WILEY | Hindawi

Research Article

Intelligent Vision-Enabled Detection of Water-Surface Targets for Video Surveillance in Maritime Transportation

Yongqi Guo,[1] Yuxu Lu,[2] Yu Guo,[2] Ryan Wen Liu [iD],[2] and Kwok Tai Chui [iD][3]

[1]*Center of Teaching Supervision, Wuhan University of Technology, Wuhan 430070, China*
[2]*School of Navigation, Wuhan University of Technology, Wuhan 430063, China*
[3]*School of Science and Technology, The Open University of Hong Kong, Ho Man Tin, Hong Kong*

Correspondence should be addressed to Ryan Wen Liu; wenliu@whut.edu.cn and Kwok Tai Chui; jktchui@ouhk.edu.hk

The timely, automatic, and accurate detection of water-surface targets has received significant attention in intelligent vision-enabled maritime transportation systems. The reliable detection results are also beneficial for water quality monitoring in practical applications. However, the visual image quality is often inevitably degraded due to the poor weather conditions, potentially leading to unsatisfactory target detection results. The degraded images could be restored using state-of-the-art visibility enhancement methods. It is still difficult to generate high-quality detection performance due to the unavoidable loss of details in restored images. To alleviate these limitations, we first investigate the influences of visibility enhancement methods on detection results and then propose a neural network-empowered water-surface target detection framework. A data augmentation strategy, which synthetically simulates the degraded images under different weather conditions, is further presented to promote the generalization and feature representation abilities of our network. The proposed detection performance has the capacity of accurately detecting the water-surface targets under different adverse imaging conditions, e.g., haze, low-lightness, and rain. Experimental results on both synthetic and realistic scenarios have illustrated the effectiveness of the proposed framework in terms of detection accuracy and efficacy.

1. Introduction

The visual information captured from video is able to provide more meaningful and important data in intelligent transportation systems. With the popularization of artificial intelligence (AI) and Internet of Things (IoT), intelligent vision techniques have witnessed the rapid developments of video surveillance in the IoT-empowered maritime transportation system [1–3]. To enhance traffic safety and maritime monitoring, the timely and accurate detection of surface-water targets (e.g., ship, garbage, and person in water) has received tremendous attention in the current literature. The traditional detection methods, such as mean shift [4], deformable part-based models (DPMs) [5], support vector machine (SVM) [6], and sparse representation [7], have been proposed to detect the targets of interest. However, the corresponding detection results easily suffer from the complicated environments, including water-surface reflect light and multimoving targets. To further improve detection results, deep learning has gained increasing attention during the past several years. The learning-based detection methods can be mainly divided into two types, i.e., two-stage and one-stage methods. In the literature, R-CNN [8], Fast R-CNN [9], and Faster R-CNN [10] belong to the representative two-stage strategy. The accurate detection results could be obtained but at the expense of a high computational cost. To guarantee real time detection, the one-stage methods have recently gained much attention in practical applications. The typical methods mainly include YOLOv1 [11], which is based on image global information, and its advanced extensions (e.g., YOLOv2 [12], YOLOv3 [13], and YOLOv4 [14]). These methods could achieve a good balance between detection accuracy and efficiency and have been widely adopted in practical applications. There is

thus a great potential to exploit these methods to detect water-surface targets in intelligent vision-enabled maritime transportation systems [15, 16].

However, it is often intractable to observe high-quality video images under different severe weather conditions (e.g., hazy, low-light imaging, and rainy), which easily occur in practice. The corresponding degraded visibility will be harmful for reliable detection of water-surface targets, even for deep learning-enhanced detection methods. The existing visibility enhancement methods can theoretically improve image visual quality. It should be noted that it is impossible to generate restored images without loss of any details [17]. Due to this essential limitation, the learning-based detection methods easily fail to accurately recognize the water-surface targets. To guarantee accurate and robust detection, the data augmentation strategies can be incorporated into existing learning-based detection frameworks. It has the capacity of significantly improving detection results under different imaging conditions. Taking detection of water-surface garbage as an example, it is able to guarantee reliable water quality monitoring for IoT-based maritime video surveillance. In current literature [18–20], most studies on water quality monitoring mainly focus on spectrum analysis and physical and chemical analysis in several IoT-empowered practical applications. To our knowledge, rare studies have been implemented on automatic detection of water-surface garbage through IoT-based maritime video surveillance.

The main contributions related to intelligent water-surface target detection and water quality monitoring are threefold in this work:

(1) An intelligent vision-enabled water-surface target detection framework with deep neural networks has been proposed for IoT-based maritime video surveillance

(2) Optimal strategies for training deep neural networks have been represented to handle the influences of different severe weather conditions on water-surface target detection

(3) Extensive detection experiments on both simulated and real-world scenarios have demonstrated that the proposed vision-enabled water-surface target detection framework could provide robust and accurate results under different imaging conditions

The main benefit of our water-surface target detection framework is that two aspects have been taken into consideration, i.e., the powerful learning capacity of deep neural networks and the data augmentation-based network learning strategy. Comprehensive experiments under different severe weather conditions demonstrate that the proposed framework could accurately and robustly detect the water-surface targets for maritime video surveillance.

The reminder of this work is composed of several sections. Section 2 briefly reviews the recent work related to IoT-based maritime video surveillance and detection of water-surface targets. In Section 3, an intelligent vision-enabled target detection framework is proposed to promote maritime surveillance under different weather conditions. Comprehensive experiments are performed to demonstrate the effectiveness of our detection framework in Section 4. This paper is finally concluded by summarizing the main contributions in Section 5.

2. Related Work

In the current literature, many efforts have been devoted to maritime surveillance and water quality monitoring. We will briefly review the intelligent maritime surveillance and water-surface target detection in this section.

2.1. Intelligent Video Surveillance in Maritime Transportation. Intelligent maritime surveillance has recently attracted tremendous attention [21], which enables the understanding of various maritime activities leading to enhanced maritime safety and security. With the recent burgeoning application of computer vision technology, visual-based maritime surveillance systems could provide more reliable applications, such as traffic safety management and water pollution monitoring. In the current literature, Bloisi et al. [22] proposed to promote existing maritime surveillance systems through the popular closed-circuit television (CCTV) camera device which could provide useful visual information. To promote maritime video surveillance under different weather conditions, many visibility enhancement methods were developed to improve imaging quality. For example, illumination decomposition-based image dehazing [23] reconstructed the natural-looking dehazed images in visual maritime surveillance. The observed maritime images captured in low-light conditions have also been improved through the Retinex theory [24] and deep learning [25]. The high-quality images in maritime video surveillance are potentially conducive to promoting maritime monitoring in practice. To achieve more effective visual maritime surveillance, the maritime objects of interest should also be robustly and accurately detected, recognized, and tracked [26–28].

With the rapid developments of low-end IoT devices and emerging AI techniques [29, 30], IoT-enabled intelligent maritime video surveillance has received increasing attention both from academics and practitioners. Liu et al. [2] proposed to improve big data quality to promote intelligent vessel traffic services in maritime IoT systems. To further improve the efficacy of maritime surveillance, an AI-enpowered maritime IoT was developed by proposing a parallel-network-driven approach [31]. Palma [32] also proposed to enable maritime IoT across the seas and oceans through the performance analysis of CoAP and 6LoWPAN over VHF links. By combining the AI-based computer vision and IoT, it is tractable to robustly and accurately detect the undesirable water-surface targets under different weather conditions. There is thus a great potential for monitoring water quality with the IoT-based maritime video surveillance. For more details on maritime IoT, please refer to [33] and references therein.

2.2. Detection of Water-Surface Targets in Maritime Surveillance. The detection of water-surface targets (i.e., garbage) is beneficial for water quality monitoring in maritime surveillance. The reliable evaluation of water quality is challenged for management activities aiming at protecting the limited water resources. Several different types of water quality monitoring systems have been developed to assist in assessing water quality and providing early warning [19]. Adu-Manu et al. [20] has reviewed the main water quality monitoring methods from traditional manual to newly developed methods. The paper-based monitor sensors and smart cell phones were combined for on-site water quality monitoring [34]. However, it becomes difficult to implement real time water quality monitoring in a large-scale region through this technology. To overcome this limitation, Chung and Yoo [35] proposed to implement remote water quality monitoring through the wireless sensor network (WSN). The network of smart sensors was proposed to implement *in situ* and *in continuous* spatio-temporal monitoring of surface-water quality [36]. For more details on the traditional water quality monitoring system, please refer to [19, 37, 38] and references therein. Video surveillance has become an emerging manner to indirectly assess water quality. For example, Serra-Toro et al. [39] tended to monitor water quality by recognizing the fish swimming behavior from video images. To promote learning-based waste detection in water bodies, a dataset (i.e., AquaTrash) [40] was developed based on existing TACO dataset [41] to assist in protecting water sources. Benefiting from the strong learning capacity of deep models, an extension of YOLOv3 [42] performs well in effective detection of vision-based water-surface garbage. The YOLOv3 network has been embedded into an intelligent water-surface cleaner robot, which is capable of accurately and real timely detecting and collecting floating garbage [43]. However, if the observed images are degraded under severe weather conditions (e.g., haze, low-lightness, and rain), these intelligent vision-based detection methods easily fail to accurately and robustly recognize the water-surface garbage, leading to unreliable water quality monitoring in maritime surveillance.

3. Intelligent Vision-Enabled Water-Surface Target Detection Framework

In this work, we mainly focus on detection of water-surface targets in vision-empowered maritime surveillance. An intelligent vision-enabled water-surface target detection framework with deep neural networks will be proposed. To enhance the accuracy and robustness of target detection, degraded images under different severe weather conditions will be synthetically generated through existing physical imaging models. These synthetically degraded images are naturally beneficial for improving the generalization abilities of our neural networks.

3.1. AI-Empowered Detection of Water-Surface Targets. We propose to develop the intelligent vision-enabled water-surface target detection framework based on the existing

CCTV system, which has been widely utilized in maritime video surveillance. With the great advancements of IoT technologies, where sensors and embedded equipment are connected to the Internet to efficiently gather and exchange maritime data [44], IoT-based maritime video surveillance has become increasingly more attractive to both academia and industry. Two typical deep neural networks, i.e., YOLOv4 [14] and Faster R-CNN [10], are incorporated into our learning-enabled detection framework to accurately monitor the water-surface garbage. To improve the generalization abilities of neural networks, a standard dataset, which contains several types of water-surface garbage, is designed to train our networks. However, the efficacy of water-surface garbage detection highly depends upon the imaging quality of CCTV. Under different weather conditions, e.g., haze, low-lightness, and rain, the observed images inevitably suffer from visibility degradation, leading to unsatisfactory detection results in practical applications. To guarantee reliable detection performance, we propose to develop two different strategies to enhance the proposed intelligent vision-enabled target detection framework, i.e.,

(1) It firstly selects the state-of-the-art visibility enhancement methods to improve the visual qualities of target images obtained under hazy, low-light, or rainy conditions. Both Faster R-CNN and YOLOv4, which are trained using the standard dataset only containing sharp images, are then adopted to automatically detect the water-surface garbage in visibility-enhanced images.

(2) To eliminate the negative effects of suboptimal enhanced images on garbage detection, the second strategy proposes to enlarge the existing standard dataset with different types of degraded images synthetically generated under different severe weather conditions. The enlarged dataset, which contains both sharp and degraded images, is beneficial for promoting the generalization abilities of our neural networks. The effectiveness and robustness of garbage detection under degraded visual environments could be enhanced accordingly.

Extensive experiments will be implemented in this work to compare these two strategies and select the optimal scenario. Once the harmful water-surface garbage is accurately detected, an automatic alarm device in our IoT-based maritime video surveillance system will automatically emit an alarm signal. The operators in charge will then perform the corresponding activities to reduce the negative effects of unwanted garbage on water quality. Therefore, the proposed intelligent vision-enabled water-surface target detection framework is capable of early detection of harmful pollution for real time monitoring water quality under different weather conditions.

3.2. Synthetically Degraded Images in Poor Weather Conditions. To enhance the learning capacities of our neural networks, the observed degraded images under different severe weather conditions will be synthetically generated

accordingly. In this work, the hazy, low-light, and rainy images can be simulated according to the following physical imaging principles.

3.2.1. Haze Imaging.

In the fields of image processing and computer vision, the atmospheric scattering model has been directly adopted to model the process of generating haze-degraded images. In particular, the observed degraded image $I(x) \in [0, 1]$ is obtained as follows:

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

where $J(x) \in [0, 1]$ is the real scene radiance, $t(x) \in (0, 1]$ represents the transmission map, $A \in (0, 1]$ is the global atmospheric light, and $x \in \Omega$ denotes the pixel index with Ω being image domain. Given the predefined J and A, the hazy images under different haze levels can be synthetically generated by manually changing the mappings of t. In particular, t can be theoretically generated according to the depth map d, i.e., $t = \exp(-\kappa d)$ with κ being a positive constant. It is able to directly measure the depth map d in practical applications.

3.2.2. Low-Light Imaging.

According to the Retinex theory, the observed low-light image $I(x) \in [0, 1]$ can be defined as follows:

$$I(x) = R(x) \circ L(x), \qquad (2)$$

where $\circ$ represents the element-wise multiplication operator and $R(x) \in [0, 1]$ and $L(x) \in [I(x), 1]$ denote the reflection and illumination maps, respectively. Theoretically, the visual quality of I is highly related to the magnitude of L with the existing high-quality image R. To simulate the low-light images, the latent sharp images are firstly transformed from RGB color space into HSV color space. By multiplying the V layer with different attenuation coefficients $\varpi \in (0, 1)$ in sharp images, we can accordingly generate the synthetically degraded images from severe to slight levels.

3.2.3. Rain Imaging.

The rain-degraded image $I(x) \in [0, 1]$ can be considered as the combination of rain-free background $B(x) \in [0, 1]$ and rain streak layer $S(x) \in [0, 1]$. The rainy image I can thus be synthetically expressed as follows:

$$I(x) = B(x) + S(x), \qquad (3)$$

where the rain-free scene B is commonly reconstructed by estimating the unwanted rain streaks S from the rainy version I. Analogous to the generation of rain streaks in [45], we introduce both salt-and-pepper noise and motion blur to synthetically simulate the rain-degraded images. The degraded images under different rain conditions are highly related to the levels of random noise [46] and types of motion blurs.

3.3. One-Stage-Based Detection Framework.

The object detector is an important part of our water-surface target detection framework. To improve detection accuracy and robustness, it is necessary to employ the reliable object detector. In the literature, deep network-based object detectors can be classified into two categorizes: (i) one-stage detectors and (ii) two-stage detectors. As an emerging one-stage-based object detection method, YOLOv4 is now receiving an increasing attention from both academia and industry. We will provide a brief overview of YOLOv4 used in our target detection framework in this subsection.

To balance the trade-off between detection accuracy and efficiency, Redmon et al. [11] originally proposed a one-stage detector called YOLO in 2016. It divides the target image into several partially overlapping regions of different sizes and predicts the bounding boxes and probabilities for each region. The object identification and bounding box extraction are jointly implemented by formulating object detection as a special regression problem. To further enhance detection performance, YOLOv2 [12] and YOLOv3 [13] were then proposed with improved precision and speed. More recently, a more powerful detection framework, named YOLOv4 [14], was presented, which combined several advanced strategies in different aspects. As observed in Figure 1, we choose CSPDarknet53 as the backbone of YOLOv4, which could promote the ability of learning invariable features. In particular, YOLOv4 utilizes Weighted-Residual-Connections (WRCs) and Cross-Stage-Partial-Connections (CSPC), takes into account Cross mini-Batch Normalization (CmBN), DropBlock regularization, and Mish-activation, and performs Self-Adversarial-Training (SAT) and Mosaic data augmentation to train detection network. In literature [47, 48], YOLOv4 has achieved superior performance in different datasets. This state-of-the-art detection results benefit from the powerful "Bag-of-Freebies" and "Bag-of-Specials" detection strategies during the real time implementation. There is thus a great potential to adopt YOLOv4 to real time detection of water-surface garbage in existing maritime video surveillance systems. For more details on YOLOv4, please refer to [14] and references therein.

3.4. Two-Stage-Based Detection Framework.

Current studies have demonstrated that two-stage detectors could produce higher detection accuracy compared with traditional one-stage detectors [49]. However, the superior accuracy was achieved at the expense of high computation, leading to slow inference speed. Among the two-stage-based detection methods, Faster R-CNN [10] has received tremendous attention from both academia and industry. It mainly consists of two modules. In particular, the first module is a fully convolutional region proposal network (RPN) which proposes candidate regions. The second module is the Fast R-CNN [9] detector which refines the proposed regions. Note that both RPN and Fast R-CNN share the same convolutional layers, resulting in joint training in practice. It is worth mentioning that RPN can simultaneously predict the object boundary and object score at each location. Furthermore, RPN accepts images (of any size) as input and obtains a set of rectangular object suggestions,
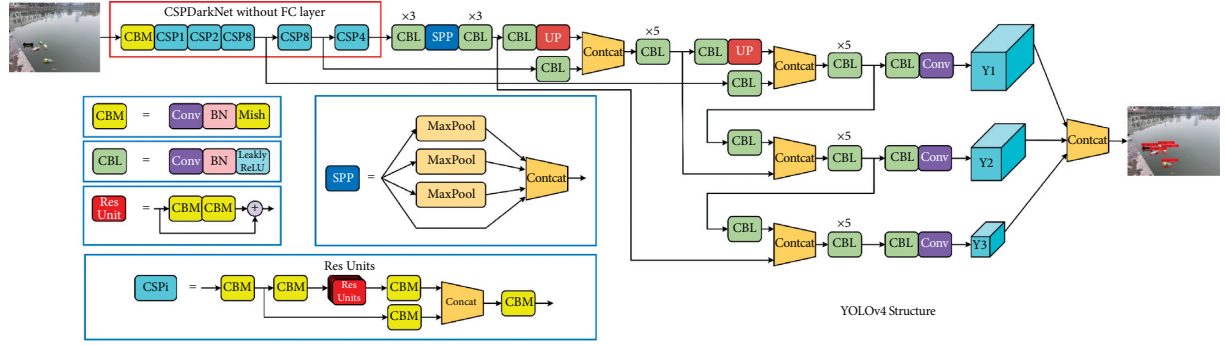
FIGURE 1: The garbage dataset, which contains 2000 color images in different scales, is developed for deep learning-based automatic detection of water-surface garbage. This current dataset is mainly composed of glass and plastic materials. More types of water-surface garbage will be incorporated into our dataset in the future work.

each of which has an objectivity score. The Faster R-CNN adopts a fully convolutional network to learn this process and guarantees end-to-end detector training on the shared convolutional features [10]. Owing to the sharing of convolutional features, it is tractable to adopt the very deep networks to yield high-quality detection results. Figure 2 visually displays the entire architecture of Faster R-CNN for water-surface garbage detection. For more implementation details on Faster R-CNN, please refer to [10] and references therein.

4. Experimental Results and Analysis

This section is mainly devoted to demonstrate the efficacy of our proposed framework on water-surface target detection under different weather conditions. Several objective criteria will be simultaneously selected to quantitatively evaluate the detection accuracy and robustness.

4.1. Experimental Settings and Dataset

4.1.1. Experimental Environment Setup.
The network training of Faster R-CNN was performed using TensorFlow 1.9.0 on a machine with Intel®CoreTM i7-8700K a 3.70 GHz × 12. The detection experiments were conducted using CUDA 9.0 and cuDNN 7.0.5 in Ubuntu18.04-based TensorFlow environment. In contrast, YOLOv4 was trained using PyTorch 1.5.1, which is a library specially developed for the deep learning model.

4.1.2. Experimental Dataset.
The original dataset adopted in our experiments contains 2000 clean maritime images collected by ourselves and downloaded from the Internet. Several typical images in our garbage dataset are visually illustrated in Figure 3. This developed dataset is mainly composed of glass and plastic materials. We manually operated the garbage-type labels and high-precision bounding boxes to annotate these experimental images. We further augmented each training image by flipping horizontally, translating randomly, and cropping randomly, which makes the training dataset 3 times larger than the original version. These images will be adopted to evaluate

the performance of our intelligent vision-enabled garbage detection method for IoT-based maritime video surveillance.

4.2. Evaluation Criteria.
To evaluate the performance of water-surface target detection under different weather conditions, we propose to simultaneously adopt several evaluation criteria in this work. The quantitative evaluation indexes mainly include the intersection over union (IoU), Precision and Recall, mean average precision (mAP), and frames per second (FPS).

4.2.1. Intersection over Union (IoU).
To objectively evaluate object detectors (i.e., YOLOv4 and Faster R-CNN in this work), IoU has become the most popular metric to compare the similarity between any two arbitrary shapes [50]. The IoU is mathematically defined as follows:

$$\text{IoU} = \frac{|A \cap B|}{|A \cup B|}, \tag{4}$$

where A and B, respectively, denote the prediction and ground-truth bounding boxes. In particular, IoU has the attractive property of scale invariance. It means that it takes into consideration the width, height, and location of the bounding boxes. Because of this appealing property, IoU has also been successfully utilized in other tasks, e.g., segmentation [51] and object tracking [52].

4.2.2. Precision and Recall.
Both Precision and Recall are adopted to measure the target detection results in this work. In particular, Precision indicates the proportion of positive detection which is actually correctly obtained through detection methods. In contrast, Recall is the proportion of actual positives which are correctly detected. To calculate both Precision and Recall, the true positive (TP), true negative (TN), false positive (FP), and false negative (FN) should be considered in practice. Mathematically, the definitions of Precision and Recall are given by

FIGURE 2: The garbage dataset, which contains 2000 color images in different scales, is developed for deep learning-based automatic detection of water-surface garbage. This current dataset is mainly composed of glass and plastic materials. More types of water-surface garbage will be incorporated into our dataset in the future work.

FIGURE 3: The water-surface garbage dataset, which contains 2000 color images in different scales, is developed for deep learning-based automatic detection of water-surface garbage. This current dataset is mainly composed of glass and plastic materials. More types of water-surface garbage will be incorporated into our dataset in the future work.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}},$$
$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}. \tag{5}$$

4.2.3. F1 Score. Sometimes, these two indicators (i.e., Precision and Recall) cannot be directly adopted to accurately evaluate the detection performance. By combining both Precision and Recall, the calculation of *F*1 score is defined as follows:

$$\frac{2}{F1} = \frac{1}{P} + \frac{1}{R}, \tag{6}$$

where *P* and *R* denote the Precision and Recall scores, respectively. In particular, the *F*1 score is exploited as the evaluation index as follows:

$$F1 = \frac{2PR}{P + R}. \tag{7}$$

4.2.4. Mean Average Precision (mAP). The commonly used evaluation index in target detection is mean average precision (mAP), which indicates the average of average precision (AP) scores over all of the types. The AP can be calculated according to the area under the Precision-Recall curve. In particular, mAP is mathematically defined as follows:

$$\text{mAP} = \frac{\sum_n (R_n - R_{n-1}) P_n}{N_c}, \tag{8}$$

where R_n and P_n, respectively, represent the Recall ratio and Precision ratio for the *n*-th preselected threshold. In addition, N_c indicates the number of target types, which is set to $N_c = 1$ in this work.

4.2.5. Frames per Second (FPS). FPS represents the number of image frames handled using a specific detection method in one second. It has been widely adopted to evaluate the detection speed in practice.

4.3. Experimental Results on Detection Accuracy. This subsection is devoted to evaluating the water-surface target detection results in terms of accuracy and robustness. In particular, both YOLOv4 and Faster R-CNN were trained directly using the original dataset. Five different objective criteria (i.e., $F1$, Recall, Precision, mAP, and FPS) are jointly adopted to quantitatively evaluate the detection accuracy and robustness. To further investigate the influences of different weather conditions on detection performance, the degraded images, synthetically generated under hazy, low-light, and rainy imaging conditions, are collected to measure the detection results. In particular, the degraded images are simulated according to the physical imaging models introduced in Section 3.2. The detailed evaluation results are summarized in Table 1. It can be observed that YOLOv4 is able to real timely implement target detection and generates slightly more accurate detection results compared with Faster R-CNN for clean images, i.e., under normal light conditions, shown in Figure 4. If the observed images are degraded due to poor imaging conditions, the detection accuracy will be obviously decreased for both YOLOv4 and Faster R-CNN.

To further evaluate the detection results, Figure 5 shows the original images and degraded versions synthetically generated under hazy, low-light, and rainy imaging conditions, respectively. The visibility-degraded images theoretically lead to reduced detection accuracy. The experimental results are visually displayed in Figure 6. If the image degradation becomes more severe, the detection accuracy and robustness will be obviously decreased, leading to unreliable water quality monitoring. This phenomenon has been confirmed by the quantitative evaluation results in Table 1.

4.4. Influences of Visibility Enhancement on Detection Results. As discussed in Section 4.3, the detection accuracy and robustness are highly depended upon the visual image quality. (Note that both YOLOv4 and Faster R-CNN were trained directly using the sharp images from original dataset. Therefore, to ensure satisfactory detection results in this case, it is necessary to guarantee high-quality observed images in our IoT-based maritime video surveillance.) If the camera images are generated under poor weather conditions, it is essential to improve image quality using existing visibility enhancement methods.

4.4.1. Haze Removal Methods. To suppress the effects of haze degradation on detection results, three typical haze removal methods, i.e., dark channel prior (DCP) [53], multiscale convolutional neural networks (MSCNNs) [54], and all-in-one dehazing network (AOD-Net) [55], are introduced in this work to enhance image visibility. The popular DCP is based upon the assumption that most local patches in haze-free images contain some pixels with greatly small intensity at least one color channel. MSCNN firstly adopts the coarse-scale network to estimate the rough transmission and then utilizes the fine-scale network to refine the rough transmission and generate final dehazed images. In contrast, AOD-Net directly reconstructs the haze-free images using an end-to-end network.

Table 2 detailedly depicts the quantitative detection results for both YOLOv4 and Faster R-CNN based on restored images yielded by different haze removal methods. As can be observed, the detection performance has been obviously improved by enhancing the visual image quality. It means that dehazed images enable more accurate detection of water-surface garbage, leading to more reliable monitoring of water quality. The detection results are visually illustrated in Figure 7. Due to the negative effects of haze degradation, it is intractable to accurately detect some small-scale objects shown in Figures 7(d) and 7(e). In contrast, Figures 7(f) and 7(g) illustrate that the popular DCP-based dehazing method [53] is able to suppress the effect of haze, leading to satisfactory detection results for both YOLOv4 and Faster R-CNN. It means that the high-quality water-surface target detection could be guaranteed by combining visibility enhancement methods and deep networks under haze conditions.

4.4.2. Low-Light Image Enhancement Methods. The adaptive histogram equalization (AHE) [56], Retinex-Net [57], and probabilistic method for image enhancement (PMIE) [58] are, respectively, adopted to reconstruct the high-quality images from their low-light versions. In particular, AHE performs well in contrast enhancement. By introducing the Retinex theory, Retinex-Net develops Decom-Net and Enhance-Net networks for image decomposition and illumination adjustment, respectively. PMIE employs a linear domain representation to simultaneously estimate both illumination and reflectance components to reconstruct latent sharp images.

The water-surface target detection results under low-light conditions are visually illustrated in Figure 8. It is obvious that the low-lightness could also generate negative effects on detection results. If the low-light images are enhanced via PMIE [58], the visual image quality is significantly improved in Figures 8(f) and 8(g), leading to more satisfactory detection performance. The importance of visibility enhancement can be further confirmed by quantitative evaluation results in Table 3. The preliminary visibility enhancement methods (e.g., AHE, Retinex-Net, and PMIE) could contribute to high-quality water-surface target detection. Compared with the detection results from original sharp images in Figures 8(b) and 8(c), the enhanced scenarios are able to achieve the comparable results, potentially leading to reliable water quality monitoring under low-light conditions.

4.4.3. Rain Removal Methods. To effectively remove the unwanted rain streaks, three state-of-the-art methods, i.e., lightweight pyramid networks (LP-Net) [59],

TABLE 1: The influences of different weather conditions on water-surface target detection. Several objective criteria (i.e., $F1$, Recall, Precision, mAP, and FPS) are simultaneously adopted to quantitatively evaluate the detection results.

Detection methods	Weather	$F1$	Recall	Precision	mAP	FPS
YOLOv4 [14]	Clean	0.94	0.8971	0.9937	0.9548	27
	Haze	0.90	0.8229	1.0000	0.9069	27
	Low-light	0.70	0.5429	0.9694	0.6529	27
	Rain	0.85	0.7543	0.9778	0.8792	27
Faster R-CNN [10]	Clean	0.93	0.9020	0.9910	0.9439	5
	Haze	0.91	0.8370	0.9911	0.9122	5
	Low-light	0.73	0.5563	0.9721	0.6735	5
	Rain	0.84	0.7426	0.9712	0.8674	5

FIGURE 4: The water-surface garbage detection results under normal light conditions. Both YOLOv4 and Faster R-CNN are capable of generating satisfactory detection performance. In contrast, Faster R-CNN only fails to implement small-scale target detection.

FIGURE 5: The synthetically degraded images selected to evaluate the detection performance. From top to bottom: original images and degraded versions generated under hazy, low-light, and rainy imaging conditions, respectively.

directional gradient-guided constraints-based model (DiG-CoM) [60], and multiscale progressive fusion network (MSPFN) [61], are simultaneously introduced to improve image quality. In particular, LP-Net adopts the mature Gaussian–Laplacian image pyramid strategy to simplify image deraining. By taking into consideration the directional gradient operator, DiG-CoM performs well in efficiently extracting rain streaks from rainy images, leading to visual quality improvement. MSPFN enhances image quality by fully exploiting the pyramid representation to collaboratively model the rain streaks from multiple scales.
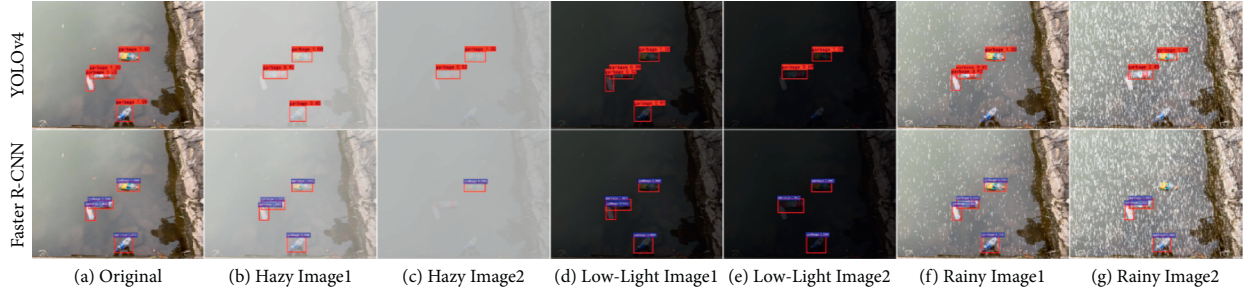
FIGURE 6: The water-surface garbage detection results under different weather conditions, i.e., hazy, low-light, and rainy conditions. It can be found that the detection accuracy will be dramatically reduced if the image degradation becomes more severe.

TABLE 2: The influences of hazy imaging condition and visibility enhancement on water-surface garbage detection.

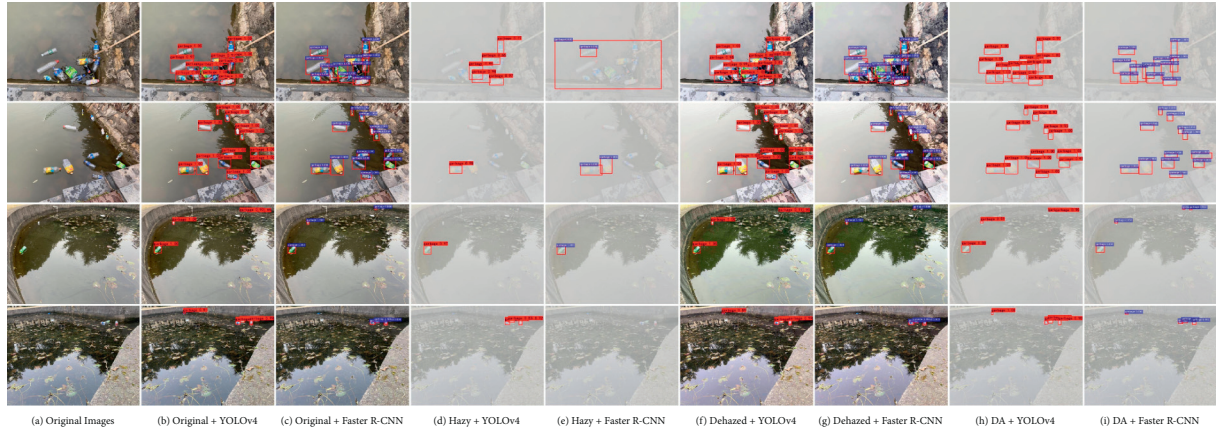| Detection methods	Enhancement methods	$F1$	Recall	Precision	mAP
YOLOv4 [14]	Hazy	0.90	0.8229	1.0000	0.9069
	DCP [53]	0.92	0.8433	0.9813	0.9237
	MSCNN [54]	0.93	0.8511	0.9912	0.9321
	AOD-Net [55]	0.93	0.8534	0.9907	0.9385
	Data augmentation	0.93	0.8729	0.9928	0.9509
Faster R-CNN [10]	Hazy	0.91	0.8370	0.9911	0.9122
	DCP [53]	0.92	0.8450	0.9872	0.9244
	MSCNN [54]	0.92	0.8620	0.9889	0.9276
	AOD-Net [55]	0.93	0.8750	0.9903	0.9322
	Data augmentation	0.94	0.8970	0.9913	0.9533

FIGURE 7: The water-surface garbage detection results under hazy conditions. From left to right: (a) original sharp images, original dataset-trained YOLOv4, and Faster R-CNN for detection results from (b-c) sharp images, (d-e) synthetic hazy images, (f-g) dehazed images, augmented dataset-trained YOLOv4, and Faster R-CNN for detection results from (h-i) synthetic hazy images, respectively. (a) Original images. (b) Original + YOLOv4. (c) Original + Faster R-CNN. (d) Hazy + YOLOv4. (e) Hazy + Faster R-CNN. (f) Dehazed + YOLOv4. (g) Dehazed + Faster R-CNN. (h) DA + YOLOv4. (i) DA + Faster R-CNN.

To evaluate the influences of visibility enhancement on target detection, the quantitative and qualitative detection results are indicated in Table 4 and Figure 9. The quantitative evaluations illustrate that rain removal methods are capable of improving detection accuracy and robustness under rainy conditions. However, it is inevitable that derained images easily suffer from loss of some fine details, unfortunately leading to suboptimal detection performance, shown in Figures 9(f) and 9(g). To guarantee reliable water-surface target detection, it is thus necessary to generate high-quality derained images. There are almost no

deraining methods adopted to adequately remove rain streaks while preserving all fine details. In this work, we will propose to adopt the widely used data augmentation (DA) strategy to retrain both YOLOv4 and Faster R-CNN to effectively improve detection results under different severe weather conditions.

4.5. Influences of Data Augmentation on Detection Results. The water-surface target detection methods introduced in Section 4.4 are essentially two-phase detection strategies, i.e.,

(a) Original Images (b) Original + YOLOv4 (c) Original + Faster R-CNN (d) Low-Light + YOLOv4 (e) Low-Light + Faster R-CNN (f) Enhanced + YOLOv4 (g) Enhanced + Faster R-CNN (h) DA + YOLOv4 (i) DA + Faster R-CNN
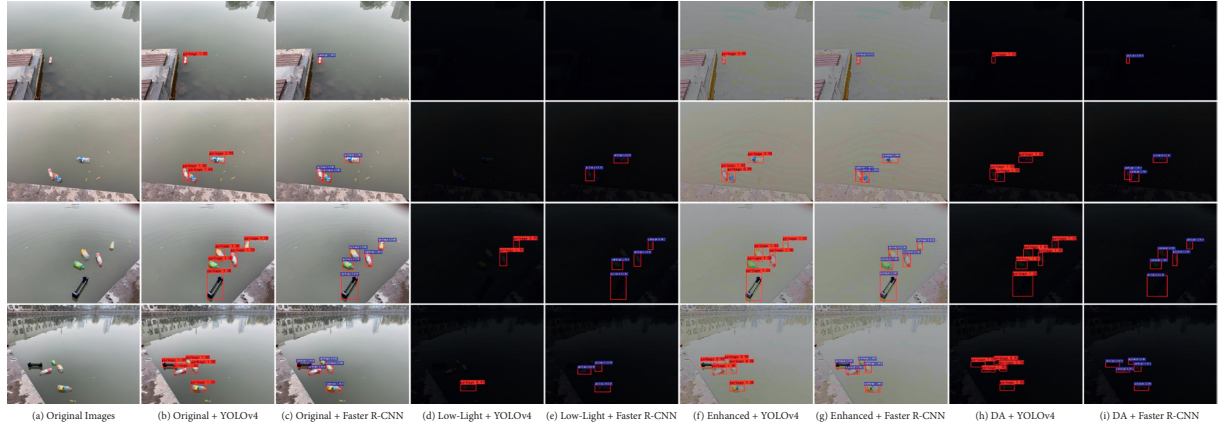
FIGURE 8: The water-surface garbage detection results under low-light conditions. From left to right: (a) original sharp images, original dataset-trained YOLOv4, and Faster R-CNN for detection results from (b-c) sharp images, (d-e) synthetic low-light images, (f-g) enhanced images, augmented dataset-trained YOLOv4, and Faster R-CNN for detection results from (h-i) synthetic low-light images, respectively. (a) Original images. (b) Original + YOLOv4. (c) Original + Faster R-CNN. (d) Low-light + YOLOv4. (e) Low-light + Faster R-CNN. (f) Enhanced + YOLOv4. (g) Enhanced + Faster R-CNN. (h) DA + YOLOv4. (i) DA + Faster R-CNN.

TABLE 3: The influences of low-light imaging condition and visibility enhancement on water-surface garbage detection.

Detection methods	Enhancement methods	$F1$	Recall	Precision	mAP
YOLOv4 [14]	Low-light	0.70	0.5429	0.9694	0.6529
	AHE [56]	0.77	0.6422	0.9722	0.8006
	Retinex-Net [57]	0.81	0.6934	0.9781	0.8255
	PMIE [58]	0.84	0.7351	0.9793	0.8737
	Data augmentation	0.92	0.8457	0.9844	0.9064
Faster R-CNN [10]	Low-light	0.73	0.5563	0.7210	0.6735
	AHE [56]	0.80	0.6799	0.9735	0.8125
	Retinex-Net [57]	0.83	0.7221	0.9741	0.8810
	PMIE [58]	0.86	0.7431	0.9788	0.8800
	Data augmentation	0.92	0.8511	0.9824	0.9123

TABLE 4: The influences of rainy imaging condition and visibility enhancement on water-surface garbage detection.

Detection methods	Enhancement methods	$F1$	Recall	Precision	mAP
YOLOv4 [14]	Rainy	0.85	0.7543	0.9778	0.8792
	LP-Net [59]	0.88	0.8127	0.9790	0.9134
	DiG-CoM [60]	0.90	0.8255	0.9811	0.9199
	MSPFN [61]	0.91	0.8429	0.9823	0.9234
	Data augmentation	0.92	0.8437	0.9837	0.9306
Faster R-CNN [10]	Rainy	0.84	0.7426	0.9712	0.8674
	LP-Net [59]	0.89	0.8245	0.9812	0.9321
	DiG-CoM [60]	0.91	0.8341	0.9836	0.9376
	MSPFN [61]	0.92	0.8625	0.9901	0.9340
	Data augmentation	0.93	0.8635	0.9920	0.9391

first visibility enhancement and then learning-based target detection. If both YOLOv4 and Faster R-CNN are trained using sharp images from the original dataset, the final detection results will thus depend on the qualities of enhanced images. However, it is almost impossible to perfectly reconstruct the high-quality maritime images, resulting in unsatisfied detection performance. In addition, this two-phase framework may suffer from long computational cost because the final total time is related to the computational costs of visibility enhancement and target detection. If we can directly and accurately detect the water-surface targets

under severe imaging conditions, it will real timely implement online water quality monitoring in practice. To achieve this goal, we will first synthetically simulate the degraded images according to the physical imaging models introduced in Section 3.2. The synthetically degraded images are collected to enlarge the existing standard dataset, which only contains original sharp images. This DA strategy could promote the volume and diversity of our training dataset, which is beneficial for enhancing the generalization abilities of our deep neural networks. The accuracy and robustness of target detection under different severe imaging conditions
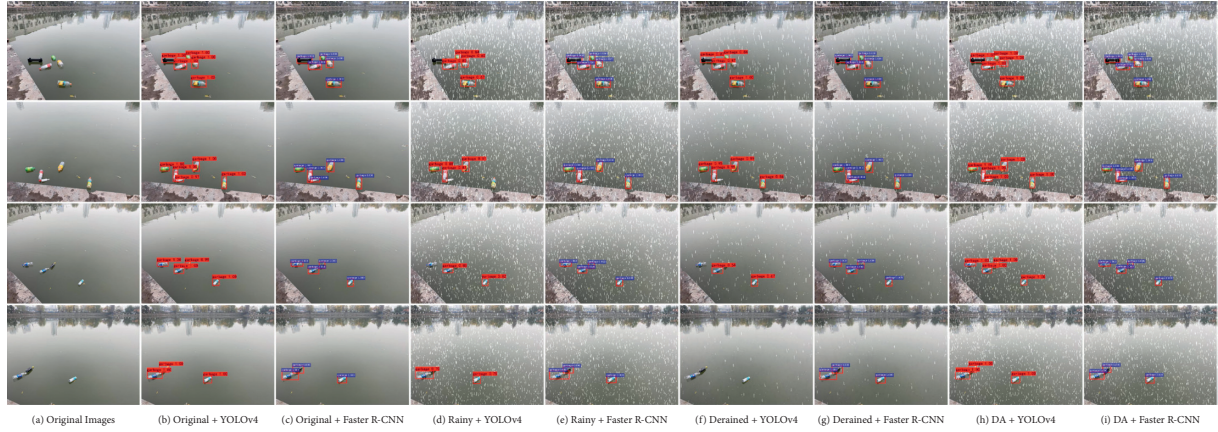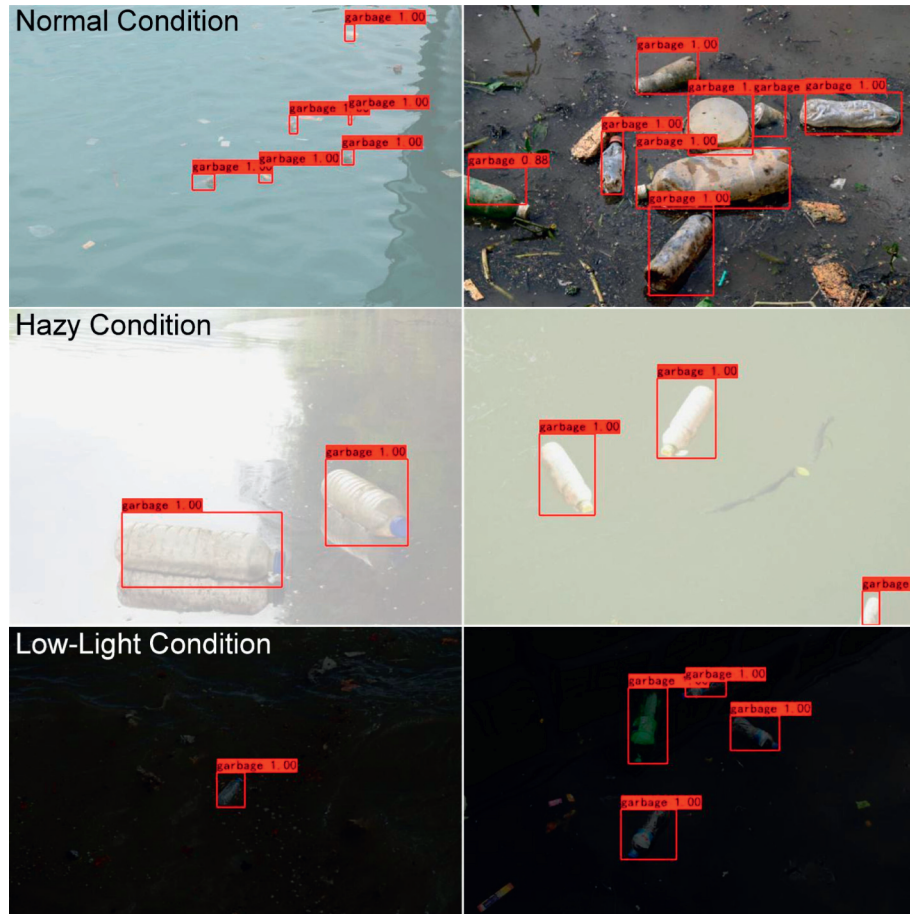
FIGURE 9: The water-surface garbage detection results under rainy conditions. From left to right: (a) original sharp images, original dataset-trained YOLOv4, and Faster R-CNN for detection results from (b-c) sharp images, (d-e) synthetic rainy images, (f-g) derained images, augmented dataset-trained YOLOv4, and Faster R-CNN for detection results from (h-i) synthetic rainy images, respectively. (a) Original images. (b) Original + YOLOv4. (c) Original + Faster R-CNN. (d) Rainy + YOLOv4. (e) Rainy + Faster R-CNN. (f) Derained + YOLOv4. (g) Derained + Faster R-CNN. (h) DA + YOLOv4. (i) DA + Faster R-CNN.

FIGURE 10: The robust detection results of our YOLOv4-based learning method in practical applications. From top to bottom: the water-surface garbage is detected from several different images obtained under normal, hazy, and low-light imaging conditions, respectively.

could be guaranteed accordingly. The experimental results are detailedly illustrated in Tables 2–4 and Figures 7–9. It can be found that DA strategy has significantly improved the representational capacities of YOLOv4 and Faster R-CNN.

The corresponding detection accuracy and robustness are comparable to, or even better than, the results obtained from the original sharp images. In particular, Faster R-CNN slightly outperforms YOLOv4 under consideration in most

of the cases. However, YOLOv4 is exploited for more efficient detection. To balance the trade-off between efficiency and accuracy, we propose to combine the DA strategy and YOLOv4 to directly detect water-surface targets from degraded images without visibility enhancement.

4.6. Experiments on Realistic Imaging-Degraded Conditions. To demonstrate the applicability of our method, we adopt the enlarged dataset, which contains original sharp images and synthetically degraded images, to train YOLOv4-based garbage detection framework. Figure 10 displays the detection results under different lighting conditions. It can be found that our IoT-based maritime video surveillance system can provide accurate and robust detection of water-surface garbage. Compared with traditional WSN or contact-type chemical sensors, our intelligent vision-enabled water quality monitoring framework is more flexible, convenient, robust, and low-cost. There is thus a huge potential to extend our intelligent framework for indirectly evaluating water quality in different water areas under different severe weather conditions.

4.7. Limitations and Future Studies. The proposed intelligent vision-enabled framework has the capacity of effectively and robustly detecting water-surface targets. However, it still suffers from some potential limitations, which constrain the further improvement of water pollution detection in maritime surveillance.

(1) The designed garbage dataset only contains two main types of pollution materials, which could constrain the detection of water-surface garbage in practice. To further improve the detection effectiveness and robustness, other types, e.g., paper, cardboard, metal, and trash, should also be considered in the future studies. The enlarged volume and diversity of training dataset are beneficial for improving the generalization abilities of neural networks, resulting in more accurate and robust water quality monitoring in maritime transportation.

(2) Both YOLOv4 and Faster R-CNN are not specifically developed for detection of water-surface garbage. Theoretically, this task is significantly different from other detection tasks, e.g., pedestrian, vessel, car, and animal detection. To further enhance the detection results, it is necessary to redesign and optimize these two neural networks according to the unique characteristics existing in water-surface targets.

Although the proposed detection framework has several limitations, it is still worthy of further investigation since it is able to achieve satisfactory detection results under severe weather conditions. The main contributions of this work show that there is a great potential for intelligent vision technique which tremendously improves water quality monitoring in maritime surveillance.

5. Conclusions

To conclude, we have proposed an intelligent vision-enabled target detection framework to automatically recognize water-surface garbage and make early warning in maritime transportation. It accordingly contributes to flexible and robust detection of harmful pollution in AI- and IoT-based maritime video surveillance. The major contributions of this paper were threefold. First, an intelligent vision-enabled water-surface target detection framework was developed to perform water quality monitoring. Second, we have designed a water-surface garbage dataset, which contains 2000 images collected by ourselves and downloaded from the Internet. A large number of synthetically degraded images were generated to further enlarge this dataset to improve the generalization abilities of our neural networks. Last, the proposed detection framework was capable of yielding timely, robust, and accurate garbage detection results. Numerous experiments on both synthetic and realistic scenarios have demonstrated the effectiveness and robustness of our water-surface target detection framework under different degraded visibility conditions. In addition, the water quality could be accordingly monitored with our intelligent maritime video surveillance.

Data Availability

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

Disclosure

Yongqi Guo and Yuxu Lu are co-first authors.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] M. M. Wang, J. Zhang, and X. You, "Machine-type communication for maritime internet of things: a design," *IEEE Communications Surveys & Tutorials*, vol. 22, no. 4, pp. 2550–2585, 2020.

[2] 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, no. 7, pp. 5374–5385, 2021.

[3] Y. Huang, Y. Li, Z. Zhang, and R. W. Liu, "GPU-accelerated compression and visualization of large-scale vessel trajectories in maritime IoT industries," *IEEE Internet of Things Journal*, vol. 7, no. 11, pp. 10794–10812, 2020.

[4] H. Zhou, Y. Yuan, and C. Shi, "Object tracking using SIFT features and mean shift," *Computer Vision and Image Understanding*, vol. 113, no. 3, pp. 345–352, 2009.

[5] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan, "Object detection with discriminatively trained part-based models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 9, pp. 1627–1645, 2009.

[6] C. F. Juang and G. C. Chen, "A TS fuzzy system learned through a support vector machine in principal component space for real-time object detection," *IEEE Transactions on Industrial Electronics*, vol. 59, no. 8, pp. 3309–3320, 2011.

[7] 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.

[8] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 580–587, Columbus, OH, USA, June 2014.

[9] R. Girshick, "Fast R-CNN," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1440–1448, Santiago, Chile, December 2015.

[10] 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.

[11] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: unified, real-time object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 779–788, Las Vegas, NV, USA, June 2016.

[12] J. Redmon and A. Farhadi, "YOLO9000: better, faster, stronger," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7263–7271, Honolulu, HI, USA, July 2017.

[13] J. Redmon and A. Farhadi, "Yolov3: an incremental improvement," 2018, https://arxiv.org/abs/1804.02767.

[14] 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.

[15] X. Chen, Y. Yang, S. Wang et al., "Ship type recognition via a coarse-to-fine cascaded convolution neural network," *Journal of Navigation*, vol. 73, no. 4, pp. 813–832, 2020.

[16] R. W. Liu, W. Yuan, X. Chen, and Y. Lu, "An enhanced CNN-enabled learning method for promoting ship detection in maritime surveillance system," *Ocean Engineering*, vol. 235, Article ID 109435, 2021.

[17] W. Lu, J. Duan, Z. Qiu, Z. Pan, R. W. Liu, and L. Bai, "Implementation of high-order variational models made easy for image processing," *Mathematical Methods in the Applied Sciences*, vol. 39, no. 14, pp. 4208–4233, 2016.

[18] Y. Chen and D. Han, "Water quality monitoring in smart city: a pilot project," *Automation in Construction*, vol. 89, pp. 307–316, 2018.

[19] S. Behmel, M. Damour, R. Ludwig, and M. J. Rodriguez, "Water quality monitoring strategies-a review and future perspectives," *Science of the Total Environment*, vol. 571, pp. 1312–1329, 2016.

[20] K. S. Adu-Manu, C. Tapparello, W. Heinzelman, F. A. Katsriku, and J.-D. Abdulai, "Water quality monitoring using wireless sensor networks," *ACM Transactions on Sensor Networks*, vol. 13, no. 1, pp. 1–41, 2017.

[21] L. Sevgi, A. Ponsford, and H. C. Chan, "An integrated maritime surveillance system based on high-frequency surface-wave radars. 1. Theoretical background and numerical simulations," *IEEE Antennas and Propagation Magazine*, vol. 43, no. 4, pp. 28–43, 2001.

[22] D. D. Bloisi, F. Previtali, A. Pennisi et al., "Enhancing automatic maritime surveillance systems with visual information," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 4, pp. 824–833, 2016.

[23] H.-M. Hu, Q. Guo, J. Zheng, H. Wang, and B. Li, "Single image defogging based on illumination decomposition for visual maritime surveillance," *IEEE Transactions on Image Processing*, vol. 28, no. 6, pp. 2882–2897, 2019.

[24] M. Yang, X. Nie, and R. W. Liu, "Coarse-to-fine luminance estimation for low-light image enhancement in maritime video surveillance," in *Proceedings of the IEEE International Conference on Intelligent Transportation Systems*, pp. 299–304, Auckland, New Zealand, October 2019.

[25] Y. Guo, Y. Lu, R. W. Liu, M. Yang, and K. T. Chui, "Low-light image enhancement with regularized illumination optimization and deep noise suppression," *IEEE Access*, vol. 8, pp. 145297–145315, 2020.

[26] Y. Zhang, Q.-Z. Li, and F.-N. Zang, "Ship detection for visual maritime surveillance from non-stationary platforms," *Ocean Engineering*, vol. 141, pp. 53–63, 2017.

[27] 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 Transactions on Intelligent Transportation Systems*, vol. 18, no. 8, pp. 1993–2016, 2017.

[28] Y. Duan, Z. Li, X. Tao, Q. Li, S. Hu, and J. Lu, "EEG-based maritime object detection for IoT-driven surveillance systems in smart ocean," *IEEE Internet of Things Journal*, vol. 7, no. 10, pp. 9678–9687, 2020.

[29] J. Zhang and D. Tao, "Empowering things with intelligence: a survey of the progress, challenges, and opportunities in artificial intelligence of things," *IEEE Internet of Things Journal*, vol. 8, no. 10, pp. 7789–7817, 2021.

[30] Y. Mehmood, F. Ahmad, I. Yaqoob, A. Adnane, M. Imran, and S. Guizani, "Internet-of-things-based smart cities: recent advances and challenges," *IEEE Communications Magazine*, vol. 55, no. 9, pp. 16–24, 2017.

[31] T. Yang, J. Chen, and N. Zhang, "AI-empowered maritime internet of things: a parallel-network-driven approach," *IEEE Network*, vol. 34, no. 5, pp. 54–59, 2020.

[32] D. Palma, "Enabling the maritime Internet of things: CoAP and 6LoWPAN performance over VHF links," *IEEE Internet of Things Journal*, vol. 5, no. 6, pp. 5205–5212, 2018.

[33] T. Xia, M. M. Wang, J. Zhang, and L. Wang, "Maritime Internet of things: challenges and solutions," *IEEE Wireless Communications*, vol. 27, no. 2, pp. 188–196, 2020.

[34] C. Sicard, C. Glen, B. Aubie et al., "Tools for water quality monitoring and mapping using paper-based sensors and cell phones," *Water Research*, vol. 70, pp. 360–369, 2015.

[35] W.-Y. Chung and J.-H. Yoo, "Remote water quality monitoring in wide area," *Sensors and Actuators B: Chemical*, vol. 217, pp. 51–57, 2015.

[36] F. Adamo, F. Attivissimo, C. G. C. Carducci, and A. M. L. Lanzolla, "A smart sensor network for sea water quality monitoring," *IEEE Sensors Journal*, vol. 15, no. 5, pp. 2514–2522, 2014.

[37] J. Dong, G. Wang, H. Yan, J. Xu, and X. Zhang, "A survey of smart water quality monitoring system," *Environmental Science and Pollution Research*, vol. 22, no. 7, pp. 4893–4906, 2015.

[38] M. Karydis and D. Kitsiou, "Marine water quality monitoring: a review," *Marine Pollution Bulletin*, vol. 77, no. 1-2, pp. 23–36, 2013.

[39] C. Serra-Toro, R. Montoliu, V. J. Traver, I. M. Hurtado-Melgar, M. Nunez-Redo, and P. Cascales, "Assessing water quality by video monitoring fish swimming behavior," in *Proceedings of the International Conference on Pattern Recognition*, pp. 428–431, Istanbul, Turkey, August 2010.

[40] H. Panwar, P. K. Gupta, M. K. Siddiqui et al., "AquaVision: automating the detection of waste in water bodies using deep transfer learning," *Case Studies in Chemical and Environmental Engineering*, vol. 2, Article ID 100026, 2020.

[41] P. F. Proença and P. Simões, "TACO: trash annotations in context for litter detection," 2020, https://arxiv.org/abs/2003.06975.

[42] X. Li, M. Tian, S. Kong, L. Wu, and J. Yu, "A modified YOLOv3 detection method for vision-based water surface garbage capture robot," *International Journal of Advanced Robotic Systems*, vol. 17, no. 3, Article ID 1729881420932715, 2020.

[43] S. Kong, M. Tian, C. Qiu, Z. Wu, and J. Yu, "IWSCR: an intelligent water surface cleaner robot for collecting floating garbage," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, pp. 1–11, 2020.

[44] Y. Yang, M. Zhong, H. Yao, F. Yu, X. Fu, and O. Postolache, "Internet of things for smart ports: technologies and challenges," *IEEE Instrumentation & Measurement Magazine*, vol. 21, no. 1, pp. 34–43, 2018.

[45] T.-X. Jiang, T.-Z. Huang, X.-L. Zhao, L.-J. Deng, and Y. Wang, "A novel tensor-based video rain streaks removal approach via utilizing discriminatively intrinsic priors," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4057–4066, Honolulu, HI, USA, July 2017.

[46] R. W. Liu, L. Shi, W. Huang, J. Xu, S. C. H. Yu, and D. Wang, "Generalized total variation-based MRI rician denoising model with spatially adaptive regularization parameters," *Magnetic Resonance Imaging*, vol. 32, no. 6, pp. 702–720, 2014.

[47] G. Du, K. Wang, S. Lian, and K. Zhao, "Vision-based robotic grasping from object localization, object pose estimation to grasp estimation for parallel grippers: a review," *Artificial Intelligence Review*, vol. 54, no. 3, pp. 1677–1734, 2021.

[48] W. Jiang, M. Liu, Y. Peng, L. Wu, and Y. Wang, "HDCB-Net: a neural network with the hybrid dilated convolution for pixel-level crack detection on concrete bridges," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 8, pp. 5485–5494, 2021.

[49] L. Liu, W. Ouyang, X. Wang et al., "Deep learning for generic object detection: a survey," *International Journal of Computer Vision*, vol. 128, no. 2, pp. 261–318, 2020.

[50] H. Rezatofighi, N. Tsoi, J. Gwak, A. Sadeghian, I. Reid, and S. Savarese, "Generalized intersection over union: a metric and a loss for bounding box regression," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 658–666, Long Beach, CA, USA, June 2019.

[51] H. Abu Alhaija, S. K. Mustikovela, L. Mescheder, A. Geiger, and C. Rother, "Augmented reality meets computer vision: efficient data generation for urban driving scenes," *International Journal of Computer Vision*, vol. 126, no. 9, pp. 961–972, 2018.

[52] W. Feng, R. Han, Q. Guo, J. Zhu, and S. Wang, "Dynamic saliency-aware regularization for correlation filter-based object tracking," *IEEE Transactions on Image Processing*, vol. 28, no. 7, pp. 3232–3245, 2019.

[53] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 12, pp. 2341–2353, 2010.

[54] W. Ren, J. Pan, H. Zhang, X. Cao, and M.-H. Yang, "Single image dehazing via multi-scale convolutional neural networks with holistic edges," *International Journal of Computer Vision*, vol. 128, no. 1, pp. 240–259, 2020.

[55] 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.

[56] S. M. Pizer, E. P. Amburn, J. D. Austin et al., "Adaptive histogram equalization and its variations," *Computer Vision, Graphics, and Image Processing*, vol. 39, no. 3, pp. 355–368, 1987.

[57] C. Wei, W. Wang, W. Yang, and J. Liu, "Deep retinex decomposition for low-light enhancement," in *Proceedings of the British Machine Vision Conference*, pp. 1–12, Newcastle upon Tyne, UK, September 2018.

[58] X. Fu, Y. Liao, D. Zeng, Y. Huang, X.-P. Zhang, and X. Ding, "A probabilistic method for image enhancement with simultaneous illumination and reflectance estimation," *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 4965–4977, 2015.

[59] X. Fu, B. Liang, Y. Huang, X. Ding, and J. Paisley, "Lightweight pyramid networks for image deraining," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 6, pp. 1794–1807, 2020.

[60] W. Ran, Y. Yang, and H. Lu, "Single image rain removal boosting via directional gradient," in *Proceedings of the IEEE International Conference on Multimedia and Expo*, pp. 1–6, London, UK, July 2020.

[61] K. Jiang, Z. Wang, P. Yi et al., "Multi-scale progressive fusion network for single image deraining," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 8346–8355, Seattle, WA, USA, June 2020.