2024 5th International Conference on Computer Vision, Image and Deep Learning (CVIDL) | 979-8-3503-7382-0/24/$31.00 ©2024 IEEE | DOI: 10.1109/CVIDL62147.2024.10603766

ClearSight: Deep Learning-Based Image Dehazing for Enhanced UAV Road Patrol

Yangyang Wang
Institute of Systems Engineering
Academy of Military Sciences
Beijing, China
wyy_961126@163.com

Jie Zhang*
Institute of Systems Engineering
Academy of Military Sciences
Beijing, China
*Corresponding author: 13488811127@163.com

Jian Zhou
Institute of Systems Engineering
Academy of Military Sciences
Beijing, China
zhoujian_sy_sjt@163.com

Mingchao Han
Institute of Systems Engineering
Academy of Military Sciences
Beijing, China
18510958675@163.com

Shaonan Li
Institute of Systems Engineering
Academy of Military Sciences
Beijing, China
tankshaonan@gmail.com

Huijie Miao
Institute of Systems Engineering
Academy of Military Sciences
Beijing, China
mhj0716@163.com

***Abstract*—Adverse weather conditions such as haze can deteriorate the quality of images captured during UAV patrols, thus affecting the effectiveness of object detection algorithms. To address this issue, this paper proposes an end-to-end image dehazing algorithm based on deep learning. Leveraging dilated convolutions and Transformers, we construct an image restoration module to adaptively learn and eliminate degradation information from the images. Furthermore, by adaptively weighting and fusing features from different layers of the network, we enhance the expression capability of image content features, effectively removing haze while preserving image content information. To validate the effectiveness of our approach, we also construct a dataset of foggy and clear images of UAV scenes using generative adversarial learning. Experimental results demonstrate that our proposed method can effectively remove haze from UAV images and significantly improve the accuracy of object detection in haze conditions, thereby enhancing the reliability of UAV patrols.**

Keywords-UAV patrols; image dehazing; deep learning; object detection

I. INTRODUCTION

With the advancement of object detection technology, UAV patrols have become a crucial method for maintaining urban safety. Not only does it significantly reduce the workload of law enforcement personnel, but it also enhances patrol efficiency. However, with the acceleration of urbanization and the continuous increase in industrial activities, haze has become a severe environmental issue faced by contemporary society. Haze not only significantly impacts people's lives and health but also poses challenges to urban management and security work. In hazy weather conditions, images captured by UAVs often suffer from high noise, low contrast, and low brightness. The edges of objects in the images are prone to be confused with the surrounding environment, making it difficult for object detection models to extract useful information and thereby reducing the effectiveness of detection algorithms, which in turn affects the normal operation of UAV patrols.

To improve the image quality of UAVs in hazy weather conditions and enhance object detection accuracy, various haze image enhancement methods have been proposed. These methods mainly include image enhancement-based methods, image restoration-based methods, and deep learning-based methods. Early image enhancement-based dehazing algorithms primarily utilized simple image processing techniques to enhance image contrast and highlight detail features, making the image appear clearer. However, these methods often overlooked the physical mechanism of imaging in foggy conditions and lacked a profound understanding of haze distribution and light propagation processes. Typical early algorithms include histogram equalization, Retinex algorithm, and wavelet transform algorithm, among others. The principle of image dehazing using histogram equalization algorithm and Retinex algorithm is similar to the aforementioned low-light image enhancement methods. Based on the theory of histogram equalization, Kim et al. [1] proposed a block overlapping histogram equalization algorithm, Stark et al. [2] proposed a generalized local histogram equalization algorithm, Xu et al. [3] proposed a contrast-limited adaptive histogram equalization algorithm, and Reza et al. [4] proposed a contrast-limited local histogram equalization algorithm. Based on the Retinex theory, Jobson et al. [5] proposed a multi-scale Retinex algorithm with color restoration, Choi et al. [6] proposed a single-scale Retinex dehazing method based on the minimum brightness difference perceivable by humans, and Fu et al. [7] proposed a variational framework-based Retinex algorithm. Wavelet transform

performs temporal and spatial transformations on images to obtain information and adjusts different frequency domains of the image to improve its visual effect. Wang et al. [8] enhanced foggy images using wavelet transform first, followed by improving image brightness using the Retinex algorithm to remove haze from the image. Russo et al. [9] added multi-scale refinement based on wavelet transform to enhance the dehazing effect on images. Although the above image enhancement-based dehazing methods have improved the visual effects of foggy images to a certain extent, these methods adjust images from a mathematical perspective to enhance contrast without considering the physical degradation process of foggy images. Therefore, they often suffer from issues such as poor dehazing effects, which affect subsequent object recognition tasks.

Image restoration-based methods analyze the degradation causes of foggy images, establish corresponding foggy image degradation models based on the characteristics of foggy images, and derive haze-free images through inverse solving based on these models. Yitzhaky et al. [10] proposed describing the influence of atmospheric particles on light scattering effects by establishing a foggy image degradation function, providing a theoretical basis for subsequent image dehazing through foggy image degradation models. However, the difficulty in obtaining function parameters hindered its widespread application. Nayar and Narasimhan [11-13] extensively studied the physical degradation processes of images under different weather conditions and proposed the most commonly used foggy image degradation model to date—the atmospheric scattering model. In recent years, many image dehazing algorithms have solved various model parameters by adding various known prior knowledge to the atmospheric scattering model and subsequently calculated haze-free images using the atmospheric scattering model formula in reverse. Tan et al. [14] observed higher contrast in haze-free images compared to foggy images of the same scene. They transformed foggy images into the chrominance space to obtain the environmental atmospheric light value and used it as prior knowledge to solve for other model parameters. However, this method led to oversaturated colors in enhanced images due to the neglect of scattering effects on image colors. Fattal et al. [15] discovered that reflectance in local regions of the image remains constant. They calculated the reflectance of all regions of the image by block processing and used it as prior knowledge to compute the atmospheric scattering model. However, this method requires significant color differences in images to achieve good results, making it unsuitable for grayscale images and dense fog scenarios. He et al. [16] analyzed a large number of images and observed that some pixel values in fog-free images tended towards zero in a color channel, termed as the dark channel prior. Using this as prior knowledge, they estimated the atmospheric transmission rate and achieved fog removal from foggy images. However, this method's computational complexity was high, resulting in slow processing speeds. Zhu et al. [17] analyzed foggy images in the HSV color space and found a positive correlation between fog density variations and differences in image saturation and brightness, which they termed as the color attenuation prior. They used this to establish a linear regression model for predicting atmospheric transmission rates and implemented image dehazing. Berman et al. [18] observed that pixel distributions in foggy images were linear in the RGB color space. Using this as prior knowledge, they estimated atmospheric transmission rates and achieved global image dehazing. However, strong illumination could attenuate color information in images, affecting dehazing algorithm effectiveness. These image restoration-based dehazing algorithms consider the fundamental causes of image degradation, resulting in more ideal and reliable dehazing results. However, these methods rely on manually designed prior knowledge and cannot adaptively update parameters based on the actual characteristics of degraded images, making it challenging to accommodate all complex and variable foggy scenes. Therefore, there is a need to explore more universally applicable dehazing methods.

With the rapid advancement of deep learning technology, many researchers have begun to apply deep learning techniques to image dehazing. Ren et al. [19] proposed a multi-scale image dehazing network that accurately estimates the image transmission rate using two scales of convolutional neural networks. Cai et al. [20] introduced DehazeNet based on convolutional neural networks, which can estimate the image transmission rate in an end-to-end manner. These dehazing algorithms estimate the image transmission rate through convolutional neural networks and then dehaze the image based on the atmospheric scattering model. Although they have achieved good results, their dehazing effects still rely on the atmospheric scattering model, thus having many limitations. In recent years, many end-to-end image dehazing algorithms have been proposed. Li et al. [21] proposed the AOD-Net network, which does not require the estimation of transmission rate and atmospheric light and can directly output the dehazed image through deep neural networks, avoiding the influence of parameter errors on dehazing effects. Mei et al. [22] proposed the FFA-Net, a U-Net-based image dehazing network, which learns the mapping function from foggy images to clear images through progressive feature fusion, achieving adaptive image dehazing. Chen et al. [23] proposed GCANet, which aggregates contextual information of each pixel using dilated residual structures and enhances feature expression by fusing features from different levels, achieving excellent image dehazing effects. These end-to-end fog networks based on deep learning have shown remarkable performance in image dehazing. However, their success largely depends on a large amount of foggy image and noisy image data from similar scenes for neural network training. Therefore, the training process of the model has a considerable threshold and requires handling a large amount of complex data.

Taking into account existing image dehazing methods, image enhancement-based techniques typically only improve the visual appearance of images without effectively handling real foggy images. Although image restoration-based methods can handle foggy images, they often require complex prior knowledge and atmospheric scattering models, making them difficult to adapt to various complex foggy scenes. In contrast, deep learning-based methods have made significant progress in image dehazing, but they demand large-scale, high-quality training datasets. Therefore, this paper proposes an end-to-end UAV image dehazing algorithm based on deep learning. It utilizes generative adversarial learning to construct foggy

image and clear image pairs adapted to real scenarios, aiming to achieve high-quality dehazing of UAV images and mitigate the impact of adverse weather conditions on image quality in UAV patrols and other applications.

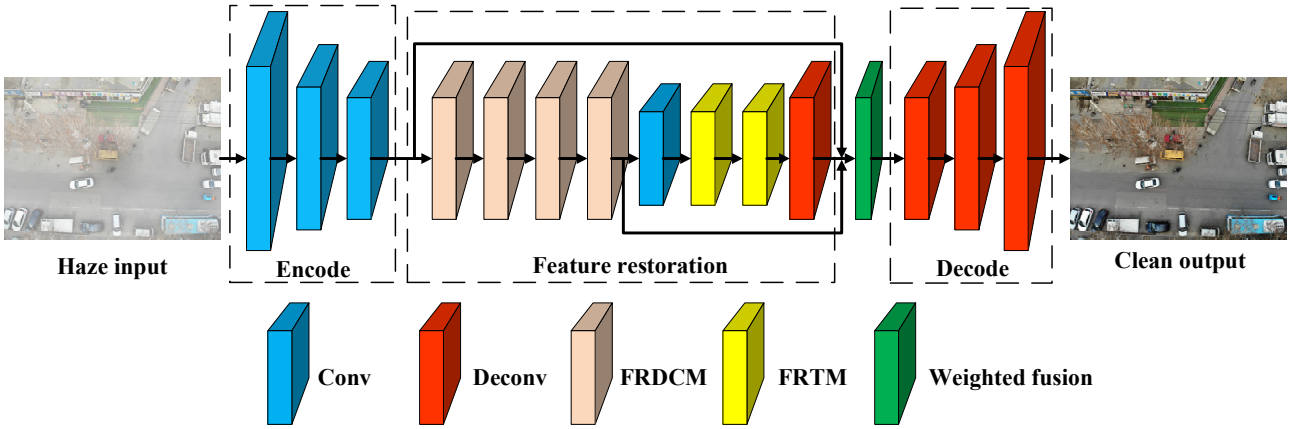

Figure 1. Overall framework of Image Dehazing Network

II. METHOD

To achieve dehazing of foggy UAV images, we designed an end-to-end image dehazing network based on deep learning techniques. The network structure is illustrated in Fig. 1, which mainly consists of three parts: an encoder module, a decoder module, and a feature restoration module.

The image dehazing network first adjusts the dimensions of the input image to multiples of 8 to ensure that the output image has the same size as the input image. Then, the input low-quality image is encoded into feature maps by the encoder module, which consists of three convolutional blocks. Each convolutional block comprises a standard 3×3 convolution followed by an Instance Normalization (IN) layer. The second and third convolutional blocks perform 1/2 downsampling on the feature maps. The feature restoration module primarily includes a Feature Restoration Dilated Convolution Module (FRDCM) and a Feature Restoration Transformer Module (FRTM), which adaptively learn and remove degradation information from the input feature maps. To preserve more content information from the input image in the output image, the network utilizes a weighted feature fusion method to merge features from different levels within the network, enhancing the expression capability of image features. Specifically, the network employs a weighted fusion module to merge input features F_I, output features F_O from the feature restoration module, and features F_M from the fourth level. The mathematical expression for this process is shown in (1).

$$\begin{aligned}(\lambda_I, \lambda_M, \lambda_O) &= \chi(F_I, F_M, F_O) \\ F &= \lambda_I * F_I + \lambda_M * F_M + \lambda_O * F_O\end{aligned} \tag{1}$$

In the equation, λ_I, λ_M and λ_O represent the fusion weights for the input features F_I, features F_M from the fourth level, and output features F_O of the feature restoration module, respectively. χ denotes the weighted fusion submodule, which is implemented through a standard 3×3 convolution in this network. Finally, the restored feature maps are decoded by the decoder to obtain the output image with the same dimensions as the input image. The decoder consists of two transpose convolution blocks and one convolution block. The first two transpose convolution blocks comprise a 4×4 transpose convolution followed by an IN layer, performing 2x upsampling on the feature maps. The last convolution block consists of a standard 3×3 convolution followed by an IN layer, which integrates channel information from the feature maps and adjusts the number of channels to 3, resulting in a clean three-channel output image.

In the feature restoration module, the FRDCM module consists of four layers of dilated convolution and IN layers, as shown in Fig. 2. This module comprises a residual structure consisting of two layers of 3×3 dilated convolution, IN layers, ReLU activation function layers, and a shortcut connection. The role of dilated convolution is to obtain a larger receptive field without increasing computational complexity, thus better restoring the degraded features of the input image. Since convolutional layers can only capture local feature information of images, which is highly limiting for tasks like image dehazing that require understanding of scene context, we designed the FRTM module based on Transformer, as illustrated in Fig. 3. The FRTM module captures global information from the feature map through Transformer, further enhancing the dehazing algorithm's effectiveness and robustness. As the computational complexity and memory usage of Transformer are quadratic with respect to the feature resolution, we first downsample the feature map by 2x using convolutional layers before FRTM to reduce the network's computational complexity and memory usage. We then use transpose convolution after FRTM to upsample the feature map to its original size for subsequent feature fusion.

To efficiently train the dehazing network, we employ the mean square error loss function as the overall loss function of the network, as shown in (2).

$$Loss = \left\| \hat{r} - r \right\|^2 \tag{2}$$

In the equation, $\hat{r}$ represents the output image, and r represents the ground truth image.

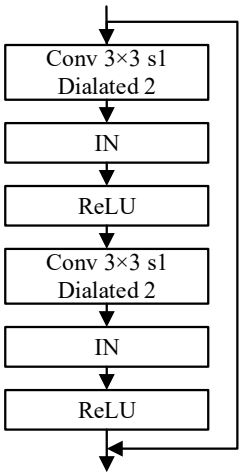

Figure 2. Feature restoration dilated convolution module

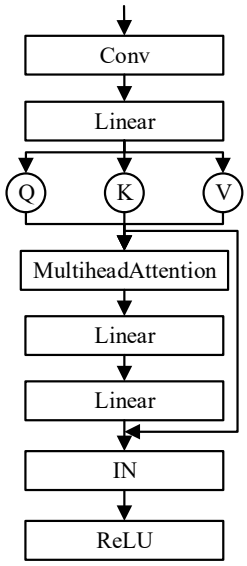

Figure 3. Feature restoration Transformer module

III. EXPERIMENT

In this section, experiments are conducted to validate the practical dehazing performance of the image dehazing algorithm and the recognition effectiveness of ground typical objects by UAVs equipped with the dehazing algorithm in foggy weather conditions.

(1) Experimental Setup

The experiments were conducted on a server equipped with an NVIDIA RTX 8000 GPU and an Intel Xeon Gold 6258R processor. The operating system used was Ubuntu 20.04 LTS. The software environment included Python 3.10 programming language, PyTorch 2.0.0 deep learning framework, Torchvision 0.15.0 deep learning toolkit, OpenCV 4.8.1 computer vision library, and CUDA 11.3 GPU driver.

(2) Experimental Parameter Settings

Regarding the image dehazing algorithm, Adam optimization algorithm is employed for model training. The maximum training epochs are set to 100, with an initial learning rate of 0.001. The learning rate is decayed by a factor of 10 at the 40th and 80th epochs, and momentum decay is set to 0.0005. The batch size is set to 4. Since the subsequent object detection algorithm requires input image resolution of 640×640, the input image resolution is adjusted to 640×640 accordingly.

For the object detection algorithm, YOLOv5s object detection model is utilized for experimental validation. The model is trained using Stochastic Gradient Descent with Momentum (SGDM) and learning rate decay. The maximum training epochs are set to 100, with a batch size of 128. Weight Decay and Momentum parameters are set to 0.005 and 0.937, respectively. The initial learning rate is set to 0.01 and gradually decayed as training epochs increase. The input image resolution is set to 640×640, and Mosaic and Mixup data augmentation methods are employed for training data augmentation.

Figure 4. Example of image dehazing dataset

(3) Experimental Dataset

Based on deep learning-based image dehazing algorithms falling within the supervised learning category, a dataset consisting of image pairs of foggy and clear weather conditions is required for model training. However, constructing such a dataset through manual collection in real-world scenarios is practically infeasible. Therefore, this study first collected a large number of clear weather images using UAVs in various urban and suburban locations and time periods. Subsequently, a dataset containing 1928 pairs of images was constructed by randomly adding different concentrations of haze to these images using generative adversarial learning methods. Partial examples of the dataset are illustrated in Fig. 4.

To evaluate the impact of integrating an image dehazing algorithm on object detection performance, this section annotates images in the image dehazing dataset with seven classes of objects: sedan, person, motor, bicycle, truck, bus, and tricycle, based on the actual needs of UAV Patrol. Table I shows the number of samples for each class in the dataset. Regarding data allocation, the dataset is randomly divided into training, validation, and testing sets in a 6:2:2 ratio.

Table I. Number of samples of each class in haze scene dataset

Class	sedan	person	tricycle	bicycle	motor	truck	bus	total
Number	26062	24312	832	6540	3543	2022	737	64048

Table II. Experimental comparison results of image dehaze

Methods	MSE ↓	PSNR ↑	SSIM ↑
DCP	0.0110	20.06	0.8569
DehazeNet	0.0437	14.58	0.8296
FFA-Net	0.0274	17.79	0.8949
MSBDN-DFF	0.0245	17.67	0.8990
Ours	**0.0003**	**35.74**	**0.9894**

(4) Experimental Results

Regarding the image dehazing algorithm, this section conducts comparative experiments with mainstream image dehazing algorithms (DCP, DehazeNet, FFA-Net, and MSBDN-DFF). Objective image quality evaluation metrics including Mean Squared Error (MSE), Peak Signal-to-Noise Ratio (PSNR), and Structural Similarity Index (SSIM) are utilized to assess the dehazing performance of the algorithms. The experimental results are presented in Table II. From the experimental results of image dehazing, it can be observed that the UAV image adaptive dehazing algorithm designed in this paper based on deep learning can effectively achieve image dehazing. In the validation set, the enhanced images achieve a PSNR of up to 35.74 compared to clear images, which is significantly higher than the DCP algorithm based on image restoration, the DehazeNet algorithm based on deep neural networks and atmospheric scattering models, and other deep learning-based algorithms (FFA-Net, MSBDN-DFF). The MSE and SSIM metrics also reach 0.0003 and 0.9894, respectively, which are also far superior to other algorithms. Fig. 5 illustrates the actual dehazing effect of the image dehazing algorithm designed in this section. From the subjective quality evaluation perspective, this method indeed achieves excellent visual results, laying a solid foundation for subsequent high-precision object detection.

Table III. Experimental results of haze image preprocessing

Class	sedan	person	tricycle	bicycle	motor	truck	bus	mAP
haze	97.0	67.3	44.5	46.3	56.2	80.3	95.9	70.3
dehaze	**97.8**	**73.0**	**67.4**	**62.4**	**68.6**	**88.8**	**96.4**	**79.2**

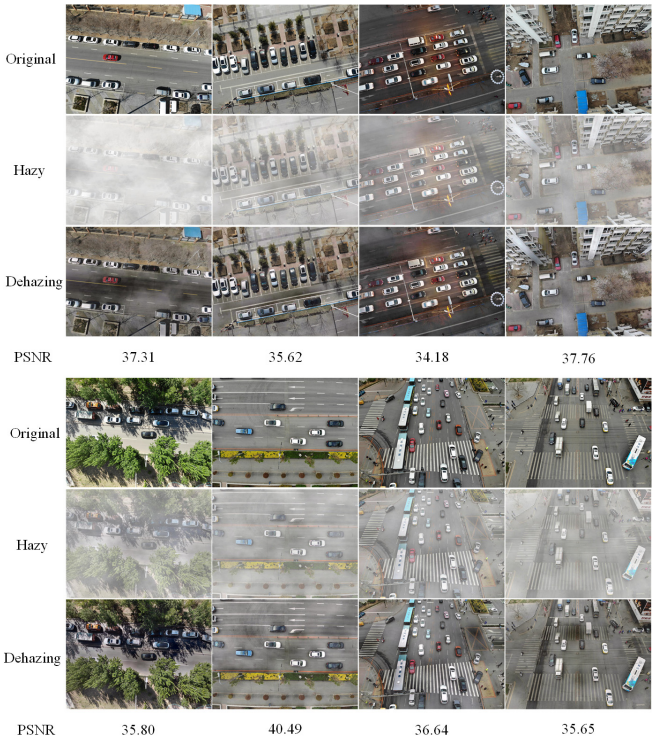

Figure 5. Experimental results of image dehazing

Figure 6. Visualization results of object detection experiments

Regarding the object detection algorithm, this section utilizes Average Precision (%) and Mean Average Precision (%) to evaluate the detection performance of the algorithm. The experimental results are presented in Table III. From the experimental results, it is evident that under identical conditions, the YOLOv5s object detection model achieves higher detection accuracy for all categories in the dehazed image dataset compared to the hazy image dataset. Particularly noticeable improvements are observed for small to medium-sized objects in categories such as person, tricycle, bicycle, and motor. The use of the image dehazing algorithm significantly enhances detection accuracy, with an average precision increase of 8.9% mAP in the dehazed image dataset compared to the hazy image dataset. Fig. 6 illustrates the actual detection results in both datasets, indicating visibly lower rates of both missed detections and false alarms in the dehazed images. Hence, the haze image quality enhancement method proposed in this paper effectively enhances the inspection quality of UAVs in adverse weather conditions such as haze.

IV. CONCLUSION

This paper addresses the impact of haze weather on UAV patrols by proposing an end-to-end image dehazing algorithm based on deep learning. It utilizes the Feature Restoration Dilated Convolution Module (FRDCM) constructed with dilated convolutions and the Feature Restoration Transformer Module (FRTM) built with Transformers to adaptively learn and eliminate degradation information in images. Additionally, it enhances the expression capability of image features by weighted fusion of features at different levels within the network, effectively removing haze while preserving image content information. Experimental results demonstrate that the proposed method outperforms mainstream image dehazing algorithms. Furthermore, the object detection experiments indicate that deploying this dehazing algorithm can significantly improve the detection accuracy of ground objects by UAVs in hazy weather conditions.

REFERENCES

[1] Kim T K, Paik J K, Kang B S. Contrast enhancement system using spatially adaptive histogram equalization with temporal filtering[J]. IEEE Transactions on Consumer Electronics, 1998, 44(1): 82-87.

[2] Stark J A. Adaptive image contrast enhancement using generalizations of histogram equalization[J]. IEEE Transactions on image processing, 2000, 9(5): 889-896.

[3] Xu Z, Liu X, Chen X. Fog removal from video sequences using contrast limited adaptive histogram equalization[C]//2009 International Conference on Computational Intelligence and Software Engineering. IEEE, 2009: 1-4.

[4] REZA A M. Realization of the contrast limited adaptive histogram equalization (CLAHE) for real-time image enhancement [J]. Journal of VLSI signal processing systems for signal, image and video technology, 2004, 38: 35-44.

[5] Jobson D J, Rahman Z, Woodell G A. A multiscale retinex for bridging the gap between color images and the human observation of scenes[J]. IEEE Transactions on Image processing, 1997, 6(7): 965-976.

[6] Choi D H, Jang I H, Kim M H, et al. Color image enhancement based on single-scale retinex with a JND-based nonlinear filter[C]//2007 IEEE International Symposium on Circuits and Systems. IEEE, 2007: 3948-3951.

[7] Fu X, Sun Y, LiWang M, et al. A novel retinex based approach for image enhancement with illumination adjustment[C]//2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2014: 1190-1194.

[8] Wang M, Zhou S. The study of color image defogging based on wavelet transform and single scale retinex[C]//International Symposium on Photoelectronic Detection and Imaging 2011: Advances in Imaging Detectors and Applications. SPIE, 2011, 8194: 111-117.

[9] Russo F. An image enhancement technique combining sharpening and noise reduction[J]. IEEE Transactions on Instrumentation and Measurement, 2002, 51(4): 824-828.

[10] Yitzhaky Y, Dror I, Kopeika N S. Restoration of atmospherically blurred images according to weather-predicted atmospheric modulation transfer functions[J]. Optical Engineering, 1997, 36(11): 3064-3072.

[11] Nayar S K, Narasimhan S G. Vision in bad weather[C]//Proceedings of the seventh IEEE international conference on computer vision. IEEE, 1999, 2: 820-827.

[12] NARASIMHAN S G, NAYAR S K. Vision and the atmosphere [J]. International journal of computer vision, 2002, 48(3): 233.

[13] Narasimhan S G, Nayar S K. Contrast restoration of weather degraded images[J]. IEEE transactions on pattern analysis and machine intelligence, 2003, 25(6): 713-724.

[14] Tan R T, Pettersson N, Petersson L. Visibility enhancement for roads with foggy or hazy scenes[C]//2007 IEEE Intelligent Vehicles Symposium. IEEE, 2007: 19-24.

[15] Fattal, Raanan. "Single image dehazing." ACM transactions on graphics (TOG) 27.3 (2008): 1-9.

[16] He K, Sun J, Tang X. Single image haze removal using dark channel prior[J]. IEEE transactions on pattern analysis and machine intelligence, 2010, 33(12): 2341-2353.

[17] ZHU Q, MAI J, SHAO L. A fast single image haze removal algorithm using color attenuation prior [J]. IEEE transactions on image processing, 2015, 24(11): 3522-3533.

[18] Berman D, Avidan S. Non-local image dehazing[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 1674-1682.

[19] Ren W, Liu S, Zhang H, et al. Single image dehazing via multi-scale convolutional neural networks[C]//Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II 14. Springer International Publishing, 2016: 154-169.

[20] CAI B, XU X, JIA K, et al. Dehazenet: An end-to-end system for single image haze removal [J]. IEEE Transactions on Image Processing, 2016, 25(11): 5187-5198.

[21] Li B, Peng X, Wang Z, et al. Aod-net: All-in-one dehazing network[C]//Proceedings of the IEEE international conference on computer vision. 2017: 4770-4778.

[22] Mei K, Jiang A, Li J, et al. Progressive feature fusion network for realistic image dehazing[C]//Computer Vision–ACCV 2018: 14th Asian Conference on Computer Vision, Perth, Australia, December 2–6, 2018, Revised Selected Papers, Part I 14. Springer International Publishing, 2019: 203-215.

[23] Chen D, He M, Fan Q, et al. Gated context aggregation network for image dehazing and deraining[C]//2019 IEEE winter conference on applications of computer vision (WACV). IEEE, 2019: 1375-1383.