




RESEARCH ARTICLE

Ship detection from coastal surveillance videos via an ensemble Canny-Gaussian-morphology framework

Xinqiang Chen,¹  Jun Ling,¹ Shengzheng Wang,² Yongsheng Yang,¹ Lijuan Luo,^{3*} and Ying Yan⁴

¹Institute of Logistics Science and Engineering, Shanghai Maritime University, Shanghai, PR China.

²Merchant Marine College, Shanghai Maritime University, Shanghai, PR China.

³AI and Data Science Application Center, School of Business and Management, Shanghai International Studies University, Shanghai, China.

⁴College of Transportation, Chang'an University, Xi'an, China.

*Corresponding author. E-mail: luolijuan@shisu.edu.cn

Received: 17 March 2020; **Accepted:** 28 May 2021; **First published online:** 9 July 2021

Keywords: ship detection, ensemble Canny-Gaussian-morphology framework, coastal surveillance video, smart ship

Abstract

Coastal surveillance video helps officials to obtain on-site visual information on maritime traffic situations, which benefits building up the maritime transportation detection infrastructure. The previous ship detection methods focused on detecting distant small ships in maritime videos, with less attention paid to the task of ship detection from coastal surveillance video. To address this challenge, a novel framework is proposed to detect ships from coastal maritime images in three typical traffic situations in three consecutive steps. First the Canny detector is introduced to determine the potential ship edges in each maritime frame. Then, the self-adaptive Gaussian descriptor is employed to accurately rule out noisy edges. Finally, the morphology operator is developed to link the detected separated edges to connected ship contours. The model's performance is tested under three typical maritime traffic situations. The experimental results show that the proposed ship detector achieved satisfactory performance (in terms of precision, accuracy and time cost) compared with other state-of-the-art algorithms. The findings of the study offer the potential of providing real-time visual traffic information to maritime regulators, which is crucial for the development of intelligent maritime transportation.

1. Introduction

Maritime surveillance video (MSV) provides visual and on-site information on waterway traffic to onshore traffic management departments, and thus helps officials to employ timely control measures to enhance maritime traffic safety (Wang et al., 2020; Yu et al., 2021). The ship detection results offer critical visual information to maritime authorities to determine law-breaking behaviours of traffic participants, such as overloading, violation of routing system, etc. The previous ship detection models can be classified into three types: radar-related techniques, visible light-based models and infrared-related methods. The first type of ship detection model, using radar, employs the echoes from objects to find ships in the monitoring area (based on the Doppler effect). Pappas et al. (2018) employed radar super-pixel to replace rectangular sliding windows for the purpose of obtaining better ship detection results from synthetic aperture radar images. They found that the performance of radar-based ship detection methods is easily degraded by sea spikes and waves, which are likely to happen in extreme weather conditions (Wang et al., 2017a, 2017b; Nova et al., 2020). Maritime surface wave radar is commonly used to identify ships around busy coastal areas by recognising ship echoes. More specifically, different ships' echoes

Table 1. Comparison of performance of ship detection techniques.

	Radar	Visible light	Infrared technique
Ship detection mechanism	Ship echoes	Visual features	Thermal imaging
Main advantage	Visibility	Informative	Large detection area
	insensitivity	spatial-temporal data	
Disadvantage/interference	Wooden ships	Extreme weather	Water clutters

vary in the maritime surface wave radar system, which helps ships' crews to efficiently determine on-site maritime traffic kinematic information. To solve the problem, Le Caillec et al. (2018) compared the performance of different temporal-spatial algorithms with different high-frequency surface wave radars, and proposed an efficient ensemble ship detection framework. Many studies have been implemented to detect ships, using methods such as removal of sea clutter, salient yet sparse ship detection, ship wake reduction, etc. (Yan et al., 2019; Zhao et al., 2019; Graziano, 2020).

Visible light-related models, the second type of ship detection techniques, employ edge operators to detect intrinsic ship contours from maritime surveillance videos (shot by cameras installed on inshore buildings and ships) (Huang et al., 2020; Liu et al., 2020b). The visible light-based techniques have shown many successes in different fields, such as vehicle detection and tracking, pedestrian detection and behaviour analysis (Liu and Gao, 2020; Liu et al., 2020a, 2020b), and thus demonstrated their potential in ship detection applications. For real-time and accurate ship detection, Shao et al. (2019) developed a saliency-aware convolution neural network-based model to extract discriminative ship features in coastal maritime images. Shafer and Harguess (2015) suppressed ship detection outliers by jointly combining the dictionary learning and group-sparsity model. Kim et al. (2018) proposed a faster region-based convolutional neural network framework to recognise ships in various maritime meteorological environments. Many studies have focused on automatic detection of ships from maritime images by exploiting both conventional hand-crafted feature detection and deep learning models (Zhang et al., 2017; Chen et al., 2020c; Ren et al., 2020).

Infrared technique is the third widely used ship detection framework. It is particularly good at detecting ships under insufficient illumination (i.e., night time). Zhu and Xue (2015) and Mumtaz et al. (2016) proposed frameworks based on salient features (including both global and local features) to discriminate ships from the non-ship background (waves, ships' wakes, etc.). However, the infrared-based models are very sensitive to object edge variations, and the detection performance may be sharply reduced (Wang et al., 2017a, 2017b). In addition, the salt-and-pepper noise (caused by sharp image intensity variation) cannot be eliminated from infrared maritime images, and this may severely deteriorate the detection performance of infrared-based methods. To address the issue, Min et al. (2018) proposed a convolutional neural network-based framework to detect ships efficiently from infrared maritime images at multi-level resolutions. Lu et al. (2006) developed a small marine target detection framework combining an edge detector and a median filter, which showed satisfactory performance in 700 infrared images. Xie et al. (2017) proposed a novel framework for ship detection by fusing spectral and thermal features, which suppressed cloud interference with a semantic segmentation model. It is not easy to obtain satisfactory ship detection results considering that each type of detection technique has both pros and cons, which are summarised in Table 1. More specifically, the radar-related techniques employ echoes to identify ships at sea; they can obtain robust detection performance under low visibility conditions, but may fail to detect wooden ships due to their faint ship echoes (Chen et al., 2020b). The visible light-related techniques implement the detection task by determining a ship's visual features from maritime images. We can exploit significant informative spatial-temporal data with the visible light-based models. Their performance is easily hindered, however, by extreme weather conditions (Biondi, 2017; Zhang et al., 2018). The infrared techniques obtain ship detection results via the principle of thermal imaging. The infrared technique has a wide coverage area (i.e., it can detect ships far away

from the infrared camera); the main disadvantage is that wave clutter may significantly deteriorate the technique's performance (Le Caillec et al., 2019).

The abovementioned advantages are the motivation to propose an ensemble Canny-Gaussian-morphology framework for solving the ship detection challenge from inland surveillance videos. The contributions of this study can be summarised as follows: (a) analysis of the cons and pros of previous studies on ship detection from varied maritime data sources (i.e., radar, infrared, and visible light images); (b) development of a novel Canny-Gaussian-morphology ship detection framework for the purpose of accurate and real-time ship detection, which involves the steps of ship edge extraction, negative ship edge suppression, and ship contour reconstruction; (c) testing of the proposed model performance under typical maritime traffic situations. More specifically, three maritime video clips (i.e., high and low traffic volume, strong wave interference) were collected with cameras installed on onshore buildings. For the purpose of model performance comparison, three ship detection models (Gaussian, support vector machine [SVM] and mask-RCNN [regions with CNN features] models) are implemented. The experimental results show that the model achieved satisfactory recall rate and precision rate with much lower time consumption than existing methods. The findings of the study can provide visual and on-the-spot maritime traffic information (traffic flow, traffic density, ship behaviour, etc.) to maritime authorities and officers-on-board. Thus, the maritime authorities can release in-time traffic control measures, and ship crew (especially the on-duty crew) can undertake early-warning sailing activities to avoid maritime traffic accident, thus enhancing maritime traffic safety.

The specialised terms used in this study are listed and briefly explained for purpose of simplicity and clarity. A 'frame' means one image from a MSV. The 'frame rate' is the number of images that are scheduled to be displayed in one second, 'frames per second'. 'Image resolution' (abbreviated as resolution) refers to the amount of pixels in a maritime frame, which is also identified as image width multiplied by height in the form of width \times height. The 'image noise' includes salt-and-pepper noises, video vibration, Gaussian noises, etc. 'Image intensity' means the grey value of a pixel, which will not exceed 255. 'Non-maximum suppression (NMS) mechanism' determines the local maximal values and discards non-maximal values. The 'eight-neighbour connectivity rule' determines a polygon with eight neighbouring points. The eight-neighbour connectivity rule was used to reconstruct ship contours in each frame. The 'structure element' of the morphology operator is used to remove small blobs and re-build ship contours. The term 'millisecond' is abbreviated as 'ms' in the study. Recall rate (Re) and precision rate (Pr) are the two statistical indicators for evaluating ship detection models.

2. Image data set

The Port of Shanghai is an important water–water transfer hub port in China; its throughput exceeded 40 million 20 ft equivalent units in 2018. Thus, it is important to enhance the management efficiency and safety of the port with the available techniques. Ship detection can accurately provide instantaneous traffic information and predict the traffic situation in the port surveillance area in advance, and thus is crucial for ensuring port efficiency. To that aim, maritime video clips were collected from cameras installed on the roof of a building located at Shanghai port. Note that the camera is robust against high-salt and high-humidity interference. Some typical frames of the collected videos are shown in [Figure 1](#). To help readers easily understand this study, the definitions are briefly introduced here of three terms: small ship, low traffic volume and high traffic volume. The term small ship is applied to the ships in each frame whose imaging size meets one or both of the following conditions: (a) the ship image size is less than 0.15% that of the current frame; (b) the imaging length or width of the ship is less than 13 pixels (Wei, 2009). Low traffic volume indicates that the number of non-small-ships in each frame is not larger than five, while high traffic volume indicates the number of non-small-ships in each frame is larger than ten.

Based on the above definitions, the test MSV clips are divided into two detection scenarios. The first detection scenario is to verify the performance of the proposed framework under different traffic states (low and high traffic volumes). The main challenge in the low traffic volume state is that ships in the



Figure 1. Typical frames of the collected image sequences for each ship detection scenario: (a) low traffic volume; (b) high traffic volume; (c) strong water clutter interference.

distance are quite similar to waters, and thus can be easily mis-detected, while in high traffic volume state ships are easily sheltered by neighbouring objects (ships, buoys, etc.). The second scenario is to test the model's performance under the interference of water ripples and ships' wakes. The first detection scenario contains two image sequences which are labelled as case-a and case-b. The frame rate of the case-a is 30 frames per second, and the resolution of each frame is 1280×720 . The frame rate and resolution for image sequences of case-b are the same as for case-a. There are 600 frames in case-a and 720 frames in case-b. The second detection scenario is labelled as case-c. The frame rate and resolution of case-c is identical with case-a, while the video length of case-c is 12 s.

3. Methodology

It is found that maritime images are composed of rigid and non-rigid objects. The rigid objects are the foreground objects (i.e., ships), and the non-rigid objects are the background objects (e.g., water, sky). Note that the contours of the rigid objects (i.e., ship contours) can be easily identified by edge descriptors. The background objects do not have obvious extractable contours due to the intrinsic imaging features of non-rigid objects. In that way, the edge detector will not find obvious edges from background objects in maritime images. Therefore, the proposed ensemble ship detection framework employs the Canny filter to detect ship contours from maritime images, and the self-adaptive Gaussian model is further introduced to suppress potential false positive ship detection edges. Moreover, the morphology model is used to connect the spitted ship edges (i.e., obtain a minimum bounding rectangle for each ship detection result). The flowchart for the proposed ship detection framework is shown in [Figure 2](#).

3.1. Ship contour extraction with Canny filter

The Canny detector is robust against wave-related imaging interference compared with the counterpart Prewitt, Sobel and Laplacian models (Canny, 1986; Chen et al., 2019). More specifically, weak ship contours can be successfully identified by the Canny detector, while the Prewitt, Sobel and Laplacian models may fail to extract the contours. The Canny detector is a multi-stage algorithm optimised for fast real-time edge detection. The advantages of using the Canny detector are as follows: (1) it is reliable for accurate detection of existent edges; (2) it produces candidate edges which are not dominant in their neighbourhood if the neighbourhoods are not considered to be edges – this is achieved by the NMS mechanism; (3) it is a non-convolutional method so it will run faster than other machine learning algorithms and does not require training data. Motivated by these advantages, in this study ship edges are determined from MSV images with the Canny detector, and more details are presented as follows.

3.1.1. Noise suppressing in ship images

Raw ship edges are obtained via the Canny filter by the steps of imaging noise suppression, gradient identification and non-maximum gradient removal. It is noted that light systematic interference may result in false alarm ship detection results. To address the issue, a two-dimensional Gaussian filter is employed to suppress potential imaging noise from maritime video clips with the help of convolution operation [see Equation (1)]. The Gaussian kernel plays the role of weighted mask when conducting

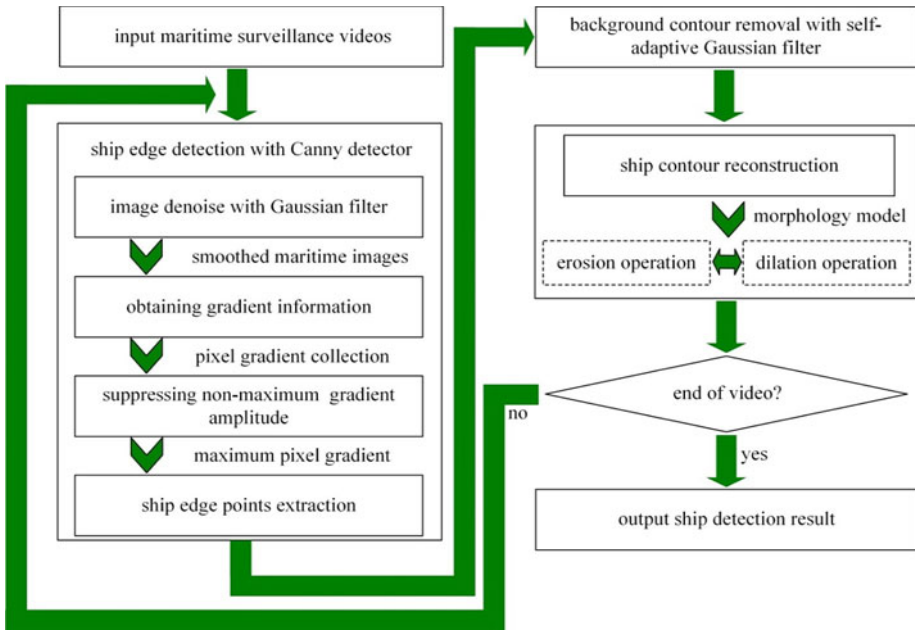


Figure 2. Overview of ship detection framework with the proposed Canny-Gaussian-morphology model.

convolution operation for suppressing noise in ship frames. More specifically, the convolution operation modifies the noise pixel intensity into normal by computing an aggregated value with adjacent pixels.

$$p(\omega, h) = \frac{1}{2\pi\sigma^2} e^{-\frac{(\omega^2+h^2)}{2\sigma^2}} \tag{1}$$

where ω is the distance to the x -axis, while h represents the distance to the y -axis. Parameter σ is the standard distribution in the Gaussian kernel.

3.1.2. Finding gradient information in each ship frame

After smoothing out noises in ship frames, the Canny filter computes gradient value and direction for every pixel in each MSV image. Two steps are implemented to obtain the gradient data: (a) constructing the convolution masks for the x and y directions. More specifically, the convolution mask for the x -direction is labelled as C_x , and C_y is the mask for the y -direction. The expressions of C_x and C_y are shown in Equations (2) and (3), respectively; (b) obtaining gradient value and direction for each pixel. After conducting the convolution operation on each pixel in each ship frame, we can obtain pixel gradient values and directions in both x - and y -axis through Equations (4)–(7). Note that the pixel gradient directions are denoted as angles; more details are provided by Lee et al. (2018) and Chen et al. (2020a).

$$C_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \tag{2}$$

$$C_y = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix} \tag{3}$$

$$p(i, j) = \frac{-I(i, j) + I(i, j + 2) - 2I(i + 1, j)}{2} + \frac{I(i + 1, j + 2) - I(i + 2, j) + I(i + 2, j + 2)}{2} \tag{4}$$

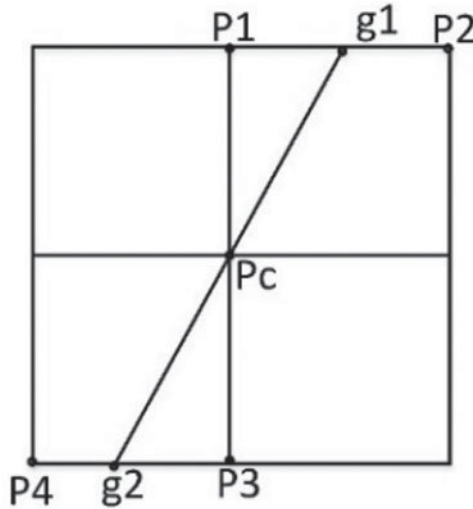


Figure 3. Edge point determination with the NMS mechanism.

$$Q(i, j) = \frac{-I(i, j) - I(i, j + 2) - 2I(i, j + 1)}{2} + \frac{2I(i + 2, j + 1) + I(i + 2, j) + I(i + 2, j + 2)}{2} \tag{5}$$

$$G(i, j) = \sqrt{P^2(i, j) + Q^2(i, j)} \tag{6}$$

$$\theta(i, j) = \arctan \frac{Q(i, j)}{P(i, j)} \tag{7}$$

where $I(i, j)$ is pixel intensity for point (i, j) . $P(i, j)$ and $Q(i, j)$ represent gradient values for x and y direction, respectively. $G(i, j)$ is gradient amplitude and θ is gradient direction. For simplicity, θ is rounded up to one of the following values ($0^\circ, 45^\circ, 90^\circ, 135^\circ$).

3.1.3. NMS mechanism for suppressing gradient amplitude

The outputs of the previous step provide a variety of candidates of edge points from potential ship contours. The NMS mechanism is employed to rule out false edge points and retain actual ship edges. As shown in Figure 3, point P_c neighbours are labelled as P_1, P_2, P_3 and P_4 for the purpose of identifying an edge point. More specifically, the points $P_1 (P_3)$ and $P_2 (P_4)$ are connected into a line which is denoted as $P_1P_2 (P_3P_4)$. The midpoint for the line $P_1P_2 (P_3P_4)$ is labelled as $g_1 (g_2)$. In that manner, we consider the line linking g_1 and g_2 as a potential gradient direction for the point P_c . Note that the point P_c is considered as an edge point when the gradient amplitudes of the points g_1 and g_2 are not larger than that of the P_c , and vice versa.

3.1.4. Extraction of ship edge points

The NMS procedure provides the collection of potential ship edge points, and the Canny filter employs the double threshold method to further smooth out false edge points, and obtain more accurate ship edges. Given two threshold values T_1 and T_2 ($T_1 > T_2$), T_1 is the predefined strong edge threshold and T_2 is the weak edge threshold. The pixel (i, j) is deemed to be a strong edge point when its gradient value G_P [see Equation (6)] is larger than T_1 . The pixel is considered as a weak edge point when G_P meets the condition $T_1 > G_P > T_2$. The obtained strong edge points are recognised as true edge points belonging to ship contours. Based on that, the strong and weak edge points are connected to form ship contours with the eight-neighbour connectivity rule (Ke et al., 2017; Chen et al., 2021), which are the ship detection results obtained by the Canny filter.

3.2. Suppressing background contours with self-adaptive Gaussian model

Though we can obtain the potential ship contour collections in MSV based on the Canny filter, edges from background objects in MSV may be wrongly extracted too. For instance, buoys in the water, aiding ship navigation, are commonly encountered background objects in MSV. The buoys' edges cannot be suppressed by the Canny detector as they are not noise edges. Besides, the sea-sky line (known as horizontal detection) is another typical challenge, which can significantly reduce the performance of the ship detection model. It is noticed that the imaging sizes of background objects are different from those of ships. More specifically, the size of buoys in MSV images is much smaller than that of ships, and the length of the sea-sky line is significantly longer than a ship. The above information stimulated the authors to employ the Gaussian related operator to rule out background edges on the obtained ship contours. The traditional Gaussian filter removes noise by setting the standard deviation σ to a constant value, which is infeasible for smoothing scale-varied noisy edges when detecting ships from MSV images. Thus, the self-adaptive Gaussian filter is introduced to suppress ship edge noises adaptively.

More specifically, we can automatically set different σ values as different image local areas possess different visual resemblance to the referred image area (Chen and Ellis, 2014). Thus, the self-adaptive Gaussian filter can automatically remove the false detected ship edges. Previous studies have shown that the noise-free detected ship edges $S_p(x, y)$ is roughly equal to the addition of raw ship contours and its second order derivative (Chen and Ellis, 2014). The ship contours, obtained from the previous step, are labelled $S_p(x, y)$. The formula of $S_p(x, y)$ is shown in Equation (8). The self-adaptive Gaussian filter's performance on suppressing background edges is determined by the optimal set of parameter β in Equation (8). However, the optimal setting of β is obtained by finding the minimum value of $E(\beta)$ in Equation (9).

$$S_p(x, y) \approx S(x, y) + \frac{S''(x, y)}{2} \beta^2 \quad (8)$$

$$E(\beta) = \int \int [(S(x, y) - G(x, y) * S(x, y))^2 + \lambda |\nabla \beta^{-1}|^2] \quad (9)$$

where β^2 is variance of the self-adaptive Gaussian filter, and operator $*$ represents convolution operation of gradient $G(x, y)$ and raw ship contours $S(x, y)$. The parameter λ controls the convergence of $E(\beta)$.

3.3. Reconstructing ship contours with morphology open operation

Small wave clutter-related edges can be observed from the output of the ship edge outlier removal procedure. The morphology model can suppress negative ship contours (i.e., wave-related edges), and thus further connect the broken edges into ship contours (i.e., ship detection results). In that manner, ship contour reconstruction performance can be measured by the ship detection results, which are quantitatively analysed with the two statistical indicators (i.e., Re and Pr). More details can be found in the following section on the experiments and results.

More specifically, the open operator of the morphology method is employed to reconstruct ship contours by connecting neighbouring detected ship edges into a rectangle (i.e., each rectangle is a detected ship in the maritime frame). The open operator is obtained by sequentially implementing the erosion and dilation operation, as the former operation aims to remove the clutter while the latter reconstructs ship contours in each frame of MSV. For a given structure element B , we can obtain morphology erosion results [see Equation (10)] by obtaining the minimum intersection pixels between the B and the local image patches (generated by traversing through the ship frame in both left-to-right and up-to-down directions).

The outputs of the erosion filter are considered as ship edges which may be disconnected from neighbouring edges. The dilation operation is thus employed to connect the split edges into a closed ship contour, which obtains the union set of pixels between the B and the local image patches. The dilation formula is shown in Equation (11). However, the dilation outputs are the final ship detection

results in the proposed framework. Previous studies suggested that the diamond element is efficient for suppressing the blob noises (Harvey et al., 2010; Deng et al., 2013; Yang et al., 2016). Motivated by that, different scales of the diamond element were tested (3×3 , 5×5 , 7×7), and it was found that the 3×3 structure element performs well at suppressing noise and keeping ship edges. Thus, the 3×3 diamond element is used as the default structure element in this study.

$$G_f \ominus B = \{(x, y) | B(x, y) \in G_f\} \quad (10)$$

$$G_f \oplus B = \{(x, y) | B(x, y) \cap G_f \neq \emptyset\} \quad (11)$$

where B is structure element and (x, y) is pixel of x - and y -coordinates, respectively. G_f is the detected ship contour set for a ship frame obtained from the previous step.

4. Experiments and results

The proposed ship detector is applied to three typical MSV clips collected for this study (i.e., case-a, case-b and case-c mentioned above). To evaluate and compare the performance of the proposed detector, the traditional Gaussian method (Deng and Cahill, 1993), the SVM model (Morillas et al., 2015), and the mask-RCNN model (Nie et al., 2018) were implemented to detect ships in the same MSV clips. The ship detection models were implemented on Windows 10 OS with setups of 8 GB RAM and 3.4 GHZ CPU. The software integrated development environment (IDE) for implementing the model was PyCharm version 2016.3, and python version 2.7 was used for fulfilling the ship detection task. The default sensitivity threshold for the Canny filter was set to 0.05 by implementing a series of sensitivity analysis experiments. Note that fine-tuning details of the parameters are did not provided here due to limitations of space. The image padding for the SVM model was set to 32×32 , and the default scale factor was set to 1.1. With help from members of the authors' group, the ground truth ship positions in each MSV image were manually labelled. The mask-RCNN was trained by following the structure in a previous study (Zimmermann and Siems, 2019), implemented with open-source libraries (including Keras and Tensorflow). More specifically, the mask size was set to 28×28 , and each image was re-sized to 1024×800 which was zero padded. ResNet-101 feature pyramid network (i.e., serving as a backbone) was employed to train the mask-RCNN network in this study, with 80% of the initial ship images as training images and the remaining 20% as testing images.

4.1. Measures of goodness of detection

The overlapping magnitude between the detected bounding box and ground truth was employed to determine the ship detection results. More specifically, the detected bounding box is considered to be a ship when the overlapping ratio exceeds a certain threshold. For more detailed explanation of overlapping calculation, refer to Dasari and Gorthi (2020). Previous research suggests that Re rate and Pr rate are two popular yet efficient metrics to measure object detection performance such as ship detection (Zhang et al., 2020). Re demonstrates correct positive ship detection ratio of the ship detection model, while Pr shows the ship detection accuracy. For that reason, the two indicators were employed to quantify various models' performance in our study. The definitions of Re and Pr are shown in Equations (12) and (13), respectively. More specifically, Re demonstrates the proportion of the positive ship targets that have been successfully detected by the models. The higher value of Re also represents a better detection result for the model. The indicator Pr shows the accuracy performance of a ship detector, where higher Pr means the ship detector misses fewer ships in a MSV frame. In addition, the time cost is employed to measure the models' computation complexity. In sum, higher Re and Pr indicate better performance for a ship detector, while lower time cost shows less complexity for the ship detection model.

$$Re = \frac{T}{T + T_F} \quad (12)$$

$$\text{Pr} = \frac{T}{T + F_T} \quad (13)$$

where T is the detected true-positive ship number. Parameter T_F represents detected false positive ship number and F_T is detected false-negative ship number. The Pr and Re indicators obtain their best performance when their values reach one, and worst at zero. The ship detection model obtains better performance when the Pr (and Re) is closer to 1, and vice versa.

4.2. Detection results and discussions

Typical ship detection results of each model for the three maritime traffic scenarios are provided in Figure 4. Figure 4(a) demonstrates the ship detection performance of each model when the maritime traffic volume was small. Ships with obvious contours can be successfully detected by various models (i.e., Gaussian, SVM, mask-RCNN and CGM [Canny-Gaussian-Morphology]). The detection models showed different performance when a ship was close to the sea-sky line (i.e., ship visual features were ambiguous). The ship detection results for case-b and case-c [see Figure 4(b) and 4(c), respectively] showed similar performance to those of case-a. Typical ship detection results are provided, considering that there was no significant difference in detection performance for different ship images for each case. More detailed explanations about the ship detection results are provided as follows.

4.2.1. Ship detection results for low traffic volume scenario

Ship detection results for each model of typical frame (second frame) under case-a were shown in Figure 4(a), where green rectangle demonstrates the ground truth ship position and red rectangle denotes the detected ship position. It is noticed that the traditional Gaussian model obtains many erroneous detection results, such as a ship being wrongly detected as several ships, mis-detection of small ships, etc. The main reason for unsatisfactory detection performance is that the Gaussian model is very sensitive to neighbouring intensity variance between background (i.e., wave pixels) and foreground (i.e., ships). More specifically, it is very difficult to set an appropriate threshold which helps the traditional Gaussian model successfully suppress wave interference, and retain the ship pixels. Thus, the false alarm rate is very high for the Gaussian model.

The SVM, mask-RCNN and CGM models show better detection performance than the traditional Gaussian method, as the number of false detected ships is significantly smaller than the Gaussian counterpart [see Figure 4(a)]. It is noted that the SVM model successfully detected all ships in the frame in Figure 4. However, small ships close to the sea-sky line were not accurately detected by the SVM model, and many ships were repeatedly detected (i.e., a small ship was wrongly detected as different ships). The mask-RCNN, a deep learning-based method, accurately detected all ships in the MSV image sequences, and the ships far away from the camera were successfully detected too. The high ship detection performance of the mask-RCNN is because the hidden layers of the model sufficiently learn and extract distinct image features of the training ship samples (at different scales, varied imaging views, etc.). The proposed CGM method showed similar performance to the mask-RCNN [see the right-most image of Figure 4(a)]. In sum, the SVM, mask-RCNN and CGM models can detect ships in images with few detection mistakes, while the Gaussian model is severely degraded by background pixels.

Table 2 shows the different models' statistical performance in case-a. The Re rate for the Gaussian model is 0.69, which is much smaller than the other three methods. The Re value for the other three models is higher than 0.90, indicating that over 90% of ships in the image sequences are successfully detected. The Re value of CGM is 6.19% higher than that of SVM, and 2.02% lower than the mask-RCNN counterpart. The Pr value shows that the four models obtain satisfactory precision performance (as the minimum Pr is 0.91). More specifically, the Pr value of the CGM model is 0.96 which is 3.23% higher than the counterpart SVM model. In addition, the mask-RCNN Pr value is 0.99 which is 3.12% higher than the proposed model. The detection results for the mask-RCNN and CGM models were carefully checked, and it was found that the mask-RCNN detects small ships near to the sea-sky line

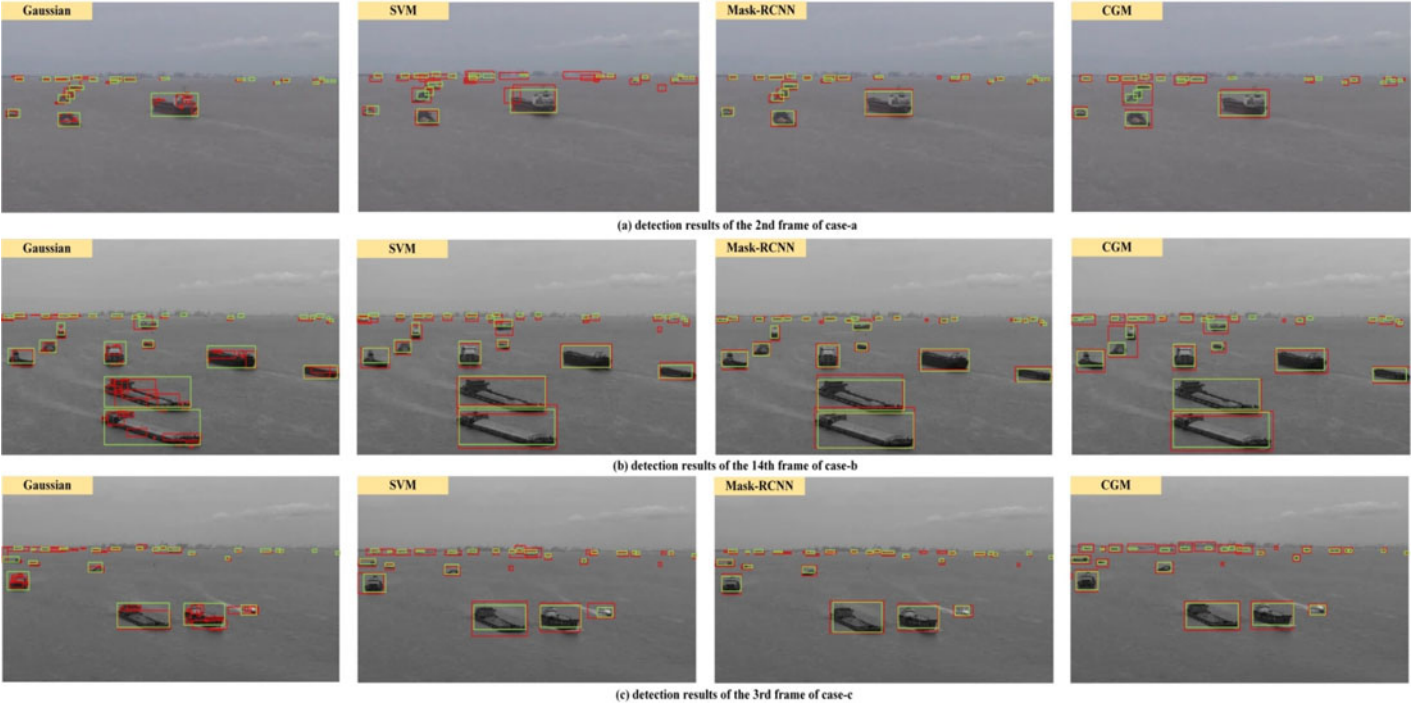


Figure 4. Ship detection results of typical frames in three cases.

Table 2. *Statistical performance of different methods in case-a.*

	Gaussian	SVM	Mask-RCNN	CGM
Re	0.69	0.91	0.99	0.97
Pr	0.87	0.93	0.99	0.96

Table 3. *Time cost for each model for three cases (ms/frame).*

	Gaussian	SVM	Mask-RCNN	CGM
Time cost at case-a	10	20	70	28
Time cost at case-b	16	28	81	35
Time cost at case-c	13	23	76	33
Average time cost	13	23.67	75.67	32

very accurately. However, the proposed CGM model sometimes detects two small ships as one large ship, and thus degrades the CGM performance. As shown in the first row of [Table 3](#), the time cost for the mask-RCNN model is 70 ms per frame, which was seven-fold (three-fold, two and a half) higher than that of the Gaussian model (SVM, the proposed CGM). The main reason is that the potential candidate ship regions are iteratively searched by convolution layers in the mask-RCNN, and this increases the model time cost. Considering that the human naked eye may fail to recognise small ships close to the sea-sky line, the CGM and mask-RCNN models' detection performances are both acceptable in the low traffic volume situation.

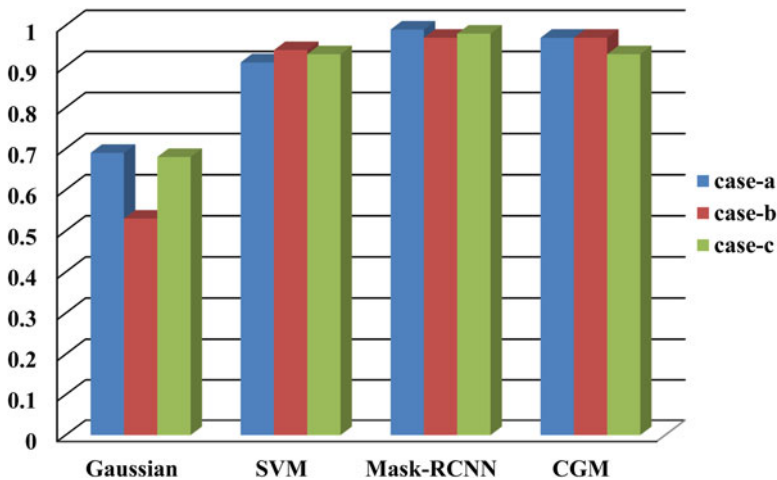
4.2.2. Ship detection results for case-b and case-c

Ship detection results of case-b (i.e., high traffic volume) are shown in subplot (b) of [Figure 4](#). It was observed that the Gaussian method obtained many false positive detected ships, which is similar to the detection performance in the low traffic volume situation. The SVM, mask-RCNN and CGM methods successfully detected ships with a few false alarms. The SVM model also showed better performance in the high traffic volume scenario than in traffic volume. The main reason is that ship-related pixels in case-b involve larger proportions of each frame, and this helped the SVM model suppress wave interference. It is noticed that a buoy was wrongly detected as a ship by SVM, indicating that a new detection outlier was introduced by SVM. Both CGM and mask-RCNN have successfully detected ships in the frame without obvious detection outliers. It was found that the ship positions obtained by mask-RCNN were closer to the ground truth data compared with those of the CGM model, which were more obviously observed near the sea-sky line area. The main reason is that the mask-RCNN model was trained with comprehensive ship samples, and thus the model fulfilled the ship detection task by identifying more advanced ship features in the MVS images. Moreover, the CGM model implemented the ship detection task by determining the salient contours, and wave pixel intensity around the ship may be wrongly detected as ship contours. For this reason, the CGM-detected ship area is usually larger than the ground truth ship imaging area.

The statistical performance for each model in case-b (see [Table 4](#)) showed similar results to those of case-a. More specifically, the Gaussian and SVM models obtained 0.53 and 0.94 Re ratio in case-b, while the mask-RCNN and CGM both achieved 0.97. The Pr indicator shows that detection results of the four models are acceptable, as the minimal and maximal Pr values are 0.81 and 0.95, respectively. From the perspective of time cost, the conventional Gaussian and SVM models outperformed the mask-RCNN and proposed CGM models, which were 16, 28, 81 and 35 ms per frame (see the second row in [Table 3](#)). In that way, the CGM computation overhead was one-fourth larger than the traditional ship detection models (i.e., Gaussian and SVM model). But, the time cost for the proposed CGM model was

Table 4. Statistical performance of different methods in case-b.

	Gaussian	SVM	Mask-RCNN	CGM
Re	0.53	0.94	0.97	0.97
Pr	0.81	0.87	0.95	0.93

**Figure 5.** Re indicator distributions for the three cases.

less than half that of the mask-RCNN model. Moreover, both the Re and Pr indicators demonstrated that mask-RCNN and CGM obtained satisfactory detection performance (i.e., the two statistical indicators were larger than 90%), and the mask-RCNN time cost was two-fold higher than that of the CGM model. According to the above analysis, it is reasonable to say that the proposed ship detection model shows reliable and real-time performance under heavy traffic situation.

The different models' performance was also evaluated by detecting ships in MSV images with strong ripple interference. As shown in the subplot (c) of Figure 4, there was a long, narrow ship wake generated by the white vessel. The Gaussian model showed similar detection performance as in case-a and case-b. More specifically, many false positive results were observed, as shown in the left-most subplot in Figure 4(c). The wave pixels neighbouring the white vessel were wrongly detected as a ship. The SVM, mask-RCNN and CGM models successfully suppressed the wave interference. It is noted that the SVM model recognised wake pixels as part of the white ship, and thus in the final detection result the white ship was significantly larger than the actual ship imaging size, while the mask-RCNN and CGM model accurately detected the white ship without containing redundant wave pixels. For the purpose of better readability, the distributions of the Re and Pr indicators for different ship detection models are provided in Figures 5 and 6, respectively. From the perspective of Re distributions, the SVM outperformed the Gaussian model under the three typical maritime situations (see Figure 5). The CGM model obtained better ship detection performance compared with those of SVM model for the three cases, which were quite close to mask-RCNN model performance. The Pr distributions confirmed this analysis, which can be found in Figure 6.

The Re and Pr indicators in Table 5 verified that the mask-RCNN obtained optimal detection performance, while the CGM model performance was close to that of the mask-RCNN. The time cost for the mask-RCNN model for case-c was significantly higher than the Gaussian, SVM and CGM models, which were 76, 13, 23 and 33 ms per frame (see the third row in Table 3). Average ship detection time cost was obtained to further quantify model performance. More specifically, the average time cost for each model was 13, 23.67, 75.67 and 32 ms per frame for the Gaussian, SVM, mask-RCNN and

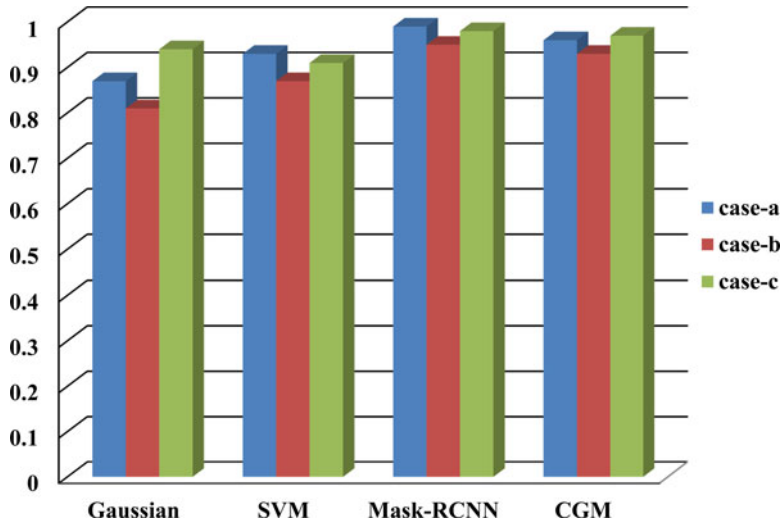


Figure 6. Pr indicator distributions for the three cases.

Table 5. Statistical performance of different methods in case-c.

	Gaussian	SVM	Mask-RCNN	CGM
Re	0.68	0.93	0.98	0.93
Pr	0.94	0.91	0.98	0.97

CGM, which indicated that the proposed model can obtain real-time ship detection results. Based on the above quantitative and qualitative analysis, it can be concluded that the proposed CGM is a reliable and efficient ship detector under various challenging ship detection situations.

5. Conclusion

Ship detection from coastal MSV is considered as a bottleneck for implementing the intelligent navigation task in the smart ship era, which may face various coupled detection challenges (e.g., ship wake, sea waves). To tackle this challenge, this paper proposed an ensemble yet efficient ship detection framework via edge detection logic. More specifically, the proposed ship detection framework was implemented via steps of ship edge detection, background edge removal and ship contour reconstruction. First, the ship edge detection step employed the Canny detector to determine potential ship edges from MSV images. Second, the background-related edges (i.e., negative positive detection results) were suppressed in the background edge removal step with the help of the self-adaptive Gaussian filter. Third, the ship contour reconstruction step was implemented to connect adjacent edges into close ship contours, and thus ship detection results were obtained. The proposed ship detection model performance was verified under three typical maritime traffic scenarios, which were further compared against three popular ship detectors (i.e., basic Gaussian, SVM and mask-RCNN). The experimental results demonstrated that the proposed framework obtained satisfactory performance considering that the average Re and Pr indicators were both 0.95 and the average time cost was 32 ms per frame. The proposed model can help a smart ship (i.e., unmanned ship) to be aware of on-site maritime traffic situations in real time, and thus advance the development of the intelligent navigation era.

In future, better performance of the model can be achieved by conducting the following explorations. First, ships to be detected in the three test scenarios were not severely overlapped by the neighbouring

objects. Testing the proposed model's performance under maritime scenarios with severe occlusions deserves further research interests. Second, the ship detection model's performance was not tested on MSV clips which involved strong camera vibration. In future, the model performance can be enhanced by integrating a background stabilisation model to tackle the ship detection task under the challenge of strong camera vibration. Third, we note that both adverse weather and low visibility can impose negative influence on ship detection performance. It would be interesting to develop a robust ship detection model against weather-related interference. Fourth, considering that the deep learning model can obtain better detection performance, merging deep learning-based models (YOLO + RCNN, inception based object detector, etc.) to improve the model is one of the potential research works in the future. Fifth, ship edge detection results can be further evaluated against varied filters to exploit model detection performance. Last but not least, the model did not tackle the challenge of ship detection under strong systematic disturbance (e.g., from seagulls). We can fulfil the task by exploiting both ship and non-ship imaging rules under strong systematic disturbance, and thus obtain satisfactory ship detection model performance under strong disturbance interference.

Acknowledgements. This work was jointly supported by the National Key R&D Program of China (2019YFB1600605), National Natural Science Foundation of China (52071200, 51978069, 52072237, 62073212, 71942003), Shanghai Committee of Science and Technology, China (18DZ1206300), Shanghai Planning Office of Philosophy and Social Science (2019EGL018), Key Research and Development Plan of Shaanxi Province (2021KWZ-09).

References

- Biondi, F.** (2017). Low-rank plus sparse decomposition and localized radon transform for ship-wake detection in synthetic aperture radar images. *IEEE Geoscience and Remote Sensing Letters*, **15**, 117–121.
- Canny, J.** (1986). A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **8**, 679–698.
- Chen, Z. and Ellis, T.** (2014). A self-adaptive Gaussian mixture model. *Computer Vision & Image Understanding*, **122**, 35–46.
- Chen, X., Wang, S., Shi, C., Wu, H., Zhao, J. and Fu, J.** (2019). Robust ship tracking via multi-view learning and sparse representation. *Journal of Navigation*, **72**, 176–192.
- Chen, X., Qi, L., Yang, Y., Luo, Q., Postolache, O., Tang, J. and Wu, H.** (2020a). Video-based detection infrastructure enhancement for automated ship recognition and behavior analysis. *Journal of Advanced Transportation*, **2020**, 1–12.
- Chen, X., Yang, Y., Wang, S., Wu, H., Tang, J., Zhao, J. and Wang, Z.** (2020b). Ship type recognition via a coarse-to-fine cascaded convolution neural network. *Journal of Navigation*, **73**, 813–832.
- Chen, Z., Chen, D., Zhang, Y., Cheng, X., Zhang, M. and Wu, C.** (2020c). Deep learning for autonomous ship-oriented small ship detection. *Safety Science*, **130**, 104812.
- Chen, X., Li, Z., Yang, Y., Qi, L. and Ke, R.** (2021). High-resolution vehicle trajectory extraction and denoising from aerial videos. *IEEE Transactions on Intelligent Transportation Systems*, **22**, 3190–3202.
- Dasari, M. M. and Gorthi, R. K. S. S.** (2020). IOU – Siamtrack: IOU Guided Siamese Network for Visual Object Tracking. Proceedings of the 2020 IEEE International Conference on Image Processing (ICIP).
- Deng, G. and Cahill, L. W.** (1993) An adaptive Gaussian filter for noise reduction and edge detection. Proceedings of the Nuclear Science Symposium & Medical Imaging Conference.
- Deng, C.-X., Wang, G.-B. and Yang, X.-R.** (2013) Image edge detection algorithm based on improved Canny operator. Proceedings of the 2013 International Conference on Wavelet Analysis and Pattern Recognition.
- Graziano, M. D.** (2020). Preliminary results of ship detection technique by wake pattern recognition in SAR images. *Remote Sensing*, **12**, 2869.
- Harvey, N. R., Porter, R. and Theiler, J.** (2010). Ship detection in satellite imagery using rank-order grayscale hit-or-miss transforms. Proceedings of the International Society for Optical Engineering, 6 April, Orlando, FL.
- Huang, Y., Li, Y., Zhang, Z. and Liu, R. W.** (2020). GPU-Accelerated Compression and visualization of large-scale vessel trajectories in maritime IoT industries. *IEEE Internet of Things Journal*, **7**, 10794–10812.
- Ke, R., Li, Z., Kim, S., Ash, J., Cui, Z. and Wang, Y.** (2017). Real-time bidirectional traffic flow parameter estimation from aerial videos. *IEEE Transactions on Intelligent Transportation Systems*, **18**, 890–901.
- Kim, K., Hong, S., Choi, B. and Kim, E.** (2018). Probabilistic ship detection and classification using deep learning. *Applied Sciences*, **8**, 936.
- Le Caillec, J., Gorski, T., Sicot, G. and Kawalec, A.** (2018). Theoretical performance of space-time adaptive processing for ship detection by high-frequency surface wave radars. *IEEE Journal of Oceanic Engineering*, **43**, 238–257.
- Le Caillec, J.-M., Habonneau, J. and Khenchaf, A.** (2019). Ship profile imaging using multipath backscattering. *Remote Sensing*, **11**, 748.
- Lee, J., Tang, H. and Park, J.** (2018). Energy efficient Canny edge detector for advanced mobile vision applications. *IEEE Transactions on Circuits and Systems for Video Technology*, **28**, 1037–1046.

- Liu, J. and Gao, Y.** (2020). 3D pose estimation for object detection in remote sensing images. *Sensors (Basel)*, **20**, 1–16.
- Liu, L., Ouyang, W., Wang, X., Fieguth, P., Chen, J., Liu, X. and Pietikäinen, M.** (2020a). Deep learning for generic object detection: A survey. *International Journal of Computer Vision*, **128**, 261–318.
- Liu, R. W., Nie, J., Garg, S., Xiong, Z., Zhang, Y. and Hossain, M. S.** (2020b). Data-driven trajectory quality improvement for promoting intelligent vessel traffic services in 6G-enabled maritime IoT systems. *IEEE Internet of Things Journal*, **8**, 5374–5385.
- Lu, J. W., He, Y. J., Li, H. Y. and Lu, F. L.** (2006). Detecting Small Target of Ship at Sea by Infrared Image. Proceedings of the IEEE International Conference on Automation Science & Engineering.
- Min, Z., Jing, M., Liu, D., Xia, Z., Zou, Z. and Shi, Z.** (2018). Multi-resolution networks for ship detection in infrared remote sensing images. *Infrared Physics & Technology*, **92**, 183–189.
- Morillas, J. R. A., García, I. C. and Zölzer, U.** (2015). Ship detection based on SVM using color and texture features. Proceedings of the 2015 IEEE International Conference on Intelligent Computer Communication and Processing (ICCP).
- Mumtaz, A., Jabbar, A., Mahmood, Z., Nawaz, R. and Ahsan, Q.** (2016). Saliency based algorithm for ship detection in infrared images. Proceedings of the International Bhurban Conference on Applied Sciences & Technology.
- Nie, S., Jiang, Z., Zhang, H., Cai, B. and Yao, Y.** (2018). Inshore Ship Detection Based on Mask R-CNN. Proceedings of the IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium.
- Nova, O., Guiffaut, C. and Reineix, A.** (2020). Method for the sea clutter characterization in HF surface wave radars from the fields diffracted by the sea surface. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, **13**, 403–413.
- Pappas, O., Achim, A. and Bull, D.** (2018). Superpixel-level CFAR detectors for ship detection in SAR imagery. *IEEE Geoscience and Remote Sensing Letters*, **15**, 1397–1401.
- Ren, Y., Yang, J., Zhang, Q. and Guo, Z.** (2020). Ship recognition based on Hu invariant moments and convolutional neural network for video surveillance. *Multimedia Tools and Applications*, **80**, 1343–1373.
- Shafer, S. and Harguess, J.** (2015). Sparsity-driven anomaly detection for ship detection and tracking in maritime video. Proceedings of the Spie Defense + Security.
- Shao, Z., Wang, L., Wang, Z., Du, W. and Wu, W.** (2019). Saliency-aware convolution neural network for ship detection in surveillance video. *IEEE Transactions on Circuits and Systems for Video Technology*, **30**, 781–794.
- Wang, X., Peng, Z., Kong, D., Ping, Z. and He, Y.** (2017a). Infrared dim target detection based on total variation regularization and principal component pursuit. *Image & Vision Computing*, **63**, 1–9.
- Wang, C., Bi, F., Zhang, W. and Chen, L.** (2017b). An intensity-space domain CFAR method for ship detection in HR SAR images. *IEEE Geoscience and Remote Sensing Letters*, **14**, 529–533.
- Wang, D., Wang, P., Zhang, X., Guo, X., Shu, Y. and Tian, X.** (2020). An obstacle avoidance strategy for the wave glider based on the improved artificial potential field and collision prediction model. *Ocean Engineering*, **206**, 107356.
- Wei, C.** (2009). *Research on Infrared Small Target Detection and Tracking Algorithm*. Harbin, PR China: Harbin Institute of Technology.
- Xie, B., Hu, L. and Mu, W.** (2017). Background Suppression Based on Improved Top-Hat and Saliency Map Filtering for Infrared Ship Detection. Proceedings of the 2017 International Conference on Computing Intelligence and Information System (CIIS).
- Yan, C. Z., Liu, C. and Pang, Y.** (2019). Multiscale saliency detection method for ship targets in synthetic aperture radar images. *The Journal of Engineering*, **2019**, 7585–7588.
- Yang, G., Jing, Y., Xiao, C. and Sun, W.** (2016). Ship wake detection for SAR images with complex backgrounds based on morphological dictionary learning. Proceedings of the IEEE International Conference on Acoustics.
- Yu, Y., Chen, L., Shu, Y. and Zhu, W.** (2021). Evaluation model and management strategy for reducing pollution caused by ship collision in coastal waters. *Ocean & Coastal Management*, **203**, 105446.
- Zhang, Y., Li, Q.-Z. and Zang, F.-N.** (2017). Ship detection for visual maritime surveillance from non-stationary platforms. *Ocean Engineering*, **141**, 53–63.
- Zhang, Y., Chen, C., Wu, Q., Lu, Q., Zhang, S., Zhang, G. and Yang, Y.** (2018). A kinect-based approach for 3D pavement surface reconstruction and cracking recognition. *IEEE Transactions on Intelligent Transportation Systems*, **19**, 3935–3946.
- Zhang, G., Li, Z., Li, X., Yin, C. and Shi, Z.** (2020). A novel salient feature fusion method for ship detection in synthetic aperture radar images. *IEEE Access*, **8**, 215904–215914.
- Zhao, J., Wen, B., Tian, Y., Tian, Z. and Wang, S.** (2019). Sea clutter suppression for shipborne HF radar using cross-loop/monopole array. *IEEE Geoscience and Remote Sensing Letters*, **16**, 879–883.
- Zhu, C. and Xue, W.** (2015). Detect ships using saliency in infrared images with sea-sky background. Proceedings of the International Conference on Digital Image Processing.
- Zimmermann, R. S. and Siems, J. N.** (2019). Faster training of mask R-CNN by focusing on instance boundaries. *Computer Vision and Image Understanding*, **188**, 102795.