



# Self-supervised zero-shot dehazing network based on dark channel prior

Xinjie Xiao<sup>1</sup> · Yuanhong Ren<sup>2</sup> · Zhiwei Li<sup>1</sup> · Nannan Zhang<sup>1</sup> · Wuneng Zhou<sup>2</sup>

Received: 2 November 2022 / Accepted: 5 February 2023  
© The Author(s) 2023

## Abstract

Most learning-based methods previously used in image dehazing employ a supervised learning strategy, which is time-consuming and requires a large-scale dataset. However, large-scale datasets are difficult to obtain. Here, we propose a self-supervised zero-shot dehazing network (SZDNet) based on dark channel prior, which uses a hazy image generated from the output dehazed image as a pseudo-label to supervise the optimization process of the network. Additionally, we use a novel multichannel quad-tree algorithm to estimate atmospheric light values, which is more accurate than previous methods. Furthermore, the sum of the cosine distance and the mean squared error between the pseudo-label and the input image is applied as a loss function to enhance the quality of the dehazed image. The most significant advantage of the SZDNet is that it does not require a large dataset for training before performing the dehazing task. Extensive testing shows promising performances of the proposed method in both qualitative and quantitative evaluations when compared with state-of-the-art methods.

**Keywords** Image dehazing · Quad-tree algorithm · Self-supervised · Zero-shot

## 1 Introduction

Haze is a special weather condition that the sky becomes blurred due to micron-sized particles suspended in the atmosphere. These suspended particles scatter and absorb light, thus deteriorating the visual clarity of an image, with image contrast degradation and color distortion. The images captured in hazy scenes significantly affect the performances

of computer vision, such as target detection [1, 2] and scene understanding [3, 4]. Therefore, image dehazing has long-term importance in computer vision.

Most deep learning-based methods [5–8] for dehazing use a supervised learning strategy, which requires a large-scale dataset and corresponding ground-truth images. These methods update network parameters by calculating the loss between the image output of the dehazing network and ground-truth images. Meanwhile, network training is time-consuming and data collection is challenging. Lightweight networks [9, 10] have been proposed to reduce training time and address the issues with conventional supervised learning, by reducing the number of parameters of the neural networks (NNs). For example, Suresh et al. [10] proposed a lightweight dehazing network that uses a pre-trained teacher network to extract multi-scale information for the student dehazing network, resulting in significant parameter reduction compared to previous networks that had millions of parameters, and thus speed-up training. However, the training process of the networks remains very time-consuming and requires a large number of ground-truth images.

Self-supervised learning [11–13] has been proposed to overcome the problem with collecting ground-truth images. For instance, Wang et al. [14] proposed a physics-enhanced deep neural network that combines NNs and

---

Xinjie Xiao and Yuanhong Ren contributed equally.

✉ Zhiwei Li  
zhiwei.li@sues.edu.cn  
Xinjie Xiao  
xinjie.xiao@sues.edu.cn  
Yuanhong Ren  
renyuanhong@sohu.com  
Nannan Zhang  
nannan.zhang@sues.edu.cn  
Wuneng Zhou  
wnzhou@dhu.edu.cn

<sup>1</sup> School of Electronic and Electrical Engineering, Shanghai University of Engineering Science, Shanghai 201620, China

<sup>2</sup> College of Information Science and Technology, Donghua University, Shanghai 201620, China

physical models. They fed a diffraction pattern into the NN and applied the Huygens-Fresnel principle to construct a pseudo-diffraction pattern as a label to supervise the learning process. Li et al. [15] proposed a zero-shot image dehazing (ZID) network that uses three sub-networks to generate a hazy image for guiding network training, but the dehazed images outputted by the ZID network are have low brightness and color distorted.

To address the problems in these previous methods, we propose a self-supervised zero-shot dehazing network (SZDNet) using dark channel prior. The image output of the NN is used to generate a hazy pseudo-label using the physical model. We update the NN parameters with a loss function to improve the dehazing ability. This method saves time and labor by removing the need for a large-scale dataset and ground-truth images to train the NN before using it to conduct dehazing. The physical model used by the SZDNet is the atmospheric scattering model (ASM) [16, 17], and we estimate the transmission map based on the dark channel prior (DCP) theory. He et al. [18] considered the mean of the globally brightest 0.1% pixels as the atmospheric light, but the method has a significant deviation. Therefore, Li and Zheng [19] estimated atmospheric light using a quad-tree method and achieved excellent performance. Nevertheless, this method is prone to failure in a scene with numerous white objects. To improve the precision and stability of locating the brightest sky region, we design a quad-tree algorithm with multiple channels, supported by the understanding that the sky region is primarily distributed in the top half of the image. Typical loss functions for deep learning dehazing algorithms [5, 6] include the mean-squared error (MSE) which achieves improved performance by accurately evaluating the pixel-level difference between the dehazed image output from the network and ground-truth image. However, the MSE ignores the variations between the row vectors and column vectors of two images. Thus, we improved the quality of dehazed images by adding the cosine distance to the loss function.

Our method differs considerably from the existing supervised deep learning dehazing methods [5–8], in that it has robust performances of the NNs and exhibits theoretical support from physical models. Notably, the concept of “zero-shot” is that feeding a hazy image into the NN results in a dehazed image. In brief, our method neither requires a large-scale dataset to train the NN before dehazing, nor requires the collection of corresponding haze-free images.

The main contributions of this work can be summarized as follows:

- We propose a novel self-supervised zero-shot dehazing network (SZDNet) based on DCP that outperforms some state-of-the-art (SOTA) methods.
- In accordance with the understanding that the sky region is primarily distributed in the top half of the image, we design a novel multichannel quad-tree algorithm to estimate atmospheric light values more accurately than the previous methods.
- The sum of the cosine distance and MSE is designed as the loss function of the NN, based on which the difference between pixels and the similarity between row vectors and column vectors of two images can be estimated, thereby improving the quality of the dehazed images.

## 2 Related works

Recently, many methods for single-image dehazing have been developed, and excellent results have been obtained. These methods can be classified into prior-based [20–22] and learning-based methods [5–8].

### 2.1 Prior-based image dehazing methods

Most prior-based dehazing methods use prior knowledge observed in the real world and apply them to the ASM [16, 17] for image dehazing. These methods include DCP [18], color attenuation prior [20], contrast maximization [21], and non-local prior [22]. Despite excellent performances achieved by prior-based methods, priors have some limitations because they are obtained under some assumptions and certain target scenes. For example, color attenuation prior [20] treats the scattering coefficient that varies with the scene depth as a constant, thereby affecting dehazing in some cases.

### 2.2 Learning-based image dehazing methods

Learning-based dehazing methods have been developed recently and achieved excellent dehazing results. Learning-based methods can be classified into two groups: The first group integrates the ASM and estimate the dehazing process parameters using NN [5, 6]. The second group learns straight mapping from hazy-to-clear images [7, 8]. For example, DehazeNet [5] combined with an ASM took hazy images as the input and estimated the medium transmission map using a NN to restore dehazed images. The enhanced pix2pix dehazing network (EPDN) [8] followed the hazy-to-clear image translation approach and generated dehazed images without relying on ASMs. However, it is difficult and expensive to collect large-scale datasets and ground-truth images sufficient for such learning-based dehazing methods.

### 2.3 Advances in self-supervised learning methods

Self-supervised learning methods [11–15] have made remarkable achievements in image processing because they do not require manual labels. For example, Hendriksen et al. [12] proposed a novel “Noise2Inverse” strategy which uses the noise model to compute multiple statistically independent reconstructions. This method uses self-supervised denoising to obtain improved denoising performance. In Ref. [13], Chen et al. proposed a context restoration strategy to better utilize unlabelled images. Self-supervised learning methods are independent of ground-truth images and have achieved promising performances. Although the self-supervised learning method is widely used in many other fields, it is rarely used in image dehazing. A ZID network [15] comprising three joint subnetworks to disentangle the input hazy image was recently proposed and has achieved excellent dehazing performance. However, the dehazed image using ZID has low brightness.

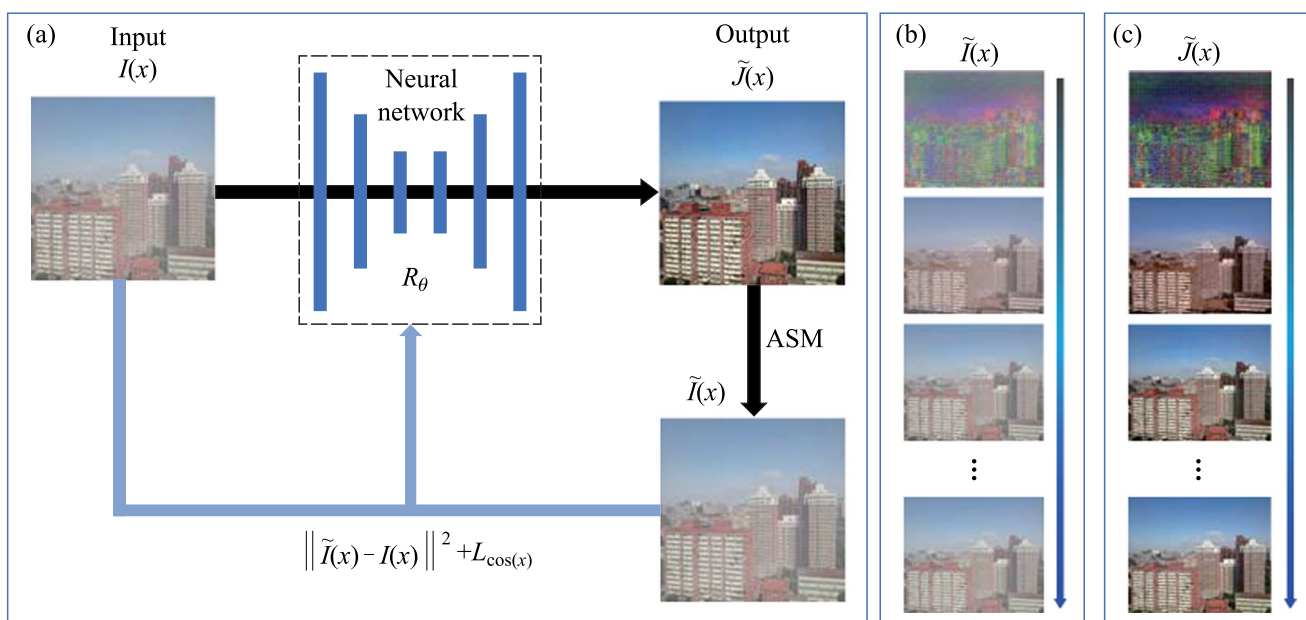
The SZDNet proposed in this study differs from these methods and does not require a large dataset for training. We only used the U-Net [23] as the backbone of the network. First, we fed a hazy image into the network, and the hazy pseudo-label was generated by processing the output of the U-Net. Then, we optimized the parameters of U-Net using the loss function. Experimental results showed that SZDNet achieved better performance than some SOTA methods.

## 3 Proposed method

The network architecture of SZDNet and its various components are described first in this section. Then, we introduce a multichannel quad-tree algorithm to find the atmospheric light value  $A$ , and finally, we construct the loss function to update the weights and biases.

### 3.1 SZDNet system architecture

Many image dehazing methods [5–8] make use of supervised learning with large-scale datasets. However, acquiring large datasets and ground-truth images is challenging. To address this problem, we proposed a self-supervised zero-shot dehazing network (SZDNet) using DCP. The system architecture and basic principle are schematically outlined in Fig. 1a. Here, we use the U-Net as the NN. First, we fed a hazy image  $I(x)$  into the NN to generate  $\tilde{J}(x)$  after NN processing. In conventional supervised dehazing NNs, the NN parameters are optimized using the loss between  $\tilde{J}(x)$  and ground-truth image  $J(x)$ . The proposed SZDNet does not rely on ground-truth images, but it uses the DCP and the multichannel quad-tree algorithm to derive the transmission map  $t(x)$  and the atmospheric light value  $A$ , respectively. The principle will be expressed in the next section. We fed  $\tilde{J}(x)$  into a physical ASM, generating a hazy pseudo-label  $\tilde{I}(x)$ . The loss between the pseudo-label  $\tilde{I}(x)$  and the input



**Fig. 1** Principle and architecture diagram of the SZDNet. **a** A hazy image  $I(x)$  is fed into the neural network, and a dehazed image  $\tilde{J}(x)$  is output after processing by the neural network, which is numerically treated to simulate the hazy image generation processes through the ASM to generate  $\tilde{I}(x)$ . We use the sum of the MSE and the cosine distance ( $L_{\cos(x)}$ ) between  $I(x)$  and  $\tilde{I}(x)$  as the loss function to update the parameter  $R_\theta$  of the neural network. **b** Evolution of  $\tilde{I}(x)$  during the optimization process. **c** Evolution of  $\tilde{J}(x)$  during the optimization process

image  $I(x)$  optimizes the network parameters, forcing the generated pseudo-label  $\tilde{I}(x)$  to converge to  $I(x)$ . Figure 1b shows the optimization process for the pseudo-label  $\tilde{I}(x)$ . With the iterative process, the optimized dehazed image  $\tilde{J}(x)$  converges to a near-optimal solution. Figure 1c shows the optimization process of the dehazed image. In summary, the proposed method adopts zero-shot learning and removes haze using only one hazy image.

U-Net has encoders and decoders and can capture a lot of spatial information. Here, the main function of U-Net is to extract image features, reconstruct clear images and refine the clear images using loss function through iterative optimization. The architecture of U-Net network is shown in Fig. 2, which consists of encoders and decoders. In the encoder network, with each down-sampling operation, the image size is reduced by half and the dimensionality is doubled. By repeating this operation, the higher-level features of the image can be fully extracted and the redundant information can be filtered out. In the decoder network, for each up-sampling operation, the image size is doubled and dimensionality is halved. A skip connection structure is used between the corresponding stages of the encoder and decoder to supplement the low-level feature information and to better restore the image detail information.

Below, we introduce the technical details of the SZDNet. In computer vision and graphics, the ASM [16, 17] accounts for hazy image generation. Its mathematical model is given as

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

where  $x$  denotes the spatial coordinate,  $I(x)$  is the hazy image,  $J(x)$  is the haze-free ground-truth image,  $t(x)$  is the transmission map,  $A$  is the atmospheric light, and  $ASM(\cdot)$  denotes the mapping function from the haze-free ground-truth image  $J(x)$  to the corresponding hazy image  $I(x)$ .

From the ASM, we obtain  $J(x)$  as

$$J(x) = ASM^{-1}(I(x)) = \frac{I(x) - A(1 - t(x))}{t(x)}. \tag{2}$$

The transmission map obtained by the DCP is given by

$$t_1 = 1 - w \min_{y \in \Omega(x)} \left[ \min_c \frac{I^c(x)}{A^c} \right], \tag{3}$$

$$t(x) = G(t_1(x)), \tag{4}$$

where  $t_1$  is the coarse transmission map,  $w$  is the correction factor and  $0 < w \leq 1$ ,  $\Omega(x)$  denotes the set of spatial

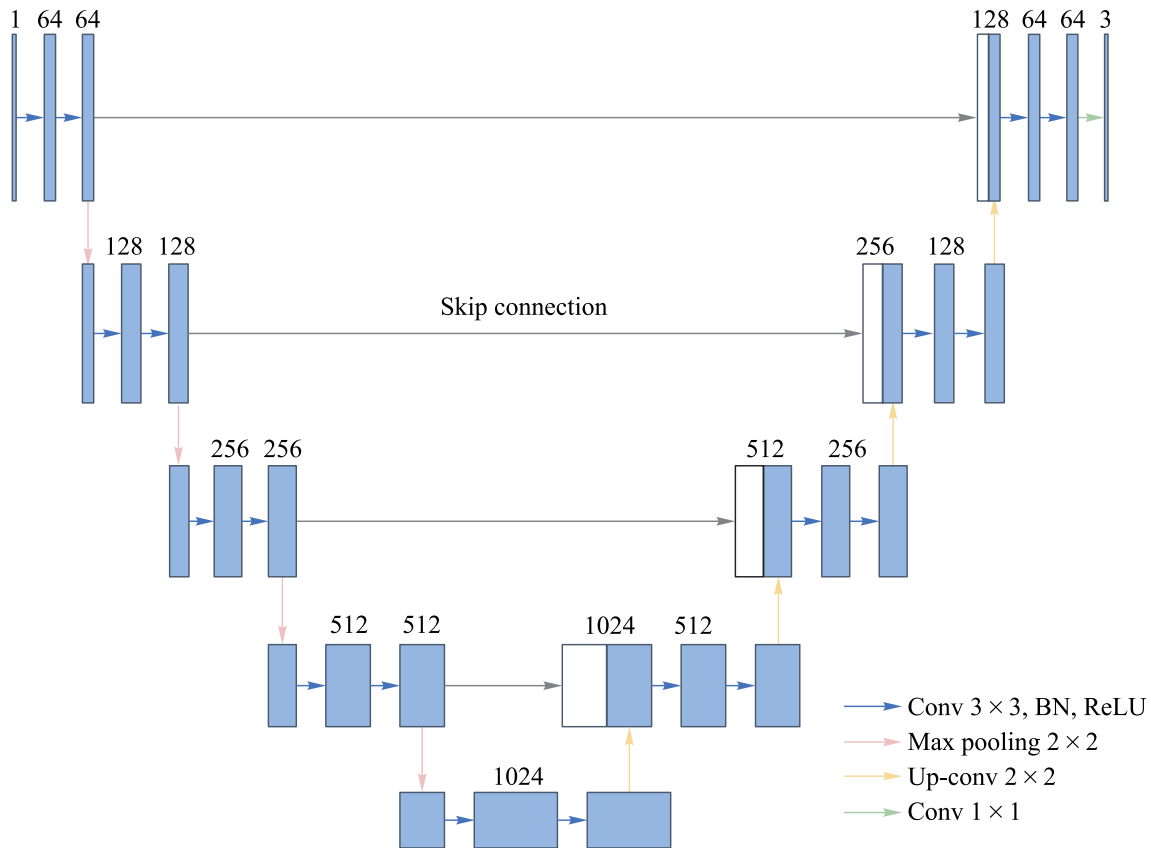


Fig. 2 Architecture diagram of the U-Net

coordinates  $x, c$  denotes the three color channels R, G, and B, and  $G(\cdot)$  denotes the guide filter operator [24, 25] that can refine the transmission map.

In the proposed method, the atmospheric light  $A$  is obtained using a multichannel quad-tree search algorithm described in the next section, and the mathematical formula for the pseudo-label  $\tilde{I}(x)$  is given by

$$\tilde{I}(x) = \tilde{J}(x)t(x) + A(1 - t(x)) = ASM(\tilde{J}(x)), \tag{5}$$

where  $\tilde{J}(x)$  denotes the dehazed image output from the neural network.

### 3.2 Estimation of atmospheric light value

Previous dehazing algorithms focused on optimizing the transmission map, while ignoring the importance of the atmospheric light value. Since a large amount of haze increases the brightness of the scene target, Tan [21] used the maximum pixel value of the dense haze region in the image as the atmospheric light value, considering the

characteristics of haze. He et al. [18] applied the mean of the globally brightest 0.1% pixels as the atmospheric light value. Li and Zheng [19] proposed a quad-tree method to estimate the atmospheric light value. Although their methods improved the accuracy of estimating atmospheric light values, some deviations exist for scenes with several white objects, and sometimes white objects are selected instead of the brightest areas of the sky. Inspired by the work of Wang et al. [26], and considering the results from the quad-tree search algorithm and the prior knowledge that the sky region lies on the upper part of an image, we developed a multichannel quad-tree search algorithm to improve the precision and stability of locating the brightest sky region. Figure 3 shows the flowchart of the multichannel quad-tree algorithm. The image is divided equally into four regions  $t_n^i$ ,  $i \in \{1, 2, 3, 4\}$  representing the left and right quarters of the top and bottom halves, respectively.  $n$  denotes the number of division level,  $n = 1$  is the first division. We define  $I(t_n^i)$  as a matrix that contains the corresponding gray values for each pixel in region  $t_n^i$ , and use the average of  $I(t_n^i)$  as the score  $M(t_n^i)$  for this region, as shown in the equation below:

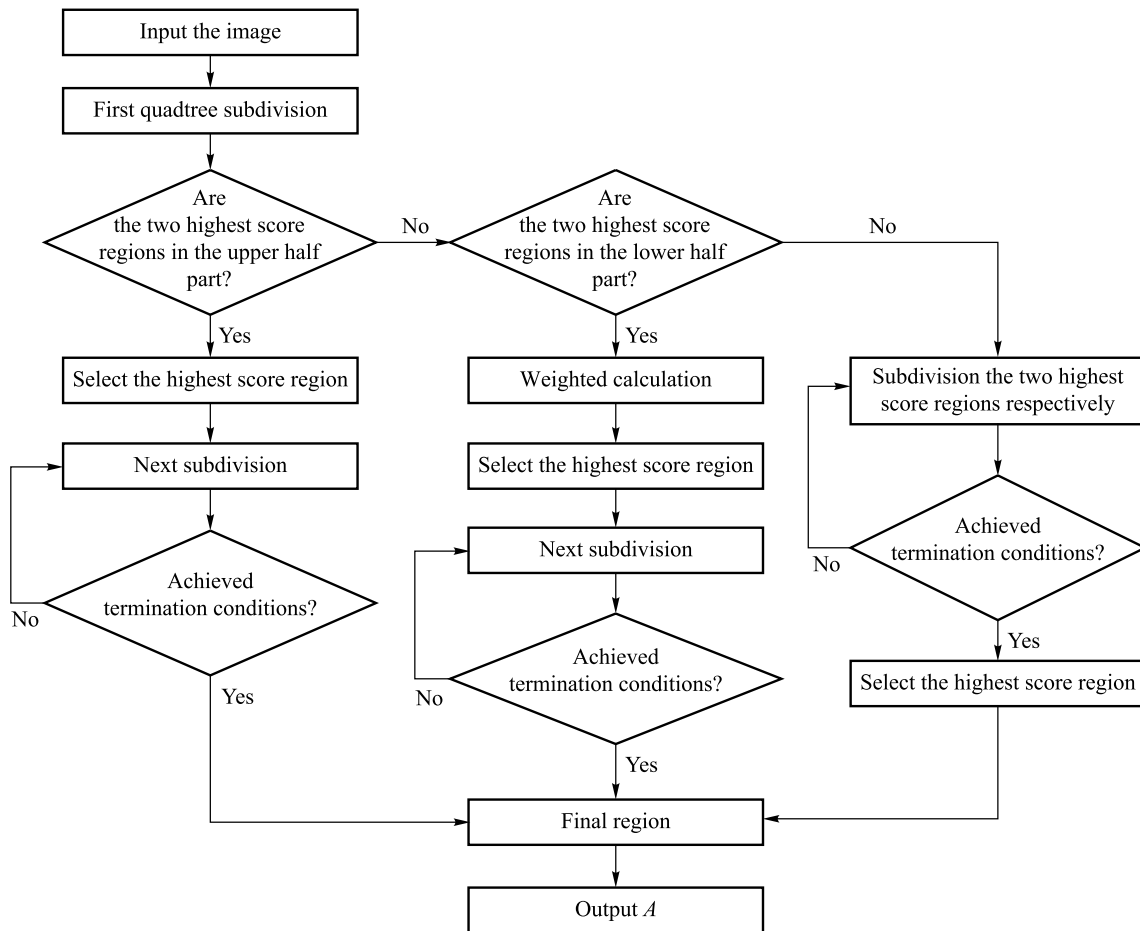


Fig. 3 Flowchart of multichannel quad-tree algorithm to estimate atmospheric light value

$$M(t_n^i) = \text{mean}(I(t_n^i)). \quad (6)$$

Suppose the two regions with the highest scores after the initial division are in the upper half of the image region  $(t_1^1, t_1^2)$ , then the region with the highest score is considered the part to be processed in the next iteration. This is done by dividing it into smaller blocks and the quad-tree decomposition method and Eq. (6) are used to calculate the scores. The iteration continues and terminates when the score becomes smaller than the predefined threshold, to get the final region  $t_{\text{final}}$ . If the two regions with the highest scores after the initial division are in the bottom half region  $(t_1^3, t_1^4)$  of the image, the upper half region must be weighted. Suppose the weighting factor is  $e (e > 1)$ , then, the scores of  $\{e \times M(t_1^1), e \times M(t_1^2), M(t_1^3), M(t_1^4)\}$  are compared, and the selected region with highest score continues to be subdivided accordingly using the quad-tree method until the termination condition is reached to obtain  $t_{\text{final}}$ . Otherwise, the two highest-scoring regions are subdivided respectively, and the highest final score region is chosen as  $t_{\text{final}}$  when the termination condition is met.

The quad-tree subdivision process has a termination condition that states that the iteration terminates if the difference between the highest and the second-highest scores is less than  $M_T$  or the width of the highest score region is less than  $w$ . The termination conditions are expressed as

$$\min |M(t_n^k) - M(t_n^m)| < M_T, \quad (7)$$

or

$$W(t_n^k) < w, \quad (8)$$

where  $\min$  is the minimum operator,  $M(t_n^k)$  denotes the highest score,  $M(t_n^m)$  denotes the second-highest score, and  $W(t_n^k)$  is the width of the region with the highest score. From the experimental data, we set  $e = 1.1$ ,  $M_T = 1$ ,  $w = 25$  for image dehazing.

### 3.3 Loss function for SZDNet

MSE loss function is an effective tool employed by the conventional supervised learning dehazing network. However, it only evaluates pixel-level changes and does not examine variations between the row vectors and column vectors of hazy and ground-truth images. Gao et al. [27] used the cosine similarity between the generated and target images to represent the similarity between the identity information of both images, thus to solve the problem of determining how to swap the identity information when dealing with the face-swapping task. Inspired by their work, we introduced the cosine distance into the

loss function in this study to address the limitations of MSE loss function. Figure 4 shows the cosine similarity between the hazy image and the pseudo-label. The horizontal and vertical axes in Fig. 4 represent the input hazy image  $I(x)$  and the dehazed image  $\tilde{J}(x)$  at the output of NN, respectively. The yellow region marked the case that the angle between  $I(x)$  and  $\tilde{J}(x)$  is a right angle and the cosine similarity is 0. The overlap of the two sector regions marked by the green dashed curves indicate that the generated pseudo-label  $\tilde{I}(x)$  is near the angle bisector between  $I(x)$  and  $\tilde{J}(x)$ , and is not similar to  $I(x)$ .

A high quality pseudo-label  $\tilde{I}(x)$  should be close enough to the input hazy image  $I(x)$ , thus should be within a small region centered on  $I(x)$ , eg., between  $\tilde{I}_1(x)$  and  $\tilde{I}_2(x)$  as shown in Fig. 4. Here, the cosines of  $\tilde{I}_1(x)$  and  $\tilde{I}_2(x)$  with  $I(x)$  are 0.9.

Thus, the loss function is as follows

$$\text{Loss} = \|\tilde{I}(x) - I(x)\|^2 + L_{\cos(x)}, \quad (9)$$

$$L_{\cos(x)} = 1 - \cos(I(x), \tilde{I}(x)), \quad (10)$$

where  $L_{\cos(x)}$  and  $\|\tilde{I}(x) - I(x)\|^2$  denote the cosine distance and the MSE, respectively.  $\cos(I(x), \tilde{I}(x))$  denotes the cosine similarity between  $I(x)$  and  $\tilde{I}(x)$ . In the optimization process, the more similar  $\tilde{I}(x)$  is to  $I(x)$ , the more the  $\cos(I(x), \tilde{I}(x))$  converges to 1, the smaller the  $L_{\cos(x)}$ , and the smaller the loss function.

## 4 Experiments

We experimented the proposed method using three datasets and compared the results with the SOTA methods based on four evaluation metrics. The experimental settings are given

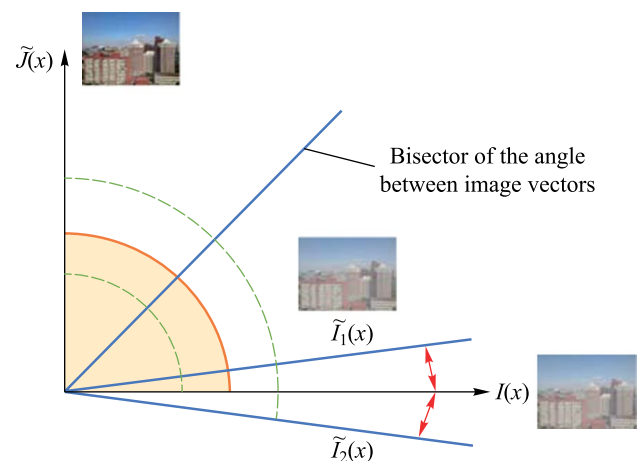


Fig. 4 Cosine similarity between the hazy image and the pseudo-label

first in this section. Next, we demonstrate the superiority of the multichannel quad-tree search algorithm in estimating atmospheric light  $A$ . We show quantitative and qualitative results on the three datasets, and finally conduct an ablation experiment.

## 4.1 Experiment settings

This section presents the datasets used, the evaluation metrics, and the implementation details.

### 4.1.1 Datasets

Our method was evaluated on two real-world datasets and one synthetic dataset. For the synthetic dataset, we employed the synthetic objective testing set (SOTS) from a large-scale benchmark REalistic Single Image DEhazing (RESIDE) [28], which contains 1000 hazy images and corresponding ground-truth images. We randomly selected 100 pairs of images from SOTS for testing. For the real-world datasets, we employed 35 and 45 hazy images and the corresponding ground-truth images from I-HAZE [29] and O-HAZE [30] benchmark, respectively. The hazy images of the real-world datasets were generated using real and machine haze. Also, both the images with and without the haze were captured under the same lighting conditions.

### 4.1.2 Evaluation metrics

To evaluate the performance of the proposed SZDNet, we employed natural image quality evaluator (NIQE) [31], fog aware density evaluator (FADE), peak signal-to-noise ratio (PSNR), and structural similarity (SSIM), all of which are common evaluation metrics in image dehazing. FADE and NIQE are non-reference metrics, whereas PSNR and SSIM were compared with ground-truth images. A smaller FADE and NIQE represent less haze residue and higher perceptual quality, respectively. We compared the proposed method with other prior-based methods (e.g., DCP [18], dark direct attenuation prior (DDAP) [32]), methods based on enhanced atmospheric scattering model (e.g., image dehazing and exposure (IDE) using an enhanced atmospheric scattering model [33]), supervised learning-based methods (e.g., DehazeNet [5], AOD-Net [6], EPDN [8], TCN [34]), and unsupervised learning-based methods (e.g., RefinedNet [35], and USID-Net [36]).

### 4.1.3 Implementation details

The experiments were conducted on a computer with Inter Core i7-11700K and NVIDIA GeForce RTX 3090. The NN was implemented with TensorFlow version 1.15 using Python 3.6.0 on an Ubuntu 20.04.1 LTS system. The training

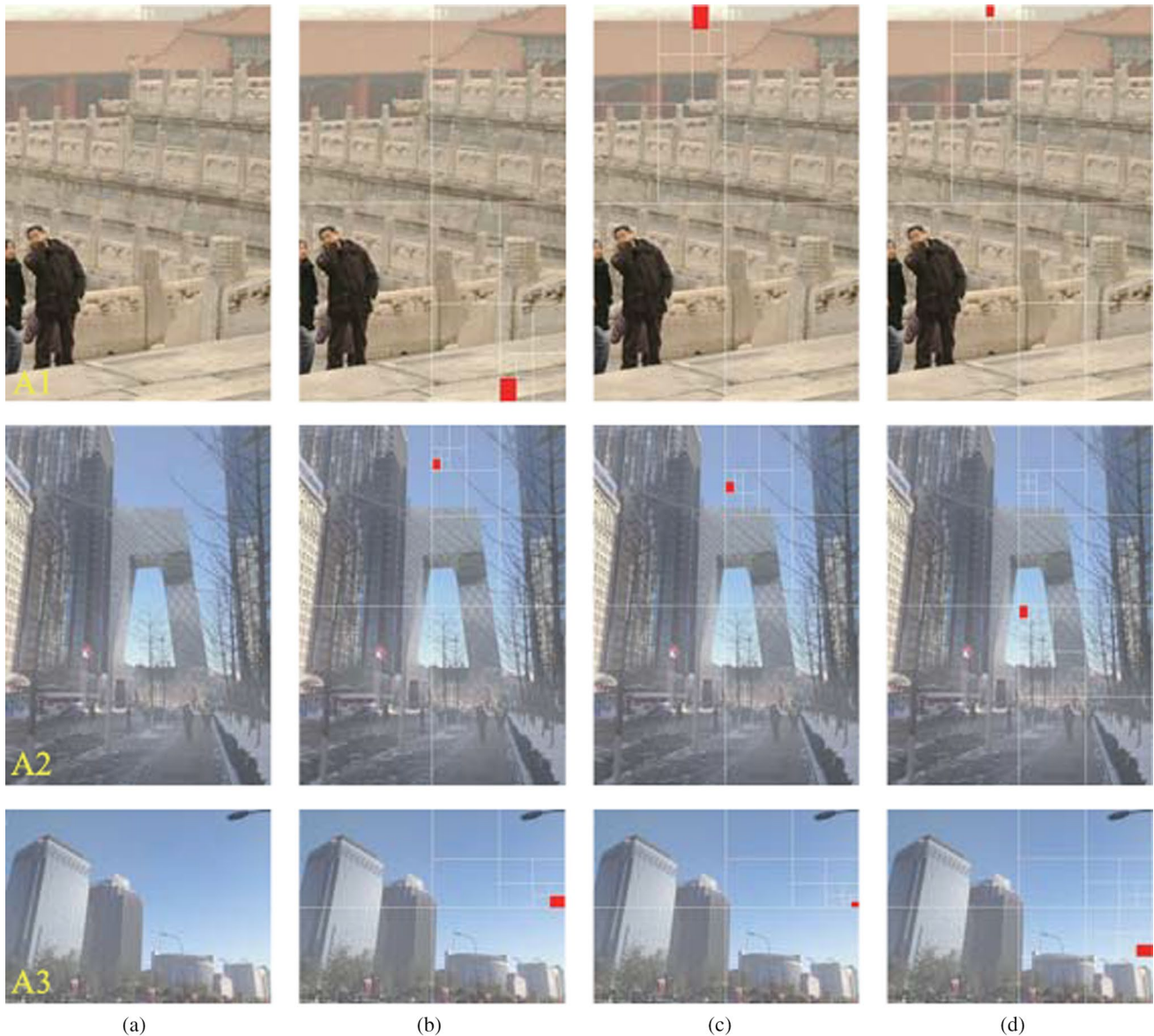
process gets the dehazed result directly from single frame training and in-depth studies reveal that the quality of the dehazed images reached a desirable performance when the number of iterations reaches 1000. The total time consumption that the proposed method to process 1000 hazy images is 44,000 s, and the average time consumption of processing each image is 44 s. We optimized the weights and biases of the NN using the Adam optimizer [37] and an initial learning rate of 0.001. Also, we added a uniformly distributed noise of 0 to 1/30 for better conversion [38]. Furthermore, we used an exponential-decay strategy with a decay rate of 0.95 to adjust the learning rate. We removed the noise and generated a dehazed image to complete the optimization.

## 4.2 Comparison of atmospheric light estimation

The dehazing results are positively impacted by the application of an accurate atmospheric light value. We developed a multichannel quad-tree search algorithm that improves the accuracy and robustness of locating the brightest sky region and estimates the atmospheric light values. We compared our results with that of the algorithms reported in Refs. [19] and [26], as shown in Fig. 5, where three experimental images A1, A2 and A3 are used and the marked red regions represent the brightest sky regions searched out. The method in Ref. [19] achieves excellent performance but fails to select the correct sky region when the image contains several white objects (see image A1 in Fig. 5b). The algorithm in [26] can locate the sky region, but not the brightest sky region (see images A2 and A3 in Fig. 5c). Our method outperformed these methods by accurately locating the brightest sky region (Fig. 5d). Moreover, our multichannel quad-tree search algorithm considers an image with the sky as the main component, therefore, it still works even the brightest sky region is located in the lower middle of the image.

## 4.3 Comparisons on a synthetic dataset

Table 1 shows the quantitative comparison on the SOTS dataset, and Fig. 6 shows the corresponding qualitative comparison. Table 1 summarizes the metrics compared with some SOTA methods using a test set of 100 randomly selected images from SOTS of RESIDE. The results show that SZDNet is superior to all methods in PSNR and SSIM, and achieves the second-best performance in FADE compared with SOTA methods. Although SZDNet is slightly inferior to IDE, DDAP, TCN, and RefinedNet in NIQE, it demonstrates the best visual performance in Fig. 6. For example, compared to the ground-truth images, IDE and TCN showed color distortions despite achieving satisfactory visual results (Fig. 6c, h). The dehazed images output by DDAP and RefinedNet are too low brightness (Fig. 6d, i). Figure 6k shows the recovered results using SZDNet. The



**Fig. 5** Comparison of atmospheric light estimation with SOTA methods. **a** Hazy images. **b–d** Brightest sky regions (marked by red rectangles) obtained using the method reported by **(b)** Li et al. [19], **(c)** Wang et al. [26], and **(d)** proposed method

**Table 1** Quantitative comparison between SZDNet and SOTA methods on the SOTS dataset. The best, second best, and third best performances are marked in red, blue, and green, respectively

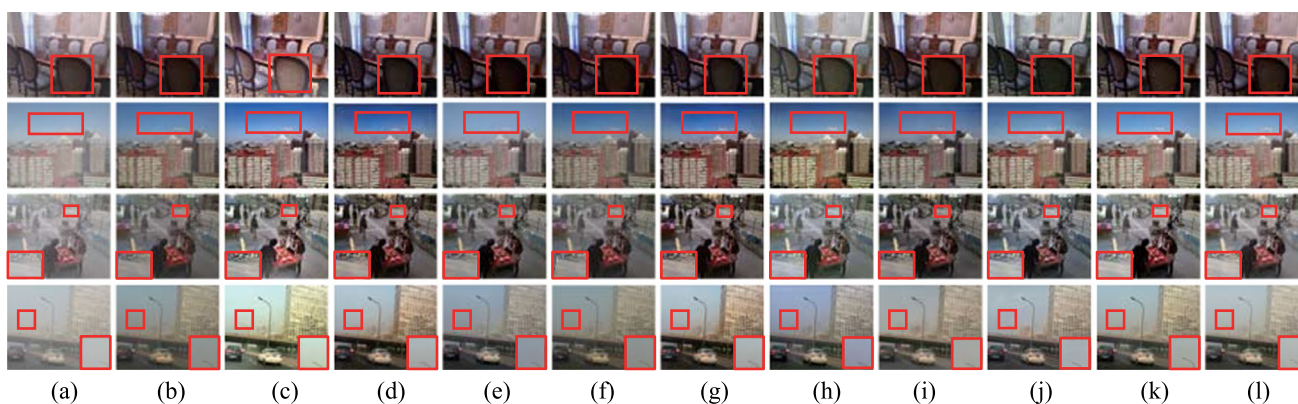
Type	Methods	PSNR	SSIM	FADE	NIQE
Prior	DCP	22.291	0.8866	2.0287	3.628
	IDE	15.375	0.7937	1.6513	3.3733
	DDAP	18.445	0.8524	<b>1.3709</b>	<b>3.2794</b>
	DehazeNet	<b>24.492</b>	<b>0.9075</b>	1.5632	3.5522
Supervised	AOD-Net	21.095	0.8922	1.6127	3.5221
	EPDN	<b>22.731</b>	0.8953	<b>1.5329</b>	3.7192
	TCN	18.681	0.8486	1.9723	<b>3.3128</b>
	RefineDNet	22.05	<b>0.8991</b>	1.6175	<b>3.1406</b>
Unsupervised	USID-Net	22.362	0.8528	1.8829	3.7472
	SZDNet	<b>24.502</b>	<b>0.9082</b>	<b>1.5201</b>	3.4759

images recovered by SZDNet have better visibility and are similar to the ground-truth images. We conclude that more attention should be paid to the results of comparison with ground-truth images for a dehazing task. By combining the results of the four evaluation metrics, SZDNet achieved a more satisfactory performance.

#### 4.4 Comparisons on real-world datasets

Tables 2 shows the quantitative comparison of SZDNet and the SOTA methods, on the I-HAZE and O-HAZE datasets.





**Fig. 6** SZDNet and well-known methods were compared qualitatively on the SOTS dataset. **a** Hazy images. **b** DCP [18]. **c** IDE [33]. **d** DDAP [32]. **e** DehazeNet [5]. **f** AOD-Net [6]. **g** EPDN [8]. **h** TCN [34]. **i** RefineDNet [35]. **j** USID-Net [36]. **k** SZDNet. **l** Ground-truth. Red rectangles indicate certain areas that are suggested to be viewed in closer detail for improved visualization and comparison

From Table 2, SZDNet achieved the best performance in PSNR and SSIM on the test dataset I-HAZE compared with the SOTA methods. The SZDNet achieved the best and second-best results in SSIM and NIQE compared with SOTA methods on the test dataset O-HAZE. SZDNet achieved better performance because atmospheric light values can be estimated more accurately, and including the cosine distance in the loss function reduces the difference between the hazy input images and pseudo-label images.

Additionally, the quantitative and qualitative comparison results are further displayed for three selected experimental images T1, T2, and T3. Figure 7 shows that the dehazed images with DCP [18], DDAP [32], DehazeNet [5], and AOD-Net [6] had lower brightness compared with the ground-truth images (see T1, Fig. 7b, d–f). The dehazed images with IDE [33], EPDN [8], TCN [34], RefineDNet [35], and USID-Net [36] showed color distortion (Fig. 7c, g–j). SZDNet performed better in dehazing, and the restored images were the most similar to the ground-truth images (T1 in Fig. 7k, l). Also, we specified the corresponding performance metrics in Table 3, showing that the proposed SZDNet achieved the best or the second-best results in PSNR and SSIM except for T2. FADE scores of our method are close to the best methods. Despite IDE, DDAP, and AOD-Net

obtaining better FADE scores, the dehazed images were color distorted or have low brightness (T1 and T3 in Fig. 7c, d, f). Thus, FADE is not the most appropriate evaluation metric. Although NIQE of our method was slightly inferior compared to other methods, our dehazed results were the closest to the ground-truth images. Upon combining all four evaluation metrics and the qualitative comparisons, SZDNet achieved a more satisfactory performance.

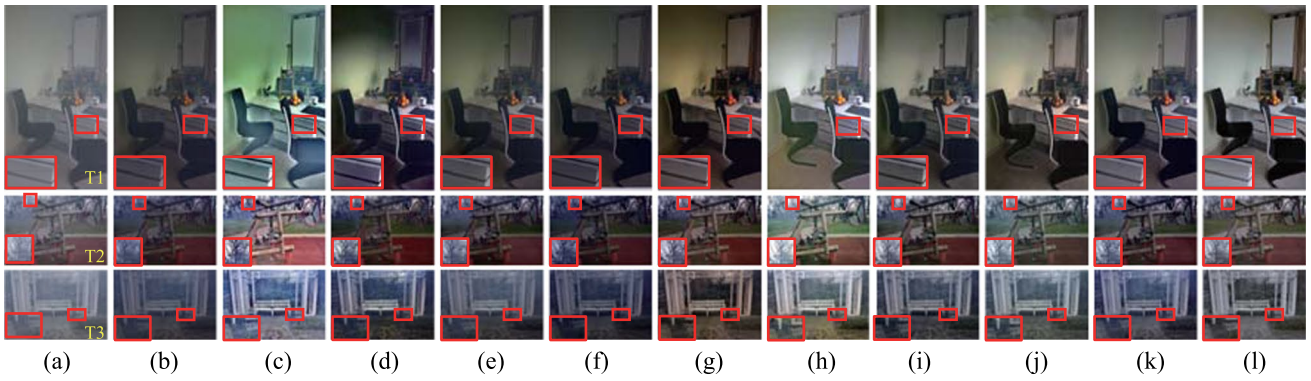
### 4.5 Comparisons on the real-shot images

As NN trained using the supervised approach perform better on the synthetic dataset than on real-shot images, we further compared SZDNet with the prior-based and learning-based approaches on the real-shot images. We did not quantitatively compare the dehazed images in Fig. 8 because the ground-truth images of real-shot hazy images were unavailable.

Figure 8 shows that all methods perform well on image R1. The dehazing effect in the marked region in the upper left corner with DCP, IDE, DehazeNet, AOD-Net, EPDN, TCN, and USID-Net was not good, whereas images processed by DDAP, RefineDNet, and the proposed SZDNet had better visibility and dehazing effect (see image R1 in Fig. 8d, i, k). As shown by the red rectangular region in image R2 of Fig. 8b, d–g, contrast of the trees in the middle of the image is sacrificed with DCP, DDAP, DehazeNet, AOD-Net, and EPDN. For IDE, the entire image showed color distortion. TCN, RefineDNet, and USID-Net could not entirely remove the haze (Fig. 8c, h–j). However, as observed in Fig. 8k, SZDNet performed the best compared with the other methods and the trees in the middle region are visible. In the marked red regions in image R3 (Fig. 8), the trees were the clearest with the SZDNet (Fig. 8k). In contrast, the brightness with DCP and AOD-Net were low that with IDE was too large, the dehazed image with DehazeNet and USID-Net had residual haze on the image, and the trees in the middle of the

**Table 2** Quantitative comparison between SZDnet and SOTA methods on the I-haze and O-haze datasets. The best, second best, and third best performances are marked in red, blue, and green, respectively

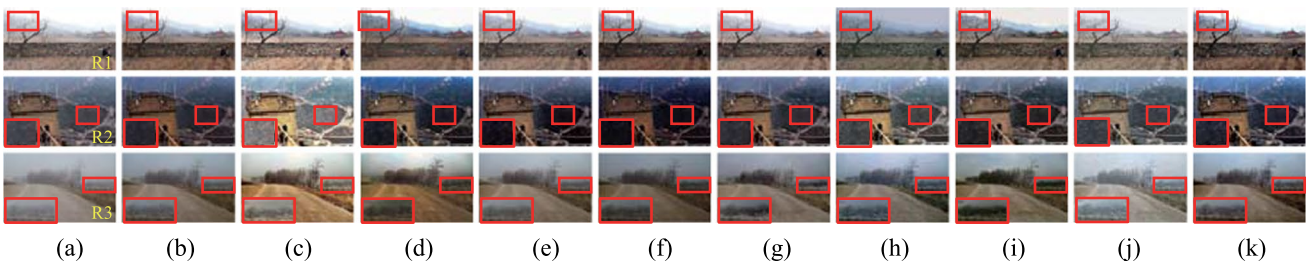
Type	Methods	I-HAZE			O-HAZE				
		PSNR	SSIM	FADE	NIQE	PSNR	SSIM	FADE	NIQE
Pior	DCP	15.259	0.6973	2.7171	4.4242	15.947	0.5583	2.0011	2.9058
	IDE	14.702	0.6998	1.9006	<b>3.2013</b>	13.423	0.6219	1.3323	2.8346
	DDAP	10.858	0.5417	<b>1.635</b>	3.3199	12.713	0.5755	<b>1.3204</b>	2.6011
Supervised	DehazeNet	15.208	0.6746	2.2773	4.0103	15.391	0.5754	2.1666	2.8111
	AOD-Net	14.732	0.6733	2.2585	3.9103	15.003	0.5187	1.5416	2.8524
	EPDN	14.947	0.6869	1.9152	3.8454	<b>16.906</b>	0.6342	1.7758	2.7913
	TCN	<b>16.45</b>	<b>0.7139</b>	2.4685	3.3312	14.704	<b>0.6436</b>	1.6671	<b>2.5448</b>
Unsupervised	RefineDN	15.781	<b>0.7121</b>	2.2313	3.2719	16.616	0.5904	1.7771	<b>2.2353</b>
	USID-Net	15.165	0.6209	2.7707	3.6779	15.594	<b>0.6446</b>	2.0286	2.7258
	SZDNet	<b>16.522</b>	<b>0.7186</b>	2.2070	3.2687	15.833	<b>0.6459</b>	1.6647	2.3929



**Fig. 7** SZDNet and well-known methods were compared qualitatively on real-world datasets. **a** Hazy images. **b** DCP [18]. **c** IDE [33]. **d** DDAP [32]. **e** DehazeNet [5]. **f** AOD-Net [6]. **g** EPDN [8]. **h** TCN [34]. **i** RefineDNet [35]. **j** USID-Net [36]. **k** SZDNet. **l** Ground-truth. Red rectangles indicate certain areas that are suggested to be viewed in closer detail for improved visualization and comparison

**Table 3** Qualitative comparison between SZDNet and the SOTA dehazing method on images. The best, second best, and third best performances are marked in red, blue, and green, respectively

Metrics	Images	Prior			Supervised				Unsupervised		
		DCP	IDE	DDAP	DehazeNet	Aod-Net	EPDN	TCN	RefineDNet	USID-Net	SZDNet
PSNR	T1	13.760	15.858	11.890	13.699	12.881	13.363	19.741	15.619	10.896	18.564
	T2	18.212	14.073	13.676	18.765	16.133	21.257	17.065	22.189	22.478	20.723
	T3	16.653	14.814	13.178	17.899	15.314	18.505	17.860	19.721	21.421	21.154
SSIM	T1	0.7059	0.7516	0.5782	0.7190	0.6854	0.7005	0.7731	0.7788	0.6618	0.8303
	T2	0.7222	0.6476	0.6232	0.7419	0.6573	0.8092	0.7591	0.8344	0.7871	0.8104
	T3	0.6367	0.7010	0.5720	0.7015	0.6563	0.7504	0.7451	0.7171	0.7754	0.8201
FADE	T1	2.941	1.719	1.750	2.469	2.211	2.789	2.505	2.356	2.518	2.285
	T2	1.826	1.264	1.260	1.493	1.212	1.514	1.593	2.358	1.840	1.335
	T3	2.230	1.331	1.175	2.014	1.462	1.576	1.475	1.818	1.982	1.592
NIQE	T1	5.462	3.913	3.770	5.050	4.478	4.442	3.937	3.903	3.711	4.105
	T2	2.609	4.085	2.372	2.423	2.377	2.052	2.037	2.123	2.451	2.688
	T3	2.598	3.615	2.874	2.065	2.142	2.105	3.0155	1.516	2.240	2.260



**Fig. 8** SZDNet and well-known methods were compared qualitatively on the real shot images. **a** Hazy images. **b** DCP [18]. **c** IDE [33]. **d** DDAP [32]. **e** DehazeNet [5]. **f** AOD-Net [6]. **g** EPDN [8]. **h** TCN [34]. **i** RefineDNet [35]. **j** USID-Net [36]. **k** SZDNet. Red rectangles indicate certain areas that are suggested to be viewed in closer detail for improved visualization and comparison

**Table 4** Results on the SOTS dataset

Method	PSNR	SSIM	FADE	NIQE
Baseline	23.251	0.8946	<b>1.4904</b>	3.8670
Baseline + A(quad-tree)	23.924	0.8954	1.4914	3.8569
SZDNet	<b>24.502</b>	<b>0.9082</b>	1.5201	<b>3.4759</b>

The bold numbers denote the best value in each category

dehazed image were distorted with DDAP, EPDN and RefinedNet (see image R3). Thus, the dehazed images with SZDNet were the closest to the original image and had better dehazing results even with a self-supervised and zero-shot method.

### 4.6 Ablation study

We performed ablation tests by analyzing various factors, including a multichannel quad-tree algorithm to estimate the atmospheric light and cosine distance ( $L_{\cos(x)}$ ) to show the efficiency of the proposed SZDNet.

We built a basic network as the baseline for reference, where the atmospheric light and transmission maps are estimated with DCP, dehazed images are obtained with U-net, and MSE are chosen as the loss function. We added the following modules to the basic dehazing network: (1) A(quad-tree), where a multichannel quad-tree approach was exploited to solve atmospheric light instead of using the mean of the globally brightest 0.1% pixels. (2) A(quad-tree) +  $L_{\cos(x)}$ , where the cosine distance ( $L_{\cos(x)}$ ) was added to the loss function besides (1), which is the so-called

**Table 5** Quantitative comparison between the proposed SZDNet and the other two methods

Images	Baseline	Baseline + A(quad-tree)	SZDNet	Ground-truth
F1	1.3929	1.3998	1.5498	1.6774
F2	1.4461	1.4415	1.4914	1.6741

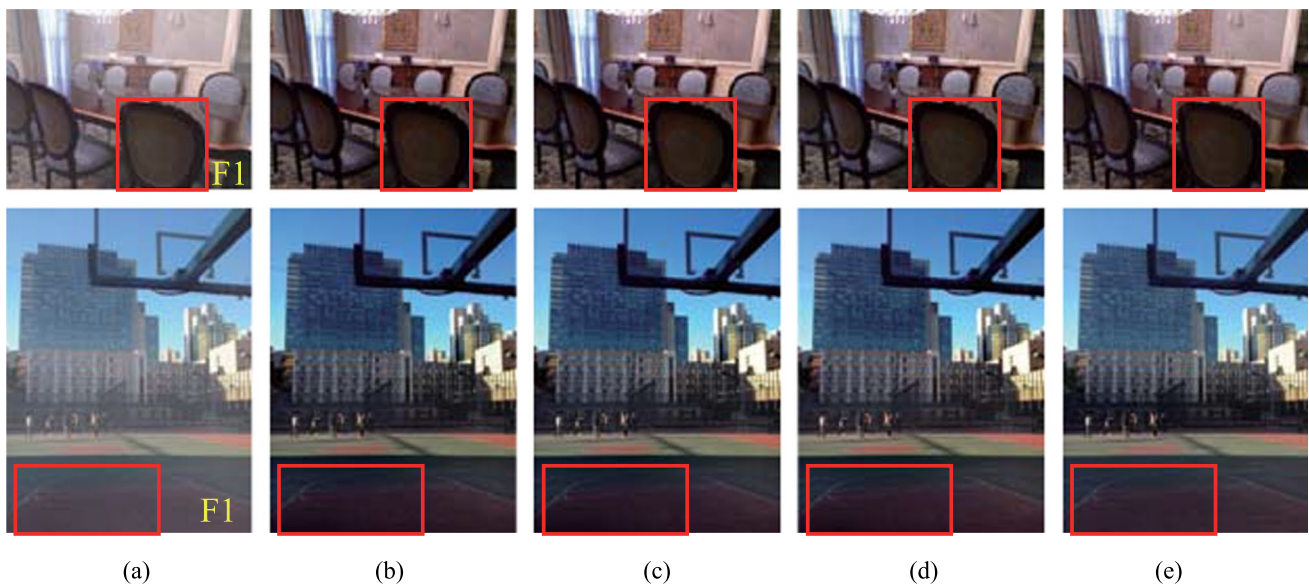
SZDNet proposed in this work. Table 4 shows the specific performances for 100 randomly selected images from SOTS of RESIDE.

#### 4.6.1 Estimating atmospheric light with a multichannel quad-tree algorithm

The atmospheric light value obtained by averaging the brightest 0.1% pixels varied significantly. The dehazing network performed better when the atmospheric light was detected by multichannel quad-tree method because it could accurately pinpoint the brightest sky region. Table 4 shows that SZDNet outperformed the basic dehazing network on PSNR, SSIM, and NIQE. Thus, the multichannel quad-tree algorithm for solving atmospheric light significantly improved dehazing effect.

#### 4.6.2 Cosine distance ( $L_{\cos(x)}$ )

Although MSE is the mostly used loss function in dehazing, it ignores the link between neighboring pixels. The row vectors and column vectors of two images were optimized when the cosine distance was added to the loss function,



**Fig. 9** Qualitative comparison between the proposed SZDNet and the other two methods (baseline and baseline + A(quad-tree)) in the two example images from SOTS Dataset. **a** Input hazy image. **b** Baseline. **c** Baseline + A(quad-tree). **d** SZDNet. **e** Ground-truth images

by examining the relationship between neighboring pixels. Table 4 shows that PSNR, SSIM, and NIQE of SZDNet were better than (2), whereas the FADE was slightly worse. The reason for the latter can be attributed to the fact that FADE might mistake the recovered shadow details for haze residue. We used two images from SOTS for further testing, and the result was given in Fig. 9 and Table 5. Despite the inferior FADE scores of the proposed SZDNet, our results outperformed the other methods and the dehazed images were closer to the ground-truth images. Figure 9d, e show that the shaded regions marked by red rectangles in our results were similar to the ground-truth images. The shaded regions in the other results have low brightness although the FADE scores were better (image F1 and F2 in Fig. 9b, c). This results showed that adding the cosine distance to the loss function improved dehazing effect.

## 5 Conclusions

This study proposed a novel self-supervised zero-shot dehazing network combining the U-Net and ASM. A new multi-channel quad-tree algorithm was developed to estimate the atmospheric light value, which outperformed previous methods in terms of accuracy and robustness in localizing the brightest sky region. The cosine distance was introduced to the MSE as the loss function to consider the entirety of all pixels. SZDNet had the advantage of generating a dehazed image by simply inputting a hazy image without dataset training. Using synthetic and actual datasets, we conducted experiments to demonstrate the dehazing ability of SZDNet. The results showed that the SZDNet outperformed other SOTA methods. The SZDNet might perform better if a more effective technique can be developed to infer the transmission map and the atmospheric light value, which will be addressed in our future research.

**Acknowledgements** This work was supported in part by the National Natural Science Foundation of China (Grant No. 61705127), and Degree Construction Project of Detection Technology and Automation Devices, Shanghai University of Engineering Science (No. 19XXK003).

**Author contributions** XJX proposed idea, wrote the code and the manuscript; YHR and ZWL guided the writing manuscript and provided language polish; NNZ conducted experiments; WNZ provided guidance. All authors read and approved the final manuscript.

**Availability of data and material** The data that support the findings of this study are available from the corresponding author, upon reasonable request.

## Declarations

**Competing interests** The authors declare that they have no competing interests.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Bai, J., Zhu, J., Zhao, R., Gu, F., Wang, J.: Area-based non-maximum suppression algorithm for multi-object fault detection. *Front. Optoelectron* **13**(4), 425–432 (2020)
- Sun, L., Zhao, S., Li, G., Liu, B.: High accuracy object detection via bounding box regression network. *Front. Optoelectron* **12**(3), 324–331 (2019)
- Sakaridis, C., Dai, D., Hecker, S., Van Gool, L.: Model adaptation with synthetic and real data for semantic dense foggy scene understanding. In: *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 687–704 (2018)
- Li, X., Liu, G., Sun, S.: Efficient point cloud segmentation approach using energy optimization with geometric features for 3d scene understanding. *JOSA A* **38**(1), 60–70 (2021)
- Cai, B., Xu, X., Jia, K., Qing, C., Tao, D.: Dehazenet: an end-to-end system for single image haze removal. *IEEE Trans. Image Process* **25**(11), 5187–5198 (2016)
- Li, B., Peng, X., Wang, Z., Xu, J., Feng, D.: Aod-net: all-in-one dehazing network. In: *2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 4770–4778 (2017)
- Chen, D., He, M., Fan, Q., Liao, J., Zhang, L., Hou, D., Yuan, L., Hua, G.: Gated context aggregation network for image dehazing and deraining. In: *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 1375–1383 (2019)
- Qu, Y., Chen, Y., Huang, J., Xie, Y.: Enhanced pix2pix dehazing network. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 8160–8168 (2019)
- Tang, G., Zhao, L., Jiang, R., Zhang, X.: Single image dehazing via lightweight multi-scale networks. In: *IEEE International Conference on Big Data (big Data)*, pp. 5062–5069 (2019)
- Suresh, A., Nisha, J., Gopi, V.P.: Rich feature distillation with feature affinity module for efficient image dehazing. *Optik* **267**, 169656 (2022)
- Xu, L., Wei, Y.: “Pyramid deep dehazing”: an unsupervised single image dehazing method using deep image prior. *Opt. Laser Technol.* **148**, 107788 (2022)
- Hendriksen, A.A., Pelt, D.M., Batenburg, K.J.: Noise2iInverse: self-supervised deep convolutional denoising for tomography. *IEEE Trans. Comput. Imaging* **6**, 1320–1335 (2020)
- Chen, L., Bentley, P., Mori, K., Misawa, K., Fujiwara, M., Rueckert, D.: Self-supervised learning for medical image analysis using image context restoration. *Med. Image Anal.* **58**, 101539 (2019)
- Wang, F., Bian, Y., Wang, H., Lyu, M., Pedrini, G., Osten, W., Barbastathis, G., Situ, G.: Phase imaging with an untrained neural network. *Light Sci. Appl.* **9**(1), 1–7 (2020)
- Li, B., Gou, Y., Liu, J.Z., Zhu, H., Zhou, J.T., Peng, X.: Zero-shot image dehazing. *IEEE Trans. Image Process* **29**, 8457–8466 (2020)

16. McCartney, E.J.: Optics of the atmosphere: scattering by molecules and particles. New York, John Wiley and Sons Inc **1976**, 421 (1976)
17. Narasimhan, S.G., Nayar, S.K.: Contrast restoration of weather degraded images. *IEEE Trans. Pattern Anal. Mach. Intell.* **25**(6), 713–724 (2003)
18. He, K., Sun, J., Tang, X.: Single image haze removal using dark channel prior. *IEEE Trans. Pattern Anal. Mach. Intell.* **33**(12), 2341–2353 (2011)
19. Li, Z., Zheng, J.: Edge-preserving decomposition-based single image haze removal. *IEEE Trans. Image Process* **24**(12), 5432–5441 (2015)
20. Satrasupalli, S., Daniel, E., Guntur, S.R.: Single image haze removal based on transmission map estimation using encoder-decoder based deep learning architecture. *Optik* **248**, 168197 (2021)
21. Tan, R.T.: Visibility in bad weather from a single image. In: 2008 IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–8 (2008). IEEE
22. Berman, D., Treibitz, T., Avidan, S.: Non-local image dehazing. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1674–1682 (2016)
23. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer-assisted Intervention, pp. 234–241 (2015). Springer
24. He, K., Sun, J., Tang, X.: Guided image filtering. *IEEE Trans. Pattern Anal. Mach. Intell.* **35**(6), 1397–1409 (2012)
25. Chen, G., Li, L., Jin, W., Qiu, S., Guo, H.: Weighted sparse representation and gradient domain guided filter pyramid image fusion based on low-light-level dual-channel camera. *IEEE Photonics J.* **11**(5), 1–15 (2019)
26. Wang, W., Yuan, X., Wu, X., Liu, Y.: Fast image dehazing method based on linear transformation. *IEEE Trans. Multimedia* **19**(6), 1142–1155 (2017)
27. Gao, G., Huang, H., Fu, C., Li, Z., He, R.: Information bottleneck disentanglement for identity swapping. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 3404–3413 (2021)
28. Li, B., Ren, W., Fu, D., Tao, D., Feng, D., Zeng, W., Wang, Z.: Benchmarking single-image dehazing and beyond. *IEEE Trans. Image Process* **28**(1), 492–505 (2019)
29. Ancuti, C., Ancuti, C.O., Timofte, R., De Vleeschouwer, C.: I-haze: a dehazing benchmark with real hazy and haze-free indoor images. In: International Conference on Advanced Concepts for Intelligent Vision Systems, pp. 620–631 (2018). Springer
30. Ancuti, C.O., Ancuti, C., Timofte, R., De Vleeschouwer, C.: O-haze: a dehazing benchmark with real hazy and haze-free outdoor images. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 754–762 (2018)
31. Mittal, A., Soundararajan, R., Bovik, A.C.: Making a “completely blind” image quality analyzer. *IEEE Signal Process. Lett.* **20**(3), 209–212 (2012)
32. Li, Z., Shu, H., Zheng, C.: Multi-scale single image dehazing using laplacian and gaussian pyramids. *IEEE Trans. Image Process* **30**, 9270–9279 (2021)
33. Ju, M., Ding, C., Ren, W., Yang, Y., Zhang, D., Guo, Y.J.: Ide: image dehazing and exposure using an enhanced atmospheric scattering model. *IEEE Trans. Image Process* **30**, 2180–2192 (2021)
34. Shin, J., Park, H., Paik, J.: Region-based dehazing via dual-supervised triple-convolutional network. *IEEE Trans. Multimedia* **24**, 245–260 (2021)
35. Zhao, S., Zhang, L., Shen, Y., Zhou, Y.: Refinednet: a weakly supervised refinement framework for single image dehazing. *IEEE Trans. Image Process* **30**, 3391–3404 (2021)
36. Li, J., Li, Y., Zhuo, L., Kuang, L., Yu, T.: Usid-net: Unsupervised single image dehazing network via disentangled representations. *IEEE Trans. Multimedia* (2022)
37. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint [arXiv:1412.6980](https://arxiv.org/abs/1412.6980) (2014)
38. Ulyanov, D., Vedaldi, A., Lempitsky, V.: Deep image prior. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 9446–9454 (2018)



**Xinjie Xiao** is currently pursuing the master's degree with the School of Electronic and Electrical Engineering, Shanghai University of Engineering Science, Shanghai, China. He is a prospective PhD student and currently focusing on deep learning and image processing.



**Yuanhong Ren** received the M.S. degree in Computer Application Technology from Wuhan University of Technology, China in 2007, and received the Ph.D. degree in Control Science and Engineering from Donghua University, China in 2021. She is currently a lecturer at Shanghai Urban Construction Vocational College, China. Her current research interests include the analysis and control of multi-agent systems and complex dynamic networks.



**Zhiwei Li** received the M.S. degree in Electronic Science and Technology from Wuhan University, China in 2006, and the Ph.D. degree from Wuhan National Laboratory for Optoelectronics, Huazhong University of Science and Technology, China in 2016. He is currently an associate Professor at Shanghai University of Engineering Science, China. His research interests include intelligent theory and control, image processing, and optoelectronics.



**Nannan Zhang** is currently pursuing the master's degree with the School of Electronic and Electrical Engineering, Shanghai University of Engineering Science, China. He is currently focusing on deep learning and image processing.



**Wuneng Zhou** received the M.S. degree from Central China Normal University, China in 1982, and the Ph.D. degree in Control Science and Engineering from Zhejiang University, China in 2005. He is currently a Professor with Donghua University, China. His research interests include system theory and control theory, control engineering, and robust control theory and application.