Spotlighting on Objects: Prior Knowledge-Driven Maritime Image Dehazing and Object Detection Framework

Yaozong Mo, Chaofeng Li, *Senior Member, IEEE*, Wenqi Ren, *Member, IEEE*, and Wenwu Wang, *Senior Member, IEEE*

***Abstract*—Maritime environments often face visibility challenges due to haze which significantly impacts detection models. However, existing maritime object detection algorithms often neglect haze conditions or the unique characteristics of the maritime environment, resulting in decreased effectiveness in hazy weather. In this article, we propose a prior knowledge-driven maritime image dehazing and object detection framework (MDD), which consists of a detection network and a restoration network. Leveraging the characteristics of the highlighted ships in the inverted dark channel prior (IDCP), the detection network incorporates a prior subnetwork to learn ship-related features, which are subsequently merged into the backbone network through an IDCP cross-attention module. During training, the restoration network is integrated to improve the clarity of the features learned by the detection network. In addition, a ship-haze enrichment strategy is implemented to emphasize ship regions in the training samples, along with a ship-aware reconstruction loss to enhance the network's ability to learn dehazed features. Moreover, we establish a maritime object recognition with haze levels (MORHL) data set to evaluate object detector performance in maritime hazy conditions. It includes 13 280 annotated images across six categories: cargo ship, container ship, fishing boat, passenger ship, island, and buoy, with haze levels categorized as light, medium, and heavy. Comprehensive experiments on the MORHL and SMD data sets demonstrate that the proposed MDD framework outperforms the state-of-the-art detectors and various combinations of dehazing and detection methods.**

***Index Terms*—Image dehazing, inverted dark channel prior (IDCP), maritime object detection, maritime object recognition with haze levels (MORHL) data set.**

I. Introduction

HAZY conditions prevalent in maritime environments often result in restricted visibility and diminished contrast in images. This deterioration in image quality poses significant challenges to the accurate detection and recognition of objects [1], including ships, buoys, and other maritime entities. Such compromised visibility elevates the risk of maritime incidents, such as collisions and grounding incidents. Accordingly, there is a compelling demand to explore image dehazing algorithms and object detection algorithms tailored to vision-based applications in maritime scenes.

Object detection algorithms currently include single-stage detectors, such as the YOLO series [2], [3], which divide images into grids and predict object classes and bounding boxes directly, and two-stage detectors, such as Faster R-convolutional neural network (R-CNN) [4], which uses a region proposal network (RPN) to generate candidate boxes before classifying and refining them. Other notable models are single shot multibox detector [5], RetinaNet [6], and Mask R-CNN [7], each with its strengths and applicability in various scenarios, such as real-time performance (YOLO), high accuracy (Faster R-CNN), and instance segmentation capabilities (Mask R-CNN). Recently, several improved networks have been explored to address the challenge of ship detection, such as sequential three-way decisions and multi-granularity detector (S3MDet) [8], cross-level feature refinement detector (CLFR-Det) [9], and rotational YOLO based model (RYM) [10]. However, these networks have not been optimized for hazy conditions, making it difficult to accurately identify and localize objects in hazy weather. In contrast, our approach integrates prior knowledge into the detection network, enabling the model to more effectively distinguish ships from the hazy background, thus improving both detection and localization under severe weather conditions.

Although there exist numerous efficient image dehazing networks, such as MSBDN [11] and FFANet [12], they are primarily optimized for terrestrial scenarios. Consequently, these dehazing methods do not effectively address the unique characteristics of maritime environments. Overall, maritime image dehazing remains relatively unexplored in the current literature. In maritime environments, images often feature extensive sky and sea areas. While ships are relatively small within these images, they are central to downstream tasks, such as detection [13] and tracking [14], [15]. Building upon these observations, we propose a prior knowledge-driven maritime image dehazing and object detection framework (MDD) which consists of a detection

Received 13 October 2024; revised 30 December 2024 and 29 January 2025; accepted 30 January 2025. Date of publication 21 April 2025; date of current version 16 July 2025. This work was supported by the National Natural Science Foundation of China under Grant 62176150. *(Corresponding author: Chaofeng Li.)*

Associate Editor: P. Ren.

Yaozong Mo and Chaofeng Li are with the Institute of Logistics Science and Engineering, Shanghai Maritime University, Shanghai 201306, China (e-mail: moyaozong@outlook.com; wxlichaofeng@126.com).

Wenqi Ren is with the School of Cyber Science and Technology, Sun Yat-sen University, Shenzhen 518000, China (e-mail: renwq3@mail.sysu.edu.cn).

Wenwu Wang is with the Center for Vision Speech and Signal Processing, Department of Electrical and Electronic Engineering, University of Surrey, GU2 7XH Surrey, U.K. (e-mail: w.wang@surrey.ac.uk).

Digital Object Identifier 10.1109/JOE.2025.3545289

network and a restoration network for effective maritime object detection under hazy conditions.

Specifically, we develop an inverted dark channel prior (IDCP) that could accentuate ships within maritime images. In dark channels [16], the sky and sea surface are typically brighter than ships due to their pixel values being similar to air light. Utilizing this characteristic, ships are highlighted when the dark channel is inverted. We further utilize min–max normalization to magnify this effect. This procedure preserves the darker areas of the sky and sea surface while enhancing the brightness of objects on the sea.

In light of the characteristics of the highlighted ships in the IDCP, we establish a prior subnetwork within the detection network. The backbone network processes the hazy image to extract contextual and texture features, while the prior subnetwork handles the corresponding inverted dark channel to learn ship-related features. To effectively integrate structural and prior knowledge features, we propose an IDCP cross-attention module that facilitates multiscale learning and information fusion.

During training, we incorporate an auxiliary restoration network to improve the clarity of the features learned by the detection network. The restoration network reconstructs the extracted features into haze-free images and is optimized by image reconstruction loss functions. By integrating the restoration network, the detection network is effectively guided to produce dehazed features, thereby improving its capability to accurately detect objects under hazy conditions. This approach ensures that the detection network not only focuses on object identification but also learns to produce features that are robust against the distortions caused by haze.

Furthermore, we implement a ship-haze enrichment strategy (SHES) during training to strengthen the network's capacity for extracting dehazed features. This strategy introduces thick haze specifically to ship areas, thereby directing the model's attention towards effectively removing haze from these regions. In addition, we design a ship-aware reconstruction loss function to prioritize ship-related features while minimizing the influence of features from the sky and sea surface. This loss function aims to promote the network in both dehazing and detection tasks by effectively preserving ship-related information in the dehazed images.

Maritime environments often face visibility challenges due to haze, which can obscure objects and substantially hinder the effectiveness of object detection models. To assess the performance improvement of object detection networks before and after image dehazing in maritime scenarios, we propose a maritime object recognition with haze levels (MORHL) data set. This data set comprises 13 280 images divided into training and test sets, encompassing six categories: cargo ship, container ship, fishing boat, passenger ship, island, and buoy, with light, medium, and heavy haze levels.

In brief, this work presents the following contributions.

1) A prior knowledge-driven MDD is developed. The MDD utilizes a prior subnetwork to learn ship-related features from the IDCP, which are then integrated into the backbone network via an IDCP cross-attention module. Besides, an additional restoration network is incorporated to enable the detection network to extract dehazed features, thus facilitating efficient object detection in maritime haze environments.
2) By leveraging the IDCP, we implement a SHES to introduce additional thick haze to ship regions in the training samples. This approach is complemented by a ship-aware reconstruction loss that enhances the network's capacity to learn dehazed features effectively.
3) We construct a MORHL data set to assess the performance improvement of object detectors before and after image dehazing in maritime scenarios. This data set comprises 13 280 annotated images across six categories: cargo ship, container ship, fishing boat, passenger ship, island, and buoy, encompassing light, medium, and heavy haze conditions.

The rest of this article is organized as follows. Section II presents an overview of relevant studies, encompassing image dehazing, object detection, and detection with restoration algorithms. Section III details the proposed MDD framework. Section IV presents a thorough analysis of experimental results, benchmarking our approach against state-of-the-art methods. Finally, Section V concludes this article.

II. Related Work

This section offers a brief summary of single image dehazing algorithms, object detection approaches, and detection with restoration methods.

A. Image Dehazing

Maritime image analysis [17] and enhancement techniques [18], [19] play a crucial role in supporting visual perception tasks in maritime scenes. These advanced methods not only enhance image interpretability but also facilitate robust feature extraction in complex maritime environments. Among these methods, image dehazing algorithms can significantly improve the clarity of maritime images under hazy weather conditions. Based on the atmospheric scattering model (ASM) [20], a hazy image can be expressed as follows:

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

Here, x denotes the pixel position, $I(x)$ refers to degraded hazy image, and $J(x)$ is haze-free image. A and $t(x)$ are the air light and the transmission. There are two main types of dehazing approaches: prior-based and learning-based. The prior-based image dehazing approache utilize statistical properties of natural images to predict transmission and air light. He et al. [16] suggested that a minimum of one channel in the RGB space tends to approach zero in a fog-free image. While these approaches demonstrate efficacy in image dehazing, they might struggle in specific scenarios. For example, the dark channel prior (DCP) [16] often encounters challenges with the sky or white buildings.

As CNNs have advanced, many researchers have developed learning-based dehazing algorithms. Cai et al. [21] exploited a network for estimating medium transmission in hazy images, featuring a specialized CNN architecture with Maxout units for

feature extraction and a novel activation function, BReLU, to enhance image quality. Ren et al. [22] introduced a multiscale neural network, which models the connection between hazy input and their transmissions.

Although supervised learning methods perform well on synthetic images, they may struggle with real-world ones due to domain shifts. Hence, Li et al. [23] devised a semi-supervised learning approach utilizing a deep CNN with distinct supervised and unsupervised branches. These branches employ varied loss functions to optimize the network across synthetic and real-world data sets. Shao et al. [24] explored a domain adaptation approach employing image translation and two dehazing modules. Initially, a bidirectional translation subnetwork aligns synthetic and real domains. Subsequently, two dehazing subnetworks are trained using translated images under a consistency constraint, enhancing domain adaptability with real hazy images.

Recently, Zheng et al. [25] introduced a variant of CycleGAN for overwater image dehazing. However, it only replaces land hazy images with overwater ones during training, leading to limited dehazing results. In contrast, our approach focuses on maritime image features, ensuring dehazing is suitable for detection tasks.

B. Object Detection

Object detection is crucial in scene recognition and modeling [26], [27], identifying targets like people, vehicles, and objects in complex scenes. Single-stage methods concurrently generate region proposals and object classification through unified encoding. YOLO [28] partitions input images into grid cells, predicting multiple bounding boxes relative to each cell's center. Nevertheless, it demonstrates reduced detection accuracy for small or occluded objects, constrained by the maximum number of objects per cell. In subsequent versions [2], [3], a sequence of enhancements has been introduced to tackle these concerns.

In two-stage methods, the process begins with proposal selection, followed by the classification and regression of region proposals. For instance, Faster R-CNN [4] introduces a RPN within the network architecture, automatically generating proposals through the RPN. Multistage methods take the two-stage approach further by repeating the steps multiple times. For example, Cascade R-CNN [29] iterates on proposal generation, classification, and regression, with the goal of progressively refining the region proposals.

The incorporation of anchors has demonstrated efficacy in improving object detectors. Nonetheless, achieving optimal coverage of ground truth labels needs a significant amount of meticulously configured anchors [30]. To address this challenge, anchor-free methods have been proposed as alternatives. CornerNet [31] detects objects by predicting their keypoints (corners) and forming bounding boxes based on these key points. CenterNet [32] directly estimates the central point of objects and regresses the bounding box size around that center. It achieves accurate detection by focusing on the center keypoint rather than traditional anchor-based methods. Taking advantage of the potent feature extraction capabilities of Transformer, Vision Transformer [33] integrates the Transformer architecture into computer vision tasks. Extending this, DEtection Transformer (DETR) [34] devises an encoder–decoder framework tailored for object detection. However, DETR involves high computational cost, which may restrict its practical applicability in computation resource limited applications. To this end, RT-DETR [35] is proposed to achieve a tradeoff in both speed and accuracy.

In summary, single-stage methods generally offer faster performance in object detection, but they tend to have slightly weaker performance compared to detectors based on region proposals.

C. Detection With Restoration

In degraded environments, it is generally recognized that restoration techniques can enhance images, leading to better performance in recognition tasks. Yim and Sohn [36] investigated how image quality affects convolutional networks in image classification and proposes a dual-channel architecture to address these issues. Dai et al. [37] presented a comprehensive study on the utility of image super-resolution (ISR) for various vision tasks beyond perceptual evaluation, evaluating six ISR methods across four vision applications.

Some restoration and detection networks are trained together on degraded images. Liu et al. [38] explored a unified framework that combines image denoising with high-level vision tasks, showcasing the benefits of semantic guidance for improved visual quality and robustness across various vision tasks without fine-tuning. Huang et al. [39] developed DSNet, a novel dual-subnet network for object detection in foggy conditions. DSNet attains significant performance improvements by jointly optimizing these tasks, demonstrating superior accuracy while maintaining high speed. Different from above, Liu et al. [40] optimized the entire network with only a detection loss. Although the approaches mentioned above yield promising results for high-level tasks, the domain of maritime image detection coupled with restoration remains relatively underexplored. The objective of our MDD framework presented in this article is to address this gap for achieving efficient object detection performance in maritime hazy scenes.

III. Proposed Method

This section introduces our prior knowledge-driven MDD framework. Initially, we provide an outline of our approach. Following this, we elaborate on the detection network and restoration network. Then, we delve into the specifics of the prior knowledge. Subsequently, we discuss the details of the loss functions utilized for network training. Finally, we provide an explanation of the proposed MORHL data set.

A. Method Overview

The proposed MDD framework is presented in Fig. 1. Initially, given a haze-free maritime image, we compute its inverted dark channel to highlight ship regions. Utilizing both the clean image and its corresponding inverted dark channel, we then apply the SHES to generate a hazy image, followed by calculating the inverted dark channel for the hazy image. Subsequently, the hazy

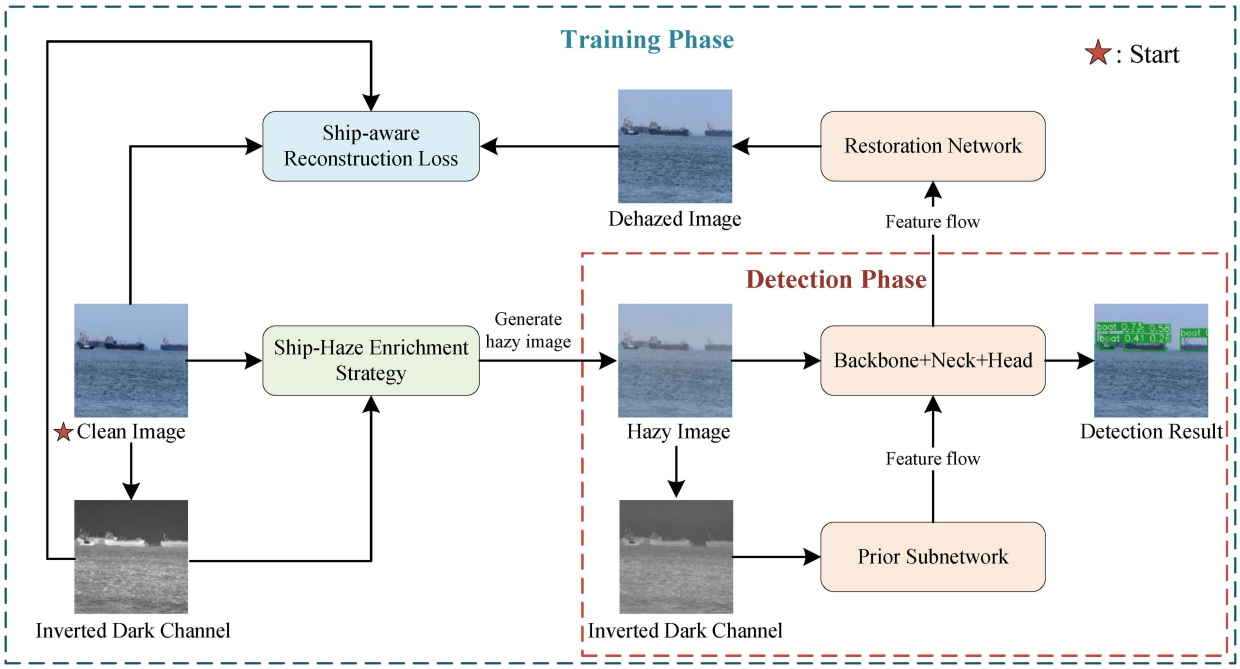

Fig. 1. Flowchart of the proposed prior knowledge-driven MDD.

image is processed by the backbone network to extract contextual features, while the prior subnetwork handles the prior image to extract ship-related features. Finally, these features are integrated using a cross-attention module for object detection.

During the training phase, we also develop an auxiliary restoration network which is designed to improve the clarity of the features extracted by the detection network. This restoration network utilizes features from the backbone network to reconstruct haze-free images. By leveraging the ship-aware reconstruction loss, we improve the detection network's capacity to produce clearer, dehazed features, thereby improving its performance in detecting objects under hazy conditions. It is crucial to emphasize that the restoration network is utilized solely during the training stage.

B. Detection Network

1) Network Architecture: The detection network, as depicted in Fig. 2, integrates an enhanced YOLOv8 [41] architecture with a prior subnetwork. Specifically, the YOLOv8 uses the CSPDarknet53 backbone and features the path aggregation network for its neck. The head generates predictions by computing bounding box coordinates, object scores, and class probabilities. To integrate ship-related features from the prior subnetwork with the backbone network, we designed IDCP cross-attention modules within the backbone network. This design enhances multiscale learning and facilitates effective information fusion within the backbone network.

The prior subnetwork contains five downsampling stages, each containing a 3×3 convolution with a stride of 2, followed by a depthwise separable convolution (DSC) [42]. This integration of DSC enables the network to be lightweight while maintaining its representational capacity. In the final three downsampling stages, the output features from each depth-separable convolution are concurrently fed into the cross-attention module within the backbone network, thereby supplying ship-related features derived from the prior map.

2) IDCP Cross-Attention: To improve the model's capability to capture and utilize contextual information from both input images and the inverted dark channel, we design a collaborative fusion mechanism called IDCP cross-attention. Fig. 2 illustrates the basic idea behind our innovative IDCP cross-attention module. The fusion process integrates the structural information token from one branch with the prior information tokens from another branch. In this process, the structural information token within its respective branch has already acquired abstract information from the input image, interacting with the prior tokens at the other branch helps to learn information at a different scale.

Specifically, the structural and prior features from the previous layer $u \in \mathbb{R}^{C\times H\times W}$, $v \in \mathbb{R}^{C\times H\times W}$ are first transformed into a query ($\boldsymbol{Q}$) and key vector ($\boldsymbol{K}$)

$$\boldsymbol{Q} = \sigma(\mathrm{Avg}(\mathrm{Conv}_{1\times 1}(u, w))) \tag{2}$$

$$\boldsymbol{K} = \sigma(\mathrm{Avg}(\mathrm{Conv}_{1\times 1}(v, w))) \tag{3}$$

where $\mathrm{Conv}_{1\times 1}$ denotes convolution operation with kernel size 1, w is convolution kernel, Avg and σ represent average pooling and ReLU activation.

After obtaining the query and key vector, we calculate the attention vector and applied to the value ($\boldsymbol{V}$)

$$\boldsymbol{V} = \mathrm{DSC}(u, w) \tag{4}$$

$$O = \boldsymbol{V} + \boldsymbol{V} \times \mathrm{Softmax}(\mathrm{Conv}_{1\times 1}(\mathrm{cat}(\boldsymbol{Q}, \boldsymbol{K}))) \tag{5}$$

where DSC denotes DSC, cat denotes concatenation operation, and O is the output feature.

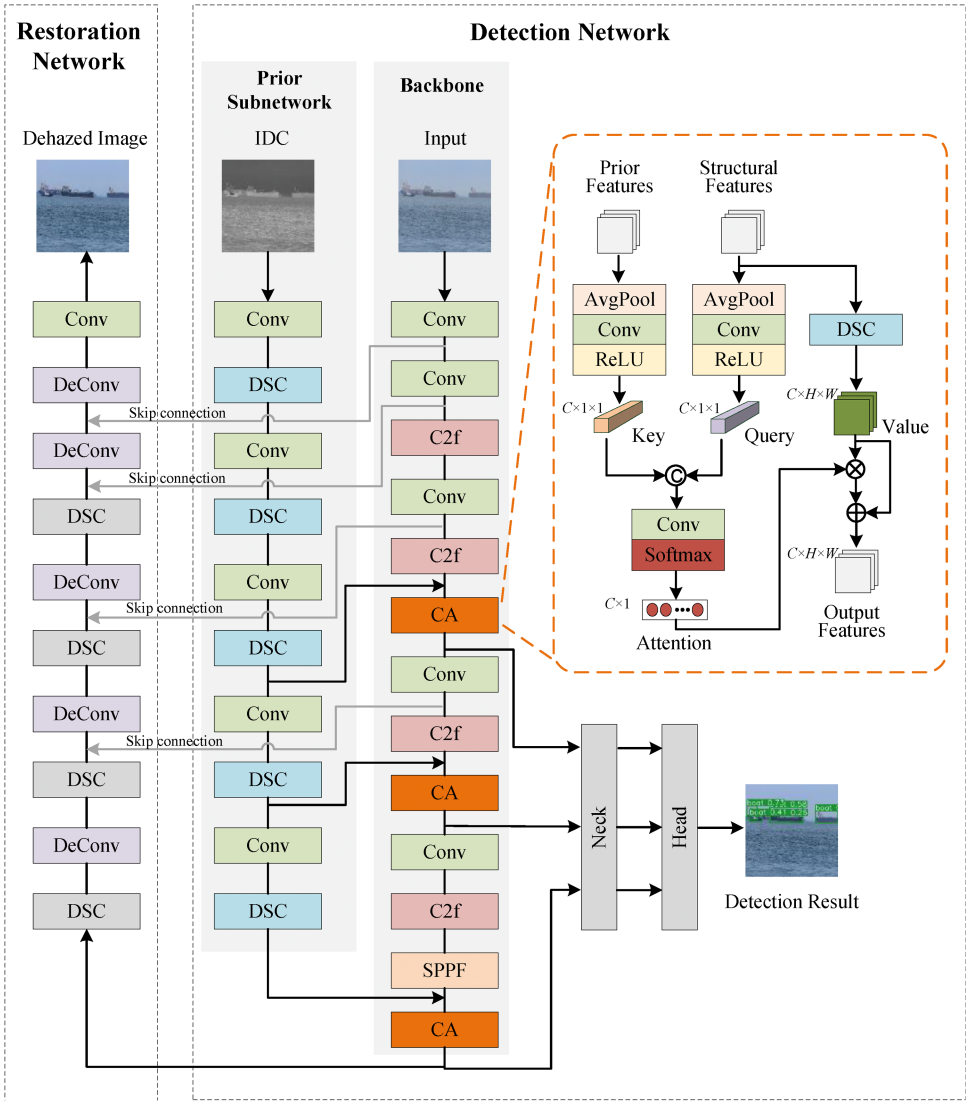

Fig. 2. Architecture of the proposed MDD framework, where "DSC" denotes depthwise separable convolution and "CA" represents the IDCP cross-attention module.

C. Restoration Network

In our MDD framework, the restoration network focuses on generating haze-free images and facilitating the backbone network to learn dehazed features, thereby improving detection accuracy under hazy conditions. To achieve this, we have implemented a decoder-like network that incorporates five upsampling stages, as shown in Fig. 2. The first four upsampling stages consist of a deconvolutional layer followed by a DSC. The final upsampling stage incorporates an additional 7×7 convolution layer, which plays a critical role in generating the final dehazed image.

It is crucial to emphasize that the restoration network is utilized solely during the training stage. Its role is to assist the backbone network in acquiring effective representations for dehazing. Once the model has been trained, the restoration network is not required for inference, as the backbone network is capable of leveraging the learned dehazed features for object detection in real-world scenarios. This design ensures that the model can perform efficiently and effectively during deployment.

D. Prior Knowledge

1) Inverted Dark Channel Prior: Marine scenes often feature expansive sky and sea areas. When objects in the maritime scene closely resemble atmospheric light, as He et al. [16] claimed, the DCP loses its effectiveness, leading to brighter sky and sea regions in the dark channel. However, this brightness contrast can be leveraged to highlight objects (such as ships and islands) by inverting the dark channel in maritime scenes. As depicted in Fig. 3, the inverted dark channel can effectively emphasize the contrast between the background and the objects.

2) Ship-Haze Enrichment Strategy: Haze that obscures ships considerably hinders model performance in object detection and tracking tasks. To simulate this effect, we introduce thick haze into the ship regions of the training images, named SHES. If the network fails to effectively restore and detect the ships, it results in higher image reconstruction and detection loss values. Consequently, this strategy drives the model to prioritize effective dehazing and accurate detection of these critical regions.

Fig. 3. Comparison of the (a) input image, (b) dark channel, and (c) inverted dark channel. To illustrate the distinctions more effectively, we utilize heatmaps for visualization.

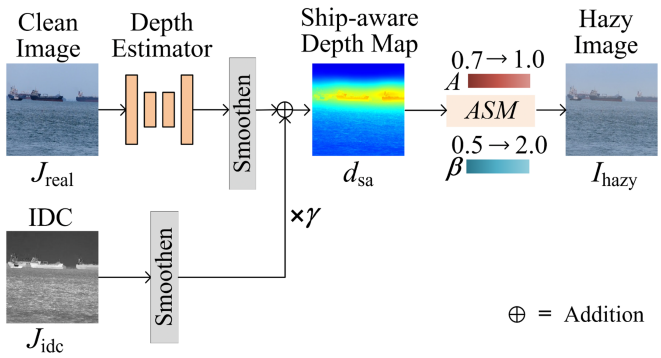

Fig. 4. Flowchart of the SHES.

Equation (1) illustrates the imaging model of hazy degraded images, where $t(x) = e^{-\beta d(x)}$ depends on scene depth $d(x)$ and haze density coefficient β. Regions with a greater depth of field will result in more severe haze. Since IDCP can effectively highlight ships, we utilize it to deepen the ship area and generate the hazy image, as illustrated in Fig. 4. We overlay the inverted dark channel J_{idc} onto the depth map to enhance the depth of the ship region. Because of the greater depth of field, the haze in the ship area will be more pronounced when generating hazy image using ASM. Here, γ denotes a weight parameter, fixed at 0.2 in our experimental configuration. In addition, the values of A and β vary within the ranges of 0.7 to 1.0 and 0.5 to 2.0, respectively, to generate distinct haze concentrations.

3) Ship-Aware Reconstruction Loss: To sharpen the network's attention on ship areas in maritime images, we introduce a novel loss function named the "ship-aware reconstruction loss." It strengthens the network's ability to accurately recover and emphasize ship-related features during image dehazing. This loss function is expressed as

$$L_{\text{sa}} = ||(G(I(x)) - J(x))J_{\text{sidc}}||_1 \tag{6}$$

where $I(x)$ and $J(x)$ represent for input hazy image and recovered clean image, G denotes dehazing net, J_{sidc} is smoothed inverted dark channel of $J(x)$, and $||\cdot||_1$ stands for L_1 norm. Fig. 5 illustrates the calculation process of ship-aware recovery loss. In L_{sa}, the emphasis lies in augmenting the loss associated

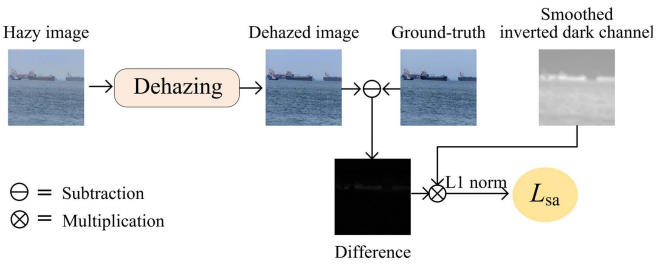

Fig. 5. Flowchart of the ship-aware recovery loss calculation.

with ship-related features while concurrently reducing the loss attributed to the sky and sea surface. As a result, the ship-aware recovery loss aims to improve the model's effectiveness in both dehazing and detection by effectively addressing and preserving the ship-related information in the dehazed images.

E. Loss Functions

The MDD loss function comprises three components: the detection loss L_{det}, the generative adversarial network (GAN) loss L_G, and the ship-aware recovery loss L_{sa} which is detailed earlier.

1) GAN Loss: The GAN loss is utilized to update both the generators and discriminators in an adversarial way. It is defined as

$$L_G = \mathbb{E}[(D(J(x)) - 1)^2] + \mathbb{E}[D(G(I(x)))^2] \tag{7}$$

where $J(x)$ is the clean image, $I(x)$ is the hazy image, G denotes the generator, D denotes the discriminator which distinguishes between translated images $G(I(x))$ and real images $J(x)$. In the MDD framework, the backbone and the restoration network collaboratively play the role of the generator to produce haze-free images.

2) Detection Loss: Detection loss comprises multiple components that work collaboratively to boost the accuracy of object localization and classification., which can be expressed as

$$L_{\text{det}} = L_{\text{conf}} + L_{\text{reg}} + L_{\text{cls}}. \tag{8}$$

The confidence loss L_{conf} measures how accurately the model predicts the presence of an object within a bounding box. The regression loss L_{reg} minimizes location errors between predicted and actual boxes, and the classification loss L_{cls} checks for discrepancies in class probabilities and labels.

Specifically, the smooth l_1 loss [43] is employed as the regression loss

$$L_{\text{smooth }L1}(x) = \begin{cases} 0.5x^2, & \text{if } |x| < 1 \\ |x| - 0.5, & \text{otherwise.} \end{cases} \tag{9}$$

The term x denotes the distance between the predicted and ground truth boxes, with the softmax loss serving as the confidence metric

$$L_{\text{conf}} = -\sum_{i=1}^{c} \text{pre}_i \log(\text{gt}_i) \tag{10}$$

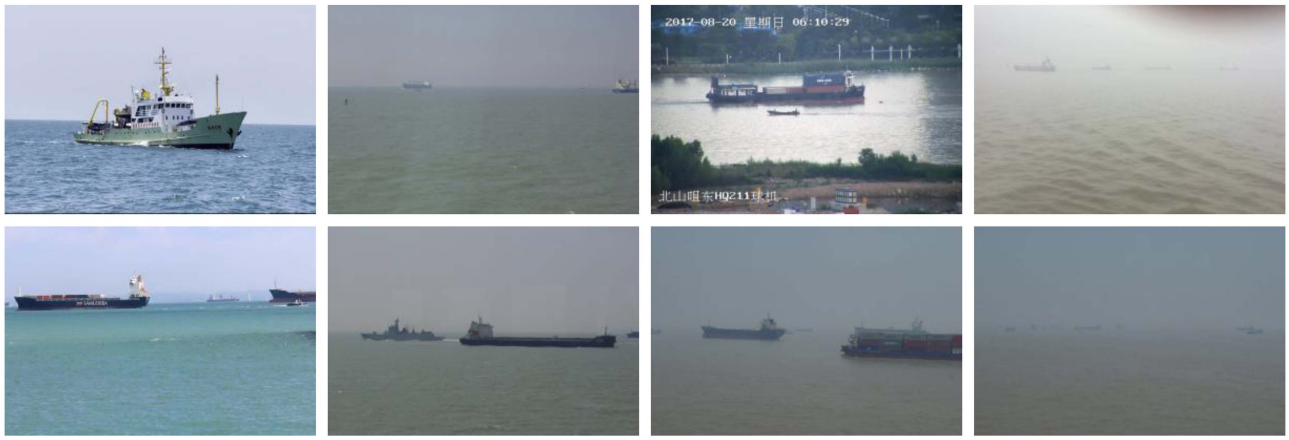

Fig. 6. Example images of the MORHL data set with different haze conditions. (a) Haze-free. (b) Light haze. (c) Medium haze. (d) Heavy haze.

where pre_i and gt_i represent the ith element of the predicted and ground truth class vectors. Finally, the cross-entropy loss is adopted as the classification loss

$$L_{\text{cls}} = -\frac{1}{N}\sum_{i=1}^{N} y_i \log(\text{p}_\text{i}) \tag{11}$$

where p_i denotes the estimated probability and y_i is the ground truth probability for class i, with N denoting the total number of classes.

Overall loss function: By aggregating all the previously mentioned losses, the total loss function is defined as

$$L = \lambda_1 L_{\text{G}} + \lambda_2 L_{\text{det}} + \lambda_3 L_{\text{sa}} \tag{12}$$

where λ_i balances the impact of individual losses. Our practical experiments demonstrate that the ship-aware recovery loss substantially enhances the dehazing performance through fine-tuning. Consequently, we establish the hyperparameters as follows to ensure stable training: setting λ_1 to 0.5, which contributes to training stability, while configuring λ_2 as 0.1 and λ_3 as 10 to optimize the impact of various loss components.

F. MORHL Data Set

The maritime environment is known for its complexity, particularly due to hazy weather conditions that present significant challenges to imaging quality when detecting objects. However, there is a notable lack of maritime data sets that are designed to assess the impact of haze and dehazing algorithms on object detection networks. Therefore, to thoroughly assess the performance of object detectors under haze conditions, we propose a MORHL data set. Our data collection process involved manual video recording and the use of search engines to obtain maritime images, ship videos, and ship images captured under various haze conditions. In addition, we integrated 7000 images from the SeaShips data set [44] and included the hazy subset of the Singapore Maritime Data Set (SMD) [45], totaling 692 images, into our MORHL data set.

Specifically, the MORHL data set is composed of 13 280 images split into a training set and a test set, with no duplication of images within each set. The training set comprises haze-free and hazy images, while the test set includes images depicting three haze levels: light, medium, and heavy haziness. This data set encompasses six distinct categories: cargo ship, container ship, fishing boat, passenger ship, island, and buoy. Figs. 6 and 7 provide visual examples of images from the data set depicting various haze conditions and distinct category labels. In addition, Table I offers statistical information into the MORHL data set.

IV. Experiments

A. Implementation Details

Our prior knowledge-driven MDD is developed using PyTorch and trained on a Linux workstation with an Nvidia GTX 4090 GPU. For synthesizing hazy images, we employ MIMNet [46] as the depth estimation network. During training, samples are resized to 512×512, and random horizontal flipping is applied for data augmentation. The training employs the ADAM optimizer with a learning rate set at 0.0001, a batch size of 4, and a cosine annealing schedule that reduces the learning rate to zero after 100 epochs.

B. Data Sets and Metrics

In this article, we train and evaluate our MDD framework using the proposed MORHL data set and the Singapore Maritime Data Set (SMD) [45]. Detailed information about the MORHL data set can be found in Section III-F. For the SMD data set, we employed the version by Moosbauer et al. [47], we then devide the SMD data set into a training set of 9200 haze-free images and a test set of 692 hazy images.

We select mean Average Precision (mAP) as our evaluation metric. It is calculated across multiple Intersection over Union (IoU) thresholds, ranging from 0.5 to 0.95 in 0.05 increments, providing a comprehensive evaluation of an object detector's performance across varying levels of overlap between the predicted and ground truth bounding boxes.

Fig. 7. Example images and annotations from the MORHL data set with different category labels. The ground truth bounding boxes are represented by colored rectangles: red for cargo ships, orange for container ships, blue for fishing boats, cyan for passenger ships, green for islands, and yellow for buoys. Please zoom in on the image for a better visualization.

TABLE I
SUMMARY OF MORHL DATA SET

	Haze level	Images	Cargo ship	Container ship	Fishing boat	Passenger ship	Island	buoy
Training	Clean	5439	17 654	1195	12 573	844	1638	4396
	Hazy	1133	1143	133	1010	157	16	156
Test_l	Light haze	1196	1082	195	355	61	20	298
Test_m	Medium haze	2512	24 548	12	0	0	0	12
Test_h	Heavy haze	3000	17 790	4009	0	0	0	0
Total	c+l+m+h	13 280	62 217	5544	13 938	1062	1674	4862

The symbols "c," "l," "m," and "h" denote different haze levels: clean, light, medium, and heavy, respectively.

TABLE II
PERFORMANCE COMPARISON WITH STATE-OF-THE-ART DETECTION METHODS ON SMD AND MORHL DATA SETS

Methods	#Params(M)	Latency(ms)	SMD(mAP)	MORHL (mAP)		
				Light haze	Medium haze	Heavy haze
YOLOv5-L	46	15.9	38.5	45.5	42.3	30.2
YOLOv8-L	43	22.9	39.3	46.2	42.9	29.8
YOLOv10-X	30	19.1	39.7	47.0	43.4	30.3
RTDETR-L	42	23.7	**41.5**[2]	**48.5**[2]	**44.1**[2]	**32.1**[2]
MDD	49	24.8	**43.1**[1]	**50.1**[1]	**46.3**[1]	**34.6**[1]

C. Performance Evaluation

1) Comparison With Different Object Detectors: We conduct experiments and compare the results with various object detection methods, including the one-stage detector YOLOv5 [48], YOLOv8 [41], YOLOv10 [49], and the transformer-based detector RT-DETR [35]. Specifically, we utilize YOLOv5-L, YOLOv8-L, and RT-DETR-L models with a similar number of parameters for performance comparison. YOLOv5 and

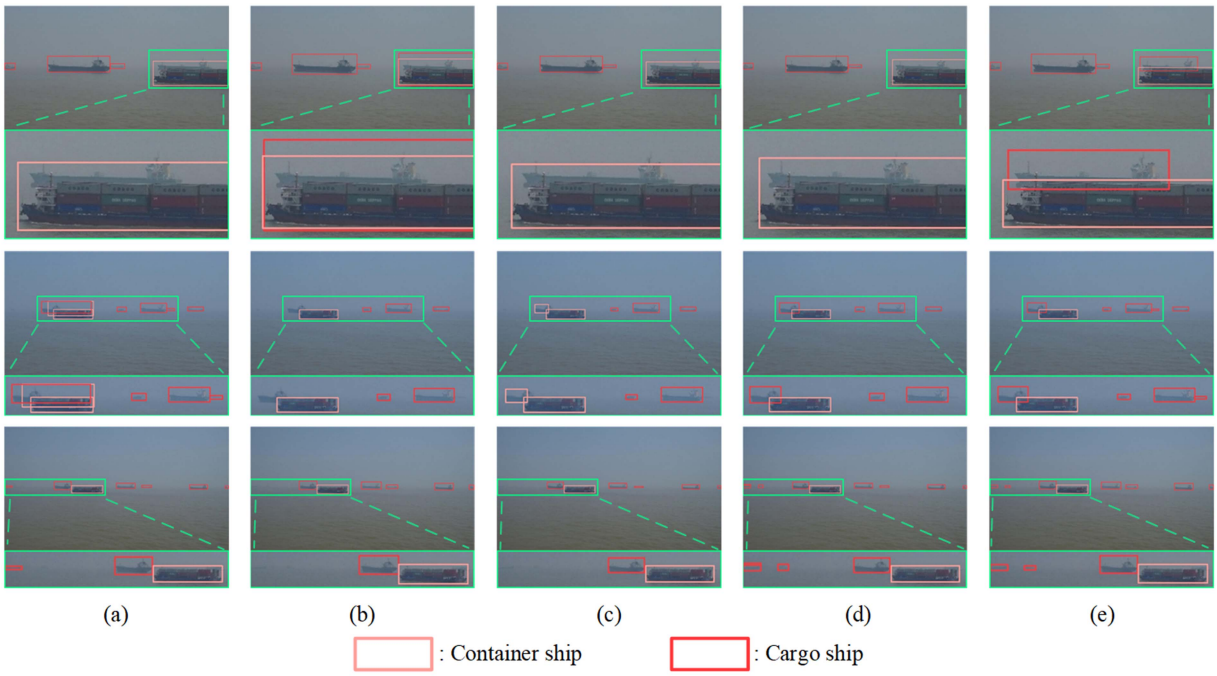

Fig. 8. Qualitative comparison of different detection methods on the MORHL test sets. (a) YOLOv5. (b) YOLOv8. (c) YOLOv10. (d) RT-DETR. (e) MDD.

TABLE III
PERFORMANCE COMPARISON ON THE SMD AND MORHL DATA SETS WITH INDEPENDENTLY USED DEHAZING AND DETECTION METHODS

Network		Dataset			
Dehazing	Detection	SMD (mAP)	MORHL (mAP)		
			Light haze	Medium haze	Heavy haze
AOD	YOLOv8-L	39.7	47.4	43.1	30.5
MSBDN	YOLOv8-L	40.9	48.1	43.6	31.5
4KDehazing	YOLOv8-L	39.5	47.9	43.4	31.4
SLAD	YOLOv8-L	40.3	46.3	42.9	31.0
RIDCP	YOLOv8-L	39.1	45.9	42.5	30.4
DEANet	YOLOv8-L	**41.5**[2]	**48.4**[2]	**43.9**[2]	**31.7**[2]
MDD		**43.1**[1]	**50.1**[1]	**46.3**[1]	**34.6**[1]

The bold values indicate the top two rankings for each metric.

YOLOv8 are chosen as baselines due to their established effectiveness and real-time performance in a wide range of object detection tasks. In addition, we include the YOLOv10, known for its lightweight design and efficiency, and select its largest variant, YOLOv10-X, to ensure a fair comparison. RT-DETR is also incorporated as it represents a real-time transformer-based detector, combining the advantages of transformer architectures with low latency.

The quantitative performance comparison of these models with our MDD framework in hazy maritime scenarios for ship detection is presented in Table II, where the bold values indicate the top two rankings for each metric. In this table, "Latency" refers to the average time required to process a single image, including preprocessing, forward propagation, and postprocessing. The results show that MDD achieved the highest mAP values across the SMD and MORHL test sets, under light, medium, and heavy haze conditions. Specifically, MDD outperformed the second-best method by 1.6%, 2.2%, and 2.5% on the light, medium, and heavy haze test sets, respectively. This consistent superiority emphasizes the effectiveness of MDD in diverse haze scenarios. Notably, the most significant improvement was observed on the heavy haze test set, further demonstrating MDD's robust ability to accurately detect ships even in dense haze conditions. Unlike other methods that lack haze-specific optimization, our MDD framework leverages IDCP knowledge to enhance ship detection in hazy environments. As haze intensity increases, the advantage of our method becomes more apparent, effectively reducing the haze's impact on feature extraction and focusing on relevant ship features. Although MDD exhibits a slight increase in latency, with a 1.1 ms increase over RT-DETR-L, it achieves a 1.6% improvement on the SMD data set and a 2.5% improvement on the MORHL heavy haze test set.

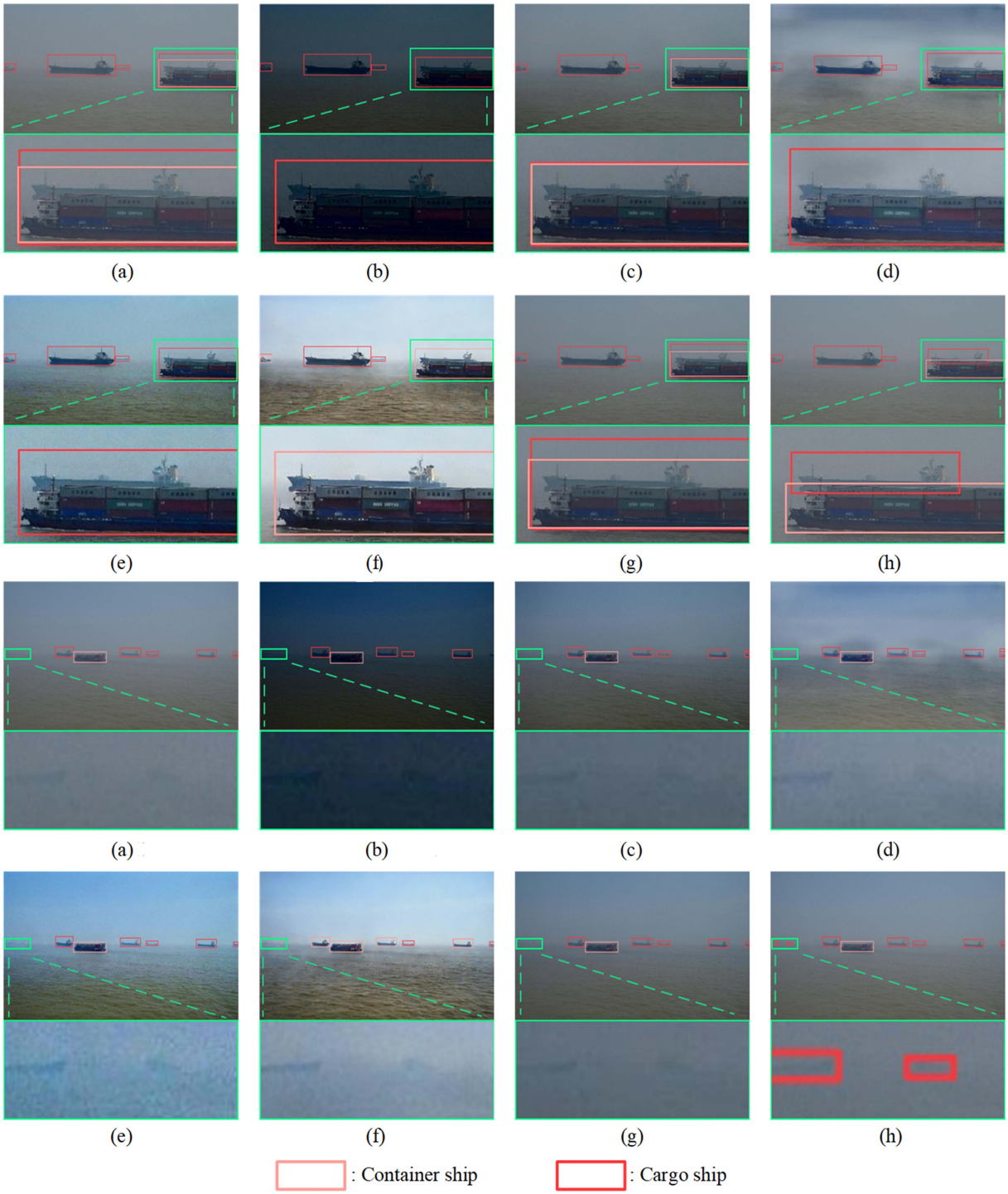

Fig. 9. Qualitative comparison on the MORHL data set with independently used dehazing and detection methods. (a) Hazy. (b) AOD. (c) MSBDN. (d) 4KDehazing. (e) SLAD. (f) RIDCP. (g) DEANet. (h) MDD.

A qualitative comparison of various object detection methods evaluated on the MORHL test sets is illustrated in Fig. 8. The comparison highlights that the MDD framework consistently outperforms other models across various challenging conditions. From the third set of comparisons, it can be observed that MDD is able to accurately detect small distant ships covered by haze. Although RT-DETR also demonstrates good detection accuracy, it encounters the issue of duplicate detections. Detecting occluded targets is challenging due to partial obstruction, and haze further complicates this by obscuring key features. However, our MDD framework overcomes these issues by leveraging prior knowledge-driven feature learning and the IDCP cross-attention module, which enhances the model's ability to focus on ship regions even in the presence of occlusion and haze. In addition, the restoration network, along with the ship-aware reconstruction loss, enables the model to

TABLE IV
ABLATION STUDY OF MDD WITH DIFFERENT SETTINGS ON SMD AND MORHL DATA SETS

Setting		1	2	3	4	5	6
SHES			✓		✓	✓	✓
L_{sa}						✓	✓
Prior-subnetwork + CA				✓	✓		✓
Restoration Network						✓	✓
$\mathrm{mAP}_{\mathrm{SMD}}(\%)$		39.3	39.5	41.5	42.3	**42.6**[2]	**43.1**[1]
$\mathrm{mAP}_{\mathrm{MORHL}}(\%)$	Light haze	46.2	46.9	47.9	48.9	**49.3**[2]	**50.1**[1]
	Medium haze	42.9	43.3	44.7	45.5	**45.6**[2]	**46.3**[1]
	Heavy haze	29.8	30.1	32.0	33.1	**33.8**[2]	**34.6**[1]

The bold values indicate the top two rankings for each metric.

better learn ship features by enhancing their clarity, thereby improving detection performance in occlusion conditions. Overall, MDD demonstrates exceptional performance in detecting ships amidst occlusion and heavy haze, as well as effectively identifying small target ships, which are typically difficult to detect.

2) Comparison With Dehazing Combined Detection Methods: We also compared our MDD with a conventional approach where dehazing is performed prior to object detection. In this comparison, a dehazing network was used as a preprocessing step, followed by the YOLOv8-L detection network. This allowed us to assess how well our integrated framework performs relative to a traditional two-step process involving separate dehazing and detection stages. We conduct the experiments on MORHL and SMD data sets and make comparison with diverse image dehazing methods including AOD [50], MS-BDN [11], 4KDehazing [51], SLAD [52], RIDCP [53], and DEANet [54].

Table III and Fig. 9 provide a comprehensive evaluation of our MDD framework against traditional approaches that first apply dehazing followed by object detection. The quantitative results in Table III indicate that, while the application of a dehazing network as a preprocessing step results in enhanced detection performance, there remains a significant performance gap when compared to our integrated MDD framework. Specifically, while these traditional dehazing methods improve the overall visual quality of the images by removing haze, they primarily focus on enhancing image clarity without considering the impact of haze removal on the feature representation required for accurate object detection. As a result, although the haze is reduced, these methods fail to optimize the detection network's ability to extract relevant features for better localization and classification. In contrast, our MDD framework integrates dehazing with detection, allowing the network to jointly learn both tasks, which leads to a more effective feature extraction process. This holistic approach ensures that the improvements in image quality directly translate into better detection performance, resulting in mAP improvements of 1.6%, 1.7%, 2.4%, and 2.9% on the SMD, light haze, medium haze, and heavy haze test sets of the MORHL data sets, respectively.

Fig. 9 complements these quantitative findings with qualitative results, highlighting that our MDD framework provides superior performance in real-world scenarios. The visual comparisons demonstrate that our framework improves detection accuracy in challenging conditions, such as occlusions and small target ships under haze. The combined evidence from both quantitative and qualitative analyses underscores the effectiveness of our unified approach, reinforcing that our integrated dehazing and detection framework is highly effective for maritime object detection in hazy conditions.

D. Model Analyses

1) Ablation Study: This section evaluates the performance of our MDD framework across various configurations. Table IV and Fig. 10 illustrate the impact of 1) the SHES, 2) the ship-aware reconstruction loss (L_{sa}), 3) the combination of the prior subnetwork and IDCP cross-attention, and 4) the restoration network on detection performance.

From Table IV, it can be observed that when the network is trained using the SHES on hazy images, the mAP shows a slight improvement compared to the baseline, with a 0.2% increase on the SMD data set. In contrast, when the network is trained using the combination of the prior subnetwork and IDCP cross-attention on clear images, a more substantial performance gain is achieved. Specifically, compared to Setting 2, the network shows a 2.0% improvement on the SMD data set, and increases of 1.0%, 1.4%, and 1.9% for light, medium, and heavy haze conditions, respectively, on the MORHL data set. These results demonstrate that even when trained on clear images, the network can still effectively capture ship target features in hazy environments by leveraging prior knowledge.

When the SHES, ship-aware reconstruction loss, and restoration network are employed, the network benefits from the combined contribution of these components. The SHES helps to simulate hazy conditions more effectively, improving the model's robustness in hazy environments. The ship-aware reconstruction loss enhances the network's ability to preserve and focus on ship-related features during the restoration process. Meanwhile, the restoration network further refines the image quality by

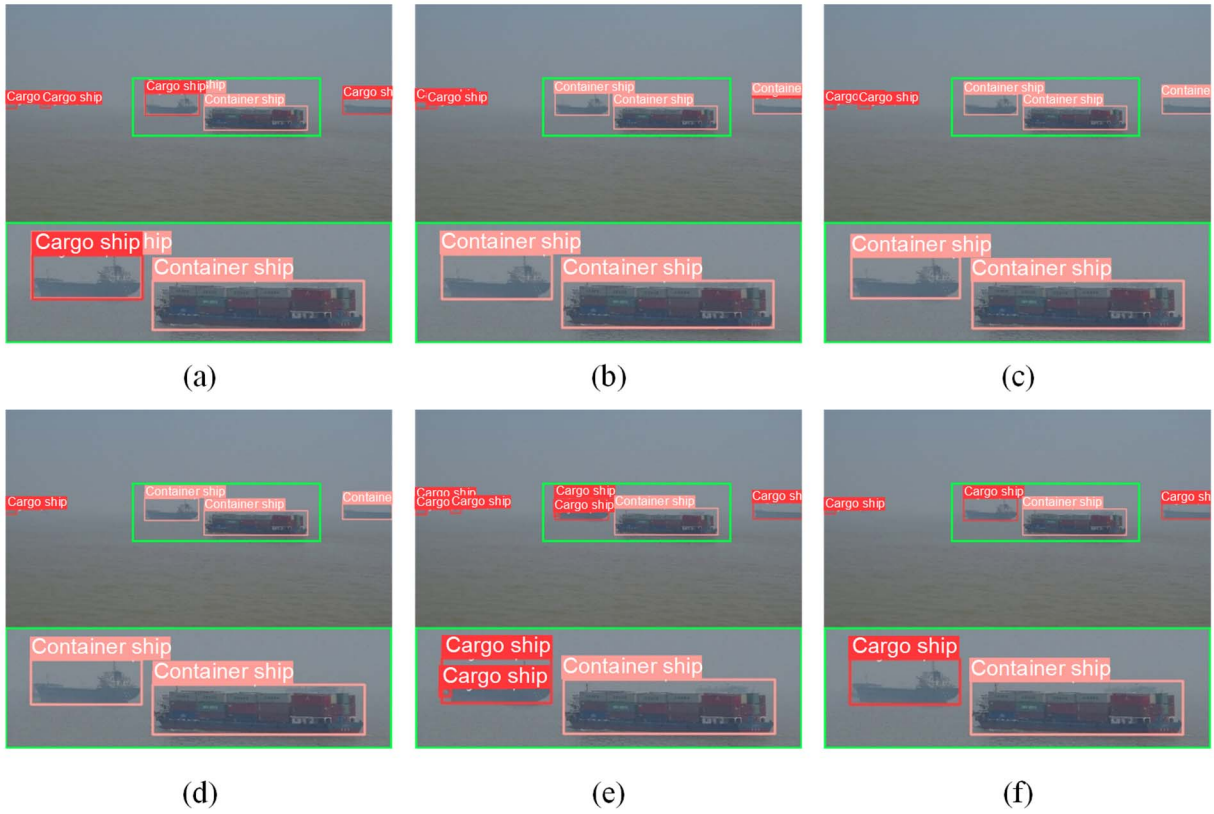

Fig. 10. Qualitative comparison of MDD detection results under different settings. (a) Setting 1. (b) Setting 2. (c) Setting 3. (d) Setting 4. (e) Setting 5. (f) Setting 6.

mitigating haze effects, ensuring better feature visibility. As a result, the network achieves the second-best performance across the four test sets, demonstrating the synergistic effectiveness of these components in improving overall performance.

The qualitative comparison of MDD results under different configurations is presented in Fig. 10. It can be observed that when all components are applied, the network achieves the best detection performance. In contrast, networks using alternative configurations encounter issues such as false detections and duplicate detections. These findings emphasize the critical role that each component plays in improving detection accuracy and minimizing errors, highlighting the effectiveness of the full configuration in optimizing performance.

In summary, while each component—the prior subnetwork and IDCP cross-attention, the SHES, the ship-aware reconstruction loss, and the restoration network—individually contributes to enhancing object detection, their combined application yields the best results. This phenomenon demonstrates that the synergy between components significantly boosts overall detection performance, confirming the efficacy of our approach in addressing challenging maritime hazy conditions.

2) Effectiveness of Different Losses: To assess the impact of different weights for the ship-aware recovery loss on the network's detection performance, we conducted a series of ablation experiments. The results presented in Table V show that the network achieves the best detection accuracy when the weight is set to 10, with performance declining when the weight is either smaller or larger than this value. Specifically, when the weight is smaller than the optimal value, the effect of the ship-aware recovery loss on the network weakens, resulting in the network's failure to effectively capture key ship features. On the other hand, when the weight exceeds 10, the network tends to prioritize enhancing the visual quality of the recovered image over detection accuracy. These findings underscore the importance of carefully balancing the weight of the ship-aware recovery loss to ensure that the network can both recover the visual quality of the image and maintain strong detection capabilities.

TABLE V
ABLATION STUDY ON THE EFFECT OF DIFFERENT WEIGHTS FOR THE SHIP-AWARE RECOVERY LOSS

Weight		1	5	10	15	20
mAP_{SMD} (%)		39.4	42.5	**43.1**	41.9	39.6
$\text{mAP}_{\text{MORHL}}$(%)	Light haze	47.0	49.2	**50.1**	48.3	47.2
	Medium haze	43.5	45.7	**46.3**	44.6	43.7
	Heavy haze	32.1	33.8	**34.6**	33.0	32.5

The bold values indicate the top two rankings for each metric.

3) Limitations: Like previous methods, the proposed approach has certain limitations. When certain areas in the image tend toward white (e.g., ocean waves), the inverted dark channel tends to suppress these areas, as demonstrated in Fig. 11. This happens because white regions typically have high pixel values across all three channels in the input image, causing them to appear bright in the dark channel [16]. Upon inverting the dark

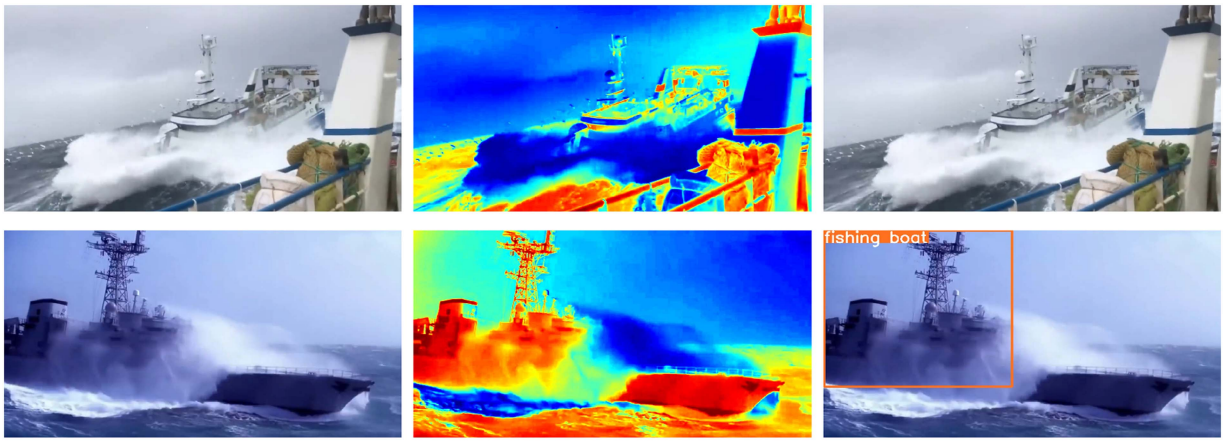

Fig. 11. Detection failure cases of the MDD. In the IDC, red indicates higher pixel values, while dark blue represents lower pixel values. (a) Input image. (b) IDC. (c) Detection result.

channel, these regions are represented as dark. Consequently, when ships are partially covered by ocean waves, these regions are also suppressed in the IDC due to their high pixel values and further impede the network's ability to extract features from these areas, leading to missed detections, as shown in Fig. 11(c). This limitation highlights the need for further refinement to more effectively handle bright regions and ship features obscured by waves in future work.

V. Conclusion

This article has presented a prior knowledge-driven MDD, which comprises both a detection network and a restoration network. To address the specific challenges of maritime scenarios, we introduced an IDCP designed to highlight the ships in maritime images. The detection network incorporates a prior subnetwork: the backbone processes hazy input images, while the prior subnetwork handles the corresponding inverted dark channels. In addition, we developed a cross-attention module within the backbone network to integrate abstract information from both sources. During training, the restoration network is incorporated to improve the clarity of the features extracted by the object detection network. Furthermore, a ship-haze enrichment strategy and ship-aware reconstruction loss are implemented to ensure that the network focuses on ship regions during training. To assess the influence of haze and dehazing techniques on object detection, we constructed the MORHL data set. Experimental results across different data sets indicate that the MDD framework outperforms state-of-the-art object detectors and dehazing-detection combinations.

In future applications, the MDD framework has the potential to be adapted for image enhancement and detection tasks across various adverse weather conditions, including low-light and rainy environments. This can be achieved by synthesizing degraded images from diverse environments and employing joint detection network training. Such extensions will enable the algorithm to effectively handle real-world scenarios, adapt to more complex weather conditions, and enhance the robustness of object detection models.

References

[1] Z. Wu, J. Wen, Y. Xu, J. Yang, and D. Zhang, "Multiple instance detection networks with adaptive instance refinement," *IEEE Trans. Multimedia*, vol. 25, pp. 267–279, 2023.

[2] J. Redmon and A. Farhadi, "YOLO9000: Better, faster, stronger," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 7263–7271.

[3] A. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," Accessed on: Jan. 3, 2024. [Online]. Available: https://arxiv.org/abs/1804.02767

[4] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017.

[5] W. Liu et al., "SSD: Single shot multibox detector," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 21–37.

[6] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 2980–2988.

[7] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 2961–2969.

[8] L. Ying, D. Miao, and Z. Zhang, "A robust one-stage detector for SAR ship detection with sequential three-way decisions and multi-granularity," *Inf. Sci.*, vol. 667, 2024, Art. no. 120436.

[9] L. Liu et al., "CLFR-DET: Cross-level feature refinement detector for tiny-ship detection in SAR images," *Knowl.-Based Syst.*, vol. 284, 2024, Art. no. 111284.

[10] X. Chen, H. Wu, B. Han, W. Liu, J. Montewka, and R. W. Liu, "Orientation-aware ship detection via a rotation feature decoupling supported deep learning approach," *Eng. Appl. Artif. Intell.*, vol. 125, 2023, Art. no. 106686.

[11] H. Dong et al., "Multi-scale boosted dehazing network with dense feature fusion," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 2157–2167.

[12] J. Luo, Q. Bu, L. Zhang, and J. Feng, "Global feature fusion attention network for single image dehazing," in *Proc. IEEE Int. Conf. Multimedia Expo Workshops*, 2021, pp. 1–6.

[13] D. P. Williams, "Fast target detection in synthetic aperture sonar imagery: A new algorithm and large-scale performance analysis," *IEEE J. Ocean. Eng.*, vol. 40, no. 1, pp. 71–92, Jan. 2015.

[14] K. Panetta, L. Kezebou, V. Oludare, and S. Agaian, "Comprehensive underwater object tracking benchmark dataset and underwater image enhancement with GAN," *IEEE J. Ocean. Eng.*, vol. 47, no. 1, pp. 59–75, Jan. 2022.

[15] T. Zhou, Y. Wang, L. Zhang, B. Chen, and X. Yu, "Underwater multitarget tracking method based on threshold segmentation," *IEEE J. Ocean. Eng.*, vol. 48, no. 4, pp. 1255–1269, Oct. 2023.

[16] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2009, pp. 1956–1963.

[17] H. Wang, K. Köser, and P. Ren, "Large foundation model empowered discriminative underwater image enhancement," *IEEE Trans. Geosci. Remote Sens.*, vol. 63, 2025, Art. no. 5609317.

[18] H. Li, H. Wang, Y. Zhang, L. Li, and P. Ren, "Underwater image captioning: Challenges, models, and datasets," *ISPRS J. Photogrammetry Remote Sens.*, vol. 220, pp. 440–453, 2025.
[19] H. Wang et al., "Inspiration: A reinforcement learning-based human visual perception-driven image enhancement paradigm for underwater scenes," *Eng. Appl. Artif. Intell.*, vol. 133, 2024, Art. no. 108411.
[20] S. Nayar and S. Narasimhan, "Vision in bad weather," in *Proc. 7th IEEE Int. Conf. Comput. Vis.*, 1999, pp. 820–827.
[21] B. Cai, X. Xu, K. Jia, C. Qing, and D. Tao, "DehazeNet: An end-to-end system for single image haze removal," *IEEE Trans. Image Process.*, vol. 25, no. 11, pp. 5187–5198, Nov. 2016.
[22] W. Ren, S. Liu, H. Zhang, J. Pan, X. Cao, and M.-H. Yang, "Single image dehazing via multi-scale convolutional neural networks," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 154–169.
[23] L. Li et al., "Semi-supervised image dehazing," *IEEE Trans. Image Process.*, vol. 29, pp. 2766–2779, 2019.
[24] Y. Shao, L. Li, W. Ren, C. Gao, and N. Sang, "Domain adaptation for image dehazing," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2020, pp. 2805–2814.
[25] S. Zheng, J. Sun, Q. Liu, Y. Qi, and S. Zhang, "Overwater image dehazing via cycle-consistent generative adversarial network," in *Proc. Asian Conf. Comput. Vis.*, 2020, pp. 1–16.
[26] Z. Wu, J. Wen, Y. Xu, J. Yang, X. Li, and D. Zhang, "Enhanced spatial feature learning for weakly supervised object detection," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 35, no. 1, pp. 961–972, Jan. 2024.
[27] Z. Wu, Y. Xu, J. Yang, and X. Li, "Misclassification in weakly supervised object detection," *IEEE Trans. Image Process.*, vol. 33, pp. 3413–3427, 2024.
[28] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 779–788.
[29] Z. Cai and N. Vasconcelos, "Cascade R-CNN: Delving into high quality object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 6154–6162.
[30] Z. Wu, C. Liu, J. Wen, Y. Xu, J. Yang, and X. Li, "Selecting high-quality proposals for weakly supervised object detection with bottom-up aggregated attention and phase-aware loss," *IEEE Trans. Image Process.*, vol. 32, pp. 682–693, 2023.
[31] H. Law and J. Deng, "CornerNet: Detecting objects as paired keypoints," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 734–750.
[32] K. Duan, S. Bai, L. Xie, H. Qi, Q. Huang, and Q. Tian, "CenterNet: Keypoint triplets for object detection," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2019, pp. 6569–6578.
[33] A. Dosovitskiy et al., "An image is worth 16x16 words: Transformers for image recognition at scale," Accessed on: Jan. 15, 2024. [Online]. Available: https://arxiv.org/abs/2010.11929
[34] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko, "End-to-end object detection with transformers," in *Proc. Eur. Conf. Comput. Vis.*, 2020, pp. 213–229.
[35] Y. Zhao et al., "DETRs beat YOLOs on real-time object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2024, pp. 16965–16974.
[36] J. Yim and K.-A. Sohn, "Enhancing the performance of convolutional neural networks on quality degraded datasets," in *Proc. Int. Conf. Digit. Image Comput.: Tech. Appl.*, 2017, pp. 1–8.
[37] D. Dai, Y. Wang, Y. Chen, and L. Van Gool, "Is image super-resolution helpful for other vision tasks?," in *Proc. IEEE Winter Conf. Appl. Comput. Vis.*, 2016, pp. 1–9.
[38] D. Liu, B. Wen, J. Jiao, X. Liu, Z. Wang, and T. S. Huang, "Connecting image denoising and high-level vision tasks via deep learning," *IEEE Trans. Image Process.*, vol. 29, pp. 3695–3706, 2020.
[39] S.-C. Huang, T.-H. Le, and D.-W. Jaw, "DSNet: Joint semantic learning for object detection in inclement weather conditions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 8, pp. 2623–2633, Aug. 2021.
[40] W. Liu, G. Ren, R. Yu, S. Guo, J. Zhu, and L. Zhang, "Image-adaptive YOLO for object detection in adverse weather conditions," in *Proc. AAAI Conf. Artif. Intell.*, 2022, pp. 1792–1800.
[41] R. Varghese and M. Sambath, "YOLOv8: A novel object detection algorithm with enhanced performance and robustness," in *Proc. Int. Conf. Adv. Data Eng. Intell. Comput. Syst.*, 2024, pp. 1–6.
[42] A. G. Howard et al., "MobileNets: Efficient convolutional neural networks for mobile vision applications," Accessed on: Jan. 10, 2024. [Online]. Available: https://arxiv.org/abs/1704.04861
[43] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2015, pp. 1440–1448.
[44] Z. Shao, W. Wu, Z. Wang, W. Du, and C. Li, "SeaShips: A large-scale precisely annotated dataset for ship detection," *IEEE Trans. Multimed.*, vol. 20, no. 10, pp. 2593–2604, Oct. 2018.
[45] D. K. Prasad, D. Rajan, L. Rachmawati, E. Rajabally, and C. Quek, "Video processing from electro-optical sensors for object detection and tracking in a maritime environment: A survey," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 8, pp. 1993–2016, Aug. 2017.
[46] Z. Xie, Z. Geng, J. Hu, Z. Zhang, H. Hu, and Y. Cao, "Revealing the dark secrets of masked image modeling," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2023, pp. 14475–14485.
[47] S. Moosbauer, D. König, J. Jäkel, and M. Teutsch, "A benchmark for deep learning based object detection in maritime environments," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops*, 2019, pp. 916–925.
[48] G. Jocher et al., "Ultralytics/YOLOv5: V3.1 - bug fixes and performance improvements," Accessed on: Jan. 15, 2024. [Online]. Available: https://github.com/ultralytics/yolov5
[49] A. Wang et al., "YOLOv10: Real-time end-to-end object detection," in *Proc. Adv. Neural Inf. Proces. Syst.*, 2024, pp. 107984–108011.
[50] B. Li, X. Peng, Z. Wang, J. Xu, and D. Feng, "AOD-Net: All-in-one dehazing network," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2017, pp. 4780–4788.
[51] Z. Zheng et al., "Ultra-high-definition image dehazing via multi-guided bilateral learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 16180–16189.
[52] Y. Liang, B. Wang, W. Zuo, J. Liu, and W. Ren, "Self-supervised learning and adaptation for single image dehazing," in *Proc. 31st Int. Joint Conf. Artif. Intell.*, 2022, pp. 1–15.
[53] R.-Q. Wu, Z.-P. Duan, C.-L. Guo, Z. Chai, and C. Li, "RIDCP: Revitalizing real image dehazing via high-quality codebook priors," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2023, pp. 22282–22291.
[54] Z. Chen, Z. He, and Z.-M. Lu, "DEA-Net: Single image dehazing based on detail-enhanced convolution and content-guided attention," *IEEE Trans. Image Process.*, vol. 33, pp. 1002–1015, 2024.

Yaozong Mo received the B.S. degree in electrical engineering and automation from Xiangnan University, Chenzhou, China, in 2019. He is currently working toward the Ph.D. degree with the Institute of Logistics Science and Engineering, Shanghai Maritime University Shanghai, China.

His research interests include image dehazing and object detection.

Chaofeng Li (Senior Member, IEEE) received the B.S. degree in coal geology, the M.S. degree in mathematical geology, and the Ph.D. degree in mineral resource prospecting and exploration from the Chinese University of Mining and Technology, Xuzhou, China, in 1995, 1998, and 2001, respectively.

From 2001 to 2003, he was finishing Postdoctoral research work with the Nanjing University of Science and Technology Nanjing, China. From 2003 to 2017, he was an Associate Professor and a Professor with the School of Internet of Things Engineering, Jiangnan University, Wuxi, China. From 2008 to 2009, he was a Visiting Research Scholar with the Laboratory for Image and Video Engineering, The University of Texas at Austin, Austin, TX, USA. He is currently a Professor with the Institute of Logistics Science and Engineering, Shanghai Maritime University, Shanghai, China. His current research interests include image processing, deep learning, and pattern recognition.

Wenqi Ren (Member, IEEE) received the Ph.D. degree in technology for computer applications from Tianjin University, Tianjin, China, in 2017.

From 2015 to 2016, he was supported by China Scholarship Council and worked with Prof. Ming-Husan Yang as a Joint-Training Ph.D. Student with the Electrical Engineering and Computer Science Department, University of California, Merced, CA, USA. He is currently a Professor with the School of Cyber Science and Technology, Sun Yatsen University, Shenzhen Campus, Shenzhen, China. His research interests include image processing and related high-level vision problems.

Wenwu Wang (Senior Member, IEEE) was born in Anhui, China. He received the B.Sc., M.E., and Ph.D. degrees, all in the area of automation, from Harbin Engineering University, Harbin, China, in 1997, 2000, and 2002, respectively.

In 2007, he joined the University of Surrey, Guildford, England, U.K., where he is currently a Professor of signal processing and machine learning, and the Co-Director of the Machine Audition Lab within the Centre for Vision Speech and Signal Processing. Before joining University of Surrey, he was with King's College London, London, England, Cardiff University, Cardiff, U.K., Tao Group, Ltd. (now Antix Labs Ltd.), and Creative Labs Reading, England. He has authored or coauthored more than 300 publications in his research areas which includes blind signal processing, sparse signal processing, audio-visual signal processing, machine learning and perception, machine audition (listening), and statistical anomaly detection.

Dr. Wang was an Associate Editor and the Senior Area Editor for IEEE TRANSACTIONS ON SIGNAL PROCESSING, from 2014 to 2018, and from 2019 to 2023, respectively. He is currently an Associate Editor for IEEE/ACM TRANSACTIONS ON AUDIO SPEECH AND LANGUAGE PROCESSING. From 2023 to 2024, he was the Elected Chair of IEEE Signal Processing Society Machine Learning for Signal Processing, and from 2025 to 2027, he is the Elected Chair of EURASIP Technical Area Committee on Audio Speech and *Music Signal Processing*.