Self-Adaptation Feature Attention Network with Multi-Step Fusion for Single Image Dehazing

JIAWEI ZHANG, XIAOCHEN LIU, DONGHUA ZHAO, CHENGUANG WANG, CHONG SHEN, JUN TANG, JUN LIU

Key Laboratory of Instrumentation Science & Dynamic Measurement, Ministry of Education, School of Instrument and Electronics, North University of China, Taiyuan 030051, P. R. CHINA
School of Instrumentation Sciences and Engineering, Southeast University, Nanjing 210096, P. R. CHINA
School of Information and Communication Engineering, North University of China, Taiyuan 030051, P. R. CHINA
Shanxi province Key laboratory of quantum sensing and precision measurement (201905D121001), North University of China, Taiyuan 030051, P. R. CHINA

Abstract—Although single image dehazing has been widely studied as a common low-level computer vision task, it still faces serious challenges such as limited ability to dehaze real foggy pictures. We propose an efficient end-to-end self-adaptation feature attention (SAFA) network with multi-step fusion for this purpose. The proposed SAFA module can adaptively expand the receptive field to obtain the key structure information in space and extract more comprehensive and accurate features. In addition, considering the lack of connection between features acquired at low and high levels in the network, we also implement a multi-step fusion module, which makes the features of different layers in the network complementary effectively in the process of image recovery. The network structure is simplified, and the required computing resources are significantly reduced by decreasing network parameters. For multiple datasets and photographs with real haze, our method demonstrates better efficiency and availability, both quantitatively and qualitatively.

Keywords— image dehazing; self-adaptation feature attention; multi-step fusion

Received: September 9, 2021. Revised: April 15, 2022. Accepted: May 12, 2022. Published: June 28, 2022.

1. Introduction

FOR a long time, input images captured in indistinct scenes show the negative impact on the performance of computer vision tasks gravely. When the environment is affected by the particles floating in the atmosphere, such as smoke, haze, and dust, human activities in nature will be influenced seriously, and our safety even will be threatened due to the lack of visibility. The images are taken outdoors tend to suffer from problems like reduced contrast, which include degraded colors and structural details.

Therefore, single image dehazing has gradually become indispensable research. The purpose of it is to effectively recover the image from the corrupted input, which means, to restore the basic information of the clean pictures. This can be used as a pre-preparation for high-level visual tasks in many fields such as real-time object detection, remote sensing, and automatic transportation. Other computer vision applications that are initially challenged by the hazy environment can also be completed.

Basically, the generation of hazy images can be described by applying the classic atmospheric scattering model [1, 2], Based on the physical atmosphere scattering model, most dehazing methods were proposed rely on prior knowledge of physics and various assumption in early studies [3, 4, 5, 6]. For instance, dark channel prior (DCP) proposed by He et al.[4] is the most representative algorithm among them. In general, this kind of method has gained some achievements in image dehazing. However, their assumptions do not precisely reflect the inherent attributes of the image. Therefore, the performance of these techniques is often actually limited.

With the rising-up and evolution of the deep learning in recent years, it has also been applied to some simple computer vision tasks such as target recognition [7] and image reconstruction [8]. Compared with traditional ways, deep learning method has extraordinary capability and robustness on dehazing capability. Besides, with the remarkable success of convolutional neural network (CNN) techniques for image dehazing, more and more research teams are tending to use the similar methods to estimate atmospheric light and transmit maps to achieve the desired effect by using external data. For instance, the transport map is decided to be in an end-to-end way in DehazeNet [9]. And in the following research [10, 11, 12, 13, 14], all kinds of novel techniques have also been

gradually added to this field to strengthen the haze removal effect of the network. Due to the strong expression of deep learning networks, these end-to-end network models often have the ability to gain much better dehazing effects than previous work. But the haze from the real world is much more complex than the simulated one, which makes it harder for these methods to process real-world haze images. On the other hand, all of them inevitably need huge computation cost to support. Previous studies [13, 15, 16, 17, 18] have paid too much attention to improve dehazing performance by greatly increasing the depth or width of models and using vast training parameters. But they have not taken into account reasonably the time consumption, memory, or computing consumption, which also makes these models can not be applied in resource-limited environments (such as mobile devices).

In this text, we propose an end-to-end self-adaptation feature attention (SAFA) network with multi-step fusion for single image dehazing. The convolution kernel with the fixed shape is usually adopted in the previous CNN-based image dehazing network, which results in the structural cues in the feature space cannot be effectively utilized. The SAFA module we proposed can adaptively adjust the deformable convolution kernel during the training process to obtain and deal with the crucial structural information in the space. In addition, the application of a multi-step fusion module makes the features of different levels in the network efficiently combined. This network not only reduces the computation cost with a compact and simplified network structure, but also shows excellent visual effects and metrics on several datasets as well as real foggy images.

Basically, the main contributions of this paper include:

A self-adaptation feature attention (SAFA) module is proposed, which integrates the attention mechanism and deformable convolution mechanism. This module is capable of paying more attention to dense haze areas and handling different kinds of complex information adaptively. Besides, the uncomplicated network structure also avoids great computational consumption.

A multi-step fusion module is developed, which is capable of fusing the features of disparate steps adaptively and supplementing each other to get the haze-free image.

An efficient end-to-end self-adaptation feature attention (SAFA) network with multi-step fusion for single image dehazing is implemented, which is combined with the above modules. Moreover, we adjust the network and carry out exhaustive experiments to get the best performance on public datasets as well as real images with fog. Abundant experimental results highlight the validity and practicability of our dehazing network compared with the state-of-the-art (SOTA) methods.

2. Related Work

In the past, most image dehazing work was relied on external information, such as available geo-referenced models [19, 20] or information obtained from other sources [21]. However, due to the unknown nature of transmission map and global atmospheric light, there is no suitable external information for image dehazing in real application. For this extremely challenging task, the current solutions are generally divided into two categories: the classical priority-based ways and the novel deep learning-based methods. But either way, the fundamental problem that how to deal with transmission map and atmospheric light is still remained.

2.1 Priority-based Image Dehazing Methods

The priority-based method for image dehazing usually depends on the atmospheric scattering model. It utilizes certain assumptions or priors to estimate the atmospheric light and transmission map and also takes advantage of additional constraints to compensate for the information lost in the process. Then the image corrupted by haze can be restored to clear. He et al. [4] realized that at least one of the color channels in the haze-affected area has a pixel intensity value close to zero. Based on this, they proposed using dark channel priors to estimate the transmission map and atmospheric light, which is a landmark method of fog removal. By generating a linear model to model the scene depth of the blurred image, Zhu and Mai [22] proposed a valid but not complicated color attenuation prior to recover the depth information to estimate the transmitted map and atmospheric light, thus obtaining the clear image. Tan et al. [23] developed a local contrast maximization dehazing technique to increase the visibility of hazy images, which depends on the theory observed that the contrast of the foggy images is often lower than that of the clean ones. In the research of Fattal et al. [24], an image production model was proposed, which can be applied for scenario transmission and surface shading in order to improve the visibility of the scenario and restore the contrast of haze-free environment. Although these methods achieved impressive results, their performance are still limited by the accuracy of the priors, which are heavily dependent on assumptions and target scenarios. However, it is realized that the lost priors will lead to poor robustness when the scene is becoming more complex. As a result, they are not capable of handling all situations as properly as they used to, such as dehazing the sky area of the image.

2.2 Learning-based Image Dehazing Methods

In recent years, as deep learning has been proven effective in image processing tasks and the availability of related synthetic image datasets, data-driven image dehazing methods have gradually become the mainstream.

Among them, early studies [9, 14, 25] usually apply neural networks for the estimation of the transmission map and atmospheric light in the physical scattering model. For example, a three-layer CNN with the coarse-to-fine method developed by Cai et al. [9], is applied to estimate the media transmission map from existing foggy images to remove haze. For the AOD-Net implemented by Li et al. [14], on the other hand, re-establishes the scattering model through the lightweight CNN with creative design and generates clean images accordingly. But these estimations are not always accurate, which will lead to serious reconstruction errors

between the reconstructed images and the clear ones, such as artifacts and distortion.

Another research strategy is to focus on the end-to-end dehazing network model which uses the neural network to learn the mapping of foggy images to clean ones straightway to complete the dehazing task [13, 15, 16, 26, 27, 28, 29]. A network based on feature fusion attention mechanism (FFA-Net) proposed by Qin et al. [16], which only utilizes a simple loss function L1 to reconstruct the loss, and the combination of different attention mechanisms makes the network more flexible when dealing with disparate information. As Ren et al. [26] presented the multi-scale convolutional neural network (MSCNN) for dehazing image, many essentially similar but fully improved networks were born on this basis, such as the gated fusion network (GFN) [27] and the multi-scale boost dehazing network (MSBDN) [13]. Compared with these methods, the enhanced pix2pix dehazing network (EPDN) implemented by Qu et al. [28] is combined with the generative adversarial network, which is able to reduce the dependence on paired datasets and restoring the haze-free images directly.

2.3 Attention Mechanism

Since the attention mechanism is capable of guiding the network model to dispose of the crucial components in images adaptively, it has been paid more and more attention and applied to a series of computer vision tasks [30, 31, 32, 33] by researchers. For instance, Liu et al. [15] combined a channel-wise attention mechanism with the end-to-end neural network and used the multi-scale estimation technology to guide information exchange and aggregation in the network flexibly. By utilizing the channel attention mechanism, a feature attention dehazing network based on pyramid channels was proposed by Zhang et al. [25] to remove fog in images. Qin et al. [16] also developed a FFA-Net, which includes both channel attention and pixel attention and has the ability to conduct different types of information efficiently.

3. Our Method

3.1 Method Overview

Inspired by the FA module from FFA-Net [16], we propose a new self-adaptive feature attention module (SAFA) as our basic module, and only five of these modules are used in the main architecture of the network. At the same time, a multi-step fusion module is adopted between each SAFA module to realize feature fusion between different steps, which dramatically reduces the memory required for calculation (compared with 57 FA modules in the original network [16]). As shown in the Fig. 1, our network first applies the downsampling operation (such as one convolution with stride 1 and one convolution layer with stride 2, both followed by the ReLU function) for making the subsequent modules obtain the capability to learn the feature representation in the low-resolution domain, and a regular convolution layer for shallow feature extraction. After continuous SAFA modules and multi-step fusion modules, one convolution layer and the related upsampling operation are used to produce the recovered haze-free image.

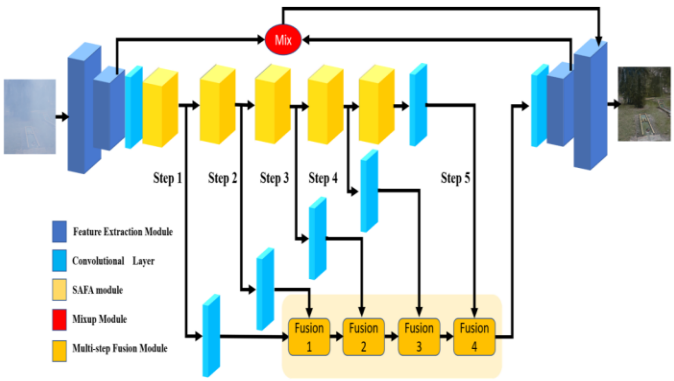

Fig. 1 The architecture of the self-adaptation feature attention network with multi-step fusion

Generally speaking, the shallow features such as edge will gradually lost with the increase of network depth. Some researches [34, 35], including the FFA-Net[16], combine the shallow features and deep features through the operation of multi-skip connection and concatenation, so as to form output. For solving the problem of lack of contact between the downsampling layer and the upsampling layer in our network, the adaptive mixup operation [36] is utilized to link the information between the two layers to maintain information flow adaptively and better restore the image. In this network, the final output of this operation can be expressed as:

$$f_{out} = \mathrm{Mix}(f_{down}, f_{up}) = \sigma(\theta) * f_{down} + (1 - \sigma(\theta)) * f_{up} \quad (1)$$

where f_{out} denotes the final output, f_{up} and f_{down} represent feature maps from upsampling and downsampling, respectively. $\sigma(\theta)$ refers to the learnable factor to combine the inputs from the two layers, which is obtained by sigmoid function σ with parameter θ.

3.2 The Self-Adaptation Feature Attention Module

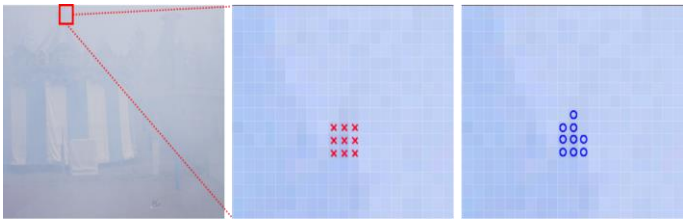

Fig. 2 Principle of deformable convolution

In early researches [13, 15, 16, 28, 29, 37], the fixed network convolution kernel as shown in the middle of Fig. 2 is usually adopted, which leads to the limitation of the receptive field and the inability to explore the structured clues in the feature space effectively. To solve this problem, it is also crucial to adjust the shape of the receptive field. As shown in the right of Fig. 2, due to the flexibility of the deformable convolution kernel, it is capable of obtaining more significant structural information adaptively. The spatial invariant

convolution kernel will lead to the destruction of image texture, which has been confirmed in previous studies [38]. As the core component of our SAFA module, we introduce two deformable convolution layers with deformable 2D kernels into the original pixel attention module [39], as shown in the Fig. 3, which implements an expansion of the receptive field adaptively and promotes the transformation ability of the model when the network focus on the calculation of thick fog pixels and high-frequency image regions. The capability to sample the unconstrained deformation of the grid also enables the network to adaptively integrate more spatial structure information and achieve a better dehazing effect. In addition, it is worth noting that in each SAFA module, the deformable convolution in deep deployment is better than in shallow deployment. On the other hand, through experiments, we notice that for the pixel attention module in our method, it is more efficient to replace the original two convolution layers and ReLU function with one 1 × 1 convolution, and the network is simplified to a certain extent. Therefore, the process can be defined as:

$$PA = F_{in} \otimes \sigma(DfConv(DfConv(Conv(F_{in})))) \quad (2)$$

where $DfConv$ refers to the deformable convolution operation and the σ represents the sigmoid function. The rest parts of the SAFA module basically keep the network structure of the FA module [16].

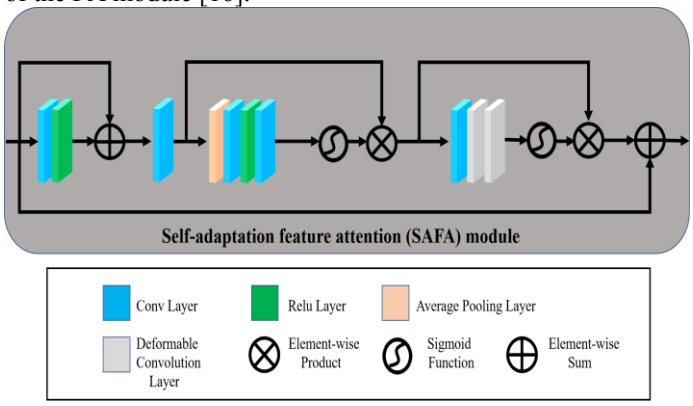

Fig. 3 The basic architecture of the SAFA module

3.3 Multi-step Fusion Module

Low-level features(i.e., step 1 and step 2 in our network), such as local information like edges, can often be easily extracted. With the increase of receptive field, the semantics of the global scope can be obtained by high-level features. In many cases such as target detection [7], image restoration [8], and other CNN-based tasks, the application of different levels of feature extraction and fusion methods has demonstrated significant effects. However, for the image dehazing field, the existing feature fusion methods do not fully consider feature fusion from disparate levels. In general, using only high-level features results in images lacking local details. By contrast, applying only low-level features, preserves the details though, but does not recover the semantics at the global level. In order to make full use of the advantages of this method, we implement a multi-step feature fusion module for the dehazing network. As shown in the Fig. 1, there are four fusion modules from left to right. The first module fuses the features from step 1 and Step 2, and the resulting fusion feature 1, as a low-level feature, continues to fuse with the high-level features of Step 3 in the second fusion module to produce fusion feature 2. Similarly, fusion feature 3 generated after the fusion of the feature in step 4 and fusion feature 2 in the third fusion module is also used for the final feature fusion module after step 5.

Basically, for each feature fusion module, there is a low-level feature and a high-level feature, respectively. Each of them passes through a convolution layer before being fused, and then fusion operation is completed by an element-wise product. Two different features are combined in the fused features, which will go through a convolution layer and a ReLU layer, and then be processed by the next fusion module in sequence. The high-level and low-level features in each fusion module are denoted as F_h and F_l, respectively, and δ means the ReLU function. F_{out} represents the final output of the whole module. Finally, this process can be expressed as:

$$F_{out} = \delta(Conv(Conv(F_h) \otimes Conv(F_l))) \quad (3)$$

3.4 Loss Function

There are three loss functions utilized to measure the deviation between the haze-free images and related clear ones to optimize the model. They are mean square error (MSE), Smooth L1 loss, and perceptual loss, and each of them plays a different role in the total loss function, respectively. The MSE is usually applied to precisely obtain some information of the low frequencies in the images, which are necessary for recovering clear images. This term is formulated as:

$$L_{mse} = \frac{1}{C\,H\,W}\|J_d - J\|^2 \quad (4)$$

where J_d means the image after dehazing, and J denotes that related ground truth image, and C, H, W represent the number of RGB channels, height, and width of the image severally.
In addition to strengthen low-frequency correctness, Smooth L1 loss is also insensitive to outliers and can be used to mitigate situations such as gradient explosion. Accordingly, this term can be expressed as:

$$L_1 = \frac{1}{C\,H\,W}\psi(J_d - J)$$
$$where\ \psi(\varepsilon) = \begin{cases} 0.5e^2, & if\ |e| < 1, \\ |e| - 0.5, & otherwise \end{cases} \quad (5)$$

To enhance network recovery of images with low-to-high semantic fidelity and better visual effect criterion, we utilize perceptual loss, which leverages multi-scale features obtained from the pre-trained neural network, to quantify the feature discrepancy between J_d and J.

$$L_{per} = \sum_{k\in\{4,9,16\}} \frac{1}{C_k H_k W_k}\|\phi_k(J_d) - \phi_k(J)\|^2 \quad (6)$$

where ϕ_k denotes the k-th feature extractor related to the three stages of the pre-trained VGG16 network associated with

restored image J_d and its clear image J, the C_k, H_k, and W_k refer to the value of three parameters mentioned above of the feature maps corresponding to the k-th layer of the VGG16 severally.

At last, the overall loss function is defined by integrating three terms above as shown below:

$$L_{total} = L_{mse} + L_1 + \alpha L_{per} \quad (7)$$

where α is a weight parameter applied to control the balance of the three terms.

4. Experiment

4.1 Datasets and Evaluation Metrics

Due to the serious difficulty of collecting the authentic hazy images and their references without haze, we first choose the outdoor training set (OTS) and synthetic objective testing set (SOTS) from the RESIDE-standard dataset [40] for training and testing goals severally. RESIDE contains plentiful synthetic hazy indoor and outdoor images as well as their related clear images (ground truth). It has been used for a long time by researchers as a benchmark in the field of image dehazing based on CNN. To further evaluate the integrated dehazing capability of our model in the scene of the real world, we also adopt Dense-Haze dataset [41] and NH-HAZE dataset [42], which included 55 pairs of images from various outdoor scenes of homogeneous, uneven fog as well as their ground truth, respectively.

The peak signal-to-noise ratio (PSNR) and the structural similarity index (SSIM)[43] are applied as the metrics for the assessment section, which also are the most common criteria for comparing the image quality in dehazing tasks.

4.2 Training Details

We implement proposed model utilizing PyTorch [44] framework, and all training and tests are performed on the platform with an Nvidia GeForce RTX 2080Ti GPU, 128GB of RAM, and the Intel XEON E5-2698V4 CPU. For the model training section, the configuration is shown below: The Adam optimizer [45] with exponential decay rates of 0.9 and 0.99 respectively is applied, and the 8 hazy-image patches with the size 240 × 240 are extracted as input of our network. Moreover, the initial learning rate is set as 0.0002 and decayed based on the cosine annealing strategy. The network is totally trained for about 130 epochs on the OTS subset. Besides, 90, 180, 270 degrees random horizontal and vertical rotations are applied as extra augmentation methods of training data.

4.3 Evaluation on the Benchmark dataset

Firstly, the proposed network is tested according to the visual effect and quantitative accuracy with the synthetic dataset SOTS [40]. We compare our way with SOTA methods in the visual effect of the recovered image, as shown in the Fig. 4. It can be clearly seen that although the haze is successfully removed in DCP [4] and MSBDN [13], it also caused the problem of color distortion. The image utilized GridDehazeNet[15] is recovered though, the brightness became too high. In comparison, AOD-Net [14] and FFA-Net [16] obtained relatively good output results, but there is still a small amount of haze in the local region of images.

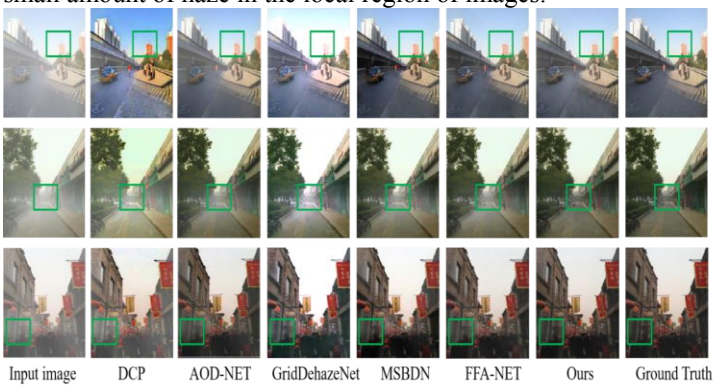

Fig. 4 Visual results comparison of images on SOTS dataset [40].

Besides, some experimental comparisons are conducted with SOTA techniques including DCP [4](the prior-based method), AOD-Net [14], GridDeHazeNet [15], MSBDN [13] and FFA-Net [16]. The quantitative results on the testing set are summarized below in Table 1:

It can be observed from the comparison with the previous FFA-Net [16] of Table 1, our SAFA network realized 0.24dB PSNR performance increase with significantly reduced parameters. Although SSIM decreased slightly by 0.0063, our method generated images more naturally.

Table 1 Quantitative comparisons of results with SOTA techniques on SOTS[40] dataset.

Methods	PSNR	SSIM
DCP[4]	15.09	0.7649
AOD-Net[14]	19.82	0.8178
GridDehazNet[15]	32.16	0.9836
MSBDN[13]	33.79	0.9840
FFA-Net[16]	36.39	**0.9886**
SAFA-Net	**36.63**	0.9823

4.4 Evaluation on real-world datasets

We also compared the test results of Dense-Haze [41] and NH-HAZE [42] datasets with other SOTA approaches. Both datasets are under much denser and more difficult-to-remove fog than the RESIDE dataset [40], especially the former.

It can be observed from the Fig. 5 and Fig. 6 that whether it is DCP [4], AOD-Net [14], GridDeHazeNet [15], and MSBDN [13], all of them have limited visual effects on removing dense haze. It is obvious that most fog still remains on the processed images, while there are particular problems such as texture loss and color degradation in FFA-Net [16]. By comparing the visual effects, our method can apparently recover more explicit images than other methods while retaining the original details and structure.

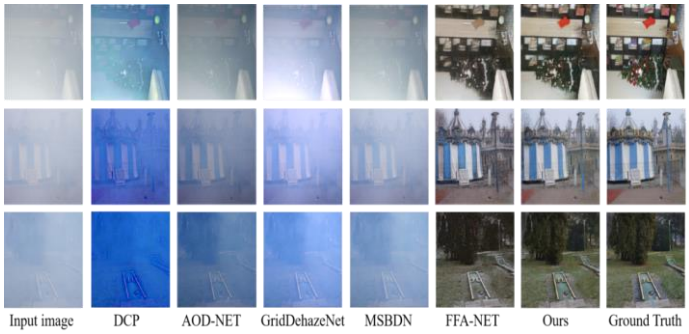

Fig. 5 Visual results comparison of images on Dense-Haze dataset [41].

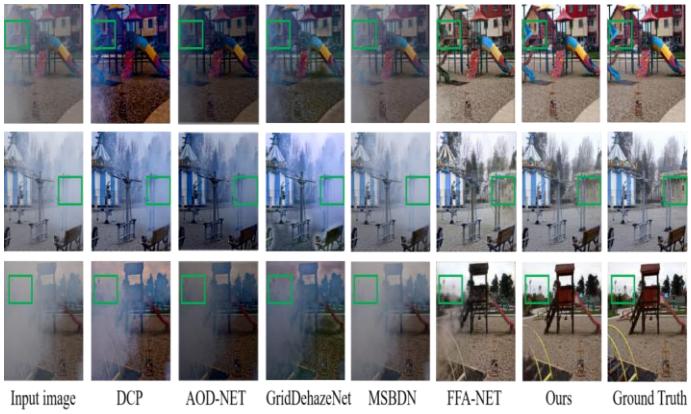

Fig. 6 Visual results comparison of images on NH-HAZE dataset [42].

As shown in Table 2 and Table 3, it is shown that the performance of our SAFA network on the Dense-Haze dataset [41] is far superior to all SOTA techniques depend on 17.34dB PSNR and 0.5817 SSIM. For the NH-HAZE dataset [42], we also obtained the highest PSNR and SSIM, which are 21.81 dB and 0.7253 severally.

Besides, it is not difficult to discover from the last row of Table 3 that the proposed network achieves relatively excellent results with fewer parameters for the trade-off between calculation parameters and image recovery metrics, which also reduces the cost of computation effectively.

Table 2 Quantitative comparisons of results with SOTA techniques on Dense-Haze[41] dataset.

Methods	PSNR	SSIM
DCP[4]	10.06	0.3856
AOD-Net[14]	13.14	0.4144
GridDehazNet[15]	14.31	0.4081
MSBDN[13]	15.47	0.4858
FFA-Net[16]	14.39	0.4725
SAFA-Net	**17.34**	**0.5817**

Table 3 Quantitative comparisons of results with SOTA techniques on NH-HAZE[42] dataset.

Methods	PSNR	SSIM	*Parameters
DCP[4]	10.57	0.5196	-
AOD-Net[14]	15.40	0.5693	0.002M
GridDehazNet[15]	13.80	0.5370	0.96M
MSBDN[13]	19.23	0.7056	31.35M
FFA-Net[16]	19.87	0.6915	4.68M
SAFA-Net	**21.81**	**0.7253**	2.37M

4.5 Evaluation on real-world hazy photographs

In order to measure the dehazing effect of our network on real foggy photographs and make it more convincing. Plentiful real hazy photographs obtained from the RTTS [40] dataset and a part of foggy day images collected by the author in the campus of the University of Kent were tested and compared. The visual results are shown in the figure. It can be seen that although the previous methods of AOD-Net [14], GridDeHazeNet [15], MSBDN[13], and FFA-Net [16] perform very well on artificial datasets, the effect of fog removal for such real images is not satisfactory enough. Besides, relatively effective DCP [4] suffers from color distortion and tends to over-enhance the images. In some cases, the results of AOD-Net [14] appeared floating shadows, and the brightness of the pictures after MSBDN [13] processing became lower. In general, our model achieves the superior visual effect in image detail recovery while maintaining the overall brightness, and the clear and haze-free images are reconstructed with good perceptual quality as well.

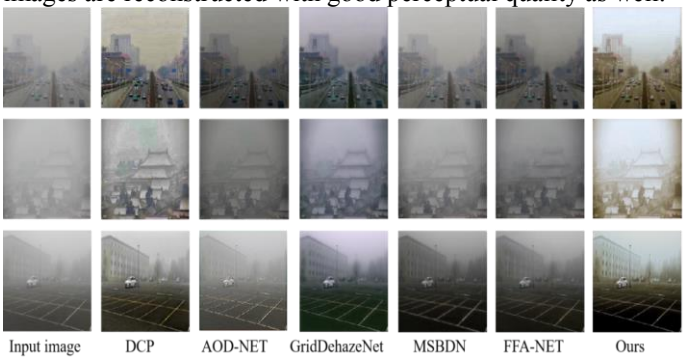

Fig. 7 Visual results comparison of real photographs with haze.

5. Experiment

In this paper, we propose an end-to-end dehazing network that consists of self-adaptation feature attention (SAFA) module and a multi-step fusion module. The former module is capable of extracting the detailed features of the hazy image adaptively, which enlarges the range of dealing with complicated information to increase the transformation ability of the network significantly. The latter one uses the features from multiple steps to obtain the benefit from their combination. We also carried out exhaustive experiments on disparate datasets, and by comparing with results of some SOTA algorithms, it is proved that the apparent advantages of this network structure in the aspect of image detail recovering with effect. In addition, as we reduce the depth and complexity in network design, the more compact network significantly reduces the computational power consumption and time required for operation. Through further research in the future, it is expected to realize real-time dehazing and the application

based on this network structure in other image restoration tasks.

References

[1] E. J. McCartney, "Optics of the atmosphere: Scattering by molecules and particles," New York, John Wiley and Sons, Inc, 1976. 421, 1976.
[2] S. G. Narasimhan and S. K. Nayar, "Vision and the atmosphere," International Journal of Computer Vision, vol. 48, no. 3, pp. 233–254, 2002.
[3] R. Fattal, "Dehazing using color-lines," ACM Transactions on Graphics, vol. 34, no. 1, pp. 1–14, 2014.
[4] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 2341–2353, 2010.
[5] R. T. Tan, "Visibility in bad weather from a single image," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–8, 2008.
[6] Q. Zhu, J. Mai, and L. Shao, "A fast single image haze removal algorithm using color attenuation prior," IEEE transactions on image processing, vol. 24, no. 11, pp. 3522–3533, 2015.
[7] B. Xu and Z. Chen, "Multi-level fusion based 3d object detection from monocular images," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2345–2353, 2018.
[8] J. Li, F. Fang, J. Li, K. Mei, and G. Zhang, "Mdcn: Multiscale dense cross network for image super-resolution," IEEE Transactions on Circuits and Systems for Video Technology, 2020.
[9] B. Cai, X. Xu, K. Jia, C. Qing, and D. Tao, "Dehazenet: An end-to-end system for single image haze removal," IEEE Transactions on Image Processing, vol. 25, no. 11, pp. 5187–5198, 2016.
[10] S. Zhao, L. Zhang, Y. Shen and Y. Zhou, "RefineDNet: A Weakly Supervised Refinement Framework for Single Image Dehazing," in IEEE Transactions on Image Processing, vol. 30, pp. 3391-3404, 2021.
[11] L. Huang, J. Yin, B. Chen and S. Ye, "Towards Unsupervised Single Image Dehazing With Deep Learning," 2019 IEEE International Conference on Image Processing (ICIP), pp. 2741-2745, 2019.
[12] Y. Zhang and Y. Dong, "Single Image Dehazing via Reinforcement Learning," 2020 IEEE International Conference on Information Technology,Big Data and Artificial Intelligence (ICIBA), pp. 123-126, 2020.
[13] H. Dong, J. Pan, L. Xiang, Z. Hu, X. Zhang, F. Wang, and M.-H. Yang, "Multi-scale boosted dehazing network with dense feature fusion," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 2157–2167, 2020.
[14] B. Li, X. Peng, Z. Wang, J. Xu, and D. Feng, "Aod-net: All-in-one dehazing network," in Proceedings of the IEEE International Conference on Computer Vision, pp. 4770–4778, 2017.
[15] X. Liu, Y. Ma, Z. Shi, and J. Chen, "Griddehazenet: Attention-based multi-scale network for image dehazing," In ICCV, 2019.
[16] X. Qin, Z. Wang, Y. Bai, X. Xie, and H. Jia, "Ffa-net: Feature fusion attention network for single image dehazing." in AAAI, pp. 11 908– 11 915, 2020.
[17] T. Guo, X. Li, V. Cherukuri, and V. Monga, "Dense scene information estimation network for dehazing," In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2019.
[18] X. Zhang, J. Wang, T. Wang and R. Jiang, "Hierarchical Feature Fusion with Mixed Convolution Attention for Single Image Dehazing," in IEEE Transactions on Circuits and Systems for Video Technology, 2021.
[19] G. Meng, Y. Wang, J. Duan, S. Xiang, and C. Pan, "Effificient image dehazing with boundary constraint and contextual regularization," in Proceedings of the IEEE International Conference on Computer Vision, pp. 617–624, 2013.
[20] C. O. Ancuti, C. Ancuti, C. Hermans, and P. Bekaert, "A fast semi-inverse approach to detect and remove the haze from a single image," in Proceedings of the Asian Conference on Computer Vision, pp. 501–514, 2010.
[21] S. G. Narasimhan and S. K. Nayar, "Interactive (de) weathering of an image using physical models," in Proceedings of the IEEE Workshop on Color and Photometric Methods in Computer Vision, vol. 6, no. 6.4, p. 1, 2003.
[22] Q. Zhu, J. Mai, and L. Shao, "A fast single image haze removal algorithm using color attenuation prior," IEEE transactions on image processing, vol. 24, no. 11, pp. 3522–3533, 2015.
[23] R. T. Tan, "Visibility in bad weather from a single image," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–8, 2008.
[24] R. Fattal, "Dehazing using color-lines," ACM Transactions on Graphics, vol. 34, no. 1, pp. 1–14, 2014.
[25] H. Zhang and V. M Patel, "Densely connected pyramid dehazing network," In CVPR, pages 3194–3203, 2018.
[26] W. Ren, S. Liu, H. Zhang, J. Pan, X. Cao, and M. Yang, "Single image dehazing via multi-scale convolutional neural networks," In ECCV, pages 154–169, 2016.
[27] W. Ren, L. Ma, J. Zhang, J. Pan, X. Cao, W. Liu, and M.-H. Yang, "Gated fusion network for single image dehazing," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3253–3261, 2018.
[28] Y. Qu, Y. Chen, J. Huang, and Y. Xie, "Enhanced pix2pix dehazing network," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 8160–8168, 2019.
[29] M. Hong, Y. Xie, C. Li, and Y. Qu, "Distilling image dehazing with heterogeneous task imitation," In CVPR, pages 3462–3471, 2020.
[30] Q. Wang, Z. Teng, J. Xing, J. Gao, W. Hu, and S. Maybank, "Learning attentions: residual attentional siamese network for high performance online visual tracking," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4854–4863, 2018.

[31] Y. Peng, X. He, and J. Zhao, "Object-part attention model for fine-grained image classification," IEEE Transactions on Image Processing, vol. 27, no. 3, pp. 1487–1500, 2017.
[32] X. Zhang, R. Jiang, T. Wang, P. Huang, and L. Zhao, "Attentionbased interpolation network for video deblurring," Neurocomputing, 2020.
[33] Y. Hu, J. Li, Y. Huang, and X. Gao, "Channel-wise and spatial feature modulation network for single image super-resolution," IEEE Transactions on Circuits and Systems for Video Technology, 2019.
[34] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," In MICCAI, pages 234–241, 2015.
[35] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," In CVPR, pages 770–778, 2016.
[36] H. Zhang, M. Cisse, Y. N Dauphin, and D. Lopez-Paz, "mixup: Beyond empirical risk minimization," arXiv preprint arXiv:1710.09412, 2017.
[37] Y. Shao, L. Li, W. Ren, C. Gao, and N. Sang, "Domain adaptation for image dehazing," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 2808–2817, 2020.
[38]X. Xu, M. Li, and W. Sun, "Learning deformable kernels for image and video denoising," arXiv preprint arXiv:1904.06903, 2019.
[39] J. Dai, H. Qi, Y. Xiong, Y. Li, G. Zhang, H. Hu, and Y. Wei, "Deformable convolutional networks," In ICCV, pages 764–773, 2017.
[40] 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, vol. 28, no. 1, pp. 492–505, 2018.
[41] C.O. Ancuti, C. Ancuti, M. Sbert, and R. Timofte, "Dense haze: A benchmark for image dehazing with dense-haze and haze-free images," In ICIP, 2019.
[42] C.O. Ancuti, C.Ancuti, and R.Timofte, "NH-HAZE: An image dehazing benchmark with nonhomogeneous hazy and haze-free images," In The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, 2020.
[43]Z. Wang, A. C Bovik, H. R Sheikh, and E.P Simoncelli, "Image quality assessment: from error visibility to structural similarity," IEEE transactions on image processing, 13(4):600–612, 2004.
[44] A. Paszke, S. Gross, S. Chintala, and G. Chanan, "Pytorch," Computer Software. Vers. 0.3, vol. 1, 2017.
[45] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," arXiv preprint arXiv:1412.6980, 2014.

Authors' Contributions : All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Jiawei Zhang and Xiaochen Liu. The first draft of the manuscript was written by Jiawei Zhang and all authors commented on previous versions of the manuscript. Donghua Zhao, Chenguang Wang, Chong Shen, Jun Tang and Jun Liu gave the guidance and help during the whole research. All authors read and approved the final manuscript.

Funding : This work was supported, in part, by the National Natural Science Foundation of China (Grant Nos. 61973281 and 51705477), the Innovative Research Group Project of the National Natural Science Foundation of China (Grant No. 51821003), the Pre-research Field Foundation (Grant No. 6140518010201), the Scientific and Technology Innovation Programs of Higher Education Institutions in Shanxi (Grant No. 201802084), the Aeronautical Science Foundation of China (Grant No. 2018ZCU0002), the Weapons and Equipment Joint Foundation (Grant No. 6141B021305), the Program for the Top Young Academic Leaders of Higher Learning Institutions of Shanxi, the Young Academic Leaders Foundation in North University of China, the Science Foundation of North University of China (Grant No. XJJ201822), and the Fund for Shanxi "1331 Project" Key Subjects Construction.