyles such as wavy or curly hair. Additionally, distributing large collections of high-resolution RGB videos of people with diverse demographics can be complicated by privacy issues.

At the same time, recent developments in 3D scanning introduced novel, publicly available datasets of a large number of scanned meshes of human heads with a high level of detail. Similarly, manually constructed, rigged meshes of human characters remain the primary representations in gaming and animation. Furthermore, the need for the methods of affordable large-scale reconstruction is motivated by the recent progress in generative models that currently have to resort to augmentation techniques to ensure the required dataset size and diversity [He et al. 2024] or rely on costly handcrafted parametric hair models [Zhou et al. 2023a].

To this end, we introduce a novel method for reconstructing hair strands directly from the scan geometry without relying on the captured photographs or the surface color. By focusing on geometric rather than color information, our approach can find more prominent guiding sharp features of the surface, thus enabling the reconstruction of high-quality 3D hair strands with higher robustness. The ability to reconstruct hair directly from 3D scans, which is available in large amounts, opens more data acquisition opportunities for diverse hair strand collections. In turn, this introduces more room for data-driven generative models for hair by reducing the need for tedious synthetic data creation [Zhou et al. 2023a] or

heavy augmentation techniques [He et al. 2024] necessary for their training.

Our method is based on optimizing the strands, guided by the orientation losses, diffusion prior, and volume guidance. Contrary to existing image-based approaches, which derive the primary supervision signal — orientation maps — by processing input RGB images, we employ two sources of orientation extraction directly from the 3D scan. First, we find sharp ridges and ravines on the mesh via Crest Lines [Yoshizawa et al. 2005] and calculate their orientations. Second, enhance orientation extraction by rendering scan-based shading from multiple viewpoints and applying a neural 2D line orientation detector [Soria et al. 2023] as auxiliary supervision. This combination of geometric and learned 2D cues provides a more accurate and reliable estimate of strand orientations, even for complex hairstyles. Furthermore, we integrate a diffusion prior trained on synthetic hair scans from the HAAR [Sklyarova et al. 2023b] dataset. Unlike in previous state-of-the-art [Sklyarova et al. 2023a; Zakharov et al. 2024], our approach refines the diffusion prior through additional diffusion steps and an optimized noise schedule, enabling better modeling of intricate hair structures. To further enhance adaptation to specific hairstyles, we condition the prior on the embeddings of a text prompt, generated by a vision language model [Li et al. 2023] for the scan renderings and aggregated over various viewpoints. This novel prompting mechanism provides greater flexibility in handling diverse hairstyles and improves the generalization of the reconstruction process.

To summarize, our contributions are as follows:

- We propose a first reconstruction method that works with colorless scans using extraction of 3D and 2D orientations from the mesh surface.
- We improve the diffusion-based score distillation sampling prior with text guidance and an improved denoising schedule.
- We use our method, named GeomHair, to transform the scans from both existing and newly collected public data to construct the Strands400 dataset, the largest publicly available

real-world dataset of hair strands with diverse hairstyles and demographics of 400 participants that have precisely registered corresponding surface geometry.

2 RELATED WORK

2D and 3D Line Detection. 3D hair reconstruction heavily relies on 3D line detection, as hair can be effectively approximated with 3D polylines. Different 2D line detectors have been explored for hair reconstruction, including edge detection filters such as Canny [Canny 1986] and Gabor [Paris et al. 2004], with the latter receiving wide adoption [Hu et al. 2015; Kuang et al. 2022; Luo et al. 2024; Rosu et al. 2022; Wu et al. 2024, 2022; Zakharov et al. 2024; Zhou et al. 2024, 2018]. However, while providing better results for RGB images, Gabor-based line orientation maps fail in low or strong directional light environments. Also, Gabor filtering heavily relies on the reflectance profile of the hair fibers [Paris et al. 2004], leading to failures for some of the data sources, such as colorless scans. Thus, unlike previous hair modeling work, our method relies on the other supervision signals based on denoised 2D edges [Soria et al. 2023] and 3D crest lines [Yoshizawa et al. 2005]. In combination, these two supervision sources allow us to achieve superior performance compared to classically used 2D Gabor filters and their 3D analogs.

Strand-based Hair Priors. Hair strands have a complicated internal structure that requires priors for accurate modeling. These priors are trained from the synthetic datasets in the form of the generative hairstyle models [He et al. 2024; Rosu et al. 2022; Sklyarova et al. 2023a,b; Zhou et al. 2023a] or strand-based hair growing models [Kuang et al. 2022; Wu et al. 2024, 2022; Zheng et al. 2023; Zhou et al. 2024]. Strand-based hair-growing is usually formulated as a conditional strand generation given a 3D field of hair directions. This model can be formulated as an algorithm that traverses this field [Paris et al. 2004] or a neural network [Wu et al. 2024, 2022; Zhou et al. 2024].

However, these approaches do not allow end-to-end usage of this prior during the strand-based hairstyle reconstruction; instead, they heavily rely on the quality of volumetric hair reconstruction. On the other hand, generative hairstyle priors are highly versatile and can work with various data modalities. They utilize neural generative approaches, such as latent diffusion [Sklyarova et al. 2023a,b], VAEs [Zhou et al. 2023a], or GANs [He et al. 2024]. These models are trained from a dataset of synthetic assets and can be used for downstream tasks, such as reconstruction [He et al. 2024; Rosu et al. 2022; Sklyarova et al. 2023a], text-conditional 3D asset generation [Sklyarova et al. 2023b] or semantic editing [Rosu et al. 2022; Zhou et al. 2023a]. In our work, we utilize a state-of-the-art text-conditional deep generative prior HAAR [Sklyarova et al. 2023b] to regularize the hair interior in our reconstructions during an end-to-end optimization process.

Strand-based Hair Reconstruction. The data domains that are used for strand-based hair reconstruction include single images [Wu et al. 2022; Zheng et al. 2023; Zhou et al. 2018], monocular videos [Luo et al. 2024; Sklyarova et al. 2023a; Wu et al. 2024; Zakharov et al. 2024; Zhou et al. 2024], multi-view captures [Rosu et al. 2022; Takimoto et al. 2024] and even CT scans [Shen et al. 2023]. Image-to-hairstyle

regression methods utilize deep neural networks to learn a regression model capable of estimating a strand-based 3D hairstyle from 2D information, such as Gabor-based orientation maps or depth maps [Wu et al. 2022; Zheng et al. 2023; Zhou et al. 2018]. This allows easy in-the-wild applications such as asset generation directly from the images, but are not as versatile as generative priors. For example, these methods rely highly on high-frequency feature extraction from the images to predict the underlying strand-based hairstyle. Monocular videos and multiview data present a richer and more constrained data source in terms of the ease of 3D reconstruction but ultimately suffer from the same issue. The closest problem setting to ours is hair modeling from CT scans. However, CT scanners provide highly accurate volumetric data, including the internal structure of the hairstyle [Shen et al. 2023], which is not available in colorless scans. Our problem setting allows our method to have a wide range of applications, starting from a reconstruction based on structured light 3D scanners and ending with the populating the existing mesh-based hair assets with high-frequency strand-based geometry.

3 METHOD

In this section, we present an overview of the method. We take inspiration from the approach used in Neural Haircut [Sklyarova et al. 2023a], which we briefly review in Subsec. 3.1. In Subsec. 3.2, we introduce GeomHair and highlight the key differences that enable direct training on a 3D scan.

3.1 Recap: Neural Haircut

In Neural Haircut [Sklyarova et al. 2023a], hairstyle is parametrized using *hair map* H with resolution 256×256 corresponding to a scalp region of the 3D head model. Each texel of the map defines a single hair strand that is represented as a 3D polyline with L points: $S_i = \{\mathbf{p}_i^l\}_{l=1}^L$. We define directions between nearest points as $\mathbf{d}_i^l = \mathbf{p}_i^{l+1} - \mathbf{p}_i^l$ and normalized $\mathbf{b}_i^l = \mathbf{d}_i^l / \|\mathbf{d}_i^l\|_2$.

The reconstruction method consists of two stages, *coarse* and *fine* reconstruction. During the coarse stage, they estimate hair and bust geometry as signed distance functions (SDFs) $f_{\text{hair}}, f_{\text{bust}} : \mathbb{R}^3 \rightarrow \mathbb{R}$ as well as train an additional surface 3D hair orientation field $\beta : \mathbb{R}^3 \rightarrow \mathbb{S}^2$. β is optimized using a minimal angular difference between its projections into camera space and Gabor orientation maps estimated from ground-truth RGB images.

During the fine stage, hairstyle in the form of *latent hairstyle map* $T = \{\mathbf{z}_i\}_{i=1}^N$ is optimized to match geometry f_{hair} and orientations β obtained from the coarse stage. After training, the hairstyle from T could be easily retrieved by using the pre-trained decoder \mathcal{G} . At each iteration, a set of points N sampled from the fitted FLAME scalp, strands embeddings $\{\mathbf{z}_i\}_{i=1}^N$ are retrieved from these locations from T and decoded into 3D space using the pre-trained decoder \mathcal{G} . Strands S_i are supervised to lie inside volume by \mathcal{L}_{vol} :

$$\mathcal{L}_{\text{vol}} = \sum_{i=1}^N \sum_{l=1}^L \mathbb{I}[f_{\text{hair}}(\mathbf{p}_i^l) > 0] (f_{\text{hair}}(\mathbf{p}_i^l))^2, \quad (1)$$

where \mathbb{I} denotes the indicator function. To ensure that the visible outer surface of f_{hair} is uniformly covered, a one-way Chamfer

distance is optimized between the sampled points \mathbf{x}_k and its nearest neighbor from the strand \mathbf{p}_k :

$$\mathcal{L}_{\text{chm}} = \sum_{k=1}^K \|\mathbf{x}_k - \mathbf{p}_k\|_2^2, \quad (2)$$

Finally, orientations of all strands that lie near the visible surface are supervised using the learned direction field β :

$$\mathcal{L}_{\text{orient}} = \sum_{m=1}^M (1 - |\mathbf{b}_m \cdot \beta(\mathbf{p}_m)|), \quad (3)$$

where \mathbf{p}_m are a set of strand points near the surface and \mathbf{b}_m - their corresponding directions. Full geometry loss can be written as follows:

$$\mathcal{L}_{\text{geom}} = \mathcal{L}_{\text{vol}} + \lambda_{\text{chm}} \mathcal{L}_{\text{chm}} + \lambda_{\text{orient}} \mathcal{L}_{\text{orient}}. \quad (4)$$

Finally, diffusion-based loss $\mathcal{L}_{\text{diff}}$ forces \mathbf{T} to be in the distribution of realistic hairstyles. Diffusion loss is applied in low-resolution \mathbf{T}_{LR} space by using a pre-trained on synthetic hairstyles diffusion prior:

$$\mathcal{L}_{\text{diff}} = \mathbb{E}_{\mathbf{T}_{\text{LR}}, \sigma, \epsilon} \left[\lambda_{\text{diff}}(\sigma) \cdot \|\mathcal{D}(\mathbf{x}, \sigma) - \mathbf{T}_{\text{LR}}\|_2^2 \right], \quad (5)$$

where $\lambda_{\text{diff}}(\sigma)$ is a weighting function, and the expectation is approximated via sampling. For more details, please follow the supplementary doc.

The final loss for the fine stage is the following:

$$\mathcal{L}_{\text{fine}} = \mathcal{L}_{\text{geom}} + \lambda_{\text{diff}} \mathcal{L}_{\text{diff}} + \lambda_{\text{render}} \mathcal{L}_{\text{render}}, \quad (6)$$

Here, $\mathcal{L}_{\text{render}}$ represents a photometric loss that enhances the hairstyle quality. Since our setup lacks color information, we describe how we adapt the first two terms to our scenario.

3.2 GeomHair

Overview. Our method reconstructs 3D hair strands by extracting geometric features from full 3D head scans. We begin by applying a segmentation algorithm [Xie et al. 2021] to isolate the hair region from the rest of the head. Once a clean hair mesh is obtained, we estimate hair orientations by integrating features from crest lines [Yoshizawa et al. 2005], applied to the hair volume, with 2D features extracted using [Soria et al. 2023] from multi-view renderings of the colorless scan. These geometric features form the basis of our strand synthesis process.

To generate 3D hair strands, we utilize the parametric model from [Rosu et al. 2022], which operates on a latent geometry texture. We enforce constraints to keep the strands near the surface, ensure dense coverage of the hair volume, and align their growth direction with our estimated orientation vectors, following a similar approach to [Sklyarova et al. 2023a]. Neural Haircut is the first method to introduce a diffusion prior as a regularization technique for generating realistic hairstyles. We refine this pipeline further by incorporating the conditional diffusion prior from [Sklyarova et al. 2023b] and introducing a novel scheduling strategy with multiple denoising steps to enhance reconstruction quality.

3.3 Estimating Hair Orientation

3D crest lines. To accurately extract 3D hair strands from the surface of our high-quality scans, we utilize crest lines—salient surface features defined by curvature extrema. Crest lines are especially effective for hair extraction, as they capture the sharp bends and curves characteristic of hair strands.

Following [Yoshizawa et al. 2005], we define crest lines on a surface S as the set of points where one of the principal curvatures reaches an extremum along its corresponding curvature direction. Mathematically, we can express this as:

$$e_{\max} = \frac{\partial k_{\max}}{\partial t_{\max}} = 0 \quad (\text{for convex crest lines}), \quad (7)$$

$$e_{\min} = \frac{\partial k_{\min}}{\partial t_{\min}} = 0 \quad (\text{for concave crest lines}), \quad (8)$$

where k_{\max} and k_{\min} are the maximum and minimum principal curvatures, respectively, and t_{\max} and t_{\min} are their corresponding principal directions. The quantities e_{\max} and e_{\min} are called extremality coefficients.

We use a local polynomial fitting approach to identify crest lines on our 3D hair scans. For each vertex in the mesh, we fit a cubic polynomial of the form:

$$h(x, y) = \frac{1}{2}(b_0x^2 + 2b_1xy + b_2y^2) + \frac{1}{6}(d_0x^3 + 3d_1x^2y + 3d_2xy^2 + d_3y^3)$$

After computing these values, we trace the crest lines across the mesh by connecting points where the extremality coefficients vanish. To ensure we capture only the most salient features of the hair structure, we employ a thresholding scheme based on the cyclideness measure:

$$C = \sqrt{|e_{\max}|^2 + |e_{\min}|^2} \quad (9)$$

Let $C = \{c_1, \dots, c_N\}$ be a set of crest lines, where each crest line c_i is a set of points $\{\mathbf{p}_i\}_{l=1}^{L_i}$, and L_i is the number of points in the i -th crest line, which may vary for different crest lines. We treat a single crest line as a guiding hair strand and compute local orientations for each.

Specifically, we estimate local curvature along each crest line c_i , which informs our adaptive window sizing. For each point p_i , we compute a local coordinate frame using PCA [Abdi and Williams 2010] on a window of neighboring points. We apply an adaptive smoothing step to refine these initial estimates, balancing noise reduction with the preservation of directional changes. The resulting normalized orientations are then used as a supervision signal for the strands-based reconstruction in the next step.

2D TEED feature extractor. Previous image-based hair strand reconstruction relies on Gabor filters [Paris et al. 2004] to obtain 2D orientation maps. Gabor filters have been effective in reconstructing realistic hairstyles. They serve as a supervision signal derived from RGB images to capture high-frequency details of hair directionality. However, this approach is limited when applied to 3D mesh inputs, as the absence of color information makes it challenging to infer fine-grained hair orientations. To address this, we adapt the concept to the more challenging case of colorless 3D scan inputs.

Specifically, we generate 2D orientation maps by applying the 2D feature extractor TEED [Soria et al. 2023] to multi-view renderings of the colorless scans. After detecting edges in the rendered views with TEED, we perform skeletonization to thin the edges to single-pixel-width paths, effectively reducing noise and removing parasitic lines while preserving the overall hair structure. These skeletonized paths are then converted into a graph representation to extract the longest continuous paths, capturing the primary flow of the hair strands. For each path segment, we compute orientation angles by analyzing the directional vectors between connected points. Finally, we lift both the graph structure and its associated orientation vectors back into 3D space, establishing a directional field that guides the subsequent hair strand reconstruction.

3.4 Improved diffusion prior.

In Neural Haircut, the authors introduce a global diffusion-based prior to enhance the realism of reconstructed hairstyles. They train their diffusion model on the USC-HairSalon [Hu et al. 2015] dataset, which contains 343 synthetic hairstyles. Building on this idea, our approach leverages HAAR [Sklyarova et al. 2023b], a text-to-hairstyle diffusion model, which is trained not only on USC-HairSalon [Hu et al. 2015] but also on CT2Hair [Shen et al. 2023] and an additional artist-curated dataset specializing in curly hairstyles.

We integrate HAAR’s conditional diffusion capabilities into our reconstruction pipeline through three key components:

First, we incorporate HAAR’s text-conditioning pipeline, where hairstyle descriptions are generated using a VQA model [Li et al. 2023] with renderings of the 3D head scan as input. These descriptions are then processed through BLIP [Li et al. 2022] to generate text embeddings, which serve as conditioning inputs for the diffusion model.

Second, we implement a denoising scheduler that begins with high noise levels and progressively reduces them throughout optimization. This provides structural guidance in the early iterations while allowing geometry loss to dominate in later stages. Additionally, we introduce multiple denoising steps to improve adherence to the prompt.

Third, we enhance scalp map estimation, which defines the regions where hair grows. To achieve this, we project segmented 3D scans onto the FLAME template model and filter out non-hair regions. In our experiments, we demonstrate that this conditional diffusion approach provides a stronger prior than the unconditional model in Neural Haircut, resulting in more accurate reconstructions of complex hairstyles.

3.5 Strands-based reconstruction.

Our strand-based reconstruction approach builds upon the fine reconstruction stage of Neural Haircut, with adjustments made to accommodate our 3D input scans. We optimize a latent hairstyle map $T = \{z_i\}_{i=1}^N$ to match the underlying geometry and orientations, utilizing various supervision signals and distance function representations.

Instead of signed distance functions (SDFs), we utilize an unsigned distance function (UDF) approximator [Zhou et al. 2023b] for representing hair geometry. This choice is motivated by the fact

that our extracted hair mesh lacks an internal structure, making the use of UDF more suitable for our scenario. We pre-train the UDF model $F_{\text{udf}} : \mathbb{R}^3 \rightarrow \mathbb{R}^+$ on our extracted hair mesh and use it to supervise the volume loss in our training pipeline:

$$\mathcal{L}'_{\text{vol}} = \sum_{i=1}^N \sum_{l=1}^L (F_{\text{udf}}(\mathbf{p}_i^l))^2, \quad (10)$$

where \mathbf{p}_i^l represents the l -th point of the i -th strand.

Our orientation supervision combines signals from two complementary sources: 3D crest lines and the lifted from 2D to 3D features from TEED extractor. The orientation loss is formulated as:

$$\mathcal{L}'_{\text{orient}} = \alpha \cdot \mathcal{L}_{\text{orient}}^{\text{3DO}} + (1 - \alpha) \cdot \mathcal{L}_{\text{orient}}^{\text{2DO}}, \quad (11)$$

where α is a weighting parameter, $\mathcal{L}_{\text{orient}}^{\text{3DO}}$ and $\mathcal{L}_{\text{orient}}^{\text{2DO}}$ follow the same angular difference structure as in Eq. 3, but use orientation fields derived from 3D crest lines (β_{3D}) and lifted TEED features (β_{2D}) respectively. Based on empirical results, we find that $\alpha = 0.5$ provides the best reconstruction quality, creating an equal balance between the 3D and 2D orientation signals. For each generated strand point near the surface, we apply this orientation loss by finding its nearest neighbor in the supervision signal.

Furthermore, we apply the Chamfer distance loss \mathcal{L}_{chm} , which ensures uniform coverage of the visible outer surface, following established formulations. The complete geometry loss looks similar to Eq. 4 with new refined losses:

$$\mathcal{L}'_{\text{geom}} = \mathcal{L}'_{\text{vol}} + \lambda_{\text{chm}} \mathcal{L}_{\text{chm}} + \lambda_{\text{orient}} \mathcal{L}'_{\text{orient}}. \quad (12)$$

As previously described, we employ HAAR’s conditional diffusion model rather than an unconditional one. This diffusion loss maintains the form shown in Eq. (15) while extending it with the text condition embedding c derived from our automated description pipeline. Furthermore, we propose to decrease the noise level with time and do several steps of denoising compared to Neural Haircut [Sklyarova et al. 2023a]. Our final optimization objective focuses on geometric and diffusion-based constraints:

$$\mathcal{L}_{\text{final}} = \mathcal{L}'_{\text{geom}} + \lambda_{\text{diff}} \mathcal{L}_{\text{diff}}, \quad (13)$$

which simplifies the loss function in Eq. 6 by omitting rendering-based components.

4 EXPERIMENTS

We use an extended version of NPHM [Giebenhain et al. 2023] dataset that consists of scans captured by Artec Eva scanners. For more details, please follow the Supplementary document. We fit FLAME [Li et al. 2017] head to each scan and segment the hair region by applying [Xie et al. 2021]. In this work, we compare our method against the state-of-the-art approaches in hair reconstruction from monocular video, such as Neural Haircut [Sklyarova et al. 2023a], Gaussian Haircut [Zakharov et al. 2024], and MonoHair [Wu et al. 2024]. For qualitative comparison, we capture 360-degree monocular video using a smartphone along with the head scan. For quantitative evaluation, we use two scenes from Cem Yuksel’s Hair Models [Yuksel et al. 2009].

4.1 Implementation details

We train our model on a single GPU for 75k steps using Adam optimizer with a learning rate of 0.001 employed with multi-step scheduler where the gamma value is 0.5. On an RTX A6000, a single training typically takes about 14 hours. For diffusion we decrease the noise level from 80 to 0.5. Also, for the first 10k iterations we make 2 denoising steps with classifier free guidance weight of 4.0.

4.2 Results

Qualitative evaluation. We compare the results of our hairstyle reconstructions from colorless scans with the one obtained from RGB images using Neural Haircut [Sklyarova et al. 2023a], Gaussian Haircut [Zakharov et al. 2024] and MonoHair [Wu et al. 2024]. For Neural Haircut [Sklyarova et al. 2023a], Gaussian Haircut [Zakharov et al. 2024], and MonoHair [Wu et al. 2024], the input is a 360° RGB video, and for our method, the input is a corresponding 3D scan without color of the same person, closely following NPHM dataset [Giebenhain et al. 2023] setup. The results are reflected in Fig. 3). We may observe that our method operates on par with Neural Haircut and Gaussian Haircut but the fitting process is more regularized and better adapts to complicated hairstyles. With MonoHair, hairstyle is generally correct but the hair length is at times erroneous. Our method also demonstrates better picking of important guiding strands and hairline.

Quantitative evaluation. We compare our method with Neural Haircut [Sklyarova et al. 2023a] and Gaussian Haircut [Zakharov et al. 2024] on two synthetic hairstyles for straight and wavy hair. To do that, we render it in Blender [Community 2018] and use the obtained images as input to baseline methods. As our method works on scans, we voxelize strands and use them as input for our method. In Table 1, we provide comparison results. Our method significantly outperforms the baselines for straight hair across most metrics. Although it shows a small decline in Precision on curly scenes, it excels in Recall and F-score compared to the others.

Ablation study. We conduct a comprehensive ablation study to highlight the significance of each component in our pipeline, see Tab. 1. First, we verify the effectiveness of orientation, volume, and diffusion-based losses: w/o $\mathcal{L}_{\text{orient}}$, w/o \mathcal{L}_{vol} , and w/o $\mathcal{L}_{\text{diff}}$. Then, we ablate the importance of using 3D orientations and 2D orientations: w/o 2DO and w/o 3DO. We then checked if replacing our proposed 3D orientations calculation with the 3D analog of Gabor filters (w/ 3DO Gabor) affects the performance and found it to deteriorate significantly. Then, we validate our proposed data terms with a baseline Neural Haircut SDS prior: w/ NH prior. After replacing NH prior with HAAR prior without introducing a proposed denoising schedule (HAAR w/o noise schedule), we notice a drop across almost all metrics compared to the baseline. The same happens when we remove text-conditioning from HAAR (HAAR w/o prompt). However, after introducing the text conditioning and denoised schedule, we observed a significant boost in metrics (Ours), validating the effectiveness of all individual components in our approach.

Applications. Lastly, we show the potential of our approach for dataset creation from mesh-based hair assets. We show the results

of our model on artist-created hairstyles; see Figure 4. We notice that our method can obtain nice-looking results purely from the provided geometry without using any color information as input.

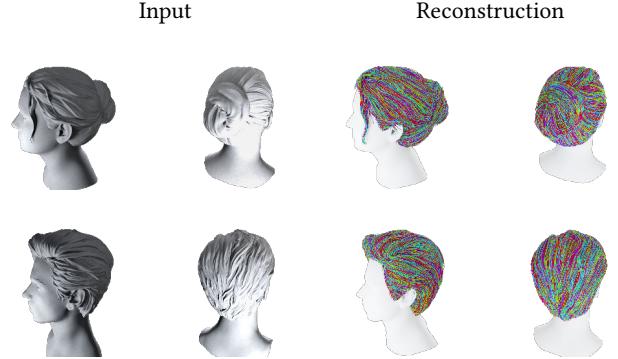


Fig. 4. Results on 3D-designed assets. Here, we demonstrate the results of fitting to a synthetic, hand-carved 3D mesh from a 3D stock (*CGTrader, author: ZStuff*). Our method is capable of reconstructing a plausible collection of strands, even when not so many guiding strands can be observed.

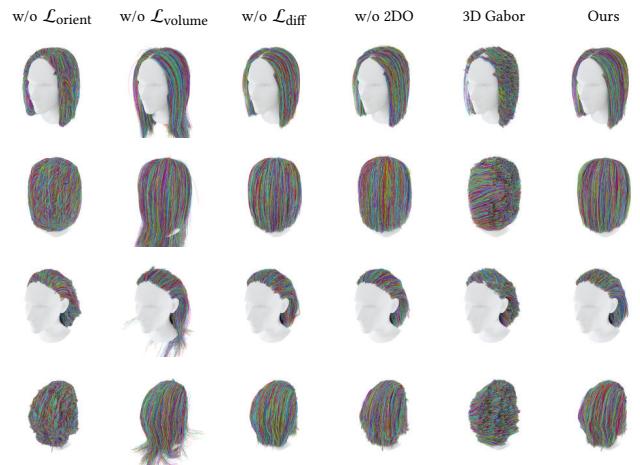


Fig. 5. Ablation over the various components of the GeomHair pipeline. Here, we demonstrate the most important components of the pipeline that provide the largest improvement. The comparison is done over the scans by Cem Yüksel made from hand-drawn hair strands. 3D Gabor corresponds to replacing Crest Lines algorithm with more simple Gabor filtering (corresponds to the w/ 3DO Gabor ablation in Table 1).

Limitations and future work. In this work, we show the possibilities of our model to extract strand-based geometry for colorless scans. This model could be used for scalable dataset generation. While our method performs well on wavy and straight hairstyles, we could see that on very curly scenes we reconstruct some unrealistic strands. One reason for that is that scans failed to capture such high-frequency details, which amplify the noisiness of proposed orientation estimators. A potential direction could be to further improve the hairstyle prior in order to better denoise the internal geometry based on surface orientations.

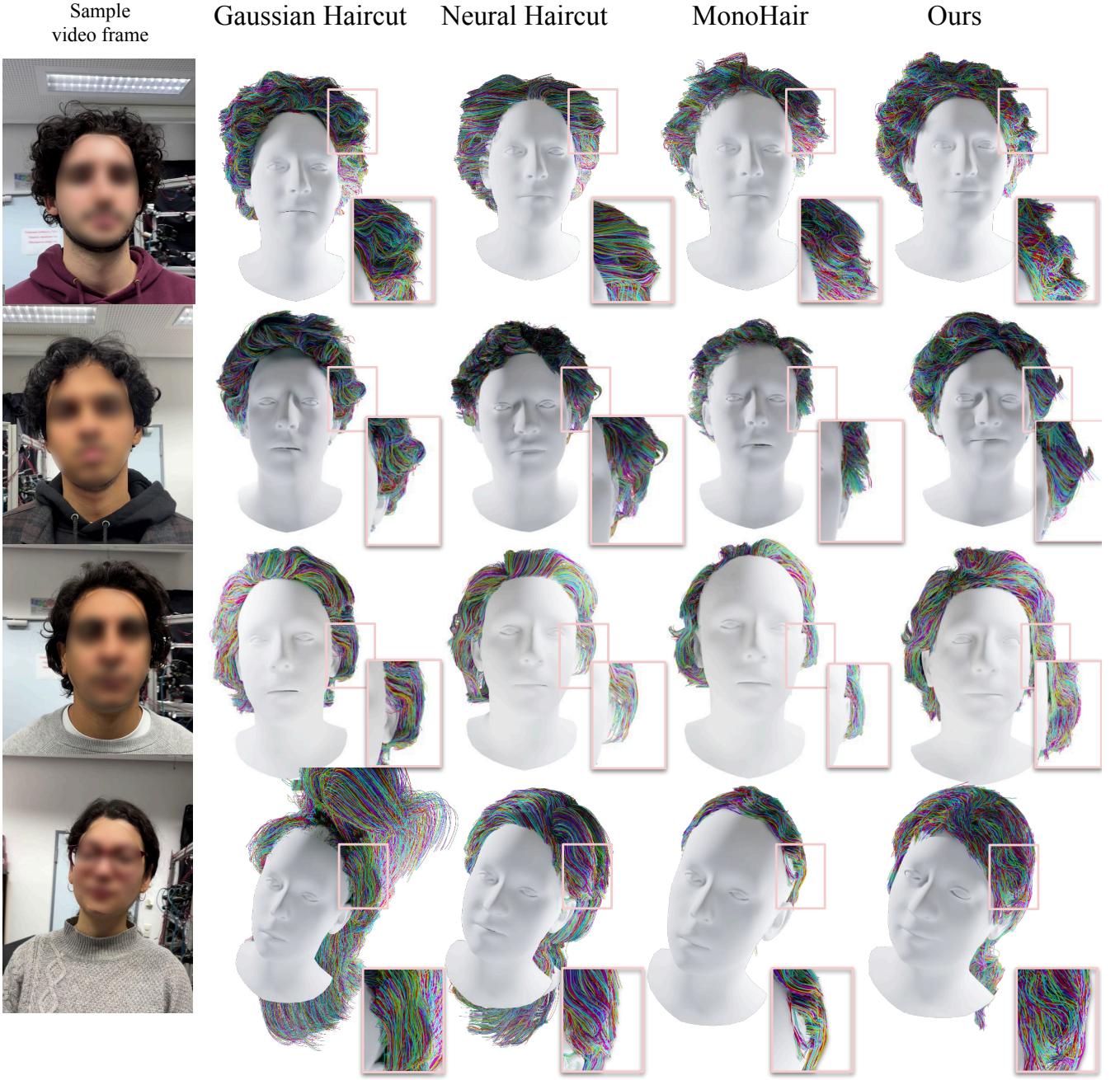


Fig. 3. Comparison of our method with state-of-the-art hair reconstruction methods across five different scenes. For Neural Haircut [Sklyarova et al. 2023a], Gaussian Haircut [Zakharov et al. 2024], and MonoHair [Wu et al. 2024], the input is a 360° RGB video, and for our method, the input is a corresponding 3D scan without color of the same person, closely following NPHM dataset [Giebenhain et al. 2023] setup. Faces blurred for anonymity purposes.

5 CONCLUSION

In this work, we propose a method for reconstructing human hair directly from a colorless 3D scan. Our main idea is to combine several training signals, including a diffusion prior, to guide strand generation inferred from the surface geometry. As demonstrated

by the quantitative and qualitative evaluations on newly collected scans and handcrafted assets, our method is capable of reconstructing challenging hairstyles and performs on par with methods that require color information. The reconstructed hairstyles are suitable for animation and modeling. Additionally, we present the Strand400

| Method | Straight hair | | | | | | | | | | | | Curly hair | | | | | | | | | | | |
|-----------------------------------|----------------|-------------|-------------|-------------|----------------|-------------|-------------|-------------|----------------|-------------|-------------|-------------|----------------|-------------|-------------|-------------|----------------|-------------|-------------|-------------|----------------|-------------|-------------|-------------|
| | 2/20 3/30 4/40 | | | | 2/20 3/30 4/40 | | | | 2/20 3/30 4/40 | | | | 2/20 3/30 4/40 | | | | 2/20 3/30 4/40 | | | | 2/20 3/30 4/40 | | | |
| | Precision | | | | Recall | | | | F-score | | | | Precision | | | | Recall | | | | F-score | | | |
| w/o $\mathcal{L}_{\text{orient}}$ | 19.2 | 36.7 | 51.8 | 18.8 | 33.2 | 50.1 | 19.0 | 34.9 | 50.9 | 15.3 | 31.2 | 47.5 | 14.5 | 29.4 | 44.9 | 14.9 | 30.3 | 46.2 | 14.9 | 30.3 | 46.2 | 14.9 | 30.3 | 46.2 |
| w/o $\mathcal{L}_{\text{volume}}$ | 57.4 | 62.5 | 64.5 | 72.9 | 85.6 | 96.1 | 64.2 | 72.2 | 77.2 | 26.9 | 49.4 | 66.7 | 30.6 | 52.7 | 69.0 | 28.6 | 51.0 | 67.8 | 28.6 | 51.0 | 67.8 | 28.6 | 51.0 | 67.8 |
| w/o $\mathcal{L}_{\text{diff}}$ | 79.4 | 90.8 | 93.9 | 69.4 | 84.2 | 90.5 | 74.1 | 87.4 | 92.2 | 27.2 | 50.5 | 68.5 | 30.0 | 52.0 | 68.5 | 28.5 | 51.2 | 68.5 | 28.5 | 51.2 | 68.5 | 28.5 | 51.2 | 68.5 |
| w/o 2DO | 81.6 | 88.5 | 91.2 | 81.1 | 88.8 | 91.6 | 81.3 | 88.6 | 91.4 | 30.2 | 52.8 | 68.9 | 32.7 | 55.1 | 70.9 | 31.4 | 53.9 | 69.9 | 31.4 | 53.9 | 69.9 | 31.4 | 53.9 | 69.9 |
| w/o 3DO | 62.3 | 83.3 | 90.4 | 48.6 | 71.1 | 83.3 | 54.6 | 76.7 | 86.7 | 22.1 | 42.3 | 59.9 | 22.6 | 42.1 | 59.1 | 22.3 | 42.2 | 59.5 | 22.3 | 42.2 | 59.5 | 22.3 | 42.2 | 59.5 |
| Crest Lines → Gabor | 2.5 | 6.5 | 13.6 | 2.4 | 7.8 | 16.6 | 2.4 | 7.1 | 15.0 | 2.1 | 5.3 | 10.3 | 1.8 | 5.1 | 10.9 | 1.9 | 5.2 | 10.6 | 1.9 | 5.2 | 10.6 | 1.9 | 5.2 | 10.6 |
| w/ NH prior | 78.1 | 88.6 | 92.3 | 71.2 | 84.4 | 90.5 | 74.5 | 86.4 | 91.4 | 27.5 | 50.8 | 68.7 | 30.2 | 52.1 | 68.3 | 28.8 | 51.4 | 68.5 | 28.8 | 51.4 | 68.5 | 28.8 | 51.4 | 68.5 |
| HAAR w/o noise schedule | 78.5 | 88.5 | 91.8 | 61.0 | 75.4 | 85.1 | 68.7 | 81.4 | 88.3 | 27.3 | 50.9 | 69.4 | 26.5 | 47.7 | 64.6 | 26.9 | 49.2 | 66.9 | 26.9 | 49.2 | 66.9 | 26.9 | 49.2 | 66.9 |
| HAAR w/o prompt | 78.6 | 88.5 | 91.9 | 60.9 | 75.3 | 85.0 | 68.6 | 81.4 | 88.3 | 28.2 | 52.3 | 70.8 | 26.9 | 48.0 | 64.8 | 27.5 | 50.1 | 67.7 | 27.5 | 50.1 | 67.7 | 27.5 | 50.1 | 67.7 |
| Ours | 80.4 | 89.0 | 91.7 | 72.6 | 85.6 | 90.8 | 76.3 | 87.3 | 91.2 | 28.2 | 52.0 | 70.0 | 29.9 | 51.8 | 68.3 | 29.0 | 51.9 | 69.1 | 29.0 | 51.9 | 69.1 | 29.0 | 51.9 | 69.1 |
| Neural Haircut | 59.9 | 84.1 | 92.1 | 13.1 | 22.7 | 31.5 | 21.5 | 35.8 | 46.9 | 45.8 | 72.1 | 84.6 | 6.4 | 12.8 | 21.0 | 11.2 | 21.7 | 33.6 | 11.2 | 21.7 | 33.6 | 11.2 | 21.7 | 33.6 |
| Gaussian Haircut | 64.2 | 72.6 | 76.6 | 46.3 | 53.9 | 59.7 | 53.8 | 61.9 | 67.1 | 35.5 | 55.8 | 69.4 | 24.3 | 42.9 | 58.4 | 28.9 | 48.5 | 63.4 | 28.9 | 48.5 | 63.4 | 28.9 | 48.5 | 63.4 |

Table 1. Quantitative evaluation of different methods and ablations. For straight hair, our method outperforms both Neural Haircut [Sklyarova et al. 2023a] and Gaussian Haircut [Zakharov et al. 2024] in both precision and F-score metrics. For curly hair, our method achieves superior recall and F-score compared to both approaches, demonstrating better overall hair strand recovery.

dataset, which was compiled by applying our algorithm to a combination of the NPHM dataset and newly collected scans.

Acknowledgments. We gratefully acknowledge the support: (Artem Sevastopolsky) by a TUM-IAS Hans Fischer Senior Fellowship and the ERC Starting Grant Scan2CAD (804724); (Vanessa Sklyarova) the Max Planck ETH Center for Learning Systems; (Egor Zakharov) by the “AI-PERCEIVE” ERC Consolidator Grant, 2021. We also thank Alexey Artemov for helpful advice and Simon Giebenhain for help with the data acquisition, as well as the participants of the NPHM dataset and newly collected captures.

REFERENCES

- Hervé Abdi and Lynne J Williams. 2010. Principal component analysis. *Wiley interdisciplinary reviews: computational statistics* 2, 4 (2010), 433–459.
- Mohiuddin Ahmed, Raihan Seraj, and Syed Mohammed Shamsul Islam. 2020. The k-means algorithm: A comprehensive survey and performance evaluation. *Electronics* 9, 8 (2020), 1295.
- John F. Canny. 1986. A Computational Approach to Edge Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-8* (1986), 679–698. <https://api.semanticscholar.org/CorpusID:13284142>
- Blender Online Community. 2018. *Blender - a 3D modelling and rendering package*. Blender Foundation, Stichting Blender Foundation, Amsterdam. <http://www.blender.org>
- Qiancheng Fu, Qingshan Xu, Yew-Soon Ong, and Wenbing Tao. 2022. Geo-Neus: Geometry-Consistent Neural Implicit Surfaces Learning for Multi-view Reconstruction. *Advances in Neural Information Processing Systems (NeurIPS)* (2022).
- Guy Gafni, Justus Thies, Michael Zollhofer, and Matthias Nießner. 2020. Dynamic Neural Radiance Fields for Monocular 4D Facial Avatar Reconstruction. *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2020), 8645–8654.
- Simon Giebenhain, Tobias Kirschstein, Markos Georgopoulos, Martin Rünz, Lourdes Agapito, and Matthias Nießner. 2023. Learning neural parametric head models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 21003–21012.
- Philip-William Grassal, Malte Prinzler, Titus Leistner, Carsten Rother, Matthias Nießner, and Justus Thies. 2021. Neural Head Avatars from Monocular RGB Videos. *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2021), 18632–18643.
- Chengan He, Xin Sun, Zhixin Shu, Fujun Luan, Soren Pirk, Jorge Alejandro Amador Herrera, Dominik Ludewig Michels, Tuanfeng Y. Wang, Meng Zhang, Holly Rushmeier, and Yi Zhou. 2024. Perrn: A Parametric Representation for Multi-Style 3D Hair Modeling. *ArXiv abs/2407.19451* (2024). <https://api.semanticscholar.org/CorpusID:271534613>
- Liwen Hu, Chongyang Ma, Linjie Luo, and Hao Li. 2015. Single-view hair modeling using a hairstyle database. *ACM Transactions on Graphics (ToG)* 34, 4 (2015), 1–9.
- Tero Karras, Miika Aittala, Timo Aila, and Samuli Laine. 2022. Elucidating the Design Space of Diffusion-Based Generative Models. In *Advances in Neural Information Processing Systems (NeurIPS)*.
- Taras Khakhulin, V. V. Sklyarova, Victor S. Lempitsky, and Egor Zakharov. 2022. Realistic One-shot Mesh-based Head Avatars. In *European Conference on Computer Vision*.
- Zhiyi Kuang, Yiyang Chen, Hongbo Fu, Kun Zhou, and Youyi Zheng. 2022. DeepMVSHair: Deep Hair Modeling from Sparse Views. In *SIGGRAPH Asia 2022 Conference Papers* (Daegu, Republic of Korea) (SA ’22). Association for Computing Machinery, New York, NY, USA, Article 10, 8 pages. <https://doi.org/10.1145/3550469.3555385>
- Chunyuan Li, Cliff Wong, Sheng Zhang, Naoto Usuyama, Haotian Liu, Jianwei Yang, Tristan Naumann, Hoifung Poon, and Jianfeng Gao. 2023. Llava-med: Training a large language-and-vision assistant for biomedicine in one day. *Advances in Neural Information Processing Systems* 36 (2023), 28541–28564.
- Junnan Li, Dongxu Li, Caiming Xiong, and Steven Hoi. 2022. BLIP: Bootstrapping Language-Image Pre-training for Unified Vision-Language Understanding and Generation. In *ICML*.
- 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. <https://doi.org/10.1145/3130800.3130813>
- 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).
- Morgan McGuire, Tiancheng Sun, Giljoo Nam, Carlos Aliaga, Christophe Hery, and Ravi Ramamoorthi. 2021. Human Hair Inverse Rendering using Multi-View Photometric data. In *Eurographics Symposium on Rendering*.
- Giljoo Nam, Chenglei Wu, Min H. Kim, and Yaser Sheik. 2019. Strand-Accurate Multi-View Hair Capture. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2019), 155–164.
- Sylvain Paris, Héctor M. Briceño, and François X. Sillion. 2004. Capture of hair geometry from multiple images. *ACM Transactions on Graphics (TOG)* 23 (2004), 712 – 719. <https://api.semanticscholar.org/CorpusID:7900484>
- Keunhong Park, U. Sinha, Jonathan T. Barron, Sofien Bouaziz, Dan B. Goldman, Steven M. Seitz, and Ricardo Martin-Brualla. 2020. Nerfies: Deformable Neural Radiance Fields. *2021 IEEE/CVF International Conference on Computer Vision (ICCV)* (2020), 5845–5854.
- Radu Alexandru Rosu, Shunsuke Saito, Ziyan Wang, Chenglei Wu, Sven Behnke, and Giljoo Nam. 2022. Neural Strands: Learning Hair Geometry and Appearance from Multi-View Images. In *Computer Vision – ECCV 2018: 15th European Conference*.
- 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. 42, 4, Article 75 (July 2023), 13 pages. <https://doi.org/10.1145/3592106>
- Vanessa Sklyarova, Jenya Chelishev, Andreea Dogaru, Igor Medvedev, Victor Lempitsky, and Egor Zakharov. 2023a. Neural Haircut: Prior-Guided Strand-Based Hair Reconstruction. In *Proceedings of IEEE International Conference on Computer Vision (ICCV)*.
- Vanessa Sklyarova, Egor Zakharov, Otmar Hilliges, Michael J. Black, and Justus Thies. 2023b. Text-Conditioned Generative Model of 3D Strand-Based Human Hairstyles. *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2023), 4703–4712. <https://api.semanticscholar.org/CorpusID:266361928>

- Xavier Soria, Yachuan Li, Mohammad Rouhani, and Angel D Sappa. 2023. Tiny and efficient model for the edge detection generalization. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 1364–1373.
- Yusuke Takimoto, Hikari Takehara, Hiroyuki Sato, Zihao Zhu, and Bo Zheng. 2024. Dr.Hair: Reconstructing Scalp-Connected Hair Strands without Pre-Training via Differentiable Rendering of Line Segments. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 20601–20611.
- Laurens Van der Maaten and Geoffrey Hinton. 2008. Visualizing data using t-SNE. *Journal of machine learning research* 9, 11 (2008).
- Ziyan Wang, Giljoo Nam, Tuur Stuyck, Stephen Lombardi, Chen Cao, Jason M. Saragih, Michael Zollhoefer, Jessica K. Hodgins, and Christoph Lassner. 2022. NeuWigs: A Neural Dynamic Model for Volumetric Hair Capture and Animation. *ArXiv* abs/2212.00613 (2022).
- Ziyan Wang, Giljoo Nam, Tuur Stuyck, Stephen Lombardi, Michael Zollhoefer, Jessica K. Hodgins, and Christoph Lassner. 2021. HVH: Learning a Hybrid Neural Volumetric Representation for Dynamic Hair Performance Capture. *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2021), 6133–6144.
- Keyu Wu, Lingchen Yang, Zhiyi Kuang, Yao Feng, Xutao Han, Yuefan Shen, Hongbo Fu, Kun Zhou, and Youyi Zheng. 2024. MonoHair: High-Fidelity Hair Modeling from a Monocular Video. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.
- 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. *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2022), 1516–1525. <https://api.semanticscholar.org/CorpusID:248572051>
- Enze Xie, Wenhui Wang, Zhiding Yu, Anima Anandkumar, Jose M Alvarez, and Ping Luo. 2021. SegFormer: Simple and efficient design for semantic segmentation with transformers. *Advances in neural information processing systems* 34 (2021), 12077–12090.
- Shin Yoshizawa, Alexander Belyaev, and Hans-Peter Seidel. 2005. Fast and robust detection of crest lines on meshes. In *Proceedings of the 2005 ACM symposium on Solid and physical modeling*, 227–232.
- Cem Yuksel, Scott Schaefer, and John Keyser. 2009. Hair meshes. In *ACM SIGGRAPH Asia 2009 Papers* (Yokohama, Japan) (*SIGGRAPH Asia ’09*). Association for Computing Machinery, New York, NY, USA, Article 166, 7 pages. <https://doi.org/10.1145/1661412.1618512>
- Egor Zakharov, Vanessa Slyarova, Michael Black, Giljoo Nam, Justus Thies, and Otmar Hilliges. 2024. Human Hair Reconstruction with Strand-Aligned 3D Gaussians. In *Computer Vision – ECCV 2024: 18th European Conference, Milan, Italy, September 29–October 4, 2024, Proceedings, Part XVI* (Milan, Italy). Springer-Verlag, Berlin, Heidelberg, 409–425. https://doi.org/10.1007/978-3-031-72640-8_23
- Yufeng Zheng, Victoria Fernández Abrevaya, Xu Chen, Marcel C. Buhler, Michael J. Black, and Otmar Hilliges. 2021. I M Avatar: Implicit Morphable Head Avatars from Videos. *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2021), 13535–13545.
- 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. *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2023), 12726–12735. <https://api.semanticscholar.org/CorpusID:257365928>
- Yufeng Zheng, Yifan Wang, Gordon Wetzstein, Michael J. Black, and Otmar Hilliges. 2022. PointAvatar: Deformable Point-based Head Avatars from Videos. *ArXiv* abs/2212.08377 (2022).
- Junsheng Zhou, Baorui Ma, Shujuan Li, Yu-Shen Liu, and Zhizhong Han. 2023b. Learning a More Continuous Zero Level Set in Unsigned Distance Fields through Level Set Projection. In *Proceedings of the IEEE/CVF international conference on computer vision*.
- Yuxiao Zhou, Menglei Chai, Alessandro Pepe, Markus H. Gross, and Thabo Beeler. 2023a. GroomGen: A High-Quality Generative Hair Model Using Hierarchical Latent Representations. *ACM Transactions on Graphics (TOG)* 42 (2023), 1 – 16. <https://api.semanticscholar.org/CorpusID:265018898>
- Yuxiao Zhou, Menglei Chai, Daoye Wang, Sebastian Winberg, Erroll Wood, Kripasindhu Sarkar, Markus Gross, and Thabo Beeler. 2024. GroomCap: High-Fidelity Prior-Free Hair Capture. *ACM Trans. Graph.* 43, 6, Article 254 (Nov. 2024), 15 pages. <https://doi.org/10.1145/3687768>
- Yi Zhou, Liwen Hu, Jun Xing, Weikai Chen, Han-Wei Kung, Xin Tong, and Hao Li. 2018. Single-View Hair Reconstruction using Convolutional Neural Networks. In *European Conference on Computer Vision*. <https://api.semanticscholar.org/CorpusID:49666680>

A STRANDS400 DATASET

Legal notice. All participants in our dataset signed an agreement form compliant with GDPR. Please note that GDPR compliance includes the right for every participant to request the timely deletion of their data, which we will enforce as part of the distribution process of our dataset.

Dataset statistics. In Fig. 6, we demonstrate the distribution of the age, stratified by gender, (left) and of ethnicity (right), reported by the participants in the Strands400 dataset. In the age histogram (left), the horizontal axis corresponds to the participants’ age bins, and the vertical axis corresponds to the number of people in the bin. Only the votes of the participants who willingly disclosed that information were taken into account.

Additionally, we analyze the content of the answers provided by the employed VQA model (LLaVA [Li et al. 2023]) after showing the rendered shading of the frontal and back views of the 3D scans in Strands400. We used the same BLIP embedder [Li et al. 2022] as in training to obtain the embeddings from the answers to all 27 questions asked in training and fit t-SNE [Van der Maaten and Hinton 2008] to them to visualize them on a 2D plane. Apart from that, to obtain the coloring of the t-SNE points, we run K-Means [Ahmed et al. 2020] over 5 clusters. The t-SNE locations are enhanced with LLaVA answers for the respective samples. We visualize the results in Fig. 9 and Fig. 10. Each second sample has been shown in the plots to provide more space for the captions.

Capture setup. Our dataset consists of two parts – the subset of the latest version of NPHM that contains 383 scans, and separately collected 17 scans with a setup similar to the NPHM setup. The latter setup consists of two handheld Artec Eva scanners, rotating over a 360° trajectory within ~ 3 seconds to capture a single person. Since our setup largely follows NPHM capture setup, we refer the reader to the NPHM paper for the remaining details regarding the capture setup [Giebenhain et al. 2023]. Additionally, for each of the 20 scans in the second part of the dataset, we have collected a 360° RGB video with a smartphone, around 30-seconds long and in 4K resolution, to be able to compare our method with the baselines. These participants in this category were selected with an emphasis on hairstyles, more challenging for reconstruction (wavy, curly, etc.), to better align the overall distribution to the overall spectrum of hairstyles.

Representative samples. More samples from the Strands400 dataset are demonstrated in Fig. 7 and in Fig. 8.

B TECHNICAL DETAILS

Questions employed in the calculation of the text prompt for the diffusion prior conditioning. We provide the list of questions for querying LLaVA model below.

- (1) *Describe in detail the bang/fringe of depicted hairstyle including its directionality, texture and coverage of face?*
- (2) *What is the overall hairstyle depicted in the image?*

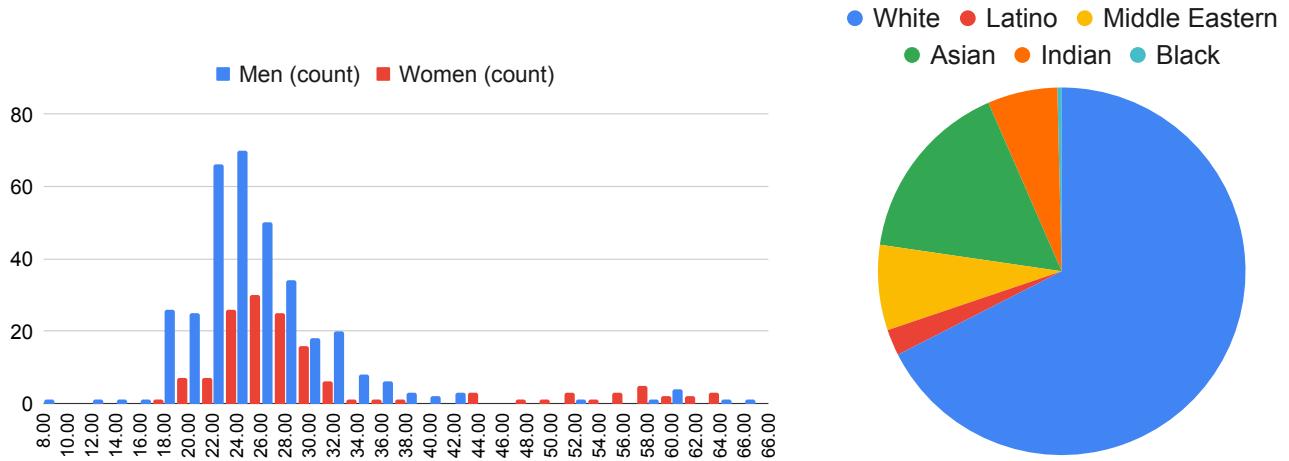


Fig. 6. The distribution of the age, stratified by gender, (left) and of ethnicity (right), reported by the participants in the Strands400 dataset. Only the votes of the participants who willingly disclosed that information were taken into account.

- (3) Does the depicted hairstyle longer than the shoulders or shorter than the shoulder?
- (4) Does the depicted hairstyle has short bang or long bang or no bang from frontal view?
- (5) Does the hairstyle has straight bang or Baby Bangs or Arched Bangs or Asymmetrical Bangs or Pin-Up Bangs or Choppy Bangs or curtain bang or side swept bang or no bang?
- (6) Are there any afro features in the hairstyle or no afro features?
- (7) Is the length of hairstyle shorter than middle of the neck or longer than middle of the neck?
- (8) What is the main geometry features of the depicted hairstyle?
- (9) What is the overall shape of the depicted hairstyle?
- (10) Is the hair short, medium, or long in terms of length?
- (11) What is the type of depicted hairstyle
- (12) What is the length of hairstyle relative to human body?
- (13) Describe the texture and pattern of hair in the image.
- (14) What is the texture of depicted hairstyle
- (15) Does the depicted hairstyle is straight or wavy or curly or kinky?
- (16) Can you describe the overall flow and directionality of strands?
- (17) Could you describe the bang of depicted hairstyle including its directionality and texture
- (18) Describe the main geometric features of the hairstyle depicted in the image
- (19) Is the length of hairstyle buzz cut, pixie, ear length, chin length, neck length, shoulder length, armpit length or mid-back length?
- (20) Describe actors with similar hairstyle type.
- (21) Does the hairstyle cover any parts of the face? Write which exactly parts.
- (22) In what ways is this hairstyle a blend or combination of other popular hairstyles?
- (23) Could you provide the most closest types of hairstyles from which this one could be blended?
- (24) How adaptable is this hairstyle for various occasions (casual, formal, athletic)?

- (25) How is this hairstyle perceived in different social or professional settings?
- (26) Are there historical figures who were iconic for wearing this hairstyle?
- (27) Could you describe the partition of this hairstyle if it is visible?

Diffusion prior. We use the Elucidating Diffusion Moldel (EDM) [Karras et al. 2022] for the denoiser \mathcal{D} . We apply HAAR [Sklyarova et al. 2023b] diffusion prior on low-resolution samples from T and define them as T_{LR} or y to have the same notation as in [Karras et al. 2022]. At each iteration we obtain a noised input: $x = y + \epsilon \cdot \sigma$, where $\epsilon \sim \mathcal{N}(0, I)$, and σ is a noise strength. We then predict a denoised input:

$$\mathcal{D}(x, \sigma) = c_{\text{skip}}(\sigma) \cdot x + c_{\text{out}}(\sigma) \cdot F(c_{\text{in}}(\sigma) \cdot x, c_{\text{noise}}(\sigma)), \quad (14)$$

where the c_{skip} , c_{out} , c_{in} and c_{noise} are part of pre-conditioning approach proposed in [Karras et al. 2022], which improves the robustness of D to the low noise strength σ , and F is a neural network. Our training objective also follows [Karras et al. 2022]:

$$\mathcal{L}_{\text{diff}} = \mathbb{E}_{y, \sigma, \epsilon} \left[\lambda_{\text{diff}}(\sigma) \cdot \| \mathcal{D}(x, \sigma) - y \|_2^2 \right], \quad (15)$$

where $\lambda_{\text{diff}}(\sigma)$ is a weighting function, and the expectation is approximated via sampling. Different from Neural Haircut [Sklyarova et al. 2023a] instead of sampling noise randomly, we first decrease it with time from max to min level and then start sampling randomly. Also, we use a conditional diffusion model HAAR [Sklyarova et al. 2023b]. To increase the strength of the prior, we propose to do several denoising steps.



Fig. 7. Sample strands reconstructions from Strands400 dataset.

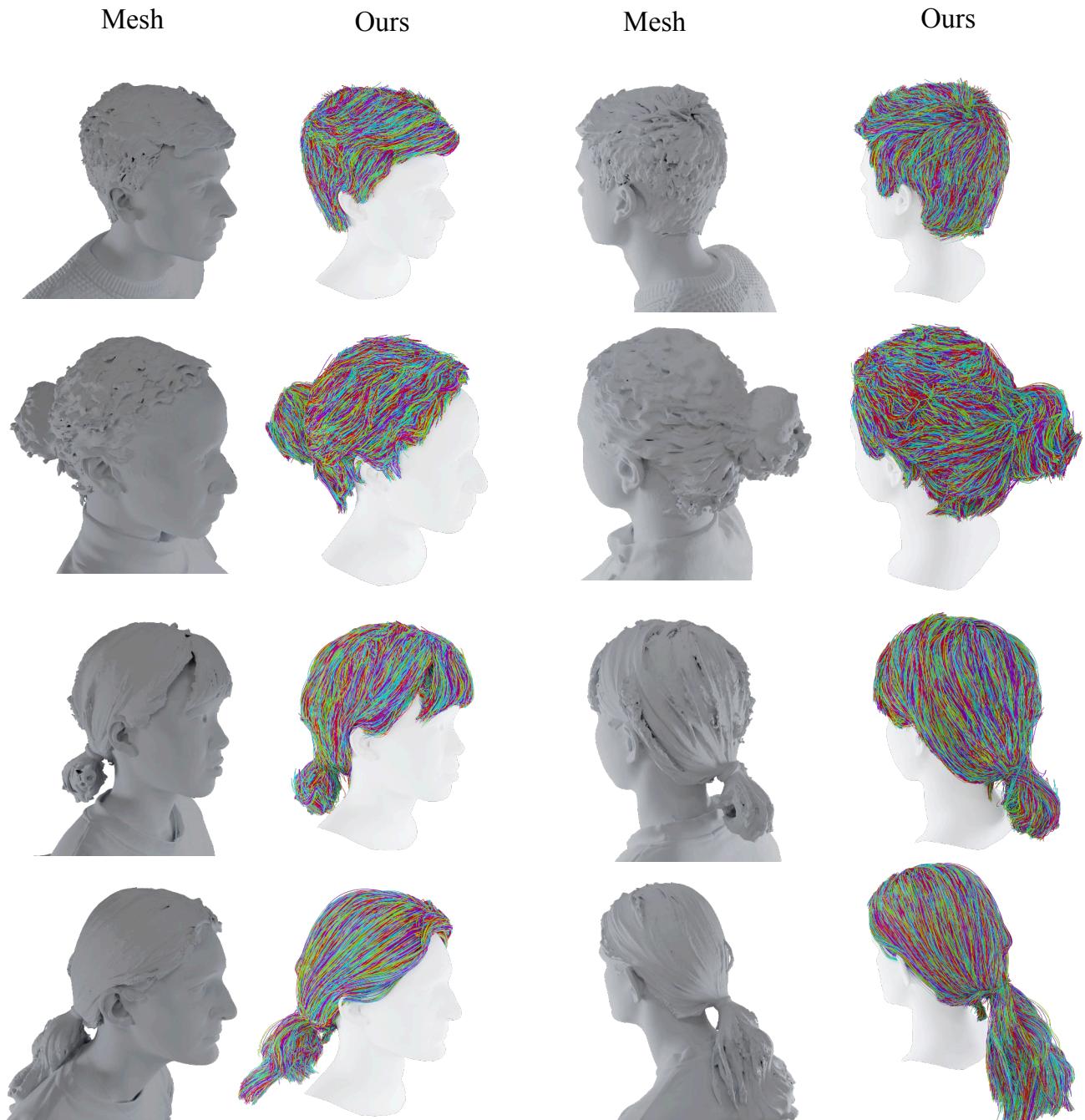


Fig. 8. Sample scans from Strands400 dataset and the corresponding reconstructions.

Strands-400: t-SNE of 768D Embeddings.
Captions are the LLaVA answers to:
What is the type of depicted hairstyle?



Fig. 9. The distribution of hair length in the Strands400 dataset. The captions are collected from the answers of a VQA model (LLaVA [Li et al. 2023]) after showing the rendered shading of the frontal and back views of the 3D scans in Strands400. The locations correspond to the t-SNE [Van der Maaten and Hinton 2008] over BLIP embeddings [Li et al. 2022] of the LLaVA answers. The colors are calculated via K-Means [Ahmed et al. 2020].

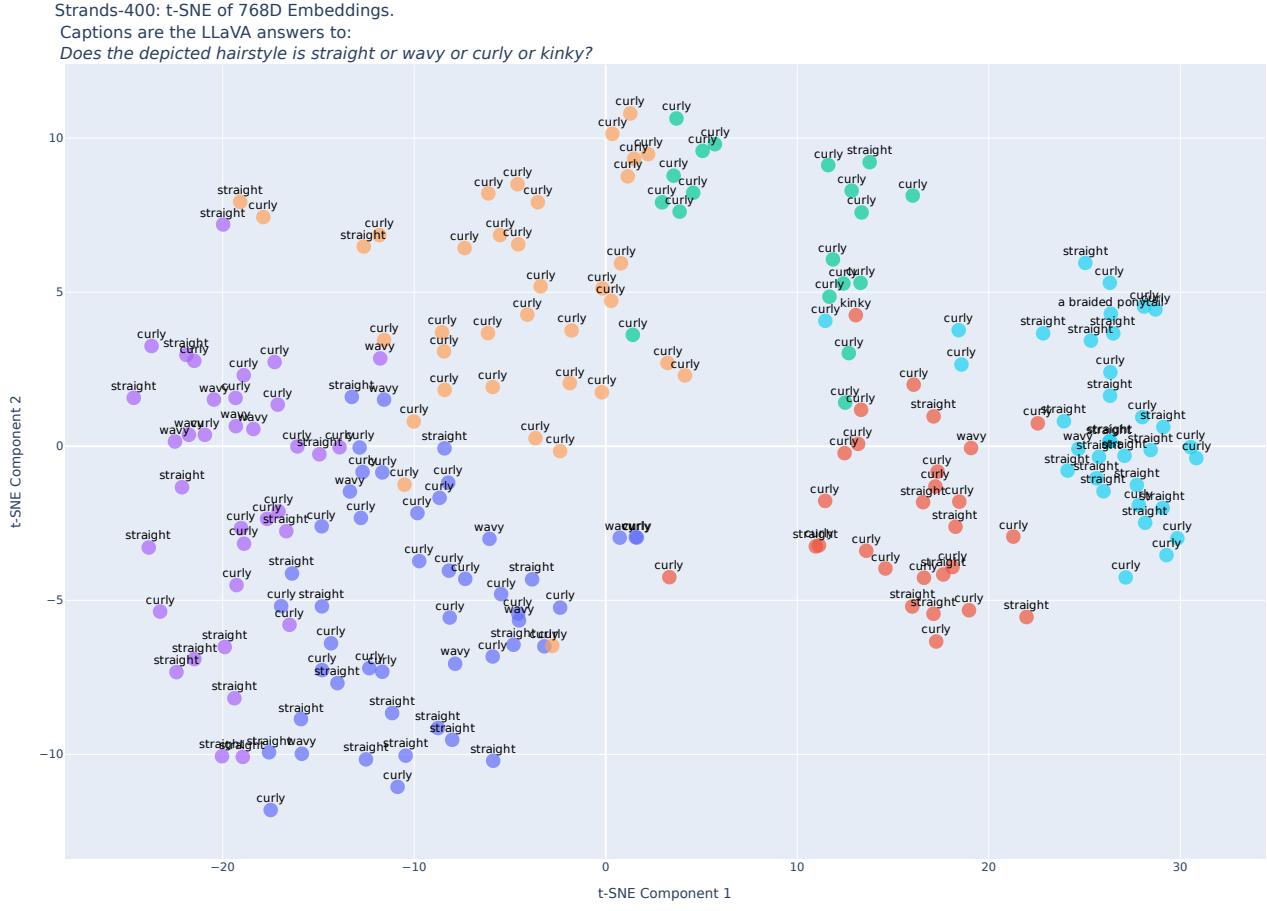


Fig. 10. The distribution of hair waviness in the Strands400 dataset. The captions are collected from the answers of a VQA model (LLaVA [Li et al. 2023]) after showing the rendered shading of the frontal and back views of the 3D scans in Strands400. The locations correspond to the t-SNE [Van der Maaten and Hinton 2008] over BLIP embeddings [Li et al. 2022] of the LLaVA answers. The colors are calculated via K-Means [Ahmed et al. 2020].