Sea-Surface Object Detection Based on Electro-Optical Sensors: A Review

Hongguang Lyu* **and Zeyuan Shao***
Are with Navigation College, Dalian Maritime University, Dalian 116026, China.
E-mail: lhg@dlmu.edu.cn; szy_dlmu@163.com

Tao Cheng
Is with SpaceTimeLab, Civil, Environmental and Geomatic Engineering, University College London, Gower Street, London WC1E 6BT, U.K. E-mail: tao.cheng@ucl.ac.uk

Yong Yin
Is with Navigation College, Dalian Maritime University, Dalian 116026, China.
E-mail: bushyin@dlmu.edu.cn

Xiaowei Gao
Is with SpaceTimeLab, Civil, Environmental and Geomatic Engineering, University College London, Gower Street, London WC1E 6BT, U.K. E-mail: ucesxwg@ucl.ac.uk

xxxxxxx

Abstract— Sea-surface object detection is critical for navigation safety of autonomous ships. Electro-optical (EO) sensors, such as video cameras, complement radar on board in detecting small obstacle sea-surface objects. Traditionally, researchers have used horizon detection, background subtraction, and foreground segmentation techniques to detect sea-surface objects. Recently, deep learning-based object detection technologies have been gradually applied to sea-surface object detection. This article demonstrates a comprehensive overview of sea-surface object-detection approaches where the advantages and drawbacks of each technique are compared, covering four essential aspects: EO sensors and image types, traditional object-detection methods, deep learning methods, and maritime datasets collection. In particular, sea-surface object detections based on deep learning methods are thoroughly analyzed and compared with highly influential public datasets introduced as benchmarks to verify the effectiveness of these approaches. The article also proposes the direction of future research for sea-surface object detection based on EO sensors.

Digital Object Identifier 10.1109/MITS.2022.3198334
Date of current version: 9 September 2022

*Corresponding authors

1939-1390/22©2022IEEE

Studies on navigation technology of autonomous ships have gradually increased in the past few decades due to the continuous development of technologies for computers, communication tools, and artificial intelligence approaches [1]. Fast and accurate perceptual information is the foundation of autonomous navigation decision making for autonomous ships. Marine radar is often used as a sensor to detect obstacles on the sea, but there are some obvious drawbacks to this technology. On the one hand, due to the existence of close blind spots, some sea-surface objects close to the ship are not easy to capture. On the other hand, it is difficult for radar to extract low radar cross-section (RCS) objects in clutter environments such as rain, snow, wind, and waves.

Electro-optical (EO) sensors are further used as an excellent supplement to the shortcomings of radar because of the following reasons. First, the image and video information generated using the EO sensor is intuitive and interpretable for the watching officers on board or the shore-based center. Second, EO sensors are more adaptable to new technologies and can apply image processing technology and computer vision to achieve more intelligence [2]. However, the use of EO sensors to detect sea-surface objects has several drawbacks [3], [4], including difficulty in detecting foreground objects under complex backgrounds [5], changing the appearance of the detected object due to distance and angle of observation [6], and sensitivity of EO sensors to weather changes (e.g., illumination and sea fog) [7].

Subsequently, several studies on improving EO sensor technology for object detection have been conducted to achieve faster and more accurate sea-surface object detection. In general, a traditional maritime object-detection system consists of three modules: horizon detection, static-background subtraction, and foreground segmentation. Each of these phases is relatively challenging because of coastal interference and the dynamic of ocean waves. Recently, deep learning using different convolutional neural network (CNN) models to extract features of maritime objects has been continuously developed and has produced significant detection results. However, comprehensive reviews of sea-surface object-detection methods in the existing literature are still lacking. Subsequently, this study offers a comprehensive overview of traditional techniques and deep learning methods, including analyzing the advantages and limitations of each technique, presenting a comprehensive collection of public maritime datasets, and providing extensive guidance for the use and verification of sea-surface object-detection methods.

Related Works

Contents of Previous Research

Moreira et al. [9] introduced several maritime vessel foreground segmentation methods. Chan [8] evaluated 37 maritime background subtraction algorithms using an established dataset. Prasad et al. [2] performed a comprehensive review and evaluation of background subtraction methods based on object detection in a maritime environment. Currently, only a limited number of studies have investigated a deep learning-based sea-surface object-detection technique. Schöller et al. [10] evaluated three deep learning maritime object-detection methods, and Wang et al. [11] proposed several classic deep learning architectures and applications for sea-surface object detection.

Objectives of Previous Research

Existing investigations of sea-surface object detection using EO sensors have various objectives, with relatively few studies reporting the comprehensive collection and evaluation of sea-surface object-detection methods. Chan [8] investigated different sea-surface object-detection algorithms primarily to improve the performance of maritime background subtraction in the object-detection step. Schöller et al. [10] used three different deep learning methods to evaluate the algorithm's detection performance and classification efficiency. In addition, Prasad et al. [2] and Moreira et al. [9] introduced object-detection steps in the maritime environment and presented a foundation for subsequent research on sea-surface object tracking.

Limitations of Previous Research

Detecting sea-surface objects using computer vision technology, specifically deep learning methods, requires high-quality datasets as benchmarks. The relevant literature reviews have provided and organized datasets for the convenience of other researchers. For instance, Mittal et al. [14] collected and sorted out low-altitude drone datasets to study the problem of detecting unmanned aerial vehicle (UAV) objects. Kanjir et al. [12] and Li et al. [13] conducted literature surveys on object-detection methods in ships based on optical remote sensing images. Among those studies, Li et al. [13] collected datasets that may be used as benchmarks to verify the object detection and classification methods in ships. Considering the sea-surface objects datasets based on visible and infrared, Wang et al. [11] collected some ship datasets. However, these datasets have not been comprehensively collected and organized for sea-surface object detection.

Therefore, our review will cover four essential aspects of using EO sensors for sea-surface object detection: EO sensors, traditional object-detection techniques, deep learning methods, and maritime datasets collection, as presented in Figure 1.

Our Contributions

This study reviews the latest EO sensors-based technologies for sea-surface object detection and evaluates the technologies, highlighting their potential to develop autonomous

FIG 1 An overview of the article's content. MARDCT: Maritime Detection, Classification, and Tracking; VAIS: a dataset for recognizing maritime imagery in the visible and infrared spectrums

navigation and maritime surveillance systems for autonomous ships. The main contributions of this study, in comparison to the existing literature, are as follows (see Table 1):

- We present a comprehensive insight of traditional technologies for sea-surface object detection by rigorously comparing the advantages and disadvantages of the three processes of traditional sea-surface object-detection methods (i.e., horizon detection, static-background subtraction, and foreground segmentation).
- We define three key steps of the deep learning-based method of sea-surface object detection, (i.e., training dataset construction, object feature extraction, and model optimization) and make comprehensive analyzation and further comparation of the characteristics of various techniques in these steps.
- We collect more comprehensive visible/infrared image datasets for sea-surface object detection and discuss the method for evaluating the performance of object-detection algorithms based on the benchmark dataset.
- We analyze and discuss the challenges and future development of EO sensors-based sea-surface object-detection methods.

Table 1. A comparison of related research and the proposed work.

Year	References	Key Contents	1	2	3	4
2014	Moreira et al. [9]	Maritime vessels detection and tacking algorithms	√	√	×	×
2017	Prasad et al. [2]	Maritime objects detection and tracking methods	√	√	×	×
2018	Kanjir et al. [12]	Ship detection and classification	×	√	√	×
2019	Schöller et al. [10]	Sea-surface object-detection methods	√	×	√	×
2020	Wang et al. [11]	Architectures and algorithms for marine object recognition	√	×	√	√
2021	Chan [8]	Marine background subtraction algorithms	√	√	×	×
2021	Li et al. [13]	Ship detection and classification	×	√	√	√
2022	This article	Sea-surface object-detection methods	√	√	√	√

1: visible/infrared image; 2: background subtraction/foreground segmentation; 3: deep learning methods; 4: dataset collection; √: considered; ×: not considered.

Overview of Object Detection in the Maritime Environment

Comparison of Sensors Used in the Maritime Environment for Object Detection

An autonomous ship needs to perceive and obtain information of the environment around the ship and ultimately to navigate reliably, autonomously, and safely. Ship-sensing systems rely on GPS, automatic identification system (AIS), sonar, marine radar, lidar, and EO sensors. Among those tools, GPS is mainly used for autonomous positioning of ships [15], while AIS systems are primarily used for ship-to-ship information sensing and are not suitable for detecting sea-surface objects that are not equipped with AIS [16], [17]. The sensors commonly used for detecting obstacles in the sea include sonar, marine radar, lidar, and EO ones. Moreover, an EO sensor is divided into visible cameras and infrared optical ones. Table 2 summarizes the advantages and disadvantages of these types of sensors. Figure 2 shows the date and methods used for the first application of each sensor to the detection of sea-surface objects in the past 20 years in this survey. For infrared cameras, Toet [18] used a morphology approach to segment maritime foreground objects. For marine radar, Panagopoulos and Soraghan [19] used signal averaging, time-frequency representation, and morphological filtering for object detection. For visible cameras, Bouma et al. [20] used the change of object pixel intensity to establish a background model to detect marine targets. For sonar, Heidarsson and Sukhatme [21] processed the collected sonar data, followed by the processing and feature extraction of the overhead imagery, and then used a binary classification framework to detect marine obstacles. For lidar, Gal and Zeitouni, [22] used probability density estimation and Bayesian filters to identify and track sea-surface objects.

Navigation radar, sonar, and lidar are commonly used detection sensors for autonomous ships. Sonar is frequently employed for detecting underwater objects, rather than for detecting objects on the water's surface [23], [24]. At present, nautical radar is the most widely used detection equipment by large merchant ships, and it has great capability to detect and track objects at sea. It can well detect most objects, including objects far away from ships. However, using radar at sea also has a series of shortcomings, chiefly as follows:

- Nautical radar has some drawbacks, including different accuracy among different directions. For example, due to the existence of blind spots, it often fails to detect objects that are too close to its own ship.
- Radar is not suitable for detecting such sea-surface objects (small wooden fishing boats, rafts, and so on) that are not equipped with radar reflectors and AIS.
- Radar is not effective in extracting targets with a low RCS. Pirate ships are often difficult to track by radar because they are very small and fast, using almost nonmetallic, rigid, inflatable boats [25], [26].
- In the presence of wind and waves, some tiny objects (e.g., buoys and pontoons; see Figure 3) [27], [28] may be shielded by radar clutter suppression.
- Radar-based systems are not suitable for object detection in densely populated areas offshore because of relatively large electromagnetic emissions [29], [30].

Table 2. A comparison of sensors in sea-surface object detection.

Sensor	Distance	Advantages/Characteristics	Disadvantages/Limitations
Sonar	~1 m to several 100 m	① Close detection distance on the sea ② Mainly used for underwater detection ③ Able to detect objects with acoustic characteristics	① Needs separate systems for small-range detections ② Requires specialized user training ③ Performs poorly for objects with weak acoustic features
Nautical radar	~40 m to 72 nmi	① Far detection distance ② Detects objects with high RCSs (mostly metallic) ③ Large onboard power supply requirement ④ Adapts to severe weather and sea conditions ⑤ Provides nearly all-weather and broad-area imagery	① There is a radar blind spot at close range ② Insufficient ability to detect small objects ③ Requires specialized user training ④ Not suitable for detecting in populated areas offshore ⑤ Susceptible to high waves and water reflectivity ⑥ Cannot detect objects with a small RCS ⑦ Cannot penetrate water
Lidar	~1 m to 200 m	① Good at near-range obstacle detection ② Good in real time ③ Less affected by rain and fog than EO sensors	① The detection range is close ② Less information about object characteristics is obtained ③ Relatively high cost ④ Cannot penetrate water
Visible range EO	~1 m to several kilometers	① Processes color information ② High resolution ③ Adaptive to new computer vision and image processing algorithms ④ Simple operation	① Sensitive to illumination and weather changes ② Relatively complicated calculation ③ Difficult to detect far objects and predict their size and distance ④ A relatively short detection distance
Infrared range EO	~1 m to several kilometers	① Suitable for night detection ② Longer range than visible-range EO ③ Adaptive to new computer vision and image processing algorithms ④ Simple operation	① Indoor or evening use only ② Relatively complicated calculation ③ Difficult to detect far objects and predict their size and distance ④ A relatively short detection distance

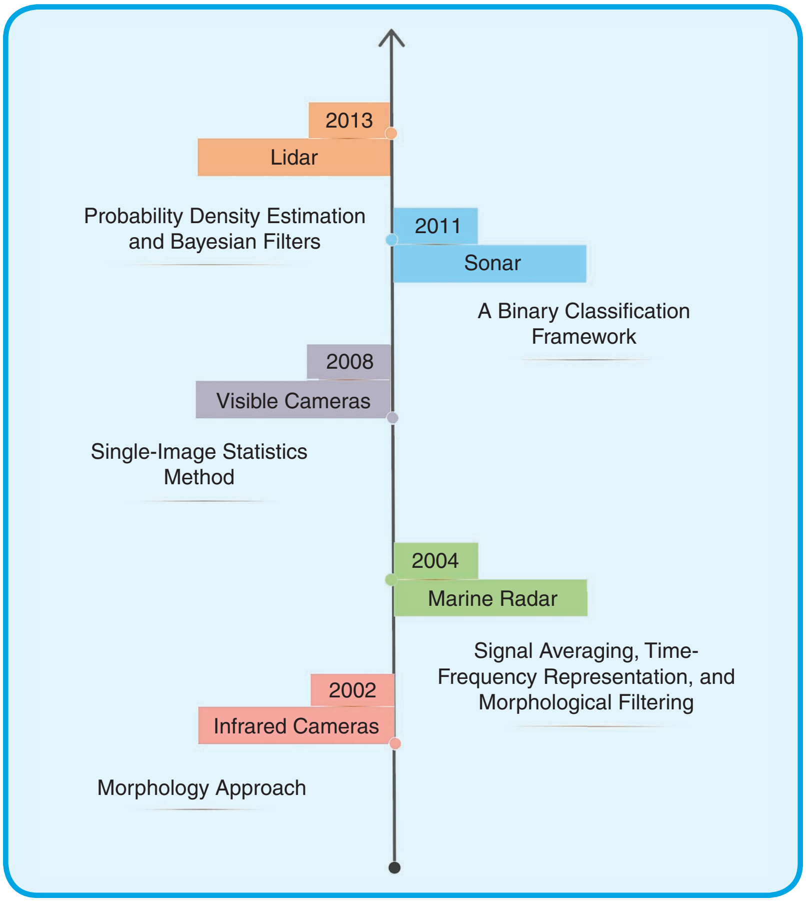

FIG 2 Sensor usage graphic milestones.

Moreover, lidar is accurate and reliable in detecting objects close to the ship; however, its cost is relatively high. In addition, the characteristic information of the obtained objects is short of intuitiveness and richness compared with the objects captured by EO sensors [27], [31], [32], [33], [34].

As presented in Table 2, EO sensors can detect sea-surface objects essential to autonomous ships. The sea-surface objects missed by radar can be detected by the images and video data acquired from the EO sensors. Such images and video data have rich feature information, intuitiveness and effectiveness, and intense immersion [25], [35], [36], [37], [38]. EO sensors can adapt to the latest image processing, computer vision, and other new technologies. In addition, EO sensors have great potential as an auxiliary observation method and have important practical significance and broad application prospects for autonomous ships [39], [40], [41], [42], [43]. Subsequently, it is necessary to introduce EO sensors into the autonomous ship-sensing system even though such sensors have some disadvantages, e.g., they are affected by illumination, weather conditions, and relatively complex calculations. Based on this, in this

FIG 3 Several sea-surface objects [2].

study, we investigated sea-surface object detection using EO sensors to provide a technical reference for the sensing module for the autonomous navigation of ships.

Sea-Surface Object Detection From EO Sensors

The overall framework of sea-surface object detection from EO sensors is shown in Figure 4. Three phases are conducted to carry out such a work: image data input, sea-surface object detection, and detection result output. First, EO sensors are used to collect the image data of maritime obstacles. An object-detection technology is further employed to detect the sea-surface object, and finally, the output is generated as information about the detected object, e.g., boats and buoys.

Using EO sensors to detect sea-surface objects depends on the object-detection technology employed. Most of the object-detection technologies are not easy to use directly to detect sea-surface objects due to complexity of the maritime environment. This study investigates various methods of sea-surface object detection divided into traditional and deep learning approaches. The traditional approach detects maritime obstacles using background subtraction technology to obtain foreground objects after preprocessing related object images (e.g., image denoising and image segmentation). The deep learning method designs a detection model for the target dataset and uses the trained model to detect the target in the image directly. The following two sections review the progress of both approaches.

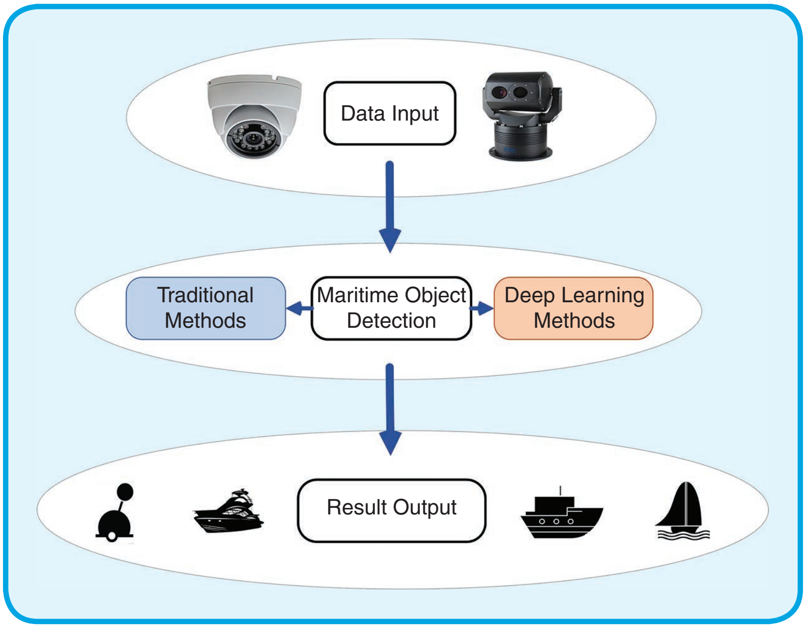

FIG 4 The framework of EO-sensors-based sea-surface object detection.

Traditional Sea-Surface Object-Detection Methods Using EO Sensors

To detect sea-surface objects using traditional techniques, three essential steps are conducted, i.e., horizon detection, static-background subtraction, and foreground segmentation. This article discusses traditional sea-surface object-detection methods from the perspective of the process (see Table 3).

Horizon Detection

Linear Fitting Approach

The linear fitting approach selects candidate points for linear fitting and then generates sea-skyline parameters. Jiang et al. [44] proposed a sea-skyline detection method based on linear fitting. First, the sea horizon information is determined by performing a histogram analysis, obtaining pixel information from near the sea horizon, and then using the linear fitting technique to remove the irrelevant pixels. Moreover, Ma et al. [45] suggested a sea-skyline extraction method based on a straight-line fitting. First, the line-segment detection method was adopted to locate the sea skyline roughly, and then the least squares approach was used to calculate the sea-skyline accurately.

The horizon-generation method based on linear fitting is less computationally intensive and has good real-time performance. Subsequently, such an approach is suitable for horizon detection in a simple sea and sky background.

Transform Domain Approach

The horizon-generation methods based on the transform domain include Hough transform, Radon transform, wavelet, and so forth. This type of process first transforms the collected image data into the corresponding transform domain and then processes the image in the transform domain. Kong [46] proposed a horizon-detection method based on the wavelet transform. Mathematical equations were established to locate the horizon's position by checking the approximate image of Haar wavelet decomposition. Tang et al. [47] used Radon transform to extract horizons based on maritime infrared image data and fuzzy comprehensive evaluation to produce the detection results. Moreover, Wei et al. [48] designed a vision-based sea-surface object-detection system in which a fast edge detection algorithm is used to generate a binary edge image of the scene. The Hough transform is further used to detect the horizon in the edge image.

Using the horizon-generation method based on the transform domain has the advantage of reliably and accurately generating the horizon in a complex environment with better robustness. The disadvantages are that the number of calculations is relatively large, and it is challenging to meet real-time requirements.

Gradient Saliency Approach

The gradient saliency approach adopts pixel characteristics of the horizon. On the other hand, the gradient amplitude of the horizon changes drastically in the vertical direction. Moreover, the pixel characteristics in the horizontal direction are the same. Wang et al. [49] proposed a gradient saliency algorithm based on a red, green, blue color space. Such an approach effectively suppressed interference factors and realized identification of the horizon through a regional growth method based on gradient direction. Lin et al. [50] used saliency of the gradient between the sea surface and the sky to enhance the image gradient through Gaussian low-pass filtering. They eliminated the influence of clouds and waves on the image gradient, determined the potential area of the horizon, and finally, used iterative polynomial fitting to generate the horizon.

The gradient-based saliency method can be carried out quickly, with a good real-time performance. However, for complex backgrounds, interference factors need to be suppressed in algorithm processing, otherwise, it is easy to cause missed detection.

Image Segmentation Approach

The horizon is generated based on the principle of image segmentation. Generally, the dividing line of the water and sky area is defined as the *horizon*. Lu et al. [51] integrated median filtering and Canny edge detection to design an adaptive threshold based on the characteristics of infrared images to segment the horizon. Liang et al. [52] used a gray-level co-occurrence matrix based on texture features to locate the sea-skyline region. It was carried out by obtaining an adaptive segmentation threshold using the maximum between-class variance method, adopting the clustering technique to select the appropriate points, and finally generating the horizon parameters by a straight-line fitting.

The horizon-generation method based on image segmentation has a small calculation and good real-time performance in a simple background. However, determining the optimal segmentation threshold of the horizon in a complex background using this approach is difficult, and the ability to resist interference factors is poor.

Information Entropy Approach

The information entropy approach is also widely used for horizon detection. Yang et al. [53] proposed an information entropy method based on variance weighting to detect horizons in infrared images, and verified the effectiveness of

Table 3. A comparison of horizon-detection methods.

Classification	References	Representative Methods	1	2	3	4
Linear fitting	Jiang et al. [44] (2010); Ma et al. [45] (2016)	Least squares method	√	×	×	×
Transform domain	Kong [46] (2016); Tang et al. [47] (2013); Wei et al. [48] (2019)	Hough, wavelet, Radon transforms	×	√	—	√
Gradient significance	Wang et al. [49] (2016); Lin et al. [50] (2020)	RGB color space	√	—	×	×
Image segmentation	Lu et al. [51] (2006); Liang et al. [52] (2015)	OTSU, canny edge detection	√	—	×	×
Information entropy	Yang et al. [53] (2006)	Edge phase encoding	×	—	√	√
Based on the feature	Jeong et al. [54] (2019); Zhan et al. [55] (2019)	CNN	√	—	√	—

1: real time; 2: detection accuracy; 3: robust; 4: complex background adaptability; √: advantages; ×: limitation; —: average. RGB: red, green, blue; OTSU: Otsu Nobuyuki.

this approach through experiments. The information entropy approach has good robustness and solid environmental adaptability, but the real-time performance is poor and the computational cost is too high.

Feature Approach

Feature-based techniques have gradually been applied to horizon detection in recent years. Jeong et al. [54] used a combination of multiscale and CNNs to detect the horizon. Zhan et al. [55] suggested a new water-boundary-determination method. First, the input image pixels were clustered into different regions through an adaptive multilevel segmentation algorithm, and a label and confidence value were assigned to each pixel. Then the obtained label map and related confidence map were input as training samples into the CNN to train the network online. Finally, the online trained CNN was adopted to segment the input image again. The experimental results show that this technique had higher accuracy and strong robustness.

Feature-based methods have a solid ability to extract targets. However, to implement such an approach, many images must be collected in advance for training and many unpredictable factors are involved in the actual scene, e.g., lousy weather and occlusion of ships.

Static-Background Subtraction

After the horizon is detected, the relevant detection area is obtained. However, due to the dynamics and complexity of the marine environment, detecting related targets in the detection area is still relatively complicated. Subsequently, first, it is necessary to remove the background of the image through background modeling to further segment the foreground and output the target-detection result [3]. Through literature analysis, the following static-background subtraction methods used for sea-surface object detection are listed in Table 4: single-image statistics, Bayesian classifier, difference operation, domain pixel, Gaussian mixture, feature based, and principal component analysis (PCA).

Single-Image Statistics Approach

The static-background-removal method based on single-image statistics is often used for the detection of sea-surface objects. Bouma et al. [20] used the change of object pixel intensity to establish a background model to detect small boats, buoys and other targets. Ren et al. [56] proposed a saliency accumulation method to detect small targets on the sea. This technique combined characteristics of the space and frequency domains and accumulated the saliency maps of consecutive frames by applying a threshold to obtain a binary saliency map. The experiment proved that the method was simple and effective. Zhou et al. [57] used a sequence-based top-hat filter model for small infrared targets at sea. Such an approach well suppressed background clutter and enhanced the detection accuracy of small infrared targets.

The method for single-image statistics is simple, does not require memory learning, and has a better effect on small-target detection. It cannot, however, solve the related problems of multimodality.

Bayesian Classifier Approach

The Bayesian classifier is a method with a low probability of misclassification and strong classification ability. Culibrk et al. [58] used a neural network system to form an unsupervised Bayesian classifier. The constructed classifier can effectively deal with complex backgrounds using motion and illumination changes. Socek et al. [59] used

Table 4. A comparison of static-background subtraction methods for sea-surface object detection.

Classification	References	Model	1	2	3	4
Single-image statistics	Bouma et al. [20] (2008); Ren et al. [56] (2012); Zhou et al. [57] (2014)	Spatial filtering, histogram	×	×	—	×
Bayes classifier	Culibrk et al. [58] (2007); Socek et al. [59] (2015)	Decision framework, evidence theory	—	—	—	×
Difference operation	Razif et al. [60] (2015)	Background difference, frames difference	√	×	—	×
Domain pixel	Borghgraef et al. [61] (2010); Adiguzel et al. [62] (2018); Tran and Le [63] (2016)	ViBe, behavior subtraction	—	—	√	—
Gaussian mixture	Wang et al.[67] (2014); Zhang et al. [68] (2012); Frost et al. [69] (2013); Zhou et al. [70] (2020)	Kernel density estimation, GMM	—	—	√	√
Based on feature	Zhu et al. [71] (2010); Nie et al. [72] (2020); Fiorini et al. [73] (2017)	Shape, texture, moment invariant	√	—	√	√
Principal component analysis	Biondi [74] (2016); Sobra et al. [75] (2015); Kajo et al. [76] (2021);	Singular value decomposition, RPCA	√	√	—	√

1: real time; 2: detection accuracy; 3: robust; 4: complex background adaptability; √: advantages; ×: limitation; —: average; ViBe: visual background extractor.

the Bayesian classifier method to estimate and suppress the marine background and evaluated the technique's applicability in a dataset of real ocean scenes. The background-removal method based on the Bayesian classifier has a simple classification, but the learning phase in this approach is relatively complicated and the training dataset is more sensitive.

Difference Operation Approach

The difference operation technique is divided into the background difference and interframe difference methods. Among them, the interframe difference method uses two consecutive frames of images in the image sequence to perform difference, and then binarizes the gray-scale difference image to extract motion information. Razif et al. [60] used the interframe difference method to remove the maritime background, which ensured a low computational complexity.

The advantage of the difference operation approach is that the algorithm is relatively simple and the operation speed is fast. However, this method does not adapt to complex environmental changes, such as the chromaticity changes caused by illumination changes.

Domain Pixel Approach

Domain pixel-based methods are also often applied to maritime background removal by considering spatial and temporal correlations. Borghgraef et al. [61] evaluated the visual background extractor (ViBe) and behavior subtraction algorithms, and the experimental results show that the background-removal effect is better than that of the traditional parametric techniques. Using the improved ViBe algorithm, Adiguzel and Ozyilmaz [62] could better remove noise in the image and augment the effect of background subtraction in the marine environment. Tran and Le [63] used the ViBe algorithm, combined with saliency-detection technology, to obtain a high detection rate on the maritime challenge dataset.

As a parameter-free method, the domain pixel technique has the advantage of less memory usage and timely initialization of the background model. The limitation is that the extraction of moving objects is incomplete in complex and changeable scenes.

Gaussian Mixture Approach

The Gaussian mixture approach can better suppress noise interference in the complex background than the other background-removal approaches [64], [65], [66]. Wang et al. [67] faced the challenge of the complex environment on the water surface and used a method based on Gaussian mixture (GMM) to remove the sea background for detecting floating objects on the sea. Zhang et al. [68] used a Gaussian mixture model (GMM) to remove the sea background with different illumination and weather. Frost and Tapamo [69] used the kernel density estimation model to remove the maritime background and obtain a good target-detection effect. Zhou et al. [70] used the Gaussian distribution in the Fourier domain to model the marine background and extracted the targets by considering the maritime background dynamics and the Gaussian discriminant coefficient, obtaining relatively accurate results on multiple marine infrared video sequences. The Gaussian mixture method can effectively remove the complex background of the sea and has high detection accuracy under a complex background, yet the modeling is more complicated and the computational cost is high.

Feature Approach

The feature-based background classifier method uses training datasets (positive and negative samples) for supervised feature learning for different target categories. Zhu et al. [71] considered ship shape and texture characteristics and obtained candidate ship targets by eliminating the background of clouds, islands, and sea clutter. Nie et al. [72] combined shape, texture, and moment-invariant features to describe ship targets more effectively, and eliminated false alarms through support vector machine (SVM) training. Fiorini et al. [73] obtained high accuracy and real-time performance on the Maritime Detection, Classification, and Tracking (MARDCT) dataset by extracting marine vessel target features, combined with a decision tree classifier.

The feature-based background classifier method has good robustness to complex backgrounds (e.g., occlusion and rotation of the target) and has been applied to various case studies. Subsequently, this approach is relatively mature. However, it has high complexity of the algorithm and poor real-time performance.

PCA Approach

PCA can well restore the data containing Gaussian noise and is often used in the analysis of video foreground and background. Biondi [74] applied the robust PCA (RPCA) technique to maritime radar images, which can reduce a large amount of redundant data. Sobral et al. [75] used a double-constrained version of the RPCA to remove the ocean background. The experimental results show that the combination of confidence maps and shape constraints can improve the effect of foreground detection. Kajo et al. [76] proposed a tensor-based singular value decomposition method for ocean background removal, which effectively handles the challenges brought by marine stationary foreground objects and ghosts by updating the separation component incremental operation and adopting a forgetting mechanism. The method based on PCA can accurately separate the foreground target in the complex image background, but the disadvantage is that it cannot maintain efficient separation continuously through the update mechanism.

Foreground Segmentation

In traditional maritime data processing, after eliminating the static background, the foreground segmentation method is used to identify the contour of maritime objects. When such a contour is further detected, the morphology approach is the most crucial foreground segmentation approach [2]. A morphological method is a nonlinear filtering technique used to process binary images and is later applied to gray-scale image processing. Four basic operations need to be carried out in the morphology approach: corrosion, expansion, and open and close operations. Based on these basic operations, various morphological algorithms can be combined and improved upon. Table 5 lists several interesting morphological methods for foreground segmentation of sea-surface objects.

Westall et al. [77] used basic morphological operations to segment the foreground of persons in distress at sea. Toet [18] used a morphological top-hat transform to identify kayaks and swimmers in complex ocean backgrounds and recommended adjusting the structural elements used in top-hat conversion to the size and shape of the target. Such an approach would significantly reduce false-detection rates. Eum et al. [78] used the Sobel edge-detection method and morphological operations to segment foreground objects to separate the maritime foreground from the background. Genitha et al. [79] offered an improved watershed segmentation algorithm using label control to avoid oversegmentation of the algorithm. This method can accurately segment ship targets in maritime remote sensing images. Kushwaha et al. [80] used wavelet transform to wavelet decomposition of images, used background modeling on approximate coefficient (LL subband), and finally, used closed-shape operators to segment marine objects.

The morphology approach extracts the corresponding shape in the image through certain structural elements, removes irrelevant structures, and finally, achieves the purpose of foreground object recognition. When using morphological methods to segment sea-surface objects, it is necessary to assume that the targets are not occluded and separated enough to ensure that the boundaries will not merge.

Sea-Surface Object-Detection Methods Based on Deep Learning

The sea-surface object-detection approach based on deep learning uses the provided sea-surface objects dataset to train the network. The network then automatically learns the parameters to detect and recognize the sea-surface objects. The previous investigations mainly focused on improving the performance of sea-surface object detection from three aspects: 1) dataset construction, 2) object feature extraction, and 3) model optimization. An extensive literature review was further carried out to analyze those existing works and is presented in Table 6.

Training Dataset Construction

Sea-surface objects data based on visible and infrared, especially image data of small fishing boats and buoys, are not extensive. Subsequently, three approaches are generally used to construct sea-surface objects training datasets: 1) specific datasets based, 2) traditional image augmentation based, and 3) generative adversarial network (GAN) based (see Table 7). These approaches promote the research of maritime obstacle object detection based on deep learning, increase the utilization rate of existing data, and further enrich the types of sea-surface objects datasets and the diversity of object shapes.

Specific Dataset Approach

CNNs can extract many distinguishable features through a series of convolution and pooling layers. However, object-detection performance is closely related to large-scale, high-quality datasets because a CNN is a data-driven method. The detection performance of some objects can be significantly improved (e.g., fish image [81], vehicle image [82], and pedestrian detection image datasets [83], [84]) by making datasets of specific objects. In addition, related research has developed the maritime obstacle object image dataset. Table 8 briefly introduces five classic maritime datasets for deep learning (the details of the published datasets are listed in the "Public Maritime Datasets" section).

Shao et al. [85] established a new large-scale ship dataset called *SeaShips*, which accurately annotated the bounding boxes of six types of sea ships and described the detailed design of the dataset. Liu et al. [86] established a sea buoy dataset named *SeaBuoys*, which contains six different types of buoys. Ribeiro et al. [87] proposed SeaGull, a sea-surface objects image dataset captured by a small drone, and used different cameras to obtain video sequences from different heights and different perspectives. Zhang et al. [88] offered a publicly available maritime image VAIS, a dataset for recognizing maritime imagery in the visible and infrared spectrums, which contains both

Table 5. Morphology-based maritime foreground segmentation methods.

References	Morphological Method	Foreground Object
Westall et al. [77] (2008)	Dilation, erosion, open and close operations	Persons in distress at sea
Toet [18] (2002)	Top-hat transform	Kayaks, swimmers
Eum et al. [78] (2015)	Sobel edge detection	Ship
Genitha et al. [79] (2020)	Watershed segmentation	Ship
Kushwaha and Srivastava [80] (2015)	Close operations	Ship

Table 6. A comparison of different sea-surface object-detection methods based on deep learning.

Classification	References	Network Model	Network Design	Dataset Enhancement	Model Optimization	1	2	3	4
Basic classification network	Pan et al. [133] (2020)	ResNet	Use attention mechanism (RMA)	—	—	×	√	×	×
	Ma et al. [125] (2020)	ResNet	Hybrid ResNet and DenseNet	—	—	×	√	×	×
	Xu et al. [134] (2020)	VGG	Use attention mechanism (MF-APN)	—	—	×	√	×	√
	Fiorini et al. [143] (2017)	VGG	Use SVM	Expand training samples	Fine-tuning of the weights	×	√	×	×
	Chen et al. [171] (2020)	CFCCNN	Introduce RHS mechanism	—	—	×	√	×	×
Based on Two stages	Chen et al. [116] (2018)	Fast R-CNN	—	—	—	×	×	×	×
	Qi et al. [94] (2019)	Faster R-CNN	Hierarchical-narrowing network	Downscaled image	—	√	√	√	×
	Farahnakian and Heikkonen [124] (2020)	Faster R-CNN	Contrast feature fusion system	—	—	×	×	×	×
	He et al. [92] (2018)	Faster R-CNN	—	Expand negative samples	—	×	√	×	√
	Gao et al. [91] (2019)	Faster R-CNN	—	Expand negative samples	—	√	√	×	√
	Fu et al. [93] (2018)	Faster R-CNN	—	Expand small-object data	—	×	√	√	√
	Shin et al. [90] (2020)	Mask R-CNN	—	New image synthesis	—	×	√	×	×
	Guo et al. [131] (2020)	Rotational Libra R-CNN	Balanced FPN	—	—	×	√	×	√
	Zhao et al. [127] (2019)	DCNet	Fusion of DNet and CNet	—	—	×	√	×	√
Based on One stage	Chen et al. [96] (2020)	YOLO	—	Image translation, rotation/adding noise	—	×	√	×	√
	Chen et al. [102] (2020)	YOLOv2	—	Based on a GAN extension	Anchor box	×	√	√	√
	Shao et al. [129] (2020)	YOLOv2	Consider coastline features	—	—	√	√	×	√
	Li et al. [126] (2020)	YOLOv3	Use spatial separation convolution	—	—	√	√	×	×
	Wang et al. [95] (2019)	YOLOv3	Introduce the CFE module	Picture flip, cut, scale adjustment	Loss function	√	√	√	√
	Li et al. [97] (2020)	YOLOv3	Fusion DenseNet	Picture brightness and rotation adjustment	—	×	√	×	√
	Huang et al. [128] (2020)	YOLOv3	Use a jump-connection mechanism	Guide filter, change picture gray scale	Anchor box	×	√	×	√
	Huang et al. [98] (2020)	RDCNN	FPN	—	Activation function	√	√	×	×
	Liu et al. [86] (2020)	YOLOv3	Join the PANet fusion structure	Picture flip, zoom	Anchor box, loss function	×	√	×	×
	Fu et al. [135] (2021)	YOLOv4	Join the CBAM module	—	—	×	√	×	×
	Liu et al. [136] (2021)	YOLOv4	Introduce the RDSC module	—	—	√	√	×	√

1: speed increase; 2: accuracy increase; 3: improvement of small-object-detection ability; 4: enhanced detection capabilities in complex backgrounds; √: considered; #: not considered; —: not mentioned; CBAM: convolutional block attention module; CFE: comprehensive feature enhancement; CFCCNN: coarse-to-fine cascaded convolution neural network; FPN: feature pyramid network; MF-APN: multi-feature attention proposal network; RDCNN: regressive deep CNN; RDSC: reverse depthwise separable convolution; RMS: random heuristic selection; RMA: ResNet-Multiscale-Attention.

Table 7. A comparison of training dataset construction methods.

Approaches	Brief Description	Advantages	Limitation
Specific dataset	Build a dataset for specific target objects in the maritime environment	☑ High image quality and uniform resolution; able to provide test benchmarks for related algorithms	☒ Expensive; requires a lot of manpower and material resources
Traditional image data augmentation	Perform image processing operations such as geometric transformation and optical transformation on the data	☑ Simple and convenient operation	☒ The original distribution of the data may be changed; some clutter information will be mixed in
GAN	Learn the distribution and structure of the original dataset through the deep network model	☑ Able to automatically generate sea-surface objects and improve the image diversity of sea-surface objects	☒ The generator and the discriminator need to be kept in sync; the interpretability is poor; the model is easy to collapse

infrared and visible images. Iancu et al. [89] built a dataset dominated by maritime vessels, containing multiple vessel types as well as some buoys and floating objects at sea.

The specific dataset approach builds a dataset for specific environments and objects. The image quality is high and the resolution is uniform. It can provide test benchmarks for various sea-surface object-detection algorithms. It does, however, consume considerable workforce and material resources.

Table 8. The construction of a maritime dataset for deep learning.

References	Datasets	Dataset Construction
Shao et al. [85] (2018)	SeaShips	Six types of ship targets (ore carrier, bulk cargo carrier, general cargo ship, container ship, fishing boat, and passenger ship)
Liu et al. [86] (2021)	SeaBuoys	Six types of buoys (buoy_1, buoy_2, buoy_green, buoy_red, buoy_blue, and buoy_yellow)
Ribeiro et al. [87] (2019)	SeaGull	Cargo ships, smaller boats (27 m long), sailing yachts, life rafts, dinghies, and a hydrocarbon slick
Zhang et al. [88] (2015)	VAIS	Merchant ships, sailing ships, medium passenger ships, medium "other" ships, tugboats, and small boats
Iancu et al. [89] (2021)	ABOShips	Contains 11 types of sea-surface objects, including nine types of vessels, seamarks and miscellaneous floaters

Traditional Image Augmentation Approach

At least five data processing methods using traditional image augmentation technology have been developed (see Figure 5). The data source expansion approach adds more related images in the image dataset, directly supplementing in quantity. The geometric transformation technique includes translation, flipping, scaling, and segmentation of the image. On the other hand, adding noise randomly adds Gaussian noise, salt-and-pepper noise, and so on to the image, whereas the optical transformation method converts image brightness, contrast, saturation, and color space. In addition, using Unity and Unreal Engine 4 software to simulate the target object can also achieve expansion of the image dataset.

Aiming at the sea-surface target image, Shin et al. [90] proposed a method to expand the sea-surface target image automatically. Gao et al. [91] and He et al. [92] realized the expansion of data sources by increasing negative samples. Based on increasing the number of negative samples, Fu et al. [93] also increased the number of small targets at sea and proved, through experiments, that data source expansion could effectively enhance the detection capabilities of ships at sea. Qi et al. [94] performed a scale-reduction operation on maritime images, Wang et al. [95] adopted geometric transformations such as flipping and shearing on maritime images, and Chen et al. [96] enhanced the dataset by adding noise to the picture. Li et al. [97]

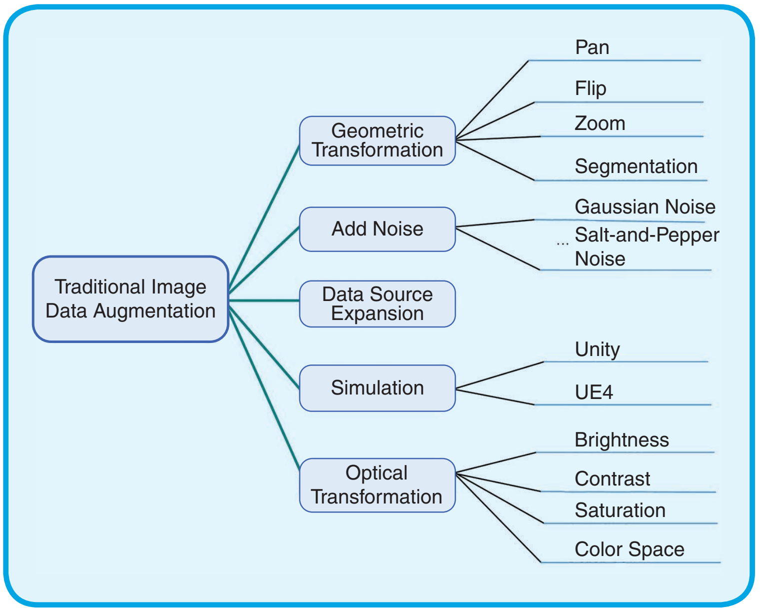

FIG 5 Traditional dataset-enhancement methods.

enhanced the dataset by changing the brightness and rotation of the picture. Huang et al. [98] used guided filtering to enhance the gray level of the picture to expand the data.

The image augmentation approach is convenient and straightforward to operate. However, the original distribution of the data is changed while expanding the data, and the introduced clutter information result in specific false alarms.

GAN Approach

Data enhancement can also be achieved by simulating the visible/infrared image of the sea through the GAN to solve the problem of insufficient target image data of some obstacles on the sea. A GAN uses a CNN to learn the internal statistical laws of sample data and obtains a probability distribution model. Such a probability model generates fake samples that can deceive the discriminator and generate more sample datasets [99].

Schwegmann et al. [100] proposed a generative confrontation network (InfoGAN) based on information maximization successfully extending a ship's remote sensing images. Zou et al. [101] used multiscale Wasserstein distance and gradient loss to improve the original GAN. The augmented network (auxiliary classifier GAN) can constantly generate high-resolution synthetic aperture radar ship images. For visible and infrared images, Chen et al. [102] used Wasserstein GAN with gradient penalty based on GAN improvement combined with GMM to expand the image data of small ships. Experimental results show that this method can generate many information-rich, small-ship samples, which indirectly improve the detection ability of small objects on the water surface.

The GAN-based approach can enhance the diversity of sea-surface objects image data by understanding the related data distribution and its potential structure. However, the training is complex and the generator and the discriminator need to be synchronized. Moreover, the algorithm results are not easy to converge in the image synthesis process.

Object Feature Extraction

The object feature extraction mainly uses a CNN to extract the high-level semantic information, shallow information in the images, and the core of the deep learning algorithm. In recent years, CNNs have made many breakthroughs. In 2012, Krizhevsky et al. [103] proposed the AlexNet model and won the ImageNet Large-Scale Visual Recognition Challenge, proving the effectiveness of CNNs under complex models, thereby establishing the leading role of CNNs in the field of computer vision. The representative models of CNNs include LeNet [104], AlexNet [105], [106], [107], VGGNet [108], Google's Inception series [109], [110], [111], ResNet [112], DenseNet [113], and so on. CNNs are commonly used for object detection and can be divided into two categories: 1) two-stage methods, e.g., region-based CNNs (R-CNNs) [114], fast R-CNNs [115], [116], and faster R-CNNs [117]; and 2) one-stage techniques, e.g., You Only Look Once (YOLO) [118], YOLOv2 [119], YOLOv3 [120], Retina-Net [121], single shot multiBox detector [122], and YOLOv4 [123].

For a target image in the complex ocean environment, feature extraction mainly solves the following three problems.

- How are object features in multimorphology, multiscale, and multiresolution situations. efficiently extracted?
- How are the feature extraction capabilities for small objects in maritime images improved? Incidentally, the international organization Society of Photo-Optical Instrumentation Engineers defines the relative size of small objects as an object area of less than 80 pixels in a 256×256 image. The COCO dataset defines the absolute size of small objects as objects with a size smaller than 32×32 pixels.
- How are features reasonably extracted to reduce the impact of complex backgrounds, especially shore buildings, foggy weather, and so on?

Considering these three problems, researchers have widely used feature fusion and the method of inserting convolution modules to improve sea-surface object detection in the object feature extraction step, as depicted in Table 9.

Feature Fusion Approach

In the process of using CNNs to extract image features, shallow feature maps have high resolution but lack semantic information, and deep feature maps have low resolution but rich semantic information. Subsequently, feature fusion methods are often used in visible-image object detection to fuse information from different feature layers to improve target classification and detection due to the variable shapes of maritime obstacle objects and the significant difference in object size. Among those approaches, ultradense connections and feature pyramids are representative of this type of method.

Ultradense Network Approach

Under the premise of ensuring the transmission of information between layers in the network, the ultradense network can make more effective use of features by adding connections. Farahnakian and Heikkonen [124] compared several classic fusion systems for the problem of feature extraction of sea-surface objects and proved the superiority of the feature fusion system. Li et al. [97] and Ma et al. [125] integrated DenseNet into the original network model and Li et al. [126] used a spatial separation convolution instead of a standard convolution to improve the feature pyramid network (FPN) based on adding DenseNet. In addition to DenseNet, several feature fusion architectures have also been applied to feature extraction of marine objects, e.g., Zhao et al. [127] combined two networks,

Table 9. A comparison of feature extraction methods for sea-surface objects

Classification	Approaches	Description	Advantages	Limitations
Feature fusion	① Ultradense network ② Feature pyramid network	Strengthen feature delivery by increasing network connections	☑ Able to effectively unify multiscale representation and semantic distribution; improve the detection ability of dense, small targets	☒ Increased network complexity; affects the balance of model calculation and accuracy
Convolution module	① Attention mechanism	By inserting the convolution module in the deep network	☑ It can effectively focus on the target, suppress other interference information in the image, and improve the accuracy of target detection	☒ Increased network complexity; affects the balance of model calculation and accuracy

DNet and CNet; Huang et al [128] introduced a jump-connection mechanism in the original network model to improve the feature extraction performance of sea-surface objects. Moreover, adding maritime saliency features to the network model can also enhance the performance of sea-surface objects feature extraction. For example, Shao et al. [129] integrated horizon features into CNNs, which reduced the extraction image area and enhanced the feature extraction capabilities of ships at sea.

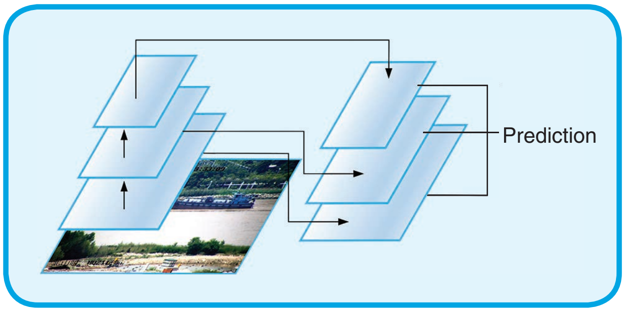

FIG 6 The basic block diagram of an FPN [130].

Feature Pyramid Approach

The FPN [130] transfers deep semantic information from top to bottom to the underlying feature map to enhance the semantic information of the underlying feature map. It is a feature fusion structure in target-detection models such as YOLOv3. Figure 6 shows the basic framework of the FPN. By improving the FPN, the problems of multiform and multiscale marine targets eliminated.

Liu et al. [86] used an across-PANet fusion structure. PANet added a bottom-up feature fusion path based on the FPN, enhancing the semantic information and location information of the feature map. Huang et al. [98] proposed an enhanced network model regressive deep CNN based on the YOLO series to improve the FPN. The augmented feature layer network can detect and predict the input sea-surface objects image on two scales. Guo et al. [131] further proposed a balanced pyramid method to solve different sizes and dense distribution of ships in an image. The ship's rotation-angle position information is finally better predicted by balancing the three networks of feature, sample, and target levels.

The method, based on feature fusion, effectively unifies semantics and multiscale representation through interlayer fusion and improves the ability to detect small targets at sea. Its limitation lies in increasing complexity of the network and reducing the speed of object detection to a certain extent.

Convolution Module Approach

Many researchers improve performance of the object-detection model by inserting a convolutional attention module [132]. This method draws on the human visual mechanism and obtains deeper semantic information by paying attention to the local information of the target. It also plays an essential role in the image detection of small sea-surface objects with few pixels.

Pan et al. [133] proposed a fine-grained root-mean square classification model classifying and recognizing different navigation marks on the water surface and using the attention mechanism to obtain deeper semantic information of the target. Xu et al. [134] replaced the APN attention mechanism model in the network with multi-feature-APN, effectively improving the accuracy of sea-surface objects recognition. Fu et al. [135] used the YOLOv4 model to detect sea-surface objects and improved the accuracy of object detection by adding an attention mechanism module, convolutional block attention module (CBAM), to the network. Liu et al. [136] also used the YOLOv4-based network framework and proposed the reverse depthwise separable convolution (RDSC) module. By using RDSC at the correct network layer, accuracy of the sea-surface target-detection model was improved as was real-time performance, to a certain extent.

The attention mechanism approach obtains deep semantic information of the sea-surface object's image and increases complexity of the network model. Maintaining the balance between the amount of calculation and accuracy is a problem with this approach.

Model Optimization

To replace the feature network connection layer and insert the convolution module, the complex network design technique also includes optimizing the loss function, classifier, and anchor box, as presented in Table 10.

Anchor-Box Design Approach

Designing the anchor box is essential to selecting the candidate area in the two stages and generating a priori a box in one stage. Choosing a suitable anchor box can reduce the false-alarm rate and the missed-detection rate of the target. The anchor-box approach needs to assign a binary label to each anchor point.

Liu et al. [86] introduced in detail three deep learning-based anchor-box design methods: 1) threshold, 2) average, and 3) select all. The experimental results proved that the average and select-all methods could effectively improve the average detection accuracy of sea-surface objects. Huang et al. [128] used the k-means++ algorithm to cluster the ship dataset. They found that clustering can accelerate convergence of the network and effectively upgrade gradient descent during the training process compared with artificially selecting anchor point values. In addition, Chen et al. [102] used a density-based clustering algorithm, density-based spatial clustering of applications with noise (DBSCAN), for poor recognition of irregular objects by the k-means algorithm and the problem of hyperparameter selection, improving the problem and saving the time of artificial adjustment.

Using anchor points in the detection model, an anchor box that is more suitable for the dataset is generated. Such a phase can effectively improve performance of the sea-surface object-detection model.

Loss Function Design Approach

Loss function of the object-detection model is composed of three parts: 1) bounding-box loss, 2) category loss, and 3) confidence loss. The purpose of the loss function is to modify the prediction box so that it is closer to the actual box. Among those methods, intersection over union (IoU) [137], improvement variants of generalized IoU (GIoU) [138], and distance IoU [139] are widely used to measure the similarity of the proper and predicted boxes in the object-detection task. The IoU calculates the intersection and union ratio of predicted box A and real box B, as shown in Figure 7.

The calculation formula can be expressed as

$$IoU = \frac{A \cap B}{A \cup B} \tag{1}$$

The loss is expressed as

$$\text{Loss} = 1 - IoU \tag{2}$$

Wang et al. [95] used the GioU to replace the IoU in the YOLOv3 model to solve the defect with no return gradient when two sea-surface objects do not intersect. Liu et al. [86] also used the GioU to redesign the loss function and verified it through experiments. The experimental results showed that this approach could effectively improve the detection performance of maritime obstacle objects.

The loss function design approach further enhances regression accuracy by considering geometric factors, e.g., the overlapping area of the bounding-box regression, and distance between the center points and aspect ratio, hence improving the performance of sea-surface object detection.

Classifier Design Approach

The use of classifiers is an essential step in sea-surface object detection and classification. The most common classifier methods include SoftMax, SVMs, k-nearest neighbors (KNNs), and extreme learning machines).

Gallego et al. [140] combined the neural network code extracted by CNNs with the KNN method to extract and classify features of ships at sea and compared the classification results obtained by KNNs with the results output by SoftMax. Kumar and Sherly [141] and Shi et al. [142] further improved overall performance of the network model by attaching the pretrained CNN to the SVM. Fiorini et al. [143] used the SVM classifier to replace the last layer in the VGG16 network to classify different ship targets on the

Table 10. A comparison of model optimization methods.

Classification	Representative Method	Brief Summary
Anchor-box design	k-means, average method, select-all method, DBSCAN	Through the design of anchor points, an anchor frame that is more suitable for the model is generated
Loss function	IoU, GIoU, DIoU	Optimize the detection model by considering bounding-box regression factors
Classifier design	SoftMax, SVM, KNN	According to the characteristics of different datasets, select the appropriate classifier

IoU: intersection over union; GIoU: generalized intersection over union; DIoU: distance intersection over union.

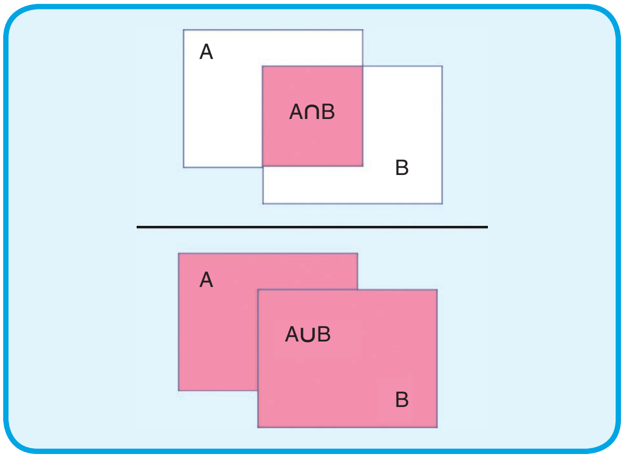

FIG 7 An IoU diagram [137].

sea and obtained high accuracy. Ji et al. [144] compared the performance of KNNs, the Bayesian classifier, and a backpropagation neural network classifier and proposed a combination of SVM strategies. The suggested approach has been proved through experiments as a powerful approach in sea-surface object detection and classification. The classifier design approach improves the classification efficiency of sea-surface objects by selecting the appropriate classifier and choosing different classifiers according to the characteristics of different datasets.

Public Maritime Datasets

Large-scale training datasets are the primary way to use deep learning methods for identifying sea-surface objects; they are also essential for research on the improvement of sea-surface object-detection technology. Most of the object-detection algorithms based on deep learning have been trained on the PASCAL VOC dataset and utilized the transfer learning method to use the PASCAL VOC dataset as the source domain dataset in the pretraining stage. The PASCAL VOC dataset comprises four categories: people, indoor objects, animals, and vehicles as well as 20 corresponding subcategories. It can be determined from the aforementioned object categories that it does not include sea-surface obstacles such as buoys, islands, and reefs. Therefore, this dataset cannot effectively detect sea-surface obstacles. To achieve better detection accuracy, maritime datasets should be used effectively. In recent years, some research institutions have publicly published their maritime datasets to promote sea-surface object-detection technology development. These maritime datasets were developed based on visible and infrared images, providing researchers with test benchmarks to verify different sea-surface object-detection algorithms. Ten image datasets can be used for sea-surface object-detection research, as presented in Table 11.

VAIS (http://vcipl-okstate.org/pbvs/bench/) is the world's first publicly available maritime dataset and was developed by pairing visible and infrared images. This dataset contains 2,865 ship images, including 1,623 visible and 1,242 infrared ones. Moreover, 16 fine-grained categories are annotated in the dataset, which can classify different types of ships at sea.

MARVEL (https://github.com/avaapm/marveldataset2016) is the largest fine-grained ship identification dataset, containing 2 million visible images of ships obtained from the foreground angle of view, which are divided into 26 categories that can be used to classify marine ship targets.

SeaShips (http://www.lmars.whu.edu.cn/prof_web/shaozhenfeng/datasets/SeaShips%287000%29.zip) is a collection of 31,455 visible images in the sea ship dataset, of which 7,000 are disclosed, including six ship types: ore carrier, bulk cargo carrier, general cargo ship, container ship, fishing boat, and passenger ship. These datasets are used for the detection and classification of sea-surface objects.

Table 11. A compilation of public maritime datasets on the whole network.

Sequence	Dataset	References	Camera Type	Format	Resolution	Brief Description	Application
1	VAIS	Zhang et al. [88] (2015)	Visible/infrared	Image	Random	Contains 2,865 ship images (1,623 visible/1,242 infrared), divided into 16 categories	Classification
2	MARVEL	Leclerc et al. [146] (2018)	Visible	Image	1,024 × 768	Contains 2 million ship images, divided into 26 categories	Classification
3	SeaShips	Shao et al. [85] (2018)	Visible	Image	1,920 × 1,080	Contains 31,455 ship images (7,000 images publicly available)	Classification/detection
4	ABOShips	Iancu et al. [89] (2021)	Visible	Image	1,920 × 720	Contains 9,880 images, including nine types of vessels, seamarks, and miscellaneous floaters	Classification/detection
5	SMD	Prasad et al. [2] (2017)	Visible/infrared	Video	1,920 × 1,080	Includes 11 videos on board and 70 videos on shore	Detection/tracking
6	MODD1	Kristan et al. [150] (2016)	Visible	Video/image	640 × 480	Composed of 12 video sequences, providing a total of 4,454 fully annotated frames	Detection/tracking
7	MODD2	Bovcon et al. [151] (2018)	Visible	Video/image	1,278 × 958	Consists of 28 video sequences, providing a total of 11,675 stero frames	DETECTION/TRACKING
8	IPATCH	Chan [8] (2021)	Visible/infrared	Video	Random	Contains 59 video sequences of anomalous activities in the sea where pirates attacked merchant ships	Detection/tracking
9	SeaGull	Ribeiro et al. [87] (2019)	Visible/infrared	Video/image	1,024 × 768	Contains more than 150,000 object images of maritime obstacles	Classification/detection/tracking
10	MARDCT	Bloisi and Iocchi [147] (2009)	Visible/infrared	Video/image	Random	Multiple sources (fixed, moving, and pan-tilt-zoom cameras)	Classification/detection/tracking

SMD: Singapore Maritime Dataset; MODD1: maritime obstacle detection dataset 1.

ABOShips (https://www.fairdata.fi/en/), as opposed to the SeaShips dataset, contains more object classes. In addition to containing nine different types of ships, it also includes two types of targets: buoys and floating objects. Among them, 11 categories of objects in the dataset are accurately annotated.

The Singapore Maritime Dataset [(SMD)—dilipprasad] (https://sites.google.com/site/dilipprasad/home/singapore-maritime-dataset) consists of 11 onboard videos and 70 onshore video sequences. It provides visual-optical (VIS) and near-infrared (NIR) video data for the verification of sea-surface object detection and tracking algorithms.

The marine obstacle detection dataset (MODD) 1 (http://www.vicos.si/Downloads/MODD) is mainly composed of 12 video sequences, providing a total of 4,454 fully annotated frames with a resolution of 640 × 480 pixels. The dataset and annotations and the MATLAB evaluation code are publicly available on the Internet. The video data are mostly taken on a 2.2 m small USV that can be used for marine semantic segmentation and the detection and tracking of maritime obstacles.

MODD2 (https://box.vicos.si/borja/viamaro/index.html) consists of 28 video sequences with a total of 11,675 images, and the image pixels are 1,278 × 958. The dataset is a multimodal marine obstacle detection dataset captured by a USV, which contains a variety of marine weather conditions, extreme situations, and a large number of small sea-surface obstacles.

IPATCH (https://drive.google.com/drive/folders/1d8VcnLj-aizWCsg0-LB3QNpBoY14gfDM) uses small speedboats to simulate pirate attacks and evaluates how the ship-sensing system detects pirate ships to determine their threat levels. IPATCH contains video sequences captured by visible and infrared cameras and can be used for evaluation and verification of maritime dynamic object detection and tracking algorithms.

SeaGull (http://vislab.isr.tecnico.ulisboa.pt/seagull-dataset) is a marine surveillance image dataset captured by a small UAV. It contains more than 150,000 images of marine obstacles and targets, including cargo ships, smaller boats (27-m long), sailing yachts, life rafts, dinghies, and a hydrocarbon slick. This dataset can be used for maritime target detection and classification based on low-altitude UAVs.

MARDCT (http://labrococo.dis.uniroma1.it/MAR/) provides data from intelligent surveillance systems in the maritime environment and fixed, mobile, and pan-tilt-zoom cameras. These datasets are used to collect videos and images in different marine scenarios containing 24 types of ship images sailing in Venice, and hence can be used for ship detection, classification, and tracking.

Relevant scholars evaluated the performance of the algorithm based on the maritime dataset benchmark. Cane and Ferryman [145] evaluated various semantic segmentation networks on the MODD1, SMD, IPATCH, and SeaGull datasets to compare the recognition performance of different network models. Zhang et al. [88] used CNNs to classify and identify the 16 ship categories in the VAIS dataset. Leclerc et al. [146] also adopted CNNs to classify ship categories on the MARVEL dataset. Bloisi and Iocchi [147] and Bloisi et al. [148] evaluated a fast and effective background-elimination independent multimodal background subtraction algorithm in the MARDCT dataset. This method is specially designed for marine scenes and can better suppress marine noise. Bloisi et al. [149] established the ARGOS classification benchmark on the MARDCT dataset used to classify ships. Prasad et al. [3], [4] verified a new horizon-detection technique based on multiscale cross-modal linear features and 23 classic background subtraction algorithms on the SMD dataset. They pointed out that a specific background subtraction method is required for complex maritime environments. Ribeiro et al. [87] evaluated sea-surface object detection and tracking algorithms on the SeaGull dataset sequence and defined the baseline performance. Kristan et al. [150] adopted the Markov random field framework in the MODD1 dataset and derived parameters of the optimization model. Bovcon et al. [151] derived equations for projecting the horizon into an image, proposed an efficient algorithm for maritime obstacle detection, and validated it on the MODD2 dataset. Chan [8] evaluated the performance of 37 background-elimination algorithms on the IPATCH dataset. The experimental results show that the multifeature category background subtraction algorithm can better adapt to challenges of the complex maritime environment and achieve the best results. In addition, Chan [152] also developed an efficient filtering method that meets the requirements of maritime vision applications. The proposed approach is based on the dark channel prior, which further improves overall performance of the background subtraction algorithm.

For deep learning object-detection algorithms, a few public maritime image benchmark datasets are evaluated. Table 12 lists several mainstream maritime object algorithms based on deep learning, all of which are strong baselines on the corresponding dataset benchmarks. Shao et al. [85] obtained the best detection accuracy on the SeaShip dataset using a faster R-CNN model with the backbone network ResNet101. Liu et al. [136] selected 7,000 images in the SeaShip dataset and tested them with the YOLOv4 model; the mean average precision (MAP) value reached 0.928, and the real-time performance was 55 frames per second. Moosbauer et al. [153] relabeled and processed the SMD dataset and established a deep learning-based maritime image dataset benchmark. Among them, the mask R-CNN algorithm shows great potential in detection accuracy, achieving scores of 0.875 and 0.877 on the VIS and NIR subsets, respectively. In addition, Iancu et al. [89] also demonstrated the superior performance of the faster R-CNN model in sea-surface object detection, and when the backbone network used Inception ResNet V2, the recognition effect reached the best performance in the ABOShips dataset.

Open datasets are beneficial to promote the development of related research and also provide benchmarks for the performance evaluation of different object-detection algorithms. The emergence of a series of maritime datasets, such as SeaShips and SMD, has effectively supported the development of maritime object-detection technology based on EO sensors and improved the comparability and reproducibility of maritime object-detection algorithms.

Summaries, Future Directions, and Conclusions

Summaries

Through the analysis of the large amount of literature mentioned previously, some developments of object detection based on EO sensors in maritime research can be summarized as follows:

- A part of the research considers the extraction of salient features at sea to improve overall performance of the object detection, mainly reflected in the extraction of horizon features.
- The accuracy and speed of the sea-surface object-detection algorithm have been considered in many articles; however, the balance between accuracy and speed is not discussed extensively.
- Aiming at the problems of video instability and missed target detection caused by the EO sensor affected by ocean waves, the performance of the maritime object-detection algorithms based on electronic image-stabilization technology has been continuously improved.
- When designing the maritime object-detection algorithm, some factors specific to the maritime environment have been considered, such as ship occlusion, coastal building interference, image blur caused by sea fog, changes in viewpoint, and illumination.
- A part of the research explored the attention mechanism model to obtain deeper semantic information of maritime obstacle objects.
- Many studies have considered the detection of sea-surface objects based on multiscales, however, for small sea-surface objects, good detection results have still not been achieved.
- Related research institutions have produced several maritime datasets to expand the image data of maritime obstacles and quantitatively compare object-detection algorithms based on the dataset.
- Given the small amount of data in some maritime obstacle datasets, training methods based on weak supervision and unsupervised have been considered.

Meanwhile, there are some primary limitations and challenges for detecting sea-surface objects based on EO sensors, which are summarized into the following aspects described in the next sections (see Figure 8).

Accurate Detection of Small Objects at Sea

Compared with vehicles on land, ships take more time making a maneuvering decision to executing it and achieving the final desired effect. Therefore, the precise identification of small objects on the sea can increase the ship's response time, thereby improving navigation safety. Traditionally, the background-elimination method using denoising, segmentation, and other preprocessing—combined with horizon extraction—identifies small objects to obtain better robustness. In deep learning, several improvement algorithms with significant effects are proposed and divided into two types: one type designs a multiscale neural network to extract features of different levels, adapting to different-size object-detection tasks. The other uses transposed convolution to expand the depth feature map. These methods better eliminate object-scale change but still cannot detect small objects well.

EO Sensors Always Shake With the Waves

When an autonomous ship is sailing at sea, the EO sensor is not only affected by wind, waves, and currents but also by the vibration caused by operation of the main engine and other mechanical equipment. At the same time, due to the small field of view of the EO sensor, slight shaking will cause the collected video image to shake and become blurred, causing the target in the image to rotate and translate on the plane, increasing the difficulty of detecting and tracking sea-surface objects.

Table 12. A baseline of maritime object-detection algorithms based on deep learning.

Dataset	Amount	Studies	Graphics Card	Models	Backbone	Indicators	MAP	Frames Per Second	F-Score
SeaShip	31,455	Shao et al. [85] (2018)	Titan Xp×4	Faster R-CNN	ResNet101	VOC	0.9240	7	—
SeaShip	7,000	Liu et al. [136] (2021)	Geforce RTX 2080TI	YOLOv4	CSPDarknet53	VOC	0.9280	55	—
SMD	31,653	Moosbauer [153] (2019)	—	Mask R-CNN	ResNet101	COCO	—	—	VIS:0.875 NIR:0.877
ABOShips	9,880	Iancu[89] (2021)	—	Faster R-CNN	Inception ResNet V2	COCO	0.3518	—	—

—: not mentioned; MAP: mean average precision.

The Effect of Strong Light Reflection From Sea Surface on Imaging

Lighting conditions directly affect imaging of the EO sensor. In addition to scenes with strong light exposure at sea, there will also be reflections on the sea surface. After the sunlight is reflected by the specular element on the water surface, it will produce a strong radiation signal. When the wind passes, the near-smooth water surface will form an inclination to produce flares or formation of flash points on top of the waves, which are all caused by specular reflection of the water body. For object detection in the maritime environment, the phenomenon of sea-surface flare is inevitable, and the radiation intensity is much greater than that of other marine objects, which easily causes saturation and distortion of the optical imaging sensor.

The Impact of Rain and Fog on Imaging

Rain and fog are common weather phenomena at sea. Autonomous ships sail in coastal waters at low-visibility conditions. Generally, information about surrounding targets is obtained through navigation aids such as radar and AIS, but the information obtained is less intuitive. Video images obtained by the EO sensor are rich in feature information, which can achieve the same effect as a visual lookout. However, the rain and sea fog that often appear at sea seriously affect performance of the autonomous ship's vision system, as illustrated in Figure 7(d) and (e). Improving the clarity of images obtained in rainy and foggy environments is an important prerequisite for autonomous ships to achieve the goal of intelligently sensing sea-surface obstacles.

Different Degrees of Occlusion of Sea Targets

Near the offshore port, the background of coastal buildings is complex and ships are dense. The objects in marine images often appear occluded and overlapped. There are two common occlusion problems: inter- and intraclass. Interclass occlusion refers to the fact that the target is occluded by other classes of targets, such as marine buoys being occluded by ships. Intraclass occlusion means that the target object is occluded by the same class, such as a partial overlap between ship targets. In complex maritime environments, different degrees of occlusion between obstacles increase the difficulty of maritime object recognition. Occlusion of the target by different types of objects will lead to a partial loss of detection target information, resulting in missed detection. Occlusion between the same type of targets often introduces a large amount of interference information, resulting in a false detection.

How to Fuse Different Sources of Information

Adverse weather and sea conditions can cause EO sensors to have difficulty and error in identifying sea-surface objects. However, how to fuse the information among different sensors is a challenging problem in the maritime environment. In addition to hardware synchronization, sensors also need to achieve time synchronization, but the

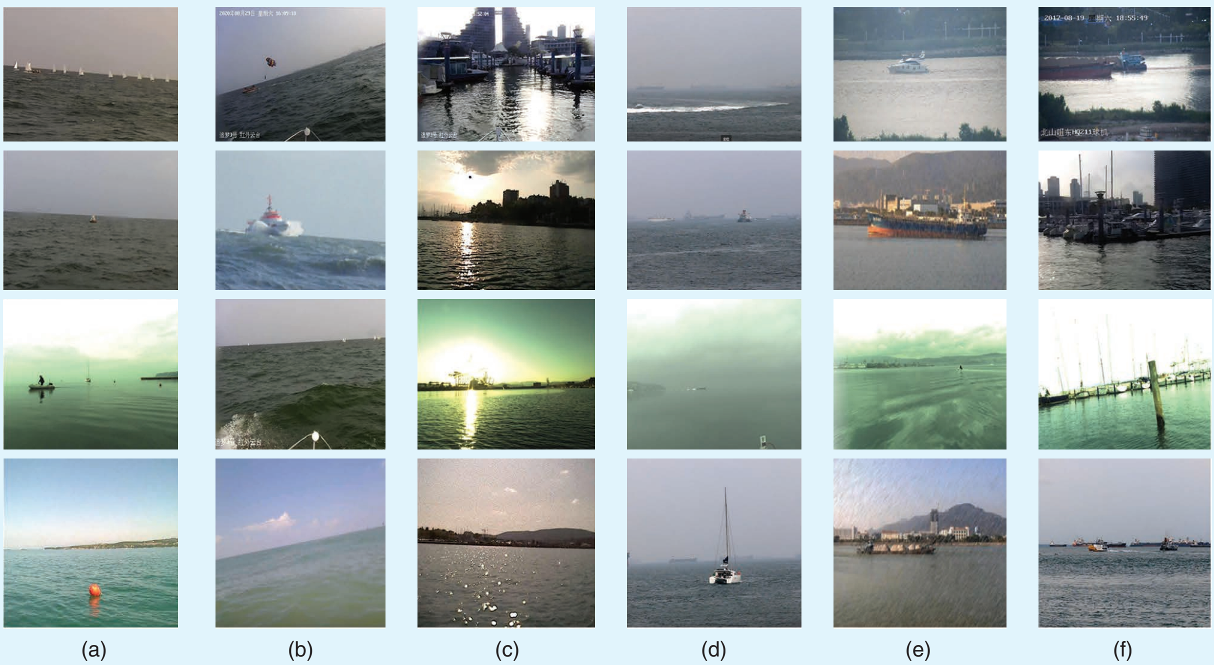

FIG 8 The limitations and challenges of sea-surface object detection [2], [38], [85], [150], [151]. (a) Small objects, (b) wave influence, (c) water reflection, (d) sea fog, (e), rain, and (f) overlap occlusion.

acquisition time of each sensor is not the same. In addition, most of the existing fusion systems usually require the participation of external users to obtain satisfactory performance in the fusion process [29].

The Balance Between Model Accuracy and Speed

Considering the high requirements for real time and the accuracy of maritime obstacle object recognition in an intelligent ship-perception system, the sea-surface object-detection model must achieve high targets in terms of accuracy and speed. Extensive literature reviews on sea-surface object detection have shown that only a portion of the literature can balance accuracy and speed of the algorithm.

The Lack of Image Data of Some Sea-Surface Objects

Adopting the dataset is exceptionally essential in using EO sensors for sea-surface objects detection. More specifically, the deep learning method requires a large amount of image data to train a multilayer neural network. Most of the existing maritime datasets are not made available to the public due to project background. The published maritime datasets contain mostly ships. The target is relatively single and lacks everyday sea-surface objects image datasets, e.g., buoys, islands, reefs, and lighthouses.

Trends and Future Directions

We have observed several trends and found topics that still need to be developed in the future, including feature extraction technology from sea-surface small objects such as the following:

- *Explore high-performance backbone networks*: Complex and intensive calculations of CNNs are extremely demanding for hardware, which makes them difficult to deploy on common hardware devices. In this case, lightweight network technology comes into being. However, the limitation of lightweight network technology is that it affects the accuracy of object detection. Maintaining sufficient accuracy while making the backbone network lightweight is the core research direction in the future.
- *Feature fusion at different scales*: The feature maps of different stages in the network model have different receptive fields and different levels of information abstraction. The FPN fuses the feature maps of different stages, which improves the performance of multiscale object detection and enhances the small-object-recognition effect to a certain extent [130]. In addition, exploring a new context network module [154] to establish information links among objects is also an effective method to enhance the detection accuracy of small sea-surface objects.
- *Explore an efficient attention mechanism model*: In the detection of sea-surface objects, there are objects such as navigation marks and ships—especially navigation marks—that contain deep navigation information. Adding a new attention mechanism model to the network model can learn image object information more efficiently and in depth, hence obtaining more fine-grained image features and acquiring a deeper understanding of maritime obstacle images.

Sea Image Defogging Technology

Existing dehazing algorithms are generally used in land scenarios, such as monitoring of urban roads. Relatively few algorithms exist for fog removal in maritime traffic perception. Single-image dehazing techniques include image-enhancement-based [155] and physical model-based methods [156]. The disadvantage of the approach based on image enhancement is that it does not consider the formation principle of fog, which leads to a lack of information in the recovered image. The method based on the physical model requires sufficient physical color information and is not suitable for dense fog scenes. He et al. [157] proposed an algorithm based on the dark channel prior, but the recovery of this method for bright regions that do not satisfy the dark channel prior theory would result in obvious distortion. As the sea image has a large area of the sky and other parts that do not conform to the dark channel prior principle, the restored image is often ineffective. By considering the characteristics of the sea fog environment and improving the computational efficiency while ensuring the quality of the restored image, the EO sensor can detect sea-surface objects more efficiently.

Detection Technology of Occluded Objects

Compared with general object detection, occlusion is more common in maritime obstacle detection, which is also one of the most concerning issues in the field of maritime object detection. Wang et al. [158] improved loss function for the first time to enable the network to continuously augment target localization performance during the automatic learning process. In the classic deep learning faster R-CNN network model, the target candidate box is divided into different parts to extract features respectively, which reduces the influence of occlusion position on the global features. According to the results shown in the existing research, the detection effect under occlusion is far worse than that under nonocclusion. A key reason is that obstacles occlude each other in dense scenes of maritime environment. In the case of limited training datasets, detectors based on supervised training cannot learn various occlusions.

Data Expansion and Enhancement

The following areas of data expansion have been identified for future development:

- *Maritime dataset expansion*: The abundance and quality of prior knowledge directly affect the quality of CNN model training. In response to the complex marine environment, relevant research institutions have established maritime datasets such as SMD, SeaGull, and SeaShips. Large-scale corpus labeling of datasets through manual methods is still one of the current mainstream techniques. In the future, more marine obstacle image datasets will be created.

- *GAN-based data enhancement*: Based on the supervised learning method, the cost of manual labeling is high, and hence, it is impossible to label all the scenes. On the other hand, the labeled data cannot adaptively and accurately recognize the new objects in the new scene. Subsequently, weak supervision [159], [160] or unsupervised [161], [162] approaches have been proposed. Rai et al. [163] proposed a semisupervised segmentation algorithm named *SemiSegSAR*, which need only label a small amount of data to obtain a satisfactory performance on public ship datasets. Chen et al. [102] successfully augmented and enhanced the maritime dataset using an improved GAN. In the future, combining prior knowledge to achieve adaptive object recognition in different scenarios is a hot spot in the research of sea-surface object detection.

Electronic Phase-Stabilization Technology for Shipborne Cameras

Object recognition of sea-surface obstacles in a static environment is the basis of visual perception. In a dynamic environment, the ship's six-degree-of-freedom motion (e.g., wind and waves) will have a huge impact on the EO sensor, making it difficult for the shipborne camera to obtain a stable video sequence. Shan et al. [164] proposed a new algorithm for sea-surface object detection based on electronic image-stabilization technology by using point-line, point classification, and image classification models. The experimental results show that the algorithm achieved high indicators in terms of mean-square error, average precision, and peak signal-to-noise ratio and demonstrates certain potential in image stabilization and ship detection. In the future, the use of onboard electronic image-stabilization technology on unmanned ships in a dynamic environment to obtain stable image data is another challenge for the development of sea-surface-detection technology.

Multisensor Fusion

A harsh maritime environment will have a more significant impact on the camera. The data fusion of multiple sensors can reduce these effects, such as inertial measurement units (IMUs), AISs, and lidar, complementing EO sensors. Bloisi et al. [29] integrate visual information into the traditional vessel traffic service system and combine radar and AIS information to obtain intuitive and higher-precision maritime object-detection results. In addition, fusing the EO sensor with more sensors can enable better maritime situational awareness, such as the use of IMU readings to estimate the location of the horizon in the image, and program calibration to improve the efficiency of maritime obstacle recognition [165], [166].

Movement Analysis of Sea-Surface Targets

Sea-surface object detection is often accompanied by motion analysis, which is a very challenging kernel task. Online applications with batch algorithms are often memory intensive when using multiple hybrid sensors for target motion analysis (TMA). Liu and Guo [167] designed a recursive estimator of batch counters for multistatic TMA using a hybrid measure of angle of arrival, time difference of arrival, and frequency difference of arrival. The proposed method did not require additional batch estimators for initialization, which improves the convenience of applying the algorithm. It is an interesting and promising direction to design an efficient and stable batch estimator for the analysis of moving objects in complex maritime environments.

Enhanced Localization of Ships

The use of sensors to accurately locate ships is also closely related to sea-surface object detection. Positioning technology is the premise for ships to detect objects on the sea surface. Accurate positioning can make the detection of objects on the sea-surface more efficient. Guo and Liu [168] proposed a stochastic model-based fusion algorithm, which considered the geometric transformation of the vehicle model of the roll and slip angles, and introduced a stochastic model-based extended Kalman filter (EKF) by embedding random noise modulated by absolute value to ensure localization accuracy of the method.

Seeing that the robustness to measurement noise is poor in complex driving environments, Liu and Guo [169] proposed an adaptive mechanism based on improved EKF and deep learning theme for vehicles, which could obtain more reliable results during GPS outages. In the maritime environment, moisture and salt spray damage the sensors quite seriously [170], so the positioning and navigation of autonomous ships at sea is full of challenges, and more intelligent positioning technology needs to be developed.

Test and Validation System Development

For object-detection requirements in maritime environments, methods such as selective parameter sharing [171], [172], data augmentation [173], and complementary feature fusion [174] have been proposed successively, and the recognition efficiency of maritime obstacles has been continuously improved. On the basis of theoretical research, R&D of the sea-surface object-detection algorithm testing and verification platform is urgent. Huang et al. [38] proposed a USV test system for the validation of maritime optical recognition algorithms. Through experiments on the self-built RZ3MOD dataset, it is proved that the proposed technical system can effectively verify the advanced sea-surface object recognition algorithm. By establishing the

USV algorithm test platform, hidden dangers can be found before the final project application, thereby effectively promoting implementation of the algorithm while reducing cost consumption.

Conclusion

Using EO sensors to detect sea-surface objects has always been a challenging issue in the maritime field, especially for autonomous ships. This article offered a comprehensive overview of the EO sensors-based sea-surface object-detection methods, providing a technical basis for subsequent object tracking and positioning. From the perspective of sea-surface object detection, the article comprehensively analyzed traditional methods, deep learning-based techniques, and typical maritime datasets as well as their research development, bottleneck, challenges, and future directions in this field. To the best of our knowledge, this is the first survey in the literature to focus on object detection using deep learning in maritime datasets. Compared with traditional sea-surface object-detection methods, deep learning technology has shown to be decisive in maritime object-detection accuracy and real-time performance. However, it requires high-quality datasets to produce reliable results. The existing two types of sea-surface object-detection technology have their advantages and limitations, and we believe that they can complement each other to some extent.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this article.

Acknowledgment

This work is supported by the National Key R&D Program of China (grant number 2019YFB1600602), the National Natural Science Foundation of China (number 52071049), and the Natural Science Foundation of Liaoning Province (number 2020-BS-070). The authors would like to express their gratitude to EditSprings for the expert linguistic services provided.

About the Authors

Hongguang Lyu (lhg@dlmu.edu.cn) earned his Ph.D. degree in transport information engineering and control in 2019 from Dalian Maritime University. He is an associate professor at Dalian Maritime University, Dalian 116026, China. His research focuses on intelligent navigation system for maritime autonomous surface ships and unmanned surface vehicles, especially in the field of autonomous collision avoidance and path planning.

Zeyuan Shao (szy_dlmu@163.com) earned his B.Eng. degree in nautical science from Zhejiang Ocean University in 2019. He is currently pursuing his Ph.D. degree in transport information engineering and control from Dalian Maritime University, Dalian 116026, China. His research interests include intelligent navigation systems for maritime autonomous surface ships and unmanned surface vehicles, maritime data mining, computer vision, and artificial intelligence.

Tao Cheng (tao.cheng@ucl.ac.uk) earned her Ph.D. degree in bioInformatics in 1999 from Wageningen Agricultural University and the International Institute for Geo-Information Science and Earth Observation, The Netherlands. She is a professor of geoinformatics and director of the SpaceTimeLab for Big Data Analytics, University College London, London WC1E 6BT, U.K. Her research interests span network complexity, geocomputation, integrated spatial-temporal analytics, and big data mining (modeling, prediction, clustering, and simulation), with applications in transport, crime, health, social media, and environmental monitoring.

Yong Yin (bushyin@dlmu.edu.cn) earned his Ph.D. degree in transport information engineering and control in 2001 from Dalian Maritime University. He is a professor at Dalian Maritime University, Dalian 116026, China. He served as the deputy director of the China Institute of Navigation from 2015–2018. His research interests include navigation simulation, navigation intelligence, and virtual reality.

Xiaowei Gao (ucesxwg@ucl.ac.uk) earned a master's degree in transport information engineering and control in 2017 from Dalian Maritime University and a master's degree in data science in 2020 from King's College London. He is currently pursuing his Ph.D. degree with SpaceTimeLab for Big Data Analytics, University College London, London WC1E 6BT, U.K. His research interests include spatial-temporal data mining, urban computing on cycling, data fusion techniques, bicycle share systems, and artificial intelligence.

References

[1] Z. Liu, Y. Zhang, X. Yu, and C. Yuan, "Unmanned surface vehicles: An overview of developments and challenges," *Annu. Rev. Control*, vol. 41, pp. 71–93, Jun. 2016, doi: 10.1016/j.arcontrol.2016.04.018.
[2] D. K. Prasad, D. Rajan, L. Rachmawati, E. Rajabally, and C. Quek, "Video processing from electro-optical sensors for object detection and

tracking in a maritime environment: A survey," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 8, pp. 1993–2016, Aug. 2017, doi: 10.1109/TITS.2016.2634580.
[3] D. K. Prasad, C. K. Prasath, D. Rajan, L. Rachmawati, E. Rajabally, and C. Quek, "Object detection in a maritime environment: Performance evaluation of background subtraction methods," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 5, pp. 1787–1802, May 2019, doi: 10.1109/TITS.2018.2836399.
[4] D. K. Prasad, C. K. Prasath, D. Rajan, L. Rachmawati, E. Rajabaly, and C. Quek, "Challenges in video based object detection in maritime scenario using computer vision," Aug. 2016. [Online]. Available: http://arXiv.org/abs/1608.01079
[5] Z. Ma, J. Wen, and X. Liang, "Video image clarity algorithm research of USV visual system under the sea fog," in *Advances in Swarm Intelligence*, vol. 7929, Y. Tan, Y. Shi, and H. Mo, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 436–444, doi: 10.1007/978-3-642-38715-9_52.
[6] K. K. Liu and J. H. Wang, "A method of detecting wave grade based on visual image taken by USV," *Appl. Mech. Mater.*, vols. 291–294, pp. 2437–2441, Feb. 2013, doi: 10.4028/www.scientific.net/AMM.291-294.2437.
[7] O. Gal, "Automatic obstacle detection for USV's navigation using vision sensors," in *Proc. 4th Int. Robot. Sailing Conf.*, A. Schlaefer and O. Blaurock, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 127–140, doi: 10.1007/978-3-642-22836-0_9.
[8] Y.-T. Chan, "Comprehensive comparative evaluation of background subtraction algorithms in open sea environments," *Comput. Vis. Image Understanding*, vol. 202, p. 103,101, Jan. 2021, doi: 10.1016/j.cviu.2020.103101.
[9] R. D. S. Moreira, N. F. F. Ebecken, A. S. Alves, F. Livernet, and A. Campillo-Navetti, "A survey on video detection and tracking of maritime vessels," *Int. J. Recent Res. Appl. Stud.*, vol. 20, no. 1, pp. 37–50, Jul. 2014.
[10] F. E. T. Schöller, M. K. Plenge-Feidenhans'l, J. D. Stets, and M. Blanke, "Assessing deep-learning methods for object detection at sea from LWIR images," *IFAC-PapersOnLine*, vol. 52, no. 21, pp. 64–71, 2019, doi: 10.1016/j.ifacol.2019.12.284.
[11] N. Wang, Y. Wang, and M. J. Er, "Review on deep learning techniques for marine object recognition: Architectures and algorithms," *Control Eng. Pract.*, vol. 118, p. 104,458, May 2020, doi: 10.1016/j.conengprac.2020.104458.
[12] U. Kanjir, H. Greidanus, and K. Oštir, "Vessel detection and classification from spaceborne optical images: A literature survey," *Remote Sens. Environ.*, vol. 207, pp. 1–26, Mar. 2018, doi: 10.1016/j.rse.2017.12.033.
[13] B. Li, X. Xie, X. Wei, and W. Tang, "Ship detection and classification from optical remote sensing images: A survey," *Chinese J. Aeronaut.*, vol. 34, no. 3, pp. 145–163, Mar. 2021, doi: 10.1016/j.cja.2020.09.022.
[14] P. Mittal, A. Sharma, and R. Singh, "Deep learning-based object detection in low-altitude UAV datasets: A survey," *Image Vis. Comput.*, vol. 104, p. 104,046, Dec. 2020, doi: 10.1016/j.imavis.2020.104046.
[15] C. Zhang, C. Guo, and D. Zhang, "Ship navigation via GPS/IMU/LOG integration using adaptive fission particle filter," *Ocean Eng.*, vol. 156, pp. 435–445, May 2018, doi: 10.1016/j.oceaneng.2018.03.012.
[16] A. Hesselbarth, D. Medina, R. Ziebold, M. Sandler, M. Hoppe, and M. Uhlemann, "Enabling assistance functions for the safe navigation of inland waterways," *IEEE Intell. Transp. Syst. Mag.*, vol. 12, no. 3, pp. 123–135, Fall 2020, doi: 10.1109/MITS.2020.2994103.
[17] V. Fernandez Arguedas, G. Pallotta, and M. Vespe, "Maritime traffic networks: From historical positioning data to unsupervised maritime traffic monitoring," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 3, pp. 722–732, Mar. 2018, doi: 10.1109/TITS.2017.2699635.
[18] A. Toet, "Detection of dim point targets in cluttered maritime backgrounds through multisensor image fusion," in *Proc. Int. Soc. Opt. Eng.*, Orlando, FL, USA, Aug. 2002, pp. 118–129, doi: 10.1117/12.478798.
[19] S. Panagopoulos and J. J. Soraghan, "Small-target detection in sea clutter," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 7, pp. 1355–1361, Jul. 2004, doi: 10.1109/TGRS.2004.827259.
[20] H. Bouma, D.-J. J. de Lange, S. P. van den Broek, R. A. W. Kemp, and P. B. W. Schwering, "Automatic detection of small surface targets with electro-optical sensors in a harbor environment," in *Proc. SPIE 7114, Electro-Opt. Remote Sens., Photon. Technol., Appl.*, Cardiff, U.K., Oct. 2008, p. 711,402, doi: 10.1117/12.799813.
[21] H. K. Heidarsson and G. S. Sukhatme, "Obstacle detection and avoidance for an Autonomous Surface Vehicle using a profiling sonar," in *Proc. IEEE Int. Conf. Robot. Automat.*, Shanghai, China, May 2011, pp. 731–736, doi: 10.1109/ICRA.2011.5980509.
[22] O. Gal and E. Zeitouni, "Tracking objects using PHD filter for USV autonomous capabilities," in *Proc. 5th Int. Robot. Sailing Conf.*, C. Sauzé, and J. Finnis, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 3–12, doi: 10.1007/978-3-642-33084-1_1.
[23] M. P. Hayes and P. T. Gough, "Synthetic aperture sonar: A review of current status," *IEEE J. Ocean. Eng.*, vol. 34, no. 3, pp. 207–224, Jul. 2009, doi: 10.1109/JOE.2009.2020853.
[24] H. K. Heidarsson and G. Sukhatme, "Obstacle detection from overhead imagery using self-supervised learning for autonomous surface vehicles," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, San Francisco, CA, USA, Sep. 2011, pp. 3160–3165, doi: 10.1109/IROS.2011.6094610.
[25] Z. L. Szpak and J. R. Tapamo, "Maritime surveillance: Tracking ships inside a dynamic background using a fast level-set," *Expert Syst. Appl.*, vol. 38, no. 6, pp. 6669–6680, Jun. 2011, doi: 10.1016/j.eswa.2010.11.068.
[26] L. Wen, J. Ding, and Z. Xu, "Multiframe detection of sea-surface small target using deep convolutional neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–16, 2022, doi: 10.1109/TGRS.2021.3122515.
[27] K. Zhang, P.-L. Shui, and Y. Feng, "Detection of sea-surface small targets masked by range sidelobes of large objects," *IEEE Trans. Aerosp. Electron. Syst. (1965–present)*, vol. 58, no. 2, pp. 1446–1461, Apr. 2022, doi: 10.1109/TAES.2021.3116120.
[28] J. Zhuang, L. Zhang, S. Zhao, J. Cao, B. Wang, and H. Sun, "Radar-based collision avoidance for unmanned surface vehicles," *China Ocean Eng.*, vol. 30, no. 6, pp. 867–883, Dec. 2016, doi: 10.1007/s13344-016-0056-0.
[29] D. D. Bloisi, F. Previtali, A. Pennisi, D. Nardi, and M. Fiorini, "Enhancing automatic maritime surveillance systems with visual information," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 4, pp. 824–833, Apr. 2017, doi: 10.1109/TITS.2016.2591321.
[30] D. D. Bloisi, L. Iocchi, D. Nardi, and M. Fiorini, "Integrated visual information for maritime surveillance," in *Clean Mobility and Intelligent Transport Systems*, M. Fiorini and J.-C. Lin, Eds. London, U.K.: Institution of Engineering and Technology, 2015, pp. 237–264, doi: 10.1049/PBTR001E_ch11.
[31] M. Blaich, S. Kohler, M. Schuster, T. Schuchhardt, J. Reuter, and T. Tietz, "Mission integrated collision avoidance for USVs using laser range finder," in *Proc. OCEANS - Genova*, Genova, Italy, May 2015, pp. 1–6, doi: 10.1109/OCEANS-Genova.2015.7271413.
[32] J. Choi, S. Ulbrich, B. Lichte, and M. Maurer, "Multi-Target Tracking using a 3D-Lidar sensor for autonomous vehicles," in *Proc. 16th Int. IEEE Conf. Intell. Transp. Syst. (ITSC)*, The Hague, The Netherlands, Oct. 2013, pp. 881–886, doi: 10.1109/ITSC.2013.6728343.
[33] A. R. J. Ruiz and F. S. Granja, "A short-range ship navigation system based on ladar imaging and target tracking for improved safety and efficiency," *IEEE Trans. Intell. Transp. Syst.*, vol. 10, no. 1, pp. 186–197, Mar. 2009, doi: 10.1109/TITS.2008.2011698.
[34] J. Villa, J. Aaltonen, and K. T. Koskinen, "Path-following with LiDAR-based obstacle avoidance of an unmanned surface vehicle in harbor conditions," *IEEE/ASME Trans. Mechatronics*, vol. 25, no. 4, pp. 1812–1820, Aug. 2020, doi: 10.1109/TMECH.2020.2997970.
[35] X. Bai, M. Liu, T. Wang, Z. Chen, P. Wang, and Y. Zhang, "Feature based fuzzy inference system for segmentation of low-contrast infrared ship images," *Appl. Soft Comput.*, vol. 46, pp. 128–142, Sep. 2016, doi: 10.1016/j.asoc.2016.05.004.
[36] Y. Lang and B. Yuan, "Algorithm application based on the infrared image in unmanned ship target image recognition," *Microprocessors Microsyst.*, vol. 80, p. 103,554, Feb. 2021, doi: 10.1016/j.micpro.2020.103554.
[37] X. Mou and H. Wang, "Wide-baseline stereo-based obstacle mapping for unmanned surface vehicles," *Sensors*, vol. 18, no. 4, p. 1085, Apr. 2018, doi: 10.3390/s18041085.
[38] Y. Huang, H. Wang, J. Ma, J. Lou, and H. Yi, "Research and practical exploration of test and validation technologies applied on unmanned surface vehicle optical recognition," in *Proc. IEEE Int. Conf. Unmanned Syst. (ICUS)*, Beijing, China, Oct. 2021, pp. 976–981, doi: 10.1109/ICUS52573.2021.9641224.
[39] Z. Gao, Y. Zhang, and Y. Li, "Extracting features from infrared images using convolutional neural networks and transfer learning," *Infrared Phys. Technol.*, vol. 105, p. 103,237, Mar. 2020, doi: 10.1016/j.infrared.2020.103237.
[40] B. Jiang, X. Ma, Y. Lu, Y. Li, L. Feng, and Z. Shi, "Ship detection in spaceborne infrared images based on Convolutional Neural Networks and synthetic targets," *Infrared Phys. Technol.*, vol. 97, pp. 229–234, Mar. 2019, doi: 10.1016/j.infrared.2018.12.040.
[41] Z. Liu, F. Zhou, X. Chen, X. Bai, and C. Sun, "Iterative infrared ship target segmentation based on multiple features," *Pattern Recognit.*, vol. 47, no. 9, pp. 2839–2852, Sep. 2014, doi: 10.1016/j.patcog. 2014.03.005.
[42] Z. Liu, X. Bai, C. Sun, F. Zhou, and Y. Li, "Infrared ship target segmentation through integration of multiple feature maps," *Image Vis. Comput.*, vols. 48–49, pp. 14–25, Apr. 2016, doi: 10.1016/j.imavis.2015.12.005.
[43] M. Zhou, M. Jing, D. Liu, Z. Xia, Z. Zou, and Z. Shi, "Multi-resolution networks for ship detection in infrared remote sensing images," *Infrared Phys. Technol.*, vol. 92, pp. 183–189, Aug. 2018, doi: 10.1016/j.infrared.2018.05.025.
[44] C. Jiang, H. Jiang, C. Zhang, and J. Wang, "A new method of sea-sky-line detection," in *Proc. 3rd Int. Symp. Intell. Inf. Technol. Security Informat.*, Jian, China, Apr. 2010, pp. 740–743, doi: 10.1109/IITSI.2010.147.
[45] T. Ma, J. Ma, and W. Fu, "Sea-sky line extraction with linear fitting based on line segment detection," in *Proc. 9th Int. Symp. Comput.*

Intell. Des. (ISCID), Hangzhou, China, Dec. 2016, pp. 46–49, doi: 10.1109/ISCID.2016.1019.
[46] X. Kong, "Automatic detection of sea-sky horizon line and small targets in maritime infrared imagery," *Infrared Phys.*, vol. 76, pp. 185–199, May 2016, doi: 10.1016/j.ijnaoe.2020.03.007.
[47] D. Tang, G. Sun, D. Wang, Z. Niu, and Z. Chen, "Research on infrared ship detection method in sea-sky background," in *Proc. SPIE Int. Soc. Opt. Eng.*, Beijing, China, Sep. 2013, p. 89072H, doi: 10.1117/12.2033039.
[48] H. Wei, H. Nguyen, P. Ramu, C. Raju, X. Liu, and J. Yadegar, "Automated intelligent video surveillance system for ships," in *Proc. SPIE Int. Soc. Opt. Eng.*, Orlando, FL, USA, May 2009, p. 73061N, doi: 10.1117/12.819051.
[49] 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, Apr. 2016, doi: 10.3390/s16040543.
[50] C. Lin, W. Chen, and H. Zhou, "Multi-visual feature saliency detection for sea-surface targets through improved sea-sky-line detection," *J. Marine Sci. Eng.*, vol. 8, no. 10, p. 799, Oct. 2020, doi: 10.3390/jmse8100799.
[51] J. Lu, J. Ren, Y. Lu, X. Yuan, and C. Wang, "A modified canny algorithm for detecting sky-sea line in infrared images," in *Proc. 6th Int. Conf. Intell. Syst. Des. Appl.*, Jian, China, Oct. 2006, vol. 2, pp. 289–294, doi: 10.1109/ISDA.2006.253848.
[52] D. Liang, W. Zhang, Q. Huang, and F. Yang, "Robust sea-sky-line detection for complex sea background," in *Proc. IEEE Int. Conf. Prog. Informat. Comput. (PIC)*, Nanjing, China, Dec. 2015, pp. 317–321, doi: 10.1109/PIC.2015.7489861.
[53] L. Yang, Y. Zhou, J. Yang, and L. Chen, "Variance WIE based infrared images processing," *Electron. Lett.*, vol. 42, no. 15, p. 857, 2006, doi: 10.1049/el:20060827.
[54] C. Jeong, H. S. Yang, and K. Moon, "A novel approach for detecting the horizon using a convolutional neural network and multi-scale edge detection," *Multidimensional Syst. Signal Process.*, vol. 30, no. 3, pp. 1187–1204, Jul. 2019, doi: 10.1007/s11045-018-0602-4.
[55] W. Zhan et al., "Autonomous visual perception for unmanned surface vehicle navigation in an unknown environment," *Sensors*, vol. 19, no. 10, May 2019, Art. no. 10, doi: 10.3390/s19102216.
[56] L. Ren, C. Shi, and X. Ran, "Target detection of maritime search and rescue: Saliency accumulation method," in *Proc. 9th Int. Conf. Fuzzy Syst. Knowl. Discovery*, Chongqing, China, May 2012, pp. 1972–1976, doi: 10.1109/FSKD.2012.6233916.
[57] J. Zhou, H. Lv, and F. Zhou, "Infrared small target enhancement by using sequential top-hat filters," in *Proc. SPIE Int. Soc. Opt. Eng.*, Beijing, China, Nov. 2014, p. 93011L, doi: 10.1117/12.2072159.
[58] D. Culibrk, O. Marques, D. Socek, H. Kalva, and B. Furht, "Neural network approach to background modeling for video object segmentation," *IEEE Trans. Neural Netw.*, vol. 18, no. 6, pp. 1614–1627, Nov. 2007, doi: 10.1109/TNN.2007.896861.
[59] D. Socek, D. Culibrk, O. Marques, H. Kalva, and B. Furht, "A hybrid color-based foreground object detection method for automated marine surveillance," in *Advanced Concepts for Intelligent Vision Systems*, vol. 3708, J. Blanc-Talon, W. Philips, D. Popescu, and P. Scheunders, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005, pp. 340–347, doi: 10.1007/11558484_43.
[60] A. M. R. Mohd, M. Mokji, and M. M. A. Zabidi, "Low complexity maritime surveillance video using background subtraction on H.264," in *Proc. Int. Symp. Technol. Manage. Emerg. Technol. (ISTMET)*, Langkawi, Malaysia, Aug. 2015, pp. 364–368, doi: 10.1109/ISTMET.2015.7359060.
[61] A. Borghgraef, O. Barnich, F. Lapierre, M. Van Droogenbroeck, W. Philips, and M. Acheroy, "An evaluation of pixel-based methods for the detection of floating objects on the sea surface," *EURASIP J. Adv. Signal Process.*, vol. 2010, no. 1, p. 978,451, Dec. 2010, doi: 10.1155/2010/978451.
[62] D. Adiguzel and L. Ozyilmaz, "Improving of ViBE algoritihm for moving target detection at maritime surveilance," in *Proc. 26th Signal Process. Commun. Appl. Conf. (SIU)*, Izmir, Turkey, May 2018, pp. 1–4, doi: 10.1109/SIU.2018.8404679.
[63] T.-H. Tran and T.-L. Le, "Vision based boat detection for maritime surveillance," in *Proc. Int. Conf. Electron., Inf., Commun. (ICEIC)*, Danang, Vietnam, Jan. 2016, pp. 1–4, doi: 10.1109/ELINFOCOM.2016.7563033.
[64] M. Cristani, M. Bicego, and V. Murino, "Integrated region- and pixel-based approach to background modelling," in *Proc. Workshop Motion Video Comput.*, Orlando, FL, USA, 2002, pp. 3–8, doi: 10.1109/MOTION.2002.1182206.
[65] A. Mittal and N. Paragios, "Motion-based background subtraction using adaptive kernel density estimation," in *Proc. 2004 IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Washington, DC, USA, vol. 2, pp. 302–309, doi: 10.1109/CVPR.2004.1315179.
[66] C. Stauffer and W. E. L. Grimson, "Learning patterns of activity using real-time tracking," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 747–757, Aug. 2000, doi: 10.1109/34.868677.
[67] Y. Wang et al., "Aquatic debris monitoring using smartphone-based robotic sensors," in *Proc. 13th Int. Symp. Inf. Process. Sensor Netw.*, Berlin, Germany, Apr. 2014, pp. 13–24, doi: 10.1109/IPSN.2014.6846737.
[68] D. Zhang, E. O'Connor, K. McGuinness, N. E. O'Connor, F. Regan, and A. Smeaton, "A visual sensing platform for creating a smarter multi-modal marine monitoring network," in *Proc. 1st ACM Int. Workshop Multimedia Anal. Ecological Data (MAED '12)*, Nara, Japan, 2012, p. 53, doi: 10.1145/2390832.2390846.
[69] D. Frost and J.-R. Tapamo, "Detection and tracking of moving objects in a maritime environment using level set with shape priors," *EURASIP J. Image Video Process.*, vol. 2013, no. 1, p. 42, Dec. 2013, doi: 10.1186/1687-5281-2013-42.
[70] A. Zhou, W. Xie, and J. Pei, "Background modeling in the Fourier domain for maritime infrared target detection," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 8, pp. 2634–2649, Aug. 2020, doi: 10.1109/TCSVT.2019.2922036.
[71] C. Zhu, H. Zhou, R. Wang, and J. Guo, "A novel hierarchical method of ship detection from spaceborne optical image based on shape and texture features," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 9, Sep. 2010, Art. no. 9, doi: 10.1109/TGRS.2010.2046330.
[72] T. Nie, X. Han, B. He, X. Li, H. Liu, and G. Bi, "Ship detection in panchromatic optical remote sensing images based on visual saliency and multi-dimensional feature description," *Remote Sens.*, vol. 12, no. 1, p. 152, Jan. 2020, doi: 10.3390/rs12010152.
[73] M. Fiorini, A. Pennisi, and D. D. Bloisi, "Optical target recognition for drone ships," in *Proc. 2017 Int. Conf. Marine Navigation Safety Sea Transp.*, TRANSNAV, pp. 371–375, doi: 10.1201/9781315099132-64.
[74] F. Biondi, "Low rank plus sparse decomposition of synthetic aperture radar data for maritime surveillance," in *Proc. 4th Int. Workshop Compressed Sens. Theory Appl. Radar, Sonar Remote Sens. (CoSeRa)*, Aachen, Germany, Sep. 2016, pp. 75–79, doi: 10.1109/CoSeRa.2016.7745703.
[75] A. Sobral, T. Bouwmans, and E. ZahZah, "Double-constrained RPCA based on saliency maps for foreground detection in automated maritime surveillance," in *Proc. 12th IEEE Int. Conf. Adv. Video Signal Based Surveillance (AVSS)*, Karlsruhe, Germany, Aug. 2015, pp. 1–6, doi: 10.1109/AVSS.2015.7301753.
[76] I. Kajo, N. Kamel, and Y. Ruichek, "Tensor-based approach for background-foreground separation in maritime sequences," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 11, pp. 7115–7128, Nov. 2021, doi: 10.1109/TITS.2020.3001687.
[77] P. Westall, J. J. Ford, P. O'Shea, and S. Hrabar, "Evaluation of maritime vision techniques for aerial search of humans in maritime environments," in *Proc. 2008 Digit. Image Comput., Techn. Appl.*, Canberra, Australia, pp. 176–183, doi: 10.1109/DICTA.2008.89.
[78] H. Eum, J. Bae, C. Yoon, and E. Kim, "Ship detection using edge-based segmentation and histogram of oriented gradient with ship size ratio," *Int. J. Fuzzy Log. Intell. Syst.*, vol. 15, no. 4, pp. 251–259, Dec. 2015, doi: 10.5391/IJFIS.2015.15.4.251.
[79] C. Heltin Genitha, M. Sowmya, and T. Sri, "Comparative analysis for the detection of marine vessels from satellite images using FCM and marker-controlled watershed segmentation algorithm," *J. Indian Soc. Remote Sens.*, vol. 48, no. 8, pp. 1207–1214, Aug. 2020, doi: 10.1007/s12524-020-01148-x.
[80] A. Kushwaha and R. Srivastava, "A framework of moving object segmentation in maritime surveillance inside a dynamic background," *Trans. Comput. Sci.*, vol. 9030, pp. 35–54, Apr. 2015, doi: 10.1007/978-3-662-47074-9_3.
[81] G. Cutter, K. Stierhoff, and J. Zeng, "Automated detection of rockfish in unconstrained underwater videos using Haar cascades and a new image dataset: Labeled fishes in the wild," in *Proc. IEEE Winter Appl. Comput. Vis. Workshops*, Waikoloa, HI, USA, Jan. 2015, pp. 57–62, doi: 10.1109/WACVW.2015.11.
[82] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? The KITTI vision benchmark suite," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Providence, RI, USA, Jun. 2012, pp. 3354–3361, doi: 10.1109/CVPR.2012.6248074.
[83] P. Dollar, C. Wojek, B. Schiele, and P. Perona, "Pedestrian detection: An evaluation of the state of the art," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 4, pp. 743–761, Apr. 2012, doi: 10.1109/TPAMI.2011.155.
[84] S. Zhang, R. Benenson, and B. Schiele, "CityPersons: A diverse dataset for pedestrian detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 4457–4465, doi: 10.1109/CVPR.2017.474.
[85] Z. Shao, W. Wu, Z. Wang, W. Du, and C. Li, "SeaShips: A large-scale precisely annotated dataset for ship detection," *IEEE Trans. Multimedia*, vol. 20, no. 10, pp. 2593–2604, Oct. 2018, doi: 10.1109/TMM.2018.2865686.
[86] T. Liu, B. Pang, S. Ai, and X. Sun, "Study on visual detection algorithm of sea surface targets based on improved YOLOv3," *Sensors*, vol. 20, no. 24, p. 7263, Dec. 2020, doi: 10.3390/s20247263.

[87] R. Ribeiro, G. Cruz, J. Matos, and A. Bernardino, "A data set for airborne maritime surveillance environments," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 29, no. 9, pp. 2720–2732, Sep. 2019, doi: 10.1109/TCSVT.2017.2775524.
[88] M. M. Zhang, J. Choi, K. Daniilidis, M. T. Wolf, and C. Kanan, "VAIS: A dataset for recognizing maritime imagery in the visible and infrared spectrums," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Boston, MA, USA, Jun. 2015, pp. 10–16, doi: 10.1109/CVPRW.2015.7301291.
[89] B. Iancu, V. Soloviev, L. Zelioli, and J. Lilius, "ABOships—An inshore and offshore maritime vessel detection dataset with precise annotations," *Remote Sens.*, vol. 13, no. 5, p. 988, Mar. 2021, doi: 10.3390/rs13050988.
[90] H.-C. Shin, K.-I. Lee, and C.-E. Lee, "Data augmentation method of object detection for deep learning in maritime image," in *Proc. IEEE Int. Conf. Big Data Smart Comput. (BigComp)*, Busan, South Korea, Feb. 2020, pp. 463–466, doi: 10.1109/BigComp48618.2020.00-25.
[91] L. Gao, Y. He, X. Sun, X. Jia, and B. Zhang, "Incorporating negative sample training for ship detection based on deep learning," *Sensors*, vol. 19, no. 3, p. 684, Feb. 2019, doi: 10.3390/s19030684.
[92] Y. He, X. Sun, L. Gao, and B. Zhang, "Ship detection without sea-land segmentation for large-scale high-resolution optical satellite images," in *Proc. IGARSS IEEE Int. Geosci. Remote Sens. Symp.*, Valencia, Spain, Jul. 2018, pp. 717–720, doi: 10.1109/IGARSS.2018.8519391.
[93] H. Fu, Y. Li, Y. Wang, and P. Li, "Maritime ship targets recognition with deep learning," in *Proc. 37th Chinese Control Conf.*, Wuhan, China, 2018, pp. 9297–9302, doi: 10.23919/ChiCC.2018.8484085.
[94] L. Qi et al., "Ship target detection algorithm based on improved faster R-CNN," *Electronics*, vol. 8, no. 9, p. 959, Aug. 2019, doi: 10.3390/electronics8090959.
[95] Y. Wang, X. Ning, B. Leng, and H. Fu, "Ship detection based on deep learning," in *Proc. IEEE Int. Conf. Mechatronics Automat. (ICMA)*, Tianjin, China, Aug. 2019, pp. 275–279, doi: 10.1109/ICMA.2019.8816265.
[96] X. Chen et al., "Video-based detection infrastructure enhancement for automated ship recognition and behavior analysis," *J. Adv. Transp.*, vol. 2020, pp. 1–12, Jan. 2020, doi: 10.1155/2020/7194342.
[97] Y. Li et al., "A novel target detection method of the unmanned surface vehicle under all-weather conditions with an improved YOLOV3," *Sensors*, vol. 20, no. 17, p. 4885, Aug. 2020, doi: 10.3390/s20174885.
[98] Z. Huang, B. Sui, J. Wen, and G. Jiang, "An intelligent ship image/video detection and classification method with improved regressive deep convolutional neural network," *Complexity*, vol. 2020, pp. 1–11, Apr. 2020, doi: 10.1155/2020/1520872.
[99] I. J. Goodfellow et al., "Generative adversarial networks," Jun. 2014. [Online]. Available: http://arXiv.org/abs/1406.2661
[100] C. P. Schwegmann, W. Kleynhans, B. P. Salmon, L. W. Mdakane, and R. G. V. Meyer, "Synthetic aperture radar ship discrimination, generation and latent variable extraction using information maximizing generative adversarial networks," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Fort Worth, TX, USA, Jul. 2017, pp. 2263–2266, doi: 10.1109/IGARSS.2017.8127440.
[101] L. Zou, H. Zhang, C. Wang, F. Wu, and F. Gu, "MW-ACGAN: Generating multiscale high-resolution SAR images for ship detection," *Sensors*, vol. 20, no. 22, p. 6673, Nov. 2020, doi: 10.3390/s20226673.
[102] Z. Chen, D. Chen, Y. Zhang, X. Cheng, M. Zhang, and C. Wu, "Deep learning for autonomous ship-oriented small ship detection," *Safety Sci.*, vol. 130, p. 104,812, Oct. 2020, doi: 10.1016/j.ssci.2020.104812.
[103] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017, doi: 10.1145/3065386.
[104] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998, doi: 10.1109/5.726791.
[105] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, 2017, doi: 10.1109/TPAMI.2016.2644615.
[106] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 4, pp. 640–651, 2017, doi: 10.1109/TPAMI.2016.2572683.
[107] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," May 2015. [Online]. Available: http://arXiv.org/abs/1505.04597
[108] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," Apr. 2015. [Online]. Available: http://arXiv.org/abs/1409.1556
[109] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," Mar. 2015. [Online]. Available: http://arXiv.org/abs/1502.03167
[110] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," Dec. 2015. [Online]. Available: http://arXiv.org/abs/1512.00567
[111] C. Szegedy et al., "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Boston, MA, USA, Jun. 2015, pp. 1–9, doi: 10.1109/CVPR.2015.7298594.
[112] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.
[113] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 2261–2269, doi: 10.1109/CVPR.2017.243.
[114] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Columbus, OH, USA, Jun. 2014, pp. 580–587, doi: 10.1109/CVPR.2014.81.
[115] R. Girshick, "Fast R-CNN," Sep. 2015. [Online]. Available: http://arXiv.org/abs/1504.08083
[116] Y. Chen, X. Chen, J. Zhu, F. Lin, and B. M. Chen, "Development of an autonomous unmanned surface vehicle with object detection using deep learning," in *Proc. IECON 44th Annu. Conf. IEEE Ind. Electron. Soc.*, Washington, DC, USA, Oct. 2018, pp. 5636–5641, doi: 10.1109/IECON.2018.8591129.
[117] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, 2017, doi: 10.1109/TPAMI.2016.2577031.
[118] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, 2016, pp. 779–788, doi: 10.1109/CVPR.2016.91.
[119] J. Redmon and A. Farhadi, "YOLO9000: Better, faster, stronger," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 6517–6525, doi: 10.1109/CVPR.2017.690.
[120] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," 2018. [Online]. Available: http://arXiv.org/abs/1804.02767
[121] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, "Focal loss for dense object detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 2, pp. 318–327, Feb. 2020, doi: 10.1109/TPAMI.2018.2858826.
[122] W. Liu et al., "SSD: Single shot multibox detector," in *Computer Vision – ECCV*, vol. 9905, B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds. Cham: Springer International Publishing, 2016, pp. 21–37, doi: 10.1007/978-3-319-46448-0_2.
[123] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal speed and accuracy of object detection," Apr. 2020. [Online]. Available: http://arXiv.org/abs/2004.10934
[124] F. Farahnakian and J. Heikkonen, "Deep learning based multi-modal fusion architectures for maritime vessel detection," *Remote Sens.*, vol. 12, no. 16, p. 2509, Aug. 2020, doi: 10.3390/rs12162509.
[125] L. Ma, W. Xie, and H. Huang, "Convolutional neural network based obstacle detection for unmanned surface vehicle," *Math. Biosci. Eng.*, vol. 17, no. 1, pp. 845–861, 2020, doi: 10.3934/mbe.2020045.
[126] Z. Li, L. Zhao, X. Han, M. Pan, and F.-J. Hwang, "Lightweight ship detection methods based on YOLOv3 and DenseNet," *Math. Problems Eng.*, vol. 2020, pp. 1–10, Sep. 2020, doi: 10.1155/2020/4813183.
[127] H. Zhao, W. Zhang, H. Sun, and B. Xue, "Embedded deep learning for ship detection and recognition," *Future Internet*, vol. 11, no. 2, p. 53, Feb. 2019, doi: 10.3390/fi11020053.
[128] H. Huang, D. Sun, R. Wang, C. Zhu, and B. Liu, "Ship target detection based on improved YOLO Network," *Math. Problems Eng.*, vol. 2020, pp. 1–10, Aug. 2020, doi: 10.1155/2020/6402149.
[129] Z. Shao, L. Wang, Z. Wang, W. Du, and W. Wu, "Saliency-aware convolution neural network for ship detection in surveillance video," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 3, pp. 781–794, Mar. 2020, doi: 10.1109/TCSVT.2019.2897980.
[130] T.-Y. Lin, P. Dollar, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, Jul. 2017, pp. 936–944, doi: 10.1109/CVPR.2017.106.
[131] H. Guo, X. Yang, N. Wang, B. Song, and X. Gao, "A rotational libra R-CNN method for ship detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 8, pp. 5772–5781, Aug. 2020, doi: 10.1109/TGRS.2020.2969979.
[132] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," Jul. 2018. [Online]. Available: http://arXiv.org/abs/1807.06521
[133] M. Pan, Y. Liu, J. Cao, Y. Li, C. Li, and C.-H. Chen, "Visual recognition based on deep learning for navigation mark classification," *IEEE Access*, vol. 8, pp. 32,767–32,775, Feb. 2020, doi: 10.1109/ACCESS.2020.2973856.
[134] Z. Xu, Y. Huo, K. Liu, and S. Liu, "Detection of ship targets in photoelectric images based on an improved recurrent attention convolutional neural network," *Int. J. Distrib. Sensor Netw.*, vol. 16, no. 3, p. 155,014,772,091,295, Mar. 2020, doi: 10.1177/1550147720912959.

[135] H. Fu, G. Song, and Y. Wang, "Improved YOLOv4 marine target detection combined with CBAM," *Symmetry*, vol. 13, no. 4, p. 623, Apr. 2021, doi: 10.3390/sym13040623.
[136] T. Liu, B. Pang, L. Zhang, W. Yang, and X. Sun, "Sea surface object detection algorithm based on YOLO v4 fused with reverse depthwise separable convolution (RDSC) for USV," *J. Marine Sci. Eng.*, vol. 9, no. 7, p. 753, Jul. 2021, doi: 10.3390/jmse9070753.
[137] J. Yu, Y. Jiang, Z. Wang, Z. Cao, and T. Huang, "UnitBox: An advanced object detection network," in *Proc. 24th ACM Int. Conf. Multimedia*, Oct. 2016, pp. 516–520, doi: 10.1145/2964284.2967274.
[138] H. Rezatofighi, N. Tsoi, J. Gwak, A. Sadeghian, I. Reid, and S. Savarese, "Generalized intersection over union: A metric and a loss for bounding box regression," Apr. 2019. [Online]. Available: http://arXiv.org/abs/1902.09630
[139] Z. Zheng, P. Wang, W. Liu, J. Li, R. Ye, and D. Ren, "Distance-IoU loss: Faster and better learning for bounding box regression," Nov. 2019. [Online]. Available: http://arXiv.org/abs/1911.08287
[140] A.-J. Gallego, A. Pertusa, and P. Gil, "Automatic ship classification from optical aerial images with convolutional neural networks," *Remote Sens.*, vol. 10, no. 4, p. 511, Mar. 2018, doi: 10.3390/rs10040511.
[141] A. S. Kumar and E. Sherly, "A convolutional neural network for visual object recognition in marine sector," in *Proc. 2nd Int. Conf. Convergence Technol.*, Mumbai, India, 2017, pp. 304–307, doi: 10.1109/i2ct.2017.8226141.
[142] Q. Shi, W. Li, F. Zhang, W. Hu, X. Sun, and L. Gao, "Deep CNN with multi-scale rotation invariance features for ship classification," *IEEE Access*, vol. 6, pp. 38,656–38,668, Jul. 2018, doi: 10.1109/ACCESS.2018.2853620.
[143] M. Fiorini, D. D. Bloisi, A. Youssef, and A. Pennisi, "Fine-grained boat classification using convolutional neural networks," *e-Navi Int. J. e-Navigation Maritime Econ.*, vol. 8, pp. 60–67, Dec. 2017.
[144] K. Ji, X. Xing, W. Chen, H. Zou, and J. Chen, "Ship classification in TerraSAR-X SAR images based on classifier combination," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Melbourne, Australia, Jul. 2013, pp. 2589–2592, doi: 10.1109/IGARSS.2013.6723352.
[145] T. Cane and J. Ferryman, "Evaluating deep semantic segmentation networks for object detection in maritime surveillance," in *Proc. 15th IEEE Int. Conf. Adv. Video Signal Based Surveillance (AVSS)*, Auckland, New Zealand, Nov. 2018, pp. 1–6, doi: 10.1109/AVSS.2018.8639077.
[146] M. Leclerc, R. Tharmarasa, M. C. Florea, A.-C. Boury-Brisset, T. Kirubarajan, and N. Duclos-Hindie, "Ship classification using deep learning techniques for maritime target tracking," in *Proc. 21st Int. Conf. Inf. Fusion (FUSION)*, Cambridge, U.K., Jul. 2018, pp. 737–744, doi: 10.23919/ICIF.2018.8455679.
[147] D. D. Bloisi and L. Iocchi, "ARGOS - A video surveillance system for boat traffic monitoring in Venice," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 23, no. 7, pp. 1477–1502, Nov. 2009, doi: 10.1142/S0218001409007594.
[148] D. D. Bloisi, A. Pennisi, and L. Iocchi, "Background modeling in the maritime domain," *Mach. Vis. Appl.*, vol. 25, no. 5, pp. 1257–1269, Jul. 2014, doi: 10.1007/s00138-013-0554-5.
[149] D. D. Bloisi, L. Iocchi, A. Pennisi, and L. Tombolini, "ARGOS-Venice boat classification," in *Proc. 12th IEEE Int. Conf. Adv. Video Signal Based Surveillance (AVSS)*, Karlsruhe, Germany, Aug. 2015, pp. 1–6, doi: 10.1109/AVSS.2015.7301727.
[150] M. Kristan, V. Sulic Kenk, S. Kovacic, and J. Pers, "Fast image-based obstacle detection from unmanned surface vehicles," *IEEE Trans. Cybern.*, vol. 46, no. 3, pp. 641–654, Mar. 2016, doi: 10.1109/TCYB.2015.2412251.
[151] B. Bovcon, R. Mandeljc, J. Perš, and M. Kristan, "Stereo obstacle detection for unmanned surface vehicles by IMU-assisted semantic segmentation," *Robot. Auton. Syst.*, vol. 104, pp. 1–13, Jun. 2018, doi: 10.1016/j.robot.2018.02.017.
[152] Y.-T. Chan, "Maritime filtering for images and videos," *Signal Process., Image Commun.*, vol. 99, p. 116,477, Nov. 2021, doi: 10.1016/j.image.2021.116477.
[153] S. Moosbauer, D. Konig, J. Jakel, and M. Teutsch, "A benchmark for deep learning based object detection in maritime environments," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Long Beach, CA, USA, Jun. 2019, pp. 916–925, doi: 10.1109/CVPRW.2019.00121.
[154] X. Tang, D. K. Du, Z. He, and J. Liu, "PyramidBox: A context-assisted single shot face detector," Aug. 2018. [Online]. Available: http://arXiv.org/abs/1803.07737.
[155] C. Land and H. Edwin, "The retinex theory of color vision," *Scientific Amer.*, vol. 237, no. 6, pp. 108–128, 1977.
[156] R. T. Tan, "Visibility in bad weather from a single image," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Anchorage, AK, USA, Jun. 2008, pp. 1–8, doi: 10.1109/CVPR.2008.4587643.
[157] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 12, pp. 2341–2353, Sep. 2010, doi: 10.1109/TPAMI.2010.168.
[158] X. Wang, T. Xiao, Y. Jiang, S. Shao, J. Sun, and C. Shen, "Repulsion loss: Detecting pedestrians in a crowd," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Salt Lake City, UT, USA, Jun. 2018, pp. 7774–7783, doi: 10.1109/CVPR.2018.00811.
[159] L. Karlinsky et al., "RepMet: Representative-based metric learning for classification and one-shot object detection," Nov. 2018. [Online]. Available: http://arXiv.org/abs/1806.04728
[160] M. Ren et al., "Meta-learning for semi-supervised few-shot classification," Mar. 2018. [Online]. Available: http://arXiv.org/abs/1803.00676
[161] B. Demirel, R. G. Cinbis, and N. Ikizler-Cinbis, "Zero-shot object detection by hybrid region embedding," May 2018. [Online]. Available: http://arXiv.org/abs/1805.06157
[162] S. Rahman, S. Khan, and F. Porikli, "Zero-shot object detection: Learning to simultaneously recognize and localize novel concepts," Mar. 2018. [Online]. Available: http://arXiv.org/abs/1803.06049
[163] M. Chendeb El Rai, J. H. Giraldo, M. Al-Saad, M. Darweech, and T. Bouwmans, "SemiSegSAR: A semi-supervised segmentation algorithm for ship SAR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022, doi: 10.1109/LGRS.2022.3185306.
[164] X. Shan, M. Pan, D. Zhao, D. Wang, F.-J. Hwang, and C.-H. Chen, "Maritime target detection based on electronic image stabilization technology of shipborne camera," *IEICE Trans. Inf. Syst.*, vol. E104.D, no. 7, pp. 948–960, Jul. 2021, doi: 10.1587/transinf.2020EDP7192.
[165] J. Muhovic, R. Mandeljc, B. Bovcon, M. Kristan, and J. Pers, "Obstacle tracking for unmanned surface vessels using 3-D point cloud," *IEEE J. Ocean. Eng.*, vol. 45, no. 3, pp. 786–798, Jul. 2020, doi: 10.1109/JOE.2019.2909507.
[166] B. Bovcon and M. Kristan, "A water-obstacle separation and refinement network for unmanned surface vehicles," in *Proc. IEEE Int. Conf. Robot. Automat. (ICRA)*, Paris, France, May 2020, pp. 9470–9476, doi: 10.1109/ICRA40945.2020.9197194.
[167] J. Liu and G. Guo, "A recursive estimator for pseudolinear target motion analysis using multiple hybrid sensors," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–9, Jul. 2021, doi: 10.1109/TIM.2021.3097400.
[168] G. Guo and J. Liu, "A stochastic model-based fusion algorithm for enhanced localization of land vehicles," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–10, 2022, doi: 10.1109/TIM.2021.3137566.
[169] J. Liu and G. Guo, "Vehicle localization during GPS outages with extended Kalman filter and deep learning," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–10, Jul. 2021, doi: 10.1109/TIM.2021.3097401.
[170] S. Thombre et al., "Sensors and AI techniques for situational awareness in autonomous ships: A review," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 1, pp. 64–83, Jan. 2022, doi: 10.1109/TITS.2020.3023957.
[171] X. Chen et al., "Ship type recognition via a coarse-to-fine cascaded convolution neural network," *J. Navigation*, vol. 73, no. 4, pp. 813–832, Jul. 2020, doi: 10.1017/S0373463319000900.
[172] R. Gonzales-Martinez, J. Machacuay, P. Rotta, and C. Chinguel, "Hyperparameters tuning of faster R-CNN deep learning transfer for persistent object detection in radar images," *IEEE Latin Amer. Trans.*, vol. 20, no. 4, pp. 677–685, Apr. 2022, doi: 10.1109/TLA.2022.9675474.
[173] Z. Zhang, T. He, H. Zhang, Z. Zhang, J. Xie, and M. Li, "Bag of freebies for training object detection neural networks," Apr. 2019. [Online]. Available: http://arXiv.org/abs/1902.04103
[174] Q. Zhao et al., "M2Det: A single-shot object detector based on multi-level feature pyramid network," Jan. 2019. [Online]. Available: http://arXiv.org/abs/1811.04533

ITS