

# Fully End-to-End learning based Conditional Boundary Equilibrium GAN with Receptive Field Sizes Enlarged for Single Ultra-High Resolution Image Dehazing

Sehwan Ki, Hyeonjun Sim, Jae-Seok Choi, Soo Ye Kim, Soomin Seo, Saehun Kim, and Munchurl Kim

School of EE, Korea Advanced Institute of Science and Technology

Daejeon, Korea

{shki, flhy5836, jschoi14, sooyekim, ssm9462, elchem96, mkimee}@kaist.ac.kr

## Abstract

*A receptive field is defined as the region in an input image space that an output image pixel is looking at. Thus, the receptive field size influences the learning of deep convolution neural networks. Especially, in single image dehazing problems, larger receptive fields often show more effective dehazing by considering the brightness and color of the entire input hazy image without additional information (e.g. scene transmission map, depth map, and atmospheric light). The conventional generative adversarial network (GAN) with small-sized receptive fields cannot be effective for hazy images of ultra-high resolution. Thus, we proposed a fully end-to-end learning based conditional boundary equilibrium generative adversarial network (BEGAN) with the receptive field sizes enlarged for single image dehazing. In our conditional BEGAN, its discriminator is trained ultra-high resolution conditioned on downscale input hazy images, so that the haze can effectively be removed with the original structures of images stably preserved. From this, we can obtain the high PSNR performance (Track 1 - Indoor: top 4<sup>th</sup>-ranked) and fast computation speeds. Also, we combine an L1 loss, a perceptual loss and a GAN loss as the generator's loss of the proposed conditional BEGAN, which allows to obtain stable dehazing results for various hazy images.*

## 1. Introduction

Images are often captured under bad weather conditions, which results in the degraded images with many obscured regions by fog, mist, and haze etc. Especially, the hazy images not only lower their aesthetical values, but also cause a significant performance degradation for object recognition. Thus, dehazing is an essential preprocessing to both aesthetic photography and computer vision applications. In general, the formulation of a hazy image can be modeled as

$$I(x) = J(x)t(x) + A(1-t(x)) \quad (1)$$

where  $I(x)$  and  $J(x)$  are an input hazy image and a clean image,  $A$  is the global atmospheric light, and  $t(x)$  is the

transmission ratio that the portion of lights reaches the camera sensors. As a result, the haze removal using only a single degraded hazy image is a very challenging and ill-posed problem. The conventional haze removal methods estimate the global atmospheric light and the transmission ratio, and they remove the haze using the estimated parameters of (1) [1]-[4]. But, this approach is not a way to optimize the perceptual quality of generated dehazed images. Also, the inaccuracies of the estimated parameters can lead to weird distortions or to poor performance of haze removal. Instead, deep-learning-based convolutional neural networks can be used to effectively remove the image haze via fully end-to-end-learning. For an effective fully-end-to-end learning, the network must be able to understand the characteristics of the entire hazy images. Especially, when the resolutions of hazy images are very large, the training of the haze removal networks with small receptive field sizes becomes difficult since the networks cannot consider the properties of the entire hazy images.

Thus, we proposed a fully end-to-end learning based single image dehazing method by extending the BEGAN as a conditional BEGAN. The proposed condition BEGAN uses relatively large receptive field sizes for downscaled input hazy images so that it can capture global property of hazy input images and yield stably dehazed output images with image structures preserved. In order to increase the receptive field size, we use the input hazy images of reduced sizes which are then processed through our conditional BEGAN, and increase the dehazed output image sizes back to their original sizes using a bicubic interpolation method. Also, we use a combined cost of L1 loss, perceptual loss and an adversarial loss to train the conditional BEGAN, which helps enhancing the perceptual quality of dehazed images. We have experimentally confirmed that the conditional BEGAN has shown not only better perceptual quality but also better PSNR performance than that of convolution neural networks trained with the L1 loss and perceptual loss. Our proposed conditional BEGAN was ranked in the 4<sup>th</sup> place in NTIRE 2018 Dehazing Challenge (Track 1: Indoor) [10].

## 2. Related works

### 2.1. Single Image Dehazing

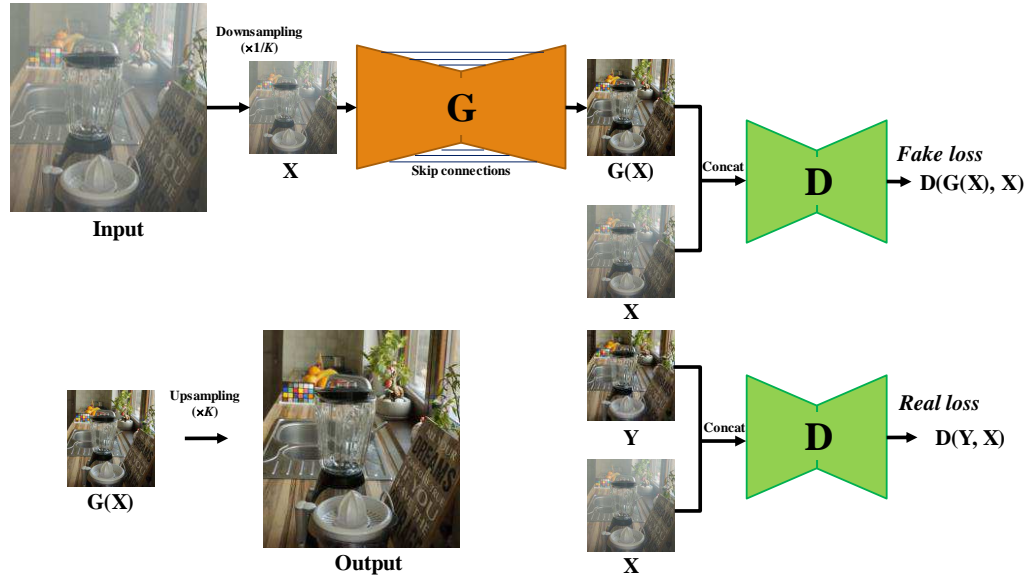


Figure 1: An illustration of the proposed conditional BEGAN

Early image dehazing methods performed by using multiple images captured under different weather conditions [1] or additional infrared versions of the hazy images [2]. But, the hazy removal should be performed by using only single hazy images since there is seldom additional information in realistic dehazing applications. Thus, single image dehazing methods have been studied extensively [3], [4], [5], [17], [18], [19], [20], [21], [22], [23]. Tan *et al.* [17] enlarges local contrast without exceeding the global atmospheric light value. He *et al.* [3] proposed a new dark channel prior based single image dehazing method. The dark channel is composed of the lowest intensity value of one pixel among three RGB channels. Using the dark channel prior, the thickness of the haze can be directly estimated. Tarel *et al.* [18] proposed a filtering method assuming that the depth-map must be continuous except along edges with large depth jumps. Ancuti *et al.* [19] presented an enhancing method based on the semi-inverse of the image. Tang *et al.* [21] proposed a framework that learns a set of features extracted from a single hazy image. Recently, the dehazing was also addressed using deep neural networks as an application of image-to-image translation. Ren *et al.* [4] proposed a multi-scale convolutional neural network for single image dehazing by learning the mapping between input hazy images and their corresponding transmission ratio maps. The coarse-scale net predicts a transmission map based on an entire hazy image, and the fine-scale net refines the dehazing results using the local property of the hazy image. But, the above conventional dehazing methods only predict the transmission ratios and the global atmospheric lights from the single input hazy images. And then, the dehazed images are separately computed according to (1) for hazy images. These methods cannot be optimized for perceptual

quality of the generated dehazed images. Thus, Swami *et al.* [5] firstly proposed a conditional adversarial networks based fully end-to-end system for single image haze removal. Also, Swami firstly applies GAN for single image haze removal. The discriminator network of GAN ensures that the generated dehazed images look as real clean images. The combination of these two approaches made it possible to optimize the perceptual quality of the generated dehazed image. But, when the size of each input hazy image is very large, the dehazing performance is degraded because it has a relatively small receptive field size.

## 2.2. U-net

For the image-to-image translation problems, an encoder-decoder network is commonly used [9], [11]-[12]. In this network, the input is passed through a series of layers that progressively downsample the output feature maps which are then passed through reverse processing. This encoder-decoder network passed only high-level features from each input hazy image to the decoder. But, there is also great deal of low-level features shared between the input images and the target images for specific image-to-image translation problems. For example, in the single image dehazing and image colorization, the input images and output images share the edge locations and prominent structures of the images. Thus, we use U-Net [6] which adds skip connections from the encoder network to the decoder network as its generator network. Each skip connection simply concatenates all channels at encoder layer  $i$  and decoder layer  $n-i$  where  $n$  is the total number of layers. Using U-Net, we can increase the receptive field sizes as well as pass the low-level features for single high resolution image dehazing.

## 2.3. Boundary Equilibrium GAN

GANs can generate very convincing images, sharper than the ones produced by auto-encoders using pixel-wise losses. But, GANs still have many unsolved problems, which is difficult to fairly train the generator and the discriminator in a balanced manner, and to appropriately determine the hyper parameters. Balancing the convergence of the generator and the discriminator is important and challengeable. Thus, Berthelot *et al.* [7] proposed a boundary equilibrium generative adversarial networks (BEGAN) which is a new equilibrium enforcing method paired with a loss derived from the Wasserstein distance in training the auto-encoder based GAN. The discriminator network of BEGAN is an auto-encoder network unlike the common discriminator networks of GAN. The output of the discriminator of BEGAN using the loss of the Wasserstein distance is neither real (0) nor fake (1). The discriminator of BEGAN is trained towards making the difference between the real image output of the discriminator and real image go to zero, and the difference between the fake image output of the discriminator and fake image go to infinite. And, by adjusting the parameters of the discriminator loss using the concept of equilibrium, the discriminator loss and the generator loss are both decreased during training. In this work, we use the U-Net as the generator network and an auto-encoder as the discriminator network with the loss derived from the Wasserstein distance with the concept of equilibrium.

## 3. Our proposed method

### 3.1. Downscale of input hazy images for large receptive field sizes

For single image dehazing, the network must consider the characteristics of the entire hazy images. If the size of an input hazy image is not large, the network can learn the entire image property based on the patch learning. However, if the size of the hazy image is very large, simple patch learning is very ineffective for haze removal. This is because a network that only observes a narrow region cannot determine whether the local region is an originally homogeneous region or a hazy region. Therefore, the performance of haze removal has a lot of influence on the receptive field sizes of the network. However, simply expanding the input patch sizes is difficult due to the lack of hardware memory sizes. Thus, we propose to reduce the input image sizes by using a simple bicubic interpolation method, and then to extend the generated dehazed images by the same interpolation method. This is a simple method, but it can easily expand the sizes of the receptive fields. Also, by reducing the input image sizes, it can greatly reduce the runtime. Since the receptive field is enlarged, the haze removal performance gets improved. In a result, the generated dehazed images tend to have more vivid color. But, there is a problem that the edges are blurred due to the processing of the hazy images with reduced resolutions and the upscaling of the output hazy images. Due to this trade-

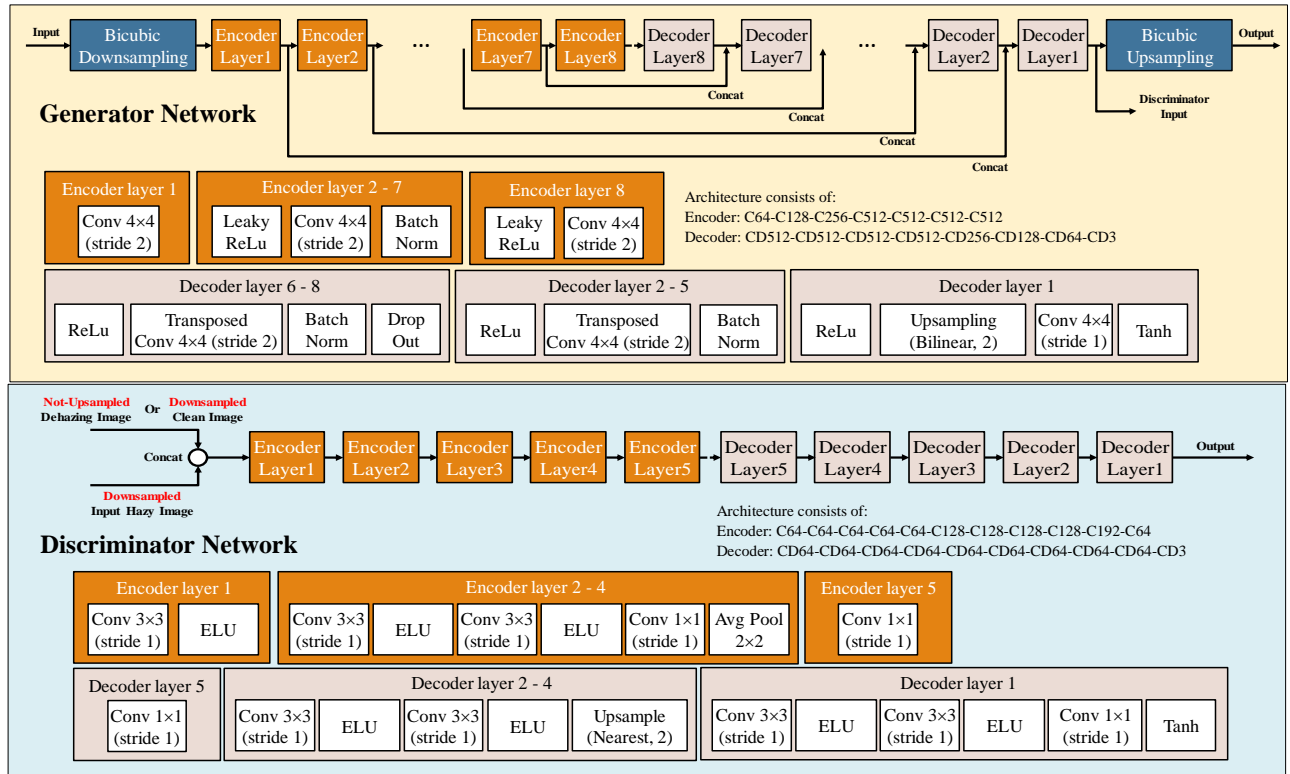


Figure 2: The generator and discriminator network architectures of the proposed conditional BEGAN.

off, it is necessary to control the resizing ratio depending on the characteristics of the input hazy images. For examples, in the case of the hazy images with many homogenous regions, increasing the resizing ratio becomes effective for haze removal because the blurred edges are small. In contrast, in the case of the hazy images with many complex texture regions, large reduction ratios adversely affect the overall perceptual image qualities due to the resulting blurred edges.

### 3.2. Proposed network architecture

As shown in Figure 1, we proposed a fully end-to-end learning based conditional BEGAN for single image dehazing. Each input image is a single hazy RGB image and the output image is a single dehazed RGB image. In order to increase the receptive fields without the loss of low-level features, we adopt U-net as the generator network which is an auto-encoder architecture with skip connections. The discriminator has also auto-encoder architecture.

As shown in Figure 2, the generator network is composed to 8 encoder layers, 8 decoder layers, and bicubic interpolator. Firstly, each input hazy image is down-sampled by a decimation factor  $K$  using bicubic interpolation where  $K$  is experimentally determined as 4 for the data set of the NTIRE 2018 Indoor Dehazing Challenge [15]. The input patches of the first encoder layer are  $512 \times 512$ -sized RGB images extracted randomly from the down-sampled input hazy images. In order to pass the low-level features to the decoder layers, the same spatial-resolution features of the encoder sides and decoder sides are concatenated. The concatenated features are input to the next decoder layers. Lastly, the generated dehazed image is up-scaled by a factor  $K$  using bicubic interpolation. It is noted that the dehazed image inputted to the discriminator network is not the up-scaled dehazed image but the dehazed image of a reduced size outputted from the generator network. By doing so, we have the same effect of enlarging the receptive field sizes to the discriminator network.

Additionally, when the convolution layer filter size is set to the multiples of the stride sizes, the checkerboard artifact from the auto-encoder network is known to be suppressed [13]. Thus, we set all convolution filter sizes to  $4 \times 4$  since the stride size is 2. We divide the transposed convolution layer into a 2-times upsampling layer using the bilinear interpolation and a convolution layer. This technique also reduces the checkerboard artifact due to upscaling using the transposed convolution [13]. As shown in Figure 2, the discriminator is also an auto-encoder as the same as the generator except skip connections. Since we use the discriminator loss derived from the Wasserstein distance, the output of the discriminator is an image map which has the same resolution as that of the input image to the discriminator. Also, we combine the conditional GAN [8] with the basic discriminator of BEGAN, which is called a conditional BEGAN. The input image of the generator can

help the discriminator distinguish whether the generated image is real or fake. In the case of the hazy removal problems, since the structure of the input hazy image and the structure of the generated dehazed image are not greatly different, only the brightness and the color are changed. Therefore, the information of the input hazy image can help the discriminator make a correct judgment. Thus, our proposed discriminator network receives the input hazy image by concatenating the input hazy image with the generated dehazed image or the clean image.

### 3.3. Objective losses

It is important to note that the proposed discriminator aims at optimizing the Wasserstein distance between auto-encoder loss distributions, not between sample distributions. The loss of the discriminator for the target clean image is defined as

$$errD_{\text{real}} = |y - D(y, x)| \quad (2)$$

where  $y$  is the target clean image,  $x$  is the input hazy image, and  $D(\cdot)$  is the output of the auto-encoder discriminator. The loss of the discriminator for the generated dehazed image is defined as

$$errD_{\text{fake}} = |G(x) - D(G(x), x)| \quad (3)$$

where  $G(x)$  is a generated dehazed image. The discriminator aims at minimizing  $errD_{\text{real}}$  and maximizing  $errD_{\text{fake}}$  simultaneously. Thus, the proposed discriminator loss,  $L_D$ , is defined as

$$L_D = errD_{\text{real}} - k_t errD_{\text{fake}} \quad (4)$$

where  $k_t$  is the control parameter that maintains the equilibrium between  $errD_{\text{real}}$  and  $errD_{\text{fake}}$ . The initial value of  $k_0$  is set to 0 and  $k_t$  is updated as follows:

$$k_{t+1} = \min(\max(k_t + \lambda_k (\gamma errD_{\text{real}} - errD_{\text{fake}}), 0), 1) \quad (5)$$

where  $\lambda_k$  is the proportional gain for  $k$  and  $\gamma$  is the diversity ratio. We used 0.005 and 0.7 for  $\lambda_k$  and  $\gamma$ , respectively, in our experiments. The loss of the generator is the weighted sum of the adversarial loss, L1 loss, and perceptual loss using pre-trained VGG-16 network [9] for optimizing the perceptual quality of the generated dehazed image. Thus, our proposed generator loss,  $L_G$ , is defined as

$$L_G = \lambda_{GAN} errD_{\text{fake}} + \lambda_{L1} |G(x) - y| + \lambda_{VGG} L_{VGG} \quad (6)$$

where  $\lambda_{GAN}$ ,  $\lambda_{L1}$ , and  $\lambda_{VGG}$  are empirically determined to be 0.2, 0.6, and 0.2, respectively. The perceptual loss based on the pre-trained VGG-16 is defined as

$$L_{VGG} = |vgg16_{relu2_2}(G(x)) - vgg16_{relu2_2}(y)| \quad (7)$$

The L1 loss plays a major role in estimating the structure of the dehazed image, and the perceptual loss helps restore the original color in the dehazed image. The GAN loss suppresses the artifacts caused by the L1 loss and the perceptual loss. Thus, the dehazed image of perceptually high quality can be generated by using the combinations of the losses.

## 4. Experiment results

### 4.1. Experiment settings

For training images, we used 30 ultra-high resolution hazy-clean image pairs obtained from the NTIRE 2018 Dehazing Challenge training and validation dataset [10], [15]. Hazy-Clean subimage pairs are extracted in random crop from the down-sized images. We use only the horizontal flip method for data augmentations. The subimages are extracted with 16 patches random crop per one training image per every epoch. The subimage sizes of the hazy image and the clean image are both set to  $512 \times 512$  pixels. The batch size is set to 1, the learning rates of the generator and the discriminator are both set to  $2 \times 10^{-4}$ . The network is trained using Adam optimizer ( $\beta_1=0.5$ ,  $\beta_2=0.999$ ) with a learning rate annealing [14]. It is necessary to vary the downscaling ratio depending on the characteristics of the input hazy image. That is, we set the downscaling ratio to 4 and 2, respectively, for the indoor and outdoor dehazing datasets [15], [16]. The number of epoch is set to 400.

### 4.2. Experimental Results and discussion

#### 4.2.1 The effects of GAN loss for PSNR performance

In general, for deep learning based image enhancement

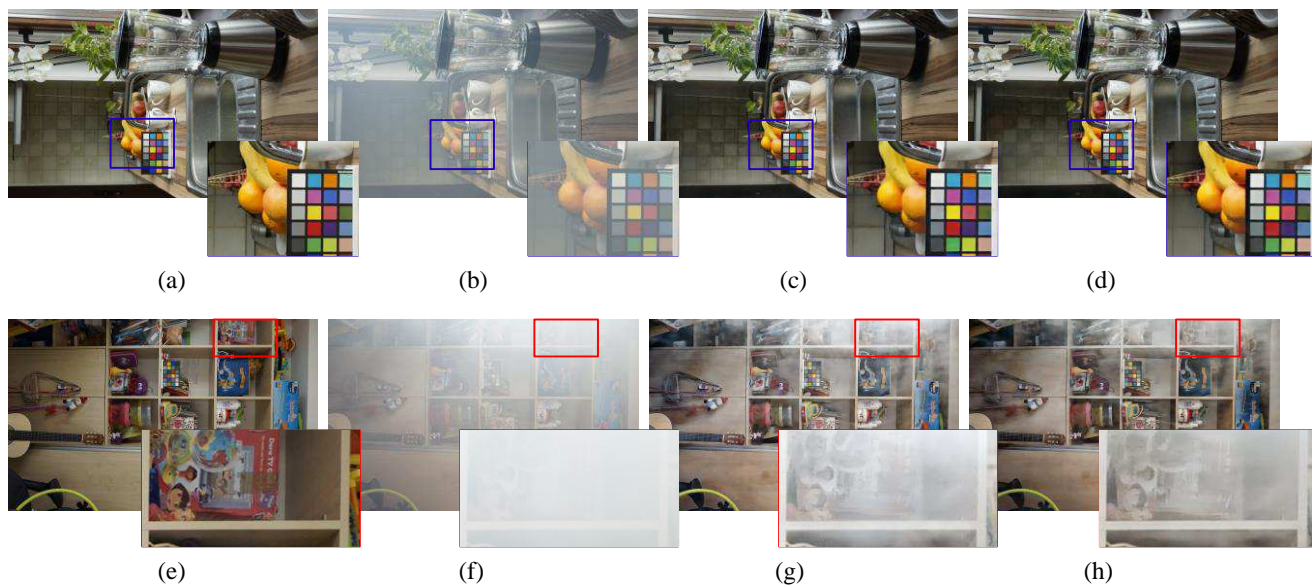


Figure 4. Subjective quality comparison between the proposed conditional BEGAN and the network without GAN loss. (a) and (e) are original images of I2 image and I3 image. (b) and (f) are hazy images of (a) and (e). (c) and (g) are the output dehazed images of the network without GAN loss. (d) and (h) are the output dehazed images of the proposed conditional BEGAN.

problems, it is known that the PSNR performance is degraded when the networks are trained with GAN loss instead of L1 or L2 norm. But, we have experimentally observed that when the generator uses a combined cost of L1 loss, perceptual loss and GAN loss, the PSNR performance is improved compared to the case without using the GAN loss. For comparison, our proposed conditional BEGAN and the CNN-based generator network without the discriminator are trained using the NTIRE 2018 Dehazing Challenge 25 training images, and are tested for the 5 validation images [15]. For fair comparison, we utilize the same generator network for our conditional BEGAN as the CNN-based generator network without the discriminator, and also use the same hyper parameters. Figure 3 shows the PSNR performance for different training losses. As shown in Figure 3, the PSNR result of the

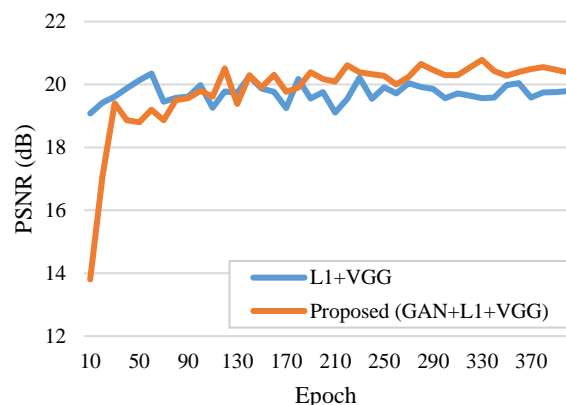


Figure 3. The average PSNR results of our proposed conditional BEGAN and the network without GAN loss for the validation hazy images.

proposed conditional BEGAN is average 0.6dB higher than that of the network without the GAN loss at 400 epochs. In Table I, it can be noted that our conditional BEGAN works considerably well for I3 and I5 images. Figure 4 shows the subjective quality comparison for the proposed conditional BEGAN and its generator-only network without GAN loss. As shown in Figure 4, I3 image looks relatively more severely hazy than I2 image. In the case of I2 image, the output dehazed image could be generated as being very similar to the original clean images based only on L1 loss and perceptual loss. However, in the case of I3 image, the network without GAN loss fails to effectively restore the shape of the obscured regions due to more severe haze, but the proposed conditional BEGAN successfully yields pleasing output dehazed image. Based on these results, it can be noted that the GAN loss can effectively enhance the perceptual quality and PSNR values of the hazy images having many obscured regions due to the severe haze.

Table I. The PSNR results for each validation image

Images	PSNR (dB) of Proposed conditional BEGAN	
	L1 + VGG losses	GAN+L1+VGG losses
I1	17.33	17.33
I2	24.80	24.21
I3	14.40	16.66
I4	22.85	22.64
I5	19.56	21.09
Avg.	19.79	20.39

#### 4.2.2 Effects of receptive field sizes with respect to dehazing performance

The receptive field sizes of a deep neural network can be effectively enlarged with down-scaling of input hazy images. That is, if the down-scaling by a decimation factor of  $N$  is applied to the input images, the resulting receptive field sizes in the input images can be increased  $N$  times than those of the original input. We use the training images of the NTIRE 2018 Dehazing Challenge which are of  $4,657 \times 2,833$  size [15]. Since our conditional BEGAN uses

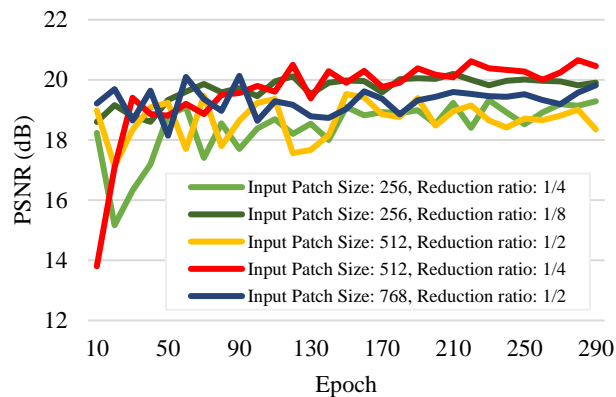


Figure 5. Average PSNR performance for different combinations of input patch sizes and downscaling factor values for the validation images of the Indoor track of NTIRE 2018 Dehazing Challenge [15].

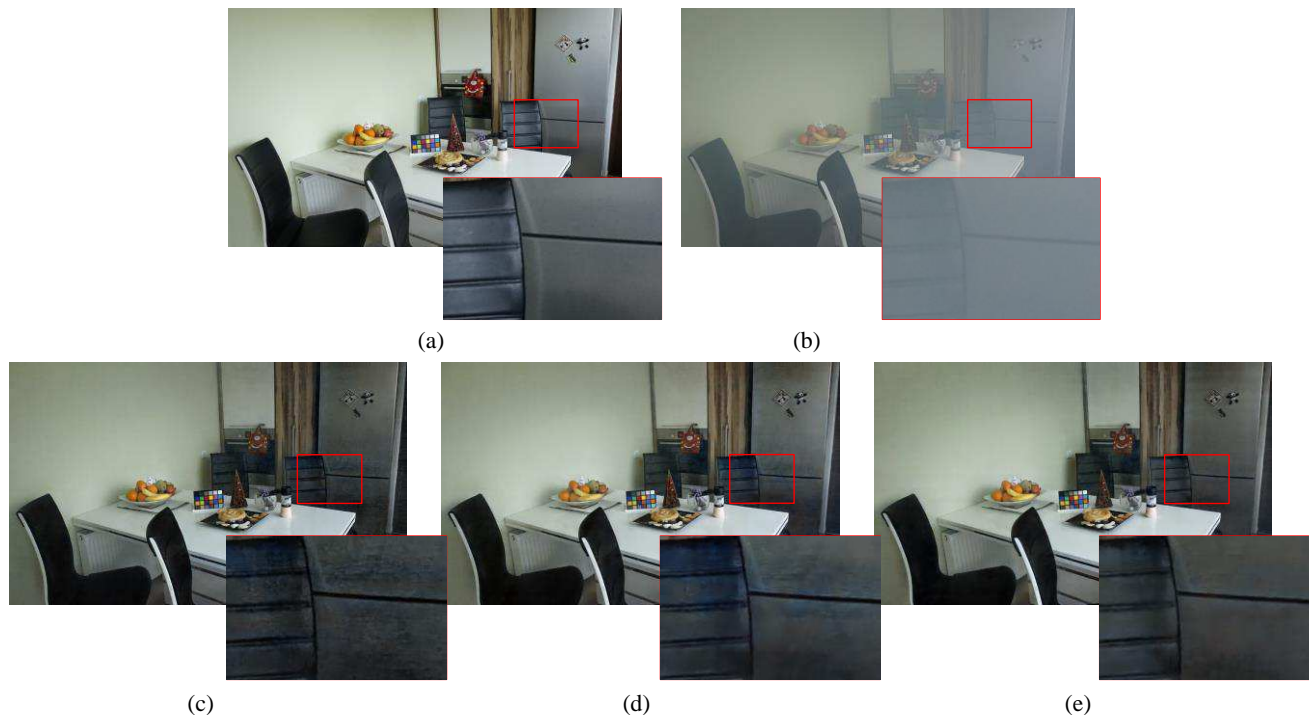


Figure 7. The subjective quality comparison according to the different discriminator networks. (a) and (b) are the original clean image of I1 image and the hazy image of I1 image. (c), (d), and (e) are the dehazed image generated by Pix2Pix, the combination of U-Net and the original BEGAN without the conditional inputs, and the proposed conditional BEGAN.

as its generator network an autoencoder having 8 convolution layers with stride 2, the sizes of the training input patches must be multiples of  $2^8$ . Thus, we have tested various cases for different combinations of the training input patch sizes and the downscaling factors with respect to the PSNR performance. Figure 5 shows the PSNR performance for different combinations of the training input patch sizes and the downscaling factors. As shown in Figure 5, the best PSNR performance is obtained when the downscaling is as small as possible and the receptive field size is as large as possible. It should be noted that if the training image samples are severely down-scaled to obtain extremely large receptive field sizes, the edges of the dehazed images get degraded with blurring artifacts. Therefore, it is necessary to consider appropriate receptive field sizes while maintaining edge information simultaneously. Thus, the input patch size of  $512 \times 512$  and the downscaling factor of 4 turns out to be best for the validation samples of the Indoor track of NTIRE 2018 Dehazing Challenge [15].

#### 4.2.3 Performance of the conditional BEGAN

Our proposed conditional BEGAN is designed by adopting the U-Net as its generator and a conditional auto-encoder network with the concept of equilibrium as the discriminator. Since the auto-encoder discriminator with the concept of equilibrium can be more appropriately trained with stable convergence than the general discriminator in the form of CNN structures [8], we adopt a condition auto-encoder for the discriminator network of our condition BEGAN. Note that a general image-to-image translator, called Pix2Pix [8], uses the U-Net as its generator and a CNN-based conditional encoder as its discriminator network. So, our first performance comparison is done between the two networks: the discriminator of ours is a conditional auto-encoder network with the concept of equilibrium and the discriminator of Pix2Pix is a CNN-based conditional encoder.

Next, our second performance comparison is for the two cases of our proposed conditional BEGAN: (i) using conditional auto-encoder as its discriminator; and (ii) using a general auto-encoder without conditional input as its discriminator. Figure 6 shows PSNR performances for our proposed conditional BEGAN and the Pix2Pix network. As shown in Figure 6, the proposed conditional BEGAN with the conditional auto-encoder as its discriminator shows the highest PSNR performance than the Pix2Pix network and our network with a general a general auto-encoder without conditional input as its discriminator. Figure 7 shows the dehazed images for three different discriminators used in their networks in Figure 6. It can also be noted in Figure 7 that the dehazed image of the proposed conditional BEGAN with the conditional auto-encoder as its discriminator shows the highest perceptual quality than the other two dehazed images. As shown in Figure 7, the dehazed image

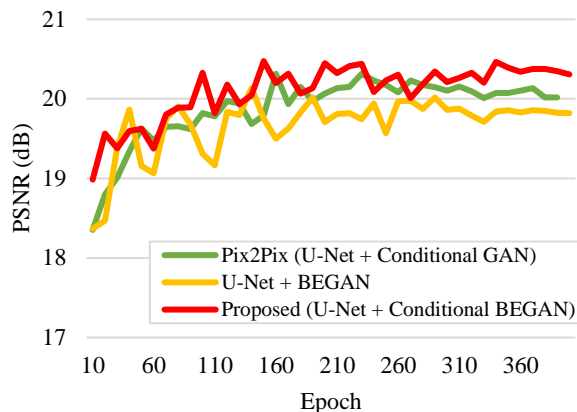


Figure 6. Average PSNR results according to the different discriminator networks for the validation images of the Indoor track of the NTIRE 2018 Dehazing Challenge [15].

of Figure 7-(c) shows a lot of stains on the refrigerator surface and the chair back. In the dehazed image of Figure 7-(d), the stains are less observed than the dehazed image of Figure 7-(c), but they still remain light.

#### 4.2.4 Comparison between the proposed conditional BEGAN and the conventional dehazing methods

We also compared the PSNR performance between the proposed conventional BEGAN and the conventional dehazing methods [3]-[4], [18], [23]-[28] for the 500 synthetic indoor hazy images of the Synthetic Objective Testing Set (SOTS) [29]. For fair comparison, our conditional BEGAN is newly trained using the REISDE training set which contains 13, 990 synthetic indoor hazy images [29]. Table II shows the average PSNR performance of our method and previous dehazing methods for 500 synthetic indoor hazy images of SOTS. Since the training set and the test dataset are composed of low-resolution hazy-clean pair images, we did not perform the down-scaling for input hazy images. As shown in Table II, it is noted that the proposed conditional BEGAN also shows the highest PSNR performance for dehazing of the small synthetic hazy images compared with the recent dehazing methods.

Table II. Average PSNR performance of our method and the other dehazing methods for 500 synthetic indoor hazy images of SOTS (red bold: 1<sup>st</sup> and blue bold: 2<sup>nd</sup>)

Dehazing Method	PSNR (dB)	SSIM
DCP (3)	16.62	0.8179
FVR (18)	15.72	0.7483
BDDR (23)	16.88	0.7913
GRM (24)	18.86	<b>0.8553</b>
CAP (25)	19.05	0.8364
NLD (26)	17.29	0.7489
DehazeNet (27)	<b>21.14</b>	0.8472
MSCNN (4)	17.57	0.8102
AOD-net (28)	19.06	0.8504
Proposed	<b>22.07</b>	<b>0.9187</b>

## 5. Conclusion

We have experimentally shown that the receptive field sizes are very important for the ultra-high resolution single image dehazing. Thus, we proposed the conditional BEGAN with the large receptive field sizes by downscaling the input hazy images. It is also observed that the PSNR performance of the proposed conditional BEGAN is higher than that of the network trained by the L1 loss and the perceptual loss without the GAN loss since the GAN loss works effectively in the severe hazy regions. Also, we experimentally showed that the proposed conditional BEGAN could generate the dehazed images of higher perceptual quality and has higher PSNR performance than the conditional GAN and the original BEGAN for the single image dehazing problems.

## 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] S. G. Narasimhan, and S. K. Nayar. Vision and Atmosphere. *Int. J. Comput. Vision*, 48, 233-254, 2002.
- [2] L. Schaul, *et al.* Color Image Dehazing using the Near-Infrared. *Proc. IEEE Int. Conf. on Computer Vision Pattern Recognition (CVPR)*, 325-332, 2001.
- [3] K. He, *et al.* Single Image Haze Removal using Dark Channel. *Proc. IEEE Int. Conf. on Computer Vision Pattern Recognition (CVPR)*, 1956-1963, 2009.
- [4] Ren, Wenqi, *et al.* Single image dehazing via multi-scale convolutional neural networks. *European conference on computer vision*. Springer, Charm, 154-169, 2016.
- [5] Swami, Kunal, and Saikat Kumar. CANDY: Conditional Adversarial Networks based Fully End-to-End System for Single Image Haze Removal. *arXiv preprint arXiv: 1081.02892*, 2018.
- [6] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. *In MICCAI*, 234-241, Springer, 2015.
- [7] Berthelot, David, Tom Schumm, and Luke Metz. Began: Boundary equilibrium generative adversarial networks. *arXiv preprint arXiv: 1703.10717*, 2017.
- [8] Isola, Phillip, *et al.* Image-to-image translation with conditional adversarial networks. *arXiv preprint*, 2017.
- [9] Johnson, Justin, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. *European Conference on Computer Vision (ECCV)*. Springer, Charm, 2016.
- [10] 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.
- [11] D. Pathak, P. Krahenbuhl, J. Donahue, T. Darrell, and A. A. Efros. Context encoders: Feature learning by inpainting. *IEEE Int. Conf. on Computer Vision Pattern Recognition (CVPR)*, 2016.
- [12] X. Wang and A. Gupta. Generative image modeling using style and structure adversarial networks. *European Conference on Computer Vision (ECCV)*, 2016.
- [13] Augustus Odena, Vincent Dumoulin, and Chris Olah. Deconvolution and Checkerboard Artifacts. <https://distill.pub/2016/deconv-checkerboard/>, 2016.
- [14] Kingma, P. Diederik, and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
- [15] 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*, 2018.
- [16] 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*, 2018.
- [17] Robby T. Tan, Visibility in bad weather from a single image. *IEEE Conference on Computer Vision and Pattern Recognition*, 2008.
- [18] J.-P. Tarel and N. Hautiere. Fast visibility restoration from a single color or gray level image. *IEEE ICCV*, 2009.
- [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. O. Ancuti and C. Ancuti, "Single image dehazing by multiscale fusion," *IEEE Transactions on Image Processing*, vol. 22(8), pp. 3271–3282, 2013.
- [21] K. Tang, J. Yang, and J. Wang. Investigating haze-relevant features in a learning framework for image dehazing. *IEEE Conference on Computer Vision and Pattern Recognition*, 2014.
- [22] S. Emberton, L. Chittka, and A. Cavallaro. Hierarchical rankbased veiling light estimation for underwater dehazing. *Proc. of British Machine Vision Conference (BMVC)*, 2015.
- [23] 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.
- [24] C. Chen, M. N. Do, and J. Wang. Robust image and video dehazing with visual artifact suppression via gradient residual minimization. *European Conference on Computer Vision*, 2016.
- [25] Q. Zhu, J. Mai, and L. Shao. A fast single image haze removal algorithm using color attenuation prior. *IEEE Transactions on Image Processing*, vol. 24, no. 11, pp. 3522–3533, 2015.
- [26] D. Berman, S. Avidan *et al.* Non-local image dehazing. *IEEE Conference on Computer Vision and Pattern Recognition*, 2016.
- [27] 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.
- [28] B. Li, X. Peng, Z. Wang, J. Xu, and D. Feng. Aod-net: All-in-one dehazing network. *IEEE International Conference on Computer Vision*, 2017.
- [29] Li, Boyi, *et al.* RESIDE: A Benchmark for Single Image Dehazing. *arXiv preprint arXiv:1712.04143*, 2017.