RainGAN: Unsupervised Raindrop Removal via Decomposition and Composition

Xu Yan
Nanyang Technological University
xuya0012@ntu.edu.sg

Yuan Ren Loke
Nanyang Technological University
yrloke@ntu.edu.sg

Abstract

Adherent raindrops on windshield or camera lens may distort and occlude vision, causing issues for downstream machine vision perception. Most of the existing raindrop removal methods focus on learning the mapping from a raindrop image to its clean content by the paired raindrop-clean images. However, the paired real-world data is difficult to collect in practice. This paper presents a novel framework for raindrop removal that eliminates the need for paired training samples. Based on the assumption that a raindrop image is a composition of a clean image and raindrop style, the proposed framework decomposes a raindrop image into a clean content image and a raindrop-style latent code. Inversely, it composes a clean content image and a raindrop style code to a raindrop image for data augmentation. The proposed framework introduces a domain-invariant residual block to facilitate the identity mapping for the clean portion of the raindrop image. Extensive experiments on real-world raindrop datasets show that our network can achieve superior performance in raindrop removal to other unpaired image-to-image translation methods, even with comparable performance with state-of-the-art methods that require paired data.

1. Introduction

Adverse weather such as rain poses a challenge for outdoor computer vision tasks. Adherent raindrops on windshield or camera lens usually distort and occlude a portion of scene, leading to degraded performance of downstream computer vision applications including self-driving cars and outdoor surveillance cameras. Therefore, it is essential to restore a clear scene first.

Many CNN-based autoencoder methods [4, 31, 15, 33, 30, 11, 6, 7, 17, 18, 32, 36, 37, 34] learns the mapping from rain images to clean images, have achieved satisfactory results on synthetic datasets Rain100H [37], DID-MDN [36] and Rain800 [32]. They mainly focus on the rain streak removal and ignore the fact that visibility is distorted or occluded by raindrops mostly (see Figure 2) instead of by the rain streak. Very few synthetic datasets such as Hao *et al*. [8] and RainCityscapes++ [28] blend the synthetic raindrop into the images with handcrafted functions. However, they are still far from the real images.

Figure 1. Without requiring paired training data, our proposed network can learn to achieve superior raindrop removal as illustrated: For the raindrop images in the left column, the images in the right column show the corresponding recovery by our decomposition generator.

Qian *et al*. [26] collected a real-world well-aligned raindrop-clean image pairs. [26, 27, 8, 29, 25, 19, 24, 1] have shown promising results on the dataset using paired training. However, this dataset is very hard to collect and has many limitations. The raindrop image and its clean image are taken at different times, so any dynamic objects and ambient light change will make the paired data not aligned. Hence, paired training methods rely on this kind of dataset can hardly generalize to driving scene image. Since it's not practical to collect a real-world driving scene raindrop-clean paired dataset, we can only leverage the unpaired dataset.

An unpaired dataset consists of two or more collections of images in different domains. Images in each domain collection do not have the exact counterparts in other domain collections. Practically we can collect real-world unpaired raindrop images and clean images. Specifically, the clean images can be taken right before and after the rain. They also can be taken right after each wipe of the windshield. Recently, generic domain transfer methods [20, 39, 13, 2, 3, 38] achieved very promising results in image-to-image translation with unpaired data. Inspired by them, [12, 21, 5] were proposed to restore clean images from a shadow, blurring, and Gaussian noise respectively with only unpaired images. However, to our best knowledge, no existing method dedicates to remove the adherent raindrops effectively with only unpaired data. Directly applying the generic domain transfer and the domain-specific image-to-image transfer methods on raindrop removal do not achieve a satisfactory result.

In this paper, we observe the fact that a clean scene can blend with infinite many kinds of raindrop styles to form infinite raindrop images. All those raindrop images share one scene. In other words, a clean image can be composed of many raindrop styles to generate many different raindrop images. The raindrop images can also be decomposed to the common scene and different raindrop styles. Hence, we formulate the raindrop removal as a many-to-one image-to-image translation problem. Moreover, we generate new realistic raindrop images by composing clean scenes and raindrop styles. The mixture of raindrop style and the clean scene is a very complicated function, which a hand-crafted function can hardly represent. Our strategy is to use an autoencoder to directly learn the decomposition function from a raindrop image to a clean scene and a raindrop style latent code while another autoencoder learns the composition function from a clean scene and a raindrop style to a raindrop image. We represent the raindrop style as a latent code that is encoded or decoded by the proposed autoencoders.

Unlike generic style transfer methods, where the entire image is modified, the raindrop removal only removes the raindrop from the input image and keeps its clean portion unchanged. Residual blocks are introduced to the autoencoders of the GANs to facilitate identity mapping for the clean portion. Rain can appear and disappear at any time randomly. The model should be domain invariant to handle the input image with and without raindrops. Thus, we enforce the output to be identical to the input when the input image is clean. We conduct extensive experiments on two real-world raindrop datasets. The results show that our method achieves comparable performance with state-of-the-art paired methods and outperforms other unpaired image-to-image translation methods. In summary, our contributions are:

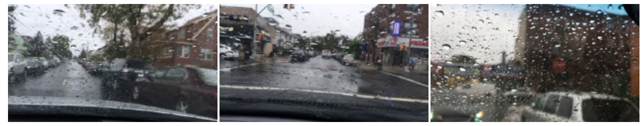


Figure 2. When driving under the rainy condition, raindrop is the major factor affects the visibility and degrades the computer vision applications performance.

- We propose an unpaired training framework, RainGAN. It is the first work that formulates the raindrop removal problem as a many-to-one image-to-image translation problem. It leverages unpaired real-world images, which makes it the first approach to be able to remove the real-world raindrop effectively.
- Through the composition of different permutations of clean scene and raindrop style, we synthesize the realistic raindrop images to further improve the performance of the raindrop removal tremendously.
- We introduce a residual block to the autoencoders on restoring the raindrop portion of the images.

2. Related Work

Rain removal problems have been studied extensively for decades. Most of works [4, 31, 15, 33, 30] focus on rain streak removal. Due to the lack of paired raindrop-clean real-world images, very few adherent raindrop removal methods have been proposed. Traditional method [35] models adherent raindrops using the law of physics and detects raindrops based on these models in combination with intensity derivatives of the input image. A hand-crafted feature is hard to generalize well on real-world data.

Raindrop removal methods. Qian *et al*. [26] collected a raindrop-clean real-world dataset and proposed Attentive-Recurrent Network, which uses LSTM [10] to learn the raindrop mask to aid the raindrop removal in GAN. [27] integrate an edge map as an attention map to the autoencoder. [1] proposed to augment Qian *et al*. [26] dataset with screen-space refraction. The augmented data is trained with VGG perceptual loss [16] and GAN loss. Although their results are quite promising, they require well-aligned paired data for training. Collecting well-aligned scenes requires the scenes to be static. Hence, the model learned from such a dataset, hard to generalize to the real-world driving scene.

Unpaired image domain transfer. Unpaired image domain transfer methods [39, 20, 13, 38, 3, 2] leverage unpaired data when the paired data is not available. CycleGAN [39] and UNIT [20] are one-to-one translation methods. They use the cycle-consistency loss to enforce the generator to keep the content information while transferring the

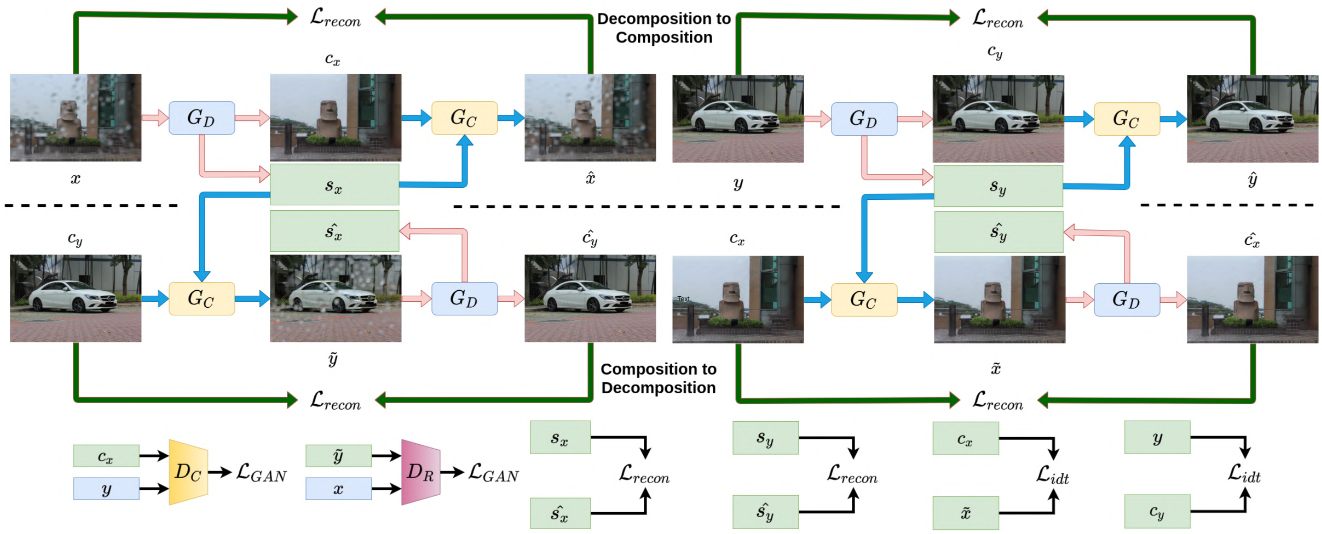


Figure 3. Illustration of the our framework. Decomposition-to-Composition (D2C) and Composition-to-Decomposition (C2D) are showd on the top and bottom of the figure, respectively. a raindrop image x and clean image y are first passed through D2C to obtain their clean content c_x and c_y, style code s_x and s_y, and reconstructed image $\hat{x}$ and $\hat{y}$. Then the style codes are swapped. c_y, s_x, c_x and s_y are passed through C2D to generates composed image $\tilde{y}$ and $\tilde{x}$, reconstructed content image $\hat{c_y}$ and $\hat{c_x}$, and style code $\hat{s}_x$ and $\hat{s}_y$. Two discriminators D_C and D_R learns to distinguish c_x from y and $\hat{y}$ from x. G_D and G_C generate realistic clean and raindrop image respectively to fool D_C and D_R.

style. The one-to-one relationship between images in two domains forces the generator to encode domain-specific information into the transferred image so it can be mapped back to the original image. The encoded domain-specific information harms the translation quality. Unpaired many-to-many image-to-image translation methods [13, 38, 3, 2] learns the content and style information separately. The same image in one domain can transfer to many other styles in another domain by injecting different style information. However, no style information is needed to transfer a raindrop image to a clean image. On top of MUNIT [13], Mask-ShadowGAN [12] uses a binary shadow mask to guide the shadow removal. DRNet [21] introduces perceptual loss and KL loss to remove the blurring. LIR [5] remove Gaussian noise with Background Consistency Module. They work well under their task, but their losses are not optimal for raindrop removal.

3. Proposed Method

We consider a raindrop image x composes of a clean image c_x and a raindrop style latent code s_x. The decomposition and composition functions can be expressed as:

$$c_x, s_x = G_D(x) \tag{1}$$

and

$$x = G_C(c_x, s_x), \tag{2}$$

where G_D and G_C are the decomposition generator and composition generator respectively. They are implemented as autoencoders. Our goal is to train G_D to learn the mapping function from raindrop images to clean images while keeping scenes unchanged. We make a few assumptions to achieve the goal. A raindrop image is decomposed into a clean image and a raindrop-style latent code. The same raindrop image should be composed back by the clean image and latent code, so all the raindrop image information is stored in the clean image and the raindrop style latent code. When the clean image is composed with another raindrop style latent code, the corresponding raindrop style should be transferred to the clean image. To fulfill the assumptions, RainGAN is devised to have two pipelines: Decomposition-to-Composition (D2C) and Composition-to-Decomposition (C2D). We randomly sample one raindrop image x and one clean image y from each domain and pass them through D2C and followed by C2D for each iteration of the training process. Figure 3 illustrates our framework.

3.1. Decomposition-to-Composition

G_D decomposes a raindrop image x to a clean image c_x and a raindrop style latent code s_x. G_C then takes c_x and s_x to generate $\hat{x}$. Similarly a clean image y is fed to the same pipeline to generate c_y, s_y and $\hat{y}$. In order to remove the raindrop, we apply LSGAN [23] adversarial loss to G_D

and D_C as:

$$\mathcal{L}_{\text{GAN}}(D_C) = \frac{1}{2}\mathbb{E}_{y\sim p_{data}(y)}[(D_C(y)-1)^2] + \frac{1}{2}\mathbb{E}_{c_x\sim p_{data}(c_x)}[(D_C(c_x))^2] \tag{3}$$

and

$$\mathcal{L}_{\text{GAN}}(G_D) = \frac{1}{2}\mathbb{E}_{c_x\sim p_{data}(c_x)}[(D_C(c_x)-1)^2], \tag{4}$$

where $p_{data}(c_x)$ is the clean image distribution generated by G_D and D_C is a clean image discriminator. By minimising $\mathcal{L}_{\text{GAN}}(D_C)$, D_C tries distinguish between a generated clean image c_x and a real clean image y. While minimising $\mathcal{L}_{\text{GAN}}(D_C)$ to enforce G_D to generate clean images that looks real. Reconstruction loss $\mathcal{L}_{\text{recon}}(x,\hat{x}) = ||x-\hat{x}||_1$ and $\mathcal{L}_{\text{recon}}(y,\hat{y})$ is applied to enforce the content and raindrop style and are being preserved. As y is a clean image, c_y should be identical to y, identity loss $\mathcal{L}_{\text{idt}}(y,c_y) = ||y-c_y||_1$ is applied to ensure G_D does not change the input image when no rain present.

3.2. Composition-to-Decomposition

The order of G_D and G_C is reversed in C2D pipeline. G_C composes s_x and c_y obtained from D2C to $\tilde{y}$, which is then decomposed to $\hat{s}_x$ and $\hat{c}_y$ by G_D. Adversarial loss is applied to G_C and D_R, is defined as:

$$\mathcal{L}_{\text{GAN}}(D_R) = \frac{1}{2}\mathbb{E}_{x\sim p_{data}(x)}[(D_R(x)-1)^2] + \frac{1}{2}\mathbb{E}_{\tilde{y}\sim p_{data}(\tilde{y})}[(D_R(\tilde{y}))^2] \tag{5}$$

and

$$\mathcal{L}_{\text{GAN}}(G_C) = \frac{1}{2}\mathbb{E}_{\tilde{y}\sim p_{data}(\tilde{y})}[(D_R(\tilde{y})-1)^2], \tag{6}$$

where $p_{data}(\tilde{y})$ is the raindrop image distribution composed by G_C and D_R is the raindrop image discriminator, which distinguishes between a generated raindrop image $\tilde{y}$ and a real raindrop image x. We optimize $\mathcal{L}_{\text{GAN}}(D_R)$ and $\mathcal{L}_{\text{GAN}}(G_C)$ to make G_C generate photo-realistic raindrop images. In another word, a raindrop style is transferred from a raindrop image to another clean image. We then can train G_D with augmented data in paired manner by applying reconstruction loss $\mathcal{L}_{\text{recon}}(c_y,\hat{c}_y)$ and $\mathcal{L}_{\text{recon}}(s_x,\hat{s}_x)$. Likewise, c_x and s_y are also fed to the pipeline. $\tilde{x}$, $\hat{s}_y$ and $\hat{c}_x$ are generated accordingly with $\mathcal{L}_{\text{recon}}(c_x,\hat{c}_x)$, $\mathcal{L}_{\text{recon}}(s_y,\hat{s}_y)$ and $\mathcal{L}_{\text{idt}}(c_x,\tilde{c}_x)$ being applied to them.

3.3. Objective Function

$\mathcal{L}_{\text{idt}}$ and $\mathcal{L}_{\text{recon}}$ are summarized as :

$$\mathcal{L}_{\text{idt}} = \mathcal{L}_{\text{idt}}(y,c_y) + \mathcal{L}_{\text{idt}}(c_x,\tilde{x}) \tag{7}$$

and

$$\mathcal{L}_{\text{recon}} = \mathcal{L}_{\text{recon}}(x,\hat{x}) + \mathcal{L}_{\text{recon}}(y,\hat{y}) + \mathcal{L}_{\text{recon}}(c_x,\hat{c}_x) + \mathcal{L}_{\text{recon}}(c_y,\hat{c}_y) + \mathcal{L}_{\text{recon}}(s_x,\hat{s}_x) + \mathcal{L}_{\text{recon}}(s_y,\hat{s}_y). \tag{8}$$

Put all losses together, our full objective function is:

$$\mathcal{L}(G_D,G_C,D_R,D_C) = \mathcal{L}_{\text{GAN}}(G_D) + \mathcal{L}_{\text{GAN}}(G_C) + \mathcal{L}_{\text{GAN}}(D_C) + \mathcal{L}_{\text{GAN}}(D_R) + \mathcal{L}_{\text{recon}} + \lambda\mathcal{L}_{\text{idt}}, \tag{9}$$

where λ controls the importance of $\mathcal{L}_{\text{idt}}$. We optimize the function as follow:

$$G_D^*, G_C^* = \arg\min_{G_D,G_C,D_R,D_C} \mathcal{L}(G_D,G_C,D_R,D_C) \tag{10}$$

Intuitively, G_D and G_C can be viewed as an encoder and decoder pair. In D2C, G_D encodes x, and store them into latent code c_x and s_x. G_C composes the $\hat{x}$ with c_x and s_x. The objective function is used to regularize the latent codes. $\mathcal{L}_{\text{recon}}(x,\hat{x})$ enforces s_x and c_x preserve the content and raindrop style which can be used to restore x, so G_C can reconstruct the x losslessly. $\mathcal{L}_{\text{GAN}}(G_D)$ enforces the G_D only encodes clean content in c_x. In C2D, $\mathcal{L}_{\text{GAN}}(G_C)$ enforces G_D only encodes raindrop style in s_x, as s_x is used to apply a raindrop style in y. Hence, the content and raindrop style are decomposed optimally from a raindrop image.

4. Implementation

The residual block [9] has been proved effective in learning identity mapping. Fog, mist, shadow, motion blur, and noises usually spread over the entire image. However, raindrops usually only distort a portion of images. Therefore, there could be a large portion of the image is clean and only require identity mapping. We introduce Decomposition Residual Block and Composition Residual Block(see Figure4) for G_D and G_C to facilitate the identity mapping in the clean part and focus only on learning the residual between raindrops and its clean scene.

4.1. Decomposition Residual Block

We define $c_x = tanh(r_x^d + x)$, where r_x^d denoted as a residual image. Firstly, G_D takes x as an input and generate r_x^d and s_x. r_x^d is then added to x followed by $tanh$ activation to obtain c_x. We can see that G_D learns to generate the residual image. Then it produces clean images indirectly. This process is called decomposition. Similarly, G_D can also take y as input to get $(c_y,s_y) = G_D(y)$.

4.2. Composition Residual Block

c_x and s_x are first concatenated and passed to G_C to generate r_x^c. r_x^c is a residual image, which is then added to c_x, followed by $tanh$ activation to produce $\hat{x}$

Through the experiment, we find out that encoding the raindrop style code to the same dimension as the output image can keep the geometry and style information. When the style code is composed with another clean image, the raindrop visual effect display on the fake raindrop image is very similar to the one in the original raindrop image. Figure 7 shows the effectiveness of raindrop transfer.

4.3. Discriminative Network

We use PatchGAN [14] architecture as D_C and D_R. Raindrop images and clean images are extracted by CNNs and downscaled by four times. LSGAN[23] loss is used for adversarial training. The real label is set to 1 and the fake label is set to 0.

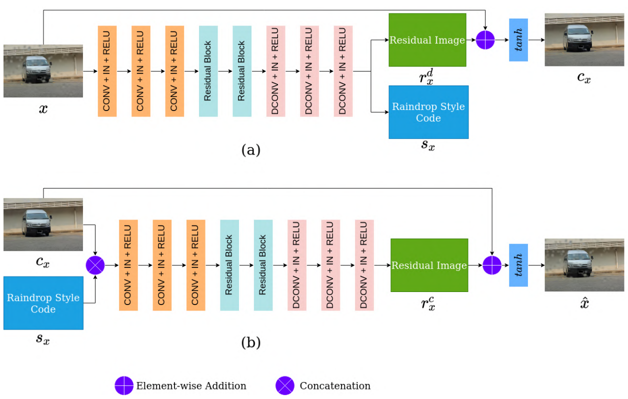


Figure 4. **The architecture of our decomposition residual block (a)** and **composition residual block (b)**. The input images are first passed through three Convolution, Instance Normalization, and Relu blocks, followed by two ResNet blocks and three Deconvolution, Instance Normalization, and Relu blocks. The outputs are the residual images which then being added to the input images, followed by a $\tanh$ activation

4.4. Training and Inference Scheme

During the training, images are randomly sampled from the clean and raindrop domains. We apply random flip to the images, followed by randomly adjusting the brightness and contrast. We use Adam optimizer to optimize generators and discriminators. The learning rate is set to 0.0002 and λ is set to 20. During inference, we simply forward the images to G_D to obtain the clean image.

5. Experiments

5.1. Dataset and Evaluation Metrics

The true unpaired real-world dataset has no ground truth for the raindrop images. To evaluate the effectiveness of raindrop removal in real-world images, we used two well-aligned real-world raindrop datasets, Qian *et al*. [26] dataset and Robotcar [25] dataset. They were collected for paired training with the ground truth. Therefore, we not only demonstrated the raindrop removal on the real-world images but also used the two common metrics, peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) with the ground truth for our evaluation. We trained the model in an unpaired manner on the real-world datasets to show the feasibility of training on other unpaired real-world datasets.

Qian *et al*. [26] dataset has a train set and two test sets. The train set comprises 861 pairs of images. **Test_a** contains 249 pairs of images and **Test_b** contains 58 pairs of images. A glass with a water droplet is used to simulate the adherent raindrop. The raindrop-clean image pairs were taken with and without the glass in front of the camera lens. Any movement or ambient change in the scene caused the misalignment between the paired images and inaccurate evaluation results. These datasets are not truly paired because the images were taken at different times. PSNR evaluates the pixel accuracy and SSIM focuses more on the structure of the contents. PSNR is affected significantly by small changes in ambient light or misalignment. Since raindrops usually change the structure of the clean images significantly, we believe that SSIM is more appropriate than PSNR in the evaluation of this problem.

Robotcar [25] dataset was collected by a stereo camera mounting on a moving car. Water was dynamically sprayed on the right camera while the left camera was clean all the way. The left camera and the right camera were calibrated to align with the scene. The entire dataset is consists of 4816 paired images in sequence. Raindrops in this dataset are much denser than the Qian *et al*. [26]. They occlude the majority of the area of the raindrop images. We randomly selected 500 pairs of images for the evaluation set and 4316 pairs of images for the training set. The training images are equally divided into two sets. Each set has 2158 pairs of images. We used the raindrop images from the first set as the raindrop domain images and the clean images from the second set were taken as the clean domain images. Hence the clean domain and the raindrop domain training images are truly unpaired sets.

5.2. Baselines

We evaluated our method on both datasets and compared it with WSRR-GAN [22], the only unpaired method dedicated for raindrop removal to our knowledge, LIR [5], an unpaired method for Gaussian noise removal, CycleGAN, a generic one-to-one style transfer method, and DRNet, an unpaired method for deblurring. One of the state-of-the-art paired methods, AttentiveGAN [26] was also included as a reference.

(a) (b) (c) (d) (e) (f) (g) (h)

Figure 5. Visual comparison among different methods. From left to right: (a) Raindrop image, (b) LIR, (c) CycleGAN, (d) AttentiveGAN, (e) WSRRGAN, (f) DRNet, (g) Ours, and (h) Ground-truth. Our method has successfully removed most raindrops while keeping the color and content almost the same as the ground truth.

Setting	Method	**Test_a**		**Test_b**	
		PSNR	SSIM	PSNR	SSIM
Paired	AttentiveGAN	**31.5700**	0.9023	24.1596	0.8292
Unpaired	CycleGAN	24.2038	0.8396	21.9647	0.7802
	WSRR-GAN	25.4624	0.8763	23.2445	0.8064
	LIR	21.3000	0.8393	20.5594	0.7990
	DRNet	24.8379	0.8616	23.0263	0.8171
	Ours	28.5517	**0.9095**	**25.6648**	**0.8627**

Table 1. Quantitative results on Qian *et al*. [26] dataset. **Test_a** and **Test_b** are used for evaluation. **Test_b** has denser raindrops than **Test_a**, so all the results are lower. Our methods outperform the other methods on both test set.

	PSNR	SSIM
Raindrop	13.04	0.43
CycleGAN	15.72	0.56
LIR	13.51	0.49
Ours	**16.66**	**0.59**

Table 2. Quantitative results on Robotcar [25] dataset. The raindrop image has a very low PSNR and SSIM score. It's difficult the estimate the ground truth with little information. Our methods estimate the best clean feature and score the highest scores

5.3. Results

Table 1 shows that our method outperforms other state-of-the-art unpaired methods by a large margin on **Test_a** on both PSNR and SSIM. Our SSIM results are even higher than AttentiveGAN's. LIR performs poorly as it only focuses on texture transfer while keeping the image structure, whereas raindrops partially distort the content image structure. CyclGAN works better than LIR as it does not impose so many structure consistency constraints. The result is still far from state-of-the-art. WSRR-GAN's results are far behind ours as well. On the harder dataset **Test_b**, which has

(a) Raindrop (b) CycleGAN (c) LIR (d) Ours (e) Groundtruth

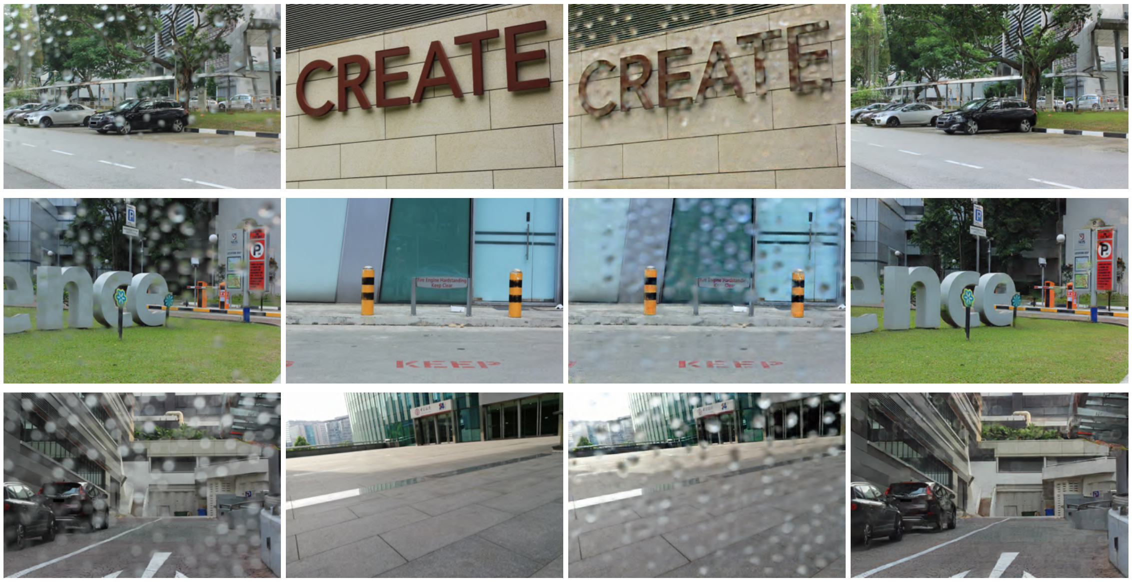
Figure 6. Qualitative evaluation on Robotcar [25]. Despite the visibility is largely occluded and distorted, our method restored coherent clean features in the raindrop images.

some misalignment. All the results are lower than **Test_a**'s. It could also because the raindrops are denser and larger in **Test_b** than **Test_a**. Despite the difficulty, our method outperforms all paired and unpaired methods on both PSNR and SSIM by a large margin. It shows that our method is very capable of restoring the structure of distorted raindrop images. As **Test_b** contains four times more images than **Test_a**, we believe that the results are less biased to the test set. It shows that our method has a better generalization over a larger test set and unseen data.

Figure 5 shows the qualitative results. LIR [5] barely removes raindrops as it is optimized for high-frequency texture noise removal, but raindrops are mostly in low-frequency content space. Due to the cycle-consistency constrain, CycleGAN is forced to encode all information including the raindrop into the output clean images. The raindrops cannot be removed optimally. AttentiveGAN largely removes raindrops, while suffering from color shift and artifacts. Our method removes most of the raindrops while creating fewer artifacts.

Figure 6 demonstrates raindrop removal performance on Robotcar [25] dataset, which raindrops are large and dense. It is very hard to estimate the content being occluded even for a human. LIR [5] somehow removes the small raindrops, while the large raindrop remains. The images restored by CycleGAN are darker than the ground truth and some artifacts can be seen. Our method successfully removes the raindrops and restores coherent content. However, the estimated coherent content could potentially be wrong when restored from a large occluded raindrop area. In practice, the windshield wiper constantly removes the raindrops. The leftover raindrops are much smaller, which true content can be easily estimated by our method.

We also measure the decomposition generator inference speed. Our model runs at 67 Hz with 384×256 resolution images on an Nvidia RTX 2080Ti GPU. Practically, our decomposition generator can be added before computer vision tasks such as object detection without creating too much overhead. Due to the 2-stage constraints, WSRR-GAN can only run at 4.8 Hz at a similar resolution.



(A) (B) (C) (D)

Figure 7. Transfer visualization. (A) Real-world raindrop image A. (B) Real-world clean image B. (C) Clean image B composed with A raindrop style. (D) Raindrop removed A. The location and the style of raindrop have been reproduced on the clean image.

5.4. Ablation Study

To evaluate the effectiveness of D2C, C2D, and residual module(RM) in our framework, we add them one by one to train and evaluate under the same setting and compare with D2C without residual module as a baseline. D2C without residual block directly maps a raindrop image to a clean image using an autoencoder. Table 3 shows the quantitative result of the ablation study. With residual module (RM), PSNR and SSIM are much higher than without it. Adding the C2D residual module, the results further improve due to the data augmentation. We can tell that both RM and C2D are very effective in removing the raindrop.

Method	PSNR	SSIM
D2C	24.5031	0.8411
D2C + RM	26.5031	0.8811
D2C + C2D + RM (ours)	28.5517	0.9095

Table 3. Adding RM and C2D, the model performance boosted significantly

5.5. Raindrop Transfer

We also study whether our decomposition generator can create a unique mapping between a raindrop style and a style code by conducting a raindrop transfer experiment. Figure 7 demonstrates that G_D not only removes the raindrops but also successfully extracts a raindrop visual style and encodes it to a content-invariant style code. With the same style code, G_C transfer it naturally to the other clean images.

6. Conclusion and future work

In this paper, we propose a concise end-to-end training framework RainGAN for adherent raindrop removal, which leverages unpaired real-world data. It fills the gap where the dedicated raindrop removal paired methods cannot be generalized well on real-world raindrop images. It is the first method that can be deployed for outdoor camera intelligent systems to remove the raindrop effectively. The decomposition generator is domain invariant and only performs raindrop removal when there are raindrops in the images. In addition, we have successfully demonstrated its capability to extract raindrop style from a raindrop image and transfer the style to another image using our proposed decomposition and composition generators. This capability can be further exploited as a way for natural noise augmentation in other applications. In the future, as the framework is generic, it can be applied to other noise removal problems such as fog and haze, where clean-noise paired images are practically impossible to collect.

7. Acknowledgement

This study is supported under the RIE2020 Industry Alignment Fund – Industry Collaboration Projects (IAF-ICP) Funding Initiative, as well as cash and in-kind contribution from Singapore Telecommunications Limited (Singtel), through Singtel Cognitive and Artificial Intelligence Lab for Enterprises (SCALE@NTU).

References

[1] Stefano Alletto, Casey Carlin, Luca Rigazio, Yasunori Ishii, and Sotaro Tsukizawa. Adherent raindrop removal with self-supervised attention maps and spatio-temporal generative adversarial networks. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops*, Oct 2019.

[2] Yunjey Choi, Minje Choi, Munyoung Kim, Jung-Woo Ha, Sunghun Kim, and Jaegul Choo. Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 8789–8797, 2018.

[3] Yunjey Choi, Youngjung Uh, Jaejun Yoo, and Jung-Woo Ha. Stargan v2: Diverse image synthesis for multiple domains. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8188–8197, 2020.

[4] Sen Deng, Mingqiang Wei, Jun Wang, Yidan Feng, Luming Liang, Haoran Xie, Fu Lee Wang, and Meng Wang. Detail-recovery image deraining via context aggregation networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14560–14569, 2020.

[5] Wenchao Du, Hu Chen, and Hongyu Yang. Learning invariant representation for unsupervised image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14471–14480, 2020.

[6] Xueyang Fu, Jiabin Huang, Xinghao Ding, Yinghao Liao, and John Paisley. Clearing the skies: A deep network architecture for single-image rain removal. *IEEE Transactions on Image Processing*, 26(6):2944–2956, 2017.

[7] Xueyang Fu, Jiabin Huang, Delu Zeng, Yue Huang, Xinghao Ding, and John Paisley. Removing rain from single images via a deep detail network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3855–3863, 2017.

[8] Zhixiang Hao, Shaodi You, Yu Li, Kunming Li, and Feng Lu. Learning from synthetic photorealistic raindrop for single image raindrop removal. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pages 4340–4349, 2019.

[9] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[10] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

[11] Xiaowei Hu, Chi-Wing Fu, Lei Zhu, and Pheng-Ann Heng. Depth-attentional features for single-image rain removal. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8022–8031, 2019.

[12] Xiaowei Hu, Yitong Jiang, Chi-Wing Fu, and Pheng-Ann Heng. Mask-shadowgan: Learning to remove shadows from unpaired data. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 2472–2481, 2019.

[13] Xun Huang, Ming-Yu Liu, Serge Belongie, and Jan Kautz. Multimodal unsupervised image-to-image translation. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 172–189, 2018.

[14] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1125–1134, 2017.

[15] Kui Jiang, Zhongyuan Wang, Peng Yi, Chen Chen, Baojin Huang, Yimin Luo, Jiayi Ma, and Junjun Jiang. Multi-scale progressive fusion network for single image deraining. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8346–8355, 2020.

[16] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *European conference on computer vision*, pages 694–711. Springer, 2016.

[17] 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 *Proceedings of the 26th ACM international conference on Multimedia*, pages 1056–1064, 2018.

[18] Xia Li, Jianlong Wu, Zhouchen Lin, Hong Liu, and Hongbin Zha. Recurrent squeeze-and-excitation context aggregation net for single image deraining. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 254–269, 2018.

[19] Huangxing Lin, Changxing Jing, Yue Huang, and Xinghao Ding. A 2 net: Adjacent aggregation networks for image raindrop removal. *IEEE Access*, 8:60769–60779, 2020.

[20] Ming-Yu Liu, Thomas Breuel, and Jan Kautz. Unsupervised image-to-image translation networks. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems 30*, pages 700–708. Curran Associates, Inc., 2017.

[21] Boyu Lu, Jun-Cheng Chen, and Rama Chellappa. Unsupervised domain-specific deblurring via disentangled representations. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 10225–10234, 2019.

[22] Wenjie Luo, Jianhuang Lai, and Xiaohua Xie. Weakly supervised learning for raindrop removal on a single image. *IEEE Transactions on Circuits and Systems for Video Technology*, 2020.

[23] Xudong Mao, Qing Li, Haoran Xie, Raymond YK Lau, Zhen Wang, and Stephen Paul Smolley. Least squares generative adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pages 2794–2802, 2017.

[24] Jiayi Peng, Yong Xu, Tianyi Chen, and Yan Huang. Single-image raindrop removal using concurrent channel-spatial attention and long-short skip connections. *Pattern Recognition Letters*, 131:121–127, 2020.

[25] Horia Porav, Tom Bruls, and Paul Newman. I can see clearly now: Image restoration via de-raining. In *2019 International Conference on Robotics and Automation (ICRA)*, pages 7087–7093. IEEE, 2019.

[26] Rui Qian, Robby T Tan, Wenhan Yang, Jiajun Su, and Jiaying Liu. Attentive generative adversarial network for raindrop removal from a single image. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2482–2491, 2018.

[27] Yuhui Quan, Shijie Deng, Yixin Chen, and Hui Ji. Deep learning for seeing through window with raindrops. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 2463–2471, 2019.

[28] Yiyang Shen, Yidan Feng, Sen Deng, Dong Liang, Jing Qin, Haoran Xie, and Mingqiang Wei. Mba-raingan: Multibranch attention generative adversarial network for mixture of rain removal from single images. *arXiv preprint arXiv:2005.10582*, 2020.

[29] Ülkü Uzun and Alptekin Temizel. Cycle-spinning gan for raindrop removal from images. In *2019 16th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, pages 1–6. IEEE, 2019.

[30] Hong Wang, Qi Xie, Qian Zhao, and Deyu Meng. A model-driven deep neural network for single image rain removal. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3103–3112, 2020.

[31] Tianyu Wang, Xin Yang, Ke Xu, Shaozhe Chen, Qiang Zhang, and Rynson W.H. Lau. Spatial attentive single-image deraining with a high quality real rain dataset. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 12262–12271, June 2019.

[32] 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 *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1357–1366, 2017.

[33] Rajeev Yasarla, Vishwanath A Sindagi, and Vishal M Patel. Syn2real transfer learning for image deraining using gaussian processes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2726–2736, 2020.

[34] Yuntong Ye, Yi Chang, Hanyu Zhou, and Luxin Yan. Closing the loop: Joint rain generation and removal via disentangled image translation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2053–2062, 2021.

[35] Shaodi You, Robby T Tan, Rei Kawakami, Yasuhiro Mukaigawa, and Katsushi Ikeuchi. Adherent raindrop modeling, detectionand removal in video. *IEEE transactions on pattern analysis and machine intelligence*, 38(9):1721–1733, 2015.

[36] He Zhang and Vishal M Patel. Density-aware single image de-raining using a multi-stream dense network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 695–704, 2018.

[37] He Zhang, Vishwanath Sindagi, and Vishal M Patel. Image de-raining using a conditional generative adversarial network. *IEEE transactions on circuits and systems for video technology*, 30(11):3943–3956, 2019.

[38] Ziqiang Zheng, Yang Wu, Xinran Han, and Jianbo Shi. Forkgan: Seeing into the rainy night. In *The IEEE European Conference on Computer Vision (ECCV)*, August 2020.

[39] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, Oct 2017.