

Lightweight multiple scale-patch dehazing network for real-world hazy image

Juan Wang^{1,2}, Chang Ding^{1,2}, Minghu Wu^{1,2*}, Yuanyuan Liu^{1,2}, Guanhai Chen^{1,2}

¹ Hubei Energy Internet Engineering Technology Research Center, Hubei University of Technology
Wuhan 430068, China

[e-mail: happywj@hbut.edu.cn or 101900169@hbut.edu.cn]

² Hubei Laboratory of Solar Energy Efficient Utilization and Energy Storage Operation Control, Hubei
University of Technology, Wuhan 430068, China

[e-mail:d8425@foxmail.com]

*Corresponding author: Minghu Wu

*Received August 20, 2021; revised October 16, 2021; accepted November 27, 2021;
published December 31, 2021*

Abstract

Image dehazing is an ill-posed problem which is far from being solved. Traditional image dehazing methods often yield mediocre effects and possess substandard processing speed, while modern deep learning methods perform best only in certain datasets. The haze removal effect when processed by said methods is unsatisfactory, meaning the generalization performance fails to meet the requirements. Concurrently, due to the limited processing speed, most dehazing algorithms cannot be employed in the industry. To alleviate said problems, a lightweight fast dehazing network based on a multiple scale-patch framework (MSP) is proposed in the present paper. Firstly, the multi-scale structure is employed as the backbone network and the multi-patch structure as the supplementary network. Dehazing through a single network causes problems, such as loss of object details and color in some image areas, the multi-patch structure was employed for MSP as an information supplement. In the algorithm image processing module, the image is segmented up and down for processed separately. Secondly, MSP generates a clear dehazing effect and significant robustness when targeting real-world homogeneous and nonhomogeneous hazy maps and different datasets. Compared with existing dehazing methods, MSP demonstrated a fast inference speed and the feasibility of real-time processing. The overall size and model parameters of the entire dehazing model are 20.75M and 6.8M, and the processing time for the single image is 0.026s. Experiments on NTIRE 2018 and NTIRE 2020 demonstrate that MSP can achieve superior performance among the state-of-the-art methods, such as PSNR, SSIM, LPIPS, and individual subjective evaluation.

Keywords: Image enhancement, Convolutional neural networks, Image processing.

1. Introduction

The visual recognition system must extract accurate image features in applications. During the propagation process, the reflected light of the target object encounters numerous suspended particles in the air and interacts there with through absorption, radiation, and scattering, causing the redistribution of light energy. The collected images appear to have reduced contrast, covered details, and significant color distortion; low-quality images seriously impact the effectiveness of the visual system due to insufficient information.

From employing the algorithm based on the prior knowledge to the algorithm based on deep learning end-to-end, the atmospheric scattering model [1] serves as a significant foundation theory which describes the relationship between hazy image information.

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

Where $\mathbf{I}(\mathbf{x})$ denotes the hazy image, $\mathbf{J}(\mathbf{x})$ is the scene radiance representing the haze-free image, \mathbf{A} denotes the global atmospheric light, and $\mathbf{t}(\mathbf{x})$ denotes the medium transmission. $\mathbf{t}(\mathbf{x})$ can be defined as follows if the global atmosphere is homogeneous,

$$\mathbf{t}(\mathbf{x}) = e^{-\beta \mathbf{d}(\mathbf{x})} \quad (2)$$

Where $\mathbf{d}(\mathbf{x})$ represents the scene depth and β is the scattering coefficient of the atmosphere.

In the early stages of dehazing works, the most widely employed methods were those based on prior knowledge, such as Dark Channel Prior (DCP) [2], Maximum Contrast (MC) [3], and Color Attenuation Prior (CAP) [4]. However, methods perform insufficiently in dense haze images or complicated situations.

Compared with the earlier methods, the performance of deep learning is superior. The early deep learning method, DehazeNet [5], adopts a convolutional neural network-based deep architecture for single image haze removal, with experiments [6] articulating that the method possesses better processing results and faster processing speed compared to most prior methods and systems utilized to restore clear images. CNN-based methods, [7]-[11] and methods based on GAN architecture, [12]-[14] have been employed in numerous scenarios, such as underwater circumstances [15], Nighttime Dehazing [16]-[19], Nonhomogeneous haze images [20]-[23], and physics-based models [24], [25].

During the collection and processing of hazy images, findings were made that the hazy images collected by the handheld camera on the ground, low-floor camera or low-altitude UAV ordinarily reveal a nonhomogeneous haze distribution. Most hazy images are targeted by existing haze removal algorithms and nonhomogeneous hazy images are synthesized by hazing algorithms. Hazy images are treated as nonhomogeneous hazy images requiring processing, but for hazy images in the real world, there are often problems such as unclear dehazing effects or color deviation.

To resolve the aforementioned problems, the first layer of the multi-patch backbone framework are utilized, and an image segmentation mode is employed in MSP, with segmentation being changed from left-right to up-down. By processing each block separated by thick and thin haze individually, the restoration of image information is enhanced.



Fig. 1. Low altitude outdoor hazy map. Images from dataset O-haze and ourselves.

The following figures articulate the frequency distribution of the histogram separated by up and down parts in the real-world haze map.

The illustration in **Fig. 1** is divided into two parts: top and bottom. As observed in **Fig. 2**, the distribution of the upper histogram is concentrated at a high level, indicating that the images are considerably covered by haze with low contrast, and a trend towards white. Although most of the histogram at the bottom is distributed between 0-150bins, there is a large span of values and a few pixels between 200-256bins, which revealing that the bottom part of the image possessed three characteristics: large color contrast, color histogram values far away from the high area, and low haze coverage.

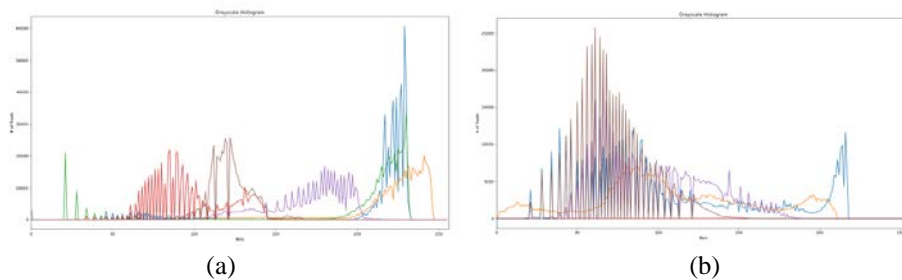


Fig. 2. Histogram of the sum of the upper(a) and lower(b) parts of the six images in **Fig. 1**.

The main contributions of the present paper include the following:

- Framework **Multiple scale-patch (MSP)** based on the original research was proposed for quick haze removals. Additionally, a multi-scale framework was employed as the backbone network, and multi-patch framework as the incidental network to supplement the image information when training. MSP utilized the negligible increase in model parameters in exchange for the retention of the fundamental original image structure information following haze removal.

- Findings were made that a haze distribution difference between the lower and upper part of the image would occur in the real-world image. To achieve enhanced dehazing of the non-homogeneous real-world images, the segmentation multi-patch method should be changed, which allowing different haze regions to be treated separately.

- The algorithm in the present study was trained through a variety of datasets and a minute quantity of training images. The experimental results demonstrate generalization performance and rapid inference of MSP which increased the feasibility of its implementation in practical projects and real-time video dehazing.

Several advanced dehazing algorithms, and the construction of the MSP model, are discussed in the following sections. In Section 2, recent related dehazing methods are introduced, while the architecture of MSP is illustrated in Section 3. The experiment is detailed in Section 4, and the comparison analysis is given in Section 5. Finally, the conclusion and future research directions are provided in Section 6.

2. Related work

The proposed dehazing methods can be classified into traditional priority-based methods and modern learning-based methods.

Regarding the traditional methods, in DCP, most non-sky local areas images in the test images outside were found to have at least one channel with a low value, with the dehazing in the sky part failing. In CAP, a phenomenon was revealed where a linear relationship between image depth and brightness exists; haze density could possibly be estimated by calculating the difference between brightness and saturation.

Recent deep learning dehazing methods can be classified into two categories, a) utilizing atmospheric scattering models, and, b) image-to-image translation. Deep learning methods employed a model to learn paired or unpaired hazy & clear images. After utilizing large and quality images for training, the model learned the residuals of the haze in the image, which was then used to cleanly restore the image. Based on the atmospheric scattering model, in DehazeNet, a model and a new nonlinear function “BReLU” were proposed to calculate transmission $t(x)$. In AOD-Net, a lightweight end-to-end model was employed, which processed images directly based on CNN, while an edge-preserving encoder with densely connected pyramids was implemented in DCPDN to accurately estimate transmission mapping. Smooth dilation was utilized in GCA-Net [26] to eliminate gridding artifacts caused by dilation convolution and extra parameters, and Gated Fusion was used for different levels features fusing. FFA-Net proposed a novelty feature attention (FA) module, which combined channel attention with pixel attention mechanism. FA dealt with different features and pixels unequally, learned feature weights adaptively from the FA module, and assigned more weights to important features. MSBDN proposed a multi-scale enhanced dehazing network with dense feature fusion based on U-NET [27], and applied “firm-operate-subtract” enhancement strategy to the decoder for dehazing work. With regard to certain algorithms based on GAN, CycleDehaze [28] applied the structure of CycleGAN [29] to dehazing, employed two pairs of generators and discriminators for image-to-image conversion. EPDN [30] was based on multi-scale GAN network, which placed the enhancing block in rear of the backbone network. Additionally, in FD-GAN, a fusion-discriminator was developed that integrates frequency information into the learning process, improved the performance of the generators, and in turn generated more realistic images from the side to obtain $t(x)$. HIDEGAN [31] was established by designing an enhanced version of CycleGAN and an enhanced conditional version of GAN to recover clean images directly.

Along with the traditional deep learning algorithms based on CNN and GAN, more algorithms combining the knowledge of other tasks have recently emerged to solve image dehazing problems. To illustrate, domain adaptation [32] was utilized to dehaze the real-world hazy map, then, the synthesized hazy map and the real hazy map were transformed into each other to learn and reprocess for better dehazing effects. Two-branch [33] utilized ensemble learning and transfer learning to enhance the fast learning and multi-learning capacity, and realized the model generalization performance in multi-class environments. AECD-Net [34] utilized novel contrast regularization (CR) technology based on contrast learning, hazy images and clear images were employed as negative samples and positive samples, respectively. To ensure the restored image was similar to the clear images, and away from the position of hazy images, CR was employed. Method RefineDNet [35], which combines prior knowledge with deep learning, restores visibility through dark channel prior and adversarial learning are utilized for second stage processing.

3. Proposed Method

In the current section, present study is introduced in detail, including the two mainly utilized architectures, multiple scale-patch framework, and the image segmentation that was used for enhanced image restoration.

As the target network of the present study was a lightweight and fast processing network, a multi-scale structure was allocated to be the main lightweight fundamental framework of MSP with a simple structure and small number of parameters when constructing. Despite utilizing a multi-scale structure, no significant dehazing effect of mutil-scale model was detected, with the appearance of color deviation. To resolve the issue, an incidental network was employed to retain the image color information. However, the incidental networks were not allowed to irreversibly influence the network setting goals.

Inspired by DMPHN [36], multi-patch mechanisms were applied to MSP as the information supplement module. To avoid increasing the number of model parameters, the image separation method of left and right in the original multi-patch structure to the upper and lower separation method was optimized according to the prior knowledge of thick and thin haze images discussed in Section 1. Algorithm was optimized by MSP in terms of image color, overall structure, and model processing effects based on the aforementioned model improvements. As each module possessed a unique function, singular module will not yield the best effects of haze removal.

3.1 Backbone Multi-scale Architecture

A three-layer network structure similar to DMPHN was utilized in the present study, but MSP changed the multi-patch method to multi-scale, and learned image features through the same encoding and decoding operation of different sizes image. In recent years, said structure has been widely employed in dehazing and low-level image processing.

The main framework was divided into three layers, each layer contained the same content but a different size hazy map input. The structure layout is as follows, the input hazy map I_{L1} of the first layer was two times and four times the hazy map I_{L2} of the second layer and the hazy map I_{L3} of the third layer. In each layer, MSP entered the images into the Encoders of -1, -2, and -3 to obtain the feature map, and then the above sampling was fused with the upper network feature information to retain more image information. Additionally, the image after processed by the Decoders of -1, -2 and -3 served as the information compensation of the input image on the upper layer to achieve the same purpose. To avoid complicating the overall model, only simple connections and information fusion strategies were employed.

Regarding the encoder in the algorithm of DMPHN across the connection, particularly the way in which information in the low level was reserved and transferred to the high-level information, was beneficial for preserving image information. Despite said benefits, considering the few scaling times and network layers, the upsampling reservation of the intermediate map was implemented to the third and second layers of the mutli-scale portion in the MSP to prevent gradient explosion during the training process.

3.2 Incidental Multi-patch Architecture

In the original backbone network experiment, the results revealed a color deviation and detail feature loss and multi-patch was also used as a reference for the DMPHN. Additionally, the original network efficiency comparison of different levels was utilized. A multi-layer patch to incidental network was employed to improve the overall efficiency and retain the lightweight goals of the model. The results of the experiment on DMPHN revealed that the size of the

structure of the model increased exponentially when the hierarchy deepened. The optimal structure of the processing speed and haze removal effect was 1_2_4, but for the attached network, a 1_2 network structure was selected in the present study to avoid increasing the complexity of the model. The results of the DMPHN experiment revealed that the 1_2 structure sacrificed 1.5% PSNR and 0.7% SSIM parameters in exchange for reducing the size of the model by half and the running time by two-thirds.

Regarding the ordinary synthetic image with haze, the haze addition algorithm utilizes homogeneous white noise points as haze in the image, while most existing algorithms directly remove homogenous haze through the atmospheric scattering model. Such algorithms are ineffective in dealing with real-world, nonhomogeneous or dense haze images.

For achieving better results the application of multi-patch as the incidental network was also to implement the purpose of processing image parts with different degrees of ambiguity separately in the original images, which also demonstrates the correctness of the present theory.

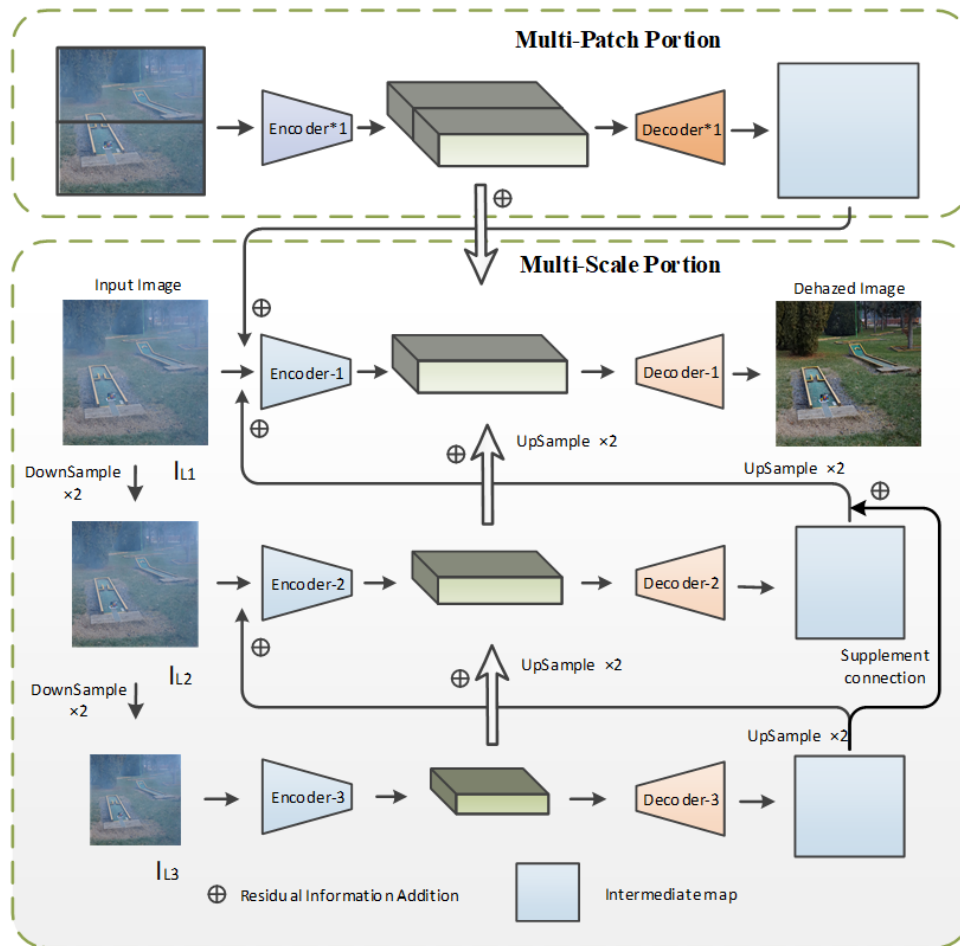


Fig. 3. The architecture of the MSP's model. Image I_{L1} is the raw input. Three-tier encoder and decoder structure, in which module information is not shared. The information in supplement connection is the same as the original information. Residual Information Addition employs concatenate operations.

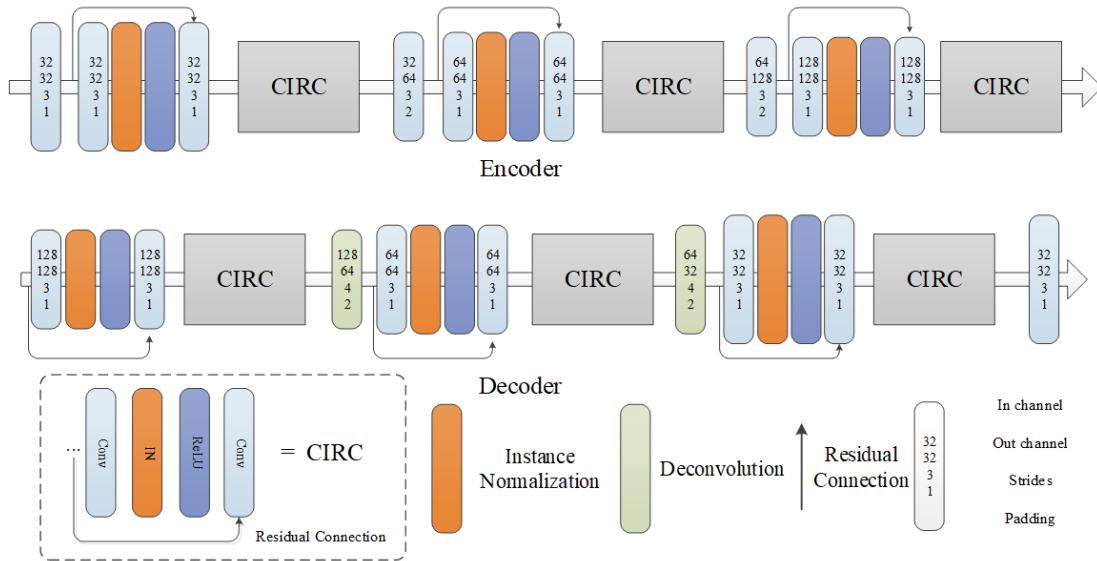


Fig. 4. The architectural of Encoder-Decoder. Each CIRC that does not display the number of channels is the same as the previous CIRC that displays the number of channels. Since the study was limited to equipment, batch size is 4 and IN has been utilized instead of BN.

The left and right segmentation method of multi-patch was changed into upper and lower segmentation, and applied to the dehazing tasks. Further, distant and near haze were separated to obtain better haze removal effect for the real-world hazy map. The novel image segmentation has overcome the sky color distortion existing in the previous algorithms, in addition to the problem of the distant object haze removal impurity or the object edge residual haze. Overall, more accurate information learned by image segmentation in the affiliated network was transmitted to the main network to complement the details of image information. By sacrificing the negligible computational complexity and the increase of inference time, a more accurate function of dehazing color restoration and object information retention could be obtained.

3.3 Encoder-Decoder Module

MSP employed an encoder consisting of 15 convolutional layers, 6 instance normalization blocks and 6 ReLU activation units. The maximum number of channels after convolution could reach 128. For the decoder, 13 convolutional layers, 2 deconvolutional layers, 6 instance normalization blocks and 6 ReLU activation units were set, among which the convolutional layer connecting the two modules became deconvolution to achieve the purpose of better image recovery.

MSP refers to the encoder-decoder structure of DMPHN, but changes the method of convolution + activation function + convolution to CIRC module for extracting image features. For MSP, IN (Instance normalization) in CIRC was chosen due to the small mini-patch input value, in addition to IN for a small amount of training data being more beneficial.

3.4 Network Architecture

Image information transmission starts from the third layer. The feature map F_3 generated after image I_{L3} through Encoder-3 will be shared and transferred to the upper layer for fusion and decoding. The generated image output O_3 after feature map is decoded by Decoder-3,

$$\begin{aligned} \mathbf{F}_3 &= \text{Encoder}_3(\mathbf{I}_{L3}) \\ \mathbf{O}_3 &= \text{Decoder}_3(\mathbf{F}_3) \end{aligned} \quad (3)$$

While \mathbf{O}_3 is taken as the input of the second layer after upscaled by factor of 2, \mathbf{O}_3 and \mathbf{I}_{L2} are fed into Encoder-2 to obtain the temporary feature map \mathbf{F}_{T2} . Then \mathbf{F}_3 and \mathbf{F}_{T2} are combined into the second feature map \mathbf{F}_2 and processed by Decoder-2 to generate output in the second layer \mathbf{O}_2 .

$$\begin{aligned} \mathbf{F}_{T2} &= \text{Encoder}_2(\mathbf{I}_{L2} + \text{Upscale}(\mathbf{O}_3)) \\ \mathbf{F}_2 &= \mathbf{F}_{T2} + \text{Upscale}(\mathbf{F}_3) \\ \mathbf{O}_2 &= \text{Decoder}_2(\mathbf{F}_2) \end{aligned} \quad (4)$$

In the first layer, MSP is completed by combining the backbone multi-scale network with multi-patch network parts. To distinguish parameters, * was added to the partial variables of multi-patch network. The input image of multi-patch part is the same as that of \mathbf{I}_{L1} , which is divided into upper and lower parts \mathbf{I}_{L1-1} and \mathbf{I}_{L1-2} . Similar to the trunk network, \mathbf{I}_{L1-1} and \mathbf{I}_{L1-2} were fed into Encoder*1 simultaneously to obtain the feature graph \mathbf{F}_{*1} , and then \mathbf{F}_{*1} obtained \mathbf{O}_{*1} through Decoder*1, in which \mathbf{F}_{*1} and \mathbf{O}_{*1} are provided to the first layer of the trunk network as the output of the incidental network.

$$\begin{aligned} \mathbf{F}_{*1} &= \text{Encoder}^*1(\mathbf{I}_{L1-1}) + \text{Encoder}^*1(\mathbf{I}_{L1-2}) \\ \mathbf{O}_{*1} &= \text{Decoder}^*1(\mathbf{F}_{*1}) \end{aligned} \quad (5)$$

In the backbone of network, upscaled \mathbf{O}_2 , \mathbf{O}_{*1} and \mathbf{I}_{L1} are combined, then fed into Encoder-1 to obtain temporary feature graph \mathbf{F}_{T1} . \mathbf{F}_{*1} , \mathbf{F}_{T1} and \mathbf{F}_2 after upscale are combined to generate the feature map \mathbf{F}_1 of the first layer. The network will process \mathbf{F}_1 through Decoder-1 to generate the final dehazed image \mathbf{I} .

$$\begin{aligned} \mathbf{F}_{T1} &= \text{Encoder}_1(\mathbf{I}_{L1} + \text{Upscale}(\mathbf{O}_2) + \mathbf{O}_{*1}) \\ \mathbf{F}_1 &= \mathbf{F}_{T1} + \text{Upscale}(\mathbf{F}_2) + \mathbf{F}_{*1} \\ \mathbf{I} &= \text{Decoder}_1(\mathbf{F}_1) \end{aligned} \quad (6)$$

3.5 Loss Function

Multiple types of loss functions were adopted and combined with different weights. The training loss for the present model was comprised of the combination of L_I and L_2 loss term, a perceptual loss term, and a TV loss term. as formulated below:

$$L_{\text{total}} = \lambda_a L_{I+2} + \lambda_b L_p + \lambda_c L_{\text{tv}} \quad (7)$$

Where the value of λ_a is set as 1, the value of λ_b is 0.006, and the value of λ_c is $2e-8$. Among said losses, the combination loss, as the pixel-level loss, is the most significant factor, and restoration can shorten the distance between pixels. However, the loss of pixels alone is not sufficient, visual loss is needed to perfect the balance between the advanced information of images, and, the total variation loss function was adopted to keep the edge of the image smooth

and suppress the image noise.

3.5.1 L1 and L2 Loss

The combination of the non-smooth L1 loss function and L2 smooth loss function preserves the fast convergence and convergence smoothing ability of the L2 function, and complements the disadvantage of L1's ability to find the optimal value but easily unstable derivative at zero. Combination loss is used here to make the dehazing image closer to the original ground truth. Weights were assigned to the two loss functions, respectively, as:

$$L_{1+2} = \lambda_1 L_1 + \lambda_2 L_2 \quad (8)$$

Where the value of λ_1 is set as 0.6 and the value of λ_2 is set as 0.4.

3.5.2 Perceptual Loss

Perceptual loss was employed for better image information restoration, which utilized to compare the feature obtained by the real image with the feature obtained by the generated image, so the high-level information (content and global structure) is close.

$$L_p = \sum_{i=1}^N \|\phi(G(I_i)) - \phi(J_i)\|_1 \quad (9)$$

$\phi()$ represents the feature maps obtained by the ReLU1_2 layer within the VGG16 network.

3.5.3 TV Loss

Total variation loss is a common regular term and needs to be added to the algorithm to maintain the smoothness of the image object edge. TV loss was employed to work with other losses to constrain hazy noise.

$$L_{tv} = \|\nabla_x \hat{I}\|_2 + \|\nabla_y \hat{I}\|_2 \quad (10)$$

4. Experiment

4.1 Datasets

As collecting a large number of real-world hazy images and corresponding haze-free counterparts is impractical, synthetic hazy images generated by algorithms have been extensively used. The number and quality of images in datasets determine the effect of the deep learning model and appropriate training dataset will improve the model performance which is superior to the improvement on algorithm.

To test the robustness of the network on low altitude real world hazy maps, NTIRE18 [37], which includes indoor and outdoor hazy images. and NH-HAZE [38], which contains 55 images of nonhomogeneous haze for the train and test dataset, were adopted. The NTIRE dataset contains 70 training images and 5 test images, and NH-HAZE contains 50 training images and 5 test images. Multi-dataset testing could demonstrate the generalization performance of the present model.

Other available datasets include: RESIDE [39], D-HAZE [40], I-HAZE [41], O-HAZE [42], HazeRD [43], KITTI [44], and BeDDE [45].

4.2 Data Preparation

Due to the small number of images in the dataset, the dataset was expanded. Each image of size 1024×1024 in dataset NTIRE2018 was divided into $16 \text{ pcs} \times 256 \times 256$ smaller images on average to be used as the training set, and the total number of images after augmentation was 1120. As for NTIRE2020, the nonhomogeneous haze dataset was augmented. Similar to the previous dataset, the original 50 training images with a size of 1600×1200 were replaced with 800 images with a size of 400×300 .

4.3 Implementation Details

The MSP model was adopted to train for 500 epochs with the Adam optimizer on NTIRE18, and train for 800 epochs on the NH-HAZE dataset. The default initial learning rate was set to 1×10^{-5} . The batch size of 4 was adopted to manipulate the images more finely. The testing part imported RGB images with size of $1024 \times 1024 \times 3$. The network was trained and tested by deep learning framework Pytorch with Nvidia GTX 1660 SUP.

4.4 Metric

The objective evaluation of the effect of image dehazing was to employ quantitative measurement methods to automatically evaluate the image quality, so as to obtain parameters reflecting the quality or the degree of loss as the evaluation result. The comparison of the result image and ground truth is the real distance and metrics.

SSIM

Structural Similarity Index Measure is an index to evaluate the loss and distortion of the fused image, and consists of three parts: correlation loss, brightness and contrast distortion.

PSNR

Peak signal to noise ratio is an objective standard for evaluating images, and is generally used for an engineering project between the maximum signal and background noise.

LPIPS

Through learned perceptual image patch similarity [46], the features extracted from the network structure of the model could be measured to obtain judgments that are more consistent with human perception. The feature difference between the real sample and the generated sample in the model was analyzed, and the difference was calculated in each channel using L2. LPIPS uses the weight sum of all channels.

4.5 Color Disk Contrast

To detect the algorithms recovery effect for image information, color disk in the test image was captured. From the perspective of the image recovery process, most of the existing methods reveal the overall image tends to dark in the image color recovery under the circumstances of dense haze. and image restoration color also possesses different degrees of deviation result from unclean haze removal. The observation of the MSP color disk after haze removal demonstrates that, due to the supplement of the incidental network in the MSP, the color information of the image was slightly deviated.

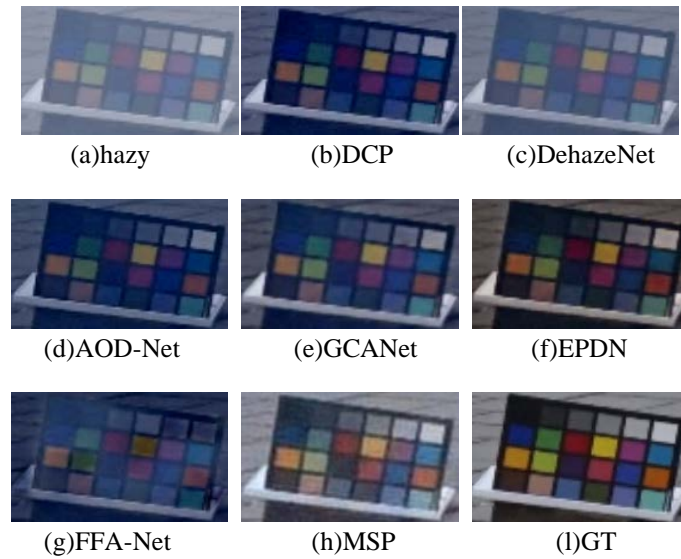


Fig. 5. The results of various dehazing algorithms on the NH-HAZE dataset.

4.6 Image Evaluation

Results on NTIRE18. From **Fig. 6**, an observation can be made that a large area of haze remained in the image after being processed by **DehazeNet** and **AOD-Net**, bringing little visual improvement to the image before and after dehazing. Image color deviation is also present. At the same time, an observation can be made from the numerical comparison in **Table 1** that **AOD-Net** and **DehazeNet** were lagging behind in indicators.

The overall color of the image tended to be blue after **GCANet**, with the haze removal processing in some small areas being incomplete and the image color being deviated. Despite improving on all three metrics, **GCANet** did not reach the level of ideal perception. Compared with **DehazeNet** and **AOD-Net**, **EPDN** achieved restored images with higher quality and construction details of the image object being well preserved. However, **EPDN** failed to effectively dehaze under dense haze conditions, and the overall color of the image after haze removal was partial to dark. Images dehazed by **FFA-Net**, there was some residual haze left in the images, where the color information was lost to a certain extent. **MSBDN** dehazing effect is not observable enough, especially in dense areas, but there is no serious color deviation phenomenon. **Two-Branch**'s image visual effect was acceptable, but the color recovery was incomplete. **MSP** significantly reduced the runtime of the dehazing process, the overall color of image was brighter, a better precondition for subsequent operation was provided. **MSP** also removed the haze in the dense haze area, and supplemented the knowledge learned from the training dataset to complete the information of the significantly lost area in the image to the greatest extent.

Results on NH-HAZE. In comparing the state-of-the-art results of the algorithms in NH-HAZE, **DehazeNet** and **AOD-Net**, which are based on prior knowledge in whole or in part, said algorithms had little effect on nonhomogeneous hazy images, and the strategy to improve the contrast had little effect on the dataset. **GCA-Net** was similar to the previous two methods, where the image contrast was excessively enhanced after dehazing, but the visualization

NTIRE18 (main)

(a)Hazy (b)DehazeNet(c)AOD-Net (d)GCANet (e)EPDN (f)FFA-Net (g)MSBDN(h)Two-Branch (i)MSP (j)GT
Fig. 6. The results of various dehazing algorithms on dataset NTIRE18

Table 1. The experimental results comparison as follows, where AR represents the average processing time of a single image, including model and image import, processing and saving time.

Methods\Metric	PSNR	SSIM	LPIPS	AR (s)	Parameters
DehazeNet	16.20	0.66	0.33	7.94	×
AOD-Net	15.21	0.62	0.35	2.94	0.0017M
GCANet	16.27	0.68	0.31	5.04	0.7M
EPDN	17.03	0.70	0.29	2.10	16M
FFA-Net	12.93	0.51	0.49	5.47	4M
MSBDN	19.70	0.70	0.32	0.78	29.9M
Two-Branch	19.50	0.58	0.53	5.56	50M
Our Model	20.11	0.67	0.19	1.29	6.8M
Ground Truth	$+\infty$	1.00	0.00	×	×

NH-HAZE



Fig. 7. The results of various dehazing algorithms on dataset NH-HAZE.

Table 2. The experimental results comparison.

Methods\Metric	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow	AR(s) \downarrow	Parameters
DehazeNet	12.34	0.42	0.56	13.30	×
AOD-Net	13.02	0.41	0.57	3.84	-
GCANet	14.25	0.51	0.49	1.67	-
EPDN	14.50	0.53	0.47	1.85	-
FFA-Net	12.56	0.42	0.61	4.69	-
MSBDN	13.91	0.48	0.54	1.28	-
Two-Branch	19.53	0.52	0.58	6.73	-
Our Model	16.69	0.61	0.34	1.31	-
Ground Truth	$+\infty$	1.00	0.00	×	×

improved, which is consistent with the numerical comparison in [Table 2](#). The results of EPDN, FFA and MSBDN are similar to those of NTIRE18.

From a subjective perspective, two-branch achieved a better visual effect. However, due to excessive attention to the local features and pixel-level information, and slight consideration to structural balance between image blocks. The performance of the two-branch method in SSIM declined. MSP obtained a brighter visual effect and the fastest processing efficiency. Only a few areas emerge hazy, and nonhomogeneous haze was decomposed and processed separately in the multi-scale-patch framework. As shown in [Table 2](#), MSP achieved prominent results on most metrics.

5. Ablation Study

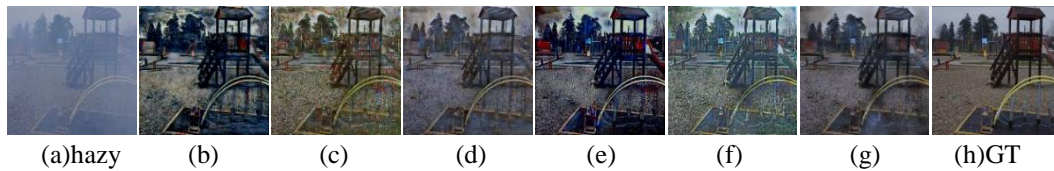


Fig. 8. The results of different network structure adjustment on dataset NTIRE18.

Table 3. Network architecture adjustment and comparison

		PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow	AR(s) \downarrow	RT(s) \downarrow
70*(1024*1024) Trainset for Epoch = 1200	Baseline Multi-s (b)	16.43	0.60	0.53	1.29	0.018
	Baseline + Multi s-p (c)	17.53	0.66	0.51	1.28	0.020
	Baseline + Multi s-p + supplement connection (d)	20.43	0.67	0.37	1.29	0.026
1120*(256*256) Trainset for Epoch = 600	Baseline Multi-s (e)	16.84	0.51	0.52	1.28	0.018
	Baseline + Multi s-p (f)	17.33	0.60	0.47	1.28	0.019
	Baseline + Multi s-p + supplement connection (g)	20.87	0.72	0.28	1.30	0.026

To demonstrate the effectiveness and superiority of the proposed multiple scale-patch framework, an ablation study was conducted by gradually modifying the baseline model and comparing the performances of all models.

For the comparison of the two datasets including 70pcs \times 1024 \times 1024 and 1120pcs \times 256 \times 256, there were three different designs: (1) The baseline: multi-scale, (2) Multi-scale-patch, and (3) MSP. For fair comparison, the same network architecture and training setting parameters were kept for all of the aforementioned models, except the modification depicted in [Table 3](#).

The experimental results are shown in [Fig. 8](#) and [Table 3](#). The test images were from the NTIRE18 dataset. Compared with the baseline, the MSP model obtained by training with the multiple scale-patch structure exhibited improvements on three metrics. The most significant improvement was 13.5% on the SSIM metric, demonstrating the model quality of the proposed framework was improved after training. In control groups (d) and (g), 24.13% PSNR, 25.22% SSIM and 61.53% LPIPS parameters were improved by sacrificing 0.005s processing time for single images and the improvement in image quality was worth the trade-off. In order to

demonstrate the robustness of MSP in the small dataset, 70 training images were used to train models with defective experimental results, but better results than other methods were achieved, which also demonstrates the feasibility of the present model training on small datasets.

6. Conclusion

This paper has summarized and discussed the issues exist in present dehazing methods. Proposed MSP has achieved superior results in generalization performance and industrial transplant ability compare to the most said algorithms supported by the current findings. The key to a lightweight and fast dehazing network is utilizing multi-scale-patch architecture, which is a simple-constructed and effective feature extractor and complementor that maintains images' detailed edges and textures. Meanwhile, an improved image segmentation method aiming at the common features of hazy map in the real world is proposed and utilized for better processing effect. Predominant experiment and ablation study demonstrate that the MSP outperforms the algorithms for comparison in terms of most metrics, especially in inference time for real-world dehazing and the frame superiority.

More information on expanded dataset and particular modules would help us to establish a greater degree of precision restoration and inference speed on this matter according to usage scenarios. The MSP-based video dehazing algorithm in the real scene is also what we are currently researching in.

Acknowledgement

Funding: The present study was supported by the National Natural Science Foundation of China (grant numbers 62006073).

References

- [1] McCartney, E. J., "Scattering phenomena. (book reviews: Optics of the atmosphere. scattering by molecules and particles)," *Science*, 196, 1084–1085, 1977.
- [2] K. He, J. Sun and X. Tang, "Single Image Haze Removal Using Dark Channel Prior," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 12, pp. 2341-2353, Dec. 2011. [Article \(CrossRef Link\)](#)
- [3] R. T. Tan, "Visibility in bad weather from a single image," in *Proc. of 2008 IEEE Conference on Computer Vision and Pattern Recognition*, 2008. [Article \(CrossRef Link\)](#)
- [4] Q. Zhu, J. Mai and L. Shao, "A Fast Single Image Haze Removal Algorithm Using Color Attenuation Prior," *IEEE Transactions on Image Processing*, vol. 24, no. 11, pp. 3522-3533, Nov. 2015. [Article \(CrossRef Link\)](#)
- [5] B. Cai, X. Xu, K. Jia, C. Qing and D. Tao, "DehazeNet: An End-to-End System for Single Image Haze Removal," *IEEE Transactions on Image Processing*, vol. 25, no. 11, pp. 5187-5198, Nov. 2016. [Article \(CrossRef Link\)](#)
- [6] A Benoit, Leonel Cuevas, Jean-Baptiste Thomas, "Deep learning for dehazing: Comparison and analysis," in *Proc. of Colour and Visual Computing Symposium (CVCS)*, 2018. [Article \(CrossRef Link\)](#)
- [7] B. Li, X. Peng, Z. Wang, J. Xu, and D. Feng, "AOD-Net: All-in-one dehazing network," in *Proc. of the IEEE International Conference on Computer Vision*, vol. 1, p. 7, 2017. [Article \(CrossRef Link\)](#)

- [8] Zhang, H., Patel, V. M., Patel, V. M. and Patel, V. M., “Densely connected pyramid dehazing network,” in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, 3194–3203, 2018. [Article \(CrossRef Link\)](#)
- [9] X.H. Liu, Y.R. Ma, Z.H. Shi and J. Chen, “GridDehazeNet: Attention-Based Multi-Scale Network for Image Dehazing,” in *Proc. of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 7314-7323, 2019. [Article \(CrossRef Link\)](#)
- [10] W.Q. Ren, L. Ma, J.W. Zhang, J.S. Pan, X.C. Cao, W. Liu, M.H. Yang, “Gated Fusion Network for Single Image Dehazing,” in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3253-3261, 2018. [Article \(CrossRef Link\)](#)
- [11] H. Dong, J.S. Pan, L. Xiang, Z. Hu, X.Y. Zhang, F. Wang, M.H. Yang, “Multi-Scale Boosted Dehazing Network With Dense Feature Fusion,” in *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2154-2167, 2020. [Article \(CrossRef Link\)](#)
- [12] Y. Dong, Y. Liu, H. Zhang, S. Chen, Y. Qiao, “FD-GAN: Generative Adversarial Networks with Fusion-discriminator for Single Image Dehazing,” in *Proc. of the AAAI Conference on Artificial Intelligence*, vol. 34(07), pp. 10729-10736. 2020. [Article \(CrossRef Link\)](#)
- [13] H. Zhu, X. Peng, V. Chandrasekhar, L. Li, J.H. Lim, “DehazeGAN: When Image Dehazing Meets Differential Programming,” *IJCAI*, 2018. [Article \(CrossRef Link\)](#)
- [14] W. Wang, A. Wang, Q. Ai, C. Liu and J. Liu, “AAGAN: Enhanced Single Image Dehazing With Attention-to-Attention Generative Adversarial Network,” *IEEE Access*, vol. 7, pp. 173485-173498, 2019. [Article \(CrossRef Link\)](#)
- [15] N. Wang, Y.B. Zhou, F.L. Han, H.Y. Zhu, Y.J. Zheng, “UWGANUnderwater GAN for Real-world Underwater Color,” *arXiv:1912.10269*. 2019. [Article \(CrossRef Link\)](#)
- [16] J. Zhang, Y. Cao, and Z.F. Wang, “Nighttime haze removal based on a new imaging model,” in *Proc. of 2014 IEEE International Conference on Image Processing (ICIP)*, pp. 4557–4561, 2014. [Article \(CrossRef Link\)](#)
- [17] Y. Li, R. T. Tan, and M. S. Brown, “Nighttime haze removal with glow and multiple light colors,” in *Proc. of the IEEE International Conference on Computer Vision (ICCV)*, pp. 226-234, 2015. [Article \(CrossRef Link\)](#)
- [18] J. Zhang, Y. Cao, S. Fang, Y. Kang, and C. W. Chen, “Fast haze removal for nighttime image using maximum reflectance prior,” in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 7418-7426, 2017. [Article \(CrossRef Link\)](#)
- [19] J. Zhang, Y. Cao, Z. J. Zha, D. C. Tao, “Nighttime Dehazing with a Synthetic Benchmark,” in *Proc. of the 28th ACM International Conference on Multimedia*, 2020. [Article \(CrossRef Link\)](#)
- [20] S. D. Das, S.t Dutta, “Fast Deep Multi-Patch Hierarchical Network for Nonhomogeneous Image Dehazing,” in *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pp. 482-483, 2020. [Article \(CrossRef Link\)](#)
- [21] T. T. Guo, X. L. Li, V. Cherukuri, and V. Monga, “Dense scene information estimation network for dehazing,” in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pp. 2122-2130, 2019. [Article \(CrossRef Link\)](#)
- [22] T. T. Guo, V. Cherukuri, and V. Monga, “Dense ’123’ color enhancement dehazing network,” in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 2131-2139, 2019. [Article \(CrossRef Link\)](#)
- [23] X. Qin, Z. L. Wang, Y. C. Bai, X. D. Xie, and H. Z. Jia, “FFA-Net: Feature fusion attention network for single image dehazing,” in *Proc. of the AAAI Conference on Artificial Intelligence*, vol. 34(07), pp. 11908-11915, 2020. [Article \(CrossRef Link\)](#)
- [24] R. T. Li, X. Y. Zhang, S. D. You, Y. Li, “Learning to Dehaze From Realistic Scene with A Fast Physics Based Dehazing Network,” *arXiv preprint arXiv:2004.08554*, 2020. [Article \(CrossRef Link\)](#)
- [25] Yang, X., Xu, Z., Luo, J., “Towards Perceptual Image Dehazing by Physics-Based Disentanglement and Adversarial Training,” in *Proc. of the AAAI Conference on Artificial Intelligence*, vol. 32(1), 2018. [Article \(CrossRef Link\)](#)

- [26] D. Chen et al., “Gated Context Aggregation Network for Image Dehazing and Deraining,” in *Proc. of 2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 1375-1383, 2019. [Article \(CrossRef Link\)](#)
- [27] Ronneberger O., Fischer P., Brox T., “U-Net: Convolutional Networks for Biomedical Image Segmentation,” In: *Navab N., Hornegger J., Wells W., Frangi A. (eds) Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. vol 9351, pp. 234-241, 2015. [Article \(CrossRef Link\)](#)
- [28] D. Engin, A. Genc, H. K. Ekenel, “Cycle-Dehaze: Enhanced CycleGAN for Single Image Dehazing,” in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pp. 825-833, 2018. [Article \(CrossRef Link\)](#)
- [29] J. Y. Zhu, T. Park, P. Isola, A. A. Efros, “Unpaired Image-To-Image Translation Using Cycle-Consistent Adversarial Networks,” in *Proc. of the IEEE International Conference on Computer Vision (ICCV)*, pp. 2242 – 2251, 2017. [Article \(CrossRef Link\)](#)
- [30] Y. Y. Qu, Y. Z. Chen, J. Y. Huang, Y. Xie, “Enhanced Pix2pix Dehazing Network,” in *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 8152-8160, 2019. [Article \(CrossRef Link\)](#)
- [31] A. Mehta, H. Sinha, P. Narang, M. Mandal, “HIDEGAN: A Hyperspectral-guided Image Dehazing GAN,” in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pp. 846-856, 2020. [Article \(CrossRef Link\)](#)
- [32] Y.J. Shao, L. Li, W. Q. Ren, C. X. Gao, N. Sang, “Domain Adaptation for Image Dehazing,” in *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2805-2814, 2020. [Article \(CrossRef Link\)](#)
- [33] Y. K. Yu, H. Liu, M. H. Fu, J. Chen, X. Y. Wang, K. Y. Wang, “A Two-Branch Neural Network for Non-Homogeneous Dehazing via Ensemble Learning,” in *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pp. 193-202, 2021. [Article \(CrossRef Link\)](#)
- [34] H. Y. Wu, Y. Y. Qu, S. H. Lin, J. Zhou, R. Z. Qiao, Z. Z. Zhang, Y. Xie, L. Z. Ma, “Contrastive Learning for Compact Single Image Dehazing,” in *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 10551-10560, 2021. [Article \(CrossRef Link\)](#)
- [35] S. Zhao, L. Zhang, Y. Shen and Y. Zhou, “RefineDNet: A Weakly Supervised Refinement Framework for Single Image Dehazing,” in *Proc. of IEEE Transactions on Image Processing (TIP)*, vol. 30, pp. 3391-3404, 2021. [Article \(CrossRef Link\)](#)
- [36] H. G. Zhang, Y. C. Dai, H. D. Li, and P. Koniusz, “Deep stacked hierarchical multi-patch network for image deblurring,” in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5971-5979, 2019. [Article \(CrossRef Link\)](#)
- [37] Ancuti, C., Ancuti, C. O., and Timofte, R., “Ntire 2018 challenge on imagedehazing: Methods and results,” in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pp. 891–901, 2018. [Article \(CrossRef Link\)](#)
- [38] C.O. Ancuti, C. Ancuti, R. Timofte, “NH-HAZE: An Image Dehazing Benchmark with Non-Homogeneous Hazy and Haze-Free Images,” in *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pp. 1798-1805, 2020. [Article \(CrossRef Link\)](#)
- [39] B. Li, W. Ren, D. Fu, D. Tao, D. Feng, W. Zeng, and Z. Wang, “Benchmarking single-image dehazing and beyond,” *IEEE Transactions on Image Processing (TIP)*, vol. 28(1), pp. 492– 505, 2019. [Article \(CrossRef Link\)](#)
- [40] C. Ancuti, C. O. Ancuti, and C. De Vleeschouwer, “D-HAZY: a dataset to evaluate quantitatively dehazing algorithms,” in *Proc. of IEEE International Conference on Image Processing (ICIP)*, IEEE, pp. 2226-2230, 2016. [Article \(CrossRef Link\)](#)
- [41] Ancuti C., Ancuti C.O., Timofte R., De Vleeschouwer C, “I-HAZE: a dehazing benchmark with real hazy and haze-free indoor images,” in *Proc. of International Conference on Advanced Concepts for Intelligent Vision Systems*, Springer, Cham, 2018. [Article \(CrossRef Link\)](#)
- [42] C. O. Ancuti, C. Ancuti, R. Timofte, C.D. Vleeschouwer, “O-HAZE: a dehazing benchmark with real hazy and haze-free outdoor images,” in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pp. 754-762, 2016. [Article \(CrossRef Link\)](#)

- [43] Zhang, Y. F., Li D., and G. Sharma, "Hazerd: an outdoor scene dataset and benchmark for single image dehazing," in *Proc. of 2017 IEEE International Conference on Image Processing (ICIP)*, IEEE, pp. 3205-3209, 2017. [Article \(CrossRef Link\)](#)
- [44] A. Gaidon, Q. Wang, Y. Cabon, E. Vig, "Virtual worlds as proxy for multi-object tracking analysis," in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4340-4349, 2016. [Article \(CrossRef Link\)](#)
- [45] S. Zhao, L. Zhang, S. Huang, Y. Shen and S. Zhao, "Dehazing Evaluation: Real-World Benchmark Datasets, Criteria, and Baselines," *IEEE Transactions on Image Processing*, vol. 29, pp. 6947-6962, 2020. [Article \(CrossRef Link\)](#)
- [46] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric," in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 586-595, 2018. [Article \(CrossRef Link\)](#)



Juan Wang received her B.S. degree from Nanyang Normal University in 2007 and M.S. degree from Xihua University in 2010. and her PhD from Tianjin University in 2015. She is currently a faculty member at Hubei University of Technology. Her re-search interests include image processing, pattern recognition, and computer vision.



Chang Ding received his B.S. degrees in Jingchu University of Technology, Jingmen, China, in 2019. He is currently pursuing his master's degree at Hubei University of Technology, Wuhan, China. His current research interests are mainly focused on deep learning and image enhancement.



Minghu Wu received his B.S. degree from Communication University of China in 1998 and M.S. degree from Huazhong University of Science and Technology in 2003. and his PhD from Nanjing University of Posts and Telecommunications in 2013. His research interests include artificial intelligence and image processing.



Yuanyuan Liu received his B.S.degrees in Hubei University of Technology, Wuhan, China, in 2021. He is currently pursuing his master's degree at Fudan university, Shanghai, China. his research interests are intelligent perception and communication.



Guanhai Chen received his B.S.degrees in Hubei Polytechnic University, Huangshi, China, in 2021. He is currently pursuing his master's degree at Hubei University of Technology, Wuhan, China. his