Unsupervised Deraining: Where Contrastive Learning Meets Self-similarity

Yuntong Ye[1], Changfeng Yu[1], Yi Chang[1*], Lin Zhu[2], Xi-le Zhao[3], Luxin Yan[1], Yonghong Tian[2]
[1]School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, China
[2]School of Artificial Intelligence, Peking University, China
[3]School of Mathematical Sciences, University of Electronic Science and Technology of China, China

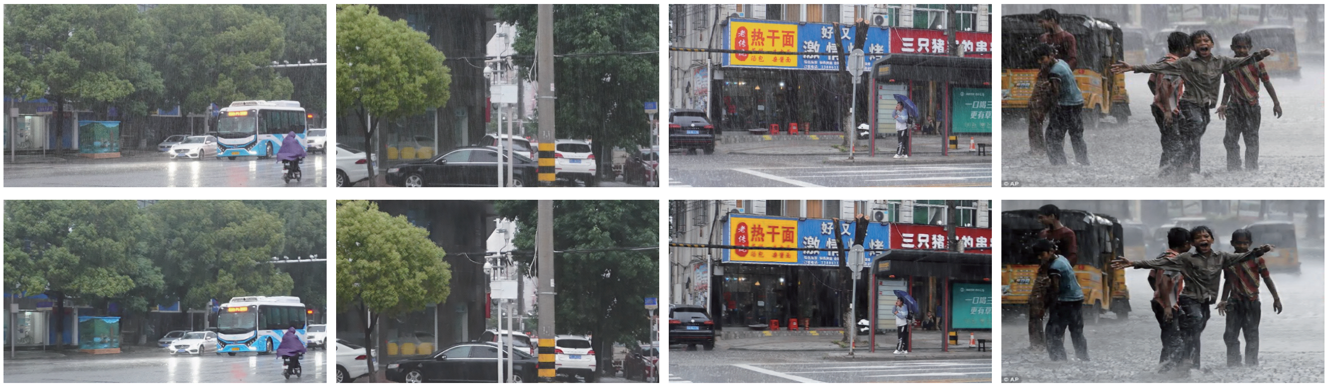


Figure 1. The proposed method can remove both the real-world rain streaks and veiling effect meanwhile well preserve image structures in an unsupervised manner. More real results and comparisons with the state-of-the-arts can be found in the supplementary.

Abstract

Image deraining is a typical low-level image restoration task, which aims at decomposing the rainy image into two distinguishable layers: clean image layer and rain layer. Most of the existing learning-based deraining methods are supervisedly trained on synthetic rainy-clean pairs. The domain gap between the synthetic and real rains makes them less generalized to different real rainy scenes. Moreover, the existing methods mainly utilize the property of the two layers independently, while few of them have considered the mutually exclusive relationship between the two layers. In this work, we propose a novel non-local contrastive learning (NLCL) method for unsupervised image deraining. Consequently, we not only utilize the intrinsic self-similarity property within samples, but also the mutually exclusive property between the two layers, so as to better differ the rain layer from the clean image. Specifically, the non-local self-similarity image layer patches as the positives are pulled together and similar rain layer patches as the negatives are pushed away. Thus the similar positive/negative samples that are close in the original space benefit us to enrich more discriminative representation. Apart from the self-similarity sampling strategy, we analyze how to choose an appropriate feature encoder in NLCL. Extensive experiments on different real rainy datasets demonstrate that the proposed method obtains state-of-the-art performance in real deraining.

1. Introduction

The existing high-level computer vision tasks such as image segmentation [6], and object detection [34] have achieved significant progress in recent years. Unfortunately, their performance would suffer from degradation under the rainy weather [1, 22, 29]. To alleviate the influence of the rain, numerous full-supervised deraining methods have been proposed [11,54,59]. Although they can achieve good results on simulated rainy image, they cannot well generalize to the real rain because of the domain gap between the simplified synthetic rain and complex real rain [56]. The goal of this work is to remove the real rain in an unsupervised manner.

To handle the real-world complex rainy images, the optimization-based methods are firstly proposed with handcrafted priors such as the sparse coding [36], low-rank [4] and Gaussian mixture model [31]. However, these handcrafted priors are of limited representation ability, especially for highly complex and varied rainy scenes. To rectify this weakness, the learning-based CNN methods [11, 28, 30, 54] have made great progresses. The key idea of these supervised learning methods tries the best to simulate the rain

[*] Corresponding author

as real as possible with sophisticated models, such as the additive model [25], screen blend model [36], heavy rain model [54], and comprehensive rain model [19], to name a few. Unfortunately, there still exist gap between these synthetic rain models and real rain degradation, since the real rainy atmosphere is usually a high-order nonlinear system.

Furthermore, the semi-supervised deraining methods have been proposed to effectively improve the robustness for real rains [20, 35, 48, 49, 55, 56], where they employ the simulated labels for good initialization and unlabeled real rains for generalization. Their performances still depend on the distribution gap between the simulated and real rainy images to some extent. Once the distributions are of large distance, these semi-supervised deraining results would be less satisfactory. Very recently, the unsupervised methods have raised more attentions for real rain removal, mainly including the CycleGAN-based unpaired image translation methods [23,50,60] and the optimization-model driven deep prior network [58]. However, the previous methods including the unsupervised ones mainly pay attention to the property of the image or rain layer independently, yet seldom consider the mutually exclusive relationship between the two layers.

To overcome these problems, we formulate the image deraining into a contrastive learning framework [7, 18] from an unsupervised perspective. The core idea of contrastive learning is that the representation of similar samples should be pulled close together, while that of dissimilar samples should be pushed far away in the embedding space [16, 52]. Figure 2 illustrates the main idea of proposed method. The image deraining is formulated as an image decomposition task, in which the clean image patches are regarded as the positives while the rain layer patches as the negatives. Thus, we not only take advantage of the properties of both image and rain layers, but also model the mutually exclusive relationship between the two layers for better decomposition. On the other hand, the proposed method does not require the clean supervision, which makes it generalize well for the real-world rainy images.

The key factor of contrastive learning is how to construct different views for both the positive and negative samples. The main stream is to augment a single instance with different transformations as the positive samples so as to learn the invariant representations [7]. However, these instance-level hand-crafted augmentations are not adequate to cover various situations. In this work, we provide a new perspective via the patch-level self-similarity within a single image. While non-local self-similarity [3] has been extensively studied in the literature, this intrinsic property for capturing the cross-patch relation in a single image with contrastive learning has barely been explored for visual representation learning.

To the best of our knowledge, we are the first to incorporate non-local self-similarity into contrastive learning for positive/negative sampling. The advantage of the proposed non-local sampling is twofold. First, the non-local self-similarity sampling strategy would naturally guarantee more compact clusters for positives and negatives respectively, which would benefit us to differ the positives from negatives. Second, these positive non-local patches are the samples searched from real images with diverse variable information, not manually generated fake samples, which would provide more faithful information for representation. Note that, the non-local strategy is not only applicable for the positive samples, but also beneficial to the negative samples. Overall, our contributions can be summarized as follow:

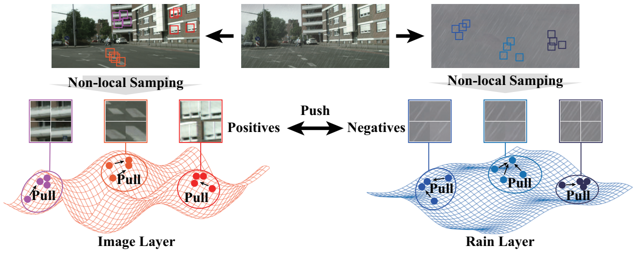


Figure 2. Most previous methods model the property of the image layer and rain layer independently in a supervised manner. In this work, we go further by considering the mutually exclusive relationship between the two layers, and propose an unsupervised non-local contrastive learning method to learn mutually exclusive relationship by pushing away the image positives and rain negatives. Moreover, the non-local self-similarity has been exploited to improve the positive/negative sampling with discriminative representation.

- We propose a non-local contrastive learning method (NLCL) for unsupervised image deraining. Compared with previous deraining methods, we not only exploit the specific property of the image and rain layers, but also model the contrastive relationship between them for better decoupling the rain layer from the clean image.
- We connect the contrastive learning with the non-local self-similarity. The non-local patch sampling strategy naturally endows the positive/negative samples with more compact and discriminative representation for better decomposition. In addition, we provide an guidance of how to design a good encoder for better embedding in NLCL.
- We conduct extensive experiments on both synthetic and real-world datasets, and show that NLCL outperforms favorably state-of-the-art methods on real image deraining.

2. Related Work

Single Image Deraining. Here, we mainly focus on the learning-based deraining methods. Most of the existing methods are full-supervised which require a large number of paired rainy and clean images as training samples [5,12,13,19,28,30,41,47,54,57,62]. Fu *et al.* [12] first introduced the end-to-end residual CNN for rain streaks removal. Latter, the multi-stage [54], multi-scale [22], density [59], and attention [30] have been widely utilized for better representation. Unfortunately, the domain gap between the complex real rain and the simplified synthetic rain would limit their generalization in real scenes. The semi-supervised deraining models [20,48,49,55,56] could alleviate this issue to some extent. For example, Wei *et al*. [48] first proposed a semi-supervised transfer learning framework via network structure sharing for real image deraining. Recently, the unsupervised deraining methods have emerged [23,50,58,60]. Yu *et al*. [58] connected the model-driven and data-driven methods via an unsupervised learning framework. In this work, we propose a novel contrastive learning framework for unsupervised deraining. Compared with previous methods, the NLCL could further take mutual exclusive relationship between image and rain layers into consideration.

Contrastive Learning. Contrastive learning (CL) has achieved promising results in unsupervised representation learning [7–9, 18, 39, 52]. The main idea is to push the features of unrelated data (as negatives) and pull the related data (as positives), so as to learn the representations which are discriminative to the negatives and invariant between the positives. CL can be effectively applied by appropriately defining the positives and negatives in terms of the tasks, including the multi-views [42, 43], temporal coherence in video sequence [17], augmented transformation [7, 18], to name a few. Recently, researches have applied the CL to low-level applications [33, 40, 51]. Wu *et al*. [51] pulled the restored image closer to ground truth (GT) and pushed them far away from the hazy image in the representation space.

Our NLCL is significantly different from [51] in three aspects. First, our method is completely unsupervised which does not need the GT. Second, we take the estimated image and rain layers as the positive and negative, respectively. Such an explicit disentanglement between the two layers would better facilitate us to decouple the rain from the clean image. Third, [51] employs a classical instance image-level samples for contrast, while we have explored the intrinsic similarity between the patches within a single image. The self-similarity within the positive or negative would further boost more compact and structural feature space.

Non-local Self-similarity. The self-similarity serves as a powerful image prior model, which has been verified in various image restoration techniques, including filtering methods [3, 10], sparse optimization models [15, 37], and deep neural networks [2, 32, 46]. The nonlocal prior reveals a general image property that the similar small patches tend to recurrently appeared within a single image. This generic property could provide group sparsity of the image with structural representation. Beneficial from capturing the correlation among the self-similarity patches, these non-local based methods have achieved the state-of-the-art performances at the time, such as the BM3D in denoising [10], WNNM in restoration [15], and kernelGAN in blind super-resolution [2]. In this work, we demonstrate how this intrinsic property benefits the contrastive learning in terms of the positive/negative sampling, and boosts the performance in low-level image deraining task.

3. Non-local Contrastive Learning

3.1. Overview of the Framework

Given a rainy image $\boldsymbol{O}$, our goal is to decompose the rainy image into a clean background layer $\boldsymbol{B}$ and a rain layer $\boldsymbol{R}$. The degradation procedure can be formulated as:

$$\boldsymbol{O} = \boldsymbol{B} + \boldsymbol{R}. \tag{1}$$

Note that, although we follow this simple decomposition framework [31], this does not mean the proposed method only handles the rain streak. The proposed method can well restore the heavy rain with haze or veil artifacts. Thus, the image deraining task can be formulated as an ill-posed inverse problem with following optimization function:

$$\mathcal{L}_{decom} = ||\boldsymbol{B} + \boldsymbol{R} - \boldsymbol{O}||_F^2 + \delta P_b(\boldsymbol{B}) + \lambda P_r(\boldsymbol{R}), \tag{2}$$

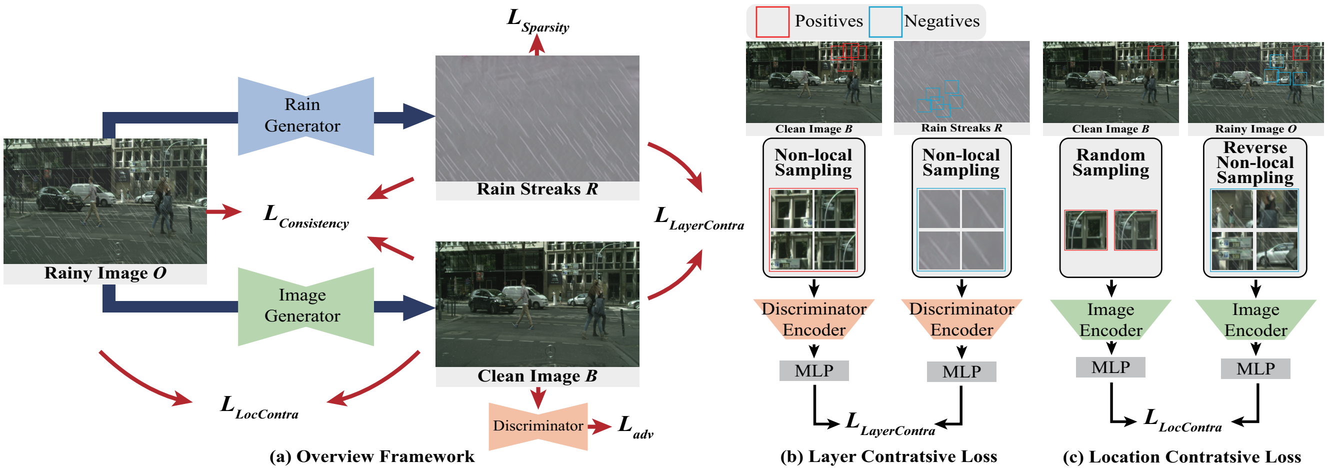
where the first term is self-consistent loss, namely the data fidelity term, P_b and P_r denote the prior knowledge for the clean image and rain streaks, respectively. Thanks to the sparsity of the rain streaks in space, in this work, we regularize the rain layer with the L_1 constraint: $P_r(\boldsymbol{R}) = ||\boldsymbol{R}||_1$ favoring the rain streaks with large discontinuities. On the other hand, for the clean images, we employ the adversarial loss [14] to learn the distribution mapping differing the rainy image from clean image:

$$P_b(\boldsymbol{B}) = \mathbb{E}_{\boldsymbol{B}}\left[\log D(\boldsymbol{B})\right] + \mathbb{E}_{\boldsymbol{O}}\left[\log(1 - D(G_{\boldsymbol{B}}(\boldsymbol{O})))\right], \tag{3}$$

where D is the discriminator, and $G_{\boldsymbol{B}}$ is the generator for the clean image. The corresponding network of the decomposition-based architecture is shown in Fig. 3(a), which consists of two branches to restore the background ($G_{\boldsymbol{B}}$) and extract the rain ($G_{\boldsymbol{R}}$).

Most of the existing restoration methods follow the decomposition framework in Eq. (2) with different handcrafted or learned priors, where they only consider the clean image or rain layer separately. That is to say, the Eq. (2) mainly focuses on modelling of the statistical property of the signal itself. However, it has neglected the relationship between clean image $\boldsymbol{B}$, rain layers $\boldsymbol{R}$, and observed image $\boldsymbol{O}$. In this work, we argue the relationship among these components can further help to distinguish them from each other. We introduce the contrastive learning to model the relationship between different components. Specifically, we evolve the relation $\mathcal{L}_{LayerCon}$ between the clean image $\boldsymbol{B}$ and rain layer $\boldsymbol{R}$, also relation $\mathcal{L}_{LocCon}$ between clean image $\boldsymbol{B}$ and observed image $\boldsymbol{O}$. Thus, the full objective function including the decomposition constraint and contrastive loss is formulated as:

$$\mathcal{L}_{overall} = \mathcal{L}_{decom} + \mu\mathcal{L}_{LayerCon}(\boldsymbol{B}, \boldsymbol{R}) + \sigma\mathcal{L}_{LocCon}(\boldsymbol{B}, \boldsymbol{O}). \tag{4}$$



(a) Overview Framework (b) Layer Contratsive Loss (c) Location Contratsive Loss

Figure 3. Overview of architecture of the proposed method. (a) The NLCL consists of two sub-networks to extract the background and the rain layers respectively with two additional contrastive constraints. (b) The layer contrastive between the clean image and rain streaks with the non-local sampling for both positives and negatives. (c) The location contrastive between the clean image and rainy image with the reverse non-local sampling for negatives.

Layer Contrastive: First, the clean image $\boldsymbol{B}$ and the rain layer $\boldsymbol{R}$ are vastly different, in which the rain streaks are simple and directional line-pattern, while the natural images are complex yet meaningful structures such as edges and textures. The *dissimilarity* between the $\boldsymbol{B}$ and $\boldsymbol{R}$, as two different categories, can be well modelled by CL as negative pairs. And it is very reasonable to take the patches in the same image as the positive samples. The important sampling strategy and encoder in CL will be discussed in next subsection. Referring to the rain patches $\boldsymbol{p}_{\boldsymbol{R}_i}$ as negatives, while the background patches as the positives $\boldsymbol{p}_{\boldsymbol{B}_i}$, the contrastive loss between the two layers $\boldsymbol{B}$ and $\boldsymbol{R}$ can be formulated:

$$\begin{aligned}\mathcal{L}_{LayerCon} = &-\frac{1}{N_B}\sum_{k=1}^{N_B}\sum_{i=1}^{N_B}\frac{\exp(\boldsymbol{f}_{\boldsymbol{B}_i}\cdot\boldsymbol{f}_{\boldsymbol{B}_k}/\tau)}{\sum_{j=1}^{N_R}\exp(\boldsymbol{f}_{\boldsymbol{B}_i}\cdot\boldsymbol{f}_{\boldsymbol{R}_j}/\tau)} \\ &-\frac{1}{N_R}\sum_{m=1}^{N_R}\sum_{j=1}^{N_R}\frac{\exp(\boldsymbol{f}_{\boldsymbol{R}_j}\cdot\boldsymbol{f}_{\boldsymbol{R}_m}/\tau)}{\sum_{i=1}^{N_B}\exp(\boldsymbol{f}_{\boldsymbol{R}_j}\cdot\boldsymbol{f}_{\boldsymbol{B}_i}/\tau)},\end{aligned} \quad (5)$$

where $\boldsymbol{f}_{\boldsymbol{B}_i} = E_D(\boldsymbol{p}_{\boldsymbol{B}_i}), \boldsymbol{f}_{\boldsymbol{R}_j} = E_D(\boldsymbol{p}_{\boldsymbol{R}_j})$, τ denotes the scale temperature parameter [7]. E_D is the encoder of contrastive network. The features $\boldsymbol{f}_{\boldsymbol{B}_k}$ are extracted from the non-local patches $\boldsymbol{p}_{\boldsymbol{B}_k}$ of $\boldsymbol{p}_{\boldsymbol{B}_i}$, while the $\boldsymbol{f}_{\boldsymbol{R}_m}$ are extracted from the non-local patches $\boldsymbol{p}_{\boldsymbol{R}_m}$ of $\boldsymbol{p}_{\boldsymbol{R}_j}$. N_B and N_R denote the sample numbers of positives and negatives. The layer contrastive could facilitate us to better push the image layer away from rain layer, and pull each layer further to different clusters.

Location Contrastive: Second, we can observe that the clean image $\boldsymbol{B}$ and the observed image $\boldsymbol{O}$ are visually close to each other, since the rain streaks $\boldsymbol{R}$ are much simpler than $\boldsymbol{B}$. The *similarity* between patches of the same location in $\boldsymbol{B}$ and $\boldsymbol{O}$, as the same view, can be well modelled as the positive samples. Consequently, we set the patches with different locations as the negative samples. Note that, here for location contrastive, there should be only one positive sample, since the location correspondence is exactly one-to-one. The encoder of image generator $E_{G_{\boldsymbol{B}}}$ is utilized to extract the patch features, denoted as $\boldsymbol{v}_{\boldsymbol{O}_i} = E_{G_{\boldsymbol{B}}}(\boldsymbol{p}_{\boldsymbol{O}_i})$, and $\boldsymbol{v}_{\boldsymbol{B}_i} = E_{G_{\boldsymbol{B}}}(\boldsymbol{p}_{\boldsymbol{B}_i})$. Thus, the location contrastive loss is formulated as:

$$\mathcal{L}_{LocCon} = \sum_{i=1}^{N}\frac{\exp(\boldsymbol{v}_{\boldsymbol{O}_i}\cdot\boldsymbol{v}_{\boldsymbol{B}_i}/\tau)}{\exp(\boldsymbol{v}_{\boldsymbol{O}_i}\cdot\boldsymbol{v}_{\boldsymbol{B}_i}/\tau)+\sum_{j=1}^{N}\exp(\boldsymbol{v}_{\boldsymbol{O}_j}\cdot\boldsymbol{v}_{\boldsymbol{B}_i}/\tau)}, \quad (6)$$

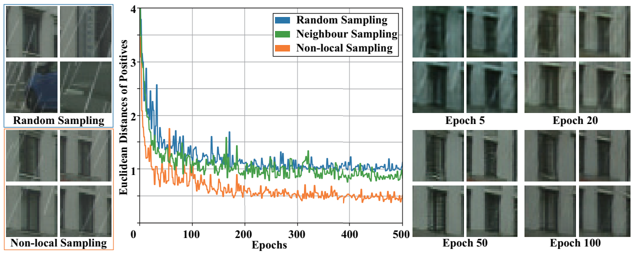
where N is the negative sample numbers. The location contrastive constrains the restored background patches $\boldsymbol{p}_{\boldsymbol{B}_i}$ at location i to be related (positive) with the corresponding input patches $\boldsymbol{p}_{\boldsymbol{O}_i}$ in comparison to other random patches $\boldsymbol{p}_{\boldsymbol{O}_j}$, so as to retain the image content.

3.2. Non-local Sampling Strategy

In contrastive learning, the negatives are the samples which should be discriminated by the learned representations, while the positives are highly related and possess the invariance in the learned representations. The previous methods usually use the augmentations to construct the single instance positives and randomly sampling as the negatives [7]. Note that, the self-similarity is a generic and powerful prior knowledge. In this work, we introduce the non-local self-similarity to automatically select both positive and negative samples within a single image. We employ the block matching [10] with L_2 Euclidian distance to measure the dissimilarity/similarity in image space:

$$Dist(\boldsymbol{p}_i, \boldsymbol{p}_{i_R}) = ||\boldsymbol{p}_i - \boldsymbol{p}_{i_\Omega}||^2, \quad (7)$$

where $\boldsymbol{p}_i$ is the query patch, $\boldsymbol{p}_{i_\Omega}$ are the searched patches in the support set Ω. We take the top-k smallest *Dist*() as the similar patches, while the top-k largest *Dist*() can be regarded as the dissimilar patches. On one hand, the non-local sampling with similar structures would greatly ease the



(a) Example of Positives (b) Positive distances decrease during epoch (c) Progressive restoration of positives

Figure 4. Effectiveness of the non-local sampling strategy. (a) Example of the randomly sampled positives with low similarity in comparison with the non-local self-similar sampled positives. (b) The Euclidean distances of positives decrease rapidly to a relative low level by non-local sampling strategy when compared with the random strategy, indicating the self-similarities are gradually learned and the patches are more relevant in the deraining procedure. (c) The similar patches guide each other to gradually restore the clean image and remove the randomly distributed rains.

learning difficulty. On the other hand, the small perturbation within the similar samples would further improve the diversity. Moreover, the patches cropped from the image itself would provide more reliable representation learning. The non-local sampling strategy can be applied for sampling the positive and negative. Here we briefly describe how we use the non-local sampling in very flexible ways.

Non-local Sampling in Layer Contrastive. In layer contrastive, the clean image and rain streaks can be regarded as two distinct categories where they have intra-class similarity and inter-class dissimilarity. Our principle is that the positive samples (clean image patches in $\boldsymbol{B}$) should be pulled together as much as possible, so is the negative samples (rain streak patches in $\boldsymbol{R}$) which can also be pulled together. That is to say, we enforce the non-local sampling on both the positive and negative samples. Compared with single positive instance, the multiple non-local positive samples would benefit us to improve the feature representation. The recent research has also shown that positives from multiple instances could improve the representations if sampled appropriately (with supervised labels [26] or multiple modalities [17]). Moreover, compared with the random negative samples, the non-local sampling could additionally model the relationship within the samples.

To illustrate this, Figure 4 shows the superiority of the non-local sampling on positives. Figure 4(a) shows an example of random and non-local sampled patches. The non-local patches possess the structure self-similarities in comparison with the random sampled ones. During training, we continually re-sample the non-local positives and calculate the similarity by Euclidean distance. Compared with random sampling or neighbour sampling which samples the surrounding patch neighbours, the distances of non-local positives decrease rapidly, and converge at a relatively low level, which indicates the self-similarities are gradually learned and the patches are more relevant in the restoration procedure [Fig. 4(b)]. Therefore, by maximizing the correlations of positives, the self-similar patches guide each other to gradually restore the clean image and remove the rain streaks. The progressive deraining results are shown in Fig. 4(c).

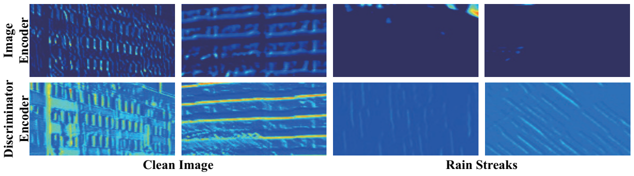


Figure 5. Effectiveness of the discriminator encoder. The first row shows the features extracted from image encoder. Although the extracted features in two different images are clear, the image generator has nearly no response to the rain streaks. The second row shows the features extracted from the discriminator encoder. The extracted features in two images and rain streaks are both clear and discriminative. This strongly supports the effectiveness of the discriminator serving as the encoder for the image and rain layer.

Non-local Sampling in Location Contrastive. The observed image $\boldsymbol{O}$ and clean image $\boldsymbol{B}$ are very similar to each. In location contrastive, the goal is to retain the image content and remove the rain streaks in observed image, which is exactly a image-to-image translation task. Thus, we follow the CUT [40] by setting the patches of the same location in $\boldsymbol{B}$ and $\boldsymbol{O}$ as the positive samples with a large batch size. The previous methods including CUT randomly select the different patches as the negative. However, it is more reasonable that the more dissimilar from the positive sample, the better the negative sample is. This motivates us to still use the non-local sampling strategy to construct the negative patches. Instead of calculating the nearest top-k samples, we choose the farthest top-k samples (the largest distance) which means they are mostly different from the target positive. We name this negative sampling as the reverse non-local sampling.

3.3. Feature Encoder

In contrastive learning, the feature encoder is to map the inputs to the embedding low-dimensional feature representation space that facilitates the measurement of the distances between positive and negative samples. It has been recognized that for different CL tasks, the choice of the encoder would vastly influence the final performance [27]. In this work, we also demonstrate that the encoder is indeed tasks dependent for low-level restoration tasks, and explore different encoders for both the layer and location contrastive constraints intuitively and experimentally.

As for the layer contrastive, the goal is to differ the rain streaks from the clean image, which has been analyzed that this is analog to a classification problem. That is to say, the encoder of the layer contrastive should extract the high-level semantic about the category information. The discriminator is in line with the layer contrastive encoder, which can differ the image from non-image component including the rain

Table 1. Quantitative comparisons with SOTA unsupervised methods on synthetic and real datasets.

Methods	RainCityscapes			SPA		
	PSNR	SSIM	NIQE	PSNR	SSIM	NIQE
DSC	24.91	0.7603	6.17	33.71	0.9127	9.82
DIP	22.45	0.6936	7.86	30.36	0.8422	9.97
CycleGAN	24.86	0.7906	3.68	33.54	0.9127	**6.67**
UDGNet	25.16	**0.8749**	5.31	29.67	0.9299	9.50
CUT	25.21	0.8225	4.08	32.97	0.9434	9.60
NLCL	**26.46**	0.8666	**3.67**	**33.82**	**0.9468**	9.55

streaks. As for the location contrastive, the clean image and the observed rainy image are very similar to each other, in which the clean image is the dominant component in rainy image. In other words, the encoder of the location contrastive should well extract the image features. The image generator can satisfactorily achieve this goal.

To verify our hypothesis, Figure 5 visualizes the embedded features map encoded by different encoders: image generator and discriminator. The first row shows the features extracted from image encoder, and the second row shows the features extracted from discriminator encoder. We select two different clean images and two different rain streaks as the example. We can observe that the image generator could effectively extract the image structures, while it cannot extract any informative information from the rain streaks. On the contrary, the line patterned rain streaks and the image structure can be clearly observed in the features extracted by the discriminator encoder. The discriminator focuses on the distinguishable features of image and non-image factors to perform the classification task, which matches the layer contrastive learning task better.

4. Experiments

4.1. Implementation Details

We utilize the same ResNet architectures [24] for both the background extraction and rain extraction. Please refer to the supplementary for details. PatchGAN [21] is employed for the discriminator. We first calculate the top-k non-local patches in image space, then obtain the multilayer features [40] from the encoder, and finally embed the non-local features through a two-layer MLP with 256 units. The sampling number N, N_B, and N_R are set as 256, 8, 256. The encoder updating follows the setting of MoCo [18], using momentum value 0.99 and temperature 0.77. The balance weights for each loss λ, δ, μ, σ are set as $0.1, 1, 1, 1$. During the training, the original images are randomly cropped into 256×256 as input without any augmentation. We adopt the Adam optimizer and train the network with learning rate 0.0001, and batch size 4 on four RTX 2080TI GPUs.

4.2. Datasets and Experimental Settings

We conduct the experiments on both synthetic dataset RainCityscapes [19] and real dataset SPA [45]. To simulated the real situation, we split the RainCityscapes with 1400 for training and 175 for testing. Note that we have no access to the ground truth and can only learn in an unsupervised manner. For the real dataset, we obtain 2000 rainy images from SPA for training and 200 rainy images for testing. For a fair comparison, we mainly select the unsupervised methods, including the optimization-based DSC [36], CNN-based DIP [44], GAN-based CycleGAN [61], contrastive learning-based CUT [40], and optimization-driven deep CNN [58]. Furthermore, we compare with state-of-the-art supervised JORDER-E [53] on the real rainy images. We employ the full-reference PSNR and SSIM to evaluate the deraining performance, and also the no-reference natural image quality evaluator (NIQE) [38] for comprehensive evaluation.

4.3. Comparisons with State-of-the-arts

In Table 1, we report the quantitative results on Cityscape and SPA. These datasets mainly contains the rain streaks with different visual appearances without the veiling in heavy rainy images. The quantitative results of NLCL mostly outperform state-of-the-art methods, which verifies the robustness of the proposed NLCL. Note that UDGNet mainly takes advantage of the directionality of the rain streaks, which is very suitable for the directional rain streaks in Cityscape. CycleGAN is an image generation method aiming at visually natural appearance, which matches the goal of NIQE, but cannot well preserve the original image content in terms of the relatively low PSNR. We do admit that our NLCL is not designed for the quantitative index on rain streaks. Instead, our philosophy is to unsupervisedly handle the real rains. To validate this, in Fig. 6, we compare with the state-of-the-art on real-world rainy images, which contains both the rain streak and veiling. NLCL consistently achieves more natural and better visual results, which not only remove the rain streaks but also the veiling artifacts. The results strongly support the effectiveness of the CL and non-local sampling for better distinguishing real rain from image texture.

4.4. Ablation Study

The Effectiveness of the Non-local Sampling Strategy. In Table 2, we compare the different sampling strategies for both the positives and negatives, including the random sampling, neighbour sampling (8 nearest neighbour patches), and the non-local sampling. These experiments are all performed on the layer contrastive. Compared with the random sampling, the non-local sampling for both the positive and negative could obviously improve the restoration results. That is to say, the non-local sampling is favorable to learn the image and rain streaks similarity, thus indeed reduces the variance within the positives and negatives, and at the same time enlarge the discrepancy between them. The neighbour sampling could slightly improve the results, while the non-local sampling still obtains the best performance.

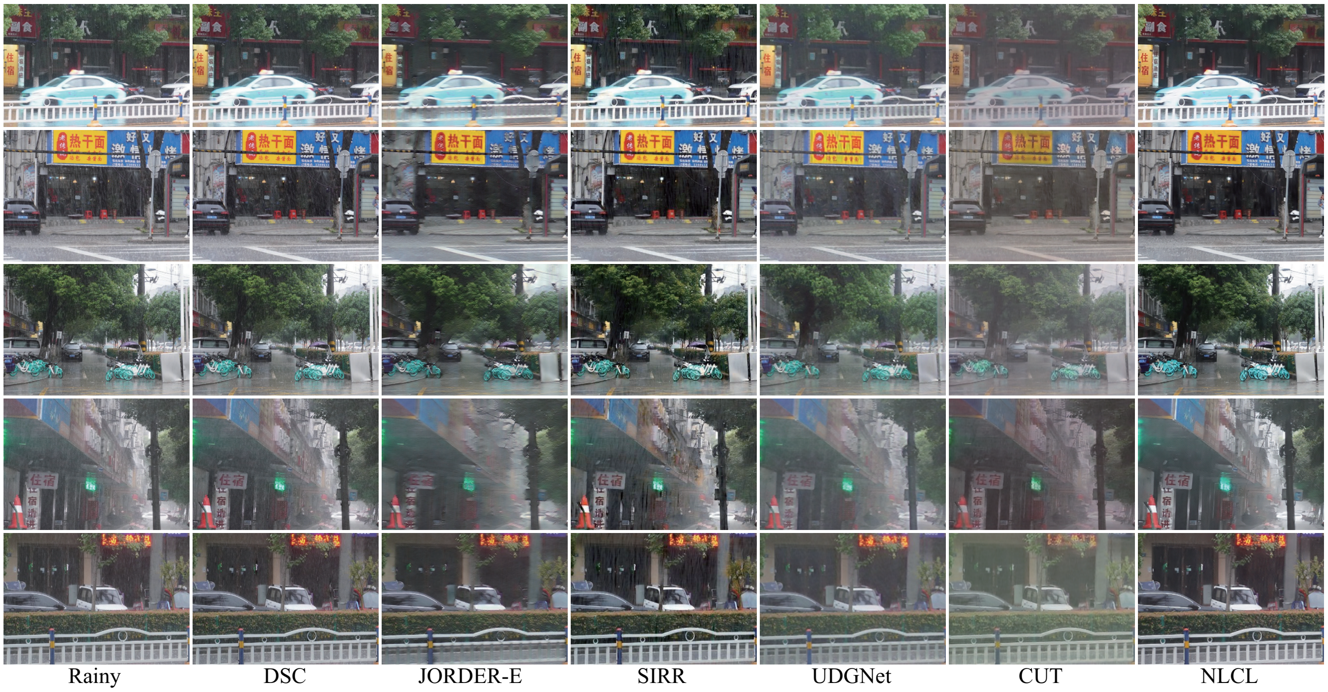


Figure 6. Visual comparisons in real rainy scenes including both rain streaks and heavy haze. We suggest to view the zoomed results on PC.

Table 2. Ablation on different sampling strategies.

Positive	Negative	PSNR	SSIM
Random	Random	25.83	0.8471
Neighbour	Random	26.03	0.8491
Neighbour	Neighbour	25.97	0.8477
Non-local	Random	26.18	0.8489
Random	Non-local	26.16	0.8531
Non-local	Non-local	**26.46**	**0.8666**

Table 3. The choice of different feature encoders.

Encoder	PSNR	SSIM	NIQE
Image Generator	24.86	0.8046	3.83
Image-Rain Generator	24.12	0.8023	3.95
Discriminator	**26.46**	**0.8666**	**3.67**

Table 4. The analysis of optimal sampling number.

Neg Pos	64	128	256	512
4	23.21	26.11	26.19	26.30
8	24.55	26.30	**26.46**	26.42
16	24.54	25.96	26.02	26.11
32	23.49	25.02	25.40	25.37

Table 5. Different strategy ablations on location contrast.

Ablations	PSNR	SSIM	NIQE
Random Sampling	26.33	0.8617	3.81
Discriminator Encoder	24.71	0.8476	3.94
Sample Number 64	25.03	0.8646	3.88
Sample Number 128	26.14	0.8604	3.75
Sample Number 512	25.98	0.8594	3.70
NLCL	**26.46**	**0.8666**	**3.67**

The Choice of Different Feature Encoders. The choice of the encoder for latent feature space is very important. In Table 3, we test different encoders for layer contrastive feature embedding. First, we take the image generator as the feature encoder for both the image and rain layers. Second, we utilize the image generator and rain generator as the feature encoder for the image and rain layer, respectively. Third, we employ the discriminator as the feature encoder for both the image and rain layers. The discriminator encoder has achieved the best result, which verifies the discriminator is suitable to distinguish the image from rain patches.

The Influence of the Non-local Sampling Number. We show how the sampling numbers affect the derain result in Table 4. The PSNR increases when the positive sizes grow to an appropriate number, and then decrease since the excessive positives are somehow dissimilar. 8 positives and 256 negatives obtain the best performance. The reason is that most of the rain have the similar line patterns, thus more non-local similar patches can be found to boost the learning than complex image patches. Moreover, the sampling number is not the larger the better, since enforcing the dissimilar patches to be similar may violate the similar assumption.

The Strategies of Location Contrast. We further study the strategies of location contrast in Table 5, which shows the improvement from reverse non-local sampling. Moreover, the image generator encoder is much better than discriminator to preserve the image content in location contrastive. 256 is an appropriate number for the sampling number.

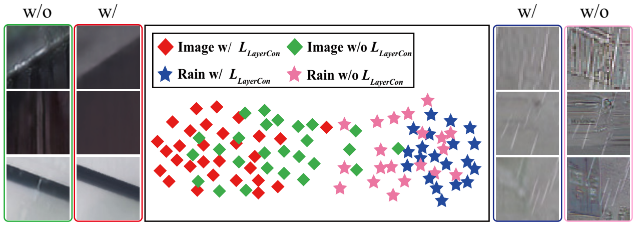
The Effectiveness of Each Loss. In Table 6, we show how

Table 6. Effectiveness of each loss in NLCL.

Model	PSNR	SSIM	NIQE
w/o $\mathcal{L}_{adv}$	21.55	0.7984	5.03
w/o $\mathcal{L}_1$	26.33	0.8566	3.74
w/o $\mathcal{L}_{LocCon}$	25.20	0.8469	3.85
w/o $\mathcal{L}_{LayerCon}$	24.12	0.8402	3.98
NLCL	**26.46**	**0.8666**	**3.67**

Table 7. The model size and inference time under image $256 * 256$.

Method	DSC	JORDER-E	CycleGAN	UDGNet	CUT	NLCL
Size(MB)	–	16.7	45.6	5.7	45.6	2.6
Time(s)	33.95	0.128	0.0144	0.0170	0.0135	0.0098



(a) Image (b) Visualization (c) Rain

Figure 7. The effectiveness of the layer contrastive for better image and rain streaks decomposition. (a) and (c) show the decoupled rain and image patches w/ and w/o the layer contrastive. (b) visualizes the low-dimensional distributions w/ and w/o the layer contrastive.

each loss contributes to the final result. The $\mathcal{L}_{LocCon}$ and $\mathcal{L}_{LayerCon}$ aim to learn the correlations between the rainy-clean images, and the rain-image layers, which could greatly improve the deraining results. The self-consistency and adversarial loss are the baseline of our model. $\mathcal{L}_1$ sparse loss could slightly improve the performance.

4.5. Analysis and Discussion

Effectiveness of Contrastive Learning. In Fig. 7(b), we perform the tSNE to visualize the distribution of the decomposed image and rain layer w/ and w/o contrastive constraint. Without the layer contrastive, the distribution of the green rhombus (image) and the pink pentacle (rain streaks) are divergent. Moreover, they are mixed with each other which means they are still indistinguishable. On the contrary, with the layer contrastive, the distribution of the red rhombus (image) and the dark blur pentacle (rain streaks) are focused and distinguishable. In Fig. 7(a) and (c), with contrastive loss, the image and rain layers are better decoupled.

Visualization of Self-similarity Patches. We visualize the top 5 non-local positives and negatives of both light and heavy rain conditions in Fig. 8. The extremely heavy rain would unavoidably increase the difficulty in non-local searching. The two-stage searching framework could be used, where the coarse clearer results are obtained before we search the non-local patches in the intermediated results. The positives and negatives are similar to that of the query key. The similarity is real and reliable with slight difference, instead of synthesis or fixed transformations. This intrinsic property facilitates us to learn discrimination representation.

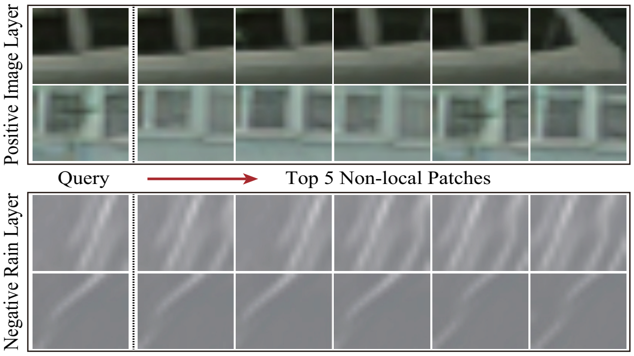


Figure 8. The visualization of the Top5 non-local searched patches.

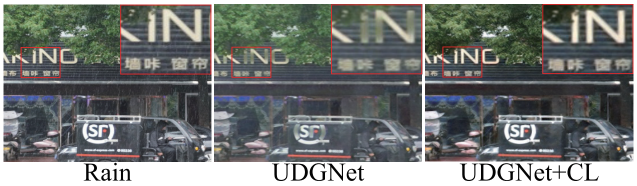


Figure 9. The benefits of the NLCL when applied to UDGNet.

The Benefit of the NLCL Strategy for Other Method. Our NLCL is a general prior which can be naturally embedded into the existing methods for better decomposition. Here, we take the unsupervised deraining method UDGNet [58] as example. As shown in Fig. 9, although UDGNet could well remove the rain streaks without NLCL, the image structures have been unexpectedly removed along with rain. The result of UDGNet + NLCL is much better, such as the text.

Model Size and Running Time. In inference phase, only the image generator is employed, making NLCL very fast and small, as shown in Table 7. But in training phase, the additional nonlocal self-similarity searching is somewhat time-consuming. Normally, we take one day and a half for training 1400 images, which is the main limitation of NLCL. Speeding up the training time is one of our future work.

5. Conclusion

In this paper, we propose a novel non-local contrastive learning method, which explores the powerful self-similarity property within the image. Our method is totally unsupervised which can automatically decouple the image from the rain artifacts. We show that our non-local sampling strategy can be used to learn meaningful representations for both positives and negatives. Especially, the proposed non-local sampling strategy enriches the faithful, diverse and structural representation for both negatives and positives. Moreover, we provide an guidance of how to select an appropriate encoder for better feature embedding. Extensive experiments demonstrate the effectiveness of the proposed method.

Acknowledgements. This work was supported by National Natural Science Foundation of China under Grant No. 61971460 and No. 62101294, and Equipment Pre-Research Foundation under Grant No. 6142113200304.

References

[1] Chris H Bahnsen and Thomas B Moeslund. Rain removal in traffic surveillance: Does it matter? *IEEE Trans. Intell. Transp. Syst.*, 20(8):2802–2819, 2018. 1

[2] Sefi Bell-Kligler, Assaf Shocher, and Michal Irani. Blind super-resolution kernel estimation using an internal-gan. In *Adv. Neural Inform. Process. Syst.*, pages 284–293, 2019. 3

[3] Antoni Buades, Bartomeu Coll, and J-M Morel. A non-local algorithm for image denoising. In *IEEE Conf. Comput. Vis. Pattern Recog.*, volume 2, pages 60–65, 2005. 2, 3

[4] Yi Chang, Luxin Yan, and Sheng Zhong. Transformed low-rank model for line pattern noise removal. In *Int. Conf. Comput. Vis.*, pages 1726–1734, 2017. 1

[5] Chenghao Chen and Hao Li. Robust representation learning with feedback for single image deraining. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 7742–7751, 2021. 2

[6] Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Eur. Conf. Comput. Vis.*, pages 801–818, 2018. 1

[7] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In *Int. Conf. on Mach. Learn.*, pages 1597–1607, 2020. 2, 3, 4

[8] Ting Chen, Simon Kornblith, Kevin Swersky, Mohammad Norouzi, and Geoffrey Hinton. Big self-supervised models are strong semi-supervised learners. *arXiv preprint arXiv:2006.10029*, 2020. 3

[9] Xinlei Chen, Haoqi Fan, Ross Girshick, and Kaiming He. Improved baselines with momentum contrastive learning. *arXiv preprint arXiv:2003.04297*, 2020. 3

[10] Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian. Image denoising by sparse 3-d transform-domain collaborative filtering. *IEEE Trans. Image Process.*, 16(8):2080–2095, 2007. 3, 4

[11] Xueyang Fu, Jiabin Huang, Xinghao Ding, Yinghao Liao, and John Paisley. Clearing the skies: A deep network architecture for single-image rain removal. *IEEE Trans. Image Process.*, 26(6):2944–2956, 2017. 1

[12] Xueyang Fu, Jiabin Huang, Delu Zeng, Yue Huang, Xinghao Ding, and John Paisley. Removing rain from single images via a deep detail network. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 3855–3863, 2017. 2

[13] Xueyang Fu, Qi Qi, Zheng-Jun Zha, Yurui Zhu, and Xinghao Ding. Rain streak removal via dual graph convolutional network. In *AAAI*, pages 1–9, 2021. 2

[14] Ian J Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron C Courville, and Yoshua Bengio. Generative adversarial networks. In *Adv. Neural Inform. Process. Syst.*, pages 1050–1060, 2014. 3

[15] Shuhang Gu, Lei Zhang, Wangmeng Zuo, and Xiangchu Feng. Weighted nuclear norm minimization with application to image denoising. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 2862–2869, 2014. 3

[16] Raia Hadsell, Sumit Chopra, and Yann LeCun. Dimensionality reduction by learning an invariant mapping. In *IEEE Conf. Comput. Vis. Pattern Recog.*, volume 2, pages 1735–1742, 2006. 2

[17] Tengda Han, Weidi Xie, and Andrew Zisserman. Self-supervised co-training for video representation learning. In *Adv. Neural Inform. Process. Syst.*, pages 5679–5690, 2020. 3, 5

[18] Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick. Momentum contrast for unsupervised visual representation learning. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 9729–9738, 2020. 2, 3, 6

[19] Xiaowei Hu, Chi-Wing Fu, Lei Zhu, and Pheng-Ann Heng. Depth-attentional features for single-image rain removal. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 8022–8031, 2019. 2, 6

[20] Huaibo Huang, Aijing Yu, and Ran He. Memory oriented transfer learning for semi-supervised image deraining. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 7732–7741, 2021. 2, 3

[21] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 1125–1134, 2017. 6

[22] 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 *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 8346–8355, 2020. 1, 2

[23] Xin Jin, Zhibo Chen, Jianxin Lin, Zhikai Chen, and Wei Zhou. Unsupervised single image deraining with self-supervised constraints. In *IEEE Int. Conf. Image Process.*, pages 2761–2765, 2019. 2, 3

[24] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *Eur. Conf. Comput. Vis.*, pages 694–711, 2016. 6

[25] Li-Wei Kang, Chia-Wen Lin, and Yu-Hsiang Fu. Automatic single-image-based rain streaks removal via image decomposition. *IEEE Trans. Image Process.*, 21(4):1742–1755, 2011. 2

[26] Prannay Khosla, Piotr Teterwak, Chen Wang, Aaron Sarna, Yonglong Tian, Phillip Isola, Aaron Maschinot, Ce Liu, and Dilip Krishnan. Supervised contrastive learning. In *Adv. Neural Inform. Process. Syst.*, volume 33, 2020. 5

[27] Phuc H Le-Khac, Graham Healy, and Alan F Smeaton. Contrastive representation learning: A framework and review. *IEEE Access*, 2020. 5

[28] Ruoteng Li, Loong-Fah Cheong, and Robby T 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

[29] Ruoteng Li, Robby T Tan, Loong-Fah Cheong, Angelica I Aviles-Rivero, Qingnan Fan, and Carola-Bibiane Schonlieb. Rainflow: Optical flow under rain streaks and rain veiling effect. In *Int. Conf. Comput. Vis.*, pages 7304–7313, 2019. 1

[30] Xia Li, Jianlong Wu, Zhouchen Lin, Hong Liu, and Hongbin Zha. Recurrent squeeze-and-excitation context aggregation net for single image deraining. In *Eur. Conf. Comput. Vis.*, pages 254–269, 2018. 1, 2, 3

[31] Yu Li, Robby T Tan, Xiaojie Guo, Jiangbo Lu, and Michael S Brown. Rain streak removal using layer priors. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 2736–2744, 2016. 1, 3

[32] Ding Liu, Bihan Wen, Yuchen Fan, Chen Change Loy, and Thomas S Huang. Non-local recurrent network for image restoration. In *Adv. Neural Inform. Process. Syst.*, pages 1680–1689, 2018. 3

[33] Rui Liu, Yixiao Ge, Ching Lam Choi, Xiaogang Wang, and Hongsheng Li. Divco: Diverse conditional image synthesis via contrastive generative adversarial network. *IEEE Conf. Comput. Vis. Pattern Recog.*, 2021. 3

[34] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C Berg. Ssd: Single shot multibox detector. In *Eur. Conf. Comput. Vis.*, pages 21–37, 2016. 1

[35] Yang Liu, Ziyu Yue, Jinshan Pan, and Zhixun Su. Unpaired learning for deep image deraining with rain direction regularizer. In *Int. Conf. Comput. Vis.*, pages 4753–4761, 2021. 2

[36] Yu Luo, Yong Xu, and Hui Ji. Removing rain from a single image via discriminative sparse coding. In *Int. Conf. Comput. Vis.*, pages 3397–3405, 2015. 1, 2, 6

[37] Julien Mairal, Francis Bach, Jean Ponce, Guillermo Sapiro, and Andrew Zisserman. Non-local sparse models for image restoration. In *Int. Conf. Comput. Vis.*, pages 2272–2279, 2009. 3

[38] Anish Mittal, Rajiv Soundararajan, and Alan C Bovik. Making a "completely blind" image quality analyzer. *IEEE Sign. Process. Letters*, 20(3):209–212, 2012. 6

[39] Aaron van den Oord, Yazhe Li, and Oriol Vinyals. Representation learning with contrastive predictive coding. *arXiv preprint arXiv:1807.03748*, 2018. 3

[40] Taesung Park, Alexei A Efros, Richard Zhang, and Jun-Yan Zhu. Contrastive learning for unpaired image-to-image translation. In *Eur. Conf. Comput. Vis.*, pages 319–345, 2020. 3, 5, 6

[41] Ruijie Quan, Xin Yu, Yuanzhi Liang, and Yi Yang. Removing raindrops and rain streaks in one go. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 9147–9156, 2021. 2

[42] Yonglong Tian, Dilip Krishnan, and Phillip Isola. Contrastive multiview coding. *arXiv preprint arXiv:1906.05849*, 2019. 3

[43] Yonglong Tian, Chen Sun, Ben Poole, Dilip Krishnan, Cordelia Schmid, and Phillip Isola. What makes for good views for contrastive learning? In *Adv. Neural Inform. Process. Syst.*, pages 6827–6839, 2020. 3

[44] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Deep image prior. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 9446–9454, 2018. 6

[45] 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 *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 12270–12279, 2019. 6

[46] Xiaolong Wang, Ross Girshick, Abhinav Gupta, and Kaiming He. Non-local neural networks. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 7794–7803, 2018. 3

[47] Yinglong Wang, Chao Ma, and Bing Zeng. Multi-decoding deraining network and quasi-sparsity based training. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 13375–13384, 2021. 2

[48] Wei Wei, Deyu Meng, Qian Zhao, Zongben Xu, and Ying Wu. Semi-supervised transfer learning for image rain removal. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 3877–3886, 2019. 2, 3

[49] Yanyan Wei, Zhao Zhang, Yang Wang, Jicong Fan, Shuicheng Yan, and Meng Wang. Deraincyclegan: A simple unsupervised network for single image deraining and rainmaking. *arXiv preprint arXiv:1912.07015*, 2019. 2, 3

[50] 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 Trans. Image Process.*, 30:4788–4801, 2021. 2, 3

[51] Haiyan Wu, Yanyun Qu, Shaohui Lin, Jian Zhou, Ruizhi Qiao, Zhizhong Zhang, Yuan Xie, and Lizhuang Ma. Contrastive learning for compact single image dehazing. *IEEE Conf. Comput. Vis. Pattern Recog.*, 2021. 3

[52] Zhirong Wu, Yuanjun Xiong, Stella X Yu, and Dahua Lin. Unsupervised feature learning via non-parametric instance discrimination. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 3733–3742, 2018. 2, 3

[53] Wenhan Yang, Robby T Tan, Jiashi Feng, Zongming Guo, Shuicheng Yan, and Jiaying Liu. Joint rain detection and removal from a single image with contextualized deep networks. *IEEE Trans. Image Process.*, 42(6):1377–1393, 2019. 6

[54] 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 *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 1357–1366, 2017. 1, 2

[55] Rajeev Yasarla, Vishwanath A Sindagi, and Vishal M Patel. Syn2real transfer learning for image deraining using gaussian processes. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 2726–2736, 2020. 2, 3

[56] Yuntong Ye, Yi Chang, Hanyu Zhou, and Luxin Yan. Closing the loop: Joint rain generation and removal via disentangled image translation. *IEEE Conf. Comput. Vis. Pattern Recog.*, 2021. 1, 2, 3

[57] Qiaosi Yi, Juncheng Li, Qinyan Dai, Faming Fang, Guixu Zhang, and Tieyong Zeng. Structure-preserving deraining with residue channel prior guidance. In *Int. Conf. Comput. Vis.*, pages 4238–4247, 2021. 2

[58] Changfeng Yu, Yi Chang, Yi Li, Xile Zhao, and Luxin Yan. Unsupervised image deraining: Optimization model driven deep cnn. In *ACM Int. Conf. Multimedia*, pages 2634–2642, 2021. 2, 3, 6, 8

[59] He Zhang and Vishal M Patel. Density-aware single image de-raining using a multi-stream dense network. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 695–704, 2018. 1, 2

[60] 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 *AAAI*, pages 9332–9339, 2019. 2, 3

[61] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Int. Conf. Comput. Vis.*, pages 2223–2232, 2017. 6

[62] Lei Zhu, Chi-Wing Fu, Dani Lischinski, and Pheng-Ann Heng. Joint bi-layer optimization for single-image rain streak removal. In *Int. Conf. Comput. Vis.*, pages 2526–2534, 2017. 2