



# EMS: 3D Eyebrow Modeling from Single-view Images

CHENGHONG LI\*, FNii, CUHKSZ, China and SSE, CUHKSZ, China

LEYANG JIN\*, SSE, CUHKSZ, China

YUJIAN ZHENG, FNii, CUHKSZ, China and SSE, CUHKSZ, China

YIZHOU YU†, The University of Hong Kong, Hong Kong

XIAOGUANG HAN†, SSE, CUHKSZ, China and FNii, CUHKSZ, China

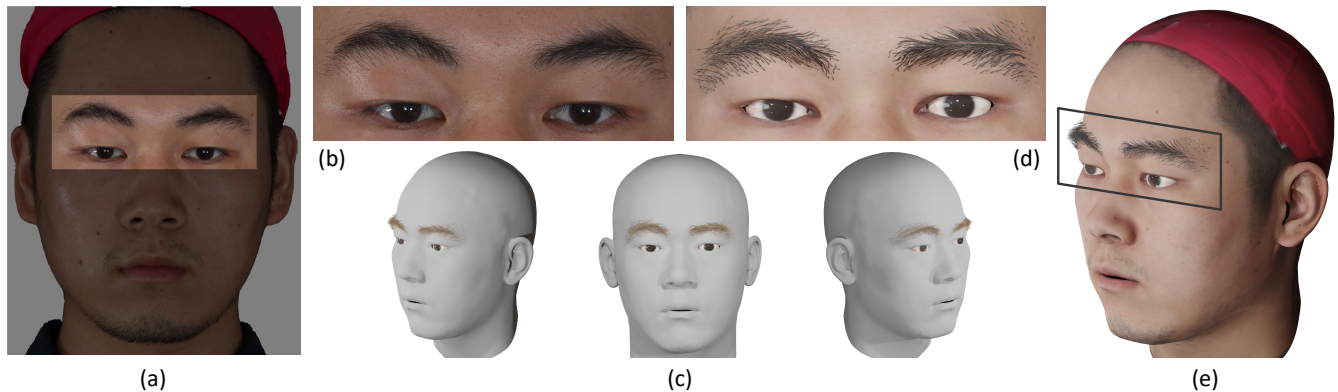


Fig. 1. Nowadays, 3D heads can be easily digitalized with high-fidelity geometry and texture, like the setup used in FaceScape. In this work, we proposed EMS, a learning-based approach, which can further reconstruct 3D eyebrows from just the frontal-view image. (a) the input image. (b) the cropped eyebrow image from (a). (c) our reconstructed fiber-level 3D eyebrow model rendered with multiple views. (d) the cropped rendering of (c) for comparison with (b). (e) putting our result on the textured 3D head can further improve the realism of face digitalization.

Eyebrows play a critical role in facial expression and appearance. Although the 3D digitization of faces is well explored, less attention has been drawn to 3D eyebrow modeling. In this work, we propose EMS, the first learning-based framework for single-view 3D eyebrow reconstruction. Following the methods of scalp hair reconstruction, we also represent the eyebrow as a set of fiber curves and convert the reconstruction to fibers growing problem. Three modules are then carefully designed: *RootFinder* firstly localizes the fiber root positions which indicate where to grow; *OriPredictor* predicts an orientation field in the 3D space to guide the growing of fibers; *FiberEnder* is designed to determine when to stop the growth of each fiber. Our *OriPredictor* directly borrows the method used in hair reconstruction. Considering the differences between hair and eyebrows, both *RootFinder* and *FiberEnder*

are newly proposed. Specifically, to cope with the challenge that the root location is severely occluded, we formulate root localization as a density map estimation task. Given the predicted density map, a density-based clustering method is further used for finding the roots. For each fiber, the growth starts from the root point and moves step by step until the ending, where each step is defined as an oriented line segment with a constant length according to the predicted orientation field. To determine when to end, a pixel-aligned RNN architecture is designed to form a binary classifier, which outputs stop or not for each growing step. To support the training of all proposed networks, we build the first 3D synthetic eyebrow dataset that contains 400 high-quality eyebrow models manually created by artists. Extensive experiments have demonstrated the effectiveness of the proposed EMS pipeline on a variety of different eyebrow styles and lengths, ranging from short and sparse to long bushy eyebrows.

\*Both authors contributed equally to this work.

†Corresponding author

Authors' addresses: Chenghong Li, chenghongli@link.cuhk.edu.cn, FNii, CUHKSZ, Shenzhen, China and SSE, CUHKSZ, Shenzhen, China; Leyang Jin, leyangjin1@link.cuhk.edu.cn, SSE, CUHKSZ, Shenzhen, China; Yujian Zheng, yujianzheng@link.cuhk.edu.cn, FNii, CUHKSZ, Shenzhen, China and SSE, CUHKSZ, Shenzhen, China; Yizhou Yu, yizhouy@acm.org, The University of Hong Kong, Hong Kong; Xiaoguang Han, hanxiaoguang@cuhk.edu.cn, SSE, CUHKSZ, Shenzhen, China and FNii, CUHKSZ, Shenzhen, China.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2023 Copyright held by the owner/author(s). Publication rights licensed to ACM. 0730-0301/2023/12-ART269 \$15.00 <https://doi.org/10.1145/3618323>

CCS Concepts: • Computing methodologies → Shape modeling; Neural networks.

Additional Key Words and Phrases: eyebrow, deep neural networks, single-view modeling, dataset

## ACM Reference Format:

Chenghong Li, Leyang Jin, Yujian Zheng, Yizhou Yu, and Xiaoguang Han. 2023. EMS: 3D Eyebrow Modeling from Single-view Images. *ACM Trans. Graph.* 42, 6, Article 269 (December 2023), 19 pages. <https://doi.org/10.1145/3618323>

## 1 INTRODUCTION

Recently, creating a digital human has become easier along with the development of capturing setups and also deep-learning based approaches. In addition to numerous studies on the digitization of

the head and human body [Bao et al. 2021; Deng et al. 2019; Feng et al. 2021; Li et al. 2017; Loper et al. 2015; Saito et al. 2020; Xiu et al. 2022; Zheng et al. 2021; Zielonka et al. 2022], there are also many works that pay attention to the details of the facial features due to their vital roles in supporting a high-fidelity face, for example, the modeling of eyes [Bérard et al. 2016, 2014; Bermano et al. 2015; Li et al. 2022; Wen et al. 2017], mouth [Dinev et al. 2018; Garrido et al. 2016; Ploumpis et al. 2022; Zoss et al. 2019, 2018] and teeth [Velinov et al. 2018; Wu et al. 2016; Yang et al. 2020a; Zhang et al. 2022]. Compared with them, facial hair (including eyebrows, eyelashes, and beards) has not received sufficient attention. Facial hair, particularly eyebrows, also plays a vital role in representing unique personal characteristics. According to [Sadr et al. 2003], eyebrows represent a significant influential attribute for facial recognition. To take high-fidelity human digitization one step further, in this work, we focus on the 3D reconstruction of eyebrows.

To the best of our knowledge, no prior work has been devoted to algorithms specifically tailored for 3D eyebrow modeling. Existing facial hair modeling methods [Beeler et al. 2012; Winberg et al. 2022] primarily focus on recovering the overall facial hair. They adopted multi-view stereo (MVS) [Seitz et al. 2006] systems customized to reconstruct 3D facial hair fibers from the appearance cues of multi-view images. These methods rely on specialized hardware and well-controlled settings, making them not widely accessible to the average consumer. To ease the capturing procedure and improve user-friendliness, several works [Herrera et al. 2010; Rotger et al. 2019] have been proposed to attempt conducting reconstruction from monocular inputs. However, as the first attempt, their methods are too straightforward and crude, making the results far from natural-looking. Specifically, rule-based methods neglect accurate root point localization and fiber growing direction determination. We argue that the formidable challenges of severe self-occlusion and single-view ambiguity greatly hinder the success of rule-based approaches. Thus, in this work, we propose to use data-driven techniques for single-view 3D eyebrow reconstruction.

Leveraging data-driven methods is a popular solution frequently employed in the problem of single-view scalp hair reconstruction [Wu et al. 2022; Zheng et al. 2023; Zhou et al. 2018]. The incorporation of learned priors from the data can provide information for occluded regions and recover more plausible results. Existing works usually use 3D strands to represent hairs and then the problem becomes the generation of those strand curves. Regarding the extreme complexity of generating a heavy amount of strands, the key idea to simplify this problem is to grow curves from the scalp surface. To be specific, for each strand, they usually considered three questions: where to start; how to grow; when to stop. As all root points of hairs cannot be observed typically, the commonly used way is to pre-sample a fixed number of points (usually 10k) densely on the scalp surface as starting positions. Then, each root point will initiate the growth of a fiber and each fiber will grow step by step, where each step is defined by an oriented line segment with a pre-defined constant length. To guide such growth, an orientation field is required, and inferring an orientation field from the input image thus becomes one of the core problems. Another important issue is determining when to stop the growth of fibers. Existing methods directly used

the 3D contour of the hairs for strand cutting, where they typically predicted a 3D occupancy field to define the scalp hair contour mesh.

For 3D eyebrows, we use fiber (similar in concept to strand with shorter length) as the output format and treat it as a curve growing problem. For the prediction of the orientation field, we directly adopt the method in [Wu et al. 2022]. However, we found it is not proper to use existing methods of root localization and strand cutting in hair reconstruction for eyebrow modeling. This is due to the inherent differences between scalp hairs and eyebrows. At first, although the root points of eyebrows may be severely occluded, they can still be partially observed. Thus, an intuitive uniform sampling will produce results that do not match the input image. Second, considering a frontal-view facial image, the fibers of the eyebrow may not only have out-of-plane growth but also in-the-plane growing manner. For in-the-plane growing directions, a 3D contour is usually not sufficient for accurate fiber length control.

To address the aforementioned challenges, we proposed two novel learning-based methods for root localization and fiber length determination individually. Due to the severe occlusion of the root points, a direct detection solution for root localization tends to fail. We thus introduce the density map as an intermediate representation of the roots. Specifically, we first construct an image-to-image translation network to map the input image into a density map, where each pixel of the density map represents the response of a root existence. Then, a density-based clustering method is employed to extract root locations from the predicted density map. Next, we lift the 2D root points to 3D positions. The above density map prediction network combined with the clustering and lifting method is termed as *RootFinder*. After all root points are confirmed, each root will grow a fiber, which will then grow step by step. Similar to the method in hair reconstruction, each growing step follows an orientation field which is predicted based on a pixel-aligned implicit learning framework [Wu et al. 2022]. The prediction method of the orientation field is termed as *OriPredictor*. For each growing fiber, we observed that determining when to stop depends not only on its location but also on the image appearances along its growth trajectory. Based on this assumption, we design a pixel-aligned stacked RNN [Giles et al. 1994] architecture to form a classifier that outputs a binary status for each growing step: stop or continue. To be specific, the network is conditioned on both a positional encoding of the sequential 3D points and an accumulated pixel-aligned image feature along the historical growing path. We term this as *FiberEnder*.

To summarize, we introduce *EMS* in this work, which is the first deep learning-based framework to reconstruct 3D eyebrow models from a single portrait image. Our *EMS* consists of three modules: *RootFinder*, *OriPredictor* and *FiberEnder*. To support both the training and evaluation of all three modules, we also contribute a dataset with well-annotated synthetic fiber-level eyebrows of 400 people in FaceScape [Yang et al. 2020b], which are manually created by artists. Extensive experiments have demonstrated the superiority of our method over existing rule-based methods and the effectiveness of all our module designs. The results of in-the-wild testing also verify the generalization ability of our method, even when it is trained using synthetic data only.

The main contributions of this work are listed as follows:

- We propose the first learning-based framework for 3D eyebrow modeling, capable of reconstructing fiber-level eyebrow geometry from only a single portrait image. Our method achieves state-of-the-art performance.
- Considering the differences among eyebrows and hairs, we propose a new method for root point localization which leverages a density map as an intermediate representation.
- We are the first to formulate fiber length determination of eyebrows as a classification problem and design a novel pixel-aligned RNN architecture for solving it.
- We contribute the first high-quality 3D synthetic eyebrow dataset *EBStore*, which will be released to benefit the research in this area.

## 2 RELATED WORK

This section provides a comprehensive review of existing works closely related to our work.

*Scalp Hair Capturing.* Scalp hair modeling has been extensively explored in computer graphics over the past few decades. Grabli et al. [2002] utilized hair scattering profiles to recover 3D hair orientation from image sequences captured under moving light sources. Follow-up works [Luo et al. 2012; Paris et al. 2004, 2008; Wei et al. 2005] estimate a dense vector field from different views or under different illuminations with 3D geometry constraints, e.g., visual hull, structure light, or epipolar constraints. To improve the hair capturing quality, various sensor modalities have also been explored, such as depth-of-focus devices [Jakob et al. 2009], thermal imaging cameras [Herrera et al. 2012], or RGB-D sensors [Hu et al. 2014b]. Luo et al. [2013] and Hu et al. [2014a] used shape primitives like ribbons and wisps to fit a fused point cloud generated by MVS for the acquisition of complete hair strands that connect to the scalp. Additionally, Xu et al. [2014] traced a motion path for each hair pixel for dynamic hair temporal coherence. Nam et al. [2019] introduced a slanted line-based MVS framework with a novel cost function. Sun et al. [2021] leveraged a per-pixel lightcode to boost the multi-view consistency of 3D hair segments and estimate hair reflectance for realistic rendering. Rosu et al. [2022] attempted to recover hair strand geometry and appearance by differentiable rendering with multi-view inputs for the first time and explored a novel representation that parameterizes a hairstyle using a latent texture map on the scalp. To simplify the data acquisition, several works investigate modeling hair from sparse views [Zhang et al. 2017, 2018] and selfie videos [Liang et al. 2018]. Recently, Kuang et al. [2022] introduced a deep learning-based framework for multi-view strand hair modeling that enables the reconstruction of view-consistent hair geometry using only sparse views. However, the complex hardware setup or long processing cycles make these techniques inaccessible to the average user.

*Facial Hair Capturing.* Compared to the extensive research on digital modeling of human scalp hair, facial hair capturing has been a relatively neglected research area. Pioneer work by [Herrera et al.

2010] attempts to model static facial hair using statistical characteristics of texture images, growing 3D hairs through particle shooting techniques, and modeling facial hair geometry roughly. To generate connected piece-wise 3D facial hair primitives on a static face, Beeler et al. [2012] proposed a passive, multi-camera system with uniform lighting, which incorporates 2D hair detection and 3D facial hair growing algorithms. While capturing individual fibers of facial hair is available, their approach involves considerable post-processing and refinement steps to generate plausible hair segments. Furthermore, their system also requires additional close-up photographs of the facial hair regions. Contemporary work by [Fyffe 2012] introduces a novel photo-consistency cost function to help identify areas containing facial hair and triangulate 2D line segments to a set of 3D oriented hair particles from multiple views. Additionally, LeGendre et al. [2017] employed asperity scattering and hair fiber geometric parameter fitting techniques along vellus hairs' backlit silhouette to model barely noticeable and fine texture vellus hair. Inspired by [Beeler et al. 2012], Rotger et al. [2019] extended the task formulation with just a single-view image as input, but with limited fidelity and realism due to the representation of 3D facial hair fibers with only a simple hair growing parametric model. Recently, Winberg et al. [2022] improved upon [Beeler et al. 2012] to integrate dense facial hair tracking and underlying skin deformation capabilities. Although their method obtains impressive results in capturing beard hair from dynamic sequences, it struggles to capture eyebrow geometry with high-frequency details. The primary cause of this limitation is the inherent self-occlusion problem in eyebrows, which is exacerbated for thicker, bushier eyebrows. Consequently, the precision of identifying hair follicle locations and hair fiber matching is diminished even with the input of multiple front-view images. These factors greatly affect the fidelity and realism of the reconstructed eyebrows and highlight the need for more specialized techniques to address the eyebrow modeling complexities.

*Single-View Scalp Hair Modeling.* In contrast to multi-view scalp hair reconstruction techniques, single-view hair modeling methods [Chai et al. 2013, 2012] demonstrate notable advantages in generality and efficiency. Later works by [Chai et al. 2016; Hu et al. 2015] provide synthetic 3D hair databases and generate satisfactory results from a single portrait image based on data-driven techniques. However, the fidelity of their hair reconstruction results is highly dependent on the quality and diversity of the hair dataset. In particular, the retrieved hair model is prone to fail if a 3D hair model with an identifiable likeness is not available in the database. Recently, the significant advancements achieved in deep neural networks across various fields have prompted the adoption of learning-based algorithms for scalp hair modeling. Early work by [Zhou et al. 2018] exploits an encoder-decoder network to infer 3D hair strands directly from a 2D orientation map of the segmented hair region. Saito et al. [2018] trained a variational autoencoder to represent 3D hairstyles as a latent space of hair volume and utilized the latent code to synthesize 3D hair from a single-view image. Later work by [Zhang and Zheng 2019] introduces a generative adversarial network with 3D convolutional layers to recover the 3D orientation field of hair strands. Furthermore, Yang et al. [2019] investigated

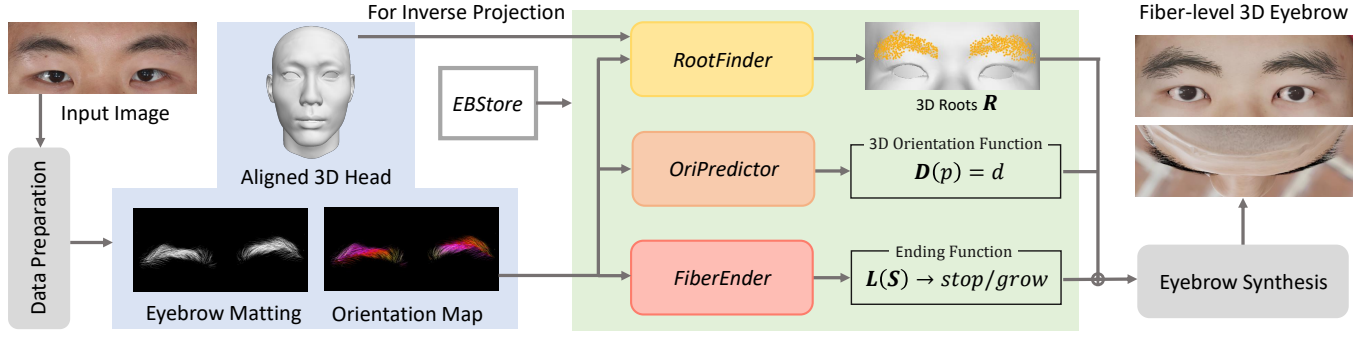


Fig. 2. The pipeline of eyebrow modeling from single-view images. Given a single-view image, our method reconstructs a fiber-level 3D eyebrow. We first prepare the intermediate data, then feed them into *RootFinder*, *OriPredictor* and *FiberEnder* to generate 3D eyebrow roots  $\mathbf{R}$ , 3D orientation function  $\mathbf{D}$  and ending function  $\mathbf{L}$ . Following the eyebrow synthesis, we finally obtain the fiber-level eyebrow model.

dynamic hair modeling from monocular video streams and inferred 3D spatial and temporal features of moving hairs via neural networks. While the aforementioned methods are capable of capturing the overall shape of hairs and predominant strand orientation, their explicit hair representation (e.g. orientation voxels) tends to reconstruct over-smoothed 3D hair geometry. To overcome this limitation, concurrent works by [Wu et al. 2022; Zheng et al. 2023] leverage implicit representation and pixel-aligned features to represent intricate hair geometry. Despite a large body of research devoted to learning-based scalp hair modeling, the application of these methods to eyebrow reconstruction is not adequate due to the notable differences in the root distributions and growing rules between scalp hair and eyebrows.

### 3 OVERVIEW

This section contains a system overview for the single-view 3D reconstruction of a fiber-level eyebrow model. We first give the definition of this problem, then briefly introduce our pipeline.

#### 3.1 Problem Definition

Given a single input image  $\mathbf{I}$ , we need to recover the corresponding fiber-level 3D eyebrow model  $\mathbf{E}$  growing on the target 3D head  $\mathbf{H}$ .

*3D eyebrow representation.* A fiber-level 3D eyebrow model  $\mathbf{E}$  is actually a set of 3D curves growing from the roots to the tips. Following state-of-the-art methods of scalp hair modeling [Wu et al. 2022; Zheng et al. 2023], we model  $\mathbf{E}$  via an implicit orientation field. Once we obtain the eyebrow roots and lengths, fiber curves can be easily grown step-by-step from this field. Thus, we define the fiber-level 3D eyebrow model as

$$\mathbf{E} = \text{Syn}(\mathbf{R}, \mathbf{D}, \mathbf{L}), \quad (1)$$

where  $\text{Syn}(\cdot)$  stands for the algorithm of eyebrow synthesis, i.e. the growing procedure.  $\mathbf{R}$ ,  $\mathbf{D}$  and  $\mathbf{L}$  represent the eyebrow roots, implicit 3D orientation function and the ending function, respectively. Specifically, the eyebrow roots  $\mathbf{R} = \{r_i, 0 \leq i < n\}$  is a 3D point cloud attached to the surface of the brow bone region on the target head  $\mathbf{H}$ , with a dimension of  $n \times 3$ . The implicit orientation function  $\mathbf{D}$  maps a 3D query point  $p$  to a unit growing vector  $d$ , which can be

formulated as  $d = \mathbf{D}(p)$ . The ending function  $\mathbf{L}$  describes whether a growing eyebrow fiber  $\mathbf{S}$  should be stopped:

$$\mathbf{L}(\mathbf{S}) = \begin{cases} 0, & \text{if } \mathbf{S} \text{ should be stopped} \\ 1, & \text{otherwise} \end{cases}. \quad (2)$$

#### 3.2 Pipeline

Fig. 2 shows the pipeline of our learning-based framework (*EMS*) for single-view eyebrow modeling. Our method takes a single portrait image as input and outputs a fiber-level 3D eyebrow model with fine details. Given the lack of a 3D eyebrow dataset, we create a high-quality synthetic 3D eyebrow dataset named *EBStore* (Sec. 4), to facilitate the data-driven approach. Based on *EBStore*, *EMS* achieves state-of-the-art performance on single-view 3D eyebrow modeling.

Starting from a cropped portrait image of the eyebrow region, we first prepare the intermediate data, including the 3D head, eyebrow matting, and the orientation map (Sec. 5.1). Following the definition in Eq. (1), we break down 3D eyebrow modeling into three steps, i.e. localizing fiber root (Sec. 5.2), predicting 3D orientation (Sec. 5.3) and determining fiber length (Sec. 5.4). After the eyebrow synthesis (Sec. 5.5), we finally obtain a fiber-level 3D eyebrow model. It should be noted that, as there is a non-negligible domain gap between real images and synthetic images, we follow [Kuang et al. 2022; Wu et al. 2022; Zhou et al. 2018] to adopt the 2D orientation map as the intermediate representation rather than input the image directly to the sub-modules.

### 4 DATASET CONSTRUCTION

Unlike the existing facial hair datasets [Dataset 2022; Wang et al. 2022; Xiao et al. 2021] with only 2D detection and alpha matting annotations, we construct the first 3D fiber-level synthetic eyebrow dataset spanning a wide range of eyebrow geometries and different local details based on FaceScape [Yang et al. 2020b]. FaceScape obtained data from a dense multi-view stereo with DSLR cameras, which provides high-quality 3D face models. These face models are registered to the template head mesh with uniform topology and located within a unified space, which facilitates the rough alignment of 3D eyebrows attached to various brow bone geometries. Besides, 68 DSLR cameras of the multi-view system are able to capture 4K



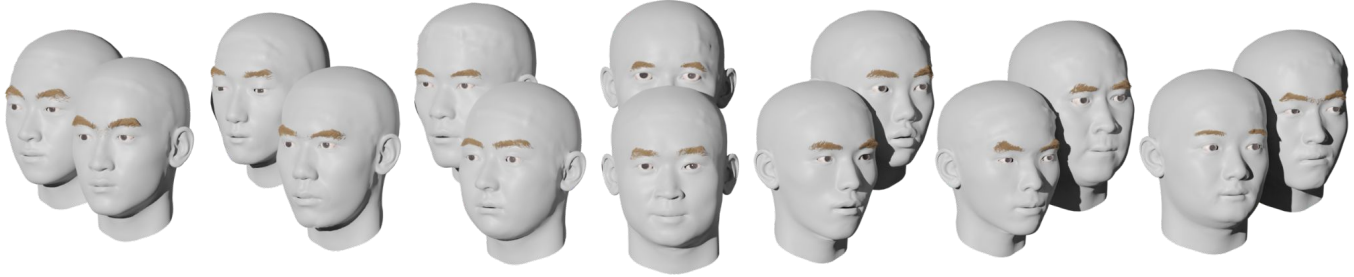


Fig. 3. *EBStore* dataset gallery with diverse 3D eyebrow models.

images, ensuring clear visibility of the eyebrows in frontal views. We carefully selected 197 males and 203 females from the large-scale FaceScape dataset to ensure gender balance in building the *EBStore* dataset. The subjects in *EBStore* range from 16 to 70 years old with diverse eyebrow styles. These 400 3D eyebrow models are manually created by artists using particle edit mode in Blender [Community 2018], guided by both the high-resolution texture map of the eyebrow region and a front-facing image of the individual. In general, a skillful artist typically needs to take around 1-2 days to create a high-quality fiber-level eyebrow model that closely matches the given image. Each eyebrow fiber in *EBStore* is comprised of 20 connected 3D points for further processing. Fig. 3 shows a gallery of our data.

## 5 METHODOLOGY

In this section, we describe the detailed designs of our approach.

### 5.1 Preparation

The initial stage of the preparation for our system is to detect eyebrow regions from single-view portrait images. We use the 3DDFA method [Guo et al. 2020] to detect the face region of the input image and then feed the face image to the pre-trained face parsing model on CelebAMask-HQ [Lee et al. 2020]. Then we crop the eyebrow region from the face parsing results and adopt the eyebrow matting method [Wang et al. 2022] to get the mask of eyebrow fibers. Next, we compute the real orientation map  $\mathbf{O}$  for the eyebrow fiber map with the mask using the method of [Luo et al. 2012], which uses a bank of rotated filters to detect the dominant orientation at each pixel. The orientation map  $\mathbf{O}$  is then enhanced with 3 passes of iterative refinement for a better signal-to-noise ratio proposed by [Chai et al. 2012]. As for acquiring the human head  $\mathbf{H}$ , we get a detailed 3D human face from a single-view 3D face prediction method [Yang et al. 2020b]. By utilizing the mapping between the 2D image and 3D face mesh, we can derive a pseudo-camera pose for eyebrow projection. However, it should be noted that the predicted 3D face mesh may not be aligned with our training head space. To solve this, we align the predicted face mesh with the average head of FaceScape by optimizing its translation, rotation and scale.

### 5.2 Localization of Fiber Root

Given the eyebrow orientation map  $\mathbf{O}$  as input, we aim to find the 3D roots  $\mathbf{R}$  of the eyebrow fibers. As previously discussed, directly detecting eyebrow roots is a tough challenge due to severe self-occlusion from visual perception. To address this issue, we proposed a density-based eyebrow root localization method, named *RootFinder*.

In our approach, the root distribution estimation of the eyebrow is viewed as the problem of density map prediction, which involves the modeling of the spatial distribution of root points as a continuous function over the image domain. The high-level idea of our approach is quite straightforward: given an eyebrow orientation map  $\mathbf{O}$ , our goal is to recover a density function  $\mathbf{P}$  that approximates the real distribution of each fiber root pixel in this image. Our notion of a density function  $\mathbf{P}$  can be interpreted as the response of a given pixel representing a root point. To prepare the training data, we employ the pre-processed orientation maps as input and utilize their corresponding cameras to project 3D ground-truth root points to the 2D image coordinates. Motivated by the previous crowd counting methods [Liu et al. 2019; Ranjan et al. 2021], we generate the ground-truth density map,  $P^* \in \mathbb{R}^{H \times W}$ , from the 2D root coordinates by using a Gaussian kernel with adaptive window size. Inspired by the encoder-decoder crowd counting framework [Jiang et al. 2019; Yan et al. 2019], we find that simply using a U-Net based architecture [Ronneberger et al. 2015] can already achieve satisfactory results. Following the standard settings, we use a pixel-wise MSE loss against  $P^*$ , which is formulated as:

$$L_{den} = \frac{1}{N} \sum_{i=1}^N \|P(O_i) - P_i^*\|^2, \quad (3)$$

where  $P(O_i)$  is the predicted density map.  $O_i$  is the  $i^{th}$  input of total  $N$  orientation maps, and  $P_i^*$  is the ground-truth density map.

Since predicting the distribution of eyebrow roots involves a rough estimation of their coordinates, the output density map at the patch level cannot be directly used for accurate pixel-level root localization. Thus we employ DBSCAN [Ester et al. 1996], a density-based clustering algorithm, to compute the number of valid fiber root clusters. Once we determine the number of clusters using DBSCAN, we apply the K-Means [Hartigan et al. 1979] clustering method to calculate the cluster center, which corresponds to the 2D root points. Next, our goal is to derive the 3D positions of the acquired 2D root coordinates. However, directly lifting the 2D root points

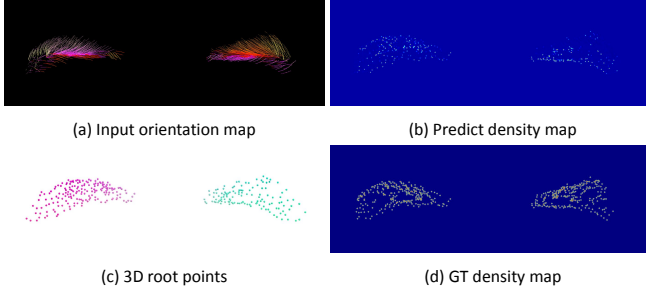


Fig. 4. Localization of the 3D root points by *RootFinder*. (a) input orientation map. (b) predict density map. (c) 3D root point cloud by clustering and lifting. (d) ground-truth density map.

to 3D space is an ill-posed problem since inverse projection leads to depth ambiguity. To overcome this issue, we first sample dense 3D points on the brow bone surface region of the face mesh and project them onto the image. For each extracted 2D root, we find its nearest projected 2D sampling points and take the corresponding 3D sampling point as the 3D root  $r_i$ . Relative maps mentioned in this subsection are illustrated in Fig. 4.

### 5.3 Prediction of 3D Orientation

To recover the growing direction of fiber  $S$ , *OriPredictor* is constructed to predict a unit vector for every point on  $S$  under the guidance of the given image  $I$ . In this task, the pre-processed 2D orientation map  $O$  from  $I$  is used to calculate the feature map  $F$ , and the 3D point  $p$  is projected to 2D image coordinate by camera projection  $\pi(\cdot)$  to exact pixel-aligned local image feature  $F(\pi(p))$ . In addition, to represent high-frequency variation in eyebrow fiber geometry, positional encoding  $\Phi(p)$  is added to transform coordinates to higher dimensional information [Mildenhall et al. 2021]. Then we can define the implicit orientation function  $D$  for arbitrary point  $p$  in 3D space as

$$D(F(\pi(p)), \Phi(p)) = d : d \in \mathbb{R}^3. \quad (4)$$

In our implementation, we apply an Hourglass filter to obtain feature maps from input orientation maps. Positional encoding will be concatenated with the image feature as a joint local feature. MLPs are designed as the decoder to output a unit 3D vector  $d$  and  $L_1$  loss is chosen to measure the error between the predicted direction  $d$  and corresponding ground-truth direction  $d^*$  during training.

### 5.4 Determination of Fiber Length

Theoretically, a growing eyebrow fiber  $S$  with  $q$  points can be discretized as a 3D points sequence  $\{p_0, p_1, \dots, p_{q-1}\}$ . *FiberEnder* is designed to determine whether  $S$  is supposed to be stopped by emitting an ending signal, which can be formulated into a binary classification problem.

It is trivial to think that one can tell the ending point of an eyebrow fiber on a close-up photo based on its surrounding pixels, so the local image feature should be considered. We also leverage 3D information by encoding fiber point coordinates into the local feature list since images may not be able to exhibit useful details in the region of self-occlusion. Hence, the concatenated growing

feature  $G$  for a single 3D point  $p$  can be given by

$$G(p) = \{F(\pi(p)), \Phi(p)\}, \quad (5)$$

where  $\pi(\cdot)$  is the camera projection.  $F(\pi(p))$  and  $\Phi(p)$  are the image feature and positional feature for  $p$ , respectively.

However, noticed that features of points along the same fiber have a strong correlation, we cannot treat a node as an isolated identity, since the prediction of whether  $S$  will be stopped at  $p_{q-1}$  is highly dependent on the features of its predecessor. Therefore, the ending function  $L$  for  $S$  can be defined as

$$\begin{aligned} L(S) &= L(\{p_0, p_1, \dots, p_{q-1}\}) \\ &= C(G(p_0), G(p_1), \dots, G(p_{q-1})) \\ &= \begin{cases} 0, & \text{if } S \text{ should be stopped at } p_{q-1} \\ 1, & \text{otherwise} \end{cases}, \quad (6) \end{aligned}$$

where  $C$  is a composite function taking the indefinite number of point feature  $G$  as ordinal inputs.

In practice, consistent with the prediction of the 3D orientation field, the Hourglass filter is employed to exact 2D image local features from input orientation maps. For organizing a point feature series with an unfixed length into a latent vector, we borrow knowledge from [Schuster and Paliwal 1997] and introduce a stacked RNN based encoder. The iteration begins with the initial zero hidden code  $h_0$  and it will be passed to a GRU cell [Cho et al. 2014] together with a new growing feature  $G(p_i)$ ,  $0 \leq i < q$  for total  $q$  times. Then the final hidden code  $h_q$  is passed to a MLP classifier to output a probability that  $S$  will continue growing. It is worth mentioning that the direction of fiber growth is not reversible, thus we use a monodirectional stacked RNN rather than a bidirectional one. The whole structure of *FiberEnder* is illustrated in Fig. 5.

Based on *EBStore*, we annotate all fibers and their sub-sequences with binary labels: the full sequence with the tip point is labeled to  $l = 0$  as the negative sample  $S^-$ , indicating the fiber will be stopped; the sub-sequences are labeled to  $l = 1$  as the positive sample  $S^+$ , suggesting the fiber will continue growing. In the training phase, we construct a data pair of  $\{S^-, S^+\}$  for each fiber in every epoch, where the positive sample is randomly picked from all its sub-sequences. Binary Cross Entropy (BCE) is chosen to calculate the loss between predicted normalized probability  $C(S)$  and labels:

$$L_{BCE} = -(l \cdot \log(C(S)) + (1-l) \cdot \log(1-C(S))). \quad (7)$$

### 5.5 Synthesis of Fiber-level Eyebrow

After  $\{R, D, L\}$  has been well-defined, the subsection is aimed to illustrate how to generate a fiber-level eyebrow  $E$ . The whole growing process can be generalized as *Syn*( $R, D, L$ ): Given  $n$  detected root points  $R = \{r_i, 0 \leq i < n\}$ ,  $n$  fibers will be generated by point-wise querying from  $r_i$ . For the  $j^{th}$  node  $p_{i,j}$  on fiber  $S_i$ , the unit growing direction  $D(p_{i,j}) = d_{i,j} \in \mathbb{R}^3$  and an ending signal  $L(S_i) = L(\{p_{i,0}, p_{i,1}, \dots, p_{i,j}\}) = l_{i,j} \in \{0, 1\}$ .  $S_i$  will keep growing as the direction of  $d_{i,j}$  with a fixed step  $\bar{s}$  if  $l_{i,j} = 1$ . Otherwise,  $S_i$  will be stopped at point  $p_{i,j}$ . The above synthesis procedure can be presented in Fig. 6. The step-wise iteration can also be summarized into an algorithm in Tab. 1 (In practice, it can be conducted in batch for all fibers of an eyebrow to achieve computational efficiency).

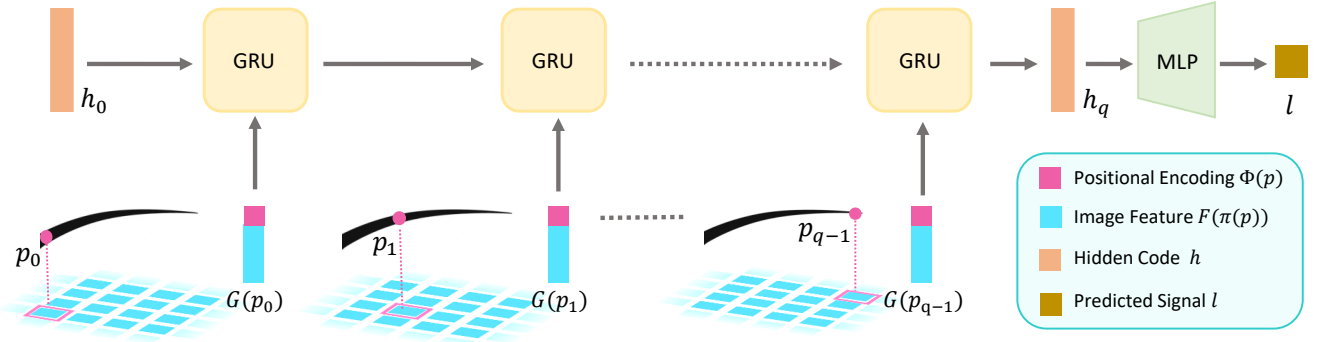


Fig. 5. The architecture of the proposed pixel-aligned stacked RNN-based *FiberEnder*. Radius change along a fiber has no particular meaning, only for intuitive illustration.

Table 1. Growing algorithm of the fiber-level eyebrow.

ALGORITHM: Fiber-level Eyebrow Synthesis	
<b>Input:</b>	root points $\mathbf{R}$ , 3D orientation function $\mathbf{D}$ , ending function $\mathbf{L}$
<b>Output:</b>	fiber-level eyebrow $\mathbf{E}$
<b>for</b> $i = 0 : n$ <b>do</b>	
build the empty point list $\mathbf{S}_i$ of the $i^{\text{th}}$ fiber	
$p_{i,0} = r_i$	
$\mathbf{S}_i = \{\mathbf{S}_i, p_{i,0}\}$	
<b>while</b> $j = 0 : \infty$ <b>do</b>	
$d_{i,j} = \mathbf{D}(p_{i,j})$	
$l_{i,j} = \mathbf{L}(\mathbf{S}_i)$	
<b>if</b> $l_{i,j} = 0$ : <b>break</b>	
$p_{i,j+1} = p_{i,j} + \bar{s}d_{i,j}$	
$\mathbf{S}_i = \{\mathbf{S}_i, p_{i,j+1}\}$	
<b>end</b>	
<b>end</b>	
$\mathbf{E} = \{\mathbf{S}_0, \mathbf{S}_1, \dots, \mathbf{S}_{n-1}\}$	

In addition, considering that the inevitable noises of filter-based orientation maps from real images cause the vibration of the 3D orientation field prediction, we also modify the predicted growing direction  $d$  to ensure smoothness and alleviate abnormal warping of fibers. Following [Shen et al. 2020], the growing direction is set to the mean value of the current and the last node if their angle difference is larger than a particular threshold  $\theta$ . However, the eyebrow fibers tend to exhibit less drastic angle variation between two adjacent points compared with hair strands, thus the given threshold in hair modeling is no more suitable for eyebrow modeling. Based on our experimental findings, we set  $\theta$  to be  $30^\circ$ .

## 6 EXPERIMENTS

### 6.1 Dataset

Our experiments are based on 400 synthetic eyebrow models in *EBStore*. For each 3D model, we employ flipping w.r.t the reflection symmetry plane of the head model to augment the 3D eyebrow

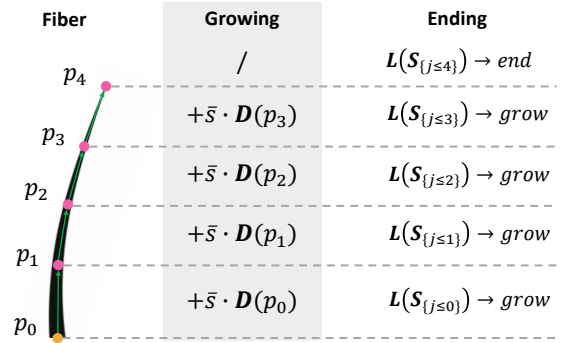


Fig. 6. Synthesis of an eyebrow fiber. Starting from a root point in  $\mathbf{R}$ , an eyebrow fiber is grown by querying the 3D orientation function  $\mathbf{D}(p)$  and the ending function  $\mathbf{L}(S)$  step-by-step. Radius change along a fiber has no particular meaning, only for intuitive illustration.

dataset and randomly select 3 views by moving the camera position in spherical coordinates with a fixed radius, azimuthal angle in  $\pm 10^\circ$  and polar angle in  $\pm 15^\circ$ . The camera ray is ensured to pass through the center point of the 3D eyebrow model. For each view of an eyebrow model, we render an orientation map. In total, we have 2202 images for training and 198 images for testing. Based on the observation that hair fibers may grow between the left and right eyebrow in some eyebrow types, we process the left and right parts simultaneously in all experiments instead of separating them for individual processing.

### 6.2 Metrics

In this subsection, we introduce fair metrics for the evaluation of 3D eyebrow modeling. We design several metrics to evaluate root localization and fiber length determination, respectively. Also, we use 3D metrics to assess the overall reconstruction performance of the whole system.

*Root localization metrics.* We first compute the nearest density error (*NDE*) and a density-aware chamfer distance (*DCD*) to evaluate

the localization of 3D eyebrow roots. The nearest density error is based on the density  $den(r, \mathbf{R})$  of root  $r$  against the root point cloud  $\mathbf{R}$ , which is calculated by counting the number of neighbors within a radius  $\varphi$ . Then the nearest density errors from predicted roots  $\mathbf{R}'$  to the ground-truth  $\mathbf{R}^*$  can be defined as

$$NDE' = \frac{1}{|\mathbf{R}'|} \sum_{x \in \mathbf{R}'} \left| den(x, \mathbf{R}') - den(x^*, \mathbf{R}^*) \right|, \quad (8)$$

where  $x^* \in \mathbf{R}^*$  is the nearest root of  $x \in \mathbf{R}'$ . Based on the density errors, we also compute a density-aware chamfer distance from predicted roots  $\mathbf{R}'$  to the ground-truth  $\mathbf{R}^*$

$$DCD' = \frac{1}{|\mathbf{R}'|} \sum_{x \in \mathbf{R}'} \left| den(x, \mathbf{R}') - den(x^*, \mathbf{R}^*) + 1 \right| \cdot \|x - x^*\|_2. \quad (9)$$

The reversed  $NDE^*$  and  $DCD^*$  can be calculated in the same way. We take their arithmetic means  $NDE = (NDE' + NDE^*)/2$  and  $DCD = (DCD' + DCD^*)/2$  as the bi-directional metrics to evaluate the root localization.

*Fiber length determination metrics.* We define the Mean Length Error (*MLE*) for evaluating the average loss per fiber for  $k$  synthetic 3D eyebrow, which can be calculated as

$$MLE = \frac{\sum_{i=1}^k \sum_{n=1}^{n_i} \bar{s} |l_{i,j} - l_{i,j}^*|}{\sum_{i=1}^k n_i}, \quad (10)$$

where  $n_i$  is the total number of fibers of the  $i^{th}$  ( $1 \leq i \leq k$ ) eyebrow,  $l_{i,j}$  represents the predicted length labels for its  $j^{th}$  ( $1 \leq j \leq n_i$ ) fiber, and  $l_{i,j}^*$  stands for its corresponding ground-truth label.  $\bar{s}$  is the pre-determined fixed step.

*3D fiber-level eyebrow reconstruction metrics.* For the 3D metrics, we use the IoU of meshes converted from reconstructed eyebrows  $\mathbf{E}'$  and ground-truth eyebrows  $\mathbf{E}^*$  via the method in [Zhu and Bridson 2005], which is designed to evaluate the overall shape, i.e. the silhouette of the reconstructed eyebrow model. To further compare the inside geometry of eyebrow growing, we evaluate the fiber-level  $L_2$  distance of 3D orientation, which is shorted as *FDO*. For the *FDO'* from  $\mathbf{E}'$  to  $\mathbf{E}^*$ , we first find the nearest fiber  $\mathbf{S}^* \in \mathbf{E}^*$  for each fiber  $\mathbf{S} \in \mathbf{E}'$  according to the distance of their roots, then compute *FDO'* as

$$FDO' = \frac{1}{|\mathbf{E}'|} \sum_{\mathbf{S} \in \mathbf{E}'} \sum_{\mathbf{p} \in \mathbf{S}, \mathbf{q} \in \mathbf{S}^*} \|\mathbf{D}(\mathbf{p}_i) - \mathbf{D}(\mathbf{q}_i)\|_2, \quad (11)$$

where  $\mathbf{D}(\mathbf{p}_i)$  and  $\mathbf{D}(\mathbf{q}_i)$  are the unit growing directions of the  $i^{th}$  points  $\mathbf{p}_i$  and  $\mathbf{q}_i$  along  $\mathbf{S}$  and  $\mathbf{S}^*$ , respectively. We use the bi-directional error  $FDO = (FDO' + FDO^*)/2$  as the metric.

### 6.3 Evaluations

This subsection is to evaluate our key modules *RootFinder*, *OriPredictor* and *FiberEnder* on the synthetic test set by both quantitative and qualitative results.

*Evaluation of eyebrow roots.* Firstly, we analyze the performances of *RootFinder* by calculating *DCD* and *NDE*. We compare our *RootFinder* with an intuitive method which first conducts the uniformly random sampling of 2D roots on 2D matting masks of eyebrow images, then

lifts 2D roots to 3D using the same method in Sec. 5.2. From Tab. 2, it is clear to see a significant reduction of two metrics with different neighborhood radius  $\varphi$ . In our setting, the target face is normalized to  $[-1, 1]$  in width. Thus, we choose  $\varphi$  as 0.04, 0.02, and 0.01 for comprehensive evaluations. Visual comparisons are illustrated by predicted 3D roots with the colorization of *DCD'* ( $\varphi = 0.02$ ) and extracted 2D roots with the background of ground-truth eyebrows (see Fig. 7). The most distinguished problem of random sampling is that randomly sampled roots cannot cover the eyebrow shape with correct sparseness. Almost all selected roots fall in dense pixel regions and barely on sparse regions, such as the tails and the fibers between two eyebrows. In contrast, *RootFinder* maintains the overall boundary very well, while also preserving the local distribution suggested by the input image.

Table 2. Quantitative comparisons for the Localization of fiber roots.

Method	<i>DCD</i>	<i>NDE</i>	<i>DCD</i>	<i>NDE</i>	<i>DCD</i>	<i>NDE</i>
	( $\varphi=0.04$ )		( $\varphi=0.02$ )		( $\varphi=0.01$ )	
<i>Random</i>	0.2147	21.57	0.0811	7.503	0.03267	2.506
<b><i>RootFinder</i></b>	0.1261	15.13	0.04811	5.078	0.02222	1.735

*Evaluation of 3D orientation.* We follow [Wu et al. 2022; Zheng et al. 2023] to use  $L_2$  error to evaluate the orientation field. The  $L_2$  error of our method is 0.0951. From Fig. 9, the results of our method have a much better alignment with input eyebrow images in fiber growing direction than previous ruled-based approaches [Herrera et al. 2010; Rotger et al. 2019], indicating that the predicted orientation field is capable of describing how the eyebrow should grow in the aforementioned implicit manner.

*Evaluation of fiber length.* To evaluate the effectiveness of our method on the determination of fiber length, we design experiments with five settings: ending fiber growing with the mean length of all fibers in *EBStore* (baseline), cutting using the ground-truth contour mesh, our method without recurrent module, our method without positional encoding  $\Phi(p)$ , and our full model *FiberEnder*. As for the mesh-cutting method, we follow [Hu et al. 2017] to first extract watertight ground-truth eyebrow meshes from ground-truth eyebrows using the method in [Zhu and Bridson 2005]. Given the ground-truth mesh, we can easily stop the growth of a fiber if it grows out of the boundary of the mesh. The quantitative results are displayed in Tab. 3. Our full model has a 45.77% reduction in *MLE* compared with the intuitive approach by setting all fibers into the mean length and 60.16% less than the mesh-cutting method, even if it uses ground-truth meshes rather than predicted meshes. Moreover, ablation of the recurrent architecture and positional information does affect the results to a certain extent, which can provide evidence to our assumption that historical growing features help in the learning of length control of eyebrow fibers. Visual results colored by *MLE* for each fiber in Fig. 8, revealing more obvious differences between these approaches. Growing with mean length tends to preserve overall characteristics but lose length details in the region that is out of distribution. Mesh-cutting can only cut off fibers correctly which should stop near the boundary of the



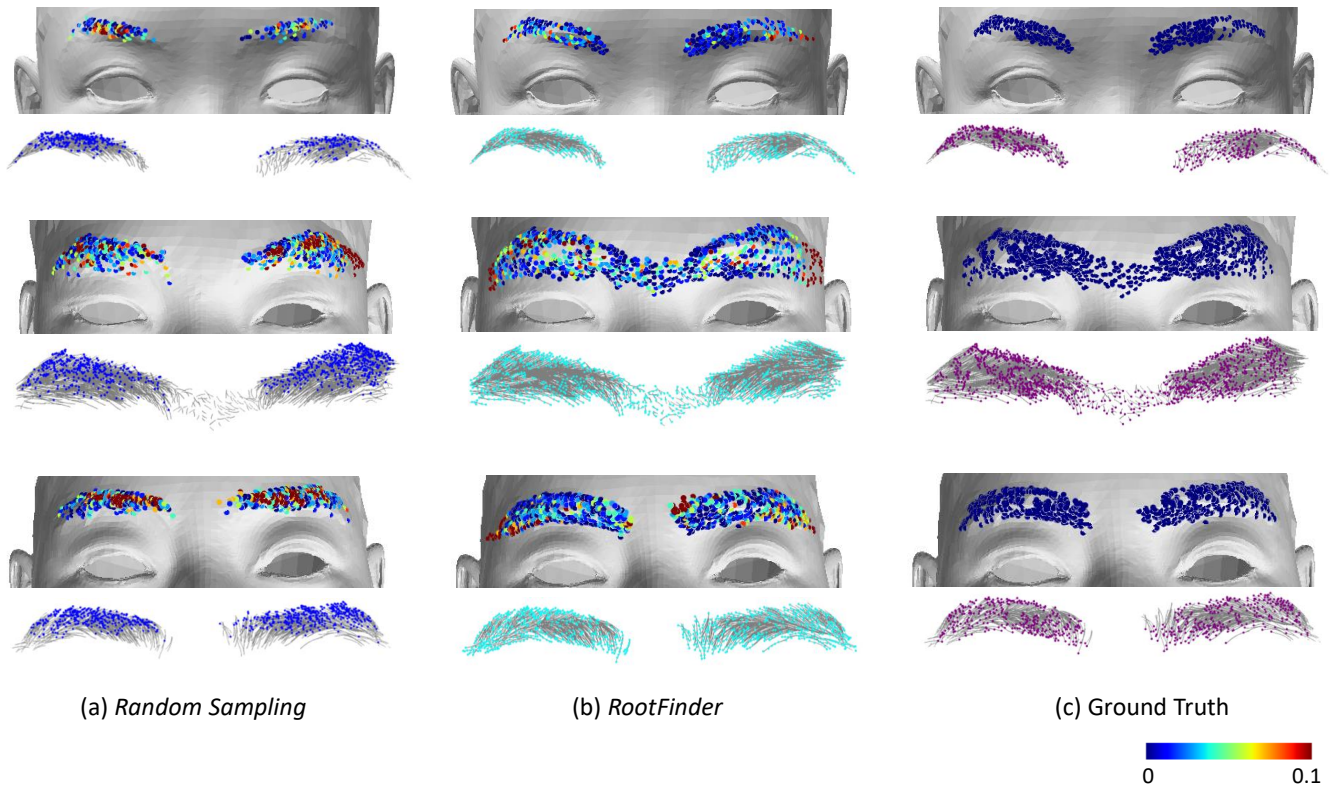


Fig. 7. Qualitative comparisons for the localization of fiber root. Every row presents the 3D root points and their corresponding projection in 2D of the same identity. From left to right: (a) Random sampling in eyebrow matting mask region, (b) *RootFinder*, (c) Ground-truth. Each 3D root point is colored according to  $DCD'$  ( $\varphi = 0.02$ ).

mesh but fails to work on fibers that should stop inside the mesh, especially in the middle region of the eyebrows. Moreover, results will be even worse if the predicted coarse meshes are used instead of the ground-truth ones. To sum up, *FiberEnder* outperforms other methods to much extent.

Table 3. Quantitative comparisons for the determination of fiber length.

Method	$MLE (\times 10^{-2}) \downarrow$
<i>End with mean length (Baseline)</i>	2.9926
<i>Cut by ground-truth mesh</i>	4.0739
<i>FiberEnder w/o recurrent module</i>	4.2130
<i>FiberEnder w/o PE</i>	1.8070 (-39.62%)
<b><i>FiberEnder</i></b>	<b>1.6229 (-45.77%)</b>

#### 6.4 Comparisons

This subsection is to compare *EMS* with existing related approaches by quantitative and qualitative results.

*Comparison with rule-based facial hair reconstruction.* Since there have been no existing learning-based attempts for 3D eyebrow modeling tasks, we implement rule-based methods in [Herrera et al.

2010] and [Rotger et al. 2019] to make comparisons. On the testing set of *EBStore*, we report IoU and *FDO* for reference in Tab. 4. With the help of *RootFinder*, *OriPredictor* and *FiberEnder*, *EMS* outperforms the traditional methods in both metrics. We also collect some in-the-wild photos with an iPhone Pro Max 13 in our local area. As shown in Fig. 9, our method achieves state-of-the-art performance. Compared with previous methods, our reconstructed eyebrows have a more similar shape matching with original images and the fiber growing pattern is more natural, thanks to the intrinsic advantages of the data-driven approach by introducing 3D priors to single view reconstruction. Results in Fig. 9 prove that our method trained on synthetic data also works for the input of real images with a non-negligible domain gap. However, we have to admit that, due to the lack of diversity in FaceScape [Yang et al. 2020b] (most identities are Asian), our dataset *EBStore* may fail to cover all properties such as eyebrow types, skin tones, and fiber colors despite our efforts to increase the variety of *EBStore*. Although this limitation may have a certain negative impact on the generalization capability of our method, the reconstructed results tested on the online public images [Pexels 2021] with different skin tones and fiber colors still look fine (Fig. 10).



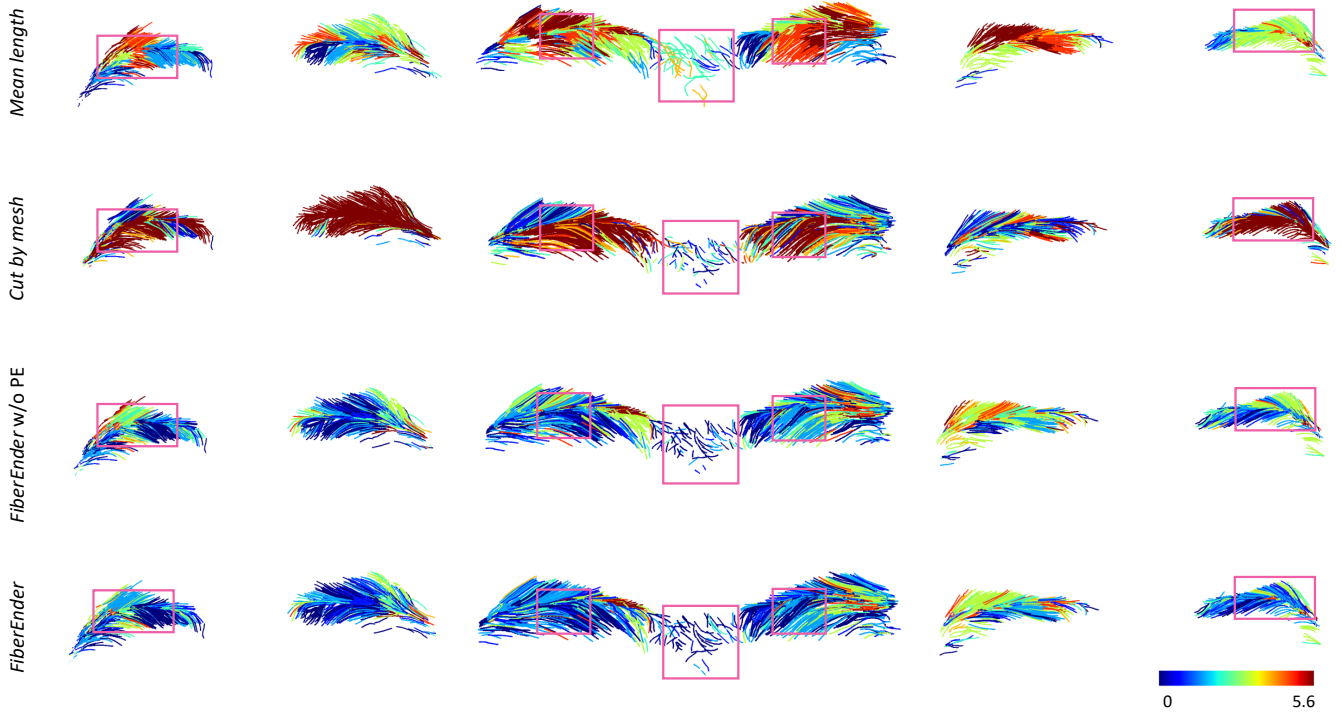


Fig. 8. Qualitative comparisons for the determination of fiber length. From top to bottom, we show the reconstructed 3D eyebrow models of ending with mean length, cutting by ground-truth mesh, *FiberEnder* w/o PE and *FiberEnder*, respectively. Every column shows the results of the same identity. Each fiber is colored according to MLE ( $\times 10^{-2}$ ).

It is worth mentioning that the accuracy of eyebrow reconstruction, especially 3D root prediction, is partially based on an off-the-shelf single-view face reconstruction method [Yang et al. 2020b]. Even when only coarse face models are obtained sometimes in in-the-wild examples, the quality of eyebrow reconstruction is still visually satisfactory. The reason may be that our method mainly relies more on the projected growing hints.

*Comparison with multi-view facial hair reconstruction.* To further illustrate the effectiveness of our single-view reconstruction framework, we compare *EMS* with a multi-view method [Beeler et al. 2012] by taking only a front-view image as input. From the result in Fig. 11, although *EMS* may not produce highly detailed results like multi-view capture systems, it has already depicted the overall appearance of a comparable level with a more handy setting.

*Comparison with single-view scalp hair reconstruction.* To make the experiments more comprehensive, we compare our method with state-of-the-art single-view scalp hair modeling methods [Wu et al. 2022; Zheng et al. 2023], which are usually designed to grow fibers from dense pre-sampled roots on the scalp according to an implicit orientation field, and cut using predicted coarse meshes. For convenience, we compare *EMS* with cutting by ground-truth meshes, which is an even more tolerant setting. IoU and *DCD* are reported in Tab. 4, where *EMS* is better thanks to module design for eyebrow characteristics especially.

Table 4. Quantitative comparisons with existing related methods.

Method	IoU $\uparrow$	<i>FDO</i> $\downarrow$
[Herrera et al. 2010]	0.5095	0.9752
[Rotger et al. 2019]	0.5126	0.9848
<i>Scalp hair SVR</i>	0.6160	0.2758
<b><i>EMS</i></b>	0.8756	0.2480

## 6.5 Ablation Study

To better demonstrate the necessity of the designing for each module, the whole *EMS* system is ablated into three settings: growing with random roots and a mean length (baseline), baseline with only *FiberEnder* and baseline with only *RootFinder*. They are all compared with each other and our full model by the metrics of IoU and *FDO* against the ground truth on the testing set.

Table 5. Quantitative comparisons for the ablation study.

Method	IoU $\uparrow$	<i>FDO</i> $\downarrow$
<i>Baseline</i>	0.5665	0.2425
<i>Baseline + RootFinder</i>	0.8440	0.2655
<i>Baseline + FiberEnder</i>	0.5918	0.2338
<b><i>Full</i></b>	0.8756	0.2480



Fig. 9. Qualitative comparisons for in-the-wild real images. From top to bottom, we show the reconstructed 3D eyebrow models of previous methods [Herrera et al. 2010], [Rotger et al. 2019], and EMS, respectively. The different colors used have no particular meaning, just for fruitful illustration of different examples.

From the comparisons of IoU in Tab. 5 and Fig. 12, we can tell that *RootFinder* exerts a significant effect on the recovery of the overall shape of 3D eyebrows. Using randomly sampled roots, the tails and the region between two eyebrows cannot be covered very well. Furthermore, *FiberEnder* also brings some bonus on both metrics, that is, controlling the length of every fiber not only preserves a better overall volume but also modifies wrong orientation by ending inappropriate growing. However, with the huge improvement of the IoU brought by *RootFinder*, a small perturbation in the accuracy of orientation is acceptable.

## 6.6 User Study

Last but not least, we make a user study on three methods in Fig. 9, with a total of 56 users involved. In the questionnaire, they are provided with 8 randomly chosen reconstructed examples by different approaches and asked to select one result that matches the given close-up photos best from three randomly sorted options. 78.57% of answers vote for EMS as the best, while 13.39% and 8.04% vote for the methods of [Herrera et al. 2010] and [Rotger et al. 2019], respectively. It is supportive evidence for the superiority of our method.



Fig. 10. Qualitative comparisons for diverse skin tones and fiber colors. We show the reconstructed 3D eyebrow models and their corresponding original images. The different colors used have no particular meaning, just for fruitful illustration of different examples.



Fig. 11. Qualitative comparisons with the multi-view method. From top to bottom, we show the reconstructed 3D eyebrow models of [Beeler et al. 2012], EMS, and its corresponding real image.

## 7 CONCLUSION AND LIMITATION

To conclude, in this work, we propose the first learning-based system for fiber-level 3D eyebrow reconstruction from single-view images. EMS consists of three components *RootFinder*, *OriPredictor*

and *FiberEnder*, aiming to tackle the key challenges of eyebrow modeling. We also contribute the first high-quality 3D synthetic eyebrow dataset *EBStore*, which we use to facilitate the whole framework and introduce objective metrics to assess the performance of each module and overall reconstruction geometries. Experiments with visual and numerical results on both synthetic data and in-the-wild real images show the effectiveness of our system.

**Limitation.** Although EMS provides realistic reconstructed results on a wide range of data, it may fail in some cases shown in Fig. 14: continuous dark eyebrow region of make-up or eyebrow tattoos (as the first row), occlusion of eyebrows by hair strands on the forehead (as the second row), and relatively slight contrast between skin tone and hair color (e.g., white eyebrows on light skin, as the third row). These artifacts are mainly caused by the eyebrow matting module [Wang et al. 2022]. In these cases, the masks we obtain may be blurry or have wrong detection of eyebrow fibers, which will further lead to low-quality orientation maps, wrong distribution of eyebrow roots and inaccuracy of orientation and length prediction. We believe that if we can find a more robust matting method or ease the dependence on it, our method is supposed to work better.

Besides, the generalization ability of our model is limited to eyebrows and may fail for other types of facial hairs, since prior knowledge learned from *EBStore* may be unsuitable to represent different growing rules. Another thing that needs to be noted is that our current model is trained using synthetic data only which takes the orientation map as the intermediate input. This makes the issue of



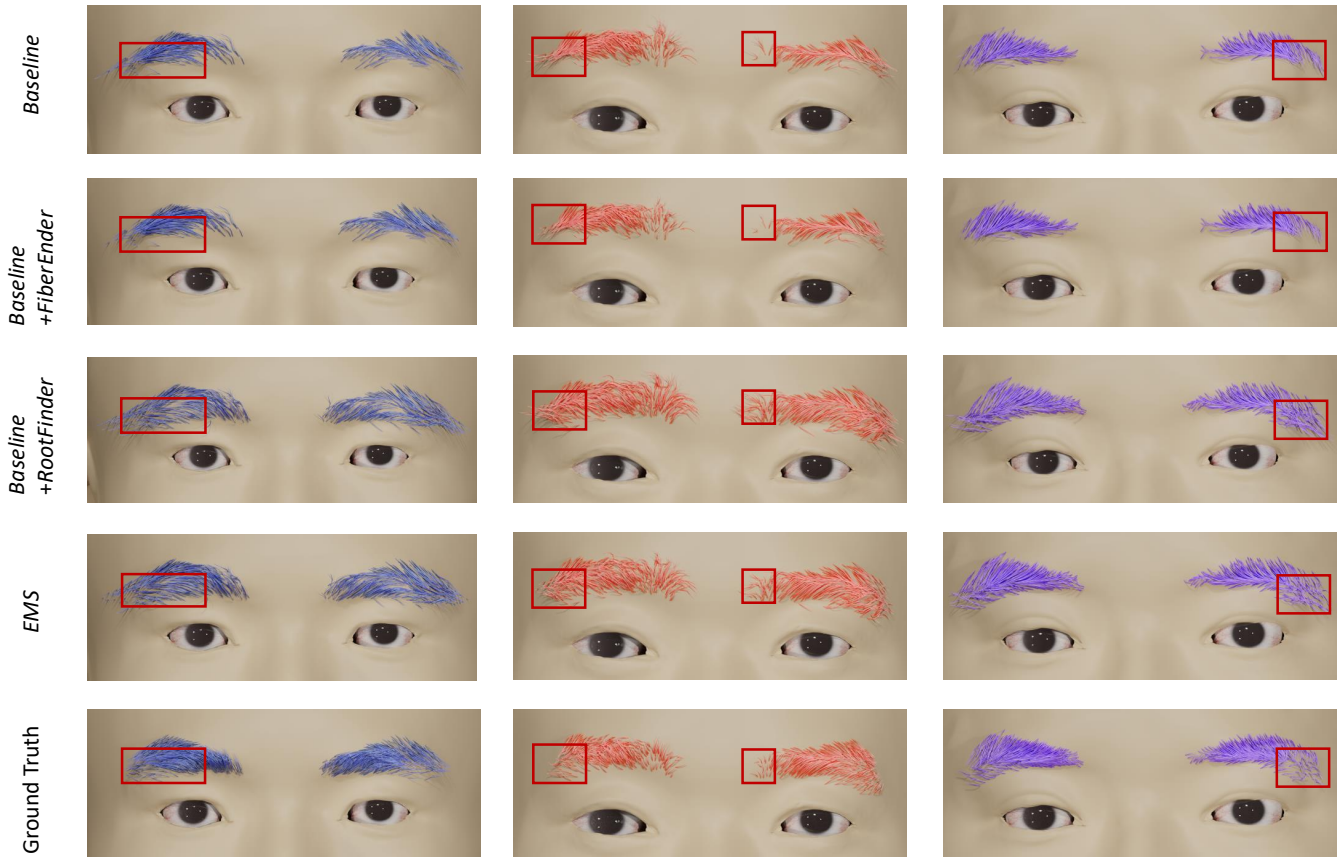


Fig. 12. Qualitative ablation study results. From top to bottom, we show the reconstructed 3D eyebrow models of the baseline, baseline with *FiberEnder* only, baseline with *RootFinder* only, our full model *EMS* and ground truth, respectively. The different colors used have no particular meaning, just for fruitful illustration of different examples.

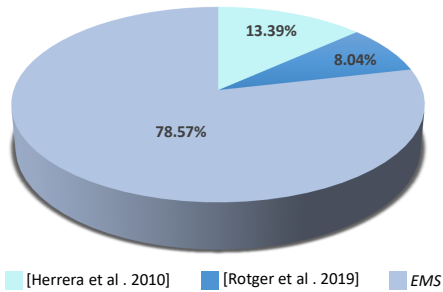


Fig. 13. Statistics of user study. The pie chart displays the percentage of votes for the best eyebrow models among three methods represented by different colors.

domain gap still exist since the orientation maps extracted from real-world images are usually noisy. We leave this as a future research direction.

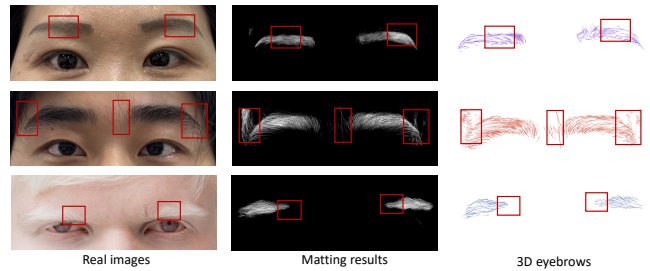


Fig. 14. Failure cases. From left to right, we show the real images, corresponding matting results and reconstructed 3D eyebrows. The different colors used have no particular meaning, just for fruitful illustration of different examples.

#### ACKNOWLEDGMENTS

The work was supported in part by NSFC with Grant No. 62293482, the Basic Research Project No. HZQB-KCZYZ-2021067 of Hetao

ShenzhenHK S&T Cooperation Zone, the National Key R&D Program of China with Grant No. 2018YFB1800800, the Shenzhen Outstanding Talents Training Fund 202002, the Guangdong Research Projects No. 2017ZT07X152 and No. 2019CX01X104, the Guangdong Provincial Key Laboratory of Future Networks of Intelligence (Grant No. 2022B1212010001), and the Shenzhen Key Laboratory of Big Data and Artificial Intelligence (Grant No. ZDSYS201707251409055). It was also partially supported by the Hong Kong Research Grants Council under General Research Funds (HKU17206218). Additionally, it received partial support from NSFC-62172348, the Outstanding Young Fund of Guangdong Province with No. 2023B1515020055, and the Shenzhen General Project with No. JCYJ20220530143604010.

## REFERENCES

- Linchao Bao, Xiangkai Lin, Yajing Chen, Haoxian Zhang, Sheng Wang, Xuefei Zhe, Di Kang, Haozhi Huang, Xinwei Jiang, Jue Wang, et al. 2021. High-fidelity 3d digital human head creation from rgb-d selfies. *ACM Transactions on Graphics (TOG)* 41, 1 (2021), 1–21.
- Thabo Beeler, Bernd Bickel, Gioacchino Noris, Paul Beardsley, Steve Marschner, Robert W Sumner, and Markus Gross. 2012. Coupled 3D reconstruction of sparse facial hair and skin. *ACM Transactions on Graphics (TOG)* 31, 4 (2012).
- Pascal Bérard, Derek Bradley, Markus Gross, and Thabo Beeler. 2016. Lightweight eye capture using a parametric model. *ACM Transactions on Graphics (TOG)* 35, 4 (2016), 1–12.
- Pascal Bérard, Derek Bradley, Maurizio Nitti, Thabo Beeler, and Markus H Gross. 2014. High-quality capture of eyes. *ACM Trans. Graph.* 33, 6 (2014), 223–1.
- Amit Bermanto, Thabo Beeler, Yera Kozlov, Derek Bradley, Bernd Bickel, and Markus Gross. 2015. Detailed spatio-temporal reconstruction of eyelids. *ACM Transactions on Graphics (TOG)* 34, 4 (2015), 1–11.
- Menglei Chai, Tianjia Shao, Hongzhi Wu, Yanlin Weng, and Kun Zhou. 2016. Autohair: Fully automatic hair modeling from a single image. *ACM Transactions on Graphics* 35, 4 (2016).
- Menglei Chai, Lvdi Wang, Yanlin Weng, Xiaogang Jin, and Kun Zhou. 2013. Dynamic hair manipulation in images and videos. *ACM Transactions on Graphics (TOG)* 32, 4 (2013), 1–8.
- Menglei Chai, Lvdi Wang, Yanlin Weng, Yizhou Yu, Baining Guo, and Kun Zhou. 2012. Single-view hair modeling for portrait manipulation. *ACM Transactions on Graphics (TOG)* 31, 4 (2012).
- Kyunghyun Cho, Bart Van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. 2014. Learning phrase representations using RNN encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078* (2014).
- Blender Online Community. 2018. *Blender - a 3D modelling and rendering package*. Blender Foundation, Stichting Blender Foundation, Amsterdam. <http://www.blender.org>
- Dataset. 2022. Eyebrows Dataset. <https://universe.roboflow.com/dataset-0dpth/eyebrows>. <https://universe.roboflow.com/dataset-0dpth/eyebrows> visited on 2023-05-12.
- Yu Deng, Jialong Yang, Sicheng Xu, Dong Chen, Yunde Jia, and Xin Tong. 2019. Accurate 3d face reconstruction with weakly-supervised learning: From single image to image set. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*. 0–0.
- Dimitar Dinev, Thabo Beeler, Derek Bradley, Moritz Bächer, Hongyi Xu, and Ladislav Kavan. 2018. User-guided lip correction for facial performance capture. In *Computer Graphics Forum*, Vol. 37. 93–101.
- Martin Ester, Hans-Peter Kriegel, Jörg Sander, Xiaowei Xu, et al. 1996. A density-based algorithm for discovering clusters in large spatial databases with noise.. In *kdd*, Vol. 96. 226–231.
- Yao Feng, Haiwen Feng, Michael J. Black, and Timo Bolkart. 2021. Learning an Animatable Detailed 3D Face Model from In-The-Wild Images. *ACM Transactions on Graphics, (Proc. SIGGRAPH)* 40, 8. <https://doi.org/10.1145/3450626.3459936>
- Graham Fyfe. 2012. High fidelity facial hair capture. In *ACM SIGGRAPH 2012 Talks*.
- Pablo Garrido, Michael Zollhöfer, Chenglei Wu, Derek Bradley, Patrick Pérez, Thabo Beeler, and Christian Theobalt. 2016. Corrective 3D reconstruction of lips from monocular video. *ACM Trans. Graph.* 35, 6 (2016), 219–1.
- C Lee Giles, Gary M Kuhn, and Ronald J Williams. 1994. Dynamic recurrent neural networks: Theory and applications. *IEEE Transactions on Neural Networks* 5, 2 (1994), 153–156.
- Stephane Grabli, Francois X. Sillion, Stephen R. Marschner, and Jerome E. Lengyel. 2002. Image-based hair capture by inverse lighting. In *Proceedings of Graphics Interface (GI)*. 51–58.
- Jianzhu Guo, Xiangyu Zhu, Yang Yang, Fan Yang, Zhen Lei, and Stan Z Li. 2020. Towards fast, accurate and stable 3d dense face alignment. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XIX*. Springer, 152–168.
- John A Hartigan, Manchek A Wong, et al. 1979. A k-means clustering algorithm. *Applied statistics* 28, 1 (1979), 100–108.
- Tomas Lay Herrera, Arno Zinke, and Andreas Weber. 2012. Lighting hair from the inside: A thermal approach to hair reconstruction. *ACM Transactions on Graphics (TOG)* 31, 6 (2012), 1–9.
- Tomas Lay Herrera, Arno Zinke, Andreas Weber, and Thomas Vetter. 2010. Toward image-based facial hair modeling. In *Proceedings of the 26th Spring Conference on Computer Graphics*.
- Liwen Hu, Chongyang Ma, Linjie Luo, and Hao Li. 2014a. Robust hair capture using simulated examples. *ACM Transactions on Graphics (TOG)* 33, 4 (2014), 1–10.
- 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).
- Liwen Hu, Chongyang Ma, Linjie Luo, Li-Yi Wei, and Hao Li. 2014b. Capturing braided hairstyles. *ACM Transactions on Graphics (TOG)* 33, 6 (2014).
- Liwen Hu, Shunsuke Saito, Lingyu Wei, Koki Nagano, Jaewoo Seo, Jens Fursund, Iman Sadeghi, Carrie Sun, Yen-Chun Chen, and Hao Li. 2017. Avatar digitization from a single image for real-time rendering. *ACM Transactions on Graphics (ToG)* 36, 6 (2017), 1–14.
- Wenzel Jakob, Jonathan T Moon, and Steve Marschner. 2009. Capturing hair assemblies fiber by fiber. *ACM Transactions on Graphics (TOG)* 28, 5 (2009), 1–9.
- Xiaolong Jiang, Zehao Xiao, Baochang Zhang, Xiantong Zhen, Xianbin Cao, David Doermann, and Ling Shao. 2019. Crowd counting and density estimation by trellis encoder-decoder networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 6133–6142.
- Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).
- Zhiyi Kuang, Yiyang Chen, Hongbo Fu, Kun Zhou, and Youyi Zheng. 2022. DeepMVSHair: Deep Hair Modeling from Sparse Views. In *ACM SIGGRAPH Asia 2022 Conference Papers*.
- Cheng-Han Lee, Ziwei Liu, Lingyun Wu, and Ping Luo. 2020. Maskgan: Towards diverse and interactive facial image manipulation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 5549–5558.
- Chloe LeGendre, Loc Hyunh, Shanhe Wang, and Paul Debevec. 2017. Modeling vellus facial hair from asperity scattering silhouettes. In *ACM SIGGRAPH 2017 Talks*.
- Gengyan Li, Abhimitra Meka, Franziska Mueller, Marcel C Buehler, Otmar Hilliges, and Thabo Beeler. 2022. EyeNeRF: a hybrid representation for photorealistic synthesis, animation and relighting of human eyes. *ACM Transactions on Graphics (TOG)* 41, 4 (2022), 1–16.
- 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>
- Shu Liang, Xiufeng Huang, Xianyu Meng, Kunyao Chen, Linda G Shapiro, and Ira Kemelmacher-Shlizerman. 2018. Video to fully automatic 3d hair model. *ACM Transactions on Graphics (TOG)* 37, 6 (2018).
- Weizhe Liu, Mathieu Salzmann, and Pascal Fua. 2019. Context-aware crowd counting. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 5099–5108.
- Matthew Loper, Naureen Mahmood, Javier Romero, Gerard Pons-Moll, and Michael J. Black. 2015. SMPL: A Skinned Multi-Person Linear Model. *ACM Trans. Graphics (Proc. SIGGRAPH Asia)* 34, 6 (Oct. 2015), 248:1–248:16.
- Linjie Luo, Hao Li, Sylvain Paris, Thibaut Weise, Mark Pauly, and Szymon Rusinkiewicz. 2012. Multi-view hair capture using orientation fields. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 1490–1497.
- Linjie Luo, Hao Li, and Szymon Rusinkiewicz. 2013. Structure-aware hair capture. *ACM Transactions on Graphics (TOG)* 32, 4 (2013).
- Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. 2021. Nerf: Representing scenes as neural radiance fields for view synthesis. *Commun. ACM* 65, 1 (2021), 99–106.
- Giljoo Nam, Chenglei Wu, Min H Kim, and Yaser Sheikh. 2019. Strand-accurate multi-view hair capture. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 155–164.
- Alejandro Newell, Kaiyu Yang, and Jia Deng. 2016. Stacked hourglass networks for human pose estimation. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part VIII 14*. Springer, 483–499.
- Sylvain Paris, Hector M. Briceno, and Francois X. Sillion. 2004. Capture of hair geometry from multiple images. *ACM transactions on graphics (TOG)* 23, 3 (2004), 712–719.
- Sylvain Paris, Will Chang, Oleg I Kozhushnyan, Wojciech Jarosz, Wojciech Matusik, Matthias Zwicker, and Frédo Durand. 2008. Hair photobooth: geometric and photometric acquisition of real hairstyles. *ACM Trans. Graph(Proc. SIGGRAPH)* 27, 3 (2008), 30.



- Pexels. 2021. *Pexels - The best free stock photos, royalty free images & videos shared by creators*. <https://www.pexels.com>
- Stylios Ploumpis, Stylios Moschoglou, Vasileios Triantafyllou, and Stefanos Zafeiriou. 2022. 3D human tongue reconstruction from single" in-the-wild" images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2771–2780.
- Viresh Ranjan, Udbhav Sharma, Thu Nguyen, and Minh Hoai. 2021. Learning to count everything. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 3394–3403.
- Olaf Ronneberger, Philipp Fischer, and Thomas Brox. 2015. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015: 18th International Conference, Munich, Germany, October 5–9, 2015, Proceedings, Part III 18*. Springer, 234–241.
- Radu Alexandru Rosu, Shunsuke Saito, Ziyang Wang, Chenglei Wu, Sven Behnke, and Giljoo Nam. 2022. Neural strands: Learning hair geometry and appearance from multi-view images. In *European Conference on Computer Vision*. Springer, 73–89.
- Gemma Rotger, Francesc Moreno-Noguer, Felipe Lumbrales, and Antonio Agudo. 2019. Single view facial hair 3D reconstruction. In *Pattern Recognition and Image Analysis*. Springer, 423–436.
- Javid Sadi, Izzat Jarudi, and Pawan Sinha. 2003. The role of eyebrows in face recognition. *Perception* 32, 3 (2003), 285–293.
- Shunsuke Saito, Liwen Hu, Chongyang Ma, Hikaru Ibayashi, Linjie Luo, and Hao Li. 2018. 3D hair synthesis using volumetric variational autoencoders. *ACM Transactions on Graphics (TOG)* 37, 6 (2018).
- Shunsuke Saito, Tomas Simon, Jason Saragih, and Hanbyul Joo. 2020. Pifuhd: Multi-level pixel-aligned implicit function for high-resolution 3d human digitization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 84–93.
- Mike Schuster and Kuldip K Paliwal. 1997. Bidirectional recurrent neural networks. *IEEE transactions on Signal Processing* 45, 11 (1997), 2673–2681.
- Steven M Seitz, Brian Curless, James Diebel, Daniel Scharstein, and Richard Szeliski. 2006. A comparison and evaluation of multi-view stereo reconstruction algorithms. In *2006 IEEE computer society conference on computer vision and pattern recognition (CVPR'06)*, Vol. 1. IEEE, 519–528.
- Yuefan Shen, Changgeng Zhang, Hongbo Fu, Kun Zhou, and Youyi Zheng. 2020. Deepsketchhair: Deep sketch-based 3d hair modeling. *IEEE transactions on visualization and computer graphics* 27, 7 (2020), 3250–3263.
- Tiancheng Sun, Giljoo Nam, Carlos Aliaga, Christophe Hery, and Ravi Ramamoorthi. 2021. Human hair inverse rendering using multi-view photometric data. (2021).
- Zdravko Velinov, Marios Papas, Derek Bradley, Paulo Gotardo, Parsa Mirdehghan, Steve Marschner, Jan Novák, and Thabo Beeler. 2018. Appearance capture and modeling of human teeth. *ACM Transactions on Graphics (TOG)* 37, 6 (2018), 1–13.
- Luyuan Wang, Hanyuan Zhang, Qinjie Xiao, Hao Xu, Chunhua Shen, and Xiaogang Jin. 2022. Effective Eyebrow Matting with Domain Adaptation. (2022).
- Yichen Wei, Eyal Ofek, Long Quan, and Heung-Yeung Shum. 2005. Modeling hair from multiple views. *ACM Transactions on Graphics (TOG)* 24, 3 (2005), 816–820.
- Quan Wen, Feng Xu, Ming Lu, and Jun-Hai Yong. 2017. Real-time 3D eyelids tracking from semantic edges. *ACM Transactions on Graphics (TOG)* 36, 6 (2017), 1–11.
- Sebastian Winberg, Gaspard Zoss, Prashanth Chandran, Paulo Gotardo, and Derek Bradley. 2022. Facial hair tracking for high fidelity performance capture. *ACM Transactions on Graphics (TOG)* 41, 4 (2022), 1–12.
- Chenglei Wu, Derek Bradley, Pablo Garrido, Michael Zollhöfer, Christian Theobalt, Markus H Gross, and Thabo Beeler. 2016. Model-based teeth reconstruction. *ACM Trans. Graph.* 35, 6 (2016), 220–1.
- Keyu Wu, Yifan Ye, Lingchen Yang, Hongbo Fu, Kun Zhou, and Youyi Zheng. 2022. NeuralHDHair: Automatic High-fidelity Hair Modeling from a Single Image Using Implicit Neural Representations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 1526–1535.
- Qinjie Xiao, Hanyuan Zhang, Zhaorui Zhang, Yiqian Wu, Luyuan Wang, Xiaogang Jin, Xinwei Jiang, Yong-Liang Yang, Tianjia Shao, and Kun Zhou. 2021. Eyelashnet: A dataset and a baseline method for eyelash matting. *ACM Transactions on Graphics (TOG)* 40, 6 (2021), 1–17.
- Yuliang Xiu, Jinlong Yang, Dimitrios Tzionas, and Michael J. Black. 2022. ICON: Implicit Clothed humans Obtained from Normals. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 13296–13306.
- Zexiang Xu, Hsiang-Tao Wu, Lvdi Wang, Changxi Zheng, Xin Tong, and Yue Qi. 2014. Dynamic hair capture using spacetime optimization. *ACM Transactions on Graphics (TOG)* 33 (2014), 6.
- Zhaoyi Yan, Yuchen Yuan, Wangmeng Zuo, Xiao Tan, Yezhen Wang, Shilei Wen, and Errui Ding. 2019. Perspective-guided convolution networks for crowd counting. In *Proceedings of the IEEE/CVF international conference on computer vision*. 952–961.
- Haotian Yang, Hao Zhu, Yanru Wang, Mingkai Huang, Qiu Shen, Ruigang Yang, and Xun Cao. 2020b. Facescape: a large-scale high quality 3d face dataset and detailed riggable 3d face prediction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 601–610.
- Lingchen Yang, Zefeng Shi, Wu Yiqian, Xiang Li, Kun Zhou, Hongbo Fu, and Youyi Zheng. 2020a. iOrthoPredictor: model-guided deep prediction of teeth alignment. *ACM Transactions on Graphics* 39, 6 (2020), 216.
- Lingchen Yang, Zefeng Shi, Youyi Zheng, and Kun Zhou. 2019. Dynamic hair modeling from monocular videos using deep neural networks. *ACM Transactions on Graphics (TOG)* 38, 6 (2019), 1–12.
- Congyi Zhang, Mohamed Elgharib, Gereon Fox, Min Gu, Christian Theobalt, and Wenping Wang. 2022. An Implicit Parametric Morphable Dental Model. *ACM Transactions on Graphics (TOG)* 41, 6 (2022), 1–13.
- Meng Zhang, Menglei Chai, Hongzhi Wu, Hao Yang, and Kun Zhou. 2017. A data-driven approach to four-view image-based hair modeling. *ACM Trans. Graph.* 36, 4 (2017), 156–1.
- Meng Zhang, Pan Wu, Hongzhi Wu, Yanlin Weng, Youyi Zheng, and Kun Zhou. 2018. Modeling hair from an rgb-d camera. *ACM Transactions on Graphics (TOG)* 37, 6 (2018).
- Meng Zhang and Youyi Zheng. 2019. Hair-GAN: Recovering 3D hair structure from a single image using generative adversarial networks. *Visual Informatics* 3, 2 (2019), 102–112.
- Yujian Zheng, Zirong Jin, Moran Li, Haibin Huang, Chongyang Ma, Shuguang Cui, and Xiaoguang Han. 2023. Hairstep: Transfer synthetic to real using strand and depth maps for single-view 3d hair modeling. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 12726–12735.
- Zerong Zheng, Tao Yu, Yebin Liu, and Qionghai Dai. 2021. Pamir: Parametric model-conditioned implicit representation for image-based human reconstruction. *IEEE transactions on pattern analysis and machine intelligence* 44, 6 (2021), 3170–3184.
- Yi Zhou, Liwen Hu, Jun Xing, Weikai Chen, Han-Wei Kung, Xin Tong, and Hao Li. 2018. Hairnet: Single-view hair reconstruction using convolutional neural networks. In *Proceedings of the European Conference on Computer Vision (ECCV)*. 235–251.
- Yongning Zhu and Robert Bridson. 2005. Animating sand as a fluid. *ACM Transactions on Graphics (TOG)* 24, 3 (2005), 965–972.
- Wojciech Zielonka, Timo Bolkart, and Justus Thies. 2022. Towards metrical reconstruction of human faces. In *Computer Vision—ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XIII*. Springer, 250–269.
- Gaspard Zoss, Thabo Beeler, Markus Gross, and Derek Bradley. 2019. Accurate markerless jaw tracking for facial performance capture. *ACM Transactions on Graphics (TOG)* 38, 4 (2019), 1–8.
- Gaspard Zoss, Derek Bradley, Pascal Bérard, and Thabo Beeler. 2018. An empirical rig for jaw animation. *ACM Transactions on Graphics (TOG)* 37, 4 (2018), 1–12.

## A IMPLEMENTATION DETAILS

In this section, we describe the details of our networks and training settings in *RootFinder*, *OriPredictor* and *FiberEnder* experiments.

**RootFinder.** We employ the U-Net [Ronneberger et al. 2015] to predict the density map of the input orientation image with the input size  $1500 \times 600$ . Adam optimizer [Kingma and Ba 2014] is selected for training with the initial learning rate  $1e^{-5}$  and decayed by the factor of 0.01. The network is trained with a batch size of 4 for 300 epochs on a single NVIDIA RTX-3090Ti GPU for roughly two days.

**OriPredictor.** We use the Hourglass [Newell et al. 2016] network with four stacks with the input size  $1500 \times 600$  to extract local image features for learning the 3D orientation field. The layer sizes of MLP as the decoder are 277, 1024, 512, 256, 128, 3. Adam optimizer is used during training and the learning rate is set to  $1e^{-4}$  and decayed by the factor of 0.1 in the 60<sup>th</sup> epoch. The network is trained with a batch size of 4 for 100 epochs on a single NVIDIA RTX-3090Ti GPU for about 15 hours.

**FiberEnder.** *FiberEnder* shares similar 2D image features with *OriPredictor*, so we load a pre-trained checkpoint to facilitate training. Despite using only one stack of feature lists for both the training and testing phases due to limited computing resources, the performance of our model demonstrates high accuracy. All fibers of one eyebrow are passed into the network in a batch. As for the RNN encoder, the size of hidden code  $h$  is  $256 \times 1$ . The layer sizes of MLP as the decoder are 277, 1024, 512, 256, 128, 1. Adam optimizer is used during training and the learning rate is initialized to  $1e^{-4}$  at the beginning and decayed by the factor of 0.1 in the 50<sup>th</sup> epoch. It takes about two days to train for 100 epochs on a single NVIDIA RTX-3090Ti GPU. It is worth mentioning that the growing directions inferred by *OriPredictor* may not exactly align with the curves on the orientation maps during fiber synthesis. To narrow the domain gap between synthetic and real, we use fibers predicted by well-trained *OriPredictor* growing from ground-truth eyebrow roots to train *FiberEnder*. In other words, the labels used for loss calculation are actually pseudo-labels according to the length of the corresponding ground-truth fiber in the raw dataset.

## B LENGTH STATISTICS OF *EBSTORE*

Our *EBStore* dataset enables *EMS* for training and evaluation. To better demonstrate data preparation for *FiberEnder* experiments, we present length statistics for 400 models in *EBStore* as Tab. 6. All fibers created by artists contain an identical quantity of 20 points for standardization. We take the longest fiber as a reference, calculate the growing step  $\bar{s} = 0.014$  and regroup all fibers into different length levels according to  $\bar{s}$ , which are labeled as the ground truth during training. The length statistics for each level are presented in Tab. 6. The mean length of fiber is 0.0714 (level 5). Since most fiber length concentrates in the middle levels, using the mean length to end fiber growing as our baseline for comparison is fair.

## C DENSITY MAP VISUALIZATION

We show the intermediate density map results of the *RootFinder* module in Fig. 15. The visual results reveal that the predicted density

maps closely match the ground-truth density maps, enabling the accurate representation of diverse distributions of hair fiber roots.

## D MORE VISUALIZATIONS OF *EBSTORE* DATASET

We provide more visualizations of Facescape [Yang et al. 2020b] front images and *EBStore* geometries for males, females and older people in Fig. 16, Fig. 17 and Fig. 18, respectively. These examples comprehensively show the diversity in eyebrow density, shape and growing pattern between different genders and ages.

## E MORE VISUAL RESULTS OF ABLATION STUDY

We provide more visual results of the ablation study under four configurations: baseline of growing with random roots and a mean length, baseline with only *FiberEnder*, baseline with only *RootFinder*, and our *EMS* full model. As illustrated in Fig. 19, *EMS* outperforms the ablated settings, with all modules working together. *RootFinder* can capture the overall shape of an eyebrow, while *FiberEnder* modifies the boundary details.

Table 6. Length statistics of *EBStore* with step = 0.014.

Length level	1	2	3	4	5	6	7	8	9	10	11	≥ 12
Counts	881	4921	39884	143449	68229	50680	37236	22976	11159	5035	2341	2656
Percentages	0.23%	1.26%	10.24%	36.82%	17.51%	13.01%	9.56%	5.90%	2.86%	1.29%	0.60%	0.72%

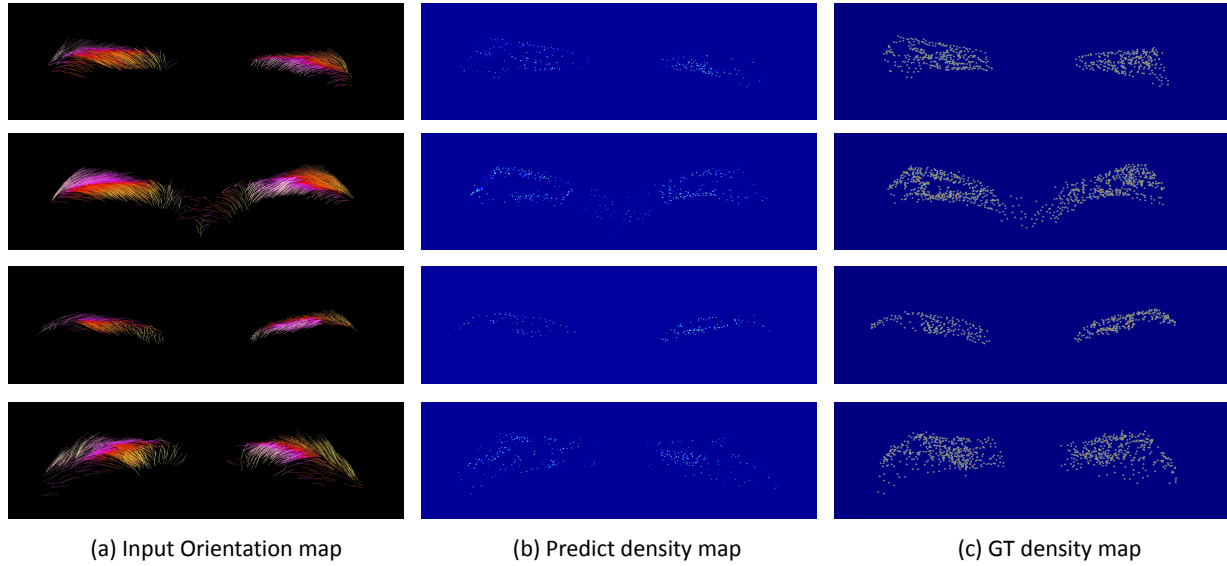


Fig. 15. Qualitative results of the *RootFinder* density map prediction. From left to right, we show the input orientation map, predict density map, and ground-truth density map, respectively.



Fig. 16. Male data samples of *EBStore* Dataset. More visualizations of Facescape front images (left) and corresponding geometries (right) for males.

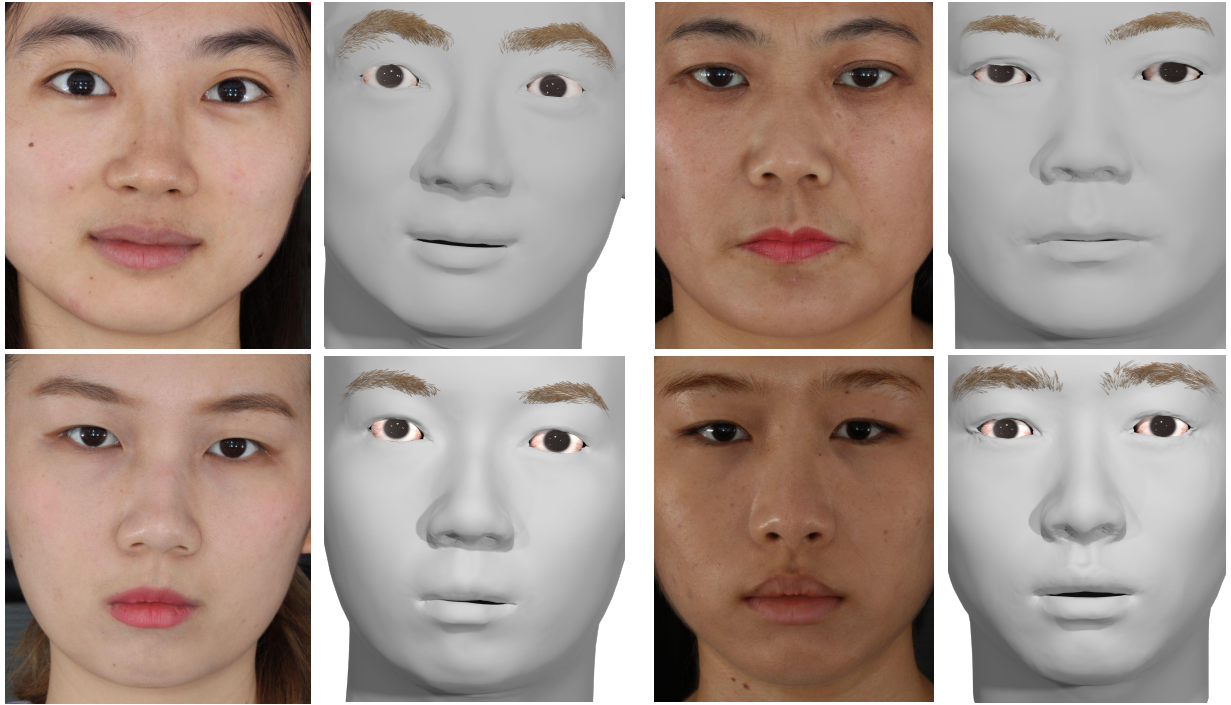


Fig. 17. **Female data samples of EBStore Dataset.** More visualizations of Facescape front images (left) and corresponding 3D geometries (right) for females without (the first row) and with (the second row) eyebrow make-up.



Fig. 18. **Data samples of older individuals in the EBStore Dataset.** More visualizations of Facescape front images (left) and corresponding 3D geometries (right) for old people.



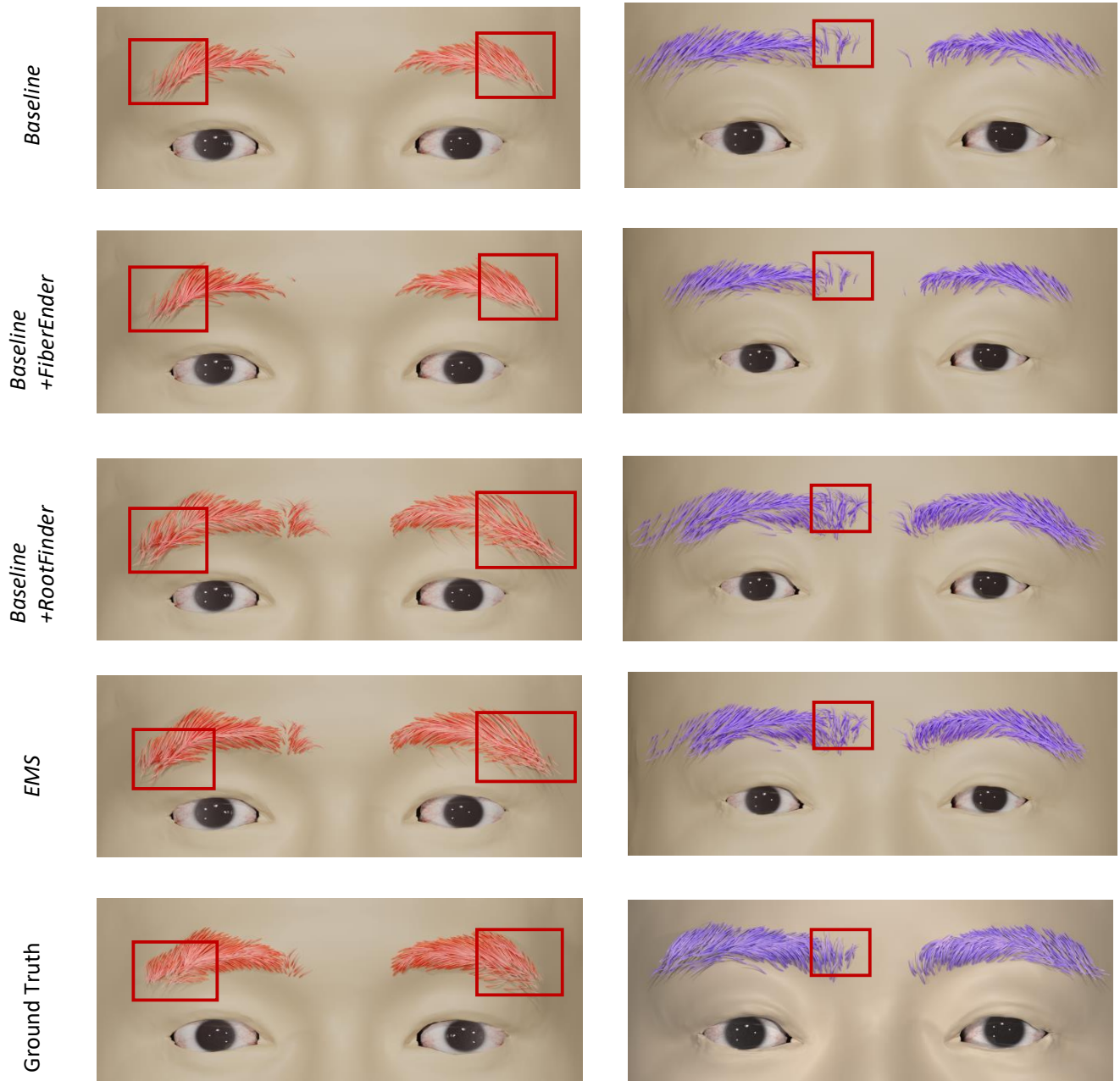


Fig. 19. Qualitative ablation study results. From top to bottom, we show the reconstructed 3D eyebrow models of the baseline, baseline with *FiberEnder* only, baseline with *RootFinder* only, our full model *EMS*, and ground truth, respectively. The different colors used have no particular meaning, just for fruitful illustration of different examples.