

Research Article

Deep Learning-Enabled Variational Optimization Method for Image Dehazing in Maritime Intelligent Transportation Systems

Xianjun Hu ¹, Jing Wang,² Chunlei Zhang,¹ and Yishuo Tong¹

¹College of Electronic Engineering, Naval University of Engineering, Wuhan 430033, China

²China Aerospace Science and Industry Corporation, Beijing 100048, China

Correspondence should be addressed to Xianjun Hu; xianjunhoo@163.com

Received 10 November 2020; Revised 30 March 2021; Accepted 18 April 2021; Published 3 May 2021

Academic Editor: Wen Liu

Copyright © 2021 Xianjun Hu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Image dehazing has become a fundamental problem of common concern in computer vision-driven maritime intelligent transportation systems (ITS). The purpose of image dehazing is to reconstruct the latent haze-free image from its observed hazy version. It is well known that the accurate estimation of transmission map plays a vital role in image dehazing. In this work, the coarse transmission map is firstly estimated using a robust fusion-based strategy. A unified optimization framework is then proposed to estimate the refined transmission map and latent sharp image simultaneously. The resulting constrained minimization model is solved using a two-step optimization algorithm. To further enhance dehazing performance, the solutions of subproblems obtained in this optimization algorithm are equivalent to deep learning-based image denoising. Due to the powerful representation ability, the proposed method can accurately and robustly estimate the transmission map and latent sharp image. Numerous experiments on both synthetic and realistic datasets have been performed to compare our method with several state-of-the-art dehazing methods. Dehazing results have demonstrated the proposed method's superior imaging performance in terms of both quantitative and qualitative evaluations. The enhanced imaging quality is beneficial for practical applications in maritime ITS, for example, vessel detection, recognition, and tracking.

1. Introduction

Maritime video surveillance system has always been an indispensable part of maritime supervision. Affected by small droplets in the air, the image captured by the maritime surveillance system's imaging equipment always tends to be of low quality. There are phenomena such as poor visibility, edge degradation, color distortion, and texture distortion. This negative impact will directly affect the implementation of advanced vision tasks, such as ship detection and tracking [1, 2]. With vigorous computer vision development, many enhancement methods for maritime images have been proposed [3, 4]. Since haze is easily generated in the maritime environment and seriously affects the visual effect, it is also necessary to research the dehazing of maritime images. Starting from the correlation of haze images, Tang et al. [5] discovered some features, including Local Max Contrast, Local Max Saturation, and Hue Disparity, different from dark channel prior (DCP). In [5], features that significantly

impact the dehazing effect are obtained by comparison and selection. The model is constructed using these combined features and trained using the synthetic dataset to optimize the parameters. However, it easily leads to noise amplification and distortion when dealing with those haze images with relatively high concentration. By proposing a new prior named color attenuation prior, Zhu et al. [6] created a linear model to estimate the scene depth under the new prior. It solved the atmospheric scattering model to obtain potential clear images. The algorithm's defect lies in that the estimation of the scene depth of hazy white images is biased, affecting the dehazing effect. Subsequently, many variations-based image dehazing methods have been proposed [7–10]. Although these methods can perform well in some situations, they cannot robustly process maritime images due to huge texture structure differences.

Based on Tang et al. [5] and Zhu et al. [6], Cai et al. [11] constructed a convolutional neural network (CNN) called DehazeNet for learning the mapping relations between hazy

images and their medium transmission maps. As the first successful usage of deep learning for image dehazing, DehazeNet is trained by building a training dataset with synthesized hazy images. The trained DehazeNet takes a hazy image as input and outputs its medium transmission map to acquire a haze-free result by a simple traditional method. Similar to DehazeNet, Ren et al. [12] developed a multiscale CNN (i.e., MSCNN) to learn the relations between hazy images and their transmission maps and synthesized an indoor dataset with different hazy images based on the NYU Depth dataset [13]. Zhao et al. [14] proposed a deep fully convolutional network for more accurate transmission estimation and developed a new outdoor synthetic training set. Compared with DehazeNet and MSCNN, the dehazing effect in [14] is improved, but it needs broad parameters, and the computational cost that results in the speed of dehaze is slow. Li et al. [15] reformulated the atmospheric scattering model and proposed a light-weight CNN, called AOD-NET. Unlike most previous indirect CNN-based works that first estimate medium transmission maps based on CNN and recover haze-free images by traditional physical methods, AOD-NET can directly dehazed images from their hazy ones. Motivated by image denoising, Du et al. [16] converted a dehazing problem into a denoising one and proposed a deep residue learning network to remove haze from hazy images. Chen et al. [17] proposed a gated context aggregation network for image dehazing and deraining and applied the smoothed dilated convolution to avoid the gridding artifacts. Most CNN-based methods leverage haze-free images to synthesize hazy datasets [17, 18]. However, some researchers thought that it could not represent the data distribution of real hazy images correctly, and some deficiencies existed in those models trained with synthesized datasets.

Recently, generative adversarial net (GAN), proposed by Goodfellow et al. [19], has been proven potent in image-to-image translation. Typically, GAN includes two subnetworks, that is, generator and discriminator, which are adversarially trained at the same time to acquire expected results. Besides, GAN cannot rely on any synthetic hazy image pairs. Yang et al. [20] proposed an end-to-end disentangled dehazing network trained by unpaired hazy images and clean images and generated haze-free images. Zhang et al. [21] proposed a new dehazing architecture, called densely connected pyramid dehazing network (DCPDN), that could jointly learn to estimate transmission map, atmospheric light, and dehazed images. Furthermore, a joint discriminator within DCPDN was designed to optimize dehazed images and transmission maps. Suarez et al. [22] proposed a stacked conditional GAN to dehaze each color channel of RGB image independently and applied multiple loss functions to optimize the network over a conditional probabilistic model. Generally, GAN-based algorithms need extensive data to train, but the training is toughed and requires higher equipment requirements. Although many dehazing methods based on the physical model and deep learning have been proposed, these methods do not consider the characteristics of maritime images. It cannot be well applied to maritime supervision tasks.

Therefore, it is necessary to propose a method for enhancing the hazy maritime image.

In this work, our contribution can be described as follows:

We propose a dehazing method based on the atmospheric scattering model. Specifically, we use a fusion strategy to estimate the transmission. Then, a two-step optimization method based on deep learning is designed to optimize the transmission.

The solution to the subproblem obtained by our proposed two-step optimization algorithm is equivalent to image denoising based on deep learning.

Our method has the best performance in synthetic and real dehazing experiments compared with other methods.

The remainder of this paper is organized into the following several sections. Section 2 briefly introduces the problem formulation related to image dehazing. The optimal estimation of the transmission map is presented in Section 3. Section 4 proposes the CNN-enabled variational optimization method and its numerical optimization algorithm. Numerous experiments on synthetic and realistic datasets are performed in Section 5. Finally, we end this paper by summarizing the main contributions in Section 6.

2. Problem Formulation

2.1. DCP. Based on the statistical analysis of massive images, He et al. [23] discovered the dark channel phenomenon and proposed the DCP. DCP believes that some pixel values in the nonsky local region of any clear image are always low in a specific color channel, even approaching zero. Figure 1 shows various maritime images and corresponding dark channel images. It is evident that the dark channel value in most regions tends to zero. In this work, the mathematical expression of DCP can be written as

$$J^{\text{dark}}(x) = \min_{c \in \{r, g, b\}} \left(\min_{y \in \Omega(x)} (J^c(y)) \right) \rightarrow 0, \quad (1)$$

where J and J^{dark} , respectively, indicate the outdoor haze-free image and the corresponding dark channel image, J^c represents the single-channel image corresponding to J in the color channel $c \in \{r, g, b\}$, and $\Omega(x)$ is the local region centred on the pixel point x .

2.2. Image Dehazing. To achieve image dehazing, we first describe the formation of hazy images. Narasimhan et al. [24] explained the hazy images' imaging process by establishing the mathematical model named atmospheric scattering model. This model can be exploited to describe the hazy image production process; that is,

$$I(x) = J(x)t(x) + A(1 - t(x)), \quad (2)$$

where I denotes the hazy image, t represents the transmission map, and A is the global atmospheric light value. Assuming that A is known, we can take minimization

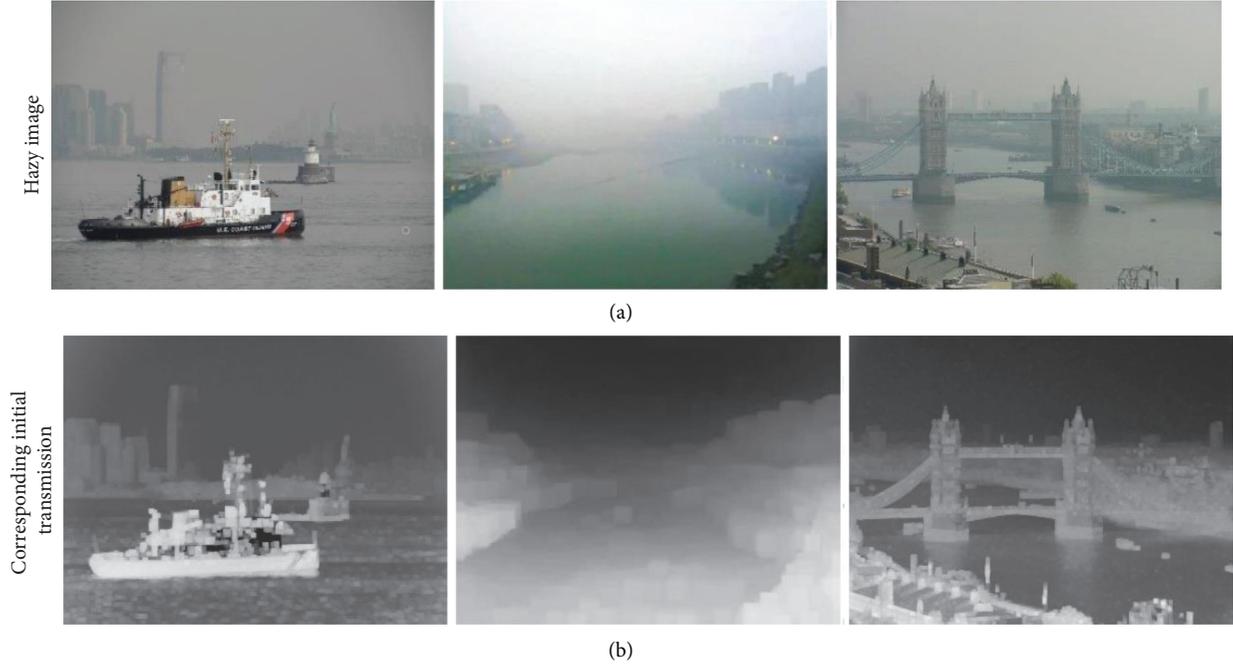


FIGURE 1: The hazy image on the water with sky region (a) and the corresponding initial transmission (b).

operation on both sides of equation (2) to obtain equation (3), which can be given by

$$\min_c \left(\min_{y \in \Omega(x)} \frac{I^c(y)}{A^c} \right) = \min_c \left(\min_{y \in \Omega(x)} \frac{J^c(y)}{A^c} \right) t(x) - (1 - t(x)). \quad (3)$$

According to equation (1), we can approximate $\min_c (\min_{y \in \Omega(x)} J^c(y)/A^c) \approx 0$. Therefore, equation (3) can be rewritten as follows:

$$t(x) = 1 - \omega \min_c \left(\min_{y \in \Omega(x)} \frac{I^c(y)}{A^c} \right), \quad (4)$$

where ω is an adjustment parameter, indicating the degree of dehazing of the image. The parameter's introduction can preserve a certain haze in the sky region to make the dehazed image more natural. The value of ω is determined by the actual situation. In general, a better result can be obtained with $\omega = 0.95$.

In the atmospheric light value estimation, the traditional approach selects the pixel with the highest brightness in the sky region as the value of A . However, the pixels' inability with the highest brightness to be accurately distributed in the sky regions will result in a failed estimation of atmospheric light. Therefore, He et al. [23] selected the pixels with the 0.1% highest brightness in the dark channel image of the hazy image and used the maximum value of these pixels corresponding to the hazy image as the estimated value of atmospheric light.

It can be seen from equation (3) that the transmission map obtained by DCP has the same value in the local regions, and the transmission changes significantly when the brightness changes suddenly. Therefore, the image restored by this transmission map will have a blocking effect. To obtain a more refined transmission map, He et al. adopt the soft matting

algorithm to optimize the initial transmission map. After obtaining the atmospheric light A and transmission map t , the potentially clear image can be restored according to the inverted atmospheric scattering model; that is,

$$J(x) = \frac{I(x) - A}{\max(t(x), t_0)} + A, \quad (5)$$

where t_0 represents the lower boundary of the transmission map. In our experiments, we set $t_0 = 0.1$ generally.

However, the above dehaze method based on transmission map estimation needs to ensure that the DCP theory is valid. When the DCP theory fails, the inaccurate transmission map estimated will lead to the restored image's poor visual effects. Theoretically, in the dark channel image corresponding to the clear maritime image, the sky and water regions' pixel value fails to approximate 0. Therefore, it is unreasonable to directly estimate the transmission map of the hazy maritime image with DCP.

3. Optimal Estimation of Transmission Map

Images captured on the water usually contain large sky regions that generally do not satisfy the DCP. The conventional algorithm introduced in Section 2 is easy to lead to inaccurate transmission estimation. To improve the dehazing effect of the proposed method in haze image on the water with sky region, we also use the soft segmentation method [15] to correct the initial transmission map in this section.

Firstly, we estimate the haze image's initial transmission map on the water based on the DCP. Secondly, we use the soft segmentation method to process the initial transmission map to obtain the transmission weight map. Thirdly, we use the brightness model proposed in [25] to estimate the sky

region's transmission map. Finally, we combine the transmission weight map to merge the transmission map of the nonsky region and the sky region to obtain the corrected initial transmission map. The revised initial transmission map in this section will help to calculate the joint optimization model.

3.1. Weight Function of Transmission Map. As for a hazy image on the water I , we can calculate the initial transmission map based on the DCP mentioned in Section 2. The initial transmission map $\tilde{t}_d(x)$ can be given by

$$\tilde{t}_d(x) = 1 - \omega \min_c \left(\min_{y \in \Omega(x)} \frac{I^c(y)}{A^c} \right), \quad (6)$$

where $\omega = 0.95$ and $\Omega(x)$ represents a 21×21 region centred on x . The atmospheric light A can be estimated by the DCP directly.

As shown in Figure 1, the sky region's values are obviously smaller than the values of other regions in the transmission map estimated by the DCP theory. Therefore, we can roughly distinguish the sky region and other hazy images based on the transmission value in the initial transmission map. The methods proposed in [25] can obtain the possibility that each pixel of a hazy image belongs to the sky region and other regions, that is, transmission weight map, based on the above transmission map. In the transmission weight map, the pixels close to 0 are considered sky region, while pixels close to 1 are considered other regions. To distinguish the sky region and other hazy image regions more accurately, we use a sigmoid function to stretch the transmission. Specifically, the transmission weight function of the initial transmission is obtained as follows:

$$w(x) = \frac{1}{1 + e^{-\theta_1(\tilde{t}_d(x) - \theta_2)}}, \quad (7)$$

where $\tilde{t}_d(x)$ is the initial transmission obtained using the DCP theory, θ_1 is the parameter to adjust the slope of the sigmoid curve, θ_2 is the centre of the horizontal coordinate set according to the $\tilde{t}_d(x)$ range, and θ_1 and θ_2 can be, respectively, given as follows:

$$\theta_1 = \frac{20}{\max(\tilde{t}_d(x)) - \min(\tilde{t}_d(x))}, \quad (8)$$

$$\theta_2 = -10 - \theta_1 \min(\tilde{t}_d(x)). \quad (9)$$

This section uses the sigmoid function to stretch the initial transmission to get the transmission weight map, shown in Figure 2. The soft segmentation method can easily distinguish the sky region and other regions. Because the water surface condition is complicated, the transmission weight map of some hazy image on the water obtained by the soft segmentation method may have deviations, which will result in insufficient transmission after merge. Therefore, the transmission after merge is further optimized to obtain the optimal value.

3.2. Transmission Estimation with Sky Regions. The transmission of the sky region estimated by the DCP theory is not accurate because the white regions (such as the sky) do not

conform to the DCP theory, which will result in the failure of dehazing the hazy image on the water with a large sky region. In order to estimate the transmission of the sky region more accurately, we estimate the transmission of the sky region based on the brightness model.

According to the hazy image degradation model described in Section 2, the relationship between the transmission of the scene and the depth information of the object is as follows:

$$t(x) = e^{-\beta(\lambda)d}, \quad (10)$$

where d represents the distance between the object and the imaging device. β is the dissipation coefficient of the medium.

The above formula shows that the scene transmission could be estimated if the depth information of the image can be obtained. In [25], Zhu et al. found that the brightness distribution in HSL color space in the hazy image is usually related to depth information through the statistics of a large number of hazy images, and the brightness of the sky region is much larger than other regions. Therefore, we can simulate the scene depth according to the brightness of the hazy image and then estimate the scene transmission as follows:

$$\tilde{t}_L(x) = e^{-\beta(\lambda)\hat{L}(x)}, \quad (11)$$

where $\tilde{t}_L(x)$ is the transmission estimated from the brightness model, $\hat{L}(x)$ is the corrected brightness, and β is the dissipation coefficient of the medium. Different wavelengths of light have different dissipation coefficients under the same medium. According to the Mie Scattering Model, the dissipation coefficients of red, green, and blue light are taken as 0.3324, 0.34333, and 0.3502, respectively.

To better simulate the real scene depth, the brightness is stretched as follows:

$$\hat{L}(x) = \frac{\tau}{L^*} L(x), \quad (12)$$

where $L(x)$ is the brightness of the hazy image, L^* takes the value at the 95% quantile of brightness, and τ represents the depth range of the real scene. The value of τ can be selected according to the haze density in the hazy image. The greater the haze, the greater the value of τ , and vice versa. In this paper, $\tau = 3.4$.

By combining equations (11) and (12), we can see that the transmission obtained based on the brightness model is

$$\tilde{t}_L(x) = e^{-\beta(\lambda)\frac{\tau}{L^*}L(x)}. \quad (13)$$

3.3. Combination of Transmission Map. The brightness model can simulate the depth of the sky region well but cannot accurately simulate the depth of other regions, which will result in the inaccurate transmission of other regions estimated. On the contrary, the initial transmission estimated based on DCP theory can better estimate the transmission of other regions. Therefore, in this section, we merge the transmissions that are, respectively, estimated

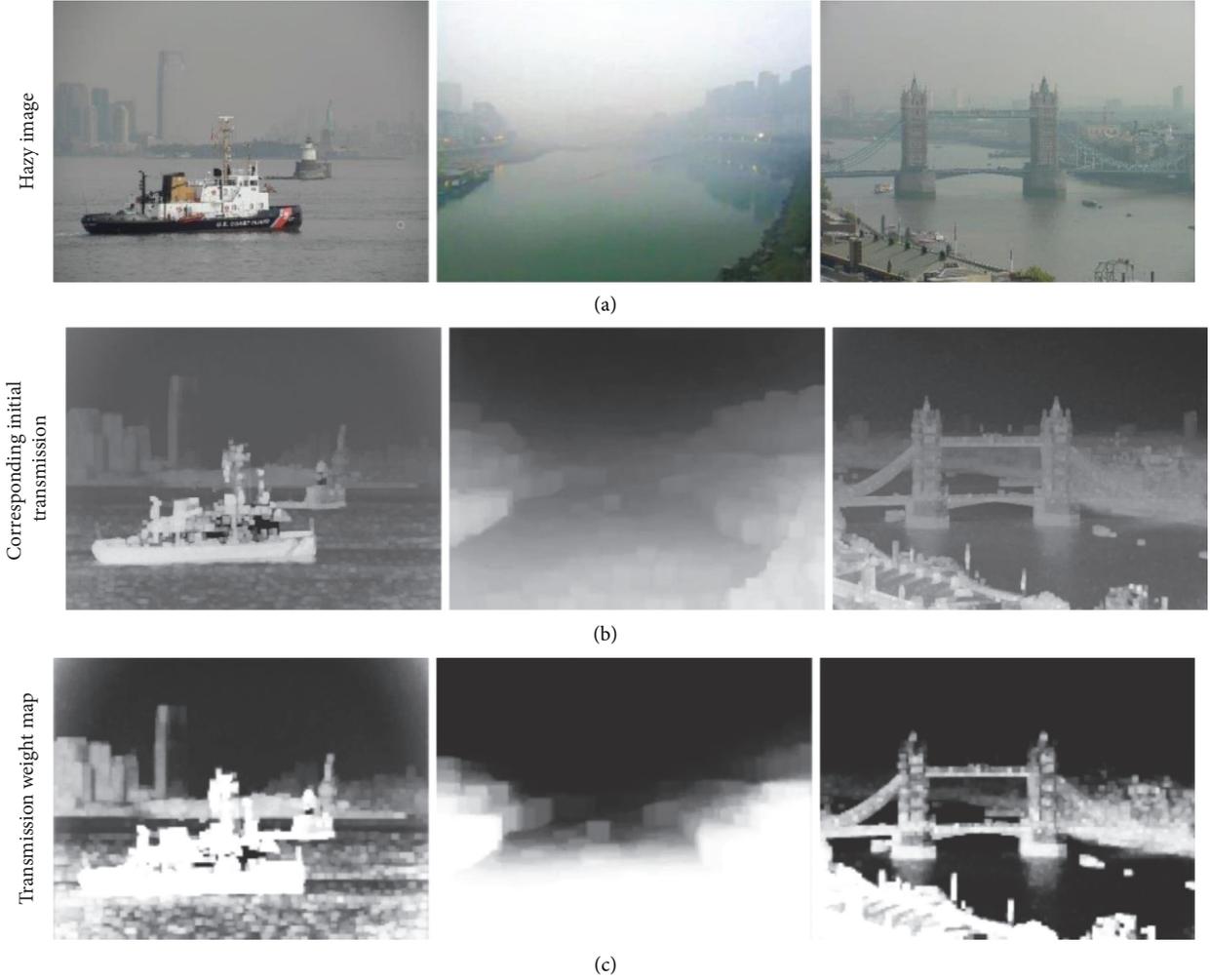


FIGURE 2: The result of the sky region segmentation. (a)-(c) From top to bottom: the hazy image on the water with a large sky region, the corresponding initial transmission, and the transmission weight map.

using the above methods through the transmission weight map to obtain a more accurate result. The corrected initial transmission t_0 is expressed as follows:

$$t_0 = w(x)\tilde{t}_d + (1 - w(x))\tilde{t}_L(x), \quad (14)$$

where \tilde{t}_d represents the initial transmission of other regions, \tilde{t}_L represents the initial transmission of the sky region, $w(x)$ represents the weight function of other regions' transmission, and $1 - w(x)$ represents the weight function of the sky region transmission.

4. CNN-Enabled Variational Optimization Method

4.1. The Unified Transmission Estimation and Image Dehazing Framework. In the previous section, the weight map of transmission is obtained by the soft segmentation and fuses the initial transmission by the weight map. Because the soft segmentation method only uses the sigmoid function to stretch the initial transmission to obtain the weight map, transmission obtained by fusion of this

weight map is still inaccurate. According to the process of image dehazing, accurate estimation of transmission is a crucial step for image dehazing to obtain satisfactory results. To estimate the transmittance as accurately as possible and restore the potential clear image, we proposed a joint optimization model that simultaneously optimizes the transmittance and the potential clear image within a unified framework. The joint optimization model is given by

$$(\hat{\mathbf{J}}, \hat{t}) = \arg \min_{\mathbf{J}, t} \mathbf{I} - \mathbf{J}t - \mathbf{A}(1 - t)_2^2 + \alpha t - t_{02}^2 + \lambda_1 \varphi(\mathbf{J}) + \lambda_2 \psi(t), \quad (15)$$

where \mathbf{J} is the haze-free image to be restored, \mathbf{I} is the captured hazy image, \mathbf{A} is the atmospheric light, t is the transmission of hazy image, and t_0 represents the corrected initial transmission. $\mathbf{I} - \mathbf{J}t - \mathbf{A}(1 - t)_2^2$ and $t - t_{02}^2$ are data fidelity terms to constraint haze-free image and its transmission. $\varphi(\mathbf{J})$ and $\psi(t)$ are the regularization terms representing the prior information of \mathbf{J} and t , respectively. α , λ_1 , and λ_2 are positive parameters.

4.2. *Numerical Optimization.* Since the haze-free image \mathbf{J} and the transmission t are independent, we tend to propose a two-step optimization method to decompose equation (15) into the following two subtasks.

4.2.1. *Estimation of Haze-Free Image \mathbf{J} .* If t^k is fixed, the optimization solution of haze-free images is as follows:

$$\mathbf{J}^{k+1} = \arg \min_{\mathbf{J}} \left(\frac{\mathbf{I} - \mathbf{A}}{t^k} + \mathbf{A} \right) - \mathbf{J}_2^{k2} + \lambda_1 \varphi(\mathbf{J}^k). \quad (16)$$

4.2.2. *Estimation of Transmission t .* In a similar way, if \mathbf{J}^k is fixed, transmission t optimization subtask is solved by

$$\begin{aligned} t^{k+1} &= \arg \min_{t} \mathbf{I} - \mathbf{J}t - \mathbf{A}(1-t)_2^2 + \alpha t - t_{02}^2 + \lambda_1 \varphi(\mathbf{J}) + \lambda_2 \psi(t) \\ &= \arg \min_t \frac{\mathbf{I} - \mathbf{A}}{\mathbf{J}^k - \mathbf{A}} - t_2^2 + \alpha t^k - t_{02}^2 + \lambda_2 \psi(t^k) \\ &= \arg \min_t (1 + \alpha) \frac{\mathbf{I} - \mathbf{A}/\mathbf{J}^k - \mathbf{A} + \alpha t_0}{1 + \alpha} - t_2^{k2} + \lambda_2 \psi(t^k). \end{aligned} \quad (17)$$

$$\begin{aligned} \mathbf{J}^{k+1} &= \arg \min_{\mathbf{J}} \left(\frac{\mathbf{I} - \mathbf{A}}{t^k} + \mathbf{A} \right) - \mathbf{J}_2^{k2} + \lambda_1 \varphi(\mathbf{J}^k) \\ &= \arg \min_{\mathbf{J}} \frac{1}{2(\sqrt{\lambda_1}/2)^2} \left(\frac{\mathbf{I} - \mathbf{A}}{t^k} + \mathbf{A} \right) - \mathbf{J}_2^{k2} + \varphi(\mathbf{J}^k), \end{aligned} \quad (18)$$

$$\begin{aligned} t^{k+1} &= \arg \min_t (1 + \alpha) \frac{\mathbf{I} - \mathbf{A}/\mathbf{J}^k - \mathbf{A} + \alpha t_0}{1 + \alpha} - t_2^{k2} + \lambda_2 \psi(t^k) \\ &= \arg \min_t \frac{1}{2(\sqrt{\lambda_2}/2(1 + \alpha))^2} \frac{\mathbf{I} - \mathbf{A}/\mathbf{J}^k - \mathbf{A} + \alpha t_0}{1 + \alpha} - t_2^{k2} + \psi(t^k). \end{aligned} \quad (19)$$

According to the research in [22], the images $\mathbf{I} - \mathbf{A}/t^k + \mathbf{A}$ in equation (13) and $\mathbf{I} - \mathbf{A}/\mathbf{J}^k - \mathbf{A} + \alpha t_0/1 + \alpha$ in equation (14) can both be considered as the denoising tasks whose noise levels are $\sqrt{\lambda_1}/2$ and $\sqrt{\lambda_2}/2(1 + \alpha)$, respectively. Therefore, any image denoiser can be solved by equations (18) and (19).

$$\mathbf{J}^{k+1} = \text{Denoiser}_1 \left(\left(\frac{\mathbf{I} - \mathbf{A}}{t^k} + \mathbf{A} \right), \sqrt{\frac{\lambda_1}{2}} \right), \quad (20)$$

$$t^{k+1} = \text{Denoiser}_2 \left(\frac{\mathbf{I} - \mathbf{A}/\mathbf{J}^k - \mathbf{A} + \alpha t_0}{1 + \alpha}, \sqrt{\frac{\lambda_2}{2(1 + \alpha)}} \right), \quad (21)$$

where Denoiser_1 and Denoiser_2 are image denoisers.

4.3. *CNN-Based Blind Denoising.* In Section 4.2, a dehazing model is proposed to jointly optimize transmission and

The above two subtasks need to define appropriate regularization terms $\varphi(\mathbf{J})$ and $\psi(t)$ to be solved. In other words, the solution of the above subtasks is limited by the regularization terms $\varphi(\mathbf{J})$ and $\psi(t)$. Considering that deep learning has a strong prior learning ability, the deep learning method is used in this paper to solve these two subtasks. It can be found that the two subtasks have the same form. According to the Bayesian probability, the above two subtasks can be equivalent to Gaussian denoising tasks [26, 27].

Firstly, equations (16) and (17) are transformed into the following form:

restore haze-free image, and the joint optimization model is converted into two subtasks. In this section, we mainly introduce a CNN-based blind denoising model to solve the above two subtasks.

4.3.1. *Image Denoising Model.* The image noise model can be expressed by

$$\mathbf{y} = \mathbf{x} + \mathbf{v}, \quad (22)$$

where \mathbf{y} is observed noisy image, \mathbf{x} is noise-free image to be restored, and \mathbf{v} is white Gaussian noise (AWGN) with standard deviation σ .

Because image denoising is an ill-posed inverse problem, previous works generally adopt the prior knowledge or regularization terms to constrain variables to solve the problem. In the Bayesian framework, equation (22) can be

solved by solving the following maximum posterior problem:

$$\hat{\mathbf{x}} = \mathbf{arg} \max_{\mathbf{x}} \log p(\mathbf{y}|\mathbf{x}) + \log p(\mathbf{x}), \quad (23)$$

where $\log p(\mathbf{y}|\mathbf{x})$ is log-likelihood of observed noise image and $\log p(\mathbf{x})$ is prior information of \mathbf{x} .

According to equation (23), recovering a high-quality noise-free image from a noise image can be regarded as a problem of minimizing the energy function as follows:

$$\hat{\mathbf{x}} = \mathbf{arg} \min_{\mathbf{x}} \frac{1}{2} \mathbf{y} - \mathbf{x}^2 + \lambda \Phi(\mathbf{x}), \quad (24)$$

where $1/2\mathbf{y} - \mathbf{x}^2$ is data fidelity term, $\Phi(\mathbf{x})$ is regularization item related to image prior information, and λ is positive parameter.

4.3.2. CBDNet Structure. After decades of research, many image denoising methods have been proposed. These methods can be divided into two categories: model optimization-based methods and supervised learning-based methods. Model optimization-based methods mainly include total variation, Gaussian mixture model, and BM3D. Most of these algorithms are computationally complex and time-consuming. Supervised learning-based methods mainly contain denoising algorithms based on deep learning such as DnCNN and FFDNet. This denoising method has the advantages of fast speed, excellent performance, and strong robustness [28]. Because deep learning has strong prior learning capabilities and deep learning-based denoising methods have better denoising performance, we adopt the convolutional blind denoising network (CBDNet) proposed by Guo et al. [29] to solve equations (20) and (21). CBDNet benefits from its two subnetworks and provides a blind denoising solution for image denoising. Therefore, we use CBDNet as an optimization algorithm to eliminate haze of different concentrations. As shown in Figure 3, the CBDNet consists of two parts: the noise estimation subnetwork (CNN_E) and the nonblind denoising subnetwork (CNN_D). The noise estimation subnetwork firstly estimates the corresponding noise level image according to the noise image. The nonblind denoising subnetwork obtains the final denoising image according to the noise image and the estimated noise level image. The noise estimation subnetwork includes five fully convolutional layers. Each convolutional layer comprises 32 convolution kernels with size 3×3 , followed by the ReLU nonlinear activation function. The nonblind denoising subnetwork is a 16-layer U-Net structure, and the input layer and output layer of the network are connected by skip connection to obtain final denoised images. Equation (20) is equivalent to denoising color images, while equation (21) is equivalent to grayscale images. Therefore, two CBDNet models denoted as $F_{\Theta_1}(\cdot)$ and $F_{\Theta_2}(\cdot)$ need to be trained with the color dataset and the grayscale dataset, respectively. Θ_1 and Θ_2 are the parameters of the networks, respectively. The two CBDNet models share the

same structure, except that their input and output channel are different.

4.4. Restoration of Latent Haze-Free Images. In summary, the process of the joint optimization dehazing model is shown in Algorithm 1.

Although the haze image can be restored directly according to the proposed model, experiments show that the haze-free image restored with the optimized transmission has better visual effects in structure and texture. In this work, the coarse transmission map is firstly estimated using a robust fusion-based strategy. A unified optimization framework is then proposed to simultaneously estimate the refined transmission map and latent sharp image. The resulting constrained minimization model is solved using a two-step optimization algorithm. To further enhance dehazing performance, the solutions of subproblems obtained in this optimization algorithm are equivalent to deep learning-based image denoising. Therefore, according to the model, we restore the haze-free image after obtaining the optimized transmittance. In summary, the flow of this algorithm is shown in Figure 4.

5. Experimental Results and Analysis

5.1. Experimental Datasets and Settings

5.1.1. Training Data. As mentioned above, we need to train two CBDNet denoising networks (i.e., F_{Θ_1} and F_{Θ_2}) to remove the unwanted noise in the original image and transmission map, respectively.

Because the denoising network F_{Θ_1} mainly focuses on denoising color images, we choose the SeaShips dataset to build a training set for F_{Θ_1} . We randomly select 500 clear images that are high-quality and noise-free from the SeaShips dataset and synthesize noisy images with different noise levels according to equation (17). F_{Θ_2} aims to optimize the transmittance map, so the depth map in the NYU Depth dataset is used to make the corresponding training set. Similarly, we firstly select 500 depth maps from the NYU Depth dataset, use equation (17) to transform them into corresponding transmission maps, and then synthesize transmission maps with different noise levels according to equation (17). Then, these synthesized images and their source images are cropped into many image blocks whose sizes are 128×128 .

5.1.2. Experimental Settings. We use the Adam Optimizer to optimize the network weight parameters in the network training stage. The batch size is set to 64, and the number of iterations (epoch) is set to 40. For the first 20 epochs, we set the learning rate to 10^{-3} , for the last 20 epochs, we set the learning rate to 10^{-4} . The network weight parameters are all initialized with a Gaussian distribution with a mean value of 0 and a variance of 0.01. The loss functions of the two denoising networks are shown in equations (20) and (21). The parameter settings in the

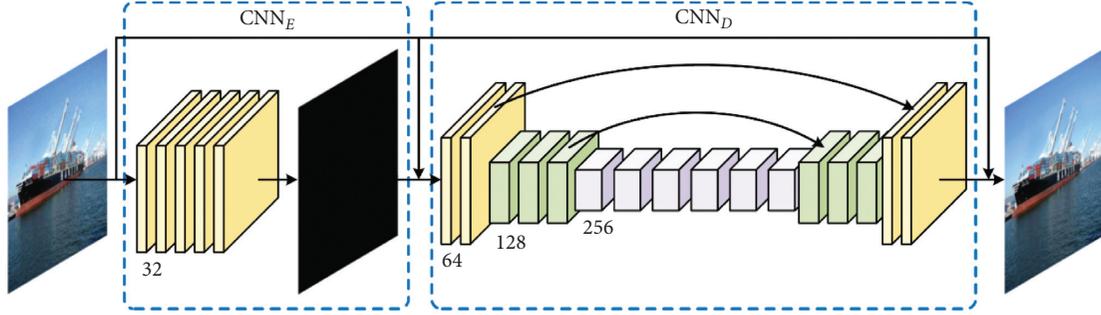


FIGURE 3: CBDNet structure.

- (1) Input: haze image I , atmosphere light A , maximum number of iterations maxiter , adjustment parameter α .
 - (2) Init: $t^0 = t_0$, $J^0 = I - A/t_0 + A$.
 - (3) While $k < \text{maxiter}$:
 - Assuming that t^k is a constant, $J^{k+1} = F_{\Theta_1}(t^k)$,
 - Assuming that J^{k+1} is a constant, $t^{k+1} = F_{\Theta_2}(J^k)$.
- Output: optimized transmission $t^* = t^{k+1}$, haze-free image $J^* = J^{k+1}$

ALGORITHM 1: Deep learning-enabled variational optimization method.

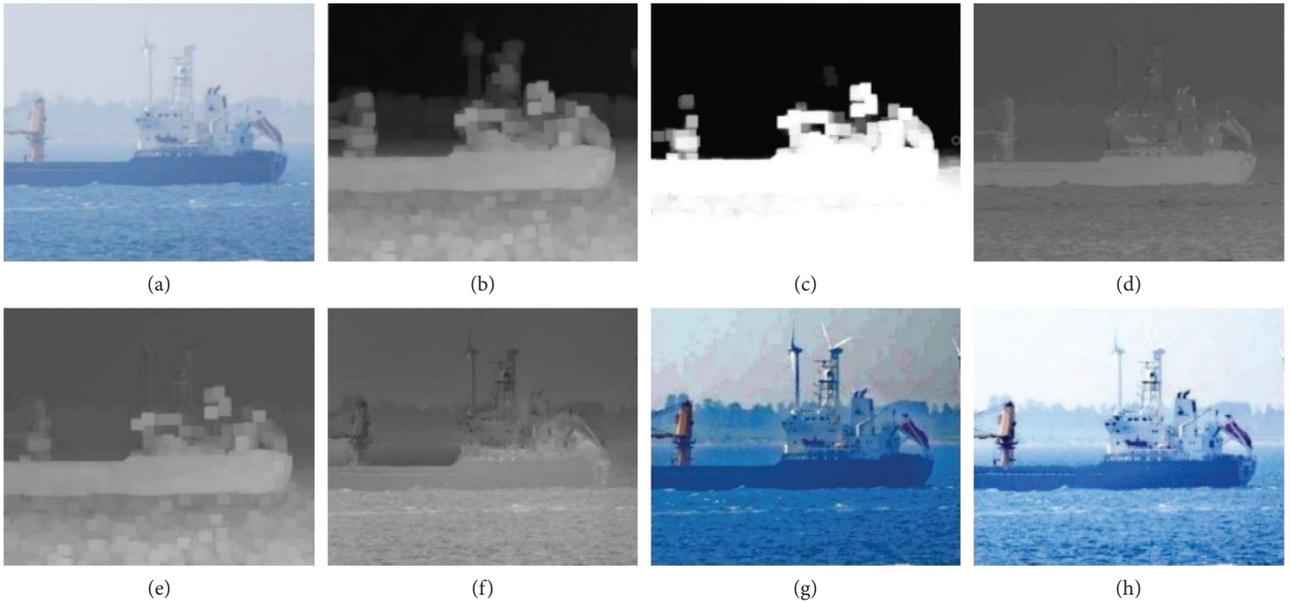


FIGURE 4: Our dehazing framework. From top-left to bottom-right: (a) input hazy image, (b) DCP-based transmission map \tilde{t}_d , (c) weight map of transmission, (d) luminance-based transmission map $\tilde{t}_L(x)$, (e) initial transmission \tilde{t}_0 , (f) optimized transmission t^* , (g) dehazing result with \tilde{t}_d , and (h) dehazing result with t^* .

algorithm are as follows: the maximum number of iterations of the joint optimization model $\text{maxiter} = 6$, and the adjustment parameter $\alpha = 0.5$. The experiment is conducted in Python 3.7 environment with PyTorch package and an Ubuntu 18.04 system, Intel(R) Core(TM) i9-9900X processor, and NVIDIA GeForce RTX 2080Ti GPU. The training of a single model can be done in about one day.

The loss function of the denoising network F_{Θ_1} is as follows:

$$\text{loss} = \frac{1}{q} \sum_{i=1}^q \|F_{\Theta_1}(J_i) - J_i^*\|^2, \quad (25)$$

where $F_{\Theta_1}(J_i)$ is the potentially noise-free clear image, J_i is the input noisy image, J_i^* is the corresponding original noise-free image, and q is the number of training images.



FIGURE 5: Twelve different clear test images.

The loss function of the denoising network F_{Θ_2} is as follows:

$$\text{loss} = \frac{1}{q} \sum_{i=1}^q \|F_{\Theta_2}(t_i) - t_i^*\|^2, \quad (26)$$

where $F_{\Theta_2}(J_i)$ is output transmission map of F_{Θ_2} , t_i is the input transmission map, and t_i^* is the noise-free transmission map of t_i .

5.2. Experiments on Synthetic Datasets. In this section, comparison and analysis are made between the proposed method and three classic image dehazing methods on synthetic images. To verify the effect of the proposed algorithm on removing different haze concentrations, we have selected three transmissions (i.e., $t = 0.1, 0.3, 0.5$), and three atmospheric light values (i.e., $A = 0.7, 0.8, 0.9$). A total of 9 kinds of different degrees of haze are synthesized on twelve test images shown in Figure 5 to validate the performance of the proposed method. As shown in Tables 1 and 2, we have calculated the PSNR and SSIM objective evaluation indicators of our proposed method and the other three methods. It can easily be seen that our method has the best performance in most cases, the hazy concentration is more extensive, and there are more obvious indicator differences.

As shown in Figures 6–8, we have conducted subjective visual analysis experiments. It is obvious that the sky area of the DCP [23] recovery results has different degrees of color distortion and artificial vignetting. The restoration results in RIVD [30] and MSCNN [12] still have a certain degree of haze, and some color distortions appeared in the sky and water. In contrast, the overall effect of the restoration results

of the algorithm in this paper is the best; while maintaining the best visual effect, it virtually eliminates the haze.

5.3. Experiments on Realistic Datasets. In this section, we choose some real hazy images to verify the superiority of our proposed method. Figure 9 shows the dehazing results of different dehazing methods on three maritime video surveillance images. It can be seen from the comparative experimental results in Figure 9 that DCP dehazing algorithm can effectively remove haze. However, at the same time, it also causes artifacts, blocking effects, and color distortion in the restoration results. RIVD and MSCNN dehazing algorithms have individual dehazing capabilities and can avoid artifacts, blocking effects, and so on. However, haze still exists in the restoration results. In contrast, the algorithm in this paper can effectively eliminate the haze, and the restoration results have richer colors, details, and other pieces of information.

The comparison experiment results in other on-water scenes containing hazy images in Figure 10 further prove the effectiveness of the proposed method. It can be seen from the experimental results in Figure 10 that the restoration results of DCP have noticeable artifacts, blocking effects, and color distortion. What is more, the restoration of the sky area is lacking. There is still an apparent haze in RIVD and MSCNN restoration results, resulting in unclear details of some objects. Evidently, the proposed method has an excellent dehazing effect on both sky and nonsky areas. The detailed information of the restoration result is more affluent, and the color is more natural.

It can be seen from the above visual experiments that the proposed method can better recover potentially clear images from hazy images in different water scenes. The restored

TABLE 2: SSIM comparisons (mean \pm std) of various dehazed methods on all test images shown in Figure 5.

PSNR	$A = 0.7$			$A = 0.8$			$A = 0.9$		
	$t = 0.1$	0.3	0.5	0.1	0.3	0.5	0.1	0.3	0.5
Haze	12.22 \pm 2.08	10.59 \pm 2.29	8.77 \pm 2.08	14.40 \pm 2.08	12.78 \pm 2.29	10.96 \pm 2.08	17.30 \pm 2.09	15.72 \pm 2.29	13.87 \pm 2.08
DCP	11.92 \pm 2.25	9.86 \pm 2.30	7.53 \pm 1.83	13.39 \pm 2.42	10.94 \pm 2.11	8.87 \pm 1.79	15.23 \pm 2.11	12.90 \pm 1.80	10.96 \pm 1.59
RIVD	14.82 \pm 1.84	13.32 \pm 2.52	11.09 \pm 2.36	17.44 \pm 2.25	19.92 \pm 2.04	21.94 \pm 3.86	16.08 \pm 4.02	17.64 \pm 3.03	19.10 \pm 2.53
MSCNN	13.91 \pm 1.97	12.26 \pm 2.44	10.17 \pm 2.23	18.70 \pm 1.55	18.08 \pm 2.87	15.68 \pm 2.90	21.28 \pm 1.76	22.70 \pm 2.15	22.20 \pm 3.91
Ours	17.71 \pm 2.02	16.93 \pm 3.33	14.10 \pm 3.20	19.63 \pm 2.13	22.77 \pm 2.79	24.27 \pm 5.16	18.86 \pm 2.91	20.56 \pm 2.30	22.29 \pm 2.07

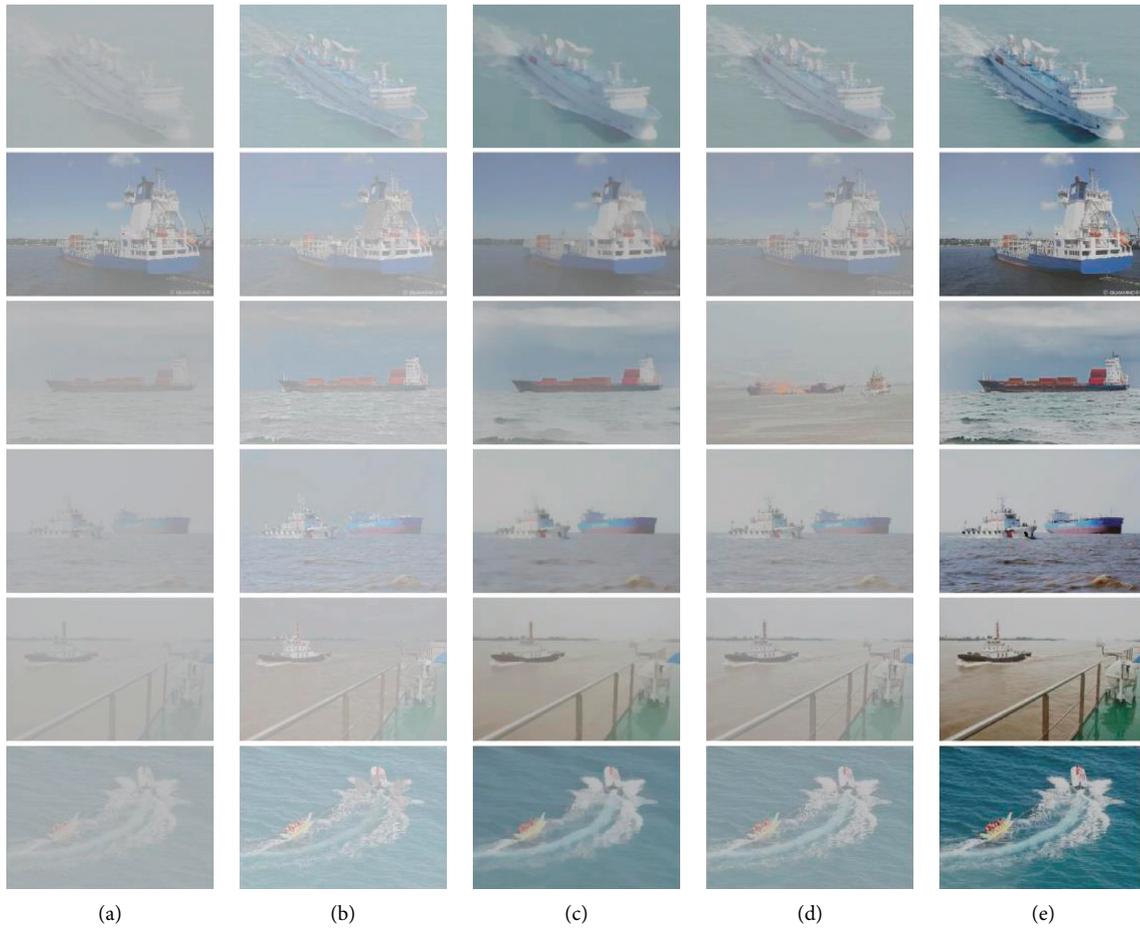


FIGURE 6: Continued.



(f)

FIGURE 6: Visual comparisons of synthetic experiments on six images with transmittance value = 0.1 and atmospheric light value = 0.7. From left to right: (a) synthetic haze image and dehazed images generated by (b) DCP [23], (c) RIVD [30], (d) MSCNN [12], (e) ours, and (f) original clear images, respectively.

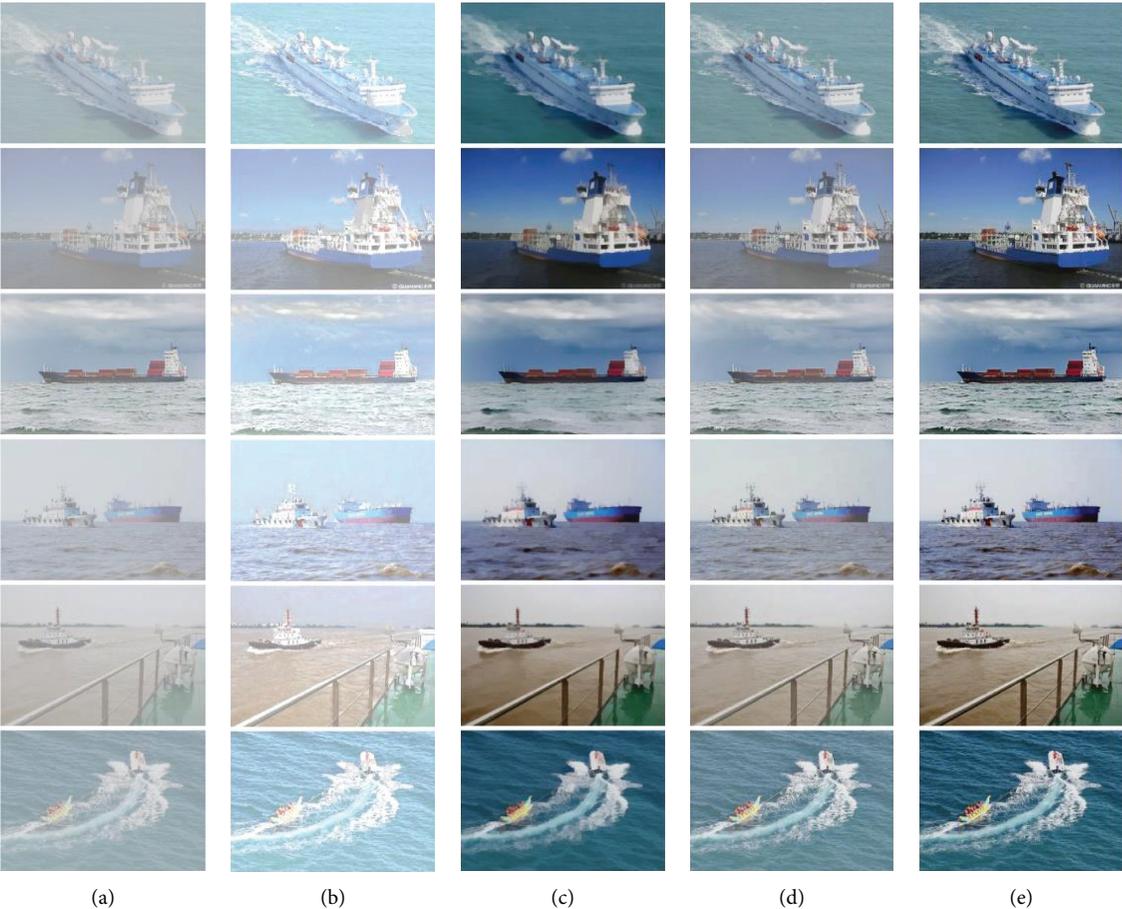


FIGURE 7: Continued.



(f)

FIGURE 7: Visual comparisons of synthetic experiments on six images with transmission value = 0.3 and atmospheric light value = 0.8. From left to right: (a) synthetic haze image and dehazed images generated by (b) DCP [23], (c) RIVD [30], (d) MSCNN [12], (e) ours, and (f) original clear images, respectively.

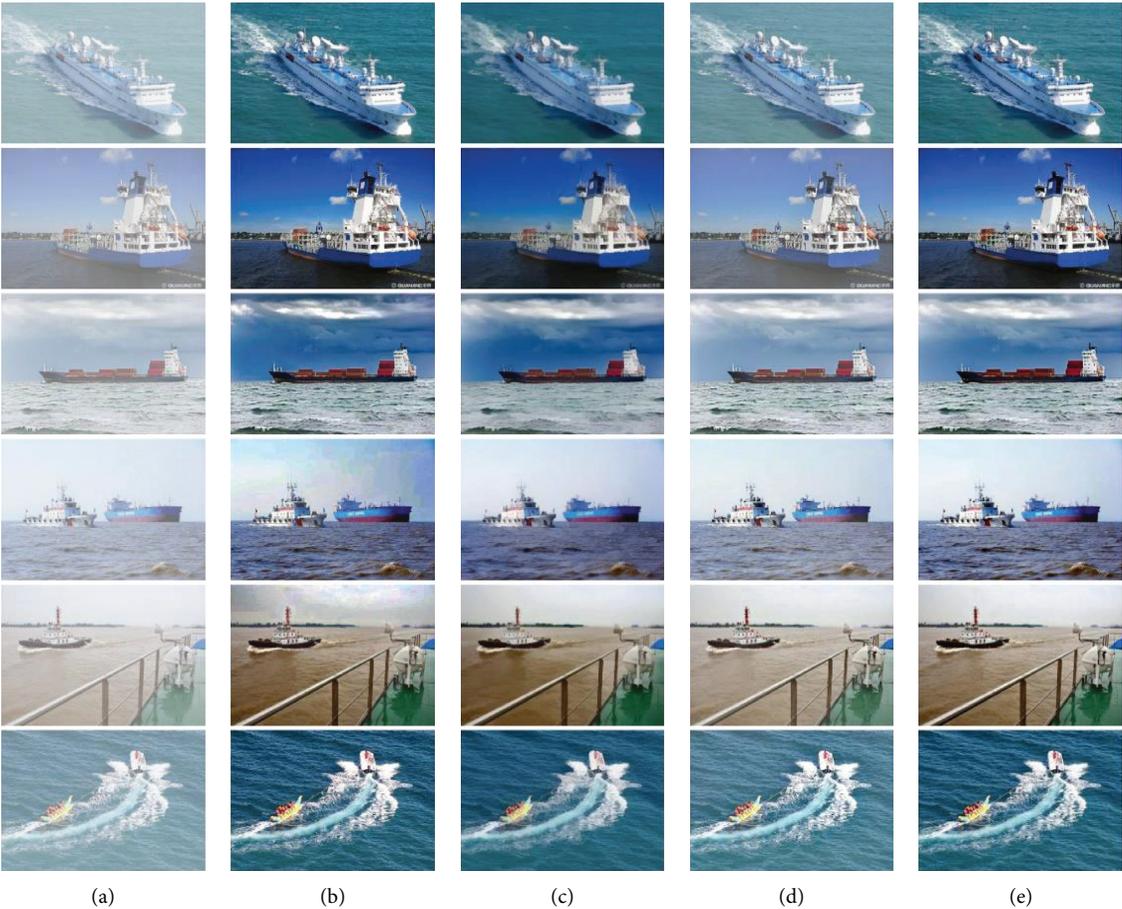
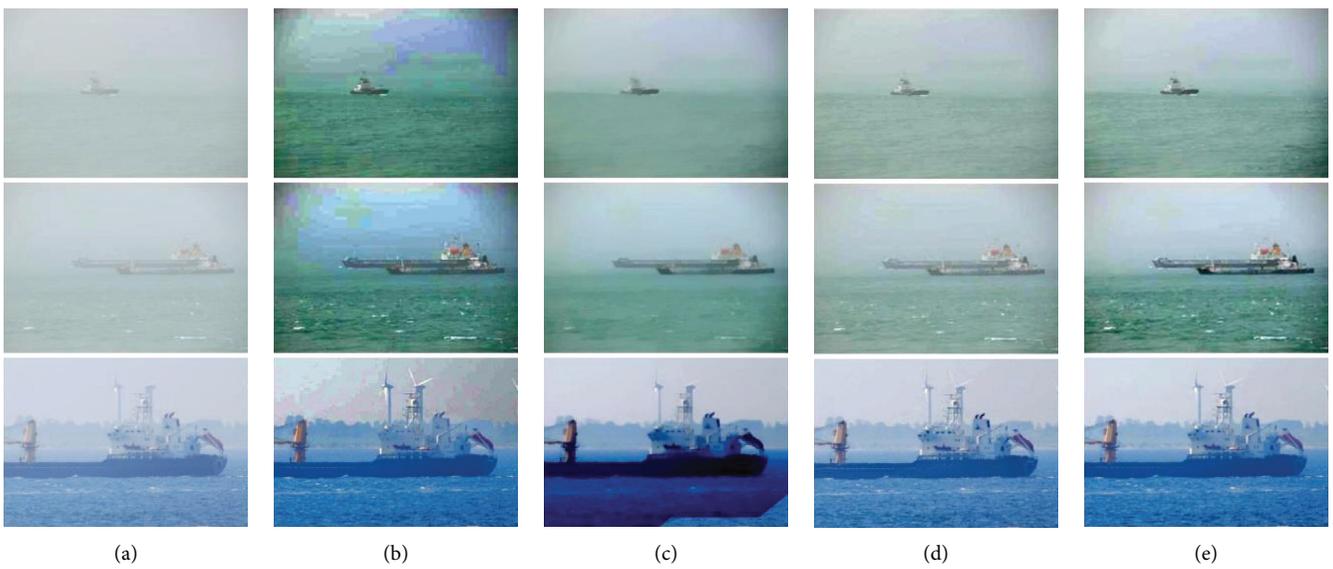


FIGURE 8: Continued.



(f)

FIGURE 8: Visual comparisons of synthetic experiments on six images with transmission value = 0.5 and atmospheric light value = 0.9. From left to right: (a) synthetic haze image and dehazed images generated by (b) DCP [23], (c) RIVD [30], (d) MSCNN [12], (e) ours, and (f) original clear images, respectively.



(a)

(b)

(c)

(d)

(e)

FIGURE 9: Visual comparisons of realistic experiments on three haze images. From left to right: (a) real haze image and dehazed images generated by (b) DCP [23], (c) RIVD [30], (d) MSCNN [12], and (e) ours, respectively.

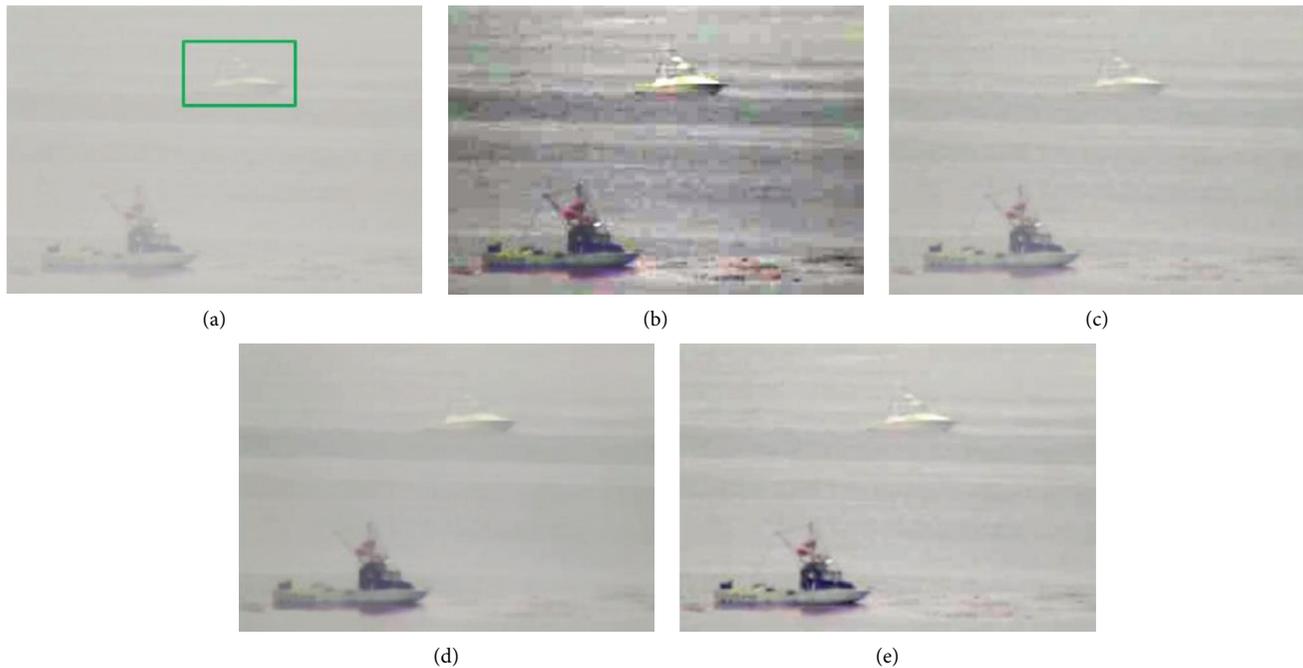


FIGURE 10: Visual comparisons of realistic dehazing experiments. The proposed method is able to significantly improve the image quality and preserve important structures. (a) Haze, (b) DCP, (c) RIVD, (d) MSCNN, (e) Ours.

images have richer detailed information, which shows the effectiveness and stability of the proposed method.

6. Conclusion

Image dehazing is an important preprocessing problem, which has great practical value in various applications in maritime ITS. In this work, a deep learning-enabled variational optimization method is proposed to reconstruct the latent haze-free image from the observed hazy version. Compared to several competing dehazing techniques, the proposed method is capable of generating superior image restoration results in terms of visual image quality and qualitative evaluation. The main benefit of our method is that it takes full advantage of the unified dehazing framework and the strong representation ability of deep learning. In practical applications, the effectiveness and robustness of vessel detection, recognition, and tracking could be significantly enhanced with the enhanced image quality.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This study was supported by the National Natural Science Foundation of China under Grant no. 41774021.

References

- [1] X. Chen, X. Xu, Y. Yang, H. Wu, J. Tang, and J. Zhao, "Augmented ship tracking under occlusion conditions from maritime surveillance videos," *IEEE Access*, vol. 8, pp. 42884–42897, 2020.
- [2] X. Chen, S. Wang, C. Shi, H. Wu, J. Zhao, and J. Fu, "Robust ship tracking via multi-view learning and sparse representation," *Journal of Navigation*, vol. 72, no. 1, pp. 176–192, 2019.
- [3] Y. Guo, Y. Lu, R. W. Liu, M. Yang, and K. T. Chui, "Low-light image enhancement with regularized illumination optimization and deep noise suppression," *IEEE Access*, vol. 8, pp. 145297–145315, 2020.
- [4] M. Yang, X. Nie, and R. W. Liu, "Coarse-to-fine luminance estimation for low-light image enhancement in maritime video surveillance," in *Proceedings of the IEEE International Conference on Intelligent Transportation Systems (ITSC)*, pp. 299–304, Auckland, New Zealand, October 2019.
- [5] K. Tang, J. Yang, and J. Wang, "Investigating haze-relevant features in a learning framework for image dehazing," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2995–3000, Columbus, OH, USA, June 2014.
- [6] Q. Zhu, J. Mai, and L. Shao, "A fast single image haze removal algorithm using color attenuation prior," *IEEE Transactions on Image Processing: A Publication of the IEEE Signal Processing Society*, vol. 24, no. 11, pp. 3522–3533, 2015.
- [7] R. W. Liu, S. Xiong, and H. Wu, "A second-order variational framework for joint depth map estimation and image dehazing," in *Proceedings of IEEE International Conference on*

- Acoustics, Speech and Signal Processing*, pp. 1433–1437, IEEE, Calgary, Alberta, Canada, April 2018.
- [8] Q. Shu, C. Wu, Z. Xiao, and R. W. Liu, “Variational regularized transmission refinement for image dehazing,” in *Proceedings of IEEE International Conference on Image Processing*, pp. 2781–2785, IEEE, Taipei, Taiwan, 2019.
- [9] Q. Shu, C. Wu, Q. Zhong, and R. W. Liu, “Alternating minimization algorithm for hybrid regularized variational image dehazing,” *Optik*, vol. 185, pp. 943–956, 2019.
- [10] Q. Shu, C. Wu, R. W. Liu, K. T. Chui, and S. Xiong, “Two-phase transmission map estimation for robust image dehazing,” in *Proceedings of the International Conference on Neural Information Processing*, pp. 529–541, Springer, Siem Reap, Cambodia, December 2018.
- [11] B. Cai, X. Xu, K. Jia, C. Qing, and D. Tao, “Dehazenet: an end-to-end system for single image haze removal,” *IEEE Transactions on Image Processing*, vol. 25, no. 11, pp. 5187–5198, 2016.
- [12] W. Ren, S. Liu, H. Zhang, J. Pan, X. Cao, and M.-H. Yang, “Single image dehazing via multi-scale convolutional neural networks,” in *Proceedings of the European Conference on Computer Vision*, pp. 154–169, Springer, Berlin, Germany, October 2016.
- [13] N. Silberman, D. Hoiem, P. Kohli, and R. Fergus, “Indoor segmentation and support inference from rgb-d images,” *Computer Vision-ECCV 2012*, Springer, in *Proceedings of the European Conference on Computer Vision*, pp. 746–760, 2012.
- [14] X. Zhao, K. Wang, Y. Li et al., “Deep fully convolutional regression networks for single image haze removal,” in *Proceedings of the IEEE Visual Communications and Image Processing*, pp. 1–4, IEEE, St. Petersburg, FL, USA, December 2017.
- [15] B. Li, X. Peng, Z. Wang et al., “Aod-net: all-in-one dehazing network,” in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 4770–4778, Venice, Italy, October 2017.
- [16] Y. Du and X. Li, “Recursive deep residual learning for single image dehazing,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 730–737, IEEE, Salt Lake City, Utah, USA, June 2018.
- [17] D. Chen, M. He, Q. Fan et al., “Gated context aggregation network for image dehazing and deraining,” in *Proceedings of the IEEE Winter Conference on Applications of Computer Vision*, pp. 1375–1383, IEEE, Waikoloa Village, HI, USA, January 2019.
- [18] S. D. Das and S. Dutta, “Fast deep multi-patch hierarchical network for nonhomogeneous image dehazing,” in *Proceedings of the Conference on Computer Vision and Pattern Recognition Workshops*, pp. 482–483, IEEE, Seattle, WA, USA, June 2020.
- [19] I. Goodfellow, J. Pouget-Abadie, M. Mirza et al., *Generative Adversarial Nets*. *Advances in Neural Information Processing Systems*, pp. 2672–2680, Montréal, Canada, 2014.
- [20] X. Yang, Z. Xu, and J. Luo, “Towards perceptual image dehazing by physics-based disentanglement and adversarial training,” in *Proceedings of the Thirty-second AAAI Conference on Artificial Intelligence*, pp. 7485–7492, New Orleans, Louisiana, USA, February 2018.
- [21] H. Zhang and V. M. Patel, “Densely connected pyramid dehazing network,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3194–3203, IEEE, Salt Lake City, UT, USA, June 2018.
- [22] P. L. Suarez, A. D. Sappa, B. X. Vintimilla et al., “Deep learning based single image dehazing,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 1169–1176, IEEE, Salt Lake City, UT, USA, June 2018.
- [23] K. He, J. Sun, and X. Tang, “Single image haze removal using dark channel prior,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 12, pp. 2341–2353, 2010.
- [24] S. G. Narasimhan and S. K. Nayar, “Chromatic framework for vision in bad weather,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 598–605, IEEE, Hilton Head Island, SC, USA, June 2000.
- [25] Y. Zhu, G. Tang, X. Zhang, J. Jiang, and Q. Tian, “Haze removal method for natural restoration of images with sky,” *Neurocomputing*, vol. 275, pp. 499–510, 2018.
- [26] K. Zhang, W. Zuo, S. Gu et al., “Learning deep CNN denoiser prior for image restoration,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3929–3938, IEEE, Honolulu, HI, USA, July 2017.
- [27] W. Lu, J. Duan, Z. Qiu, Z. Pan, R. W. Liu, and L. Bai, “Implementation of high-order variational models made easy for image processing,” *Mathematical Methods in the Applied Sciences*, vol. 39, no. 14, pp. 4208–4233, 2016.
- [28] B. Goyal, A. Dogra, S. Agrawal, B. S. Sohi, and A. Sharma, “Image denoising review: from classical to state-of-the-art approaches,” *Information Fusion*, vol. 55, pp. 220–244, 2020.
- [29] S. Guo, Z. Yan, K. Zhang et al., “Toward convolutional blind denoising of real photographs,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1712–1722, IEEE, Long Beach, California, USA, June 2019.
- [30] C. Chen, M. N. Do, and J. Wang, “Robust image and video dehazing with visual artifact suppression via gradient residual minimization,” in *Proceedings of the European Conference on Computer Vision*, pp. 576–591, Springer, Amsterdam, The Netherlands, October 2016.