A GAN-Based Input-Size Flexibility Model for Single Image Dehazing

Shichao Kan[a,b,c,1], Yue Zhang[b,c,1], Fanghui Zhang[b,c] and Yigang Cen[b,c,*]

[a]*School of Computer Science and Engineering, Central South University, 410083, Changsha, Hunan, China*
[b]*Institute of Information Science, Beijing Jiaotong University, Beijing 100044, China*
[c]*Beijing Key Laboratory of Advanced Information Science and Network Technology, Beijing 100044, China*

ARTICLE INFO

ABSTRACT

Image-to-image translation based on generative adversarial network (GAN) has achieved state-of-the-art performance in various image restoration applications. Single image dehazing is a typical example, which aims to obtain the haze-free image of a haze one. This paper concentrates on the challenging task of single image dehazing. Based on the atmospheric scattering model, a novel model is designed to directly generate the haze-free image. The main challenge of image dehazing is that the atmospheric scattering model has two parameters, i.e., transmission map and atmospheric light. When they are estimated respectively, the errors will be accumulated to compromise the dehazing quality. Considering this reason and various image sizes, a novel input-size flexibility conditional generative adversarial network (cGAN) is proposed for single image dehazing, which is input-size flexibility at both training and test stages for image-to-image translation with cGAN framework. A simple and effective U-connection residual network (UR-Net) is proposed to combine the generator and adopt the spatial pyramid pooling (SPP) to design the discriminator. Moreover, the model is trained with multi-loss function, in which the consistency loss is a novel designed loss in this paper. Finally, a multi-scale cGAN fusion model is built to realize state-of-the-art single image dehazing performance. The proposed models receive a haze image as input and directly output a haze-free one. Experimental results demonstrate the effectiveness and efficiency of the proposed models.

1. Introduction

Haze removal [9] is a classical ill-posed image restoration problem, which plays an important role in intelligent transportation systems, e.g., object detection under haze conditions [18, 19, 23]. Haze is defined as some particles such as dust that obscure the clarity of the atmosphere. Dehazing is to remove the veil of haze from a haze image and restore a corresponding haze-free image. In recent years, because the development of deep learning has greatly improved the performance of image processing compared with non-learning-based technology, the problem of dehazing attracts more and more attentions in image restoration research community. Various image dehazing methods based on deep learning technology have been proposed, including: (1) Generating medium transmission map [3] or haze-free image [17, 23, 43] by a convolutional neural network (CNN); (2) generating transmission map [29] or haze-free image [24, 39, 42] based on encoder-decoder structure without adversary training; (3) reconstructing haze-free image with paired image-to-image translation models based on generative adversary network (GAN) [20, 28, 30, 41, 45]; (4) reconstructing haze-free image with unpaired image-to-image translation models based on cycle GAN (CGAN) [5, 22, 40].

In order to directly generate medium transmission map, [3] and [29] proposed an end-to-end learnable CNN model. To generate haze-free image from a haze one via an end-to-end manner, [17],[23] and [43] proposed light-weight and fast CNNs. [24], [39] and [42] incorporated some modern technologies into CNNs based on encoder-decoder structure. Usually, the real of image restoration is sub-optimal based on these models. In order to merge GAN [7] and image dehazing, supervised learning model with paired and unpaired samples based on adversary training are developed. [20], [28], [30], [41] and [45] are GAN-based end-to-end learnable models that trained with paired synthetic dataset, while [40], [5] and [22] are cycle-consistency models that trained with unpaired training dataset.

From these deep learning-based methods, we can see that: (1) Because the end-to-end dehazing models [17, 20, 24, 30, 39, 41, 42, 43, 45] can directly generate the haze-free image without additional parameter estimation, they are generally more efficient than non-end-to-end dehazing models [3, 29]; (2) Due to down-sampling and up-sampling process are not used before and after image dehazing with input-size flexibility models [17, 24], the information loss can be minimized throughout the restoration process. Thus, images generated with input-size flexibility models have a better visual effect than images generated with input-size fixed models [30, 41, 42, 43]; (3) Because the paired samples have definitive supervised information, the training of the network can be truly supervised when the paired samples are used. Thus, paired image-to-image translation models [20, 28, 30, 41, 45] are usually more effective than unpaired image-to-image translation models [5, 22, 40]; (4) Various

[*]Corresponding author
✉ kanshichao@csu.edu.cn (S. Kan);
17112065@bjtu.edu.cn (Y. Zhang); 18112013@bjtu.edu.cn (F. Zhang); ygcen@bjtu.edu.cn (Y. Cen)
ORCID(s): 0000-0003-0097-6196 (S. Kan)
[1]The first two authors (Shichao Kan and Yue Zhang) contribute equally.

works [20, 28, 30, 45] focus on exploring single image dehazing with GAN-based models and achieve promising performance. Considering these properties, we propose an end-to-end input-size flexibility conditional generative adversarial network (cGAN) for single image dehazing. The proposed model can not only remove the haze as much as possible but also preserve the clear content of an image.

The method proposed in this paper has obtained the state-of-the-art results on the datasets of the intelligent traffic video image enhancement processing competition[2] of ICIG 2019 and the more large scale REalistic Single Image DEhazing (RESIDE) dataset [19] for single image dehazing. The performance improvement primarily comes from the input-size flexibility training and test, multi-loss supervised training and the designed end-to-end framework.

Our works have the following contributions.

- An end-to-end input-size flexibility cGAN model is proposed for single image dehazing. The size of feature map in each layer of the generator is automated calculated based on the size of the input image. Based on our model, input-size flexibility mode can be applied to both adversary training and test stages and the image dehazing performance can be improved greatly.

- In our framework, a UR-Net structure is designed based on the popular U-Net [38] structure and residual learning [12], which is simple and effective. The generator is the iteration of UR-Net between two adjacent convolutional layers. Moreover, in order to realize input-size flexibility adversary training, the spatial pyramid pooling (SPP) [10] structure is embedded into the discriminator.

- Training with multi-loss functions is also an important part of our framework. We proposed a novel consistency loss to keep the transformation consistency between the generated dehazing image and the real input image, and combined adversary loss, L_1 loss, the structural similarity (SSIM) loss and a new peak signal to noise ratio (PSNR) loss to train our network. The effectiveness of these loss functions is verified by ablation studies.

The rest of this paper is organized as follows. In Section 2, related works about learning-based single image dehazing are reviewed. The idea, framework and details of the proposed input-size flexibility cGAN for single image dehazing are presented in Section 3. In Section 4, datasets, evaluation metrics and the experimental results are presented. Section 5 concludes the paper.

2. Related Work

Single image dehazing is a difficult vision task and has a long research history. Traditional single image dehazing methods are based on the handcrafted priors [6], e.g.,

dark channel prior [9, 31, 32], color attenuation prior [47] and non-local prior [2, 21], which are usually simple and effective for many scenes. However, prior-based methods are limited when describing specific statistics. In recent few years, learning-based methods are becoming popular because they can overcome the limitations of specific priors [4, 26]. We also oriented to study learning-based single image dehazing in this paper. Here, works related to them are reviewed in detail, including learning-based dehazing without and with GAN methods, respectively.

2.1. Learning-based Dehazing Without GAN

Learning-based dehazing methods become more and more popular since the learning idea was proposed by Tang *et al.*[37]. The original idea was learning a regression model based on random forests from prior-based haze-relevant features, such as dark channel [9], local max contrast [36], hue disparity [1], and local max saturation [37]. Subsequently, more powerful learning dehazing models were proposed, especially CNN-based end-to-end learning methods. Song *et al.* [35] proposed a ranking CNN to capture the statistical and structural attributes of hazy images, simultaneously. However, it is not an end-to-end learning system. Cai *et al.* [3] proposed an end-to-end learning system to directly generate a medium transmission map, which is based on the CNN framework and called DehazeNet. Ren *et al.* [29] proposed a coarse-to-fine multi-scale CNN (MSCNN) model to predict transmission maps. Although the two models can be learned via an end-to-end manner, they are not end-to-end dehazing models.

In 2017, Li *et al.* [17] proposed a light-weight, effective and fast end-to-end learning model for image dehazing, called AOD-Net, which can directly generate a haze-free image from a haze one. Since then, the end-to-end dehazing idea is favored by researchers. Based on the AOD-Net framework, Liu *et al.* [23] investigated various loss functions and demonstrated that training with perception-driving loss can further boost the performance of dehazing. Zhang *et al.* [42] proposed a multi-scale image dehazing method using a perceptual pyramid deep network based on an encoder-decoder structure with a pyramid pooling module. In this model, the designed network is based on dense blocks [13] and residual blocks [12], the perceptual loss is also incorporated into the training process. Xu *et al.* [39] proposed an instance normalization unit and embedded it into the VGG-based [33] U-Net [38] with an encoder-decoder structure. Liu *et al.* [24] proposed a generic model-agnostic CNN (GMAN) for signal image dehazing, which is based on the fully convolutional idea and is not rely on the atmosphere scattering model. Both Xu *et al.* and Zhang *et al.* are based on the mean squared error (MSE) and VGG-feature-based perceptual loss to train the network. Recently, Zhang and Tao [43] proposed a fast and accurate multi-scale end-to-end dehazing network called FAMED-Net, which is lightweight and computationally efficient.

Inspired by the success of these models, our proposed framework is based on the U-Net structure and residual learn-

---

[2]http://icig2019.csig.org.cn/?page_id=328

ing, which is also an end-to-end dehazing one. Different from the previous idea, our network is designed for generalized image restoration, especially for different sizes of images, which can accept input images of any size during both training and test processes.

2.2. Learning-based Dehazing With GAN

The idea of GAN was first proposed in [7], which is designed to synthesize realistic images via an adversarial process. Latter, it is widely extended to a variety of image generation tasks, such as conditional image generation [25], paired image-to-image translation [15], unpaired image-to-image translation [46], etc. Now, it is also becoming popular in single image dehazing. Zhang and Patel [41] proposed to jointly learn the transmission map, atmospheric light, and dehazing based on GAN framework, which is called densely connected pyramid dehazing network (DCPDN) and is an end-to-end single image dehazing model. Zhu *et al.* [45] formulated the atmospheric scattering model into a GAN framework and proposed a DehazeGAN, which can be used to learn the global atmospheric light and the transmission coefficient simultaneously. In order to generate realistic clear images, Li *et al.* [20] directly estimates the haze-free image based on an end-to-end trainable cGAN with encoder-decoder architecture. Ren *et al.* [30] adopted a fusion-based strategy to fuse three inputs from an original hazy image and proposed an end-to-end gated fusion network (GFN) for single image dehazing, which is trained with MSE and adversarial loss. Qu *et al.* [28] directly generate a haze-free image from a haze one without the physical scattering model, which is called enhanced pix2pix dehazing network (EPDN), and multi-loss function optimization idea is also used to train the network, including adversarial loss. All of these models are based on paired image-to-image translation framework.

Moreover, the unpaired image-to-image translation framework can be also found in single image dehazing. Yang *et al.* [40] proposed an end-to-end disentangled dehazing network to generate a haze-free image based on unpaired supervision. Engin *et al.* [5] completed the dehazing task based on unpaired supervision, which did not rely on the atmospheric scattering model and trained by combining cycle-consistency and perceptual losses. Liu *et al.* [22] developed an end-to-end learning system that uses unpaired fog and fog-free training images to generate a fog-free image, which also uses adversarial discriminators and cycle-consistency losses to train the whole framework. The advantage of unpaired supervision training is that the training process does not need to rely on synthetic dataset, because unpaired samples are easy to obtain. However, because these frameworks do not rely on the paired training data, the performance to restore realistic images is limited.

Therefore, our designed framework is based on paired cGAN, which is also incorporating multi-loss function optimization in it.

3. Input-Size Flexibility Conditional Generative Adversarial Network

Most of the previous single image dehazing models are based on the atmospheric scattering model, which tends to estimate the parameters of transmission map and atmospheric light. However, parameter estimation usually introduces estimation errors, which reduces the quality of restoration image. We thus develop an end-to-end and image-to-image translation model for single image dehazing, which is independent of the atmospheric scattering model and there is no additional parameter estimation. The proposed model directly produces a haze-free image from a haze one and is input-size flexibility at both training and test stages. In the following, the atmospheric scattering model is first analyzed and then each component of our proposed input-size flexibility conditional generative adversarial network is presented, respectively, i.e., generator, discriminator and loss functions.

3.1. The Analysis of Atmospheric Scattering Model

The famous atmospheric scattering model [27] can be formulated as follows:

$$I_{re}(x) = J_{re}(x)t(x) + \alpha(1 - t(x)), \tag{1}$$

where $I_{re}(x)$ is the real haze image that need to be restored, $J_{re}(x)$ is the expected haze-free image that could be recovered from $I_{re}(x)$, $t(x)$ is the medium transmission map, α is the global atmospheric light and x is the indexes of the pixels corresponding to an image (I_{re}, J_{re} and t). In real tasks, only $I_{re}(x)$ of Eq.(1) is known, the other three variables are unknown. Because the final goal is to estimate $J_{re}(x)$, thus if $t(x)$ and α can be estimated, then one can directly obtain the $J_{re}(x)$ according the following formula:

$$J_{re}(x) = \frac{1}{t(x)} I_{re}(x) + \alpha(1 - \frac{1}{t(x)}) \tag{2}$$

However, estimating $t(x)$ is a complex task because $t(x)$ is related with both the distance $d(x)$ from the scene point to the camera and the scattering coefficient β of the atmosphere, which can be formulated as follows:

$$t(x) = e^{-\beta d(x)} \tag{3}$$

Moreover, there will always exist an error in the estimation of each parameter. Suppose that δ_1 and δ_2 are the average estimation errors of parameters t and α, respectively. When Eq.(2) is used to obtain a haze-free image, if the total average estimated error is δ, then we have:

$$\delta = \delta_1 + \delta_2 + \delta_1 \cdot \delta_2 \tag{4}$$

From Eq.(4), only both $\delta_1 \rightarrow 0$ and $\delta_2 \rightarrow 0$, we can obtain $\delta \rightarrow 0$. When the estimating parameters are more than one in a system, the estimated error of each parameter is usually difficult to control simultaneously. In order to estimate them by an end-to-end manner, a new framework
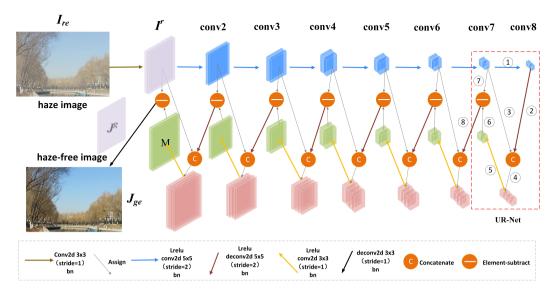
Figure 1: The designed generator (UR-Net-7) of the proposed input-size flexibility conditional generative adversarial network.

with a novel consistency loss (Eq.(7)) is designed. Next, we will analyze it.

Through log transformation, Eq.(2) can be transformed to the following form:

$$\log(J_{re}(x) - \alpha) = \log(I_{re}(x) - \alpha) - \log(t(x)), \quad (5)$$

By setting $J^g(x) = \log(J_{re}(x) - \alpha)$, $I^r(x) = \log(I_{re}(x) - \alpha)$ and $M(x) = \log(t(x))$, Eq.(5) can be rewritten as follows:

$$J^g(x) = I^r(x) - M(x), \quad (6)$$

In this paper, Encoder-decoder idea is used to realize dehazing task. According to Eq.(6), if we assume that I^r is one layer output of the encoder network with input I_{re}, and J^g is one layer output of the decoder network. We can obtain J_{ge} according to the following rule: $I^r = \log(I_{re} - \alpha) \Rightarrow \alpha = I_{re} - \exp(I^r)$ then $J^g = \log(J_{ge} - \alpha) \Rightarrow J_{ge} = I_{re} - [exp(I^r) - exp(J^g)]$. In the following, we design our framework based on this observation and Eq.(6).

3.2. The Generator of the Proposed Input-size Flexibility cGAN

As derived in Eq.(6), residual idea is an important component of our generator. An U-connection residual network (UR-Net) is designed for single image dehazing, and the whole generator is the iteration of UR-Net between two adjacent convolutional layers. Fig. 1 is the framework of our designed generator.

The unit of the red dotted rectangle in Fig. 1 is an example of the designed UR-Net. Step ① is a convolutional layer with kernel of 5×5 and stride 2. Suppose that the shape of conv7 is $(1, c_7, h_7, w_7)$ and the shape of conv8 is $(1, c_8, h_8, w_8)$, then we have $h_8 = \lceil \frac{h_7}{2} \rceil$, $w_8 = \lceil \frac{w_7}{2} \rceil$, where $\lceil \cdot \rceil$ is an up-round symbol. This is implemented by using the "same" padding operation in TensorFlow. Step ② is a de-convolutional layer with a kernel of 5×5 and stride 2.

Suppose that the output shape of step ② is $(1, c_8^o, h_8^o, w_8^o)$, then we have $h_8^o = h_7$, $w_8^o = w_7$ by setting $h_8^o = \frac{h_8+1}{2}$, $w_8^o = \frac{w_8+1}{2}$. Next, we concatenate the output of step ② and conv7 in the channel dimension, the output of step ④ is the concatenated result. This is the idea of U-Net for the purpose of fine information recovery. In order to realize residual learning between the input (conv7 in this example) and output of the penultimate layer of UR-Net, we need to ensure that the output channel dimension of the penultimate layer equals to the input. Thus, in step ⑤, a convolutional layer with kernel of 3×3 and stride 1 is used to reduce the channel dimension of step ④, and the output size of step ⑤ is equal to its input size by using the "same" padding operation in TensorFlow. Finally, the residual can be obtained by the subtraction between conv7 and the output of step ⑤. Moreover, batch normalization (bn) [14] is used to each convolutional and de-convolutional layer in our framework for the purpose of fast convergence. The activation function of the last layer is tanh(·), other layers are leak ReLU (Lrelu) and the value of leak is set to 0.2.

In order to provide noise to realize conditional input, dropout operation is used at both training and test stage after the de-convolutional layers corresponding to conv6, conv7 and conv8. The dropout rate is set as 0.5. In Fig. 1, the height h_i and width w_i of each conv$i$ is related to the height h and width w of an input image, the calculation formulas are $h_i = \frac{h+2^i-1}{2^i}$ and $w_i = \frac{w+2^i-1}{2^i}$. The designed generator is an encoder-decoder structure, conv1 to conv8 form the encoder, the other parts form the decoder.

For convenience, in the following, UR-Net-K is used to indicate that the number of UR-Net structure in the generator is K (e.g., the generator of Fig. 1 has 7 UR-Net structure, thus we call it UR-Net-7). At the same time, UR-Net-K^* is used to represent that there is no subtraction process in the last UR-Net of the generator (the last UR-Net is located at
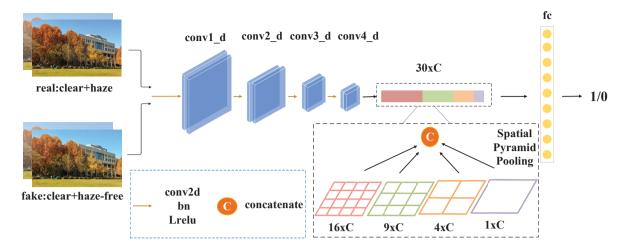
Figure 2: The designed discriminator of the proposed input-size flexibility conditional generative adversarial network.

the last de-convolutional layer).

The purpose of generator (G) is generating a haze-free output image J_{ge} based on the input haze image I_{re} and random noise Z, i.e., $G : \{I_{re}, Z\} \rightarrow J_{ge}$.

3.3. The Discriminator of the Proposed Input-size Flexibility cGAN

The discriminator D is an important part of our proposed input-size flexibility cGAN model, which is used to discriminate the input sample is a "real image" (J_{re}) or a "generated image" (J_{ge}). As shown in Fig. 2, it consists of an input layer, 4 convolutional layers, a spatial pyramid pooling (SPP) [11] layer and a fully convolutional layer (fc). One image of the input layer is the concatenation of real clear image and real haze image across the channel dimension, another one is the concatenation of real clear image and the generated haze-free image across the channel dimension. The first three convolutional layers have a convolutional operation with kernel of 5×5 and stride 2, the last convolutional layer have a convolutional operation with kernel of 5×5 and stride 1. The SPP layer is designed to pool different sizes of input feature maps into vectors of the same length (the level of SPP is set as 4), thus training with input-size flexibility can be realized. The fc layer is a classifier to discriminate whether the input sample is real or fake (generated).

3.4. Multi-loss Function

The idea of multi-loss function optimization is widely used in various CNN-based systems, which is proved effective in different kinds of applications. It is also used in our framework. Next, we will define them one-by-one.

To ensure that $I^r = \log(I_{re} - \alpha)$ and $J^g = \log(J_{ge} - \alpha)$, we need to constrain $I_{re} - \exp(I^r) = J_{ge} - \exp(J^g)$. Thus, we define a consistency loss as follows:

$$\mathcal{L}_{Consistency}(G) = ||I_{re} - \exp(I^r) - J_{ge} + \exp(J^g)||_1 \quad (7)$$

This consistency loss function is to ensure that the transformations of I_{re} and J_{ge} in the network are approximated to

the log transformation with parameter α, which is novel and important for our framework. Instead of learning the parameter of α, by this consistency loss, a convolutional layer is developed to estimate the transformation of $\log(I_{re} - \alpha)$ and the inverse transformation of $\log(J_{ge} - \alpha)$, respectively.

Then, we adopt the general cGAN loss function [15] in our model, which is defined as follows:

$$\mathcal{L}_{cGAN}(G, D) = \mathbb{E}_{I_{re}, J_{re}}[\log D(I_{re}, J_{re})] + \\ \mathbb{E}_{I_{re}, Z}[\log(1 - D(I_{re}, G(I_{re}, Z)))] \quad (8)$$

At the training stage, the generator G is trained to produce outputs that cannot be distinguished as "fakes" by the discriminator D, and D is trained to distinguish the generated example as "fakes". Thus, G tries to minimize objective (8) against an adversarial D that tries to maximize it, i.e., $G^* = arg \min_G \max_D \mathcal{L}_{cGAN}(G; D)$. In the last term of (8), minimizing G is equivalent to maximizing $\log(D(I_{re}, G(I_{re}, Z)))$, which is adopted at the implementation stage.

Because the L_1 loss function can constrain the output of the generator absolutely equal to the expected output thus reduce the blur. We also introduce it as one of our loss functions, as follows:

$$\mathcal{L}_{L_1}(G) = \mathbb{E}_{I_{re}, J_{re}, Z}[||J_{re} - G(I_{re}, Z)||_1] \quad (9)$$

Moreover, perception-driving losses are verified effective in various image restoration tasks. Thus, in order to make the generated haze-free images have a good visual effect, we adopt SSIM and PSNR loss to construct our perception losses. In our model, the calculation formula of SSIM is the same as [44]. PSNR is defined as:

$$PSNR(J_{re}, J_{ge}) = 10 \cdot log_{10}(\frac{(\max(J_{re}) - \min(J_{re}))^2}{MSE(J_{re}, J_{ge})}), \quad (10)$$

where $MSE(J_{re}, J_{ge})$ is the mean of $(J_{re} - J_{ge})^2$ and can be formulated as $MSE(J_{re}, J_{ge}) = mean((J_{re} - J_{ge})^2)$.
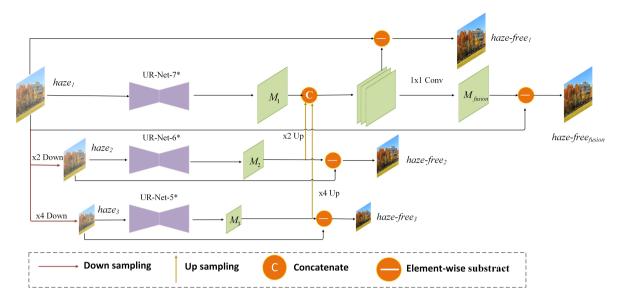
Figure 3: The fusion model of multi-scale generator.

According to the above formula, SSIM and PSNR losses are defined as follows:

$$\mathcal{L}_{SSIM}(G) = 1 - SSIM(J_{re}, J_{ge}), \tag{11}$$

$$\mathcal{L}_{PSNR}(G) = 1 - \frac{PSRN(J_{re}, J_{ge})}{thresh}, \tag{12}$$

where *thresh* is a threshold that is set to 40 in our experiments.

Finally, the overall loss function of our model is defined as follows:

$$\mathcal{L} = \mathcal{L}_{Consistency}(G) + \lambda_1 \mathcal{L}_{cGAN}(G, D) + \lambda_2 \mathcal{L}_{L_1}(G) + \\ \lambda_3 \mathcal{L}_{SSIM}(G) + \lambda_4 \mathcal{L}_{PSNR}(G) + \lambda ||w||_2^2, \tag{13}$$

where λ_1, λ_2, λ_3 and λ_4 are weights of their corresponding loss functions, which are set to 1, 100, 100 and 100 in our experiments, respectively. The final goal is to minimize (13). The last term is only used to multi-scale training stage (Section 3.5), in which w is the weights of the generator and λ is the weight of this term.

3.5. Multi-scale Generator Fusion

Multi-scale fusion is verified effective in image dehazing by Zhang and Tao [43]. In their model, a Gaussian pyramid architecture with a late fusion module is designed to fuse different estimated feature maps. Here, the Gaussian pyramid architecture is also used to realize multi-scale generator fusion model, as shown in Fig. 3. This fusion model aims to show the generalization of our model to multi-scale framework. It should be noted that FAMED-Net[43] is trained without adversarial, our multi-scale generator is trained based on our cGAN framework. The input of the generator includes one haze image (haze$_1$) and the corresponding down-sampled images ($\frac{1}{2}$ scale (haze$_2$) and $\frac{1}{4}$ scale (haze$_3$)). The

output of the generator includes 4 haze-free images (haze-free$_1$, haze-free$_2$, haze-free$_3$, and haze-free$_{fusion}$ in Fig. 3), which corresponds to the original scale of input haze image, $\frac{1}{2}$ scale, $\frac{1}{4}$ scale, and multi-scale fused output. The module of multi-scale fusion is performed based on the concatenation of haze maps (M_1, M_2, M_3 in Fig. 3) of original scale, 2× up-sampling of $\frac{1}{2}$ scale, 4× up-sampling of $\frac{1}{4}$ scale. The fused haze map (M_{fusion}) is obtained after applying a convolutional layer with 1×1 kernel to the concatenated haze maps. In this fusion model, the down-sampling and up-sampling operations are performed with bicubic interpolation. The generators of original scale, $\frac{1}{2}$ down-sampling scale and $\frac{1}{4}$ down-sampling scale of haze images are UR-Net-7*, UR-Net-6* and UR-Net-5*, respectively.

The discriminator is also vital for the fusion generator. Because the designed discriminator is input-size flexibility, thus we have two alternative of discriminator for the fusion generator, i.e., with and without sharing parameters for each output of the generator. Although sharing parameters of discriminator can reduce the model size, it can not reduce the number of computations. In order to enhance the discriminant ability of this fusion generator, we directly adopt the discriminator model without sharing parameters, i.e., each output of the fusion generator is discriminated by different discriminators.

For the loss function of the fusion generator, we apply objective (13) to each output of the generator and corresponding discriminator.

3.6. Model Training

The model is implemented based on TensorFlow, and is trained with minibatch SGD (Stochastic Gradient Descent). The Adam solver [16] with a learning rate of 0.0002 and momentum parameters $\beta_1 = 0.5$, $\beta_2 = 0.999$ is applied to optimize our model. All parameters are trained from scratch

and the batch size is set as 1. The hyper-parameter λ in objective (13) is set as 0.001. In order to better maintain the convergence balance between the generator and the discriminator, we update parameters of the discriminator once every 4 iterations. Because our model is input-size flexibility, we can train this model by images with different sizes to obtain a better haze-free image. However, training with different sizes of input is slow, thus the model is first trained with fixed size of input images and then fine-tuned based on different sizes of images.

4. Experiments

We conduct experiments on the dataset of intelligent traffic video image enhancement processing competition of ICIG 2019 (we call it ICIG2019 for convenience) and the large scale REalistic Single Image DEhazing (RESIDE) dataset [19] for single image dehazing.

The **ICIG2019** dataset contains 5500 clear images of real scene and corresponding synthetic haze ones. The training and validation sets contain 5000 and 500 image pairs, respectively. In the experiments, we use the training set to train our models and use the validation set to test the trained models. Ablation studies are conducted on this dataset.

The **RESIDE** dataset is one of the largest single image dehazing datasets, which contains 110,500 synthetic hazy indoor images (ITS) and 313,950 synthetic hazy outdoor images (OTS) in the training set. The synthetic objective testing set (SOTS) contains 500 indoor images and 500 outdoor images. The hybrid subjective testing set (HSTS) contains 10 real-world images and 10 synthetic images. In the training dataset, each clear image corresponds to multiple haze images of different concentrations. For each clear image, we randomly select a corresponding haze image from the training samples to form our training set.

We use 4 evaluation metrics that have been realized in the skimage package of python to evaluate the performance of single image dehazing, which are MSE (the smaller the better ↓), normalization root mean-squared error (NRMSE) (the smaller the better ↓), PSNR (the larger the better ↑) and SSIM (the larger the better ↑). We also re-test the compared methods by running the corresponding released models. All the results reported for the compared methods and our methods are evaluated using the standard evaluation interface of python for a fair comparison.

4.1. Ablation Studies

The experiments of ablation study is conducted on the ICIG2019 dataset. We first verify the effectiveness of each loss function in Eq.(13) based on the model of UR-Net-7 (Fig. 1) and the input image with size of 256×256. The maximum training epoch is set as 16. Based on the $\mathcal{L}_{Base} = \mathcal{L}_{Consistency}(G) + \mathcal{L}_{cGAN}(G, D)$ loss function in our framework, we verify the combinations of $\mathcal{L}_{Base} + \mathcal{L}_{L_1}$, $\mathcal{L}_{Base} + \mathcal{L}_{SSIM}$, $\mathcal{L}_{Base} + \mathcal{L}_{PSNR}$, $\mathcal{L}_{Base} + \mathcal{L}_{L_1} + \mathcal{L}_{SSIM}$ (without $\mathcal{L}_{PSNR}$), $\mathcal{L}_{Base} + \mathcal{L}_{L_1} + \mathcal{L}_{PSNR}$ (without $\mathcal{L}_{SSIM}$), $\mathcal{L}_{Base} + \mathcal{L}_{SSIM} + \mathcal{L}_{PSNR}$ (without $\mathcal{L}_{L_1}$), and $\mathcal{L}_{Base} + \mathcal{L}_{L_1} + \mathcal{L}_{SSIM} + \mathcal{L}_{PSNR}$ ($\mathcal{L}$). The experimental results are shown in Table 1.

Table 1

Results of Different Losses on The ICIG2019 Dataset.

loss	MSE ↓	NRMSE ↓	PSNR ↑	SSIM ↑
$\mathcal{L}_{Base}$	637.6	0.168	21.54	0.789
$\mathcal{L}_{Base} + \mathcal{L}_{L_1}$	382.8	0.135	23.21	0.866
$\mathcal{L}_{Base} + \mathcal{L}_{SSIM}$	383.3	0.135	23.26	0.894
$\mathcal{L}_{Base} + \mathcal{L}_{PSNR}$	336.3	0.126	24.03	0.890
without $\mathcal{L}_{PSNR}$	321.1	0.123	24.15	0.899
without $\mathcal{L}_{SSIM}$	299.2	0.119	24.45	0.898
without $\mathcal{L}_{L_1}$	337.8	0.126	23.87	0.902
$\mathcal{L}$	**287.9**	**0.116**	**24.58**	**0.904**

Table 2

Results of Different Input Sizes With and Without Input-size Flexibility Fine-tuning on The ICIG2019 Dataset.

Training mode	MSE ↓	NRMSE ↓	PSNR ↑	SSIM ↑
256×256	287.9	0.116	24.58	0.904
256×256 + IFF	245.1	0.108	25.04	0.905
368×544	300.8	0.118	24.53	0.898
368×544 + IFF	**219.6**	**0.101**	25.58	0.905
512×512	317.1	0.116	24.67	0.891
512×512 + IFF	223.4	**0.101**	**25.67**	**0.906**

From Table 1, it can be seen that the best performance is obtained when all the losses are used. The performances of $\mathcal{L}_{Base} + \mathcal{L}_{L_1}$, $\mathcal{L}_{Base} + \mathcal{L}_{SSIM}$ and $\mathcal{L}_{Base} + \mathcal{L}_{PSNR}$ are much better than the performance of $\mathcal{L}_{Base}$, which verified the effectiveness of each loss function after combined with $\mathcal{L}_{Base}$. The performance without $\mathcal{L}_{PSNR}$ is much better than the performances of $\mathcal{L}_{Base} + \mathcal{L}_{L_1}$ and $\mathcal{L}_{Base} + \mathcal{L}_{SSIM}$, the performance without $\mathcal{L}_{SSIM}$ is much better than the performances of $\mathcal{L}_{Base} + \mathcal{L}_{L_1}$ and $\mathcal{L}_{Base} + \mathcal{L}_{PSNR}$, and the performance without $\mathcal{L}_{L_1}$ is much better than the performances of $\mathcal{L}_{Base} + \mathcal{L}_{SSIM}$ and $\mathcal{L}_{Base} + \mathcal{L}_{PSNR}$, which verified the effectiveness of combining any two loss functions with $\mathcal{L}_{Base}$. Moreover, we notice that the values of MSE and PSNR of $L_{Base} + L_{PSNR}$ are much better than $L_{Base} + L_{L_1}$ and $L_{Base} + L_{SSIM}$, which shows that the proposed PSNR loss is much better than the L_1 loss and L_{SSIM} loss when they combined with L_{Base}, respectively.

The second ablation experiments are the generator network with fixed sizes of input images and fine-tuned with input-size flexibility images, which aims to verify the effectiveness of input-size flexibility. The experimental results are shown in Table 2.

In Table 2, the IFF is the abbreviation of input-size flexibility fine-tuning. The training mode indicates the sizes of training input. The test results are based on the mode of input-size flexibility, i.e., the output size of an image equals to the size of the input image. From Table 2, we can see that with the input-size flexibility fine-tuning, better performances can be obtained. Moreover, the best MSE is obtained with the training mode of 368×544 + IFF, the best

Table 3
Comparison With The State-of-the-art Methods on The Validation of ICIG2019 Dataset.

Methods	MSE ↓	NRMSE ↓	PSNR ↑	SSIM ↑
MSCNN [29]	1292	0.250	17.33	0.810
DCPDN [41]	971.2	0.218	19.06	0.848
GFN [30]	766.7	0.176	20.96	0.828
De-cGAN [20]	764.4	0.174	21.02	0.857
AOD-Net [17]	646.8	0.175	20.73	0.868
GMAN [24]	290.2	0.118	24.37	0.887
GMAN fine-tuned	287.1	0.118	24.43	0.891
FAMED-Net [43]	249.5	0.107	25.17	0.909
UR-Net-7	**223.4**	**0.101**	**25.67**	0.906
Multi-scale cGAN	**213.6**	**0.099**	**25.89**	**0.912**

Table 4
Comparison With The State-of-the-art Methods on The Outdoor of SOTS Dataset.

Methods	MSE ↓	NRMSE ↓	PSNR ↑	SSIM ↑
MSCNN [29]	812.2	0.202	20.02	0.880
DCPDN [41]	828.1	0.204	19.93	0.858
GFN [30]	676.2	0.172	21.47	0.849
De-cGAN [20]	611.1	0.160	21.96	0.868
AOD-Net [17]	693.0	0.185	20.47	0.899
FAMED-Net [43]	199.6	0.098	26.17	**0.925**
UR-Net-7	**160.5**	**0.089**	**26.96**	0.910
Multi-scale cGAN	**152.3**	**0.086**	**27.28**	**0.925**

Table 5
Comparison With The State-of-the-art Methods on The Synthetic of HSTS Dataset.

Methods	MSE ↓	NRMSE ↓	PSNR ↑	SSIM ↑
MSCNN [29]	1164.2	0.233	18.47	0.813
DCPDN [41]	841.5	0.197	20.21	0.852
GFN [30]	527.6	0.147	22.83	0.887
De-cGAN [20]	498.5	0.145	22.85	0.869
AOD-Net [17]	711.1	0.181	20.56	0.887
FAMED-Net [43]	168.9	0.089	26.68	**0.922**
UR-Net-7	**146.3**	**0.083**	**27.04**	0.908
Multi-scale cGAN	**105.9**	**0.071**	**28.38**	0.919

PSNR and SSIM are obtained with the training mode of 512×512 + IFF. The size of 368×544 is the mean size of the training images (368 is the mean of heights and 544 is the mean of widths). Moreover, we can see that when IFF is not used, the best MSE is obtained by the model trained with input size of 256×256. These experimental results show that if IFF is used, the process of pre-training with larger input size can lead to higher PSNR and SSIM, otherwise, the model trained with smaller input size can lead to a smaller MSE. However, both pre-training and fine-tuning are complex processes. The conclusion that can be determined from the experiments is that the performances with IFF are better than the performances without IFF. Other conclusions may be related to the experimental parameter settings.

4.2. Comparison With the State-of-the-Art Methods

We compare the proposed method with the state-of-the-art CNN-based dehazing methods. They are AOD-Net [17], MSCNN [29], GMAN [24], DCPDN [41], De-cGAN [20], GFN [30], and recently proposed FAMED-Net [43]. The comparison results of ICIG2019 dataset are shown in Table 3.

In Table 3, the method of GMAN fine-tuned means the fine-tuned model of GMAN on the ICIG2019 dataset based on the pre-trained GMAN model. From Table 3, we can see that the proposed UR-Net-7 is much better than the previous proposed methods for the evaluations of MSE, NRMSE and PSRN. The best SSIM is obtained by the proposed multi-scale cGAN, followed by the method of FAMED-Net. Moreover, we notice that after the GMAN model is fine-tuned (GMAN fine-tuned) on the ICIG2019 dataset, the performance are better than the GMAN without fine tuning (GMAN (SPL19)).

In these comparison methods, both AOD-Net and GMAN are input-size flexibility at the test stage. But they are not input-size flexibility at the training stage, one reason is that the batch-size of them is greater than 1 to obtain a good performance. The performance of them will drop a lot if the batch-size is set as 1 for input-size flexibility purpose at

the training stage, because batch-normalization is adopted to realize good performance by using large batch-size. The FAMED-Net is designed based on the AOD-Net, it can be also changed to input-size flexibility mode at the test stage, because late fusion idea is adopted, better performance can be obtained. Different from these works, the proposed model is GAN-based input-size flexibility, which is input-size flexibility at both training and test stages. Moreover, the proposed multi-scale cGAN obtained the best single image dehazing performance based on the evaluations in this paper, which also proved the effectiveness of image late fusion. Different from previous fusion idea, the proposed fusion framework is based on cGAN, which is a cGAN fusion framework.

Table 4 and Table 5 are the comparisons of the outdoor of SOTS and HSTS on the RESIDE dataset. It can be seen that the proposed multi-scale cGAN obtains the best results for the evaluations of MSE, NRMSE and PSNR. For the evaluation of SSIM, the best value is obtained by the method of FAMED-Net (both in Table 4 and Table 5) and multi-scale cGAN (in Table 4). Although the SSIM of the proposed multi-scale cGAN is less than the FAMED-Net 0.03% in Table 5, the MSE, NRMSE, and PSNR values of the proposed multi-scale cGAN are much higher than the FAMED-Net. In particular, the PSNR is 1.34dB higher.

Fig. 4 is the subjective comparisons on synthetic hazy images from the ICIG2019 validation set. From these de-
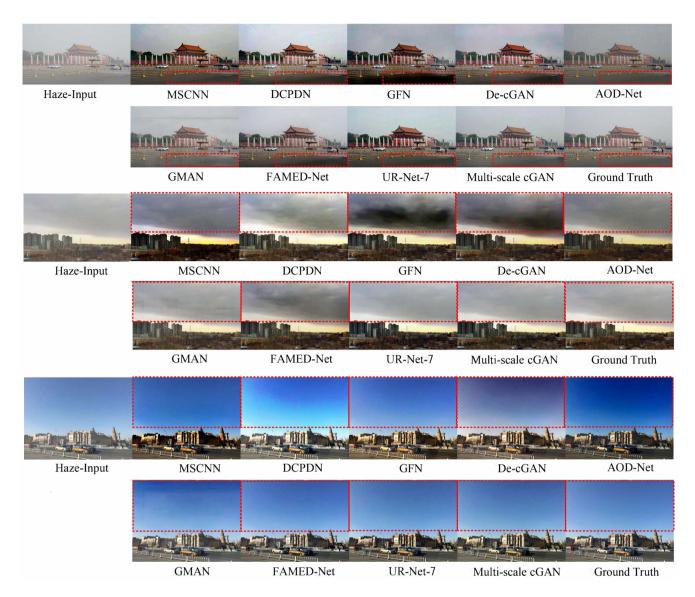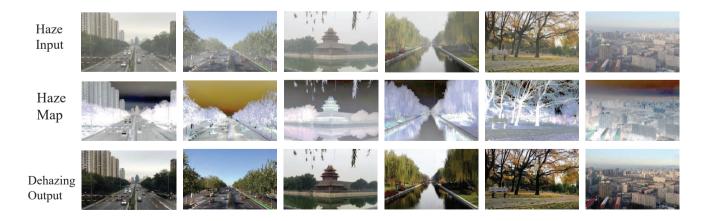
Figure 4: Subjective comparisons between the proposed methods and the most related state-of-the-art methods on synthetic hazy images from ICIG2019 validation set. Best viewed in color.
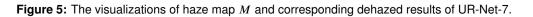

Figure 5: The visualizations of haze map M and corresponding dehazed results of UR-Net-7.

Figure 6: The visualizations of haze map M_{fusion} and corresponding dehazed results of Multi-scale cGAN.

Table 6
Comparison With The State-of-the-art Methods on The Indoor of SOTS Dataset.

Methods	MSE ↓	NRMSE ↓	PSNR ↑	SSIM ↑
MSCNN [29]	2097.5	0.383	16.00	0.780
AOD-Net [17]	1144.8	0.271	19.07	0.824
GFN [30]	443.0	0.175	22.48	0.888
FAMED-Net [43]	361.4	0.153	23.63	**0.901**
UR-Net-7	**274.1**	**0.139**	**24.42**	0.881
Multi-scale cGAN	**265.3**	**0.132**	**24.56**	0.900

Table 7
Learnable Parameters and Time Spent of Different Methods.

Methods	Params	Time(second)
MSCNN [29]	8,014	0.04
AOD-Net [17]	1,833	0.004
De-cGAN [20]	1.23×10^8	0.05
DCPDN [41]	6.69×10^7	0.04
GFN [30]	514,415	0.05
FAMED-Net [43]	17,991	0.03
UR-Net-7	8.59×10^7	0.04
Multi-scale cGAN	2.06×10^8	0.1

hazed images, we can see that our methods (especially multi-scale cGAN) are relatively good for the ground, the clouds and the sky.

Table 6 is the comparisons of the indoor of SOTS on the RESIDE dataset. According to Table 6, we can see that the best SSIM is obtained by the FAMED-Net. However, the best values of MSE, NRMSE, and PSNR are obtained by the proposed UR-Net-7 and multi-scale cGAN.

Table 7 summarizes the learnable parameters and time spent of different models based on the Tesla K40c GPU. The numbers of learnable parameters of our methods are

larger than other methods. This is because our model includes discriminators. However, the discriminators are not used during test. The time spent of our UR-Net-7 is similar to most of other state-of-the-art learning-based methods.
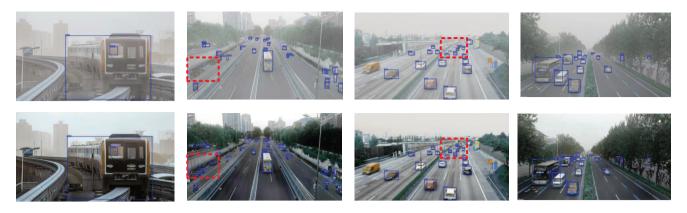
4.3. Discussions

As analysis in Section 3.1, the general atmospheric scattering model can be simplified to Eq.(6). According to Eq.(6), one haze image (I) can be seen as a clear image (J) plus a content related noise image M, which is a general additive noise model. For CNN-based image denoising or restoration, most of the noise models can be transformed into an additive noise model, e.g., the multiplicative noise model can be transformed into an additive noise model by logarithmic transformation. Thus, the proposed input-size flexibility cGAN is a general image restoration model.

The haze map (M) of Eq.(6) can be thought of as a kind of content related noise in a haze image. The visualizations of M are shown in Fig. 5. The haze maps in Fig 5 is the transformed results of M, which is same as the transformation of J^g, i.e., add 1 and multiply by the mean. From Fig. 5, it can be seen that the haze maps relate with the color, illumination and the concentration of haze, also the content of the corresponding haze images. Similar to the haze of real scenes, there is no specific rule for these generated haze maps. Moreover, the visualizations of M_{fusion} for multi-scale cGAN are also shown in Fig. 6, the characteristics of these haze maps are similar to those in Fig. 5.

Considering the applicability, image dehazing can usually be used to the preprocess step of other computer vision tasks. The proposed image dehazing algorithm can be used to assist object detection, as shown in Fig.7, which is the comparison of object detection results before and after dehazing with the proposed UR-Net-7. The detection algorithm is SNIPER [34], we only use the released code[3] and the pre-trained model for detection. From the detection results of the two images in the middle of Fig.7, we can see

---

[3]https://github.com/mahyarnajibi/SNIPER

Figure 7: The detection results before (the first line) and after (the second line) dehazing with UR-Net-7.
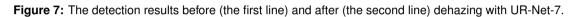

Figure 8: The dehazing results for images of nighttime or low-light conditions based on the UR-Net-7 model trained on the RESIDE outdoor training set.

Table 8

Summary of Major Contributions for Performance Improvement Based on The Experimental Results of ICIG2019 Dataset.

Methods	MSE ↓	NRMSE ↓	PSNR ↑	SSIM ↑
Baseline	637.6	0.168	21.54	0.789
+ Mulit-loss	287.9	0.116	24.58	0.904
+ IFF	223.4	0.101	25.67	0.906

that more objects can be detected after dehazing with UR-Net-7 (the objects in the red rectangle).

Moreover, the UR-Net-7 model trained on the RESIDE outdoor training set is used to test haze images in nighttime or low-light conditions. Results are shown in Fig. 8, it can be seen that the proposed model can be generalized to the haze scene under special conditions.

Finally, the major contributions of our work for performance improvement based on the experimental results of ICIG2019 dataset are summarized in Table 8. We can see that optimization with multi-loss functions greatly improves the performance of baseline, boosted 3.04 dB and 12.5% for PSNR and SSIM evaluation metrics, respectively. Moreover, input-size flexibility fine-tuning (IFF) can further improve the PSNR about 1.09 dB.

5. Conclusions

In this paper, an input-size flexibility cGAN with multi-loss function training model is developed for single image dehazing, experimental results proved the effectiveness of input-size flexibility and multi-loss function optimization. Moreover, a multi-scale image restoration fusion framework based on cGAN was proposed and verified for single image dehazing. Experimental results showed that we obtained the best single image dehazing performance on the ICIG2019 and RESIDE datasets. On the ICIG2019 dataset, the PSNR has been improved 0.5dB and 0.72dB for UR-Net-7 and Multi-scale cGAN compared with the FAMED-Net, respectively. On the outdoor of SOTS dataset, the PSNR has been improved 0.79dB and 1.11dB for UR-Net-7 and Multi-scale cGAN compared with the state-of-the-art methods, respectively. Our basic idea is to realize image restoration based on Eq.(6), thus the proposed framework can be also used to other image restoration tasks, such as image denoising, image deblurring, and image fusion [8]. Future works could be focused on extending our methods to other image restoration tasks.

References

[1] Ancuti, C.O., Ancuti, C., Hermans, C., Bekaert, P., 2010. A fast semi-inverse approach to detect and remove the haze from a single image, in: Computer Vision - ACCV 2010 - 10th Asian Conference on Computer Vision, Queenstown, New Zealand, November 8-12, 2010, Revised Selected Papers, Part II, pp. 501–514.

[2] Berman, D., Treibitz, T., Avidan, S., 2016. Non-local image dehazing, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016, pp. 1674–1682.

[3] Cai, B., Xu, X., Jia, K., Qing, C., Tao, D., 2016. Dehazenet: An end-to-end system for single image haze removal. IEEE Trans. Image Processing 25, 5187–5198.

[4] Ding, X., Liang, Z., Wang, Y., Fu, X., 2021. Depth-aware total variation regularization for underwater image dehazing. Signal Process. Image Commun. 98, 116408.

[5] Engin, D., Genç, A., Ekenel, H.K., 2018. Cycle-dehaze: Enhanced cyclegan for single image dehazing, in: 2018 IEEE Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops 2018, Salt Lake City, UT, USA, June 18-22, 2018, pp. 825–833.

[6] Gao, Y., Hu, H., Li, B., Guo, Q., Pu, S., 2019. Detail preserved single image dehazing algorithm based on airlight refinement. IEEE Trans. Multimedia 21, 351–362.

[7] Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A.C., Bengio, Y., 2014. Generative adversarial nets, in: Advances in Neural Information Processing Systems 27: Annual Conference on Neural Information Processing Systems 2014, December 8-13 2014, Montreal, Quebec, Canada, pp. 2672–2680.

[8] Guo, X., Nie, R., Cao, J., Zhou, D., Mei, L., He, K., 2019. Fusegan: Learning to fuse multi-focus image via conditional generative adversarial network. IEEE Trans. Multimedia 21, 1982–1996.

[9] He, K., Sun, J., Tang, X., 2011. Single image haze removal using dark channel prior. IEEE Trans. Pattern Anal. Mach. Intell. 33, 2341–2353.

[10] He, K., Zhang, X., Ren, S., Sun, J., 2015a. Spatial pyramid pooling in deep convolutional networks for visual recognition. IEEE Trans. Pattern Anal. Mach. Intell. 37, 1904–1916.

[11] He, K., Zhang, X., Ren, S., Sun, J., 2015b. Spatial pyramid pooling in deep convolutional networks for visual recognition. IEEE Trans. Pattern Anal. Mach. Intell. 37, 1904–1916.

[12] He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016, pp. 770–778.

[13] Huang, G., Liu, Z., van der Maaten, L., Weinberger, K.Q., 2017. Densely connected convolutional networks, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017, pp. 2261–2269.

[14] Ioffe, S., Szegedy, C., 2015. Batch normalization: Accelerating deep network training by reducing internal covariate shift, in: Proceedings of the 32nd International Conference on Machine Learning, ICML 2015, Lille, France, 6-11 July 2015, pp. 448–456.

[15] Isola, P., Zhu, J., Zhou, T., Efros, A.A., 2017. Image-to-image translation with conditional adversarial networks, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017, pp. 5967–5976.

[16] Kingma, D.P., Ba, J., 2015. Adam: A method for stochastic optimization, in: 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings.

[17] Li, B., Peng, X., Wang, Z., Xu, J., Feng, D., 2017. Aod-net: All-in-one dehazing network, in: IEEE International Conference on Computer Vision, ICCV 2017, Venice, Italy, October 22-29, 2017, pp. 4780–4788.

[18] Li, B., Peng, X., Wang, Z., Xu, J., Feng, D., 2018a. End-to-end united video dehazing and detection, in: Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-18), the 30th innovative Applications of Artificial Intelligence (IAAI-18), and the 8th AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI-18), New Orleans, Louisiana, USA, February 2-7, 2018, pp. 7016–7023.

[19] Li, B., Ren, W., Fu, D., Tao, D., Feng, D., Zeng, W., Wang, Z., 2019. Benchmarking single-image dehazing and beyond. IEEE Trans. Image Processing 28, 492–505.

[20] Li, R., Pan, J., Li, Z., Tang, J., 2018b. Single image dehazing via conditional generative adversarial network, in: 2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018, Salt Lake City, UT, USA, June 18-22, 2018, pp. 8202–8211.

[21] Liu, Q., Gao, X., He, L., Lu, W., 2018a. Single image dehazing with depth-aware non-local total variation regularization. IEEE Trans. Image Processing 27, 5178–5191.

[22] Liu, W., Hou, X., Duan, J., Qiu, G., 2019a. End-to-end single image fog removal using enhanced cycle consistent adversarial networks. arXiv preprint, abs/1902.01374 .

[23] Liu, Y., Zhao, G., Gong, B., Li, Y., Raj, R., Goel, N., Kesav, S., Gottimukkala, S., Wang, Z., Ren, W., Tao, D., 2018b. Improved techniques for learning to dehaze and beyond: A collective study. arXiv preprint, abs/1807.00202 .

[24] Liu, Z., Xiao, B., Alrabeiah, M., Wang, K., Chen, J., 2019b. Single image dehazing with a generic model-agnostic convolutional neural network. IEEE Signal Process. Lett. 26, 833–837.

[25] Mirza, M., Osindero, S., 2014. Conditional generative adversarial nets. arXiv preprint, abs/1411.1784 .

[26] Nair, D., Sankaran, P., 2021. A modular architecture for high resolution image dehazing. Signal Process. Image Commun. 92, 116113.

[27] Narasimhan, S.G., Nayar, S.K., 2003. Contrast restoration of weather degraded images. IEEE Trans. Pattern Anal. Mach. Intell. 25, 713–724.

[28] Qu, Y., Chen, Y., Huang, J., Xie, Y., 2019. Enhanced pix2pix dehazing network, in: 2019 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019, 2019.

[29] Ren, W., Liu, S., Zhang, H., Pan, J., Cao, X., Yang, M., 2016. Single image dehazing via multi-scale convolutional neural networks, in: Computer Vision - ECCV 2016 - 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II, pp. 154–169.

[30] Ren, W., Ma, L., Zhang, J., Pan, J., Cao, X., Liu, W., Yang, M., 2018. Gated fusion network for single image dehazing, in: 2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018, Salt Lake City, UT, USA, June 18-22, 2018, pp. 3253–3261.

[31] Salazar-Colores, S., Arreguín, J.M.R., Echeverri, C.J.O., Cabal-Yepez, E., Ortega, J.C.P., Rodríguez-Reséndiz, J., 2018. Image dehazing using morphological opening, dilation and gaussian filtering. Signal Image Video Process. 12, 1329–1335.

[32] Salazar-Colores, S., Arreguín, J.M.R., Ortega, J.C.P., Rodríguez-Reséndiz, J., 2019. Efficient single image dehazing by modifying the dark channel prior. EURASIP J. Image Video Process. 2019, 66.

[33] Simonyan, K., Zisserman, A., 2015. Very deep convolutional networks for large-scale image recognition, in: 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings.

[34] Singh, B., Najibi, M., Davis, L.S., 2018. SNIPER: efficient multi-scale training, in: Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems 2018, NeurIPS 2018, 3-8 December 2018, Montréal, Canada., pp. 9333–9343.

[35] Song, Y., Li, J., Wang, X., Chen, X., 2018. Single image dehazing using ranking convolutional neural network. IEEE Trans. Multimedia 20, 1548–1560.

[36] Tan, R.T., 2008. Visibility in bad weather from a single image, in: 2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2008), 24-26 June 2008, Anchorage, Alaska, USA.

[37] Tang, K., Yang, J., Wang, J., 2014. Investigating haze-relevant features in a learning framework for image dehazing, in: 2014 IEEE

Conference on Computer Vision and Pattern Recognition, CVPR 2014, Columbus, OH, USA, June 23-28, 2014, pp. 2995–3002.

[38] Tang, Z., Peng, X., Li, K., Metaxas, D.N., 2019. Towards efficient u-nets: A coupled and quantized approach. IEEE Transactions on Pattern Analysis and Machine Intelligence , 1–1.

[39] Xu, Z., Yang, X., Li, X., Sun, X., 2018. The effectiveness of instance normalization: a strong baseline for single image dehazing. arXiv preprint, abs/1805.03305 .

[40] Yang, X., Xu, Z., Luo, J., 2018. Towards perceptual image dehazing by physics-based disentanglement and adversarial training, in: Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-18), the 30th innovative Applications of Artificial Intelligence (IAAI-18), and the 8th AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI-18), New Orleans, Louisiana, USA, February 2-7, 2018, pp. 7485–7492.

[41] Zhang, H., Patel, V.M., 2018. Densely connected pyramid dehazing network, in: 2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018, Salt Lake City, UT, USA, June 18-22, 2018, pp. 3194–3203.

[42] Zhang, H., Sindagi, V., Patel, V.M., 2018. Multi-scale single image dehazing using perceptual pyramid deep network, in: 2018 IEEE Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops 2018, Salt Lake City, UT, USA, June 18-22, 2018, pp. 902–911.

[43] Zhang, J., Tao, D., 2020. Famed-net: A fast and accurate multi-scale end-to-end dehazing network. IEEE Trans. Image Process. 29, 72–84.

[44] Zhao, H., Gallo, O., Frosio, I., Kautz, J., 2017. Loss functions for image restoration with neural networks. IEEE Trans. Computational Imaging 3, 47–57.

[45] Zhu, H., Peng, X., Chandrasekhar, V., Li, L., Lim, J., 2018. Dehazegan: When image dehazing meets differential programming, in: Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI 2018, July 13-19, 2018, Stockholm, Sweden., pp. 1234–1240.

[46] Zhu, J., Park, T., Isola, P., Efros, A.A., 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks, in: IEEE International Conference on Computer Vision, ICCV 2017, Venice, Italy, October 22-29, 2017, pp. 2242–2251.

[47] Zhu, Q., Mai, J., Shao, L., 2015. A fast single image haze removal algorithm using color attenuation prior. IEEE Trans. Image Processing 24, 3522–3533.