IET Image Processing

Research Article

IET Journals
The Institution of Engineering and Technology

Image dehazing with uneven illumination prior by dense residual channel attention network

ISSN 1751-9659
Received on 18th July 2019
Revised 1st May 2020
Accepted on 30th June 2020
doi: 10.1049/iet-ipr.2019.0873
www.ietdl.org

Shibai Yin[1,2,3], Jin Xin[1], Yibin Wang[4,5] ✉, Anup Basu[5]

[1]*Southwestern University of Finance and Economics, Department of of Economics and Information Engineering, Chengdu, People's Republic of China*

[2]*Key Laboratory of Financial Intelligence and Financial Engineering of Sichuan Province, Southwestern University of Finance and Economics, Chengdu 611130, People's Republic of China*

[3]*Collaborative Innovation Center for the Innovation and Regulation of Internet-based Finance, Southwestern University of Finance and Economics, Chengdu 611130, People's Republic of China*

[4]*Sichuan Normal University, Department of Engineering, Chengdu, People's Republic of China*

[5]*Department of Computing Science, University of Alberta, Edmonton, Canada*

✉ *E-mail: yibeen.wong@gmail.com*

Abstract: Existing dehazing methods based on convolutional neural networks estimate the transmission map by treating channel-wise features equally, which lacks flexibility in handling different types of haze information, leading to the poor representational ability of the network. Besides, the scene lights are predicted by an even illumination prior which does not work for a real situation. To solve these problems, the authors propose a dense residual channel attention network (DRCAN) for estimating the transmission map and use an image segmentation strategy to predict scene lights. Specifically, DRCAN is built based on the proposed dense residual block (DRB) and dense residual channel attention block (DRCAB). DRB extracts the hierarchical features with increasing receptive fields. DRCAB makes the network focus on the features containing heavy haze information. After the transmission map is estimated, fuzzy partition entropy combined with graph cuts is used to segment the transmission map into scene regions covered with varying scene lights. This strategy not only considers the fuzzy intensities of the low-contrast transmission map but also takes spatial correlation into account. Finally, a clear image is obtained by the transmission map and varying scene lights. Extensive experiments demonstrate that our method is comparable to most of existing methods.

1 Introduction

Intelligent technologies, including artificial intelligence and machine intelligence, are constantly evolving. Artificial intelligence empowers computers to perform human-like tasks more efficiently, while machine learning aids the computer to finish these tasks by breaking traditional rules. Hence, many electronic systems have rapidly emerged utilising machine intelligence. As a component of machine intelligence, machine vision relies on image pre-processing technology to extract information from images for performing visual tasks in electronic systems [1, 2]. For example, SAR imaging system uses kurtosis wavelet energy function to form texture features and use support vector machines (SVMs) to implement texture recognition [3]. Remote sensing systems utilise spectral clustering methods to obtain the accurate texture and colour features for segmenting polarimetric synthetic aperture radar images [4]. Since the outdoor images usually suffer from degradations, e.g. low contrast and low saturation, image dehazing technologies have become a necessary image processing method for intelligent systems [5–7]. Especially, automated parking systems rely on image dehazing to generate clear images for recognising vehicles and pedestrians [8]. Marine monitoring systems depend on image dehazing to obtain contrast-enhanced images for identifying ships [9].

To achieve image dehazing, single image dehazing methods based on atmospheric scattering models have been extensively studied [10–12]. Physical image processing is formulated as

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

where x is the index representing a pixel. I represents the observed hazy image. ρ denotes the scene radiance and also represents the clear image which needs to be estimated. A is the global atmospheric light and is usually set to a constant. $t(x)$ is the transmission map describing the portion of light that is not scattered and reaches a camera.

However, predicting ρ from (1) is difficult due to its ill-posed nature. It can be observed from (1) that multiple solutions can be found given a known hazy image I. Most existing methods attempt to solve this problem by estimating the transmission map and the atmospheric light via handcrafted priors. As an example, He *et al.* [11] presented the famous dark channel prior (DCP) based on the observation that some pixels in a local patch have very low intensity values in at least one of the RGB channels. Zhu *et al.* [12] proposed a colour attenuation prior (CAP) to estimate the transmission map using the scene depth. Fattal [13] and Berman *et al.* [14] further proposed the non-local prior (NLP) to estimate the transmission map. However, these priors depending on the statistical properties of a hazy image to estimate the transmission map cannot work well for all the cases, leading to inaccurate predicted results. Typically, DCP cannot predict accurate transmission values for the white objects in the scene which have similar colours as the sky regions. For refining the transmission map, Chen *et al.* [15] proposed gradient residual minimisation (GRM) for recovering a clear image after refining the transmission map via depth-edge-aware smoothing. Meng *et al.* [16] explored the boundary constraint and contextual regularisation (BCCR) on patch-wise transmission map for recovering a high-quality haze-free image. Zhao *et al.* [17] further proposed the multi-scale optimal fusion (MOF) model to fuse pixel-wise and path-wise transmission maps effectively, avoiding erroneously estimated transmission regions and halo artefacts. Xu *et al.* [18] proposed a fusion model to combine patch-wise and pixel-wise dehazing operators for overcoming halos and over-saturation. Although, these handcrafted prior based methods can improve the accuracy of the predicted transmission map, the possibility of uneven illumination in the scene is not considered at all. Consider Fig. 1 as an example, under even illumination, the camera in Fig. 1*a* only

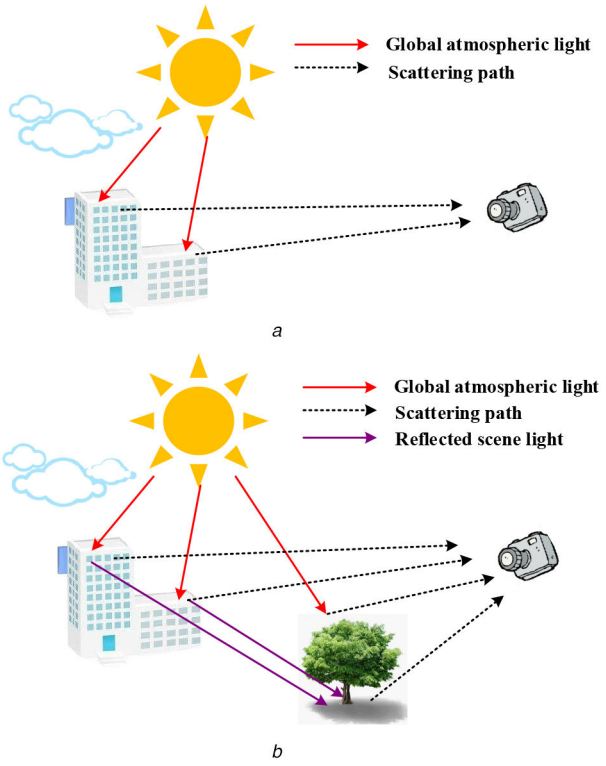


Fig. 1 *Physical atmospheric scattering model for different cases*
(a) Even illumination case, ***(b)*** Uneven illumination case

receives the atmospheric light reflected from objects, which leads to a fixed A in the $A\rho(x)t(x)$ term of (1). In reality, the reflected scene light and the atmospheric light all contribute to the final image. Hence, the camera not only receives the atmospheric light but also receives scene light reflected from objects which are not exposed in the atmospheric light directly (e.g. the trunk of the tree in Fig. 1*b*). To better describe the imaging process, Yoon *et al.*[19] proposed a wavelength-adaptive physical model (WAPM) without using a even illumination assumption and estimate the transmission map by an image segmentation strategy. As a further step, Ju *et al.* [20] proposed an improved atmospheric scattering model (IASM), where the atmospheric light and the varying scene light are calculated for addressing the uneven illumination problem. However, the transmission map is predicted by a linear model which cannot properly describe the relationships between the distribution of haze and the image properties.

Recently, deep learning-based methods have been widely used for many computer vision tasks, e.g. change detection, depth estimation, image restoration and image classification [21–23]. Generally speaking, these learning methods can be classified into two categories: one is supervised learning and the other is unsupervised learning. Deep belief networks (DBNs), which consists of multiple restricted Boltzmann machines (RBMs), are considered as an example which exploits both supervised and unsupervised approaches to learn network parameters. Specifically, RBMs are trained first, in an unsupervised manner, to initialsze the network parameters. Then, the parameters of the whole network are learned in a supervised approach based on an evaluation metric. Motivated by the growth of DBNs, it has been applied to the change detection field. For example, Samadi *et al.* [24] train the DBN with diverse data provided by morphological operators and employ the trained DBN to achieve change detection in SAR images. Owing to the advantage of providing diverse training data, this strategy can also be used in image dehazing.

On the other hand, convolutional neural networks (CNNs) are viewed as a supervised approach. In CNNs, it is necessary to have the labels of all the data for the training network. Hence, one common strategy is to build network architecture for learning the mapping relationship between a hazy image and a labelled clear image. Li *et al.* [25] first proposed the all-in-one dehazing network (AODN) to estimate the haze-free image without all the intermediate processing. Zhang *et al.* [26] proposed a perceptual pyramid deep network (PPDN) to directly learn a non-linear function between a hazy and a clear image. Qu *et al.* [27] proposed the enhanced pix2pix dehazing network (EPDN), motivating by the success of generative adversarial networks in image translation. Later, for enabling the mapping relationship learned between the hazy image and the clear image to be more reliable, more methods incorporate statistical regularities or information fusion strategy into the dehazing model. Zhang *et al.* [28] proposed the fully point-wise CNN (FPCNet) for modelling statistical regularities in hazy images. Zhang *et al.* [29] proposed a fast and accurate multi-scale end-to-end dehazing network (FAMED-Net) by comprising encoders at three scales and a fusion module for multi-scale information fusion. Ren *et al.* [30] proposed a gated fusion network (GFN) to estimate a clear image by fusing three derived images of the original hazy image effectively. Based on GFN, Liu *et al.* [31] further proposed the GridDehazeNet for image dehazing, which adopt the pre-processing module to convert a hazy image to several derived images for information fusion, and then introduce the post-processing module for improving the quality of the clear image. Comparing with the GFN, all the modules in the GridDehazeNet are fully trainable, which is in line with the performance of data-driven methods. Although these network models can automatically obtain a clear image from a hazy image, some physical parameters in the atmospheric scattering model, e.g. atmospheric light and transmission map, are not estimated separately. Once the transmission map or the atmospheric light needs to be explored in computer vision tasks, the end-to-end dehazing network cannot meet practical requirements. Hence, some methods are prone to estimate the transmission map and atmospheric light, separately. For instance, the densely connected pyramid dehazing network (DCPDN) proposed by Zhang and Patel [32] jointly learns the atmospheric light, transmission map and dehazing result simultaneously. The DehazeNet proposed by Cai *et al.* [33] maps a hazy image to a transmission map, and uses the empirical rules to acquire the atmospheric light. For improving the precision of the transmission map, Ren *et al.* [34] used a multi-scale CNN (MSCNN) to predict the transmission map and optimise it later by a refinement stage. However, there are still two factors hindering the performance of these methods. First, existing CNN based methods treat all the channel-wise features equally which lacks flexibility in handling different types of information among discriminative channels, leading to the poor representational ability of the network. In reality, different channel-wise features carry different information. For the case of estimating the transmission map, some channel-wise features contain more information related to heavy haze concentration, while others may carry more information about light haze concentration. Since heavy haze removal is more difficult than light haze removal, we make the network focus on heavy haze information more. Hence, the interdependencies among channels should be explored for accurate image dehazing. Second, with the predicted transmission map and atmospheric light, this strategy still estimates the clear image following (1) which does not consider uneven illumination.

To mitigate the first issue, we resort to the attention mechanism which can be viewed as a strategy to allocate the available computational resources towards the most formative components of the input. Due to attention weights being assigned to channel-wise features, the useful components of the input are found and paid more attention to for achieving different computer vision tasks, such as object recognition, image classification and image restoration [22, 23, 35, 36]. However, few researchers have investigated the effect of channel-wise attention mechanism in image dehazing. Considering that the transmission map is related to haze concentration, we propose the Dense residual channel attention network (DRCAN) which can assign more computational resources towards informative channel-wise features, e.g. the features carrying more information related to heavy haze concentration. To address the second issue, we combine the maximum fuzzy entropy with graph cuts to obtain different regions for predicting varying scene lights. To be specific, DRCAN is designed based on the encoder–decoder architecture, where the proposed dense residual block (DRB) and dense residual channel

attention block (DRCAB) are used as the basic building modules. DRB in the encoder is proposed based on the inspiration of the residual dense block (RDB) from [37], which contains feature processing, dense connections and the adaptive residual learning mechanism. DRCAB obtained by incorporating a channel-wise attention mechanism into DRB captures more informative channel-wise features for guiding the decoding process. After the transmission map is estimated by DRCAN (Fig. 2*b*), the IASM proposed by Ju *et al.* [20] is adopted for estimating the clear image. This model allows us to estimate a clear image by varying scene lights. Here, an image segmentation strategy is used for segmenting the transmission map into the distant, medium and nearby scenes, since the scene light varies with depth. Common SAR image segmentation strategies which can successfully segment degraded SAR images by capturing texture features, can also be used here, e.g. CNN and multilayer perceptron (CNN-MLP) [9] and SVM [3]. However, considering the fuzzy intensities of nearby and distant scenes, a maximal fuzzy entropy segmentation combined with graph cuts optimisation is selected for segmenting the transmission map (Fig. 2*c*). Such a strategy not only considers the fuzzy intensity of the low-contrast transmission map, but it also takes the spatial correlation into account. Finally, a clear image (Fig. 2*e*) can be obtained by the predicted transmission map (Fig. 2*b*) and optimal scene lights (Fig. 2*d*). Extensive experiments demonstrate that our method can achieve better accuracy and visual results over state of the art methods. The flowchart of the proposed method is shown in Fig. 2. We believe our method can work well in vehicular systems, marine monitoring systems, remote sensing systems and so on. For example, Sharifzadeh *et al.* [9] proposed a ship classification strategy by using a pre-processor, a detector and a hybrid CNN-MLP. To reduce false alarms, we can incorporate the proposed DRCAB into CNN-MLP to get rid of interfering clutter edges and speckles, and pay attention to the texture features of SAR images for classification.

There are three contributions of our network:

- We propose the DRCAN for accurately estimating the transmission map. By applying the proposed DRCAB in the network, the informative channel-wise features can be given much more attention.
- For taking uneven illumination into account, the varying lights are calculated from its covered regions. Here, we combine the maximal fuzzy entropy with a graph cuts strategy to segment the transmission map into different regions. Such a strategy not only considers the fuzzy intensity of the low-contrast transmission map, but it also takes the spatial correlation into account, to estimate varying lights from accurate scenes.
- Extensive experiments demonstrate that our method can achieve better dehazing effect over state of the art methods by using a predicted transmission map and varying scene lights.

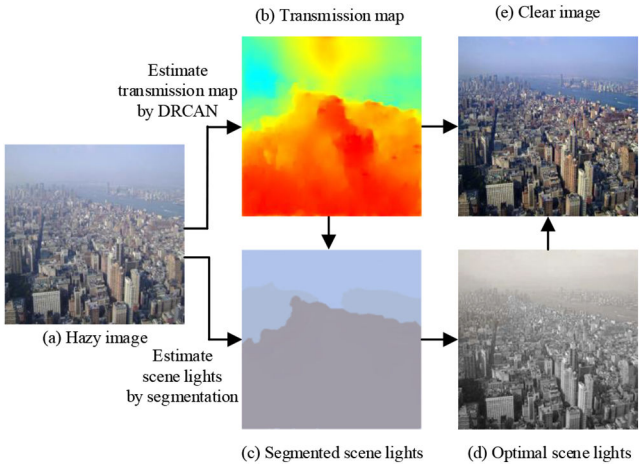


Fig. 2 *Flowchart of the proposed method*
(a) Hazy image, ***(b)*** Transmission map, ***(c)*** Segmented scene lights, ***(d)*** Optimal scene lights, ***(e)*** Clear image

2 Proposed method

2.1 Atmospheric scattering model with uneven illumination prior

For overturning the even illumination assumption in (1), Ju *et al.* [20] took varying scene lights into account and redefined the atmospheric scattering model as

$$I(x) = R_i\rho(x) + A(1 - t(x)), x \in E_i, \tag{2}$$

where x is the index of a pixel, E_i refers to the pixel set of the ith scene, R_i denotes the scene light in the ith scene, A is the atmospheric light and t is the transmission map. As only the observed image I is known, recovering the scene radiance ρ from (2) is an ill-posed problem. We need to estimate ρ based on I by estimating A, R_i and t.

2.2 DRCAN for estimating the transmission map

In this section, we propose the DRCAN based on the encoder–decoder architecture for directly estimating a transmission map from a hazy image. As observed in Fig. 3, our DRCAN model has four parts: shallow feature extraction, encoder, decoder and mapping layer. Now, we give more details on these components in DRCAN.

Shallow feature extraction: Previous research has shown that shallow features such as edges, textures and contours, are crucial for image restoration [22]. Similar to existing CNNs which utilise the convolution and pooling operations to extract edges and texture, we leverage the same strategy to implement shallow feature extraction. The related operation can be expressed as:

$$F_s = H_{SF}(I), \tag{3}$$

where I is a hazy input. H_{SF} is the composite function of shallow feature extraction which actually includes the pooling, Rectified Linear Unit (ReLU), Batch Normalisation and a convolutional layer with stride equal to 2. The output of shallow feature extraction denoted by F_s also serves as the input to the subsequent encoder.

Encoder: The encoder is used for extracting the hierarchical features with increasing receptive field size. By taking advantage of the dense and residual net, DRB is proposed as the basic building model in the encoder. Besides, for capturing more informative channel-wise features, the proposed DRCAB is applied at the end of the encoder, after the last DRB (see Fig. 3). More

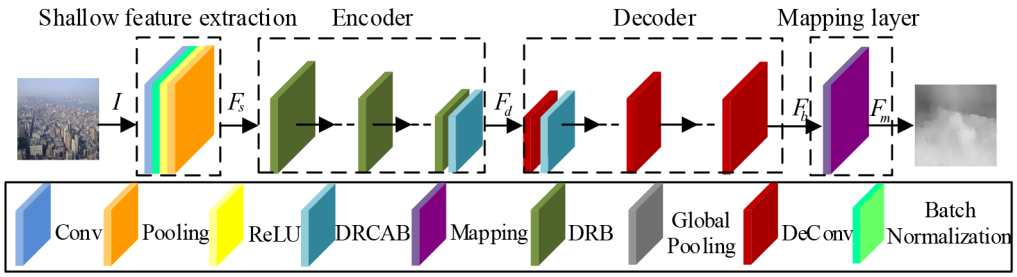


Fig. 3 *Overall architecture of the proposed DRCAN model*

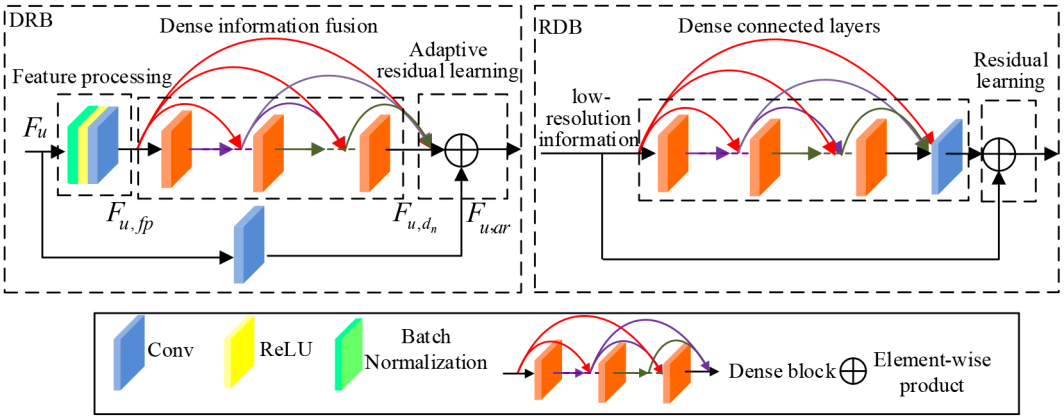


Fig. 4 *Proposed DRB and previous RDB*

details about DRB and DRCAB are given in Sections 2.3 and 2.4, respectively. We formulate the encoder as

$$\begin{aligned} F_d &= H_{\mathrm{EN}}(F_{\mathrm{S}}) \\ &= H_{\mathrm{DRBCAB}}(H_{\mathrm{DRB},d}(\cdots(H_{\mathrm{DRB},1}(F_{\mathrm{S}}))\cdots)), \end{aligned} \tag{4}$$

where $H_{\mathrm{EN}}(\cdot)$ is the function of the encoder which contains d DRBs and 1 DRCAB. $H_{\mathrm{DRB},d}$ denotes the operations of the dth DRB. H_{DRCAB} denotes the function of DRCAB. F_d is the output of the encoder.

Decoder: The decoder expands F_d by symmetrical de-convolutional layers. For further exploiting the inter-dependencies among channel-wise features and guiding the decoding process with informative input, the first de-convolutional layer in the beginning of the decoder is equipped with the proposed DRCAB. The output of the decoder F_d can be calculated as

$$\begin{aligned} F_b = H_{\mathrm{DE}}(F_d) = H_{\mathrm{DECOV},b}(\cdots H_{\mathrm{DECOV},3}(\\ H_{\mathrm{DECOV},2}(H_{\mathrm{DRCAB}}(H_{\mathrm{DECOV},1}(F_d))))\cdots), \end{aligned} \tag{5}$$

where $H_{\mathrm{DE}}(\cdot)$ denotes the function of the decoder which contains 1 DRCAB and b de-convolutional layers. F_b is the output of the decoder. $H_{\mathrm{DECOV},b}$ represents the bth de-convolution function.

Mapping layer: One convolutional layer is used at the tail of the network to map the learned features into a transmission map. Such a post-processing design can refine the transmission map and obtain accurate results

$$F_{\mathrm{m}} = H_{\mathrm{MAP}}(F_b), \tag{6}$$

where $H_{\mathrm{MAP}}(\cdot)$ and F_{m} denote the mapping function and the corresponding output, respectively.

2.3 DRB architecture

Inspired by the advantages of Dense Net [38] and Residual Net [39], Zhang and co-authors proposed a RDB to explore hierarchical features via dense connected layers and a residual learning layer (see RDB model in Fig. 4). Due to the convincing advantages of RDB, we also propose DRB based on RDB as the basic building module for the encoder. However, there are two main differences between RDB and DRB: First, RDB is designed for image super-resolution. Hence, each layer in the densely connected layers of RDB has direct access to the original low-resolution information for implicit deep supervision, as the red line in the densely connected layers in RDB. Our proposed DRB is designed for image dehazing and predicts the transmission map by learning the residuals between the hazy image and the ground truth of the transmission map. Therefore, we mainly focus on dense information fusion and residual learning, as the DRB in Fig. 4. Second, the stacked RDBs in the network [37] are used for extracting features with a fixed scale, enabling the network to obtain the higher resolution features for image super-resolution. We embed the DRB into the encoder–decoder architecture for extracting hierarchical features with varying feature size, leading to an increasing receptive field for image dehazing.

Based on the above discussion, DRB consists of feature processing, dense information fusion and adaptive residual learning, as shown in Fig. 4.

Feature processing is applied for rescaling the feature size and obtaining global information. Since the DRB used for building the encoder is expected to have the ability of adaptively rescaling feature size, we adopt a feature processing layer to reduce the feature size of the input and extract the global information. In the implementation, one possible solution is to perform the pooling operation for obtaining a larger image scope. However, this strategy may loose much of the details. To alleviate this problem, we use a convolutional layer with stride equal to 2 to implement feature processing. Let the input feature be denoted by F_u, the related operation is defined as:

$$F_{u,\mathrm{fp}} = H_{\mathrm{FP}}(F_u), \tag{7}$$

where H_{FP} denotes the composite function of Batch Normalisation, ReLU and one convolutional layer with stride equal to 2. $F_{u,\mathrm{fp}}$ is the output of the feature processing layer in the current RDB.

Dense information fusion is performed by a six-layer dense block, which is provided in a pre-trained dense-net121 [38]. Serving as an extractor, it can extract hierarchical features during the encoding. The related function can be defined as

$$F_{u,d_n} = H_{\mathrm{DI},n}([F_{u,\mathrm{fp}}, F_{u,d_1}, F_{u,d_2}\ldots, F_{u,d_{n-1}}]), 1 \le n \le 6, \tag{8}$$

where $[F_{d,\mathrm{fp}}, F_{u,d_1}, F_{u,d_2}\ldots, F_{u,d_{n-1}}]$ refers to the concatenations of the feature-maps produced by the $n-1$ convolutional layers in the dense block. $H_{\mathrm{DI},n}$ denotes the function of the 1×1 convolutional layer in the nth layer of the dense block. F_{u,d_n} is the corresponding output of the current nth convolutional layer.

Adaptive residual learning is implemented by adding input F_u into the output of the dense information fusion. Considering that the feature size of F_{u,d_n} is half the size of the input feature F_u, a convolutional layer with the stride of 2 is used before implementing residual learning. Such an operation can make the size of the input F_u half the size of its original resolution and enable the residual learning to be performed smoothly. Owing to the adaptive rescaling of the feature size, we refer to this residual learning as adaptive residual learning and the related operation is defined as

$$F_{u,\mathrm{ar}} = F_{u,d_n} + H_{\mathrm{AR}}(F_{d-1}), \tag{9}$$

where H_{AR} denotes the function of the convolutional layer with stride equal to 2, and $F_{u,\mathrm{ar}}$ is the corresponding output.

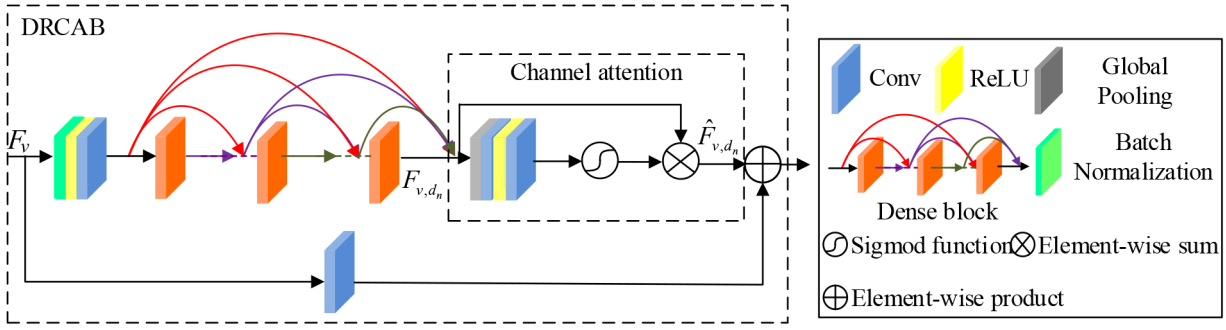


Fig. 5 *Proposed DRCAB*

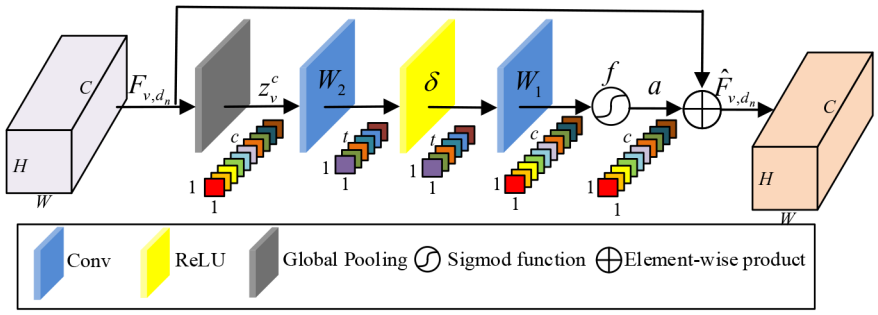


Fig. 6 *Channel attention mechanism*

2.4 DRCAB architecture

Differing from existing decoders which employ the deconvolutional layers to reconstruct the transmission map without a channel-wise distinction, we apply the DRCAB at the end of the encoder and at the beginning of the decoder for exploring the inter-dependencies across the channel-wise features (see Fig. 3)

From Fig. 5, we observe that DRCAB is obtained by inserting the channel attention mechanism into the slot between the dense information fusion and adaptive residual learning in DRB. More details on the channel attention mechanism can be found in Fig. 6. Let F_v be the input of DRCAB in the decoder, $F_{v,d_n} \in R^{H\times W\times C}$ in Fig. 5 can be obtained after F_v goes through the dense information layer which consists of a six-layer dense block, where H, W and C are height, width and the number of channels, respectively. For exploiting the inter-dependencies among channel-wise features, we first squeeze the spatial information into a channel descriptor. In our implementation, global average pooling is used for generating a set of local aggregated information. Hence, the local channel descriptor can be obtained by the channel-wise average value z_v^c via

$$z_v^c = \frac{1}{H\times W}\sum_{i=1}^{H}\sum_{j=1}^{W} F_{v,d_n}^c(i,j), \tag{10}$$

where $F_{v,d_n}^c(i,j)$ is the value at position (i,j) of the feature F_{v,d_n}^c in c channel, and z_v^c is the aggregated vector. Such a design allows the descriptor to express the whole image.

Then, we introduce a gating mechanism (the sigmoid function) for generating the attention weights α for each channel. Instead of using all the features, we focus on the informative features that have higher attention weights. The related operation is defined as

$$\alpha = f(W_1\delta(W_2 z_v)), \tag{11}$$

where $W_1 \in R^{t\times c}$ and $W_2 \in R^{c\times t}$ are learning weights of two fully connected layers; t is the reduction ratio, $\delta(\cdot)$ is the ReLU function and $f(\cdot)$ is the sigmoid function. By multiplying the attention weights α with the features F_{v,d_n}, the useful features can be adaptively rescaled and highlighted, obtaining output $\hat{F}_{v,d_n}$ in Fig. 5. Meanwhile, $\hat{F}_{v,d_n}^c$ in c channel is calculated by

$$\hat{F}_{v,d_n}^c = F_{v,d_n}^c \cdot \alpha^c \tag{12}$$

2.5 Loss function

Inspired by the success of the combined loss function used in PPDN [26], the DRCAN is trained by minimising the combined loss function, including the standard L_2 loss function, the gradient L_G loss function and feature edge L_F loss function. Formally, the loss function can be defined as

$$L = L_2 + \lambda L_G + \beta L_F, \tag{13}$$

where λ and β are the regulation coefficients for terms L_G and L_F. Let $\{I^i, t^i\}_{i=1}^N$ represent a pair of haze image and its corresponding transmission map and H_{DRCAN} denote the function of DRCAN. Then, we can compute L_2 as

$$L_2 = \frac{1}{N}\sum_{i=1}^{N} \| H_{DRCAN}(I^i) - t^i \|_2 \tag{14}$$

Similarly, given G_v and G_h, which are the gradient functions along the vertical and horizontal directions, respectively, we can obtain

$$L_G = \frac{1}{N}\left(\sum_{i=1}^{N} \| G_v(H_{DRCAN}(I^i) - G_v(t^i) \|_2 + \sum_{i=1}^{N} \| G_h(H_{DRCAN}(I^i) - G_h(t^i) \|_2\right) \tag{15}$$

L_F is defined using hierarchical features extracted by the pre-trained VGG-16 network [40] following (16). Where F_{f_1} and F_{f_2} are extracted high-level features, such as edge and texture information, from the first and second layers of the VGG-16 network

$$L_{\mathrm{F}} = \frac{1}{N}\Big(\sum_{i=1}^{N} \| F_{f_1}(H_{\mathrm{DRCAN}}(I^i) - F_{f_1}(t^i) \|_2 + \sum_{i=1}^{N} \| F_{f_2}(H_{\mathrm{DRCAN}}(I^i) - F_{f_2}(t^i) \|_2\Big). \tag{16}$$

2.6 Estimation of the scene incident light and atmospheric light

As mentioned in Section 2.1, the scene incident light R_i should be estimated from the regions it covers. Here, we try to segment the predicted transmission map into distant, intermediate and nearby scenes, because of the light variation with depth. Differing from Yoon *et al.*'s [19] method, which utilises a threshold-based strategy (fuzzy partition entropy) to partition scene regions, we combine fuzzy partition entropy with graph cuts to implement segmentation. Such a design not only considers the fuzzy intensity in a low-contrast transmission map, but it also takes the spatial correlation into account. Consider Fig. 2*b* as an example, compared to performing the fuzzy partition entropy directly (Fig. 7*a*), our strategy can remove isolated noise and avoid an object with low transmission values in the nearby scene being misclassified into the sky regions in the distant scene (Fig. 7*b*).

In our previous papers [41, 42], a segmentation strategy composed of the fuzzy partition entropy and graph cuts optimisation was proposed for multilevel segmentation. In addition, an iterative scheme was utilised for improving the computational efficiency of fuzzy partition entropy. Here, the same segmentation approach is adopted to segment the transmission map. Since the goal is to segment the transmission map into three fuzzy sets, referred as the distant scene set E_{d}, the intermediate scene set E_m and the nearby scene set E_c, we select the *S*-function, *Z*-function and *M*-function to build the fuzzy three-partition entropy (see Appendix 1 for the definitions of *S*-function, *Z*-function and *M*-function). Then, fuzzy set probabilities P_c, P_m and P_{d} of sets E_c, E_m and E_{d} are defined as

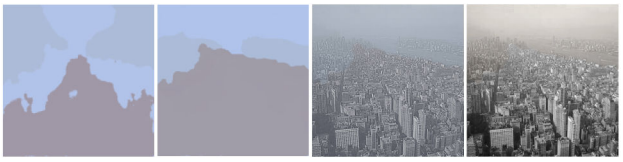
$$\begin{cases} P_c = \sum_{k=0}^{255} h(k)S(k), \\ P_m = \sum_{k=0}^{255} h(k)M(k), \\ P_{\mathrm{d}} = \sum_{k=0}^{255} h(k)Z(k), \end{cases} \tag{17}$$

where $h(\cdot)$ represents the normalised histogram function, which calculates the occurrence probability of an assigned grey level k. Thus, the total fuzzy entropy of the transmission map can be obtained as

$$H(u_1, v_1, w_1, u_2, v_2, w_2) = -P_c\log(P_c) - P_m\log(P_m) - P_d\log(P_d), \tag{18}$$

where $u_1, v_1, w_1, u_2, v_2, w_2$ are parameters of the *S*-function, the *Z*-function and the *M*-function.

The most appropriate fuzzy three-partition probabilities can be obtained by maximising (18). Here an iterative scheme proposed in our previous work [41, 42] is used to search the maximal entropy efficiently, which performs iterative calculation by three separated sum operations with two parameters, hence the time complexity is $O(n^2)$. To make a further optimisation, multi-label graph cuts is performed based on the obtained fuzzy set probabilities to ensure the spatial correlation. To be specific, the data cost T_{data} of graph cuts is designed by P_c, P_m and P_{d}



a b c d

Fig. 7 *Estimation of scene lights in an aerial image*
(a) Scene lights obtained by the maximal fuzzy partition entropy segmentation, ***(b)*** Scene lights obtained by the fuzzy partition entropy and graph cuts optimisation, ***(c)*** Optimal scene lights obtained with window size $w_{\mathrm{d}} = 2$ in a regularisation optimisation strategy, ***(d)*** Optimal scene lights obtained with window size $w_{\mathrm{d}} = 7$ in a regularisation optimisation strategy

$$\begin{cases} T_{\mathrm{data}}(l_p = \text{`close scene'}) = -\log(P_c), \\ T_{\mathrm{data}}(l_p = \text{`medium scene'}) = -\log(P_m), \\ T_{\mathrm{data}}(l_p = \text{`distant scene'}) = -\log(P_{\mathrm{d}}), \end{cases} \tag{19}$$

where l_p is the label of pixel p.

The smoothing term T_{smooth} of graph cuts is defined based on the definition in reference [42] as

$$T_{\mathrm{smooth}} = \exp(-\frac{(k_p - k_q)^2}{2\sigma^2}) \times \frac{1}{\mathrm{dist}(p,q)}, \tag{20}$$

where k_p and k_q denote the grey levels of adjacent pixels p and q respectively, $\mathrm{dist}(p,q)$ represents the Euclidean distance between pixels p and q. σ refers to the level of variation between adjacent pixels in the image and stays in the range $[0, 1]$. After T_{data} and T_{smooth} are set for all the pixels in the graph cuts model, the optimal label assignment for E_i is achieved by the α–β swap algorithm [43], see Appendix 2 for details.

Once the segmented E_i is obtained, we choose the top 1% brightest pixels in a hazy image from the E_i region. Then, the coarse R_i can be obtained by calculating the average intensity of the pixels in the corresponding E_i region. Consider Fig. 2*b* as an example, the coarse R is displayed in Fig. 7*b*. Note that the scene light covering the distant scene region is selected as the atmospheric light A.

Furthermore, for refining and conforming R closely to realistic ambient light, we adopt the regularisation optimisation from [20] for enhancing the edge information

$$\hat{R} = \mathrm{argmin}\Big\{\| R - \hat{R} \|_2^2 + \theta_1 \| \bigtriangledown(\hat{R}) - \bigtriangledown(G) \|_2^2\Big\}, \tag{21}$$

where θ_1 is the regularisation parameter which by default is set to 0.4. G refers to the luminance component of the hazy image I. $\bigtriangledown$ is the gradient operation. The first term of (21) ensures that the optimal scene light $\hat{R}$ approximates the segmentation result R, while the second term imposes the edge features in $\hat{R}$ corresponding to the ones in the guiding image G. For solving (21) effectively, the iterative strategy from [20, 44] is adopted and the specific iterative form is expressed as

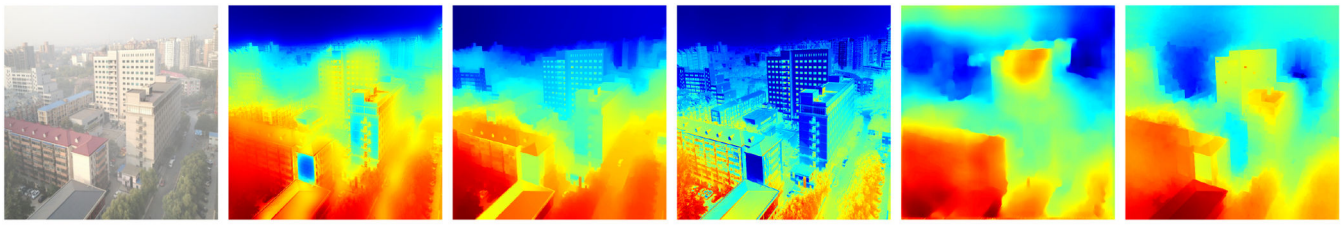
$$\hat{R}_k(x) = \Big(R_k(x) - \theta_1\Big(\sum_{x \in N} (\hat{R}_{k-1}(x) - G(x)) + |N|G(x)\Big)\Big)/(1 - \theta_1|N|) \tag{22}$$

where x is the index representing a pixel; $\hat{R}_k(x)$ and $\hat{R}_{k-1}(x)$ represent the $\hat{R}$ of the $(k-1)$th and the kth iteration, respectively; $x \in N$ denotes a $w_d \times w_d$ local region N with x in the centre. $|N|$ defines the number of pixels in N. We set the initial state $\hat{R}_0$ as R and G as the luminance component of the hazy image I. The iteration is finished when the Euclidean Distance between $\hat{R}_k$ and $\hat{R}_{k-1}$ is <0.001. Consider an aerial image (see Fig. 2*a*) as an

a b c d

Fig. 8 *Estimation of scene lights in a flower image*

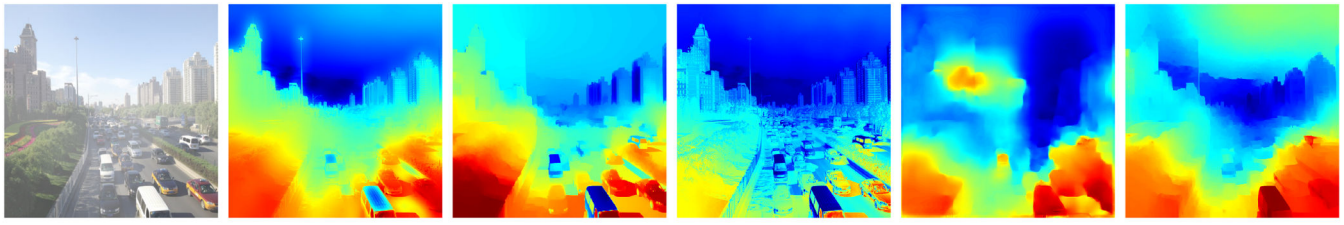
(a) Original hazy image, ***(b)*** Scene lights obtained by the fuzzy partition entropy and graph cuts optimisation, ***(c)*** Optimal scene lights obtained with window size $w_d = 2$ in a regularisation optimisation strategy, ***(d)*** Optimal scene lights obtained with window size $w_d = 7$ in a regularisation optimisation strategy



SSIM	0.809	0.859	0.742	0.911	0.969
a	*b*	*c*	*d*	*e*	*f*

Fig. 9 *Transmission results for the synthetic hazy image 1*

(a) Input, ***(b)*** DCP [11], ***(c)*** IASM [20], ***(d)*** WAPM [19], ***(e)*** PPDN [26], ***(f)*** Proposed work



SSIM	0.818	0.733	0.695	0.847	0.974
a	*b*	*c*	*d*	*e*	*f*

Fig. 10 *Transmission results for the synthetic hazy image 2*

(a) Input, ***(b)*** DCP [11], ***(c)*** IASM [20], ***(d)*** WAPM [19], ***(e)*** PPDN [26], ***(f)*** Proposed work

example, Fig. 7*a* is the fuzzy three-partition result of the transmission map in Fig. 2*b*, while Fig. 7*b* is the graph cuts result of Fig. 7*a*. Figs. 7*c* and *d* show optimal scene lights obtained with different window sizes w_d in the regularisation optimisation strategy. Note that the parameter w_d is important for obtaining the optimal scene light. A small window size with $w_d = 2$ results in an obvious science light boundary, e.g. Fig. 7*c*, while a big window size such as $w_d = 25$ leads to over-smooth transition for the scene light in different the scene regions. Hence, it is necessary to set a proper w_d value for obtaining the scene lights. For Fig. 2*a*, where depth changes gradually and the corresponding close, medium and distant scenes are not easy to discriminate, the best w_d is set to 7 empirically and the corresponding result is shown in Fig. 7*d*. As can be seen, not only do the scene lights in different segmented regions have different luminance values but also the border among these scene lights are natural. For further testing the robustness of the regularisation optimisation strategy, a flower image (see Fig. 8*a*) with obvious close, medium and distant scenes is also selected for estimating scene lights with different w_d. After performing extensive experiments with different w_d, we find a similar conclusion also fits this kind of images. As displayed in Fig. 8, the satisfying scene light (Fig. 8*d*) can be obtained based on the segmented result (Fig. 8*b*) with window size w_d=7, while the inferior scene light (Fig. 8*c*) with unnatural border is obtained with window size w_d=2. Hence, the widow size w_d is set to 7 in our experiments.

2.7 Recovering the scene radiance

With the estimated transmission map, the varying scene light and the atmospheric light, we can recover the scene radiance according to (23). Simultaneously, to avoid producing too much noise and preserve a small amount of haze in the scene radiance, we restrict the transmission map t to a lower bound t_0. Then, the scene radiance can be expressed by

$$\rho = \frac{I - A(1-t)}{\hat{R}\ \max(t, t_0)}, \tag{23}$$

3 Experiments

In this section, we compare our proposed method with five state of the art image dehazing methods DCP-based dehazing method proposed by He *et al.* [11], WAPM-based dehazing method proposed by Yoon *et al.* [19], IASM-based dehazing method proposed by Ju *et al.* [20], AODN by Li *et al.* [25], and PPDN proposed by Zhang *et al.* [26]. Among these algorithms, DCP [11], WAPM [19] and IASM [20] are handcrafted prior based methods. While AODN [25] and PPDN [26] are CNN-based methods. In order to verify the performance of different methods, we conduct tests on synthetic data sets and real world images with visual effects (see Figs. 9–19) and quantitative measures (see Tables 1 and 2).

3.1 Data sets

Recently, a public RESIDE data set which collects abundant synthetic hazy images, depth images and corresponding clear images was released for single image dehazing [45]. For training DRCAN and testing the performance of the proposed method effectively, we randomly select 4000 synthetic outdoor images with $\beta \in \{0.04, 0.06, 0.08, 0.1, 0.12, 0.16, 0.2\}$ and $A \in \{0.8, 0.85, 0.9\}$ to create a training data set. Simultaneously, the corresponding

a b c d e f

Fig. 11 *Synthetic hazy images and its corresponding ground truths*
(a) Synthetic hazy image 1, ***(b)*** Ground truth of image 1, ***(c)*** Zoom-in details of the ground truth of image 1, ***(d)*** Synthetic hazy image 2, ***(e)*** Ground truth of image 2, ***(f)*** Zoom-in details of the ground truth of image 2

19.11/0.923 a | 17.75/0.894 b | 27.57/0.977 c | 20.55/0.911 d | 17.99/0.842 e | **28.25/0.982** f

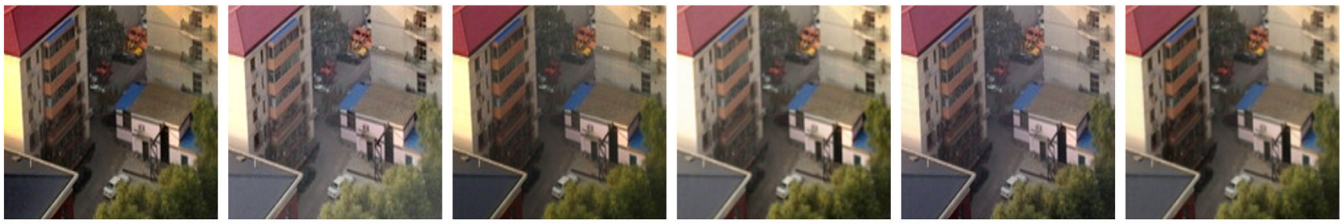
Fig. 12 *Dehazing results of the synthetic hazy image 1*
(a) DCP [11], ***(b)*** IASM [20], ***(c)*** WAPM [19], ***(d)*** PPDN [26], ***(e)*** AODN [25], ***(f)*** Proposed work. The numbers (PSNR/SSIM) are given under each method and the best results are highlighted in bold

Table 1 Quantitative SSIM results for the transmission map and SSIM/PSNR results for scene radiance on the synthetic testing data set

	DCP [11]	IASM [20]	WAPM [19]	PPDN [26]	Proposed work
transmission	0.8736	0.8961	0.8864	0.9543	0.9742
image	0.8243/17.59	0.8596/19.54	0.8469/18.79	0.9312/23.87	0.9510/28.56

Table 2 Average PSNR and SSIM comparisons on the outdoor images of SOTS. italic, bold and bold-italic fonts are used to indicate top first, second and third best performance

	DehazeNet [33]	AODN [25]	DCPDN [32]	GFN [30]	EPDN [27]	MSCNN [34]	FAMED-Net [29]	FPCNet [28]
PSNR	22.46	20.29	19.93	21.55	22.57	22.06	**29.03**	22.75
SSIM	0.8514	0.8765	0.8449	0.8444	0.8630	0.9078	**0.9570**	0.9014
	GridDehazeNet [31]	MOF [17]	DCP [11]	NLP [14]	CAP [12]	BCCR [16]	GRM [15]	proposed work
PSNR	*30.86*	18.61	19.13	19.26	18.88	14.06	19.15	***27.62***
SSIM	*0.9819*	0.873	0.8148	0.6190	0.7980	0.5660	0.8633	***0.9505***



a b c d e f

Fig. 13 *Zoom-in details of Fig. 12*
(a) DCP [11], ***(b)*** IASM [20], ***(c)*** WAPM [19], ***(d)*** PPDN [26], ***(e)*** AODN [25], ***(f)*** Proposed work

transmission maps which are calculated by providing depth images are also used for training. Similarly, another 400 outdoor images with the corresponding transmission maps and clear images are randomly selected for testing. In addition, the sub-data set SOTS from RESIDE containing 500 outdoor images with different haze concentration and corresponding clear images is also employed as the testing data set.

3.2 Comparison of transmission map

In this subsection, we first give the implementation details of DRCAN in Section 3.2.1. Then, we further investigate the effects of different components and basic network parameters of DRCAN in Section 3.2.2. Finally, the accuracy of estimated transmission maps is verified by comparing DRCAN with existing methods.

3.2.1 Implementation details of DRCAN: We use three DRBs, one DRCAB in the encoder and three de-convolutional layers and one DRCAB in the decoder. Note that the last RDB in the encoder and the first de-convolutional layer in the decoder are equipped with DRCABs for capturing informative data. In each DRB and DRCAB, we used a six-layer dense block with a growth rate of 32 for dense information fusion. All the kernel sizes of convolution

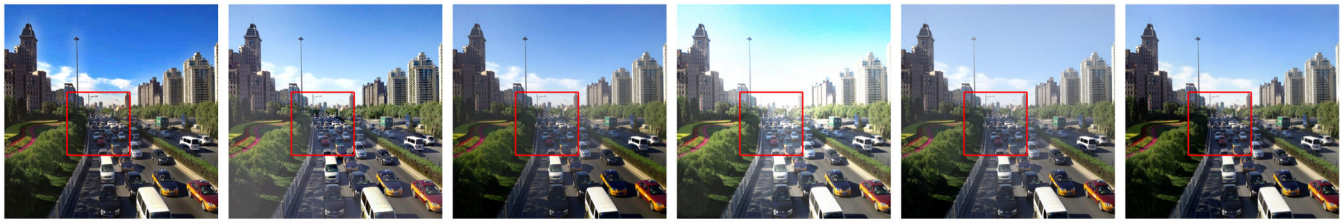


Fig. 14 *Dehazing results of the synthetic hazy image 2*
(a) DCP [11], ***(b)*** IASM [20], ***(c)*** WAPM [19], ***(d)*** PPDN [26], ***(e)*** AODN [25], ***(f)*** Proposed work. The numbers (PSNR/SSIM) are given under each method and the best results are highlighted in bold

Fig. 15 *Zoom-in details of Fig. 14*
(a) DCP [11], ***(b)*** IASM [20], ***(c)*** WAPM [19], ***(d)*** PPDN [26], ***(e)*** AODN [25], ***(f)*** Proposed work

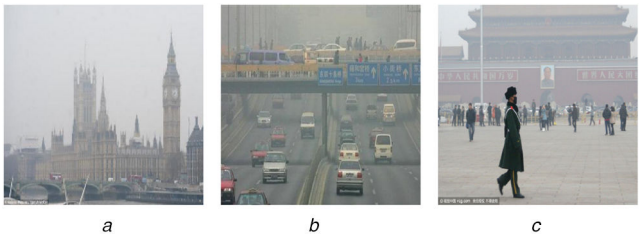


Fig. 16 *Real hazy images*
(a) Real building image, ***(b)*** Real overbridge image, ***(c)*** Real Tian An Men image

Fig. 17 *Dehazing result of real building image*
(a) DCP [11], ***(b)*** WAPM [19], ***(c)*** IASM [20], ***(d)*** PPDN [26], ***(e)*** AODN [25], ***(f)*** Proposed work

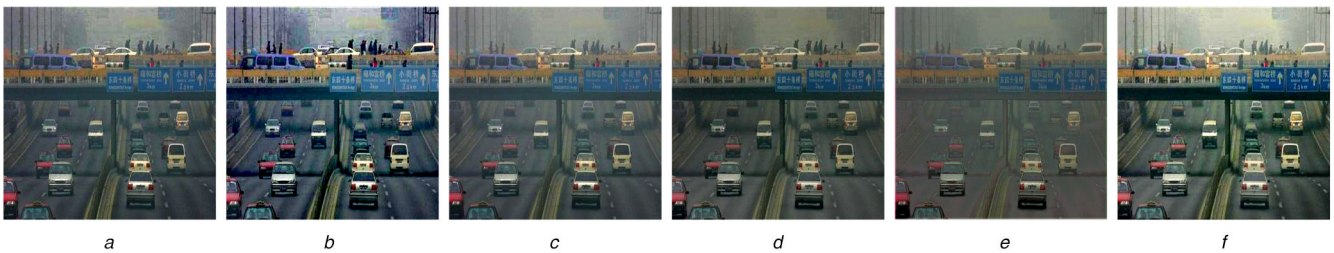


Fig. 18 *Dehazing result of real overbridge image*
(a) DCP [11], ***(b)*** WAPM [19], ***(c)*** IASM [20], ***(d)*** PPDN [26], ***(e)*** AODN [25], ***(f)*** Proposed work

and deconvolutional layers are set to 3×3, except for the convolution layer in the shallow feature extraction, where the kernel size is 7×7. During training, we use ADMA as the optimisation algorithm with $\beta_1 = 0.5$, $\beta_2 = 0.999$ and $\epsilon = 10^{-8}$. The original learning rate and batch size are set to 0.002 and 2. All the training images are resized to 512×512 before training. Training DRCAN roughly takes two days on the Pytorch platform with NVIDIA RTX 2080 Ti GPU for 800,000 iterations.

3.2.2 Ablation study: The proposed DRB and DRCAB are significant contributions of this work. Therefore, in the ablation study, we first verify the effectiveness of these two modules. Then,

a *b* *c* *d* *e* *f*

Fig. 19 *Dehazing result of real Tian An Men image*
(a) DCP [11], ***(b)*** IASM [19], ***(c)*** WAPM [20], ***(d)*** PPDN [26], ***(e)*** AODN [25], ***(f)*** Proposed work

Table 3 Quantitative SSIM for ablation study evaluated on the synthetic testing data set

Experiment index	1	2	3	4	5	6	7
Adaptive residual learning	×	√	×	√	√	√	√
channel-wise attention mechanism	×	×	√	√	√	√	√
number of DRCAB in encoder	1	1	1	1	2	3	0
number of DRCAB in decoder	1	1	1	1	2	3	0
SSIM	0.9541	0.9687	0.9628	**0.9742**	0.9654	0.9419	0.9517

Bold values indicate the best results.

we further verify the effects of basic parameters in the DRCAN, e.g. adaptive residual learning in DRB and DRCAB, channel-wise attention mechanism in DRCAB and number of DRCAB in the network.

To test the effectiveness of DRB and DRCAB, we keep all the other configurations the same as DRCAN, except for replacing DRB with RDB [37] and replacing DRCAB with RDCAB, which is generated by incorporating the channel-wise attention mechanism into RDB. Since RDB captures hierarchical features with fixed scale, the output features of the encoder have the same spatial size of the input image. For keeping symmetrical feature size, we further use convolution layer with a stride of 1 to take place with the de-convolution layer in the decoder. Then, this variant model denoted as RDCAN is obtained. Compared to our DRCAN whose average SSIM is 0.9742 on 400 outdoor testing images, RDCAN performs very poorly with SSIM = 0.9329. This result demonstrates that using RDBs and RDCABs in our dehazing network cannot capture hierarchical features with varying scales and results in poor performance.

Based on the above conclusion, we further verify the effects of basic parameters in our DRCAN. To this end, seven configurations of the proposed network are evaluated on the testing data set and related results are shown in Table. 3. In order to explicitly describe each configuration, special notations, e.g. ×, √ and numbers, are used in Table 3. Where the component marked with × means this component is removed for testing, otherwise the component marked with √ means it is kept in the network. The numbers in Table 3 represent the number of different components in the current configuration.

To demonstrate the effects of DRB and DRCAB, we first remove the adaptive residual learning and the channel-wise attention mechanism from DRB and DRCAB in Experiment 1. Thus, both of these two blocks are degraded into a dense block. Simultaneously, we set the number of DRCAB in the encoder and the decoder to 1 and 1. From Table 3, we find that Experiment 1 works poorly with SSIM=0.9541. We further use Experiment 1 as the baseline for testing the effect of adaptive residual learning in DRB and DRCAB. For Experiment 2, we add the adaptive residual learning into DRB and DRCAB, then the SSIM score reaches 0.9687. For Experiment 3, we only add the channel-wise attention mechanism into DRCAB, and SSIM is improved from 0.9591 in the baseline to 0.9628. This indicates that simply stacking dense blocks in the encoder and the decoder cannot achieve optimal performance for image dehazing. The performance would increase with additional adaptive residual learning or channel-wise attention mechanism. By using both components in Experiment 4, even better performance with SSIM=0.9742 is obtained.

Based on the above discussion, we use the design of DRB and DRCAB in Experiment 4 for testing the effects of the number of DRCAB in the decoder. Because DRCAB uses the channel-wise attention mechanism to extract useful information with a large receptive field, it is applied at the end of the encoder and the beginning of the decoder. Based on this design principle, in Experiment 5, we use two DRCABs at the end of the encoder and two DRCABs at the beginning of the decoder. In Experiments 6, 3 DRCABs are used in both the encoder and the decoder. Besides, we also remove all the DRCABs in Experiment 7 to test the effect of DRCAB. Comparing the results of Experiments 4, 5, 6 and 7, we find that Experiment 4 which uses one DRCAB in both the encoder and the decoder, achieves the best result.

3.2.3 Comparison of transmission map on testing data set: Our synthetic testing data set is used to further evaluate the estimated transmission maps. Since the ground truths of the transmission map in the testing data set are available, we are able to evaluate the results visually and quantitatively. Visual results of the two samples in our testing data set are displayed in Figs. 9 and 10 with SSIM marked under each image. Besides, for clearly examining the results, all the intensities of the pixels in the transmission map are transformed to RGB values. From Figs. 9 and 10, we find that comparable methods such as DCP [11], WAPM [19] and IASM [20] can estimate the transmission map with structural details, but they also tend to generate erroneous estimation. For example, the results of DCP proposed by He *et al.* [11] displayed in Figs. 9*b* and 10*b* show that the buildings and vehicles have similar colours as the sky region, since the DCP views all the white objects as atmospheric light. IASM [20] and WAPM [19] which predict the transmission map via luminance and threshold-based segmentation respectively still cannot handle the white object in the foreground (see the buildings in Fig. 10*c* and vehicles in Fig. 10*d*). In contrast, the results from PPDN [26] and our DRCAN in Figs. 9*e*, 10*e* and Figs. 9*f*, 10*f* are much closer to the ground truths, due to the higher SSIM values marked under these results. However, our results preserve more necessary details and have highest SSIM values. The quantitative SSIM evaluated on 400 test images are tabulated in the second row of Table 1, firmly demonstrating that our DRCAN achieves the best performance with highest SSIM. It is also reasonable for DRCAB to perform well because it assigns more computational resources on informative channel-wise features by using DRCAB in the encoder and the decoder.

3.3 Comparison of dehazing results on synthetic images

The visual dehazing results of two challenging samples are displayed in Figs. 11*a* and *d*. As well, the corresponding ground

truths and zoomed-in details of regions enclosed in red rectangles in clear images are shown in Figs. 11*b*, *e* and *c*, *f*. The dehazing results of different algorithms are displayed in Figs. 12–15. From Fig. 12*a*, we find that the results of DCP [11] suffer from severe colour distortions, since DCP cannot handle the white objects. For example, the sky colour in Fig. 12*a* has turned yellow. Besides, the colour of sky regions in Fig. 14*a* are also darker than that of the ground truth (see Fig. 11*e*). Although, zoomed-in details shown in Fig. 13*a* have stronger contrast and better visual effect than our result displayed in Fig. 13*f*, the colour distortions still exist. For example, the roof in the left corner of Fig. 13*a* are much darker than the ground truth shown in Fig. 11*c*. While, our zoomed-in details (see Fig. 13*f*) are close to the ground truth (see Fig. 11*c*). The results by IASM [20] and WAPM [19] are either too dark or too white. For example, the colours of the red roof in Figs. 13*b* and *c* are pale and dark, respectively. This colour distortion arising from transmission estimation errors cannot be avoided effectively in the dehazing results, due to the adopted handcrafted prior. In IASM [20], Ju *et al.* use a linear model to describe the transmission map which causes inaccurate results. In WAPM [19], Yoon *et al.* use a threshold based strategy to partition different regions covered with varying scene lights without considering the spatial correlation, this results in the misclassification of regions. On the other hand, the dehazing results by AODN [25] still have some haze residuals and colour distortions (see Figs. 12*e* and 14*e*). Obvious residual haze and colour distortion can be observed in the zoomed-in details in Figs. 13*e* and 15*e*. The dehazing results by PPDN [26] are clearer than the above mentioned results. However, upon detailed inspection, this strategy reveals over dehazed results, e.g. over white sky regions in Figs. 12*d* and 14*d*. In contrast, our results (Figs. 12*f* and 14*f*) with less colour distortion are closest to the ground truths (Figs. 11*b* and 11*e*). The PSNR/SSIM marked under each image and the quantitative results tested on 400 test images displayed in third row of Table 1 further demonstrate the effectiveness of the proposed method.

In addition, for further testing the robustness of our method, we compare our algorithm with several state of the art methods on the outdoor images of SOTS, including the CNNs based methods, e.g. DehazeNet [33], AODN [25], DCPDN [32], GFN [30], EPDN [27], MSCNN [34], FAMED-Net [29], FPCNet [28], GridDehazeNet [31] and the handcrafted priors based method, e.g. DCP [11], NLP [14], CAP [12], BCCR [16], GRM [15] and MOF [17]. The related quantitative results displayed in Table 2. From Table 2, we find that GridDehazeNet [31] and FAMED-Net [29] achieve the best and the second performance, respectively. Our method ranks third. However, note that the GridDehazeNet consists of a pre-processing module, a dehazing module and a post-processing module for image dehazing. FAMED-Net [29] includes encoders at three scales and a fusion module to fuse multi-scale information. Hence GridDehazeNet and FAMED-Net have the first and second competitive performance. In contrast, our model removes haze with uneven illumination prior by DRCAN, which neither adopts pre-processing operation, post-processing operation or adopts multi-scale information fusion. Further, our method even outperforms some complicated networks, e.g. EPDN, FPCNet, DCPDN.

3.4 Evaluation on real-world images

To further investigate the generalisation ability of the proposed method, we conduct visual comparisons on real-world images (see Fig. 16). As displayed in Figs. 17–19, DCP [11] and WAPM [19] fail to recover the real colour. For example, the sky regions in Figs. 18*a* and *b* are much darker than the real colour. IASM [20] and AODN [25] not only darken the images, but also leave residual haze in the results (see Figs. 17*c*, 18*c*, 19*c* and Figs. 17*e*, 18*e* and 19*e*). Although, the results by PPDN [26] are clearer than that of other algorithms, it still leaves a small amount of residual haze in the dehazing results, which leads to loss of detailed information, e.g. the words on the Tian An Men in Fig. 19*d* cannot be recognised. By comparison, our method removes haze with fine details and realistic colour shifts (see Figs. 17*f*, 18*f* and 19*f*).

3.5 Analysis of execution time

The run time of our algorithm is mainly the sum of two parts: the first part is spent on estimating the transmission map by the proposed DRCAN, which is discussed in Section 2.2. Since there are only three DRBs, one DRCAB in the encoder and 3 de-convolutional layers and 1 DRCAB in the decoder, the architecture of DRCAN is simple and the network depth is shallow. The average run time for a 512×512 testing image is around 0.2 s. The second part is spent on calculating the scene lights by segmenting the transmission map via fuzzy partition entropy and graph cuts, also including the time spent on refining the scene lights with regularisation optimisation. This part is discussed in Section 2.6. For the fuzzy partition entropy operation, the iterative calculation method [41, 42] used for searching the maximal fuzzy partition entropy is performed by three separate sum operations with two parameters. Hence, the run time is not influenced by the image size and is around 0.2 s for each image. In additions, graph cuts is also computationally efficient by using α–β swap operator [43] and the run time on one image is about 3 s. Finally, the iterative scheme [20, 44] for solving regularisation optimisation takes around 0.5 s. Hence, the total run time for one image is around 3.9 s.

4 Conclusion

In this work, we proposed an image dehazing method with uneven illumination prior by using a novel DRCAN. First, we proposed a DRCAN for learning the mapping between a hazy image and a transmission map. To be specific, the encoder is formed by the proposed DRB which helps extract the hierarchical features with increasing receptive fields. In addition, the symmetrical de-convolutional layers are adopted to build the decoder. To enable the decoder to be guided by meaningful information (e.g. the features containing heavy haze information), the last DRB in the encoder and first de-convolutional layers in the decoder are equipped with the proposed DRCAB. Furthermore, to calculate the varying scene lights under an uneven illumination prior, fuzzy partition entropy combined with graph cuts is used for segmenting the transmission map into different regions covered with different scene lights. This segmentation strategy not only considers the fuzzy intensities in the low-contrast transmission map but also takes the spatial correlation into account. After calculating the scene lights in its corresponding regions, a clear image can be obtained by the transmission map and scene lights. Extensive experiments on synthetic images and real-world images demonstrate promising performance for our image dehazing method. We believe that our model can be applied not only to image dehazing but also to other image restoration tasks, e.g. image deraining and image deblurring. We will further investigate the effect of the proposed model on various image restoration tasks where the proposed DRCAB learns to capture the most relevant information from the degraded images.

5 Acknowledgments

This work was supported by the National Natural Science Foundation of China (No. 61502396), the Education Department Foundation of Sichuan Province (No. 18ZB0484), the Fundamental Research Funds for the Central Universities (No. JBK2002026) and the funding from UAHJic at the University of Alberta. In addition, this work is also supported by the Key Laboratory of Financial Intelligence and Financial Engineering of Sichuan Province, the China Scholarship Council (CSC) and 2020 Educational Reform Funds of the Central Universities.

6 References

[1] Vicente, A., Raveendran, R., Huang, B.*, et al.*: 'Computer vision system for froth-middlings interface level detection in the primary separation vessels', *Comput. Chem. Eng.*, 2019, **123**, pp. 357–370

[2] Xiao, J., Zhu, L., Zhang, Y.*, et al.*: 'Scene-aware image dehazing based on sky-segmented dark channel prior', *IET Image Process.*, 2017, **11**, (12), pp. 1163–1171

[3] Akbarizadeh, G.: 'A new statistical-based kurtosis wavelet energy feature for texture recognition of sar images', *IEEE Trans. Geosci. Remote Sens.*, 2012, **50**, (11), pp. 4358–4368

[4] Akbarizadeh, G., Rahmani, M.: 'Efficient combination of texture and color features in a new spectral clustering method for polsar image segmentation', *Natl. Acad. Sci. Lett.*, 2017, **40**, (2), pp. 117–120
[5] Wang, W., He, C., Xia, X.G.: 'A constrained total variation model for single image dehazing', *Pattern Recognit.*, 2018, **80**, pp. 196–209
[6] He, L.Y., Kun, L., Zhao, J.Z., *et al.*: 'Visibility restoration of single foggy images under local surface analysis', *Neurocomputing*, 2019, **341**, pp. 212–226
[7] Gao, Y., Chen, H., Li, H., *et al.*: 'Single image dehazing using local linear fusion', *IET Image Process.*, 2017, **12**, (5), pp. 637–643
[8] Heimberger, M., Horgan, J., Hughes, C., *et al.*: 'Computer vision in automated parking systems: design, implementation and challenges', *Image Vis. Comput.*, 2017, **68**, pp. 88–101
[9] Sharifzadeh, F., Akbarizadeh, G., Kavian, Y.S.: 'Ship classification in sar images using a new hybrid cnn–mlp classifier', *J. Indian Soc. Remote Sens.*, 2019, **47**, (4), pp. 551–562
[10] Anwar, M.I., Khosla, A.: 'Vision enhancement through single image fog removal', *Eng. Sci. Technol. Int. J.*, 2017, **20**, (3), pp. 1075–1083
[11] He, K., Sun, J., Tang, X.: 'Single image haze removal using dark channel prior', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2010, **33**, (12), pp. 2341–2353
[12] Zhu, Q., Mai, J., Shao, L.: 'A fast single image haze removal algorithm using color attenuation prior', *IEEE Trans. Image Process.*, 2015, **24**, (11), pp. 3522–3533
[13] Fattal, R.: 'Single image dehazing', *ACM Trans. Graphics (TOG)*, 2008, **27**, (3), p. 72
[14] Berman, D., Avidan, S.: 'Non-local image dehazing'. Proc. IEEE Conf. on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 2016, pp. 1674–1682
[15] Chen, C., Do, M.N., Wang, J.: 'Robust image and video dehazing with visual artifact suppression via gradient residual minimization'. European Conf. on Computer Vision Springer, Amsterdam, the Netherlands, 2016, pp. 576–591
[16] Meng, G., Wang, Y., Duan, J., *et al.*: 'Efficient image dehazing with boundary constraint and contextual regularization'. Proc. IEEE Int. Conf. on Computer Vision, Sydney, NSW, Australia, 2013, pp. 617–624
[17] Zhao, D., Xu, L., Yan, Y., *et al.*: 'Multi-scale optimal fusion model for single image dehazing', *Signal Process., Image Commun.*, 2019, **74**, pp. 253–265
[18] Xu, L., Zhao, D., Yan, Y., *et al.*: 'Iders: iterative dehazing method for single remote sensing image', *Inf. Sci.*, 2019, **489**, pp. 50–62
[19] Yoon, I., Jeong, S., Jeong, J., *et al.*: 'Wavelength-adaptive dehazing using histogram merging-based classification for uav images', *Sensors*, 2015, **15**, (3), pp. 6633–6651
[20] Ju, M., Gu, Z., Zhang, D.: 'Single image haze removal based on the improved atmospheric scattering model', *Neurocomputing*, 2017, **260**, pp. 180–191
[21] Li, B., Dai, Y., He, M.: 'Monocular depth estimation with hierarchical fusion of dilated cnns and soft-weighted-sum inference', *Pattern Recognit.*, 2018, **83**, pp. 328–339
[22] Zhang, Y., Li, K., Li, K., *et al.*: 'Residual non-local attention networks for image restoration'. Proc. Int. Conf. on Learning Representations, New Orleans, LA, USA, 2019, pp. 1–8
[23] Wang, F., Jiang, M., Qian, C., *et al.*: 'Residual attention network for image classification'. Proc. IEEE Conf. on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 2017, pp. 3156–3164
[24] Samadi, F., Akbarizadeh, G., Kaabi, H.: 'Change detection in sar images using deep belief network: a new training approach based on morphological images', *IET Image Process.*, 2019, **13**, (12), pp. 2255–2264
[25] Li, B., Peng, X., Wang, Z., *et al.*: 'Aod-net: all-in-one dehazing network'. Proc. IEEE Int. Conf. on Computer Vision, Venice, Italy, 2017, pp. 4770–4778
[26] Zhang, H., Sindagi, V., Patel, V.M.: 'Multi-scale single image dehazing using perceptual pyramid deep network'. Proc. IEEE Conf. on Computer Vision and Pattern Recognition Workshops, Salt Lake City, UT, USA, 2018, pp. 902–911
[27] Qu, Y., Chen, Y., Huang, J., *et al.*: 'Enhanced pix2pix dehazing network'. Proc. IEEE Conf. on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 2019, pp. 8160–8168
[28] Zhang, J., Cao, Y., Wang, Y., *et al.*: 'Fully point-wise convolutional neural network for modeling statistical regularities in natural images'. Proc. 26th ACM Int. Conf. on Multimedia, Seattle, WA, USA, 2018, pp. 984–992
[29] Zhang, J., Tao, D.: 'Famed-net: a fast and accurate multi-scale end-to-end dehazing network', *IEEE Trans. Image Process.*, 2019, **29**, pp. 72–84
[30] Ren, W., Ma, L., Zhang, J., *et al.*: 'Gated fusion network for single image dehazing'. Proc. IEEE Conf. on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 2018, pp. 3253–3261
[31] Liu, X., Ma, Y., Shi, Z., *et al.*: 'Griddehazenet: attention-based multi-scale network for image dehazing'. Proc. IEEE Int. Conf. on Computer Vision, Seoul, Republic of Korea, 2019, pp. 7314–7323
[32] Zhang, H., Patel, V.M.: 'Densely connected pyramid dehazing network'. Proc. IEEE Conf. on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 2018, pp. 3194–3203
[33] Cai, B., Xu, X., Jia, K., *et al.*: 'Dehazenet: an end-to-end system for single image haze removal', *IEEE Trans. Image Process.*, 2016, **25**, (11), pp. 5187–5198
[34] Ren, W., Liu, S., Zhang, H., *et al.*: 'Single image dehazing via multi-scale convolutional neural networks'. European Conf. on Computer Vision, Amsterdam, the Netherlands, 2016, pp. 154–169
[35] Zhang, X., Wang, T., Qi, J., *et al.*: 'Progressive attention guided recurrent network for salient object detection'. Proc. IEEE Conf. on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 2018, pp. 714–722
[36] Hu, J., Shen, L., Sun, G.: 'Squeeze-and-excitation networks'. Proc. IEEE Conf. on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 2018, pp. 7132–7141
[37] Zhang, Y., Tian, Y., Kong, Y., *et al.*: 'Residual dense network for image super-resolution'. Proc. IEEE Conf. on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 2018, pp. 2472–2481
[38] Huang, G., Liu, Z., Van Der Maaten, L., *et al.*: 'Densely connected convolutional networks'. Proc. IEEE Conf. on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 2017, pp. 4700–4708
[39] He, K., Zhang, X., Ren, S., *et al.*: 'Deep residual learning for image recognition'. Proc. IEEE Conf. on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 2016, pp. 770–778
[40] Simonyan, K., Zisserman, A.: 'Very deep convolutional networks for large-scale image recognition'. Int. Conf. on Learning Representations, Banff, Canada, 2014, pp. 1–6
[41] Yin, S., Zhao, X., Wang, W., *et al.*: 'Efficient multilevel image segmentation through fuzzy entropy maximization and graph cut optimization', *Pattern Recognit.*, 2014, **47**, (9), pp. 2894–2907
[42] Yin, S., Qian, Y., Gong, M.: 'Unsupervised hierarchical image segmentation through fuzzy entropy maximization', *Pattern Recognit.*, 2017, **68**, pp. 245–259
[43] Boykov, Y., Veksler, O., Zabih, R.: 'Fast approximate energy minimization via graph cuts', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2001, **23**, (11), pp. 1222–1239
[44] Nan, D., Bi, D., Ma, S., *et al.*: 'Single image dehazing method based on scene depth constraint', *Acta Electron. Sin.*, 2015, **43**, (3), pp. 500–504
[45] Li, B., Ren, W., Fu, D., *et al.*: 'Benchmarking single-image dehazing and beyond', *IEEE Trans. Image Process.*, 2018, **28**, (1), pp. 492–505

6 Appendices

6.1 Appendix 1

The S -function is defined as

$$S(x)=\begin{cases}1 & x\le u\\ 1-\dfrac{(x-u)^2}{(w-u)(v-u)} & u<x\le v\\ \dfrac{(x-w)^2}{(w-u)(w-v)} & v<x\le w\\ 0 & x>w\end{cases} \tag{24}$$

where u,v,w are the parameters determining the shape of S -function.

The Z-function is the opposite of the S-function which can be written as $Z(x,u,v,w)=1-S(x,u,v,w)$ and defined as

$$Z(x)=\begin{cases}0 & x\le u\\ \dfrac{(x-u)^2}{(w-u)(v-u)} & u<x\le v\\ 1-\dfrac{(x-w)^2}{(w-u)(w-v)} & v<x\le w\\ 1 & x>w\end{cases} \tag{25}$$

The M-function is derived from the S-function and the Z-function

$$M(x)=\begin{cases}Z(x,u_1,v_1,w_1) & x\le w_1\\ S(x,u_2,v_2,w_2) & x>w1\end{cases} \tag{26}$$

After substituting (24) and (25) into (26), the M-function can be rewritten as

$$M(x)=\begin{cases}0 & x\le u_1\\ \dfrac{(x-u_1)^2}{(w_1-u_1)(v_1-u_1)} & u_1<x\le v_1\\ 1-\dfrac{(x-w)^2}{(w_1-u_1)(w_1-v_1)} & v_1<x\le w_1\\ 1 & w_1<x\le u_2\\ 1-\dfrac{(x-u_2)^2}{(w_2-u_2)(v_2-u_2)} & u_2<x\le v_2\\ \dfrac{(x_2-w_2)^2}{(w_2-u_2)(w_2-v_2)} & v_2<x\le w_2\\ 0 & x>w_2\end{cases} \tag{27}$$

1. Start from an arbitrary labeling l
2. Set success :=0
3. For each pair of labels $\{\alpha,\beta\}\subset$ {"close scene","medium scene","distant scene"}
repeat
3.1. Find $\bar{l}$=argmin $T(l)$ among $\bar{l}$ within one α-β swap of l
3.2. if $T(\bar{l}) < T(l)$, set $l := \bar{l}$ and success :=1
until 4. success=1 go to 2
5. Return l

Fig. 20 *Algorithm 1: α–β swap*

where $u_1, v_1, w_1, u_2, v_2, w_2$ are parameters determining the shape of the *M*-function, which need to satisfy the constraint that $0 \le u_1 < v_1 < w_1 < u_2 < v_2 < w_2 \le 255$.

6.2 Appendix 2

For the multi-label problem, the goal is to find a labelling l which assigns each pixel p a proper label l_p. Hence, for obtaining the optimal label assignment for close, medium and distant scene sets in the transmission map, we seek the labelling l that minimises the energy function of a multi-label graph cuts

$$T(l) = T_{\text{smooth}}(l) + T_{\text{data}}(l) \tag{28}$$

where T_{data} is defined as in (19) and T_{smooth} is defined as in (20). Labelling l can be viewed as a partition of pixels $l = \{l_p \mid l_p \in \text{closescene}, \text{mediumscene}, \text{distantscene}\}$.

Here, we adopt the α–β swap algorithm to solve a multi-label graph cuts problem. Specifically, given a pair of labels α and β, a move from a partition l to a new partition $\bar{l}$ is called an α–β swap, where the pixels labelled by α in l are now labelled by β in $\bar{l}$ or some pixels labelled by β in l are now labelled by α in $\bar{l}$. Hence, the aim of the α–β swap algorithm is to find a new label $\bar{l}$ which can minimise (28) over all labellings within one α–β swap of l. The key steps of the α–β swap algorithm from [43] are summarised in Algorithm 1 (see Fig. 20.).

From Algorithm 1 (Fig. 20), we find that steps 2, 3 and 4 form a cycle. In each cycle, the α–β swap algorithm performs an iteration for every pair of labels. A cycle is successful when a better labelling is found at any iteration. Otherwise, the algorithm stops once the first unsuccessful cycle is obtained since there is no further improvement.