Information Sciences 541 (2020) 16–35

Contents lists available at ScienceDirect

Information Sciences

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

Deep joint neural model for single image haze removal and color correction

Tianlun Zhang [a,c], Xi Yang [c], Xizhao Wang [a,e], Ran Wang [b,d,*]

[a] College of Computer Science and Software Engineering, Shenzhen University, Shenzhen 518060, China
[b] College of Mathematics and Statistics, Shenzhen University, Shenzhen 518060, China
[c] College of Information Science and Technologe, Dalian Maritime University, Dalian 116026, China
[d] Shenzhen Key Laboratory of Advanced Machine Learning and Applications, Shenzhen University, Shenzhen 518060, China
[e] The Guangdong Key Laboratory of Intelligent Information Processing, Shenzhen University, Shenzhen 518060, China

ARTICLE INFO

Article history:
Received 9 July 2019
Received in revised form 20 May 2020
Accepted 26 May 2020
Available online 31 May 2020

Keywords:
Deep neural network
Single image de-hazing
Color adjustment
Physics-driven model
Joint optimization
Cyclic restoration

ABSTRACT

The quality of an image affects the performance of computer vision applications. The presence of haze often greatly depreciates the visual effect of images. It is a traditional and critical vision challenge to remove haze from a single image. This paper proposes a trainable end-to-end de-hazing connectionist model with a special design. First, feature learning is conducted using hierarchical convolutional layers with nested structures. Cascaded haze-relevant tasks are then sequentially performed via a physics-driven sub-network. In particular, to break the assumption of a homogeneous atmosphere, a branch of the sub-network estimates the scattering factor in the form of a two-dimensional tensor. Finally, a chromatic adaptation layer is proposed for color adjustment, which is often neglected in existing de-hazing methods. In addition, we integrate different training criteria based on the characteristics of the haze-relevant variables in our model. For a fully actionable optimization, an asynchronous learning paradigm is designed for the fusion of different de-hazing tasks, and the joint model is further facilitated by a cyclic restoration. The effectiveness of the proposed de-hazing model was verified via extensive experiments, and most results of our method are remarkable.

© 2020 Elsevier Inc. All rights reserved.

1. Introduction

In hazy environments, the light reflected from scene points is scattered and absorbed by turbid media, such as suspended particles and water droplets in the atmosphere. Meanwhile, the ambient light reflected from the atmospheric media becomes considerably more intense than it should be. Such atmospheric phenomena often lead to degraded images with low scene contrast and color fidelity, as shown by the *hazy input* in Fig. 1. When images are suffering from visual quality degradation, most outdoor vision systems, e.g., surveillance and autonomous navigation, may fail to perform favorably. Thus, it becomes highly desirable to develop effective approaches for image haze removal.

In research and applications, hazy images are described by attenuated and distorted visual characteristics. Conventionally, the visual features of a hazy image are improved via image enhancement. Various techniques, such as contrast restoration [11] and saturation adjustment [24], have been devoted to this purpose. However, the performance of enhancement

* Corresponding author at: College of Mathematics and Statistics, Shenzhen University, Shenzhen 518060, China.
E-mail address: wangran@szu.edu.cn (R. Wang).

https://doi.org/10.1016/j.ins.2020.05.105
0020-0255/© 2020 Elsevier Inc. All rights reserved.

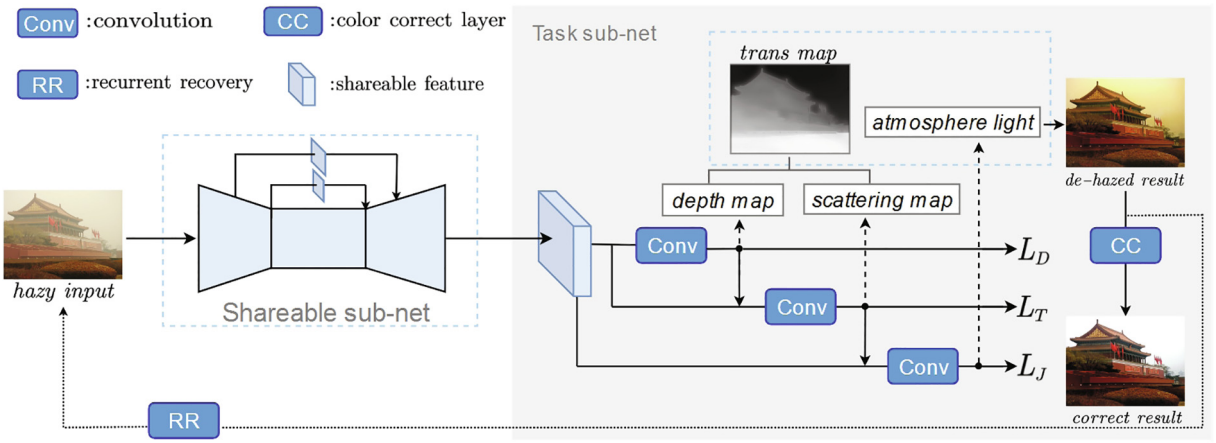


Fig. 1. Architecture of the proposed de-hazing model. The shareable sub-net learns the representative features from the hazy input. The task sub-net (in the gray region) is designed to recover haze-free images in a part-wise manner. Here, intermediate results depend on the previous steps. Then, L_D, L_T, and L_J are the optimization objects for the supervised learning of scene distance, transmission map, and de-hazed image, respectively.

methods is barely satisfactory since the degradation of hazy images is spatially variant. In other words, the problem with haze removal is an ill-posed problem if the necessary spatial information is unknown. Some studies have addressed this problem using multiple images or additional knowledge. For instance, an instant de-hazing method [32] can be used with two independent images taken through a polarizer at different orientations. Similarly, another method [27] can be used with multiple images taken from a scene under different weather conditions to obtain more constraints. In addition, user inputs and 3D models can be used to provide rough scene depth for de-hazing tasks [20]. However, such information is not always practically available; therefore, single image de-hazing methods have emerged as popular tools to attain high levels of perceived quality.

Typically, considerable development in single image haze removal processes depends on stronger assumptions and priors. According to the empirical statistics on natural haze-free images, He et al. [13] found the well-known dark channel prior (DCP), which defines dark pixels as the low intensity pixels in at least one color channel of each local region. The remarkable advantage of DCP is that the dark pixel can directly infer the proportion of scene light captured by the camera. By applying this prior to an image de-hazing model, high-quality haze-free results can be realized. For available scene structures, Zhu et al. [45] introduced the color attenuation prior (CAP) into a linear model, which can estimate the spatially variant depth using the difference between brightness and saturation. This prior provides a simple but powerful approach for building the bridge between a hazy image and its unknown spatial information. In addition, the physically solid model related to color lines promotes research findings on the intensity of scattered light [7]. In particular, through localized assumptions and global image statistics, Sulami et al. [35] derived two procedures to recover the orientation and magnitude of the atmospheric light. By adopting this method, several de-hazing works [1] have achieved favorable results since accurate atmospheric light can avoid low brightness and over-saturation while maintaining color fidelity. Pioneering works demonstrate that the useful haze-relevant information and heuristic cues can be acquired from a single hazy image. Based on this, one may attempt to associate single image de-hazing work with deep learning methods. Recent developments regarding deep learning have demonstrated that a deep connectionist model can automatically learn the representation of data with different levels of abstraction. These deep representation learning methods have created an immense opportunity to improve the state-of-the-art performance in various fields. In particular, there are many breakthroughs in image processing [30]. In fact, a few deep learning methods have been proposed for single image haze removal and have achieved outstanding results due to feature representation learning, further discussed in the next section.

Despite the achievements and advancements, there are still a few problems in existing image de-hazing studies. First, although the low and blurred visibility of scenes can be effectively addressed, the unbalanced color tends to be neglected. Different from visibility degradation, the color cast problem is caused by two factors: 1) the absorption effect of the atmospheric media leads to energy reduction in a certain color channel; 2) the airlight reflected from the turbid particles is often blended in the incoming sight line. Consequently and inevitably, the de-hazed results have shifted scene colors, as shown by the *de-hazed result* in Fig. 1. For more pleasing results, a haze removal method that can provide satisfactory color correction should be developed. Second, existing de-hazing methods have a general assumption of homogeneous atmosphere [13,35,45]. Therefore, the scattering coefficient of the atmosphere is uniform in a hazy image, and the attenuation is only dependent on the scene depth. However, in most practical scenarios, the scattering effect often leads to random attenuation since the scattering coefficient may vary at different positions. To improve the robustness with respect to random attenuation, the de-hazing model should be able to describe the scattering amount as a function of both scene depth and scattering coefficient, which are spatially variant. Finally, the relationships between haze-relevant variables have inherent rules, which cannot be accurately expressed by the common activation functions used in neural networks. Moreover, de-hazing networks

tend to be complex non-linear mappings with immense parameter sizes. To achieve better optimization, a well-established paradigm can be adopted to pre-train the model parameters via auxiliary tasks or highly related training samples. This optimization paradigm has been proven to be effective [3] since the pre-trained model can transfer useful information to the target task. Based on the problems mentioned above, we focus on conducting further investigation into improving the visual appeal of hazy images. Fig. 1 shows the entire de-hazing system proposed in this paper. More concretely, our work can be summarized as follows.

(1) An innovative de-hazing neural model is proposed. In this model, the shareable sub-network is a holistically nested structure for feature learning. The shareable features are propagated into the task sub-network, which is a downstream structure with cascaded network branches. This sub-network helps estimate haze-relevant variables, particularly, the spatially variant scattering coefficient map. In addition, the neural model ends with a color constancy layer (represented as CC in Fig. 1), which is employed to correct the shifted colors caused by atmospheric absorption.
(2) It is interesting that some imaging models in this field are differentiable during back propagation. Owing to this, we use optical rules as viable alternatives of traditional non-linear activation functions. Thus, the proposed network is a physics-driven model for haze removal. Moreover, this de-hazing system is designed as a multi-task optimization in which each supervised criterion is a task-dependent term conditioned on an edge-aware regularizer.
(3) We demonstrate a reasonable learning paradigm to fuse different haze-relevant tasks in a unified connectionist model. Moreover, a task of recurrent recovery for hazy observation is designed to reinforce the joint optimization. With the cyclic restoration and fusion of learning patterns, this model achieves competitive performance for single image haze removal.

The rest of this paper is organized as follows. In Section 2, we review some basic concepts related to our model. In Section 3, we present the details of the proposed de-hazing method. In Section 4, comprehensive experiments are presented. Finally, the conclusions and future works are discussed in Section 5.

2. Preliminaries

2.1. Hazy imaging model

In the field of computer vision, the atmospheric scattering model [28] has been employed in various de-hazing studies. This imaging model is a linear superimposition of two components, i.e., scene radiance attenuation and airlight (ambient light in the line of sight), as shown in Fig. 2. Mathematically, the model describes the formation of a hazy image as follows:

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

where x indexes the location of a pixel, I denotes the hazy image intensity, α is the global atmospheric light, t refers to the medium transmission, which represents the proportion of light that is not scattered via media and is captured by the camera, and J is the scene radiance. Note that $I(x)$, $J(x)$, and α are vectors in $\mathcal{R}^3$ with one intensity value per color channel. The term $J(x)t(x)$ is the direct attenuation describing the degradation of scene radiance, and the other term $\alpha(1 - t(x))$ is the airlight

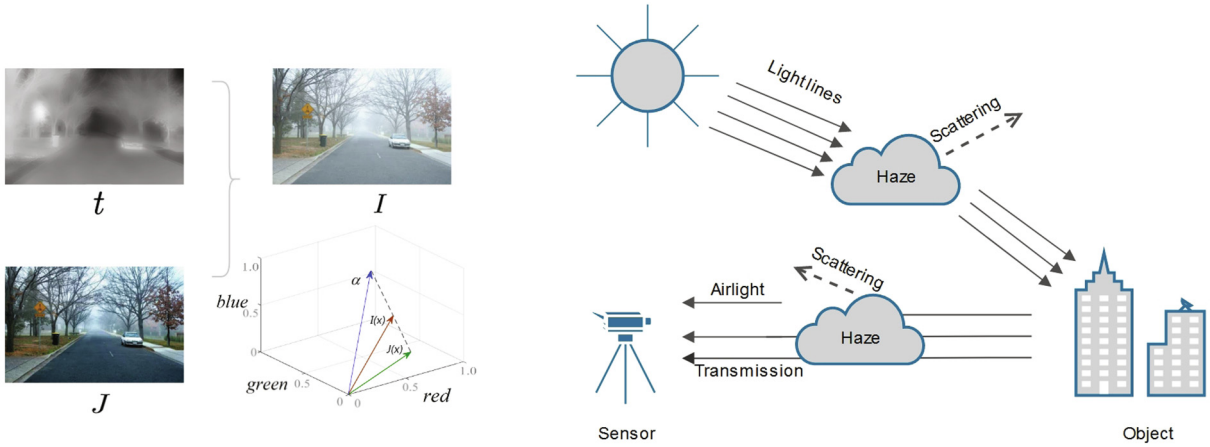


Fig. 2. Left: atmospheric scattering model. I is the hazy image, J is the haze-free image, t is the transmission, and α is the atmospheric light. *red*, *green* and *blue* denote the color channels. Right: illustration of the imaging in hazy environment. Transmission attenuation is the scene radiance decayed in media, while airlight is formed by the scattered light and results in the chromatic shift. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

caused by light scattering, which leads to scene color shift. According to Eq. (1), the haze-free scene radiance $J(x)$ can be restored by

$$J(x) = \frac{I(x) - \alpha}{t(x)} + \alpha. \tag{2}$$

Here, the haze removal of an image relies on the estimations of t and α. Based on the Lambert–Beer law [5], the light traveling through transparent objects will be attenuated exponentially, thus $t(x)$ can be formulated as

$$t(x) = e^{-\beta d(x)} \tag{3}$$

where d denotes the scene depth and β is the scattering coefficient of the atmosphere. Essentially, β is crucial for haze removal. However, many de-hazing methods [13,35,45] assume that the atmosphere is homogeneous. Thus, for a given hazy image, β is considered as a constant and the scene radiance is only degraded with respect to d, namely,

$$0 = \lim_{d(x)\to+\infty} J(x)t(x). \tag{4}$$

When $d(x)$ approaches infinity, $t(x)$ theoretically goes to zero and the atmospheric light α approximately equals to $I(x)$. However, this is an ideal case for estimating the atmospheric light. In practice, an alternative way is to predict it accdfording to the low value t_0 in the transmission map [13], i.e.,

$$\alpha = \max_{y\in\{x|t(x)\leqslant t_0\}} I(y). \tag{5}$$

Here, y is a pixel location on which the intensity value approximates the atmospheric light. In the natural appearance of a hazy image, scene radiance is impacted by two main factors. Besides the scattering effect mentioned above, the light attenuation also results from absorption, which is modeled as a function of both scene distance and wavelength. As a result, the image chrominance tends to have sharp changes.

A popular approach for color correction is white balancing, which has a large number of available variants [6]. In particular, Finlayson et al. [8] proposed a Shades-of-Grey method to calculate the scene illumination of each channel through the Minkowski p-norm. When p is 1, this method is a special case of the Gray-World, and when p tends to infinity, it behaves in a manner similar to the White-Patch hypothesis [6]. Finlayson et al. also went a step further to investigate a light-weight color constancy approach, which yields comparative results with complicated white balancing algorithms, e.g., the image statistics-based method [9]. Furthermore, Gijsenij et al. [18] proposed a color constancy method based on the Grey-Edge hypothesis that is similar to Shades-of-Grey, and this method can be formulated via the extended form of the Minkowski p-norm. Limare et al. [23] proposed the simplest white balancing algorithm in which a naive color correction is conducted using a pixel-wise affine transformation, i.e.,

$$f(J^c(x)) = \frac{(J^c(x) - V_{\min}) \times (J^c(x)_{\max} - J^c(x)_{\min})}{V_{\max} - V_{\min}} + J^c(x)_{\min}, \tag{6}$$

where $c \in \{red, green, blue\}$ is the channel index, $J^c(x)_{\max}$ and $J^c(x)_{\min}$ denote the highest and lowest values, respectively, in one channel with size N, $V_{\min}$ and $V_{\max}$ are values taken at positions $N \times s_1$ and $N \times (1 - s_2) - 1$, respectively, from an ascending sequence of all pixel values in this channel, and $s_1, s_2 \in [0, 1]$ denote the saturation levels. The pixels with $J^c(x) \leqslant V_{\min}$ (resp. $J^c(x) \geqslant V_{\max}$) are then updated as $V_{\min}$ (resp. $V_{\max}$).

Based on the aforementioned imaging models, we design a deep joint neural network for single image haze removal. Without any assumption or prior, this method can independently learn how to estimate the unknown variables related to the de-hazing work. Different from most existing works, our method models the scattering coefficient as a spatial variable rather than a single constant. Moreover, our method can perform chromatic adaptation for some extreme cases via color shift. To this end, we develop the neural model by incorporating a few special designs, described in Section 3.

2.2. Relevant deep learning based works

In the last decade, tremendous efforts have been made in the field of deep learning, which has experienced continuous development in a wide range of applications, e.g., computer vision [29], natural language processing [41], and graph data [39,40]. In this section, we will first review a deep neural structure related to our proposed model, then introduce the haze removal methods based on deep learning.

With regard to deep structure, a remarkable model is the convolutional neural network (CNN), which considers the shared local connection and pooling kernel as the fundamental processing units for computational layers. The initial CNN was proposed as a viable alternative to hand-crafted features in image recognition [4]. Following this idea of feature learning, various convolutional structures have been proposed in the past few years. In particular, GoogLeNet emerged as an important member in the family of CNNs due to its convincing performance in the classification challenge of 2014 ILSVRC [36]. An interesting observation is that the number of parameters in GoogLeNet is considerably smaller compared with those in its predecessor AlexNet [21] and its concurrent competitor VGGNet [34]. These compelling characteristics of GoogLeNet benefit from one architectural innovation, i.e., the inception structure, which conducts multiple convolutional operations with dif-

ferent filter bank sizes, and in the end concatenates highly correlated neurons. In the multi-scale processing, a technique of dimension reduction is employed to maintain the computational budget constant. According to several principles of architecture design, inception-style models and their variants further spur research on high-performance CNNs [37].

Recently, some CNN-based methods have been proposed for image haze removal. Cai et al. [2] proposed an early relevant work, i.e., DehazeNet, which is used to learn the mapping relationship between a hazy image and its associated medium transmission map. Different from classical CNNs, DehazeNet has a special structural design. In particular, the first convolutional layer with a max-out unit can simulate different types of filters, such as the opposite, round, and all-pass filters. The next layer is an inception-style model for the fusion of multi-scale features. Owing to the crafted structural design, DehazeNet can effectively learn haze-relevant knowledge, e.g., the dark channel feature, maximum contrast, color attenuation, and hue disparity, which are useful for estimating the transmission. Subsequently, the estimation result is refined by the guided filter [14], and the haze-free image is recovered via traditional atmospheric models as mentioned in the previous section. For an end-to-end haze removal model, Zhang et al. [43] proposed a densely connected pyramid de-hazing network (DCPDN). The framework of DCPDN can be divided into two sub-models, i.e., a generative model and an adversarial model [10]. The generative model is formed by two parallel encoder–decoder networks to obtain a haze-free image. The adversarial model is a discriminator for evaluating the generative results. Benefiting from the generative-adversarial framework, DCPDN can utilize the pixel-, feature-, and symbol-level information to improve performance. In addition, the motivation of cross-layer concatenation roots in the dense convolutional network [15] can facilitate the gradient backpropagation while maximizing the feature reuse. Ren et al. [31] proposed another end-to-end trainable model called a gated fusion network (GFN). This model comprises an encoder and a decoder. For more contextual information in larger spatial regions, dilation convolution [42] is adopted to expand the receptive field in the layers of the encoder block. Through the deconvolutional layers of the decoder, three confidence maps are learned for the gated fusion of three enhanced versions derived from the raw hazy image, and then the combination result yields the haze-free outcome. The gating operation is an effective strategy of image fusion that can preserve the high-quality visual regions while reducing patch-based artifacts.

3. Proposed de-hazing model

The proposed de-hazing model is a single and unified network that incorporates an atmospheric scattering model and color correction using activation functions and empirical criteria specialized for de-hazing problems. Moreover, this model is a fully convolutional inception structure [36] composed of two sub-networks. The first one is designed for representative learning based on multi-pipeline feature fusion, and the second one is a cascaded CNN for characteristic learning of heterogeneous but related haze removal tasks. Additionally, a cyclic restoration loss facilitates the joint optimization of different de-hazing tasks.

3.1. Sub-network for representative learning

As illustrated in Fig. 3, the first 16 shareable convolutional blocks construct the representative model, which involves two hierarchical feature in-networks. On the bottom-up pathway, the feature maps are proportionally scaled with a fixed scaling

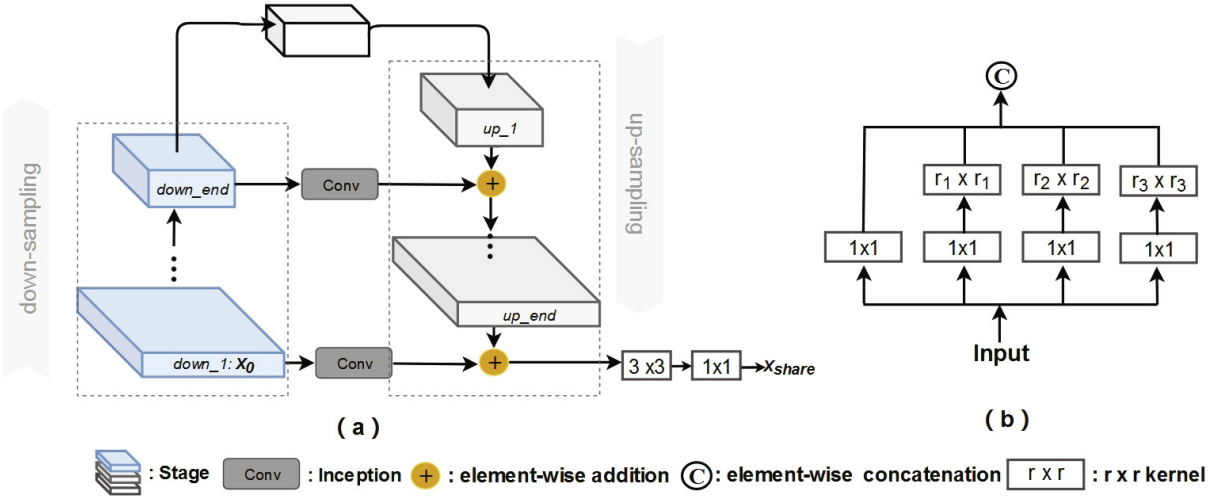


Fig. 3. Illustration of representative network. (a) shows the nested structure with down-sampling, up-sampling, and lateral connection. X_0 is the hazy input, and X_{share} is the learned features shared for the vision tasks. (b) is the inception block with two convolutional architectures. The first layer comprises 1×1 convolution, and the filters in the second layer have different kernel sizes.

step of 2. In particular, the layers with the same hierarchy level are defined as a network stage in which a convolution block, i.e., $\phi(X)$, is defined as

$$\phi(X) = [\delta(X), \tau_{r_1}(X), \tau_{r_2}(X), \tau_{r_3}(X)], \tag{7}$$

where

$$\delta(X) = \xi(W_{1\times1} * X), \tag{8}$$

and

$$\tau_r(X) = \xi(W_{r\times r} * \delta(X)). \tag{9}$$

In Eqs. (7)–(9), X is the input feature map, $W_{r\times r}$ denotes the convolution kernel with size of $r \times r$, $[\cdot]$ represents a concatenation operation, $*$ is a convolutional operation, and ξ is the batch normalization [17] followed by a rectified linear unit [26]. The resolution of each dimension in the top stage is one-sixteenth of that in the input observation, and the top stage outputs the highest-level semantic features in this down-sampling in-network. Contrarily, fractionally strided poolings are employed to recover the original resolution of the input image. Despite being semantically stronger, the feature maps become spatially coarser from the top to the bottom. For more exact activations and locations, these coarser-resolution maps are enhanced via lateral connections, each of which merges features with the same spatial size from the top-down or bottom-up direction. One lateral connection in the up-sampling in-network can be defined as

$$X_{8+i} = up(X_{7+i}) \oplus \phi(\phi(X_{8-i})), \tag{10}$$

where $\oplus$ denotes the element-wise addition, the subscript of X indexes the layer ($i = 2, 4, 6, 8$), $up(\cdot)$ doubles the size of the feature map, and $X_0 = I$ is the hazy observation. The output X_{16} directly leads to a block artifact in the desired results. To remove the aliasing effect caused by up-sampling, a 3×3 convolution layer followed by a 1×1 convolution kernel is appended at the end of the last stage. The final feature representation for the hazy image can thus be given as

$$X_{share} = \delta(\xi(W_{3\times3} * X_{16})). \tag{11}$$

This sub-network fuses together differing semantic knowledge from multi-scale receptive fields via the holistically nested architecture (i.e., these separate pipelines), which leverages more context.

Meanwhile, the computational complexity of our model can be significantly decreased using the inception structure as a basic block. The time and spatial complexities of one convolution are well known to be described as $O(wh \cdot r^2 \cdot n_{in} \cdot n_{out})$ and $O(r^2 \cdot n_{in} \cdot n_{out} + wh \cdot n_{out})$, respectively, where wh is the spatial dimension and n_{in} and n_{out} denote the numbers of input and output maps, respectively. An example in Table 1 clearly compares an inception structure with its naive structure. Besides the additional 1×1 convolution, the other convolutions in an inception structure have the same setting as those in the naive structure. When the input and output are fixed, the computational complexity of the naive structure is roughly seven times as expensive as that of the inception structure. The main reason of complexity reduction in inception structures is that the additional 1×1 convolution scales down the input channels of the 3×3, 5×5, and 11×11 convolutions. Through stacking several similar inception blocks, the time and spatial complexities of this holistically nested architecture are 28 billion floating-point operations in number of multiply–adds (FLOPS) and 5.6M parameters, respectively.

3.2. Sub-network for cascaded multi-task

Single image haze removal is an ill-posed problem since a haze-free scenario may be connected to multifarious hazy environments. Moreover, when presented with heavy haze, this connection will be more ambiguous since the fine details of scene have little evidence in the corresponding hazy versions. According to Eq. (2), this problem can become well posed as long as the relevant spatial variables have been acquired. Hence, we present a cascaded sub-network that can restore the scene radiance using Eq. (2). As illustrated by the gray framework in Fig. 1, this sub-network consists of several task-dependent branches. By taking the shareable feature X_{share} as input, the prediction of the first branch can be formulated as

$$D = \xi(W_{r\times r} * X_{share}), \tag{12}$$

Table 1
Comparison between an inception structure and its naive structure. The input size and output size are represented as *width* × *height* × *channel*. The inception adopts the structure shown by Fig. 2(b), and $r_1 = 3$, $r_2 = 5$, and $r_3 = 11$. *params* denotes the spatial complexity in terms of the parameter size, and *ops* is the time complexity in terms of the floating-point operation. M and B are the abbreviations for million and billion, respectively.

Structure	Input size	#1 × 1	#1 × 1 reduce	#3 × 3	#1 × 1 reduce	#3 × 3	#1 × 1 reduce	#11 × 11	Output size	*params*	*ops*
Naive	52 × 52 × 256	32	–	64	–	64	–	64	52 × 52 × 224	2.5M	6.8B
Inception	52 × 52 × 256	32	32	64	32	64	32	64	52 × 52 × 224	0.35M	0.95B

where D is the scene depth map with single channel, W is to match the depth in practice, $r \times r$ indicates the dimension of each kernel in this branch, and $\xi(\cdot)$ limits the output in the range of $[0, \infty)$. Subsequently, X_{share} and D are simultaneously propagated into the next branch that outputs the atmosphere scattering coefficient as

$$B = \xi(W_{r\times r} * [D, X_{share}]), \tag{13}$$

where B is the scattering coefficient map, i.e., $B(x)$ is the single scattering value at pixel x, and the concatenation $[\cdot]$ provides diverse-level inputs that can facilitate the position-wise alignment between variables, i.e., D and the scattering coefficient map B. Without the assumption of homogeneous atmosphere in local or whole observation, this branch aims at a random distribution of the haze thickness. Consequently, the output B can present a more natural visibility degradation that tends to vary at different positions. Based on the outputs of the first two branches, the medium transmission map can be obtained as

$$T = e^{(-B\otimes D)}, \tag{14}$$

where $\otimes$ denotes the element-wise multiplication and Eq. (3) is used as the nonlinear activation function. The relationship between T and the two spatial variants, namely B and D, is then developed. Subsequently, the following branch focuses on the estimation of the atmospheric light. In particular, a three-channel map is first calculated by

$$A^c = \xi(W_{r\times r} * [T, X_0, X_{share}]), c \in \{red, green, blue\} \tag{15}$$

where A^c is a set of candidates for the atmospheric light value in the c channel and X_0 is the hazy observation. The final output of this branch is then the pooled response to each map in A, i.e.,

$$\alpha^c = \frac{1}{wh}\sum_x A^c(x). \tag{16}$$

The 3-dimensional vector $\alpha = (\alpha^{red}, \alpha^{green}, \alpha^{blue})$ is the estimated global atmospheric light across all color channels. With the intermediate results T and α, this sub-network can recover a clean image through

$$J = \frac{X_0 - \alpha}{T} + \alpha, \tag{17}$$

where the activation function roots from Eq. (2). Finally, to further correct the color shift, a white balancing layer is proposed as

$$wb(J) = f(BReLU(J)) \tag{18}$$

where $BReLU$ is a bilateral restraint function: $J(x) \rightarrow \min(\max(J(x), V_{\min}), V_{\max})$. We can use this function to restrict pixel values into the range of $[V_{\min}, V_{\max}]$. As mentioned in Section 2.1, $V_{\min}$ and $V_{\max}$ are the lower and upper bounds derived from the sorted pixels of J, and $f(\cdot)$ is the affine transformation for color correction. Similar to Eqs. (14) and (17), $wb(\cdot)$ is a specialized activation function that can represent inherent characteristics of the relationships in the current task. Finally, $wb(J)$ is the haze-free output of this sub-network.

3.3. Optimization of the proposed network

As discussed above, the proposed de-hazing model is a multi-task learning system. However, it is difficult to simultaneously learn related but different tasks. For pragmatic training, a three-stage alternating optimization is adopted. In the first stage, the model is initialized with random parameters; we then train the shareable sub-network and the branch of D by minimizing the following loss.

$$L_D = l_\partial(D, D') + l_G(D, D') \tag{19}$$

where $D\prime$ is the expected scene depth, $l_\partial(\cdot)$ is an empirical criterion to adjust the prediction result, and $l_G(\cdot)$ is an edge-preserving regularization. In particular, $l_\partial(\cdot)$ is the scale-invariant loss that can be defined as

$$l_\partial(D, D') = \frac{1}{(nwh)^2}\sum_{x,y}((\log D(x) - \log D(y)) - (\log D'(x) - \log D'(y)))^2 = \frac{1}{nwh}\sum_x \partial_x^2 - \frac{1}{(nwh)^2}\sum_{x,y}\partial_x\partial_y, \tag{20}$$

where n is the number of channels and $\partial_x = \log D(x) - \log D'(x)$ is the difference in log space. From the first view, $l_\partial(\cdot)$ ensures that the relation between a pair of elements in D is the same as that of the pair at the corresponding locations in D'. In contrast, it is difficult to predict D in an element-wise way, thus Eq. (20) allows a set of solutions with $l_\partial(\cdot) = 0$ in which case, ∂_x becomes the constant $\hat{\partial}$ and $D = e^{\hat{\partial}}D'$.

Since the regularization term is to penalize higher loss in the edge regions, $l_G(\cdot)$ can be defined as

$$l_G(D, D') = \frac{1}{nwh}(\|G_v(D) - G_v(D')\|_F + \|G_h(D) - G_h(D')\|_F), \tag{21}$$

where G_v and G_h are the gradient operators along vertical and horizontal directions, and F is the Frobenius norm. Under the constraint of $l_G(\cdot)$, several details of object edges with higher gradients can be retained.

In the second stage, at early steps we truncate the gradient propagation of the depth branch and train the next branch of scattering coefficient via the optimization item

$$L_T = l_\theta(T, T') + l_G(T, T'). \tag{22}$$

The empirical criterion $l_\theta(\cdot)$ focuses on structural differences from three aspects [38]: Luminance (U), Contrast (C), and Structure (S), which can be represented as follows:

$$U(T,T') = \frac{2\mu_T\mu_{T'} + \varepsilon_1}{\mu_T^2 + \mu_{T'}^2 + \varepsilon_1}; \quad C(T,T') = \frac{2\sigma_T\sigma_{T'} + \varepsilon_2}{\sigma_T^2 + \sigma_{T'}^2 + \varepsilon_2}; \quad S(T,T') = \frac{\sigma_{TT'} + \varepsilon_3}{\sigma_T\sigma_{T'} + \varepsilon_3}, \tag{23}$$

where μ_T and $\mu_{T'}$ are the mean signal intensities, σ_T and $\sigma_{T'}$ are the standard deviations, and $\sigma_{TT'}$ refers to the covariance. The non-negative constants ε_1, ε_2, and ε_3 are included to avoid division by zero, and we set $\varepsilon_2 = 2\varepsilon_3$. Thus, the comprehensive empirical criterion (illustrated in Fig. 4) can be defined as

$$l_\theta(T,T') = 1 - U(T,T')C(T,T')S(T,T') = 1 - \frac{(2\mu_T\mu_{T'} + \varepsilon_1)(2\sigma_T\sigma_{T'} + \varepsilon_2)}{(\mu_T^2 + \mu_{T'}^2 + \varepsilon_1)(\sigma_T^2 + \sigma_{T'}^2 + \varepsilon_2)}. \tag{24}$$

Evidently, B's behavior can tend toward the unknown scattering coefficient map when T and D are perfect. Thus, the learning driven by Eqs. (19) and (22) facilitates the unsupervised optimization of B. Moreover, in the case of $\hat{\partial} \neq 0, B$ can be a scalar, which equals to the true scattering coefficient map multiplied by $e^{-\hat{\partial}}$. Thus, the unsupervised optimization of B can compensate the imperfect cases in Eq. (20). The final step in this stage then aims to fine-tune these layers including shareable convolutions: branches of D and B.

In the last stage, we first freeze the branches with respect to D and B, and just optimize the branch of α, which can be learned in the manner similar to the learning of B when minimizing the distance between J and the ground-truth image J', i.e.,

$$L_J = l_\theta(J, J') + l_G(J, J'). \tag{25}$$

We then develop a cyclic restoration loss to reinforce the performance of the joint optimization regarding the three abovementioned branches, such loss can be described as

$$L_I = l_\theta(J'T + \alpha(1 - T), X_0). \tag{26}$$

Further, the final output $wb(J)$ is optimized by a perceptual loss, which is a popular approach to maintain the salient features of an image. A comparison is shown in Fig. 5, where the differences in detail are significant. Concretely, this perceptual discrepancy can be described as

$$L_P = l_\rho(wb(J), J') \tag{27}$$

where the error metric $l_\rho(\cdot)$ is

$$l_\rho(J, J') = \frac{1}{nwh}\|\varphi_j(J) - \varphi_j(J')\|_F. \tag{28}$$

Here, $\varphi_j(J)$ is the output of the jth layer in the perceptual network φ for J. Same as the work proposed by Johnson et al. [19], the perceptual network is the VGG net [34] pre-trained on the ImageNet dataset.

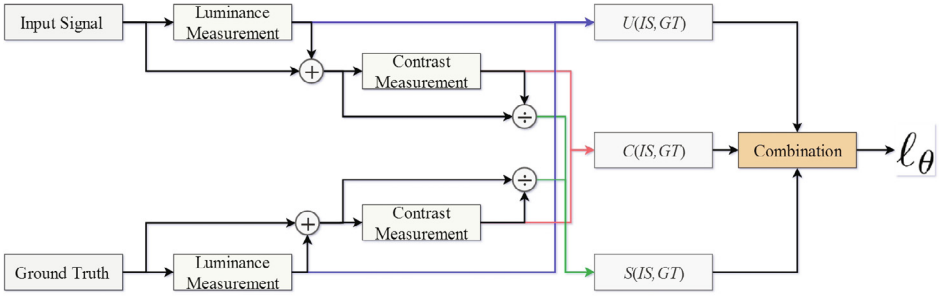


Fig. 4. Diagram of the image characteristic measurement. The restrictive signal IS and its ground truth GT are taken part in the calculation of structural differences $l_\theta(\cdot)$ with respect to the luminance $U(IS, GT)$, contrast $C(IS, GT)$, and structure $S(IS, GT)$.

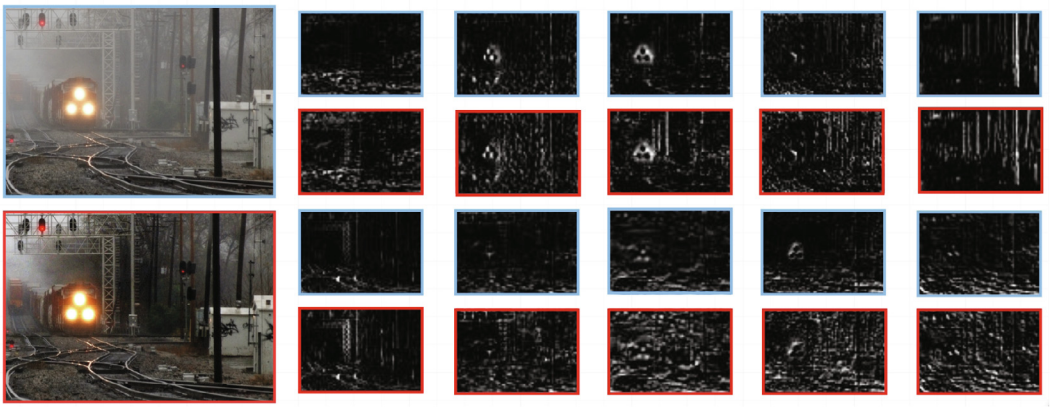


Fig. 5. Comparison of salient perceptual information. Left: the hazy image (in blue frame) and its de-hazed result (in red frame). Right: the feature maps of the hazy image (in blue frames) and the corresponding maps of the de-hazed image (in red frames). All features are extracted from the pool3 layer of the trained VGG network. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

4. Experiments

In this section, comprehensive experiments conducted on artificial and natural hazy images are discussed. The optimization method for the proposed connectionist model is a momentum optimizer, with the momentum set as 0.9, the initial learning rate as 0.0125, and the decay factor as 0.1. Moreover, the saturation levels s_1 and s_2 are set to 0.5%. The implementation of the proposed de-hazing model is conducted on Python3.5, TensorFlow1.8, GeForce GTX TITAN with 12 GB RAM.

4.1. Training data preparation

Since pairs of haze and haze-free images from natural scenes are not extensively available, it is necessary to synthesize hazy images to train the deep neural model. To this end, we developed our training data based on a depth image dataset named MegaDepth [22], which contains several color images and depth map pairs. The authors of MegaDepth employed a few effective techniques, e.g., data cleaning and semantic segmentation, to address the common problems caused by noise and unreconstructable objects. Thus, the examples in MegaDepth have high-quality depth information. A few examples are shown in Fig. 6. Based on the depth maps, we generate twenty different transmission maps by using Eq. (3) in which β is a random value within [0.1, 0.9]. The corresponding hazy images are then synthesized via Eq. (1) in which α is a random value within [0.5, 1.1]. To augment the size of the training data, five patches are cropped from each hazy image along with the corresponding ground truth, depth, and transmission. As a result, there are a total of 349,200 cells $(I,\ J,\ D,\ T)$ in the training dataset. We demonstrate the training loss of independent stages in Fig. 7. Through the mean and variance of multiple independent repetitions, one can note that the model is not sensitive to the initialization.

Fig. 6. Images with their depth maps. The top images are used in the training dataset, and the bottom pairs are some examples from the evaluation dataset. For all depth maps, the regions close to camera are marked in blue, whereas the yellow regions are far from the camera. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

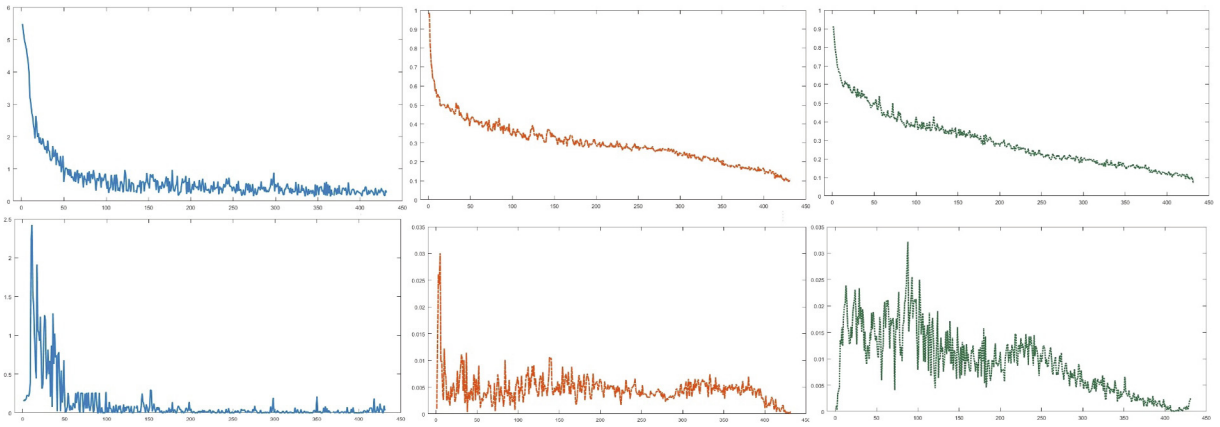


Fig. 7. Average training loss and variance of independent stages in the alternating optimization. Each optimization is repeated by adopting three different initializations. The top lines are training loss curves, and the bottom lines reflect the variance. The blue line is the first stage for depth estimation, the orange line is the second stage training process of the transmission map, and the green line is the last stage for image restoration. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

4.2. Evaluation on synthetic dataset

In this part, the proposed de-hazing model is compared with several state-of-the-art haze removal methods that we denote as DCP [13], CAP [45], ATM [35], DHN [2], DCPDN [43], and GFN [31]. We evaluate these methods on a synthetic dataset generated from the NYU-depth2 dataset [33] including 1449 stereo images (see Fig. 6 for some examples).

To quantitatively assess these de-hazed results, three evaluation criteria, i.e., structural similarity (SSIM) [38], peak signal-to-noise ratio (PSNR) [16], and feature similarity (FSIMc) [44], are employed to measure the difference between the de-hazed result and the corresponding ground truth image. For SSIM and FSIMc, a value close to 1.0 indicates a perfect performance, as for PSNR, a higher value means a better image reconstruction result. The comparison results are reported in Table 2. Our method can obtain higher SSIM and PSNR, demonstrating that it achieves favorable characteristic-level and pixel-level similarities. Regarding FSIMc, DHN obtains the best score, while the results of our method, DCPDN, and GFN are similar. Higher results on FSIMc indicate that data-driven models can retain crucial information regarding salient features reflecting human perception. In addition, we compared this model with its simplified version (denoted as Simple in Table 2. Relative to the original model, the representative learning network in the simplified model has fewer layers. To this end, one convolution block is cut in each network stage, and in the end, eight blocks are retained. This comparison demonstrates that the deeper network can achieve a better performance. The de-hazing work can be considered as image restoration, which is an ill-posed problem; multiple complex non-linear regressions are conducted in our model. To fit these tasks, a deep network is necessary, because a deep neural network tends to have a better ability to mutually promote the learning of such regression tasks, and we use the lateral connection to compensate the spatial loss in deep features. In any case, this connection is similar to the skip connection in the residual structure, which can facilitate gradient propagation from deep layers to shallow layers. Therefore, when going deeper, the network with such a connection can avoid some factors leading to performance decline.

4.3. Evaluation on natural hazy images

In this section, we will investigate the performances of the aforementioned methods on different practical hazy scenes, which are obtained from existing publications. The visual comparisons of different de-hazed results are shown in Fig. 9. DCP achieves several appealing results; however, it tends to poorly handle the sky region and certain scene objects (see b-9, b-16, and b-17). The reason is that DCP may underestimate the transmission when the brightness of such a region is similar to the atmospheric light. Moreover, a halo is observed at the position with depth change (see b-11) since the estimation in a local

Table 2
Comparison of evaluation datasets. Three full reference assessments, i.e., SSIM, PSNR, and FSIMc, are used to measure the performances of different de-hazing methods. Each value is the average result of 1449 pairs of de-hazed/ground truth images.

Methods	ATM	CAP	DCP	DHN	DCPDN	GFN	Simple	Ours
SSIM	0.6426	0.6616	0.7238	0.7221	0.7082	0.7465	0.7399	**0.7604**
PSNR	13.92	16.40	17.61	15.79	14.99	17.45	16.31	**17.95**
FSIMc	0.8598	0.8953	0.8978	**0.9147**	0.9038	0.9013	0.9033	0.9140

For each assessment measure, the bold value represesnts the best performance.

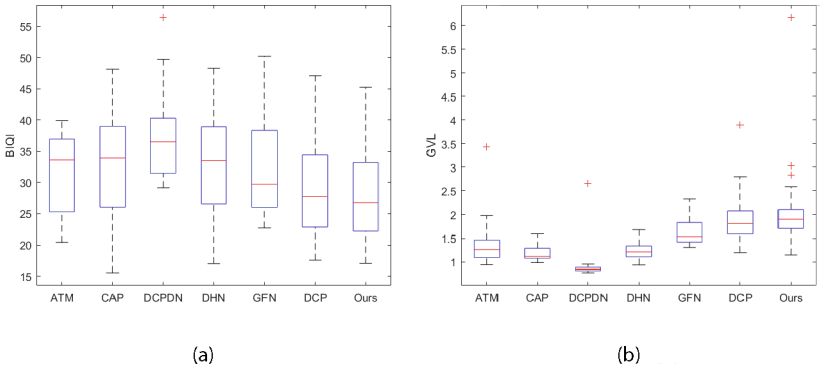


Fig. 8. Box plot of BIQI and GVL. The red line is the median value, the horizontal edges of the blue box are the 25th and 75th percentiles, the whiskers extend to the most extreme data points, and the outliers are plotted individually. All abbreviations of de-hazing methods are list at the lateral axis. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

image patch is prone to be the transmission value of the near objects. Compared with DCP, the proposed method can retain the characteristics of the sky region. This advantage can be attributed to the optimization tasks regarding structural differences and perceptual loss. The structural differences can effectively reflect the visual deviation through pixel-level statistical characteristics. In contrast, the perceptual optimization inherently emphasizes semantic and texture similarities based on high-level feature representations. Under the joint constraint of structural and perceptual differences, when image distortion happens, visibility errors will be captured, and a larger error will penalize such failure cases. According to the principle of minimizing the penalty value, we can obtain an error-sensitive model that encourages reliable reconstruction in feature and pixel levels. In addition, the proposed method is robustly resistant to drastic changes of scene distance. This advantage mainly depends on the loss term in Eq. (19) that makes the de-hazing model aware of depth. Furthermore, our model conducts the prediction in an element-wise manner without the assumption of a homogeneous atmosphere in a local image patch. Thus, it is unnecessary to share the same prediction result with adjacent pixels. Similarly, CAP is an advanced model for depth estimation; however, its robustness is generally affected by a manually obtained value of the scattering coefficient in implementation. For ATM, some results are over-saturated, and it is sensitive to the white region (see c-16). With regard to deep learning methods (i.e., DHN, GFN, DCPDN, and Ours), our model exhibits a more robust performance in most cases. This advantage mainly benefits from the multi-task learning that has been proven to be effective in several computer vision problems. In particular, the proposed method simulates the atmospheric model to restore the haze-free image based on some crucial relevant variables. To this end, we adopt the three-stage alternating optimization, and each stage facilitates its following optimization task. Thus, a complex non-linear regression can be implemented in a task-wise manner, and in the joint optimization, multiple tasks can be learned in a mutually reinforced manner. In addition, another benefit comes from the representative network that can enhance the coarser-resolution maps via lateral connections; consequently, high-quality features can be utilized by the multi-task learning. Finally, through visual inspection, one can note that the proposed model achieves better results in terms of color correction (see h22-26), which benefits from the design of the white balancing layer.

To quantify human visual perception regarding these de-hazing results, blind image quality index (BIQI) [25] is used as a reference-free criterion. The lowest score of BIQI stands for the best visual effect of an image without blockness, blur, corner outliers, and noise. The quantitative comparisons are reported in Fig. 8 and Table 3. The proposed method and DCP obtain superior scores, whereas DCP has higher median and mean values than our method due to the distortion on sky region and the failure on color correction.

Furthermore, we employ the geometric mean ratios of visibility level (GVL) [12] to assess the periodic patterns and textures, which convey crucial visibility information. Theoretically, a higher GVL reflects that more details have been identified. Fig. 8 and Table 3 report the results. Compared with other de-hazing methods, our model obtains a better GVL score, which indicates a favorable performance on image visibility improvement. Besides the abovementioned perceptual loss and structural differences, the edge-aware regularizer plays an important role in this advantage. This regularizer defines the average gradient distance that can represent the detail loss in the restored image. By minimizing such loss, an edge-preserving technique can be developed in our de-hazing model.

5. Conclusions

In this paper, we present a single image haze removal model based on the deep joint neural network, in which different but related vision tasks are sequentially performed to realize de-hazing. Two sub-networks of our model make significant contributions to representative and task-driven learning. Representative learning is realized using the first sub-network with a holistically nested structure, and the learned features are then shared by the following haze-relevant tasks (i.e., estimations of scene depth, atmosphere scattering coefficient, medium transmission, restoration of the haze-free image, and correction

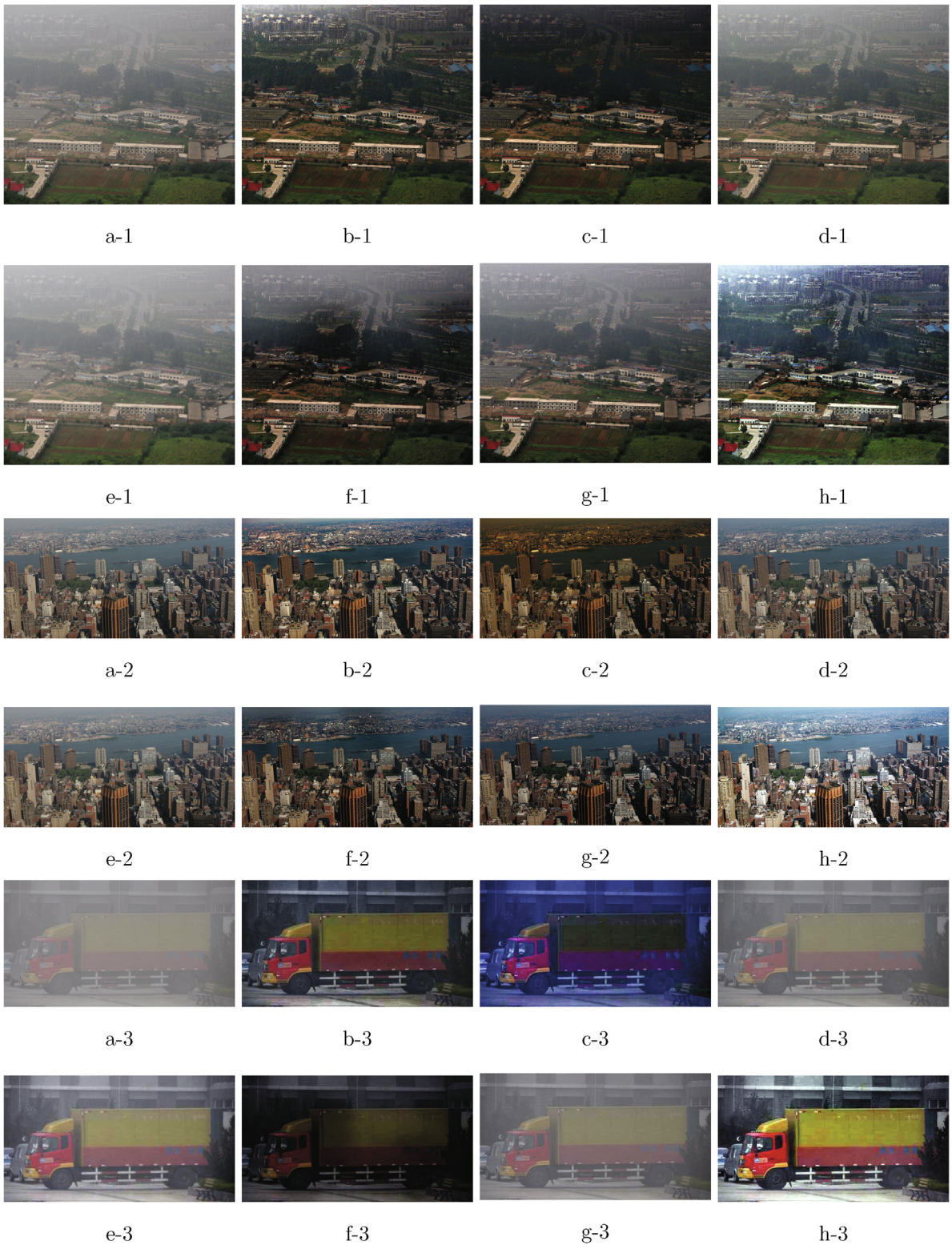


Fig. 9. De-hazed results on natural images: (a) hazy image, (b) DCP, (c) ATM, (d) CAP, (e) DCPDN, (f) GFN, (g) DHN and (h) Ours.

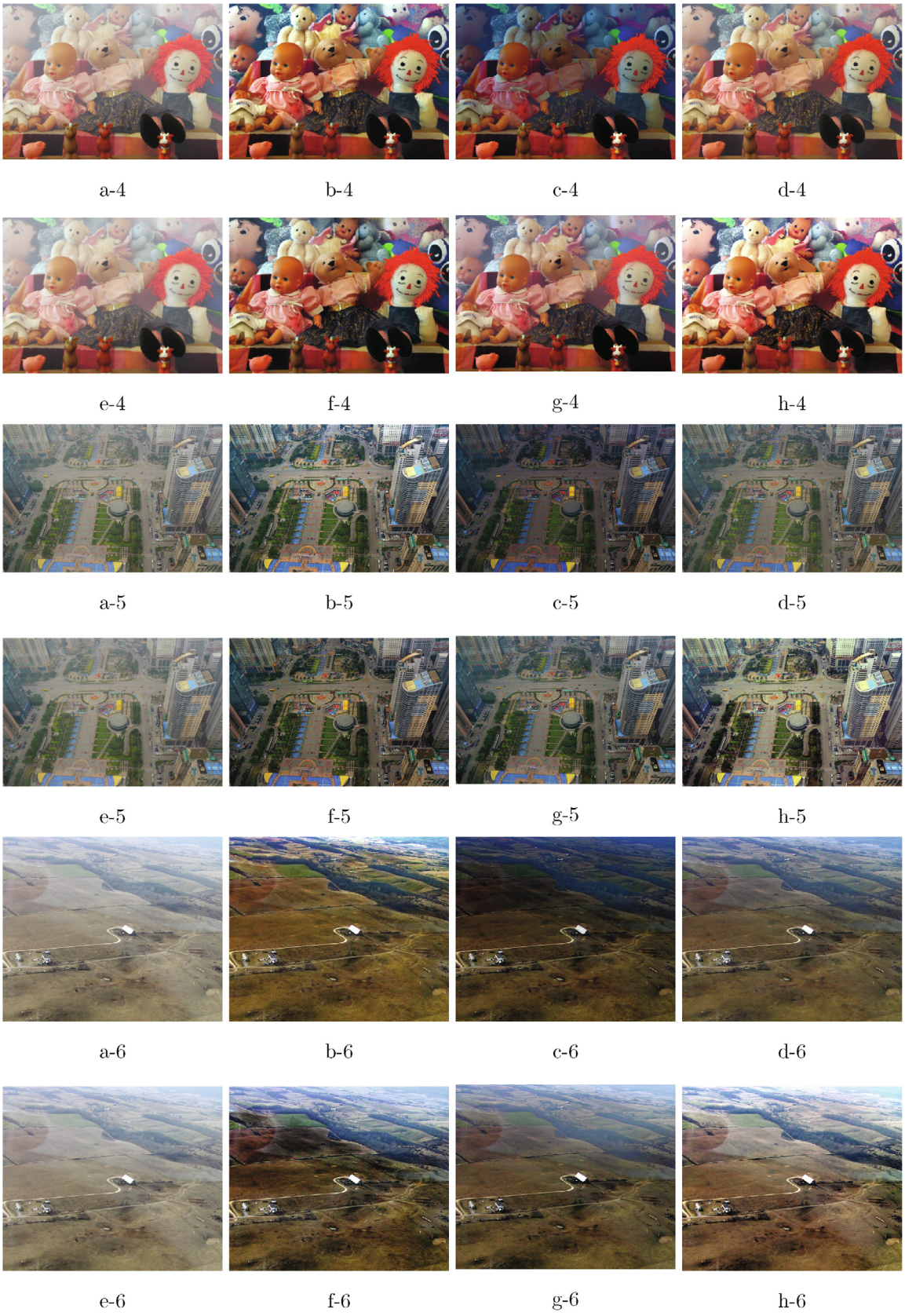


Fig. 9 (*continued*)

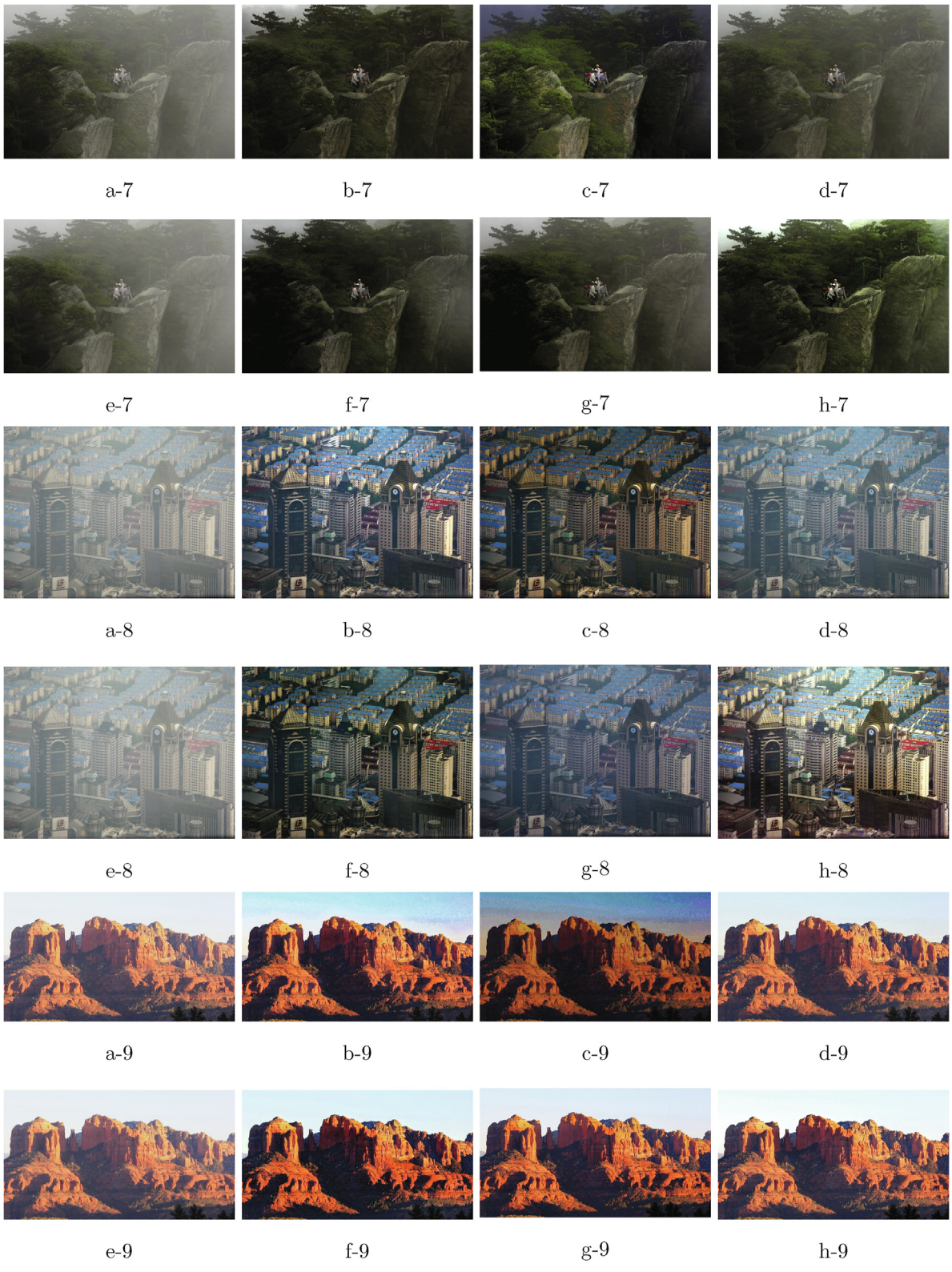


Fig. 9 (*continued*)

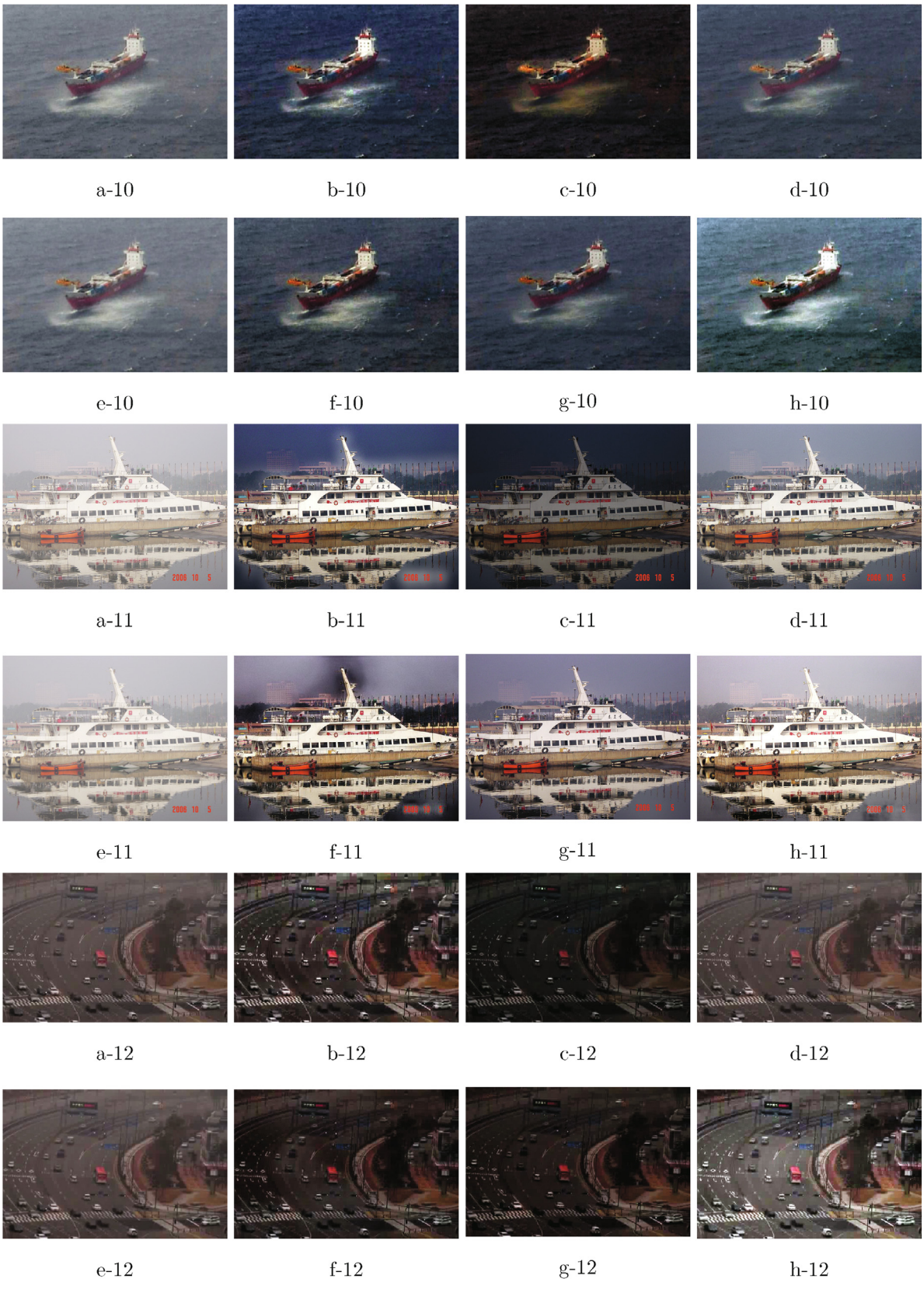


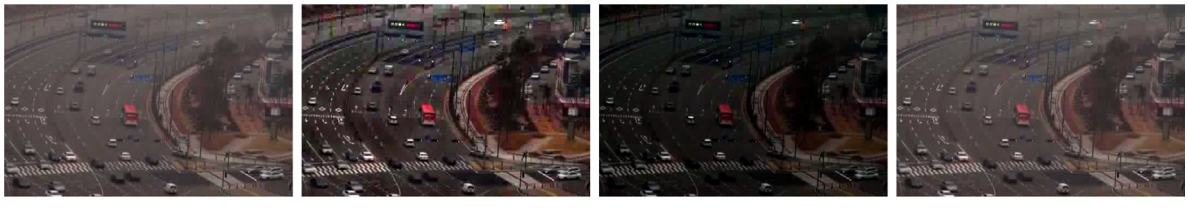
Fig. 9 (*continued*)

a-10 b-10 c-10 d-10

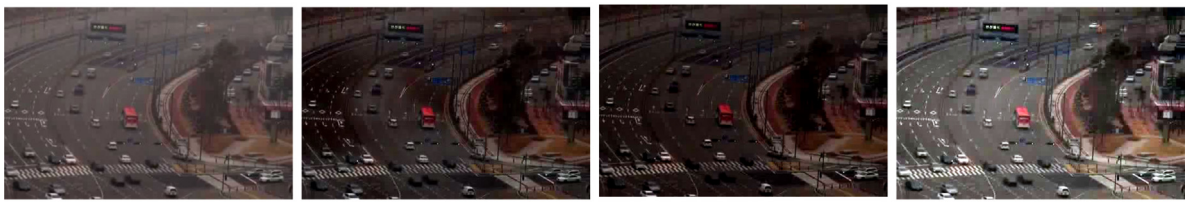
e-10 f-10 g-10 h-10

a-11 b-11 c-11 d-11

e-11 f-11 g-11 h-11



a-12 b-12 c-12 d-12



e-12 f-12 g-12 h-12

Fig. 9 (*continued*)

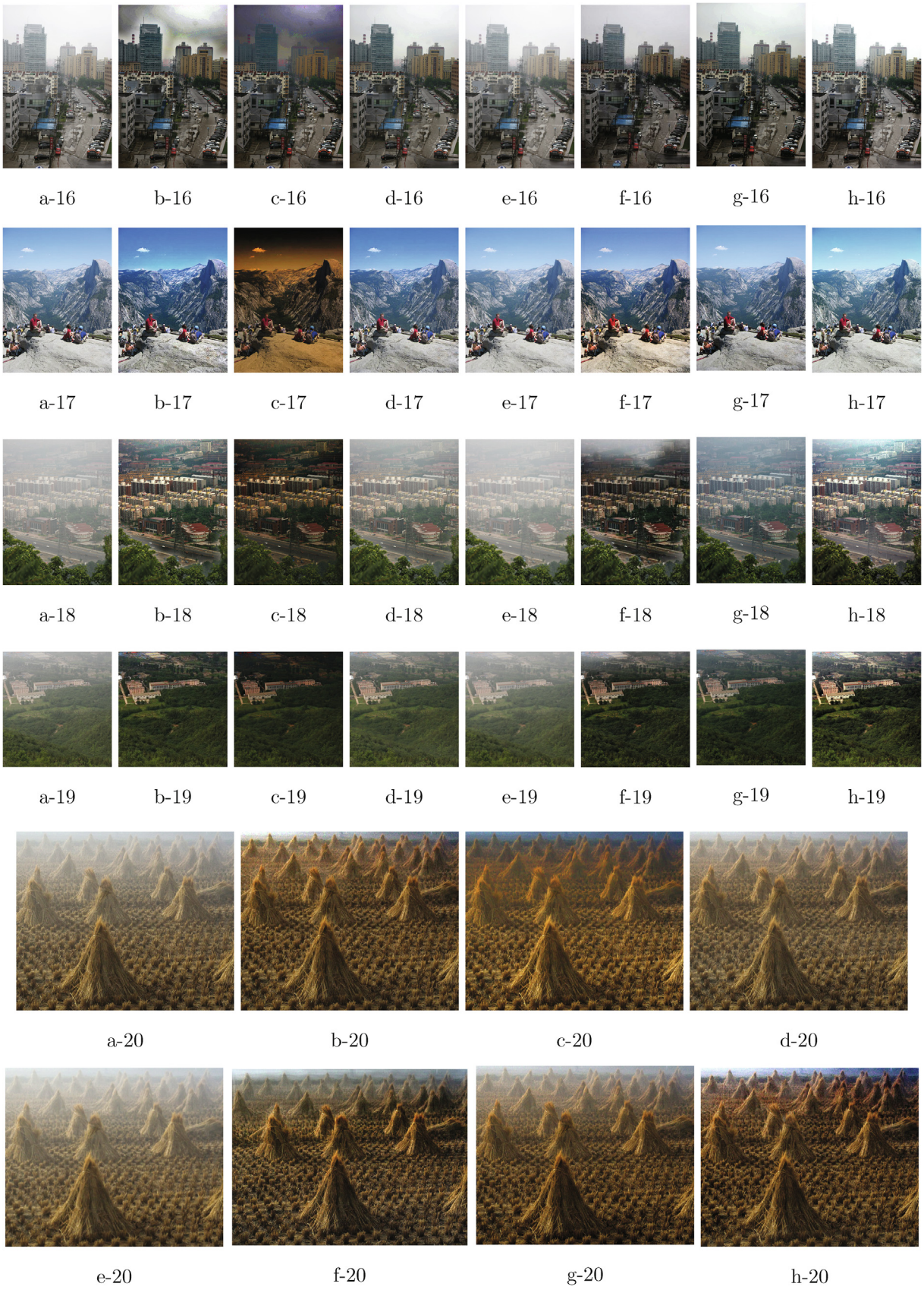


Fig. 9 (*continued*)

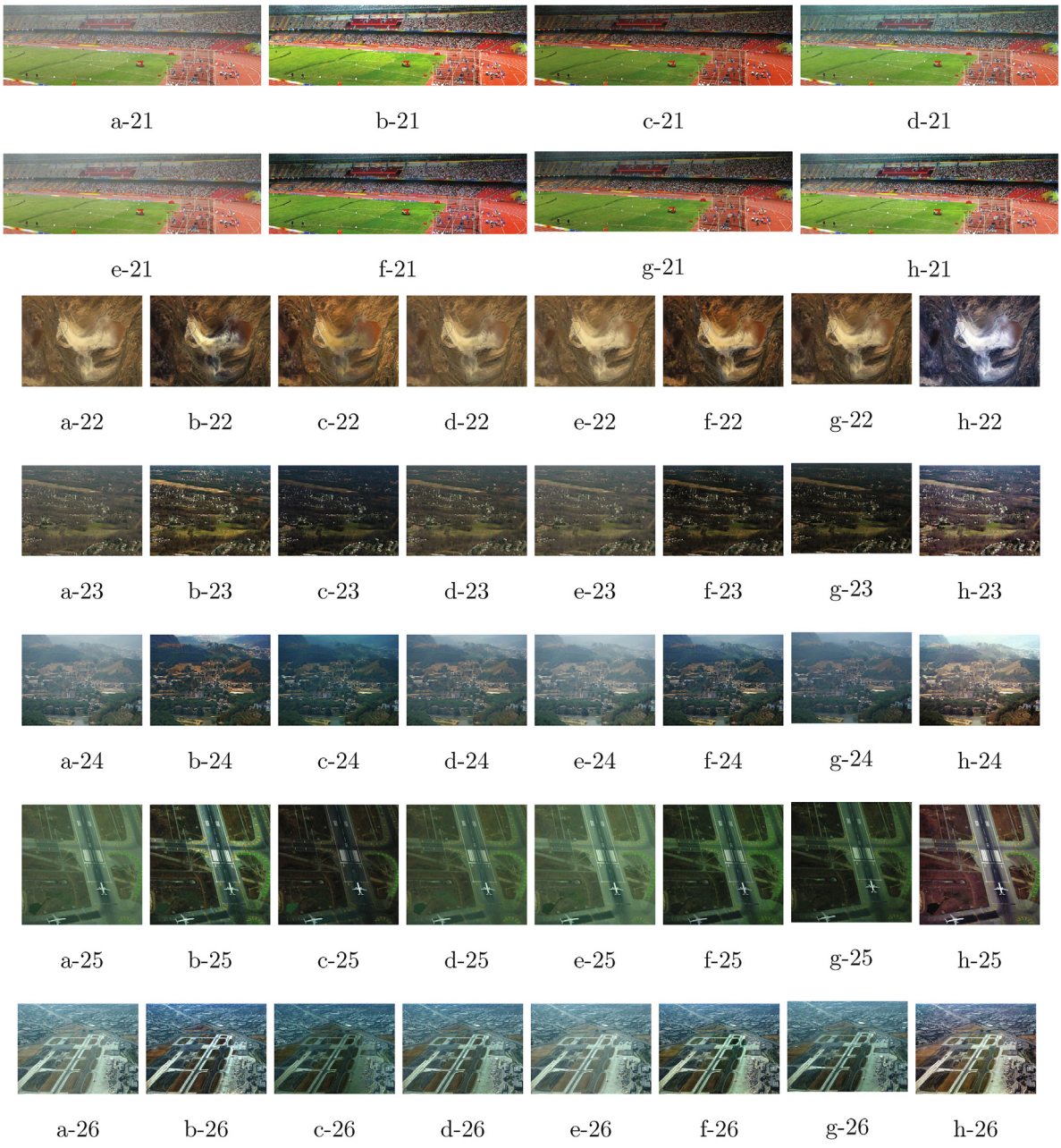


Fig. 9 (*continued*)

Table 3
Comparison of natural hazy images. Two no reference assessments, i.e., BIQI and GVL, are used to reflect the image quality. Each value is the average result on 26 de-hazed images. Typically, a higher GVL and lower BIQI indicate a better de-hazed result.

Methods	ATM	CAP	DCP	DHN	DCPDN	GFN	Ours
BIQI	31.37	32.60	29.01	31.62	37.46	31.99	**27.98**
GVL	1.368	1.187	1.928	1.232	0.917	1.671	**2.070**

For each assessment measure, the bold value represesnts the best performance.

of shifted colors), which are implemented using cascaded branches and task-specific activation functions in the second sub-network. To optimize the proposed model, a three-stage training technique is adopted for the multi-task learning, and different learning patterns are incorporated into this technique. In addition, a cyclic restoration loss is designed to reinforce the

joint optimization. Comprehensive experiments have been conducted on synthetic and natural hazy images, and the de-hazed results are evaluated by several popular full-reference and no-reference assessments. By comparing state-of-the-art de-hazing methods and the proposed model, the latter is revealed to have impressive vision appeal in most cases and the effectiveness of this model is verified. In the future, we will focus on combining the proposed de-hazing method with other computer vision applications, such as object detection and semantic segmentation, which have a strong demand for images with high visual quality.

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.

CRediT authorship contribution statement

Tianlun Zhang: Conceptualization, Methodology, Software, Validation, Writing - original draft. **Xi Yang:** Data curation, Software, Formal analysis, Investigation. **Xizhao Wang:** Supervision, Project administration, Funding acquisition. **Ran Wang:** Visualization, Validation, Writing - review & editing, Funding acquisition.

Acknowledgment

This work was supported in part by the National Natural Science Foundation of China (Grants 61772344, 61976141, 61732011, 61811530324, and 61402460), in part by the Natural Science Foundation of SZU (Grants 827-000140 and 827-000230), and in part by the Interdisciplinary Innovation Team of Shenzhen University.

References

[1] D. Berman, T. Treibitz, S. Avidan. Air-light estimation using hazelines. in: Proceedings of IEEE Int. Conf. Comput. Photogr., 2017, pp. 1–9..
[2] B.L. Cai, X. Xu, K. Jia, C. Qing, D. Tao, Dehazenet: an end-to-end system for single image haze removal, IEEE Trans. Image Process. 25 (2016) 5187–5198.
[3] R. Caruana, Multitask learning, Mach. Learn. 28 (1) (1997) 41–75.
[4] Yann Le Cun, L. Bottou, Y. Bengio. Reading checks with multilayer graph transformer networks, in: IEEE International Conference on Acoustics, Speech, and Signal Processing, 1997, pp. 21–24..
[5] S.Q. Duntley, A.R. Boileau, R.W. Preisendorfer, Image transmission by the troposphere I, J. Opt. Soc. Am. 47 (1957) 499–506.
[6] M. Ebner. Color Constancy, first ed., Wiley, 2007..
[7] R. Fattal, Single image dehazing, ACM Trans. Graphics 27 (2008) 1–9.
[8] G.D. Finlayson, E. Trezzi, Shades of gray and color constancy, in: Proceedings of Color Imaging Conference: Color Science, System and Applications, Society for Imaging Science and Technology, 2004.
[9] A. Gijsenij, T. Gevers, Color constancy using natural image statistics and scene semantics, IEEE Trans. Pattern. Anal. Mach. Intell. 33 (2011) 687–698.
[10] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, and B. Xu. Generative adversarial nets. in: Proceedings of the 27th International Conference on Neural Information Processing Systems, vol. 2, 2014, pp. 2672–2680..
[11] K. Gu, G. Zhai, W. Lin, M. Liu, The analysis of image contrast: From quality assessment to automatic enhancement, IEEE Trans. Cybernet. 46 (1) (2016) 284–292.
[12] N. Hautiere, J.P. Tarel, et al., Blind contrast enhancement assessment by gradient rationing at visible edges. Image Anal. Stereol., 27 (2008) 87–95..
[13] K.M. He, J. Sun, X.O. Tang, Single image haze removal using dark channel prior, IEEE Trans. Pattern Anal. Mach. Intell. 33 (12) (2011) 2341–2353.
[14] K.M. He, J. Sun, X.O. Tang, Guided image filtering, IEEE Trans. Pattern Anal. Mach. Intell. 6 (2013) 1397–1409.
[15] G. Huang, Z. Liu, L.V.D. Maaten, et al, Densely connected convolutional networks, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2017.
[16] Q. Huynh-Thu, M. Ghanbari, Scope of validity of psnr in image/video quality assessment, Electron. Lett. 44 (2008) 800, 601.
[17] S. Ioffe, C. Szegedy, Batch normalization: accelerating deep network training by reducing internal covariate shift, arXiv preprint arXiv:1502.03167, 2015..
[18] A. Gijsenij, J. van de Weijer, T. Gevers, Edge based color constancy, IEEE Trans. Image Process. 16 (2007) 2207–2214..
[19] J. Johnson, A. Alahi, L. Feifei, Perceptual losses for real-time style transfer and super-resolution, in: Proceedings of European Conference on Computer Vision, 2016, pp. 694–711..
[20] J. Kopf, B. Neubert, B. Chen, M. Cohen, D. Cohen-Or, O. Deussen, M. Uyttendaele, D. Lischinski, Deep photo: Model-based photograph enhancement and viewing, ACM Trans. Graph 27 (5) (2008) 1–10.
[21] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, in: Proceedings of the Advances in Neural Information Processing Systems, 2012..
[22] Z.Q. Li, N. Snavely, Megadepth: learning single-view depth prediction from internet photos, arXiv:1804.00607, 2018..
[23] N. Limare, J. Lisani, J. Morel, A. Petro, C. Sbert, Simplest color balance, Image Process. On Line 1 (2011) 297–315.
[24] J.-E. McDonald, S.K. Nayar, The saturation adjustment in numerical modelling of fog, J. Atmos. Sci. 20 (5) (1963) 476–478.
[25] A.K. Moorthy, A.C. Bovik, A two-step framework for constructing blind image quality indices, IEEE Signal Process. Lett. 17 (2010) 513–516.
[26] V. Nair, G.E. Hinton, Rectified linear units improve restricted Boltzmann machines, in: Proceedings of International Conference on Machine Learning, 2010, pp. 807–814..
[27] S.G. Narasimhan, S.K. Nayar, Contrast restoration of weather degraded images, IEEE Trans. Pattern Anal. Mach. Intell. 25 (6) (2003) 713–724.
[28] S.K. Nayar, S.G. Narasimhan, Vision in bad weather, in: Proceedings of IEEE International Conference on Computer Vision, 1999, pp. 820–827..
[29] D. Ren, W. Zuo, Q. Hu, P. Zhu, D.Y. Meng, Progressive image deraining networks: a better and simpler baseline, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019.
[30] S.Q. Ren, K.M. He, R. Girshick, J. Sun, Faster r-cnn: towards real-time object detection with region proposal networks, IEEE Trans. Pattern Anal. Mach. Intell. 39 (6) (2017) 1137–1149.
[31] W. Ren, L. Ma, J. Zhang, J. Pan, X. Cao, W. Liu, M.-H. Yang, Gated fusion network for single image dehazing, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2018.

[32] Y.Y. Schechner, S.G. Narasimhan, S.K. Nayar, Instant dehazing of images using polarization, in: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2001, pp. 8–14..
[33] N. Silberman, D. Hoiem, P. Kohli, R. Fergus, Indoor segmentation and support inference from rgbd images, in: Proceedings of European Conference on Computer Vision, 2012, pp. 746–760..
[34] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, Online: arXiv:1409.1556, 2014..
[35] M. Sulami, I. Glatzer, R. Fattal, M. Werman, Automatic recovery of the atmospheric light in hazy images, in: Proceedings of IEEE Int. Conf. Comput. Photogr., 2014, pp. 1–11..
[36] C. Szegedy, W. Liu, et al, Going deeper with convolutions, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015.
[37] C. Szegedy, S. Loffe, et al., Inception-v4, inception-resnet and the impact of residual connections on learning, Online: arXiv:1602.07261, 2016..
[38] Z. Wang, A.C. Bovik, H.R. Sheikh, E.P. Simoncelli, Image quality assessment: from error visibility to structural similarity, IEEE Trans. Image Process. 13 (2004) 600–612.
[39] Y. Xie, M.G. Gong, A.K. Qin, et al, Tpne: Topology preserving network embedding, Inform. Sci. 504 (2019) 20–31.
[40] Y. Xie, M.G. Gong, S. Wang, et al, Sim2vec: node similarity preserving network embedding, Inform. Sci. 495 (2019) 37–51.
[41] L. Yang, M. Zhang, Y. Liu, M.S. Sun, et al, Joint pos tagging and dependence parsing with transition-based neural networks, IEEE/ACM Trans. Audio Speech Language Process. 26 (8) (2018) 1352–1358.
[42] F. Yu, V. Koltun, Multi-scale context aggregation with dilated convolutions, in: Proceedings of the International Conference on Learning Representations, 2016.
[43] H. Zhang, V.M. Patel, Densely connected pyramid dehazing network, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2018.
[44] L. Zhang, L. Zhang, X. Mou, D. Zhang, Fsim: a feature similarity index for image quality assessment, IEEE Trans. Image Process. 20 (2011) 2378–2386.
[45] Q.S. Zhu, L. Shao, J. Mai, et al, A fast single image haze removal algorithm using color attenuation prior, IEEE Trans. Image Process. 24 (11) (2015) 3522–3533.