

Research Article

Extracting Vessel Speed Based on Machine Learning and Drone Images during Ship Traffic Flow Prediction

Jiansen Zhao ¹, Yanjun Chen ¹, Zhenzhen Zhou ¹, Jingying Zhao ¹,
Shengzheng Wang ¹ and Xinqiang Chen ²

¹College of Merchant Marine, Shanghai Maritime University, Shanghai 201306, China

²Institute of Logistics Science and Engineering, Shanghai Maritime University, Shanghai 201306, China

Correspondence should be addressed to Xinqiang Chen; chenxinqiang@stu.shmtu.edu.cn

Received 26 June 2022; Revised 6 August 2022; Accepted 13 September 2022; Published 4 October 2022

Academic Editor: Wen Liu

Copyright © 2022 Jiansen Zhao et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In the water transportation, ship speed estimation has become a key subject of intelligent shipping research. Traditionally, Automatic Identification System (AIS) is used to extract the ship speed information. However, transportation environment is gradually becoming complex, especially in the busy water, leading to the loss of some AIS data and resulting in a variety of maritime accidents. To make up for this deficiency, this paper proposes a vessel speed extraction framework, based on Unmanned Aerial Vehicle (UAV) airborne video. Firstly, YOLO v4 is employed to detect the ship targets in UAV image precisely. Secondly, a simple online and real time tracking method with a Deep association metric (Deep SORT) is applied to track ship targets with high quality. Finally, the ship motion pixel is computed based on the bounding box information of the ship trajectories, at the same time, the ship speed is estimated according to the mapping relationship between image space and the real space. Exhaustive experiments are conducted on the various scenarios. Results verify that the proposed framework has an excellent performance with average speed measurement accuracy is above 93% in complex waters. This paper also paves a way to further predict ship traffic flow in water transportation.

1. Introduction

In recent years, the development of economy has made the shipping industry develop rapidly, while the shipping system [1, 2] has gradually become intelligent with the help of the computer technology. To further ensure the safety of ship navigation, ship speed estimation has become an important topic in the study of intelligent shipping, and the popularity of ship automatic identification System (AIS) [3] makes it one of the main channels to obtain ship speed information. However, the rapid development of shipping makes the maritime navigation environment more complex [4], and the AIS signals are prone to be interfered, resulting in data loss. Therefore, it is of great significance to propose an auxiliary method for real-time and accurate estimation of ship speed.

At present, the ship speed estimation technologies besides AIS also include radar technology [5], laser technology [6], sonar technology [7, 8], video-based technology [9],

and so on. Among them, infrared technology is susceptible to weather. Radar and laser meters can avoid these problems, but they are usually more expensive and maintain frequently. Sonar technology takes advantages of the low-cost reflection characteristics for velocity measurement, but the measurement time is quite long and the effective distance is small. In contrast, video-based technology does not require complex and expensive hardware. It is not only easy to install and maintain, but also has a larger monitoring range. With the development of computer vision technology in recent years, the method is becoming increasingly popular in intelligent monitoring systems.

Most video-based speed estimation systems are based on fixed cameras [10]. Reference [11] proposed a vehicle speed estimation method for extracting traffic speed information. The authors used a detection and tracking algorithm in deep learning technology to capture the vehicle position in the image, and then calculated the relationship between image

space and real world to forecast the ship speed. In [12], a novel two-camera-based vehicle speed framework was presented. They estimated the speed of a vehicle by calculating the specific geometry relationship in two camera spaces. This fixed camera-based method had a limitation in monitoring range and might not be cost effective in a busy road segment.

Furthermore, based on the UAV video, reference [13] proposed the average speed measurement method, that is, the optical flow and clustering algorithm, which were used to cluster the optical flow vectors to calculate the average traffic flow speed. In reference [14], a speed estimation method was designed by searching the correspondence between scale Invariant Feature Transform (SIFT) features in video frames, and the above methods have achieved relatively excellent results in the process of road traffic speed measurement. In addition, some methods [15, 16] for measuring ship speed were developed based on the ship motion characteristics in synthetic aperture radar (SAR) images. Reference [17] proposed a method to estimate ship speed based on the relationship between the wavelengths that constitute the Kelvin mode and ship speed. It can estimate the ship speed accurately when the main features and the turbulent wakes of a ship are clearly displayed in SAR images. However, the applicability of this method may be limited in the case of ship displacement.

In the above studies, the UAV remote sensing videos generally have the advantages of high pixel and large monitoring range [18], and are suitable for estimating ship speed in the field of waterway traffic. However, due to the slow speed of the ship, the motion trends are fuzzy and the navigation trajectories are complex [19, 20]. There are few studies on ship speed estimation based on UAV remote sensing images, and the real-time and continuity need to be improved.

To make up for the shortcomings of the existing ship speed measurement technology, a multi-ship speed measurement method based on the UAV videos is proposed after carefully learning previous researches. In our framework, a self-built UAV dataset containing 18,655 images is built, the detection and tracking algorithms based on deep learning technology are used to obtain a series of positions in the images, the transformation relationship between image space and real space is calculated to provide the basis for ship velocity measurement.

2. Methodology

2.1. The Framework. The framework contains three parts: the first part is ship detection, the second part is ship tracking, and the third part contains camera calibration, ship trajectory extraction, and ship speed estimation, as shown in Figure 1.

In the first part, the images of ships are captured by using the UAV camera in a variety of environments. Then, the image annotation tool called *labelImg* is used to annotate the ground truth box of ship, and thus a dataset is formed. Considering there are many small ships in the image, YOLO v4 algorithm is employed for training model. When the loss function drops rapidly and stably, the weight and bias are saved, and the ship detection model is obtained. Finally,

the detection model is used to detect different ships and obtain a series of ship targets.

In the second part, according to the bounding box information in the previous frame, firstly, the Kalman filter algorithm is used to predict and update the tracks. Then, the Intersection-over-Union (IoU) and matching cascade are used to associate the tracks with the bounding boxes in the current frame. Finally, ship tracking is achieved and many stable ship positions in the image are obtained.

In the third part, based on the track information obtained in the part two, the ship trajectories are extracted by calculating the pixel difference between the tracks in the previous and current frame. Then, the camera inner parameters and the mapping relationship matrix are obtained by using camera calibration. Finally, combining the mapping relationship and an average speed formula, the real distance and the ship speed can be estimated.

2.2. Ship Detection. To detect the ship targets [21, 22] in the drone video stably, we focus on the deep learning detection algorithm. Because of the high position of UAV, the drone video has a wide field of view without any target occlusion. However, as many small ships sail in the waters, the background in the image is complex, and the motion blurring will be caused by the vibration of the UAV [23]. Therefore, choosing a detection algorithm with high accuracy is critical. Currently, several detection algorithms that have been proposed and mainly divided into two categories, two-stage algorithms and one-stage algorithms. In two-stage, the representative algorithms include SPPNet [24] and faster RCNN [25], while in one-stage, YOLO [26] and SSD [27] are usually used. The advantage of the two-stage algorithm is excellent accuracy, but for many devices with limited storage and computing capacity, the computing cost is high. In contrast, the one-stage algorithm achieves the current optimum in terms of detection speed and accuracy. Being one-stage, YOLO v4 algorithm [26] has obvious advantages for small targets in accuracy and efficiency. Therefore, YOLO v4 is selected for ship detection in this paper, which consists of 5 parts: input, backbone, neck, prediction, and output, as shown in Figure 2.

To achieve a better balance in terms of the input image resolution, the convolutional layer number, and the hyperparameter number, the CSPDarknet53 is employed as the backbone of YOLO v4. The CSPDarknet53 network consists of 53 convolutional layers and 23 residual blocks and is formed by applying the Cross-Stage-Partial (CSP) strategy to Darknet53. After the image passes to the CSPDarknet53, a series of feature maps at different levels are obtained. The neck can be composed of the spatial pyramid pooling (SPP) module and path aggregation network (PANet). The SPP module extracts the spatial feature information of different levels by increasing the receptive field and separates the most important contextual features. The PANet introduced a bottom-up path to make low-layer information easier to spread to the top layers. In the prediction stage, the multiclass classification and bounding box regression are applied to multiple scales. To measure the difference between the prediction and the actual and to optimize the

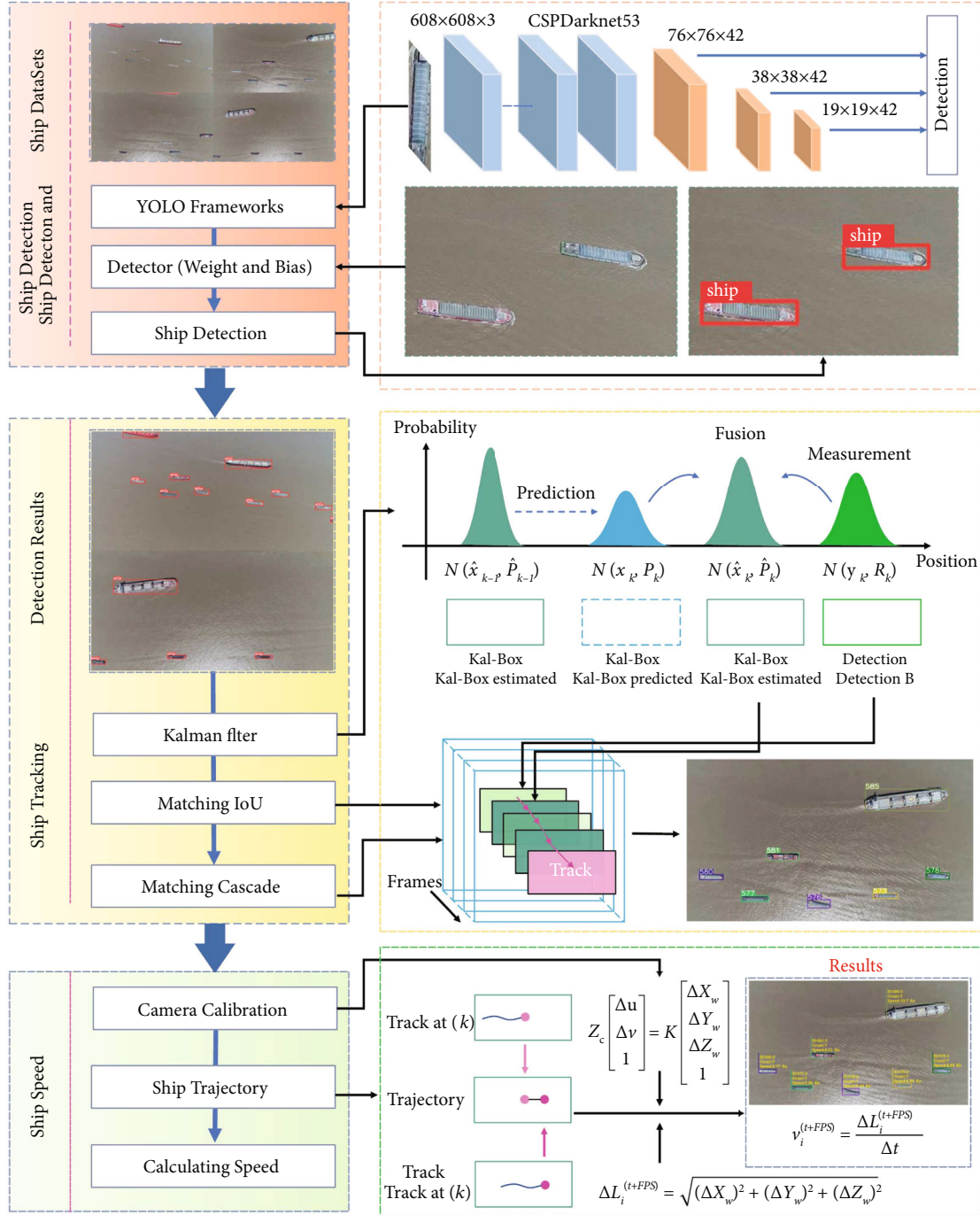


FIGURE 1: Research framework.

weights, the value of the loss function needs to be calculated. The loss function in YOLO v4 concludes bounding box regression loss, classification loss, and confidence loss, whose expressions are as follows:

$$Loss = L_{loc} + L_{cls} + L_{conf}. \quad (1)$$

In bounding box regression, taking the overlapping area, center-of-mass distance, and aspect ratio simultaneously into count, the mean square error (MSE) method is removed

and replaced by the Complete-Intersection-over-Union (CIoU) algorithm. Thus, better convergence speed and accuracy can be obtained in the training model. Finally, the fast non-maximum suppression (Fast NMS) method is used to retain the local maximum value of the confidence score to obtain the final result.

2.3. Ship Tracking. Based on the bounding box predicted in the detect stage, Deep SORT [28] tracking algorithm is considered to track the ship targets in the UAV images. Deep SORT is originated from SORT, which uses Kalman filter and Hungarian

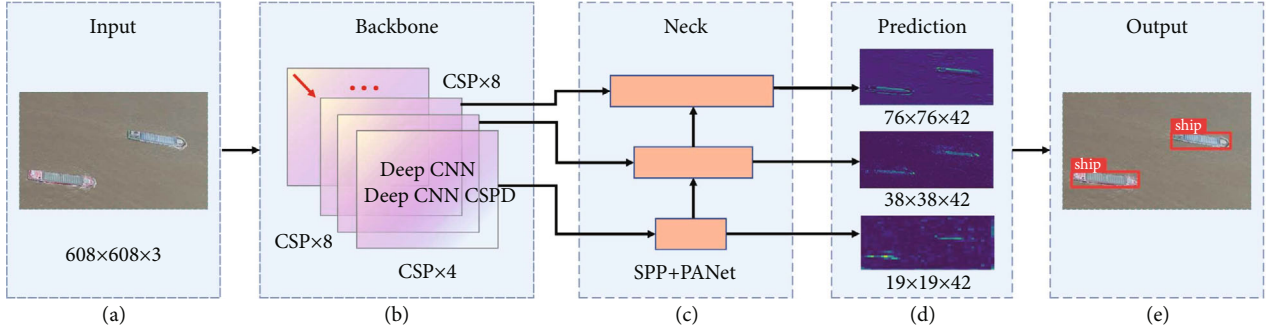


FIGURE 2: Schematic diagram of ship detection.

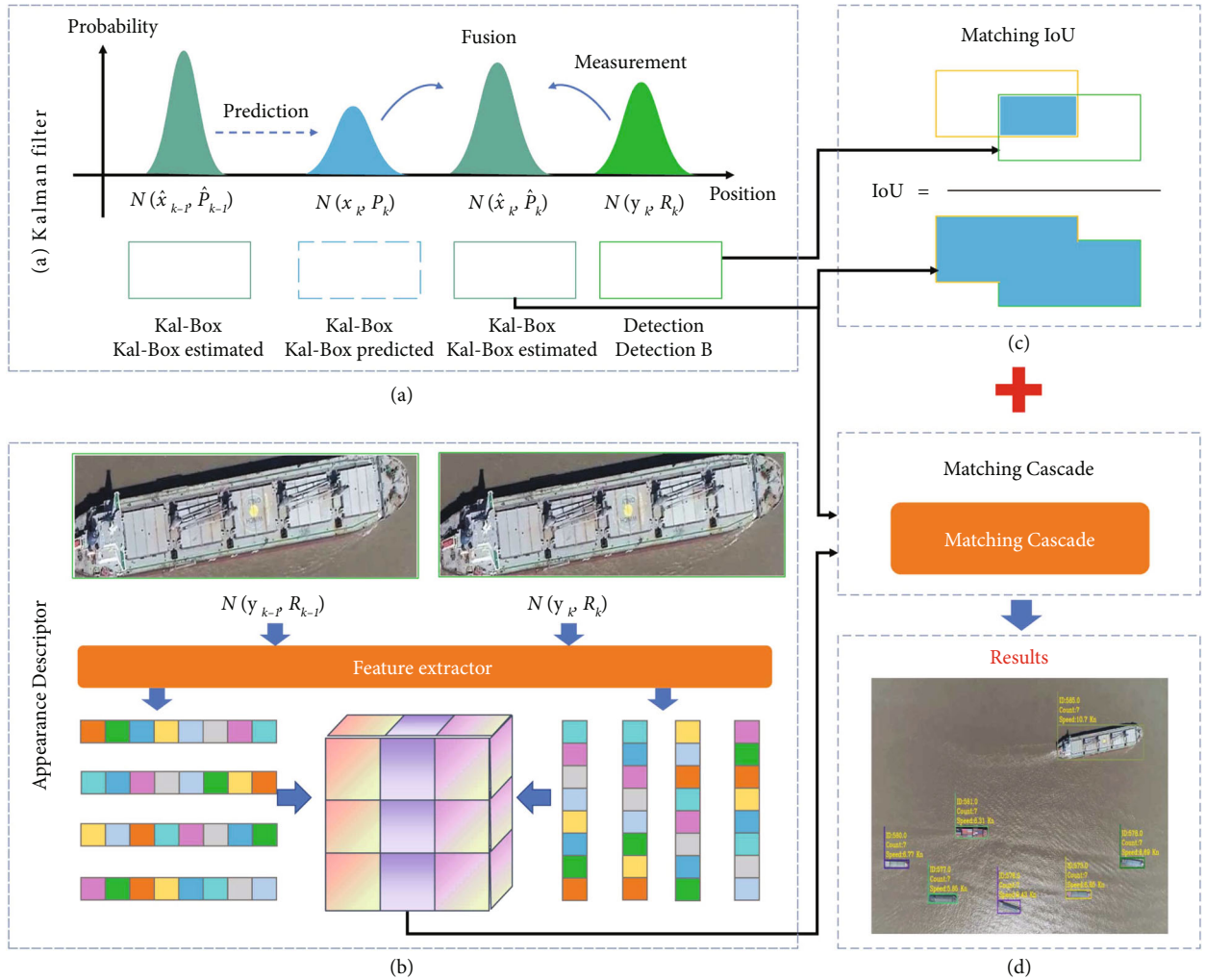


FIGURE 3: Schematic diagram of ship tracking.

algorithm to handle motion prediction and data association problems [29]. Deep SORT consists of four parts: Deep Appearance Descriptor, Kalman filter, Matching Cascade, and Matching Intersection-over-Union, as shown in Figure 3.

In the first part, The ReID network is employed to extract the appearance feature of the target. The ReID architecture is shown in Table 1. It should be noted that the global feature map with 128 dimensions is calculated in dense layer 10 and normalized by using L_2 normalization at the last layer.

Part two, according to the motion state of ship targets in the previous frame, the positions of the targets in the current frame are predicted by Kalman filter by using a linear isokinetic model $(x, y, r, h, x', y', r', h')$, containing the observed variables (x, y, r, h) , the bounding box center position (x, y) , aspect ratio r , height h , and their respective velocities.

In part three, the matching cascade is used for fusing target motion and appearance information to decrease the uncertainty of Kalman filter, when the target is occluded

TABLE 1: Overview of the ReID architecture.

Name	Patch size/stride	Output size
Conv1	3 × 3/1	32 × 128 × 64
Conv2	3 × 3/1	32 × 128 × 64
Max Pool 3	3 × 3/2	32 × 64 × 32
Residual 4	3 × 3/1	32 × 64 × 32
Residual 5	3 × 3/1	32 × 64 × 32
Residual 6	3 × 3/2	64 × 32 × 16
Residual 7	3 × 3/1	64 × 32 × 16
Residual 8	3 × 3/2	128 × 16 × 8
Residual 9	3 × 3/1	128 × 16 × 8
Dense 10		128
Batch and L_2 normalization		128

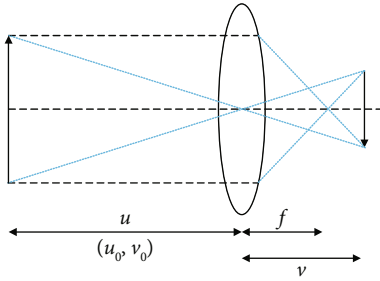


FIGURE 4: Keyhole imaging diagram.

for a long time. In the matching cascade, Mahalanobis distance is used to correlate Kalman filter prediction state and the YOLO v4 detection state, the motion information of the targets $d_{(i,j)}^1$ is shown in the following formula:

$$d_{(i,j)}^1 = (d_j - y_i)^T S_i^{-1} (d_j - y_i), \quad (2)$$

wherein, d_j denotes the motion state of the j_{th} detection box, y_i represent the predicted motion state of the i_{th} tracker, and S_i^{-1} is the covariance matrix in the observation space.

Appearance metric $d_{(i,j)}^2$ measures the smallest cosine distance between i_{th} tracker and j_{th} detection bounding box in the appearance space, as shown in the following formula:

$$d_{(i,j)}^2 = \min \left\{ 1 - r_j^T r_k^{(i)} \mid r_k^{(i)} \in R_k \right\}, R_k = \left\{ r_k^{(i)} \right\}_1^{k=100}, \quad (3)$$

where the r_j is the appearance feature descriptor, $|r_j| = 1$, and the R_k is a gallery to store a series of feature descriptors.

The correlation metric $c_{i,j}$ is combined as:

$$c_{i,j} = \lambda d_{(i,j)}^1 + (1 - \lambda) d_{(i,j)}^2, \quad (4)$$

where the λ is the weight coefficient.

In the last part, the matching IoU is applied for calculating the intersection and concurrency between the detection bounding box and the predicted box by Kalman filter, when there is no occlusion or new target appears. The detection

will be identified if the matching value is greater than IoU_{\min} , and the mathematical formula is

$$IoU = \frac{|area(A) \cap area(B)|}{|area(A) \cup area(B)|}, \quad (5)$$

where A is the detection bounding box, and B represents the prediction bounding box of the candidate trajectory.

2.4. Ship Speed Measurement. In this section, we assume that T_k and T_{k+FPS} are the trajectories in the frame k_{th} and $(k + FPS)_{th}$ frame firstly. Then, the pixel difference Δu and Δv in axis x and y are computed based on the center point of the bounding box of the trajectories, and the mathematical formula is as follows:

$$\Delta u = |u_{T_{k+FPS}} - u_{T_k}|, \quad (6)$$

$$\Delta v = |v_{T_{k+FPS}} - v_{T_k}|. \quad (7)$$

In a camera model, keyhole imaging is a relatively simple and commonly used model. Keyhole imaging principle as shown in Figure 4. And the related formula is as follows:

$$\frac{1}{f} = \frac{1}{u} + \frac{1}{v}, \quad (8)$$

where f represents the focal length, v is the image distance, u is the object distance.

The coordinate systems involved in solving Keyhole imaging parameters including world coordinate system, camera coordinate system, image coordinate system, and pixel coordinate system. The coordinate definition of the points in the specific coordinate system and the relationship between coordinate systems are as follows:

- (1) World coordinate system. An absolute coordinate system in a three-dimensional world. The position of the object is expressed in (X_w, Y_w, Z_w)
- (2) Camera coordinate system. The camera's center of light is the origin point, and the camera coordinates are represented by (X_c, Y_c, Z_c)
- (3) Image coordinate system. The center of the image plane is the coordinate origin, and the image coordinates are represented by (x, y)
- (4) Pixel coordinate system. The vertex in the upper left corner of the image plane is the origin, and the pixel coordinates are represented by (u, v)

The relationship among the image coordinate system, the camera coordinate system and the world coordinate system are as shown in Figure 5.

The transformation formula of the camera coordinate system to the world coordinate system is as follows:

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} = \begin{bmatrix} R & t \\ 0^T & 1 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix}, \quad (9)$$

where R is orthogonal rotation matrix and t is a three dimensions translation vector.

The conversion formula of image coordinate system to camera coordinate system is as follows:

$$x = f \frac{X_c}{Z_c}, \quad (10)$$

$$y = f \frac{Y_c}{Z_c}. \quad (11)$$

The above relationship is represented by a matrix form:

$$Z_C \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix}. \quad (12)$$

The relationship between the image coordinate system and the pixel coordinate system is as shown in Figure 6.

The transformation relationship formula between the image coordinate system and the pixel coordinate system is:

$$u = \frac{x}{d_x} + u_0, \quad (13)$$

$$v = \frac{y}{d_y} + v_0, \quad (14)$$

where (u_0, v_0) is the representation of the origin of the image coordinate system in the pixel coordinate system. d_x and d_y are the scale factor between two coordinate systems in direction x and y . The upper formula is represented by a matrix form:

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{d_x} & 0 & u_0 \\ 0 & \frac{1}{d_y} & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}, \quad (15)$$

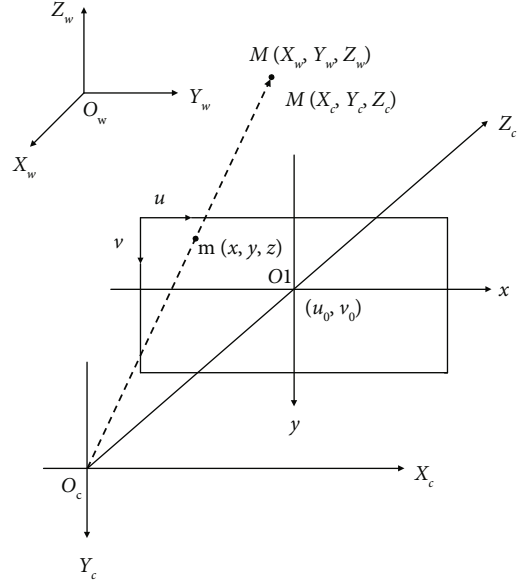


FIGURE 5: The relationship diagram among the image, the camera, and the world coordinate system.

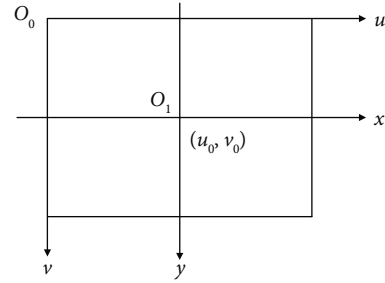


FIGURE 6: The relationship diagram between the image and the pixel coordinate system.

comprehensive:

$$Z_C \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{d_x} & 0 & u_0 \\ 0 & \frac{1}{d_y} & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} R & t \\ 0^T & 1 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix}, \quad (16)$$

thereinto:

$$a_x = \frac{f}{d_x}, \quad (17)$$

$$a_y = \frac{f}{d_y}, \quad (18)$$

$$K = \begin{bmatrix} a_x & 0 & u_0 & 0 \\ 0 & a_y & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \quad (19)$$

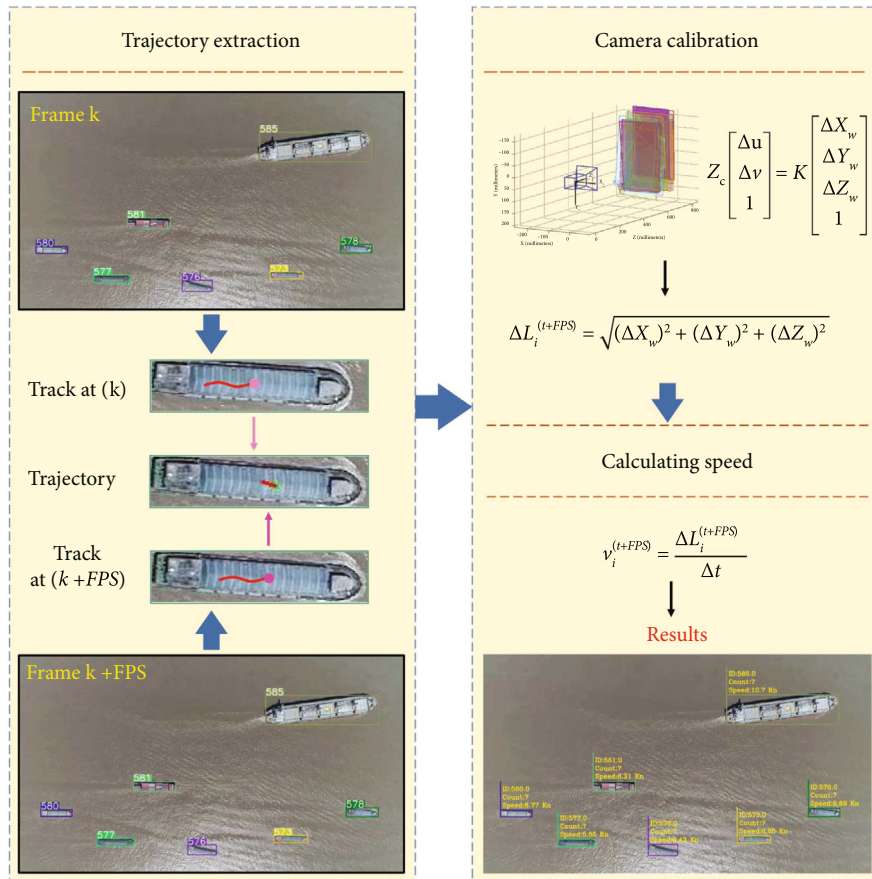


FIGURE 7: Elementary diagram of the speed measurement.



FIGURE 8: Self-built ship dataset.

$$M = \begin{bmatrix} R & t \\ 0^T & 1 \end{bmatrix}, \quad (20)$$

$$Z_C \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} a_x & 0 & u_0 & 0 \\ 0 & a_y & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} R & t \\ 0^T & 1 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} = KM \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix}, \quad (21)$$

among them, (a_x, a_y, u_0, v_0) is the inner parameter, K is called internal parameter matrix. M is called the camera's outer parameter matrix, R and t are external parameters.

Based on the value of the $(\Delta u, \Delta v)$, the mapping relationship between two-dimensional and three-dimensional space is calculated by the camera calibration [30]. Since there is no rotation and translation, the equation of spatial mapping relationship is as follows:

$$Z_C \begin{bmatrix} \Delta u \\ \Delta v \\ 1 \end{bmatrix} = K \begin{bmatrix} \Delta X_w \\ \Delta Y_w \\ \Delta Z_w \\ 1 \end{bmatrix}, \quad (22)$$

where the Z_C is the displacement of the camera from the object, K is the conversion matrix, the $(\Delta X_w, \Delta Y_w, \Delta Z_w)$ donates the displacement in the actual space, and the speed of the ship can be estimated according to the following equation:

$$\Delta L_i^{(t+FPS)} = \sqrt{(\Delta X_w)^2 + (\Delta Y_w)^2 + (\Delta Z_w)^2}, \quad (23)$$

$$v_i^{(t+FPS)} = \frac{\Delta L_i^{(t+FPS)}}{\Delta t}, \quad (24)$$

where $\Delta L_i^{(t+FPS)}$ and $v_i^{(t+FPS)}$ represent the real distance and ship speed in frame $t + FPS$, respectively. The ship speed estimation process is shown in Figure 7.

3. Procedure

In this part, we introduce our dataset for ship detection in the airborne video and some parameter values set by us. Our dataset is originated from the UAV video (FPS is equal to 23.98) taken by DJI Genie 4 Pro, in Shanghai, China. The dataset comprises 18,655 images, including three scenes with four shooting angles and two lighting conditions, and the image resolution is 3840×2160 . Besides, the UAV flying height and focal length are 500 m and 24 mm, respectively. Dataset image and schematic diagram of the shooting angle are all shown in Figure 8. Some images of our dataset have been published on the website: <https://github.com/NZII>. Those experiments are operated on a computer with an Intel I7-11800H @ 4.6 GHz processor and 6G memory.

In part of ship detection and tracking, YOLO v4 and Deep SORT algorithm are used and the parameters of the original model are retained to ssaccelerate the convergence of our model. Then, according to the video information, the interval of speed measurement is set to 1 s. Next, according to the UAV flying height, transformation matrix K is obtained, as is shown in Equation (25), that is, 1 pixel in the image is approximately equal to 0.205 m in real world. Finally, the ship speed is estimated successfully.

$$K = \begin{bmatrix} 2608.5 & 0 & 1934.4 \\ 0 & 2608.9 & 1105.7 \\ 0 & 0 & 1 \end{bmatrix}. \quad (25)$$

Owing to verify the performance of our system, the quality assessment experiment and method comparison experiment are construed. In these assessment experiments, the results of ship speed estimated by our method are compared with the ship AIS data and an excellent accuracy of our method is got successfully.

4. Results

4.1. Multiscenario Detection and Tracking. In this section, the detection and tracking performance of our method are evaluated by selecting 3 scenes with the resolutions 3840×2160 shown in Figure 9. Moreover, these scenes contain 3 shooting angles and 3 brightness conditions.

To test the detection and tracking stability, according to the video # 1 with 45° and 60° shooting angles, scene 1, grey-sea-surface background, bright illumination, and some of the ships in different positions of the image are conducted.

Scene 2, background and illumination are similar to scene 1, but the overexposed appears in the bottom of the image, which is originated from video # 2 with the shooting angle 90° to verify the detection and tracking accuracy in general conditions. Moreover, we can also clearly see that some ships pass through the overexposed area and others are not.

Referring to scene 3 with dark background and overexposed area in the bottom of the image, it is constructed based on video # 3 with the shooting angle 90° . It is used to test the accuracy of detection and tracking in dark conditions.

The detection and tracking results are shown in Figure 9, three scenes include many ships with three sizes like small, middle, and big, and different kinds of ship distributions. In scene 1, our detection model has an excellent accuracy because of the mosaic data enhance of the YOLO v4 network, which especially improves the ability of detection small targets in the image. # 1272 and # 1680 are taken in two shooting angles (45° and 60°), the Kalman filter algorithm and Hungarian algorithm in Deep SORT can predict and match different directions of detection box, thus even if the angle has changed in the process of ship moving, we also can obtain the ship position.

In scene 2, we can clearly distinguish the trajectories of different ships, both the shooting angles and the

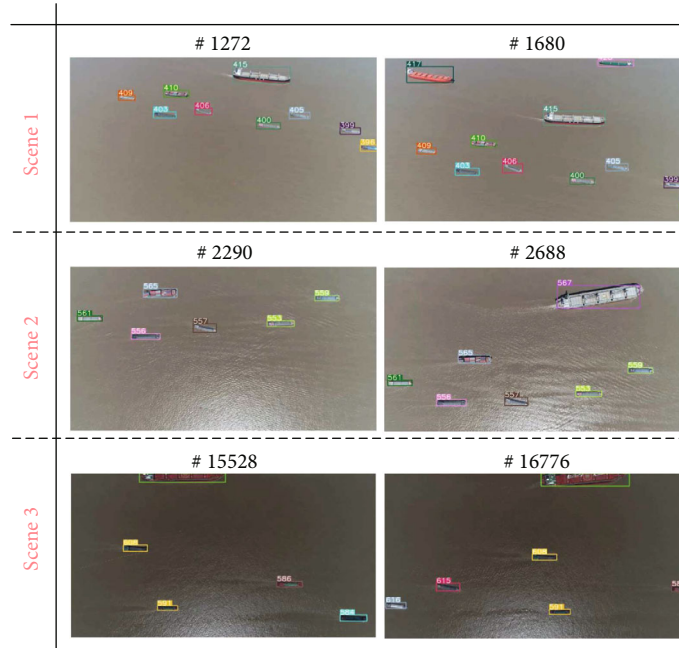


FIGURE 9: Multiscene ship detection and tracking results.

illumination conditions are also excellent. Due to the CSP strategy in the YOLO v4 network, a series of ship feature map information in different kinds of levels are obtained. Besides, Kalman filtering is very suitable for linear prediction scenarios, which ensure our model having better robustness and accurate in a good condition.

In scene 3, the illumination is dark, which will affect the detection accuracy and tracking stability. The SPP and SPPNet modules in YOLO v4 can separate the most important contextual features and make low-layer information easier to spread to the top layers. Furthermore, the appearance information in Deep SORT can decrease the uncertainty of Kalman filter, when the target is missed for a long time due to the bad environment condition. This experiment proves that our method has an excellent detection and tracking accuracy, which is crucial for the next step calculating vectors and extracting speed.

4.2. Multi-Ship Speed Quality Analysis and Method Comparison. In this part, AIS data is set as the truth value of ship speed and scenario 3 shown in Figure 9 is selected. AIS data of Baosteel terminal on April 5, 2021 is downloaded from the website: <http://www.shipxy.com>. Because of AIS data, a discrete spatiotemporal sequence cannot be matched with the continuous ship speed value estimated by our method. Therefore, based on the FPS of drone video, linear interpolation is employed to make AIS data continuous. Then, two factors, the sailing position and the illumination condition, are considered in the measurement process of ship speed, HUA QING and HUAI CHANG in scenario 3 are selected eventually.

Finally, AIS information is set as a standard, our method is compared with the visual line method. The speed comparison results and error curves of the three ships are shown in

Figure 10. Moreover, mean absolute error (MAE) and mean squared error (MSE) are introduced to estimate our velocimetry method, comprehensively. The final evaluation results are shown in Table 2. Evaluation indicator calculation formula is as follows:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2, \quad (26)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|, \quad (27)$$

where n is the number of speed measurements, y_i is the value of ship speed estimated by our system, and \hat{y}_i is the value of ship speed in AIS data.

As shown in the curves of HUA QING, we can clearly see the ship move in a stable speed. The ship speed measured by visual line is quite different from the truth value, MAE is 0.91 Kn and MSE is 2.0 Kn. The ship speed estimated by our method is relatively close to the true value and appears some reasonable fluctuation, MAE is 0.59 Kn and MSE is 1.25 Kn.

For the tachymetry curves of HUAI CHANG, the visual line method has been greatly improved in terms of accuracy, the MAE is 0.41 Kn and the MSE is 0.40 Kn. Significantly, our method still maintains an excellent performance, MAE is 0.31 Kn and MSE is 0.34 Kn.

Overall, the performance of the visual line method is uncontrollable in the process of ship speed measurement, the main reason is the visual line method only can estimate an interval average speed in a period of time, which cannot represent a real-time speed information and easily cause a global error. On the contrary, our method shows real-time ship speed information with a relatively good accuracy.

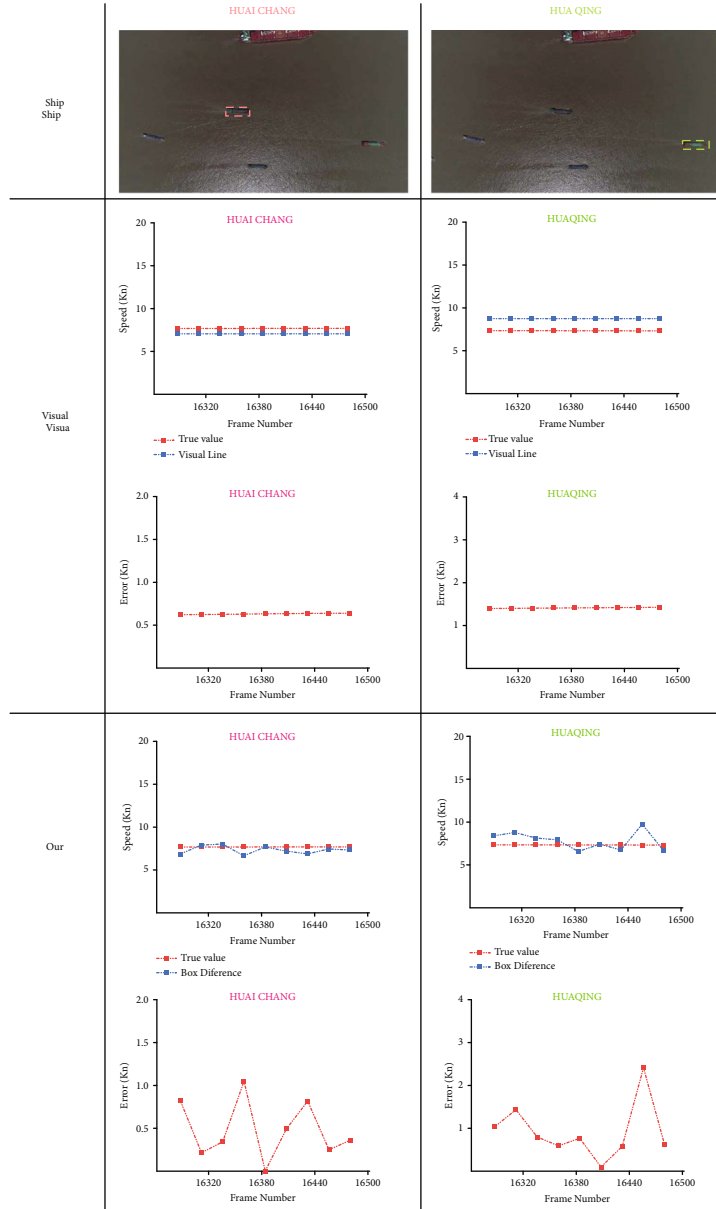


FIGURE 10: Multimethod of ship speed measurement comparison results.

TABLE 2: Evaluation results of multimethod of velocimetry.

Evaluation indicators	Ship name	Speed measurement	
		Visual line	Our
MAE	HUA QING	0.91	0.59
[Kn]	HUAI CHANG	0.41	0.31
MSE	HUA QING	2.00	1.25
[Kn]	HUAI CHANG	0.40	0.34

Although our method is relying on the information of the bounding box, which may lead to some fluctuations and have an influence on the ship speed accuracy, ship speed data can be obtained in real time and the average accuracy is above 93% in the complex scenario.

5. Conclusions

In this paper, we propose a methodological method for extracting ship speed by combing the machine vision algorithm and the UAV images. In the detection and tracking experiments, we take into account the shooting angles and lighting conditions to test the robustness of the model. The detection and tracking results show that our model still has good detection accuracy in dense scenes with poor lighting scenes, and provides the basis for subsequent speed measurements. In the speed measurement experiments where the AIS is set to true value, we compare our method with the virtual line method by selecting three ships with different sizes in the same scene. The results show that our method has obvious advantages in terms of accuracy, stability, and applicable scenarios. Specially, the speed accuracy is above

the 93%, which is a significant prerequisite for prediction ship speed and ship traffic flow in a serious manner.

Data Availability

The <https://github.com/NZII> data used to support the findings of this study are included within the article.

Conflicts of 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 paper.

Acknowledgments

This work was jointly supported by the National Key Research and Development Program of China (Grant No. 2021YFC2801004), National Natural Science Foundation of China (No. 52102397, 52071199), China Postdoctoral Science Foundation (No. 2021M700790).

References

- [1] R. W. Liu, M. Liang, J. Nie, W. Y. B. Lim, Y. Zhang, and M. Guizani, "Deep learning-powered vessel trajectory prediction for improving smart traffic services in maritime internet of things," *IEEE Transactions on Network Science and Engineering*, vol. 9, no. 5, pp. 3080–3094, 2022.
- [2] R. W. Liu, M. Liang, J. Nie et al., "STMGCN: mobile edge computing-empowered vessel trajectory prediction using spatio-temporal multi-graph convolutional network," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 11, pp. 7977–7987, 2022.
- [3] D. Yang, L. Wu, S. Wang, H. Jia, and K. X. Li, "How big data enriches maritime research – a critical review of automatic identification system (AIS) data applications," *Transport Reviews*, vol. 39, no. 6, pp. 755–773, 2019.
- [4] G. Xiao, T. Wang, X. Chen, and L. Zhou, "Evaluation of ship pollutant emissions in the ports of Los Angeles and Long Beach," *Journal of Marine Science and Engineering*, vol. 10, no. 9, p. 1206, 2022.
- [5] Y. Sato, "Radar speed monitoring system," in *Proceedings of VNIS'94-1994 Vehicle Navigation and Information Systems Conference*, pp. 89–93, Yokohama, Japan, 1994.
- [6] E. Musayev, "Laser-based large detection area speed measurement methods and systems," *Optics and Lasers in Engineering*, vol. 45, no. 11, pp. 1049–1054, 2007.
- [7] M. Lobur and Y. Darnoby, "Car speed measurement based on ultrasonic Doppler's ground speed sensors," in *2011 11th International Conference The Experience of Designing and Application of CAD Systems in Microelectronic*, pp. 392–393, Polyana, Ukraine, 2011.
- [8] E. Odat, J. S. Shamma, and C. Claudel, "Vehicle classification and speed estimation using combined passive infrared/ultrasonic sensors," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 5, pp. 1593–1606, 2018.
- [9] D. C. Luvizon, B. T. Nassu, and R. Minetto, "A video-based system for vehicle speed measurement in urban roadways," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 6, pp. 1393–1404, 2017.
- [10] W. Jianping, L. Zhaobin, L. Jinxiang, G. Caidong, S. Maoxin, and T. Fangyong, "An algorithm for automatic vehicle speed detection using video camera," in *s*, pp. 193–196, Nanning, China, 2009.
- [11] J. X. Wang, "Research of vehicle speed detection algorithm in video surveillance," in *ICALIP 2016-2016 International Conference on Audio, Language and Image Processing-Proceedings*, pp. 349–352, Shanghai, China, 2017.
- [12] D. F. Llorca, C. Salinas, M. Jimenez et al., "Two-camera based accurate vehicle speed measurement using average speed at a fixed point," in *IEEE Conference on Intelligent Transportation Systems, Proceedings*, pp. 2533–2538, Rio de Janeiro, Brazil, 2016.
- [13] R. Ke, Z. Li, S. Kim, J. Ash, Z. Cui, and Y. Wang, "Real-time bidirectional traffic flow parameter estimation from aerial videos," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 4, pp. 890–901, 2017.
- [14] T. Moranduzzo and F. Melgani, "Car speed estimation method for UAV images," in *International Geoscience and Remote Sensing Symposium*, pp. 4942–4945, Quebec City, QC, Canada, 2014.
- [15] R. Pelich, N. Longepe, G. Mercier, G. Hajduch, and R. Garello, "Vessel refocusing and velocity estimation on SAR imagery using the fractional fourier transform," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 3, pp. 1670–1684, 2016.
- [16] A. Renga and A. Moccia, "Use of Doppler parameters for ship velocity computation in SAR images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 7, pp. 3995–4011, 2016.
- [17] A. Panico, M. D. Graziano, and A. Renga, "SAR-Based vessel velocity estimation from partially imaged kelvin pattern," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 11, pp. 2067–2071, 2017.
- [18] K. Kanistras, G. Martins, M. J. Rutherford, and K. P. Valavanis, "A survey of unmanned aerial vehicles (UAVs) for traffic monitoring," in *International Conference on Unmanned Aircraft Systems*, pp. 2643–2666, 2015.
- [19] M. Liang, R. W. Liu, S. Li, Z. Xiao, X. Liu, and F. Lu, "An unsupervised learning method with convolutional auto-encoder for vessel trajectory similarity computation," *Ocean Engineering*, vol. 225, article 108803, 2021.
- [20] J. Zhao, J. Lu, X. Chen, Z. Yan, Y. Yan, and Y. Sun, "High-fidelity data supported ship trajectory prediction via an ensemble machine learning framework," *Physica A: Statistical Mechanics and its Applications*, vol. 586, article 126470, 2022.
- [21] X. Chen, J. Ling, S. Wang, Y. Yang, L. Luo, and Y. Yan, "Ship detection from coastal surveillance videos via an ensemble canny-Gaussian-morphology framework," *Journal of Navigation*, vol. 74, no. 6, pp. 1252–1266, 2021.
- [22] X. Han, L. Zhao, Y. Ning, and J. Hu, "ShipYOLO: an enhanced model for ship detection," *Journal of Advanced Transportation*, vol. 2021, Article ID 1060182, 11 pages, 2021.
- [23] C. Xu, D. Zhang, Z. Zhang, and Z. Feng, "BgCut: automatic ship detection from UAV images," *The Scientific World Journal*, vol. 2014, Article ID 171978, 11 pages, 2014.
- [24] P. Msonda, S. A. Uymaz, and S. S. Karaağaç, "Spatial pyramid pooling in deep convolutional networks for automatic tuberculosis diagnosis," *Traitement du Signal*, vol. 37, no. 6, pp. 1075–1084, 2020.
- [25] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: towards real-time object detection with region proposal networks,"

- IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, 2017.
- [26] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, “YOLOv4: optimal speed and accuracy of object detection,” 2020, <http://arxiv.org/abs/2004.10934>.
- [27] W. Liu et al., “SSD: Single shot multibox detector,” in *Lecture Notes in Computer Science*, pp. 21–37, Springer, 2016.
- [28] N. Wojke, A. Bewley, and D. Paulus, “Simple online and real-time tracking with a deep association metric,” in *Proceedings-International Conference on Image Processing*, pp. 3645–3649, Beijing, China, 2018.
- [29] C. Liu and J. Li, “Self-correction ship tracking and counting with variable time window based on YOLOv3,” *Complexity*, vol. 2021, Article ID 2889115, 9 pages, 2021.
- [30] Z. Zhang, “A flexible new technique for camera calibration,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 11, pp. 1330–1334, 2000.