

Research Article

Dual-Channel and Two-Stage Dehazing Network for Promoting Ship Detection in Visual Perception System

Ting Liu  and Baijun Zhou 

Dalian Maritime University, Dalian, China

Correspondence should be addressed to Ting Liu; liuting15348@126.com

Received 29 December 2021; Revised 31 March 2022; Accepted 8 April 2022; Published 6 May 2022

Academic Editor: Bekir Sahin

Copyright © 2022 Ting Liu and Baijun Zhou. 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.

Maritime video surveillance of visual perception system has become an essential method to guarantee unmanned surface vessels (USV) traffic safety and security in maritime applications. However, when visual data are collected in a foggy marine environment, the essential optical information is often hidden in the fog, potentially resulting in decreased accuracy of ship detection. Therefore, a dual-channel and two-stage dehazing network (DTDNet) is proposed to improve the clarity and quality of the image to guarantee reliable ship detection under foggy conditions. Specifically, an upper and lower sampling structure is introduced to expand the original two-stage dehazing network into a two-channel network, to further capture the image features from different scale. Meanwhile, the attention mechanism is combined to provide different weights for different feature maps to maintain more image information. Furthermore, the perceptual function is constructed with the MSE-based loss function, so that it can better reduce the gap between the dehazing image and the unhazy image. Extensive experiments show that DTDNet has a better dehazing performance on both visual effects and quantitative index than other state-of-the-art dehazing networks. Moreover, the dehazing network is combined with the problem of ship detection under a sea-fog environment, and experiment results demonstrate that our network can be effectively applied to improve the visual perception performance of USV.

1. Introduction

The light vision perception system of unmanned surface vessels (USV) can obtain more meaningful visual data, which plays an important role in intelligent navigation. In real-world imaging conditions, the visual image quality is often degraded due to complicated weather conditions, such as haze, rain, and others [1]. In hazy weather, the serious degradation of image quality not only reduces the view value but also affects the effectiveness of visual perception tasks, such as ship detection. Many efforts have been devoted to ship detection under haze condition.

For ship detection, a series of convolution neural networks-based methods have been proposed [2–8]. However, most methods are developed under normal sea conditions, the ship detection accuracy would be reduced under haze condition for the degraded image cannot provide efficient features. A few studies have considered foggy conditions by improving the detection network to learn the characteristics

of ship images under foggy conditions [9]. However, the impact of sea fog for ship detection of USV is still unsolved [10]. In this paper, a single image dehazing problem aiming to recover the clear image from the corrupted input is studied for promoting ship detection; this will be the pre-processing step of the ship detection task.

Single image dehazing is a fundamental low-level vision task that has attracted increasing attention in the computer vision community and artificial intelligence companies over the past few decades [11]. However, how to achieve single image dehazing is still challenging because the problem is seriously ill-posed. The most direct and the simplest method is to adjust image parameters such as contrast, brightness, and gamma correction [12, 13]. However, the image quality can be improved by brightness adjustment, but it does not realize hazy removal. Besides, as a global operation, the unnecessary information will be enhanced with the region of interest by contrast adjustment, making it difficult to distinguish the target from the background [14]. Meanwhile,

only a particular range of intensity levels can be enhanced by Gamma correction. Therefore, it is difficult to determine reasonable parameters to achieve dehazing with no prior information [15].

A series of traditional dehazing algorithms are proposed based on different assumptions and prior knowledge [16], such as histogram equalization, Dark Channel Prior (DCP) [17–20], wavelet transform [21], homomorphic filtering [22], and Retinex [23, 24]. For example, DCP is one of the outstanding dark channels prior-based dehazing methods, which assume that image patches of haze-free images often have low-intensity values in at least one channel. However, the prior-based methods only have good performance in a specific environment for each prior-based method inevitably has some limitations, such that image dehazing by DCP has the problem of fuzziness and residual image in the sky region, and the image processed by wavelet transform has low contrast and distortion [25, 26].

In recent years, the deep convolutional neural network (CNN) has been proven to succeed in image dehazing for its large model capacities and strong feature learning abilities. In 2016, the pioneering work of DehazeNet, which is an end-to-end media transmission fog removal network based on trainable CNN, is proposed by Bolun Cai et al [27]. The relationship between the hazy image blocks and the corresponding media transmission map is studied and estimated. In 2017, another CNN-based dehazing network called AOD-Net is proposed by Boyi Liet et al [28]. Specifically, the hazy-free image is directly generated from the hazy input. This novel end-to-end design makes it easy and fast to embed AOD-Net into other deep models. Up to now, there are still many dehazing networks that are improved based on DehazeNet and AOD-Net [29–31]. Compared to traditional methods, CNN-based methods try to learn the intermediate transmission map or the final hazy-free image without consideration of the image degradation mechanism, which achieves superior performance with big data applied. However, the previous CNN-based dehazing networks treat the features in the feature map equally and lack the ability to pay attention to the task related features, which limited the representation ability of deep learning networks, resulting in poor color recovery after image dehazing and incomplete dehazing problems.

Therefore, the emergence of multiscale CNN (MSCNN) has made it possible to provide more information with multiscale features [32]. Different scale features pay attention to different details of the image, so as to extract more distinctive dehazing image features for clear image reconstruction. Moreover, in order to obtain better multiscale features, a feature extraction module with a certain depth is designed to replace the convolution layer. Besides, an end-to-end Laplacian pyramid defog network (LapDehazeNet) is proposed [33], in which Taylor's infinite approximation theorem is utilized to reconstruct low-frequency and high-frequency information of images by obtaining multiscale features through N branch networks of the pyramid network corresponding to N constraints in Taylor's theorem. The network uses Taylor's infinite approximation theorem to reconstruct the low-frequency information and

high-frequency information of the image from a multiscale perspective by corresponding N constraint terms in Taylor's theorem through the N branch network of the pyramid network.

A coarse to fine two-stage fog removal network (FCTF-NET) [34] is proposed by Yufeng Li and Xiang Chen in 2020 to better restore image details. Specifically, a RDB model is combined to capture multiscale features with fog, especially in the area of nonuniform haze. However, the increase of network depth will lead to the loss of information and the disappearance of gradient, resulting in the color distortion. Therefore, channel expansion is used in this paper to increase the width of the network model to obtain multiscale information.

The increase number of multiscale features can effectively improve the quality of dehazing image reconstruction, but there is redundancy between a large numbers of multiscale features, which greatly reduces the efficiency of the network. Therefore, the fusion of multiscale features remains a problem. When a large amount of input information is input into the network model, selecting the information related to the task for processing can effectively improve the network efficiency on the basis of making full use of multiscale features. A different attention mechanism which utilizes the principle that the human visual system selectively focuses on important areas has been widely used to make the network more sensitive to the effective features [30, 35]. It is combined in our network to describe a priority among channels to create a more powerful representation.

Overall, a dual-channel and two-stage ship dehazing network is proposed in this paper to fully extract and utilize the features of fog image to obtain a dehazing image closer to the human eye. Specifically, a self-built dataset is generated based on the atmospheric scattering model and used to train the network to make up for the lack of large open-source sea image dehazing dataset. Both simulated images and real-world images are used to provide a qualitative and quantitative evaluation comparing with another the-state-of-art dehazing method to evaluate the performance of the image dehazing network. The main contributions of this paper are as follows:

- (1) A dual-channel and two-stage ship image dehazing network is designed to better obtain the multiscale features suitable for sea fog image reconstruction. Multiscale features were extracted from hazy images in different dimensions by combining network channel explanation. Besides, the channel attention mechanism was introduced to fuse the multiscale features by assigning different weight values to further capture image features under different background colors and different haze concentrations and improve the dehazing performance of the network.
- (2) Considering that the current open-source optical ship images are all taken under normal sea conditions and do not meet the requirements of conditions on foggy weather, there is a serious lack of ship data in foggy conditions. In this paper, the atmospheric

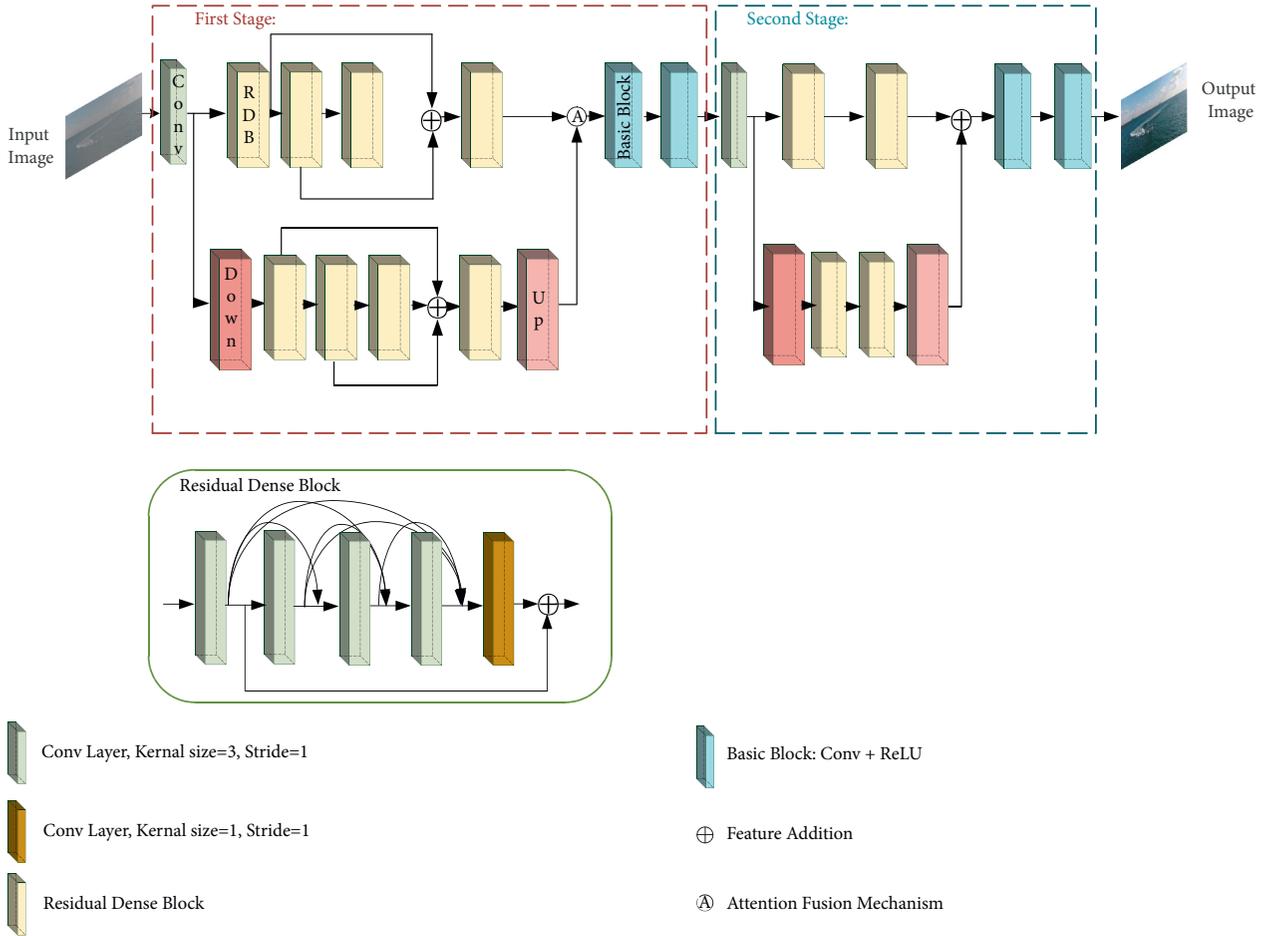


FIGURE 1: The structure of DTDNet.

scattering model is used to synthesize a set of hazy and hazy-free images on the collected images of ships on the sea surface. The model can fully consider the nonuniformity and depth of fog in the image, which is closer to the actual situation.

- (3) A new loss function is conducted by combining the perceptual loss function to the MSE-based loss function. Based on solving the outlier gradient explosion problem, the visual difference between the hazy image and the hazy-free image is quantified through the extracted multiscale features to solve the problem that the difference value of MSE loss function is unstable caused by outliers.

Therefore, the dehazing network is elaborated in detail in the second part of this paper. Moreover, a variety of experiments are conducted in the third part to verify the performance of the proposed method, including a comparison with other state-of-the-art methods. The qualitative and quantitative experimental results show that our model performs outstandingly in ship image dehazing. Besides that, ablation studies were also conducted in the third part to demonstrate the effectiveness of different parts of the newly designed network. Finally, conclusions are drawn in the fourth part.

2. Materials and Methods

In this section, the overall structure and improvement details of the dual-channel and two-stage dehazing network designed in this paper are shown in detail, including the expansion of network channels, attention mechanism, and the design of loss function.

2.1. DTDNet. The DTDNet is proposed on FCTF-Net and the overall structure of DTDNet is shown in Figure 1. Firstly, the same encoder- and decoder-based two-stage coarse-to-fine architecture as FCTF-Net is used to extract multiscale features in DTDNet. Besides, the network channel is extended to capture deeper multiscale image features. Moreover, the attention mechanism was used to generate different weights for the feature extracted by the two channels to fuse the multiscale features. The acquired features were screened to improve the effectiveness of the network. In addition, a new loss function is proposed by combining the MSE loss function with the perceptual loss function to overcome that the difference value of the MSE loss function is too large due to the existence of outliers.

The network is composed of a convolutional layer, RDB module, attention mechanism, and basic block. The RDB

module is used to link all the matching feature maps together and reuse the underlying features at a high level to improve the accuracy of dehazing. Among them, low-level features have high resolution. They contain more location information and detail information, but because of less convolution, low semantic, and more noise, high-level features have strong semantic information but low resolution. Through feature fusion, the features of high and low levels are fused so that the captured features can be complementary and fully utilized. For more details on how the RDB module works, see [20] for more details. Finally, features extracted from the two channels will be transmitted to the basic block after proportional fusion through the attention module; each basic block is composed of a 3×3 convolution layer and ReLU activation function; the structure of basic block is shown in Figure 2.

2.2. Expansion of Network Channel. The dehazing effect of FCTF-Net has been improved with two stages of hazy image processing. However, the hazy image has the problem of asymmetrical feature distribution. Therefore, a fixed number of channels will lead to different depth levels characteristics that cannot be captured. Therefore, the number of channels in the network is expanded and changed into a two-channel network that can extract features of different depths to make the image features more distinctive.

Since the FCTF-Net is divided into two stages, we expand the channel in the two stages respectively. In the first stage, the hazy image was processed through the first convolutional layer; the output of the convolutional layers will be used as the input of the RDB module and the second channel. The image features of the second channel were scaled by introducing the up- and downsampling modules to increase the perception range. Specifically, the size of the feature map is changed to half of the original by convolution operation with ReLU in the downsampling layer to better preserve the edge information. The lower scale feature map is used as the input of RDB module, while the output of RDB module is used as the input of the upsampling layer. The main operations in the upsampling layer include bilinear interpolation and convolution. After the upsampling, the feature map is changed to be the same with the first channel. Bilinear interpolation is used to solve the problem of filling pixels in the upsampling, as shown in Figure 3. The grayscale values of the four known neighboring pixels $Q_{11}, Q_{21}, Q_{12}, Q_{22}$ of the pending pixel P are used to supplement the pixels by linear interpolation in two directions, as shown in the following equations:

$$\begin{aligned} f(R_2) &= \frac{x_2 - x}{x_2 - x_1} f(Q_{12}) + \frac{x - x_1}{x_2 - x_1} f(Q_{22}), \\ f(R_1) &= \frac{x_2 - x}{x_2 - x_1} f(Q_{11}) + \frac{x - x_1}{x_2 - x_1} f(Q_{21}), \\ f(p) &= \frac{y_2 - y}{y_2 - y_1} f(R_1) + \frac{y - y_1}{y_2 - y_1} f(R_2). \end{aligned} \quad (1)$$

Thus, the feature maps of the two channels will be combined with the attention mechanism. Similarly, at the

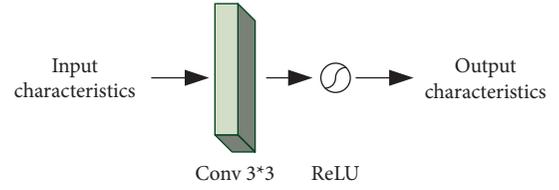


FIGURE 2: The structure of basic block.

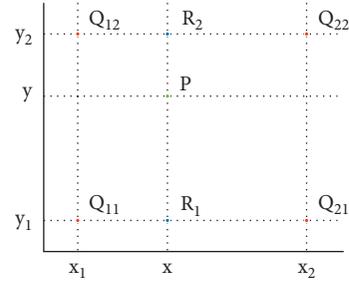


FIGURE 3: The principal of bilinear interpolation.

beginning of the second stage, a two-channel structure was also introduced after the convolution layer. Unlike the channel expansion in the first stage, the attention mechanism is not used to fuse the outputs of the two channels in the second stage. Because the input of the second stage is the image processed after the first stage, the difference of features processed by the attention mechanism is already very sharp. In addition, RDB modules in the second stage expansion channel are reduced, and the module connection mode is changed to step-by-step connection to avoid overfitting. By expansion of network channel, features of different depths can be obtained for learning which has benefit for image dehazing.

2.3. Attention Fusion Mechanism. Deep features are usually too generic to a specific visual task; it is clear that some channel features are more significant than others. When a large amount of input information is processed by the neural network, only some pivotal inputs are selected to improve the efficiency of the network. The attention mechanism takes advantage of the principle that the human visual system selectively pays attention to significant areas. In order to make the network pay more attention to these important channel features, an attention mechanism is introduced in the first stage, with the goal of further screening out the more significant features by explicitly modeling the interdependencies between the channels of its convolutional features. The framework of the attention fusion mechanism is shown in Figure 4.

The attention fusion mechanism allows the model to perform feature recalibration which can learn to fuse the image features captured by the two channels according to the proportion of weight value to further emphasize informative features and suppress useless ones. The main process of attention mechanism is divided into two steps. The first step is used to calculate the attention distribution on all input information, while the other step is used to calculate the

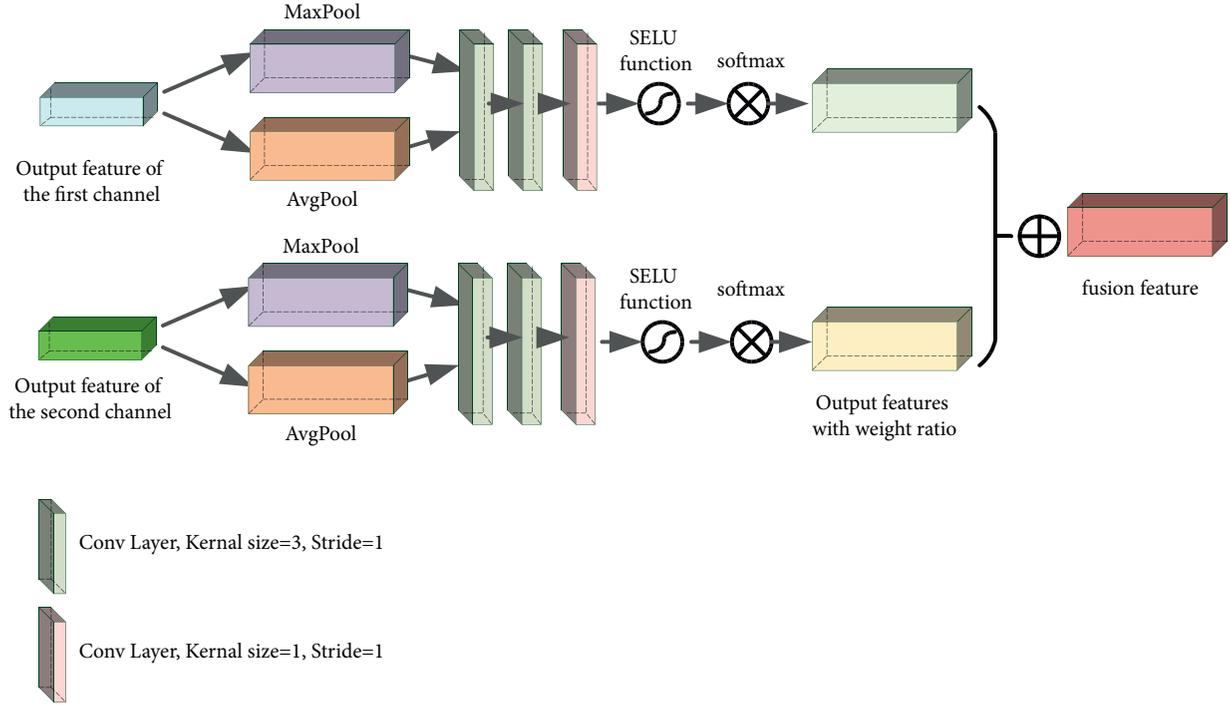


FIGURE 4: The structure of attention fusion mechanism.

weighted average of input information according to the attention distribution [36, 37].

Given the input features $X = [x_1, x_2, \dots, x_C] \in \mathbb{R}^{H \times W \times C}$, the attention module learns the weight W to transform X to $U = [u_1, u_2, \dots, u_C] \in \mathbb{R}^{H \times W \times C}$. The transformation can be written as

$$U = X * W. \quad (2)$$

For DTDNet, different scale features can be obtained through the two channels in the first stage by processing different size feature maps. Thus, it is difficult to distinguish the primary and secondary features. The attention mechanism is utilized to extract relatively important features according to the weight distribution of features as the input of the next stage, so as to facilitate further detailed capture and extraction of features. The final dual-channel fusion feature can be expressed as

$$X_{\text{fusion}} = X_0 * W_1 + X_{up} * W_2, \quad (3)$$

where X_0 and X_{up} are the original channel and extended channel features, while W_1 and W_2 are the weights calculated on the two features, respectively.

Specifically, the attentional mechanism is not used in the second stage to avoid overfitting. Attention module can be used alone or as a component of the neural network to establish an accurate mapping from the low-dimensional solution space to the high-dimensional solution space by calculating the feature weights. By introducing attention mechanism, more information is used to reconstruct the clear image while the processing task of the network can be reduced and more important features are fully utilized.

2.4. Loss Function. MSE function is the most commonly used loss function for dehazing network at present, which overcomes the problem of low stability and unstable solution of the L_1 loss function. The MSE loss function is defined as

$$L_{\text{MSE}} = \frac{1}{N} \sum_{i=1}^N (Y(i)_{\text{true}} - Y(i)_{\text{pred}})^2, \quad (4)$$

where Y_{true} and Y_{pred} represent the true value and predicted value, respectively, and N is the number of values. However, incorrect differences will be produced by the MSE loss function when there are outliers, which is quite different from the real difference. Therefore, the perceptual loss function is introduced in this paper.

The perceptual loss function can quantify and estimate the visual difference between the hazy image and the hazy-free image through the pretrained VGG16, which reduces the possibility of abnormal difference value of the original MSE loss function. The calculation formula of perceptual loss function L_A is defined as follows:

$$L_A = \frac{1}{CWH} \sum_{c=1}^C \sum_{w=1}^W \sum_{h=1}^H \|V_{c,w,h}(Y_{\text{true}}) - V_{c,w,h}(Y_{\text{pred}})\|_2^2, \quad (5)$$

where C , W , and H are the number of output channels, width, and height, respectively. Besides, V indicates the pretrained VGG16 model used in this paper, while $V(\cdot)$ is the output of VGG16 model. Finally, the total loss function applied in this paper is defined as

$$L_D = L_{\text{MSE}} + \alpha L_A, \quad (6)$$

where α is the weight value to adjust the MSE and the perceptual loss function. In our experiments, α is set to 0.06.

Through the new designation of the loss function, the problem of large loss function values due to outliers can be solved and the stability of the network can be improved.

3. Experiment and Results

In this section, extensive experiments are conducted to evaluate the dehazing performance of the DTDNet against several state-of-the-art dehazing networks with both quantitative results and qualitative visual effects. To objectively evaluate the dehazing effect of DTDNet, the peak signal-to-noise ratio (PSNR) and structural similarity index measure (SSIM) are selected for quantitative evaluation of synthetic hazy images. Furthermore, for the quantitative analysis of the actual hazy image dehazing effect, image information entropy and image contrast are selected. The information entropy represents the overall characteristics of the information source in an average sense and the unit is bit/pixel. There is only one entropy of information for a particular source. Thus, image information entropy is used to measure the average amount of information carried by each image [38].

All the experiments are carried out in the workstation configured by the laboratory. The operating system is Windows 10 Professional Edition, while the processor is Intel(R) Xeon(R) Silver 4210R, and the graphics card is NVIDIA GeForce RTX 2080 Ti. The software used is the version of PyCharm 2020.1, in which the internal environment is PyTorch with Python 3.7.

3.1. Dataset Generation. Reasonable dataset preparation is the key factor for deep learning network-based applications. Currently, the publicly available datasets for ship detection mainly include Marvel [39], VAIS [40], and FleetMon datasets [41]. These datasets are all taken under normal sea conditions, while part of the images lacks a sea surface environment due to the large size of the ship target. However, the problem focused in this paper is the ship detection performance with sea fog.

Besides, the role of the network is to remove sea fog which is the preprocessing of ship detection. The end-to-end reconstruction from hazy image to hazy-free image is realized by the proposed dehazing network. However, there is no public open-source dataset for dehazing network training and testing, especially for sea ship detection applications. In addition, it is unrealistic to obtain a large number of real datasets. Therefore, the synthetic dataset is prepared to evaluate the dehazing performance and the ship detection effect under foggy conditions.

In the dataset creation stage, clear ship images and their corresponding ship images in foggy days after being simulated are needed in this paper, and drones are used (see as Figure 5(a)) to take pictures of the navigational condition of “Lan Xin” uncrewed surface vessel (see as Figure 5(b)) which is self-built by our laboratory to get hazy-free images.

The atmospheric scattering model [42], which provides a simple approximation of the haze effect, is utilized to

generate hazy sea ship images. The mathematical model is defined as

$$I(x) = J(x)t(x) + A(1 - t(x)), \quad (7)$$

where $I(x)$ and $J(x)$ are the observed hazy and corresponding hazy-free image of the pixel x , respectively. A is the global atmospheric light, and $t(x) \in [0, 1]$ is the medium transmission which represents the percentage of the scene radiance reaching the camera. Thus, $t(x)$ can be further expressed in an exponential decay term as

$$t(x) = e^{\beta d(x)}, \quad (8)$$

where β and $d(x)$ are atmospheric scattering parameters and scene depth. The purpose of single image dehazing is to restore $J(x)$, A , and $t(x)$ from $I(x)$.

According to the parameter settings in [42], hazy images for the collected sea surface-navigation images of “Lan Xin” are generated. In total, 1000 pairs of training images and 200 pairs of testing images are prepared as our dataset. Specifically, we assume that (I) the random global atmospheric light $A \in [0.7, 1]$; (II) the atmospheric attenuation coefficient β ranges from 0.6 to 2.8 (including haze thickness from light to heavy); (III) the RGB channels of a hazy image have the same medium transmission and global atmospheric light values; and (IV) the scene depth $d(x)$ is estimated by the two-dimensional projection position of the image, which is a simple way to estimate the image depth map. Examples of the hazy-free images and corresponding hazy images are shown in Figure 6. All images collected are $566 * 217$ in size.

However, the synthesized hazy image is not real enough due to the lack of real depth map information in the above dataset synthesis process based on atmospheric scattering model. Therefore, RESIDE dataset [43] in which the hazy image is generated by clean image with its corresponding depth map on atmospheric scattering model is used as the second dataset for further verification. The image depth is combined in RESIDE dataset to generate the fog distribution based on the atmospheric scattering model, which is more consistent with the actual situation and more authentic. An example of the dataset is shown in Figure 7. The size of all images in the RESIDE dataset is $541 * 407$.

3.2. Comparative Experiment with Other State-of-The-Art Methods. The proposed dehazing networks are trained on our self-built dataset and RESIDE dataset, respectively. A batch size of 8 and 100 epochs are used to train our DTDNet on both datasets. The learning rate was set as 0.001. Adam optimizer is utilized to accelerate the training: Adam is generally considered quite robust to one’s choice of hyper parameter values, so we kept many of the balanced default values provided by Tensor flow, and the default values of β_1 and β_2 are 0.9 and 0.999, respectively. The dehazing effectiveness and applicability are evaluated on testing images by being further compared with histogram equalization, Retinex, DCP, DehazeNet [27], AOD-NET [28], MSCNN [32], LapDehazeNet [33], and FCTF-Net [34]. The codes for all the above networks are derived from the download link provided in the corresponding papers or downloaded from



FIGURE 5: Devices for collecting datasets in the prior period. (a) Drones and (b) “Lan Xin” unmanned surface vessel.

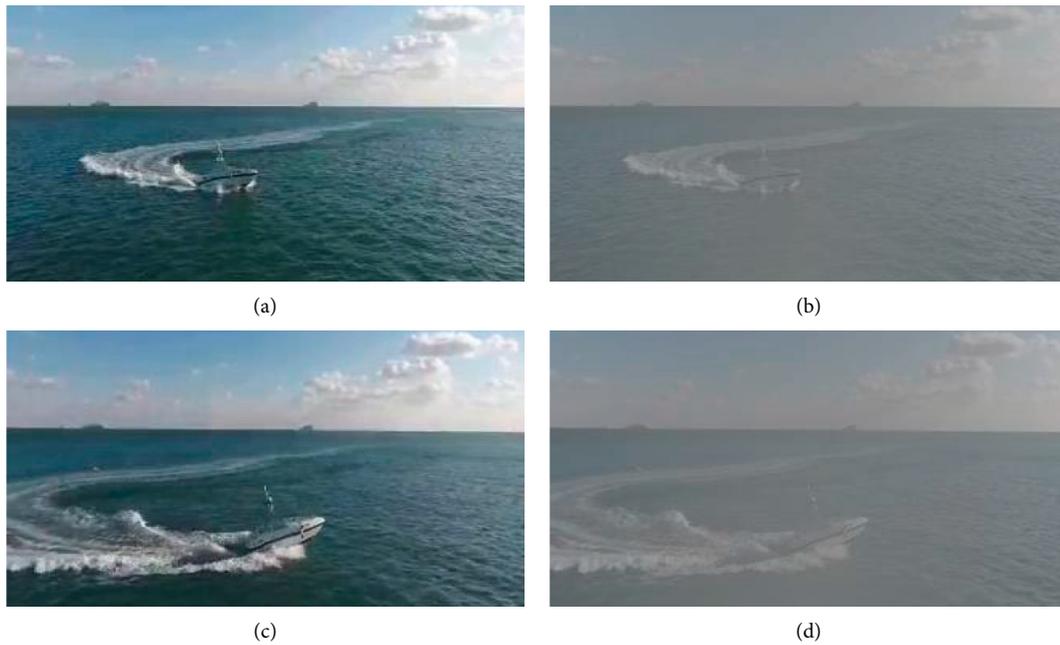


FIGURE 6: The examples of our dataset. (a) An original clear image. (b) The corresponding hazy image of (a). (c) Another original image. (d) The corresponding hazy image of (c).

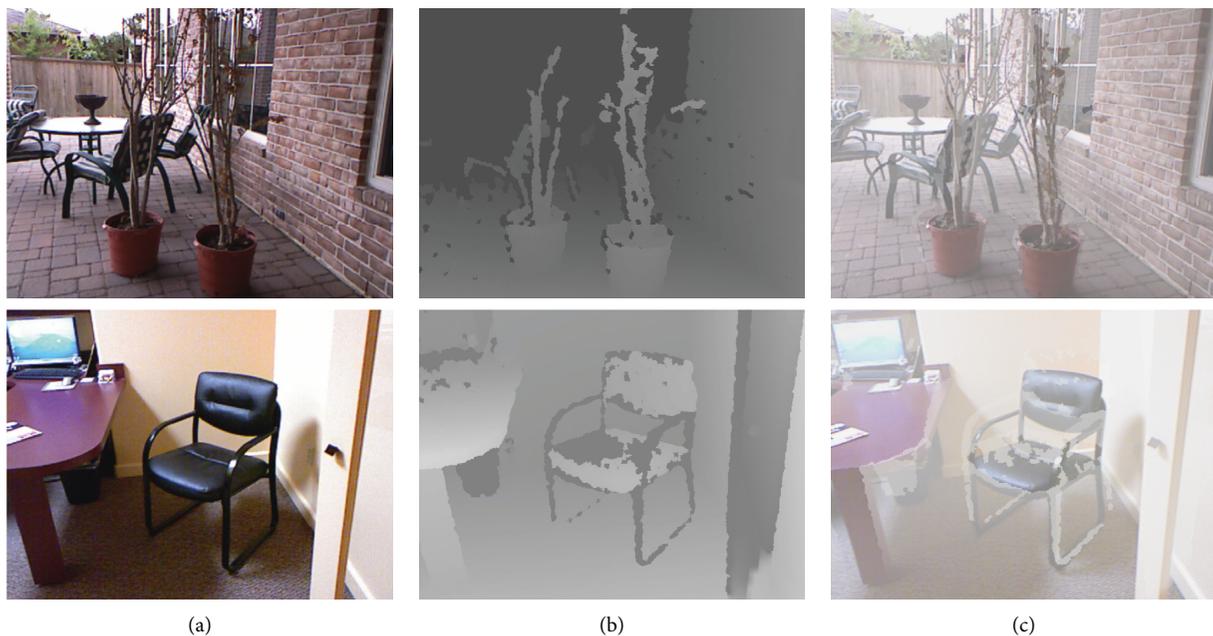


FIGURE 7: The examples of the REDIDE dataset. (a) Original image. (b) The depth-maps of (a). (c) Corresponding hazy images.

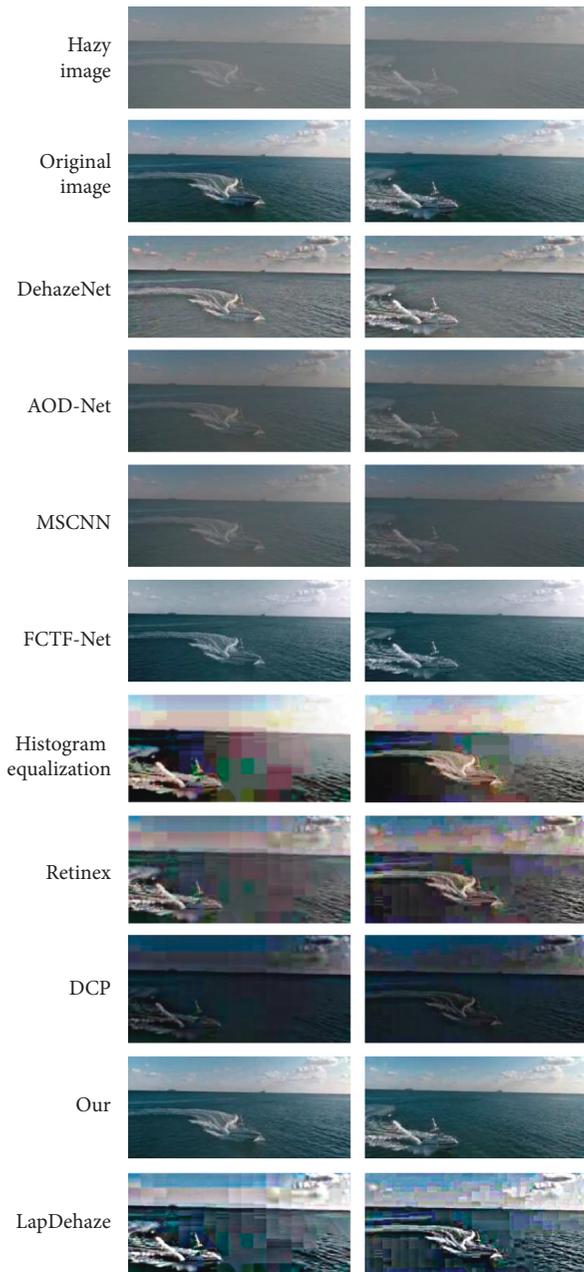


FIGURE 8: Qualitative comparison on synthetic dataset.

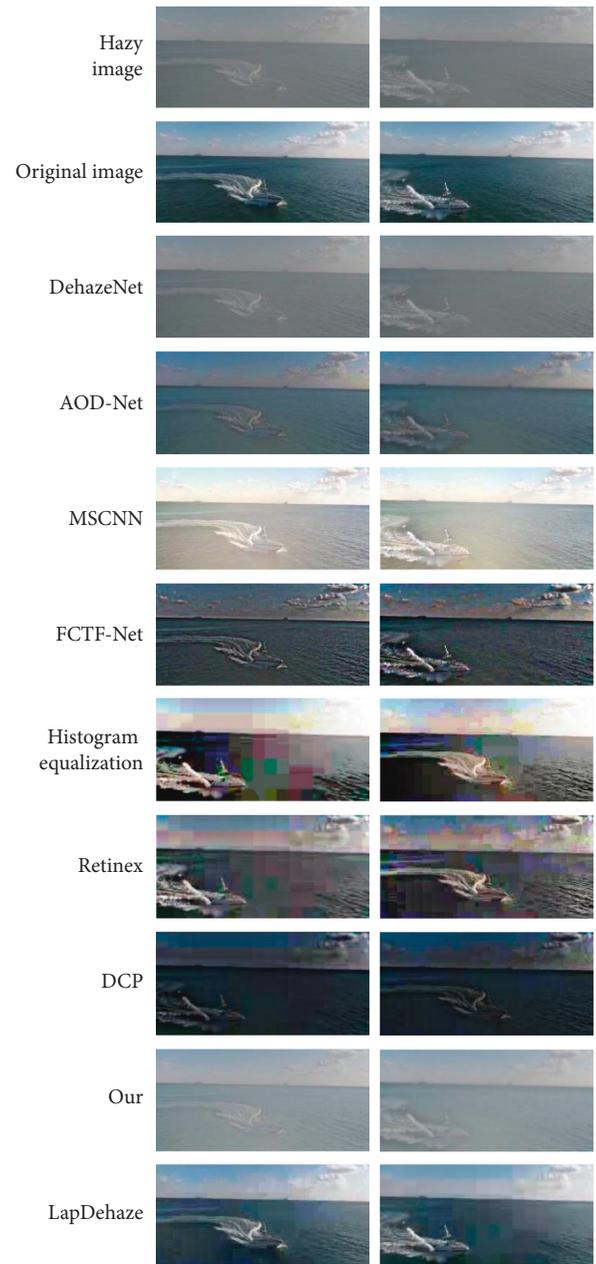


FIGURE 9: Qualitative comparison on the RESIDE dataset.

the author's personal GitHub account. All the networks including the proposed network have been trained with the same training dataset and the same training parameters are adopted for training.

The visual dehazing results are shown in Figures 8 and 9 while the average PSNR and SSIM results between each pair of hazy-free images and dehazing results on the synthetic dataset are shown in Tables 1 and 2. Besides, the processing time, which is the average time taken by each method to process a single image, is calculated and shown in Tables 1 and 2. Among them, the pixel size of each image when calculating the average processing time is $566 * 217$.

From the results, it can be seen that the DTDNet has a better dehazing effect than other networks for both visual

effect and evaluation criteria. Besides, the performance of learning-based method is better than the traditional image processing method overall. Specifically, the dehazing image of DTDNet is the closest to the real hazy-free image. DTDNet and FCTF-Net can achieve a more effective dehazing effect than AOD-Net, DehazeNet, and LapDehazeNet. In particular, there are many fuzzy regions in the result of LapDehazeNet. Compared with FCTF-Net, the method in this paper is more effective for the restoration of sky color. However, the processing time of DTDNet is a little longer than other networks, which is caused by the number of modules and channels in this network. In the future, lightweight processing will be studied to improve the processing speed to better adapt to real-time ship detection.

TABLE 1: Quantitative comparison of different methods on our synthetic dataset.

	DehazeNet	AOD-net	MSCNN	FCTF-net	Histogram equalization	Retinex	DCP	LapDehaze	Our
PSNR	24.89	25.35	23.47	27.31	23.17	21.35	24.39	28.22	30.28
SSIM	0.912	0.921	0.904	0.939	0.820	0.792	0.901	0.927	0.964
Processing speed (s)	2.72	2.62	2.79	3.28	0.85	6.24	1.02	2.15	4.11

TABLE 2: Quantitative comparison of different methods on RESIDE dataset.

	DehazeNet	AOD-net	MSCNN	FCTF-net	Histogram equalization	Retinex	DCP	LapDehaze	Our
PSNR	26.12	27.36	25.61	26.91	23.17	21.35	24.39	28.63	29.78
SSIM	0.92	0.934	0.91	0.902	0.82	0.792	0.901	0.931	0.958
Processing speed (s)	2.33	2.47	2.25	3.34	0.8449	6.2379	1.0211	3.19	3.98

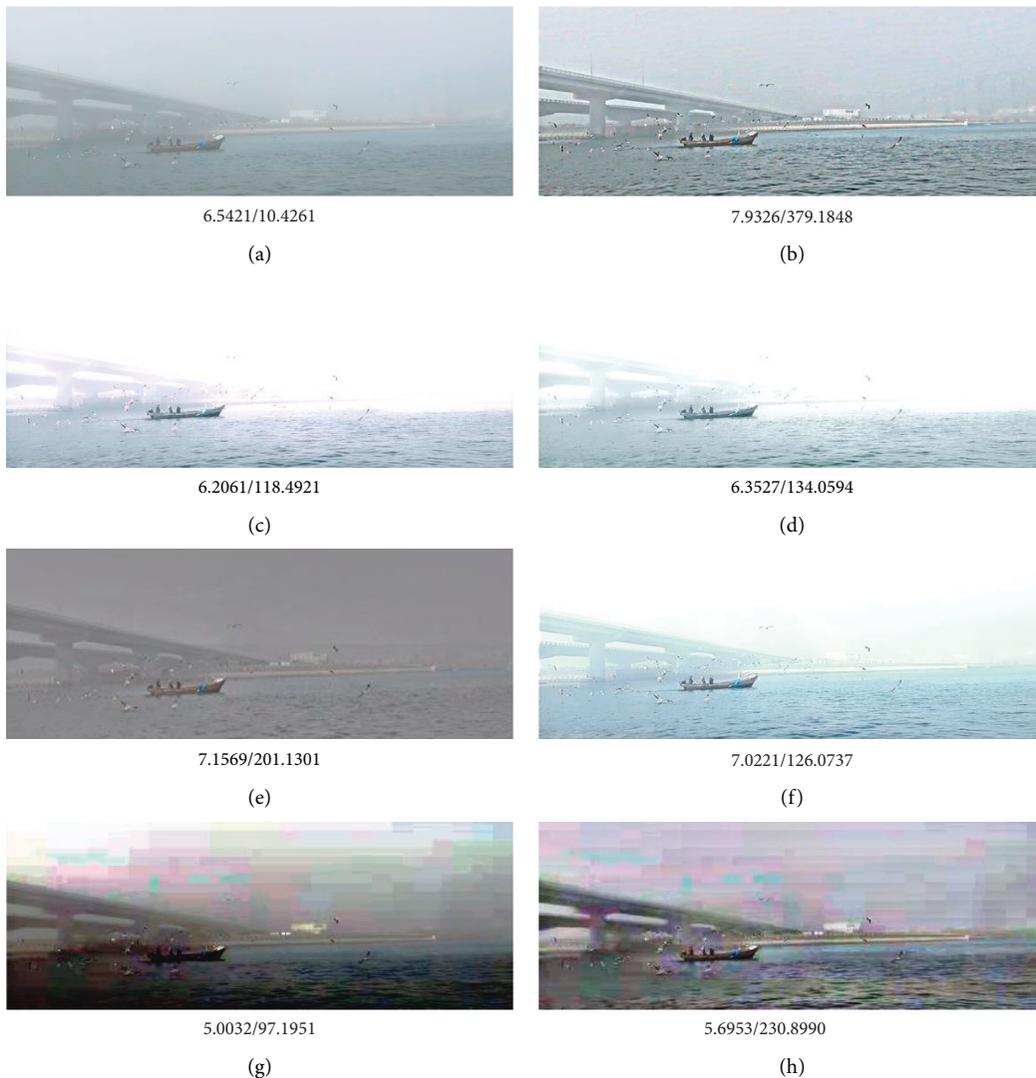


FIGURE 10: Continued.



FIGURE 10: Qualitative comparison on real hazy ship data we collected. (a) The real hazy image taken by camera. (b) The image processed by the designed network. (c) The image processed by DehazeNet. (d) The image processed by AOD-Net. (e) The image processed by MSCNN. (f) The image processed by FCTF-Net. (g) The image processed by histogram equalization. (h) The image processed by Retinex. (i) The image processed by DCP. (j) The image processed by LapDehazeNet.



FIGURE 11: Continued.

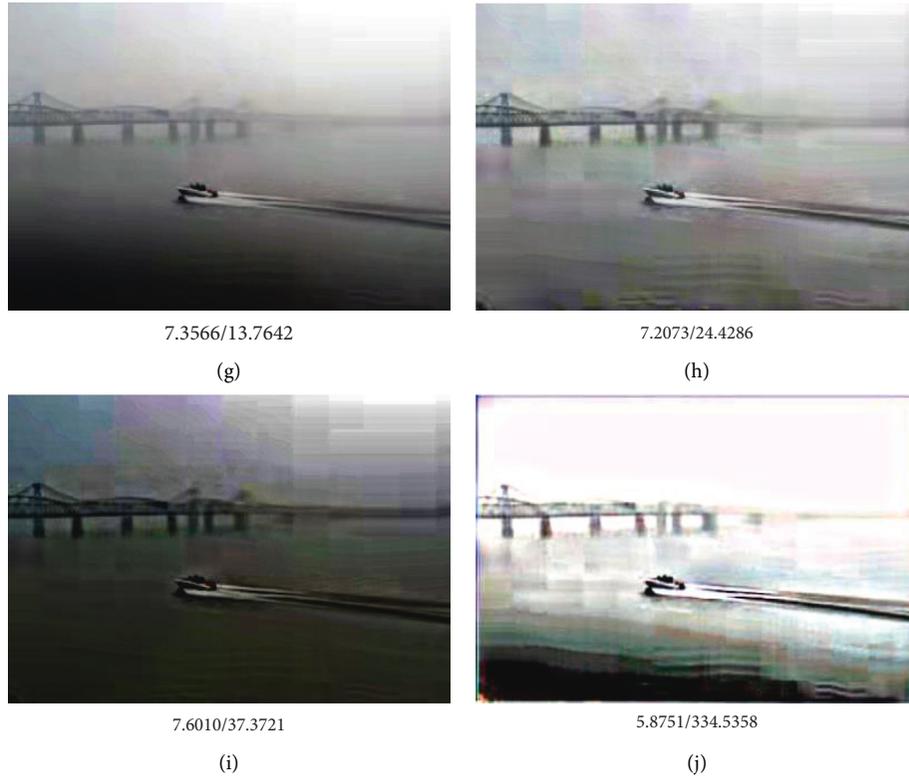


FIGURE 11: Qualitative comparison on real hazy ship data from Internet. (a) The hazy image. (b) Dehazy image of DTDNet. (c) Dehazy image of DehazeNet. (d) Dehazy image of AOD-Net. (e) Dehazy image of MSCNN. (f) Dehazy image of FCTF-Net. (g) Dehazy image of histogram equalization. (h) Dehazy image of Retinex. (i) Dehazy image of DCP. (j) Dehazy image of LapDehazeNet.

Moreover, the dehazing performance of DTDNet is further verified on the real-world hazy ship image and the visual results are shown in Figures 10 and 11. It can be seen that the dehazing image obtained by our method effectively avoids the loss of details, and its visual effect is much closer to the real-world image. Besides, the values of image information entropy and image contrast of each real image are explained in the image annotation. In addition to the contrast in visual effects, through the comparison of the two numerical values of each dehazing image, it can be seen that the image information entropy and image contrast calculated by the dehazing network designed in this paper are higher than those of other networks, which proves once again that the network designed in this paper has a good dehazing effect.

To further verify the application occasions of the network designed in this paper, a variety of real hazy images of nonsea ships are selected for comparative analysis. The comparison results including the values of the image information entropy and image contrast of each image are listed as shown in Figure 12. It can be seen from the figure that the proposed dehazing network is suitable not only for ship images but also for the preprocessing of other real hazy images. The original color and clarity of the image are better maintained, while the image information entropy processed by our designed network is higher than that processed by other methods and the contrast decreases slightly but the difference is small. Furthermore, the comparison of the

dehazing effect of each network trained by the RESIDE dataset is shown in Figure 13. It can be seen from the comparison results that the contrast value of FCTF net is the highest, followed by the network designed in this paper. However, from the visual effect, the output image of FCTF network has excessive distortion. Therefore, our network shows relatively better performance for visual and quantitative indicators.

3.3. Ablation Studies. In this part, the key components including attention mechanism, the second layer network of the first stage and the second stage, and the loss function of the proposed DTDNet are investigated and comparative results are shown in Table 3. From the results, it can be seen that all the key components have a positive effect on the final dehazing performance.

In order to further verify the proposed dual-channel and two-stage dehazing network, two different structural changes are conducted by changing the position of the key components. Firstly, the two-layer network in the second stage with the attention mechanism is used in Network A, while two stages are transformed into a two-layer network and the attention mechanism is only applied to the second stage in Network B. The structure diagrams of the two networks are shown in Figure 14, and the quantitative comparison results with the network designed in this paper are shown in Table 4.



FIGURE 12: Qualitative comparison on real hazy images under synthetic dataset.

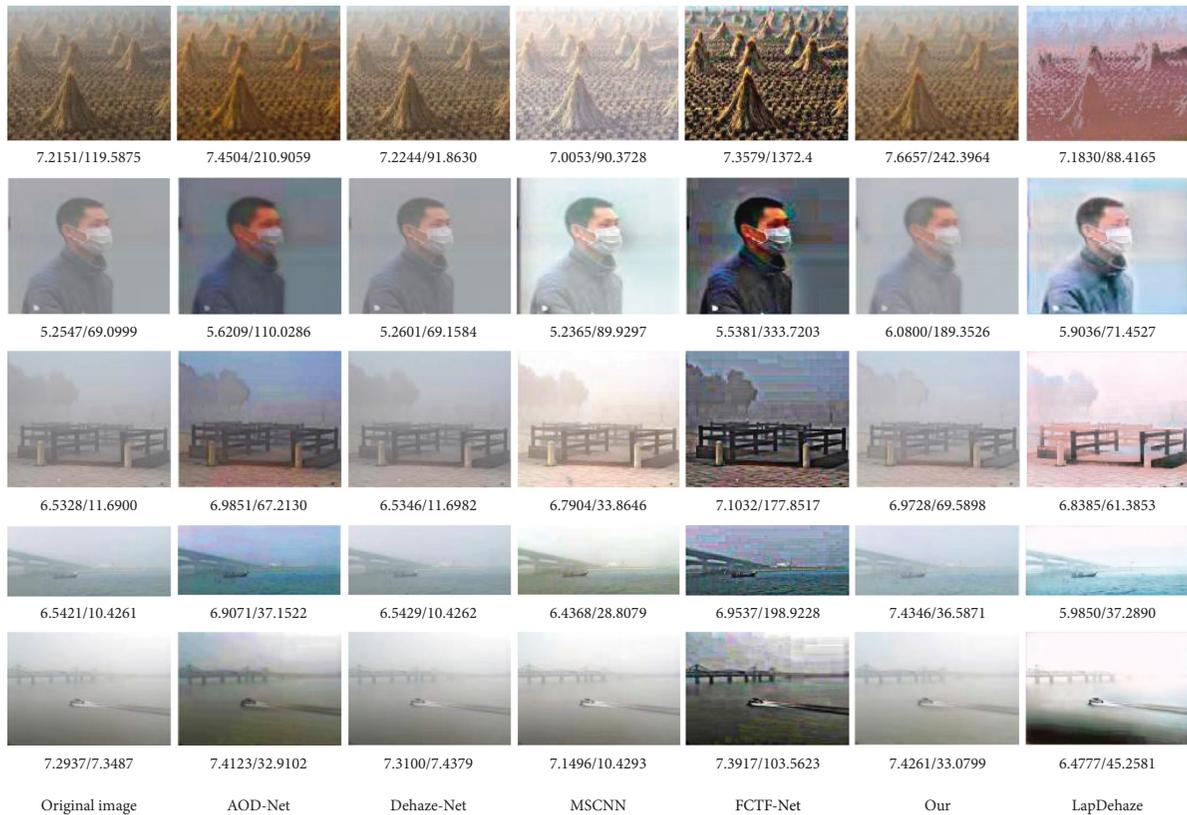


FIGURE 13: Qualitative comparison on real hazy images under the RESIDE dataset.

It can be seen from the above experimental results that the dehazing effect for Network A in which both phases are changed into a two-layer structure with an attention mechanism is similar to that of the original network. It is indicated that the introduction of the attention mechanism in both stages will result in a too small difference in the weight value of the feature, which will not fully exert the best effect of the attention

mechanism. Instead, the dehazing results for Network B in which the attention mechanism is only used in the second stage is much better. Both PSNR and SSIM are higher than FCTF-Net but lower than the dual-channel and two-stage dehazing network proposed in this paper. In summary, the best dehazing effect is achieved by the proposed DTDNet among all related state-of-the-art methods.

TABLE 3: Results of ablation studies based on synthetic datasets.

Attention mechanism	Double network of the first-phase	Double network of the second-phase	Improved loss function	PSNR/SSIM
×	√	√	√	27.413/0.941
×	×	√	√	26.823/ 0.925
√	√	×	√	28.702/ 0.948
√	√	√	×	29.087/ 0.952

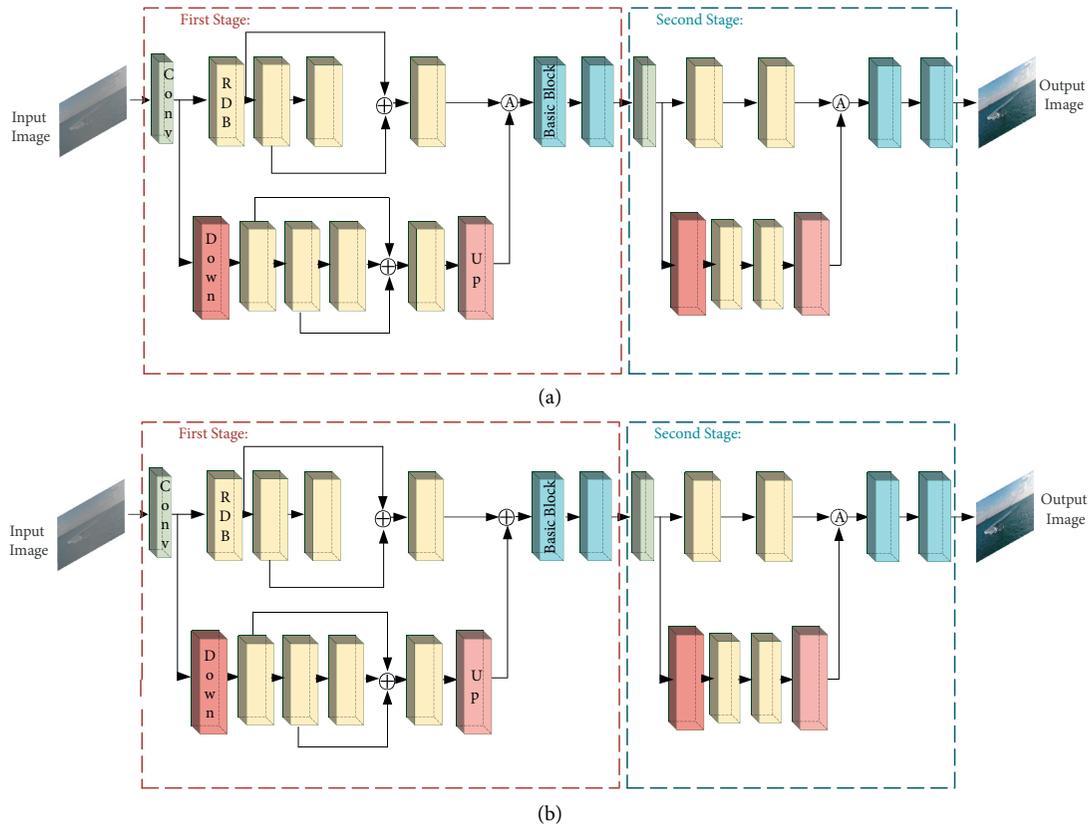


FIGURE 14: Structure diagrams of two contrast networks. (a) The structure of Network A. (b) The structure of Network B.

TABLE 4: Quantitative comparison of different structure networks.

	Network A	Network B	Our
PSNR	23.47	27.312	30.279
SSIM	0.904	0.939	0.964

3.4. Application in Ship Detection. In this paper, the dehazing network is regarded as a preprocessing step of the ship detection problem. Therefore, experiments are constructed to verify the performance of ship detection in fog. For ship detection, the YOLO V5 network which was introduced in June 2020 is used in our previous work [44]. It has been fully proved that YOLO V5 has strong applicability in ship detection under normal weather, and its detection accuracy can reach 98%. However, ship detection under complex sea weather, such as sea fog, has not been studied. In this paper, the same network structure and experimental setup as [45]

are used to verify the availability of the dehazing algorithm in ship detection.

The test set 1 contains 100 hazy ship images, and the corresponding hazy-free images estimated by DTDNet are stored as the test set 2. The images of the two datasets are input into YOLO V5 to obtain the ship detection results, respectively. According to the detection results, the confidence degree of the ship image processed by the dehazing network designed in this paper is 0.23 higher than that of the original fogged image directly detected. It is proved that the accuracy of ship detection performance can be improved with the dehazing network.

4. Discussion and Conclusions

The visual perception system performance of USV is affected by the acquired image quality. However, the image quality obtained in hazy weather is seriously degraded. In this paper,

a CNN-based dual-channel and two-stage image dehazing network is proposed as the processing procedure for high-level visual perception problems, such as ship detection. For DTDNet, which is modified on FCTF-Net, up/down-sampling is introduced to expand the network channels to acquire deeper multiscale features. At the same time, the attention mechanism is utilized to realize the adaptive fusion of multichannel features. Finally, the improved MSE loss function using the perceptual loss function effectively solves the instability problem caused by the difference of the MSE loss function due to the outliers. In addition, a self-built dataset is created based on the atmospheric scattering model. The experimental results on the synthetic dataset and real-world hazy images show that the proposed DTDNet is superior to other networks in both visual and quantitative indicators, which can be effectively used to improve the ship detection performance on the foggy sea.

However, the proposed network designed in this paper still has many shortcomings, such as processing speed and color distortion problem, which need further research. First, lightweight networks and parallel computing will be considered to improve the processing speed to adapt to real-time applications. New feature extraction network with reduced unnecessary modules to better acquire hazy image details will be studied in the future. Secondly, the performance of supervised convolutional neural network-based model is always limited by the dataset. In practice, it is difficult to obtain large-scale and high-quality sea ship fog image training dataset. In the future, semisupervised or unsupervised learning will be considered. Moreover, the combination of learning-based dehazing network and physical-model-based traditional methods will be considered. Finally, large-scale hazy images of ships closer to the actual hazy condition are needed for network training. The dataset used in network training in this paper is synthesized by the commonly used atmospheric scattering model in the field of image dehazing. After training, the network shows a good image dehazing effect on the actual hazy image. However, the dataset synthesis process fails to take into account the difference between fog on the sea and fog on the land, such as the difference of air humidity. Therefore, the construction of synthetic datasets closer to the actual situation and the accumulation of real datasets will be further studied, so as to improve the effect of network model.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare no conflicts of interest.

References

- [1] Y. Guo, Y. Lu, and R. Liu, "Lightweight deep network-enabled real-time low-visibility enhancement for promoting vessel

- detection in maritime video surveillance," *Journal of Navigation*, vol. 75, no. 1, pp. 1–21, 2021.
- [2] X. Yang, H. Sun, X. Sun, M. Yan, Z. Guo, and K. Fu, "Position detection and direction prediction for arbitrary-oriented ships via multitask rotation region convolutional neural network," *IEEE Access*, vol. 6, Article ID 50849, 2018.
- [3] Y. You, J. Cao, Y. Zhang, F. Liu, and W. Zhou, "Nearshore ship detection on high-resolution remote sensing image via scene-mask R-CNN," *IEEE Access*, vol. 7, Article ID 128444, 2019.
- [4] Z. Li, Y. You, and F. Liu, "Analysis on saliency estimation methods in high-resolution optical remote sensing imagery for multiscale ship detection," *IEEE Access*, vol. 8, Article ID 194496, 2020.
- [5] Y. Mao, Y. Yang, Z. Ma, M. Li, H. Su, and J. Zhang, "Efficient low-cost ship detection for SAR imagery based on simplified U-net," *IEEE Access*, vol. 8, Article ID 69753, 2020.
- [6] Z. Zhou, S. Yu, and K. Liu, "A real-time algorithm for visual detection of high-speed unmanned surface vehicle based on deep learning," in *Proceedings of the IEEE International Conference on Signal, Information and Data Processing (ICSIDP)*, pp. 1–5, Chongqing, China, December 2019.
- [7] Y. Shan, X. Zhou, S. Liu, Y. Zhang, and K. Huang, "SiamFPN: a deep learning method for accurate and real-time maritime ship tracking," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 99, p. 1, 2020.
- [8] K. Liu, S. Yu, and S. Liu, "An improved InceptionV3 network for obscured ship classification in remote sensing images," *Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, pp. 4738–4747, 2020.
- [9] R. W. Liu, W. Q. Yuan, X. Q. Chen, and Y. X. Lu, "An enhanced CNN-enabled learning method for promoting ship detection in maritime surveillance system," *Ocean Engineering*, vol. 235, Article ID 109435, 2021.
- [10] R. Wang, Y. You, Y. Zhang, W. Zhou, and J. Liu, "Ship detection in foggy remote sensing image via scene classification R-CNN," in *Proceedings of the International Conference on Network Infrastructure and Digital Content (IC-NIDC)*, pp. 81–85, Guiyang, China, August 2018.
- [11] J. Liu, R. W. Liu, J. N. Sun, and T. Y. Zeng, "Rank-one prior: toward real-time scene recovery," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Nashville, TN, USA, June 2021.
- [12] Z. Zhu, H. Wei, G. Hu, Y. Li, G. Qi, and N. Mazur, "A novel fast single image dehazing algorithm based on artificial multiexposure image fusion," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–23, 2021.
- [13] M. Zheng, G. Qi, Z. Zhu, Y. Li, H. Wei, and Y. Liu, "Image dehazing by an artificial image fusion method based on adaptive structure decomposition," *IEEE Sensors Journal*, vol. 20, no. 14, pp. 8062–8072, 2020.
- [14] A. Acharya and A. V. Giri, "Contrast improvement using local gamma correction," in *Proceedings of the 6th International Conference on Advanced Computing and Communication Systems (ICACCS)*, Tokyo, Japan, November 2020.
- [15] H. Singh, A. Kumar, L. K. Balyan, and G. K. Singh, "Dark image enhancement using optimally compressed and equalized profile based parallel gamma correction," in *Proceedings of the International Conference on Communication and Signal Processing (ICCSP)*, pp. 1299–1303, Chennai, India, April 2017.
- [16] Z. Zhu, Y. Luo, H. Wei, Y. Li, and P. Li, "Atmospheric light estimation based remote sensing image dehazing," *Remote Sensing*, vol. 13, no. 13, p. 2432, 2021.

- [17] 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.
- [18] S. Lee, S. Yun, and J. H. Nam, "A review on dark channel prior based image dehazing algorithms," *EURASIP Journal on Image and Video Processing*, vol. 2016, no. 1, pp. 1–23, 2016.
- [19] R. M. Yousaf, H. A. Habib, Z. Mehmood, A. Banjar, and O. Aboulola, "Single image dehazing and edge preservation based on the dark channel probability-weighted moments," *Mathematical Problems in Engineering*, vol. 2019, Article ID 9721503, 11 pages, 2019.
- [20] R. M. Yousaf, H. A. Habib, Z. Mehmood, and M. Bilal, "Image dehazing based on dark channel spatial stimuli gradient model and image morphology," *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, pp. 1–13, 2020.
- [21] Z. Rong and W. L. Jun, "Improved wavelet transform algorithm for single image dehazing," *Optik*, vol. 125, no. 13, pp. 3064–3066, 2014.
- [22] H. Wu and Z. Tan, "An image dehazing algorithm based on single-scale retinex and homomorphic filtering," in *Proceedings of the International Conference in Communications, Signal Processing, and Systems*, pp. 1482–1493, Springer, Beijing, China, September 2019.
- [23] W. Wang and X. Yuan, "Recent advances in image dehazing," *IEEE/CAA Journal of Automatica Sinica*, vol. 4, no. 3, pp. 410–436, 2017.
- [24] J.-B. Wang, N. He, and L.-L. Zhang, "Single image dehazing with a physical model and dark channel prior," *Neurocomputing*, vol. 149, pp. 718–728, 2015.
- [25] W. Mei and X. Li, "Single image dehazing using dark channel fusion and haze density weight," in *Proceedings of the IEEE 9th International Conference on Electronics Information and Emergency Communication (ICEIEC)*, pp. 579–585, Beijing, China, July 2019.
- [26] D. Yue, Y. Huang, H. Tang, and X. Huang, Z. Qin, "Single image dehazing jointly utilizing dark channel prior and guided filtering in dual-tree complex wavelet domain," in *Proceedings of the 17th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP)*, pp. 10–13, Chengdu, China, December 2020.
- [27] 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.
- [28] B. Li, X. Peng, Z. Wang, J. Xu, and D. A. O. D. N. Feng, "All-in-One dehazing network," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp. 4780–4788, Venice, Italy, October 2017.
- [29] C. Li, J. Guo, F. Porikli, H. Fu, and Y. Pang, "A cascaded convolutional neural network for single image dehazing," *IEEE Access*, vol. 6, Article ID 24887, 2018.
- [30] J. Li, G. Li, and H. Fan, "Image dehazing using residual-based deep CNN," *IEEE Access*, vol. 6, Article ID 26831–26842, 2018.
- [31] W. Wang, F. Chang, T. Ji, and X. Wu, "A fast single-image dehazing method based on a physical model and gray projection," *IEEE Access*, vol. 6, pp. 5641–5653, 2018.
- [32] G. Tang, L. Zhao, R. Jiang, and X. Zhang, "Single image dehazing via lightweight multiscale networks," in *Proceedings of the IEEE International Conference on Big Data (Big Data)*, pp. 5062–5069, Los Angeles, CA, USA, December 2019.
- [33] B. Xiao, Z. Zheng, X. Chen, C. Lv, Y. Zhuang, and T. Wang, "Single UHD Image Dehazing via Interpretable Pyramid Network," p. 4, 2022, <https://arxiv.org/abs/2202.08589>.
- [34] Y. Li and X. Chen, "A coarse-to-fine two-stage attentive network for haze removal of remote sensing images," *IEEE Geoscience and Remote Sensing Letters*, vol. 18, no. 10, pp. 1751–1755, 2021.
- [35] Z. Zhu, Y. Luo, G. Qi, J. Meng, and N. Mazur, "Remote sensing image defogging networks based on dual self-attention boost residual octave convolution," *Remote Sensing*, vol. 13, no. 16, p. 3104, 2021.
- [36] S. Jiao, J. Wang, G. Hu, Z. Pan, L. Du, and J. Zhang, "Joint attention mechanism for person Re-identification," *IEEE Access*, vol. 7, Article ID 90506, 2019.
- [37] W. Wang, X. Yuan, X. Wu, and Y. Liu, "Dehazing for images with large sky region," *Neurocomputing*, vol. 238, pp. 365–376, 2017.
- [38] X. Li, G. Liu, and J. Ni, "Autofocusing of ISAR images based on entropy minimization," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 35, no. 4, pp. 1240–1252, 1999.
- [39] E. Gundogdu, S. Berkan, Y. Veysel, and K. Aykut, "Marvel: a largescale image dataset for maritime vessels," in *Proceedings of the 13th Asian Conference on Computer Vision (ACCV)*, pp. 165–180, Taipei, Taiwan, China, November 2016.
- [40] M. M. Zhang, J. Choi, K. Daniilidis, M. T. Wolf, and C. V. A. I. S. Kanan, "A dataset for recognizing maritime imagery in the visible and infrared spectrums," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 10–16, Boston, MA, USA, June 2015.
- [41] "Vessel photo database," p. 2, 2020, <https://www.fleetmon.com/community/photos>.
- [42] J. P. Oakley and B. L. Satherley, "Improving image quality in poor visibility conditions using a physical model for contrast degradation," *IEEE Transactions on Image Processing*, vol. 7, no. 2, pp. 167–179, 1988.
- [43] B. Li, W. Ren, D. Fu et al., "A Benchmark for Single Image Dehazing," vol. 28, p. 1, 2019, <https://arxiv.org/abs/1712.04143>, 2017.
- [44] T. Liu, B. J. Zhou, Y. S. Zhao, and S. Yan, "Ship detection algorithm based on improved YOLO V5," in *Proceedings of the 2021 6th International Conference on Automation, Control and Robotics Engineering (CACRE)*, pp. 483–487, Dalian, China, July 2021.
- [45] B. Xiao, Z. Zheng, X. Chen, C. Lv, Y. Zhuang, and T. Wang, "Single UHD Image Dehazing via Interpretable Pyramid Network," 2022, <https://arxiv.org/abs/2202.08589>.