



Feature aggregation and modulation network for single image dehazing

Fei Tan¹ · Xiaoyuan Yu¹ · Renjie Wang¹ · Baoquan Ai¹ · Fengguo Li¹

Received: 7 May 2023 / Revised: 25 July 2023 / Accepted: 3 October 2023 /
Published online: 4 November 2023

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2023

Abstract

Deep learning-based methods have recently achieved satisfying results in image dehazing. However, we observe that various researchers devote themselves to learning haze-free images directly, while often paying no attention to the physical features of the hazy image formation process. For single image dehazing, a suitable transmission map and global atmospheric light guidance proved effective. Meanwhile, for many dehazing networks, deep and non-adjacent feature information is not utilized which can likewise affect the effectiveness of image recovery. Therefore, we develop an effective feature aggregation and modulation network for image dehazing called FAM-Net. Specifically, the proposed FAM-Net first uses CNN to estimate the transmission map and global atmospheric light, and then embeds the output features into the overall network for joint dehazing. A feature aggregation and modulation module is proposed to fuse the extracted features of atmospheric light and transmission map into the network. Moreover, the attention guidance aggregation module is designed as a replacement for the skip connection. Furthermore, a novel edge-preserving loss function is proposed for training the network, preserving more details of the reconstructed images. Experimental results indicate that FAM-Net outperforms existing dehazing methods in quantitative and qualitative aspects.

Fei Tan and Xiaoyuan Yu contributed equally to this work

✉ Fengguo Li
lifengguo@m.scnu.edu.cn

Fei Tan
tanf@m.scnu.edu.cn

Xiaoyuan Yu
xiaoyuanyu@scnu.edu.cn

Renjie Wang
2022022233@m.scnu.edu.cn

Baoquan Ai
aibq@scnu.edu.cn

¹ School of Physics and Telecommunication Engineering, South China Normal University, Guangzhou 510006, China

Keywords Physical model · Single image dehazing · Feature aggregation and modulation · Attention mechanism

1 Introduction

Haze is a common atmospheric phenomenon, usually caused by tiny suspended particles in the atmosphere. Images captured in hazy scenes have significant quality degradation, as the visibility of objects is reduced after tiny suspended particles absorb and scatter light. With these images as input, many subsequent tasks are susceptible to dramatic performance degradation with pre-trained models, such as computer vision tasks like detection, tracking, classification, and segmentation. Therefore, several dehazing approaches [1–3] are proposed to improve the visibility of hazy images for these issues. According to the atmospheric scattering model [4, 5], the formation of a hazy image can be expressed as:

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

where $I(x)$ is the hazy input image, $J(x)$ is the recovered haze-free image, A denotes global atmospheric light, and $t(x)$ denotes the transmission map defined by $t(x) = e^{-\beta d(x)}$. Here, β is the atmospheric scattering coefficient, and $t(x)$ is the scene depth. x is the position of the pixel in the image. The purpose of single image dehazing is to restore a haze-free image $J(x)$ from a given hazy image $I(x)$, which is a highly ill-posed problem.

Previous prior-based methods [1, 6, 7] used different hand-crafted image priors to estimate transmission maps. For example, in DCP [1], the atmospheric light is estimated from the dark channel prior. Nevertheless, it is worth noting that these conventional techniques tend to be associated with a significant time overhead in the inference phase. In recent years, learning-based methods for single image dehazing have emerged as a promising alternative and verified excellent performance. Some of these methods [8–10] learn by embedding (1) directly into the network, while estimating transmission maps, global atmospheric light and haze-free images. However, some studies fail to fully leverage the physical information available, resulting in features that are not sufficiently independent when aggregated for each physical quantity. As a result, these features can be easily impacted and confounded by one another. In addition, the self-attention models have also shown promising results, but their network structure is huge and the edges are less retained. Therefore, additional research is necessary to develop attention guidance without increasing the network parameters and retain more edges.

This study presents a novel deep learning-based framework for addressing the aforementioned problems related to haze removal. The proposed framework, called the feature aggregation and modulation network (FAM-Net), is shown in Fig. 1. In particular, we introduce the feature aggregation and modulation module (FAMM), which aggregates and modulates the transmission map, atmospheric light and hazy image features by unfolding the (1). Furthermore, multiple modules are combined and designed to work progressively to facilitate the preservation of image details and further restore visibility. As such, features from transmission map and atmospheric light can better guide the network to generate higher definition images, and play a crucial role in guiding various image regions and preserving details. Moreover, the attention guidance aggregation module (AGAM) is designed to replace the original skip connection. Additionally, this study proposes a novel edge-preserving loss function to maintain the sharp edge of the estimated haze-free images without halo-shaped artifacts. The proposed loss function is combined with several existing loss functions to jointly train our network.

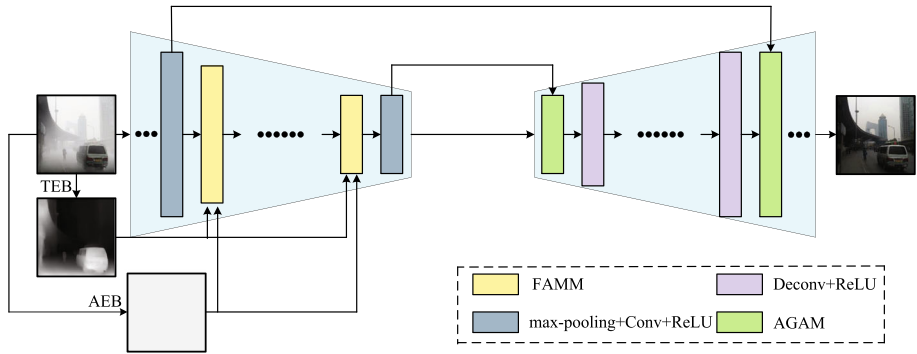


Fig. 1 Illustration of our method. the transmission map estimation branch and the atmospheric light estimation branch are embedded in the encoder by the progressive FAMMs for joint training

The contributions are recapitulated as follows:

- We develop an efficient single image dehazing network, called FAM-Net, which is a framework for simultaneously estimating the transmission map, atmospheric light, and haze-free images.
- The proposed feature aggregation and modulation module (FAMM) extracts deep level information and facilitates the interaction and fusion between features from different levels. Additionally, the attention guided aggregation module (AGAM) adaptively fuses the weights of different feature channels to improve the dehazing effect.
- A novel edge-preserving loss function is proposed to retain more details and edge information of the estimated haze-free images.

The outline of this paper is as follows. We summarize the research progress on single image dehazing in Section 2. And Section 3 details the proposed FAM-Net architecture and loss function. In Section 4, experiments are conducted on challenging datasets with different scenarios to illustrate the effectiveness of our proposed network. In addition, ablation studies are reported in this section. Finally, some conclusions are drawn in Section 5.

2 Related work

The related work is divided into three categories. These include physical model-based approaches, feature aggregation-based approaches and attention mechanism-based approaches. The physical model-based approaches [1, 8, 10] require priors or CNNs to obtain the transmission map and global atmospheric light. These approaches then use the atmospheric scattering model to recover the mathematical inversion process for the haze-free image. On the other hand, physical model-free approaches are to learn the mapping from hazy image to haze-free image directly. Common approaches for this include feature aggregation [11–13] and attention mechanisms [3, 14, 15].

2.1 Physical model-based dehazing

Methods that rely on the atmospheric scattering model are divided into prior-based method and learning-based method. The prior-based method obtains prior knowledge based on hand-crafted statistics regarding the differences between the haze-free and hazy images.

In DCP [1], the author considers that, for outdoor images, there is at least one color channel with intensity close to zero in the local area of the haze-free image. Additionally, some methods [6, 7] analyze the linear relationship of the haze-free channels in RGB color space from different perspectives to estimate the transmission map. Traditional prior-based methods usually have a good dehazing ability. However, the dehazed images are prone to distortion and artifacts in the regions that do not satisfy the priors, and the image realism is low. With the development of deep learning, learning-based methods have become the primary methods for image dehazing in recent years. For example, MSCNN-HE [16] obtained a sharp-edged transmission map using CNN and estimated the global atmospheric light using traditional methods. However, estimating both separately does not guarantee that the final solution is jointly optimal. In contrast, DCPDN [8] fully embeds the atmospheric scattering model into the end-to-end physical dehazing model of the overall optimization framework. This network jointly learns the dehazed image, transmission map, and atmospheric light to obtain a jointly optimal solution. Similarly, HRGAN [17] proposed a unified network for joint estimation of transport maps, atmospheric light, and haze-free images as a generator network. By reformulating the atmospheric scattering model, AOD-Net [18] predicted only a single output $K(x)$ that has no physical significance. Similar ideas are also found in several approaches [19, 20]. PFDN [10] embeds the proposed feature dehazing units into an encoder structure with residual learning for end-to-end training. Based on the physical model, the formation process of hazy images can be described, which facilitates the achievement of a dehazing effect. However, existing methods encounter limitations in effectively aggregating and modulating the physical features, and they struggle to seamlessly embed the features into the main dehazing path. Consequently, the restored images may still retain some hazy residuals. As a result, there remains ample room for improvement in physical model-based methods.

2.2 Feature aggregation-based dehazing

Feature fusion and aggregation have been widely used in network design, improving the performance by exploring different levels of features. The fusion and aggregation strategy provides a new idea of using features with hazy inputs. Riaz et al [21] implemented a depth graph estimation technique based on guided fusion, effectively reducing the possibility of failure of DCP [1] and recovery artifacts. In GFN [11], the author proposed a multi-scale gated fusion network to generate dehazed images using an encoder-decoder architecture. MSBDN-DFF [22] utilized enhancement strategies and back-projection techniques to enhance feature fusion. GCANet [12] efficiently aggregated contextual information by dividing the feature map into branches and introducing gating mechanisms between the branches. RefinedNet [13] proposed an effective perceptual fusion strategy to fuse different dehazing outputs. However, these methods do not explicitly exploit the features and properties of non-adjacent layers and deep layers, and are not easily applied to other architectures. Furthermore, several networks [8, 17, 18] referenced in Section 2.1 conducted the results-oriented aggregation at the end of the network using the physical model. Conversely, Wang et al [23] have experimentally substantiated that the process-oriented aggregation is effective in circumventing suboptimal dehazing outcomes.

2.3 Attention mechanisms-based dehazing

Attention mechanisms have been increasingly applied in image dehazing due to their reliable ability in feature extraction and image recovery processes [24, 25]. GridDehazeNet [14]

employed a three stages attention-based grid network to restore haze-free images. In FFA-Net [3], the author proposed a new feature attention module that fuses channel attention with pixel attention to obtain enhance the effectiveness of haze removal. LapDehazeNet [26] used attention sharing weights K to approximate higher order Taylor terms, thus avoiding the overhead associated with direct convolution of UHD images. In recent years, the Transformer, an encoder-decoder architecture based on the self-attentive mechanism, has been widely used in image restoration [15, 27, 28]. Although transformer has achieved better results, certain limitations still persist. Among these limitations, computationally complex and high data requirements are prominent issues. Moreover, it is difficult to handle spatial information effectively, which is critical for image-based tasks. Additionally, an excessive focus on global information may result in neglecting image details, such as textures and edges.

3 Proposed method

In this section, we first present the designed haze removal framework. Figure 1 illustrates the general architecture of the proposed FAM-Net. In addition, we employ multiple loss functions to further constrain our network to achieve better single image dehazing. The network consists of five main components: the encoder-decoder based dehazing main path, FAMM, AGAM, TEB, and AEB. Specifically, the main path acquires the primary content information of the hazy objects, and our proposed FAMM is used for feature aggregation and modulation. AGAM is an effective way to connect the encoder and the decoder. Moreover, TEB and AEB are used to obtain corresponding transmission map and global atmospheric light from the hazy image, respectively. We will describe the algorithm's details in the following.

3.1 The overall network structure

As shown in Fig. 1, FAM-Net is designed based on an encoder-decoder structure, integrating the transmission map and the global atmospheric light from TEB and AEB into the image dehazing process. This method encodes the hazy images first and then decodes them to obtain haze-free images. In this process, the features of the modulated transmission maps and atmospheric lights are combined with the features of the main path using the proposed FAMM. Simultaneously, the multi-level feature residuals from the encoder are fused with the features in the decoder using AGAM.

3.2 Feature aggregation and modulation module (FAMM)

The process of fusing features extracted at different levels is a critical step. However, due to the differences in scales and dimensions of the features, traditional fusion methods such as summation, multiplication, or concatenation are often less effective. Therefore, we develop the feature aggregation and modulation module (FAMM) shown in Fig. 2 to fuse the features from TEB, AEB, and the main path of the encoder. We rewrite (1) as follows:

$$J(x) = \frac{1}{t(x)} (I(x) - A) + A \quad (2)$$

In the process of refining the transmission map, we use the features from TEB as a guide and modulate them using FAMM. We apply sequentially stacked convolutional layers, each of which is followed by a ReLU nonlinearity. A max-pooling layer follows the five convolutional

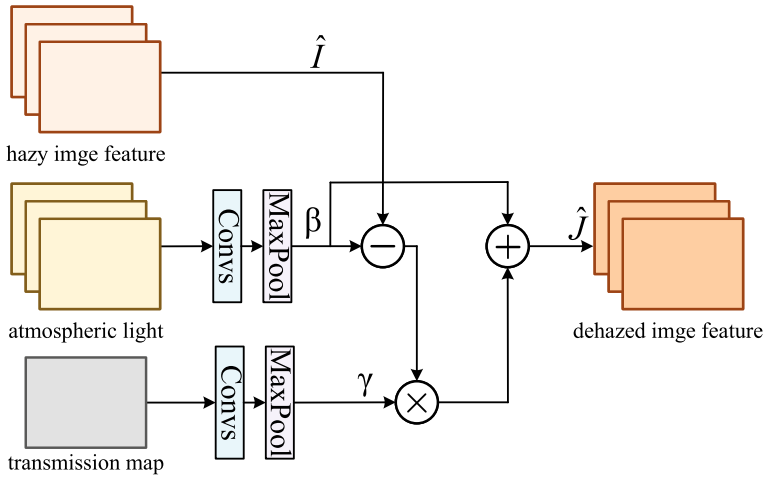


Fig. 2 structure of FAMM. This module extracts deep level information, and facilitates the interaction and fusion between features from transmission maps, atmospheric light, and hazy images at different levels

layers to reduce them to the same size as the main path. This provides us with the modulation parameter γ for the feature map $\frac{1}{t}$. Additionally, since the number of features at the output of the atmospheric light branch is small, only two convolutional layers with ReLU and a max-pooling layer are needed to complete the feature modulation, giving the modulation parameter β from the feature map A . The hazy image feature \hat{I} of the main path has already undergone feature extraction by the encoder so that can be directly input to FAMM without feature modulation. Consequently, we can express the output \hat{J} of FAMM as:

$$\hat{J}(x) = \gamma \otimes (\hat{I}(x) \ominus \beta) \oplus \beta \tag{3}$$

Where \otimes denotes element-wise multiplication, \ominus denotes element-wise subtraction, and \oplus denotes element-wise addition. The above equation is used to perform an element-wise subtraction of the input feature I with the modulation parameter β , it performs the multiplication operation with the modulation parameter γ and sum with β . FAMM plays a crucial role in the extract and aggregate deep level features from the input data to guide the image dehazing process. The resulting feature maps are used to compute the modulation parameters γ and sum with β , which are used to modulate the image features. FAMM can dynamically adjust the weighting of different features, allowing the network to focus on the most relevant information for the dehazing task.

Meanwhile, although the \hat{J} generated by a single FAMM aggregates part of the haze-free information of the image, it also contains some hazy residuals, so a single FAMM cannot effectively explore the useful image dehazing features. To better recover haze-free image, we construct FAMM as a progressive aggregation. As shown in Fig. 3, we construct a progressive aggregation structure containing 2 FAMMs. First, we define the output $F_{famm}^i(x)$ of the i -th FAMM as:

$$F_{famm}^i(x) = FAMM \left(F_{main}^i(x), t(x), A \right) \tag{4}$$

where F_{main}^i is the input to the i -th FAMM from the encoder main path, $t(x)$ is the transmission map estimated by TEB, and A denotes the the global atmospheric light estimated by AEB. Here, x is the position of the pixel in the image. After inputting $F_{famm}^i(x)$ feedback to

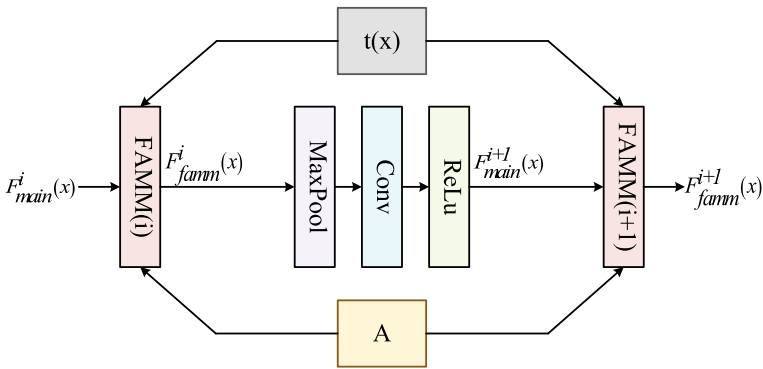


Fig. 3 Network architecture of the proposed progressive FAMMs at the i -th and $i + 1$ -th level of the encoder. The i -th FAMM is used to aggregate and modulate the features of the encoder main path F^i_{main} , $T(x)$, A to obtain F^i_{famm} , which is fed back to the encoder main path

the encoder main path, it passes through a max-pooling layer and a Convolution layer with ReLU. Features after downsampling by this series of operations, become the input $F^{i+1}_{main}(x)$ to the $(i + 1)$ -th FAMM.

3.3 Attention guidance aggregation module (AGAM)

The proposed AGAM enhances feature representation by aggregating the residuals from the encoder and features from the decoder using a Squeeze-and-Excitation (SE) mechanism, inspired by several networks [28–30] that utilize attention mechanisms. This allows the model to learn the interdependencies between feature channels and adaptively emphasize clear channels while suppressing hazy ones, resulting in improved dehazing performance.

As shown in Fig. 4, AGAM takes the residuals x_1 from the encoder and the feature x_2 from the decoder as inputs. Firstly, x_1 is transformed by the F_{tr} operation, which involves a 1×1 convolution with a ReLU activation function, to extract more useful features and obtain \hat{x}_1 .

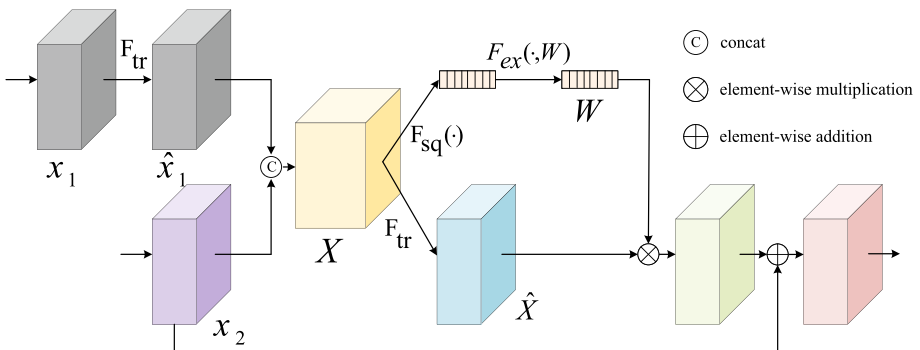


Fig. 4 Network architecture of attention guidance aggregation module. AGAM adaptively emphasize clear channels while suppressing hazy ones

Then, \hat{x}_1 is concatenated with the high-level feature information feature x_2 along the channel dimension, to create a richer feature X . The formulas are as follows:

$$X = \text{cat}([\delta(\text{Conv}(x_1)), x_2]) \quad (5)$$

Where δ is ReLU activation function. Next, the feature X is squeezed using $F_{sq}(\cdot)$, which applies global average pooling over the spatial dimensions of the feature map $H \times W$ and compresses the feature map to 1×1 . The outcome of this process is the squeeze of global spatial information into the channel descriptor, effectively expanding the receptive field. The resulting feature map passes through the $F_{ex}(\cdot)$ excitation, which includes a 1×1 convolution with a Sigmoid activation function, to obtain the channel attention weights W . The Sigmoid activation function enables the learning of nonlinear interactions between channels and facilitates the emphasis of multiple channels simultaneously [29]. The weight W can be denoted as:

$$W = \text{Sigmoid}(\text{Conv}(\text{GAP}(X))) \quad (6)$$

At the same time, F_{tr} operation transforms feature X again to obtain a richer representation \hat{X} . The channel attention weights W is multiplied by \hat{X} to obtain the re-weighted feature representation, thereby enhancing the overall quality of the features. Finally, the output Y is derived by adding the residuals x_2 of the decoder to the re-weighted feature representation:

$$Y = W \otimes \delta(\text{Conv}(\hat{X})) + x_2 \quad (7)$$

Where \otimes denotes element-wise multiplication. This output Y is then fed back to the decoder as input for the next level of AGAM.

3.4 TEB and AEB

The transmission map provides valuable information about the haze density, which helps to perform proper image dehazing. The densely connected encoder-decoder network in DCPDN [8] is used as the transmission map estimation branch (TEB). The dense blocks used in the branch maximize the flow of information along these features and ensure better convergence by connecting all layers.

The atmospheric light is also a factor affecting the network dehazing output. We choose the network structure in IPUDN [9] as the atmospheric light estimation branch (AEB). This branch calculates three atmospheric light values corresponding to each color channel. Inspired by the idea of emphasizing high-intensity pixel estimation of atmospheric light in DCP [1], the branch uses a global max-pooling layer to better estimate atmospheric light.

3.5 Loss function

The network output of the end-to-end model may contain artifacts such as over-smoothing, lack of contrast, or traces of convolutional operations. To obtain more realistic dehazed images, we propose a combined loss function consisting of the following:

Edge-preserving loss The loss function of the existing dehazing network constructs to treat all pixels in the output image equally, which may lose edge details. Some image dehazing counts use the gradient of the image to characterize the edge information with satisfying results. For example, Bai et al. [31] presents the gradient information with the conventional Sobel operator. However, the classical Sobel operator can only detect a limited number of edge

directions and has a low noise immunity. For this purpose, we construct the edge retention loss as the sum of gradients in eight directions 0° , 45° , 90° , 135° , 180° , 225° , 270° , and 315° , which can detect edges in multiple directions, defined as :

$$L_{edge} = \frac{1}{N} \sum_{i=1}^N \|\nabla f(I_i) - \nabla J_i\|_1 \quad (8)$$

∇ is the image gradient operator.

Reconstruction loss To ensure that the output image of the network is close to the ground truth, we use the L_1 loss as the reconstruction loss. Given an input hazy image I_i , the network output is $f(I_i)$, and the ground truth is J_i . Then the L_1 loss for N samples can be expressed as:

$$L_{res} = \sum_{i=1}^N \|f(I_i) - J_i\|_1 \quad (9)$$

It measures the distortion between the dehazed image and the ground truth in the image pixel space.

SSIM loss To make the network produce visually pleasing results, we use structural similarity loss. The structural similarity between the two images can be as follows:

$$ssim(f(I), J) = \frac{2\mu_{f(I)}\mu_J + C_1}{\mu_{f(I)}^2 + \mu_J^2 + C_1} \cdot \frac{2\sigma_{f(I)J} + C_2}{\sigma_{f(I)}^2 + \sigma_J^2 + C_2} \quad (10)$$

where $\mu_{f(I)}$ and $\sigma_{f(I)}^2$ are the mean and variance of $f(I)$, respectively. $\sigma_{f(I)J}$ is the covariance of $f(I)$ and J . C_1 and C_2 are constants used to maintain stability. SSIM takes values from 0 to 1. SSIM loss is given by:

$$L_{SSIM} = 1 - ssim(f(I), J) \quad (11)$$

Total loss Based on the above considerations, we combine the edge-preserving loss, reconstruction loss and SSIM loss to regularize the dehazing network, and the total loss defines as:

$$L = \lambda_1 L_{edge} + \lambda_2 L_{res} + \lambda_3 L_{SSIM} \quad (12)$$

where λ_1 , λ_2 , and λ_3 are the weight parameters for balancing the different losses.

4 Experiments

This section presents the experimental setting, including the datasets, evaluation metrics, and implementation details. We evaluate the haze removal effectiveness of the method qualitatively and quantitatively, then compare it with other state-of-the-art methods. In addition, the proposed method for the ablation studies gives in this section for a more in-depth analysis.

4.1 Experimental setting

We will introduce the selected dataset, evaluation metrics, and implementation details in this part separately.

4.1.1 Dataset

It is difficult to obtain simultaneously hazy scene images with their ground truth counterparts in the real world. Therefore, we usually need to train the network in a supervised method by mainly synthetic hazy datasets. We train our FAM-Net on the large synthetic dataset, Haze4k [33], and test on the Haze4k test set. The train set of Haze4k contains 3000 hazy images with ground truth images, and the test set contains 1000 hazy images with ground truth images. In addition, to further verify the dehazing effect of our method on real-world hazy images, we select the real outdoor benchmark dataset BeDDE [35]. This dataset collected 208 images from 23 cities in China under different weather conditions.

4.1.2 Evaluation metrics

To evaluate the performance of our network, we used three reference metrics, peak signal to noise ratio (PSNR), structural similarity index (SSIM), and learned perceptual image patch similarity (LPIPS) [36]. And two non-reference metrics, the visibility index (VI) and the realness index (VI) [35], respectively. These metrics are usually used in haze removal tasks to evaluate image quality criteria.

4.1.3 Implementation details

For the training of our FAM-Net, the Adam optimizer implements an initial learning rate $\alpha = 10^{-4}$. The ADAM algorithm [37] with default parameter settings uses as the optimizer. In addition, the Cosine Annealing strategy used in [38] uses to adjust the learning rate of FAM-Net training. We set the epoch to 100 and the initial learning rate to e^{-4} . In particular, for clear and hazy images, each image was cropped to a size of 512×512 . In addition, the loss function parameters in (12) set to $\lambda_1=0.1$, $\lambda_2=1$, $\lambda_3=0.5$, respectively. In addition, the network was trained and tested on Ubuntu 18.0.3 LTS using Pytorch 0.4.1 with one NVIDIA Tesla P40 GPU and batch size set to 10.

4.2 Comparisons with state-of-the-art methods

We test the proposed method on synthetic hazy images and real-world hazy images, compared it qualitatively and quantitatively with several state-of-the-art image dehazing methods. We also show in Table 1 the number of parameters, computation, and runtime of the network, using data obtained from 512×512 input RGB images.

Results on synthetic datasets In this section, we compared FAM-Net with other state-of-the-art methods on the Haze4k test set. We selected three evaluation metrics: PSNR, SSIM, and LPIPS, and the quantitative evaluation results are presented in Table 1. The comparison shows that our method achieves the best performance on all metrics, which proves the effectiveness of our feature aggregation and modulation network for haze removal. For a better visual comparison of different models, we present the dehazed images in Figs. 5 and 6,

Table 1 Quantitative comparison with the state-of-the-art methods on Haze-4k dataset

Method	Publication	PSNR↑	SSIM↑	LPIPS↓	Params(M)	Flops(GMac)	Runtime(ms)
DCP [1]	TPAMI'11	16.93	0.853	0.126	—	—	—
AOD-Net [18]	ICCV'17	17.21	0.831	0.107	0.0017	0.47	1.00
GCANet [12]	WACV'19	24.10	0.925	0.059	0.69	60.89	68.15
DM ² -F-Net [32]	ICCV'19	24.57	0.943	0.032	92.14	187.57	36.32
MSBDN-DFF [22]	CVPR'20	22.99	0.850	—	31.35	78.56	—
FFA-Net [3]	AAAI'20	26.97	0.950	—	4.46	1153.34	200.64
DMT-Net [33]	ACMMM'21	28.53	0.960	—	51.79	—	—
Yu et al. [34]	CVPR'21	28.70	0.971	0.025	49.35	78.37	115.07
SGID-PFF [31]	TIP'22	25.90	0.965	0.033	13.87	611.91	183.59
DeHamer [15]	CVPR'22	28.30	0.963	0.031	132.40	—	30.83
Ye et al. [30]	ECCV'22	29.88	0.978	0.016	—	—	—
FAM-Net	Ours	31.27	0.981	0.014	45.18	328.48	16.67

FAM-Net performs favorably against all other methods in all metrics, where ↑ means the higher the better, and ↓ means the lower the better

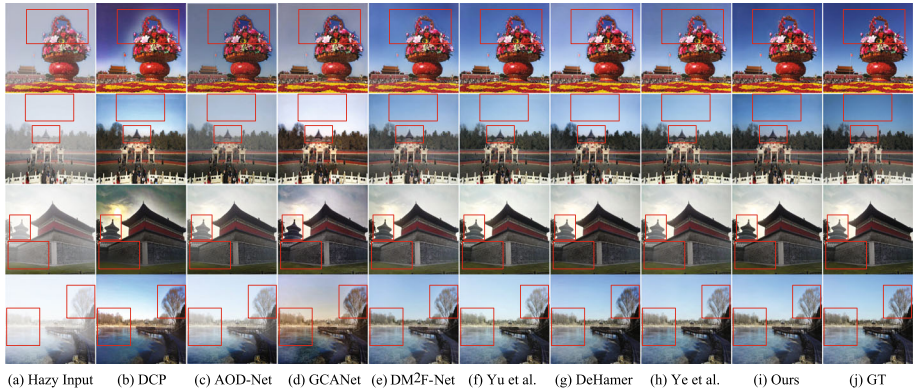


Fig. 5 Visual comparison of the dehazing effect of various methods on synthetic outdoor-hazy images, these images from the Haze4k test dataset. Zooming in on the image area, will show the effectiveness of our method

corresponding to outdoor and indoor scenes, respectively. Comparing the different methods in Fig. 5, it is evident that our method produces the least amount of haze. It also exhibits good haze removal ability while maintaining rich details and color information. Notably, for the sky area of each outdoor image, existing methods fail to deliver ideal results and cause varying degrees of color distortion, as evidenced in the images in the first and second rows. In contrast, our method generates output that is closest to the true value of the ground truth. In addition, for the indoor hazy images in Fig. 6, most of the existing methods achieve good results due to the scene's homogeneity compared to outdoor images. However, our method is capable of recovering the most realistic image possible, even in such scenarios. For example, for the recovery of the wall in the third row, many methods generate images with artifacts except for ours. Additionally, in the first row, our method retrieves the texture details of the desktop that are closest to the ground truth. We attribute this to the proposed edge-preserving loss. Overall, for the recovery of synthetic hazy images, both outdoor and indoor, our method achieves the best visual results.

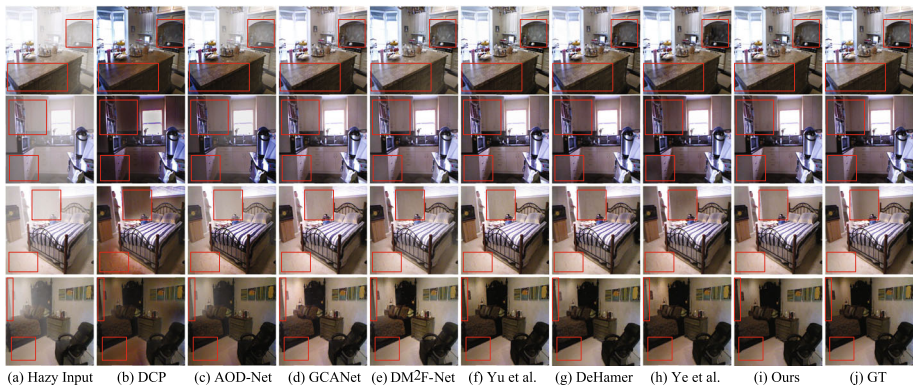


Fig. 6 Visual comparison of the dehazing effect of various methods on synthetic indoor-hazy images, these images from the Haze4k test dataset. Zooming in on the image area, such as the cropped area in the box, will show the effectiveness of our method

Results on the real-world datasets To further verify the dehazing effect of different methods on real-world hazy images, we used VI and RI for the quantitative evaluation of the dehazing effect of 167 images from BeDDE, as shown in Table 2. Our FAM-Net achieved the second best result after DCP in VI and the best result in RI. Meanwhile, we selected four representative real hazy images as samples. Figure 7 depicts the comparison with state-of-the-art methods for realistic image dehazing. Similar to synthetic image dehazing, other methods have more residual haze in the dehazing output and different degrees of artifacts. From the visual comparisons, DCP [1] shows constant color distortion in color haze images due to inappropriate priors. MSCNN [39], DehazeNet [2], AOD-Net [18], and DeHamer [15] retain some dehazing capability but still show slight color distortion that darkens the scene. The results of DCPDN [8], Yu et al. [34], and Ye et al. [30] are relatively good visually, but there are some haze residue, and the recovery ability for object edge details is not as good as that of our method. Overall, the results of our proposed method are of the highest fidelity, which is perceptually closer to the real-world. Although our method is trained on synthetic dataset only, it still obtains desirable dehazing results on real-world hazy images, proving that our model is robust.

In Figs. 5, 6 and 7, we find that the images generated by FAM-Net have great dehazing effects, in which few local artifacts are observed. We believe that this may be a result of physical feature guidance.

4.3 Ablation study

This subsection introduces some ablation experiments to illustrate the utility of different constructions and loss function. These experiments are all conducted on the Haze4k dataset used previously. The detailed experiments that will be discussed can be found as follows.

Table 2 Quantitative comparison with the state-of-the-art methods on BeDDE dataset

Method	Publication	VI↑	RI↑
DCP [1]	TPAMI'11	0.90466	0.96294
Nishino et al. [40]	ICCV'12	0.82582	0.93419
CAP [6]	TIP'15	0.84971	0.94812
NLD [41]	CVPR'16	0.82188	0.95520
MSCNN [39]	ECCV'16	0.88190	0.96667
DehazeNet [2]	TIP'16	0.87938	0.96910
AOD-Net [18]	ICCV'17	0.88785	0.96780
GFN [11]	CVPR'18	0.85380	0.96188
DCPDN [8]	CVPR'18	0.88445	0.96891
Yu et al. [34]	CVPR'21	0.89145	<u>0.96918</u>
SGID-PFF [31]	TIP'22	0.87838	0.96845
Ye et al. [30]	ECCV'22	0.83771	0.96465
DeHamer [15]	CVPR'22	0.87825	0.96787
FAM-Net	Ours	<u>0.89507</u>	0.97044

↑ means the higher the better. **Bold** and underlined indicate the best and the second best results, respectively

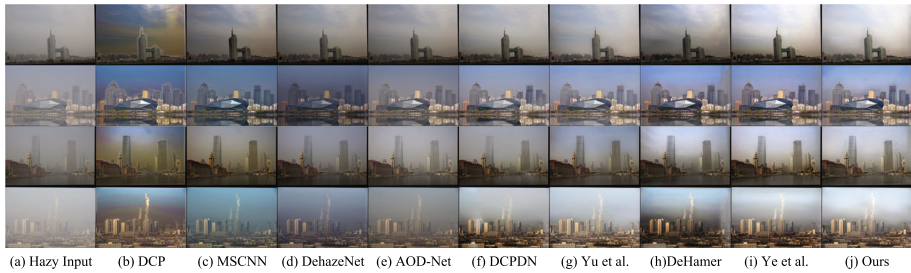


Fig. 7 Visual comparison of the dehazing effect of various methods on real-world hazy images from the BeDDE [35] dataset

4.3.1 Contribution of each core part

In this section, we conducted an extensive ablation study. We employed three different model variations to validate the effectiveness of each component of the proposed FAM-Net. The experiments that discussed in detail are shown in Table 3, which reports quantitative results, including PSNR, SSIM, and LPIPS performance.

Baseline The designed Baseline is a joint dehazing network based on an atmospheric scattering model. It comprises an encoder-decoder main path to learn the haze-free features, as well as TEB and AEB to estimate the transmission map and global atmospheric light, respectively. At the end of the network, the dehazed image is obtained using (2). As shown in Table 3, the dehazing effect of Baseline has surpassed many existing methods. This suggests that the end-to-end deep network joint estimation of the dehazed image, along with a suitable physical feature guidance mechanism, is beneficial for haze removal. As shown in Fig. 8, the hazy image processed by Baseline has more haze left.

Baseline + FAMM According to [23], a process-oriented aggregation approach can prevent suboptimal dehazing. Therefore, unlike the Baseline, our proposed network does not employ a result-oriented concatenation at the end of the network. Instead, we introduced the proposed FAMM by integrating it into the encoder network to aggregate and modulate the outputs of the main path, TEB and AEB to obtain the final dehazing results. As shown in Table 3, FAMM plays a crucial role in improving the output results. The three evaluation metrics improved by 1.61, 0.005, and 0.0092.

Baseline + FAMM + AGAM To evaluate the effectiveness of the AGAM method, we compared it to the commonly used concatenation method. The results demonstrate that the AGAM method outperforms the concatenation method in terms of network performance, with PSNR

Table 3 Ablation study of each core part and mechanism used in our proposed networks

Ablation study variant	Settings			Haze4k		
	Baseline	FAMM	AGAM	PSNR	SSIM	LPIPS
1	✓			29.02	0.976	0.0228
2	✓	✓		30.63	0.981	0.0136
3	✓	✓	✓	31.27	0.981	0.0135

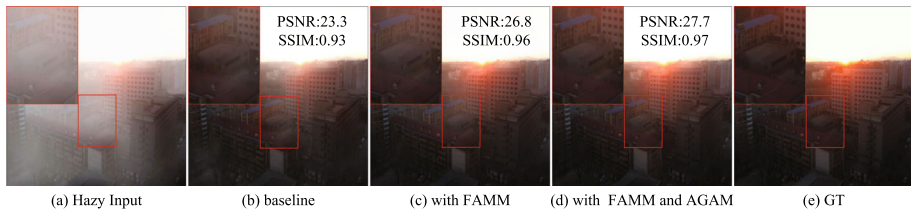


Fig. 8 Visual of each core part and mechanism used in our proposed networks. We have enlarged the selected area and placed it in the top left corner

improvement of 0.64. These results suggest that the AGAM method can be a useful alternative to the concatenation method for improving the performance of encoder-decoder architectures. As shown in Fig. 8, the dehazed image has the least haze residue and is closest to the ground truth.

4.3.2 Impact of the progressive FAMMs

We conducted a comprehensive ablation study of the progressive aggregation modulation design process. The study comprises two parts: a comparison between result-oriented and process-oriented concatenation of FAMMs, and a comparison between a single FAMM and multiple progressive FAMMs.

The result-oriented concatenation requires embedding the FAMM at the end of the network and inputting the output of TEB, AEB, and encoder-decoder together into the FAMM to obtain the output dehazed image. On the other hand, process-oriented concatenation involves embedding the FAMM into the encoder, as illustrated in Fig. 1, to aggregate and modulate the physical features. As can be seen from the first two rows of Table 4, process-oriented concatenation achieved better results than result-oriented concatenation when only one FAMM was used.

Furthermore, utilizing process-oriented concatenation, we conducted a detailed ablation study on the progressive aggregation modulation design process. We used three combinations of FAMM designs to analyze the models performance relative to aggregation and modulation. The performance evaluation with the different numbers of FAMM combinations is presented in Table 4. The quantitative experimental results show that the progressive feature aggregation and modulation of multiple FAMMs yield better results than single FAMM. Among the various progressive FAMMs designed in this section, the combination of two FAMMs is most suitable for our network.

Table 4 Impact of progressive FAMMs on FAM-Net based image dehazing

Number of progressive FAMM	Concatenation	Haze4k		
		PSNR	SSIM	LPIPS
One	Result-oriented	28.66	0.970	0.0249
One	Process-oriented	31.09	0.980	0.0138
Two	Process-oriented	31.27	0.981	0.0135
Three	Process-oriented	31.08	0.981	0.0137

Table 5 Impact of loss function on FAM-Net based image dehazing

Different combinations	loss function					Haze4k	
	L_{res}	L_{SSIM}	$L_{edge}(V1)$	$L_{edge}(V2)$	$L_{edge}(V3)$	PSNR	SSIM
1	✓					30.53	0.973
2	✓	✓				30.96	0.979
3	✓	✓	✓			31.05	0.980
4	✓	✓		✓		31.06	0.980
5	✓	✓			✓	31.27	0.981

4.3.3 Impact of loss function

In this subsection, we conduct an extensive ablation study to validate the effectiveness of the proposed loss function.

Loss function Inspired by the edge feature information to improve the performance of the dehazing network, FAM-Net adds a third loss term, L_{edge} , to (12) compared to the conventional loss function based on MSE and SSIM. The quantitative experimental results in Table 5 show that FAM-Net obtains the highest PSNR and SSIM values for the loss function $L=L_{res}+0.5L_{SSIM}+0.1L_{edge}$, indicating the improvement of the image dehazing effect in using L_{edge} with a suitable λ . Figure 9 shows the dehazed results, which indicate that the method without L_{edge} is less effective than the proposed method and that its result suffers from some residual haze and poor preservation of objects' edges.

Edge-preserving loss As discussed in Section 3.5, we apply a novel edge-preserving loss function for haze removal. We perform an ablation study using FAM-Net. Table 5 shows the experimental results without the edge-preserving loss function and with three different designs of the edge-preserving loss function.

- **Version 1** uses traditional vertical and horizontal edge detection.

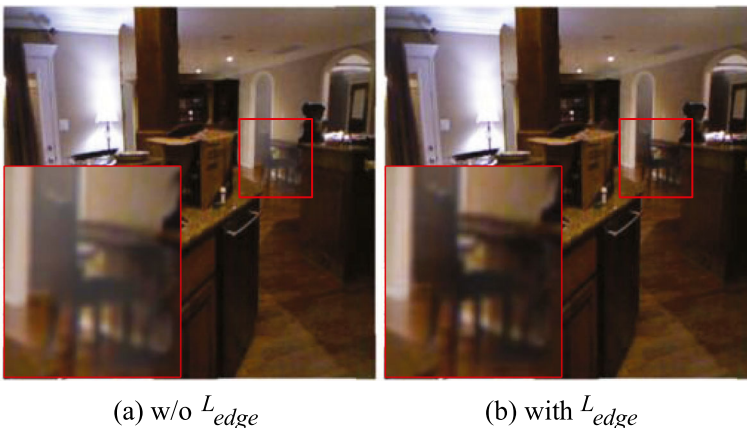


Fig. 9 We have enlarged the selected area and placed it in the lower left corner. The method without L_{edge} is less effective than the proposed method

Table 6 Average PSNR and SSIM values of FAM-Net on validation set under different hyperparameter settings in (12)

λ_1	λ_2	λ_3	PSNR	SSIM
0.1	1	0.5	33.26	0.985
0.1	1	1	33.12	0.985
1	1	0.5	33.24	0.984

- **Version 2** adds two other edge detection in the diagonal direction to version 1.
- **Version 3** goes a step further and uses Sobel operators in eight directions of 0° , 45° , 90° , 135° , 180° , 225° , 270° , and 315° for edge detection.

Based on the experimental results, it was found that using the edge-preserving loss function can improve the network's performance. Among the three proposed designs, the one that yielded the best results was constructing the edge-preserving loss function using eight directions of the Sobel operators to calculate the gradient. Alternatively, using two or four directional Sobel operators resulted in decreased performance.

Hyperparameter To explore the impact of hyperparameters in (12) on the network's performance, we calculated the average PSNR and SSIM for different loss weights on FAM-Net, as presented in Table 6. For this analysis, FAM-Net underwent parameter optimization on the validation set, where approximately 1/10th of the data from the train set was randomly selected for validation. It is essential to note that we chose the L_1 loss as the primary criterion for image reconstruction, and following the guidance from existing methods [17, 31], we did not require additional hyperparameter balancing for the reconstruction loss. Thus, we set λ_2 to a fixed value of 1 and proceeded to explore various combinations of λ_1 and λ_3 to weigh the respective loss functions. Subsequently, we performed three sets of comparison tests.

The results presented in Table 6 indicate that the hyperparameter combinations of 0.1, 1, and 0.5 yielded the most favorable outcomes in terms of both PSNR and SSIM. Consequently, we adopted these specific parameter combinations to achieve a balanced approach to losses in each part during this experiment.

5 Conclusion

This study proposes an effective feature aggregation and modulation network (FAM-Net) method for haze removal. Our method can preserve texture and edges of the image better and significantly improve the overall image quality. In FAM-Net, FAMMs, AGAM, TEB, and AEB are introduced to the network to effectively suppress image content loss. Our proposed progressive FAMMs exploit deep and non-adjacent layers of physical features generated by hazy images and modulate them, which are aggregated with clear regions to obtain complete dehazed images. AGAM replaces the original skip connection to improve the dehazing effect. Extensive experiments on standard datasets demonstrate the superiority of our method over some representative haze removal methods.

Acknowledgements This work was supported by the National Natural Science Foundation of China (Grant No. 12075090), and by the Funding by Science and Technology Projects in Guangzhou (Grant No. 2023A04J1686), and by the GuangDong Basic and Applied Basic Research Foundation (Grant No. 2022A1515110119).

Declarations

Competing interests The authors have no competing interests to declare that are relevant to the content of this article.

References

1. He K, Sun J, Tang X (2010) Single image haze removal using dark channel prior. *IEEE Trans Pattern Anal Mach Intell* 33(12):2341–2353
2. Cai B, Xu X, Jia K et al (2016) Dehazenet: An end-to-end system for single image haze removal. *IEEE Trans Image Process* 25(11):5187–5198
3. Qin X, Wang Z, Bai Y et al (2020) Ffa-net: Feature fusion attention network for single image dehazing. In: *Proceedings of the AAAI conference on artificial intelligence*, pp 11908–11915
4. McCartney EJ (1976) *Optics of the atmosphere: scattering by molecules and particles*. New York
5. Narasimhan SG, Nayar SK (2000) Chromatic framework for vision in bad weather. In: *Proceedings IEEE conference on computer vision and pattern recognition. CVPR 2000 (Cat. No. PR00662)*, IEEE, pp 598–605
6. Zhu Q, Mai J, Shao L (2015) A fast single image haze removal algorithm using color attenuation prior. *IEEE Trans Image Process* 24(11):3522–3533
7. Berman D, Treibitz T, Avidan S (2017) Air-light estimation using haze-lines. In: *2017 IEEE International conference on computational photography (ICCP)*, IEEE, pp 1–9
8. Zhang H, Patel VM (2018) Densely connected pyramid dehazing network. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp 3194–3203
9. Kar A, Dhara SK, Sen D et al (2020) Progressive update guided interdependent networks for single image dehazing. [arXiv:2008.01701](https://arxiv.org/abs/2008.01701)
10. Dong J, Pan J (2020) *Physics-based feature dehazing networks*. In: *European conference on computer vision*, Springer, pp 188–204
11. Ren W, Ma L, Zhang J et al (2018) Gated fusion network for single image dehazing. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp 3253–3261
12. Chen D, He M, Fan Q et al (2019) Gated context aggregation network for image dehazing and deraining. In: *2019 IEEE winter conference on applications of computer vision (WACV)*, IEEE, pp 1375–1383
13. Zhao S, Zhang L, Shen Y et al (2021) Refinednet: A weakly supervised refinement framework for single image dehazing. *IEEE Trans Image Process* 30:3391–3404
14. Liu X, Ma Y, Shi Z et al (2019) Griddehazenet: Attention-based multi-scale network for image dehazing. In: *Proceedings of the IEEE/CVF international conference on computer vision*, pp 7314–7323
15. Guo CL, Yan Q, Anwar S et al (2022) Image dehazing transformer with transmission-aware 3d position embedding. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp 5812–5820
16. Ren W, Pan J, Zhang H et al (2020) Single image dehazing via multi-scale convolutional neural networks with holistic edges. *Int J Comput Vis* 128(1):240–259
17. Pang Y, Xie J, Li X (2018) Visual haze removal by a unified generative adversarial network. *IEEE Trans Circ Syst Vid Tech* 29(11):3211–3221
18. Li B, Peng X, Wang Z et al (2017) Aod-net: All-in-one dehazing network. In: *Proceedings of the IEEE international conference on computer vision*, pp 4770–4778
19. Zhu H, Peng X, Chandrasekhar V et al (2018) Dehazegan: When image dehazing meets differential programming. In: *IJCAI*, pp 1234–1240
20. Zhang J, Tao D (2019) Famed-net: A fast and accurate multi-scale end-to-end dehazing network. *IEEE Trans Image Process* 29:72–84
21. Riaz I, Yu T, Rehman Y et al (2016) Single image dehazing via reliability guided fusion. *J Vis Commun Image Represent* 40:85–97
22. Dong H, Pan J, Xiang L et al (2020) 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
23. Wang C, Shen HZ, Fan F et al (2021) Eaa-net: A novel edge assisted attention network for single image dehazing. *Knowl-Based Syst* 228:107279
24. Zhang X, Wang T, Wang J et al (2020) Pyramid channel-based feature attention network for image dehazing. *Comput Vis Image Underst* 197:103003

25. Wang T, Zhao L, Huang P et al (2021) Haze concentration adaptive network for image dehazing. *Neuro-computing* 439:75–85
26. Xiao B, Zheng Z, Zhuang Y et al (2022) Single uhd image dehazing via interpretable pyramid network. Available at SSRN 4134196
27. Wang Z, Cun X, Bao J et al (2022) Uformer: A general u-shaped transformer for image restoration. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp 17683–17693
28. Song Y, He Z, Qian H et al (2023) Vision transformers for single image dehazing. *IEEE Trans Image Process* 32:1927–1941
29. Hu J, Shen L, Sun G (2018) Squeeze-and-excitation networks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp 7132–7141
30. Ye T, Zhang Y, Jiang M et al (2022) Perceiving and modeling density for image dehazing. In: Part XIX (ed) *Computer vision-ECCV 2022: 17th European conference, Tel Aviv, Israel, October 23–27, 2022, proceedings*. Springer, pp 130–145
31. Bai H, Pan J, Xiang X et al (2022) Self-guided image dehazing using progressive feature fusion. *IEEE Trans Image Process* 31:1217–1229
32. Deng Z, Zhu L, Hu X et al (2019) Deep multi-model fusion for single-image dehazing. In: *Proceedings of the IEEE/CVF international conference on computer vision*, pp 2453–2462
33. Liu Y, Zhu L, Pei S et al (2021) From synthetic to real: Image dehazing collaborating with unlabeled real data. In: *Proceedings of the 29th ACM international conference on multimedia*, pp 50–58
34. Yu Y, Liu H, Fu M et al (2021) A two-branch neural network for non-homogeneous dehazing via ensemble learning. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp 193–202
35. Zhao S, Zhang L, Huang S et al (2020) Dehazing evaluation: Real-world benchmark datasets, criteria, and baselines. *IEEE Trans Image Process* 29:6947–6962
36. Zhang R, Isola P, Efros AA et al (2018) The unreasonable effectiveness of deep features as a perceptual metric. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp 586–595
37. Kingma DP, Ba J (2014) Adam: A method for stochastic optimization. arXiv preprint [arXiv:1412.6980](https://arxiv.org/abs/1412.6980)
38. He T, Zhang Z, Zhang H et al (2019) Bag of tricks for image classification with convolutional neural networks. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp 558–567
39. Ren W, Liu S, Zhang H et al (2016) Single image dehazing via multi-scale convolutional neural networks. In: *European conference on computer vision*, Springer, pp 154–169
40. Nishino K, Kratz L, Lombardi S (2012) Bayesian defogging. *Int J Comput Vis* 98:263–278
41. Berman D, Avidan S et al (2016) Non-local image dehazing. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp 1674–1682

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.