DENOISING ADVERSARIAL NETWORKS FOR RAIN REMOVAL AND REFLECTION REMOVAL

Qian Zheng[1,*], *Boxin Shi*[2,*], *Xudong Jiang*[1], *Ling-Yu Duan*[2], *and Alex C. Kot*[1]

[1] School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore

[2] National Engineering Laboratory for Video Technology, School of EE & CS Peking University, Beijing, China

csqianzheng@gmail.com, {shiboxin, lingyu}@pku.edu.cn, {exdjiang, eackot}@ntu.edu.sg

ABSTRACT

This paper presents a novel adversarial scheme to perform image denoising for the tasks of rain streak removal and reflection removal. Similar to several previous works, the proposed method first estimates a prior image and then uses it to guide the inference of noise-free image. The novelty of our approach is to jointly learn the gradient and noise-free image based on an adversarial scheme. More specifically, we use the gradient map as the prior image. The inferred noise-free image guided by an estimated gradient is regarded as a negative sample, while the noise-free image guided by the ground truth of a gradient is taken as a positive sample. With the anchor defined by the ground truth of noise-free image, we play a min-max game to jointly train two optimizers for the estimation of the gradient and the inference of noise-free images. We show that both prior image and noise-free image can be accurately obtained under this adversarial scheme. Our state-of-the-art performance achieved on two public benchmark datasets validate the effectiveness of our approach.

Index Terms— Rain Streak Removal, Reflection Removal, Adversarial Learning, Deep Learning

1. INTRODUCTION

In recent years, deep convolutional neural networks (CNNs) have achieved unprecedented success in the problem of image denoising, *e.g.*, dehazing [1], rain drops removal [2], rain streak removal [3], and reflection removal [4]. Rain streak removal and reflection removal are problems which attempt to recover or infer rain- or reflection-free images from rain- or reflection-contaminated images.

Inferring a noise-free image from a noise-contaminated image mixed by it and a noisy image is an ill-posed problem since both noise-free image and noise image are unknown. To make it tractable, several assumptions have been made in the literature. Traditional methods make explicit assumptions based on the priors of images, such as prior on noise-free images (*e.g.*, sparsity [5], gradient [6]), or prior on noise images (*e.g.*, low-rank [7], ghosting effects [8]). However, these assumptions may not always be held in practice due to the complexity of actual situations [9]. And methods based on them tend to provide over-smooth results [10].

In order to avoid limitations above, learning based methods try to learn such priors from data. A recent study has shown that directly optimizing the regression between noise-contaminated images and noise-free images can bring the problem of underfitting [9]. Therefore, researchers propose to use a prior image [9] to guide the inference of noise-free images. The process of denoising is then composed by concatenating the stages of estimating the prior image and inferring the noise-free image. The definition of such prior image is always motivated by traditional methods, *e.g.*, gradient map [11, 12], detail map [9], rain streak or mask [3], levels of rain streak [10], etc. However, there is no universal strategy to jointly optimize the prior image and the noise-free image. Almost all of these works above [9, 3, 10, 12, 11] elaborately design specialized strategy for their prior images, leading to the difficulty of improving or modifying the network structures.

To this end, this paper presents a novel method which jointly optimizes the prior image and the noise-free image in an adversarial scheme. We adopt deep neural networks to jointly optimize the prior image estimator and the sample constructor. Benefitting from the independence between these two optimizers, their design can be more flexible. That is, the architectures of these two networks can be adjusted separately to meet different requirements, *e.g.*, modifying inputs or outputs of neural networks or simply making them deeper.

Our major contributions are in two folds: 1) we propose denoising adversarial networks in which accurate prior images and noise-free images can be acquired simultaneously, 2) we apply our approach to the problems of rain streak removal and reflection removal and achieve the state-of-the-art performance on public benchmark datasets.

2. THE PROPOSED METHOD

In this section, we first introduce the key idea of our approach and show that it can simultaneously achieve optimal solutions for both the prior image and the noise-free image in Section 2.1. We then detail our implementation and summarize our algorithm in Section 2.2.

*Corresponding authors.

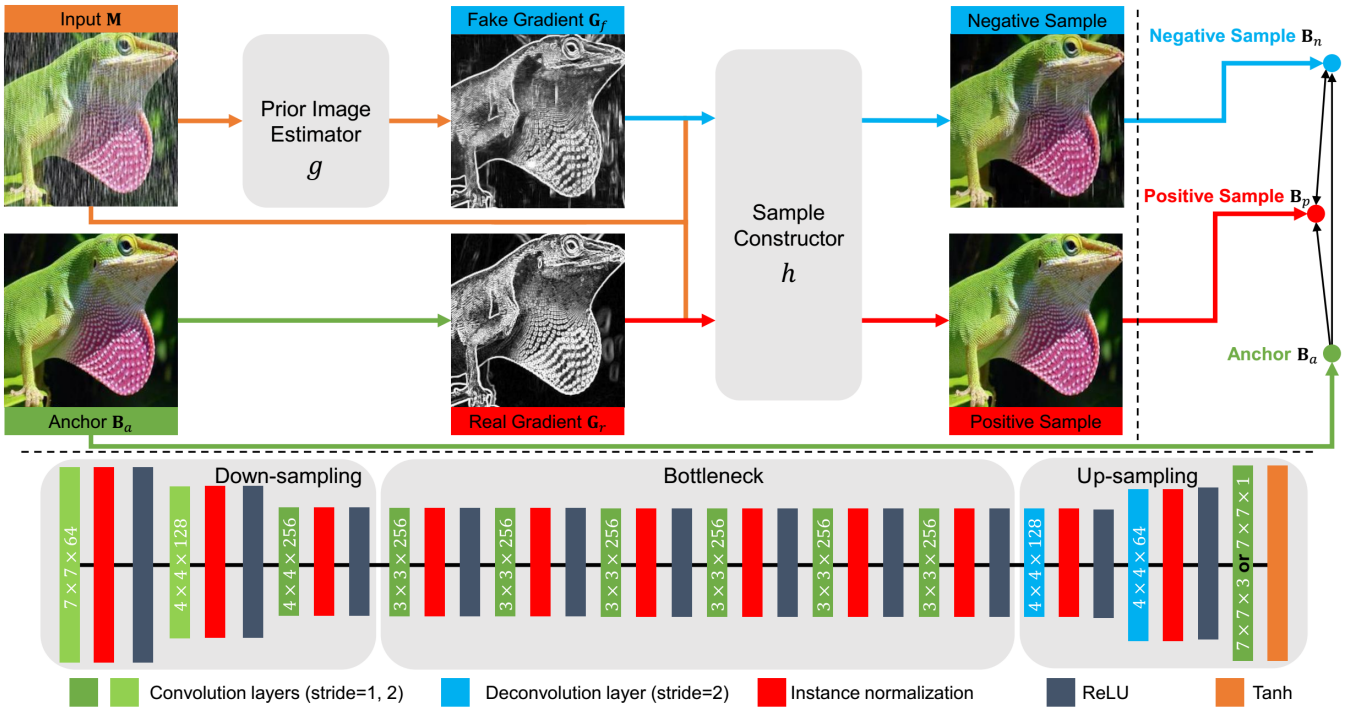

Fig. 1. The framework of the proposed approach is illustrated through the task of rain streak removal as an example (top). Boxes in silver represent two deep neural networks sharing the same architecture which is displayed at the bottom.

2.1. An Adversarial Scheme for Image Denoising

The framework of our adversarial scheme is shown in Figure 1 (top). We define the ground truth of a noise-free image as an *anchor*, the inferred sample guided by a *fake* or estimated prior image as a *negative sample*, and the inferred sample guided by a *real* or ground truth of prior image as *positive sample*. We use g and h to represent the neural networks of the prior image estimator and the sample constructor, $\mathbf{M}$, $\mathbf{B}$, and $\mathbf{G}$ to represent a given noise-contaminated image, a noise-free image, and the prior image or gradient map of $\mathbf{B}$, respectively. Our method optimizes g and h by the following objective function,

$$\begin{aligned} &\min_g \max_h d(\mathbf{B}_a, \mathbf{B}_n) - d(\mathbf{B}_a, \mathbf{B}_p), \text{s.t.}, d(\mathbf{B}_a, \mathbf{B}_p) < \delta, \\ &\mathbf{B}_p = h(\mathbf{M}, \mathbf{G}_r), \mathbf{B}_n = h(\mathbf{M}, \mathbf{G}_f) = h(\mathbf{M}, g(\mathbf{M})), \end{aligned} \tag{1}$$

where subscripts 'a', 'p', 'n', 'r', and 'f' are the abbreviation of 'anchor', 'positive', 'negative', 'real', and 'fake' respectively, δ is a small scalar quantity, d is a pre-defined metric to measure the difference between two images.

Triplet loss. In the stage of optimizing h, our approach maximums the distance between an anchor and a negative sample, while it minimizes that between the anchor and a positive sample (top, Figure 1). This idea is consistent with triplet loss, which has been proven to be effective on the problem of recognition [13]. Note that our approach contains an extra stage (the optimization of g) as compared to triplet loss.

Min-max optimization. The min-max game in Equation (1) attempts to find an optimal h to distinguish whether an inferred sample is guided by a fake gradient or a real gradient. At the same time, it optimizes an optimal g to fool the h. This idea is consistent with the actor-critic algorithms [14] (critic h and actor g in our case), which has been proven to be effective on reinforcement learning. Introducing the Triangle Inequality, we have

$$d(\mathbf{B}_n, \mathbf{B}_p) = \max_h d(\mathbf{B}_a, \mathbf{B}_n) - d(\mathbf{B}_a, \mathbf{B}_p). \tag{2}$$

According to the inequality above, the min-max optimization in (1) is essentially to minimize the difference between two inferred images guided by a fake and a real gradient map,

$$\min_g d(\mathbf{B}_n, \mathbf{B}_p) = \min_g d(h(\mathbf{M}, \mathbf{G}_r), h(\mathbf{M}, \mathbf{G}_f)). \tag{3}$$

That is, optimizer g tries to estimate a fake gradient to be similar as to the real gradient under the metric determined by an optimal h and a pre-defined d. Once the gradient is accurately estimated, the constraint in (1) ensures the accuracy of the inferred sample guided by it.

Differences from GAN. Both our approach and the Generative adversarial nets share the similar spirit of actor-critic optimizations. There are at least two key differences between our approach and GAN. On the application aspect, GAN is employed to generate data and focuses on the output of the actor, while our approach is to remove noise and focuses on the outputs of both actor and critic. On the theoretical aspect,

Algorithm 1: Minibatch stochastic gradient descent training of denoising adversarial networks. The number of steps to apply to the h, k, is a hyperparameter. We used $k = 5$ in our implementation.

for number of training iterations **do**
  for k steps **do**
    Sample minibatch of m data pairs from dataset. $\{(\mathbf{M}^{(1)}, \mathbf{B}_a^{(1)}), ..., (\mathbf{M}^{(m)}, \mathbf{B}_a^{(m)})\}$.
    Update h by ascending its stochastic gradient: $\nabla_{\theta_h} \frac{1}{m} \sum_{i=1}^{m} [d(\mathbf{B}_a, \mathbf{B}_n) - \alpha d(\mathbf{B}_a, \mathbf{B}_p)]$.
  end for
  Update g by descending its stochastic gradient: $\nabla_{\theta_g} \frac{1}{m} \sum_{i=1}^{m} d(\mathbf{B}_a, \mathbf{B}_n)$.
end for

GAN is to minimize the KL-divergence between the distributions of generated data and real data. While our approach, as introduced above, is to minimize the difference between a fake and a real gradient with a given metric, *i.e.*, d (to be introduced below). And such a metric is determined by the critic aim of obtaining an accurate inference of noise-free image.

2.2. Implementation and Algorithm

Constraint. We loose the constraint $d(\mathbf{B}_a, \mathbf{B}_p) < \delta$ in optimization (1) in our implementation so that the optimization (1) is replaced by,

$$\min_g \max_h d(\mathbf{B}_a, \mathbf{B}_n) - \alpha d(\mathbf{B}_a, \mathbf{B}_p), \tag{4}$$

where α is a weight set to 100 in our experiments.

Pre-defined metric d. Metric d is defined differently in different tasks due to their different evaluation metrics (*e.g.*, rain streak removal [10, 15], reflection removal [16, 11]). More specifically, for the problem of rain streak removal, we define $d = \ell_{per} + \lambda_1 \ell_{ssim} + \lambda_2 \ell_{psnr}$, while for the problem of reflection removal, we define $d = \ell_{per} + \lambda_1 \ell_{ssim} + \lambda_1 \ell_{si}$, where ℓ_{per}, ℓ_{ssim}, ℓ_{psnr}, and ℓ_{si} are perceptual loss [17], structural similarity index (SSIM) [18], peak signal-to-noise ratio (PSNR) [19], and structural index (SI) [20] that measures the difference between two images. λ_1 and λ_2 are set to 0.5 and $\frac{1}{60}$ in our implementation.

Network architectures Both networks g and h are composed of two convolutional layers with the stride size of two for down-sampling, six residual blocks [21], and two interpolating convolutional layers with the stride size of two for up-sampling, with instance normalization [22]. The output channels for g and h are 1 and 3 respectively. The detailed architecture is displayed in Figure 1 (bottom).

Our algorithm is summarized in Alg. 1. Both neural networks g and h are trained using Adam solver [23] with $\beta_1 = 0.5$ and $\beta_2 = 0.999$.

3. EXPERIMENTS

In this section, we report our performance compared with state-of-the-art methods for the tasks of rain streak removal and reflection removal.

3.1. Rain Removal

Datasets. We use DIDMDN-DATA [10], which is one of the most recently published datasets, to evaluate our performance for the task of rain streak removal. There are 12000 samples for training and 1200 samples for testing. We set the learning rates for g and h to 0.0002 for the first 50 epochs and decay to 0.0001 for the next 50 epochs.

Comparison methods. We compare our method against four state-of-the-art methods, including deep detail network (FH17) [9], joint rain detection and removal (YT17) [3], density-aware single image de-raining using a multi-stream dense network (ZP18) [10], and non-locally enhanced encoder-decoder network (LH18) [24]. SSIM and PSNR are used for evaluation on the luminance channel (*i.e.*, Y channel of YCbCr space) as suggested by [24].

Table 1. Comparison of quantitative results in terms of SSIM and PSNR on DIDMDN-DATA [10].

Metrics	FH17	YT17	ZP18	LH18	Ours
SSIM	0.7057	0.8763	0.8707	0.9192	**0.9331**
PSNR	23.53	30.35	28.30	33.16	**33.43**

Performance. The quantitative comparison with four state-of-the-art methods is displayed in Table 1. The numbers are the average over all the testing images. Our method consistently achieves the best performance among all the methods. Two examples of visual comparison are shown in Figure 2. A clear recovery can be produced by both LH18 [24] and our method, however, LH18 [24] loses details as compared with ours as shown in the close-up views. Both quantitative and visual results validate the effectiveness of our method.

3.2. Reflection Removal

Datasets. We follow the protocol of several state-of-the-art methods [12, 11, 16] on this topic and perform training on synthetic data and testing on real data. Same mixture model as in [11] is adopted, *i.e.*, $\mathbf{M} = \alpha\mathbf{B} + \beta\mathbf{R}$, to synthesize our training data. The noise-free images are selected from PLACES365 dataset [25][1] and they are mixed with images randomly selected from 5552 reflection images provided by [11]. With α and β being randomly sampled from uniform distributions of $\alpha \sim U(0.7, 1)$, $\beta \sim U(0.1, 0.6)$, we create a total of 19214 data pairs for training. We use a recently published SIR^2 benchmark dataset [4] for evaluation, which contains 499 reflection-contaminated images as well as corresponding reflection-free images. We set the learning rates

[1]Images are from four scenes, OFFICE, PARKING_GARAGE-INDOOR, RESTAURANT_PATIO and STREET, excluding the gray images.

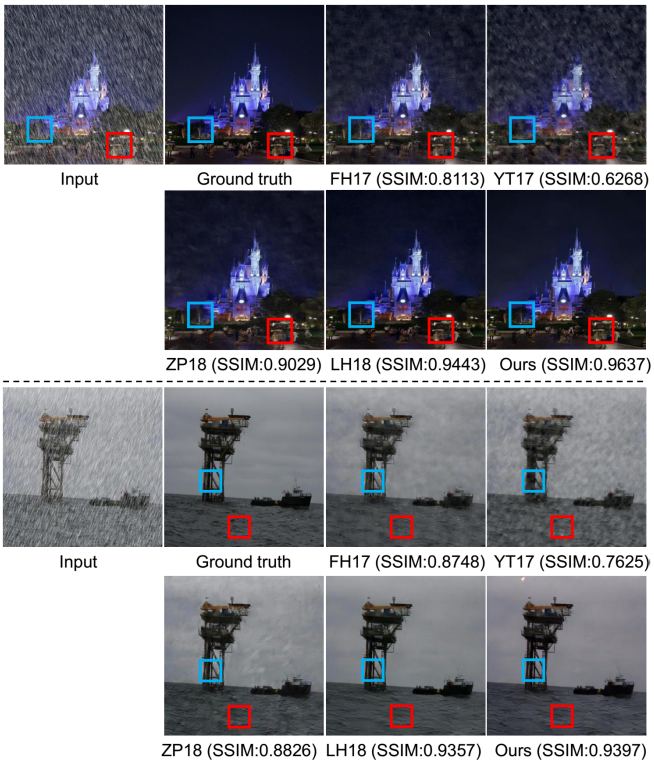

Fig. 2. Visual comparison of rain streak removal results on DIDMDN-DATA [10]. Zoom in for better details.

for both g and h to 0.0002 for the first 100 epochs and decay to 0.0001 for the next 100 epochs in this experiment.

Comparison methods. We compare our method with four state-of-the-art methods, including Laplacian ℓ_0 minimization (AA17) [26], cascaded edge and image learning network (FY17) [12], concurrent reflection removal network (WS18) [11], and reflection separation with perceptual loss (ZN18) [16]. Results from these methods are produced based on the authors' implementations. We learn new models for all the learning-based methods (FY17 [12], WS18 [11], ZN18 [16]) using our training data for fair comparisons. SSIM and SI are used to evaluate our method as previous works [11, 4] did.

Table 2. Comparison of quantitative results in terms of SSIM and SI on SIR^2 [4].

Metrics	AA17	FY17	WS18	ZN18	Ours
SSIM	0.8614	0.8649	0.8907	0.8981	**0.9022**
SI	0.8979	0.8896	0.9160	0.9150	**0.9229**

Performance. The quantitative comparison with four state-of-the-art methods is displayed in Table 2, in which the numbers are the average over all the testing images. Our method consistently achieves the best performance among all the methods. Two examples of visual comparison are shown in Figure 3. As can be observed, all these compared methods fail to remove reflection while our method produces a clear recovery for these examples. Both quantitative and visual results validate the effectiveness of our method.

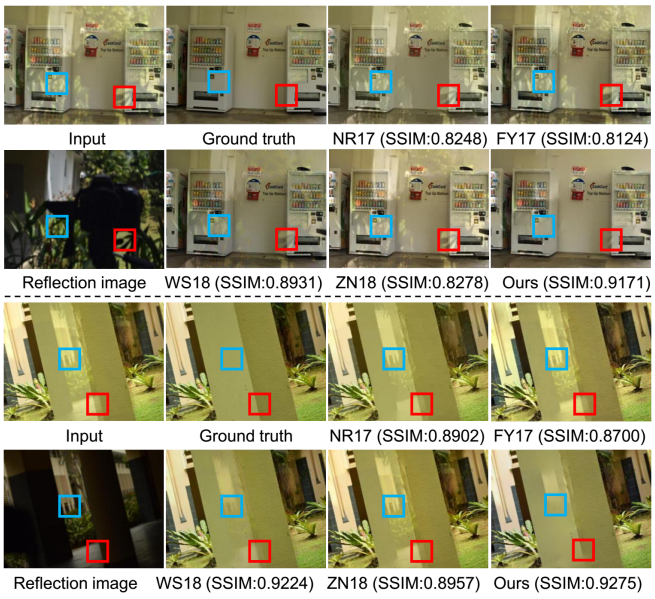

Fig. 3. Visual comparison of reflection removal results on SIR^2 [4]. Zoom in for better details.

4. CONCLUSION

This paper presents a novel adversarial scheme to perform image denoising for the tasks of rain streak removal and reflection removal. The gradient map is used as a prior image to perform a jointly min-max optimization. We show that the proposed adversarial scheme provides very good solutions for both the gradient map and the noise-free image. The superior performances on two public benchmark datasets validate the effectiveness of the proposed approach.

Exploiting other types of prior images in addition to the gradient map (*e.g.*, rain streak image) and applying our approach to other denoising tasks will be our future works.

Acknowledgment

The authors thank Renjie Wan for providing the results of [11] for our experiments. This research was carried out at the Rapid-Rich Object Search (ROSE) Lab at the Nanyang Technological University, Singapore, supported by the National Research Foundation, Singapore, under its Interactive Digital Media (IDM) Strategic Research Programme. This research was also supported in part by Singapore Ministry of Education Academic Research Fund under Grant 2015-T1-002-140, MOE Tier 1 RG 123/15, the National Research Foundation, Prime Minister's Office of Singapore, under Grant NRF2016NRF-NSFC001-098, the National Natural Science Foundation of China under Grant 61872012 and 61661146005, and the NTU-PKU Joint Research Institute through the Ng Teng Fong Charitable Foundation.

5. REFERENCES

[1] Bolun Cai, Xiangmin Xu, Kui Jia, Chunmei Qing, and Dacheng Tao, "Dehazenet: An end-to-end system for single image haze removal," *IEEE Transactions on Image Processing*, vol. 25, no. 11, pp. 5187–5198, 2016.

[2] Rui Qian, Robby T Tan, Wenhan Yang, Jiajun Su, and Jiaying Liu, "Attentive generative adversarial network for raindrop removal from a single image," in *Proc. of CVPR*, 2018.

[3] Wenhan Yang, Robby T Tan, Jiashi Feng, Jiaying Liu, Zongming Guo, and Shuicheng Yan, "Deep joint rain detection and removal from a single image," in *Proc. of CVPR*, 2017.

[4] Renjie Wan, Boxin Shi, Ling-Yu Duan, Ah-Hwee Tan, and Alex C Kot, "Benchmarking single-image reflection removal algorithms," in *Proc. of ICCV*, 2017.

[5] Renjie Wan, Boxin Shi, Ling-Yu Duan, Ah-Hwee Tan, Wen Gao, and Alex C Kot, "Region-aware reflection removal with unified content and gradient priors," *IEEE Transactions on Image Processing*, vol. 27, no. 6, pp. 2927–2941, 2018.

[6] Anat Levin and Yair Weiss, "User assisted separation of reflections from a single image using a sparsity prior," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 9, 2007.

[7] Yu Luo, Yong Xu, and Hui Ji, "Removing rain from a single image via discriminative sparse coding," in *Proc. of ICCV*, 2015.

[8] YiChang Shih, Dilip Krishnan, Fredo Durand, and William T Freeman, "Reflection removal using ghosting cues," in *Proc. of CVPR*, 2015.

[9] Xueyang Fu, Jiabin Huang, Delu Zeng, Yue Huang, Xinghao Ding, and John Paisley, "Removing rain from single images via a deep detail network," in *Proc. of CVPR*, 2017.

[10] He Zhang and Vishal M Patel, "Density-aware single image de-raining using a multi-stream dense network," in *Proc. of CVPR*, 2018.

[11] Renjie Wan, Boxin Shi, Ling-Yu Duan, Ah-Hwee Tan, and Alex C Kot, "Crrn: Multi-scale guided concurrent reflection removal network," in *Proc. of CVPR*, 2018.

[12] Qingnan Fan, Jiaolong Yang, Gang Hua, Baoquan Chen, and David P Wipf, "A generic deep architecture for single image reflection removal and image smoothing.," in *Proc. of ICCV*, 2017.

[13] Florian Schroff, Dmitry Kalenichenko, and James Philbin, "Facenet: A unified embedding for face recognition and clustering," in *Proc. of CVPR*, 2015.

[14] Vijay R Konda and John N Tsitsiklis, "Actor-critic algorithms," in *NeurIPS*, 2000.

[15] Xia Li, Jianlong Wu, Zhouchen Lin, Hong Liu, and Hongbin Zha, "Recurrent squeeze-and-excitation context aggregation net for single image deraining," in *Proc. of ECCV*, 2018.

[16] Xuaner Zhang, Ren Ng, and Qifeng Chen, "Single image reflection separation with perceptual losses," in *Proc. of CVPR*, 2018.

[17] Alexey Dosovitskiy and Thomas Brox, "Generating images with perceptual similarity metrics based on deep networks," in *NeurIPS*, 2016.

[18] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[19] Quan Huynh-Thu and Mohammed Ghanbari, "Scope of validity of psnr in image/video quality assessment," *Electronics letters*, vol. 44, no. 13, pp. 800–801, 2008.

[20] Shao-Hua Sun, Shang-Pu Fan, and Yu-Chiang Frank Wang, "Exploiting image structural similarity for single image rain removal," in *Proc. of ICIP*, 2014.

[21] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Deep residual learning for image recognition," in *Proc. of CVPR*, 2016.

[22] D Ulyanov, A Vedaldi, and V Lempitsky, "Instance normalization: the missing ingredient for fast stylization.," *arXiv preprint arXiv:1607.08022*, 2017.

[23] Diederik P Kingma and Jimmy Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[24] Guanbin Li, Xiang He, Wei Zhang, Huiyou Chang, Le Dong, and Liang Lin, "Non-locally enhanced encoder-decoder network for single image de-raining," in *Proc. of ACM Multimedia*, 2018.

[25] Bolei Zhou, Agata Lapedriza, Aditya Khosla, Aude Oliva, and Antonio Torralba, "Places: A 10 million image database for scene recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 6, pp. 1452–1464, 2018.

[26] Nikolaos Arvanitopoulos, Radhakrishna Achanta, and Sabine Süsstrunk, "Single image reflection suppression," in *Proc. of CVPR*, 2017.