All-In-One Image Restoration for Unknown Corruption

Boyun Li[1], Xiao Liu[2], Peng Hu[1], Zhongqin Wu[2], Jiancheng Lv[1], Xi Peng[1*]
[1] College of Computer Science, Sichuan University.
[2] TAL Education, China.
{liboyun.gm, penghu.ml, pengx.gm}@gmail.com; {liuxiao15, wuzhongqin}@tal.com;
lvjiancheng@scu.edu.cn

Abstract

In this paper, we study a challenging problem in image restoration, namely, how to develop an all-in-one method that could recover images from a variety of unknown corruption types and levels. To this end, we propose an All-in-one Image Restoration Network (AirNet) consisting of two neural modules, named Contrastive-Based Degraded Encoder (CBDE) and Degradation-Guided Restoration Network (DGRN). The major advantages of AirNet are two-fold. First, it is an all-in-one solution which could recover various degraded images in one network. Second, AirNet is free from the prior of the corruption types and levels, which just uses the observed corrupted image to perform inference. These two advantages enable AirNet to enjoy better flexibility and higher economy in real world scenarios wherein the priors on the corruptions are hard to know and the degradation will change with space and time. Extensive experimental results show the proposed method outperforms 17 image restoration baselines on four challenging datasets. The code is available at https://github.com/XLearning-SCU/2022-CVPR-AirNet.

1. Introduction

Single image restoration aims to generate a visually pleasant high-quality image from a given degraded correspondence, *e.g.*, noisy, rainy or hazy image. During past years, image restoration has been widely used in a number of real world applications, ranging from autopilot to medical imaging and surveillance.

Although promising results have been achieved in a specific area, such as denosing [5, 27, 40, 53–55], deblurring [12, 32–34], deraining [10, 11, 17, 46, 50, 52] and dehazing [1, 7, 9, 15, 19, 35, 37, 38], image restoration has encountered the following obstacles in practice. On the one hand, it is necessary to know the correct corruption (*i.e.*, degradation) for selecting a competitive model because almost all existing approaches could handle a specific degradation only. Once the degradation type even corruption ratio changed, the model would achieve undesirable performance due to the inconsistency between the real case and the prior adopted for model construction or training. On the other hand, the degradation usually changes in complex environment, *e.g.*, self-driving cars may suffer from the rainy and hazy weather consecutively even simultaneously. In summary, it is highly expected to develop an all-in-one method that is able to recover images from a variety of unknown[1] corruption types and levels, as shown in Figure 1. To the best of our knowledge, such a unspecific image restoration

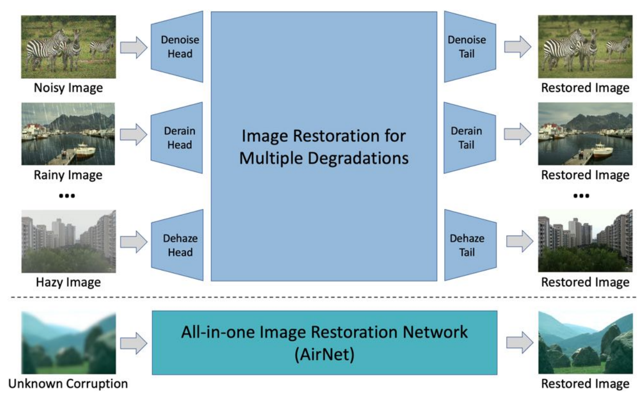

Figure 1. Illustrations of our basic idea. As shown, most of the existing multiple degradations methods handle each corruption by sending the input into a specifically designed head and using the output of the corresponding tail. Thus, they require the corruption information in advance to specify the corrected head and tail. Differently, our all-in-one image restoration network (AirNet) is free from the prior of corruption types and levels, thus enjoying better flexibility and higher economy in real world scenarios.

[1]Noticed, in this paper, the "unknown" refers to unspecific rather than unseen corruptions, and the "multiple degradations" refers to that a given image only contain a degradation but the data set will contain multiple degradations.

*Corresponding author

problem has been barely touched so far.

To tackle the aforementioned problem, we propose All-in-one Image Restoration Network (AirNet) which consists of two modules. To be specific, Contrastive-Based Degraded Encoder (CBDE) is designed to learn the degradation representation by leveraging the consistency of the images with the same degradation and the inconsistency existing into different degradations. Under the guidance of the degradation representations learned by CBDE, Degradation-Guided Restoration Network (DGRN) aims to restore the images with various degradations. Thanks to the corporation of CBDE and DGRN, AirNet enjoys two highly expected merits, *i.e.*, i) it provides an all-in-one solution to recover the images with different corruption types and ratios; ii) it is free from the prior of the corruption type and ratio. Notably, the referred all-in-one solution is different from existing so-called unified image restoration methods [3, 8, 23] in given aspects. On the one hand, the methods [3, 8, 23] have to specify the corruption type and ratio, whereas our method does not. On the other hand, they usually treat multiple degradations as a multi-task learning problem with multiple input and output heads, where each input and output head corresponds to a predetermined corruption with a given corruption ratio. In contrast, AirNet is a single pass network which does not differentiate different corruption types and ratios, thus enjoying better flexibility and higher economy.

To summarize, the contribution and novelty of this study are as below:

- As far as we know, AirNet could be one of the first methods to recover images from multiple corruptions in an all-in-one fashion. As our method does not require any degradation information for restoration in advance, it might be closer to the real world scenario.

- AirNet works in a dual manner, which contrastively learns the degradation representation from the observed images and then uses the learned degradation representation to restore the clean image. It should be pointed out that the success of contrastive learning heavily relies on the construction of positive and negative pairs. In this paper, we show a novel method that is effective to capture the inherited characteristics of multi-degradations.

- Without loss of generalizability, we conduct extensive experiments to verify the effectiveness of AirNet in denoising, deraining and dehazing, comparing with 17 baselines.

2. Related Works

In this section, we will briefly review some recent developments in the problem and the method concerned in this paper, namely, image restoration and contrastive learning.

2.1. Image Restoration

According to the focus of this paper, the existing image restoration methods could be classified into two families, *i.e.*, image restoration for single (IRSD) and multiple degradations (IRMD).

Image Restoration for Single Degradation: IRSD aims to recover a clean image from the degraded observation which is corrupted by only a specific degradation type with a fixed corruption ratio. For instance, as one of pioneering deep denoising methods, DnCNN [53] cannot handle the multi-degradation case even be failed when the noise ratio is unseen during training. Other image restoration tasks have also faced the similar challenge, such as debluring [2,12,29,32–34,36], deraining [10,17,24,42,46,49,50,52], and dehazing [1,15,20,25,28,35,37,38]. In recent, some works [13,26,39,51] show certain generalizability to different degradations. However, they need to train different models for different degradations, which are not all-in-one solutions as expected in practice.

Image Restoration for Multiple Degradations: Recently, there are some works [3, 23] shift their attention to IRMD by adopting a multi-input and -output network structure. For example, Li *et al.* [23] proposed an all-in-one model to handle multiple bad weather degradations (*e.g.* rain, fog and snow) and each degradation is specifically tackled by an encoder. Chen *et al.* [3] proposed a transformer-based image restoration method which handles multiple-degradations by using an architecture of multi-heads and multi-tails. The most similar method with our approach may be [8]. However, the method still needs to know some priors of the input (*e.g.*, noise ratio and JPEG quality) for parametrizing the network in a meta-learning manner. To summarize, although the above methods have stepped towards IRMD, they still require the degradation information in advance so that the input could be sent into the corrected head or the meta information could be generated.

2.2. Contrastive Learning

Contrastive learning [4,14,41] is the state-of-the-art unsupervised representation learning method, which aims at maximizing the similarity between positive pairs while minimizing that of negative pairs, where the positive and negative pairs are obtained through data augmentations. In recent, some studies have shown the effectiveness of contrastive learning in image restorations [43,47]. Notably, although DASR [43] and our AirNet both leverage contrastive learning to capture the degradation information, they are remarkably different in given aspects. First, the definition of positive and negative pairs is different. In fact, the success of contrastive learning heavily relies on the construction of

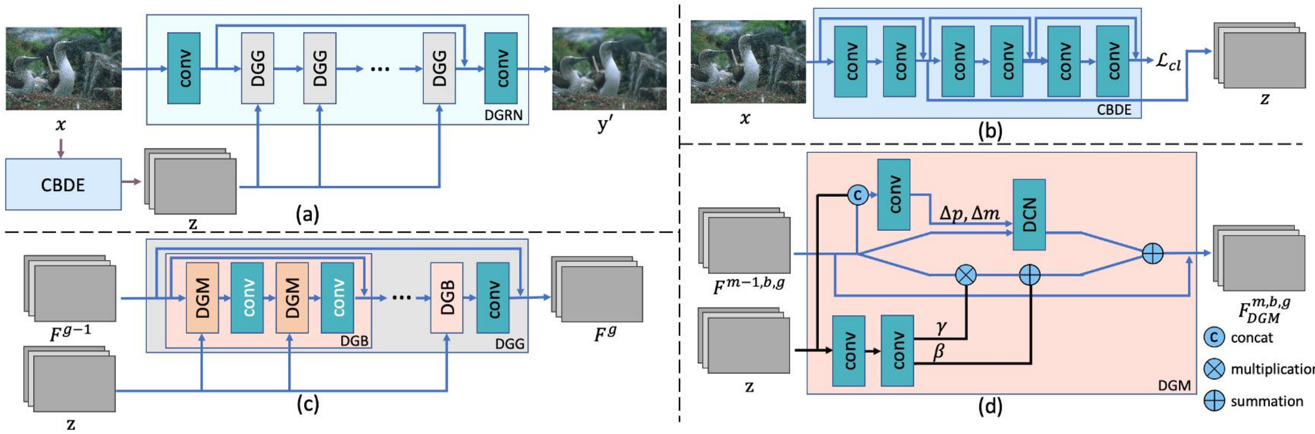

Figure 2. Architecture of the proposed AirNet. (a) All-in-one Image Restoration Network (AirNet); (b) Contrastive-Based Degradation Encoder (CBDE); (c) Degradation-Guided Groups (DGG); (d)Degradation-Guided Module (DGM).

positive and negative pairs which is thus the focus of a number of works [4]. Second, the task is different. In brief, DASR is specifically designed for image super-resolution, whereas AirNet is proposed to handle multi-degradation in all-in-one manner. Third, despite the difference in the task, DASR needs specifying the image super-resolution scale, whereas AirNet does not need any degradation parameters.

3. The Proposed Method

In this section, we elaborate on the proposed method which consists of Contrastive-Based Degradation Encoder (CBDE, $f_C(\cdot)$) and Degradation-Guided Restoration Network (DGRN, $f_D(\cdot)$) as shown in Figure 2.

For a given degraded image x, AirNet first feeds it into $f_C(\cdot)$ to learn the latent degradation representation $z = f_C(x)$. Then, x and z are further passed through $f_D(\cdot)$ to obtain the recovered image $y' = f_D(x, z)$. Without loss of generalizability, we consider three popular degradations as a showcase in this paper, namely, noise, haze, and rain. In the following, we will first introduce the overall loss function and then elaborate on the two subnetworks with the corresponding loss.

3.1. The Objective Function

To remove the corruption from the observed image, we propose the following objective function:

$$\mathcal{L} = \mathcal{L}_{Rec} + \mathcal{L}_{cl}, \tag{1}$$

where $\mathcal{L}_{Rec}$ is the reconstruction loss between the ground-truth y and the recovered clean image y'. The second loss $\mathcal{L}_{cl}$ is the contrastive loss for CBDE.

For a given degraded image x, $\mathcal{L}_{Rec}$ aims to minimize the L_1-distance between y and the recovered clean image $y' = f(x)$ through AirNet. Mathematically,

$$\mathcal{L}_{Rec} = \frac{1}{T} \sum_{i=1}^{T} |f(x_i) - y_i| \tag{2}$$

where T is the number of x's pixel, and i is the index of the pixel.

Different from $\mathcal{L}_{Rec}$, $\mathcal{L}_{cl}$ is the specific loss of CBDE, which aims to learn representations for different degradations while preserving their possible difference. More details will be presented in Section 3.2.

3.2. Contrastive-Based Degradation Encoder

The contrastive-based degradation encoder aims to extract the latent degradation representation z from the input x. To enable AirNet tackle multiple unspecific degradations, z is expected to enjoy the following two properties.

First, z should be adaptive to different degradations. In other words, for the inputs with different degradations, the corresponding z should be different even though the image contents are the same. To this end, we leverage contrastive learning to learn z by maximizing the consistency of two inputs with the same degradation (*i.e.*, positive samples), while minimizing the consistency between different degradations (*i.e.*, negative samples). To be specific, for a degradation representation q, k^+ and k_i^- are the corresponding positive and negative counterpart, respectively. Then, $\mathcal{L}_{cl}$ could be reformulated as,

$$\mathcal{L}_{cl} = -\log \frac{\exp(q \cdot k^+ / \tau)}{\sum_{i=0}^{K} exp(q \cdot k_i^- / \tau)} \tag{3}$$

where τ is a temperature hyper-parameter per [14, 48] and K denotes the number of negative samples.

In details, for a given input x, we randomly crop two patches from x, named x_q and x_{k^+}. As the degradations in the same image should be consistent, we treat x_q and x_{k^+} as positive pairs. On the contrary, the patches from other images are treated as negative x_{k^-} w.r.t. x_q. With the obtained pairs, we pass them through CBDE and get the corresponding intermediate representation v_q, v_{k^+} and v_{k^-} which are further fed into a two-layer MLP to get q, k^+ and k^-. To learn a degradation space wherein the discrimination of different degradations is preserved, Eq. 3 is used.

Thanks to our contrastive learning based solution, the learned degradation representation embraces the following advantage. To be exact, it does not rely a mathematical model that explicitly defines the relationship between the corrupted and clean images as existing methods [1, 37]. Therefore, it avoids the knowledge on such a prior and its performance is irrelevant with the exact definition. Especially, our method is more competitive when the relationship is always unknown or inexact due to mixed multiple degradations or the degradation comes from nature, *e.g*., rain and haze. On the other hand, our method unifies different degradations into the same subspace while preserving their difference. In contrast, the existing single/multi-degradation methods learn representations for different degradations from different subspaces, thus losing the comparability and relationship of degradations. For example, the Gaussian noise with corruption rate of 0.1 and 0.2 should be close in the latent space, comparing with the haze corruptions. Clearly, our contrastive degradation representation could own such a property which is crucial to handle the data with multiple degradations.

Second, z should preserve as much as possible space structure to favor image restoration. To this end, we adopt the output of the first instead of last layer of CBDE as z. In other words, z is a tensor instead of vector and thus could preserve the contextual information. In addition, as z is with the same dimension of the input and the outputs of intermediate layers, it is flexible to concat with other features and compatible to existing neural networks such as DCN [6] and SFT [44].

3.3. Degradation-Guided Restoration Network

With z learned by the CBDE, DGRN is used to restore the clean image from the input with unknown degradation. As shown in Fig. 2, DGRN builds up by five Degradation-Guided Groups (DGG) each of which further consists of five Degradation-Guided Blocks (DGB). Within each DGB, two Degradation-Guided Modules (DGM) are adopted to restore clean images under the guidance of z.

As elaborated above, DGM is the basic module of DGRN, which consists of a Deformable Convolution (DCN) layer and Spatial Feature Transform (SFT) layer. Mathematically,

$$\begin{aligned} F_{DGM}^{m,b,g} &= \Phi_{DGM}^{m,b,g}(F^{m-1,b,g}, z) \\ &= \Phi_{DCN}^{m,b,g}(F^{m-1,b,g}|z) + \Phi_{SFT}^{m,b,g}(F^{m-1,b,g}|z), \end{aligned} \tag{4}$$

where $\Phi_{DGM}^{m,b,g}$ is the m-th DGM w.r.t. the b-th DGB of the g-th DGG, and $F^{m-1,b,g}$ denotes the output of the $(m-1)$-th DGM w.r.t. b-th DGB of the g-th DGG. Φ_{DCN} and Φ_{SFT} are the DCN and SFT layer, respectively.

DGM is designed to achieve the following two goals. On the one hand, as different degradations should have different receptive fields, it is highly expected that the model could be adaptive to different degradations. To this end, DGM employs the deformable convolution (DCN) [56] which could dynamically adjust the receptive field based on the modulating offsets and masks. To be specific, given a deformable convolution kernel of K sampling locations, let w_k and $p_k \in \{(-1,-1), (-1,0), \cdots, (1,1)\}$ denote the weight and the pre-defined offsets for the k-th location, then the DCN layer used in DGM as defined by:

$$\Phi_{DCN}^{m,b,g}(F^{m-1,b,g}|z) = \sum_{k=1}^{K} w_k \cdot F^{m-1,b,g}(p+p_k+\Delta p_k) \cdot \Delta m_k, \tag{5}$$

where $F^{m-1,b,g}(p)$ denotes the features at location p from the feature maps $F^{m-1,b,g}$. Δp_k and Δm_k are the learnable offset and modulation scalar for location k, respectively. In our implementations, AirNet learns Δp_k and Δm_k using a convolution layer $conv(\cdot)$ whose input is the concatenation of $F^{m-1,b,g}$ and z. Namely,

$$(\Delta p_k, \Delta m_k) = conv(concat(F^{m-1,b,g}, z)), \tag{6}$$

where $concat(\cdot)$ is the concatenation operator.

On the other hand, as different degraded images have different latent distributions, the proposed model is expected to narrow the distribution gap for stronger multi-degradations restoration capability. To this end, DGM adopts SFT [44] as a component to adjust the distributions of F based on z, i.e.,

$$F_{SFT}^{m,b,g} = \Phi_{SFT}^{m,b,g}(F^{m-1,b,g}|z). \tag{7}$$

In details, the SFT layer aims to learn a mapping function $\mathcal{M}$ that outputs a modulation parameters (γ and β) for a given z. Then, SFT conducts affine transformation by scaling and shifting feature $F^{m-1,b,g}$ with (γ and β). Mathematically,

$$F_{SFT}^{m,b,g} = \Phi_{SFT}^{m,b,g}(F^{m-1,b,g}|\gamma, \beta) = \gamma \odot F^{m-1,b,g} + \beta, \tag{8}$$

where $\odot$ denotes the element-wise multiplication, and $(\gamma, \beta) = \mathcal{M}(z)$. In our experiments, we implement $\mathcal{M}$ using two convolution layers.

Table 1. Quantitative results of image denoising on the BSD68 and Urban100 datasets. The best results are shown in boldface.

Method	BSD68			Urban100		
	$\sigma = 15$	$\sigma = 25$	$\sigma = 50$	$\sigma = 15$	$\sigma = 25$	$\sigma = 50$
CBM3D [5]	33.50/0.9215	30.69/0.8672	27.36/0.7626	33.93/0.9408	31.36/0.9092	27.93/0.8404
DnCNN [53]	33.89/0.9290	31.23/0.8830	27.92/0.7896	32.98/0.9314	30.81/0.9015	27.59/0.8331
IRCNN [54]	33.87/0.9285	31.18/0.8824	27.88/0.7898	27.59/0.8331	31.20/0.9088	27.70/0.8396
FFDNet [55]	33.87/0.9290	31.21/0.8821	27.96/0.7887	33.83/0.9418	31.40/0.9120	28.05/0.8476
BRDNet [40]	34.10/0.9291	31.43/0.8847	28.16/0.7942	**34.42**/0.9462	31.99/0.9194	28.56/0.8577
AirNet	**34.14/0.9356**	**31.48/0.8928**	**28.23/0.8057**	34.40/**0.9487**	**32.10/0.9240**	**28.88/0.8702**

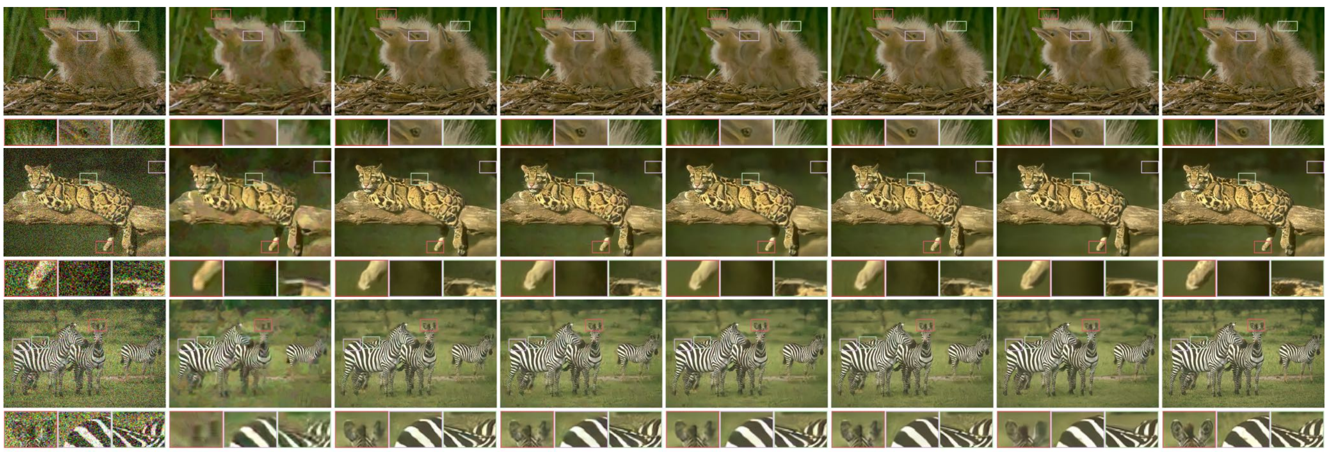

Figure 3. Comparisons with the SOTA denoising methods on the BSD68 database. Some areas are highlighted in colored rectangles and zooming-in is recommended for a better visualization and comparisons.

4. Experiments

In this section, we evaluate the proposed method on four widely-used datasets by comparing with 17 baselines. In the following, we will first introduce the experimental setting and then show the qualitative and quantitative results on benchmarks. Finally, we will conduct some ablation studies to verify the effectiveness of our method.

4.1. Experimental Settings

In this section, we introduce the details of the used datasets, baselines, evaluation metrics, and implementations details.

Datasets: In our experiments, we use the following six datasets for evaluations, *i.e.*, BSD400, BSD68 [31], WED [30], and Urban100 for denoising; Rain100L [49] for deraining; and RESIDE [22] for dehazing. To be specific, BSD400 consists of 400 clean natural images and BSD68 includes 68 natural images. WED contains 4,744 natural images collected from Internet, and Urban100 has 100 clean images. For image denoising, we use the combination of BSD400 and WED as training set, and that of BSD68 and Urban100 as testing sets like [55]. By following [40,53–55], the noisy images are generated by manually adding white Gaussian noises to the clean images with three corruption levels, *i.e.*, $\sigma = 15, 25, 50$. For image deraining, we conduct experiments on Rain100L which consists of 200 rainy-clean training pairs and 100 testing image pairs. For image dehazing, we conduct experiments on the RESIDE dataset [22] consisting of Outdoor Training Set (OTS) and Synthetic Objective Testing Set (SOTS) which are used for training and testing, respectively. In brief, OTS consists of 72,135 outdoor hazy-clean image pairs and SOTS contains 500 outdoor hazy-clean image pairs.

Baselines: For comprehensive comparisons, we compare our method with five denoising methods, five deraining methods, five dehazing methods, one image restoration method and one IRMD method. To be specific, the denoising baselines contain CBM3D [5], DnCNN [53], IRCNN [54], FFDNet [55] and BRDNet [40]. the deraining baselines are DIDMDN [52], UMRL [50], SIRR [46], MSPFN [17], and LPNet [11]. The dehazing baselines are DehazeNet [1], MSCNN [37], AOD-Net [21], EPDN [35] and FDGAN [7]. The image restoration baseline is MPRNet [51]. The IRMD baseline is the Decouple Learning (DL) [8]. To comprehensive demonstrate the effectiveness

Table 2. Quantitative results of image deraining on the Rain100L dataset. The best results are shown in boldface.

Metrics	DIDMDN [52]	UMRL [50]	SIRR [46]	MSPFN [17]	LPNet [11]	AirNet
PSNR	23.79	32.39	32.37	33.50	33.61	**34.90**
SSIM	0.7731	0.9210	0.9258	0.9480	0.9583	**0.9660**

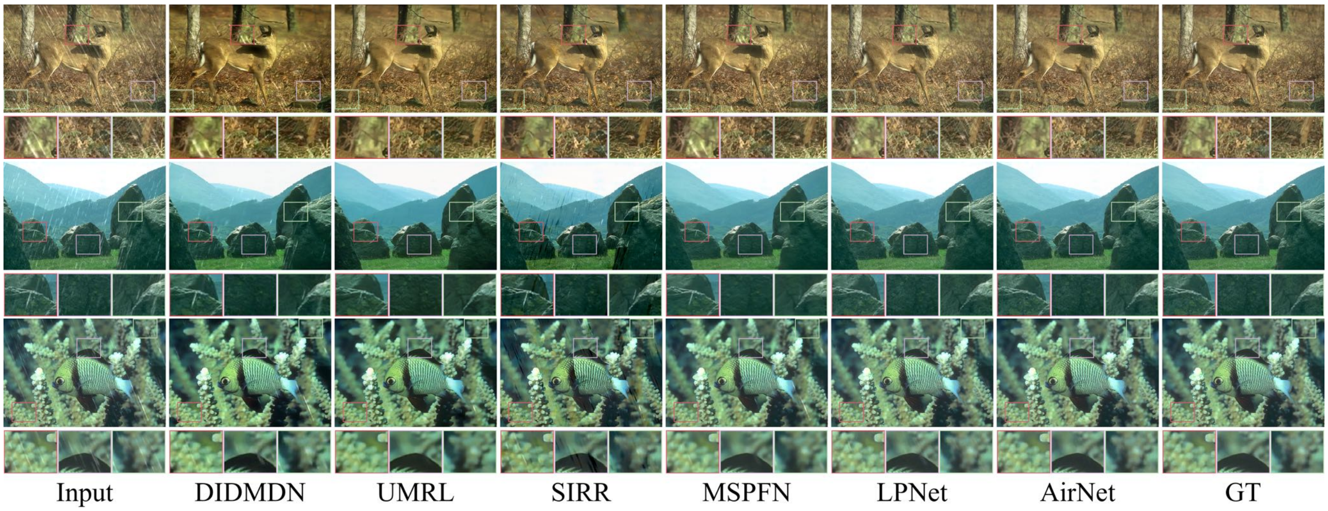

Figure 4. Comparisons of the SOTA derain methods on the Rain100L database. Some areas are highlighted in colored rectangles and zooming-in is recommended for a better visualization and comparisons.

Table 3. Quantitative results of image dehazing on the SOTS dataset. The best results are shown in boldface.

Metrics	DehazeNet [1]	MSCNN [37]	AODNet [21]	EPDN [35]	FDGAN [7]	AirNet
PSNR	22.46	22.06	20.29	22.57	23.15	**23.18**
SSIM	0.8514	0.9078	0.8765	0.8630	**0.9207**	0.9000

of our method, two different settings are examined, *i.e.*, train AirNet on a specified degradation one-by-one (OBO) and train AirNet on all degradations in an all-in-one fashion (AIO). In other words, AirNet under AIO is the model trained on the collection of all the datasets that consists of three corruptions (*i.e.*, noise, rain, and haze) with different degradation levels (*i.e.*, $\sigma = 15, 25, 50$).

Evaluation metrics: Following [7, 11, 40], two popular metrics are used for quantitative comparisons, *i.e.*, Peak Signal-to-Noise Ratio (PSNR) [16] and Structure Similarity (SSIM) [45]. Higher value of these metrics indicates better performance of the methods.

Training details: We conduct experiments in PyTorch on NVIDIA GeForce RTX 2080Ti GPUs. To optimize AirNet, we employ the ADAM optimizer [18] with the default $\{\beta_i\}_{i=1}^{2}$. We set 400 iterations as one epoch and train the model with 1,500 epochs. To warm up, we first train CBDE by the optimizing $\mathcal{L}_{cl}$ for 100 iterations. Then, we train the whole network with $\mathcal{L}$ for 1,400 iterations. The learning rate is initialized to 0.001 and then decreased to 0.0001 after 60 epochs. After that, the learning rate is decreased to half after each 125 epochs. In the experiments, we train our model with the batch size of N and the patch size of 128, where N is number of degradation types.

4.2. Comparisons on Single Degradation

In this section, we show the quantitative and qualitative results on three separated image restoration tasks, *i.e.*, denoising, deraining, and dehazing.

Denoising: Table 1 reports the results on BSD68 and Urban100 comparing with five denoising methods under the one-by-one setting. From the results, one could find that AirNet achieves the best result in almost all tests. Besides the dominance in quantitative evaluations, AirNet also shows superiority in qualitative comparisons as shown in Figure 3. Due to space limitations, we leave more results in supplementary materials.

Deraining: From Table 2 and Figure 4, one could observe that AirNet also remarkably outperforms all deraining baselines. For example, AirNet is 1.4 and 0.0074 higher than the best method under the OBO setting in PSNR and SSIM, respectively.

Dehazing As shown in Table 3 and Figure 5, AirNet is slightly better than the best baseline in PSNR. More specif-

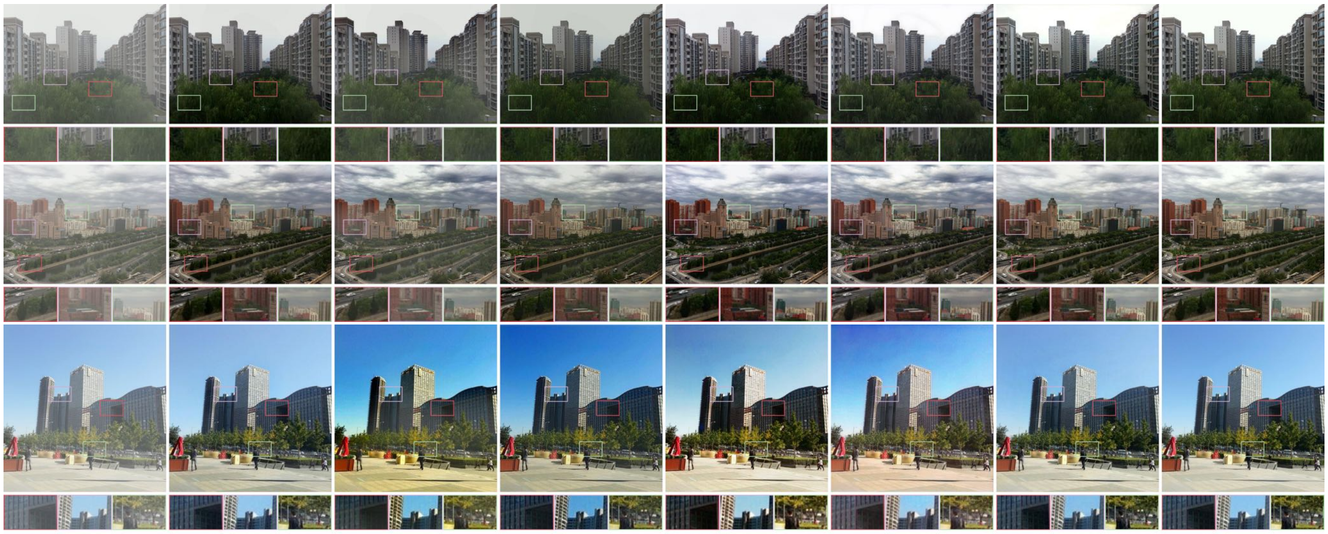

Figure 5. Comparisons of the SOTA dehaze methods on the SOTS database. Some areas are highlighted in colored rectangles and zooming-in is recommended for a better visualization and comparisons.

Table 4. Performance comparisons on three challenging datasets. The best results are shown in boldface.

Type	Method	Denoise			Derain	Dehaze	Average
		BSD68 ($\sigma = 15$)	BSD68 ($\sigma = 25$)	BSD68 ($\sigma = 50$)	Rain100L	SOTS	
One-By-One	BRDNet [40]	34.10/0.9291	31.43/0.8847	28.16/0.7942	33.15/0.9490	23.31/0.9116	30.30/0.8937
	LPNet [11]	32.31/0.9236	27.87/0.8674	25.71/0.7656	33.61/0.9583	21.43/0.8631	28.19/0.8756
	FDGAN [7]	31.11/0.9147	29.57/0.8770	27.12/0.7895	31.14/0.9422	23.15/0.9207	28.42/0.8888
	MPRNet [51]	34.01/0.9334	31.34/0.8892	28.10/0.8014	**38.26/0.9816**	**28.21/0.9672**	**31.98/0.9146**
	DL [8]	33.25/0.9225	30.38/0.8679	26.68/0.7415	33.94/0.9456	24.68/0.9243	29.79/0.8804
	AirNet	**34.14/0.9356**	**31.48/0.8928**	**28.23/0.8057**	34.90/0.9660	23.18/0.9000	30.38/0.9000
All-In-One	BRDNet [40]	32.26/0.8977	29.76/0.8355	26.34/0.6934	27.42/0.8952	23.23/0.8952	27.80/0.8434
	LPNet [11]	26.47/0.7780	24.77/0.7477	21.26/0.5522	24.88/0.7837	20.84/0,8277	23.64/0.7379
	FDGAN [7]	30.25/0.9103	28.81/0.8682	26.43/0.7757	29.89/0.9329	24.71/0.9294	28.02/0.8833
	MPRNet [51]	33.54/0.9274	30.89/0.8797	27.56/0.7792	33.57/0.9542	25.28/0.9545	30.17/0.8990
	DL [8]	33.05/0.9140	30.41/0.8606	26.90/0.7401	32.62/0.9314	26.92/0.9314	29.98/0.8755
	AirNet	**33.92/0.9329**	**31.26/0.8884**	**28.00/0.7974**	**34.90/0.9675**	**27.94/0.9615**	**31.20/0.9095**

ically, AirNet is 0.03 higher than FDGAN in PSNR. However, the visual results show that AirNet could recover more details that are human favorable.

4.3. Comparisons on Multiple Degradations

The most attractive point of AirNet is the capacity of handling different unknown degradations in an all-in-one framework. In this section, we conduct experiments to verify the effectiveness of AirNet under such a settings. To this end, we choose five IRSD methods (*i.e.*, BRDNet [40], LPNet [11], FDGAN [7] and MPRNet [51]) and one IRMD method (*i.e.*, DL [8]) as baselines. For fair and extensive comparisons, We re-train these methods with the two aforementioned settings, *i.e.*, One-By-One and All-In-One. As shown in Table 4, one could observe that AirNet is superior to all the baselines in most cases. It should be pointed out that although DL could also handle multiple degradations, it needs to know the corruption types and levels so that the correct head and tail of the network could be specified.

4.4. Results on Combined Degradations

In this section, we train AirNet with different combinations of multiple degradations to analyze how the performance influenced by the corrupted dataset. As shown in Table 5, more degradations will lead to more difficulties in denoising, whereas the same conclusion cannot be derived from the deraining and dehazing tasks. Interesting, the deraining will be helpful to denoising, and the dehazing is benefited from the combination of all degradations. More empirical studies and theoretical analysis are expected in the future.

4.5. Results on Spatially Variant Degradation

In this section, we carry out experiment to demonstrate the effectiveness of AirNet on spatially variant degradation, *i.e.*, different areas of the same image are with different cor-

Table 5. Ablation study on the combinations of degradations. In the table, "✓" denotes the AirNet with the degradation, "−" indicates unavailable results, and the best results are shown in boldface.

Degradation			Denoise			Derain	Dehaze
Noise	Rain	Haze	BSD68 ($\sigma = 15$)	BSD68 ($\sigma = 25$)	BSD68 ($\sigma = 50$)	Rain100L	SOTS
✓			**34.14/0.9355**	**31.49/0.8928**	**28.23/0.8058**	-	-
	✓		-	-	-	34.90/0.9657	-
		✓	-	-	-	-	23.18/0.9000
✓	✓		34.11/0.9352	31.46/0.8923	28.19/0.8042	**38.31/0.9824**	-
✓		✓	33.97/0.9337	31.32/0.8891	28.06/0.7992	-	23.72/0.9375
	✓	✓	-	-	-	32.50/0.9465	26.78/0.9577
✓	✓	✓	33.92/0.9330	31.26/0.8881	28.01/0.7976	34.90/0.9675	**27.94/0.9615**

Table 6. Quantitative results on the spatially variant degradation. The best results are shown in boldface.

Method	CBM3D [5]	DnCNN [53]	IRCNN [54]	FFDNet [55]	BRDNet [40]	DL [8]	Ours
PSNR	25.09	23.83	22.78	22.71	27.26	26.10	**31.42**
SSIM	0.6457	0.5037	0.3883	0.3790	0.7410	0.7528	**0.8922**

Table 7. Ablation study on BSD68 and Urban100. The best results are shown in boldface.

Dataset	BSD68			Urban100		
Noise Level	$\sigma = 15$	$\sigma = 25$	$\sigma = 50$	$\sigma = 15$	$\sigma = 25$	$\sigma = 50$
w.o. SFT	34.12/0.9354	31.47/0.8924	28.22/0.8051	34.37/0.9484	32.07/0.9234	28.84/0.8691
w.o. DCN	34.03/0.9342	31.36/0.8898	28.08/0.7985	34.21/0.9467	31.83/0.9197	28.49/0.8597
Ours	**34.14/0.9355**	**31.49/0.8928**	**28.23/0.8058**	**34.40/0.9487**	**32.10/0.9241**	**28.88/0.8699**

ruption levels. To this end, we synthesize a degraded version of BSD68 which is with spatially variant noises. In detail, we divide each clean image into four regions wherein the gaussian noises with $\sigma \in \{0, 15, 25, 50\}$ are respectively added. Table 6 shows that AirNet is also effective in recovering the latent clean image under such an evaluation protocol.

4.6. Ablation Study

To demonstrate the effectiveness of our network structure, we conduct an ablation study on the BSD68 by removing one of the DCN layer and the SFT layer. From Table 7, one could see that both the DCN layer and SFT layer are important to improve the performance of AirNet.

5. Conclusion

In this paper, we proposed an all-in-one image restoration network (AirNet) which is free from the prior of corruption type and level. Meanwhile, the method is an all-in-one solution to restore images from different corruptions, which is competitive to a variety of practical scenarios wherein the prior is hard to foreknow or the degradation might change with time and space. Extensive experimental results show that the superiority of AirNet in both qualitative and quantitative comparisons.

6. Shortcomings and Broader Impact

Although AirNet experimentally shows superiority in three image restoration tasks and their combinations, it is unclear how its performance with other corruptions such as blurring and snowing. In addition, it is also worthy to further explore why different combined degradations lead to different results w.r.t. the single task as illustrated in Section 4.5. In a broader vision, although AirNet could be adaptive to different corruptions and avoid multiple models of the same algorithm on different degradations, it still needs a large amount of resources to optimize the method, thus resulting in carbon emission and indirectly climate warming.

Acknowledgments

This work was supported in part by the National Key R&D Program of China under Grant 2020AAA0104500; in part by NSFC under Grant U21B2040, U19A2081, 61836006, and U19A2078; in part by the Fund of Sichuan University Tomorrow Advancing Life.

References

[1] Bolun Cai, Xiangmin Xu, Kui Jia, Chunmei Qing, and Dacheng Tao. DehazeNet: An End-to-End System for Single Image Haze Removal. *IEEE Transactions on Image Processing*, 25(11):5187–5198, Jan. 2016. 1, 2, 4, 5, 6

[2] Xiaochun Cao, Wenqi Ren, Wangmeng Zuo, Xiaojie Guo, and Hassan Foroosh. Scene text deblurring using text-specific multiscale dictionaries. *IEEE Transactions on Image Processing*, 24(4):1302–1314, 2015. 2

[3] Hanting Chen, Yunhe Wang, Tianyu Guo, Chang Xu, Yiping Deng, Zhenhua Liu, Siwei Ma, Chunjing Xu, Chao Xu, and Wen Gao. Pre-Trained Image Processing Transformer. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 12299–12310, Online, June 2021. 2

[4] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A Simple Framework for Contrastive Learning of Visual Representations. In *International Conference on Machine Learning*, pages 1597–1607, Online, July 2020. 2

[5] Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian. Color Image Denoising via Sparse 3D Collaborative Filtering with Grouping Constraint in Luminance-Chrominance Space. In *International Conference on Image Processing*, pages 313–316, San Antonio, TX, Sept. 2007. 1, 5, 8

[6] Jifeng Dai, Haozhi Qi, Yuwen Xiong, Yi Li, Guodong Zhang, Han Hu, and Yichen Wei. Deformable Convolutional Networks. In *International Conference on Computer Vision*, pages 764–773, Venice, Italy, Oct. 2017. 4

[7] Yu Dong, Yihao Liu, He Zhang, Shifeng Chen, and Yu Qiao. FD-GAN: Generative Adversarial Networks with Fusion-discriminator for Single Image Dehazing . In *AAAI Conference on Artificial Intelligence*, pages 10729–10736, New York, NY, Jan. 2020. 1, 5, 6, 7

[8] Qingnan Fan, Dongdong Chen, Lu Yuan, Gang Hua, Nenghai Yu, and Baoquan Chen. A General Decoupled Learning Framework for Parameterized Image Operators. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(1):33–47, Jan. 2021. 2, 5, 7, 8

[9] Xin Fan, Yi Wang, Xianxuan Tang, Renjie Gao, and Zhongxuan Luo. Two-layer gaussian process regression with example selection for image dehazing. *IEEE Transactions on Circuits and Systems for Video Technology*, 27(12):2505–2517, 2016. 1

[10] 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 Conference on Computer Vision and Pattern Recognition*, pages 1715–1723, Honolulu, HI, June 2017. IEEE. 1, 2

[11] Xueyang Fu, Borong Liang, Yue Huang, Xinghao Ding, and John Paisley. Lightweight Pyramid Networks for Image Deraining. *IEEE Transactions on Neural Networks and Learning Systems*, 31(6):1794–1807, June 2020. 1, 5, 6, 7

[12] Hongyun Gao, Xin Tao, Xiaoyong Shen, and Jiaya Jia. Dynamic Scene Deblurring with Parameter Selective Sharing and Nested Skip Connections. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 3848–3856, Long Beach, CA, June 2019. 1, 2

[13] Yuanbiao Gou, Boyun Li, Zitao Liu, Songfan Yang, and Xi Peng. CLEARER: Multi-Scale Neural Architecture Search for Image Restoration. In *Neural Information Processing Systems*, pages 17129–17140, Online, Dec. 2020. 2

[14] Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick. Momentum Contrast for Unsupervised Visual Representation Learning. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 9726–9735, Online, June 2020. 2, 3

[15] Kaiming He, Jian Sun, and Xiaoou Tang. Single Image Haze Removal Using Dark Channel Prior. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1956–1963, Miami, FL, June 2009. 1, 2

[16] Quan Huynh-Thu and Mohammed Ghanbari. Scope of validity of PSNR in image/video quality assessment. *Electronics letters*, 44(13):800–801, 2008. 6

[17] 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 Conference on Computer Vision and Pattern Recognition*, pages 8346–8355, Online, June 2020. 1, 2, 5, 6

[18] Diederik P Kingma and Jimmy Lei Ba. Adam: A Method for Stochastic Optimization. In *International Conference on Learning Representations*, San Diego, CA, May 2015. 6

[19] Boyun Li, Yuanbiao Gou, Shuhang Gu, Jerry Zitao Liu, Joey Tianyi Zhou, and Xi Peng. You Only Look Yourself: Unsupervised and Untrained Single Image Dehazing Neural Network. *International Journal of Computer Vision*, 129(5):1754–1767, May 2021. 1

[20] Boyun Li, Yuanbiao Gou, Jerry Zitao Liu, Hongyuan Zhu, Joey Tianyi Zhou, and Xi Peng. Zero-Shot Image Dehazing. *IEEE Transactions on Image Processing*, 29:8457–8466, Aug. 2020. 2

[21] Boyi Li, Xiulian Peng, Zhangyang Wang, Jizheng Xu, and Dan Feng. AOD-Net: All-in-One Dehazing Network. In *International Conference on Computer Vision*, pages 4780–4788, Venice, Italy, Oct. 2017. 5, 6

[22] Boyi Li, Wenqi Ren, Dengpan Fu, Dacheng Tao, Dan Feng, Wenjun Zeng, and Zhangyang Wang. Benchmarking Single Image Dehazing and Beyond. *IEEE Transactions on Image Processing*, 28(1):492–505, Jan. 2019. 5

[23] Ruoteng Li, Robby T Tan, and Loong-Fah Cheong. All in One Bad Weather Removal using Architectural Search. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 3175–3185, Online, June 2020. 2

[24] Xia Li, Jianlong Wu, Zhouchen Lin, Hong Liu, and Hongbin Zha. Recurrent Squeeze-and-Excitation Context Aggregation Net for Single Image Deraining . In *European Conference on Computer Vision*, pages 262–277, Munich, Germany, Sept. 2018. 2

[25] Yan Li, De Cheng, Jiande Sun, Dingwen Zhang, Nannan Wang, and Xinbo Gao. Single Image Dehazing with An Independent Detail-Recovery Network. *arXiv.org*, Sept. 2021. 2

[26] Ding Liu, Bihan Wen, Yuchen Fan, Chen Change Loy, and Thomas S Huang. Non-Local Recurrent Network for Image Restoration. In *Neural Information Processing Systems*, pages 1680–1689, Montréal, Canada, Dec. 2018. 2

[27] Jianzhao Liu, Jianxin Lin, Xin Li, Wei Zhou, Sen Liu, and Zhibo Chen. LIRA: Lifelong Image Restoration from Unknown Blended Distortions. In *European Conference on Computer Vision*, pages 616–632, Online, Aug. 2020. 1

[28] Risheng Liu, Xin Fan, Minjun Hou, Zhiying Jiang, Zhongxuan Luo, and Lei Zhang. Learning Aggregated Transmission Propagation Networks for Haze Removal and Beyond. *IEEE Transactions on Neural Networks and Learning Systems*, 30(10):2973–2986, Oct. 2019. 2

[29] Risheng Liu, Yi He, Shichao Cheng, Xin Fan, and Zhongxuan Luo. Learning collaborative generation correction modules for blind image deblurring and beyond. In *Proceedings of the 26th ACM international conference on Multimedia*, pages 1921–1929, 2018. 2

[30] Kede Ma, Zhengfang Duanmu, Qingbo Wu, Zhou Wang, Hongwei Yong, Hongliang Li, and Lei Zhang. Waterloo Exploration Database: New Challenges for Image Quality Assessment Models. *IEEE Transactions on Image Processing*, 26(2):1004–1016, Feb. 2017. 5

[31] David Martin, Charless Fowlkes, Doron Tal, and Jitendra Malik. A Database of Human Segmented Natural Images and its Application to Evaluating Segmentation Algorithms and Measuring Ecological Statistics. In *International Conference on Computer Vision*, pages 416–425, Vancouver, Canada, July 2001. 5

[32] Seungjun Nah, Tae Hyun Kim, and Kyoung Mu Lee. Deep Multi-scale Convolutional Neural Network for Dynamic Scene Deblurring. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 257–265, Honolulu, HI, June 2017. 1, 2

[33] Jinshan Pan, Zhe Hu, Zhixun Su, and Ming-Hsuan Yang. Deblurring Text Images via L0-Regularized Intensity and Gradient Prior. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 2901–2908, Columbus, OH, June 2014. IEEE. 1, 2

[34] Jinshan Pan, Deqing Sun, Hanspeter Pfister, and Ming-Hsuan Yang. Blind Image Deblurring Using Dark Channel Prior. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1628–1636, Las Vegas, NV, June 2016. 1, 2

[35] Yanyun Qu, Yizi Chen, Jingying Huang, and Yuan Xie. Enhanced Pix2pix Dehazing Network. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 8160–8168, Long Beach, CA, June 2019. 1, 2, 5, 6

[36] Wenqi Ren, Xiaochun Cao, Jinshan Pan, Xiaojie Guo, Wangmeng Zuo, and Ming-Hsuan Yang. Image deblurring via enhanced low-rank prior. *IEEE Transactions on Image Processing*, 25(7):3426–3437, 2016. 2

[37] Wenqi Ren, Si Liu, Hua Zhang, Jinshan Pan, Xiaochun Cao, and Ming-Hsuan Yang. Single Image Dehazing via Multi-Scale Convolutional Neural Networks. In *European Conference on Computer Vision*, pages 154–169, Amsterdam, The Netherlands, Oct. 2016. 1, 2, 4, 5, 6

[38] Wenqi Ren, Lin Ma, Jiawei Pan, Xiaochun Cao, Wei Liu, and Ming-Hsuan Yang. Gated Fusion Network for Single Image Dehazing. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 3253–3261, Salt Lake City, UT, June 2018. 1, 2

[39] Ying Tai, Jian Yang, Xiaoming Liu, and Chunyan Xu. MemNet: A Persistent Memory Network for Image Restoration. In *International Conference on Computer Vision*, pages 4549–4557, Venice, Italy, Oct. 2017. 2

[40] Chunwei Tian, Yong Xu, and Wangmeng Zuo. Image denoising using deep CNN with batch renormalization. *Neural Networks*, 121:461–473, Jan. 2020. 1, 5, 7, 8

[41] Aaron van den Oord, Yazhe Li, and Oriol Vinyals. Representation Learning with Contrastive Predictive Coding. *arXiv.org*, July 2018. 2

[42] Hong Wang, Qi Xie, Qian Zhao, and Deyu Meng. A Model-driven Deep Neural Network for Single Image Rain Removal. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 3100–3109, Online, June 2020. 2

[43] Longguang Wang, Yingqian Wang, Xiaoyu Dong, Qingyu Xu, Jungang Yang, Wei An, and Yulan Guo. Unsupervised Degradation Representation Learning for Blind Super-Resolution. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 10581–10590, Online, June 2021. 2

[44] Xintao Wang, Ke Yu, Chao Dong, and Chen Change Loy. Recovering Realistic Texture in Image Super-resolution by Deep Spatial Feature Transform. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 606–615, Salt Lake City, UT, June 2018. 4

[45] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004. 6

[46] Wei Wei, Deyu Meng, Qian Zhao, Zongben Xu, and Ying Wu. Semi-supervised Transfer Learning for Image Rain Removal. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 3877–3886, Long Beach, CA, June 2019. 1, 2, 5, 6

[47] Haiyan Wu, Yanyun Qu, Shaohui Lin, Jian Zhou, Ruizhi Qiao, Zhizhong Zhang, Yuan Xie, and Lizhuang Ma. Contrastive Learning for Compact Single Image Dehazing. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 10551–10560, Online, June 2021. 2

[48] Zhirong Wu, Yuanjun Xiong, Stella X Yu, and Dahua Lin. Unsupervised Feature Learning via Non-Parametric Instance Discrimination. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 3733–3742, Salt Lake City, UT, June 2018. 3

[49] 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 Transactions on Pattern Analysis and Machine Intelligence*, 42(6):1377–1393, June 2020. 2, 5

[50] Rajeev Yasarla and Vishal M Patel. Uncertainty Guided Multi-Scale Residual Learning-using a Cycle Spinning CNN for Single Image De-Raining. In *IEEE Conference on Com-*

puter Vision and Pattern Recognition, pages 8405–8414, Long Beach, CA, June 2019. 1, 2, 5, 6

[51] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Multi-Stage Progressive Image Restoration. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 14821–14831, Online, June 2021. 2, 5, 7

[52] He Zhang and Vishal M Patel. Density-aware Single Image De-raining using a Multi-stream Dense Network. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 695–704, Salt Lake City, UT, June 2018. 1, 2, 5, 6

[53] Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang. Beyond a Gaussian Denoiser: Residual Learning of Deep CNN for Image Denoising. *IEEE Transactions on Image Processing*, 26(7):3142–3155, July 2017. 1, 2, 5, 8

[54] Kai Zhang, Wangmeng Zuo, Shuhang Gu, and Lei Zhang. Learning Deep CNN Denoiser Prior for Image Restoration. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 2808–2817, Honolulu, HI, July 2017. 1, 5, 8

[55] Kai Zhang, Wangmeng Zuo, and Lei Zhang. FFDNet: Toward a Fast and Flexible Solution for CNN based Image Denoising. *IEEE Transactions on Image Processing*, 27(9):4608–4622, Sept. 2018. 1, 5, 8

[56] Xizhou Zhu, Han Hu, Stephen Lin, and Jifeng Dai. Deformable ConvNets v2: More Deformable, Better Results. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 9308–9316, Long Beach, CA, June 2019. 4