Singe Image Rain Removal with Unpaired Information: A Differentiable Programming Perspective

Hongyuan Zhu,[1] **Xi Peng,**[2] **Joey Tianyi Zhou,**[3*]
Songfan Yang,[4] **Vijay Chanderasekh,**[1] **Liyuan Li,**[1] **Joo-Hwee Lim**[1]
[1]Institute for Infocomm Research, A*STAR, Singapore,
[2]College of Computer Science, Sichuan University, China
[3]Institute of Performance Computing, A*STAR, Singapore
[4]AI Lab, TAL Education Group, China

Abstract

Single image rain-streak removal is an extremely challenging problem due to the presence of non-uniform rain densities in images. Previous works solve this problem using various hand-designed priors or by explicitly mapping synthetic rain to paired clean image in a supervised way. In practice, however, the pre-defined priors are easily violated and the paired training data are hard to collect. To overcome these limitations, in this work, we propose RainRemoval-GAN (RR-GAN), the first end-to-end adversarial model that generates realistic rain-free images using only unpaired supervision. Our approach alleviates the paired training constraints by introducing a physical-model which explicitly learns a recovered images and corresponding rain-streaks from the differentiable programming perspective. The proposed network consists of a novel multiscale attention memory generator and a novel multiscale deeply supervised discriminator. The multiscale attention memory generator uses a memory with attention mechanism to capture the latent rain streaks context at different stages to recover the clean images. The deeply supervised multiscale discriminator imposes constraints at the recovered output in terms of local details and global appearance to the clean image set. Together with the learned rain-streaks, a reconstruction constraint is employed to ensure the appearance consistent with the input image. Experimental results on public benchmark demonstrates our promising performance compared with nine state-of-the-art methods in terms of PSNR, SSIM, visual qualities and running time.

Introduction

Rain, snow, and fog are common visual artifacts which affects many vision-based applications (Zhu, Vial, and Lu 2017; Zhu et al. 2016; 2018a; 2015; Zhu, Weibel, and Lu 2016), such as drone-based video surveillance and self-driving cars. The performance of many computer vision systems often exhibits significant drop when they are presented with images that contain these artifacts. Hence, it is highly practical and expected to develop automatic artifacts removal methods (Li et al. 2017; Zhang et al. 2017b). In this paper, we mainly focus on the problem of rain streak removal from a single image.

*Corresponding author: Joey Tianyi Zhou

(a) Input (b) Ground truth (c) Our result

Figure 1: A visual illustration of single image deraining. The target is to recover a clean image from the input rain image. Our method produces a recovered image with faithful color and structure without using any paired training data.

A rain image I can be modeled as an image composition problem (Luo, Xu, and Ji 2015):

$$I = X + R \tag{1}$$

where X and R denote the desired clean image and the rain streaks, respectively. Single image rain-streak removal aims to estimate the hidden X and R from a given I, which is an unconstrained problem because we need to estimate two unknown variables and there are infinitely solutions if no further regularization are imposed.

To make the rain streak removal problem trackable, existing methods can be roughly grouped into two categories: prior-based and data-driven. The prior-based methods estimate the rain-streak based on various priors or assumptions. Typical priors include but not limited to sparse-coding prior (Luo, Xu, and Ji 2015), low-rank prior (Chang, Yan, and Zhong 2017) and Gaussian prior (Li et al. 2016). Despite remarkable progress achieved by adopting these priors, they are easily violated in practice given the rain-streak does not strictly follow the Gaussian or sparse distribution and the background scenes is cluttered and contains complex illuminations.

In recent, the interest has shifted to data-driven approaches which utilize labeled data with paired rain image and rain streak to learn a deep neural network (Fu et al. 2017; Yang et al. 2017b; Zhang and Patel 2018; Li et al. 2018). Given an input image, the rain streaks or clean image can be regressed with given input rain image. More recently, inspired by the huge success of generative adversarial networks (GAN) in image-to-image translation tasks (Goodfellow et al. 2014; Isola et al. 2017), (Zhang, Sindagi, and Patel 2017) propose a GAN-based image deraining method which employs a generator to map input image to the ground-truth clean image. One major limitation of the method is the necessity of paired training data and it is extremely challenging to collect paired training data given the changing environment.

However, collecting paired training data is extremely challenging given the changing environment. One popular solution is simulating rain in controlled environments. However, most existing simulations are usually too simplified to depict the complexity of real-world. Another common solution is using photo editing tools to add streak to clean natural images with a various rain-density levels with different orientations and scales, but this method limits the scale of training data.

In fact, one could observe that it is easier to collect a large number of rain/rain-free images from Internet despite these images are not in pair. Thus, it is highly excepted to develop derain model which can utilize unpaired supervision. Note that, CycleGAN (Zhu et al. 2017a) recently has become popular to learn cross-style image translation by using bidirectional constrains with adversarial learning. Although CycleGAN has achieved impressive performance in style translation, it is designed for style translation problem and may not preserve the appearance consistency in the translated result.

Based on the above observations, we propose a novel single image deraining method called RainRemoval-GAN (RR-GAN) which is specifically designed based on rain image composition model in Eq.1. The proposed RR-GAN is significantly different from CycleGAN and its variants in structure and application. More specifically, we propose a novel generator which uses an attention memory to captures the latent rain streaks contexts in a recurrent fashion to recover clean image X. Moreover, we propose a novel deeply-supervised multiscale discriminator to regularize the recovered image to look as realistic as possible to the rain-free images. Furthermore, these latent factors are composited together as in Eq.1 to reconstruct the original rain image to preserve faithful color and structure, thus avoiding the inefficient CycleGAN's bi-directional consistency training paradigm.

The major contributions of this work are summarized as follows:

- To the best of our knowledge, **this is one of first works to marriage CycleGAN for single image deraining**, so that makes rain-removal training with unpaired data possible. The novel GAN consists of a novel generator and discriminator which is specifically designed by incorporating the rain image generation model with un-paired training information.
- A novel multi-scale attention memory generator is proposed with an attention memory to fuse the contexts from coarse-scale and fine-scale densely connected network to recurrently learns the rain-streaks to recover the clean image using un-paired training data.
- A novel multiscale deeply-supervised discriminator is proposed to regularize the generated recover image as realistic as possible to the target image in terms of both low-level details and high-level structures.
- Extensive experiments on public benchmark demonstrates our method's efficiency and effectiveness in terms of quantitative and qualitative performance.

Related Work

In this section, we briefly review several recent related works on single image de-raining and generative adversarial networks.

Single Image De-raining

Image deraining has been a classic image restoration problem for years. Some early methods (Zhang et al. 2006; Garg and Nayar 2007; Santhaseelan and Asari 2015; Tripathi and Mukhopadhyay 2014) exploit photometric consistency and temporal dynamics for rain removal. Although these methods achieve promising performance, they are not applicable to the single image setting.

Unlike video-based methods, prior-based methods have been proposed to handle single image deraining problem by assuming the rain-streaks follows certain assumption. There are many hand-crafted priors have been proposed, e.g. sparse coding-based prior (Luo, Xu, and Ji 2015), low-rank prior (Chang, Yan, and Zhong 2017) and gaussian prior (Li et al. 2018), just to name a few. One major limitation of these methods is that the priors are easily violated which results in over-/under-estimation of the rain-streaks (Zhang and Patel 2018).

Recently, deep learning has become popular in both high-level and low-level vision tasks (Cai et al. 2016; Ren et al. 2016; Yang et al. 2017a), several CNN-based methods have also been proposed for image de-raining (Fu et al. 2017; Yang et al. 2017b; Zhang and Patel 2018; Li et al. 2018). In these methods, the idea is to learn a mapping between input rainy images and their corresponding rain-streaks using a CNN structure. (Yang et al. 2017b) design a deep dilated network to joint detect and remove rain streaks. (Li et al. 2018) propose incorporate recurrent neural network to the dilaeted network of Yang *et al.* to preserve multi-stage contexts. (Zhang and Patel 2018) propose to use multi-scale densely connected network trained with additional rain density classifier to predict the rain streaks. In summary, these methods aim to learn a mapping using synthesized rain/clean image pairs which are hard to collect in large scale.

Our methods's generator is similar to (Zhang and Patel 2018) as we combines coarse- and fine-scale network for feature extraction, however we introduce a novel attention memory network to learn the rain-streaks automatically from data in a recurrent fashion. The attention memory network is inspired by recent work in language translation (Gehring et al. 2017) which shows that the performance of machine translation models could be significantly improved by solely using an attention model instead of using additional gate in recurrent networks as in (Li et al. 2018), which significantly improve the inference speed. Moreover, our generator can explicitly learns two disentangled latent parameters which corresponds to clean image and rain streaks from data without paired information, hence significantly reduce the labeling efforts.

Generative Adversarial Networks

Recently generative adversarial networks (Goodfellow et al. 2014; Arjovsky, Chintala, and Bottou 2017; Zhao, Mathieu, and LeCun 2017) have achieved significant progress. The basic idea of GANs is transforming the white noise (or other specified prior) through a parametric model to generate candidate samples with the help of a discriminator and a generator. By optimizing a minimax two-player game, the generator aims to learn the training data distribution, and the discriminator aims to judge that a sample comes from the training data or the generator. An image can be translated into an output by the conditional generator, which has various applications, such as image super resolution (Ledig et al. 2017), text2image (Zhang et al. 2017a), image2image (Yi, Zhang, and Gong 2017). However, conditionalGAN requires paired training images, whose ground-truth can be very difficult to get.

This paper is inspired by the recent popular CycleGAN (Zhu et al. 2017a) which can translate one image from one domain to another domain without paired information. However, our work is remarkably different from CycleGAN. In brief, our architecture is specifically designed for image deraining so that the original image's color and structure could be well preserved. To the best of our knowledge, this could one of the first works to develop specific single image deraining model that incorporates un-paired adversarial learning. Our network is remarkably distinct from CycleGAN in following aspects. First, we proposes a novel attention memory generator using multiscale densely connected networks and attention memory in a recurrent fashion to learn discriminative rain features and latent rain streaks to recover the clean image. Furthermore, our network contains of a novel deeply-supervised multi-scale discriminator training regime, so that the local details to global image appearance is enforced to look realistic to the clean image set. Furthermore, these latent factors are composited together as in Eq.1 to reconstruct the original rain image to preserve faithful color and structure, thus avoiding the inefficient CycleGAN's bi-directional consistency training paradigm.

Differentiable Programming

Our work belongs to the family of differentiable programming which treats a program as neural network such that the program can be parametrized, automatically differentiated and optimizable. The first well-known work of differentiable programming is the Learned ISTA (LISTA) (Gregor and LeCun 2010), which unfolds the popular l_1 solver ISTA as a simple RNN such that the number of layers corresponds to the iteration number and the weight corresponds to dictionary. The LISTA's RNN paradigm have been applied in a wide range of tasks, e.g. hashing (Wang, Ling, and Huang 2016), classification (Wang et al. 2016), sparse coding (Zhou et al. 2018), image dehazing (Zhu et al. 2018b) and etc.

Different from conventional differentiable programming using RNN with difficult-to-interpret variables, our method reformulate the image degradation model using a feed-forward convolutional neural network with prior knowledge. Therefore, our formulation is more interpreatable and efficient then conventional RNN based DP solver.

The Method

As shown in Fig.2, our RR-GAN consists of two networks, namely, a multiscale attention memory generator (MAMG) and a multiscale deeply-supervised discriminator (MDSD) are specifically designed for single image deraining. In brief, MAMG uses an attention memory to recurrently attend to the rain-streak regions for the purpose of recovery of the clean image. MDSD employs a deeply supervision to recover image from the generator by enforcing it look as realistic as possible to the clean image set.

Multiscale Attention Memory Network

The multiscale attention memory network consists of an attentive memory network and a U-Net autoencoder with skip-connection. The former uses a multiscale feature extractor G_f to extract rain-relevant regions features. To achieve better performance, G_f combines the complementary coarse-scale and dense-scale context using recently proposed dense connection (Huang et al. 2017). The fine-scale dense network applies densely connected modular with small kernels to capture fine-scale structures relevant to rain regions with a structure of C(1, 3)-C(3, 3)-C(5, 3)-C(7, 3) , where $C(k_1, k_2)$ denotes the convolution with a filter of size $k_1 \times k_1$, an output channel number of k_2 and a stride of 1 with ReLu output (Krizhevsky, Sutskever, and Hinton 2012). Each layer are densely connected with other layers. The coarse-scale branch applies a similar architecture but with larger kernels to capture longer-range context, whose structure is of C(7, 3)-C(9, 3)-C(11, 3)-C(1, 3). The feature maps X_c and X_f from coarse-scale and fine-scale networks respectively, are concatenated and fed through the attention memory network to learn rain regions.

Visual attention models are adopted to localize relevant regions in an image so that the task-specific features are obtained. In this paper, we introduce visual attention to learn where the rain-regions should be focused on. As shown in our architecture in Fig. 2, we present an attention memory network which is inspired by recent convolutional sequence-to-sequence learning task in machine translation (Gehring et al. 2017). Our attention memory network sequentially uses attention to replace the recurrent networks and the attention map is iteratively refined given the accumulated context. At each time stamp t, our attention memory network concatenates X_c, X_f, M_{t-1}, and S_{t-1}, where X_c is the input from coarse-scale network, X_f is from fine-scale network, M_{t-1} is the rain streak mask at the previous stage $t-1$, and S_t denotes the state memory in previous stage. Note that, the initial M_0 and S_0 are set to zero. The detailed updating rule is as follows:

$$\begin{aligned} a_t &= \sigma(W_1[X_c, X_f, M_{t-1}, S_{t-1}] + b_1) \\ S_t &= a_t \cdot S_{t-1} \\ M_t &= tanh(W_2 * S_t + b_2) \end{aligned} \tag{2}$$

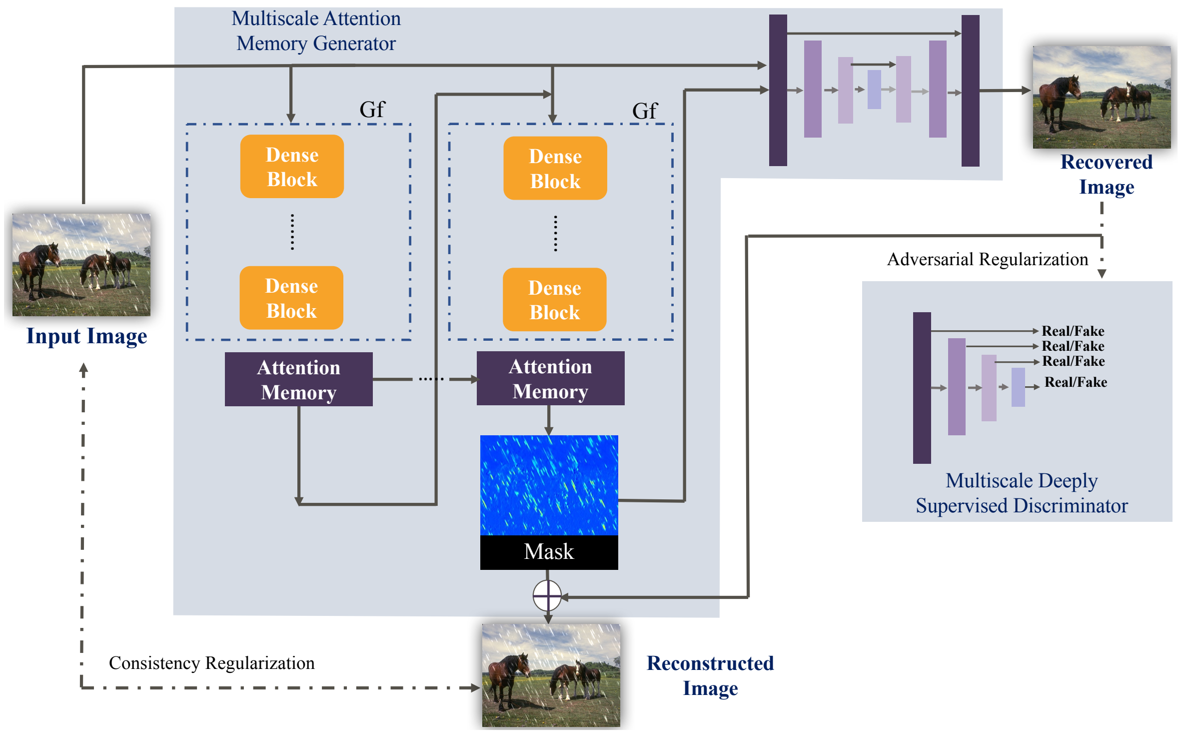

Figure 2: The pipeline of our method. Our RR-GAN consists of a multi-scale attention memory generator and a multiscale deeply-supervised discriminator. The generator uses un-paired rain and clean images to train an attention memory to aggregate the latent rain streak context from different stages. The rain image together with the latent rain-streak imasks are used as an input to the U-Net to generate the recovered image. This recovered image is regularized by a deeply-supervised multiscale discriminator so that its appearance is as close as possible to the clean images used for training in terms of both low-level details and global-level structures.

where W_1 and b_1 are the weight and bias of a convolutional layer with 3×3 kernels of one neuron, σ denotes the sigmoid function $\sigma(x) = \frac{1}{1+exp(-x)}$. W_2 and b_2 are the weight and bias of a deconvolutional layer with 3×3 kernels of one neuron, which are used to output the rain-streak mask. In Fig.2, one can observe that the rain mask is recurrently improved by using the contexts from the previous stages. The input image together the final attention map are concatenated to feed into the U-Net auto-encoder for generating the rain-free image. Our deep autoencoder includes 16 convolutional (relu) blocks, which adopts skip connections to prevent blurred outputs. To show the effectiveness of our method, Fig. 3 shows the visual results of different generators, the one which combines attention memory and skip U-Net auto-encoder yielded best visual result.

Multiscale Deeply-Supervised Discriminator

The function of discriminator is to regularize the output from the generator so that the generated image looks as realistic as possible to the target set clean images. Recently, a shallowly-supervised discriminator with discriminator with supervision at the last layer has become popular in recent popular gan architecrtures (e.g CycleGAN (Zhu et al. 2017b) and ConditionalGAN (Isola et al. 2017)) which achieves good image translation performance. However, such shallowly-supervised discriminator overlooks the low-level details and global structures which are useful for image enhancement tasks.

To overcome the above disadvantage, we propose a novel deeply-supervised discriminator which consists of four convolutional layers, which is with the structure of C(3, 64)-C(3, 128)-C(3, 256)-C(3, 512), where each convolutional layer has a stride of two. Each convolutional feature map will be passed through an instance normalization layer with Leaky-ReLu activations and then be fed into the next convolutional layer. Moreover, each side-output follows by an additional convolution layer of C(1, 1) with sigmoid to output the probability of each patch to the clean images. Hence we will have four predictions which provide multiple level supervision to provide regularization at to make the generated image as realistic as the ground-truth image in terms of low-level details and high-level structures.

Objective Function

To stabilize the training of discriminator D, we minimize an objective function L_d that consists of two terms $\mathcal{L}_{D_{real}}$ and

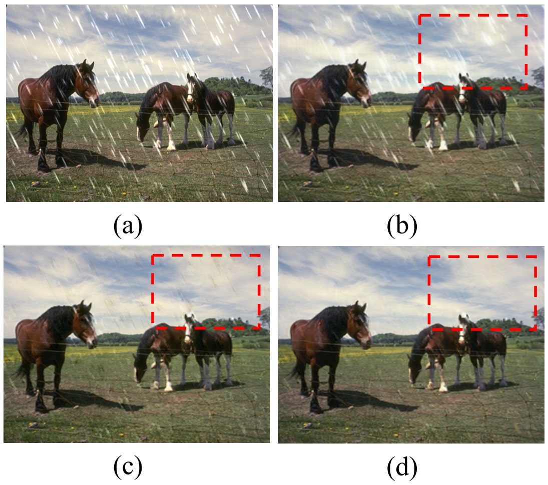

Figure 3: A visual comparison on the effectiveness of different generators. (a) Input; (b) the result of using context auto-encoder alone; (c) the result of our generator without using attention memory; (d) the result of our multiscale attention memory generator.

$\mathcal{L}_{D_{fake}}$ as in (Mao et al. 2017):

$$\begin{aligned}\mathcal{L}_d(G, D) =& \mathcal{L}_{D_{real}} + \mathcal{L}_{D_{fake}} \\ =& \frac{1}{2K}\sum_{i=1}^{K}\{\mathbb{E}_y(1 - D_i(x, y))^2 + \\ & \mathbb{E}_x(D_i(G(x))^2\}\end{aligned} \tag{3}$$

which aims to train the discriminator D_i so that it can try to differentiate the fake examples $G(x)$ from generator G and the real clean image y, where $\mathbb{E}_x$ is the expectation of x and $D_{i\in\{1,2,\ldots,K\}}$ is the classifier at ith level side-output.

On the other hand, when we train the generator G, we optimize an objective function L_g which consists of a MSE loss L_r (Ren et al. 2016) and an adversarial learning loss L_{adv}:

$$\mathcal{L}_g = \mathcal{L}_r + \gamma\mathcal{L}_{adv} \tag{4}$$

These three terms are designed for minimizing the reconstruction error and improving the perceptual quality, respectively. γ is a trade-off factors set according to cross-validation.

The **MSE Loss** L_m encourages the network to enforce the appearance consistency between the recovered image X and input image I_l using the estimated rain mask R by minimizing the discrepancy between the composite image $I_t = X + R$ and the input image I_l:

$$\mathcal{L}_r = \frac{1}{C \times W \times H}\sum_{i=1}^{W}\sum_{j=1}^{H}\sum_{c=1}^{C}\|I_t^{i,j,c} - I_l^{i,j,c}\|_2 \tag{5}$$

The W, H, and C are the width, height, and channel number of the input image I_t.

The **Adversarial Loss** $\mathcal{L}_{adv}$ encourages the generator G to recover image $G(x)$ as realistic as the ground-truth image y by assigning a real label 1 to recovered example $G(x)$:

$$\mathcal{L}_{adv}(G, D) = \frac{1}{K}\sum_{i=1}^{K}\mathbb{E}_x(1 - D_i(G(x)))^2 \tag{6}$$

where K is the layer number of discriminator, D_i is i-th level classifier to classify whether an image is from the clean image or generator $G(x)$. The term penalizes the details of recovered image that looks different from clean images.

Training Details

During training, a 286×286 image is randomly cropped from the input image (or its horizontal flip) of size 256×256. Adam is used as optimization algorithm with a mini-batch size of 1. The learning rate starts from 0.001. The models are trained for up to 10 epochs to ensure convergence. We use a weight decay of 0.0001 and a momentum of 0.9. The entire network is trained using the Pytorch framework. During training, we set $\gamma = 1$. All the parameters are defined via cross-validation using the validation set.

Experiments

We evaluate our method on public benchmark. We quantitatively evaluate the rain-removal performance using two commonly used metrics, including peak signal to noise ratio (PSNR) (Huynh-Thu and Ghanbari 2008) and structure similarity index (SSIM) (Wang et al. 2004) for evaluation. We also provide ablation study of each component proposed to demonstrate their effectiveness.

We use **Rain800** (Zhang and Patel 2018) for benchmarking. The **Rain800** dataset contains 700 synthesized images for training and 100 images for testing using randomly sampled outdoor images.

Comparing Methods We compare our proposed approach with 7 state-of-the-art methods, including image decomposition (ID) (Kang, Lin, and Fu 2012), discriminative sparse coding (DSC) (Luo, Xu, and Ji 2015), layer priors and (LP) (Li et al. 2016), DetailsNet (Fu et al. 2017), and joint rain detection and removal (JORDER) (Yang et al. 2017b) and RESCAN (Li et al. 2018).

Results on Synthetic Datasets Table 1 shows quantitative results on the **Rain800**. One can observe that our RR-GAN considerably achieves promising results in terms of both PSNR and SSIM on Rain800.

From Table 1, data-driven methods, especially the deep learning based methods (Fu et al. 2017; Yang et al. 2017b; Zhang and Patel 2018; Li et al. 2018) significantly outperform the prior-based methods (Kang, Lin, and Fu 2012; Luo, Xu, and Ji 2015; Li et al. 2016). The comparison demonstrates that the features learned by the deep neural networks is much helpful for the rain-removal task.

On the one hand, the proposed RR-GAN is trained with unpaired training data, which outperform the JORDER method on the Rain800 dataset. Note that, JORDER is trained using the paired supervised data. In other words, it utilizes a stronger supervisor than our method does. The result demonstrates that it is feasible and promising to develop deraining model using unpaired data which could be easily accessed in practice.

In terms of running time, our method is also much faster than other deep learning methods and achieve a running time of 0.03s per image, which demonstrates our method's real-time performance.

Dataset	R800		Time (s)
Measure	PSNR	SSIM	
ID	18.88	0.5832	120s
DSC	18.56	0.5996	189.3s
LP	20.46	0.7297	674.8s
DetailNet	21.16	0.7320	0.3s
JORDER	22.24	0.7763	1.5s
RESCAN	**24.09**	**0.8410**	0.4s
RR-GAN	<u>23.51</u>	<u>0.7566</u>	<u>0.03s</u>

Table 1: Quantitative experiments on Rain800. Best results are marked in bold and the second best results are underlined.

Analysis of RR-GAN To demonstrate the effectiveness of our generator used in RR-GAN, we compare it with other network architectures. The first baseline is dense connection. To be specific, we replace the dense block with residual block (ResNet) and fix other blocks in our model. One could find that the performance was dropped significantly from 23.51/0.7566 (PSNR/SSIM) to 22.82/0.6991.

Besides the ablation study in generator, we also test the effectiveness of our deeply supervised multiscale discriminator. To the end, we adopt a new discriminator which only enforces the supervision at the last layer (shallowly-supervised discriminator). One could see that the performance is dropped to 21.43/0.7254.

Component	Baseline	PSNR	SSIM
Generator	ResNet	22.82	0.6991
Discriminator	Shallow-supervision	21.43	0.7254
Our Method	RR-GAN	**23.51**	**0.7566**

Table 2: Ablation study on RR-GAN.

Analysis of Training Paradigm We analyze how the presence of corresponding rain/rain-free scenes in the training samples can aeffect the model training. We first randomly split the Rain100H dataset into two halves (split 1 and 2). We use different combinations of the images for model training, and then test the model on the rain images in split 2. Specifically, we train under 2 settings: the rain and clean image are paired from split1 (setting 1), the clean images are randomly drawn from split 1 (setting 2), whose testing performance on split 2's rain images can be observed in Table.3.

From the table, one can observe that the best deraining results can be obtain when both corresponding rain and rain-free images are used for training. However, other settings can still achieve comparable perfor- mance. This implies that it is not necessary to have paired rain and rain-free scenes during training (see setting 2).

Setting	Pair(P)/ UnPair(U)	PSNR	SSIM
1	P	23.79	0.7951
2	U	23.51	0.7566

Table 3: Analysis of Training Paradigm.

Analysis of Loss Function To better demonstrate the effectiveness of our objective function, we conduct an ablation study by considering the combinations of the proposed MSE loss $\mathcal{L}_r$ and the adversarial loss $\mathcal{L}_{adv}$. Figure 5 and Table 4 demonstrate qualitative and quantitative results on an sample image, respectively. One can observe that by using MSE loss alone without regularization from the adversarial loss, the quality of the recover image is very poor due to that there is not sufficient infomation to allow generator to learn what makes the clean image. Using adversarial loss alone can significantly improve the visual quality as the information from clean image set help the generator lean use cues to recover clean image, however the tone recovered image is bit yellowish as there is no consistency constraint between the recover image and the input image. By combining these two terms together, our method can produce recovered images with realistic color and structures.

Metrics	$\mathcal{L}_r$	$\mathcal{L}_{adv}$	$\mathcal{L}_r + \mathcal{L}_{adv}$
PSNR	17.08	21.33	23.51
SSIM	0.3760	0.7328	0.7566

Table 4: Quantitative studies on different losses.

Conclusion

This paper proposed a novel RR-GAN for end-to-end single image deraining without using paired training data. The proposed method includes a novel multiscale attention memory network and a novel deeply supervied multiscale discriminator. The attention memory network uses a state memory to learn a latent rain-streak mask in a recurrent fashion to aggreate the rain context information from diffrent stages. Together with the input image, the generator will try to generate the recovered image by iusing un-paired rain and clean images training data. The proposed deeply supervised multiscale discriminator can effectively regularize the output from the generator to look as realistic as possible to the clean image sets in terms of local details and global structures. Extensive experiments have demonstrated the promising performance of our method in terms of PSNR, SSIM, running time and visual quality.

Acknowledgements

This was funded by the Deep learning 2.0. program at I2R, A*STAR. The work of X. Peng was supported by the Fundamental Research Funds for the Central Universities under Grant YJ201748, by NFSC under Grant 61806135, 61432012 and Grant U1435213, and by the Fund of Sichuan University—Tomorrow Advancing Life.

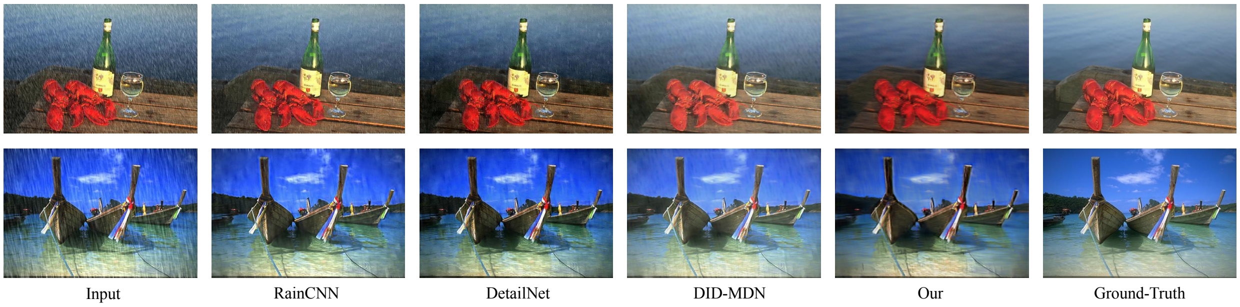

Figure 4: Qualitative comparison with the state-of-the-arts on two randomly sampled Rain800.

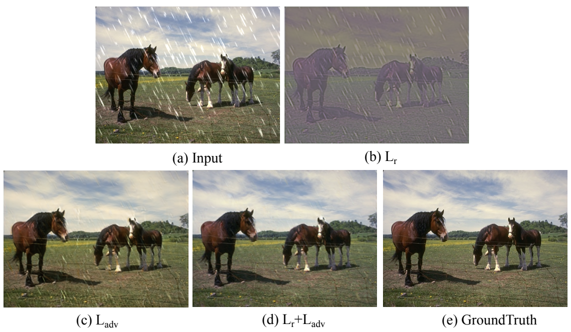

Figure 5: Qualitative studies on different loss.

References

Arjovsky, M.; Chintala, S.; and Bottou, L. 2017. Wasserstein GAN. In *International Conference on Machine Learning*.

Cai, B.; Xu, X.; Jia, K.; Qing, C.; and Tao, D. 2016. Dehazenet: An end-to-end system for single image haze removal. *IEEE Transactions on Image Processing* 25(11):5187–5198.

Chang, Y.; Yan, L.; and Zhong, S. 2017. Transformed low-rank model for line pattern noise removal. In *ICCV*.

Fu, X.; Huang, J.; Zeng, D.; Huang, Y.; Ding, X.; and Paisley, J. 2017. Removing rain from single images via a deep detail network. In *CVPR*.

Garg, K., and Nayar, S. K. 2007. Vision and rain. *IJCV* 75(1):3–27.

Gehring, J.; Auli, M.; Grangier, D.; Yarats, D.; and Dauphin, Y. N. 2017. Convolutional sequence to sequence learning. In *ICML*.

Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; and Bengio, Y. 2014. Generative Adversarial Nets. In *Advances in Neural Information Processing Systems*, 2672–2680.

Gregor, K., and LeCun, Y. 2010. Learning fast approximations of sparse coding. In *ICML*, 399–406.

Huang, G.; Liu, Z.; van der Maaten, L.; and Weinberger, K. Q. 2017. Densely connected convolutional networks. In *CVPR*.

Huynh-Thu, Q., and Ghanbari, M. 2008. Scope of validity of psnr in image/video quality assessment. *Electronics Letters* 44(13):800–801.

Isola, P.; Zhu, J.; Zhou, T.; and Efros, A. A. 2017. Image-to-image translation with conditional adversarial networks. In *CVPR*, 5967–5976.

Kang, L.; Lin, C.; and Fu, Y. 2012. Automatic single-image-based rain streaks removal via image decomposition. *IEEE TIP* 21(4):1742–1755.

Krizhevsky, A.; Sutskever, I.; and Hinton, G. E. 2012. Imagenet classification with deep convolutional neural networks. In *NIPS*, 1106–1114.

Ledig, C.; Theis, L.; Huszár, F.; Caballero, J.; Cunningham, A.; Acosta, A.; Aitken, A.; Tejani, A.; Totz, J.; Wang, Z.; and Shi, W. 2017. Photo-realistic single image super-resolution using a generative adversarial network. In *IEEE Conference on Computer Vision and Pattern Recognition*, 105–114.

Li, Y.; Tan, R. T.; Guo, X.; Lu, J.; and Brown, M. S. 2016. Rain streak removal using layer priors. In *CVPR*.

Li, B.; Peng, X.; Wang, Z.; Xu, J.; and Feng, D. 2017. An all-in-one network for dehazing and beyond. *CoRR* abs/1707.06543.

Li, X.; Wu, J.; Lin, Z.; Liu, H.; and Zha, H. 2018. Recurrent squeeze-and-excitation context aggregation net for single image deraining. In *ECCV*.

Luo, Y.; Xu, Y.; and Ji, H. 2015. Removing rain from a single image via discriminative sparse coding. In *ICCV*.

Mao, X.; Li, Q.; Xie, H.; Lau, R. Y. K.; Wang, Z.; and Smolley, S. P. 2017. Least squares generative adversarial networks. In *ICCV*.

Ren, W.; Liu, S.; Zhang, H.; Pan, J.; Cao, X.; and Yang, M.-H. 2016. Single image dehazing via multi-scale convolutional neural networks. In Leibe, B.; Matas, J.; Sebe, N.; and Welling, M., eds., *ECCV*.

Santhaseelan, V., and Asari, V. K. 2015. Utilizing local phase information to remove rain from video. *IJCV* 112(1):71–89.

Tripathi, A. K., and Mukhopadhyay, S. 2014. Removal of rain from videos: a review. *Signal, Image and Video Processing* 8(8):1421–1430.

Wang, Z.; Bovik, A. C.; Sheikh, H. R.; and Simoncelli, E. P.

2004. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing* 13(4):600–612.

Wang, Z.; Yang, Y.; Chang, S.; Ling, Q.; and Huang, T. S. 2016. Learning A deep l_∞ encoder for hashing. In *IJCAI*, 2174–2180.

Wang, Z.; Ling, Q.; and Huang, T. S. 2016. Learning deep l_0 encoders. In *AAAI*, 2194–2200.

Yang, H.; Pan, J.; Yan, Q.; Sun, W.; Ren, J.; and Tai, Y.-W. 2017a. Image dehazing using bilinear composition loss function. *arXiv preprint arXiv:1710.00279*.

Yang, W.; Tan, R. T.; Feng, J.; Liu, J.; Guo, Z.; and Yan, S. 2017b. Deep joint rain detection and removal from a single image. In *CVPR*.

Yi, Z.; Zhang, H.; and Gong, P. T. M. 2017. DualGAN: Unsupervised Dual Learning for Image-to-Image Translation. In *IEEE International Conference on Computer Vision*, 502–510.

Zhang, H., and Patel, V. M. 2018. Density-aware single image de-raining using a multi-stream dense network. In *CVPR*.

Zhang, X.; Li, H.; Qi, Y.; Leow, W. K.; and Ng, T. K. 2006. Rain removal in video by combining temporal and chromatic properties. In *ICME*.

Zhang, H.; Xu, T.; Li, H.; Zhang, S.; Huang, X.; Wang, X.; and Metaxas, D. N. 2017a. StackGAN - Text to Photo-realistic Image Synthesis with Stacked Generative Adversarial Networks. In *IEEE International Conference on Computer Vision*.

Zhang, K.; Zuo, W.; Chen, Y.; Meng, D.; and Zhang, L. 2017b. Beyond a gaussian denoiser: Residual learning of deep CNN for image denoising. *IEEE Trans. Image Processing* 26(7):3142–3155.

Zhang, H.; Sindagi, V.; and Patel, V. M. 2017. Image deraining using a conditional generative adversarial network. *CoRR* abs/1701.05957.

Zhao, J.; Mathieu, M.; and LeCun, Y. 2017. Energy-based Generative Adversarial Network. In *International Conference on Learning Representations*.

Zhou, J. T.; Di, K.; Du, J.; Peng, X.; Yang, H.; Pan, S. J.; Tsang, I.; Liu, Y.; Qin, Z.; and Goh, R. S. M. 2018. Sc2net: Sparse lstms for sparse coding. In *AAAI*.

Zhu, H.; Lu, S.; Cai, J.; and Lee, G. 2015. Diagnosing state-of-the-art object proposal methods. In *Proceedings of the British Machine Vision Conference 2015, BMVC 2015, Swansea, UK, September 7-10, 2015*, 11.1–11.12.

Zhu, H.; Lu, J.; Cai, J.; Zheng, J.; Lu, S.; and Magnenat-Thalmann, N. 2016. Multiple human identification and cosegmentation: A human-oriented CRF approach with poselets. *IEEE Trans. Multimedia* 18(8):1516–1530.

Zhu, J.; Park, T.; Isola, P.; and Efros, A. A. 2017a. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *ICCV*.

Zhu, J.; Park, T.; Isola, P.; and Efros, A. A. 2017b. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *ICCV*.

Zhu, H.; Vial, R.; Lu, S.; Peng, X.; Fu, H.; Tian, Y.; and Cao, X. 2018a. Yotube: Searching action proposal via recurrent and static regression networks. *IEEE Transactions on Image Processing* 27(6):2609–2622.

Zhu, H.; Peng, X.; Chandrasekhar, V.; Li, L.; and Lim, J. 2018b. Dehazegan: When image dehazing meets differential programming. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI 2018, July 13-19, 2018, Stockholm, Sweden.*, 1234–1240.

Zhu, H.; Vial, R.; and Lu, S. 2017. TORNADO: A spatio-temporal convolutional regression network for video action proposal. In *IEEE International Conference on Computer Vision, ICCV 2017, Venice, Italy, October 22-29, 2017*, 5814–5822.

Zhu, H.; Weibel, J.; and Lu, S. 2016. Discriminative multi-modal feature fusion for RGBD indoor scene recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016*, 2969–2976.