Both Diverse and Realism Matter: Physical Attribute and Style Alignment for Rainy Image Generation

Changfeng Yu[1], Shiming Chen[2,3], Yi Chang[1,*], Yibing Song[4], Luxin Yan[1]
[1]National Key Laboratory of Science and Technology on Multispectral Information Processing, School of Artificial Intelligence and Automation, Huazhong University of Science and Technology
[2]Carnegie Mellon University [3]Mohamed bin Zayed University of Artificial Intelligence
[4]AI[3] Institute, Fudan University
ycf@hust.edu.cn, {gchenshiming, owuchangyuo, yibingsong.cv}@gmail.com, yanluxin@hust.edu.cn

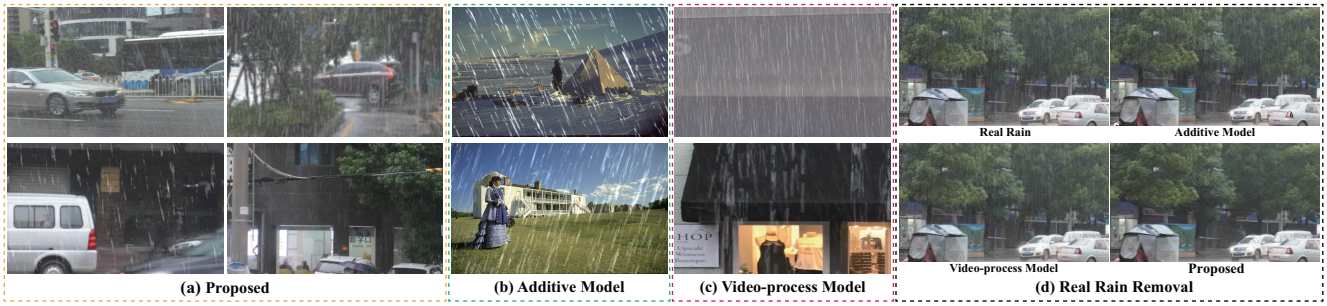


Figure 1: Comparision of different rain image generation model. (a) The proposed method can generate diverse and realistic images by manipulating attributes of rain streak. In contrast, the additive model [49] generates diverse but less realistic images (shown in (b)), while video-process model [43] produces realistic but homogeneous images (shown in (c)). Furthermore, (d) shows the real rain removal of classical deraining methods, trained on datasets generated by different rain generation models.

Abstract

Although considerable progress has been made in image deraining under synthetic data, real rain removal is still a tough problem due to the huge domain gap between synthetic and real data. Besides, difficulties in collecting and labeling diverse real rain images hinder the progress of this field. Consequently, we attempt to promote real rain removal from rain image generation (RIG) perspective. Existing RIG methods mainly focus on diversity but miss realistic, or the realistic but neglect diversity of the generation. To solve this dilemma, we propose a physical alignment and controllable generation network (PCGNet) for diverse and realistic rain generation. Our key idea is to simultaneously utilize the controllability of attributes from synthetic and the realism of appearance from real data. Specifically, we devise a unified framework to disentangle background, rain attributes, and appearance style from synthetic and real data. Then we collaboratively align the factors with a novel semi-supervised weight moving strategy for attribute, an explicit distribution modeling method for real rain style. Furthermore, we pack these aligned factors into the generation model, achieving physical controllable mapping from the attributes to real rain with image-level and attribute-level consistency loss. Extensive experiments show that PCGNet can effectively generate appealing rainy results, which significantly improve the performance under synthetic and real scenes for all existing deraining methods.

1. Introduction

The objective of single image deraining (SID) is to reclaim a clean image from the rainy image, a task often deemed an indispensable preliminary stage within outdoor computer vision tasks such as flow estimation [15, 47], detection [33], and segmentation [9]. Top performing SID methods [13, 49, 39, 43, 44, 19, 28, 48, 40, 29, 41, 50, 10] are suffering from a common issue: the scarcity of paired real rainy-clean images, thereby resulting in an obvious performance drop in the real rain removal. To overcome this issue, tremendous efforts have been made to advance the real image deraining era. However, most of them [44, 52, 7, 23, 45, 54, 55, 58] focus on the network architecture for real rain removal, while fewer works [43, 53, 42] have paid attention to datasets via paired clean-rain image generation. In this work, we attempt to promote the real rain

removal of existing SID methods from a data perspective.

Currently, most of the existing rain datasets are simulated by a physical synthetic model. These simulators [49, 32, 19, 28, 16] utilize hand-crafted prior such as motion blur to simulate the rain streaks and then add them to the clean images. The diverse rainy images are generated by empirically modulated synthetic parameters. However, empirically setting through human subjective assumptions would restrict the generated rain types [42]. To get rid of empirical restrictions, some supervised-learning-based methods are proposed [42, 37] to further promote diversity of generation results. The intrinsic of these methods is utilizing the powerful representation of deep CNN to overfit the physical synthetic rain generation model. Although these diverse synthetic datasets can be used to train deep derainers to some extent, the physical and learning-based rain simulators simplified the degradation of the real rain, resulting in an obvious domain gap between these synthetic datasets and complicated real rain and inevitably bringing performance drop in the real rain removal [55].

To reduce the gap between the synthetic simulator and real degradation, some works [53] treat the rain image generation as a style transfer task through adversarial learning with real rain. However, the GAN-based methods suffer from model collapse [1] and uncontrollable inference, resulting in undesired artifacts and unitary appearance in the generation results. Meanwhile, some works [43, 2] utilize the temporal information in the video of real rain to obtain the real paired rainy/clean image. Though the rain images are realistic, there are many limitations to generating the corresponding 'clean' image. On the one hand, the clean image generation is very rough, which may be left residual streaks in the generated clean one [42]. On the other hand, the procedure of clean image generation requires that the video scene is strictly static, imposing restrictions on the diversity of the background [43].

In general, existing rain generation methods either learn from the synthetic rainy dataset with controllable diversity but neglect the photorealism, or learn from the real rainy dataset with realism but miss the diversity. However, the degradation process of real rain is complex and morphology varies. As shown in Fig. 1, we argue that diversity and photorealism are equally important for rain generation methods. For this goal, we propose physical alignment and controllable generation network (PCGNet) to generate diverse and realistic rainy-clean image pairs. Specifically, we first learn to physically disentangle the rain images into the background, attributes of rain steaks, and style space. Since the attribute of real rain is unlabeled, we propose a novel semi-supervised weight-moving strategy to enable mutual interaction between real and synthetic data. For style space, we explicitly model the appearance style of real rain in the latent space and align the style of real and synthetic images through distribution resampling. Furthermore, we pack these aligned factors into the generation model, achieving physical controllable mapping from the attributes to real rain with image-level and attribute-level consistency loss. In summary, our contributions are mainly three folds:

- Different from existing real rain removal methods which focus on refining network architectures, we attempt to promote real rain removal from a data perspective. We establish a connection between synthetic and authentic rain images within a semi-supervised framework, facilitating the concurrent generation of diverse and realistic outputs, which offers a new insight into the rain image generation community.
- We find the physical relationship of the rain attribute and style between synthetic and real data. Thus, we propose a novel physical disentangle module to decouple the synthetic and real rain into attributes and style, and align them in a semi-supervised strategy. Furthermore, we propose a controllable generation model, achieving physical mapping from attributes to real rain and enabling realism and diverse generations.
- Extensive experiments show that our method achieves more appealing generation, and significantly improves all the deraining methods both in the synthetic and real scenes. Moreover, PCGNet enables the attribute fine-grained manipulation of the real rain.

2. Related Works

2.1. Rain Removal

Rain removal is a highly ill-posed inverse problem. Pioneer works design hand-crafted prior to decomposing a rainy image into the background and rain layer [24, 34, 8, 6, 31]. Recently, benefiting from the powerful representation of CNN, deep learning-based methods have achieved remarkable progress. The CNN-based single image rain removal methods can be mainly classified into the following categories: full-supervised [13, 30, 48, 41, 56, 46], semi-supervised [44, 52, 53], and unsupervised [7, 23, 45, 54, 55]. Most existing methods are full-supervised where synthetic paired rainy/clean images are required. Fu *et al*. [13] first introduced the end-to-end residual CNN to solve the rain streaks removal problem. The multi-stage and multi-scale architecture networks [38, 51, 22] have been extensively studied for better feature representation. To better generalize the real rain, the researchers employed the semi-supervised learning paradigm. For example, apart from the supervised loss, Wei *et al*. [44] additionally enforced a parameterized GMM distribution on real rain streaks. To get rid of the limitation of the paired synthetic-clean training data, the unsupervised methods have raised attention. Wei *et al*. [45] extended the classical CycleGAN into the DerainCycleGAN using unpaired data for real image derain-

ing. Yu *et al*. [55] took the prior knowledge into consideration and connected the model-driven and data-driven methods via an unsupervised learning framework. Although unsupervised methods can eliminate the dependence on simulated data to some extent, they still rely on a large amount of diverse and realistic rainy images for training, which requires significant cost. In this work, we attempt to promote real rain removal from data generation perspective.

2.2. Rain Generation

The datasets play an important role in the deep learning era. The researchers have made great progress in the rain synthetic model to provide numerous rain datasets, mainly including hand-crafted synthetic model [14, 49, 32, 19, 28, 16], CNN generation model [53], and video-based generation model. The pioneering work was proposed by Nayar *et al*. [14], with geometric and photometric analysis for real rain appearance. Similarly, researchers synthesize different rain streaks with a physical additive degradation model, then add the streaks layer on the clean image to construct paired samples such as Rain100L [49]. To further improve the diversity of synthetic data, Wang *et al*. [42] utilize CNN to fit the synthetic rain process in a supervised way. Though these synthetic generation methods could modulate synthetic parameters to achieve diverse generation results, there is still a huge domain gap between the synthetic and real data. Along with the research, some CNN generation-based methods [53, 45] were proposed to get more photo-realism rain. Ye *et al*. [53] treat rain image generation as a style transfer task through adversarial learning with real rain. However, due to the uncontrollable generation procedure, the results are relatively homogeneous. Another realism generation methods are video-based [43, 2]. Wang *et al*. [43] first proposed a real rain dataset with the corresponding 'clean' image, which was a semi-automatic method from real rain videos. But there still exists the diversity challenge since the harsh shooting conditions. In this work, we propose a physical alignment and controllable generation network with realistic and diverse rainy/clean image results.

2.3. Controllable Generative Model

There has been significant progress in the development of controllable generative models. Several approaches have been proposed to achieve controllable generation, including VAE-based disentangled representations [18, 27, 25], and GAN-based conditional generative models [35, 3, 4]. These models have been widely applied in face editing, person image synthesis, style transfer, and so on. However, there still exists a very practical challenge that the trade-off between controllability and quality of results [11]. Compared to existing generative methods, we are the first to simultaneously consider the diversity and realism of rainy image generation in a semi-supervised way. This is a more complex task due to the coupling of degradation and background. Furthermore, our goal for rainy image generation is to improve the performance of existing deraining methods in real rain scenes, rather than focusing solely on visual effects.

3. Physical Alignment and Controllable Generation Network

3.1. Framework Overview

Given a clean image and the desired attribute of rain streaks such as angle, length, density, and brightness, our rainy image generation aims to achieve a controllable mapping between the physical attribute space and realistic rainy image as expected properties. The key problems are that the attributes of real rain are unlabeled, which will lead the procedure of generation hard to control, and the coupling of the degradation and background, which will lead to realistic generation results is hard to guarantee.

To tackle these problems, we propose to jointly consider the controllable and realistic within a physical alignment and controllable generation framework. The core idea is to connect the diversity of synthetic and realism of real data in a semi-supervised way. Specifically, as shown in Fig. 2, the proposed PCGNet consists of two components: physical disentangle and alignment (PDA), controllable generation module (CGM). The PDA module takes the synthetic rainy image I^{Syn} and real rainy image I^{Real} as inputs and disentangles them into physical variable spaces such as background, attributes, and appearance style of the rain. Further knowledge transfer is then conducted between synthetic and real data in both attribute and style spaces to achieve the distribution alignment. Moreover, the aligned parameters (including the background, attributes, and rain style) of synthetic and real are fed into a weight-shared controllable generation model to respectively generate I^{Gen} and I^{Rec}. We construct image consistency between the I^{Rec} and I^{Real}, and attribute consistency between the I^{Gen} and I^{Syn}. Such progressive multi-stage strategy would significantly ease difficulties of controllable and realistic generation.

3.2. Physical Disentangle and Alignment Module

Rainy weather degradation is a complex and varied process influenced by environmental parameters (such as lighting and wind direction) and camera settings. These factors result in a wide range of visual appearances in the image. We analyze its various attributes from its specific physical imaging process. On rainy days, raindrops appear as elliptical shapes [14], randomly distributed in space, with varying density depending on the rainfall intensity. Influenced by gravity and wind direction, raindrops exhibit rapid motion in a specific direction. During the imaging process, since the long exposure time of the camera and fast motion of raindrops, the visual appearance of the raindrops in image space is presented as severely motion-blurred rain streaks, with the length and width varying based on the speed and

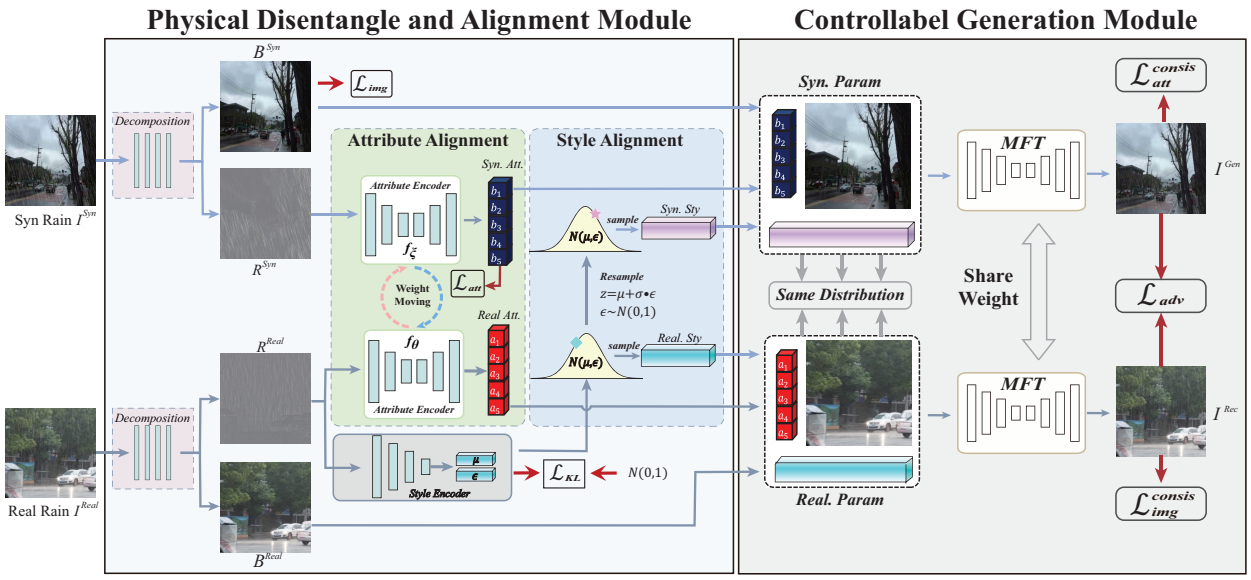


Figure 2: Overview of physical alignment and controllable generation network. The proposed method consists of two parts: physical disentangle and alignment module (PDA), controllable generation module(CGM). The PDA decouples synthetic and real rainy images into background, rain attributes, and appearance styles, and makes them aligned in the physical space. The CGM achieves physically controllable mapping from attributes to real rain, resulting in realistic and diverse generation.

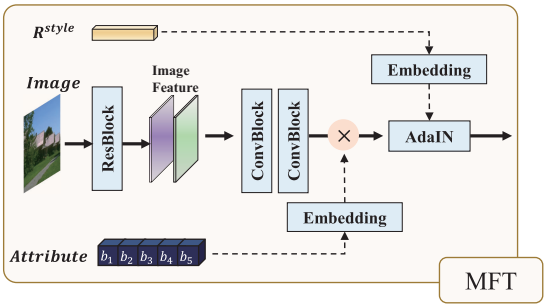


Figure 3: The architecture of MFT module. MFT could simultaneously modulates attribute features along with transforming feature style to real rain style.

size of the raindrop. Additionally, the brightness of the rain streaks is affected by environmental illumination. Our work focuses on manipulating these attributes (including density, location, length, angle, and brightness) to achieve diverse results in generating realistic rain images.

Atrribute alignment: Due to the high costs of manual annotation, we lack labeled attributes for real rain. In contrast, synthetic data generated by a simplified physical degradation model can record the physical properties of rain in real-time. Therefore, we attempt to guide the regression of real rain attributes using the labels of synthetic data. As shown in Fig. 2, we first use a decomposition module trained on simulated data to decompose rain images into two parts: rain layer and background. After obtaining the rain layer components for both synthetic R^{Syn} and real R^{Real}, we regress the corresponding attributes using an attribute encoder and pre-train the encoder with the attribute labels of the synthetic data. However, due to the distribution gap between synthetic and real data, training the attribute encoder solely on synthetic cannot accurately regress the attributes of real rain, which will affect the quality of generations.

To alleviate this, we propose a semi-supervised weight-step moving training strategy (SMT) inspired by MOSS [20]. Precise regression of the attributes of real-world data is crucial for improving the quality of subsequent rain image reconstruction. Therefore, we not only use the attribute labels of synthetic data as a supervised loss $\mathcal{L}_{att}$ but also employ image reconstruction loss $\mathcal{L}_{img}^{consis}$ of the subsequent image generation results to simultaneously constrain the gradient of the attribute regressor in a momentum-iteration manner. Specifically, we denote the f_θ and f_ξ as attribute regressors for synthetic and real data respectively. To achieve more accurate real attribute regression, we update the weight of θ and ξ in a step moving as follows:

$$\begin{aligned} \xi_t &= (1-v_1)\theta_{t-1} + v_1 \nabla \mathcal{L}_{img}^{consis} \\ \theta_t &= (1-v_2)\xi_{t-1} + v_2 \nabla \mathcal{L}_{attri} \end{aligned} \tag{1}$$

where the ξ and θ are the parameters of synthetic and real attribute encoder separately, v_1 and v_2 are initial update speed, t and t-1 represent the current and previous update stages.

Style alignment: Despite having similar physical properties in rain streaks, synthetic and real rainy images still exhibit significant discrepancies in their visual style, which are difficult to depict in statistics. To generate more realis-

tic results, it is important to focus on the style of real rain data rather than synthetic. Therefore, to explicitly model the style of real rain images, we regress a new latent variable S^{Real} using a style encoder. We assume that S^{Real} follows an isotropic Gaussian prior distribution as suggested in [12] and then use the Kullback-Leibler divergence to constrain it. Thus, the style of real rain images can be considered as different sampling points on the distribution as follows

$$\begin{aligned} S^{Real} &= \mu + \sigma \cdot \epsilon \\ \epsilon &\sim \mathcal{N}(s \mid 0, I_t) \end{aligned} \tag{2}$$

By resampling in this way, we can sample multiple sets of real rain styles with the same distribution, as shown in Fig. 2. To achieve style alignment between real rain and synthetic data, we assign the resampled Syn.Sty and Real.Sty as the style of synthetic and real streams respectively.

3.3. Controllable Generation Module

After decomposing in the PDA module, we packaged the backgrounds, attributes, and styles of the synthetic and real data as Syn.param and Real.param for the input of subsequent controllable generation module. Since the alignment between synthetic and real data has been achieved in terms of both attributes and styles. And both backgrounds belong to the real world, we can recognize they have the same distribution. Thus Syn.param and Real.param share the same distribution. Besides, the weights of the generator are also shared between the synthetic and real. Consequently, the output should naturally have the same distribution. Furthermore, we used both the real rain image consistency loss $\mathcal{L}_{img}^{consis}$ and adversarial loss $\mathcal{L}_{adv}$ to constrain outputs of the generator to have the style of the real rain. At the same time, we used the synthetic attribute consistency loss $\mathcal{L}_{attr}^{consis}$ to ensure the results of the generator have the desired attributes, thus achieving controllable generation.

Modulative feature translation: For the controllable generation model, we propose a modulative feature translation module (MFT), which simultaneously utilizes the attributes and style of the rain to modulate the properties of the result. Specifically, as shown in Fig. 3, we first extract high-level features $\mathbf{F}_b$ from the background through residual blocks. Then the MFT learns a mapping function $\mathcal{M} : (attri, style) \rightarrow (\alpha, \gamma, \beta)$ where the modulative parameter α is obtained by attribute conditions as $\alpha = \mathcal{M}_1(attri)$, and the affine transformation parameter γ, β is obtained by real rain style as $(\gamma, \beta) = \mathcal{M}_2(R^{style})$, where $\mathcal{M}_1$ and $\mathcal{M}_2$ are the embedding block. After obtaining the modulative and transform parameters α, γ, and β, the MFT first performs the feature modulation on the background feature map F_b as: $F_{mod} = F_b \otimes \alpha$. Then to transfer the feature to a real rain style, we utilize AdaIN [21] on the modulated feature as follows:

$$AdaIN(F_{mod}, \gamma, \beta) = \gamma\left(\frac{F_{mod} - \mu(F_{mod})}{\sigma(F_{mod})}\right) + \beta \tag{3}$$

Table 1: Photorealism comparison on FCRealRain.

Index	Rain100L	Rendering	VRGNet	JRGR	SPA	PCGNet
FID	141.84	128.61	128.92	122.5	**117.32**	**108.28**

With the MFT module, we can collaborate the attribute conditions and style to both modulate generation properties and transform the generation similar to real.

3.4. Loss Function

In this section, we give descriptions of loss functions in PCGNet, which can be roughly divided into three classes: disentangle loss, consistency loss, and adversarial loss.

Physical disentangle loss. Firstly, for the decomposition module, we need to separate the rain image into background and rain layer. Since there are no corresponding background labels available for real rain, we rely on synthetic rainy/clean image pairs for supervised training of the decomposition module, primarily utilizing the L2 loss:

$$\mathcal{L}_{img} = \left\| B^{Syn} - B^{GT} \right\|_2 \tag{4}$$

where B^{Syn} is background of synthetic rain image, B^{GT} is the corresponding clean image. Subsequently, we share weights trained on synthetic and real data. It should be noted that our concern mainly lies on modeling the interaction and knowledge transfer between simulated and real rain, so the residual rain streaks in background have little effects on the estimation of rain attributes A and modeling of their distribution in real rain.

For the attribute regression of rain streaks, since synthetic data possess attribute labels whereas the real data does not, we utilize the semi-supervised Step-moving training strategy to constrain the network training by supervised loss from synthetic data and the subsequent reconstruction loss of real rain. We define the attribute disentangle loss as:

$$\mathcal{L}_{attr}^{syn} = \left\| \mathcal{F}(R^{Syn}; W_{attr}) - A^{GT} \right\|_2 \tag{5}$$

$$\mathcal{L}_{attr} = \mathcal{L}_{attr}^{syn} + \left\| \mathcal{F}(R^{Real}; W_{attr}, W_{MFT}) - R^{Real} \right\|_2 \tag{6}$$

where the $\mathcal{F}(\bullet)$ is the network transformation, W_{attr} and W_{MFT} are attribute encoder parameter and MFT module parameter, R_{Syn} and R_{Real} are the rain layer of synthetic and real, A_{GT} is the corresponding synthetic attribute labels.

For the style regression of real rain, we impose the Gaussian prior distribution on style and the KL divergence to constrain the real rain style close to the prior $p(z)$ as:

$$p(z) \sim \mathcal{N}(s \mid 0, I_t) \tag{7}$$

$$\begin{aligned} \mathcal{L}_{KL}(S^{Real}; \mu, \sigma) &= \mathcal{D}_{KL}[q(S|R^{Real}) \mid\mid p(z)] \\ &= \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{M_z} (1 + \log \sigma_{ij} - \mu_{ij}^2 - \sigma_{ij}^2) \end{aligned} \tag{8}$$

where $I_t \in \mathcal{R}^{t \times t}$ is the unit matrix, and M_z is the dimension of the style latent vector S^{Real}.

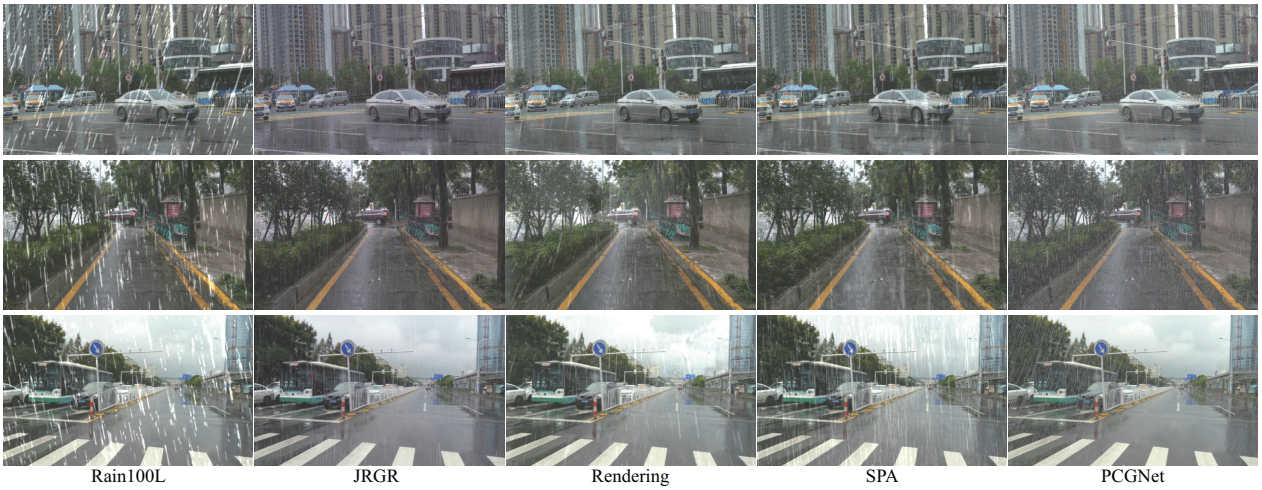


Figure 4: The qualitative comparison on generation results.

Figure 5: The qualitative generation results. The first column are two forms of input in the proposed methods. When taking clean images as input, our methods can generate diverse results by randomly sampling attribute parameters. When taking rainy images as input, PCGNet can manipulate real rain by modulating attribute parameters based on original attributes.

Consistency loss. To ensure the generation results follow the same style as the real rain, we employ image-level reconstruction consistency loss $\mathcal{L}_{img}^{consis}$ for reconstruction results I^{Rec} and real rain. For controllability of the generated results, we utilize attribute-level reconstruction consistency loss $\mathcal{L}_{attr}^{consis}$ for the synthetic generation I^{Gen} and synthetic rain. We define the $\mathcal{L}_{img}^{consis}$ and $\mathcal{L}_{attr}^{consis}$ as follows:

$$\mathcal{L}_{img}^{consis} = \left\| \mathcal{F}(R^{Real}, B^{Real}; W_{attri}, W_{sty}, W_{MFT}) - I^{Real} \right\|_2 \quad (9)$$

$$\mathcal{L}_{attr}^{consis} = \left\| \mathcal{F}(I^{Gen}; W_{attri}) - \mathcal{F}(I^{Syn}; W_{attri}) \right\|_2 \quad (10)$$

Adversarial loss. In addition to consistency losses, we introduce the adversarial loss between the synthetic generation I^{Gen} and the real reconstruction rain I^{Rec}, the loss function is defined as below:

$$\mathcal{L}_{adv} = \mathbb{E}[\log(1 - D(I_{gen}))] + \mathbb{E}[\log D(I_{Rec})] \quad (11)$$

Overall objective function. The full objective function contains three losses as follows:

$$\begin{aligned} \mathcal{L}_{Total} = \lambda_{Dis}(\mathcal{L}_{img} + \mathcal{L}_{attr} + \mathcal{L}_{KL}) + \\ \lambda_{Consis}(\mathcal{L}_{img}^{consis} + \mathcal{L}_{attr}^{consis}) + \lambda_{Adv}\mathcal{L}_{adv} \end{aligned} \quad (12)$$

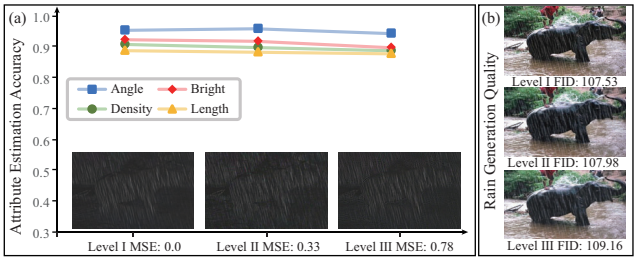


Figure 6: Impact of improper decomposed results on attribute estimation and rain generation quality.

where λ_{Dis}, λ_{Consis} and λ_{Adv} are the balance weight. With the full loss function, our framework collaborates physical disentangle and controllable generation modules with each other to achieve controllable and realistic generation.

3.5. Implementation Details.

The training images are randomly cropped into 128 as input of PCGNet. We first pre-train the decomposition block and attribute regression on the synthetic dataset for 50

Table 2: Quantitative comparison on promoting deraining performance with different paired datasets. 'O', 'C','V','J', and 'P' represent the original data, augmented data by classical augmentations, VRGNet, JRGR, and PCGNet respectively.

Dataset	Index	DDN					JORDER-E					MPRNet				
		O	C	V	J	P	O	C	V	J	P	O	C	V	J	P
Rain1400	PSNR	28.98	28.72	30.71	27.31	**31.25**	32.33	32.35	32.88	31.26	**33.11**	34.43	34.27	34.59	33.29	**34.81**
	SSIM	-0.8897	0.8885	0.9172	0.8802	**0.9233**	0.9256	0.9258	0.9215	0.9105	**0.9355**	0.9443	0.9435	0.9460	0.9232	**0.9483**
	ΔPSNR↑	0	-0.17	1.12	-1.58	**2.36**	0	0.02	0.55	-1.07	**0.78**	0	-0.16	0.16	-1.14	**0.38**
SPA	PSNR	36.16	36.23	37.01	35.46	**38.89**	40.24	39.81	40.64	39.76	**40.85**	47.21	47.15	47.49	46.53	**48.02**
	SSIM	0.9463	0.9471	0.9488	0.9329	**0.9555**	0.9824	0.9800	0.9825	0.9799	**0.9832**	0.9908	0.9902	0.9911	0.9871	**0.9921**
	ΔPSNR↑	-	0.07	0.85	-0.7	**2.73**	-	-0.43	0.4	-0.48	**0.61**	-	-0.06	0.28	-0.67	**0.81**

Table 3: Quantitative comparison on promoting the performance of unpaired real rain removal.

Methods	Rain	DDN					JORDER-E					MPRNet				
		O	C	V	J	P	O	C	V	J	P	O	C	V	J	P
NIQE↓	4.159	4.612	4.587	4.132	3.987	**3.541**	4.816	4.626	5.043	4.035	**3.926**	4.780	4.722	4.598	4.209	**4.012**
User study↑	1.000	4.021	4.233	3.739	5.024	**5.891**	4.295	4.165	4.362	5.132	**6.519**	3.552	3.427	4.930	5.616	**6.312**

Table 4: Promotion on transformed-based method [46].

Dataset	Index	O	C	V	J	P
Rain1400	PSNR↑	33.39	33.36	33.48	32.75	**33.74**
	SSIM ↑	0.9384	0.9378	0.9392	0.9217	**0.9418**
FCRealRain	User study↑	4.21	4.07	4.89	5.61	**6.33**

epochs. Then we joint train PCGNet using paired synthetic data and unpaired real data for 100 epochs, in which the balance weight λ_{Dis}, λ_{Consis}, λ_{Adv} are empirically set as 2, 4, 1. The Adam optimizer [26] is adopted with batch size 12. In testing, we could generate the desired rainy image with clean image and parameters as inputs, and manipulate its attributes with the real rainy image as input.

4. Experiments

4.1. Diverse and Realistic Image Generation

Dataset and experimental setting. We simulate rainy images following the screen blend model [34] with different properties based on the place2 dataset [57] and randomly select 10035 pairs as the synthetic training data and select the FCRealRain [5] as our real data, which has no corresponding clean image and attribute labels. Then we collect 600 clear images of the traffic scene for a generation. For generation methods, we conduct the hand-crafted model Rain100L [49], physical model Rendering [16], generation model JRGR [53], and collected real rain SPA [43] to generate the rain layer on the collected clear image. Importantly, since the Rendering datasets have not released synthetic code and SPA datasets are the real rain, for a fair comparison, we randomly select the rain layer and add it to the collected clear images by the screen blend model.

Qualitative results. Figure 4 presents samples generated by different existing generation models. On one hand, PCGNet has an obvious improvement in photorealism compared with hand-crafted synthetic models such as Rain100L and Rendering. On the other hand, PCGNet could generate more diverse rainy images including different angles, lengths, and brightness by modulating the attribute parameters while the results of JRGR and SPA almost are the same properties. Moreover, compared to existing rainy image generation methods such as JRGR and VRGNet, which can only feed a rainy image to the model, our methods could not only generate the desired rain image with a clean image and parameters as inputs, but also manipulate its attributes with the real rainy image as input such as shown in Fig. 5.

Quantitative results. For a quantitative comparison, we utilize the Frechet Inception Distance (FID) [17] to demonstrate the photorealism and diversity of the generation results. Table 1 shows the FID score on FCRealRain, which is better with a smaller value. As seen, PCGNet generation results achieve the best performance. This is also an expected result since our model has seen the FCRealRain in the training stage and studied the distribution of real rain.

Impact of improper decomposed result on later stage. Note that, subsequent attribute estimation is a coarse prediction problem with four attribute scalars, which inherently are robust to input disturbance. We artificially add image contents onto rain layer to assess the robustness of proposed method in Fig. 6. Then evaluate the impact of improper decomposed results via subsequent attribute regression accuracy and final image generation quality. With decomposed error greatly increasing even to 0.78, where the corresponding restored image PSNR is 28 (most deraining methods could achieve higher), different attributes regression accuracy keep almost unchanged. We further examine final image generation quality, where FID and visual results are also robust based on the improper decomposed result.

4.2. The Promotion for Image Deraining

Dataset and experiment setting. The goal of rain image generation (RIG) is to facilitate the existing datasets for boosting the performance of deraining methods. For paired data, similar to the augmented strategy in [16], we

Figure 7: Performance comparison on the unpaired real rain FCRealRain, including rainy image, deraining result from DSC, and deep SOTAs trained on the original datasets SPA (1^{st} row) and generated dataset by PCGNet (2^{nd} row).

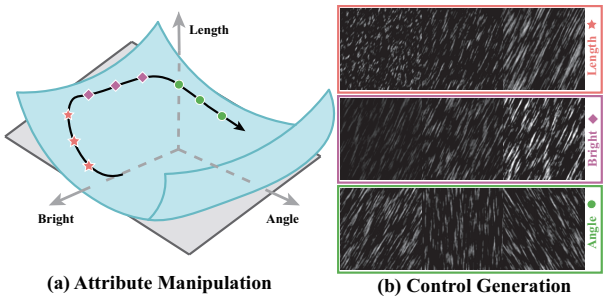
(a) w/o Style Alignment (b) w/ Style Alignment

Figure 8: Effectiveness of style alignment for generation.

conduct classical augmentations (including rotation, contrast, scale), VRGNet [42], JRGR [53] as compared methods to augment the existing datasets with a ratio of 0.5. Note that, since the VRGNet is trained in a supervised way, for a fair comparison, JRGR and PCGNet also utilizes the corresponding synthetic data as our unlabeled training data in the physical decomposition. We evaluate the effectiveness of the augmentation strategy through the latest single image deraining methods, DDN [13], JORDER-E [48], and MPRNet [56] based on commonly paired datasets including Rain1400 [13] and SPA [43]. For unpaired real data, we conduct the real deraining experiment on FCRealRain with the corresponding released and augmented models. In the following, we use the notation 'O', 'C','V','J', and 'P' to denote the original data, augmented data by classical augmentations, VRGNet, JRGR, and PCGNet.

Effectiveness of diverse augmented datasets. For paired data, the diverse training datasets could alleviate the bias of original data and achieve better performance on the testing dataset. Table 2 quantitatively compares the promoting performance of competing methods. On the one hand, the classical augmentation and JRGR could not significantly improve the deraining methods since the linear transform and monomorphic generation. On the other hand, compared to VRGNet, PCGNet can bring more significant improvement since the more diversity of generation data.

Effectiveness of realistic augmented datasets. For unpaired data, the key to improving rain removal methods is to enable the generator results with a similar appearance style to the real rain. In Table 3, we employ the non-reference NIQE [36] and user studies to quantitatively evaluate the visual quality of deraining results on FCRealRain [5]. As for the user study, we randomly select 100 samples and shuffle them, subsequently inviting 30 students to give a score from 1 to 10. The higher user studies and lower NIQE, the better deraining result is. As seen, the training datasets are vital for CNN methods and PCGNet can facilitate a significant improvement for all deraining methods in the real scene. Trained with generation pairs of PCGNet, even the early method DDN can achieve very comparable performance on real rain removal, which can further demonstrate the importance of diverse and photorealism training datasets. Figure 7 shows the qualitative comparison of deraining results be-



(a) Attribute Manipulation (b) Control Generation

Figure 9: Linear manipulation generation. (a) Sampling the value along with the specific dimensions linearly in attribute space. (b) PCGNet can generate rain images in image space that are linearly correlated with attribute parameters.

fore and after augmented by PCGNet. Here we also list optimization-based method DSC [34] for comprehensive comparisons. We can see that results of augmented model possess less rain residual retain more image details.

Promoting on latest deraining method. In order to fully prove the promotion of PCGNet to existing methods, we supplement the experimental results of transformer deraining method IDT [46] on the Rain1400 dataset with different augmentation methods in Table 4. Even with the powerful transformer-based method, incorporating the data generated by our approach (P) can further improve the performance compared to the baseline. This strongly validates the general and practical ability of the proposed method.

4.3. Ablation Study and Discussing

Effectiveness of style alignment and SMT for generation. As for style alignment, Fig. 8 shows samples generated by PCGNet with or without style alignment. We observe a large improvement in photorealism when style alignment is added. As for SMT, The accuracy of attribute estimation on real data will directly affect the quality of real rain reconstruction. To verify the impact of SMT on attribute estimation, in Tabel 5, we show the accuracy of attribute estimation. We manually annotated attributes of 500 images from test set of Rain1400 for attribute estimation validation, which is different from our synthetic data. We can observe that SMT can significantly help attribute encoder to regress better rain properties.

Impact of style alignment and SMT for rain removal. The key goal of rain generation is to facilitate rain removal. Here, we further validate the impact of style alignment and SMT on rain removal performance of JORDER-E [48]. As shown in Table 6, style alignment facilitates learning distribution of Rain1400 with strong improvement, and SMT enhances sufficient promotion by better attribute regression.

Effectiveness of consistency and adversarial losses. Although final loss functions may appear complex in Eq. (12), it can be divided into three main categories: physical disentangle loss, consistency loss, and adversarial loss. In Table 7, we analyze the effect of $\mathcal{L}_{attr}^{consis}$ and $\mathcal{L}_{adv}$ from two aspects: rain attributes accuracy (measurement of controllable ability) and image quality (FID, measurement of realism degree) of generation results. We can conclude that $\mathcal{L}_{attr}^{consis}$ significantly improves controllable generation accuracy, while $\mathcal{L}_{adv}$ enhances the realism of the results.

Table 5: Effects of SMT on attribute estimation accuracy.

Index	Angle	Length	Brightness	Density
w/o SMT	92.3%	82.3%	84.5%	70.2%
w/ SMT	93.5%	88.2 %	85.1%	84.7%

Table 6: Effects of style alignment and SMT for deraining.

Index	SA(×) SMT(×)	SA(√) SMT(×)	SA(√) SMT(√)
PSNR/SSIM	30.15/0.9092	32.56/0.9298	**33.11/0.9355**

Table 7: Effects of consistency and adversarial losses.

Index	$\mathcal{L}_{attr}^{consis}(\times)$ $\mathcal{L}_{adv}(\times)$	$\mathcal{L}_{attr}^{consis}(\times)$ $\mathcal{L}_{adv}(\surd)$	$\mathcal{L}_{attr}^{consis}(\surd)$ $\mathcal{L}_{adv}(\times)$	$\mathcal{L}_{attr}^{consis}(\surd)$ $\mathcal{L}_{adv}(\surd)$
Accuracy↑	0.711	0.739	0.862	**0.895**
FID↓	138.5	115.4	125.7	**108.28**

Attribute interpolation. Different from existing RIG methods implicitly control such as VRGNet [42], proposed methods can explicitly and linearly control rain attributes. As shown in Fig. 9, when operating attribute (such as angle, length, and brightness) varies linearly, physical attributes of rain streak in results are also changed linearly, which demonstrates superior controllable generation of PCGNet.

Limitations. Currently, we mainly focus on enhancing realism of rain through explicit modeling of its semantic attributes. However, authenticity of rain should encompass not only rain semantic but also scene semantic. We will take 3D scene semantic into consideration in the future works.

5. Conclusion

In this paper, we attempt to facilitate real rain removal from data generation perspective. To achieve this goal, we propose a physical alignment and controllable generation network (PCGNet) with diverse and realistic generation results. The proposed PDA decouples synthetic and real rainy images into background, rain attributes, and appearance styles, and makes them aligned in physical space by weight Step-moving and distribution resampling strategy. The proposed CGM achieves physically controllable mapping from attributes to real rain for realistic and diverse rain generation. Extensive comparison results have shown appealing generation results, which significantly improve the performance of other derain methods on real data.

Acknowledgements. This work was supported in part by the National Natural Science Foundation of China under Grant 61971460 and Grant 62101294, in part by JCJQ Program under Grant 2021-JCJQ-JJ-0060, and in part by the Fundamental Research Funds for the Central Universities, HUST: 2022JYCXJJ001.

References

[1] M. Arjovsky, S. Chintala, and L. Bottou. Wasserstein generative adversarial networks. In *Int. Conf. Mach. Learn.*, pages 214–223, 2017. 2

[2] Y. Ba, H. Zhang, E. Yang, A. Suzuki, A. Pfahnl, C. Chandrappa, C. de Melo, S. You, S. Soatto, A. Wong, and A. Kadambi. Not just streaks: Towards ground truth for single image deraining. In *Eur. Conf. Comput. Vis.*, pages 723–740, 2022. 2, 3

[3] J. Bao, D. Chen, F. Wen, H. Li, and G. Hua. Cvae-gan: Fine-grained image generation through asymmetric training. In *Int. Conf. Comput. Vis.*, pages 2745–2754, 2017. 3

[4] J. Bao, D. Chen, F. Wen, H. Li, and G. Hua. Towards open-set identity preserving face synthesis. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 6713–6722, 2018. 3

[5] Y. Chang, Y. Guo, Y. Ye, C. Yu, L. Zhu, X. Zhao, L. Yan, and Y. Tian. Unsupervised deraining: Where asymmetric contrastive learning meets self-similarity. *arXiv preprint arXiv:2211.00837*, 2022. 7, 8

[6] Y. Chang, L. Yan, and S. Zhong. Transformed low-rank model for line pattern noise removal. In *Int. Conf. Comput. Vis.*, pages 1726–1734, 2017. 2

[7] X. Chen, J. Pan, K. Jiang, Y. Li, Y. Huang, C. Kong, L. Dai, and Z. Fan. Unpaired deep image deraining using dual contrastive learning. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 2017–2026, 2022. 1, 2

[8] Y. Chen and C. Hsu. A generalized low-rank appearance model for spatio-temporally correlated rain streaks. In *Int. Conf. Comput. Vis.*, pages 1968–1975, 2013. 2

[9] C. Cheng, A. Koschan, C. Chen, D. Page, and M. Abidi. Outdoor scene image segmentation based on background recognition and perceptual organization. *IEEE Trans. Image Process.*, 21(3):1007–1019, 2011. 1

[10] S. Deng, M. Wei, J. Wang, Y. Feng, L. Liang, H. Xie, F. Wang, and M. Wang. Detail-recovery image deraining via context aggregation networks. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 14560–14569, 2020. 1

[11] Y. Deng, J. Yang, D. Chen, F. Wen, and X. Tong. Disentangled and controllable face image generation via 3d imitative-contrastive learning. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 5154–5163, 2020. 3

[12] C. Doersch. Tutorial on variational autoencoders. *arXiv preprint arXiv:1606.05908*, 2016. 5

[13] X. Fu, J. Huang, D. Zeng, Y. Huang, X. Ding, and J. Paisley. Removing rain from single images via a deep detail network. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 3855–3863, 2017. 1, 2, 8

[14] K. Garg and S. Nayar. Vision and rain. *Int. J. Comput. Vis.*, 75(1):3–27, 2007. 3

[15] A. Geiger, P. Lenz, and R. Urtasun. Are we ready for autonomous driving? the kitti vision benchmark suite. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 3354–3361, 2012. 1

[16] S. Halder, J. Lalonde, and R. Charette. Physics-based rendering for improving robustness to rain. In *Int. Conf. Comput. Vis.*, pages 10203–10212, 2019. 2, 3, 7

[17] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. In *Adv. Neural Inform. Process. Syst.*, pages 6629–6640, 2017. 7

[18] I. Higgins, L. Matthey, A. Pal, C. Burgess, X. Glorot, M. Botvinick, S. Mohamed, and A. Lerchner. Beta-vae: Learning basic visual concepts with a constrained variational framework. In *Int. Conf. Learn. Represent.*, pages 1–22, 2017. 3

[19] X. Hu, C. Fu, L. Zhu, and P. Heng. Depth-attentional features for single-image rain removal. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 8022–8031, 2019. 1, 2, 3

[20] H. Huang, A. Yu, and R. He. Memory oriented transfer learning for semi-supervised image deraining. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 7732–7741, 2021. 4

[21] X. Huang and S. Belongie. Arbitrary style transfer in real-time with adaptive instance normalization. In *Int. Conf. Comput. Vis.*, pages 1501–1510, 2017. 5

[22] K. Jiang, Z. Wang, P. Yi, C. Chen, B. Huang, Y. Luo, J. Ma, and J. Jiang. Multi-scale progressive fusion network for single image deraining. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 8346–8355, 2020. 2

[23] X. Jin, Z. Chen, J. Lin, Z. Chen, and W. Zhou. Unsupervised single image deraining with self-supervised constraints. In *IEEE Int. Conf. Image Process.*, pages 2761–2765, 2019. 1, 2

[24] L. Kang, C. Lin, and Y. Fu. Automatic single-image-based rain streaks removal via image decomposition. *IEEE Trans. Image Process.*, 21(4):1742–1755, 2012. 2

[25] H. Kim and A. Mnih. Disentangling by factorising. In *Int. Conf. Mach. Learn.*, pages 2649–2658, 2018. 3

[26] D. Kingma and J. Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 7

[27] A. Kumar, P. Sattigeri, and A. Balakrishnan. Variational inference of disentangled latent concepts from unlabeled observations. *arXiv preprint arXiv:1711.00848*, 2017. 3

[28] R. Li, L. Cheong, and R. Tan. Heavy rain image restoration: Integrating physics model and conditional adversarial learning. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 1633–1642, 2019. 1, 2, 3

[29] S. Li, I. Araujo, W. Ren, Z. Wang, E. Tokuda, R. Junior, R. Cesar-Junior, J. Zhang, X. Guo, and X. Cao. Single image deraining: A comprehensive benchmark analysis. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 3838–3847, 2019. 1

[30] X. Li, J. Wu, Z. Lin, H. Liu, and H. Zha. Recurrent squeeze-and-excitation context aggregation net for single image deraining. In *Eur. Conf. Comput. Vis.*, pages 254–269, 2018. 2

[31] Y. Li, R. Tan, X. Guo, J. Lu, and M. Brown. Rain streak removal using layer priors. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 2736–2744, 2016. 2

[32] J. Liu, W. Yang, S. Yang, and Z. Guo. Erase or fill? deep joint recurrent rain removal and reconstruction in videos. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 3233–3242, 2018. 2, 3

[33] W. Liu, G. Ren, R. Yu, S. Guo, J. Zhu, and L. Zhang. Image-adaptive yolo for object detection in adverse weather conditions. In *AAAI Conf. Artif. Intell.*, pages 1792–1800, 2022. 1

[34] Y. Luo, Y. Xu, and H. Ji. Removing rain from a single image via discriminative sparse coding. In *Int. Conf. Comput. Vis.*, pages 3397–3405, 2015. 2, 7, 9

[35] M. Mirza and S. Osindero. Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*, 2014. 3

[36] A. Mittal, R. Soundararajan, and A. Bovik. Making a "completely blind" image quality analyzer. *IEEE Sign. Process. Letters*, 20(3):209–212, 2012. 8

[37] S. Ni, X. Cao, T. Yue, and X. Hu. Controlling the rain: From removal to rendering. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 6328–6337, 2021. 2

[38] B. Pang, D. Zhai, J. Jiang, and X. Liu. Single image deraining via scale-space invariant attention neural network. In *ACM Int. Conf. Multimedia*, pages 375–383, 2020. 2

[39] D. Ren, W. Zuo, Q. Hu, P. Zhu, and D. Meng. Progressive image deraining networks: A better and simpler baseline. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 3937–3946, 2019. 1

[40] G. Wang, C. Sun, and A. Sowmya. Erl-net: Entangled representation learning for single image de-raining. In *Int. Conf. Comput. Vis.*, pages 5644–5652, 2019. 1

[41] H. Wang, Q. Xie, Q. Zhao, and D. Meng. A model-driven deep neural network for single image rain removal. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 3103–3112, 2020. 1, 2

[42] H. Wang, Z. Yue, Q. Xie, Q. Zhao, Y. Zheng, and D. Meng. From rain generation to rain removal. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 14791–14801, 2021. 1, 2, 3, 8, 9

[43] T. Wang, X. Yang, K. Xu, S. Chen, Q. Zhang, and R. Lau. Spatial attentive single-image deraining with a high quality real rain dataset. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 12270–12279, 2019. 1, 2, 3, 7, 8

[44] W. Wei, D. Meng, Q. Zhao, Z. Xu, and Y. Wu. Semi-supervised transfer learning for image rain removal. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 3877–3886, 2019. 1, 2

[45] Y. Wei, Z. Zhang, Y. Wang, M. Xu, Y. Yang, S. Yan, and M. Wang. Deraincyclegan: Rain attentive cyclegan for single image deraining and rainmaking. *IEEE Trans. Image Process.*, 30:4788–4801, 2021. 1, 2, 3

[46] J. Xiao, X. Fu, A. Liu, F. Wu, and Z. Zha. Image de-raining transformer. *IEEE Trans. Pattern Anal. Mach. Intell.*, pages 1–13, 2022. 2, 7, 9

[47] W. Yan, A. Sharma, and R. Tan. Optical flow in dense foggy scenes using semi-supervised learning. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 13259–13268, 2020. 1

[48] W. Yang, R. Tan, J. Feng, Z. Guo, S. Yan, and J. Liu. Joint rain detection and removal from a single image with contextualized deep networks. *IEEE Trans. Pattern Anal. Mach. Intell.*, 42(6):1377–1393, 2019. 1, 2, 8, 9

[49] W. Yang, R. Tan, J. Feng, J. Liu, Z. Guo, and S. Yan. Deep joint rain detection and removal from a single image. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 1357–1366, 2017. 1, 2, 3, 7

[50] W. Yang, R. Tan, S. Wang, Y. Fang, and J. Liu. Single image deraining: From model-based to data-driven and beyond. *IEEE Trans. Pattern Anal. Mach. Intell.*, 43(11):4059–4077, 2020. 1

[51] R. Yasarla and V. Patel. Uncertainty guided multi-scale residual learning-using a cycle spinning cnn for single image deraining. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 8405–8414, 2019. 2

[52] R. Yasarla, V. Sindagi, and V. Patel. Syn2real transfer learning for image deraining using gaussian processes. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 2726–2736, 2020. 1, 2

[53] Y. Ye, Y. Chang, H. Zhou, and L. Yan. Closing the loop: Joint rain generation and removal via disentangled image translation. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 2053–2062, 2021. 1, 2, 3, 7, 8

[54] Y. Ye, C. Yu, Y. Chang, L. Zhu, X. Zhao, L. Yan, and Y. Tian. Unsupervised deraining: Where contrastive learning meets self-similarity. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 5821–5830, 2022. 1, 2

[55] C. Yu, Y. Chang, Y. Li, X. Zhao, and L. Yan. Unsupervised image deraining: Optimization model driven deep cnn. In *ACM Int. Conf. Multimedia*, pages 2634–2642, 2021. 1, 2, 3

[56] S. Zamir, A. Arora, S. Khan, M. Hayat, F. Khan, M. Yang, and L. Shao. Multi-stage progressive image restoration. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 14821–14831, 2021. 2, 8

[57] B. Zhou, A. Lapedriza, A. Khosla, A. Oliva, and A. Torralba. Places: A 10 million image database for scene recognition. *IEEE Trans. Pattern Anal. Mach. Intell.*, 40(6):1452–1464, 2017. 7

[58] H. Zhu, X. Peng, J. Zhou, S. Yang, V. Chanderasekh, L. Li, and J. Lim. Singe image rain removal with unpaired information: A differentiable programming perspective. In *AAAI Conf. Artif. Intell.*, pages 9332–9339, 2019. 1