Makeup Prior Models for 3D Facial Makeup Estimation and Applications

Supplementary Material

In this supplemental material, we first briefly review the two methods proposed by Yang *et al*. [84] in Sec. A.1; the texture decomposition-based 3D facial makeup extraction and the optimization-based residual makeup model fitting, which we named "Yang-Ext" and "Yang-Res", respectively. Then, we compare the Yang-Res and our PCA model in Sec. A.2, which aims for a better understanding of the differences.

For the supplemental experiments, we first provide additional details on the ablation studies in Sec. B, as promised in the main text. This includes two comparisons; one is an experiment on the refinement process using inferred coefficients for initialization in Sec. B.1, and the other is the details of quantitative experiments in Sec. B.2. Subsequently, we present more results related to makeup extraction, 3D face reconstruction, makeup interpolation, and makeup transfer in Sec. C.

A. Previous Makeup Estimation Methods

A.1. Yang-Ext and Yang-Res

Yang *et al*. [84] tackled the problem of 3D-aware makeup extraction for the first time, in which they proposed two variants, i.e., what we call "Yang-Ext" and "Yang-Res". We first explain the texture decomposition-based method, "Yang-Ext", which consists of the following four coarse-to-fine steps:

Step 1 servers as coarse facial texture extraction via regression-based inverse rendering using a 3D morphable model (3DMM). The facial textures are represented as UV maps for the 3DMM. This step extends the method by Deng *et al*. [17] to estimate not only face shape, albedo, and diffuse shading but also specular shading. The rendering method for this 3DMM reconstruction includes both the Lambertian and Blinn-Phong reflection models. The pre-trained 3D face reconstruction network E_{FLAME} in our method leverages the same network as theirs.

Step 2 employs inpainting in UV maps. Their method tries to extract detailed information by fitting the 3DMM to facial images, but the images might be partially hidden due to self-occlusion. It thus utilize a face UV inpainting technique [40].

Step 3 refines the inpainted coarse textures via optimization to obtain a high-resolution albedo map that contains bare skin and makeup without the effect of illumination.

Step 4 separates the high-resolution albedo map into bare skin, makeup, and alpha matte based on makeup transfer. The combination of these decomposed textures follows Eq. (2) and Eq. (3)

As an application, they perform principal component analysis on the extracted makeup textures to create a three-channel morphable makeup model follows Eq. (4). Therefore, they can implement an optimization-based fitting of the makeup model to makeup images, which we named "Yang-Res".

The entire method of Yang-ext is a long pipeline, making the algorithm unstable and time-consuming. The refinement of albedo in Step 3 is performed using the coarse albedo obtained in Step 1, which does not contain makeup, affecting subsequent accuracy.

A.2. Differences between Yang-Res and Our PCA

Briefly, Yang-Res has the following key limitations:

1. **Undecoupled approach**: It models makeup as a residual via subtraction (see Sec. 7.1 in their paper [84]). The residual closely ties the makeup to its corresponding bare skin. This approach is neither easily generalizable nor versatile for transferring to different faces, as it may carry excessive residuals that do not match the new face's features, as shown in Fig. 7.
2. **Analysis-by-synthesis optimization**: The optimization incurs significant computational costs due to several hundreds of iterations, as opposed to the single forward pass of our regression network.
3. **Lacks makeup-specific regularization**: As shown in Fig. 6, without the makeup-specific regularization to constrain PCA parameters, makeup becomes unconstrained and thus overfits the input images.

Our PCA addresses these issues through the following improvements:

1. **Alpha blending model**: We adopt the same makeup model as Yang-Ext, *i.e*., the alpha blending model, which reduces the risk of overfitting by independently estimating makeup bases and opacity, thereby decreasing reliance on specific features. Our $\mathbf{M}_b$ and $\mathbf{M}_a$ correspond to Yang-Ext's $\mathbf{D}_m^m$ and $(1 - \mathbf{A})$, respectively.
2. **Regression network**: We trained a regression network to accelerate the inference.
3. **Makeup consistency module**: We designed a novel architecture to regularize makeup PCA parameters. This module transfers makeup across different identities, expressions, and poses while maintaining cycle consistency. This enhances the disentanglement of makeup and

bare skin, enabling us to handle various scenarios, such as occluded faces resulting from large poses.

It is notable that even though we could employ Yang-Res to implement the acceleration in inference (Item 2), Yang-Res still suffers from the shortcomings of Items 1 and 3.

B. Additional Details on Ablation Studies

B.1. Coefficient Initialization for Refinement

In Fig. S.1, we demonstrate that initialization with inferred coefficients obtained from the makeup estimation network has a large impact on makeup refinement. For this demonstration, we chose blush as an example. It is important to note that without using these initial values, the quality of makeup estimation does not show significant improvement.

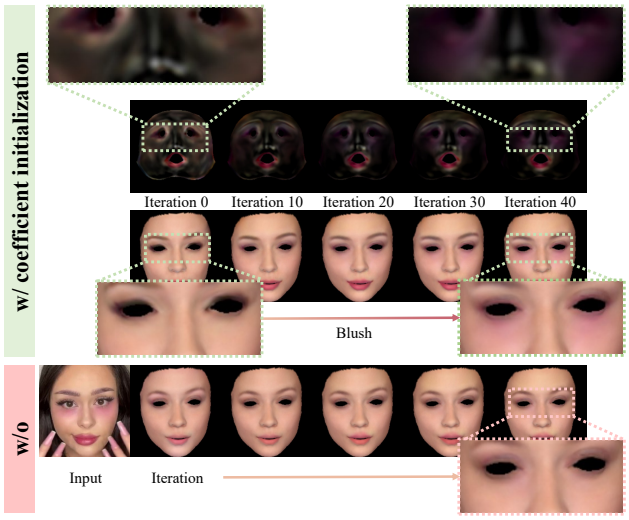

Figure S.1. Qualitative comparison of makeup refinement with and without initialization of coefficients inferred by the makeup estimation network. With initialization, detailed makeup features can be achieved with minimal iterations of refinement, as exemplified by the enhancement of blush details.

B.2. Details of Quantitative Evaluations

In the quantitative evaluations, we compare the differences between input images and the rendering results of 3D face reconstruction. Since there is no ground truth data to evaluate the makeup of invisible face regions, our experiments are based on visible areas. As shown in Fig. S.2, we first perform segmentation on the input images to obtain masks for the face, eyes, and lips. We employ morphological dilation with a 15×15 kernel, iterating three times, to expand the regions of the eyes and eyebrows. This approach ensures coverage of most of the eyeshadow area. Regarding the metrics, we use the eye regions and lip regions for the Histogram Matching (HM) metric, while we use the face region for the other metrics.

Figure S.2. Examples of segmented masks for quantitative experiments.

HM (eyes) ↓	**0.0028**	0.0029	0.0030	0.0030
HM (lips) ↓	0.0079	0.0082	0.0079	**0.0077**
RMSE ↓	**0.0775**	0.0782	0.0787	0.0797
SSIM ↑	**0.6118**	0.6074	0.6056	0.5989
LPIPS ↓	**0.0660**	0.0669	0.0676	0.0697

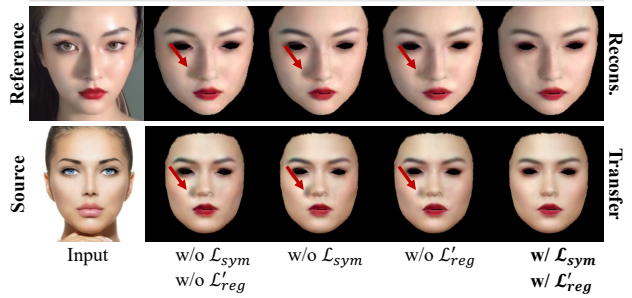

Figure S.3. Example of quantitative and qualitative comparison for loss functions. The improved quantitative metrics do not necessarily benefit downstream makeup-related tasks.

Tab. S.1 presents a quantitative ablation study on the loss functions. While $\mathcal{L}_{cyc}$, $\mathcal{L}_{sym}$, and $\mathcal{L}'_{reg}$ induce some fluctuations in quantitative metrics, they actually help the model avoid overfitting, qualitatively resulting in more natural and realistic visual makeup effects. In other words, the impact of these loss functions may not be easily captured by traditional quantitative measures but is significant in terms of visual outcomes. Therefore, we consider qualitative and quantitative evaluations as complementary assessment methods. For instance, as shown in Fig. S.3, overfitting to the shadows beside the nose as part of the makeup patterns might enhance quantitative metrics. However, this can hinder the effectiveness of makeup transfer across different individuals. The same rationale applies to user-friendly editing scenarios.

C. Additional Results

To the best of our knowledge, apart from Yang *et al*., no other relevant methods address makeup for 3DMM or model-based single image 3D makeup estimation. Hence, we primarily compared against Yang *et al*. as a baseline method. Figs. S.4 and S.5 show additional comparative results of makeup estimation, particularly in scenarios involving self-occluded faces. Figs. S.6 and S.7 display additional comparative results for model-based 3D facial reconstruction with makeup. We selected challenging makeup

Table S.1. **Comparative analysis of loss functions in 3DMM-based 3D face reconstruction.** Values in **bold** represent the best results.

Method	**Wild** [36]					**BeautyFace** [79]				
	HM(eyes)↓	HM(lips)↓	RMSE↓	SSIM↑	LPIPS↓	HM(eyes)↓	HM(lips)↓	RMSE↓	SSIM↑	LPIPS↓
w/o $\mathcal{L}_{cyc}$ (PCA)	0.0041	0.0078	**0.0604**	**0.6119**	**0.0667**	0.0035	**0.0076**	**0.0684**	**0.5021**	**0.0718**
w/ $\mathcal{L}_{cyc}$ (PCA)	0.0041	0.0078	0.0609	0.6111	0.0681	0.0035	0.0078	0.0690	0.5013	0.0733
w/o $\mathcal{L}_{cyc}$ (StyleGAN2 w/o Refine)	0.0042	0.0086	0.0634	0.6078	0.0707	0.0036	0.0087	0.0707	0.5021	0.0763
w/ $\mathcal{L}_{cyc}$ (StyleGAN2 w/o Refine)	0.0042	**0.0083**	**0.0618**	**0.6091**	**0.0685**	**0.0035**	**0.0084**	**0.0695**	**0.5031**	**0.0741**
w/o $\mathcal{L}_{sym}$ w/o $\mathcal{L}'_{reg}$ (StyleGAN2)	**0.0034**	**0.0071**	**0.0464**	**0.6386**	**0.0534**	**0.0029**	**0.0066**	**0.0621**	**0.5293**	**0.0601**
w/o $\mathcal{L}_{sym}$ (StyleGAN2)	0.0035	0.0072	0.0488	0.6313	0.0559	0.0030	0.0067	0.0633	0.5206	0.0628
w/o $\mathcal{L}'_{reg}$ (StyleGAN2)	0.0035	0.0072	0.0493	0.6309	0.0596	0.0030	0.0067	0.0637	0.5214	0.0654
w/ $\mathcal{L}_{sym}$ w/ $\mathcal{L}'_{reg}$ (StyleGAN2)	0.0036	0.0073	0.0517	0.6240	0.0608	0.0031	0.0068	0.0650	0.5134	0.0673

styles, such as uncommon colors and gradational eyeshadows, blushes, and lipsticks for comparison. We use DECA [24] as a reference that lacks a makeup prior model. As shown in Figs. S.8 and S.9, we demonstrate the makeup interpolation and transfer using makeup coefficients for both the PCA and StyleGAN2 models. It can be observed that we can achieve multi-interpolation by blending makeup coefficients. In our examples, we utilize bilinear interpolation for illustration.

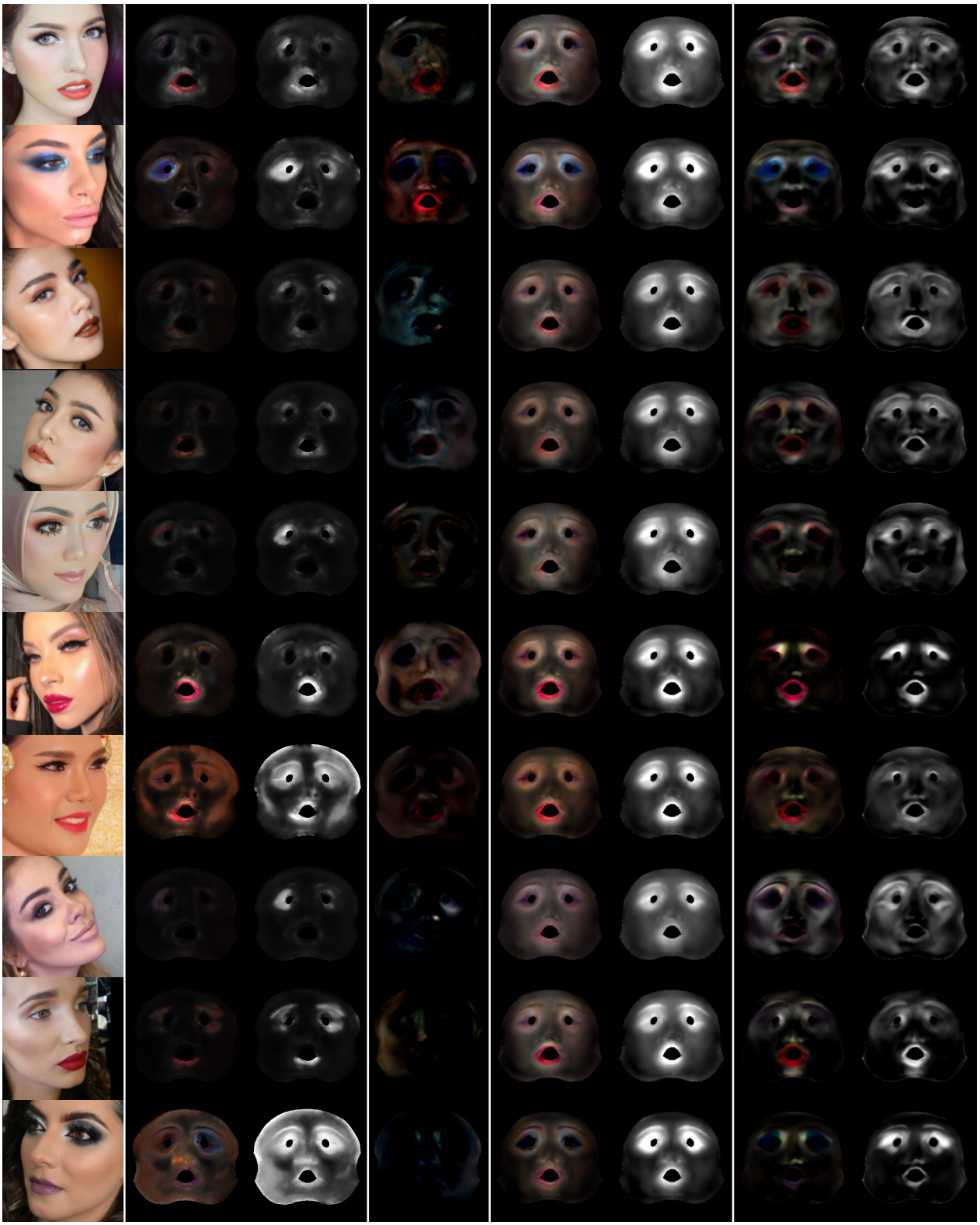

Figure S.4. **Comparison to previous work.** Our methods (PCA and StyleGAN2) outperform both Yang-Ext and Yang-Res [84], which show limitations in handling self-occluded faces.

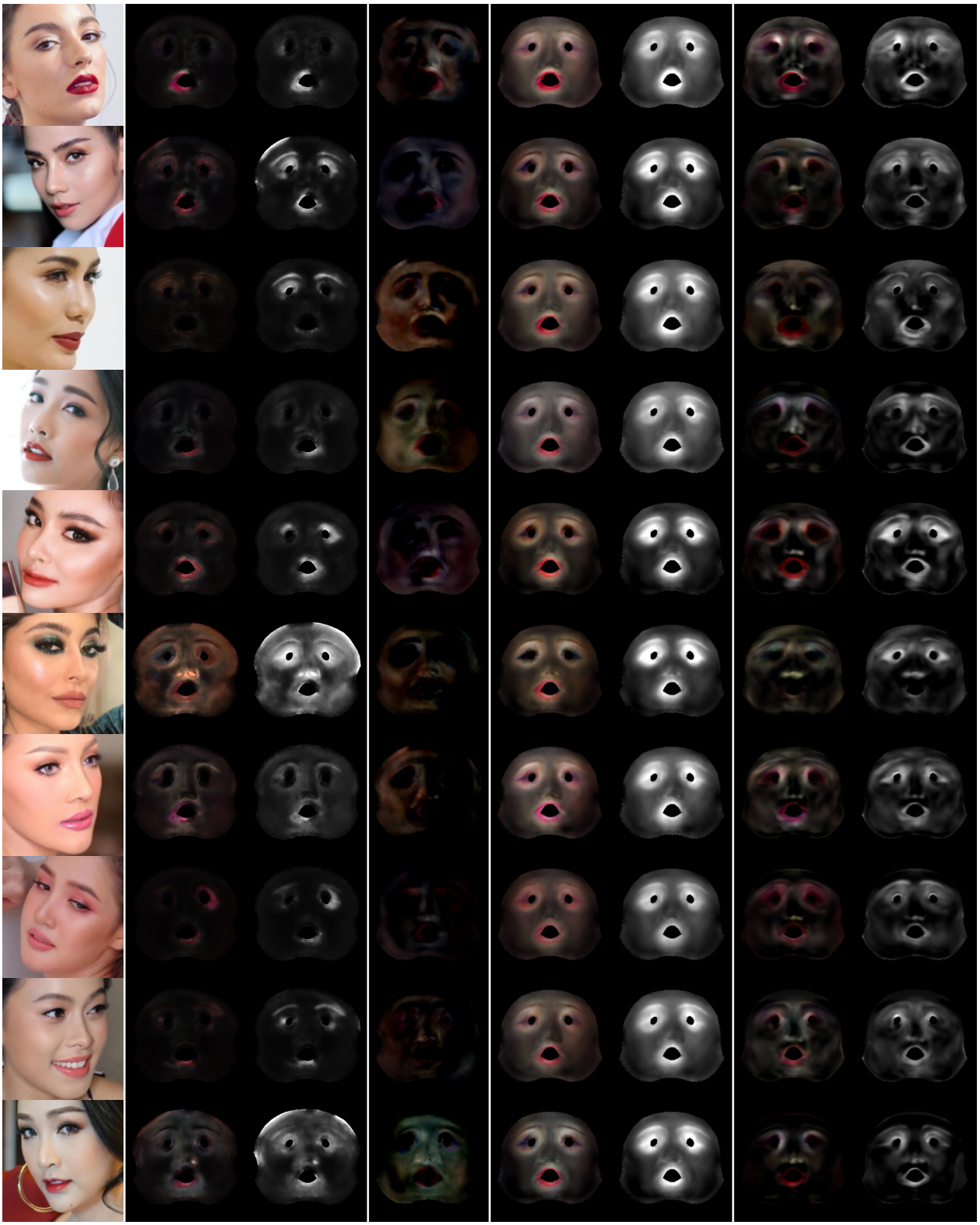

Figure S.5. **Additional comparison to previous work.** Our methods (PCA and StyleGAN2) outperform both Yang-Ext and Yang-Res [84], which show limitations in handling self-occluded faces.

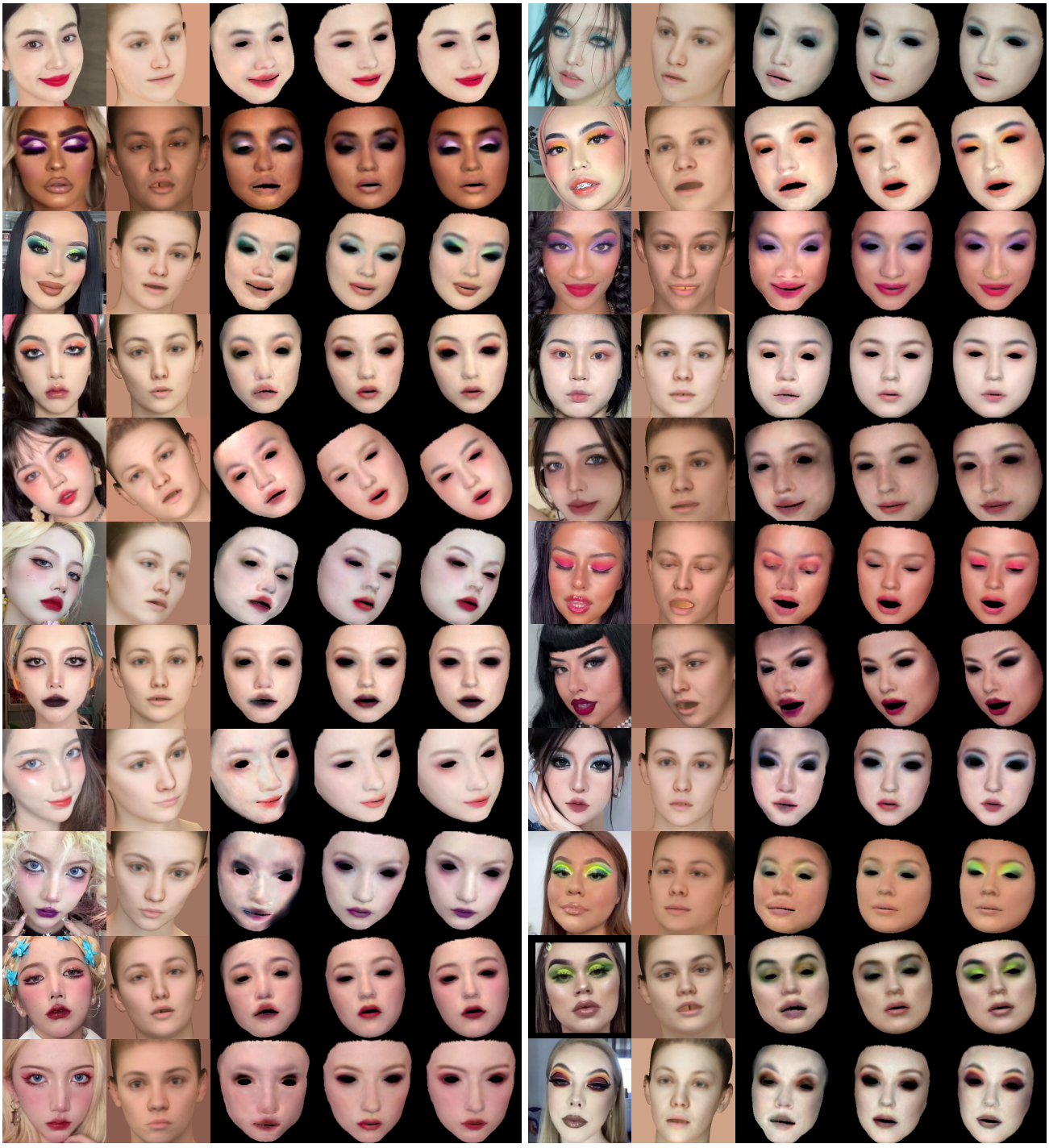

Figure S.6. **Comparison with 3D face reconstruction methods using our makeup prior models.** Our methods successfully reconstruct facial makeup. Specifically, our PCA model is capable of broadly recovering makeup colors, while our StyleGAN2 model achieves precise replication of complex makeup features, such as blush and gradational eyeshadow.

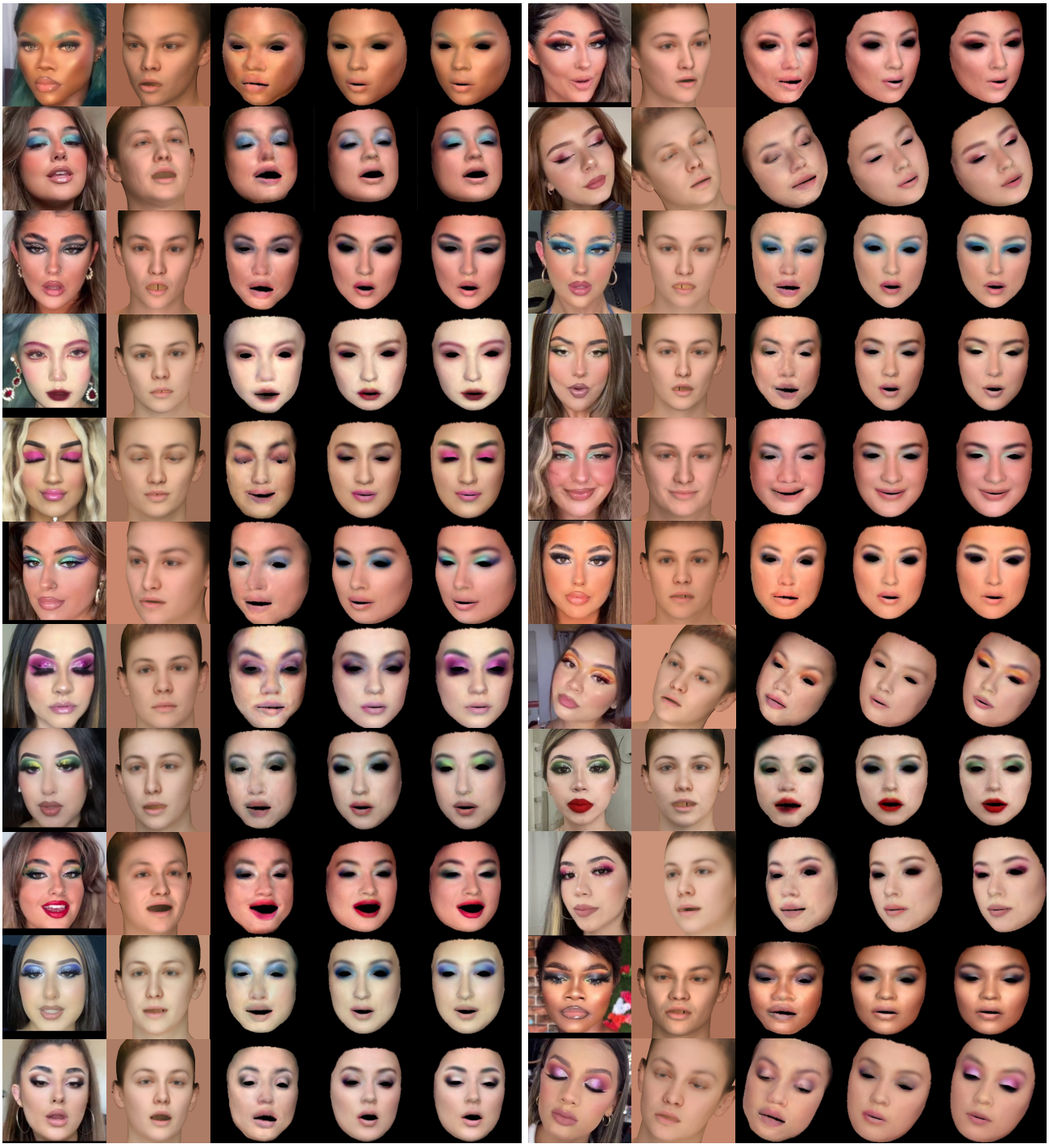

Figure S.7. **Additional comparison with 3D face reconstruction methods using our makeup prior models.** Our methods successfully reconstruct facial makeup. Specifically, our PCA model is capable of broadly recovering makeup colors, while our StyleGAN2 model achieves precise replication of complex makeup features, such as blush and gradational eyeshadow.

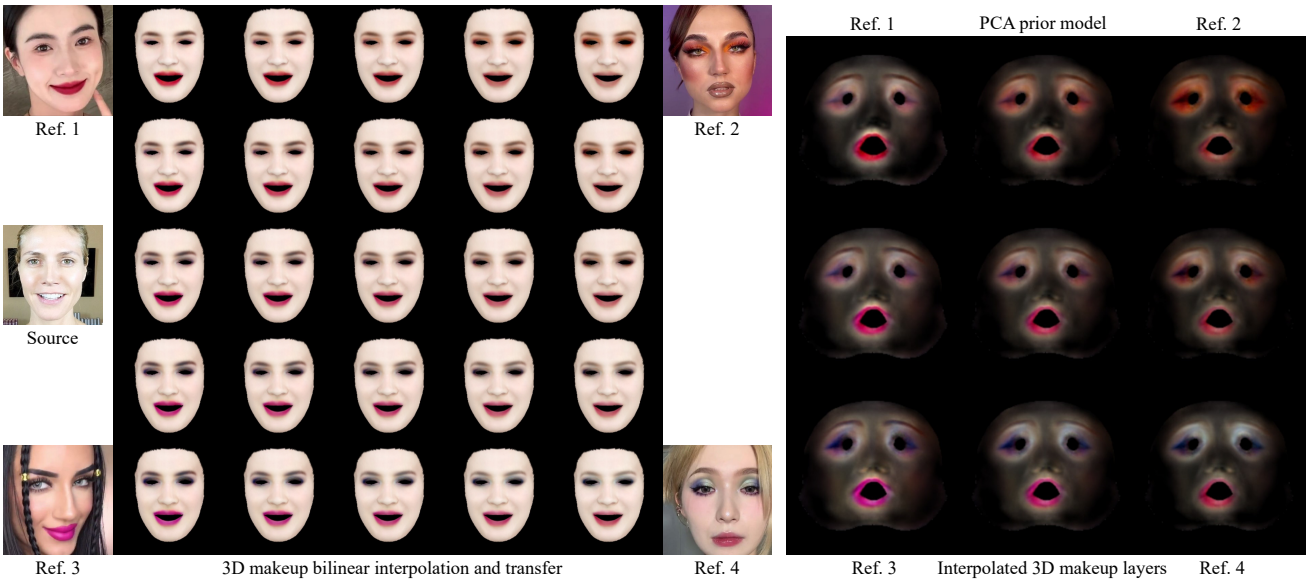

Figure S.8. **Results of 3D makeup interpolation and transfer using our PCA model.** Left: the bilinear makeup interpolation between four makeup styles. Right: the estimated makeup layers and the interpolated results using the makeup coefficient υ.

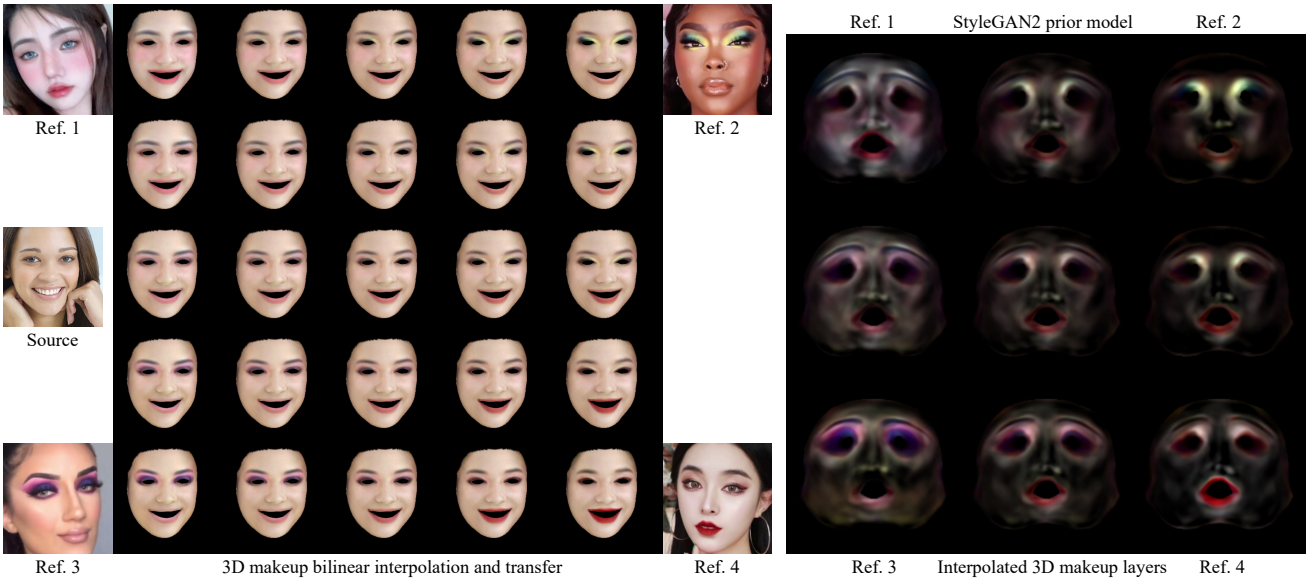

Figure S.9. **Results of 3D makeup interpolation and transfer using our StyleGAN2 model.** Left: the bilinear makeup interpolation between four makeup styles. Right: the estimated makeup layers and the interpolated results using the makeup coefficient $\mathbf{w}$.