

Article

Salient Ship Detection via Background Prior and Foreground Constraint in Remote Sensing Images

Jianming Hu ^{1,†} , Xiyang Zhi ^{1,*,†} , Wei Zhang ¹, Longfei Ren ² and Lorenzo Bruzzone ³ 

¹ Research Center for Space Optical Engineering, Harbin Institute of Technology, Harbin 150001, China; hujianming@stu.hit.edu.cn (J.H.); wzhang@hit.edu.cn (W.Z.)

² Provincial Key Lab of Information Coding and Transmission, Southwest Jiaotong University, Chengdu 610031, China; renlf@my.swjtu.edu.cn

³ Remote Sensing Laboratory, Department of Information Engineering and Computer Science, University of Trento, 38123 Trento, Italy; lorenzo.bruzzone@ing.unitn.it

* Correspondence: zhixiyang@hit.edu.cn; Tel.: +86-0451-86414883

† These authors contributed equally to this work.

Received: 22 September 2020; Accepted: 11 October 2020; Published: 15 October 2020



Abstract: Automatic ship detection in complicated maritime background is a challenging task in the field of optical remote sensing image interpretation and analysis. In this paper, we propose a novel and reliable ship detection framework based on a visual saliency model, which can efficiently detect multiple targets of different scales in complex scenes with sea clutter, clouds, wake and islands interferences. Firstly, we present a reliable background prior extraction method adaptive for the random locations of targets by computing boundary probability and then generate a saliency map based on the background prior. Secondly, we compute the prior probability of salient foreground regions and propose a weighting function to constrain false foreground clutter, gaining the foreground-based prediction map. Thirdly, we integrate the two prediction maps and improve the details of the integrated map by a guided filter function and a wake adjustment function, obtaining the fine selection of candidate regions. Afterwards, a classification is further performed to reduce false alarms and produce the final ship detection results. Qualitative and quantitative evaluations on two public available datasets demonstrate the robustness and efficiency of the proposed method against four advanced baseline methods.

Keywords: ship detection; visual saliency; background prior; foreground constraint; optical remote sensing

1. Introduction

Maritime ship detection in remote sensing images is a crucial technique for various military and civilian applications, such as maritime dynamic monitoring [1], oceanic rescue services and vessel traffic management. From the perspective of remote sensing data sources, ship detection can be addressed by using either synthetic aperture radar (SAR) images [2] or optical images. Compared with the SAR detection mode, the optical mode has advantages in capturing abundant spectral information and preserving high-resolution edge details, thus becoming a popular technical pattern for marine wide-area fast search, discovery, recognition and confirmation. However, ship detection in optical remote sensing images is a challenging task due to the changeable appearances of targets and the complicated oceanic background. Actually, ships in optical images are quite different in intensity [3], size, direction, color, texture and other object characteristics. Meanwhile, ships are easily confused with the interferences of heavy clouds, islands, ocean waves and other uncertain sea state conditions, resulting in the difficulty of target detection increasing. Therefore, the detection of different types

of ships in a complex maritime background from optical remote sensing images is still an open challenging problem.

In order to achieve an efficient ship detection in optical images, a variety of ship detection algorithms have been presented over the last decades. Most current existing methods adopt a two-stage detection mechanism based on the selection of candidate region [4] and the ship target discrimination [5]. In the first stage, potential regions of candidate ship targets are extracted by minimizing missed alarms. The next stage consists in discriminating whether candidates generated by the first stage contain ship targets and in locating their positions in the image accurately [6]. Generally speaking, the traditional ship detection methods can be roughly classified into two types [7]. The first type includes several threshold-based methods which are mainly based on the gray-level information statistics of the sea background. The regions beyond the gray threshold are considered as target regions. In [8], a threshold estimation process is defined to eliminate cloud interference and improve ship detection performance. The Otsu threshold selection is made to segment the possible target regions and clutter pixels in [9]. However, these methods are only suitable when the sea background has uniform texture distribution and the ship target and the sea background have a high contrast. The detection performance reduces greatly in complex sea background. The second type consists of methods based on target feature representation, which usually utilize shape, texture and spectral characteristics to distinguish between the target and the sea background. For instance, in [10] the shape and texture characteristics of maritime targets are studied and the false alarms are removed by a support vector machine (SVM) classifier. Furthermore, the wake signature is used to identify the moving ship in [11]. In addition, geometric features including ship head detection, ship width and ship length are computed in [12] to locate the ship regions. These methods based on target representation can provide relatively high detection accuracy. However, they have poor adaptability to the changes of target characteristics in different illumination and scale conditions, leading to some false alarms. In order to guarantee the detection accuracy, a large amount of prior information of the ship target in different conditions is needed. Moreover, these methods usually require an exhaustive search by using sliding window on potential target location in the image, leading to a large amount of computation and time consumption.

Recently, deep learning approaches have become significant tools for ship detection. Different excellent deep learning networks have been applied for feature extraction. For instance, in [13] a residual dense network is exploited to learn the features of different levels. Similarly, a pyramid structure is used to learn features of multi-scale rotating targets, which can accurately detect dense targets in various complex scenes [14]. Furthermore, a singular value decomposition network is proposed to automatically learn the multi-scale features of the image scene elements in [15], thus providing a reliable input for the subsequent classifier. In [16], an improved network derived from the classic YOLOv2 architecture is presented to collect the different fine-grained features, which contributes to detect ships in different directions under complex ocean scenes. In addition, a network structure is used to learn the cross-domain invariant features of ships in [17], which can effectively solve the problem of small sample classification. Generally speaking, different multi-layer neural networks have advantages in representing targets in the feature extraction step. Moreover, these methods can effectively resist the interference of complex backgrounds, and are suitable to address the detection of large-scale and high-contrast targets in natural scenes. However, in the case of large sea surfaces and small ship targets, the detection accuracy decreases. In addition, deep learning methods require many training samples as well as complex training phases. In general, a good detection method should consider both accuracy and calculation speed. Therefore, methods based on hand-designed features and shallow classifiers are still meaningful, especially for small platforms without GPU environment like Unmanned Airborne Vehicles (UAVs).

It is well known that humans can determine the significance of different image regions and concentrate on crucial areas routinely, even in highly cluttered scenes. Since the ships in optical images generally occupy the relatively low proportion of pixels compared with the whole image scene and can

be readily captured by the visual attention system, the ships can be regarded as salient objects even in complex scenes [18]. Inspired by the perceptual mechanism of the human vision systems, ship detection methods based on visual saliency have gained great research interest. In terms of computation, these methods based on visual saliency need neither exhaustive search by sliding window like traditional methods, nor long training on a large number of images like deep learning methods. Consequently, many saliency-based methodologies are introduced to locate the image parts of interest quickly that may comprise salient objects. For instance, in [3], a saliency model via the scene statistical properties analysis is presented to predict possible target regions and then a discriminative classifier is trained to confirm the correct ships. In [19], a saliency methodology via wavelet-domain transform is employed to detect the multi-scale ships under various complicated weather conditions. In recent years, several scholars attempted to mine prior information to explicitly segment foreground and background parts. In [20], a center-prior is exploited to calculate the distribution of background parts and the part saliency is defined by weighting the color contrast. Similarly, a background-prior map is constructed based on some corner cues in [21] and corresponding regions of interest are calculated by reverse-measurement principle. Furthermore, the convex hull of feature points is utilized to predict foreground elements in [22] and then a smooth process is performed by minimizing a energy function. In addition, in [23], the convex hull is used to estimate the coarse foreground regions, thus computing the object distribution and the observation likelihood. Although several prior-based researches have been conducted to predict the region of interest quickly and accurately, there are still some problems to be solved. On the one hand, most previous saliency methodologies via background prior mainly employ the global or local contrast to distinguish possible distributions of ship targets, but they usually calculate contrasts by considering the neighboring area or the image border area as the background part. Since this definition of background is not necessarily correct, the subsequent calculated contrast has some errors accordingly, leading to an unsatisfactory result of the final prediction map. In addition, these methods are based on the strict assumption that the target is near the image center, nevertheless, there are limited studies focused on the random positions of ships, which produces obvious prior errors especially for multi-target scenes or scenes when the target is far from the central field. On the other hand, since the invariant features of the target are seldom considered, approaches via foreground cues usually introduce extra background noise.

To address these issues mentioned above, it is required to define strategies for more accurately modeling background prior and suppressing interferences misjudged as foreground patches. Different from previous methods, we propose a non-center-prior information mining method, and fully consider the random position of ship target in the image to obtain a reliable background prior. Moreover, in order to effectively reduce noise caused by foreground extraction, we integrate the invariant features of the target into the final foreground prediction, and further adopt the refinement and discrimination strategies to obtain the final well-constrained results. Specifically, we propose a novel ship detection framework derived from visual saliency that can concurrently meet the demand of accurate and fast detection. First, we select superpixels near the image edges as the elements of the initial border set. Additionally, we further measure the boundary properties of each element to eliminate possible foreground noise, which is especially critical when there are ships near the edge in the multi-target scenes. The new border set provides reliable background prior information and we then gain the background-based prediction result by computing the saliency differences between the prior element and other image superpixels. This can effectively reduce the influences of many complex interference factors such as clouds and sea clutters. Second, we extract the foreground salient parts through convex hull segmentation and propose a weight strategy according to characteristic differences between the ship and non-ship elements in terms of eccentricity, symmetric regularity and regional solidity. By using this strategy, the regions with lower similarity to the real target characteristics tend to achieve more obvious inhibition effects. After restricting these false alarm regions, we gain the prediction map derived from well-constrained foreground regions. Third, considering that the background-based prediction map puts emphasis on the significance of all the potential salient regions,

whereas the foreground-based result can better restrain some mistaken candidate parts, we merge the two prediction maps and improve the details of the combined map by using a guided function and a wake suppression function. Thus, we gain the final saliency map, which is equivalent to completing the process of candidate region extraction. Finally, a discrimination classifier, which is robust to different ship sizes and directions, is employed to confirm real ships. The main steps of the proposed method are shown in Figure 1.

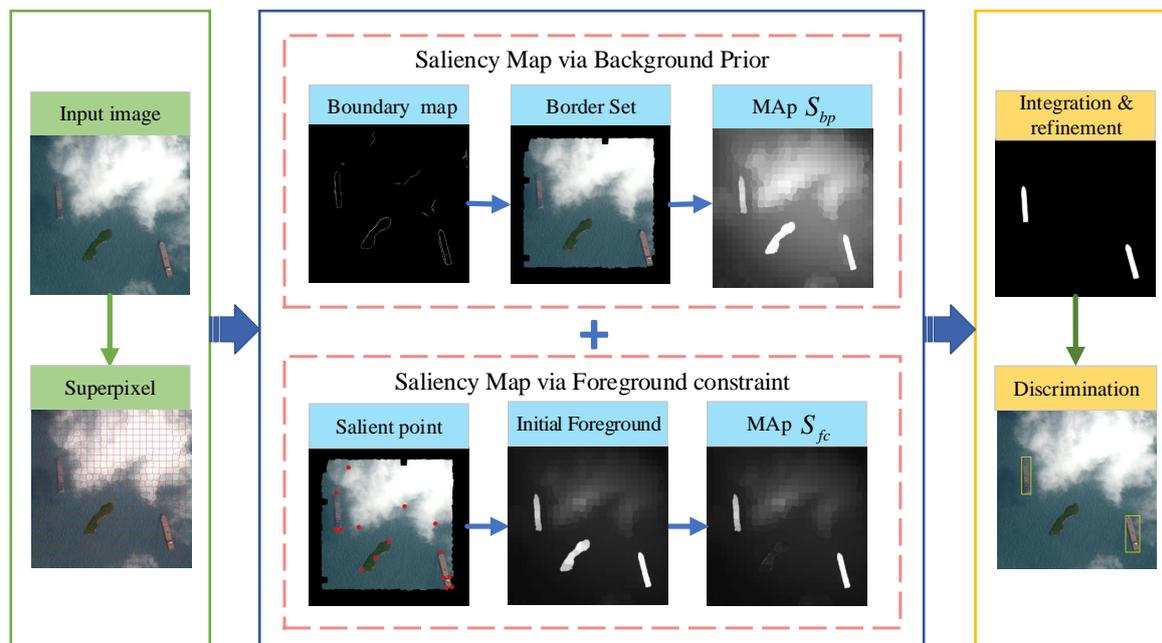


Figure 1. Illustration of the proposed salient ship detection method (Best viewed in color).

The main contributions of the proposed method consist in:

- A novel and stable framework for ship detection derived from visual saliency, which can efficiently detect multiple targets of different scales even in different complicated conditions with the interferences of heavy sea clutter, thick clouds, heavy wake, island and reef.
- A reliable background prior extraction method adaptive for the random locations of targets, which can automatically extract reliable background prior, making salient regions outstand more stably even in complex marine background.
- An efficient foreground constraint strategy combined with invariant characteristics of ship targets, which can effectively remove false alarm as well as highlight the correct salient regions.

The remainder of this paper is organized as follows. Section 2 elaborates the methodology of the proposed two-stage framework, namely the salient candidate region extraction stage, which consists of several steps including background prior extraction, foreground constraint and saliency map integration and refinement, and the ship target discrimination stage. Section 3 reports the experimental results of the proposed method and several advanced baseline techniques on two datasets. Finally, conclusions are drawn in Section 4.

2. Proposed Method

2.1. Saliency Prediction via Background Prior

From the general habit of image acquisition, the ship observed is usually distributed at or near the center of an image. However, when there are multiple ship targets (see Figure 2a), this prior principle is not necessarily accurate. Another prior principle, which is generally correct, is that regions near the

edge of the image are likely to contain the background information. Under this premise, we can extract the area along the image borders as background prior regions.

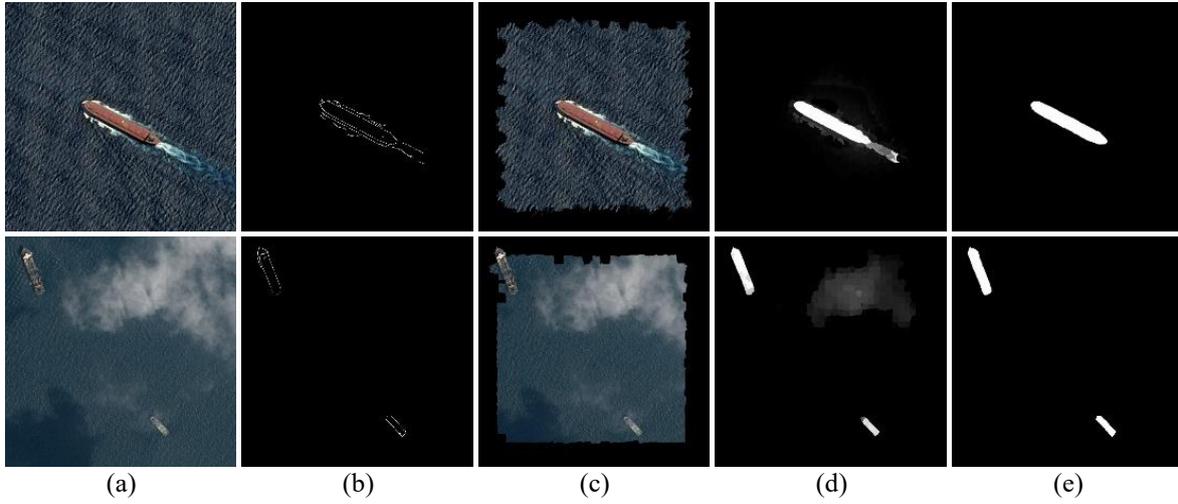


Figure 2. Illustration of prediction maps via background prior. (a) Input image, (b) probability of boundary (PB) map, (c) final border set, (d) saliency map via background prior and (e) ground truth.

Given an input image, the construction of a pixel-level graph-based model is computationally demanding. To capture the structural information and improve computation efficiency, we adopt superpixel as the minimum computing unit in an input image instead of single pixel. Let us assume that an input image is segmented into N compact and edge-preserving superpixels denoted as $\{S\}_{n=1}^N$ by the simple linear iterative clustering (SLIC) algorithm. Let the v_i^{lab} , h_i^{lab} and v_i^{cen} be the mean CIELab color vector, the CIELab histogram vector and the centroid location vector of the i -th superpixel, respectively. We build the border set by extracting the superpixels with a certain centroid distance from the image edges. By analyzing the characteristics of these border set, we can roughly obtain the prior characteristic of the background regions.

As previously described, for complex scenes, especially when there are multiple ships, some ships may appear in the border regions, leading to saliency detection error. We consider ships in the border set as foreground noise, and therefore we need to reduce the effect of foreground noise, while preserving the correct background regions. We employ the probability of boundary (PB) index [24] to measure the edge properties of superpixels (see Figure 2b). The PB index of the i -th superpixel s_i is calculated as

$$PB_i = \frac{1}{|S_i|} \sum_{I \in S_i} E_{(x,y)}^{pb} \quad (1)$$

where S_i represents the edge contour pixel set of the superpixel s_i , $|S_i|$ is the pixel number of the edge pixel set. (x, y) is the coordinate position of pixel element E , the PB value of the pixel E is calculated as

$$E_{(x,y)}^{pb} = \max_{\theta} \left[\sum_{p=1}^P \alpha_p G_p(x, y, \theta) \right] \quad (2)$$

where p is the feature channel index of the total channels P , for an RGB image, we usually consider $p = 4$ (i.e., brightness, texture, color a and color b). $G_p(x, y, \theta)$ represents the histogram gradient magnitude [25] in the direction of angle θ at point (x, y) . α_p is a parameter used to measure the impact of each channel magnitude and is assigned by analyzing the gradient change of the F-measure on some given images. In our experiments, θ is sampled in six directions within the interval $[0, \pi]$.

The smaller the PB_i value, the more likely the superpixel s_i belongs to the background regions. After selecting an adaptive gray threshold by using the OTSU segmentation algorithm [26], we can

effectively segment the background and foreground noise and then obtain a new image border set containing more reliable background information (see Figure 2c).

To distinguish high-contrast difference between background and salient regions, we choose a superpixel s_i from the border set and define the saliency distance between the background superpixel s_i and the superpixel to be analyzed s_j as

$$\begin{aligned} dis(s_i, s_j) = & \lambda_1 \|v_i^{lab} - v_j^{lab}\| + \lambda_2 \kappa^2(h_i^{lab}, h_j^{lab}) \\ & + \lambda_3 \|v_i^{cen} - v_j^{cen}\| \end{aligned} \quad (3)$$

where λ_1 , λ_2 and λ_3 are weighting parameters all normalized to $[0,1]$, $\kappa^2(h_i^{lab}, h_j^{lab})$ denotes the chi-squared distance between the two histograms, which is defined as

$$\kappa^2(h_i^{lab}, h_j^{lab}) = \frac{1}{2} \sum_{k=1}^K \frac{[h_i^{lab}(k) - h_j^{lab}(k)]^2}{h_i^{lab}(k) + h_j^{lab}(k)} \quad (4)$$

where K is the bin number of the histogram.

Considering the randomness of the positions of the ship and the background elements, we calculate the saliency maps in top, bottom, left and right directions, respectively. Let us consider a Gaussian similarity function which is defined as

$$w_{ij} = \exp\left(\frac{-dis(s_i, s_j)}{\sigma^2}\right) \quad (5)$$

where σ indexes a factor which adjusts the decay level. Then the map in the top direction can be derived as

$$S_{top} = \sum_{j \neq i} dis(s_i, s_j) w_{ij} \quad (6)$$

Similarly, by choosing the superpixel in the bottom, left and right border set in turn, we can compute the other three maps S_{bot} , S_{lef} and S_{rig} , respectively. Finally, the saliency map via background prior S_{bg} is obtained as

$$S_{bp} = S_{top} \cdot S_{bot} \cdot S_{lef} \cdot S_{rig} \quad (7)$$

Figure 2 shows some examples of background-based prediction results, from which we can see that the maps can highlight the ship candidates in single-target and multi-target images with the background suppressed to some extent.

2.2. Saliency Prediction via Foreground Constraint

After obtaining background-based saliency maps, from the perspective of remote sensing scene, we can basically extract the regions of ship candidates via the background prior knowledge, even when there are thin clouds and heavy sea clutters in the image scene. However, this may introduce serious errors for some input sea images, such as when the image scene has some islands or heavy clouds (see Figure 3a,b). As a result that some characteristics of islands and clouds are close to ships, we consider some non-ship objects as salient targets. Therefore, to generate robust saliency maps for more complex marine scene images, we propose a foreground constraint method to reduce background noise and enhance foreground contours. The main idea of the method is as follow: Firstly, we roughly locate the salient foreground parts, avoiding exhaustive searching and inefficient computation. Secondly, we compute the prior probability of each salient region derived from the Bayes rule to gain the preliminary foreground-based prediction map. Thirdly, we adopt a weighting strategy to constrain some possible mistaken salient parts, providing more accurate prediction results with less background noise.

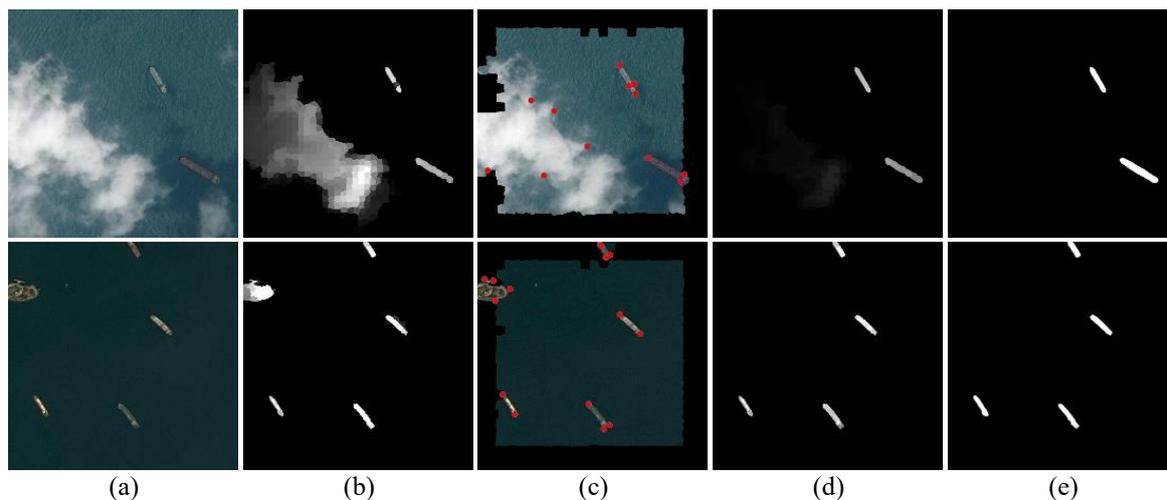


Figure 3. Illustration of prediction maps via foreground constraint. (a) Input image, (b) saliency map via background prior, (c) salient point map, (d) saliency map via foreground constraint and (e) ground truth.

Specifically, we firstly utilize the Harris point operator with color boosting [27] to generate salient points in the images with the border set removed (as shown in Figure 3c). The salient point is usually the feature point reflecting the local outline characteristics in the image. Let $I_x = (R_x, G_x, B_x)$ be the derivative vector of the image I . We give the color boosting function $g(\cdot)$ by considering the probability of the vector as

$$p(I_{x_1}) = p(I_{x_2}) \leftrightarrow \|g(I_{x_1})\| = \|g(I_{x_2})\| \tag{8}$$

where $p(\cdot)$ represents the probability of a vector. The salient point map is computed as

$$S_{harris} = H [g(I_x), g(I_y)] \tag{9}$$

where $H(\cdot)$ represents the Harris transformation filter.

Then, we extract the convex hull of the salient point set, thus estimating the rough distributions of different salient foreground parts. Considering that k-means approach is a widely used and efficient unsupervised clustering algorithm, we use it to divide the superpixels inside and outside the convex hull into several clusters, respectively. Subsequently, we select the inner cluster with the maximum color-domain distinction to the outer clusters as saliency cluster.

Assume that the selected cluster contains Q superpixel units and the tag of the q -th ($q \in 1, 2, \dots, Q$) unit is t_q ($t_q \in 1, 2, \dots, N$). According to Bayes theory, the saliency of foreground superpixel elements can be calculated as

$$S_{fg} = \frac{1}{Q} \left[S_{temp} + \frac{1}{Q-1} \sum_{q=1}^Q \delta(n, t_q) S_{temp} \right] \tag{10}$$

where $\delta(n, t_q)$ is a potential function which measures the similarity between n and t_q and is defined as

$$\delta(n, t_q) = \begin{cases} 1 & n = t_q \\ 0 & n \neq t_q \end{cases} \tag{11}$$

and S_{temp} is defined as

$$S_{temp} = \sum_{q=1, t_q \neq n}^Q \frac{\lambda_n}{d(v_n^{lab}, v_{t_q}^{lab}) + \eta d(v_n^{cen}, v_{t_q}^{cen})} \tag{12}$$

where λ and η are weighting factors used to balance the impact of color distance and spatial distance.

Let $\{C_1, C_2, \dots, C_M\}$ be the salient foreground clusters, We define a weighting function to constrain the foreground clusters as

$$L = \sum_{m=1}^M \alpha_1 \Phi_{ecc} + \sum_{m=1}^M \alpha_2 \Phi_{sym} + \sum_{m=1}^M \alpha_3 \Phi_{sol} \quad (13)$$

where α_1 , α_2 and α_3 are weighting parameters all normalized to (0,1) that adjust the importance of each term. Their values can be obtained through image training. Φ_{ecc} refers to the eccentricity of an ellipse with the same standard second-order central moment as the region. Since the shape of a ship is generally slender, the eccentricity can be regarded as an important shape feature. Φ_{sym} is a measure of regional symmetric regularity, which is defined as

$$\Phi_{sym} = 1 - \frac{|A_{0-l/2} - A_{l/2-l}|}{A} \quad (14)$$

where A is the area of the salient cluster region, l represents the major axis of the above ellipse, $A_{0-l/2}$ and $A_{l/2-l}$ are parts of the regional area on both sides of the major axis. Φ_{sol} is a measure of regional solidity, which is defined as the ratio of the area A to the area of the smallest convex polygon containing the region. These three indexes can effectively distinguish ship from island, cloud and other background elements. Actually, from the training process, we found that symmetry has the most important effect on the final saliency result and set $\alpha_2 = 0.6$ can obtain the best map. α_1 and α_3 are usually both set to 0.2.

Then we calculate the final saliency map via foreground constraint as

$$S_{fc} = S_{fg} \cdot L \quad (15)$$

Figure 3 shows some examples of saliency map results, from which we can see that the maps via background prior are not effective enough to suppress the effects of heavy clouds and islands. However, after the foreground constraint processing, the detection of the salient ship candidates has been significantly improved.

2.3. Fine Selection of Candidate Regions

After obtaining the saliency maps via background prior and via foreground constraint, it is necessary to integrate them and then refine the integrated map for the fine selection of candidate regions. Actually, the background-based saliency map puts emphasis on the significance of all the potential salient areas, while the foreground-based map better eliminates some regions of non-ship candidates. Therefore, we merge them into a unified map according to the following equation:

$$S = S_{fc} \cdot (1 - \exp(-\eta \cdot S_{bp})) \quad (16)$$

where η is a balancing factor. Similar to [28], we set η to 6 in our experiments.

Through the integration step, we can get good saliency maps in most images. However, there are still some flaws in some details, such as the edge of the salient region is not smooth enough and sometimes the wake of the ship is not filtered out. Consequently, we adopt two steps to refine the unified map:

(1) To smooth the salient region edge, we introduce a guided filter function [29], which can significantly smooth the structural edges and suppress most of the image artifacts, such as the abnormal points and gradient reversion. For all of our experiments, the regularization parameter in the function is set to 4.

(2) Considering that there is a gradual change in the brightness of ship wake in the optical image, generally, the wake near the ship is brighter than the wake far away. Moreover, this kind of brightness gradient still appears after the image is divided into superpixels. Although it is difficult to remove the

wake completely in the complex sea scene, we use an adjustment function to reduce the influence of some wakes which is defined as:

$$A(x) = \begin{cases} \frac{x \cdot x_{\max}}{x_{\text{thre}}} & x \geq x_{\text{thre}} \\ x^2 & x < x_{\text{thre}} \end{cases} \quad (17)$$

where x represents the gray brightness of each superpixel and it is normalized to (0,1), x_{\max} and x_{thre} represent the maximum and the threshold values of gray brightness, respectively. In our experiments, we set $x_{\text{thre}} = 0.65 \cdot x_{\max}$.

After the integration and refinement steps, we obtain the final saliency map S_{final} containing more accurate candidate region information.

2.4. Ship Target Discrimination

The final saliency map contains several highlighted regions, providing possible candidate ship positions. The purpose of the discrimination step is to locate all real ships in a given image and place a bounding box around them. To discriminate real ships, we use the advanced spatial-frequency channel feature (SFCF) descriptor in [30], which considers simultaneously the rotation-invariant characteristics in the frequency domain and the useful spatial characteristics such as color and gradient. Compared with the widely used histogram of oriented gradients (HOG) and FourierHOG descriptors, the SFCF descriptor enriches the representations of those features in different channels, thus it is more effective to describe the characteristics (e.g., direction and size) of ships under complex background conditions. The representation of SFCF descriptor can be written as

$$D_{\text{SFCF}} = \left[\Omega_1(I)_{C_1}, \dots, \Omega_1(I)_{C_j}, \dots, \Omega_2(I)_{C_1}, \dots, \right. \\ \left. \Omega_2(I)_{C_j}, \dots, \Omega_3(I)_{C_1}, \dots, \Omega_3(I)_{C_j}, \dots \right] \quad (18)$$

where $\{\Omega_i(I)\}_{i=1}^3$ represents the different feature sets in the color channels, gradient magnitude channels and rotation-invariant channels. $\Omega_i(I)_{C_j}$ represents the region-based features by using the j -th convolution kernel.

With respect to the classifier, we adopt the random forest classifier [31], which estimates the priority of different features efficiently derived from bagging and ensemble learning methods with low computational cost. It was proven to perform the best among 179 classifiers on 121 UCI datasets from different domains [32]. The primary solution of the random forest classifier is to build a set of decision trees by using some randomly selected subsets of SFCF descriptor features and then to perform prediction by taking the majority vote of these decision trees as final results of the ship detection.

3. Experimental Results

In this section, we first make descriptions about the datasets exploited, the baseline approaches and the evaluation indexes. Then, to clarify the superiority of the proposed method, both qualitative and quantitative evaluations are performed and analyzed. Finally, the limitation of the proposed method is discussed. Most of experiments are conducted by using MATLAB 2019b software on a PC with 2.93-GHz Intel i7 CPU. Some experiments involving deep learning network training are run on NVIDIA GeForce GTX TITAN X GPU.

3.1. Dataset Description and Experimental Setup

3.1.1. Datasets

In order to demonstrate the effectiveness and robustness of our method, we conduct comparative experiments on two public datasets: Airbus ship detection challenge dataset (it is available from: <https://www.researchgate.net/publication/325111111>).

[//www.kaggle.com/c/airbus-ship-detection](https://www.kaggle.com/c/airbus-ship-detection)) and MARitime SATellite Imagery (MASATI) dataset [33]. The Airbus dataset is the largest sample set in the ship detection research community. It contains 192556 training images and 15606 testing images whose sizes are 768×768 pixels. It is worth noting that the image scenes of this dataset cover various complex environment elements such as heavy sea clutters, thin and thick clouds, wake and islands, etc. The MARitime SATellite Imagery dataset includes 6212 optical images. About half of the dataset images contain one or multiple ships labeled in various weather and illumination conditions. The average image size is around 512×512 pixels. In the two datasets, the rotated bounding boxes of the ships are given.

3.1.2. Experimental Settings

In the proposed method, obtaining the prediction map derived from the saliency model is the core step and an important premise of ship detection. Therefore, we conduct experiments to assess the performance of our proposed saliency model first and then demonstrate the overall detection performance.

Baseline methods for saliency model performance: We compare the proposed prediction model with seven representative saliency approaches, including robust background weighted contrast (wCtr) [34], saliency filter (SF) [35], manifold ranking (MR) [36], markov absorption probabilities (MAP) [37], background and foreground seed selection (BFS) [24], multiscale CNN features learning (MDF) [38] and multiscale-context saliency detection (MC) [39]. These baseline techniques cover the mainstream and emerging models for predicting regions of salient objects. More precisely, the first five methods are mainly based on the traditional feature extraction, and the last two are mainly based on the deep learning networks. It is noted that the ground truth maps for the two datasets are manually annotated. Three researchers, who mainly study saliency detection, are asked to annotate salient ships and then draw an accurate contour for each ship target.

Baseline methods for overall algorithm performance: To prove the validity of the proposed method, four state-of-the-art ship detection algorithms have been selected as comparison techniques from two categories: approaches based on the hand-designed features and the learning-based approaches. For the former, we considered ship detection via multi-scale analysis (MSA) [40] and ship detection via saliency segmentation and structure-local binary pattern feature (SSS-LBPF) [18]. For the latter, we selected the Faster R-CNN [41] and hierarchical selective filtering network (HSF-Net) [42]. These four advanced methods are good representatives of commonly used approaches to solve the ship detection problem.

3.1.3. Evaluation Criteria

For saliency model performance evaluation, two metrics widely used for saliency detection task are adopted: The precision–recall curve (PR curve) and the F-measure. The PR curve is obtained by binarizing the saliency map with an adaptive threshold varying from 0 to 255. The precision, recall and F-measure are defined as

$$precision = \frac{|S \cap G|}{|S|} \quad (19)$$

$$recall = \frac{|S \cap G|}{|G|} \quad (20)$$

$$F = \frac{(1 + \beta^2) \cdot precision \cdot recall}{\beta^2 \cdot precision + recall} \quad (21)$$

where S and G represent the saliency map and ground truth, respectively, and β^2 is a control parameter used to balance the importance of precision and recall. Similar to [28,35], we set β^2 to 0.3 to emphasize the precision, which is more important in most applications.

For overall ship detection algorithm performance evaluation, in addition to the detection rate, we also consider the index of false alarm rate, which is a concerned indicator in practical detection

applications. Therefore, these two indicators are adopted to evaluate the final performance of ship target discrimination, which are defined as

$$Accuracy = \frac{\text{number of correctly detected ships}}{\text{number of real ships}} \quad (22)$$

$$False\ ratio = \frac{\text{number of detected false alarms}}{\text{number of detected candidates}} \quad (23)$$

3.2. Results

3.2.1. Analysis of the Contributions of Different Steps to Performance

We first carry out experiments on the two datasets to analyze the contribution of each step in our saliency map generation stage. Figure 4a,b show the PR curves obtained on the two datasets. We can observe that after we use the foreground constraint step, the precision is significantly improved compared with that yielded using only the background prior information. Moreover, the precision is also sharply improved after the integration and refinement steps, confirming that the suppression of ship wakes and map smoothness are beneficial to the overall ship detection performance.

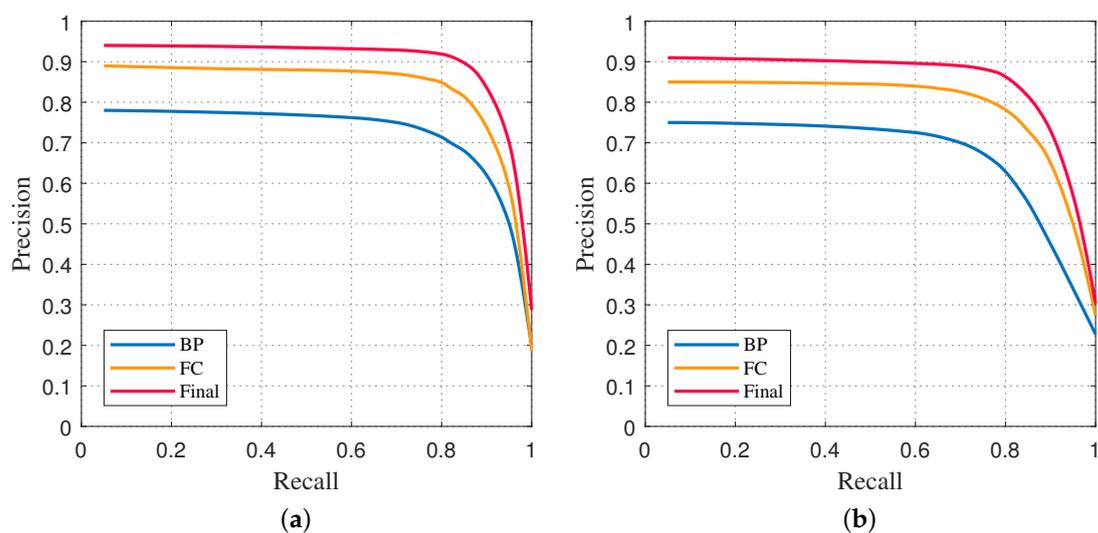


Figure 4. Precision–recall (PR) curves of different steps on the two considered datasets. (a) Airbus dataset and (b) MARitime SATellite Imagery (MASATI) dataset. Note that BP, FC and Final represent background prior, foreground constraint and final refinement steps, respectively (Best viewed in color).

3.2.2. Performance Comparison of Saliency Maps

In this part, we evaluate the performance of the proposed saliency model. For a fair comparison, we add the refinement step to the seven baseline methods. Figure 5 shows some representative examples of saliency maps provided by the seven baseline approaches and by the proposed method. These examples cover thin clouds, thick clouds, heavy sea clutters and strong ship wakes as well as multi-target application scenarios. From the perspective of qualitative visual effect, we can see that the proposed method highlights more uniformly the entire salient candidate parts.

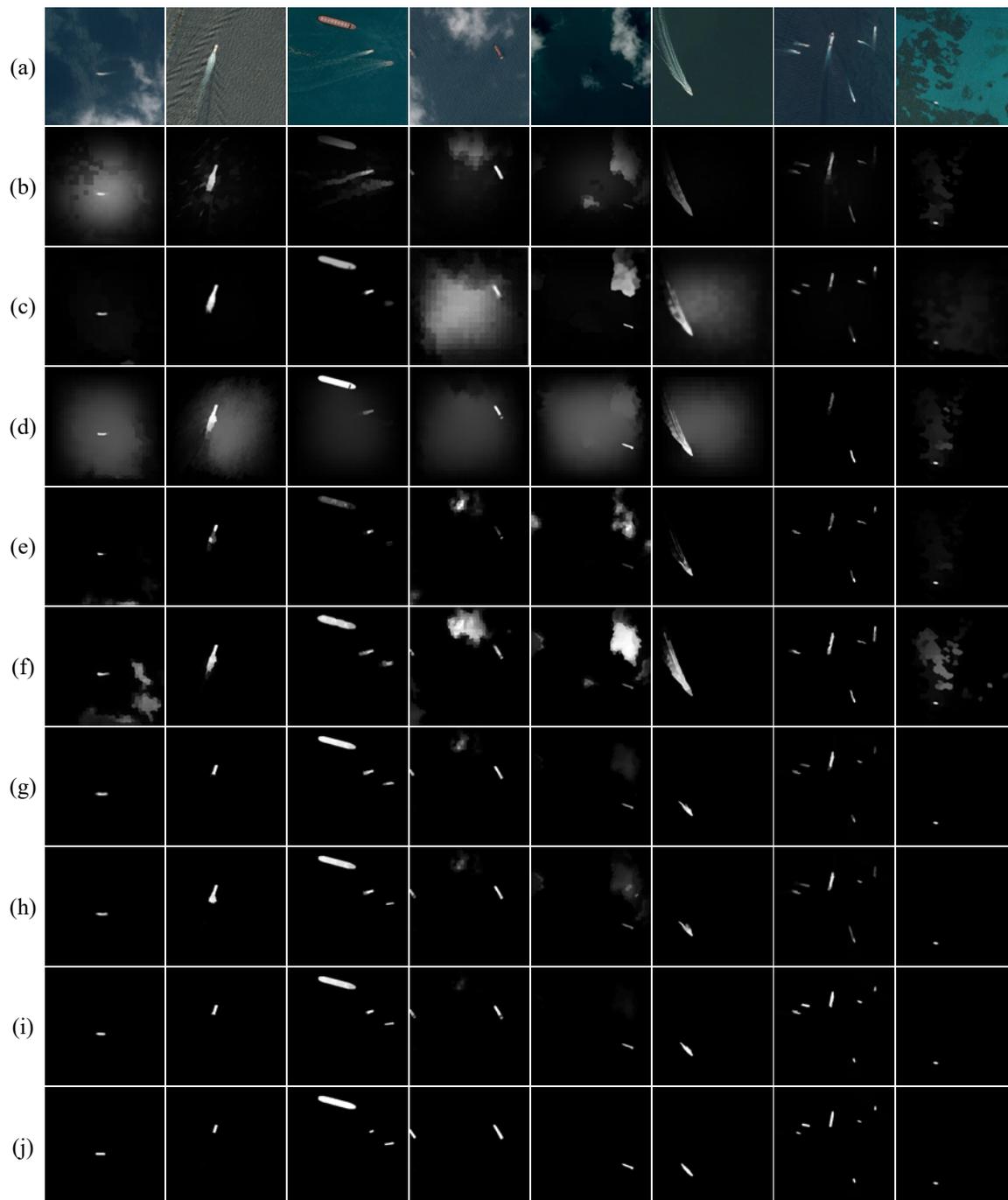


Figure 5. Examples of input images and their corresponding results provided by different saliency methods on the two datasets (Columns 1–5 from Airbus dataset, Columns 6–8 from MASATI dataset): (a) Input image, (b) BFS [24], (c) MAP [37], (d) MR [36], (e) SF [35], (f) wCtr [34], (g) MDF [38], (h) MC [39], (i) proposed method and (j) ground truth.

Specifically, the BFS and MAP methods, which also extract the characteristic information of the image boundary superpixels as background knowledge, can homogeneously emphasize the salient regions, but sometimes fail to suppress the background noise completely. The MR method, which computes the graph-based manifold ranking to measure the foreground and background cues, does not perform well in highlighting small salient regions in the multi-target scenarios, achieving the worst performance. The SF method, which uses contrast-based filter, detects only high-contrast edges and attenuates smoothness of the interior regions. The performance of the wCtr method, which uses the

boundary connectivity to measure the background information, is accurate and stable for simple sea scenes, but it is significantly reduced in the complex environment with more clouds or sea clutter interferences. The two learning-based methods and our method are relatively stable as a whole and can effectively highlight salient regions while suppressing background interferences.

Figures 6 and 7 show the comparisons of the PR curves and F-measure metrics. One can observe from the quantitative evaluations on the two datasets that the performance of the two approaches derived from learning network is obviously better than that of the five traditional baseline methods in term of precision, recall and F-measure metrics. Moreover, the MDF method is slightly better than MC. In addition, compared with the advanced MDF method, the proposed method always achieves higher precision, but may slightly decrease the recall (see Figure 7b). This is because the proposed method puts emphasis on the saliency of detected candidate regions, whereas the MDF method focuses on the coverage of detected regions. More importantly, the F-measure score of the proposed method on the two datasets is higher, indicating that our method obtains a better overall saliency map quality.

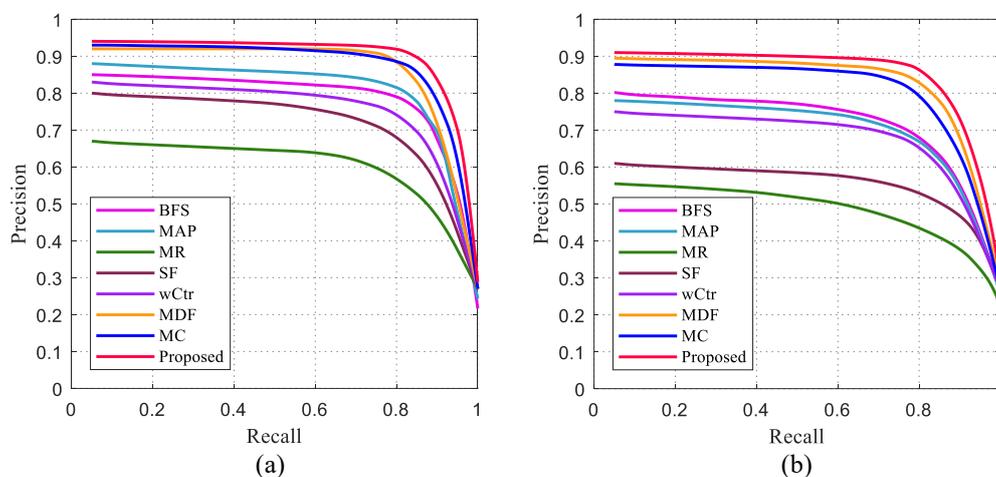


Figure 6. PR curves of saliency maps on the two considered datasets. (a) Airbus dataset and (b) MASATI dataset. (Best viewed in color).

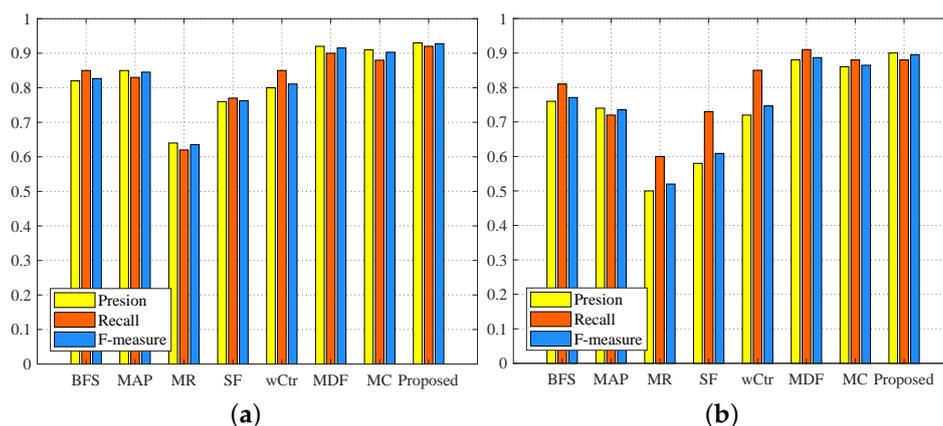


Figure 7. Precision, recall and F-measures of all applied methods on the two considered datasets. (a) Airbus dataset and (b) MASATI dataset. For these indexes, the larger value is better.

3.2.3. Comparison of Overall Detection Results

In this part, we evaluate the overall performance of our proposed method. Figure 8 shows several representative detection instances obtained by the four baseline methods and our method on the two considered datasets. For achieving a clear visualization, the ships marked with the red rectangle bounding boxes are enlarged. The first column of Figure 8 shows some thick clouds in the image

scene and a ship near the edges of clouds. One can see that the MSA, SSS-LBPF and Faster R-CNN methods miss the ship in the shadow, on the contrary, the HSF-Net method mistakenly detects a ship, whereas our method can detect it correctly. The second column of Figure 8 shows a ship with strong wake. All the bounding boxes of the four baseline methods are inaccurate, whereas the proposed method reduces the influence of wakes, obtaining a correct bounding box. The third column of Figure 8 shows a small ship with island and reef interferences. The detection results indicate that all the four compared methods mistakenly detect at least one ship, whereas the proposed method can mitigate these interference factors and provide an accurate result. For heavy sea clutter situations, such as those demonstrated in the fourth column of Figure 8, one can observe that the MSA and Faster R-CNN methods miss a small ship, the SSS-LBPF and HSF-Net methods mistakenly detect an object, whereas our algorithm can better complete the task of ship detection in the cluttered background compared with the other methods.

Table 1 shows the quantitative results obtained by the different methods on the two datasets in terms of accuracy and false ratio. As one can observe, the proposed method obtained highest accuracy and lowest false ratio compared with the other four considered approaches. Among all considered approaches, the MSA approach achieves the highest false ratio. This method, similar to the proposed method, also utilizes a saliency model to primarily select candidate regions. However, it has little consideration on the prior relationship between the foreground and background, leading to relatively high false alarms. On the contrary, our method exploits saliency prior first to remove background clutters and then to constrain the foreground candidate regions, thus obtaining high detection rate and few false alarms. In addition, we use the advanced SFCF descriptor and the random forest classifier to discriminate real ships. This results in further decreasing false alarms caused by objects with similar features to the targets. The SSS-LBPF method yields the worst detection accuracy. This is because it ignores the similarity between some background elements (e.g., island and reef) and the ship target in complex background. Moreover, it is limited by the representation ability of target distinctness. Although the detection performance can be enhanced by deep network modeling, the Faster R-CNN and HSF-Net methods hold slightly lower performance than the proposed technique. This mainly depends on their low robustness to tiny ships and ship wakes.

We further calculate the average time cost of each image for all detection methods, as shown in Table 2. Let us now analyze the average running time of different methods. The average running time of the two methods applying deep learning network are shorter than those of other approaches. This is because the two learning-based methods are run in a GPU environment. When computing resources are limited, these two methods become time demanding due to the complex image training phases. The running time of the MSA and SSS-LBPF methods are longer than that of our method. In addition, the average running time of the proposed method on the two datasets is 1.06 seconds and 0.70 seconds, respectively. This indicates that our method meets the needs of a real-time detection tasks. Consequently, through the presented qualitative and quantitative evaluations, we can draw the conclusion that the proposed method can work accurately and satisfactorily in complex maritime image scenes with clouds, islands, ship wakes and sea clutters interferences, outperforming the other considered state-of-the-art techniques.

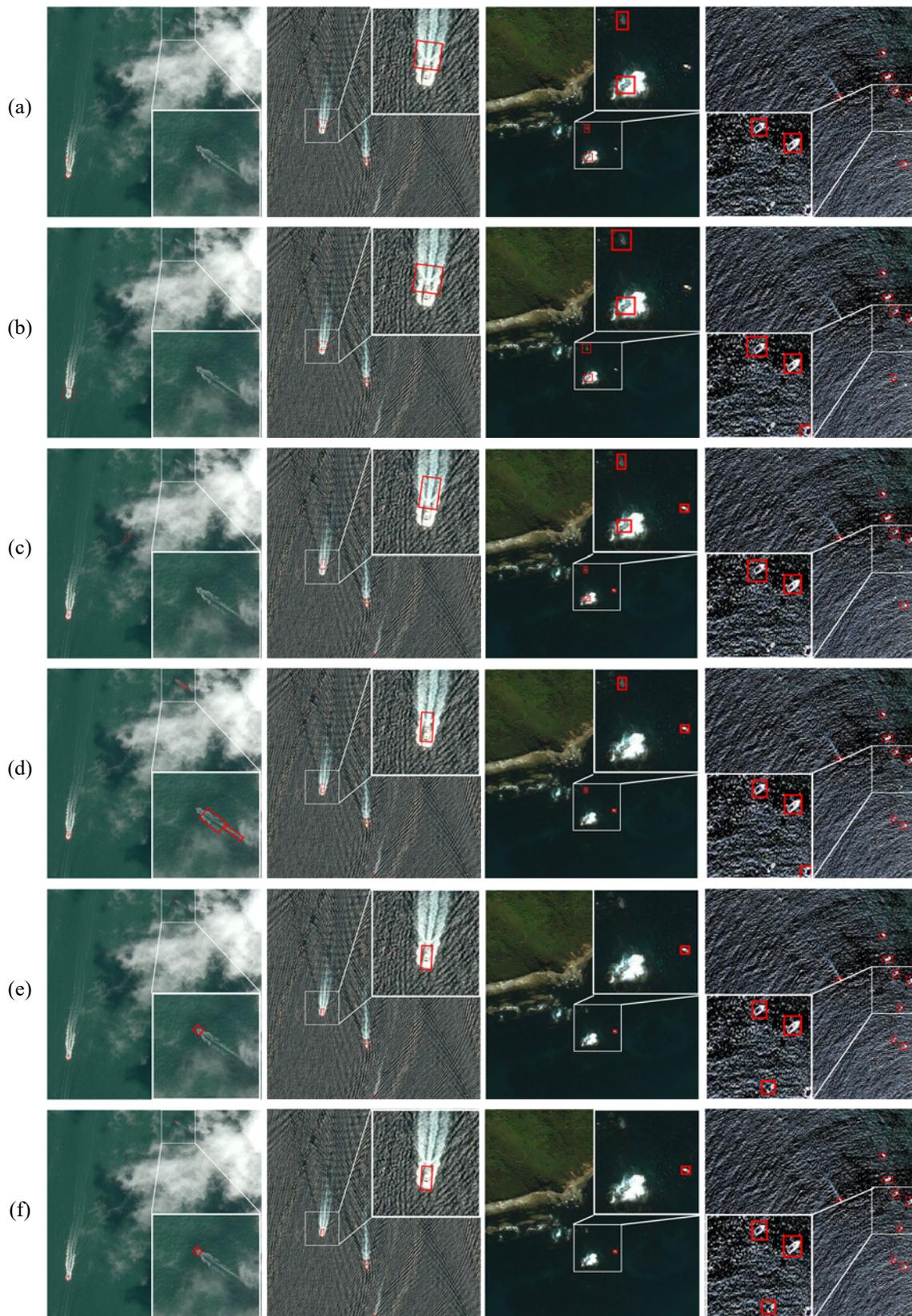


Figure 8. Examples of detection instances of different methods on the two considered datasets (Columns 1 and 2 from Airbus dataset, Columns 3 and 4 from MASATI dataset): (a) MSA, (b) SSS-LBPF, (c) Faster R-CNN, (d) HSF-Net, (e) proposed method and (f) ground truth (best viewed in color).

Table 1. Detection performance evaluation of all methods on the two considered datasets.

Datasets	Metrics	MSA	SSS-LBPF	Faster R-CNN (GPU)	HSF-Net (GPU)	Proposed
Airbus	Accuracy (%)	84.22	79.64	87.40	92.10	93.40
	False ratio (%)	8.92	8.39	8.31	6.52	4.84
MASATI	Accuracy (%)	81.60	78.32	86.55	90.54	92.04
	False ratio (%)	9.40	8.60	8.38	6.75	5.52

Table 2. Average running time (second) of all methods on the two considered datasets.

Datasets	MSA	SSS-LBPF	Faster R-CNN (GPU)	HSF-Net (GPU)	Proposed
Airbus	1.39	1.11	0.16	0.22	1.06
MASATI	1.04	0.81	0.13	0.16	0.70

4. Conclusions

In this paper, we have presented a novel and accurate salient ship detection method via background prior and foreground constraint in optical remote sensing images. First, we analyze the boundary probability of the image border set regardless of the random positions of ships and generate the saliency map containing reliable background prior. Then, we locate the foreground salient regions and propose a weighting strategy to constrain the foreground clusters, obtaining the foreground-driven saliency map. Furthermore, to make full use of the two kind of prediction maps, we merge them into a converged map and improve the details, making fine selection of candidate regions. An efficient discrimination classifier is then used to reduce false alarms on the candidate parts extracted by the final saliency map and the ultimate ship detection results are obtained. The experimental results demonstrate that the proposed method can work effectively and robustly in complicated maritime image scenes with thick clouds, islands, strong wake and heavy sea clutters interfere.

Although the proposed method outperforms most of the exiting algorithms in terms of detection accuracy and false ratio, it mainly focuses on suppressing the interference of complex natural environment factors such as cloud and sea clutter. However, the foreground constraint strategy can be affected by the presence of many artificial structures, thus limiting the performance of our method for the inshore ship detection in complicated harbor scenes. As future development, we will study more sophisticated features and more effective strategies to reduce the false ratio under these complex scene conditions.

Author Contributions: J.H. wrote the manuscript and designed the comparative experiments; X.Z. performed the experimental analysis and revised the manuscript; W.Z. supervised and gave comments to the study; L.R. contributed to the algorithm implementation. L.B. revised the manuscript and gave suggestions to the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China under Grant 61975043 and the China Scholarship Council under grant NO.201906120110.

Acknowledgments: The authors would like to thank the Kaggle competition organizers for providing the Airbus dataset and Antonio-Javier Gallego from the University of Alicante for providing the MASATI dataset.

Conflicts of Interest: The authors declare that there is no conflict of interest.

References

1. Nie, T.; Han, X.; He, B.; Li, X.; Liu, H.; Bi, G. Ship Detection in Panchromatic Optical Remote Sensing Images Based on Visual Saliency and Multi-Dimensional Feature Description. *Remote Sens.* **2020**, *12*, 152. [[CrossRef](#)]
2. Graziano, M.D. Preliminary Results of Ship Detection Technique by Wake Pattern Recognition in SAR Images. *Remote Sens.* **2020**, *12*, 2869. [[CrossRef](#)]
3. Dong, C.; Liu, J.; Xu, F. Ship detection in optical remote sensing images based on saliency and a rotation-invariant descriptor. *Remote Sens.* **2018**, *10*, 400. [[CrossRef](#)]

4. Chen, H.; Gao, T.; Chen, W.; Zhang, Y.; Zhao, J. Contour Refinement and EG-GHT-Based Inshore Ship Detection in Optical Remote Sensing Image. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 8458–8478. [[CrossRef](#)]
5. Qi, S.; Ma, J.; Lin, J.; Li, Y.; Tian, J. Unsupervised ship detection based on saliency and S-HOG descriptor from optical satellite images. *IEEE Geosci. Remote Sens. Lett.* **2015**, *12*, 1451–1455.
6. He, H.; Lin, Y.; Chen, F.; Tai, H.M.; Yin, Z. Inshore ship detection in remote sensing images via weighted pose voting. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 3091–3107. [[CrossRef](#)]
7. Kanjir, U.; Greidanus, H.; Oštir, K. Vessel detection and classification from spaceborne optical images: A literature survey. *Remote Sens. Environ.* **2018**, *207*, 1–26. [[CrossRef](#)]
8. Corbane, C.; Najman, L.; Pecoul, E.; Demagistri, L.; Petit, M. A complete processing chain for ship detection using optical satellite imagery. *Int. J. Remote Sens.* **2010**, *31*, 5837–5854. [[CrossRef](#)]
9. Leng, X.; Ji, K.; Xing, X.; Zhou, S.; Zou, H. Area ratio invariant feature group for ship detection in SAR imagery. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 2376–2388. [[CrossRef](#)]
10. Zhu, C.; Zhou, H.; Wang, R.; Guo, J. A novel hierarchical method of ship detection from spaceborne optical image based on shape and texture features. *IEEE Trans. Geosci. Remote Sens.* **2010**, *48*, 3446–3456. [[CrossRef](#)]
11. Graziano, M.D.; Renga, A.; Moccia, A. Integration of Automatic Identification System (AIS) Data and Single-Channel Synthetic Aperture Radar (SAR) Images by SAR-Based Ship Velocity Estimation for Maritime Situational Awareness. *Remote Sens.* **2019**, *11*, 2196. [[CrossRef](#)]
12. Liu, G.; Zhang, Y.; Zheng, X.; Sun, X.; Fu, K.; Wang, H. A new method on inshore ship detection in high-resolution satellite images using shape and context information. *IEEE Geosci. Remote Sens. Lett.* **2013**, *11*, 617–621. [[CrossRef](#)]
13. Chen, L.; Shi, W.; Fan, C.; Zou, L.; Deng, D. A Novel Coarse-to-Fine Method of Ship Detection in Optical Remote Sensing Images Based on a Deep Residual Dense Network. *Remote Sens.* **2020**, *12*, 3115. [[CrossRef](#)]
14. Yang, X.; Sun, H.; Fu, K.; Yang, J.; Sun, X.; Yan, M.; Guo, Z. Automatic ship detection in remote sensing images from google earth of complex scenes based on multiscale rotation dense feature pyramid networks. *Remote Sens.* **2018**, *10*, 132. [[CrossRef](#)]
15. Zou, Z.; Shi, Z. Ship detection in spaceborne optical image with SVD networks. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 5832–5845. [[CrossRef](#)]
16. Liu, W.; Ma, L.; Chen, H. Arbitrary-oriented ship detection framework in optical remote-sensing images. *IEEE Geosci. Remote Sens. Lett.* **2018**, *15*, 937–941. [[CrossRef](#)]
17. Rostami, M.; Kolouri, S.; Eaton, E.; Kim, K. Deep transfer learning for few-shot sar image classification. *Remote Sens.* **2019**, *11*, 1374. [[CrossRef](#)]
18. Yang, F.; Xu, Q.; Li, B. Ship detection from optical satellite images based on saliency segmentation and structure-LBP feature. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 602–606. [[CrossRef](#)]
19. Xu, F.; Liu, J.; Dong, C.; Wang, X. Ship detection in optical remote sensing images based on wavelet transform and multi-level false alarm identification. *Remote Sens.* **2017**, *9*, 985. [[CrossRef](#)]
20. Li, L.; Zhou, F.; Zheng, Y.; Bai, X. Saliency detection based on foreground appearance and background-prior. *Neurocomputing* **2018**, *301*, 46–61. [[CrossRef](#)]
21. Xia, C.; Zhang, H.; Gao, X. Combining multi-layer integration algorithm with background prior and label propagation for saliency detection. *J. Vis. Commun. Image Represent.* **2017**, *48*, 110–121. [[CrossRef](#)]
22. Yang, C.; Zhang, L.; Lu, H. Graph-regularized saliency detection with convex-hull-based center prior. *IEEE Signal Process. Lett.* **2013**, *20*, 637–640. [[CrossRef](#)]
23. Tang, W.; Wang, Z.; Zhai, J.; Yang, Z. Salient object detection via two-stage absorbing Markov chain based on background and foreground. *J. Vis. Commun. Image Represent.* **2020**, *71*, 102727. [[CrossRef](#)]
24. Wang, J.; Lu, H.; Li, X.; Tong, N.; Liu, W. Saliency detection via background and foreground seed selection. *Neurocomputing* **2015**, *152*, 359–368. [[CrossRef](#)]
25. Arbelaez, P.; Maire, M.; Fowlkes, C.; Malik, J. Contour detection and hierarchical image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2010**, *33*, 898–916. [[CrossRef](#)]
26. Otsu, N. A threshold selection method from gray-level histograms. *IEEE Trans. Syst. Man Cybern. Syst.* **1979**, *9*, 62–66. [[CrossRef](#)]
27. Xie, Y.; Lu, H.; Yang, M.H. Bayesian saliency via low and mid level cues. *IEEE Trans. Image Process.* **2012**, *22*, 1689–1698.
28. Wang, Z.; Xiang, D.; Hou, S.; Wu, F. Background-driven salient object detection. *IEEE Trans. Multimedia* **2016**, *19*, 750–762. [[CrossRef](#)]

29. Ham, B.; Cho, M.; Ponce, J. Robust guided image filtering using nonconvex potentials. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *40*, 192–207. [[CrossRef](#)]
30. Wu, X.; Hong, D.; Tian, J.; Chanussot, J.; Li, W.; Tao, R. ORSIIm Detector: A novel object detection framework in optical remote sensing imagery using spatial-frequency channel features. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 5146–5158. [[CrossRef](#)]
31. Pal, M. Random forest classifier for remote sensing classification. *Int. J. Remote Sens.* **2005**, *26*, 217–222. [[CrossRef](#)]
32. Fernández-Delgado, M.; Cernadas, E.; Barro, S.; Amorim, D. Do we need hundreds of classifiers to solve real world classification problems? *J. Mach. Learn. Res.* **2014**, *15*, 3133–3181.
33. Gallego, A.J.; Pertusa, A.; Gil, P. Automatic ship classification from optical aerial images with convolutional neural networks. *Remote Sens.* **2018**, *10*, 511. [[CrossRef](#)]
34. Zhu, W.; Liang, S.; Wei, Y.; Sun, J. Saliency optimization from robust background detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 2814–2821.
35. Perazzi, F.; Krähenbühl, P.; Pritch, Y.; Hornung, A. Saliency filters: Contrast based filtering for salient region detection. In Proceedings of the 2012 IEEE conference on computer vision and pattern recognition, Providence, RI, USA, 16–21 June 2012; pp. 733–740.
36. Yang, C.; Zhang, L.; Lu, H.; Ruan, X.; Yang, M.H. Saliency detection via graph-based manifold ranking. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR, USA, 23–28 June 2013; pp. 3166–3173.
37. Sun, J.; Lu, H.; Liu, X. Saliency region detection based on Markov absorption probabilities. *IEEE Trans. Image Process.* **2015**, *24*, 1639–1649. [[CrossRef](#)] [[PubMed](#)]
38. Li, G.; Yu, Y. Visual saliency detection based on multiscale deep CNN features. *IEEE Trans. Image Process.* **2016**, *25*, 5012–5024. [[CrossRef](#)]
39. Zhao, R.; Ouyang, W.; Li, H.; Wang, X. Saliency detection by multi-context deep learning. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 1265–1274.
40. Dong, C.; Liu, J.; Xu, F.; Liu, C. Ship Detection from Optical Remote Sensing Images Using Multi-Scale Analysis and Fourier HOG Descriptor. *Remote Sens.* **2019**, *11*, 1529. [[CrossRef](#)]
41. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster r-cnn: Towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 1137–1149. [[CrossRef](#)]
42. Li, Q.; Mou, L.; Liu, Q.; Wang, Y.; Zhu, X.X. HSF-Net: Multiscale deep feature embedding for ship detection in optical remote sensing imagery. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 7147–7161. [[CrossRef](#)]

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).