Unpaired Deep Image Deraining Using Dual Contrastive Learning

Xiang Chen [1,2], Jinshan Pan [2], Kui Jiang [3], Yufeng Li [1], Yufeng Huang [1],
Caihua Kong [1], Longgang Dai [1], Zhentao Fan [1]

[1] Shenyang Aerospace University, [2] Nanjing University of Science and Technology, [3] Wuhan University

https://cxtalk.github.io/projects/DCD-GAN.html

Abstract

Learning single image deraining (SID) networks from an unpaired set of clean and rainy images is practical and valuable as acquiring paired real-world data is almost infeasible. However, without the paired data as the supervision, learning a SID network is challenging. Moreover, simply using existing unpaired learning methods (e.g., unpaired adversarial learning and cycle-consistency constraints) in the SID task is insufficient to learn the underlying relationship from rainy inputs to clean outputs as there exists significant domain gap between the rainy and clean images. In this paper, we develop an effective unpaired SID adversarial framework which explores mutual properties of the unpaired exemplars by a dual contrastive learning manner in a deep feature space, named as DCD-GAN. The proposed method mainly consists of two cooperative branches: Bidirectional Translation Branch (BTB) and Contrastive Guidance Branch (CGB). Specifically, BTB exploits full advantage of the circulatory architecture of adversarial consistency to generate abundant exemplar pairs and excavates latent feature distributions between two domains by equipping it with bidirectional mapping. Simultaneously, CGB implicitly constrains the embeddings of different exemplars in the deep feature space by encouraging the similar feature distributions closer while pushing the dissimilar further away, in order to better facilitate rain removal and help image restoration. Extensive experiments demonstrate that our method performs favorably against existing unpaired deraining approaches on both synthetic and real-world datasets, and generates comparable results against several fully-supervised or semi-supervised models.

1. Introduction

Images captured under complicated rain weather environments often suffer from unfavorable visibility by rain streaks. Such these degraded images usually affect many computer vision tasks (including detection [20], segmen-
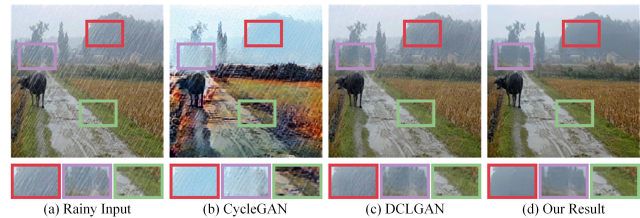

Figure 1. Derained results on a real-world rainy image by unpaired learning approaches. Directly employing existing unpaired learning methods, *e.g.*, CycleGAN and DCLGAN, does not effecively remove rain streaks.

tation [38] and video surveillance [28], etc.) with drastic performance drop. Thus, it is of great interest to develop an effective algorithm to recover high-quality rain-free images.

In general, the rain process is usually modeled by the following linear superimposition model:

$$O = B + R, \tag{1}$$

where O, B, and R denote the rainy image, clean image, and rain streaks, respectively. The goal of single image deraining (SID) is to estimate clean image B from the input rainy image O. This is an ill-posed problem as only rainy image is known. To make the problem well-posed, conventional methods [17, 19, 23, 47] usually impose certain priors on the clean images and rain components. Although decent results have been achieved, the priors are based on empirical statistical results and may not model the inherent properties of the clean images and rain components, which thus do not effectively remove the rain well.

Recently, numerous data-driven learning methods have developed [15, 27, 42, 45, 49]. Although remarkable performance has been achieved, these fully-supervised methods need paired synthetic data which does not model the real-world degradation well. Therefore, these methods usually do not perform well when handling the real-world rainy images due to the domain gap between the training and test data. Moreover, obtaining large scale paired real-world data for training in complex rainy environments is challenging.

To overcome these problems, the semi-supervised [13,

22, 36, 43] and unsupervised [9, 51] learning have been proposed for SID. These methods [13, 16] either focus on domain invariant features by leveraging the limited labeled data and introducing the auxiliary optimization objectives or develop domain adaption strategies [9, 37] to improve the generalization capabilities of the deep models. However, without suitable constraints for rain streaks and clean images, existing unpaired learning methods [8, 52] do not effectively restore high-quality derained results, as shown in Fig. 1 (b) and (c). Note that, these methods mainly consider the mapping relationship in the image space but ignore the potential relationship in the feature space, which does not fully excavate the useful feature information for SID. Since the ground truth labeled data is not fully available, how to model the latent-space representation by exploring the relationship between the rainy inputs and clean outputs is important for the deep learning-based methods. In addition, given that clean images can be easily obtained, it is also important to develop an effective method that can explore properties of the clean exemplars to facilitate image restoration when paired data is not available.

Towards this end, we develop a Dual Contrastive Derain-GAN (DCD-GAN) method which incorporates recent contrastive learning with adversarial framework to explore useful features from the rainy images and unpaired clean images so that the extracted features can better facilitate rain removal. Ideally, we note that if a deep model can accurately restore a clean image from rainy one, the features that are for clean image reconstruction would have mutual information with the ones from the ground truth rain-free images by the same deep model. This motivates us to introduce a contrastive learning method to mining the mutual features of rainy images and clean ones in the deep feature space, so that we can use the features from the clean images to better guide the image restoration. Therefore, our proposed DCD-GAN includes two main interactive branches: Bidirectional Translation Branch (BTB) and Contrastive Guidance Branch (CGB). Specifically, BTB is equipped with bidirectional mapping to mine rain-related or clean-cue features, and the use of adversarial training produces rich exemplar pairs during the optimization. In addition, CGB implicitly constrains the latent space of corresponding patches to guide deraining by encouraging the positives (i.e, similar feature distributions) closer while keeping the negatives (i.e, dissimilar ones) further away. To summarize, the main contributions of this paper are summarized as follows:

- We formulate an effective DCD-GAN which leverages dual contrastive learning to encourage the model to explore mutual features while distinguish the dissimilar ones between the rainy domain and the rain-free domain in the deep feature space.

- The proposed DCD-GAN is performed without paired

training information, where the features from the unpaired clean exemplars can facilitate rain removal. The learned latent restoration can boost the cross-domain deraining generalization performance.

- Experimental results on challenging datasets considerably show that our developed method performs favorably against existing unpaired deraining approaches, and achieves comparable performance against several fully-supervised or semi-supervised models.

2. Related Work

2.1. Single Image Deraining

Existing deep learning-based SID methods are categorized into paired (fully-supervised), semi-supervised and unpaired (without paired supervised) approaches [41].

For the paired deraining methods, Fu *et al.* [5] first employ the Derain Net with multi-layer CNN to extract and remove the rain layer, and further introduce deep detail network (DDN) [6] that directly removes the rain streaks by reducing the mapping range. Later, Zhang *et al.* [48] present a density-aware multi-stream densely connected CNN algorithm, termed DID-MDN, to characterize rain streaks by estimating rain density. The classic work of RESCAN [18] introduces a convolutional and recurrent neural network-based way to make full use of the contextual information. Spatial attentive network (SPANet) [34] captures the spatial contextual information based on the recurrent network and applies a branch to obtain the spatial details in a local-to-global manner. By replacing low-quality features by latent high-quality features, a robust representation learning network is proposed by [1] to address deraining model errors.

For the semi-supervised learning-based methods, Wei *et al.* [36] utilize a semi-supervised learning framework (SSIR) to analyze the residual difference between domains. The work of Syn2Real [43] develops a Gaussian process-based semi-supervised method, which can increase the network ability by using synthetic images and unlabeled real rainy images. In [22], a rain direction regularizer is designed to constrain the network by combining the semi-supervised learning part and the knowledge distillation part.

For the unpaired deraining methods, CycleGAN [52] has been widely employed to address the unpaired image-to-image translation issue, to achieve deraining process. Recent works [9, 16, 37, 51] all use the improved CycleGAN architecture and constrained transfer learning to jointly the rainy and rain-free image domains. As mentioned before, it is difficult for these methods to learn accurate transformation between two domains using only cycle-consistency constraints, since their domain knowledge is asymmetrical. Instead, our approach can overcome this limitation and incorporates contrastive learning with adversarial training to further improve deraining robustness in unpaired setting.
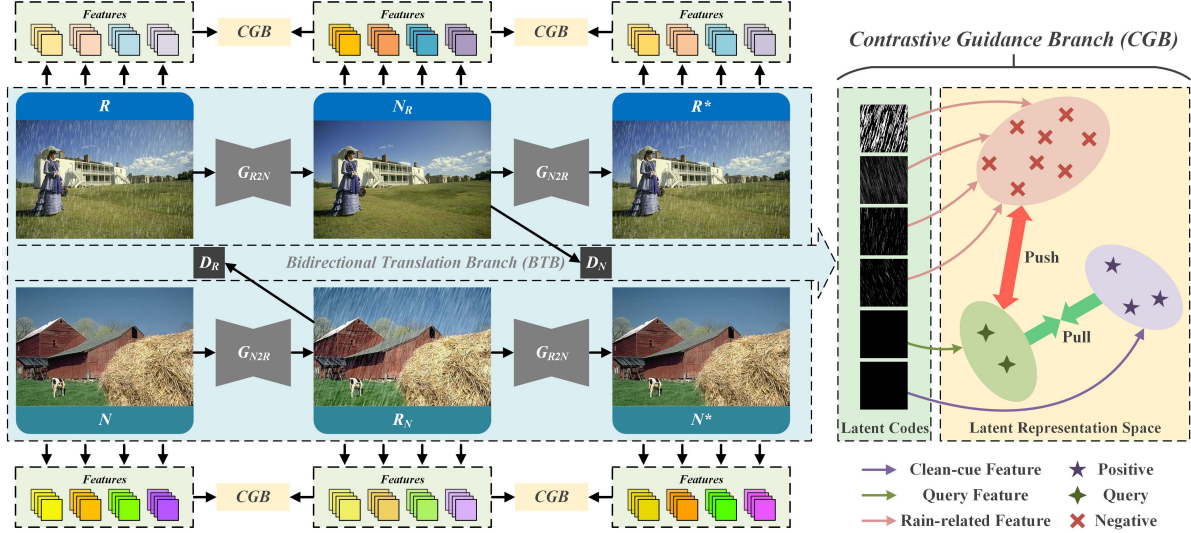
Figure 2. The overall framework of our Dual Contrastive Derain-GAN (DCD-GAN). There are two cooperative branches, bidirectional translation branch (BTB) and contrastive guidance branch (CGB). In the feature space, images generated from close ("positive") latent codes are visually similar, while images generated from far-away ("negative") latent codes are visually disimilar. Our developed CGBs aim to learn a representation to pull similar feature distribution (*e.g.*, clean →← clean) and push disimilar (*e.g.*, rain ←→ clean) apart.

2.2. Contrastive Learning

Contrastive learning has witnessed a significant progress in the field of self-supervised and unsupervised representation learning. Instead of using fixed targets and pre-defined ones, contrastive regularization maximizes the mutual features between different domains by exploiting both the information of positive pairs and negative pairs. Specifically, it aims to learn suitable embeddings by pulling the exemplar close to positive samples while pushing it far away from negative samples in the related representation space. Recent studies have applied contrastive losses in a host of low-level vision tasks and achieved SOTA performance, *e.g.*, image dehazing [39], image denoising [4], image super-resolution [50] and underwater image restoration [7]. Similarly, the recent work of DCLGAN [8] takes the advantages of both CUT [26] and CycleGAN [52] to cope with unsupervised image-to-image translation problem. Different from its mutual mapping in complex image space, we show the use of contrastive constraints to perform unpaired learning in deep feature space, which can better help image restoration.

3. Proposed Method

Fig. 2 presents the overall architecture of our proposed DCD-GAN. It consists of BTB and CGB branches. BTB is used to guide transfer ability and latent feature distributions between the rainy domain and the rain-free domain. Furthermore, CGBs are adopted to constrain the mutual feature information between two domains, as well as the multiple loss functions to provides jointly optimization. We describe the details about the proposed network in the following.

3.1. Bidirectional Translation Branch (BTB)

Let $\mathbb{D}_R$ denotes the set of the rainy images without ground truth labels and $\mathbb{D}_N$ denotes the set of clean exemplar images, our final goal is to learn a deep model to explore the features from these unpaired data $\mathbb{D}_R$ and $\mathbb{D}_N$ without the supervision of the ground truth labels to estimate the rain-free images. To this end, we introduce a bidirectional translation branch (BTB) as the backbone of the proposed DCD-GAN. Specifically, BTB contains two generators G_{R2N}, G_{N2R} generating the clean and rainy images respectively and two discriminators D_R, D_N distinguishing between fake derained images and real clean images. Continuing along the trajectory of [37, 44], the overall pipeline of BTB covers two circuits of rain generation and rain removal: (1) rainy to rainy forward cycle-consistency transformation $R \rightarrow N_R \rightarrow R^*$; (2) rain-free to rain-free backward cycle-consistency transformation $N \rightarrow R_N \rightarrow N^*$.

Due to the circulatory architecture of bidirectional learning, the exemplars generated by BTB could be mined rain-related or clean-cue features during the optimization, thus forming latent feature distributions to explore regularization. Notably, unpaired clean exemplars from the backward cycle can facilitate rain removal. We will show the influence of the clean exemplars in Section 4.3. In fact, it will be not sufficient to utilize cycle-consistency to remove rain, in that this constraint is weak. Hence, we develop contrastive constraints to formulate multiple contrastive guidance branches (CGBs) that serve as internal bridges into the BTB, yielding more faithful solutions (good generation and good reconstruction). The CGB component is described below.
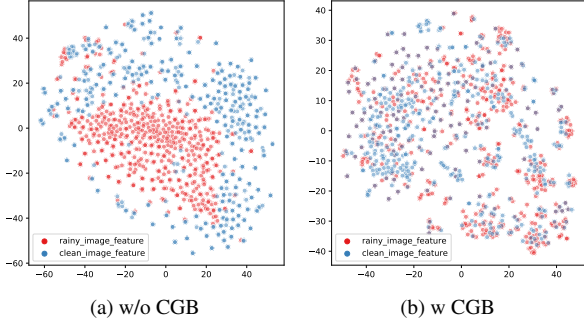
(a) w/o CGB   (b) w CGB

Figure 3. The t-SNE visualization of features learned with/without our CGB. The red round point denotes the deep features extracted from rainy images for clean image reconstruction. The blue round points denote the deep features extracted from clean images. With CGBs, the embeddings of the features from the clean image and the rainy image have low distances in the latent space. Thus, using the rainy features by CGBs is able to generate clean images.

3.2. Contrastive Guidance Branch (CGB)

To guide in better deraining, CGBs are involved between $\mathbb{D}_R$ and $\mathbb{D}_N$ to exploit advantage of the learned domain-relevant feature distribution. Our primary insight is that the features extracted from $\mathbb{D}_R$ share some mutual properties to the features extracted from $\mathbb{D}_N$. If pairs of patches generated from different domains with low feature disparity are similar (referred to as "positive"), their embeddings in latent space should also have a low distance, and vice-versa. Therefore, we propose to learn the image restoration in the deep feature space by enforcing contrastive constraints.

Following the advantage of dual learning setting in stability training [46], we take the auxiliary encoder E as separate embeddings to extract features from domain $\mathbb{D}_R$ and $\mathbb{D}_N$, respectively. As suggested in [21,26], by using off-the-shelf generators to extract feature representations in contrastive learning, excessive modifications to the original architecture can be avoided. Consequently, we directly extract features of images from the L encoding layers of the generator (G_{R2N} and G_{N2R}) as the auxiliary encoder of CGB and send them to a two-layer multi-layer perceptron (MLP) projection head. Here, we do not share weights, thus capturing variability in both domains and learning better embeddings. After having different stacks of features, our CGB constrains the contrastive distribution learning with two distinct embeddings, with the goal of associating a query generated exemplar and its similar feature distribution (*e.g.*, clean $\rightarrow\leftarrow$ clean) of the representations, in contrast to other dissimilar ones (*e.g.*, rain $\leftarrow\rightarrow$ clean) at the same time for implicitly supervising the deraining process.

To constrain mutual features in their embeddings, let us define the query code $\hat{x}$ sampled from one domain, the positive code $\hat{x}^+$ and k negative codes $\left\{\hat{x}_i^-\right\}_{i-1}^k$ sampled from the other domain. Different from previous methods [26,39] that only select negative samples within the input image, we

focus on both internal and external latent codes concatenating the two sides of the CGB within the feature space, and experiments demonstrate our strategy achieves better benefits for SID (see Section 4.3). The constraint of CGB is enforced by encouraging the positives closer while keeping the negatives further away, so as to help the deraining generators guide correctly in transforming the rainy inputs to the clean outputs. To understand the effect of such CGBs, we further use t-SNE [32] to visualize learned features in Fig. 3. Ideally, if a deep model can accurately restore a clean image from rainy one, the features that are for clean image reconstruction and the feature of ground truth rain-free images will completely overlap in the deep feature space. Obviously, with CGBs, these two types of features are pulled together, which become well distinguished and able to capture the relationship between $\mathbb{D}_R$ and $\mathbb{D}_N$. Mathematically, these corresponding feature representations are then formalized as: $f = E(\hat{x})$, $f^+ = E\left(\hat{x}^+\right)$, $f_i^- = E\left(\hat{x}_i^-\right)$. Thus, the **Contrastive Loss** is developed to regularize the captured images' feature distributions, which can be written as:

$$L_{cont}\left(G_{R2N}, G_{N2R}\right) =$$
$$\mathbb{E}_{r\sim R, n\sim N}\left[-\log \frac{\text{sim}\left(f, f^+\right)}{\text{sim}\left(f, f^+\right) + \sum_{i=1}^N \text{sim}\left(f, f_i^-\right)}\right],$$
(2)

where $\text{sim}(u, v) = \exp\left(\frac{u^T v}{\|u\|\|v\|\tau}\right)$ is the cosine similarity function that computes the similarity between two normalized feature vectors, and τ denotes a temperature parameter.

3.3. Loss Function

As the ground truths are not available, it is vital to develop effective loss function to constrain the DCD-GAN for better SID. In addition to the contrastive loss, we further introduce other constraints to regularize the network training.

Color Cycle-Consistency Loss. It is worth noting that previous CycleGAN-based methods mostly adopt the cycle-consistency loss to preserve image content. However, there exists the problem of "channel pollution" [29,30] in cycle-consistency loss, because generating a whole image at one time makes different channels interact with each other, resulting in unpleasant artifacts in final derained results. To alleviate it, we develop a simple yet effective solution by constructing the consistence loss for each channel independently. Here, we calculate the pixel loss of the red, green and blue channels between the input and the reconstructed image separately, and then sum the three distance losses as the color cycle-consistency loss, which can be expressed as:

$$L_{colorcyc} = \sum_{i\in\{r,g,b\}} \mathbb{E}_{n\sim N}^i\left[\|G_{R2N}\left(G_{N2R}(n)\right) - n\|_1\right]$$
$$+ \sum_{i\in\{r,g,b\}} \mathbb{E}_{r\sim R}^i\left[\|G_{N2R}\left(G_{R2N}(r)\right) - r\|_1\right],$$
(3)

Table 1. Comparison of quantitative results on four benchmark datasets. Bold and *italic* indicate the best and second-best results.

Datasets		*Test100*		*Test1200*		*Test1400*		*Test1000*	
Metrics		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Prior-based methods	DSC [23]	18.56	0.599	24.24	0.827	27.31	0.837	34.95	0.941
	GMM [19]	20.46	0.730	25.66	0.817	26.87	0.808	34.30	0.942
Paired / Supervised methods	DDN [6]	21.16	0.732	27.93	0.853	28.00	0.873	34.70	0.934
	DID-MDN [48]	21.89	0.795	29.66	0.899	26.38	0.835	34.68	0.930
	RESCAN [18]	24.09	0.841	*32.25*	0.907	**32.03**	*0.917*	34.71	0.937
	SPANet [34]	*24.37*	**0.861**	30.05	*0.934*	29.83	0.904	35.13	*0.944*
Semi-supervised methods	SIRR [36]	/	/	30.57	0.910	30.01	0.907	34.85	0.936
	Syn2Real [43]	23.74	0.799	/	/	/	/	/	/
Unpaired / Without paired supervised methods	CycleGAN [52]	21.92	0.683	24.17	0.729	24.34	0.776	31.59	0.896
	RR-GAN [51]	23.51	0.757	/	/	/	/	/	/
	DerainCycleGAN [37]	24.30	*0.854*	/	/	/	/	*35.20*	**0.950**
	Ours	**25.61**	0.813	**32.46**	**0.937**	*31.65*	**0.918**	**35.30**	0.943

Table 2. Descriptions of different benchmark training datasets.

Datasets	Rain800	DID-Data	DDN-Data	SPA-Data
Train-Set	700	12000	12600	28500
Test-Set	100	1200	1400	1000
Type		Synthetic		Real-world
Name	*Test100*	*Test1200*	*Test1400*	*Test1000*

where i denotes different channels. $r \in \mathbb{D}_R$, $n \in \mathbb{D}_N$ are a rainy image and an unpaired clean image.

Adversarial Loss. We observed that employing the adversarial loss is able to generate the realistic images. Similar to [51], the adversarial loss in $\mathbb{D}_N$ for SID is defined as:

$$L_{adv}(G_{R2N}, D_N) = \mathbb{E}_{n \sim N}[\log D_N(n)] \\ + \mathbb{E}_{r \sim R}[\log(1 - D_N(G_{R2N}(r)))], \quad (4)$$

where G_{R2N} minimizes the objective function to make the generated clean images look similar to real samples. In contrast, D_N maximizes the loss to distinguish generated clean images and real clean images. Here, we omit same adversarial loss in $\mathbb{D}_R$. The overall adversarial loss is calculated by $L_{adv} = L_{adv}(G_{R2N}, D_N) + L_{adv}(G_{N2R}, D_R)$.

Frequency Loss. Inspired by [2, 12], we take advantage of frequency domain and exploit a frequency loss to further preserve the content and structure information of the image:

$$L_{freq} = \mathbb{E}_{n \sim N}[\|FT(G_{R2N}(G_{N2R}(n))) - FT(n)\|_2^2] \\ + \mathbb{E}_{r \sim R}[\|FT(G_{N2R}(G_{R2N}(r))) - FT(r)\|_2^2], \quad (5)$$

where FT denotes one-sided Fourier transform to obtain their corresponding frequency domain.

Based on above considerations, the final loss function is:

$$L_{total} = \lambda_1 L_{cont} + \lambda_2 L_{colorcyc} + \lambda_3 L_{adv} + \lambda_4 L_{freq}, \quad (6)$$

where λ_i is the trade-off weight. In our experiments, we empirically set $\lambda_1 = 2$, $\lambda_2 = 1$, $\lambda_3 = 1$, and $\lambda_4 = 0.1$.

4. Experiments

4.1. Experimental Settings

Datasets. We use four challenging benchmark datasets including Rain800 [49], DID-Data [48], DDN-Data [6], and SPA-Data [34] with various rain streaks of different sizes, shapes and directions. The detailed descriptions of the used training and testing datasets are tabulated in Tab. 2.

Implementation Details. The overall architecture of DCD-GAN is based on CycleGAN, a Resnet-based [10] generator with 9 residual blocks and a PatchGAN [14] discriminator. During training, we use the Adam optimizer with $\beta_1 = 0.5$ and $\beta_2 = 0.999$ and the models are trained for total 600 epochs. The model is trained from the scratch for 300 epochs with the learning rate of 0.0001, followed by another 300 epochs with the learning rate linearly decayed to 0. In the CGB, we select 255 internal and 256 external negative samples for one query. The temperature parameter τ is set to 0.07. We use a batch size of 1 and instance normalization [31]. The entire network is performed using PyTorch framework on two NVIDIA Tesla V100 GPUs. It should be noted that all training images are randomly cropped to 256×256 patches in an unpaired learning manner. To better facilitate contrastive learning, we have implemented specific data enhancement strategies for the above benchmarks.

Comparison Methods. We compare our method with those of two prior-based algorithms (i.e., DSC [23] and GMM [19]), four paired supervised methods (i.e., DDN [6], DID-MDN [48], RESCAN [18], and SPANet [34]), two semi-supervised approaches (i.e., SSIR [36] and Syn2Real [43]), and three unsupervised deep nets (i.e., CycleGAN [52], RR-GAN [51], and DerainCycleGAN [37]). As the code of the methods [37, 51] is not available, we refer to some comparison results presented in their paper. For other methods, if there are no pre-trained models, we retrain them with the implementations provided by authors, otherwise we evaluate them directly with the online available codes.

Evaluation Metrics. For the images with ground truth, we evaluate each method by two well-known used metrics: PSNR and SSIM [35]. For the cases without ground truth, we use no-reference quality metrics NIQE [24] and BRISQUE [25] to perform quantitative comparison.
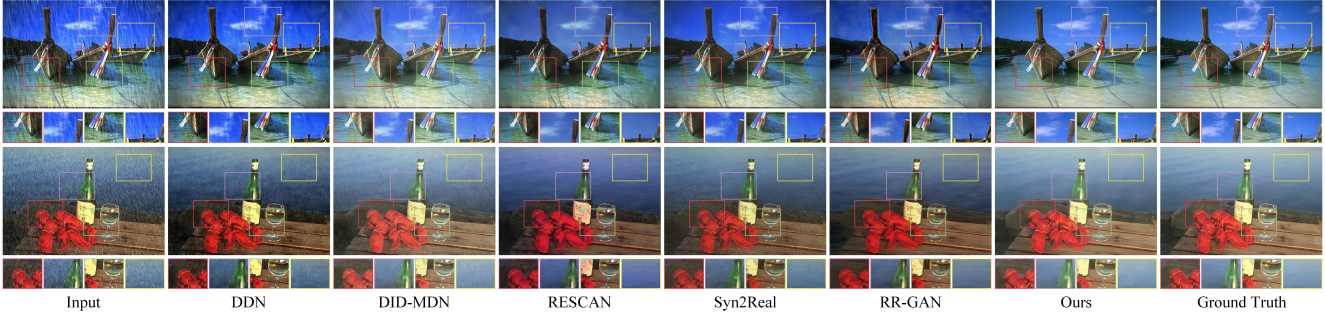
Figure 4. Qualitative evaluations on the Test100 dataset. Zooming in the figures offers a better view at the deraining capability.
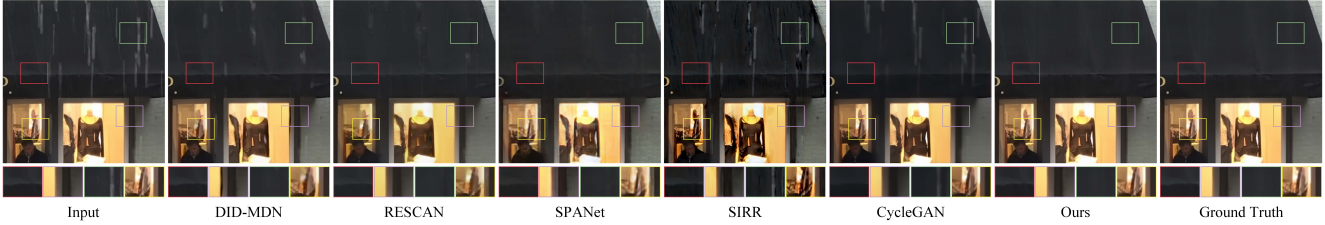

Figure 5. Qualitative evaluations on the Test1000 dataset. Zooming in the figures offers a better view at the deraining capability.

Figure 6. Qualitative evaluations on a real rainy image. Zooming in the figures offers a better view at the deraining capability.

4.2. Experimental Results

Synthetic Data. Tab. 1 reports the quantitative results of different methods on synthetic datasets including Test100, Test1200, and Test1400. We can observe that: (1) compared with unpaired deraining methods, our net achieves state-of-the-art results of PSNR, which reflects the excellent performance of DCD-GAN. (2) For semi-supervised methods, there is still a performance gap with fully-supervised models, even worse than our unpaired net. (3) Our approach can deliver competitive results to several existing paired supervised models, thanks to the additional constraints provided by CGBs. Besides, we show some hard examples for visual observation comparisons in Fig. 4. According to the ground truth, all the other works do an undesired work in details and color restoration. In contrast, our developed method not only be able to remove various types of rain streaks but also preserve color and structural properties of underlying objects to a substantial degree.

Real-world Data. For further general verification in practical use, we conduct comparisons against other algorithms on the real-world rainy dataset. Since SPA-Data [34] has its ground truth by human labeling, it can be evaluated using

Table 3. Comparison of quantitative results on real-world rainy images, and note that lower scores indicate better image quality.

Method	RESCAN	SPANet	SSIR	Syn2Real	CycleGAN	Ours
NIQE ↓	4.571	4.317	4.263	4.354	4.689	**4.032**
BRISQUE ↓	25.446	20.220	27.418	15.920	23.801	**13.311**

the numerical metrics. According to the numerical value in the last column of Tab. 1, our proposed model of DCD-GAN gets the best values on PSNR metric and obtains competitive SSIM values compared with other related nets on Test1000. From Fig. 5, it can be observed that the proposed method recovers clearer images with faithful structures.

Furthermore, we choose 30 real-world rainy images from Google image search. Fig. 6 illustrates the deraining performance of all competing approaches on one sample image. As rainy images from natural scenarios are more complex and challenging, our method still exhibits remarkable performance with less rain streaks residue. In contrast, other approaches fail to achieve the desired results. According the average values of NIQE and BRISQUE in Tab. 3, our method gets the lower scores, which means a high-quality output with higher naturalness and better fidelity against other comparative models on real-world images.
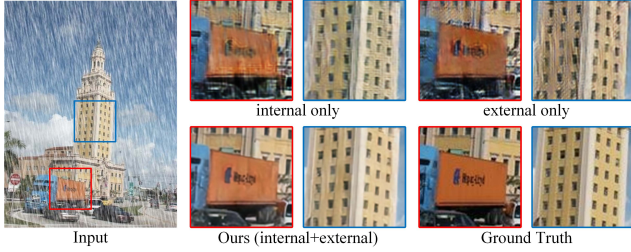
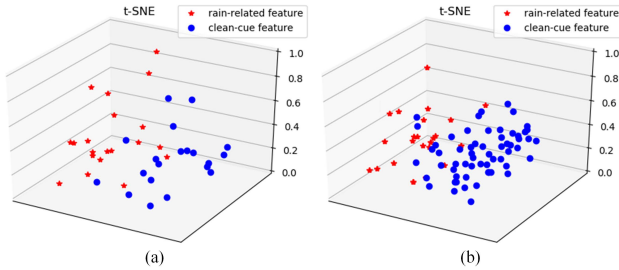Figure 7. Qualitative analysis of negatives selection strategies.

internal only | external only
Input | Ours (internal+external) | Ground Truth

Figure 8. Visualization results of t-SNE. (a) w/o backward cycle; (b) w forward+backward cycle.

Table 4. Ablation study for different loss functions. * indicates that $L_{colorcyc}$ is replaced by L_{cyc}.

Model	w/o L_{adv}	w/o $L_{colorcyc}$	w L_{cyc}*
PSNR / SSIM	24.03 / 0.754	25.08 / 0.782	25.32 / 0.795
Model	w/o L_{cont}	w/o L_{freq}	Ours
PSNR / SSIM	22.84 / 0.731	25.40 / 0.799	**25.61 / 0.813**

Table 5. Ablation study for different ways of selecting negatives.

Model	internal only	external only	Ours (internal+external)
PSNR / SSIM	25.47 / **0.815**	25.09 / 0.798	**25.61** / 0.813

Table 6. Ablation study for different variants of our model.

Model	w/o backward cycle	Ours (forward+backward cycle)
PSNR / SSIM	24.15 / 0.776	**25.61 / 0.813**
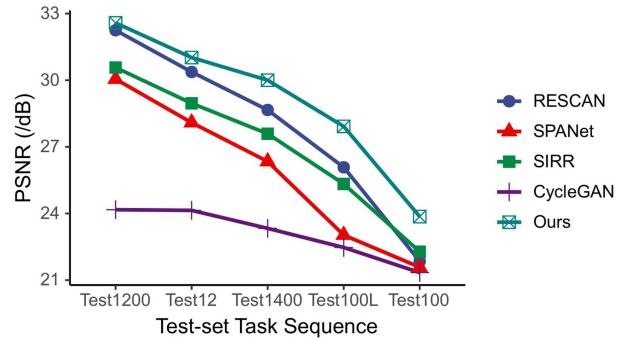
Figure 9. Quantitative analysis on cross-domain generalization.

4.3. Ablation Study

Effectiveness of Proposed Loss Function. In order to evaluate the effectiveness of our hybrid loss function, we conduct an ablation study on Test100. To be specific, we regularly remove one component to each configuration at one time. To ensure the fair comparison, the same training settings are kept for all models testing, except the modification depicted in Tab. 4. The best performance achieves 25.61 dB by using all the above component, so that each loss terms we consider has its own contribution in deraining process. We further investigate our color cycle-consistency loss effect by comparing with commonly used cycle-consistency loss (L_{cyc}). Through the last column of Tab. 4, we can observe that the performance drops with the replacement of color cycle-consistency loss, which confirms its advantages for improving the quality of recovery results.

Influences of Selecting Negatives. The most important design in contrastive learning is how to select the negatives. At each iteration, our developed DCD-GAN can generate four different stacks of features. Due to the advantages of dual learning setting, we concatenate the two stacks of features on both sides of the CGB together to provide more negative samples. Here, we explore the influence of this strategy by comparing using only 255 internal or 255 external negatives from one side. In Tab. 5, the quantitative results on Test100 show that our strategy achieves the highest value of PSNR, while the method using only internal negatives obtains the best SSIM. However, zooming the color boxed in Fig. 7, we observe that the output quality of the comparison

models are not as good as that of our strategy, that is, their derained results tend to make the image blur.

Influences of Clean Exemplars. To further verify the influences of the exemplars from the unparied clean images, we remove the backward cycle as our comparison model. The quantitative results on Test100 are recorded in Tab. 6. As one can see, the absence of the backward cycle substantially degrades results. In addition, we visualize the corresponding features using t-SNE [32] in Fig. 8. Interestingly, we further observe that more clean-cue features are produced during the optimization. This is due to the backward cycle enriches clean exemplars, resulting in these unpaired data to better facilitate rain removal.

4.4. Cross-Domain Generalizability Analysis

We also conduct experiments to demonstrate that DCD-GAN can promote its cross-domain generalization. Specifically, the model trained on the DID-Data [48] is used to sequentially test the task sequence Test1200 [48] → Test12 [19] → Test1400 [6] → Test100L [40] → Test100 [49], and the result is reported in Fig. 9. It is evident that the performance of the fully-supervised approaches decreases dramatically due to domain variation. Thanks to the learned latent restoration in the deep feature space, our method could be easily adapted without requiring any paired data in the new domain, which greatly boosts its cross-domain deraining generalization.
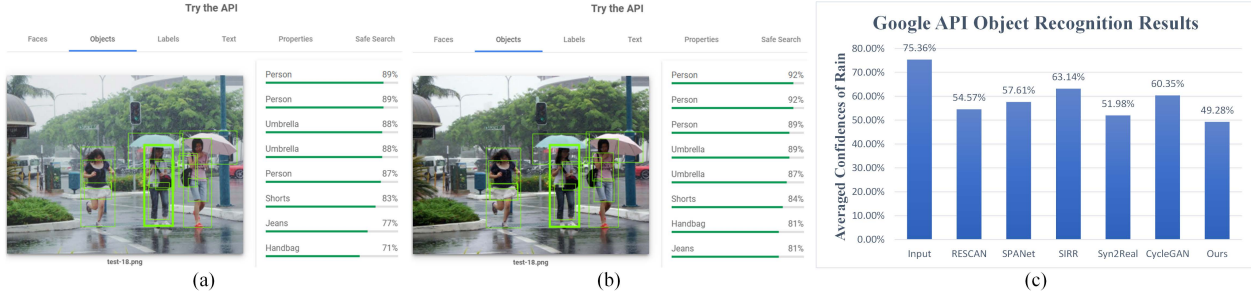
Figure 10. Comparison results tested on the Google Vision API. (a-b) Object recognition results for the input rainy image and our derained image; (c) The averaged confidences in recognizing rain. Note that lower scores indicate better performance.

Figure 11. Comparison results when trained on the mixed dataset.

4.5. Discussions with the Closely-Related Methods

To better solve the SID problem, we implement a mixed training manner. Here, we assemble a mixture of Rain800 [49] and 356 real rainy images (selected elaborately and cropped manually to highlight the rainy regions), which can provide more rain streak information for model learning. We note that [52] introduces a cycle-consistent constraint for unpaired image-to-image translation. In this work, to better achieve SID, we further develop multiple CGBs to provide complementary constraints. Fig. 11 (b) shows that CycleGAN fails to remove rain streaks, and it tends to blur the contents and cause color distortion. In contrast, we construct an effective contrastive regularization to explore the correlation of the latent space between the rainy and rain-free images, which thus leads to better derained results.

In addition, we also note that [7] recently develops a dual contrastive learning method for unsupervised image-to-image translation based on the CycleGAN framework. Different from this method learns domain-invariant information in complex image space, we focus on domain-specific properties in deep feature space. As displayed in Fig. 11 (c) and (d), the recovery result of DCLGAN has some residual rain streaks, while our proposed DCD-GAN can deal with majority of rain streaks. Since the mapping in the low-dimensional feature space is in principle much easier to learn than in the high-dimensional image space [33], another benefit can be found is being good at avoiding unnecessary image structure distortion. Due to the introduction of real rainy images in the mixing training, it also supplies more negative samples to CGB to some extent. One

more thing, cycle-consistent constraint and contrastive constraint share some commonalities [11], which will be a double benefit for encouraging deraining performance.

4.6. Application

To examine whether our method benefits outdoor vision-based applications, e.g., object recognition, we apply the detection method, i.e., Google Vision API, to evaluate the derained results. Fig. 10 (a) and (b) show that the recognition accuracy is improvement by using our derained output, suggesting that DCD-GAN can effectively improve the subsequent detection performance. Similar to [3, 45], we test 30 sets of real-world rainy images and derained images of six different methods, as shown in Fig. 10 (c). The confidence of rain indicates that the probability of rainy weather. As one can see, the averaged confidences in recognizing rain from our output results are significantly reduced, which indicates that DCD-GAN can restore high-quality images against other approaches. Therefore, it is meaningful to our developed unpaired deraining method to promote the practicality and scalability of real-world applications.

5. Concluding Remarks

In this paper, we have presented an effective unpaired learning, i.e., DCD-GAN, for SID. We incorporate the bidirectional translation branch and the contrastive guidance branch into a unified framework to achieve the joint training. Contrastive learning is first introduced into SID task to explore mutual features while distinguish the dissimilar ones in the deep feature space for better image restoration. Our approach enables the features from the unpaired clean exemplars to facilitate rain removal. Experimental results on synthetic and real rainy datasets considerably show that the effectiveness and generalization of our model. In future work, we will explore the potential of our proposed learning scheme in other unsupervised low-level vision tasks.

Limitation. Our method has limitation in SID performance when training on the small-scale datasets, due to contrastive learning tends to require a large number of sample pairs to achieve excellent performance.

References

[1] Chenghao Chen and Hao Li. Robust representation learning with feedback for single image deraining. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 7742–7751, 2021. 2

[2] Ka Leong Cheng, Yueqi Xie, and Qifeng Chen. Iicnet: A generic framework for reversible image conversion. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1991–2000, 2021. 5

[3] 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 Conference on Computer Vision and Pattern Recognition*, pages 14560–14569, 2020. 8

[4] Nanqing Dong, Matteo Maggioni, Yongxin Yang, Eduardo Pérez-Pellitero, Ales Leonardis, and Steven McDonagh. Residual contrastive learning for joint demosaicking and denoising. *arXiv preprint arXiv:2106.10070*, 2021. 3

[5] 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. 2

[6] 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. 2, 5, 7

[7] Junlin Han, Mehrdad Shoeiby, Tim Malthus, Elizabeth Botha, Janet Anstee, Saeed Anwar, Ran Wei, Lars Petersson, and Mohammad Ali Armin. Single underwater image restoration by contrastive learning. *arXiv preprint arXiv:2103.09697*, 2021. 3, 8

[8] Junlin Han, Mehrdad Shoeiby, Lars Petersson, and Mohammad Ali Armin. Dual contrastive learning for unsupervised image-to-image translation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 746–755, 2021. 2, 3

[9] Kewen Han and Xinguang Xiang. Decomposed cyclegan for single image deraining with unpaired data. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 1828–1832, 2020. 2

[10] 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. 5

[11] Chih-Hui Ho and Nuno Vasconcelos. Contrastive learning with adversarial examples. *arXiv preprint arXiv:2010.12050*, 2020. 8

[12] Huaibo Huang, Aijing Yu, Zhenhua Chai, Ran He, and Tieniu Tan. Selective wavelet attention learning for single image deraining. *International Journal of Computer Vision*, 129(4):1282–1300, 2021. 5

[13] Huaibo Huang, Aijing Yu, and Ran He. Memory oriented transfer learning for semi-supervised image deraining. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 7732–7741, 2021. 1, 2

[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. 5

[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 Conference on Computer Vision and Pattern Recognition*, pages 8346–8355, 2020. 1

[16] Xin Jin, Zhibo Chen, Jianxin Lin, Zhikai Chen, and Wei Zhou. Unsupervised single image deraining with self-supervised constraints. In *IEEE International Conference on Image Processing*, pages 2761–2765, 2019. 2

[17] Li-Wei Kang, Chia-Wen Lin, and Yu-Hsiang Fu. Automatic single-image-based rain streaks removal via image decomposition. *IEEE Transactions on Image Processing*, 21(4):1742–1755, 2011. 1

[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*, pages 254–269, 2018. 2, 5

[19] Yu Li, Robby T Tan, Xiaojie Guo, Jiangbo Lu, and Michael S Brown. Rain streak removal using layer priors. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2736–2744, 2016. 1, 5, 7

[20] Li Liu, Wanli Ouyang, Xiaogang Wang, Paul Fieguth, Jie Chen, Xinwang Liu, and Matti Pietikäinen. Deep learning for generic object detection: A survey. *International Journal of Computer Vision*, 128(2):261–318, 2020. 1

[21] Rui Liu, Yixiao Ge, Ching Lam Choi, Xiaogang Wang, and Hongsheng Li. Divco: Diverse conditional image synthesis via contrastive generative adversarial network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 16377–16386, 2021. 4

[22] Yang Liu, Ziyu Yue, Jinshan Pan, and Zhixun Su. Unpaired learning for deep image deraining with rain direction regularizer. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 4753–4761, 2021. 1, 2

[23] Yu Luo, Yong Xu, and Hui Ji. Removing rain from a single image via discriminative sparse coding. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 3397–3405, 2015. 1, 5

[24] Anish Mittal, Anush Krishna Moorthy, and Alan Conrad Bovik. No-reference image quality assessment in the spatial domain. *IEEE Transactions on Image Processing*, 21(12):4695–4708, 2012. 5

[25] Anish Mittal, Rajiv Soundararajan, and Alan C Bovik. Making a "completely blind" image quality analyzer. *IEEE Signal Processing Letters*, 20(3):209–212, 2012. 5

[26] Taesung Park, Alexei A Efros, Richard Zhang, and Jun-Yan Zhu. Contrastive learning for unpaired image-to-image translation. In *Proceedings of the European Conference on Computer Vision*, pages 319–345, 2020. 3, 4

[27] Dongwei Ren, Wangmeng Zuo, Qinghua Hu, Pengfei Zhu, and Deyu Meng. Progressive image deraining networks: A

better and simpler baseline. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3937–3946, 2019. 1

[28] Waqas Sultani, Chen Chen, and Mubarak Shah. Real-world anomaly detection in surveillance videos. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6479–6488, 2018. 1

[29] Hao Tang, Wei Wang, Dan Xu, Yan Yan, and Nicu Sebe. Gesturegan for hand gesture-to-gesture translation in the wild. In *Proceedings of the 26th ACM international conference on Multimedia*, pages 774–782, 2018. 4

[30] Hao Tang, Dan Xu, Hong Liu, and Nicu Sebe. Asymmetric generative adversarial networks for image-to-image translation. *arXiv preprint arXiv:1912.06931*, 2019. 4

[31] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Instance normalization: The missing ingredient for fast stylization. *arXiv preprint arXiv:1607.08022*, 2016. 5

[32] Laurens Van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of Machine Learning Research*, 9(11), 2008. 4, 7

[33] Ziyu Wan, Bo Zhang, Dongdong Chen, Pan Zhang, Dong Chen, Jing Liao, and Fang Wen. Bringing old photos back to life. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2747–2757, 2020. 8

[34] Tianyu Wang, Xin Yang, Ke Xu, Shaozhe Chen, Qiang Zhang, and Rynson WH Lau. Spatial attentive single-image deraining with a high quality real rain dataset. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 12270–12279, 2019. 2, 5, 6

[35] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004. 5

[36] Wei Wei, Deyu Meng, Qian Zhao, Zongben Xu, and Ying Wu. Semi-supervised transfer learning for image rain removal. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3877–3886, 2019. 1, 2, 5

[37] Yanyan Wei, Zhao Zhang, Yang Wang, Mingliang Xu, Yi Yang, Shuicheng Yan, and Meng Wang. Deraincyclegan: Rain attentive cyclegan for single image deraining and rainmaking. *IEEE Transactions on Image Processing*, 30:4788–4801, 2021. 2, 3, 5

[38] Zbigniew Wojna, Vittorio Ferrari, Sergio Guadarrama, Nathan Silberman, Liang-Chieh Chen, Alireza Fathi, and Jasper Uijlings. The devil is in the decoder: Classification, regression and gans. *International Journal of Computer Vision*, 127(11):1694–1706, 2019. 1

[39] Haiyan Wu, Yanyun Qu, Shaohui Lin, Jian Zhou, Ruizhi Qiao, Zhizhong Zhang, Yuan Xie, and Lizhuang Ma. Contrastive learning for compact single image dehazing. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 10551–10560, 2021. 3, 4

[40] 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. 7

[41] Wenhan Yang, Robby T Tan, Shiqi Wang, Yuming Fang, and Jiaying Liu. Single image deraining: From model-based to data-driven and beyond. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020. 2

[42] Wenhan Yang, Shiqi Wang, Dejia Xu, Xiaodong Wang, and Jiaying Liu. Towards scale-free rain streak removal via self-supervised fractal band learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 12629–12636, 2020. 1

[43] Rajeev Yasarla, Vishwanath A Sindagi, and Vishal M Patel. Syn2real transfer learning for image deraining using gaussian processes. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2726–2736, 2020. 1, 2, 5

[44] 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 Conference on Computer Vision and Pattern Recognition*, pages 2053–2062, 2021. 3

[45] Qiaosi Yi, Juncheng Li, Qinyan Dai, Faming Fang, Guixu Zhang, and Tieyong Zeng. Structure-preserving deraining with residue channel prior guidance. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 4238–4247, 2021. 1, 8

[46] Zili Yi, Hao Zhang, Ping Tan, and Minglun Gong. Dualgan: Unsupervised dual learning for image-to-image translation. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2849–2857, 2017. 4

[47] He Zhang and Vishal M Patel. Convolutional sparse and low-rank coding-based rain streak removal. In *IEEE Winter Conference on Applications of Computer Vision*, pages 1259–1267, 2017. 1

[48] 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. 2, 5, 7

[49] 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. 1, 5, 7, 8

[50] Jiahui Zhang, Shijian Lu, Fangneng Zhan, and Yingchen Yu. Blind image super-resolution via contrastive representation learning. *arXiv preprint arXiv:2107.00708*, 2021. 3

[51] Hongyuan Zhu, Xi Peng, Joey Tianyi Zhou, Songfan Yang, Vijay Chanderasekh, Liyuan Li, and Joo-Hwee Lim. Singe image rain removal with unpaired information: A differentiable programming perspective. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 9332–9339, 2019. 2, 5

[52] 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*, pages 2223–2232, 2017. 2, 3, 5, 8