Liu, Xintong; Hu, Yutian; Ji, Huiting; Zhang, Mingyang; Yu, Qing

A Deep Learning Method for Ship Detection and Traffic Monitoring in an Offshore Wind Farm Area

Article

A Deep Learning Method for Ship Detection and Traffic Monitoring in an Offshore Wind Farm Area

Xintong Liu [1,†] , **Yutian Hu** [1,†] , **Huiting Ji** [1,†] , **Mingyang Zhang** [2] and **Qing Yu** [1,3,*]

1   Maritime Risk & Behavioral Sciences Lab, School of Navigation, Jimei University, Xiamen 361021, China; 202212861022@jmu.edu.cn (X.L.); 202021011031@jmu.edu.cn (Y.H.); 202021012061@jmu.edu.cn (H.J.)
2   Marine Technology Group, Department of Mechanical Engineering, Aalto University, 00076 Espoo, FI, Finland; mingyang.0.zhang@aalto.fi
3   Hubei Key Laboratory of Inland Shipping Technology, Wuhan University of Technology, Wuhan 430063, China
*   Correspondence: qing.yu@jmu.edu.cn
†   These authors contributed equally to this work.

Abstract: Newly built offshore wind farms (OWFs) create a collision risk between ships and installations. The paper proposes a real-time traffic monitoring method based on machine vision and deep learning technology to improve the efficiency and accuracy of the traffic monitoring system in the vicinity of offshore wind farms. Specifically, the method employs real automatic identification system (AIS) data to train a machine vision model, which is then used to identify passing ships in OWF waters. Furthermore, the system utilizes stereo vision techniques to track and locate the positions of passing ships. The method was tested in offshore waters in China to validate its reliability. The results prove that the system sensitively detects the dynamic information of the passing ships, such as the distance between ships and OWFs, and ship speed and course. Overall, this study provides a novel approach to enhancing the safety of OWFs, which is increasingly important as the number of such installations continues to grow. By employing advanced machine vision and deep learning techniques, the proposed monitoring system offers an effective means of improving the accuracy and efficiency of ship monitoring in challenging offshore environments.

Keywords: traffic safety; offshore wind farms; YOLOv7; stereo vision; deep learning

1. Introduction

Motivated by the demands for clean energy in the context of ongoing climate change, the number of offshore wind farms (OWFs) grows rapidly in lots of coastal countries [1]. The present development shows the advantages of offshore wind farms, for example, delivering secure, affordable, and clean energy while fostering industrial development and job creation. Based on the annual report from the Global Wind Energy Council, 2021 becomes the best year in the offshore wind industry, in which 21.1 GW of offshore wind capacity was commissioned, bringing global offshore wind capacity reach to 56 GW, three times more than in 2020. The GWEC Market Intelligence forecasts that 260 GW of new offshore wind capacity could be added during 2022–2030 under the current positive policies, allowing the total global offshore wind installations to rise from 23% growth in 2021 to at least 30% growth by 2031. In the year 2021, China constructed 80% of new offshore installations worldwide becoming the world's largest offshore market [2].

However, this trend of increasing numbers of offshore wind farms puts pressure on local water traffic management. For example, the newly added obstacles increase the difficulties for both the navigation of passing vessels [3] and of Search and Rescue. Once an accident happens, it results in water pollution, significant damage to facilities and other catastrophic casualties and economic losses. For the year 2021, the Global Offshore Wind Health and Safety Organisation (G+) report pointed out that there were a total of

204 high-potential incidents and injuries recorded [4]. Recently, some collision accidents were also reported for the UK, China, and the Netherlands, leading to shipping hull and turbine damage, and electric power loss, especially for construction and fishing ships. For instance, on 31 January 2021, the drifting bulk carrier Julietta D collided with a transformer platform in the Hollandse Kust Zuid wind farm, which was under construction. On 2 July 2022, a dragging accident in southern waters led to 25 casualties and a vessel sank. Therefore, monitoring vessels in the offshore wind farm waters and detecting potential hazards becomes an urgent question for stakeholders of offshore wind farms.

The current monitoring system for water traffic is the Vessel Traffic Services (VTS), which uses the Automatic Identification System (AIS), radar and other detection sensors to show the water traffic situation dynamically. The topics of technology development, information collection, communication, and system design have been studied for the VTS. However, the limitations of its use in offshore wind farm waters are noted due to trespasses and the deliberate turning off of the AIS. Developing a reliable monitoring system can aid the safety of navigation of passing vessels and offshore wind farms. To improve the reliability of using VTS in offshore wind farm water, previous studies proposed several novel methodologies to develop monitoring models. Li et al. reviewed 153 papers related to ship detection and classification and suggested that the current ship detection methods could be divided into two categories, feature design-based methods and Convolutional Neural Network (CNN) based methods [5]. For instance, Zhao et al. developed a coupled CNN model to detect small ships in waterways with a high density of shipping water using a clustered Synthetic Aperture Radar image approach [6]; Zhang et al. proposed a fast region-based convolutional neural network (RCNN) method to detect ships from high-resolution remote sensing images [7]. Using feature design-based methods, Hu et al. employed dilated rate selection and attention guided feature representation strategies for ship detection, in which the dilated convolution selection strategy was applied to a multi-branch extraction network, extracting context information of different receptive fields [8]. Yu et al. employed an advanced ship detection method using Synthetic Aperture Radar images based on the YOLOv5 [9]. The results of this study showed a great improvement in the feature expression ability and sensitivity for target detection after adding a coordinate attention mechanism and a feature fusion network structure. Priakanth et al. proposed a hybrid system by using wireless and IoT technology to avoid boundary invasions [10]. Ouelmokhtar et al. suggested using Unmanned Surface Vehicles (USV) to monitor waters, in which an onboard RADAR is used to detect the targets [11]. Nyman discussed the possibility of using satellites for monitoring, which allows for the visual surveillance of a large area [12]. The relatively low cost of data acquisition makes the use of satellites appealing, however, some prior theory or knowledge is needed to sort through the vast collection of satellite data and images. Although these studies have their advantages, these technologies may typically focus on large water areas, increasing the financial burden for offshore wind farms. Therefore, low-cost, high efficiency and reliable systems are still needed. The low cost of equipment makes video surveillance technologies a possible way to monitor water traffic at a close distance. However, original video surveillance requires human supervision to achieve continuous monitoring, consuming a large number of human resources. In addition, the challenges faced by offshore wind farm operators and managers in the environment and the high workload of their daily tasks may lead to error-prone and tedious situations. Nowadays, automated surveillance systems observing the environment utilizing cameras are widely used. These automatic systems can accomplish a multitude of tasks, which include detection, interpretation, understanding, recording, and creating alarms based on the analysis. They are widely used in different situations, for instance, in road traffic, Pawar and Attar got detection and localization of road accidents from traffic surveillance videos [13]. Vieira et al. and Thallinger et al. quantitatively evaluated the safety of traffic by utilizing security cameras [1,14]. A video surveillance-based system that can detect the pre-events, with an automatic alarm generated in the control room was proposed for improving road safety [15].

Motivated by the above-mentioned difficulties and advantages, this study aimed to pioneer the use of machine vision technologies to aid traffic monitoring in offshore wind farm waters. Specifically, a vision-based monitoring system was developed for OWFs to improve the reliability and efficiency of ship traffic detection and tracking. The system trains a "YOLOv7" machine visual model using automatic identification system (AIS) data. The fine-trained model can identify the passing ship in OWF waters, not only providing the identified target, but also the degree of confidence during the monitoring process. Then a stereo vision algorithm is applied in the model to locate and track the positions of the passing ships. So that the dynamic information (e.g., speed, course, and position) for each target can be provided by the system. In addition, the proposed system was validated by comparing the provided dynamic information with AIS data to ensure the reliability of the results. The contributions of this work are highlighted as follows. The study pioneers the uses of machine vision to aid traffic monitoring in offshore wind farm waters, covers the gap for the current VTS system and provides a novel way for offshore water traffic management. In addition, the model can provide more efficient and reliable dynamic data for individual and overall traffic that can be expanded to providing early warning for ship risk and accident prevention. The main contributions of this work are summarized as follows:

(1) This paper evaluates the performance of using the YOLOv7 model in offshore wind farm waters. It shows that the YOLOv7 model has a high accuracy rate for vessel detection. The results prove that the YOLOv7 models have higher dynamicity for ship detection compared to other ship detection methods.

(2) An optimization strategy for training a visual-based identification model is presented. The study collects hybrid data sources (e.g., AIS data, images) to develop the model training database and validate the model by comparing the positions between AIS data and the detection results. consequently, this ensures the performance of continuous target detection with significant accuracy.

(3) The proposed target identification system was further tested in an offshore case. By using an embedded device, the inference time reached real-time performance (less than 0.1 s) and the overall processing time for one frame was 0.76 s, proving the possibility of implementation of the system in real-time ship traffic monitoring.

The remainder of this paper is organized as follows: Section 2 outlines a review of related research. In Section 3, the framework of the system is introduced in detail. The experimental data and the test results are reported in Section 4. The obtained results are discussed in Section 5. Finally, conclusions are drawn in Section 6.

2. Literature Review

This section can be categorized into three groups based on the method used in the studies: ship monitoring, machine vision in target detections, and machine vision in target tracking, respectively.

2.1. Ship Monitoring Technology

Implemented to promote safe and efficient marine traffic, VTS are typically the most widely used technology for ship supervision. It is a shoreside service within a country's territorial waters that, by detecting and tracking the ship, aims to monitor the traffic, assist the traffic control and manage navigational matters, and provide support and required information for passing ships. The VTS collects dynamic data via two sensors, the radar and AIS; however, both have their limitations. For instance, the radar echo can be disturbed in an environment with the external noise of RF interference and clutter, which creates potentially dangerous situations and decrease VTS functionality. Regarding this matter, Root proposed high-frequency radar ship detection by clutter elimination [16]. Dzvonkovskaya et al. pioneered a new detection algorithm in ship detection and tracking but ignored the external influences position [17]. Another question concerns the limitations of detecting small ships in coastal waters (e.g., offshore wind farm water). From this point

of view, Margarit et al. proposed a ship operation monitoring system based on Synthetic Aperture Radar image processing to achieve inferred ship monitoring and classification information, which further improves the Synthetic Aperture Radar image from the new sensor data [18]. Moreover, radar is unable to provide sufficient static information such as ship type and size, which means other systems (e.g., AIS) are needed. The AIS is another important piece of information for ship monitoring, which can compensate for the limitations of radar. It achieves the automatic exchange of ship information and navigation status between ships and shore. As a type of reliable data source, AIS data has been widely used in lots of studies to analyze ship traffic, making it important for water traffic management. For instance, Brekke et al. combined AIS data with satellite Synthetic Aperture Radar images to detect ship dynamic information (e.g., speed, course) [19]. To improve the reliability of radar, Stateczny collected sets of data from AIS and radar, and then applied a numerical model to compare the covariance between the two types of data [20]. Pan and Deng proposed a real-time monitoring system for shore-based monitoring of ship traffic [21]. Although the AIS data are valuable for ship traffic management, several limitations remain. For instance, the AIS can be closed manually. The AIS is not mandatory for some small ships such as dinghies and fishing boats [22].

In coastal waters, the offshore wind farm is a relatively new type of installation that influences the existing ship traffic, not only occupying the navigable waters but also creating blind areas by blocking radio signals, reducing the ability to detect and track small targets. Relevant studies using traditional data sources (e.g., AIS, radar), including Yu et al., used AIS to analyze the characteristics of the ship traffic in the vicinity of offshore wind farms [23] and then developed models to assess the risk for individual ships [24] or for the ship traffic flows [25]. However, as a highly accurate and reliable model for ship detection is still the basis for ship monitoring in offshore waters, traditional ship detection methods (e.g., AIS, radar) were updated to overcome the potential uncertainties, such as small target detection, visual monitoring, and unpredicted invasions.

Consequently, some vision-based technologies have been applied for ship detection, ship tracking, and ship monitoring. For instance, to overcome the difficulty of remote ship control and monitoring in high traffic waters, Liu et al. designed a portable integrated ship monitoring and commanding system [26]. To test the data availability, Shao et al. used image data captured from surveillance cameras to achieve target detection [27]. To improve the function of the target automatic monitoring and tracking, Chen et al. proposed a mean-shift ship monitoring and tracking system [28], which shows the possibility of using machine visual technologies for water traffic monitoring.

2.2. Applications of Machine Vision in Target Detections

Machine vision technology enables a machine with a visual perception system, with the help of hardware (e.g., camera, infrared thermal imager, night vision device) and a software program. It can recognize and manipulate the activities and perform image-based process control and surveillance for traffic monitoring, manufacture inspection, autopilot, and other scene perception usages [29,30]. A widely used application of machine vision is the Tesla driverless system, which was equipped with the hardware needed for Autopilot and the software program to realize Full Self-Driving.

Central to machine vision is the target recognition and detection algorithms. Figure 1 shows the development of recognition and detection algorithms that are used for object detection. The early studies of machine vision came up with Scale Invariant Feature Transform, which involved five steps to match the similarity of two images and to detect targets [31]. Then SIFT was upgraded to the Viola-Jones detection algorithm [32], histogram of oriented gradients (HOG) [33], Data Management Platform (DMP), and so on.
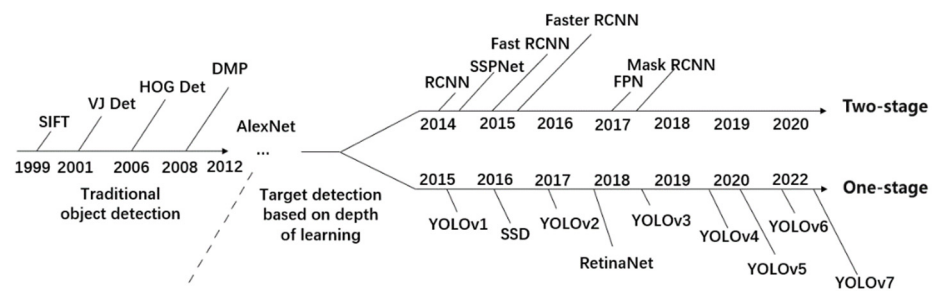
Figure 1. Development of machine vision technologies.

However, the above-mentioned algorithms extract target features manually, and can only perform well when they are guaranteed to extract sufficiently accurate features, so that they are inapplicable where a large number of targets exist. They are replaced by applying deep learning approaches to detect targets. Deep learning-based detection algorithms have the advantage of being able to extract features from complex images. The deep learning-based methods can be grouped into two categories based on the way they extract target features: the anchor-based methods (i.e., Convolutional Neural Network (CNN) methods and You Only Look Once (YOLO) methods) and the anchor-free methods (e.g., adaptively spatial feature fusion methods [34], Corner Net methods [35]). The Anchor-based algorithms are further classified into single-stage detection and two-stage detection. Due to stability and accuracy, anchor-based methods have become more popular in recent years. Typical methods include YOLOv1-v5 [16,36–40], single shot multibox detector [41], and Region CNN etc. For instance, Girshick et al. [42] proposed a novel method of R-CNN for target detection. The method uses image segmentation combining regions and CNNs to improve accuracy; however, it requires a large database to train the detection model, which reduces the detection speed. To improve detection speed, Meng et al. developed an improved Mask R-CNN, which ignores the RolAlign layer in the R-CNN [43]. Zhao et al. suggested enhancing the relationship among non-local features and refining the information on different feature maps to improve the detection performance of R-CNN [44]. Redmon and Farhadi proposed a joint training method to improve the traditional YOLOv1 model [45]. The upgraded YOLOv3 model uses binary cross-entropy loss and scale prediction to improve the accuracy of the model while ensuring the detection speed of detection [16]. The YOLOv3 model was adopted in vision detection studies including Gong et al. [46]. Patel et al. proposed a novel deep learning approach, which combines the capabilities of Graph Neural Networks (GNN) and the You Look Only Once (YOLOv7) deep learning framework; the results show an increased detection accuracy [47], Li et al. [48] and etc., which prove its fast speed in the convergence and detection process. The applications of vision detection have been done in various domains, as well as water traffic management. To design a deep learning-based detector for ship detection, Li et al. applied the Faster-CNN algorithm to train the ship target detection model, which achieves higher accuracy [49]. To address the shortcomings of the region proposal computation, Ren et al. introduced a region proposal network (RPN) by sharing the convolutional features of Fast R-CNN and RPN to further merge the two into one network [50].

2.3. Applications of Machine Vision in Target Position and Tracking

Machine vision methods used for target tracking can be categorized as monocular vision and binocular vision based on the tracking mechanism [51,52]. The monocular vision was first proposed by Davison [53], who used overall decomposition sampling to solve the challenge of real-time feature initialization. The basis of the monocular vision systems is the simultaneous localization and mapping (SLAM) method, which calculates the distance of the target within the camera's field of view. Although monocular visual localization is simple to operate and does not require feature point matching, it is less accurate and only suitable for specific environments. Therefore, it is not suitable for use in complex environments such as maritime target localization and tracking. To solve these

problems, binocular vision positioning has been proposed and is widely used in many fields. However, binocular vision pre-localization only works on a flat surface and cannot accurately localize objects, so scholars have extended binocular vision to stereoscopic vision [54,55].

Binocular stereo vision technology can simulate the human eye to perceive the surrounding environment in three dimensions, making it widely used in various fields [56–59]. To reduce errors in the localization part of binocular stereo vision systems due to interference from complex environments, Zou et al. proposed a binocular stereo vision system based on a virtual manipulator and the localization principle [60]. They designed a binocular stereo vision measurement system to achieve an accurate estimation of target object positions. Zuo et al. used binoculars to capture point and line features and selected orthogonality as the minimum parameter for feature extraction, which solved the problem of unreliability of binocular stereo vision in detecting objects [61]. Therefore, compared to monocular vision techniques, binocular stereo vision is a more effective technique for target tracking. It is more accurate, simpler to operate, and suitable for dynamic environments, making binocular vision systems more appropriate for ship supervision in offshore wind farms than monocular vision systems. Video tracking allows for continuous monitoring of the ship, which is very helpful.

Based on the above studies, it can be noted that traditional approaches of using AIS and radar data for ship identification, tracking and localization have their drawbacks. For example, radar has difficulty in detecting small targets and fails to provide some important information (i.e., ship type). The AIS provides dynamic and static information for individual ships but it can be turned off manually, which cause difficulties in ship tracking and invasion detection. Hence, this study suggests a novel way for ship detection and tracking by using a visual-based target detection model which not only overcomes the difficulties of current ship detection and tracking approaches but also provides new ideas to enhance dynamic ship monitoring in coastal waters.

3. Methods

This section proposes a novel framework for ship detection and tracking in the waters near OWFs. The framework uses the vanilla YOLOv7 algorithm as a critical component to detect ships in dynamic situations while applying binocular stereo vision to track the ships. The framework structure is shown in Figure 2.

The framework consists of seven steps. (1) to collect real-time ship video from the waters in the vicinity of the OWFs; (2) to process the collected ship video and picture information; (3) to set the relevant parameters; (4) to construct the training database from the collected video and AIS data; (5) to train the ship detection model using the YOLOv7 approach; (6) to map the ship position from videos into the physical world with the aid of binocular stereo vision and (7) to validate and output the results. The details of those steps are introduced as follows.

Step 1: Cameras placed at wind turbines are used to provide real-time videos of ship traffic in the vicinity of offshore waters. Then, images containing ships are collected to develop the training databases. In the database, images that may have ships that can be overlapped with another image are deleted.

Step 2: This step defines the parameters used in the monitoring system. For the detection model (i.e., YOLOv7), the three parameters learning rate, momentum, and decay are used. The learning rate is a hyperparameter used by the optimizer to control the step size of the parameter update. It can influence the adjustment of the parameter value. Momentum is a regulating hyperparameter in the optimizer, which is used to control the direction and speed of parameter updates. By setting a suitable momentum hyperparameter, the parameter updates can be made more stable and avoid problems such as excessive oscillation or too slow descent during training. Decay or decay rate is a regulating hyperparameter in the optimizer that is used to make the learning rate gradually

decrease. It allows for a more detailed search of the parameter space, making it easier for the model to reach a minimum and produce more accurate predictions.

In addition, the camera includes internal and external parameters. The internal parameters are focal, pixel, and resolution. They are mainly used as the basis to transform the ship's position from pixel coordinates to camera coordinates. The external parameters cover camera positions, and high, horizontal, and vertical angles, which convert the ship's position from camera coordinates to world coordinates. By calibrating the left and right cameras, those internal and external camera parameters are crucial for ship positioning.

Step 3: Establish the initial training samples for YOLO model training, which is used to obtain the object in the port videos (i.e., generate a bounding box for each ship in the video). Before training, the ship pictures need to be processed through annotation using an image annotation tool. Each ship in the picture is selected in this study, and the corresponding AIS data are input to develop the database. The database includes the position coordinates of the corners of the ship's box, as well as the width and height of the ship in the picture. In the training process, standard techniques such as multi-scale training, data augmentation, and batch normalization are used to train the ship detector.
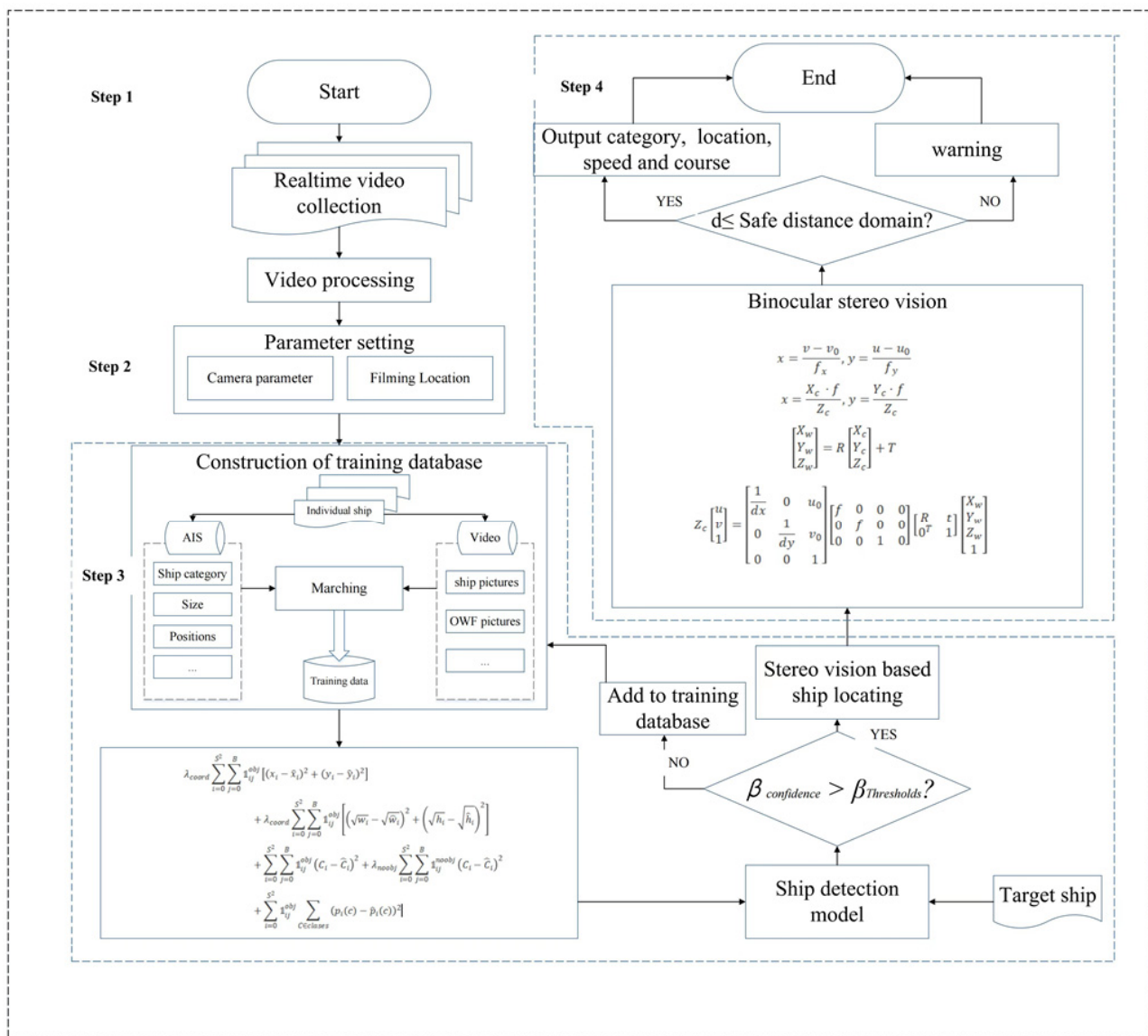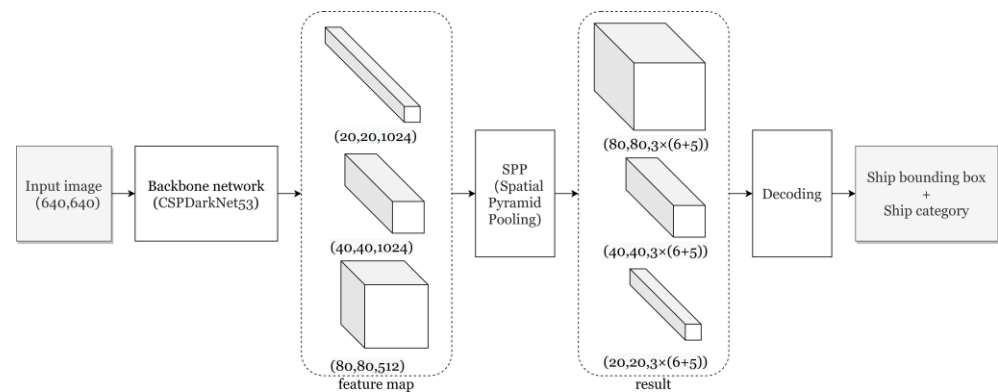

**Figure 2.** The framework for the monitoring system.

Figure 3 shows the training process of the YOLOv7 model, which consists of the backbone network, a convolutional feature fusion network, and the decoding processing.


Figure 3. The steps to train the YOLOv7 model.

The YOLOv7 model consists of a backbone, neck, and a head. The input image is scaled to a size of 640 × 640 with a constant aspect ratio, and the blank part of the image is filled with gray. As shown in Figure 3, the input image is grided as 640 × 640 to detect the target. The backbone is a convolutional neural network responsible for extracting features from a given input image; the neck mainly generates a feature pyramid and passes various scales of features to the head, which generates prediction boxes and ship categories as outputs for each object. In addition, sensitivity and accuracy can be affected by the size of the grids. As shown in Figure 4, a scale of 20 × 20 is used for big target detection, a medium scale of 40 × 40 is used to detect the middle target, and a small scale of 80 × 80 for the small target.

(a)　　　　　　　　　　(b)　　　　　　　　　　(c)
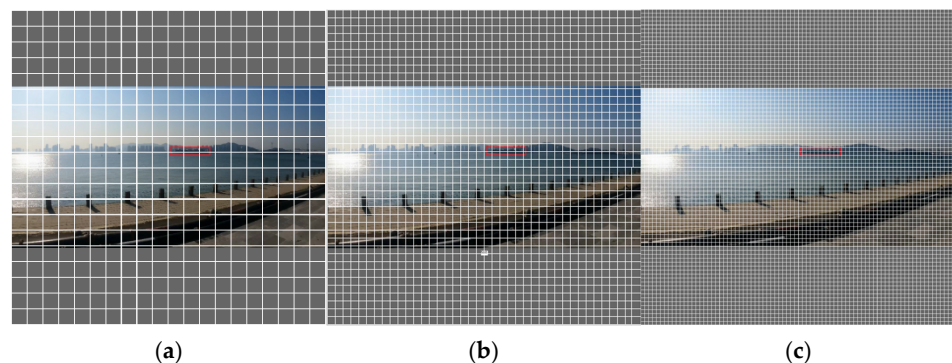
Figure 4. The sizes of the grids to capture ships: (**a**) 20 × 20, large ships; (**b**) 40 × 40, medium ships; (**c**) 80 × 80, small ships.

Step 4: This study uses a 3D reconstruction technique called binocular stereo vision to achieve ship positioning. Stereo vision is a technique that involves detecting objects using two or more cameras. By simultaneously calibrating the left and right cameras, the internal and external parameters of both cameras can be determined. To obtain the position of the ship, the target coordinates are mapped from the video to the physical world using imaging principles, the whole structure is shown in Figure 5.
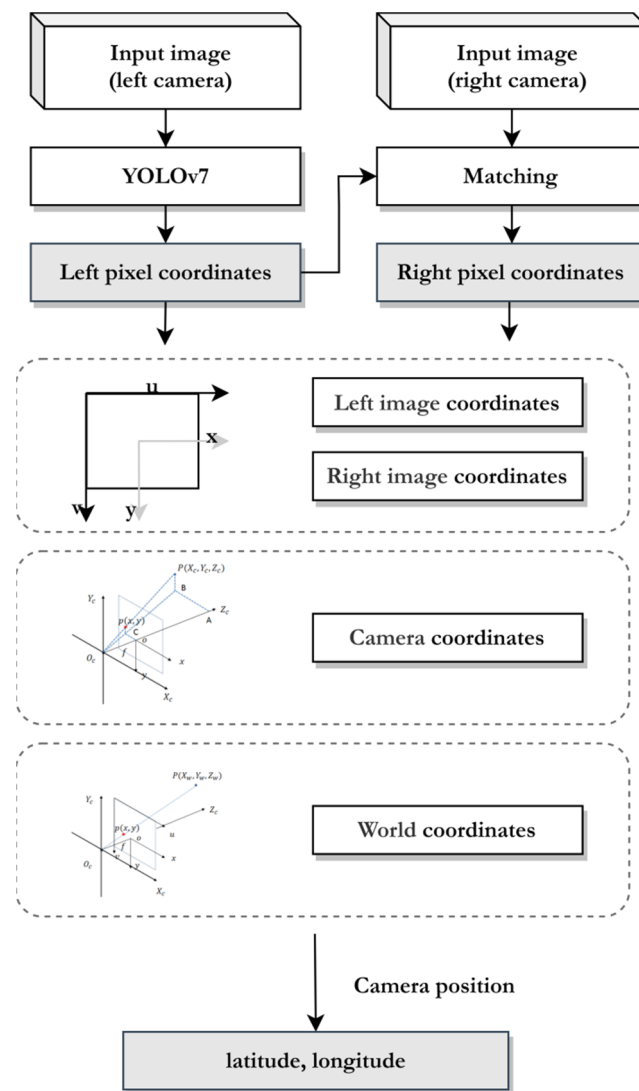
Figure 5. The steps for ship detection and positioning.

In the proposed system, the left pixel coordinate $\left(u_{left}, v_{left}\right)$ is obtained from the input image captured by the left camera into YOLOv7. Meanwhile, the right pixel coordinate $\left(u_{right}, r_{right}\right)$ is obtained after matching the image captured by the right camera and left camera.

The pixel point is denoted by $\boldsymbol{m} = [u, v]^T$. The world point is denoted by $\boldsymbol{M} = [X, Y, Z]^T$. We use $\widetilde{x}$ to denote the augmented vector by adding 1 as the last element: $\widetilde{\boldsymbol{m}} = [u, v, 1]^T$, $\widetilde{\boldsymbol{M}} = [X, Y, Z, 1]^T$. The relationship between the world point $\boldsymbol{M}$ and its pixel projection $\boldsymbol{m}$ is given by:

$$s\widetilde{\boldsymbol{m}} = \boldsymbol{A}[\boldsymbol{R}\ \boldsymbol{T}]\ \widetilde{\boldsymbol{M}}, \text{ with } \boldsymbol{A} = \begin{bmatrix} \alpha & \gamma & u_0 \\ 0 & \beta & v_0 \\ 0 & 0 & 1 \end{bmatrix} \tag{1}$$

where s is an arbitrary scale factor, the extrinsic parameters $(\boldsymbol{R}, \boldsymbol{T})$ is the rotation and translation, which related the world coordinate system to the camera coordinate system, and $\boldsymbol{A}$ is the camera intrinsic matrix, with (u_0, v_0) the coordinates of the principal point, α and β the scale factors in u and v axes, and γ the parameter describing the skew of the two axes.

In order to obtain the relative positions between any two coordinate systems, the rotation $\boldsymbol{R}$ and translation $\boldsymbol{T}$ need to be acquired by calibrating the left and right cameras,

simultaneously. This calibration process involves capturing images of a checkerboard pattern at different orientations, as shown in Figure 6. The images are then processed using the "Stereo Camera Calibrator" tool in MATLAB to obtain the camera parameters, including the rotation and translation matrices.
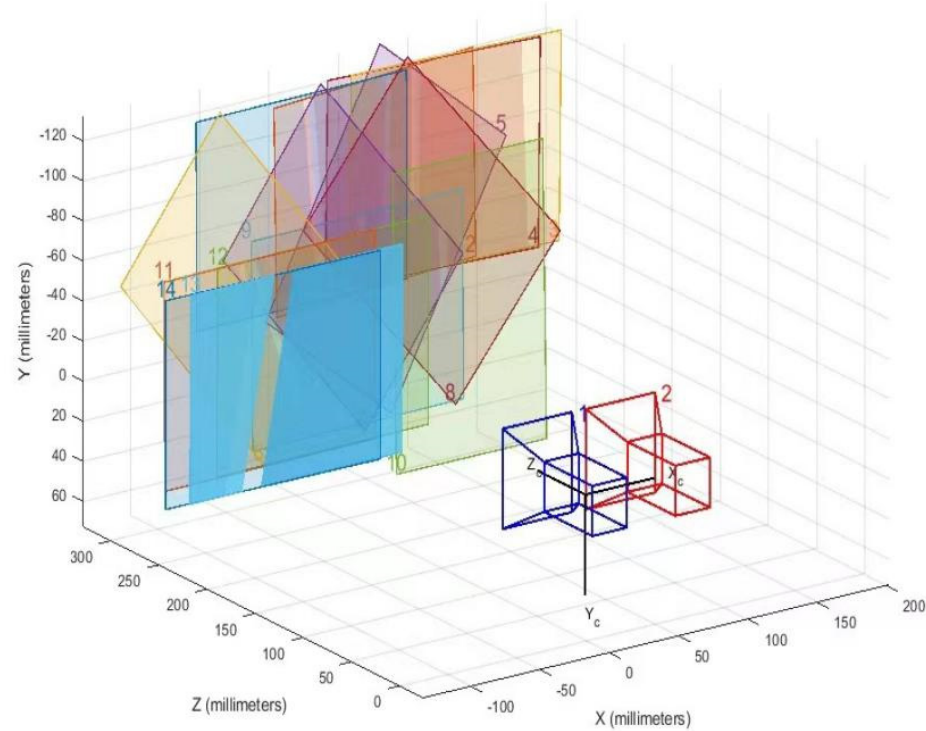
Figure 6. The checkerboard calibration.

As a result, the world coordinate $[X, Y, Z]^T$ of the ship is obtained. The latitude–longitude coordinate of the ship is denoted by $\boldsymbol{P} = \left[P_{latitude}, P_{longitude} \right]^T$ can be formulated as follows:

$$P = P_{camera} + \begin{bmatrix} X \\ Y \end{bmatrix} \qquad (2)$$

where $\boldsymbol{P_{camera}} = \left[P_{latitude}^{camera}, P_{longitude}^{camera} \right]^T$ is the latitude–longitude coordinate of the camera.

4. Case Study

4.1. Database and Processing

In this experiment, our team collected a total of 1000 images of marine ships to form the MYSHIP database[1]. It had a resolution of either 1920×1080 or 2840×2160. Because the training of a convolutional neural network requires a considerable number of samples, we also added the SEASHIP database to our database. The SEASHIP database consists of a total of 7000 images with a resolution of 1920×1080. As shown in Figure 7, the database is divided into six categories of ships: ore carriers, bulk cargo carriers, general cargo ships, container ships, fishing boats, and passenger ships.

Ore carrier	General cargo ship

Bulk cargo carrier	Fishing boat

Passenger ship	Container ship

Figure 7. The six types of ships included in the training database.

We divide the database into three parts: the training set, the validation set, and the test set in the proportion 6:2:2. The division of the database is shown in Table 1.

Table 1. The number of images in each database.

Database	Number of Samples		Number of Ships	
	MYSHIP	**SEASHIP**	**MYSHIP**	**SEASHIP**
Training set	600	4200	1334	4934
Validation set	200	1400	444	1610
Test set	200	1400	446	1669
Total	1000	7000	2224	8213

4.2. Parameter Setting

In this experiment, the parameters of the camera used to capture ship video and image data are shown in Table 2.

Table 2. The hardware parameters of the camera.

Parameter Name	Parameter	Unit
Focal length	12.5–775	mm
Pixel	3	mp
Resolution	2048*1536	ppi
Wide dynamic range	120	db

The internal and external parameters of the two cameras are shown in Table 3.

Table 3. The calibration parameters.

Setting	Intrinsic Matrix	Rotation	Translation	Distortion
	Pixel	**m**	**m**	**m**
Left camera	$\begin{bmatrix} 534.31 & 0.00 & 342.64 \\ 0.00 & 534.31 & 234.42 \\ 0.00 & 0.00 & 1.00 \end{bmatrix}$	$\begin{bmatrix} 1.00 & 1.00 & 1.00 \\ 1.00 & 1.00 & 1.00 \\ 1.00 & 1.00 & 1.00 \end{bmatrix}$	$\begin{bmatrix} 1.00 \\ 1.00 \\ 1.00 \end{bmatrix}$	$\begin{bmatrix} -0.29 \\ 0.11 \\ 0.00 \end{bmatrix}, \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix}$
Right camera	$\begin{bmatrix} 537.39 & 0.00 & 326.62 \\ 0.00 & 537.01 & 250.58 \\ 0.00 & 0.00 & 1.00 \end{bmatrix}$	$\begin{bmatrix} 1.00 & 0.00 & 0.01 \\ -0.00 & 1.00 & 250.58 \\ -0.01 & 0.01 & 1.00 \end{bmatrix}$	$\begin{bmatrix} -99.72 \\ 1.27 \\ 0.05 \end{bmatrix}$	$\begin{bmatrix} -0.29 \\ 0.10 \\ 0.00 \end{bmatrix}, \begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix}$

The experiments were carried out on a computer platform configured with 64G memory, an Intel Core i9-12900kF CPU and a NVIDIA GeForce RTX 3090 Ti GPU for training and testing. The system of the experiment platform was Windows 10. Referring to previous studies [62–64], the training parameters of the model were set as follows: the initial learning rate was 0.001, the attenuation coefficient was 0.0005, and the stochastic gradient descent rate was 0.9.

4.3. Construction of the Training Database

In this paper, we used the image annotation tool labellmg to manually annotate the boxes of each ship in the images (https://github.com/heartexlabs/labelImg (accessed on 18 June 2023)). Labellmg is the most widely used image annotation tool for creating custom databases. Once the images are annotated, a dataset is generated that contains the category of the ship, the position of the corners of the ship's box, as well as the width and height of the ship.

To prevent overfitting and improve target detection accuracy, data augmentation strategies are applied to the images in the database, which increased sample diversity and improved the robustness of the model. In the experiment, several augmentation techniques such as horizontal flipping, vertical flipping, random rotation, Mosaic, and cutout were applied to enrich the training samples.

4.4. Detection

The proposed model was then tested with four scenarios to assess its performance. The details for the video of each scenario is shown in Table 4.

Table 4. The details for the videos.

No.	Frame Rate (fps)	Resolution	Duration(s)	Category	Ship Status
video#1	10	720 × 1280	150	Container ship	Underway
video#2	10	2160 × 3840	60	Bulk cargo carrier	Moored
video#3	10	2160 × 3840	480	Passenger ship	Moored
video#4	10	2160 × 3840	60	Passenger ship Fishing boat	Underway Anchored

Video 1 shows an underway container ship navigating a traffic route. The frame rate of the video is 10 per second (fps), the duration is 150 s, and the resolution is 720 × 1280.

Video 2 tests the performance of the model to detect a static target, therefore a moored bulk cargo ship was selected. The frame rate of the video is 10 per second (fps), the duration is 60 s, and the resolution is 2160 × 3840.

Video 3 is a berth with several moored ships. It tests whether the model could detect the passenger ship within a complex water environment. The frame rate of video is 10 per second (fps), the duration is 480 s, and the resolution is 2160 × 3840.

Video 4 shows a scenario of two ships. One is a passenger ship passing a bridge and another is an anchored fishing boat. It can be considered as a typical condition to test whether the model can detect ships when obstacles exist. The frame rate of the video is 10 per second (fps), the duration is 60 s, and the resolution is 2160 × 3840.

The video clips are shown in Figure 8.

Figure 8. The four scenarios used to test the model: video 1; video 2; video 3 and video 4.

It can be seen from Figure 8, that with use of the proposed model, the targets in the different scenarios can be detected; the associated ship type and confidence rates are given in Table 5.

**Table 5.** The confidence of detect results for videos.

No.	Minimum Confidence	Maximum Confidence	Average Confidence
Video 1	0.95	1.00	0.97
Video 2	0.99	1.00	0.99
Video 3	0.50	0.94	0.76
Video 4	0.53	0.84	0.72

In Table 5, the minimum confidence, maximum confidence, and average confidence for the target in Video 1 and Video 2 are (0.95, 1.00, 0.97) and (0.99, 1.00, 0.99), respectively, showing high accuracy and reliability of the model to detect ships underway and moored. Based on this, it can be concluded that the proposed system provides satisfactory detection performance and is capable of successfully detecting target objects in a typical offshore wind farm. However, in Video 3 and Video 4, although the target can be detected by the proposed model, the minimum confidence decreased to 0.50 and 0.53, and the average confidence decreased to 0.76 and 0.72. The main reason for this reduced confidence is the impacts from the complex environment conditions and surroundings, which can be overcome by (1) providing more training data to enhance the model detection probability, (2) using additional segmentation methods before detecting the ship, however, the time-consumption will become another problem.

4.5. Position and Tracking Results

Figures 9 and 10 show typical object position and tracking results (i.e., Video 1 and Video 2). The results shown in Figure 9 detect an underway container ship, with a confidence degree of 0.99. After synchronizing images from two cameras and matching them with binocular stereo vision model, the ship's position was obtained as 118.745 E and 24.4749 N. In addition, the ship's dynamic information of speed and course was calculated as well, where the speed was 11.61 knots and the course was 290.01 degrees.
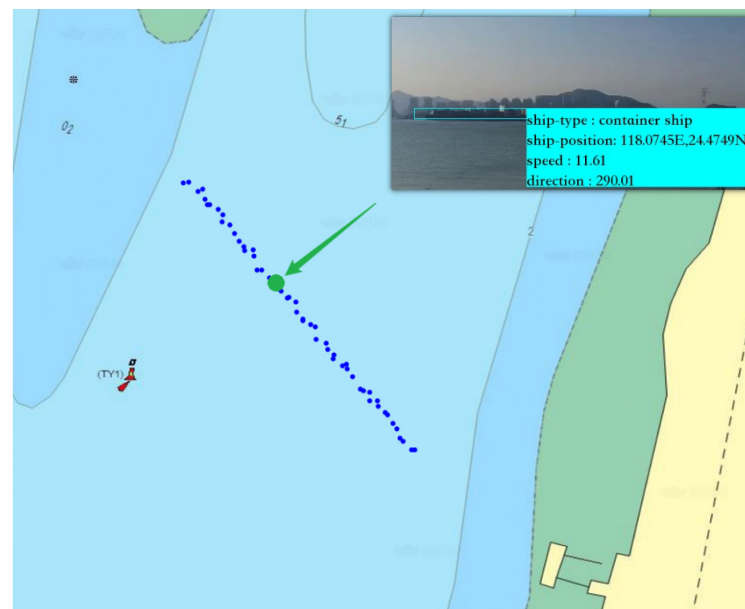
Figure 9. Position and tracking result for Video 1.
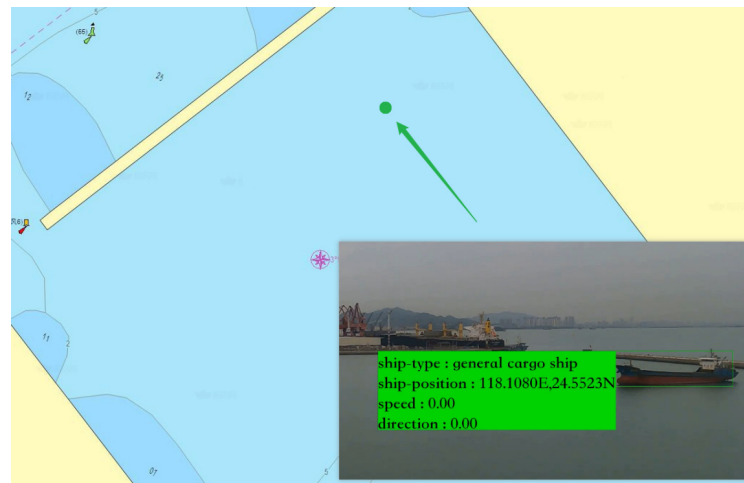
Figure 10. Position and tracking result for Video 2.

In Figure 10, an anchored bulk ship was detected; the model assigns the confidence degree of 0.99 and gives the position of the ship as longitude of 118.1074 E and latitude of 24.5521 N. As it is an anchored ship, the speed and the course for this ship is 0.

5. Validation

5.1. Detection Validation

To validate the proposed model, the two databases MYSHIP and SEASHIP were used. The detection capabilities were evaluated by the following indices: accuracy rate (P), recall rate (R), false alarm rate (F), miss alarm rate (M), and average precision (AP), which were defined as follows:

- ■ Precision (P): proportion of samples that are correctly detected in all test results.
- ■ Recall rate (R): proportion of actual positive samples that are correctly detected.
- ■ False alarm rate (F): proportion of negative samples that are incorrectly detected as positive samples.
- ■ Miss alarm (M): proportion of actual positive samples that are incorrectly detected as negative samples.
- ■ Average precision (AP): the integral value of the Precision Rate-Recall rate curve (P-R curve).
- ■ *Average precision$_{50}$* (AP_{50}): the average accuracy of the test when the IOU threshold is 0.5.
- ■ *Average precision$_{50:95}$* ($AP_{50:95}$): the average accuracy of the test when the IOU threshold is 0.5–0.95.

The relationship between these indices are mathematically described as function 3:

$$\begin{cases} P = \frac{TP}{TP+FP} \\ R = \frac{TP}{TP+FN} \\ F = \frac{FP}{TP+FP} = 1 - P \\ M = \frac{FN}{TP+FN} = 1 - R \\ AP = \int_0^1 P(R)dR \end{cases} \tag{3}$$

where *TP* is the counts of the true positive that correctly predicts positive samples, *TN* is the counts of the true negative that correctly predicts negative samples, *FP* is the counts of the false positive that incorrectly predicts negative samples and *FN* is the counts of false negative that incorrectly predicts positive samples.

5.2. Validating the Proposed Models

By using the validation index, the performance of the model is shown in Table 6.

Table 6. The performance of the YOLOv7 model on the database.

Database	P	R	F	M	AP_{50}	$AP_{50:95}$
MYSHIP	74.38%	78.20%	21.80%	25.62%	97.10%	72.61%
SEASHIP	62.49%	70.97%	29.03%	37.52%	96.40%	61.61%

In general, the system has high recognition accuracy and precision for ships. The rates of detection accuracy were relatively high (74.38% for the MYSHIP and 62.46% for SEASHIP datasets). However, by comparing the evaluation results, Table 6 shows the YOLOv7 model demonstrated better performance on the latter database (i.e., SEASHIP database). This is due to the MYSHIP database including more distant and overlapping ships, making identification more challenging.

Then, to validate the accuracy of the model's predictions, the ship's real-time AIS data in the four cases (Video 1–4) was collected. The detected results from the proposed model were compared with information provided in the AIS data, which is considered more accurate and reliable. Table 7 displays the comparison of different detection systems for ships.

**Table 7.** The results comparing different detections systems for ships.

	Video #1	Video #2	Video #3	Video #4
System result	Container ship	General cargo ship	Passenger ship $\times$ 5 General cargo ship $\times$ 2	Fishing boat Passenger ship
AIS data	Container ship	General cargo ship	Passenger ship $\times$ 2	Passenger ship

Table 7 shows the proposed system can detect ship types with high accuracy, while showing advantages of ship detection that are not provided by AIS data. For instance, in Video 3, only two passenger ships have AIS data but in reality, there were five passenger ships and two general cargo ships but some of them had turned off their AIS or were missing signals. Similarly, Video 4 shows a fishing boat that did not provide its AIS data, but it was detected by the video system.

The accuracy of the positioning results was compared as well. As the AIS transmits ship position with different rates (e.g., 2–10 s when ships are underway and 180 s when ships are moored or anchored), this study linked all the AIS positions with time sequences to obtain the ship trajectories. In this study, the update frequency of the proposed model was set as 2 s each time and all positions were linked following the time sequences to obtain the trajectories. Then, the deviation between the AIS based trajectories and the model detected trajectories in Video 1 were calculated and the results are given in Figure 11 and the deviations are given in Table 8.

Table 8. The comparison between the AIS data and the system output result.

		Sample#1	Sample#2	Sample#3	Sample#4
System output	Longitude	118.0796 E	118.1074 E	118.1081 E	118.1117 E
	Latitude	24.4806 N	24.5521 N	24.5466 N	24.5579 N
AIS data (°)	Longitude	118.07962 E	118.10737 E	118.10814 E	118.11169 E
	Latitude	24.48064 N	24.55212 N	24.54661 N	24.55791 N
Deviation (°)	Longitude	0.00002	−0.00003	0.00004	−0.00001
	Latitude	0.00004	0.00002	0.00001	0.00001
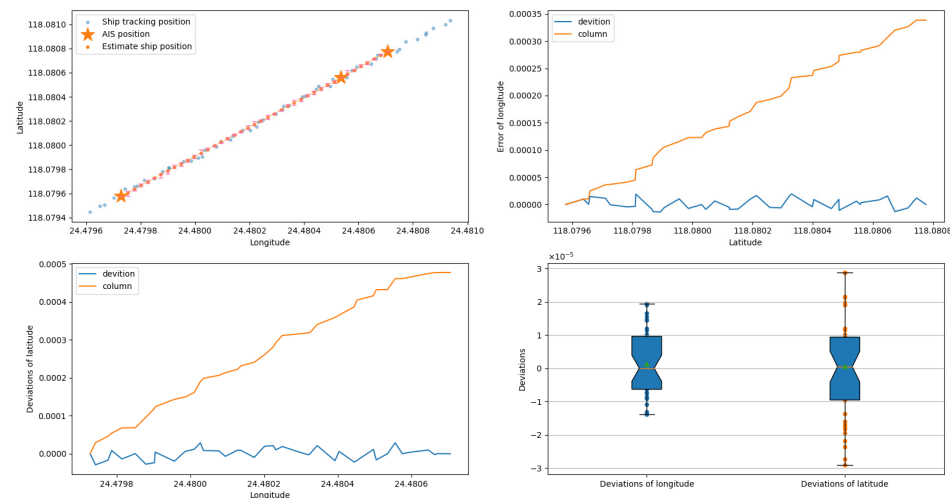
Figure 11. The comparison between the AIS based trajectories and model detected trajectories.

It can be seen in Figure 11 that due to the low update rate, only three AIS records were received for the ship in Video 1. Therefore, three AIS positions were linked and the deviations on the longitude and latitude were calculated (see Table 8). Notably, although the deviation between the two trajectories existed, they were all less than ±0.0001 degree (11 m approximately), which is broadly acceptable.

6. Discussion

6.1. Comparing the Current System Used for Ship Detection and Tracking

Based on the case study results, the advantages of use of visual based technologies for ship detection are as follows:

(1) The proposed approach provides a way for ship monitoring with satisfactory accuracy and reliability. The developed system can obtain necessary information about ship traffic (e.g., speed, position, course), so that it becomes a novel supplement for VTS and OWF managers. In addition, in contrast to AIS and radar, this system uses cameras to collect ship videos, and the CCTVs are already fixed in most of the OWFs worldwide. Therefore, the cost of developing this system in practice can be very low.

(2) As already mentioned, the proposed system has a high frequency of updating the ship information (i.e., 10 fps) and high accuracy in individual ship tracking (the tracking errors less than 15 m/0.0001 degrees). Thus, the developed system reasonably provides ship dynamic data and ensures collision risk assessment for ships in the vicinity of OWFs.

(3) The proposed framework hybridizes several visual technologies. The study not only proves the possibility of using these technologies to aid ship monitoring in offshore wind farm waters but also can be considered as evidence to apply the machine visual model for ship detection and tracking in other similar waters, such as narrow channels, bridges, on a river, etc.

Moreover, the advantages and disadvantages of current ship detection approaches were compared, and the results are listed in Table 9.

Table 9 presents several highlights between the current ship detection approaches. The radar tracking system can accurately measure and track moving or stationary targets, but it has a blind spot, and the echoes are susceptible to environmental factors, which can lead to target loss. The AIS-based ship reporting system can compensate for these shortcomings of radar, but AIS cannot identify targets that are not equipped with AIS or have the AIS turned off, so that leads to missed tracking. Both the CNN-based and the YOLO-based model show high accuracy for target detection. However, The YOLO algorithm is time consuming due to its faster detection speed and frequency.

Table 9. The comparison between the existing system and our system.

Method Name	Advantage	Disadvantage
AIS based ship reporting system	1. Capable of automatically obtaining target position, status, speed, and other information 2. High accuracy of ship dynamic data 3. Strong performance against rain, snow, and wave interference 4. Eliminate the blocking area of some radars	1. Operates only at VHF 2. Cannot identify targets without AIS devices installed 3. The data are not continuous and poor timeliness
Radar tracking system	1. Capable of automatic tracking and automatic warning 2. Accurate measurement and display of distance, bearing, speed and size of the target	1. Blind spots exist 2. The echo is easily affected by weather, sea state and terrain obstruction, etc. 3. In the multi-target dense area is prone to false tracking and target loss phenomenon
YOLO model	1. Has a high recognition speed, and can efficiently identify multiple targets 2. Superiority in real-time target detection 3. Suitable for detecting objects of various shapes and sizes	1. Global information has a good performance, but poor performance on a small range of information 2. Compared to some recognition algorithms, the accuracy rate is slightly worse
RCNN model	1. Powerful adaptability for recognition and classification of complex data 2. Feature classification works well	1. Large number of operations, large sample size and time are required to train the model 2. Lack of annotation resources for specific target samples, high quality for the database

6.2. YOLO Series

In this study, the YOLOv7 model was trained with AIS data and incorporated binocular vision to improve the accuracy of system detection. The proposed system can complete ship detection and target tracking, which significantly increases the model's accuracy and applicability. However, the YOLO series are developing rapidly. To evaluate the performance of detecting targets in maritime domain by using the YOLO series, this study selected the YOLOv5 model as the benchmark. In the comparison, both methods use the same database and hardware platforms. The results of the comparison experiment is shown in Table 10. The original YOLOv5 model achieved 73.80% of detection precision, while the YOLOv7 model achieved 77.48%, which is 3.68% higher than the YOLOv5 model. The recall rates for the YOLOv5 and YOLOv7 models were not very different (73.41% and 73.19%). The false alarm rate of the YOLOv5 model was 3.68%, which was higher than for the YOLOv7 model. For the index of miss alarm rate, the YOLOv5 and YOLOv7 models had similar performance, with 26.6% and 26.81%, respectively. Then the study set the IOU threshold at 0.5 to test the index of average accuracy. The YOLOv5 model obtained an average accuracy of 95.76%, while the YOLOv7 model reached 97.03%. Similarly, setting the IOU threshold in the interval of 0.5–0.95, the results showed the average accuracy for the YOLOv5 model dropped to 70.37% and for the YOLOv7 model it was 71.51%, which is 1.14% higher than the YOLOv5 model. Consequently, based on these comparisons, it can be concluded that the proposed method can be effective for ship detection and real-time traffic monitoring.

Table 10. The results of the comparison experiment between the YOLOv5 and YOLOv7 models.

Methods	P	R	F	M	AP_{50}	$AP_{50:95}$
YOLOv5	73.80%	73.41%	26.20%	26.60%	95.76%	70.37%
YOLOv7	77.48%	73.19%	22.52%	26.81%	97.03%	71.51%

However, the system proposed above also needs to be improved. First, there are still some ships that may not be well detected in the proposed system (current accuracy is 77.48%) due to their position and orientation. When the ship drives independently and laterally, it is easily detected by the system. Moreover, when the ship is longitudinal, and multiple ships overlap, the system may regard these ships as obstacles and ignore them. For example, the ship in Video 3 has the lowest confidence. This is because, among the

vessels detected by the system, those with a low confidence level are characterized by facing the camera end on. This orientation makes the ship's characteristics less obvious, so it is easily ignored by the system. Second, the database needs to be expanded to include more training data. Third, the potential effects of critical weather conditions (e.g., snow, heavy fog, and rain) were not tested in this study. Therefore, further studies can be carried out providing more labelled samples, and we will collect more data to enrich the system database and train better classifiers in following research. Meanwhile, the system will be tested under critical environments in practice soon.

7. Conclusions

This paper proposed a ship detection and positioning system based on the YOLOv7 algorithm and stereo vision technologies and introduces the framework and detailed methods used in this system. In addition, this study suggests a novel concept of using AIS data as the training resource for model training, which improves the accuracy of using the YOLOv7 model and stereo vision algorithms in ship detection and tracking. Applying the proposed model in a real ship case study validates the possibility of using the YOLOv7 algorithm to track and identify ships when the stereo vision algorithm is applied to locate ship positions.

The benefit of the proposed system is that it can detect vessels automatically and achieve real-time tracking and positioning. The system not only eases the workload of OWF operators during CCTV monitoring but also provides a possible way for ship traffic management in the water in the vicinity of OWFs. Moreover, the novel system shares the idea of using machine vision technology for ship collision prevention. In addition, the proposed method applies not only to offshore wind farm waters but can be applied in the future to any place of interest. Based on the analysis of the proposed system, further study can investigate applying the proposed system in ships to achieve situation forecasting.

Author Contributions: Conceptualization, validation, writing, original draft, methodology and formal analysis—H.J., X.L. and Y.H.; writing, review and editing, supervision, funding acquisition—Q.Y. and M.Z. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

Note

[1] 1000 images of inland ships were captured at Bay Park, Xiamen Bridge, and Gao Qi Wharf in Xiamen City, Fujian Province.

References

1. Vieira, M.; Henriques, E.; Snyder, B.; Reis, L. Insights on the impact of structural health monitoring systems on the operation and maintenance of offshore wind support structures. *Struct. Saf.* **2022**, *94*, 102154. [CrossRef]
2. Musial, W.; Spitsen, P.; Duffy, P.; Beiter, P.; Marquis, M.; Hammond, R.; Shields, M. *Offshore Wind Market Report: 2022 Edition*; National Renewable Energy Lab. (NREL): Golden, CO, USA, 2022.
3. Torres-Rincón, S.; Bastidas-Arteaga, E.; Sánchez-Silva, M. A flexibility-based approach for the design and management of floating offshore wind farms. *Renew. Energy* **2021**, *175*, 910–925. [CrossRef]
4. Brady, R.L. *Offshore Wind Industry Interorganizational Collaboration Strategies in Emergency Management*; Walden University: Minneapolis, MN, USA, 2022.
5. Li, B.; Xie, X.; Wei, X.; Tang, W. Ship detection and classification from optical remote sensing images: A survey. *Chin. J. Aeronaut.* **2021**, *34*, 145–163. [CrossRef]

6. Zhao, J.; Guo, W.; Zhang, Z.; Yu, W. A coupled convolutional neural network for small and densely clustered ship detection in SAR images. *Sci. China Inf. Sci.* **2018**, *62*, 42301. [CrossRef]
7. Zhang, S.; Wu, R.; Xu, K.; Wang, J.; Sun, W. R-CNN-Based Ship Detection from High Resolution Remote Sensing Imagery. *Remote Sens.* **2019**, *11*, 631. [CrossRef]
8. Hu, J.; Zhi, X.; Shi, T.; Yu, L.; Zhang, W. Ship Detection via Dilated Rate Search and Attention-Guided Feature Representation. *Remote Sens.* **2021**, *13*, 4840. [CrossRef]
9. Yu, C.; Shin, Y. SAR ship detection based on improved YOLOv5 and BiFPN. *ICT Express* **2023**. [CrossRef]
10. Priakanth, P.; Thangamani, M.; Ganthimathi, M. *WITHDRAWN: Design and Development of IOT Based Maritime Monitoring Scheme for Fishermen in India*; Elsevier: Amsterdam, The Netherlands, 2021.
11. Ouelmokhtar, H.; Benmoussa, Y.; Benazzouz, D.; Ait-Chikh, M.A.; Lemarchand, L. Energy-based USV maritime monitoring using multi-objective evolutionary algorithms. *Ocean Eng.* **2022**, *253*, 111182. [CrossRef]
12. Nyman, E. Techno-optimism and ocean governance: New trends in maritime monitoring. *Mar. Policy* **2019**, *99*, 30–33. [CrossRef]
13. Pawar, K.; Attar, V. Deep learning based detection and localization of road accidents from traffic surveillance videos. *ICT Express* **2022**, *8*, 379–387. [CrossRef]
14. Thallinger, G.; Krebs, F.; Kolla, E.; Vertal, P.; Kasanický, G.; Neuschmied, H.; Ambrosch, K.-E. Near-miss accidents–classification and automatic detection. In Proceedings of the Intelligent Transport Systems—From Research and Development to the Market Uptake: First International Conference, INTSYS 2017, Hyvinkää, Finland, 29–30 November 2017; Proceedings 1; pp. 144–152.
15. Pramanik, A.; Sarkar, S.; Maiti, J. A real-time video surveillance system for traffic pre-events detection. *Accid. Anal. Prev.* **2021**, *154*, 106019. [CrossRef] [PubMed]
16. Root, B. HF radar ship detection through clutter cancellation. In Proceedings of the 1998 IEEE Radar Conference, RADARCON'98, Challenges in Radar Systems and Solutions, Dallas, TX, USA, 14 May 1998; pp. 281–286.
17. Dzvonkovskaya, A.; Gurgel, K.-W.; Rohling, H.; Schlick, T. Low power high frequency surface wave radar application for ship detection and tracking. In Proceedings of the 2008 International Conference on Radar, Adelaide, SA, Australia, 2–5 September 2008; pp. 627–632.
18. Margarit, G.; Barba Milanés, J.; Tabasco, A. Operational Ship Monitoring System Based on Synthetic Aperture Radar Processing. *Remote Sens.* **2009**, *1*, 375–392. [CrossRef]
19. Brekke, C. *Automatic Ship Detection Based on Satellite SAR*; FFI-rapport: Kjeller, Norway, 2008.
20. Kazimierski, W.; Stateczny, A. Fusion of Data from AIS and Tracking Radar for the Needs of ECDIS. In Proceedings of the 2013 Signal Processing Symposium (SPS), Serock, Poland, 5–7 June 2013; pp. 1–6.
21. Pan, Z.; Deng, S. Vessel Real-Time Monitoring System Based on AIS Temporal Database. In Proceedings of the 2009 International Conference on Information Management, Innovation Management and Industrial Engineering, Kuala Lumpur, Malaysia, 3–5 April 2009; pp. 611–614.
22. Lin, B.; Huang, C.-H. Comparison between Arpa Radar and Ais Characteristics for Vessel Traffic Services. *J. Mar. Sci. Technol.* **2006**, *14*, 7. [CrossRef]
23. Yu, Q.; Liu, K.; Chang, C.-H.; Yang, Z. Realising advanced risk assessment of vessel traffic flows near offshore wind farms. *Reliab. Eng. Syst. Saf.* **2020**, *203*, 107086. [CrossRef]
24. Chang, C.-H.; Kontovas, C.; Yu, Q.; Yang, Z. Risk assessment of the operations of maritime autonomous surface ships. *Reliab. Eng. Syst. Saf.* **2021**, *207*, 107324. [CrossRef]
25. Yu, Q.; Liu, K.; Teixeira, A.; Soares, C.G. Assessment of the influence of offshore wind farms on ship traffic flow based on AIS data. *J. Navig.* **2020**, *73*, 131–148. [CrossRef]
26. Liu, Y.; Su, H.; Zeng, C.; Li, X. A Robust Thermal Infrared Vehicle and Pedestrian Detection Method in Complex Scenes. *Sensor* **2021**, *21*, 1240. [CrossRef] [PubMed]
27. Shao, Z.; Wang, L.; Wang, Z.; Du, W.; Wu, W. Saliency-aware convolution neural network for ship detection in surveillance video. *IEEE Trans. Circuits Syst. Video Technol.* **2019**, *30*, 781–794. [CrossRef]
28. Chen, Z.; Li, B.; Tian, L.F.; Chao, D. Automatic detection and tracking of ship based on mean shift in corrected video sequences. In Proceedings of the 2017 2nd International Conference on Image, Vision and Computing (ICIVC), Chengdu, China, 2–4 June 2017; pp. 449–453.
29. Rahmatov, N.; Paul, A.; Saeed, F.; Hong, W.-H.; Seo, H.; Kim, J. Machine learning–based automated image processing for quality management in industrial Internet of Things. *Int. J. Distrib. Sens. Netw.* **2019**, *15*, 1550147719883551. [CrossRef]
30. Chávez Heras, D.; Blanke, T. On machine vision and photographic imagination. *AI Soc.* **2020**, *36*, 1153–1165. [CrossRef]
31. Cho, S.; Kwon, J. Abnormal event detection by variation matching. *Mach. Vis. Appl.* **2021**, *32*, 80. [CrossRef]
32. Lowe, D.G. Object recognition from local scale-invariant features. In Proceedings of the Seventh IEEE International Conference on Computer Vision, Kerkyra, Greece, 20–27 September 1999; pp. 1150–1157.
33. Orr, G.B.; Müller, K.-R. *Neural Networks: Tricks of the Trade*; Springer: Berlin/Heidelberg, Germany, 1998.
34. Qi, S.; Ma, J.; Lin, J.; Li, Y.; Tian, J. Unsupervised ship detection based on saliency and S-HOG descriptor from optical satellite images. *IEEE Geosci. Remote Sens. Lett.* **2015**, *12*, 1451–1455.
35. Zhu, C.; He, Y.; Savvides, M. Feature selective anchor-free module for single-shot object detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 840–849.

36. Law, H.; Deng, J. CornerNet: Detecting objects as paired keypoints. In Proceedings of the European conference on computer vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 734–750.
37. Bochkovskiy, A.; Wang, C.-Y.; Liao, H.-Y.M. Yolov4: Optimal speed and accuracy of object detection. *arXiv* **2020**, arXiv:2004.10934.
38. Chen, Z.; Liu, C.; Filaretov, V.F.; Yukhimets, D.A. Multi-Scale Ship Detection Algorithm Based on YOLOv7 for Complex Scene SAR Images. *Remote Sens.* **2023**, *15*, 2071. [CrossRef]
39. Cen, J.; Feng, H.; Liu, X.; Hu, Y.; Li, H.; Li, H.; Huang, W. An Improved Ship Classification Method Based on YOLOv7 Model with Attention Mechanism. *Wirel. Commun. Mob. Comput.* **2023**, *2023*, 7196323. [CrossRef]
40. Yasir, M.; Zhan, L.; Liu, S.; Wan, J.; Hossain, M.S.; Isiacik Colak, A.T.; Liu, M.; Islam, Q.U.; Raza Mehdi, S.; Yang, Q. Instance segmentation ship detection based on improved Yolov7 using complex background SAR images. *Front. Mar. Sci.* **2023**, *10*. [CrossRef]
41. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.-Y.; Berg, A.C. Ssd: Single shot multibox detector. In Proceedings of the Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, 11–14 October 2016; Proceedings, Part I 14. pp. 21–37.
42. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR, USA, 23–28 June 2013; pp. 580–587.
43. Meng, C.; Bao, H.; Ma, Y.; Xu, X.; Li, Y. Visual Meterstick: Preceding Vehicle Ranging Using Monocular Vision Based on the Fitting Method. *Symmetry* **2019**, *11*, 1081. [CrossRef]
44. Zhao, Y.; Zhao, L.; Xiong, B.; Kuang, G. Attention receptive pyramid network for ship detection in SAR images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 2738–2756. [CrossRef]
45. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 779–788.
46. Gong, H.; Li, H.; Xu, K.; Zhang, Y. Object detection based on improved YOLOv3-tiny. In Proceedings of the 2019 Chinese automation congress (CAC), Hangzhou, China, 22–24 November 2019; pp. 3240–3245.
47. Patel, K.; Bhatt, C.; Mazzeo, P.L. Improved Ship Detection Algorithm from Satellite Images Using YOLOv7 and Graph Neural Network. *Algorithms* **2022**, *15*, 473. [CrossRef]
48. Li, J.; Qu, C.; Shao, J. Ship detection in SAR images based on an improved faster R-CNN. In Proceedings of the 2017 SAR in Big Data Era: Models, Methods and Applications (BIGSARDATA), Beijing, China, 13–14 November 2017; pp. 1–6.
49. Li, Y.; Rong, L.; Li, R.; Xu, Y. Fire Object Detection Algorithm Based on Improved YOLOv3-tiny. In Proceedings of the 2022 7th International Conference on Cloud Computing and Big Data Analytics (ICCCBDA), Chengdu, China, 22–24 April 2022; pp. 264–269.
50. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster r-cnn: Towards real-time object detection with region proposal networks. *Adv. Neural Inf. Process. Syst.* **2015**, *28*, 1137–1149. [CrossRef]
51. Campbell, J.; Sukthankar, R.; Nourbakhsh, I.; Pahwa, A. A robust visual odometry and precipice detection system using consumer-grade monocular vision. In Proceedings of the 2005 IEEE International Conference on Robotics and Automation, Barcelona, Spain, 18–22 April 2005; pp. 3421–3427.
52. Matthies, L.; Shafer, S. Error modeling in stereo navigation. *IEEE J. Robot. Autom.* **1987**, *3*, 239–248. [CrossRef]
53. Davison, A.J. Real-time simultaneous localisation and mapping with a single camera. In Proceedings of the Computer Vision, IEEE International Conference, Nice, France, 13–16 October 2003; p. 1403.
54. Schwartzkroin, P.A. Neural Mechanisms: Synaptic Plasticity. Molecular, Cellular, and Functional Aspects. Michel Baudry, Richard F. Thompson, and Joel L. Davis, Eds. MIT Press, Cambridge, MA, 1993. xiv, 263 pp., illus. $50 or£ 44.95. *Science* **1994**, *264*, 1179–1180. [CrossRef] [PubMed]
55. Howard, I.P.; Rogers, B.J. *Binocular Vision and Stereopsis*; Oxford University Press: Oxford, UK, 1995.
56. Xu, Y.; Zhao, Y.; Wu, F.; Yang, K. Error analysis of calibration parameters estimation for binocular stereo vision system. In Proceedings of the 2013 IEEE International Conference on Imaging Systems and Techniques (IST), Beijing, China, 22–23 October 2013; pp. 317–320.
57. Yu, Y.; Tingting, W.; Long, C.; Weiwei, Z. Stereo vision based obstacle avoidance strategy for quadcopter UAV. In Proceedings of the 2018 Chinese Control And Decision Conference (CCDC), Shenyang, China, 9–11 June 2018; pp. 490–494.
58. Sun, X.; Jiang, Y.; Ji, Y.; Fu, W.; Yan, S.; Chen, Q.; Yu, B.; Gan, X. Distance measurement system based on binocular stereo vision. In Proceedings of the IOP Conference Series: Earth and Environmental Science, Hainan, China, 25–26 April 2009; p. 052051.
59. Wang, C.; Zou, X.; Tang, Y.; Luo, L.; Feng, W. Localisation of litchi in an unstructured environment using binocular stereo vision. *Biosyst. Eng.* **2016**, *145*, 39–51. [CrossRef]
60. Zou, X.; Zou, H.; Lu, J. Virtual manipulator-based binocular stereo vision positioning system and errors modelling. *Mach. Vis. Appl.* **2012**, *23*, 43–63. [CrossRef]
61. Zuo, X.; Xie, X.; Liu, Y.; Huang, G. Robust visual SLAM with point and line features. In Proceedings of the 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Vancouver, BC, Canada, 24–28 September 2017; pp. 1775–1782.
62. Hong, Z.; Yang, T.; Tong, X.; Zhang, Y.; Jiang, S.; Zhou, R.; Han, Y.; Wang, J.; Yang, S.; Liu, S. Multi-Scale Ship Detection From SAR and Optical Imagery Via A More Accurate YOLOv3. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 6083–6101. [CrossRef]

63. Cao, C.; Wu, J.; Zeng, X.; Feng, Z.; Wang, T.; Yan, X.; Wu, Z.; Wu, Q.; Huang, Z. Research on Airplane and Ship Detection of Aerial Remote Sensing Images Based on Convolutional Neural Network. *Sensor* **2020**, *20*, 4696. [CrossRef]

64. Yang, R.; Wang, R.; Deng, Y.; Jia, X.; Zhang, H. Rethinking the Random Cropping Data Augmentation Method Used in the Training of CNN-Based SAR Image Ship Detector. *Remote Sens.* **2020**, *13*, 34. [CrossRef]