IMAGE PROCESSING ON LINE

Published in Image Processing On Line on 2022–10–12.
Submitted on 2022–09–28, accepted on 2022–09–29.
ISSN 2105–1232 
This article is available online with supplementary materials, software, datasets and online demo at
https://doi.org/10.5201/ipol.2022.429

# Experiments on Deep Single-Image Portrait Relighting

Quentin Bammey

Université Paris-Saclay, ENS Paris-Saclay, Centre Borelli, Gif-sur-Yvette, France
quentin.bammey@ens-paris-saclay.fr

*Communicated by* Jean-Michel Morel *Demo edited by* Quentin Bammey

**Abstract**

This work describes deep single-image portrait relighting, a method to change the lighting of an image. The method is based on an hourglass convolutional neural network, which encodes the image and its estimated original lighting features, then replaces the lighting features with the target light. We highlight the overall good results produced by this method, as well as its limitations and the lighting artifacts that sometimes appear on relighted images.

**Source Code**

The source code and documentation for this algorithm are available from the web page of this article[1]. Usage instructions are included in the README.txt file of the archive. The authors' original implementation of the method is available here[2].
This is an MLBriefs article, the source code has not been reviewed!

**Keywords:** image forensics; forgery detection; convolutional neural network

# 1 Introduction

Image relighting, in particular portrait relighting, aims at simulating a modification of the light sources on an image. Originally tackled by optimization-based inverse rendering [2, 1, 3, 9], the emergence of deep learning saw this field get more interest in the past years [7, 8, 4, 3, 6, 12, 10], in addition to parallel works on video portrait relighting [11].

In this work, we briefly describe and analyze deep single-image portrait relighting (DPR), a method by Zhou et al. [12] to relight images. It is the first deep learning-based method that is able to process relatively high-resolution images (up to $1024 \times 1024$ pixels). Using the implementation from the authors, we experiment with the method to analyze its results and limitations.

[1] https://doi.org/10.5201/ipol.2022.429
[2] https://zhhoper.github.io/dpr.html

# 2 Method

The method consists of a convolutional neural network, that estimates the original lighting condition and changes it in the image. The structure of the network is presented in detail in Figures 1 for the global network structure and 2 for the basic blocks.

**Encoding.** The network takes as input a grayscale image, or the luminance channel of a color image, resized to $512 \times 512$ pixels with bilinear interpolation. The encoder progressively extracts features and downsamples the input to $32 \times 32$. A total of 128 image features and 27 lighting features are extracted.

**Light estimation and relighting.** The 27 lighting features are then spatially averaged, and two fully-connected layers yield the 9-parameters lighting estimation of the image. These 9 parameters correspond to second-order spherical harmonic coordinates. To relight the image, the 9-parameters target lighting is processed into 27 lighting features with 2 fully-connected layers, then spatially-repeated into a $32 \times 32$ uniform image.

**Decoding.** The original 27 lighting features are replaced by the 27 ones from the target. Then, the decoder progressively upsamples and creates the relighted image. The spatial features are reconstructed using skip connections from the encoder in addition to the decoding.

**Output.** The neural network yields a $512 \times 512$ one-channel image. The image is resized into the original size with bilinear interpolation. In case of a color input, the output luminance channel is inserted back into the image.

$1024 \times 1024$ **model** The original model works on $512 \times 512$ images. Zhou et al. [12] provide another model to work on $1024 \times 1024$ inputs. It is very similar in structure to the original model, with the addition of a 2-downsampling after the first convolution and a 2-upsampling after the last skip-connection (before the last convolution block in Figure 2). The model thus does not really work at a true $1024 \times 1024$ resolution, it only manages the super-resolution from $512 \times 512$ to $1024 \times 1024$ in the last block, without using the base resolution of the original image to do so.

**Training.** The network is trained on images from the CelebA [5] dataset with synthetic lighting. Example images from this dataset can be seen in Figure 3.

The images are synthetically relit using a relighting pipeline, that estimates the original reflectance as well as a face mesh, fits it with detected facial features and generates the final image with modified lighting conditions. Details from the training can be found in the original article [12]. In the experiments, we use the weights provided by the authors[3].

# 3 Experiments

In our experiments, we test the method against images not coming from the training dataset, so as to test the generality of the method. Indeed, the CelebA [5] dataset, used to train the network, features portraits of celebrities, a specific type of portrait made in optimal settings. For this purpose, we use various images found on flickr[4]. Specific credits for each image can be found in the last section.

[3] https://zhhoper.github.io/dpr.html
[4] flickr.com/

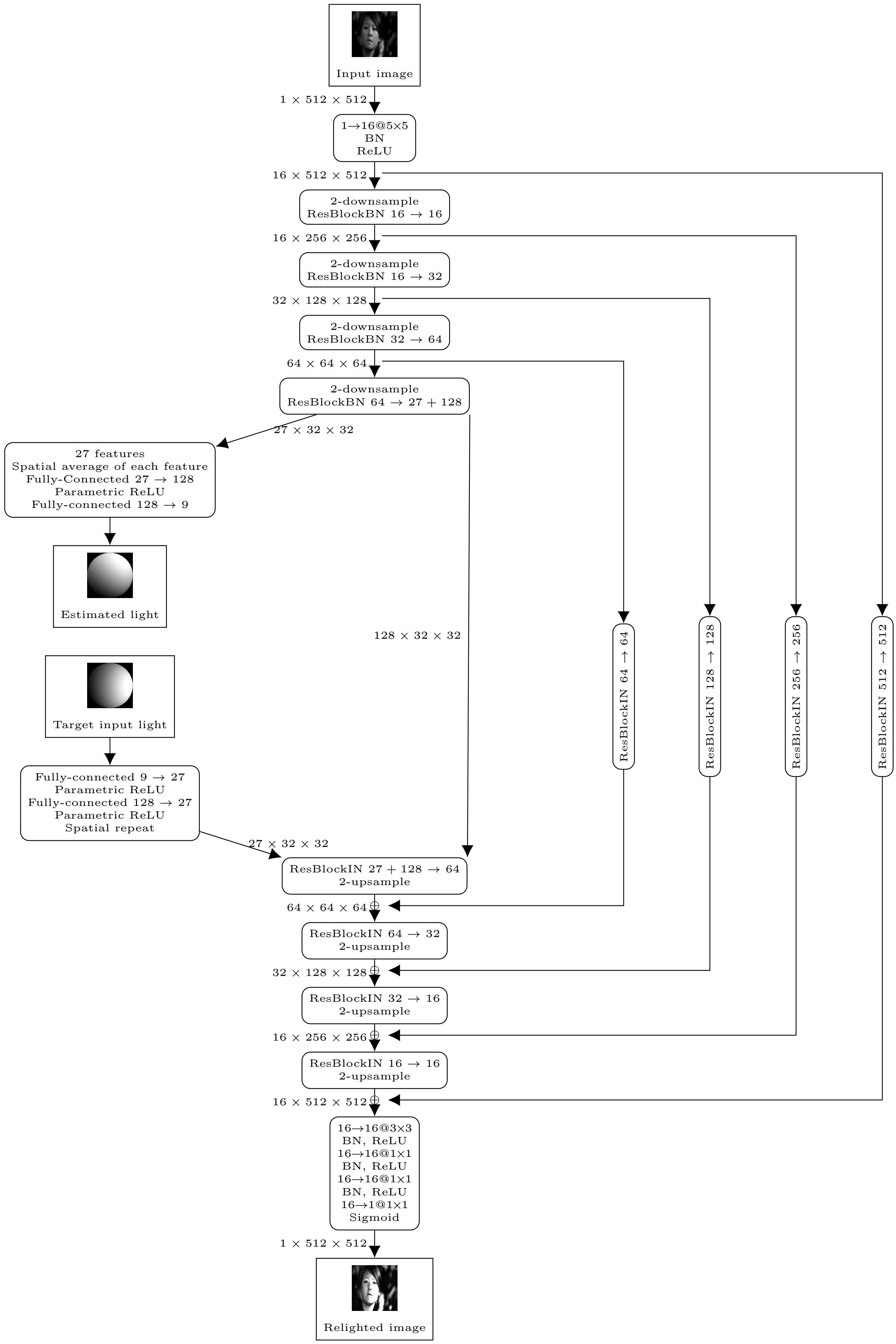



Figure 1: Network structure. All convolutions are without bias. See Figure 2 for the structure of the ResBlockIN and ResBlockBN blocks.

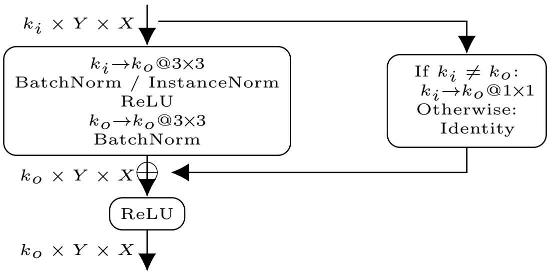



Figure 2: Structure of ResBlockIN and ResBlockBN, using batch normalization (ResBlockBN) or instance normalization (ResBlockIN).


Figure 3: Example images from the CelebA dataset. Source: http://mmlab.ie.cuhk.edu.hk/projects/CelebA.html

## 3.1 Light Estimation

The method estimates the lighting features from the image, and replaces them with target features. As a consequence, a correct estimation of the light features seems crucial for a successful relighting. We thus start by examining the ability of the method to estimate the original lighting of a portrait. As can be seen in Table 1, in many cases, the method is able to provide a good estimation of the lighting, despite a few errors, seen in Table 2. One particular limitation is that the second-order spherical harmonics used as representation cannot represent more complex shadings such as shadows from a hat or external elements.

| Image | Estimated light | Image | Estimated light | Image | Estimated light |
|---|---|---|---|---|---|

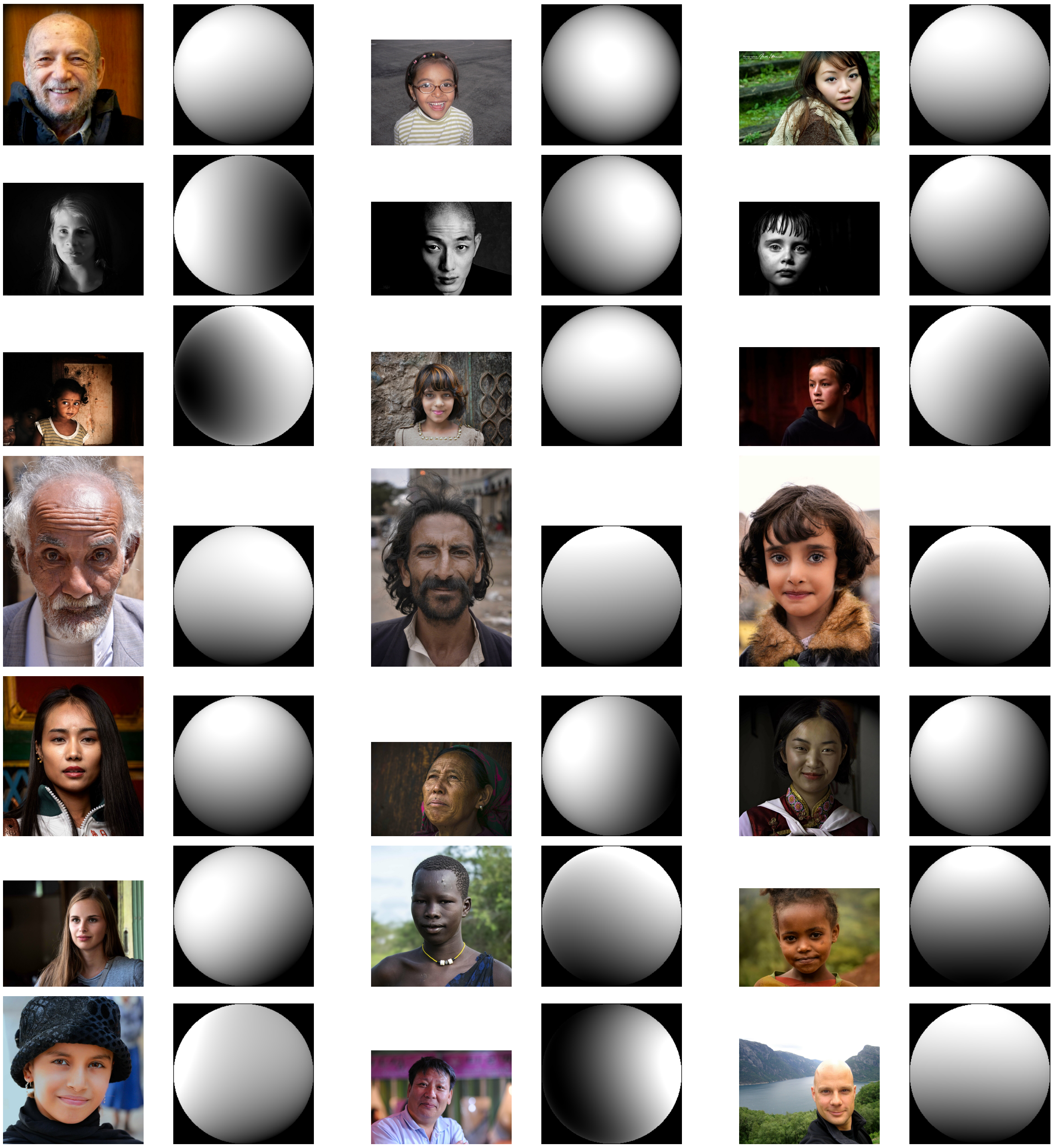


Table 1: In most cases, the method is able to provide good estimations of the image lighting, even when the face is occluded or if the background is not neutral.

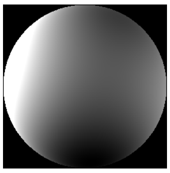
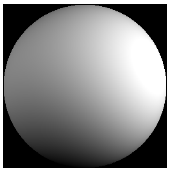
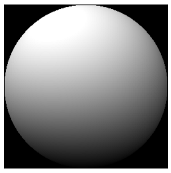
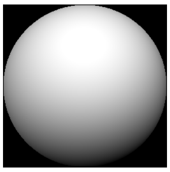
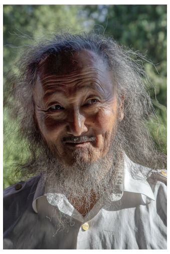
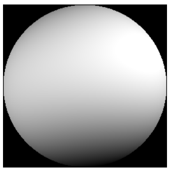
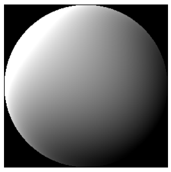
| Image | Estimated light | Image | Estimated light | Image | Estimated light |
|---|---|---|---|---|---|
|  |  |  |  |  |  |
|  |  |  |  |  |  |

Table 2: Examples of failures to estimate the original lighting. In the first row, this is due to the incapacity of the chosen representation to represent complex lightings such as those from a hat or external elements.

## 3.2 Relighting

We now study the ability of the method to relight images. As seen in Figure 4, the method can correctly relight the image in many cases. However, it is not without failures.

As stated earlier, the use of second-order spherical harmonics limits the representativeness of the lighting estimation. It also makes it difficult to correctly recast external shadows, such as those due to hair or hats. Such issues can be seen in Figure 5.

The method is not an interpolation method either; as a consequence, and as seen in Figure 6, it cannot reconstruct details that were missing from the original image due to shadow or light saturation, but should theoretically be visible from the target lighting.

Another issue is caused by the resolution of the method. Since the image is resized to $512 \times 512$ pixels for processing, and rescaled to the original size afterwards, many details are lost on big images. As seen in Figures 7 and 8, this even creates artifacts in some images. Zhou et al. [12] propose another network which takes and outputs $1024 \times 1024$ images, encapsulating the original network and managing the resampling. Nevertheless, large images still need to be resampled. Although this is the first method to provide relighting on more than thumbnail-sized images, we note that a decent resampling could easily improve performances on large images. Ideally, though, one could attempt to restore the resolution perfectly using features from the original image.

As can be seen in Figure 9, the method sometimes fails to properly generate the desired lighting orientation. This may be due to a failure to correctly capture the initial lighting features.

More importantly, the method tends to generate many lighting artifacts, seen in Figures 10, 11 and 12. The background is sometimes modified with incoherent bands of highlight or shadow. This appears mostly, but not only, on non-uniform backgrounds. This may be due to the training dataset featuring mainly images with no or little background. As the normal map is diffused from the face to the background, the new background can be incorrectly modified. Artifacts also appear directly on the face, with highlights or shadows appearing – or not disappearing correctly – in target lighting that should not produce them.

| Original image | Light estimation | Target light | Relight |
|---|---|---|---|

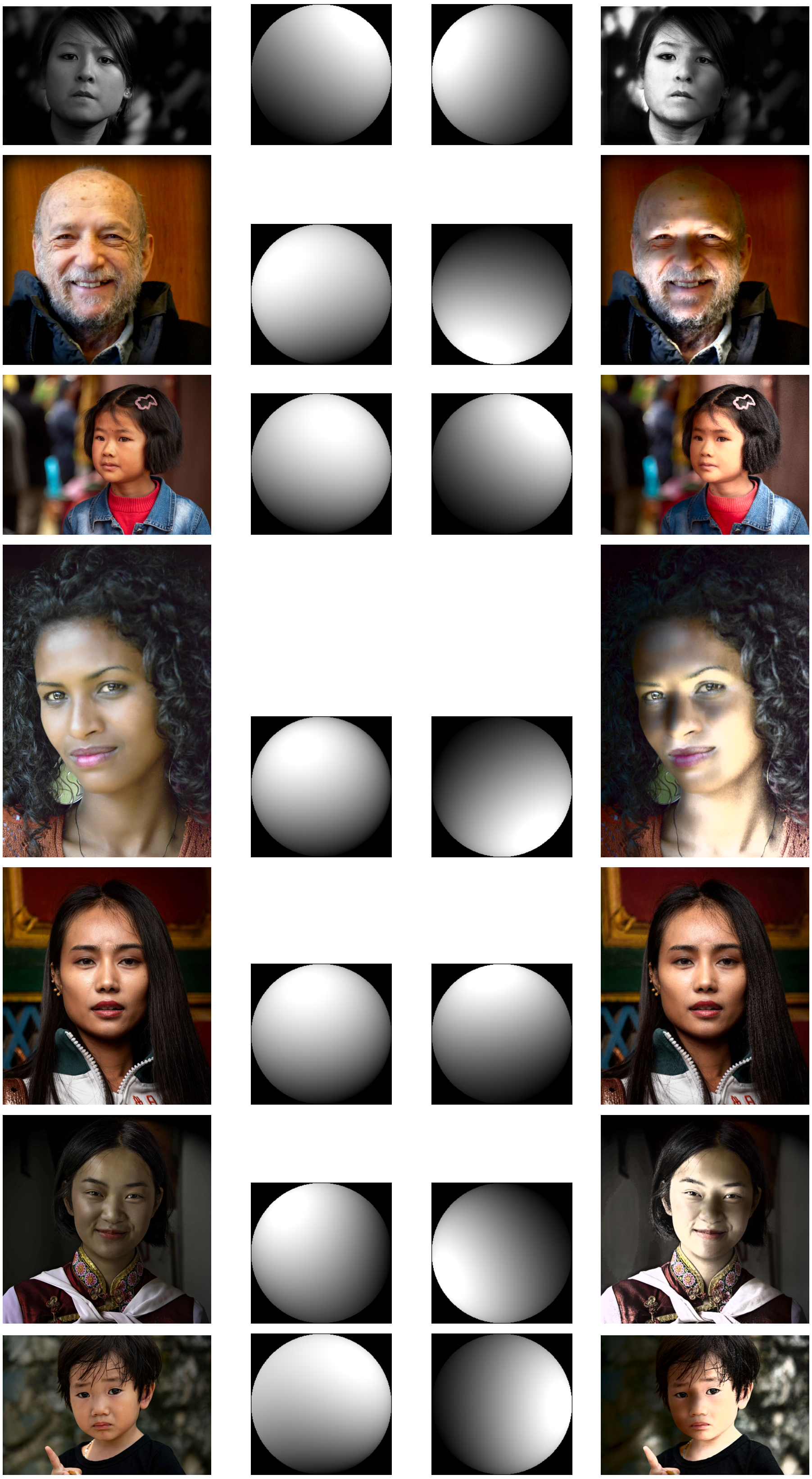


Figure 4: Examples of successful relighting

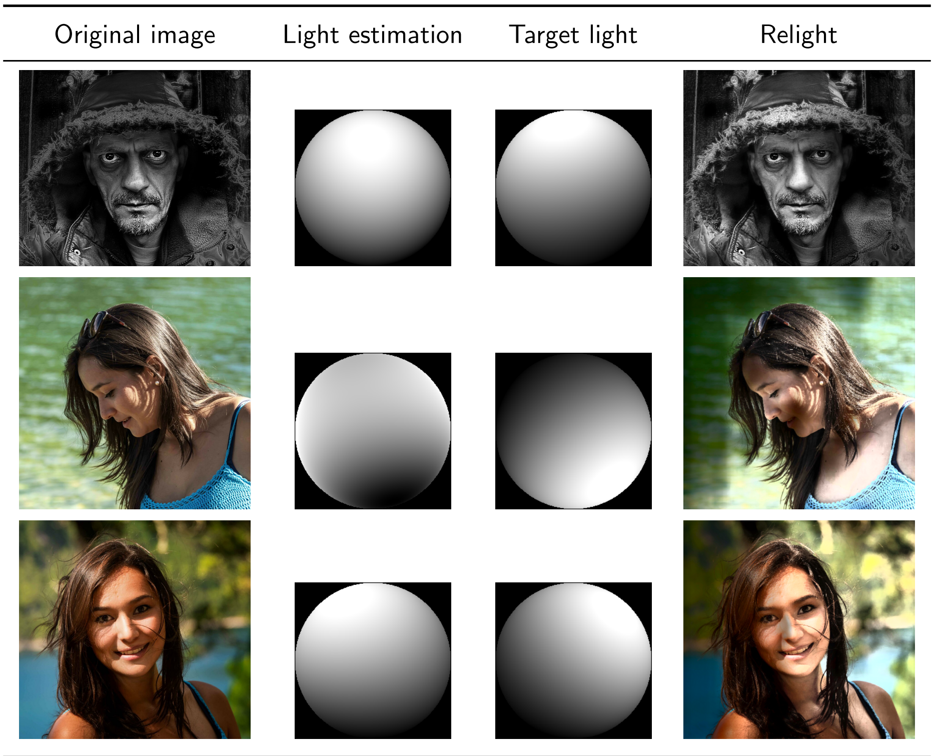



Figure 5: In those examples, the second-order spherical harmonics representation prevents the method from recasting shadows due to external elements such as hair or hats.

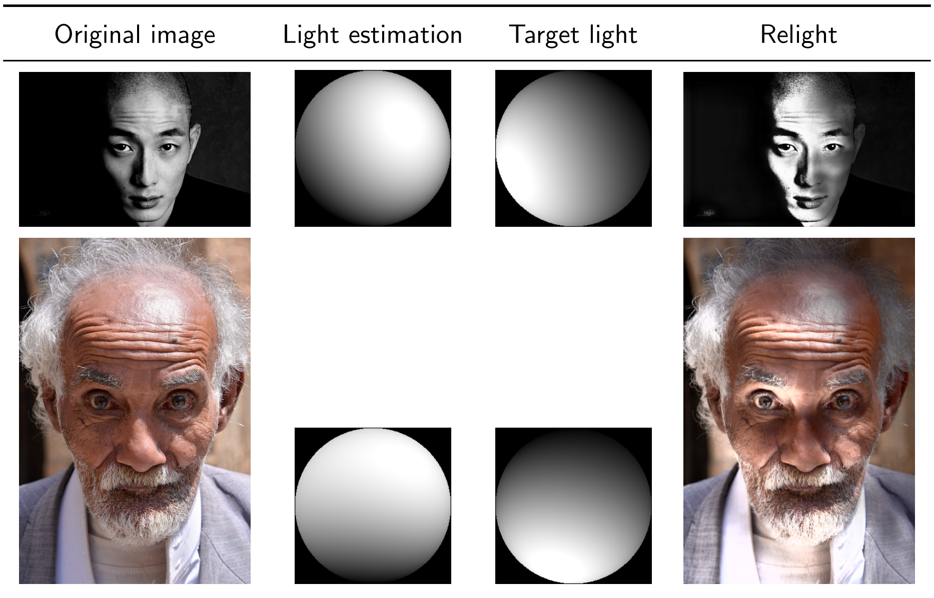



Figure 6: When the original image is saturated due to strong shadows or highlights, the saturated regions cannot be reconstructed even when they should be visible under another lighting.

Original image Light estimation

Relight ($512 \times 512$ network).

Relight ($1024 \times 1024$ network).

Figure 7: Working on $512 \times 512$ images, details are lost and artifacts created when large-scale images need to be resized by bilinear resampling. The $1024\times1024$ network only partly alleviates this issue. Full resolution versions of the images in the middle and bottom rows can be found in the demo's archive, at https://ipolcore.ipol.im/demo/clientApp/demo.html?id=429&archive=516158 and https://ipolcore.ipol.im/demo/clientApp/demo.html?id=429&archive=516159, respectively.

| Original image | Light estimation | Target light | Relight $512 \times 512$ | Relight $1024 \times 1024$ |
|---|---|---|---|---|

Figure 8: Possibly due to the working size, sharp noise appears throughout this example with the $512 \times 512$ model, but mostly disappears with the $1024 \times 1024$ network

| Original image | Light estimation | Target light | Relight |
|---|---|---|---|

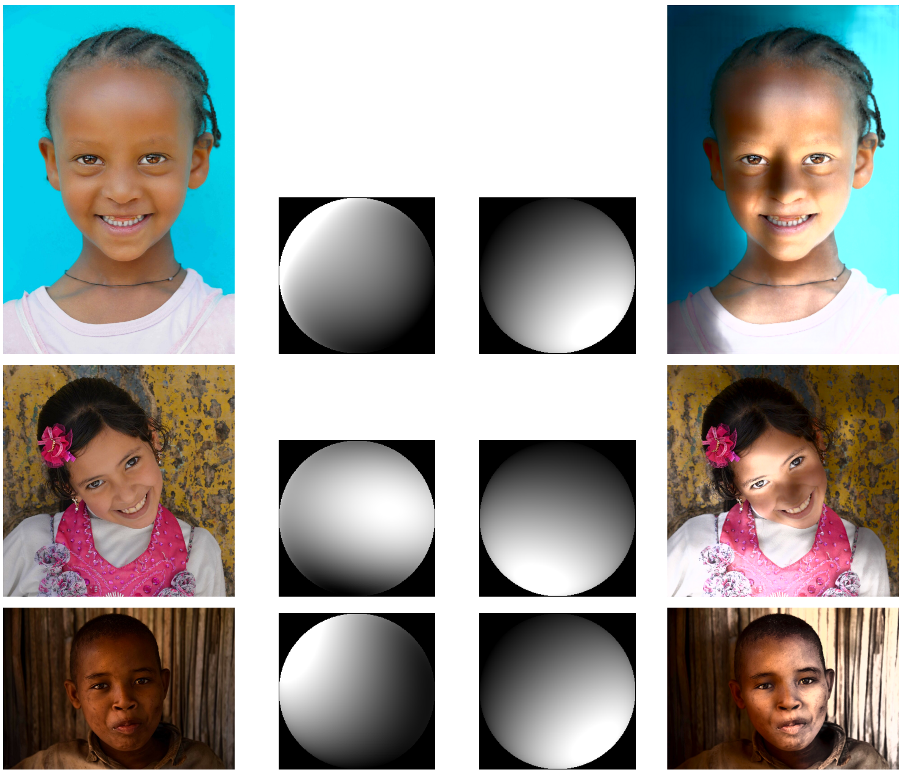


Figure 9: In those examples, the target light orientation is not perfectly generated.

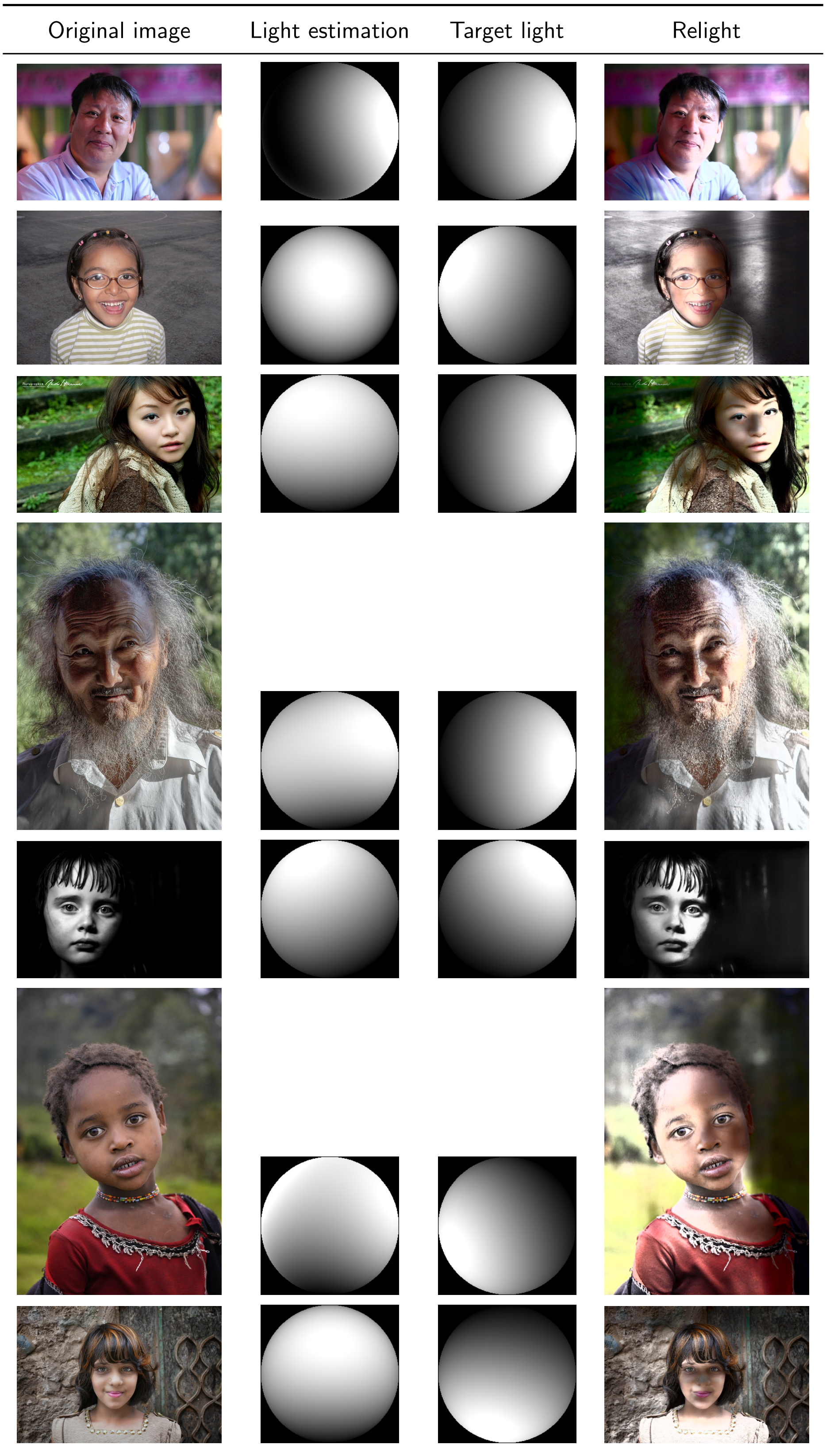



Figure 10: The method sometimes creates lighting artifacts. They especially appear in non-neutral backgrounds, where lines or bands of light or shadows are created when none existed in the original image.

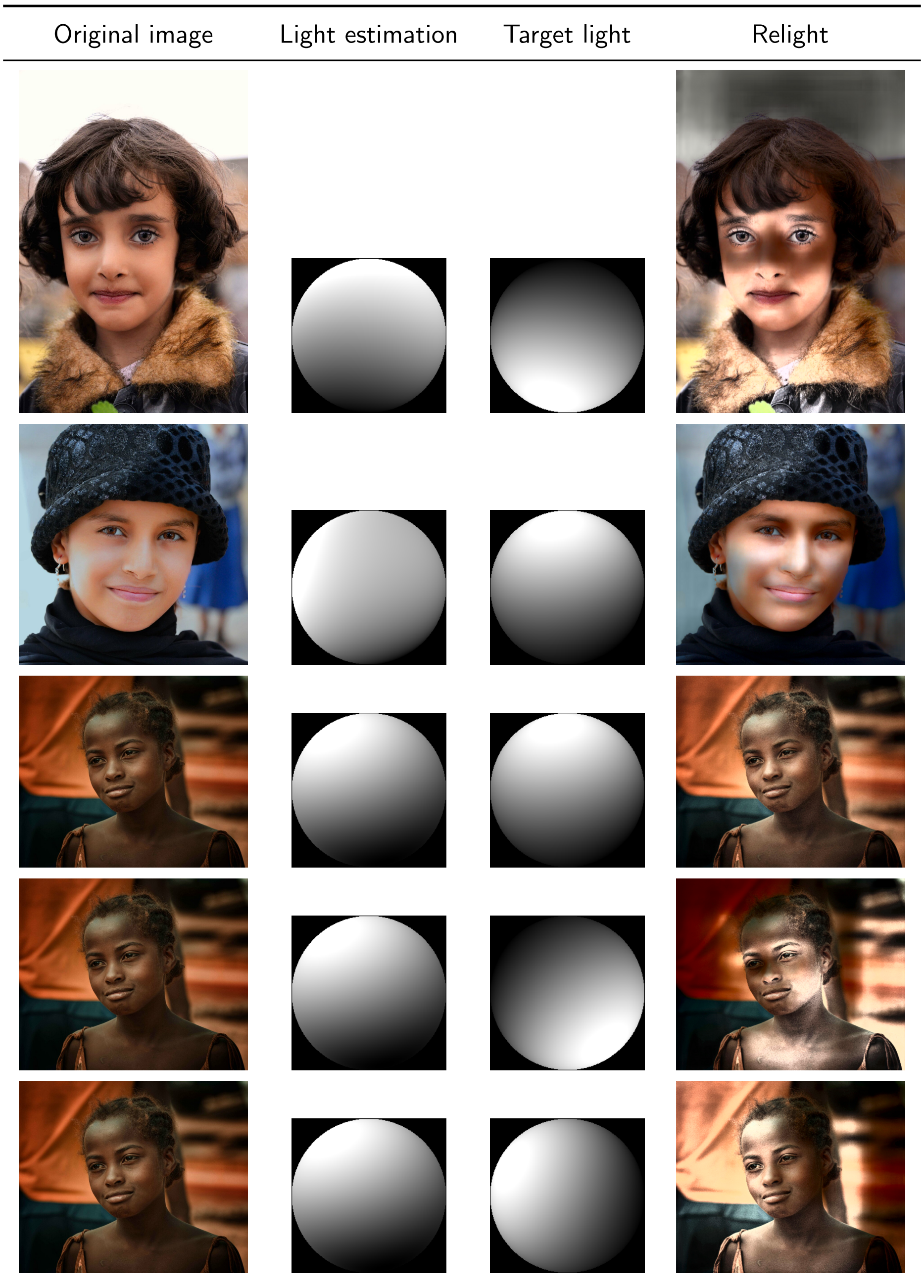



Figure 11: More lighting artifacts created by the method, seemingly not due to background-related causes. In the bottom three rows, lighting artifacts appear at the forehead, where the method is unable to remove the highlights when needed, despite this region not being saturated.


Figure 12: In this example, the method creates strange highlights on the eyebrows.

# 4 Conclusion

The presented method relights a portrait by modifying the lighting features extracted by an encoder-decoder network. It works well overall, but could be improved by a better resampling for bigger images, especially considering it would be possible to make use of the original image at full resolution for this purpose. Its choice of second-order spherical harmonics, while reasonable in most cases, could be improved when more complex shadows need to be recast, such as shadows from hair, hats or external elements. The main issue for the method is the appearance of lighting artifacts in different scenarios, that give the impression of unnatural lighting.

# Image Credits

 Rod Waddington[5], CC BY SA 2.0[6].

 Nithi Anand[7], CC BY 2.0[8].

 Curo Ninja[9], CC BY NC 2.0[10].

 yooperann[11], CC BY NC ND 2.0[12].

[5]https://www.flickr.com/photos/rod_waddington/
[6]https://creativecommons.org/licenses/by-sa/2.0/
[7]https://www.flickr.com/photos/nithiclicks/
[8] https://creativecommons.org/licenses/by/2.0/
[9]https://www.flickr.com/photos/curoninja
[10] https://creativecommons.org/licenses/by-nc/2.0/
[11]https://www.flickr.com/photos/yooperann
[12]https://creativecommons.org/licenses/by-nc-nd/2.0/

 oneVillage Initiative[13], CC BY SA 2.0[6].

 AehoHikaruki[14], CC BY SA 2.0[6].

 https://www.flickr.com/photos/123_456[15], CC BY NC ND 2.0[12].

 frankyu[16], CC BY NC 2.0[10].

 Oskari Kettunen[17], CC BY 2.0[8].

 Steve Austin[18], CC BY ND 2.0[19].

 M. G. N. Marcel ON OF[20], CC BY NC 2.0[10].

 Luigi Morante[21], CC BY 2.0[8].

 nebojsa mladjenovic[22], CC BY NC ND 2.0[12].

 Mt Page[23], CC BY NC ND 2.0[12].

 Tatiana[24], CC BY SA 2.0[6].

 Sage Ross[25], CC BY SA 2.0[6].

## References

[1] O. Aldrian and W.A.P. Smith, *Inverse Rendering of Faces with a 3D Morphable Model*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 35 (2013), pp. 1080–1093. https://doi.org/10.1109/TPAMI.2012.206.

[2] V. Blanz and T. Vetter, *A morphable model for the synthesis of 3D faces*, in 26th Annual Conference on Computer Graphics and Interactive Techniques, 1999, pp. 187–194. https://doi.org/10.1145/311535.311556.

[3] B. Egger, S. Schönborn, A. Schneider, Ad. Kortylewski, A. Morel-Forster, C. Blumer, and T. Vetter, *Occlusion-aware 3D morphable models and an illumination prior for face image analysis*, International Journal of Computer Vision, 126 (2018), pp. 1269–1287. https://doi.org/10.1007/s11263-018-1064-8.

[13] https://www.flickr.com/photos/1village
[14] https://www.flickr.com/photos/aehohikaruki
[15] https://www.flickr.com/photos/123_456
[16] https://www.flickr.com/photos/frankyu/
[17] https://www.flickr.com/photos/aok/
[18] https://www.flickr.com/photos/bruvva/
[19] https://creativecommons.org/licenses/by-nd/2.0/
[20] https://www.flickr.com/photos/marcelxxl/
[21] https://www.flickr.com/photos/eudaimos/
[22] https://www.flickr.com/photos/mladjenovic_n/4601084173/
[23] https://www.flickr.com/photos/113318113@N04/
[24] https://www.flickr.com/photos/kulitat/
[25] https://www.flickr.com/photos/ragesoss/

[4] K. Genova, F. Cole, A. Maschinot, A. Sarna, D. Vlasic, and W.T. Freeman, *Unsupervised Training for 3D Morphable Model Regression*, in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2018. http://dx.doi.org/10.1109/CVPR.2018.00874.

[5] T. Karras, T. Aila, S. Laine, and J. Lehtinen, *Progressive growing of GANs for improved quality, stability, and variation*, 2017. https://doi.org/10.48550/arXiv.1710.10196.

[6] S. Sengupta, A. Kanazawa, C.D. Castillo, and D.W. Jacobs, *SfSNet: Learning Shape, Reflectance and Illuminance of Faces 'in the Wild'*, in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2018. https://doi.org/10.1109/CVPR.2018.00659.

[7] Z. Shu, E. Yumer, S. Hadap, K. Sunkavalli, E. Shechtman, and D. Samaras, *Neural face editing with intrinsic image disentangling*, in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 2017. https://doi.org/10.1109/CVPR.2017.578.

[8] A. Tewari, M. Zollhofer, H. Kim, P. Garrido, F. Bernard, P. Perez, and C. Theobalt, *MoFA: Model-Based Deep Convolutional Face Autoencoder for Unsupervised Monocular Reconstruction*, in IEEE International Conference on Computer Vision (ICCV) Workshops, Oct 2017. https://openaccess.thecvf.com/content_ICCV_2017/papers/Tewari_MoFA_Model-Based_Deep_ICCV.

[9] 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 (2009), pp. 1968–1984. https://doi.org/10.1109/TPAMI.2008.244.

[10] Z. Wang, X. Yu, M. Lu, Q. Wang, C. Qian, and F. Xu, *Single image portrait relighting via explicit multiple reflectance channel modeling*, ACM Transactions on Graphics, 39 (2020). https://doi.org/10.1145/3414685.3417824.

[11] L. Zhang, Q. Zhang, M. Wu, J. Yu, and L. Xu, *Neural video portrait relighting in real-time via consistency modeling*, in IEEE/CVF International Conference on Computer Vision (ICCV), October 2021, pp. 802–812. https://doi.org/10.1109/ICCV48922.2021.00084.

[12] H. Zhou, S. Hadap, K. Sunkavalli, and D.W. Jacobs, *Deep single-image portrait relighting*, in IEEE/CVF International Conference on Computer Vision (ICCV), October 2019. https://doi.org/10.1109/ICCV.2019.00729.