REPORT ON BIOFACENET: DEEP BIOPHYSICAL FACE INTERPRETATION

1. RESULTS FROM PYTORCH IMPLEMENTATION

Input :	3 X 64 X 64
BatchSize :	25
Training Set Images :	9000
Optimizer :	Adam
Learning Rate :	1e-05
Output :	3 X 64 X 64 (Reconstructed Image)

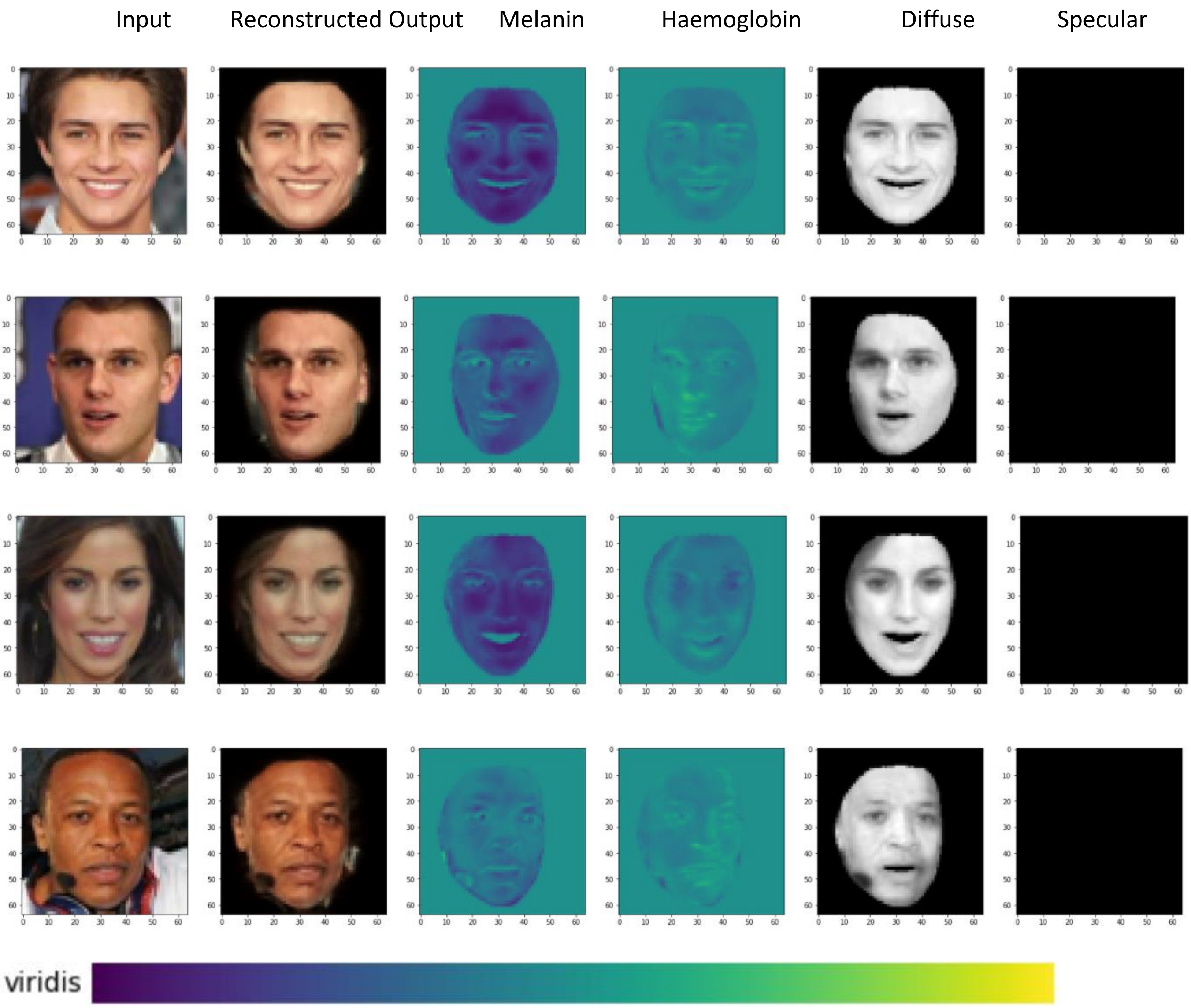

Figure 1: Rows 1-4 show input, reconstructed, melanin, haemoglobin, diffuse and specular maps respectively, the 5th row shows Viridis color map used for melanin and haemoglobin maps

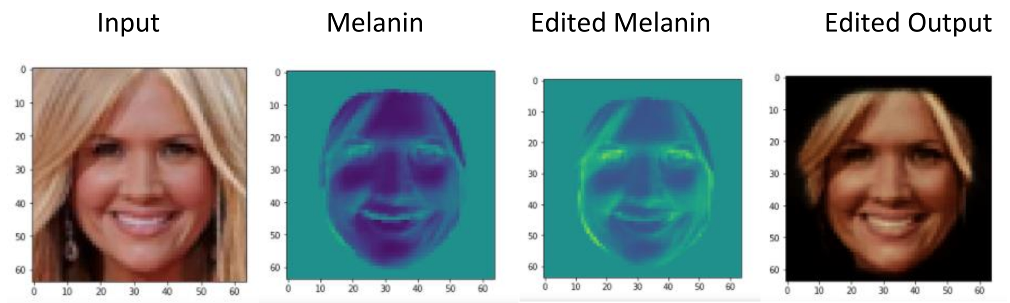

Figure 2: Input, melanin, melanin increased by 0.4 and output constructed from the edited melanin map

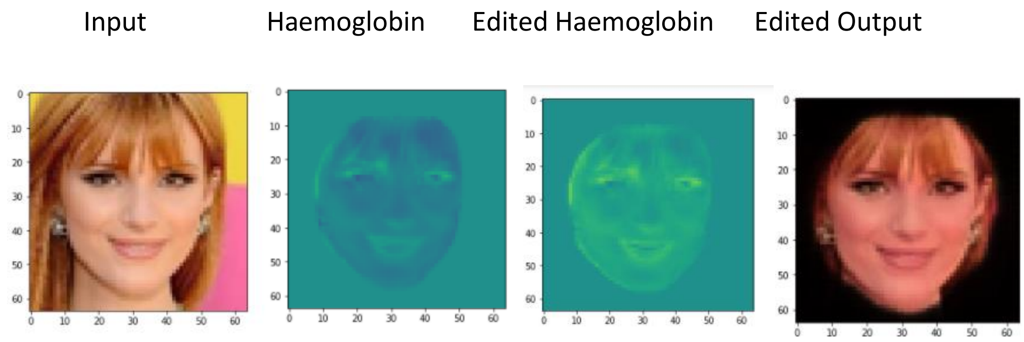

Figure 3: Input, haemoglobin, haemoglobin increased by 0.3 and output constructed from the edited haemoglobin map

2. PAPER INTERPRETATION

The paper has 2 goals, Firstly, to change the method of constructing faces by including specular component. This in the past was ignored, as the face construction was no different than any other object. The consequence of including a specular component gives an acceptable level of detail caused by specular flections make it more plausible. Secondly, to edit face models by using biophysical maps, i.e given the input picture to produce a plausible reconstruction by changing haemoglobin, melanin, reflecting (diffuse and specular) components.

Implementation of model from paper: The model from the paper feeds input to a U-Net type architecture[1]. A interesting feature of the U-Net is the usage of skip connections, the concatenation of feature maps from the layers involved in reducing spatial resolution to the layers increasing spatial resolution. The forward path passes through a trainable encoder, from which is sent to 4 decoders (one for each map- melanin, haemoglobin, diffuse and shading maps) and fully connected layers to predict camera model and illuminant spectral power density. The maps along with the camera parameters and illuminant spectral power density are then fed to a hand-crafted decoder to reconstruct the image.

3. PIPELINE ISSUES

- According to the paper, the dataset used for training is CelebA taken from [2]. According to the paper the BioFacenet[6] was trained on 224 x 224 size images, yet the dataset currently from the source is 64 x 64

- MATLAB implementation of the model does not seem to have gamma correction applied to the reconstructed images yet mentioned in the paper.

4. SUGGESTED IMPROVEMENTS

- One of the loss components is L1 sparsity on the specular shading leading to the Specularity values to be low (i.e. mostly black pixel values) which gives inaccurate results as compared to the pseudo ground truth. Refer figure 3 taken the from the paper[6].

Figure 4: Column 1 shows ground truth and column 2 shows Specular map from BIOFACENET[6]

Training the model with a lower weight for sparsity loss yields better for specular map. The output map still retains physical meaning while the L1 sparsity on specular shading plays a less significant role to the new loss. Figure 4 shows specular map from Pytorch model when sparseweight (weight for L1 sparsity on specular shading) is 1e-8 run on 10 epochs on 1000 images

Figure 5: Column 1 shows Input and column 2 shows Specular map after dividing by the maximum value

- Provide a specular loss component loss from the pseudo ground truth in H. Kim et al[3]
- L2 norm used for appearance loss can be replaced with superior methods such MS-SSIM, a multi-scale version of SSIM which correlate better with human visual systems[4]
- Atrous Spatial Pyramid Pooling (ASSP) was introduced by L-C Chen et al[5]. Improvements may arise by making use of this method.

5. REFERENCES

[1] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation

[2] Zhixin Shu, Ersin Yumer, Sunil Hadap, Kalyan Sunkavalli, Eli Shechtman, and Dimitris Samaras. Neural face editing with intrinsic image disentangling. In Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 2017

[3] 2. H. Kim, H. Jin, S. Hadap, and I. Kweon. Specular reflection separation using dark channel prior. In CVPR, pages 1460– 1467, 2013

[4] https://research.nvidia.com/sites/default/files/pubs/2017-03_Loss-Functions-for/NN_ImgProc.pdf

[5] 6.L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, “Encoder Decoder with Atrous Separable Convolution for Semantic Image Segmentation.” in European Conference on Computer Vision (ECCV), 2018