

Feature Attention Parallel Aggregation Network for Single Image Haze Removal

CUILI LI, (Member, IEEE), YUFENG HE, (Member, IEEE), AND XU LI
College of Information Engineering, Tarim University, Aral City 843300, China

Corresponding author: Yufeng He (heyufeng@hnu.edu.cn)

This work was supported in part by the Natural Science Project of the Presidential Foundation of Tarim University under Grant TDZKSS202140 and Grant TDZKSS202132; and in part by the Joint Funds of China Agricultural University and Tarim University under Grant ZNLH202103.

ABSTRACT Images captured in hazy weather often suffer from color distortion and texture blur due to turbid media suspended in the atmosphere. In this paper, we propose a Feature Attention Parallel Aggregation Network (FAPANet) to restore a clear image directly from the corresponding hazy input. It adopts the encoder-decoder structure while incorporating residual learning and attention mechanism. FAPANet consists of two key modules: a novel feature attention aggregation module (FAAM) and an adaptive feature fusion module (AFFM). FAAM recalibrates features by integrating channel attention and pixel attention in parallel to stimulate useful information and suppress redundant features. The shallow and deep layers of neural networks tend to characterize the low-level and high-level semantic features of images, respectively, so we introduce AFFM to fuse these two features adaptively. Meanwhile, a joint loss function, composed of L1 loss, perceptual loss, and structural similarity (SSIM) loss, is employed in the training stage for better results with more vivid colors and richer details. Comprehensive experiments on both synthetic and real-world images demonstrate the impressive performance of the proposed approach.

INDEX TERMS Attention mechanism, deep learning, feature fusion, image restoration, single image dehazing

I. INTRODUCTION

PHOTOS taken in hazy weather are usually degraded in contrast and color fidelity, as shown in Fig. 1. There are two main reasons for this image quality degradation phenomenon. For one thing, the scene radiance is attenuated due to the absorption and scattering of small floating particles in the air. Another is that the ultimate irradiance captured by the camera is blended with the ambient light, which is reflected by atmospheric particles. The presence of the haze may impair the performance of subsequent computer vision algorithms and industrial applications such as object detection and tracking in traffic monitoring systems. Therefore, it is important to recover the clear scene from hazy images.

According to the atmospheric scattering model [1, 2], the generation of a hazy image can be mathematically formulated as follows:

\[ I(x) = J(x)t(x) + A(1 - t(x)) \]  

(1)

where \( I(x) \) and \( J(x) \) are the hazy images and the clear counterparts to be recovered, respectively. The variable \( A \) represents the global atmospheric light, which describes the intensity of the ambient light, \( x \) represents the pixel position, and \( t(x) \) is the medium transmission, which means the rate of the scene radiance that is not scattered or absorbed. Furthermore, the transmission matrix \( t(x) \) is a distance-dependent parameter that can be calculated using the formula \( t(x) = e^{-\beta d(x)} \), where \( \beta \) is the atmospheric scattering coefficient, and \( d(x) \) is the scene depth. Single image haze removal aims to recover the clear image \( J(x) \) from a given hazy image \( I(x) \). Obviously, it is an ill-posed problem due to the unknown atmospheric light \( A \) and transmission \( t(x) \).

Some researchers tried to supplement auxiliary information to reconstruct the original clean image in early attempts at hazy image restoration. For example, photos of the same scene taken with different degrees of polarization [3, 4] or captured under diverse weather conditions [5] were used to remove the haze. In addition, depth information was also utilized for dehazing [6, 7]. These multiple-image-based methods can effectively solve the degradation problem of hazy images. However, extra information is not always available in practical applications. Therefore, haze removal using a single
image deserves more attention. Image enhancement methods [8, 9, 10, 11, 12, 13] usually do not consider the atmospheric scattering model but rather improve the visual quality of degraded images by modifying the image’s contrast, color, and brightness. Therefore, these enhancement methods cannot adequately restore clear pictures because they do not consider the fundamental causes of image degradation. Neither directly supplementing the additional information nor ignoring the physical model, prior-based image restoration methods [1, 14, 15, 16, 17] add constraints to the physical model through priors derived from rational hypothesis, observation, and statistics, thereby restoring clear and haze-free images. Unfortunately, these methods may not work in some particular scenarios where the priors do not hold.

With the great success of deep learning in various visual tasks, many dehazing methods based on convolutional neural networks (CNNs) [18, 19, 20, 21, 22, 23] have emerged in recent years. These methods exploit massive amounts of training data and the powerful learning capability of CNN to capture the intricate patterns of hazy images. Many previous studies have evaluated the unknown transmission maps and the global atmospheric light using CNN, ultimately obtaining the original haze-free images by referring to the existing physical model. However, the performance of this kind of two-stage approach [18, 19] rapidly decreases when the intermediate variables cannot be assessed accurately. For this reason, many end-to-end CNN-based methods [20, 21, 22, 23] have been proposed to remove haze, which directly learn the potential sophisticated mapping between hazy images and corresponding clear counterparts.

In this paper, we propose an effective neural network for single image haze removal that is inspired by the following three main ideas: Firstly, some recent dehazing methods [21, 23, 24] based on encoder-decoder structure have achieved good results, and UNet [25], as a classic encoder-decoder structure network, has an excellent performance in some other low-level image processing tasks such as image segmentation. However, in traditional UNet-like neural networks, multiple downsampling and upsampling operations will lose some low-level information of the input images and cause feature misalignment, which results in the performance degradation of the whole network. Therefore, the network proposed in this paper adopts a simplified UNet structure. Only two strided convolution layers are used in the encoder for downsampling twice. In the decoder, two strided deconvolution layers transform the feature resolution to be consistent with the input. Secondly, the attention mechanism is widely used in various visual tasks as a powerful feature recalibration tool that can effectively capture non-local information in images. This paper proposes a novel and practical feature attention aggregation module that connects channel attention and spatial attention in parallel. A skip connection between the module’s input and output is used to aid gradient flow. Finally, with the number of intermediate layers in the deep neural network increasing, the receptive field of the feature map gradually expands. But the hierarchical information between different spatial scales, which has been proven helpful for image reconstruction in previous research [21, 24, 26, 27], is not fully utilized. Therefore, we introduce an adaptive feature fusion module, which learns the weights of the feature maps in shallow and deep layers and adds them together by weight, finally achieving the delicate aggregation of features with different receptive fields.

The contributions of this research can be summarized as follows:

- We propose an end-to-end convolutional neural network, named FAPANet, to restore the clean images from the hazy inputs. And we introduce a joint loss function that consists of L1 loss, perceptual loss, and SSIM loss to train the proposed FAPANet, which dramatically enhances the details of the restored images.
- To enhance the features of the network, the feature attention aggregation module (FAAM), an effective hybrid attention scheme, is proposed in this paper, which recalibrates the feature maps by parallel channel attention and spatial attention. In this module, we use the maximum, minimum, and mean values of each channel of the features simultaneously to calculate the feature channel attention. Also, the spatial attention part of the module uses a simple, parameter-free strategy.
- A simple and effective feature fusion strategy is presented to assemble the features at different spatial scales. It makes full use of the shallow and deep features, which substantially works in image reconstruction.
- Extensive experiments on synthetic and natural images have been conducted. The quantitative and qualitative results compared with various state-of-the-art methods demonstrate the effectiveness of the proposed method. In addition, the adequate ablation study reveals the importance of each sub-module in the network.
II. RELATED WORK
For many years, the problem of image dehazing has attracted much attention, and many excellent methods have been proposed. In this section, we provide a brief review of existing techniques. As mentioned previously, we classify haze removal methods into four categories: multi-image-based methods, image enhancement methods, prior-based methods, and deep-learning-based methods, the latter three of which we individually introduce the representative studies of each of these four methods.

A. MULTI-IMAGE-BASED METHODS
Early methods added constraints to the ill-posed equation via supplementary information for the existing atmospheric scattering model, thereby restoring haze-free images. Kopf et al. [6] and Narasimhan et al. [7] removed haze with the assistance of the 3D model of the scene or depth information provided interactively by users. Nayar and Narasimhan [5] restored the visibility of hazy images using multiple images of the same scene captured under diverse weather conditions. In [28], Schaul et al. presented a near-infrared-based method for image dehazing. In addition, some polarization-based techniques have been proposed in [3, 4]. However, these methods are often difficult to deploy in real-world applications since the geometrical information and multiple images are not always available.

B. IMAGE ENHANCEMENT METHODS
Faced with the difficulty of reliably evaluating the transmission map and ambient light through a single hazy input, some researchers have attempted to bypass the physical model and use image enhancement methods to improve the visual quality of images. With some minor adjustments to the original histogram equalization algorithm, Stark et al. [8] and Joung-Youn Kim et al. [9] proposed adaptive histogram equalization and partially overlapped sub-block histogram-equalization to improve image contrast. Furthermore, some fusion-based methods [10, 11] and retinex-based methods [12, 13] have also been presented to enhance the visibility of hazy images.

C. PRIOR-BASED METHODS
Prior-based methods estimate the albedo of the scene and the atmospheric light using various hypotheses, statistical rules, and observations. Tan et al. [14] found that photos taken in hazy weather always have low color contrast, therefore, they maximized the contrast of image patches to enhance image visibility. He et al. [1] estimated the transmission map using a dark channel prior, which was derived from the intriguing observation that the patches of a haze-free image often have one color channel whose intensity is close to 0. Fattal [15] observed that the local pixels of image patches typically exhibit a one-dimensional distribution and introduced a color-line method based on this discovery. Berman et al. [16] proposed an algorithm based on a non-local prior, which presupposes that the colors of a haze-free image are well approximated by a few hundred distinct hues. Zhu et al. [17] offered a dehazing method based on color attenuation prior, which assumes that the difference between brightness and saturation is positively correlated with haze concentration. In general, these prior-based methods rely heavily on the validity of handcrafted features. For this reason, they would fail in some real-world scenes where the assumption is invalid.

D. DEEP-LEARNING-BASED METHODS
Over the past several years, many learning-based approaches have been introduced to cope with the degradation problem of hazy images. Cai et al. [18] constructed DehazeNet to predict transmission maps via a novel nonlinear activation function known as the Bilateral Rectified Linear Unit (BReLU). Similarly, Ren et al. [19] employed a multiscale convolutional neural network (MSCNN) to estimate and refine transmissions. Li et al. [20] proposed the AOD-Net by reformulating the physical model and incorporating transmissions and atmospheric light into a new K-module. The Densely Connected Pyramid Dehazing Network (DCPDN) [21] has been proposed to jointly learn transmission maps and atmospheric light. All of these CNN-based methods have achieved promising improvements for single image dehazing, but compared with traditional methods, they scarcely consider haze-related priors to constrain the learning space. To that end, Yang et al. [22] presented a CNN-based dehazing approach known as Proximal Dehaze-Net that learns both dark channel and transmission priors via their introduced energy functions. Different from these methods, Ren et al. [23] first combined fusion strategy with deep neural networks to dehaze, exploiting an encoder-decoder network to nonlinearly fuse three inputs derived from the same hazy image by applying white balance (WB), contrast enhancing (CE), and gamma correction (GC). In addition, given the great success of generative adversarial networks (GANs) in various computer vision tasks, many GAN-based dehazing methods [29, 30] have also been proposed. Recently, Kuanar et al. [31] proposed a deep learning-based model, DeGlow–DeHaze, to solve the problem of dehazing for images taken at night.

The method proposed in this study uses a single image to remove haze without auxiliary geometrical information. Furthermore, it is an end-to-end haze removal network that can potentially learn the complex mapping between haze and haze-free images without strictly referring to the ideal atmospheric scattering model, thereby avoiding the error magnification caused by separately estimating transmission maps and atmospheric ambient light.

III. PROPOSED METHOD
In this section, we introduce FAPANet in detail. First, we describe the architecture of our network and the motivation behind it. Second, we present two novel sub-modules, i.e., the Feature Attention Aggregation Module (FFAM), which integrates the features enhanced by channel and spatial attention in parallel for feature recalibration progressively, and the Adaptive Feature Fusion Module (AFFM) for the fusion...
of shallow and deep features by learnable weight. Finally, the joint loss function is introduced.

A. NETWORK ARCHITECTURE
As shown in Fig. 2, FAPANet is a classic encoder-decoder structure network, which can be roughly divided into 3 stages. Firstly, four consecutive conventional and strided convolutional layers are used as an encoder to encode the inputs, in which the strided convolutional layers are used to significantly reduce the feature resolution, thereby saving a tremendous amount of training memory, and the convolution layers gradually expand the receptive field of features to boost the representation of high-dimensional semantic features. After that, the features acquired by the encoder are progressively enhanced by 10 cascaded FAAMs. Then, two AFFMs reasonably fuse the multi-scale spatial features. A detailed description of these two essential modules will be presented in Section III-B and Section III-C. Finally, the decoder is a symmetrical structure to the encoder, which restores the resolution of the feature map to that of the original input via two strided deconvolution operations, and then decodes the feature maps to haze-free results.

In the process of feature encoding and decoding, the original UNet employs multiple upsampling and downsampling operations, which inevitably leads to the loss of some image information, especially low-level features that are crucial for image recovery. While feature mapping using the original resolution of the input image, instead of multiple samplings, requires too much memory in the training stage and causes excessive information redundancy. Therefore, FAPANet is designed as a simplified UNet structure, which performs two downsampling and upsampling operations in the encoder and decoder, respectively. But at the same time, it has feature attention aggregation modules and feature fusion modules to compensate for the information loss caused by upsampling and downsampling as much as possible, to obtain representative hazy image features.

The general convolution layer with a fixed grid kernel could lead to image texture degradation and artifacts [32]. In contrast, the deformable convolution operation has a dynamic and flexible kernel, which allows it to capture more information [24, 33]. Therefore, we introduce two deformable convolution blocks in FAPANet to strengthen the feature representative capability and get more realistic dehazing results. Following the previous work [24], the deformable convolution is cascaded after the FAAMs to achieve better performance.

B. FEATURE ATTENTION AGGREGATION MODULE
Although only two downsampling operations are used in the encoder of the proposed FAPANet to compress the resolution of the feature map to a quarter of the input resolution, it also causes low-level information loss. To compensate for this loss, we introduce the Feature Attention Aggregation Module (FAAM) and cascade 10 FAAMs to enhance the features progressively. The attention mechanism is a common feature enhancement method in computer vision, which stimulates the valuable features and suppresses the useless ones from the perspective of the channel or spatial pixel, called channel attention and spatial attention, respectively. The FAAM integrates the channel and spatial attention in parallel, which simultaneously achieves channel-wise and pixel-wise feature recalibration. It is known that the probability of gradient vanishes and gradient explosion in the training phase of a neural network increases as the network deepens. Fortunately, by adding a shortcut connection between the input and output of the network layer, the ResNet [35] explicitly specifies the network layers to learn the residual of the input feature maps, which makes the deep network easier to train. Similarly, we also deploy a shortcut connection in the FAAM.
Previous studies [21, 24, 26, 27] have shown that the fusion of features at different spatial scales is conducive to improving network performance. In this paper, we propose the Adaptive Feature Fusion Module (AFFM) to achieve a fine-grained fusion of deep and shallow features. As shown in Fig. 4, the AFFM consists of three parallel processing pipelines. First of all, two feature maps with the same resolution from the shallow and deep layers in FAPANet are concatenated along the channel dimension. And then, the merged feature map is fed into a convolutional layer to reduce its channels and learn the potential interaction between the deep and shallow features. Finally, the outputs of the sigmoid activation layer are used as coefficients to linearly combine the deep and shallow features, and this combination is pixel-by-pixel with learnable coefficients. The module can be formulated as:

\[
w = \text{Sigmoid}(F_{\text{conv}}(\text{concat}(x_{\text{deep}}, x_{\text{shallow}}))) \\
y = x_{\text{shallow}} \cdot w + x_{\text{deep}} \cdot (1 - w)
\]

(2)

Where \(x_{\text{deep}}\) and \(x_{\text{shallow}}\) represent the input deep and shallow features, \(F_{\text{conv}}\) represents an abstract function, which is used to map the implicit relationship between \(x_{\text{deep}}\) and \(x_{\text{shallow}}\). The concatenation operation and sigmoid activation function are denoted by \(\text{concat}\) and \(\text{Sigmoid}\), respectively. Therefore, \(y\) indicates the result of a linear combination of \(x_{\text{deep}}\) and \(x_{\text{shallow}}\) with the coefficient \(w\).

Unlike existing feature fusion methods [21, 26, 27] that directly merge feature maps of different scales by addition or concatenation operations, our approach achieves fine-grained feature fusion by explicitly modeling the weights of each position of the deep and shallow features, thereby obtaining effective feature information.
D. LOSS FUNCTION

To restore realistic haze-free images, we propose a joint loss function that linearly combines L1 loss, SSIM loss, and perceptual loss by weight. In the following, we present a detailed introduction to these three kinds of loss functions. For uniform expression, we specify that $I$ and $J$ represent the hazy images and the corresponding haze-free images, respectively.

**L1 Loss.** L1 loss pixel-wisely measures the mean absolute error between the recovered image and its corresponding ground truth, a technique that is widely used in image reconstruction. Following the previous studies [22, 24, 27, 40], we also adopt L1 loss to optimize our model. $\mathcal{L}_1$ is used to denote L1 loss, which can be formulated as follows:

$$\mathcal{L}_1 = \frac{1}{NHW} \sum_{i=1}^{N} |\mathcal{F}_{\text{FAPANet}}(I_i) - J_i|$$

where $N$, $H$, and $W$ represent the batch size, height, and width of the image, respectively. The variable $i$ indicates the sample in a batch training data, and $\mathcal{F}_{\text{FAPANet}}$ expresses the latent mapping function between the hazy images and the clear images which is learned by the proposed FAPANet. Accordingly, $\mathcal{F}_{\text{FAPANet}}(I_i)$ denotes the restored image corresponding to the hazy input $I_i$.

**SSIM Loss.** Since the L1 loss function has equal gradient values in most cases, even small loss values have large gradients, which is not conducive to the convergence of the model. Also, the model trained with L1 loss alone tends to produce texture collapse and halo artifacts. To avoid these problems, we combine L1 loss with SSIM loss function following the strategy in literature [40]. By assessing images from the perspective of perceived differences, the SSIM loss function helps restore visually clear and realistic images. The SSIM value of the given pixel $p$ between the hazy image and the clear image can be calculated as follows:

$$\text{SSIM}(p) = \frac{2\mu_x \mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \cdot \frac{2\sigma_{xy} + C_2}{\sigma_x^2 + \sigma_y^2 + C_2}$$

where $x$ and $y$ represent the image patches in the hazy image $I$ and the clear image $J$, respectively, while $p$ is the central pixel of the image patch. The variables $\mu_x$, $\sigma_x^2$, and $\sigma_{xy}$ represent the mean of $x$, the variance of $x$, and the covariance of $x$ and $y$, respectively. Similarly, $\mu_y$ and $\sigma_y^2$ are the mean and variance of image patch $y$, respectively. According to the above equation, the SSIM loss between the recovered image and the corresponding ground truth can be defined as

$$\mathcal{L}_{\text{SSIM}} = \frac{1}{N} \sum_{i=1}^{N} \left(1 - \frac{1}{M} \sum_{p=1}^{M} \text{SSIM}(p)\right)$$

where $N$ is the batch size and $M$ is the total number of pixels in a single training sample.

**Perceptual Loss.** It is well known that shallow layers of CNN often capture low-level image features, such as edges and contours, which are critical for image dehazing. Therefore, we introduce the perceptual loss function [21, 29], which can be expressed as

$$\mathcal{L}_{\text{VGG}} = \frac{1}{N} \sum_{i=1}^{N} \left\| \mathcal{V}_j(\mathcal{F}_{\text{FAPANet}}(I_i)) - \mathcal{V}_j(J_i) \right\|_2^2$$

Table 1. The detailed description of the FAPANet structure. "Conv2d" and "ConvT2d" denote the convolution layer and the deconvolution layer. "CA" means Channel Attention, and "SA" means Spatial Attention. In the last column, "ic", "oc", "ks", "st", and "p" indicate the input channel, output channel, kernel size, stride, and the padding value, respectively.

Here, we inherit the notations of $\mathcal{F}_{\text{FAPANet}}$, $N$, $i$, $I$ and $J$ in $\mathcal{L}_1$, and let $\mathcal{V}_j$ be the feature map calculated by the $j$-th layer of the pre-trained VGG network [41]. The "relu1-1" and "relu2-1" layers of VGG-16 were used in this study. Essentially, $\mathcal{L}_{\text{VGG}}$ estimates the mean squared error between the restored image and the degraded image in feature space, and therefore can be used as an effective complement to the original MSE loss for image reconstruction tasks.

**Total Loss Function.** The proposed FAPANet is trained by minimizing a total loss function which combines the described above L1 loss, SSIM loss and perceptual loss by weight as follows:

$$\mathcal{L}_{\text{total}} = \alpha \mathcal{L}_1 + \beta \mathcal{L}_{\text{SSIM}} + \gamma \mathcal{L}_{\text{VGG}}$$
IV. EXPERIMENTS

The implementation specifics of the proposed FAPANet are initially introduced in this section. The usefulness of the presented technique is then demonstrated using synthetic datasets and real-world images. We compared our method with several state-of-the-art algorithms, which include DCP (CVPR’2009) [1], CEP (TIP’2017) [42], HazeLine (TPAMI’2018) [16], IDE (TIP’2021) [43], AOD (ICCV’2017) [20], DehazeNet (TIP’2016) [18], GDN (ICCV’2019) [27], GCANet (WACV’2019) [44], and MSBDN (CVPR’2020) [26], in terms of quantitative and qualitative evaluation. In addition, we conducted a series of detailed ablation experiments to verify the effectiveness of its sub-modules.

A. IMPLEMENTATION DETAILS

We implement the proposed FAPANet using the Pytorch framework. All convolutional layers in FAPANet are followed by the ReLU activation function, except for the last layer, which adopts the Tanh activation function. Similarly, the filter size is set as 3×3 in all convolutional layers, except for the first and last convolutional layers, which use 7×7 convolutional kernels. The detailed network structure is described in Table 1. The FAPANet was trained on the RESIDE dataset [45], and all training samples were cropped to 256x256 randomly. During the training stage, we utilized the Adam algorithm as the optimizer, and set the initial learning rate to 0.001, and set the batch size to 16. The network was trained 120 epochs with a learning rate decay of 0.1 for every 30 epochs. Following the practice of previous studies [21, 29] and adjusting in our experiments, we set α=5, β=1, and γ=1 in the total loss function. The training and testing was performed on a same personal computer with the acceleration by a RTX 3090 GPU.

B. HAZE REMOVAL ON SYNTHETIC IMAGES

To fairly evaluate the dehazing performance of the proposed FAPANet, we compared the dehazing performance of the FAPANet with nine state-of-the-art methods on synthetic datasets of both indoor and outdoor scenes. We adopt the Synthetic Objective Testing Set (SOTS) subset of the RESIDE dataset as the test set, which contains 500 synthetic indoor hazy images and 500 synthetic outdoor hazy images, called SOTS-indoor and SOTS-outdoor, respectively. All the test images do not overlap with samples from either the indoor or outdoor training set of the RESIDE dataset.

Due to inaccuracies in atmospheric light and transmission map evaluations, some color distortions appear in the results of the prior-based DCP and CEP algorithms, as shown in Fig. 5(a). The Hazeline and IDE methods, on the other hand, can only remove local haze, leaving some fog behind. DehazeNet and AOD are unable to thoroughly remove the haze, resulting in unclean results. GDN, MSBDN, and PRENet produce better results in terms of overall quality, but we notice that GDN and MSBDN cause minor color distortion in the first image’s floor area and the third image’s ceiling portion, particularly around the mirror, while GCANet has some haze remaining in the leftmost part of the third image. In contrast, our method achieves a better dehazing performance for the indoor scene.

Fig. 5(b) shows the dehazing results of various dehazing methods for outdoor images. DCP and CEP cause some obvious color distortion in the third test image, and slight artifacts appear in the sky of the first and second test images. In addition, significant local contrast reduction is observed in all the results of the CEP and DCP algorithms. The IDE method makes the results brighter by exposure processing, but it does not remove the haze well, as shown in the first test image. Similar to the results in Fig. 5(a), HazeLine, DehazeNet, and AOD can not completely eliminate haze either. In general, GCANet, GDN, MSBDN, and FAPANet have better dehazing effects in outdoor scene images.

Furthermore, we quantitatively compare our proposed FAPANet with the state-of-the-art methods using two authoritative full-reference metrics, the peak signal to noise ratio (PSNR) and the structural similarity index (SSIM). The PSNR is the most commonly used metric to evaluate the difference between an image and its reference image in terms of pixel values. Generally, a higher PSNR indicates that the reconstructed image is of higher quality. The SSIM index is widely used to predict the perceived quality of a given image by measuring the similarity between an image and its distortion-free reference image in terms of structure and texture. The higher the SSIM value, the more similar the restored image is to the reference clear image.

As shown in Table 2, our method achieves the best values of PSNR and SSIM on the SOTS-Indoor benchmark. Its PSNR is 35.907 dB, exceeding the result of the second-best model (MSBDN) by 10.45%, while its SSIM reaches 0.993, meaning that our FAPANet can restore high-quality haze-free images. For the sake of fairness, all of the data in this table is obtained by repeated experiments of all compared methods. The results of DehazeNet, AOD, and HazeLine are almost the same as those released in [45], and our GDN and MSBDN results are also consistent with those reported in the corresponding papers [26, 27]. It should be noted that our PSNR and SSIM data are calculated by the scikit-image library1 [46] with default parameters. For the SOTS-Outdoor dataset, the PSNR and SSIM results of FAPANet are only slightly lower than those of MSBDN and GDN, but it also produces satisfying dehazed images.

C. HAZE REMOVAL ON NTIRE DATASETS

In recent years, the New Trends in Image Restoration and Enhancement workshop and challenges on image and video processing (NTIRE) has become the most popular academic activity in the field of low-level image and video processing, and haze removal is one of its most important topics. NTIRE 2018, NTIRE 2019, and NTIRE 2020 released four image dehazing datasets, named I-HAZE [47], O-HAZE

1https://scikit-image.org/
FIGURE 5. Qualitative comparisons of dehazing results on SOTS benchmark. Subfigure (a) shows the samples from SOTS-Indoor, while subfigure (b) are samples from SOTS-Outdoor. The letters in the upper right corner of the image indicate the corresponding algorithm: (a) Input image, (b) DCP[1], (c) CEP[42], (d) HazeLine[16], (e) IDE[43], (f) AOD[20], (g) GCANet[44], (h) DehazeNet[18], (i) GDN[27], (j) MSBDN[26], (k) Ours, (l) Ground-truth.
TABLE 2. Quantitative comparisons of dehazing results on SOTS benchmark and the best evaluation values are marked in bold.

| Dataset | Metrics | DCP[1] | CEP[42] | HazeLine[16] | IDE[43] | AOD[20] | GCANet[44] | DehazeNet[18] | GDN[27] | MSBDN[26] | FAPANet |
|---------|---------|--------|---------|-------------|--------|---------|------------|-------------|--------|-----------|--------|
| I-HAZE  | PSNR    | 12.166 | 11.100  | 13.269      | 12.938 | 13.980  | 20.262     | 14.969      | 16.620 | 23.930    | 22.221 |
|         | SSIM    | 0.625  | 0.549   | 0.683       | 0.611  | 0.732   | 0.868      | 0.709       | 0.787  | 0.891     | 0.860  |
| O-HAZE  | PSNR    | 15.858 | 13.186  | 16.144      | 14.873 | 15.030  | 24.164     | 17.111      | 18.920 | 24.360    | 25.317 |
|         | SSIM    | 0.680  | 0.543   | 0.710       | 0.707  | 0.539   | 0.876      | 0.676       | 0.749  | 0.846     |        |
| DenseHaze | PSNR    | 12.085 | 11.602  | 11.621      | 10.775 | 13.140  | 14.688     | 13.840      | 13.310 | 15.370    | 15.969 |
|         | SSIM    | 0.359  | 0.334   | 0.331       | 0.314  | 0.414   | 0.567      | 0.425       | 0.368  | 0.486     | 0.488  |
| NH-HAZE | PSNR    | 11.483 | 10.273  | 11.001      | 11.355 | 15.400  | 16.951     | 16.620      | 13.800 | 19.230    | 21.724 |
|         | SSIM    | 0.495  | 0.410   | 0.485       | 0.552  | 0.569   | 0.729      | 0.534       | 0.537  | 0.706     | 0.730  |

TABLE 3. Quantitative comparisons of dehazing results on NTIRE benchmarks and the best evaluation values are marked in bold.

| Dataset | Metrics | DCP[1] | CEP[42] | HazeLine[16] | IDE[43] | AOD[20] | GCANet[44] | DehazeNet[18] | GDN[27] | MSBDN[26] | FAPANet |
|---------|---------|--------|---------|-------------|--------|---------|------------|-------------|--------|-----------|--------|
| indoor  | PSNR    | 19.414 | 16.118  | 17.292      | 14.331 | 18.734  | 30.230     | 21.347      | 32.147 | 32.510    | 35.907 |
|         | SSIM    | 0.854  | 0.766   | 0.777       | 0.755  | 0.829   | 0.980      | 0.862       | 0.983  | 0.977     | 0.993  |
| outdoor | PSNR    | 17.001 | 13.488  | 18.068      | 15.529 | 19.628  | 25.761     | 22.936      | 30.867 | 32.165    | 28.950 |
|         | SSIM    | 0.849  | 0.698   | 0.849       | 0.799  | 0.892   | 0.947      | 0.900       | 0.981  | 0.968     | 0.960  |

[48], DenseHaze [49], and NH-HAZE [50], respectively. I-HAZE contains 35 pairs of hazy and corresponding haze-free indoor images, O-HAZE and NH-HAZE are both composed of outdoor paired hazy samples, and DenseHaze contains 50 pairs of indoor and outdoor images that have extremely thick haze. Different from the RESIDE dataset, which consists of synthetic hazy images, the NTIRE datasets were constructed from images with real haze generated by a professional haze machine. Therefore, the performance of the dehazing algorithm on the NTIRE datasets is possibly more indicative of its dehazing ability in the real world, so in addition to RESIDE, we also conducted quantitative experiments on four NTIRE datasets.

We still use PSNR and SSIM to evaluate the performance of dehazing methods, and the results are shown in Table 3. The proposed FAPANet obtains the best PSNR results on three datasets: O-HAZE, DenseHaze, and NH-HAZE, and the second-best PSNR results on the I-HAZE dataset after the MSBDN algorithm. In terms of SSIM metrics, the FAPANet algorithm performs best on the NH-HAZE dataset, third-best on the I-HAZE dataset after MSBDN and GCANet, and second-best on both the O-HAZE and DenseHaze datasets after the GCANet algorithm. It can be seen that GCANet has good results in SSIM metrics, while our FAPANet has better PSNR values along with competitive SSIM scores. The quantitative experiment results show the powerful dehazing ability of FAPANet for real-world hazy images, and we believe that this strong dehazing ability benefits from the FAAM, AFFM, and joint loss function proposed in this paper.

D. HAZE REMOVAL ON REAL-WORLD IMAGES

On four NTIRE datasets, we conducted comparison experiments with existing state-of-the-art algorithms. However, the haze in the samples of NTIRE datasets was generated by a haze machine, so verification in a natural hazy scene is required. FAPANet was used to dehaze nine challenging real-world images collected from the internet\(^2\), and the results were compared to those of the nine state-of-the-art methods. Although the prior-based DCP, CEP, and HazeLine methods dehaze well, the dehazed results have low color saturation and significant haloes, as shown in the first, third, and eighth tests in Fig. 6. The neural-network-based dehazing algorithm performs well on thin haze images, but not so well on dense haze scenes, as demonstrated in the second, third, and eighth tests. The IDE produces visually pleasing results due to its combination of exposure processing and dehazing, but its inability to handle thick haze images is revealed in the third and eighth tests. Our method, on the other hand, can remove haze and restore clear, natural images with vivid colors and high visual quality, as well as produce competitive results even when the input images have dense haze.

For the real-world degraded images, it is difficult to obtain the corresponding haze-free ground truth, so we cannot use the full-reference assessments to measure the performance of the dehazing algorithm. Therefore, we additionally use four commonly used no-reference evaluation metrics to measure the restoration quality of the results of each dehazing algorithm. Therefore, we additionally use four commonly used no-reference evaluation metrics to measure the restoration quality of the results of each dehazing algorithm, which are $e$, $r$, and $\sigma$, proposed in [51], and the Fog Aware Density Evaluator (FADE) [52], respectively. The $e$ metric indicates the capability of the dehazing algorithm to restore new visual edges in the dehazed image; the $r$ metric estimates the visibility enhancement level of the restored image; the $\sigma$ calculates the ratio of saturated pixels. FADE predicts the fog density of a hazy image based on statistical features. The $\sigma$ and FADE indicate better image visual quality with small values, while $e$ and $r$ are opposite. For fairness, we list the averages of the no-reference evaluation scores in Table 4 for all comparison algorithms dehazing on all nine test samples shown in Fig. 6. The proposed FAPANet

\(^2\)http://www.cs.huji.ac.il/~raananf/projects/dehaze_cl/results/
FIGURE 6. Visual comparisons of dehazing results on real-world hazy images. The images in each row from left to right are haze removal results of the following algorithms: (a) Input hazy image, (b) DCP[1], (c) CEP[42], (d) HazeLine[16], (e) IDE[43], (f) AOD[20], (g) GCANet[44], (h) DehazeNet[18], (i) GDN[27], (j) MSBDN[26], (k) Ours. Best viewed on a high-resolution display with zoom-in.

TABLE 4. No-reference assessment scores of dehazing results for 9 challenging real-world hazy images shown in Fig. 6. All the data listed in the table are the average scores of the evaluation results of the corresponding dehazing algorithms on all of these 9 test images. Texts in red, blue, and green indicate the best, the second-best, and the third-best performance, respectively. Symbol ↑ indicates that the higher the value, the better the result, while symbol ↓ indicates the opposite.

| Metrics | DCP[1]  | CEP[42] | HazeLine[16] | IDE[43] | AOD[20] | GCANet[44] | DehazeNet[18] | GDN[27] | MSBDN[26] | FAPANet |
|---------|---------|---------|-------------|---------|---------|------------|--------------|---------|-----------|---------|
| e ↑     | 1.443   | 1.488   | 1.169       | 1.006   | 0.865   | 0.626      | 0.721        | 0.736   | 0.626     | 1.374   |
| r ↑     | 2.001   | 1.905   | 2.368       | 3.028   | 1.424   | 1.670      | 1.257        | 1.316   | 1.260     | 2.724   |
| σ ↓     | 0.006   | 0.389   | 1.389       | 0.326   | 1.305   | 0.003      | 2.175        | 0.002   | 0.008     | 0.043   |
| FADE ↓  | 0.295   | 0.208   | 0.219       | 0.346   | 0.393   | 0.591      | 0.442        | 0.577   | 0.576     | 0.270   |

achieves good scores in terms of e, r, and FADE metrics while maintaining a respectable σ result, which shows that our algorithm reconstructs clear images with nice visuality.

E. ABLATION STUDY

We conducted sufficient ablation experiments concerning three factors: the feature enhancement strategy, the feature fusion technique, and the loss function used in the training stage, denoted in Table 5 by "FF," "FE," and "Loss," respectively. According to previous research [24, 25], we used the simplified UNet with two downsampling operations as a baseline, then combined it with various feature enhancement and feature fusion methods to create a series of variant models, which we then trained with different loss functions. Finally, all models were quantitatively evaluated on the SOTS-Indoor dataset using by PSNR and SSIM. Table 5 displays all of the comparison results.

1) EFFECTIVENESS OF THE FAAM

We demonstrate the effectiveness of the FAAM proposed in Section III-B through the performance comparison of models (a), (b), (c), (d), (e), (f), and (g). The model (a) is the baseline, a simplified UNet in which the output layer of the encoder is directly connected to the input layer of
| Serial | FE | FF | Loss | PSNR | SSIM | Params(M) |
|--------|----|----|------|------|------|-----------|
| (a)    | w/o add | L1  | 23.23 | 0.9002 | 0.76 |
| (b)    | FFA*6 add | L1  | 32.82 | 0.9861 | 9.29 |
| (c)    | RB*6 add | L1  | 32.16 | 0.9825 | 9.14 |
| (d)    | CBAM*6 add | L1  | 33.31 | 0.9883 | 9.19 |
| (e)    | SACA*6 add | L1  | 33.53 | 0.9884 | 12.29 |
| (f)    | CASA*6 add | L1  | 33.20 | 0.9880 | 12.29 |
| (g)    | FAAM*6 add | L1  | 33.85 | 0.9892 | 12.29 |
| (h)    | FAAM*8 add | L1  | 34.19 | 0.9901 | 15.70 |
| (i)    | FAAM*10 add | L1  | 34.44 | 0.9902 | 19.11 |
| (j)    | FAAM*12 add | L1  | 34.03 | 0.9898 | 22.51 |
| (k)    | FAAM*10 w/o | L1  | 33.79 | 0.9877 | 19.11 |
| (l)    | FAAM*10 concat | L1  | 33.80 | 0.9883 | 19.27 |
| (m)    | FAAM*10 AFFM | L1  | 34.58 | 0.9905 | 20.58 |
| (n)    | FAAM*10 AFFM L2 |  | 34.15 | 0.9889 | 20.58 |
| (o)    | FAAM*10 AFFM ssim |  | 35.57 | **0.9946** | 20.58 |
| (p)    | FAAM*10 AFFM JLoss |  | **35.93** | 0.9932 | 20.58 |

TABLE 5. Quantitative results of the ablation experiments on SOTS-Indoor. Each row of the table represents a variant model, where model (m) is the FAPANet proposed in this paper. The backbone of all models listed in table is a simplified UNet.

3) EFFECTIVENESS OF THE AFFM
Similarly, we also verified the effectiveness of the AFFM module proposed in Section III-C. We constructed models (i) and (m) based on model (i) by replacing its addition part with the concatenation operation and AFFM, respectively. And we removed the addition operation of model (i) to build model (k), which had no feature fusion module. The quantitative comparison results show that AFFM helps the network recover better quality images.

4) EFFECTIVENESS OF THE JOINT LOSS
In this experiment, we compared the performance differences between the FAPANets trained with L1 loss (model (m)), L2 loss (model (n)), SSIM Loss (model (o)), and the joint loss (model (p)) proposed in Section III-D, respectively. As shown in Table 5, the joint loss, abbreviated as "JLoss" in the table, contributes to a 5.2% improvement in the PSNR evaluation of the model compared to L2 loss, while keeping a competitive SSIM value.

V. ANALYSIS AND DISCUSSIONS
In this section, we first show the effect of the feature attention aggregation module proposed in this paper by the visualization technology, followed by the discussion of the running time, and some failure cases are presented at the end.

A. VISUALIZATION OF THE FEATURE ATTENTION
The design of the FAAM is one of our main contributions, and the ablation study in Section IV-E2 quantitatively verifies the usefulness of the module. To further demonstrate its effect in feature enhancement, we use visualization techniques to explicitly show the visual comparison of feature maps before and after FAAM enhancement. We use models (b), (c), (d), (e), (f), and (g), described in Table 5, to remove the haze in two hazy images, which are samples with heterogeneous and thick haze from the NH-HAZE[50] dataset. We then show the original features calculated by the encoder of each model and the features recalibrated by the first feature enhancement block, which follows the encoder.

As shown in Fig. 7, no obvious feature excitation or suppression happens in the results of models (b) and (c), while model (d), which employs CBAM[34] as the feature enhancement module, over-excites features in the global range and blurs the texture, thus leaving a large amount of haze or halo artifacts in dehazed images, as shown in the red region of the first sample and the green region of the second sample, respectively. In the last row of Fig. 7, model (g), which uses the proposed FAAM, exhibits suitable feature excitation in the dense haze region and thus achieves good dehazed results.

B. RUNNING TIME
To compare the computational time of the proposed method with some state-of-the-art algorithms [1, 16, 20, 26, 27, 42,
FIGURE 7. Visualization results of feature attention enhancement. The first row of the figure shows the input hazy images and the corresponding ground truths, and the subsequent rows 2 to 7 show the visualization results of the feature enhancement by models (b), (c), (d), (e), (f), and (g) described in Table 5, respectively. Columns (a) and (b) indicate the features before and after enhancement, and column (c) indicates the dehazed images of the models mentioned above. Best viewed on a high-resolution display with zoom-in.

| Methods       | Platform       | Time (Seconds) |
|---------------|----------------|----------------|
| DCP[1]        | Matlab(CPU)    | 1.395          |
| CEP[42]       | Python(CPU)    | 0.061          |
| HazeLine[16]  | Matlab(CPU)    | 3.906          |
| IDE[43]       | Matlab(CPU)    | 0.720          |
| AOD[20]       | Pytorch(GPU)   | 0.017          |
| GCANet[44]    | Pytorch(GPU)   | 0.071          |
| GDN[27]       | Pytorch(GPU)   | 0.076          |
| MSBDN[26]     | Pytorch(GPU)   | 0.064          |
| FAPANet(ours) | Pytorch(GPU)   | 0.062          |

TABLE 6. Comparison of the average time for processing 500 images with a resolution of 460 × 620 by different methods.
spatial attention in FAAM is parameterless and plain, which dramatically reduces the computational complexity of the overall model. The result shows that our algorithm is competitive compared with other state-of-the-art methods in time overhead, but at the same time, the dehazing capability is significantly improved.

C. FAILURE CASES

While our algorithm works well on most daytime images, it fails if the input image is taken at night or in dense foggy weather. As shown in the first row of Fig. 8, our method cannot remove the haze well and even produces halos in the area marked by the green rectangle. This is because the nighttime images are significantly different in light distribution from the daytime images, and the physical degradation models of haze images are different for day and night. Our algorithm is trained on a large number of daytime images and therefore does not generalize well to nighttime scenes. For scenes with very thick haze, the loss of image detail is so severe that our model is unable to recover enough image content from subtle cues. As shown in the second example in Fig. 8, our algorithm removes the haze in the near part of the scene, as marked by the red rectangle, but does not help with the dense fog in the distance. We will try to address these two issues in future work.

VI. CONCLUSION

In this study, we have presented an effective single image dehazing method. This new approach removes haze in an end-to-end way, thereby potentially reducing the risk of error amplification caused by the independent evaluation of global atmospheric light and the transmission map. To extract more effective haze-relevant features, we utilize the specifically designed feature attention aggregation module to progressively enhance the feature map of the hazy image. The feature fusion mechanism is then employed to effectively merge the features calculated from the shallow and deep layers of the network. In order to recover realistic clear images, we combine L1 loss, perceptual loss, and SSIM loss to train our network. Quantitative and qualitative experiment results demonstrate that the proposed algorithm outperforms existing advanced algorithms in the dehazing of both synthetic and real-world hazy images.

REFERENCES

[1] Kaiming He, Jian Sun, and Xiaou Tang. Single image haze removal using dark channel prior. IEEE transactions on pattern analysis and machine intelligence, 33(12):2341–2353, 2010.
[2] Earl J McCartney. Optics of the atmosphere: scattering by molecules and particles. New York, 1976.
[3] Tali Treibitz and Yoav Y Schechner. Polarization: Beneficial for visibility enhancement? In 2009 IEEE Conference on Computer Vision and Pattern Recognition, pages 525–532. IEEE, 2009.
[4] Yoav Y Schechner, Srinivasa G Narasimhan, and Shree K Nayar. Instant dehazing of images using polarization. In CVPR (1), pages 325–332, 2001.
[5] Srinivasa G Narasimhan and Shree K. Nayar. Contrast restoration of weather degraded images. IEEE transactions on pattern analysis and machine intelligence, 25(6):713–724, 2003.
[6] Johannes Kopf, Boris Neubert, Billy Chen, Michael Cohen, Daniel Cohen-Or, Oliver Deussen, Matt Uyttendaele, and Dani Lischinski. Deep photo: Model-based photograph enhancement and viewing. ACM transactions on graphics (TOG), 27(5):1–10, 2008.
[7] Srinivasa G Narasimhan and Shree K. Nayar. Interactive (de) weathering of an image using physical models. In IEEE Workshop on color and photometric Methods in computer Vision, volume 6, page 1. France, 2003.
[8] J Alex Stark. Adaptive image contrast enhancement using generalizations of histogram equalization. IEEE Transactions on image processing, 9(5):889–896, 2000.
[9] Jong-Youn Kim, Lee-Sup Kim, and Seung-Ho Hwang. An advanced contrast enhancement using partially overlapped sub-block histogram equalization. IEEE transactions on circuits and systems for video technology, 11(4):475–484, 2001.
[10] Codruta Orniana Ancuti and Cosmin Ancuti. Single image dehazing by multi-scale fusion. IEEE Transactions on Image Processing, 22(8):3271–3282, 2013.
[11] Adrian Galdran, Javier Vazquez-Corral, David Pardo, and Marcelo Bertalmio. Fusion-based variational image dehazing. IEEE Signal Processing Letters, 24(2):151–155, 2016.
[12] Jinbao Wang, Ke Lu, Jian Xue, Ning He, and Ling Shao. Single image dehazing based on the physical model and msrcr algorithm. IEEE Transactions on Circuits
and Systems for Video Technology, 28(9):2190–2199, 2018.

[13] Adrian Galdran, Aitor Alvarez-Gila, Alessandro Bria, Javier Vazquez-Corral, and Marcelo Bertalmio. On the duality between retinex and image dehazing. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 8212–8221, 2018.

[14] Robby T Tan. Visibility in bad weather from a single image. In 2008 IEEE Conference on Computer Vision and Pattern Recognition, pages 1–8. IEEE, 2008.

[15] Raanan Fattal. Dehazing using color-lines. ACM transactions on graphics (TOG), 34(1):13, 2014.

[16] Dana Berman, Tali Treibitz, and Shai Avidan. Single image dehazing using haze-lines. IEEE transactions on pattern analysis and machine intelligence, 42(3):720–734, 2018.

[17] Qingsong Zhu, Jiaming Mai, and Ling Shao. A fast single image haze removal algorithm using color attenuation prior. IEEE transactions on image processing, 24(11):3522–3533, 2015.

[18] Bolun Cai, Xiangmin Xu, Kui Jia, Chunmei Qing, and Dacheng Tao. Dehazenet: An end-to-end system for single image haze removal. IEEE Trans. Image Process., 25(11):5187–5198, 2016.

[19] Wenqi Ren, Si Liu, Hua Zhang, Jinshan Pan, Xiaochun Cao, and Ming-Hsuan Yang. Single image dehazing via multi-scale convolutional neural networks. In European conference on computer vision, pages 154–169. Springer, 2016.

[20] Boyi Li, Xiulian Peng, Zhangyang Wang, Jizheng Xu, and Dan Feng. Aod-net: All-in-one dehazing network. In Proceedings of the IEEE international conference on computer vision, pages 4770–4778, 2017.

[21] He Zhang and Vishal M Patel. Densely connected pyramid dehazing network. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 3194–3203, 2018.

[22] Dong Yang and Jian Sun. Proximal dehaze-net: a prior learning-based deep network for single image dehazing. In Proceedings of the European Conference on Computer Vision (ECCV), pages 702–717, 2018.

[23] Wenqi Ren, Lin Ma, Jiawei Zhang, Jinshan Pan, Xiaochun Cao, Wei Liu, and Ming-Hsuan Yang. Gated fusion network for single image dehazing. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 3253–3261, 2018.

[24] Haiyan Wu, Yanyun Qu, Shaohui Lin, Jian Zhou, Ruizhi Qiao, Zhizhong Zhang, Yuan Xie, and Lizhuang Ma. Contrastive learning for compact single image dehazing. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 10551–10560, 2021.

[25] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention, pages 234–241. Springer, 2015.

[26] Hang Dong, Jinshan Pan, Lei Xiang, Zhe Hu, Xinyi Zhang, Fei Wang, and Ming-Hsuan Yang. Multi-scale boosted dehazing network with dense feature fusion. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 2157–2167, 2020.

[27] Xiaohong Liu, Yongrui Ma, Zhihao Shi, and Jun Chen. Griddehazenet: Attention-based multi-scale network for image dehazing. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pages 7314–7323, 2019.

[28] Lex Schaul, Clément Fredembach, and Sabine Süsstrunk. Color image dehazing using the near-infrared. In 2009 16th IEEE International Conference on Image Processing (ICIP), pages 1629–1632. IEEE, 2009.

[29] Runde Li, Jinshan Pan, Zechao Li, and Jinhui Tang. Single image dehazing via conditional generative adversarial network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 8202–8211, 2018.

[30] Yanyun Qu, Yizi Chen, Jingying Huang, and Yuan Xie. Enhanced pix2pix dehazing network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 8160–8168, 2019.

[31] Shiba Kuanar, Dwarkanath Mahapatra, Monalisa Bilas, and KR Rao. Multi-path dilated convolution network for haze and glow removal in nighttime images. The Visual Computer, pages 1–14, 2021.

[32] Xiangyu Xu, Muchen Li, and Wenzhu Sun. Learning deformable kernels for image and video denoising. arXiv preprint arXiv:1904.06903, 2019.

[33] Jifeng Dai, Haozhi Qi, Yuwen Xiong, Yi Li, Guodong Zhang, Han Hu, and Yichen Wei. Deformable convolutional networks. In Proceedings of the IEEE international conference on computer vision, pages 764–773, 2017.

[34] Sanghyun Woo, Jongchan Park, Joon-Young Lee, and In So Kweon. Cbam: Convolutional block attention module. In Proceedings of the European conference on computer vision (ECCV), pages 3–19, 2018.

[35] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 770–778, 2016.

[36] Jongchan Park, Sanghyun Woo, Joon-Young Lee, and In-So Kweon. Bam: Bottleneck attention module. In British Machine Vision Conference (BMVC), 2018.

[37] Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 7132–7141, 2018.

[38] Qilong Wang, Banggu Wu, Pengfei Zhu, Peihua Li, Wangmeng Zuo, and Qinghua Hu. Eca-net: efficient channel attention for deep convolutional neural net-
works, 2020 ieee. In CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2020.

[39] Lingxiao Yang, Ru-Yuan Zhang, Lida Li, and Xiaohua Xie. Simam: A simple, parameter-free attention module for convolutional neural networks. In International Conference on Machine Learning, pages 11863–11874. PMLR, 2021.

[40] Hang Zhao, Orazio Gallo, Iuri Frosio, and Jan Kautz. Loss functions for image restoration with neural networks. IEEE Transactions on Computational Imaging, 3(1):47–57, 2016.

[41] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. In 3rd International Conference on Learning Representations, ICLR 2015, 2015.

[42] Trung Minh Bui and Wonha Kim. Single image dehazing using color ellipsoid prior. IEEE Transactions on Image Processing, 27(2):999–1009, 2017.

[43] Mingye Ju, Can Ding, Wenqi Ren, Yi Yang, Dengyin Zhang, and Y Jay Guo. Ide: Image dehazing and exposure using an enhanced atmospheric scattering model. IEEE Transactions on Image Processing, 30:2180–2192, 2021.

[44] Dongdong Chen, Mingming He, Qingnan Fan, Jing Liao, Liheng Zhang, Dongdong Hou, Lu Yuan, and Gang Hu. Gated context aggregation network for image dehazing and deraining. In 2019 IEEE Winter Conference on Applications of Computer Vision (WACV), pages 1375–1383. IEEE, 2019.

[45] Boyi Li, Wenqi Ren, Dengpan Fu, Dacheng Tao, Dan Feng, Wenjun Zeng, and Zhangyang Wang. Benchmarking single-image dehazing and beyond. IEEE Transactions on Image Processing, 28(1):492–505, 2018.

[46] Stefan Van der Walt, Johannes L Schönberger, Juan Nunez-Iglesias, François Boulogne, Joshua D Warner, Neil Yager, Emmanuelle Gouillart, and Tony Yu. scikit-image: image processing in python. PeerJ, 2:e453, 2014.

[47] Cosmin Ancuti, Codruta O Ancuti, Radu Timofte, and Christophe De Vleeschouwer. I-haze: A dehazing benchmark with real hazy and haze-free indoor images. In International Conference on Advanced Concepts for Intelligent Vision Systems, pages 620–631. Springer, 2018.

[48] Codruta O Ancuti, Cosmin Ancuti, Radu Timofte, and Christophe De Vleeschouwer. O-haze: A dehazing benchmark with real hazy and haze-free outdoor images. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pages 754–762, 2018.

[49] Codruta O Ancuti, Cosmin Ancuti, Mateu Shert, and Radu Timofte. Dense-haze: A benchmark for image dehazing with dense-haze and haze-free images. In 2019 IEEE International Conference on Image Processing (ICIP), pages 1014–1018. IEEE, 2019.

[50] Codruta O Ancuti, Cosmin Ancuti, and Radu Timofte. Nh-haze: An image dehazing benchmark with non-homogeneous hazy and haze-free images. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, pages 444–445, 2020.

[51] Nicolas Hautiere, Jean-Philippe Tarel, Didier Aubert, and Eric Dumont. Blind contrast enhancement assessment by gradient ratioing at visible edges. Image Analysis & Stereology, 27(2):87–95, 2008.

[52] Lark Kwon Choi, Jaehie You, and Alan Conrad Bovik. Referenceless prediction of perceptual fog density and perceptual image defogging. IEEE Transactions on Image Processing, 24(11):3888–3901, 2015.

[53] Xu Qin, Zhilin Wang, Yuanchao Bai, Xiaodong Xie, and Huizhu Jia. Ffa-net: Feature fusion attention network for single image dehazing. In Proceedings of the AAAI Conference on Artificial Intelligence, volume 34, pages 11908–11915, 2020.