FSAD-Net: Feedback Spatial Attention Dehazing Network

Yu Zhou, Zhihua Chen, Ping Li, *Member, IEEE*, Haitao Song, C. L. Philip Chen, *Fellow, IEEE*, and Bin Sheng, *Member, IEEE*

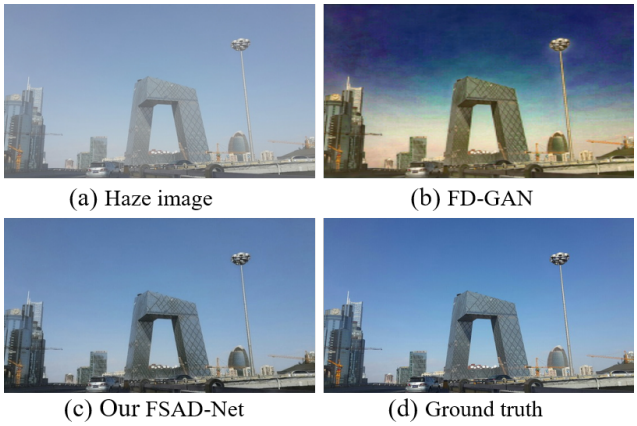
***Abstract*—Recent dehazing networks learn more discriminative high-level features by designing deeper networks or introducing complicated structures, while ignore inherent feature correlations in intermediate layers. In this paper, we establish a novel and effective end-to-end dehazing method, named feedback spatial attention dehazing network (FSAD-Net). The FSAD-Net is based on the recurrent structure and consists of four modules: a shallow feature extraction block (SFEB), a feedback block (FB), multiple advanced residual blocks (ARBs), and a reconstruction block (RB). FB is designed to handle the feedback connections, and it can improve the dehazing performance by exploiting the dependencies of deep features across stages. The ARB implements a novel attention-based estimation on a residual block to adapt to pixels with different distributions. Finally, RB helps to restore the haze-free images. It can be seen from the experimental results that the FSAD-Net almost outperforms the state-of-the-arts in terms of five quantitative metrics. Moreover, the qualitatively comparisons on real-world images also demonstrate the superiority of the proposed FSAD-Net. Considering the efficiency and effectiveness of FSAD-Net, it can be expected to serve as a suitable image dehazing baseline in the future.**

***Index Terms*—Image dehazing, recurrent structure, spatial attention mechanism, dehazing network.**

I. Introduction

HAZE is a commonly atmospheric phenomenon, caused by turbid medium suspended in air, like dust, mist, fumes and water droplets, which often leads to captured images have poor visibility [1], [2]. The degradation of the image quality existing in haze images potentially challenges many vision-based applications [3] (such as segmentation [4], detection [5] and recognition of object [6], traffic monitoring, advanced driving assistance system etc.). Therefore, haze removal is a practical preprocessing method to benefit the performance of various computer vision systems. Image haze removal, known as image dehazing, is a technique that aims to restore a clear haze-free image from a single or multiple haze images. In atmospheric scattering model [7] widely used to describe the formation of a haze image, the amount of scattering depends on two unknown components: the medium transmission and the atmospheric light. Since there are more unknown values than known values, the restoration of haze-free image is an under-constraint problem. Many algorithms have been proposed, which exploited either handcrafted features (e.g. reasonable assumptions) [8] or learning-based features [9] to solve the ill-posed problem in recent years. Handcrafted features used simple linear-mapping transformations and were not robust to variations of the input, e.g., images with various haze densities, depths. Features extracted by deep networks improved the limitations of handcrafted features, and achieved promising dehazing performances.

Although early deep learning-based dehazing methods (e.g. MSCNN [9], DehazeNet [11]) have distinct advantages in providing a uniform feature extraction framework to free users from troublesome handcrafted feature extraction, most CNN-



(a) Haze image (b) FD-GAN (c) Our FSAD-Net (d) Ground truth

Fig. 1: A haze image and the results recovered by different dehazing methods. (a) Haze image, dehazing results by (b) FD-GAN [10] and (c) our FSAD-Net, respectively. (d) Ground truth. Our recurrent structure-based method generates cleaner images with fewer artifacts and color distortion. The basic idea of recurrent structure is to make predictions in an iterative manner. This provides a core advantage of enabling optimization of the final dehazed result by iterations.

Manuscript received September 26, 2020; revised November 09, 2021; accepted January 22, 2022. This work was supported in part by the National Natural Science Foundation of China under Grant 61672228, Grant 61872241, and Grant 61572316, in part by the National Key Research and Development Program of China under Grant 2019YFB1703600, and in part by The Hong Kong Polytechnic University under Grant P0030419, Grant P0030929, and Grant P0035358. *(Corresponding authors: Zhihua Chen, Bin Sheng.)*

Y. Zhou and Z. Chen are with the Department of Computer Science and Engineering, East China University of Science and Technology, Shanghai 200237, China (Email: czh@ecust.edu.cn).

P. Li is with the Department of Computing, The Hong Kong Polytechnic University, Kowloon, Hong Kong (Email: p.li@polyu.edu.hk).

H. Song is with the Artificial Intelligence Institute, Shanghai Jiao Tong University, Shanghai 200240, China (Email: allen5@sjtu.edu.cn).

C. L. P. Chen is with the School of Computer Science and Engineering, South China University of Technology, Guangzhou 510006, China; with the Navigation College, Dalian Maritime University, Dalian 116026, China; and also with the Faculty of Science and Technology, University of Macau, Macau 999078, China (Email: philip.chen@ieee.org).

B. Sheng is with the Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai 200240, China (Email: shengbin@sjtu.edu.cn).

based haze removal methods lack the considerations of the information from the original haze images and the inherent feature correlations in intermediate layers, leading relatively low dehazing performances. Some methods hoped to learn more discriminative high-level features by designing deeper or wider networks while rarely focused on the inherent feature correlations in intermediate layers. As Zeiler and Fergus [12] pointed, it does not each layer of CNNs contribute for robust feature extraction nor more number filters would guarantee better system performances. These methods occupied huge storage resources and hindered the representational ability of CNNs. Some methods introduced complicated structures in their models, with more and more complicated structures are introduced in image dehazing networks, it is difficult to analyze contributions of different modules such as dense blocks [13], dilated convolution [14], residual blocks [15], squeeze-and-excitation [16], and generative adversarial networks (GAN) [17] in a dehazing network.

In this paper, we aim to propose a network with simple baseline and higher computational efficiency that can ease the difficulty of training CNNs and exploit the feature correlations. To reduce network parameters, the recurrent structure is often employed, which can share the information through feedback connections [18], [19]. In the field of cognition, feedback connections as a core role which link the cortical visual areas, and transmit response signals of higher-order areas to lower-order areas [20]. Recently, some studies have applied the feedback connections in their architectures to carry high-level features back to previous layers and refine low-level encoded information. Motivated by these structures, we establish a fully end-to-end FSAD-Net, which introduces a feedback block for passing the high-level information back to previous layers to correct feature representations. Moreover, inspired by the success of the attention mechanism in image restoration task, we incorporate the spatial attention mechanism to residual block to exploit feature correlations through spatial information.

Comprehensive experiments have been conducted to evaluate different image dehazing networks. In addition to PSNR (peak-signal-to-noise-ratio) and SSIM (structure similarity index), the two commonly used to quantify dehazed image restoration quality, we also use CIEDE2000, UQI (Universal Quality Index) [21] and LPIPS (Learned Perceptual Image Patch Similarity) [22] for evaluating dehazing results. On the widely used dehazing benchmark datasets RESIDE [23] and HZERD [24], our FSAD-Net surpasses most previous methods in both quantitative and qualitative dehazing results. The visually pleasing dehazing results on real haze images have also proved the generalization ability of the FSAD-Net model. Fig. 1 shows the visual effect of our method. Although FD-GAN [10] can remove haze relatively effectively, haloes and artifacts are not avoided in bright regions. Compared with FD-GAN [10], the proposed FSAD-Net can remove most haze, generating promising dehazing quality in terms of color and details. In sum, our approach has the following contributions:

- We propose a novel network for image dehazing named FSAD-Net, which based on recurrent structure and does not rely on the atmosphere scattering model. Our FSAD-Net can unfold into multiple iterations through recurrent structure, and improve the dehazing performances iteration-by-iteration.
- We introduce the feedback block (FB) to our dehazing network to learn the dependencies of features. The feedback connection passes hidden state from previous iteration into FB at current iteration to correct middle-level representations.
- We propose advanced residual blocks (ARBs) by introducing spatial attention mechanism and skip fusion into residual blocks. The ARB can focus on the inherent spatial correlations of features, and enhance discriminative learning ability of network in dealing with different types of information.

It is worthy of note that some methods had already introduced the recursive architecture and iteration algorithm in their dehazing networks. Liu *et al.* [1] proposed learning aggregated transmission propagation networks (DPATN) for haze removal. Liu *et al.* [25] proposed an iteration algorithm with existing CNNs to learn deep priors for image dehazing. However, our FSAD-Net is inherently different from them in following aspects. First, both DPATN [1] and LDP [25] are based on atmospheric scattering model. DPATN proposed an aggregated residual architecture for propagating transmission, meanwhile it estimated atmospheric light through improving DCP. LDP developed existing CNNs to learn the image prior terms of atmospheric light, transmission and clear image. Such method of separately estimating the unknown parameters in the atmospheric scattering model may lead to accumulation of errors. Our FSAD-Net is designed for directly estimating a haze-free image from a haze image and can avoid error accumulation. Second, DPATN [1] and LDP [25] did not utilize attention mechanism in their architecture, while our model takes advantage of spatial attention mechanism for making effective use of the haze-relevant features, the ablation studies on spatial attention block show the effectiveness of spatial attention mechanism. The structure of this paper is organized as follows. We first briefly review some dehazing methods and make a simple classification and summary in Section II. In Section III, we present the details of the proposed FSAD-Net. Section IV describes experimental results as well as comparisons with the state-of-the-art dehazing methods. At last, we make the conclusion in Section V.

II. Related Work

In this section, we discuss the prior-based and learning-based single image dehazing methods.

A. Prior-Based Dehazing

Based on observation that a haze-free image must have higher contrast compared with the input hazy image, Tan [26] enhanced the visibility of a haze image by maximizing its local contrast. Fattal [27] estimated the transmission map according to the prior knowledge that there is no locally statistically correlation between object surface shading and transmission map. He *et al.* [8] proposed a dark channel prior (DCP) based on the statistical observation that in majority of non-sky regions in clear images, the minimum of all the color channels

has extremely low intensities in some pixels. This method combined with a soft-matting operation, effectively estimates the transmission map and atmospheric light. However, the DCP may fail in some particular regions of which the intensity are similar to atmospheric light, leading the underestimation of transmission. Thus, a number of methods based on DCP are proposed to improve this problem. Liu *et al.* [28] reformulated the image dehazing as a trade off between luminance and contrast in order to achieve contrast values superior to that obtained using other methods for a given brightness level. Peng *et al.* [29] removed haze by using the depth-dependent color change and the difference between the observed intensity and the ambient light. Tang *et al.* [30] trained a regression model based on random forests from a variety of multi-scale haze-relevant features including dark channel feature, local max contrast, hue disparity and local max saturation to obtain an accurate transmission map. However, this feature fusion method relied largely on the dark channel features. Zhu *et al.* [31] proposed a color attenuation prior (CAP) by making use of HSV color space. This method assumed that saturation of the image patch is inversely proportional to the density of haze to obtain an accurate transmission map using likelihood learning. From statistical analysis, Chio *et al.* [32] used natural scene statistics (NSS) and fog aware features of hazy images to enhance the visibility of haze images and achieved the assessment of existing defogging algorithms. Chen *et al.* [33] introduced a gradient residual and error layer into the image recovery process with minimizing possible visual artifacts in it. After observing that colors of a haze-free image can be well approximated by a few hundred distinct colors, Berman *et al.* [34] proposed a non-local image dehazing method based on this prior. Zhang *et al.* [35] proposed a novel and effective maximum reflectance prior for Nighttime image dehazing, which assumes in majority of daytime haze-free image patches, each color channel has very high intensities in some pixels. Bui and Kim [36] fitted a haze pixel cluster in RGB space to a statistical color ellipsoid, and this method is robust to image signal randomness. Wu *et al.* [37] also proposed a joint framework to avoid the weakness of noise sensitivity when recovering the scene radiance and transmission map by introducing a transmission-aware nonlocal regularization and a semantic-guided regularization.

B. Learning-Based Dehazing

Ren *et al.* [9] estimated the transmission map by proposing a multi-scale convolutional neural network (MSCNN), which consisted of a coarse-scale network and a fine-scale network for coarse-to-fine regression of transmission map. Cai *et al.* [11] proposed an trainable architecture called DehazeNet for transmission estimation. Rather than estimating the intermediate transmission and the atmospheric light separately, Li *et al.* [38] proposed an end-to-end CNN model called the all-in-one dehazing network (AOD-Net) based on a reformulated atmospheric scattering model for learning the clear image from a haze one. Li *et al.* [17] designed an encoder-decoder architecture based on a conditional generative adversarial network (cGAN), which is an end-to-end model that generates realistic haze-free images. Zhang and Patel [13] proposed a densely connected pyramid dehazing network (DCPDN) to learn the transmission map, atmospheric light, and dehazed image jointly. They also exploited an adversarial loss based on a GAN to supervise their encoder-decoder dehazing architecture. To address the dehazing problem, Ren *et al.* [39] recommended a fusion-wise strategy using an encoder-decoder architecture called gated fusion network (GFN), this method applied white balance, contrast enhancing, and Gamma correction to an original image to derive three input images. Yang and Sun [40] proposed the proximal dehaze-net, which combined the haze imaging model, the DCP, and the transmission prior by building the energy function. Chen *et al.* [41] designed a multi-scale U-module architecture named patch map selection network (PMS-Net) to adaptively choose the patch size corresponding to each pixel when dehazing. Qu *et al.* [42] proposed an Enhanced Pix2pix Dehazing Network (EPDN), which restored haze-free images without relying on the atmospheric scattering model. Ren *et al.* [43] proposed a new multi-scale CNN for single image dehazing based on their preliminary work, and introduced a novel holistic edge guided network for refinement of edge of the estimated transmission map. Dong *et al.* [10] proposed a GAN with fusion-discriminator (FD-GAN) for image dehazing, which made effective use of GAN and prior knowledge. Liu *et al.* [1] designed a novel residual architecture to aggregate both prior and data information to propagate transmissions. Li *et al.* [44] proposed a level-aware progressive network (LAP-Net) for single image dehazing, which can progressively learn the gradually aggravating haze. Hong *et al.* [45] proposed a knowledged-distill dehazing network, which contained a teacher and student. The teacher network was proposed for image reconstruction, the student network imitated the task of image reconstruction and achieved image dehazing. Our approach also adopts the deep CNN for image dehazing. In contrast to those above methods, the proposed FSAD-Net is specifically devised through the feedback connections for single image dehazing, from which the haze can be gradually removed iteration-by-iteration.

C. Attention Mechanism

The important idea of the attention mechanism is to guide the allocation of available computational resources to the most useful and informative features and ignore the less informative features [16]. Recent works has introduced the attention mechanism into deep learning frameworks to carry different computer vision tasks such as image classification [46] and image processing [47], [48]. Hu *et al.* [16] proposed the squeeze-and-excitation (SE) block to boost the representational capacity of a deep learning model via assigning different weights to channel-wise features. Cao *et al.* [49] solved the problem of cross-modal recipe retrieval through parallel- and cross-attention networks learning. Yan *et al.* [50] proposed spatial-temporal attention mechanism (STAT) for video captioning, which made full use of the spatial and temporal structures in an input video. STAT first used the 2D-CNN, 3D-CNN and region-based CNN as encoder to extract three

types of features, then fused these features via spatial attention mechanism and temporal attention mechanism to indicate the characteristics of key frames, and finally utilized LSTM to generate sentences. Shu *et al.* [51] proposed a skeleton-joint co-attention recurrent neural networks (SC-RNN) for human motion prediction, SC-RNN proposed a novel spatiotemporal co-attention (SCA) to learn the joints-relevant spatial-attention factors and the skeletons-relevant temporal-attention factors. Hu *et al.* [52] established a channel-wise and spatial feature modulation (CSFM) network for image super-resolution. In the CSFM, the channel-wise and spatial attention residual block can adaptively capture more informative features. Li *et al.* [53] proposed a single image deraining network, which was based on recurrent neural networks and squeeze-and-excitation blocks. Liu *et al.* [47] proposed an attention-based multi-scale network called GridDehazeNet to remove haze. Qin *et al.* [15] proposed a feature fusion attention network (FFA-Net) to restore a clear image from a haze image. Although STAT [50] and SC-RNN [51] are methods based on the recurrent structure and the attention mechanism, our method is different from them. Specifically, since their methods are based on video sequences, their attention mechanisms contain spatial attention mechanism and temporal attention mechanism. Our method is proposed for single image, and our FSAD-Net only include the spatial attention mechanism.

III. The Proposed Method

The development of CNNs based computer vision made the effects of image dehazing a truly breakthrough. Traditional CNNs consist of many stacked feedforward layers, imitating the bottom-up path of the human visual cortex, where the features flow from the shallower layers to deeper ones. The shallow layers tend to learn biologically plausible feature, while deeper ones try to respond to concrete visual objects, with the layers of CNNs become deeper, the more abstract representations of the input data are learned. The connections in feedforward systems learn unidirectional feature flow, the nolinear mapping is not directly influenced by the generated output, which hindering the restoration ability to some extent. Unlike the conventional feed-forward networks, feedback systems are adopted to influence the input based on rerouting the output of the system. The feedback mechanism in deep networks aims to refine the low level features by propagating deep features to the shallow layers. With the help of high-level information, low-level features become more representative and informative. It has been widely applied for various high-level vision tasks but has rarely exploited in image dehazing.

Iterativeness and propagating the high-level information extracted from deep layers to shallow ones are two core requirements in a feedback architecture [54]. As shown in Fig. 2, the output of the feedback block travels back to guide its input helps to apply the principle of the feedback scheme to image dehazing: the deep representation propagated by recurrent states can guide a haze image to restore a better dehazing image. In this section, we mainly introduce our feedback spatial attention dehazing network, the end-to-end network is proposed by considering network architecture, input and output, and loss functions, aims to combine only a few simple modules for single image dehazing.

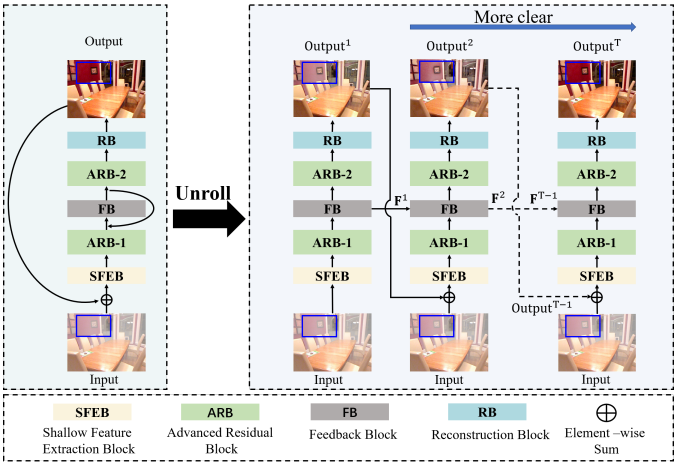


Fig. 2: Overview of the proposed FSAD-Net. The FSAD-Net consists of a shallow feature extraction block (SFEB), multiple advanced residual blocks (ARBs), a feedback block (FB), and a reconstruction block (RB), and can run for many iterations for a single training example. In our network, A feedback block is implemented using convolutional LSTM, whose output is fed back to guide its input. The feedback block receives the output of ARB-1 and hidden state F^{t-1} from last iteration, and then passes its hidden state F^t to the next iteration and ARB-2. Such recurrent structure with feedback connections provides strong dehazing ability.

A. Feedback Spatial Attention Dehazing Net

To meet the demand of feedback network, our proposed FSAD-Net is naturally designed as a convolutional recurrent network, and can be unrolled to *T* time steps, in which each iteration *t* is temporally ordered from 1 to *T*. Besides, at each time step, the sub-network can be regarded as an independent network which aims at restoring haze-free image using an input haze image. As can be seen in Fig. 2, the sub-network in each iteration *t* mainly consists of four parts: a shallow feature extraction block (SFEB), multiple advanced residual blocks (ARBs), a feedback block (FB), and a reconstruction block (RB). The weight set of each block are shared across time. The communication between two iterations from *t*-1 to *t* is a forward pass in time achieved by feedback connections. We adopt global residual learning in the sub-network at each iteration *t* to restore a residual image while input a haze image.

The SFEB is responsible for extracting the shallow feature information F_{sf}^t from the original haze image I at the iteration *t*:

$$F_{sf}^t = f_{SFEB}(O^{t-1}, I) \tag{1}$$

where f_{SFEB} denotes the operations of the SFEB, it contains a 3×3 convolutional layer and an ReLu activation. We note that f_{SFEB} takes the concatenation of the stage-wise estimation O^{t-1} and haze image I as input. The combination of input haze image can further improve the dehazing performance.

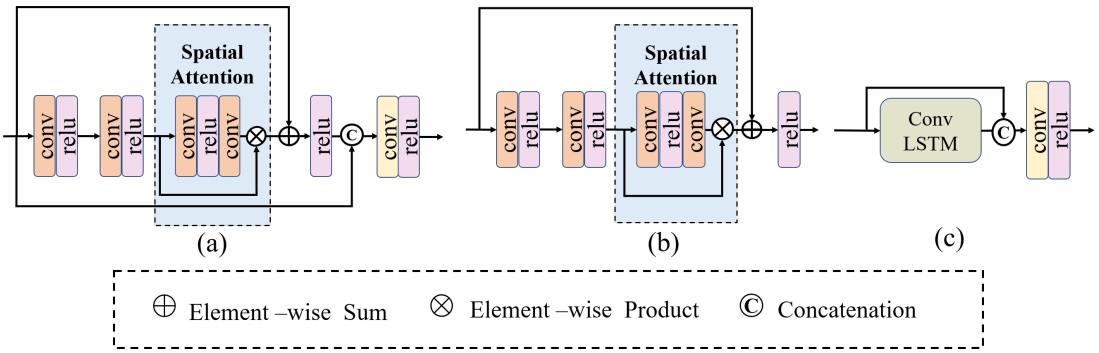


Fig. 3: The structure of the basic unit of FSAD-Net. (a) RFFB. (b) RSAB. (c) FB. RFFB combines the skip fusion and SAB module. SAB is a spatial attention mechanism and the goal is to increase representation power by focusing on important informative features (e.g. dense-hazed patches and high-frequency image regions) and suppressing unnecessary ones. The FB consists of skip fusion and ConvLSTM.

Then the ARB-1 takes the shallow feature F_{sf}^t as input to extract middle representation:

$$F_{in}^t = f_{ARB-1}(F_{sf}^t) \tag{2}$$

where f_{ARB-1} represents the operations of the ARB-1, F_{in}^t represents the middle-level features and is fed to the FB, combined with the output of the FB from the previous iteration t-1 through a feedback connection. The mathematical operations of the FB is f_{fb}, and the process of feedback can be presented as:

$$F_{fb}^t = f_{fb}(F_{in}^t, F_{fb}^{t-1}) \tag{3}$$

where F_{fb}^t denotes the output of the FB at iteration t. At t = 1, when there is no feedback, the hidden state is initialised with zero.

The ARB-2 receives the feature maps F_{fb}^t learned from FB to learn the discriminative global features. The generation of high-level features at every iteration is as below:

$$F_{gf}^t = f_{ARB-2}(F_{fb}^t) \tag{4}$$

where F_{gf}^t denotes the global features.

In reconstruction block, the extracted high-level features recover a residual image through the operations of RB. Then, the estimated residual image is added to the original haze image I using a global residual skip connection to reconstruct the clear image O^t at the t-th time stage. The mathematical formulation of the reconstruction block can be expressed as below:

$$O^t = f_{rb}(F_{gf}^t) + I \tag{5}$$

where f_{rb} denotes the function of the RB.

B. Advanced Residual Block

In this section we give more details about ARB, which composed of residual feature fusion block (RFFB) and residual spatial attention block (RSAB). As Fig. 3 shows, an important unit of RFFB and RSAB is the spatial attention block (SAB). To pay more attention to informative features of various haze image patches (e.g. dense-hazed patches and high-frequency image regions), we integrate a spatial attention block (SAB) into a basic ResNet to produce RSAB. Attention not only tells the deep learning model where to focus, but also improves the representation of interests [55], [56]. The significance of attention has been studied extensively in the previous literatures [57]. Our SAB is designed for increasing representation power by using attention mechanism, it focuses on important haze-relevant features and suppresses unnecessary ones. The input features of SAB are first passed through two convolution layers with ReLu and sigmoid activation function to produce a spatial descriptor. Then, the output of SAB obtained by utilizing element-wise multiplication for input features and the spatial descriptor. Since the haze distribution is nonuniform on the different image regions, the introduction of SAB can assist the module adapt to different image patches. Considering the SAB is the major part of RFFB and RSAB for feature extracting, we compared FSAD-Nets with and without spatial attention block in Fig. 4 to prove the effective of the SAB. Experimental results show that the SAB can help the model

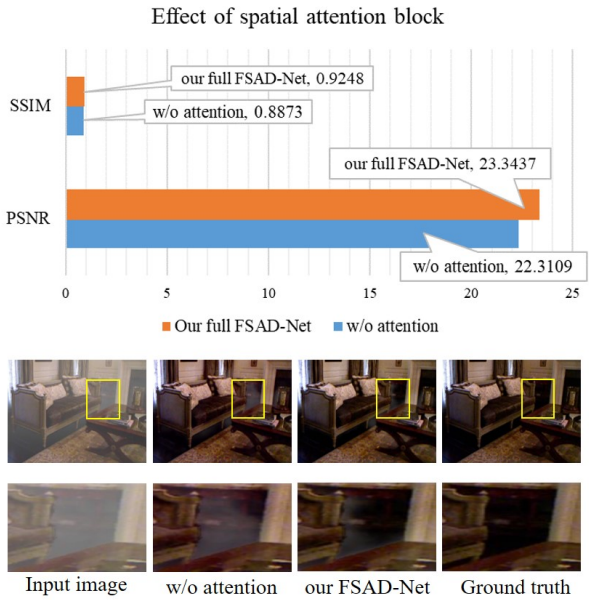


Fig. 4: Comparison of FSAD-Net models with and without spatial attention block (SAB).

(a) Haze image (b) w/o ARB-1 nor ARB-2 (c) ARB-1 only (d) ARB-2 only (e) Our full FSAD-Net (f) Ground truth

Fig. 5: Qualitative comparison of FSAD-Nets with different ARBs on synthetic indoor image. (a) Haze image, haze-free images restored by (b) FSAD-Net without ARB-1 nor ARB-2, (c) FSAD-Net with ARB-1 only, (d) FSAD-Net with ARB-2 only, and (e) our full FSAD-Net. (f) Ground truth.

to pay more attentions to more useful components.

Enlightened by the powerful feature representational ability of residual learning in computer vision tasks [46], we designed the advanced residual block (ARB) for feature extracting. The skip connection in residual block can transmit abundant low-frequency features to next block, while ARB's spatial attention can extract the spatial information in pixels to further emphasis more informative components in an input haze image. In our network, f_{ARB-1} contains a RFFB and a RSAB, f_{ARB-2} includes two RFFBs. As Fig. 3(b) shows, the first two layers of our RSAB hold 3×3 sized convolutional kernels followed by a SAB. To go a further step to residual learning, the RSAB's input is added to the output of SAB, then is fed into ReLu activation. As Fig. 3(a) shows, RFFB concatenates input features with the output of RSAB and applies a 1×1 compression convolution layer for feature fusion.

We also analyze the effect of each ARB, Fig. 5 and Table I show the qualitative and quantitative comparison results, respectively. The FSAD-Net without ARB-1 nor ARB-2 has the lowest PSNR and SSIM, as Fig. 5(b) shows, the color of restored image obtained by FSAD-Net without using any ARB tends to yellow. It is obvious that ARBs can extract more informative features such as color and structure information. Similarly, when the FSAD-Net only contains a single ARB-1 or ARB-2, the dehazed images will lose details as red rectangles in Fig. 5(c) and Fig. 5(d) denoted. The quantitative results and visual example prove the introduction of ARB-1 and ARB-2 will improve the representational power of our model and produce much better texture details.

TABLE I: Average PSNR and SSIM of FSAD-Nets with different ARBs.

Model	**PSNR**	**SSIM**
w/o ARB-1 nor ARB-2	22.3162	0.8747
ARB-1 only	23.1464	0.9118
ARB-2 only	23.1550	0.8982
Our full FSAD-Net	**23.3437**	**0.9248**

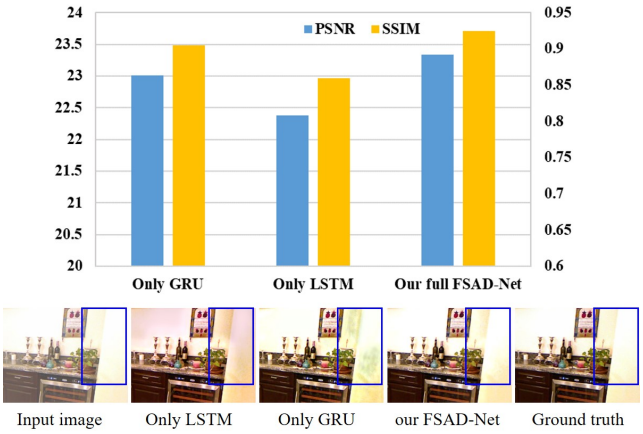


Fig. 6: Comparison of FSAD-Net models with different recurrent structures.

C. Feedback Block

Fig. 3(c) shows the feedback block, f_{fb} is implemented using convolutional LSTM, all the convolutions in which have 32 input channels and 32 output channels. f_{fb} at t iteration receives the concatenation of the feedback information F_{fb}^{t-1} and middle-level features F_{in}^{t} as input, and then passes more powerful representations F_{fb}^{t} as its output to the RB and next iteration. The feedback connection can use hidden state from previous iteration to correct middle-level representations at current stage. f_{rb} uses a RSAB and a 1-layer convolution to reconstruct the residual image. Although recurrent frameworks with unit ConvLSTM [57], ConvGRU [58] have been employed in many tasks due to their good performance, our FSAD-Net based on residual learning, spatial attention mechanism and recurrent structure have its own merits. Our ARB with attention mechanism and residual learning can improve the feature extraction ability of the model. The FB helps the information flow across iteration. To reveal the difference between ConvLSTM, ConvGRU and our FSAD-Net, we perform investigations on different recurrent structures. Fig. 6 show the comparison results, "only GRU" and "only LSTM" are two shallow residual networks using the ConvLSTM and ConvGRU as feedback block, respectively. From the quantitative results shown in the upper of Fig. 6, we can conclude that our full FSAD-Net with FB and SAB performs better than "only GRU" and "only LSTM".

D. Loss Function

Some loss functions have been widely used for training dehazing networks, such as perceptual loss, GAN loss, L_2 loss, L_1 loss. Some methods also exploited hybrid loss functions, however, the composition of different loss will increase the

burden of hyper-parameter tuning. Different with the complex loss, our FSAD-Net benefited from Feedback block will be optimized with a concise loss. It is commonly acknowledged that the Euclidean loss (L_2 loss) tends to blur the final result [13], [47]. Hence, inaccurate estimation of the restored image using the L_2 loss may result in the loss of details, leading to the artifacts in the restored color image, the L_1 can efficiently improve this issue. For a model with *T* iterations, we have *T* dehazed images. We impose supervision on the output O^t, and the L_1 loss can be written as:

$$L_1 = \sum_{i=1}^{w}\sum_{j=1}^{h} ||O^t(i,j) - I^{gt}(i,j)||_1 \tag{6}$$

where (*i*,*j*) denotes pixel location, I^{gt} is the corresponding ground truth, *w* and *h* denote width and height of the image, respectively.

We also adopt the negative SSIM to optimize our architecture:

$$L_{ssim} = -\frac{(2\mu_{O^t}\mu_{I^{gt}} + a_1)(2\sigma_{O^tI^{gt}} + a_2)}{(\mu_{O^t}^2 + \mu_{I^{gt}}^2 + a_1)(\sigma_{O^t}^2 + \sigma_{I^{gt}}^2 + a_2)} \tag{7}$$

where μ_{O^t} and $\mu_{I^{gt}}$ are the average value of O^t and I^{gt}, respectively. $\sigma_{O^tI^{gt}}$ is the covariance of O^t and I^{gt}. $\sigma_{O^t}^2$ and $\sigma_{I^{gt}}^2$ are the variance of O^t and I^{gt}, respectively. a_1 and a_2 are two constants to maintain stability.

We adopt the following loss function to train FSAD-Net:

$$L = \sum_{t=1}^{T} \delta_t (L_1(O^t, I^{gt}) + \lambda L_{ssim}(O^t, I^{gt})) \tag{8}$$

where λ is a hyper-parameter to balance L_1 and L_{ssim}, the value of it depends on experience. δ_t is tradeoff parameter for indicating the contribution of the output at iteration t. The experiment results section show the experiments concerning the loss function in recursive supervision or only supervised in final output. It is worth mentioning that when $\delta_{t=1...(T-1)} = 0$, $\delta_{t=T} = 1$, the loss function shown in Eq. (8) can be rewritten as $L = L_1(O^T, I^{gt}) + \lambda L_{ssim}(O^T, I^{gt})$, which means the dehazing network can only be supervised by final output. In Section IV-F, we will discuss the differences between applying supervision on the final output and outputs of all iterations.

IV. Experimental Results

In this section, we first introduce datasets used for training and testing, implementation details and evaluation baselines. Next, we conduct extensive quantitative and qualitative visual experiments to demonstrate that the proposed FSAD-Net performs favorably against the state-of-the-arts (DCP [8],FD-GAN [10], cGAN [17], LDP [25], HL [34], GFN [39], PMS-Net [41], SemiDN [59], RefineDNet [60], LPQC [61], DDIP [62], PMHLD [63]) on synthetic datasets. In addition, we conduct assessments on real haze images by comparing the eight methods. Furthermore, we perform ablation studies to verify the main effectiveness of our methods. Finally, we show the user study results and discuss limitations of our method. The source code will be made publicly available.

Algorithm 1 Training Details at Each Iteration

Input: Synthetic training dataset *R*, hyper-parameter λ
Output: Updated FSAD-Net

for $i < training\ iterations$ **do**
  Randomly choose haze/clear image pairs I/I^{gt} from *R*;
  for $t < recurrent\ iterations\ T$ **do**
    obtain shallow feature information F_{sf}^t by SFEB;
    obtain middle-level features F_{in}^t by ARB-1;
    obtain F_{fb}^t by Eq. (3) from F_{sf}^t and F_{in}^t;
    obtain global feature F_{gf}^t by ARB-2;
    obtain clear image O^t at the t-th time iteration by Eq. (5) from F_{gf}^t and I;
    $t \leftarrow t + 1$;
  end for
  obtain L_1 by Eq. (6) from $\{O^T, I^{gt}\}$;
  obtain L_{ssim} by Eq. (7) from $\{O^T, I^{gt}\}$;
  back propagate FSAD-Net by $L = L_1 + \lambda L_{ssim}$;
  $i \leftarrow i + 1$;
end for

A. Datasets

Li *et al.* [23] proposed an image dehazing benchmark consisting of both synthetic and real-world hazy images, called REalistic Single Image DEhazing (RESIDE). In this paper, we use RESIDE [23] to train our FSAD-Net due to its large scale and diverse data sources and image contents. The Indoor Training Set (ITS) and The Outdoor Training Set (OTS) of RESIDE [23] corresponding indoor and outdoor scenarios, respectively. ITS contains a total of 13990 hazy indoor images, the depth maps $d(\boldsymbol{x})$ are obtained from the NYU Depth V2 [64] and Middlebury Stereo datasets [65], the global atmosphere light are ranged from 0.8 to 1.0, and the scatter parameters randomly selected from 0.04 to 0.2. OTS used 8477 clean images and the corresponding labeled depth maps to generate a total of 296695 hazy outdoor images. The atmospheric light is randomly sampled in [0.8, 1.0], and the scattering coefficient is randomly set within [0.04, 0.2]. For testing, we use Synthetic Objective Testing Set (SOTS) of RESIDE [23] and HAZERD [24] for quantitative evaluations, the SOTS contains 500 indoor and 500 outdoor synthetic hazy images (non-overlapping with RESIDE [23] Training sets). HazeRD [24] simulates different conditions from light to thick haze. We use the Real-world Task-driven Testing Set (RTTS) of RESIDE-β [23] for visual comparisons, the RTTS consisted of 4322 real-world hazy images crawled from the web, covering mostly traffic and driving scenarios.

B. Implementation Details

Our FSAD-Net is end-to-end trainable without the need of pre-training for sub-modules, and is implemented using Pytorch, trained on a PC equipped with only one NVIDIA GTX 1080Ti GPU. In our experiments, the patch size is 100×100. The ADAM optimizer is adopted to train the models with a batch size of 18, where $\beta 1$ and $\beta 2$ take the default values of 0.9 and 0.999, respectively. The initial learning rate

Fig. 7: Qualitative comparison of different methods on synthetic indoor image. (a) Haze images with different haze density, haze-free images restored by (b) DCP [8], (c) HL [34], (d) SemiDN [59], (e) LDP [25], (f) GFN [39], (g) RefineDNet [60], (h) FD-GAN [10], and (i) our FSAD-Net. (j) Ground truths. Our FSAD-Net is robust to different haze density.

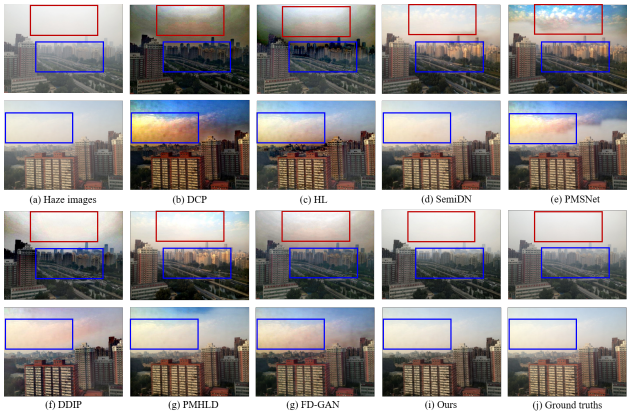


Fig. 8: Qualitative comparison of different methods on synthetic outdoor image. (a) Haze image, haze-free images restored by (b) DCP [8], (c) HL [34], (d) SemiDN [59], (e) PMS-Net [41], (f) DDIP [62], (g) PMHLD [63], (h) FD-GAN [10], and (i) our FSAD-Net. (j) Ground truths.

is set to 0.001 and decayed by multiplying 0.2 when reaching 30, 50 and 80 epochs. For ITS, we train the network for 100 epochs in total, for OTS, the network is only trained for 20 epochs. We present the training details when updating the FSAD-Net in Algorithm 1.

C. Evaluation Baselines

We will compare FSAD-Net with twelve competing methods, including two prior-based methods and ten deep CNN-based methods. Among the prior-based methods, the first one is the DCP [8], this is a commonly used baseline approach in most dehazing papers. DCP affirmed that values of pixels in a clear image patch close to zero in at least one color channel. The second is haze lines (HL) [34], which is a non-local approach approximating colors of a haze-free image by several hundreds of distinct colors. The CNN-based methods include learning a patch quality comparator (LPQC) [61], GFN [39], double deep image priors (DDIP) [62], cGAN [17], FD-GAN [10], LDP [25], SemiDN [59], RefineDNet [60], PMS-Net [41] and PMHLD [63]. These CNN-based methods achieved haze removal from different points. LPQC [61] designed a patch quality comparator for image dehazing, the comparator compared different output patches with the input hazy version. GFN [39] proposed a dehazing network based on fusion strategy. DDIP [62] is an unsupervised learning framework based on deep image prior [66]. cGAN [17] and FD-GAN [10] aimed to restore a haze-free image from a input haze image by GAN architectures. LDP [25] proposed an iteration algorithm to learn haze-relevant priors for haze removal. SemiDN [59] and RefineDNet [60] want to utilize merits of image prior and learning-based approaches and performed dehazing in a semi-supervised and weakly supervised manner. PMS-Net [41] proposed a patch size selection network for image dehazing, PMHLD [63] improved the PMS-Net [41] and presented a patch map-based hybrid dehazing network. For the fairness of comparisons, all results for comparisons were obtained by using the original codes from the home pages of the authors.

TABLE II: Average PSNR and SSIM comparison on the SOTS. Bold font is used for indicating our results.

Methods	PSNR		SSIM	
	Indoor	*Outdoor*	*Indoor*	*Outdoor*
DCP [8]	18.71	18.34	0.7918	0.7918
HL [34]	18.46	18.25	0.7237	0.8086
LPQC [61]	21.50	23.14	0.8714	0.8647
GFN [39]	23.00	23.50	0.9033	0.8251
DDIP [62]	16.47	16.27	0.5891	0.5085
cGAN [17]	26.63	25.02	0.9422	0.8671
FD-GAN [10]	21.99	23.54	0.8402	0.8549
LDP [25]	21.65	24.50	0.8472	0.8584
SemiDN [59]	23.52	26.10	0.9189	0.9039
RefineDNet [60]	21.95	23.71	0.8655	0.8775
PMS-Net [41]	21.93	22.40	0.8836	0.8699
PMHLD [63]	24.81	22.78	0.8573	0.7201
Our FSAD-Net	**23.4105**	**26.3876**	**0.9336**	**0.9292**
Our FSAD-Net*	**24.5424**	**26.0134**	**0.9459**	**0.9325**

D. Evaluation on Synthetic Datasets

In this section, we will compare FSAD-Net with twelve state-of-the-arts image dehazing algorithms. Since PSNR and SSIM are widely used in image objective evaluation, we first adopt PSNR and SSIM for quantitative assessment of the dehazed outputs. We also use three additional image quality metrics, the first one is CIEDE2000, which evaluates color of dehazed image, value of which is in the range [0,100] with smaller values indicating better color preservation. The second metric is UQI [21], which is designed to measure structual distortion of restored image, including loss of correlation, luminance distortion and contrast distortion. The value of this metric ranged from -1 to 1, and higher UQI [21] means better

TABLE III: Average CIEDE2000, UQI [21] and LPIPS [22] comparison on the SOTS. Bold font is for indicating our results.

Methods	CIEDE2000			UQI [21]			LPIPS [22]		
	Indoor	*Outdoor*	*Average*	*Indoor*	*Outdoor*	*Average*	*Indoor*	*Outdoor*	*Average*
DCP [8]	12.1918	15.9610	14.0764	0.7062	0.5922	0.6492	0.1377	0.1960	0.1669
HL [34]	12.3533	10.1487	11.251	0.6952	0.6901	0.69265	0.1108	0.1534	0.1321
LPQC [61]	6.9905	8.6935	7.842	0.8176	0.8190	0.8183	0.0772	0.0748	0.076
GFN [39]	6.8496	6.8044	6.827	0.7825	0.6817	0.7321	0.0678	0.1393	0.1035
DDIP [62]	10.8871	8.5829	9.735	0.6758	0.6739	0.6749	0.3762	0.2300	0.3031
cGAN [17]	4.2191	7.7086	5.9639	0.7792	0.5272	0.6532	0.0765	0.1686	0.1226
FD-GAN [10]	9.3007	7.7413	8.5210	0.7056	0.7307	0.71815	0.1303	0.1370	0.1337
LDP [25]	11.8186	10.400	11.1093	0.2113	0.1646	0.18795	0.0997	0.1251	0.1124
SemiDN [59]	5.8816	4.4511	5.16635	0.8100	0.7513	0.78065	0.0605	0.1090	0.08475
RefineDNet [60]	7.7726	7.4029	7.58775	0.7437	0.7176	0.73065	0.1239	0.1191	0.1215
PMS-Net [41]	5.8884	6.7702	6.3293	0.7728	0.7838	0.7783	0.1003	0.1540	0.12715
PMHLD [63]	5.0979	10.9725	8.0352	0.7471	0.5958	0.67145	0.0713	0.2268	0.14905
Our FSAD-Net	**5.7991**	**5.1609**	**5.48**	**0.8162**	**0.8637**	**0.83995**	**0.0660**	**0.0556**	**0.0608**
Our FSAD-Net*	**5.1325**	**4.7190**	**4.92575**	**0.8385**	**0.8489**	**0.8437**	**0.0540**	**0.0724**	**0.0632**

quality. The last one is LPIPS [22], evaluating the distance between image patches, lower LPIPS [22] means more similar to ground truths. Fig. 7 and Fig. 8 show the synthetic indoor and outdoor dehazing results obtained by different methods, respectively. Fig. 7(a) and Fig. 8(a) are synthetic hazy images to be dehazed, Fig. 7(j) and Fig. 8(j) are corresponding ground truths. It is worth point that, to prove the proposed the suitability of our method for each distinct haze density, we choose the thin haze image and thick haze image synthesized by the same haze-free indoor image for comparison. As shown in Fig. 7, the first row of Fig. 7(a) is dense haze image, the second row of Fig. 7(a) is thin haze image. It can be seen from Fig. 7(b) and Fig. 7(c) that DCP [8] and HL [34] tend to produce over-dehazed results, and the phenomenon become more severe when haze is thin. Fig. 7(d) and Fig. 7(g) show the results of SemiDN [59] and RefineDNet [60], for these methods, the dehazed performances of images with different haze densities are very different. When the haze is dense, the dehazed results still remained some haze, while when the haze is thin, the dehazed results suffered over-enhancement and possessed dimmer colors. This also illustrated that these methods are not robust to haze density. For GFN [39], with the change of haze density, the difference between the dehazing results is small. GFN [39] is a CNN-based method, which was based on three derived images and reconstructed a haze-free image by fusing them using a confidence map, while the derived images introduced estimation errors and leading over-enhancement of dehazed images, such as the blue box denotes in Fig. 7(f). Dehazing results of LDP [25] preserve structural details of objects in image with thin haze, however, as shown in two images of Fig. 7 (e), haze still exists at regions denoted by the red boxes. It can be observed that color shift existed in the results by FD-GAN [10].

We also choose the synthetic outdoor images with dense haze for comparison. Fig. 8(b) depicts the results of DCP [8], most details of the scenes and objects are well restored. However, the results significantly suffer from over-enhancement (for instance, the sky region of the outdoor image exist color distortion and halo). This is because DCP [8] has an inherent problem of overestimating the transmission. The same phenomenon also appeared in results of HL [34] as shown in Fig. 8(c). Although PMS-Net can generate a appropriate patch map for DCP and relieved the color deviation, the results shown in Fig. 8(e) also exist color distortions in sky regions. The results of DDIP [62] and FD-GAN [10] suffered from different degrees of visual artifact, such as the halo in sky region as shown in fist images of Fig. 8(f) and Fig. 8(g). Methods from SemiDN [59], PMHLD [63] and our method have the most competitive visual results. However, by looking closer, it can be observed that SemiDN [59] and PMHLD [63] produced unrealistic color shifts such as the color of sky compared with the ground truth in Fig. 8(d) and Fig. 8(g). On the contrary, on the synthetic indoor and outdoor datasets, our method is able to generate more natural and realistic colors while better removing haze. Our dehazed results of indoor images illustrated the proposed FSAD-Net can restore more similar to the ground truth compared with the results obtained by other methods. Our results do not show any negative effects and can maintain the original tones when haze density change. This can be seen from Fig. 7(i) and Fig. 8(i). The reason why our FSAD-Net can achieve more natural results and rich details is that our feedback mechanism can refine low level representations with high-level information.

We performed quantitative evaluations on the SOTS in terms of average PSNR and SSIM values as illustrated in Table II. It can be seen that CNN-based methods [10], [17], [25], [39], [41], [59], [60], [61], [62] and [63] generally outperformed the prior-based methods [8] and [34]. CNN-based dehazing methods provided powerful capability to represent and establish a more complex dehazing mapping function, while having

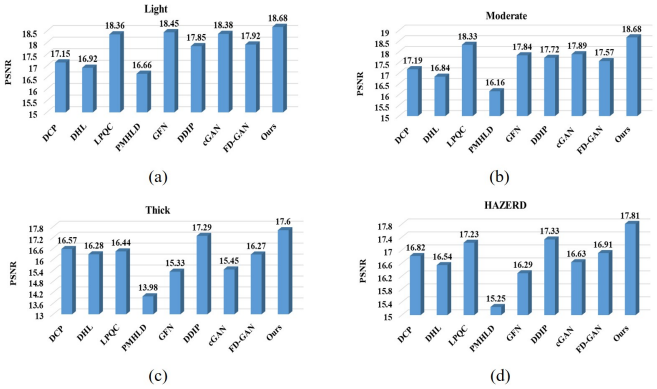


Fig. 9: Performance of different dehazing methods on HazeRD [24]. (a) Qualitative comparison on 'Light' condition of HAZERD [24]. (b) Qualitative comparison on 'Moderate' condition of HAZERD [24]. (c) Qualitative comparison on 'Thick' condition of HAZERD [24]. (d) Qualitative comparison on HAZERD [24].

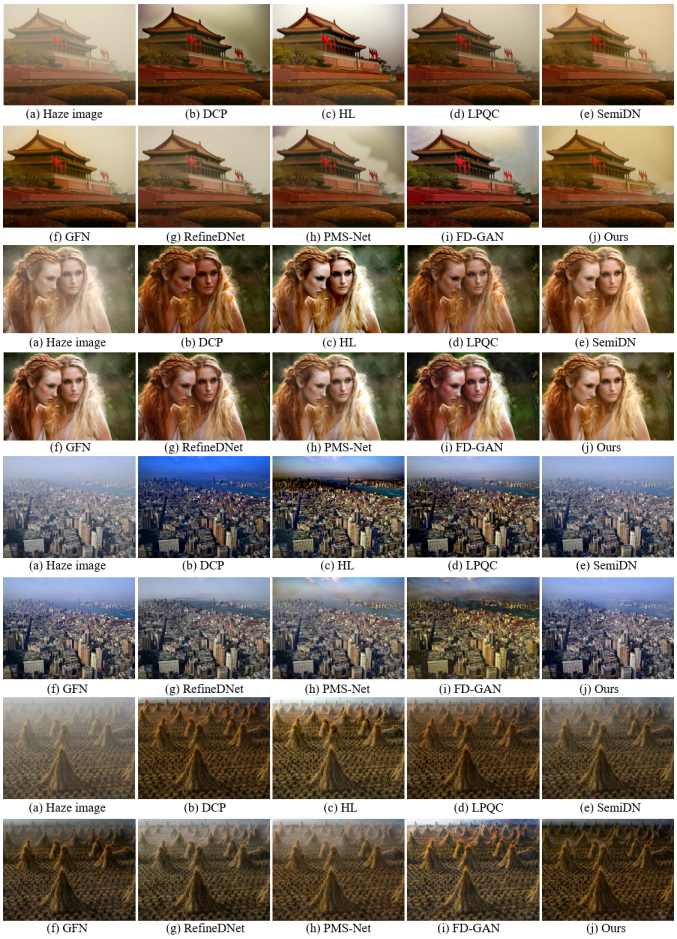


Fig. 10: Qualitative comparison of different methods on real-world images. (a) Haze image, haze-free images restored by (b) DCP [8], (c) HL [34], (d) LPQC [61], (e) SemiDN [59], (f) GFN [39], (g) RefineDNet [60], (h) PMS-Net [41], (i) FD-GAN [10], and (j) our FSAD-Net. Our FSAD-Net effectively removes haze from a single hazy image while preserving fine details and realistic colors.

stronger restoration performance than prior-based methods. As shown in Table II, our proposed FSAD-Net outperforms nine different state-of-the-art methods in terms of PSNR and SSIM. In addition, we use average CIEDE2000, UQI [21] and LPIPS [22] as full-reference image quality assessment for quantitative performance evaluation to further prove the robustness of the proposed FSAD-Net. As indicated in Table III, compared to other state-of-the-art methods, FSAD-Net almost achieved best performance on the CIEDE2000, UQI [21], and LPIPS [22] metric for all datasets. The quantitative results also illustrated complicated network architectures are difficult to train and suffered from the overfitting problem. For example, the CIEDE2000s of FD-GAN [10] in indoor images and outdoor images existed a big gap. Apparently, the proposed method realizes the best color-balanced and recovers more effective details when dehazing. We then evaluate our method on the HazeRD [24]. To illustrate our method is robust to different haze density, we choose 'Light', 'Moderate' and 'Thick' weather conditions in HAZERD [24]. From Fig. 9(a) to Fig. 9(c) show the comparisons on 'Light', 'Moderate' and 'Thick' in term of PSNR. Fig. 9(d) show the evaluation on HAZERD [24]. The results show our algorithm performs favorably against the state-of-the-art dehazing methods.

After carefully dissecting the proposed architecture of FSAD-Net and comparing it with state-of-the-art architectures, we can make the following conclusions. 1) recurrent structure plays a key role in our method, it can achieve haze removal step-by-step, meanwhile, the spatial attention mechanism is very effective for compensating the information loss during convolutional operations. 2) modeling the dehazing task in an end-to-end manner is beneficial, and the outstanding dehazing results does not necessarily depend on the atmospheric scattering model.

E. Evaluation on Real-World Datasets

We further compare the proposed FSAD-Net against twelve state-of-the-arts on the real-world images. Here we shall only make qualitative comparisons since the haze-free counterparts of the real-world hazy images are not available. It is discovered that all the methods have made a great progress on dehazing task as shown in Fig. 10, while some methods are susceptible to bright regions, whose color are mistakenly considered as haze. Such as the sky regions of the images restored by DCP [8], HL [34] and FD-GAN [10] are over-enhanced, and prone to produce color distortions and blurred vision as shown in fourth row of Fig. 10. The same phenomenon appeared in women's faces. For SemiDN [59] and RefineDNet [60], there are still some remaining haze in dehazed images. LPQC [61] outperformed these methods but has limited ability to deal with distance region and exhibits color artifacts as illustrated in Fig. 10(d). Is is obvious that PMS-Net [41] produced halo in first image of Fig. 10(h). The dehazed results by GFN [39] are closer to the proposed method, however it still retained residual haze in dense haze regions, e.g., the last row of Fig. 10(f). In comparison to eight state-of-the-art methods, though the proposed FSAD-Net is trained on synthesis haze images,

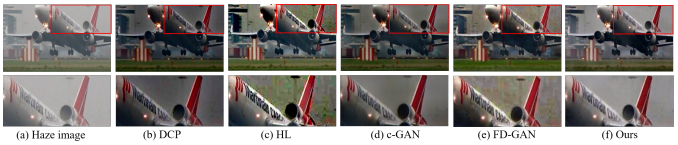


Fig. 11: Qualitative comparison of different methods on real-world haze image with dense haze. (a) Haze image, haze-free images restored by (b) DCP [8], (c) HL [34],(d) cGAN [17], (e) FD-GAN [10], and (f) our FSAD-Net. The images in the second row are the close-up of the red boxes in the first row. Our FSAD-Net generate a sharper and more pleasing result.

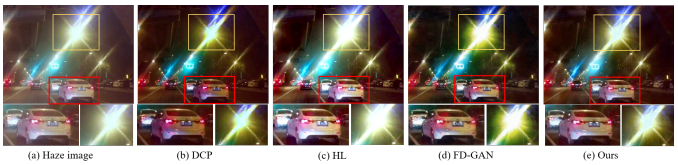


Fig. 12: Qualitative comparison of different methods on night-time haze image. (a) Haze image, haze-free images restored by (b) DCP [8], (c) HL [34],(d) FD-GAN [10], and (e) our FSAD-Net. The images in the second row are the close-up of the red and yellow boxes in the first row. Our FSAD-Net remove the haze for texture regions, and produce pleasing result.

it more effective in haze removal on the real-world dataset and has the remarkable performance on generalization ability.

To compare different methods generalization ability, we random selected 50 daytime haze images and 50 nighttime images from RTTS for comparisons. Fig. 11 and Fig. 12 show the results of day image and night image respectively. Results of DCP [8] and cGAN [17] show that the sky are become dimmer than input haze image. HL [34] and FD-GAN [10] have limited ability to deal with distance region and exhibits color artifacts as illustrated in Fig. 11(c) and Fig. 11(e). Our FSAD-Net, in contrast, produces restored images with minimal observable artifacts and color shifts. In addition, we get more contrast-enhanced images than the other state-of-the-art methods. From the close-up views of nighttime dehazed images shown in second row of Fig. 12, DCP [8] restored darker subjects than other methods. The light sources in the outputs restored by HL [34] and FD-GAN [10] are magnified as can be seen from the light of street lamp of Fig. 12(c) and Fig. 12(d). The proposed method completely removed the haze without color distortion. The comparisons in nighttime images show that our FSAD-Net's potential for haze removal in non-uniform atmosphere light, e.g., artificial ambient light in nighttime haze environment.

F. Ablation Analysis

To further demonstrate the effectiveness of the proposed FSAD-Net, we conduct an ablation study by considering the combination of five keys: 1) residual block with spatial attention and skip fusion, 2) feedback block, 3) multi-stage iteration, 4) loss function, 5) input and output. To make fair comparison, we keep the same training settings for all ablation studies. We first take the study of the influence of the recurrent

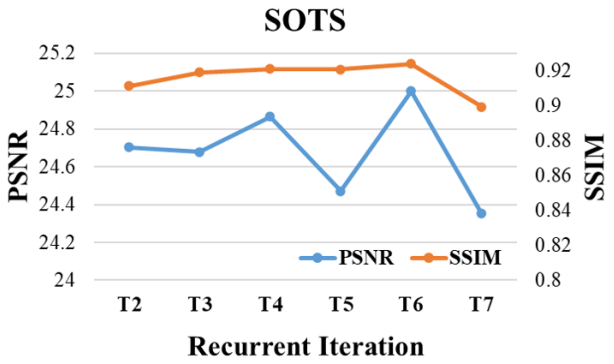


Fig. 13: Comparison of FSAD-Net models with different T iterations. The test performance increases when the number of iterations increases, but the performance of dehazing seems to diminish after $T = 6$.

iteration number T. Fig. 13 provides the information about the PSNR and SSIM values of six FSAD-Net models with stages $T = 2, 3, 4, 5, 6, 7$ on SOTS. It can be observed from the comparison results that the FSAD-Net with more iterations (from 2 iterations to 6 iterations) usually have higher average PSNR and SSIM values. This further prove that with the help of feedback connections, the dehazing quality can be improved stage-by-stage. But the phenomenon will be changed when the iteration is too large, this is because networks with large iteration have difficulties to train. As the Fig. 13 shows, the PSNR and SSIM values become smaller when the proposed network with 7 iterations. Based on this study, we set the iteration number to 6 in the experiments in this paper.

The iteration T can change the maximum effective depth of the proposed network. To dig deeper into the influence of T, we train the model with different iterations and visualize the output of every iteration, the results are shown in illustrated in Fig. 14. From Fig. 14, we have two observations. The first observation is that the test performance increases when the number of unfolding steps increases, but the performance of dehazing seems to diminish after $T = 6$. Compared with the feedforward network ($T = 1$), outputs acquired from the feedback network contain more details, showing a stronger effect of suppressing the smooth area of the input image. This reflects the feedback network has more powerful dehazing ability than the feedforward one. Secondly, feedback network can obtain well-dehazed outputs at the initial iteration, and it is allowed to devote most of its efforts to take a self-correcting process. As shown in Fig. 14, the high-frequency components (i.e. edges and contours) are gradually restored by iteration.

To demonstrate the effectiveness and superiority of attention mechanism and skip fusion, we train the FSAD-Net without attention mechanism and skip fusion, and calculate the PSNR and SSIM of dehazing results of it as shown in the forth column of Table IV. Compared with our results in the last column of Table IV, It is obviously that the introduction of attention mechanism and skip fusion does benefit the dehazing performance in terms of PSNR and SSIM. We also examine the baseline that FSAD-Net without multi-stage iterations (t

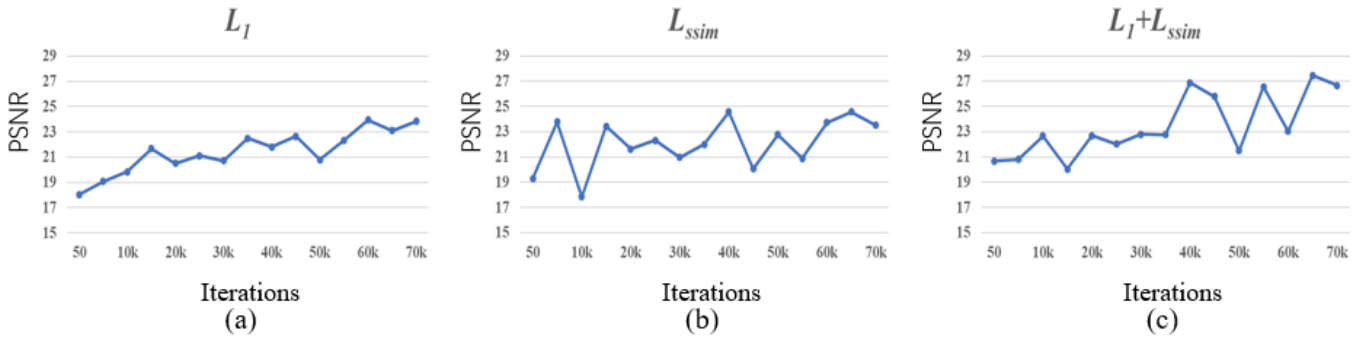


Fig. 15: Comparison of different Loss. (a) PSNR on training data using loss function L_1 , (b) PSNR on training data using loss function L_{ssim}, (c) PSNR on training data using loss function $L_{ssim} + L_1$. The $L_{ssim} + L_1$ leads to better results.

Fig. 14: (a) Haze images. (b)-(h) Dehazing results by FSAD-Net at iteration T = 1, 2, 3, 4, 5, 6, 7, respectively. (i) Ground truth. We also display the the PSNR and SSIM of each restored image, these values are shown below every dehazed image. The values of evaluation metrics generally gets better as the number of iterations increases. A FSAD-Net can remove most haze at the first iteration, and then remaining haze can be progressively removed by iterations, leading to promising dehazing quality at the sixth iteration. However, larger iteration T also makes the proposed network to face the problem of overfitting. It can be seen from the output of FSAD-Net at iteration T = 7, the model suffers from the remaining haze.

= 1). Although the quantitative metrics values are lower than FSAD-Net (t = 6), the results of FSAD-Net from the first iteration shown in fifth column of Table IV are still superiority, compared with the results of state-of-the-art methods displayed in Table II. We present the comparison of FSAD-Net with and without feedback block to show the effectiveness of FB, the results show that FB plays an important role in the network performance. Table IV also lists the PSNR and SSIM values of

TABLE IV: Ablation study results on SOTS indoor testset for different configurations.

Attention+Skip fusion	√	√		√	√
Feedback Block			√	√	√
Multi-Stage Iteration		√	√		√
PSNR	22.1694	21.5958	21.1391	21.4320	**23.3437**
SSIM	0.8858	0.8236	0.8456	0.8063	**0.9248**

TABLE V: Ablation study results on SOTS dataset for different loss functions and inputs.

Models	SSIM	PSNR	CIDE2000	UQI [21]	LPIPS [22]
FSAD-Net-L_1	0.8923	22.5145	7.1289	0.6857	0.1359
FSAD-Net-L_{ssim}	0.9201	**23.6745**	6.4059	0.8084	**0.0689**
FSAD-Net-In	0.9079	22.3455	7.5595	0.7891	0.0802
Our FSAD-Net	**0.9248**	23.3437	**5.9955**	**0.8144**	0.0705

FSAD-Net without feedback block and multi-stage iterations, it is reasonable to see that dehazing quality can improved by iterations. In addition, we can clearly see that even if we use the structure shown in Table IV, our network can be very competitive compared with eight state-of-the-art methods.

In addition, we analyze the effect of each supervised loss. Fig. 15 shows the PSNR on training data using different losses. Table V presents the average performance of different loss functions on the SOTS in terms of the SSIM, PSNR, CIEDE2000, and UQI [21]. As shown in Table V and Fig. 15, the comparisons show the method using the combination of L_1 and L_{ssim} performs well. Finally, we conduct several ablation studies about input and output to analyse the influence of the input and output to our FSAD-Net. We send the stage-wise estimation O^{t-1} without addition with haze image I to each sub-network in each iteration t, and denote the model as "FSAD-Net-In". The quantitative evaluation of FSAD-Net-In on SOTS are shown in forth row of Table V. The results show that

TABLE VI: Average PSNR and SSIM comparison on the SOTS. The results indicate that a single loss on the final stage is sufficient to train progressive networks.

Models	Indoor		Outdoor		HazeRD	
	PSNR	*SSIM*	*PSNR*	*SSIM*	*PSNR*	*SSIM*
FSAD-Net	23.4105	0.9336	**26.3876**	0.9292	**17.8144**	**0.5889**
FSAD-Net*	**24.5424**	**0.9459**	26.0134	**0.9325**	17.3417	0.5773

(a) Haze images (b) DCP (c) FD-GAN (d) Ours

Fig. 16: Failure cases. (a) Haze images. Dehazing results by (b) DCP [8], (c) FD-GAN [10] and (d) FSAD-Net.

adding O^{t-1} and I as the input to each sub-network is a better choice for our FSAD-Net. As we mentioned before, when $\delta_{t=1...(T-1)} = 0$, $\delta_{t=T} = 1$, the loss function shown in Eq. (8) can be rewritten as $L = L_1(O^T, I^{gt}) + \lambda L_{ssim}(O^T, I^{gt})$, which means the dehazing network can only be supervised by final output. To investigate the effect of output in each iteration, we show the comparison between our network supervised by final output (FSAD-Net) and our network supervised by outputs of each stage (FSAD-Net*) in Table VI. From Table VI, FSAD-Net performs moderately inferior to FSAD-Net* in SOTS dataset, while FSAD-Net performs better in HAZERD dataset compared with FSAD-Net*. We can conclude that the total loss applied only on the final stage is sufficient to train our dehazing network. Considering the completeness of our method, we use FSAD-Net* as an enhanced version of our method and show the quantitative results on SOTS in Table II and Table III.

G. User Study

To measure the visual quality of dehazed images obtained by different methods, we conducted a subjective user study to quantify the user-preference. We collect 25 synthetic images from RESIDE dataset and 25 real-world haze images from related works. Some corresponding images have been shown in Figs. 7, Fig. 8 and Fig. 10. We randomly order the results obtained by 13 methods and separately display them to 20 participants. Each participant grade results obtained by each method from 1 to 10 subjectively according to visibility and natural color of dehazed images (1 represents the worst visibility and severe color distortion and 10 represents the best image quality). Table VII lists the result of our user study, for synthetic and real-world image, we show the subjective scores of visibility and color balance, the final subjective score denoted as "Quality" are obtained by averaging scores of visibility and color balance. From the subjective scores

TABLE VII: User study results.

Methods	Synthetic images			Real-world images		
	Visibility	*Color*	*Quality*	*Visibility*	*Color*	*Quality*
DCP [8]	7.2	6.15	6.68	8.02	7.07	7.55
HL [34]	7.37	6.97	7.17	7.92	7.46	7.69
LPQC [61]	7.86	7.73	7.8	8.06	7.89	7.98
GFN [39]	8.14	7.95	8.05	8.06	8.08	8.07
DDIP [62]	7.86	7.59	7.73	7.97	7.37	7.67
cGAN [17]	7.91	8.14	8.03	8.15	8.12	8.14
FD-GAN [10]	7.87	7.76	7.82	7.81	7.4	7.61
LDP [25]	8.34	8.29	8.32	7.83	8.22	8.03
SemiDN [59]	8.12	8.04	8.08	7.75	8.09	7.92
RefineDNet [60]	8.43	8.36	8.4	7.97	7.83	7.9
PMS-Net [41]	7.87	7.27	7.57	7.7	8.09	7.90
PMHLD [63]	8.19	8.29	8.24	8.2	8.49	8.35
Our FSAD-Net	**8.9**	**8.93**	**8.92**	**8.45**	**8.56**	**8.51**

illustrated in Table VII, our FSAD-Net yields best score in two aspects, which indicates that our method can produce more clear haze-free images and avoid color distortion comparing with other methods.

H. Limitation

In Fig. 16 we show some failure cases of our FSAD-Net. Our method can effectively handle most images with thick haze (as shown in Fig. 11). However, As Fig. 16 illustrated, our FSAD-Net performs poorly on sky regions with heavy haze. The result in first row of Fig. 16(d) show that our method cannot uniformly remove the haze in the sky area. The second dehazed image of Fig. 16(d) still remained a large amount of haze. Heavy haze is also an important factor interfered performances of other CNN-based methods [9] [17] [39]. The first reason that causes our method to fail in sky region with heavy haze is the dependence on synthetic data. Our FSAD-Net is trained on the RESIDE [23] which is created based on atmospheric scattering model [7], while the model cannot hold in sky areas with heavy haze. The second reason is the structure and texture features of the sky regions with heavy haze are rare, it is difficult for the algorithm to distinguish the haze and sky. We aim to handle these problems in the future.

V. Conclusion and Future Work

In this paper, we propose a novel end-to-end network for image dehazing called feedback spatial attention dehazing network (FSAD-Net). We introduce the recurrent structure to deep image dehazing architecture. We take the stage-wise result to each iteration to improve the dehazing performance. The feedback block included in FSAD-Net not only extract the feedback feature, but effectively handle the information reuse, benefiting the information flow across time. Moreover, the residual block with spatial attention mechanism and skip fusion which can help network pay more attentions on low frequency regions bypassing the unimportant information. Extensive experiments demonstrate that FSAD-Net achieves remarkably high efficiency comparing with twelve state-of-the-art methods. The comparisons on real-world images further

validate the proposed model has a powerful advantage in the restoration of image content and faithful color. Although our FSAD-Net based on recurrent architecture has achieved good performance compared with twelve state-of-the-art methods, our method has its shortcoming in dealing with bright region with heavy haze. Besides, the color deviations still exist in our results compared with corresponding ground truths. In our future works, we aim to solve this problem by improving network structure, loss functions, and input and output.

REFERENCES

[1] R. Liu, X. Fan, M. Hou, Z. Jiang, Z. Luo, and L. Zhang, "Learning aggregated transmission propagation networks for haze removal and beyond," *IEEE TNNLS*, vol. 30, no. 10, pp. 2973–2986, 2019.

[2] J. Zhang and D. Tao, "FAMED-Net: A fast and accurate multi-scale end-to-end dehazing network," *IEEE TIP*, vol. 29, pp. 72–84, 2020.

[3] S. E. Kim, T. H. Park, and I. K. Eom, "Fast single image dehazing using saturation based transmission map estimation," *IEEE TIP*, vol. 29, pp. 1985–1998, 2020.

[4] K. Xu, L. Wen, G. Li, L. Bo, and Q. Huang, "Spatiotemporal cnn for video object segmentation," in *IEEE CVPR*, 2019, pp. 1379–1388.

[5] K.-H. Shih, C.-T. Chiu, J.-A. Lin, and Y.-Y. Bu, "Real-time object detection with reduced region proposal network via multi-feature concatenation," *IEEE TNNLS*, vol. 31, no. 6, pp. 2164–2173, 2020.

[6] X. Lei, L. He, Y. Tan, K. X. Wang, X. Wang, Y. Du, S. Fan, and Z. Yu, "Direct object recognition without line-of-sight using optical coherence," in *IEEE CVPR*, 2019, pp. 11 729–11 738.

[7] S. Narasimhan and S. K. Nayar, "Vision and the atmosphere," *IJCV*, vol. 48, no. 3, pp. 233–254, 2002.

[8] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," *IEEE TPAMI*, vol. 33, no. 12, pp. 2341–2353, 2011.

[9] W. Ren, S. Liu, H. Zhang, J. Pan, X. Cao, and M.-H. Yang, "Single image dehazing via multi-scale convolutional neural networks," in *ECCV*, 2016, pp. 154–169.

[10] Y. Dong, Y. Liu, H. Zhang, S. Chen, and Y. Qiao, "FD-GAN: Generative adversarial networks with fusion-discriminator for single image dehazing," in *AAAI*, vol. 34, no. 07, 2020, pp. 10 729–10 736.

[11] B. Cai, X. Xu, K. Jia, C. Qing, and D. Tao, "DehazeNet: An end-to-end system for single image haze removal," *IEEE TIP*, vol. 25, no. 11, pp. 5187–5198, 2016.

[12] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *ECCV*, 2014, pp. 818–833.

[13] H. Zhang and V. M. Patel, "Densely connected pyramid dehazing network," in *IEEE CVPR*, 2018, pp. 3194–3203.

[14] Y. Pang, J. Xie, and X. Li, "Visual haze removal by a unified generative adversarial network," *IEEE TCSVT*, vol. 29, no. 11, pp. 3211–3221, 2019.

[15] X. Qin, Z. Wang, Y. Bai, X. Xie, and H. Jia, "FFA-Net: Feature fusion attention network for single image dehazing," in *AAAI*, vol. 34, no. 07, 2020, pp. 11 908–11 915.

[16] J. Hu, L. Shen, S. Albanie, G. Sun, and E. Wu, "Squeeze-and-excitation networks," *IEEE TPAMI*, vol. 42, no. 8, pp. 2011–2023, 2020.

[17] R. Li, J. Pan, Z. Li, and J. Tang, "Single image dehazing via conditional generative adversarial network," in *IEEE CVPR*, 2018, pp. 8202–8211.

[18] K. Chen, L. Yao, D. Zhang, X. Wang, X. Chang, and F. Nie, "A semisupervised recurrent convolutional attention model for human activity recognition," *IEEE TNNLS*, vol. 31, no. 5, pp. 1747–1756, 2020.

[19] D. Ren, W. Zuo, Q. Hu, P. Zhu, and D. Meng, "Progressive image deraining networks: A better and simpler baseline," in *IEEE CVPR*, 2019, pp. 3932–3941.

[20] Z. Li, J. Yang, Z. Liu, X. Yang, G. Jeon, and W. Wu, "Feedback network for image super-resolution," in *IEEE CVPR*, 2019, pp. 3862–3871.

[21] Z. Wang and A. Bovik, "A universal image quality index," *IEEE Signal Processing Letters*, vol. 9, no. 3, pp. 81–84, 2002.

[22] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric," in *IEEE CVPR*, 2018, pp. 586–595.

[23] B. Li, W. Ren, D. Fu, D. Tao, D. Feng, W. Zeng, and Z. Wang, "Benchmarking single-image dehazing and beyond," *IEEE TIP*, vol. 28, no. 1, pp. 492–505, 2019.

[24] Y. Zhang, L. Ding, and G. Sharma, "HazeRD: An outdoor scene dataset and benchmark for single image dehazing," in *IEEE ICIP*, 2017, pp. 3205–3209.

[25] Y. Liu, J. Pan, J. Ren, and Z. Su, "Learning deep priors for image dehazing," in *IEEE ICCV*, 2019, pp. 2492–2500.

[26] R. T. Tan, "Visibility in bad weather from a single image," in *IEEE CVPR*, 2008, pp. 1–8.

[27] R. Fattal, "Single image dehazing," *ACM TOG*, vol. 27, no. 3, pp. 72:1–72:9, 2008.

[28] P.-J. Liu, S.-J. Horng, J.-S. Lin, and T. Li, "Contrast in haze removal: Configurable contrast enhancement model based on dark channel prior," *IEEE TIP*, vol. 28, no. 5, pp. 2212–2227, 2019.

[29] Y.-T. Peng, K. Cao, and P. C. Cosman, "Generalization of the dark channel prior for single image restoration," *IEEE TIP*, vol. 27, no. 6, pp. 2856–2868, 2018.

[30] K. Tang, J. Yang, and J. Wang, "Investigating haze-relevant features in a learning framework for image dehazing," in *IEEE CVPR*, 2014, pp. 2995–3002.

[31] Q. Zhu, J. Mai, and L. Shao, "A fast single image haze removal algorithm using color attenuation prior," *IEEE TIP*, vol. 24, no. 11, pp. 3522–3533, 2015.

[32] L. K. Choi, J. You, and A. C. Bovik, "Referenceless prediction of perceptual fog density and perceptual image defogging," *IEEE TIP*, vol. 24, no. 11, pp. 3888–3901, 2015.

[33] C. Chen, M. Do, and J. Wang, "Robust image and video dehazing with visual artifact suppression via gradient residual minimization," in *ECCV*, 2016, pp. 576–591.

[34] D. Berman, T. Treibitz, and S. Avidan, "Single image dehazing using haze-lines," *IEEE TPAMI*, vol. 42, no. 3, pp. 720–734, 2020.

[35] J. Zhang, Y. Cao, S. Fang, Y. Kang, and C. W. Chen, "Fast haze removal for nighttime image using maximum reflectance prior," in *IEEE CVPR*, 2017, pp. 7016–7024.

[36] T. M. Bui and W. Kim, "Single image dehazing using color ellipsoid prior," *IEEE TIP*, vol. 27, no. 2, pp. 999–1009, 2018.

[37] Q. Wu, J. Zhang, W. Ren, W. Zuo, and X. Cao, "Accurate transmission estimation for removing haze and noise from a single image," *IEEE TIP*, vol. 29, pp. 2583–2597, 2020.

[38] B. Li, X. Peng, Z. Wang, J. Xu, and D. Feng, "AOD-Net: All-in-one dehazing network," in *IEEE ICCV*, 2017, pp. 4780–4788.

[39] W. Ren, L. Ma, J. Zhang, J. Pan, X. Cao, W. Liu, and M.-H. Yang, "Gated fusion network for single image dehazing," in *IEEE CVPR*, 2018, pp. 3253–3261.

[40] D. Yang and J. Sun, "Proximal dehaze-net: A prior learning-based deep network for single image dehazing," in *ECCV*, 2018, pp. 729–746.

[41] W.-T. Chen, J.-J. Ding, and S.-Y. Kuo, "PMS-Net: Robust haze removal based on patch map for single images," in *IEEE CVPR*, 2019, pp. 11 673–11 681.

[42] Y. Qu, Y. Chen, J. Huang, and Y. Xie, "Enhanced pix2pix dehazing network," in *IEEE CVPR*, 2019, pp. 8152–8160.

[43] W. Ren, J. Pan, H. Zhang, X. Cao, and M.-H. Yang, "Single image dehazing via multi-scale convolutional neural networks with holistic edges," *IJCV*, vol. 128, no. 1, pp. 240–259, 2020.

[44] Y. Li, Q. Miao, W. Ouyang, Z. Ma, H. Fang, C. Dong, and Y. Quan, "LAP-Net: Level-aware progressive network for image dehazing," in *IEEE ICCV*, 2019, pp. 3275–3284.

[45] M. Hong, Y. Xie, C. Li, and Y. Qu, "Distilling image dehazing with heterogeneous task imitation," in *IEEE CVPR*, 2020, pp. 3459–3468.

[46] F. Wang, M. Jiang, C. Qian, S. Yang, C. Li, H. Zhang, X. Wang, and X. Tang, "Residual attention network for image classification," in *IEEE CVPR*, 2017, pp. 6450–6458.

[47] X. Liu, Y. Ma, Z. Shi, and J. Chen, "GridDehazeNet: Attention-based multi-scale network for image dehazing," in *IEEE ICCV*, 2019, pp. 7313–7322.

[48] C. Chen, D. Gong, H. Wang, Z. Li, and K.-Y. K. Wong, "Learning spatial attention for face super-resolution," *IEEE TIP*, vol. 30, pp. 1219–1231, 2021.

[49] D. Cao, J. Chu, N. Zhu, and L. Nie, "Cross-modal recipe retrieval via parallel- and cross-attention networks learning," *Knowledge-Based Systems*, vol. 193, pp. 105 428–105 439, 2020.

[50] C. Yan, Y. Tu, X. Wang, Y. Zhang, X. Hao, Y. Zhang, and Q. Dai, "STAT: Spatial-temporal attention mechanism for video captioning," *IEEE TMM*, vol. 22, no. 1, pp. 229–241, 2020.

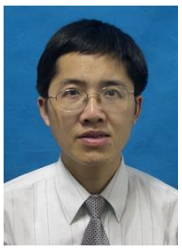
[51] X. Shu, L. Zhang, G.-J. Qi, W. Liu, and J. Tang, "Spatiotemporal co-attention recurrent neural networks for human-skeleton motion prediction," *IEEE TPAMI*, pp. 1–16, 2021.

[52] Y. Hu, J. Li, Y. Huang, and X. Gao, "Channel-wise and spatial feature modulation network for single image super-resolution," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 11, pp. 3911–3927, 2020.

[53] X. Li, J. Wu, Z. Lin, H. Liu, and H. Zha, "Recurrent squeeze-and-excitation context aggregation net for single image deraining," in *ECCV*, 2018, pp. 262–277.
[54] A. R. Zamir, T.-L. Wu, L. Sun, W. B. Shen, B. E. Shi, J. Malik, and S. Savarese, "Feedback networks," in *IEEE CVPR*, 2017, pp. 1808–1817.
[55] L. Yang, Q. Song, Y. Wu, and M. Hu, "Attention inspiring receptive-fields network for learning invariant representations," *IEEE TNNLSS*, vol. 30, no. 6, pp. 1744–1755, 2019.
[56] Z.-J. Zha, C. Wang, D. Liu, H. Xie, and Y. Zhang, "Robust deep co-saliency detection with group semantic and pyramid attention," *IEEE TNNLS*, vol. 31, no. 7, pp. 2398–2408, 2020.
[57] G. Zhu, L. Zhang, L. Yang, L. Mei, S. A. A. Shah, M. Bennamoun, and P. Shen, "Redundancy and attention in convolutional lstm for gesture recognition," *IEEE TNNLS*, vol. 31, no. 4, pp. 1323–1335, 2020.
[58] K. Cho, B. V. Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using RNN encoder–decoder for statistical machine translation," in *EMNLP*, 2014, pp. 1724–1734.
[59] L. Li, Y. Dong, W. Ren, J. Pan, C. Gao, N. Sang, and M.-H. Yang, "Semi-supervised image dehazing," *IEEE Transactions on Image Processing*, vol. 29, pp. 2766–2779, 2020.
[60] S. Zhao, L. Zhang, Y. Shen, and Y. Zhou, "RefineDNet: A weakly supervised refinement framework for single image dehazing," *IEEE Transactions on Image Processing*, vol. 30, pp. 3391–3404, 2021.
[61] S. Santra, R. Mondal, and B. Chanda, "Learning a patch quality comparator for single image dehazing," *IEEE TIP*, vol. 27, no. 9, pp. 4598–4607, 2018.
[62] Y. Gandelsman, A. Shocher, and M. Irani, ""Double-DIP": Unsupervised image decomposition via coupled deep-image-priors," in *IEEE CVPR*, 2019, pp. 11 018–11 027.
[63] W.-T. Chen, H.-Y. Fang, J.-J. Ding, and S.-Y. Kuo, "PMHLD: Patch map-based hybrid learning dehazeNet for single image haze removal," *IEEE TIP*, vol. 29, pp. 6773–6788, 2020.
[64] N. Silberman, D. Hoiem, P. Kohli, and R. Fergus, "Indoor segmentation and support inference from RGBD images," in *IEEE CVPR*, 2012, pp. 746–760.
[65] D. Scharstein and R. Szeliski, "High-accuracy stereo depth maps using structured light," in *IEEE CVPR*, 2003, pp. 195–202.
[66] D. Ulyanov, A. Vedaldi, and V. S. Lempitsky, "Deep image prior," in *IEEE CVPR*, 2018, pp. 9446–9454.

Yu Zhou received the M.Eng. degree in computer technology from the Jiangxi University of Finance and Economics, Nanchang, China, in 2019.

She is currently pursuing the Ph.D. degree in computer science with the Department of Computer Science and Engineering, East China University of Science and Technology, Shanghai, China. Her current research interests include image dehazing, computer vision, and machine learning.



Zhihua Chen received the Ph.D. degree in computer science from Shanghai Jiao Tong University, Shanghai, China, in 2006.

He is currently a Full Professor with the Department of Computer Science and Engineering, East China University of Science and Technology, Shanghai, China. His current research interests include image/video processing and computer vision.

Ping Li (Member, IEEE) received the Ph.D. degree in computer science and engineering from The Chinese University of Hong Kong, Shatin, Hong Kong, in 2013.

He is currently a Research Assistant Professor with The Hong Kong Polytechnic University, Kowloon, Hong Kong. He has published many top-tier scholarly research papers and has excellent research project reported worldwide by *ACM TechNews*. His current research interests include artistic rendering and synthesis, and creative media.

Haitao Song received the M.Sc. degree in computer science from Shanghai Jiao Tong University, Shanghai, China, in 2016.

He is currently a Deputy Dean with the Artificial Intelligence Institute, Shanghai Jiao Tong University, Shanghai, China. He is also the CEO with the Shanghai Artificial Intelligence Research Institute Co., Ltd, Shanghai, China. His current research interests include electronics and information, industrial Internet, and integrated energy system.

C. L. Philip Chen (Fellow, IEEE) received the Ph.D. degree in electrical engineering from Purdue University, West Lafayette, IN, USA, in 1988.

He is currently a Chair Professor and the Dean of the School of Computer Science and Engineering, South China University of Technology, Guangzhou, China. Being a Program Evaluator of the Accreditation Board of Engineering and Technology Education (ABET) in the U.S., for computer engineering, electrical engineering, and software engineering programs, he successfully architects the University of Macau's Engineering and Computer Science programs receiving accreditations from Washington/Seoul Accord through Hong Kong Institute of Engineers (HKIE), of which is considered as his utmost contribution in engineering/computer science education for Macau as the former Dean of the Faculty of Science and Technology. He is a Fellow of IEEE, AAAS, IAPR, CAA, and HKIE; a member of Academia Europaea (AE), European Academy of Sciences and Arts (EASA), and International Academy of Systems and Cybernetics Science (IASCYS). He received IEEE Norbert Wiener Award in 2018 for his contribution in systems and cybernetics, and machine learnings. He is also a highly cited researcher by Clarivate Analytics in 2018 and 2019.

His current research interests include systems, cybernetics, and computational intelligence. Dr. Chen was a recipient of the 2016 Outstanding Electrical and Computer Engineers Award from his alma mater, Purdue University (in 1988), after he graduated from the University of Michigan at Ann Arbor, Ann Arbor, MI, USA in 1985. He was the IEEE Systems, Man, and Cybernetics Society President from 2012 to 2013, the Editor-in-Chief of the IEEE Transactions on Cybernetics (2020-2021) and the IEEE Transactions on Systems, Man, and Cybernetics: Systems (2014-2019), and currently, an Associate Editor of the IEEE Transactions on Fuzzy Systems. He was the Chair of TC 9.1 Economic and Business Systems of International Federation of Automatic Control from 2015 to 2017, and currently is a Vice President of Chinese Association of Automation (CAA).

Bin Sheng (Member, IEEE) received the Ph.D. degree in computer science and engineering from The Chinese University of Hong Kong, Shatin, Hong Kong, in 2011.

He is currently a Full Professor with the Shanghai Jiao Tong University, Shanghai, China. He is an Associate Editor of the IEEE Transactions on Circuits and Systems for Video Technology. His current research interests include virtual reality and computer graphics.