Inclusive & photo-realistic albedo editing through latent space translation

Swati Kanwal
University of British Columbia

Introduction

Figure 1: A human face and its corresponding albedo map [3]

Problem Statement:

- Given a 4096 x 4096 face albedo texture image, generate a corresponding albedo image for different expressions and ages

Motivation:

- Skilled digital artists invest substantial effort in fine-tuning various spatially varying shading parameters to accurately model and modify rendered human faces
- There is an even greater need to train models that are gender, race, age inclusive and truly showcase human diversity whilst preserving photorealism

Challenges:

- Human skin appearance depends on biophysical factors & and intricate interactions of light within skin tissues and thus sis complex to model
- Facial skin appearance varies with age, gender, race, and expressions, among other factors
- Creating high-resolution 3D datasets containing albedo images of the face is prohibitively expensive
- Translated albedo must align with the mesh such that their UV coordinates match
- Some 3D datasets do exist, they are often limited in size and not easily accessible to the public
- No 3D dataset with albedos that has variations age and emotion for the same individual

Background

Human Skin two layer model: Rendering a human face realistically is complex because it depends on biophysical factors and intricate interactions of light within the skin tissues. The two main chromophores that dictate the appearance of the skin are:

- Melanin
- Haemoglobin

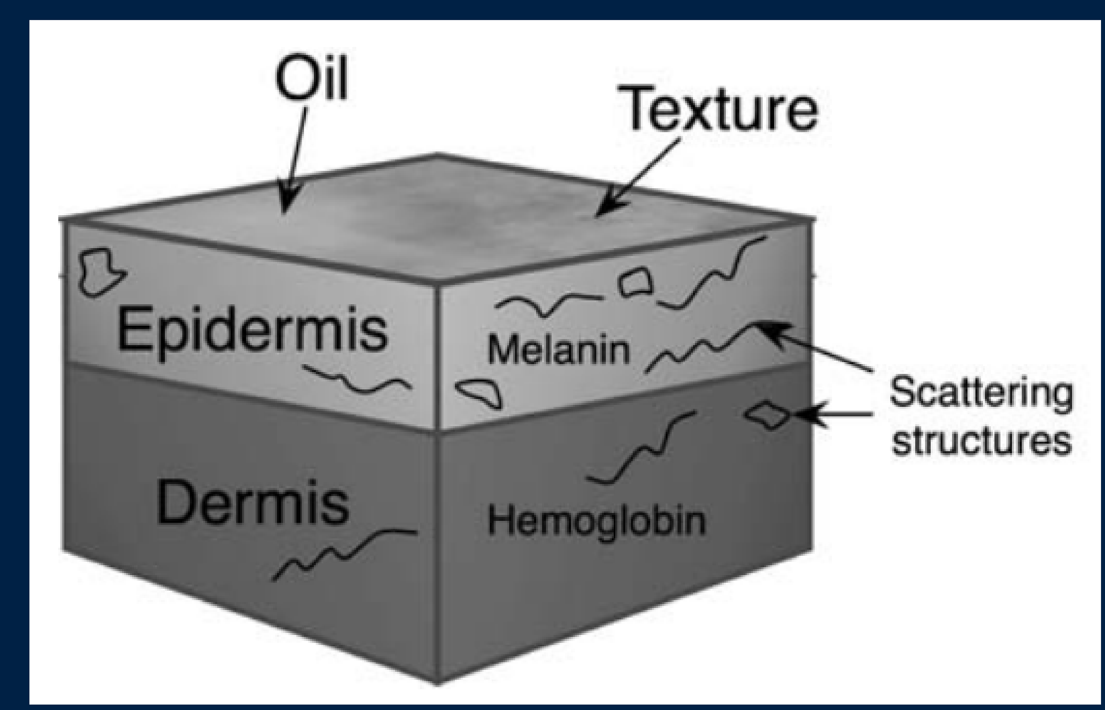

Figure 2: Two layer model of the skin [6]

Spectral Data: Recent research [1] has proposed an encoder-decoder-based solution that learns a mapping between RGB and chromophore (biophysical) space, enabling the manipulation of human face albedos at the chromophore level. While this approach is effective for different skin types it does not account for age, expression, and other factors that influence skin appearance.

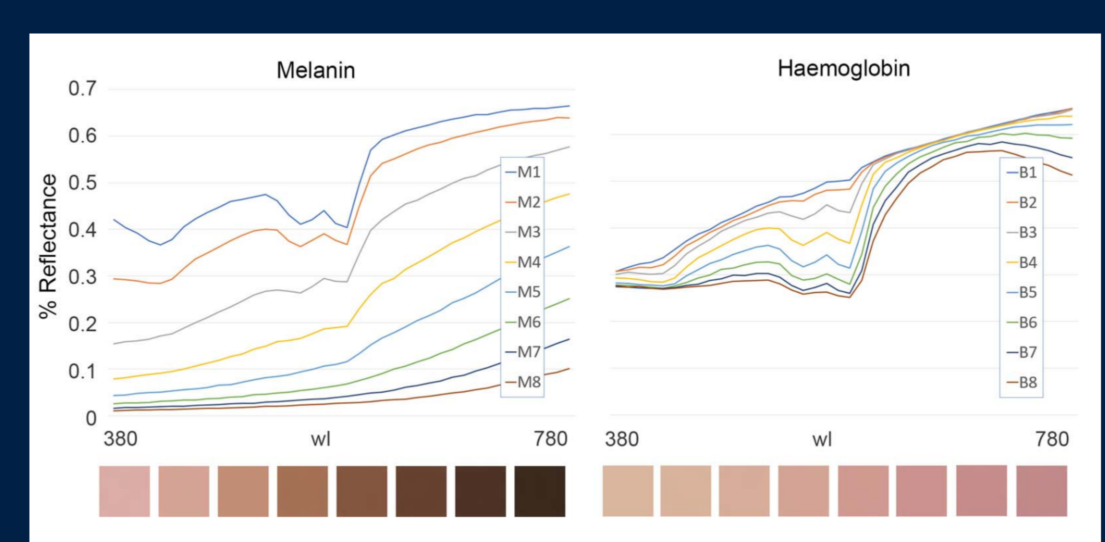

Figure 3: Relationship between Melanin, Haemoglobin and reflectance as reported in [1]

InterFaceGAN: There has been past work that utilizes style transfer for human faces to modify their expression, race, expression etc. However, not a lot of work exits for albedos

Figure 4: Results of style transfer using InterFaceGAN for human faces [4]

Methodology

StyleGAN 3: is an extension of the GAN architecture that uses a mapping network to map points in the latent space to an intermediate latent space in order to control the style at each point in the generator model

FFHQ-UV dataset: is a large-scale facial UV-texture dataset that contains over 50,000 high-quality texture UV-maps with even illuminations, neutral expressions, and cleaned facial regions, which are desired characteristics for rendering realistic 3D face models under different lighting conditions.

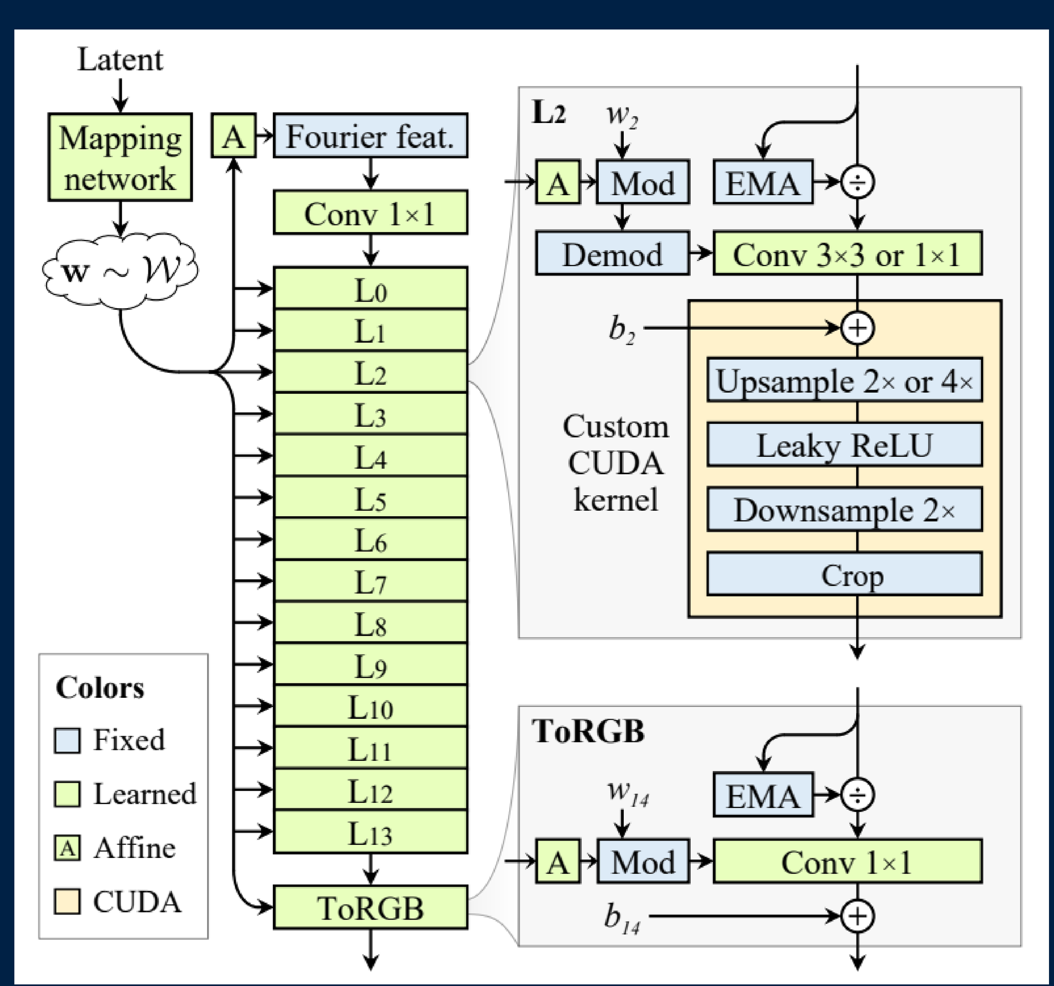

Figure 5: StyleGAN architecture [2]

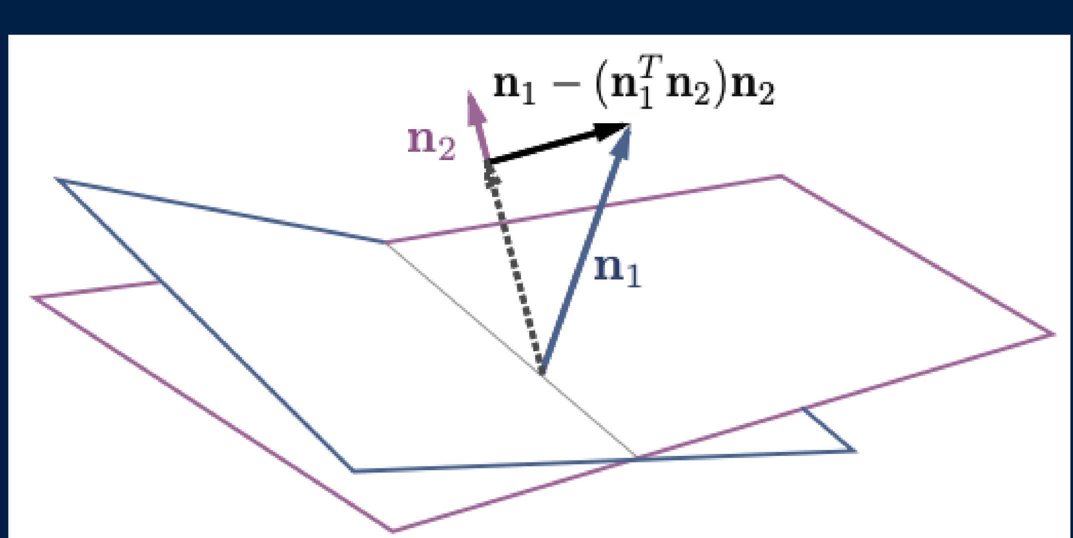

Figure 6: Latent Space Translation [3]

To train our model, we employ transfer learning to train a pre-trained FFHQ-StyleGAN model with FFHQ-UV albedos to learn the albedo latent space. Subsequently, we use boundaries learned by InterFaceGAN[4], specifically age and other attribute boundaries, to facilitate latent space transformations for generating textures with attribute variations.

The rationale behind this approach is straightforward: it capitalizes on the similarity between a human face and it's albedo. Since the age and other attributes of the human face correspond to those of the albedo, the same boundaries apply in the latent space

Results

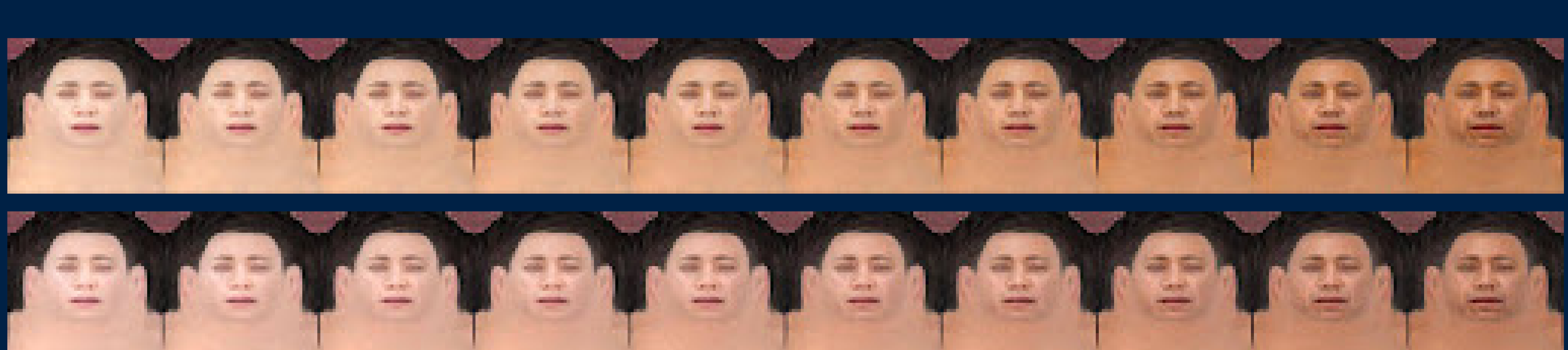

Figure 5: Two sample results of Age & Race translation given random latent codes

Future Work

This solution can be further expanded by incorporating an encoder into the albedo space.

This encoder would allow embedding any albedo into the space and subsequently translating it within the latent space to produce photorealistic edits.

References

[1]Aliaga, C., Hery, C., & Xia, M. (2022). Estimation of Spectral Biophysical Skin Properties from Captured RGB Albedo.
[2]Karras, T., Laine, S., & Aila, T. (2018). A Style-Based Generator Architecture for Generative Adversarial Networks.
[3]Bai, H., Kang, D., Zhang, H., Pan, J., & Bao, L. (2022). FFHQ-UV: Normalized Facial UV-Texture Dataset for 3D Face Reconstruction.
[4]Shen, Y., Yang, C., Tang, X., & Zhou, B. (2020). InterFaceGAN: Interpreting the Disentangled Face Representation Learned by GANs.
[5] Donner & Jensen (2006), A spectral BSSRDF for shading human skin

Acknowledgements

This work was done as part of an internship at Huawei R&D Canada. Indebted to Wei Sen Loi and Abraham Beauferris for their research guidance and support.