Towards High Fidelity Face Relighting with Realistic Shadows
Supplementary Materials

Figure 1: **Performance of [5] on FFHQ**. (a) input image, (b) target lighting, (c) Nestmeyer et al. [5]. The method of [5] does not seem to generalize well to in-the-wild images, which is likely caused by the limited subject diversity and night-time setting of their training set.

1. Qualitative Results on FFHQ for Nestmeyer et al. [5]

We include qualitative results on FFHQ for Nestmeyer et al. [5] (See Fig. 1). Overall, their model seems to generalize poorly to in-the-wild images, which is likely caused by the limited number of subjects (21) in their training set and the night-time setting of their images. Therefore, we chose not to compare with them qualitatively on FFHQ.

2. Ablations on L_{face}, L_{gradient}, L_{DSSIM}, and $L_{\text{adversarial}}$

We perform 4 additional ablations to show that the addition of each loss function improves our model's relighting performance. We train 4 additional models, each of which excludes one of L_{face}, L_{gradient}, L_{DSSIM}, and $L_{\text{adversarial}}$, and compare the performance with our proposed model. We evaluate their performance quantitatively on Multi-PIE [3] (See Tab. 1). We find that our proposed model, which includes all loss functions, achieves the best overall performance across our 3 evaluation metrics.

We further demonstrate the benefits of including these 4 losses qualitatively on FFHQ [4] (See Fig. 2). We find that

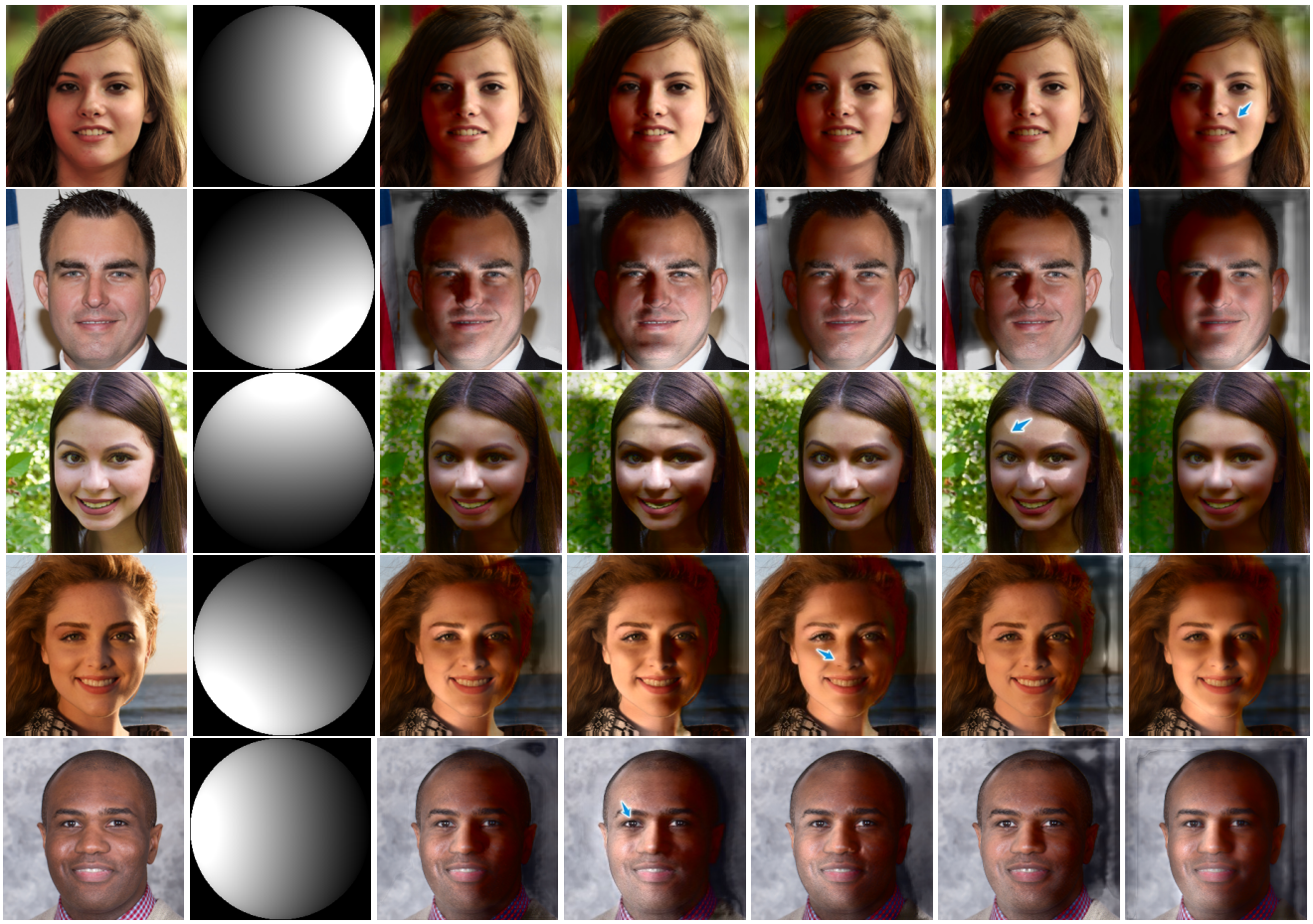
Method	Si-MSE	MSE	DSSIM
w/o L_{face}	0.0262	0.0346	0.1734
w/o L_{gradient}	0.0223	0.0297	**0.1589**
w/o L_{DSSIM}	0.0227	0.0293	0.1703
w/o $L_{\text{adversarial}}$	0.0275	0.0361	0.1800
Proposed	**0.0220**	**0.0292**	0.1605

Table 1: **Additional Ablation Studies**. We perform ablation studies on L_{face}, L_{gradient}, L_{DSSIM}, and $L_{\text{adversarial}}$. We find that the proposed model (which includes all loss functions) achieves the best overall performance.

excluding L_{face} lowers the model's ability to preserve the subject's facial details, as expected since L_{face} is responsible for ensuring face feature consistency for the same subject under different lighting. Removing L_{gradient} can lead to smoother edges on the face, which is understandable given it encourages the edges of the predicted and target ratio images to be similar. Another effect is the quality of the cast shadows generally decreases, which makes sense because shadow borders are edges in the image. Excluding L_{DSSIM} can lead to unnatural transitions from illuminated to shadowed regions of the face: rather than transitioning gradually, the shift between the regions can be abrupt. This behavior can arise since the SSIM metric measures if the local image patterns of the predicted image match the target image. Without the corresponding loss L_{DSSIM}, these unnatural transitions may occur. Finally, we find that removing $L_{\text{adversarial}}$ noticeably reduces the visual quality of the relit images and makes the output more blurry. This verifies that the inclusion of PatchGAN [2] discriminators helps the model capture high frequency details and improve the photorealism of the results. We thus assert that qualitatively, our proposed model produces the best results.

3. Ratio vs Relit Image Estimation

We provide more insights concerning the advantages of estimating the ratio image instead of the relit image directly. Regressing a ratio image is easier for the network since it contains less high-frequency detail than the target image. This can be shown by the power spectrums of a target Multi-



(a) Source Image (b) Target Lighting (c) Proposed (d) w/o L_{face} (e) w/o L_{gradient} (f) w/o L_{DSSIM} (g) w/o $L_{\text{adversarial}}$

Figure 2: **Qualitative Results on FFHQ for Ablation Models.** We compare the qualitative performance of our proposed model with 4 additional ablation models, each of which removes one loss function. Removing L_{face} degrades the model's ability to preserve the subject's facial features, removing L_{gradient} causes the edges on the face to become smoother (fourth row, around the nose) and lowers the quality of the produced cast shadows, removing L_{DSSIM} can lead to unnatural transitions from illuminated regions of the face to shadowed regions (third row), and removing $L_{\text{adversarial}}$ leads to a noticeably more blurry output. The proposed model that uses all of these losses yields the best qualitative performance.

PIE image and its corresponding ratio image in Fig. 3. It's clear from the power spectrums that the target image contains more high-frequency details than the ratio image. The ratio image contains more lower frequency information, and is therefore easier to estimate. We can thus more easily preserve facial details by regressing ratio images instead of relit images directly.

4. Shadow Removal

As shadow removal is a challenging task in face relighting, we demonstrate our model's capacity to remove hard cast shadows. Fig. 4 shows relit results on FFHQ subjects using a frontal target lighting, which should remove facial shadows. Our model either removes hard shadows completely or softens them.

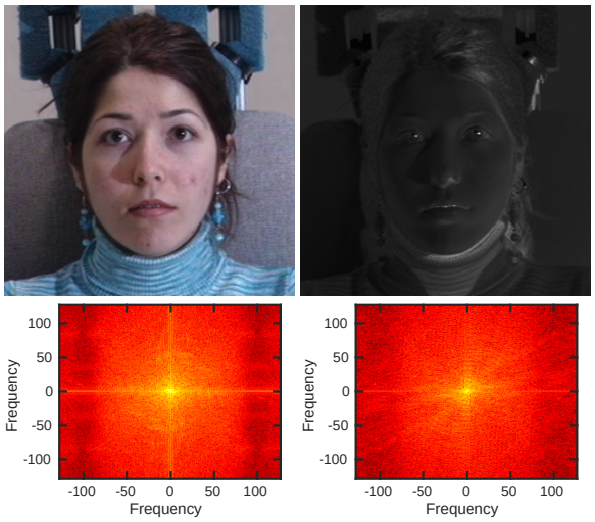
5. Non-frontal Poses

To demonstrate our model's performance on face images with large poses, we apply our model to FFHQ subjects with non-frontal poses. As shown in Fig. 5, our model can handle large poses gracefully.

6. Lighting Estimation

We provide more details on how we estimate the groundtruth lightings for the DPR [7], Yale [1], and Multi-PIE [3] datasets.

The DPR images provide the SH coefficients as groundtruth lighting. The Yale dataset provides the lighting direction for each image, and we treat each light as a directional light. Multi-PIE provides positions for every light,



(a) Target Image (b) Ratio Image

Figure 3: **Power spectrums**. Comparing the power spectrums of a target Multi-PIE image and its corresponding ratio image, it is clear that the ratio image contains less high-frequency detail than the target image. The ratio image is thus easier for our model to regress than the target image, which indicates that it is an easier way to preserve facial details.

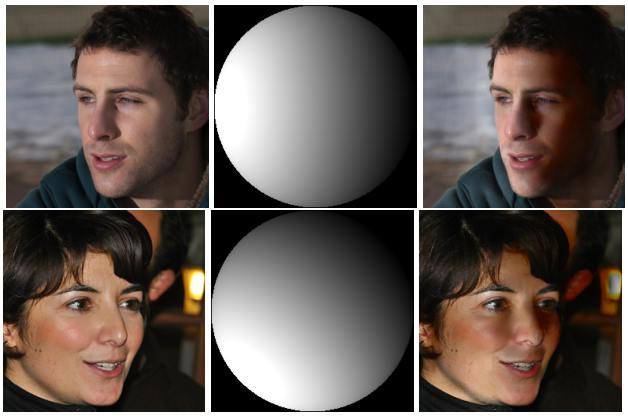
(a) Input Image (b) Target Lighting (c) Result

Figure 4: **Shadow Removal**. Our model is able to remove or significantly soften hard cast shadows.

which we treat as point lights. For Yale and Multi-PIE, we project the light source to SH basis functions on the unit sphere to get the SH coefficients. The ambient component is estimated using shadow masks, as explained in Sec. 3.5 of the main paper.

7. Inference Time

We compare our model's inference time with SfSNet [6] and Nestmeyer *et al.* [5], two methods with available code and significantly different architectures from our work.



(a) Input Image (b) Target Lighting (c) Result

Figure 5: **Non-frontal poses**. Our model is able to properly relight images with large, non-frontal poses.

Running on Multi-PIE, [6] takes 0.0087 seconds per image, [5] takes 0.0750 seconds per image, and our model is the fastest at 0.0075 seconds per image.

8. Qualitative Video

We include a video with 5 FFHQ subjects to show that our model can faithfully perform face relighting across many different target lighting directions and many diverse subjects. Our video also shows that our model can handle varying lighting intensities, as seen when we move the light source closer to the last subject at the end of the video.

References

[1] Athinodoros Georghiades, Peter Belhumeur, and David Kriegman. From Few to Many: Illumination Cone Models for Face Recognition under Variable Lighting and Pose. *PAMI*, 2001. 2

[2] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *NeurIPS*, 2014. 1

[3] Ralph Gross, Iain Matthews, Jeffrey Cohn, Takeo Kanade, and Simon Baker. Multi-PIE. *Image and Vision Computing*, 2010. 1, 2

[4] Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In *CVPR*, 2019. 1

[5] Thomas Nestmeyer, Jean-Francois Lalonde, Iain Matthews, and Andreas Lehrmann. Learning Physics-guided Face Relighting under Directional Light. In *CVPR*, 2020. 1, 3

[6] Soumyadip Sengupta, Angjoo Kanazawa, Carlos D. Castillo, and David W. Jacobs. SfSNet: Learning shape, refectance and illuminance of faces in the wild. In *CVPR*, 2018. 3

[7] Hao Zhou, Sunil Hadap, Kalyan Sunkavalli, and David W Jacobs. Deep single-image portrait relighting. In *ICCV*, 2019. 2