

Review Article

Survey on Deep Learning-Based Marine Object Detection

Ruolan Zhang , Shaoxi Li , Guanfeng Ji, Xiuping Zhao, Jing Li, and Mingyang Pan 

Navigation College, Dalian Maritime University, Dalian 116026, China

Correspondence should be addressed to Shaoxi Li; lishaoxi@dmlu.edu.cn

Received 17 July 2021; Accepted 4 October 2021; Published 25 November 2021

Academic Editor: Chunjia Han

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

We present a survey on marine object detection based on deep neural network approaches, which are state-of-the-art approaches for the development of autonomous ship navigation, maritime surveillance, shipping management, and other intelligent transportation system applications in the future. The fundamental task of maritime transportation surveillance and autonomous ship navigation is to construct a reachable visual perception system that requires high efficiency and high accuracy of marine object detection. Therefore, high-performance deep learning-based algorithms and high-quality marine-related datasets need to be summarized. This survey focuses on summarizing the methods and application scenarios of maritime object detection, analyzes the characteristics of different marine-related datasets, highlights the marine detection application of the YOLO series model, and also discusses the current limitations of object detection based on deep learning and possible breakthrough directions. The large-scale, multisenario industrialized neural network training is an indispensable link to solve the practical application of marine object detection. A widely accepted and standardized large-scale marine object verification dataset should be proposed.

1. Introduction

Information technology and intelligent development have changed the operation mode and direction of many industries. Traditional maritime shipping industry also has gradually been advanced from digitization and informatization to intelligence [1]. As a major advance in machine learning over the last decades, the deep learning approach is becoming the most powerful technique for intelligent transportation system [2]. The deep learning methodologies are applied in various fields in the maritime industry such as ship classification, object detection, collision avoidance, risk perception, and anomaly detection. The main application directions can be summarized as maritime surveillance and autonomous ship navigation.

Currently, most of the research focuses on some aspects of the deep learning technique that performance much higher than humans; however, that technique is unable to complete complex tasks. So far, although seafarers have some limitations and many failure examples in the process of completing shipping transportation, humans are still the most reliable executors. Therefore, it is necessary to survey the applications of deep learning-based technologies in the

maritime field to explore how computer vision replaces or even surpasses humans in a real-world application, especially the object detection task, which has exploded in recent years, and most of the evaluation index has already made great progress.

Humans perceive the external objects' size, brightness, color, and movement state through their eyes; 80% of human perception information (situation awareness) is obtained through vision. With the limitation of the seafarer's lookout at the ship bridge, the visual perception of the horizon cannot be done excluding the solar direction [3] and bad weather often affects the seafarer's judgment. Most collisions and grounding are due to wrong interpretation or disregard of improper lookout (COLREGS-1972) [4].

Computer vision is an interdisciplinary scientific field that obtains and completes a series of image information processing from digital images or videos [5]. From the perspective of engineering, it seeks to perceive, understand, and automate tasks that the human visual system does.

Visual perception is an information-based approach to understanding biological and artificial vision [6]. It refers to the process of organizing, identifying, and interpreting visual information in environmental expression and

understanding. According to this definition, the goal of computer vision is to express and understand the environment. The core issue of visual perception is to study how to organize the input image information, identify objects and scenes, explain the content of the image.

A number of surveys of general object detection have been published in recent years. Zou et al. [7] reviewed more than 400 papers on the development of object detection technology from 1998 to 2018. This survey includes historical milestone detectors, detection datasets, measurement methods, and the latest detection methods. This article also reviews some important detection applications, such as pedestrian detection, face detection, and text detection, and conducts an in-depth analysis of the challenges and technological improvements in recent years. Jiao et al. [8] analyzed the existing typical object detection model and methods and discussed how to construct an effective and efficient system architecture based on the current detection model. Wu et al. [9] systematically analyzed the existing deep learning-based object detection frameworks and organized the survey into three major parts: (i) detection components, (ii) learning strategies, and (iii) applications and benchmarks. This survey covers a variety of elements affecting detection performance, such as detector architectures, feature learning, proposal generation, and sampling strategies. Chen et al. [10] analyzed the characteristics of the imbalance problem in different kinds of deep detectors and experimentally compared the performance of some state-of-the-art solutions on the COCO benchmark. Qiao et al. [11] combined the visual perception tasks required for maritime surveillance with those required for intelligent ship navigation to form a marine computer vision-based situational awareness complex and investigated the key technologies they have in common. This review focuses on the ship detection by the ship's own equipment and does not include the influence of other possible objects and backgrounds at sea, as well as the problems that may arise in industrial applications.

Computer vision-based marine object detection, as one of the most fundamental and challenging issues in maritime intelligent transportation, has received great attention over the last decades. As shown in Figure 1, it indicates the increasing number of publications in marine object detection from 2012 to 2021 July, the growing number of papers that their title is associated with "marine object detection" and "deep learning" over the past decades. The three advancements of digital data collection, computing power, and algorithm have promoted this deep learning research and application boom in maritime fields [12–15].

In recent years, deep learning-based visual perception has been widely applied to autonomous ship navigation and maritime transportation surveillance for intelligent transportation systems (ITS). The survey articles in the maritime field-related applications of computer vision are as follows: Qiao et al. [11] summarized the progress made in four aspects: full scene parsing of an image, ship reidentification, ship tracking, and multimodal data fusion with different visual sensors. Prasad et al. [16] provided a comprehensive overview of various approaches of video processing for object detection in the maritime environment. It consists of three modules: horizon detection, static background

subtraction, and foreground segmentation. Moniruzzaman et al. [17] described the use of deep learning for underwater imagery analysis, and deep learning architectures have been highlighted. Hashmani et al. [18] presented a survey on edge detection-based and machine learning-based marine horizon line detection; each study is presented with a recommendation for their suitability for a specific application in the marine environment.

Also, projection-based, region-based, hybrid, and artificial neural network (ANN) based methods for sea horizon detection have been discussed [19]. The researches of ANN methods in maritime surveillance made the horizon line detection easy, accurate, and robust. For optical remote sensing images applied in maritime, Li et al. [20] summarized the detection and classification of ship optical remote sensing images. Both methods were analyzed for traditional feature-designed methods and the deep convolutional neural networks (CNN).

The main difference between this paper and the above surveys is summarized as follows: (i) this paper only focuses on the task of deep learning-based object detection in computer vision. (ii) It analyzes the state-of-the-art of marine object detection in maritime surveillance, autonomous ship navigation, and other related applications. (iii) It analyzes and discusses the factors that affect the state-of-the-art solutions, especially the mainstream datasets and the milestone detectors.

As far as we know, the aim of this paper is to provide a survey of the most important approaches in the field of deep learning-based object detection for the maritime transportation system. This survey focuses on describing and analyzing deep learning-based marine object detection tasks. We contribute to the following:

Literature evaluation: we summarize the existing application scenarios of visual object detection in the maritime field. (2) Comparison of existing datasets. In practical engineering problems, big data plays an important role in realizing industrial applications. (3) Special emphasis on the role and development direction of visual detection in the autonomous ship navigation scenario. (4) Discussing the current limitations of object detection based on deep learning and possible breakthrough directions.

The rest of this paper is organized as follows. Section 2 highlights the state-of-the-art methods for general object detection. Section 3 introduces the application of object detection based on deep learning in various subdivisions of maritime affairs. The state-of-the-art backbone-based models and important datasets are described in Section 4.

2. State-of-the-Art of Object Detection

The definition of object detection is the task of detecting instances of targets of a certain class within an image or video.

Generally speaking, the detection task consists of two subtasks. One is the category information and probability of the target, and it is a classification task. The second is the specific location information of the target, which is a positioning task.

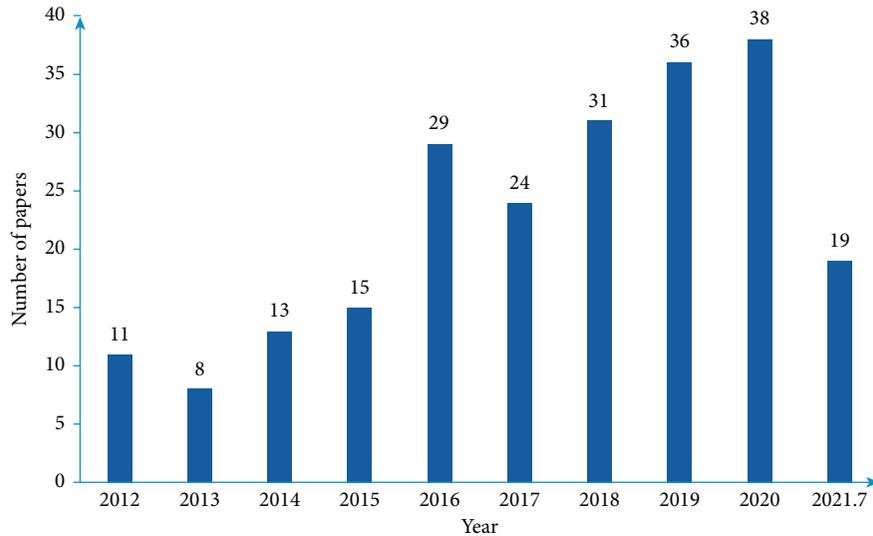


FIGURE 1: The increasing number of publications in marine object detection from 2012 to 2021 July (data from Google scholar advanced search).

As one of the most popular research fields of computer vision, object detection research prospered, which is the basic idea changed from traditional artificial feature design, shallow classifiers to deep neural network-based feature autonomous learning.

In the nondeep learning era, many tasks are not solved at once but require multiple steps, such as [21]. In the deep learning era, many tasks use the end-to-end framework, that is, input a picture and output the final result. The algorithm details and learning process are all completed through neural networks. It is particularly obvious in the field of object detection.

Under the deep learning architecture, whether it is a clear step-by-step process or the end-to-end method, the object detection algorithm must have three modules. The first is the selection of the detection window, the second is the extraction of image features, and the third is the design of the classifier.

As shown in Figure 2, the milestone of the neural network backbone and SOTA methods of object detection is listed in the timeline. 2012 was a critical point. Although CNN was proposed many years before, it was still hidden by other machine learning algorithms. After 2012, various neural networks and modules were combined, and deep learning-based methods suddenly left other methods behind. Deep learning-based application research in various fields can be widely carried out.

2.1. Review of Traditional Object Detection Methods. In 2001, [22, 23] proposed the Viola-Jones object detection framework. Based on the AdaBoost algorithm [24], the Viola-Jones framework uses Haar-like wavelet features and integral graph technology to perform face detection. This is the first detection method based on Haar + AdaBoost. It is also the first real-time framework for detection. Before the advent of deep learning technology, the Viola-Jones detector has always been the mainstream framework for face detection algorithms [25, 26].

Histogram of oriented gradient (HOG) [27] calculates the histogram not based on the color value but based on the gradient. It constructs the feature by calculating the gradient direction histogram of the local area of the image. HOG features combined with SVM classifiers have been widely used in image recognition, especially in pedestrian detection [28, 29]. Many further related researches have been presented, such as invariant histograms of oriented gradients (Ri-HOG) [30], which adopt annular spatial bins type cells and apply radial gradient transform (RGT) to attain gradient binning invariance for feature descriptors.

The DPM [31] algorithm adopts the detection ideas of improved HOG, SVM classifier, and sliding window. For the multiview problem of the target, it adopts the strategy of multicomponent. For the deformation problem of the target itself, it adopts the component model strategy of pictorial structure. DPM is a component-based detection method, which has strong robustness to the deformation of the target. At present, DPM has become the core of many classification, segmentation, pose estimation, and other deep learning-based algorithms [32–35].

In some specific application scenarios, object detection algorithms based on machine learning can still maintain good advantages. In [36], the image data were divided into smaller blocks and represented with a vector. These feature vectors are created by adding the subfeatures extracted from the color and texture properties of the images one after another. 99.62% classification success was achieved by using the Random Forest method. An average of 3.4 times acceleration was achieved by running each method on 1 master +4 workers clustering architecture on Apache Spark.

2.2. Deep Learning-Based Object Detection. CNN is one of the representative algorithms of deep learning [37]. It is the cornerstone of the current great success of deep learning, and it is a type of Feed forward Neural Networks (FNN) that includes convolution calculations and has a deep structure.

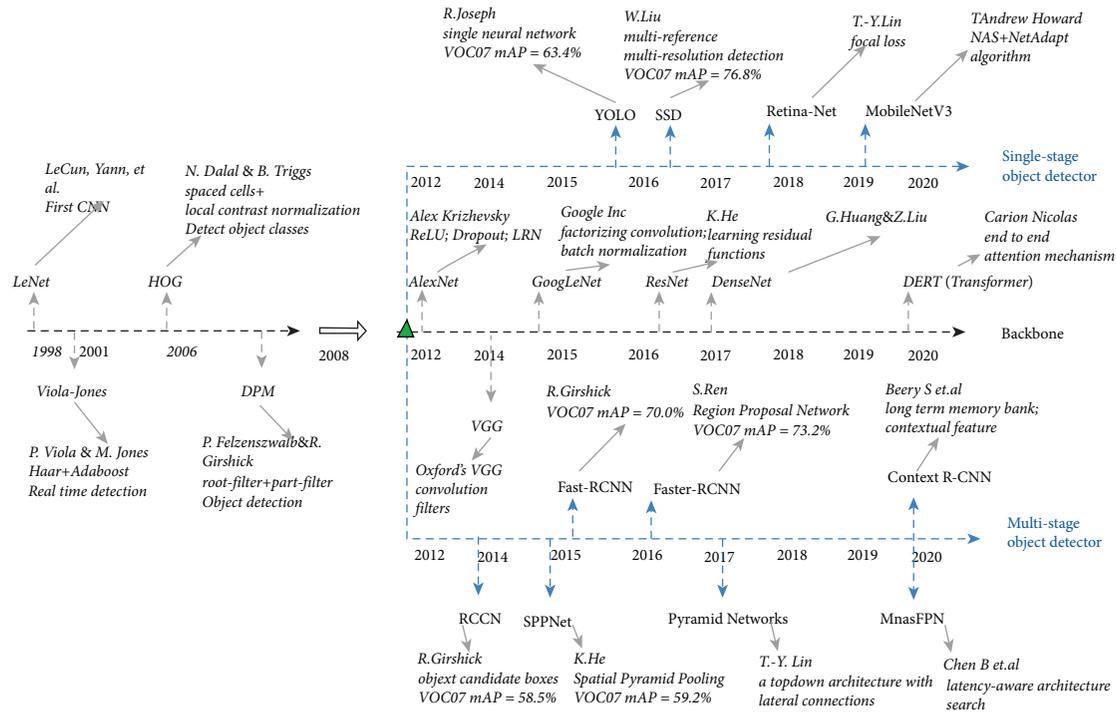


FIGURE 2: Deep learning-based visual perception development of neural network backbone and object detector.

CNN has the ability of representation learning, and it can perform shift-invariant classification of input information according to its hierarchical structure.

LeNet is one of the earliest CNN. Since 1988 [38], after many successful iterations, this pioneering result completed by Yann LeCun was named LeNet5. The architecture of LeNet5 is based on the view that the features of an image are distributed across the entire image, and the convolution of learnable parameters is an effective way to extract similar features in multiple locations with a small number of parameters.

In the nearly 20 years since LeNet was proposed, neural networks were once surpassed by other machine learning methods, such as support vector machines. Although LeNet can achieve good results on early small datasets, its performance on larger real datasets is not satisfactory. Computationally complex and insufficient computing power are the two main reasons for limiting its performance.

In 2012, Alex Krizhevsky proposed AlexNet [39]. Specifically, there are the following four innovations: (a) the GPU is used for network acceleration training for the first time. (b) ReLU activation function is used instead of the traditional sigmoid activation function and tanh activation function. (c) LRN local is used for response normalization. (d) In the first two layers of the fully connected layer, the dropout method is used to randomly inactivate neurons in a certain proportion to reduce overfitting.

AlexNet adds 3 convolutional layers on the basis of LeNet. VGG [40] proposed the idea of building a deep model by reusing simple basic blocks. The convolutional layers of vgg-block have the same structure, which means that the input size is equal to the output size. VGG proposed the idea of building a deep model by reusing a basic vgg-block. All VGG-block configurations are designed using the same

principles, the filter (kernel) adapted with a very small receptive field (3×3), the convolution stride is fixed to 1 pixel, the padding is used to maintain the image resolution after convolution, and the max-pooling is performed over a 2×2 pixel window, with stride 2. This design increased network depth to improve classification accuracy.

In 2014, GoogLeNet [41] proposed the inception network structure, which is to construct a “basic neuron” structure to build a network structure with sparseness and high computational performance. The following two innovations should be highlighted: (a) using factorization into small convolution can reduce the number of parameters, reduce overfitting, and increase the nonlinear expression ability; (b) using the Inception Module, multiple branches extract high-level features with different levels of abstraction, which can enrich the expressive ability.

In some practice, the training error tends to increase instead of decreasing after adding too many layers. Even if the numerical stability brought by batch normalization makes it easier to train deep models, the problem still exists. He et al. [42, 43] presented a residual block (ResNet) to solve this problem. The ResNet can train an effective deep neural network through the cross-layer data channel; it deeply influenced the design of later deep neural networks [44–47].

The cross-layer connection design in ResNet has led to several follow-up works, and DenseNet [48] is one of the representative innovations. The main building blocks of DenseNet are dense block and transition layer. The former defines how the input and output are connected, and the latter is used to control the number of channels so that it is not too large.

In the field of computer vision, CNN has always occupied the mainstream position. However, researchers continue to try to introduce the transformer model in the field of natural

language processing (NLP) into computer vision, propose a new Vision Transformer model, and achieve performance close to the current SOTA method on multiple image process benchmarks. DERT [49] demonstrated that the transformer model for NLP can also be used for image pretraining and object detection tasks. Han et al. [50] surveyed the research of transformer-based computer vision.

Deep learning-based object detection models still have to solve the three problems of region selection, feature extraction, and classification regression. Generally speaking, it can be divided into two categories: single-stage methods and multistage methods.

The multistage methods have high localization and object recognition accuracy, and the example models include R-CNN [51], SPPNet [52], fast R-CNN [53], faster R-CNN [54], mask R-CNN [55], and cascade R-CNN [56]. The R-CNN framework is a typical representative of the multistage method. It uses selective search to generate candidate regions and then the detection process, and the number of candidate windows is controlled at about 2000. After selecting these image frames, the corresponding frames can be resized and then sent to CNN for training. Due to the very powerful nonlinear characterization ability of CNN, it can perform good feature expressions for each region. The final output of CNN uses multiple classifiers for classification judgment. This method increases the detection rate on PASCAL VOC [57] from 35.1% to 53.7%, which is equivalent to AlexNet's breakthrough in classification tasks in 2012 and has a profound impact on the field of target detection. Subsequently, Fast R-CNN proposed RoI Pooling to select regional features from the convolutional feature map corresponding to the entire image, which solved the problem of repeated feature extraction. Faster R-CNN proposes region proposal, anchors divide the image into $n \times n$ regions, and each region gives 9 proposals with different ratios and scales, which solves the problem of repeatedly extracting candidate proposals. Other representative multistage object detectors also include SPPNet [52], pyramid networks [58], context R-CNN [59], and MnasFPN [60].

Single-stage methods prioritize inference speed, and example models include YOLO [61], SSD [32], RetinaNet [62], and MobileNetV3 [63]. YOLO is the representative single-stage model; there is no explicit bounding box extraction process. First, it resizes the image with a fixed size, divides the input image as a 7×7 grid, predicts 2 bounding boxes per grid, and classifies and locates for each bounding box. The YOLO model has also undergone many versions of development and is currently developed to YOLOv5. YOLO's approach is fast, but there will be many missing objects, especially tiny objects. So, single shot multibox detector (SSD) adds the concept of anchor from Faster R-CNN on the basis of YOLO and combines the features of different convolutional layers to make predictions. The main contribution of SSD is the multireference and multi-resolution detection techniques, which significantly improve the detection accuracy of a one-stage detector, especially for some tiny objects [64]. Although the methods of the YOLO and SSD series do not have the extraction of region proposals and it becomes faster, they inevitably lose information

and accuracy. The more representative single-stage object detector also includes RetinaNet [62] and MobileNet [63] series models.

In the field of computer vision, commonly used datasets include Microsoft Common Objects in Context (MSCOCO) [65], Karlsruhe Institute of Technology and Toyota Technological Institute (KITTI) [66], Visual Genome [67], Dataset for Object deTectioN in Aerial Images (DOTA) [68], and PASCAL Visual Object Classes Challenge (PASCAL VOC) [57]. The most popular object detection benchmark is the MSCOCO dataset. Models are typically evaluated according to a mean average precision metric.

In the field of marine object detection, there are few specialized datasets related to maritime supervision and autonomous ship navigation. Zhang et al. [69] proposed the use of generative adversarial networks (GANs) to solve insufficient marine data when training some object detection neural network. In [70], the novel idea is extracting the mask of the foreground object and combining it with the new background to automatically generate the location information and object information. Marine object detection-related dataset will be introduced in the subsequent section, separately.

2.3. Optimization Methods. The marine environment is complex and changeable, and the visual data has its own characteristics. Therefore, most of the researchers optimize the model and enhance the data based on the characteristics of the marine environment, to improve the accuracy and speed of marine object detection.

Chen et al. [71] presented a novel hybrid deep learning algorithm that combines improved generative adversarial network (GAN) and CNN-based detection methods for small ship detection. It uses Gaussian Mixture Wasserstein GAN with gradient penalty to generate sufficient informative artificial samples of small ships and uses raw and generated data to approach high accuracy tiny object detection. Ren et al. [72] proposed an effective ship image recognition method, which combines Hu invariant moment features and CNN features to achieve superior ship image recognition. Hu moment invariant feature joint to the last pooling layer achieves the highest recognition accuracy on self-built and VAIS datasets. Cao et al. [73] proposed a ship recognition method based on Morphological Watershed image segmentation and Zemike moment; although the Hu moment and Zernike moment are geometrically invariant, the Hu moment is unstable when the scale changes and the Zemike moment has a better stability. Using rotation to enhance the dataset causes errors in object detection tasks, Dong et al. [74] proposed a multiangle box-based rotation insensitive object detection structure (MRI-CNN) that improves the robustness of the model and reduces the detection performance impact due to the insufficient dataset.

3. Marine Target Detection Application

3.1. Maritime Surveillance. The interest in maritime surveillance has been increased in the last decades, and it is a significant issue for assuring the safety and security of

international transportation and defense mission. Despite being an important activity, how to efficiently conduct maritime surveillance is still a difficult problem for all countries. Computer vision-based digital maritime surveillance can solve most of this situation awareness issues, which can be divided into three categories: (1) detection and location (e.g., manmade pollution, oil spills, maritime hazardous event, noxious substances, and crashed plane debris), (2) tracking (e.g., ships, shipwrecks, lifeboats, illegal fisheries, illegal ballast water discharge, and smuggling), and (3) behavioral recognition (e.g., abnormal path confirmation, ships rendezvous, and high-speed objects on maritime surface). Most researchers focused on shore-based maritime surveillance, high-resolution satellite image surveillance, synthetic aperture radar (SAR) remote sensing, and so on.

3.1.1. Shore-Based Surveillance. In the current social environment, the traditional marine video surveillance technology simply relies on a large number of maritime managers who are no longer able to meet the needs of safe navigation. Computer vision combined with image processing technology has become the mainstream of maritime surveillance. In [75], experiments show that the ship detection based on YOLOv3 has high accuracy in the face of different scenes such as small traffic flow, foggy ship navigation, large traffic flow, and small imaging scale. The YOLOv3 algorithm uses the k -means algorithm to predict the bounding box and combines the multiscale features for ship identification; YOLOv3 can adapt to port scenarios with different traffic flow by using a multiscale detection mechanism, which has strong generalization ability.

Shao et al. [76] proposed a ship detection model based on a saliency-aware CNN framework that realizes real-time detection through the monitoring video taken by the camera. It can predict the category and position of the ship and use the global contrast-based salient region detection to correct the location. Based on the YOLOv2 pipeline, a saliency-aware CNN framework is proposed to improve the accuracy and robustness of ship detection under complex coastal conditions. Liu et al. [77] improved the YOLOv3 anchor method and feature fusion structure, respectively, GIOU loss was added to the loss function, and cross PANet was proposed to replace the FPN structure in YOLOv3. The results show that the proposed method can significantly improve the accuracy of YOLOv3 detecting sea surface objects. The SeaBuoys dataset was established according to the actual sea surface conditions, and comparative experiments were carried out with the existing SeaShips dataset [78].

Li et al. [79] proposed a new ship detection from visual image (SDVI) algorithm, named enhanced YOLOv3 tiny network for real-time ship detection. The convolution layer, instead of the max-pooling layer and expanding the channels of the prediction network, introduced an attention module named CBAM into the backbone network, which makes the model more focus on the target. The algorithm has a 9.6% improvement in mAP and has a faster detection speed.

Huang et al. [80] used k -means++ clustering on the dimensions of bounding boxes to prioritize the model, improve the YOLOv3-Darnet53 network, increase jump connection mechanism, decrease feature redundancy, and improve the ability of tiny ship detection. On the premise of ensuring real-time performance, the precision of ship identification is improved by 12.5%, and the recall rate is increased by 11.5%.

In [12], a “reference model” pretrained with Pascal VOC image dataset and a “proposed model” trained with a specific maritime dataset (Singapore Maritime Dataset, SMD), the same structure of the “reference model” compared with the “proposed model,” experiments show that, in SMD verification dataset, the proposed model is about twice as accurate as the reference model in terms of IoU and recall rate. Cane et al. [81] evaluated semantic segmentation networks in the context of an object detection system for maritime surveillance. The authors indicate that the SegNet and ENet achieve higher detection accuracy and precision. Considering the maritime surveillance actual condition, the ENet model would be the most suitable model.

3.1.2. High-Resolution Satellite Image Surveillance. High-resolution color remote sensing ship images taken from short distances provide advantages in ship detection applications. But the analysis of these high-dimensional images is complicated and requires a long time [36]. Synthetic aperture radar (SAR) is an active side-looking radar that can overcome weather interference and provide high-resolution images. SAR creates two-dimensional images or three-dimensional reconstructions of objects; it is typically mounted on a moving platform, such as an aircraft or spacecraft, and has its origins in an advanced form of side-looking airborne radar (SLAR).

Ghosh [82] proposed an efficient onboard detection system connected with a medium resolution wide amplitude optical camera and solved the problem of limited satellite coverage and limited simulation and equipment. Tian et al. [83] proposed a detection framework based on remote sensing image combining image enhancement module and dense feature reuse module to improve the object detection capability. Chen et al. [84] proposed an improved YOLOv3 based on an attention mechanism for fast and accurate ship detection, which accelerates detection speed to achieve real-time detection effect and improves the level of maritime surveillance.

Wang et al. [85] proposed an improved YOLOv3 algorithm for ship detection in optical remote sensing images. Adding the squeeze-and-excitation (SE) structure to the backbone improves the feature extraction capabilities and improves the detection accuracy by the fusion of multiscale feature maps. It achieves detection speeds of about 27 fps on NVIDIA RTX2080ti, with recall (R) = 95.32% and precision (P) = 95.62%. Cao et al. [86] conducted a similar study, the feature pyramid structure is introduced to combine the deep semantic information with the shallow semantic information, and the multiscale feature mapping is integrated to improve the detection ability of small objects.

Tang et al. [87] proposed a ship detection method based on noise classification and target extraction. The method consists of three modules: NLC (noising level classifying) module, STPAE (SAR target potential area extraction module) module, and the recognition module based on YOLOv5. The advantage of this model is that it can reduce the noise interference from the coast to ship detection. Tang et al. [88] introduced a novel high-resolution image network-based approach based on the preselection of a region of interest (RoI). It designs an HSV (hue, saturation, and value) module composed of four cores: background removal, noise removal, box-finding, and noise deletion, which can obtain useful RoI in a short time.

3.1.3. Airborne Maritime Surveillance. The speed of maritime surveillance ships is difficult to operate in complex seas and/or dispatching in busy ports. At present, although some maritime regulatory agencies carry manned helicopters, manned helicopters cannot take off to ensure the safety of personnel under hazardous sea conditions. At the same time, the cost of use is high; it is unable to meet the high-density, high-intensity maritime surveillance requirements.

Unmanned aerial vehicles (UAVs) can be remotely controlled or fly in the air in autonomous mode. It is a miniaturized and intelligent flight platform that can complete one or more tasks by carrying different task modules. It also has great potential in maritime applications. Solving the problems of drone flight stability, data transmission, ship-board electromagnetic compatibility, and convenient take-off and landing in the marine environment can enable drones to play a greater role in maritime applications. At present, according to the principle of flight, UAVs suitable for maritime applications mainly include fixed-wing UAVs, unmanned helicopters, multirotor UAVs, and vertical take-off and landing fixed-wing UAVs. Various types of UAVs have their own unique advantages in maritime applications.

Ribeiro et al. [89] presented an airborne maritime surveillance dataset captured by a small size UAV. This dataset presents object examples ranging from cargo ships, small boats, and life rafts to oil spill. Due to the continuous shaking of the UAV's camera, it is very difficult to label data on the acquired video dataset. The authors proposed a new labeling tool, which is developed in C++, and the OpenCV library is used to create labels manually. Reference [90] presents an approach to detect boats in a maritime surveillance scenario using a small UAV. This work relies on CNN to perform robust detection even in the presence of distractors like wave crests and sun glare. Reference [91] explores maritime search and rescue missions by using experimental UAV data to detect the sea surface object. Reference [92] addresses the development of an integrated system to support maritime situation awareness based on UAVs, emphasizing the role of the automatic detection subsystem.

Xiu et al. [93] contributed a system that includes a maritime unmanned aerial vehicle (Mar-UAV) with a high-resolution camera and an Automatic Identification System (AIS). Multifeature information, including position, scale,

heading, and speed, is used to match between real-time image and AIS message. The results demonstrate that the proposed algorithm and the Mar-UAV system are very significant for achieving autonomous maritime surveillance. Reference [94] presents a method to learn spatial and temporal features from video sequences; temporal features attempt to improve the maritime objects detection ability, which contain strong distractors such as glare and wakes. The proposed method is composed of two main parts, one spatial feature extractor based on the VGG network and one recurrent layer, the ConvLSTM.

3.1.4. Satellite Radar Image Surveillance. Ship detection in synthetic aperture radar (SAR) images has been widely studied due to its indispensable role in military intelligence acquisition, maritime management, and many civil fields. However, due to the limitations of bandwidth and computer computing power in satellite scenarios, SAR image-based ship detection deployment is largely hindered. Another reason is that searching for targets of interest in massive SAR images by eyes becomes time-consuming and often impractical. Therefore, lightweight neural network training models are widely used.

Chen et al. [95] proposed a novel learning scheme for training a lightweight ship detector called Tiny YOLO-Lite, which simultaneously (1) reduces the model storage size; (2) decreases the floating-point operations (FLOPs) calculation; and (3) guarantees the high accuracy with faster speed. Reference [96] proposes a lightweight CNN-LiraNet combining dense connections, residual connections, and group convolution. It uses a two-layer predictor and adds residual models to transmit features easier; experimental results show that the Lira-YOLO network has less complexity, only 2.980 Bflops. The parameters only have 4.3 MB. The mean average accuracy (mAP) index of the Mini-RD and SARShip detection dataset (SSDD) reaches 83.21% and 85.46%, respectively, which is comparable to the tiny-YOLOv3.

Yang et al. [97] proposed a one-stage object detection framework based on RetinaNet and rotatable bounding box (RBox) for the problems such as feature scale mismatch and task contradiction. Experimental results show that the average accuracy improved 13.26%, 9.29%, 8.92%, 8.55%, and 4.55% compared to the other four advanced RBox-based ship detection methods at the IoU threshold of 0.5. In this paper, scale calibration is proposed to make the proportion distribution of the main feature map and the object feature map consistent.

In arctic waters, a vast majority of objects are icebergs drifting in the ocean and can be mistaken for ships in terms of navigation and ocean surveillance. Hass and Jokar Arsanjani [98] presented a YOLOv3-based deep learning model that uses SAR images to discriminate icebergs and ships, which could be used for mapping ocean objects ahead of a journey.

To solve the problem of small objects and multiobject ship detection in complex scenarios, [99] proposes a detection method based on an optimized feature pyramid network (FPN) model. The results show that the small ship

detection accuracy reaches 98.62%, and the proposed model has higher accuracy and better comprehensive performance compared with YOLO.

3.1.5. Other Applications. For military defense and intelligent early warning, an infrared intrusion object detection algorithm based on a neural network is proposed [100]. The extended CNN designed by this algorithm can fuse and expand the image features, enhance object filtering, and improve background suppression. Xie et al. [101] proposed an inspection system based on tracking technology, which can automatically process ship inspection video and predict suspicious areas where cracks may exist. Intelligent computer vision is the most important technology for the development and utilization of deep-sea resources. Han et al. proposed a combination of the max-RGB method and shades of the gray method that is applied to achieve the enhancement of underwater vision [102]. In [103], vision-based object detection for underwater robots has been proposed. In order to overcome the limitations of cameras and to make use of the advantages of image data, a number of approaches have been tested. The topics include color restoration algorithm for the degraded underwater images, detection, and tracking methods for underwater target objects.

3.2. Vision-Based Autonomous Ship Navigation. Object detection and vision-based ship navigation is an essential task for autonomous ship navigation. However, sunlight reflection, camera motion, and illumination changes may cause false object detection in the maritime environment. Farahnakian and Heikkonen [104] proposed three fusion architectures (pixel-level, feature-level, and decision-level) to fuse two imaging modes (visible and infrared); they employed deep learning for performing fusion and detection. Pan et al. [105] proposed the navigation mark classification and identification model based on deep learning (RMA: ResNet-Multiscale-Attention), which can identify different navigation marks finely. It can identify the nuances of the navigation mark; no additional supervision information is required except for the label, and it is end-to-end training.

3.2.1. Horizon Detection. Marine horizon detection is the most significant semantic boundary for segmenting the image into sea and sky. References [18, 19] have summarized the marine horizon line detection. In the past research, many robust marine horizon generation methods have been proposed. For the marine horizon model of the straight line, the traditional methods include the following: (a) linear fitting: the selection of candidate points of this method is easily susceptible to the complex sea and sun glint [106]. (b) Image segmentation: the optimal segmentation threshold of this method is difficult to be adaptively determined [107]. The algorithm's anti-noise ability is insufficient. (c) Gradient significance, each interference factor has abundant edges, and these edges have gradient values similar to or even

higher than the marine horizon, which is easy to cause false detection [108].

Typical marine horizon detection relies on edge information, which requires two important issues to be overcome: unstable edge detection and complex marine environment with shore background and weather conditions. Jeong et al. [109] proposed a novel method for horizon detection that combines a multiscale approach and CNN; it has a median positional error (MPE) of less than 1.7 pixels from the center of the horizon and a median angular error (MAE) of approximately 0.1 degrees. This method is one of the methods for horizon detection with high speed and high accuracy, but it may have failed detection in some scenarios such as an absence of obvious line feature. Prasad et al. [110] presented a novel method called multiscale consistence of weighted edge Radon transform, abbreviated as MuSCoWERT. It has a median error of about 2 pixels (less than 0.2%) from the center of the actual horizon and a median angular error of less than 0.4 deg. Compared with traditional methods (ENIW [111], FGSL [112], MuSMF [113], IntGF, IntG, Hough [114], and GWR [115]), MuSCoWERT has excellent performance. Jeong et al. [116] proposed a fast method for detecting the horizon line in maritime scenarios by combining a multiscale approach and region-of-interest detection. Experimental results show that the proposed method can accurately identify the region of interest on the moving platform and ensure the robustness of sea-sky-line detection. And it is less affected by ships, light changes, waves, and wakes. In [117], a novel algorithm based on probability distribution and physical characteristics is introduced. The authors designed a hybrid method, which consists of sea-sky region extraction and horizon estimation based on the information of color, texture, and context. The proposed algorithm precisely detects the horizon not only from fine images but also from blurred image, even with a splashed camera.

As shown in Table 1, mean height deviation (MHD) and angle deviation (AD) as the evaluation standard have been recognized by most of the researchers. Nondeep learning methods occupy the majority of marine horizon detection. In recent years, marine horizon detection based on the Singapore Maritime Dataset has gradually increased, which is conducive to the comparison between different algorithms. Although it is still affected by the objective environment such as the different computing power of computers, it will play a role in pointing the direction of future research.

Accurate identification, tracking, and positioning of the marine horizon, as well as an accurate description of water boundary lines, are the basic requirements for safe driving of autonomous ship navigation. However, a large number of current researches mainly focus on pure marine horizon detection, and there is no in-depth research on marine horizon tracking and positioning and accurate description.

From the perspective of the marine horizon detection process, this survey summarizes and analyzes the key points of the existing marine horizon detection methods and summarizes the content that still needs to be studied in the future. It is suggested that, in complex water environment

TABLE 1: Examples of typical marine horizon detection algorithms.

| Methods | Dataset | Advantage | Disadvantage |
|---|--------------------------------|--|--|
| Zhang et al. [118] | MAV | Correct ratio >99.9%. | — |
| Lipschutz et al. [119], probability distribution edge detection Hough transform | — | Visible-light images: MHD ~3.69 pixels; AD ~0.28 degrees. Infrared images: MHD ~1.49 pixels; AD ~0.14 degrees | — |
| Gershikov et al. [114], H-REM | — | MHD ~2.28 pixels; AD ~0.19 degree; mean run time ~0.14 s | — |
| Prasad et al. [110], weighted edge Radon multiscale consistence | SMD MAR-DCT Buoy dataset | MPE ~2 pixels; MAE ~0.4 degrees | Does not work well in certain scenarios |
| Jeong et al. [116], multiscale approach region of interest (RoI) | SMD Buoy dataset | 15 fps; MPE <2 pixels; MAE ~0.15 degrees | Performance reduction (edges related to horizon) |
| Sun et al. [120], coarse-fine-stitched hybrid filtering Random sample | SMD Marine Obstacle | MHD ~0.89 pixels; MAD ~0.19 degrees | — |
| Liang et al. [117], probability distribution physical characteristics | SMD Buoy dataset | MPE ~7.6 pixels; MAE ~0.4 degrees | Ineffective (large area occlusion) |
| Jeong et al. [109], multiscale approach NN | SMD | MPE <1.7 pixels; MAE ~0.1 degrees | Line features absent |
| Yang et al. [121], probabilistic graphical expectation-max Gaussian models | Marine obstacle | — | Reflection and illumination |

and engineering applications with high real-time requirements, marine horizon detection is facing severe challenges. In future work, we should improve the algorithm to improve the real-time performance and environmental adaptability of the algorithm.

3.2.2. Surface Moving Object Detection. Over the last decades, a lot of researchers have worked on the big challenge of detection of moving ships in various complex marine environments. Reference [122] presents a ship object detection algorithm to achieve efficient visual maritime surveillance from nonstationary surface platforms.

The maritime target detection represented by YOLO has made great achievements in recent years. Chen et al. [123] proposed a YOLO-based integrated framework to detect ships from maritime surveillance videos and accurately identify ship behavior in continuous frames. The average check rate reaches 92.85%, and the registration rate reaches 93.91%, respectively. It shows that the proposed method identifies the historical behavior of the detected object successfully, helps managers understand the historical navigation, predicts the future navigation trajectory, implements early warning measures to ensure maritime traffic safety. Li et al. [124] proposed a lightweight ship detection model (LSDM) based on YOLOv3 and DenseNet, in which the backbone network is improved by using dense connection inspired from DenseNet, and the feature pyramid networks are improved by using spatial separation convolution to replace the original convolution network. In the proposed model, only one-third of the parameters of the YOLOv3 network can reach average accuracy of 94% for ship detection, and in the LSDM tiny network, just one-

eighth of the parameters of the YOLOv3 network can reach double detection speed and average accuracy of 93.5%.

Qiao et al. [125] proposed a detection framework based on YOLOv3, which integrates multimodel and multicue (M^3C) pipeline. Multimodel is used to solve the problem of unstable tracking of target maneuverability in traditional single-model Kalman tracker (such as CV model), and multicue solves the problem of frequent IDS caused by motion blurring and occlusion. The two public maritime datasets showed that the proposed method achieved state-of-the-art performance, not only in identity switches (IDS) but also in frame rates. Huang et al. [126] solved the problem of low recognition rate on a small dataset and improved the real-time performance of ship detection. It provides a high-precision, real-time ship detection for smart port management and USV visualization.

The author discovered that the current research of marine moving object detection has flaws, and the dataset from the perspective of the ship bridge is difficult to obtain. So far, there is no suitable benchmark. Figure 3 shows an example of an onboard visual navigation dataset from the author's lab.

Benchmark datasets containing various marine scenes from the perspective of ship bridges need to be presented, and all relevant studies should have unified standards and recognized evaluation mechanisms. Figure 4 shows an example result of onboard object detection. In the verification dataset, the missed detection rate of small targets should be included, which is essential for the autonomous navigation of large ships.

As shown in the upper left corner of Figure 4, the obstacle of the marine surface and navigation aid signs should be clearly identified, and the trained model needs to

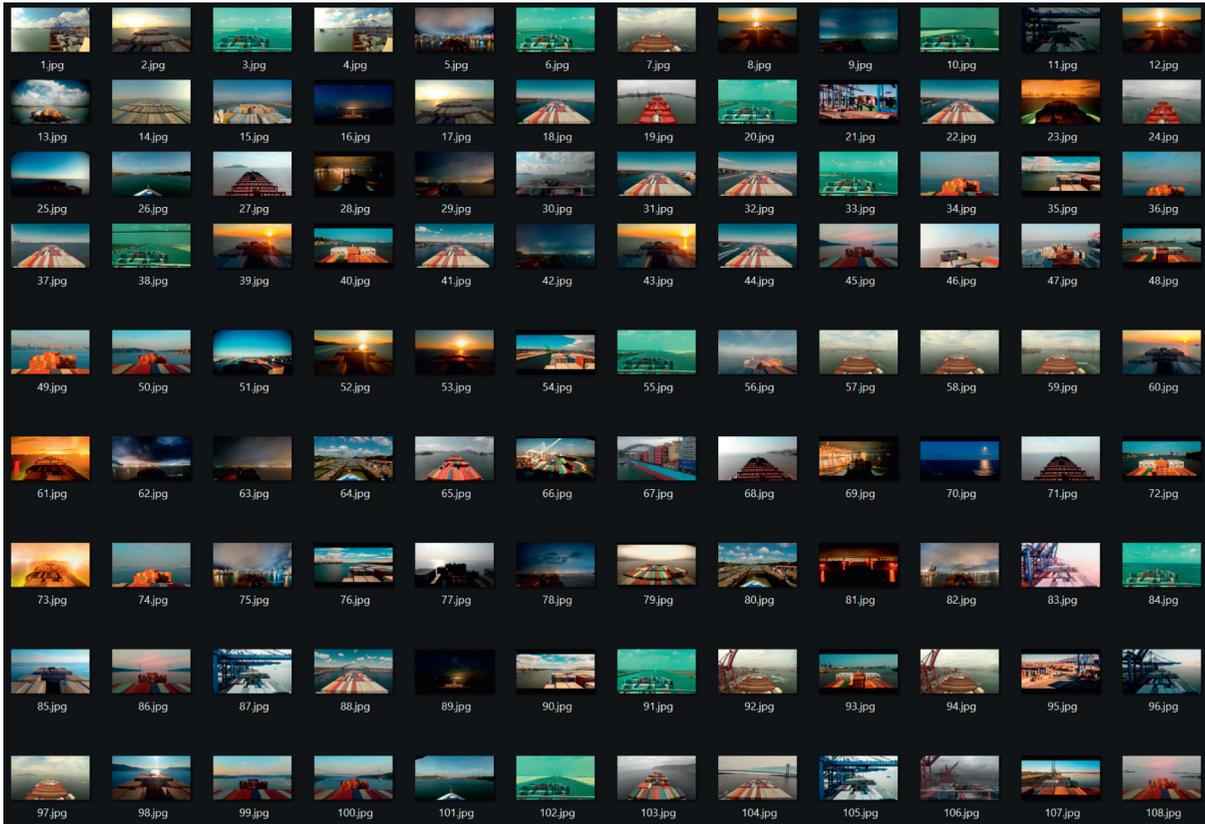


FIGURE 3: An example of onboard visual navigation dataset.

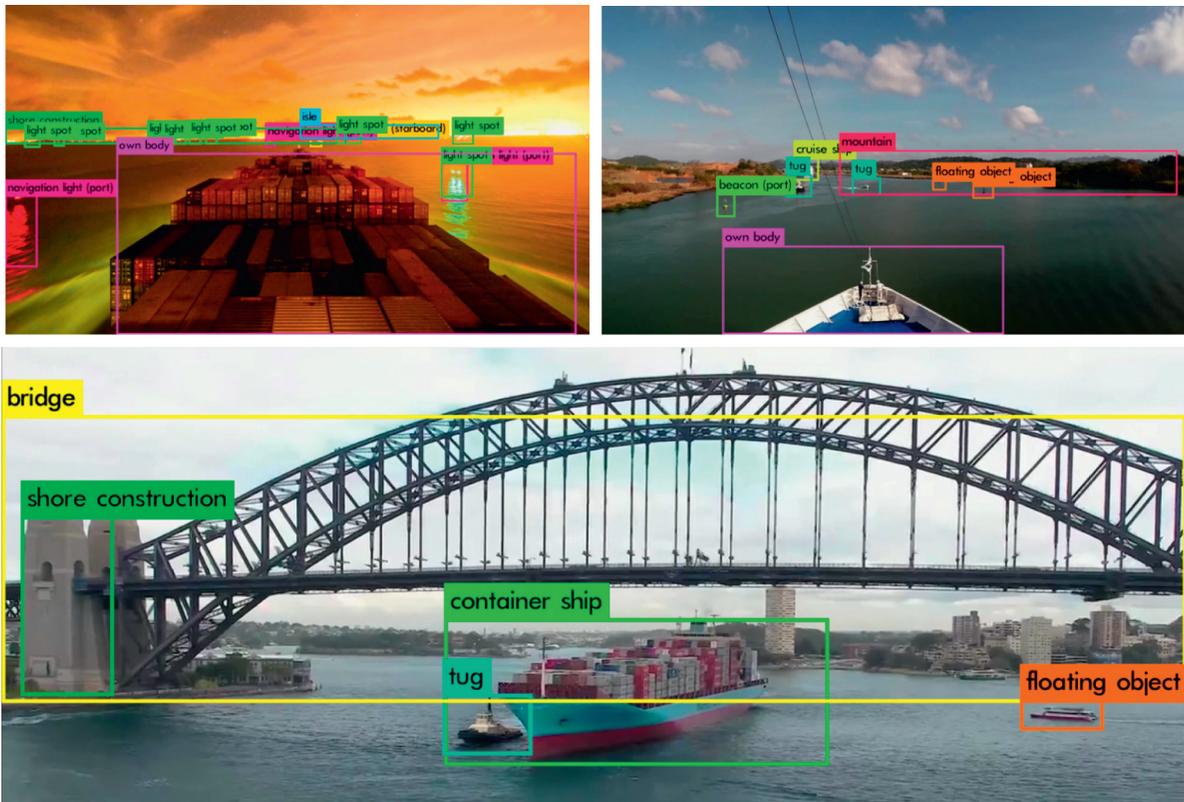


FIGURE 4: An example of onboard object detection.

understand the different meanings of different objects for navigation. The identification of near-shore constructions and moving objects on the water, the background lights on the shore, and the lights on the ship is still a critical problem for object detection.

3.2.3. Background Subtraction. In the characteristics of dynamic marine environment, the detector needs to subtract the dynamically changing objects from the backgrounds; meanwhile, there are a large number of linear features and constantly changing lighting conditions. Even the advanced sea level detection technology and video frame registration technology are facing challenges. Many background subtraction and object detection methods are very difficult in the video stream.

For example, [84] designs a multiclass ship dataset (MSD) to highlight the difference between the ship and the background; it can improve the accuracy of tiny ship detection.

Prasad et al. [127] provided a benchmark of the performance of 23 classical and state-of-the-art background subtraction algorithms on visible range and near-infrared range videos in the Singapore Maritime Dataset. This paper indicates the limitations of the conventional performance evaluation criteria for maritime vision and proposes new performance evaluation criteria that are better suited to this problem.

Although these 23 methods have been successful, the recall and accuracy are extremely low. Even the most advanced BS technology cannot deal well in the marine environment. This means that the new BS algorithm needs to be formulated for maritime vision. The traditional performance evaluation index IoU is modified to a new evaluation index IOG, and a new index bottom edge proximity (BEP) is proposed to judge whether the bottom of detection object (DO) and ground truth (GT) are close. This indicator enables more extensive detection in the presence of trails.

Zhang et al. [122] proposed a discrete cosine transform (DCT) based ship detection algorithm which can extract the sea regions accurately for complex background modeling. The main contribution is to provide more accurate detection results within the complex sea surface background, which is of vital importance for ship-/buoy-based surveillance applications in the presence of large waves. The independent detectors for sky and sea regions increase the detection sensitivity to small objects around the horizon.

The lighting environment at sea is ever-changing; one method or a model suitable for one weather and lighting condition is ineffective. Establishing a model that can seamlessly select models and methods for different lighting conditions is essential for the practical application of maritime treatment. Prasad et al. [13] discussed the technical challenges in maritime image processing and machine vision problems for video streams generated by cameras. Challenges are arising from the dynamic nature of the background, unavailability of static cues, presence of small objects at distant backgrounds, and illumination effects.

Chan et al. [128] compared thirty-seven nonstatic electrooptical sensor (combine visible-light and infrared cameras)-based background subtraction methods; the results indicate that background subtraction algorithms of the multiple features category can better handle maritime challenges, thereby realizing higher accuracy when analyzing visible-light and infrared cameras.

3.2.4. Other Applications. Augmented reality (AR) can combine computer-generated graphic information with real camera views and is an effective display technology. Reference [129] used additional location data retrieved from the AIS device to improve retrieval performance based on the characteristics of the sea-sky-line boundary and used the k -means clustering algorithm and pixel contour to distinguish the sea-sky-line. The author also emphasized that the proposed system is based on CCTV and computer image processing; therefore, the performance is influenced by sea conditions, for example, the low light condition such as foggy, dark-night, and heavy rainy days.

4. Discussion

4.1. Model Comparison. The accuracy and real-time requirements of object detection for autonomous ship navigation and maritime surveillance are important. It is necessary to propose a maritime environment image/video perception based on an improved regressive deep convolution network. YOLO series architecture is always the first neural network to be considered, for example, [12, 71, 74, 75, 77, 79, 82, 84, 88, 123, 126, 130–132]; these improvements contributed to a stronger baseline cross YOLO series detector.

As shown in Table 2, we collect some YOLO backbone network-based marine object detection models. Based on the advantages of YOLO in detection efficiency and speed, most of the researches focus on ship detection task. Experiments on public datasets (such as SMD and SeaShips) show that most of the enhanced YOLO series models have improved performance in different levels.

4.2. Marine Datasets Comparison. Moosbauer et al. [144] proposed a benchmark that is based on the Singapore Maritime Dataset (SMD). As shown in Table 3, this dataset included onshore and onboard objects in the marine environment; it provides Visual-Optical and Near-Infrared videos along with annotations for object detection. The authors evaluate two state-of-the-art object detection models for the applicability in the maritime domain: Faster R-CNN and Mask R-CNN. The SMD-based dataset can be used as a benchmark that encourages reproducibility and comparability for object detection in maritime environments. Recent research [12, 70, 81, 110, 127, 144–151] reflects this characteristic.

To advance object detection research in Earth Vision, also known as Earth Observation and Remote Sensing, [68] introduces a large-scale Dataset for Object deTectioN in Aerial images (DOTA). There are many studies using this

TABLE 2: Application of improved YOLO backbone network in maritime object detection.

| Algorithms (backbone) | Datasets | Scenarios | Improved method | Effect |
|-----------------------|----------------------------------|------------------------------------|--|--|
| YOLOv2 [71] | Small ship dataset | Small ship detection | Density-based spatial clustering (DBSCAN) | AUC: 0.960 TPR: 98.3% FPR: 3.5% |
| YOLOv3 [77] | SeaShip dataset Buoy dataset | Ship detection | Loss function (GIOU) PANet replaces FPN | mAP (SeaShip): 98.37% mAP (Buoy dataset): 90.58% |
| YOLOv2 and CNN [12] | Pascal VOC SMD | Ship detection | — | Recall: 77.12% IoU: 66.69% |
| YOLOv3 [75] | Shanghai port surveillance video | Ship detection | — | Average acc.: 0.84 |
| YOLOv3 [79] | SeaShip dataset | Ship detection | CBAM | mAP increase 9.6% |
| YOLO [123] | Self-collected | Ship detection | — | Average acc.: 92.85% |
| YOLOv3 tiny [124] | From Internet | Ship detection | Dense connection spatial separate conv. | LSDM average acc.: 94% LSDM tiny: 93.5% mAP@0.25 IoU: 0.97 mAP@0.50 IoU: 0.90 mAP@0.75 IoU: 0.29 |
| YOLOv3 [132] | LWIR | Object detection | — | mAP: 41.2% |
| YOLOv3 [125] | SMD; PETS 2016 | Ship tracking | — | Recall: 73.86% |
| YOLOv2 [130] | Pascal VOC SMD | Object detection Ship detection | Pass through layer Transfer learning | IoU: 60.79% |

TABLE 3: Comparison of representative marine object detection datasets.

| Datasets | Resolution (Pixels) | Usage Scenarios | References | Data sources |
|--------------------|---------------------|--|--------------------------------------|---------------------|
| <i>SMD</i> | 1080 × 1920 | Autonomous ship navigation Marine horizon detection Maritime surveillance Object detection Object tracking and so on | [13, 16, 18, 70, 109, 110, 133, 134] | Onboard and onshore |
| <i>DOTA</i> | 4000 × 4000 | Object detection Remote sensing Maritime safety Maritime surveillance | [135–138] | Aerial images |
| <i>SeaShips</i> | 1920 × 1080 | Marine object detection Target tracking Maritime surveillance | [78, 139, 140] | Onshore |
| <i>AIR-SARShip</i> | 3000 × 3000 | Object detection Remote sensing Maritime surveillance | [141–143] | GF-3 satellite |

dataset in the field of maritime remote surveillance [135, 138, 152] and so on.

SeaShips is a large ship dataset. The dataset consists of 11,126 images, covering 6 common ship categories (ore ships, bulk carriers, general cargo ships, container ships, fishing ships, and passenger ships). All images come from about 5400 real video clips, collected by 156 surveillance cameras in the coastline video surveillance system. Some research uses this dataset to train their model or improve the model's performance [139, 140].

Some other datasets need to be highlighted. Spagnolo et al. [153] presented a boat Re-ID dataset composed of 107 classes, and each class represents a different boat with a total of 5523 images. In order to verify the superiority of the proposed dataset, the authors give the results of training

CNN by using this dataset, and the research results can be used as a benchmark for future comparisons.

Bovcon et al. [154] introduced the MaSTr1325 dataset for training deep USV obstacle detection models in small-sized coastal USV. They also proposed a data augmentation protocol to address slight appearance differences. The dataset is applied to three popular semantic segmentation architectures: U-Net, PSPNet, and Deeplabv2, among which Deeplabv2 performs best in obstacle detection. In [148], the authors used 4K videos for maritime video surveillance and proposed an approach that attempts to leverage both temporal and spatial video information for achieving fast and accurate object extraction. Multiscale texture discrimination algorithm carried out key video locations to achieve final object extraction.

4.3. Current Challenges and Future Works. Computer vision is the subject of studying image information organization, object and scenario recognition, and interpreting events by taking images (video) as input and aiming at representation and understanding of the environment. Judging from the current research status, the research mainly focuses on the organization and recognition of image information, and the interpretation of events is rarely involved, at least at a very preliminary stage.

The relationship between artificial intelligence and computer vision is as follows: artificial intelligence puts more emphasis on reasoning and decision-making, but at least computer vision is still mainly at the stage of image information expression and object recognition. Object recognition, environment perception, and scenario understanding also involve reasoning and decision-making from image features, but they are fundamentally different from the reasoning and decision-making of artificial intelligence.

4.3.1. Current Challenges. Specific maritime engineering applications belong to systemic issues, affected by many objective factors, for example, equipment shaking, model dependence, and light interference on shore.

- (i) Shaking problem of imaging equipment: in actual marine engineering applications (onshore and onboard), the effect of the model is often much lower than the accuracy and speed obtained in the laboratory. In the actual marine environment, the shaking of the equipment is the main reason for small object missed or object false detection. Even in the tracking task, the tracking fails due to the same reason.
- (ii) Model dependence: at present, all models require fixed scene training; the environmental changes have a large impact on the recognition accuracy of the model. Weather changes will change the external photosensitive environment, which will lead to bad results for marine object detection, and even equipment updates can cause model detection to fail.
- (iii) Background light pollution on the shore is an important issue, and there are few research papers related to the extraction of background light on the shore. Even experienced seafarers are still prone to think of shore lights as lights moving on the sea and make inappropriate decision-making. This is an urgent problem in the field of autonomous ship navigation.

4.3.2. Future Works

(1) Online Training. The current models are trained first and then deployed. Applications such as autonomous ship navigation require reasoning and decision-making based on environmental information in real time. One of the future trends is to solve this problem.

(2) Build a Maritime Data Sharing Center. (1) Unified model algorithm evaluation mechanism: at present, maritime surveillance and intelligent transportation need a public benchmark for different researchers who proposed various models. (2) Construct various marine scenarios and sea condition dataset share platform. The actual detection task of marine objects requires training with a lot of data in their respective sea conditions; the current datasets cannot complete this task. We will put more energy into the work of data sorting and build a maritime data sharing platform.

5. Conclusions

This survey covers most of the application scenarios of object detection for maritime surveillance and autonomous ship navigation. In recent years, a large number of marine object detection models based on deep learning have been proposed, but due to the lack of universal evaluation criteria, it is difficult to compare different improved models. According to the characteristics of the maritime environment, this paper summarized the advantages of the computer vision milestone model and proposed different application scenarios of the single-stage model and the multistage model under different development routes. The most popular YOLO series models are compared in different dimensions, and the importance of public dataset benchmarks is proposed. We also discussed the urgency of building a maritime proprietary dataset platform that satisfies different scenarios and model training in practical engineering applications. This work will put forward feasible suggestions for future research directions of deep learning-based marine object detection.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported in part by the Fundamental Research Funds for the Central Universities, Grant nos. 3132021130 and 3132019400.

References

- [1] A. Felski and K. Zwolak, "The ocean-going autonomous ship-challenges and threats," *Journal of Marine Science and Engineering*, vol. 8, no. 1, p. 41, 2020.
- [2] A. Noel, K. Shreyanka, and K. Gowtham, "Autonomous ship navigation methods: a review," in *Proceedings of the Conference Proceedings of ICMET OMAN. 2019*, Muscat, Oman, November 2019.
- [3] M. Furusho, "Visual environment and sight-line displacements of navigation officers for good lookout," *Journal of Light and Visual Environment*, vol. 25, no. 1, pp. 43–48, 2001.
- [4] X.-Y. Zhou, J.-J. Huang, and F.-W. Wang, "A study of the application barriers to the use of autonomous ships posed by the good seamanship requirement of COLREGs," *Journal of Navigation*, vol. 73, no. 3, pp. 710–725, 2020.
- [5] T. Morris, *Computer Vision and Image processing*, Palgrave Macmillan Ltd, London, UK, 2004.

- [6] T. Cornsweet, *Visual perception*, Academic Press, Cambridge, MA, USA, 2012.
- [7] Z. Zou, Z. Shi, and Y. Guo, "Object detection in 20 years: a survey," 2019, <https://arxiv.org/abs/1905.05055>.
- [8] L. Jiao, F. Zhang, and F. Liu, "A survey of deep learning based object detection," *IEEE access*, vol. 7, pp. 128837–128868, 2019.
- [9] X. Wu, D. Sahoo, and S. C. H. Hoi, "Recent advances in deep learning for object detection," *Neurocomputing*, vol. 396, pp. 39–64, 2020.
- [10] J. Chen, "Foreground-background imbalance problem in deep object detectors: a review," in *Proceedings of the 2020 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR)*, IEEE, Tokyo, Japan, August 2020.
- [11] D. Qiao, G. Liu, and T. Lv, "Marine vision-based situational awareness using discriminative deep learning: a survey," *Journal of Marine Science and Engineering*, vol. 9, no. 4, p. 397, 2021.
- [12] S. J. Lee, M. I. Roh, and H. W. Lee, "Image-based ship detection and classification for unmanned surface vehicle using real-time object detection neural networks," in *Proceedings of the 28th International Ocean and Polar Engineering Conference*, OnePetro, Sapporo, Japan, June 2018.
- [13] D. K. Prasad, C. K. Prasath, and D. Rajan, "Challenges in video based object detection in maritime scenario using computer vision," 2016, <https://arxiv.org/abs/1608.01079>.
- [14] S. Thombre, Z. Zhao, and H. Ramm-Schmidt, "Sensors and ai techniques for situational awareness in autonomous ships: a review," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, 2020.
- [15] M. N. Chapel and T. Bouwmans, "Moving objects detection with a moving camera: a comprehensive review," *Computer Science Review*, vol. 38, Article ID 100310, 2020.
- [16] 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.
- [17] M. Moniruzzaman, S. M. S. Islam, and M. Bennamoun, "Deep learning on underwater marine object detection: a survey," in *Proceedings of the International Conference on Advanced Concepts for Intelligent Vision Systems*, pp. 150–160, Springer, Antwerp, Belgium, September 2017.
- [18] M. A. Hashmani, M. Umair, and S. S. H. Rizvi, "A survey on edge detection based recent marine horizon line detection methods and their applications," *IEEE*, in *Proceedings of the 2020 3rd International Conference on Computing, Mathematics and Engineering Technologies (iCoMET)*, pp. 1–5, Sindh, Pakistan, January 2020.
- [19] M. Petković, I. Vujović, and I. Kuzmanić, "An overview on horizon detection methods in maritime video surveillance," *Transactions on Maritime Science*, vol. 9, no. 1, pp. 106–112, 2020.
- [20] B. Li, X. Xie, and X. Wei, "Ship detection and classification from optical remote sensing images: a survey," *Chinese Journal of Aeronautics*, vol. 34, 2020.
- [21] Z. Liu, C. K. Loo, and K. Pasupa, "Meta-cognitive recurrent kernel online sequential extreme learning machine with kernel adaptive filter for concept drift handling," *Engineering Applications of Artificial Intelligence*, vol. 88, Article ID 103327, 2020.
- [22] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition*, vol. 1, Kauai, HI, USA, December 2001.
- [23] P. Viola and M. J. Jones, "Robust real-time face detection," *International Journal of Computer Vision*, vol. 57, no. 2, pp. 137–154, 2004.
- [24] G. Rätsch, T. Onoda, and K. R. Müller, "Soft margins for AdaBoost," *Machine Learning*, vol. 42, no. 3, pp. 287–320, 2001.
- [25] B. Yang, J. Yan, and Z. Lei, "Aggregate channel features for multi-view face detection," in *Proceedings of the IEEE International Joint Conference on Biometrics*, pp. 1–8, IEEE, Seoul, Korea, August 2014.
- [26] M. Cerf, J. Harel, and W. Einhäuser, "Predicting human gaze using low-level saliency combined with face detection[J]," *Advances in Neural Information Processing Systems*, vol. 20, pp. 1–7, 2008.
- [27] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proceedings of the 2005 IEEE computer conference on computer vision and pattern recognition (CVPR'05)*, vol. 1, pp. 886–893, IEEE, San Diego, CA, USA, June 2005.
- [28] Z. R. Wang, Y. L. Jia, and H. Huang, "Pedestrian detection using boosted hog features," in *Proceedings of the 2008 11th International IEEE Conference on Intelligent Transportation Systems*, pp. 1155–1160, IEEE, Beijing, China, October 2008.
- [29] G. Gan and J. Cheng, "Pedestrian detection based on HOG-LBP feature," in *Proceedings of the 2011 Seventh International Conference on Computational Intelligence and Security*, pp. 1184–1187, IEEE, Sanya, China, December 2011.
- [30] Z. Luo, "Rotation-invariant histograms of oriented gradients for local patch robust representation," in *Proceedings of the 2015 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA)*, IEEE, Hong Kong, China, December 2015.
- [31] P. Felzenszwalb, D. McAllester, and D. Ramanan, "A discriminatively trained, multiscale, deformable part model," in *Proceedings of the 2008 IEEE conference on computer vision and pattern recognition*, pp. 1–8, Anchorage, AK, USA, June 2008.
- [32] W. Liu, D. Anguelov, and D. Erhan, "Ssd: single shot multibox detector," in *Proceedings of the European Conference on Computer Vision*, pp. 21–37, Springer, Amsterdam, The Netherlands, October 2016.
- [33] C. Szegedy, W. Zaremba, and I. Sutskever, "Intriguing properties of neural networks," 2013, <https://arxiv.org/abs/1312.6199>.
- [34] M. Oquab, L. Bottou, and I. Laptev, "Learning and transferring mid-level image representations using convolutional neural networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1717–1724, Columbus, OH, USA, June 2014.
- [35] A. Newell, K. Yang, and J. Deng, "Stacked hourglass networks for human pose estimation," in *Proceedings of the European Conference on Computer Vision*, pp. 483–499, Springer, Zurich, Switzerland, September 2016.
- [36] B. Dolapci and C. Özcan, "Automatic ship detection and classification using machine learning from remote sensing images on Apache Spark," *Journal of Intelligent Systems: Theory and Applications*, vol. 4, no. 2, pp. 94–102, 2021.
- [37] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*, vol. 1, MIT press, Cambridge, MA, USA, 2016.
- [38] Y. LeCun, L. Bottou, and Y. Bengio, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

- [39] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in Neural Information Processing Systems*, vol. 25, pp. 1097–1105, 2012.
- [40] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, <https://arxiv.org/abs/1409.1556>.
- [41] C. Szegedy, W. Liu, and Y. Jia, "Going deeper with convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–9, Boston, MA, USA, June 2015.
- [42] K. He, X. Zhang, and S. Ren, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778, Las Vegas, NV, USA, June 2016.
- [43] K. He, X. Zhang, and S. Ren, "Identity mappings in deep residual networks," in *Proceedings of the European Conference on Computer Vision*, pp. 630–645, Springer, Amsterdam, The Netherlands, October 2016.
- [44] J. Y. Zhu, T. Park, and P. Isola, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2223–2232, Venice, Italy, October 2017.
- [45] A. Vaswani, N. Shazeer, and N. Parmar, "Attention is all you need," in *Proceedings of the Advances in Neural Information Processing Systems*, pp. 5998–6008, Long Beach, CA, USA, December 2017.
- [46] C. Szegedy, S. Ioffe, and V. Vanhoucke, "Inception-v4, inception-resnet and the impact of residual connections on learning," in *Proceedings of the Thirty-first AAAI conference on artificial intelligence*, San Francisco, CA, USA, February 2017.
- [47] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proceedings of the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7132–7141, Salt Lake City, UT, USA, December 2018.
- [48] G. Huang, Z. Liu, and L. Van Der Maaten, "Densely Connected Convolutional networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4700–4708, Honolulu, HI, USA, July 2017.
- [49] N. Carion, F. Massa, and G. Synnaeve, "End-to-end object detection with transformers," in *Proceedings of the European Conference on Computer Vision*, pp. 213–229, Springer, Glasgow, UK, August 2020.
- [50] K. Han, Y. Wang, and H. Chen, "A survey on visual transformer," 2020, <https://arxiv.org/abs/2012.12556>.
- [51] R. Girshick, J. Donahue, and T. Darrell, "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.
- [52] K. He, X. Zhang, and S. Ren, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 9, pp. 1904–1916, 2015.
- [53] R. Girshick, "Fast r-cnn," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1440–1448, Santiago, Chile, December 2015.
- [54] S. Ren, K. He, and R. Girshick, "Faster r-cnn: towards real-time object detection with region proposal networks," *Advances in Neural Information Processing Systems*, vol. 28, pp. 91–99, 2015.
- [55] K. He, G. Gkioxari, and P. Dollár, "Mask r-cnn," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2961–2969, Venice, Italy, October 2017.
- [56] Z. Cai and N. Vasconcelos, "Cascade R-Cnn: delving into high quality object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 6154–6162, Salt Lake City, UT, USA, June 2018.
- [57] M. Everingham, L. Van Gool, and C. K. I. Williams, "The pascal visual object classes (voc) challenge," *International Journal of Computer Vision*, vol. 88, no. 2, pp. 303–338, 2010.
- [58] T. Y. Lin, P. Dollár, and R. Girshick, "Feature Pyramid Networks for Object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2117–2125, Honolulu, HI, USA, July 2017.
- [59] S. Beery, G. Wu, and V. Rathod, "Context r-cnn: long term temporal context for per-camera object detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 13075–13085, Seattle, WA, USA, June 2020.
- [60] B. Chen, G. Ghiasi, and H. Liu, "Mnasfpn: learning latency-aware pyramid architecture for object detection on mobile devices," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 13607–13616, Seattle, WA, USA, June 2020.
- [61] J. Redmon, S. Divvala, and R. Girshick, "You only look once: unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 779–788, Seattle, WA, USA, June 2016.
- [62] T. Y. Lin, P. Goyal, and R. Girshick, "Focal loss for dense object detection," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2980–2988, Venice, Italy, October 2017.
- [63] A. Howard, M. Sandler, and G. Chu, "Searching for mobilenetv3," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 1314–1324, Seoul, South Korea, October 2019.
- [64] Á. Morera and A. B. Moreno, "SSD vs. Yolo for detection of outdoor urban advertising panels under multiple variabilities," *Sensors*, vol. 20, no. 16, p. 4587, 2020.
- [65] T. Y. Lin, M. Maire, and S. Belongie, "Microsoft coco: Common objects in context," in *Proceedings of the European conference on computer vision*, pp. 740–755, Springer, Zurich, Switzerland, September 2014.
- [66] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? the kitti vision benchmark suite," in *Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3354–3361, IEEE, Providence, RI, USA, June 2012.
- [67] R. Krishna, Y. Zhu, and O. Groth, "Visual genome: connecting language and vision using crowdsourced dense image annotations," 2016, <https://arxiv.org/abs/1602.07332>.
- [68] G. S. Xia, X. Bai, and J. Ding, "DOTA: a large-scale dataset for object detection in aerial images," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3974–3983, Salt Lake City, UT, USA, June 2018.
- [69] R. L. Zhang and M. Furusho, "Developing generative adversarial nets to extend training sets and optimize discrete actions," *TransNav: International Journal on Marine Navigation and Safety of Sea Transportation*, vol. 13, 2019.
- [70] H. C. Shin, K. I. Lee, and C. E. Lee, "Data augmentation method of object detection for deep learning in maritime image," in *Proceedings of the 2020 IEEE International Conference on Big Data and Smart Computing (BigComp)*, pp. 463–466, IEEE, Busan, Korea, February 2020.

- [71] Z. Chen, D. Chen, and Y. Zhang, "Deep learning for autonomous ship-oriented small ship detection," *Safety Science*, vol. 130, Article ID 104812, 2020.
- [72] Y. Ren, J. Yang, and Q. Zhang, "Ship recognition based on Hu invariant moments and convolutional neural network for video surveillance," *Multimedia Tools and Applications*, vol. 80, no. 1, pp. 1343–1373, 2021.
- [73] X. Cao, S. Gao, and L. Chen, "Ship recognition method combined with image segmentation and deep learning feature extraction in video surveillance," *Multimedia Tools and Applications*, vol. 79, no. 13, pp. 9177–9192, 2020.
- [74] Z. Dong and B. Lin, "Learning a robust CNN-based rotation insensitive model for ship detection in VHR remote sensing images," *International Journal of Remote Sensing*, vol. 41, no. 9, pp. 3614–3626, 2020.
- [75] X. Chen, L. Qi, and Y. Yang, "Port ship detection in complex environments," in *Proceedings of the 2019 International Conference on Sensing and Instrumentation in IoT Era (ISSI)*, pp. 1–6, IEEE, Piscataway, NJ, USA, August 2019.
- [76] Z. Shao, L. Wang, and Z. Wang, "Saliency-aware convolution neural network for ship detection in surveillance video," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 3, pp. 781–794, 2019.
- [77] T. Liu, B. Pang, and S. Ai, "Study on visual detection algorithm of sea surface targets based on improved YoloV3," *Sensors*, vol. 20, no. 24, p. 7263, 2020.
- [78] Z. Shao, W. Wu, and Z. Wang, "Seaships: a large-scale precisely annotated dataset for ship detection," *IEEE Transactions on Multimedia*, vol. 20, no. 10, pp. 2593–2604, 2018.
- [79] H. Li, L. Deng, and C. Yang, "Enhanced Yolo v3 tiny network for real-time ship detection from visual image," *IEEE Access*, vol. 9, pp. 16692–16706, 2021.
- [80] H. Huang, D. Sun, and R. Wang, "Ship target detection based on improved Yolo network," *Mathematical Problems in Engineering*, vol. 2020, Article ID 6402149, 10 pages, 2020.
- [81] T. Cane and J. Ferryman, "Evaluating deep semantic segmentation networks for object detection in maritime surveillance," in *Proceedings of the 2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, pp. 1–6, IEEE, Auckland, New Zealand, November 2018.
- [82] S. Ghosh, P. K. Konugurthi, and G. Shankar Rao Singurapu, "On-board ship detection for medium resolution optical sensors," *Sensors*, vol. 21, no. 9, p. 3062, 2021.
- [83] L. Tian, Y. Cao, and B. He, "Image enhancement driven by object characteristics and dense feature reuse network for ship target detection in remote sensing imagery," *Remote Sensing*, vol. 13, no. 7, p. 1327, 2021.
- [84] L. Chen, W. Shi, and D. Deng, "Improved YoloV3 based on attention mechanism for fast and accurate ship detection in optical remote sensing images," *Remote Sensing*, vol. 13, no. 4, p. 660, 2021.
- [85] Q. Wang, F. Shen, and L. Cheng, "Ship detection based on fused features and rebuilt YoloV3 networks in optical remote-sensing images," *International Journal of Remote Sensing*, vol. 42, no. 2, pp. 520–536, 2021.
- [86] C. Cao, J. Wu, and X. Zeng, "Research on airplane and ship detection of aerial remote sensing images based on convolutional neural network," *Sensors*, vol. 20, no. 17, p. 4696, 2020.
- [87] G. Tang, Y. Zhuge, and C. Claramunt, "N-Yolo: A SAR ship detection using noise-classifying and complete-target extraction," *Remote Sensing*, vol. 13, no. 5, p. 871, 2021.
- [88] G. Tang, S. Liu, I. Fujino, and H. Yolo, "A single-shot ship detection approach based on region of interest preselected network," *Remote Sensing*, vol. 12, no. 24, p. 4192, 2020.
- [89] R. Ribeiro, G. Cruz, and J. Matos, "A dataset for airborne maritime surveillance environments," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 29, no. 9, pp. 2720–2732, 2017.
- [90] G. Cruz and A. Bernardino, "Aerial detection in maritime scenarios using convolutional neural networks," in *Proceedings of the International Conference on Advanced Concepts for Intelligent Vision Systems*, pp. 373–384, Springer, Lecce, Italy, October 2016.
- [91] C. D. Rodin, L. N. de Lima, and F. A. de Alcantara Andrade, "Object classification in thermal images using convolutional neural networks for search and rescue missions with unmanned aerial systems," in *Proceedings of the 2018 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–8, IEEE, Rio de Janeiro, Brazil, July 2018.
- [92] M. M. Marques, V. Lobo, and A. P. Aguiar, "An unmanned aircraft system for maritime operations: the automatic detection subsystem," *Marine Technology Society Journal*, vol. 55, no. 1, pp. 38–49, 2021.
- [93] S. Xiu, Y. Wen, and H. Yuan, "A multi-feature and multi-level matching algorithm using aerial image and ais for vessel identification," *Sensors*, vol. 19, no. 6, p. 1317, 2019.
- [94] G. Cruz and A. Bernardino, "Learning temporal features for detection on maritime airborne video sequences using convolutional LSTM," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 9, pp. 6565–6576, 2019.
- [95] S. Chen, R. Zhan, and W. Wang, "Learning slimming SAR ship object detector through network pruning and knowledge distillation," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 1267–1282, 2020.
- [96] Z. Long, W. Suyuan, and C. U. I. Zhongma, "Lira-Yolo: a lightweight model for ship detection in radar images," *Journal of Systems Engineering and Electronics*, vol. 31, no. 5, pp. 950–956, 2020.
- [97] R. Yang, Z. Pan, and X. Jia, "A novel CNN-based detector for ship detection based on rotatable bounding box in SAR images," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 1938–1958, 2021.
- [98] F. S. Hass and J. Jokar Arsanjani, "Deep learning for detecting and classifying ocean objects: application of YoloV3 for iceberg–ship discrimination," *ISPRS International Journal of Geo-Information*, vol. 9, no. 12, p. 758, 2020.
- [99] P. Chen, Y. Li, and H. Zhou, "Detection of small ship objects using anchor boxes cluster and feature pyramid network model for SAR imagery," *Journal of Marine Science and Engineering*, vol. 8, no. 2, p. 112, 2020.
- [100] Y. Lang and B. Yuan, "Algorithm application based on the infrared image in unmanned ship target image recognition," *Microprocessors and Microsystems*, vol. 80, Article ID 103554, 2021.
- [101] J. Xie, E. Stensrud, and T. Skramstad, "Detection-based object tracking applied to remote ship inspection," *Sensors*, vol. 21, no. 3, p. 761, 2021.
- [102] F. Han, J. Yao, and H. Zhu, "Underwater image processing and object detection based on deep CNN method," *Journal of Sensors*, 2020.
- [103] D. Lee, G. Kim, and D. Kim, "Vision-based object detection and tracking for autonomous navigation of underwater robots," *Ocean Engineering*, vol. 48, pp. 59–68, 2012.

- [104] F. Farahnakian and J. Heikkonen, "Deep learning based multi-modal fusion architectures for maritime vessel detection," *Remote Sensing*, vol. 12, no. 16, p. 2509, 2020.
- [105] M. Pan, Y. Liu, and J. Cao, "Visual recognition based on deep learning for navigation mark classification," *IEEE Access*, vol. 8, pp. 32767–32775, 2020.
- [106] Y. Zhou, Y. Lu, and Y. Shen, "Polarized remote inversion of the refractive index of marine spilled oil from PARASOL images under sunglint," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 4, pp. 2710–2719, 2019.
- [107] A. Martins, A. Dias, and J. Almeida, "Field experiments for marine casualty detection with autonomous surface vehicles," in *Proceedings of the 2013 OCEANS-San Diego*, pp. 1–5, IEEE, San Diego, CA, USA, September 2013.
- [108] B. Wang, Y. Su, and L. Wan, "A sea-sky line detection method for unmanned surface vehicles based on gradient saliency," *Sensors*, vol. 16, no. 4, p. 543, 2016.
- [109] C. Jeong, H. S. Yang, and K. D. Moon, "A novel approach for detecting the horizon using a convolutional neural network and multi-scale edge detection," *Multidimensional Systems and Signal Processing*, vol. 30, no. 3, pp. 1187–1204, 2019.
- [110] D. K. Prasad, D. Rajan, and L. Rachmawati, "MuSCoWERT: consistence of weighted edge Radon transform for horizon detection in maritime images," *JOSA A*, vol. 33, no. 12, pp. 2491–2500, 2016.
- [111] S. M. Ettinger, M. C. Nechyba, and P. G. Ifju, "Vision-guided flight stability and control for micro air vehicles," *Advanced Robotics*, vol. 17, no. 7, pp. 617–640, 2003.
- [112] S. Fefilyayev, D. Goldgof, and M. Shreve, "Detection and tracking of ships in open sea with rapidly moving buoy-mounted camera system," *Ocean Engineering*, vol. 54, pp. 1–12, 2012.
- [113] H. Bouma, D. J. J. de Lange, and S. P. van den Broek, "Automatic detection of small surface targets with electro-optical sensors in a harbor environment," *International Society for Optics and Photonics*, vol. 7114, Article ID 711402, 2008.
- [114] E. Gershikov, T. Libe, and S. Kosolapov, "Horizon line detection in marine images: which method to choose?" *International Journal on Advances in Intelligent Systems*, vol. 6, no. 1, 2013.
- [115] B. A. Alpatov, P. V. Babayan, and N. Y. Shubin, "Weighted Radon transform for line detection in noisy images," *Journal of Electronic Imaging*, vol. 24, no. 2, Article ID 023023, 2015.
- [116] C. Y. Jeong, H. S. Yang, and K. D. Moon, "Fast horizon detection in maritime images using region-of-interest," *International Journal of Distributed Sensor Networks*, vol. 14, no. 7, 2018.
- [117] D. Liang and Y. Liang, "Horizon detection from electro-optical sensors under maritime environment," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 1, pp. 45–53, 2019.
- [118] H. Zhang, P. Yin, and X. Zhang, "A robust adaptive horizon recognizing algorithm based on projection," *Transactions of the Institute of Measurement and Control*, vol. 33, no. 6, pp. 734–751, 2011.
- [119] I. Lipschutz, E. Gershikov, and B. Milgrom, "New methods for horizon line detection in infrared and visible sea images," *International Journal of Computer Engineering Research*, vol. 3, no. 3, pp. 1197–1215, 2013.
- [120] Y. Sun and L. Fu, "Coarse-fine-stitched: a robust maritime horizon line detection method for unmanned surface vehicle applications," *Sensors*, vol. 18, no. 9, p. 2825, 2018.
- [121] W. Yang, H. Li, and J. Liu, "A sea-sky-line detection method based on Gaussian mixture models and image texture features," *International Journal of Advanced Robotic Systems*, vol. 16, no. 6, 2019.
- [122] 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.
- [123] X. Chen, L. Qi, and Y. Yang, "Video-based detection infrastructure enhancement for automated ship recognition and behavior analysis," *Journal of Advanced Transportation*, vol. 2020, Article ID 7194342, 12 pages, 2020.
- [124] Z. Li, L. Zhao, and X. Han, "Lightweight ship detection methods based on YoloV3 and DenseNet," *Mathematical Problems in Engineering*, vol. 2020, Article ID 4813183, 10 pages, 2020.
- [125] D. Qiao, G. Liu, and J. Zhang, "M3C: multimodel-and-multicue-based tracking by detection of surrounding vessels in maritime environment for USV," *Electronics*, vol. 8, no. 7, p. 723, 2019.
- [126] Z. Huang, B. Sui, and J. Wen, "An intelligent ship image/video detection and classification method with improved regressive deep convolutional neural network," *Journal of Complexity*, vol. 2020, Article ID 1520872, 11 pages, 2020.
- [127] D. K. Prasad, C. K. Prasath, and D. Rajan, "Object detection in a maritime environment: performance evaluation of background subtraction methods," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 5, pp. 1787–1802, 2018.
- [128] Y. T. Chan, "Comprehensive comparative evaluation of background subtraction algorithms in open sea environments," *Computer Vision and Image Understanding*, vol. 202, Article ID 103101, 2021.
- [129] J. M. Lee, K. H. Lee, and B. Nam, "Study on image-based ship detection for AR navigation," in *Proceedings of the 2016 6th International Conference on IT Convergence and Security (ICITCS)*, pp. 1–4, IEEE, Prague, Czech, September 2016.
- [130] S. J. Leela, M. I. Roh, and M. J. Ohb, "Image-based ship detection using deep learning," *Ocean Systems Engineering*, vol. 10, 2020.
- [131] S. T. Westlake, T. N. Volonakis, and J. Jackman, "Deep learning for automatic target recognition with real and synthetic infrared maritime imagery," *International Society for Optics and Photonics*, vol. 11543, Article ID 1154309, 2020.
- [132] F. E. T. Schöller, M. K. Plenge-Feidenhans, and J. D. Stets, "Assessing deep-learning methods for object detection at sea from LWIR images," *IFAC-PapersOnLine*, vol. 52, no. 21, pp. 64–71, 2019.
- [133] M. Nalamati, N. Sharma, and M. Saqib, "Automated monitoring in maritime video surveillance system," in *Proceedings of the 2020 35th International Conference on Image and Vision Computing New Zealand (IVCNZ)*, pp. 1–6, IEEE, Wellington, New Zealand, November 2020.
- [134] H. Feng, J. Guo, and H. Xu, "SharpGAN: dynamic scene deblurring method for smart ship based on receptive field block and generative adversarial networks," *Sensors*, vol. 21, no. 11, p. 3641, 2021.
- [135] X. Hou, W. Ao, and Q. Song, "FUSAR-Ship: building a high-resolution SAR-AIS matchup dataset of Gaofen-3 for ship detection and recognition," *Science China Information Sciences*, vol. 63, no. 4, pp. 1–19, 2020.
- [136] Y. You, Z. Li, and B. Ran, "Broad area target search system for ship detection via deep convolutional neural network," *Remote Sensing*, vol. 11, no. 17, 2019.

- [137] M. Zhang, Y. Chen, and X. Liu, "Adaptive anchor networks for multi-scale object detection in remote sensing images," *IEEE Access*, vol. 8, pp. 57552–57565, 2020.
- [138] Y. Zhuang, L. Li, and H. Chen, "Small sample set inshore ship detection from VHR optical remote sensing images based on structured sparse representation," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, pp. 2145–2160, 2020.
- [139] L. Zhou, Z. Wang, and Y. Luo, "Separability and compactness network for image recognition and super-resolution," *IEEE transactions on neural networks and learning systems*, vol. 30, no. 11, pp. 3275–3286, 2019.
- [140] Z. Xu, R. Hu, and J. Chen, "Semisupervised discriminant multimanifold analysis for action recognition," *IEEE transactions on neural networks and learning systems*, vol. 30, no. 10, pp. 2951–2962, 2019.
- [141] S. U. N. Xian, W. Zhirui, and S. U. N. Yuanrui, "AIR-SARShip-1.0: high-resolution SAR ship detection dataset," *Radar Journal*, vol. 8, no. 6, pp. 852–862, 2019.
- [142] Y. Mao, Y. Yang, and Z. Ma, "Efficient low-cost ship detection for SAR imagery based on simplified U-net," *IEEE Access*, vol. 8, pp. 69742–69753, 2020.
- [143] F. Gao, Y. He, and J. Wang, "Anchor-free convolutional network with dense attention feature aggregation for ship detection in SAR images," *Remote Sensing*, vol. 12, no. 16, p. 2619, 2020.
- [144] S. Moosbauer, D. Konig, and J. Jakel, "A benchmark for deep learning based object detection in maritime environments," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, Long Beach, CA, USA, June 2019.
- [145] C. Y. Jeong, H. S. Yang, and K. D. Moon, "Horizon detection in maritime images using scene parsing network," *Electronics Letters*, vol. 54, no. 12, pp. 760–762, 2018.
- [146] R. Spraul, L. Sommer, and A. Schumann, "A comprehensive analysis of modern object detection methods for maritime vessel detection," *International Society for Optics and Photonics*, vol. 11543, Article ID 1154305, 2020.
- [147] U. Ganbold and T. Akashi, "The real-time reliable detection of the horizon line on high-resolution maritime images for unmanned surface-vehicle," in *Proceedings of the 2020 International Conference on Cyberworlds (CW)*, pp. 204–210, IEEE, Caen, France, September 2020.
- [148] V. Marié, I. Bechar, and F. Bouchara, *Towards Maritime Videosurveillance using 4K Videos*, Springer, Cham, Switzerland, pp. 123–133, 2018.
- [149] C. Lin, W. Chen, and H. Zhou, "Multi-visual feature saliency detection for sea-surface targets through improved sea-sky-line detection," *Journal of Marine Science and Engineering*, vol. 8, no. 10, p. 799, 2020.
- [150] V. Soloviev, F. Farahnakian, and L. Zelioli, "Comparing cnn-based object detectors on two novel maritime datasets," in *Proceedings of the 2020 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*, IEEE, London, UK, July 2020.
- [151] V. Marie, I. Bechar, and F. Bouchara, "Real-time maritime situation awareness based on deep learning with dynamic anchors," in *Proceedings of the 2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, pp. 1–6, IEEE, Auckland, New Zealand, November 2018.
- [152] S. Li, Z. Zhang, and B. Li, "Multiscale rotated bounding box-based deep learning method for detecting ship targets in remote sensing images," *Sensors*, vol. 18, no. 8, p. 2702, 2018.
- [153] P. Spagnolo, F. Filieri, and C. Distanto, "A new annotated dataset for boat detection and re-identification," in *Proceedings of the 2019 16th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, pp. 1–7, IEEE, Taipei, Taiwan, September 2019.
- [154] B. Bovcon, J. Muhovič, and J. Perš, "The mastr1325 dataset for training deep usv obstacle detection models," in *Proceedings of the 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 3431–3438, IEEE, Venetian Macao, Macau, November 2019.