Pacific Graphics 2022
N. Umetani, E. Vouga, and C. Wojtan
(Guest Editors)

Volume 40 (2021), Number 7

Effective Eyebrow Matting with Domain Adaptation

Luyuan Wang Hanyuan Zhang Qinjie Xiao Hao Xu Chunhua Shen Xiaogang Jin[†]

State Key Lab of CAD&CG, Zhejiang University, Hangzhou 310058, China

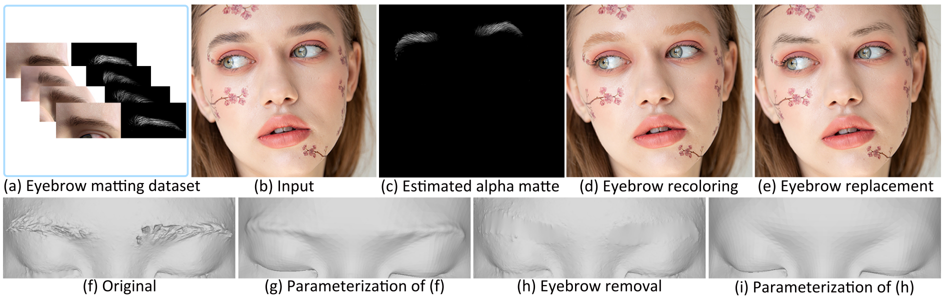


Figure 1: *We create the first synthetic eyebrow matting dataset (a). This enables semi-supervised training of a domain adaptation eyebrow matting network. The network can learn domain-robust features from synthetic data (a) together with unlabeled real-world eyebrow images without using any real matting data and estimate high-quality eyebrow alpha matte (c) from a real RGB image (b) only without any prior. The eyebrow matting allows us to automatically remove the interference of eyebrows during the multi-view stereo (MVS) based 3D face reconstruction process, and therefore largely enhances the efficiency and efficacy of the reconstruction of eyebrow regions. Without eyebrow removal, the reconstructed eyebrow geometry (f) often induces noises and artifacts when fitting the eyebrow during 3D parametric face reconstruction (g), which requires very expensive manual repair in hours. In contrast, eyebrow matting facilitates the easy attainment of better geometry (h) and more faithful parameterization of the eyebrow region (i). Furthermore, our eyebrow matting method can be used for cosmetic design purposes such as eyebrow recoloring (d) and eyebrow replacement (e).*

Abstract

We present the first synthetic eyebrow matting datasets and a domain adaptation eyebrow matting network for learning domain-robust feature representation using synthetic eyebrow matting data and unlabeled in-the-wild images with adversarial learning. Different from existing matting methods that may suffer from the lack of ground-truth matting datasets, which are typically labor-intensive to annotate or even worse, unable to obtain, we train the matting network in a semi-supervised manner using synthetic matting datasets instead of ground-truth matting data while achieving high-quality results. Specifically, we first generate a large-scale synthetic eyebrow matting dataset by rendering avatars and collect a real-world eyebrow image dataset while maximizing the data diversity as much as possible. Then, we use the synthetic eyebrow dataset to train a multi-task network, which consists of a regression task to estimate the eyebrow alpha mattes and an adversarial task to adapt the learned features from synthetic data to real data. As a result, our method can successfully train an eyebrow matting network using synthetic data without the need to label any real data. Our method can accurately extract eyebrow alpha mattes from in-the-wild images without any additional prior and achieves state-of-the-art eyebrow matting performance. Extensive experiments demonstrate the superior performance of our method with both qualitative and quantitative results.

CCS Concepts

• ***Computing methodologies*** $\rightarrow$ ***Image processing;***

1. Introduction

Eyebrow is crucial in representing personal characteristics and emotions. As a result, eyebrow modeling and editing play an im-

[†] Corresponding Author: jin@cad.zju.edu.cn

portant role in many applications, such as digital human reconstruction, portrait editing, cosmetic design, etc. To model or edit the eyebrow in the captured portraits, a common workflow [BBN*12] in the modern computer vision and computer graphics industry consists of two steps: 1) matting/masking out the eyebrow region from the captured portraits; and 2) using the eyebrow matte/mask as prior information for further manipulations. While the latter step has attracted significant interest from the research community, such as hairstyle editing [XYH*21, ZAFW21], eyelash matting [XZZ*21], hair reconstruction [BBN*12, NWKS19] and modeling [YSZZ19], etc., the former still suffers from tedious manual efforts. Previous works [XYH*21, ZAFW21] manually mask out rough head hair masks to edit hairstyle using GAN-based generation methods. However, eyebrow contains much intricate sub-pixel alpha information. Using binary masks for strand-accurate hair editing in captured portraits may generate noticeable artifacts since no transparency information is taken into account. An improved solution is to use the hair matte as priors. It has been proved that using the hair matte leads to better performance than using the mask in hair generation tasks [XYH*21] and eyelash editing [XZZ*21].

However, predicting the eyebrow mattes from captured portraits is a very challenging task. It is almost impossible to obtain accurate eyebrow mattes by manual annotations since it is extremely labor-intensive. A more efficient way is to perform image matting. Nevertheless, current image matting methods [LL20, QLY*20, LRS*21, AOP*18] may produce unsatisfactory matting results due to the lack of eyebrow matting datasets for training. Existing methods for building matting datasets require specifying solid color backgrounds to infer detailed foreground objects, such as bluescreening [RRW*09, SB96], green-screen keying [AAPS16, DJSXN19], etc. However, as eyebrow is grown on the skin (background), it is intractable to apply those methods in the case of eyebrow matting. The fluorescent labeling system developed by Xiao et al. [XZZ*21] can highlight eyelashes (foreground) as priors to infer alpha matte. However, they need to color the eyelash very carefully to prevent applying to the skin. Hence, the coloring process is extremely time-consuming, especially when coloring a large amount of eyebrows. Furthermore, current methods for generating ground-truth matting datasets suffer from tedious manual annotations or capturing processes. This limits the scale of the matting datasets [RRW*09, XPCH17, QLY*20, STT21]. To summarize, constructing a large-scale eyebrow matting dataset from real-world portraits remains a great challenge.

Thus, one of the primary goals of the work here is to overcome the dataset limitation in eyebrow matting. Different from existing methods that often rely on the large-scale annotated dataset, our approach leverages a synthetic eyebrow matting dataset generated by rendering avatars instead of annotating eyebrow matting ground-truth data from captured portraits. Our method using synthetic images offers several advantages:

- *Photo-realistic*. First, the rendering images are realistic and of very high quality.
- *Scalable*. Second, compared to annotating real-world images, it is much cheaper to obtain a large-scale, high-quality matting dataset.
- *Flexible*. Third, the data diversity can be fully controlled, such as expressions, illuminations, eyebrow styles, poses, etc.

While a synthetic dataset can solve the problem of the lack of a real-world matting dataset, it is not difficult to distinguish between the rendered and real-world images. That is, there exists a domain gap between the virtual data and real data, making the model trained on synthetic data perform poorly on real data. The domain difference can be alleviated by a progressive training strategy [XZZ*21]. However, such a method requires an accurate eyelash mask as priors, which is difficult and labor-intensive to obtain. To tackle the domain gap between the rendered and real-world eyebrow images, we present a semi-supervised semantic matting network inspired by [RL18] to learn domain-robust alpha-matte features from synthetic eyebrow images while adapting the feature representation to real-world images using *adversarial learning*. The network contains an encoder, a decoder, and a discriminator, and is trained by alternating training. In one round of training, the encoder first extracts features from the rendered image and the real image, respectively. Next, we freeze the encoder and decoder, input the rendered image features and the real image features into the discriminator, respectively, and employ the adversarial discriminant loss to train the discriminator such that the features extracted by the encoder in the two domains become closer. Second, we freeze the discriminator, input the rendered graph features into the decoder, generate the estimated eyebrow mask, and employ the mean absolute error loss to train the encoder and decoder. We iteratively perform this alternate training until the network converges. In the testing phase, the image data are passed through the encoder-decoder to generate the estimated eyebrow masks. In addition, we manually annotate a test eyebrow matting dataset containing various eyebrow images from the Internet. We validate our method by comparing against the state-of-the-art methods quantitatively and qualitatively, and the experimental results show that our method achieves the best performance on eyebrow matting. To summarize, our paper makes the following main contributions:

- We present a semi-supervised eyebrow matting network that uses adversarial domain adaptation to learn domain-robust feature representation from virtual matting data and unlabeled in-the-wild images using an alternating training strategy. The network allows for automatic and accurate eyebrow matting on portrait images with a variety of eyebrow styles, colors, genders, ages, expressions, illuminations, and poses.
- We build the first eyebrow matting dataset generated by rendering avatars. To validate our method, we manually annotate a test in-the-wild eyebrow matting dataset. The source code and the eyebrow dataset have been made publicly available at `https://github.com/Dancingmader/DAM-Net`.

2. Related Work

Matting datasets are the key component for supervised learning-based image matting. A few datasets were released to the community. The matting benchmark on `alphamatting.com` [RRW*09] intrigues further data-driven-based matting methods. To overcome the dataset's scale limit of `alphamatting.com` [RRW*09], which has only 27 foregrounds (FG) training data, larger matting datasets such as composition-1k [XPCH17] (250 dissimilar FG objects), Dinstinctions-646 [QLY*20] (646 dissimilar FG objects) were created to improve the matting quality for the image matting task. Sun et al. [STT21] further present a large-scale semantic image matting dataset with careful consideration of data balancing across different semantic classes to learn richer semantic information for image matting. Matting datasets for significant FG objects such as portraits [STG*16] and humans [CGX*18] were also explored. These datasets focus on the overall structure of objects instead of local and detailed components such as eyebrows. Local objects like the eyebrow masks in CelebAMask-HQ [LLWL20] and eyelash matting datasets [XZZ*21] were released too. However, these datasets are not suitable for eyebrow matting due to the lack of high-quality eyebrow masks. In summary, eyebrow matting has not been addressed by prior works, and there is no eyebrow matting dataset available thus far. On the other hand, creating a high-quality eyebrow matting dataset is rather challenging. Because the eyebrow is grown on the skin, it is impossible to build eyebrow matting datasets using the aforementioned existing methods [RRW*09, SB96, AAPS16, DJSXN19]. Likewise, Xiao et al. [XZZ*21] may easily induce noises to the skin when coloring the eyebrows as addressed before. Moreover, it is tedious and time-consuming to manually annotate detailed eyebrows compared with other objects of interest such as humans [CGX*18].

In our work, we create a synthetic eyebrow matting dataset by rendering avatars for eyebrow matting. Moreover, we train our network in a semi-supervised manner and can learn domain-robust features from the synthetic matting dataset and unlabeled real-world eyebrow images without annotating any real matting data.

Image matting is a fundamental problem that has been actively studied in computer vision and graphics. Affinity-based methods [AOAP17] that rely on pixel similarity metrics are proposed for image matting. Sampling-based methods [FLZ16] have been investigated as well. Sun et al. [SLKS06] extract mattes using a pair of flash/no-flash images in a sufficiently distant background scene. Note that these methods often generate unsatisfactory results when applied to delicate objects such as eyebrows.

Recent deep learning methods have achieved encouraging results on natural image matting [LDSX19, LL20, STT21], portrait matting [STG*16, DLS21], and human matting [CGX*18, LYH*20, LRS*21, SJC*20, YZZ*21]. Prior-based methods [XPCH17, AOAP17, YZZ*21, LXZ*21] take an image and a prior (mask, trimap, etc.) to estimate the detailed alpha matte. Li and Lu [LL20] present a guided content attention network to weaken the requirements of a trimap and attain better results than affinity-based methods [AOAP17]. However, labeling accurate trimaps or masks of eyebrows requires tedious user interaction, which can be infeasible in practice. Recently, fully automatic matting methods [ZGF*19, QLY*20, LYH*20, STT21] have been proposed to overcome the limitation of annotating trimaps or mask inputs.

Domain adaptation methods are proposed to reduce the domain shift between different datasets and can be applied to various applications such as image segmentation [CWPC19, KKS*21, HWYY18, SBJ*18], image classification [THSD17], image translation, and generation [HTP*18, LCD*19], and learning generalizable high-level visual representations [RL18], etc. Thus far, few domain adaptation methods pay attention to the task of image matting.

Xiao et al. [XZZ*21] propose a progressive training strategy to reduce the domain gap between the synthetic eyelash matting data and captured real eyelash data. Nevertheless, to achieve plausible matting results and avoid over-fitting, their method requires capturing a strong eyelash prior, which is unfeasible for eyebrows. Different from their methods, we present a domain adaptation matting network to learn domain-invariant mid-level alpha features from a carefully-designed synthetic eyebrow matting dataset and unlabeled real-world images based on *adversarial learning*. Next, we present our approach.

3. Domain Adaptation Matting Network

We first demonstrate the U-Net baseline for eyebrow matting, then introduce the adversarial learning-based domain adaptation network. Finally, we introduce the setups of the synthetic eyebrow matting dataset.

3.1. Baseline Matting Network

The baseline network is a U-net structure [RFB15] (see Fig. 2 in [LL20]), which is an encoder-decoder network with a few stacked residual blocks [HZRS16]. The input to the baseline network is an RGB image. The output is an alpha matte estimation of the foreground objects. The baseline network leverages an alpha prediction loss, defined as the average difference between the ground truth and the estimated alpha matte to update the network's parameters $\phi\boldsymbol{E}, \phi\boldsymbol{R}$:

$$L_\alpha(\phi\boldsymbol{E}, \phi\boldsymbol{R}|\hat{\boldsymbol{\alpha}}) = \frac{1}{|\hat{\boldsymbol{\alpha}}|}\sum_k |\hat{\alpha}_k - \alpha_k|, \tag{1}$$

where $\hat{\boldsymbol{\alpha}}$ is the estimated alpha matte. $\hat{\alpha}_k \in \hat{\boldsymbol{\alpha}}$ and α_k are the estimated and ground truth values of the alpha matte at position k. The loss metric is ℓ_1 loss.

3.2. Domain Adaptation Matting Network

In this section, a domain adaptation matting network (DAM-Net) is designed to learn semantic features from synthetic eyebrow data and adapts the feature representation to real-world images. The network consists of an encoder $\boldsymbol{E}$, a decoder $\boldsymbol{R}$, and a discriminator $\boldsymbol{D}$. Fig. 2 illustrates the network pipeline of the domain adaptation network. We divide the structure of the baseline network into two parts, the encoder and decoder of DAM-Net.

We perform an alternating training approach to train the network. For a given real-world eyebrow dataset $\boldsymbol{Y}$, a synthetic eyebrow image dataset $\boldsymbol{X}$ and its corresponding alpha matte dataset $\boldsymbol{A}$, the training process of the DAM-Net is divided into two steps (as shown in Algorithm 1). In the first step, we randomly select a synthetic matting data $\boldsymbol{x}_i \in \boldsymbol{X}, \boldsymbol{\alpha}_i \in \boldsymbol{A}$ and a real-world eyebrow image $\boldsymbol{y}_i \in \boldsymbol{Y}$ as inputs to DAM-Net. The encoder infers alpha features from the synthetic image $\boldsymbol{z_x} = \{\boldsymbol{z}_{\boldsymbol{x}_i} | \boldsymbol{z}_{\boldsymbol{x}_i} = \boldsymbol{E}(\boldsymbol{x}_i)\}$ and the real-world image $\boldsymbol{z_y} = \{\boldsymbol{z}_{\boldsymbol{y}_j} | \boldsymbol{z}_{\boldsymbol{y}_j} = \boldsymbol{E}(\boldsymbol{y}_j)\}$. The discriminator $\boldsymbol{D}$ is employed to determine whether the input alpha features originate from a synthetic image or a real image. We find that such simple setups are sufficient to make the encoder to learn domain-invariant features from synthetic and real-world eyebrow images. Then we minimize the

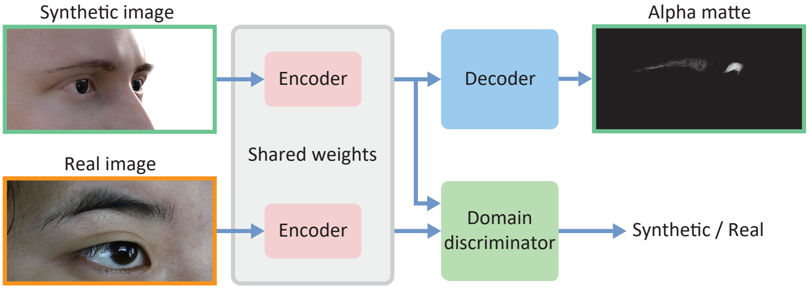


Figure 2: *An overview of the domain adaptation matting network, which is made up of an encoder, a decoder, and a domain discriminator. During the training stage, the encoder first infers alpha features from a synthetic image and then from a real-world image. The domain discriminator is then trained to accurately distinguish whether the input alpha feature is from a real or synthetic image. Finally, the network updates the encoder to deceive the discriminator, preventing features from the two domains from being easily classified, while also updating the decoder to learn good features for eyebrow matting. The encoder-decoder network takes an RGB image as input and outputs an estimated alpha matte during inference.*

parameters $\phi \boldsymbol{D}$ of the discriminator $\boldsymbol{D}$ using the following binary cross-entropy loss:

$$L_D(\phi \boldsymbol{D}|z_{\boldsymbol{x}}, z_{\boldsymbol{y}}) = -\sum_i \log \boldsymbol{D}(z_{x_i}) - \sum_j \log(1 - \boldsymbol{D}(z_{y_j})). \quad (2)$$

In the second step, the decoder infers alpha mattes $\hat{\boldsymbol{\alpha}}_{\boldsymbol{x}} = \{\hat{\boldsymbol{\alpha}}_{\boldsymbol{x}_i}\}$ from the synthetic images. We freeze the discriminator $\boldsymbol{D}$ and update the parameters $\phi \boldsymbol{E}, \phi \boldsymbol{R}$ of the encoder $\boldsymbol{E}$ and decoder $\boldsymbol{R}$ by minimizing the following loss:

$$\begin{aligned} L_{ER}(\phi \boldsymbol{E}, \phi \boldsymbol{R}|\hat{\boldsymbol{\alpha}}_{\boldsymbol{x}}, z_{\boldsymbol{x}}) = &-\lambda_d \sum_i \log(\boldsymbol{D}(1 - z_{x_i})) \\ &+ \frac{\lambda_a}{|\hat{\boldsymbol{\alpha}}_{\boldsymbol{x}}|} \sum_i L_\alpha(\phi \boldsymbol{E}, \phi \boldsymbol{R}|\hat{\boldsymbol{\alpha}}_{\boldsymbol{x}_i}), \end{aligned} \quad (3)$$

where λ_d, λ_a are the weights for the discriminative loss and the alpha prediction loss, respectively. L_{ER} updates the encoder $\boldsymbol{E}$ to fool the discriminator $\boldsymbol{D}$ into thinking that the alpha features extracted from a synthetic image are from a real-world image, while also updating the decoder R, such that the features are good for estimating alpha mattes from real-world images.

Algorithm 1 Pseudocode of the domain adaptation network's training process.

Input: Synthetic images and their corresponding alpha mattes $\boldsymbol{X}, \boldsymbol{A}$, real-world images $\boldsymbol{Y}$, maximum iterations T
Output: domain adaptation encoder $\boldsymbol{E}$, decoder $\boldsymbol{R}$, discriminator $\boldsymbol{D}$

for t = 1 to T **do**
  Sample a batch of synthetic images $\boldsymbol{x} = \{\boldsymbol{x}_i\}$
  Sample a batch of real images $\boldsymbol{y} = \{\boldsymbol{y}_j\}$
  Extract features from synthetic images $z_{\boldsymbol{x}} = \{z_{x_i} | z_{x_i} = \boldsymbol{E}(\boldsymbol{x}_i)\}$
  Predict alpha mattes $\hat{\boldsymbol{\alpha}}_{\boldsymbol{x}} = \{\hat{\boldsymbol{\alpha}}_{\boldsymbol{x}_i}\}$ for synthetic images, and calculate loss $L_{ER}(\phi \boldsymbol{E}, \phi \boldsymbol{R}|\hat{\boldsymbol{\alpha}}_{\boldsymbol{x}}, z_{\boldsymbol{x}})$
  Extract features from real-world image: $z_{\boldsymbol{y}} = \{z_{y_j} | z_{y_j} = \boldsymbol{E}(\boldsymbol{y}_j)\}$
  Calculate loss $L_D(\phi \boldsymbol{D}|z_{\boldsymbol{x}}, z_{\boldsymbol{y}})$
  Keep $\boldsymbol{E}, \boldsymbol{R}$ frozen, update $\boldsymbol{D}$ through $L_D(\phi \boldsymbol{D}|z_{\boldsymbol{x}}, z_{\boldsymbol{y}})$
  Keep $\boldsymbol{D}$ frozen, update $\boldsymbol{E}$, $\boldsymbol{R}$ through $L_{ER}(\phi \boldsymbol{E}, \phi \boldsymbol{R}|\hat{\boldsymbol{\alpha}}_{\boldsymbol{x}}, z_{\boldsymbol{x}})$
end for

3.3. Synthetic Eyebrow Matting Dataset

We construct a synthetic eyebrow matting dataset created by rendering virtual avatars. The avatars are created and rendered using DAZ 3D [Daz21]. We use seventeen avatars (two white females, four white males, two black females, two black males, one Asian female, three Asian males, two brown females, and one brown male) with various variations, such as race, gender, eyebrow style, age, pose, expression, and illumination (Table 1). Two white elderly men and one Asian elderly man are specifically included to provide unique data to the elderly on eyebrow shape and color (e.g., gray, white, and grizzled). We continuously vary their head poses from -150° to 150° for the yaw angle and -45° to 45° for the pitch angle. Initially, we set 22 eyebrow shapes and 29 eyebrow colors for female avatars, and 23 eyebrow shapes and 26 eyebrow colors for male avatars in the keyframes of animation sequences (Fig. 3). We completely customize dozens of expressions for these avatars (Fig. 4), and the eyebrow shape changes smoothly as the expression varies. As a result, the number of eyebrow shapes in the synthetic dataset is larger than the initial number. With such a setup, the rendered eyebrow dataset contains the most common poses that may appear in common portrait photos. We use detailed 3D models of eyebrows instead of 2D eyebrow textures to enhance the eyebrow diversity in multiple views and simulate the self-shadows of the eyebrows. We render eyebrow images in RGBA format and then extract the alpha channel as alpha mattes.

Finally, we construct 1,000 alpha matting data instances with varying genders, races, eyebrow colors, eyebrow styles, poses, expressions, etc. The dataset is split into a training dataset (800 rendered data) and a test dataset (200 rendered data), and will be made publicly available.

4. Experiments

We first introduce the implementation and data preparation details. Then, we conduct ablation studies to validate the effectiveness of our method. After that, we qualitatively and quantitatively evaluate the performance of our method to demonstrate its effectiveness. We use four metrics for quantitative evaluation, including the mean

Table 1: *Our synthetic eyebrow matting dataset includes skin tone, gender, initial eyebrow shape, and eyebrow color combinations. Some of the initial eyebrow shapes and colors listed in the table are reused among different avatars (for example, most brows have the black color option).*

Skin tone	Gender	#Avatar	#Initial eyebrow shape	#Eyebrow color
White	Female	2	2	1
	Male	4	6	9
Asian	Female	1	15	3
	Male	3	17	17
Brown	Female	2	17	4
	Male	1	14	10
Black	Female	2	4	21
	Male	2	15	10

Figure 3: *Examples of various eyebrow style combinations. We render unusual eyebrow shapes such as the elders' sparse eyebrows and unibrows. In addition to the common eyebrow color, we offer uncommon eyebrow colors such as white, salt-pepper, gray, gold, ginger, red, and so on.*

square error (MSE), the sum of absolute difference (SAD), the gradient (GRAD), and the connectivity error (Conn) proposed in *alphamatting.com* [RRW*09]. Finally, we apply our work to several practical eyebrow editing applications.

Table 2: *Network structure of discriminator.*

Module	Module Type	I/O Channel	Input
$\boldsymbol{D}_{Conv1}$	conv3x3+BN+ReLU	512/512	Encoder
$\boldsymbol{D}_{Conv2}$	conv1x1+BN+ReLU	512/256	$\boldsymbol{D}_{Conv1}$
$\boldsymbol{D}_{Conv3}$	conv1x1+`sigmoid`	256/1	$\boldsymbol{D}_{Conv2}$

4.1. Implementation

Network implementation details. In terms of the baseline network, the encoder $\boldsymbol{E}$ has 3 *conv* layers and 4 residual blocks, while the decoder $\boldsymbol{R}$ has 4 residual blocks, 1 *deconv* layer, and 1 *conv* layer. The skip connections, which provide lower-level features to the estimated alpha matte, are built using five shortcut layers. Two guided contextual attention (GCA) modules extract similarity information from low-level image features to refine the alpha features (see GCA [LL20]). The discriminator in our network is a ResNet-based network with three convolutional layers and an activation layer. The final layer employs `sigmoid` as the activation function and returns true or false, indicating whether the input alpha features are from a real or synthetic image. Table 2 shows the structure of the discriminator. The discriminator is initialized to a constant, and the encoder and decoder are initialized using the customized ResNet-34 [HZRS16] backbone (trained on ImageNet [DDS*09]) of [LL20]. The network is trained for 200,000 iterations with batch size 4, taking about 33.75 hours. We use the 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 [GDG*18, LH17, HZZ*19]. We set hyperparameters $\lambda_d = 0.25$, $\lambda_a = 3$ for training DAM-Net. The input image size for training is 512×512. We utilize Pytorch (version 1.11.0) [PGM*19] to train the networks and run all comparisons on a desktop PC with a single NVIDIA GTX 2080 (8GB memory), Intel Xeon W-2123 3.60 GHz CPU, and 32GB RAM.

Dataset. We use 800 synthetic eyebrow matting data and 1,215 Internet images as the training dataset to train our network. Evaluating the performance of a matting algorithm usually requires a realistic and accurate ground truth eyebrow dataset to serve as a baseline testing dataset. However, it is almost impossible to obtain the accurate eyebrow alpha matte from captured real-world data. A compromise is using manual annotations. Manually labeling is a very time-consuming process: it takes about 30 minutes to annotate one eyebrow matting ground-truth image for a professional artist.

In our work, we use 200 synthetic eyebrow matting data and 68 manually annotated real-world eyebrow matting data (Fig. 5) as baseline testing dataset for evaluation.

Data augmentation. The rendering scenes limit the variety of synthetic eyebrow images. Creating and rendering a large number of scenes, on the other hand, is time-consuming and labor-intensive. Hence, we further apply data augmentation to enhance the data diversity. We first 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 eyebrow 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.

4.2. Comparison

We perform both qualitative and quantitative evaluations to validate the proposed domain adaptation matting network. We train the baseline network and DAM-Net on the proposed eyebrow matting dataset. Then we compare against the method of Li and Lu [LL20] (trained on the Adobe Image Matting dataset [XPCH17]), the progressive training method of Xiao et al. [XZZ*21] (trained on

Figure 4: *An exemplar of rendered data with 10 expressions. From left to right and top to bottom, the eye expressions are neutral, closed eyes, frown, looking down, looking left, looking right, angry, raising eyebrows, simper, surprised, respectively.*

Figure 5: *Exemplars of manually labeled (columns 1-3) and synthetic (columns 4-5) eyebrow matting data through rendering 3D avatars.*

the eyebrow matting dataset), the state-of-the-art on eyelash matting, and the baseline network, respectively. Specially, we provide trimaps for Li and Lu [LL20] for a fair comparison. We evaluate the results of the above four methods on the baseline testing dataset.

Table 3 shows the quantitative results. Compared with Li and Lu [LL20], Xiao et al. [XZZ*21], and the baseline network, our results achieve marginal improvements in MSE and SAD metrics on the Internet image test dataset while being slightly worse than the baseline network on the synthetic test dataset. This is expected since the bias of the network to the synthetic data becomes larger after performing domain adaptation. It is also desirable, as we generally apply the trained model to real Internet images instead of synthetic images.

Fig. 6 presents the qualitative comparisons. Our results are visually better than those from Li and Lu [LL20], Xiao et al. [XZZ*21] and the baseline network on the Internet images (from row 1 to 4, while the results (from rows 5 to 6) of our method on the synthetic images are visually acceptable compared to Li and Lu [LL20], Xiao et al. [XZZ*21] and the baseline network. The results in both Table 3 and Fig. 6 demonstrate the effectiveness of our method. Since the eyebrows are relatively sparse and the eyebrow matte contains a large number of weak alpha regions, the connectivity metrics may not be suitable for evaluating the quality of the eyebrow matte. Hence, although our qualitative results are visually better, the quantitative results are worse on the Internet test dataset than the method of Xiao et al. [XZZ*21] and the baseline network in Conn metric.

In addition, we compare the inference speed to previous methods. For a single 512*512 resolution input, GCA [LL20], Xiao et al. [XZZ*21], baseline, Sun et al. [STT21], and our inference times are 93.3 ms, 90.6 ms, 91.3 ms, 10.5 ms, and 91.7 ms, respectively.

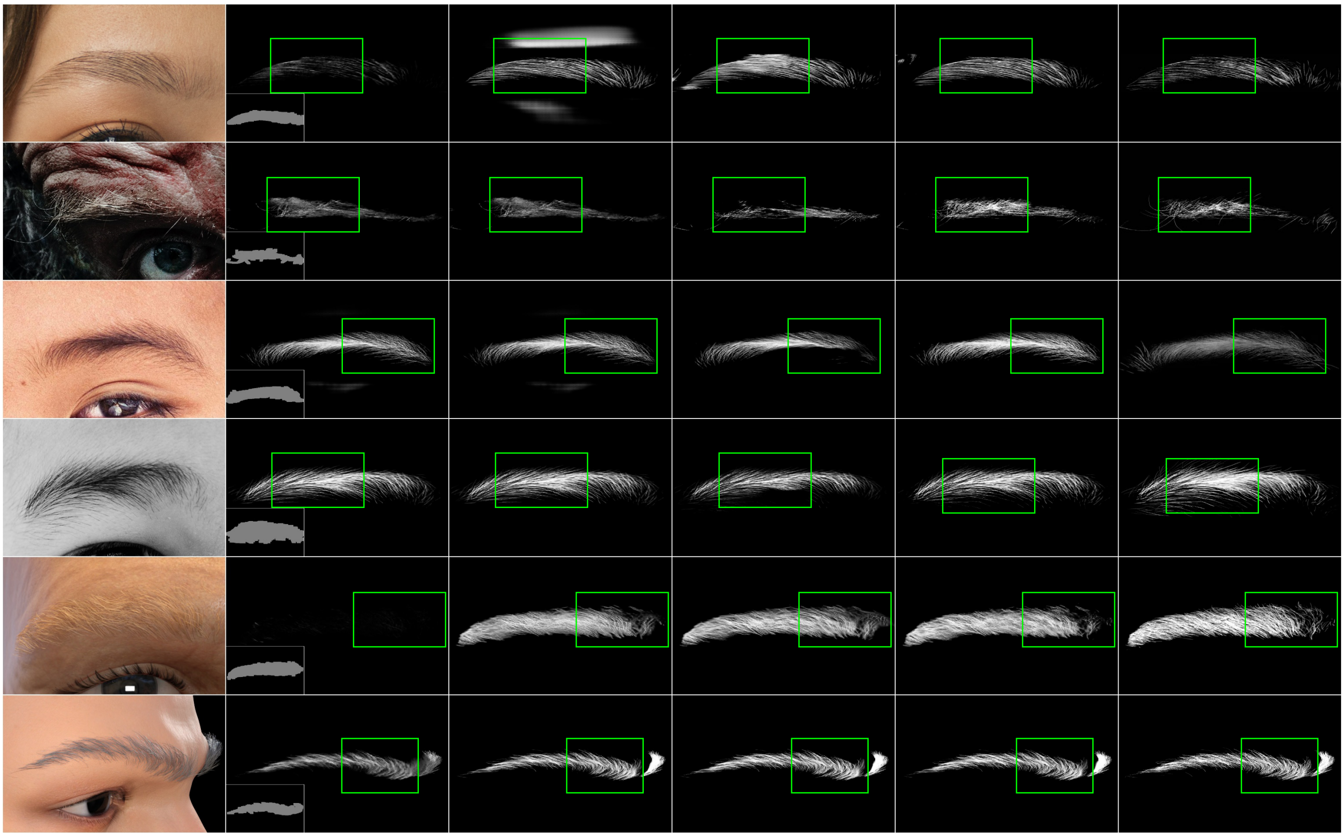
4.3. Effectiveness of Domain Adaptation

The results of the baseline network reported in Table 3 and Fig. 6(d) represent the results of removing the domain adaptation (removing the discriminator and not using real unlabeled data). As we can see, on the real Internet images, the proposed DAM-Net indeed shows better accuracy and qualitative results (highlighted in green boxes of Fig. 6(e)) than that without domain adaption. While on the synthetic test images, DAM-Net works worse than the baseline. As explained above, this is since that the baseline network is overfitting the synthetic data.

Fig. 7 shows additional qualitative comparison results to Sun et al. [STT21], the state-of-the-art on natural image matting on the Internet test dataset. This method takes an RGB image and a trimap as input, while our method uses an RGB image only. We use their release code and model to generate the matting results, as shown in Fig. 7c. Our results are significantly better than those of Sun et al. [STT21]. Experiments show that current general-purpose matting method may generate unsatisfactory results for eyebrow matting.

4.4. Results on Internet Images

Fig. 8 shows the qualitative results of our method on the Internet images. Since the synthetic eyebrow images contain images in different poses, illuminations, ages, races, eyebrow shapes, etc., our network is able to estimate high-quality eyebrow alpha mattes for portrait images taken in different environments on the Internet. The network can still generate plausible results when the eyebrow regions are blurred. Fig. 9 shows additional matting results on daily captured images with various variations, such as poses, illuminations, shadows, ages, races, and so on. The test images in Fig. 1,



(a) Input (b) Li and Lu [LL20] (c) Xiao et al. [XZZ*21] (d) Baseline network (e) DAM-Net (Ours) (f) Ground truth

Figure 6: *Exemplars of the qualitative results of 4 methods (Li and Lu [LL20], Xiao et al. [XZZ*21], the baseline network, and ours, respectively) applying on the baseline testing dataset (containing both synthetic and Internet images). From left to right, we present the input images, the results of Li and Lu [LL20] and its trimap inputs (white boxes in the left corner), Xiao et al. [XZZ*21], the baseline network, and ours, respectively. The details are highlighted in green boxes. Clearly, 'Ours' works significantly better than the 'Baseline' (see the first 4 rows, which are the real input images), showing the effectiveness of the proposed domain adaptation approach.*

Table 3: *The quantitative results of 4 methods (Li and Lu [LL20], Xiao et al. [XZZ*21], the baseline network and ours, respectively) applying on the Internet test dataset and the synthetic test dataset, respectively. Note that Li and Lu [LL20] use manually labeled trimaps (b) as input, while our method does not need such a time-consuming labeling process. Even so, our approach still exhibits better overall performance for eyebrow matting.*

Datasets	Internet image test dataset				Synthetic test dataset			
Methods	Li and Lu [LL20]	Xiao et al. [XZZ*21]	Baseline network	Ours	Li and Lu [LL20]	Xiao et al. [XZZ*21]	Baseline network	Ours
MSE	0.0814	0.0066	0.0067	**0.0055**	0.0962	0.0025	**0.0024**	0.0028
SAD	10.0597	6.3385	6.1414	**5.5842**	11.7339	3.3504	**3.1914**	3.3627
GRAD	8.5853	**3.3739**	4.1891	3.7007	8.5264	2.3407	**2.1364**	2.1881
Conn	**3.3124**	3.7289	3.7938	3.5920	5.1747	2.6902	**2.5787**	2.6347

(a) Input (b) Trimap (c) Sun et al. [STT21] (d) DAM-Net (Ours) (e) Ground truth

Figure 7: *Exemplars of the quantitative results of 2 methods (Sun et al. [STT21] and Ours) applying on the Internet testing dataset. From left to right, we present the input images, the corresponding trimap inputs, Sun et al. [STT21] and Ours, respectively. 'Ours' works significantly better than Sun et al. [STT21], showing the effectiveness of the proposed domain adaptation approach.*

Figure 8: *Exemplars of eyebrow alpha matte estimation on Internet images under different variations, such as poses, illuminations, gender, ages, races, etc.*

Fig. 5, Fig. 6 (rows 1-4), Fig. 7, Fig. 8, Fig. 9, and Fig. 10 are courtesy of unsplash.com [Uns22] and flickr.com [Fli22].

Applications. Eyebrow matting can be used in a variety of scenarios, including eyebrow editing and removal. We demonstrate several eyebrow editing applications, including eyebrow recoloring (Fig. 1d) and eyebrow replacement (Fig. 1e). Furthermore, eyebrow matting is useful for high-quality 3D face reconstruction (Fig. 1f-i). People typically scan a set of expression keyframes (e.g., 100+ expression scans) or continuous expression sequences (thousands of frames using MVS 4D scanner) in a MVS-based high-quality 3D face reconstruction pipeline. If the eyebrows are not removed, they may introduce artifacts during 3D parametric face reconstruction, necessitating costly manual repair in hours. However, eyebrow matting makes it simple to improve the geometry of the eyebrow region.

5. Discussion

Different from other general natural image matting methods [STT21, LXZ*21], our method is specifically designed for eyebrow matting. For completeness, we show additional qualitative comparison results to these two methods in the supplementary material. An advantage of our method is that the encoder-decoder network in DAM-Net can be easily replaced with other matting networks (e.g., [STT21, LXZ*21]). While our eyebrow matting network can achieve high-quality eyebrow matte estimation, it still has some restrictions. Due to the lack of ground truth for real-world eyebrow images (which is almost impossible to obtain), we use pseudo ground truth as previous works [XPCH17, QLY*20, STT21] for evaluation, thus we cannot provide an accurate test of the networks. Our model may matte out hair that covers the eyebrows, which is reasonable as eyebrows and hair are similar in terms of the underlying hair structure. However, in some cases, our matting network may fail (Fig. 10). For example, if the lighting is too dim, our model may fail. This can be fixed by increasing the brightness of the input images. Rare color lighting and unique eyebrow makeup are also difficult situations for our method. Despite the fact that our synthetic dataset includes training data for the elderly, our model may fail on images of old people's eyebrows with distinctive grizzled eyebrow color or deep skin wrinkles. However, we believe that increasing the diversity of the training dataset can improve the failure cases described above.

Figure 9: *Exemplars of alpha matte estimation on daily-captured images.*

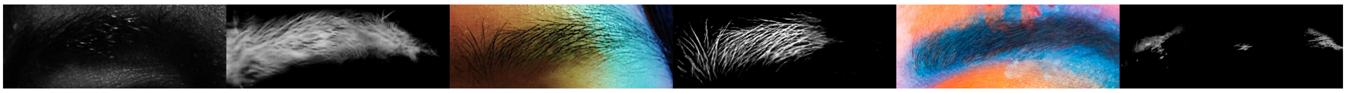


Figure 10: *Exemplars of failure cases. Our method may fail in cases of dim light, rare color lighting, or special makeup around the eyebrows.*

6. Conclusion and Future Work

In this paper, we render avatars to create the first eyebrow matting dataset. The dataset includes 1,000 synthetic eyebrow matting data and 1,215 unlabeled real-world eyebrow images for semi-supervised learning. We also present a domain adaptation matting network that learns semantic features from synthetic images and uses adversarial learning to adapt its representation to real-world images. The overall system is self-supervised, requiring no manual annotation of real matting data while achieving state-of-the-art eyebrow matting performance. Our method can estimate high-quality eyebrow alpha mattes from real-world portrait images with a variety of eyebrow styles, gender, races, expressions, illuminations, and poses. Furthermore, we manually annotate 68 eyebrow matting data from real-world images as a test dataset to validate our method. Through extensive experiments, we have demonstrated the effectiveness of our network and dataset. Results show that the network trained on our dataset outperforms prior methods and can estimate accurate eyebrow mattes on Internet images. We also show how our work can be applied to high-fidelity 3D face reconstruction and high-quality eyebrow editing. Our work makes a step towards high-quality eyebrow matting, and it has the potential to inspire other works in the field. In the future, we are interested in extending our method to other semantic matting tasks, such as portrait matting, human matting, etc.

Acknowledgements

Xiaogang Jin was supported by the Key Research and Development Program of Zhejiang Province (Grant No. 2020C03096), the National Natural Science Foundation of China (Grant No. 61972344), and the Ningbo Major Special Projects of the "Science and Technology Innovation 2025" (Grant No. 2020Z007).

References

[AAPS16] Aksoy Y., Aydin T. O., Pollefeys M., Smolić A.: Interactive high-quality green-screen keying via color unmixing. *ACM Trans. Graph. 35*, 5 (aug 2016), 152:1–152:12. 2, 3

[AOAP17] Aksoy Y., Ozan Aydin T., Pollefeys M.: Designing effective inter-pixel information flow for natural image matting. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (July 2017), pp. 228–236. doi:10.1109/CVPR.2017.32. 3

[AOP*18] Aksoy Y., Oh T., Paris S., Pollefeys M., Matusik W.: Semantic soft segmentation. *ACM Trans. Graph. 37*, 4 (2018), 72:1–72:13. doi:10.1145/3197517.3201275. 2

[BBN*12] Beeler T., Bickel B., Noris G., Beardsley P. A., Marschner S., Sumner R. W., Gross M. H.: Coupled 3d reconstruction of sparse facial hair and skin. *ACM Trans. Graph. 31*, 4 (2012), 117:1–117:10. 2

[CGX*18] Chen Q., Ge T., Xu Y., Zhang Z., Yang X., Gai K.: Semantic human matting. In *Proceedings of the 26th ACM International Conference on Multimedia* (2018), pp. 618–626. doi:10.1145/3240508.3240610. 3

[CWPC19] Chang W.-L., Wang H.-P., Peng W.-H., Chiu W.-C.: All about structure: Adapting structural information across domains for boosting semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2019), pp. 1900–1909. doi:10.1109/CVPR.2019.00200. 3

[Daz21] Daz Productions Inc: Daz3d - model, render, & animate. https://www.daz3d.com/, 2021. 4

[DDS*09] Deng J., Dong W., Socher R., Li L.-J., Li K., Fei-Fei L.: Imagenet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition* (2009), pp. 248–255. doi:10.1109/CVPR.2009.5206848. 5

[DJSXN19] Derouet-Jourdan A., Salvati M., Xing X., Nishikawa T.: Efficient mask expansion for green-screen keying using color distributions. In *ACM SIGGRAPH 2019 Posters* (New York, NY, USA, 2019), SIGGRAPH '19, Association for Computing Machinery, pp. 68:1–68:2. doi:10.1145/3306214.3338595. 2, 3

[DLS21] Dai Y., Lu H., Shen C.: Towards light-weight portrait matting via parameter sharing. *Computer Graphics Forum 40*, 1 (2021), 151–164. doi:10.1111/cgf.14179. 3

[Fli22] Flickr: Find your inspiration. https://www.flickr.com/, 2022. 9

[FLZ16] Feng X., Liang X., Zhang Z.: 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* (2016), vol. 9906 of *Lecture Notes in Computer Science*, Springer, pp. 204–219. doi:10.1007/978-3-319-46475-6_13. 3

[GDG*18] Goyal P., Dollár P., Girshick R., Noordhuis P., Wesolowski L., Kyrola A., Tulloch A., Jia Y., He K.: Accurate, large minibatch sgd: Training imagenet in 1 hour, 2018. arXiv:1706.02677. 5

[HTP*18] Hoffman J., Tzeng E., Park T., Zhu J.-Y., Isola P., Saenko K., Efros A., Darrell T.: Cycada: Cycle-consistent adversarial domain adaptation. In *International conference on machine learning* (2018), PMLR, pp. 1989–1998. doi:10.48550/arXiv.1711.03213. 3

[HWYY18] Hong W., Wang Z., Yang M., Yuan J.: Conditional generative adversarial network for structured domain adaptation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (June 2018). doi:10.1109/CVPR.2018.00145. 3

[HZRS16] He K., Zhang X., Ren S., Sun J.: Deep residual learning for image recognition. In *IEEE Conf. Computer Vision and Pattern Recognition* (2016), pp. 770–778. doi:10.1109/CVPR.2016.90. 3, 5

[HZZ*19] He T., Zhang Z., Zhang H., Zhang Z., Xie J., Li M.: Bag of tricks for image classification with convolutional neural networks. In *2019 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019* (2019), Computer Vision Foundation / IEEE, pp. 558–567. doi:10.1109/CVPR.2019.00065. 5

[KKS*21] Kundu J. N., Kulkarni A., Singh A., Jampani V., BABU R. V.: Generalize then adapt: Source-free domain adaptive semantic segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)* (October 2021), pp. 7046–7056. doi:10.1109/ICCV48922.2021.00696. 3

[LCD*19] LI J., CHEN E., DING Z., ZHU L., LU K., HUANG Z.: Cycle-consistent conditional adversarial transfer networks. In *Proceedings of the 27th ACM International Conference on Multimedia* (2019), pp. 747–755. doi:10.1145/3343031.3350902. 3

[LDSX19] LU H., DAI Y., SHEN C., XU S.: Indices matter: Learning to index for deep image matting. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)* (October 2019), pp. 3265–3274. doi:10.1109/ICCV.2019.00336. 3

[LH17] LOSHCHILOV I., HUTTER F.: Sgdr: Stochastic gradient descent with warm restarts, 2017. arXiv:1608.03983. 5

[LL20] LI Y., LU H.: Natural image matting via guided contextual attention. In *Proc. AAAI Conf. Artificial Intelligence* (2020), vol. 34, pp. 11450–11457. doi:10.1609/aaai.v34i07.6809. 2, 3, 5, 6, 7

[LLWL20] LEE C., LIU Z., WU L., LUO P.: Maskgan: Towards diverse and interactive facial image manipulation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2020* (2020), IEEE, pp. 5548–5557. doi:10.1109/CVPR42600.2020.00559. 3

[LRS*21] LIN S., RYABTSEV A., SENGUPTA S., CURLESS B. L., SEITZ S. M., KEMELMACHER-SHLIZERMAN I.: Real-time high-resolution background matting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (June 2021), pp. 8762–8771. doi:10.1109/CVPR46437.2021.00865. 2, 3

[LXZ*21] LIU Q., XIE H., ZHANG S., ZHONG B., JI R.: *Long-Range Feature Propagating for Natural Image Matting.* Association for Computing Machinery, New York, NY, USA, 2021, p. 526–534. doi:10.1145/3474085.3475203. 3, 9

[LYH*20] LIU J., YAO Y., HOU W., CUI M., XIE X., ZHANG C., HUA X.: Boosting semantic human matting with coarse annotations. In *IEEE/CVF Conf. Computer Vision and Pattern Recognition* (2020), pp. 8560–8569. doi:10.1109/CVPR42600.2020.00859. 3

[NWKS19] NAM G., WU C., KIM M. H., SHEIKH Y.: Strand-accurate multi-view hair capture. In *IEEE Conference on Computer Vision and Pattern Recognition* (2019), Computer Vision Foundation / IEEE, pp. 155–164. doi:10.1109/CVPR.2019.00024. 2

[PGM*19] PASZKE A., GROSS S., MASSA F., LERER A., BRADBURY J., CHANAN G., KILLEEN T., LIN Z., GIMELSHEIN N., ANTIGA L., DESMAISON A., KÖPF A., YANG E., DEVITO Z., RAISON M., TEJANI A., CHILAMKURTHY S., STEINER B., FANG L., BAI J., CHINTALA S.: Pytorch: An imperative style, high-performance deep learning library, 2019. arXiv:1912.01703. 5

[QLY*20] QIAO Y., LIU Y., YANG X., ZHOU D., XU M., ZHANG Q., WEI X.: Attention-guided hierarchical structure aggregation for image matting. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2020), IEEE, pp. 13673–13682. doi:10.1109/CVPR42600.2020.01369. 2, 3, 9

[RFB15] RONNEBERGER O., FISCHER P., BROX T.: 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* (2015), vol. 9351 of *Lecture Notes in Computer Science*, Springer, pp. 234–241. doi:10.1007/978-3-319-24574-4_28. 3

[RL18] REN Z., LEE Y. J.: Cross-domain self-supervised multi-task feature learning using synthetic imagery. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2018), pp. 762–771. doi:10.1109/CVPR.2018.00086. 2, 3

[RRW*09] RHEMANN C., ROTHER C., WANG J., GELAUTZ M., KOHLI P., ROTT P.: A perceptually motivated online benchmark for image matting. In *IEEE Conference on Computer Vision and Pattern Recognition* (2009), pp. 1826–1833. doi:10.1109/CVPR.2009.5206503. 2, 3, 5

[SB96] SMITH A. R., BLINN J. F.: Blue screen matting. In *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH 1996* (1996), ACM, pp. 259–268. doi:10.1145/237170.237263. 2, 3

[SBJ*18] SANKARANARAYANAN S., BALAJI Y., JAIN A., LIM S. N., CHELLAPPA R.: Learning from synthetic data: Addressing domain shift for semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (June 2018). doi:10.1109/CVPR.2018.00395. 3

[SJC*20] SENGUPTA S., JAYARAM V., CURLESS B., SEITZ S. M., KEMELMACHER-SHLIZERMAN I.: Background matting: The world is your green screen. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2020), pp. 2288–2297. doi:10.1109/CVPR42600.2020.00236. 3

[SLKS06] SUN J., LI Y., KANG S. B., SHUM H.-Y.: Flash matting. *ACM Trans. Graph. 25*, 3 (July 2006), 772–778. doi:10.1145/1179352.1141954. 3

[STG*16] SHEN X., TAO X., GAO H., ZHOU C., JIA J.: Deep automatic portrait matting. In *Computer Vision – ECCV 2016* (2016), Leibe B., Matas J., Sebe N., Welling M., (Eds.), Springer International Publishing, pp. 92–107. doi:10.1007/978-3-319-46448-0_6. 3

[STT21] SUN Y., TANG C.-K., TAI Y.-W.: Semantic image matting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (June 2021), pp. 11120–11129. doi:10.1109/CVPR46437.2021.01097. 2, 3, 6, 8, 9

[THSD17] TZENG E., HOFFMAN J., SAENKO K., DARRELL T.: Adversarial discriminative domain adaptation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (July 2017). 3

[Uns22] UNSPLASH: Unsplash: Photos for everyone. https://unsplash.com/, 2022. 9

[XPCH17] XU N., PRICE B., COHEN S., HUANG T.: Deep image matting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (July 2017), pp. 311–320. doi:10.1109/CVPR.2017.41. 2, 3, 5, 9

[XYH*21] XIAO C., YU D., HAN X., ZHENG Y., FU H.: Sketchhairsalon: Deep sketch-based hair image synthesis. *ACM Trans. Graph. 40*, 6 (dec 2021), 216:1–216:16. doi:10.1145/3478513.3480502. 2

[XZZ*21] XIAO Q., ZHANG H., ZHANG Z., WU Y., WANG L., JIN X., JIANG X., YANG Y.-L., SHAO T., ZHOU K.: Eyelashnet: A dataset and a baseline method for eyelash matting. *ACM Trans. Graph. 40*, 6 (dec 2021), 217:1–217:17. doi:10.1145/3478513.3480540. 2, 3, 5, 6, 7

[YSZZ19] YANG L., SHI Z., ZHENG Y., ZHOU K.: Dynamic hair modeling from monocular videos using deep neural networks. *ACM Trans. Graph. 38*, 6 (nov 2019), 235:1–235:2. doi:10.1145/3355089.3356511. 2

[YZZ*21] YU Q., ZHANG J., ZHANG H., WANG Y., LIN Z., XU N., BAI Y., YUILLE A.: Mask guided matting via progressive refinement network. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (June 2021), pp. 1154–1163. doi:10.1109/CVPR46437.2021.00121. 3

[ZAFW21] ZHU P., ABDAL R., FEMIANI J., WONKA P.: Barbershop: Gan-based image compositing using segmentation masks. *ACM Trans. Graph. 40*, 6 (dec 2021), 215:1–215:13. doi:10.1145/3478513.3480537. 2

[ZGF*19] ZHANG Y., GONG L., FAN L., REN P., HUANG Q., BAO H., XU W.: A late fusion cnn for digital matting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (June 2019), pp. 7469–7478. doi:10.1109/CVPR.2019.00765. 3