

Physics-Based Generative Adversarial Models for Image Restoration and Beyond

Jinshan Pan, Jiangxin Dong, Yang Liu, Jiawei Zhang,
Jimmy Ren, Jinhui Tang, Yu-Wing Tai, and Ming-Hsuan Yang

Abstract—We present an algorithm to directly solve numerous image restoration problems (e.g., image deblurring, image dehazing, and image deraining). These problems are ill-posed, and the common assumptions for existing methods are usually based on heuristic image priors. In this paper, we show that these problems can be solved by generative models with adversarial learning. However, a straightforward formulation based on a straightforward generative adversarial network (GAN) does not perform well in these tasks, and some structures of the estimated images are usually not preserved well. Motivated by an interesting observation that the estimated results should be consistent with the observed inputs under the physics models, we propose an algorithm that guides the estimation process of a specific task within the GAN framework. The proposed model is trained in an end-to-end fashion and can be applied to a variety of image restoration and low-level vision problems. Extensive experiments demonstrate that the proposed method performs favorably against state-of-the-art algorithms.

Index Terms—Generative adversarial network, physics model, low-level vision, image restoration.

1 INTRODUCTION

Numerous image restoration and low-level vision problems (e.g., image deblurring, image super-resolution, image dehazing, and image deraining) aim to estimate a clear image x from a given input y . The fundamental assumption is that the estimated x should be consistent with the input y under the image formation model¹:

$$y = \mathcal{H}(x), \quad (1)$$

where the operator \mathcal{H} maps the unknown result x to the observed image y . For example, \mathcal{H} corresponds to the blur operation if (1) describes an image deblurring problem. As estimating x from y is ill-posed, it is necessary to introduce additional constraints to regularize x . One widely-used approach is based on the maximum a posterior (MAP) framework, where x can be solved by

$$x^* = \arg \max_x p(x|y) = \arg \max_x p(y|x)p(x). \quad (2)$$

In the above equation, $p(y|x)$ and $p(x)$ are probability density functions, which are usually referred to as

- J. Pan and J. Tang are with School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing, 210094, China. E-mail: sduuran@gmail.com, jinhuitang@njust.edu.cn.
- J. Dong and Y. Liu are with School of Mathematical and Sciences, Dalian University of Technology, Dalian, 116024, China. E-mail: {lewisyangliu, dongjxjx}@gmail.com.
- J. Zhang and J. Ren are with SenseTime Research, Shenzhen, 518000, China. E-mail: zhjw1988@gmail.com, rensijie@sensetime.com.
- Y.-W. Tai is with YouTu research lab, Shenzhen, 518000, China. E-mail: yuwwing@gmail.com.
- M.-H. Yang is with School of Engineering, University of California, Merced, CA, 95344. E-mail: mhyang@ucmerced.edu.

1. The image formation process is modeled by (1). In this paper, we refer to (1) as the image degradation model or physics model.

the likelihood term and image prior. In recent years, several deep models have been developed to deal with image restoration and related low-level vision tasks, e.g., image super-resolution [3], [4], [5], [6], [7], [8], image filtering [9], [10], noise removal [11], [12], [13], image deraining [14], and dehazing [15], [16], [17], to name a few. Mathematically, these methods directly learn the mapping functions between x and y with

$$x^* = \mathcal{G}(y), \quad (3)$$

where \mathcal{G} is the mapping function. The function \mathcal{G} can be regarded as an inverse operator of \mathcal{H} in (1). Theoretically, $\mathcal{G}(y)$ should be close to the ground truth if the mapping function can be estimated well. However, due to the complexity of the problem, e.g., the solution space of the corresponding problem is large, a simple model with a random initialization is not sufficient to estimate the mapping function well. Thus, only using a feed-forward network to learn the inverse operator \mathcal{G} does not generate good results as we will demonstrate below.

Recently, generative adversarial networks (GANs) [18] have been applied to image restoration problems, e.g., image super-resolution [19], image deraining [1], and image deblurring [20]. The GAN framework contains a generative model and a discriminative model, where the discriminative model is used to regularize the generative model, such that the distribution of outputs is close to that of realistic images. However, the adversarial loss does not ensure that the contents of outputs are consistent with those of the inputs. Although several algorithms [1], [19] use a pixel-wise loss function based on the ground truths, and a perceptual loss function [21] based on pre-

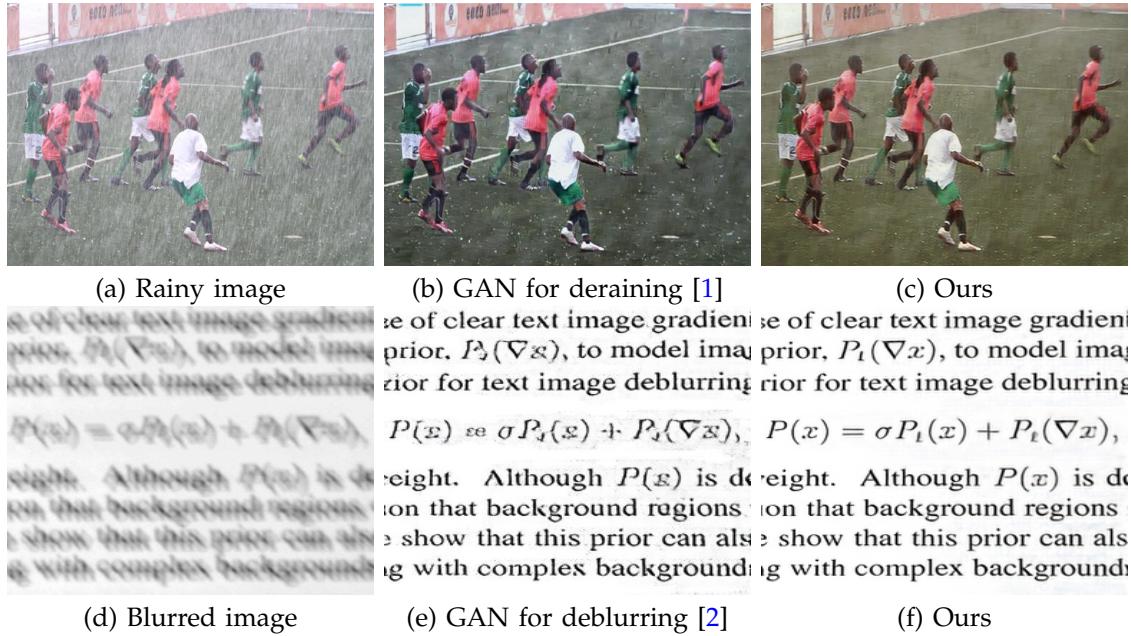


Fig. 1. Two image restoration problems. Our method is motivated by a key observation that the restored results should be consistent with the observed inputs under the degradation process (i.e., physics model). Enforcing this fundamental constraint in the generative adversarial network (GAN) can generate better results than those without using the constraint (e.g., (b)). We show that the proposed method can be applied to several image restoration and low-level vision problems and performs favorably against state-of-the-art algorithms.

trained VGG features as the constraints in the GAN formulation, these algorithms still do not perform well in image restoration as shown in Figure 1(b).

We note that the aforementioned methods only aim to estimate the mapping function (3) but do not guarantee whether the solutions satisfy the physics model (1) or not. Without this constraint, the methods based on the feed-forward networks do not restore correct results, e.g., main structures and details of the generated images are incorrect as shown in Figure 1(b) and (e). Thus, it is important to develop an algorithm that can model both the mapping function (i.e., (3)) and physics model (i.e., (1)) in a unified framework to address image restoration and related problems.

In this paper, we propose a GAN model constrained by a physics model for image restoration and low-level vision tasks. The physics model ensures that the estimated result (i.e., $\mathcal{G}(y)$) should be consistent with the observed image y . The proposed physics constrained GAN model is jointly trained in an end-to-end fashion. We show that the proposed algorithm can be effectively applied to a variety of image restoration and low-level vision problems (see Figure 1).

2 RELATED WORK

In this section, we review the methods closely to this work in proper context.

Generative adversarial networks. Goodfellow et al. [18] propose the GAN framework to generate realistic images from random noise. Motivated by this framework, numerous methods [2], [22], [23], [24], [25] have been proposed for vision tasks. Recently, the

GAN framework has also been applied to low-level vision problems [1], [19], [20], [26], [27]. In this work, we propose an efficient algorithm constrained by the physics model to improve the GAN framework for image deblurring, image dehazing, and related tasks.

Image deblurring. Numerous methods based on statistical priors have been developed to address this ill-posed problem. In contrast to conventional methods, Schuler et al. [28] develop a multi-layer perceptron approach to remove noise and artifacts in the deblurring process. Xu et al. [29] develop a convolutional neural network based on the singular value decomposition to deal with outliers. As these methods are designed for non-blind image deblurring, it is not clear how these approaches can be extended to blind image deblurring. For blind image deblurring, some approaches [30], [31] first use convolutional neural networks to estimate blur kernels and then deblur images with the conventional image restoration methods. To directly restore clear images, several end-to-end trainable neural networks [32], [33] have been developed. Although these methods alleviate the complex blur kernel estimation step, the generated results usually do not satisfy the image model and the structures of the recovered images are not preserved well as shown in Figure 1(e).

Image dehazing. The success of the conventional dehazing methods is due to effective design of hand-crafted features for estimating transmission maps, e.g., dark channel [34]. Recent deep learning-based methods [15], [16] first use neural networks to estimate transmission maps and then restore clear images

based on the traditional schemes [34]. Different from these methods, we propose an end-to-end trainable network to solve the image dehazing problem, which can directly restore clear images from hazy inputs.

Image deraining. For the image deraining, conventional algorithms are usually developed based on the statistical properties of rainy streaks [35], [36], [37], [38]. Recently, Eigen et al. [14] develop a neural network to remove rain/dirt in images. Motivated by the success of the ResNet [39], Fu et al. [40] develop a deep network for image deraining. On the other hand, Zhang et al. develop a GAN-based method [1] with a perceptual loss function [21] and Yang et al. [41] develop a multi-task network for rain detection and removal. As the physical formation model is not considered, these deep models do not effectively solve the deraining problem (see Figure 1(b)).

Image super-resolution and low-level vision tasks. Significant progress has been made in super-resolution due to the use of deep network models [5], [6], [19], [42]. In [5], Dong et al. develop an end-to-end trainable network for super-resolution (SRCNN). As the SRCNN algorithm is less effective in restoring image details, Kim et al. [6] propose a deeper network with residual learning. To generate more realistic images, Ledig et al. [19] develop a GAN for image super-resolution. In addition, the deep learning methods have been applied to other low-level vision problems, such as image filtering [9], [10] and image denoising [11], [12], [13]. Different from these approaches, we propose a GAN-based method constrained by the physics model for image restoration and low-level vision problems.

3 IMAGE RESTORATION WITH GAN

The GAN algorithm learns a generative model via an adversarial learning process. It simultaneously trains a generative network and a discriminative network by optimizing

$$\min_{\mathcal{G}} \max_{\mathcal{D}} \mathbb{E}_{x \sim p_{\text{data}}(x)} [\log \mathcal{D}(x)] + \mathbb{E}_{z \sim p_z(z)} [\log(1 - \mathcal{D}(\mathcal{G}(z)))] \quad (4)$$

where z denotes random noise, x is a real image, and \mathcal{D} represents a discriminative network. For simplicity, we also use \mathcal{G} to denote a generative network.

In the training process, the generator draws samples (i.e., $\mathcal{G}(z)$) that can fool the discriminator, while the discriminator learns to distinguish the real data and those from the generator. The discriminator is a binary classifier. If we use the observed image y as the input of the generator, the adversarial loss is

$$\max_{\mathcal{D}} \mathbb{E}_{x \sim p_{\text{data}}(x)} [\log \mathcal{D}(x)] + \mathbb{E}_{y \sim p_{\text{data}}(y)} [\log(1 - \mathcal{D}(\mathcal{G}(y)))] \quad (5)$$

Thus, the value of (5) is close to zero if the distribution of the generated image $\mathcal{G}(y)$ is significantly different from that of the clear one and larger otherwise.

Taking the negative log operation on (2), the image restoration problem (2) can be solved by

$$x^* = \arg \min_x \rho(x, y) + \varphi(x), \quad (6)$$

where $\rho(x, y)$ is the data term to ensure that the recovered image y is consistent with the input image y under the corresponding image degradation model (1), and $\varphi(x)$ is the regularization of x modeling the properties of x (e.g., sparse gradient distribution [43]).

For vision tasks, e.g., image deblurring, $\varphi(x)$ acts as a discriminator and the value is much smaller if x is clear and larger otherwise [44]. In other words, optimizing the objective function (6) will make the value of $\varphi(x)$ smaller. Thus, the estimated intermediate image will be much clearer. Motivated by this, $\varphi(x)$ can be achieved by a binary classifier, which has been used in image deblurring [45]. Therefore, the adversarial loss can be used as a prior to regularize the solution space of image restoration as demonstrated by [45].

We note that GAN models with the observed data y as the input have shown promising results in image super-resolution [19], image deraining [1], and image deblurring [20], [27]. However, unconstrained GAN models do not guarantee whether the solutions satisfy the image degradation model (1) or not (i.e., a GAN model does not consider the effect of the data term in (6)) and thus do not generate clear images as illustrated in Section 1. In the following, we propose a physics-based GAN model (i.e., based on (1)) for image restoration and low-level vision tasks.

4 PROPOSED ALGORITHM

To ensure the output of a GAN (i.e., $\mathcal{G}(y)$) is consistent with the input y based on the model (1), we introduce an additional discriminative network. Figure 2 shows the main modules of the proposed algorithm: two discriminative networks, one generative network, and one image degradation model (i.e., (1)). We use the image deblurring problem as an example. Let $\{x_i\}_{i=1}^N$ and $\{y_i\}_{i=1}^N$ denote the clear and corresponding blurred images. The generative network learns the mapping function \mathcal{G} and generates the intermediate deblurred image $\mathcal{G}(y_i)$ from the input y_i . Then, we apply the image degradation model (1) to $\mathcal{G}(y_i)$

$$\tilde{y}_i = k_i \otimes \mathcal{G}(y_i), \quad (7)$$

where k_i denotes a blur kernel used only in the training process², and \otimes denotes the convolution operator. The discriminative network \mathcal{D}_h takes the blurred image y_i and the regenerated image \tilde{y}_i as the inputs to classify whether the generated results satisfy the blur model or not. The other discriminative network \mathcal{D}_g takes the ground truth x_i and the intermediate deblurred image $\mathcal{G}(y_i)$ as the inputs to classify whether $\mathcal{G}(y_i)$ is clear or not.

2. Note that the blur kernel k_i in (7) is known, which is also used to generate the blurred image y_i from the clear image x_i when synthesizing the training data. Thus, the physics model used in the proposed algorithm is not stochastic.

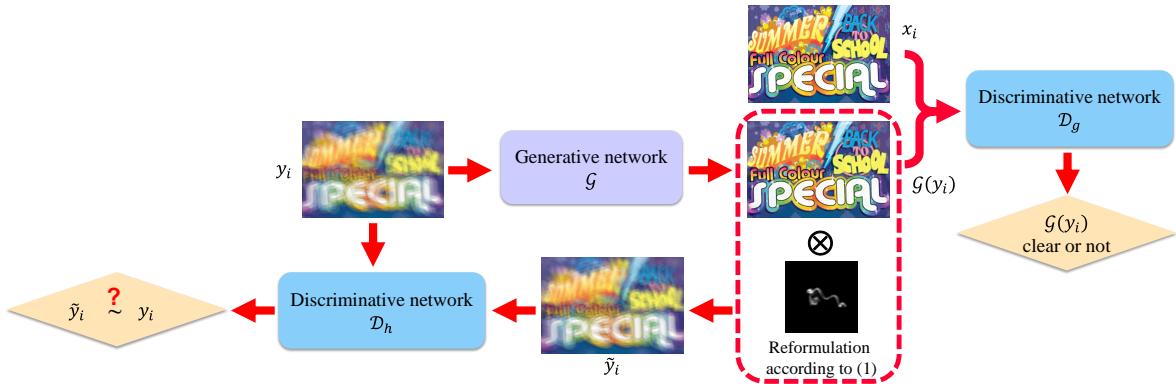


Fig. 2. Proposed algorithm. The discriminative network \mathcal{D}_g is used to classify whether the distributions of the outputs from the generator \mathcal{G} are close to those of the ground truth images or not. The discriminative network \mathcal{D}_h is used to classify whether the regenerated result \tilde{y}_i is consistent with the observed image y_i or not. All the networks are jointly trained in an end-to-end manner.

For image dehazing, the physics model is $\tilde{y}_i = \mathcal{G}(x_i)t_i + A_i(1 - t_i)$, where t_i is the transmission map and A_i is the atmospheric light. In image super-resolution, the physics model is $\tilde{y}_i = \mathcal{B}(\mathcal{G}(x_i))$, where \mathcal{B} denotes the down-sampling and filtering operator. In this paper, we use the bicubic interpolation operation. For other applications, we can use the corresponding physics models for different vision tasks.

We note that although the proposed network is trained in an end-to-end manner, it is constrained by a physics model and thus is not fully blind in the training stage. With the learned generator \mathcal{G} , the test stage is blind. We can directly obtain the final results by applying it to the input images.

4.1 Network Architecture

Once the physics model constraint is determined for each vision task, we can use existing network architectures for the generative and discriminative modules.

Generative network. The generative network \mathcal{G} is used to render the final results. Numerous generative networks have been developed for image restoration and low-level vision problems (e.g., super-resolution [21], [19] and image editing [22]). In this work, we use network model similar to that of the CycleGAN [22] as our generative module. Table 1 shows the network parameters.

Discriminative network. The PatchGANs [19], [22], [2] have fewer parameters than a full image discriminator and achieve state-of-the-art results in many vision problems. In this work, the discriminative models \mathcal{D}_h and \mathcal{D}_g are similar to those of the PatchGANs [2], [19], [22]. We use the same network parameters for both \mathcal{D}_h and \mathcal{D}_g . The discriminator model parameters are shown in Table 1.

4.2 Loss Function

A straightforward GAN model for image restoration is to use (4). However, the contents of the generated images based on this training loss may be different from the ground truth images as demonstrated

by [19]. To ensure that the contents of the generated results from the generative network are close to those of the ground truth images and also consistent with those of the inputs under the physics model (1), we use the L_1 norm regularized pixel-wise loss functions

$$\mathcal{L}_p = \sum_i \|\tilde{y}_i - y_i\|_1, \quad (8)$$

and

$$\mathcal{L}_g = \sum_i \|\mathcal{G}(y_i) - x_i\|_1, \quad (9)$$

in the training stage. To make the generative network learning process more stable, we use the loss function

$$\tilde{\mathcal{L}}_g = \sum_i \|\mathcal{G}(\tilde{y}_i) - x_i\|_1, \quad (10)$$

to regularize the generator \mathcal{G} .

In this work, we propose the objective function

$$\begin{aligned} \mathcal{L}_a = & \sum_i [\log(\mathcal{D}_g(x_i))] + [\log(1 - \mathcal{D}_g(\mathcal{G}(y_i)))] \\ & + [\log(\mathcal{D}_h(y_i))] + [\log(1 - \mathcal{D}_h(\tilde{y}_i))], \end{aligned} \quad (11)$$

to ensure that the output of the proposed GAN model is consistent with the observed input under the image formation model (1).

Based on the above considerations, the \mathcal{G} , \mathcal{D}_g , and \mathcal{D}_h models are trained by solving

$$(\mathcal{G}^*, \mathcal{D}_g^*, \mathcal{D}_h^*) = \min_{\mathcal{G}} \max_{\mathcal{D}} \frac{1}{N} \left\{ \mathcal{L}_a + \lambda \mathcal{L}_g + \gamma \mathcal{L}_p + \mu \tilde{\mathcal{L}}_g \right\}, \quad (12)$$

where λ , γ , and μ are weight parameters. To make the training process stable, we use the least square GAN loss [25] in \mathcal{L}_a in this work.

5 EXPERIMENTAL RESULTS

In this section, we evaluate the proposed algorithm on several image restoration tasks including image deblurring, dehazing, super-resolution, and deraining. We show the main experimental results in the main paper and present more analysis as well as applications in the supplementary material. The trained models, datasets, and source code are available at <https://jspan.github.io/physicsgan/>.

TABLE 1

Network configurations. ‘‘CIR’’ denotes the convolutional layer with the instance normalization (IN) [46] and ReLU; ‘‘ResBlock’’ represents the residual block [39] which contains two convolutional layers with the IN and ReLU; ‘‘CTIR’’ denotes the fractionally-strided convolutional layers with the IN and ReLU; ‘‘CILR’’ describes the convolutional layer with IN and LeakyReLU; ‘‘CT’’ describes the convolutional layer with the Tanh function.

Parameters of the generative network									
Layers	CIR ₁	CIR ₂	CIR ₃	ResBlock ₁ -ResBlock ₉			CTIR ₁	CTIR ₂	CT
Filter size	7	3	3		3		3	3	7
Filter numbers	64	128	256	256		256	128	64	3
Stride	1	2	2	1		1	2	2	1

Parameters of the discriminative network					
Layers	CILR ₁	CILR ₂	CILR ₃	CILR ₄	CILR ₅
Filter size	4	4	4	4	4
Filter numbers	64	128	256	512	1
Stride	2	2	2	1	1

5.1 Datasets

For image deblurring, we use the training dataset by Hradiš et al. [32], which consists of images with both defocus blur generated by anti-aliased disc and motion blur synthesized based on random walk. We randomly crop one million 256×256 blurred image patches from the dataset for training and use the test dataset [32] which includes 100 clear images for evaluation. In addition, with the same blur kernels [32], we randomly choose 50,000 clear face images from the CelebA dataset [47] and 50,000 natural images from the COCO dataset [48] to generate the training data for the face image deblurring and natural image deblurring. We add random noise into each blurred image, where the noise level ranges from 0 to 10%.

For image dehazing, we use the NYU depth dataset [49] and select 1,449 clear images and corresponding depth maps to generate hazy images according to the hazy model [34]. As the images in the NYU depth dataset [49] are indoor images, we also randomly choose 964 outdoor images from the Make3D dataset [50] to synthesize outdoor hazy images as the dataset. Thus, there are 2,413 synthetic images. All the images are resized to the size of 512×512 pixels. To evaluate image dehazing methods, we randomly select 241 images as the test dataset and the remaining images are used for training. In addition to these synthetic images, we also compare with the state-of-the-art methods using real captured hazy images.

For the image deraining task, we use the dataset by Zhang et al. [1] to train and evaluate the proposed algorithm against the state-of-the-art methods.

5.2 Training

We train the models using the Adam optimizer [51] with an initial learning rate of 0.0002, and the updating strategy of learning rate is the same as [22]. We set the batch size to be 1 and the slope of the LeakyReLU is 0.2. We use the same weight initialization method as [22]. After obtaining the generator \mathcal{G} , as we know the paired training data $\{x_i, y_i\}$ and corresponding physics model parameters (e.g., k_i in

image deblurring) that are used to synthesize y_i from x_i , we apply the same physics model parameters to $\mathcal{G}(y_i)$ and generate \tilde{y}_i . The discriminator \mathcal{D}_g then takes x_i and $\mathcal{G}(y_i)$ as the input while the discriminator \mathcal{D}_h takes y_i and \tilde{y}_i as the input. Similar to [22], [52], we update the discriminators using a history of generated images instead of the ones by the latest generative networks according to [22]. The update ratio between the generator and the discriminators is set to be 1. We empirically set γ and μ to be 10 and λ to be 50, respectively.

5.3 Image Deblurring

We evaluate the proposed deblurring algorithm against the state-of-the-art methods based on statistical priors [44], [53], [54] and deep neural networks [20], [32], [33], [55]. We note that recent CNN-based deblurring algorithms [20], [33], [55] are designed for natural images. For deblurring text and face images, we retrain these algorithms using the same training datasets for fair comparisons. For natural image deblurring, we use the trained models [20], [33], [55] for fair comparisons. We note that the pix2pix [2] and CycleGAN [22] algorithms are designed for image-to-image translation which can be applied to restoration. We retrain these two algorithms using the same training datasets as the proposed method for fair comparisons. Since the original CycleGAN algorithm [22] is trained with unpaired training data, the CycleGAN model trained with the paired images is referred to as the PCycleGAN method.

Synthetic blurred images. We evaluate the proposed algorithm using the text image dataset described in Section 5.1. Table 2 shows that the proposed algorithm performs favorably against state-of-the-art methods in terms of PSNR and SSIM³. Note that the method by [33] uses a multi-scale CNN with adversarial learning. However, it is less effective than the proposed algorithm Table 2. The pix2pix [2] method performs

3. As the implementation of [56] is not available, we do not compare this method in this paper. When computing the PSNR and SSIM values, we consider the possible shift that is caused by blur kernels [54].

TABLE 2

Quantitative evaluations with the state-of-the-art methods on the text image deblurring dataset by [32].

Methods	Input	Xu [53]	Pan [54]	Pan [44]	CNN [32]	Nah [33]	pix2pix [2]	CycleGAN [22]	PCycleGAN	Ours
PSNR	18.52	17.52	18.19	18.47	26.53	22.57	23.33	11.92	19.71	28.80
SSIM	0.6658	0.4186	0.6270	0.6127	0.9422	0.8924	0.9170	0.2792	0.5833	0.9744

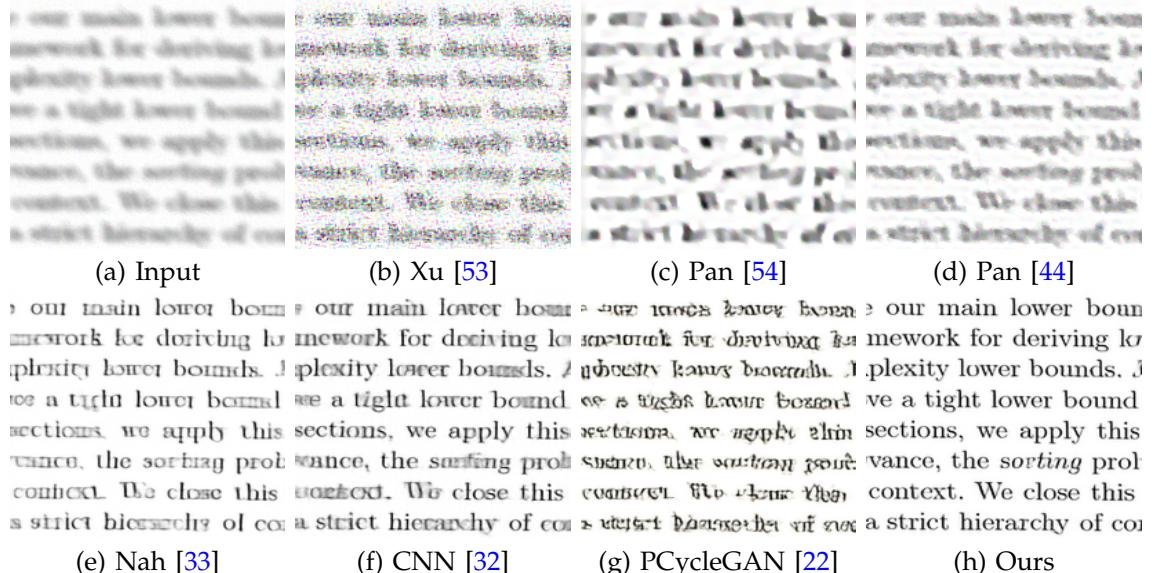


Fig. 3. One synthetic blurred image from the text image deblurring dataset [32]. The proposed method generates images with much clearer characters.

similar as [33] since both schemes are based on GANs with the pixel-wise loss function (9). In contrast, the proposed algorithm exploits a physics model which is effective image deblurring.

Figure 3 shows a blurred text image from the test dataset. The conventional algorithms [44], [53], [54] does not generate clear images. Although the CNN-based method [32] performs better, the deblurred images contain significant blur residual as the feed-forward network does not consider the consistency between the estimated results and blurred inputs. We note that the CycleGAN method [22] uses the cycle consistency constraint for the image-to-image translation task. However, it does not generate clear images as shown in Figure 3(g) although it is retrained with the paired data. In contrast, the proposed algorithm is able to generate clearer images with recognizable characters as shown in Figure 3(h), which demonstrates the importance of the image degradation model constraint.

To evaluate the proposed method on face image deblurring, we first randomly select 160 clear face images from CelebA dataset [47] and generate 160 blur kernels according to [58], where the clear images and generated blur kernels are not used in the training process. We then synthesize 160 blurred face images based on clear face images and generated kernels for test. We compare the proposed algorithm against state-of-the-art methods based on statistical priors [44], [53], [54] and deep neural networks [20], [33], [55]. Table 3 shows that the proposed algorithm

performs favorably against state-of-the-art methods in terms of PSNR and SSIM and Figure 4 shows some deblurred results. As the input image contains significant blur, several evaluated methods [22], [33], [55], [57] do not generate clear images. We note neither the DeblurGAN nor PCycleGAN method performs well on the face image shown in Figure 4. In contrast, the proposed algorithm is able to effectively deblur this image which can be attributed to the image degradation model constraint.

We further evaluate the proposed method on natural image deblurring. Similar to the settings used in face image deblurring task, we randomly select 160 clear images from COCO dataset [48] and generate 160 blur kernels according to [58], where the clear images and generated blur kernels are not used in the training process. We synthesize 160 blurred natural images based on clear natural images and generated kernels for test. The results in Table 4 show that the proposed method performs well on natural image deblurring. In addition, although the proposed network is trained on the uniform blurred images, it also performs well the dynamic scenes [33] as the blur in dynamic scenes can be approximated by locally uniform blur model [59]. More results are presented in the supplementary material.

Real blurred images. Figure 6(a) shows one of the real blurred images in the experiments. The deblurred image by the proposed method are clearer than other algorithms with sharper characters as shown in Figure 6(h). More results on real blurred images are

TABLE 3
Quantitative evaluations with the state-of-the-art deblurring methods on face images.

Methods	Xu [53]	Pan [57]	Pan [44]	Zhang [55]	Nah [33]	DeblurGAN [20]	CycleGAN [22]	PCycleGAN	Ours
PSNR	18.84	18.85	21.51	23.22	22.48	19.18	20.73	21.82	24.17
SSIM	0.4054	0.4652	0.4263	0.6832	0.4962	0.2563	0.5978	0.6018	0.7705

(a) Input
(b) Xu [53]
(c) Pan [57]
(d) Pan [44]

(e) Zhang [55]
(f) DeblurGAN [32]
(g) PCycleGAN
(h) Ours

Fig. 4. Face image deblurring results. The proposed method generates images with fewer artifacts.

TABLE 4
Quantitative evaluations with the state-of-the-art deblurring methods on natural images.

Methods	Xu [53]	Pan [54]	Pan [44]	Zhang [32]	Nah [33]	DeblurGAN [2]	CycleGAN [22]	PCycleGAN	Ours
PSNR	20.11	19.97	20.72	22.48	20.89	20.10	19.98	21.86	22.63
SSIM	0.3802	0.4419	0.3450	0.5982	0.4878	0.4596	0.5963	0.6198	0.7151

(a) Input
(b) Xu [53]
(c) Pan [54]
(d) Pan [44]

(e) Zhang [55]
(f) DeblurGAN [20]
(g) PCycleGAN
(h) Ours

Fig. 5. Natural image deblurring results. The proposed method generates images with fewer artifacts.

TABLE 5
Quantitative evaluations with the state-of-the-art methods on the proposed image dehazing dataset.

Methods	Input	He [34]	Berman [60]	Ren [15]	Cai [16]	pix2pix [2]	Ours
PSNR	13.10	18.01	16.67	17.73	20.20	24.75	24.78
SSIM	0.6958	0.7784	0.7350	0.7719	0.8134	0.8225	0.8657

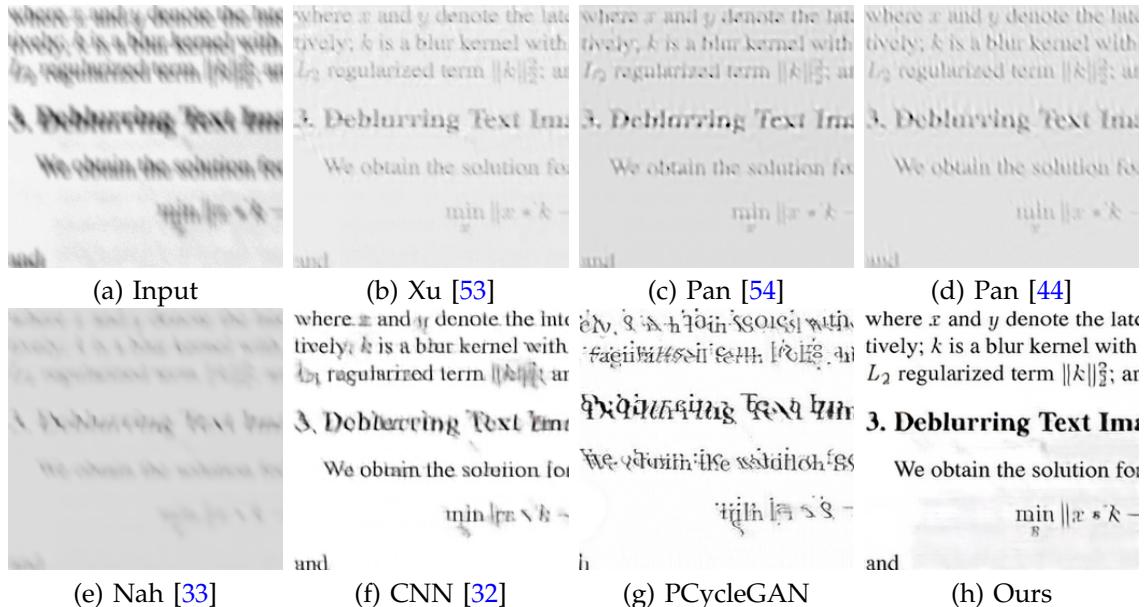


Fig. 6. Real text image deblurring results. The proposed method generates images with much clearer characters.

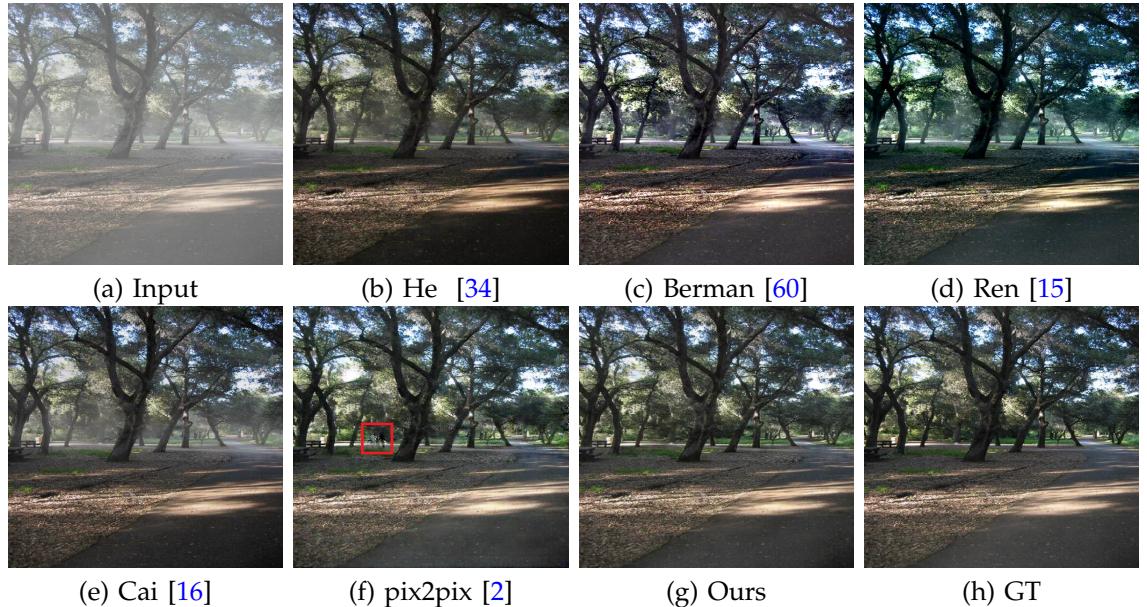


Fig. 7. One synthetic hazy image from the proposed synthetic hazy dataset. The structures enclosed in the red box in (f) are not preserved well. The proposed method does not need to estimate the transmission map and atmospheric light and generates a much clearer image which is visually close to the ground truth image.

presented in the supplementary material.

5.4 Image Dehazing

For image dehazing, we evaluate the proposed algorithm against the state-of-the-art methods based on statistical priors [34], [60], [61] and deep models [15], [16], [62]. As the methods based on neural networks [15], [16] estimate transmission maps and the training code is not available [16], we use the provided models for comparisons. We also retrain pix2pix [2] and CycleGAN [22] using the proposed training dataset for fair comparisons.

Synthetic hazy images. We first evaluate our method on the synthetic dataset as mentioned in Section 5.1.

The test dataset contains 241 images including both indoor and outdoor scenes. Table 5 shows that the proposed algorithm performs favorably against state-of-the-art methods in terms of PSNR and SSIM.

We show one example from the test dataset in Figure 7. The conventional methods [34], [60] need to estimate both the transmission map and atmospheric light based on manually designed features. The colors of the generated images of [34] and [60] are slightly different from those of the ground truth images due to inaccurate estimations of the transmission map and atmospheric light. The CNN-based methods [15], [16] first estimate the transmission maps from hazy images and use the conventional schemes to restore clear

images. The restored images contain hazy residual as shown in Figure 7(d)-(e). Although the GAN-based methods (e.g., pix2pix [2]) are able to generate realistic images, the structures are not preserved well. In contrast, due to the use of the image degradation model constraint, the proposed algorithm generates a much clearer image with fine details, which is visually close to the ground truth image.

We further evaluate the proposed algorithm on the outdoor dataset [63] which contains 45 images. Although the hazy images are captured in hazy environments, we refer to this dataset as a synthetic dataset because the haze is manmade. As the resolution of the original test images is too large, we resize each test image to the size of 512×512 pixels for comparisons. Table 6 and Figure 8 show that the proposed algorithm performs favorably against state-of-the-art dehazing methods.

Real hazy images. We evaluate the proposed algorithm on real hazy images and show some results against the state-of-the-art methods in Figure 9. The colors of the restored image by the proposed method are more vivid than the other algorithms.

5.5 Image Super-resolution

The proposed algorithm can be applied to image super-resolution by replacing the physics model (1) with the image formulation of super-resolution as stated in Section 4. We randomly choose 50,000 images from the COCO dataset [48] to train the proposed algorithm and evaluate it on the images from the “Set5” against the state-of-the-art algorithms including SRCNN [3], ESPCN [65], VDSR [6], SRGAN [19], and EDSR [41]. We fine-tune the methods based on deep neural networks using the proposed training dataset and choose the best models for fair comparisons. Quantitative evaluation results are shown in Table 7. Although the proposed algorithm is not designed for super-resolution, it achieves competitive results compared to the state-of-the-art methods.

5.6 Image Deraining

We further apply the proposed algorithm to image deraining problem which aims to remove rainy streaks or dirties from the input images. The image deraining model is based on a linear superimposition of clear image and rainy streak. We evaluate the proposed algorithm against the conventional image prior based method [38] and state-of-the-art deep learning based methods [1], [41], [67] using the test dataset by [1] for comparisons. We also retrain pix2pix [2], CycleGAN [22], and PCycleGAN using the same training dataset [1] for fair comparisons.

Figure 10 shows a real rainy image. The method [38] based on handcrafted image priors does not remove rainy streaks well. We note that the image deraining algorithm [1] is based on a GAN and a perceptual loss function [21]. However, there exist significant color

distortions in the restored image (e.g., the part in the red box in Figure 10(c)). As discussed in Section 1, this is mainly because the algorithm [1] does not consider the image degradation model constraint. The scheme by Fu et al. [67] decomposes the input image into a detail layer and a base layer, where the detail layer is estimated by a CNN. However, this method is less effective for the scenes with heavy rain. The method by Yang et al. [41] based on a multi-task network does not remove the rainy streaks well as shown in Figure 10(e). The pix2pix [2] and PCycleGAN algorithms do not generate clear images. In contrast, the proposed method is able to remove rainy streaks and generates a much clearer image.

We further evaluate the proposed method on other related low-level vision problems and present more experimental results in the supplementary material.

6 ANALYSIS AND DISCUSSION

In this section, we analyze the performance of the proposed algorithm against the state-of-the-art methods.

Effect of the image degradation model constraint. Our method without the image degradation model constraint reduces to the GAN model [18] with the loss function (9). To analyze the effect of this constraint with fair comparisons, we remove the physics model and corresponding loss functions (BaseGAN for short) in our implementation and train the baseline model on the same image dehazing dataset which contains 6,900 hazy images. We randomly choose 495 synthetic hazy images (which are not used in the training stage) for evaluations. Table 8 shows the experimental results of all the baseline algorithms and one example is shown in Figure 11. As shown in Figure 11(b), the BaseGAN method does not generate clear images. The structures of the restored image (e.g., the sky) by the BaseGAN method are significantly different from the input when the physics model is not used. In contrast, the proposed method with the image degradation model constrained learning process generates much clearer images.

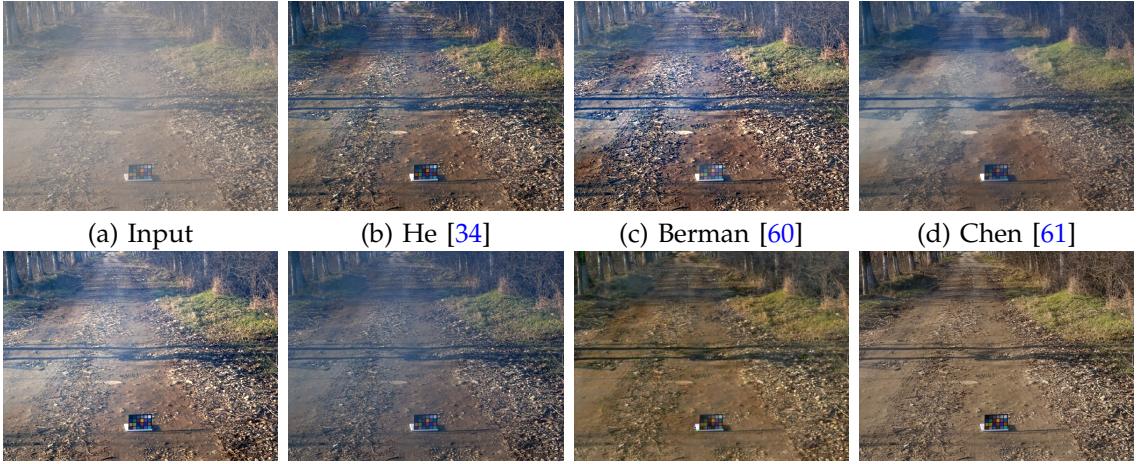
We show the effect of the proposed loss functions in Table 8. We note that the method without loss functions (8), (9), or (10) is less effective and using one of these helps obtain better results. The restored images in Figure 11(c)-(f) demonstrate that the proposed algorithm with the proposed loss function is able to generate clear images.

Relation with GAN-based methods. Recently, several methods have been proposed to improve the GAN [18], e.g., CycleGAN [22], DiscoGAN [23], and DualGAN [24]. The CycleGAN algorithm [22] uses two generators and two discriminators for image-to-image translation when the paired training data is not available. In addition, a cycle consistency loss is used to train the proposed CycleGAN network. The other methods [1], [2], [19], [26], [27] explore GANs in the

TABLE 6

Quantitative evaluations with the state-of-the-art methods on the outdoor image dehazing dataset [63]. The reported results in this table are generated by the provided implementations or models.

Methods	Input	He [34]	Berman [60]	Chen [61]	Meng [64]	Ren [15]	Cai [16]	Zhang [62]	Ours
PSNR	13.89	16.60	15.76	15.92	16.04	17.44	15.66	13.08	17.76
SSIM	0.6296	0.6915	0.7483	0.6322	0.7374	0.7551	0.6744	0.7301	0.7795



(a) Input
(b) He [34]
(c) Berman [60]
(d) Chen [61]

(e) Ren [15]
(f) Cai [16]
(g) Ours
(h) GT

Fig. 8. A hazy image from the dataset [63]. The proposed method generates a much clearer image which is visually close to the ground truth.

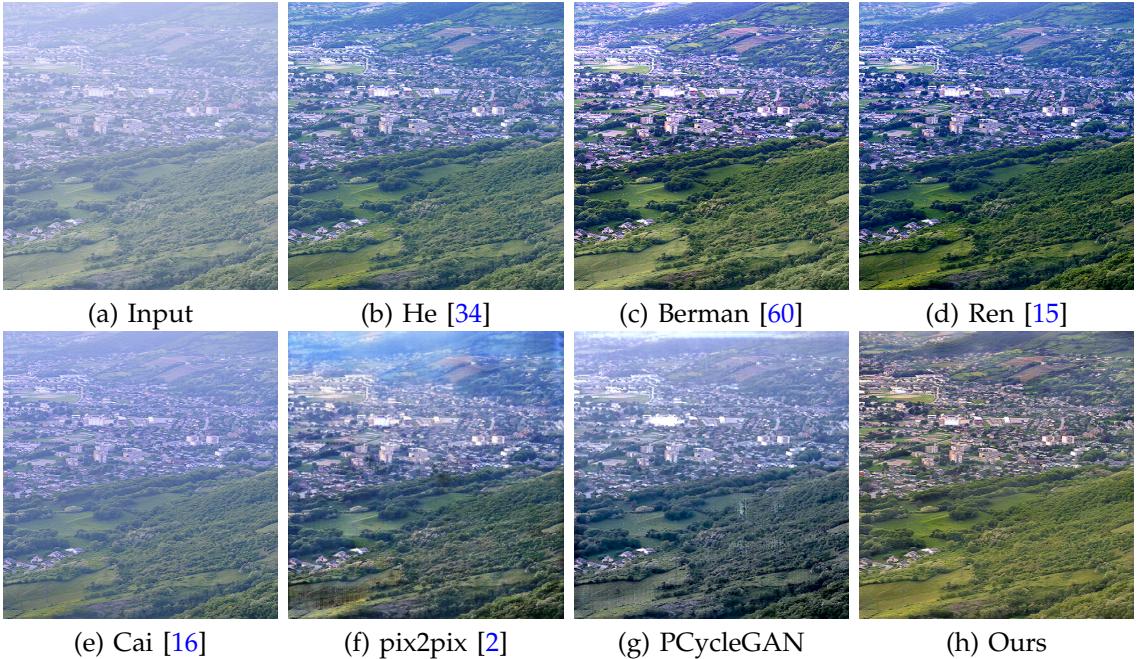


Fig. 9. Results on a real hazy image. The proposed method generates a clearer image.

TABLE 7

Quantitative evaluations with the state-of-the-art methods on the image super-resolution ($\times 4$) problem.

Methods	Bicubic	SRCCN [3]	ESPCN [65]	VDSR [6]	SRGAN [19]	EDSR [66]	Ours
PSNR	28.42	30.50	30.27	31.41	32.05	32.46	30.03
SSIM	0.8104	0.8629	0.8540	0.8840	0.8910	0.8968	0.9030

conditional settings and use a pixel-wise loss function (i.e., (9)) with a perceptual loss function [21] to ensure that the outputs of the generative network are close to the ground truth images in the training stage.

Although these methods are not designed for the image restoration problems addressed in this work, we train the most related algorithms using the same training datasets as the proposed approach to clarify

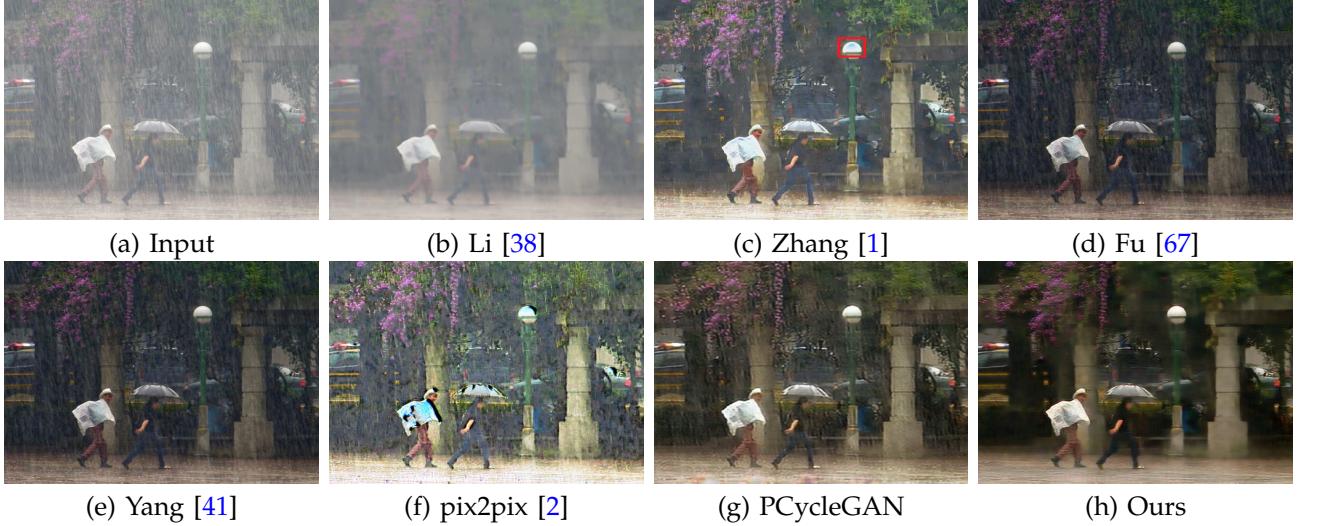


Fig. 10. Results on a real rainy image. Due to the heavy rain, both the CNN-based [67], [41] and GAN-based [1], [2], [22] methods are not able to generate clear images. The part enclosed in the red box in (c) contains significant color distortions (best viewed on high-resolution displays with zoom-in).

the differences. As these algorithms directly learn the inverse process (i.e., (3)) by end-to-end trainable networks without considering the image formation process, they are not able to generate clear images as demonstrated in Section 5.3-5.6.

Using the notations in the proposed method, the CycleGAN algorithm assumes that $\mathcal{F}(\mathcal{G}(y_i)) = y_i$ and $\mathcal{G}(\mathcal{F}(x_i)) = x_i$, where \mathcal{F} is a generator which has the similar effect to the mapping function \mathcal{H} in the physics model (1). With this assumption, this algorithm is likely to converge to a trivial solution as the identity mapping functions always hold for this assumption. This is the main reason why the results by the CycleGAN algorithm are similar to the inputs, e.g., the dehazing results are similar to the hazy inputs (Figure 9(g)).

Different from the CycleGAN algorithm, our method does not learn the generator \mathcal{F} as the physics model is known. Thus, our method is able to avoid trivial solutions and performs favorably against the state-of-the-art algorithms on each task.

As the proposed method uses loss functions (8), (9), and (10), it is of interest to analyze the performance gain if the same loss functions are used in the PCycleGAN method. Using the same image dehazing datasets, we train a model using the PCycleGAN method with the same loss functions as the proposed algorithm for evaluation. Table 8 and Figure 11(g) show that the PCycleGAN method with the same loss functions does not generate clear images.

Robustness to image noise. We evaluate the proposed method using 100 clear text images from [32], where each sample is blurred and added with random noise ranging from 0% to 10%. Figure 12(a) shows that the proposed method performs well even when the noise level is high.

Convergence property. As our algorithm needs to

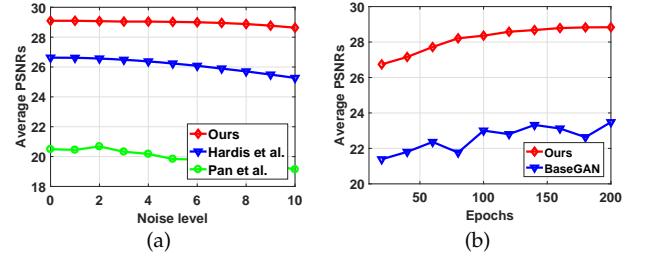


Fig. 12. (a) Quantitative evaluations on the blurred text images with random noise. The proposed method is robust to image noise in image deblurring. (b) Quantitative evaluations of the convergence property on the blurred text images.

jointly train generative and discriminative networks, a natural question is whether our method converges well or not. We quantitatively evaluate the convergence properties of our method on the text image deblurring dataset [32]. Figure 12(b) shows that the proposed method converges well within 200 epochs in terms of PSNR.

We note that although using multiple discriminators in GANs may increase complexity in the training stage, our numerical results in Figure 12(b) show that using the physics model makes the training process more stable and leads to better convergence performance compared to the GAN algorithm with one discriminator and generator.

Sensitivity analysis. The proposed objective function involves the weight parameter λ which is important for restoration tasks. We analyze the effect of this parameter using blurred face images from the proposed test dataset by setting its values from 10 to 100 with the step size of 10. Table 9 shows that the proposed method performs well when λ is set to be within a wide range (i.e., [40, 100]) and slightly better when this parameter is set to be 50.

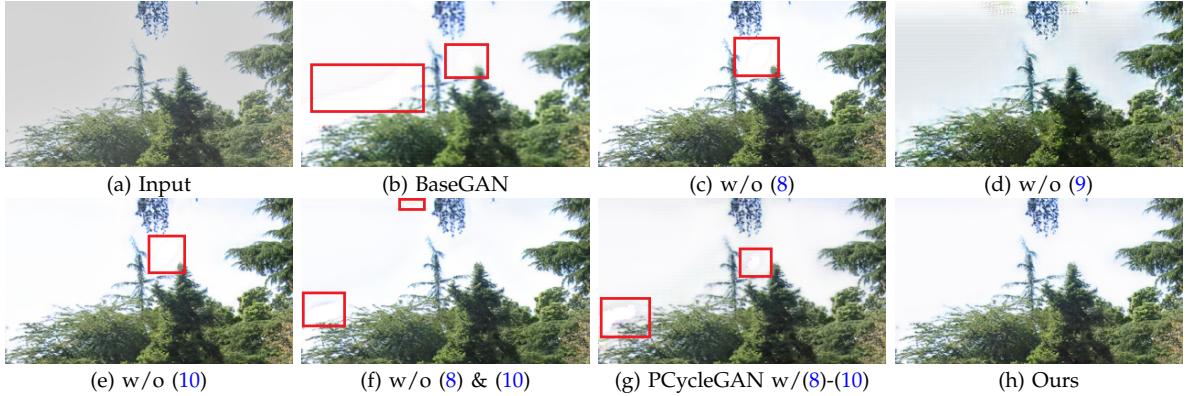


Fig. 11. Effectiveness of the proposed method and loss function on image dehazing. The structures of the restored images (e.g., the sky and the part enclosed in the red boxes) by the baseline methods are not estimated well.

TABLE 8
Effectiveness of the proposed algorithm and loss functions on the image dehazing task.

Input	BaseGAN	w/o (8)	w/o (9)	w/o (10)	w/o (8) & (10)	PCycleGAN w/ (8)-(10)	Ours
PSNR	18.52	29.90	31.93	21.24	31.71	30.53	27.46
SSIM	0.8357	0.9541	0.9718	0.8096	0.9693	0.9694	0.9722

TABLE 9
Sensitivity analysis with respect to the weight parameter λ and the number of ResBlocks
Effect of weight parameter λ on image deblurring

λ	10	20	30	40	50	60	70	80	90	100
PSNR	22.92	23.77	24.06	24.25	24.57	24.36	24.55	24.54	24.41	24.55
SSIM	0.6782	0.7103	0.7378	0.7489	0.7531	0.7528	0.7489	0.7521	0.7477	0.7488

Influence of the number of ResBlocks (#) on image deblurring										
ResBlock (#)	3	6	9	20	25	30	35	40	45	50
PSNR	20.45	20.49	20.77	20.98	20.33	20.82	21.22	20.93	20.88	20.83
SSIM	0.5513	0.5628	0.6005	0.6120	0.5483	0.6137	0.6173	0.6062	0.6018	0.6075

Ablation study w.r.t. the number of ResBlocks. The proposed algorithm contains several ResBlocks. To analyze the effects of the number of ResBlocks, we evaluate the proposed network using blurred face images from the proposed test dataset by setting the number of ResBlocks from 3 to 50. Table 9 demonstrates that the proposed method is insensitive to the number of ResBlocks. We empirically use 9 ResBlocks as a trade-off between accuracy and speed.

Limitations. Although the proposed method is able to restore images from degraded ones, it is still less effective for those examples which are caused by multiple degradation factors, e.g., both rain and haze. For these cases, the physics model does not describe the complex image formation process well. Figure 13 shows an example where the proposed algorithm is not able to remove rain/snow from the input image due to the complex degradation process. Our future work will consider jointly using semi-supervised or unsupervised learning algorithms to address these issues.

7 CONCLUDING REMARKS

Motivated by the observation that the estimated results should be consistent with the observed inputs under the physical formation process in image



Fig. 13. The proposed algorithm is less effective for the images where the physics model does not describe the complex degradation process well.

restoration problems, we enforce the physics model in a generative adversarial network for image restoration. As the physics model is derived from the image formation process of low-level problems, the proposed algorithm can be applied to a variety of tasks. With an end-to-end network formulation, the proposed algorithm performs favorably against the state-of-the-art methods for image restoration and low-level vision problems.

ACKNOWLEDGMENTS

This work has been supported in part by the National Natural Science Foundation of China (Nos. 61922043, 61872421, 61732007), the Natural Science Foundation

of Jiangsu Province (No. BK20180471), and National Science Foundation CAREER (No. 1149783).

REFERENCES

- [1] H. Zhang, V. Sindagi, and V. M. Patel, "Image de-raining using a conditional generative adversarial network," *CoRR*, vol. abs/1701.05957, 2017. [1](#) [2](#) [3](#) [5](#) [9](#) [11](#)
- [2] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *CVPR*, 2017, pp. 1125–1134. [2](#) [4](#) [5](#) [6](#) [7](#) [8](#) [9](#) [10](#) [11](#)
- [3] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE TPAMI*, vol. 38, no. 2, pp. 295–307, 2016. [1](#) [9](#) [10](#)
- [4] C. Dong, C. C. Loy, and X. Tang, "Accelerating the super-resolution convolutional neural network," in *ECCV*, 2016, pp. 391–407. [1](#)
- [5] C. Dong, C. C. Loy, K. He, and X. Tang, "Learning a deep convolutional network for image super-resolution," in *ECCV*, 2014, pp. 184–199. [1](#) [3](#)
- [6] J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," in *CVPR*, 2016, pp. 1646–1654. [1](#) [3](#) [9](#) [10](#)
- [7] ——, "Deeply-recursive convolutional network for image super-resolution," in *CVPR*, 2016, pp. 1637–1645. [1](#)
- [8] R. Liao, X. Tao, R. Li, Z. Ma, and J. Jia, "Video super-resolution via deep draft-ensemble learning," in *ICCV*, 2015, pp. 531–539. [1](#)
- [9] L. Xu, J. S. J. Ren, Q. Yan, R. Liao, and J. Jia, "Deep edge-aware filters," in *ICML*, 2015, pp. 1669–1678. [1](#) [3](#)
- [10] S. Liu, J. Pan, and M.-H. Yang, "Learning recursive filters for low-level vision via a hybrid neural network," in *ECCV*, 2016, pp. 560–576. [1](#) [3](#)
- [11] C. Dong, Y. Deng, C. C. Loy, and X. Tang, "Compression artifacts reduction by a deep convolutional network," in *ICCV*, 2015, pp. 576–584. [1](#) [3](#)
- [12] V. Jain and H. S. Seung, "Natural image denoising with convolutional networks," in *NIPS*, 2008, pp. 769–776. [1](#) [3](#)
- [13] J. Xie, L. Xu, and E. Chen, "Image denoising and inpainting with deep neural networks," in *NIPS*, 2012, pp. 350–358. [1](#) [3](#)
- [14] D. Eigen, D. Krishnan, and R. Fergus, "Restoring an image taken through a window covered with dirt or rain," in *ICCV*, 2013, pp. 633–640. [1](#) [3](#)
- [15] W. Ren, S. Liu, H. Zhang, J. Pan, X. Cao, and M.-H. Yang, "Single image dehazing via multi-scale convolutional neural networks," in *ECCV*, 2016, pp. 154–169. [1](#) [2](#) [7](#) [8](#) [10](#)
- [16] B. Cai, X. Xu, K. Jia, C. Qing, and D. Tao, "Dehazenet: An end-to-end system for single image haze removal," *IEEE TIP*, vol. 25, no. 11, pp. 5187–5198, 2016. [1](#) [2](#) [7](#) [8](#) [10](#)
- [17] B. Li, X. Peng, Z. Wang, J. Xu, and D. Feng, "Aod-net: All-in-one dehazing network," in *ICCV*, 2017, pp. 4770–4778. [1](#)
- [18] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. C. Courville, and Y. Bengio, "Generative adversarial nets," in *NIPS*, 2014, pp. 2672–2680. [1](#) [2](#)
- [19] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, "Photo-realistic single image super-resolution using a generative adversarial network," in *CVPR*, 2017, pp. 4681–4690. [1](#) [2](#) [3](#) [4](#) [9](#) [10](#)
- [20] O. Kupyn, V. Budzan, M. Mykhailych, D. Mishkin, and J. Matas, "Deblurgan: Blind motion deblurring using conditional adversarial networks," in *CVPR*, 2018, pp. 8183–8192. [1](#) [2](#) [3](#) [5](#) [6](#) [7](#)
- [21] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *ECCV*, 2016, pp. 694–711. [1](#) [3](#) [4](#) [9](#) [10](#)
- [22] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *ICCV*, 2017, pp. 2223–2232. [2](#) [4](#) [5](#) [6](#) [7](#) [8](#) [9](#) [11](#)
- [23] T. Kim, M. Cha, H. Kim, J. K. Lee, and J. Kim, "Learning to discover cross-domain relations with generative adversarial networks," in *ICML*, 2017, pp. 1857–1865. [2](#) [9](#)
- [24] Z. Yi, H. Zhang, P. Tan, and M. Gong, "Dualgan: Unsupervised dual learning for image-to-image translation," in *CoRR*, 2017, pp. 2849–2857. [2](#) [9](#)
- [25] X. Mao, Q. Li, H. Xie, R. Y. K. Lau, and Z. Wang, "Multi-class generative adversarial networks with the L2 loss function," *CoRR*, vol. abs/1611.04076, 2016. [2](#) [4](#)
- [26] C. K. Sønderby, J. Caballero, L. Theis, W. Shi, and F. Huszár, "Amortised MAP inference for image super-resolution," in *ICLR*, 2017. [2](#) [9](#)
- [27] X. Xu, D. Sun, J. Pan, Y. Zhang, H. Pfister, and M.-H. Yang, "Learning to super-resolve blurry face and text images," in *ICCV*, 2017, pp. 251–260. [2](#) [3](#) [9](#)
- [28] C. J. Schuler, H. C. Burger, S. Harmeling, and B. Schölkopf, "A machine learning approach for non-blind image deconvolution," in *CVPR*, 2013, pp. 1067–1074. [2](#)
- [29] L. Xu, J. S. J. Ren, C. Liu, and J. Jia, "Deep convolutional neural network for image deconvolution," in *NIPS*, 2014, pp. 1790–1798. [2](#)
- [30] J. Sun, W. Cao, Z. Xu, and J. Ponce, "Learning a convolutional neural network for non-uniform motion blur removal," in *CVPR*, 2015, pp. 769–777. [2](#)
- [31] C. J. Schuler, M. Hirsch, S. Harmeling, and B. Schölkopf, "Learning to deblur," *IEEE TPAMI*, vol. 38, no. 7, pp. 1439–1451, 2016. [2](#)
- [32] M. Hradis, J. Kotera, P. Zemcik, and F. Sroubek, "Convolutional neural networks for direct text deblurring," in *BMVC*, 2015, pp. 6.1–6.13. [2](#) [5](#) [6](#) [7](#) [8](#) [11](#)
- [33] S. Nah, T. Hyun Kim, and K. Mu Lee, "Deep multi-scale convolutional neural network for dynamic scene deblurring," in *CVPR*, 2017, pp. 3883–3891. [2](#) [5](#) [6](#) [7](#) [8](#)
- [34] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," in *CVPR*, 2009, pp. 1956–1963. [2](#) [3](#) [5](#) [7](#) [8](#) [10](#)
- [35] L.-W. Kang, C.-W. Lin, and Y.-H. Fu, "Automatic single-image-based rain streaks removal via image decomposition," *IEEE TIP*, vol. 21, no. 4, pp. 1742–1755, 2012. [3](#)
- [36] Y. Luo, Y. Xu, and H. Ji, "Removing rain from a single image via discriminative sparse coding," in *ICCV*, 2015, pp. 3397–3405. [3](#)
- [37] Y.-L. Chen and C.-T. Hsu, "A generalized low-rank appearance model for spatio-temporally correlated rain streaks," in *ICCV*, 2013, pp. 1968–1975. [3](#)
- [38] Y. Li, R. T. Tan, X. Guo, J. Lu, and M. S. Brown, "Rain streak removal using layer priors," in *CVPR*, 2016, pp. 2736–2744. [3](#) [9](#) [11](#)
- [39] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *CVPR*, 2016, pp. 770–778. [3](#) [5](#)
- [40] X. Fu, J. Huang, D. Zeng, Y. Huang, X. Ding, and J. Paisley, "Removing rain from single images via a deep detail network," in *CVPR*, 2017, pp. 3855–3863. [3](#)
- [41] W. Yang, R. T. Tan, J. Feng, J. Liu, Z. Guo, and S. Yan, "Deep joint rain detection and removal from a single image," in *CVPR*, 2017, pp. 1357–1366. [3](#) [9](#) [11](#)
- [42] M. S. M. Sajjadi, B. Schölkopf, and M. Hirsch, "Enhancenet: Single image super-resolution through automated texture synthesis," in *ICCV*, 2017, pp. 4501–4510. [3](#)
- [43] A. Levin, Y. Weiss, F. Durand, and W. T. Freeman, "Understanding and evaluating blind deconvolution algorithms," in *CVPR*, 2009, pp. 1964–1971. [3](#)
- [44] J. Pan, D. Sun, H. Pfister, and M.-H. Yang, "Blind image deblurring using dark channel prior," in *CVPR*, 2016, pp. 1628–1636. [3](#) [5](#) [6](#) [7](#) [8](#)
- [45] L. Li, J. Pan, W.-S. Lai, C. Gao, N. Sang, and M.-H. Yang, "Learning a discriminative prior for blind image deblurring," in *CVPR*, 2018, pp. 6616–6625. [3](#)
- [46] D. Ulyanov, A. Vedaldi, and V. S. Lempitsky, "Instance normalization: The missing ingredient for fast stylization," *CoRR*, vol. abs/1607.08022, 2016. [5](#)
- [47] Z. Liu, P. Luo, X. Wang, and X. Tang, "Deep learning face attributes in the wild," in *ICCV*, 2015, pp. 3730–3738. [5](#) [6](#)
- [48] T.-Y. Lin, M. Maire, S. J. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft COCO: common objects in context," in *ECCV*, 2014, pp. 740–755. [5](#) [6](#) [9](#)
- [49] N. Silberman, D. Hoiem, P. Kohli, and R. Fergus, "Indoor segmentation and support inference from RGBD images," in *ECCV*, 2012, pp. 746–760. [5](#)
- [50] A. Saxena, M. Sun, and A. Y. Ng, "Make3d: Learning 3d scene structure from a single still image," *IEEE TPAMI*, vol. 31, no. 5, pp. 824–840, 2009. [5](#)

- [51] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *CoRR*, vol. abs/1412.6980, 2014. [5](#)
- [52] A. Shrivastava, T. Pfister, O. Tuzel, J. Susskind, W. Wang, and R. Webb, "Learning from simulated and unsupervised images through adversarial training," *CoRR*, vol. abs/1612.07828, 2016. [5](#)
- [53] L. Xu, S. Zheng, and J. Jia, "Unnatural L_0 sparse representation for natural image deblurring," in *CVPR*, 2013, pp. 1107–1114. [5, 6, 7, 8](#)
- [54] J. Pan, Z. Hu, Z. Su, and M.-H. Yang, " L_0 -regularized intensity and gradient prior for deblurring text images and beyond," *IEEE TPAMI*, vol. 39, no. 2, pp. 342–355, 2017. [5, 6, 7, 8](#)
- [55] J. Zhang, J. Pan, J. Ren, Y. Song, L. Bao, R. W. Lau, and M.-H. Yang, "Dynamic scene deblurring using spatially variant recurrent neural networks," in *CVPR*, 2018, pp. 2521–2529. [5, 6, 7](#)
- [56] L. Xiao, J. Wang, W. Heidrich, and M. Hirsch, "Learning high-order filters for efficient blind deconvolution of document photographs," in *ECCV*, 2016, pp. 734–749. [5](#)
- [57] J. Pan, Z. Hu, Z. Su, and M.-H. Yang, "Deblurring face images with exemplars," in *ECCV*, 2014, pp. 47–62. [6, 7](#)
- [58] U. Schmidt, C. Rother, S. Nowozin, J. Jancsary, and S. Roth, "Discriminative non-blind deblurring," in *CVPR*, 2013, pp. 604–611. [6](#)
- [59] J. Pan, Z. Hu, Z. Su, H.-Y. Lee, and M.-H. Yang, "Soft-segmentation guided object motion deblurring," in *CVPR*, 2016, pp. 459–468. [6](#)
- [60] D. Berman, T. Treibitz, and S. Avidan, "Non-local image dehazing," in *CVPR*, 2016, pp. 1674–1682. [7, 8, 10](#)
- [61] C. Chen, M. N. Do, and J. Wang, "Robust image and video dehazing with visual artifact suppression via gradient residual minimization," in *ECCV*, 2016, pp. 576–591. [8, 10](#)
- [62] H. Zhang and V. M. Patel, "Densely connected pyramid dehazing network," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 3194–3203. [8, 10](#)
- [63] <http://www.vision.ee.ethz.ch/ntire18/>. [9, 10](#)
- [64] G. Meng, Y. Wang, J. Duan, S. Xiang, and C. Pan, "Efficient image dehazing with boundary constraint and contextual regularization," in *ICCV*, 2013, pp. 617–624. [10](#)
- [65] W. Shi, J. Caballero, F. Huszar, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, and Z. Wang, "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network," in *CVPR*, 2016, pp. 1874–1883. [9, 10](#)
- [66] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, "Enhanced deep residual networks for single image super-resolution," in *CVPR Workshops*, 2017, pp. 1132–1140. [10](#)
- [67] X. Fu, J. Huang, X. Ding, Y. Liao, and J. Paisley, "Clearing the skies: A deep network architecture for single-image rain removal," *IEEE TIP*, 2017. [9, 11](#)



Jinshan Pan is a professor of School of Computer Science and Engineering, Nanjing University of Science and Technology. He received the Ph.D. degree in computational mathematics from the Dalian University of Technology, China, in 2017. His research interest includes image deblurring, image/video analysis and enhancement, and related vision problems.



Jiangxin Dong is currently working toward the PhD degree in the School of Mathematical Sciences, Dalian University of Technology. She received the BS degree in 2014. Her research interests include image deblurring and denoising.



Yang Liu is currently working toward the PhD degree in the School of Mathematical Sciences, Dalian University of Technology. He received the B.Sc degree from Dalian University of Technology in 2016. His research interests include image deblurring, image super-resolution and related computer vision problems.



Jiawei Zhang is a research scientist at SenseTime Research. He received a PhD degree from City University of Hong Kong in 2018, a master degree from Institute of Acoustics, Chinese Academy of Sciences in 2014 and a bachelor degree from University of Science and Technology of China in 2011 respectively.



Jimmy Ren is a senior research scientist at SenseTime Research. He received Ph.D from City University of Hong Kong. His research interests include: machine learning and deep learning with focus on deep convolutional neural networks, recurrent neural networks and their applications in computer vision, image processing and business.



Jinhu Tang is a professor in the School of Computer Science and Engineering, Nanjing University of Science and Technology. He received the BE and PhD degrees from the University of Science and Technology of China, in 2003 and 2008, respectively. He received the Best Student Paper Award in MMM 2016, and Best Paper Awards in ACM MM 2007, PCM 2011, and ICIMCS 2011. He is a senior member of the IEEE and a member of the ACM.



Yu-Wing Tai is a research director of YouTu research lab of Social Network Group of Tencent. He received his BEng (First Class Honors) and MPhil degrees from the Department of Computer Science and Engineering, HKUST in 2003 and 2005 and PhD degree from the National University of Singapore in 2009. He was an associate professor at the Korea Advanced Institute of Science and Technology (KAIST) from July 2009 to August 2015. He was awarded the Microsoft Research Asia Fellowship in 2007, and the KAIST 40th Anniversary Academic Award for Excellent Professor in 2011 respectively.



Ming-Hsuan Yang is a professor of Electrical Engineering and Computer Science with the University of California, Merced, CA, USA. He received the Ph.D. degree in computer science from the University of Illinois at Urbana-Champaign, USA, in 2000. He served as an Associate Editor of the IEEE Transactions on Pattern Analysis and Machine Intelligence from 2007 to 2011, and is an Associate Editor of the International Journal of Computer Vision, Image and Vision Computing, and Journal of Artificial Intelligence Research. He received the NSF CAREER Award in 2012, and the Google Faculty Award in 2009. He is a fellow of the IEEE.