Hindawi
Journal of Advanced Transportation
Volume 2020, Article ID 7194342, 12 pages
https://doi.org/10.1155/2020/7194342

Research Article

Video-Based Detection Infrastructure Enhancement for Automated Ship Recognition and Behavior Analysis

Xinqiang Chen,[1] **Lei Qi**,[1] **Yongsheng Yang**,[1] **Qiang Luo**,[2] **Octavian Postolache**,[3] **Jinjun Tang**,[4] **and Huafeng Wu**[5]

[1] *Institute of Logistics Science and Engineering, Shanghai Maritime University, Shanghai 201306, China*
[2] *School of Civil Engineering, Guangzhou University, Guangzhou 510006, China*
[3] *ISCTE – Instituto Universitário de Lisboa, Lisbon University Institute, Lisbon, Portugal*
[4] *School of Traffic and Transportation Engineering, Central South University, Changsha 410075, China*
[5] *Merchant Marine College, Shanghai Maritime University, Shanghai 201306, China*

Correspondence should be addressed to Jinjun Tang; jinjuntang@csu.edu.cn

Received 31 July 2019; Revised 5 October 2019; Accepted 28 October 2019; Published 20 January 2020

Guest Editor: Hao Yu

Video-based detection infrastructure is crucial for promoting connected and autonomous shipping (CAS) development, which provides critical on-site traffic data for maritime participants. Ship behavior analysis, one of the fundamental tasks for fulfilling smart video-based detection infrastructure, has become an active topic in the CAS community. Previous studies focused on ship behavior analysis by exploring spatial-temporal information from automatic identification system (AIS) data, and less attention was paid to maritime surveillance videos. To bridge the gap, we proposed an ensemble you only look once (YOLO) framework for ship behavior analysis. First, we employed the convolutional neural network in the YOLO model to extract multi-scaled ship features from the input ship images. Second, the proposed framework generated many bounding boxes (i.e., potential ship positions) based on the object confidence level. Third, we suppressed the background bounding box interferences, and determined ship detection results with intersection over union (IOU) criterion, and thus obtained ship positions in each ship image. Fourth, we analyzed spatial-temporal ship behavior in consecutive maritime images based on kinematic ship information. The experimental results have shown that ships are accurately detected (i.e., both of the average recall and precision rate were higher than 90%) and the historical ship behaviors are successfully recognized. The proposed framework can be adaptively deployed in the connected and autonomous vehicle detection system in the automated terminal for the purpose of exploring the coupled interactions between traffic flow variation and heterogeneous detection infrastructures, and thus enhance terminal traffic network capacity and safety.

1. Introduction

SHIP behavior recognition and prediction is very important for the early warning of risky behavior, identifying potential ship collision, improving maritime traffic efficiency, etc., and thus is a very active topic in the intelligent maritime navigation community. Currently, we primarily rely on the AIS data to explore traffic flow knowledge under varied maritime traffic situations. The large-scale available AIS data support research of probing maritime spatial-temporal traffic patterns, and recognizing ship behaviors by inferring from ship locations, ship heading directions, ship speeds (i.e., speed over ground and speed over water), etc. Wang et al. obtained the spatial-temporal traffic tensors by mining the high-resolution AIS data, and then employed a sparse multi-linear decomposition method to predict ship behaviors [1]. Li et al. developed a multi-dimension scaling model to explore spatial similarity among extensive ship trajectories, and then an improved density spatial clustering algorithm is proposed to acquire the optimal AIS clusters and recognize potential abnormal ship behaviors at a fixed time interval [2]. Zhao et al. combined the Douglas-Peucker-based compression model and density clustering method to discover maritime traffic patterns [3]. Arguedas et al. proposed a two-layer artificial neural network to represent structured maritime traffic patterns, which provide maritime regulators a high-efficiency tool for perceiving real-time maritime situation, and fulfill the automatic maritime traffic monitoring task [4].

The AIS data can efficiently model ship trajectories at sea, and help maritime relevant participants (maritime officials, ship crew, etc.) take early actions to avoid potential accidents. Zhang et al. proposed different frameworks to recognize possible near miss ship-ship collisions from AIS data [5, 6]. Bye et al. analyzed maritime accidents by mining the inshore ship AIS data considering varied maritime static and kinematic information, such as sailed nautical miles, accumulated engine working hours, port call number, ship type, flag state, gross tonnage, etc. [7]. Integrating the AIS data with other maritime sources (synthetic aperture radar (SAR), radar, etc.) to fulfill accurate ship behavior recognition task have shown numerous successes. Mazzarella et al. fused the space-borne SAR images and AIS information to explore maritime traffic knowledge by deeply exploiting historical ship trajectories, cross-validate ship positions detected in satellite imagery and recognize those ships that deliberately hide their sailing information [8]. Habtemariam et al. developed a measurement-level fusion algorithm by merging radar data and AIS messages with a novel joint probabilistic data association framework [9]. The data fusing relevant methods obtain comprehensive information for both, on and off-site maritime traffic, and support robust ship behavior exploration.

Though the AIS dataset contains rich information for ensuring maritime safety, security and efficiency, the following critical weaknesses reduce AIS based techniques performance when analyzing ship behaviour: (1) Some ships (e.g., fishing boats) may not be equipped with AIS relevant facility, and some ships sailing at sea may attempt to deactivate (or even shut down) their AIS transmitters (smuggling ships, warships, etc.) [10]. (2) AIS equipment broadcasts the host ship static and kinematic information at a fixed frequency (usually varied from 2 to 10 seconds) when the ship is sailing at coastal channels, and the data broadcasting interval can extend to 3 minutes when the ship is in anchoring state, which leads to significant challenge of formatting AIS database (creating, retrieving, updating and deleting operations). (3) We can hardly obtain the visual spot traffic information straightforwardly from AIS system. More specifically, we need to manually recover the original maritime traffic situations by inferring from ship trajectories with the support of historical AIS data, which is very time consuming and labor-intensive. The Long Range Identification and Tracking (LRIT) technique is another useful method for obtaining ship positions. But, the LRIT data is private and confidential, and thus it is not easy for the public to access the data.

The extensive deployment of maritime sensors and rapid development of computer vision techniques help us easily collect, store and analyze the on-site maritime transportation data. Currently, public accessible maritime image sources include SAR image, infrared data, and closed-circuit television (CCTV) videos. The SAR images sharply scale down the original ship-to-ship distances as it shoots maritime images at a very high attitude (approximately 5000 km above the earth). The image quality may be severely degraded by strong clouds, and thus may fail to provide us high-resolution imageries for conducting accurate maritime traffic situations analyses [11]. The infrared image resolution is not high due to the intrinsic infrared imaging technique bottlenecks, which impose great challenge of extracting high-fidelity traffic information for analyzing the small ship (i.e., ship size in images are small) behaviours [12]. Besides, the performance of infrared based techniques is easily interfered by the wave and ship engine temperature variations [13].

The CCTV data sources provide us rich and real time on-spot traffic information (traffic volume, ship speed, heading angle, etc.), and thus support high-fidelity ship behavior analysis researches. Valsamis et al. employed traditional machine learning algorithms to extract ship trajectories from CCTV videos [14]. Ship tracking and detection are the two popular topics for implementing ship behaviors recognition task via CCTV data sources. Zhang et al. presented a ship detection framework to remove vibration interference generated by non-stationary surface platform (buoys, sailing ships, etc.), and yielded trustable visual maritime surveillance results [15]. Yao et al. proposed a local visual saliency map to detect ships from GF-4 satellite sequential images, and the local peak signal-to-noise ratio indicator was introduced to quantitatively evaluate the model performance [16]. Kang et al. proposed a self-selective correlation filtering method to solve the ship scale variation challenge for the purpose of ship tracking [17]. The deep learning methods have shown great potential in object detection and tracking field, which were pre-trained by the public-access benchmarks, and the models were then fine-tuned with customized data to obtain satisfactory ship behavior recognition performance. Woo et al. developed a long short-term memory based recurrent neural network structure to detect and predict kinematic behaviors of unmanned surface vehicles [18]. Gao et al. developed an online real-time ship behaviour prediction model by constructing a bidirectional long short-term memory recurrent deep learning neural network [19]. Similar researches can be found in [20–22].

After carefully reviewing the previous ship behavior related studies, we found the following disadvantages significantly challenge the ship behavior recognition performance (from maritime video data): (1) ships sailing far from monitoring camera can be severely interfered by background imaging pixels, especially for the ships have similar intensity with background. More specifically, the ship visual features may be contaminated by background, which may not be easily extracted by the feature detectors; (2) ships in the maritime images are quite easily sheltered by obstacles (such as sea clutters, neighboring ships), leading to a big challenge of accurate extracting high-fidelity ship imaging positions. To address the issue, we proposed a novel framework to achieve accurate ship behavior recognition with four consecutive steps. More specifically, the ensemble YOLO framework was developed to accurately determine ship positions in consecutive maritime images, and then ship trajectories was modeled and recognized based on geometry knowledge.

The findings in the research provide us on-site ship kinematic information which significantly benefits automated terminal data stream interaction for enhancing terminal logistics efficiency. More specifically, the ship, port, and vehicles in terminal districts (e.g., container truck, automated guided vehicle) are closely connected for the purpose of safe and efficient cargo container trafficking. After obtaining ship kinematic information (displacement, moving speed, sailing angle, etc.) via maritime surveillance video, maritime participants can take early initiative activities to identify (and avoid) potential traffic collision through the manner of maneuvering ships, sending out risky information, etc. Meanwhile, the automated terminal management center can generate production planning and scheduling solution in advance, and then the autonomous vehicles are dispatched to the ship anchoring area to prepare for unloading the on-board containers and transmitting cargos to destinations (terminal yard, cargo receiver, etc.).

Our primary contributions were summarized as follows: (1) we have analyzed the pros and cons of automated ship recognition and ship behavior analysis via varied maritime data sources (AIS, LRIT, maritime surveillance videos, etc.). It is found that the video based methods provide us with more high-fidelity immediate and understandable on-site traffic situation awareness information in realistic applications compared to the other popular maritime data; (2) we employed a YOLO based ensemble framework to collect ship spatial-temporal dataset from maritime videos. More specifically, we extract high-fidelity ship kinematic information (i.e., moving displacements, speeds, course angle, accelerations) from maritime videos, which provides instantaneous traffic information to the maritime involved participants for taking early-warning measurements to avoid potential ship collisions; (3) we have collected ship video clips on two typical traffic scenarios (i.e., irregular ship turning motion, moving straight), from which we can extract both of microscopic and macroscopic maritime data supporting further maritime traffic flow knowledge discovery. Considering, a few ship video benchmarks are open for public, we are willing to share the collected ship videos with potential interested readers (by sending request email to jinjuntang@csu.edu.cn).

2. Data Description

The lack of public accessible ship videos (due to the data confidential and sensitivity) imposes additional challenge of evaluating performance of ship behavior recognition framework. To this end, we have shot several maritime surveillance videos from coastal areas near Shanghai terminal in China. The collected ship images are denoted as case-1 and case-2 scenarios, which are classified with ship sailing directions. More specifically, the first scenario focuses on ship moving-straight situation, and the second case involves with consistently irregular ship turning situation. It is noted that ships moves far away from camera may be very difficult to be recognized by human beings, and thus the small size ships (cannot be recognized by human beings) in maritime images are suppressed for further analysis. Readers are recommended to refer to [23] for the small size ship definition.

To enhance the framework generalization performance, we employed data augmentation techniques to generate more ship images by applying common augmentation operations. More specifically, we can obtain 20 variant ship images for each input training image with the help of data augmentation technique (with operations of translation, rotation, color shifting, etc.). The ship variant samples are manually selected from the generated 20 images where the visual ship features (edges, contours, color, etc.) obviously differ from the original input image (see Figure 1). We have collected 970 maritime frames in the case-1 and 2030 images in case-2. Overall, the ship datasets (for the two video clips) have 3000 maritime surveillance pictures, with 70% of them being used for training sets, and the rest as a validation set. More specifically, we select 679 frames in case-1 and 1421 images in case-2 for model training purpose, while the remaining 291 and 609 frames in case-1 and case-2 are used for the test purpose. Following the rules in the previous studies [24], we manually rectified training image resolutions into 720 × 480, and the ship image resolutions in the validation dataset were formatted into 416 × 416. The frame rate of each video is 30 frames per second (fps). The ground truth ship positions in each frame are manually labeled by our group member (i.e., undergraduate and graduate students).

3. Methodology

The overall framework developed for recognizing ship behaviors includes four steps: ship feature extraction, bounding box generation, ship position identification, and ship behavior analysis. The first step employs the YOLO network to extract ship features from the input training data at different scales. In the second step, our framework predicts a bunch of bounding boxes which are considered as potential ship positions in each image. The third step aims to remove the interference from irrelevant bounding boxes, the K-means method is introduced to obtain anchor boxes (i.e. potential ships), and the binary cross-entropy cost function is then solved to determine the final ship detection results. The third step employs geometry theory to recognize consecutive ship positions in each image (i.e., positions from same ship), and determine ship behaviors by analyzing the ship sailing angle variation. The flowchart of the proposed framework is shown in Figure 2.

3.1. Ship Feature Extraction. The YOLO model is introduced to learn the distinct ship features from input images by using a convolution neural network [24, 25]. More specifically, the YOLO model is introduced to explore ship features at different scales by varied scaled filters and obtain ship feature pyramids, which are composed by the high-resolution features (fine-grained level features). The convolutional neural network in the proposed ensemble YOLO framework is nested with convolutional layers, and the deeper layers in the nested network can exploit more discriminative ship features than those in the previous layers, which greatly benefits ship detection accuracy for the ensemble YOLO model. The obtained ship feature is shown as a matrix as follows:

$$sf_v^t = f\left(\sum_k S_k^{t-1} * \rho_{kv}^t + W_v^t\right), \quad (1)$$

where S_k^{t-1} is the kth input ship features from the $(t-1)$th convolutional layer, ρ_{kv}^t is the weight matrix between the vth and the kth ship feature layer. The parameter W_v^t is the bias of the vth output ship features at the tth convolutional network layer, and f represents the activation model used for activating

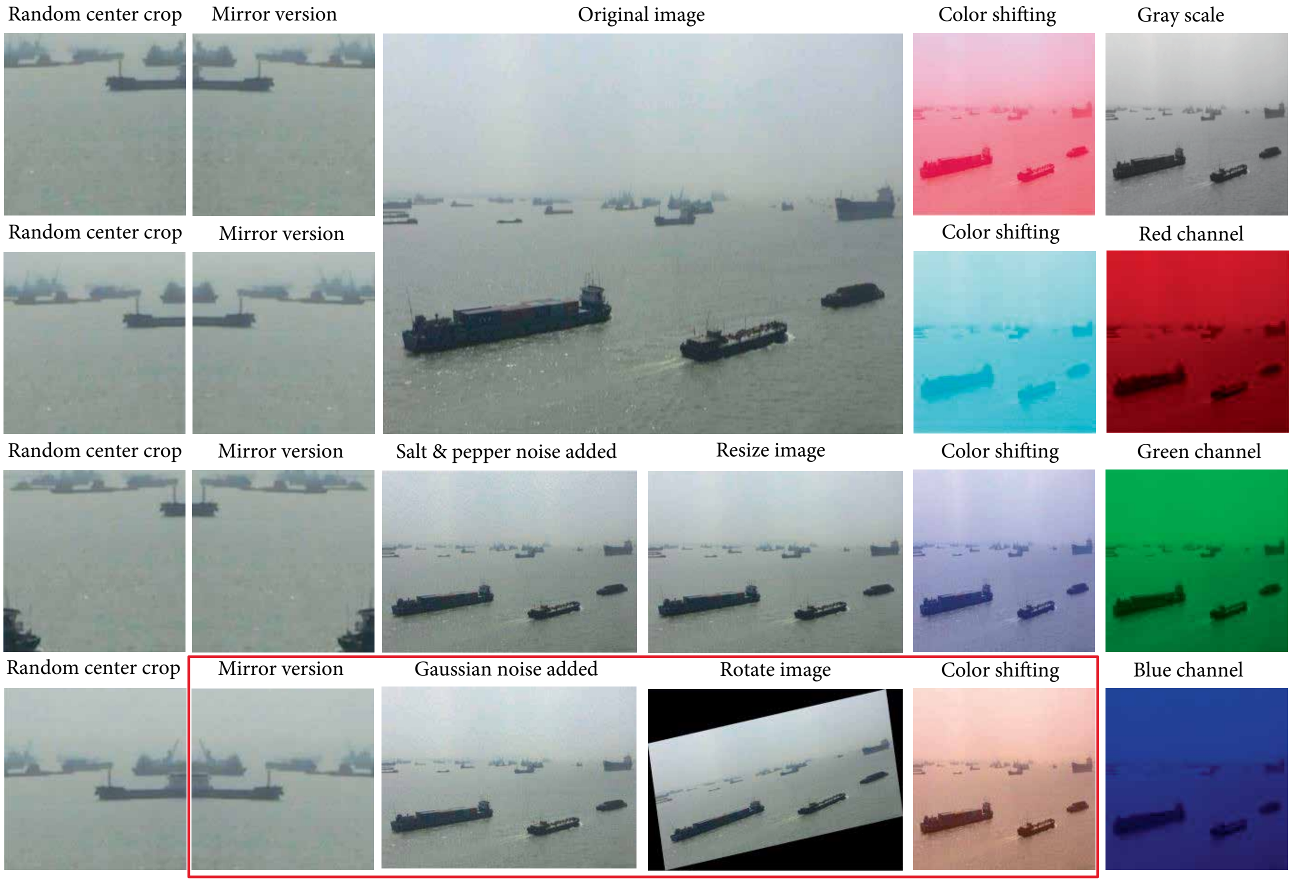


FIGURE 1: The original collected and data-augmentation generated ship images (the output images are labeled by red rectangle).

neurons at the tth layer. The sf_v^t is the vth output features at same layer.

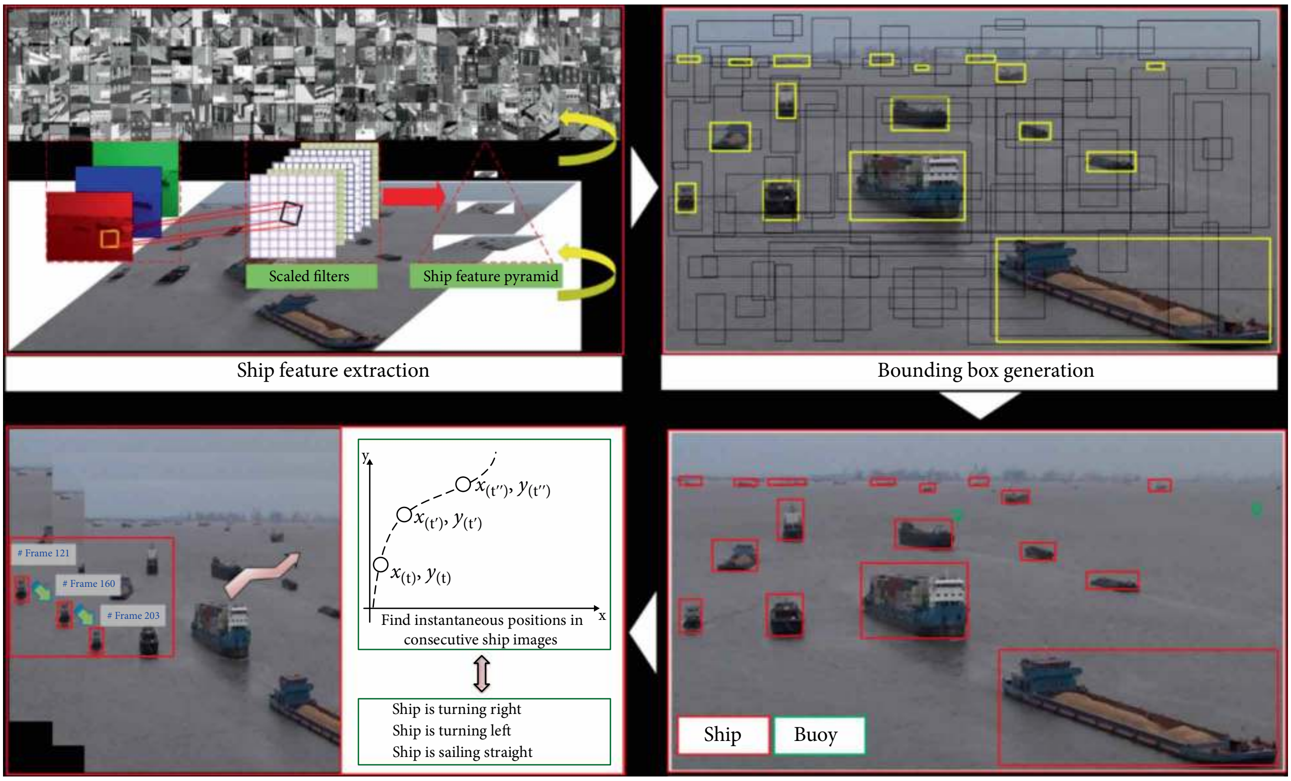
3.2. Bounding Box Generation. The ensemble YOLO framework predicts the coordinates of bounding boxes directly using fully connected layers on top of the convolutional feature extractor. More specifically, with the features extracted from the previous step, the original input image is split into $M \times M$ grids. The center of grid cell is used to predict the object location and class when the object image center falls in the grid center. The grid cell outputs a confidence score for depicting the object category, which is defined as $Cr \times \text{IOU}_{pre}^{gth}$. The confidence score is obtained as the IOU value between the ground truth box and the ensemble YOLO model predicted area (see Figure 3), and the IOU calculation formula is shown in Equation (3). Note that each grid cell outputs the object category confidence. The parameter Cr is set to 1 when the object center locates in the grid cell center, otherwise the parameter Cr is set to zero. Thus, the grid cell detection results are positively related with IOU value. A larger IOU shows the bounding box (i.e., detected ship position) is closer to the ground truth, and vice versa. The proposed framework predicts each grid cell which is represented by x-coordinate t_x^p, y-coordinate t_y^p, width t_w^p, and height t_h^p of a bounding box, and ship confidence level t_o^p . The ensemble YOLO framework detected bounding boxes' information are presented as follows [25]:

$$\begin{aligned} b_x^p &= \sigma\left(t_x^p\right) + v_x, \\ b_y^p &= \sigma\left(t_y^p\right) + v_y, \\ b_w^p &= p_w^a e^{t_w^p}, \\ b_h^p &= p_h^a e^{t_h^p}, \end{aligned} \tag{2}$$

$$t_o^p = \frac{S_{gth} \cap S_{bbox}}{S_{gth} \cup S_{bbox}}, \tag{3}$$

where b_x^p and b_y^p are the ensemble framework detection result, b_w^p and b_h^p are the width and height of the detected box, respectively. The v_x and v_y are the horizontal and vertical distances between the grid cell center point to the input image top left corner, respectively. The p_w^a and p_h^a are the weight matrices on the width and height, respectively. The $\sigma\left(t_y^p\right)$ and $\sigma\left(t_x^p\right)$ are the sigmoid function outputs based on the predicted bounding box information of the grid cell. The parameter S_{gth} is the square of ground truth box for the target ship in a maritime image, and the S_{bbox} is the counterpart of the detected bounding box. The symbol $\cap$ depicts the overlapping operation, while the $\cup$ is the union operator.

3.3. Ship Detection and Recognition. The outputs of the previous step generate a lot of bounding boxes with many of them being false alarms, and thus we employ the K-means



Ship behavior analysis

Anchor box determination and ship detection

FIGURE 2: Ship behavior analysis for the proposed framework workflow.

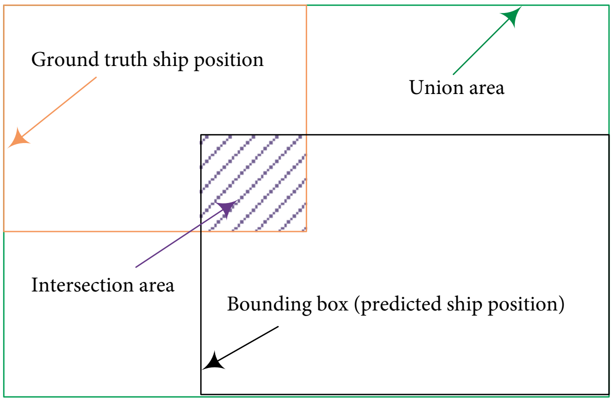

Ground truth ship position
Union area
Intersection area
Bounding box (predicted ship position)

FIGURE 3: Sketch map of calculating IOU.

clustering method to obtain anchor boxes by suppressing the interference boxes [26]. More specifically, the K-means algorithm randomly selects K bounding boxes as the initial clustering center, and the distance between each bounding box to the cluster center is calculated. The conventional K-means algorithm employs the Euclidean distance as the evaluation measurement, and larger boxes (ships with larger imaging size) contain higher detection false alarms than the smaller ones. To mitigate the negative influence, we employ the distance in Eq. (4) for the box clustering rule in the K-means algorithm. The K-means algorithm obtains the width (height) ratio r_w (r_h) by dividing the bounding box width (height) with the image width, which are shown as Equations (5) and (6), respectively. Then, the K-means algorithm clusters the (r_w, r_h) into K classes, and the each cluster center is considered as the anchor box. Readers are suggested to refer to [27] for more details about K-means algorithm. In the training procedure, the anchor box with maximum IOU is considered as detected ship positions. We employ the binary cross-entropy cost function to determine the class of the box (see Eq. (7)).

$$d(bbox, center) = 1 - \mathrm{IOU}(bbox, center), \tag{4}$$

$$r_w = \frac{w_{bbox}}{w_{img}}, \tag{5}$$

$$r_h = \frac{h_{bbox}}{h_{img}}, \tag{6}$$

$$L_y = -\frac{1}{N}\sum_{z=1}^{N} y_z * \log\left(p\left(y_z\right)\right) + \left(1 - y_z\right) * \log\left(1 - p\left(y_z\right)\right), \tag{7}$$

where $d(bbox, center)$ is the distance between the bounding box and the center box, and $\mathrm{IOU}(bbox, center)$ is the intersection over union between the two boxes. The w_{bbox} and w_{img} are the widths of the bounding box and image, respectively. The h_{bbox} and h_{img} are the height of the bounding box and image, respectively. The N is the bounding box number, y_z is

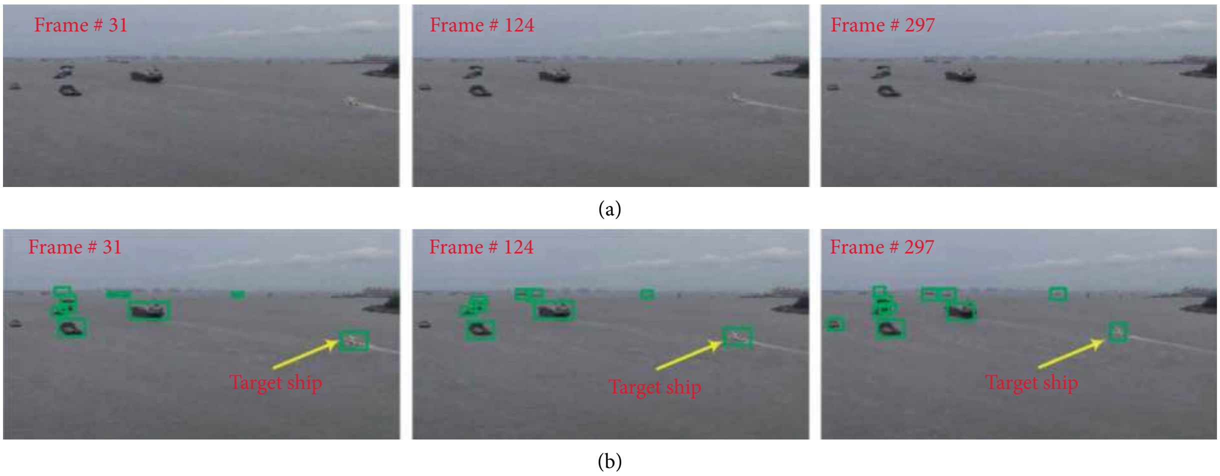


Figure 4: Ship detection results on typical frames of case-1. (a) Initial ship images, (b) Ship detection results.

the zth detected bounding box in the image, and $p(y_z)$ shows the predicted probability for the bounding box belonging to the class.

3.4. Ship Behavior Analysis. After obtaining ship locations in each maritime image via the previous steps, connecting positions from neighboring images of the same ship is very crucial for the ship behavior analysis. Considering that ships are rigid objects, and bounding box on the same ship should share the same motion, we determine the ship location with the spatial-temporal constraints based method. Note that each bounding box is presented by its center point. Assume the ship location in the image i is presented as l_{xi}, l_{yi}, θ_i, where the l_{xi} and l_{yi} denote the x and y-coordinate of the bounding box center point in the ith maritime image, and the θ_i is the ship sailing direction in the same image. We consider the positions in the neighboring ith and the $(i+1)$th images belong to the same ship when the group constraints in Eq. (8) are met. In addition, the ship behavior analysis is implemented by analyzing variation tendency of θ_i in the consecutive images. More specifically, the variation between neighboring θ_i with a decreasing trend shows that the ship is turning left, and an increasing tendency implying the ship is turning right. The ship is considered as sailing straight when the θ_i variation keeps in slight fluctuations (see Eqs. (9) and (10)).

$$\begin{gathered} d(x) = \left|l_{xi} - l_{x(i+1)}\right| < \alpha, \\ d(y) = \left|l_{yi} - l_{y(i+1)}\right| < \beta, \\ d(s) = \sqrt{\left(\left|l_{xi} - l_{x(i+1)}\right|\right)^2 + \left(\left|l_{yi} - l_{y(i+1)}\right|\right)^2} < \gamma, \end{gathered} \tag{8}$$

$$\theta_i = \frac{\left(l_{yi} - l_{y(i+1)}\right)}{\left(l_{xi} - l_{x(i+1)}\right)}, \tag{9}$$

$$d(\theta) = \left|\theta_i - \theta_{(i+1)}\right| < \varphi, \tag{10}$$

where $d(x)$ and $d(y)$ determine the ship moving distance in the x and y direction, respectively. The parameters α and β are thresholds of determining the maximum pixel distance in the x and y axis. The parameter $d(s)$ is ship displacement between neighboring frames, and the threshold γ determines maximal neighboring ship moving displacement. The parameter d(θ) indicates ship sailing direction variation tendency, and the parameter φ is the corresponding threshold.

3.5. Detection Goodness Measurements. To evaluate the proposed framework detection performance, we compare ship detection results with manually labeled ship positions (i.e., ground truth data) in each maritime image. Following the rules in the previous studies [28], two statistical indicators are employed to demonstrate the framework performance, which are recall rate (R_r) and precision rate (P_r). The indicator R_r demonstrates the miss-detection performance of the proposed framework. More specifically, the lower value of the indicator R_r implies that fewer objects in the maritime images (such as ships, buoys, etc.) are miss-detected by the framework. The parameter P_r shows the precision detection rate for the proposed framework. More specifically, the larger P_r demonstrates less detection error, and thus indicting better detection performance for our proposed framework. The definitions of R_r and P_r indicators are shown as follows:

$$R_r = \frac{T_t}{T_t + T_f}, \tag{11}$$

$$P_r = \frac{T_t}{T_t + f_T}, \tag{12}$$

where T_t is the number of ships positively detected by the proposed framework. The parameter T_f is the miss-detected ship number and the f_T is false-detected ship number.

4. Experiments

4.1. Experimental Settings. The proposed framework is applied to the two collected maritime video clips, which have been

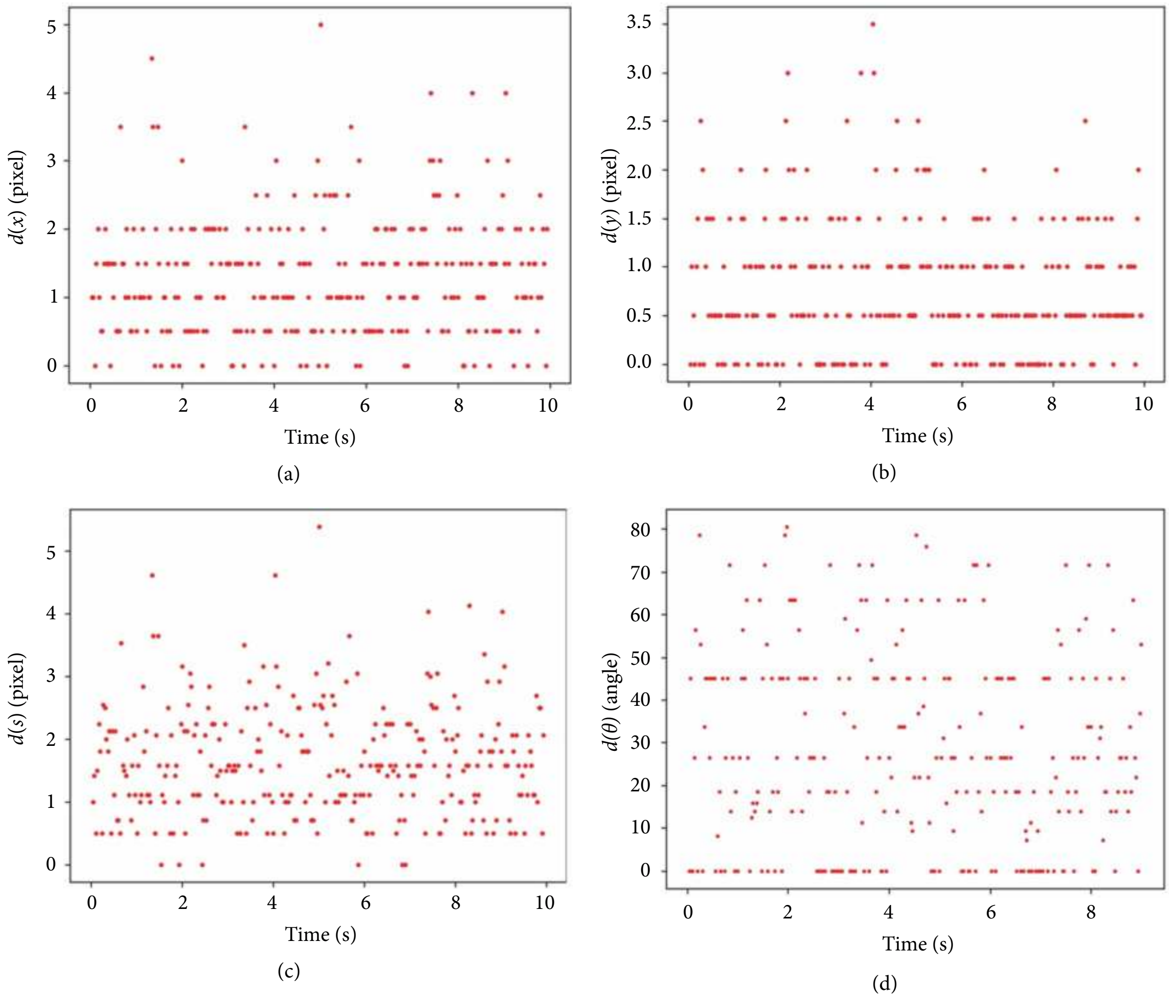
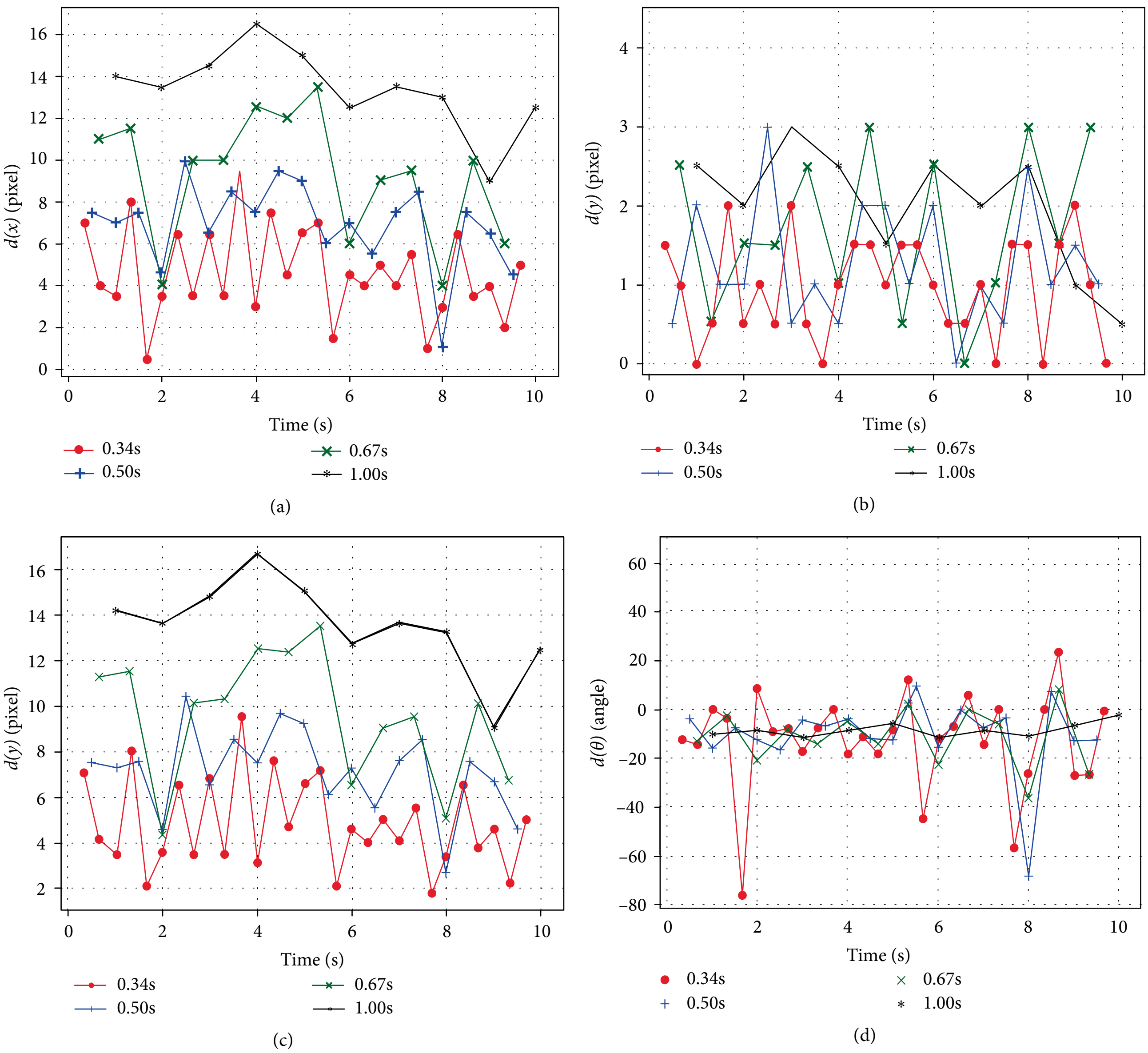


Figure 5: Ship kinematic data variation tendency at time interval 0.03 s; (a) ship position variation tendency at x-axis; (b) ship position variation tendency at y-axis; (c) ship moving displacement variation; (d) ship sailing direction variation.

detailed described in the above sections. Our framework was developed on the Win10 OS with 32G RAM and 3.2 GHz CPU. The GPU version is NVIDIA GeForce GTX 1080 Ti, which contains 11 GB RAM. Besides, the simulation platform is Tensorflow implemented on the Python (3.7 version). Though different versions of YOLO detector are public accessible, we developed our framework based on YOLO v3 for the purpose of accurate ship detection and behavior analysis. We set the anchor box number to 3 considering the tradeoff between time consumption and detection performance. The cluster number was set to 3 considering that ship, aiding facility (buoy, light beacon, etc.), and the obstacles (rock, bridges, etc.) are the three common types of objects in maritime images. The more detailed YOLO model setups are suggested to refer to [24].

4.2. Experimental Results for Case-1. The outputs of the proposed framework are presented in detail to reveal the model results. The ships in each maritime image were detected by the ensemble YOLO model, and detection samples were shown in Figure 4. It is observed that the majority of ships in each image was successfully detected, and partial small-size ships were miss-detected by the proposed framework (i.e., ships sailing close to the water-sky-line maybe miss-detected). The main reason is that we did not mark out all the small-size ships in the training images, and thus the proposed model was not trained by such ship samples. More specifically, considering the small size ships cannot be 100% correctly recognized by our naked eyes, we only marked out the ships with discernible visual features (contours, edges, etc.) in the training dataset when we fine-tuned the YOLO model settings in the proposed framework.

Table 1 presents the ship detection performance of the proposed framework. Both of the R_r and P_r indicators are higher than 90%, indicating that more than 90% ships in the case-1 were successfully detected by our proposed framework. More specifically, the R_r value in the case-1 is 93.52%, which implies that less than 7% ships are failed to be detected by the proposed framework. It is found that the miss-detected ships are the small size ships which are quite far from the camera shooting area. The P_r value is 94.16% which indicates that more than 94% detected ships in case-1 are positive results (i.e., over 94% detected ships are the true targets). After carefully checking the framework detection results, we found the false-detected ships mainly consisted of navigation aiding facilities.



(a)

(b)

(c)

(d)

Figure 6: Parameter variation tendency with different time intervals. (a) $d(x)$ variation at different time intervals. (b) $d(y)$ variation at different time intervals. (c) $d(s)$ variation at different time intervals. (d) $d(\theta)$ variation at different time intervals.

Table 1: Ship detection performance for case-1 and case-2.

	R_r (%)	P_r (%)
Case-1	93.52	94.16
Case-2	92.17	93.65
Average	92.85	93.91

Both R_r and P_r values have shown that our framework have obtained satisfied ship detection performance, and can provide accurate ship positions for the ship behavior analysis.

With the obtained ship positions in the previous steps, we then connected positions from same ship based on the criterion in Equation (8). It is worth noting that parameter setup is crucial for identifying consecutive ship positions. We first estimated the parameter settings per frame (i.e., time interval was set to 0.03 s). As shown in each subplot in Figure 5, each parameter in Equation (8) (i.e., $d(x)$, $d(y)$, $d(s)$ and $d(\theta)$) presents random and even unrealistic variation tendency. The main reason is that when ship moving distance is very small within 0.03 s (as ship speed is very slow), and thus ship kinematic information in images is very sensitive to the small measurement imaging error. According to the thumb of rule, we divided the sample rate of collecting ship positions into 0.34 s, 0.50 s, 0.67 s, and 1 s, and the parameter distributions at each sample rate were shown in Figure 6. It is noticed that larger sample rate provides smoother ship position variation, and obtains more reasonable results. For instance, at the time interval 0.34 s, the maximum $d(x)$ value reaches 10.0 pixels, which was nine-fold higher than that of the minimum $d(x)$ (see Figure 6(a)). But, the $d(x)$ ranges from 9 to 16 pixels (at time interval 1 s), with fluctuation magnitude sharply decreases to 43.75%. We can observe similar findings from $d(y)$ and $d(s)$ variations as shown in Figures 6(b) and 6(c), respectively.

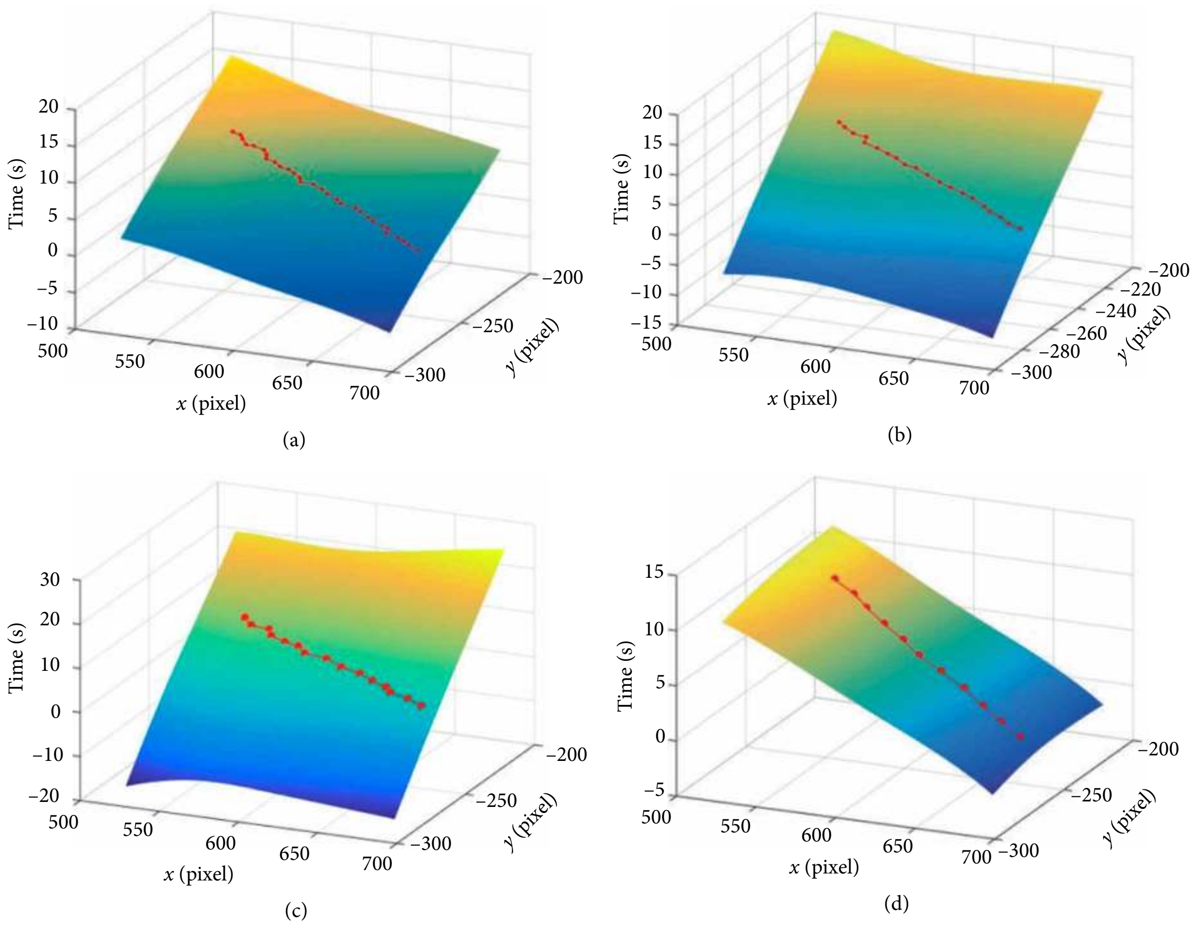


Figure 7: Ship spatial-temporal trajectory map at different time intervals in case-1. (a) Ship spatial-temporal trajectory map at time interval 0.34 s, (b) Ship spatial-temporal trajectory map at time interval 0.50 s, (c) Ship spatial-temporal trajectory map at time interval 0.67 s, (d) Ship spatial-temporal trajectory map at time interval 1.00 s.

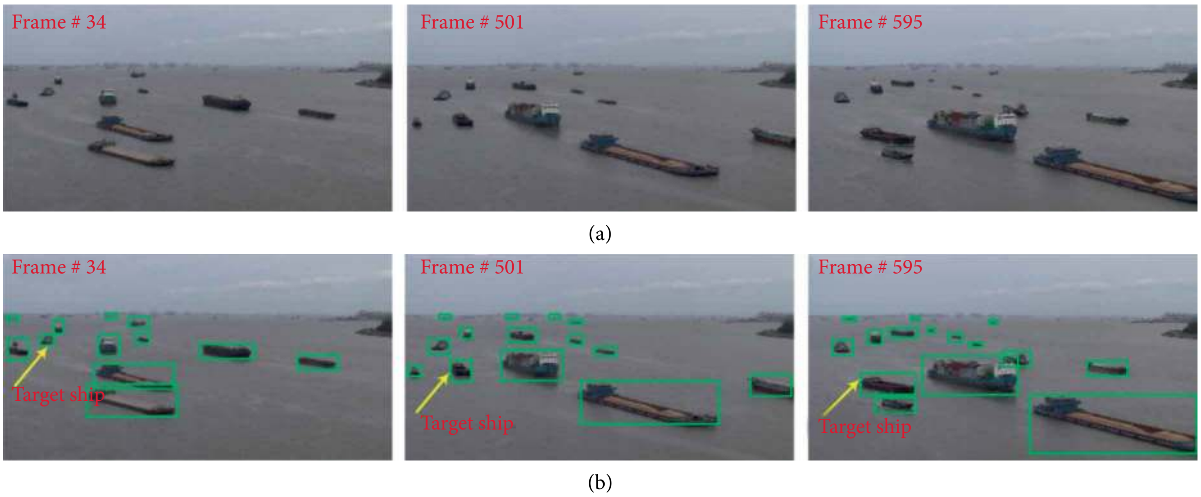


Figure 8: Ship detection results on typical frames of case-2. (a) Initial ship images, (b) Ship detection results.

The $d(\theta)$ variation at different sample rates has confirmed the variation tendency of the other parameters (see Figure 6(d)). In fact, the $d(\theta)$ indicates ship sailing direction variation in time domain, and the neighboring ship sailing direction is supposed to be smoothly changed considering ship displacement in neighboring frames is small. It is observed that 0.34 s time interval leads to sharp ship sailing angle variation, which is mainly caused by minor error of ship position measurement. As shown in the Figure 6(d), $d(\theta)$ variation at the 0.5 s time interval fluctuation is less sharply as that of the 0.34 s. However, the $d(\theta)$ variation at the 1 s presents a smooth variation, and indeed the unrealistic ship moving status has been successfully

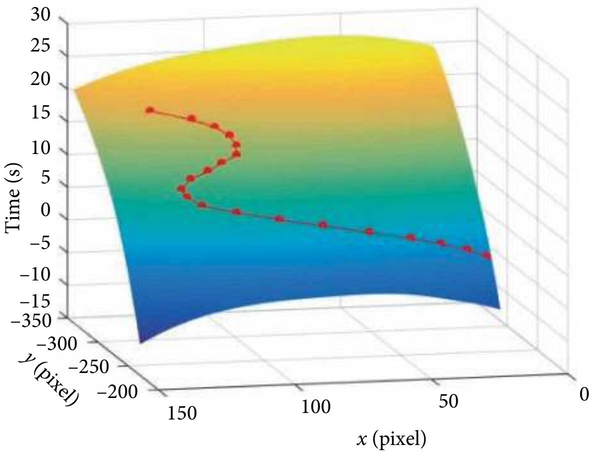


Figure 9: Ship spatial-temporal trajectory map in case-2.

suppressed. The results in Figure 6 suggest that a larger data sample rate leads to more reasonable results in our model, and thus 1 s was set to the default value (which is applied in following section without specific illustrations). The thresholds used for recognizing ship behaviors are set as follows: α is set to 16.5, β is set to 2.5, γ is set to 16.7, and the φ is set to 12.

To sum up, we have tested varied parameter setups for the purpose of obtaining optimal model performance following the rules in previous studies [24, 25, 29]. It is observed that ship detection results are very robust to the YOLO parameter variations. But, the time span is crucial for accurately measuring ship moving displacements and analyzing ship behaviors, and thus the sensitivity analysis was conducted to determine the optimal time interval (under 0.03 s, 0.34 s, 0.50 s, 0.67 s, 1 s). We did not compare our framework training results with the pretrained weights trained by the ship samples in public datasets (COCO, ImageNet, etc. [30, 31]). The main reason is that ship training and detection challenges in the collected maritime videos are significantly different from the previous public benchmarks (i.e., ships in the COCO and ImageNet datasets are not severely contaminated by environments).

The target ship (i.e., the white ship shown in Figure 4) trajectories in spatial-temporal map at different time intervals are shown in Figure 7. It is observed that several ship trajectory points overlapped when the time interval is small, due to that ship moving displacement is very small at a small time interval. More specifically, we can only recognize ship sailing direction with moving straight-forward tendency (see Figures 7(a)–7(c)). The ship behavior can be clearly recognized as moving straight when the sample rate becomes larger, which is obviously confirmed when the time interval it arrives at is 1 s (see Figure 7(d)). After checking the initial maritime video clip, we found the white ship is coast guard ship which was monitoring the traffic on-site at a high speed.

4.3. Experimental Results for Case-2. The proposed framework was also applied on the case-2 where ships have different travelling behaviors. It is found that the ship with clear visual features in maritime images was successfully detected by our method. But, the ship detection performance under case-2 is not as good as that of the case-1, as the P_r and R_r statistical indicators were both lower than the counterparts in case-1 (see Table 1). More specifically, the P_r of case-2 is 92.17%, which is 1.46% lower than that of the case-1. After carefully checking the ship detection results in each frame (detection samples are shown in Figure 8), we found that many ships were overlapped in the maritime sequences, and thus the proposed framework cannot accurately detect ship positions. More specifically, some ships may be labeled by a larger or smaller bounding box in ship overlapping area, which cannot be successfully detected by our naked eyes. The target ship spatial-temporal map was shown in Figure 9, which have shown varied ship behavior patterns. More specifically, the ship behaviors can be divided into three stages, which are moving straight, turning right and finally turning left. We have checked the initial maritime video in case-2, and it is observed that a ship following the ship in the same channel moved very fast which may trigger traffic accident. According to the rule in the international regulation for preventing collision at sea [32], the target ship is supposed to take initiative activities to avoid the accident. Thus, the target ship changed her sailing direction toward left waterway, and provided wider navigation area for the following ship. After that, the target ship navigated straight for the purpose of surpassing the neighboring ships. By obeying the maritime traffic separation law, the target ship changed her sailing direction towards starboard to move along the channel.

It can be observed that the maritime video data supports us to exploit the on-spot microscopic kinematic ship traffic information (ship platoon speeds, ship maneuvering directions, distance to closest point of approach, time of closest point of approach) which cannot be easily obtained from AIS data [33, 34]. The on-duty maritime officials' professional level affects the maritime monitoring performance as they are assumed to ensure maritime traffic safety by consistently watching at the real-time on-spot maritime surveillance videos. More specifically, the maritime accidents may happen when staff fails to send out early-warning alert on potential ship collision risks (due to staff fatigue, careless in work). The automatic video processing based methods can largely reduce the possibility of such types of maritime accidents. With the help of proposed automatic video processing technique, we can automatically exploit high-resolution microscopic ship platoon kinematic information (moving speed, sailing direction, etc.) which benefits smart shipping development.

5. Conclusion

Ship behavior recognition is crucial for the intelligent shipping development, which needs to overcome many environmental challenges at different ship navigation situations. We proposed an ensemble YOLO based framework to detect ships from maritime images, and accurately recognize ship behaviors in consecutive frames. The framework was implemented in four steps, which are ship feature exploration, bounding box generation, ship position determination, and ship behavior recognition. The ship feature exploration step aimed for extracting multi-scale distinct ship features based on YOLO detection model. The second step generates several bounding boxes which are considered as potential ship positions. The third step determined true ship positions by applying geometry theory. The fourth step connected same ship positions in consecutive images and ship spatial-temporal behaviors were recognized according to variation tendency of group constraints. We have implemented our framework on two typical maritime situations with different ship sailing behaviors. More specifically, the first video clip involved with ship moving-straight situation, and second video is relevant with ship turning behavior. From the perspective of recall and precision rates, our proposed framework achieved satisfied performance, with the average R_r and P_r obtained 92.85% and 93.91% in the ship detection procedure. The spatial-temporal map provided us clear ship behavior results, which helped us learn the ship historical travel patterns, predict ship future travel trajectories, and thus early warning measurements can be implemented to ensure maritime traffic safety.

Though our method has achieved satisfied performance on ship behavior analysis, some remaining research work can be done to further improve our method performance. First, we have collected high-resolution ship positions with the customized YOLO detector (e.g., by feeding the collected ship videos as training dataset). The obtained ship position data and collected ship videos support us deeply explore on the topics of maritime traffic flow analysis (ship speed variation pattern, traffic density distribution, etc.), traffic safety evaluation (i.e., inshore high-risk channel area identification, early-warning on ship collision, etc.), reducing ship platoon fuel consumption, etc. [33, 34]. Second, the two test videos were collected at good weather conditions (without obvious interference from storm, strong wind, fog, etc.) Analyzing ship behaviors at varied maritime videos shot at extreme weather conditions can be an interesting expansion in future. Third, we have not tested our proposed framework on background vibrated maritime videos, which impose additional challenge of connecting adjacent ship positions in consecutive frames. Fourth, verifying and testing our model performance with different training ship images and weights can be considered as a potential exploration in future. Last but not the least, considering ships far away from the maritime images may impose trivial interference, we can enhance our model robustness by introducing saliency based model focusing on extracting ship information at regions of interest in maritime surveillance videos.

Data Availability

Readers can access our data in the findings by sending an email to Xinqiang Chen (chenxinqiang@stu.shmtu.edu.cn).

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China (51579143, 51509167), Shanghai Committee of Science and Technology China (18040501700, 18295801100, and 17595810300).

References

[1] J. Wang, C. Zhu, Y. Zhou, and W. Zhang, “Vessel spatio-temporal knowledge discovery with AIS trajectories using co-clustering,” *The Journal of Navigation*, vol. 70, no. 6, pp. 1383–1400, 2017.

[2] H. Li, J. Liu, K. Wu, Z. Yang, R. W. Liu, and N. Xiong, “Spatio-temporal vessel trajectory clustering based on data mapping and density,” *IEEE Access*, vol. 6, pp. 58939–58954, 2018.

[3] L. Zhao and G. Shi, “A trajectory clustering method based on Douglas-Peucker compression and density for marine traffic pattern recognition,” *Ocean Engineering*, vol. 172, pp. 456–467, 2019.

[4] V. F. Arguedas, G. Pallotta, and M. Vespe, “Maritime traffic networks: from historical positioning data to unsupervised maritime traffic monitoring,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 3, pp. 722–732, 2018.

[5] W. Zhang, F. Goerlandt, J. Montewka, and P. Kujala, “A method for detecting possible near miss ship collisions from AIS data,” *Ocean Engineering*, vol. 107, pp. 60–69, 2015.

[6] W. Zhang, F. Goerlandt, P. Kujala, and Y.nhai Wang, “An advanced method for detecting possible near miss ship collisions from AIS data,” *Ocean Engineering*, vol. 124, pp. 141–156, 2016.

[7] R. J. Bye and A. L. Aalberg, “Maritime navigation accidents and risk indicators: an exploratory statistical analysis using AIS data and accident reports,” *Reliability Engineering & System Safety*, vol. 176, pp. 174–186, 2018.

[8] F. Mazzarella, M. Vespe, and C. Santamaria, “SAR ship detection and self-reporting data fusion based on traffic knowledge,” *IEEE Geoscience and Remote Sensing Letters*, vol. 12, no. 8, pp. 1685–1689, 2015.

[9] B. Habtemariam, R. Tharmarasa, M. McDonald, and T. Kirubarajan, “Measurement level AIS/radar fusion,” *Signal Processing*, vol. 106, pp. 348–357, 2015.

[10] X. Chen, S. Wang, C. Shi, H. Wu, J. Zhao, and J. Fu, “Robust ship tracking via multi-view learning and sparse representation,” *Journal of Navigation*, vol. 72, no. 1, pp. 176–192, 2019.

[11] R. W. Jansen, R. G. Raj, L. Rosenberg, and M. A. Sletten, “Practical multichannel SAR imaging in the maritime environment,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 7, pp. 4025–4036, 2018.

[12] D. K. Prasad, D. Rajan, L. Rachmawati, E. Rajabally, and C. Quek, “Video processing from electro-optical sensors for object detection and tracking in a maritime environment: a survey,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 8, pp. 1993–2016, 2017.

[13] B. Wang, Y. Motai, L. Dong, and W. Xu, “Detecting infrared maritime targets overwhelmed in sun glitters by antijitter spatiotemporal saliency,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 7, pp. 5159–5173, 2019.

[14] A. Valsamis, K. Tserpes, D. Zissis, D. Anagnostopoulos, and T. Varvarigou, “Employing traditional machine learning algorithms for big data streams analysis: the case of object trajectory prediction,” *Journal of Systems and Software*, vol. 127, pp. 249–257, 2017.

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

[16] Y. Liu, L. Yao, W. Xiong, and Z. Zhou, "GF-4 satellite and automatic identification system data fusion for ship tracking," *IEEE Geoscience and Remote Sensing Letters*, vol. 16, no. 2, pp. 281–285, 2019.

[17] X. Kang, B. Song, J. Guo, X. Du, and M. Guizani, "A self-selective correlation ship tracking method for smart ocean systems," *Sensors*, vol. 19, no. 4, p. 821, 2019.

[18] J. Woo, J. Park, C. Yu, and N. Kim, "Dynamic model identification of unmanned surface vehicles using deep learning network," *Applied Ocean Research*, vol. 78, pp. 123–133, 2018.

[19] M. Gao, G. Shi, and S. Li, "Online prediction of ship behavior with automatic identification system sensor data using bidirectional long short-term memory recurrent neural network," *Sensors*, vol. 18, no. 12, p. 4211, 2018.

[20] L. Zhang, L. Zhang, and B. Du, "Deep learning for remote sensing data: a technical tutorial on the state of the art," *IEEE Geoscience and Remote Sensing Magazine*, vol. 4, no. 2, pp. 22–40, 2016.

[21] M. Kang, K. Ji, X. Leng, and Z. Lin, "Contextual region-based convolutional neural network with multilayer fusion for SAR ship detection," *Remote Sensing*, vol. 9, no. 8, p. 860, 2017.

[22] H. Lin, Z. Shi, and Z. Zou, "Fully convolutional network with task partitioning for inshore ship detection in optical remote sensing images," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 10, pp. 1665–1669, 2017.

[23] X. Chen, H. Chen, Y. Yang et al., "Detecting ships from coastal surveillance videos with a canny-gaussian-morphology framework," submitted to IEEE ACCESS.

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

[25] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, IEEELas Vegas, NV, USA, 2016.

[26] D. T. Nguyen, T. N. Nguyen, H. Kim, and H.-J. Lee, "A high-throughput and power-efficient FPGA implementation of YOLO CNN for object detection," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 27, no. 8, pp. 1861–1873, 2019.

[27] G. Gan and M. K.-P. Ng, "K-means clustering with outlier removal," *Pattern Recognition Letters*, vol. 90, pp. 8–14, 2017.

[28] X. Chen, Z. Li, Y. Wang et al., "Anomaly detection and cleaning of highway elevation data from google earth using ensemble empirical mode decomposition," *Journal of Transportation Engineering, Part A: Systems*, vol. 144, no. 5, p. 04018015, 2018.

[29] M. J. Shafiee, B. Chywl, F. Li, and A. Wong, "Fast YOLO: a fast you only look once system for real-time embedded object detection in video," 2017, https://arxiv.org/abs/1709.05943.

[30] T.-Y. Lin, M. Maire, S. Belongie et al., "Microsoft coco: common objects in context," in *European Conference on Computer Vision*, Springer, Cham, 2014.

[31] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: a large-scale hierarchical image database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, Miami, FL, USA, 2009.

[32] K. Saravanan, S. Aswini, and R. Kumar, "How to prevent maritime border collision for fisheries?-A design of real-time automatic identification system," *Earth Science Informatics*, vol. 12, no. 2, pp. 241–252, 2019.

[33] A. D. May, *Traffic Flow Fundamentals*, Prentice Hall, Upper Saddle River, NJ, USA, 1990.

[34] R. P. Roess, E. S. Prassas, and W. R. McShane, *Traffic Engineering*, Pearson/Prentice Hall, Upper Saddle River, NJ, USA, 2004.