DOI: 10.1111/cgf.14762
EUROGRAPHICS 2023 / K. Myszkowski and M. Nießner
(Guest Editors)

COMPUTER GRAPHICS *forum*
Volume 42 (2023), Number 2

Makeup Extraction of 3D Representation via Illumination-Aware Image Decomposition

Xingchao Yang,[1,2] Takafumi Taketomi,[1] and Yoshihiro Kanamori[2]

[1]CyberAgent, AI Lab, Japan [2]University of Tsukuba, Japan

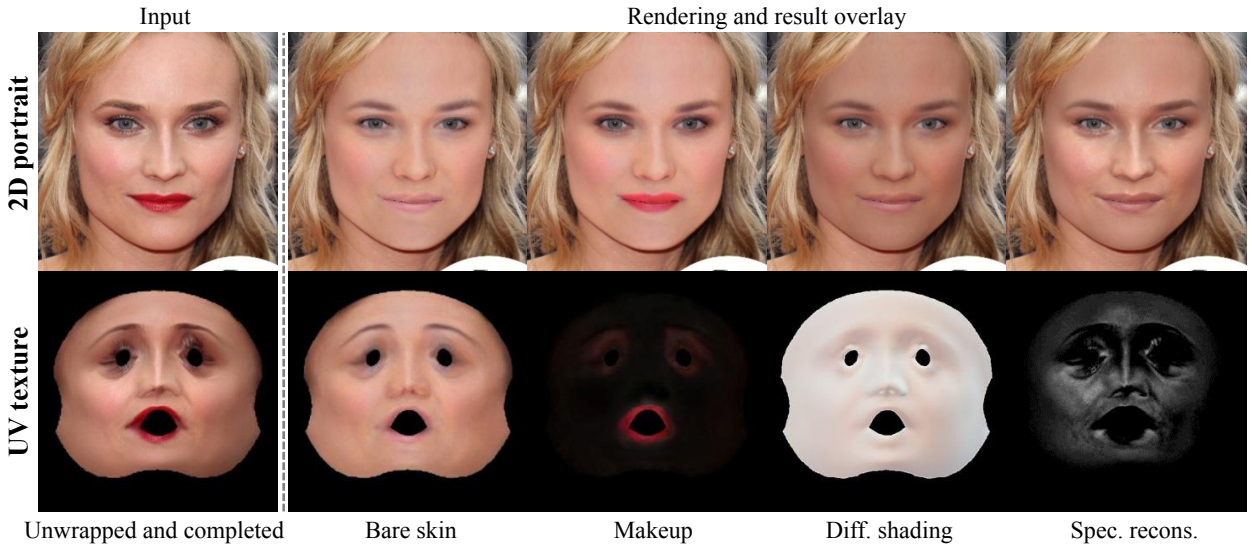

Figure 1: *Makeup-aware facial inverse rendering and component-wise reconstruction. The top row displays a makeup portrait input and overlaid rendering images (from left to right: bare skin only, bare skin plus makeup, bare skin multiplied by diffuse shading, and plus specular reconstruction) whereas the bottom row shows disentangled materials in the UV space.*

Abstract

Facial makeup enriches the beauty of not only real humans but also virtual characters; therefore, makeup for 3D facial models is highly in demand in productions. However, painting directly on 3D faces and capturing real-world makeup are costly, and extracting makeup from 2D images often struggles with shading effects and occlusions. This paper presents the first method for extracting makeup for 3D facial models from a single makeup portrait. Our method consists of the following three steps. First, we exploit the strong prior of 3D morphable models via regression-based inverse rendering to extract coarse materials such as geometry and diffuse/specular albedos that are represented in the UV space. Second, we refine the coarse materials, which may have missing pixels due to occlusions. We apply inpainting and optimization. Finally, we extract the bare skin, makeup, and an alpha matte from the diffuse albedo. Our method offers various applications for not only 3D facial models but also 2D portrait images. The extracted makeup is well-aligned in the UV space, from which we build a large-scale makeup dataset and a parametric makeup model for 3D faces. Our disentangled materials also yield robust makeup transfer and illumination-aware makeup interpolation/removal without a reference image.

CCS Concepts

• ***Computing methodologies*** → ***Computer graphics; Computer vision; Machine learning;***

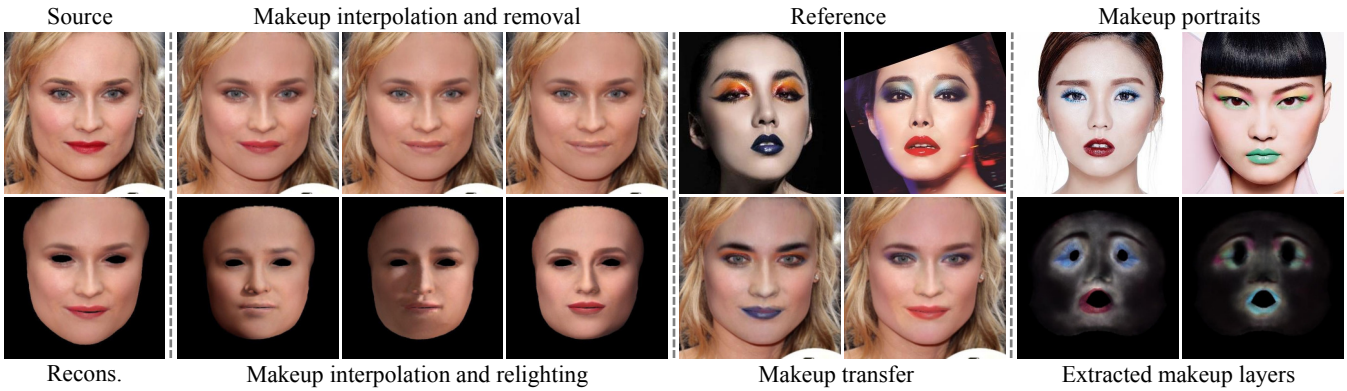

Figure 2: *Example applications of our method. From left to right, makeup controllable 3D face reconstruction, illumination-aware makeup interpolation/removal/relighting, illumination-aware makeup transfer, and makeup data collection.*

1. Introduction

Facial makeup is an art of enhancing human appearance, dating back to ancient times. Currently, it is quite commonly used for beautification purposes to improve the quality of life. Furthermore, facial makeup enriches the user experience of face-related applications such as VR/AR, video games, online commerce, and social camera apps. These trends have been driving active research on facial makeup in computer graphics and computer vision. In particular, research on makeup for 3D facial models has been gaining increasing attention in the movie and advertising industries for digital humans.

To obtain facial makeup for 3D characters, the following three approaches are currently available; 1) direct painting, 2) capturing of real-world makeup, and 3) makeup transfer from 2D images. Direct painting is labor-intensive for makeup artists and quite costly. Capturing real-world makeup requires special devices [SRH*11, HLHC13] and thus is also costly and not scalable. 2D makeup transfer [LQD*18,MZWX21] is the current mainstream of facial makeup research, exploiting a myriad of facial makeup photos available on the Internet. However, most existing studies have focused on 2D-to-2D transfer and struggled with physical constraints. For example, faces of in-the-wild photos frequently contain lighting effects such as specular highlights and shadows and occlusions with hands, possibly with various facial expressions and head poses. Recent 3D face reconstruction techniques [SKCJ18, EST*20] can handle these physical constraints, but none of the existing techniques focus on facial makeup.

In this paper, we propose the first integrated solution for extracting makeup for 3D facial models from a single portrait image. Fig. 1 shows example outputs of our method. Our method exploits the strong facial prior of the 3D morphable model (3DMM) [LBB*17] via regression-based inverse rendering and extracts coarse facial materials such as geometry and diffuse/specular albedos as UV textures. Unlike the existing regression-based techniques for facial inverse rendering [SKCJ18, EST*20], we further refine the coarse facial materials via optimization for higher fidelity. To alleviate the inherent skin color bias in the 3DMM, we also integrate skin color adjustment inspired by color transfer. From the refined diffuse albedo, we extract the bare skin, facial makeup, and an alpha matte. The alpha matte plays a key role in various applications such as manual tweaking of the makeup intensity and makeup interpolation/removal. The extracted makeup is well aligned in the UV space, from which we build a large-scale makeup texture dataset and a parametric makeup model using principal component analysis (PCA) for 3D faces. By overlaying rendered 3D faces onto portrait images, we can achieve novel applications such as illumination-aware (*i.e., relightable*) makeup transfer, interpolation, and removal, working on 2D faces (see Fig. 2).

The key contributions are summarized as follows:

- We present the first method to achieve illumination-aware makeup extraction for 3D face models from in-the-wild face images.
- We propose a novel framework that improves each of the following steps; (1) an extended 3D face reconstruction network that infers not only diffuse shading but also specular shading via regression, (2) a carefully designed inverse rendering method to generate high-fidelity textures without being restricted by the limited lighting setup, and (3) a novel procedure that is specially designed for extracting makeup by leveraging the makeup transfer technique. The UV texture representation effectively integrates these three modules into a single framework.
- Our extracted illumination-independent makeup of the UV texture representation facilitates many makeup-related applications. The disentangled maps are also editable forms. We employ the extracted makeup to build a PCA-based makeup model that is useful for 3D face reconstruction of makeup portraits.

2. Related Work

Our framework consists of three steps for extracting makeup from a portrait. It is related to recent approaches in terms of three aspects. First, we discuss facial makeup-related research. Subsequently, we review intrinsic image decomposition which can separate the input image into several elements. Finally, we discuss 3D face reconstruction methods that generate a 3D face model from a single portrait image.

2.1. Facial Makeup

Facial makeup recommendation and makeup transfer methods have been developed in the computer graphics and computer vision research fields. Makeup recommendation methods can provide appropriate makeup for faces. The method in [SRH*11] captured 56 female faces with and without facial makeup. Suitable makeup was recommended by analyzing the principal components of the makeup and combining them with the facial appearance. An examples-rules guided deep neural network for makeup recommendation has been designed [AJWF17]. Professional makeup knowledge was combined with the corresponding before and after makeup data to train the network. However, it is challenging to collect a large-scale makeup dataset by following the existing methods. Recently, we proposed BareSkinNet [YT22], which can remove makeup and lighting effects from input face images. However, BareSkinNet cannot be used for subsequent makeup applications because it discards the makeup patterns and specular reflection. In contrast, our method can automatically extract the makeup layer information from portrait image inputs. As a result, a large-scale makeup dataset can be constructed.

The aim of makeup transfer is to transfer the makeup of a person to others. Deep learning technologies have significantly accelerated makeup transfer research. BeautyGAN [LQD*18] is a method based on CycleGAN [ZPIE17] that does not require a before and after makeup image pair. A makeup loss function was also designed to calculate the difference between makeups. The makeup loss is a histogram matching loss that approximates the color distribution of the relative areas of different faces. Subsequent approaches that have used BeautyGAN as the baseline can solve more challenging problems. LADN [GWC*19] achieves heavy makeup transfer and removal, whereas PSGAN [JLG*20] and SCGAN [DHC*21] solve the problem of makeup transfer under different facial expressions and head poses. CPM [NTH21] is a color and pattern transfer method and SOGAN [LDP*21] is a shadow and occlusion robust GAN for makeup transfer. EleGANt [YHXG22] is a locally controllable makeup transfer method. Makeup transfer can also be incorporated into the Virtual-Try-On applications [KGPB20, KJB*22]. However, existing methods do not consider illumination and can only handle 2D images. Our method takes advantage of the makeup transfer technique. The extracted makeup is disentangled into bare skin, facial makeup, and illumination in the UV space. Furthermore, the extracted makeup is editable.

2.2. Intrinsic Image Decomposition

We mainly review the recent intrinsic image decomposition methods relating to the portrait image input. An intrinsic image decomposition method [LZL15] has been employed, thereby enabling accurate and realistic makeup simulation from a photo. SfSNet [SKCJ18] is an end-to-end network that decomposes face images into the shape, reflectance, and illuminance. This method uses real and synthetic images to train the network. Inspired by SfSNet, Relighting Humans [KE18] attempts to infer a light transport map to solve the problem of light occlusion from an input portrait. The aforementioned methods can handle only diffuse reflection using spherical harmonic (SH) [RH01] lighting. Certain methods use a light stage to obtain a large amount of portrait data with illumination [SBT*19, POL*21, WYL*20]. The global illumination can be inferred by learning these data. As data collection using a light stage requires substantial resources, several more affordable methods have been proposed [TKE21, LSY*21, JYG*22, WWR22, TFM*22, YNK*22]. In this study, we focus on makeup with the aim of decomposing a makeup portrait into bare skin, makeup, diffuse, and specular layers without using light stage data.

2.3. Image-Based 3D Face Reconstruction

3D face reconstruction from a single-view portrait is challenging because in-the-wild photos always contain invisible areas or complex illumination. Existing 3D face reconstruction methods [BV03, THMM17, TZK*17, TZG*18, GZC*19, GCM*18, SBFB19, DYX*19, SSL*20, DBA*21, FFBB21, DBB22, ZBT22] generally use a parametric face model (known as a 3DMM) [BV99, LBB*17, GMB*18, BRP*18] to overcome this problem. In general, 3D face reconstruction is achieved by fitting the projected 3DMM to the input image. These methods estimate the SH lighting while inferring the shape and texture. We recommend that readers refer to [EST*20, SL09], as these surveys provide a more comprehensive description of the 3DMM and 3D face reconstruction.

The 3DMM-based 3D face reconstruction method estimates coarse facial materials and cannot achieve a high-fidelity facial appearance. A detailed 3D face reconstruction method was proposed in [DBA*21], which can reconstruct the roughness and specular compared to the previous methods. This method involves the setup of a virtual light stage of illumination and the use of a two-stage coarse-to-fine technique to refine the facial materials. Inspired by [DBA*21], we employ coarse facial materials and take specular into consideration. Furthermore, we extend and train a deep learning network using the method in [DYX*19] as the backbone with a large-scale dataset. We optimize the textures in UV space to refine facial materials, to solve the problem of self-occlusion and obstacles of the face. The completed UV texture is advantageous because the complete makeup can be extracted to achieve high-fidelity makeup reconstruction.

The generation of a completed facial UV texture is a technique for obtaining refined facial materials. Several methods use UV texture datasets to train an image-translation network for the generation of the completed texture via supervised learning [SWH*17, YSN*18, DCX*18, GPKZ19, LL20, LL20, LMG*20, BLC*22]. Other methods [CHS22, GDZ21] use the GAN inversion [XZY*22] technique to generate faces in different directions for completion. However, both of these methods are limited by the training dataset. We adopt a state-of-the-art UV completion method known as DSD-GAN [KYT21] to obtain a completed high-fidelity face texture, which is to be used as the objective for the optimization of the refined facial materials. This method employs self-supervised learning to fill the missing areas without the need for paired training data.

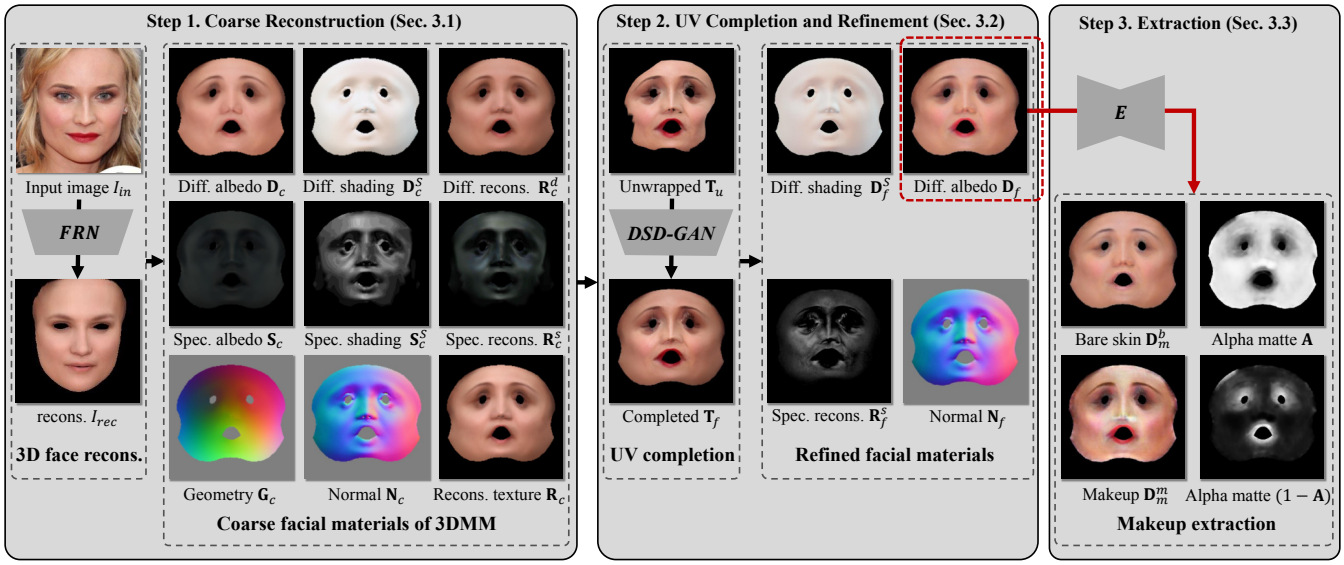

Figure 3: *Overview of our framework. Given a makeup portrait, we extract an illumination-independent bare skin and makeup in the UV space via the following three steps; First, we reconstruct a 3D face to estimate the coarse facial materials using a 3D face reconstruction network FRN [DYX*19] (Sec. 3.1). Second, we refine the coarse facial materials, which may have missing pixels owing to occlusion. We apply an inpainting network DSD-GAN [KYT21] and then apply optimization (Sec. 3.2). Finally, we extract a bare skin, makeup, and an alpha matte from the refined diffuse albedo using makeup extraction network E (Sec. 3.3).*

3. Approach

We design a coarse-to-fine texture decomposition process. As shown in Fig. 3, our framework is composed of three steps. 1) We estimate the coarse 3D facial materials using 3D face reconstruction by 3DMM fitting. In order to handle highlights, we extend a general 3DMM fitting algorithm. The reconstructed coarse facial materials are used for the refinement process in the next step (Sec. 3.1). 2) We propose an inverse rendering method to obtain refined 3D facial materials via optimization. First, we use the 3DMM shape that is obtained in the previous step to sample the image in UV space. Subsequently, a UV completion method is employed to obtain a high-fidelity entire face texture. Finally, using the completed texture as the objective, and coarse facial materials as priors, the refined facial materials are optimized. (Sec. 3.2). 3) The refined diffuse albedo that is obtained from the previous step is used as the input. Inspired by the makeup transfer technique, we design a network to disentangle bare skin and makeup. The key idea is that the makeup albedo can be extracted using an alpha blending manner for bare skin and makeup (Sec. 3.3). The details of each step are described in the following subsections.

3.1. Coarse Facial Material Reconstruction

We obtain the coarse facial materials using a 3D face reconstruction method based on regression-based inverse rendering. We use the FLAME [LBB*17] model with specular albedo from AlbedoMM [SSD*20]. The diffuse and specular albedo of FLAME are defined in the UV texture space. We only use the facial skin region in our study. Compared to existing methods, we extend the capability of the 3D face reconstruction network [DYX*19] (*FRN*) to estimate the shape, diffuse albedo, and diffuse shading, as well as the specular albedo and specular shading. Inspired by [DBA*21], a simplified virtual light stage with regular icosahedral parallel light sources is set up to infer the specular shading. The intensity of 20 light sources is predicted during the reconstruction process, and the direction of the light sources can be adjusted slightly. Furthermore, In order to eliminate the limited color range of the skin tone of FLAME, we estimate the skin tone adjustment parameters to ensure a diffuse albedo that is similar to the original image. The skin tone ablation study is depicted in Fig. 10. This process can improve the diversity of the diffuse albedo representation capability of FLAME.

The coarse shape geometry $\mathbf{G}_c$, diffuse albedo $\mathbf{D}_c$, and specular albedo $\mathbf{S}_c$ of the 3DMM are defined as follows:

$$\mathbf{G}_c = \bar{\mathbf{G}} + \mathbf{B}_{id}\alpha + \mathbf{B}_{ex}\beta, \tag{1}$$

$$\mathbf{D}_c = \bar{\mathbf{D}} + \mathbf{B}_d\gamma \odot \mathbf{C}_{gain} + \mathbf{C}_{bias}, \tag{2}$$

$$\mathbf{S}_c = \bar{\mathbf{S}} + \mathbf{B}_s\delta, \tag{3}$$

where $\bar{\mathbf{G}}$, $\bar{\mathbf{D}}$, and $\bar{\mathbf{S}}$ are the average geometry, diffuse albedo, and specular albedo, respectively. $\odot$ denotes the Hadamard product. The subscript c indicates coarse facial materials. Moreover, $\mathbf{B}_{id}$, $\mathbf{B}_{ex}$, $\mathbf{B}_d$, and $\mathbf{B}_s$ are the PCA basis vectors of the identity, expression, diffuse albedo, and specular albedo, respectively, whereas $\alpha \in \mathbb{R}^{200}$, $\beta \in \mathbb{R}^{100}$, $\gamma \in \mathbb{R}^{100}$, and $\delta \in \mathbb{R}^{100}$ are the corresponding parameters for controlling the geometry and reflectance of a 3D face. Finally, $\mathbf{C}_{gain}$ and $\mathbf{C}_{bias}$ are the skin tone adjustment parameters.

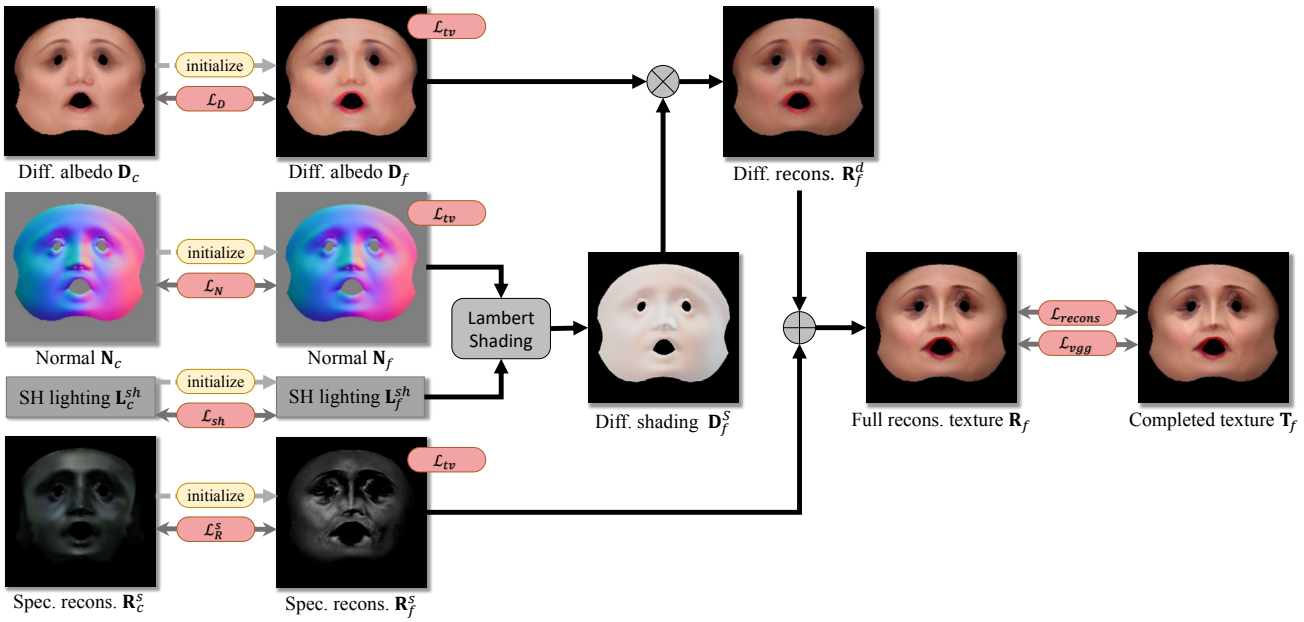

Figure 4: *Facial material optimization module for Step 2 (Sec. 3.2). We optimize the coarse facial materials* $\mathbf{D}_c, \mathbf{N}_c, \mathbf{R}_c^s$ *and SH lighting* $\mathbf{L}_c^{sh}$ *so that the full reconstruction* $\mathbf{R}_f$ *resembles the completed texture* $\mathbf{T}_f$*. The refined diffuse albedo* $\mathbf{D}_f$*, normal* $\mathbf{N}_f$*, and specular reconstruction* $\mathbf{R}_f^s$ *are the outputs.* $\oplus$ *and* $\otimes$ *denote the per-pixel addition and multiplication, respectively.*

Our reconstructed texture can be formulated as follows:

$$\mathbf{R}_c = \mathbf{R}_c^d + \mathbf{R}_c^s \tag{4}$$

$$= \mathbf{D}_c \odot \mathbf{D}_c^S + \mathbf{S}_c \odot \mathbf{S}_c^S, \tag{5}$$

where $\mathbf{R}_c^d$, $\mathbf{R}_c^s$, $\mathbf{D}_c^S$, and $\mathbf{S}_c^S$ are the diffuse reconstruction, specular reconstruction, diffuse shading, and specular shading, respectively. In this paper, a geometrically-derived shading component is referred to as "shading," wheras a multiplication with reflectance is dubbed "reconstruction." Using the normal $\mathbf{N}_c$ and second-order SH lighting coefficients $\mathbf{L}_c^{sh} \in \mathbb{R}^{27}$, the diffuse shading is calculated following the Lambertian reflectance model [BJ03]. The normal $\mathbf{N}_c$ is computed from the geometry $\mathbf{G}_c$. The specular shading is calculated following the Blinn–Phong reflection model [Bli77]. We define the 20 light sources of the virtual light stage with the light intensity $\mathbf{L}_i \in \mathbb{R}^{20}$ and light direction $\mathbf{L}_d \in \mathbb{R}^{60}$. $\rho \in \mathbb{R}^{20}$ is the exponent that controls the shininess. We employ the 3D face reconstruction network implementation of [DYX*19], and extend it to regress the parameters $\chi = (\alpha, \beta, \gamma, \delta, \mathbf{C}_{gain}, \mathbf{C}_{bias}, \mathbf{r}, \mathbf{t}, \mathbf{L}_c^{sh}, \mathbf{L}_i, \mathbf{L}_d, \rho)$, where $\mathbf{r} \in \mathbb{R}^3$ is the face rotation, and $\mathbf{t} \in \mathbb{R}^3$ is the face translation. Using the geometry $\mathbf{G}_c$, reconstructed texture $\mathbf{R}_c$, rotation $\mathbf{r}$, and translation $\mathbf{t}$, the reconstructed image I_{rec} can be rendered.

We refer to [DYX*19] to set up the loss functions for the model training as follows:

$$\begin{aligned}\mathcal{L}_c(\chi) = &\omega_{photo}\mathcal{L}_{photo}(\chi) + \omega_{lan}\mathcal{L}_{lan}(\chi) \\ &+ \omega_{skin}\mathcal{L}_{skin}(\mathbf{C}_{gain}, \mathbf{C}_{bias}) \\ &+ \omega_{reg}\mathcal{L}_{reg}(\alpha, \beta, \gamma, \delta, \mathbf{L}_i, \mathbf{L}_d),\end{aligned} \tag{6}$$

where $\mathcal{L}_{photo}(\chi)$ is the L1 pixel loss between the skin region of input image I_{in} and reconstructed image I_{rec}. $\mathcal{L}_{lan}(\chi)$ is the L2 loss between detected landmarks from I_{in} and the projected landmarks of the 3DMM. $\mathcal{L}_{skin}(\mathbf{C}_{gain}, \mathbf{C}_{bias})$ is the L1 loss for computing the mean color error between the skin region of I_{in} and diffuse albedo $\mathbf{D}_c$. This is our specially designed loss term to adjust the skin tone. $\mathcal{L}_{reg}(\alpha, \beta, \gamma, \delta, \mathbf{L}_i, \mathbf{L}_d)$ is the regulation loss for preventing a failed face reconstruction result. In contrast to the previous method [DYX*19], we extend constraints on the specular related coefficients of δ, $\mathbf{L}_i$, and $\mathbf{L}_d$. $\mathcal{L}_{reg}(\alpha, \beta, \gamma, \delta, \mathbf{L}_i, \mathbf{L}_d)$ is determined as follows:

$$\begin{aligned}\mathcal{L}_{reg} = &\omega_\alpha \|\alpha\|_2^2 + \omega_\beta \|\beta\|_2^2 + \omega_\gamma \|\gamma\|_2^2 \\ &+ \omega_\delta \|\delta\|_2^2 + \omega_L \|\mathbf{L}_i\|_2^2 + \omega_L \|\mathbf{L}_d\|_2^2,\end{aligned} \tag{7}$$

where $\|\cdot\|_2$ denotes the L2 norm. The balance weights are set to $\omega_{photo} = 19.2$, $\omega_{lan} = 5$, $\omega_{skin} = 3$, $\omega_{reg} = 3 \times 10^{-4}$, $\omega_\alpha = 1.0$, $\omega_\beta = 0.8$, $\omega_\gamma = 1.7 \times 10^{-2}$, $\omega_\delta = 1.0$, and $\omega_L = 1.0$ in all experiments.

3.2. UV Completion and Facial Material Refinement

The goal of this step is to obtain the disentangled refined facial materials. We use the geometry $\mathbf{G}_c$ to sample the colors from the input image I_{in} and project them to the UV texture space $\mathbf{T}_u$. $\mathbf{T}_u$ contains the missing area owing to self-occlusion or obstacles. As opposed to the 3D vertex color sampling approach used in related work [NTH21, CHS22], we adopt the image-to-UV rendering approach used in DSD-GAN [KYT21] to obtain a high-quality texture. The direct use of incomplete textures will cause many problems, such as noise and error. Thereafter, we fill the missing areas following DSD-GAN to obtain the completed UV

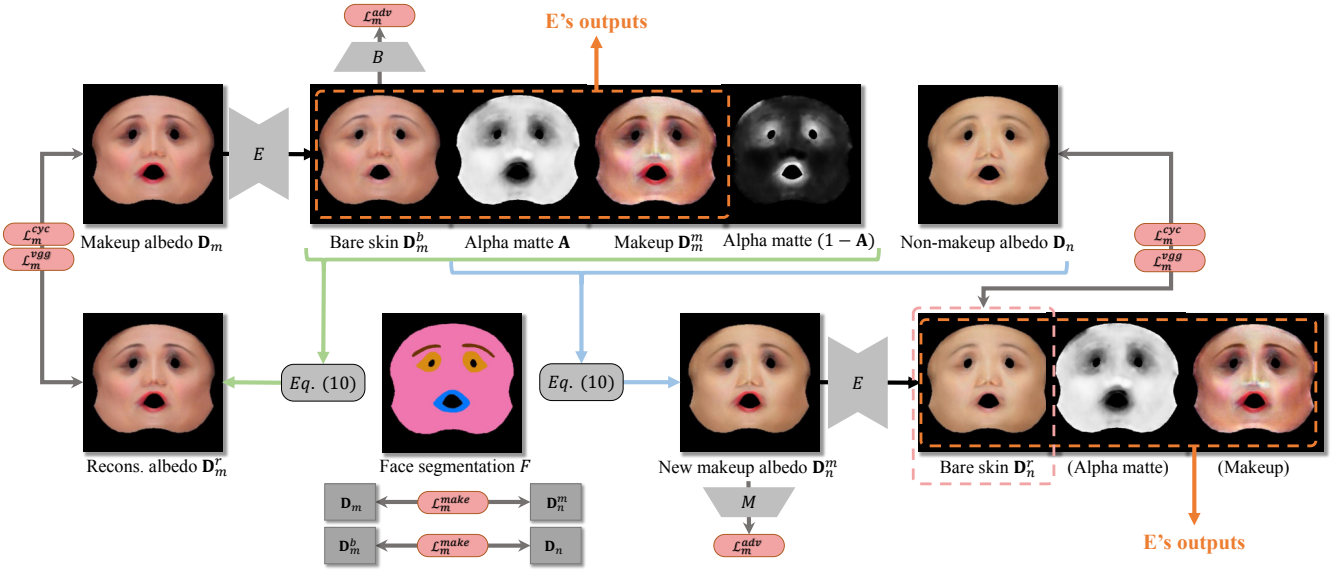

Figure 5: *Makeup extraction with network E for Step 3 (Sec. 3.3). Considering alpha blending in mind, the network E decomposes the refined diffuse albedo into bare skin, alpha matte, and makeup.*

texture $\mathbf{T}_f$. The subscript f indicates refined facial materials. The difference between our method and DSD-GAN is that we use the FLAME model for implementation, whereas DSD-GAN uses the BFM model [GMB*18]. The results of the UV completion are presented in Fig. 6.

Using the completed UV texture as an objective with facial details, the optimization-based refinement module (see Fig. 4) is designed. Given the target texture $\mathbf{T}_f$, the coarse prior $\mathbf{D}_c$, $\mathbf{N}_c$, $\mathbf{L}_c^{sh}$, and $\mathbf{R}_c^s$ are used for initialization, and we optimize the refined materials of $\mathbf{D}_f$, $\mathbf{N}_f$, $\mathbf{L}_f^{sh}$, and $\mathbf{R}_f^s$ to reconstruct the refined texture $\mathbf{R}_f$. The diffuse shading is calculated following the Lambertian reflectance model [BJ03]. For the specular, we directly optimize the specular reconstruction $\mathbf{R}_f^s$, rather than the specular albedo and specular shading, which is not restricted by the light source settings of the virtual light stage in Sec. 3.1. Note that the specular reconstruction $\mathbf{R}_c^s$ is three channels image due to the specular albedo $\mathbf{S}_c$ of 3DMM, while $\mathbf{R}_f^s$ is converted to one channel image for stable and efficient optimization.

The loss functions for the optimization are calculated as follows:

$$\begin{aligned}\mathcal{L}_f(\Psi) = \omega_{recons}\mathcal{L}_{recons}(\mathbf{R}_f) + \omega_{vgg}\mathcal{L}_{vgg}(\mathbf{R}_f) \\ + \omega_{tv}\mathcal{L}_{tv}(\mathbf{D}_f, \mathbf{N}_f, \mathbf{R}_f^s) \\ + \omega_{prior}\mathcal{L}_{prior}(\mathbf{D}_f, \mathbf{N}_f, \mathbf{R}_f^s, \mathbf{L}_f^{sh}),\end{aligned} \tag{8}$$

where $\Psi = (\mathbf{R}_f, \mathbf{D}_f, \mathbf{N}_f, \mathbf{R}_f^s, \mathbf{L}_f^{sh})$ is a new parameter set for optimization. $\mathcal{L}_{recons}(\mathbf{R}_f)$ ensures L1 consistency between $\mathbf{R}_f$ and $\mathbf{T}_f$. $\mathcal{L}_{vgg}(\mathbf{R}_f)$ is the perceptual loss [JAF16] that aims to preserve the facial details. $\mathcal{L}_{tv}(\mathbf{D}_f, \mathbf{N}_f, \mathbf{R}_f^s)$ is the total variation loss [GEB16] that encourages spatial smoothness in the optimized textures $\mathbf{D}_f$, $\mathbf{N}_f$, and $\mathbf{R}_f^s$.

$$\begin{aligned}\mathcal{L}_{prior}(\mathbf{D}_f, \mathbf{N}_f, \mathbf{R}_f^s, \mathbf{L}_f^{sh}) = \omega_D\mathcal{L}_D(\mathbf{D}_f) + \omega_N\mathcal{L}_N(\mathbf{N}_f) \\ + \omega_R^s\mathcal{L}_R^s(\mathbf{R}_f^s) + \omega_{sh}\mathcal{L}_{sh}(\mathbf{L}_f^{sh}),\end{aligned} \tag{9}$$

where $\mathcal{L}_{prior}(\mathbf{D}_f, \mathbf{N}_f, \mathbf{R}_f^s, \mathbf{L}_f^{sh})$ regulates the optimized texture to be similar to the coarse prior. We compute the L1 loss over the diffuse albedo, normal, and specular reconstruction as $\mathcal{L}_D(\mathbf{D}_f)$, $\mathcal{L}_N(\mathbf{N}_f)$, and $\mathcal{L}_R^s(\mathbf{R}_f^s)$, respectively. The coarse textures are resized to the same resolution as that of refined textures. We note that the coarse and refined textures are not exactly equal; they differ in clarity, and with or without makeup. Therefore, before calculating the losses between the coarse and refined textures, we simply use a Gaussian blur filter K to blur the refined textures in each iteration. We use the blurred refined texture to calculate the loss, while the original refined texture is not changed. $\mathcal{L}_{sh}(\mathbf{L}_f^{sh})$ is the L2 loss over the SH lighting. We normalize $\mathbf{N}_f$ to [-1, 1] following each optimization iteration to guarantee the correctness of the normal. The parameters $\omega_{recons} = 40$, $\omega_{vgg} = 5$, $\omega_{tv} = 10$, $\omega_{prior} = 1.0$, $\omega_D = 4$, $\omega_N = 1.0$, $\omega_R^s = 1.0$, and $\omega_{sh} = 1.0$ balance the importance of the terms. The refined textures of our coarse-to-fine optimization step are illustrated in Figs. 6, 7, and 8.

3.3. Makeup Extraction

In this step, we only use the refined diffuse albedo $\mathbf{D}_f$, and decompose the texture into the makeup, bare skin, and an alpha matte. To train the network, we create two diffuse albedo datasets with and without makeup following the previous process; the makeup albedo $\mathbf{D}_m$ and non-makeup albedo $\mathbf{D}_n$, respectively.

As shown in Fig. 5, we design a makeup extraction network based on alpha blending. Our network consists of a makeup extractor E, a makeup discriminator M, and a bare skin discriminator B.

M and B attempt to distinguish the makeup and non-makeup image. The core idea is that we regard the diffuse albedo as a combination of makeup and bare skin that is achieved by alpha blending. Therefore, E extracts the bare skin $\mathbf{D}_m^b$, makeup $\mathbf{D}_m^m$, and alpha matte $\mathbf{A}$. $\mathbf{A}$ is used to blend the extracted makeup $\mathbf{D}_m^m$ and a non-makeup albedo $\mathbf{D}_n$ to generate a new makeup albedo $\mathbf{D}_n^m$ which has the identity from $\mathbf{D}_n$ and makeup from $\mathbf{D}_m$. The reconstructed makeup albedo $\mathbf{D}_m^r$ and the reconstructed bare skin $\mathbf{D}_n^r$ should be consistent with their original input. The reconstructed $\mathbf{D}_m^r$ and generated $\mathbf{D}_n^m$ are formulated as follows:

$$\begin{aligned} \mathbf{D}_m^r &= \mathbf{A} \odot \mathbf{D}_m^b + (1-\mathbf{A}) \odot \mathbf{D}_m^m, \\ \mathbf{D}_n^m &= \mathbf{A} \odot \mathbf{D}_n + (1-\mathbf{A}) \odot \mathbf{D}_m^m, \end{aligned} \tag{10}$$

where $(1-\mathbf{A})$ is an inverted version of $\mathbf{A}$ with pixel values of $[0,1]$. The discriminators ensure that the generated bare skin $\mathbf{D}_m^b$ and makeup $\mathbf{D}_n^m$ are reliable. As the network takes advantage of the uniformity of the UV space, the makeup transfer is straightforward, and the problem of face misalignment does not need to be considered.

The loss function is given by:

$$\begin{aligned} \mathcal{L}_m(\Phi) = {} & \omega_m^{cyc}\mathcal{L}_m^{cyc}(\mathbf{D}_m^r, \mathbf{D}_n^r) + \omega_m^{vgg}\mathcal{L}_m^{vgg}(\mathbf{D}_m^r, \mathbf{D}_n^r) \\ & + \omega_m^{adv}\mathcal{L}_m^{adv}(\mathbf{D}_m^b, \mathbf{D}_n^m) + \omega_m^{tv}\mathcal{L}_m^{tv}(\mathbf{D}_m^b, \mathbf{D}_n^m) \\ & + \omega_m^{make}\mathcal{L}_m^{make}(\mathbf{D}_m^b, \mathbf{D}_n^m), \end{aligned} \tag{11}$$

where $\Phi = (\mathbf{D}_m^r, \mathbf{D}_n^r, \mathbf{D}_m^b, \mathbf{D}_n^m)$, $\mathcal{L}_m^{cyc}(\mathbf{D}_m^r, \mathbf{D}_n^r)$ and $\mathcal{L}_m^{vgg}(\mathbf{D}_m^r, \mathbf{D}_n^r)$ are the L1 loss and perceptual loss for the reconstruction, respectively. $\mathcal{L}_m^{adv}(\mathbf{D}_m^b, \mathbf{D}_n^m)$ is the adversarial loss for the discriminators and generators. $\mathcal{L}_m^{tv}(\mathbf{D}_m^b, \mathbf{D}_n^m)$ is the total variation loss for $\mathbf{D}_m^b$ and $\mathbf{D}_m^m$ to provide smooth texture generation, whereas $\mathcal{L}_m^{make}(\mathbf{D}_m^b, \mathbf{D}_n^m)$ is the makeup loss introduced by BeautyGAN [LQD*18]. We defined the corresponding face region F of the brows, eyes, and lips in the UV texture space to compute makeup loss, and the compared textures are $\mathbf{D}_m$ and $\mathbf{D}_n^m$, and $\mathbf{D}_n$ and $\mathbf{D}_m^b$. Note that, in contrast with previous makeup transfer methods, the skin region is not calculated for the makeup loss because we believe that the difference in skin tone between individuals cannot be considered as makeup, while it can also be a restriction if the foundation of the makeup changes the entire face color. This is a trade-off between skin tone and makeup color. In this work, we assume that the makeup will not drastically change the skin tone. Moreover, as indicated in Fig. 7, even without the makeup loss of the skin region, our network precisely extracts the makeup of the cheeks using the alpha blending approach because we use two discriminators to distinguish the makeup and non-makeup textures. The details are discussed in Sec. 5.

We use $\omega_m^{cyc} = 20$, $\omega_m^{vgg} = 2$, $\omega_m^{adv} = 5$, $\omega_m^{tv} = 8$, and $\omega_m^{make} = 1$ as the balancing terms. Our makeup extraction results are presented in Figs. 6, 7, and 8.

4. Implementation Details

4.1. Coarse Facial Material Reconstruction

We followed the training strategy of [DYX*19] and use the same datasets for approximately 260K face images. We used [BT17] to detect and crop the faces for alignment. The image size was 256×256 and the resolution of FLAME albedo textures was 256×256. We initialized the network with the weights of the pretrained [RDS*15] and modified the last layer to estimate our own 3DMM parameters. The batch size was 8, and the learning rate was 1×10^{-4} using an Adam optimizer with 20 training epochs. For the shininess parameter ρ, we set the initial value to 200 to achieve highlight effects.

4.2. UV Completion and Facial Material Refinement

We sampled approximately 100K images from FFHQ [KLA19] and CelebA-HQ [KALL18] in the UV space to train DSD-GAN for the UV completion, and the resolution of the UV textures was 512×512. Prior to sampling, the face images were segmented using the method of [YWP*18] and only the skin region of the face was used. Subsequently, we performed a manual cleanup to remove the low-quality textures. Eventually, 60,073 textures remained for training.

The optimization-based refinement process was executed with 500 iterations for each texture. The learning rate of the Adam optimizer was set to 1×10^{-2} with a 0.1 learning rate decay. The kernel size of the Gaussian blur filter K was set to 11.

4.3. Makeup Extraction

We combined two makeup datasets, namely the MT dataset [LQD*18] and LADN dataset [GWC*19], which consisted of 3,070 makeup images and 1,449 non-makeup images. A total of 300 makeup images were randomly selected for testing. By implementing the previous steps, both the makeup and non-makeup images were processed into albedo textures to train our network. We trained the network with 40 epochs and batch size of 1. The Adam optimizer used a learning rate of 1×10^{-4}. The makeup extractor E had the same architecture as the generator of DSD-GAN, and PatchGAN [IZZE17] was used for the discriminators M and B.

We used Nvdiffrast [LHK*20] for the differentiable renderer and trained the networks using a single NVIDIA GeForce RTX 2080 Ti GPU. Our training required approximately 3 days for the coarse facial reconstruction network and approximately 1 day for the makeup extraction network. Approximately 1 minute was required to process a texture in the refinement step.

5. Experiments

We evaluated the results of our approach. First, we present the intermediate outputs of each step of the framework (see Fig. 6). Thereafter, we discuss the final outputs (see Fig. 7). Subsequently, we analyze the albedo texture that was associated with the makeup and observe how the makeup changed in the albedo texture (see Fig. 8), Finally, we provide several examples with complex illumination and examine the decomposition (see Fig. 9). Note that the original images of the specular reconstruction were too dark to display, so we adjusted the contrast for better display. Please refer to the supplemental material to see the original images.

Intermediate outputs of each step. Fig. 6 depicts the intermediate outputs, which are related to the final textures in (d), to illustrate the effectiveness of each step. Four makeup portraits are presented

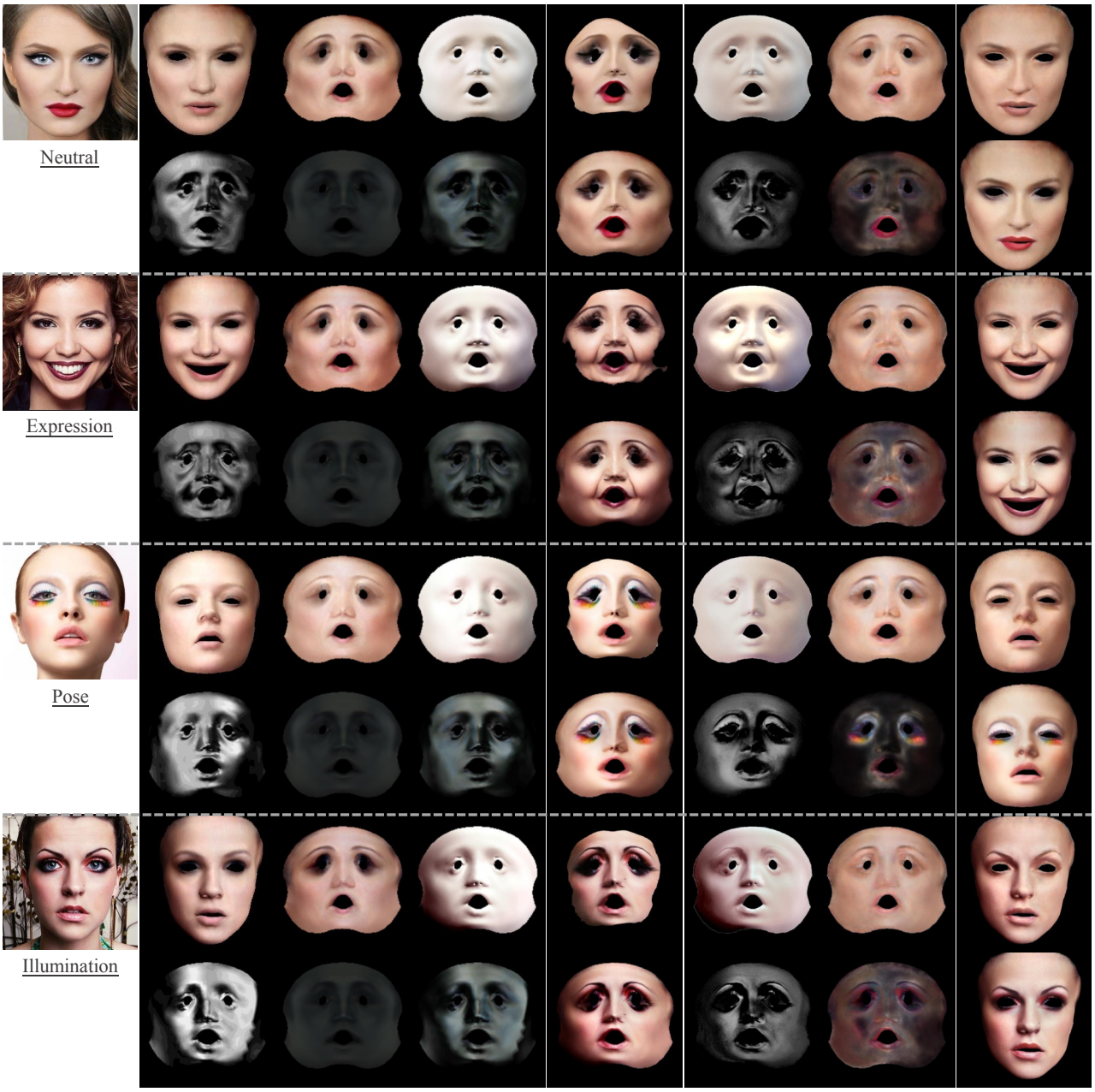

Figure 6: *Intermediate outputs for each step. Left to right and top to bottom: (a) input makeup portraits, (b) facial inverse rendering (diffuse albedo, diffuse shading, specular shading, specular albedo, and specular reconstruction), (c) unwrapped and completed textures, (d) final refined facial materials (diffuse shading, bare skin, specular reconstruction, and makeup), and (e) rendered bare skin face and rendered makeup face using textures from (d).*

in (a) with different facial features: a neutral face, a face with expression, a face pose with an angle, and a face with uneven illumination. The results of the 3D face reconstruction are shown in (b). It can be observed that the diffuse albedo of the 3DMM contained only a coarse texture without makeup. The specular shading was affected by the limited light source, and thus, the global illumination could not be recovered. Furthermore, the specular albedo was a coarse texture; therefore, the final specular reconstruction is not detailed. Although it was not possible to obtain a reconstruction of the makeup using only the 3D face reconstruction method, these

Figure 7: *Final outputs of our framework. (a) Input makeup portraits, and (b) fully reconstructed renderings and corresponding textures. The columns from (c) to (f) are similar to those of Fig. 1.*

coarse facial materials were helpful for subsequent steps. We used the original texture and completed it to obtain the refined facial materials as well as makeup. The UV completion results are presented in (c). The final textures were refined using the outputs of the previous steps, as illustrated in (d). For better visualization, we show the blended makeup, which was calculated by $(1-\mathbf{A})\odot\mathbf{D}_m^m$. It can be observed that the makeup and bare skin were disentangled, the diffuse shading was adjusted, and the specular reconstruction was more detailed. We combined the final textures to achieve effects such as rendering a bare skin face and a makeup face while the lighting remained invariant. These results demonstrate that our method is robust to different makeup, expressions, poses, and illumination.

The final outputs. The final outputs are depicted in Fig. 7. We rendered the disentangled final UV textures separately for improved visualization: (c) bare skin, (d) bare skin with makeup, (e) bare skin with diffuse shading, and (f) bare skin with the full illumination model. Furthermore, we rendered the makeup faces in (b) using these textures which could be considered as a 3D face reconstruction of makeup portraits. It can be observed that the layers of bare skin, makeup, diffuse, and specular are disentangled. For example, a comparison of (c) and (d) indicates that makeup was added while no illumination was involved, and the makeup around the eyebrows, eyes, and lips was disentangled. The presence of makeup on the cheeks can also be clearly observed in the third identity from the top.

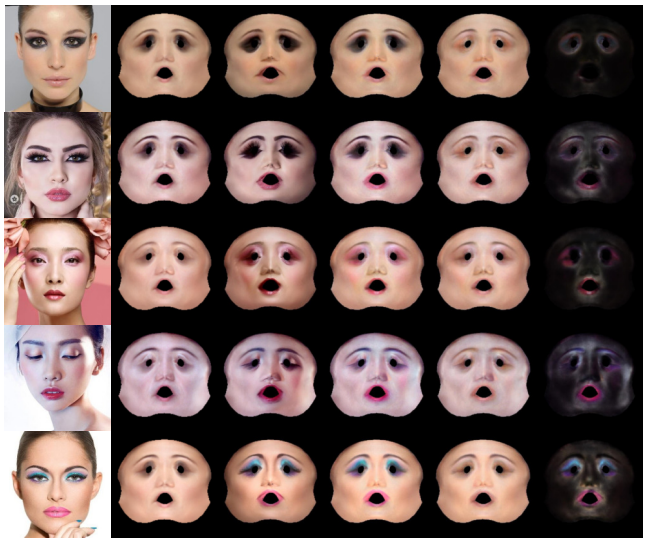

Figure 8: *Texture outputs relating to makeup. Left to right: (a) input makeup portraits, (b) coarse diffuse albedo, (c) completed UV textures, (d) refined diffuse albedo, (e) bare skin, and (f) makeup.*

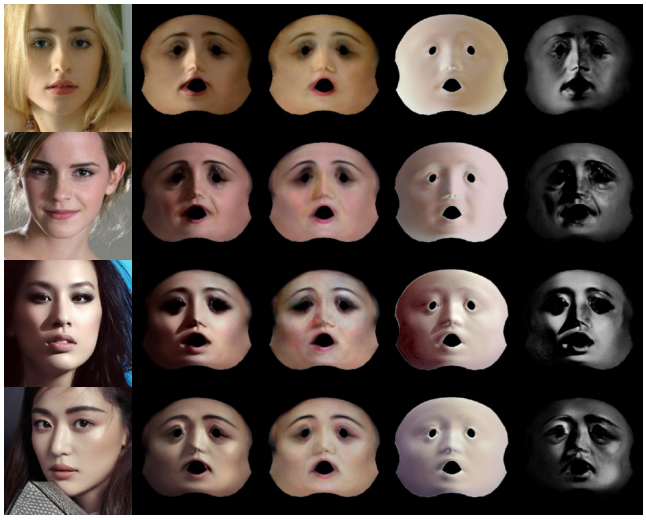

Figure 9: *Outputs of decomposed refined facial materials in complex illumination conditions. Left to right: (a) input makeup portraits, (b) completed UV textures, (c) diffuse albedo, (d) diffuse shading, and (e) specular reconstruction.*

Makeup in diffuse albedo. The textures relating to the makeup are depicted in Fig. 8 to demonstrate the effect of the makeup changes on the diffuse albedo textures. This process represents the main

Table 1: *Quantitative ablation study for makeup face reconstruction results.*

Complete data	RMSE	SSIM	LPIPS
FRN (w/o specular)	0.056	0.953	0.091
FRN (w/ specular)	0.053 ↓	0.956 ↑	0.088 ↓
Rendered result	0.060	0.968 ⇑	0.062 ⇓
Test data	**RMSE**	**SSIM**	**LPIPS**
FRN (w/o specular)	0.058	0.948	0.102
FRN (w/ specular)	0.055 ↓	0.951 ↑	0.099 ↓
Rendered result	0.063	0.964 ⇑	0.066 ⇓

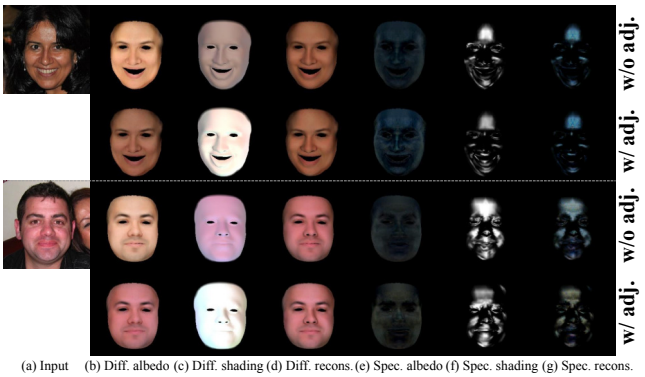

Figure 10: *Qualitative ablation study for skin tone adjustment. For each identity, the upper and lower rows display the results before and after the skin tone adjustment, respectively. Left to right: (a) input face images, (b) diffuse albedo, (c) diffuse shading, (d) diffuse reconstruction, (e) specular albedo, (f) specular shading, and (g) specular reconstruction.*

concept under consideration for extracting the makeup. The reconstructed coarse diffuse albedo could not preserve the makeup effectively; although the reconstruction results exhibited some black makeup around the eyes, it was difficult to preserve the makeup beyond the scope of the 3DMM texture space (see (b)). Thus, we used the original texture directly. The completed UV textures, which contained makeup and involved illumination, are depicted in (c). Subsequently, the illumination was removed. A comparison of (d), (e), and (f) demonstrates the validity of our makeup extraction network. Moreover, the final row presents an example of a makeup portrait with occlusion, for which our method was still effective in extracting the makeup. As we only used the sampled skin region while excluding the others, the missing regions were filled to become complete.

Decomposition of illumination. We evaluate the results of the refined facial materials to demonstrate the capabilities of our illumination-aware makeup extraction. It can be observed from the outputs in Fig. 9 that portraits containing uneven lighting and shadows (particularly around the bridge of the nose) were captured by diffuse shading, whereas the highlights were reflected in the specular reconstruction. Thus, the diffuse albedo textures became clean and flat, and our makeup extraction was more precise after removing the illumination.

The entire process of our method is executed in the UV space. UV-represented makeup offers numerous advantages. First, 3D makeup avatar creation becomes accessible when the same UV coordinates are used. Furthermore, such makeup can be extended to scanned 3D faces, in which case makeup without illumination will be helpful. Second, the makeup can be further divided into several parts using a corresponding face segmentation mask, which will enable specification of which makeup region is to be used. Third, the textures can be directly edited and incorporated into a traditional rendering pipeline. Finally, the disentangled makeup maps can be collected to form a makeup dataset, which will be useful for 3D makeup face reconstruction or makeup recommendation. We explore several applications in Sec. 7.

6. Ablation Studies

We conducted ablation studies on the 3D face reconstruction step. We performed a comparative experiment with and without skin tone adjustment and assessed the influence on the components of 3DMM. We selected two images from FFHQ [KLA19] with strong illumination. As illustrated in Fig. 10, the skin tone adjustment had only a slight effect, or none at all, on the specular. Although the diffuse reconstruction did not change, the balance of the diffuse albedo and diffuse shading was significantly adjusted. Limited by the 3DMM texture, the diffuse albedos of the two faces before adjusting the skin tones had almost similar colors, resulting in an incorrect estimation of the diffuse shading. This would be misleading for the subsequent refinement and makeup extraction step. To mitigate this error, we adjusted the skin tone so that the average color was the same as that of the original image.

Tab. 1 displays the results of the quantitative evaluation of the 3D makeup face reconstruction. We trained two face reconstruction networks with different illumination models: one used SH lighting without specular estimation, and the other was the full model. The rendered results using the bare skin, makeup, diffuse shading, and specular reconstruction textures are also listed for reference. We used the complete makeup, which was not used to train the 3D face reconstruction network for the general evaluation. We used a gray color to mark the rendered results, because the generated maps for rendering were trained on the complete dataset. Furthermore, the test makeup dataset, which was not used to train the makeup extraction network, was separated to evaluate the rendered results. The network containing the specular illumination model improved the accuracy of the reconstruction, thereby demonstrating the effectiveness of our model. Our final rendered makeup face exhibited improvement in terms of the perceptual similarity. Thus, the reconstructed results were more compatible with the visual evaluation, as we believe that makeup significantly influences human perception.

7. Applications

We explore several makeup-related applications using the results of our method and compare them with state-of-the-art methods.

Table 2: *Taxonomy of state-of-the-art makeup transfer methods. "Misalignment": faces with different poses. "Shade": controlling the shade. "Control": selection of the face region to be transferred. "Edit": editing within arbitrary areas. "Occlusion": robust to occlusion. "Illumination": controlled illumination during transfer.*

Method	Misalignment	Shade	Control	Edit	Occlusion	Illumination
BeautyGAN [LQD*18]						
LADN [GWC*19]		✓				
PSGAN [JLG*20]	✓	✓	✓			
SCGAN [DHC*21]	✓	✓	✓			
CPM [NTH21]	✓	✓	✓	✓		
SOGAN [LDP*21]	✓	✓	✓	✓	✓	
EleGANt [YHXG22]	✓	✓	✓	✓	✓	
Ours	✓	✓	✓	✓	✓	✓

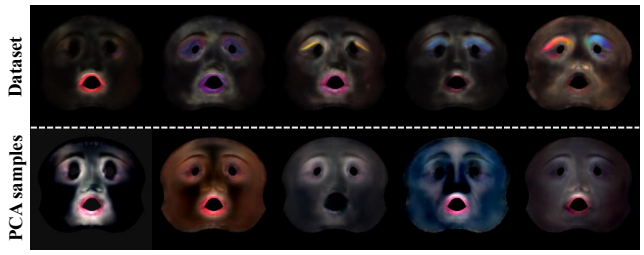

Figure 11: *Makeup dataset and PCA-based makeup model. We used the dataset to create a PCA-based makeup model. The top row presents the extracted makeup textures from image input, and the bottom row shows randomly sampled makeup textures along the principal components from the makeup model.*

7.1. 3D Face Reconstruction of Makeup

First, we demonstrate how the extracted makeup dataset can be used to enhance the 3D face reconstruction of makeup. The same process as the diffuse albedo model construction of FLAME [LBB*17] was followed, and the collected makeup textures $(1-\mathbf{A})\odot\mathbf{D}_m^m$ were used to construct a PCA-based statistical model of makeup. The randomly sampled textures from the makeup model are depicted in Fig. 11. The makeup model is an extension of the diffuse albedo model, and the new model $\mathbf{D}_c^{'}$ can be formulated as:

$$\mathbf{D}_c^{'} = \mathbf{D}_c + \mathbf{D}_c^m, \tag{12}$$

where $\mathbf{D}_c^m$ is the makeup model. We used an optimization-based manner to reconstruct the 3D face from the makeup portraits because no large-scale makeup dataset is available to train a neural network. A comparison of the 3D makeup face reconstruction using different albedo models $\mathbf{D}_c$ and $\mathbf{D}_c^{'}$ is presented in Fig. 12. $\mathbf{D}_c^{'}$ could recover the makeup, and improve the accuracy of the entire reconstruction, especially for lipsticks and eye shadows. The shape of the lips and eyes were also matched more effectively by extending the ability of the diffuse albedo.

We believe that this makeup database can be explored further; for example, by using advanced image generation techniques such as StyleGAN/StyleGAN2 [KLA19, KLA*20] or diffusion model [HJA20]. The accuracy of the reconstruction results also requires further quantitative evaluation, and we consider these as future research topics.

7.2. Illumination-Aware Makeup Transfer

We used the extracted makeup for makeup transfer by employing the same method as that for the makeup extraction network. Equation (10) was followed to blend a new makeup face, which was subsequently projected onto the original image. The functionalities of our method and previous methods are summarized in Tab. 2. Similar to the approach of CPM [NTH21], the makeup textures are UV representations to solve the misalignment of faces. The face region that is used for transfer can be specified, and the makeup is editable. In addition to these advantages, our approach extracts makeup and uses a completed UV representation, which can handle occlusion and illumination.

Our makeup transfer results are depicted in Fig. 13. The reference images contained various complex factors, including occlusion, face misalignment, and lighting conditions. We extracted the makeup from the reference image. Subsequently, we transferred the makeup to the source image while maintaining the original illumination of the source image.

A qualitative comparison with state-of-art makeup transfer methods is depicted in Fig. 14, in which the source and reference images have different illumination. As existing makeup transfer methods do not consider illumination, they transfer not only the makeup, but also the effect of the illumination. The first two rows present the exchanged makeup results of two identities, one was evenly illuminated, whereas the other had shadows on the cheeks. The existing methods could not retain the illumination from the source image, which resulted in a mismatch between the face and surrounding environment. Note that our method preserved the illumination and shade of the original images in the cheeks. The final row shows an example in which the source and reference images have contrasting illumination. Our method enabled a natural makeup transfer, whereas the other methods were affected by the illumination, resulting in undesirable results. Note that the pioneering method known as BeautyGAN [LQD*18] (a) was relatively stable. However, as it is not suitable for transferring eye shadow, the makeup in the source image was not cleaned up.

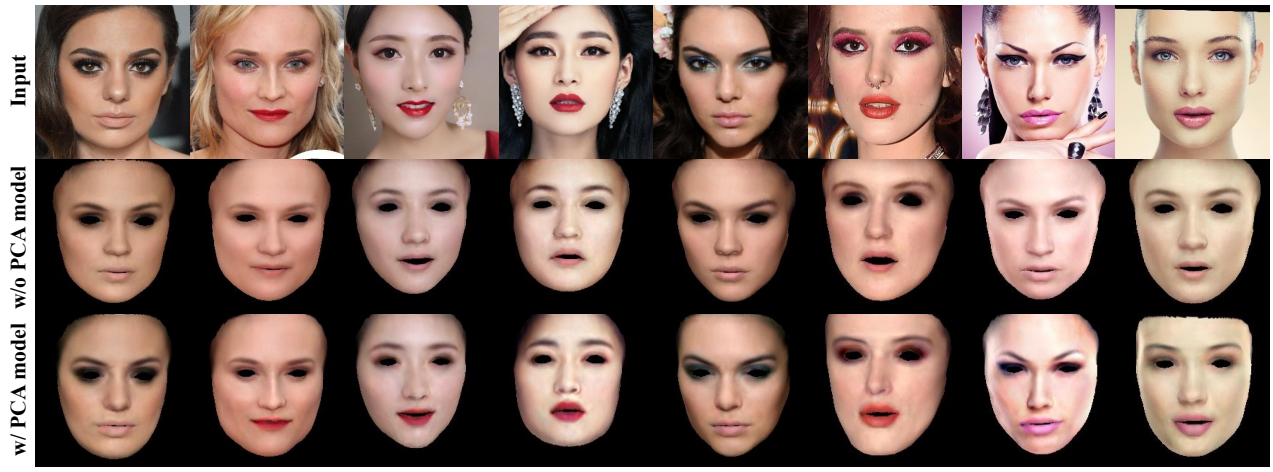

Figure 12: *3D face reconstruction results using 3DMM. The PCA-based makeup model significantly improves the fidelity of synthesized makeup.*

Figure 13: *Our makeup transfer can handle occlusion, illumination, and face misalignment.*

7.3. Illumination-Aware Makeup Interpolation and Removal

As illustrated in Fig. 15, compared to existing makeup transfer methods, our method could achieve makeup interpolation and removal without a reference image. Moreover, our makeup interpolation maintained constant illumination conditions and achieved natural makeup interpolation. The first two rows present the interpolation and removal results of the face images. The specular and diffuse shading were not changed while the makeup was adjusted from heavy to light. The third row shows how the alpha matte changed in the makeup interpolation process. The alpha matte $\mathbf{A}$ is adjusted to $\mathbf{A}_\sigma$ as follows:

$$\mathbf{A}_\sigma(p) = clamp(\mathbf{A}(p) + \sigma, 0, 1), \tag{13}$$

where $\mathbf{A}_\sigma(p)$ and $\mathbf{A}(p)$ are values of $\mathbf{A}_\sigma$ and $\mathbf{A}$ at pixel p, respectively, and $\sigma \in [0,1]$. $\mathbf{A}_\sigma$ was eventually clipped to between 0 and 1. It can be observed that for the change in $\mathbf{A}_\sigma$, the makeup was mainly lipstick and eye shadow in this sample. Thus, the original $\mathbf{A}$ was the black color around the lips and eyes, which means that this part of the skin color was not used, while the makeup is mainly applied. With the increase in
sigma, $\mathbf{A}_\sigma$ became whiter; thus the use of makeup was reduced and makeup interpolation was achieved. As the illumination was disentangled from the makeup, it was possible to relight the face while adjusting the makeup. The results are presented in the final row.

In addition to the above applications, the extracted makeup can be used for makeup recommendations, please refer to [SRH*11, AJWF17]. The makeup textures can also facilitate the subsequent processing of traditional graphics pipelines such as physically-based makeup rendering.

8. Limitations and Conclusions

Limitations. Although we integrated a color adjustment to alleviate the inherent skin color bias, the 3DMM skin colors still exhibited a problem. Namely, because the skin colors of the FLAME model [LBB*17] were obtained by unwrapping face images of the FFHQ dataset [KLA19], they contained baked-in lighting effects. Therefore, our coarse albedos that were obtained using the FLAME model contained shading effects, which further caused errors in our refinement step and makeup extraction. Whereas our method yielded satisfactory results in most cases, we would like to explore a better albedo model.

Our network erroneously extracts a makeup-like material from a makeup-less face because our network is currently not trained with paired data of makeup-less inputs and makeup-less outputs.

Although our refinement step greatly improves the diffuse albe-

Figure 14: *Qualitative comparison of makeup transfer between two faces. Left to right: source images providing the identity and illumination, the reference images providing the makeup, (a) BeautyGAN [LQD*18], (b) PSGAN [JLG*20], (c) CPM [NTH21], (d) SCGAN [DHC*21], (e) EleGANt [YHXG22], and (f) our results, which retain the lighting effects.*

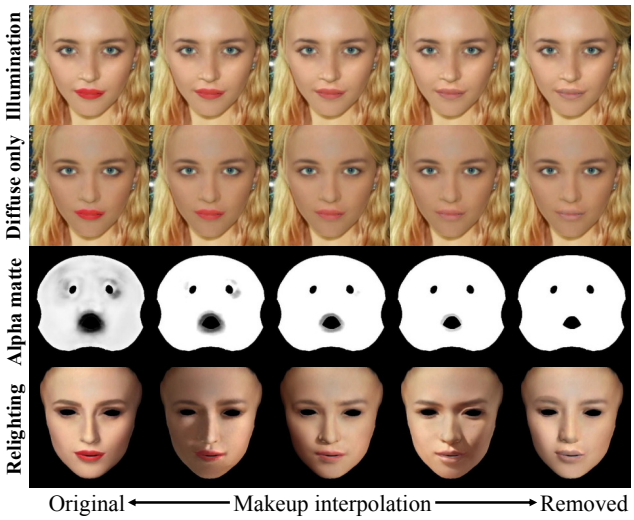

Figure 15: *Results of illumination-aware makeup interpolation and removal. Left to right: the original makeup was interpolated and finally removed. The first row presents the interpolation results with constant illumination. The second row displays the results using only diffuse shading. The third row shows the alpha matte* **A** *for controlling the balance of the bare skin and makeup. The final row presents the rendered faces with relighting.*

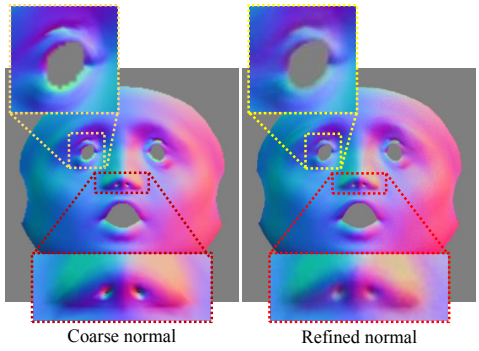

Figure 16: *Comparison of coarse and refined normal.*

dos (see Fig. 8(b)(d)), the difference in the face geometry is subtle before and after refinement. Fig. 16 shows that the normals around the eyes and mouth were smoothed. We would like to improve the face geometry as well.

At present, we extract makeup from diffuse albedos, but in the real world, makeup contains specular albedos. We would like to account for makeup BRDFs for a more physically-plausible makeup transfer. Quantitative evaluation is difficult in makeup-related research because no public ground-truth dataset of accurately aligned face images before and after makeups is available. Therefore, it is also essential to establish a quantitative evaluation criterion.

Conclusions. We have presented the first method for extracting makeup for 3D face models from a single makeup portrait, which consists of the following three steps; 1) the extraction of coarse facial materials such as geometry and diffuse/specular albedos via extended regression-based inverse rendering using 3DMM [LBB*17], 2) a newly designed optimization-based refinement of the coarse materials, and 3) a novel network that is designed for extracting makeup. Thanks to the disentangled outputs, we can achieve novel applications such as illumination-aware (i.e., relightable) makeup transfer, interpolation, and removal. The resultant makeup is well aligned in the UV space, from which we built a large-scale makeup texture dataset and a PCA-based makeup model. In future work,

we would like to overcome the current limitations and explore better statistical models for facial makeup.

Acknowledgements

We thank Prof. Yuki Endo in University of Tsukuba for the suggestions and comments on our research. We also thank the reviewers for their constructive feedback and suggestions, which helped us improve our paper.

References

[AJWF17] ALASHKAR T., JIANG S., WANG S., FU Y.: Examples-rules guided deep neural network for makeup recommendation. In *AAAI 2017* (2017), pp. 941–947. 3, 12

[BJ03] BASRI R., JACOBS D. W.: Lambertian reflectance and linear subspaces. *Transactions on Pattern Analysis and Machine Intelligence 25*, 2 (2003), 218–233. 5, 6

[BLC*22] BAO L., LIN X., CHEN Y., ZHANG H., WANG S., ZHE X., KANG D., HUANG H., JIANG X., WANG J., YU D., ZHANG Z.: High-fidelity 3D digital human head creation from RGB-D selfies. *Transactions on Graphics 41*, 1 (2022), 3:1–3:21. 3

[Bli77] BLINN J. F.: Models of light reflection for computer synthesized pictures. In *Proc. of SIGGRAPH 1977* (1977), pp. 192–198. 5

[BRP*18] BOOTH J., ROUSSOS A., PONNIAH A., DUNAWAY D. J., ZAFEIRIOU S.: Large scale 3D morphable models. *International Journal of Computer Vision 126*, 2-4 (2018), 233–254. 3

[BT17] BULAT A., TZIMIROPOULOS G.: How far are we from solving the 2D & 3D face alignment problem? (and a dataset of 230,000 3D facial landmarks). In *ICCV 2017* (2017), pp. 1021–1030. 7

[BV99] BLANZ V., VETTER T.: A morphable model for the synthesis of 3D faces. In *Proc. of SIGGRAPH 1999* (1999), Waggenspack W. N., (Ed.), pp. 187–194. 3

[BV03] BLANZ V., VETTER T.: Face recognition based on fitting a 3D morphable model. *Transactions on Pattern Analysis and Machine Intelligence 25*, 9 (2003), 1063–1074. 3

[CHS22] CHEN J., HAN H., SHAN S.: Towards high-fidelity face self-occlusion recovery via multi-view residual-based GAN inversion. In *AAAI 2022* (2022), pp. 294–302. 3, 5

[DBA*21] DIB A., BHARAJ G., AHN J., THÉBAULT C., GOSSELIN P. H., ROMEO M., CHEVALLIER L.: Practical face reconstruction via differentiable ray tracing. *Computer Graphics Forum 40*, 2 (2021), 153–164. 3, 4

[DBB22] DANECEK R., BLACK M. J., BOLKART T.: EMOCA: Emotion driven monocular face capture and animation. In *CVPR 2022* (2022), pp. 20311–20322. 3

[DCX*18] DENG J., CHENG S., XUE N., ZHOU Y., ZAFEIRIOU S.: UV-GAN: adversarial facial UV map completion for pose-invariant face recognition. In *CVPR 2018* (2018), pp. 7093–7102. 3

[DHC*21] DENG H., HAN C., CAI H., HAN G., HE S.: Spatially-invariant style-codes controlled makeup transfer. In *CVPR 2021* (2021), pp. 6549–6557. 3, 11, 13

[DYX*19] DENG Y., YANG J., XU S., CHEN D., JIA Y., TONG X.: Accurate 3D face reconstruction with weakly-supervised learning: From single image to image set. In *CVPR 2019 Workshops* (2019), pp. 285–295. 3, 4, 5, 7

[EST*20] EGGER B., SMITH W. A. P., TEWARI A., WUHRER S., ZOLLHÖFER M., BEELER T., BERNARD F., BOLKART T., KORTYLEWSKI A., ROMDHANI S., THEOBALT C., BLANZ V., VETTER T.: 3D morphable face models - past, present, and future. *Transactions on Graphics 39*, 5 (2020), 157:1–157:38. 2, 3

[FFBB21] FENG Y., FENG H., BLACK M. J., BOLKART T.: Learning an animatable detailed 3D face model from in-the-wild images. *Transactions on Graphics 40*, 4 (2021), 88:1–88:13. 3

[GCM*18] GENOVA K., COLE F., MASCHINOT A., SARNA A., VLASIC D., FREEMAN W. T.: Unsupervised training for 3D morphable model regression. In *CVPR 2018* (2018), pp. 8377–8386. 3

[GDZ21] GECER B., DENG J., ZAFEIRIOU S.: OSTeC: One-shot texture completion. In *CVPR 2021* (2021), pp. 7628–7638. 3

[GEB16] GATYS L. A., ECKER A. S., BETHGE M.: Image style transfer using convolutional neural networks. In *CVPR 2016* (2016), pp. 2414–2423. 6

[GMB*18] GERIG T., MOREL-FORSTER A., BLUMER C., EGGER B., LÜTHI M., SCHÖNBORN S., VETTER T.: Morphable face models - an open framework. In *Proceedings of International Conference on Automatic Face & Gesture Recognition* (2018), pp. 75–82. 3, 6

[GPKZ19] GECER B., PLOUMPIS S., KOTSIA I., ZAFEIRIOU S.: GANFIT: generative adversarial network fitting for high fidelity 3D face reconstruction. In *CVPR 2019* (2019), pp. 1155–1164. 3

[GWC*19] GU Q., WANG G., CHIU M. T., TAI Y., TANG C.: LADN: local adversarial disentangling network for facial makeup and de-makeup. In *ICCV 2019* (2019), pp. 10480–10489. 3, 7, 11

[GZC*19] GUO Y., ZHANG J., CAI J., JIANG B., ZHENG J.: CNN-based real-time dense face reconstruction with inverse-rendered photo-realistic face images. *Transactions on Pattern Analysis and Machine Intelligence 41*, 6 (2019), 1294–1307. 3

[HJA20] HO J., JAIN A., ABBEEL P.: Denoising diffusion probabilistic models. In *NeurIPS 2020* (2020), vol. 33, pp. 6840–6851. 11

[HLHC13] HUANG C.-G., LIN W.-C., HUANG T.-S., CHUANG J.-H.: Physically-based cosmetic rendering. In *Proceedings of the ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games* (2013), p. 190. 2

[IZZE17] ISOLA P., ZHU J., ZHOU T., EFROS A. A.: Image-to-image translation with conditional adversarial networks. In *CVPR 2017* (2017), pp. 5967–5976. 7

[JAF16] JOHNSON J., ALAHI A., FEI-FEI L.: Perceptual losses for real-time style transfer and super-resolution. In *ECCV 2016* (2016), vol. 9906, pp. 694–711. 6

[JLG*20] JIANG W., LIU S., GAO C., CAO J., HE R., FENG J., YAN S.: PSGAN: pose and expression robust spatial-aware GAN for customizable makeup transfer. In *CVPR 2020* (2020), pp. 5193–5201. 3, 11, 13

[JYG*22] JI C., YU T., GUO K., LIU J., LIU Y.: Geometry-aware single-image full-body human relighting. In *ECCV 2022* (2022). 3

[KALL18] KARRAS T., AILA T., LAINE S., LEHTINEN J.: Progressive growing of GANs for improved quality, stability, and variation. In *Proceedings of International Conference on Learning Representations* (2018). 7

[KE18] KANAMORI Y., ENDO Y.: Relighting humans: occlusion-aware inverse rendering for full-body human images. *Transactions on Graphics 37*, 6 (2018), 270. 3

[KGPB20] KIPS R., GORI P., PERROT M., BLOCH I.: CA-GAN: weakly supervised color aware GAN for controllable makeup transfer. In *ECCV 2020* (2020), vol. 12537, pp. 280–296. 3

[KJB*22] KIPS R., JIANG R., BA S. O., DUKE B., PERROT M., GORI P., BLOCH I.: Real-time virtual-try-on from a single example image through deep inverse graphics and learned differentiable renderers. *Computer Graphics Forum 41*, 2 (2022), 29–40. 3

[KLA19] KARRAS T., LAINE S., AILA T.: A style-based generator architecture for generative adversarial networks. In *CVPR 2019* (2019), pp. 4401–4410. 7, 10, 11, 12

[KLA*20] KARRAS T., LAINE S., AITTALA M., HELLSTEN J., LEHTINEN J., AILA T.: Analyzing and improving the image quality of StyleGAN. In *CVPR 2020* (2020), pp. 8107–8116. 11

[KYT21] KIM J., YANG J., TONG X.: Learning high-fidelity face texture completion without complete face texture. In *CVPR 2021* (2021), pp. 13970–13979. 3, 4, 5

[LBB*17] LI T., BOLKART T., BLACK M. J., LI H., ROMERO J.: Learning a model of facial shape and expression from 4D scans. *Transactions on Graphics 36*, 6 (2017), 194:1–194:17. 2, 3, 4, 11, 12, 13

[LDP*21] LYU Y., DONG J., PENG B., WANG W., TAN T.: SOGAN: 3D-aware shadow and occlusion robust GAN for makeup transfer. In *Proceedings of International Conference on Multimedia* (2021), pp. 3601–3609. 3, 11

[LHK*20] LAINE S., HELLSTEN J., KARRAS T., SEOL Y., LEHTINEN J., AILA T.: Modular primitives for high-performance differentiable rendering. *Transactions on Graphics 39*, 6 (2020), 194:1–194:14. 7

[LL20] LEE G., LEE S.: Uncertainty-aware mesh decoder for high fidelity 3D face reconstruction. In *CVPR 2020* (2020), pp. 6099–6108. 3

[LMG*20] LATTAS A., MOSCHOGLOU S., GECER B., PLOUMPIS S., TRIANTAFYLLOU V., GHOSH A., ZAFEIRIOU S.: AvatarMe: Realistically renderable 3D facial reconstruction "in-the-wild". In *CVPR 2020* (2020), pp. 757–766. 3

[LQD*18] LI T., QIAN R., DONG C., LIU S., YAN Q., ZHU W., LIN L.: BeautyGAN: Instance-level facial makeup transfer with deep generative adversarial network. In *Proceedings of International Conference on Multimedia* (2018), pp. 645–653. 2, 3, 7, 11, 13

[LSY*21] LAGUNAS M., SUN X., YANG J., VILLEGAS R., ZHANG J., SHU Z., MASIÁ B., GUTIERREZ D.: Single-image full-body human relighting. In *Proceedings of Eurographics Symposium on Rendering* (2021), pp. 167–177. 3

[LZL15] LI C., ZHOU K., LIN S.: Simulating makeup through physics-based manipulation of intrinsic image layers. In *CVPR 2015* (2015), pp. 4621–4629. 3

[MZWX21] MA X., ZHANG F., WEI H., XU L.: Deep learning method for makeup style transfer: A survey. *Cognitive Robotics 1* (2021), 182–187. 2

[NTH21] NGUYEN T., TRAN A. T., HOAI M.: Lipstick ain't enough: Beyond color matching for in-the-wild makeup transfer. In *CVPR 2021* (2021), pp. 13305–13314. 3, 5, 11, 13

[POL*21] PANDEY R., ORTS-ESCOLANO S., LEGENDRE C., HÄNE C., BOUAZIZ S., RHEMANN C., DEBEVEC P. E., FANELLO S. R.: Total relighting: learning to relight portraits for background replacement. *Transactions on Graphics 40*, 4 (2021), 43:1–43:21. 3

[RDS*15] RUSSAKOVSKY O., DENG J., SU H., KRAUSE J., SATHEESH S., MA S., HUANG Z., KARPATHY A., KHOSLA A., BERNSTEIN M. S., BERG A. C., FEI-FEI L.: ImageNet large scale visual recognition challenge. *International Journal of Computer Vision 115*, 3 (2015), 211–252. 7

[RH01] RAMAMOORTHI R., HANRAHAN P.: An efficient representation for irradiance environment maps. In *Proc. of SIGGRAPH 2001* (2001), pp. 497–500. 3

[SBFB19] SANYAL S., BOLKART T., FENG H., BLACK M. J.: Learning to regress 3D face shape and expression from an image without 3D supervision. In *CVPR 2019* (2019), pp. 7763–7772. 3

[SBT*19] SUN T., BARRON J. T., TSAI Y.-T., XU Z., YU X., FYFFE G., RHEMANN C., BUSCH J., DEBEVEC P., RAMAMOORTHI R.: Single image portrait relighting. *Transactions on Graphics 38*, 4 (2019). 3

[SKCJ18] SENGUPTA S., KANAZAWA A., CASTILLO C. D., JACOBS D. W.: SfSNet: Learning shape, reflectance and illuminance of faces 'in the wild'. In *CVPR 2018* (2018), pp. 6296–6305. 2, 3

[SL09] STYLIANOU G., LANITIS A.: Image based 3D face reconstruction: a survey. *International Journal of Image and Graphics 9*, 2 (2009), 217–250. 3

[SRH*11] SCHERBAUM K., RITSCHEL T., HULLIN M., THORMÄHLEN T., BLANZ V., SEIDEL H.-P.: Computer-suggested facial makeup. *Computer Graphics Forum 30*, 2 (2011). 2, 3, 12

[SSD*20] SMITH W. A. P., SECK A., DEE H., TIDDEMAN B., TENENBAUM J. B., EGGER B.: A morphable face albedo model. In *CVPR 2020* (2020), pp. 5010–5019. 4

[SSL*20] SHANG J., SHEN T., LI S., ZHOU L., ZHEN M., FANG T., QUAN L.: Self-supervised monocular 3D face reconstruction by occlusion-aware multi-view geometry consistency. In *ECCV 2020* (2020), Vedaldi A., Bischof H., Brox T., Frahm J., (Eds.), vol. 12360, pp. 53–70. 3

[SWH*17] SAITO S., WEI L., HU L., NAGANO K., LI H.: Photorealistic facial texture inference using deep neural networks. In *CVPR 2017* (2017), pp. 2326–2335. 3

[TFM*22] TAN F., FANELLO S., MEKA A., ORTS-ESCOLANO S., TANG D., PANDEY R., TAYLOR J., TAN P., ZHANG Y.: VoLux-GAN: A generative model for 3D face synthesis with HDRI relighting. In *ACM SIGGRAPH 2022 Conference Proceedings* (2022), pp. 58:1–58:9. 3

[THMM17] TRAN A. T., HASSNER T., MASI I., MEDIONI G. G.: Regressing robust and discriminative 3D morphable models with a very deep neural network. In *CVPR 2017* (2017), pp. 1493–1502. 3

[TKE21] TAJIMA D., KANAMORI Y., ENDO Y.: Relighting humans in the wild: Monocular full-body human relighting with domain adaptation. *Computer Graphics Forum 40*, 7 (2021), 205–216. 3

[TZG*18] TEWARI A., ZOLLHÖFER M., GARRIDO P., BERNARD F., KIM H., PÉREZ P., THEOBALT C.: Self-supervised multi-level face model learning for monocular reconstruction at over 250 Hz. In *CVPR 2018* (2018), pp. 2549–2559. 3

[TZK*17] TEWARI A., ZOLLHÖFER M., KIM H., GARRIDO P., BERNARD F., PÉREZ P., THEOBALT C.: MoFA: Model-based deep convolutional face autoencoder for unsupervised monocular reconstruction. In *CVPR 2017* (2017), pp. 3735–3744. 3

[WWR22] WIMBAUER F., WU S., RUPPRECHT C.: De-rendering 3D objects in the wild. In *CVPR 2022* (2022), pp. 18490–18499. 3

[WYL*20] WANG Z., YU X., LU M., WANG Q., QIAN C., XU F.: Single image portrait relighting via explicit multiple reflectance channel modeling. *Transactions on Graphics 39*, 6 (2020), 220:1–220:13. 3

[XZY*22] XIA W., ZHANG Y., YANG Y., XUE J., ZHOU B., YANG M.: GAN inversion: A survey. *Transactions on Pattern Analysis and Machine Intelligence* (2022), 1–17. 3

[YHXG22] YANG C., HE W., XU Y., GAO Y.: EleGANt: Exquisite and locally editable GAN for makeup transfer. In *ECCV 2022* (2022). 3, 11, 13

[YNK*22] YEH Y., NAGANO K., KHAMIS S., KAUTZ J., LIU M., WANG T.: Learning to relight portrait images via a virtual light stage and synthetic-to-real adaptation. *Transactions on Graphics* (2022). 3

[YSN*18] YAMAGUCHI S., SAITO S., NAGANO K., ZHAO Y., CHEN W., OLSZEWSKI K., MORISHIMA S., LI H.: High-fidelity facial reflectance and geometry inference from an unconstrained image. *Transactions on Graphics 37*, 4 (2018), 162. 3

[YT22] YANG X., TAKETOMI T.: BareSkinNet: De-makeup and De-lighting via 3D Face Reconstruction. *Computer Graphics Forum* (2022). 3

[YWP*18] YU C., WANG J., PENG C., GAO C., YU G., SANG N.: Bisenet: Bilateral segmentation network for real-time semantic segmentation. In *ECCV 2018* (2018), vol. 11217, pp. 334–349. 7

[ZBT22] ZIELONKA W., BOLKART T., THIES J.: Towards metrical reconstruction of human faces. In *ECCV 2022* (2022), pp. 250–269. 3

[ZPIE17] ZHU J., PARK T., ISOLA P., EFROS A. A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In *ICCV 2017* (2017), pp. 2242–2251. 3