Blind Single Image Reflection Suppression for Face Images using Deep Generative Priors

Paramanand Chandramouli
University of Siegen
Paramanand.Chandramouli@uni-siegen.de

Kanchana Vaishnavi Gandikota
University of Siegen
kanchana.gandikota@student.uni-siegen.de

Abstract

The goal of single image reflection removal is to suppress unwanted merging of radiances from different surfaces in the scene. This is an inherently ill-posed and challenging problem. Conventional approaches use different assumptions and constraints on the background and reflected layers to solve this problem. Recently, deep learning-based approaches have been applied to this task. These methods require extensive amount of realistic data for training. In this paper, we propose to incorporate class-specific prior models for reducing the ill-posedness of the reflection separation task. Specifically, we use a pre-trained deep face-generative model for reflection supression from face images. We design an optimization scheme that effectively leverages the deep generative model and leads to a constrained solution space. Our method does not require training data corresponding to reflection separation task. We evaluate our proposed approach using both synthetic and real world facial images containing reflections and compare with existing state-of-the-art techniques. The results demonstrate advantages of our approach over the current state-of-the-art in single image reflection separation from faces.

1. Introduction

Photographs taken through transparent/translucent surfaces contain reflections in addition to the desired background scene. The presence of such reflections limits the visibility of the desired scene. Moreover, presence of reflections in facial images can severely limit the performance of face-specific computer vision algorithms. Furthermore, separating reflections from facial images could be useful in surveillance and security applications. This motivates the need to effectively remove unwanted reflections from scenes, in particular face images captured through a glass pane.

An observed image through a glass pane $\mathbf{I}$ can be modeled as a superposition of a background scene $\mathbf{B}$ and a reflected component $\mathbf{R}$

$$\mathbf{I} = \mathbf{B} + \mathbf{R} \tag{1}$$

The goal of reflection removal is to suppress unwanted reflections $\mathbf{R}$ so that background scene $\mathbf{B}$ can be reliably recovered. This is an inherently ill-posed problem, which requires additional information to obtain a reliable solution. One approach is to make use of multiple images [5, 8, 11, 37] and exploit additional cues so that the problem becomes more tractable. When multiple images are not available, reflection removal from a single image using conventional approaches [20, 26, 2, 32] use prior assumptions and constraints on the background and reflection layers to make the problem less ill-posed. However, such prior assumptions may not always hold, in which case, these conventional approaches cannot yield satisfactory results.

Recently, deep learning based approaches have also been applied to the problem of single image reflection separation in [7, 40, 14, 38, 35, 36], which improved the state-of-the-art in reflection suppression. However, capturing large scale real datasets for the task of single image reflection suppression is highly challenging. Due to lack of availability of large scale real datasets for training neural networks for the task of single image reflection removal, deep learning based reflection removal techniques such as [7, 40, 14, 38, 35, 36] use synthetic datasets for learning reflection removal. When synthesized datasets cannot adequately capture the variations in real-world reflection corrupted images, deep learning based techniques can suffer from decline in performance while suppressing reflections in real-world images.

In this paper, we propose a novel optimization-based unsupervised blind reflection suppression method for facial images using deep generative priors for the background layer. In recent years, significant success has been achieved by neural network-based deep generative models in learning the representation for different kinds of data [17, 10, 29, 15, 23]. Our objective is to exploit face-specific priors learnt by a generative model to guide the optimization of the background scene radiance. We observe that such an approach aids in reducing the ill-posedness of the problem.

Input [40] [36] [14] [39] Ours

Figure 1. Visual comparison on real images from prior work. The input and output images are of size 64×64 for our algorithm, while they are of size 256×256 for the methods in [14, 40, 36, 39].

Our scheme does not require extensive training datasets for the task of reflection suppression.

In Fig. 1, we see two real examples of face images with reflections along with the results of the techniques in [14, 40, 36, 39] and our result. Except for one of the result of [40], we observe that these recent state-of-the-art reflection suppression techniques are unable to remove reflection. This indicates that the current models are still not adequate to handle such difficult tasks. In contrast, our method is successful to some extent. However, there is a limitation of our method. As we see subsequently, the resolution of the output of our method is limited by the generative prior that we utilize.

2. Prior Work

Since reflection suppression is a highly ill-posed problem, additional information is required to reliably recover the desired background scene. Existing approaches to reflection suppression can be broadly classified as single image-based and multiple image-based approaches. Use of multiple images can make the problem of reflection separation more tractable. Multiple image based reflection suppression exploiting motion cues are used in [37], specifically motion cues such as translative motion in [5], affine transformations in [8], homography in [11]. There are also multiple image based reflection removal methods utilizing images taken under different imaging conditions such as images with flash and no-flash [1], images with different focus settings [25], and images with different polarization [18].

Often, multiple images are not available, and therefore reflection suppression from single image becomes necessary. For single images, conventional reflection removal algorithms incorporate additional constraints and priors on the background and reflected layers to perform reflection separation. When imaging through glass, it is more likely that the desired background is in focus and hence sharp, whereas the reflection layer could be out of focus and therefore blurry. The authors in [20] exploit separable sparse gradients prior due to different levels of blur in background and reflection layers. Similarly, in [32], the authors estimate depth of field to separate the background and reflection layers. The authors in [2] use Laplacian data fidelity term and an l_0 gradient sparsity prior for the background layer for reflection suppression. These methods [20, 32, 2] assume that reflection layer is much blurred compared to the desired background, and therefore can breakdown when this assumption is not met. Often reflected component is predominant in only a part of the image. For such images, the authors in [30] first estimate a reflection map based on different blur levels for the two layers, and use a content prior based on the internal patch recurrence to recover regions in the background image contaminated by reflections. When images are captured through thick glass panes, the reflected layer contains shifted attentuated versions of the same scene, which is referred to as ghosting effect. The authors in [26, 13, 12], make use of such ghosting cues to separate the reflection component and background layer from single image. Recently, the authors in [39] formulated reflection suppression as a convex model, and obtained its solution by gradient thresholding and solving a partial differential equation using Discrete Cosine Transform. However, this approach tends to fail when edges in the reflection layer are sharp and strong.

Recently, deep learning based approaches have been applied to the task of single image reflection suppression. Fan *et al*. [7] proposed the first convolutional neural network (CNN) to solve single image reflection suppression and to handle artifacts due to saturation. A compact neural network for reflection separation was proposed in [14], where color ambiguity and saturation are handled explicitly in synthetic data generation. Authors in [40] trained a deep network for reflection suppression by including a perceptual feature loss, an adversarial loss for background layer and a novel pixel-level exclusion loss and showed improved performance in reflection removal. Yang *et al*. [38] proposed a bidirectional network which estimates both background scene and reflected image in a cascade neural network to improve reflection suppression. Due to lack of sufficiently large real datasets for the task of single image reflection removal, the works [7, 14, 40, 38] use synthetic datasets for training. This can lead to decline in performance while suppressing reflections in real world images. Authors in [31, 33] used a deep network which combine image appearance information and multi-scalegradient information to effectively suppress reflections from superposed images. Additionally, they also created a new Reflection Image Dataset containing 3250 images for training. Since creating large datasets for the task of reflection removal is difficult, recent works, [36, 16] instead attempt to improve the synthetic data generation to incorporate more realistic reflection scenarios. Wen *et al*. [36] use an adverserially trained synthesis network which outputs an alpha-blending mask to generate realistic reflection contaminated images. A reflection removal network is then trained with such synthetically generated training data. Kim *et al*. [16] use displacement mapping and path tracing to simulate the depth-dependent light transportation for synthesizing a realistic dataset for the reflection removal problem. A deep network is then trained using the generated data for reflection separation. Since imaging aligned pairs of images with and without reflections is difficult, Wei *et al*. [35] instead created a dataset with misaligned pairs of images, and trained a deep network for reflection removal using such misaligned data by incorporating specific network enhancements and high-level feature losses which are invariant to misalignments to improve reflection suppression. In the absence of training sets, Gandelsman *et al*. [9] used a coupled deep image prior networks for unsupervised layer decomposition tasks such as image-dehazing, foreground/background segmentation, watermark-removal with single image. However, for the more challenging decomposition task of reflection separation, [9] requires two different mixture images of the same background and reflection images.

The problem of reflection removal in images of specific classes is addressed in [24, 19, 34]. The authors in [24] use an optimization based approach to remove reflections from frontal face images with eye glasses using tight constraints which characterize eye glass reflection. In [19], categories of both transmitted background and reflected layers are assumed to be known, and a generative adverserial network (GAN) based network is used to jointly estimate both the layers. In contrast, in our approach, we assume the category of only background scene *i.e*. faces, with no additional assumption on reflected layer. Moreover, ours is an optimization based approach which does not require training data for reflection suppression. We note that, in the recent work [34], reflection suppression from facial images is performed using deep neural networks. In contrast to our unsupervised optimization based approach for face reflection removal, the approach in [34] uses very strong supervision during training, requiring two input images with reflection and ground truth background face along with corresponding label for face recognition task. Obtaining such highly supervised dataset is challenging and therefore authors in [34] synthesize face images with reflections by combining separately captured reflections and labelled background faces. Moreover, this network needs additional image for guiding reflection removal, whereas our work uses only single image for reflection suppression.

In recent years, deep generative models beginning with Variational Autoencoders (VAEs) [17] and GANS [10] have demonstrated impressive ability in learning representations by generating new images from the distribution of training data. Wasserstein autoencoders (WAEs) [29] which minimize a penalized form of the Wasserstein distance between the model and target distributions have shown improved performance over traditional VAEs which minimize Kullback-Leibler distance between these distributions. Recently, new powerful generative models such as progressive GANs (PGANs) [15] and Vector Quantized Variational AutoEncoder (VQ-VAE) [23] have been proposed which can generate very high quality images. Generative models have also been used in compressive sensing in [6], semantic photo manipulation [4], and class-specific image restoration tasks such as face inpainting [21] and deblurring [3]. In our work, we utilize face-specific priors learnt by a WAE model pretrained on face images to guide the optimization based recovery of background faces.

Highlights: The highlights of our work can be summarized as follows: i) We propose a single image reflection removal scheme without any need of realistic training data. ii) We develop an optimization scheme that suitably exploits a pre-trained generative model (WAE) for face images. This can be extended by using more powerful generative models such as PGANs, VQ-VAE to obtain high resolution results. iii) Experiments show that our scheme can achieve a superior level of performance in reflection suppression, which is not possible by existing methods, albeit with a limit on image resolution.

3. Optimization using Generative Priors

Since separation of background image $\mathbf{B}$ from the superposed image $\mathbf{I}$ in equation. (1), is ill-posed, we propose to use the prior knowledge that the background image $\mathbf{B}$ belongs to the specific class of face images. We incorporate this knowledge using a generative model $\mathbf{G}$ (a WAE) pretrained on facial images. We develop an optimization scheme that suitably exploits this pretrained generative model.

3.1. Optimizing latent representation

A possible approach could be to directly consider $\mathbf{B} = \mathbf{G}(z)$, where z is the corresponding representation of a face image in the latent space. Then, one could solve for the latent representation of background layer as

$$\hat{z} = \arg\min_{z} \|\mathbf{I} - \mathbf{G}(z) - \mathbf{R}\|_2^2 \tag{2}$$

However, this is still ill-posed, as the reflection layer $\mathbf{R}$ is still unknown. Moreover, even if $\mathbf{R}$ is known, solving equation. (2) is still ill-posed as noted in [4], where additional perceptual losses are also incorporated. Even then, when $\mathbf{G}$ is not able to generate images resembling $\mathbf{B}$, there will be representation error as noted in [6]. This is unavoidable because finding a latent code z that can accurately recover an arbitrary face image which is not similar to the training set is hard. Therefore, direct optimization on latent space of generative model is not feasible for the problem of single image reflection suppression. We therefore propose an Encoder-Decoder based optimization approach, which circumvents these problems by leveraging the encoding ability of a variational autoencoder.

3.2. Encoder-Decoder based optimization

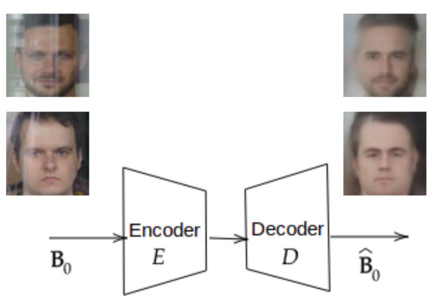

Figure 2. Projection of initial estimate of background face image into face-image space of the generative model. We see that the reflective components get removed when the the estimates are fed forward through the WAE.

Given an face image observation $\mathbf{I}$ corrupted by reflection components, we consider the input observation $\mathbf{I}$ as the initial estimate for both the background layer $\mathbf{B}$ and the reflected layer $\mathbf{R}$, since there is no other information available. Let E and D denote the encoder and decoder models of the WAE [29] pretrained using face images. Given an estimate of the background layer $\mathbf{B}_t$, where t denotes the iteration index, we can project it to the set corresponding to the range of the decoder D, by feeding it to the encoder and decoder as $\hat{\mathbf{B}}_t = D(E(\mathbf{B}_t))$. While the estimate of $\mathbf{B}_t$ would contain components of the reflected layer, the projected image $\hat{\mathbf{B}}_t$ would be devoid of the reflected components. This is because the decoder has been trained to generate face images. However, there would be discrepancies between the facial features of the generated projection and the true background face. In Figure. 2, we see examples of the result of such a projection when face images with reflection are fed through WAE. Due to reflections, the encoded representation $E(\mathbf{B}_t)$ would also not correspond to the latent representation of the true background face image. In order to arrive at an estimate which is consistent with the observed image, as well as to enforce the class-specific model, we propose a cost function $\mathcal{E}(\mathbf{B}_t)$ consisting of two main components: the data consistency loss $\mathcal{C}_{\text{loss}}(\mathbf{B}_t)$ and the generative prior loss $\mathcal{G}_{\text{loss}}(\mathbf{B}_t)$. These loss functions are defined as

$$\begin{aligned}\mathcal{C}_{\text{loss}}(\mathbf{B}_t) &= \lambda_{\text{VGG}} \sum_{i=1}^{N} \frac{1}{M_i} \|F^i\left(\hat{\mathbf{B}}_t + \mathbf{R}_t\right) - F^i(\mathbf{I})\|_2^2 + \\ &\qquad \|\hat{\mathbf{B}}_t + \mathbf{R}_t - \mathbf{I}\|_2^2 \\ \mathcal{G}_{\text{loss}}(\mathbf{B}_t) &= \|\mathbf{B}_t - \hat{\mathbf{B}}_t\|_2^2 + \lambda_{\text{VGG}} \sum_{i=1}^{N} \|F^i(\mathbf{B}_t) - F^i\left(\hat{\mathbf{B}}_t\right)\|_2^2,\end{aligned} \tag{3}$$

where $F^i()$ corresponds to i-th layer with M_i features in the VGG16 network [27]. In addition, we also minimize the correlation $\mathcal{R}_{\text{loss}}$ between the gradients of the two layers $\nabla\mathbf{B}_t$ and $\nabla\mathbf{R}_t$.

$$\mathcal{R}_{\text{loss}}(\mathbf{B}_t) = \|\nabla\mathbf{B}_t \odot \nabla\mathbf{R}_t\|_1, \tag{4}$$

where $\odot$ denotes point-wise multiplication. This is based on the observation that the edges of reflected and transmitted layer are distinct from each other [40]. We further incorporate total-variation prior on the background layer $TV_{\text{loss}}(\mathbf{B}) = \|\nabla\mathbf{B}_t\|_1$ to encourage sharpness of the background layer. Our combined loss function can be written as

$$\mathcal{E}_{\text{loss}} = \mathcal{C}_{\text{loss}} + \mathcal{G}_{\text{loss}} + \mathcal{R}_{\text{loss}} + TV_{\text{loss}} \tag{5}$$

Note that in equation 3, we use $\hat{\mathbf{B}}_t = D(E(\mathbf{B}_t))$ instead of $\mathbf{B}_t$ because initially $\mathbf{B}_t$ contains significant correlation with $\mathbf{R}_t$. Due to finite representation error of the generative model, $\hat{\mathbf{B}}$ and $\mathbf{B}_t$ are not equal. The use of perceptual loss components in our cost function, helps in improving perceptual similarity between $\hat{\mathbf{B}}$ and $\mathbf{B}_t$. At each iteration t, we first update the background layer using RMSprop [28] keeping current estimate of the reflected layer fixed. Subsequently, we update the reflected layer as $\mathbf{R}_{t+1} = \mathbf{I} - \mathbf{B}_{t+1}$. Algorithm 1 summarizes our approach.

4. Experiments

We evaluate our approach to reflection suppression qualitatively and also quantitatively using normalized cross correlation metric. We first pretrain a Wassestein Autoencoder using CelebA [22] training dataset with face images of size 64x64. We use this pretrained WAE as our generative prior.

Figure 3. Synthetic experiment with defocused reflecting layer and images of size 256×256. Row 1: input. Rows 2-4: outputs of [14], [39], and [40], respectively. Last row: true background layer.

Algorithm 1 Encoder-Decoder based Optimization

Require: Face image with reflection $\mathbf{I}$
Ensure: clean image $\mathbf{B}$
Initialize $\mathbf{B}_0 = \mathbf{I}, \mathbf{R}_0 = \mathbf{I}$.
for $t = 0$ to T **do**
  $\hat{\mathbf{B}}_t = D\left(E\left(\mathbf{B}_t\right)\right)$.
  Compute the gradients of $\mathcal{E}_{\text{loss}}$ w.r.t. $\mathbf{B}_t$.
  Update $\mathbf{B}_{t+1}$ using RMSprop [28].
  Update $\mathbf{R}_{t+1} = \mathbf{I} - \mathbf{B}_{t+1}$.
end for
$\mathbf{B} = \mathbf{B}_T$.

Consequently, in all experiments with our approach, input and output images should be of size 64×64.

We evaluate our method on both real and synthetic observations. For synthetic experiments, we use faces from CelebA test dataset for the background layer $\mathbf{B}$. To generate reflected layer, we make use of cropped and resized images from PASCAL VOC 2012 dataset. The reflected layer is given as $R = \alpha \mathbf{R}_1 * h$, where $\mathbf{R}_1$ are the images from dataset, where we use $\alpha = 0.3$ or 0.5 as the scaling factors, h is the blur kernel, which is either an implulse function or a Gaussian filter with variance 3.5 and $*$ represents convolution operation. To generate superposed images $\mathbf{I}$, we consider $\mathbf{I} = f(\mathbf{B} + \mathbf{R})$, where f is a normalizing function that scales its input so that there is no saturation. For real experiments, we captured images of people with reflections and then we manually cropped regions corresponding to faces. The cropped regions were of sizes close to 350×350. To apply our scheme, we resized these cropped regions to 64×64. For both synthetic and real experiments, we apply our iterative optimization scheme for 6000 iterations with a learning rate of $1e - 4$ using alternate update steps for background layer and reflection layer as discussed in section. 3.2.

We compare the performance of our method with that of the techniques in [14, 40, 36, 39]. We use the implementations publicly shared by the authors of these papers. The deep network-based models of [14, 40, 36] consider that presence of a significant extent of receptive field. Therfore, we used inputs of higher resolution to these networks as well as to the method of [39].

Synthetic experiments We generate 50 observations in our evaluation by using different images for background and

Figure 4. Synthetic experiment with sharp reflecting layer and images of size 256×256. Row 1: input. Rows 2-4: outputs of [14], [39], and [40], respectively.

Figure 5. Our results with sharp reflecting layer. Row1: input images when the reflecting layer was scaled by 0.5. Row2: output of our method with Row1 as input. Row3: input images when the reflecting layer was scaled by 0.3. Row4: our method with Row3 as input.

reflective layer. For comparing with other methods, the observations are generated by resizing the two layers to 256×256. For our method, we resize the images to 64×64. Fig. 3 shows sample images and results from other methods when the reflected layer was defocused and scaled by a factor 0.3. In this experiment, we observe that results of the methods of [14, 40, 39] do remove the reflections by some extent. In some cases, the estimates are accurate and in some cases they are not. We also observed that the method of [36] was not working well on this data. We did try all the models that were made available by the authors of [36].

We next repeated the experiment (Fig. 4) with the same set of images but without defocusing the reflected layer. The reflected layer was weighted by a factor of 0.3. For this scenario, we observe that in most of the cases, the two layers were not getting separated. The experiments of Figs. 3 and 4, indicate that layer separation becomes harder when the reflected layer is also sharp.

We applied our method on the same set of images but with resizing. We quantified the performance by evaluating the normalized cross-correlation (NCC) between our estimate of the background face image and the true face image. When the reflected layer was defocused and weighted by a factor of 0.3, the average NCC value was 0.92. This value reduced to 0.90 when the reflected layer was kept sharp. One can see sample input and output images for this scenario in rows 3 and 4 of Fig. 5, respectively. We also tried a more difficult scenario wherein the reflected layer was scaled by 0.5 (Row1 of Fig. 5). Even for this scenario our results (Row2 of Fig. 5), indicate significant suppression of the reflected layer. In this case, the average NCC value dropped to 0.87.

Real experiments We qualitatively evaluate our method on real images in Figs. 1 and 6 by visual comparison with the approaches of [14, 40, 36, 39]. From Figs. 1 and 6 we can observe that these recent state-of-the-art reflection suppression techniques are unable to remove strong reflections from facial images. Only the approach in [40] second column in Figs. 1 and 6 is able to remove reflections from facial images, but only when the reflection component is not sharp (corresponding to row 1 in Fig. 1 and row 3 in Fig. 6). None of the approaches [14, 40, 36, 39] are able to suppress strong or sharp reflections from facial images. In contrast, our face generative prior based approach is able to better handle even such strong reflections albeit at low resolutions. In Fig. 6, we also show the effect of representation error by plotting the output $\hat{\mathbf{B}}$ of the decoder in the final iteration. The discrepancies between the faces shown in the last two columns is because it is hard for a generative model to accurately represent any arbitrary face image.

5. Conclusions

Blind reflection separation from single images is a challenging problem. Capturing large datasets for supervised learning of reflection removal is difficult. We have presented a novel optimization-based unsupervised blind reflection suppression method from a single facial image leveraging deep generative priors. Our approach does not require of large datasets for the task of reflection suppression. Our experimental results show that learning-based approaches trained on general reflection corrupted images do not generalize well to facial images. Comparisons with prior work using both synthetic and real experiments shows that our method outperforms prior state of art reflection removal approaches in removing reflections from facial images. Specifically, our approach can remove strong and sharp reflections, where all other methods fail. This shows that optimization using generative priors could be an effective alternative, especially when constructing supervised training dataset is tedious. However, our approach is limited by the resolution of generative prior model. This can be remedied by the use of deep priors based on more powerful generative models.

References

[1] A. Agrawal, R. Raskar, S. K. Nayar, and Y. Li. Removing photography artifacts using gradient projection and flash-exposure sampling. In *ACM Transactions on Graphics (TOG)*, volume 24, pages 828–835. ACM, 2005.

[2] N. Arvanitopoulos, R. Achanta, and S. Susstrunk. Single image reflection suppression. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4498–4506, 2017.

[3] M. Asim, F. Shamshad, and A. Ahmed. Blind image deconvolution using deep generative priors. *arXiv preprint arXiv:1802.04073*, 2018.

[4] D. Bau, H. Strobelt, W. Peebles, J. Wulff, B. Zhou, J.-Y. Zhu, and A. Torralba. Semantic photo manipulation with a generative image prior. *ACM Trans. Graph.*, 38(4):59:1–59:11, July 2019.

[5] E. Be'Ery and A. Yeredor. Blind separation of superimposed shifted images using parameterized joint diagonalization. *IEEE Transactions on Image Processing*, 17(3):340–353, 2008.

[6] A. Bora, A. Jalal, E. Price, and A. G. Dimakis. Compressed sensing using generative models. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, pages 537–546. JMLR. org, 2017.

[7] Q. Fan, J. Yang, G. Hua, B. Chen, and D. Wipf. A generic deep architecture for single image reflection removal and image smoothing. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 3238–3247, 2017.

[8] K. Gai, Z. Shi, and C. Zhang. Blind separation of superimposed moving images using image statistics. *IEEE transactions on pattern analysis and machine intelligence*, 34(1):19–32, 2011.

[9] Y. Gandelsman, A. Shocher, and M. Irani. Double-DIP: Unsupervised Image Decomposition via Coupled Deep-Image-Priors. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.

[10] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Gen-

Figure 6. Visual comparison on real images from prior work. The input and output images are of size 64×64 for our algorithm, while they are of size 256×256 for the methods in [14, 40, 36, 39].

erative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014.

[11] X. Guo, X. Cao, and Y. Ma. Robust separation of reflection from multiple images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2187–2194, 2014.

[12] S. Hsieh, Y. Yang, C. Yeh, S. Huang, and Y. Lu. Subpixel-level-accurate algorithm for removing double-layered reflections from a single image. In *2018 25th IEEE International Conference on Image Processing (ICIP)*, pages 395–399, Oct 2018.

[13] Y. Huang, Y. Quan, Y. Xu, R. Xu, and H. Ji. Removing reflection from a single image with ghosting effect. *IEEE Transactions on Computational Imaging*, 2019.

[14] M. Jin, S. Süsstrunk, and P. Favaro. Learning to see through reflections. In *2018 IEEE International Conference on Computational Photography (ICCP)*, pages 1–12. IEEE, 2018.

[15] T. Karras, T. Aila, S. Laine, and J. Lehtinen. Progressive growing of gans for improved quality, stability, and variation. *arXiv preprint arXiv:1710.10196*, 2017.

[16] S. Kim, Y. Huo, and S.-E. Yoon. Single image reflection removal with physically-based rendering. *arXiv preprint arXiv:1904.11934*, 2019.

[17] D. P. Kingma and M. Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.

[18] N. Kong, Y.-W. Tai, and J. S. Shin. A physically-based approach to reflection separation: from physical modeling to constrained optimization. *IEEE transactions on pattern analysis and machine intelligence*, 36(2):209–221, 2013.

[19] D. Lee, M.-H. Yang, and S. Oh. Generative single image reflection separation. *arXiv preprint arXiv:1801.04102*, 2018.

[20] Y. Li and M. S. Brown. Single image layer separation using relative smoothness. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2752–2759, 2014.

[21] Y. Li, S. Liu, J. Yang, and M.-H. Yang. Generative face completion. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3911–3919, 2017.

[22] Z. Liu, P. Luo, X. Wang, and X. Tang. Deep learning face attributes in the wild. In *Proceedings of International Conference on Computer Vision (ICCV)*, December 2015.

[23] A. Razavi, A. v. d. Oord, and O. Vinyals. Generating diverse high-fidelity images with vq-vae-2. *arXiv preprint arXiv:1906.00446*, 2019.

[24] T. Sandhan and J. Young Choi. Anti-glare: Tightly constrained optimization for eyeglass reflection removal. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017.

[25] Y. Y. Schechner, N. Kiryati, and R. Basri. Separation of transparent layers using focus. *International Journal of Computer Vision*, 39(1):25–39, 2000.

[26] Y. Shih, D. Krishnan, F. Durand, and W. T. Freeman. Reflection removal using ghosting cues. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3193–3201, 2015.

[27] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. In *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, 2015.

[28] T. Tieleman and G. Hinton. Rmsprop gradient optimization. *URL http://www. cs. toronto. edu/tijmen/csc321/slides/lecture_slides_lec6. pdf*, 2014.

[29] I. Tolstikhin, O. Bousquet, S. Gelly, and B. Schoelkopf. Wasserstein auto-encoders. *arXiv preprint arXiv:1711.01558*, 2017.

[30] R. Wan, B. Shi, L.-Y. Duan, A.-H. Tan, W. Gao, and A. C. Kot. Region-aware reflection removal with unified content and gradient priors. *IEEE Transactions on Image Processing*, 27(6):2927–2941, 2018.

[31] R. Wan, B. Shi, L.-Y. Duan, A.-H. Tan, and A. C. Kot. CRRN: Multi-scale guided concurrent reflection removal network. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.

[32] R. Wan, B. Shi, T. A. Hwee, and A. C. Kot. Depth of field guided reflection removal. In *2016 IEEE International Conference on Image Processing (ICIP)*, pages 21–25, Sep. 2016.

[33] R. Wan, B. Shi, H. Li, L. Duan, A. Tan, and A. K. Chichung. Corrn: Cooperative reflection removal network. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019.

[34] R. Wan, B. Shi, H. Li, L.-Y. Duan, and A. C. Kot. Face image reflection removal. *arXiv preprint arXiv:1903.00865*, 2019.

[35] K. Wei, J. Yang, Y. Fu, D. Wipf, and H. Huang. Single image reflection removal exploiting misaligned training data and network enhancements. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 8178–8187, 2019.

[36] Q. Wen, Y. Tan, J. Qin, W. Liu, G. Han, and S. He. Single Image Reflection Removal Beyond Linearity. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.

[37] T. Xue, M. Rubinstein, C. Liu, and W. T. Freeman. A computational approach for obstruction-free photography. *ACM Transactions on Graphics (TOG)*, 34(4):79, 2015.

[38] J. Yang, D. Gong, L. Liu, and Q. Shi. Seeing deeply and bidirectionally: A deep learning approach for single image reflection removal. In *The European Conference on Computer Vision (ECCV)*, September 2018.

[39] Y. Yang, W. Ma, Y. Zheng, J.-F. Cai, and W. Xu. Fast single image reflection suppression via convex optimization. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 8141–8149, 2019.

[40] X. Zhang, R. Ng, and Q. Chen. Single Image Reflection Separation with Perceptual Losses. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4786–4794, 2018.