The Visual Computer
https://doi.org/10.1007/s00371-019-01774-8

ORIGINAL ARTICLE

DRCDN: learning deep residual convolutional dehazing networks

Shengdong Zhang[1] · **Fazhi He**[1]

Abstract
Single image dehazing, which is the process of removing haze from a single input image, is an important task in computer vision. This task is extremely challenging because it is massively ill-posed. In this paper, we propose a novel end-to-end deep residual convolutional dehazing network (DRCDN) based on convolutional neural networks for single image dehazing, which consists of two subnetworks: one network is used for recovering a coarse clear image, and the other network is used to refine the result. The DRCDN firstly predicts the coarse clear image via a context aggregation subnetwork, which can capture global structure information. Subsequently, it adopts a novel hierarchical convolutional neural network to further refine the details of the clean image by integrating the local context information. The DRCDN is directly trained using complete images and the corresponding ground-truth haze-free images. Experimental results on synthetic datasets and natural hazy images demonstrate that the proposed method performs favorably against the state-of-the-art methods.

Keywords Residual learning · Dehazing · Image restoration · Global structure information · Deep learning

1 Introduction

The presence of haze dramatically reduces the contrast and color of the captured images, thereby resulting in the performance degradation of modern visual systems of detection and segmentation, which assume that the input is a clean image. These degradations are highly ill-posed because the lights reaching the camera are always reflected or absorbed by the particles in the air.

Based on this observation, the captured hazy image I can be described as a linear combination of the clean image and airlight contributions, which can be mathematically expressed as follows [1]:

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

where $I(x)$ represents the observed hazy image, $J(x)$ denotes the corresponding haze-free scene radiance to be recovered, A is the global atmospheric light, and $t(x)$ is the transmission map, which can be defined as follows:

$$t(x) = e^{-\beta d(x)}, \tag{2}$$

✉ Fazhi He
fzhe@whu.edu.cn

[1] School of Computer Science, Wuhan University, Wuhan, China

where β is the scattering coefficient of the atmosphere, and $d(x)$ is the distance between the scene point and the camera. Based on the model (1), we are able to recover the haze-free image as follows:

$$J(x) = \frac{1}{t(x)} I(x) - A\frac{1}{t(x)} + A. \tag{3}$$

As only a single hazy image $I(x)$ is available, recovering the corresponding haze-free image $J(x)$ is one of the typically ill-posed problems. To solve this problem efficiently, traditional methods have been proposed by using additional prior information [2] or multiple images [3–6] as that in many image processing researches [7–9]. For example, the depth information is recovered from multiple images taken in different weather conditions [4]. Polarization-based methods impose more constraints by considering image with different degrees of polarization [3,5]. In [6], a depth-based method was proposed, which extracts the depth information from user inputs or known 3D models. Although these methods have been successfully applied in dehazing, the requirement of additional information limits the application of these methods.

To solve this problem, single image dehazing methods based on various image priors have been proposed. Based on model (1), it is a natural way to solve the dehazing problem with two steps. The first step involves estimating the

Published online: 29 November 2019

transmission map and the atmospheric light by using various image priors. The second step involves recovering the final dehazed image by using Eq. (3). Based on the above two-step methodology, traditional single image dehazing methods [2,10–15] that estimate the transmission map by utilizing visual cues or statistical properties of hazy images have been proposed. In recent years, learning-based methods have been widely applied in many areas, such as object tracking [16], low-light image enhancement [17], image deblur [18], feature classification [19,20], image recolor [21], shadow removal [22], intelligent computing [23,24] and so on [25,26]. Researchers employ CNNs (convolutional neural networks) to estimate the transmission map [27,28] or to estimate the atmospheric light [29,30]. All the above-mentioned methods treat the dehazing problem as a separate optimization problem, in which the inaccurate estimation of the atmospheric light or the transmission map may thus affect the final dehazing quality.

To solve the aforementioned problems of the two-step methodology and unify the process, end-to-end dehazing methods were proposed [31–34]. Li et al. proposed an end-to-end dehazing network [31] (AOD-Net) based on fusion of the transmission map and atmospheric light into one new parameter, which can be expressed as follows:

$$J(x) = K(x)I(x) - K(x) + b. \tag{4}$$

where

$$K(x) = \frac{\frac{1}{t(x)}(I(x) - A) + (A - b)}{I(x) - 1}. \tag{5}$$

However, this method is limited by the small receptive field of the model. Furthermore, the dehazing performance of the AOD-Net is also highly limited by the accuracy of estimation of the new parameter $K(x)$. Although the DCPDN [34] achieves end-to-end dehazing by directly embedding the image degradation model (1) into the optimization framework via math operation modules, the generalization ability of the DCPDN is limited by model (1), and this aspect results in the natural images being far from the optimal results. Furthermore, the method cannot easily scaled up/down for different sizes of input images. GFN [32] is a deep end-to-end trainable neural network based on a novel fusion-based strategy, which is an extension of prior work [35]. Compared with the GFN and [35], the proposed technique has three major differences.

First, as a preprocessing step, the GFN and the model reported in [35], respectively, derive three and two inputs from an original hazy image. If these derivations do not contain sufficient information for removing the haze, a low-quality dehazing result could be generated. Furthermore, preprocessing itself easily results in information loss, which may incur the loss of certain details in the dehazed result. In contrast, our work avoids the information loss by extracting features from the hazy image directly.

Second, the model reported in [35] designs a complex blending based on the luminance, chromatic and saliency maps. The GFN [32] improves the fusion by blending the three derived input images. In contrast, our work avoids the complex blending by performing end-to-end training.

Third, the GFN [32,35] employ the image pyramid to remove artifacts and color distortions. In contrast, our work employs both the low-level and high-level features to remove the artifacts and color distortions. Furthermore, we utilize an end-to-end trainable network to avoid the preprocessing step and immediate parameters by learning the mapping from a hazy input to a haze-free output directly. Prior CNN-based methods [28,31,32,34] are usually trained on synthetic indoor haze datasets. The generalization ability of these methods is limited, and these methods cannot effectively deal with real hazy images.

The above discussion indicates that the most intrinsic problems for the further promotion of the single image dehazing methods pertain to the extraction of the effective features for removing the haze from a single image, efficient preservation of the object details, and avoidance of color distortions.

To solve these problems, we propose a novel end-to-end Deep Residual Convolutional Dehazing Networks, i.e., the DRCDN, based on convolutional neural networks (CNNs) similar to that reported in [31–34]. The DRCDN takes the complete hazy image as the input and outputs the corresponding haze-free image directly, hierarchically removing the haze from the global view to the local contexts, and from the coarse scale to fine scales. First, we use a context aggregation network to learn a coarse haze-free result. Second, we use shallow layers to learn more image details of the haze-free image. Finally, we use the coarse haze-free image and image details to reconstruct the final clean image.

Major contributions can be summarized as follows:

- We present a hierarchically removing haze model, which first reconstructs a coarse haze-free image from the global view, then learns the corresponding details from local contexts, finally recovers the dehazed result via these information.
- In order to recover more colorful and photo-realistic results, we develop a new loss function based on perceptual loss and $L1$ loss.
- We conduct extensive experiments on two synthetic haze datasets, one real haze dataset and real-world hazy images to compare our method with the state-of-the-art single image dehazing methods quantitatively and qualitatively, which demonstrate the effectiveness of the proposed method.

The rest of the paper is organized as follows. Section 2 reviews related work. Section 3 presents our method in detail. Section 4 covers our experimental results. Finally, we conclude the paper in Sect. 5.

2 Related work

2.1 Separate estimation methods

Based on model (3), there exist two key factors for single image dehazing: (1) accurate estimation of the transmission map, and (2) accurate estimation of the atmospheric light.

However, the estimation of the transmission map or atmospheric light is an ill-posed problem. Thus, many methods [2,10–15,36,37] use priors to estimate the transmission map.

Based on the observation that a clean image has higher contrast than that of a hazy image, Tan [10] proposed a method that maximizes the contrast of the dehazing result. Based on the system statistics of experiments performed on clean images, He et al. [2] noted that a local patch in clean image contains some pixels with very low intensity. Using a dark channel prior, the atmospheric light can be estimated; next, the transmission map can be estimated, and the final dehazing result can be acquired by using model (3). Meng et al. [12] proposed a more general prior boundary constraint. Berman et al. [15] proposed a method based on a haze-line (the pixels with the same color in a clean image will form a line in a hazy image due to the effect of haze).

All the priors are acquired based on certain statics and thus cannot be generalized for all cases. To improve the accuracy of estimating the transmission map, some methods drew lessons from learning-based methods [38,39]. Cai et al. [27] designed a trainable end-to-end network to estimate the transmission map. This work extracts four features (e.g., dark channel, hue disparity, and color attenuation) in the first layer. Next the sequential layers use multiscale mapping, local extremum, and nonlinear regression to estimate the final transmission map.

Ren et al. [28] proposed a multiscale network to estimate the transmission map. Their technique firstly estimates a coarse transmission map based on the entire image, and later refines it via the local context information. Furthermore, Tang proposed the use of four haze-relevant features (e.g., dark channel, hue disparity, and color attenuation) to learn a suitable combination to estimate the transmission map.

However, all these methods exhibit the same problem that they are separate approaches to address the haze removal problem, and the inaccurate estimation of the transmission map affects the estimation of the clear image.

In addition to the estimation of the transmission map, some other methods [29,30] focused on estimating the atmospheric light were proposed. However, estimating the atmospheric light encounters the same problems as the case of estimating the transmission map.

2.2 Joint learning-based methods

To overcome the problems associated with estimating the atmospheric light and transmission map separately, a natural approach is to jointly estimate the atmospheric light and transmission map. CNNs have achieved considerable success in image classification [40] and detection [41–43]. Inspired by the use of end-to-end deep learning methods in image restoration [44–47], end-to-end methods have been proposed for single image dehazing [31–34,48].

For example, by fusing the transmission map and atmospheric light into a new variable, Li et al. propose an end-to-end dehazing network, named AOD-Net [31]. Although this method can generate high-quality dehazing for a lightly hazy image, the dehazing ability of this method is limited by the small receptive size, and it cannot effectively deal with heavily hazy images.

DCPDN is a novel end-to-end jointly optimizable dehazing network formulated by directly embedding model (1) into the optimization framework via math operation modules. Consequently, this technique allows the network to optimize the transmission map, atmospheric light and dehazed image jointly. However, the dehazing ability of this technique is limited by using the atmospheric scattering model, which cannot suitably describe a naturally hazy image.

The GFN is a fusion-based dehazing method and generate visually pleasing results for most cases. However, the GFN cannot handle hazy images of large haze area. Furthermore, the dehazing ability of the GFN is limited by its derivations from the inputs, which are assumed to be containing clear cues to reconstruct the haze-free image.

To overcome the above-mentioned limitations, we propose to employ dilation convolutional layer to increase the receptive size and capture more context information to reconstruct the final dehazed result.

3 Proposed method

In this section, we describe the proposed method, including the motivation, network architecture, loss function and network training. The network architecture, which includes a structure learning subnetwork and a detail (residual) refining subnetwork, is shown in Fig. 1.

3.1 Motivations

Our first motivation is that objects can appear at vastly different scales. The coverage of many object scales is a critical problem for single image dehazing. Cai et al. introduced

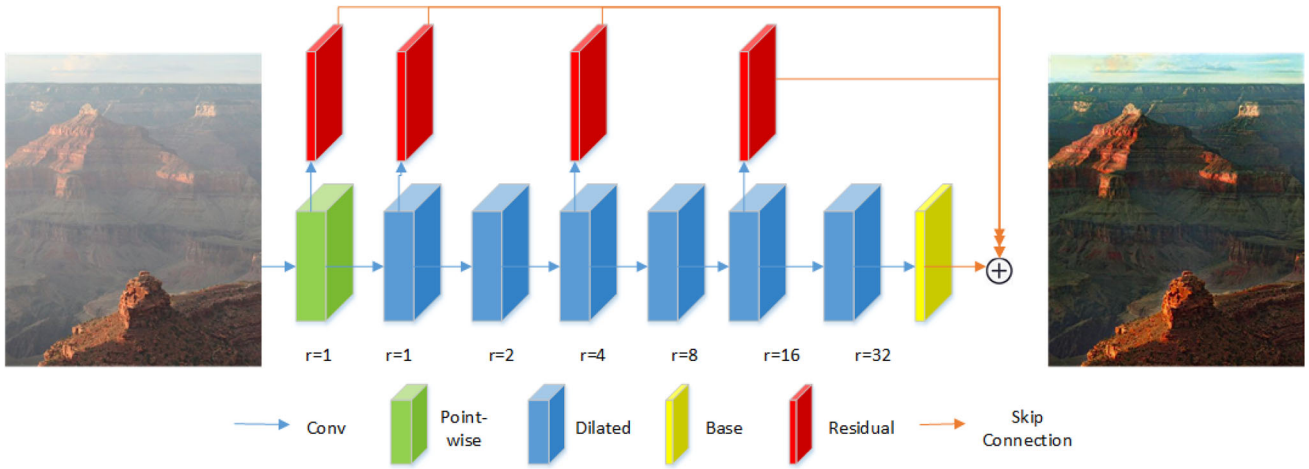

Fig. 1 The framework of the proposed DRCDN. Dilated convolution layers are used to extract representative features from hazy input image. Then the residual layers are used to recover image details from dilated features, which capture multiscale information well. Base layer is designed to reconstruct the coarse result and residual layer is designed to restore the image details

a multiscale layer in the DehazeNet to improve the dehazing quality. However, we note that the receptive field of the DehazeNet cannot match the spatial support of the object well, which results in the degradation of the dehazing performance. Dilation and skip connection techniques have also been used for single image dehazing [33]. However, the FEED-Net cannot effectively deal well with densely hazy images because its receptive field cannot effectively match that of small objects.

Our second motivation is that the use of dilation convolution has been proved to cause gridding artifacts [49]. When applying dilation convolution in dehazing, we note that it causes artifacts as shown in Fig. 3. Applying residual network to derain problems [50,51], Fu et al. extracted the base layer using a guided filter, and later used the detail layer to recover the final clear image. Furthermore, residual learning can be used to recover the image details and remove the halo artifacts.

In order to overcome the above-mentioned issues, we use the dilation convolution and residual learning techniques to recover the global structure and remove haze.

Dilation convolution is used to increase the receptive field of the model, which can ensure that the model captures the global structure and recovers a coarse dehazed result.

Residual learning is used to refine the dehazed result. Gridding artifacts can be considered as the lost details of a clear image, and they can be removed via residual learning. To recover the image details, our model learns the residual aspects from the shallow layers. We use a dilation contextual aggregation subnetwork to extract the base layer of the clear image.

Subsequently, we construct the detail layers from the shallow layers of the dilation contextual aggregation subnetwork and recover the final dehazed result by adding the detail layers to the base layer.

3.2 Network architecture

Based on the motivations described in Sect. 3.1, we propose a fully end-to-end deep learning-based method for single image dehazing. As shown in Fig. 1, our model consists of two subnetworks: The first subnetwork is designed to extract the coarse dehazed result, and the second subnetwork is used to recover the detail layers of the dehazed result. Subsequently, we add the coarse result to the multiscale detail layers to recover the final dehazing result.

Our coarse result recovering network is adopted from Yu et al.'s model [52], which is used for semantic segments. In this paper, we use the dilation contextual aggregation subnetwork to construct the coarse result, which requires the consideration of more high-level features. Thus, we recover the image detail layers from the first, second, fourth and sixth shallow layers. Because neighbor layers contain duplicate information, we select the sparse layers to reconstruct the image details.

Our coarse dehazed result reconstructing subnetwork consists of nine convolution layers, and the detail constructing subnetwork consists of four convolution layers; these layers are followed by an element-wise sum layer to reconstruct the final dehazing layer. The coarse image contains more structure information, and thus we recover the coarse image by using high-level feature maps.

As shown in Fig. 1, the convolution layers masked by red color following the dilation convolution layer are used to reconstruct the image detail layers. To recover the base layer of the clean image, we introduce the context information into the dehazing process via the dilation convolution, which

has been proved to enlarge the receptive field of the model and capture more contextual information. We use the dilation convolution technique to progressively enlarge the dilation rate, which provides multiscale context information for the base layer reconstruction.

To increase the flexibility of the network, we use a pointwise convolution in the first layer of the network. As mentioned in 3.1, the use of dilation results in gridding artifacts. We treat the gridding artifacts as the lost details of the clean image and recover them via a multiscale subnetwork, which restores the image details from the shallow layers, which contain more low-level information. Our model uses the high-level features to construct the coarse image layer and the low-level features to restore the image details.

3.3 Loss function

To train the deep learning network, the Euclidean loss ($L2$ loss) is widely used for regression problems. However, the Euclidean loss often generates a blurry result. To avoid the problem associated with the Euclidean loss, we use the $L1$ loss to train our network and define the L1 loss as follows:

$$\mathcal{L}_1 = \frac{1}{N}\sum_{i=1}^{N} \left\| \tilde{J}_i - J_i \right\|_1, \tag{6}$$

where $\tilde{J}_i$ denotes the predicted result, J_i represents the ground-truth haze-free image, N represents the number of train samples.

To improve the dehazing quality, we use the perceptual loss as an auxiliary loss, which is used to realize the style transfer and super-resolution, as proposed in [53]. The perceptual loss consists of the feature reconstruction loss and style reconstruction loss. Instead of encouraging the pixels of the dehazing image $\tilde{J}$ to exactly match the pixels of the ground-truth image, the feature reconstruction loss encourages these pixels to have similar feature representations. The feature reconstruction loss can be defined as follows:

$$\mathcal{L}^{\phi,j}_{\text{feature}} = \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{K} \left\| \phi_j(\tilde{J}_i) - \phi_j(J_i) \right\|_2^2, \tag{7}$$

where N denotes the number of pixels in features map, ϕ presents the VGG-19 network, which is trained on ImageNet [54] and j denotes the layer number.

We select the layers 'conv1-2', 'conv2-2', 'conv3-2', 'conv4-2', and 'conv5-2' in the VGG-19 network to compute the feature reconstruction loss. The style reconstruction loss can be defined as follows:

$$\mathcal{L}^{\phi,j}_{\text{style}} = \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{K} \left\| G_j(\tilde{J}_i) - G_j(J_i) \right\|_2^2, \tag{8}$$

where G denotes the *Gram matrix*, which can be defined as follows:

$$\mathcal{G}^{\phi}_{j}(x)_{c,c'} = \frac{1}{C_j W_j H_j}\sum_{h=1}^{H_j}\sum_{w=1}^{W_j} \left\| \phi_j(x)_{h,w,c} - \phi_j(x)_{h,w,c'} \right\|_2^2, \tag{9}$$

where W, H and C are the shape of feature in jth layer of the VGG network.

We select the layers 'conv1-2' and 'conv2-2' in the VGG-19 network to compute the style reconstruction loss. We define the perceptual loss as follows:

$$\mathcal{L}_{\text{per}} = \lambda_1 \sum_{j=1}^{5} \mathcal{L}^{\phi,j}_{\text{feature}} + \lambda_2 \sum_{j=1}^{2} \mathcal{L}^{\phi,j}_{\text{style}}, \tag{10}$$

where we use λ_1 and λ_2 to control the importance of feature reconstruction loss and style reconstruction loss, respectively.

Our overall loss function is:

$$\mathcal{L}(\tilde{J}, J) = \mathcal{L}_1 + \mathcal{L}_{\text{per}}. \tag{11}$$

3.4 Training and testing data

Training a deep learning-based model needs vast labeled data. However, collecting such a large dataset [40,56–58] is extremely expensive. To realize the training of a dehazing model, it is considerably more difficult to collect pairs of hazy images and the corresponding haze-free images.

Cai et al. proposed a method based on two assumptions: The image context is independent of the medium transmission, and the medium transmission is locally constant in an image patch. Based on these two assumptions and the physical haze formation model [1], Cai et al. synthesized the training pairs of hazy and haze-free image patches. However, this training dataset works well only for estimating the transmission map, and it cannot be used for end-to-end dehazing methods

Ren et al. proposed a method based on the NYU Depth dataset. The authors randomly selected clean images and the corresponding depth maps; subsequently, they used the depth information to generate the transmission map and synthesize the hazy images. Furthermore, Ren et al. proposed a multiscale network to estimate the transmission map.

To train our deep learning model, we adopt the same strategy as that used by Ren et al. to synthesize the training dataset. However, the synthesized indoor hazy image is not suitable for an end-to-end dehazing network. Consequently, we propose a new method to generate the outdoor hazy image dataset. Our dataset improves the accuracy of the synthesized

outdoor hazy image compared with that obtained using Cai et al.'s method.

Comparing with Ren et al.'s method, our dataset is more suitable for outdoor dehazing task. In order to generate a hazy image, we have to compute the depth of a haze-free image. During single clean image depth estimating, we estimate the depth using Liu et al.'s method [59] and then generate the corresponding transmission map using Eq. (2). We choose a random atmospheric light $A = [k, k, k]$, where $k \in [0.7, 1]$, and then we synthesize the final hazy image.

Compared with Ren et al.'s method, our dataset is more suitable for the outdoor dehazing task. To generate a hazy image, we must compute the depth of a haze-free image. To estimate the depth of a single clean image, we use Liu et al.'s method [59] , and we later generate the corresponding transmission map by using Eq. (2). We choose a random atmospheric light, $A = [k, k, k]$, where $k \in [0.7, 1]$, and subsequently, we synthesize the final hazy image.

To train our model, we generate a synthesized hazy image dataset corresponding to model (1) by using the proposed method. We first select 1, 100 clean images from the SYSU-Scene dataset, which has been used for semantic segments. Subsequently, we generate the corresponding depths for all the clean images by using the CNN method [59]. For a clean image, we generate 10 random medium scattering coefficients whose values range within $[0, 0.2]$. We do not use a large medium scattering coefficient because the resulting transmission maps tend to zero. Therefore, we obtain 11,000 hazy images and the corresponding haze-free images.

3.5 Implementation details

In this subsection, we describe the details of model training. The identity initializer [52] has been proved to be effective for the dilation convolution initialization. Consequently, we initialize all the dilation convolution layers by using the identity initializer, and we initialize the remaining layers by using Gaussian random variables.

ReLU has been widely used as an activation function. However, we note that the LReLU is more effective for our models. We choose 16 filters for all the layers except for the structure and detail layers. The last layer used for image dehazing is an element-wise sum layer. We use a batch size of two with an image size of (500×500) and learning rate of 0.0001.

During training, we use ADAM as the optimization algorithm to train our network. In this paper, we use the VGG-19 network trained on the ImageNet to calculate the perceptual loss, which helps our model recover a photo-realistic dehazing result.

We compute the feature reconstruction loss by using the layers 'conv1-2', 'conv2-2', 'conv3-2', 'conv4-2', and 'conv5-2' in the VGG-19 network. We compute the style reconstruction loss by using the layers 'conv1-2', 'conv2-2' in the VGG-19 network.

We set $\lambda_1 = 0.0005$ and $\lambda_2 = 0.001$ to balance the contribution of the losses. The model is trained for approximately 50 epochs for convergence, and it usually obtains sufficiently well dehazed results after 50 epochs.

4 Experimental results

In this section, we describe the quantitative and qualitative evaluation of the proposed method against several state-of-the-art methods on synthetic haze datasets, real haze datasets and real-world hazy images. We present and discuss the results obtained using the proposed method and compare them with the dehazing results of several state-of-the-art methods. We also present the time complexities of the proposed method and other methods, which demonstrates the high efficiency of our method.

4.1 Quantitative evaluation

Synthetic hazy image datasets have been widely used to evaluate the performance of dehazing methods [60,61]. To demonstrate the generalization ability of our method, we test its performance on two public hazy datasets [55,60]. These two datasets include indoor and outdoor hazy images. Indoor hazy images have been previously used to evaluate dehazing owing to the lack of a suitable outdoor hazy image dataset.

Due to the development of sensors, high-quality depth values have been captured and highly realistic hazy outdoor images have been generated [55]. We use these images to evaluate the dehazing methods. Furthermore, we test our model on two real haze datasets (I-HAZE and O-HAZE datasets).

RESIDE dataset A public synthetic indoor hazy dataset is available to evaluate the performance qualitatively as well as quantitatively. We evaluate our method on the SOTS [60] and compare it with several state-of-the-art single image dehazing methods by considering the peak signal to noise ratio (PSNR) and Structural Similarity Index (SSIM). Table 1 indicates that our method achieves higher scores in comparison with other state-of-the-art methods. It should be noted that we resize the ground truth to 512×512 and compute the PSNR and SSIM with the output of the DCPDN. Our method benefits considerably from the use of end-to-end learning, which avoids the estimation of the transmission map and atmospheric light.

Our method also overcomes the problem of inaccurate estimation of the transmission map and atmospheric light, which tends to degrade the final dehazing performance.

Table 1 Average PSNR/SSIM of dehazed results on the SOTS dataset from RESIDE

	DCP	CAP	NLD	MSCNN	DehazeNet	AOD-Net	DCPDN	GFN	PDNet	DRCDN
PSNR	16.62	19.05	17.29	17.57	21.14	19.06	15.86	22.30	22.83	*23.15*
SSIM	0.82	0.84	0.75	0.81	0.85	0.85	0.82	0.88	0.91	*0.92*

It means that DRCD achieves the best result when being compared with other methods

Table 2 Average PSNR/SSIM of dehazed results on the HazeRD dataset [55]

	DCP	CAP	NLD	MSCNN	DehazeNet	AOD-Net	DCPDN	GFN	PDNet	DRCDN
PSNR	17.66	18.56	17.47	19.10	19.53	18.13	18.82	19.18	20.14	*20.32*
SSIM	0.84	0.83	0.79	0.85	0.85	0.83	0.89	0.86	0.89	*0.90*

It means that DRCD achieves the best result when being compared with other methods

Table 3 Quantitative evaluation on the I-HAZE dataset

	DCP	CAP	NLD	MSCNN	DehazeNet	AOD-Net	DCPDN	GFN	PDNet	DRCDN
PSNR	15.29	15.94	15.94	17.28	15.06	15.71	15.71	15.30	15.00	17.78
SSIM	0.71	0.71	0.77	0.79	0.77	0.61	0.71	0.72	0.65	0.77

Compared with the AOD-Net, GFN and PDNet approaches, whose performance is limited by the receptive size, the dilation subnetwork of the proposed model can help our model increase the receptive size and suitably capture the global structure.

Hazerd dataset Haze is an outdoor phenomenon, and thus performing an evaluation on the synthetic outdoor hazy image is more suitable for the dehazing process. To realize the evaluation of the dehazing methods on the outdoor dataset, the Proximal DehazeNet [61] simulates a real outdoor hazy image dataset, which includes 128 images with different haze levels. Since all the learning-based methods do not include images in the Hazerd as training data, it is a fair approach to evaluate these methods on the Hazerd. We refer to the Proximal DehazeNet as PDNet. As shown in Table 2 , our deep learning-based method achieves the highest PSNR and SSIM on the Hazerd and outperforms the second best learning-based method Proximal DehazeNet [61] by 0.18 dB in terms of the PSNR.

Compared with other methods trained on indoor hazy images or synthetic outdoor hazy patches, our method can generally manage synthetic outdoor hazy images well. The DehazeNet is trained on synthetic outdoor hazy patches, which induces halo artifacts in the estimated transmission map.

To eliminate the halo artifacts, DehazeNet employs a guided filter to smooth the estimated transmission map, which results in an inaccurate transmission map. The GFN, which is trained on indoor hazy images, generates a satisfactory result on the indoor hazy image dataset. However, it cannot deal with the outdoor hazy images well.

We also note that the PDNet performs satisfactorily for the indoor and outdoor hazy images. We infer that the dark channel prior is driven from outdoor images, which results in satisfactory generalization on the outdoor images, and because the model is trained on indoor images, a satisfactory generalization is obtained for the indoor images. Our model is trained on outdoor images, and the large receptive size can effectively capture the structure and haze distribution; these aspects help our model generate excellent results for both indoor and outdoor hazy images.

I-HAZE and O-HAZE dataset In contrast to the Hazerd and RESIDE data, I-HAZE and O-HAZE generate haze by using a professional haze machine, which leads to a higher resemblance with real hazy images. I-HAZE provides 35 hazy indoor images with the corresponding haze-free (ground-truth) indoor images. O-HAZE provides 45 hazy outdoor images with the corresponding haze-free (ground-truth) images. To quantitatively evaluate the performance of the proposed method and the state-of-the-art methods, we directly compare the output with the ground-truth (haze-free) images by using the PSNR and SSIM. As shown in Tables 3 and 4 , our method can suitably deal with real hazy images and demonstrate a competitive performance.

From the results of the three experiments described above, we can see that the proposed method can achieve a high quality on both the indoor and outdoor hazy images. Although our model is trained on outdoor images, it effectively captures the distribution of haze, which helps our model effectively to generate indoor hazy images. Furthermore, our model can generate real hazy images and achieve state-of-the-art performance.

Table 4 Quantitative evaluation on the O-HAZE dataset

	DCP	CAP	NLD	MSCNN	DehazeNet	AOD-Net	DCPDN	GFN	PDNet	DRCDN
PSNR	16.59	17.62	16.61	19.07	16.21	16.72	15.62	18.30	18.52	19.15
SSIM	0.74	0.71	0.75	0.77	0.67	0.68	0.62	0.72	0.75	0.77

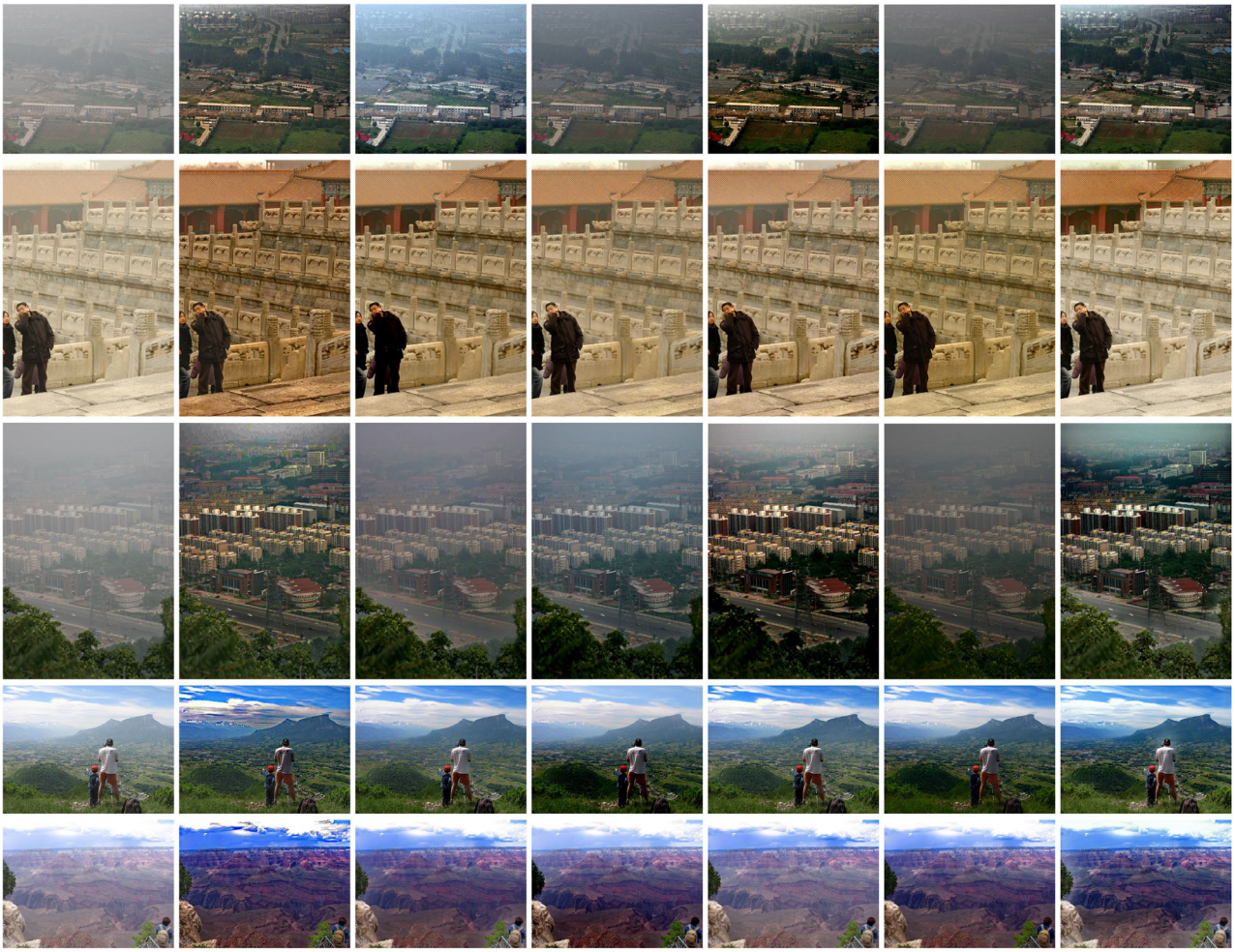

Fig. 2 Visual comparison on real-world images. We can notice that AOD-Net tends to leave haze in dehazed results. The results of BCCR contain some color distortion and over enhancement. MSCNN loses some details, which can be seen in leave area in third row. In contrast, our method can generate a haze-free result while preserve the image detail well

4.2 Qualitative comparison

Although we demonstrated the effectiveness of our model on synthetic hazy images datasets, the performance of removing the haze from a natural hazy image is critical for dehazing methods. To demonstrate the generalization ability of the proposed method when dealing with real-world natural hazy images, we evaluate the proposed method on several challenging naturally hazy images provided using previous methods and collected from the Internet. The dehazed results are compared against those of five state-of-the-art methods.

First, we compare our method with the state-of-art methods on five challenging hazy images [12,14,27,28,31]. The results for the twelve sample images obtained from the previous methods are shown in Fig. 2. The methods [12] tend to generate color distortion, which also reduces the performance of the model visual system. The methods [27,28,31] tend to retain haze in the final restored results. The CNN benefits considerably from the use of a large amount of data and large receptive size. Compared with the DehazeNet and AOD-Net, the MSCNN is more effective in removing haze due to its large receptive size and training on a large indoor

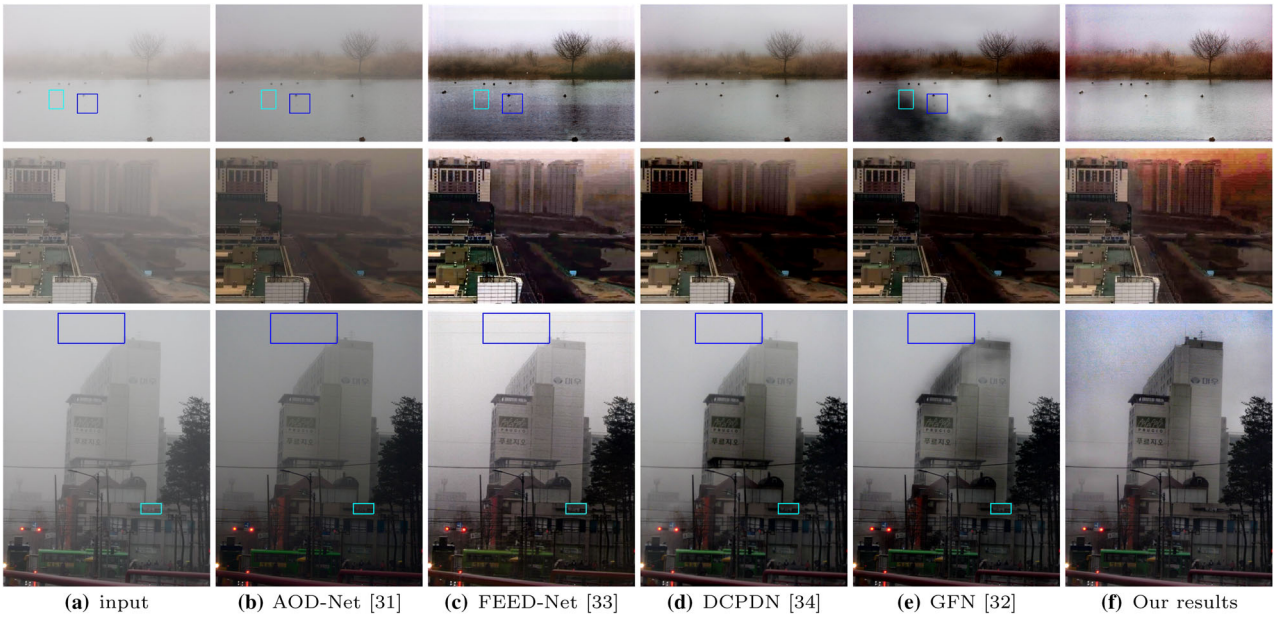

Fig. 3 Visual comparison on real-world images. FEED-Net generates white artifacts, which can be seen in "lake" image. DCPDN loses image details, which is shown in "riverside" image. The results of GFN contain color distortion. In contrast, our method can remove haze well with preserving image detail

hazy dataset, which can be seen from the second and fourth rows. The CAP estimates the transmission map based on the pixel information, which means that it cannot effectively deal with dense haze. In contrast, our model has a large perceptive size and is trained on large outdoor hazy images, and consequently, it can generate realistic colors while removing haze in a moderate manner.

Second, we demonstrate the high performance of the proposed method by comparing it with four end-to-end methods [31–34] on heavily hazy images. As shown in Fig. 3, the AOD-Net tends to retain haze in the dehazed result. Compared with the AOD-Net, the GFN tends to generate color distortion, which indicates that its ability of removing haze is better than the AOD-Net, thereby introducing the color distortion. Furthermore, we note that the perceptive size of the GFN is small, which results in the GFN being unable to capture the global structure well. The FEED-Net tends to generate a certain amount of noise in the dehazed result, which is induced by the dilation convolution. The DCPDN loses some details in the dehazed result. In contrast, our method can remove haze well and preserve the image details.

Last, we provide more examples for natural hazy images and the dehazed results of the state-of-the-art methods [2,12,14,27,36,37]. As shown in Fig. 4, the FVR [36] tends to generate color distortion and halo artifacts. The BCCR and DCP tend to over enhance the result and generate the color distortion. The RF [37] and CAP overestimate the transmission map and retain the haze in the final dehazed result. The DehazeNet tends to lose details, which can be seen in the red rectangles. In contrast, our method can remove the haze and retain the image details.

4.3 Ablation study

To validate the performances obtained by each module in the proposed network, we perform an ablation study involving the following five experiments:

(1) We replace the dilation convolutional layer with a traditional convolutional one, which demonstrates the effectiveness of the dilation convolutional layer. We denote this case as w/o dilation;
(2) We design a module to extract a detail layer from the hazy image directly and add it to base layer. We denote this case as DFH;
(3) We also train a network without perceptual loss and denote this case as w/o perceptual;
(4) This case involves the full proposed model with all the modules introduced in Sect. 3.

As shown in Table 5, our model can capture the structure effectively and recover more image details.

4.4 Comparison of running times

The lightweight structure of the DRCDN results in real-time dehazing. We select 50 images from our synthetic outdoor hazy image dataset to test all the methods on the same

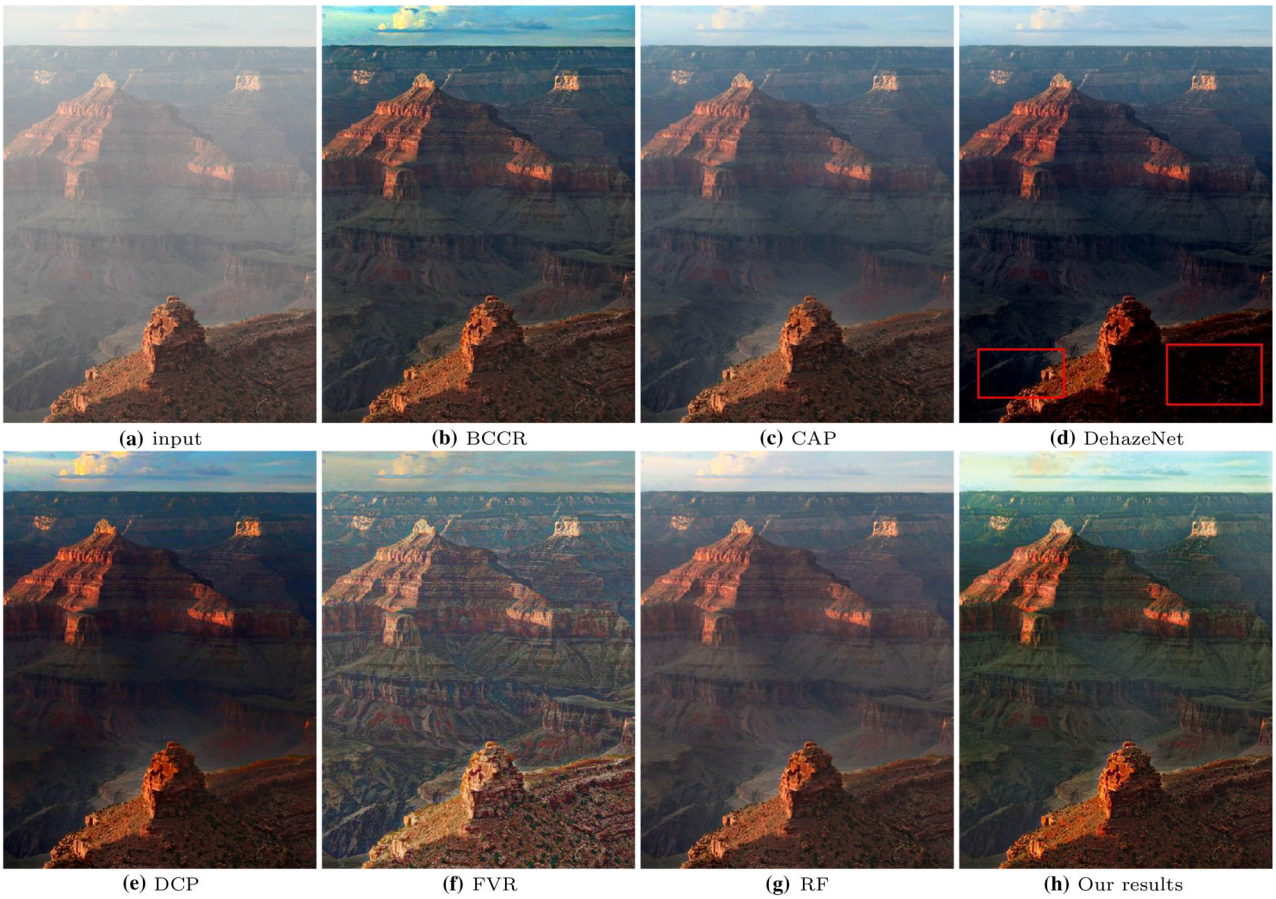

Fig. 4 Visual comparison on real-world images. The proposed method generates much clearer images with finer details

Table 5 Quantitative PSNR and SSIM results on the synthetic dataset using different loss configurations

	W/O dilation	DFH	W/O perceptual	Full model
PSNR	19.53	22.5	22.85	23.15
SSIM	0.82	0.88	0.89	0.92

Table 6 Average running time on the synthetic outdoor hazy images

Method	He	Meng	Berman	Ren	Cai	Our
Platform	Matlab					Tensorflow
Time (s)	25.08	3.52	8.41	2.45	2.56	2.02

machine (Intel(R) Core(TM) i5-6300HQ CPU@2.3 GH and 8 GB memory).

The per-image average running times of all the methods are listed in Table 6. The result demonstrates the promising efficiency of our method.

5 Conclusion and future work

In this paper, we present a fully end-to-end dehazing method, which restores a coarse clean image using high-level feature maps. Subsequently, the approach recovers the image details using low-level feature maps. Compared with existing learning-based methods that learn a relation between the hazy images and corresponding clean images, our method is more generalizable and can generate more visually pleasing results.

The results of experiments performed on the public synthetic outdoor and indoor image datasets demonstrate the effectiveness of the proposed method. The extensive experimental comparisons on natural hazy images show that the proposed method generates dehazed results with higher quantitative and qualitative aspects compared to those generated by using state-of-the-art methods. In future work, we will extend the proposed idea to other topics of visual computing [62–64].

Acknowledgements We thank anonymous reviewers very much for their suggestive comments. This work is partially supported by the NSFC (No. 61472289, 41571436).

Compliance with ethical standards

Conflict of interest The authors declares that there is no conflict of interest.

References

1. Narasimhan, S.G., Nayar, S.K.: Vision and the atmosphere. Int. J. Comput. Vision **48**(3), 233–254 (2002)
2. He, K., Sun, J., Tang, X.: Single image haze removal using dark channel prior. IEEE Trans. Pattern Anal. Mach. Intell. **33**(12), 2341–2353 (2011)
3. Schechner, Y.Y., Narasimhan, S.G., Nayar, S.K : Instant dehazing of images using polarization. In: Computer Vision and Pattern Recognition, vol. 1, pp. 325–332 (2001)
4. Narasimhan, S.G., Nayar, S.K.: Contrast restoration of weather degraded images. IEEE Trans. Pattern Anal. Mach. Intell. **25**(6), 713–724 (2003)
5. Shwartz, S., Namer, E., Schechner, Y.Y : Blind haze separation. In: Computer Vision and Pattern Recognition, vol. 2, pp. 1984–1991 (2006)
6. Kopf, J., Neubert, B., Chen, B., Cohen, M., Cohen-Or, D., Deussen, O., Uyttendaele, M., Lischinski, D.: Deep photo: model-based photograph enhancement and viewing. In: ACM transactions on graphics, vol. 27, Article No. 116 (2008)
7. Chen, X., He, F.: A matting method based on full feature coverage. Multimedia Tools Appl. **78**(9), 11173–11201 (2019)
8. Yu, H., He, F.: A novel segmentation model for medical images with intensity inhomogeneity based on adaptive perturbation. Multimedia Tools Appl. **78**(9), 11779–11798 (2019)
9. Haiping, Y., He, F., Pan, Y.: A novel region-based active contour model via local patch similarity measure for image segmentation. Multimedia Tools Appl. **77**(18), 24097–24119 (2018)
10. Tan, R.T: Visibility in bad weather from a single image. In: Computer Vision and Pattern Recognition (2008)
11. Fattal, R.: Single image dehazing. ACM Trans. Gr. **27**(3), 72 (2008)
12. Meng, G., Wang, Y., Duan, J., Xiang, S., Pan, C.: Efficient image dehazing with boundary constraint and contextual regularization. In: International Conference on Computer Vision, pp. 617–624 (2013)
13. Fattal, R.: Dehazing using color-lines. ACM Trans. Gr. **34**(1), 13 (2014)
14. Zhu, Q., Mai, J., Shao, L.: A fast single image haze removal algorithm using color attenuation prior. IEEE Trans. Image Process. **24**(11), 3522–3533 (2015)
15. Berman, D., Avidan, S., et al.: Non-local image dehazing. In: Computer Vision and Pattern Recognition, pp. 1674–1682 (2016)
16. Li, K., He, F., Haiping, Y., Chen, X.: A parallel and robust object tracking approach synthesizing adaptive bayesian learning and improved incremental subspace learning. Front. Comput. Sci. **13**(5), 1116–1135 (2019)
17. Ren, W., Liu, S., Ma, L., Qianqian, X., Xiangyu, X., Cao, X., Junping, D., Yang, M.-H.: Low-light image enhancement via a deep hybrid network. IEEE Trans. Image Process. **28**(9), 4364–4375 (2019)
18. Ren, W., Zhang, J., Ma, L., Pan, J., Cao, X., Zuo, W., Liu, W., Yang, M.-H.: Deep non-blind deconvolution via generalized low-rank approximation. In: Advances in Neural Information Processing Systems, pp. 297–307 (2018)
19. Li, H., He, F., Yan, X.: IBEA-SVM an indicator-based evolutionary algorithm based on pre-selection with classification guided by SVM. Appl. Math.-A J. Chin. Univ. **34**(1), 1–26 (2019)
20. Li, H., He, F., Liang, Y., Quan, Q.: A dividing-based many-objective evolutionary algorithm for large-scale feature selection. Soft Comput. (2019). https://doi.org/10.1007/s00500-019-04324-5
21. Yan, Y., Ren, W., Cao, X.: Recolored image detection via a deep discriminative model. IEEE Trans. Inf. Forensics Secur. **14**(1), 5–17 (2018)
22. Ding, B., Long, C., Zhang, L., Xiao, C.: ARGAN: attentive recurrent generative adversarial network for shadow detection and removal. In: International Conference on Computer Vision (2019)
23. Yong, J., He, F., Li, H., Zhou, W.: A novel bat algorithm based on cross boundary learning and uniform explosion strategy. Appl. Math.-A J. Chin. Univ. (2019). https://doi.org/10.1007/s11766-019-3714-1
24. Luo, J., He, F., Yong, J.: An efficient and robust bat algorithm with fusion of opposition-based learning and whale optimization algorithm. Intell. Data Anal. 24(3: to appear in this issue) (2020)
25. Zhang, W., Xiao, C: PCAN: 3D attention map learning using contextual information for point cloud based retrieval. In: the IEEE Conference on Computer Vision and Pattern Recognition, pp. 12436–12445 (2019)
26. Hou, N., He, F., Zhou, Y., Chen, Y.: An efficient GPU-based parallel tabu search algorithm for hardware/software co-design. Front. Comput. Sci. (2020). https://doi.org/10.1007/s11704-019-8184-3
27. Cai, B., Xiangmin, X., Jia, K., Qing, C., Tao, D.: Dehazenet: an end-to-end system for single image haze removal. IEEE Trans. Image Process. **25**(11), 5187–5198 (2016)
28. Ren, W., Liu, S., Zhang, H., Pan, J., Cao, X., Yang, M.-H.: Single image dehazing via multi-scale convolutional neural networks. In: European Conference on Computer Vision, pp. 154–169 (2016)
29. Sulami, M., Glatzer, I., Fattal, R., Werman, M.: Automatic recovery of the atmospheric light in hazy images. In: IEEE International Conference on Computational Photography, pp. 1–11 (2014)
30. Berman, D., Treibitz, T., Avidan, S.: Air-light estimation using haze-lines. In: IEEE International Conference on Computational Photography, pp. 1–9 (2017)
31. Li, B., Peng, X., Wang, Z., Xu, J., Feng, D.: Aod-net: all-in-one dehazing network. In: International Conference on Computer Vision, pp. 4770–4778 (2017)
32. Ren, W., Ma, L., Zhang, J., Pan, J., Cao, X., Liu, W., Yang, M.-H.: Gated fusion network for single image dehazing. In: Computer Vision and Pattern Recognition, pp. 3253–3261 (2018)
33. Zhang, S., Ren, W., Yao, J.: Feed-net: Fully end-to-end dehazing. In: IEEE International Conference on Multimedia and Expo, pp. 1–6 (2018)
34. Zhang, H., Patel, V.M.: Densely connected pyramid dehazing network. In: Computer Vision and Pattern Recognition, pp. 3194–3203 (2018)
35. Ancuti, C.O., Ancuti, C.: Single image dehazing by multi-scale fusion. IEEE Trans. Image Process. **22**(8), 3271–3282 (2013)
36. Tarel, J.-P., Hautiere, N.: Fast visibility restoration from a single color or gray level image. In: International Conference on Computer Vision, pp. 2201–2208 (2009)
37. Tang, K., Yang, J., Wang, J.: Investigating haze-relevant features in a learning framework for image dehazing. In: Computer Vision and Pattern Recognition, pp. 2995–3000 (2014)
38. Pan, Y., He, F., Yu, H.: A correlative denoising autoencoder to model social influence for top-n recommender system. Front. Comput. Sci. (2019). https://doi.org/10.1007/s11704-019-8123-3
39. Pan, Y., He, F., Yu, H.: Learning adaptive trust strength with user roles of truster and trustee for trust-aware recommender systems. Appl. Intell. (2019). https://doi.org/10.1007/s10489-019-01542-0
40. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems, pp. 1097–1105 (2012)
41. Szegedy, C., Toshev, A., Erhan, D.: Deep neural networks for object detection. In: Advances in Neural Information Processing Systems, pp. 2553–2561 (2013)

42. Ren, S., He, K., Girshick, R., Sun, J.: Faster R-CNN: towards real-time object detection with region proposal networks. In: Advances in Neural Information Processing Systems, pp. 91–99 (2015)
43. Yu, J., Jiang, Y., Wang, Z., Cao, Z., Huang, T.: Unitbox: an advanced object detection network. In: Proceedings of the 2016 ACM on Multimedia Conference, pp. 516–520 (2016)
44. Xie, J., Xu, L., Chen, E.: Image denoising and inpainting with deep neural networks. In: Advances in Neural Information Processing Systems, pp. 341–349 (2012)
45. Dong, C., Loy, C.C., He, K., Tang, X.: Xiaoou: image super-resolution using deep convolutional networks. IEEE Trans. Pattern Anal. Mach. Intell. **38**(2), 295–307 (2016)
46. Kim, J., Kwon Lee, J., Mu Lee, K.: Accurate image super-resolution using very deep convolutional networks. In: Computer Vision and Pattern Recognition, pp. 1646–1654 (2016)
47. Liu, D., Wen, B., Liu, X., Huang, T.S.: When image denoising meets high-level vision tasks: a deep learning approach. In: International Joint Conferences on Artificial Intelligence, pp. 842–848 (2017)
48. Zhang, S., He, F., Ren, W., Yao, J.: Joint learning of image detail and transmission map for single image dehazing. Vis. Comput. (2018). https://doi.org/10.1007/s00371-018-1612-9
49. Yu, F., Koltun, V., Funkhouser, T.A: Dilated residual networks. In:Computer Vision and Pattern Recognition, vol. 2, p. 3 (2017)
50. Fu, X., Huang, J., Zeng, D., Huang, Y., Ding, X., Paisley, J.: Removing rain from single images via a deep detail network. In:Computer Vision and Pattern Recognition, pp. 3855–3863 (2017)
51. Mehta, S., Rastegari, M., Caspi, A., Shapiro, L., Hajishirzi, H.: Espnet: efficient spatial pyramid of dilated convolutions for semantic segmentation. In: European Conference on Computer Vision, pp. 552–568 (2018)
52. Yu, Fi., Koltun, V.: Multi-scale context aggregation by dilated convolutions. In: International Conference on Learning Representations (2016). arXiv:1511.07122
53. Johnson, J., Alahi, A., Fei-Fei, L.: Perceptual losses for real-time style transfer and super-resolution. In: European Conference on Computer Vision, pp. 694–711 (2016)
54. Russakovsky, O., Deng, J., Hao, S., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M.: Imagenet large scale visual recognition challenge. Int. J. Comput. Vision **115**(3), 211–252 (2015)
55. Zhang, Y., Ding, L., Sharma, G.: Hazerd: an outdoor scene dataset and benchmark for single image dehazing. In: International Conference on Image Processing, pp. 3205–3209 (2017)
56. Li, K., He, F., Yu, H.: Robust visual tracking based on convolutional features with illumination and occlusion handing. J. Comput. Sci. Technol. **33**(1), 223–236 (2018)
57. Mbelwa, J.T., Zhao, Q., Wang, F.: Visual tracking tracker via object proposals and co-trained kernelized correlation filters. Vis. Comput. (2019). https://doi.org/10.1007/s00371-019-01727-1
58. Pan, Y., He, F., Haiping, Y.: A novel enhanced collaborative autoencoder with knowledge distillation for top-n recommender systems. Neurocomputing **332**, 137–148 (2019)
59. Liu, F., Shen, C., Lin, G., Reid, I.: Learning depth from single monocular images using deep convolutional neural fields. IEEE Trans. Pattern Anal. Mach. Intell. **38**(10), 2024–2039 (2016)
60. Li, B., Ren, W., Fu, D., Tao, D., Feng, D., Zeng, W., Wang, Z.: Benchmarking single image dehazing and beyond. In: IEEE Transactions on Image Processing, pp. 492–505 (2018)
61. Yang, D., Sun, J.: Proximal dehaze-net: A prior learning-based deep network for single image dehazing. In: European Conference on Computer Vision, pp. 702–717 (2018)
62. FazlErsi, E., Kazemi Nooghabi, M.: Revisiting correlation based filters for low-resolution and long-term visual tracking. Vis. Comput. **35**(10), 1447–1459 (2019)
63. Doyle, L., David Mould, D.: Augmenting photographs with textures using the laplacian pyramid. Vis. Comput. **35**(10), 1489–1500 (2019)
64. Umer, S., Dhara, B.C., Chanda, B.: NIR and VW iris image recognition using ensemble of patch statistics features. Vis. Comput. **35**(9), 1327–1344 (2019)

Shengdong Zhang is currently a Ph.D. candidate in School of Computer Science, Wuhan University. His research interests are image processing, computer graphics and deep learning.

Fazhi He received bachelors, masters and Ph.D. degrees from Wuhan University of Technology. He was a post-doctor researcher in The State Key Laboratory of CAD&CG at Zhejiang University, a visiting researcher in Korea Advanced Institute of Science & Technology and a visiting faculty member in the University of North Carolina at Chapel Hill. Now he is a professor in School of Computer Science, Wuhan University. His research interests are artificial intelligence, intelligent computing, computer graphics, image processing, computer-aided design, computer-supported cooperative work and co-design of software/hardware.