The Visual Computer
https://doi.org/10.1007/s00371-018-1612-9

ORIGINAL ARTICLE

Joint learning of image detail and transmission map for single image dehazing

Shengdong Zhang[1] · Fazhi He[1] · Wenqi Ren[2] · Jian Yao[3]

Abstract
Single image haze removal is an important task in computer vision. However, haze removal is an extremely challenging problem due to its massively ill-posed, which is that at each pixel we must estimate the transmission and the global atmospheric light from a single color measurement. In this paper, we propose a new deep learning-based method for removing haze from single input image. First, we estimate a transmission map via joint estimation of clear image detail and transmission map, which is different from traditional methods only estimating a transmission map for a hazy image. Second, we use a global regularization method to eliminate the halos and artifacts. Experimental results on synthetic dataset and real-world images show our method outperforms the other state-of-the-art methods.

Keywords Joint learning · Dehazing · Image detail estimating · Non-local regularization · Transmission estimating

1 Introduction

Images captured in bad weather conditions often suffer from poor visibility and low contrast due to haze, small particles in the air that scatter the light in the atmosphere. The light is absorbed and scattered by the particles in the atmosphere, and it is also fused with airlight reflected from other directions. This process reduces the contrast and fades the color of the haze images, and the haze images show poor visual vividness. As shown in Fig. 1a, the haze image loses color fidelity and contrast, which makes it hard to distinguish the objects from the scene. As stated in [1], the amount of scattering depends on the distance between the camera and the scene points; the image degradation increases with the distance from the camera.

Restoring the image captured in these bad weather conditions has caught great interest in the last years due to its wide applications. Haze removal can significantly improve the contrast and visibility of the scene and correct the color shift caused by airlight. As such, the dehazed image is often more visually friendly and is highly needed in computer vision applications. Most computer vision algorithms [2–7] often assume that the input image/video is haze-free. The performance of the computer vision algorithm inevitably suffers from low contrast and the color shift of haze image. Last, haze removal can benefit advanced image editing [8] and many computer vision algorithm.

Our goal is to recover the RGB values of the haze-free image for each pixel from hazy image. However, removing haze is an extremely challenging problem as the haze is dependent on the unknown depth. Obviously, this is a severely under-constrained problem, especially when the input is only a single hazy image, because there are only three equations available and at least four unknowns for each pixel, and there is inherent ambiguity between haze and object radiance. To address this ambiguity, some previous methods used additional information such as more images or depth map, while others assumed an image prior to solve the problem from a single image.

The hazy images can be model as [9]:

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

where $\mathbf{I}$ represents the hazy image, $\mathbf{J}$ represents the haze-free image, $\mathbf{A}$ denotes the global atmospheric light, t represents

✉ Fazhi He
fzhe@whu.edu.cn

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

2 Institute of Information Engineering, Chinese Academy of Sciences, Beijing, China

3 School of Remote Sensing and Information Engineering, Wuhan University, Wuhan, China

Published online: 10 November 2018

(a) **(b)**

Fig. 1 Haze removal using our method. **a** Input haze image. **b** Image recovered by our method

the transmission describing the probability of light reaching the camera. To remove haze from the haze image is equal to solve the $\mathbf{A}$, t from $\mathbf{I}$. The term $\mathbf{J} \times t(i, j)$ we call it as direct attenuation, and the term $(1 - t(i, j)) \times \mathbf{A}$ we call it as airlight $\mathbf{A}$ contributions.

Based on Eq. (1), a lot of methods [1,10–18] have been proposed. Recently, dramatic progress has been made on single image dehazing with the advent of powerful learning-based technology [16,19–22]. Transmission estimation has been applied to dehazing successfully [16,19–21]. However, it is still in the initial stage for dehazing using image detail estimation [22]. Although these methods have been applied to dehazing and achieve great success, if the transmission map is not well estimated, they will accordingly interfere the final recovered image.

Joint learning aims to learn two tasks simultaneously to boost the performance of main task or all ones. It has been successfully applied to pose estimation. Despite the success of joint learning in various vision problems, there is no comprehensive study of joint learning for dehazing. In this paper, we study dehazing as a joint learning problem where image detail is the main task with transmission estimation being the an auxiliary one. We answer the questions on how and why transmission estimation can help single image dehazing. In order to use joint learning, we rewritten Eq. (1) as follows:

$$\mathbf{I}(x) - \mathbf{A}(x) = (\mathbf{J}(x) - \mathbf{A}(x)) \times t(x), \tag{2}$$

we call $(\mathbf{J} - \mathbf{A})$ as clear image detail and t as transmission map.

Based on Eq. (2), we find that a hazy image can be used to estimate transmission map and details of clear image. It is not trivial to formulate the joint estimation of detail and transmission, in which the two tasks could benefit each other. In this paper, we propose a novel method to train a network to joint estimate the detail and transmission.

In this paper, we propose a new framework to integrate transmission map and clear image detail using unified CNN by joint learning. The main contributions of our paper are listed as follows:

- To the best of our knowledge, we are the first to train a network by estimating transmission map and image details simultaneously, which can be used to improve the dehazing result.
- We propose a new global regularization. In this paper, we show that global regularization can be used to remove block artifacts and improve the quality of dehazing.
- Extensive experiments are conducted to show the state-of-the-art performance of our method. We conduct experiments on two synthetic datasets and one real-world image dataset, and the result shows that our method outperforms the other state-of-the-art methods.

2 Related works

Due to the wide application of dehazing, a lot of methods have been proposed. In this section, we review the methods related to dehazing.

2.1 Traditional methods

Based on the prior or assumption, a lot of state-of-the-art methods [1,10–17,23–25] were proposed based on the statistical prosperities of clear images. However, these methods may loss effect when the proposed assumptions or priors are not satisfied in some specific cases, for example, DCP [1] cannot deal large white area well. All these state-of-the-art methods are limited by the haze-relevant priors or assumption. In this paper, we propose a method to learn more effective features for haze removal using deep convolution neutral networks.

2.2 Learning-based methods

Tang et al. [19] investigate the different haze-relevant feature of hazy image and use the best suitable feature combination to estimate the corresponding transmission map for a hazy image. Zhu et al. [16] propose a learning-based method which considers the transmission as a linear combination of the saturation and brightness of pixels in a hazy image. They use a learning strategy to get the parameters of the model. The most relative works to ours are Cai et al.[21] and [20], which are also a deep learning-based method for estimating of transmission map. Ren et al. [20] propose a multi-scale deep convolutional neural networks to dehazing. In contrast, our method estimates a haze-free image from hazy image directly. Compared with other learning-based methods, our network is much simpler and generates high-quality dehazed results. Li et al. [26] propose an end-to-end dehazing networks, but we find that it may fail for some images.

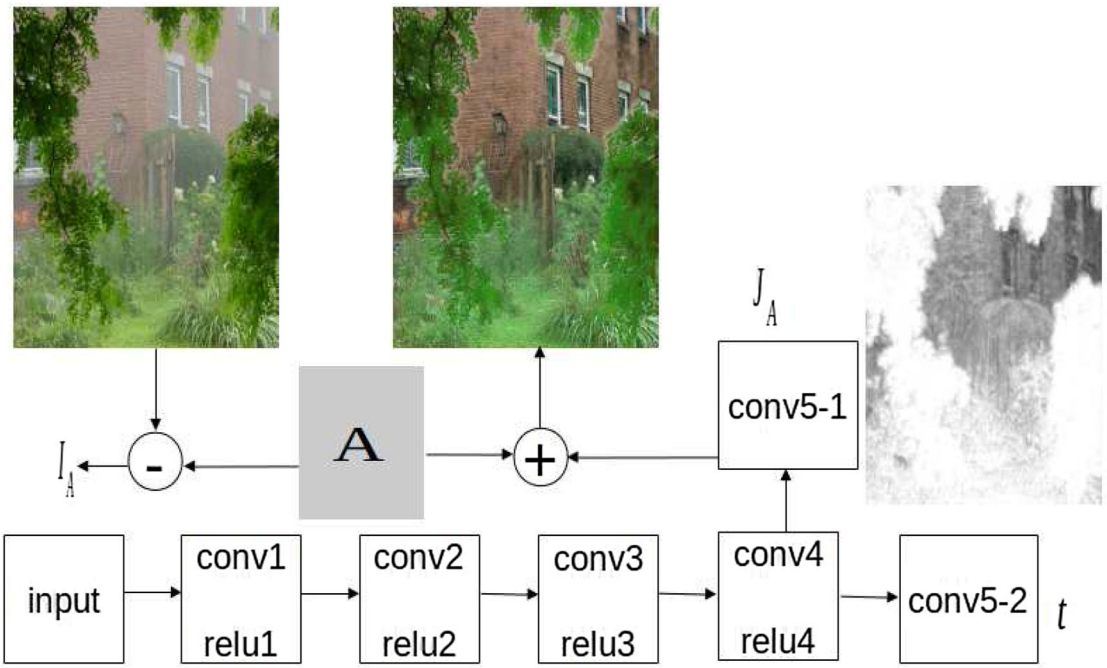

Fig. 2 CNN architecture used for haze removal experiments. Our CNN model consists of six layers. The first layer extracts 16 feature maps from input hazy image using 9×9 kernel size. The second to fourth layer is of the same type: 16 filters of the size $3 \times 3 \times 16$. The conv5-1 is used to predict image detail using 3 filters of the size $3 \times 3 \times 16$, and conv5-2 is used to predict transmission map using 1 filters of the size $3 \times 3 \times 16$

3 Detail and transmission estimating

In this section, we describe our method in detail and show the architecture of networks in Fig. 2. In order to describe our method clearly, we firstly introduce some notations.

3.1 Notation

We estimate A using the methods [1,27]. We define I_A,$r(x)$ and J_A as follows:

$$\mathbf{I}_A(x) = \mathbf{I}(x) - \mathbf{A}, \tag{3}$$
$$r(x) = ||\mathbf{I}_A(x)||, \tag{4}$$
$$\mathbf{J}_A(x) = \mathbf{J}(x) - \mathbf{A}, \tag{5}$$
$$\tilde{r}(x) = ||\mathbf{J}_A(x)||, \tag{6}$$

where $\mathbf{I}_A$ translates the 3D RGB coordinate of hazy image to sphere coordinate, in which the airlight is at the origin; we call $\mathbf{I}_A$ as hazy image detail. $r(x)$ represents the distance in RGB space of each pixel in hazy image to the airlight, $\mathbf{J}_A$ translates a clear image $\mathbf{J}$ to sphere coordinate, and we call $\mathbf{J}_A$ as clear image detail. $\tilde{r}(x)$ represents the distance in RGB space of each pixel in haze-free image to the airlight.

3.2 Motivation

Based on Eq. (2), we find that transmission map and clear image detail can be inferred from hazy image detail. However, we only find transmission map is used in convolution neural networks [20,21]. Recently, clear image detail is used to dehazing in convolution neural networks [22]. But all methods ignore the relation between the clear image detail and transmission map. In this paper, we exploit the relation between the clear image detail and transmission map to remove haze from single hazy image.

In order to study the relation between transmission map and the clear image detail, Eq. (2) can be rewritten as follows:

$$t = \mathbf{I}_A / \mathbf{J}_A, \tag{7}$$

where $\mathbf{J}_A$ is the clear image detail and $\mathbf{I}_A$ is the hazy image detail. Based on Eq. (7), transmission map can be inferred from clear image detail; transmission map and clear image detail can be treated as the same thing. In [1], He et al. assume that $\mathbf{A}$, $\mathbf{I}$ and $\mathbf{J}$ are coplanar and their end points are collinear in RGB color space, based on this assumption 7 can be rewritten as follows:

$$t(x) = ||\mathbf{I}_A(x)||/||\mathbf{J}_A(x)|| = \frac{\mathbf{I}^c(x) - \mathbf{A}^c(x)}{\mathbf{J}^c(x) - \mathbf{A}^c(x)}, \tag{8}$$

where c denotes one channel of R, G, B, $\mathbf{I}^c(x)$ represents the c channel value of $\mathbf{I}(x)$, $\mathbf{J}^c(x)$ represents the c channel value of $\mathbf{J}(x)$, $\mathbf{A}^c(x)$ represents the c channel value of $\mathbf{A}(x)$.

Based on Eq. (8), we can get the following equation:

$$\mathbf{I}^c(x) - \mathbf{A}^c(x) = t \times (\mathbf{J}^c(x) - \mathbf{A}^c(x)) \tag{9}$$

we can find for three channels, the length ratio will be same. However, only using convolution neural networks can satisfy this restriction well.

Compared with dehazing using transmission map, we find clear image detail is more robust. For a pixel with a small transmission value $t = 0.2$ and pixel intensity value [0.65, 0.70, 0.84], atmospheric light value is [0.8, 0.8, 0.8], and the corresponding clear pixel value is [0.05, 0.3, 1]. If error is 0.1, we can get two transmission 0.1 and 0.3, we can get the corresponding dehazing pixel value [−0.7, −0.2, 1.2] and [0.3, 0.47, 0.93], and we find that the error is large. However, the clear image detail does not have this problem; it can deal this problem well. In other words, the small error in transmission map when the transmission is less than a value(0.3) will boost the error in dehazing result. We can get a conclusion that when the transmission is small, the clear image detail is more suitable for dehazing.

Based on the above two observations, we find that the clear image detail and transmission map are correlated with each other. Further more, transmission map can serve as haze density map, which can make our model understand where and how much haze needs to remove. We use joint learning to solve the problem of only use transmission or clear image detail. By joint learning clear image detail and transmission map, the performance of estimation of clear image detail can be boosted, and the dehazing quality can be improved.

3.3 Formulation

The goal of dehazing is to remove haze from single hazy image. However, this goal is hard to achieve. In order to improve the performance of dehazing, we use a joint learning to train a deep convolution neural networks. Considering a single hazy image, we first get $\mathbf{I}_A(x)$ by Eq. (3), which is the only preprocessing we perform. Our goal is to recover from $\mathbf{I}_A(x)$ the initial clear image $F_{\text{detail}}(\mathbf{I}_A)$ and the corresponding transmission map $F_t(\mathbf{I}_A)$, which are as similar as possible to the truth clear image $\mathbf{J}_A$ and transmission map t. For the ease of presentation, we still call $\mathbf{I}_A(x)$ as hazy image, although it is $\mathbf{I}(x) - \mathbf{A}$.

In order to estimate clear image exactly, our goal can be expressed as follows:

$$\min ||F_{\text{detail}}(\mathbf{I}_A) - \mathbf{J}_A||^2, \tag{10}$$

$$\min ||F_t(\mathbf{I}_A) - t||^2, \tag{11}$$

$F(\mathbf{I}_A)$ represents the output of the network, and $\mathbf{J}_A$ represents the truth image detail.

In order to effectively solve the dehazing problem, we utilize a preprocessing to input hazy image. This preprocessing converts our problem from image color domain to image details domain. This preprocessing is inspired by classification using DCNN. Our DCNN method can benefit from this preprocessing, which makes our network easy to train.

It is worthy to point out that it is hard to train a network for predicting a haze-free image from hazy image directly. By training such a network, and testing the output, we find that the output cannot be regarded as haze-free image. So our method benefits much from learning $\mathbf{J}_A$ from $\mathbf{I}_A$, which makes our method remove haze from hazy possible.

3.4 Joint learning CNN

One good model is required for any CNN-based dehazing method. In this paper, we adopt the CNN architecture [22], which has been proved to work well on dehazing. However, we make some modifications to improve the performance of dehazing and add joint learning. First, we reduce the six layers to five layers, which can reduce the time of forward time. Second, we add transmission map to network. Our convolutional network architecture can be expressed as:

$$F^0(Y) = Y, \tag{12}$$

$$F^n(Y) = \max(W^n * F^{(n-1)}(Y) + B^n, 0), n = 1, \ldots, 4 \tag{13}$$

$$F_{\text{detail}}(Y) = W^n * F^{(n-1)}(Y) + B^n, n = 5$$

$$F_t(Y) = W^n * F^{(n-1)}(Y) + B^n, n = 5 \tag{14}$$

where $F_{\text{detail}}(Y)$ is the output of the networks, which represents the estimating result of clear image detail. $F_t(Y)$ is the output of the networks, which represents the transmission map of hazy image.

Context information is very important for dehazing. In order to use context information well, we use large kernel size (9×9) with 16 filters in the first layer of network. From the second to fourth layer, we use 16 filters with kernel size (3×3). For the fifth layer, 3 filters with kernel size (3×3) as output are used to predict the clear image detail and 1 filter with kernel size (3×3) to predict the hazy image transmission. Comparing with [22], our model has a large receptive field, which can help our model capture more contextual information. By adding a sibling layer to estimate transmission, it makes our model understand where and how much haze needs to be removed. As shown in Fig. 2, we adopt large kernel size (9×9) to increase the receptive field of our model. By adding a sliding layer to estimate the transmission map, it makes our model understand the density of haze well.

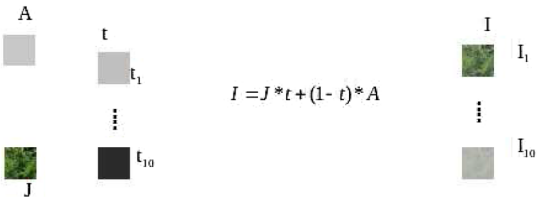

Fig. 3 Our method to prepare the hazy data

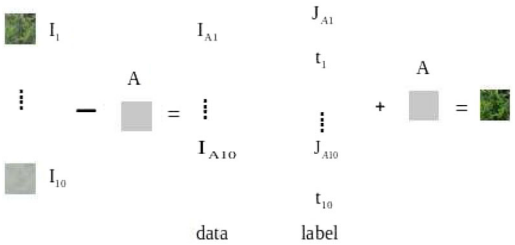

Fig. 4 Our method to prepare the training data

3.5 Training data preparation and network training

It is hard to collect the vast amount of labeled data for training dehazing CNN [21], so we adopt the same strategy as [16,19,21]. Based on the two assumptions [19]: first, the transmission in a local patch is constant. Second, image content has no relation with the transmission. As shown in Fig. 3, we crop an image patch from haze-free image and generate ten transmission values and atmospheric light values; then, we use Eq. (1) to get ten hazy image patches. As shown in Fig. 4, we can get a pair of $\mathbf{I}_A(x)$ and $\mathbf{J}_A(x)$ via Eqs. (3) and (5), and these data are used to train our network. As shown in Fig. 4, we model the haze removal problem as layer separation. Therefore, we can estimate $\mathbf{A}$ via a CNN network and then get a haze-free image.

In this paper, we train a model for each value of A. In order to consider more contextual information, we use patch size 32×32, which provides more contextual information. We implement our CNN architecture using Caffe package [28], which is a popular tool of deep learning. We use stochastic gradient descent (SGD) to train our model. We set learning rate to 0.0001 and batch size to 64.

It is worth nothing to point out that the commonly used constant assumption on the transmission in a local patch results in block artifacts [1]. Because we use the constant assumption on the transmission in a local patch to prepare the training data, our result also exists block artifacts, and we use non-local regularization and guided filter [29] to eliminate the block artifacts. Our model also can be extended to fully end-to-end dehazing via dehazing by using DCNN to remove block artifacts.

4 Dehazing using image detail and non-local regularization

In this section, we describe how to use image detail and global regularization to get a haze-free image. Firstly, we introduce how to use image detail to get a transmission map. Secondly, we introduce global regularization. We assume the $\mathbf{A}$ is known for all hazy images.

4.1 Convert image detail to transmission map

We find that recovering haze-free image using image detail directly often results color distortion and block artifacts. However, removal of artifacts and color distortion in color domain are very difficult, so we solve this problem using transmission map.

In order to use the output of CNN well, we convert image detail to transmission map. Based on Eq. (8), we can recover the transmission map easily. However, we find this result is not good for lost of global consistent. Next we will introduce how to incorporate the global regularization into dehazing.

4.2 Non-local regularization

It is obvious that two pixels with the same color in hazy image should have the same color in clear image. We can express this as:

$$I(x) = I(y) \Rightarrow F_t(I(x)) = F_t(I(y)), \tag{15}$$

where x and y represent the location in hazy image. F_t represents the predicted transmission of hazy image. However, the CNN cannot guarantee it well, so we introduce global regularization to provide a image-level consistency. In oder to implement global regularization efficiently, we use pixel voting. For pixels with value (r, g, b), we can collect all transmission values from transmission map and choose the transmission with most pixels voting as the final transmission. Then, we use guided filter to get a smooth transmission map.

4.3 Dehazing

Once the transmission map is estimated, we can recover the haze-free image by inversing Eq. (1):

$$J(x) = \frac{\mathbf{I}(x) - \mathbf{A}(x)}{t(x)} + \mathbf{A}(x) \tag{16}$$

In Fig. 5, it is an example of our method. From Fig. 5c, g, we can see that by using global regularization the result can get image-level consistence, but the result also lacks local

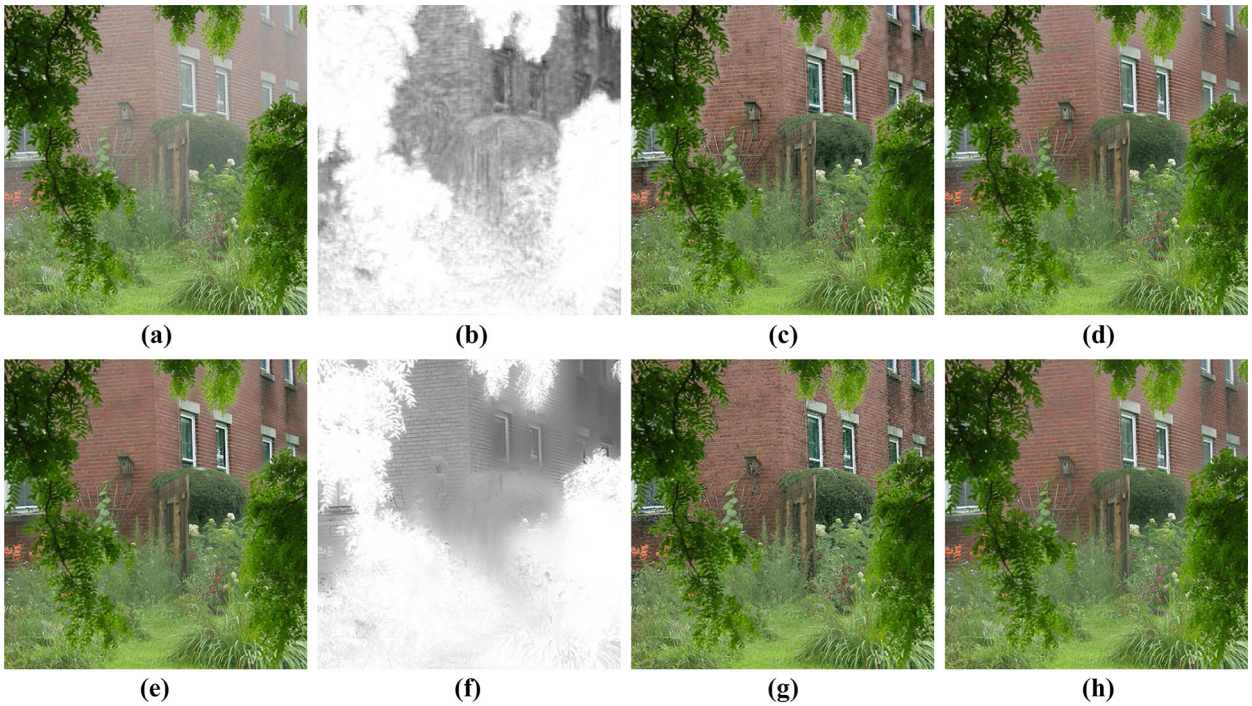

Fig. 5 Intermediate and the final results of our method: **a** An input hazy image; **b** the initial estimating transmission map; **c** the result without global regularization and guided filter; **d** the result with guided filter, but without global regularization; **e** our final result; **f** our final transmission map; **g** the result with global regularization, but without guided filter; **h** the result with global regularization and guided filter

smooth, so we need to use guided filter to get local smooth result. From Fig. 5d, h, the result in Fig. 5h is more haze-free around the leaves of tree.

5 Analysis and discussion

5.1 Effect of joint learning

In this section, we will analyze the effect of joint learning and show why they improve the performance of single image dehazing. In order to show the effect of joint learning, we train two models with single task network. The first one model is trained with detail, and the other is with transmission map. In order to show the performance of this model, we generate a hazy indoor images dataset using RGBD dataset [30,31]. This dataset is based on the indoor RGBD dataset [30], we use $\mathbf{A} = [0.78, 0.78, 0.78]$ and choose three values for β as 0.06, 0.3 and 0.54 to generate hazy image [20], and this dataset contains 4347 hazy images. Our hazy dataset is accompanied with ground truth haze-free images, which enable us to evaluate models using $PSNR$ and $SSIM$, and we also use this dataset to evaluate the performance of the state-of-the-art methods in 6.1. All models are trained with same hyper parameters, the learning rate is 0.0001, the momentum is 0.9, the weight decay is 0.0005, and the convolution layer is initialed with "xavier." As shown in Table 1, we can see that the model trained with image detail is better than the one trained with transmission map, and the model trained with joint learning is the best.

Table 1 Comparison with different models

Type	Transmission	Image detail	Joint learning
PSNR	21.02	**21.1**	*21.65*
SSIM	0.828	**0.834**	*0.865*

Italic indicates the best performance, and bold indicates the second best performance

5.2 Effect of non-local regularization

In this subsection, we analyze the effect of non-local regularization via comparison between with and without non-local regularizations. As shown in Table 2, we find non-local regularization improves the dehazing quality slightly. We also perform the experiment on indoor dataset. However, we find that the regularization cannot improve the dehazing and even results in degradation of the performance of our method. For this phenomenon, one possible reason is that indoor image has more local artifact which can be deal effectively via

Table 2 Fattal's dataset

Type	Without	With
Outdoor	**5.17**	*5.15*

Italic indicates the best performance and bold indicates the second best performance

guided filter, and our non-local regularization considers more of image-level information.

6 Experiment results

In this section, we evaluate our method on a large dataset containing both synthetic and natural images and compare our performance to state-of-the-art methods. In this section, we use the $L1err = \frac{1}{N}\sum(\sum_{c\in R,G,B}|\mathbf{J}^c - \mathbf{G}^c|)$ as metric, where $\mathbf{J}$ represents the dehazing result image and $\mathbf{G}$ represents the ground truth image. We convert the dehazing result and ground truth image into range of [0, 1].

6.1 Synthetic hazy images

In this subsection, we compare our method with state-of-the-art methods on both indoor and outdoor synthetic hazy images. An outdoor synthetic hazy images dataset was introduced by [32], which is available online. This dataset contains eleven haze-free images and corresponding simulated hazy images. These images (its intensity scaled to [0, 1]) were added an identically distributed zero-mean Gaussian noise with three different noise levels: $\sigma = 0.01, 0.025, 0.05$. Evaluating the dehazing on indoor hazy image is very popular, and we generate an indoor hazy dataset with fixed A.

6.1.1 Outdoor hazy image

An outdoor synthetic hazy images dataset was introduced by [32], which is available online. This dataset contains eleven haze-free images and corresponding simulated hazy images. These images (its intensity scaled to [0, 1]) were added an identically distributed zero-mean Gaussian noise with three different noise levels: $\sigma = 0.01, 0.025, 0.05$. For this dataset, we sum all L1 errors as quality standard. We compare our method with some state-of-art methods [1,32,33] on all images in dataset and list all results in Tables 4, 5, 6, 7, 8 and 9. As shown in Table 3, we can see that our results are the best in general. For each image , we also get very high performance. We compare our method with [32] and [1], and we can see that He et al.'s result exits a lot of color distortion, and Fattl's result generates a more pleased than He et al.'s. Our method can generate a high-quality result which is similar to ground truth haze-free image (Fig. 6).

Table 3 Fattal's dataset

Type	He	Fattal	Berman	Zhang	Ours
Outdoor	8.40	6.67	8.81	**5.75**	*5.17*

Italic indicates the best performance and bold indicates the second best performance

Table 4 Church in Fattal's dataset with sky pixels

Image	He	Fattal	Berman	Zhang	Ours
D1	0.215	0.208	0.448	**0.203**	*0.200*
D2	0.184	0.184	0.186	**0.178**	*0.151*
D3	0.174	**0.133**	0.265	0.135	*0.102*
S10	0.215	0.240	0.200	**0.198**	*0.177*
S25	**0.250**	0.257	0.361	0.262	*0.232*
S50	0.340	0.592	0.361	**0.334**	*0.325*
Church	0.184	**0.173**	0.186	0.178	*0.151*

Italic indicates the best performance, and bold indicates the second best performance

Table 5 Lawn1 in Fattal's dataset with sky pixels

Image	He	Fattal	Berman	Zhang	Ours
D1	0.205	**0.124**	0.411	*0.121*	0.157
D2	0.193	0.116	**0.085**	0.086	*0.081*
D3	0.207	*0.072*	0.150	0.099	**0.090**
S10	0.203	0.136	*0.105*	**0.111**	*0.105*
S25	0.239	0.230	0.180	**0.172**	*0.169*
S50	0.317	0.397	0.334	*0.277*	**0.278**
Lawn1	0.193	0.119	0.100	**0.086**	*0.081*

Italic indicates the best performance, and bold indicates the second best performance

Table 6 Mansion in Fattal's dataset with sky pixels

Image	He	Fattal	Berman	Zhang	Ours
D1	0.1392	**0.100**	0.409	0.109	*0.090*
D2	0.129	**0.076**	0.149	0.079	*0.054*
D3	0.134	*0.053*	0.301	0.065	**0.059**
S10	0.121	0.095	0.171	**0.090**	*0.070*
S25	0.134	0.157	0.219	**0.127**	*0.113*
S50	0.225	0.240	0.289	**0.200**	*0.193*
Mansion	0.129	**0.072**	0.149	0.079	*0.054*

Italic indicates the best performance, and bold indicates the second best performance

6.1.2 Result without sky pixels

The $L1$ errors on non-sky pixels of dehazing images have been used in [32–34], which can exclude errors of sky in dark channel prior. We use same metric to compare our method with [1,32–34] and show the result in Tables 10 11 12 13

Table 7 Raindeer in Fattal's dataset with sky pixels

Image	He	Fattal	Berman	Zhang	Ours
D1	0.210	*0.130*	0.477	0.206	**0.160**
D2	0.205	0.105	0.134	**0.08**	*0.071*
D3	0.202	**0.071**	0.233	*0.057*	0.074
S10	0.197	0.126	0.146	**0.090**	*0.083*
S25	0.201	0.161	0.190	**0.120**	*0.119*
S50	0.256	0.250	0.278	*0.185*	**0.192**
Raindeer	0.205	0.101	0.137	**0.081**	*0.071*

Italic indicates the best performance, and bold indicates the second best performance

Table 8 Road1 in Fattal's dataset with sky pixels

Image	He	Fattal	Berman	Zhang	Ours
D1	0.176	*0.115*	0.442	0.157	**0.134**
D2	0.145	**0.099**	0.117	0.122	*0.078*
D3	0.133	**0.080**	0.220	**0.080**	*0.074*
S10	0.166	**0.116**	0.133	0.137	*0.100*
S25	0.209	0.198	0.185	**0.180**	*0.155*
S50	0.299	0.346	0.286	**0.266**	*0.257*
Road1	0.145	**0.098**	0.117	0.122	*0.078*

Italic indicates the best performance, and bold indicates the second best performance

Table 9 Result of left images in Fattal's dataset with sky pixels

image	He	Fattal	Berman	Zhang	Ours
Couch	0.123	0.160	*0.092*	**0.111**	0.118
Flower1	0.294	*0.056*	**0.067**	0.107	0.123
Flower2	0.283	*0.045*	0.136	0.096	**0.094**
Lawn2	0.203	0.117	0.100	**0.099**	*0.085*
Moebius	0.439	0.227	0.242	**0.138**	*0.091*
Road2	0.177	0.131	**0.120**	0.122	*0.084*

Italic indicates the best performance, and bold indicates the second best performance

Table 10 Road1 in Fattal's dataset without sky pixels

Image	He	Fattal	Berman	Bui	Ours
D1	0.064	**0.041**	0.136	*0.039*	0.046
D2	0.052	0.034	**0.031**	0.038	*0.026*
D3	0.046	*0.026*	0.029	0.031	**0.027**
S10	0.065	**0.053**	0.058	0.056	*0.033*
S25	0.082	0.079	0.075	**0.071**	*0.051*
S50	0.094	0.102	0.097	*0.082*	**0.085**

Italic indicates the best performance and bold indicates the second best performance

Table 11 Lawn1 in Fattal's dataset without sky pixels

Image	He	Fattal	Berman	Bui	Ours
D1	0.070	0.052	0.119	**0.048**	*0.047*
D2	0.063	0.042	**0.031**	0.046	*0.026*
D3	0.066	**0.037**	0.049	0.043	*0.031*
S10	0.051	0.041	**0.040**	0.052	*0.033*
S25	0.071	0.068	**0.063**	0.071	*0.052*
S50	0.093	0.105	0.095	*0.085*	**0.087**

Italic indicates the best performance, and bold indicates the second best performance

and 14. We can see that our method outperforms previous state-of-the-art methods in most cases and handles noise well. We also find that for indoor images Mansion and Raindeer, we can see our method gets the highest performance in Tables 13 and 12. For the left outdoor images, we find that our method performs well in most cases. *DCP* cannot perform as well as other methods [32–34] and our method. However, at high haze levels, the DCP is much better than [33] and almost comparable to [32,34] and our method. Even our model is not trained noise, our method is robust to noise and generate higher quality results than [32–34].

In this subsection, we compare our method with state-of-the-art methods on Fattal's dataset. We show the error with and without sky. From the result, we can see that our method is robust and can generate high-quality dehazing result.

Fig. 6 Comparison on outdoor hazy images

Table 12 Mansion in Fattal's dataset without sky pixels

Image	He	Fattal	Berman	Bui	Ours
D1	0.047	0.043	0.107	**0.042**	*0.031*
D2	0.043	0.033	0.050	**0.037**	*0.018*
D3	0.044	**0.026**	0.042	0.032	*0.020*
S10	0.048	**0.036**	0.056	0.046	*0.024*
S25	**0.061**	0.067	0.082	0.064	*0.037*
S50	**0.077**	0.085	0.097	0.079	*0.064*

Italic indicates the best performance and bold indicates the second best performance

Table 13 Raindeer in Fattal's dataset without sky pixels

Image	He	Fattal	Berman	Bui	Ours
D1	0.070	0.063	0.293	**0.059**	*0.053*
D2	0.068	0.053	0.057	**0.052**	*0.024*
D3	0.067	**0.044**	0.068	0.047	*0.025*
S10	0.063	0.049	0.053	**0.044**	*0.028*
S25	0.068	0.062	0.070	**0.058**	*0.040*
S50	0.083	0.094	0.096	**0.082**	*0.064*

Italic indicates the best performance and bold indicates the second best performance

Table 14 Church in Fattal's dataset without sky pixels

Image	He	Fattal	Berman	Bui	Ours
D1	0.047	0.039	0.146	*0.037*	**0.039**
D2	0.048	0.034	*0.021*	0.035	**0.027**
D3	0.061	**0.029**	0.049	0.032	*0.025*
S10	0.065	0.058	0.052	**0.047**	*0.034*
S25	0.078	0.086	0.064	**0.058**	*0.052*
S50	0.100	0.113	0.091	*0.076*	**0.087**

Italic indicates the best performance, and bold indicates the second best performance

6.1.3 Indoor hazy image

Our method is based on $\mathbf{A}$ which has been given, so we set $\mathbf{A} = [0.78, 0.78, 0.78]$ for methods [20,21,27,33]. Table 15 shows the average $PSNR$ and $SSIM$ results on our dataset, respectively. Since our model is trained from end to end under the MSE loss, it is not amazing to see our method has higher $PSNR$ performance than others. More appealing is the observation that our model gets even higher $SSIM$ performance over all competitors, even though $SSIM$ is not an optimization criterion in our training. When compared with [26], our method also gets higher performance with $PSNR$ and $SSIM$ (Table 16).

Table 16 Quantitative comparison on the HIFA dataset w/o the ground truth atmospheric lights A

Type	Li [26]	Ours
PSNR	*18.64*	17.22
SSIM	0.747	**0.778**

Italic indicates the best performance and bold indicates the second best performance

6.2 Quantitative evaluation on benchmark natural images dataset

In this subsection, we evaluate the dehazing methods on a few natural image examples that are significantly more challenging than general outdoor images. In Fig. 7, we compare our method with state-of-the-art methods[1,20–22,26,27,32,33]. Some of the results are provided by Fattal [32] , Berman [33] and Cai [21], which are online. We also get some results via the program provided by Ren [20]. As shown in Fig. 7, we compare all methods in three hazy images. For "waterfall" image, we find [20,21,26] cannot remove haze completely and [27] loses some color detail. In contrast, our result and [22] are more colorful and recover more detail. However, Zhang's result [22] loses some detail in red box. For "canyon" image, we find He [1,27] exits the intolerable artifacts. CAP, MSCNN and Dehazenet alleviate the problem. Li[26] alleviates this problem furthermore; however, we also notice some artifacts in green box. For "waterfall" image, we find that He [1] and Zhu [16] cannot remove haze in red box well. Dehazenet, color line [32] and haze line [33] lose some color in blue box. For "cityscape" image, [19,21,27,35] can yield an excellent result in general, but lack some micro-contrast details when compared to [32] and ours. This is obvious in the zoomed-in buildings shown in Cityscape results, where in our result and [32] the windows are more clear than in other methods. In contrast, our method can deal these situations very well. In general, our method can recover more image details and vivid color than other state-of-the-art methods.

Table 15 Quantitative comparison on the indoor dataset with the ground truth atmospheric lights A

	BCCR [27]	NLD [33]	Dehazenet [21]	MSCNN [20]	AOD [26]	Ours
PSNR	18.70	17.38	**19.06**	18.64	18.64	*21.65*
SSIM	0.794	0.764	0.789	**0.800**	0.747	*0.865*

Italic indicates the best performance and bold indicates the second best performance

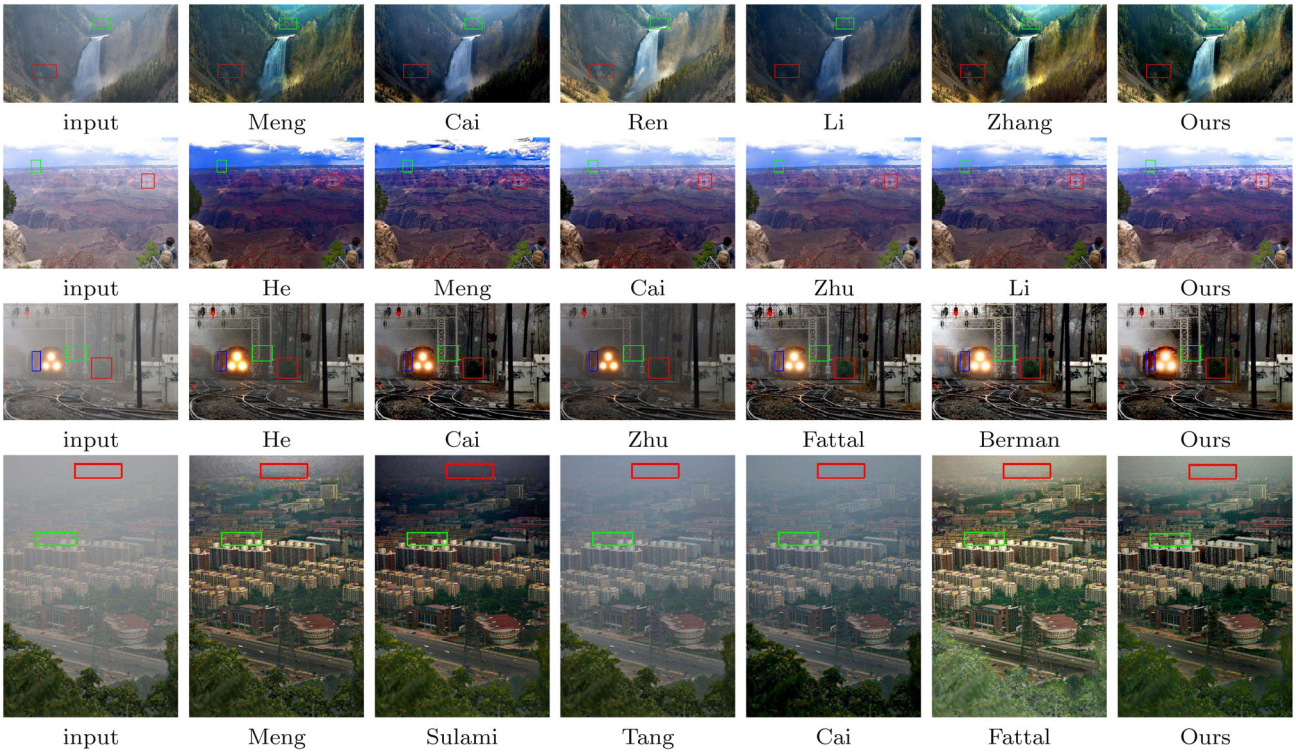

Fig. 7 Comparison on natural images: [Left] Input images. [Right] Our result. Middle columns display results by several methods, since each paper reports results on a different set of images

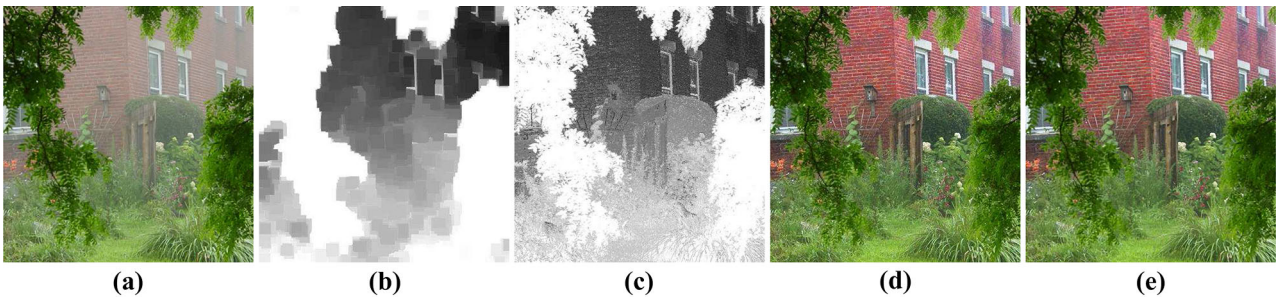

Fig. 8 An example of our global regularization

6.3 An example of non-local regularization

In this section, we provide an example of global regularization, which can get a high-quality result without guided filter, and which is shown in Fig 8. Figure 8 shows our new regularization can recover a lot of detail from the rough transmission map and recover a high-quality result. However, we also can find some artifacts from the result, and we remove it by local smooth operation. We find that non-local regularization improves the dehazing quality by keeping the similar in result as in hazy image.

6.4 Limitation

Our method assumes the atmospheric light has been accurately estimated and trained on outdoor images, and we find that our method could not perform well on indoor hazy image without the ground truth atmospheric light. We show the comparison result in Table 17, and we can see that our method also performs well than Cai et al.'s. It worth nothing to point out that our model trained with outdoor images which are more complex and have more depth jumps. Li et al.'s and Ren et al.'s models are trained with indoor images which are similar to our generated test dataset. This is the main reason for these two methods to get high performance with $PSNR$ metric.

Table 17 Quantitative comparison on the HIFA dataset w/o the ground truth atmospheric lights A

	Cai [21]	Ren [20]	Li [26]	Ours
PSNR	16.81	**17.68**	*18.64*	17.22
SSIM	0.745	*0.788*	0.747	**0.778**

Italic indicates the best performance and bold indicates the second best performance

7 Conclusions

In this paper, we address the haze removal problem using a deep network which improves the dehazing results via joint learning of image detail and transmission map. The proposed learning-based method learns more effective feature for dehazing via joint solve the image detail and transmission map. Our method can get a haze-free image directly from hazy image. We also propose a new contextual regularization, which can be used to get more image-level consistent. Experimental results on synthetic dataset prove that joint learning two tasks, which can be converted from each other via simple formula, can be used to boost the haze removal quality. We also show that CNN cannot keep some priors well. For example, the pixels with same value in hazy image have same transmission. Experimental results on synthetic outdoor, indoor and real images show the effect of the proposed algorithm. Extensive experimental comparisons show our method yields higher quantitative and qualitative performance than existing state-of-the-art methods. In future work, we will extend the proposed methods to medical image process [36,37].

Acknowledgements This study was funded by National Natural Science Foundation of China (Grant Numbers 61472289 and 41571436).

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

1. He, K., Sun, J., Tang, X.: Single image haze removal using dark channel prior. IEEE Trans. Pattern Anal. Mach. Intell. **33**(12), 2341 (2011)
2. 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 (2018)
3. Li, K., He, F., Yu, H., Chen, X.: A correlative classifiers approach based on particle filter and sample set for tracking occluded target. Appl. Math. A J. Chin. Univ. **32**(3), 294 (2017)
4. Chen, X., He, F., Yu, H.: A matting method based on full feature coverage. Multimed. Tools Appl. (2018). https://doi.org/10.1007/s11042-018-6690-1
5. Li, K., He, F., Yu, H., Chen, X.: A parallel and robust object tracking approach synthesizing adaptive Bayesian learning and improved incremental subspace learning. Front. Comput. Sci. (2018). https://doi.org/10.1007/s11704-018-6442-4
6. Zhang, D., He, F., Han, S., Zou, L., Wu, Y., Chen, Y.: An efficient approach to directly compute the exact Hausdorff distance for 3D point sets. Integr. Comput. Aided Eng. **24**(3), 261 (2017)
7. Chen, Y., He, F., Wu, Y., Hou, N.: A local start search algorithm to compute exact Hausdorff distance for arbitrary point sets. Pattern Recognit. **67**, 139 (2017)
8. Fan, X., Wang, Y., Gao, R., Luo, Z.: Haze editing with natural transmission. Vis. Comput. **32**(1), 137 (2016)
9. Harald, K.: Theorie der horizontalen Sichtweite: Kontrast und Sichtweite. Keim & Nemnich, Munich (1924)
10. Tan, R.T.: Visibility in bad weather from a single image. In: IEEE Conference on Computer Vision and Pattern Recognition. IEEE, pp. 1–8 (2008)
11. Tarel, J.P., Hautiere, N.: Fast visibility restoration from a single color or gray level image. In: IEEE international conference on computer vision. IEEE, pp. 2201–2208 (2009)
12. Fattal, R.: Single image dehazing. ACM Trans. Graph. (TOG) **27**(3), 1 (2008)
13. Kratz, L., Nishino, K.: Factorizing scene albedo and depth from a single foggy image. IEEE Int. Conf. Comput. Vis. **30**(2), 1701 (2009)
14. Nishino, K., Kratz, L., Lombardi, S.: Bayesian defogging. Int. J. Comput. Vis. **98**(3), 263 (2012)
15. Gibson, K.B., Nguyen, T.Q.: An analysis of single image defogging methods using a color ellipsoid framework. Eur. J. Image Video Process. **2013**(4), 1 (2013)
16. Zhu, Q., Mai, J., Shao, L.: A fast single image haze removal algorithm using color attenuation prior. IEEE Trans. Image Process. **24**(11), 3522 (2015)
17. Li, Z., Zheng, J.: Edge-Preserving decomposition-based single image haze removal. IEEE Trans. Image Process. **24**(12), 5432 (2015)
18. Khmag, A., Al-Haddad, S., Ramli, A.R., Kalantar, B.: Single image dehazing using second-generation wavelet transforms and the mean vector L2-norm. Vis. Comput. **34**(5), 675 (2018)
19. Tang, K., Yang, J., Wang, J.: Investigating haze-relevant features in a learning framework for image dehazing. In: IEEE conference on computer vision and pattern recognition (CVPR). IEEE, pp. 2995–3002 (2014)
20. 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. Springer, pp. 154–169 (2016)
21. Cai, B., Xu, X., Jia, K., Qing, C., Tao, D.: DehazeNet: An end-to-end system for single image haze removal. arXiv preprint arXiv:1601.07661 (2016)
22. Zhang, S., He, F., Yao, J.: Single image dehazing using deep convolution neural networks. In: Pacific Rim conference on multimedia. Springer, pp. 315–325 (2017)
23. Xiao, C., Gan, J.: Fast image dehazing using guided joint bilateral filter. Vis. Comput. **28**(6–8), 713 (2012)
24. Ling, Z., Li, S., Wang, Y., Shen, H., Lu, X.: Adaptive transmission compensation via human visual system for efficient single image dehazing. Vis. Comput. **32**(5), 653 (2016)
25. Ju, M., Zhang, D., Wang, X.: Single image dehazing via an improved atmospheric scattering model. Vis. Comput. **33**(12), 1613 (2018)

26. Li, B., Peng, X., Wang, Z., Xu, J., Feng, D.: Aod-net: All-in-one dehazing network. IEEE International Conference on Computer Vision. **1**(4), 4770–4778 (2017)
27. Meng, G., Wang, Y., Duan, J., Xiang, S., Pan, C.: Efficient image dehazing with boundary constraint and contextual regularization. In: IEEE international conference on computer vision, pp. 617–624 (2013)
28. Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S., Darrell, T.: Caffe: convolutional architecture for fast feature embedding. In: ACM international conference on multimedia. ACM, pp. 675–678 (2014)
29. He, K., Sun, J., Tang, X.: Guided image filtering. IEEE Trans. Pattern Anal. Mach. Intell. **35**(6), 1397 (2013)
30. Silberman, N., Hoiem, D., Kohli, P., Fergus, R.: Indoor segmentation and support inference from RGBD images. In: European conference on computer vision. Springer, pp. 746–760 (2012)
31. Li, W., Saeedi, S., McCormac, J., Clark, R., Tzoumanikas, D., Ye, Q., Huang, Y., Tang, R., Leutenegger, S.: InteriorNet: mega-scale multi-sensor photo-realistic indoor scenes dataset. In: British machine vision conference (2018)
32. Fattal, R.: Dehazing using color-lines. ACM Trans. Graph. (TOG) **34**(1), 13 (2014)
33. Berman, D., treibitz, T., Avidan, S.: Non-local image dehazing. In: IEEE conference on computer vision and pattern recognition (CVPR), pp. 1674–1682 (2016)
34. Bui, T.M., Kim, W.: Single image dehazing using color ellipsoid prior. IEEE Trans. Image Process. **27**(2), 999 (2018)
35. 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)
36. Yu, H., He, F., Pan, Y.: A novel segmentation model for medical images with intensity inhomogeneity based on adaptive perturbation. multimedia tools and applications (2018). https://doi.org/10.1007/s11042-018-6735-5
37. Yu, H., He, F., Pan, Y.: A novel region-based active contour model via local patch similarity measure for image segmentation. Multimed. Tools Appl. **77**(18), 24097 (2018)

Shengdong Zhang A Ph D student in Whu, main study in image dehazing.

Fazhi He Professor, PhD Supervisor, Wuhan University State Key Lab of Software Engineering, School of computer science, P.R. China

Wenqi Ren Assistant Professor in Institute of Information Engineering, Chinese Academy of Sciences.

Jian Yao Professor, PhD Supervisor.