



Unsupervised single image dehazing with generative adversarial network

Wei Ren^{1,2} · Li Zhou¹ · Jie Chen¹

Received: 23 August 2021 / Accepted: 9 September 2021 / Published online: 17 February 2022
© The Author(s) 2022

Abstract

Most recent learning algorithms for single image dehazing are designed to train with paired hazy and corresponding ground truth images, typically synthesized images. Real paired datasets can help to improve performance, but are tough to acquire. This paper proposes an unsupervised dehazing algorithm based on GAN to alleviate this issue. An end-to-end network based on GAN architecture is established and fed with unpaired clean and hazy images, signifying that the estimation of atmospheric light and transmission is not required. The proposed network consists of three parts: a generator, a global test discriminator, and a local context discriminator. Moreover, a dark channel prior based attention mechanism is applied to handle inconsistency haze. We conduct experiments on RESIDE datasets. Extensive experiments demonstrated the effectiveness of the proposed approach which outperformed previous state-of-the-art unsupervised methods by a large margin.

1 Introduction

Haze is a traditional atmospheric phenomenon caused by small particles absorbing and scattering the light in the atmosphere. Images of outdoor scenes captured from hazy fields typically suffer from low contrast and poor visibility. As shown in Fig. 1, in sunny weather, the wavelength of visible light is much larger than size of air molecules; hence, the scattering is minimal, the final imaging effect is good, and the contrast between the target light and the background light is apparent. However when the air is humid, the combustion products, sea salt and other particles form aerosol particles in the air, thereby increasing atmospheric scattering and absorption. Furthermore, hazy images can significantly impact computer vision applications, including object detection [30, 31], face detection [33, 34], and

semantic segmentation [32, 39]. Hence, single image dehazing has gain increasing attention over the past few years.

The generation mechanism of atomized images needs to be accurately described to accurately process images during a haze. An atmospheric scattering model was derived by McCaeny [1] in 1975 to describe the haze mechanism. Over the succeeding decades, several traditional methods [2–8] re proposed basing on this model. Single-image dehazing based on the atmospheric scattering model is an underconstrained problem that depends on an unknown depth and thus a highly ill-posed inverse problem. Therefore, traditional dehazing methods generally make additional assumptions and priors to restrict the model boundary and conditions. However, these assumptions may lose effectiveness and damage the quality of recovered image.

Recently learning-based methods have been extensively used in image dehazing [9–11, 14, 15, 35]. Despite great progress, these CNN-based methods heavily rely on large-scale paired datasets. Each sample consists of a hazy image and a corresponding synthesized haze-free image as ground truth. However, it is impractical to capture both the clear and corresponding hazy image of the same scene simultaneously. One way to solve this issue is few-shot learning which complete training from a handful of data rather than millions of data [42–46]. Another way to fix this problem is using generative adversarial network and its variants [12, 13, 18–20], among which models based on CycleGAN [21] is most prominent. However, these

✉ Wei Ren
renwei@ime.ac.cn

✉ Jie Chen
jchen@ime.ac.cn

Li Zhou
zhouli@ime.ac.cn

¹ Institute of Microelectronics, Chinese Academy of Sciences, Beijing, China

² University of Chinese Academy of Sciences, Beijing, China

Fig. 1 Comparison between haze and clear weather



unsupervised methods still fail to reach a high standard while undertaking image dehazing. Problems remain like color distortion, loss of image detail and incomplete dehazing particularly in thick fog images. Thus, an effective approach operated with unpaired hazy and clear images is necessary.

This paper proposes a generative adversarial network training with unpaired hazy and clear images and has achieved a state-of-the-art result compared to other unsupervised methods. A cyclic consistent loss is not used in our model, making it easier to train and converge the model. This paper's main contributions are as follow:

1. An end-to-end generative adversarial network training based on U-net architecture is proposed. We apply unsupervised hazy and clear images to train the model. Adversarial loss aside, we introduce perceptual loss to assess the differences between VGG features of hazy and dehazing images.
2. We adopt global–local discriminators. The global discriminator looks at the entire image to evaluate its overall consistency, while the local discriminator only looks at small areas centered on the completion area to ensure the local consistency of the generated patches. A global–local discriminator can help model to deal with spatially varying haze and generate cleaner images
3. We propose an attention mechanism inspired by dark channel prior to further process the sharply changing local area in the image. The dark channel map of hazy image is extracted and scaled to fuse with the features in the model. The proposed work is more robust and can better retain the details of images.

2 Related work

2.1 Single image dehazing

Prior-based methods are mainly based on prior assumptions to restore haze images, where atmospheric scattering models is widely used, as Eq. (1) shows. Where $I(x)$ is the hazy image, $J(x)$ is the clear image, $t(x)$ is the transmission map, and A is the global atmospheric light which is a constant overall the image. Considering prior assumptions, the parameters of atmospheric scattering model can be established.

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

He [5] discovered dark channel prior (DCP), one of the most popular hand crafted dehazing features based on empirical statistics of experiments on haze-free images. Dark channel prior assumes that the image patches of outdoor free images have low-intensity values in at least one channel. Zhu [8] developed a color attenuation prior, established a linear scene depth model for the hazy image, and supervised learning the model parameters. The assumptions and priors did not always hold, so they might fail in some instance.

With the latest progress of deep learning, several CNN-based methods have been proposed. Cai [9] proposed a trained model DehazeNet to estimate the transmission map from hazy images. Ren [10] proposed MSCNN that could directly regress the dehazed images with a coarse-grained model and a fine-tuned model. This algorithm reformulated the atmospheric model to integrate the transmission matrix and atmospheric light into a single variate. Li et al. [11] proposed AOD-Net, a light weighted linear model with dense

connection to recover the haze images. DCPDN [15] is also a U-net [40] architecture which estimate transmission map and global atmospheric light respectively. Deep-DCP [16] put forward a new loss function inspired by dark channel prior and trained the model unsupervised.

2.2 GAN

With the rise of GANs introduced by Goodfellow et al. [17], remarkable progress has been made in directly generating corresponding images end-to-end. Isola et al. [18] utilize the conditional generative adversarial network (CGAN) to complete the style transformation of paired images. GANs are also often used in dehazing. Visual Geometry Group Network (VGG) features was introduced by Li et al. [19]. The L1-regularized gradient of CGAN was used for image dehazing. To deal with the issue, a hazy image and the corresponding haze-free image were used as a supervised input and label. However, synthesized hazy images have a domain shift to the real-world scene. Moreover, the characteristics of image depths and hazy areas may change frequently. So Engin et al. used CycleDehaze [20], an enhanced version of CycleGAN architecture which trained networks with unpaired images. CycleGAN [21] is even more challenging to train compare to GAN, and CycleDehaze outputs haze-free images with color distortions. The above methods allow the network itself to learn the inner characteristics of hazy and haze-free images to perform end-to-end transformation. Thereby resolving the relationship between hazy and haze-free features at a deeper level and improving the overall dehazing effect on the image. However, the limitations of the overall style migration of the GANs resulted in the model being unable to have smooth transition in areas without haze or with slight haze, neither unable to achieve different conversion degrees for areas with varying haze concentrations. Thus this paper proposes an approach that trains an encoder-decoder architecture with no cyclic generator and no consistency loss. Specifically, it designs two discriminators rather than one to discriminate the quality of a generated image. A global discriminator scans the entire image and a local discriminator scans a patch input to get a high quality reconstruction. On this basis, this paper introduces the attention mechanism as a function of network, designing an attention map which can obtain satisfactory dehazing effects in various situations.

2.3 Proposed method

The haze area and concentrations are typically local and uneven, so we introduce an attention mechanism into the dehazing approach. Inspired by dark channel that can effectively reflect the area and concentrations, we put forward a dark-channel attention mechanism. When the scene’s depth gets deeper, the haze in the image gets heavier, while the value in dark-channel map gets larger. We introduce dark channel as attention map to focus on the specific areas in the images.

The proposed method employs an encoder-decoder architecture to generate dehazed images. Compared to CycleDehaze, it introduces only one generator without a cyclic conversion process and is easier to train. So we call this model “SingleDehaze”. We introduce two discriminators to make a high quality reconstruction and a dark-channel attention map to focus on dehazing areas during transformation. The dark-channel attention map is enhanced with a coefficient γ to improve the adaptability during training.

2.4 Single GAN model

We adopt the architecture from Johnson et al. [22] as our generator G which is also used in CycleGAN. An overview of layers in network model architecture is listed in Table 1. The difference is there are two generators in CycleGAN while we have only one. Our goal is to learn a direct mapping from the hazy images domain X to the dehazed images domain Y , given training samples $\{x_i\}_{i=1}^N$ where $x_i \in X$ and $\{y_j\}_{j=1}^M$ where $y_j \in Y$. We denote the data distribution as $x \sim p_{\text{data}}(x)$ and $y \sim p_{\text{data}}(y)$. As shown in Fig. 1, our model includes a mapping $G : X \rightarrow Y$ and it’s discriminator D distinguishes between the dehazed image $\{G(x)\}$ from $\{y\}$, where $\{x, y\}$ refers to hazy and ground truth unpaired image set. We use LSGAN [41] loss as an objective:

$$\min_D \text{Loss}_{\text{GAN}}(D) = \frac{1}{2} E_{y \sim p_{\text{data}}(y)} [(D(y) - 1)^2] + \frac{1}{2} E_{x \sim p_{\text{data}}(x)} [(D(G(x)))^2], \tag{2}$$

$$\min_G \text{Loss}_{\text{GAN}}(G) = \frac{1}{2} E_{x \sim p_{\text{data}}(x)} [(G(x) - 1)^2], \tag{3}$$

where G tries to generate $G(x)$ that look like haze-free images, while D aims to distinguish between dehazed image $G(x)$ from unpaired ground truth samples y . D aims $D(G(x))$ to get close to 0, while G aims close to 1.

Table 1 PSNR/SSIM results on SOTS datasets

Evaluation	DCP	DehazeNet	AOD-Net	CycleDehaze	Deep-DCP	Proposed method
PSNR	16.94	20.63	19.25	17.17	21.36	29.04
SSIM	0.7925	0.8495	0.8193	0.8103	0.8497	0.8705

2.5 Attention mechanism

Attention mechanism [36, 37] has been extensively used in image transformation. To focus on objects of interests that require transformation, Mejja et al. [23] built an attention network based on CycleGAN and utilized an attention map to label objects. In the dehazing task, a problem exists in which haze is uneven. When the local effect is better, the overall effect typically biased, or some areas are ignored. To resolve this problem, we seek to find an attention map that is related to the haze concentration. Image intensity may be a simple choice. Inspired by dark channel prior by He et al., Chen et al. [24] use dark-channel feature in CycleGAN. They used the dark-channel map as transmission and restored the image with atmosphere scattering model. Eventually the mapping happened only at the last layer. In our case, we scale the dark-channel map to various sizes and produce an elementwise product with the features of each layer. Figure 2 illustrate the process.

The dark-channel attention map is obtained as follows:

$$J^{\text{dark}}(x) = \min_{y \in \Omega(x)} \left(\min_{c \in \{R,G,B\}} J^c(y) \right), \tag{4}$$

$$\text{dark}_{\text{attention}} = \min \{ v * J^{\text{dark}}(x), 255 \}, \tag{5}$$

where J^{dark} is the original dark-channel method, $\Omega(x)$ is the area centered on x , and J^c is the color channel. As the intensity of dark-channel is relatively low compare to hazy

images. An enhancing coefficient v is thus trained to make the dark-channel more adaptable as Eq. (5) indicates.

The dark channel attention map is consecutively down sampled with maxpooling for four times to obtain the characteristic maps of different scales, which are at2, at3, at4 and at5, as shown in Fig. 2b. Then they are used as attention initialization factors in the decoder, multiplied by the feature map after maxpooling in the generative network encoder, and added to the up sampled feature map of the corresponding scale in the decoder to complete the skip connection. Here the bilinear interpolation is used for up sampling. Then deconvolution is carried out to extract features to complete the latent and the original input as follows to obtain the final output.

$$\text{output} = \text{latent} + \text{attention} \times \text{input}. \tag{6}$$

2.6 Local discriminators

Iizuka et al. [25] introduced global–local discriminators in image completion and improve the restoration of low-resolution images. The downscaling and upscaling process in the generator network while dehazing can lead to a degradation of the image quality, so we build two discriminators to make full use of both context and local context. While training, the global discriminator takes the entire images rescaled to 320×320 as inputs. It consists of six convolution layers and ends with a 512-dimension fully-connected layer. All the

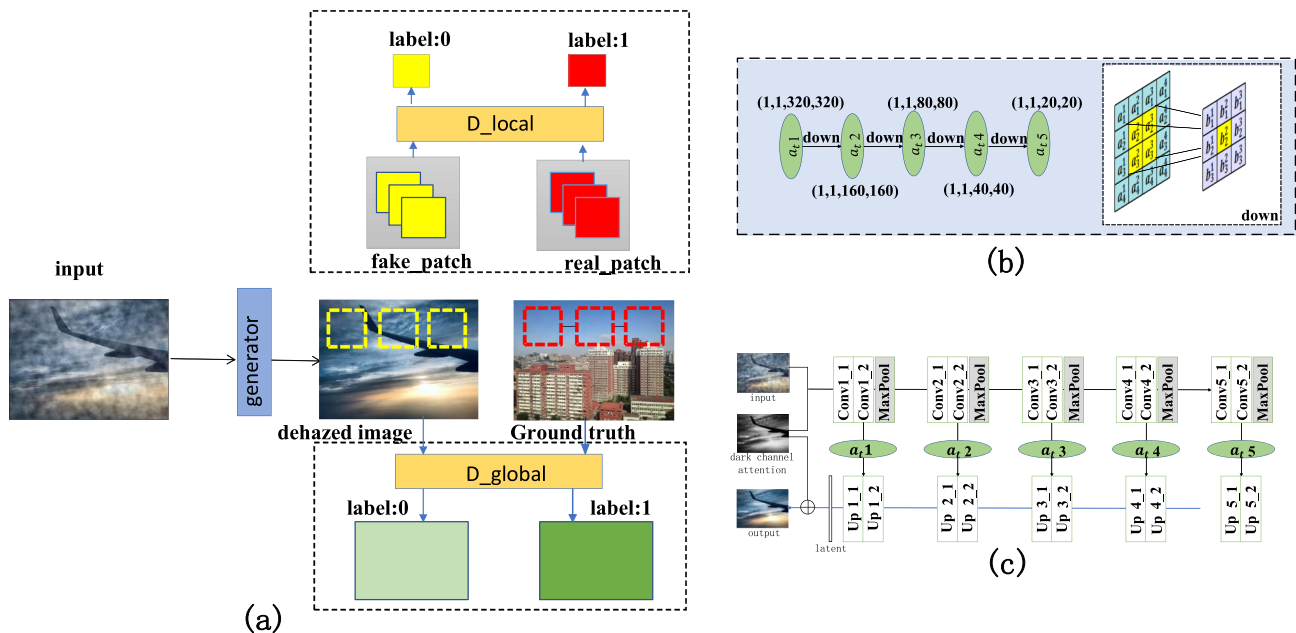


Fig. 2 The proposed SingleDehaze model with local and global discriminators and attention mechanism. α_i represents the dark channel attention map. **a** Is the overall model. **b** Is the attention map scaling process. **c** Is the generator with attention mechanism

convolutional layers employ a stride of 2×2 pixels to decrease the image resolution while increasing the number of output filters, and they use 3×3 kernels. The local discriminator has a similar architecture taking only a patch as input. These patches are 32×32 pixels randomly cropped from the entire images. Meanwhile, the local discriminator has five convolution layers and a 512-dimension fully-connected layer. The global and local discriminator each has an objective:

$$\min_{D_{\text{global}}} \text{Loss}_D^{\text{global}} = \frac{1}{2} E_{y \sim p_{\text{data}}(y)} [(D_{\text{global}}(y) - 1)^2] + \frac{1}{2} E_{x \sim p_{\text{data}}(x)} [(D_{\text{global}}(G(x)))^2], \tag{6}$$

$$\min_{D_{\text{local}}} \text{Loss}_D^{\text{local}} = \frac{1}{2} E_{y \sim p_{\text{data}}(y)} [(D_{\text{local}}(y') - 1)^2] + \frac{1}{2} E_{x \sim p_{\text{data}}(x)} [(D_{\text{local}}(G'(x)))^2], \tag{7}$$

where y' is randomly cropped patch from the unpaired ground truth, and $G'(x)$ is the patch of $G(x)$.

2.7 Loss function

Only generate adversarial loss is insufficient to recover all textural information, since the dehazing task is a pixel-level transformation. Various studies have shown the improvement of the quality of dehazed images with the help of perceptual loss [38], whether through supervised or unsupervised methods. The external data and the pre-trained model render an evident increase on performance. Perceptual loss can preserve image structure by comparing the features at different level of VGG16, a pre-trained a classification network. Perceptual loss function can be expressed as follows:

$$\text{Loss}_{\text{perceptual}} = \|\theta(z) - \theta(G(x))\|_2^2, \tag{8}$$

where θ is a VGG16 feature extractor from 3rd and 5th pooling layers. x is the hazy image sample and z is the corresponding clear image sample. Considering the illumination invariance of classification network, the hazy and dehazing images share the similar construction. We can see that the reconstructed image remain the content and spatial structure from the reconstruction of higher layers, but loses the exact color and texture. In our unsupervised model, the clear image sample corresponding to a hazy image does not exist. Hence we use the haze image and its corresponding dehazed image to formulate this modified perceptual loss. By experiments comparison, we find it is also effective in this case. Our perceptual loss of the whole image and image patch are as follow:

$$\text{Loss}_{\text{perceptual}}^{\text{global}} = \|\theta(x) - \theta(G(x))\|_2^2, \tag{9}$$

$$\text{Loss}_{\text{perceptual}}^{\text{local}} = \|\theta(x') - \theta(G'(x))\|_2^2, \tag{10}$$

where x' is the patch from x .

The overall loss objective is:

$$\text{Loss} = \text{Loss}_D^{\text{global}} + \text{Loss}_D^{\text{local}} + \text{Loss}_{\text{perceptual}}^{\text{local}} + \text{Loss}_{\text{perceptual}}^{\text{global}}. \tag{11}$$

3 Experiments

In this section, we assess our models alongside various unsupervised network models, including CycleGAN [21], pix-2pix [18], and Mejjati et al. [23]. Furthermore, the proposed method is compared with traditional method DCP and some learning-based methods. Among them, CycleDehaze and Deep-DCP are unsupervised methods, while the others are supervised. We conducted the experiments on two datasets and analyzed the results. The indicators used to evaluate the models are the peak signal-to-noise ratio (PSNR), structure similarity (SSIM) [26].

3.1 Datasets and training settings

RESIDE (Realistic Single Image Dehazing) dataset [27] is a large-scale dataset that contains synthesized and real-world paired images of indoor and outdoor scene. RESIDE training set contains 5 subsets. They are ITS (Indoor Training Set), OTS (Outdoor Training Set), SOTS (Synthetic Object Testing Set), URHI (Unannotated Real Hazy Images), and RTTS (Real Task-driven Testing Set). We remove redundant images from the same scenes and selected 9000 hazy/real image pairs from OTS and 6000 indoor image pairs to train a model adapted to different scenes. During training, we take one image from haze domain and randomly take another image from haze-free domain to guarantee the two images are unpaired.

3.2 Implementation

Our generator is similar to the original CycleGAN architecture. The difference is that we abandon one generator and the cycle-consistent loss to lower the difficulty of convergence and preserve only one generator during training. And we add three more modules into the network, namely dark-channel attention map, local discriminator and perceptual loss.

We used Pytorch framework for training and testing phases, and python to resize the images. During data augmentation, we cropped the images b selecting crop sizes and pixel coordinates randomly. Then the crops were resized to 320×320 and randomly flipped horizontally or vertically as the inputs of our network.

We trained our model with 4 NVIDIA titan X graphics card. We performed around 200 epochs on each dataset in order to ensure convergence. Our testing time is about 200 ms per image on Intel Core i7 CPU. We use the Adam optimizer with a learning rate of 0.0002. The learning rate was kept for the first 100 epochs and linearly decayed to zero over the next 100 epochs.

3.3 Experiments on synthetic datasets

We have conducted our algorithm on the benchmark SOTS dataset against state-of-the-art methods based on the hand-crafted priors (DCP) supervised networks and unsupervised networks and illustrate the performance. Figure 3 illustrates the dehazing images of the above approaches on SOTS dataset. DCP method can restore detail features of images. But it introduces halo artifacts in the edge area and a small amount of mist remains. Under the limitation of dark channel prior, the images get a serious color distortion on background of sky, so as much noises that cause textures and blocking artifacts. Furthermore, the color of shadows caused by foreground is deepened, which makes a big difference with the clean images. Supervised networks like DehazeNet, AOD-Net exhibit the obvious advantage in dehazing performance,

however, remain insufficient defogging. CycleDehaze method is a classic trial to train with unpaired images with a cyclegan model, however the quality of dehazing images are unideal. Deep-DCP method can remove the dehaze clearly, while the overall color is dark, and image details are not well preserved. From the result, our algorithm can get higher quality of dehazing images with no obvious color distortion and more closer to the ground truth. Compared by zooming in, our results can clearly recover the detail of image.

The quantitative comparison of results are shown in Table 2. As demonstrated, the proposed method achieves the highest PSNR and SSIM values on SOTS datasets. In comparison with the start-of-the-art unsupervised dehazing method Deep-DCP, our method obtains an improvement of 7.68Db and 0.02 in terms of PSNR and SSIM respectively on SOTS datasets.

Compared with the other five methods, the method in this paper has the best effect, the color is consistent with the original image, the overall contrast of the image is high, and the changing regions of the image can be well highlighted. The details in the image are more, the edges are more prominent and clearer. Compared with other dehazing methods based on traditional methods and deep learning, the proposed method is better in terms of dehazing performance,

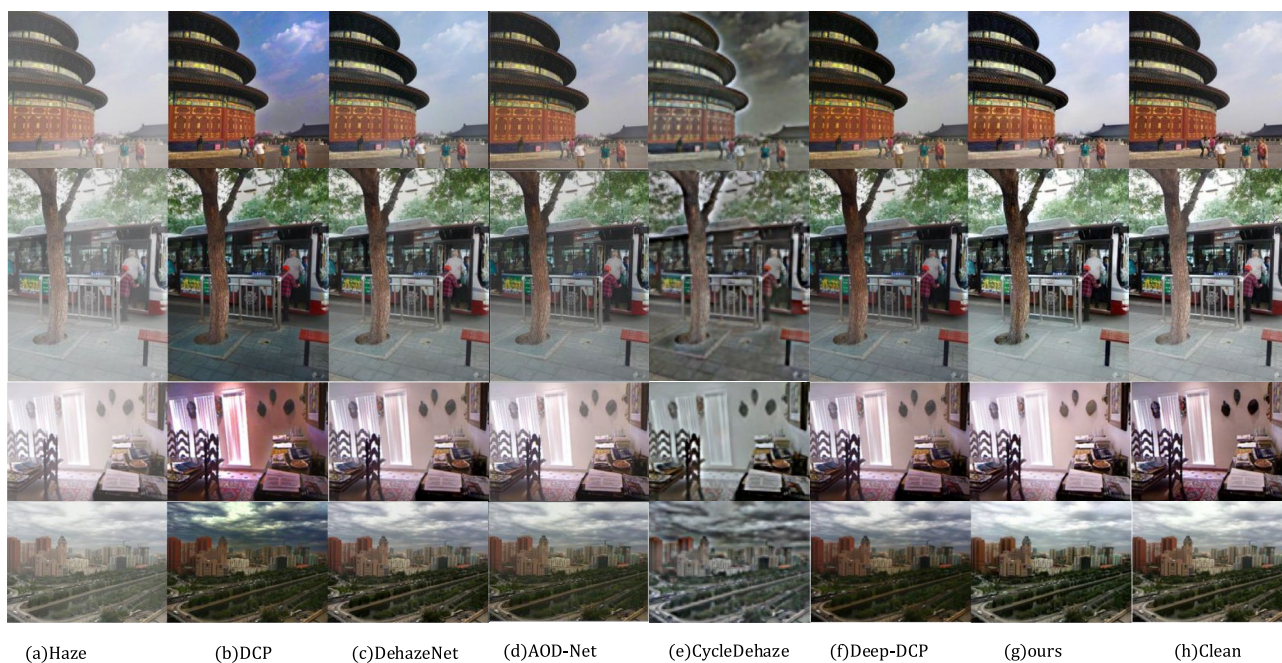


Fig. 3 dehazing images on SOTS datasets

Table 2 mAP results on RTTS Datasets

Evaluation	Haze	DCP	DehazeNet	AOD-Net	DCPDN	CycleDehaze	Deep-DCP	Ours
mAP	37.58	39.36	40.54	37.47	61.28	42.53	56.19	62.87

color recovery and image brightness. Through subjective visual evaluation, it can give people a good visual effect and achieve a good purpose of dehazing.

3.4 Experiments on real images

To verify the effectiveness of the model proposed in this section for dehazing in real-world images, we conducted tests on the real task-driven data set RTTS. The RTTS test set contains 4322 haze images of real-world scenes, a data set designed based on the target detection task. Vehicles, pedestrians and other objects related to the traffic scene are labeled with bounding boxes. The test set is evaluated as follows: the defogged images restored by the defogging algorithm were detected by using a Faster R-CNN model pre-trained on the VOC2007 dataset, and then the

mean Average Precision (mAP) of detection is calculated. This evaluation method can directly reflect whether the fog image is helpful to the improvement of target detection. As shown in Table 3, both the DCPDN [15] algorithm with better performance and the algorithm in this paper use VGG16 to construct the perception loss, while the feature extraction network in Faster RCNN is the VGG16 network. Perception loss applied in the fog removal stage produces a similar effect to pre-training for the network. Notably that CycleDehaze's fog removal visual effect is relatively poor among the comparison algorithms for comparison in this paper. However, since it also uses perceptual loss in its training process, it actually outperforms some algorithms with better visual performance in task-driven evaluation of the RTTS test set. Figure 4 presents visual results of state-of-the-art algorithms on RTTS dataset. The dehazing images from our model tend to be sharper and brighter, while results from other methods suffer color distortion or haze residual.

Table 3 Running time of different models

Method	Platform	Time(s)
DCP	matlab	1.23
DehazeNet	matlab	1.68
AOD-Net	caffe	0.40
MSCNN	matconvnet	3.05
CycleGAN	tensorflow	2.28
Deep-DCP	pytorch	0.52
Ours	pytorch	1.13

3.5 Running time test

In computer vision system, algorithm execution efficiency is one of the important evaluation criteria of system performance. As an image preprocessing technology, image defogging algorithm is often applied in some real-time applications, such as monitoring system, vehicle supervision system, automatic driving system, etc. Therefore, the time



Fig. 4 visual results on RTTS dataset

spent by the image defogging algorithm is a very important factor to consider. If it can provide a good visual effect in a relatively short time, the performance of the defogging algorithm is considered to be better.

To compare the real-time performance of the image defogging algorithm, we tested the average running time of different methods on the SOTS indoor composite image, where the image size was 512×512 . The running time of the algorithm was tested with the same equipment configuration. The experimental configuration is as follows: CPU ES-2678 v3@3.4 GHz, 48 core processor, 64G memory, GPU graphics card NVIDIA GTX 1080Ti. The following table shows the time (in seconds) taken by each algorithm to perform defogging on the original fog diagram. As shown in Table 3, except AOD-NET and deep-DCP, the method in this chapter is more efficient than other algorithms. Aod-net and deep-DCP models are smaller and sacrifice performance. In addition, aod-net is a model running on C++ platform, which also improves its operation efficiency.

3.6 Ablation study

3.6.1 Validity of loss function

This section analyzes the role of perceptual loss, local discriminator and attention mechanism in the training process on the SOTS dataset. We design three experiments by removing the components of local discriminator and dark channel attention, respectively. And compare the results with that of the complete model with global–local discriminator, dark channel attention and perceptual loss. As shown in Fig. 5, the first column is the results produced by SingleDehaze without dark channel attention. The second to fourth columns show the images from SingleDehaze with no perceptual loss, no local discriminator and no attention mechanism respectively. The last column is produced by the full SingleDehaze model. As presented Fig. 5, the network model with the best defogging effect can be obtained by combining the above three components. Due to an absence of constraint of perceptual loss, the transformed dehazing images suffer from a profound style loss. When local discriminator or attention module is removed from the model, the restored image tend to contain local regions of underexposure, color distortion and artifacts. These local regions mainly appear in the sky or the road regions. In contrast, the dehazing images of full SingleGAN contain realistic color

and eliminate most of the artefacts. The restoration in the sky region is stable and visually pleasing, demonstrating the proposed model produce images of better consistency. The metrics PSNR and SSIM are adopt to evaluate the quality of restored images of the above cases as well, as shown in Table 4. The results on the SOTS dataset also validate the effectiveness of global–local discriminator and dark channel attention mechanism.

3.6.2 Ablation study

To better understand the contribution of jump connection to image defogging in network model, the network structure is improved.

The Ablation study was performed. Two models are trained to compare with the network structure proposed in this chapter: (1) remove the jump connection between the encoder and decoder in the generated network; (2) feature addition is adopted instead of feature series in the implementation of jump connection. As shown in Table 5, the recovery effect of the model with skip connection is better than that without skip connection. Because skip connections reuse features in the encoder, they help train the network. For the implementation of jump connection, the result of feature series is slightly higher than that of feature addition. Since feature addition will fuse features, which is a special form of feature series, although it will reduce parameters and computation, it will cause feature loss. Therefore, feature series is selected to realize jump connection in this chapter.

4 Conclusion

We proposed an unsupervised image dehazing network, which provides a training process using unpaired hazy and clear images. It is an end-to-end model that directly outputs dehazing images without the estimation of parameters of the physical model. We introduce a modified perceptual loss in adapt to this unsupervised manner without cyclic consistency loss. The global–local discriminator and attention mechanism help improve the quality of the recovery images. We test the model on benchmark datasets and demonstrate the effectiveness of the method.

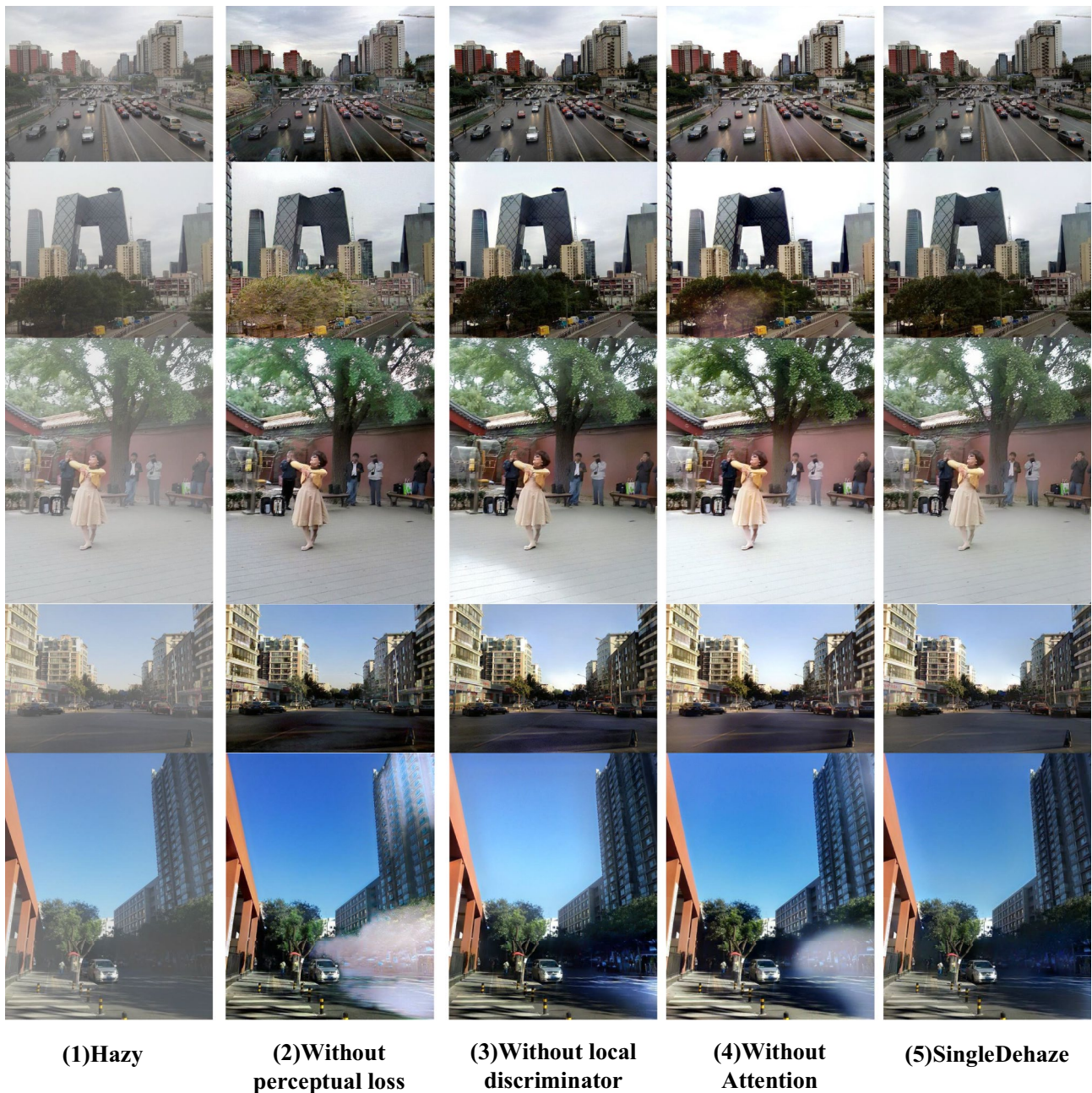


Fig. 5 Visual comparison from ablation study of SingleGAN. The images from left to right are samples as follows respectively: **a** hazy images, **b** results from SingleDehaze without perceptual loss, **c** results from SingleGAN without local discriminator, **d** results from

SingleGAN without attention mechanism, **e** results from the final version of SingleGAN. Images in **b–d** suffer from severe inconsistency and color distortion, while the final version that contains both local discriminator and attention mechanism gains the best visual effects

Table 4 Results of different models on SOTS dataset

Loss function	PSNR	SSIM
Without perceptual loss	28.87	0.8252
Without local discriminator	29.22	0.8714
Without attention mechanism	28.34	0.8241
SingleDehaze	29.24	0.8851

Table 5 Loss of network structure validity

Models	PSNR	SSIM
No skip connection	27.59	0.8173
Feature addition	28.68	0.8649
Skipping connection	29.24	0.8851

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. McCartney, E.J.: Scattering phenomena (book reviews: optics of the atmosphere. scattering by molecules and particles). *Science* **196**, 1084–1085 (1977)
2. Ancuti, C., Ancuti, C.O., De Vleeschouwer, C., Bovik, A.C.: Night-time dehazing by fusion. In: *IEEE International Conference on Image Processing (ICIP)*, pp. 2256–2260. IEEE (2016)
3. Ancuti, C.O., Ancuti, C., Hermans, C., Bekaert, P.: A fast semi-inverse approach to detect and remove the haze from a single image. In: *Asian Conference on Computer Vision*, pp. 501–514. Springer (2010)
4. Emberton, S., Chittka, L., Cavallaro, A.: Hierarchical rank-based veiling light estimation for underwater dehazing
5. He K, Sun J, Tang X. Single image haze removal using dark channel prior[C]// 2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2009), 20-25 June 2009, Miami, Florida, USA. IEEE, 2009
6. Meng, G., Wang, Y., Duan, J., Xiang, S., Pan, C.: Efficient image dehazing with boundary constraint and contextual regularization. In: *IEEE International Conference on Computer Vision (ICCV)*, pp. 617–624. IEEE (2013)
7. Tarel, J.-P., Hautiere, N.: Fast visibility restoration from a single color or gray level image. In: *IEEE International Conference on Computer Vision*, pp. 2201–2208. IEEE (2009)
8. Zhu, Q., Mai, J., Shao, L.: A fast single image haze removal algorithm using color attenuation prior. *IEEE Trans. Image Process.* **24**(11), 3522–3533 (2015)
9. Cai, B., Xu, X., Jia, K., Qing, C., Tao, D.: DehazeNet: an end-to-end system for single image haze removal. *IEEE Trans. Image Process.* **25**, 5187–5198 (2016)
10. Ren, W., Liu, S., Zhang, H., Pan, J., Cao, X., Yang, M.-H.: Single image dehazing via multi-scale convolutional neural networks. In: *Lecture Notes in Computer Science*. Springer Science and Business Media, Cham, pp. 154–169 (2016)
11. Li, B., Peng, X., Wang, Z., Xu, J., Dan, F.: AOD-Net: all-in-one dehazing network. In: *2017 IEEE International Conference on Computer Vision (ICCV)*. IEEE (2017)
12. Swami, K., Das, S.K.: Candy: Conditional adversarial networks based fully end-to-end system for single image haze removal. *arXiv preprint arXiv:1801.02892* (2018)
13. Yang, X., Xu, Z., Luo, J.: Towards perceptual image dehazing by physics-based disentanglement and adversarial training. In: *The Thirty-Second AAAI Conference on Artificial Intelligence (AAAI-18)* (2018)
14. Zhang, H., Sindagi, V., Patel, V.M.: Joint transmission map estimation and dehazing using deep networks. *arXiv preprint arXiv:1708.00581* (2017)
15. He, Z., Patel, V.M.: Densely connected pyramid dehazing network. In: *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE (2018)
16. Golts, A., Freedman, D., Elad, M.: Unsupervised single image dehazing using dark channel prior loss. *IEEE Trans. Image Process* **99**, 1 (2019)
17. Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., Bing, X., Bengio, Y.: *Generative Adversarial Nets*. MIT Press, Cambridge (2014)
18. Isola, P., Zhu, J.-Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. In: *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, 21–26 July 2017, pp. 5967–5976 (2017)
19. Li, R., Pan, J., Li, Z., Tang, J.: Single image dehazing via conditional generative adversarial network. In: *Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18–22 June 2018, pp. 8202–8211 (2018)
20. Engin, D., Genc, A., Ekenel, H.K.: Cycle-dehaze: enhanced CycleGAN for single image dehazing. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, Salt Lake City, 18–22 June 2018, pp. 938–9388 (2018)
21. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. *IEEE* (2017)
22. Johnson, J., Alahi, A., Fei-Fei, L.: Perceptual losses for real-time style transfer and super-resolution. In: *European Conference on Computer Vision*. Springer, Cham (2016)
23. Mejjati, Y.A., Shen, Z., Snower, M., Gokaslan, A., Wang, O., Tompkin, J. et al. Generating object stamps (2020)
24. Chen, J., Wu, C., Chen, H., Cheng, P.: Unsupervised dark-channel attention-guided cyclegan for single-image dehazing. *Sensors* **20**(21), 6000 (2020)
25. Iizuka, S., Simo-Serra, E., Ishikawa, H.: Globally and Locally Consistent Image Completion. *SIGGRAPH* (2017)
26. Wang, Z.: Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Process* (2004)
27. Li, B., Ren, W., Fu, D., Tao, D., Feng, D., Zeng, W. et al. Benchmarking single image dehazing and beyond (2017)
28. Berman, D., Treibitz, T., Avidan, S.: Non-local image dehazing. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE (2016)
29. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. *Computer Science* (2014)
30. Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., Berg, A.C.: SSD: single shot multibox detector. In: *European Conference on Computer Vision*. Springer, pp. 21–37 (2016)
31. Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You only look once: unified, real-time object detection. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 779–788 (2016)
32. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3431–3440 (2015)
33. Jiang, H., Learned-Miller, E.: Face detection with the faster r-cnn. In: *2017 12th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2017)*, pp. 650–657. IEEE (2017)
34. Yang, S., Luo, P., Loy, C.-C., Tang, X.: Wider face: a face detection benchmark. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5525–5533 (2016)
35. Qu, Y., Chen, Y., Huang, J., Xie, Y.: Enhanced pix2pix dehazing network. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 8160–8168 (2019)
36. Kong, F., Li, J., Jiang, B., Wang, H., Song, H.: Integrated generative model for industrial anomaly detection via bi-directional LSTM and attention mechanism. *IEEE Trans. Ind. Inform.* **99**, 1 (2021)

37. Ranjan, A., Behera, V., Reza, M.: Using a bi-directional LSTM model with attention mechanism trained on midi data for generating unique music (2020)
38. Yang, J., Wang, C., Jiang, B., Song, H., Meng, Q.: Visual perception enabled industry intelligence: state of the art, challenges and prospects. *IEEE Trans. Ind. Inform.* **99**, 1 (2020)
39. Edoh, T.: Smart cities: foundations, principles, and applications. *Comput. Rev.* **59**(12), 652–652 (2018)
40. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer International Publishing (2015)
41. Mao, X., Li, Q., Xie, H., Lau, R., Smolley, S.P.: Least squares generative adversarial networks. In: *2017 IEEE International Conference on Computer Vision (ICCV)*. IEEE (2017)
42. Yang, Y., Zhang, Z., Mao, W., et al.: Radar target recognition based on few-shot learning. *Multimed. Syst.* (2021). <https://doi.org/10.1007/s00530-021-00832-3>
43. Liu, S., Tang, Y., Tian, Y., et al.: Visual driving assistance system based on few-shot learning. *Multimed. Syst.* (2021). <https://doi.org/10.1007/s00530-021-00830-5>
44. Li, Y., Yang, J.: Few-shot cotton pest recognition and terminal realization. *Comput. Electron. Agric.* **169**, 105240 (2020)
45. Peng, Z., Li, Z., Zhang, J. et al. Few-shot image recognition with knowledge transfer. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 441–449 (2019)
46. Li, Y., Yang, J.: Meta-learning baselines and database for few-shot classification in agriculture. *Comput. Electron. Agric.* **2**, 2 (2021)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.