Computer Vision and Image Understanding 197–198 (2020) 103003

Contents lists available at ScienceDirect

Computer Vision and Image Understanding

journal homepage: www.elsevier.com/locate/cviu

Pyramid Channel-based Feature Attention Network for image dehazing

Xiaoqin Zhang, Tao Wang, Jinxin Wang, Guiying Tang, Li Zhao *

College of Computer Science and Artificial Intelligence, Wenzhou University, Wenzhou, 325035, China

ARTICLE INFO

Keywords:
Image dehazing
Deep neural network
Channel attention

ABSTRACT

Traditional deep learning-based image dehazing methods usually use the high-level features (which contain more semantic information) to remove haze in the input image, while ignoring the low-level features (which contain more detail information). In this paper, a Pyramid Channel-based Feature Attention Network (PCFAN) is proposed for single image dehazing, which leverages complementarity among different level features in a pyramid manner with channel attention mechanism. PCFAN consists of three modules: a three-scale feature extraction module, a pyramid channel-based feature attention module (PCFA), and an image reconstruction module. The three-scale feature extraction module simultaneously captures the low-level spatial structural features and the high-level contextual features in different scales. The PCFA module utilizes the feature pyramid and the channel attention mechanism, which effectively extracts interdependent channel maps and selectively aggregates the more important features in a pyramid manner for image dehazing. The image reconstruction module is used to reconstruct features to recover a clear image. Meanwhile, a loss function that combines a mean square error loss part and an edge loss part is employed in PCFAN, which can better preserve image details. Experimental results demonstrate that the proposed PCFAN outperforms existing state-of-the-art algorithms on standard benchmark datasets in terms of accuracy, efficiency, and visual effect. The code will be made publicly available.

1. Introduction

Image dehazing problem, which aims at recovering a clear image from a given hazy input, is one of the classical image processing problems. It has attracted significant attention in the fields of image processing and computer vision in recent decades, as the techniques of image dehazing are required in many higher-level vision tasks (Zhang et al., 2020b; Yuan et al., 2017; Liu et al., 2018; Zhang et al., 2015).

Most successful methods depend on the atmosphere scattering model (Narasimhan and Nayar, 2002), which provides an estimate of the haze-free image. It is formulated as:

$$I(x) = t(x)J(x) + A(x)(1 - t(x)), \tag{1}$$

where x refers to the pixel coordinates in the image plane, I denotes the observed image that is degraded by haze, and J is the haze-free scene image. The matrix A represents the global atmospheric light, and the transmission map t is the medium transmission rate which describes the portion of the light that reaches the camera sensors without being scattered. The transmission map t can be expressed as $t(x) = e^{-\beta d(x)}$, where β is the scatting coefficient of the atmosphere and $d(x)$ is the scene depth. However, the transmission map and the atmospheric light are unknown in practice. Therefore, many image dehazing methods estimate t and A from a hazy image I, and then obtain the unknown clear image J via the atmosphere scattering model.

Previous image dehazing approaches concentrate more on restoring the clear image using priors such as dark-channel prior, contrast color-lines, and haze-line prior. For example, He et al. (2010) propose a dark channel prior (DCP) based method for estimating the transmission map. Although these prior-based methods have achieved considerable success, their performances are limited because not all the images of real scenes are compatible with the predefined priors. Recently, deep learning has exhibited effectiveness in various computer vision tasks. Various convolutional neural network (CNN) based methods have been proposed to estimate the transmission map and the atmospheric light. Once the transmission map and the atmospheric light are estimated, the dehazed image is restored through the atmosphere scattering model. Generally speaking, low-level features in a CNN partly refer to the detail information, and high-level features contain more semantic information. Both of them are important for recovering a clear image, but most CNN-based methods usually use high-level features to achieve image dehazing. Moreover, these methods are based on the atmosphere scattering model. If the estimated transmission map and atmospheric light are not accurate, then the dehazed result will be of low quality.

* Corresponding author.
E-mail addresses: zhangxiaoqinnan@gmail.com (X. Zhang), taowangzj@gmail.com (T. Wang), jxwang@stu.wzu.edu.cn (J. Wang), guiyingtang9503@163.com (G. Tang), lizhao@wzu.edu.cn (L. Zhao).

https://doi.org/10.1016/j.cviu.2020.103003
Received 11 December 2019; Received in revised form 16 April 2020; Accepted 29 May 2020
Available online 2 June 2020

Fig. 1. Examples of image dehazing results. Top left: input hazy image. Top right and bottom left: restored haze-free images using DCP and EPDN respectively. Bottom right: dehazed image generated by our method. Zoom in for better visibility.

In this work, we propose a novel end-to-end framework called Pyramid Channel-based Feature Attention Network (PCFAN) for single image dehazing, which leverages complementarity among different level features in a pyramid manner with channel attention mechanism. Specifically, PCFAN consists of three modules: a three-scale feature extraction module, a pyramid channel-based feature attention (PCFA) module, and an image reconstruction module. First, the three-scale feature extraction module extracts features at three different scales. Then, these features are fed into the PCFA module. This module extracts more important attention features by the channel-attention blocks and fuses these attention features in different levels. Finally, based on the output of PCFA, the image reconstruction module is used to restore a clear image. In addition, we introduce a training loss function that consists of two terms: the MSE loss and the Edge loss. The MSE loss is utilized to measure the pixel-wise distance, while the Edge loss promotes to generate a clean image with more details. As shown in Fig. 1, the proposed PCFAN produces a more realistic image with more details.

The main features of the proposed image dehazing method are summarized as follows.

- We propose an end-to-end Pyramid Channel-based Feature Attention Network for single image dehazing, which does not need to explicitly estimate the transmission map and the atmospheric light.
- The PCFA module can extract more informative features by the channel attention block, and fuse the complementary features in different levels in a pyramid manner.
- A loss function that combines a mean square error loss part and an edge loss part is employed in PCFAN, which can better preserve image details.
- Extensive experiments on standard benchmark datasets demonstrate that the proposed PCFAN performs favorably compared with state-of-the-art methods, in terms of quantitative accuracy and qualitative visual effect.

The rest of this paper is structured as follows: A brief review of image dehazing and the attention mechanism is given in Section 2. The proposed PCFAN method is discussed in Section 3. The experimental results are presented in Section 4, and Section 5 is a conclusion of this paper.

2. Related work

In this section, we introduce related work on both the image dehazing and the attention mechanism as follows.

Image dehazing. Recent years have witnessed great advancements in the task of single image dehazing. Many classical methods have been proposed in the existing literature to tackle this well-known ill-posed problem (Zhao et al., 2019; Hodges et al., 2019; Alajarmeh et al., 2018; He et al., 2010; Ren et al., 2016). These methods can be generally classified as either image prior-based dehazing methods or deep learning-based dehazing methods. He et al. (2010) propose a novel prior-based method called dark channel prior (DCP) to accurately estimate the transmission map. The DCP applies if at least one color channel in the RGB color space has a very low intensity within a haze-free image without sky or bright regions. In the work by Zhu et al. (2015), the efficiency and effectiveness of the color attenuation prior in the single image dehazing task is demonstrated. This method estimates the transmission and restores the scene radiance to remove the haze from a single image. A linear model is used in a supervised fashion to build the bridge between the hazy image and its corresponding depth image. Berman et al. (2016) propose an algorithm, the computational complexity of which is linear in the size of an image, for image dehazing based on non-local priors. This algorithm assumes that the colors of a haze-free image can be approximated by typical colors that are clustered in the RGB color space. To solve suppressing artifacts in dehazing image, Chen et al. (2016) utilize the gradient residual minimization (GRM) to suppresses edges in the dehazing images that do not exist in the input images. Although these aforementioned methods have achieved success in haze removal, their dehazing performances are not always satisfying because of the assumptions on which they rely.

Recently, data-driven deep learning methods have demonstrated their superior capability in feature representation in many computer vision tasks (Krizhevsky et al., 2012; He et al., 2016; Kupyn et al., 2018; Li et al., 2018b). Diverse methods based on deep learning have been proposed for single image dehazing (Cai et al., 2016; Ren et al., 2016; Li et al., 2017; Zhang and Patel, 2018; Qu et al., 2019). Cai et al. (2016) introduce an end-to-end system called DehazeNet. It first estimates a medium transmission map, then restores a haze-free image via the classical atmosphere scattering model. In addition, the authors design special Maxout layers of units for feature extraction and a bilateral rectified linear unit to recover high quality images in DehazeNet. Ren et al. (2016) adopt a multi-scale deep neural network (MSCNN) to estimate the scene transmission maps. The algorithm consists of two parts: the coarse-scale network is used to predict the transmission maps, while the fine-scale network is utilized to locally refine the results for better haze removal. A CNN-based image dehazing model called the All-in-One Dehazing Network (AOD-Net) is presented by Li et al. (2017). Their light-weight network directly generates haze-free images rather than separately estimating atmospheric light and the transmission matrix for haze removal. In contrast with existing methods, Zhang and Patel (2018) propose the Densely Connected Pyramid Dehazing Network (DCPDN) for image dehazing in an end-to-end manner. The network simultaneously learns the transmission map, atmospheric light, and dehazed image, and then recovers the haze-free image. Chen et al. (2019) propose an end-to-end gated context aggregation network (GCA) to directly restore the final haze-free image. Treating the single-image dehazing problem as a terse image-to-image translation problem, EPDN (Qu et al., 2019) directly generates a haze-free image from the input hazy image, using a generative adversarial network and a novel multi-scale enhancer.

Attention mechanism. As a significant property of the human perception system (Itti et al., 1998), the attention mechanism can be considered as the guidance that leads an individual's sight to focus on the most important and informative parts of an input scene, rather than processing the whole scene at once. Recently, attention mechanism has

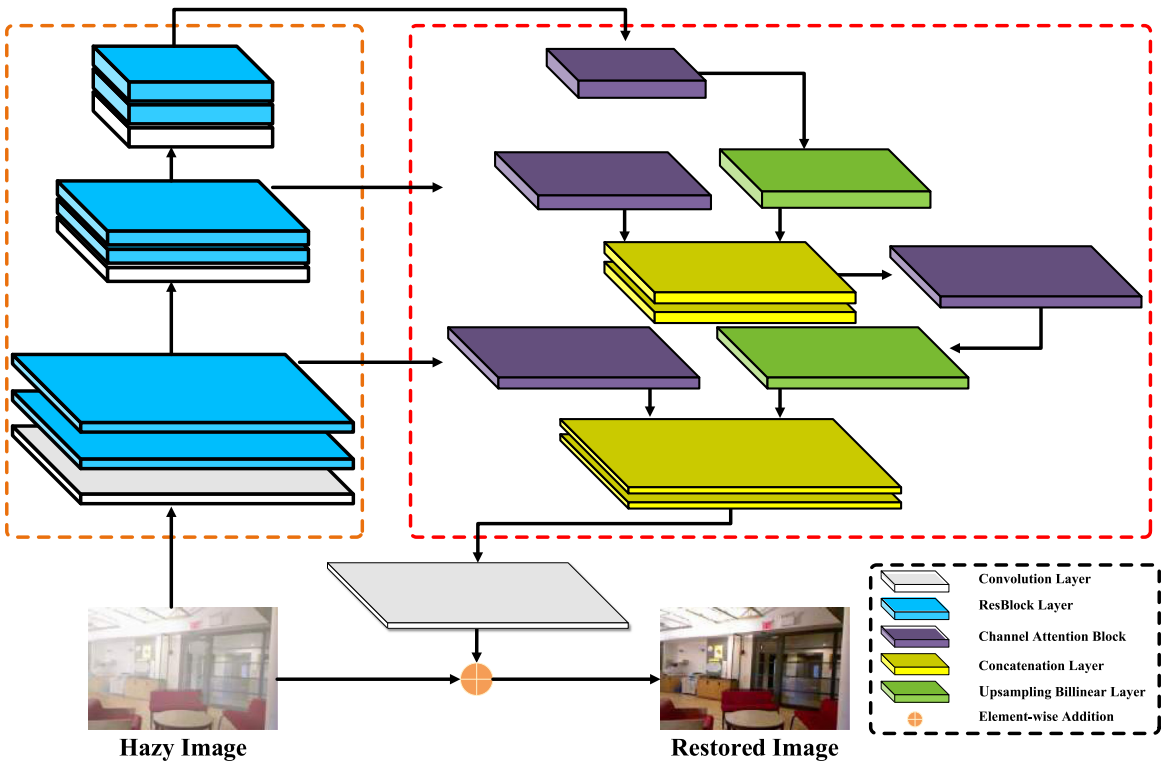


Fig. 2. Overall architecture of PCFAN: (1) Extract multi-scale features using the proposed three-scale feature extraction module (denoted by the yellow dotted line). Every feature extraction stage of the module consists of three components, namely a 3 × 3 convolution layer and two Resblocks; (2) The three-scale features generated by the feature extraction module are then fed into the proposed pyramid channel-based feature attention module (denoted by the dotted red line). Three channel attention blocks are used to process the features at different scales in a top-down pyramid fashion. This makes it possible to capture more crucial and informative features to predict better dehazed results; (3) Finally, the image reconstruction module, including a convolution operation and a simple element-wise addition operation, is adopted to restore the dehazed single image. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

been introduced into deep learning methods to handle many computer vision tasks (Zhang et al., 2020a; Cao et al., 2015; Jaderberg et al., 2015; Bluche, 2016; Zhang et al., 2018; Liu et al., 2019). After expressing some computer vision tasks as sequential decision tasks, Mnih et al. (2014) propose recurrent models for visual attention and fully optimize the non-differentiable model to learn task-specific policies using a policy gradient algorithm. Instead of processing the whole input image at once, the attention-based model sequentially and adaptively decides which regions should be focused on and processed. Spatial transformer networks are presented by Jaderberg et al. (2015), and the differentiable module is introduced to alleviate diverse problems in input images or multi-channel feature maps, including object rotation, scale transformation, translation, and clutter. To adaptively extract informative, high-frequency, channel-attention features in image super-resolution, Zhang et al. (2018) adopt the channel attention mechanism to enhance the representational ability of a very deep residual network. Using the classical global average pooling operation in the channel attention module, the useful channel-wise global spatial information is taken into consideration. Fu et al. (2019) propose the Dual Attention Network (DANet) based on the self-attention mechanism for the scene segmentation task. The proposed position attention module is designed to selectively learn the spatial interdependencies of features, while the channel attention module is utilized to emphasize channel interdependencies. Thus, precise segmentation results can be achieved with the two attention modules. The GridDehazeNet, proposed by Liu et al. (2019) for image dehazing, is a kind of multi-scale network with a channel-wise attention module. The channel-wise attention is utilized to reconstruct features of diverse scales, as well as to alleviate the bottleneck issue that occurs in some multi-scale networks.

3. Pyramid channel-based feature attention network

3.1. Network architecture

In this work, we combine the benefits of the channel-attention and pyramid operation, and propose a pyramid channel-based feature attention network (PCFAN) for image dehazing. The overall framework of PCFAN is illustrated in Fig. 2. The PCFAN consists of three modules, namely the three-scale feature extraction module, the pyramid channel-based feature attention module, and the image reconstruction module. The three-scale feature module contains three stages: The first feature extraction stage is composed of a 3 × 3 convolution layer and two Resblocks (He et al., 2016). In this stage, the depth (the number of channels) of feature maps is increased to 32. The following two stages both consist of a 3 × 3 convolution with stride 2 and two Resblocks. They increase the depth of the feature maps to 64 and 128, and reduce the resolution of the feature maps by half, respectively. Unlike previous works that only use the output features of the third stage, all the outputs of three stages are fed into the pyramid channel-based feature attention module that is constructed by many channel-attention blocks. The channel-attention block is used to dehaze features in both spatial and channel dimensions. Finally, an image reconstruction network consisting of only one convolution layer is utilized to reconstruct the clear image. The core components in PCFAN are the channel-attention block and the pyramid channel-based feature attention block, which are detailed in the following parts.

Channel attention block. In this work, to ensure that the network capture more informative features, the channel attention mechanism (Zhang et al., 2018) is employed to explore the interdependencies among feature channels.

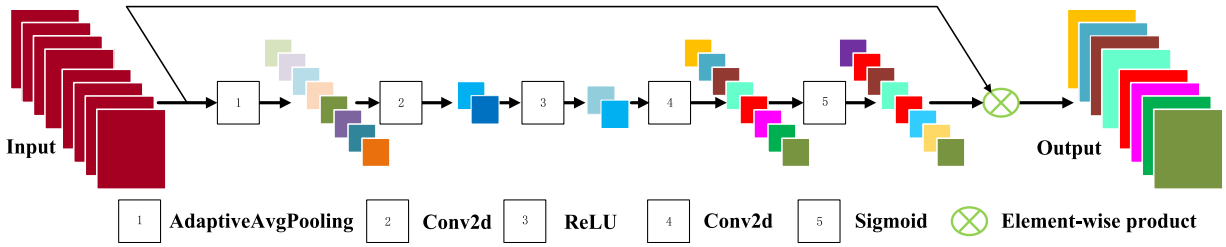


Fig. 3. Detailed structure of the channel attention block.

The channel attention block is presented in Fig. 3. Suppose that features $f \in \mathbb{R}^{C\times W\times H}$, $f = [f_1, f_2, \ldots, f_C]$ are given, where $f_i \in \mathbb{R}^{W\times H}$ is the i-th sub feature of f, and C is the set of channel numbers of f. First, the global channel-wise statistic of f is obtained by the global average pooling operation as follows:

$$v_c = \frac{1}{WH}\sum_{i=1}^{W}\sum_{j=1}^{H} f_c(i,j), \mu = [v_1, v_2, \ldots, v_C], \tag{2}$$

where v_c denotes the channel-wise feature, W and H are the width and height of the feature, respectively, $f_c(i,j)$ refers to the value of the c-th feature f_c at location (i,j), $[.,.,.]$ is concatenation operation, and therefore μ is the concatenation of $v_k(k = 1, \ldots, C)$. After that, two convolutions with the ReLU and Sigmoid activation function are used to learn linear and nonlinear interactions between channels. These operations can capture channel-wise dependencies among the aggregated features. This is formulated as:

$$\tilde{f} = \sigma(\phi_2(\eta(\phi_1(\mu)))), \tag{3}$$

where ϕ, η, and σ refer to the convolution layer, ReLU, and Sigmoid activation function, respectively. ϕ_1 aims to reduces the channels of input features. After being activated by ReLU η, the features are then increased to the original width (the number of channels) with a convolution layer ϕ_2. The final output feature F_{out} of this block is obtained by

$$F_{out} = \tilde{f} \otimes f, \tag{4}$$

where $\otimes$ is element-wise product, and f is original feature.

Pyramid channel-based feature attention module. As stated in Girshick (2015) and He et al. (2015), the pyramid operation can extract features from multiple layers of CNN and simultaneously fuse them to generate more effective features. However, these methods usually use an intuitive fusion strategy like addition or concatenation. Thus, we propose a pyramid channel-based feature attention module (PCFA), which combines the benefits of the feature pyramid and channel attention mechanism.

As shown in Fig. 2, PCFA consists of four channel-attention blocks, two upsampling layers, and two concatenation layers. There are two pathways in PCFA: bottom up and top down. For the bottom-up pathway, features from three layers are fed into corresponding channel attention blocks. Each channel-attention block processes the features in the corresponding channel, which selectively captures the important channel maps for the feature reconstruction. From the top-down pathway, PCFA first upsamples the size of features by a factor of 2, then integrates them. While from the down-top pathway, PCFA reconstructs higher spatial resolution from the semantically rich layers. The features fusion between the bottom-up and top-down pathways improves the feature representation ability, and effectively learn the importance of features from different levels with channel-attention mechanism. Thus, PCFA can fully exploit the complementary information between low-level and high-level features for image dehazing.

3.2. Loss function

To optimize the proposed network, two loss functions are utilized, namely the MSE loss $\mathscr{L}_{mse}$ and the Edge loss $\mathscr{L}_{edge}$.

MSE loss. The Mean Square Error (MSE) loss is used to measure differences in the pixel-wise aspect between the clear image and the output dehazed image. The MSE is defined by:

$$\mathscr{L}_{mse} = \frac{1}{CWH}\sum_{c=1}^{C}\sum_{i=1}^{W}\sum_{j=1}^{H}(I_{c,i,j}^{clear} - \tilde{I}_{c,i,j}^{dehazed})^2, \tag{5}$$

where C, W, and H represent the channel number, width, and height of an image, respectively, $I_{c,i,j}^{clear}$ is the value of ground truth at location (i,j) of channel c, and $\tilde{I}_{c,i,j}^{dehazed}$ corresponds to the value of the dehazed image generated by PCFAN.

Edge loss. To recover the clear image with more detail, we introduce an Edge loss function to the network. First, the convolution operation $Conv$ with Laplace operator (Trudinger, 1983) is used to obtain the edge images of the clear and dehazed images. Then, the $Tanh$ activation function is used to map values of edge image to $[0, 1]$. Finally, the pixel-wise distance (L_1 Norm) is used to measure the differences between clear and dehazed edge images. The Edge loss function is given by:

$$\begin{aligned}\mathscr{L}_{edge} =& \| Tanh(Conv(I^{clear}, k_{laplace})) - \\ & Tanh(Conv(\tilde{I}^{dehazed}, k_{laplace})) \|_1 .\end{aligned} \tag{6}$$

Total loss. In the training stage, the total loss is defined by combining these two loss functions, and is given by:

$$\mathscr{L} = \mathscr{L}_{mse} + \alpha \cdot \mathscr{L}_{edge}, \tag{7}$$

where α is a hyper-parameter that is used to yield the final loss $\mathscr{L}$. In this work, α is set to 0.01.

4. Experiments

In this section, extensive experiments are conducted on both a synthetic dataset and a real world dataset to demonstrate the effectiveness of the proposed network. The proposed network is compared with state-of-the-art image prior-based methods and learning-based methods, including DCP (He et al. CVPR'09), DehazeNet (Cai et al. TIP'16), MSCNN (Ren et al. ECCV'16), AOD-Net (Li et al. ICCV'17), GFN (Ren et al. CVPR'18), DCPDN (Zhang et al. CVPR'18), EPDN (Qu et al. CVPR'19) and FAMEDNet (Zhang TIP'20). In addition, an ablation study is conducted to verify the effectiveness of the Edge loss function and the pyramid channel-based feature attention module.

4.1. Experimental settings

Datasets. It is difficult to collect a large number of real-world hazy images and their haze-free counterparts. Thus, data-driven methods often rely on synthetic hazy images, which are generated from clear images based on the atmosphere scattering model using the proper scattering coefficient β and atmospheric light A. In this work, a large-scale synthetic dataset called RESIDE (Li et al., 2018a) is used to train and test the proposed PCFAN. RESIDE is divided into five different

Fig. 4. Exhibition of multi-channel feature maps after using the PCFA module. The first row depicts the input hazy image and its six typical channels selected from the multi-channel feature maps handled by PCFA. The same six channels of feature maps generated from the dehazed image and corresponding ground truth are presented in the second and third rows, respectively.

subsets: Indoor Training Set (ITS), Outdoor Training Set (OTS), Synthetic Objective Testing Set (SOTS), Real-World Task-Driven Testing Set (RTTS), and Hybrid Subjective Testing Set (HSTS). ITS, OTS, and SOTS are synthetic datasets, images in RTTS are from real scenes, and HSTS contains both synthetic and real-world images. ITS contains 13990 hazy images generated from 1399 clear images, and SOTS consists of 500 indoor hazy images and 500 outdoor hazy images. In this work, ITS and SOTS are used as training set and testing set, respectively. The settings are the same as those used in a previous method (Qu et al., 2019). In addition, some experiments are conducted on RTTS to demonstrate the generalization ability of the proposed network.

Implementation. When training the proposed network, all images are processed in the RGB space. To optimize the proposed network, the Adam (Kingma and Ba, 2014) optimizer with a batch size of 1 is adopted, where the values of β_1 and β_2 are 0.5 and 0.999, respectively. The learning rate is set as 0.0001. The hyper-parameter of the loss function is set as $\alpha = 0.01$. The proposed network is implemented with the PyTorch framework. Additionally, training and testing are also conducted on a PC with an Intel Xeon Silver 4114 CPU, 32 GB RAM, and an NVIDIA Tesla P100 GPU. For fair comparisons, the quantitative results of PSNR in this paper are calculated using the PYTHON code based on the dehazed results. $PSNR = 10\times\log_{10}(MSE)^2/MSE$, where MAX is the maximum pixel value of each image. The MSE refers to the error between a clear image and dehazed image.

Quality Measures. To evaluate the performance of the proposed network, two aspects are considered in this work: one is the objective measurement, and the other is the subjective evaluation. For the former aspect, two evaluation criteria are examined: the Peak Signal to Noise Ratio (PSNR) and the Structural Similarity index (SSIM) (Wang et al., 2004). For the latter, the proposed network is compared visually with six state-of-the-art methods on the SOTS and RTTS datasets.

4.2. Ablation study

To further demonstrate the effectiveness of the proposed PCFAN, an ablation study is conducted to verify whether all parts of the proposed PCFAN are effective. The core components of the proposed PCFAN are the pyramid channel-based feature attention module (PCFA), the channel attention block, and the Edge loss function. Therefore, an ablation study is conducted by considering the different channel attention blocks in the PCFA module and the Edge loss function. As discussed in Section 3, there are four important channel attention blocks that greatly influence the performance of PCFAN. These blocks are indicated by the purple blocks shown in Fig. 2 and marked as Att1, Att2, Att3, and Att4 (from left to right, from bottom to top). The following network variants are constructed: (1) Backbone network: the four attention blocks are removed from the PCFAN. The resulting network processes

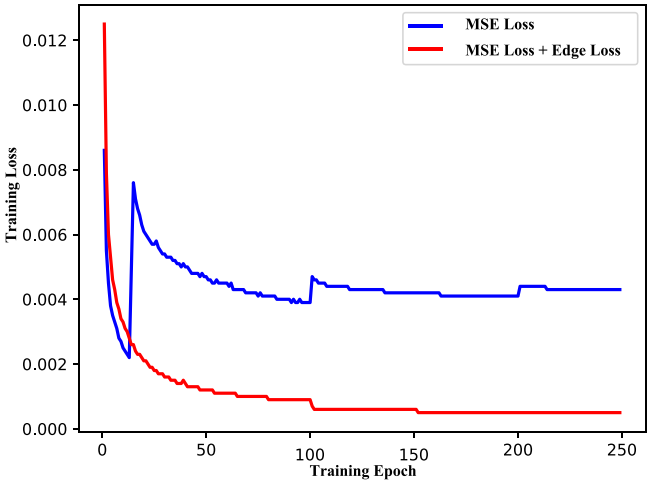


Fig. 5. The training loss of the proposed network. When training the proposed network without using Edge loss, the loss curve fluctuates and the model does not converge to a point in the parameter space. If the Edge loss is included, then the Edge loss curve is more stable and smooth, and the optimization of the network is successful.

features without attention mechanism; (2) Backbone + Att1: the first channel attention block is added into the Backbone. This helps the network only learn information features from high-level features; (3) Backbone + Att12: the first and the second channel attention blocks are added to the Backbone network; (4) Backbone + Att123: the first three channel attention blocks are added to the Backbone network; (5) PCFAN-: all channel attention blocks are added to the Backbone network, which helps the network fully exploit the complementary information between low and high level features; (6) PCFAN: this final model consists of all the channel attention blocks. It is trained with both the MSE loss and Edge loss. The network variants (1)-(5) are trained only with the MSE loss. The detailed configurations of the ablation study are listed in Table 2.

The results of the ablation study are presented in Table 3. The digital values are the image dehazing results of PSNR and SSIM on the outdoor and indoor datasets of SOTS. It is noted that PCFAN has the best performance on both the indoor and outdoor datasets among all network variants. The backbone network achieves the worst results in terms of PSNR and SSIM. The performances of the Backbone+Att1, the Backbone+Att12, and the Backbone+Att123 are improved by adding the channel attention blocks. It can be see that both considering low-level and high-level features are important for image dehazing. Moreover,

Table 1
Quantitative comparison results of the seven state-of-the-art methods and the PCFAN on the SOTS set. The best and the second best results are marked in red and blue text, respectively.

Method		DCP	DehazeNet	MSCNN	AOD-Net	DCPDN	GFN	EPDN	FAMEDNet	PCFAN
Indoor	PSNR	16.62	21.14	19.84	19.06	15.85	22.30	25.06	25.00	31.39
	SSIM	0.8179	0.8472	0.8327	0.8504	0.8175	0.8800	0.9232	0.9172	0.9868
Outdoor	PSNR	19.13	22.46	22.06	20.29	19.93	21.55	22.57	29.03	24.01
	SSIM	0.8148	0.8514	0.9078	0.8765	0.8449	0.8444	0.8630	0.9570	0.9350
–	Size	–	–	–	–	256 MB	45.6 MB	66 MB	86.3 Kb	0.9 MB

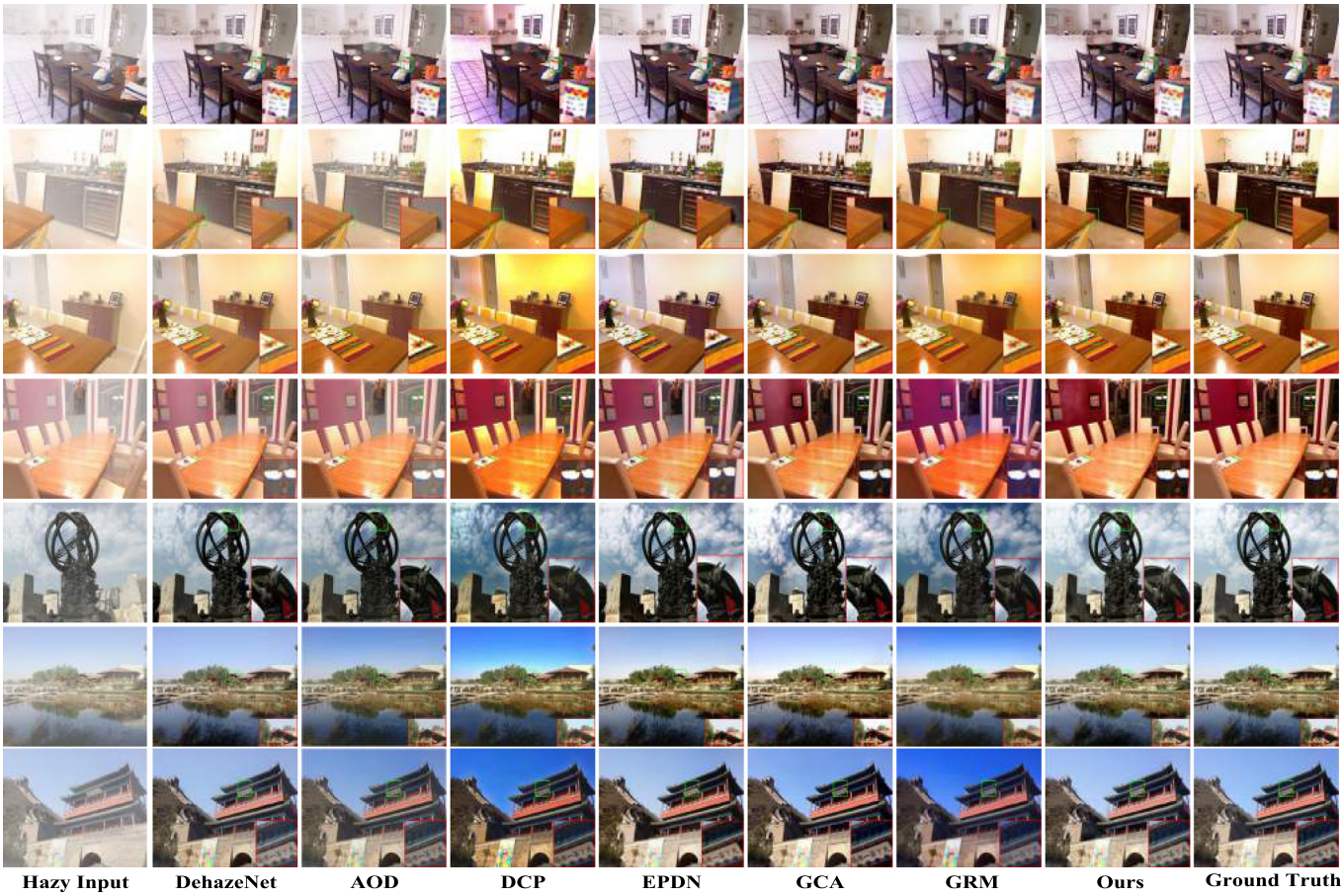


Fig. 6. Visual comparison results on the SOTS dataset. The first column presents the hazy images. The results of six representative state-of-the-art single-image dehazing methods are illustrated separately. The dehazed results of the proposed method and the ground truth images are shown in the last two columns. The upper four rows show the results of the indoor subset, while the last three rows are dehazed images of the outdoor subset. Zoom in for better visibility..

Table 2
Ablation study configurations. Note: ✓indicates that the model includes a component, and the — indicates that the model does not include a component.

Models	Att1	Att2	Att3	Att4	EdgeLoss
Backbone	–	–	–	–	–
Backbone + Att1	✓	–	–	–	–
Backbone + Att12	✓	✓	–	–	–
Backbone + Att23	✓	✓	✓	–	–
PCFAN-	✓	✓	✓	✓	–
PCFAN	✓	✓	✓	✓	✓

Table 3
Comparisons on both indoor and outdoor datasets of SOTS of variants of the proposed PCFAN.

Variants	Indoor		Outdoor	
	PSNR	SSIM	PSNR	SSIM
Backbone	27.45	0.968	21.98	0.914
Backbone + Att1	28.91	0.976	23.31	0.931
Backbone + Att12	29.11	0.977	23.14	0.937
Backbone + Att123	27.49	0.968	23.56	0.940
PCFAN-	30.80	0.971	23.68	0.940
PCFAN	31.39	0.987	24.01	0.940

the results of the outdoor dataset show that considering different channel information from different level features can improve the robustness of the model.

The influence of the Edge loss on the network is also explored. Compared with the variant PCFAN-, PCFAN achieves higher values of PSNR and SSIM. The PSNR is especially higher on both the indoor and outdoor dataset. This demonstrates that the Edge loss is vital for training the proposed PCFAN. From Fig. 5, it is clear that the training loss converges faster to a small value with the help of the Edge loss function. In addition, some attention maps of the hazy image, outputs of the PCFAN, and the ground truth are shown in Fig. 4. The attention maps in Fig. 4 display significant differences that arise from different information, such as edges, textures, and other details. It also proves that the effectiveness of the proposed PCFA. The ablation study shows that the PCFAN benefits from all of the channel attention blocks and the Edge loss.

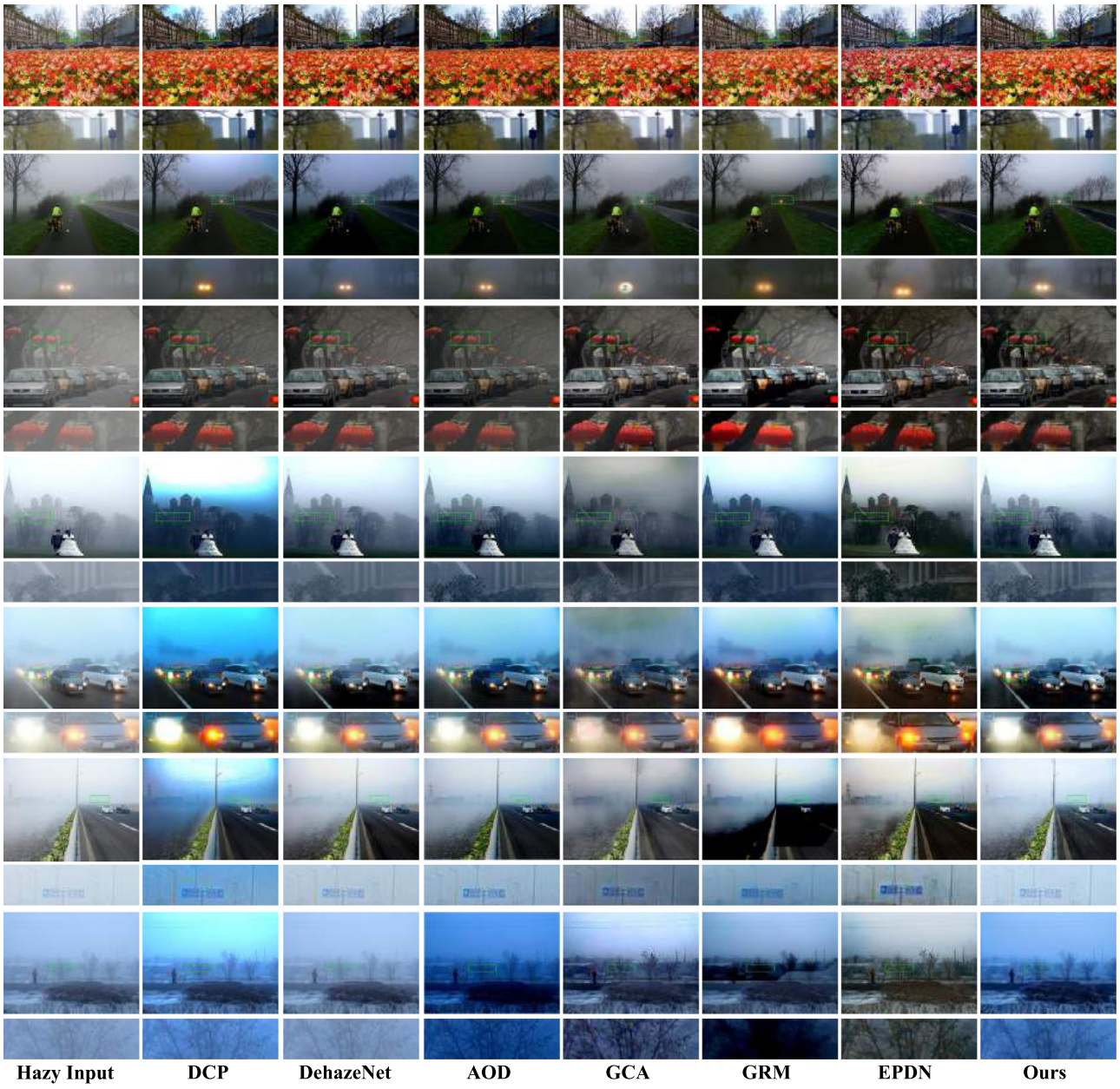


Fig. 7. Visual comparison with state-of-the-art dehazing methods on the RTTS dataset. Zoom in for better visibility..

4.3. Comparisons with state-of-the-art methods

To demonstrate the superiority of the proposed network, we also compare it with several state-of-the-art image dehazing methods on the synthetic dataset and on the real-world dataset, both quantitatively and qualitatively.

Synthetic dataset. The results of the proposed network on the synthetic dataset are compared with the state-of-the-art methods. Some methods, such as DCP, DehazeNet and GRM, first estimate the transmission map and the atmospheric light, then resort to the atmosphere scattering model to restore the dehazed image. Other methods, such as DCPDN and EPDN, directly learn a map between the hazy image and the dehazed image, and use this map to restore the dehazed image. Fig. 6. shows the qualitative comparisons of the visual effect on the indoor and outdoor datasets of SOTS. The prior-based methods, such as DCP and GRM, tend to produce darker images compared with the ground truth, as these methods often fail to accurately estimate the hazy thickness of images. Additionally, DCP and GRM suffer from the problem of color distortions, which degrade the quality of their recovered images. (*e.g.,* the building, the sky, the floor, and the chair in Fig. 6 (DCP, GRM)). For the learning-based methods, there is a greater amount of haze in the results of DehazNet and GCA. This leads to color distortions problem. Although the AOD-Net reduces color distortion, it suffers from a halo effect. (see, *e.g.,* the boundaries of the chair and building in Fig. 6 (AOD)). Although EPDN achieves better results, there remain some haze and color distortions. Compared with these methods, the proposed method achieves the best visual performance in terms of haze removal.

The quantitative comparison results are presented in Table 1, in which the digital values are the results on the SOTS database in terms of average PSNR and SSIM. The results demonstrate that the PCFAN achieves the best performance for image dehazing. Specifically, on the indoor dataset of SOTS, the PCFAN ranks first among the compared methods. As compared with the second-best method, the results of PCFAN present increments of 6.33 dB and 0.0298 in PSNR and SSIM, respectively. On the outdoor dataset of SOTS, PCFAN outperforms the

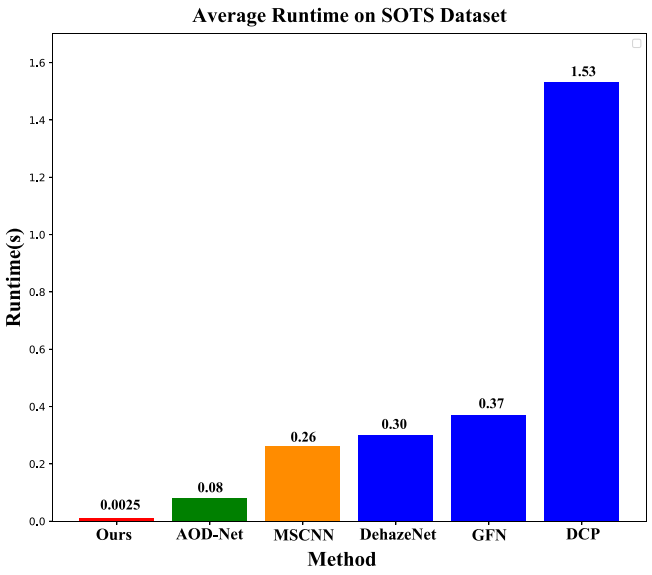


Fig. 8. Runtime comparison of different dehaizng method on SOTS dataset.

state-of-the-art methods in terms of SSIM, and ranks second in term of PSNR. Although FAMEDNet achieves the best performance on the outdoor dataset, it should be noted that it uses both ITS and OTS as training sets, which are much larger than the training set used by the proposed method. In addition, the proposed model achieves a balance between efficiency and complexity. A superior performance of image dehazing is obtained from a lightweight network.

Real-world dataset. To verify the robustness of the proposed PCFAN, additional experiments are conducted on the real-world dataset. A comparison of the visual effects on real hazy images is presented in Fig. 7. It can be observed that DCP suffers from color distortions (*e.g.,* the sky in Fig. 7). For DehazNet, AOD, GCA, haze removal is incomplete in a dense haze situation. However, the dehazed image produced by PCFAN is a bit blurred. The results of EPDN looks more natural. This is because we do not use the same training method as EPDN that is trained with generative adversarial scheme. With the help of adversarial learning, it recovers more realistic images from the real-world dataset. In general, the proposed PCFAN is more effective than existing methods in removing haze and preserving texture details.

Running-time. In addition to the superior PSNR and SSIM of the proposed model, we also show the comparison result of running time. Fig. 8 shows the average run times of different state-of-the-art methods for dehazing one image from SOTS. The proposed method is the most efficient, in that it is 32 times faster than the second one.

5. Conclusion

In this paper, we introduce a novel end-to-end dehazing network called pyramid channel-based feature attention network (PCFAN) to tackle the challenging single image dehazing problem. PCFAN consists of a three-scale extraction module, a pyramid channel-based feature attention module, and an image reconstruction module. PCFAN is able to efficiently restore the haze-free image directly. In addition, we propose a novel Edge loss to help the network learn more detailed information. The PCFAN is lightweight and can be easily put into practice. Extensive experiments on the synthetic and real-world images demonstrate the effectiveness and efficiency of the proposed PCFAN.

CRediT authorship contribution statement

Xiaoqin Zhang: Funding acquisition, Project administration, Supervision, Conceptualization, Writing - review & editing. **Tao Wang:** Methodology, Investigation, Software, Writing - original draft. **Jinxin Wang:** Software, Data curation. **Guiying Tang:** Conceptualization , Resources, Visualization, Validation. **Li Zhao:** Supervision, Writing - review, Formal analysis.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China [grant no. 61922064], in part by the Zhejiang Provincial Natural Science Foundation, China [grant nos. LR17F030001, LQ19F020005], in part by the Project of science and technology plans of Wenzhou City, China [grant nos. C20170008, G20150017, ZG2017016].

References

Alajarmeh, A., Salam, R., Abdulrahim, K., Marhusin, M., Zaidan, A., Zaidan, B., 2018. Real-time framework for image dehazing based on linear transmission and constant-time airlight estimation. Inform. Sci. 436, 108–130.

Berman, D., Avidan, S., et al., 2016. Non-local image dehazing. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. pp. 1674–1682.

Bluche, T., 2016. Joint line segmentation and transcription for end-to-end handwritten paragraph recognition. In: Proceedings of Advances in Neural Information Processing Systems. pp. 838–846.

Cai, B., Xu, X., Jia, K., Qing, C., Tao, D., 2016. Dehazenet: An end-to-end system for single image haze removal. IEEE Trans. Image Process. 5187–5198.

Cao, C., Liu, X., Yang, Y., Yu, Y., Wang, J., Wang, Z., Huang, Y., Wang, L., Huang, C., Xu, W., et al., 2015. Look and think twice: Capturing top-down visual attention with feedback convolutional neural networks. In: Proceedings of IEEE International Conference on Computer Vision. pp. 2956–2964.

Chen, C., Do, M.N., Wang, J., 2016. Robust image and video dehazing with visual artifact suppression via gradient residual minimization. In: Proceedings of European Conference on Computer Vision. Springer, pp. 576–591.

Chen, D., He, M., Fan, Q., Liao, J., Zhang, L., Hou, D., Yuan, L., Hua, G., 2019. Gated context aggregation network for image dehazing and deraining. In: Proceedings of IEEE Winter Conference on Applications of Computer Vision. IEEE, pp. 1375–1383.

Fu, J., Liu, J., Tian, H., Li, Y., Bao, Y., Fang, Z., Lu, H., 2019. Dual attention network for scene segmentation. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. pp. 3146–3154.

Girshick, R., 2015. Fast r-cnn. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 1440–1448.

He, K., Sun, J., Tang, X., 2010. Single image haze removal using dark channel prior. IEEE Trans. Pattern Anal. Mach. Intell. 2341–2353.

He, K., Zhang, X., Ren, S., Sun, J., 2015. Spatial pyramid pooling in deep convolutional networks for visual recognition. IEEE Trans. Pattern Anal. Mach. Intell. 37 (9), 1904–1916.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. pp. 770–778.

Hodges, C., Bennamoun, M., Rahmani, H., 2019. Single image dehazing using deep neural networks. Pattern Recognit. Lett. 128, 70–77.

Itti, L., Koch, C., Niebur, E., 1998. A model of saliency-based visual attention for rapid scene analysis. IEEE Trans. Pattern Anal. Mach. Intell. 1254–1259.

Jaderberg, M., Simonyan, K., Zisserman, A., et al., 2015. Spatial transformer networks. In: Proceedings of Advances in Neural Information Processing Systems. pp. 2017–2025.

Kingma, D.P., Ba, J., 2014. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. In: Proceedings of Advances in Neural Information Processing Systems. pp. 1097–1105.

Kupyn, O., Budzan, V., Mykhailych, M., Mishkin, D., Matas, J., 2018. Deblurgan: Blind motion deblurring using conditional adversarial networks. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. pp. 8183–8192.

Li, B., Peng, X., Wang, Z., Xu, J., Feng, D., 2017. Aod-net: All-in-one dehazing network. In: Proceedings of IEEE International Conference on Computer Vision. pp. 4770–4778.

Li, B., Ren, W., Fu, D., Tao, D., Feng, D., Zeng, W., Wang, Z., 2018a. Benchmarking single-image dehazing and beyond. IEEE Trans. Image Process. 28 (1), 492–505.

Li, B., Yan, J., Wu, W., Zhu, Z., Hu, X., 2018b. High performance visual tracking with siamese region proposal network. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. pp. 8971–8980.

Liu, X., Ma, Y., Shi, Z., Chen, J., 2019. GridDehazeNet: Attention-based multi-scale network for image dehazing. In: Proceedings of IEEE International Conference on Computer Vision. pp. 7314–7323.

Liu, S., Sun, Y., Zhu, D., Ren, G., Chen, Y., Feng, J., Han, J., 2018. Cross-domain human parsing via adversarial feature and label adaptation. In: Proceedings of AAAI Conference on Artificial Intelligence.

Mnih, V., Heess, N., Graves, A., et al., 2014. Recurrent models of visual attention. In: Proceedings of Advances in Neural Information Processing Systems. pp. 2204–2212.

Narasimhan, S.G., Nayar, S.K., 2002. Vision and the atmosphere. Proc. Int. J. Comput. Vision 48 (3), 233–254.

Qu, Y., Chen, Y., Huang, J., Xie, Y., 2019. Enhanced pix2pix dehazing network. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. pp. 8160–8168.

Ren, W., Liu, S., Zhang, H., Pan, J., Cao, X., Yang, M.-H., 2016. Single image dehazing via multi-scale convolutional neural networks. In: Proceedings of European Conference on Computer Vision. pp. 154–169.

Trudinger, N.S., 1983. Elliptic Partial Differential Equations of Second Order. Springer-Verlag.

Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P., et al., 2004. Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process. 13 (4), 600–612.

Yuan, Y., Liang, X., Wang, X., Yeung, D.-Y., Gupta, A., 2017. Temporal dynamic graph LSTM for action-driven video object detection. In: Proceedings of IEEE International Conference on Computer Vision. pp. 1801–1810.

Zhang, X., Hu, W., Xie, N., Bao, H., Maybank, S., 2015. A robust tracking system for low frame rate video. Int. J. Comput. Vis. 115, 279–304.

Zhang, X., Jiang, R., Wang, T., Huang, P., Zhao, L., 2020a. Attention-based interpolation network for video deblurring. Neurocomputing.

Zhang, Y., Li, K., Li, K., Wang, L., Zhong, B., Fu, Y., 2018. Image super-resolution using very deep residual channel attention networks. In: Proceedings of European Conference on Computer Vision. pp. 286–301.

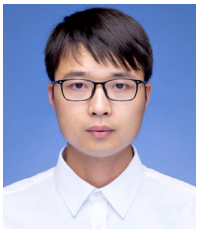
Zhang, H., Patel, V.M., 2018. Densely connected pyramid dehazing network. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. pp. 3194–3203.

Zhang, X., Wang, D., Zhou, Z., Ma, Y., 2020b. Robust low-rank tensor recovery with rectification and alignment. IEEE Trans. Pattern Anal. Mach. Intell..

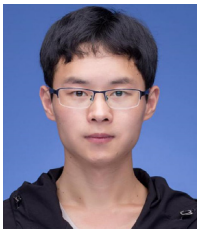
Zhao, D., Xu, L., Yan, Y., Chen, J., Duan, L.-Y., 2019. Multi-scale optimal fusion model for single image dehazing. Signal Process., Image Commun. 74, 253–265.

Zhu, Q., Mai, J., Shao, L., 2015. A fast single image haze removal algorithm using color attenuation prior. IEEE Trans. Image Process. 3522–3533.

Xiaoqin Zhang received the B.Sc. degree in electronic information science and technology from Central South University, China, in 2005 and Ph.D. degree in pattern recognition and intelligent system from the National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, China, in 2010. He is currently a professor in Wenzhou University, China. His research interests are in pattern recognition, computer vision and machine learning. He has published more than 80 papers in international and national journals, and international conferences, including IEEE T-PAMI, IJCV, IEEE T-IP, IEEE T-IE, IEEE T-C, ICCV, CVPR, NIPS, IJCAI, AAAI, and among others.



Tao Wang is currently a graduate student at College of Computer Science and Artificial Intelligence, Wenzhou University, China. He received the B.Sc. degree in information and computing science from Hainan Normal University, China, in 2018. His research interests include several topics in computer vision and machine learning, such as object tracking, image/video quality restoration, adversarial learning, image-to-image translation and reinforcement learning.



Jinxin Wang is currently a graduate student at College of Computer Science and Artificial Intelligence, Wenzhou University, China. He received his bachelor's degree in information and computing science at Wenzhou University. His research interests include visual tracking, image generation and deep learning.

Guiying Tang is currently a graduate student at College of Mathematics and Physics, Wenzhou University, China. She received the B.Sc. degree in the College of Mathematics and Software Science, Sichuan Normal University, China, in 2017. Her main research interest is in computer vision and deep learning, such as image quality restoration, object tracking.

Li Zhao received the B.Sc. degree in automation in 2005 and MEng degree in control theory and control engineering in 2008 from Central South University, China. She is currently an assistant researcher in Wenzhou University. Her research interests are in pattern recognition, computer vision, and machine learning.