

Research Article

CNN-Enabled Visibility Enhancement Framework for Vessel Detection under Haze Environment

Yuxu Lu ^{1,2}, Yu Guo ^{1,2} and Maohan Liang ^{1,2}

¹Hubei Key Laboratory of Inland Shipping Technology, School of Navigation, Wuhan University of Technology, Wuhan, China

²National Engineering Research Center for Water Transport Safety, Wuhan, China

Correspondence should be addressed to Maohan Liang; mhliang@whut.edu.cn

Received 16 January 2021; Revised 5 March 2021; Accepted 24 April 2021; Published 20 May 2021

Academic Editor: Yi-Sheng Lv

Copyright © 2021 Yuxu Lu 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.

Maritime images captured under haze environment often have a terrible visual effect, making it easy to overlook important information. To avoid the failure of vessel detection caused by fog, it is necessary to preprocess the collected hazy images for recovering vital information. In this paper, a novel CNN-enabled visibility dehazing framework is proposed, consisting of two subnetworks, that is, Coarse Feature Extraction Module (C-FEM) and Fine Feature Fusion Module (F-FFM). Specifically, C-FEM is a multiscale haze feature extraction network, which can learn information from three scales. Correspondingly, F-FFM is an improved encoder-decoder network to fuse multiscale information obtained by C-FEM and enhance the visual effect of the final output. Meanwhile, a hybrid loss function is designed for monitoring the multiscale output of C-FEM and the final result of F-FFM simultaneously. It is worth mentioning that massive maritime images are considered the training dataset to further adapt the vessel detection task under haze environment. Comprehensive experiments on synthetic and realistic images have verified the superior effectiveness and robustness of our CNN-enabled visibility dehazing framework compared to several state-of-the-art methods. Our method preprocesses images before vessel detection to demonstrate our framework has the capacity of promoting maritime video surveillance.

1. Introduction

1.1. Background and Related Work. It is well known that the maritime surveillance system is an indispensable part of vessel traffic services [1]. As an efficient, convenient, and intuitive monitoring method, Closed Circuit Television (CCTV) is thus widely applied to critical regions, for example, ports and waterways. As shown in Figure 1, significant information in the images, however, is easily buried under the haze. Therefore, it is difficult for maritime regulatory authority to effectively extract detailed information (e.g., monitoring targets and water traffic conditions) from degraded images, which seriously affects maritime supervision efficiency. Besides, the low-quality images collected under haze environment have also brought severe challenges to intelligent surveillance methods based on vessel detection, recognition, and tracking [2–5]. To improve the maritime safety surveillance capability under haze environment, it is

necessary to restore images under CCTV monitoring. In current literature, dehazing methods can be categorized into image enhancement-based methods, physical model-based methods, and deep learning-based methods.

1.1.1. Image Enhancement-Based Methods. Early research mainly enhanced the contrast of hazy images to highlight the scene characteristics of the interest region. Histogram Equalization (HE) [6] is a classic enhancement method devoted to enhancing the contrast by stretching the dynamic range of image pixel values. In current literature, HE-based methods can be divided into two categories, that is, global and local histogram equalization. Since the global histogram equalization can enhance the entire image by single mapping, it has the characteristics of simple principle and fast calculation. However, these methods often ignore the local information, resulting in the haze-free images having poor performance. To solve this problem, Stark et al. [7] proposed



FIGURE 1: Maritime images captured under haze environment.

an adaptive local histogram equalization method. Subsequently, Kim et al. [8] proposed a nonoverlapping subblock histogram equalization method to reduce the blocky effect and computational complexity. Retinex theory-based image dehazing method is devoted to separating the illumination and reflection from the hazy image and enhancing the image by reducing the illumination impact. Jobson et al. [9] first used the Gaussian filter to obtain a smooth illumination according to the Retinex theory and thus proposed a single-scale Retinex (SSR). To avoid color distortion, Rahman et al. [10] proposed a multiscale Retinex algorithm with color restoration (MSRCR) by introducing a color compensation factor. To sum up, the image generated by these methods has higher contrast and color fidelity, but the halo often appears on the edge of the interest object.

1.1.2. Physical Model-Based Methods. These methods are proposed based on a certain physical model that describes the process of image degradation under haze weather. Because these methods use mathematical methods to describe the haze formation process based on light scattering, the final restored target is clear and natural. Physical model-based methods include the following categories, that is, depth-based method and prior-based method. The depth-based methods mainly obtain depth information through a specific method and then get stable model parameters. Finally, the potentially clear image can be obtained by the atmospheric scattering model. For instance, Oakley et al. [11] first used radar and other types of equipment to measure the shooting scene depth. Hautiere et al. [12] proposed an image dehazing algorithm based on the 3D geographic model for vehicle vision systems. Although these methods have an excellent dehazing effect, they heavily rely

on distance measuring equipment. Therefore, Liu et al. [13] proposed a dehazing method to estimate the depth map through a second-order variational framework. In contrast, the prior-based method mainly analyzes haze formation and relies on specific prior information to achieve image dehazing. Dark channel prior (DCP) [14] and its improvements [15–17] have an excellent performance in the image dehazing task. Through numerous statistics on outdoor haze-free images, He et al. proposed DCP based on the assumption that most local color blocks contain some pixels with very low intensity in at least one color channel. Zhu et al. proposed a novel linear color attenuation prior [18], based on the difference between the brightness and the saturation of pixels within the hazy image. Subsequently, a nonlocal prior dehazing method [19] is employed to obtain the nonlocal transmission map from the haze-line property. To reduce halo and unnatural artifacts, a low-complexity color ellipsoid prior [20] is designed to accurately and swiftly estimate the transmission map. In current literature, several variational model-based transmission estimation methods [15, 21, 22] are also proposed. Although prior-based methods have verified excellent dehazing performance, they may cause a loss in color fidelity under certain circumstances and fail to obtain pleasing visual effects on maritime images.

1.1.3. Deep Learning-Based Method. Affected by the excellent results of machine learning technology in computer vision, the CNN-enabled dehazing method has gradually become a new research direction. For instance, Tang et al. [23] proposed a learning-based approach to systematically investigate different haze-relevant features and identify the best feature combination for image dehazing. Zhu et al. [18] designed the color attenuation prior model to estimate the

scene depth. Meanwhile, a supervised learning method is used to obtain the scene depth and the atmospheric scattering model parameters. However, this method fails to estimate the scene depth in the white environment. To further improve the deep learning-based dehazing methods performance, Cai et al. [24] first constructed a convolutional neural network (DehazeNet) to learn the mapping relationship between hazy images and transmission. DehazeNet employed artificially synthesized hazy images of different concentrations as the dataset. The trained DehazeNet can directly estimate the corresponding transmission and restore the potentially clear image according to the traditional atmospheric light scattering model. Ren et al. [25] proposed a Multiscale Convolutional Neural Network (MSCNN). MSCNN used the New York University indoor image depth database [26] to synthesize different hazy images as the dataset, making the hazy image more realistic. Subsequently, Zhao et al. [27] used outdoor scenes synthetic images and proposed a fully convolutional neural network model to estimate the transmission. Compared with DehazeNet and MSCNN, this network has better visual performance. However, it produces more parameters and calculations. To simplify the calculation, Li et al. [28] designed an end-to-end light-weight convolutional neural network (AOD-Net) that effectively balances calculation speed and visual effects. Inspired by image denoising, Du et al. [29] proposed a Deep Residual Learning (DRL) network to reconstruct the potential image. Besides, Chen et al. [30] proposed an end-to-end gated context aggregation network to directly restore the final haze-free image. It is worth noting that if the training datasets do not contain the geometric features presented in the haze-free target, it is usually difficult to produce satisfactory image quality. Therefore, it is necessary to design a CNN-enabled visibility enhancement framework for vessel detection under haze environment to further improve maritime video surveillance efficiency.

1.2. Contributions. This paper presents a CNN-enabled framework for practically solving vessel detection problem under haze environment. The main contribution of our method differs from others in the following aspects:

- (i) A CNN-enabled visibility dehazing framework is proposed to improve the visibility of maritime images. Specifically, this framework includes a *Coarse Feature Extraction Module* (C-FEM) for capturing multiscale features and *Fine Feature Fusion Module* (F-FFM) for information fusion and enhancement.
- (ii) To improve the generalization of the proposed network, we design a novel hybrid loss function to supervise the multiscale outputs of C-FEM and the final output of F-FFM simultaneously.
- (iii) Image dehazing and vessel detection experiments under haze conditions are conducted to verify our superior performance compared to several state-of-the-art methods.

1.3. Construction. The remainder of this paper is divided into the following sections. Section 2 mainly describes the problem formulation related to the imaging model. In Section 3, a CNN-enabled visibility enhancement framework is proposed to improve the visual effect of hazy images. Implementation details and experiments are implemented in Section 4. Finally, we conclude our main contributions in Section 5.

2. Problem Formulation

2.1. Atmospheric Scattering Model. Video images collected by maritime video surveillance system under haze conditions often have poor visual quality. As shown in Figure 2, Narasimhan et al. [31] proposed the atmospheric scattering model to divide the light irradiance into the incident light attenuation part $J(x)e^{-\beta d(x)}$ and the atmospheric light imaging part $A_{\infty}(1 - e^{-\beta d(x)})$. The incident light attenuation model considers that the reflected light by the vessel surface is scattered and attenuated by particulate impurities in the air, reducing the intensity of light reaching the imaging system. Note that as the propagation distance increases, the reflected light intensity decays exponentially. On the contrary, the atmospheric light imaging model believes that light intensity scattered by natural light enters the imaging system to participate in imaging. As the propagation distance increases, the scattered light intensity will gradually increase. Finally, the images collected by the imaging system under haze environments exhibit degradation phenomena such as low contrast, blurred images, and color distortion under the combined action of these two models. Mathematically, the atmospheric light scattering model can be expressed as

$$I(x) = J(x)e^{-\beta d(x)} + A_{\infty}(1 - e^{-\beta d(x)}), \quad (1)$$

where J and I , respectively, denote the hazy image and haze-free image, A_{∞} and β represent the atmospheric light value and scattering coefficient, x is the image pixel index, and d is the distance between the scene point and the imaging system, that is, field depth. When we set $t(x) = e^{-\beta d(x)}$, equation (1) can thus be rewritten as follows:

$$I(x) = J(x)t(x) + A_{\infty}(1 - t(x)), \quad (2)$$

with t being the transmission. According to equation (2), the restoration haze-free image can be easily obtained by

$$J(x) = \frac{I(x) - A_{\infty}}{t(x)} + A_{\infty}. \quad (3)$$

2.2. Transformed Formula. According to equation (3), we can obtain a satisfactory haze-free image J by accurately estimating A_{∞} and t . However, it is intractable to estimate two parameters simultaneously. For the sake of better performance of the end-to-end network, Li et al. [28] proposed the transformed atmospheric scattering model, which can be given by

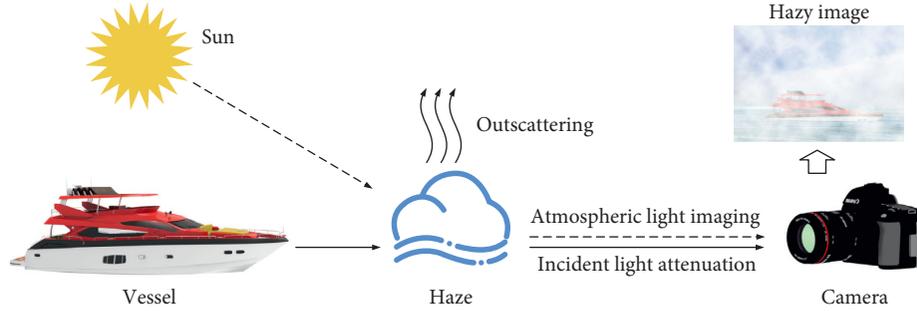


FIGURE 2: The principle of the atmospheric scattering model.

$$J(x) = \mathcal{K}(x)I(x) - \mathcal{K}(x) + 1, \quad (4)$$

where \mathcal{K} is a particular parameter to integrate A_∞ and t ; that is, $\mathcal{K}(x) = ((1/t(x))(I(x) - A_\infty) + A_\infty - 1/I(x) - 1)$. It is worth noting that hazy maritime images usually contain background (i.e., sky and water regions). Many statistical features-based methods, for example, DCP and maximum local contrast, often fail to obtain ideal transmission maps. Deep learning-based methods do not rely on these statistical features and can learn the mapping of hazy and haze-free images. Therefore, we will propose the CNN-enabled visibility enhancement network to effectively improve the quality of hazy maritime images and improve vessel detection accuracy.

3. CNN-Enabled Visibility Enhancement Framework

In this section, a CNN-enabled visibility enhancement framework is proposed to process hazy maritime images shown in Figure 3. This framework consists of two sub-networks, that is, *Coarse Feature Extraction Module* (C-FEM) and *Fine Feature Fusion Module* (F-FFM). In this work, C-FEM is introduced to learn multiscale hazy features. Meanwhile, F-FFM, an improved encoder-decoder network, is proposed to fuse and enhance the hazy image and the multiscale output obtained by C-FEM. Once our method gets the sharp image, it can easily detect the vessel containing the image by any target detection method.

$$I * w_{(t)} = \sum_{p+q=t} I(p)w(q), \quad (5)$$

$$I * {}_d w_{(t)} = \sum_{p+lq=t} I(p)w(q), \quad (6)$$

3.1. C-FEM. C-FEM is a module for initial extracting the features of the hazy image. In particular, C-FEM can perform mapping learning on three scales (i.e., 1, 1/4, and 1/16) to obtain coarse feature information with different resolutions simultaneously. Figure 4 shows the network architecture of C-FEM under one resolution, which is only composed of six convolutions. In this work, dilated convolution is embedded to increase the reception field of C-FEM. According to our research, dilated convolution can

reduce the loss of spatial features without reducing the receptive field. However, the use of dilated convolution may increase the risk of spatially continuous information loss, destroying image feature information (especially edges). To alleviate the interference caused by dilated convolution, we combine standard convolution (Conv) [32] and dilated convolution (DConv) with improving the detailed information extraction ability. DConv can effectively solve this difficult problem with different receptive fields by adjusting the dilation rate value. Formally, standard convolution and dilated convolution are, respectively, defined as follows: where I is the discrete signal, w is convolution kernel, subscript (\cdot) is the position of a discrete signal, $d \in \mathbb{Z}^+$ is the dilation factor, and $*_d$ is dilated convolutions with a factor d . The only difference between standard convolution and dilated convolution is the influence of the dilation factor d on the multiplication position of $I(p)$ and $w(q)$. Dilated convolution benefits from $w(q)$ are no longer limited to a fixed receptive field, and the dilation factor d can be adjusted to have a larger receptive field. In this work, the method of fusing Conv and DConv can reduce the loss of spatial information caused by the excessive dilated rate and fully consider long- and short-distance information to present a better visual effect. Furthermore, Instance Normalization (IN) [33] and Rectified Linear Unit (ReLU) [34] are deployed after each Conv layer. Meanwhile, the feature map channels of the first five convolution outputs are set to 32.

$$\mathcal{L}^{\text{MAE}}(\hat{J}, J) = |\hat{J} - J|. \quad (10)$$

According to our research, most deep learning-based dehazing methods rely on more complex network models to obtain better visual effects. When the model is relatively simple, it is usually hard to learn the fog feature, causing information damage to potential images. In contrast, the imaging model $J(x) = \mathcal{K}(x)I(x) - \mathcal{K}(x) + 1$ reduces the algorithm complexity, making it easier for the network to extract information. The introduction of this model makes it possible for a simple network model to extract potential multiscale features from the original image. It is worth noting that the output of C-FEM only introduced to provide a prior is not used as the final result. Simultaneously, it has a faster calculation speed and can satisfy the needs of real-time processing.

3.2. F-FFM. Coarse feature maps of three resolutions (i.e., 1, 1/4, and 1/16) have been obtained by C-FEM, which

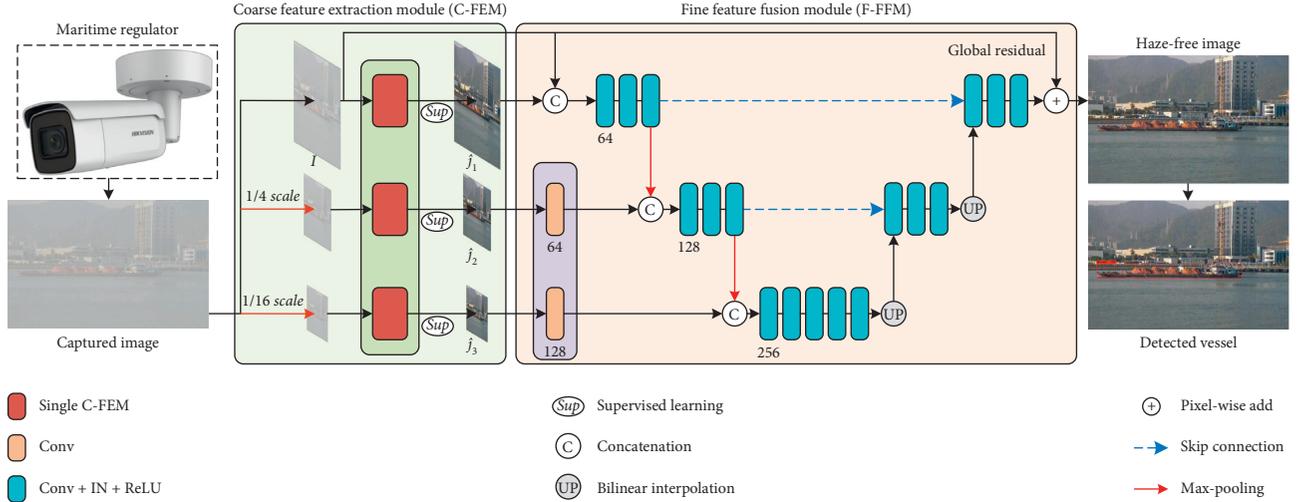


FIGURE 3: The flowchart of our method. From top to bottom in C-FEM: subimages reduced to 1, 1/4, and 1/16 of the original image, respectively.

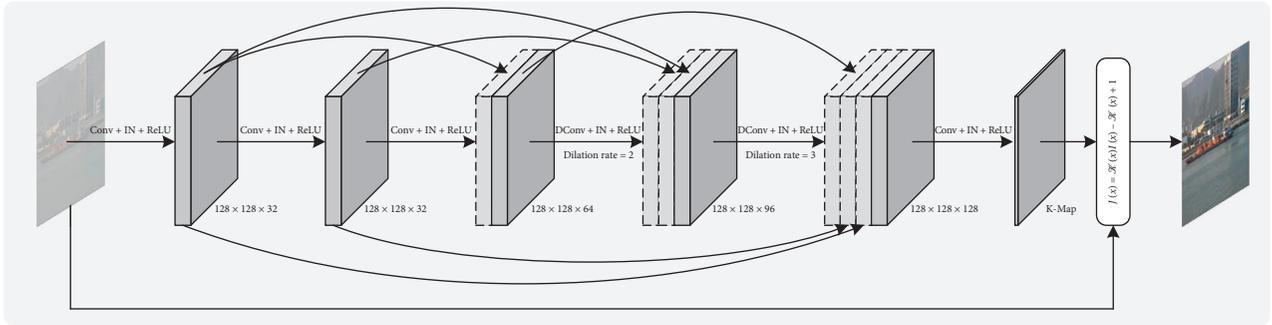


FIGURE 4: Network architecture of single C-FEM.

contains most of the potentially clear image information. The feature map information obtained by a standard encoder-decoder CNN is usually found irregularly. When the prior information obtained by C-FEM is introduced to the encoder, we believe that F-FFM can obtain better parameters and accelerate the convergence speed. When the feature maps are fused at different scales, the deep network can further extract edge detail information. Table 1 shows that the architecture of Fine Feature Fusion Modul (F-FFM) is a special encoder-decoder structure. Specifically, F-FFM only performs two downsampling operations and merges with the corresponding output of C-FEM. Both the encoder and the decoder consist of the same module, that is, a 3×3 convolution filter (Conv) [32], Instance Normalization (IN) [33], and Rectified Linear Unit (ReLU) [34]. Maximum pooling and bilinear interpolation are exploited to perform down- and upsampling operations on the feature map, respectively. Different from traditional encoder-decoder structures, our F-FFM encoder integrates the output of C-FEM. This strategy can guide F-FFM to learn the mapping of hazy images and haze-free targets. To better preserve the boundary details of the input, we adopt a global skip connection strategy to further ensure the details of the output image. In other words, the output of the last convolution and the input image is directly added as the output of F-FFM,

and we find that it can significantly improve the dehazing effect through comparative experiments.

3.3. Loss Function. To robustly learn the multiscale mapping relationship between hazy image and haze-free image, a specific loss function \mathcal{L}^{C-FEM} is proposed. As shown in Figure 3, C-FEM has three scale outputs (i.e., \hat{J}_1 , \hat{J}_2 , and \hat{J}_3). These three images sequentially have 1, 1/4, and 1/16 of the original image size. Subsequently, the maximum pooling operation is used to obtain clear images with three scales named J_1 , J_2 , and J_3 , which, respectively, correspond to the scale of \hat{J}_1 , \hat{J}_2 , and \hat{J}_3 . In this work, Mean Square Error (MSE) loss function is employed to constrain each scale output of C-FEM; that is,

$$\mathcal{L}^{C-FEM} = \lambda_1 \mathcal{L}^{MSE}(\hat{J}_1, J_1) + \lambda_2 \mathcal{L}^{MSE}(\hat{J}_2, J_2) + \lambda_3 \mathcal{L}^{MSE}(\hat{J}_3, J_3), \quad (7)$$

where $\mathcal{L}^{MSE}(\hat{J}_*, J_*) = (\hat{J}_* - J_*)^2$, λ_1 , λ_2 , and λ_3 are trade-off parameters of corresponding loss functions. To further preserve the high-frequency details of the potential haze-free image while eliminating boundary artifacts, a hybrid loss function \mathcal{L}^{F-FFM} is introduced to limit the ground truth J and the predicted restored image \hat{J} ; that is,

TABLE 1: Network architecture of F-FFM.

Input	Layers	Type	Channels	Filter	Stride	Output	Size
$I + \hat{J}_1$	1	Convolutional	64	3×3	1	\downarrow_1	128×128
\downarrow_1	3	Convolutional	64	3×3	1	\downarrow_2	128×128
\downarrow_2		Max pooling	64	3×3	2	\downarrow_3	
$\downarrow_3 + \hat{J}_2$	1	Convolutional	128	3×3	1	\downarrow_4	64×64
\downarrow_4	3	Convolutional	128	3×3	1	\downarrow_5	64×64
\downarrow_5		Max pooling	128	3×3	2	\downarrow_6	
$\downarrow_6 + \hat{J}_3$	1	Convolutional	256	3×3	1	\downarrow_7	32×32
\downarrow_7	5	Convolutional	256	3×3	1	\downarrow_8	32×32
\downarrow_8		Bilinear interpolation				\downarrow_9	
$\downarrow_9 + \downarrow_4$		Skip connection	128	1×1	1	\downarrow_{10}	64×64
\downarrow_{10}	1	Convolutional	128	3×3	1	\downarrow_{11}	64×64
\downarrow_{11}	3	Convolutional	128	3×3	1	\downarrow_{12}	64×64
\downarrow_{12}		Bilinear interpolation				\downarrow_{13}	
$\downarrow_{13} + \downarrow_1$		Skip connection	64	1×1	1	\downarrow_{14}	128×128
\downarrow_{14}	1	Convolutional	64	3×3	1	\downarrow_{15}	128×128
\downarrow_{15}	3	Convolutional	64	3×3	1	\downarrow_{16}	128×128
$\downarrow_{16} + I$		Global residual	3	3×3	1	\hat{J}	128×128

$$\begin{aligned} \mathcal{L}^{F-FFM} = & \lambda_4 \mathcal{L}^{MS-SSIM}(\hat{J}, J) + \lambda_5 \mathcal{L}^{MAE}(\hat{J}, J) \\ & + \lambda_6 \mathcal{L}^{Edge}(\hat{J}, J) + \lambda_7 \mathcal{L}^{TV}(\hat{J}), \end{aligned} \quad (8)$$

with λ_4 , λ_5 , λ_6 , and λ_7 being the penalty weights. Multiscale structural similarity (MS-SSIM) [35] is firstly employed to constrain the structure, brightness, and contrast of the image. The MS-SSIM loss function can be defined as follows:

$$\mathcal{L}^{MS-SSIM}(\hat{J}, J) = 1 - \text{MSSSIM}(\hat{J}, J), \quad (9)$$

with MSSSIM being the calculation operation of the multiscale structural similarity index between two images. The hazy image inevitably has a low contrast phenomenon in local regions, resulting in color distortion. To solve this problem, the Mean Absolute Error (MAE) loss function \mathcal{L}^{MAE} is introduced as a part of \mathcal{L}^{F-FFM} , which can reduce the color distortion problem to a certain extent. In particular, \mathcal{L}^{MAE} is defined as

The high-frequency detail information is easily destroyed in the process of image dehazing. To further improve the fidelity and authenticity of details, we propose an additional edge loss function [36] to limit the high-frequency components, for example, edge and texture. \mathcal{L}^{Edge} can be written as

$$\mathcal{L}^{Edge}(\hat{J}, J) = \sqrt{(\text{Lap}(\hat{J}) - \text{Lap}(J))^2 + \varepsilon^2}, \quad (11)$$

where $\text{Lap}(J)$ and $\text{Lap}(\hat{J})$ represent edge maps extracted from J and \hat{J} via the Laplacian operator, respectively. The penalty coefficient ε is empirically set to 10^{-3} . In addition, the Total Variation (TV) loss function [37] is exploited to suppress the pixel-jump problem, which can be given by where ∇_h and ∇_v represent the operators of the horizontal and vertical gradients, respectively. We refer interested readers to [35–37] for more details on calculations of MS-SSIM, edge loss, and TV. To sum up, the total loss function can be written as follows:

$$\mathcal{L}^{\text{Total}} = \gamma_1 \mathcal{L}^{C-FEM} + \gamma_2 \mathcal{L}^{F-FFM}, \quad (13)$$

where γ_1 and γ_2 are the penalty coefficient of \mathcal{L}^{C-FEM} and \mathcal{L}^{F-FFM} . By comparative experiment, we manually selected the optimal weight of all loss functions; that is, $\lambda_1 = 1$, $\lambda_2 = 1$, $\lambda_3 = 1$, $\lambda_4 = 0.8$, $\lambda_5 = 0.2$, $\lambda_6 = 0.05$, $\lambda_7 = 0.05$, $\gamma_1 = 1$, and $\gamma_2 = 1$.

4. Experimental Results and Analysis

This section will describe all the implementation details of network training, including dataset construction and network parameter settings. We will compare our method with several state-of-the-art dehazing methods on both synthetic and realistic hazy maritime images. To prove that our method can improve detection accuracy, our proposed framework will be employed in vessel detection tasks under haze environment.

4.1. Comparison Methods and Evaluation Indicators. Our method will be compared with four handcrafted prior-based methods and three deep learning-based methods. For the sake of fair comparison, the parameters of other competing dehazing methods are provided by the authors' code.

- (1) DCP: *Dark Channel Prior-Based Method* [14]. Through numerous statistics on outdoor haze-free images, DCP is proposed based on the assumption that most local color blocks contain some pixels with very low intensity in at least one color channel. According to this statistic prior and the haze imaging model, a high-quality haze-free image can be directly obtained.
- (2) GRM: *Gradient Residual Minimization-Based Method* [16]. This method first proposes the depth-edge-aware smoothing algorithm to refine the transmission map generated by local priors. Meanwhile, Gradient Residual Minimization (GRM) is

introduced during the image recovery process. By comparison, the GRM-based method can jointly recover the haze-free image and explicitly minimize possible visual artifacts in it.

- (3) HL: *Haze-Lines-Based Method* [38]. This method finds that the pixel values of a hazy image can be modeled as lines intersecting at the air light. Based on this prior condition, a novel haze-lines-based method is proposed to restore the hazy image better. It is worth noting that the complexity of HL is linear in the number of pixels, having higher computational efficiency.
- (4) F-LDCP: *Fusion of Luminance and Dark Channel Prior-Based Method* [39]. To make the sky region more natural in long-shot images, a Fusion of Luminance and Dark Channel Prior (F-LDCP) method is proposed. The transmission maps estimated by the brightness model and the DCP model are fused through a soft segmentation.
- (5) MSCNN: *Multiscale Convolutional Neural Networks* [25]. To learn the practical features of a hazy image, a multiscale deep network (MSCNN) is designed to address the image dehazing problem. MSCNN can be divided into the coarse-scale network and fine-scale network. The coarse-scale network can learn a holistic estimation of the scene transmission, and the fine-scale network is used to optimize the obtained transmission. Finally, the haze-free image can be obtained by the atmospheric scattering model.
- (6) AOD-Net: *All-in-One Dehazing Network* [28]. AOD-Net, a light-weight CNN, is designed according to the reformulated atmospheric scattering model. This network replaces the atmospheric light value and transmission with one parameter. It is worth mentioning that AOD-Net has been embedded in other deeper models (e.g., Faster R-CNN) to improve the advanced tasks of hazy images.
- (7) GCA-Net: *Gated Context Aggregation Network* [30]. GCA-Net is an end-to-end Gated Context Aggregation Network. In particular, the latest smoothed dilation technology is designed to eliminate gridding artifacts caused by the extensive-used dilated convolution with negligible additional parameters.

$$\mathcal{L}^{\text{TV}}(\hat{J}, J) = (\nabla_h \hat{J})^2 + (\nabla_v \hat{J})^2, \quad (12)$$

In synthetic and realistic experiments, we will compare these methods with our proposed method. In addition, three full-reference indicators, that is, Peak-Signal-to-Noise Ratio (PSNR) [40], SSIM [41], and Feature Similarity (FSIM) [42], are introduced to evaluate the dehazing performance in the synthetic experiment. Meanwhile, one popular no-reference image quality assessment method, that is, Natural Image Quality Evaluator (NIQE) [43], is also exploited to perform dehazing quality evaluation in the real experiment.

Theoretically, higher values of PSNR, SSIM, FSIM, and lower values of NIQE indicate better visual performance.

4.2. Experimental Datasets and Settings. To guarantee high-quality dehazing results, we tend to select 7000 haze-free maritime images as the dataset and randomly cropped these images into 34000 patches of size 256×256 . In this work, our network is trained for 80 epochs. The learning rate of the first 40 epochs is 10^{-3} and the learning rate of the last 40 epochs is 10^{-4} to increase the convergence rate. In each epoch, the hazy synthetic versions are obtained by equation (2), that is, atmospheric scattering model. In particular, the transmission t and atmospheric light value A_∞ are random constants ranging between (0.2, 0.6) and (0.8, 0.9). All numerical experiments and model training are conducted in Python 3.7 and Matlab2019a environment running on a PC with Intel(R) Core (TM) i7-9750H CPUa 2.60 GHz and a Nvidia GeForce GTX 2080Ti GPU. It takes about 10 hours to train our network with the Pytorch package [44]. The Python source code is available at https://github.com/LouisYuxuLu/JAT_DeHazing.

4.3. Experiments on Synthetic Maritime Datasets. This subsection is devoted to comparing our proposed method with seven popular dehazing methods, that is, DCP [14], GRM [16], HL [38], F-LDCP [39], MSCNN [25], AOD-Net [28], and GCA-Net [30]. In this work, six clear maritime images are exploited to conduct synthetic experiments shown in Figure 5. In particular, we tend to produce 36 degraded images by setting $t \in \{0.2, 0.4, 0.6\}$ and $A_\infty \in \{0.8, 0.9\}$ according to equation (2). To quantitatively evaluate the dehazing performance, three full-reference metrics (i.e., PSNR, SSIM, and FSIM) are employed in this experiment.

For the sake of better visual comparisons, the dehazing versions of hazy images with different degrees obtained by various methods are shown in Figure 6. It can be clearly observed that DCP and HL often make the color unnatural. Meanwhile, due to the incomplete dehazing, the results obtained by DCP easily suffer from the interference of boundary artifacts around the object. Although GRM can get satisfactory visual effects, it requires complex calculations and has the risk of excessive smoothness. F-LDCP can excellently solve the blocking artifacts and halo problems in the sky regions, but the color fidelity in the water regions needs improvement. MSCNN and AOD-Net can handle the low-concentration hazy image. However, the restored versions of the high-concentration hazy images (i.e., hazy images with $t = 0.2$) usually have a poor visual effect. GCA-Net fails in the synthetic experiment, resulting in a nonuniform distribution of fog remaining in the results. By comparison, our method can not only make the restored image visually more natural but also ensure the color reproduction of the sky and water regions.

To further confirm the superiority of our method, the quantitative results of PSNR, SSIM, and FSIM are shown in Figure 7 and Table 2. PSNR, SSIM, and FSIM values are

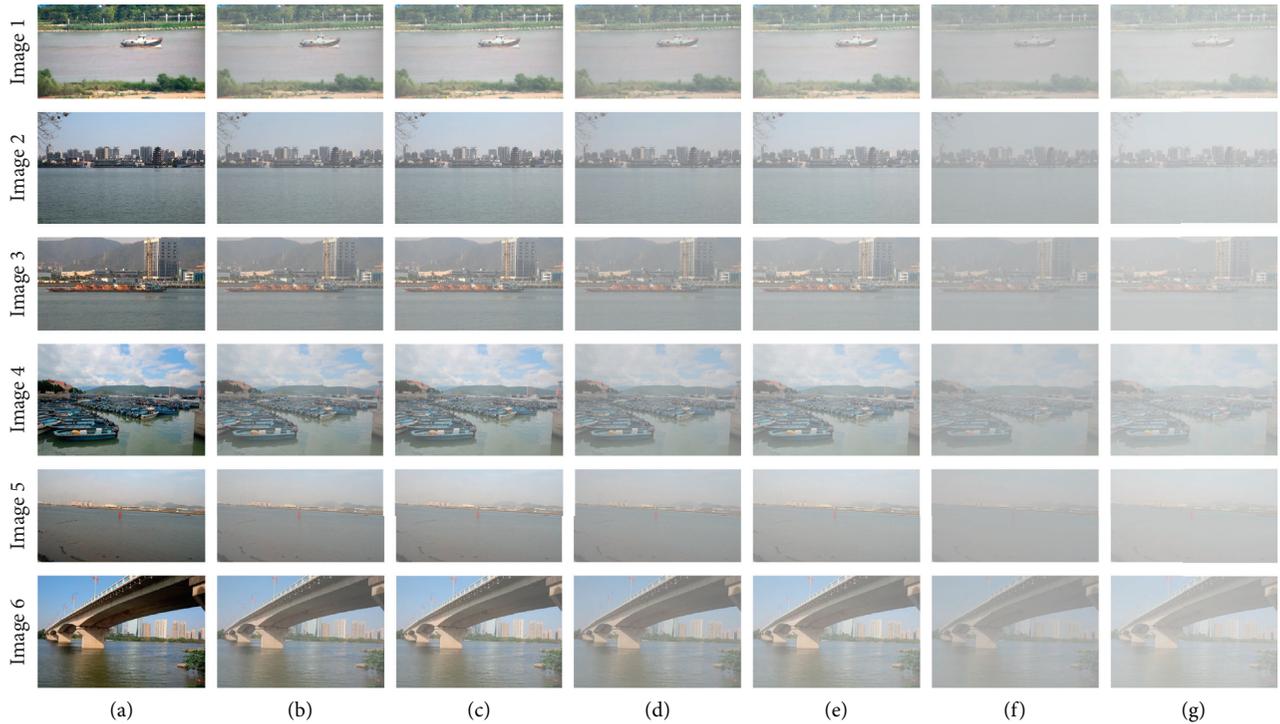


FIGURE 5: Six different original sharp maritime images selected to generate synthetic hazy images. From left to right: (a) original images; hazy images with (b) $A_{\infty} = 0.8/t = 0.6$, (c) $A_{\infty} = 0.9/t = 0.6$, (d) $A_{\infty} = 0.8/t = 0.4$, (e) $A_{\infty} = 0.9/t = 0.4$, (f) $A_{\infty} = 0.8/t = 0.2$, and (g) $A_{\infty} = 0.9/t = 0.2$, respectively.

illustrated using box-plot in Figure 7. It can be seen that our method has higher index values in most cases. Particularly for high-concentration hazy images, our method can stably obtain high-quality restored versions. Besides, Table 2 shows three metrics value comparisons of various image enhancement methods on 36 hazy images. In particular, we display the best result of each metric in bold. Due to the highest values of PSNR, SSIM, and FSIM, our method has the best dehazing performance. Meanwhile, the standard deviation calculated by the SSIM and FSIM is the smallest, which verifies that our method has excellent robustness.

4.4. Experiments on Realistic Maritime Datasets. This subsection will verify the reliability of several methods in realistic hazy maritime images due to the distinctness between synthetic and realistic versions. Meanwhile, NIQE is introduced to describe the naturalness of visual effects quantitatively, and our proposed method will be compared with seven dehazing methods, that is, DCP [14], GRM [16], HL [38], F-LDCP [39], MSCNN [25], AOD-Net [28], and GCA-Net [30]. Figure 8 shows several dehazing results to reflect the imaging performance more intuitively.

From the visual comparisons, DCP and HL have serious color distortion problems and blocking effects in the sky regions. Recovery results obtained by GRM have the risk of

low contrast, especially in the recovery task of Image 9. F-LDCP and AOD-Net fail to correct the color of the image. GCA-Net not only has the problem of overexposure in the sky region but also has nonuniform fog remaining in the image. Although MSCNN has better visual effects than other methods, our method has pleasing color and can remove fog more fully. Our superior performance can be further confirmed by the quantitative results NIQE shown in Table 3.

4.5. Experiments on Vessel Detection under Haze Environment. In the maritime imaging system, the harsh imaging environment severely restricts the regular operation of the visible light imaging sensor, reduces vessel detection accuracy, and leads to incorrect identification. To prove this phenomenon, we, respectively, used YOLOv4 [45] and Faster-RCNN [46] to detect vessels in haze and haze-free images. As shown in Figure 9, it is easily found that the haze image has low contrast and massive useful information is obscured, which leads to problems, for example, identification errors or missing identification during the target detection process. After dehazing, the vessel target is effectively captured and recognized, and the recognition accuracy is significantly increased. Therefore, dehazing the degraded hazy image by our method can improve vessel detection performance. The computer and the related workers can make correct decisions in time.

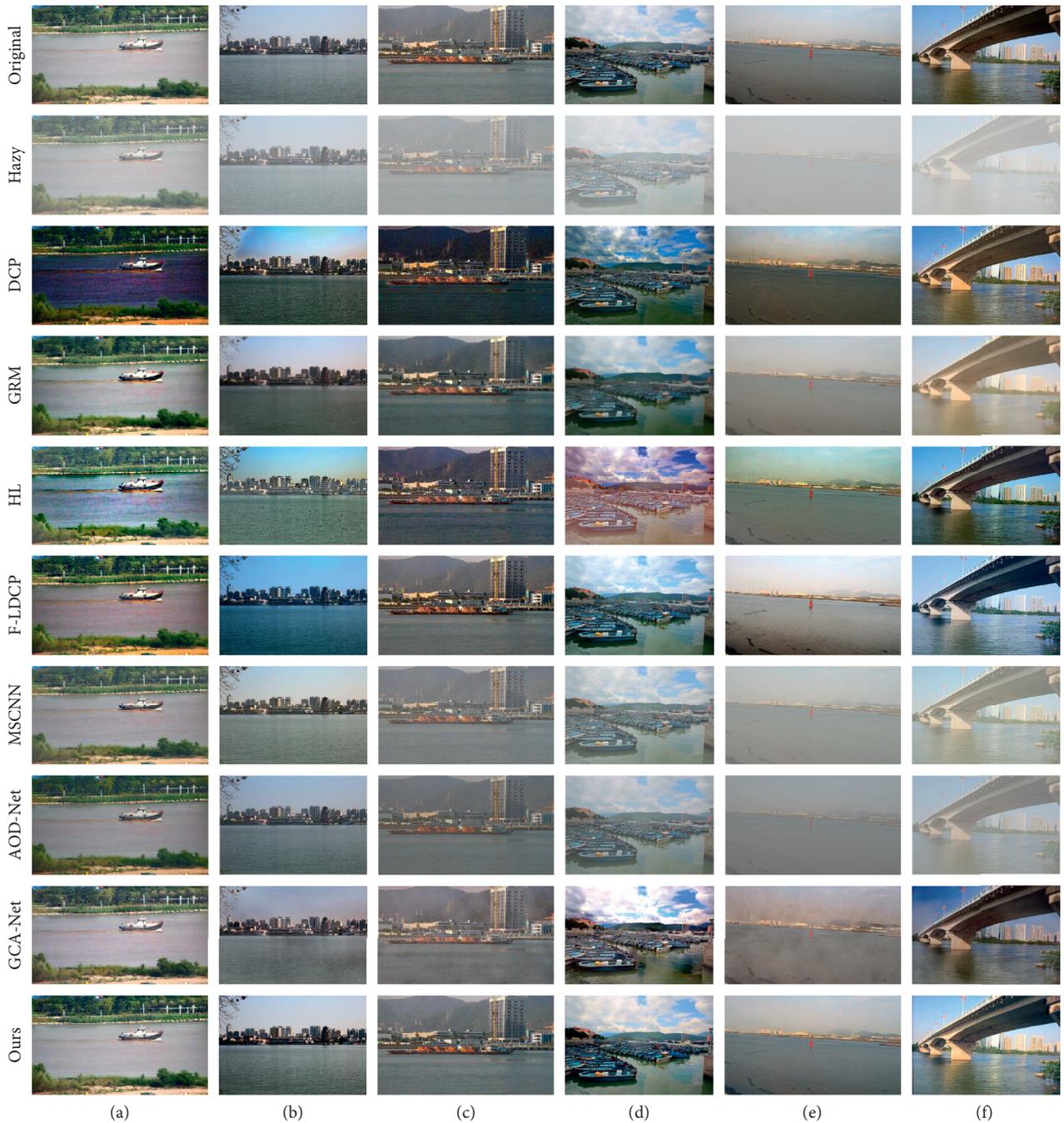
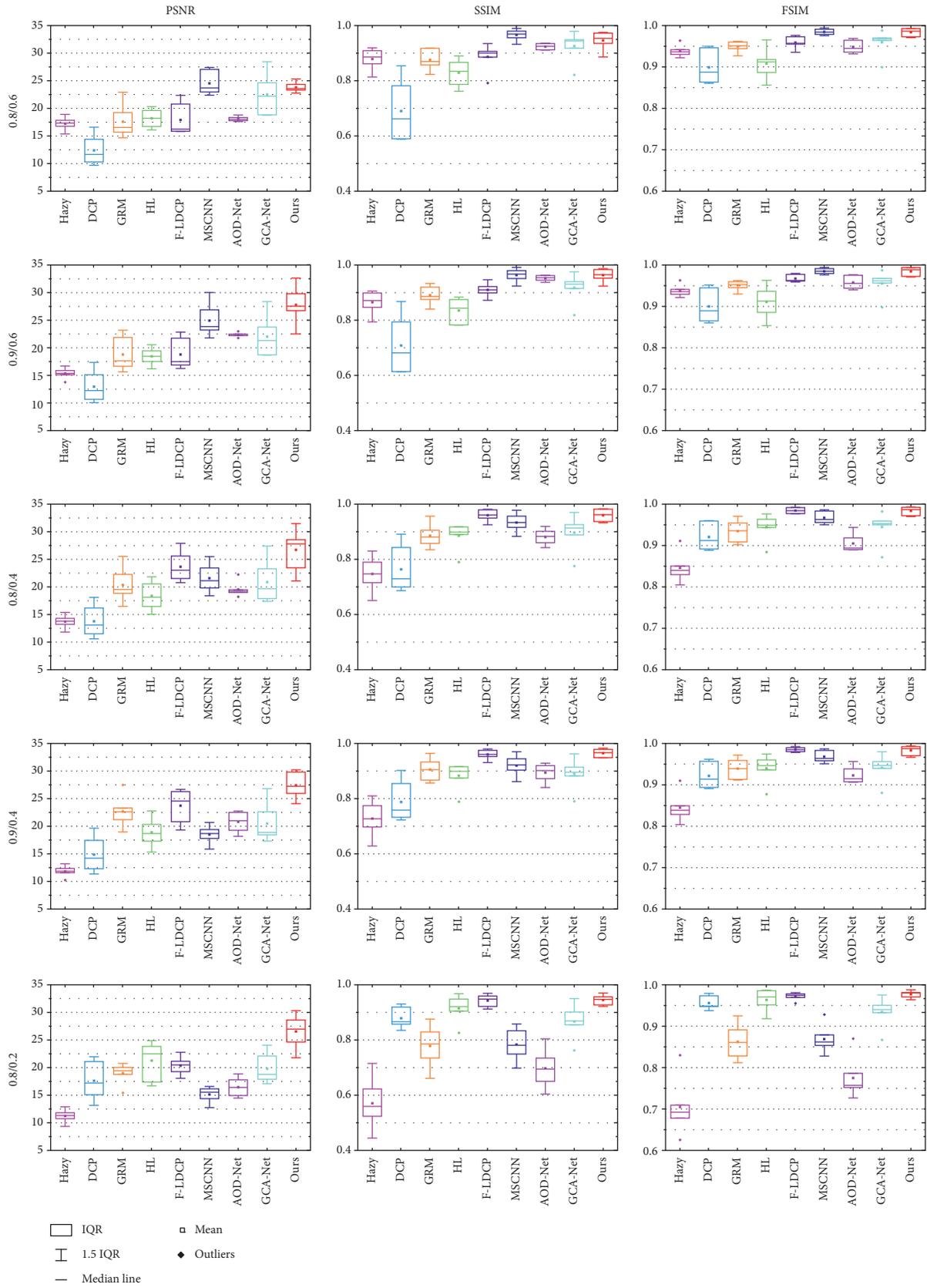


FIGURE 6: Comparisons of synthetic experiments on six typical hazy images in Figure 5. Each column represents the hazy image synthesized by different combinations of A_{∞} and t , the original image corresponding to the hazy image, and the dehazing results obtained by all methods. From left to right in the second line: (a) hazy Image 1 with $A_{\infty} = 0.8/t = 0.6$, (b) hazy Image 2 with $A_{\infty} = 0.9/t = 0.6$, (c) hazy Image 3 with $A_{\infty} = 0.8/t = 0.4$, (d) hazy Image 4 with $A_{\infty} = 0.9/t = 0.4$, (e) hazy Image 5 with $A_{\infty} = 0.8/t = 0.2$, and (f) hazy Image 6 with $A_{\infty} = 0.9/t = 0.2$, respectively.



(a)

FIGURE 7: Continued.

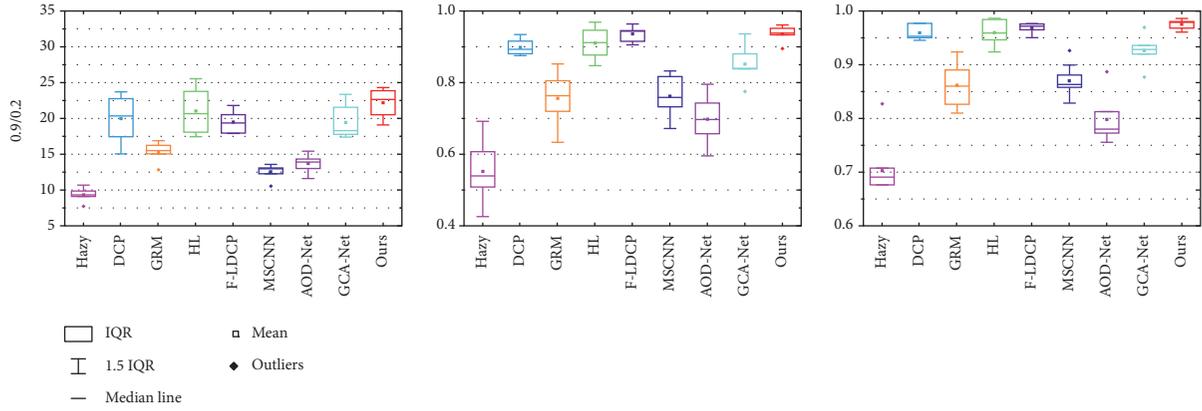


FIGURE 7: The box-plot of PSNR, SSIM, and FSIM values of various methods on all hazy images shown in Figure 5. From top to bottom: the synthetic hazy experiment with $A_{\infty} = 0.8/t = 0.6$, $A_{\infty} = 0.9/t = 0.6$, $A_{\infty} = 0.8/t = 0.4$, $A_{\infty} = 0.9/t = 0.4$, $A_{\infty} = 0.8/t = 0.2$, and $A_{\infty} = 0.9/t = 0.2$, respectively. Note that IQR represents interquartile range.

TABLE 2: PSNR, SSIM, and FSIM comparisons (mean \pm std) of various image enhancement methods on all test images shown in Figure 5. The best results are highlighted in bold.

Methods	PSNR	SSIM	FSIM
Hazy	13.17 \pm 2.90	0.727 \pm 0.138	0.830 \pm 0.105
DCP	15.23 \pm 2.93	0.789 \pm 0.086	0.933 \pm 0.022
GRM	18.96 \pm 3.21	0.845 \pm 0.070	0.912 \pm 0.046
HL	19.46 \pm 1.71	0.874 \pm 0.042	0.937 \pm 0.027
F-LDCP	20.58 \pm 3.00	0.933 \pm 0.032	0.972 \pm 0.011
MSCNN	19.71 \pm 5.19	0.890 \pm 0.091	0.940 \pm 0.055
AOD-net	18.70 \pm 3.30	0.848 \pm 0.113	0.889 \pm 0.079
GCA-net	21.26 \pm 1.38	0.899 \pm 0.027	0.947 \pm 0.012
Ours	25.80 \pm 3.01	0.954 \pm 0.014	0.982 \pm 0.003

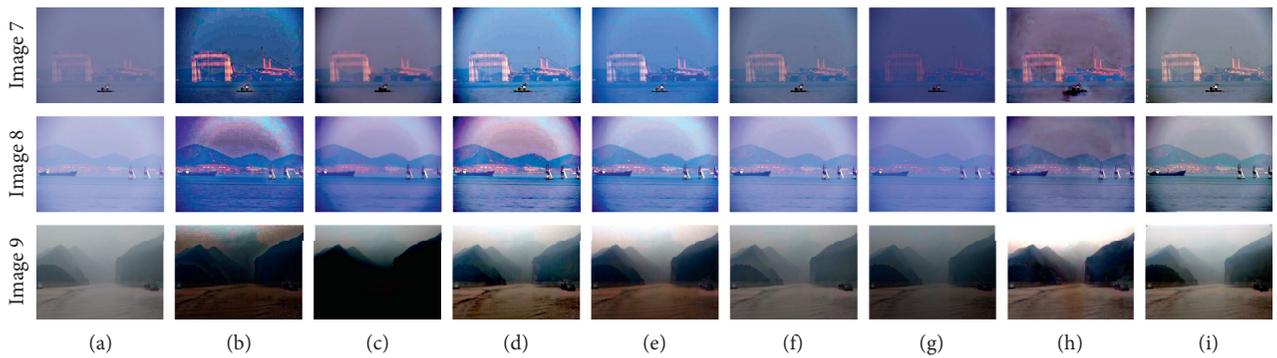


FIGURE 8: Comparisons of realistic experiments on three hazy maritime images. (a) Hazy, (b) DCP, (c) GRM, (d) HL, (e) F-LDCP, (f) MSCNN, (g) AOD-Net, (h) GCA-Net, and (i) ours.

TABLE 3: NIQE comparison of various image enhancement methods on all test images shown in Figure 6. The best results are highlighted in bold.

	DCP	GRM	HL	F-LDCP	MSCNN	AOD-Net	GCA-Net	Ours
Image 7	5.448	5.780	5.296	5.932	6.224	7.220	5.080	5.057
Image 8	5.525	10.297	4.961	5.868	5.826	5.716	5.694	5.122
Image 9	4.856	5.281	5.204	5.330	5.255	5.409	4.626	5.103
Average	5.276 ± 0.299	7.119 ± 2.299	5.154 ± 0.141	5.710 ± 0.270	5.768 ± 0.398	6.115 ± 0.792	5.133 ± 0.438	5.094 ± 0.027

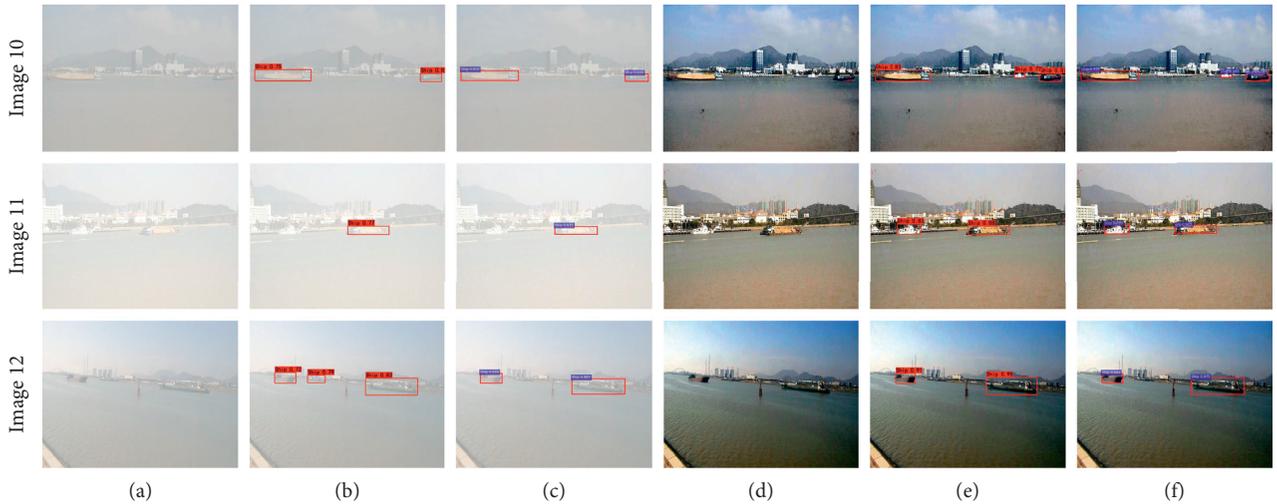


FIGURE 9: Vessel detection experiments under haze environment. (a) Hazy image, (b) hazy image (YOLOv4), (c) hazy image (Faster-RCNN), (d) dehazed image, (e) dehazed image (YOLOv4), and (f) dehazed image (Faster-RCNN).

5. Conclusion

In this paper, a novel CNN-enabled visibility dehazing framework was proposed, which could significantly improve the visual effect of images captured by the maritime camera under haze environment. In particular, this framework is composed of two subnetwork named Coarse Feature Extraction Module (C-FEM) and Fine Feature Fusion Module (F-FFM). C-FEM is an initial multiscale feature extraction network containing three simple six-layer convolutional networks, that is, Single C-FEM. C-FEM can obtain coarse feature maps from 1, 1/4, and 1/16 of the original image pixel size. F-FFM is a special encoder-decoder structure used to fuse and enhance the multiscale information obtained by C-FEM and original hazy image. To further improve the network performance, a corresponding loss function is proposed to simultaneously supervise the multiscale output of C-FEM and the final result of F-FFM. Furthermore, our dataset contains massive maritime images to complete the vessel detection task under haze environment successfully. Both qualitative and quantitative experiments have illustrated the effectiveness of our proposed framework.

Data Availability

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

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors' Contributions

Yuxu Lu and Yu Guo are co-first authors.

Acknowledgments

This work was supported by the National Key R&D Program of China (No. 2018YFC0309602).

References

- [1] R. W. Liu, J. Nie, S. Garg, Z. Xiong, Y. Zhang, and M. S. Hossain, "Data-driven trajectory quality improvement for promoting intelligent vessel traffic services in 6G-enabled maritime IoT systems," *IEEE Internet of Things Journal*, vol. 8, pp. 1–12, 2021.
- [2] X. Chen, L. Qi, Y. Yang et al., "Video-based detection infrastructure enhancement for automated ship recognition and behavior analysis," *Journal of Advanced Transportation*, vol. 2020, Article ID 7194342, 12 pages, 2020.
- [3] 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.

- [4] 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.
- [5] X. Chen, X. Xu, Y. Yang, Y. Huang, J. Chen, and Y. Yan, "Visual ship tracking via a hybrid kernelized correlation filter and anomaly cleansing framework," *Applied Ocean Research*, vol. 106, Article ID 102455, 2020.
- [6] E. D. Pisano, S. Zong, B. M. Hemminger et al., "Contrast limited adaptive histogram equalization image processing to improve the detection of simulated spiculations in dense mammograms," *Journal of Digital Imaging*, vol. 11, no. 4, pp. 193–200, 1998.
- [7] J. A. Stark, "Adaptive image contrast enhancement using generalizations of histogram equalization," *IEEE Transactions on Image Processing*, vol. 9, no. 5, pp. 889–896, 2000.
- [8] T. K. Kim, J. K. Paik, and B. S. Kang, "Contrast enhancement system using spatially adaptive histogram equalization with temporal filtering," *IEEE Transactions on Consumer Electronics*, vol. 44, no. 1, pp. 82–87, 1998.
- [9] D. J. Jobson, Z. Rahman, and G. A. Woodell, "Properties and performance of a center/surround retinex," *IEEE Transactions on Image Processing*, vol. 6, no. 3, pp. 451–462, 1997.
- [10] Z. Rahman, D. J. Jobson, and G. A. Woodell, "Multi-scale retinex for color image enhancement," in *Proceedings of the IEEE International Conference on Image Processing*, vol. 3, pp. 1003–1006, Lausanne, Switzerland, September 1996.
- [11] 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, 1998.
- [12] N. Hautiere, J. P. Tarel, and D. Aubert, "Towards fog-free in-vehicle vision systems through contrast restoration," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–8, Minneapolis, MN, USA, June 2007.
- [13] R. W. Liu, S. Xiong, and H. Wu, "A second-order variational framework for joint depth map estimation and image dehazing," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 1433–1437, Calgary, Canada, September 2018.
- [14] 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, 2011.
- [15] 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, Siem Reap, Cambodia, December 2018.
- [16] 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, Amsterdam, The Netherlands, October 2016.
- [17] Q. Liu, X. Gao, L. He, and W. Lu, "Single image dehazing with depth-aware non-local total variation regularization," *IEEE Transactions on Image Processing*, vol. 27, no. 10, pp. 5178–5191, 2018.
- [18] 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.
- [19] D. Berman and S. Avidan, "Non-local image dehazing," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1674–1682, Las Vegas, NV, USA, June 2016.
- [20] T. M. Bui and W. Kim, "Single image dehazing using color ellipsoid prior," *IEEE Transactions on Image Processing*, vol. 27, no. 2, pp. 999–1009, 2018.
- [21] Q. Shu, C. Wu, Z. Xiao, and R. W. Liu, "Variational regularized transmission refinement for image dehazing," in *Proceedings of the IEEE International Conference on Image Processing*, pp. 2781–2785, Taipei, Taiwan, August 2019.
- [22] 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.
- [23] 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, GA, USA, September 2014.
- [24] 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.
- [25] 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, Cham, Amsterdam, October 2016.
- [26] N. Silberman, D. Hoiem, P. Kohli, and R. Fergus, "Indoor segmentation and support inference from RGBD images," in *Proceedings of the European Conference on Computer Vision*, pp. 746–760, Berlin, Germany, October 2012.
- [27] X. Zhao, K. Wang, Y. Li, and J. Li, "Deep fully convolutional regression networks for single image haze removal," in *Proceedings of the IEEE Visual Communications and Image Processing*, pp. 1–4, St. Petersburg, FL, USA, December 2017.
- [28] B. Li, X. Peng, Z. Wang, J. Xu, and D. Feng, "AOD-net: all-in-one dehazing network," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 4770–4778, Venice, Italy, October 2017.
- [29] 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, Salt Lake City, UT, USA, June 2018.
- [30] 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, Waikoloa Village, HI, USA, January 2019.
- [31] 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, Hilton Head Island, SC, USA, June 2000.
- [32] Y. Pang, M. Sun, X. Jiang, and X. Li, "Convolution in convolution for network in network," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 5, pp. 1587–1597, 2018.
- [33] X. Huang and S. Belongie, "Arbitrary style transfer in real-time with adaptive instance normalization," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1510–1519, Venice, Italy, October 2017.
- [34] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, 2017.
- [35] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multi-scale structural similarity for image quality assessment," in

- Proceedings of the Asilomar Conference on Signals, Systems and Computers*, pp. 1398–1402, Pacific Grove, CA, USA, November 2003.
- [36] K. Jiang, Z. Wang, P. Yi et al., “Multi-scale progressive fusion network for single image deraining,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 8346–8355, Seattle, WA, USA, June 2020.
- [37] L. I. Rudin, S. Osher, and E. Fatemi, “Nonlinear total variation based noise removal algorithms,” *Physica D Nonlinear Phenomena*, vol. 60, no. 1–4, pp. 259–268, 1992.
- [38] D. Berman, T. Treibitz, and S. Avidan, “Air-light estimation using haze-lines,” in *Proceedings of the IEEE International Conference on Computer Photography*, pp. 1–9, Stanford, CA, USA, May 2017.
- [39] Y. Zhu, G. Tang, X. Zhang, J. Jiang, and Q. Tian, “Haze removal method for natural restoration of images with sky,” *Neurocomputing*, vol. 275, no. 31, pp. 499–510, 2017.
- [40] Z. Wang and A. C. Bovik, “Mean squared error: love it or leave it?,” *IEEE Signal Processing Magazine*, vol. 26, no. 1, pp. 98–117, 2009.
- [41] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image quality assessment: from error visibility to structural similarity,” *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.
- [42] L. Zhang, L. Zhang, X. Mou, and D. Zhang, “FSIM: a feature similarity index for image quality assessment,” *IEEE Transactions on Image Processing: A Publication of the IEEE Signal Processing Society*, vol. 20, no. 8, pp. 2378–2386, 2011.
- [43] A. Mittal, R. Soundararajan, and A. C. Bovik, “Making a “completely blind” image quality analyzer,” *IEEE Signal Processing Letters*, vol. 20, no. 3, pp. 209–212, 2013.
- [44] P. Adam, G. Sam, and C. Soumith, “Automatic differentiation in PyTorch,” in *Proceedings of the Conference and Workshop on Neural Information Processing Systems*, pp. 1–4, Long Beach, CA, USA, December 2017.
- [45] A. Bochkovskiy, C. Y. Wang, and H. Y. M. Liao, “YOLOv4: optimal speed and accuracy of object detection,” 2020, <https://arxiv.org/abs/2004.10934>.
- [46] S. Ren, K. He, R. Girshick, and J. Sun, “Faster R-CNN: towards real-time object detection with region proposal networks,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, 2016.