J. Vis. Commun. Image R. 72 (2020) 102897

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

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

Full Length Article

# Weakly supervised single image dehazing

Cong Wang [a,1], Wanshu Fan [b,a,1,*], Yutong Wu [a], Zhixun Su [a]

[a] School of Mathematical Sciences, Dalian University of Technology, No. 2 Linggong Road, Ganjingzi District, Dalian 116000, China
[b] School of software Engineering, Dalian University, NO.10 University Avenue, Economic and Technological Development Area, Dalian 116622, China

ARTICLE INFO

*Keywords:*
Image dehazing
Weakly supervised
Convolutional neural network (CNN)
Multi-level multi-scale block

ABSTRACT

Single image dehazing is a critical image pre-processing step for many practical vision systems. Most existing dehazing methods solve this problem utilizing various of hand-crafted priors or by supervised training on the synthetic hazy image information (such as haze-free image, transmission map and atmospheric light). However, the assumptions on the hand-crafted priors are easily violated and collecting realistic transmission map and atmospheric light are unpractical. In this paper, we propose a novel weakly supervised network based on the multi-level multi-scale block. The proposed network reduces the constraint on the training data and automatically estimates the transmission map and the atmospheric light as well as the intermediate haze-free image without using any realistic transmission map and atmospheric light as supervision. Moreover, the estimated intermediate haze-free image helps to generate accurate transmission map and atmospheric light by embedding the physical-model, which presents reliable restoration of the final haze-free image. In particular, our network also can be trained on the real-world dataset to fine-tune the model and the fine-tuning operation improves the dehazing performance on the real-world dataset. Quantitative and qualitative experimental results demonstrate the proposed method performs on par with the supervised methods.

## 1. Introduction

Images and videos taken under severe hazy conditions are usually degraded in visibility and contrast when the light diffuses into the atmosphere. These image quality degradations in turn may jeopardize the performance of many computer vision systems, especially for some scene understanding and recognition tasks, such as traffic detection and environmental monitoring. Numerous image and video-based dehazing algorithms [1–9] have been developed to solve this problem, as a challenging instance of image restoration and enhancement.

Image dehazing aims to recover a clear image from a hazy image which is caused by haze, fog or smoke. The hazing process can be modeled as

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

where $I(x)$ and $J(x)$ denote the hazy image and the haze-free image, respectively. $A$ is the global atmospheric light indicating the intensity of the ambient light, $t(x)$ is the transmission map describing the portion of light and $x$ is the pixel index. Assuming that the haze is homogeneous, the transmission map can be estimated by $t(x) = e^{-\beta d(x)}$, where $\beta$ is the notations of scattering coefficient and $d(x)$ is the scene depth.

As only $I(x)$ is available, recovering $J(x)$, $t(x)$ and $A$ from a hazy image $I(x)$ is a well-known ill-posed problem as multiple solutions to satisfy Eq. (1). As can be seen from Eq. (1), we note that there remains two major issues in the dehazing process, one is accurately estimate the transmission map and the other one is accurately estimate the atmospheric light. When the transmission map and the atmospheric light are obtained, the final haze-free image can be computed by

$$J(x) = \frac{I(x) - \tilde{A}(1 - \tilde{t}(x))}{\tilde{t}(x)}, \tag{2}$$

where $\tilde{t}(x)$ and $\tilde{A}$ denote the estimated transmission map and the estimated atmospheric light, respectively.

To make image dehazing well-posed, most early methods focus on utilizing hand-crafted priors to estimate transmission map and atmospheric light in a separate manner [1,10–13]. The commonly used hand-crafted priors try to use some visual cues to capture deterministic and statistical properties of hazy images, such as dark channel prior [14], color-line prior [15] and haze-line prior [16]. Nevertheless, these methods fail to generate visually pleasing results for some cases since the assumptions on the priors do not always hold.

Recently, the amazing success of deep learning based methods [2,5,6,9,17–22] provide new ideas for image dehazing. Ren et al. [5] and Cai et al. [6] develop the end-to-end convolution neural networks to estimate the transmission map and then follow the traditional dehazing

* Corresponding author.
*E-mail address:* fan921amber@163.com (W. Fan).
[1] The first two authors contributed equally to this work.

https://doi.org/10.1016/j.jvcir.2020.102897
Received 17 February 2020; Received in revised form 17 July 2020; Accepted 29 August 2020
Available online 7 September 2020

method to estimate the atmospheric light for recovering the haze-free image. These two methods perform well and neglect the atmospheric light or haze-free image. Instead of estimating the transmission map and atmospheric light separately, Li et al. [8] reformulate the physical scattering model and propose an all-in-one dehazing network (AOD-net) by integrating the transmission map and the atmospheric light into a single variable. All the aforementioned methods require pairs of hazy and corresponding ground-truth images for training. Different form these methods, Engin et al. [21] and Dudhane and Murala [22] utilize unpaired training methods for image dehazing. Although these CNN-based methods have achieved better performance than the prior based ones, they are restricted in practical application because the model training requires a large amount of data for supervision.

Most existing dehazing methods almost depend on more synthetic information for estimating the atmospheric light $t(x)$ and the transmission atmospheric light $A$, which uses the haze-free image, the artificial atmospheric light and the artificial transmission map as their ground-truth information for supervision. But the realistic transmission map and atmospheric light are more complex than the artificial transmission map and atmospheric light. Moreover, the artificial atmospheric light and transmission map do not meet the current unsupervised or weakly supervised trends, while a better solution, which can generate the atmospheric light and the transmission map without their supervised information, is needed. In this paper, we propose a novel weakly supervised dehazing network, which trains on the pairs of hazy and the corresponding haze-free image without any supervised information about the atmospheric light and the transmission map. Instead of feeding the network with a large amount of training data, the proposed network only uses the ground-truth of haze-free images for supervision. The atmospheric light and transmission map networks utilize loss function Eq. (7) to estimate the transmission map and the atmospheric light in an automatic learning manner. Moreover, our network also can be trained on the real-world dataset as the semi-supervision to fine-tune the model and the fine-tuning operation improves the dehazing performance on the real-world dataset. Fig. 1 is a sample dehazing image using the proposed method. It provides a new viewpoint for future unsupervised haze removal research.

The contributions of this work are as follows:

- We propose an end-to-end trainable network based on the multi-level multi-scale block to solve image dehazing problem.
- The proposed network reduces the constraint on the training data and simultaneously estimates the transmission map and the atmospheric light by automatic learning. Our network is trained in a weakly supervised manner and generates reliable restoration of the haze-free image by embedding the physical-model.
- Our network can be trained on the real-world dataset as the semi-supervision to fine-tune the model and the fine-tuning operation improves the dehazing performance on the real-world dataset.
- Quantitative and qualitative experiments on both synthesized datasets and real-world hazy images demonstrate the proposed method performs favorably against the state-of-the-art dehazing methods.

## 2. Related work

In this section, we first briefly review several relevant dehazing methods and then we introduce the multi-scale learning manner.

### 2.1. Image dehazing

Most of the existing single image dehazing methods can be roughly divided into two categories: the prior based methods and the deep learning based methods.

**Prior based methods:** Various image priors have been developed for single image dehazing, these methods mainly rely on making assumptions on atmospheric light, transmission map, or clear images [4, 14–16]. Based on these priors, they are to extract haze-relevant features. Tan [10] propose a patch-based contrast-maximization method for image dehazing inspired by the observation that the haze-free images must have higher contrast, but this method generates an over-saturated recovered image. To address this problem, He et al. [14] propose a dark channel prior (DCP) to estimate the transmission map in general cases, which assumes that the local minimum of the dark channel in a haze-free image is close to zero. Zhu et al. [4] create a linear model for formulating the scene depth of the hazy image based on the hand-crafted features. The depth information plays a critical role in many vision problems [23–25]. Recently, some color-based priors have been proposed as well for boosting the dehazing performance [15, 16]. In [15], a color-line prior is explored to characterize the haze-free image based on the observation that the small image patches may exhibit the 1D distribution of pixels in RGB color space. Similarly, Berman et al. [16] develop a haze line prior to describe the hazy image. These color-based methods fail to recover hazy images with a dense haze. Though all of the above priors are powerful and present strong effectiveness in helping haze removal, these priors are not robust to the unconstraint environment because they are designed under the observation of specific image properties.

**Deep learning based methods:** Different form the prior based methods, deep learning based methods directly estimate the transmission map or the atmospheric light, which have been made remarkable progress in image dehazing problem. Cai et al. [6] introduce an end-to-end convolutional neural network (CNN) to estimate the transmission map with a novel BReLU layer. In [5], a coarse-scale network is designed to learn the mapping between the hazy images and their corresponding transmission maps directly, and then a fine-scale network is used to refine the transmission map. These two methods perform well but they limit their capabilities by only considering the transmission maps in their CNN frameworks. To overcome this problem, In [8], Li et al. develop an all-in-one dehazing network which needs to compute an intermediate variable integrating both atmospheric light and transmission map. More recently, Zhang and Patel [26] present a densely connected pyramid dehazing Network to jointly learn the transmission map, atmospheric light and dehazed image all together, and then a joint discriminator-based GAN [27] is utilized to refine the estimated transmission map and the dehazed image. Li et al. [28] propose a conditional generative adversarial network to solve the image dehazing problem. Ren et al. [9] use an encoder–decoder network and adopt a novel fusion-based strategy, the dehazed image is produced as a fusion of the contrast enhanced, white balance and gamma corrected image, and then the final haze-free image is generated by gating these important features of the derived inputs. Qu et al. [29] consider the image dehazing problem as an image-to-image translation problem, and develop an Enhanced Pix2pix Dehazing Network, which generates the haze-free image without relying on the physical scattering model. Chen et al. [30] propose a novel method based on a new feature-patch map for image dehazing, which can adaptively select the patch size for each pixel.

### 2.2. Multi-scale learning manner

The multi-scale strategy has been successfully applied in many low- and high-level computer vision tasks and achieves better performance, such as object detection [31], pose estimation [32], optical flow estimation [33], scene parsing [34] and visual recognition [35]. The pooling operation is often used to obtain multi-scale feature information. Recently, some researchers combine the properties of both the traditional spatial pyramid architectures as well as the convolutional neural network to deal with various practical vision tasks. Ranjan and Black [33] develop a spatial pyramid network for the optical flow estimation, which can estimate the large motions in a coarse-to-fine strategy by warping one image of a pair at each pyramid level. In [31], Lin et al. propose a feature pyramid network for object detection, they exploit

(a) (b) (c)

**Fig. 1.** A sample image dehazed result using the proposed method. (a) Hazy input. (b) Transmission map estimation result. (c) Our result.

the inherent multi-scale, pyramidal hierarchy of deep convolutional networks for constructing feature pyramids with marginal extra cost. A top-down architecture with lateral connections is designed to build high-level semantic feature maps at all scales. Chen et al. [32] propose a cascaded pyramid network for multi-person pose estimation, which can successfully localize the key points (such as eyes and hands). More recently, Zhang and Patel [26] develop a densely connected pyramid network for single image dehazing, they adopt a multi-level pyramid pooling module to refine the learned features by considering the global structural information into the optimization [36]. Based on the above analysis, many works verify that the multi-scale strategy is effective for various computer vision tasks and it is necessary to introduce the multi-scale learning manner to handle image dehazing problem.

## 3. The proposed network

This section presents the details of the proposed network, we define this network as Multi-Level Multi-Scale Dehazing Network. The proposed network can automatically estimates the transmission map, the atmospheric light and the intermediate haze-free image all together by only using the ground-truth of haze-free images for supervision. With the estimated intermediate haze-free image, the transmission map and the atmospheric light can be estimated accurately via the physical model Eq. (1). The architecture of the proposed network is illustrated in Fig. 2.

### 3.1. Overall framework

The proposed network architecture consists of the following three modules: (1) The intermediate haze-free image estimation net, (2) The transmission map estimation net, and (3) The atmosphere light estimation net. The final haze-free is obtained by the estimated transmission map $\tilde{t}(x)$ and estimated atmospheric light $\tilde{A}$ and the input hazy image by solving Eq. (2). we exploit the encoder–decoder networks with multi-level multi-scale information aggregation, which are designed based on the U-Net [37] framework. Skip connections between encoder and decoder are applied to enable the computation of long-range spatial dependencies as well as efficient usage of the feature activation of proceeding layers. In the following, we explain these modules in details.

### 3.2. Multi-Level Multi-Scale Block (MLMSB)

The multi-scale architecture has been successfully applied into many computer vision tasks [26,31–33] and achieves better performance. Although the multi-scale learning manner performs well in these domains, the multi-level is still ignored. In order to generate the perceptually pleasing haze-free image, we propose a multi-level multi-scale block (MLMSB), as shown in Fig. 3. The proposed MLMSB not only obtains the multi-scale information via several pooling operations, but also learns the features with different levels, i.e., multi-level information. Benefiting from designing the MLMSB, the proposed network can estimate the transmission map, the atmospheric light and the intermediate haze-free image, and generate the final haze-free image via Eq. (2). The multi-scale block (MSB) can be expressed as follows:

Firstly, we utilize *Pooling* operation with different size of kernels and strides to obtain the multi-scale features:

$$y_i = Pooling_i(x), i = 1, 2, 4, 8, \tag{3}$$

where $Pooling_i$ denotes *Pooling* operation with $i \times i$ kernel and stride.

Secondly, all the scales are fused and feed into two convolution layers then added the original input signal $x$ to learn the residual:

$$z = H(Cat[Up_1(y_1), \ldots, Up_i(y_i)]) + x, \tag{4}$$

where $Up_i$ denotes $i\times$ Upsampling operation and $Cat$ denotes concatenation operation at the channel dimension. $H$ denotes a series of operations that consists of two $3 \times 3$ and one $1 \times 1$ convolution operations. The MSB can learn features with different scales and the all different features are fused to learn the primary feature.

Several MSBs make up our proposed multi-level multi-scale block that cascades a series of MSBs to learn features with different levels. We will give detailed analysis on the levels of MLMSB in Section 4.4.1.

### 3.3. Loss function

In the proposed weakly supervised network, the estimation of transmission map and atmospheric belong to unsupervised learning, and the estimation of the intermediate haze-free image can be regarded as supervised learning, i.e., the estimated transmission map $\tilde{t}(x)$ and estimated atmospheric light $\tilde{A}$ do not have the corresponding ground-truth. $\tilde{t}(x)$ and $\tilde{A}$ can be better learned from the network automatically by designing the loss functions.

The overall network consists of two loss functions: $\mathcal{L}_{dehaze}$ and $\mathcal{L}_{cycle}$, which denote supervised model and unsupervised model, respectively.

The traditional $L_1$ loss is used to learn the intermediate haze-free image $\tilde{J}(x)$.

$$\mathcal{L}_{dehaze} = \| \hat{J}(x) - \tilde{J}(x) \|_1, \tag{5}$$

where $\hat{J}(x)$ and $\tilde{J}(x)$ denote the ground-truth of the haze-free image and the intermediate haze-free image, respectively.

The traditional $L_1$ loss is used to learn the accurate transmission map $\tilde{t}(x)$ and atmospheric light $\tilde{A}$.

$$\mathcal{L}_{cycle} = \| \tilde{I}(x) - I(x) \|_1, \tag{6}$$

where $\tilde{I}(x) = \tilde{J}(x)\tilde{t}(x) + (1 - \tilde{t}(x))\tilde{A}$, $I(x)$ denote the input hazy image.

The overall loss function $\mathcal{L}$ is defined as follows:

$$\mathcal{L} = \mathcal{L}_{dehaze} + \lambda \mathcal{L}_{cycle}, \tag{7}$$

where $\lambda$ is a hyper-parameter and the analysis on it will be discussed in Section 4.4.2.

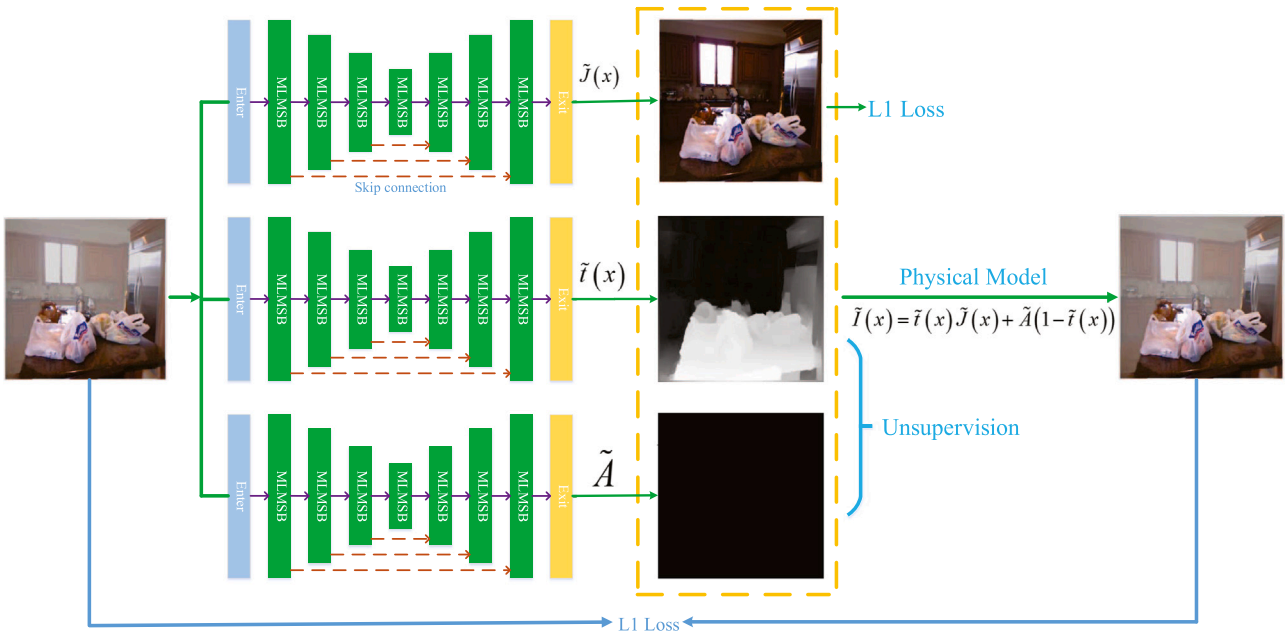



**Fig. 2.** An overview of the proposed image dehazing method. The final haze-free image is obtained via estimated $\tilde{t}(x)$ and $\tilde{A}$ combined the input hazy image by solving Eq. (2).

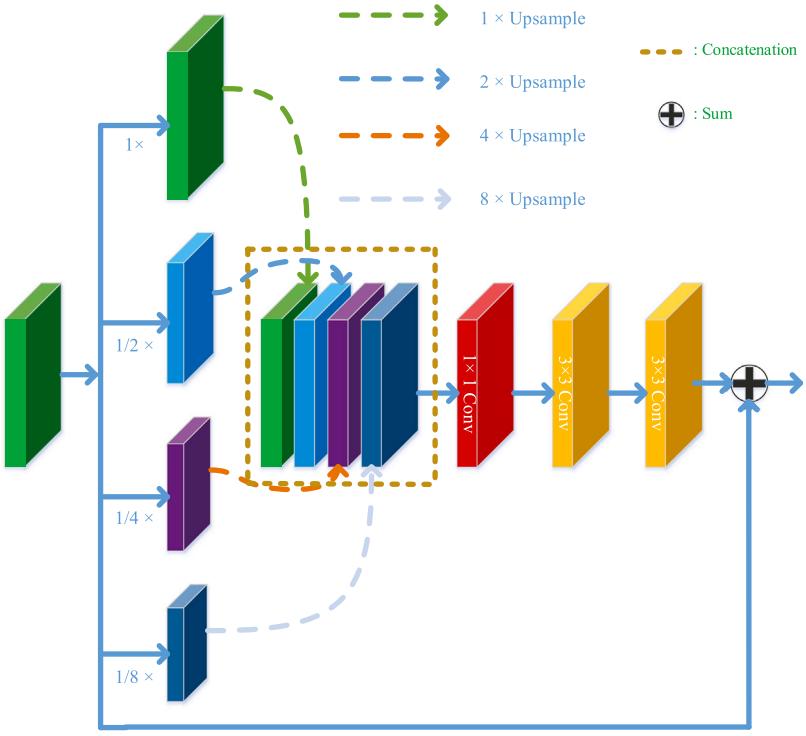



**Fig. 3.** Multi-Scale Block (MSB). The multi-level multi-scale block consists of several MSBs.

## 4. Experimental results

In this section, we evaluate the proposed dehazing method on both synthetic datasets and real-world hazy images, with comparisons to the state-of-the-art methods in terms of accuracy and visual effect. We first demonstrate the experimental settings and criterions of quality evaluation in Section 4.1. Quantitative comparisons on synthetic datasets are shown in Section 4.2 and visual comparisons on real-world images are provided in Section 4.3. All the results are compared with seven state-of-the-art methods, including DCP (CVPR'09) [14], CAP (TIP'15) [4], NLD (CVPR'16) [16], MSCNN (ECCV'16) [5], GFN (CVPR'18) [9], DCPDN (CVPR'18) [26], EPDN (CVPR'19) [29].

### 4.1. Experiment settings

**Training dataset**. For the training data, we use the training dataset TrainA provided by Zhang and Patel [26] to train our model. This dataset TrainA contains 4000 indoor synthetic images.

**Testing datasets**. We evaluate the performance of our method on three test datasets. One is introduced by Zhang and Patel [26], which

**Table 1**
Quantitative experiments evaluated on the synthetic dataset Test I. Best results are marked in bold.

| Metric | [14] CVPR'09 | [4] TIP'15 | [16] CVPR'16 | [5] ECCV'16 | [9] CVPR'18 | [26] CVPR'18 | [29] CVPR'19 | **Ours** |
|---|---|---|---|---|---|---|---|---|
| PSNR | 13.67 | 16.10 | 16.68 | 19.58 | 22.12 | 26.32 | 24.80 | **30.53** |
| SSIM | 0.7428 | 0.7569 | 0.7332 | 0.8462 | 0.8768 | 0.9155 | 0.9209 | **0.9660** |
| CIEDE2000 | 11.97 | 16.04 | 11.98 | 7.90 | 6.35 | 3.82 | 4.99 | **2.18** |


**Fig. 4.** Visual and quantitative comparisons on a synthetic example. (a) Input. (b) He et al. [14]. (c) Zhu et al. [4]. (d) Berman et al. [16]. (e) Ren et al. [5]. (f) Ren et al. [9]. (g) Zhang et al. [26]. (h) Qu et al. [29]. (i) Ours. Compared with other results, our dehazed result is more pleasing.

contains 400 images for testing, denoted as Test I. Another one is SOTS-outdoor test set [38] from **RESIDE**, it contains 500 outdoor images, denoted as Test II. The last one is HSTS test set [38] from **RESIDE**, which contains 10 real-world images, denoted as Test III. The three test datasets include both indoor and outdoor scenes. The real-world hazy images we use for testing are provided by the previous methods. We use them to demonstrate the generalization ability of our network.

**Training settings**. In our network, we randomly crop each training image pairs to 256 × 256 patch pairs. We use ADAM [39] optimizer with a batch size 24 for training. For each multi-scale block (MSB) of the proposed MLMSB, the nonlinear activation we used is LeakyReLU with $\alpha = 0.2$ after each convolution layer. The initial learning rate is $1 \times 10^{-3}$, and is updated twice by a rate of 1/10 at 45 000 and 60 000 iterations. Our entire network is trained on three Nvidia GTX 1080Ti GPUs based on PyTorch. To process a 512 × 512 pixels hazy image, the testing process only takes 0.0608 s on a PC with a GTX 1080Ti GPU.

As the proposed network is a weakly supervised dehazing method and it can be trained on the real-world dataset, we train our model on the real-world dataset via Eq. (6). Eq. (6) only involves the hazy image without any other auxiliary information. After training on the synthetic dataset, we continually train our model on the real-world dataset. We set 25 epochs and the batchsize is 1 to fine-tune our network on the real-world dataset RTTS [38] from **RESIDE**.

**Quality measures**. To evaluate the quality of the dehazed results in comparison with ground-truth images for a specific algorithm, we adopt three metrics: Peak signal to noise ratio (PSNR), structure similarity index (SSIM), CIEDE2000. PSNR and SSIM are widely used to evaluate the quality of restored results with ground-truth, i.e., estimated dehazing result and ground-truth. The higher its value, the better the restored image will be. SSIM is consistent with human perception, which is a measure of similarity between two images. The value of SSIM is closer to 1, the more similar the two images are. CIEDE2000 measures the color difference, the small value of CIEDE2000 indicates better color preservation. The three metrics only computed for synthetic datasets. For the real-world images, we use the Natural Image Quality Evaluator (NIQE) and visual comparisons to evaluate the performance of our method. The NIQE calculates the no ground-truth image quality score for the real-world hazy image. The lower the image quality is, the higher NIQE is.

### 4.2. Results on synthetic datasets

Our synthesized hazy images are accompanied with ground-truth images, enabling us to compare those dehazed results in terms of PSNR, SSIM and CIEDE2000. Table 1 displays the results on the synthetic datasets Test I. On the synthetic dataset Test I, the proposed method generates the results with higher PSNR, SSIM and CIEDE2000 among all the compared methods.

To provide visual comparisons, we test our method on some images from Test I and show three challenging examples in Figs. 4–6. As can be seen in Fig. 4, the methods of NLD [16], MSCNN [5], GFN [5], DCPDN [26] and EPDN [29] tend to underestimate haze concentration so that the dehazed results have some remaining haze. The results by He et al. [14] and Zhu et al. [4] are visually better than the results by Berman et al. [16], Zhang and Patel [26], Ren et al. [5,9] and Qu et al. [29]. However, by looking closer, there still exist haze residues in some regions.

Fig. 5 shows another synthetic example. The dehazed results by the prior based methods DCP [14], CAP [4], and NLD [16] contain significant color distortion as they usually assume that the atmospheric light is constant, as shown in Fig. 5(b)–(d). The learning based methods MSCNN [5], GFN [9], DCPDN [26] and EPDN [29], use end-to-end trainable networks to directly estimate the haze-free images, these methods generate better results than the prior based ones, but the dehazing results still contain color distortions in some regions (for instance, the man's jeans).

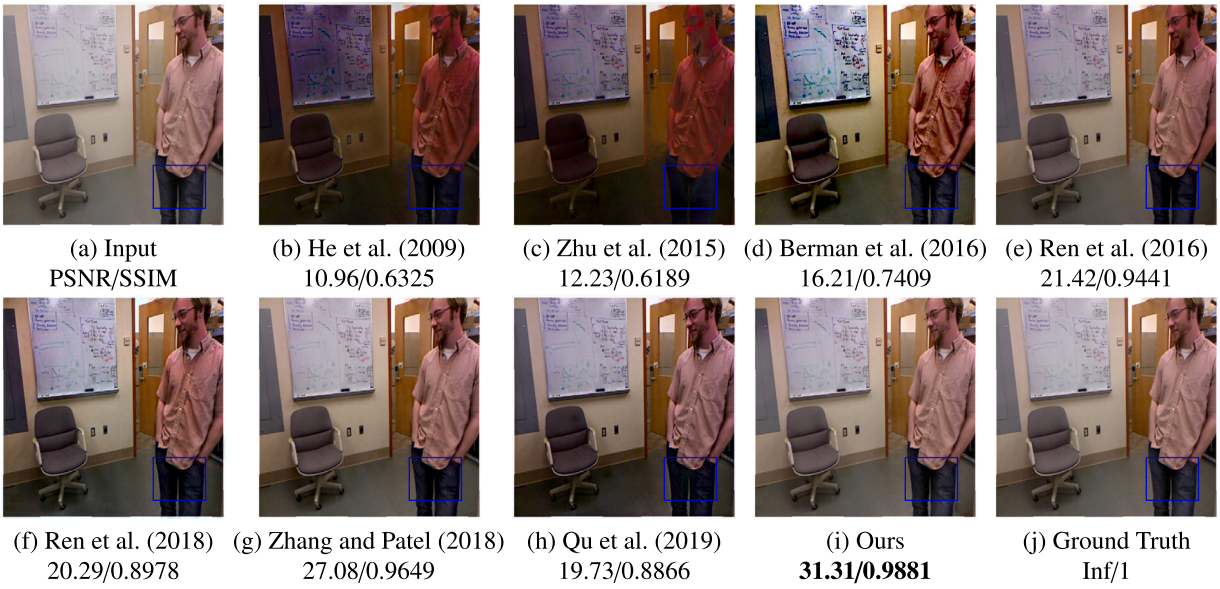



**Fig. 5.** Visual and quantitative comparisons on a synthetic example. (a) Input. (b) He et al. [14]. (c) Zhu et al. [4]. (d) Berman et al. [16]. (e) Ren et al. [5]. (f) Ren et al. [9]. (g) Zhang et al. [26]. (h) Qu et al. [29]. (i) Ours. Visually, the proposed method generates the cleanest result with the least color distortions.

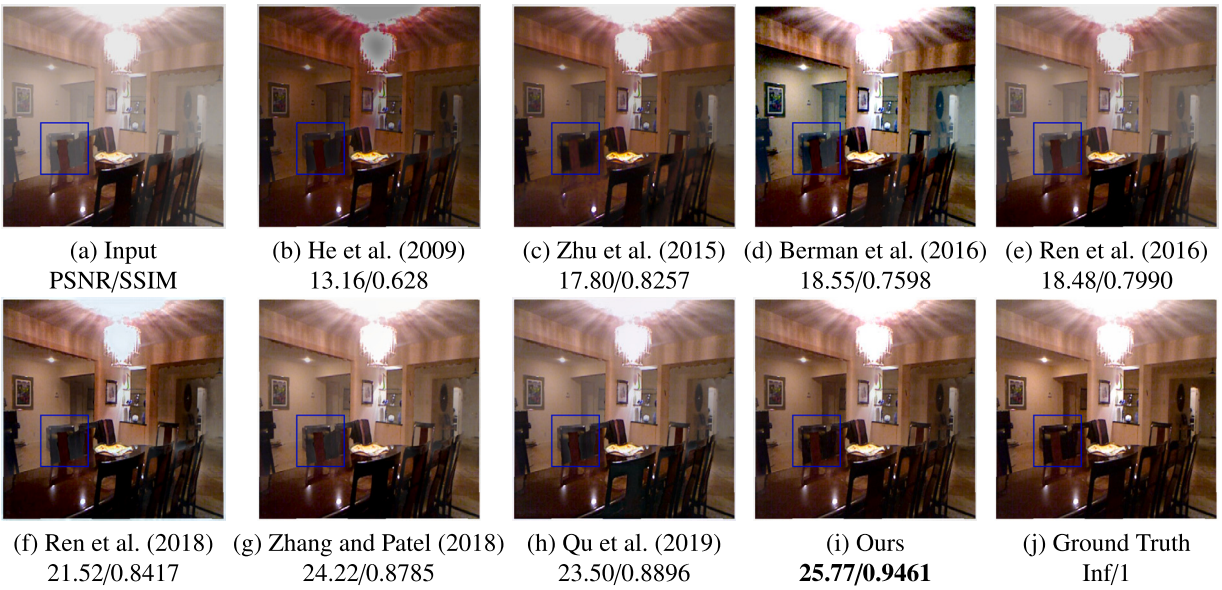



**Fig. 6.** Visual and quantitative comparisons on a synthetic example. (a) Input. (b) He et al. [14]. (c) Zhu et al. [4]. (d) Berman et al. [16]. (e) Ren et al. [5]. (f) Ren et al. [9]. (g) Zhang et al. [26]. (h) Qu et al. [29]. (i) Ours. Our method generates a clearer image visually closer to the ground-truth image.

The dehazed results for another image are shown in Fig. 6. The methods DCP [14], CAP [4], NLD [16], and GFN [9] overestimate the thickness of the haze and produce dark results with color distortions as shown in Fig. 6(b), (c), (d) and (f). Although the dehazed results by Ren et al. [5], Zhang and Patel [26] and Qu et al. [29] are closer to ground-truth than the results by He et al. [14], Zhu et al. [4], Berman et al. [16] and Ren et al. [9], there are still some remaining haze. Overall, our dehazed results have higher visual quality and fewer color distortions compared with other methods.

### 4.3. Results on real-world images

We further evaluate the performance of the proposed method on a series of real-world hazy images. Fig. 7 presents one example. The results of DCP [14], CAP [4] and NLD [16] shown in Fig. 7(b)–(d) significantly suffer from over-enhancement. The background buildings are much darker than they should be and the recovered images look unnatural. The methods MSCNN [5], GFN [9], DCPDN [26] and EPDN [29] suffer from various degree of color distortions as shown in Fig. 7(e)–(h). In comparison, our result shown in Fig. 7(i) are visually the best.

Another real-world example is shown in Fig. 8. There are still exist haze residues or color distortion in the recovered images by the methods [4,5,9,14,16,26,29], while our method generates more natural result with vivid colors as shown in Fig. 8(i)

### 4.4. Analysis and discussions

#### 4.4.1. Analysis on the number of the levels

In the proposed network, the number of levels plays a critical role for the dehazed results. In order to choose a suitable number of levels,


**Fig. 7.** Example results evaluated on real-world image datasets. (a) Input. (b) He et al. [14]. (c) Zhu et al. [4]. (d) Berman et al. [16]. (e) Ren et al. [5]. (f) Ren et al. [9]. (g) Zhang et al. [26]. (h) Qu et al. [29]. (i) Ours. Compared with the state-of-the-art dehazing methods, our method generates the dehazed result visually better than the other methods.

**Table 2**
Quantitative experiments for analyzing the hyper-parameters $\lambda$.

| $\lambda$ | 0.001 | 0.1 | 0.25 | 0.5 | 0.75 | 1 |
|---|---|---|---|---|---|---|
| PSNR | 29.01 | 29.32 | 30.53 | 29.36 | 29.24 | 29.27 |
| SSIM | 0.9560 | 0.9571 | 0.9660 | 0.9614 | 0.9615 | 0.9615 |

**Table 3**
Analysis on whether fine-tune.

| Metric | w/o fine-tune | fine-tune |
|---|---|---|
| PSNR | 16.01 | 24.18 |
| SSIM | 0.6988 | 0.9316 |

we carried out experiments on the synthetic test datasets [26] to test 5 different values of levels varying from 1 to 5. The statistical results are summarized in Fig. 9. In the subfigures, the two horizontal axes denote the number of levels, and the two vertical axes denote the average PSNR and average SSIM, respectively. In Fig. 9(a), we observe that the optimal number of levels is 3 and it achieves the average PSNR higher than 30. Fig. 9(b) demonstrates that when the number of levels is 3, the proposed network generates results with a relatively larger average SSIM. Based on the above observations, we think the number of levels equal to 3 is a good choice and set the number of levels to be 3 in all of our network.

#### 4.4.2. *Effect of the loss functions*

To generate high quality dehazed images, we propose a loss function which includes two terms with a important hyper-parameter $\lambda$. we evaluate the performance of our method by varying the hyper-parameter from 0.001 to 1 and compute the average PSNR and average SSIM to measure the dehazed results on the test dataset Test I. Table 2 shows that when $\lambda = 0.25$, the proposed method generates the best results in terms of both PSNR and SSIM. So we choose $\lambda = 0.25$ as our network setting.

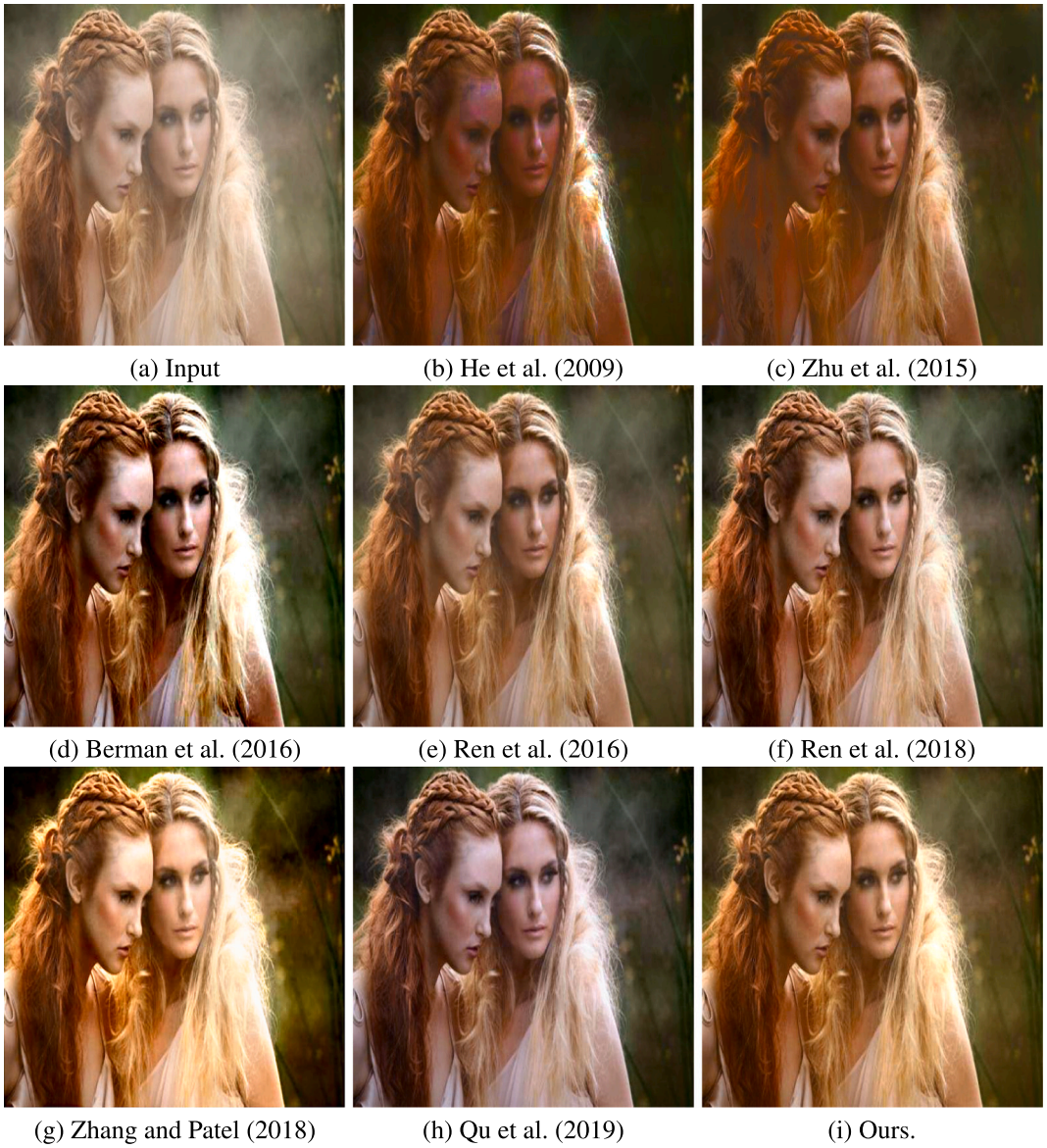



**Fig. 8.** Example results evaluated on real-world image datasets. (a) Input. (b) He et al. [14]. (c) Zhu et al. [4]. (d) Berman et al. [16]. (e) Ren et al. [5]. (f) Ren et al. [9]. (g) Zhang et al. [26]. (h) Qu et al. [29]. (i) Ours. Our method removes haze and generates clearer image with vivid colors.

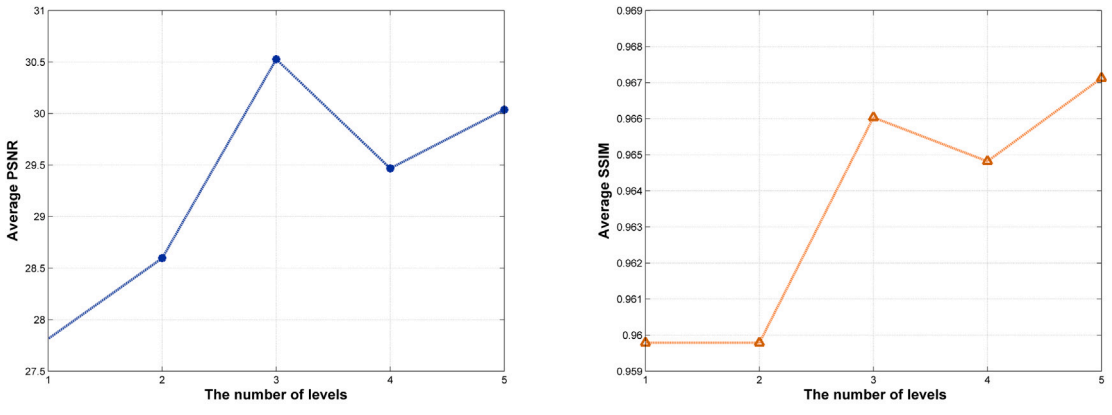
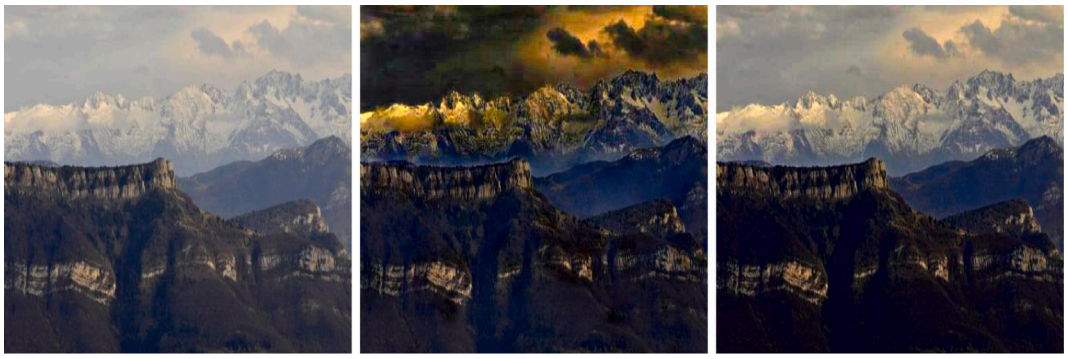



**Fig. 9.** The results on the number of the levels. (a) The curve of average PSNR. (b) The curve of average SSIM.



(a) (b) (c)

**Fig. 10.** (a) Hazy input. (b) Result without fine-tune on a real-world dhazy image. (c) Result after fine-tune on a real-world hazy image. The result with the fine-tune operation is visually better than the result without fine-tune operation.

**Table 4**
Quantitative experiments evaluated on the real-world images. Best results are marked in bold.

| | He *et al.* [14] | Zhu *et al.* [4] | Berman *et al.* [16] | Ren *et al.* [5] | Ren *et al.* [9] | Zhang *et al.* [26] | Qu *et al.* [29] | **Ours** |
|---|---|---|---|---|---|---|---|---|
| NIQE | 4.2223 | 3.7256 | 3.9841 | 3.8429 | 3.6653 | 4.3049 | 3.8379 | **3.5108** |

**Table 5**
Quantitative experiments evaluated on the synthetic datasets Test II. Best results are marked in bold.

| Metric | [14] | [4] | [16] | [5] | [9] | [26] | [29] | **Ours** |
|---|---|---|---|---|---|---|---|---|
| PSNR | 18.08 | 14.83 | 18.21 | 19.60 | 21.66 | 18.94 | 22.98 | **24.18** |
| SSIM | 0.8489 | 0.7881 | 0.7843 | 0.8750 | 0.8513 | 0.8700 | 0.9033 | **0.9316** |
| CIEDE2000 | 10.25 | 13.86 | 9.86 | 7.93 | 6.39 | 8.63 | 6.93 | **5.36** |

### *4.5. Analysis on the fine-tune operation*

We set 25 epochs to fine-tune our network on the real-world dataset RTTS [38] from **RESIDE**. The qualitative results on Test II (outdoor images) are reported in Table 3. Our method without the fine-tune operation is limited on the out-door dataset Test II because it is trained on the indoor dataset TrainA [26]. The fine-tune operation improves the dehazed results. We also perform experiments on the real-world dataset HSTS [38] from **RESIDE** and the Natural Image Quality Evaluator (NIQE) is used as the performance evaluation criterion on this dataset. The comparative quantitative results are shown in Table 4. It demonstrates that our method achieves the best performance of image dehazing in terms of NIQE on real-world dataset.

Moreover, a real-world visual example is presented in Fig. 10. The distribution of light is uneven without the fine-tune operation and the dehazed result is more natural after the fine-tune operation.

Quantitative comparisons between the proposed method and other state-of-the-art methods on the dataset Test II are shown in Table 5. In terms of the three evaluation criterions, our dehazed results are the best among all the methods.

## 5. Conclusion

In this paper, we demonstrate the feasibility of weakly supervised image dehazing. Our network can be trained only using hazy and haze-free image pairs as supervision and is able to generate perceptually appealing dehazed results by estimating transmission map and atmospheric light automatically. What is more, our network can be trained on the real-world dataset as the semi-supervision to fine-tune the model and the fine-tuning operation improves the dehazing performance on the real-world dataset. It provides a new viewpoint for future unsupervised haze removal research.

## Declaration of competing interest

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

## Acknowledgments

This work was supported by the Natural Science Foundation of China [grant numbers 61976041]; National Science and Technology Major Project [grant number 2018ZX04041001]; National Key R&D Program of China [grant number 2018AAA0100300].

## References

[1] G. Meng, Y. Wang, J. Duan, S. Xiang, C. Pan, Efficient image dehazing with boundary constraint and contextual regularization, in: ICCV, 2013, pp. 617–624, http://dx.doi.org/10.1109/ICCV.2013.82.

[2] K. Tang, J. Yang, J. Wang, Investigating haze-relevant features in a learning framework for image dehazing, in: CVPR, 2014, pp. 2995–3002, http://dx.doi.org/10.1109/CVPR.2014.383.

[3] Z. Li, P. Tan, R.T. Tan, D. Zou, S.Z. Zhou, L. Cheong, Simultaneous video defogging and stereo reconstruction, in: CVPR, 2015, pp. 4988–4997, http://dx.doi.org/10.1109/CVPR.2015.7299133.

[4] Q. Zhu, J. Mai, L. Shao, A fast single image haze removal algorithm using color attenuation prior, TIP 24 (11) (2015) 3522–3533, http://dx.doi.org/10.1109/TIP.2015.2446191.

[5] W. Ren, S. Liu, H. Zhang, J. Pan, X. Cao, M. Yang, Single image dehazing via multi-scale convolutional neural networks, in: ECCV, 2016, pp. 154–169, http://dx.doi.org/10.1007/978-3-319-46475-6_10.

[6] B. Cai, X. Xu, K. Jia, C. Qing, D. Tao, Dehazenet: An end-to-end system for single image haze removal, TIP 25 (11) (2016) 5187–5198, http://dx.doi.org/10.1109/TIP.2016.2598681.

[7] C. Chen, M.N. Do, J. Wang, Robust image and video dehazing with visual artifact suppression via gradient residual minimization, in: ECCV, 2016, pp. 576–591, http://dx.doi.org/10.1007/978-3-319-46475-6_36.

[8] B. Li, X. Peng, Z. Wang, J. Xu, D. Feng, AOD-Net: All-in-one dehazing network, in: ICCV, 2017, pp. 4780–4788, http://dx.doi.org/10.1109/ICCV.2017.511.

[9] W. Ren, L. Ma, J. Zhang, J. Pan, X. Cao, W. Liu, M. Yang, Gated fusion network for single image dehazing, in: CVPR, 2018, pp. 3253–3261, http://dx.doi.org/10.1109/CVPR.2018.00343, URL: http://openaccess.thecvf.com/content_cvpr_2018/html/Ren_Gated_Fusion_Network_CVPR_2018_paper.html.

[10] R.T. Tan, Visibility in bad weather from a single image, in: CVPR, 2008, http://dx.doi.org/10.1109/CVPR.2008.4587643.

[11] J. Tarel, N. Hautière, Fast visibility restoration from a single color or gray level image, in: ICCV, 2009, pp. 2201–2208, http://dx.doi.org/10.1109/ICCV.2009.5459251.

[12] M. Sulami, I. Glatzer, R. Fattal, M. Werman, Automatic recovery of the atmospheric light in hazy images, in: ICCP, 2014, pp. 1–11, http://dx.doi.org/10.1109/ICCPHOT.2014.6831817.
[13] D. Berman, T. Treibitz, S. Avidan, Air-light estimation using haze-lines, in: ICCP, 2017, pp. 115–123, http://dx.doi.org/10.1109/ICCPHOT.2017.7951489.
[14] K. He, J. Sun, X. Tang, Single image haze removal using dark channel prior, in: CVPR, 2009, pp. 1956–1963, http://dx.doi.org/10.1109/CVPRW.2009.5206515.
[15] R. Fattal, Dehazing using color-lines, TOG 34 (1) (2014) 13:1–13:14, http://dx.doi.org/10.1145/2651362.
[16] D. Berman, T. Treibitz, S. Avidan, Non-local image dehazing, in: CVPR, 2016, pp. 1674–1682, http://dx.doi.org/10.1109/CVPR.2016.185.
[17] Z. Xu, X. Yang, X. Li, X. Sun, Strong baseline for single image dehazing with deep features and instance normalization, in: BMVC, 2018, p. 243, URL: http://bmvc2018.org/contents/papers/0821.pdf.
[18] X. Yang, Z. Xu, J. Luo, Towards perceptual image dehazing by physics-based disentanglement and adversarial training, in: AAAI, 2018, pp. 7485–7492, URL: https://www.aaai.org/ocs/index.php/AAAI/AAAI18/paper/view/17047.
[19] L. Huang, J. Yin, B. Chen, S. Ye, Towards unsupervised single image dehazing with deep learning, in: ICIP, 2019, pp. 2741–2745, http://dx.doi.org/10.1109/ICIP.2019.8803316.
[20] W. Ren, J. Pan, H. Zhang, X. Cao, M. Yang, Single image dehazing via multi-scale convolutional neural networks with holistic edges, IJCV (2019) 1–20, http://dx.doi.org/10.1007/s11263-019-01235-8.
[21] D. Engin, A. Genç, H.K. Ekenel, Cycle-dehaze: Enhanced cyclegan for single image dehazing, in: CVPR Workshops, 2018, pp. 825–833, http://dx.doi.org/10.1109/CVPRW.2018.00127, URL: http://openaccess.thecvf.com/content_cvpr_2018_workshops/w13/html/Engin_Cycle-Dehaze_Enhanced_CycleGAN_CVPR_2018_paper.html.
[22] A. Dudhane, S. Murala, Cdnet: Single image de-hazing using unpaired adversarial training, in: WACV, 2019, pp. 1147–1155, http://dx.doi.org/10.1109/WACV.2019.00127.
[23] C. Su, L.K. Cormack, A.C. Bovik, Color and depth priors in natural images, TIP 22 (6) (2013) 2259–2274, http://dx.doi.org/10.1109/TIP.2013.2249075.
[24] A. Saxena, S.H. Chung, A.Y. Ng, Learning depth from single monocular images, in: NIPS, 2005, pp. 1161–1168, URL: http://papers.nips.cc/paper/2921-learning-depth-from-single-monocular-images.
[25] C. Chen, J. Wei, C. Peng, W. Zhang, H. Qin, Improved saliency detection in RGB-D images using two-phase depth estimation and selective deep fusion, TIP 29 (2020) 4296–4307, http://dx.doi.org/10.1109/TIP.2020.2968250.
[26] H. Zhang, V.M. Patel, Densely connected pyramid dehazing network, in: CVPR, 2018, pp. 3194–3203, http://dx.doi.org/10.1109/CVPR.2018.00337, URL: http://openaccess.thecvf.com/content_cvpr_2018/html/Zhang_Densely_Connected_Pyramid_CVPR_2018_paper.html.
[27] I.J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A.C. Courville, Y. Bengio, Generative adversarial nets, in: NeurIPS, 2014, pp. 2672–2680, URL: http://papers.nips.cc/paper/5423-generative-adversarial-nets.
[28] R. Li, J. Pan, Z. Li, J. Tang, Single image dehazing via conditional generative adversarial network, in: CVPR, 2018, pp. 8202–8211, http://dx.doi.org/10.1109/CVPR.2018.00856, URL: http://openaccess.thecvf.com/content_cvpr_2018/html/Li_Single_Image_Dehazing_CVPR_2018_paper.html.
[29] Y. Qu, Y. Chen, J. Huang, Y. Xie, Enhanced pix2pix dehazing network, in: CVPR, 2019, pp. 8160–8168, http://dx.doi.org/10.1109/CVPR.2019.00835, URL: http://openaccess.thecvf.com/content_CVPR_2019/html/Qu_Enhanced_Pix2pix_Dehazing_Network_CVPR_2019_paper.html.
[30] W. Chen, J. Ding, S. Kuo, PMS-Net: Robust haze removal based on patch map for single images, in: CVPR, 2019, pp. 11681–11689, URL: http://openaccess.thecvf.com/content_CVPR_2019/html/Chen_PMS-Net_Robust_Haze_Removal_Based_on_Patch_Map_for_Single_CVPR_2019_paper.html.
[31] T. Lin, P. Dollár, R.B. Girshick, K. He, B. Hariharan, S.J. Belongie, Feature pyramid networks for object detection, in: CVPR, 2017, pp. 936–944, http://dx.doi.org/10.1109/CVPR.2017.106.
[32] Y. Chen, Z. Wang, Y. Peng, Z. Zhang, G. Yu, J. Sun, Cascaded pyramid network for multi-person pose estimation, in: CVPR, 2018, pp. 7103–7112, http://dx.doi.org/10.1109/CVPR.2018.00742, URL: http://openaccess.thecvf.com/content_cvpr_2018/html/Chen_Cascaded_Pyramid_Network_CVPR_2018_paper.html.
[33] A. Ranjan, M.J. Black, Optical flow estimation using a spatial pyramid network, in: CVPR, 2017, pp. 2720–2729, http://dx.doi.org/10.1109/CVPR.2017.291.
[34] H. Zhao, J. Shi, X. Qi, X. Wang, J. Jia, Pyramid scene parsing network, in: CVPR, 2017, pp. 6230–6239, http://dx.doi.org/10.1109/CVPR.2017.660.
[35] K. He, X. Zhang, S. Ren, J. Sun, Spatial pyramid pooling in deep convolutional networks for visual recognition, in: ECCV, 2014, pp. 346–361, http://dx.doi.org/10.1007/978-3-319-10578-9_23.
[36] H. Zhao, J. Shi, X. Qi, X. Wang, J. Jia, Pyramid scene parsing network, in: CVPR, 2017, pp. 6230–6239, http://dx.doi.org/10.1109/CVPR.2017.660.
[37] O. Ronneberger, P. Fischer, T. Brox, U-Net: Convolutional networks for biomedical image segmentation, in: MICCAI, 2015, pp. 234–241, http://dx.doi.org/10.1007/978-3-319-24574-4_28.
[38] B. Li, W. Ren, D. Fu, D. Tao, D. Feng, W. Zeng, Z. Wang, Benchmarking single-image dehazing and beyond, TIP 28 (1) (2019) 492–505, http://dx.doi.org/10.1109/TIP.2018.2867951.
[39] D.P. Kingma, J. Ba, Adam: A method for stochastic optimization, in: ICLR, 2015, URL: http://arxiv.org/abs/1412.6980.