Single Image Dehazing via Conditional Generative Adversarial Network

Runde Li* Jinshan Pan* Zechao Li Jinhui Tang†
School of Computer Science and Engineering, Nanjing University of Science and Technology

Abstract

In this paper, we present an algorithm to directly restore a clear image from a hazy image. This problem is highly ill-posed and most existing algorithms often use hand-crafted features, e.g., dark channel, color disparity, maximum contrast, to estimate transmission maps and then atmospheric lights. In contrast, we solve this problem based on a conditional generative adversarial network (cGAN), where the clear image is estimated by an end-to-end trainable neural network. Different from the generative network in basic cGAN, we propose an encoder and decoder architecture so that it can generate better results. To generate realistic clear images, we further modify the basic cGAN formulation by introducing the VGG features and an L_1-regularized gradient prior. We also synthesize a hazy dataset including indoor and outdoor scenes to train and evaluate the proposed algorithm. Extensive experimental results demonstrate that the proposed method performs favorably against the state-of-the-art methods on both synthetic dataset and real world hazy images.

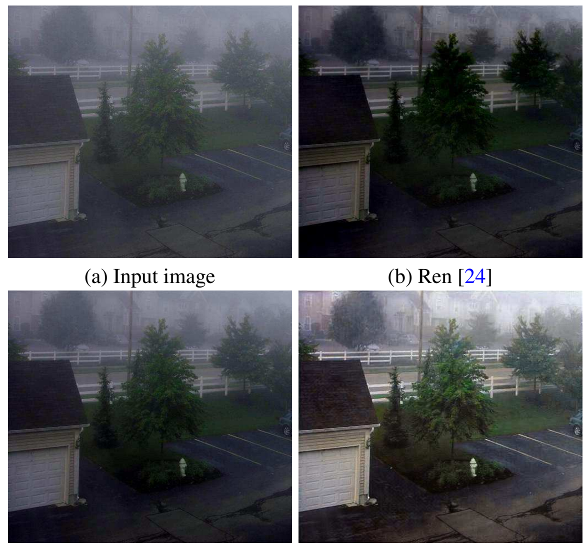
1. Introduction

Image dehazing aims to restore a clear image from a hazy image which is corrupted by haze, fog or smoke. This process can be formulated by [11, 33]

$$I(x) = J(x)t(x) + A(1 - t(x)) \tag{1}$$

where $I(x)$ and $J(x)$ represent the hazy image and the scene radiance, respectively. A is the global atmospheric light, and $t(x)$ is the medium transmission map. If the haze is homogeneous, the transmission map can be expressed as $t(x) = e^{-\rho d(x)}$, where ρ is the medium extinction coefficient and $d(x)$ is the scene depth. x indexes pixels in an image. As only the observed image $I(x)$ is known, recovering the scene radiance $J(x)$ is highly ill-posed.

In recent years, we have witnessed significant advances in image dehazing mainly due to using hand-crafted features to estimate transmission maps and atmospheric lights [1, 3, 4, 10, 11, 19, 20, 22, 40]. The commonly used hand-crafted features are mainly based on chromatic, textural and contrast properties. However, methods based on these features do not work well for some cases since the assumptions on the features do not always hold. For example, He *et al.* [11] assume that the values of dark channel in clear images are close to zero and then use it to estimate the transmission map. However, it does not work well for the scene objects which are similar to the atmospheric light. Recently, some deep learning-based methods have been proposed to solve image dehazing. These methods first use convolution neural networks to estimate the transmission map and then follow the conventional method to estimate the atmospheric light to recover clear images. However, if the transmission map is not well estimated, they will accordingly interfere the estimation of atmospheric light. Therefore the final recovered image usually contains color distortions or artifacts. As most existing algorithms estimate the transmission map and atmospheric light separately, it is of great interest



(a) Input image (b) Ren [24] (c) Cai [3] (d) Ours

Figure 1: Image dehazing example. Existing methods usually estimate transmission maps and atmospheric lights separately, which cannot effectively solve the problem. Our proposed end-to-end trainable method avoids this problem and generates much better images.

†Corresponding author: jinhuitang@njust.edu.cn. * equal contribution.

to jointly estimate the transmission map and atmospheric light. To that end, we propose an end-to-end trainable network to solve the aforementioned problems.

Our end-to-end trainable network is based on cGAN, where the generator contains an encoder and decoder architecture so that it can capture more useful information and generate much better outputs. The discriminator is used to distinguish whether outputs from the generator are fake or not. To preserve details of the output from generator, we use the pre-trained VGG features as the perceptual loss. As the final output from generator usually contains artifacts, we further propose an L_1-regularized gradient prior to remove artifacts while preserving important details.

The contributions of this work are as follows:

- We propose an end-to-end trainable network based on cGAN to solve image dehazing problem.
- To generate much better dehazed results from generator, we develop an encoder and decoder architecture in the generative network so that it can capture more useful information.
- To generate realistic clear images and remove artifacts, we develop a new loss function based on the pre-trained VGG features and an L_1-regularized gradient prior.
- We synthesize a hazy image dataset which includes both indoor and outdoor images and show that our algorithm achieves the state-of-the-art performance on the proposed hazy dataset and real-world images.

Figure 1 shows that our algorithm generates a better clear image than those of the convolutional neural network-based methods [3, 24].

2. Related Work

In this section, we briefly review the most related single image dehazing methods and the applications of the conditional generative adversarial network on image processing.

2.1. Single Image Haze Removal

Single image dehazing methods can be roughly divided into the adaptive color contrast enhancement-based method and the regularization-based method. The adaptive color contrast enhancement-based method usually suffers from visual artifacts such as color blocking and aliasing, which are invisible in the input and obvious in the output [6, 33]. The regularization-based method is mainly based on the physical haze formation model, and kinds of features or image priors are developed to estimate clear images. As the hazy model involves the transmission map and atmospheric light, several methods employ priors on the scene depth. For example, Nishino *et al.* point out that the clear images and the corresponding depths are statistically independent and could be jointly estimated on the basis of priors [21]. The other methods assume that the clear images and the corresponding depths are piece-wise constant and use some priors based on statistical properties of local image patches [5, 7, 11]. The learning-based methods have been developed to estimate the transmission map. Tang *et al*. estimate the the transmission map by learning multi-scale haze relevant features [35].

Motivated by the success of the CNN in object detection, recognition and related tasks [8, 13, 34], CNN has been applied in image dehazing [3, 24]. These methods first estimate the transmission map and then use conventional methods to recover clear images. Thus, if the transmission map was not well estimated, it would affect the clear image estimation. To overcome this problem, Li *et al*. [18] jointly estimate the transmission map and atmospheric light by a CNN. Different from [18], we develop an end-to-end dehazing method based on a cGAN.

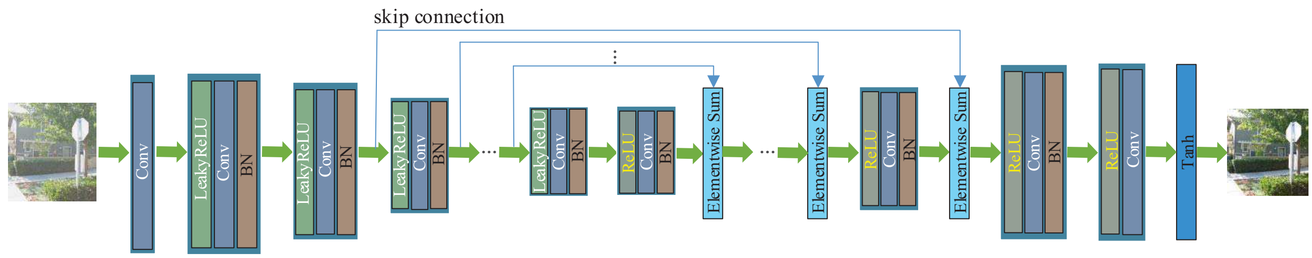
2.2. Conditional Generative Adversarial Network

In [9], Goodfellow *et al.* propose the GAN framework to generate realistic-looking images from random noise via an adversarial learning. However, GAN is not stable in training process and often produces some artifacts such as noise and color shift in the synthesized images. Incorporating conditional information in GAN results in more effective learning [32]. The conditioning variables augmenting information increase the stability of learning process and improve the representation capability of the generator.

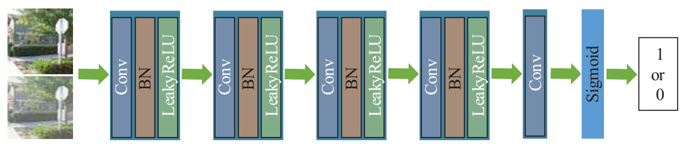
Different from original GAN [9], the cGAN algorithm learns to generate a clear image J from an input image I and random noise z by optimizing the following objective function

$$\min_G \max_D \mathbb{E}_{I,z}[\log(1 - D(I, G(I, z)))] + \mathbb{E}_{I,J}[\log D(I, J)] \quad (2)$$

The cGAN has been made great progress in image processing field such as super-resolution [17], image inpainting [37] and style transfer [15]. Raymond *et al.* [37] propose a semantic image inpainting algorithm using a cGAN. In [23], Isola *et al*. develop a deep architecture and GAN formulation to bridge these advances in text and image modeling translating visual concepts from characters to pixels. The method generates interesting images of flowers and birds by conditioning on text descriptions. In image super-resolution, Ledig *et al*. [17] modify the GAN formulation by introducing pixel-wise content loss and perceptual loss [15] to generate high quality images. Zhang *et al*. [39] use the pixel-wise content loss and perceptual loss in cGAN to solve image deraining problem. Different from these methods, we proposed an effective image dehazing algorithm based on cGAN in this paper.



(a) Generator



(b) Discriminator

Figure 2: The network structure of the proposed method. The generator network contains an encoder and decoder process. The same color used in the rectangles denotes the same operation. "Conv" and "BN" denote the convolution and the batch normalization operations.

3. Proposed Method

In this section, we introduce the architecture of the proposed network including the generator and the discriminator. To generate clear images from hazy inputs, we modify the cGAN framework by a new loss function including adversarial loss, perceptual loss and L_1-regularized gradient prior. In the following, we introduce the proposed network and loss function in details.

3.1. Generator

The function of the generator is to generate a clear image from an input hazy image. Therefore it should not only preserve the structure and detail information of an input image but also remove the haze as much as possible. Motivated by "ResNet" [14] and "U-Net" [25], we introduce skip connections of the symmetric layers to break through the bottleneck of information in decoding process. Instead of simply concatenating all the channels of the symmetric layers, we adopt a summation method to capture more useful information. The difference between concatenation and summation is discussed in section 5.1. As shown in Figure 2(a), the generator contains an encoding process and a decoding process. The encoding process is mainly based on the down-sampling operations and provides feature maps to the symmetric layer of the decoding process. The decoding process mainly uses the up-sampling operations and a non-linear space transfer. The details of the generator structures and parameter settings are shown in Table 1.

3.2. Discriminator

The discriminator is used to distinguish whether an image is real or fake. Similar to the network in [39], we develop a neural network, where the basis operations are convolutional, batch normalization, and LeakyReLU activation. For the final layer of the discriminator, we apply a sigmoid function to the feature maps so that the probability score can be normalized into [0,1]. The architecture of the discriminator is shown in Figure 2(b). The details of the discriminator structures and parameter settings can included in Table 2.

3.3. Loss Function

Let $\{I_i, i = 1, 2, ..., N\}$ and $\{J_i, i = 1, 2, ..., N\}$ denote the hazy images and the corresponding clear images. A straightforward way to train the generative network is to directly utilize the original cGAN formulation in (2) which can be expressed as

$$L_A = \frac{1}{N}\sum_{i=1}^{N}\log(1 - D(I_i, \widetilde{J}_i)), \tag{3}$$

where $\widetilde{J}_i$ is the output of the generator G. However, we find that the cGAN algorithm using this function is not able to remove the haze well and also generates some artifacts and color distortion on generated images. As shown in the following, both the visual results (Figure 4(f) and Figure 6(b)) and the quantitative results (Table 5) indicate that only using (3) does not generate clear images.

In order to recover realistic images, we introduce the perceptual loss based on the pre-trained VGG features to constrain the generator, which is defined as

$$L_P = \frac{1}{N}\sum_{i=1}^{N}\|\mathcal{F}_i(G(I_i)) - \mathcal{F}_i(J_i)\|_2^2 \tag{4}$$

Here, $\mathcal{F}_i$ represents the feature maps of the i-th layer of the VGG network [31] which is pre-trained on ImageNet [26]. The effect of the perceptual loss has been demonstrated in super-resolution, image restoration and other relative fields [2, 15, 17]. Different from these applications, we find that using (4) is able to help the details restoration and haze removal but it accordingly introduces artifacts in the recovered images. This inevitably degrades quality of the recovered images. We will show the effect of this loss function in Section 5.2.

To remove the artifacts and preserve details and structures, we introduce L_1-regularization gradient prior on the output of the generator and content-based pixel-wise loss, which is defined as

$$L_T = \frac{1}{N}\sum_{i=1}^{N}(\|G(I_i) - J_i\|_1 + \lambda\|\nabla G(I_i)\|_1) \tag{5}$$

where $\|\nabla G(I_i)\|_1$ denotes the total variation regularization, $\|G(I_i) - J_i\|_1$ is the content-based pixel-wise loss, and λ is the regularization weight. This loss function is able to remove the artifacts and preserve the details. We will show the effect of this loss function in Section 5.2.

Finally, we combine the adversarial loss, perceptual loss, L_1-regularized gradient prior and content-based pixel-wise loss to regularize the proposed generative network, which is defined as

$$\mathcal{L} = \alpha L_A + \beta L_P + \gamma L_T \tag{6}$$

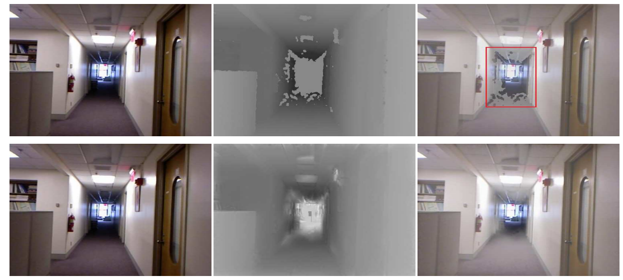
where α, β and γ are the positive weights. The generator G is trained by minimizing (6).

After obtaining the intermediate generator G, we update the discriminator D by

$$\max_D \frac{1}{N}\sum_{i=1}^{N}\Big(\log(1 - D(I_i, \widetilde{J}_i)) + \log(D(I_i, J_i))\Big) \tag{7}$$

4. Experimental Results

In this section, we quantitatively and qualitatively evaluate our method against several state-of-the-art algorithms on synthetic dataset and real-world images. The source code and datasets used in the paper are publicly available at the website: https://github.com/hong-ye/dehaze-cGAN. More experimental results are included in the supplemental material.



(a) Clear image (b) Transmission map (c) Synthesized image

Figure 3: The proposed method for synthesizing hazy images. The first row shows that synthesizing hazy images with original depths will lead to artifacts. The second row shows that using image guided filtering method [12] to remove the holes in depths can generate better hazy images.

4.1. Synthetic Dataset

As there exist few hazy datasets in image dehazing, we synthesize a new dataset including both indoor and outdoor images to train the network. Similar to [24], we use the NYU Depth dataset [30] only including the indoor images. In addition, the Make3D datasets [27, 28, 29] are employed as the outdoor images. We randomly choose 2,400 synthesized images and extract 240 testing images. Given a clear image $\boldsymbol{J}$ and the corresponding ground truth depth d, we synthesize a hazy image $\boldsymbol{I}$ according to (1). We generate the random atmospheric light $A = [n_1, n_2, n_3]$, where $n \in [0.8, 1.0]$, and use the random value $\rho \in [0.8, 1.6]$ for each image. The clean image and the corresponding scene depth are resized to the canonical size of 512×512 pixels before they are fused. However, directly synthesizing hazy images according to (1) usually leads to significant artifacts as there exist holes in the provided depths (Figure 3(b)). In order to remove these artifacts, we use the image guided filtering method [12] (where the clear image is the guidance) to remove the holes in the depth (Figure 3(b)). With the filtered depth, we can generate much better hazy images (Figure 3(c)).

4.2. Experimental Settings

The detail architectures and parameter settings of the proposed network are presented in Table 1. Each layer of the encoding process consists of the convolution, batch normalization and LeakyReLU. Each layer of the decoding process is composed of deconvolution (fractionally-strided convolution [38]), batch normalization and ReLU. The size of the input and output in the generator is set to be $256 \times 256 \times 3$. The size of the input in the discriminator is set to be $256 \times 256 \times 6$ and the size of its output is $256 \times 256 \times 1$. In training process, we empirically set $\alpha = 1$, $\beta = 150$, $\gamma = 150$, $\lambda = 10^{-5}$. The learning rate is set to be 2×10^{-4}. The update ratio of generator G and dis-

Table 1: Architecture of the generator and parameter setting. “conv” denotes the convolution, “uconv” denotes the deconvolution (fractionally-strided convolution [38]), “Tanh” denotes an hyperbolic tangent function.

Generator											
Encoding	Layer	conv	conv	conv	conv	conv	conv	conv	conv	conv	conv
	Kernel Size	5×5	3×3	4×4	4×4	4×4	4×4	4×4	4×4	4×4	4×4
	Stride	1×1	1×1	2×2	2×2	2×2	2×2	2×2	2×2	2×2	2×2
	Pad	2×2	1×1	1×1	1×1	1×1	1×1	1×1	1×1	1×1	1×1
	Channel	64	64	64	128	256	512	1024	1024	1024	1024
Decoding	Layer	uconv	uconv	uconv	uconv	uconv	uconv	uconv	uconv	conv	Tanh
	Kernel Size	4×4	4×4	4×4	4×4	4×4	4×4	4×4	4×4	3×3	-
	Stride	2×2	2×2	2×2	2×2	2×2	2×2	2×2	2×2	1×1	-
	Pad	1×1	1×1	1×1	1×1	1×1	1×1	1×1	1×1	1×1	-
	Channel	1024	1024	1024	512	256	128	64	64	3	-

criminator D is set to be 1. We use the Adam optimization method [16] to train our network. The proposed algorithm is implemented in Torch7 on a computer with a Nvidia Titan-X GPU.

Table 2: Architecture of the discriminator and parameter setting. “Sigmoid” denotes a sigmoid function.

Discriminator						
Layer	conv	conv	conv	conv	conv	Sigmoid
Kernel Size	3×3	3×3	3×3	3×3	3×3	-
Stride	1×1	1×1	1×1	1×1	1×1	-
Pad	1×1	1×1	1×1	1×1	1×1	-
Channel	48	96	192	384	1	-

4.3. Quantitative Evaluation

We evaluate our algorithm on the synthetic dataset and compare it with several state-of-the-art single image dehazing methods using Peak Signal to Noise Ratio (PSNR) and Structural Similarity Index (SSIM). We also retrain the original cGAN with the same parameter settings for fair comparisons. The quantitative evaluation results are shown in Table 3. The proposed method generates the results with higher PSNR and SSIM values than those of other algorithms.

In Figure 4, we show three examples from the synthetic testing dataset for illustration. The dehazing results by He *et al*. [11] have some color distortions and blocking artifacts when the scene objects are similar to the atmospheric light (e.g., the part enclosed in red box on the image from the second row). This is mainly caused by inaccurate transmission maps. The deep learning-based methods by Ren *et al*. [24] and Cai *et al* [3] use CNN to estimate transmission maps, which overcome the limitations of [11] to some extent. Thus, the dehazing results with fewer artifacts than those of He *et al*. [11] on the outdoor dataset. However, there are still some hazy residuals in the estimated images. The method by Li *et al*. [18] jointly estimates the transmission map and the atmospheric light by a CNN. However, the dehazing results still contains some hazy residuals. The dehazing results by the original cGAN contain some artifacts and color distortion. In contrast, the proposed dehazing method generates much clearer images with fewer artifacts and finer details. In addition, compared to the baseline method, i.e., cGAN, the proposed method introduces new loss functions. The results show that the proposed loss function is able to help the image dehazing problem.

Table 3: Quantitative comparisons on the synthetic testing dataset

	He [11]	Ren [24]	Cai [3]	Li [18]	cGAN	Ours
PSNR	32.72	31.11	32.42	31.38	31.09	**33.61**
SSIM	0.814	0.793	0.842	0.814	0.802	**0.915**

4.4. Real Image Haze Removal

Although the proposed network is trained on synthetic haze images, we show that it can be generalized to handle real-world haze images. Figure 5 shows three real hazy images and the corresponding dehazing results generated by state-of-the-art algorithms. Although the dark channel prior-based method [11] is able to remove some haze, it also generates some artifacts as shown in Figure 5(b). The deep learning methods by Cai *et al*. [3] and Ren *et al*. [24] use a CNN to estimate the transmission map and then use the conventional method to recover clear images. However, the dehazing results still contain some artifacts and haze residuals due to the imperfect transmission map estimation. The method proposed by Li *et al*. [18] directly estimate clear images from hazy images. However, the method fails to generate clear images as shown in Figure 5(c).

Different from these methods, the proposed algorithm is based on an end-to-end trainable network which avoids the transmission map estimation and the atmospheric light estimation thus facilitating haze removal. The images generated by the proposed method are much clearer than those of other algorithms as shown in Figure 5(f).

4.5. Run Time

As our network contains twenty layers, a natural question is that whether the proposed algorithm is fast or not. We evaluate the proposed method using the synthetic test

(a) Input PSNR/SSIM | (b) He [11] 35.67/0.766 | (c) Ren [24] 26.91/0.797 | (d) Cai [3] 26.51/0.807 | (e) Li [18] 25.17/0.759 | (f) cGAN 30.35/0.744 | (g) Ours **44.76/0.824** | (h) Ground truth $+\infty/1$

(a) Input PSNR/SSIM | (b) He [11] 46.15/0.853 | (c) Ren [24] 32.55/0.886 | (d) Cai [3] **49.25**/0.812 | (e) Li [18] 47.25/0.899 | (f) cGAN 40.63/0.828 | (g) Ours 39.86/**0.953** | (h) Ground truth $+\infty/1$

(a) Input PSNR/SSIM | (b) He [11] 33.69/0.848 | (c) Ren [24] **34.93**/0.878 | (d) Cai [3] 33.53/0.919 | (e) Li [18] 33.04/0.906 | (f) cGAN 31.78/0.836 | (g) Ours 31.18/**0.950** | (h) Ground truth $+\infty/1$

Figure 4: Three instances of synthetic hazy dataset and haze removal results using several state-of-the-art dehazing methods. Please see the detail differences in rectangles in amplified figures. (a) Synthetic hazy images, (h) Ground truth.

dataset on the a computer (Intel(R) Core(TM) i7-6700 CPU @3.40GHz). We also compare with the state-of-the-art algorithms [39, 18, 24, 3]. The algorithm including thirteen layers [39], a light-weight model AOD-Net [18] and our network are accelerated with a Titan GPU. Table 4 shows the implementation platform and the average run time of several state-of-the-art dehazing methods on synthetic test dataset.

Table 4: Average run time (second) on the synthetic test dataset

Image Size	Ren [24]	Cai [3]	Li [18]	Zhang [39]	Ours
Platform	Matlab	Matlab	Pycaffe	Torch7	Torch7
512×512	1.89	1.78	0.015	0.059	0.052

5. Analysis and Discussions

In this section, we further analyze and discuss the effect of the proposed algorithm including the network architectures and loss functions. We also show the robustness of the proposed algorithm on the image noise. Finally, we discuss the limitations of the proposed methods.

5.1. Effect of the Proposed Network

In the generator, we use the summation method to capture more useful information. To show the effect of concatenation and summation methods, we train the networks with these two methods in the same settings for fair comparisons. Figure 7 shows the quantitative comparisons of concatenation and summation methods. Although the methods with concatenation and summation strategies tend to generate the similar results, the maximum PSNR and SSIM values (33.61dB, 0.9152) of summation strategy are larger than those (33.45dB, 0.9092) of the concatenation strategy. Therefore we use the summation strategy in the generator.

We also note that several methods develop GANs [39] to solve image deraining and image super-resolution [36]. For fair comparisons, we retrained the model by [39] using the same dataset and parameter settings.

5.2. Effect of Loss Functions

To generate high quality dehazing images, we propose a loss function which includes several terms. In order to evaluate the effect of the loss function, we show the effect of each term in Table 5. The quantitative evaluations are conducted on the proposed synthetic test dataset with the same settings. For simplicity, we denote L_1 as the term of L_T only using the first term.

We note that the method with L_T loss generates the results with higher PSNR values compared to the method with L_1, which indicates the effectiveness of the L_1-regularized gradient prior in image dehazing. The results from the first column and the second column show that L_P helps to improve the SSIM value indicating that it is able to preserve the structures of images.

Figure 6 shows some dehazing results examples with different loss functions and corresponding quantitative results. The method with the proposed loss function generates better results. The GAN method by [39] is able to remove some

(a) Nature images (b) He [11] (c) Li [18] (d) Ren [24] (e) Cai [3] (f) Ours

Figure 5: Real world hazy images and corresponding dehazing results from several state-of-the-art methods. The second row, the fourth row and the sixth row are the close-ups from the first row, the third row and the fifth row in red rectangles, respectively.

Table 5: Quantitatively evaluate the effect of the different loss functions in the proposed method

Loss	L_A	$L_A + L_P$	$L_A + L_1$	$L_A + L_T$	$L_A + L_1 + L_P$	$L_A + L_T + L_P$
PSNR	31.08	31.19	31.76	33.81	32.82	33.61
SSIM	0.761	0.8966	0.887	0.883	0.908	0.915

hazy, but the recovered images still contain haze residuals and some artifacts, as shown in supplementary material.

5.3. Robustness to Image Noise

The proposed method is robust to image noise. In order to evaluate the robustness of the proposed method, we add random noise with noise level from 0.5% to 3% to all test samples. Figure 8 shows quantification results of several state-of-the-art methods on the synthetic test dataset. Our method performs well even when the noise level increases.

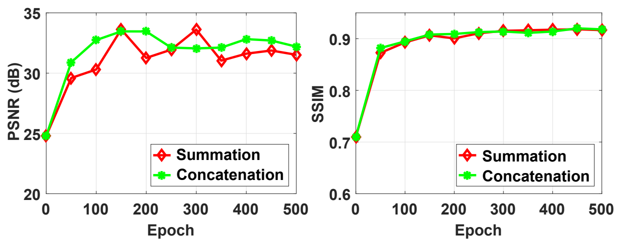
5.4. Limitations

The proposed method learns the mapping functions from hazy images to corresponding clear images and is trained based on the synthetic dataset. However, if the hazy model does not hold for hazy images, the proposed method will

(a) Input psnr/ssim | (b) 27.87/0.867 | (c) 29.35/0.936 | (d) 31.65/0.913 | (e) 31.89/0.902 | (f) 31.21/0.9306 | (g) **32.47/0.939** | (h) Ground truth $+\infty/1$

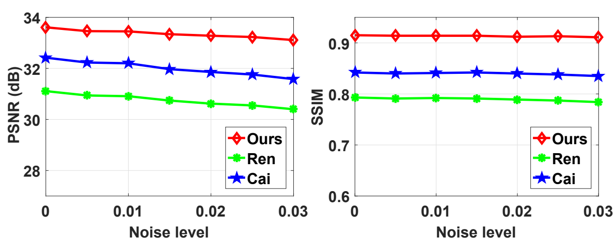
(a) Input psnr/ssim | (b) 30.95/0.697 | (c) 31.88/0.940 | (d) 29.48/0.908 | (e) 33.81/0.896 | (f) 35.31/0.924 | (g) **35.55/0.945** | (h) Ground truth $+\infty/1$

Figure 6: The effect of the proposed network with different loss functions. (b) L_A loss. (c) $L_A + L_P$ loss. (d) $L_A + L_1$ loss. (e) $L_A + L_T$ loss. (f) $L_A + L_1 + L_P$ loss. (g) $L_A + L_T + L_P$ loss.



(a) PSNR (b) SSIM

Figure 7: Quantitative comparisons of concatenation and summation methods.



(a) PSNR (b) SSIM

Figure 8: Quantitative evaluations of several dehazing methods on the test dataset with different noise level.

not be able to generate clear images. Figure 9 shows that the proposed method does not work well on light hazy images and night hazy images. This is probably because our training dataset does not include similar samples. Therefore the hazy model can not learn the corresponding mapping function. We will solve these problems by dedicating to collect more comprehensive haze samples and optimize the model.

6. Conclusion

In this paper, we adopt a conditional generative adversarial network for single image haze removal. The proposed network is trained in an end-to-end manner, which avoids to estimate the transmission map and atmospheric light separately. To generate better results, we have proposed an encoder and decoder architecture so that it can capture more useful information. We further modify the basic cGAN formulation by introducing new loss functions to generate realistic clear images. We also synthesize a hazy data including indoor and outdoor scenes to train and evaluate the proposed algorithm. The proposed method performs favorably against several state-of-the-art methods on both synthetic dataset and real world hazy images.

(a) Light hazy image (b) Output

(a) Night hazy image (b) Output

Figure 9: The proposed method does not work well when the hazy images cannot be modeled by the hazy model.

Acknowledgements. This work has been partially supported by the 973 Program (No. 2014CB347600), the National Natural Science Foundation of China (No. 61522203, 61732007 and 61772275), the Natural Science Foundation of Jiangsu Province (No. BK20170033), and the National Ten Thousand Talent Program of China (Young Top-Notch Talent).

References

[1] C. O. Ancuti and C. Ancuti. Single image dehazing by multi-scale fusion. *IEEE Transactions on Image Processing*, 22(8):3271–3282, 2013. 1

[2] J. Bruna, P. Sprechmann, and Y. Lecun. Super-resolution with deep convolutional sufficient statistics. In *International Conference on Learning Representations (ICLR)*, 2016. 4

[3] B. Cai, X. Xu, K. Jia, C. Qing, and D. Tao. Dehazenet: An end-to-end system for single image haze removal. *IEEE Transactions on Image Processing*, 25(11):5187–5198, 2016. 1, 2, 5, 6, 7

[4] C. Chen, M. N. Do, and J. Wang. Robust image and video dehazing with visual artifact suppression via gradient residual minimization. In *European Conference on Computer Vision (ECCV)*, pages 576–591, 2016. 1

[5] R. Fattal. Dehazing using color-lines. *ACM Trans. Graph.*, 34(1):13:1–13:14, 2014. 2

[6] A. Galdran, J. Vazquez-Corral, D. Pardo, and M. Bertalm. Enhanced variational image dehazing. *SIAM Journal on Imaging Sciences*, 8(3):1519–1546, 2015. 2

[7] K. B. Gibson and T. Q. Nguyen. An analysis of single image defogging methods using a color ellipsoid framework. *EURASIP Journal on Image and Video Processing*, 2013(1):37, 2013. 2

[8] R. B. Girshick. Fast R-CNN. In *IEEE International Conference on Computer Vision (ICCV)*, pages 1440–1448, 2015. 2

[9] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. In *Neural Information Processing Systems (NIPS)*, pages 2672–2680. 2014. 2

[10] N. Hautiere, J. P. Tarel, and D. Aubert. Towards fog-free in-vehicle vision systems through contrast restoration. In *Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, 2007. 1

[11] K. He, J. Sun, and X. Tang. Single image haze removal using dark channel prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(12):2341–2353, 2011. 1, 2, 5, 6, 7

[12] K. He, J. Sun, and X. Tang. Guided image filtering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(6):1397–1409, 2013. 4

[13] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016. 2

[14] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016. 3

[15] J. Johnson, A. Alahi, and L. Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *European Conference on Computer Vision (ECCV)*, pages 694–711, 2016. 2, 4

[16] D. P. Kingma and J. L. Ba. Adam: A method for stochastic optimization. In *International Conference on Learning Representations (ICLR)*, 2015. 5

[17] C. Ledig, L. Theis, F. Huszr, 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 *Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 105–114, 2017. 2, 4

[18] B. Li, X. Peng, Z. Wang, J. Xu, and D. Feng. Aod-net: All-in-one dehazing network. In *International Conference on Computer Vision (ICCV)*, pages 4780–4788, 2017. 2, 5, 6, 7

[19] Z. Li, P. Tan, R. T. Tan, D. Zou, S. Zhiying Zhou, and L.-F. Cheong. Simultaneous video defogging and stereo reconstruction. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4988–4997, 2015. 1

[20] G. Meng, Y. Wang, J. Duan, S. Xiang, and C. Pan. Efficient image dehazing with boundary constraint and contextual regularization. In *International Conference on Computer Vision (ICCV)*, pages 617–624, 2013. 1

[21] K. Nishino, L. Kratz, and S. Lombardi. Bayesian defogging. *International Journal of Computer Vision*, 98(3):263–278, 2012. 2

[22] S. C. Pei and T. Y. Lee. Nighttime haze removal using color transfer pre-processing and dark channel prior. In *International Conference on Image Processing (ICIP)*, pages 957–960, 2012. 1

[23] S. Reed, Z. Akata, X. Yan, L. Logeswaran, B. Schiele, and H. Lee. Generative adversarial text to image synthesis. In *International Conference on International Conference on Machine Learning (ICML)*, pages 1060–1069, 2016. 2

[24] W. Ren, S. Liu, H. Zhang, J. Pan, X. Cao, and M.-H. Yang. Single image dehazing via multi-scale convolutional neural networks. In *European Conference on Computer Vision (ECCV)*, pages 154–169, 2016. 1, 2, 4, 5, 6, 7

[25] O. Ronneberger, P.Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 234–241, 2015. 3

[26] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. S. Bernstein, et al. Imagenet large scale visual recognition challenge. *International Journal of Computer Vision*, 115(3):211–252, 2015. 4

[27] A. Saxena, S. H. Chung, and A. Y. Ng. Learning depth from single monocular images. In *Neural Information Processing Systems (NIPS)*. 2006. 4

[28] A. Saxena, S. H. Chung, and A. Y. Ng. 3-d depth reconstruction from a single still image. *International Journal of Computer Vision*, 76(1):53–69, 2008. 4

[29] A. Saxena, M. Sun, and A. Y. Ng. Make3d: Learning 3d scene structure from a single still image. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(5):824–840, 2009. 4

[30] N. Silberman, D. Hoiem, P. Kohli, and R. Fergus. Indoor segmentation and support inference from rgbd images. In *European Conference on Computer Vision (ECCV)*, pages 746–760, 2012. 4

[31] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. In *International Conference on Learning Representations (ICLR)*, 2015. 4

[32] K. Sohn, H. Lee, and X. Yan. Learning structured output representation using deep conditional generative models. In *Neural Information Processing Systems (NIPS)*, pages 3483–3491. 2015. 2

[33] R. T. Tan. Visibility in bad weather from a single image. In *Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, 2008. 1, 2

[34] J. Tang, X. Shu, Z. Li, G. Qi, and J. Wang. Generalized deep transfer networks for knowledge propagation in heterogeneous domains. *ACM Transactions on Multimedia Computing*, 12(4s):68:1–68:22, 2016. 2

[35] K. Tang, J. Yang, and J. Wang. Investigating haze-relevant features in a learning framework for image dehazing. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2995–3002, 2014. 2

[36] J. P. Y. Z. H. P. M.-H. Y. Xiangyu Xu, Deqing Sun. Learning to super-resolve blurry face and text images. In *International Conference on Computer Vision (ICCV)*, pages 251–260, 2017. 6

[37] R. A. Yeh, C. Chen, T. Yian Lim, A. G. Schwing, M. Hasegawa-Johnson, and M. N. Do. Semantic image inpainting with deep generative models. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6882–6890, 2017. 2

[38] M. D. Zeiler, G. W. Taylor, and R. Fergus. Adaptive deconvolutional networks for mid and high level feature learning. In *International Conference on Computer Vision (ICCV)*, pages 2018–2025, 2011. 4, 5

[39] H. Zhang, V. Sindagi, and V. M. Patel. Image de-raining using a conditional generative adversarial network. *arXiv preprint arXiv:1701.05957*, 2017. 2, 3, 6

[40] Q. Zhu, J. Mai, and L. Shao. A fast single image haze removal algorithm using color attenuation prior. *IEEE Transactions on Image Processing*, 24(11):3522–3533, 2015. 1