

# S<sup>3</sup>-Face: SSS-Compliant Facial Reflectance Estimation via Diffusion Priors

## Supplementary Material

In this supplementary material, we first provide a detailed description of the datasets (Sec. 1) used in the manuscript. Next, we outline the details of our inference pipeline (Sec. 2), including UV completion and UV stitching. Then, we present additional experimental results (Sec. 3) to demonstrate the effectiveness of our method. Finally, we discuss the limitations of our current work and our plans for future research (Sec. 4). We provide a supplementary **video** demonstrating how our generated facial reflectance maps can be directly applied in CG pipelines for realistic relighting.

### 1. Datasets

**Captured Data.** We use Light Stage system to capture our reflectance raw data. Our system based on photometric stereo, achieves high-precision geometric reconstruction and pore-level reflectance map asset reconstruction. Our system is composed of 46 DSLR cameras with 12MP resolution and 7 DSLR cameras with 25MP resolution. During capturing, the system performs 15 types of lighting in 2 seconds, including one full-light scenario where all cameras capture images for high-precision geometric reconstruction; the other 14 types of polarized light are captured only by the 5 cameras with 25MP resolution, used for computing pore-level diffuse, tangent normal, specular, roughness, hemoglobin, melanin. The original capture under different polarized light is shown in Fig. 1. Our captured dataset contains 20160 images for each reflectance domain.

**3D Scan Store Albedos.** We buy 142 albedo maps from the 3D scan store, including 50 males and 92 females. These albedos have a balanced distribution regarding races.

**VISIA Skin Data.** To further validate the accuracy of our method in predicting hemoglobin and melanin, we collected skin analysis images from 128 real patients using the VISIA skin analyzer. The VISIA skin analyzer is a skin detection device widely used in medical aesthetics institutions, dermatology clinics, and the cosmetics industry. It utilizes cross-polarized light imaging and advanced image analysis techniques to detect and evaluate various skin characteristics, including surface pigmentation, capillary dilation, and inflammation conditions. Given that its imaging principle aligns with our acquisition system, we can use these real patient data for evaluation. As shown in Fig. 2, our method is able to accurately synthesize hemoglobin and melanin images. The inflammatory areas in the hemoglobin map and the pigmented areas in the melanin map are highly consistent with the ground truth.



Figure 1. Raw data plots from 7 DSLR cameras with 25 MP resolution under different polarized light.

Domain	PSNR $\uparrow$	SSIM $\uparrow$	LPIPS $\downarrow$
Diffuse Albedo	38.26	0.9169	0.0171
Specular Albedo	35.13	0.8748	0.0279
Normal	32.75	0.8477	0.0542
Hemoglobin	36.79	0.9099	0.0186
Melanin	36.29	0.8528	0.0230

Table 1. Demonstration of pre-trained VAE reconstruction capability for each reflectance domain.

### 2. Inference Details

**UV Completion.** For single-view input, we use OSTeC [3] for UV completion. This method uses a StyleGANv2 [4] model as a generator, iteratively predicts faces from different views through an iterative process, and then maps the results back into UV space for completion.

**UV Stitching.** For multi-view inputs, we obtain UV maps by stitching together the generated multi-view reflectance images. Experiments show that our stitching method performs robustly when the face angles in the input multi-view images are in the range of approximately  $-45^\circ$  to  $+45^\circ$ . Fig. 3 illustrates the visualization of the masks used in our stitching process across different views.

### 3. More Experimental Results

We present additional experimental results to validate the effectiveness of our model. First, we validated the reconstruction capability of the Variational Autoencoder [6] (VAE) in Stable Diffusion [5]. Tab. 1 demonstrates that SD’s VAE achieves excellent reconstruction quality, accurately capturing and reproducing details of the input images. This finding provides a solid foundation for our method.

Then, we verified the generalization ability of the model. Our model can not only effectively infer facial images but also process UV texture maps. Notably, our model demonstrated a powerful ability to handle UV textures with various topological structures, as shown in Fig. 4. This flexibility allows our method to adapt to different input formats and



Figure 2. Comparison of hemoglobin and melanin images. The first and third rows show the GT data, while the second and fourth rows are our estimated results. The first row presents two Rosacea cases, and the third row shows two Chloasma cases.



Figure 3. Example of three-view UV stitching masks in the inference pipeline. please note that these mask are generated by face shape unwrapping.

structures, greatly expanding its range of applications.

Besides, we showcase how the model generates the final UV reflectance maps from multi-angle inputs. As shown in Fig. 5, this not only verifies the robustness of our method but also showcases its advantages in handling complex and diverse inputs.

Finally, we showcase additional relighting results using various HDRI environments in Fig. 7. Our approach generates high-quality UV reflectance maps from input images, producing impressive rendering outcomes. We also include video demonstrations to highlight the dynamic nature of our relighting effects.

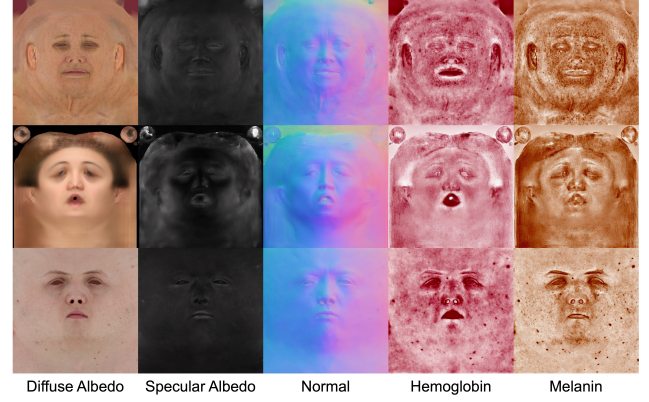


Figure 4. Examples of multiple UV texture inference with different topologies. The topological structures shown are: HiFi3D [1] (first row), FLAME [2] (second row), and OSTeC [3] (third row).

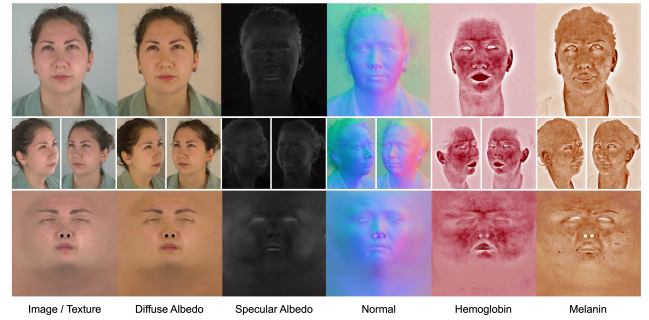


Figure 5. Example of multi-view inference. Given input images from multiple views, we predict all reflectance domains for each viewpoint separately, then stitch them together to obtain the final reflectance map.

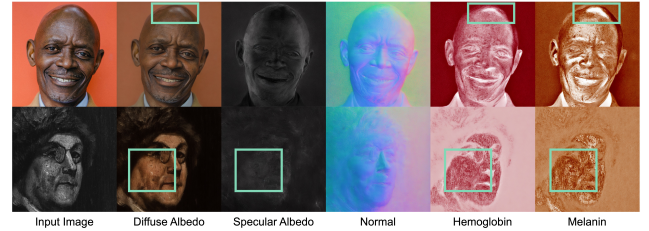


Figure 6. Limitations of our method. We demonstrate how bright spots and noise in input images affect our estimation results, both of which lead to errors in the hemoglobin and melanin maps.

## 4. Limitation and Future Works

Although our method can achieve excellent results, input images under extreme lighting conditions still pose a challenge. For example, our method struggles to remove bright spots in diffuse albedo, which in turn affects the estimation of hemoglobin and melanin. Given that our method is a pixel-level aligned estimation, low-quality or blurry images



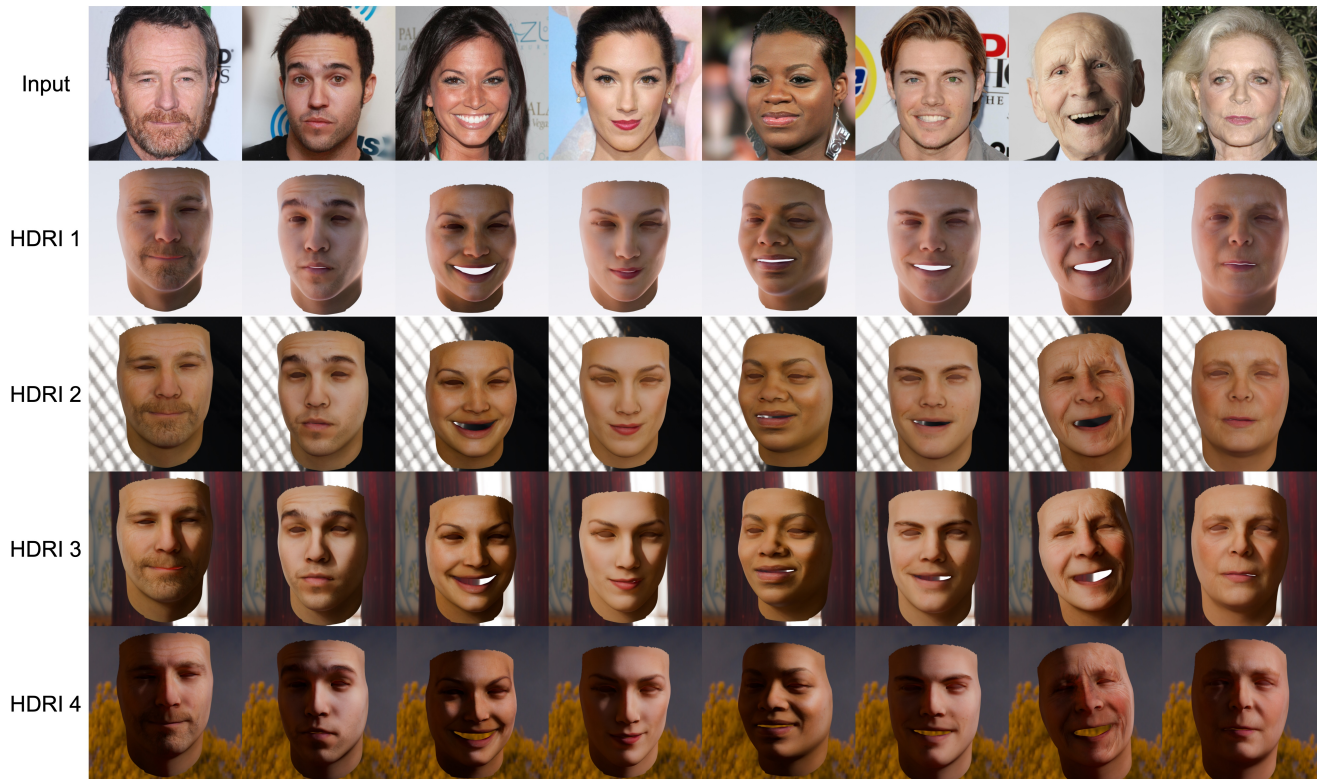


Figure 7. More relighting results. Each row of images maintains the same HDRI illumination.

may also lead to errors in reflectance estimation, such as misinterpreting noise as melanin, as shown in Fig. 6.

In the future, we will introduce more high-quality data to improve the model performance and explore more applications in the field of dermatology.

## References

- [1] Haoran Bai, Di Kang, Haoxian Zhang, Jinshan Pan, and Linchao Bao. Ffhq-uv: Normalized facial uv-texture dataset for 3d face reconstruction. In *CVPR*, 2023. 2
- [2] Yao Feng, Haiwen Feng, Michael J Black, and Timo Bolkart. Learning an animatable detailed 3d face model from in-the-wild images. *TOG*, 2021. 2
- [3] Baris Gecer, Jiankang Deng, and Stefanos Zafeiriou. Ostec: One-shot texture completion. In *CVPR*, 2021. 1, 2
- [4] Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. Analyzing and improving the image quality of stylegan. In *CVPR*, 2020. 1
- [5] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *CVPR*, 2022. 1
- [6] Aaron Van Den Oord, Oriol Vinyals, et al. Neural discrete representation learning. In *NeurIPS*, 2017. 1