Enhanced Pix2pix Dehazing Network

Yanyun Qu[1*] Yizi Chen[1*] Jingying Huang[1] Yuan Xie[2,3†]
[1]Fujian Key Laboratory of Sensing and Computing for Smart City,
School of Information Science and Engineering, Xiamen University, Fujian, China
[2]School of Computer Science and Software Engineering,
East China Normal University, Shanghai, China
[3]Institute of Advanced Artificial Intelligence in Nanjing, Horizon Robotic, Nanjing, China
yyqu@xmu.edu.cn, atavism@msn.cn, jyhuang33@outlook.com, yxie@sei.ecnu.edu.cn

Abstract

In this paper, we reduce the image dehazing problem to an image-to-image translation problem, and propose Enhanced Pix2pix Dehazing Network (EPDN), which generates a haze-free image without relying on the physical scattering model. EPDN is embedded by a generative adversarial network, which is followed by a well-designed enhancer. Inspired by visual perception global-first [5] theory, the discriminator guides the generator to create a pseudo realistic image on a coarse scale, while the enhancer following the generator is required to produce a realistic dehazing image on the fine scale. The enhancer contains two enhancing blocks based on the receptive field model, which reinforces the dehazing effect in both color and details. The embedded GAN is jointly trained with the enhancer. Extensive experiment results on synthetic datasets and real-world datasets show that the proposed EPDN is superior to the state-of-the-art methods in terms of PSNR, SSIM, PI, and subjective visual effect.

1. Introduction

Haze is a typical atmospheric phenomenon, and it causes color distortion, blurring and low contrast for the photographed image, which results in the difficulties of subsequent tasks, such as object recognition and image understanding. Thus, more and more attentions are attracted to image dehazing.

Most of the successful approaches depend on the physical scattering model [12] [14], which is formulated as

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

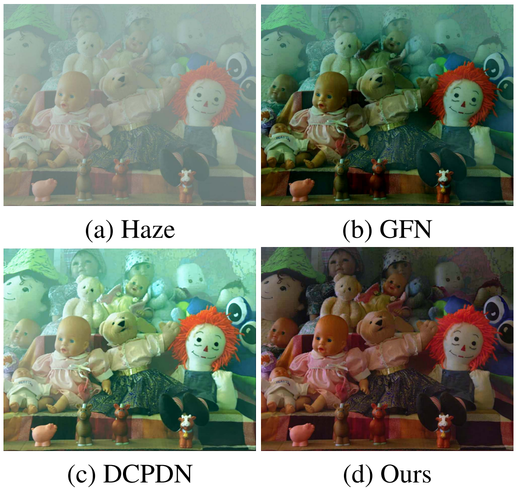
(a) Haze (b) GFN (c) DCPDN (d) Ours

Figure 1. A single image haze removal example. Our method produces a haze-free image with faithful color and rich details compared with GFN [16] and DCPDN [21].

where I is the observed hazy image, J is the scene radiance, t is the transmission map, A is the atmospheric light and z is the pixel location. The solution of the haze-free image depends on the estimation of the atmospheric light and the transmission map. Early dehazing methods are mostly prior-based methods, e.g. DCP [7], which estimates the transmission map by investigating the dark channel prior. These prior-based methods can achieve good dehazing effect to a certain extent. However, the prior may be easily violated in practice, which leads to an inaccurate estimation of transmission map so that the quality of the dehazing image is not desirable. With the rising up of deep learning, the estimations of the transmission map or the atmospheric light are estimated by the convolutional neural network rather than relying on priors. Some methods utilize the deep convolutional neural network to estimate the transmission map [3] [15], some employ the deep convolutional neural network to jointly estimate the atmospheric light and

[*]Equal contribution
[†]Corresponding author

the transmission map [20] [23]. Either early dehazing methods or exsiting deep learning based ones almost depend on the physical scattering model, and the estimation accuracies of the atmospheric light and the transmission map greatly influence the quality of the dehazing image.

In order to disentangle image dehazing from the physical scattering model, we try to transform a hazy image to a haze-free image pixel by pixel directly. Motivated by the success of generative adversarial networks (GANs) in image-to-image translation [9] [19] [24], we marry GAN to image dehazing. However, GANs for image-to-image translation can not be directly applied to image dehazing because image haze is the depth-dependent noise and nonuniform. Directly application will produce undesirable results which are overcolored and lack of details. It is known as the visual perception global-first theory [5], an object or scene is discriminated only in a global way, not necessarily depending on the details of realistic images, while the creation of a realistic image must be dependent on details as more as possible.

In this paper, we propose Enhanced Pix2pix Dehazing Network (EPDN). EPDN includes three parts: the discriminator, the generator, and the enhancer. The GAN module with generator and discriminator is embedded in EPDN, in which the discriminator just supervises the intermediate result of EPDN. The enhancer following the generator will reinforce the output of the GAN, which is designed according to the receptive field model. To the best of our knowledge, EPDN is the first work to embed GAN for image dehazing according to the visual perception theory. Moreover, regarding the proposed architecture, we develop a joint training scheme which alternatively optimizes the embedding GAN (generator and discriminators) and the generator along with the enhancer.

The proposed method can generate a more realistic image in terms of color and details. Fig. 1 shows the visual effect of our method. Compared with GFN [16] and DCPDN [21], our method achieves more realistic dehazing effect with faithful color and structures. Moreover, we introduce the Perceptual Index (PI) to evaluate the performance of image dehazing including Peak Signal to Noise Ratio (PSNR) and Structural Similarity (SSIM). As we know, visual effect is a subject estimation, which is not convenient for computation. PI is computable by simulating the visual perception.

The contributions of our work are as follow:

1. EPDN is proposed for image dehazing, which does not rely on the physical scattering model, while adopts the way of image-to-image translation alternatively.

2. Inspired by the global-first property of visual perception, the embedded GAN and enhancer are designed to produce a perceptually pleasing images with more details.

3. A joint training scheme is developed for updating the embedded GAN and enhancer through reasonably combining four kinds of loss functions.

4. The Perceptual Index (PI) is introduced for quantitative evalution from the perceptual perspective. In addition, extensive experiments on both synthetic datasets and real-world dataset indicate that EPDN performs favourably against the state-of-the-art methods. Especially, our results are outstanding in visual perception.

2. Related Work

Our work is related to two topics: single image dehazing and generative adversarial networks which are briefly discussed in this section.

Single image dehazing. Most of the existing dehazing methods depend on the physical scattering model [12] [14] Eq. (1), which are divided into two classes: the prior based methods and the learning-based methods. The physical model contains two important factors: the transmission map and the atmospheric light. Efforts are made to estimate the two factors for a solution of haze removal.

Prior-based dehazing. Tan *et al.* [17] made a model to maximize the contrast of an image for image dehazing because it is observed that haze-free images have higher contrast than hazy images. He *et al.* [7] [8] proposed a dark channel prior for the estimation of the transmission map. Zhu *et al.* [25] recovered depth information via color atenuation prior. Tang *et al.* [18] systematically investigated a variety of haze-relevant priors in a regression framework to learn the best prior combination for image dehazing. A haze line prior is proposed by Dana Berman *et al.* [1], which assumes that colors of a haze-free image are well approximated by a few hundred distinct colors. Though the prior-based dehazing methods have achieved promising results, the prior is not robust to the unconstraint environment in the wild, thus, the dehazing performance is not always desirable.

Learning-based dehazing. Different from the prior-based methods, the learning-based methods directly estimate the transmission map or atmospheric light rather than relying on the priors. Cai *et al.* [3] proposed an end-to-end dehazing model based on convolutional neural network (CNN) named DehazeNet, which estimates the transmission map. Ren *et al.* [15] proposed a multi-scale deep model to estimate the transmission map. Li *et al.* [10] reformulated the physical scattering model, and design AOD-Net to learn a mapping function based on CNN. Ren *et al.* [16] used an encoder-decoder network and adopted a novel fusion-based strategy to directly restore a clear image from a hazy image.

Generative Adversarial Networks (GANs). Recently, great progress is made in GAN [6]. GAN includes two parts: the discriminator and the generator. They are trained simultaneously in an adversarial way so that the generator could produce a realistic image which confuses the discriminator. GAN is widely used in many computer vision applications. Especially, GAN has achieved promising results in image synthesis [9] [19] [24]. Inspired by the success of GAN, we utilize it for image dehazing. DCPDN [21] implements GAN on image dehazing which learns transmission map and atmospheric light simultaneously in the generators by optimizing the final dehazing performance for haze-free images. Yang *et al.* [20] proposed the disentangled dehazing network, which uses unpaired supervision. The GAN proposed by Yang *et al.* [20] contains three generators: the generator for the haze-free image, the generator for the atmospheric light, and the generator for transmission map. DehazeGAN [23] draws lessons from the differential programming to use GAN for simultaneous estimations of the atmospheric light and the transmission map. The marriage of GAN and image dehazing is still in the beginning. The current dehazing methods via GAN all depend on the physical scattering model. Until now, little is discussed how to deal with image dehazing independent of the physical scattering model. As discussed in Introduction, it is meaningful to investigate a model-free dehazing method via GAN.

3. Proposed Method

3.1. The Architecture of EPDN

In this paper, we cast the single image dehazing problem as a task of image-to-image translation. Hazy images and haze-free images are regarded as two different image styles. The framework of EPDN is shown in Fig. 2, which consists of a multi-resolution generator module, the enhancer module, and the multi-scale discriminator module. The GAN architecture similar to pix2pixHD [19] is embedded in EPDN, followed by the enhancer. The enhancer contains two well-designed enhancing blocks, each of which is built depending on the receptive field model. And a shot-cut is employed to maintain the color information of original images. In the following, we detail the architecture of EPDN.

Multi-resolution generator. The multi-resolution generator of GAN module consists of global sub-generator G_1 and a local sub-generator G_2, as shown in Fig. 2. Both G_1 and G_2 include a convolutional front-end, three residual blocks, and a transposed convolutional back-end. The input of G_1 is $2\times$ downsampled from the original hazy images. G_1 is embedded in G_2, and the element-wise sum of the output of G_1 and the feature maps obtained by the convolutional front-end of G_2 is fed into the residual block of G_2. The multi-resolution structure has been proven successful in image-to-image translation. The global sub-generator creates an image on a coarse scale, while the local sub-generator creates an image on a fine scale. And the combination of the two sub-generators produce an image from coarse-to-fine.

Multi-scale discriminator. The embedded GAN module contains a multi-scale discriminator module which contains two-scale discriminators named D_1 and D_2. D_1 and D_2 have the same architecture, and the input of D_2 is $2\times$ downsampled from the input of D_1. The output of the generator is fed into D_1. The multi-scale discriminators could guide the generator from coarse-to-fine. On the one hand, D_2 guides the generator to produce a global pseudo realistic image on a coarse scale. On the other hand, D_1 guides the generator on a fine scale.

Enhancing block Even though pix2pixHD utilizes the coarse-to-fine feature, the results obtained from only pix2pixHD still lack details and are overcolored. One possible reason is that the existing discriminator is limited in guiding the generator to create realistic details. In other words, the discriminator should merely direct the generator to restore structure imfromation rather than details.

To efficiently solve this problem, a pyramid pooling block [21] [22] is implemented to make sure the details of features from different scales are embedded in the final result. We name it enhancing block. Drawing lesson from global context information in object recognition, details of features in various scales are needed. Thus, the enhancing block is designed according to the recepive field model which can extract information on different scales. The enhancing block is shown in Fig. 3. In detail, there are two 3×3 front-end convolution layers in the enhancing block. The output of the front-end convolution layer is downsampled by factors of $4\times$, $8\times$, $16\times$, $32\times$ to build a four-scale pyramid. Feature maps on different scales provide different receptive fields, which helps to reconstruct an image on various scales. And then, 1×1 convolution is implemented for dimension reduction. Actually, 1×1 convolution implies the attention mechanism which weights the channel adaptively. After that, we upsample the feature maps to the original size and concatenate them together with the output of the front-end convolution layer. Finally, the 3×3 convolution is implemented on the concatenation of the feature maps.

In EPDN, the enhancer includes two enhancing blocks. Moreover, the first enhancing block is fed by the concatenation of the original image and the feature maps of the generator which are also fed to the second enhancing block.

3.2. Overall Loss Function

In order to optimize EPDN, we utilize four loss functions as Eq. (2), the adversarial loss L_A, the feature matching loss L_{FM}, the perceptual loss L_{VGG}, and the fidelity loss L_F. The adversarial loss together with the feature matching loss is used to make the GAN module learn the global information and recover the original image structure by using multi-

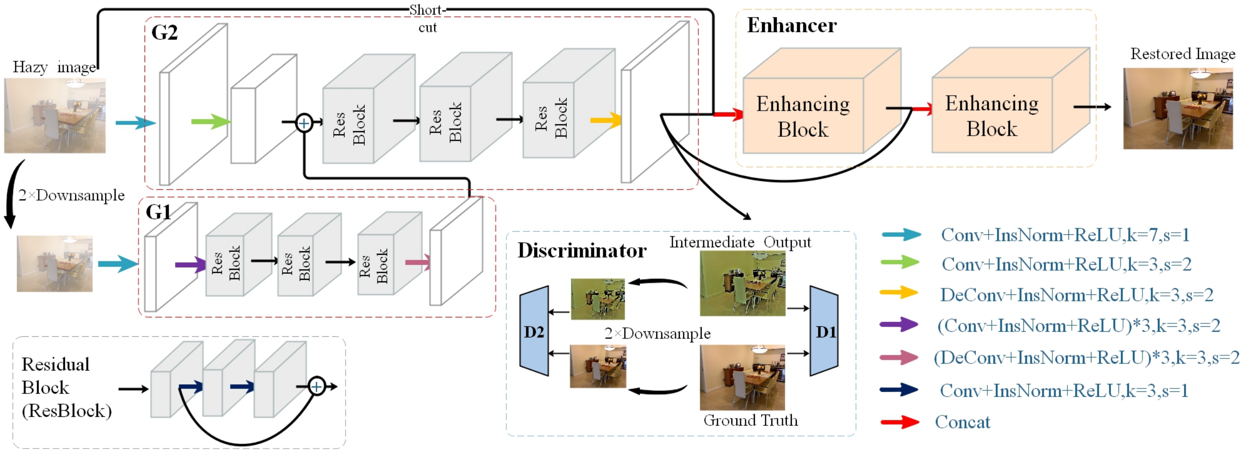

Figure 2. The architecture of EPDN. EPDN includes three parts: the multi-resolution generator, i.e. G_1 and G_2, the muti-scale discriminator, i.e. D_1 and D_2, and the enhancer.

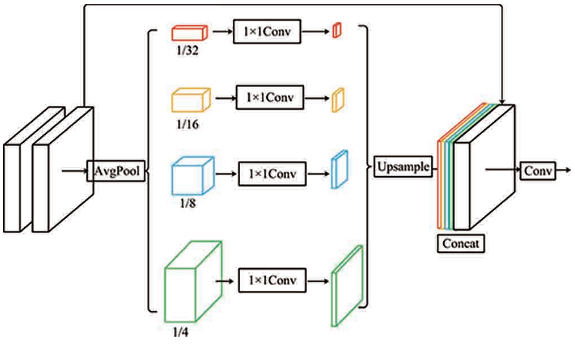

Figure 3. The structure of enhancing block

scale features. Perceptual loss and fidelity loss are used to reinforce the fine features and preserve original color information. To simplify the model, thes coefficients of the feature matching loss and perceptual loss are set to be the same.

$$L_{EP} = L_A + \lambda L_{FM} + \lambda L_{VGG} + L_F. \tag{2}$$

Adversarial loss. We adopt the adversarial loss of GAN. The generator is initialized to translate a hazy image to the haze-free image, while the discriminator aims to distinguish whether an image is real or fake. Considering there are the two-scale discriminators D_1, D_2, the adversarial loss is formulated as a multi-task learning loss

$$L_A = \min_{\widetilde{G}}[\max_{D_1,D_2} \sum_{k=1,2} \ell_A(\widetilde{G}, D_k)], \tag{3}$$

where $\ell_A(\widetilde{G}, D_k)$ is the single adversarial loss of the k-th discriminator D_k, it is formulated as,

$$\ell_A(\widetilde{G}, D_k) = E_{(X)}[\log D_k(X)] + E_{(X)}[\log(1 - D_k(\widetilde{G}(\hat{X})))], \tag{4}$$

and X and $\hat{X}$ denote real haze-free images and hazy images. $\widetilde{G}(\hat{X})$ represents the output produced by the generator of the GAN module, but not the finally result of EPDN.

Feature matching loss. In order to make a realistic image, the adversarial loss is improved by incorporating a feature matching loss based on the discriminator. We use this loss to make the generator produce natural multi-scale statistical information. The intermediate feature maps are learned to match between the real and the synthesized image. The feature matching loss function is formulated as

$$L_{FM} = \min_{\widetilde{G}}[\sum_{k=1,2} \ell_{FM}(\widetilde{G}, D_k)]. \tag{5}$$

$\ell_{FM}(\widetilde{G}, D_k)$ is the feature matching loss with the k-th discriminator D_k.

$$\ell_{FM}(\widetilde{G}, D_k) = E_{(X)} \sum_{i=1}^{T} \frac{1}{N_i} [\|D_k^{(i)}(X) - D_k^{(i)}(\widetilde{G}(\hat{X}))\|_1], \tag{6}$$

where T is a total number of layers used for feature extraction, N_i is a number of elements in each layer, $D_k^{(i)}$ is the operator of the feature extraction of the i-th layer in D_k.

Perceptual loss. In order to keep the perceptual and semantic fidelity, we use perceptual loss function to measure high-level difference between the hazy image and its counterpart dehazing image. Based on a pre-trained VGGNet for image classification, we extract the activations of the i-th layers of VGGNet, denoted by $\phi_i()$, which are treated as the perceptual feature. We use the pixel-wise distance to measure the difference between the perceptual features of the hazy and dehazing image. The perceptual loss function is as follow

$$L_{VGG}^{\phi,i}(\hat{Y}, X) = \frac{1}{C_i H_i W_i} \|\phi_i(\hat{Y}) - \phi_i(X)\|_1, \quad (7)$$

where $\hat{Y}$ is the final result of EPDN. H_i and W_i are the height and width of the i-th feature map, and C_i indicates the channel.

Fidelity loss. The Euclidean distance between the haze-free image X and final output $\hat{Y}$ is regarded as the fidelity loss, which is defined as

$$L_F = \|X - \hat{Y}\|_2. \quad (8)$$

3.3. Training

Algorithm 1: EPDN training algorithm

```
Input:
$n_b \leftarrow$ the batch size;
$\lambda \leftarrow$ the hyper-parameter;
for $num = 1; num \leq trainingiterations$ do
    Sample hazy examples $\hat{X} = \{\hat{x}^{(1)}, ..., \hat{x}^{(n_b)}\}$;
    Sample clean examples $X = \{x^{(1)}, ..., x^{(n_b)}\}$;
    $M \leftarrow \widetilde{G}(\hat{X})$, the output of the muti-resolution generator;
    $Y \leftarrow EP(\hat{X})$, the output of EPDN;
    $M_k \leftarrow 2^{k-1}$ time $downsample(M)$;
    $X_k \leftarrow 2^{k-1}$ time $downsample(X)$;
    for $k = 1, 2$ do
        Update the discriminators $D_k$ by ascending the gradient of Eq. (3);
    Update the multi-resolution generator($\widetilde{G}$) by descending the gradient of the sum of Eq. (3) and Eq. (5);
    Update $\widetilde{G}$ and enhancer by descending the gradient of the sum of Eq. (7) and Eq. (8);
```

Because GAN is only a part of the whole architecture of EPDN, we cannot implement the training scheme of GAN directly on EPDN. We develop a new training scheme. We adopt the alternative iteration algorithm. Firstly, the GAN architecture is optimized with the adversarial loss function Eq. (3) and the feature matching loss function Eq. (5). In detail, we first update the multi-scale discriminator by ascending its gradient and then update the multi-resolution generator by descending its gradient. Secondly, the enhancer and the multi-resolution generator is optimized by descending the gradient of the sum of perceptual loss Eq. (7) and the fidelity loss Eq. (8). We summarize the algorithm as Algorithm 1. Different from the original GAN training scheme, our generator is updated twice respectively with the discriminator and the enchancer, which satisfies the global-first theory.

Table 1. Ablation study settings.

Method	Enhancing block	short-cut	Embedded
GAN+E	1	-	✓
GAN+E+S	1	✓	✓
GAN+EE	2	-	✓
GAN$^+$	2	✓	-
GAN	0	-	-
Ours	2	✓	✓

Table 2. Comparison of variants with different components on the outdoor dataset of SOTS.

Method	PSNR	SSIM	PI
GAN+E	20.56	0.7553	3.2394
GAN+E+S	18.66	0.7636	2.2374
GAN+EE	21.47	0.7992	3.1153
GAN$^+$	21.73	0.8716	2.556
GAN	20.78	0.7455	2.7397
Ours	22.57	0.8630	2.3858

4. Experiments

In this section, we implement the proposed method on both the synthesis dataset and the real-world dataset to demonstrate the effectiveness of the proposed method. We compare our proposed method with five state-of-the-art methods: DCP [7] (He CVPR'09), DehazeNet [3] (Cai TIP'16), AOD-Net [10](Li ICCV'17), GFN [16] (Ren CVPR'18), and DCPDN [21] (Zhang CVPR'18). For the fairness of comparison, the source codes of the compared methods are presented publicly by the authors. In addition, we do an ablation study to demonstrate the effectiveness of our embedding GAN and the enhancing block.

4.1. Experiment Settings

Dataset. RESIDE [11] is a new large-scale hazy image dataset and it consists of five subsets: Indoor Training Set (ITS), Outdoor Training Set (OTS), Synthetic Objective Testing Set (SOTS), Real World task-driven Testing Tet (RTTS), and Hybrid Subjective Testing Set (HSTS). Among the five subsets, ITS, OTS, SOTS are synthetic datasets, RTTS is the real-world dataset, both synthetic data and real-word hazy data are involved in HSTS. On the one hand, ITS and SOTS which contain both indoor and outdoor hazy images are respectively employed for our training and testing. On the other hand, we collect the real-world images used by

Table 3. Comparison results of the state-of-the-art dehazing methods on SOTS.

Method		DCP [7]	DehazeNet [3]	AOD-NET [10]	DCPDN [21]	GFN [16]	Ours
indoor	PSNR	16.62	21.14	19.06	15.85	22.30	25.06
	SSIM	0.8179	0.8472	0.8504	0.8175	0.8800	0.9232
	PI	3.9535	4.0458	3.6961	4.7485	4.1146	4.0620
outdoor	PSNR	19.13	22.46	20.29	19.93	21.55	22.57
	SSIM	0.8148	0.8514	0.8765	0.8449	0.8444	0.8630
	PI	2.5061	2.4346	2.4280	2.7269	2.1608	2.3858

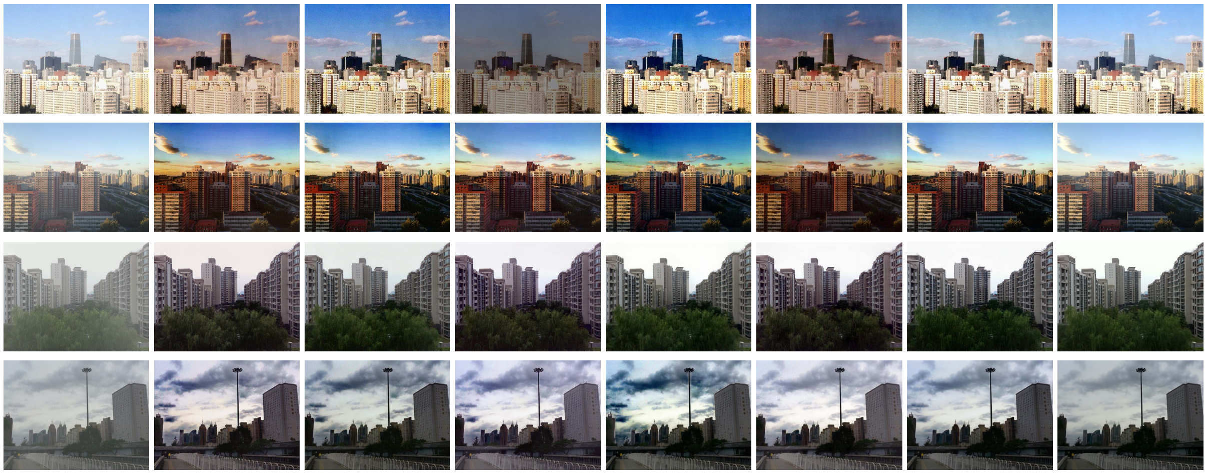

Figure 4. Comparison results of variants with different components in visual effect on outdoor images.

previous methods, and compare our method with the state-of-the-art methods on this dataset.

Training Details. During training, ITS is used as the training dataset which is also used as the training dataset for the compared methods. We adopt Adam optimizer with a batch size of 1, and set a learning rate as 0.0002, the exponential decay rates as $(\beta_1, \beta_2) = (0.6, 0.999)$. The hyper-parameter of loss function is set as $\lambda = 10$. We implement our model with the PyTorch framework and a TITAN GPU.

Quality Measures. To evaluate the performance of our method, we adopt three metrics: the Peak Signal to Noise Ratio (PSNR), the Structural Similarity index (SSIM) and Perceptual Index (PI). As we know, image qualification evaluation is very important for image restoration. It includes the objective measurement and subjective measurement. For the former, PSNR and SSIM are usually used in image dehazing. For the latter, visual effect is used to evaluate the image dehazing performance. However, it is not convenient to use for image qualification evaluation. PI is a new criterion which bridges the visual effect with computable index. And it has been recognized to be effective in image super-resolution [2]. In the experiment, we use PI to evaluate the performance of image dehazing. The lower the image quality is, the higher PI is. PI is formulated as

$$PI = \frac{1}{2}((10 - Ma) + NIQE),$$

where Ma and $NIQE$ are two image qualification indexes which are detailed in [4] and [13].

4.2. Ablation Study

To better demonstrate the effectiveness of the architecture of our method, we conduct an ablation study by considering the combination of four factors: GAN, one enhancing block, two enhancing blocks, and the short-cut skip. We construct the following variants with different component combinations: 1) **GAN**: only pix2pixHD [19] is used; 2) **GAN+E**: only one enhancing block follows the embedding pix2pixHD; 3) **GAN+E+S**: the variant GAN+E combines a short-cut skip which connects the original image to the enhancer; 4) **GAN+EE**: two enhancing blocks follow the embedding pix2pixHD; 5) **GAN$^+$**: the whole architecture is a GAN, in which the the whole generator is the combination of the generator of pix2pixHD and two enhancing blocks and the short-cup skip connects the original image to the first enhancing block. The ablation configurations are given in Table. 1.

We compare EPDN with five variants with different components on the outdoor dataset of SOTS. The results are shown in Table. 2 and Fig. 4. It demonstrates that the proposed EPDN achieves the best performance of image dehazing in PSNR and SSIM and the best visual effects. Compared with pix2pixHD [19], the improvement gains in

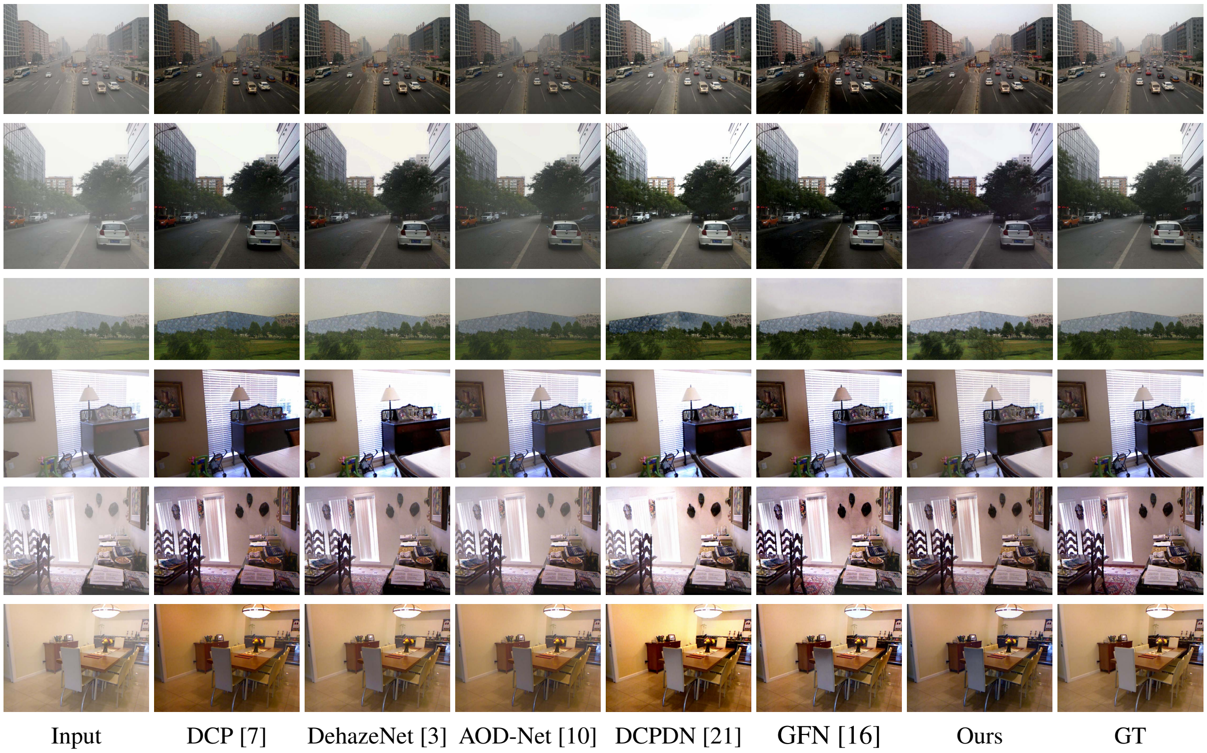

Figure 5. Comparison of the state-of-the-art dehazing methods on SOTS. The upper three rows show the dehazing results on outdoor images and the bottom three rows show the dehazing results on indoor images.

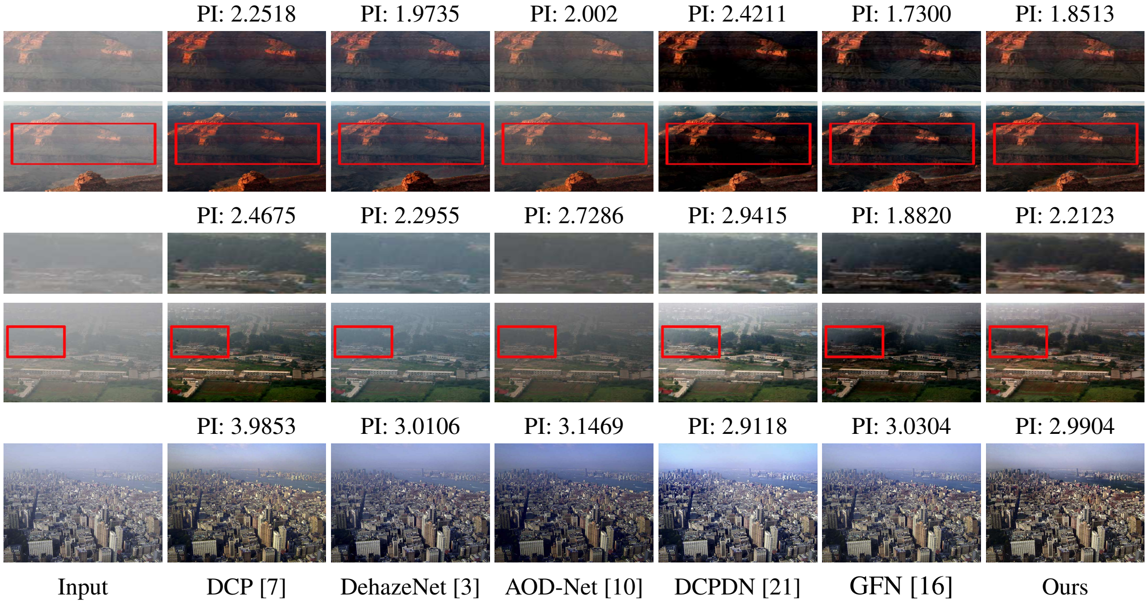

Figure 6. Comparison of the state-of-the-art dehazing methods on the real dataset. The images in the first row and the third row are the close-up of the red boxes in the second row. PIs are shown at the top of each image.

PSNR and SSIM are 2.02 dB and 0.11 respectively, which shows that EPDN is better than pix2pixHD. GAN+E+S is the worst of the ablation study variant, because the hazy image through the short-cut skip add more noises on the output of generator. GAN$^+$ is superior to GAN+E+S, because two enhancing blocks dehaze more effectively than one. We also compare the variant with and without the short-cut skip and observe that the variant with the short-cut skip is better than those without the short-cut skip in PI, because the re-enter of the original image keep the faithful color and details. EPDN and GAN$^+$ are in about the same performance, but EPDN is better than GAN$^+$ in two of three criteria, especially it is better in PI, thus, we adapt the architecture in this paper. From Fig. 4, we observe that the proposed EPDN achieve the closest result to the ground truth. Though GAN$^+$ achieves the similar performance to ours in PSNR and SSIM, but it's visual effect is inferior to ours. GAN$^+$ is overcolored obviously. Moreover, without the short-cut-skip the dehazing results looks a little darker in the first row of results, which demonstrate the effectiveness of the short-cut skip.

These ablation study demonstrates that the enhancing blocks, the short-cut skip, and the embedded structure are effective for image dehazing.

4.3. Comparisons with State-of-the-art Methods

Results on synthesis dataset. The comparison results are shown in Table. 3 in which the digital valuse are the averages of the results on SOTS in terms of PSNR, SSIM, and PI. It demonstrates that EPDN achieves the best performance of image dehazing in terms of both PSNR and SSIM on the indoor dataset of SOTS, and it achieves the breakthrough gain with 2.76 dB in PSNR and 0.0432 in SSIM compared with the second place method GFN [16].

On the outdoor dataset of SOTS, EPDN achieves the best performance in PSNR, ranks the second among the compared methods in SSIM and PI. GFN [16] rank the second in PSNR and the first in PI. The distance between the best and the second best is 0.11 dB and 0.01 in PSNR and SSIM, which is smaller than the counterpart distance on indoor data.

Fig. 5 gives the comparison of visual effect in which the comparison results on outdoor data are shown in the upper three row and those on indoor data are shown in the bottom three row. It is observed that there remains some haze in the dehazing images. DCP [7] suffers from color distortion where the results are usually darker than the ground truth images. DCPDN [21] also suffers from color distortion and it fails in details restoration. Most of the color information has been lost in the GFN [16], at the same time, it generates some artifacts. EPDN makes the dehazing image look more like the ground truth image. Furthermore, it is obvious that our model exactly outperforms the above-mentioned methods in details recovery, and it improves the dehazing results qualitatively and quantitatively.

Results on a real-world dataset. Fig. 6 shows the comparison results of visual effects on real hazy images. It is observed that: 1) Though the proposed EPDN is trained on synthesis data, it still achieves desirable dehazing results on the real-world dataset, which show the robustness of EPDN. 2) DCP [7] results in color distortion in the sky area and suffers from blur. DehazeNet [3] and AOD-Net [10] cannot remove haze effectively. DCPDN [21] and GFN [16] can not remove haze effectively in heavily hazy scene. 3) Our method is comparable in terms of PI among those state-of-the-art methods, and achieves best visual effect. Comparing to the results of five state-of-the-art methods, it can be seen that our results (EPDN) are superior in both visual effect and quantitative criteria.

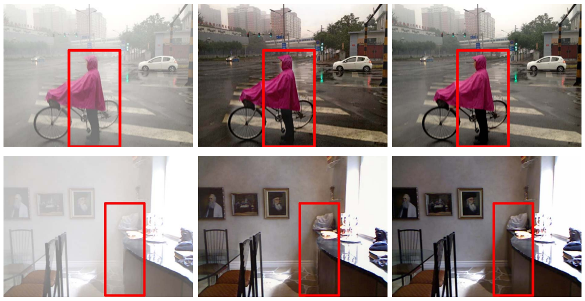
(a) Input (b) Ours (c) GT

Figure 7. A dehazing example for heavily hazy scene. Our method is not robust enough when the haze is extremely thick in the original input.

4.4. Limitation

Our method is not very robust for heavily hazy scene. As shown in Fig. 7, the edges of objects in heavily haze can not be recovered naturally. The limitation might be solved by applying more enhancing blocks in our network.

5. Conclusion

In this paper, we propose Enhanced Pix2pix Dehazing Network (EPDN) which does not rely on the estimations of the transmission map and atmospheric light. We transform the problem of image dehazing to the problem of image-to-image translation. Draw lessons from the global-first [5] theory of visual perception, we embed a GAN in our architecture, which is followed by two well-designed enhancing blocks, and the discriminator only guides the output of the multi-resolution generator. Experimental results on both the synthesis dataset and the real-world dataset demonstrate that the proposed method achieves the best performance of image dehazing in both the quantitative and qualitative evaluations. Especially, it keeps the faithful color and structures.

6. Acknowledgements

This work was supported by the National Natural Science Foundation of China under Grant 61876161, Grant 61772524, and Grant U1065252, and partly by the Beijing Municipal Natural Science Foundation under Grant 4182067.

References

[1] Dana Berman, Shai Avidan, et al. Non-local image dehazing. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1674–1682, 2016.

[2] Yochai Blau, Roey Mechrez, Radu Timofte, Tomer Michaeli, and Lihi Zelnik-Manor. 2018 PIRM challenge on perceptual image super-resolution. *arXiv preprint arXiv:1809.07517*, 2018.

[3] Kui Jia Chunmei Qing Dacheng Tao Bolun Cai, Xiangmin Xu. Dehazenet: an end-to-end system for single image haze removal. *IEEE Transactions on Image Processing*, 25(11):5187–5198, 2016.

[4] Xiaokang Yang Ming Hsuan Yang Chao Ma, Chih-Yuan Yang. Learning a no-reference quality metric for single-image super-resolution. *Computer Vision and Image Understanding*, 2017.

[5] Lin Chen. The topological approach to perceptual organization. *Visual Cognition*, 12(4):553–637, 2005.

[6] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014.

[7] Kaiming He, Jian Sun, and Xiaoou Tang. Single image haze removal using dark channel prior. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1956–1963, 2009.

[8] Kaiming He, Jian Sun, and Xiaoou Tang. Single image haze removal using dark channel prior. *IEEE Trans Pattern Anal Mach Intell*, 33(12):2341–2353, 2011.

[9] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1125–1134, 2017.

[10] Boyi Li, Xiulian Peng, Zhangyang Wang, Jizheng Xu, and Dan Feng. Aod-net: All-in-one dehazing network. In *IEEE International Conference on Computer Vision*, pages 4780–4788, 2017.

[11] Boyi Li, Wenqi Ren, Dengpan Fu, Dacheng Tao, Dan Feng, Wenjun Zeng, and Zhangyang Wang. Benchmarking single-image dehazing and beyond. *IEEE Transactions on Image Processing*, 28(1):492–505, 2019.

[12] Earl J McCartney. Optics of the atmosphere: scattering by molecules and particles. *New York, John Wiley and Sons, Inc., 1976. 421 p.*, 1976.

[13] Anish Mittal, Rajiv Soundararajan, and Alan C Bovik. Making a "completely blind" image quality analyzer. *IEEE Signal Processing Letters*, 20(3):209–212, 2013.

[14] Srinivasa G Narasimhan and Shree K Nayar. Vision and the atmosphere. *International journal of computer vision*, 48(3):233–254, 2002.

[15] Wenqi Ren, Si Liu, Hua Zhang, Jinshan Pan, Xiaochun Cao, and Ming-Hsuan Yang. Single image dehazing via multi-scale convolutional neural networks. In *European conference on computer vision*, pages 154–169. Springer, 2016.

[16] Wenqi Ren, Lin Ma, Jiawei Zhang, Jinshan Pan, Xiaochun Cao, Wei Liu, and Ming-Hsuan Yang. Gated fusion network for single image dehazing. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.

[17] Robby T. Tan. Visibility in bad weather from a single image. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–8, 2008.

[18] Ketan Tang, Jianchao Yang, and Jue Wang. Investigating haze-relevant features in a learning framework for image dehazing. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2995–3000, 2014.

[19] Ting-Chun Wang, Ming-Yu Liu, Jun-Yan Zhu, Andrew Tao, Jan Kautz, and Bryan Catanzaro. High-resolution image synthesis and semantic manipulation with conditional gans. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.

[20] Xitong Yang, Zheng Xu, and Jiebo Luo. Towards perceptual image dehazing by physics-based disentanglement and adversarial training. In *AAAI*, 2018.

[21] He Zhang and Vishal M Patel. Densely connected pyramid dehazing network. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.

[22] Hengshuang Zhao, Jianping Shi, Xiaojuan Qi, Xiaogang Wang, and Jiaya Jia. Pyramid scene parsing network. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6230–6239, 2017.

[23] Hongyuan Zhu, Xi Peng, Vijay Chandrasekhar, Liyuan Li, and Joo-Hwee Lim. Dehazegan: when image dehazing meets differential programming. In *IJCAI*, pages 1234–1240, 2018.

[24] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. *arXiv preprint*, 2017.

[25] Qingsong Zhu, Jiaming Mai, and Ling Shao. A fast single image haze removal algorithm using color attenuation prior. *IEEE Transactions on Image Processing*, 24(11):3522–3533, 2015.