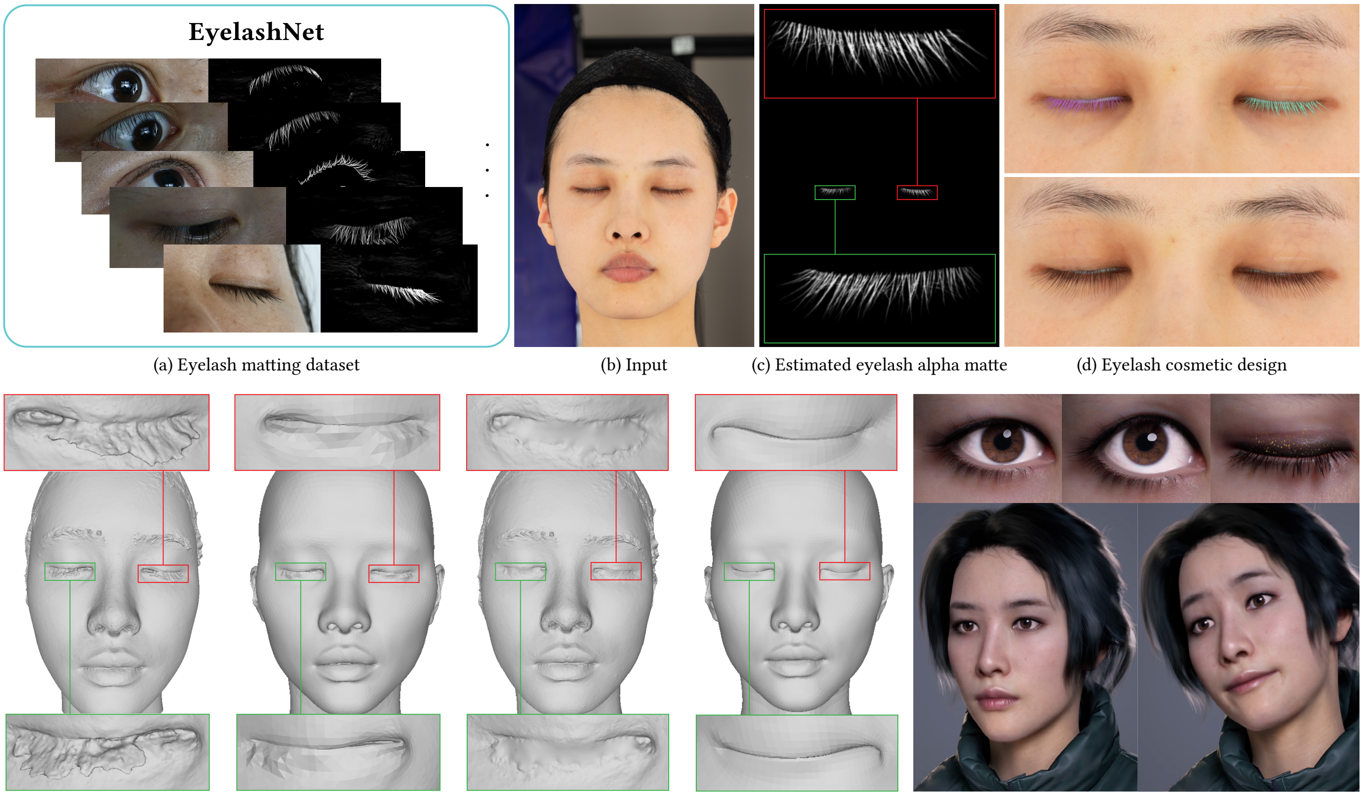
EyelashNet: A Dataset and A Baseline Method for Eyelash Matting

QINJIE XIAO, HANYUAN ZHANG, ZHAORUI ZHANG, YIQIAN WU, and LUYUAN WANG, State Key Lab of CAD&CG, Zhejiang University
XIAOGANG JIN*, State Key Lab of CAD&CG, Zhejiang University; ZJU-Tencent Game and Intelligent Graphics Innovation Technology Joint Lab
XINWEI JIANG, Tencent NExT Studios
YONG-LIANG YANG, University of Bath
TIANJIA SHAO and KUN ZHOU, State Key Lab of CAD&CG, Zhejiang University



(a) Eyelash matting dataset (b) Input (c) Estimated eyelash alpha matte (d) Eyelash cosmetic design

(e) Original (f) Parameterization of (e) (g) Eyelash removal (h) Parameterization of (g) (i) Avatar

Fig. 1. We create EyelashNet, the first eyelash matting dataset (a). This allows training a deep matting network that can automatically estimate high-quality eyelash alpha mattes (c) from an input portrait (b), where the alpha matte of the left/right eye is shown in the green/red box. In a high-fidelity avatar reconstruction pipeline, the eyelash alpha matting enables us to remove the interference of eyelashes during the multi-view stereo (MVS) based 3D face reconstruction process, and therefore largely enhances the efficacy and efficiency of the reconstruction of eye regions. Without eyelash removal, the reconstructed eyelash geometry (e) often induces noises and artifacts when fitting the eyelid during 3D parametric face reconstruction (f), which requires very expensive manual repair in hours. In contrast, eyelash matting helps to easily achieve a better geometry of the eye region (g). As a result, more faithful eyelids with much higher quality can be reconstructed. We show the fully rigged avatar in (i) for completeness. In addition, our eyelash alpha matting method can be applied for cosmetic design such as eyelash recoloring (d, top) and eyelash editing (e.g, lengthening the eyelashes) (d, bottom).

*Corresponding author.

Authors' addresses: Qinjie Xiao; Hanyuan Zhang; Zhaorui Zhang; Yiqian Wu; Luyuan Wang, State Key Lab of CAD&CG, Zhejiang University; Xiaogang Jin, State Key Lab of CAD&CG, Zhejiang University; ZJU-Tencent Game and Intelligent Graphics Innovation Technology Joint Lab; Xinwei Jiang, Tencent NExT Studios; Yong-Liang Yang, University of Bath; Tianjia Shao; Kun Zhou, State Key Lab of CAD&CG, Zhejiang University.

Eyelashes play a crucial part in the human facial structure and largely affect the facial attractiveness in modern cosmetic design. However, the appearance and structure of eyelashes can easily induce severe artifacts in high-fidelity multi-view 3D face reconstruction. Unfortunately it is highly challenging to remove eyelashes from portrait images using both traditional and learning-based matting methods due to the delicate nature of eyelashes and the lack of eyelash matting dataset. To this end, we present EyelashNet, the first eyelash matting dataset which contains 5,400 high-quality eyelash matting data captured from real world and 5,272 virtual eyelash matting data created by rendering avatars. Our work consists of a capture stage and an inference stage to automatically capture and annotate eyelashes instead of tedious manual efforts. The capture is based on a specifically-designed fluorescent labeling system. By coloring the eyelashes with a safe and invisible fluorescent substance, our system takes paired photos with colored and normal eyelashes by turning the equipped ultraviolet (UVA) flash on and off. We further correct the alignment between each pair of photos and use a novel alpha matte inference network to extract the eyelash alpha matte. As there is no prior eyelash dataset, we propose a progressive training strategy that progressively fuses captured eyelash data with virtual eyelash data to learn the latent semantics of real eyelashes. As a result, our method can accurately extract eyelash alpha mattes from fuzzy and self-shadow regions such as pupils, which is almost impossible by manual annotations. To validate the advantage of EyelashNet, we present a baseline method based on deep learning that achieves state-of-the-art eyelash matting performance with RGB portrait images as input. We also demonstrate that our work can largely benefit important real applications including high-fidelity personalized avatar and cosmetic design.

CCS Concepts: • **Computing methodologies → Computer graphics**.

Additional Key Words and Phrases: eyelash matting, dataset, deep learning

ACM Reference Format:
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 Trans. Graph.* 40, 6, Article 217 (December 2021), 16 pages. https://doi.org/10.1145/3478513.3480540

1 INTRODUCTION

High-quality personalized digital humans are crucial to a wide range of applications, including virtual reality, movies, games, etc. To create a high fidelity human face, a common workflow in the modern CG industry for virtual human creation includes two steps: 1) reconstructing a 3D parametric face model from the raw scan; and 2) generating face rigs for the parametric model. While the latter step has attracted huge interest from the research community and high-quality face rigs now can be produced automatically [Garrido et al. 2016; Li et al. 2010, 2020; Ma et al. 2016; Song et al. 2020], the former step still suffers from tedious labor works, especially in the scenario of very high accuracy face reconstruction (e.g. the face geometry contains about 7 million vertices and 15 million faces). One key challenge comes from the eyelashes. As shown in Fig. 1 (e) and (f), with the existence of eyelashes, noticeable artifacts will occur on the eyelids on the parametric face model. As a result, an artist typically needs to take around 5 hours to repair these eyelids, which is extremely labor-intensive.

In this paper, we seek to eliminate the negative effects of eyelashes on high-accuracy parametric face reconstruction, by removing the eyelashes from images before reconstruction and adding them back as a post-processing step after reconstruction. However, eyelash removal is a very challenging task. Manually removing eyelashes is often labor-intensive and time-consuming, as eyelashes are very tiny and complex. It is very difficult to separate eyelashes from fuzzy and self-shadow backgrounds (e.g. pupils) with similar colors via manual labeling. Previous methods [Beeler et al. 2012; Bermano et al. 2015; Nam et al. 2019] utilize Gabor filters to remove eyelashes, but these methods cannot deal with images with various eye expressions, illuminations and shadows (e.g. eyelashes are covered by the shadow of other objects or, fall in the shadow area). A feasible solution is to perform image matting. However, current image matting methods [Aksoy et al. 2018; Li and Lu 2020; Lin et al. 2020; Qiao et al. 2020] may yield poor matting results due to the lack of eyelash matting dataset.

To solve the above problems, we introduce EyelashNet, the first high-quality eyelash matting dataset built from authentic captured photos, which enables automatic and accurate eyelash matting on internet images with different eye expressions, illuminations, and shadows, using a baseline network. Establishing such a dataset requires addressing the following challenges. First, existing methods for building matting datasets [Rhemann et al. 2009; Smith and Blinn 1996] extract detailed foreground objects with the help of bluescreening. However, as eyelashes are covering the eyeballs and eyelids, it is impossible to use bluescreening for eyelashes. Second, bluescreening based methods can compute the alpha matte by triangulation [Rhemann et al. 2009] using the photos of the foreground object on four single-colored backgrounds (i.e. black, red, green, and blue), but in the eyelash case, we are not capable of obtaining the alpha values in this way.

To tackle the first challenge, we propose a novel eyelash capture system to accurately extract the eyelashes from the image. The key insight of our capture system is, instead of marking the background with a pure color like the blue screen, we can conversely mark the foreground object, which is more feasible for eyelashes. To this end, we design a novel eyelash capture system which marks the subject's eyelashes using a harmless fluorescent substance [MOVA 2021]. The fluorescent substance is invisible normally (Fig. 3 (a)) and is only visible when exposed to a UVA flash (Fig. 3 (b)). By taking two photos under the same pose using the double shooting mode with the UVA flash on and off, we get a normal eyelash image and a colored eyelash image. After further image alignment with image warping, an accurate eyelash mask (Fig. 3 (f)) can be extracted by subtracting the two images (Fig. 3 (a, b)). To tackle the challenge of alpha matte computation, we introduce an alpha matte inference network that takes the extracted eyelash mask and corresponding image (Fig. 3 (f, a)) as input to infer the alpha values. However, training the neural network is also difficult, because there is no real training data with ground truth. Even with the help of a synthetic eyelash dataset, the performance is still limited by the covariate shift [Ioffe and Szegedy 2015] between the synthetic eyelash dataset and captured dataset. We propose a progressive training strategy to overcome this difficulty by reducing the covariate shift between the synthetic eyelash dataset and captured dataset. We first warm up the inference network with a synthetic eyelash dataset with ground truth by rendering avatars. After warming up, we carefully check and select the perceptually correct alpha matte results using perceptual selection (a weak labeling process). Then we add the selected data to the synthetic dataset to train the alpha matte inference network. We perform the selecting and training process iteratively. Such a simple strategy can quickly adapt the inference network to real eyelash data after 2 rounds of training. The trained network is able to compute high-quality alpha mattes from the extracted eyelash masks and corresponding images.

Based on the proposed capture system and alpha matte inference network, we obtain the EyelashNet dataset composed of 5,400 high-quality eyelash matting data, captured from 50 subjects with a variety of ages (18-30), genders, head poses (with yaw angles of -60°, -30°, 0°, 30°, and 60°, and pitch angles of -45°, 0°, and 45°), facial expressions and eye expressions (e.g., open eyes, close eyes, etc.). With such a captured eyelash matting dataset, we train a baseline network which can generate accurate eyelash matting from a single portrait photo. The network architecture is adopted from [Li and Lu 2020], which is a CNN-based encoder-decoder network with multiple guided contextual attention modules. Such a network enables automatic and accurate eyelash removal for high-quality parametric face model reconstruction, and generalizes well on in-the-wild images. To summarize, this paper makes the following contributions:

- To the best of our knowledge, we propose the first high-quality eyelash matting dataset built from authentic captured photos, which enables automatic and accurate eyelash matting on portrait images with different eye expressions, illuminations, and shadows. The EyelashNet dataset will be made publicly available.
- We design a novel eyelash capture system using a fluorescent substance, which can automatically extract accurate eyelashes from the captured portrait images. Our progressive training strategy can effectively infer the alpha mattes based on the extracted eyelashes.
- We validate the performance of EyelashNet dataset with a baseline network, and compare our results with state-of-the-art methods. Our results outperform others in both qualitative and quantitative comparison.

2 RELATED WORK

Matting dataset is the key component for deep-learning based image matting. Many datasets have been released to the community. Rhemann et al. [2009] present the first image matting benchmark, which intrigues further data-driven based matting methods. Xu et al. [2017] create the composition-1k dataset, a large-scale matting dataset based on image composition. Chen et al. [2018] create a human matting dataset containing 35,513 images and their corresponding alpha mattes, but the dataset is not publically available. Qiao et al. [2020] release the Dinstinctions-646, a matting dataset composed of 646 foreground images with manually annotated alpha mattes for encouraging future research on trimap-free image matting. These datasets focus on the overall structure of objects instead of local and detailed components such as eyelashes. Lee et al. [2020] create a face image dataset containing segmentation masks of facial attributes, but eyelash data is not supported. In summary, eyelashes have not been addressed by prior works and there is no eyelash dataset available so far. On the other hand, creating a high-quality eyelash matting dataset is rather challenging. Unlike existing methods [Rhemann et al. 2009; Smith and Blinn 1996; Xu et al. 2017] that can edit the background variations of static objects to obtain foreground alpha mattes with fine details, the human eyes are dynamic and the background of eyelashes such as eyelids and eyeballs cannot be easily edited, replaced, or removed. Moreover, it is tedious and time-consuming to manually annotate detailed eyelashes compared with other objects such as humans [Chen et al. 2018]. Although using virtual eyelashes is a feasible solution to create high-quality matting data containing fine details, virtual results require costly modelling and rendering expertise to appear realistic, while the quality gap still remains when comparing with real eyelashes. To this end, we develop a novel fluorescent labeling system to automatically capture accurate eyelash matting data without any human annotation. Based on it, we present EyelashNet, an eyelash matting dataset to effectively estimate the matte of eyelashes.

Image matting is a fundamental problem that has been actively studied in image processing. Affinity-based methods [Aksoy et al. 2017] that rely on pixel similarity metrics are proposed for image matting. Sampling-based methods [Feng et al. 2016] have been investigated as well. Sun et al. [2006] extract mattes using a pair of flash/no-flash images in a sufficiently distant background scene. However, the performance of these methods is not good enough for matting delicate objects such as eyelashes. Recent deep learning methods have achieved promising results on natural image matting [Cai et al. 2019; Forte and Pitié 2020; Li and Lu 2020; Lu et al. 2019] and human matting [Chen et al. 2018; Liu et al. 2020; Qiao et al. 2020]. Trimap based methods take an image and a trimap (foreground, background, and unknown regions are marked) as a prior to estimate the detailed alpha matte [Aksoy et al. 2017; Xu et al. 2017]. Li et al. [2020] present a guided content attention network to weaken the requirements of trimap, and achieve better results compared with the affinity-based method [Aksoy et al. 2017]. However, labelling accurate trimap of eyelashes requires tedious user interaction, which is not feasible in practice. Recently, fully automatic matting methods [Qiao et al. 2020; Zhang et al. 2019] are proposed to overcome the limitation of trimap inputs. Zhang et al. [2019] present a fusion CNN for automatic matting. Liu et al. [2020] present a coarse-to-fine framework for semantic human matting that overcomes the lack of detailed matting dataset by using the combination of large-scale data with coarse annotations and small-scale matting data with fine details. Qiao et al. [2020] reconcile advanced semantic information with low-level appearance cues to refine the foreground details. However, these methods cannot deal with mixed eyelashes and pupils. Background matting methods [Lin et al. 2020; Sengupta et al. 2020] have achieved outstanding performance to process high-resolution videos firmly in real time. However, their methods require a background image as input, which is hard to obtain for eyelash matting.

High-quality eye reconstruction plays an important role in face reconstruction. It has wide applications in the film and entertainment industry, and is highly anticipated by the academic and business sectors. Existing works have focused on various aspects of eyes, including eye shape/motion reconstruction and parameterization [Bermano et al. 2015; Li et al. 2017; Trutoiu et al. 2011; Wen et al. 2017b,a], eyelid wrinkle reconstruction [Cao et al. 2015],

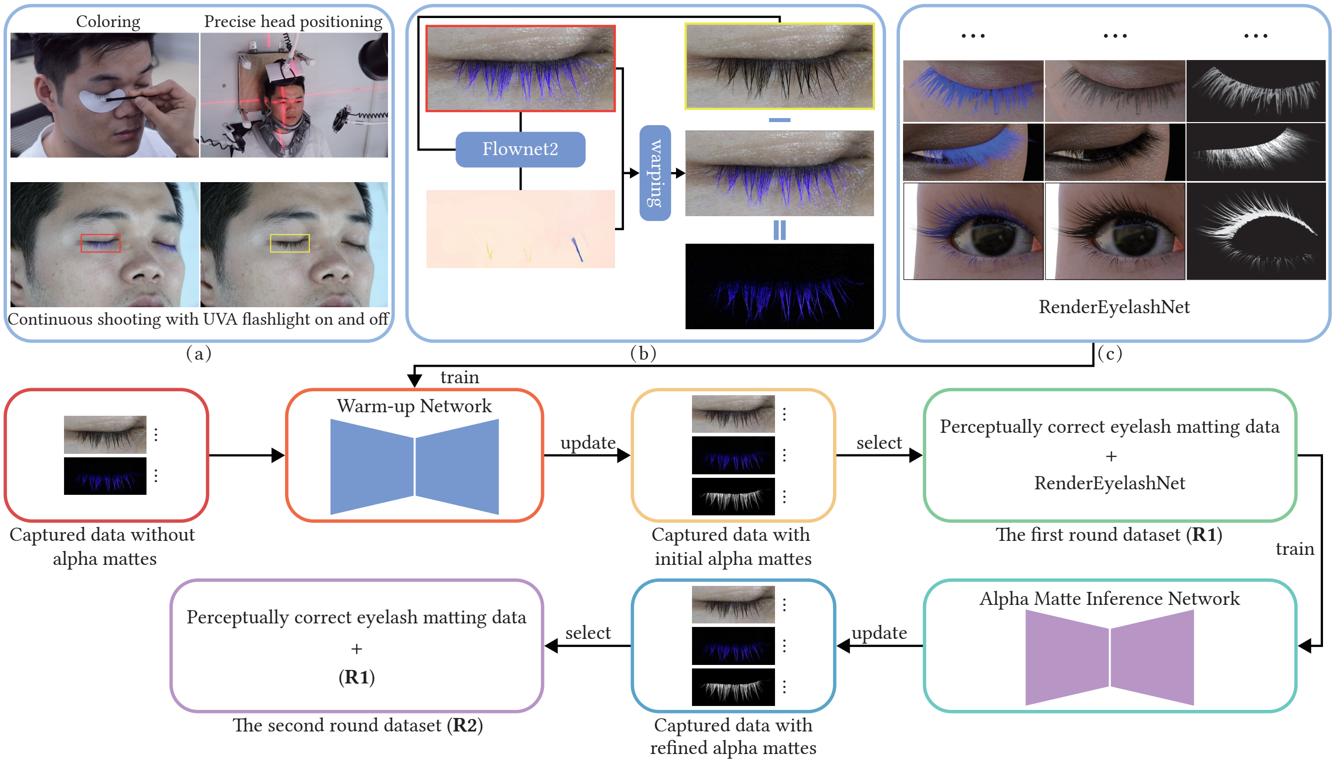

Coloring
Precise head positioning
Continuous shooting with UVA flashlight on and off
(a)
Flownet2
warping
(b)
RenderEyelashNet
(c)
train
Captured data without alpha mattes
Warm-up Network
update
Captured data with initial alpha mattes
select
Perceptually correct eyelash matting data + RenderEyelashNet
The first round dataset (R1)
train
Alpha Matte Inference Network
update
Captured data with refined alpha mattes
select
Perceptually correct eyelash matting data + (R1)
The second round dataset (R2)

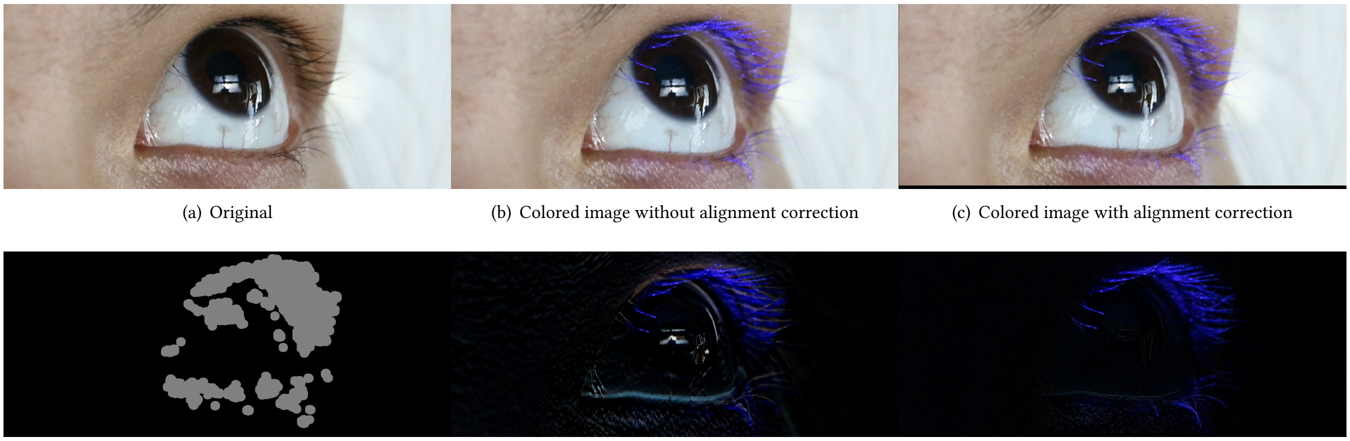
Fig. 2. An overview of our system, which consists of a capture stage (top) and an inference stage (bottom). (a) In the capture stage, we first color the subject's eyelashes with an invisible fluorescent substance, and precisely control the head position. Each camera focuses on the left and the right eye of the subject, and takes two photos with colored and normal eyelashes using continuous shooting mode with a UVA flash on and off. (b) Then we correct the alignment of these two captured photos based on a warp field estimated from a learnt FlowNet2 model, and perform subtraction to get an eyelash mask that can be used as a prior input of our baseline network to generate the refined eyelash alpha matte. (c) To complement real captured data, we also construct RenderEyelashNet, an eyelash matting video dataset created by rendering virtual avatars. In the inference stage, we employ a progressive training strategy that leverages the advantages of synthetic and captured data to achieve enhanced eyelash matting performance.

photo-realistic gaze and eye contact reconstruction [Kim et al. 2019; Nair et al. 2020; Schwartz et al. 2020; Wang et al. 2016; Whitmire et al. 2016; Wood et al. 2016], eye editing for portrait images [Shu et al. 2017], and retinal imaging [Huang et al. 2014; Swedish et al. 2015]. However, few works pay attention to eyelashes. Menon et al. [2020] recover eyelashes from images with very low resolution. GAN-based methods [Choi et al. 2020; Liu et al. 2016] enable eye style transfer, but cannot support high-quality eyelash manipulations. Beeler et al. [2012] and Nam et al. [2019] utilize Gabor filters to extract eyelashes for hair reconstruction. Nevertheless, the performance is limited when dealing with varying head poses, illuminations, and eye expressions. This refers to the fundamental problem of eyelash matting, which is mainly due to the lack of feasible eyelash matting solutions and the related eyelash matting dataset. Our work aims to fill this gap. We first present a novel fluorescent labeling system to automatically capture high-quality eyelash matting data. We also propose a deep-learning based method for effective eyelash matting.

3 OVERVIEW

In this work, we propose EyelashNet, the first eyelash matting dataset, produced with a novel fluorescent-based capture system for eyelash extraction and a novel alpha matte inference network for alpha matte computation. Thanks to the EyelashNet dataset, we can train a baseline network to achieve accurate eyelash matting on internet images. Fig. 2 gives an overview of our system, which is composed of a capture stage and an inference stage. During the capture stage (Section 4), we first brush the subject's eyelashes with an invisible fluorescent substance. While the subject sits still, the cameras focus on her/his left and right eye respectively to capture the raw data. For each eye, our system takes two photos of eyelashes using the continuous shooting mode with the UVA flash on and off. The fluorescent substance is activated and appears purple when the UVA flash is on, otherwise it is invisible. Given the two captured photos with the normal eyelashes and purple eyelashes, the accurate eyelash mask can be naturally obtained by image subtraction. However, as the eyelashes are very tiny and the subject is not able to sit perfectly still, direct subtraction on the two continuously captured photos will produce unneglectable errors (as shown in Fig. 3 (e)). Hence we align the two photos with an estimated optical flow first and then perform subtraction on the aligned photo pairs.

During the inference stage (Section 4.3), the alpha matte inference network takes as input the extracted eyelash mask as well as the


(a) Original

(b) Colored image without alignment correction

(c) Colored image with alignment correction

(d) Trimap

(e) Eyelash mask of (b) and (a)

(f) Eyelash mask of (c) and (a)

Fig. 3. Exemplars of the captured data, alignment correction, and the corresponding trimap. We present the original image (a), the colored image without (b) and with (c) alignment correction, a trimap example (d), and the differences between the colored and original images without (e) and with alignment correction (f), respectively.

corresponding eyelash image, and infers the alpha matte values for the eyelash. As there is no real training data with ground truth to supervise the network training, we introduce a progressive training strategy. The inference network is firstly warmed up (noted as warm-up network) with RenderEyelashNet (Section 6.2), a rendered eyelash dataset. Afterwards, we carefully check the alpha matte results on the captured eyelash images, and pick the perceptually correct results and add them to the training data to run the next round of training on the alpha matte inference network. The inference network can be quickly adapted to real eyelash data after 2 rounds of training, and yields high-quality alpha mattes from the eyelash masks and corresponding images.

Based on the proposed capture system and alpha matte inference network, we create an eyelash matting dataset named EyelashNet (Section 6). With EyelashNet as training data, we are able to train a baseline network for fully automatic eyelash matting from a single image. We demonstrate the qualitative and quantitative evaluations of the proposed eyelash matting dataset in Section 7.

4 EYELASH MATTING DATA GENERATION

In this section, we first present the system design principles for generating the eyelash matting data, and analyze the practical challenges. Then we elaborate on the capturing system in detail. Finally, we present the structure of our eyelash alpha matte inference network, and how we apply a progressive training strategy to enhance the matting performance.

4.1 Design Principles

Mathematically, the natural image $\boldsymbol{I}$ is defined as a convex combination of foreground image $\boldsymbol{F}$ and background image $\boldsymbol{B}$ at each pixel j as:

$$\boldsymbol{I}_j = \alpha_j \cdot \boldsymbol{F}_j + (1 - \alpha_j) \cdot \boldsymbol{B}_j, \alpha_j \in [0, 1], \tag{1}$$

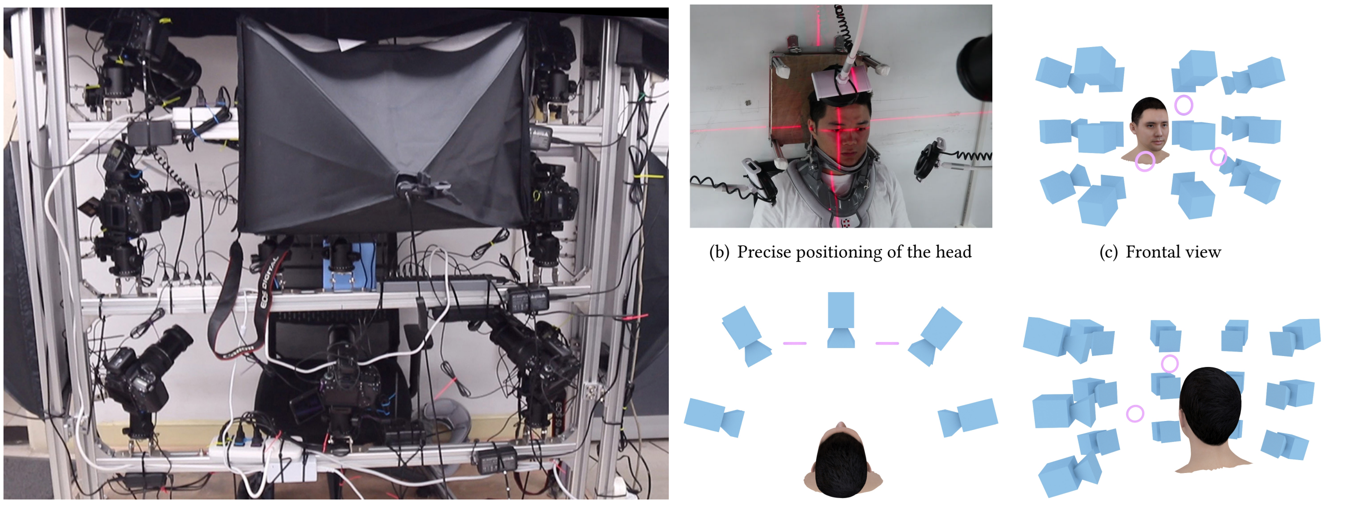
where α_j is the alpha value at pixel j that denotes the opacity of the foreground object. If α_j is not 0 or 1, then the image at pixel j is mixed. Alpha matting is an ill-defined problem since the foreground color $\boldsymbol{F}_j$, background color $\boldsymbol{B}_j$, and the alpha value α_j are unknown.

Prior work [Rhemann et al. 2009; Smith and Blinn 1996] obtains the alpha matte of a still object by laying it over at least three solid color backgrounds. However, these methods are infeasible for eyelashes as eyelashes are covering the eyelids and eyeballs. To this end, we propose to use two strictly aligned eyelash photos with different colored eyelashes to extract the alpha matte of eyelashes. Let $\boldsymbol{I}^c$ be the colored eyelash image, and $\boldsymbol{I}$ be the normal eyelash image that is strictly aligned with $\boldsymbol{I}^c$ in head pose and illumination. According to Equation 1, the alpha value at each pixel j can be formulated as:

$$\boldsymbol{I}^c_{j,k} - \boldsymbol{I}_{j,k} = \alpha_j \cdot (\boldsymbol{F}^c_{j,k} - \boldsymbol{F}_{j,k}), \tag{2}$$

where $\boldsymbol{I}^c_{j,k}, \boldsymbol{I}_{j,k}, \boldsymbol{F}^c_{j,k}, \boldsymbol{F}_{j,k}$ are the value of pixel j at channel $k \in \{R, G, B\}$ in the colored image $\boldsymbol{I}^c$, the normal image $\boldsymbol{I}$, the colored foreground $\boldsymbol{F}^c$ and the normal foreground $\boldsymbol{F}$, respectively. $\boldsymbol{F}^c, \boldsymbol{F}$ are usually unknown and related to complicated shading equation. Therefore solving α_j is a complex nonlinear problem. As we cannot use the triangulation method [Smith and Blinn 1996] with multiple single-color backgrounds to compute the alpha values, in this work we conduct a matting network to perform the alpha matte estimation.

The above design allows us to extract eyelash alpha mattes from the captured image pair $\boldsymbol{I}^c_{j,k}, \boldsymbol{I}_{j,k}$. However, it is rather challenging to obtain such an image pair, since the eyelashes and the corresponding backgrounds (skins, eyeballs, etc.) are dynamic over time and challenging to keep strictly still. This requires us to add a single color on eyelashes and take two photos with and without that color in a short period of time (0.2 second or less), such that the normal



(a) Acquisition device

(b) Precise positioning of the head

(c) Frontal view

(d) Up view

(e) Back view

Fig. 4. Our acquisition system contains 15 cameras and 3 UVA flashes (a). To control the head position precisely, we utilize a cross laser to guide a subject's head in the camera focus regions and a medical neck immobilizer to restrict head poses (b). In the second row (from left to right), we show the frontal view (c), up view (d), and back view (e) of the device layout, respectively. Cameras are illustrated with blue icons, and UVA flashes are illustrated with purple rings.

and colored eyelash images are well aligned. However, it is not possible to complete the whole process in such a short time under a normal camera and (environment) lighting setup.

Inspired by the fluorescence response where a fluorescent substance can present different colors under different lighting conditions (such as visible light, UVA light, etc.), we design a fluorescent labeling system that can capture two different visual states of eyelashes in a very short time. Since the normal eyelash images are required in our method, we use an invisible fluorescent substance, a type of pigment that can only be highlighted under the UVA light and is invisible under the visible light. We further design a capture device consisting of a camera and a specifically configured UVA flash, allowing the continuous shooting of two photos. Specifically, the UVA flash is turned on and off immediately to capture two continuous images with and without colored eyelashes. Fig. 3 shows an example of captured eyelash images.

4.2 The Capturing System

The capturing system is designed to extract eyelash masks by capturing a pair of images with normal and colored eyelashes. To generate high-quality eyelash masks, two captured photos should be strictly aligned with each other to eliminate the effect of the pose, expression, and illumination variations, etc. Besides, how to apply the invisible fluorescent substance and the UVA light device should be carefully taken into consideration. The eyelashes should be accurately and evenly colored with special care. We design a multi-camera data capturing system as shown in Fig. 2 (a) to achieve the above goals. For a subject, we first brush her/his eyelashes using an invisible fluorescent substance. Then we locate the head of the subject at the focus area of the cameras. Then we take the colored images $\boldsymbol{I}^c$ and the original images $\boldsymbol{I}$ of the subject with the UVA light on and off from different views. During the capturing process, the head is kept still.

Eyelash coloring. The subject's left and right eyelashes are colored with an invisible fluorescent substance, resulting in a colored appearance under the UVA light and a normal appearance under the visible light. Other physical properties such as viscosity, gloss, hardness, etc. are almost preserved. The visibility of the fluorescent substance under UVA light should be appropriate, neither too strong that leads to halo effects, nor too weak that cannot be observed. Besides, the fluorescent substance should be harmless to the eyes, skins, eyelashes, etc. According to the official instructions and the *Safety Data Sheet*[1] of *Noris 110UV ink* [Noris 2021] and after several attempts (on mannequins first, and then human subjects), it turns out that *Noris 110UV ink* is a good substance to meet our requirements. A low-power 365nm UVA light is suitable to capture the eyelashes accordingly. For hygiene reasons, we use disposable eyelash brushes to color the eyelashes of each subject. During the capturing process, in the rare case where the substance accidentally enters the eyes, we use normal saline to quickly wash the eyes. We cancel the capture process if the subject feels uncomfortable when touching the eyelashes.

Acquisition device. Our multi-view capturing system contains 15 cameras, 3 UVA flashes, and 3 photographic lamps. Fig. 4 (a) shows the setup of the capturing system. Specifically, each of the UVA flash is placed about 0.5m away from the subject's eyes, and consists of 48 ultraviolet LED lamps with the power of 0.06 Watt, which is much lower than the standard of International Electrotechnical Commission [Commission [n.d.]], so it is harmless to the subject. We provide additional analysis for the UVA flashes in the

[1] http://www.norisusa.com/110UV_Blue_MSDS_Noris_Ink.pdf

supplementary materials to further alleviate potential security concerns. To prevent the visible light emitted by UVA flashes which may change the environment illumination, we utilize 365nm ultraviolet transmittance filter to suppress visible light as much as possible. We set the cameras to focus on a specified region and then fix their focal lengths, and use wireless shutters to ensure efficient capture of the eyelash data. The cameras are set on the continuous shooting mode to take two photos in about 1/7 second (twice of the continuous shooting speed of *Canon 80D* cameras). An additional camera is set on the single-shot mode to turn the UVA flash on in the first shot, and off in the second shot during the capture process. Since moving cameras will reduce the capturing efficiency dramatically, we place the cameras in fixed positions.

Capture environment. The 16 cameras are synchronized based on the wireless shutter, among which 15 cameras are used to capture data under fast burst shooting, and the remaining camera is used to trigger the UVA flashes. To control the illumination, we cover the acquisition device with a blackout cloth to create a darkroom-like environment and control the light with 3 photographic lamps. We add filters to the UVA flashes to reduce the intensity of visible light to minimize the color differences between the two photos and the stimulated response of the eyes to the flashes.

Precise positioning of the head. For those cameras on the side, it is impossible to acquire clear photos of both eyes because they do not appear on the same focus plane. Hence we capture the left eye and the right eye separately. We precisely control the subject's head located in the camera focus regions. We use an adjustable chair to control the upper and lower body, and a cervical vertebra tractor to control the head orientation of the subject. We put adjustable boards behind the subject's head to control the front and rear positions. Finally, we slightly adjust the subject's body position to align the pupil of the subject's right/left eye with the center of the medical positioning laser (as shown in Fig. 4 (b)). We require the subject to remain as still as she/he can during one continuous shooting. Even so, the involuntary movement of the subject still affects the result. We use a medical neck immobilizer to further restrict head poses.

After the head is carefully positioned, we take the colored images $\boldsymbol{I}^c$ and the normal images $\boldsymbol{I}$ of the subject with the UVA light on and off. The difference between the colored image and normal image is obtained as follows:

$$D(\boldsymbol{I}^c, \boldsymbol{I}) = \max(\boldsymbol{I}^c - \boldsymbol{I}, 0), \tag{3}$$

where $\max(\cdot, \cdot)$ is the pixel-wise maximum operation. Fig. 3 (a, b, e) shows an example of the colored images, the original images, and the differences noted as eyelash mask between them, respectively. In total, we capture 12 facial and eye expressions of the subject from 15 views. On average, the whole capturing process takes about 15 minutes to apply the invisible substance, 5 minutes for head positioning, 10 minutes for capturing data, and 5 minutes to remove the fluorescent substance.

Alignment correction. Although our system is carefully designed to allow accurate eyelash mask estimation, in practice it is extremely difficult to avoid subtle pose and expression changes of the subject during the capturing process. This often causes misalignment between paired eyelash images and results in erroneous eyelash masks. To resolve this, we employ image warping [Shih et al. 2019] to correct the misalignment between paired images. The image warping is guided by the optical flow between the normal and colored eyelash image pairs estimated by FlowNet2 [Ilg et al. 2017]. Specially, we fine-tune the FlowNet2 model on an optical flow dataset created from RenderEyelashNet (Section 6.2). The optical flow dataset consists of continuously rendered normal and colored image sequences with affine transformations being applied between each pair of image frames [Dosovitskiy et al. 2015]. We treat the data as a failure case with noticeable noise in the eyelash masks before and after alignment correction (as shown in Fig. 5).

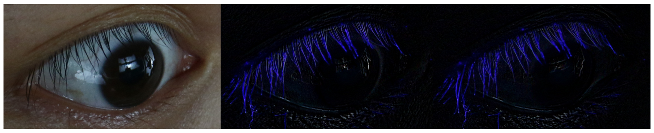


Fig. 5. An exemplar of the optical flow alignment's failure case.

4.3 Alpha Matte Inference Network

Inspired by [Li and Lu 2020], we construct an inference network to estimate the alpha matte from the extracted eyelash mask and the corresponding eyelash image. To make the paper more self-contained, we first demonstrate the network structure of [Li and Lu 2020]. Then we describe the details of our inference network. Finally, we present the progressive training strategy that iteratively adapts the network to the real eyelash data.

The network of [Li and Lu 2020] is a U-net like structure [Ronneberger et al. 2015] , i.e., an encoder-decoder network containing stacked residual blocks [He et al. 2016]. The encoder contains 3 *conv* layers and 4 residual blocks, and the decoder consists of 4 residual blocks, 1 *deconv* layer and 1 *conv* layer. Five shortcut layers are used to build the skip connections, which provide lower-level features to the estimated alpha matte. Two guided contextual attention (GCA) modules extract the similarity information from the low-level image features to refine the alpha features. The input to their network is an RGB image and a trimap. In their implementation, a trimap is an 8-bit image which defines three regions: a definite foreground (with value 255), a definite background (with value 0), and an unknown region (with value 127). In the alpha matte inference network (Section 4.3), the trimap will be replaced by an eyelash mask (as shown in Fig. 3(f)). In the baseline network (Section 5), the trimap will be replaced by a trimap with all pixels set to 127. The output is an alpha matte estimation of the foreground in the unknown regions. Their network leverages one alpha prediction loss defined as the average difference between the ground truth and the estimated alpha matte over the unknown regions:

$$L_{MAE} = \frac{1}{|\boldsymbol{U}|} \sum_{i \in \boldsymbol{U}} |\hat{\alpha}_i - \alpha_i|, \tag{4}$$

where $\boldsymbol{U}$ represents the unknown regions indicated in the trimap, $\hat{\alpha}_i$ and α_i are the estimated and ground truth values of the alpha matte at position i. For the image feature $\boldsymbol{I}_{gca}$ in the GCA module, they extract 3×3 patches from the whole image feature. For a patch $\boldsymbol{P}_{x,y}$ centered on (x, y), they define an attention score for the patch

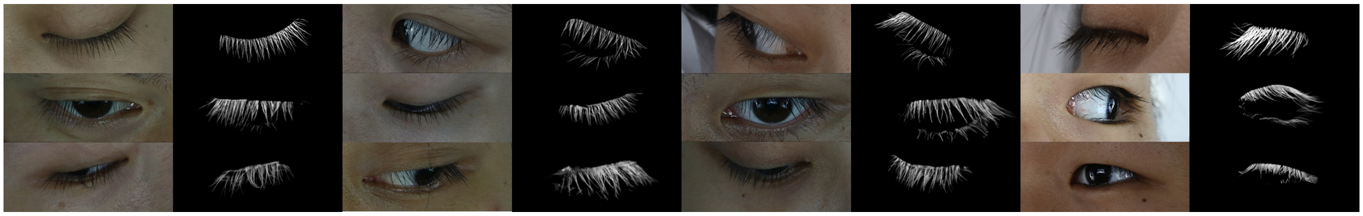


Fig. 6. Exemplars of perceptually reasonable (first row), borderline (second row), and unreasonable cases (third row), respectively.

based on the trimap as follows:

$$w(\boldsymbol{U},\boldsymbol{K},x,y) = \begin{cases} clamp(\sqrt{\frac{|\boldsymbol{U}|}{|\boldsymbol{K}|}}), & \boldsymbol{P}_{x,y} \in \boldsymbol{U}; \\ clamp(\sqrt{\frac{|\boldsymbol{K}|}{|\boldsymbol{U}|}}), & \boldsymbol{P}_{x,y} \in \boldsymbol{K}, \end{cases} \tag{5}$$

$$clamp(w) = \min(\max(w, 0.1), 10), \tag{6}$$

where $\boldsymbol{K}$ represents the known regions. $clamp(\cdot)$ is a function that limits the value of attention score $w(\boldsymbol{U},\boldsymbol{K},x,y)$ within 0.1 to 10. The attention score is used to leverage the feature similarity between different patches. Please refer to [Li and Lu 2020] for more details on the network structure and the GCA module.

Our alpha matte inference network shares almost the same structure and setup as [Li and Lu 2020], except that the input, the attention score function and data augmentation (described in Section 7.1) are specifically designed to fit the format of our captured data. The input to the alpha matte inference network comprises an RGB image and an eyelash mask (as shown in Fig. 3(a, f)). We use the same objective function as in Li and Lu [2020]. The only difference is that we treat the whole input image region as the unknown region in Eq. 4.

More specifically, as the extracted eyelash mask (Eq. 3) contains floating values around the foreground boundaries instead of binary values, it provides more detailed information than the trimap used in Li and Lu [2020] to help the GCA module generate more effective attention map for the inference network. As a result, the inference network can estimate eyelashes in fuzzy regions (e.g., pupils) based on the eyelash mask prior. We first downsample the eyelash mask to the same size of the image feature $\boldsymbol{I}_{gca}$ in the GCA module. Then we let D^B be the blue channel of the down-sampled eyelash mask (which has the strongest response), and use it to compute the attention score for the patch $\boldsymbol{P}_{x,y}$ of $\boldsymbol{I}_{gca}$ as follows:

$$w_c(x,y) = \frac{1}{9}\sum_{i,j} D^B(i,j), i \in \{x-1,x,x+1\}, j \in \{y-1,y,y+1\}. \tag{7}$$

4.4 Progressive Training

Based on the alpha matte inference network, we adopt a progressive training strategy to improve the matting performance on the real captured eyelash data. Since our work aims at detailed matting results, high precision matting data are required for inferring the eyelash alpha matte. For our captured data, we select image pairs that are visually aligned well as the starting dataset of progressive training. During training, we first warm up the alpha matte inference network with the synthetic dataset RenderEyelashNet (Section 6.2). After warming up, we use the network to test the real captured data. Then, we carefully check and select the estimated alpha matte results through perceptual selection (described in the next paragraph). We treat the perceptually correct results as pseudo ground truth and add them back to the training data to run the next round of training. We denote the above merged dataset as the first round of eyelash dataset (R1). For the second round, we train our inference network on R1 to update the alpha mattes for all of the captured data (including the selected captured data in R1). Similarly, we select the perceptually correct results as pseudo ground truth and add them to R1 to run the next round of training (denoted as R2). For each round of training, the inference network follows the initialization and training setup described in Section 7.1. Such a simple strategy can quickly adapt the network to real eyelash data after 2 rounds of training. The trained network is able to yield accurate alpha mattes with the eyelash mask and corresponding image as input. Such a progressive training strategy helps achieve better eyelash matting performance (see Section 7.3 for qualitative and quantitative evaluations).

Perceptual selection. We do not come up with an automatic method as no sufficiently large labeled dataset is available for satisfactory classification. The compromise is a perceptual selection, a weak labeling process following the criterion that the alpha matte should cover almost all eyelashes but not non-eyelash areas. In practice, five raters are invited to observe and judge whether the matting data are qualified. We choose the data with a majority of qualifying scores. Each matting result takes about 5 seconds for a rater to observe and judge. We treat the eyelash matting data with defocus blur, observable unmasked eyelashes, or observable masked background, etc., as failure cases. Fig. 6 shows examples of perceptually reasonable, borderline, and unreasonable cases.

5 BASELINE NETWORK

We introduce a baseline network to infer the eyelash alpha matte using an RGB image only as input without any prior (e.g., eyelash mask, trimap, etc.). Our baseline network follows the same network structure and objective function (Eq. 4) as that of Li and Lu [2020] except that we set all the pixel values of the trimap to 127. We employ the data augmentation method described in Section 7.1 to increase the diversity of the captured data.

6 DATASET

The availability of image matting dataset [Qiao et al. 2020; Rhemann et al. 2009; Xu et al. 2017] has greatly enhanced the research on matting techniques. However, due to the lack of eyelash matting data, the performance of the existing learned model for eyelash matting is far from satisfactory (see Fig. 12). As such, eyelash matting data

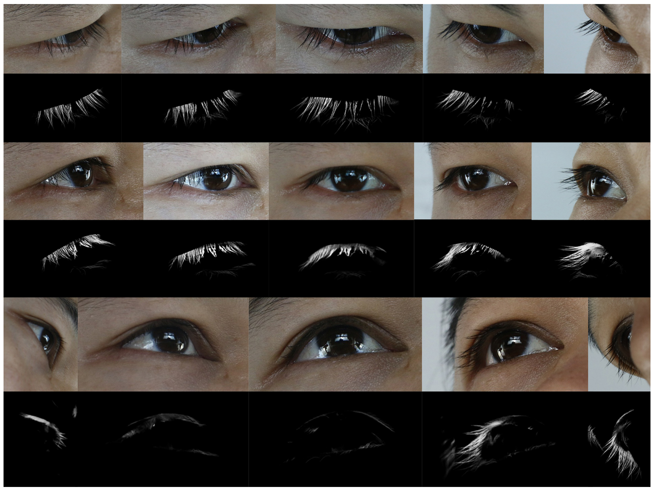


Fig. 7. An exemplar of captured data from 15 views. Rows 1, 3, and 5 are the captured photos of eyelashes, and rows 2, 4, and 6 are the corresponding alpha mattes. In each row (from left to right), we show views at head yaw angles of -60°, -30°, 0°, 30°, and 60°, respectively. In each column (from top to bottom), we show the vertical view (rows 1, 2), front view (rows 3, 4), and bottom view (rows 5, 6) of eyelashes, respectively.

are largely required to not only fill in the literature gap on image matting, but also benefit the downstream applications such as face performance capture, personalized avatar, eyelash editing, cosmetic design, etc.

In this section, we introduce EyelashNet, the first eyelash matting dataset generated with our system. As EyelashNet is captured within the lab environment with limited subjects which may not fully reflect the in-the-wild conditions (e.g., illuminations, skins, etc), we further augment EyelashNet with RenderEyelashNet generated by rendering avatars. Moreover, we build a baseline test dataset (noted as EyelashNet-Base) to evaluate the quality of EyelashNet.

6.1 EyelashNet

We collect the eyelashes of 50 Asians (25 males and 25 females) aging from 18 to 30 under the academic usage agreement. The capturing system's risks are evaluated and approved by an ethics board in advance. The participants gave informed consent before participating. We capture (from 15 views) the left and right eyelash images under 12 facial and eye expressions, including neutral, close eyes, frown, raise eyebrows, look left, look right, look up, look down, squeeze eyes, smile, simper, and anger, respectively (as shown in Fig. 8). Fig. 7 shows an example of the 15 multi-view eyelash images of an individual. From left to right, we capture the eyelash images with the views corresponding to -60°, -30°, 0°, 30°, 60° yaw angles of head pose, respectively. From top to bottom is the vertical view, front view, bottom view of eyelashes. We finally create EyelashNet, a high-quality eyelash matting dataset composed of 5,400 real captured data after progressive training. EyelashNet is split into a training dataset (4,860 captured data) and a test dataset (540 captured data).

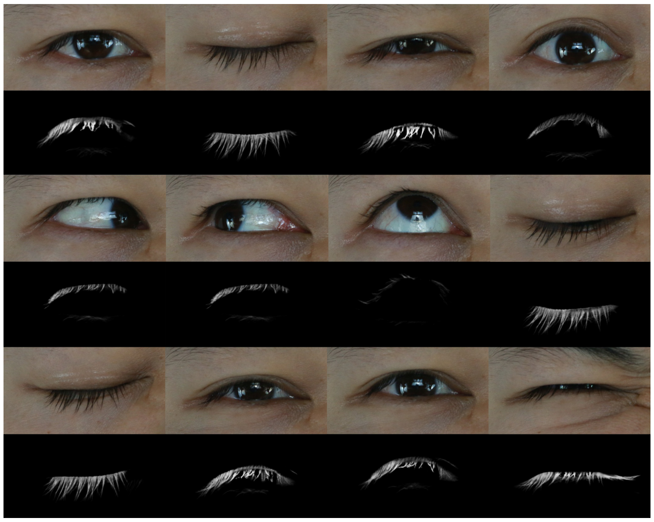


Fig. 8. An exemplar of captured data with 12 eye expressions. From left to right and top to bottom, the eye expressions are neutral, close eyes, frown, raise eyebrows, look left, look right, look up, look down, squeeze eyes, smile, simper, and anger, respectively.

6.2 RenderEyelashNet

Despite the visual difference from real eyelash images, there are clear advantages of synthetic eyelash images created by rendering avatars. Firstly, it is much cheaper to obtain accurate eyelash alpha mattes under variations of genders, head poses, races, illuminations, etc. In contrast, since the camera positions of our capturing system are fixed, the head poses of captured images are restricted. Also, the variations of the captured images are limited to the recruited subjects and the lab illumination condition. To complement the real captured data, we construct RenderEyelashNet, an eyelash matting video dataset created by rendering virtual avatars. The avatars are created and rendered using DAZ 3D [Inc 2021]. We use four avatars (one white woman, one black woman, one Asian man, and one white man) with pose and expression variations. We continuously change their head poses from -120° to 120° yaw angles, and -60° to 60° pitch angles. The facial and eye expressions are also changing smoothly in the videos. With such a setup, the RenderEyelashNet contains most common poses that may appear in common portrait photos. Moreover, we simulate the fluorescent capture system, and render additional fluorescent-like videos. Three weaker purple light sources are created to simulate the light noise caused by the visible part of UVA light (UV filters can not completely remove the visible light portion). We render part of the avatars with very small perturbations to simulate the subtle expressions and pose noises. We also render the background image (faces without eyelashes) of the corresponding foreground eyelashes. Fig. 9 shows several key frames of RenderEyelashNet. From top to bottom, we present the simulated colored images, the original image, the background, and the alpha mattes, respectively. Finally, we construct 5,272 alpha matting data instances with varying genders, races, poses, expressions, etc. RenderEyelashNet is split into a training dataset (4,965 rendered data) and a test dataset (307 rendered data).

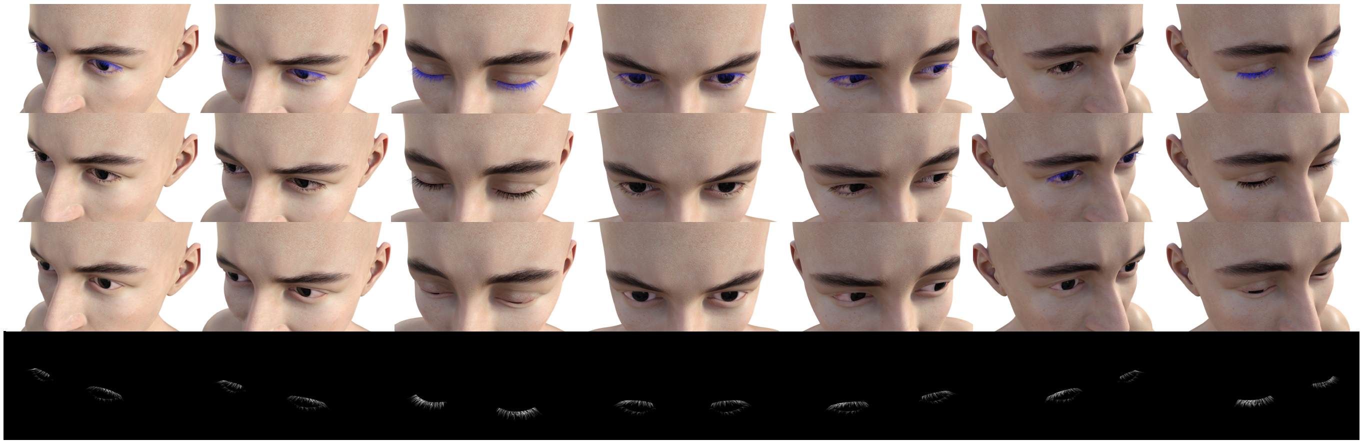


Fig. 9. Several key frames of eyelash matting data created by rendering videos of the avatar. The video contains pose and expression variations of the avatar. We render the fluorescent data of the avatar to assist EyelashNet generation and train a FlowNet2 [Ilg et al. 2017] model to correct the misaligned original and colored eyelash image pairs.

Note that for clarity, we present RenderEyelashNet and EyelashNet separately in this section. In practice, RenderEyelashNet is included as part of EyelashNet for data augmentation. Moreover, the augmented EyelashNet will be made publicly available.

7 EXPERIMENTS

In this section, we first present the implementation and data preparation details. Then we conduct ablation studies to validate the effectiveness of the data capture stage and the alpha matte inference stage, along with the rationality of our capturing system's eyelash coloring solution. After that, we qualitatively and quantitatively evaluate the performance of our baseline matting network on EyelashNet to demonstrate its effectiveness. We use four metrics for quantitative evaluation, including the mean square error (MSE), the sum of absolute error (SAD), the gradient (Grad), and the connectivity error (Conn) proposed in [Rhemann et al. 2009]. Finally, we apply our work to important practical applications including avatar generation, eyelash recoloring, and eyelash editing.

7.1 Implementation and Data Preparation Details

Network implementation details. We follow the training setups in [Li and Lu 2020]. Both the alpha matte inference network and the baseline network are initialized using the customized ResNet-34 [He et al. 2016] backbone (trained on ImageNet [Deng et al. 2009]) of [Li and Lu 2020] and trained for 160,000 epochs with batch size 4, taking about 16 hours. We use Adam optimizer with $\beta_1 = 0.5, \beta_2 = 0.99$. The learning rate is initialized to 4×10^{-4} and adaptively adjusted with warm-up and cosine decay [Goyal et al. 2018; He et al. 2019; Loshchilov and Hutter 2017]. We utilize Pytorch (version 1.2) [Paszke et al. 2019] to train the networks on a desktop PC with single NVIDIA GTX 1080 (8GB memory), Intel Xeon 3.6 GHz CPU, and 16GB RAM.

Data augmentation. Since EyelashNet is captured from a lab-based environment, the diversity of the captured data is limited to the experiment scene. Hence we further apply data augmentation to enhance our baseline model. Since the accuracy of Eyelashes is sensitive to the image quality, we use a Gaussian blur with a random kernel size from 5 to 30 to blur the foreground and the alpha matte with a probability of 0.2. Then a gamma correction (the value is randomly sampled from 0.5 to 1.4) is applied to adjust the illumination of eyelash images. Next, a random affine transformation is applied to the foreground image and the corresponding alpha matte image. We generate a random rotation, scaling, shearing as well as vertical and horizontal flipping in the affine transformation. After that, the foreground images are then converted into the HSV color space, and different jitters are imposed on the hue, saturation and value. For each eyelash foreground, we choose its background for data augmentation according to the following probability distribution. We use its original image directly (i.e., its original background) as input image with a probability of 0.5, and select an image randomly from EyelashNet and the MS COCO dataset [Lin et al. 2014] with a probability of 0.4 and 0.1, respectively, as the background for composition.

Baseline test dataset. Normally, a ground truth eyelash matting dataset is needed to evaluate the quality of EyelashNet. However, it is almost impossible to obtain the ground truth eyelash alpha mattes from the captured images. A compromise is to manually segment out the eyelashes as masks. But this is also a very labor-intensive and time-consuming work, even for a professional artist. In our work, it takes about 30 minutes for an artist to segment eyelashes from one captured image. We annotate 94 and 31 eyelash masks from captured images (noted as captured test dataset) and Internet images (noted as Internet test dataset), respectively, as the approximated ground truth. We use them as the baseline test dataset to evaluate the performance of our method for EyelashNet generation, so as the baseline network.

Device setup. In our capturing system, we use *Canon 80D* cameras with the continuous shooting mode (7 fps) to take normal and colored photos. Base on *Travor* LED flash (0.01 second exposure time), we replace the original LED lamps by *CZINELIGHT* LED

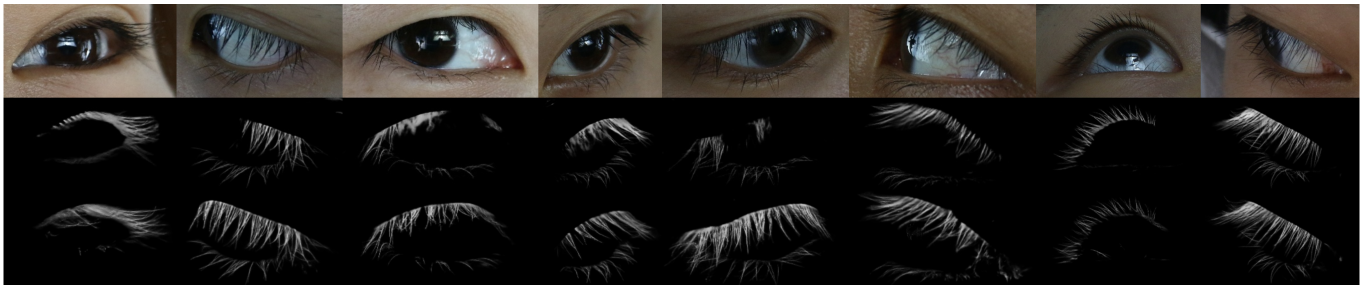


Fig. 10. From top to bottom, we show the original image, the estimated alpha mattes of the baseline network (takes an RGB image as input), the inference network (takes a eyelash mask and an RGB image as input), respectively.

Table 1. The quantitative results on MSE, SAD, Grad, Conn metrics of 3 stages in progressive training.

	MSE			SAD			Grad			Conn		
	R0	R1	R2	R0	R1	R2	R0	R1	R2	R0	R1	R2
Captured test dataset	0.00389	0.00278	0.00242	3.32	2.48	2.30	3.68	2.84	2.54	1.47	1.18	1.11

lamps (0.06 Watt and 365nm). Moreover, we use a 365nm ultraviolet transmittance filter (UV pass filter type: ZWB2/UG1) to filter out as much visible light as possible. We use *MZlaser* medical positioning laser (power from 0.35 mW to 0.45 mW) for positioning the head, and the positioning process usually takes less than 2 seconds. We strongly advise that the subject closes his/her eyes and does not look at the light source directly to eliminate safety concerns. It is safe enough even if the medical positioning laser illuminates the subject's eye carelessly.

Comparison. We compare our results with a confidence map of Gabor filter that is used to extract hair in Nam et al. [2019]. As the source codes of the hair extraction method of Nam et al. [2019] are not open, we re-implement their hair extraction part with our own source codes by setting parameters: $ksize = 5, \lambda = 4, \sigma = 2.3$ for the Gabor filter. This may bring differences between our results and the results of their implementation. We open our re-implementation source codes in the supplemental material to make the comparisons more reasonable.

7.2 Ablation Study

We perform an ablation study to evaluate the efficacy of the eyelash mask (Fig. 3 (f)) and the alpha matte inference network. As the progressive training is based on RenderEyelashNet, we give preference to networks with better domain adaptation [Wang and Deng 2018] between RenderEyelashNet and the captured data. Thus we only need to evaluate the warm-up network's performance used in the progressive training. Since the trimap is usually unavailable for the eyelash images, we ablate the inference network with the baseline network, which takes an RGB image as input. We train the alpha matte inference network (takes an eyelash mask and an RGB image as input) and the baseline network on RenderEyelashNet, and test on 20 randomly sampled captured data, respectively. We perform a similar perceptual selection strategy as used in the progressive training. 20 raters were asked to vote the better eyelash alpha matte estimation between the test results of the inference network and the baseline network. In the end, there were 295 (73.75%) and 105 (26.25%) preferences from the inference network and the baseline network, respectively. Fig. 10 presents 8 random examples from the 20 captured data samples. The results of the inference network (Fig. 10, row 3) show better performance than those of the baseline network (Fig. 10, row 2), including the semantic completeness of eyelashes, the quality of eyelashes on pupil regions, etc. Using the eyelash mask further reduces the covariate shift between the synthetic eyelash dataset and the captured dataset.

7.3 Evaluation of EyelashNet Generation

The progressive training strategy is designed to progressively adapt the matting inference network to real captured data. In our work, we first train a warm up network on RenderEyelashNet, and then we perform two rounds of progressive training, resulting in two intermediate networks and datasets. We denote three different training datasets including RenderEyelashNet, the first and the second round of datasets as R0, R1, R2 for simplicity. Table 1 shows the quantitative results. Compared with the warm up network trained on RenderEyelashNet (R0), the mean square error (MSE), sum of absolute error (SAD), gradient error (Grad) and connectivity error (Conn) on the captured test dataset decrease gradually during two round (R1 and R2) of progressive training. The qualitative result in Fig. 11 also clearly demonstrates the effectiveness of the progressive training strategy.

We also conducted a user study to estimate the influence of the invisible fluorescent substance on the status of eyelashes. We randomly selected 30 normal eyelash images (painted with the fluorescent substance) from EyelashNet, and asked 20 raters to view these pictures and answer "Is there anything visible painted on the eyelashes?". 582 (97%) vote "NO", and 18 (3%) vote "YES", which shows that putting an invisible fluorescent substance on eyelashes has little impact on the eyelash appearance.

To evaluate the illumination influence of the UVA flash, we calculate the MAE of each pixel in the captured colored and normal

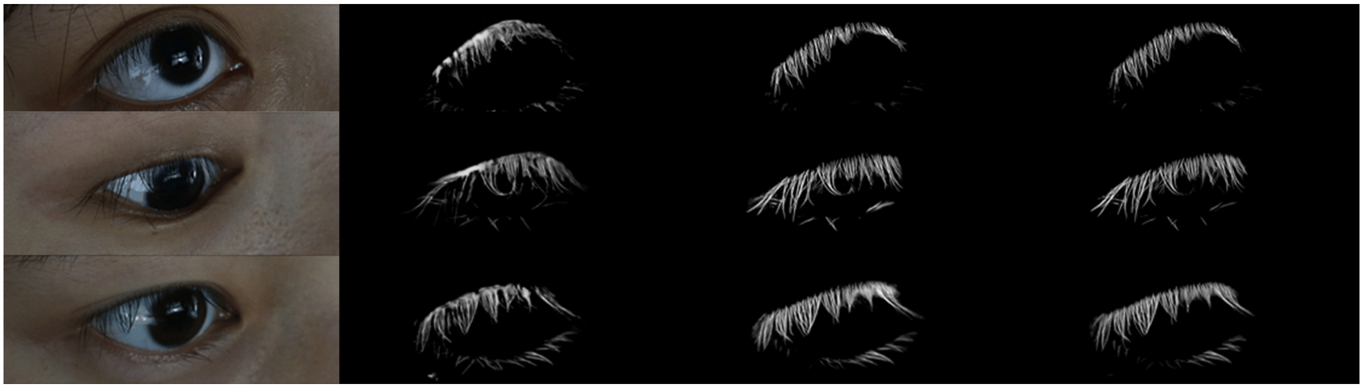
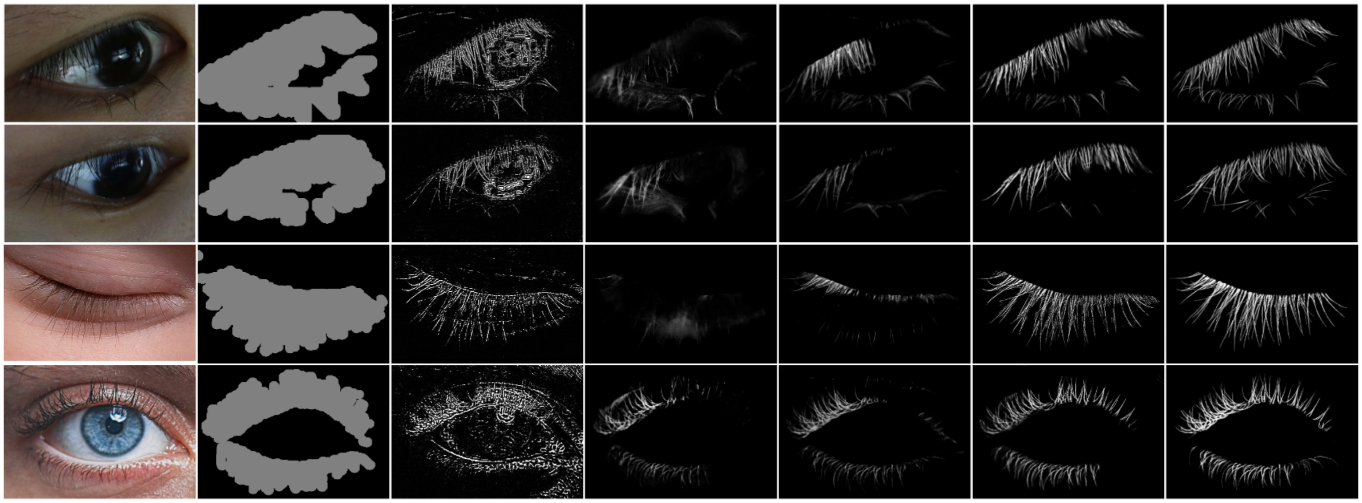


Fig. 11. Exemplars of progressive training. From left to right, we show the original image, the estimated alpha mattes of R0, R1, R2, respectively.

Table 2. The quantitative results of 4 methods (Nam et al. [2019], Li and Lu [2020], RenderEyelashNet and Ours (EyelashNet), respectively) applying on the captured test dataset and the Internet test dataset, respectively. Note that Li and Lu [2020] use a manually labeled trimap (b) as input while our method does not need such a time consuming labelling process. Even so, our approach still has better overall performance for eyelash matting.

	Captured test dataset				Internet test dataset			
	SAD	MSE	Grad	Conn	SAD	MSE	Grad	Conn
Nam et al. [2019]	6.88	0.01032	6.76	1.11	10.12	0.0205	20.31	1.46
Li and Lu [2020]	3.68	0.00534	5.15	0.88	5.02	0.0093	10.49	0.83
RenderEyelashNet	2.91	0.00352	3.57	1.28	3.80	0.0035	6.73	1.60
Ours	**2.47**	**0.00259**	**2.57**	1.23	**3.00**	**0.0011**	**4.35**	1.48



(a) Input (b) Trimap (c) Nam et al. [2019] (d) Li and Lu [2020] (e) RenderEyelashNet (f) Ours (g) Ground truth

Fig. 12. Exemplars of the quantitative results of 4 methods (Nam et al. [2019], Li and Lu [2020], RenderEyelashNet and EyelashNet (Ours), respectively) applying on the baseline testing dataset (containing both captured and Internet images). From left to right, we present the input images, the corresponding trimaps, the results of Nam et al. [2019], Li and Lu [2020], RenderEyelashNet, and ours, respectively.

eyelash images in EyelashNet, resulting in 5.46, 2.98, 2.90 MAE error and 1.94, 0.166, 0.154 variance for R, G, B channels of 24-bit images, respectively, showing that the UVA flash has little impact to the illumination of the captured images.

7.4 Evaluation of EyelashNet

We perform both qualitative and quantitative evaluations on EyelashNet. We train the baseline network on EyelashNet and RenderEyelashNet. Then we compare with [Li and Lu 2020] (trained

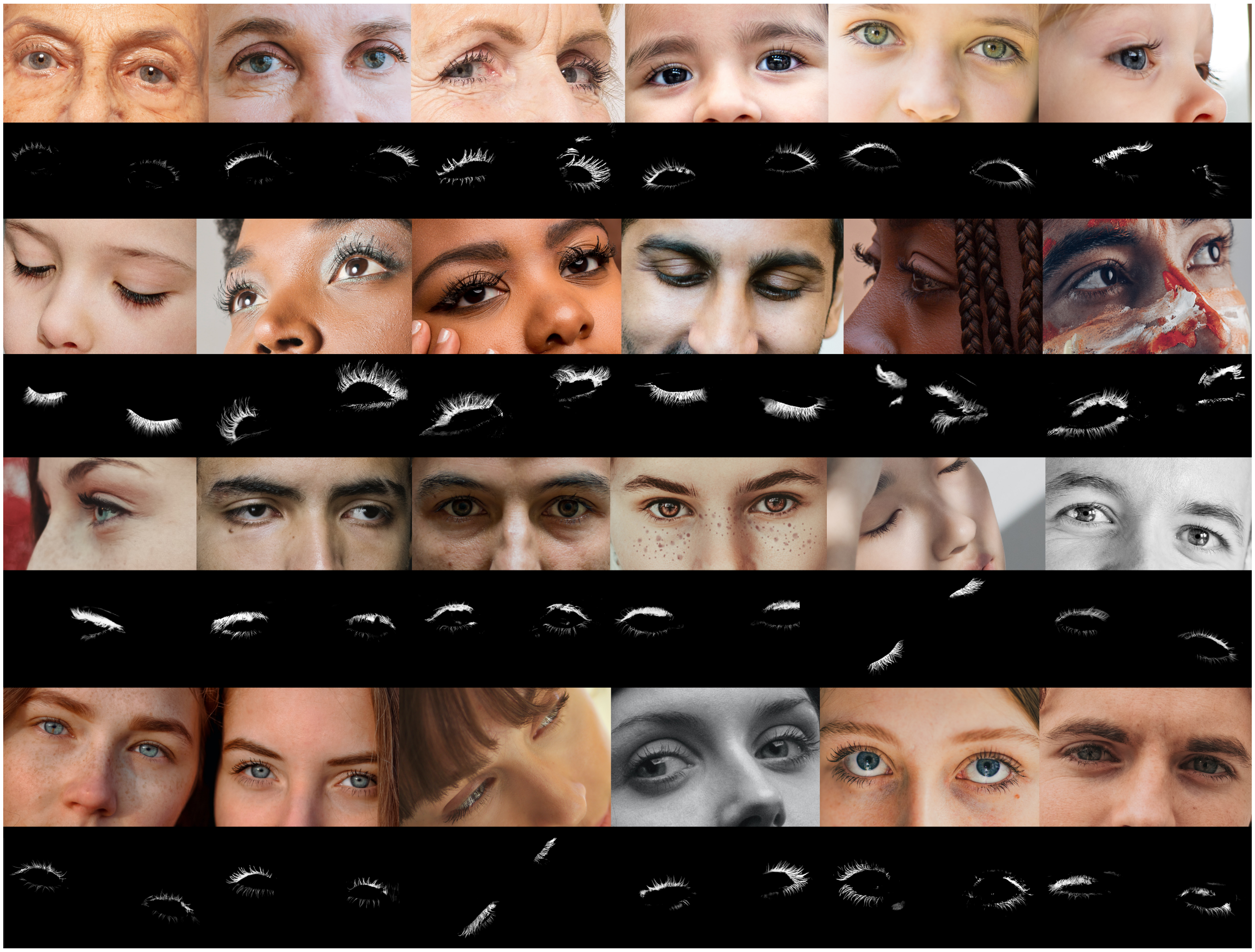


Fig. 13. Exemplars of alpha matte estimation on daily-captured images under different variations, such as poses, illuminations, shadows, ages, races, etc.

on the Adobe Image Matting dataset [Xu et al. 2017]), the state-of-the-art on natural image matting, and a Gabor-filter-based method [Nam et al. 2019], respectively. We evaluate the results of the above four methods on the baseline testing dataset. Table 2 shows the quantitative results. Compared with [Nam et al. 2019] and [Li and Lu 2020], our results achieve significant improvements. And our results are better than those from the baseline model trained on RenderEyelashNet. Fig. 12 presents the qualitative comparisons. Our results are visually better than RenderEyelashNet's results and have remarkable improvement over those from [Nam et al. 2019] and [Li and Lu 2020]. The results in both Table 2 and Fig. 12 demonstrate the effectiveness of our EyelashNet dataset.

7.5 Results on Daily-captured Images

Even though we capture the eyelash images in a lab-based environment, the baseline network trained on EyelashNet (augmented with RenderEyelashNet) can estimate eyelash alpha mattes from in-the-wild images under variations of illuminations, ages, races, etc.. Fig. 13 shows the results of the baseline network trained using EyelashNet on daily-captured images. We can observe that the baseline network is able to estimate high-quality eyelash alpha matte with a single RGB image without additional inputs. The eyelashes inside the fuzzy and self-shadow regions are properly estimated. The test images in Fig. 13 and Fig. 15 are courtesy of unsplash.com [Unsplash 2021].

7.6 Applications

Our method enables automatic and high-quality eyelash matting. It can be easily incorporated in various eyelash related applications, in particular for processing 2D portrait images and 3D virtual humans. First, eyelashes may induce noise and artifacts in MVS-based high-quality 3D face parametric reconstruction. The artists have to remove eyelashes and reconstruct eyelids manually, which can easily cost several hours. Our system can be used to automatically remove

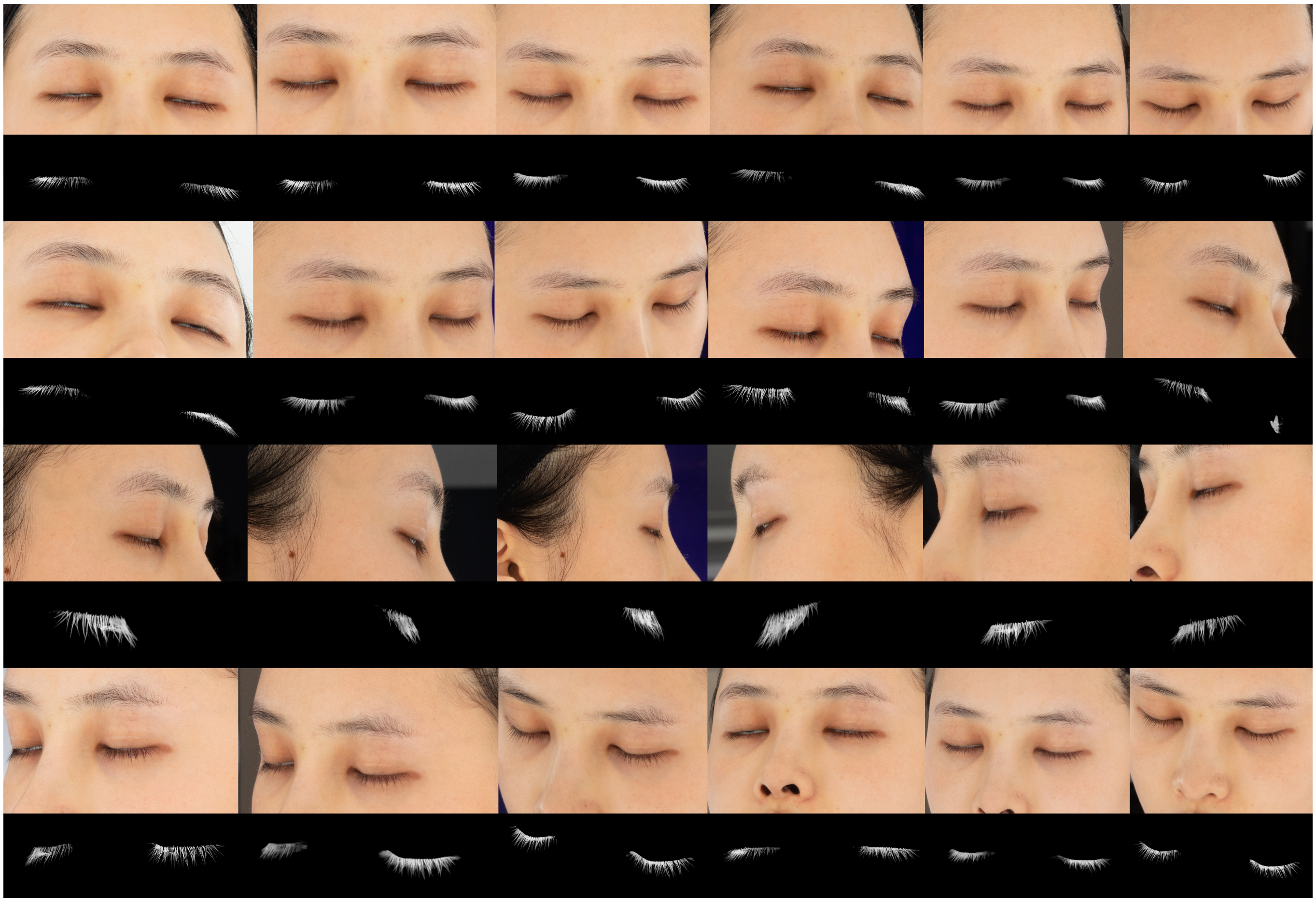
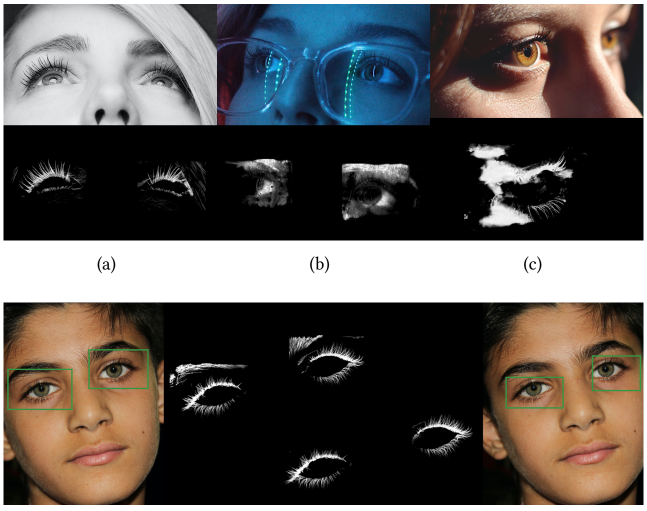


Fig. 14. Exemplars of alpha matte estimation on multi-view images.

eyelashes from multi-view images, resulting in better reconstructed geometry that can be further used for parametric reconstruction with better efficacy. Fig. 14 shows an example of eyelash alpha matte estimation in multi-view images, where our method can provide pose-free, high-quality eyelash alpha matte estimation that is particularly suitable for multi-view reconstruction applications. For example, as shown in Fig. 1 (e-i), without our eyelash removal process, the reconstructed eyelash geometry (see Fig. 1 (e)) may induce noises and artifacts when fitting the eyelid during the parametric reconstruction (see Fig. 1 (f)), which require very expensive manual repair. Our eyelash matting enables us to automatically remove the eyelashes in the input multi-view images, and reconstruct a better geometry of the eye region as shown in Fig. 1 (g). As a result, more faithful and high-quality eyelids can be reconstructed after the parametric reconstruction of the 3D face. We show the full rigged avatar in Fig. 1 (i) for completeness. Also, high-quality eyelash matting can be applied for cosmetic design as shown in Fig. 1 (d), where the eyelashes are recolored and lengthened under fine control with the help of eyelash alpha mattes. The results show that our eyelash matting system can provide fine-grained eyelash manipulation, and hence has wide applications on eyelash cosmetic design, high-quality face parametric reconstruction, etc.

8 DISCUSSION

While our eyelash matting system can achieve high-quality eyelash matte estimation, it still has some restrictions. First, due to the lack of ground truth for captured image (which is almost impossible to obtain), we use pseudo ground truth for evaluation, thus we cannot provide an accurate test of the networks. Second, the performance of eyelash matting could be affected in some extreme cases. For example, for portraits with occlusions (Fig. 15 (a)) or glasses (Fig. 15 (b)), artifacts may arise. The estimated eyelash alpha mattes in very strong shadow regions may not be accurate (Fig. 15 (c)). Since EyelashNet is captured from a laboratory environment, eyelids and eyeball may be detected by mistake in rare cases. This can be reduced by increasing the diversity of the dataset. Our model may matte out eyebrows, which is reasonable as eyebrows and eyelashes are similar in terms of the underlying hair structure. This can be resolved by using facial landmarks [Wu et al. 2018] to exclude the eyebrow region from the input image (as shown in the last row of Fig. 15 (d)).


(a) (b) (c)

(d)

Fig. 15. Example failure cases. Our method may cause artifacts due to occlusions (a), and may fail to estimate accurate eyelashes for portraits with glasses (b), and strong shadows (c). Our model may also matte out eyebrows in addition to eyelashes, depending on the region of interest for matting (d).

Third, for safety concerns and the aim of collecting clean data, sometimes the eyelashes were colored inadequately. On the one hand, the progressive training strategy refines the alpha matte inference network on its perceptually correct results. The weak supervision in the perceptual selection reduces the inference network's bias to part of the captured dataset. On the other hand, overfitting may arise for the inference network when performing more rounds of progressive training, limiting the inference network's generalization capability.

9 CONCLUSION AND FUTURE WORK

In this work, we create EyelashNet, the first eyelash matting dataset based on a newly developed fluorescent-based capture system and an alpha matte inference network. Our dataset consists of 10,672 training data (4,860 captured image pairs and 4,965 rendered image pairs) and 847 testing data (540 captured image pairs and 307 rendered image pairs). Each pair consists of an eyelash image and its corresponding eyelash alpha matte. With the dataset, we train a baseline matting network that can automatically estimate high-quality eyelash alpha mattes from real-world portrait images with various eye expressions, illuminations, and shadows. Our method is able to accurately extract eyelash alpha mattes from fuzzy and self-shadow regions such as pupils, which is almost impossible by manual annotations. Through extensive experiments, we have demonstrated the effectiveness of the capture system and inference network. Results show that the baseline network trained on our dataset outperforms prior methods and can estimate accurate eyelash mattes on internet images. We also demonstrate the applications of our work on high-quality virtual human reconstruction and cosmetic design. Our work makes a step towards high-fidelity eyelash matting, and it has the potential to inspire other works in the field. In the future, we will explore 3D eyelash reconstruction methods based on the estimated eyelash alpha mattes, and GAN-based methods to refine the estimated eyelash alpha matte to improve the eyelash details in fuzzy regions. We are also interested in extending our method to estimate high-quality eyebrow and beard alpha mattes.

ACKNOWLEDGMENTS

Xiaogang Jin was supported by the National Key R&D Program of China (Grant No. 2017YFB1002600), the Key Research and Development Program of Zhejiang Province (Grant No. 2018C03055), the National Natural Science Foundation of China (Grant Nos. 61972344, 61732015), and the Ningbo Major Special Projects of the "Science and Technology Innovation" (Grant No. 2020Z007). Yong-Liang Yang was supported by RCUK grant CAMERA (EP/M023281/1, EP/T022523/1), and a gift from Adobe. Tianjia Shao was supported by the National Natural Science Foundation of China (Grant Nos. 61772462, U1736217), and the 100 Talents Program of Zhejiang University. Kun Zhou was supported by the National Natural Science Foundation of China (Grant No. 61890954). We thank Yong Huang, Jingxiang Li, Cheng Ge, and other colleagues in NExT Studios for modeling, and rendering the digital human in Fig. 1 and the accompanying video. We thank Xinran Yao and Yuqing Zhang for kindly helping us to capture and label the eyelash matting data. We thank the anonymous reviewers for their insightful comments.

REFERENCES

Yagiz Aksoy, Tunç Ozan Aydin, and Marc Pollefeys. 2017. Designing Effective Inter-Pixel Information Flow for Natural Image Matting. In *2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*. IEEE Computer Society, 228–236.

Yagiz Aksoy, Tae-Hyun Oh, Sylvain Paris, Marc Pollefeys, and Wojciech Matusik. 2018. Semantic soft segmentation. *ACM Trans. Graph.* 37, 4 (2018), 72:1–72:13.

Thabo Beeler, Bernd Bickel, Gioacchino Noris, Paul A. Beardsley, Steve Marschner, Robert W. Sumner, and Markus H. Gross. 2012. Coupled 3D reconstruction of sparse facial hair and skin. *ACM Trans. Graph.* 31, 4 (2012), 117:1–117:10.

Amit Bermano, Thabo Beeler, Yeara Kozlov, Derek Bradley, Bernd Bickel, and Markus H. Gross. 2015. Detailed spatio-temporal reconstruction of eyelids. *ACM Trans. Graph.* 34, 4 (2015), 44:1–44:11.

Shaofan Cai, Xiaoshuai Zhang, Haoqiang Fan, Haibin Huang, Jiangyu Liu, Jiaming Liu, Jiaying Liu, Jue Wang, and Jian Sun. 2019. Disentangled Image Matting. In *2019 IEEE/CVF International Conference on Computer Vision, ICCV 2019*. IEEE, 8818–8827.

Chen Cao, Derek Bradley, Kun Zhou, and Thabo Beeler. 2015. Real-time high-fidelity facial performance capture. *ACM Trans. Graph.* 34, 4 (2015), 46:1–46:9.

Quan Chen, Tiezheng Ge, Yanyu Xu, Zhiqiang Zhang, Xinxin Yang, and Kun Gai. 2018. Semantic Human Matting. In *2018 ACM Multimedia Conference on Multimedia Conference, MM 2018*. ACM, 618–626.

Yunjey Choi, Youngjung Uh, Jaejun Yoo, and Jung-Woo Ha. 2020. StarGAN v2: Diverse Image Synthesis for Multiple Domains. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2020*. IEEE, 8185–8194.

International Electrotechnical Commission. [n.d.]. Photobiological safety of lamps and lamp systems. https://webstore.iec.ch/publication/7076

Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. 2009. ImageNet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*. 248–255.

Alexey Dosovitskiy, Philipp Fischer, Eddy Ilg, Philip Häusser, Caner Hazirbas, Vladimir Golkov, Patrick van der Smagt, Daniel Cremers, and Thomas Brox. 2015. FlowNet: Learning Optical Flow with Convolutional Networks. In *2015 IEEE International Conference on Computer Vision, ICCV 2015*. IEEE Computer Society, 2758–2766.

Xiaoxue Feng, Xiaohui Liang, and Zili Zhang. 2016. A Cluster Sampling Method for Image Matting via Sparse Coding. In *Computer Vision - ECCV 2016 - 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II (Lecture Notes in Computer Science, Vol. 9906)*. Springer, 204–219.

Marco Forte and François Pitié. 2020. *F*, *B*, Alpha Matting. arXiv:2003.07711 [cs.CV]

Pablo Garrido, Michael Zollhöfer, Dan Casas, Levi Valgaerts, Kiran Varanasi, Patrick Pérez, and Christian Theobalt. 2016. Reconstruction of Personalized 3D Face Rigs from Monocular Video. *ACM Trans. Graph.* 35, 3 (2016), 28:1–28:15.

Priya Goyal, Piotr Dollár, Ross Girshick, Pieter Noordhuis, Lukasz Wesolowski, Aapo Kyrola, Andrew Tulloch, Yangqing Jia, and Kaiming He. 2018. Accurate, Large

Minibatch SGD: Training ImageNet in 1 Hour. arXiv:1706.02677 [cs.CV]

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep Residual Learning for Image Recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016*. IEEE Computer Society, 770–778.

Tong He, Zhi Zhang, Hang Zhang, Zhongyue Zhang, Junyuan Xie, and Mu Li. 2019. Bag of Tricks for Image Classification with Convolutional Neural Networks. In *2019 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019*. Computer Vision Foundation / IEEE, 558–567.

Fu-Chung Huang, Gordon Wetzstein, Brian A. Barsky, and Ramesh Raskar. 2014. Eyeglasses-free display: towards correcting visual aberrations with computational light field displays. *ACM Trans. Graph.* 33, 4 (2014), 59:1–59:12.

Eddy Ilg, Nikolaus Mayer, Tonmoy Saikia, Margret Keuper, Alexey Dosovitskiy, and Thomas Brox. 2017. FlowNet 2.0: Evolution of Optical Flow Estimation with Deep Networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*. IEEE Computer Society, 1647–1655.

Daz Productions Inc. 2021. Daz3D - Model, Render, & Animate. https://www.daz3d.com/.

Sergey Ioffe and Christian Szegedy. 2015. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. In *Proceedings of the 32nd International Conference on Machine Learning (Proceedings of Machine Learning Research, Vol. 37)*, Francis Bach and David Blei (Eds.). PMLR, Lille, France, 448–456.

Joohwan Kim, Michael Stengel, Alexander Majercik, Shalini De Mello, David Dunn, Samuli Laine, Morgan McGuire, and David Luebke. 2019. *NVGaze: An Anatomically-Informed Dataset for Low-Latency, Near-Eye Gaze Estimation.* Association for Computing Machinery, New York, NY, USA, 1–12.

Cheng-Han Lee, Ziwei Liu, Lingyun Wu, and Ping Luo. 2020. MaskGAN: Towards Diverse and Interactive Facial Image Manipulation. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2020*. IEEE, 5548–5557.

Hao Li, Thibaut Weise, and Mark Pauly. 2010. Example-based facial rigging. *ACM Trans. Graph.* 29, 4 (2010), 32:1–32:6.

Jiaman Li, Zhengfei Kuang, Yajie Zhao, Mingming He, Karl Bladin, and Hao Li. 2020. Dynamic facial asset and rig generation from a single scan. *ACM Trans. Graph.* 39, 6 (2020), 215:1–215:18.

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 Trans. Graph.* 36, 6 (2017), 194:1–194:17.

Yaoyi Li and Hongtao Lu. 2020. Natural Image Matting via Guided Contextual Attention. In *The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020, The Thirty-Second Innovative Applications of Artificial Intelligence Conference, IAAI 2020, The Tenth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2020*. AAAI Press, 11450–11457.

Shanchuan Lin, Andrey Ryabtsev, Soumyadip Sengupta, Brian Curless, Steve Seitz, and Ira Kemelmacher-Shlizerman. 2020. Real-Time High-Resolution Background Matting. arXiv:2012.07810 [cs.CV]

Tsung-Yi Lin, Michael Maire, Serge J. Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick. 2014. Microsoft COCO: Common Objects in Context. In *Computer Vision - ECCV 2014 - 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V (Lecture Notes in Computer Science, Vol. 8693)*. Springer, 740–755.

Jinlin Liu, Yuan Yao, Wendi Hou, Miaomiao Cui, Xuansong Xie, Changshui Zhang, and Xian-Sheng Hua. 2020. Boosting Semantic Human Matting With Coarse Annotations. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2020*. IEEE, 8560–8569.

Si Liu, Xinyu Ou, Ruihe Qian, Wei Wang, and Xiaochun Cao. 2016. Makeup like a Superstar: Deep Localized Makeup Transfer Network. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence* (New York, New York, USA) *(IJCAI'16)*. AAAI Press, 2568–2575.

Ilya Loshchilov and Frank Hutter. 2017. SGDR: Stochastic Gradient Descent with Warm Restarts. arXiv:1608.03983 [cs.LG]

Hao Lu, Yutong Dai, Chunhua Shen, and Songcen Xu. 2019. Indices Matter: Learning to Index for Deep Image Matting. In *2019 IEEE/CVF International Conference on Computer Vision, ICCV 2019*. IEEE, 3265–3274.

Wan-Chun Ma, Mathieu Lamarre, Etienne Danvoye, Chongyang Ma, Manny Ko, Javier von der Pahlen, and Cyrus A. Wilson. 2016. Semantically-aware blendshape rigs from facial performance measurements. In *SIGGRAPH ASIA 2016, Macao, December 5-8, 2016 - Technical Briefs*. ACM, 3.

Sachit Menon, Alexandru Damian, Shijia Hu, Nikhil Ravi, and Cynthia Rudin. 2020. PULSE: Self-Supervised Photo Upsampling via Latent Space Exploration of Generative Models. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2020*. IEEE, 2434–2442.

MOVA. 2021. MOVA Contour Facial Capture. http://www.mova.com/.

Nitinraj Nair, Rakshit Kothari, Aayush K Chaudhary, Zhizhuo Yang, Gabriel J Diaz, Jeff B Pelz, and Reynold J Bailey. 2020. RIT-Eyes: Rendering of near-eye images for eye-tracking applications. In *ACM Symposium on Applied Perception 2020*. 1–9.

Giljoo Nam, Chenglei Wu, Min H. Kim, and Yaser Sheikh. 2019. Strand-Accurate Multi-View Hair Capture. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019*. Computer Vision Foundation / IEEE, 155–164.

Noris. 2021. Noris Color GmbH. http://www.norisusa.com/110UVInk.html.

Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Köpf, Edward Yang, Zach DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. 2019. PyTorch: An Imperative Style, High-Performance Deep Learning Library. arXiv:1912.01703 [cs.LG]

Yu Qiao, Yuhao Liu, Xin Yang, Dongsheng Zhou, Mingliang Xu, Qiang Zhang, and Xiaopeng Wei. 2020. Attention-Guided Hierarchical Structure Aggregation for Image Matting. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2020*. IEEE, 13673–13682.

Christoph Rhemann, Carsten Rother, Jue Wang, Margrit Gelautz, Pushmeet Kohli, and Pamela Rott. 2009. A perceptually motivated online benchmark for image matting. In *2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2009)*. IEEE Computer Society, 1826–1833.

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 (Lecture Notes in Computer Science, Vol. 9351)*. Springer, 234–241.

Gabriel Schwartz, Shih-En Wei, Te-Li Wang, Stephen Lombardi, Tomas Simon, Jason M. Saragih, and Yaser Sheikh. 2020. The eyes have it: an integrated eye and face model for photorealistic facial animation. *ACM Trans. Graph.* 39, 4 (2020), 91.

Soumyadip Sengupta, Vivek Jayaram, Brian Curless, Steven M. Seitz, and Ira Kemelmacher-Shlizerman. 2020. Background Matting: The World Is Your Green Screen. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2020*. IEEE, 2288–2297.

Yi-Chang Shih, Wei-Sheng Lai, and Chia-Kai Liang. 2019. Distortion-free wide-angle portraits on camera phones. *ACM Trans. Graph.* 38, 4 (2019), 61:1–61:12.

Zhixin Shu, Eli Shechtman, Dimitris Samaras, and Sunil Hadap. 2017. EyeOpener: Editing Eyes in the Wild. *ACM Trans. Graph.* 36, 1 (2017), 1:1–1:13.

Alvy Ray Smith and James F. Blinn. 1996. Blue Screen Matting. In *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH 1996*. ACM, 259–268.

Steven L. Song, Weiqi Shi, and Michael Reed. 2020. Accurate face rig approximation with deep differential subspace reconstruction. *ACM Trans. Graph.* 39, 4 (2020), 34.

Jian Sun, Yin Li, Sing Bing Kang, and Heung-Yeung Shum. 2006. Flash Matting. *ACM Trans. Graph.* 25, 3 (July 2006), 772–778.

Tristan Swedish, Karin Roesch, Ikhyun Lee, Krishna Rastogi, Shoshana Bernstein, and Ramesh Raskar. 2015. eyeSelfie: self directed eye alignment using reciprocal eye box imaging. *ACM Trans. Graph.* 34, 4 (2015), 58:1–58:10.

Laura C. Trutoiu, Elizabeth J. Carter, Iain Matthews, and Jessica K. Hodgins. 2011. Modeling and Animating Eye Blinks. *ACM Trans. Appl. Percept.* 8, 3, Article 17 (Aug. 2011), 17 pages.

Unsplash. 2021. Unsplash: Photos for everyone. https://unsplash.com/.

Congyi Wang, Fuhao Shi, Shihong Xia, and Jinxiang Chai. 2016. Realtime 3D eye gaze animation using a single RGB camera. *ACM Trans. Graph.* 35, 4 (2016), 118:1–118:14.

Mei Wang and Weihong Deng. 2018. Deep visual domain adaptation: A survey. *Neurocomputing* 312 (2018), 135–153.

Quan Wen, Feng Xu, Ming Lu, and Jun-Hai Yong. 2017b. Real-time 3D eyelids tracking from semantic edges. *ACM Trans. Graph.* 36, 6 (2017), 193:1–193:11.

Quan Wen, Feng Xu, and Jun-Hai Yong. 2017a. Real-Time 3D Eye Performance Reconstruction for RGBD Cameras. *IEEE Trans. Vis. Comput. Graph.* 23, 12 (2017), 2586–2598.

Eric Whitmire, Laura Trutoiu, Robert Cavin, David Perek, Brian Scally, James Phillips, and Shwetak Patel. 2016. EyeContact: Scleral Coil Eye Tracking for Virtual Reality. In *Proceedings of the 2016 ACM International Symposium on Wearable Computers* (Heidelberg, Germany) *(ISWC '16)*. Association for Computing Machinery, New York, NY, USA, 184–191.

Erroll Wood, Tadas Baltrušaitis, Louis-Philippe Morency, Peter Robinson, and Andreas Bulling. 2016. Learning an Appearance-Based Gaze Estimator from One Million Synthesised Images. In *Proceedings of the Ninth Biennial ACM Symposium on Eye Tracking Research & Applications*. 131–138.

Wayne Wu, Chen Qian, Shuo Yang, Quan Wang, Yici Cai, and Qiang Zhou. 2018. Look at Boundary: A Boundary-Aware Face Alignment Algorithm. In *2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018*. IEEE Computer Society, 2129–2138.

Ning Xu, Brian L. Price, Scott Cohen, and Thomas S. Huang. 2017. Deep Image Matting. In *2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*. IEEE Computer Society, 311–320.

Yunke Zhang, Lixue Gong, Lubin Fan, Peiran Ren, Qixing Huang, Hujun Bao, and Weiwei Xu. 2019. A Late Fusion CNN for Digital Matting. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019*. Computer Vision Foundation / IEEE, 7469–7478.