Unpaired Adversarial Learning for Single Image Deraining with Rain-Space Contrastive Constraints

Xiang Chen, Jinshan Pan, Kui Jiang, Yufeng Huang, Caihua Kong, Longgang Dai, Yufeng Li

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

Abstract

Deep learning-based single image deraining (SID) with unpaired information is of immense importance, as relying on paired synthetic data often limits their generality and scalability in real-world applications. However, we noticed that direct employ of unpaired adversarial learning and cycle-consistency constraints in the SID task is insufficient to learn the underlying relationship from rainy input to clean outputs, since the domain knowledge between rainy and rain-free images is asymmetrical. To address such limitation, we develop an effective unpaired SID method which explores mutual properties of the unpaired exemplars by a contrastive learning manner in a GAN framework, named as CDR-GAN. The proposed method mainly consists of two cooperative branches: Bidirectional Translation Branch (BTB) and Contrastive Guidance Branch (CGB). Specifically, BTB takes full advantage of the circulatory architecture of adversarial consistency to exploit latent feature distributions and guide transfer ability between two domains by equipping it with bidirectional mapping. Simultaneously, CGB implicitly constrains the embeddings of different exemplars in rain space by encouraging the similar feature distributions closer while pushing the dissimilar further away, in order to better help rain removal and image restoration. During training, we explore several loss functions to further constrain the proposed CDR-GAN. Extensive experiments show that our method performs favorably against existing unpaired deraining approaches on both synthetic and real-world datasets, even outperforms 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 (Liu et al. 2020), segmentation (Wojna et al. 2019) and video surveillance (Sultani, Chen, and Shah 2018), 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:

\[ O = B + R, \]

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 (Kang, Lin, and Fu 2011; Luo, Xu, and Ji 2015; Li et al. 2016; Zhang and Patel 2017) 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 (Ren et al. 2019; Li, Cheong, and Tan 2019; Zhang, Sindagi, and Patel 2019; Jiang et al. 2020; Yang et al. 2020b; Fu et al. 2021). 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 the above problems, the semi-supervised (Wei et al. 2019a; Yasarla, Sindagi, and Patel 2020; Huang, Yu, and He 2021) and unsupervised (Zhu et al. 2019; Han and Xiang 2020) learning have been proposed for SID. These methods (Jin et al. 2019; Huang, Yu, and He 2021) either focus on domain invariant features by leveraging the limited labeled data and introducing the auxiliary optimization objectives or develop domain adaption strategies (Wei 2019b; Song et al. 2020). However, direct employ of unpaired adversarial learning and cycle-consistency constraints in the SID task is insufficient to learn the underlying relationship from rainy input to clean outputs, since the domain knowledge between rainy and rain-free images is asymmetrical. To address such limitation, we develop an effective unpaired SID method which explores mutual properties of the unpaired exemplars by a contrastive learning manner in a GAN framework, named as CDR-GAN. The proposed method mainly consists of two cooperative branches: Bidirectional Translation Branch (BTB) and Contrastive Guidance Branch (CGB). Specifically, BTB takes full advantage of the circulatory architecture of adversarial consistency to exploit latent feature distributions and guide transfer ability between two domains by equipping it with bidirectional mapping. Simultaneously, CGB implicitly constrains the embeddings of different exemplars in rain space by encouraging the similar feature distributions closer while pushing the dissimilar further away, in order to better help rain removal and image restoration. During training, we explore several loss functions to further constrain the proposed CDR-GAN. Extensive experiments show that our method performs favorably against existing unpaired deraining approaches on both synthetic and real-world datasets, even outperforms several fully-supervised or semi-supervised models.
et al. 2019b; Han and Xiang 2020) to improve the generalization capabilities of the deep models. However, without suitable constraints for rain streaks and clean images, existing unpaired learning methods (Zhu et al. 2017; Han et al. 2021b) do not effectively restore high-quality results, as shown in Fig. 1 (b) and (c). 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 Contrastive DeRain-GAN (CDR-GAN) method which employs the contrastive learning (CL) to explore useful features from the rainy images and unpaired clean images so that the extracted features can better facilitate rain removal. Our primary finding is that the features extracted from rainy image patches share some mutual information to the features extracted from rain-free image patches. This motivates us to develop a CL algorithm to explore the similar features while distinguish the dissimilar ones. The proposed CDR-GAN includes two main interactive branches: Bidirectional Translation Branch (BTB) and Contrastive Guidance Branch (CGB). Specifically, BTB is derived from the CycleGAN framework (Zhu et al. 2017) that is used to mine rain-related or clean-cue features by equipping it with bidirectional mapping. 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. This process ensures that the proposed method models the relationship between rainy input and clean outputs in the deep feature space. Moreover, to obtain better restored results, we combine multiple loss functions as the task-specific proxy for the better jointly optimization.

To summarize, the main contributions of this paper are summarized as follows:

- We propose an effective CDR-GAN which adopts CL to explore mutual features while distinguish the dissimilar ones between the rainy domain and the rain-free domain in the deep feature space.
- The developed CDR-GAN is performed without paired training information, where the features from the unpaired clean exemplars can facilitate rain removal and help image restoration.
- Experimental results on several challenging datasets considerably demonstrate that our method performs favorably against existing unpaired deraining approaches, and is also competitive to other related fully-supervised or semi-supervised models.

2 Related Work

2.1 Single Image Deraining

Existing learning-based SID methods are categorized into paired (fully-supervised), semi-supervised and unpaired (without paired supervised) approaches (Yang et al. 2020a).

For the paired deraining methods, (Fu et al. 2017a) first employs the Derain Net with multi-layer CNN to extract and remove the rain layer, and further introduces deep detail network (DDN) (Fu et al. 2017b) that directly removes the rain streaks by reducing the mapping range. (Zhang and Patel 2018) presents a density-aware multi-stream densely connected CNN algorithm, termed DID-MDN, to better characterize rain streaks by estimating rain density. The work of RESCAN (Li et al. 2018) introduces a convolutional and recurrent neural network-based way to make full use of the contextual information. Spatial attentive network (SPANet) (Wang et al. 2019) 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 (Chen and Li 2021) to address deraining model errors.

For the semi-supervised learning-based methods, (Wei et al. 2019a) utilizes a semi-supervised learning framework (SSIR) to analyze the residual difference between the derained images and the input images. The work of Syn2Real (Yasarla, Sindagi, and Patel 2020) proposes a Gaussian process-based semi-supervised learning method, which can increase the network ability by using synthetic images and unlabeled real rainy images.

For the unpaired deraining field, CycleGAN (Zhu et al. 2017) has been applied to address the unpaired images translation, to achieve deraining process. The RR-GAN (Zhu et al. 2019), DerainCycleGAN (Wei et al. 2019b) and DCycleGAN (Han and Xiang 2020) all use the improved CycleGAN structure 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 the limitation and incorporates CL with adversarial training to further improve deraining robustness in unpaired setting.

2.2 Contrastive Learning

CL has witnessed a significant progress in the field of unsupervised representation learning. Instead of using a fixed target and pre-defined, CL 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 low-level vision tasks and achieved SOTA performance, e.g., image dehazing (Wu et al. 2021), image denoising (Dong et al. 2021), image super-resolution (Zhang et al. 2021) and underwater image restoration (Han et al. 2021a). Similarly, the recent work of DCLGAN (Han et al. 2021b) takes the advantages of both CUT (Park et al. 2020) and CycleGAN (Zhu et al. 2017) to achieve unsupervised image-to-image translation. Different from its mutual mapping in complex image space, we show the use of contrastive constraints to perform unpaired learning in simple rain space, which can better help rain removal.
3 Methodology

Fig. 2 shows the overall architecture of our proposed CDR-GAN. It consists of BTB and CGB branches. BTB is used to guide transfer ability between the rainy domain and the rain-free domain. CGB is used to constrain the mutual information between two domains, as well as the multiple loss functions to provide jointly optimization. We present the details about the proposed network in the following.

3.1 Bidirectional Translation Branch (BTB)

Let $D_R$ denotes the set of the rainy images without ground truth labels and $D_N$ denotes the set of clean exemplar images, our goal is to learn a deep model to explore the features from these unpaired data $D_R$ and $D_N$ without the supervision of the ground truth labels to estimate the rain-free images. To this end, we develop a bidirectional translation branch (BTB) as the backbone of the proposed CDR-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 de-rained images and real clean images. Continuing along the trajectory of (Wei et al. 2019b; Ye et al. 2021), the 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 exemplar generated by BTB could be mined rain-related or clean-cue features, 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 directly utilize bidirectional cycle-consistency to remove rain, in that this constraint is weak. Hence, we introduce CL 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.

3.2 Contrastive Guidance Branch (CGB)

To guide in better deraining, CGB is involved between $D_R$ and $D_N$ to exploit advantage of the learned domain-relevant feature distribution. In general, these domain-relevant features are embedded into two representation spaces: a domain-specific attribute (i.e., rain-component) space and a domain-invariant content (i.e., clean-background) space. Our primary insight is that the features extracted from $D_R$ share some mutual properties to the features extracted from $D_N$. Since rain space is relatively simpler, it is easier to capture feature disparity. If pairs of patches generated from different domains with low disparity in rain space are similar (referred to as “positive”), their embeddings in latent space should also have a low distance, and vice-versa.

Technically, we use the auxiliary encoder as separate embeddings to extract features from domain $D_R$ and $D_N$, respectively. As suggested in (Liu et al. 2021), by using off-the-shelf generators to extract feature representations in CL, excessive modifications to the original architecture can be avoided. Therefore, 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. Inspired by (Han et al. 2021b), we do not share weights, thus capturing variability in both domains and learning better embeddings. Intuitively, 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 supervising the deraining process. That is, CGB takes the generated images from BTB
as inputs to improve the capacity of the auxiliary encoder, which in turn also promotes the rain generation quality.

To constrain features with the discovered pairs of patches in their embeddings, let us define the query exemplar $\hat{x}$ sampled from rain space, the positive patch $\hat{x}^+$ and $k$ negative patches $\{\hat{x}^-\}_{i=1}^k$ sampled from latent space. The constraint of CGB is enforced by encouraging the positives closer while keeping the negatives further away, so as to indirectly help the deraining generators guide correctly in transforming the rainy input to the clean output. Through the contrasting process, these corresponding feature representations are then formalized as: $f = E(\hat{x})$, $f^+ = E(\hat{x}^+)$, $f_i^- = E(\hat{x}^-_i)$. Such latent feature distributions are expected to be discriminable and can be captured the relationship between $\mathbb{D}_R$ and $\mathbb{D}_N$. Therefore, we introduce a contrastive regularization term to regularize the captured images’ feature distributions. Finally, this regularization term is employed in the form of Contrastive Loss, which can be written as:

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

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 $\tau$ denotes a temperature parameter.

### 3.3 Loss Function

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

#### Color Cycle-Consistency Loss

Since no paired data is provided for supervision, previous CycleGAN-based methods mostly adopt the cycle-consistency loss to enforce forward-backward consistency. By doing so, it is capable of limiting generated samples’ space and preserving image content. However, there exists the problem of “channel pollution” (Tang et al. 2018) in cycle-consistency loss, because generating a whole image at one time makes different channels interact with each other, resulting in unpleasant artifacts in the derained results. To reduce this influence, 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:

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

where $i$ denotes different channels. $r \in \mathbb{D}_R$, $n \in \mathbb{D}_N$ are a rainy image and an unpaired clean image. We will demonstrate the effect of this loss function in Section 4.3.

### Adversarial Loss

We note that using the adversarial loss is able to generate the realistic images. Similar to (Zhu et al. 2019), the adversarial loss in $\mathbb{D}_N$ for SID is defined as:

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

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 $L_{\text{adv}} = L_{\text{adv}}(G_{R2N}, D_N) + L_{\text{adv}}(G_{N2R}, D_R)$.

#### Identity Preserving Loss

To further preserve the identity information such as color invariance between the input and the output, we add an identity preserving loss. Considering training speed and time complexity, we do not use Patch-NCE loss (Park et al. 2020) as identity preserving loss. Instead, we define it as follows:

$$L_{\text{idt}} = \mathbb{E}_{r \sim R} \left[\|G_{N2R}(n) - n\|_1\right] + \mathbb{E}_{n \sim N} \left[\|G_{R2N}(r) - r\|_1\right].$$

#### Perceptual Loss

Intuitively, we exploit the perceptual loss to restrict the reconstruction of image details. The perceptual loss can be defined as:

$$L_p = \|\varphi_l(r) - \varphi_l(G_{N2R}(G_{R2N}(r)))\|_2^2 + \|\varphi_l(n) - \varphi_l(G_{R2N}(G_{N2R}(n)))\|_2^2,$$

where $\varphi_l(\cdot)$ denotes the features extracted from $l$-th layer of VGG-16 network (Simonyan and Zisserman 2014) pretrained on ImageNet (Deng et al. 2009).

#### Total Variation Loss

To remove the artifacts in the restored images, we apply the total variation to $N_R$:

$$L_{\text{tv}} = \|\partial_h N_R\|_1 + \|\partial_v N_R\|_1,$$

where $\partial_h$ and $\partial_v$ represent the horizontal and vertical gradient operators, respectively.

Based on above considerations, the final loss function is:

$$L_{\text{total}} = \lambda_{\text{adv}} L_{\text{adv}} + \lambda_{\text{cont}} L_{\text{cont}} + \lambda_{\text{cyc}} L_{\text{color cyc}} + \lambda_{\text{idt}} L_{\text{idt}} + \lambda_p L_p + \lambda_{\text{tv}} L_{\text{tv}},$$

where $\lambda_{\text{adv}}$, $\lambda_{\text{cont}}$, $\lambda_{\text{cyc}}$, $\lambda_{\text{idt}}$, $\lambda_p$, and $\lambda_{\text{tv}}$ are trade-off weight. In our experiments, we empirically set $\lambda_{\text{adv}} = 1$, $\lambda_{\text{cont}} = 2$, $\lambda_{\text{cyc}} = 10$, $\lambda_{\text{idt}} = 1$, $\lambda_p = 5$, and $\lambda_{\text{tv}} = 0.01$.

### 4 Experiments

#### 4.1 Experimental Settings

**Datasets.** We use four challenging benchmark datasets including Rain800 (Zhang, Sindagi, and Patel 2019), DID-Data (Zhang and Patel 2018), DDN-Data (Fu et al. 2017b), and SPA-Data (Wang et al. 2019) with various rain streaks of different sizes, shapes and directions. The detailed descriptions of the used datasets are tabulated in Tab. 2.

**Implementation Details.** The overall architecture of CDR-GAN is based on CycleGAN, a Resnet-based (He et al. 2016) generator with 9 residual blocks and a PatchGAN
(Isola et al. 2017) 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. The temperature parameter $\tau$ is set to 0.07. We use a batch size of 1 and instance normalization (Ulyanov, Vedaldi, and Lempitsky 2016). The entire network is performed using Pytorch framework on one NVIDIA Tesla V100 GPU. It should be noted that all trainings are randomly cropped to $256 \times 256$ patches in an unpaired manner. To enhance the stability of adversarial learning, we keep an image buffer that stores the 50 previously created images. We will release our source code and trained models to the public.

**Evaluated Methods.** We compare our method with those of two prior-based algorithms (i.e., DSC (Luo, Xu, and Ji 2015) and GMM (Li et al. 2016)), four paired supervised methods (i.e., DDN (Fu et al. 2017b), DID-MDN (Zhang and Patel 2018), RESCAN (Li et al. 2018), and SPA-Net (Wang et al. 2019)), two semi-supervised approaches (i.e., SSIR (Wei et al. 2019a) and Syn2Real (Yasarla, Sindagi, and Patel 2018)), two semi-supervised methods (i.e., DDN (Fu et al. 2017b), DID-MDN (Zhang and Patel 2018), RESCAN (Li et al. 2018), and SPA-Net (Wang et al. 2019)), and two unpaired deep nets (i.e., CycleGAN (Zhu et al. 2017), and RR-GAN (Zhu et al. 2019)). Due to RR-GAN (Zhu et al. 2019) has not released code, 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 commonly used metrics: PSNR and SSIM (Wang et al. 2004). For the images without ground truth, we use no-reference quality metrics NIQE (Mittal, Moorthy, and Bovik 2012) and BRISQUE (Mittal, Soundararajan, and Bovik 2012).

### 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     |
| Prior-based methods | DSC     | 18.56    | 0.599    | 24.24    | 0.827    | 27.31    | 0.837    | 34.95    | 0.941    |
|           | GMM     | 20.46    | 0.730    | 25.66    | 0.817    | 26.87    | 0.808    | 34.30    | 0.942    |
| Paired / Supervised methods | DDN     | 21.16    | 0.732    | 27.93    | 0.853    | 28.00    | 0.873    | 34.70    | 0.934    |
|           | DID-MDN | 21.89    | 0.795    | 29.66    | 0.899    | 26.38    | 0.835    | 34.68    | 0.930    |
|           | RESCAN  | 24.09    | 0.841    | 32.25    | 0.907    | 32.03    | 0.917    | 34.71    | 0.937    |
|           | SPANet  | 24.37    | 0.861    | 30.05    | 0.934    | 29.83    | 0.904    | 35.13    | 0.944    |
| Semi-supervised methods | SIRR    | 30.57    | 0.910    | 30.01    | 0.907    | 34.85    | 0.935    |
|           | Syn2Real| /        | /        | /        | /        | /        | /        | /        | /        |
| Unpaired / Without paired supervised methods | CycleGAN | 22.95    | 0.783    | 28.68    | 0.875    | 27.74    | 0.852    | 31.59    | 0.886    |
|           | RR-GAN  | 23.51    | 0.799    | 23.74    | 0.810    | 32.58    | 0.937    | 31.87    | 0.921    |
|           | Ours    | 26.43    | 0.810    | 32.58    | 0.937    | 31.87    | 0.921    | 35.36    | 0.943    |

### Table 2: Descriptions of different benchmark 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 |

### 4.2 Experimental Results

**Synthetic Data.** Tab. 1 shows the quantitative evaluations of different methods on synthetic datasets including Test100, Test1200, and Test1400. We can observe that: (1) compared with unpaired deraining approaches, our net achieves state-of-the-art results of both PSNR and SSIM, which reflect the excellent performance of CDR-GAN. (2) For semi-supervised methods, there is still a performance gap with fully-supervised models, even worse than our unpaired net. (3) Our method can deliver competitive results to several existing paired supervised models, thanks to the additional constraints provided by CGB. Besides, we show some hard examples for visual observation comparisons in Fig. 3. According to the ground truth, all the other works do an undesired work in details and color restoration. In contrast, our 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 has the corresponding ground truth, it can be evaluated using the numerical metrics. According to the numerical value in the last column of Tab. 1, our proposed model of CDR-GAN gets the best values on PSNR metric and obtains competitive SSIM values compared with other related nets on Test1000. From Fig. 4, it can be observed that the proposed method recovers more clean images with faithful structures.

Furthermore, we choose another example from real rainy scenarios. Fig. 6 illustrates the deraining performance of all competing approaches on real images. As rainy images from natural scenes are more complex and challenging, our method still exhibits remarkable performance with less rain streaks left in the derained image. In comparison, other approaches fail to achieve the desired results. According the value of NIQE/BRISQUE under the images, the proposed CDR-GAN gets the lower scores, which means a high-quality output against the other comparative methods on real-world images.
4.3 Ablation Study

**Effectiveness of Proposed Loss Function.** In order to evaluate the effectiveness of our multiple loss functions, we conduct an ablation study on Test100. Correspondingly, 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. 3. And the PSNR is employed to test the performance of each deraining models. The best performance achieves 26.43 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. 3, 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 Clean Exemplars.** To further verify the influences of the exemplars from the unpaired clean images, we remove the backward cycle as our comparison model. In Tab. 4, the absence of the backward cycle substantially degrades results. This is due to the backward cycle enriches clean exemplars, resulting in these unpaired data to better facilitate rain removal.

**Discussions with the Closely-Related Methods**

We note that (Zhu et al. 2017) introduces a cycle-consistent constraint for unpaired image-to-image translation. In this work, we also employ the cycle-consistent constraint for the SID problem. However, to better achieve SID, we further develop CGBs to provide additional constraints. To compare with (Zhu et al. 2017), we train this model using the same training dataset as the proposed method. Fig. 5 (b) shows that CycleGAN fails to remove the rain streaks, and it tends to blur the contents and cause color distortion. In contrast, we construct an effective CL to explore the correlation of the latent deep feature space between the rainy and rain-free images, which thus leads to better derained results.

In addition, we also note that Han et al. (Han et al. 2021a) recently develops an effective dual CL method for unsu-
supervised image-to-image translation based on the Cycle-GAN framework. Different from this method learns domain-invariant information in complex image space, we focus on domain-specific rain-component feature in simple rain space. To better solve the SID problem, we implement a mixed training manner. Here, we assemble a mixture of Rain800 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. Due to the positive effect of this mixed dataset, it is conducive to bringing it closer to practical applications. As displayed in Fig. 5 (c) and (d), the recovery result of DCL-GAN has some residual rain streaks, while our proposed CDR-GAN can deal with majority of rain streaks. Since our CGB performs contrastive constraints in rain space, 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 contrast constraint share some commonalities, which will be a double benefit for encouraging deraining performance. In short, our method could be easily adapted without requiring any paired data in the new domain, which greatly facilitates its real-world generalization.

4.5 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. 7 (a) and (b) show that the recognition accuracy is improvement by using our derained output, suggesting that CDR-GAN can effectively improve the subsequent detection performance. Similar to (Deng et al. 2020), we test 30 sets of real-world rainy images and derained images of six different methods, as shown in Fig. 7 (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 CDR-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., CDR-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 rain 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 demonstrate that the effectiveness and generalization of our model. In future work, we plan to explore the potential of our proposed learning scheme in other unpaired low-level vision tasks.
2019b. DerainCycleGAN: A simple unsupervised network for single image deraining and rainmaking. arXiv e-prints, arXiv:1912.
Zhu, H.; Peng, X.; Zhou, J. T.; Yang, S.; Chanderasekh, V.; Li, L.; and Lim, J.-H. 2019. Single image rain removal with unpaired information: A differentiable programming perspective. In Proceedings of the AAAI Conference on Artificial Intelligence, 9332–9339.
Wang, T.; Yang, X.; Xu, K.; Chen, S.; Zhang, Q.; and Lau, R. W. 2019. Spatial attentive single-image deraining with a high quality real rain dataset. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 12270–12279.
Yasarla, R.; Sindagi, V. A.; and Patel, V. M. 2020. Syn2Real transfer learning for image deraining using Gaussian processes. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2726–2736.
Liu, L.; Ouyang, W.; Wang, X.; Fieguth, P.; Chen, J.; Liu, X.; and Pietikäinen, M. 2020. Deep learning for generic object detection: A survey. International Journal of Computer Vision, 128(2): 261–318.
Han, K.; and Xiang, X. 2020. Decomposed cycleGAN for single image deraining with unpaired data. In IEEE International Conference on Acoustics, Speech and Signal Processing, 1828–1832.
Jiang, K.; Wang, Z.; Yi, P.; Chen, C.; Huang, B.; Luo, Y.; Ma, J.; and Jiang, J. 2020. Multi-scale progressive fusion network for single image deraining. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 8346–8355.
Shao, Y.; Li, L.; Ren, W.; Gao, C.; and Sang, N. 2020. Domain adaptation for image dehazing. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2808–2817.
Zhou, Y.; and Yan, K. 2020. Domain adaptive adversarial learning based on physics model feedback for underwater image enhancement. arXiv preprint arXiv:2002.09315.
He, K.; Fan, H.; Wu, Y.; Xie, S.; and Girshick, R. 2020. Momentum contrast for unsupervised visual representation learning. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 9729–9738.
Deng, S.; Wei, M.; Wang, J.; Feng, Y.; Liang, L.; Xie, H.; Wang, F. L.; and Wang, M. 2020. Detail-recovery image deraining via context aggregation networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 14560–14569.
Chen, T.; Kornblith, S.; Norouzi, M.; and Hinton, G. 2020. A simple framework for contrastive learning of visual representations. In International Conference on Machine Learning, 1597–1607.
Park, T.; Efros, A. A.; Zhang, R.; and Zhu, J.-Y. 2020. Contrastive learning for unpaired image-to-image translation. In Proceedings of the European Conference on Computer Vision, 319–345.
Yang, W.; Tan, R. T.; Wang, S.; Fang, Y.; and Liu, J. 2020a. Single image deraining: From model-based to data-driven and beyond. IEEE Transactions on Pattern Analysis and Machine Intelligence.
Yang, W.; Wang, S.; Xu, D.; Wang, X.; and Liu, J. 2020b. Towards scale-free rain streak removal via self-supervised fractal band learning. In Proceedings of the AAAI Conference on Artificial Intelligence, 12629–12636.
Fu, X.; Qi, Q.; Zha, Z.-J.; Zhu, Y.; and Ding, X. 2021. Rain streak removal via dual graph convolutional network. In Proceedings of the AAAI Conference on Artificial Intelligence, 1–9.
Huang, H.; Yu, A.; and He, R. 2021. Memory oriented transfer learning for semi-supervised image deraining. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 7732–7741.
Shyam, P.; Yoon, K.-J.; and Kim, K.-S. 2021. Towards domain invariant single image dehazing. arXiv preprint arXiv:2101.10449.
Li, M.; Li, C.-G.; and Guo, J. 2021. Cluster-guided asymmetric contrastive learning for unsupervised person re-identification. arXiv preprint arXiv:2106.07846.
Chen, C.; and Li, H. 2021. Robust representation learning with feedback for single image deraining. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 7742–7751.
Han, J.; Shoeiby, M.; Petersson, L.; and Armin, M. A. 2021b. Dual contrastive learning for unsupervised image-to-image translation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 746–755.
Wu, H.; Qu, Y.; Lin, S.; Zhou, J.; Qiao, R.; Zhang, Z.; Xie, Y.; and Ma, L. 2021. Contrastive learning for compact single image dehazing. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 10551–10560.
Dong, N.; Maggioni, M.; Yang, Y.; Pérez-Pellitero, E.; Leonards, A.; and McDonagh, S. 2021. Residual contrastive learning for joint demosaicking and denoising. arXiv preprint arXiv:2106.10070.
Zhang, J.; Lu, S.; Zhan, F.; and Yu, Y. 2021. Blind image super-resolution via contrastive representation learning. arXiv preprint arXiv:2107.00708.
Han, J.; Shoeiby, M.; Malthus, T.; Botha, E.; Anstee, J.; Anwar, S.; Wei, R.; Petersson, L.; and Armin, M. A. 2021a. Single underwater image restoration by contrastive learning. arXiv preprint arXiv:2103.09697.
Ye, Y.; Chang, Y.; Zhou, H.; and Yan, L. 2021. Closing the Loop: Joint rain generation and removal via disentangled image translation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2053–2062.
Liu, R.; Ge, Y.; Choi, C. L.; Wang, X.; and Li, H. 2021. Divuco: Diverse conditional image synthesis via contrastive generative adversarial network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 16377–16386.