High-Resolution Image Dehazing with respect to Training Losses and Receptive Field Sizes

Hyeonjun Sim, Sehwan Ki, Jae-Seok Choi, Soo Ye Kim, Soomin Seo, Saehun Kim, and Munchurl Kim
School of EE, Korea Advanced Institute of Science and Technology, Korea
{ flhy5836, shki, jschoi14, sooyekim, ssm9462, elchem96, mkimee}@kaist.ac.kr

Abstract

Haze removal is one of the essential image enhancement processes that makes degraded images visually pleasing. Since haze in images often appears differently depending on the surroundings, haze removal requires the use of spatial information to effectively remove the haze according to the types of image region characteristics. However, in the conventional training, the small-sized training image patches could not provide spatial information to the training networks when they are relatively very small compared to the original training image resolutions. In this paper, we propose a simple but effective network for high-resolution image dehazing using a conditional generative adversarial network (CGAN), which is called DHGAN, where the hazy patches of scale-reduced training input images are applied to the generator network of the DHGAN. By doing so, the DHGAN can capture more global features of the haziness in the training image patches, thus leading to improved dehazing performance. Also, the discriminator of the DHGAN is trained based on the largest binary cross entropy loss among the multiple outputs so that the generator network of the DHGAN can favorably be trained in accordance with perceptual quality. From extensive training and test, our proposed DHGAN was ranked in the second place for the NTIRE2018 Image Dehazing Challenge Track2: Outdoor.

1. Introduction

Image enhancement is a crucial process before consuming the degraded images. As damaged images degrade the visual perception of both human and machines, it is important to remove the disturbing part of the images. In many cases, image enhancement is required as a preprocessing. One of these challenging tasks is a haze removal. Haze does not appear consistently across different images and differs from various regions even within a single image. So, it is often non-uniformly distributed, depending on the surrounding atmosphere.

From an image process point of view, haze removal is not a problem of generating something because the geometry of an input hazy image should be preserved and only haze should be removed in the input image during the dehazing process, like an image-to-image translation.

Recently, deep learning-based methods have succeeded for image-to-image translations [6], [15], [22]-[23]. The deep learning-based method enables end-to-end learning for image translation if the source and target images are provided for training. The most basic auto-encoder networks and many variations of generative adversarial networks (GANs) [1] have been applied to the image translation problems. GANs were first introduced to generate a new image from a noise and have proved to be beneficial in transforming the input images to the new images in their target domain [3]-[6], [22]-[23].

In haze removal problems, each input image is not a well-controlled single object. Input is a picture that the real world in which many objects coexist in the scene. This means that we need to consider the scene context in generating the images with haze removal. As mentioned earlier, the thickness (amount) of haze varies across different regions in an image, but adjacent objects or regions often share the haze of similar thickness. In order to make the networks learn the contextual information, the sizes of the receptive fields should be very large for the high-resolution images.

As another important point of image translation, the reconstructed (or translated) images are often judged by the most prominent regions or points that may differ mostly from their usual appearance. It is important to take it into account for network training.

In this paper, we propose an effective network for high-resolution image dehazing using a conditional generative adversarial network (CGAN), which is called DHGAN. In the proposed DHGAN, the hazy patches of scale-reduced training input images are input to its generator network to effectively enlarge the receptive field sizes, and its generator and discriminator are trained focused on the worst region of outputs. This paper is organized as follows: Section 2 provides a review of image dehazing works; Section 3 describes our proposed DHGAN in details with receptive field sizes and training losses for dehazing problems; Experimental results are provided in Section 4; and finally we conclude our work in Section 5.

2. Related works

2.1. Single Image Haze Removal

Single image haze removal is more challenging than the one with multiple images since depth can be estimated more precisely with multiple images of a scene. The depth information, the distance between a camera and the subjects, is directly related to the haze thickness by its nature. Hence, many previous works have studied the formula related to the depth or transmission information to get haze-free images from hazy images [18]-[21], [25]-[27]. He *et al.* [18] defined a novel dark channel prior which is obtained from image statistics. The dark channel prior was then used to estimate transmission maps. Ancuti *et al.* [19] introduced computation of semi-inverse images to detect haze in pixel levels so that the haze in images can be effectively removed on a per-pixel basis. Ancuti and Ancuti [20] employed a multi-scale fusion based haze removal with appropriate weight maps and inputs. Meng *et al.* [21] considered the boundary conditions of transmission maps for hazy images, which turned out to be helpful for haze removal.

As deep learning-based models have drawn much attention in image processing, some researchers have tried to learn a mapping from a single hazy image to a clean image using neural networks [3]-[5], [25]-[27]. Zhang *et al.* [3] have built a network using a GAN to predict the transmission map for the input image and to remove its haze jointly. Li *et al.* [4] also proposed a GAN-based model that predicts coarse and fine transmission maps serially and then concatenates them to generate haze-free images. These networks were allowed for end-to-end learning with ground truth of transmission maps and haze-free images. However, measuring such ground truth transmission maps is very expensive and impractical. Alternatively, we only utilize paired hazy and haze-free images to train the networks. Swami and Kumar [5] first introduced a fully end-to-end learning-based GAN for single image dehazing without transmission maps. However, it fails to dehaze high-resolution hazy images because it uses a small-sized receptive field size which is not effective for haze removal of high-resolution hazy images.

2.2. Image translation with conditional GANs

Recently, GANs are proposed to resolve image generation tasks [3]-[6], [22]-[23]. Their discriminators learn to distinguish between real samples from target domain and fake made by their generators. The generators learn to fool the discriminators so that the generated images become close to the samples taken from their target domains. The generators and discriminators are trained in an adversarial fashion and are finally kept in a balance between them.

Isola *et al.* [6] combined an adversarial loss of a conditional GAN and a pixel-level reconstruction loss between the images generated by U-net [7] and their ground truth. Hence, the paired input and ground truth images should be available for training. This model, named as pix2pix, was applied to translation tasks such as graphic maps to aerial photos, and semantic labels to real photos while maintaining the inherent identity similar to haze removal.

3. Our proposed method

Our proposed haze removal network that adopts a CGAN, called DHGAN, is based on the pix2pix network [6] that was applied to image-to-image translation. Note that the haze removal from a hazy image can also be regarded as an image translation problem. The generator of our proposed DHGAN consists of a series of six 2-strided convolutional layers as an encoder and six 2-strided transposed convolutional layers as a decoder. The outputs of each layer in the encoder are feedforward and concatenated with the inputs of the corresponding layers in the decoder, which is a similar structure as U-net [7]. The discriminator, which is similar to the patchGAN structure in [6], consists of a series of four 2-strided convolutional layers. The discriminator yields its final output in a form of 32×32 score map. The total numbers of parameters in the generator and discriminator are about 5M and 44k, respectively.

3.1. Adjusting image scales

The NTIRE2018 Challenge dataset consists of very high-resolution images [16], [17]. We investigate the effective sizes of receptive fields for haze removal of high-resolution hazy images.

Seeing only narrow part of an image is not sufficient to learn enough spatial information. For examples, some of 256×256 patches as used in [6] can often contain all flat and solid color regions in high-resolution hazy images, which is lack of spatial information. They can be small parts of a white wall or heavy fog of the images. It is beneficial to utilize global context for effective haze removal as done in other image translation processes [8], [9]. This problem can be alleviated by using large-sized training patches which can contain more global spatial information for haze characteristics. In general, for the larger the receptive field sizes, the better the quality performance is obtained. However, enlarging the receptive fields requires increasing the filter sizes and the depth of convolutional layers, which entails the increase in computational complexity and memory space. Also, it should be pointed out that two pixels with a distance larger than receptive field size in an input image do not affect their corresponding output pixels each other. Hence, another way to enlarge the effective receptive fields is to down-sample the original training images and then crop them into small-sized training patches. That is, for examples, if an input is scaled to the half size of its width and height, the receptive field can be enlarged four times, roughly speaking. Increasing training patch sizes and

down-scaling the input enlarge the effective receptive fields which can significantly improve restoration performance.

Another issue for scaling of input images sizes is about testing phases. Similar to the training phase, down-sampling test images helps consider wider context information in the inference phase. However, the inference requires up-scaling the network output back to the original resolution of the input. We use a Lanczos interpolation for output up-scaling as well as for input down-scaling. Our DHGAN was trained with an empirically found down-scaling factor of 1/4 for both Indoor and Outdoor datasets of The NTIRE2018 Challenge. For testing, the down-scaling and up-scaling factor of 1/4 (1/2) were used for Indoor (Outdoor) dataset. The patch size was set to 512×512.

3.2. Loss functions

Isola *et al.* [6] introduced a concept called patchGAN that contains a different discriminator compared to those of general GANs. Originally, the discriminators in general GANs take an image as an input and outputs a single scalar that determines whether it is real or fake. That is, it is likely for the single scalar to take the entire image patch as a receptive field. However, the discriminator of the patchGAN outputs N×N values where each element corresponds to a small region in the input image as a receptive field. So, the discriminator of the patchGAN judges multiple overlapping regions by assigning multiple probabilities for the regions. To compute the loss at the output layer of the discriminator in general GANs, the elementwise binary cross entropy losses are averaged, which is expressed as

$$L_{DAvg} = E_{n,h,w}[-\log(1 - D_{nhw}(G(x), x)) - \log(D_{nhw}(y, x))] \quad (1)$$

$$L_{GAvg} = E_{n,h,w}[-\log(D_{nhw}(G(x), x))] \quad (2)$$

where x is an input hazy image, y is its corresponding ground truth, $G(\cdot)$ is the generator's output, $D(\cdot,\cdot)$ is the discriminator's output, E is the expectation operator, and n, h, and w represent the sample index in a batch, the height and width dimension, respectively.

Based on the human's perception characteristic with a focus of attention on the worst (the most prominent) degraded regions in quality assessment, we define a new loss function by modifying (1) and (2) as follows:

$$L_{DMax} = E_n[\max_{h,w}[-\log(1 - D_{nhw}(G(x), x))]] + E_n[\max_{h,w}[-\log(D_{nhw}(y, x))]] \quad (3)$$

$$L_{GMax} = E_n[\max_{h,w}[-\log(D_{nhw}(G(x), x))]] \quad (4)$$

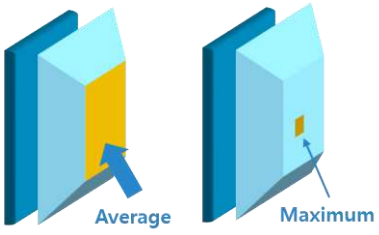
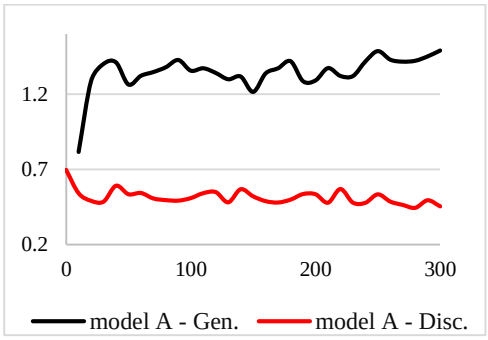


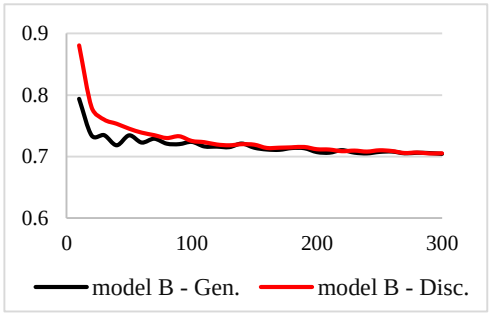
Figure 1. An illustration of the adversarial loss functions according to average loss in (1) and (2), and max loss (3) and (4).

Figure 1 illustrates the adversarial loss functions of the discriminators according to the average loss in (1) and (2), and max loss (3) and (4). While (1) and (2) penalize all regions equally by taking an average, (3) and (4) backpropagate only the error for the worst part of the output by taking the maximum over the spatial dimension. In the generator's perspective, the loss function penalizes the region for which the discriminator outputs the highest probability to be fake. On the other hands, in discriminator's perspective, the discriminator's loss function penalizes the areas that are the most misjudged.

In order to see the effectiveness of the proposed loss functions in (3) and (4), we perform a toy experiment with a simple GAN [6] where L1 and adversarial loss are only used. To trace the learning process, we trained the same GAN in two ways: (i) L1 loss and average adversarial loss in (1) and (2), (ii) L1 loss and max adversarial loss in (3) and (4). Figure 2 shows the learning curves for the average adversarial loss and the max adversarial loss. As shown in



(a) average adversarial loss



(b) proposed max adversarial loss

Figure 2. Average adversarial loss versus max adversarial loss during the training.

Figure 2-(a), the generator is overwhelmed by the discriminator while in Figure 2-(b), the discriminator and the generator are well balanced during the training. We observed this trend for various cases with different network architectures and parameters.

In addition to the basic pixel level L1 loss, a feature-level loss can also be used. It is also named as a perceptual loss, L_{VGG}, [10] which is defined as L1 norm between features of the generated image and the ground truth image. The features are extracted from a pre-trained network. We used 'relu2_2' layers of VGG-16 [11], which is a convolutional neural network for image classification. It is reported in [12] and [13] that the PSNR performance becomes worse when adversarial and L1 losses are used together than the case with only L1 loss for the rain removal task. Then, they found that the combination of adversarial, L1 and perceptual losses could yield better performance. We present the experimental results in Section 4.2 with all combinations of the loss functions and compared them one another in terms of PSNR performance. Our final loss term can be represented as a weighted sum of the loss functions as follows:

$$L_G = \lambda_1 L_1 + \lambda_{VGG} L_{VGG} + \lambda_{Avg} L_{GAvg} + \lambda_{Max} L_{GMax} \quad (5)$$

$$L_D = \lambda_{Avg} L_{DAvg} + \lambda_{Max} L_{DMax} \quad (6)$$

where L_G is the loss for the generator and L_D is the loss for the discriminator. The λ's are the weights for their corresponding losses and are empirically determined

4. Experiment results

4.1. Experiment settings

Training images are from NTIRE2018 Image Dehazing Challenge training dataset [16], [17]. All 25 indoor training images and 33 outdoor training images are used. Note that 2 out of the 35 outdoor images were not used because the heights of the images are less than 2048 and cannot be used in 512×512 size after down-scaling them by a factor of 4. Therefore, the training dataset consists of 25 indoor and 33 outdoor images, targeting at the joint training of our DHGAN for both indoor and outdoor datasets. To validate the training process of our DHGAN, 5 indoor and 5 outdoor validation hazy images are used.

The total 58 training images are down-sampled so that the widths and heights are one-fourth of their original image scales. To augment data, 512×512-sized patches are randomly cropped, flipped and rotated from the down-sampled training images at every iteration. The indoor and outdoor test images are down-sampled to quarter and half, respectively before being fed into the trained DHGAN. Finally, the output images are up-sampled to their original image sizes. Note that the down-sampling factors of 4 and 2 for indoor and outdoor images were empirically resulted. We trained the DHGAN for 500 epochs. One patch is sampled per image at each epoch. Adam optimizer [14] is used with $\beta_1 = 0.5$ and a learning rate of 0.0002. After 100 epochs, the learning rate linearly decreases towards zero.

The batch size is one of the most important hyper parameters in our task. In our DHGAN, we found that the batch size larger than 1 yields inferior results with stains in the generated images. So, we used the batch size of 1 for training.

4.2. Analysis of image scales for dehazing performance

We trained the DHGAN with various down-scaling factors mentioned in Section 3.1. To see only the effect of the image scales, we set the weight parameters in (5) and (6) as $\lambda_1 = 1$, $\lambda_{VGG} = 5$ and $\lambda_{Avg} = 0$ and $\lambda_{Max} = 0.01$. Given training images, the width and height are reduced to $1/n$ of their original sizes for a down-scaling factor of n. Then, the training patches are randomly cropped at every iteration. The patch size is fixed to either 256 or 512 during training. The test images are reduced to $1/m$ for a down-scaling factor of m. After feed-forward image translation, we calculate PSNR between a generated image up-scaled back to the original size and its corresponding ground truth as the main evaluation metric. Table I shows the average PSNR performance of our DHGAN for 10 validation images under given patch sizes and image scale factors. It is noted in Table I that n and m are the down-scaling factors for training and test, respectively. As shown in Table I, the DHGAN trained with the 512×512-sized patches gives higher performance with average 1dB or higher than the cases trained with the 256×256-sized patches, if other settings are identical. In addition, the DHGAN trained with

Table I. Average PSNR of generated images with each patch size, down-scale factor n for training images and down-scale factor m for validation images.

Patch sizes	m / n	1		2		4	
		Indoor	Outdoor	Indoor	Outdoor	Indoor	Outdoor
512	4	19.75	22.61	19.85	**22.93**	**20.08**	22.70
	2	18.34	21.75	18.27	22.15	18.47	21.43
256	4	17.95	21.46	17.73	21.86	17.66	21.51
	2	16.64	20.21	16.78	20.36	17.06	19.81

the down-scaling factor, n, of 4 is superior to the one with 2. It can be implied from Table I that enlarging the receptive fields helps the DHGAN see a wider range of the hazy scene and get well trained for dehazing. Figure 3 shows one of the generated validation images. Compared to ground truth in Figure 3-(b), the heavy haze covers the whole area of input hazy image in Figure 3-(a). When the patch size is set to 256×256 with n = 1/2 and m = 1/2, haze still remains in the upper side of the output image. When the training image scale is further reduced and the patch size gets bigger, it can be observed in Figure 3-(c) and -(d) that the haze in the upper side was more eliminated, so resulting in an improved PSNR performance.

For the down-scaling factor of m in the testing phase, we only need to find the best m when n is 1/4 and the patch size is 512. One can expect that the same value of n and m will give a better result since the difference of n and m can cause to increase a heterogeneity between training and test data. In our experiments, we found that the DHGAN did not yield the best PSNR performance for the test hazy input images without down-scaling. Moreover, the down-scaling factor of m for the testing images yields less impact on the PSNR performance of the DHGAN than the down-scaling factor of n for the training images. Down-scaling the input hazy images enlarges the receptive fields but entails losing the high frequency information from the resulting down-scaling. Therefore, it is important to find an appropriate down-scaling factor of m depending on the datasets.

4.3. Analysis of loss functions

Table II. Average PSNR performance of the DHGAN for different combinations of various loss functions.

λ_1	λ_{VGG}	λ_{Avg}	λ_{Max}	Indoor	Outdoor	Avg.
●				18.12	22.03	20.07
●	●			19.54	22.69	21.12
●		●		17.92	22.12	20.02
●			●	17.82	22.10	19.96
●	●	●		**20.56**	21.93	21.24
●	●		●	19.85	**22.93**	**21.39**

In this section, to test only the effect of loss functions, the down-scaling factors of n and m are fixed to 1/4 and 1/2, respectively. We use six combinations of different loss functions to analyze their effects on PSNR performance. As a basic pixel-level L1 loss, we set λ_1 =1. The other weighting factors of the losses are set to λ_{VGG} = 5, λ_{Avg} = 0.01 and λ_{Max} = 0.01, if they are used. Table II shows the PSNR performance of the DHGAN for different combinations of various loss functions with 10 validation images. As shown in Table II, the most effective loss function turned out to be the perceptual loss. In most cases, PSNR increased when

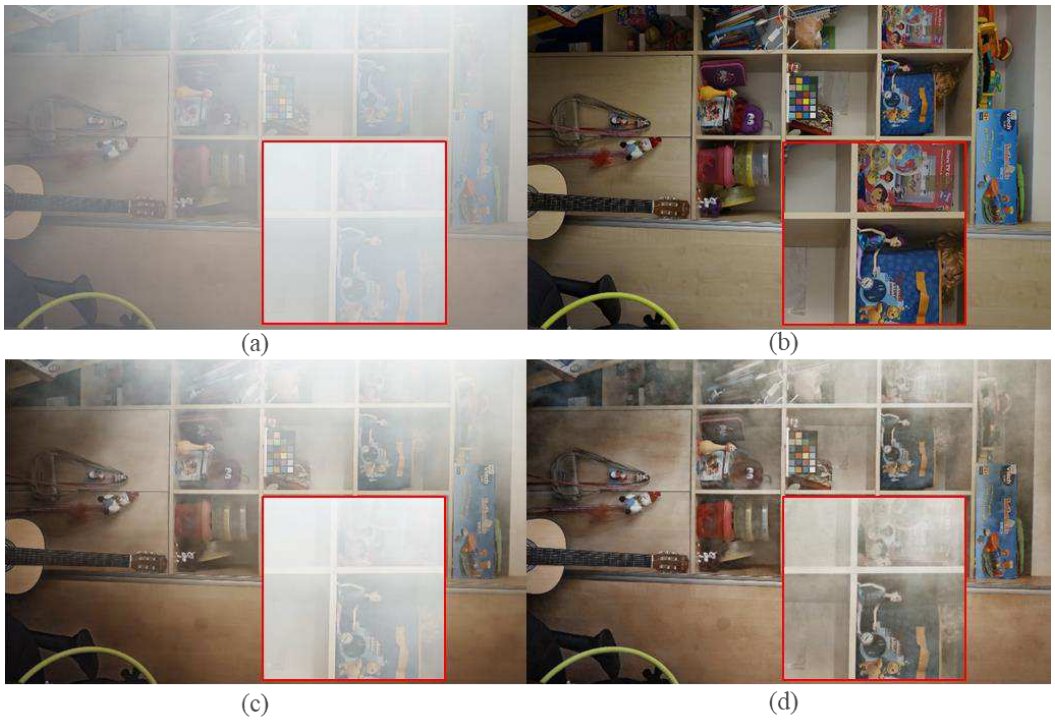
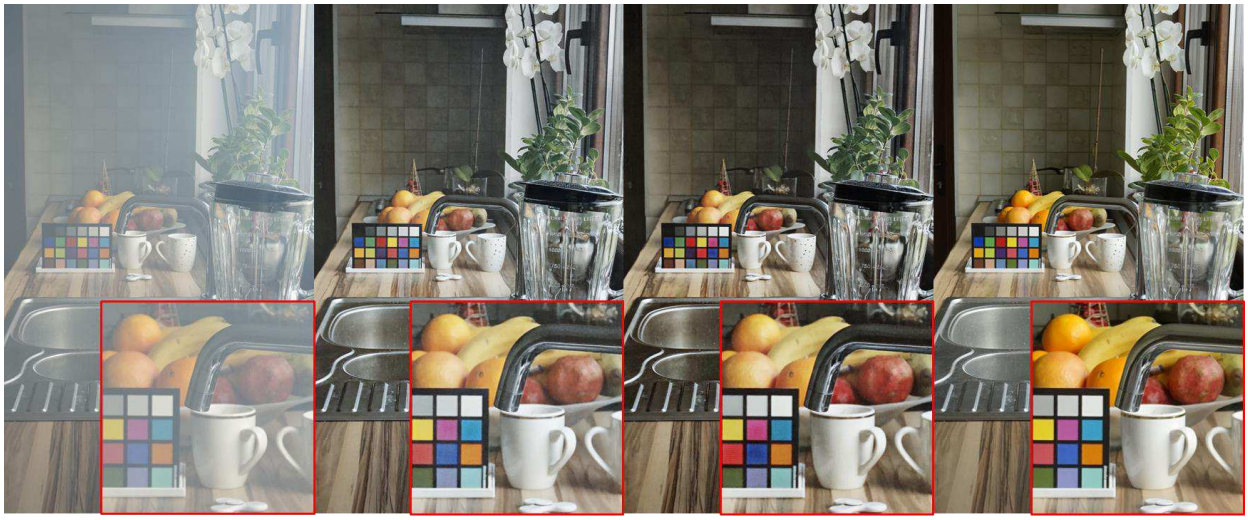


Figure 3. The generated images with image scales. (a) Input hazy image, (b) Ground truth, (c) 256-patch with n=1/2, m=1/2 PSNR=11.28 and SSIM=0.6895, (d) 512-patch with n=1/4, m=1/2 PSNR=16.28 and SSIM=0.7681.



(a) (b) (c) (d)

Figure 4. Some generated images using different loss functions: (a) Input hazy image, (b) L_1+L_{VGG}+L_{GAvg}, PSNR=23.74 and SSIM=0.8253, (c) L_1+L_{VGG}+L_{GMax}, PSNR=21.40 and SSIM=0.7998, (d) Ground truth.

the perceptual loss was added. Meanwhile, using the adversarial loss without perceptual loss did not increase PSNR performance, which is consistent with the results in [12] and [13]. However, the highest PSNRs were obtained when the perceptual loss and adversarial loss are used in a combination. For the five indoor validation images, the average adversarial loss in (1) and (2) has led to better PSNR performance than our proposed max adversarial loss in (3) and (4). For the five outdoor validation images, our alternative max adversarial loss yields better PSNR performance, based on which the DHGAN has taken the 2nd rank PSNR performance for *NTIRE2018 Image Dehazing Challenge Track2: Outdoor*. Figure 4 shows some dehazed images by the DHGAN for which average adversarial loss outperforms the case with our proposed max adversarial loss in terms of PSNR. Although PSNR of Figure 4-(b) is higher than that in Figure 4-(c), Figure 4-(c) looks better visually pleasing while the color of Figure 4-(b) looks washed out.

Figure 5 shows some dehazed images for an indoor validation image using our DHGAN trained with different loss functions. It should be noted that the original (clean) image in Figure 5-(e) contains the flat regions in a large portion unlike the one in Figure 4-(d). Figure 5-(b) contains much of stains appearing in the smooth areas for which the DHGAN was trained only with L1 loss. On the other hand, when the DHGAN is trained with L1 and perceptual losses in combination, most of the stains disappeared, which is visually more pleasing. Nevertheless, the DHGAN trained with L1 and perceptual losses yields dehazed images with

(a) (b) (c) (d) (e)

Figure 5. Some generated images with different loss functions (a) Input hazy image, (b) L_1, PSNR=18.42 and SSIM=0.7322, (c) L_1+L_{VGG}, PSNR=21.47 and SSIM=0.7807, (d) L_1+L_{VGG}+L_{GMax}, PSNR=22.53 and SSIM=0.7917, (e) Ground truth

Table III. Average PSNR and SSIM results of each dataset and each dehazing methods

	Metric	DCP [18]	BCCR [21]	DehazeNet [25]	MSCNN [26]	AOD-Net [27]	DHGAN
SOTS	PSNR	16.62	16.88	21.14	17.57	19.06	**25.60**
	SSIM	0.8179	0.7913	0.8472	0.8102	0.8504	**0.9419**
HSTS	PSNR	14.84	15.08	**24.48**	18.64	20.55	24.04
	SSIM	0.7609	0.7382	**0.9153**	0.8168	0.8973	0.9048

color washed out, which still bothers pleased perception of visual quality. This is because both L1 and perceptual loss functions are based on a pixel- and feature-level comparison between ground truth and generated one, which tries to find the best fit in an average sense [6]. On the other hand, if our max adversarial loss is additionally used, the color washed-out problem is improved, thus leading to perceptually more pleasing dehazed images.

4.4. Comparison to the previous dehazing methods

We quantitatively compared the performance of our network and the previous dehazing methods. Li *et al*. [24] provided a large-scale public benchmark dataset for single image dehazing, called Realistic Single Image Dehazing (RESIDE). RESIDE include three subsets; Indoor Training Set (ITS), Synthetic Objective Testing Set (SOTS) and Hybrid Subjective Testing Set (HSTS). ITS is a training dataset and consists of synthetic indoor hazy images and ground truth images. For the test dataset, SOTS consists of pairs of synthetic indoor hazy images and ground truth images. Another test dataset HSTS consists of both real outdoor hazy images and the pairs of synthetic outdoor hazy and ground truth images. The authors of [24] evaluated several state-of-the-art haze removal algorithms [18], [21], [25]-[27] in terms of subjective and objective metrics. The data-driven algorithms were trained with the common training dataset ITS in [24]. In order to evaluate our DHGAN under the same condition, we trained DHGAN with training dataset ITS from scratch. Since the image resolution of ITS is 620×460 which is relatively lower than the resolution of NTIRE2018 Image Dehazing Challenge dataset [16], [17], we did not down-scale the training images and just cropped 448×448-sized patches as large as possible. Also, the test images in SOTS and HSTS were not down-scaled. The weight parameters in (5) and (6) were set as $\lambda_1 = 1$, $\lambda_{VGG} = 5$ and $\lambda_{Avg} = 0$ and $\lambda_{Max} = 0.01$. Then we computed objective metrics, PSNR and SSIM for the test datasets SOTS and synthetic images of HSTS. Table III shows average PSNR and SSIM performance of the previous methods and our DHGAN for each test dataset. The first and second top values are highlighted as bold type and blue in Table III, respectively. From both PSNR and SSIM perspective, our DHGAN achieved the best and the second-best performance among the dehazing algorithms for the SOTS and HSTS test dataset, respectively. Note that the other algorithms utilize the relation between the transmission maps and the haze while proposed DHGAN

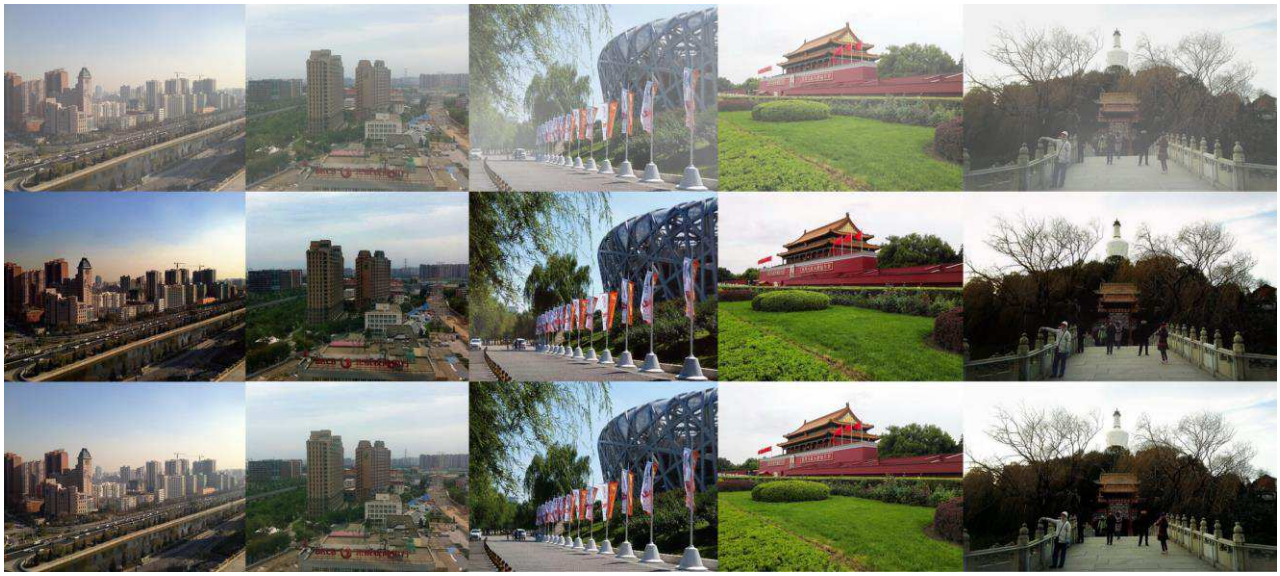


Figure 6. Some examples of (top row) input synthetic hazy images, (middle row) dehazed images by DHGAN and (bottom row) ground truth images of HSTS.

generates dehazed images directly from the hazy images. Figure 6 shows some examples of input synthetic hazy images, dehazed images by DHGAN, and ground truth images of HSTS. In Figure 6, the haze of the scenes were removed successfully in the generated images by DHGAN (middle row). There is even more haze in the real photo (bottom row) than dehazed one in the first and second examples.

5. Conclusion

In this paper, we proposed a CGAN-based high-resolution image dehazing network, where it can capture more global features of the haziness in the training image patches by using scale-reduced training input images. This leads to improved dehazing performance. Also, we proposed a max adversarial loss to train the DHGAN, which picks up the maximum values of adversarial losses among the multiple outputs. From extensive training and test, our proposed DHGAN was ranked in the second place for the NTIRE2018 Image Dehazing Challenge Track2: Outdoor.

Acknowledgement

This work was supported by Institute for Information & communications Technology Promotion (IITP) grant funded by the Korea government (MSIT) (No. 2017-0-00419, Intelligent High Realistic Visual Processing for Smart Broadcasting Media).

References

[1] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A.Courville, and Y. Bengio. Generative adversarial nets. *Advances in neural information processing systems* (*NIPS*), 2014

[2] R. Timofte, *et al*. New trends image restoration and enhancement workshop and challenge on super-resolution, dehazing, and spectral reconstruction in conjunction with CVPR 2018. http://www.vision.ee.ethz.ch/ntire18, 2018.

[3] H. Zhang, V. Sindagi, and V. M. Patel. Joint Transmission Map Estimation and Dehazing using Deep Networks. *arXiv preprint arXiv:1708.00581*, 2017.

[4] C. Li, J. Guo, F. Porikli, C. Guo, H. Fu, and X. Li. DR-Net: Transmission Steered Single Image Dehazing Network with Weakly Supervised Refinement. *arXiv preprint arXiv:1712.00621*, 2017.

[5] K. Swami, and S. K. Das. CANDY: Conditional Adversarial Networks based Fully End-to-End System for Single Image Haze Removal. *arXiv preprint arXiv: 1081.02892*, 2018.

[6] P. Isola, J.-Y Zhu, T. Zhou, and A. A. Efros. Image-to-image translation with conditional adversarial networks. *CVPR*, 2017.

[7] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. *MICCAI*, 234-241, Springer, 2015.

[8] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia. Pyramid scene parsing network. *IEEE Conf. on Computer Vision and Pattern Recognition* (*CVPR*), 2017

[9] S. Iizuka, E. Simo-Serra, and H. Ishikawa. Globally and locally consistent image completion. *ACM Transactions on Graphics* (*TOG*), 2017

[10] J. Johnson, A. Alahi, and L. Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. *European Conference on Computer Vision* (*ECCV*), Springer, Charm, 2016.

[11] K. Simonyan, and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *ICLR*, 2015.

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

[13] C. Wang, C. Xu, C. Wang, and D. Tao. Perceptual adversarial networks for image-to-image transformation. *arXiv preprint arXiv:1706.09138*, 2017.

[14] D. P. Kingma, and J. Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[15] Q. Chen, and V. Koltun. Photographic image synthesis with cascaded refinement networks. *IEEE International Conference on Computer Vision* (*ICCV*), vol. 1, 2017.

[16] C. O. Ancuti, C. Ancuti, R. Timofte, and C. De Vleeschouwer. I-HAZE: a dehazing benchmark with real hazy and haze-free indoor images. *arXiv preprint arXiv:1804.05091*, 2018.

[17] C. O. Ancuti, C. Ancuti, R. Timofte, and C. De Vleeschouwer. O-HAZE: a dehazing benchmark with real hazy and haze-free outdoor images. *arXiv preprint arXiv:1804.05101*, 2018

[18] K. He, J. Sun, and X. Tang. Single image haze removal using dark channel prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (*TPAMI*), 2011.

[19] C. O. Ancuti, C. Ancuti, C. Hermans, and P. Bekaert. A fast semi-inverse approach to detect and remove the haze from a single image. *ACCV*, 2010.

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

[21] G. Meng, Y. Wang, J. Duan, S. Xiang, and C. Pan. Efficient image dehazing with boundary constraint and contextual regularization. *IEEE Int. Conf. on Computer Vision*, 2013.

[22] Y. Taigman, A. Polyak, and L. Wolf. Unsupervised cross-domain image generation. *arXiv preprint arXiv:1703.10593*, 2017,

[23] J. Y. Zhu, T. Park, P. Isola, and A. A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. *arXiv preprint arXiv:1703.10593*, 2017

[24] B. Li, W. Ren, D. Fu, D. Tao, D. Feng, W. Zeng, and Z. Wang. Benchmarking single image dehazing and beyond. *arXiv preprint arXiv:1712.04143*, 2017

[25] 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*, vol. 25, no. 11, pp. 5187-5198, 2016.

[26] W. Ren, S. Liu, H. Zhang, J. Pan, X. Cao, and M.-H Yang. Single image dehazing via multi-scale convolutional neural networks. *European Conference on Computer Vision* (*ECCV*), 2016

[27] B. Li, X. Peng. Z. Wang, J. Xu, and D. Feng. Aod-net: All-in-one dehazing network. *IEEE International Conference on Computer Vision* (*ICCV*), 2017