



Knowledge-guided multi-perception attention network for image dehazing

Peng Liu¹ · Jianlei Liu¹

Accepted: 6 November 2023 / Published online: 2 December 2023

© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2023

Abstract

Image dehazing is an important computer vision task that aims to restore clear images from blurry, hazy images. Most of the existing deep learning dehazing methods are result-oriented, ignoring the intermediate steps and it is also difficult to deploy cumbersome deep models on devices with limited resources. In addition, the attention mechanisms based on convolution kernels within a fixed window size cannot provide additional flexibility for mapping from hazy images to clear images. This work presents a novel knowledge distillation method that guides the intermediate process of dehazing to improve the performance of image dehazing networks. Specifically, we train a teacher network on clear images, which can learn useful features from clear images (ground truth), and then we select the deep layers of the network, i.e., the decoding process, to transfer these features to a lightweight student network. Moreover, we design a multi-perception attention module and apply a heterogeneous design to this module for the teacher network and the student network to extract multiscale and multilevel features of hazy images, thus enhancing the expressive ability of the student network. We conduct experiments on several public image dehazing datasets, and the results show that our method achieves a good trade-off between reducing the parameter size and maintaining a high-quality dehazing effect compared with other algorithms.

Keywords Attention · Image dehazing · Knowledge guidance · Multi-perception

1 Introduction

Image dehazing is a fundamental task in computer vision and image processing that aims to recover the underlying clear image from a hazy image corrupted by atmospheric scattering. It has important practical applications in outdoor surveillance, autonomous driving, and remote sensing, where the visibility of the image is often impaired by haze or mist. However, image dehazing is a challenging problem due to the complex physical models underlying haze formation, which makes the problem ill-posed and nonlinear.

To model the physical process of haze formation, the atmospheric scattering model (ASM) was proposed [1, 2]. The ASM assumes that the observed hazy image is a linear combination of a clear image and atmospheric light attenuated by

the scattering medium. This can be represented mathematically as:

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

where $I(x)$ is the observed hazy image, $J(x)$ is the clear image, $t(x)$ is the transmission map, A is the atmospheric light, and x is the pixel coordinate.

Early dehazing methods focused on handcrafted prior-based methods such as the dark channel prior (DCP) [3], haze-line prior (HLP) [4] and color attenuation prior (CAP) [6]. The DCP utilizes prior knowledge to estimate $t(x)$ and A to obtain $J(x)$. However, these methods suffer from limitations such as oversimplification of the physical models. Therefore, these prior-based physical models do not perform well on real-world images. In addition, there are also some methods based on image fusion, such as AMEF [29], but they face problems with unsatisfactory dehazing effects.

Recently, deep learning-based methods have shown promising results in image dehazing by learning the underlying mapping between hazy and clear images. [7–9, 30, 31,

✉ Jianlei Liu
15269169075@163.com

¹ School of Cyber Science and Engineering, Qufu Normal University, Jining 273165, China

36–38] are some of the most representative deep learning-based methods. These methods use convolutional neural networks (CNNs) to learn the complex nonlinear mapping between hazy and clear images. Alternatively, they estimate the unknown variables in the atmospheric scattering model through the network, indirectly obtaining dehazed images. Other recent works have also explored various techniques for improving image dehazing, such as using generative adversarial networks (GANs) [16] and contrastive loss [24]. Although end-to-end deep learning-based methods have achieved great success, they mainly focus on the mapping between the input hazy image and the output haze-free image, without much consideration of the intermediate steps involved in the dehazing process. At the same time, there are difficulties in obtaining paired hazy and clear images, and the information of clear images is not fully explored. In addition, attention mechanisms designed only based on fixed-size convolutional kernels and single perception cannot endow the network with more flexibility, so they are unfavorable for learning the mapping from hazy images to clear images.

At the same time, as models become more cumbersome and complex, it is difficult to deploy these deep models on devices with limited resources. Knowledge distillation can effectively learn a small student model from a large teacher model, which provides the possibility of deploying these models on resource-limited devices such as mobile phones and embedded devices. According to the survey [39], the learning of student networks in knowledge distillation is influenced by three key factors: knowledge types, distillation strategies and teacher-student architectures. Knowledge types refer to the different kinds of information that the teacher model can provide to the student model, which can be categorized into response-based knowledge, feature-based knowledge and relation-based knowledge. Distillation strategies can be categorized into offline distillation, online distillation, and self-distillation, based on whether the teacher model and the student model are updated simultaneously. The teacher-student structure serves as the fundamental framework for knowledge distillation, playing a crucial role in the acquisition and transfer of knowledge. The selection or design of suitable teacher-student structures remains an area that requires further exploration.

Knowledge distillation was formally proposed by Hinton et al. [20] and has since been progressively applied to diverse domains, including image classification [40], face recognition [41], object detection [42], and even image dehazing [43–47]. Regarding image dehazing, [43] employs knowledge distillation to transfer knowledge from a pre-trained dehazing network (RefineDNet) to an unsupervised learning branch, resulting in an enhanced dehazing effect. [44] proposed a single image dehazing network that combines physical model and self-distillation, effectively utilizing the advantages of model-based and model-free methods. [45]

utilizes transmission map information to guide the image transformation process and leverages features extracted from the transformed clear images to aid in transmission map estimation. [46] proposed an online knowledge distillation network (OKDNet) for single image dehazing, which combines the advantages of model-based and model-free dehazing methods through online knowledge distillation, achieving excellent dehazing effect. [47] proposed an image dehazing network based on multiple priors and offline knowledge distillation, namely dark channel prior and non-local dehazing prior, to guide the learning and optimization of dehazing network.

Based on the above analysis, we propose a knowledge-guided multi-perception attention network (KMAN) for image dehazing. Overall, we employ an autoencoder-like structure, combining network deconvolution (ND) and shallow and deep feature fusion (SDF) modules to form the basic architecture of the network [22]. By incorporating the concept of knowledge distillation, we not only focus on the output results of the network but also use the teacher network to provide guidance information for the image decoding process of the student network. To give the network more flexible adjustment abilities, we design a multi-perception attention mechanism. We also use perceptual loss and contrastive loss as regularization terms, which constrain the training process in multiple respects and enhance the robustness of the model.

In summary, our contributions are as follows:

- (1) We propose a knowledge-guided multi-perception attention network for image dehazing. We train a teacher network that learns from clear images to guide the decoding process of the student network, leveraging the strong prior knowledge of the teacher network to help gradually restore hazy images to haze-free images.
- (2) We employ a multi-perception attention (MPA) module that fuses information from different receptive fields and combines channel-level global and local information. Through adaptive learning, MPA assigns higher weights to more important features.
- (3) Our method (KMAN) provides a new paradigm for designing teacher-student network architectures for image dehazing, while achieving a good trade-off between reducing the parameter size and maintaining high-quality dehazing results, which may provide satisfactory performance on platforms with limited computing resources (such as mobile devices).

2 Related work

In this section, we review related work focusing on two main topics: attention mechanisms (channel, spatial) and

their application in image dehazing, and knowledge distillation and its application in image dehazing.

2.1 Attention mechanism

The attention mechanism is a method that strongly focuses on certain features while ignoring other irrelevant features in deep learning models. In recent years, attention mechanisms have achieved significant results in computer vision tasks. They are mainly divided into channel attention, spatial attention, and self-attention [17] mechanisms. We will introduce two types of attention here: channel attention and spatial attention.

Channel Attention Channel attention mainly focuses on learning the correlation between different channels. Hu et al. proposed a channel attention module called SENet [18], which improves the model performance by learning the dependency between channels.

Spatial Attention The spatial attention mechanism emphasizes the importance of local features in images. Wang et al. proposed a spatial attention module called a nonlocal neural network [19], which can capture long-distance dependencies in images.

Attention-based methods aim to selectively attend to the most informative regions of the input hazy image to guide the dehazing process. Zhang et al. proposed a densely connected pyramid dehazing network (DCPDN) [9], which incorporates a multiscale and multilevel perceptual attention module to extract and fuse multiscale and multilevel features of hazy images. SADnet [13] combines channel attention, spatial attention, and self-attention to propose a channel-spatial self-attention (CSSA) mechanism. MSAFF-Net [14] uses channel attention and spatial attention of different scales to achieve feature fusion, while MARG-UNet [15] combines channel attention and pixel attention to propose a multimodal attention residual group (MARG) as one of the basic modules of the UNet structure to complete the dehazing task. These methods have achieved good results in their applications.

Various attention mechanisms endow the feature map adjustment with more flexibility. Image dehazing, due to the complexity of the scene, is still a very challenging problem. To improve the flexibility and perceptual ability of the pixel adjustment process taking hazy images to clear images, we propose a multi-perception attention module, which is used for learning the mapping from the latent features of hazy images to the latent features of clear images at the bottleneck of the network.

2.2 Knowledge distillation

Knowledge distillation is a method of training a small model to mimic the performance of a large model. It trains a small

model (called a student model) by transferring the knowledge of a large model (called a teacher model) [20]. Knowledge distillation has achieved success in many computer vision tasks, such as image classification and object detection [21].

In recent years, the common dehazing algorithms based on knowledge distillation are mainly divided into two categories, one is online distillation and the other is offline distillation. Self-parameter distillation [11] can be regarded as a special form of online distillation, where the teacher network and the student network are the same network. It first uses haze-free images to train the network and extract scene content features. Then, it uses hazy images to continue training the network, using the previously learned scene content information to remove the haze. In the training process, it uses a parameter interpolation method to perform self-distillation. The other category is the offline distillation method, such as [10, 12], which trains a teacher network with clear images and aligns the features at the bottleneck of the teacher network and the student network when training the dehazing student network. Whether self-parameter distillation or alignment of features at the bottleneck is used, these methods lack sufficient guidance for the image decoding process near the output layer.

In our method, the student and teacher networks are designed heterogeneously, where the more complex teacher network learns the mapping between clear images and provides hierarchical guidance for the student network in the decoding process to obtain the dehaze image. The feature information at different levels of the clear image will provide key information guidance in the process of reconstructing the dehazed image in the student network, which enables the student network to gradually produce the restored image with the teacher's guidance rather than directly producing it, and more easily restore the clear dehazed image.

3 Method

3.1 Architecture

In this section, we primarily introduce the overall structure of the network and several key components, including knowledge guidance, two designs for the multi-perception module, and the loss function.

Knowledge Guidance If a network can effectively reconstruct clear images, we consider it to have learned the information and structure of clear images, and it can provide guidance to the dehazing network by transferring its learned prior knowledge and feature representations. Based on this, we design a dual-network dehazing structure, both networks of which use a pyramid encoding–decoding framework [22], as shown in Fig. 1. First, we train a teacher network using

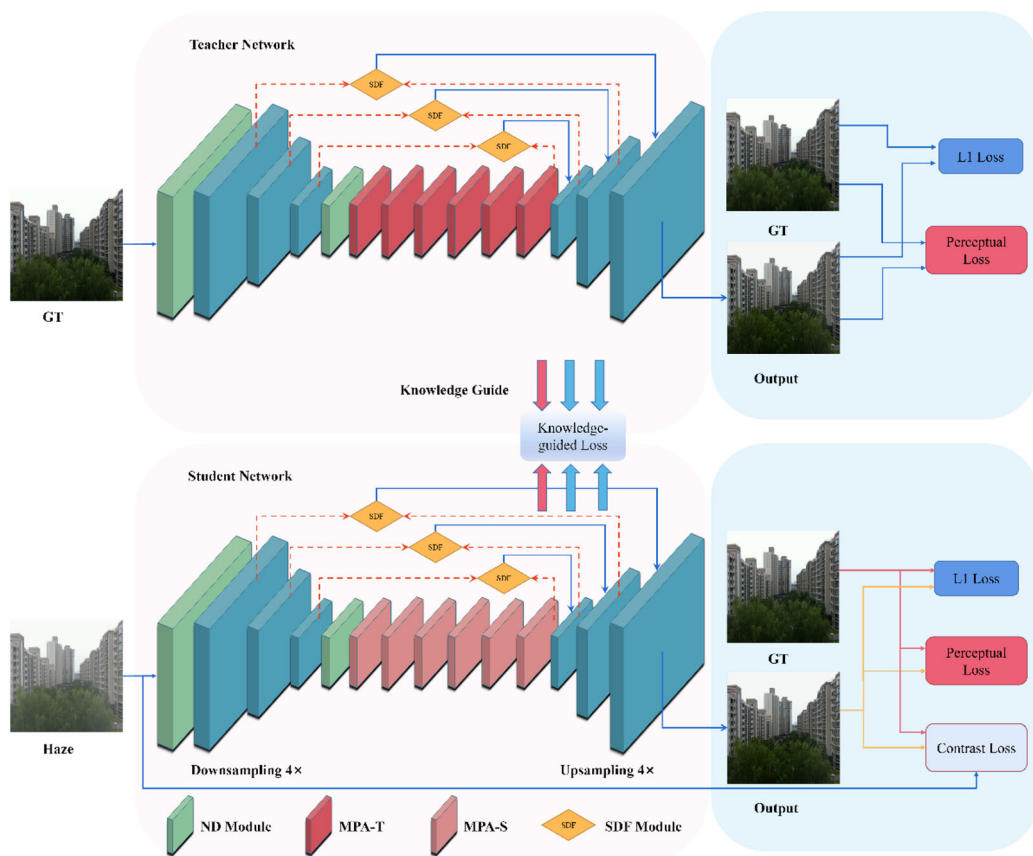


Fig. 1 The network architecture of the proposed KMAN

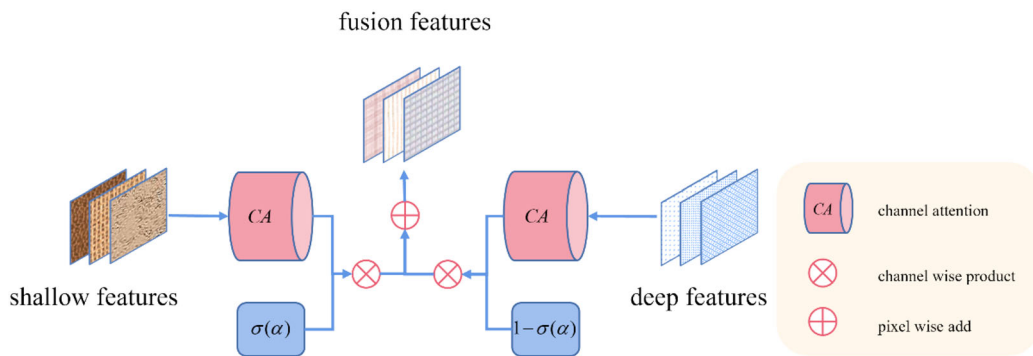


Fig. 2 The structure of the SDF module, where α is a learnable parameter

clear images to assist the dehazing network in achieving better haze removal. Then, we train a dehazing student network that is guided step-by-step by the teacher network during the decoding process to obtain clear images. Specifically, we use the pyramid encoder–decoder structure with the shallow and deep feature fusion (SDF) module (as shown in Fig. 2) [22] and the network deconvolution (ND) module [5] as the basic framework of the network. At the bottleneck of the network, we design a multi-perception attention module as a fundamental component for mapping between the latent space

representations of hazy and clear images. In terms of the loss function, in addition to the knowledge-guided regularization loss, the student network adds a contrastive learning regularization loss. The two networks also have different specific tasks: the input of the teacher network is clear images, and the focus is on providing knowledge guidance for the haze removal network; the input of the student network is hazy images, and the focus is on haze removal tasks.

Multi-perception Attention Module Haze in an image leads to decreased visibility and loss of details. To adapt to the

complexity of image dehazing scenarios, it is necessary to effectively capture and utilize information from different levels and scales of the image. Therefore, we propose the multi-perception module, which utilizes dilated convolutions to capture information from different receptive fields. Additionally, it employs channel wise average pooling and max pooling to capture global and local information, respectively. This enables the algorithm to better understand the structure and content of the scene, thereby improving the dehazing results.

The MPA modules in the two networks have slightly different structures, with the multi-perception attention module in the student network (MPA-S) being lighter than the multi-perception attention module in the teacher network (MPA-T). This design choice results in a more lightweight model. The structure of the multi-perception attention module in the teacher network is illustrated in Fig. 3a. First, the features undergo parallel 1×1 convolutional transformations. Then, they pass through the multi-perception module and the channel attention module, and the results are multiplied together to obtain the final attention weights, att^t . These weights are used to attentively weight the original features x^t of the teacher network, producing the final output y_{mpa}^t . The entire process is depicted in Eqs. (2) and (3).

$$att^t = mp(conv_{1*1}(x^t)) * ca(conv_{1*1}(x^t)) \tag{2}$$

$$y_{mpa}^t = x^t * att^t \tag{3}$$

$mp(*)$ and $ca(*)$ in the equation correspond to the multi-perception module and channel attention module shown in Fig. 3. The structure of the multi-perception attention module in the student network is illustrated in Fig. 3b. First, the features undergo a 1×1 convolutional transformation. Then, they pass through the multi-perception module, resulting in the final attention weights, att^s . These weights are used to attentively weight the original features x^s of the student network after a 1×1 convolutional transformation, producing the final output y_{mpa}^s . The entire process is depicted in Eqs. (4) and (5).

$$att^s = mp(conv_{1*1}(x^s)) \tag{4}$$

$$y_{mpa}^s = conv_{1*1}(x^s) * att^s \tag{5}$$

The main difference between these two MPA modules is that the MPA module in the student network (MPA-S) does not employ channel attention as in the teacher network (MPA-T), which makes the attention weighting process more concise. This simplifies our network structure to some extent.

3.2 Loss function

The loss function consists of four components: the reconstruction loss L_{rec} , perceptual loss L_{perc} , contrastive loss $L_{contrast}$, and knowledge-guided loss L_{kg} , as shown in Eq. (6).

$$L_{total} = L_{rec} + \alpha L_{perc} + \beta L_{contrast} + \gamma L_{kg} \tag{6}$$

Reconstruction Loss We use L1 loss as the training loss function, as shown in Eq. (7). In the task of image dehazing, we aim to preserve more details and texture information in the image, avoiding color and texture distortion. Generally, compared to L2 loss, L1 loss has a lower sensitivity to outliers, making it more effective in preserving image details, including texture and color information. Moreover, [32] demonstrated that training with L1 loss achieves better performance in terms of the PSNR and SSIM metrics in many image restoration tasks.

$$L_{rec} = |y - D(I)|_1 \tag{7}$$

y represents the ground truth, I is the hazy image input to the network, and $D(*)$ represents the dehazing network.

Perceptual loss To obtain network outputs that are similar in style to the ground truth, we should not only focus on pixel differences between the network outputs and the ground truth but also take into account the differences in features and styles between them. Therefore, we introduce perceptual loss [23] as a regularization term in the image dehazing task to measure the similarity between the dehazed results and the real image features, as shown in Eq. (8).

$$L_{perc} = \frac{1}{C_j H_j W_j} \|\varphi_j(\hat{y}) - \varphi_j(y)\|_2^2 \tag{8}$$

\hat{y} represents the network's output dehazed image, y represents the ground truth, φ is a pretrained network (VGG16), and the subscript j denotes the j -th layer of the network, with the corresponding feature shape of $C_j \times H_j \times W_j$.

Contrastive Loss From contrastive learning, we can regard the hazy images as negative samples and the clear images as positive samples. To make the network output close to the positive samples and far from the negative samples, [24] proposed a contrastive regularization method, which uses the VGG network to extract deep features from the network output, hazy images, and corresponding clear images, as shown in Eq. (9). We introduce contrastive loss $L_{contrast}$ to the dehazing task as one of the regularization terms in the loss function, further enhancing the robustness of the model.

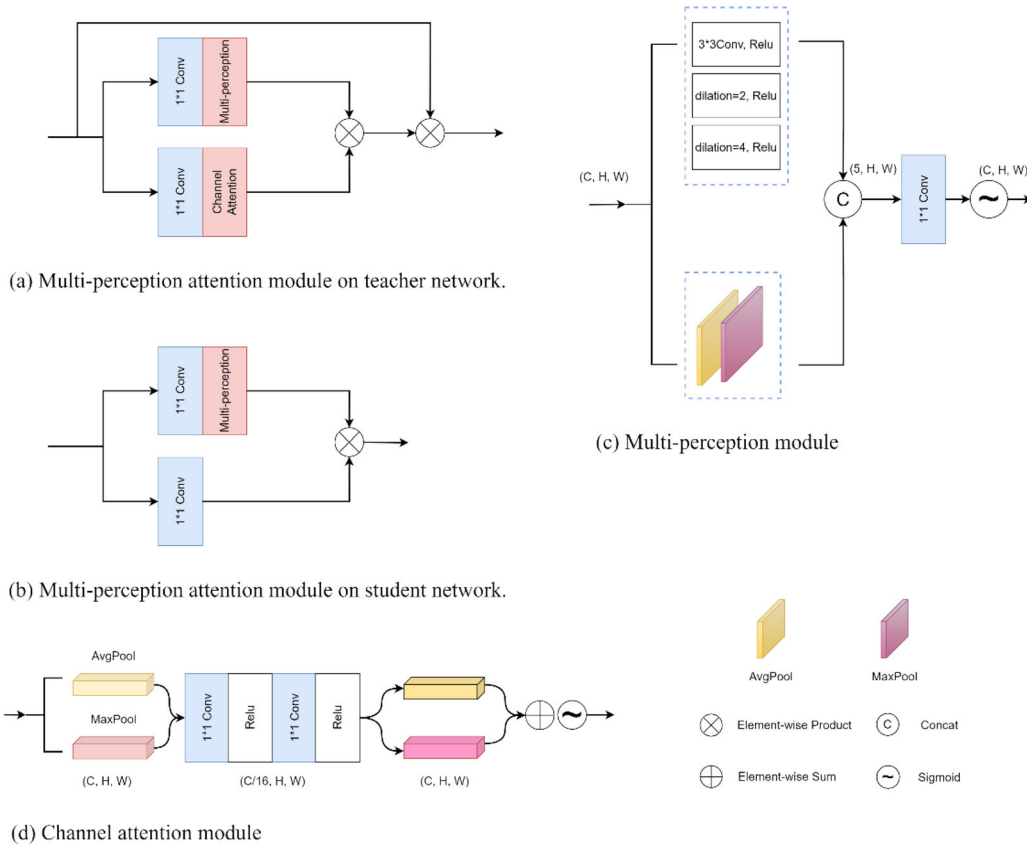


Fig. 3 Multi-perception attention module

$$L_{contrast} = \sum_{i=1}^n \omega_i \cdot \frac{Dis(\varphi_i(J), \varphi_i(D(I)))}{Dis(\varphi_i(I), \varphi_i(D(I)))} \tag{9}$$

$\varphi_i, i = 1, 2, 3, \dots, n$, are the hidden layer features extracted by a fixed pretrained model φ for the i -th layer, and ω_i are the corresponding coefficients. $Dis(a, b)$ computes the L1 loss between features a and b .

Knowledge-Guided Loss If we regard the image reconstruction as an encoding–decoding process, most deep learning-based dehazing methods only use the clear image itself as the ground truth for the loss function design. To provide guidance for the dehazing process, we trained an autoencoder network with clear images as a teacher network to provide guidance at the feature level of the clear image decoder. Specifically, when we train the dehazing student network with paired haze and clear images, we also input the clear images into the pre-trained teacher network, then, use the knowledge-guided regularization term to reduce the distance between the decoder features of the student network and the teacher network. In addition, we give a weight coefficient to each level of features to further increase the flexibility of guidance, as shown in Eq. (10).

$$L_{kg} = \alpha_1 |T_{mpa} - S_{mpa}|_1 + \alpha_2 |T_{up1/2} - S_{up1/2}|_1 + \alpha_3 |T_{up1} - S_{up1}|_1 \tag{10}$$

In this equation, the subscripts $mpa, up1/2$, and $up1$ represent features of the last multi-perception attention module, the first and second upsampling modules, which upsample the features to half and full of the input size respectively. α_1, α_2 and α_3 are the corresponding coefficients.

4 Experiment

In this section, we describe the experiments conducted to evaluate our proposed method. We first introduce the datasets used in the experiments, including the division of the training and testing sets. Then, we describe the experimental setup, including the training environment and configuration of the training parameters. We compare our experimental results with the results produced by currently popular methods, including both qualitative and quantitative comparisons, and we compare the parameter count of our deep learning method. Finally, we conduct ablation experiments to demonstrate the effectiveness of each module in the network.

4.1 Datasets

We utilize following datasets, namely, NYU [26], RESIDE-SOTS [25], I-haze/O-haze [34, 35], NH-haze [27, 28] and Dense-haze [33], to train our model.

We tested the performance of different algorithms on the following synthetic hazy image datasets: The NYU dataset contains 1449 pairs of hazy and clear images, of which 1159 randomly selected pairs are used for training, and the remaining 290 pairs are used for testing. For the SOTS dataset, the outdoor set consists of 500 paired images, of which 450 pairs are randomly selected for training and the remaining 50 pairs for testing. Additionally, we use 500 clear images from SOTS-outdoor to train our teacher network. The indoor set also contains 500 paired images, with 450 pairs used for training and the remaining 50 pairs for testing.

To test the performance of different algorithms on real hazy images, we evaluated them on the following real hazy image datasets: I-HAZE/O-HAZE, which contain images with homogeneous haze; NH-HAZE, which contains images with nonhomogeneous haze; Dense-HAZE, which contains images with dense haze; and RESIDE-HSTS, which contains unpaired real hazy images. We combined 30 pairs of hazy and haze-free images from the I-HAZE dataset and 45 pairs of hazy and clear images from the O-HAZE dataset, obtaining a total of 75 pairs of images. We randomly selected 60 pairs for training and 15 pairs for testing. The NH-HAZE dataset comprises 55 pairs of real nonhomogeneous hazy and haze-free images. We randomly selected 44 pairs for training and 11 pairs for testing. The Dense-HAZE dataset comprises 55 pairs of real dense hazy and haze-free images. We randomly selected 44 pairs for training and 11 pairs for testing.

4.2 Training

We train our proposed KMAN end-to-end by minimizing the loss function given in Eq. (6). The input images are resized to 256×256 before being fed into the network. The model is trained for 100 epochs using an NVIDIA RTX 3090 GPU. The initial learning rate is set to 0.001, and a cosine annealing strategy is employed for updates. The regularization coefficients α , β , and γ are set to 0.1, and the knowledge guidance coefficients α_1 , α_2 , and α_3 are set to 0.125, 1, and 1, respectively. The batch size is fixed at 10. All algorithms use the same test set, and all supervised learning algorithms use the same training set. In addition, to compare the generalization abilities of different algorithms on other datasets, we train the supervised learning algorithms on the SOTS-outdoor dataset and test them on other datasets.

4.3 Evaluation experiments

We conducted quantitative and qualitative comparisons on multiple datasets to evaluate the performance of our proposed image dehazing algorithm in terms of dehazing quality, performance, and parameter count. We used PSNR and SSIM as standard quantitative evaluation metrics and compared the parameter count of our algorithm with those of other deep learning methods. Moreover, our method exhibited significant advantages in visual effects, indicating its capability to remove haze while preserving more image details.

To fully test the dehazing abilities of different algorithms, we designed the following experiments. First, we split each dataset into training and testing sets and evaluated the performance of each algorithm on each testing set. Then, we trained all the supervised learning algorithms on the SOTS-outdoor dataset and tested them on the other datasets. The results are shown in Tables 1 and 2.

According to Table 1, our algorithm achieves state-of-the-art results on most of the artificially synthesized image dehazing datasets and the real-world image dehazing datasets. This indicates that our algorithm is better suited to handling relatively complex scenes, exhibiting favorable performance on real hazy images with varying levels of homogeneous, nonhomogeneous, and heavy haze. From Table 2, our algorithm demonstrates superior generalization performance on both the artificially synthesized dataset and the datasets with homogeneous and nonhomogeneous haze concentrations. Considering the results from Tables 1 and 2, we hypothesize that the difficulty in generalizing our algorithm to dense haze scenarios stems from the training on artificially synthesized hazy images with homogeneous concentrations in the SOTS-outdoor dataset. Furthermore, we compared the parameter counts of different algorithms. The results show that our algorithm has a parameter count of 1.51 million, which, although not the lowest, strikes a good balance between performance and parameter count. Compared to other algorithms, our method achieves better dehazing effects and visual perception while maintaining fewer parameters.

The dehazing results on paired image dehazing datasets are illustrated in Fig. 4. The dehazing results of our algorithm exhibit superior performance in terms of texture clarity and haze removal. The algorithm does not produce significant texture damage or color distortion, and it performs well in scenes with homogeneous, nonhomogeneous, and high-density haze in real-world scenarios.

In addition to comparing the performance on datasets, we further demonstrate the superiority of our method in terms of generalization on real-world hazy images. As shown in Figs. 5 and 6, we selected unpaired real hazy images from RESIDE-HSTS and several unpaired real-world hazy images

Table 1 Quantitative comparison results on different datasets

Method	SOTS-outdoor		SOTS-indoor		NYU		I-haze/O-haze		NH-haze		Dense-haze		Param
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	
DCP [5]	19.97	0.903	21.04	0.860	14.87	0.756	15.90	0.713	12.81	0.474	12.42	0.414	–
AOD [7]	22.31	0.882	19.93	0.833	13.75	0.624	17.86	0.714	14.51	0.504	12.82	0.383	0.002 M
AMEF [29]	20.01	0.890	18.24	0.799	13.32	0.718	17.99	0.792	13.39	0.542	13.30	0.399	–
CAP [6]	23.57	0.939	19.97	0.843	13.16	0.717	17.32	0.779	12.38	0.510	12.33	0.446	–
RefineDNet [31]	24.79	0.901	28.37	0.885	16.99	0.740	18.46	0.740	13.27	0.489	13.44	0.425	65.80 M
KTDN	23.83	0.795	29.94	0.896	23.95	0.830	20.75	0.648	17.83	0.424	15.68	0.428	49.35 M
T-Net	25.91	0.954	27.56	0.943	22.62	0.871	21.47	0.860	17.73	0.721	14.61	0.537	2.55 M
PAN [22]	27.29	0.960	23.92	0.902	19.29	0.830	20.96	0.860	16.85	0.690	14.74	0.529	3.18 M
Ours	28.37	0.964	34.41	0.971	23.82	0.899	21.55	0.862	18.45	0.718	15.30	0.551	1.51 M

Bold indicates the highest indicators

Table 2 The results of using SOTS-outdoor as the training set

Method	SOTS-outdoor		SOTS-indoor		NYU		I-haze/O-haze		NH-haze		Dense-haze		Param
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	
AOD	22.53	0.879	17.40	0.788	11.07	0.637	15.75	0.648	12.26	0.434	11.96	0.384	0.002 M
RefineDNet	24.95	0.901	19.09	0.790	12.86	0.674	16.38	0.717	12.62	0.486	12.37	0.446	65.80 M
KTDN	26.66	0.836	18.09	0.725	11.81	0.607	16.26	0.597	12.31	0.363	11.45	0.346	49.35 M
T-Net	25.54	0.945	17.32	0.820	10.47	0.657	16.19	0.707	12.33	0.476	11.92	0.372	2.55 M
PAN	27.51	0.957	19.06	0.843	11.16	0.661	15.96	0.711	12.17	0.487	12.05	0.375	3.18 M
Ours	30.18	0.972	19.74	0.856	11.87	0.683	16.07	0.724	16.34	0.501	10.60	0.353	1.51 M

Bold indicates the highest indicators

as test samples. The results show that our algorithm is better able to restore image details, colors, and contrast while effectively removing haze.

Based on the comprehensive results of our experiments, our algorithm demonstrates outstanding performance in various respects, including superior dehazing effects, performance, and visual perception.

4.4 Ablation experiments

Our image dehazing algorithm consists of a multi-perception attention module and a knowledge guidance mechanism provided by a teacher network for the decoding process. To validate the impacts of these two modules on the algorithm's performance, we conducted ablation experiments on the multi-perception attention module and the teacher network, as shown in Table 3.

Backbone We removed both the multi-perception attention module and the knowledge guidance mechanism and used the remaining modules as the backbone network for image

dehazing. The experiments were conducted on the SOTS-outdoor dataset, with PSNR and SSIM used as evaluation metrics. The results showed a significant performance drop compared to our complete algorithm. The PSNR decreased from 28.37 to 27.57, and the SSIM decreased from 0.964 to 0.961. This indicates that the multi-perception attention module and the teacher network play a crucial role in improving the performance of image dehazing.

Backbone + KG In the ablation experiment on the multi-perception attention module, we removed the module from our algorithm while retaining the knowledge guidance mechanism for image dehazing. Similarly, experiments were conducted on the SOTS-outdoor dataset, with PSNR and SSIM used as evaluation metrics. The results showed a noticeable performance drop after removing the multi-perception attention module. The PSNR decreased from 28.37 to 27.75, and the SSIM decreased from 0.964 to 0.963. This demonstrates the crucial role of the multi-perception attention module in improving the performance of image dehazing.



Fig. 4 Qualitative comparison on different datasets. From the first row to the sixth row, they correspond to the images from the datasets NYU, RESIDE-SOTS, I-haze/O-haze, NH-haze and Dense-haze respectively

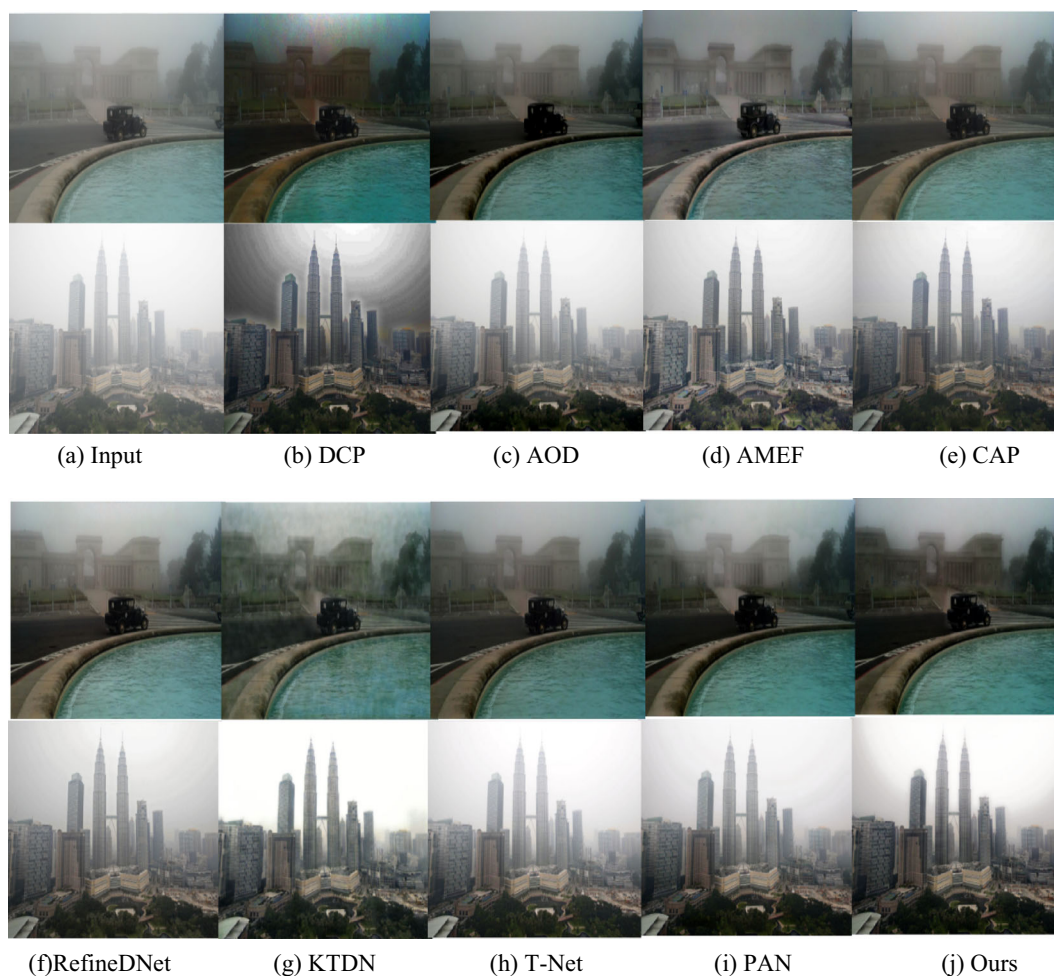


Fig. 5 Qualitative comparison on HSTS real-world images

Backbone + MPA In the ablation experiment on the knowledge guidance loss, we removed the teacher network module while retaining the multi-perception attention module for image dehazing. Again, experiments were conducted on the SOTS-outdoor dataset, with PSNR and SSIM used as the evaluation metrics. The results showed a slight performance drop after removing the teacher network module. The PSNR decreased from 28.37 to 28.15. This indicates that the teacher network module plays an important role in providing knowledge guidance for the decoding process, helping the decoder module to better restore the original image.

Loss Function We tested and recorded the values of PSNR and SSIM during the training process of 100 epochs on the SOTS-outdoor dataset. The results are shown in Fig. 7, which shows that the performance of the model gradually improves as we gradually add regularization terms to the loss function.

In addition, to further demonstrate the effectiveness of the modules and methods we propose, we conducted a qualitative analysis of the results on real-world images, as shown in Fig. 8. It can be seen that both the MPA module and

knowledge guidance significantly improve the haze removal performance of the images, and they can adapt to different complex scenes.

In summary, we can conclude that both the multi-perception attention module and the teacher network module are key components in our algorithm, and their combination is crucial for improving the performance of image dehazing. The multi-perception attention module helps the algorithm better focus on useful information and features in the image while adapting to different scenes and haze densities. The teacher network module provides valuable knowledge guidance for the decoding process, aiding the algorithm in better restoring the original image.

5 Conclusion

In this paper, we propose a knowledge-guided multi-perception attention network (KMAN) for image dehazing. We conducted experiments on multiple datasets to evaluate

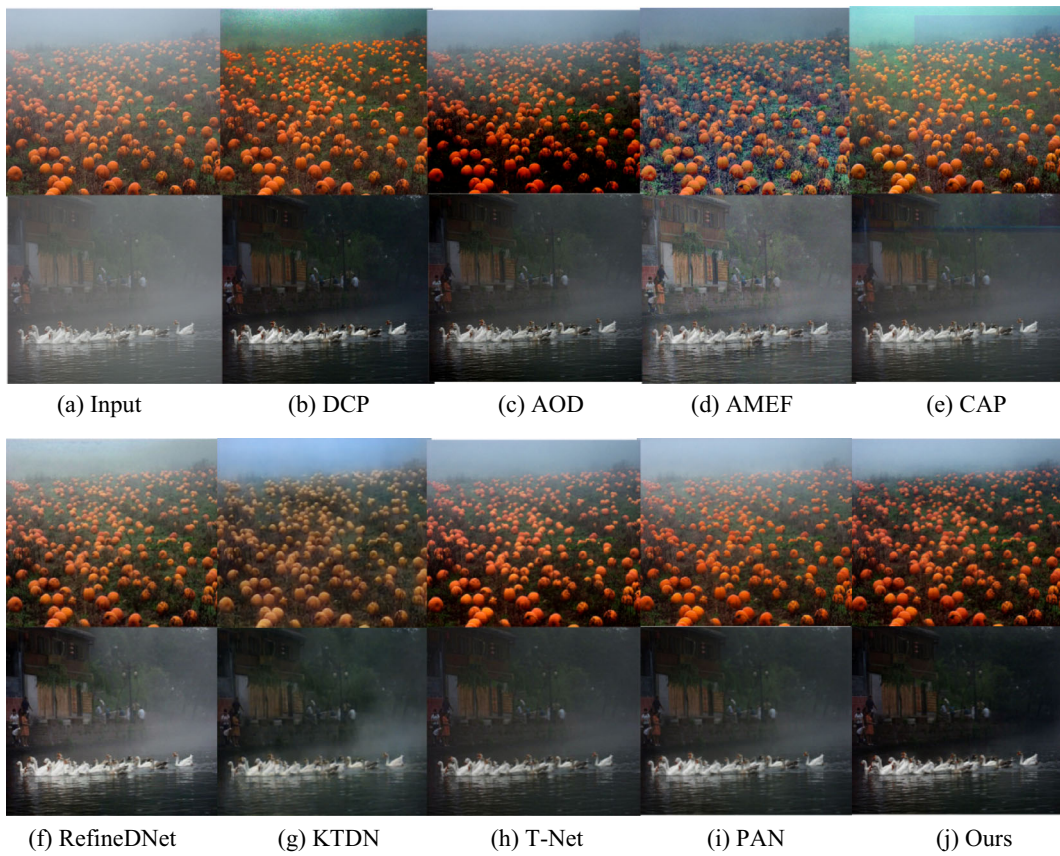


Fig. 6 Qualitative comparison on real-world images

Table 3 Ablation results

	Backbone	Backbone + KG	Backbone + MPA	Backbone + MPA + KG(ours)
PSNR	27.57	27.75	28.15	28.37
SSIM	0.961	0.963	0.965	0.964

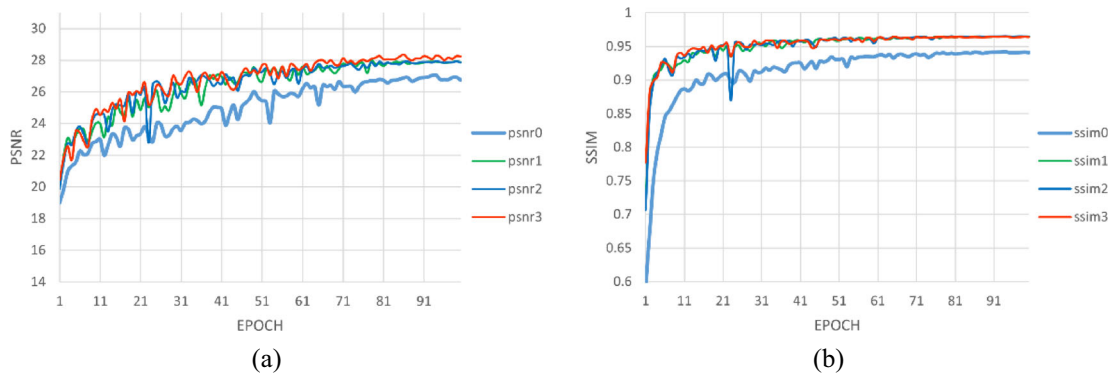


Fig. 7 Ablation experiment results for the loss function. **a** and **b** are the values of PSNR and SSIM in 100 epochs: “psnr0” and “ssim0” correspond to the results of retaining only the L1 loss, “psnr1” and “ssim1” correspond to adding perceptual loss on the basis of L1 loss, “psnr2”

and “ssim2” correspond to adding perceptual loss and contrastive loss on the basis of L1 loss, and “psnr3” and “ssim3” correspond to adding perceptual loss, contrastive loss and knowledge-guided loss on the basis of L1

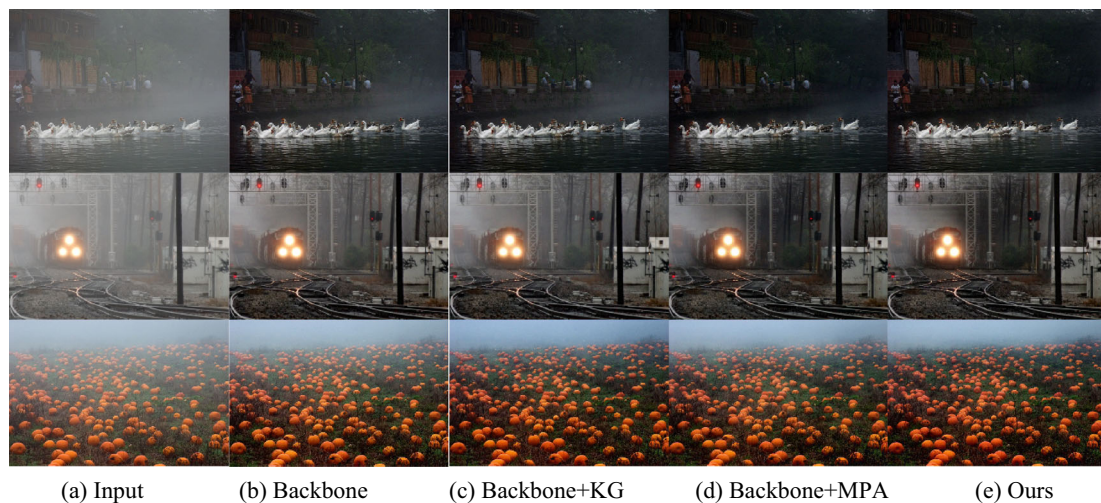


Fig. 8 The results of ablation experiments on real images

the performance of our proposed algorithm in terms of the dehazing effect, performance, and parameter size. The results show that our algorithm performs well in all respects, achieving a good balance between performance and parameter size while outperforming other deep learning methods in terms of average PSNR and SSIM scores and visual effects. Furthermore, our algorithm performs particularly well in challenging scenes and images, demonstrating its generalization ability on real-world hazy images. In the future, we plan to further optimize our algorithm and continue to explore more effective algorithms for image restoration tasks.

Funding This study was funded by the National Natural Science Foundation of China (Grant No. 61773243).

Data availability Data sharing is not applicable to this article, as no new data were created or analysed in this study.

Declarations

Conflict of interest The authors declare no conflicts of interest.

References

1. McCartney, E.J.: Optics of the atmosphere: scattering by molecules and particles. New York. (1976)
2. Narasimhan, S.G., Nayar, S.K.: Vision and the atmosphere. *Int. J. Comput. Vision* **48**, 233 (2002)
3. He, K., Sun, J., Tang, X.: Single image haze removal using dark channel prior. *IEEE Trans. Pattern Anal. Mach. Intell.* **33**, 2341–2353 (2010)
4. Berman, D., Treibitz, T., Avidan, S.: Single image dehazing using haze-lines. *IEEE Trans. Pattern Anal. Mach. Intell.* **42**, 720–734 (2018)
5. Ye, C., Evanusa, M., He, H., Mitrokhin, A., Goldstein, T., Yorke, J.A., Fermüller, C., Aloimonos, Y.: Network deconvolution. *arXiv preprint arXiv:1905.11926*. (2019)
6. Zhu, Q., Mai, J., Shao, L.: A fast single image haze removal algorithm using color attenuation prior. *IEEE Trans. Image Process.* **24**, 3522–3533 (2015)
7. Li, B., Peng, X., Wang, Z., Xu, J., Feng, D.: Aod-net: All-in-one dehazing network. In: *Proceedings of the IEEE International Conference on Computer Vision*. pp. 4770–4778 (2017)
8. Cai, B., Xu, X., Jia, K., Qing, C., Tao, D.: Dehazenet: an end-to-end system for single image haze removal. *IEEE Trans. Image Process.* **25**, 5187–5198 (2016)
9. Zhang, H., Patel, V.M.: Densely connected pyramid dehazing network. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. pp. 3194–3203 (2018)
10. Wu, H., Liu, J., Xie, Y., Qu, Y., Ma, L.: Knowledge transfer dehazing network for nonhomogeneous dehazing. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*. pp. 478–479 (2020)
11. Kim, G., Kwon, J.: Self-parameter distillation dehazing. *IEEE Trans. Image Process.* **32**, 631–642 (2022)
12. Hong, M., Xie, Y., Li, C., Qu, Y.: Distilling image dehazing with heterogeneous task imitation. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 3462–3471 (2020)
13. Sun, Z., Zhang, Y., Bao, F., Wang, P., Yao, X., Zhang, C.: Sadnet: semi-supervised single image dehazing method based on an attention mechanism. *ACM Trans. Multimed. Comput. Commun. Appl. (TOMM)* **18**, 1–23 (2022)
14. Lin, C., Rong, X., Yu, X.: Msaff-net: multiscale attention feature fusion networks for single image dehazing and beyond. *IEEE Trans. Multimed.* (2022)
15. Guo, H.-F., Piao, J.-C.: MARG-UNet: a single image dehazing network based on multimodal attention residual group. In: *2022 IEEE 2nd International Conference on Information Communication and Software Engineering (ICICSE)*. pp. 105–109. IEEE (2022)
16. Zheng, Y., Su, J., Zhang, S., Tao, M., Wang, L.: Dehaze-AGGAN: unpaired remote sensing image dehazing using enhanced attention-guide generative adversarial networks. *IEEE Trans. Geosci. Remote Sens.* **60**, 1–13 (2022)
17. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I.: Attention is all you need. *Adv. Neural Inf. Process. Syst.* **30**, (2017)
18. Hu, J., Shen, L., Sun, G.: Squeeze-and-excitation networks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. pp. 7132–7141 (2018)

19. Wang, X., Girshick, R., Gupta, A., He, K.: Non-local neural networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 7794–7803 (2018)
20. Hinton, G., Vinyals, O., Dean, J.: Distilling the knowledge in a neural network. arXiv preprint [arXiv:1503.02531](https://arxiv.org/abs/1503.02531). (2015)
21. Chen, G., Choi, W., Yu, X., Han, T., Chandraker, M.: Learning efficient object detection models with knowledge distillation. *Adv. Neural Inf. Process. Syst.* **30**, (2017)
22. Liu, J., Liu, P., Zhang, Y.: Multi-scale feature fusion pyramid attention network for single image dehazing. *IET Image Process.* (2023)
23. Johnson, J., Alahi, A., Fei-Fei, L.: Perceptual losses for real-time style transfer and super-resolution. In: Computer Vision—ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part II 14. pp. 694–711. Springer (2016)
24. Wu, H., Qu, Y., Lin, S., Zhou, J., Qiao, R., Zhang, Z., Xie, Y., Ma, L.: Contrastive learning for compact single image dehazing. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 10551–10560 (2021).
25. Li, B., Ren, W., Fu, D., Tao, D., Feng, D., Zeng, W., Wang, Z.: Benchmarking single-image dehazing and beyond. *IEEE Trans. Image Process.* **28**, 492–505 (2018)
26. Silberman, N., Hoiem, D., Kohli, P., Fergus, R.: Indoor segmentation and support inference from rgbd images. *ECCV* **5**(7576), 746–760 (2012)
27. Ancuti, C.O., Ancuti, C., Timofte, R.: 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. pp. 444–445 (2020)
28. Ancuti, C.O., Ancuti, C., Vasluianu, F.-A., Timofte, R.: Ntire 2020 challenge on nonhomogeneous dehazing. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. pp. 490–491 (2020)
29. Galdran, A.: Image dehazing by artificial multiple-exposure image fusion. *Signal Process.* **149**, 135–147 (2018)
30. Ren, W., Liu, S., Zhang, H., Pan, J., Cao, X., Yang, M.-H.: Single image dehazing via multi-scale convolutional neural networks. In: Computer Vision—ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part II 14. pp. 154–169. Springer (2016)
31. Zhao, S., Zhang, L., Shen, Y., Zhou, Y.: RefineDNet: a weakly supervised refinement framework for single image dehazing. *IEEE Trans. Image Process.* **30**, 3391–3404 (2021)
32. Zhao, H., Gallo, O., Frosio, I., Kautz, J.: Loss functions for image restoration with neural networks. *IEEE Trans. Comput. Imag.* **3**, 47–57 (2016)
33. Ancuti, C.O., Ancuti, C., Sbert, M., Timofte, R.: Dense-Haze: A benchmark for image dehazing with dense-haze and haze-free images. In: 2019 IEEE International Conference on Image Processing (ICIP). pp. 1014–1018 (2019). <https://doi.org/10.1109/ICIP.2019.8803046>
34. Ancuti, C., Ancuti, C.O., Timofte, R., De Vleeschouwer, C.: I-HAZE: A Dehazing Benchmark with Real Hazy and Haze-Free Indoor Images. In: Blanc-Talon, J., Helbert, D., Philips, W., Popescu, D., Scheunders, P. (eds.) *Advanced concepts for intelligent vision systems*, pp. 620–631. Springer International Publishing, Cham (2018). https://doi.org/10.1007/978-3-030-01449-0_52
35. Ancuti, C.O., Ancuti, C., Timofte, R., De Vleeschouwer, C.: O-HAZE: a dehazing benchmark with real hazy and haze-free outdoor images. In: Presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (2018)
36. Yi, W., Dong, L., Liu, M., Hui, M., Kong, L., Zhao, Y.: MFAF-Net: image dehazing with multi-level features and adaptive fusion. *Vis. Comput.* (2023). <https://doi.org/10.1007/s00371-023-02917-8>
37. Chen, Z., Hu, Z., Sheng, B., Li, P., Kim, J., Wu, E.: Simplified non-locally dense network for single-image dehazing. *Vis. Comput.* **36**, 2189–2200 (2020). <https://doi.org/10.1007/s00371-020-01929-y>
38. Zheng, L., Li, Y., Zhang, K., Luo, W.: T-Net: deep stacked scale-iteration network for image dehazing. *IEEE Trans. Multimed.* (2022). <https://doi.org/10.1109/TMM.2022.3214780>
39. Gou, J., Yu, B., Maybank, S.J., Tao, D.: Knowledge distillation: a survey. *Int. J. Comput. Vis.* **129**, 1789–1819 (2021). <https://doi.org/10.1007/s11263-021-01453-z>
40. Li, Z., Hoiem, D.: Learning without forgetting. *IEEE Trans. Pattern Anal. Mach. Intell.* **40**, 2935–2947 (2017)
41. Luo, P., Zhu, Z., Liu, Z., Wang, X., Tang, X.: Face model compression by distilling knowledge from neurons. In: Proceedings of the AAAI Conference on Artificial Intelligence (2016)
42. Li, Q., Jin, S., Yan, J.: Mimicking very efficient network for object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 6356–6364 (2017)
43. Lan, Y., Cui, Z., Su, Y., Wang, N., Li, A., Li, Q., Zhong, X., Zhang, C.: SSKDN: a semisupervised knowledge distillation network for single image dehazing. *J. Electron. Imag.* **32**, 013002–013002 (2023)
44. Lan, Y., Cui, Z., Su, Y., Wang, N., Li, A., Han, D.: Physical-model guided self-distillation network for single image dehazing. *Front. Neurobot.* **16**, 1036465 (2022)
45. Su, Y.Z., He, C., Cui, Z.G., Li, A.H., Wang, N.: Physical model and image translation fused network for single-image dehazing. *Pattern Recogn.* **142**, 109700 (2023)
46. Lan, Y., Cui, Z., Su, Y., Wang, N., Li, A., Zhang, W., Li, Q., Zhong, X.: Online knowledge distillation network for single image dehazing. *Sci. Rep.* **12**, 14927 (2022)
47. Wang, N., Cui, Z., Li, A., Su, Y., Lan, Y.: Multi-priors guided dehazing network based on knowledge distillation. In: Chinese Conference on Pattern Recognition and Computer Vision (PRCV). pp. 15–26. Springer (2022)

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.



Peng Liu is pursuing his M.S. degree at the School of Cyber Science and Engineering, Qufu Normal University. He received his B.S. degree in Computer Science and Technology from Shandong University of Science and Technology. His research interests include deep learning and image processing.



Jianlei Liu is a professor at the School of Cyber Science and Engineering, Qufu Normal University. He received his B.S. degree from Shandong Agricultural University, his M.S. degree from Qufu Normal University, and his Ph.D. degree from Xidian University. His research interests include machine learning, intelligent transportation, and image processing.