Neural Face Editing with Intrinsic Image Disentangling

SUPPLEMENTARY MATERIAL

Zhixin Shu[1] Ersin Yumer[2] Sunil Hadap[2] Kalyan Sunkavalli[2] Eli Shechtman [2] Dimitris Samaras[1,3]

[1]Stony Brook University [2]Adobe Research [3] CentraleSupélec, Université Paris-Saclay

[1]{zhshu,samaras}@cs.stonybrook.edu [2]{yumer,hadap,sunkaval,elishe}@adobe.com

1. Implementation: more details

In this section, we provide more details regarding the implementation of the rendering layers f_{shading} and $f_{\text{image-formation}}$ as described in the paper.

1.1. Shading Layer

The shading layer is rendered with a spherical harmonics illumination representation [6, 2, 7, 1].

The forward process is described by equations (3),(4), and (5) in the main paper. We now provide the backward process, i.e., the partial derivatives $\frac{\partial S_e^i}{\partial n_x}$, $\frac{\partial S_e^i}{\partial n_y}$,$\frac{\partial S_e^i}{\partial n_z}$ and $\frac{\partial S_e^i}{\partial L_j}$ as follows:

$$\frac{\partial S_e^i}{\partial n_x} = 2(c_1 L_9 n_x + c_1 L_5 n_y + c_1 L_8 n_z + c_2 L_4) \quad (1)$$

$$\frac{\partial S_e^i}{\partial n_y} = 2(c_1 L_5 n_x - c_1 L_9 n_y + c_1 L_6 n_z + c_2 L_2) \quad (2)$$

$$\frac{\partial S_e^i}{\partial n_z} = 2(c_1 L_8 n_x + c_1 L_6 n_y + c_3 L_7 n_z + c_2 L_3) \quad (3)$$

$$\frac{\partial S_e^i}{\partial L_1} = c_4 \quad (4)$$

$$\frac{\partial S_e^i}{\partial L_2} = 2c_2 n_y \quad (5)$$

$$\frac{\partial S_e^i}{\partial L_3} = 2c_2 n_z \quad (6)$$

$$\frac{\partial S_e^i}{\partial L_4} = 2c_2 n_x \quad (7)$$

$$\frac{\partial S_e^i}{\partial L_5} = 2c_1 n_x n_y \quad (8)$$

$$\frac{\partial S_e^i}{\partial L_6} = 2c_1 n_y n_z \quad (9)$$

$$\frac{\partial S_e^i}{\partial L_7} = c_3 n_z^2 - c_5 \quad (10)$$

$$\frac{\partial S_e^i}{\partial L_8} = 2c_1 n_x n_z \quad (11)$$

$$\frac{\partial S_e^i}{\partial L_9} = c_1 n_x^2 - c_1 n_y^2 \quad (12)$$

where

$$\begin{aligned} &c1 = 0.429043 \quad c2 = 0.511664 \\ &c3 = 0.743125 \quad c4 = 0.886227 \quad c5 = 0.247708 \end{aligned} \quad (13)$$

1.2. Image Formation Layer

The forward process of the image formation layer (for the foreground) is simply a per-element product (see equation (2) in the main paper), therefore the backward process (partial derivatives) is:

$$\frac{\partial I_{fg}}{\partial A_e^i} = S_e^i \quad (14)$$

and

$$\frac{\partial I_{fg}}{\partial S_e^i} = A_e^i \quad (15)$$

2. Quantitative Experiments

In order to evaluate the illumination estimation of our network, we utilize the Multi-PIE [4] dataset where controlled illumination is available[1]. We randomly sample 7,000 images with different identities and poses under 20 controlled light sources. We measure the variance of illumination coefficients L_i $(i = 1, ..., 9)$ within an illumination condition. The average variance of a 3D Morphable

[1]Controlled illumination in this context means that the different subjects have been illuminated under the same lighting conditions. Hence, the correspondences across subjects for the same lighting is known, but the actual lighting setup, or conditions are not known.

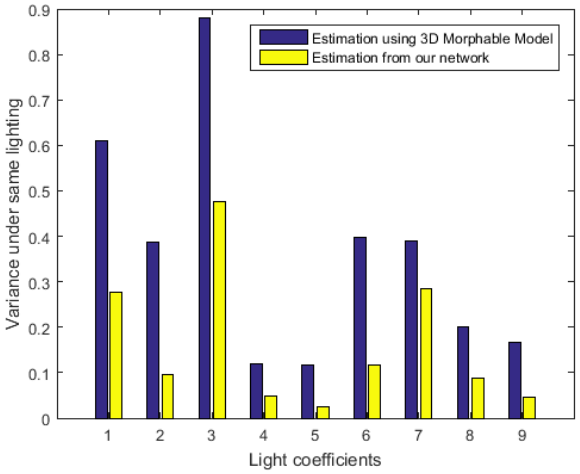

Figure 1. Stability of illumination estimation using our network. The illumination estimation from our network in the form of spherical harmonics lighting coefficients is more stable, i.e., exhibits less variance while measuring illumination of face images under identical lighting in Multi-PIE [4], than the estimation from a least squares morphable model.

Model with least square estimation is 0.36 while the average variance of our network is 0.16. In Fig. 1, we show the average (among 20 lighting conditions and RGB) variance of each lighting coefficient. Note that our model is only trained with CelebA dataset while the images from the Multi-PIE datasets are only used for quantitative evaluation.

We evaluate the quality of our normal reconstruction vs. a direct 3D Morphable Model (3DMM) fit. We create input images for five different individuals (two women, three men) using light stage data captured by Weyrich et al. [9]. We fit a 3DMM to these images and compute normals. We also pass these image as inputs to our disentangling network to get a normal reconstruction. In Fig. 2, we compare the ground truth normals to our estimates and the 3DMM normals. Even though our network was trained on normals from morphable model fits, the additional reconstruction losses enable it to expand beyond this subspace. This is especially apparent for the reconstruction of the two women, whose faces look man-like in the 3DMM fits.

3. Additional Results

We present more face editing results, as well as relighting results using our proposed method described in section 5.3 of the main paper.

Eye-glasses. Figure 4 shows results of *wear eye-glasses*. For this experiment, we sample 2000 images from the CelebA [5] of faces *wearing eye-glasses* as $\{\mathbf{x}_p\}$ and 2000 images of faces *not wearing eye-glasses* as $\{\mathbf{x}_n\}$. We compute the edits on the A_i manifold only, with $\lambda = 0.02$.

Since eye-glasses only affect the reflectance in the image, no geometry or warping is associate with this editing, therefore the natural choice of manifold-to-be-edited is Z_{A_i}. We show in figure 5 a comparison of *adding eye-glasses* via different manifolds (using same the $\lambda = 0.02$). We notice that (1) editing Z_{N_i} (Fig. 5-(c)) almost has no effect on the image; (2) editing Z_{UV} slightly *aged* the face (Fig. 5-(d)), mainly since senior people are more likely to wear eye-glasses; (3) manipulating all the manifolds leads to changes in geometry and appearance (beards, nostrils and shape of noses in Fig. 5-(e)); (4) editing through Z_{A_i} generates faithful results of *wearing eye-glasses* and has little effect on other attributes of the face.

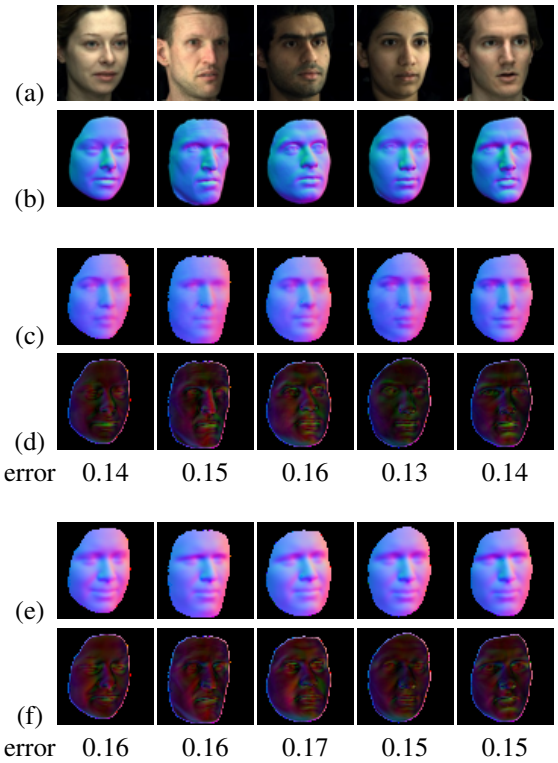

Figure 2. Comparison of normal reconstruction to ground truth. Input images (a), and ground truth normals (b). Our reconstruction and error (c,d). 3DMM fit and error (e,f)

Beards. In figures 6, 7, 8, we present additional results of *grow beard*. For this experiment, we sample 2000 images from the CelebA [5] of *male* faces *with beard* as $\{\mathbf{x}_p\}$ and 2000 images of *male* faces *without beard* as $\{\mathbf{x}_n\}$. We compute the traversal on the A_i manifold only, with $\lambda = 0.03$, 0.02, and 0.01 respectively.

Aging. Figures 9, 10, show additional results of *aging*. We compute the traversal on manifolds Z_{A_i}, Z_{N_i}, and Z_{UV}, with $\lambda = 0.05$, 0.03, and 0.02 respectively.

Smiling. Figures 11, 12 show additional results of *smiling* as described in Section 5.2 of the main paper. The data is as described in the main paper. We compute the traversal on manifolds Z_{N_i}, and Z_{UV}, with $\lambda = 0.07$, 0.05, and 0.03

(a) target (b) source (c) ours (d) SH (e) EPF

Figure 3. Comparison of re-lighting with spherical harmonics-based radiance maps (SH) [8] and edge-preserving filters (EPF) [3].

respectively.

Relighting. We present additional relighting results in Fig. 13. In addition, in Fig. 3, we provide comparisons to two previous techniques for re-lighting [8, 3]. Our results apply to the full face, capture the target lighting, and have fewer artifacts. Our general face editing technique is able to produce results that are qualitatively similar, or even better than other techniques specifically designed for this particular task.

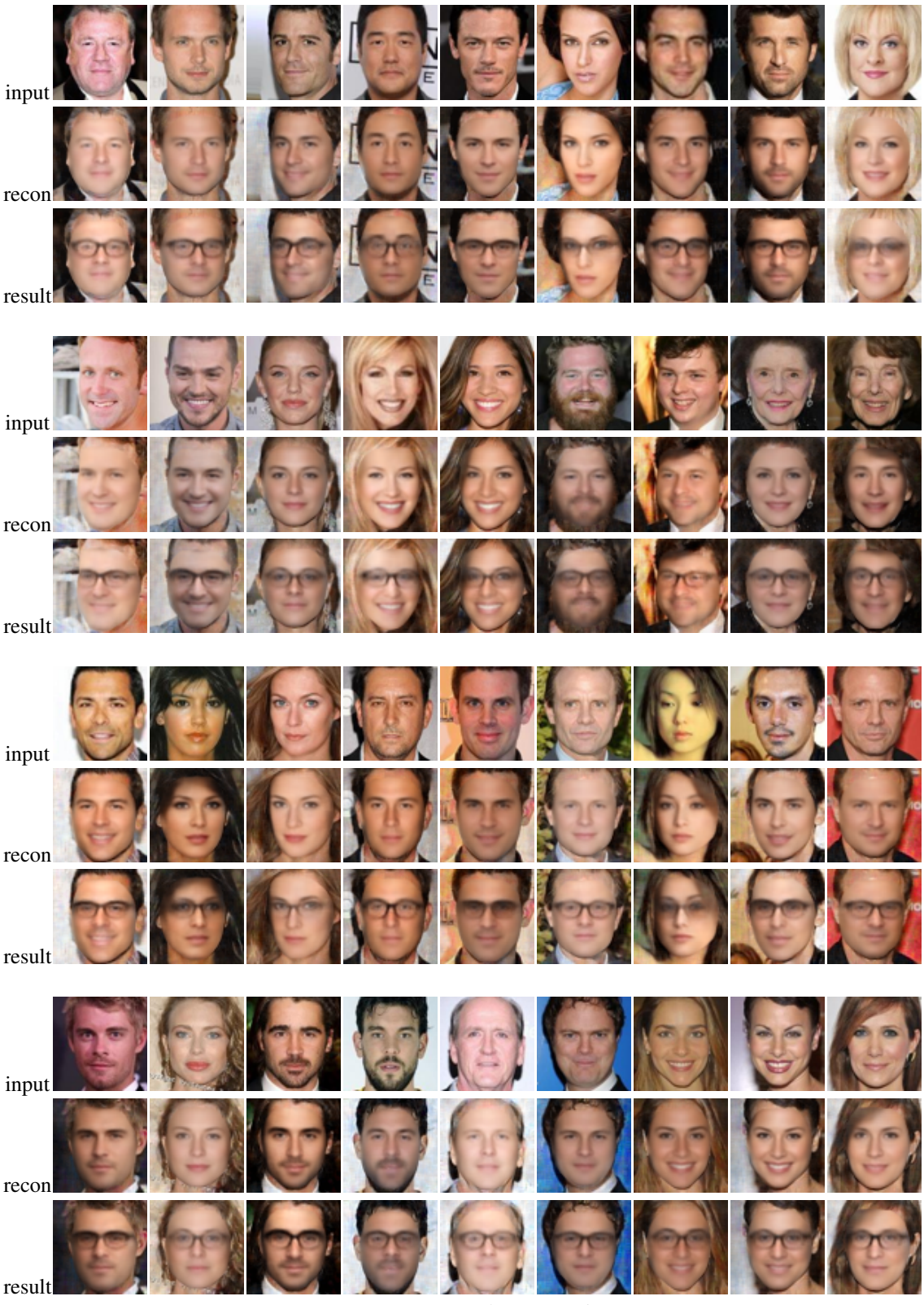

Figure 4. Adding eye-glasses to the input faces using our proposed method. Edits are computed on the Z_{A_i} manifold only and $\lambda = 0.02$.

(a) recon (b) Z_{A_i} (c) Z_{N_i} (d) Z_{UV} (e) all Zs

Figure 5. Comparison of adding eye-glasses via editing different manifold(s). Editing through Z_{A_i} generates faithful results of wearing eye-glasses (b). Almost no effects show up when manipulating Z_{N_i} (c). Editing through Z_{UV} slightly *aged* the face (d). Editing all three manifolds (e) not only adds eye-glasses, but also changes shape and appearance of the face. $\lambda = 0.02$ for all results in this figure.

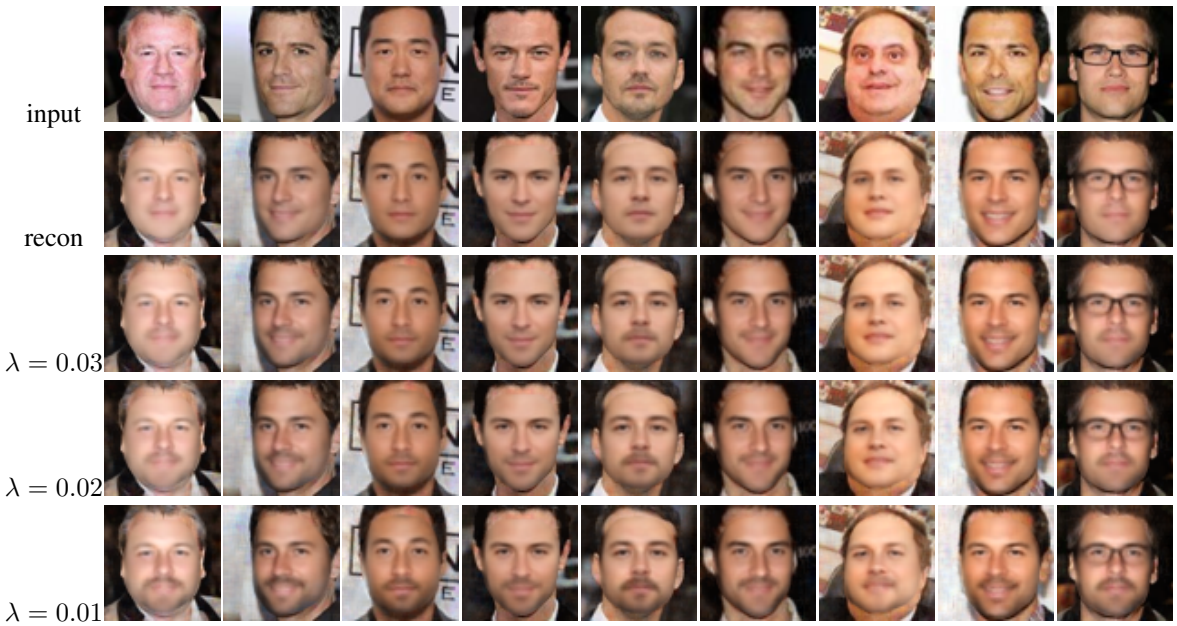

Figure 6. *Growing beard* onto the input faces. Edits/traversals are computed on the manifold Z_{A_i} only with $\lambda = 0.03$, 0.02, and 0.01 respectively.

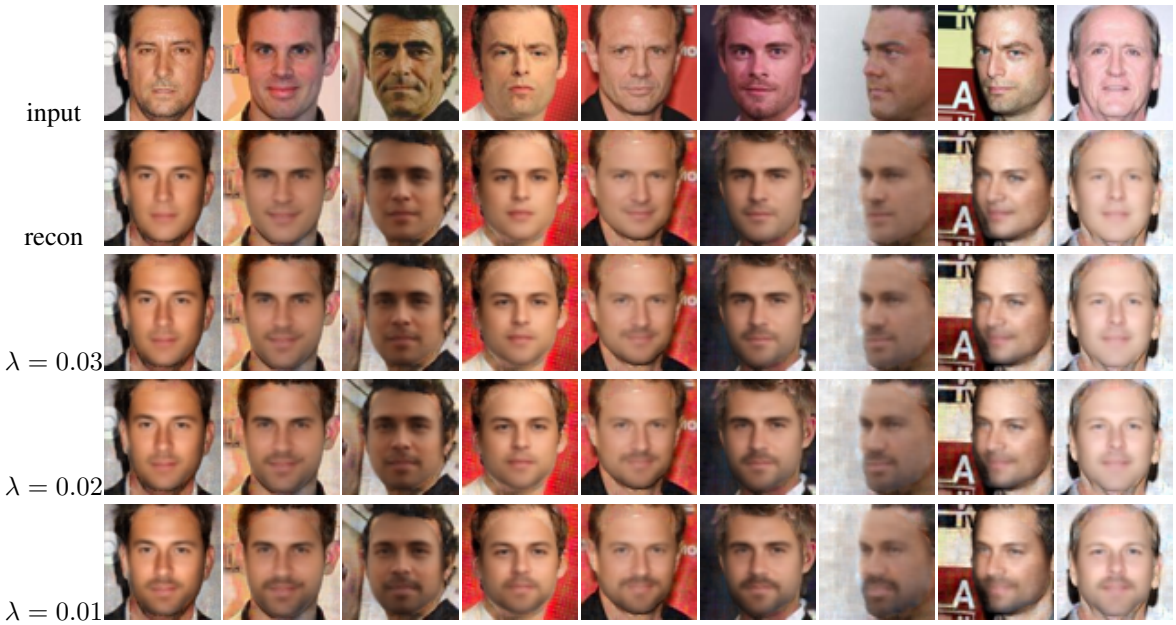

Figure 7. *Growing beard* onto the input faces. Edits/traversals are computed on the manifold Z_{A_i} only with $\lambda = 0.03$, 0.02, and 0.01 respectively.

Figure 8. *Growing beard* onto the input faces. Edits/traversals are computed on the manifold Z_{A_i} only with $\lambda = 0.03$, 0.02, and 0.01 respectively.

Figure 9. *Aging*. Edits/traversals are computeg on manifolds Z_{A_i}, Z_{N_i}, and Z_{UV}, with $\lambda = 0.05, 0.03$, and 0.02 respectively.

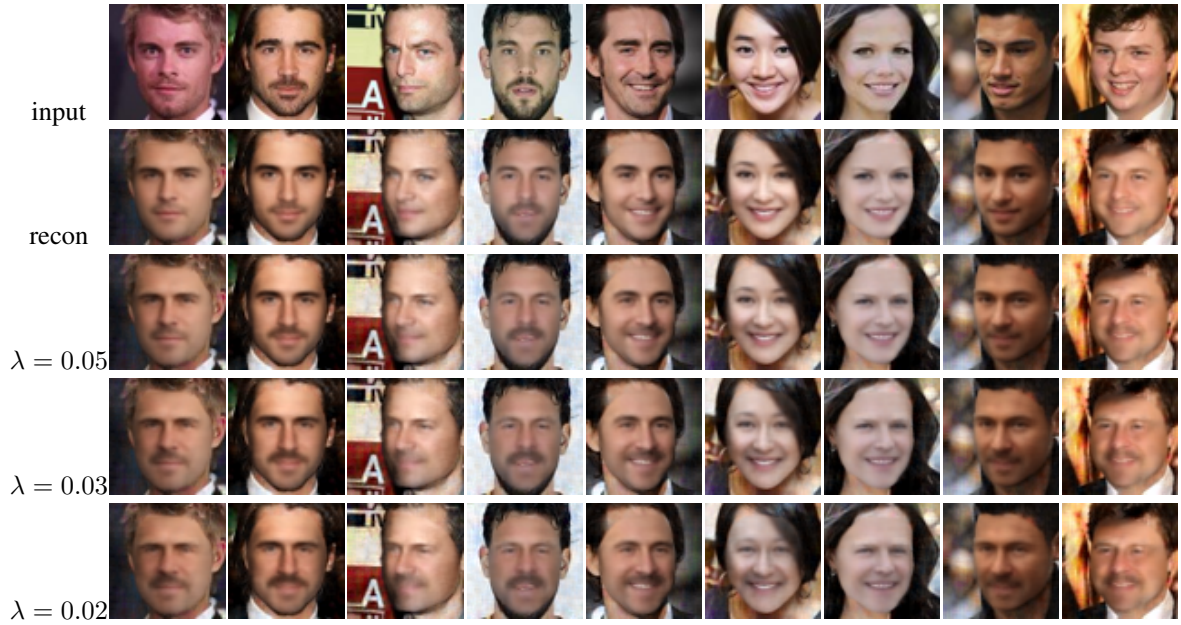

Figure 10. *Aging*. Edits/traversals are computed on manifolds Z_{A_i}, Z_{N_i}, and Z_{UV}, with $\lambda = 0.05$, 0.03, and 0.02 respectively.

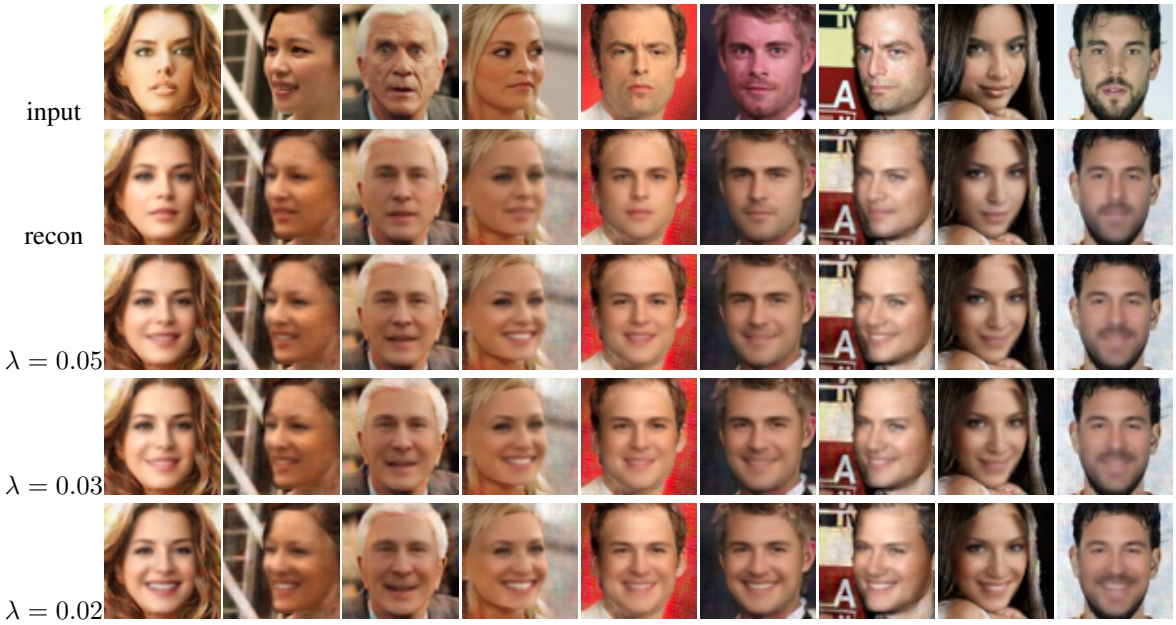

Figure 11. *Smiling*. Edits/traversals are computed on the manifolds Z_{N_i} and Z_{UV}, with $\lambda = 0.07, 0.05$, and 0.03 respectively.

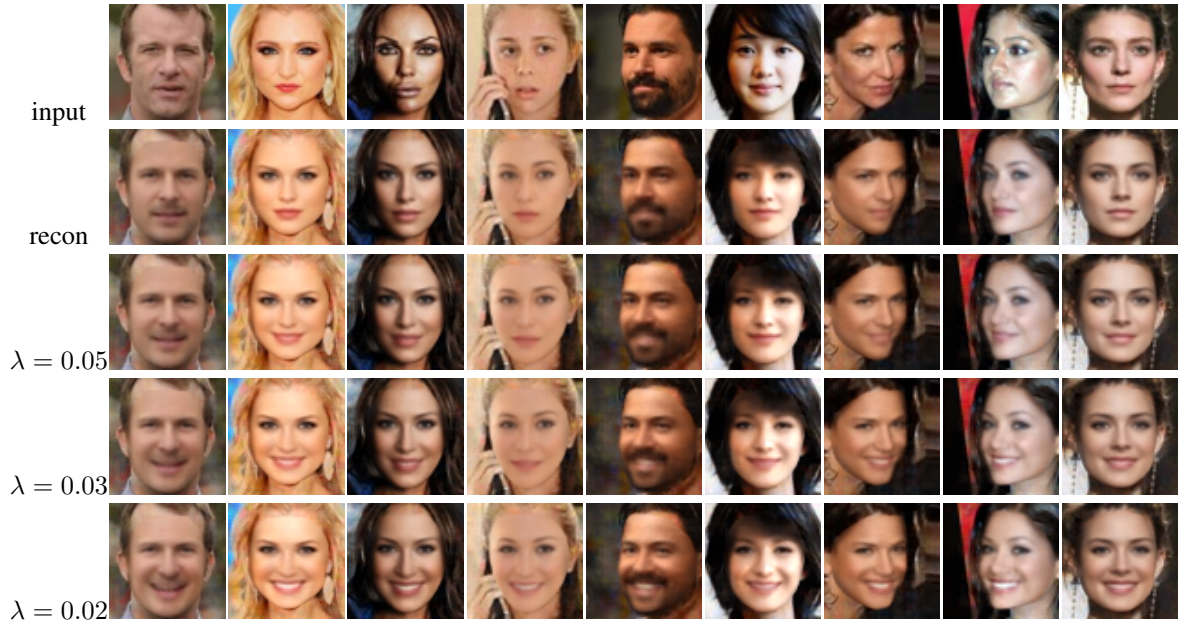

Figure 12. *Smiling*. Edits/traversals are computed on manifolds Z_{N_i} and Z_{UV}, with $\lambda = 0.07$, 0.05, and 0.03 respectively.

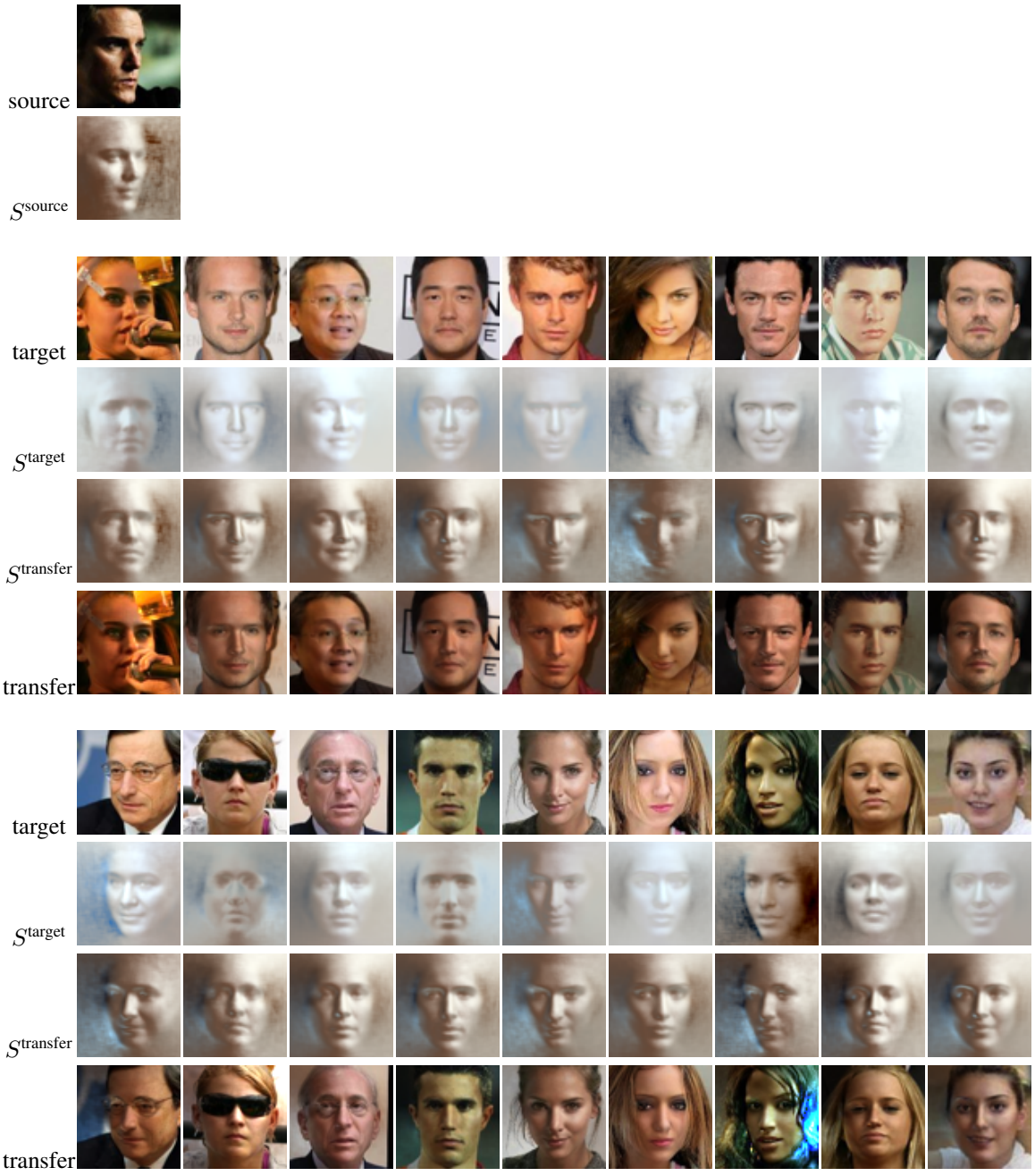

Figure 13. *Relighting*. Transfer of the lighting condition from the source image to target images using the method described in Section 5.3 of the main paper.

References

[1] J. T. Barron and J. Malik. Shape, illumination, and reflectance from shading. *IEEE transactions on pattern analysis and machine intelligence*, 37(8):1670–1687, 2015.

[2] R. Basri and D. W. Jacobs. Lambertian reflectance and linear subspaces. *IEEE transactions on pattern analysis and machine intelligence*, 25(2):218–233, 2003.

[3] X. Chen, M. Chen, X. Jin, and Q. Zhao. Face illumination transfer through edge-preserving filters. In *CVPR*, pages 281–287. IEEE, 2011.

[4] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker. Multi-pie. *Image and Vision Computing*, 28(5):807–813, 2010.

[5] Z. Liu, P. Luo, X. Wang, and X. Tang. Deep learning face attributes in the wild. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 3730–3738, 2015.

[6] R. Ramamoorthi and P. Hanrahan. On the relationship between radiance and irradiance: determining the illumination from images of a convex lambertian object. *JOSA A*, 18(10):2448–2459, 2001.

[7] Y. Wang, L. Zhang, Z. Liu, G. Hua, Z. Wen, Z. Zhang, and D. Samaras. Face relighting from a single image under arbitrary unknown lighting conditions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(11):1968–1984, 2009.

[8] Z. Wen, Z. Liu, and T. S. Huang. Face relighting with radiance environment maps. In *CVPR*, volume 2, pages II–158. IEEE, 2003.

[9] T. Weyrich, W. Matusik, H. Pfister, B. Bickel, C. Donner, C. Tu, J. McAndless, J. Lee, A. Ngan, H. W. Jensen, and M. Gross. Analysis of human faces using a measurement-based skin reflectance model. *ACM Trans. Graph.*, 25(3):1013–1024, July 2006.