Supplementary Material: Deep Portrait Delighting

Joshua Weir*[0000−0002−4445−7278], Junhong Zhao*[0000−0001−7031−3828], Andrew Chalmers[0000−0001−6457−7341], and Taehyun Rhee*[0000−0002−6150−0637]

Computational Media Innovation Centre, Victoria University of Wellington, New Zealand
{josh.weir, j.zhao, andrew.chalmers, taehyun.rhee}@vuw.ac.nz
https://www.wgtn.ac.nz/cmic

1 OLAT and Target synthesis

In sec. 3.1 of our paper, we describe how we process our data for training. Here, we provide a deeper overview of how we remove the effect of the room lights in the Multi-PIE [2] flash images, and synthesize the target de-lit image.

Since the images with no camera-flashes (room-lit) are available, we can simply remove their effect with the following equation.

$$\mathbf{I_{ratio}} = \mathbf{1} - \left(\frac{\mathbf{L(I_{roomlights})}}{\mathbf{L(I_{flash})}}\right) \tag{1}$$

$$\mathbf{I_{no\_ambient}} = \mathbf{I_{ratio}} \odot \mathbf{I_{flash}} \tag{2}$$

Where $\odot$ is pixel-wise multiplication, $\mathbf{L(A)}$ refers to the luminance channel of $\mathbf{A}$ in Lab color space. $\mathbf{I_{roomlights}}$ and $\mathbf{I_{flash}}$ are the room-lit and flash images respectively.

Another problem is that the room lights create sharp specular reflections on the faces of some subjects. Since the brightness of these regions is near full (images were taken with a LDR camera), removing these from our flash images results in dark blemishes across the face (Fig. 1 (c)). We detect these regions using the following equation:

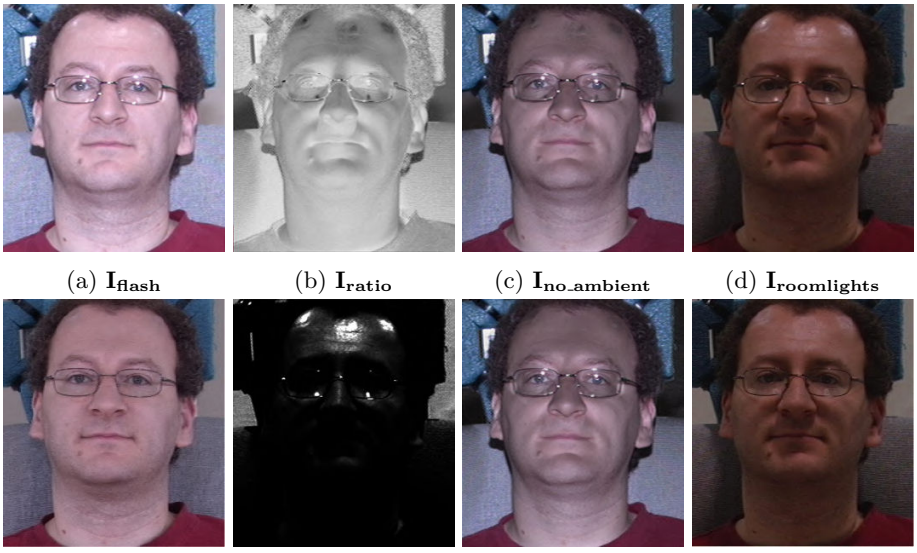
$$\mathbf{I_{specular}} = \mathbf{min}(\mathbf{1}, \frac{\mathbf{I_{roomlights}}^{\mathbf{2}}}{\mathbf{I_{average}}})^{\mathbf{4}}), \tag{3}$$

where $\mathbf{I_{average}}$ is the mean average over all flash images. We correct these small regions using Navier-Strokes [1] inpainting to produce $\mathbf{I_{OLAT}}$. An illustration of each step is shown in Fig. 1.

To synthesize the target de-lit image, we approximate ambient lighting as the average over all our 18 OLAT images, producing $\mathbf{I_{dlt}}$. From here, we modify $\mathbf{I_{dlt}}$ using the following equation:

$$\mathbf{L(I_{dlt})} = \mathbf{L(I_{dlt})} + 6\mathbf{L(I_{room\_nospec})}, \tag{4}$$

where $\mathbf{I}_{\mathbf{room\_nospec}}$ is $\mathbf{I}_{\mathbf{roomlights}}$ with specularities removed (like $\mathbf{I}_{\mathbf{OLAT}}$). Adding the luminance of the room-lit image back into $\mathbf{I}_{\mathbf{dlt}}$ makes the lighting appear more uniform.



(a) $\mathbf{I}_{\mathbf{flash}}$ (b) $\mathbf{I}_{\mathbf{ratio}}$ (c) $\mathbf{I}_{\mathbf{no\_ambient}}$ (d) $\mathbf{I}_{\mathbf{roomlights}}$

(e) $\mathbf{I}_{\mathbf{dlt}}$ (f) $\mathbf{I}_{\mathbf{specular}}$ (g) $\mathbf{I}_{\mathbf{OLAT}}$ (h) $\mathbf{I}_{\mathbf{room\_nospec}}$

Fig. 1: Illustration of our data synthesis stage.

2 More comparisons

In sec. 4.2 of our paper, we showed qualitative comparisons against TR and EMR against our testing dataset. However, due to different ground-truth modalities (*i.e.* albedo quality and targeted regions), we don't compare quantitatively against the pretrained models of TR and EMR in sec. 4.3. As an extension to Fig. 4 in the main paper, we provide a more in-depth qualitative comparison.

Upper-body portraits In Fig. 3 We provide examples from our testing dataset of images containing significantly large clothing regions. Since EMR wasn't trained on such examples, we do not present their results in this section, and instead show only EMR (retrained).

Face portraits In Fig. 4, we show more comparisons of images cropped around the face region. Here, results of all prior works mentioned in the paper are shown.

3 More soft-shadow loss examples

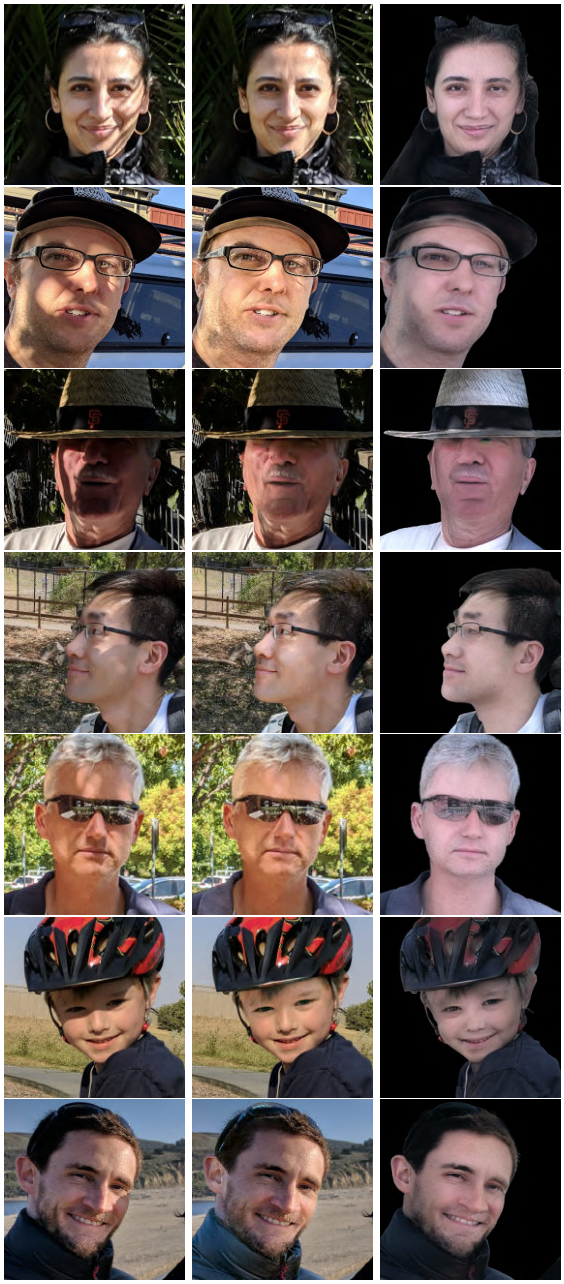
In Fig. 5, we show more qualitative examples of our soft-shadow regularization (see sec. 3.2 of our paper). Input images are from our **Alt. Lighting** dataset (see sec. 4.4 of our paper).

4 Shadow Removal

The paper of Zhang *et al.* [3] focuses only on a subset of the delighting task: shadow removal/softening, hence we did not compare against their method in the main paper. We include this comparison in Fig. 2 so that viewers can qualitatively assess our shadow removal. In their work, they design an extensive dataset of shadowed/un-shadowed image pairs, in which shadows are cast by irregularly shaped objects outside the image (foreign shadows). While no such shadows were present in our training data, our model holds up well, while also achieving the overall delighting task including reflection removal.

References

1. Bertalmio, M., Bertozzi, A.L., Sapiro, G.: Navier-stokes, fluid dynamics, and image and video inpainting. In: Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001. vol. 1, pp. I–I. IEEE (2001)
2. Gross, R., Matthews, I., Cohn, J., Kanade, T., Baker, S.: Multi-pie. Image and vision computing **28**(5), 807–813 (2010)
3. Zhang, X., Barron, J.T., Tsai, Y.T., Pandey, R., Zhang, X., Ng, R., Jacobs, D.E.: Portrait shadow manipulation. ACM Transactions on Graphics (TOG) **39**(4), 78–1 (2020)

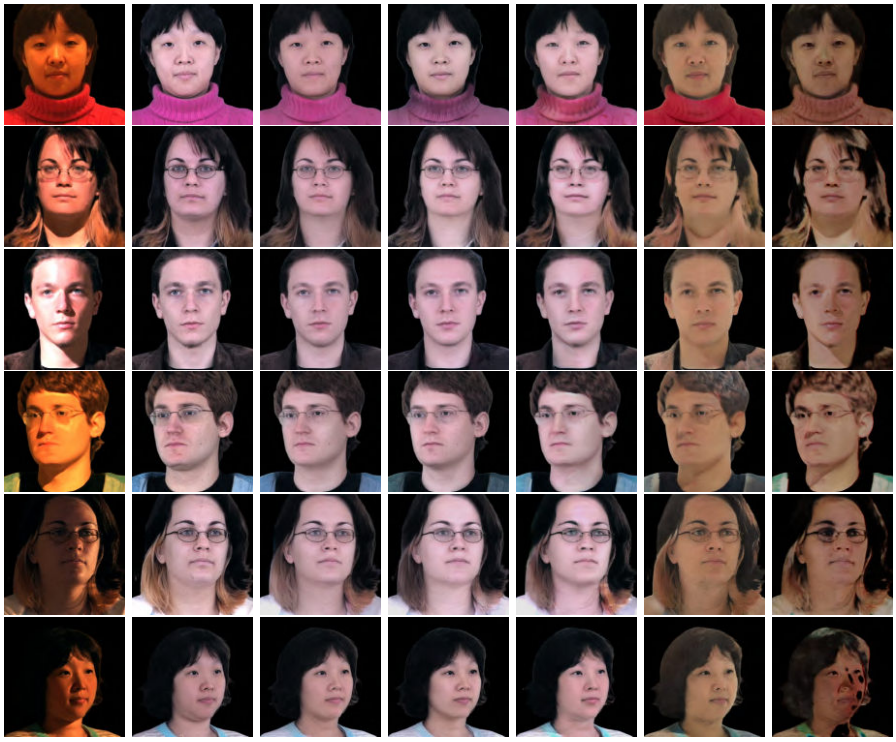


(a) Input (b) Zhang *et al.* [3] (c) Ours

Fig. 2: **Shadow Removal Comparison 1:** Comparison with the method of Zhang *et al.* [3] on irregularly shaped shadows. All images in (a) and (b) are taken from the author's website.

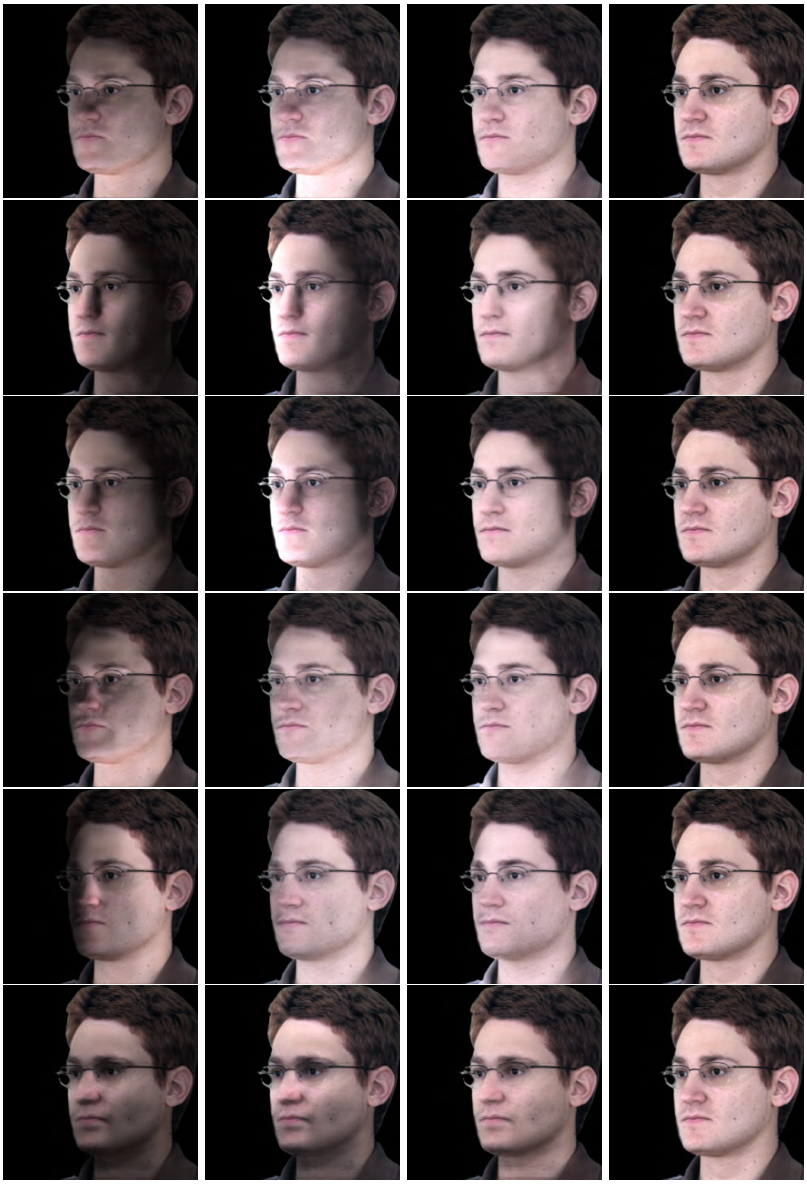
(a) Input (b) Ground truth (c) Ours (d) TR (retrained) (e) EMR (retrained) (f) TR

Fig. 3: We show more comparisons against other methods using upper body images from our testing dataset.



(a) Input (b) Ground truth (c) Ours (d) TR (retrained) (e) EMR (retrained) (f) TR (g) EMR

Fig. 4: We show more comparisons against other methods using face cropped images from our testing dataset.



(a) Input (b) w/o soft-shadow (c) w/ soft-shadow (d) Ground truth

Fig. 5: **Soft-shadow ablation:** Input images from **Alt. Lighting** dataset.