In-the-wild Facial Highlight Removal via Generative Adversarial Networks

Zhibo Wang[1,2], Ming Lu[3], Feng Xu[1,4], and Xun Cao[2]

[1] BNRist and School of Software, Tsinghua University, China
[2] School of Electronic Science and Engineering, Nanjing University, China
[3] Intel Labs, China
[4] Beijing Laboratory of Brain and Cognitive Intelligence, Beijing Municipal Education Commission

Abstract. Facial highlight removal techniques aim to remove the specular highlight from facial images, which could improve image quality and facilitate tasks, *e.g.*, face recognition and reconstruction. However, previous learning-based techniques often fail on the in-the-wild images, as their models are often trained on paired synthetic or laboratory images due to the requirement on paired training data (images with and without highlight). In contrast to these methods, we propose a highlight removal network, which is pre-trained on a synthetic dataset but finetuned on the unpaired in-the-wild images. To achieve this, we propose a highlight mask guidance training technique, which enables Generative Adversarial Networks (GANs) to utilize in-the-wild images in training a highlight removal network. We have an observation that although almost all in-the-wild images contain some highlights on some regions, small patches without highlight can still provide useful information to guide the highlight removal procedure. This motivates us to train a region-based discriminator to distinguish highlight and non-highlight for a facial image and use it to finetune the generator. From the experiments, our technique achieves high-quality results compared with the state-of-the-art highlight removal techniques, especially on the in-the-wild images.

Keywords: Generative adversarial networks, highlight removal.

1 Introduction

Highlights always lead to deviations from skin colors in facial images, which causes problems in many tasks including face recognition and reconstruction. However, capturing specular-free images in the wild is almost impossible, as these images require special equipments such as polarizing filters or *Light Stage* [3], which can only be aquired in a laboratory environment. Hence, portrait highlight removal techniques are necessary to facilitate consumer-level access to specular-free in-the-wild facial images. Traditional methods on highlight removal [31, 30] usually require strict assumptions on lighting conditions. However, facial images are normally captured in natural environments and thus the violations of the assumptions lead to obvious artifacts in the results. [17] develops a portrait highlight removal method by using some weak lighting assumptions and achieves state-of-the-art results by a complicated model of facial skins. However, it is time-consuming to solve the non-linear optimization of their facial skin model.

Recently, more efficient methods based on deep neural networks [32, 34] have been proposed to remove highlight from facial images. As they require paired images (images with and without highlight) for training, they both construct a synthetic dataset to pre-train their network in a supervised manner. To make the models more applicable to real images, [32] finetunes their network on real images, constrained by the low-rank property of diffuse components. However, this property does not hold for facial images due to the complexity of facial reflectance, which limits the diversity of their output diffuse colors. [34] further captures 150 paired real images under laboratory environment for network finetuning and achieves good performance on their laboratory testing images. Due to the limitations of their real training samples, their methods do not work well on the in-the-wild images.

In this paper, we propose a facial highlight removal method based on the Generative Adversarial Network (GAN) [6]. We first pre-train the generator using a high-quality synthetic dataset from [29]. Then, we develop a highlight mask guidance training technique to utilize in-the-wild facial images in the GAN training. As trained with real in-the-wild facial images, our method can generate convincing results on these inputs. However, adopting traditional GAN in the highlight removal cannot directly utilize in-the-wild images, as it requires real and fake image samples: one with highlight and one without highlight. Therefore, directly introducing GAN is hindered by the difficulty in capturing highlight-free in-the-wild images. In this paper, we have an observation that for a daily-recorded facial image, there are always some local regions without highlight. These regions can be indicated by a highlight mask and used for training a discriminator. When the discriminator learns the prior knowledge from the non-highlight facial regions, it lets the generator generate photo-realistic non-highlight facial regions. To extract these highlight masks, we use a traditional highlight extraction method [31]. We observe that although [31] may generate artifacts for facial highlight removal, its extracted highlight shares a strong correlation with the real highlight distribution as shown in Figure 2. We use a tactful strategy to convert the extracted highlight to highlight mask.

Our method is evaluated through extensive experiments. We evaluate the highlight mask guidance training technique with comprehensive ablation studies. We compare our methods with state-of-the-art techniques both qualitatively and quantitatively. To validate the effectiveness of our method, we present a large amount of highlight removal results on the in-the-wild images.

2 Related Work

2.1 Highlight Removal

There are many previous works about separating highlights from the input image. Early methods [1, 13] rely on color segmentation. However, algorithms based on color segmentation cannot handle complicate textured images. Based on the observation that highlight pixels contain useful information, [20] proposes to combine illumination-based constraints and image inpainting to remove highlights. [11] uses the dark channel prior to generate a specular-free image. [31] proposes a real-time highlight removal method based on bilateral filtering.

When the above methods are applied to face images, they usually generate obvious artifacts or lose the photo-realism. The reason is that they are not designed for face images and utilize some assumptions which do not work on highlights on faces. [17] proposes a method based on a skin model of human faces, which achieves the state-of-the-art highlight removal effect for facial images. However, it is time-consuming to solve the optimization. [32, 34] speed up facial highlight removal techniques based on deep neural networks. They both use synthetic images to pre-train their networks. [32] uses a low-rank assumption [7] when finetuning their network on real images but this assumption does not hold well on facial images. [34] introduces GAN into highlight removal methods. However, as they require paired training data, the real images used in their training are captured in a laboratory environment with limited numbers and lighting conditions.

Our method takes one step further. With a highlight mask guidance, our method facilitates the usage of in-the-wild images in the GAN training. Thus, our network can be trained with images captured under much more varieties of environments and lightings.

2.2 Generative Adversarial Networks

Generative models have always been an important topic in the field of computer vision. In recent years, more and more studies on deep neural networks show their potential in image generation [5, 12]. Generative adversarial networks [6] is an important method to train a certain generative model. The adversarial training between the generator and the discriminator enables the generator to output photo-realistic images [4, 22]. An important application of GANs is image-to-image translation [8, 27]. It can be applied to a variety of tasks such as super-resolution [15] and style transfer [9]. Instead of directly generating images, [25, 14] generate the corresponding residual image between the input and output. Residual learning helps to stabilize the training process. We also adopt the residual learning inspired by these pioneering works.

Specialized in face image processing, [19] uses conditional GAN [18] to control the attributes of the faces in output images. MaskGAN [16] provides more freedom in facial image manipulation. StyleGAN [10] can synthesize high resolution photorealistic facial images. [34] also uses GAN in the task of highlight removal. Indicated by CycleGAN like methods [8, 33], GANs have the potential to be trained using unpaired images. To the best of our knowledge, training GAN using unparied images has never been explored in the field of highlight removal. The proposed method fills the gap in this field and achieves visually pleasing results.

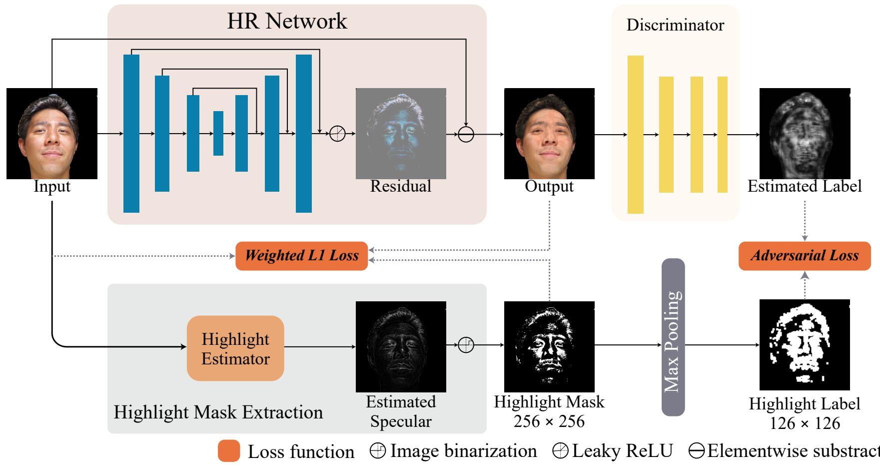


Fig. 1. The overall architecture of our method. Our network has a generator $\mathcal{G}$ and a discriminator $\mathcal{D}$. The generator is first pre-trained on a synthetic dataset and finetuned on the in-the-wild images using our highlight mask guided training technique. We use the highlight mask extraction module to generate binary highlight masks. A Leaky ReLU layer is added at the end of the generator to encourage positive outputs.

2.3 Face Color Transfer

Highlights have always been a limitation in the field of face color transfer. In [23], quotient image technique is proposed to change the lightings in facial images. Using this technique, the lightings of the input images can be changed by multiplying the ratio of two reference images captured under the target lighting and the source lighting. 3D reconstruction can also be used to edit the lighting in the facial images. This method uses a three-dimensional face morphable model [2] to fit the face geometry in the input image and estimate the albedo of faces and environment lighting and output a relighted facial image by re-rendering the facial image under the target lighting. However, the spherical harmonic light model [21] and the Lambertian surface assumption make it impossible to recover albedo under extreme lighting such as highlights. To further handle this problem, [28] unless a Markov random field to model face texture. Mass transport approaches are also used to transfer color from a reference facial image to the target one. Using mass transport, [26] remaps the color of the input image according to the facial geometry. By constructing a synthetic face dataset with multiple reflectance channels, [29] generates convincing specular and shadow of target lightings.

3 Our Method

In this section, we will present the details of our technique. The pipeline of our method is illustrated in Figure 1, we use a Highlight Mask Extraction module to extract the highlight masks as guidance for GAN training. We first assume the highlight masks as existing values and introduce the highlight mask guided GAN in Section 3.1. Then, we introduce Highlight Mask Extraction module in Section 3.2. Our implementation details are given in Section 3.3.

3.1 Highlight Mask Guided GAN

Given an input facial image I, facial highlight removal aims to change the pixel values of the facial highlight regions to show the original skin colors while preserving the values of other pixels. As shown in Figure 1, we propose a highlight removal architecture using generative adversarial networks. The generator $\mathcal{G}$ outputs the residual between the input image I and the estimated highlight removal result $\hat{I}_d$:

$$\hat{I}_d = I - \mathcal{G}(I). \tag{1}$$

The generator $\mathcal{G}$ is composed of an encoder-decoder architecture. Considering that highlight components in an image will always have positive values, we add a leaky rectified linear unit (Leaky ReLU) activation function in the last layer of $\mathcal{G}$.

For the discriminator, directly using traditional GAN methods in highlight removal requires the discriminator to tell whether the input image contains or does not contain highlights. However, we observe that a whole face region without any highlights is difficult to capture in a non-laboratory environment, so it is difficult to collect enough positive samples to train a discriminator that can determine whether an input facial image contains highlights or not. To empower the GAN training in highlight removal with in-the-wild images, we design our network architecture based on the following consideration. As most natural facial images contain both highlight and highlight-free regions, we can train a region-based discriminator $\mathcal{D}$ which outputs a low-resolution highlight mask M_L to indicate highlight and highlight-free regions. To be noted, we call the low-resolution highlight mask "highlight label map" to distinguish it from the original highlight mask M_H. By minimizing the size of the highlight regions obtained by the discriminator, the generator could be trained to perform the highlight removal task.

We also adopt the two-stage training strategy from [7, 3]. Before trained on real images with highlight mask guidance, the network is pre-trained in a supervised manner on a synthetic dataset [29], which contains about 270,000 paired images with and without highlight.

Supervised Loss Given a synthetic image I^s and its corresponding image without highlight I_d^s, the estimated $\hat{I}_d^s$ should be similar to I_d^s. We directly employ L_1 loss as follows,

$$\mathcal{L}_{sup} = \|\hat{I}_d^s - I_d^s\|_1. \tag{2}$$

Weighted-L1 Loss When trained on real images I^r, we use a weighted-L_1 loss to constrain the highlight removal results $\hat{I}_d^r$ should have the same content as I^r. As highlights have different intensities in I^r, the intensity of each pixel in the residual $\mathcal{G}(I^r)$ should varies with facial regions. Therefore, for non-highlight regions, large weights are applied. For highlight regions, we use small weights for highlight regions. With the highlight mask indicating highlight regions, the weighted-L_1 loss are defined as follows,

$$\mathcal{L}_{wL_1} = \lambda_H \|M_H \odot (\hat{I}_d^r - I^r)\| + \lambda_{NH} \|(1 - M_H) \odot (\hat{I}_d^r - I^r)\|, \tag{3}$$

where $\odot$ represent the element-wise multiplication operator.

Adversarial Loss We apply the different adversarial loss in training $\mathcal{G}$ and $\mathcal{D}$,

$$\mathcal{L}_{adv} = \mathcal{L}_{\mathcal{G}} + \mathcal{L}_{\mathcal{D}}, \tag{4}$$

where $\mathcal{L}_{\mathcal{G}}$ is for training the generator $\mathcal{G}$ and $\mathcal{L}_{\mathcal{D}}$ is for training the discriminator $\mathcal{D}$. As the non-highlight regions change little after the highlight removal, the discriminator should recognize these regions as real samples before and after the highlight removal. For highlight regions, the discriminator should not only distinguish them in the input image but also recognize the fake non-highlight regions generated by the generator. Therefore, the highlight regions should always be treated as fake samples. When training the generator, we require that the highlight removal result $\hat{I}_d^r$ can fake the discriminator over all facial regions. We define the adversarial loss functions as follows,

$$\mathcal{L}_{\mathcal{D}} = \mathrm{BCE}(\mathcal{D}(I^r), M_L) + \mathrm{BCE}(\mathcal{D}(\hat{I}_d^r), M_L), \tag{5}$$

$$\mathcal{L}_{\mathcal{G}} = \mathrm{BCE}(\mathcal{D}(\hat{I}^r), \mathbf{0}), \tag{6}$$

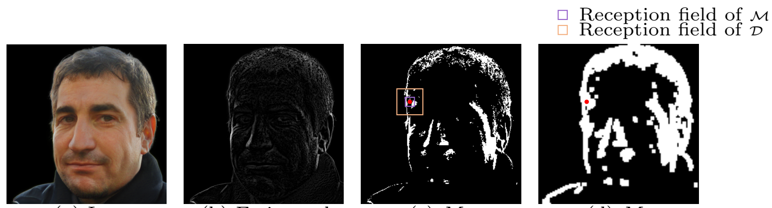
where BCE represent the Binary Cross Entropy loss and $\mathbf{0}$ represent a map filled with 0 with the same size as the highlight label map M_L. Here, 0 represent real while 1 represents fake.

Overall, our total loss function is expressed as,

$$\mathcal{L} = \lambda_{sup}\mathcal{L}_{sup} + \mathcal{L}_{wL_1} + \mathcal{L}_{adv}. \tag{7}$$

3.2 Highlight Mask Extraction

In this subsection, we detail how we extract the highlight mask of a given input image. Although previous methods [31] fail to remove highlight from facial images, their extracted highlights can still represent an approximate distribution of highlights. As shown in Figure 1, 2, we first estimate the specular component from the input facial using [31]. Then, we use a threshold t to convert the estimated specular into a binary highlight mask M_H. As for the low-resolution highlight label map M_L, it is converted from the highlight mask M_H using a



(a) Input (b) Estimated specular (c) M_H (d) M_L

Fig. 2. The intermediate results and the final output of the highlight mask extraction module. The center point (indicated by the red points) of the reception field of the max pooling layer $\mathcal{M}$ is set to be the same as that of the discriminator $\mathcal{D}$. Besides, M_L and $\mathcal{D}$ have the same size.

max pooling layer $\mathcal{M}$. To be noticed, although the reception field of the max pooling layer $\mathcal{M}$ is smaller than that of the discriminator D, the center points of both $\mathcal{M}$ and $\mathcal{D}$ are set to be the same. This can make the highlight label map has the same size as the output of $\mathcal{D}$. We apply the highlight mask extraction on real images from the FFHQ dataset [10] and use them in the finetuning stage.

3.3 Implementation Details

To train our highlight removal network, we use a two-stage training strategy. In the pre-training stage, we set the learning rate to 0.002 only with the supervised loss on the synthetic dataset. Then, in the finetuning stage, the learning rate is decreased to 0.0002 with our total loss function $\mathcal{L}$. Our model is implemented in Pytorch and trained using one NVIDIA GTX 1080Ti GPU for 20 hours. We set the threshold t in the highlight mask extraction module to 0.02, λ_{sup} to 10, λ_H to 10 and λ_L to 0.1. In the max pooling layer $\mathcal{L}$, we set its kernel size to 6, stride to 2 and padding size to 0. The average running time is about 8ms during testing with a 512×512 input image.

4 Experiments

In this section, we will first evaluate the key contributions of our proposed method. Then, we compare our method with the state-of-the-art highlight removal methods qualitatively and quantitatively. Especially, we present the comparisons on the photorealistic in-the-wild images generated by StyleGAN [10] to show that our method can improve the performance on the in-the-wild images. For quantitative experiments, as most previous methods do not release their codes or models and highlight-free images are difficult to capture, we perform the experiments based on the results reported in their paper.

4.1 Ablation Study

To illustrate the effectiveness of our highlight mask guided training strategy, we present the results of different setups. The baseline method is trained only with

Fig. 3. Qualitative evaluation of the proposed highlight masked guided GAN training on in-the-wild images. (b) shows the results of a highlight removal network only trained with supervised loss $\mathcal{L}_{sup}$. (c) represents the incorporation of traditional GAN training. (d) indicates our final method.

the supervised loss $\mathcal{L}_{sup}$. Then, we add a traditional adversarial loss, in which the discriminator distinguishes whether its input image is in the real image distribution or generated by the generator. We compare these two methods with our final method.

We present some qualitative comparisons in Figure 3. The results in Figure 3(c) demonstrate that directly implementing traditional GAN training in the highlight removal task will lead to meaningless results. Since the synthetic dataset [29] cannot explicitly model the eyes in their data synthesis, the network trained only with the supervised synthetic data fail to handle the highlights in the eyes as shown in the first row in Figure 3(b). Besides, when highlight leads to information loss in the input, the highlight masked guided GAN training enables the network to recover the missing content and generating visually pleasing results. However, only using the supervised loss $\mathcal{L}_{sup}$ itself will lead to color deviation or unresolved highlight as shown in Figure 3(b). These comparisons manifest that with the help of the highlight masked guided GAN training, we improve the highlight removal effect of our method.

4.2 Comparisons with Previous Techniques

To demonstrate the effect of the proposed method, we present both quantitative and qualitative comparisons with previous highlight removal techniques. The compared methods include some traditional general highlight removal methods [31, 24] and highlight removal methods designed for facial highlight removal [17, 32].

For quantitative experiments, since [17, 32] do not release their codes, we directly compare our methods with previous techniques based on laboratory images presented in [32]. To be noted, the RMSE and SSIM metrics differ from those reported in [32] as we crop the image for better visualization and recompute

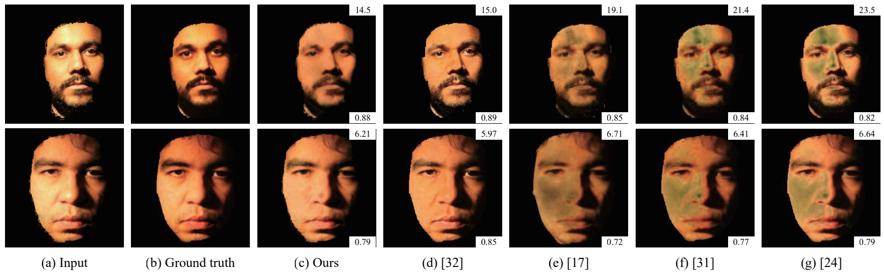


Fig. 4. Quantitative evaluation comparisons with the state-of-the-art on laboratory images. We compare our methods (c) with general highlight removal methods (f) [31] and (g) [24], and facial highlight removal techniques (d) [32] and (e) [17]. The performance is measured by RMSE (upper right) and SSIM (lower right). As the codes of [32, 17] are not released, the input, ground truth and the results of these methods are from the paper of [32] for comparison convenience.

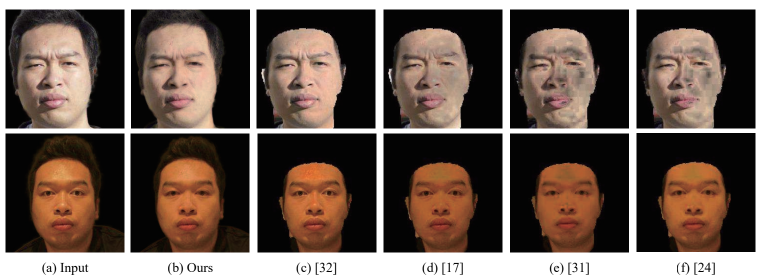


Fig. 5. Qualitative evaluation comparisons with the state-of-the-art on in-the-wild images. As our network is trained with highlight mask guided GAN on the in-the-wild images, our results outperforms all the state-of-the-art methods on the in-the-wild images.

these metrics between all the results and the ground truth for a fair comparison. The results are shown in Figure 4. As we can see, our method achieves better performance compared to [31, 24, 17]. When compared to [32], our method generates comparable result in the first row of Figure 4. Although the metrics in the second row are not as good as [32], our method still generates visually pleasing results.

The above quantitative experiments are performed on laboratory images while our method mainly focuses on the in-the-wild images. Therefore, we further make comparisons based on the in-the-wild images from [32] to demonstrate the advantages of our methods. As shown in Figure 5, our method generates more visually pleasing results compared to all the state-of-the-art methods. Although [32] works well on the laboratory images, it fails to handle the highlight around the nose and the mouth region in these in-the-wild images. Benefitting from our

Fig. 6. Limitations of our methods. Although our method works well on facial regions, white hair and white clothes might be recognized as highlight by our network.

proposed highlight mask guided training technique, our method is still robust on these inputs and generates convincing results.

4.3 More Results

We also provide more highlight removal results in our supplemental material to demonstrate the robustness of our method. In these results, we can see that our method can handle in-the-wild images with different races, genders, ages, and input lightings.

4.4 Limitations

Our method improves the highlight removal effect on the facial regions, especially on the in-the-wild images. However, as the highlight mask extraction module is designed for facial regions, other white objects in an image might be recognized as highlight. This will lead to color deviation in these regions. As shown in Figure 6, white hair and the white collar turn yellow after the highlight removal. If we add extra image parsing information in the training, we believe our method can achieve more visually pleasing results.

5 Conclusions

In this paper, we proposed a highlight removal method based on generative adversarial networks. Our method achieves high-quality results in facial highlight removal compared with the state-of-the-art, especially on the in-the-wild images. To avoid the lack of images that have no highlights, with the proposed highlight mask guided GAN training technique, our network is finetuned on the unpaired in-the-wild images. Experiments on laboratory images and in-the-wild images validate the effectiveness of our method. We hope the proposed approach can inspire future works on related problems.
Acknowledgements. This work was supported by the National Key R&D Program of China 2018YFA0704000, the NSFC (No.61822111, 61727808, 61671268, 62025108) and Beijing Natural Science Foundation (JQ19015, L182052). Xun Cao is the corresponding author.

References

1. Bajcsy, R., Lee, S.W., Leonardis, A.: Detection of diffuse and specular interface reflections and inter-reflections by color image segmentation. International Journal of Computer Vision **17**(3), 241–272 (1996)
2. Blanz, V., Vetter, T.: A morphable model for the synthesis of 3d faces. In: Proceedings of the 26th annual conference on Computer graphics and interactive techniques. pp. 187–194 (1999)
3. Debevec, P., Hawkins, T., Tchou, C., Duiker, H.P., Sarokin, W., Sagar, M.: Acquiring the reflectance field of a human face. In: Proceedings of the 27th annual conference on Computer graphics and interactive techniques. pp. 145–156 (2000)
4. Denton, E.L., Chintala, S., Fergus, R., et al.: Deep generative image models using a laplacian pyramid of adversarial networks. In: Advances in neural information processing systems. pp. 1486–1494 (2015)
5. Dosovitskiy, A., Springenberg, J.T., Brox, T.: Learning to generate chairs with convolutional neural networks. In: Computer Vision and Pattern Recognition (CVPR), 2015 IEEE Conference on. pp. 1538–1546. IEEE (2015)
6. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. In: Advances in neural information processing systems. pp. 2672–2680 (2014)
7. Guo, J., Zhou, Z., Wang, L.: Single image highlight removal with a sparse and low-rank reflection model. In: Proceedings of the European Conference on Computer Vision (ECCV). pp. 268–283 (2018)
8. Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 1125–1134 (2017)
9. Johnson, J., Alahi, A., Fei-Fei, L.: Perceptual losses for real-time style transfer and super-resolution. In: European Conference on Computer Vision. pp. 694–711. Springer (2016)
10. Karras, T., Laine, S., Aila, T.: A style-based generator architecture for generative adversarial networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 4401–4410 (2019)
11. Kim, H., Jin, H., Hadap, S., Kweon, I.: Specular reflection separation using dark channel prior. In: Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on. pp. 1460–1467. IEEE (2013)
12. Kingma, D.P., Welling, M.: Auto-encoding variational bayes. arXiv preprint arXiv:1312.6114 (2013)
13. Klinker, G.J., Shafer, S.A., Kanade, T.: The measurement of highlights in color images. International Journal of Computer Vision **2**(1), 7–32 (1988)
14. Kupyn, O., Budzan, V., Mykhailych, M., Mishkin, D., Matas, J.: Deblurgan: Blind motion deblurring using conditional adversarial networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 8183–8192 (2018)
15. Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z., et al.: Photo-realistic single image super-resolution using a generative adversarial network. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 4681–4690 (2017)
16. Lee, C.H., Liu, Z., Wu, L., Luo, P.: Maskgan: Towards diverse and interactive facial image manipulation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (June 2020)

17. Li, C., Lin, S., Zhou, K., Ikeuchi, K.: Specular highlight removal in facial images. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 3107–3116 (2017)
18. Mirza, M., Osindero, S.: Conditional generative adversarial nets. arXiv preprint arXiv:1411.1784 (2014)
19. Perarnau, G., van de Weijer, J., Raducanu, B., Álvarez, J.M.: Invertible conditional gans for image editing. arXiv preprint arXiv:1611.06355 (2016)
20. Quan, L., Shum, H.Y., et al.: Highlight removal by illumination-constrained inpainting. In: Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on. pp. 164–169. IEEE (2003)
21. Ramamoorthi, R., Hanrahan, P.: A signal-processing framework for inverse rendering. Computer Graphics Proceedings, SIGGRAPH 2001 pp. 117–128 (2001)
22. Salimans, T., Goodfellow, I., Zaremba, W., Cheung, V., Radford, A., Chen, X.: Improved techniques for training gans. In: Advances in Neural Information Processing Systems. pp. 2234–2242 (2016)
23. Shashua, A., Riklin-Raviv, T.: The quotient image: Class-based re-rendering and recognition with varying illuminations. IEEE Transactions on Pattern Analysis and Machine Intelligence **23**(2), 129–139 (2001)
24. Shen, H.L., Zheng, Z.H.: Real-time highlight removal using intensity ratio. Applied optics **52**(19), 4483–4493 (2013)
25. Shen, W., Liu, R.: Learning residual images for face attribute manipulation. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 1225–1233. IEEE (2017)
26. Shu, Z., Hadap, S., Shechtman, E., Sunkavalli, K., Paris, S., Samaras, D.: Portrait lighting transfer using a mass transport approach. ACM Transactions on Graphics (TOG) **37**(1), 2 (2018)
27. Taigman, Y., Polyak, A., Wolf, L.: Unsupervised cross-domain image generation. arXiv preprint arXiv:1611.02200 (2016)
28. Wang, Y., Zhang, L., Liu, Z., Hua, G., Wen, Z., Zhang, Z., Samaras, D.: Face relighting from a single image under arbitrary unknown lighting conditions. IEEE Transactions on Pattern Analysis and Machine Intelligence **31**(11), 1968–1984 (2008)
29. Wang, Z., Yu, X., Lu, M., Wang, Q., Qian, C., Xu, F.: Single image portrait relighting via explicit multiple reflectance channel modeling. ACM Transactions on Graphics (TOG) **39**(6), 1–13 (2020)
30. Yang, Q., Tang, J., Ahuja, N.: Efficient and robust specular highlight removal. IEEE transactions on pattern analysis and machine intelligence **37**(6), 1304–1311 (2015)
31. Yang, Q., Wang, S., Ahuja, N.: Real-time specular highlight removal using bilateral filtering. In: European conference on computer vision. pp. 87–100. Springer (2010)
32. Yi, R., Zhu, C., Tan, P., Lin, S.: Faces as lighting probes via unsupervised deep highlight extraction. In: Proceedings of the European Conference on computer vision (ECCV). pp. 317–333 (2018)
33. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE international conference on computer vision. pp. 2223–2232 (2017)
34. Zhu, T., Xia, S., Bian, Z., Lu, C.: Highlight removal in facial images. In: Chinese Conference on Pattern Recognition and Computer Vision (PRCV). pp. 422–433. Springer (2020)